UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

EVALUATION OF FLASH DROUGHT IDENTIFICATION WITH MACHINE LEARNING TECHNIQUES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

STUART EDRIS
Norman, Oklahoma
2024

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340537

EVALUATION OF FLASH DROUGHT IDENTIFICATION WITH MACHINE LEARNING TECHNIQUES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Amy McGovern, Chair

Dr. Jeffrey Basara

Dr. Jason Furtado

Dr. Xiangming Xiao

Dr. Jordan Christian

© Copyright by STUART EDRIS 2024
All Rights Reserved.

Acknowledgements

I would like to give thanks to and acknowledgements to my committee members for their advise, assistance, and feedback. In particular, I would like to give thanks to Dr. Jeff Basara and Dr. Amy McGovern. Dr. Basara has acted as my advisor and giving advise since my undergraduate career and saw me through my master's thesis. He has also offered ideas, feedback, and advice for much of this project until, and even after, he moved half way across the country. Dr. McGovern graciously took over as advisor after Dr. Basara left the university, and brought me into the IDEA Lab research group, and offered regular council, helping to push this project past the finish line. My gratitude also goes to Dr. Christian for helping with the data collection and for helping get some of the FD identification methods to work, and for verifying some of them. I also would like to give thanks to the rest of the committee for their feedback and assistance given on the project and manuscripts for this project. Thanks is also given to friends and family who have offered their support and encouragement throughout this project.

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 No. ICER-2019758. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation or the USDA Southern Great Plains Climate Hub.

Table of Contents

chapterAcknowledgementsiv

List Of Tables **vii**

List Of Figures **viii**

Abstract **xvii**

1 Introduction **1**

2 Background **6**
2.1 Traditional Drought 6
2.2 Machine Learning in Drought 10
2.3 Flash Drought 14
2.4 Machine Learning and Flash Drought 22

3 Datasets, Machine Learning Algorithms, and Statistical Methods **25**
3.1 Datasets 25
3.1.1 North American Regional Reanalysis 25
3.1.2 European Centre Atmospheric Reanalaysis v5 27
3.2 Input Data 28
3.3 Output Data 29
3.4 Machine Learning Algorithms 31
3.4.1 Standard Machine Learning Algorithms 33
3.4.2 Deep Learning Algorithms 34
3.5 Machine Learning Verification and Interpretation 42

4 Standard Machine Learning Performance **45**
4.1 Statistical Performance 45
4.2 Predicted Climatology 58
4.3 Case Studies 62

5 Deep Learning Performance **84**
5.1 Statistical Performance 84
5.2 Predicted Climatology 93
5.3 Case Studies 97

6 Machine Learning on Global Flash Drought **117**
6.1 True Climatologies 117
6.2 Bulk Statistics 118
6.3 Climatology Predictions 124
6.4 Case Studies 140

7 Conclusions **156**

List Of Tables

3.1 Class weights applied to each ML algorithm for identifying FD in each identification method. Non-FD labels were given a class weight of 1, and sea values were a class weight of 0. Non-parenthetical values denote class weights applied for the NARR dataset. Parenthetical values denote class weights applied to the global ERA5 dataset. 34

3.2 Miscellaneous parameters that were applied to each neural network. These parameter values did not vary from network to network (except dropout), nor from one FD identification method to another. λ_2 denotes the L2 regularization parameter. Activation functions were tanh for LSTM layers and elu for dense and convolution layers (except for the last convolution layer, which had tanh to prevent gradient explosion). . 35

3.3 Sample weights applied to each neural network for identifying FD in each identification method. Sample weights were applied to all positive FD labels equally to mimic the effect of class weights. Negatives FD labels were given sample weights of 1, and sea values were a sample weights of 0. Non-parenthetical values denote sample weights applied for the NARR dataset. Parenthetical values denote sample weights applied to the global ERA5 dataset. 35

4.1 True skill statistic over all grid points and pentads for each standard ML algorithm and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for is identification method is bolded. 47

5.1 True skill statistic over all grid points and pentads for each DL algorithm and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for is identification method is bolded. 85

6.1 True skill statistic over all grid points and pentads for Ada boosted trees and RNNs across the globe and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for an identification method is bolded. 123

List Of Figures

2.1 Illustration of different types of drought and levels of impacts. [Figure 2 in Dikshit et al. (2022a).] 7
2.2 Illustration of hydrologic and statistical forecasting of drought events, and the steps involved. Hybrid forecasting involves a combination of these two approaches. [Figure 5 in Hao et al. (2018).] 11
2.3 Illustration of the surface interactions that create FD events. [Figure 2a in Tyagi et al. (2022).] 16
2.4 Illustrations of FD identification methods for the (a) Christian et al. (2019b, 2023), (b) Liu et al. (2020b), and (c) Noguera et al. (2020) methods. Note the methods cluster around identifying rapid intensification, and determining if the variable is below a threshold at the end of the intensification period. [(a) Figure 2 in Christian et al. (2019b), (b) Figure 2b in Liu et al. (2020b), and (c) Figure 1c in Noguera et al. (2020).] 20

3.1 Experiment design for this project. The steps for data preparation can be broken down into data collection, preprocessing (exact nature of preprocessing depends on the dataset), splitting into training, validation, and testing folds, and standardization. 26
3.2 Neural network architectures used to identify FD. a) artificial neural network architecture, b) recurrent neural network architecture. Spatial dimensions (N_{space}) for the NARR grid was 20,580 (210 $\times$ 98). ERA5 spatial dimensions without ocean grid points was 266,723. Time dimension (N_{time}) was 43 pentads per fold. 36
3.3 U-network architecture used to identify FD. Note the U-network has spatial dropout of 0.2 applied after every convolutional layer (not shown). Spatial dimensions (N_{space}) for the NARR grid was 210 $\times$ 98. Time dimension (N_{time}) was 43 pentads per fold. 37

4.1 Percentage of years from 1971 – 2021 in which FD occurred according to the a) C23 method, b) N20 method, c) P20 method, d) L20 method, and e) O21 method. 46
4.2 Composite mean difference of true minus predicted FD labels for each FD identification method using Ada boosted tree predictions. 48
4.3 95% statistical significance of composite mean differences between the true FD labels of the Ada boosted tree predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 49
4.4 Composite mean difference of true minus predicted FD labels for each FD identification method using RF predictions. 50

4.5 95% statistical significance of composite mean differences between the true FD labels of the RF predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 51
4.6 Composite mean difference of true minus predicted FD labels for each FD identification method using SVM predictions. 52
4.7 95% statistical significance of composite mean differences between the true FD labels of the SVM predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 53
4.8 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 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations). 55
4.9 Average feature importance for the SVM models according to the GINI method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations). 56
4.10 Average feature importance for the RF models according to the GINI method (left), Shapely values (center). Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations). . . . 57
4.11 Predicted frequency climatology of FDs by the Ada boosted trees for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 59
4.12 Predicted frequency climatology of FDs by the RFs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 60
4.13 Predicted frequency climatology of FDs by the SVMs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 61
4.14 FDs 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. 63
4.15 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the Ada boosted trees. 64
4.16 FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the RFs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 65
4.17 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the RFs. 66

4.18 FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the SVMs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 67
4.19 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the SVMs. 68
4.20 FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 70
4.21 FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs. 71
4.22 FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs. 72
4.23 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 1988 case study. 73
4.24 FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 75
4.25 FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs. 76
4.26 FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs. 77
4.27 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2011 case study. 78
4.28 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 80

4.29 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs. 81
4.30 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs. 82
4.31 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2012 case study. 83

5.1 Average feature importance for the ANNs according to the SHAP method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations). 86
5.2 Composite mean difference of true minus predicted FD labels for each FD identification method using RNN predictions. 87
5.3 95% statistical significance of composite mean differences between the true FD labels of the RNN predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 88
5.4 Composite mean difference of true minus predicted FD labels for each FD identification method using U-net predictions. 89
5.5 95% statistical significance of composite mean differences between the true FD labels of the U-net predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 90
5.6 Composite mean difference of true minus predicted FD labels for each FD identification method using ANN predictions. 91
5.7 95% statistical significance of composite mean differences between the true FD labels of the ANN predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000). 92
5.8 Predicted frequency climatology of FDs by the RNNs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 94
5.9 Predicted frequency climatology of FDs by the U-nets for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 95
5.10 Predicted frequency climatology of FDs by the ANNs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 96
5.11 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. 98

5.12 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the RNNs. 99
5.13 FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the U-nets (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 100
5.14 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the U-nets. 101
5.15 FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the ANNs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. 102
5.16 True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the ANNs. 103
5.17 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 105
5.18 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets. 106
5.19 FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs. 107
5.20 FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 109
5.21 FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets. 110
5.22 FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs. 111

5.23 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2003 case study. 112
5.24 FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 113
5.25 FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets. 114
5.26 FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs. 115
5.27 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2017 case study. 116

6.1 Percentage of years from 1971 – 2021 in which FD occurred across the globe according to the a) C23 method, b) N20 method, c) P20 method, d) L20 method, and e) O21 method. 119
6.2 Composite mean difference of true minus predicted FD labels for each FD identification method using global RNN predictions. 121
6.3 Composite mean difference of true minus predicted FD labels for each FD identification method using global Ada boosted tree predictions. . . 122
6.4 Average feature importance for the global 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 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations). 125
6.5 Global FD frequency climatology predictions from the RNNs. 126
6.6 Global FD frequency climatology predictions from the Ada boosted trees. 126
6.7 Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the C23 method, and true C23 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the C23 method. 130
6.8 Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the N20 method, and true N20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by Ada boosted trees for the N20 method. . . . 131

6.9 Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the P20 method, and true P20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the P20 method. 132
6.10 Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the L20 method, and true L20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by Ada boosted trees for the L20 method. 133
6.11 Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the O21 method, and true O21 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the O21 method. 134
6.12 Average monthly FD coverage (i.e., FD seasonality) for the C23 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the C23 method. 135
6.13 Average monthly FD coverage (i.e., FD seasonality) for the N20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by Ada boosted trees for the N20 method. 136
6.14 Average monthly FD coverage (i.e., FD seasonality) for the P20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the P20 method. 137
6.15 Average monthly FD coverage (i.e., FD seasonality) for the L20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by Ada boosted trees for the L20 method. 138

6.16 Average monthly FD coverage (i.e., FD seasonality) for the O21 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the O21 method. 139
6.17 FD case study for 2001 in India for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 141
6.18 FD case study for 2001 in India for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 142
6.19 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2001 case study over India. 143
6.20 FD case study for 2010 in southern Russia for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 145
6.21 FD case study for 2010 in southern Russia for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 146
6.22 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2010 case study over southern Russia. 147
6.23 FD case study for 2015 – 2016 in the Amazon for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 149
6.24 FD case study for 2015 – 2016 in the Amazon for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 150
6.25 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2015 – 2016 case study over the Amazon. 151
6.26 FD case study for 2016 in eastern Africa for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs. 152

6.27 FD case study for 2016 in eastern Africa for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees. 153
6.28 Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2016 case study over eastern Africa. 154

Abstract

Droughts are extreme dry events that decrease an ecosystem's and society's availability of water resources, leading to impacts on vegetation health and agricultural production and food shortages. Of particular note are droughts that develop on a more rapid timescale (about 1 month), termed flash droughts. Flash droughts have gained increasing attention in the past decade, because they can result in more rapid desiccation, or deterioration in crop health, than what would normally be expected. Research into flash drought events have found certain key variables, such as soil moisture and evaporation from the soil and plants, and potential evaporation are among the key variables driving flash drought events. Varying approaches have resulted in the creation of multiple methods for identifying and quantifying flash droughts, each using different variables and thresholds (for the rapid intensification) to define them.

Machine learning (ML) techniques have been growing in popularity in the environmental sciences due to their ability to accurately represent different environmental phenomena, including drought. However, the use of ML for rapidly developing droughts remains largely unexplored. Thus, this dissertation aims to investigate the ability of various ML techniques to identify flash drought phenomena. Because ML use in flash drought is largely unexplored, this dissertation explores how multiple ML algorithms, such as random forests, support vector machines, several deep learning methods (e.g., several types of artificial neural networks), can identify flash drought events. The ML algorithms were trained on key variables known to drive flash drought events – soil moisture, evaporation, potential evaporation, temperature, precipitation, and the change in soil moisture, evaporation, and potential evaporation. Lastly, feature importance (the importance of each variable to the ML algorithms) was determined from Shapely values and permutation importance methods to give the ML algorithms interpretability and explainability. Results showed ML is capable of representing flash drought events, with boosted trees and recurrent neural networks showing the most

skill. However, the ML algorithms thought flash droughts were more active in the hotspot regions, and the more likely in the late growing season than observations actually show. Feature importance in the ML algorithms showed that the algorithms were relying heavily on soil moisture, precipitation, and potential evaporation to predict flash drought, explaining why they over-emphasized the seasonality. Global representation of flash drought were also be investigated to determine how well the ML models generalize and represent flash drought over vastly varying ecosystems, and to examine rarely examined flash drought events. Global results showed that ML models trained at local scales were significantly more skillful than at the global scale. The skill ML has shown can allow us to not only push our understanding of flash droughts forward, but can also help represent flash droughts in numerical models and allow us to identify flash droughts in real time, and potentially lead to flash drought predictions.

Chapter 1

Introduction

Droughts are extreme dry events that decrease an ecosystem's and society's availability of water resources, leading to impacts on vegetation health and most primarily agricultural loss and food shortages, though dry conditions can also increase the likelihood for fires, and lead to or exacerbate heatwaves. In extreme cases, such as in eastern Africa in the mid 2010s and early 2020s, strong drought can cause famine or exacerbate humanitarian crises as a result of the water deficit (Ribeiro et al. 2021; Dikshit et al. 2022a; Lesk et al. 2022; Krishnamurthy R et al. 2022; Runde et al. 2022; Palmer et al. 2023). In historical records, disappearing civilizations, such as the Ancestral Pueblo civilization in the southwestern United States, has been attributed to strong drought events (Cook et al. 2016). As a result, being able to understand, identify, and predict drought events becomes essential in being able to adapt to and mitigate drought impacts.

Of particular note are droughts that develop on a more rapid timescale (about 1 month), termed flash droughts. Flash droughts have gained increasing attention in the past decade as drought events that develop much faster than expected, and which can also result in more notable desiccation, or deterioration in crop health, due to the rapid water loss. For example, the 2012 flash drought in the central United States had impacts up to $17 billion due to agricultural losses (Otkin et al. 2016). In addition, with future projections showing flash droughts becoming more common in a warming climate, potentially even the new normal for droughts (Christian et al. 2021; Shah et al.

2022; Yuan et al. 2023), more research has started to focus on flash drought events. This research has found certain key variables, such as soil moisture, evaporation from the soils and plants, and potential evaporation are among the key variables driving flash drought events (Otkin et al. 2018; Christian et al. 2019b; Chen et al. 2019; Otkin et al. 2021; Tyagi et al. 2022).

In addition, research on flash drought has started to approach a consensus on what it takes to define or quantify a flash drought event (a period of rapid intensification, in which a variable undergoes rapid change from normal conditions to drought conditions in about 1 month; e.g., Christian et al. 2019b, 2023; Noguera et al. 2020; Liu et al. 2020b; Otkin et al. 2021). However, varying approaches have resulted in the creation of multiple methods for identifying and quantifying flash droughts, each using different variables and thresholds (for the rapid intensification) to define them. These different flash drought identification methods each give different prerogatives, identifying different hotspots (regions where flash drought occurs more frequently) and different seasonalities (months or season in which flash drought is more likely) relating to the variable used to identify flash drought.

Alongside flash drought research, research into machine learning applications in environmental sciences has also been growing in popularity and use. This is because of machine learning's ability to accurately represent and predict different environmental phenomena. In particular, machine learning has proved useful in being able to represent drought events, and some machine learning predictions for drought have even outperformed some of the existing methods for drought prediction (Rhee and Im 2017; Rahmati et al. 2020). This representation of drought with machine learning has further improved as scientists have started to explore more complicated machine learning tools such as different types of artificial neural networks (Dikshit et al. 2022b; Hsieh 2022).

However, the use of applying machine learning to rapidly developing droughts remains largely unexplored Dikshit et al. (2022b); Tyagi et al. (2022).

This dissertation contributions are the application of various machine learning techniques towards predicting flash droughts, and investigating their performance and which technique is best performing. Because there is not yet any consistent definition of flash drought and to give an unbiased approach, this dissertation explores the ability of machine learning algorithms to predict flash drought according to various different flash drought identification methods or definitions. In addition, because machine learning use in flash drought is largely unexplored, this dissertation also explores how multiple different machine learning algorithms, such as random forests, support vector machines, artificial neural networks, recurrent neural networks, and U-shape convolutional networks, can represent flash drought events. Variables used to train the machine learning algorithms were key variables known to drive flash drought events – soil moisture, evaporation, and potential evaporation, the changes in these variables, and temperature and precipitation. The machine learning models were evaluated using several different approaches; by examining the bulk statistics, by examining the climatological predictions of the ML models, and by examining the local predictions of the ML models (i.e., the ML predictions of specific case studies). Lastly, feature importance (the importance of each variable to the machine learning algorithms) and feature attribution (the contribution of each variable to the machine learning algorithms) was determined from Shapely values, permutation importance, and GINI (for tree based models) methods to give the machine learning algorithms interpretability and explainability.

The first two parts of this dissertation used a regional reanalysis over the contiguous United States to investigate flash drought representation over a single country. It was hypothesized that machine learning models could accurately represent FD based on

their success with traditional droughts. And the results showed machine learning is capable of representing flash drought events, with boosted trees, and recurrent neural networks showing the most skill. However, the machine learning algorithms tended to over emphasize hotspots and seasonalities (i.e., the machine learning algorithms thought flash droughts were more active in the hotspot regions and they think flash droughts are more likely on their favorable months than observations actually show). Feature importance in the machine learning algorithms showed that the algorithms were over reliant on only one or two variables (typically soil moisture and precipitation), and did not incorporate the full suite of variables to learn surface interactions. This can result in some cases, such as with the Ada boosted trees and random forests, where the model seems to learn more about the climatology of a single variable as opposed to flash drought patterns, and explains why certain summer months were over emphasized.

The final part of this project incorporated the global ERA5 dataset to investigate how well the machine learning models generalize to the globe and represent flash drought over vastly varying ecosystems. Therefore, in addition to investigating global machine learning performance, this project also builds on Noguera et al. (2020), Pendergrass et al. (2020), Liu et al. (2020b), and Otkin et al. (2021) by examining global climatology for these identification methods. It was hypothesized in this final part that machine learning models should be able to generalize flash drought outward if they did well in the United States (given the highly heterogeneous climate regimes within the United States), and help deliver flash drought predictions to various parts of the globe. Differing levels of impact and human response to droughts and flash droughts were also expected from the case studies based on recent events (e.g., food crises in eastern Africa versus drought response in the United States).

The results from these showed, however, that both RNNs and boosted tree algorithms struggled more than expected, and focused on predicting flash drought for

specific parts of the globe, such as the tropics, Mid-Latitudes, or high latitudes depending on identification method being predicted, than finding patterns across the whole globe. In this case, the results showed that for some machine learning algorithms, training at the local level may be necessary for accurate predictions.

Specific case studies across the globe also showed a secondary element to drought and flash drought impacts that is rarely discussed in the literature – human response. Societies that have more safeguards to guard against droughts and societies that respond to them (such as limiting water use, or being able to increase food imports from other locations) tend to be more weakly impacted by droughts and flash droughts. In contrast, societies at greatest risk to drought and flash drought impacts are those that do not have as many safeguards against these events, or have governments that respond poorly (e.g., by withholding relief supplies). Overall, this dissertation takes some of the first steps in incorporating machine learning to flash drought identification and prediction. The skill of machine learning has shown can allow us to not only push our understanding of flash droughts forward, but can also help represent flash drought more accurately in numerical models and allow us to identify flash droughts in real time, and potentially to predict them (two facets that are beyond our current ability).

Chapter 2

Background

The topics of machine learning (ML) and flash drought (FD) have, with only two known exceptions (Zhang et al. 2022; Foroumandi et al. 2024), not been put together in the literature. The little cross section between the two topics can result in many people know aspects of one topic, but not the other. In order to ensure all aspects of the dissertation are understandable to any reader, this chapter gives a brief overview of drought and how research on it has evolved, of ML and how it has been applied to drought monitoring and prediction, and of FD and how research on that topic has progressed. Lastly, this chapter concludes with an overview of the project goals, research questions, and brief overview on how the research was conducted.

2.1 Traditional Drought

Droughts describe a dry extreme in moisture available to the environment to use. Drought is typically divided into four separate categories, each describing different forms of impacts and types of moisture shortages (Fig. 2.1; Hao et al. 2018; AghaKouchak et al. 2022; Dikshit et al. 2022a). The first level is meteorological drought, which stems from a precipitation deficit (and which can end quickly via precipitation events). Second is agricultural drought, which is due to a soil moisture deficit (this can be due to a long enough precipitation shortage that the soil moisture starts to dry out). As the name implies, agricultural drought is when agricultural output is impacted by

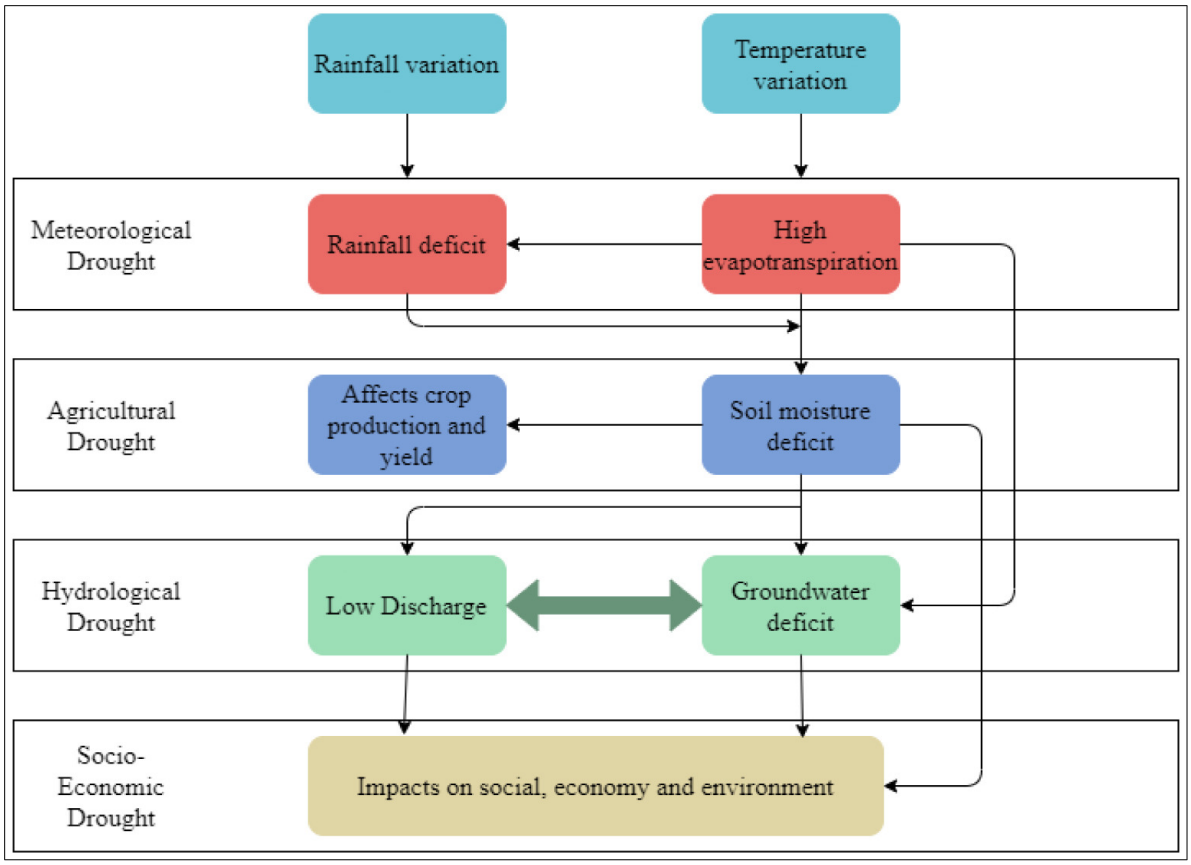


Figure 2.1: Illustration of different types of drought and levels of impacts. [Figure 2 in Dikshit et al. (2022a).]

the water shortage. This is hydrologic drought, resulting from a decreased groundwater, and has impacts on streamflow, runoff, and reservoirs. Lastly, socio-economic droughts describe droughts that impacts the supply and demand of water, food, and other such commodities, and are known to have a more profound impact on the society due to the water loss.

Social impacts of drought have led to numerous studies on the phenomenon and different ways in which to examine and identify them. As a note, there is no consensus on a precise definition for drought, particularly when it comes to quantifying the

events, and definitions can vary, each using different metrics (or indices) and thresholds (Hao et al. 2018; AghaKouchak et al. 2022; Dikshit et al. 2022a). One of the early studies of drought derived an index by which drought could be investigated, termed the Palmer Drought Severity Index (PDSI; Palmer 1965; Alley 1984). The PDSI incorporates precipitation and potential evaporation to estimate the entire water budget and determine the moisture availability in the atmosphere, which is then used to derive the index value. The use of the whole water budget has helped to make the PDSI one of the most popular metrics for identifying and monitoring drought, however the PDSI has a few disadvantages. For example, the PDSI can move into extreme values more easily than expected, it responds slowly to change, and its value does not transfer well from one region to another (e.g., a PDSI of -1 in the southwestern United States may not have the same physical meaning as a PDSI of -1 in the southeast of the United States). These were addressed with the calibrated PDSI outlined in Wells et al. (2004), which normalized moisture anomalies with moisture percentiles and allowed more variable duration factors. However, this last point (retaining the meaning of a drought index from one location to another) has led to the use of standardized indices – drought indicators, such as precipitation, transformed from their natural statistical distribution into a standard normal distribution.

While numerous drought metrics or indices have since been published (on the order of tens to hundreds; Zargar et al. 2011), the most popular index has been the Standardized Precipitation Index (SPI; Guttman 1999), and the Standardized Precipitation Evaporation Index (SPEI; Vicente-Serrano et al. 2010). The SPI is simply the precipitation transformed onto a standard normal distribution (so it changes rapidly with precipitation and describes the atmospheric moisture supply). The SPEI is the precipitation minus potential evaporation, transformed onto a standard normal distribution (thus describing the atmospheric moisture available). As a note, while the

different types of drought assume shortage of different types of moisture (precipitation, soil moisture, and groundwater), most drought studies are conducted via precipitation and potential evaporation, though many studies in more recent times have begun to incorporate soil moisture into their analysis, also resulting in the creation in soil moisture indices (e.g., Hunt et al. 2009; Cook et al. 2018; Baek et al. 2019; Kim and Raible 2021 and Stager et al. 2021 to name a few). Further, different types of drought are examined by examining drought variables over different time scales (e.g., precipitation with a 60-month running mean may be used to get the 60-month SPI and to investigate annual to multi-annual drought; Kim and Raible 2021; Kim et al. 2022). In terms of a global approach, these three drought metrics (PDSI, SPI, and SPEI), have been shown to represent atmospheric moisture supply and availability well enough that they are often taken as a sort of "truth value" when conducting drought studies across diverse locations around the globe (e.g., Jiménez-Muñoz et al. 2016; Rhee and Im 2017; Mohamadi et al. 2020; Mehr et al. 2022) as well as on the global scale (e.g., Cook et al. 2018; Stevenson et al. 2018).

As discussed, identification of drought can be difficult given the plethora of available indices, heterogeneity of drought, and differing levels of impacts depending on the time scale and type of drought. The general 'gold standard' of drought identification and monitoring in United States is considered to be the United States Drought Monitor (USDM; Svoboda et al. 2002), which uses a suite of drought metrics, observations of moisture conditions on the ground, and expert opinions to determine when there is drought, how widespread it is, and what type of drought it is. This process shows the complicated nature of drought, and the wide array of variables that have to be considered to accurately monitor it.

Along with the discussion of drought identification is the discussion of drought prediction, which is typically done in one of three ways: (1) hydrologic modeling, (2)

statistical modeling, or (3) a hybrid of the two (Fig. 2.2; Hao et al. 2018; AghaKouchak et al. 2022; Dikshit et al. 2022a). Hydrologic modeling involves attempting to predict moisture variables as a whole using physics-based numerical models and using those to predict drought. However, the numerical models could be inaccurate (moisture variables are known to be among the most difficult to represent in models and to predict), the moisture budget may not be closed in the model, computations may be difficult and computational expensive to perform, and so on. Statistical modeling attempts to predict drought metrics using statistical correlations (e.g., using climate indices in a multilinear model after determining how correlated or how strong of a driver the climate index is in driving the drought). Hybrid models combine the two approaches. Most notably among the statistical approach to drought prediction is the emergence of machine learning in drought forecasting.

2.2 Machine Learning in Drought

Machine learning incorporates a set of inputs into iterative statistical models to create a predictive model. The model is then trained to deliver to deliver realistic and reliable predictions (i.e., the model makes predictions, estimates the error, then changes parameters in the model to reduce the error, and repeats until the error reaches a minimum value, ideally a global minimum). A more complete overview of ML and the ML models used in this dissertation can be found in chapter 3.4, and details on ML methods and terminology can be found in (Chase et al. 2022, 2023). ML models have a number of benefits, such as their ability to learn non-linear and non-stationary patterns in datasets (Dikshit et al. 2022b; Park et al. 2016; Rhee and Im 2017; Dikshit and Pradhan 2021), and they have been shown to improve on existing methods in making predictions (e.g., in predicting temperature extremes in Boulaguiem et al.

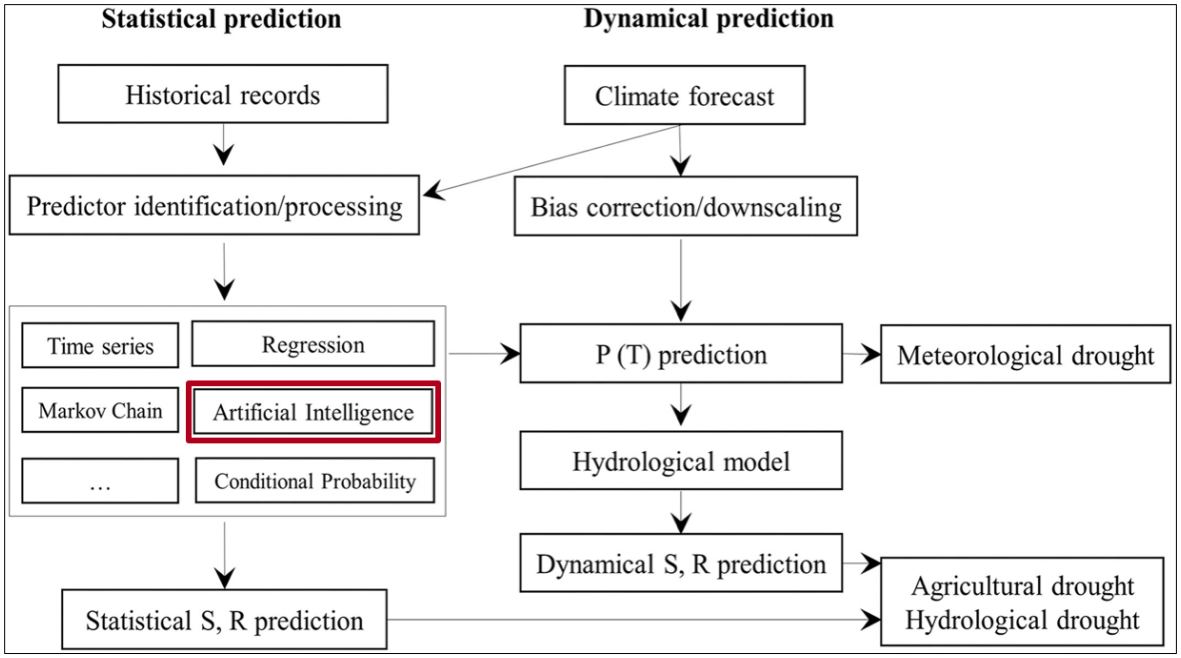


Figure 2.2: Illustration of hydrologic and statistical forecasting of drought events, and the steps involved. Hybrid forecasting involves a combination of these two approaches. [Figure 5 in Hao et al. (2018).]

2022, investigating atmospheric rivers, tropical cyclones, and fronts in Liu et al. 2016, and in predicting evaporation and soil moisture in Babaeian et al. 2022; Li et al. 2022; Liu et al. 2022). ML in drought studies have primarily been used a purely statistical models (hence their reference as a statistical predictor in Figure 2.2), however they can also be constrained by physical laws or features to help improve their performance (in this approach ML models could be argued as a more hybrid approach to drought predictions than purely statistical), though this approach is still largely unexplored in drought predictions (Hsieh 2022).

While the use of ML in environmental sciences has shown a great deal of promise, they are not foolproof. Once trained, the ML model parameters have been heavily adjusted from their original values, meaning the exact nature of the model and how it makes its predictions can be unknown, even to the designer of the ML model. Combined with the fact that ML models are highly reliant on the dataset they are trained on, learning any inherent bias, known and unknown, within them can make it difficult to determine if the model is making correct predictions for the right reason, or if the ML models are being ethical or can be fully trusted (McGovern et al. 2019, 2022; Watson 2022; Mamalakis et al. 2022; Flora et al. 2024). This particular issue, on making ML models understandable, explainable, interpretable, and trustworthy has become important enough to spawn its own field of study termed explainable AI, or XAI, which incorporates a number of methods to determine how ML models make their predictions (e.g., see Mamalakis et al. 2022; Flora et al. 2024).

The use of ML to investigate and identify FD is still a new topic, with most studies focused on investigating traditional, long-term drought events. Initially, much of the research in ML and drought has focused primarily on standard ML approaches to predict drought, such as various types of support vector regressors (e.g., Ganguli and Reddy 2013; Mohamadi et al. 2020; Dikshit and Pradhan 2021), boosted decision trees

and forests (e.g., Park et al. 2016; Rahmati et al. 2020), random forests (e.g., Park et al. 2016; Rhee and Im 2017; Rahmati et al. 2020; Prodhan et al. 2021), and other tree based approaches (such as cubist models in Park et al. 2016). In addition, research into this topic has almost entirely been treated as a regression problem, where a drought index is predicted and used to represent drought (e.g., predicting SPI or SPEI), which has proven useful given there is not a concrete definition of drought (Mohamadi et al. 2020; Dikshit and Pradhan 2021; Dikshit et al. 2022b). Predictors for drought have also varied with each study, with most focusing on climate indices (e.g., Ganguli and Reddy 2013; Rhee and Im 2017; Mohamadi et al. 2020; Dikshit and Pradhan 2021), and/or satellite derived variables (e.g., Park et al. 2016; Rhee and Im 2017; Dikshit and Pradhan 2021; Prodhan et al. 2021). The results of these studies have shown ML models recreate drought indices well, though performance varies from one region to another (as a note, many of these studies are conducted across various locations, but ML drought prediction studies are more commonly published in China, India, and Iran; Dikshit et al. 2022b). Overall, ML models have performed comparatively well to other deployed methods for predicting droughts (Rhee and Im 2017; Rahmati et al. 2020).

Predictions of drought with ML methods have improved with the addition of neural network based models (Dikshit et al. 2022b; Hsieh 2022). However, most drought studies have focused on the standard, densely connected artificial neural networks (ANNs; Mohamadi et al. 2020; Dikshit and Pradhan 2021; Dikshit et al. 2022b; Hsieh 2022). With the inclusion of ANNs and variants of them (the most popular variant of ANNs in drought predictions being the adaptive neuro-fuzzy interface system or ANFIS), drought predictions with ML methods have improved and become competitive with the current methods of drought prediction. However, the full potential of ML methods in drought are still largely unexplored, as many neural network types have not

been thoroughly explored in drought predictions, including recurrent and convolutional neural networks (RNNs and CNNs), despite the prominence of such networks in others atmospheric and environmental sciences (e.g., in predictions of atmospheric extremes and gross primary production with CNNs in Liu et al. 2016; Marcolongo et al. 2022, 2m temperature, relative humidity, and pressure with RNNs in Singh et al. 2019, and temperature extremes with a generative adversarial network or GAN in Boulaguiem et al. 2022). Indeed, most uses of neural network or deep learning (DL) methods have come in monitoring and predicting proxies for drought, such as using RNNs and GANs to monitor and predict soil moisture (Liu et al. 2020a; Foroumandi et al. 2024), or in predicting evaporation with a mix of convolutional and recurrent networks (Babaeian et al. 2022). In like manner, use of XAI methods is still being adopted in drought predictions with ML. In addition, subsets of drought, such as FDs, have not been greatly explored using ML methods (Dikshit et al. 2022b; Tyagi et al. 2022).

2.3 Flash Drought

Flash droughts describe the rapid evolution of drought conditions (on the order of 3 – 6 weeks depending on the definition of FD used; Otkin et al. 2018; Lisonbee et al. 2021; Tyagi et al. 2022). Emphasis of FD impacts has mostly been placed around agricultural impacts (Mahto and Mishra 2020; Christian et al. 2021; Gavahi et al. 2022; Mahto and Mishra 2023), however they are known to develop dry conditions that also lead into traditional, long-term droughts, increase ecosystem stress, increase fire risk (Otkin et al. 2018, 2021; Lisonbee et al. 2021), and exacerbate, or lead into conditions favorable for heatwaves (e.g., the 2010 Russian heatwave; Christian et al. 2020). However, FDs have proven more complicated to monitor and predict than standard, long-term drought due to the inherent non-linearity involved in the surface

feedback loop that characterized FDs, the strong heterogeneity in FD climatology, and the fact that FD can develop from near or above normal moisture conditions (Otkin et al. 2018, 2021; Basara et al. 2019; Christian et al. 2019a, 2021; Osman et al. 2021; Tyagi et al. 2022). This is exacerbated further by the fact that rapid development of FD means that long-term indices typically used for drought would not be effective in monitoring FD events.

FDs are a relatively new discovery. The term was first popularized in 2002 (Svoboda et al. 2002; Lisonbee et al. 2021), however studies into FD did not begin until 2013 with Anderson et al. (2013) (Otkin et al. 2018; Lisonbee et al. 2021; Tyagi et al. 2022). Early research into FD focused on using evaporative stress to investigate drought, creating the Evaporative Stress Index (ESI), and finding it was useful in being able to characterize FD conditions, and in acting as a potential precursor for FD events by up to two weeks (Anderson et al. 2013; Otkin et al. 2013, 2014). Following this, other methods for developing early warning for FD were developed using the Rapid Change Index (RCI; Otkin et al. 2014), which was built off of the ESI, and Evaporative Drought Demand Index (EDDI; Hobbins et al. 2016; McEvoy et al. 2016), which transforms potential evaporation into a standard normal distribution. These marked some of the first indices created for the purpose of investigating and monitoring FD conditions. Alongside these, research also focused on what characterized and drove FD events, finding a positive feedback loop involving surface interactions (Tyagi et al. 2022; Christian et al. 2024). In this loop, high atmospheric demand (often driven by high temperatures and/or low precipitation) would drive high evaporation from the soils and vegetation. This would decrease available soil moisture, decreasing latent heat flux and, to keep the surface energy budget balanced, increasing sensible heat flux. This would increase the temperature, thus increasing the potential evaporation and atmospheric demand, which in turn drive more evaporation (Fig. 2.3).

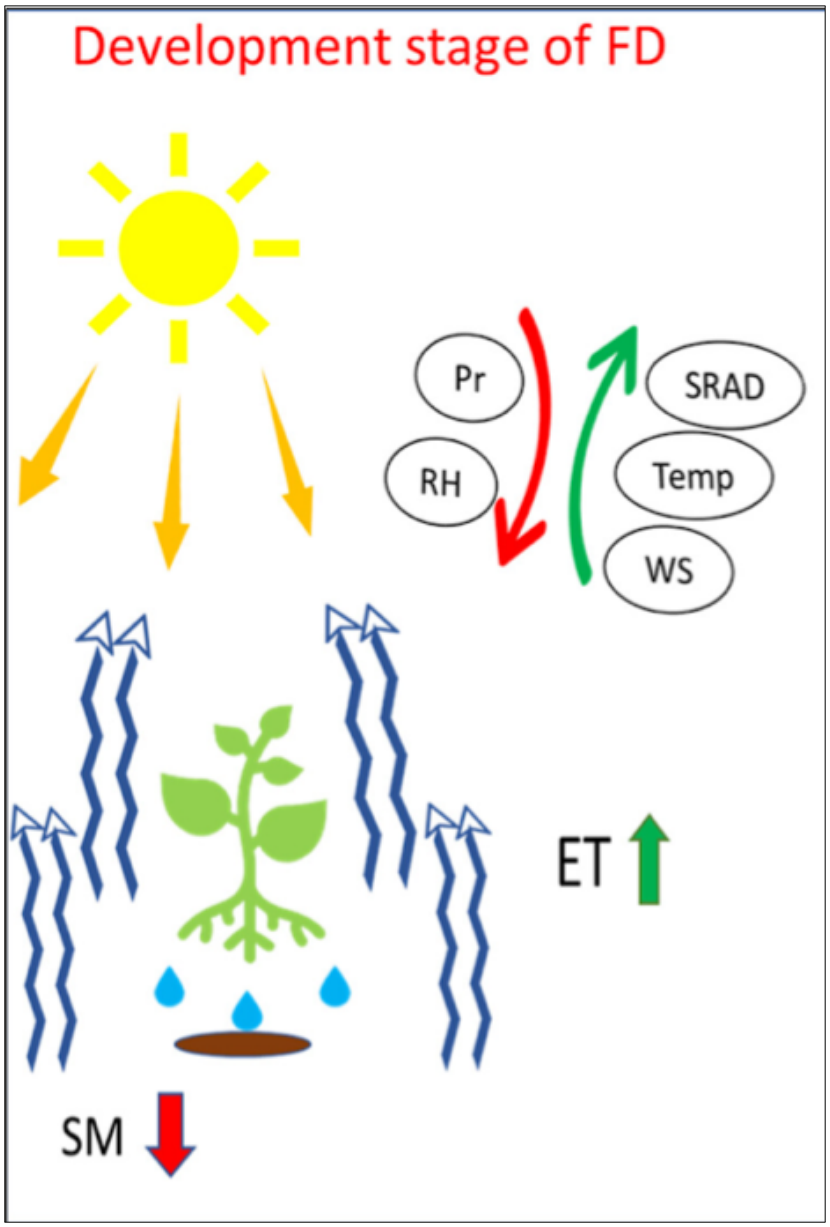


Figure 2.3: Illustration of the surface interactions that create FD events. [Figure 2a in Tyagi et al. (2022).]

Following this, a number of studies had been published about the mechanisms of FD, however there were few attempts to properly define FD or quantify it. Therefore, Otkin et al. (2018) challenged the FD community to work on quantifying FD and determine definitions for FD, stressing the focus on the rapid intensification of drought conditions, and on ensuring that the environment was actually in drought after the rapid intensification period (in which the authors suggested the standard metric of a moisture variable being below the 20th percentile). Numerous methods for identifying and quantifying FD published as a result of this (e.g., Christian et al. 2019b, 2021, 2023; Noguera et al. 2020; Pendergrass et al. 2020; Li et al. 2020; Liu et al. 2020b; Otkin et al. 2021; Osman et al. 2021).

The first of these was Christian et al. (2019b), which was later refined and improved in Christian et al. (2022, 2023). This method focused on evaporative stress and the change in evaporative stress to identify FD (the change in the evaporative stress had to fall below certain percentiles to define the rapid intensification period, and the evaporative stress had to fall below the 20th percentile at the end of the rapid intensification to be determined as FD[1]; Fig. 2.4). Other methods used a similar approach of requiring a metric had to change within a rapid time frame and fall below the 20th percentile to be considered FD. For example, the method employed in Noguera et al. (2020) uses rapid changes in the SPEI to define the rapid intensification period, and SPEI to determine

[1]More precisely, the method in Christian et al. (2023) used the standardized evaporative stress ratio to determine FD by requiring the intensification period is at least 30 days, the overall change in the standardized evaporative stress ratio was below the 25th percentile, and that same variable was below the 20th percentile at the end of the intensification period.

if the region is in drought[2]. Pendergrass et al. (2020) employed two methods to identify FD using rapid increases in the USDM categories, and rapid increases in EDDI[3], while Ford and Labosier (2017) focused on soil moisture percentiles, which was later incorporated in modified forms in Liu et al. (2020b), Yuan et al. (2023), Mahto and Mishra (2023), and others. The idea behind these methods was to use soil moisture percentiles. The rapid intensification began when the percentiles drop below the 40th percentile and end when a threshold of percentile decrease was not being met. For example, Liu et al. (2020b) required soil moisture percentiles to decrease by at least 6.5 percentiles per week to be considered rapid intensification. If the soil moisture percentiles were below the 20th percentile at the end of the rapid intensification, then it was flash drought. Lastly, the USDM was also used as a way to identify FD in Chen et al. (2019) as well, which determined FD if a region experienced an increase of 2 or more drought categories in a few weeks.

By the end of 2021, there were numerous published definitions for FD, each using some different metrics and schemes, some assuming FDs last only 6 - 8 weeks, and others allowing it to develop into traditional long-term drought, and others still not focusing on rapid intensification at all (all methods published prior to 2021 are listed in Lisonbee et al. 2021). Afterwards, more methods for FD identification were published, such as Otkin et al. (2021) and Osman et al. (2021), which used two different approaches with soil moisture. The plethora of different approaches to FD allows for

[2]The method in Noguera et al. (2020) required that the intensification period was at least 30 days, that the SPEI droped at least 2 standard deviations in that time, and the SPEI was 1.28 standard deviations below normal, the 10th percentile of a standard normal distribution, at the end of that intensification period.

[3] Pendergrass et al. (2020) required the EDDI to increase by 50 percentiles in 2 weeks and maintain that high EDDI for at least 2 more weeks to be considered FD.

numerous different ways to investigate FD, with each method giving its own "perspective". For example, the Christian et al. (2019b) method emphasizes FD more in moisture and energy transition regions and agricultural regions (due to the higher evaporation of crops) than other methods might (Osman et al. 2021; Alencar and Paton 2022).

In addition to this, studies on FD have begun to focus down on a select few variables for investigating and identifying FD. Namely, studies have begun to focus heavily on soil moisture (generally root-zone soil moisture) as one of the key variables in the feedback loop, but also evaporation and potential evaporation (Chen et al. 2019; Christian et al. 2019b, 2021; Pendergrass et al. 2020; Otkin et al. 2021; Tyagi et al. 2022). Other variables, such as high temperatures and low precipitation are also known to drive FD and help initiate that feedback loop (Otkin et al. 2018; Tyagi et al. 2022). Similar to traditional droughts, metrics focusing on these variables have also been created to investigate FD events, with the ESI, RCI, and EDDI already discussed, but also the Standardized Evaporative Stress Ratio (SESR; uses the ratio of evaporation to potential evaporation; Christian et al. 2019b; Lowman et al. 2023), the Standard Evaporative Deficit Index (SEDI; uses the difference between evaporation and potential evaporation; Vicente-Serrano et al. 2018; Li et al. 2020), and the Flash Drought Intensity Index (FDII; uses soil moisture percentiles, how rapidly they decrease, and the length of time they are below the 20th percentile to determine FD and FD intensity; Otkin et al. 2021), to name a few.

With the publication of FD identification methods also came studies examining the climatological characteristics of FD. For example, Christian et al. (2019b) examined the frequency of FD occurrence in the United States, finding several hotspot regions, such as the areas of intensive agriculture in the central United States, on the southeastern coast of the United States, and in the upper Mississippi River floodplains. The

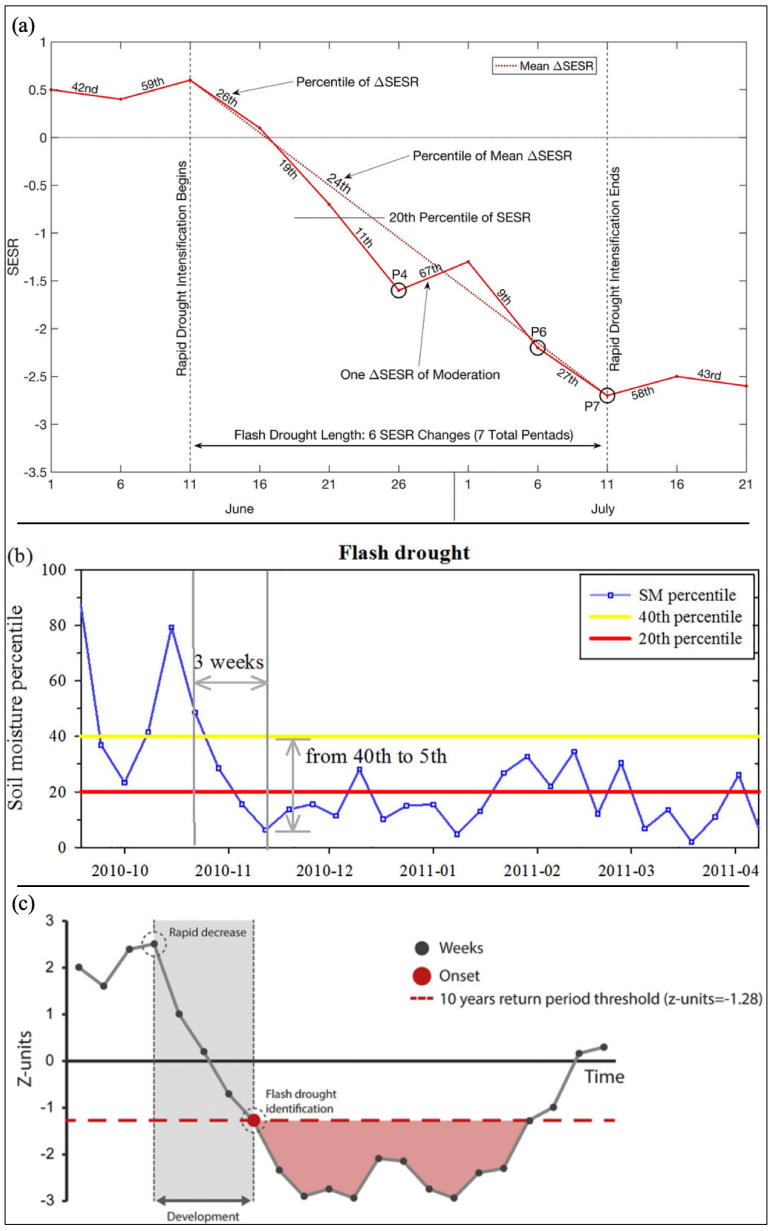


Figure 2.4: Illustrations of FD identification methods for the (a) Christian et al. (2019b, 2023), (b) Liu et al. (2020b), and (c) Noguera et al. (2020) methods. Note the methods cluster around identifying rapid intensification, and determining if the variable is below a threshold at the end of the intensification period. [(a) Figure 2 in Christian et al. (2019b), (b) Figure 2b in Liu et al. (2020b), and (c) Figure 1c in Noguera et al. (2020).]

seasonality of FD (that is, the months more favorable for FD development/months in which FD is more likely to occur), in the U.S. was also found to be more focused on the spring and early summer seasons (Christian et al. 2019a; Chen et al. 2019). Other FD climatologies studies have been conducted in the United States finding similar results (e.g., Christian et al. 2019a; Otkin et al. 2021).

FD climatologies in other countries have also been examined, though not as frequently. For example, studies have been conducted in China (Liu et al. 2020b; Li et al. 2020), in India (Mahto and Mishra 2020), and in Spain (Noguera et al. 2020). In India, for example, many of the FDs were attributed to delayed or weak monsoons, though temperature driven FDs were found in southern India and the upper Himalayas, with similar seasonalities to the monsoon (Mahto and Mishra 2020). Note that most of the FD studies have been focused on the United States and, to a lesser degree, in China, with only a few studies in other locations.

To help with this, some of the research has discussed FD on a global scale (Christian et al. 2021, 2023; Qing et al. 2022; Yuan et al. 2023; Mahto and Mishra 2023). These identified several regions across the world, such as the Iberian Peninsula, western Russia, Northern Australia, southeast Asia, he eastern Amazon, and other agricultural locations as being hotspots for FD development (Christian et al. 2021; Qing et al. 2022; Mahto and Mishra 2023). Christian et al. (2021) also examined the drivers of FDs based on region, focusing on precipitation driven and PET driven FDs, with results varying vastly from region to region (e.g., FD in the European hotspots were normally driven by high potential evaporation or increased atmospheric demand, while man of the hotspots in the Americas had FDs driven by precipitation deficits). This study has been complemented with Qing et al. (2022) and Mahto and Mishra (2023), which found similar results and emphasize humid locations as being more vulnerable to FD, and focused on the global agricultural impacts and responses to FD.

Finally, with increased understanding of FDs, what drives them, how they develop, their climatologies, and methods to quantify them, studies in recent years have begun to focus on how FD will evolve in time (i.e., under climate change). While early investigations of FD in the Coupled Model Intercomparison Project, phase 5 (CMIP5) yielded mixed results (Yuan et al. 2019; Hoffmann et al. 2021), later investigations of FD in CMIP phase 6 (CMIP6) show more robust and realistic results in representing FD (Christian et al. 2023; Yuan et al. 2023). In keeping with the heterogeneity of FD, projections of FD into the future have shown some areas will increase in FD occurence (e.g., in the Iberian Peninsula, eastern Europe, and the eastern Amazon; Shah et al. 2022; Christian et al. 2023), and other areas where FD will decrease (e.g., in India; Christian et al. 2023). Another study has also shown that FD is steadily replacing traditional, long-term drought as the default mode for drought initiation, and showed FD will become more common in the future (Yuan et al. 2023), a trend that has already been observed in global climatologies (Qing et al. 2022). This issue is expected to worsen further with increasing populatios and increasing demands of agriculture and water supples (Gavahi et al. 2022; Iglesias et al. 2022; Shah et al. 2022; Yuan et al. 2023; Foroumandi et al. 2024). Thus, being able to monitor and predict FD becomes of high importance. One of the promising avenues in this, one that remains untouched except for a few papers (Zhang et al. 2022; Foroumandi et al. 2024), is the use of ML techniques for FD identification and prediction Tyagi et al. (2022); Dikshit et al. (2022b).

2.4 Machine Learning and Flash Drought

Overall, a great deal of research has been done on droughts, drought predictions, the use of ML models to predict droughts, and about FDs, their drivers, mechanisms, and

physical characteristics. However, there has been very little work to put these two topics of ML predictions and FD together. Briefly, the studies by Zhang et al. (2022) and Foroumandi et al. (2024) are the only known studies that explicitly attempts to investigate FD using ML methods. Zhang et al. (2022) first identified droughts and FDs using soil moisture and a method similar to Ford and Labosier (2017) for FD identification, then they trained different ML methods to differentiate between droughts and FDs, a feature the ML skillfully accomplished (and in which random forests outperformed other ML methods, including RNNs). Foroumandi et al. (2024) used a modified form of a GAN (called a conditional GAN or CGAN) to generate maps of the Soil Moisture Stress Index (SSI), which was then correlated with FD events. The CGANs showed considerable skill in being able to recreate the SSI maps, and learned complex non-linearities in the soil moisture, though there were regions, such as the northeastern United States and Great Lakes region, where the skill of the model dropped a little.

There have also been other studies that have had similar ideas to FD and ML, however these studies were predicting FD indicators, and monitoring and predicting FD was not the goal of those studies. For example, other studies, similar to Foroumandi et al. (2024), have sought to predict soil moisture using DL methods, such as Liu et al. (2022) (which used random forests to predict soil moisture). Babaeian et al. (2022) used recurrent and convolutional-recurrent neural networks to predict evaporation patterns, with the convolurtional-recurrent network showing the greater skill.

However, aside from Foroumandi et al. (2024) (whose explicit goal was to monitor FD via SSI), there has been no direct or explicit investigation of the ability of ML models to represent and predict FD. Therefore, the contributions of this dissertation project was the determination of the abilities of ML, such as random forests (RFs), Ada boosted trees, support vector machines (SVMs), ANNs, RNNs, and convolutional

U-nets, to represent FD events directly, determining which ML models perform best for this task and where the ML models need to be improved. The ML models were first trained locally, on a domain of the Contiguous United Stats (CONUS), with the best performing models being carried over to a global domain to determine the ability of the ML models to generalize to the heterogeneous patterns of FD. Because there is no single definition for FD, the ML models were trained to identify FD according to five different identification methods, and evaluated for each one. The use of a global analysis allowed for the opportunity to investigate global FD patterns using identification methods that have previously only been examined in certain regions, as well as investigate FDs in regions that are rarely discussed in the literature, and thus see how impacts of FD may vary from region to region. Therefore, in addition to the contributions stated above, this dissertation also builds on the work done by previous investigations of regional FD climatology by expanding their methods to the global domain and adding additional global perspectives and patterns of FD to those made in Christian et al. (2021). This dissertation will provide some of the early steps into incorporating ML methods in FD research, help set some of the initial goal posts in ML performance for FD, and thence improve our ability to monitor and predict FD events.

Chapter 3

Datasets, Machine Learning Algorithms, and Statistical Methods

3.1 Datasets

For this project, two datasets were used to train the ML models. The North American Regional Reanalysis was used to collect data for the contiguous United States (CONUS) and training the datasets thereon, while the European Centre Medium-Range Forecast's Atmospheric reanalysis was collected to train the ML models on the global scale. Figure 3.1 shows the design process, including variables collected from each dataset, and some steps applied prior to ML training. The exact nature of the preprocessing step changed depending on the dataset used, and is detailed below.

3.1.1 North American Regional Reanalysis

CONUS scale data for this project originate for the from North American Regional Reanalysis (NARR) dataset (Mesinger et al. 2006). The NARR provides high resolution data at the $0.3^\circ \times 0.3^\circ$ or approximately 32 km $\times$ 32 km over North America and for over 40 years of data (from 1979 to present). Surface variables in the NARR were determined by the Noah land surface model (Ek et al. 2003). For the NARR, the SM collected represents a 0 – 40 cm average in SM. While the NARR provides data for all of North America, for this project it was subsetted to focus on CONUS.

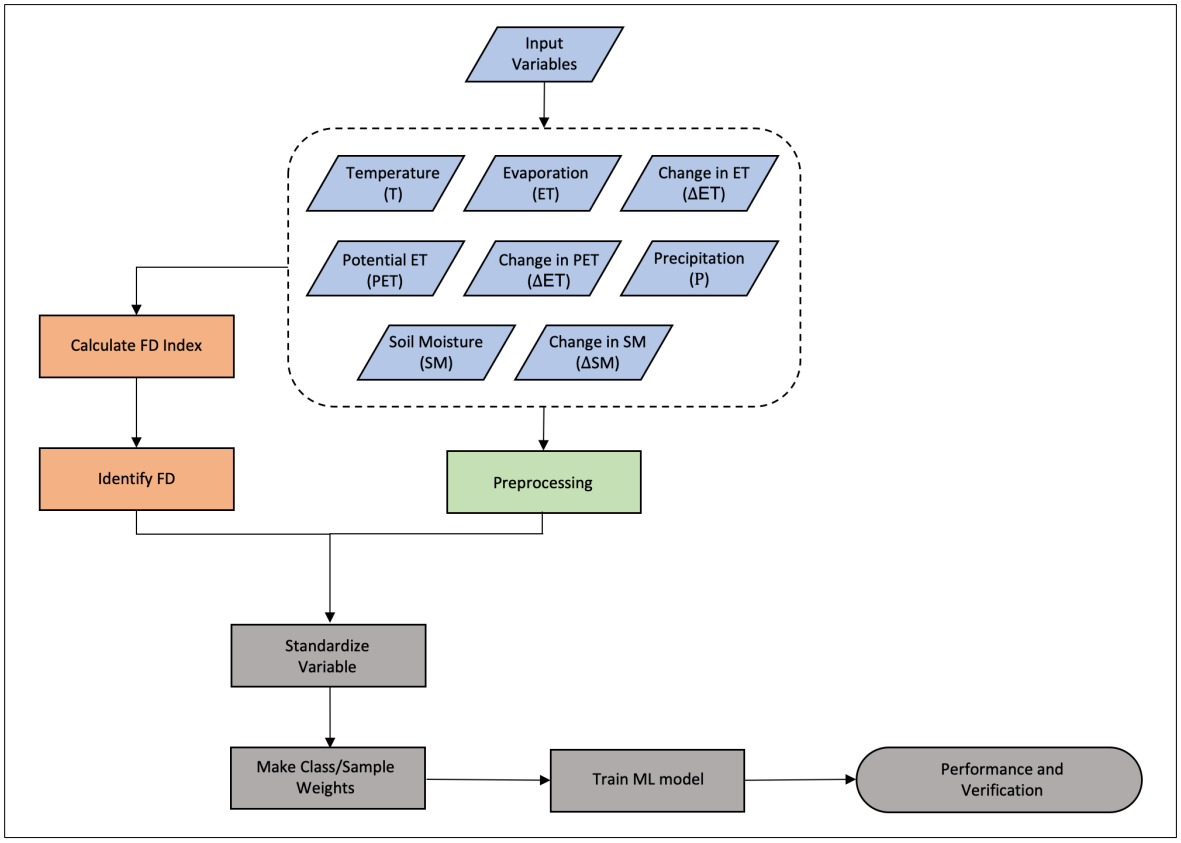


Figure 3.1: Experiment design for this project. The steps for data preparation can be broken down into data collection, preprocessing (exact nature of preprocessing depends on the dataset), splitting into training, validation, and testing folds, and standardization.

The seasonality of FD in CONUS favors spring and summer seasons (Chen et al. 2019; Christian et al. 2019b, 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, the growing seasons for 1979 – 2021 were chosen and the data was averaged into pentads (5-day, non-overlapping mean). The pentad averaging was performed to retain the signal of FD predictors and minimize day-to-day variations.

3.1.2 European Centre Atmospheric Reanalaysis v5

Global data was collected from the European Centre for Medium-Range Weather Forecasts (EMCWF) Atmospheric Reanalysis version 5 (ERA5) dataset (Hersbach 2023). The ERA5 dataset provides high resolution, reliable data with high time scales that have been shown to be effective in identifying FD (e.g., Christian et al. 2021; Shah et al. 2022; Mahto and Mishra 2023; Yuan et al. 2023). The ERA5 data was collected for the same years as the NARR data (1979 – 2021, and into 2022 ofor the Southern Hemisphere). SM data from the ERA5 represent a 0 – 28 cm average in SM. Compared to the NARR, additional levels of preprocessing was needed for the ERA5 dataset in order to load the full dataset for ML training, and because of the nature of a global dataset versus a local one.

First, the NARR dataset focused on a growing season of April to October, however this is not consistent across all parts of the globe (more specifically across both hemispheres). Christian et al. (2021) divided the globe into three parts, with the growing season running from March to October north of 30°N, September to March south of 30°S, and a year-round growing season between 30°N and 30°S. However, the results of Christian et al. (2021) still showed that tropical FD was generally favored in the

months when the ITCZ was present (i.e., during the hemispheric summer season). To this end, this study simplifies the global dataset by defining the growing season as April to October in the Northern Hemisphere, and September to March in the Southern Hemisphere[1].

In addition, to help manage the size of the datasets, the ERA5 was averaged down to $0.5^\circ \times 0.5^\circ$ resolution, and the sea grid points were removed prior to training the ML models to reduce the data size by approximately 70% and were re-inserted post hoc for plotting the results. The data was also standardized for each grid point prior to be used for training the ML model. Lastly, the poles were excluded from the study as those regions are too cold throughout the growing season to experience higher evaporation rates and sufficient soil moisture depletion Christian et al. (2021, 2023). An aridity index mask was also applied to mask out low moisture environments for similar reasons (Christian et al. 2021, 2023; Zomer et al. 2022)[2].

3.2 Input Data

ML models were trained on a set of 8 variables that have been identified as key variables in FD identification, monitoring, and driving the feedback process that characterizes FD events. These variables are temperature (T), precipitation (P), evaporation (ET), potential evaporation (PET), and soil moisture (SM). For the key FD variables that are involved in the feedback process and are regularly used for FD identification (ET, PET, and SM; e.g., Christian et al. 2019b, 2023; Li et al. 2020; Liu et al. 2020b; Pendergrass

[1]The Southern Hemisphere growing season was allowed to extend two pentads into April to keep it the same number of pentads per growing season as the Northern Hemisphere, allowing the two to be merged within the code.

[2]Regions masked by the aridity index or cold environments were treated the same as sea and ocean values and given class/sample weights of 0 so that they were ignored in training.

et al. 2020; Otkin et al. 2021), the *change* in those variables from one pentad to the next were used to capture the rapid intensification component of FD. For this project, the temporal change in these three main drivers were calculated as:

$$\Delta x_p = x_p - x_{p-1}, \tag{3.1}$$

where x is either SM, ET, or PET, p is the pentad, and $p-1$ is the previous pentad. Additionally, a 6-pentad leading moving average was applied to each change variable so that ΔSM, ΔET, and ΔPET all represent the mean 30 day change in their respective variable. Thus, the input parameters used to identify FD, in the order they are given to the ML algorithms, are T, P, ET, ΔET, PET, ΔPET, SM, and ΔSM. Moreover, prior to training an ML model, each variable was standardized for each grid point.

3.3 Output Data

For this project, the FD identification problem was treated as a classification problem. Thus, the output datasets consisted of two values or labels[3]; 0 (i.e., there was no FD for that grid point and pentad), or 1 (i.e., there was FD for that grid point and pentad). However, because there is no consensus on a definition for FD, five different FD identification methods were chosen for this project, and the ML algorithms were trained to identify FD for each method separately.

The first method for FD identification used was the method outline 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 the satellite derived Land Surface Water Index (LSWI), which was later incorporated into

[3]There is also a hidden third class applied to all sea or ocean grid points, whose class/sample weight was set to 0 so that only land values were used in training.

SESR Christian et al. (2023). 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 employed the improved and simplified version of FD identification in Christian et al. (2023) (hereafter C23). This method requires that the rapid intensification even lasts a minimum of 30 days, that the average change in the filtered SESR is below the 25th percentile (quantification of rapid intensification), and that the filtered SESR is below the 20th percentile at the end of the intensification period (quantification of drought).

The second method used is outlined in Noguera et al. (2020) (hereafter N20). N20 used SPEI on the weekly timescale 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 (24 days; 30 days on the pentad timescale), that SPEI decreased by at least 2 standard deviations (or z-units in the N20 paper) in the intensification period (quantification of rapid intensification), and that the SPEI was 1.28 standard deviations below normal (i.e., an SPEI value of -1.28) at the end of the intensification period (quantification of drought).

The third method is given in Pendergrass et al. (2020) (hereafter P20). P20 used EDDI to identified FD when the EDDI increased by 50 percentiles over two weeks (or 3 pentads; rapid intensification quantification), and that the increased EDDI was maintained for another two weeks. Because the EDDI increased by a large margin (50 percentiles), the area was assumed to be in drought after the intensification period. P20 also used another method of FD identification which involved how rapidly drought intensified according to the USDM. The USDM is still relatively new, having started in 2002 (Svoboda et al. 2002), and its methods for categorizing drought has 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 (i.e., only the first definition of P20 is intended when discussed here).

Liu et al. (2020b) (hereafter L20) detail the forth 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 (4 pentad) window of rapid intensification, the change in SM percentiles must average -6.5 percentiles/week (or a percentile/pentad decrease) or have a minimum of -10 percentiles/week (or -7.5 percentiles/pentad; rapid intensification quantification), and that the minimum SM percentile was below the 20th percentile (drought quantification).

The final FD identification method used in this project was from Otkin et al. (2021) (hereafter O21). O21 defined the FDII to quantify the occurrence and intensity of FD using SM percentiles at the pentad timescale. The O21 method quantified rapid intensification and drought separately and identified FD by multiplying the two components together. The calculation of FDII is more involved than the previous methods, and the reader is referred to O21 for details. For simplicity, FD was said to occur when FDII > 0.

3.4 Machine Learning Algorithms

Traditional ML algorithms for this project were created using the sklearn package in Python (Pedregosa et al. 2011; Buitinck et al. 2013), while neural network or deep learning models were built and trained using the TensorFlow package and the Keras API (Chollet et al. 2015; Abadi et al. 2015). To fine tune model parameters, a subset of the data (10 folds and 4 rotations) were used. The parameters with the best validation performance 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 (for sklearn models) or sample weights (for TensorFlow models) used to account for class imbalance. Class and sample are weights that were applied to labels in the error functions being minimized when training ML models. The difference between the two is class weights were applied to the classes (e.g., class weights are applied to all 0 and 1 labels after their errors are summed), while sample weights applied the weight to each sample prior to summing the error. Class weights were applied to sklearn models, and sample weights to TensorFlow models. This is due to a limitation in TensorFlow, which does not apply class weights for multidimensional outputs. For TensorFlow models each sample weight was made uniform for a given class to mimic the effects of class weights. The class/sample weights being the most impactful parameter may be expected, considering one importance 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/sample weights were modified to fit each identification method. Descriptions of these ML algorithms and the associated terminology can be found in Chase et al. (2022, 2023).

This project explored the use of multiple different ML algorithms towards FD identification and prediction. The standard ML algorithms employed were random forests (RFs), Ada boosted decision trees, and support vector machines (SVMs), while the NN architectures explored were standard artificial neural networks (ANNs), recurrent neural networks (RNNs), and convolutional U-nets.

3.4.1 Standard Machine Learning Algorithms

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. A 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 in drought identification and prediction (Rhee and Im 2017; Rahmati et al. 2020; Dikshit and Pradhan 2021; Hsieh 2022). Tests showed that 150 trees, each allowed to grow 15 branches deep, was 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).

Boosting algorithms are based on fitting an ML model (in this case decision trees) to a dataset, 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 in traditional drought studies, with XGBoosting being the most favored (Park et al. 2016; Rahmati et al. 2020; Prodhan 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}.

Lastly, SVMs are similar to perceptrons, with a model that is minimized by a loss function similar to a linear model. Various kernels can be used for SVMs, with the RBF kernel being most common for drought studies for its ability to represent non-linear data more effectively (Ganguli and Reddy 2013; Mohamadi et al. 2020; Dikshit and Pradhan 2021). However, for this study, RBF kernels did not have a significant improvement over linear kernels, whereas it increased the computation time by an order

	C23	N20	P20	L20	O21
RF	25	3.3	105	135	4.5
Ada	55 (23)	4.5 (7.1)	160 (95)	800 (18)	8 (4)
SVM	26.5	5	120	200	6.5

Table 3.1: Class weights applied to each ML algorithm for identifying FD in each identification method. Non-FD labels were given a class weight of 1, and sea values were a class weight of 0. Non-parenthetical values denote class weights applied for the NARR dataset. Parenthetical values denote class weights applied to the global ERA5 dataset.

of magnitude. Thus, linear SVMs were used for this project with the squared hinge loss function minimized, L2 regularization, and a stopping error for 10^{-6}.

The class weights used for these ML models can be found in Table 3.1. For all ML algorithms, each grid and pentad was treated as an example of FD for the ML model to learn from.

3.4.2 Deep Learning Algorithms

Some of the neural network, or deep learning, model parameters can be found in Table 3.2[4]. These are parameters that were kept the same across all DL experiments for simplicity (regularization, dropout, optimizer, learning rate, and loss function), while more primary parameters, such as sample weights and network architecture were modified and tuned for FD. Neural network architectures used for this dissertation can be found in Figure 3.2, and final sample weights can be found in table 3.3.

[4]For the global RNN experiments for the N20 method, dropout was reduced to 0.2, as this was found to help with an issue in training.

Dropout	λ_2	Learning Rate	Optimizer	Loss Function
0.6 (U-net: 0.2)	10^{-3}	10^{-3}	Adam ($\beta_1 = 0.9$, $\beta_2 = 0.99$, decay $= 0$, $\epsilon =$ None)	Categorical Cross-Entropy

Table 3.2: Miscellaneous parameters that were applied to each neural network. These parameter values did not vary from network to network (except dropout), nor from one FD identification method to another. λ_2 denotes the L2 regularization parameter. Activation functions were tanh for LSTM layers and elu for dense and convolution layers (except for the last convolution layer, which had tanh to prevent gradient explosion).

	C23	N20	P20	L20	O21
ANN	80	55	750	900	22
RNN	33 (16.3)	21 (32)	100 (57)	200 (27)	4.5 (6)
U-net	17	5	100	116	4.8

Table 3.3: Sample weights applied to each neural network for identifying FD in each identification method. Sample weights were applied to all positive FD labels equally to mimic the effect of class weights. Negatives FD labels were given sample weights of 1, and sea values were a sample weights of 0. Non-parenthetical values denote sample weights applied for the NARR dataset. Parenthetical values denote sample weights applied to the global ERA5 dataset.

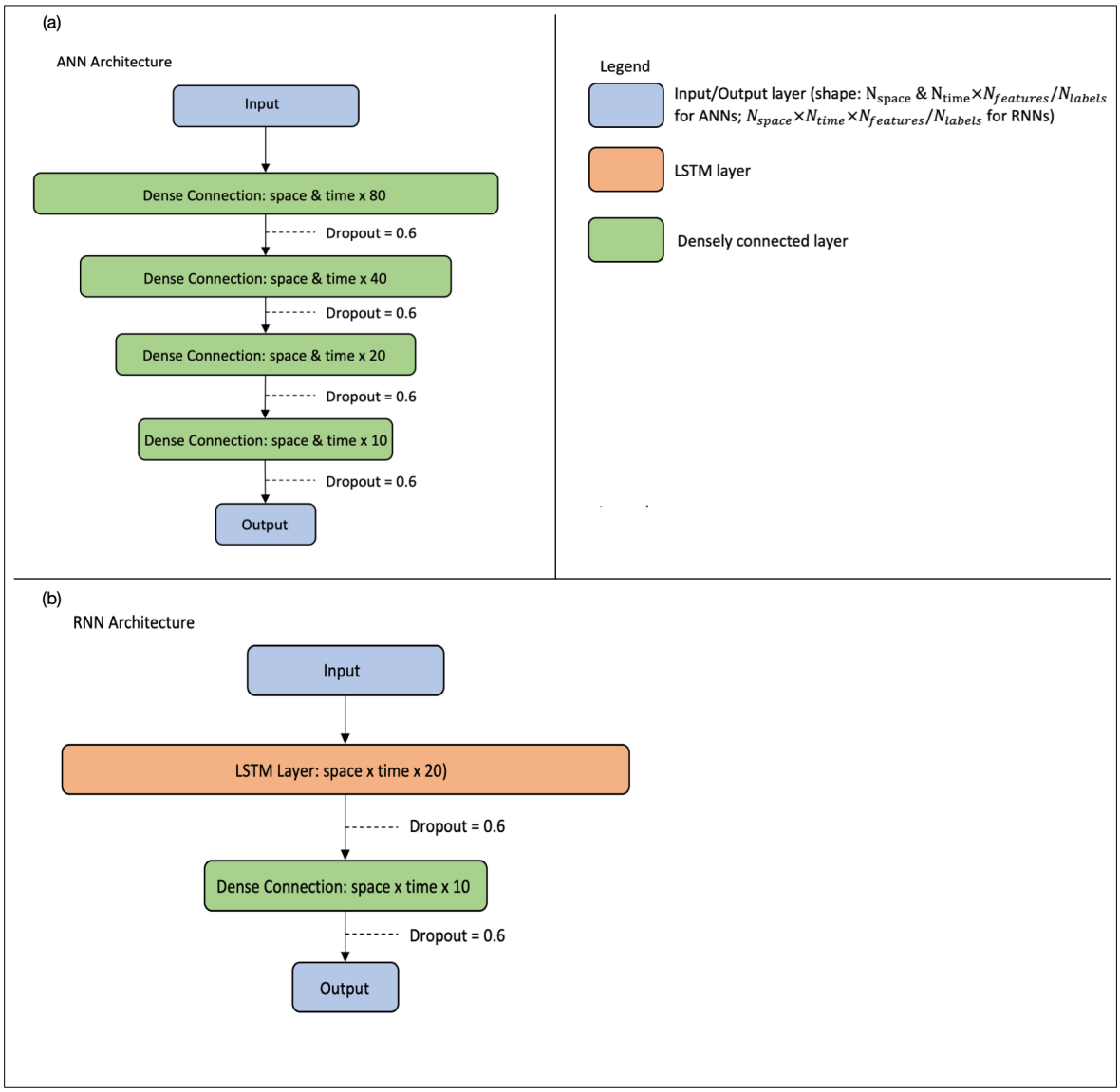


Figure 3.2: Neural network architectures used to identify FD. a) artificial neural network architecture, b) recurrent neural network architecture. Spatial dimensions (N_{space}) for the NARR grid was 20,580 (210 × 98). ERA5 spatial dimensions without ocean grid points was 266,723. Time dimension (N_{time}) was 43 pentads per fold.

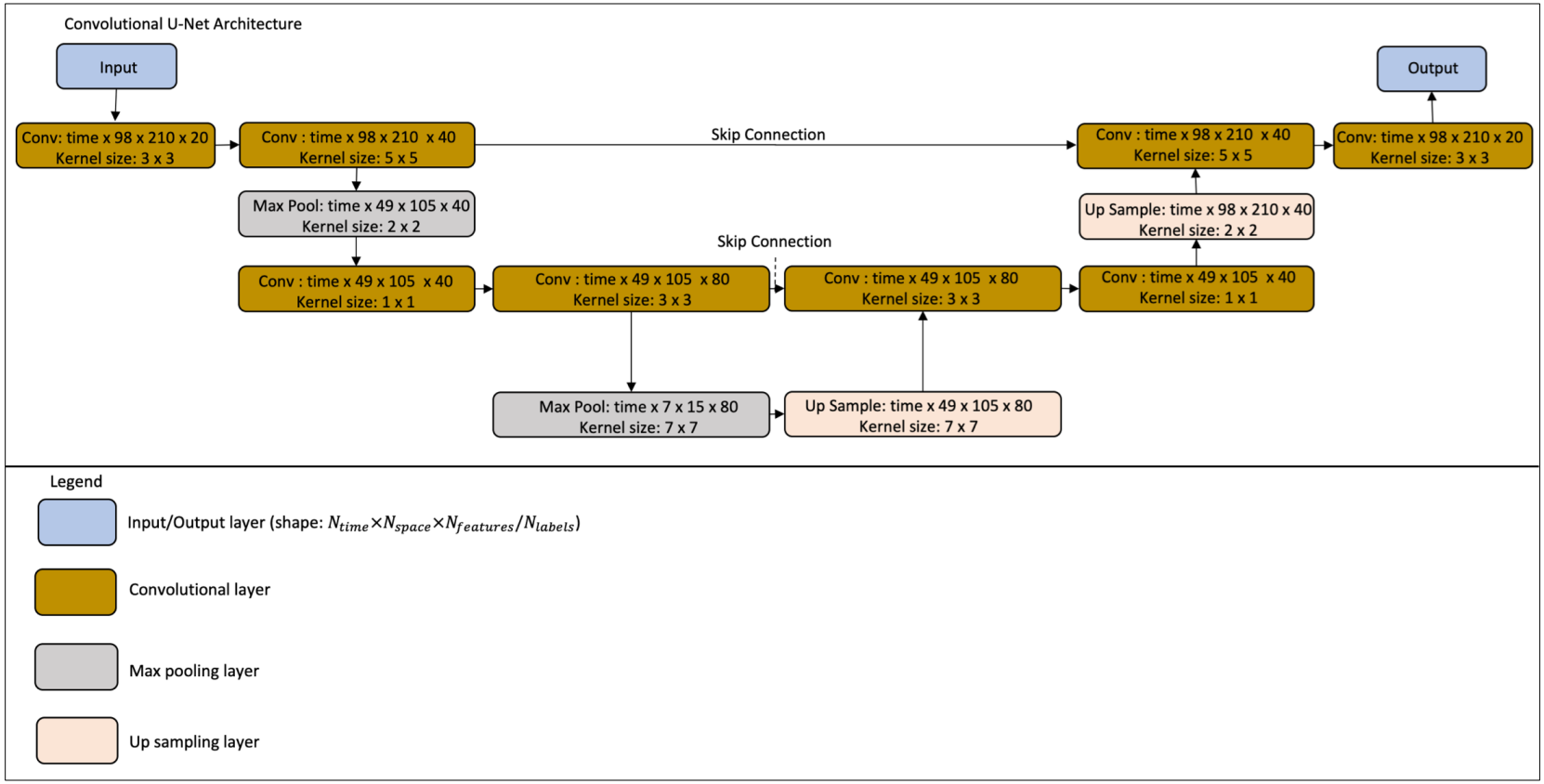


Figure 3.3: U-network architecture used to identify FD. Note the U-network has spatial dropout of 0.2 applied after every convolutional layer (not shown). Spatial dimensions (N_{space}) for the NARR grid was 210 × 98. Time dimension (N_{time}) was 43 pentads per fold.

ANNs, sometimes called densely connected neural networks (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. 2022b; 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 1 or 2 hidden layers (Dikshit and Pradhan 2021; Dikshit et al. 2022b; 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 following 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. 3.2). 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 neural network. If there were too many parameters, the network would get stuck in a simple 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 would get stuck in a local minimum in the cross-entropy loss function). These simple solutions, predictions of only one value for the entire dataset, are referred to as trivial solution throughout dissertation (stemming from the terminology in linear algebra, where a trivial solution for a set linear equations refers to solution that is all zeros) because they are trivial (predicting a single value is a simple way to get low error when predicting extreme events), and because they are neither useful nor good predictions. Conversely, if there were too few parameters, the network would not be able to sufficiently learn anything. From experimentation, approximately 2,000 to 7,000 parameters struck a balance where the networks could learn spatial and temporal features of FD with out getting trapped in a trivial solution (however, approximately 11% - 21% of rotations for the P20, L20,

and O21 methods still got trapped in trivial solutions, as seen in 5). The proposed architecture for this study contained 5,023 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 to the hidden layers.

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 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 module (LSTM), and the gated recurrent unit module (GRU) (Hochreiter and Schmidhuber 1997; Sak et al. 2014; Zaremba et al. 2014; 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. 2022b; 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 a simple RNN structure, with one recurrent (LSTM) layer followed by a densely connected layer after the time series had been learned (Fig. 3.2). This model had fewer parameters than the ANN (2,563 parameters total), and training the model an improvement over the ANNs in terms of computation time and resources.

The RNNs were faster to train (assuming no parallel processing) and did not have as many issues in getting stuck in trivial solutions (though this still occurred with the N20 method for some of the rotations). RNNs cannot be trained in parallel (due to its recurrent nature), however, and can thus be slower to train than other algorithms that employ parallel processing for training. For the RNNs, each grid point was treated as a sample, while the input features were connected to each unit, and the recursive axis was trained along the time axis.

Convolutional neural networks differ from ANNs in that not all of the nodes are connected to all of the subsequent nodes. Rather, a set of nodes from the previous layer connects to a node in the next layer in a convolutional operation. In a 2 dimensional 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, and so forth 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 layer (e.g., a kernel size of 3×3 means a sliding box of 3×3 grids will be used for inputs in the next layer). In addition, convolutional networks incorporate pooling layers, or layers that use a $n \times m$ kernel and either the average or maximum value (depending on the type of pooling used) of that kernel for the grid value in the next layer. For meteorological purposes, zero-padding (adding zeroes around the edges of the map) is typically used so that convolution layers do not change the grid size (which this project followed), while pooling is used to reduce the grid size by a prime factor (typically 2; Liu et al. 2016; Boulaguiem et al. 2022; Marcolongo et al. 2022; Foroumandi et al. 2024). An operation, termed up sampling, can be used to go in the opposite direction (increasing the grid size by a prime factor). Using a sequence of convolutional and pooling layers followed by a sequence of convolutional and up sampling layers forms a network termed a U-network (because the architecture resembles a U).

On top of convolutional networks, which are known to effectively learn spatial features, U-nets have been shown to effectively recreate and predict maps of meteorological 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 Figure 3.2. Skip connections (i.e., concatenation layers that connect layers just before dropout to layers just after up sampling) were employed to improve model training (having a copy of inputs later in the model can help jump start it when many of initial parameters or bias values are initially close to zero), and to help with model performance. Note skip connections acted as additional connections within the model (i.e., layers with skip connections skip part of U via skip connections, but were also connected to the dropout layers that followed them). This model was deeper than the other networks explored here, and contained a larger number of parameters to train (257,875 parameters), but was able to adeptly identify FD regardless. Additionally, the nature of the computations for convolutional layers are simpler than recurrent and densely connected layers, making them more efficient to train and use. For example, the U-net used for this study took roughly the same time to train as the RNNs and used similar resources, despite having orders of magnitude more parameters to train. 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 that densely connects the input variables, similar to the nodes in ANNs, except only along this third dimension). The U-nets used the time axis as the number of training examples. Note this means the U-nets had fewer samples to train on than the other networks (a training set had 41 folds, with 43 pentads per fold for 1,763 samples compared to 20,580 samples for when each grid point in the NARR was treated as an example).

3.5 Machine Learning Verification and Interpretation

Verification of ML models was performed via 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 datasets cannot be divided randomly since some of the variables, such as T, PET, and especially SM, are autocorrelated to various degrees. However, because there is approximately a 5 month difference between each growing season, it was assumed that the autocorrelation is negligible between one growing season and the next. Hence, each growing season is statistically independent and considered 1 fold. For each ML experiment, 1 fold was reserved for validation, 1 fold for testing, and the remaining folds for training. For the global datasets, 20 folds were used for training as a result of some computational limitations. Thus each dataset (CONUS and global) was divided into 43 folds, and there were then 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.

The skill of the ML algorithms was determined using the true skill statistic, or Peirce's skill score (Woodcock 1976). The true skill statistic uses all four components of a confusion matrix[5] to determine the skill score. This score is independent of climatology event frequency, can expressed similar to the equitable threat score, and can be expressed as the probability of detection of the event (POD) minus the probability of false detection (POFD). The true skill score can be expressed as:

$$skill = \frac{TP}{TP + FP} - \frac{FP}{FP + TN} = POD - POFD, \tag{3.2}$$

[5]Also called truth tables or contingency tables.

where TP is the number of true positives (or correct FD predictions), FP is the number of false positives (or false alarms), and TN is the number of true negatives (or correct non-FD predictions). 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 (note this could be interpreted as the other skill metrics did not detect a significant difference in model performance).

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 minutes the predicted labels averaged in time. Statistical significance of the differences was determined from the bootstrap method[6] at the $p < 0.05$ level, with a difference of zero used as the null hypothesis (therefore the desired result is no difference, or that there is not statistically significant difference). The Monte-Carlo bootstrapping was applied to each grid point individually (i.e., a distribution was determined for each grid point, and the p-value was determined relative to that distribution) so that each p-value is equally applicable across all grid points. Spatial distributions of other metrics, such as accuracy, precision, recall, and area under the ROC curve were examined with consistent results found.

Finally, feature importance was obtained via multiple methods. The first method used Shapely values, determined from the SHAP package Lundberg and Lee (2017).

[6]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.

Shapely values assume 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 value of the attributions. A second set of feature importances were 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 metrics (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 tree criterion caused by certain features (Pedregosa et al. 2011; McGovern et al. 2019).

Feature importance calculations are computationally expensive, and because of this, some simplifying assumptions were made. For example, calculations for the Shapely values were performed on the spatially averaged time series (i.e., data averaged over the domain) for test folds for each rotation, 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. Feature importance for ANNs was only calculated using Shapely files, due to a limitation in the permutation importances, and feature importances were omitted from RNN and U-net calculations as the calculations for 3 dimensions yield ambiguous results.

Chapter 4

Standard Machine Learning Performance

The frequency of FD events according to each method can be seen in Figure 4.1. 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 each method highlights some hotpots more strongly than others (some omit hotpot regions altogether), agreeing with the general findings in Osman et al. (2021) and Alencar and Paton (2022). These frequency climatologies are one of the targets that the ML algorithms in this chapter and chapter 5 should be able to predict.

4.1 Statistical Performance

The skill scores for the standard ML algorithms are shown in Table 4.1. Overall, of the three standard ML algorithms explored, the best performing in most cases was the Ada boosted decision trees, with the exception being with the N20 and P20 methods for which RFs performed best. 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, including the DL algorithms). This was likely due to mistiming of the ML algorithms of FD events by a few pentads up to two 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. 4.2). Regions that generally experienced more frequent FD events, most notably the Great Plains, were over predicted by the ML

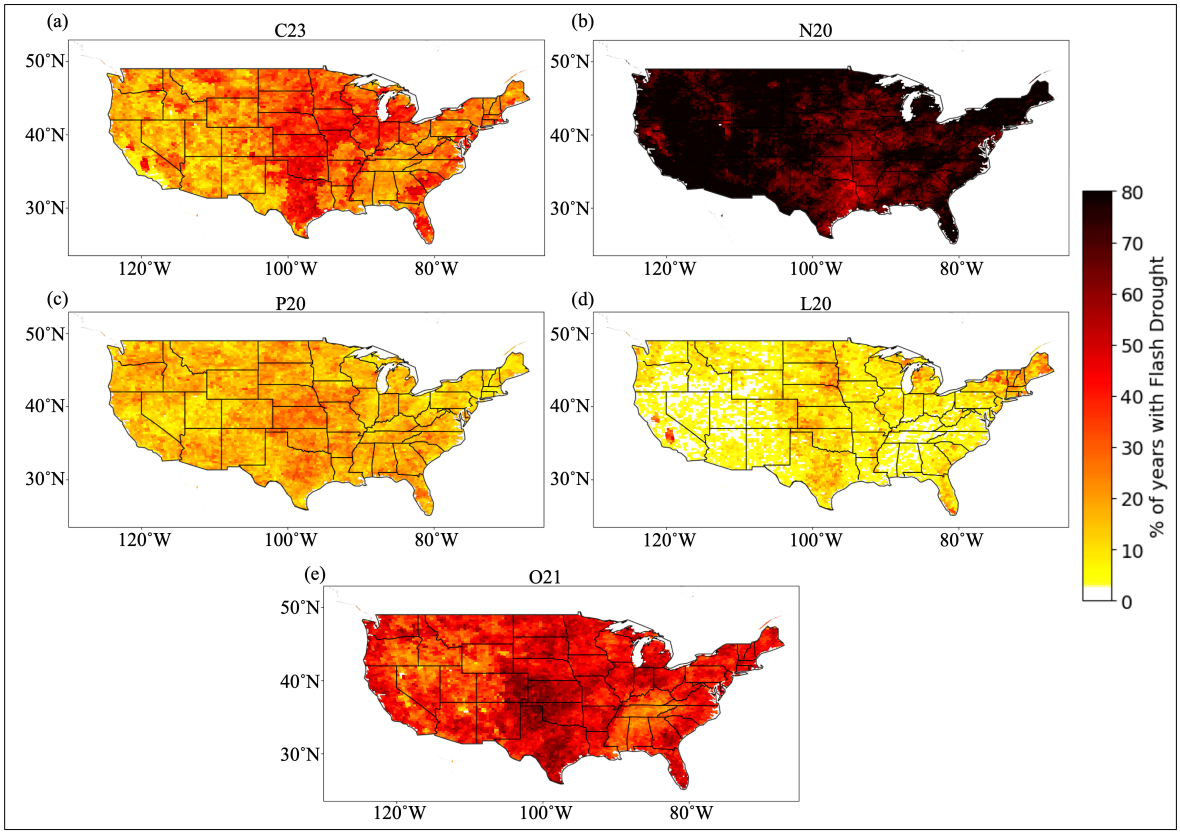


Figure 4.1: Percentage of years from 1971 – 2021 in which FD occurred according to the a) C23 method, b) N20 method, c) P20 method, d) L20 method, and e) O21 method.

	RFs	Ada Boosting	SVMs
C23	0.20 (±0.016)	**0.33 (±0.019)**	0.13 (±0.016)
N20	0.17 (±0.013)	0.13 (±0.007)	**0.22 (±0.012)**
P20	**0.11 (±0.016)**	0.06 (±0.008)	0.05 (±0.013)
L20	0.15 (±0.014)	**0.27 (±0.016)**	0.07 (±0.011)
O21	0.14 (±0.009)	**0.28 (±0.016)**	0.08 (±0.011)

Table 4.1: True skill statistic over all grid points and pentads for each standard ML algorithm and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for is identification method is bolded.

algorithms (these over predictions are also where the difference in true and predicted labels were statistically significant; Fig 4.3). In contrast, the ML algorithms performed better outside of climatological hotspots for some FD identification methods (C23, P20, and L20), and under predicted FD in other identification methods (N20 and O21). These same patterns were also seen in RFs and SVMs, though the RFs handled over predictions better than the Ada boosted Trees, and SVMs placed more emphasis on the southern portions of the Great Plains (Fig. 4.4 and 4.6). These same patterns were also found in standard performance metrics (e.g., accuracy, precision, recall, and AUC; not shown).

The last feature to note is the interpretability of the ML algorithms (Fig. 4.8 for Ada boosted trees). For this study, several approaches at estimating feature importance were examined and compared. The ranking of the most importance feature importance varied notably from one FD identification method to another. Where some had P as the most important variable, and PET as the second most important, others ranked SM as the most important. The SHAP feature importance and permutation importance

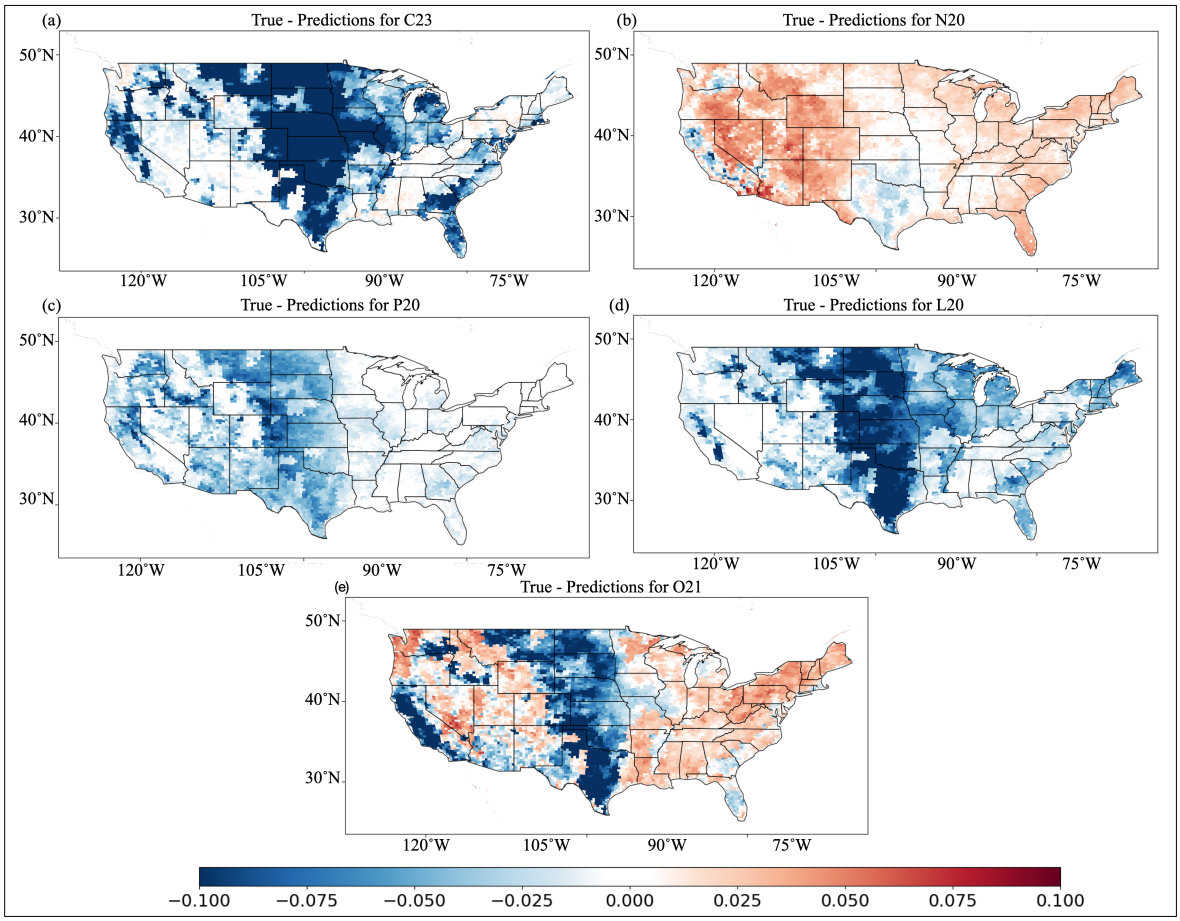


Figure 4.2: Composite mean difference of true minus predicted FD labels for each FD identification method using Ada boosted tree predictions.

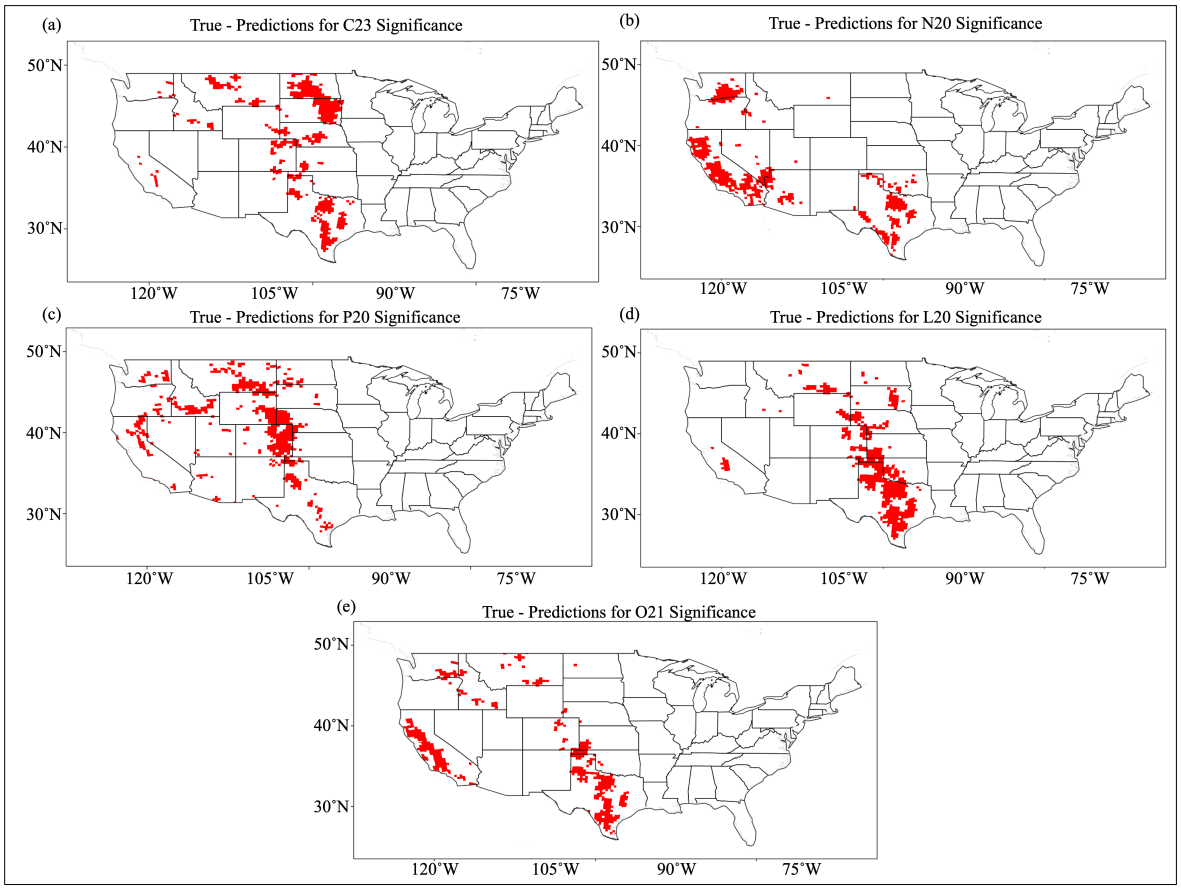


Figure 4.3: 95% statistical significance of composite mean differences between the true FD labels of the Ada boosted tree predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

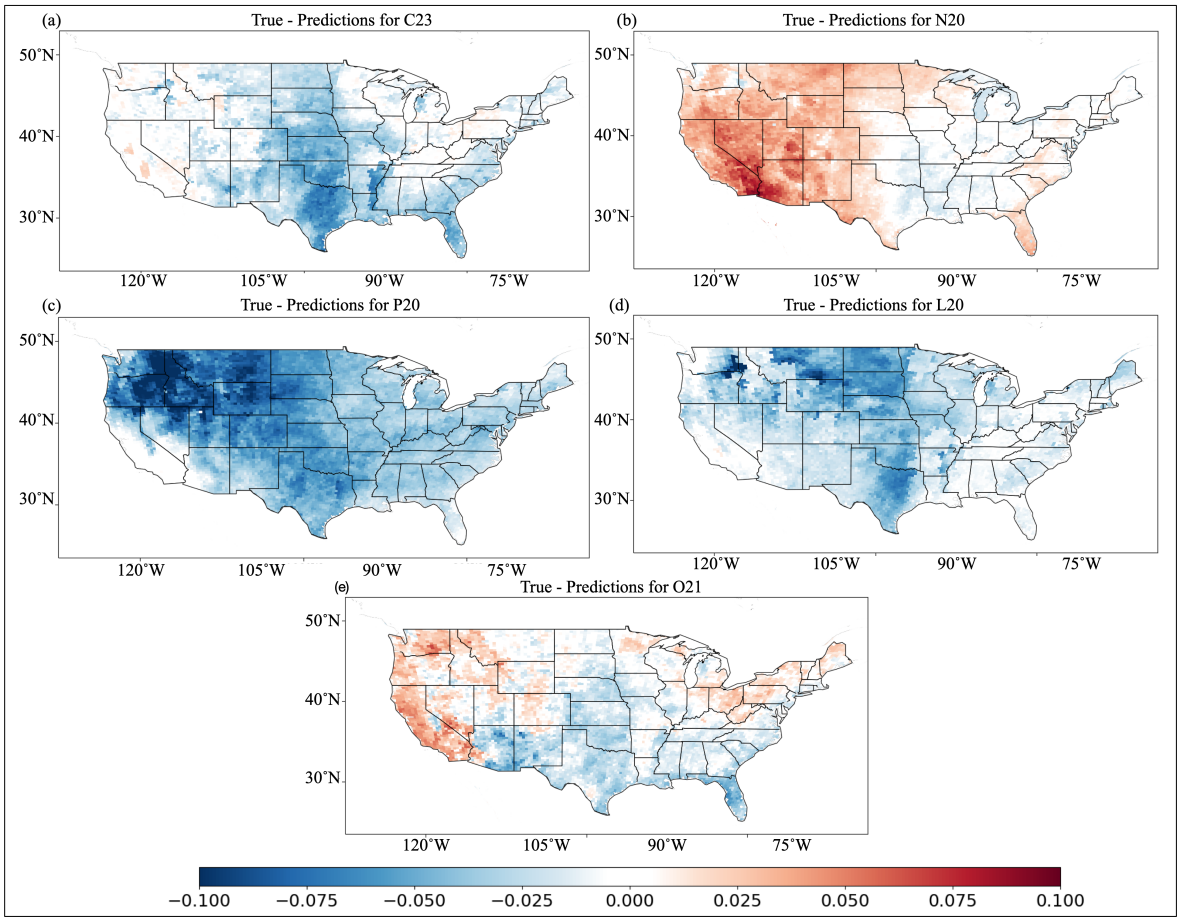


Figure 4.4: Composite mean difference of true minus predicted FD labels for each FD identification method using RF predictions.

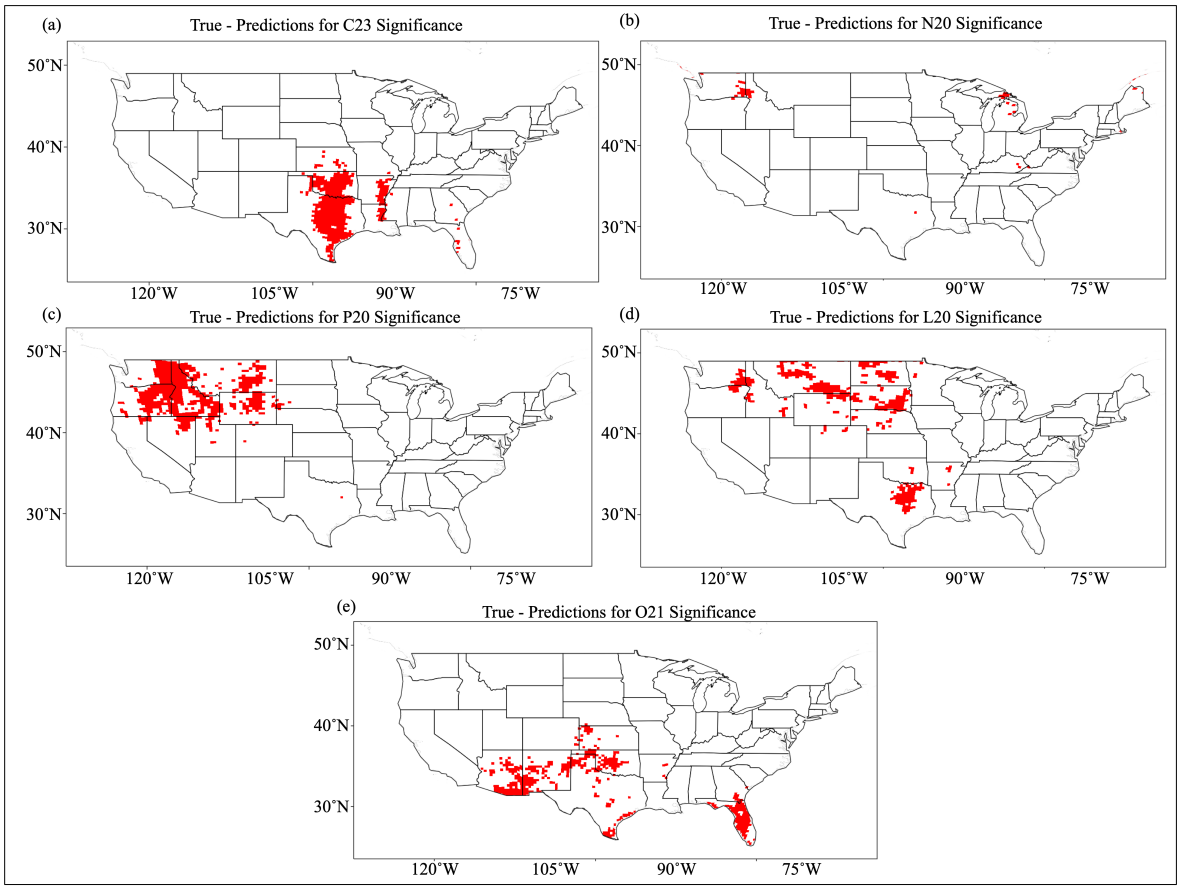


Figure 4.5: 95% statistical significance of composite mean differences between the true FD labels of the RF predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

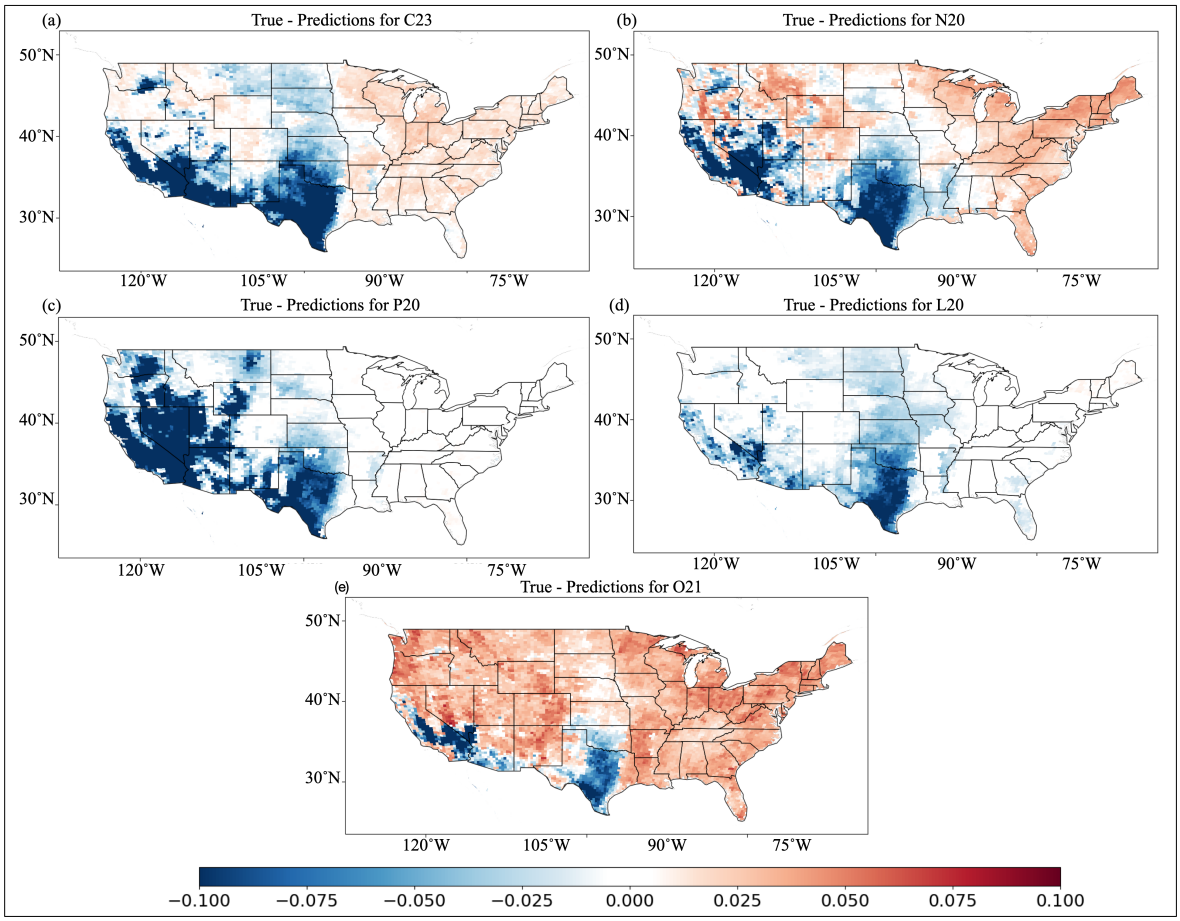


Figure 4.6: Composite mean difference of true minus predicted FD labels for each FD identification method using SVM predictions.

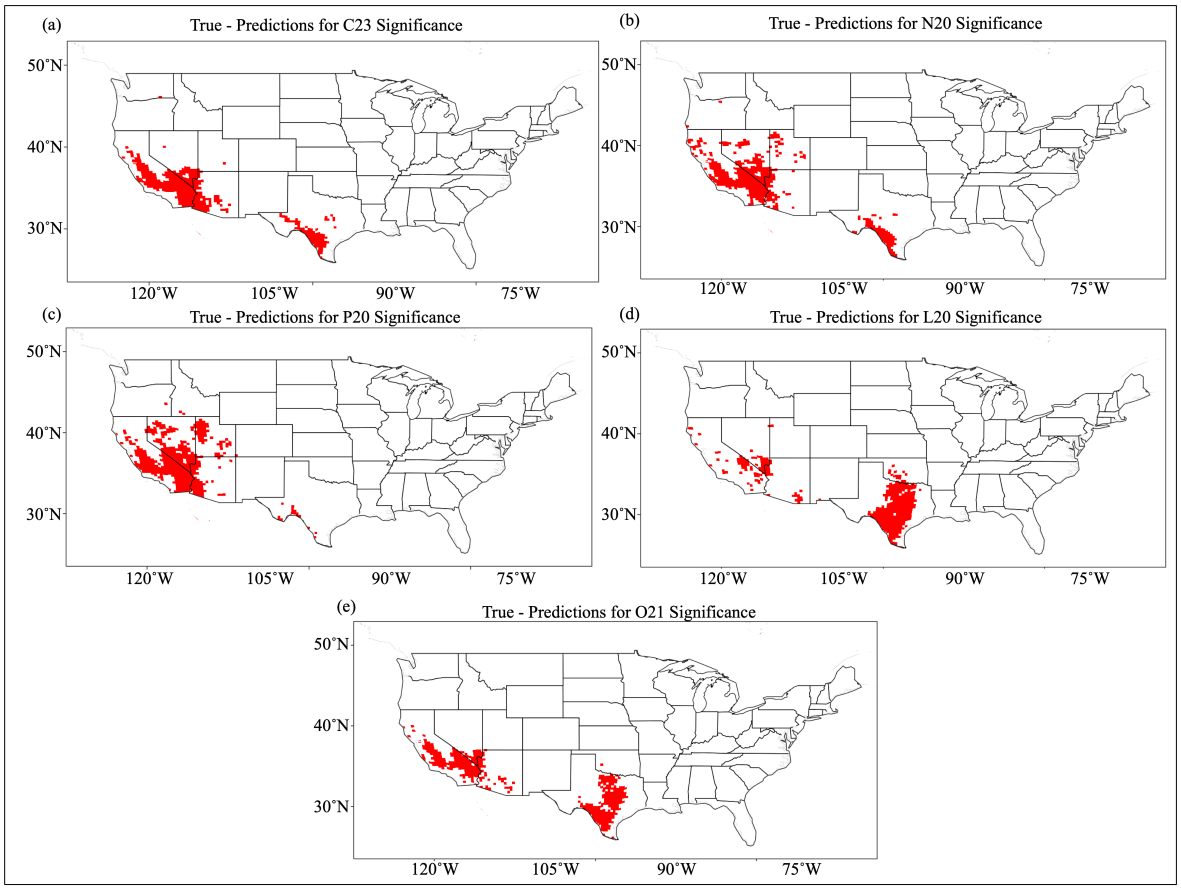


Figure 4.7: 95% statistical significance of composite mean differences between the true FD labels of the SVM predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

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 < 0.2) and statistically identical to 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. The feature importance can also explain the performance of the other models. For example, the SVMs only learned on details based on SM and δSM (Fig. 4.9). This may explain why they focused on the southern Great Plains, and much of the SVMs poorer performance (they were not incorporating patterns or synthesizing information from other variables almost at all). In contrast, the RFs had varying results, though the GINI importances had similar results to the Ada boosted trees (Fig. 4.10). However, the SHAP importances for the RFs showed multiple results, with the RFs were also reliant on ΔPET, and T as well as SM (C23, L20, and O21 methods) or P (P20 method). This suggests the RFs may have had trouble learning the surface interaction that characterizes FD given it is reliant on simpler variables (T, P, and PET), resulting in their lower skill scores though they did learn the correlation between FD and SM.

Furthermore, the standard ML algorithms learned to predict FD for some identification methods better than others. This finding may have been 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 feature importance; Fig. 4.8, 4.9), and 4.10. In addition, SESR (for the C23 method) is strongly correlated with SM, allowing an emphasis on SM to produce a larger skill score. Meanwhile, predictions for

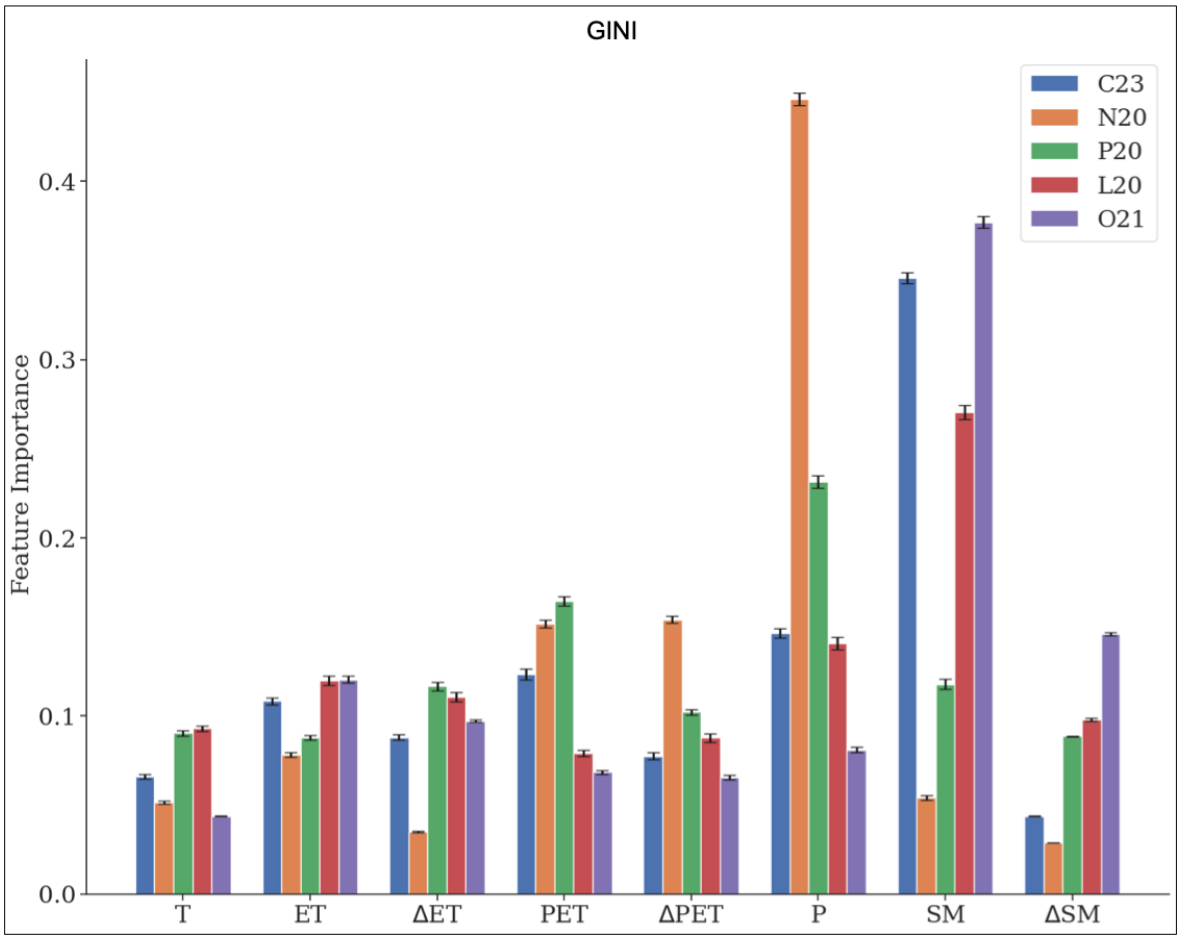


Figure 4.8: Average feature importance for the Ada boosted decision trees according to the GINI method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations).

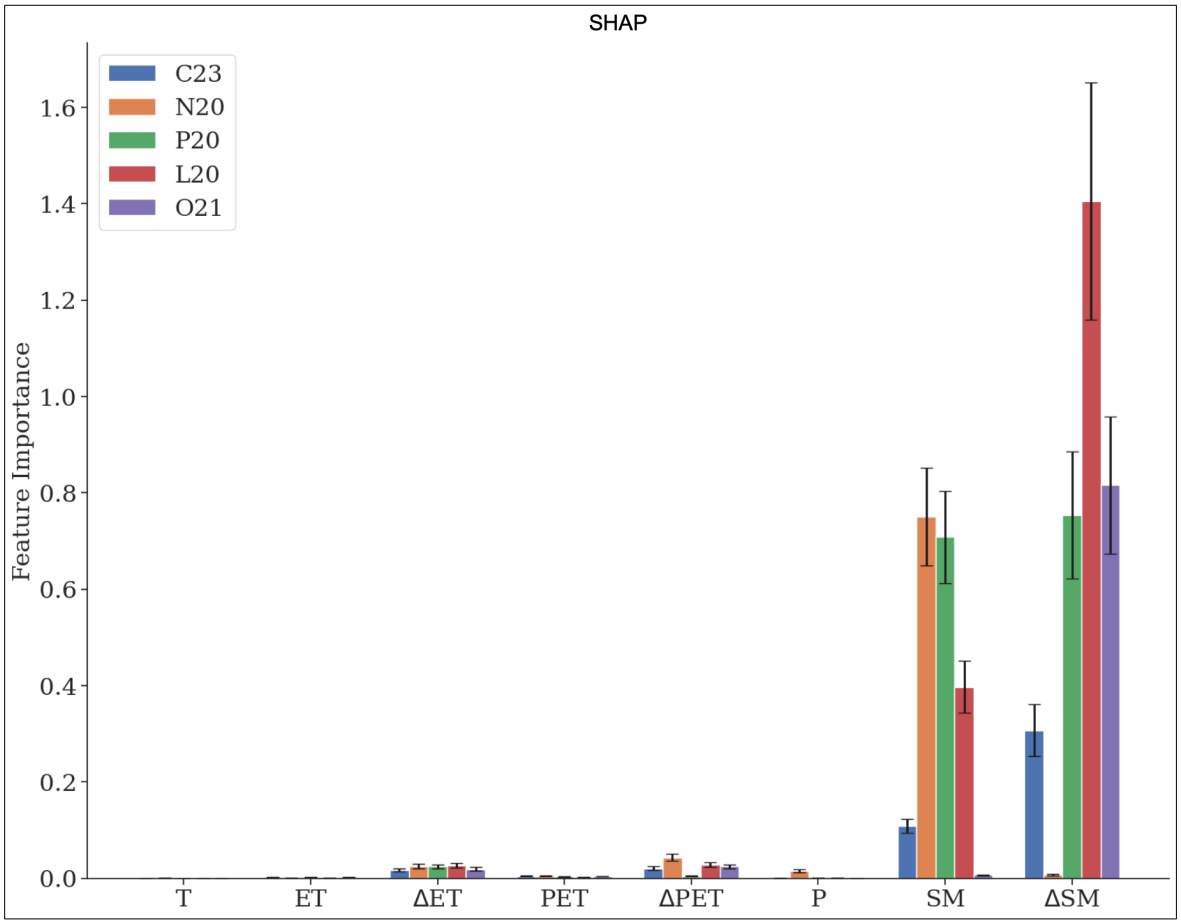


Figure 4.9: Average feature importance for the SVM models according to the GINI method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations).

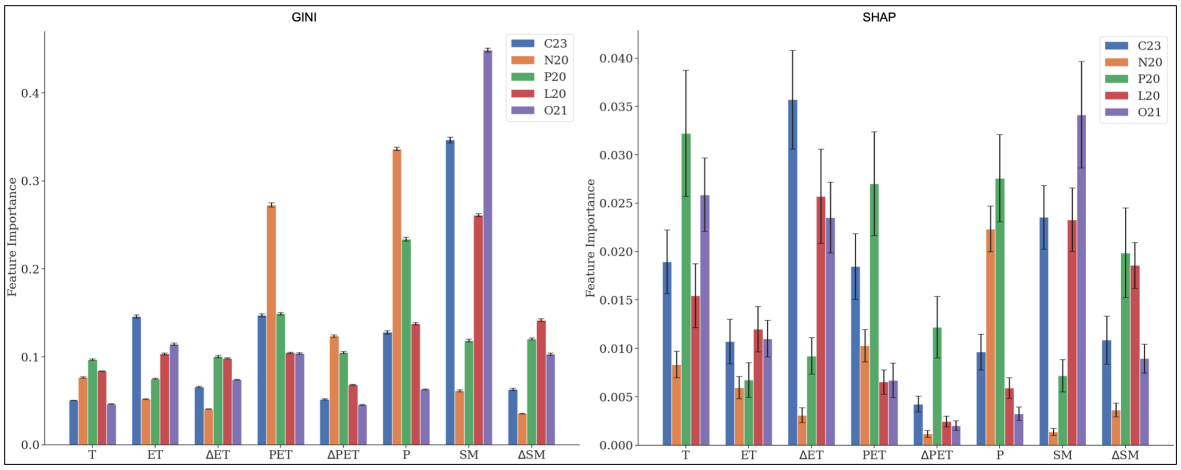


Figure 4.10: Average feature importance for the RF models according to the GINI method (left), Shapely values (center). Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations).

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 (i.e., it was relying on atmospheric supply to make predictions and did not use atmospheric demand with it), or learning as much of the surface interactions while RFs only incorporate P, and SVMs only used SM, 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 4.1). One of the FD identification methods the ML algorithms learned most successfully was the C23 method, which had the highest skill statistic for two of the ML algorithms and had fewer under predictions of FD events, despite the ML algorithms not incorporating ET or PET (the variables used for SESR) as strongly for that method.

4.2 Predicted Climatology

Apart from performance metric, climatological predictions of the ML algorithms were also examined to determine how well they reproduce the statistical patterns of FD in space and time. The predicted climatologies (Fig. 4.11) showed similar patterns to those discussed. Compared to the true labels (Fig. 4.1), the ML algorithms over predicted FD events in climatological hotspots. The standard ML algorithms also highlighted major agricultural regions in the western United States (such as Central Valley in California and southern Idaho) and over predicted FD in these locations as well. Given the development of FD via evaporative feedback from the vegetation (i.e., crops that tend to transpire more than local vegetation), it makes sense that these regions would be hotspots for FD, as is the case for other major agricultural regions (e.g., Christian et al. 2023; Lowman et al. 2023; Mahto and Mishra 2023; Yuan et al. 2023), but they were not present in the true labels (Fig. 4.1). However, this suggests that the ML algorithms were able to learn these hotspots (i.e., it learned what spatial regions experience FD more frequently), and they learned to identify FD more frequently for major agricultural regions. While the composite differences showed the RFs not emphasizing hotspots as much as Ada boosted trees, this was not mirrored as well in the predicted climatologies of RFs and SVMs exaggerated hotspots more than Ada boosted trees (Fig 4.12 and 4.13).

The standard ML algorithms overemphasized the seasonality of FD events, identifying FD more frequently during favored months and under predicting FD in months less emphasized in that climatology (Fig. 4.14). That said, this also showed 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

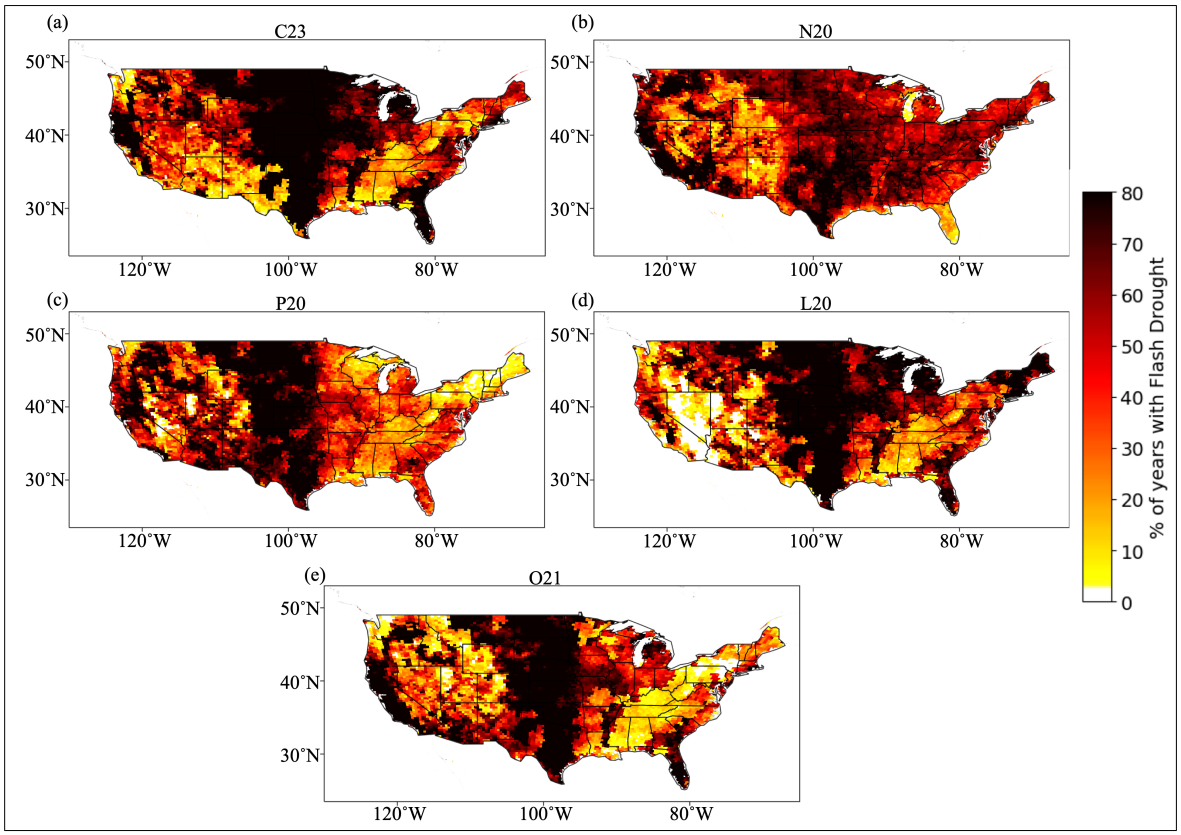


Figure 4.11: Predicted frequency climatology of FDs by the Ada boosted trees for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

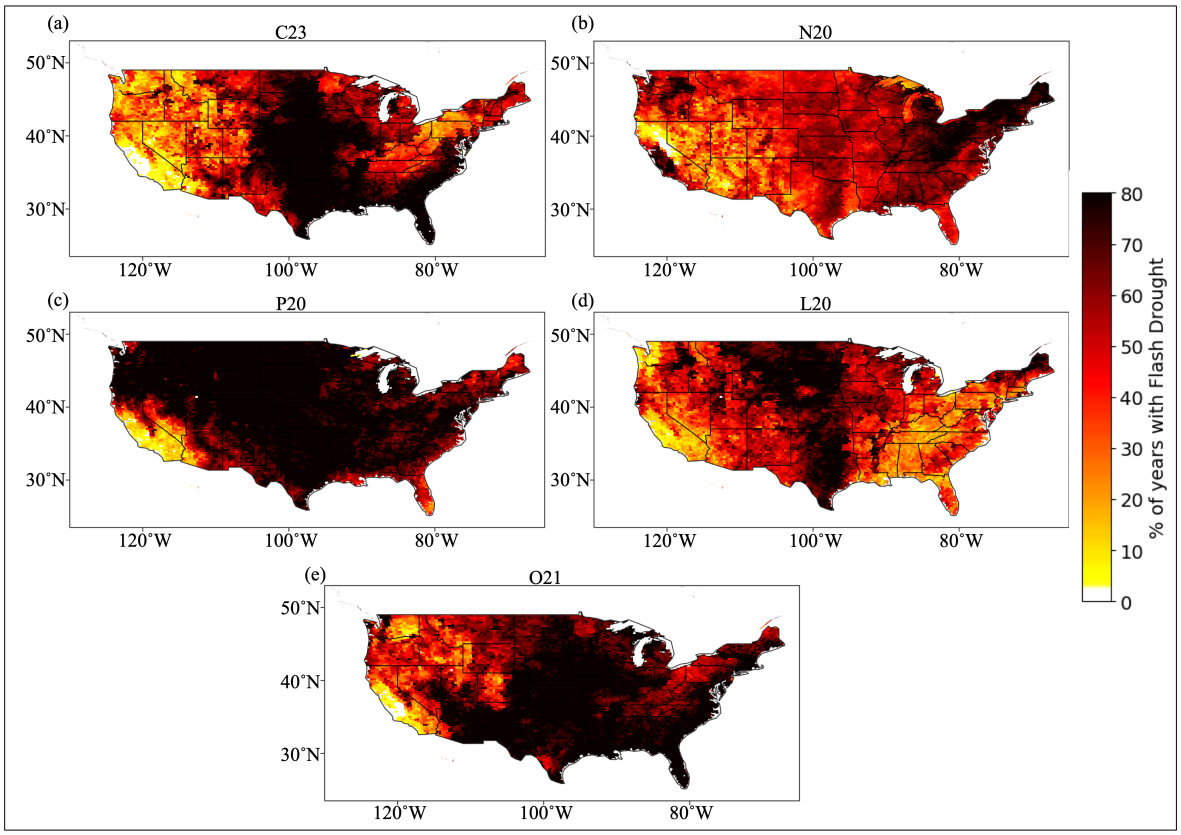


Figure 4.12: Predicted frequency climatology of FDs by the RFs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

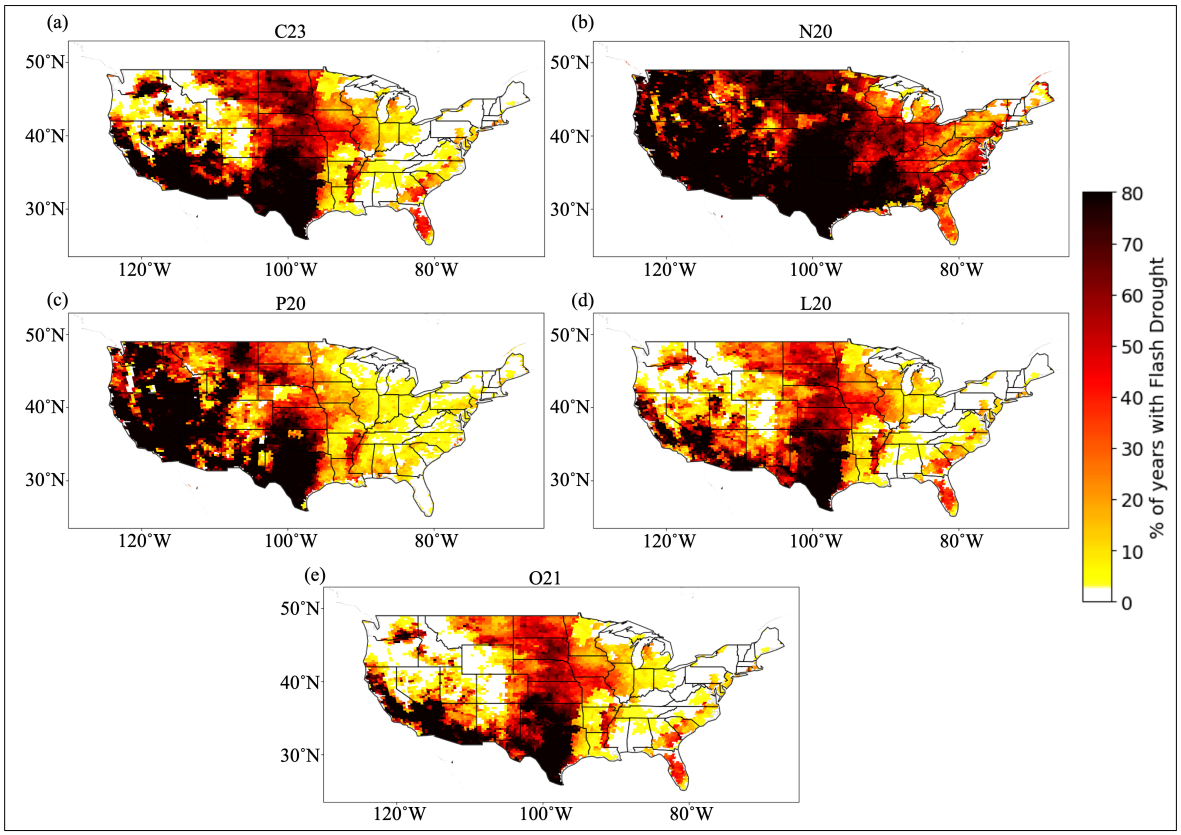


Figure 4.13: Predicted frequency climatology of FDs by the SVMs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

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 are over predicted, as these would see larger declines in SM due to intensive agriculture and irrigation than regions dominated by native vegetation. In terms of the FD coverage in time, the ML algorithms learned the general patterns of FD coverage, 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. 4.15). RFs and SVMs had similar seasonality and temporal results as the Ada boosted trees, except that RFs were more successful in learning FD seasonality for the N20 method (Fig. 4.16, 4.17, 4.18, and 4.19).

4.3 Case Studies

Lastly, the ability of the ML algorithms to represent more specific, individual events was also investigated. 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.

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. 4.20). By the end of the summer, the majority of the United States, especially the north central United States,

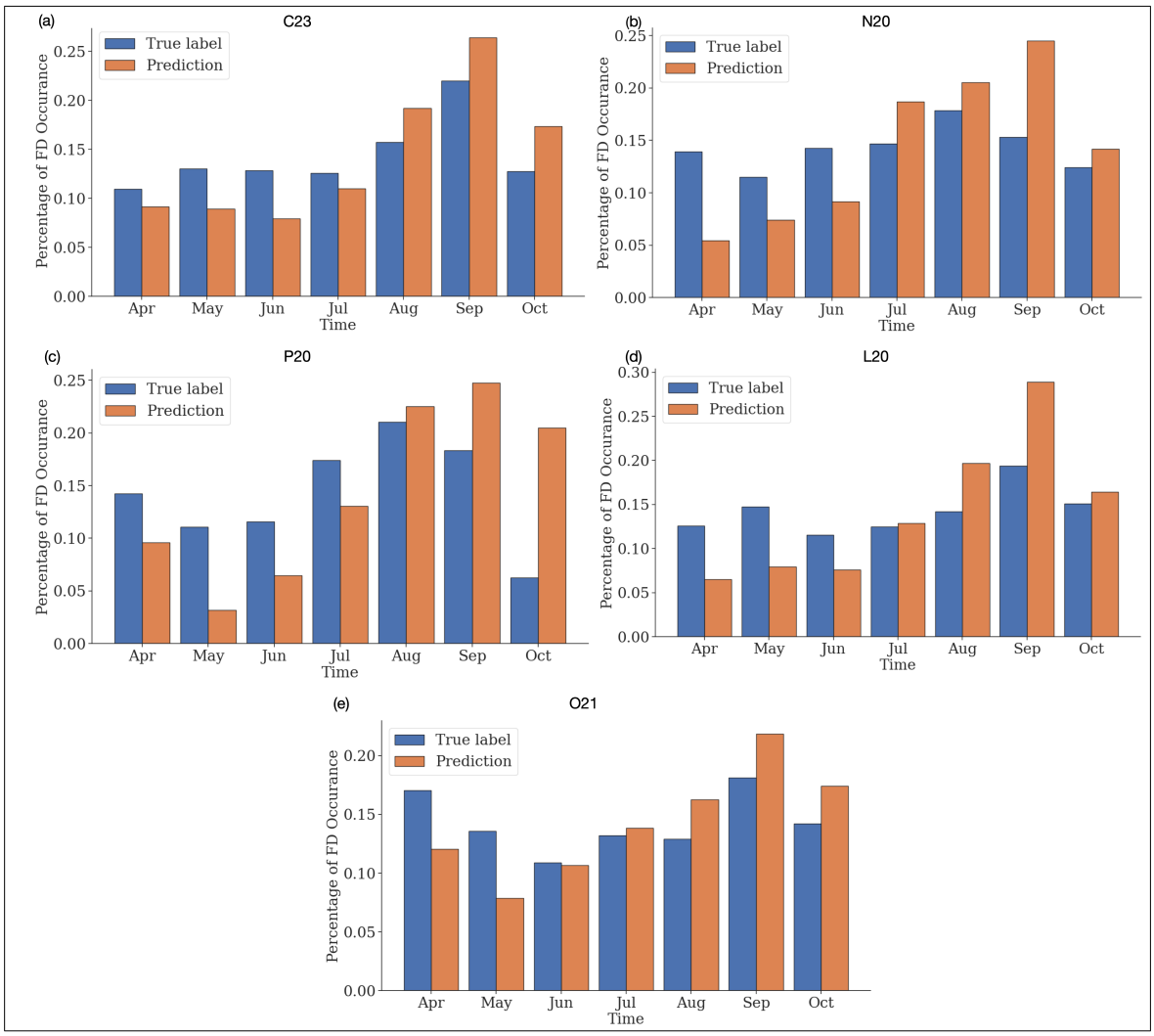


Figure 4.14: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the Ada boosted trees (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

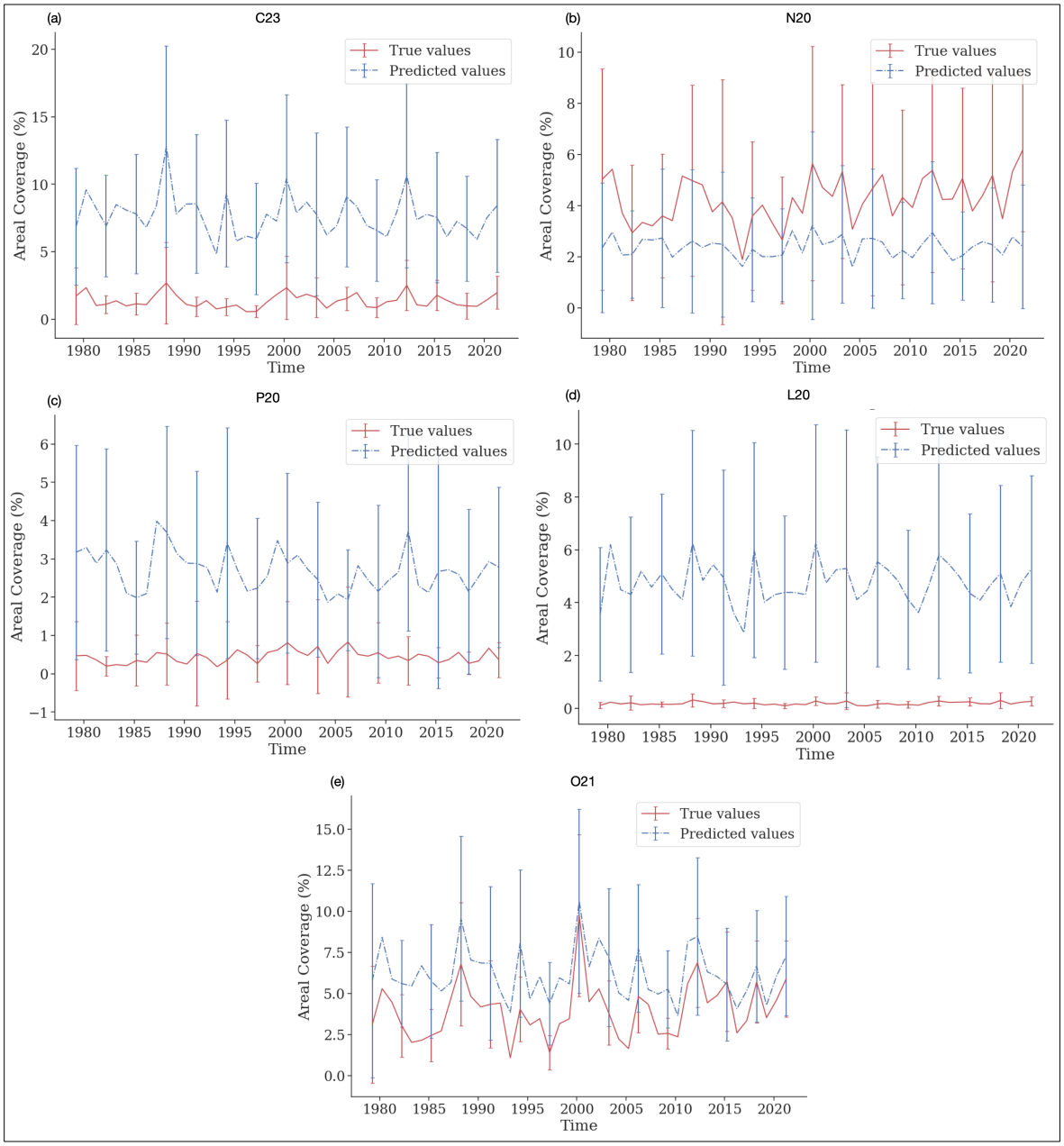


Figure 4.15: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the Ada boosted trees.

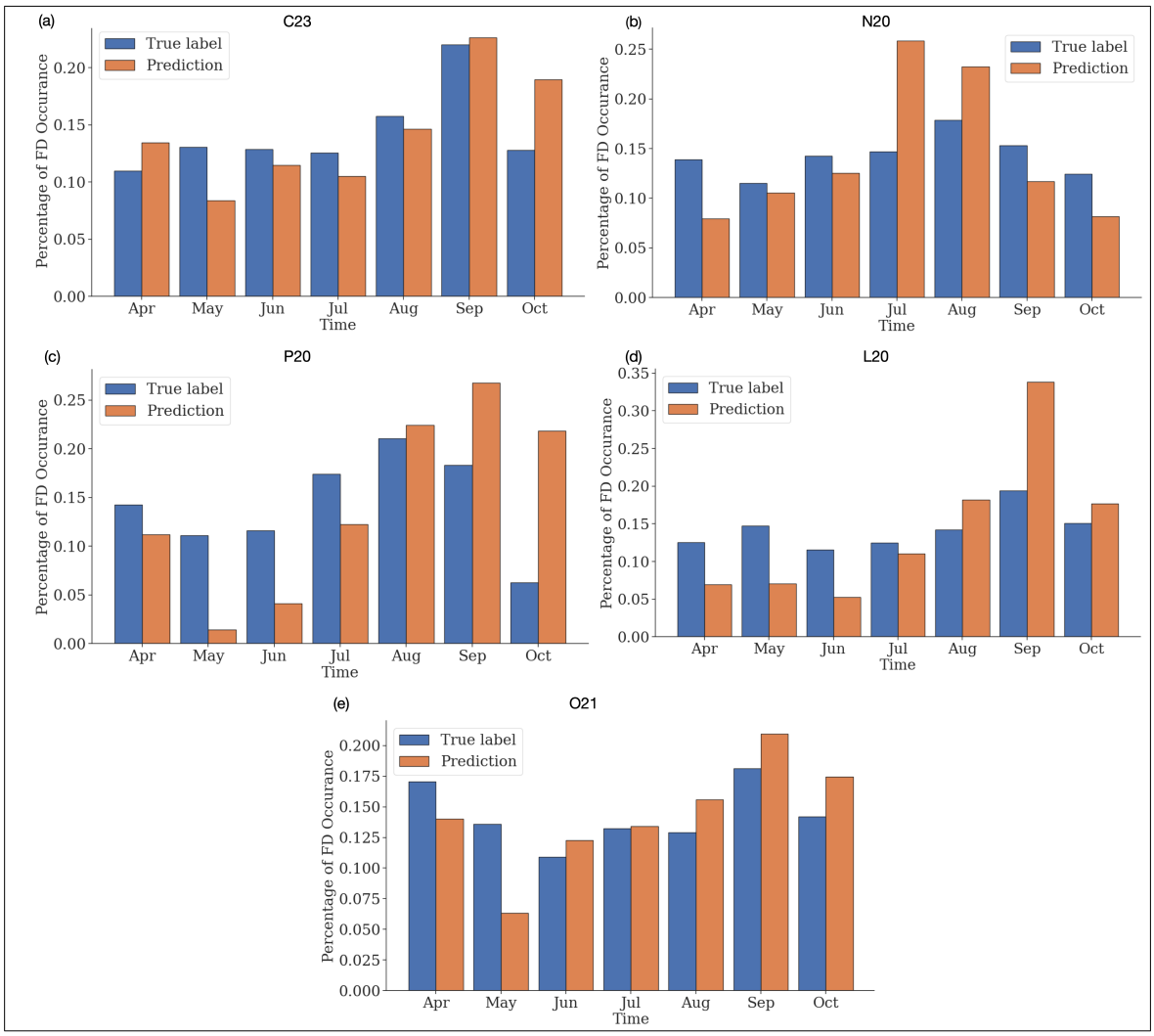


Figure 4.16: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the RFs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

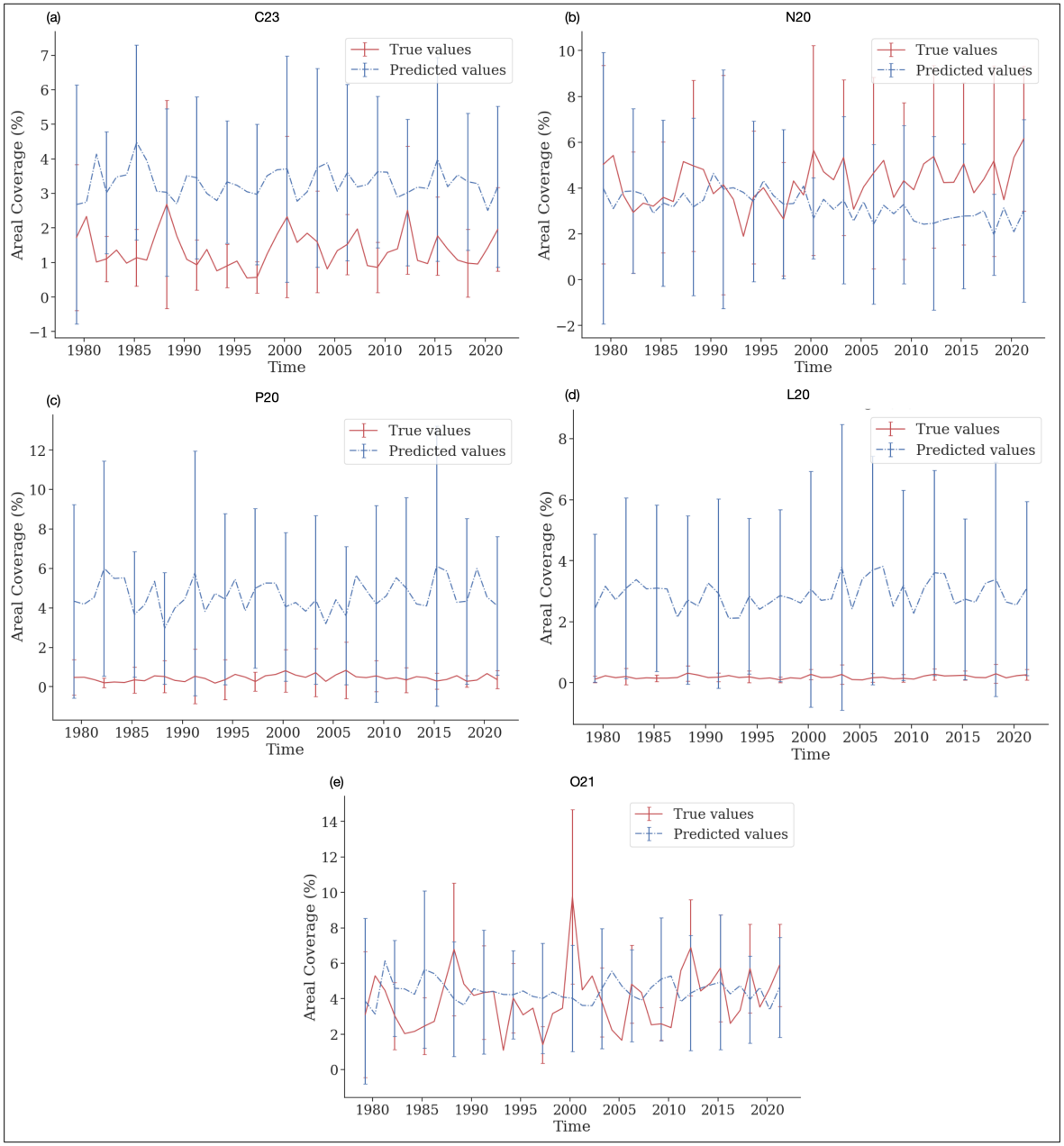


Figure 4.17: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the RFs.

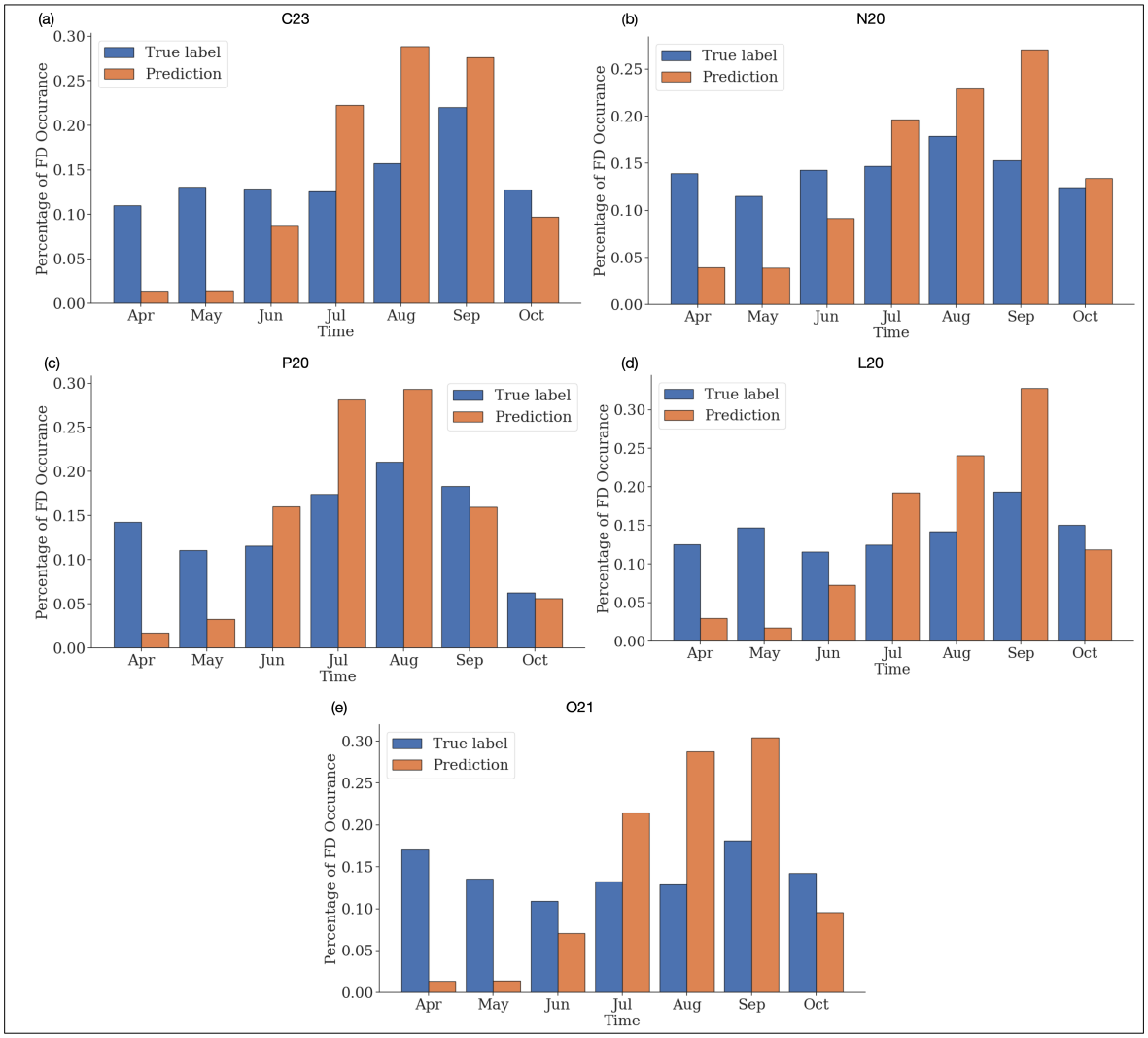


Figure 4.18: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the SVMs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

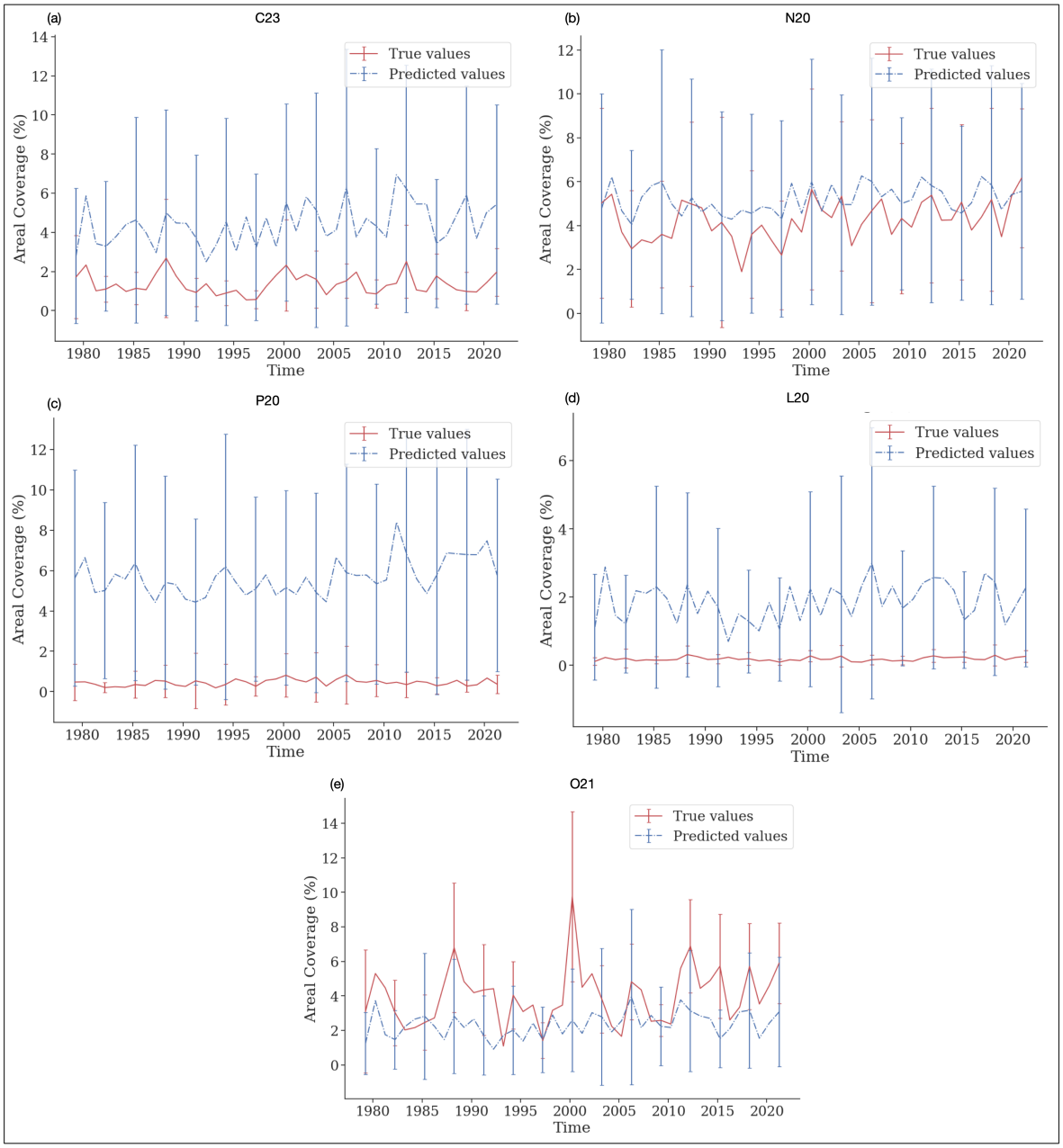


Figure 4.19: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the SVMs.

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 literature. FD was recorded for all FD identification methods, primarily in May and June when the SM started to rapidly decline in response to the increased ET due to increased T and PET (Fig. 4.23). The scope of the rapid drying, however, was different between different methods. 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.

Some of the seasonality bias can be seen as the standard ML algorithms predicted FD up to one to two months late for some methods (C23 and N20), and prematurely early in another (P20), the bias being towards the late seasonality the late seasonality the Ada boosted trees predicted as the large decline in SM wasn't until mid to late June (Fig. 4.23). The spatial coverage of the 1988 event was generally well represented by the ML algorithms, with the exception that the ML algorithms liked to make continuous patches of FD as compared to discontinuous grid in the L20 method. The RF and SVM reliance on SM can be seen in this case study, as both waited until June and July, when was SM at a minimum, to identify FD (Fig. 4.21 and 4.22). Further, the RF struggled to capture the spatial extent of the FD for the C23 method, and SVM struggled with capturing the spatial extent overall.

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

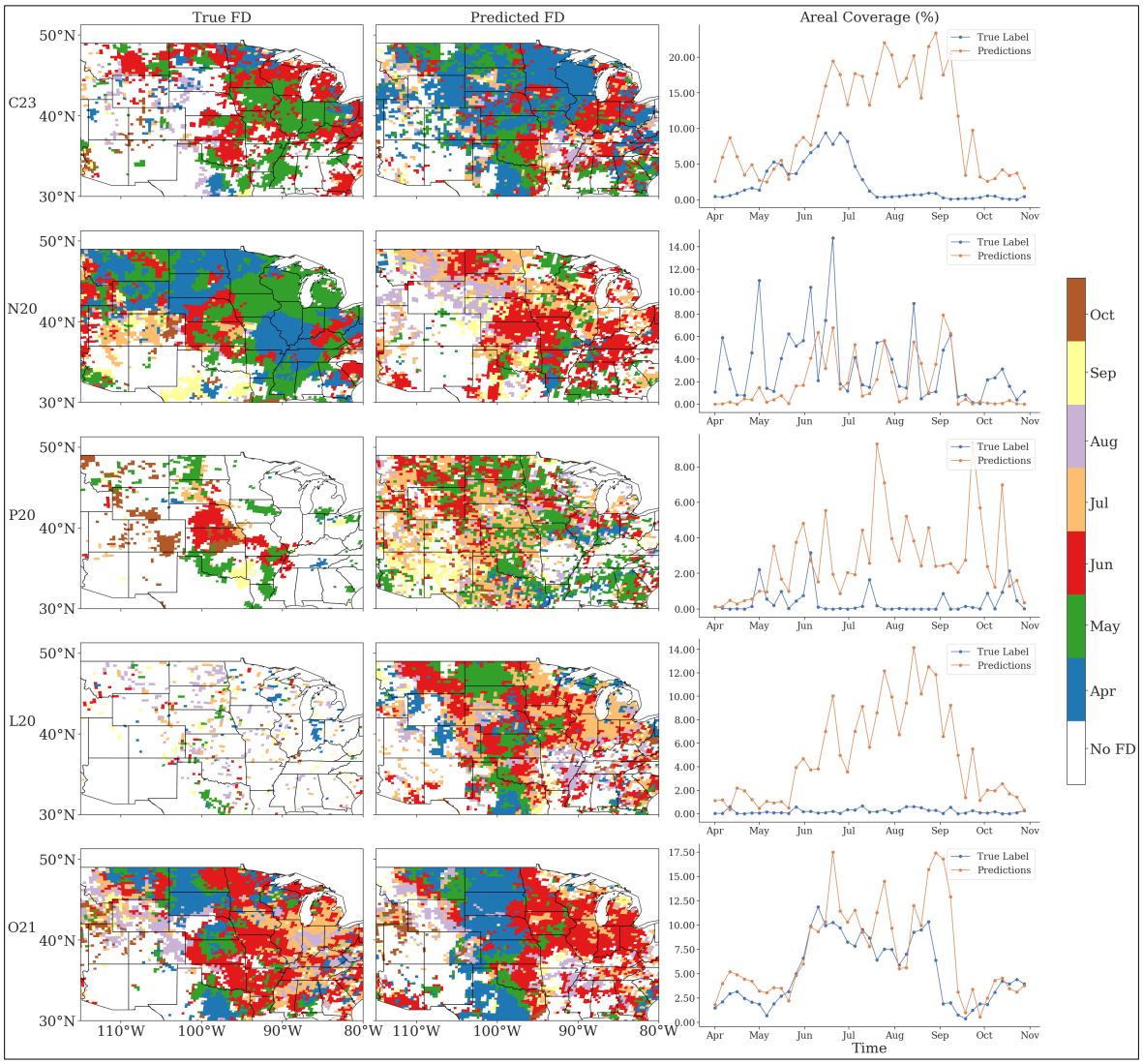


Figure 4.20: FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.

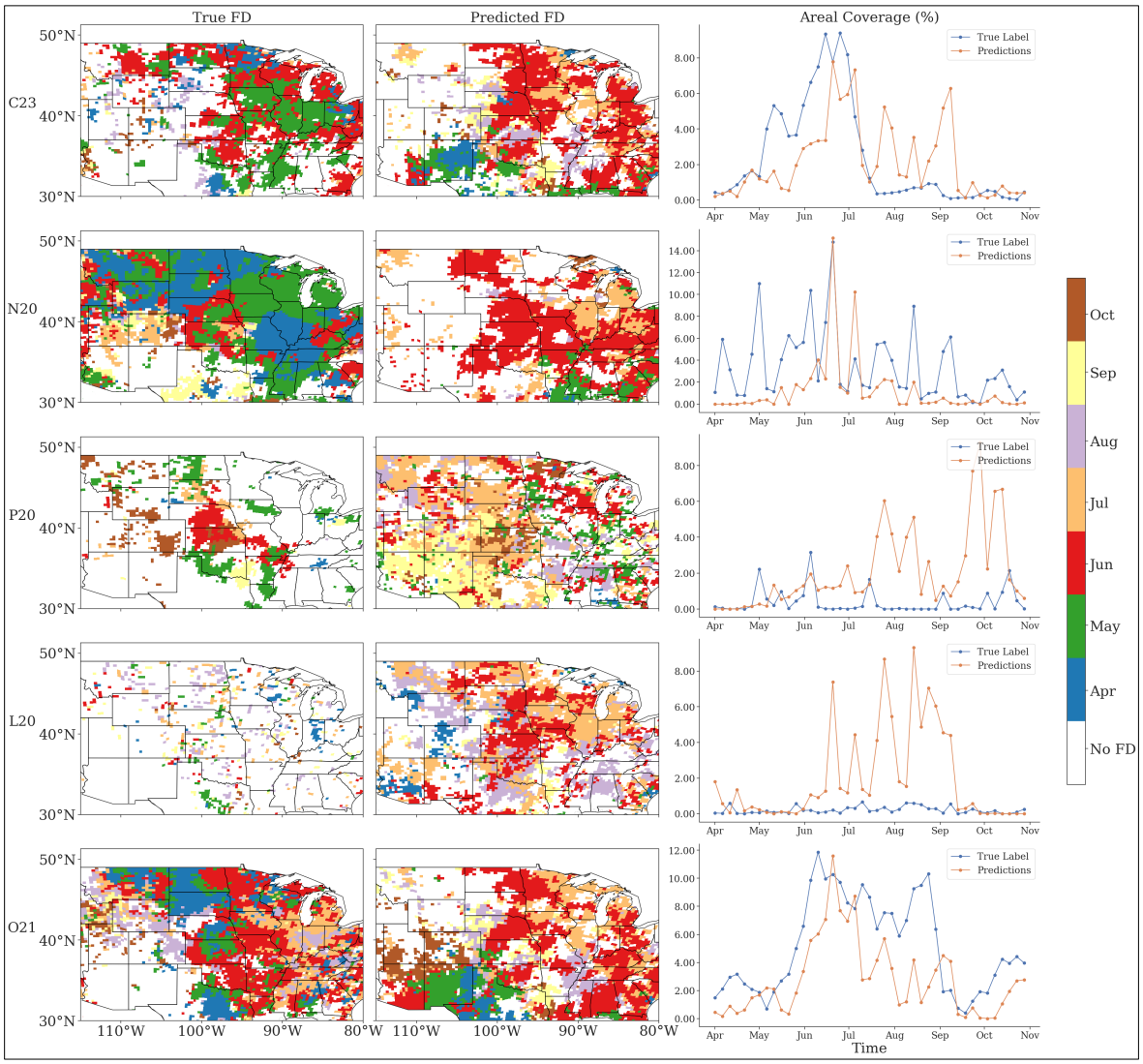


Figure 4.21: FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs.

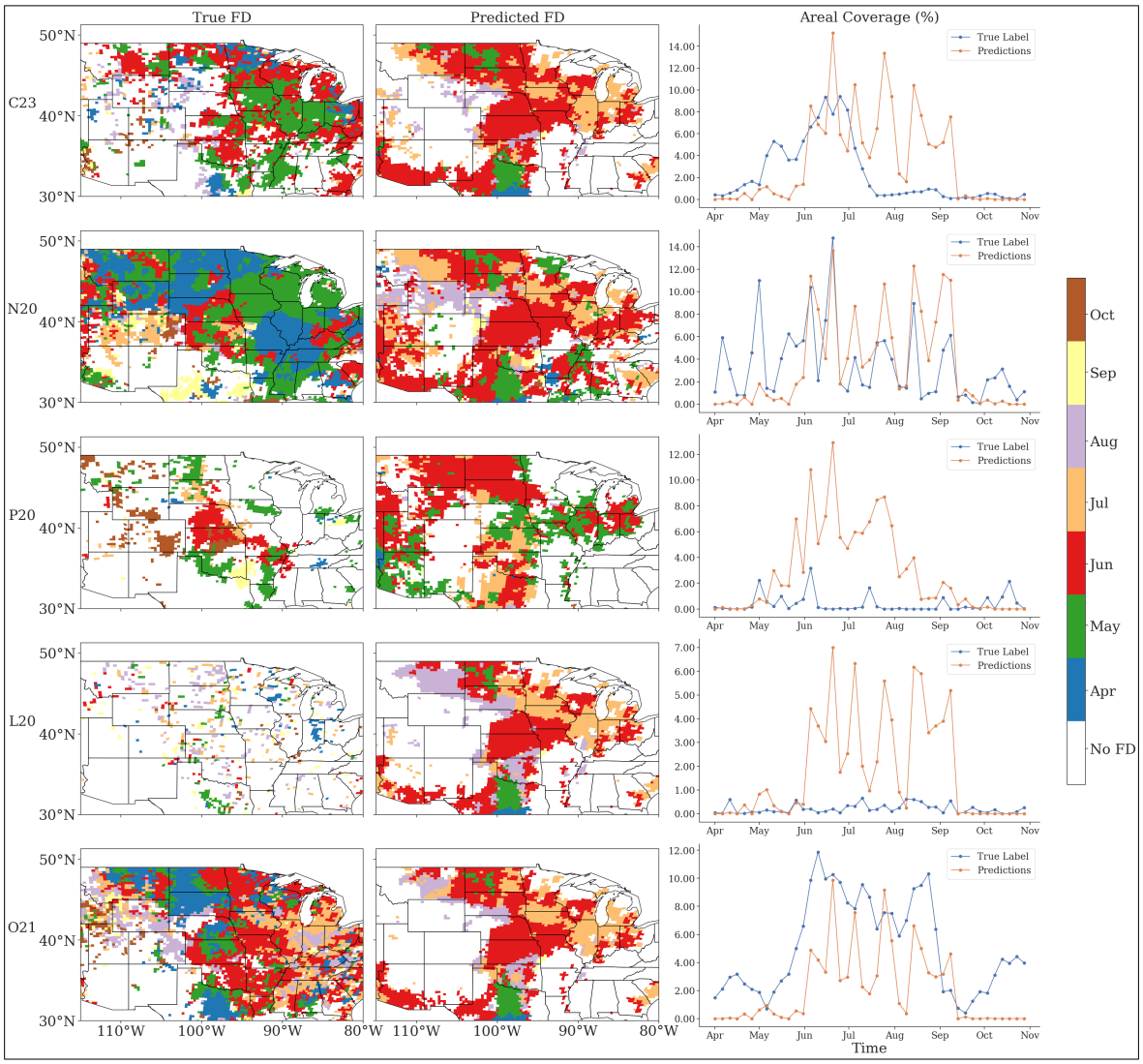


Figure 4.22: FD case study for 1988 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs.

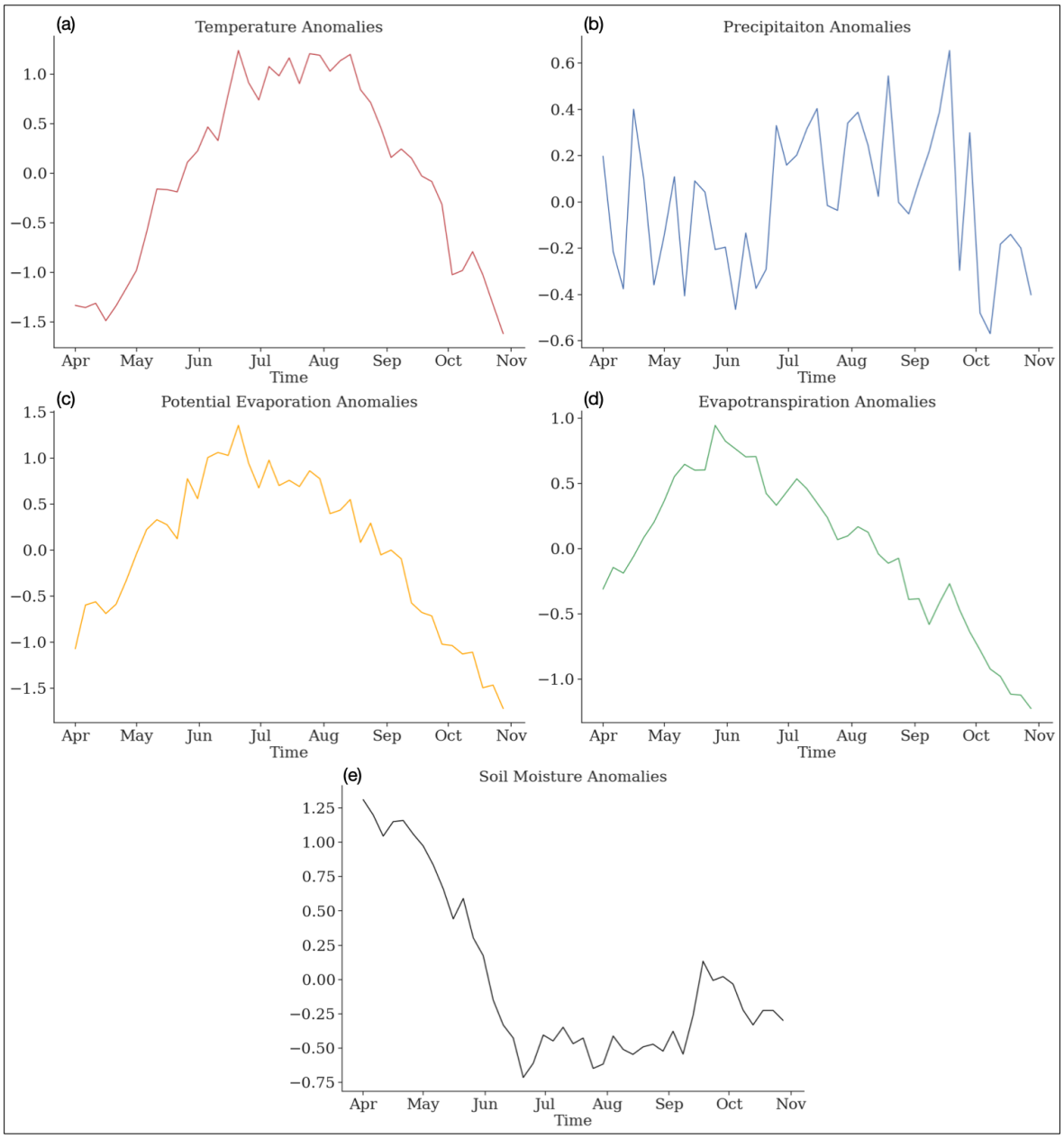


Figure 4.23: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 1988 case study.

southeastern United States settling into drought for the remaining growing season; Fig 4.24; Otkin et al. 2013; Ford et al. 2015; McEvoy et al. 2016). Note this was a largely moisture demand and soil moisture deficit driven FD (decreases for SESR in the C23 method, for example, would be due to increases in PET), as the evaporation did not start to decrease until September, despite the desiccation in the vegetation (Fig. 4.27).

For the C23 method, the ML algorithms captured the timing for this event fairly well (likely due to the early SM decline, but potentially due to the increased PET as well), but largely over predicted 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 Ada boosted trees identified FD up to two months late in many locations. For the remaining FD identification methods, the Ada boosted trees either over predicted (P20 and L20) or under predicted (N20) FD coverage and they mistimed FD occurrence by one to two months. Given the Ada boosted tree reliance on SM for the L20 method, and P for the P20 and N20 methods (Fig. 4.8), these late predictions of FD may be in response 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 Ada boosted trees predicting the L20 method, ΔSM was less important than P; Fig. 4.27b, e). The RFs struggled to identify FD in Texas, and had trouble capturing the spatial extent and timing of the event (Fig. 4.25). The SVMs did better with the timing and spatial coverage (Fig. 4.26), which might be due to the SM rapidly decreasing early the growing season, to which the SVMs responded (4.9).

In 2012, part 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

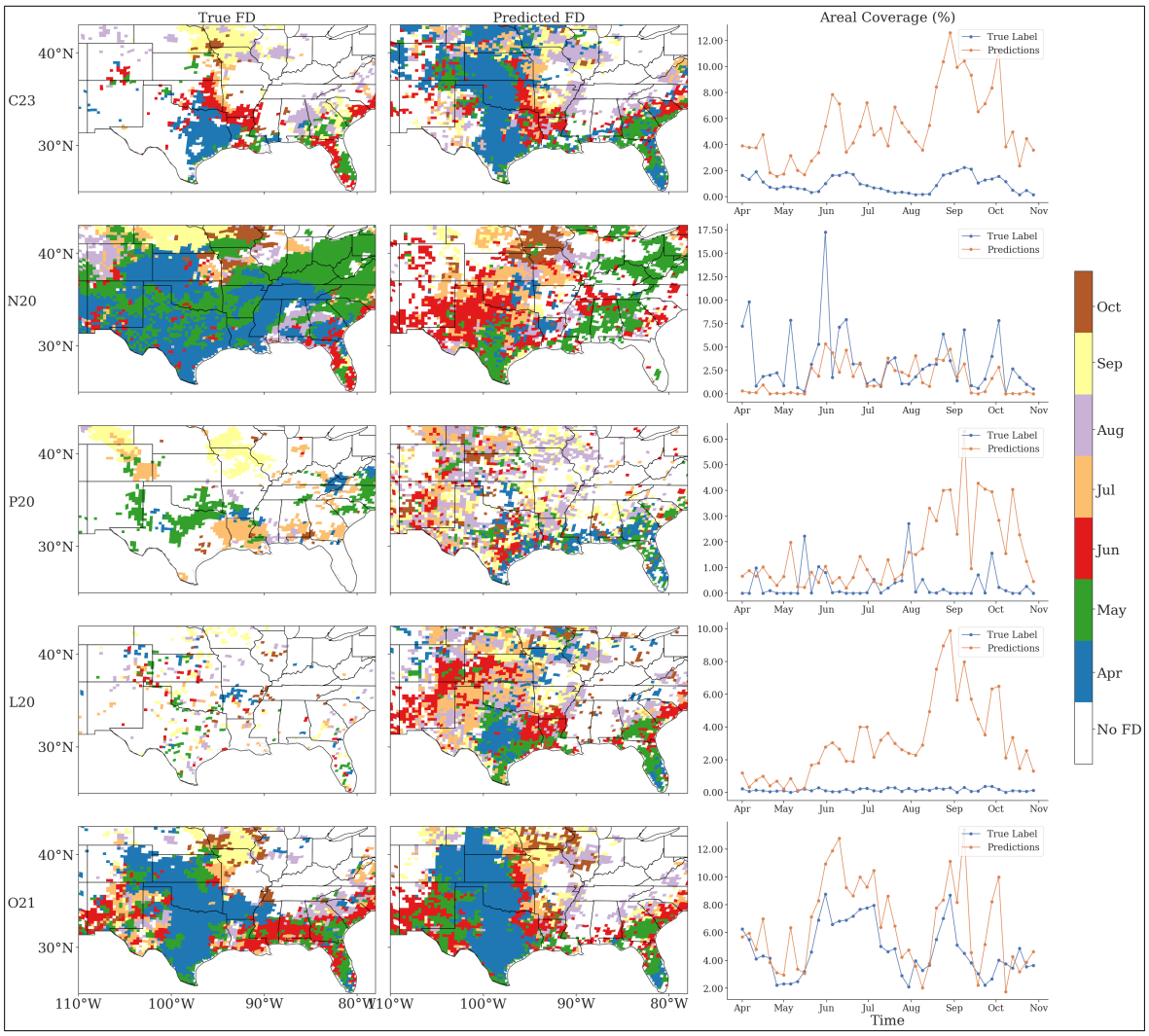


Figure 4.24: FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.

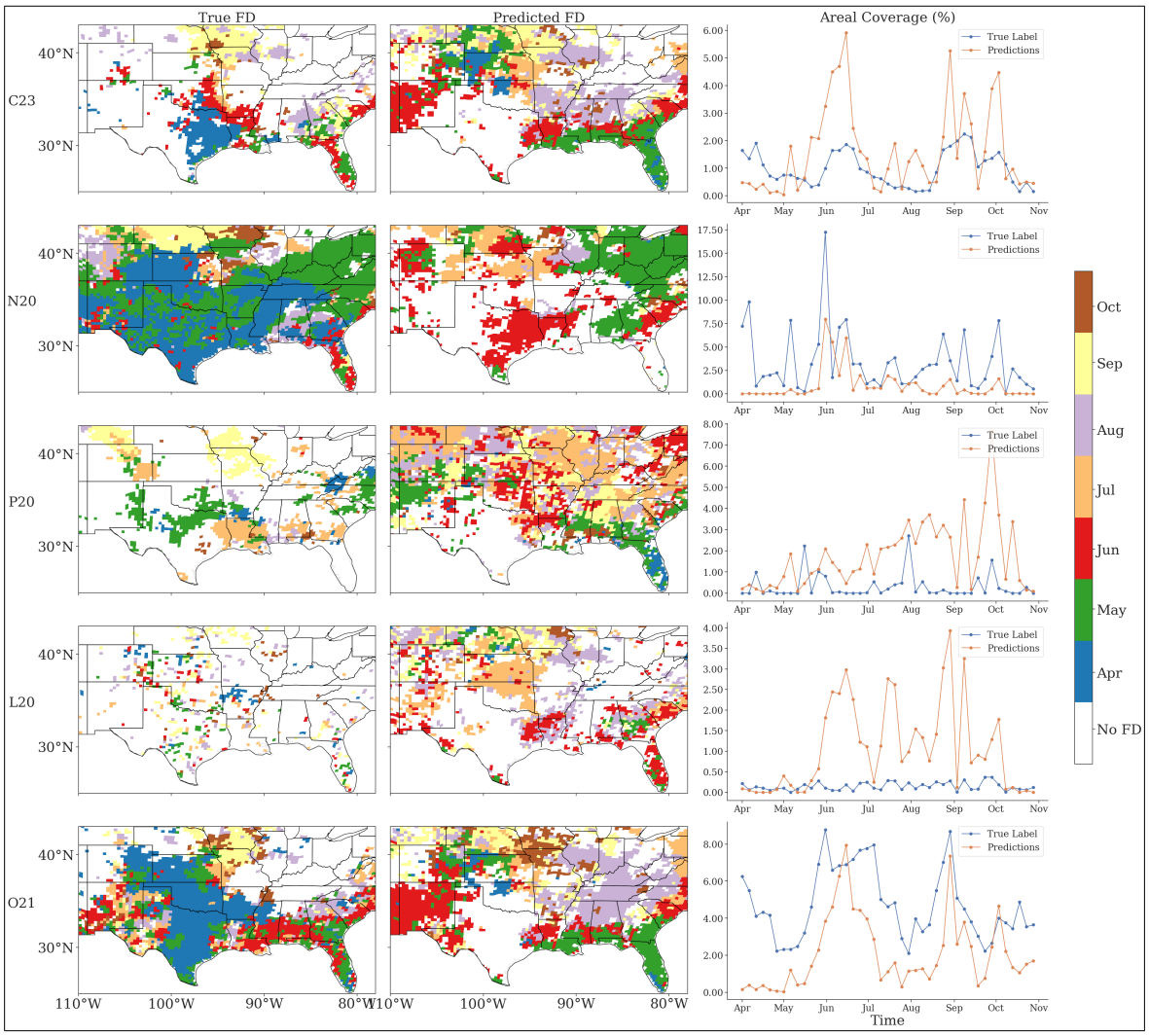


Figure 4.25: FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs.

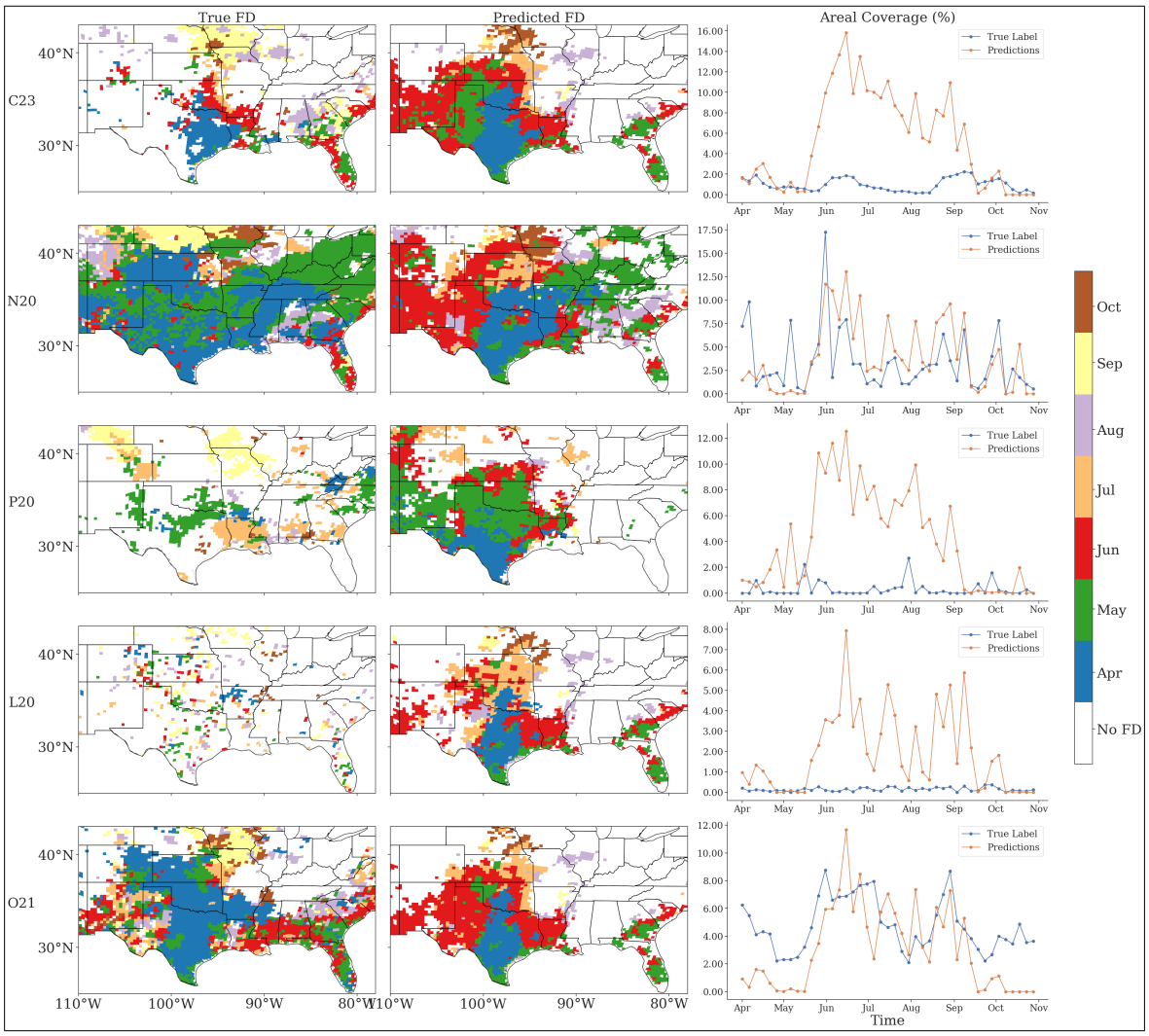


Figure 4.26: FD case study for 2011 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs.



Figure 4.27: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2011 case study.

desiccation and crop loss (Fig. 4.28; 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 Ada boosted trees generally identified FD development a month later than truth (in June and July when SM first reached its minimum, Fig. 4.31, 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 over predicted the extent of the FD for the C23, P20, and L20 methods. However, the ML algorithms were able to represent the coverage of the FD relatively well for the N20 and O21 methods. In comparison, the RFs focused heavily on June and July for when predicting FD, resulting in numerous timing issues (Fig. 4.29), when SM and P were at their minimum and PET at its maximum (based on GINI importance; Fig 4.10) and when the change in ET was also at its maximum (important in the SHAP importance; Fig. 4.10). However, the RFs capture the extent of the FD fairly well for the N20 method, and for the central Great Plains for the C23 and O21 methods (but not for Nebraska and the Dakotas). SVMs also focused on June and July (Fig. 4.30), in this case it is likely because the SM was minimum and change in SM was maximum at that time (Fig. 4.31). The SVMs also struggled with the spatial extent of the FD, identifying effectively the extent for all identification methods.

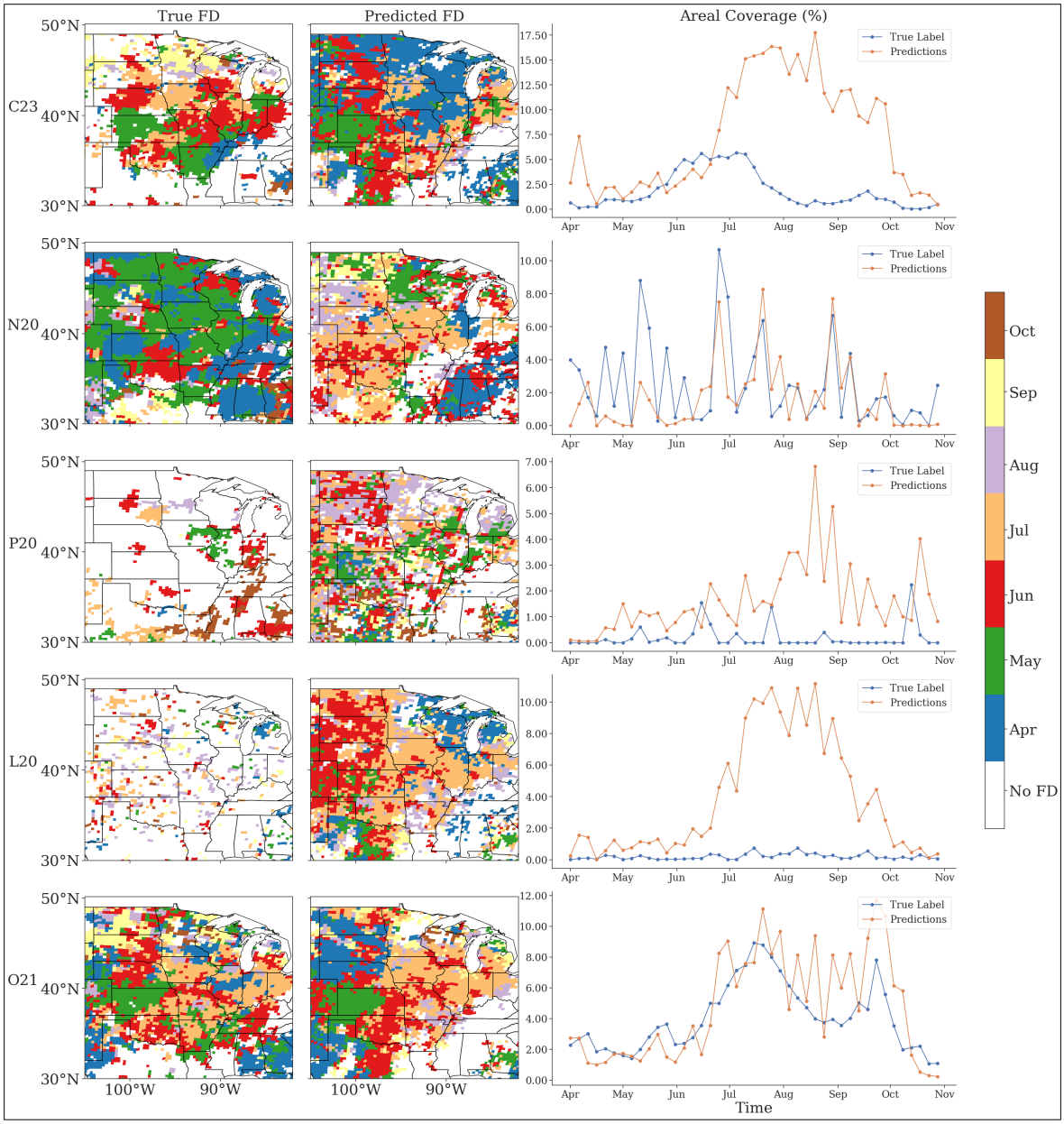


Figure 4.28: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.

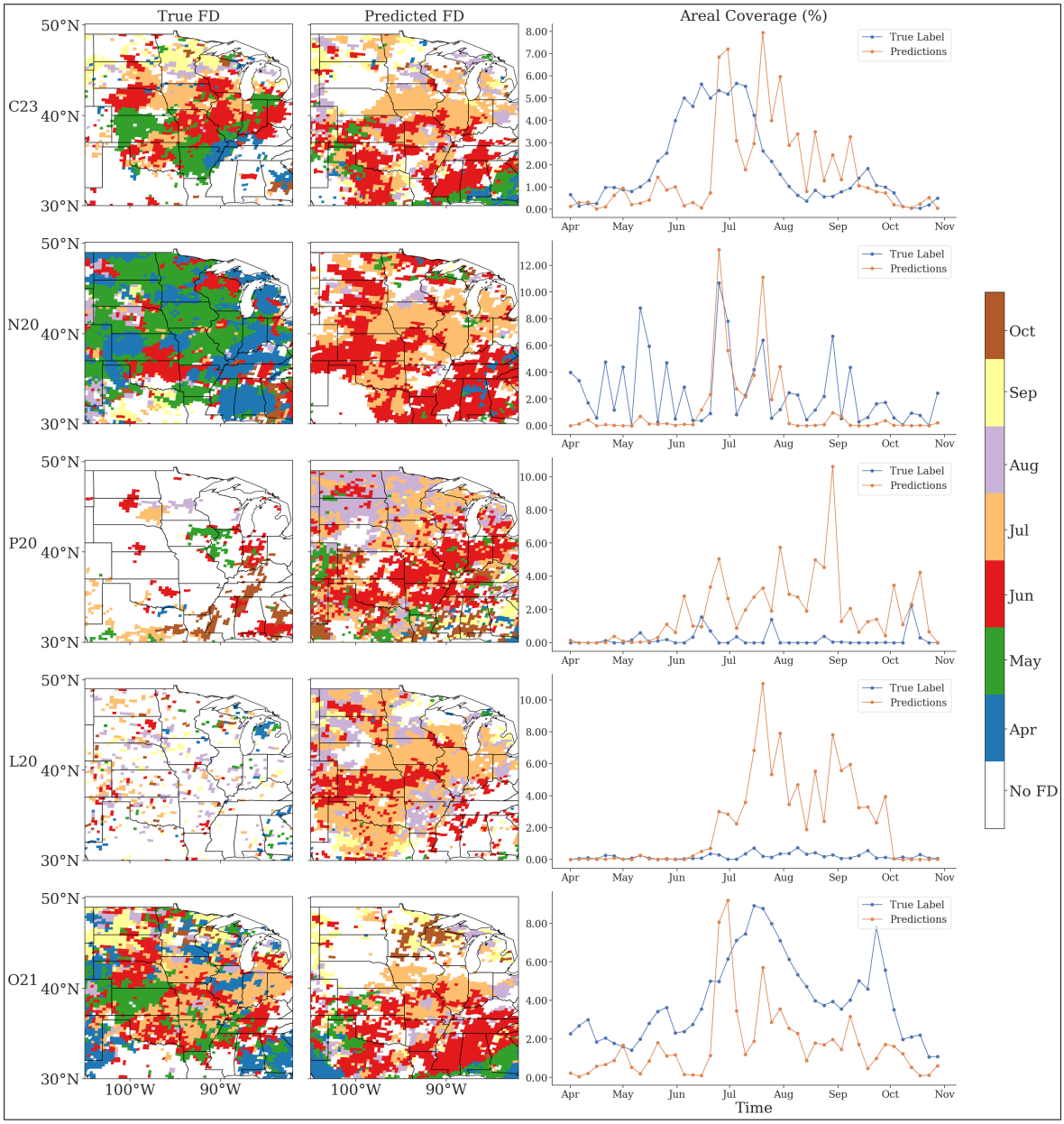


Figure 4.29: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RFs.

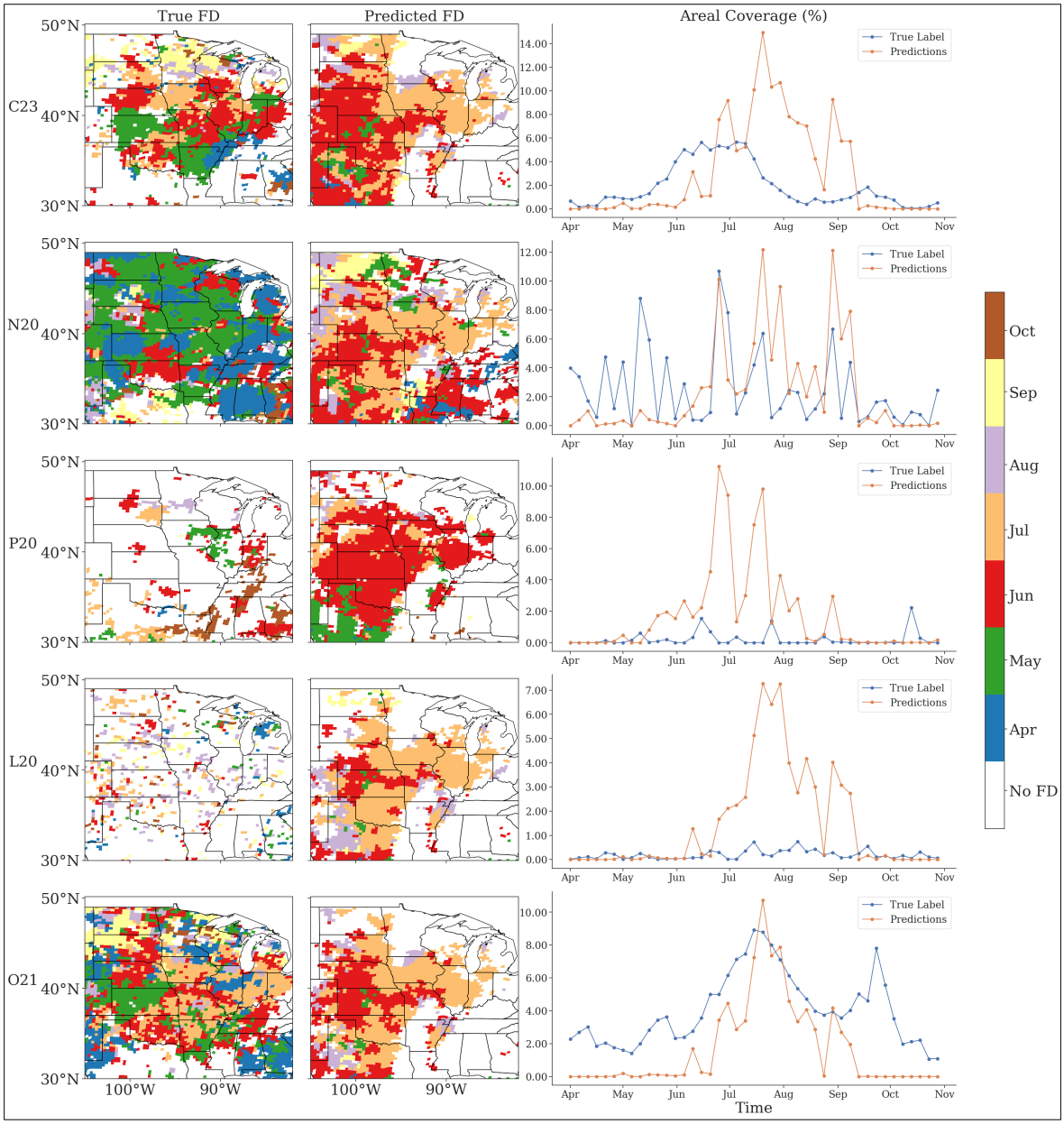


Figure 4.30: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the SVMs.

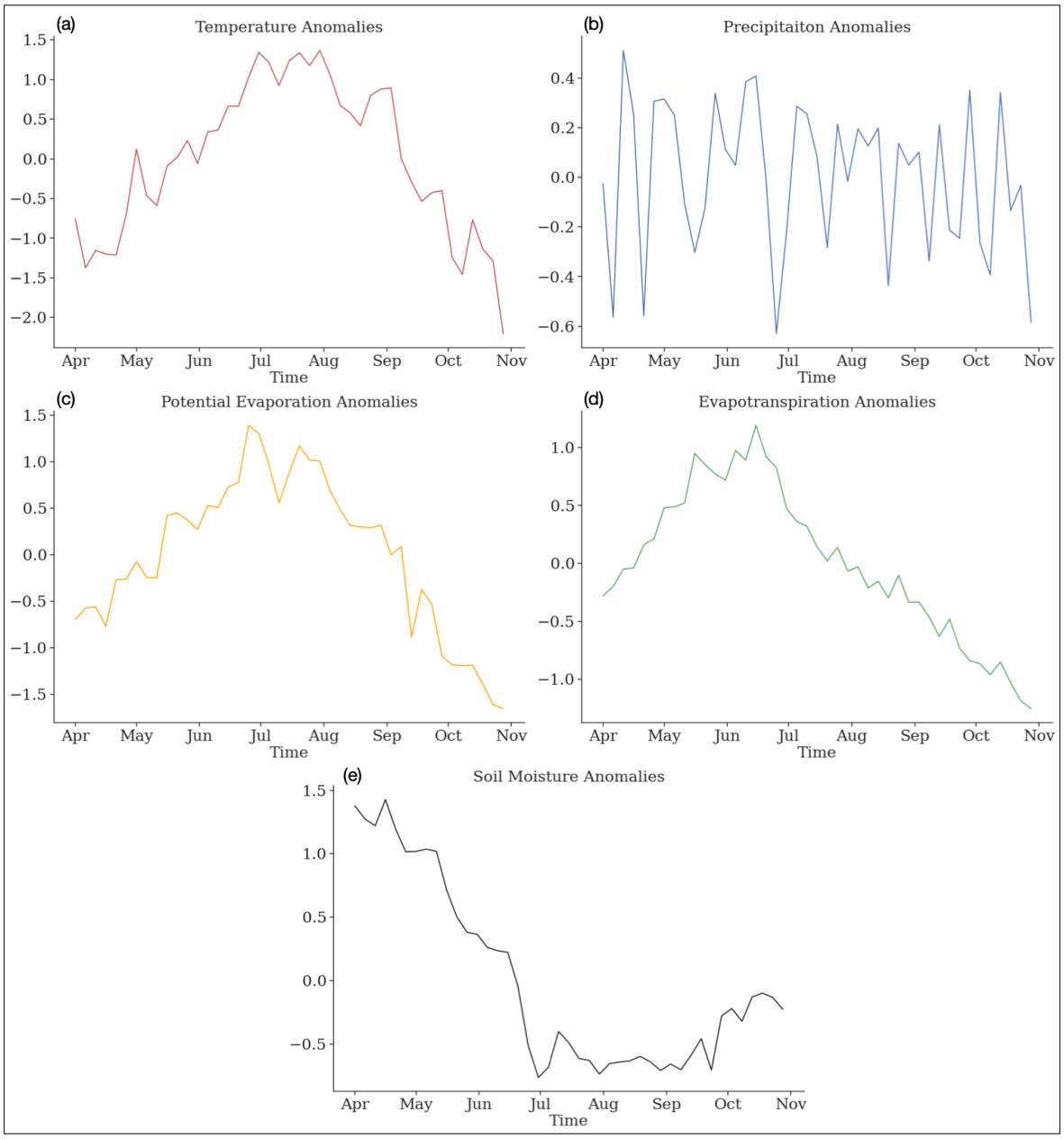


Figure 4.31: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2012 case study.

Chapter 5

Deep Learning Performance

In conjunction to investigation of FD with standard ML techniques, DL methods were also employed to determine how well they compared to the standard method, in particular to Ada boosted trees and RFs as they have managed to perform equally well as DL, and even surpass them, before (e.g., Prodhan et al. 2021; Zhang et al. 2022). In this chapter, DL and their results were examined in the same manner as in chapter 4, to preserve the thoroughness of examining the statistics, climatological predictions, and local scale case-by-case predictions, and for ease of comparison with the standard ML results.

5.1 Statistical Performance

The true skill statistic for all NN structures are given in Table 5.1. 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). In comparison, the standard ANN structure struggled to make FD predictions and the U-nets overemphasized spatial patterns more than RNNs. ANN results could be partly due to the tendency of the ANNs to get trapped in a trivial solution more frequently than the RNNs, U-nets, or standard ML algorithms, because the ANNs did not consider spatial structure, and also because it over emphasized SM above all other variables,

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

Table 5.1: True skill statistic over all grid points and pentads for each DL algorithm and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for is identification method is bolded.

similar to the SVMs (Fig. 5.1), and generally did not seem to learn the surface feature interactions involved in FD events.

In examining the difference between the true and predicted FD labels for the RNNs (Fig 5.2), a widespread improvement can be seen in the climatology predictions compared to the Ada boosted trees for four of the identification methods. In each case, the amount of over prediction of FD in hotspot regions was notably reduced, with fewer cases of under prediction except for the Great Lakes region and the Northeast in the C23 method, and southeastern California in the N20 method. The consistent under predictions of FD in the O21 method suggest that the RNNs focused heavily on the events in the southern U.S. 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 were statistically significant (Fig. 5.3), 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. Composite differences for the U-nets

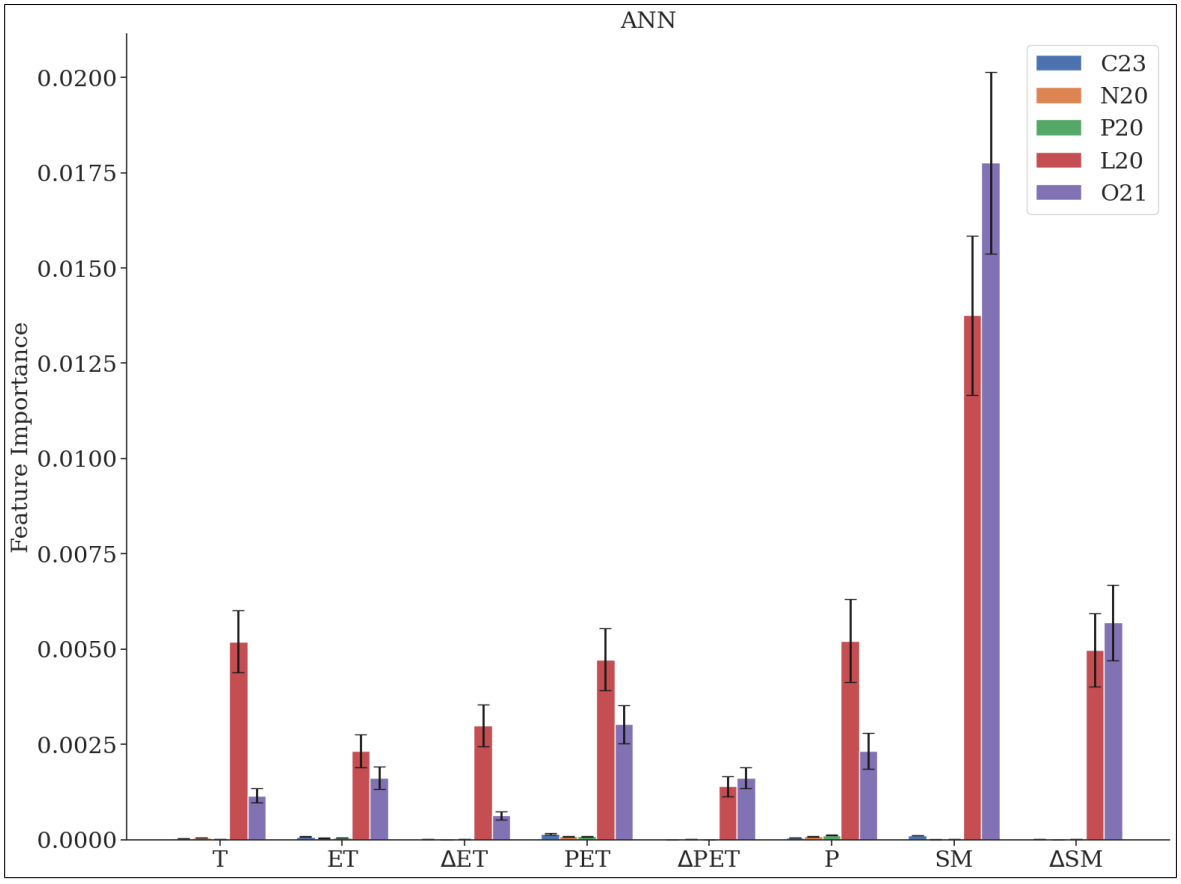


Figure 5.1: Average feature importance for the ANNs according to the SHAP method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations).

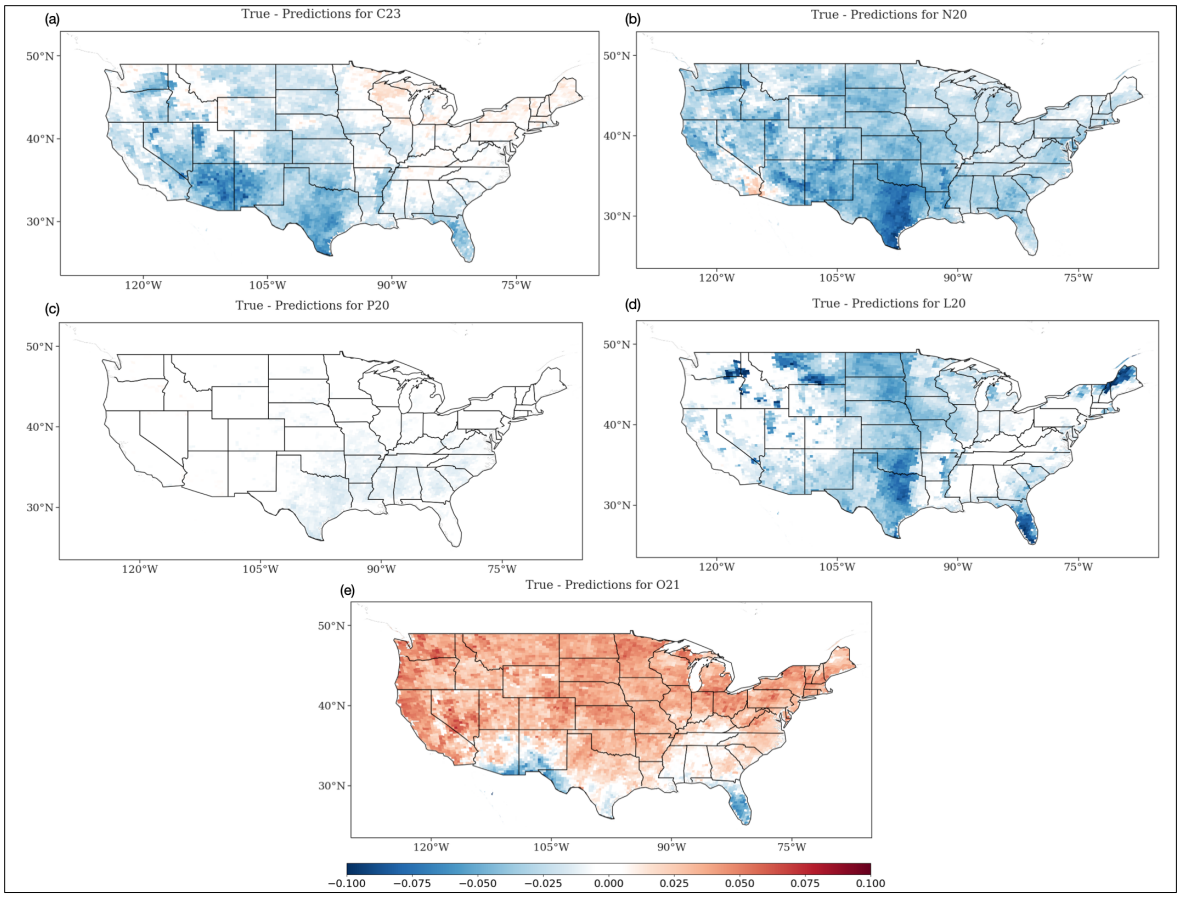


Figure 5.2: Composite mean difference of true minus predicted FD labels for each FD identification method using RNN predictions.

showed relatively good results for the L20 method except in the Great Lakes region and the Northeast (Fig. 5.4d). However, U-net results for other methods were more mixed, where the P20 method saw consistent over predictions, and the C23, N20, and O21 methods had a mixture of under predictions and over predictions (for the N20 method, the under/over prediction was spatially split between the eastern and western United States; Fig. 5.4). The ANNs, however, either over predicted (P20, L20, and O21) or under predicted (C23, and N20) FD in all locations, depending on the identification method (Fig. 5.6).



Figure 5.3: 95% statistical significance of composite mean differences between the true FD labels of the RNN predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

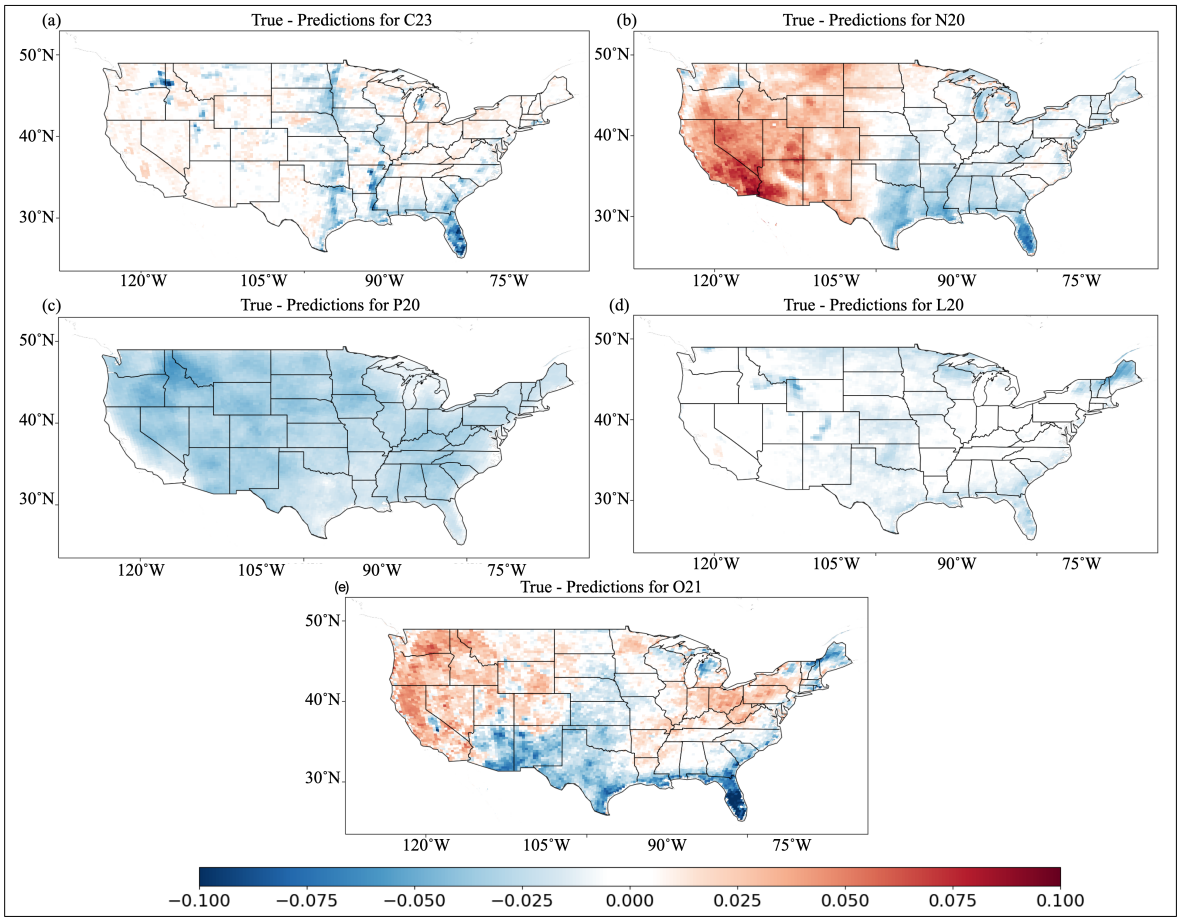


Figure 5.4: Composite mean difference of true minus predicted FD labels for each FD identification method using U-net predictions.

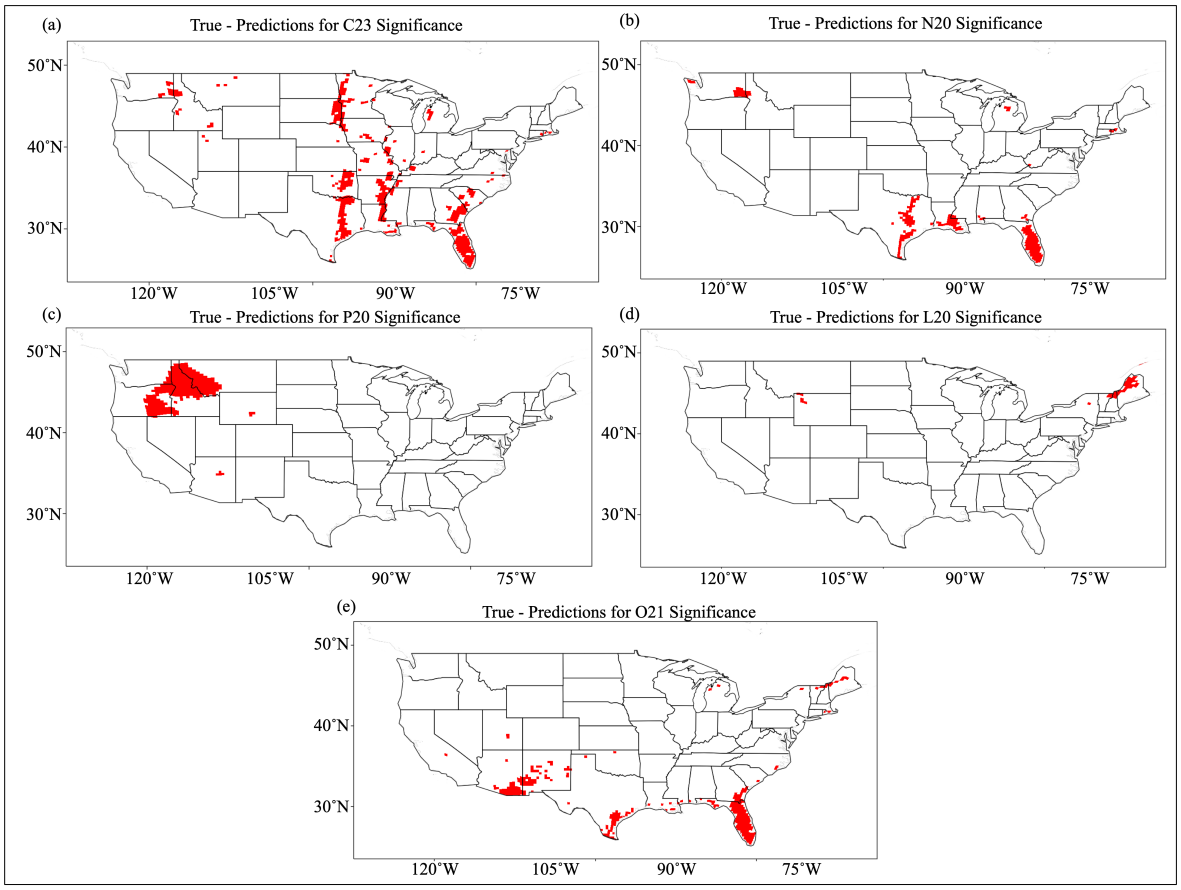


Figure 5.5: 95% statistical significance of composite mean differences between the true FD labels of the U-net predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

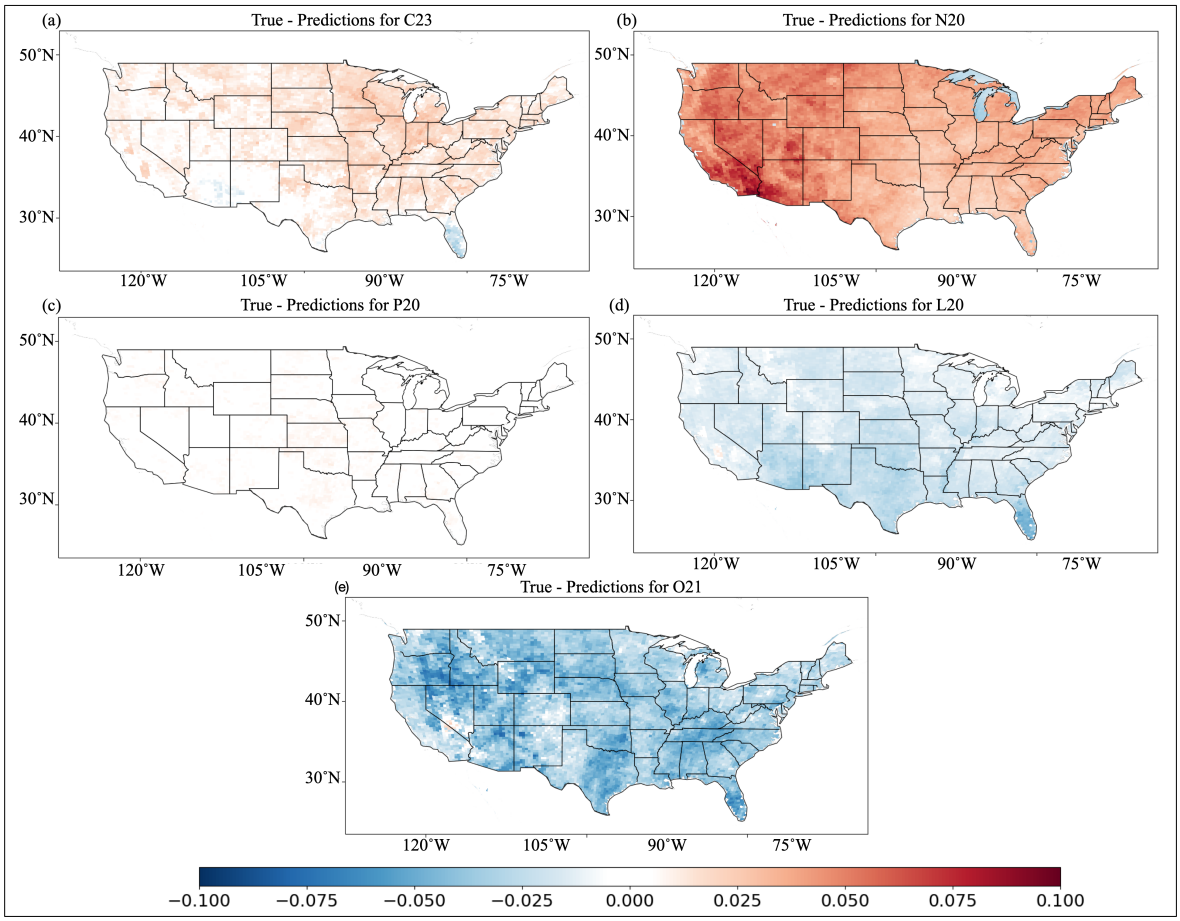


Figure 5.6: Composite mean difference of true minus predicted FD labels for each FD identification method using ANN predictions.

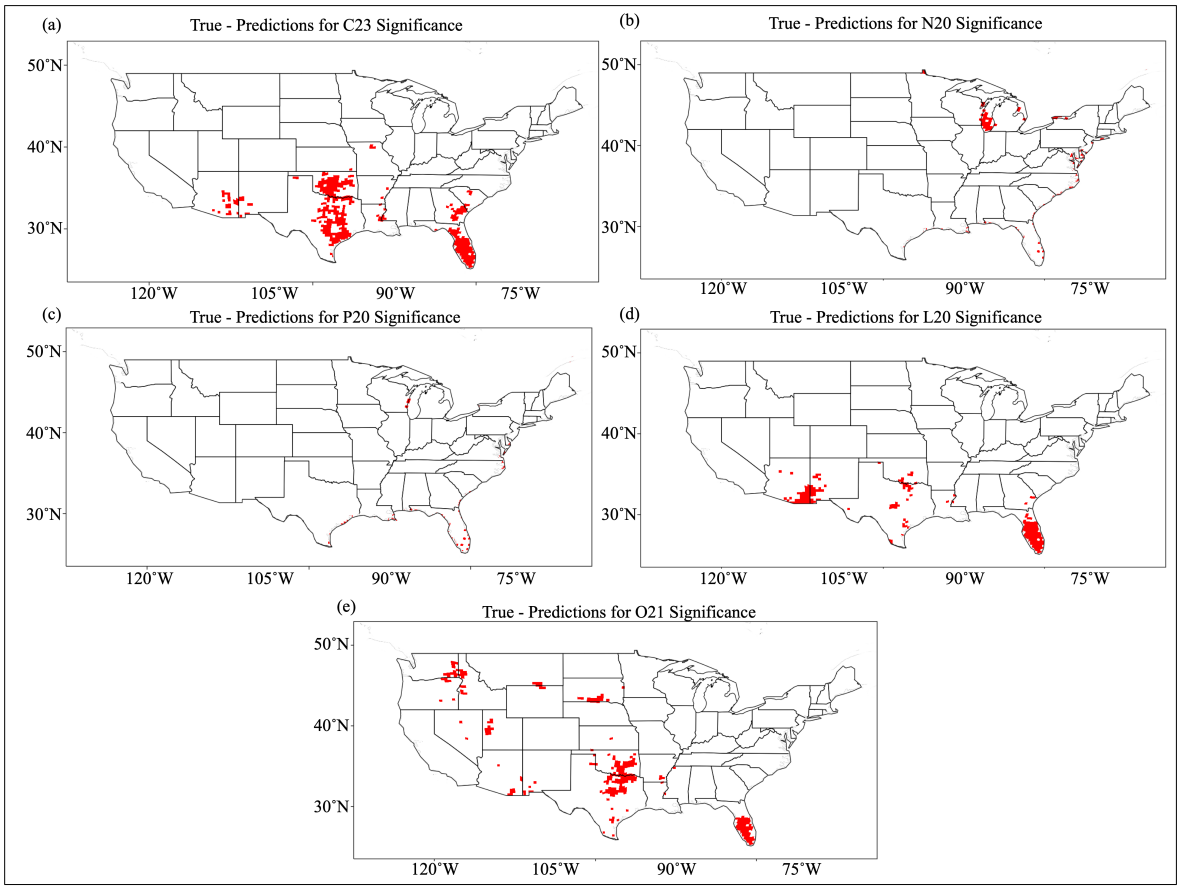


Figure 5.7: 95% statistical significance of composite mean differences between the true FD labels of the ANN predictions. Statistical significance determined by the Monte-Carlo bootstrapping method (N = 5000).

5.2 Predicted Climatology

Figure 5.8 shows the predicted FD frequency climatology by the RNNs. The RNNs showed notable improvement in the predicted climatology compared to the Ada boosted trees, with the over emphasis on hotspot regions reduced in the RNNs compared to the Ada boosted trees. The agricultural regions that the Ada boosted trees identified as hotspots were still present in some RNN predictions (e.g., for the C23 method), and absent in others (e.g., in the P20 method). In addition, the over emphasis in Florida was still present for some FD identification methods (i.e., C23, L20, and O21). Further, while the RNNs created hotspots in some agricultural regions (e.g., eastern Washington and southern Idaho), the signal there was much more muted compared to what the standard ML algorithms learned. In exchange, however, the RNNs learned to identify an additional hotspot in Arizona and western New Mexico. This was particularly interesting as while this it not a hotspot for FD in the true labels (Fig. 4.1), it was a hotspot for rapid intensification of dry conditions without going into drought (Edris et al. 2023). In particular, 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 the RNNs, for the O21 method, may have been over reliant on T and/or PET for FD prediction for that identification method, given its similarity to the T and PET climatologies in the United States.

As the RNNs seemed to over predict FD in similar locations as the Ada boosted trees for all methods except O21, they may have relied on similar variables as the Ada boosted trees to make predictions (SM, P, and PET), but the performance in over predictions and emphasis on other locations also suggests the RNNs more effectively incorporated information from other variables and were not as strongly reliant on a select few as the Ada boosted trees were. The U-nets also reproduced frequency

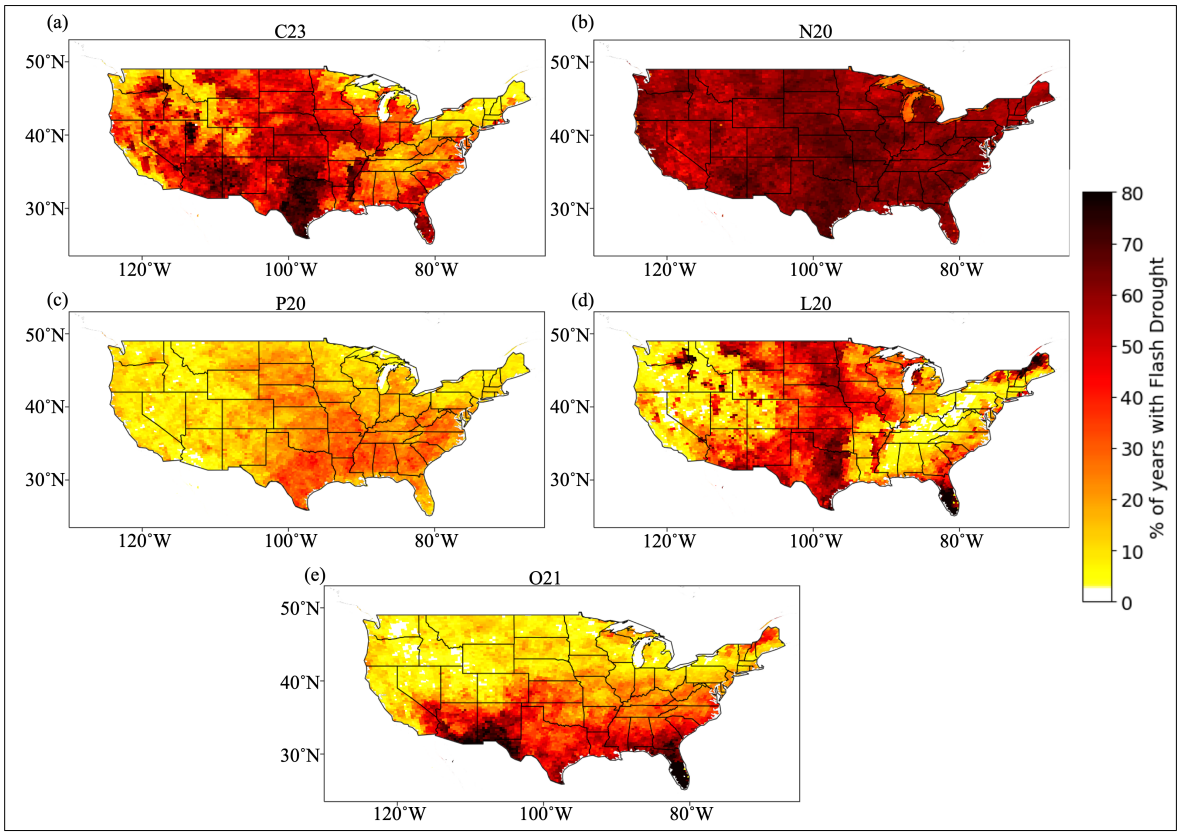


Figure 5.8: Predicted frequency climatology of FDs by the RNNs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

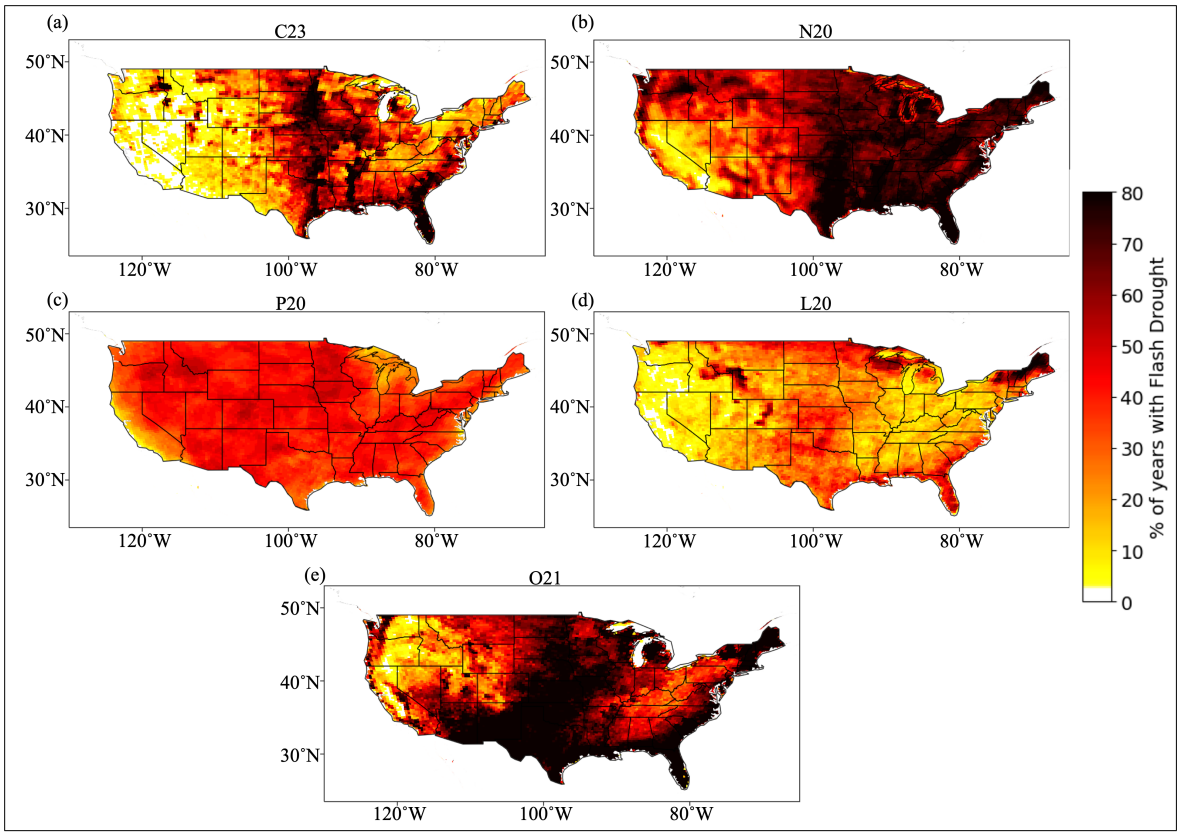


Figure 5.9: Predicted frequency climatology of FDs by the U-nets for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

climatologies for some methods well (C23, N20, and L20), while the ANNs produced a more realistic frequency climatology for the O21 method (Fig. 5.9). However, the U-nets nets struggled to learn patterns for the P20 method, and over predicted FD for the O21 method similar to the standard ML algorithms, and the ANNs struggled to learn FD for methods other than O21 as a whole (Fig. 5.10).

Also in contrast to the Ada boosted trees, the RNNs may also have learned the seasonality of FD more effectively. The 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. 5.11). The RNN seasonalities also matched the true labels more closely in some cases (E.g., the C23, N20, and L20, though it had trouble

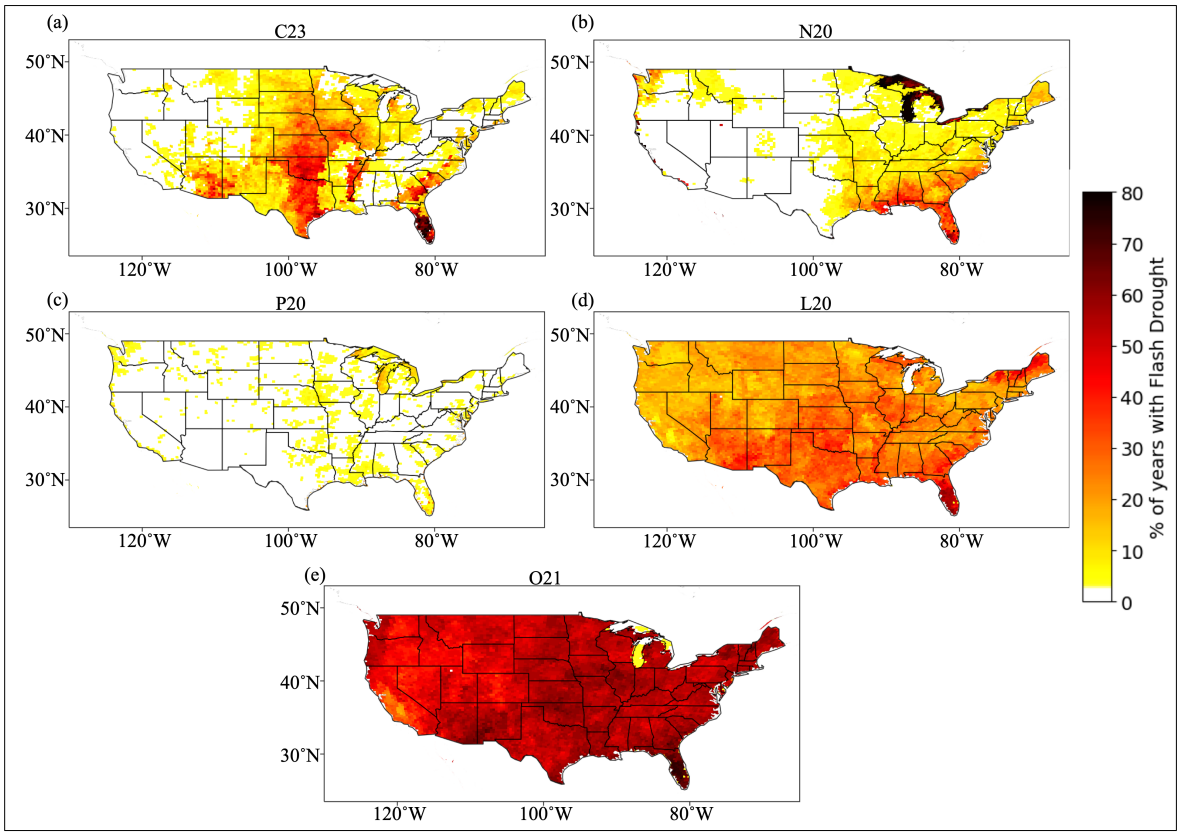


Figure 5.10: Predicted frequency climatology of FDs by the ANNs for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

finding early season FDs in C23), but that attempt at learning the seasonality also resulted in poorer results in the P20 and O21 methods (Fig. 5.11c, e). In addition, the resulting time series of FD coverage had more mixed results compared to the Ada boosted trees.

The RNNs in general did not learn as well which years had larger FD events as well, resulting in some years with much more widespread FD predicted when the true labels showed otherwise (Fig. 5.12). Note years when no FD was predicted for the N20 method represent the RNN experiments (K-Fold cross validation requires training the RNNs K, 42 in this case, times) that predicted trivial solutions (e.g., in 1983 – 1985; Fig. 5.12b). In comparison, the U-nets also captured the seasonality (Fig. 5.15) and time series (Fig. 5.14) of the C23 method well (even finding some of the early season FDs), and the seasonality of the O21 method (though the U-nets struggled getting some of the early season FDs with that method), and it had difficulty learning the time series for the O21 method (Fig. 5.15e). The ANNs were able to learn some of the seasonal patterns of FD for the O21 method, however they over emphasized the end of season FDs (Fig. 5.15), and struggled notably with the time series, creating an erroneous large area covered by FD in 2000 – 2010 (Fig. 5.16).

5.3 Case Studies

A case study for the FD of 2012 is given here (Fig. 5.17) as a direct comparison to the predictions from the Ada boosted trees in Chapter 4. In this case, the boosted trees performed relatively well and captured much of the spatial 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

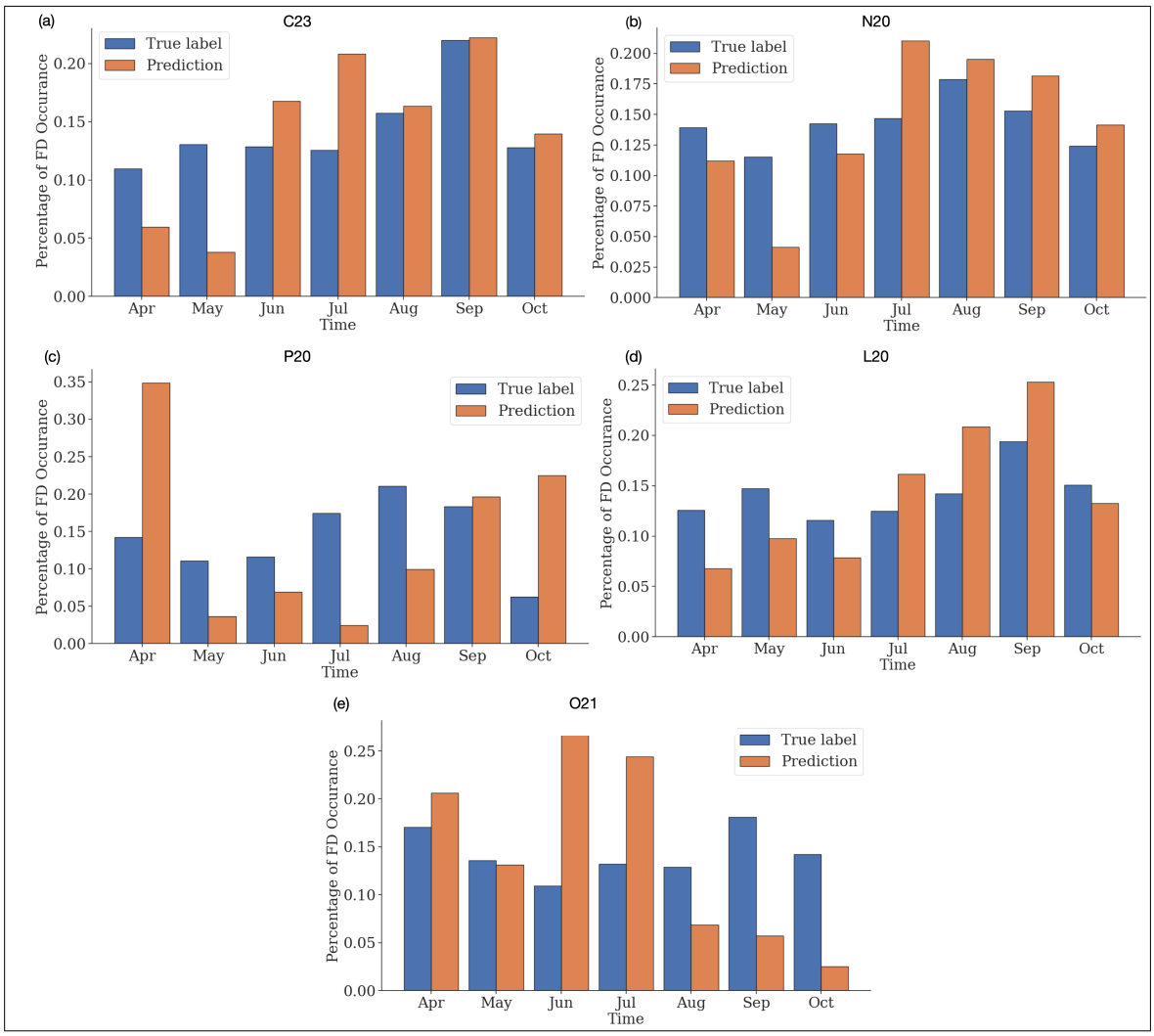


Figure 5.11: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the RNNs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

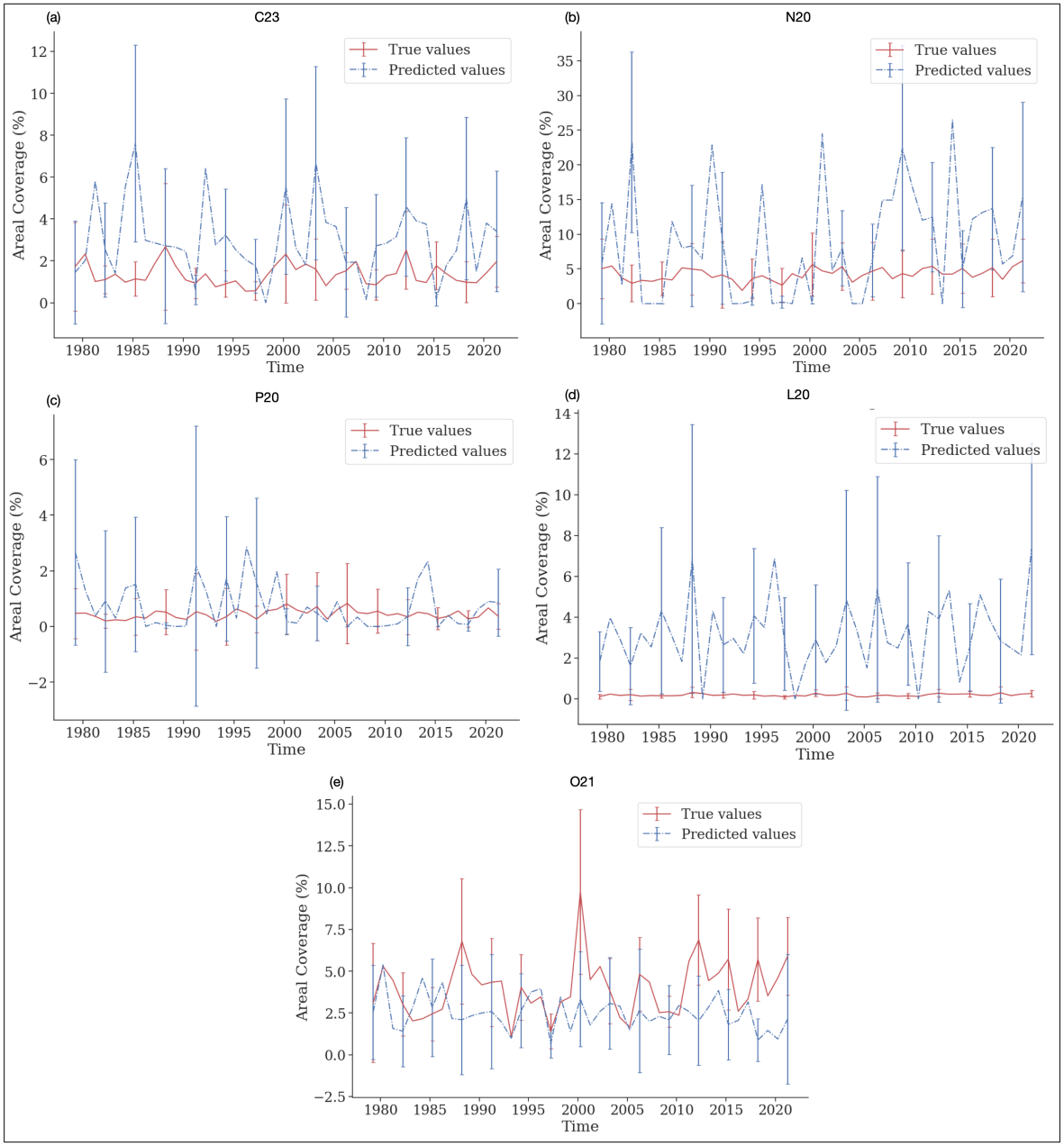


Figure 5.12: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the RNNs.

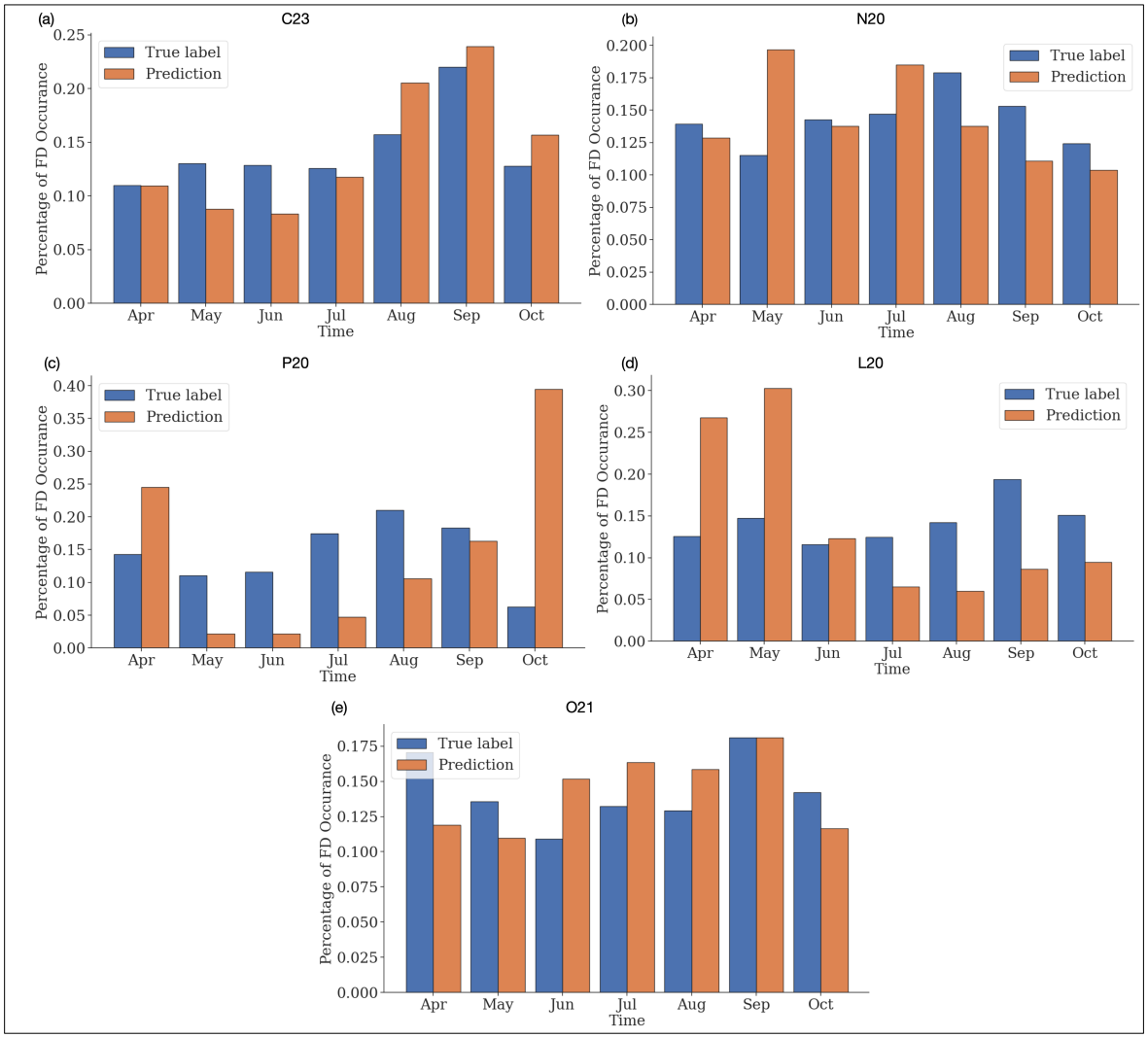


Figure 5.13: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the U-nets (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

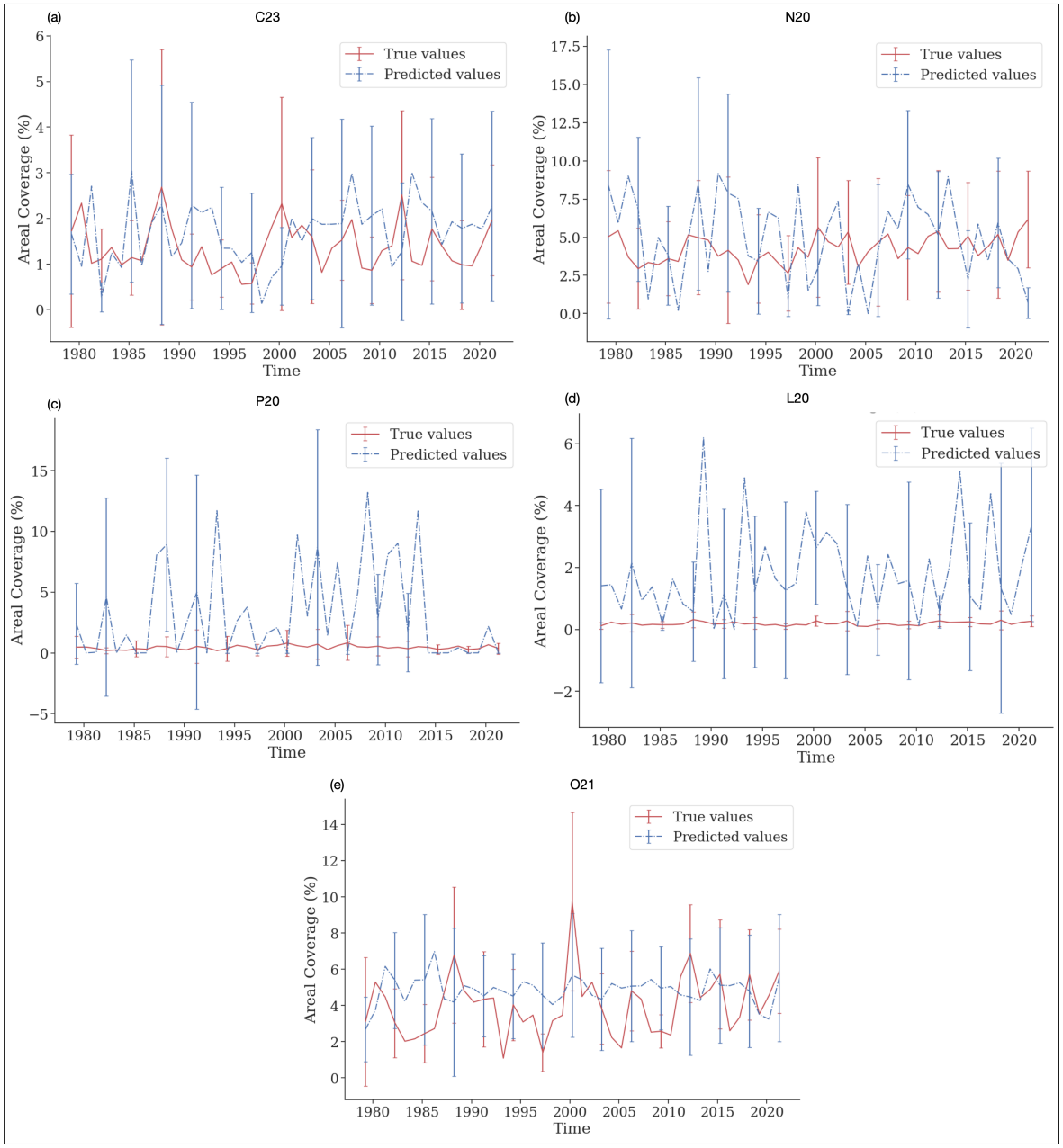


Figure 5.14: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the U-nets.

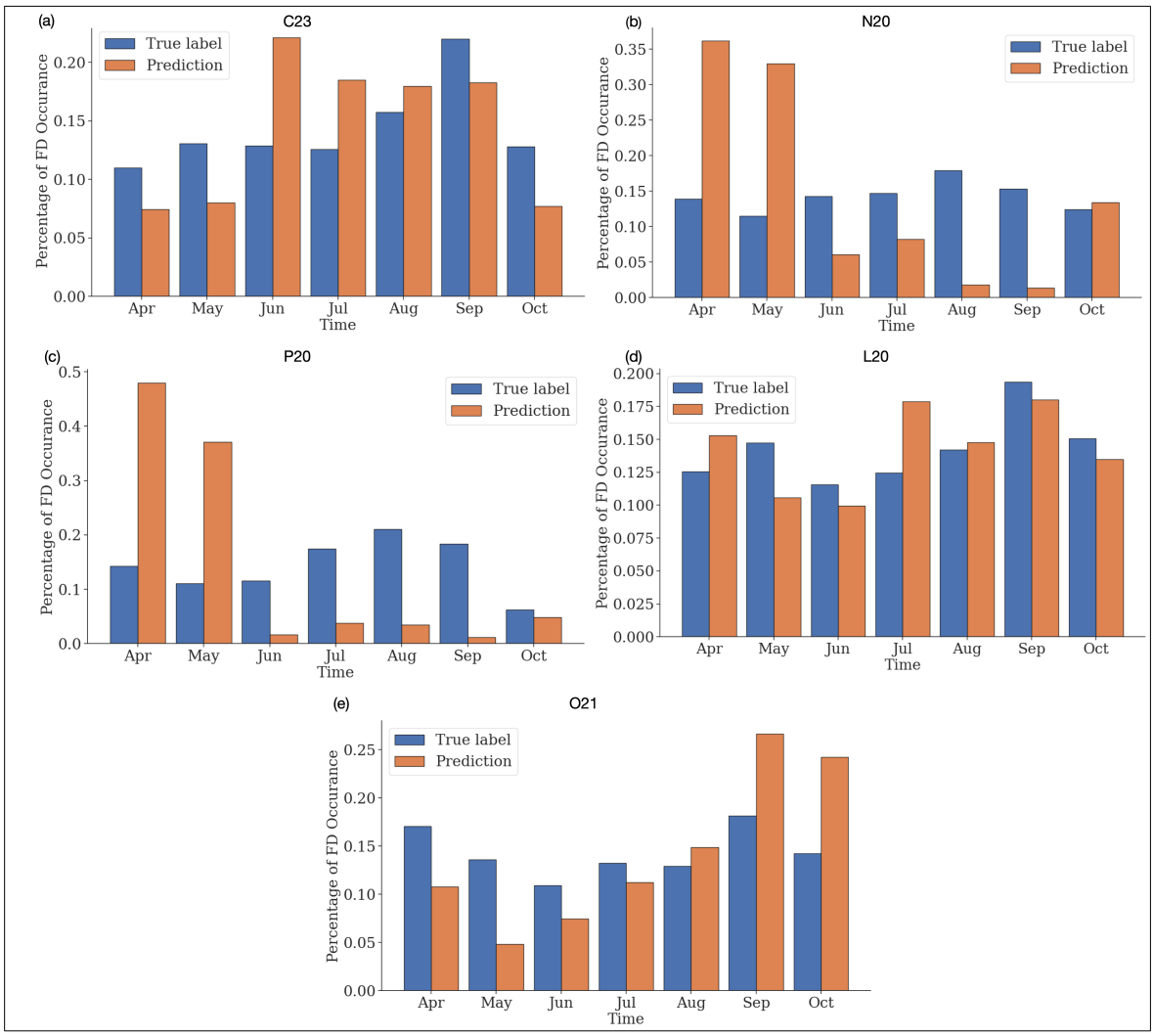


Figure 5.15: FDs seasonality (percentage of FD occurrence) for true labels (blue) and predicted labels by the ANNs (orange) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods.

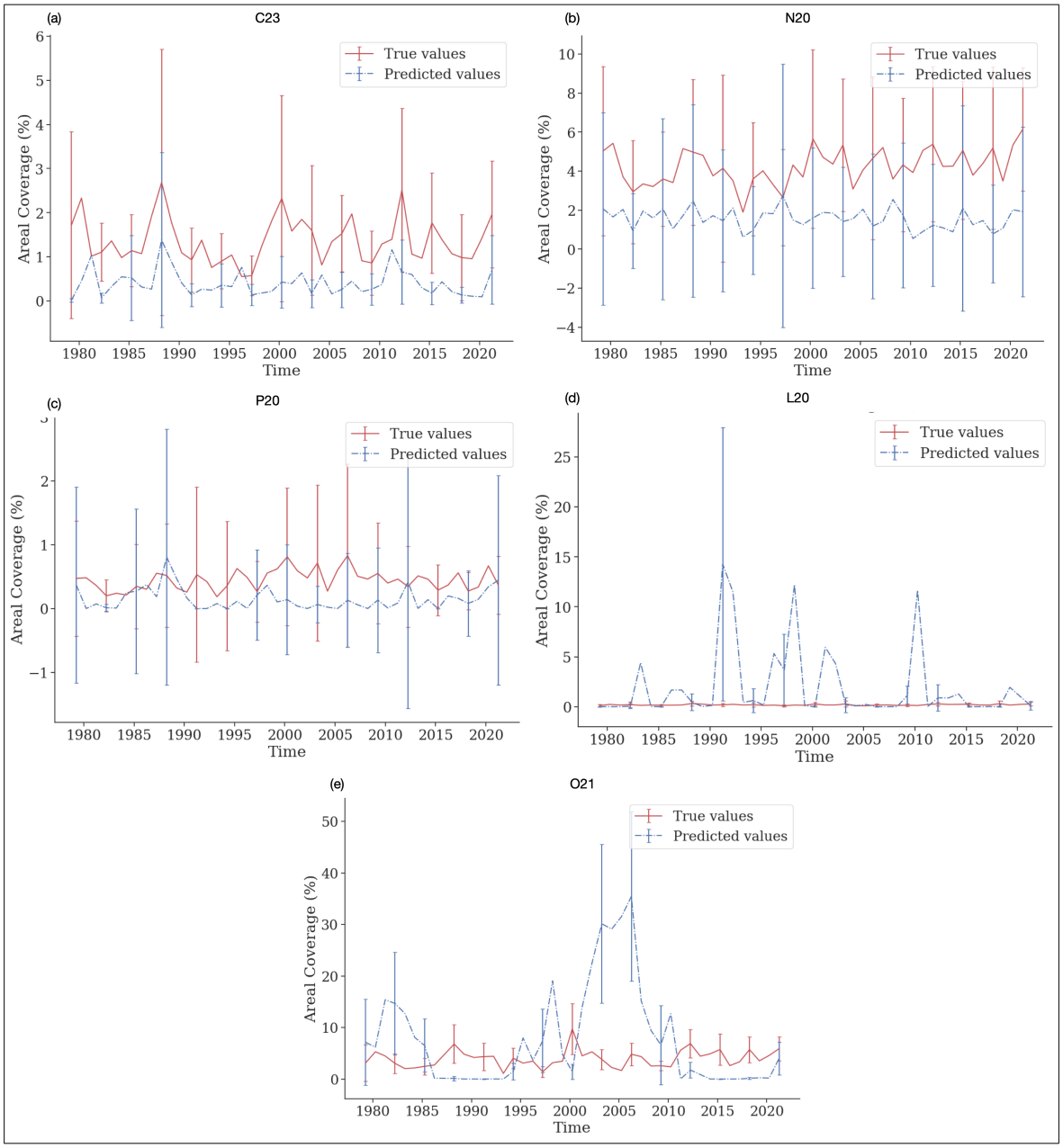


Figure 5.16: True (red) and predicted (blue) annual average in spatial coverage of FD across the domain (CONUS) for the (a) C23, (b) N20, (c) P20, (d) L20, and (e) O21 methods. Error bars denote 1 standard deviation in the annual average. Predicted labels were made by the ANNs.

struggled with the exact timing of the FD onset (Fig. 5.17). Unlike the Ada boosted trees, which would identify FD too early or too late, depending on the FD identification method, the RNNs consistently identified FD a month or two late as this case study, shifting the bulk of the identification down by four to eight pentads (Fig. 5.17, right column). In contrast, U-nets were able to learn the timing of this event more effectively for the C23 and O21 methods, but under estimated the spatial extent of the FD (Fig. 5.18). For the N20 method, the U-nets also struggled with the timing, and in the P20 method, and struggled more with the P20 and L20 methods. The ANNs struggled to identify FD for all identification methods except O21 (a repeating pattern for all case studies), where they struggled with the spatial coverage of the FD and timing (Fig. 5.19).

Another case study that was examined was in 2003, in which FD developed early in the growing season, then propagated into the central portion of the Great Plains into summer as the SM declined and vegetation mortality resulted in decreased ET rates (Fig. 5.23) throughout the rest of the growing season, creating substantial crop damage (Fig. 5.20; Otkin et al. 2014). This was an interesting case study because the RNNs learned the coverage and timing of the FD more effectively, having the bulk of the FD identified at the same time as the true labels for the C23, N20, and L20 methods. However, in the C23 and N20 methods, the RNNs failed to find regions where FD initiated early, and thus did not necessarily capture the propagation of FD (Fig. 5.20, rows 1 and 2). Combined with the good representation of FD coverage for the same three identification methods, the RNNs well outperformed the Ada boosted trees in this case study (not shown). U-nets also performed well for the C23 and O21 methods, but struggled to identify FD for the N20 and L20 methods for this case study (Fig. 5.21), and tried to create continuous patches of FD in the P20 method. For the O21 method, the ANNs over represented FD, likely responding to the deteriorating SM

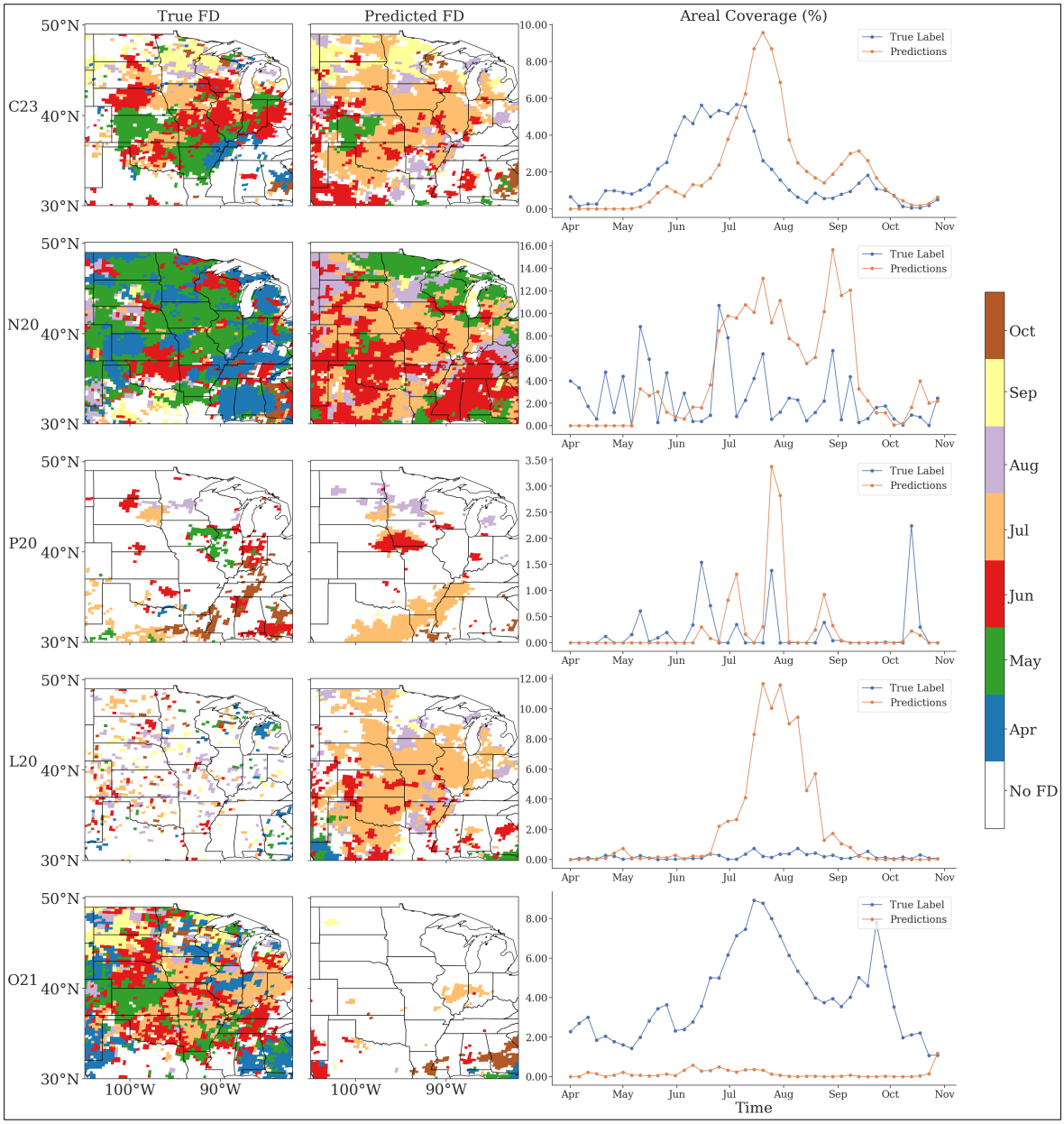


Figure 5.17: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

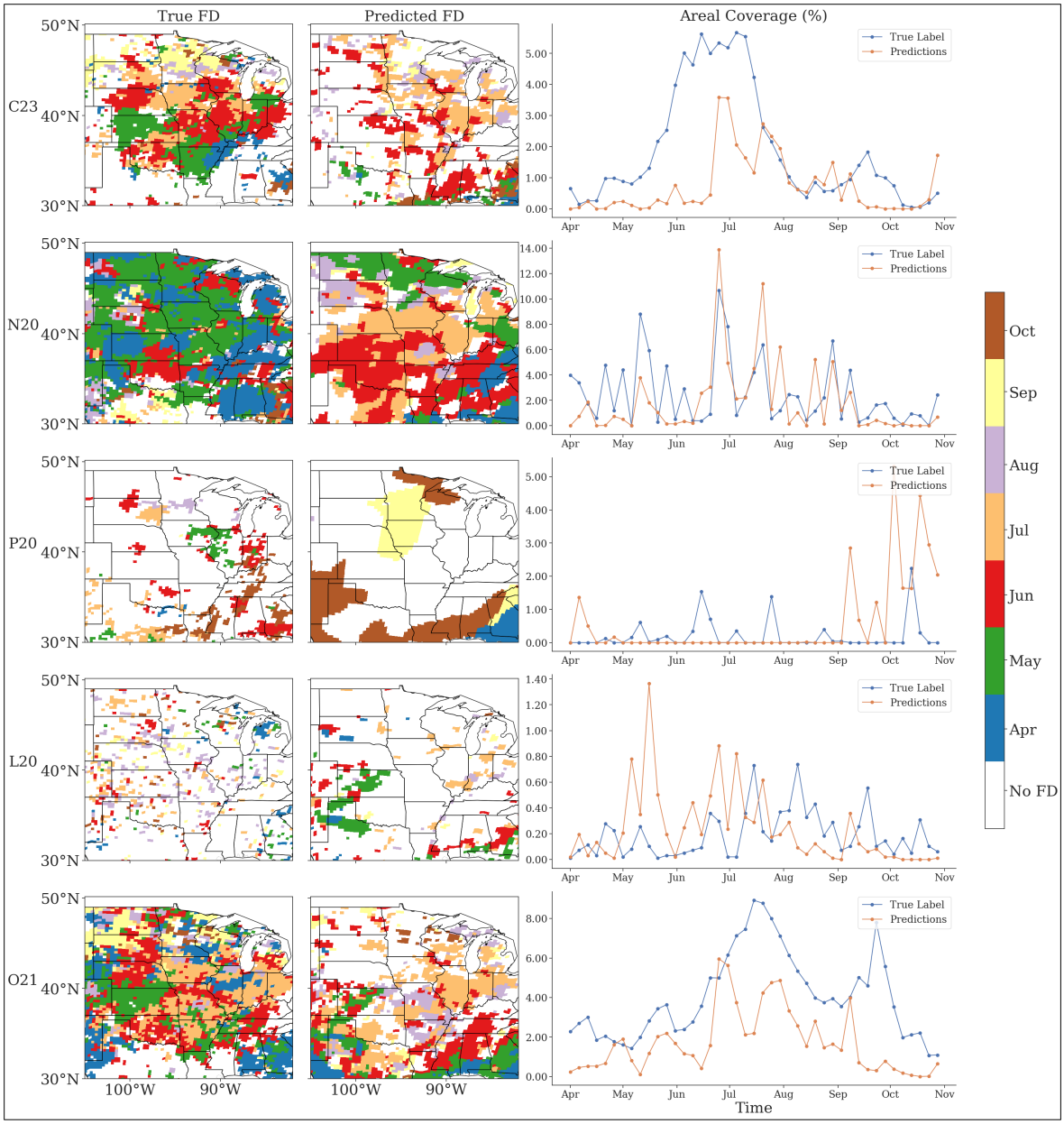


Figure 5.18: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets.

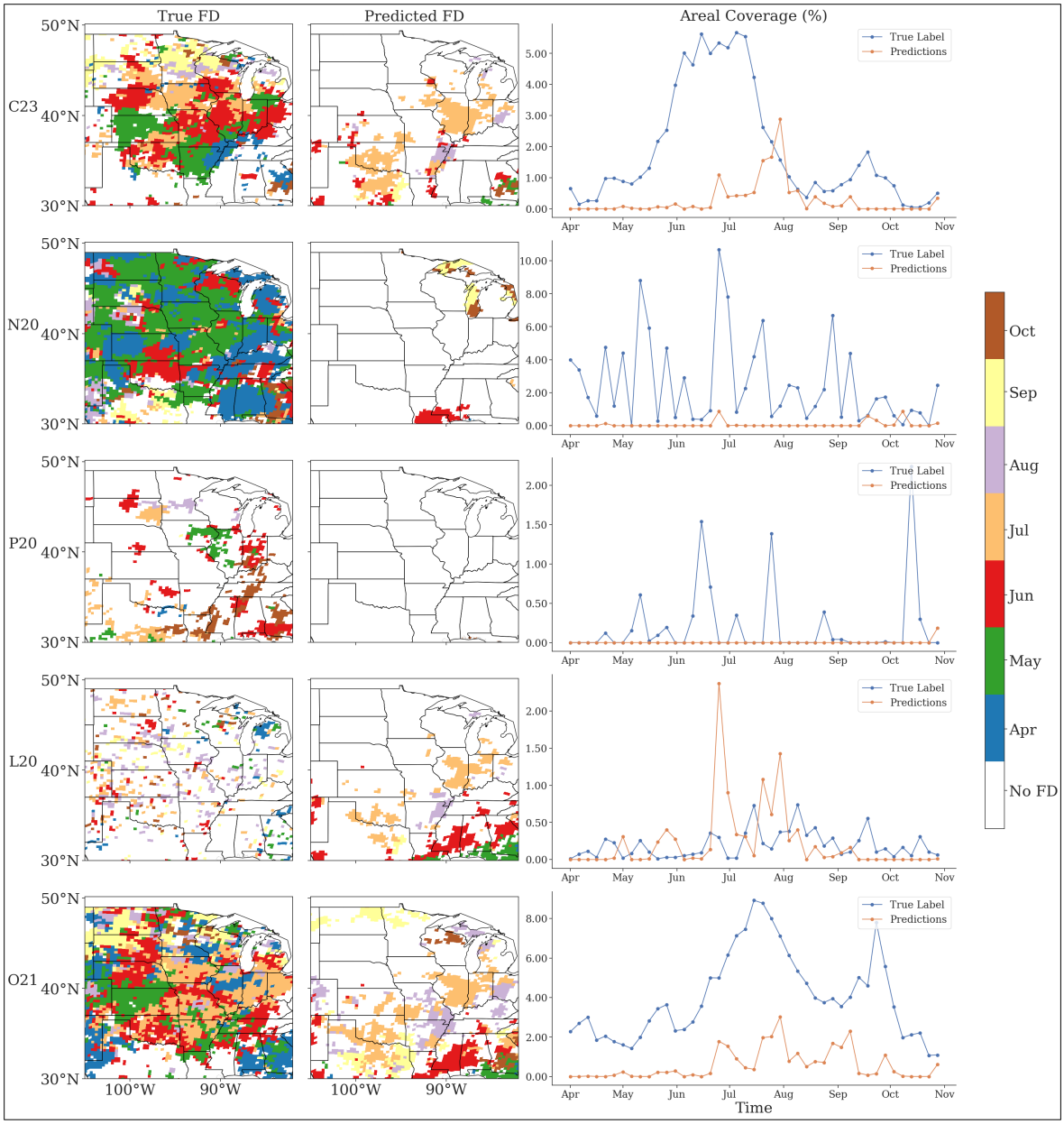


Figure 5.19: FD case study for 2012 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs.

over the region (Fig. 5.23e). The ANNs gave increasing FD coverage (and identified FD late in many cases) as it responded to SM decline.

The final case study examined was in 2017, in which large parts of eastern Montana and western North and South Dakota fell into severe drought for much of the growing season, resulting in notable agricultural impacts (up to 2.6 billion U.S. dollars in 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) as SM and ET decreased, and P20 identified the FD at the end of the growing season as evaporative demand rapidly increased (Fig. 5.24 and 5.27). 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. U-nets also capture the the FD for C23 and O21 methods fairly well, though they showed a tendency to simplify or smoothed patches of FD for all identification methods except P20 (where they failed to find FD; Fig. 5.25). In contrast, the ANNs failed to identify FD for this case study for all methods, including the O21 method (Fig. 5.26).

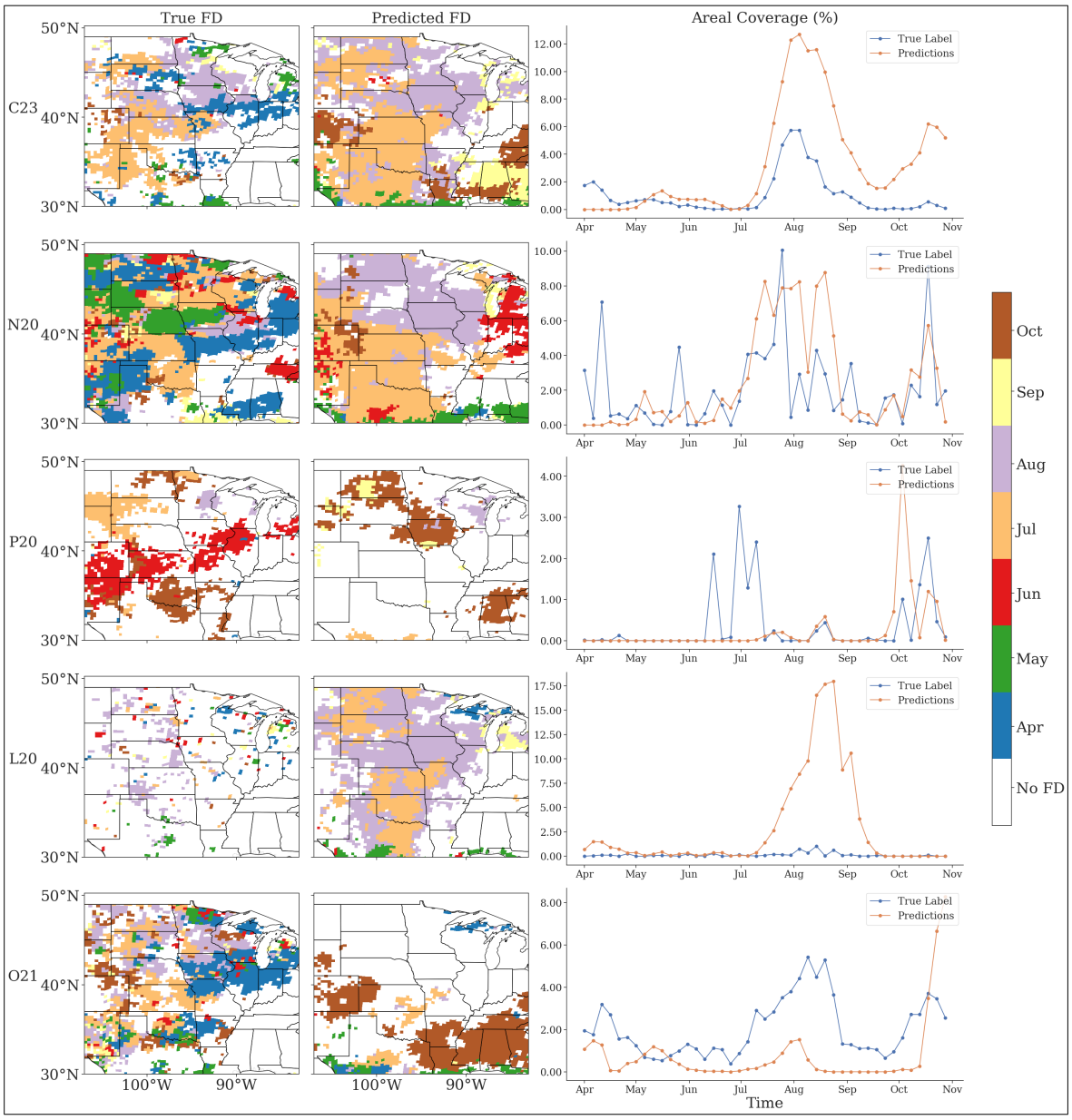


Figure 5.20: FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

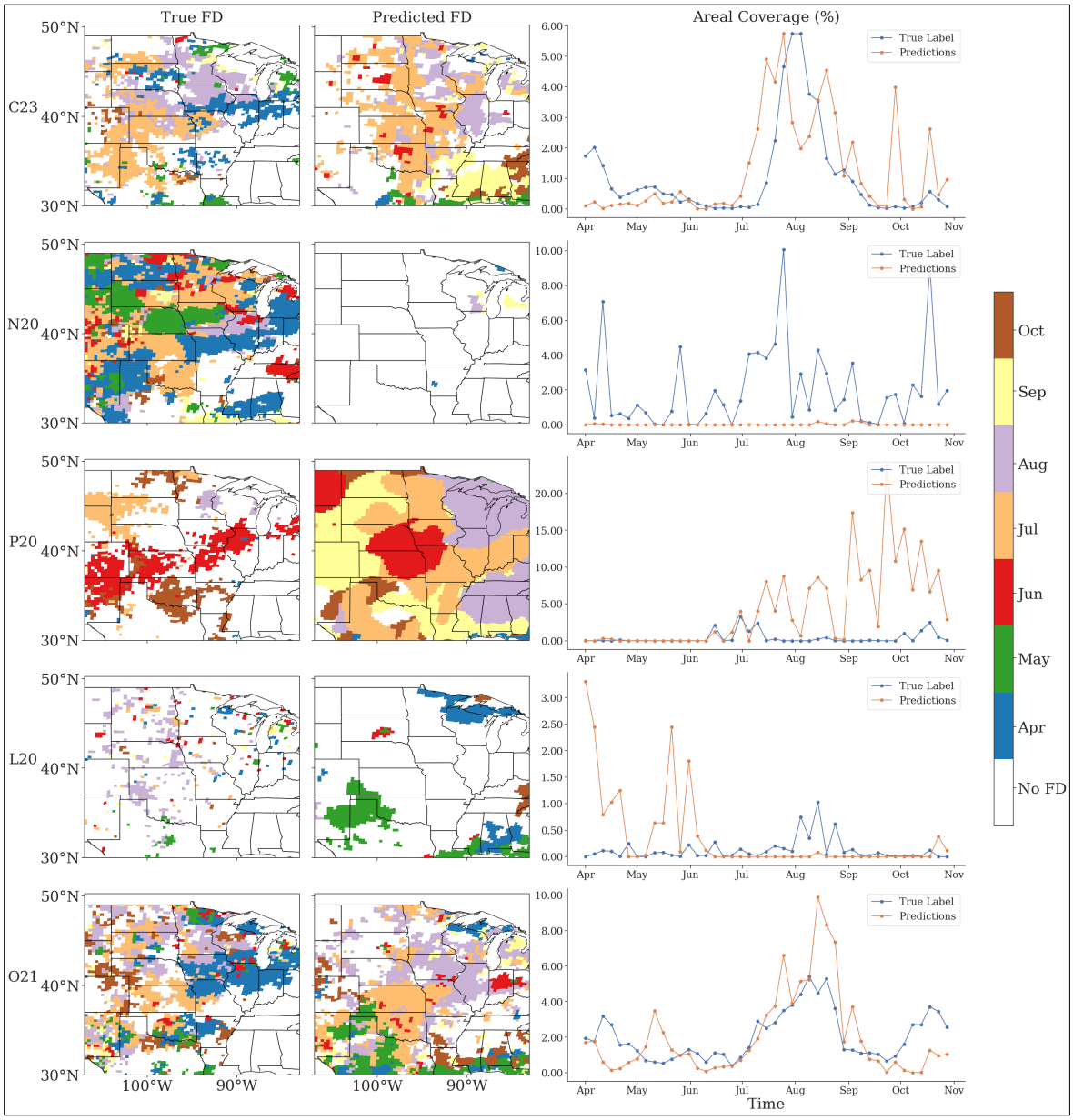


Figure 5.21: FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets.

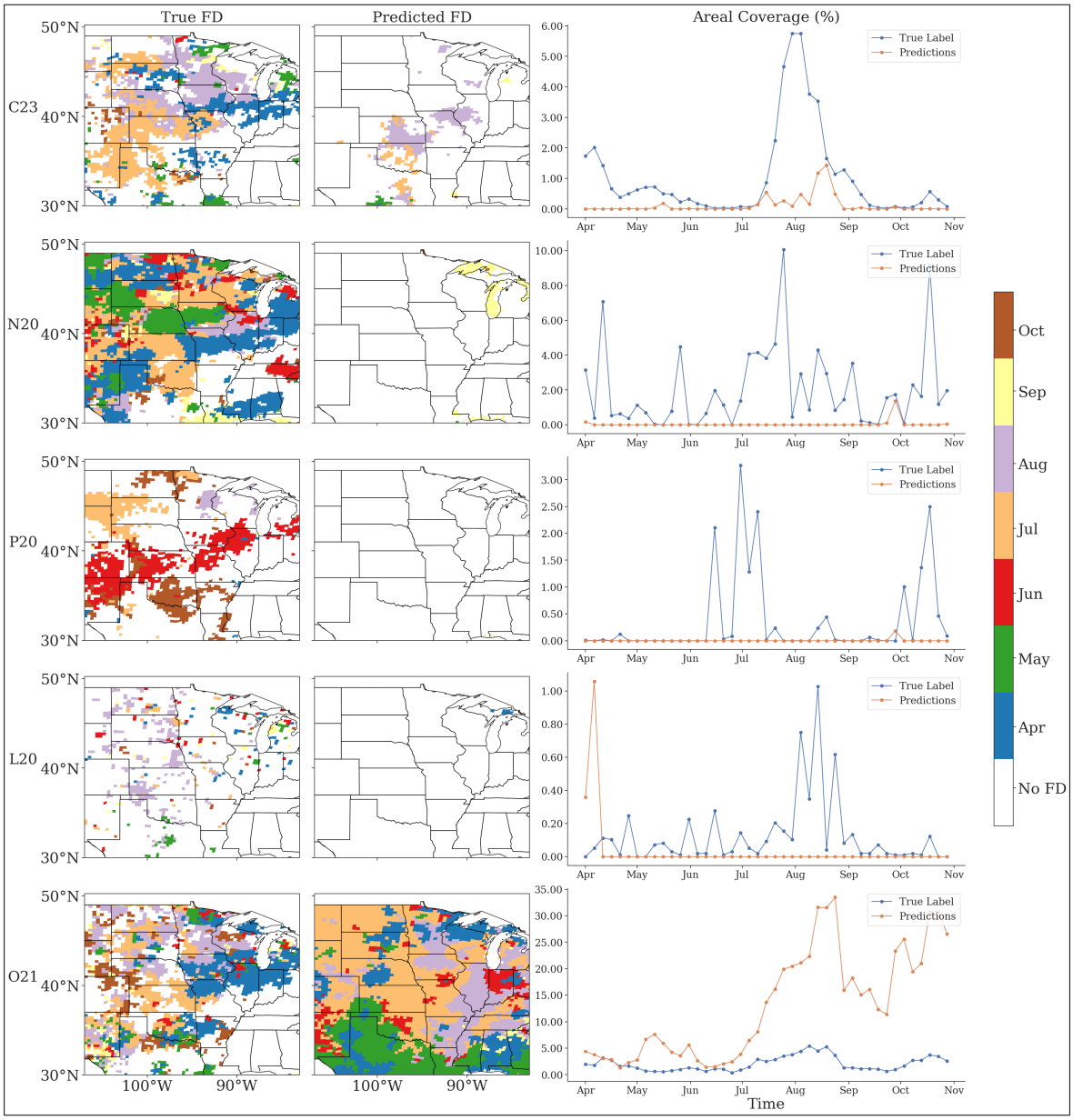


Figure 5.22: FD case study for 2003 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs.



Figure 5.23: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2003 case study.

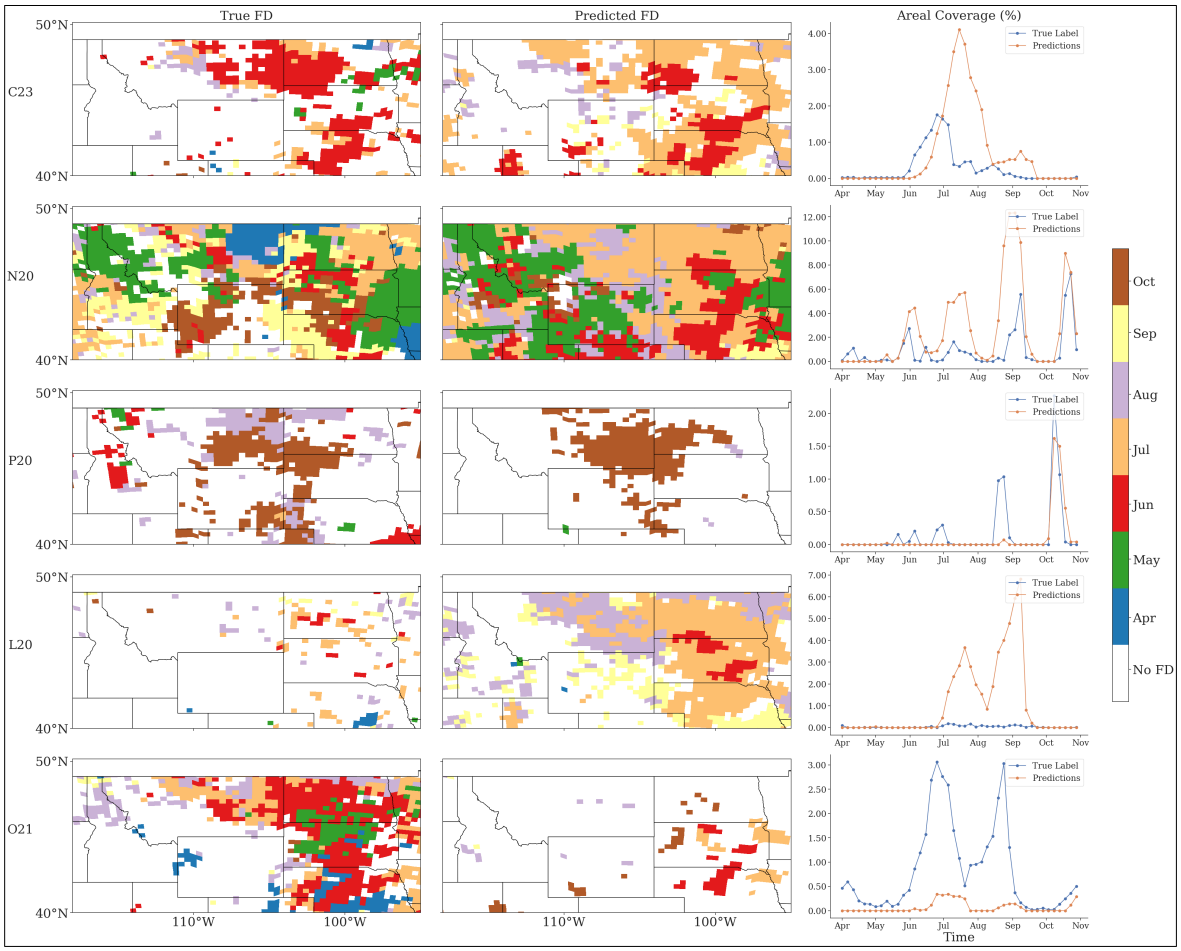


Figure 5.24: FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

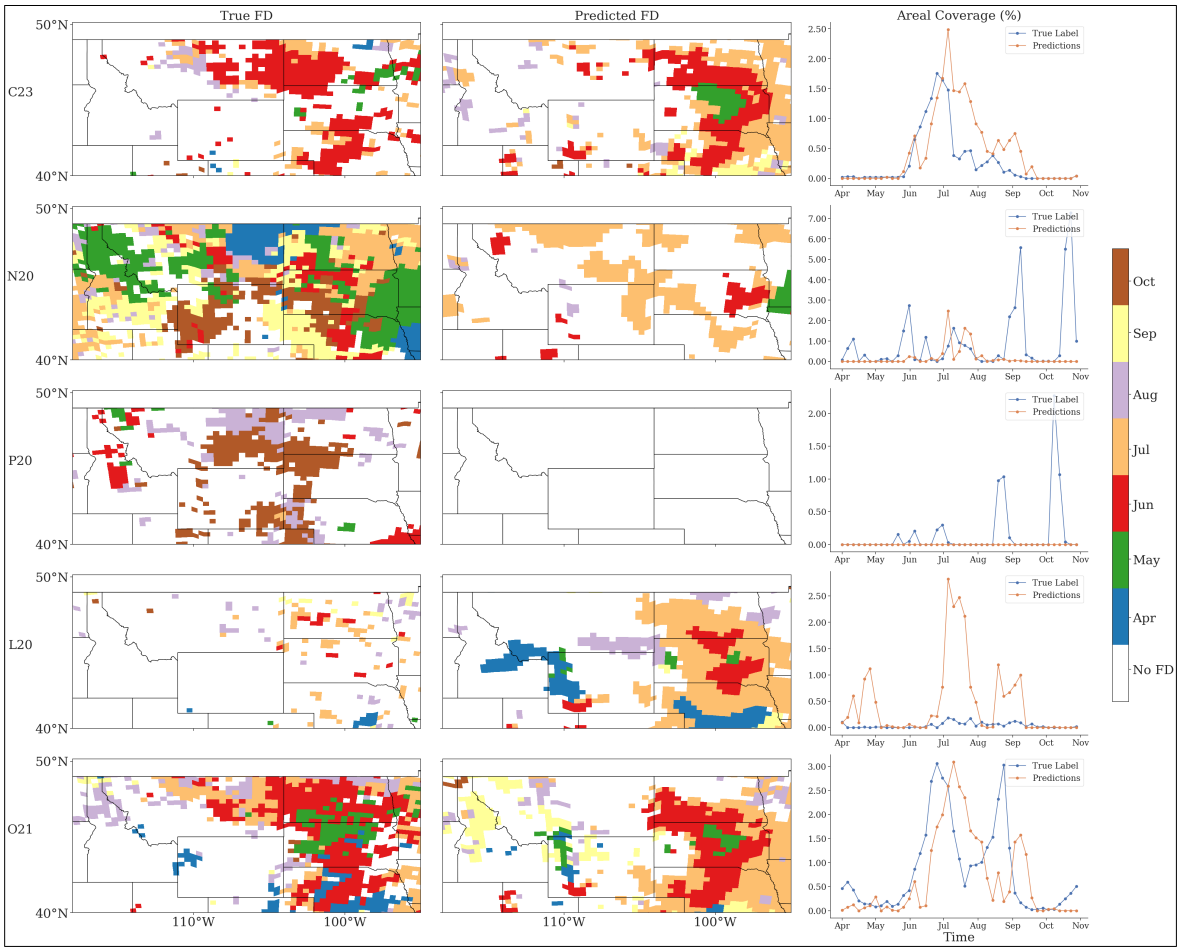


Figure 5.25: FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the U-nets.

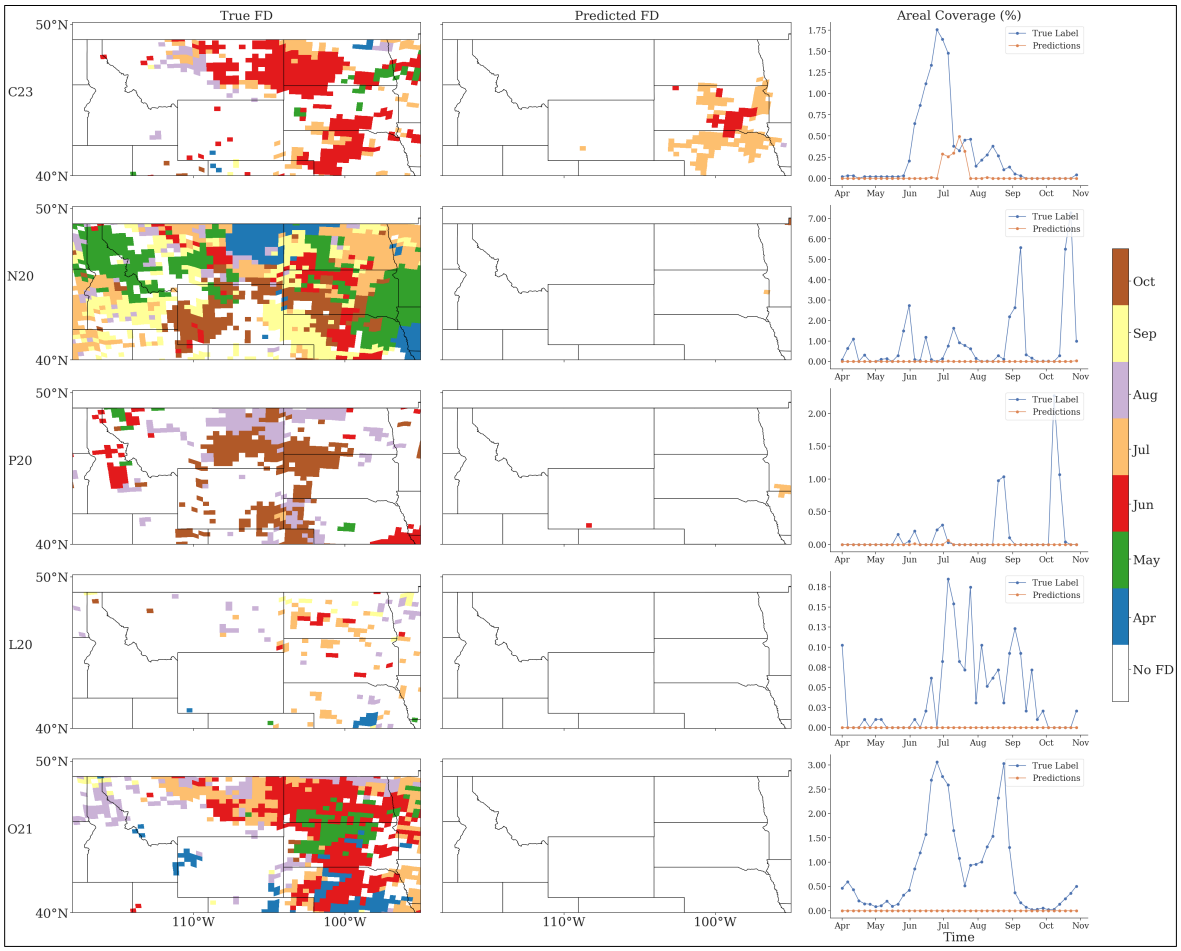


Figure 5.26: FD case study for 2017 for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the ANNs.

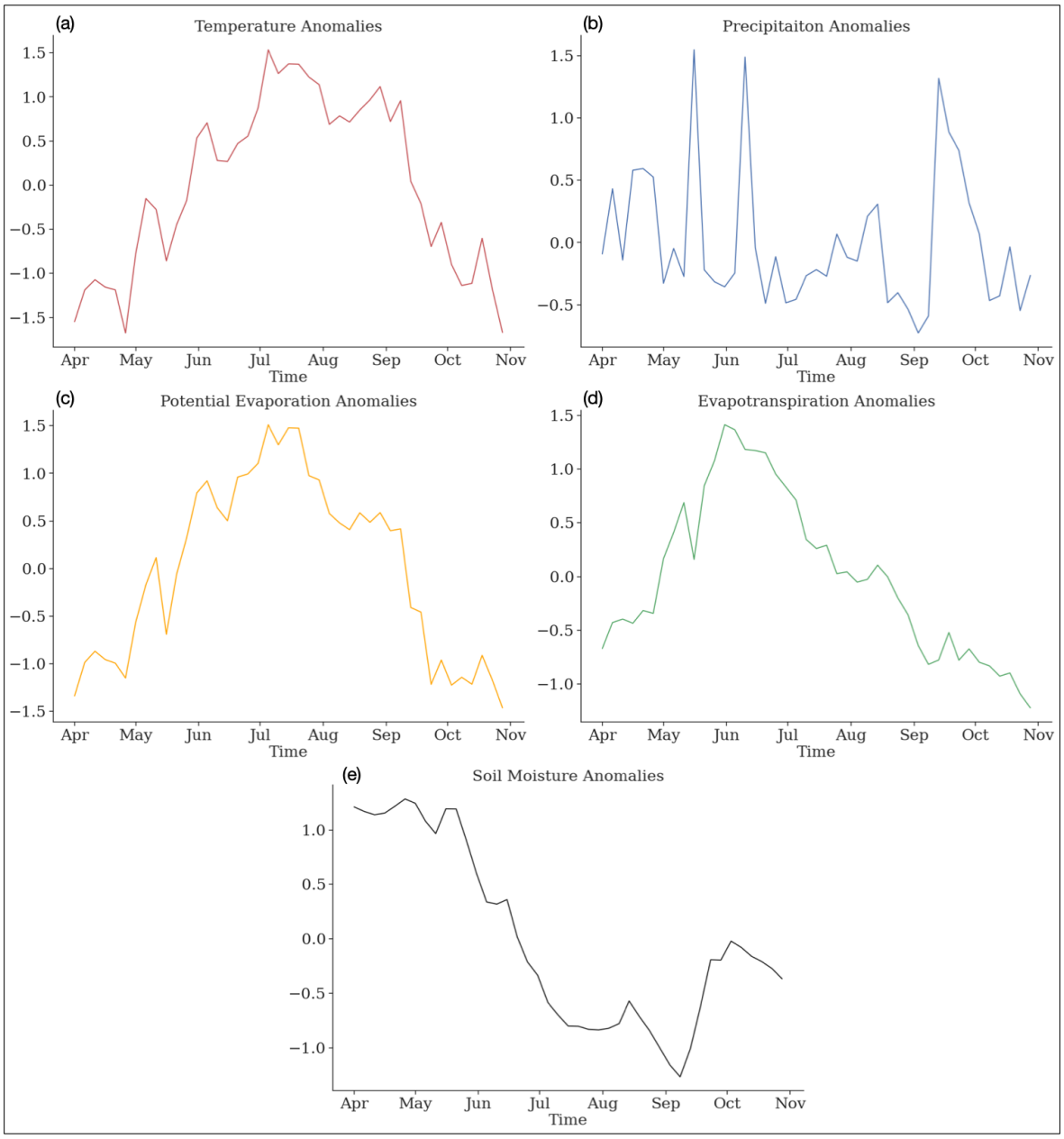


Figure 5.27: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2017 case study.

Chapter 6

Machine Learning on Global Flash Drought

6.1 True Climatologies

The climatology of the truth labels, that the ML models were trying to predict are shown in Figure 6.1. Global patterns and trends for the C23 method have already been studied in Christian et al. (2021). In addition, Christian et al. (2021) identified 15 regions across the globe that represented FD hotspots. Because these are areas that are prone to FD or have more impactful FD (typically due to the presence of intensive agriculture), global time series and seasonalities were examined with respect to these regions rather than the overall global time series (which would not represent the heterogeneous trends; e.g., Fig. 6.7).

In examining the climatologies, some similar patterns were found across all FD identification methods, including FD hotspots in the eastern Amazon, western Europe (in the Iberian Peninsula), northern Australia, central United States, and eastern and southern Asia (to varying degrees). Some contrasts between the different methods were also found. The L20 method, for example, placed emphasis on FD in the tropics. This is potentially due to it relying on declines in SM to measure FD, and given the tropics have heavily leached soils (so the soil moisture declines more easily and frequently). In contrast, the O21 method quantified SM decline differently and required that the FD last longer than a month (as that method sought to measure impactful FDs) making the distribution of FD from this method more uniform. However, the

O21 method still focused on similar hotspots to the C23 method (that is, regions of moisture and precipitation gradients, such as the Sahel and central Eurasia). The N20 method showed a high rate of FD identification, suggesting rapid declines in atmospheric moisture availability (measured via SPEI) was more common than other variables, and that many regions were sensitive to this type of moisture variation. Nonetheless, some signals could be distinguished, such as western China and southeast Asia, western Amazon, western Europe, and mountain slopes (e.g., eastern slopes of the Andes and Rocky Mountains, and the western slopes of the Rocky Mountains in Canada and Alaska). Lastly, the P20 method highlights many of the regions all the other methods do, but more weakly given it is looking for large increases in PET, which can be difficult to cause in areas that do not experience rapid increases in moisture demand (such as the tropics due to its consistent temperatures).

6.2 Bulk Statistics

The bulk skill scores for the global ML models are given in Table 6.1. Skill scores showed a notable decrease in the ML model skill compared to the Ada boosted trees and RNNs in Chapters 4 and 5, except for the P20 and O21 methods for the RNNs, which saw only a marginal decrease in skill (by 0.01). As a whole, this was not too surprising given the heterogeneous nature of the world, and the increased difficulty of moving from a regional to a global scale. However, the level of decreased skill in the Ada boosted trees and RNNs (for the C23, N20, and L20 methods) was notable, and some important trends could be seen in the composite mean differences (Fig. 6.2 and 6.3). In general, both ML models struggled noticeably with capturing global trends in FD. In general, the ML models tended to focus on a select few locations at the expense of the rest of the globe. For example, for the C23 method, ML models focused

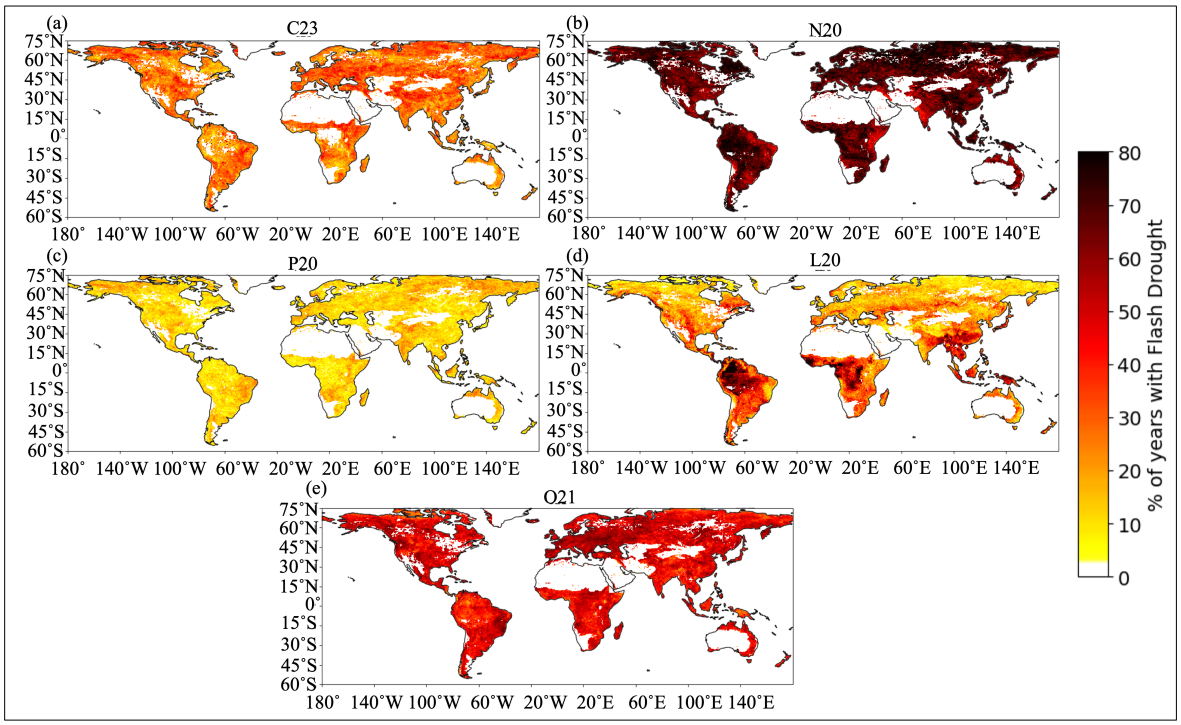


Figure 6.1: Percentage of years from 1971 – 2021 in which FD occurred across the globe according to the a) C23 method, b) N20 method, c) P20 method, d) L20 method, and e) O21 method.

on Mid-Latitude FD at the expense of under predicting FD in the tropics. For the L20 method, ML models focused heavily on the tropics at the expense of the rest of the globe (where they predicted almost no FD), and the Ada boosted trees focused more on the High Latitudes for the N20 and P20 methods. Note the RNNs had great difficulty predicting FD for the N20 method (see Section 6.3), and ultimately settled on a solution of predicting identical flash drought values everywhere, resulting in constant over predictions of FD everywhere (composite mean difference $<$ -0.1; Fig. 6.2b). For the O21 method, the ML models under predicted FD across the globe except for regions that were supposed to be masked out by the aridity index, where the ML models did predict FD and thus had over predictions (see Section 6.3 for a discussion). In some cases, the ML models captured some signals. For example, the strong over emphasis on the tropics for the L20 method came from the strong emphasis on the tropics in the true labels, which the ML models over emphasized as chapters 4 and 5 showed they were prone to do, and the C23 method focused more on areas of energy and moisture transitions in the Mid-Latitudes, which the Ada boosted trees focused more on (see Section 6.3).

These singular trends of emphasizing performance in certain latitudes dominated ML model predictions above any other pattern or signal. While problematic in terms of global predictions, this did suggest areas where ML models may be more effectively implemented for individual methods. For example, studying FD using the P20 and N20 methods via Ada boosted trees may have better results focusing on the High Latitudes, or the Low Latitudes if studying the L20 method with either ML method, or the Mid-Latitudes for the C23 method. While the patterns from the composite mean differences can be interesting, it is worth noting that none of them were statistically significant for either ML model (i.e., statistical significance produced empty maps; not shown), and as such, these results need additional evidence to support them.

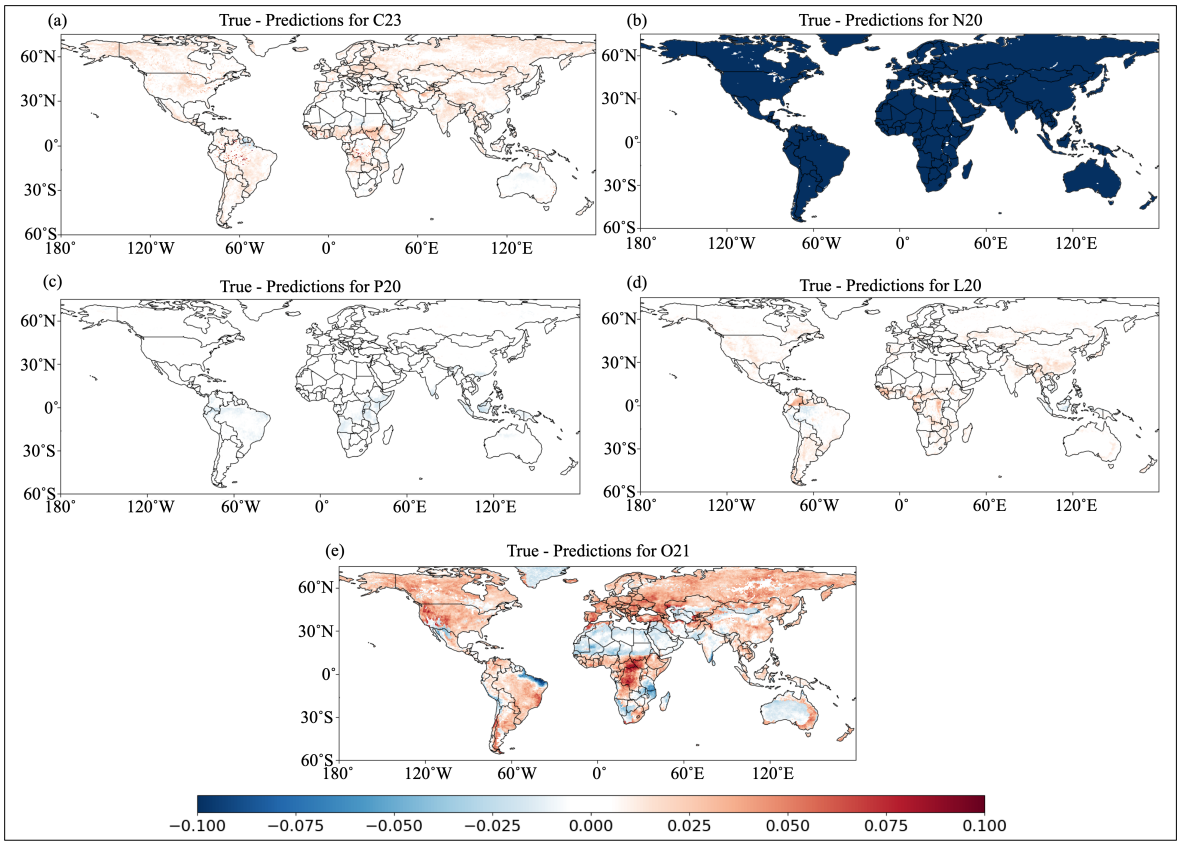


Figure 6.2: Composite mean difference of true minus predicted FD labels for each FD identification method using global RNN predictions.

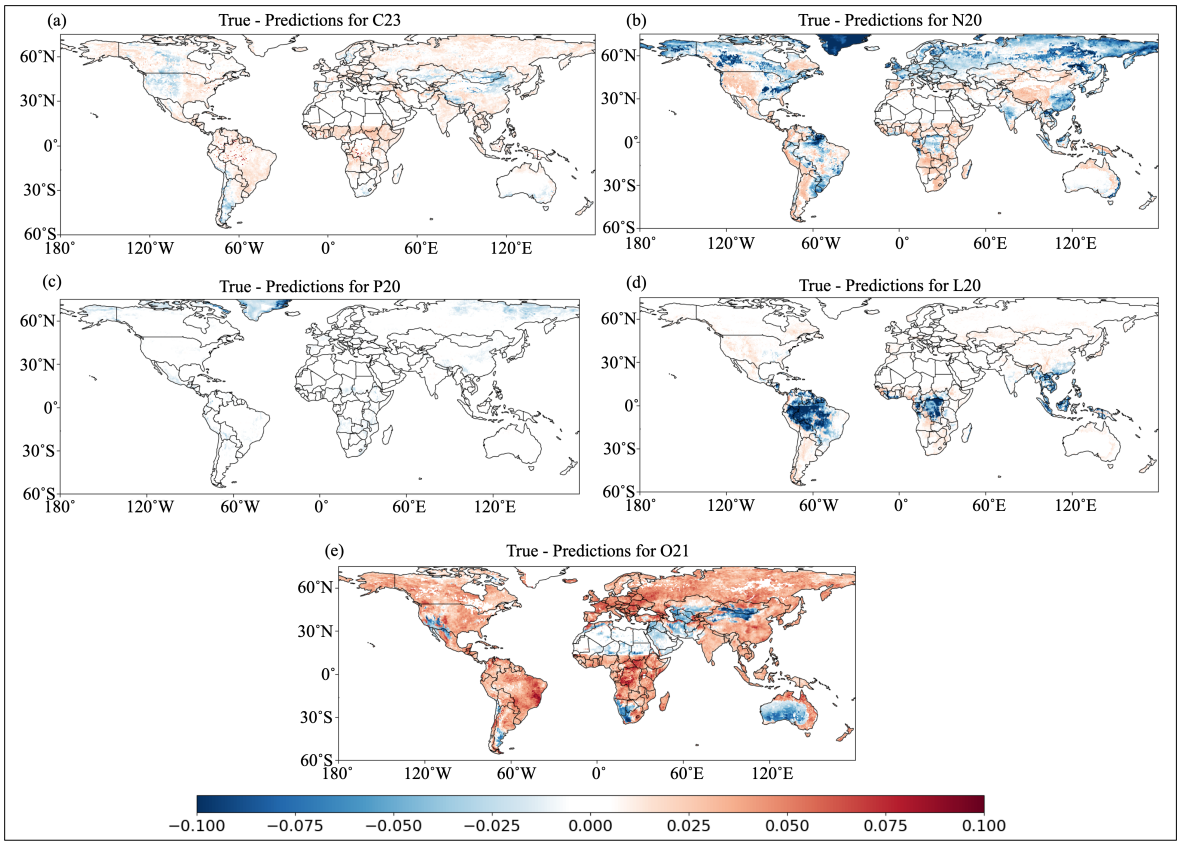


Figure 6.3: Composite mean difference of true minus predicted FD labels for each FD identification method using global Ada boosted tree predictions.

	Ada Boosting	RNNs
C23	0.03 (±0.003)	**0.05 (±0.013)**
N20	**0.24 (±0.017)**	0.12 (±0.047)
P20	0.04 (±0.007)	**0.06 (±0.011)**
L20	**0.09 (±0.007)**	0.06 (±0.015)
O21	0.04 (±0.006)	**0.09 (±0.021)**

Table 6.1: True skill statistic over all grid points and pentads for Ada boosted trees and RNNs across the globe and FD identification method. Numbers in parentheses indicate 95% confidence intervals derived from a 1-sample t-test (calculated across all folds). Highest skill score for an identification method is bolded.

The lower skill scores in Table 6.1 can also be attributed to the difficulties in capturing heterogeneous temporal patterns of FD across the globe (see Section 6.3). Lastly, the performance of the RNNs for the N20 method should be noted. As a whole, the RNNs had struggled with this method more than the others, and frequently produced test experiments that would either predict no FD anywhere (one type of trivial solution) or FD everywhere (another type of trivial solution). While alternate parameters were being explored, hyperparameter tuning was cut short due to lack of time and parameters yielding some non-trivial experiments were chosen. However, these parameters still broadly resulted in trivial solutions in the test datasets for most of the rotations for this method, resulting in the poor performance found in Figure 6.2 and in the next section, while over prediction of FDs resulted in a seemingly higher skill that was found in Table 6.1.

The performance of the Ada boosted trees can be partly understood via the GINI feature importances (Fig. 6.4), which showed a mixture of variables ranking most important. This includes PET and SM for the C23 method, P for the N20 method,

P and PET for the P20 method, T and ET for the L20 method (hence the vast over emphasis on the tropics), and P and SM for the O21 method (which is likely why this method had some of the best climatological predictions as the Ada boosted trees tried to learn SM patterns for the O21 method). In short, the Ada boosted trees appeared to have been unable to find general patterns in FD that generalized to the globe, and defaulted to some of the climatological patterns of the most important variables (e.g., the low P and PET in the poles, and high T and ET in the tropics). The Ada boosted trees that put more emphasis on SM seemed to produce more realistic predictions, in part because of SM's more heterogeneous climate patterns, but also because they did so for methods that were correlated with SM (C23 and O21).

6.3 Climatology Predictions

The climatologies predicted by the RNNs are shown in Figure 6.5 and by the Ada boosted trees in Figure 6.6. The spatial distribution of predicted FD events discussed in Section 6.2 can be seen, with over emphasis of FD in certain locations at the expense of others (though the RNNs generally under predicted FD). The climatologies that the two ML models gave reasonably resembled true global climatologies, with distinguishable spatial patterns and features, for the C23, P20, and O21 methods. Climatolgical predictions for the L20 method by the RNNs seemed reasonable, however the RNNs still heavily focused on the tropics at the expense of identifying almost no FD at higher latitudes.

For the C23 method, the ML models captured a lot of Mid-Latitude patterns, though the ML models almost failed to find FD in the tropics altogether, despite FD being present in those locations in the truth labels (Fig. 6.1a; the RNNs predicted FD in the tropics for the C23 method better than the Ada boosted trees, but still had blank

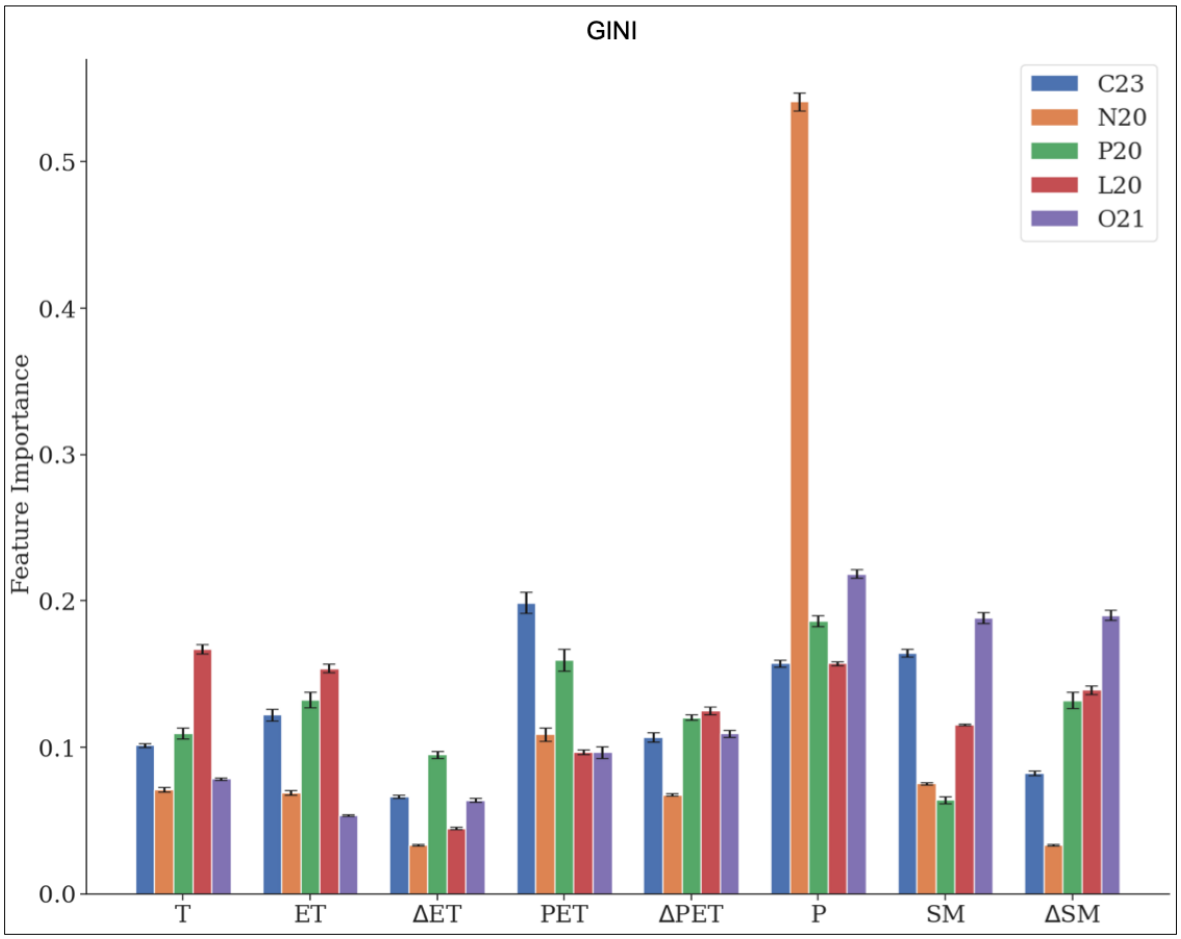


Figure 6.4: Average feature importance for the global Ada boosted decision trees according to the GINI method. Feature importance was determined for all rotations and bars show the average importance across each rotation. Error bars indicate 95% confidence intervals from a 1 sample t-test (average and standard deviation taken across all rotations).

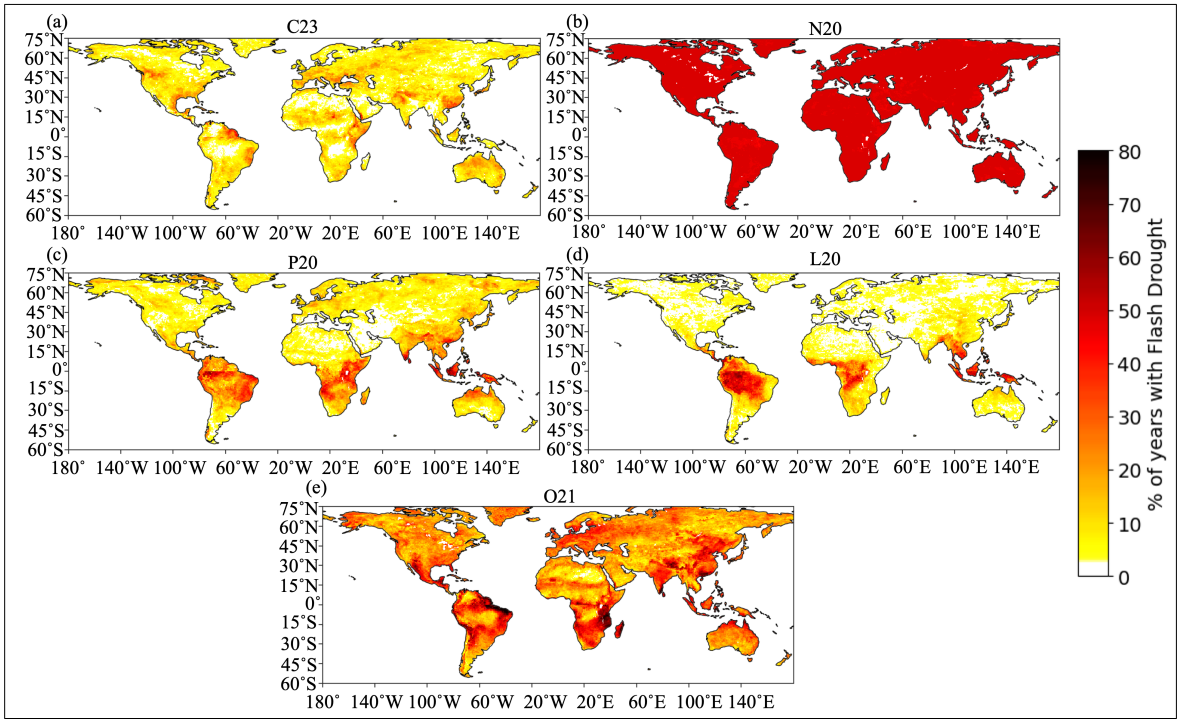


Figure 6.5: Global FD frequency climatology predictions from the RNNs.

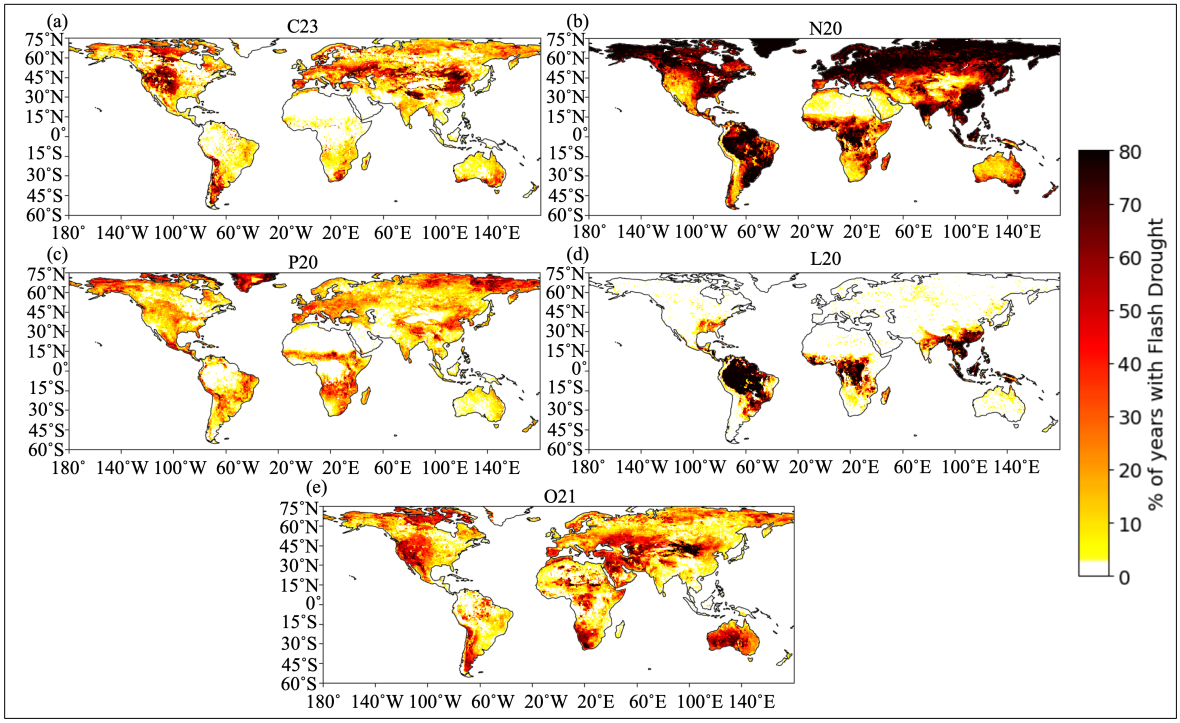


Figure 6.6: Global FD frequency climatology predictions from the Ada boosted trees.

patches of FD in the tropics). Spatial patterns in Ada boosted predictions for the C23 method also followed areas that are sensitive to precipitation deficits, such as the dry western United States, Himalayan Mountains, central Eurasia, transition regions from the humid Black and Caspian Seas to dry Siberia, and Argentina, on the lee side of the Andes Mountains. However, this can be difficult to interpret as learning precipitation deficits, as these are areas that were masked out by the aridity index that the Ada boosted trees were supposed to ignore, suggesting the Ada boosted trees seemed to think the masked out locations had higher frequencies of FD (in general, RNNs learned to ignore or have few FDs in these masked locations, though Ada boosted trees would sometimes emphasize them). Alternatively, it could be argued the Ada boosted trees learned about areas of precipitation gradients in the Mid-Latitudes for FD predictions. Given the Ada boosted trees' emphasis on PET, P, and SM for the C23 method (Fig. 6.4), the latter explanation seems more likely. A similar argument could be made for the O21 method, which identified similar hotspots in arid regions, as the Ada boosted trees focused on P, SM, and ΔSM for that method.

For the P20 method, note the Ada boosted trees heavily over predicted FD in eastern Siberia, Alaska, and Greenland (Fig. 6.6c; areas with some of the coldest global temperatures, and lowest precipitation values). Given the Ada boosted trees for the P20 method focused on P, it is likely the boosted trees simply learned precipitation climatology rather than FD patterns. Interestingly, the RNNs had the opposite problem of focusing on the tropics. In both cases, the ML models failed to identify the similar distribution of FD the true P20 labels identified, and instead focused on a specific region. That said, the ML models still learned some patterns for FD, such as eastern and southern Asia, western Europe (more central Europe in the RNNs), and the eastern Amazon (Fig. 6.5c and 6.6c). This suggests the ML models were capable of learning some of the global spatial patterns for this method.

Lastly, both ML models under predicted FD for large swaths of the globe for the O21 method, but learned distinct signals similar to the C23 method (Fig. 6.5e and 6.6e). However the Ada boosted trees also learned some patterns in the tropics for the O21 method, such as the Congo Basic (eastern Africa for the RNNs), western and central Australia (northern Australia in the RNNs), and the central and northern Amazon. In general, Ada boosted trees and RNNs looked to be learning similar patterns for the O21 method as for the C23 method, except Ada boosted trees extended further into the tropics and dry soil locations than for the C23 method. While this helped the Ada boosted trees make more realistic predictions for the O21 method, it ran at the risk of creating FD hotspots that were not present in the truth labels (Fig. 6.1e), such as in western Australia, southwestern Africa, and much of the Middle-East. In contrast, the RNNs more successfully learned to ignore areas filtered by the aridity index, preventing them creating many new hotspots, but also resulting in wider under predictions of FD compared to the Ada boosted trees.

In terms of predicted time series and seasonalities of FD, both ML models struggled significantly with these. When examining the overall time series in the FD prone regions identified in Christian et al. (2021) (Fig. 6.7, 6.8, 6.9, 6.10, and 6.11), the ML models notably struggled to capture the temporal patterns, either under predicting FD (for the C23 and P20 methods, the L20 method outside of the tropics, and the O21 method after about half way through the time series), or over predicting FD (N20 method for the Ada boosted trees, and L20 method within the tropics). ML models seemed to take some consideration of cliamte change, but not in a way that helped the models. For example, in all time series for the O21 method, FD was over predicted by RNNs, but then started to predict fewer or less expansive FD about half way through the time series, resulting in large under predictions. This change from over prediction to under prediction of FD was most likely due to changing temperatures and soil

moisture conditions resulting from climate change. However, the ML models failed to learn climate change patterns that would help them or were present in the truth labels, such as the increasing trend in FD coverage in Asia Minor (Fig. 6.7 through 6.11). Interestingly, the Ada boosted tree algorithms also created U, and inverted U-shaped, time series for several methods (e.g., in the central United States and Argentina in the N20 method, tropical locations in the L20 method, and Asia Minor and Iberia in the O21 method; Fig. 6.8, 6.10, and 6.11).

Another notable feature was the seasonality of FD predicted by the ML models (Fig. 6.12, 6.13, 6.14, 6.15, and 6.16). For all methods, the ML models predicted the seasonality of FD to be at the start, or at the end, of the growing season (depending on location and identification method). This was most likely connected to the variables emphasized by the ML models responding to shifts in P, T (and PET by extension), and SM. Overall, this showed that the Ada boosted trees and RNNs failed to learn the temporal, and most especially the seasonality, of FD, with the ML algorithms over generalizing trends and thus not learning the heterogeneous patterns present across the globe.

An interesting note to make is that the seasonal patterns of FD did not significantly change in the truth labels across FD identification methods (Fig. 6.12 through 6.16), and neither did temporal trends (Fig. 6.7 through 6.11). For example, all identification methods showed increasing trends in FD coverage in Asia Minor and western Europe, and decreasing trends in India. So while the FD identification methods showed some variation in spatial hotspots, they all had similar temporal and seasonal patterns.

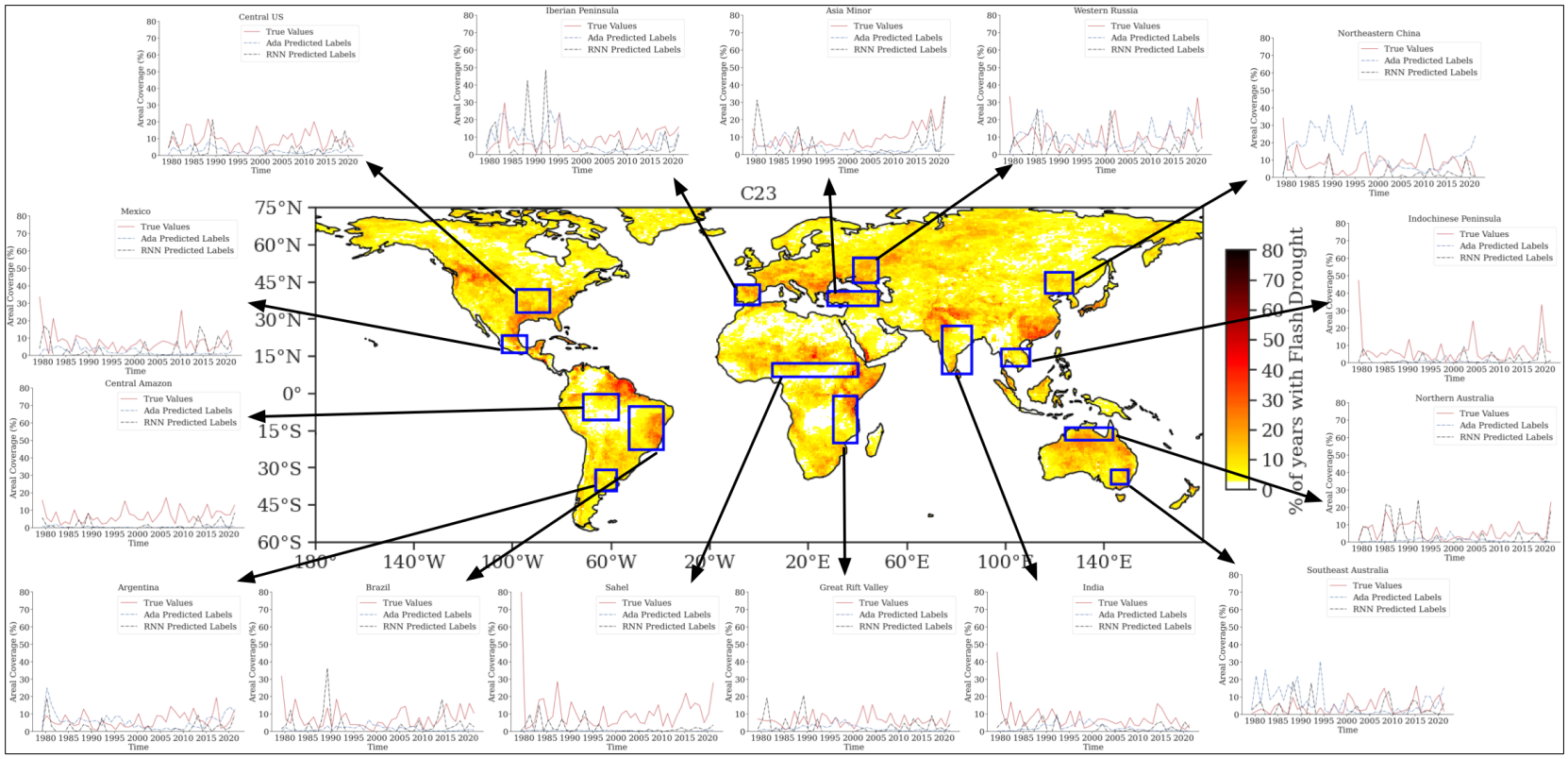


Figure 6.7: Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the C23 method, and true C23 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the C23 method.

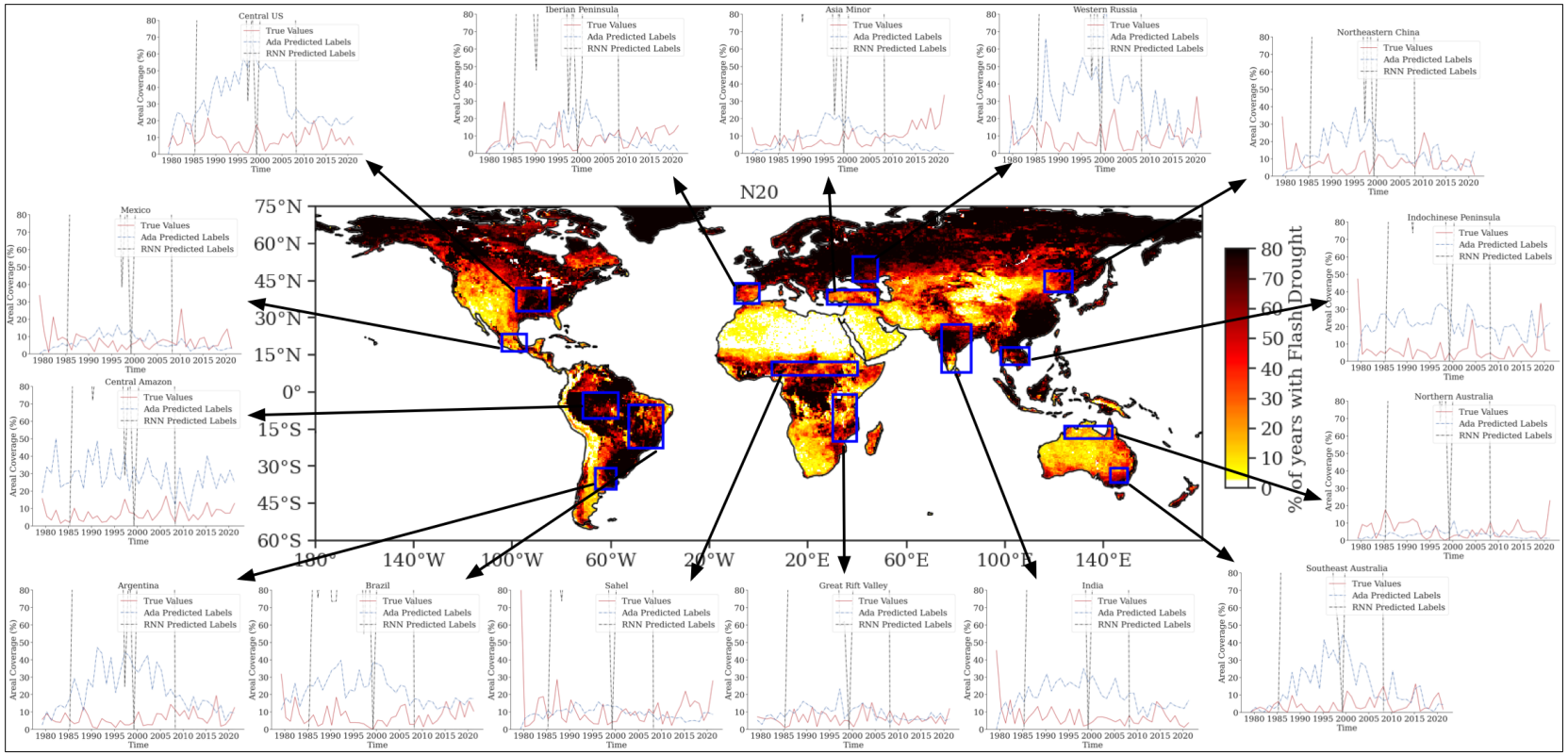


Figure 6.8: Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the N20 method, and true N20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by Ada boosted trees for the N20 method.

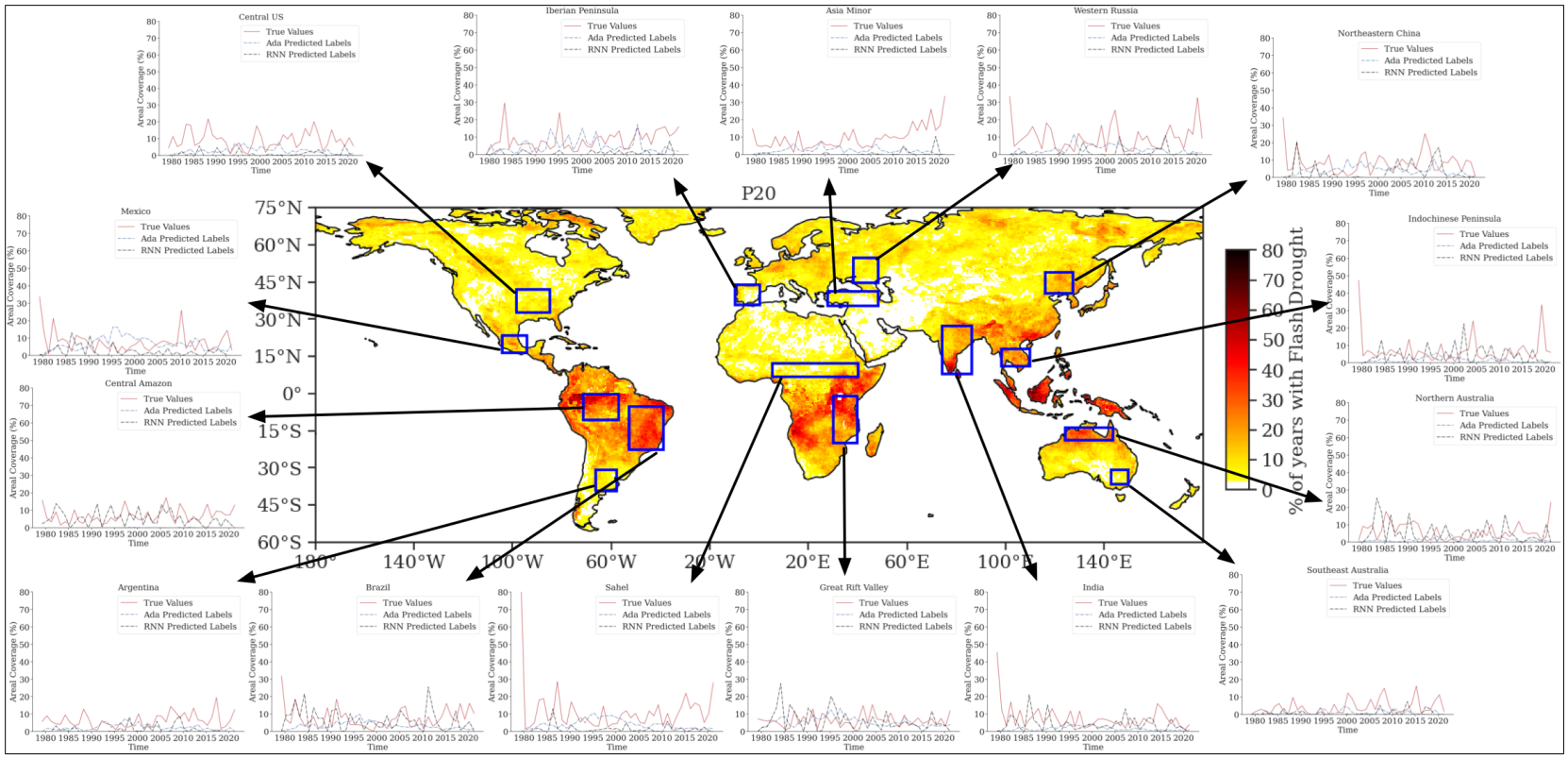


Figure 6.9: Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the P20 method, and true P20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the P20 method.

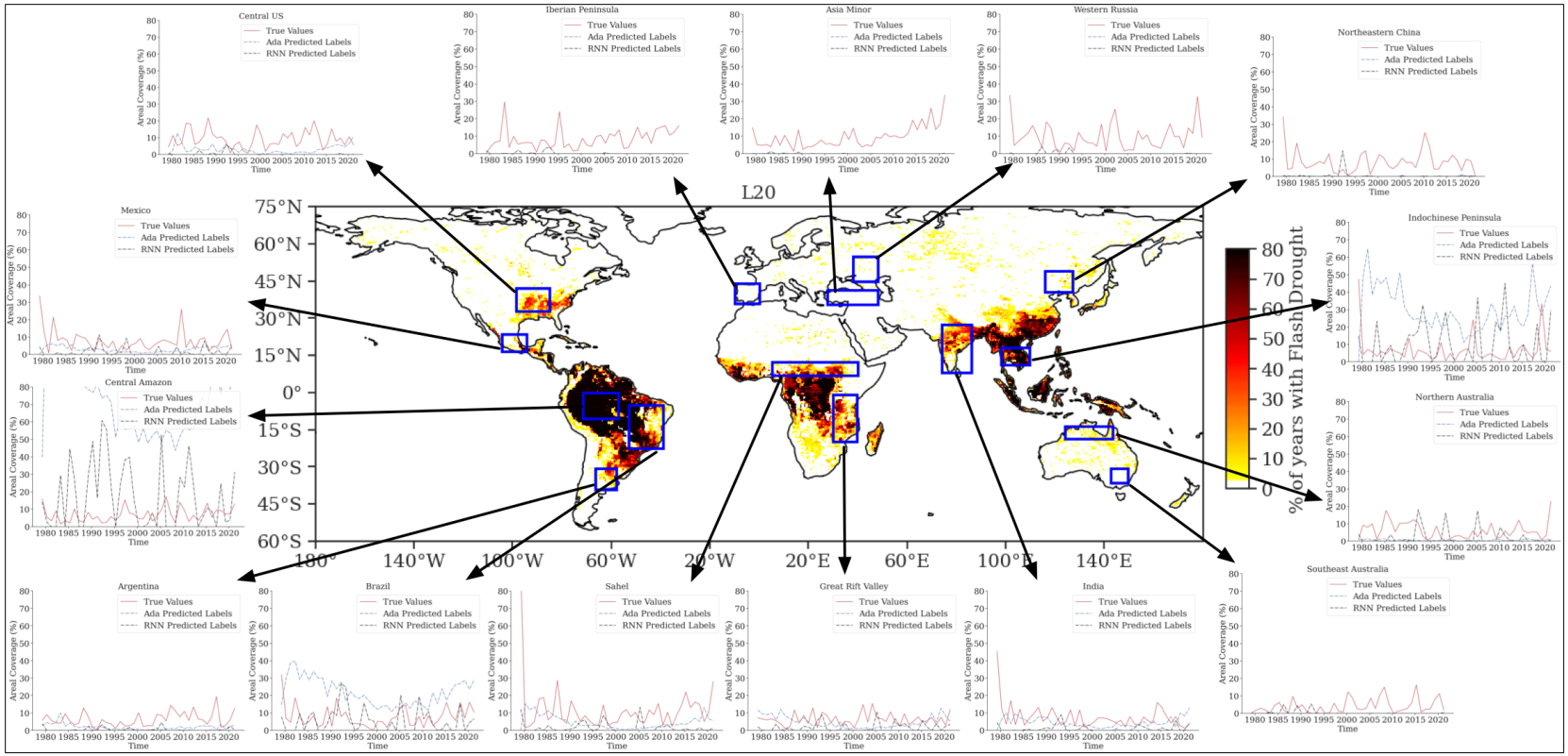


Figure 6.10: Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the L20 method, and true L20 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by Ada boosted trees for the L20 method.

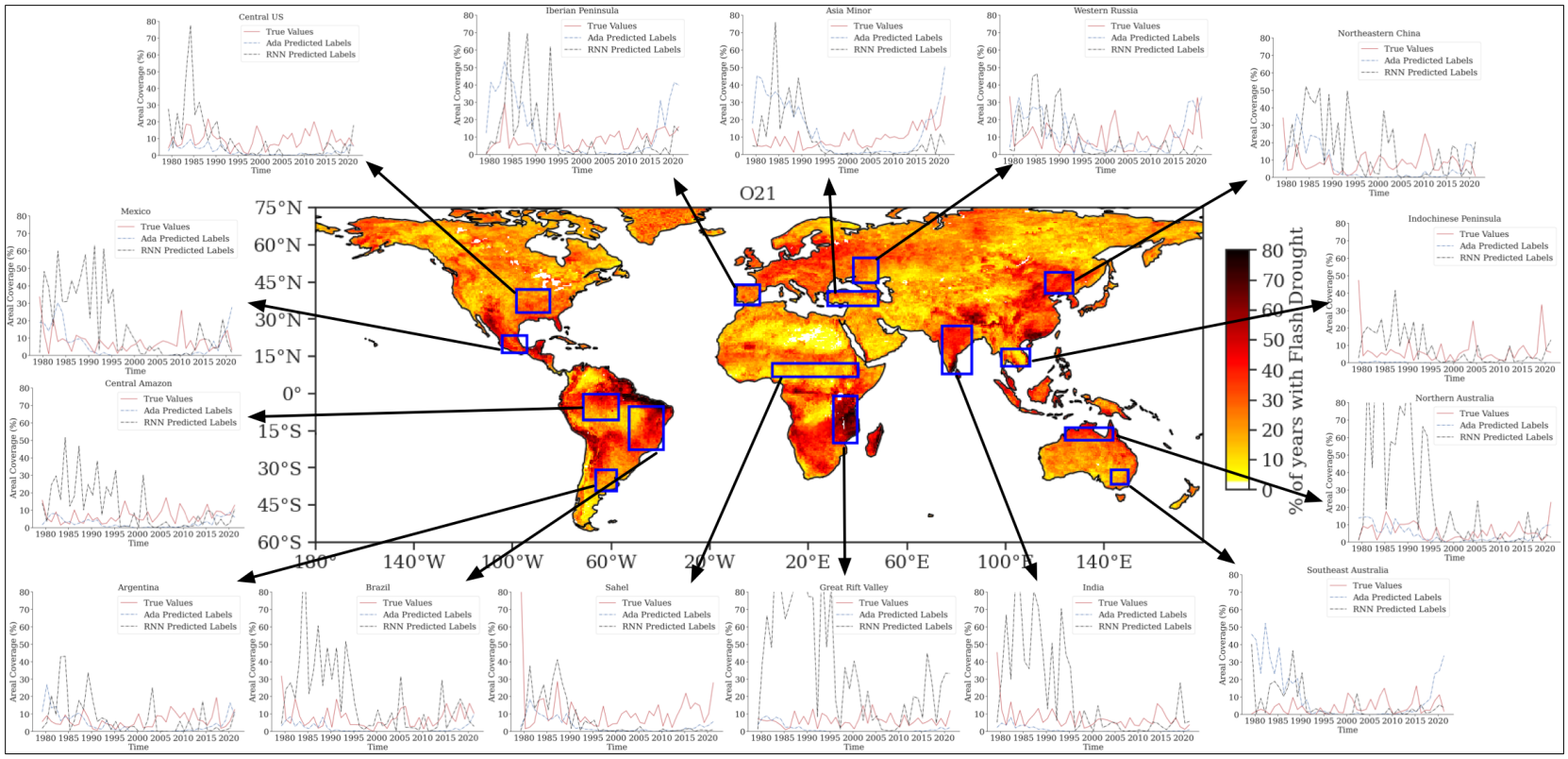


Figure 6.11: Time series of FD coverage according to Ada boosted tree (blue dashed line) and RNN (black dashed line) predictions for the O21 method, and true O21 labels (solid red line) for each flash drought hotspot region identified in Christian et al. (2021). FD climatology shown is the FD climatology prediction by RNNs for the O21 method.

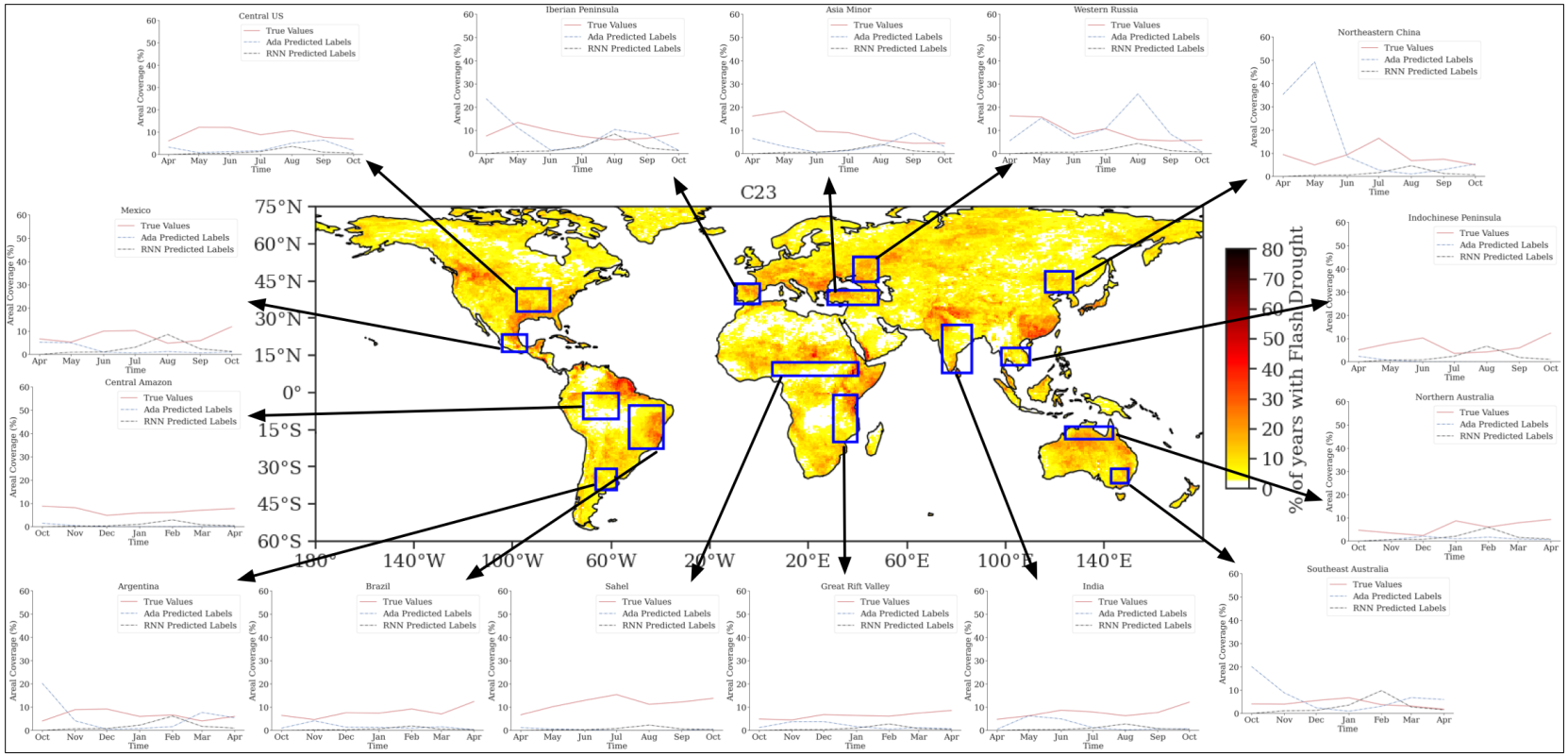


Figure 6.12: Average monthly FD coverage (i.e., FD seasonality) for the C23 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the C23 method.

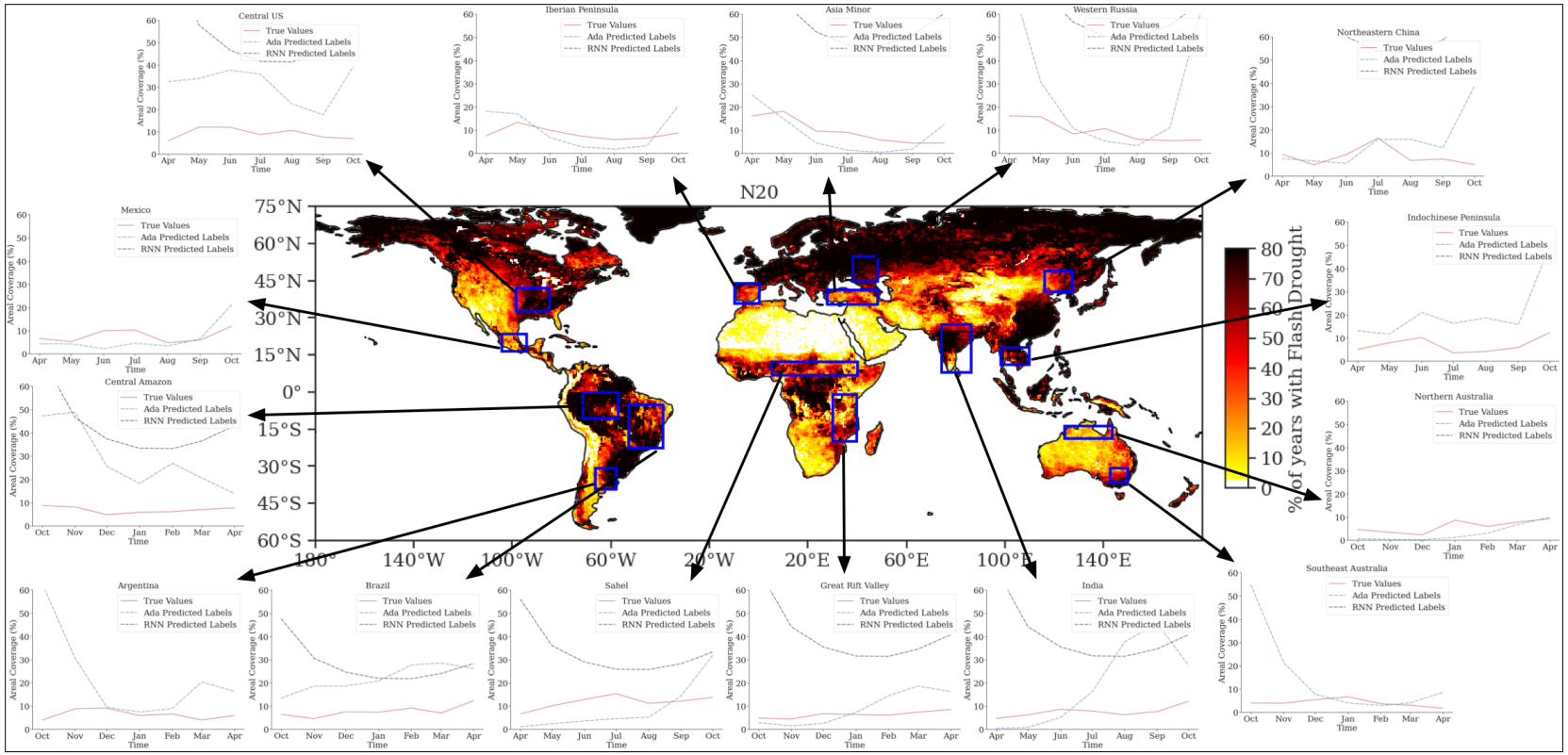


Figure 6.13: Average monthly FD coverage (i.e., FD seasonality) for the N20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by Ada boosted trees for the N20 method.

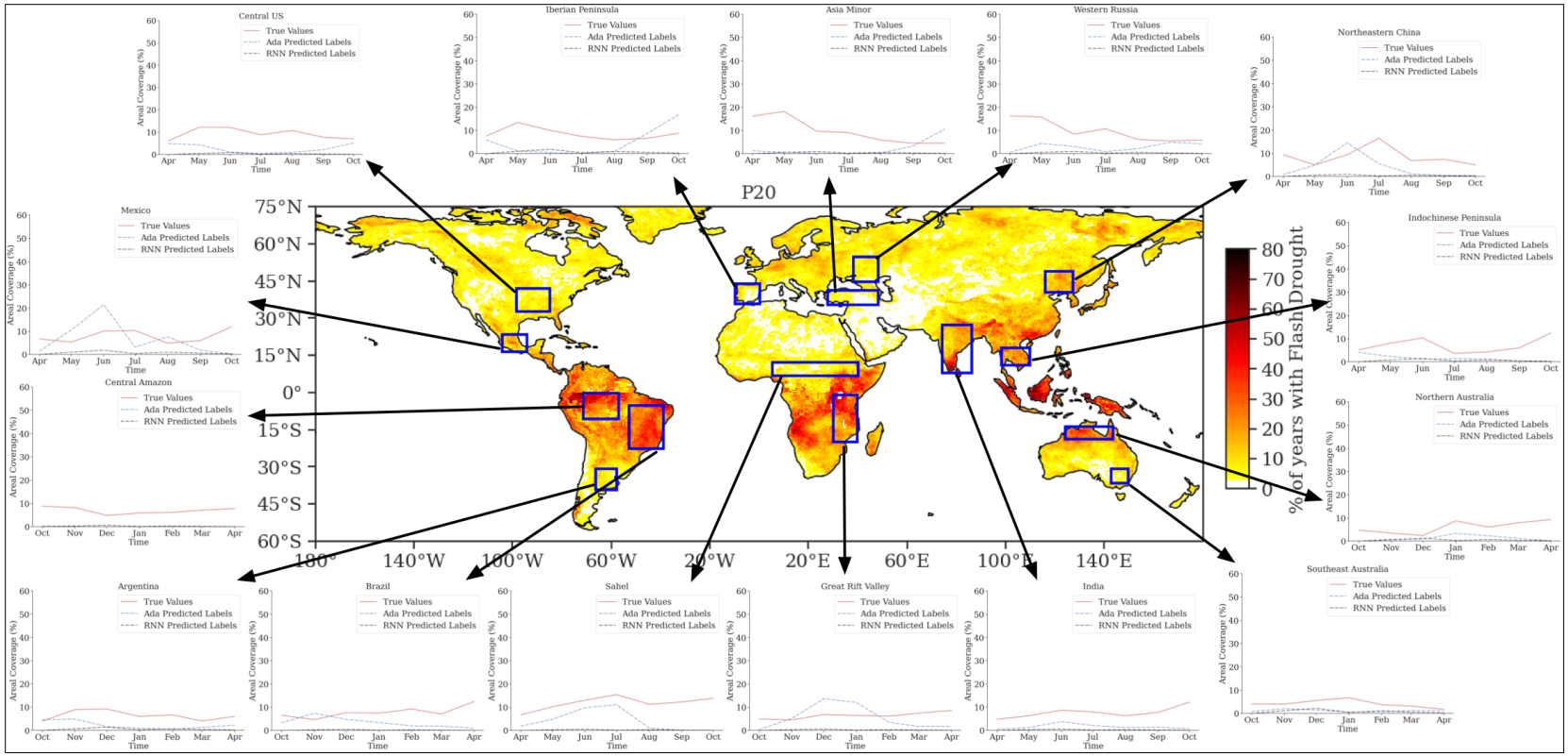


Figure 6.14: Average monthly FD coverage (i.e., FD seasonality) for the P20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the P20 method.

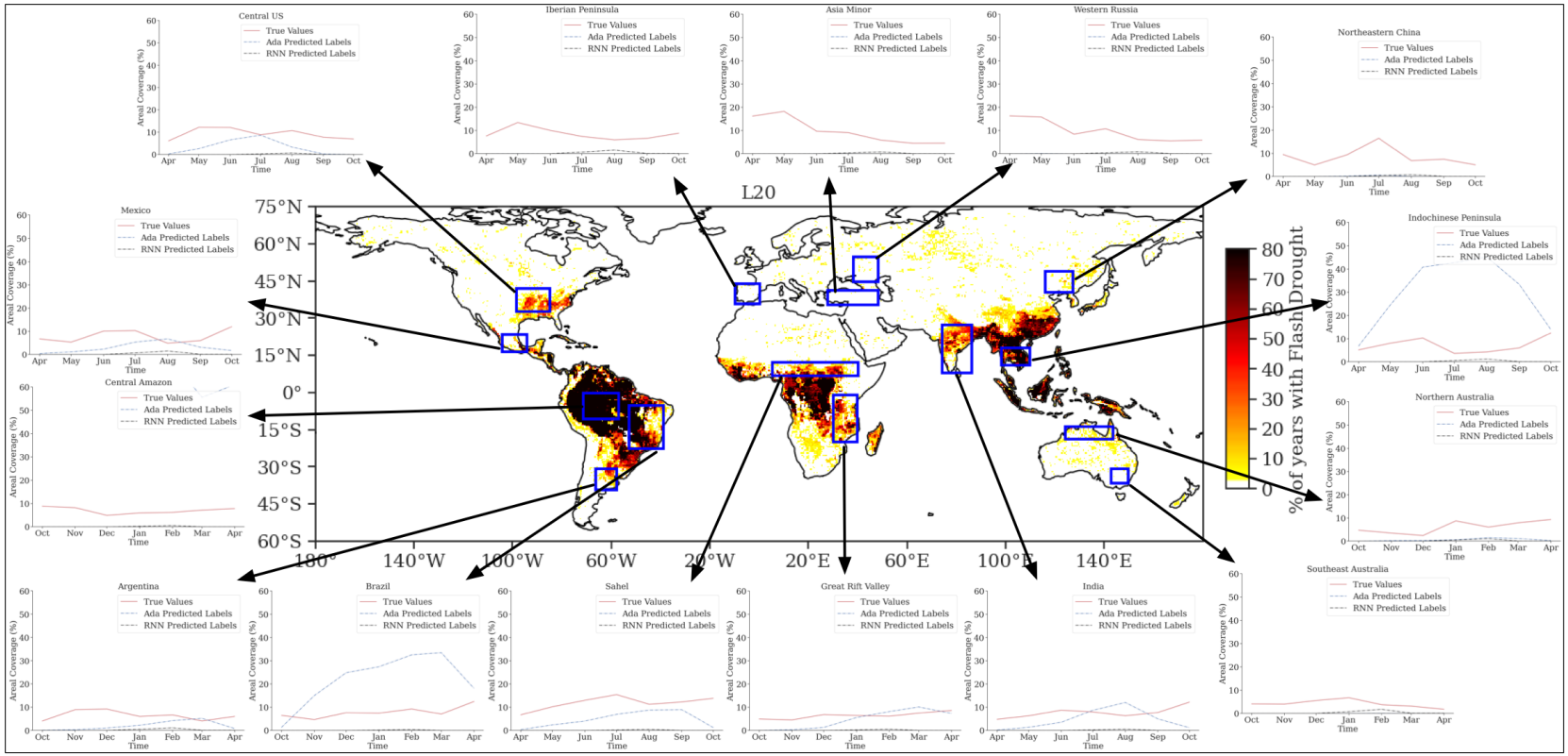


Figure 6.15: Average monthly FD coverage (i.e., FD seasonality) for the L20 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by Ada boosted trees for the L20 method.

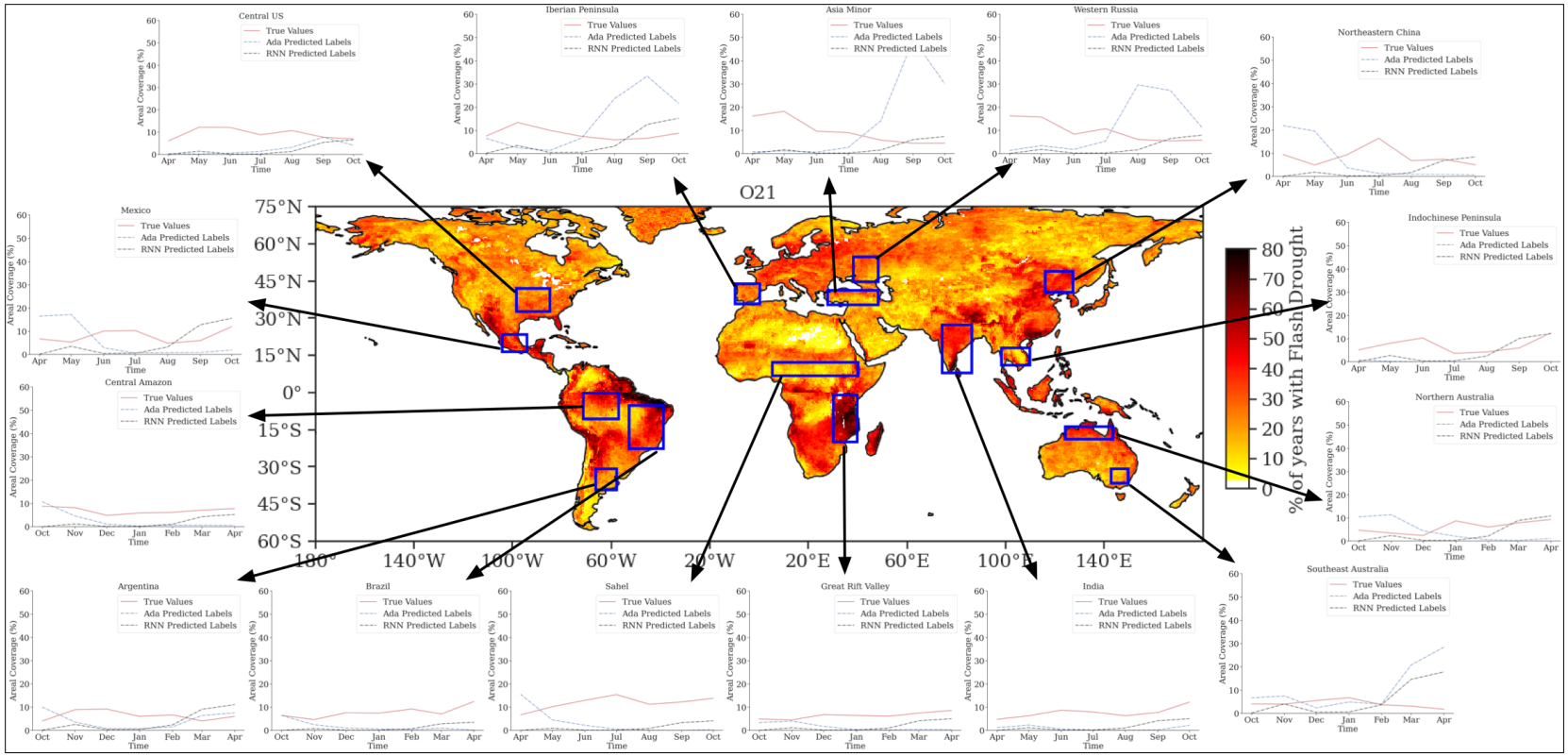


Figure 6.16: Average monthly FD coverage (i.e., FD seasonality) for the O21 method for each flash drought hotspot region identified in Christian et al. (2021) according to the truth labels (solid red line), Ada boosted tree predictions (blue dashed line), and RNN predictions (black dashed line). FD climatology shown is the FD climatology prediction by RNNs for the O21 method.

6.4 Case Studies

Finally, global case studies provide an opportunity to examine events across the globe and their impacts to see how said impacts might vary. For example, in 2001 a severe FD struck north central India and parts of southern India (Fig. 6.18 and 6.17; Mahto and Mishra 2020). As seen with other FDs in the United States, the timing of the FD event can depend on the FD identification method used. The FD developed in southern India in the early growing season, and later in northern India late in the growing season. In general, there is agreement in the location and timing of FD between the C23 P20, and O21 methods. This can be considered a case when FD did not propagate, but rather FD developed in another area due to a break in the monsoonal season (Mahto and Mishra 2020). The ML models in this case had a difficult time capturing this FD, though the Ada boosted trees for the N20 method captured much of the spatial extent of the FD, and when the FD peaked in spatial coverage (Fig. 6.18). However, the Ada boosted trees struggled with the other methods in capturing this event. The RNNs generally failed to identify the first onset of rapid drying for this event (except for the O21 method, where they prematurely identified FD too far east of the true labels). But the RNNs were able to represent the location, extent, and coverage of the second onset of rapid drought (in August to October) relatively well (Fig. 6.17, right column). Notably, this FD did not have a great deal of direct impact on people, but saw large impacts on the crop production, with over 25% of rice and maize production affected by this event (Mahto and Mishra 2020).

A second example can be seen in southern Russia, during 2010. During the early growing season FD began to develop in the north and propagate into southern Russia, worsening the impacts of a heatwave that also impacted Russia during the 2010 summer (Fig. 6.21 and 6.20; Christian et al. 2020). By July, the FD had moved into central

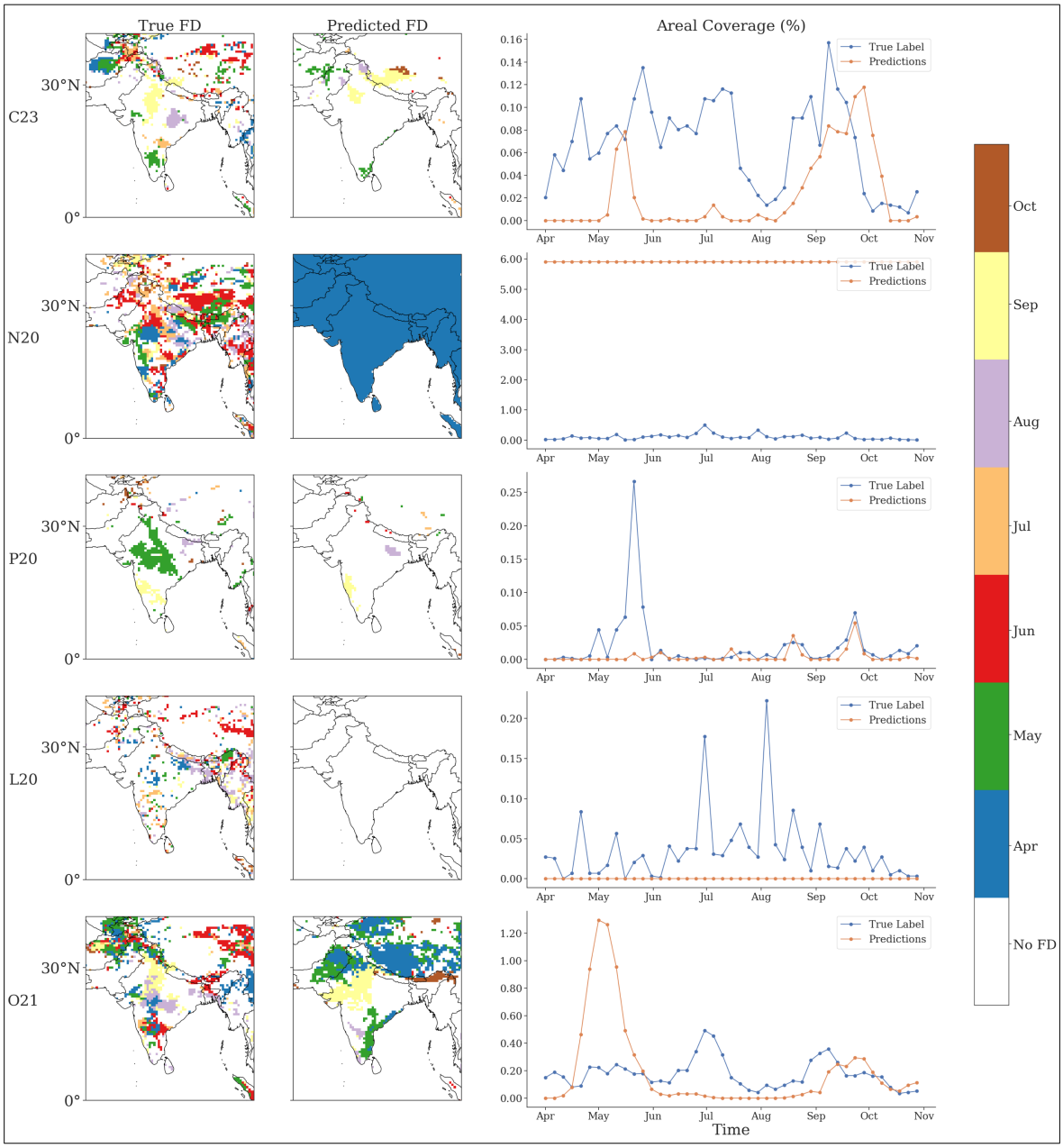


Figure 6.17: FD case study for 2001 in India for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

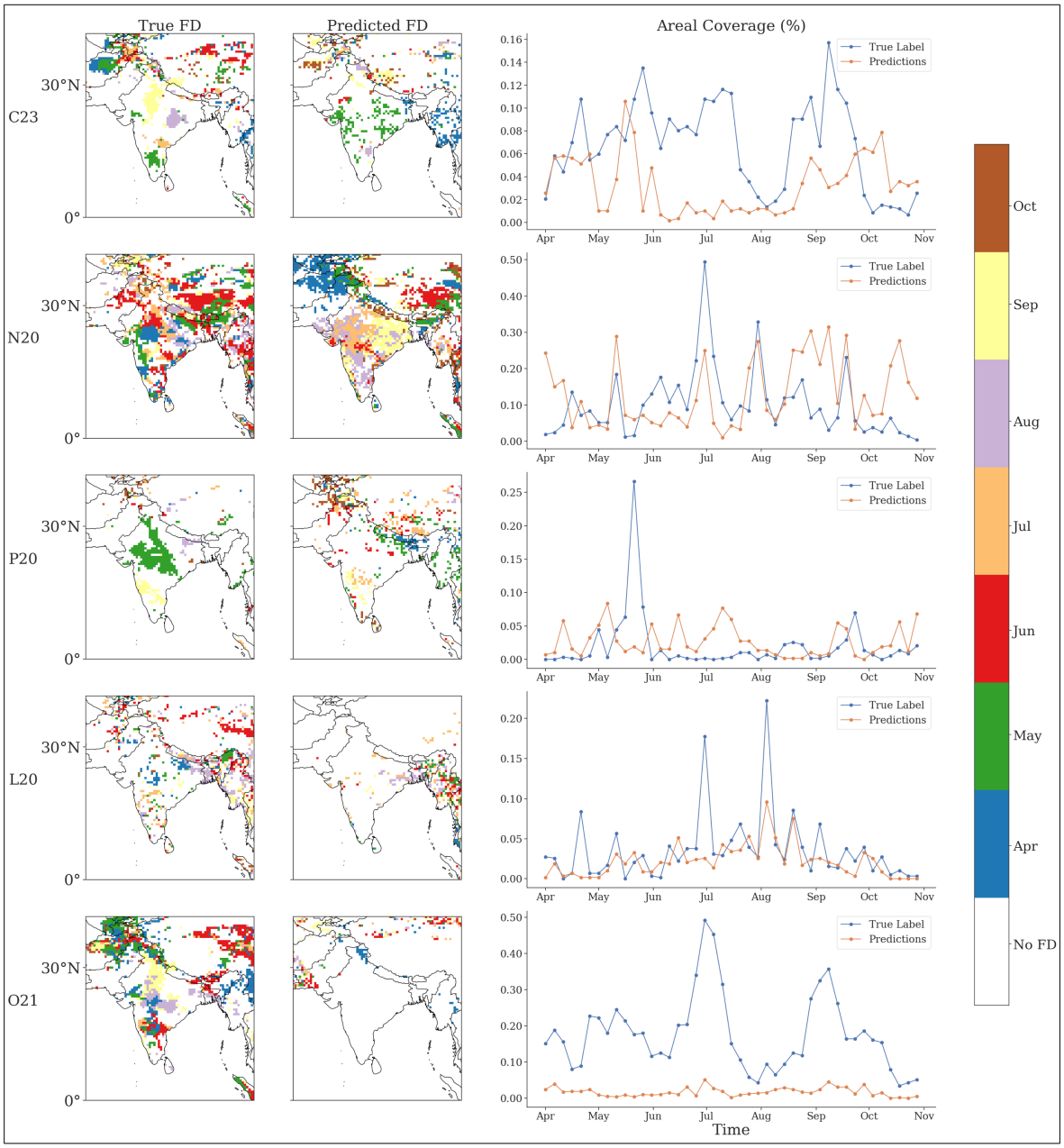


Figure 6.18: FD case study for 2001 in India for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.



Figure 6.19: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2001 case study over India.

Russia and exacerbated the heatwave and its impacts. Again the ML models had a difficult time in representing the FD event (in this case with both ML models and for all FD identification methods). In particular, the RNNs struggled to identify FD (Fig. 6.20), while the Ada boosted trees struggled with the timing of the event (Fig. 6.21). In contrast to the previous FD event discussed, the impacts of this FD were more diverse. While there was a drop in crop production during 2010, the FD amplified a severe heatwave, resulting negative impacts on public health and even causalities as a result of the heat to the Russian populace (Christian et al. 2020).

A different type of impact can be seen in the 2015 – 2016 drought in the Amazon. This event has been described as a traditional, long-term drought (Jiménez-Muñoz et al. 2016; Ribeiro et al. 2021), though elements of rapid intensification were detected in this event during the last two months of the Southern Hemisphere growing season (Fig. 6.23 and 6.24). The location of that rapid intensification varied depending on the FD identification method, with the C23 and O21 methods focusing on the central and eastern Amazon, and the remaining methods more on the eastern Amazon. However, the main impacts of the drought were more in the eastern and northern Amazon (Jiménez-Muñoz et al. 2016; Ribeiro et al. 2021), suggesting the soil depletion quantified in O21 and the evaporative stress in C23 were more inline with the impacts for this event. Given the Ada boosted trees were performing better in the tropics climatologically for N20 and L20 methods, there was little surprise that they performed better for this event for those methods, capturing both the timing and spatial extent shown in the true labels (though the Ada boosted trees, and RNNs, over predicted FD for the L20 method; Fig. 6.24). For the other methods (C23, P20, and O21) the Ada boosted trees struggled to identify FD for this event, consistent with Ada boosted trees under predicting FD in the tropics for those methods. In contrast, the RNNs did relatively well with the L20 method, though it did not find the FD in the eastern

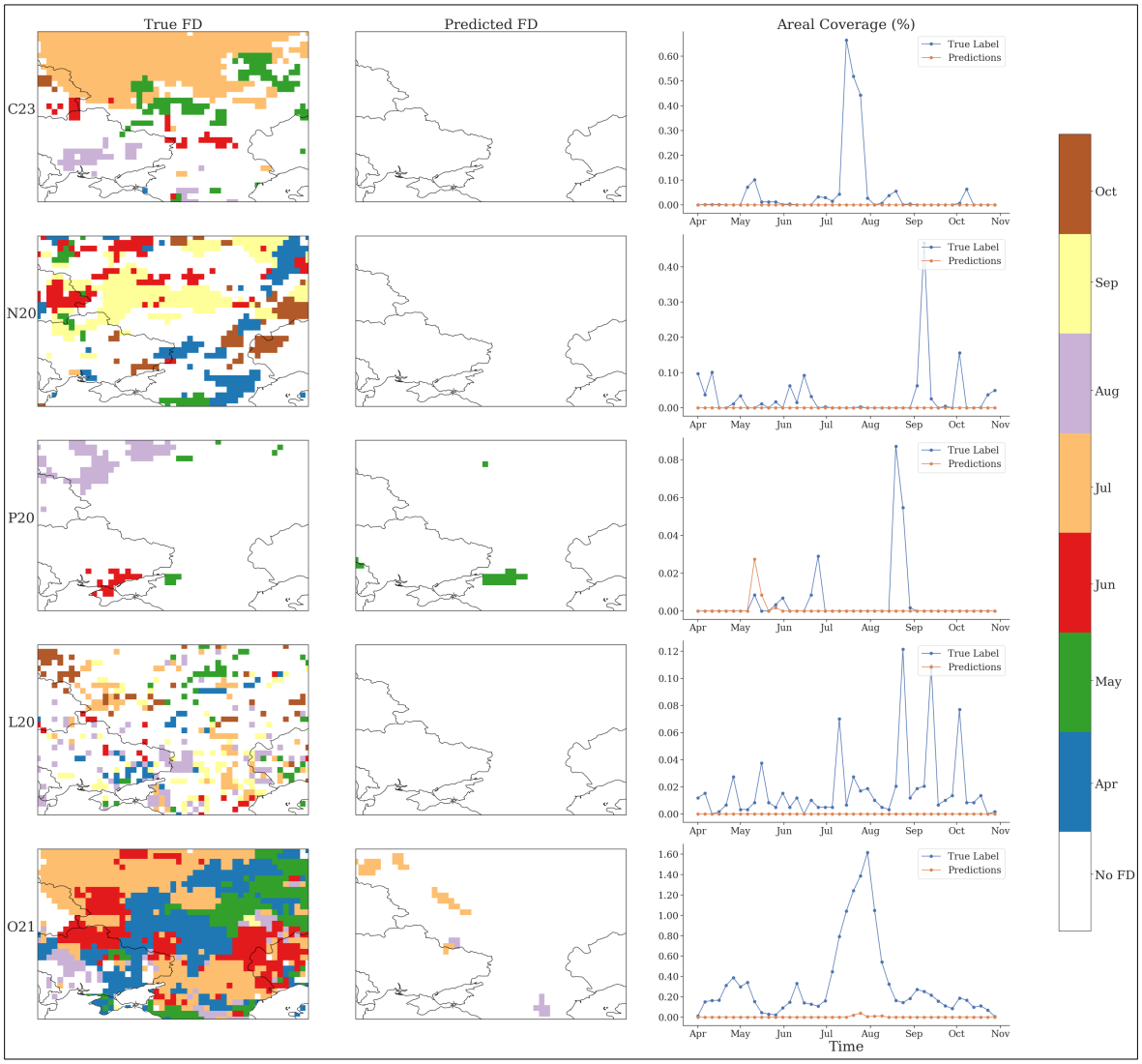


Figure 6.20: FD case study for 2010 in southern Russia for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

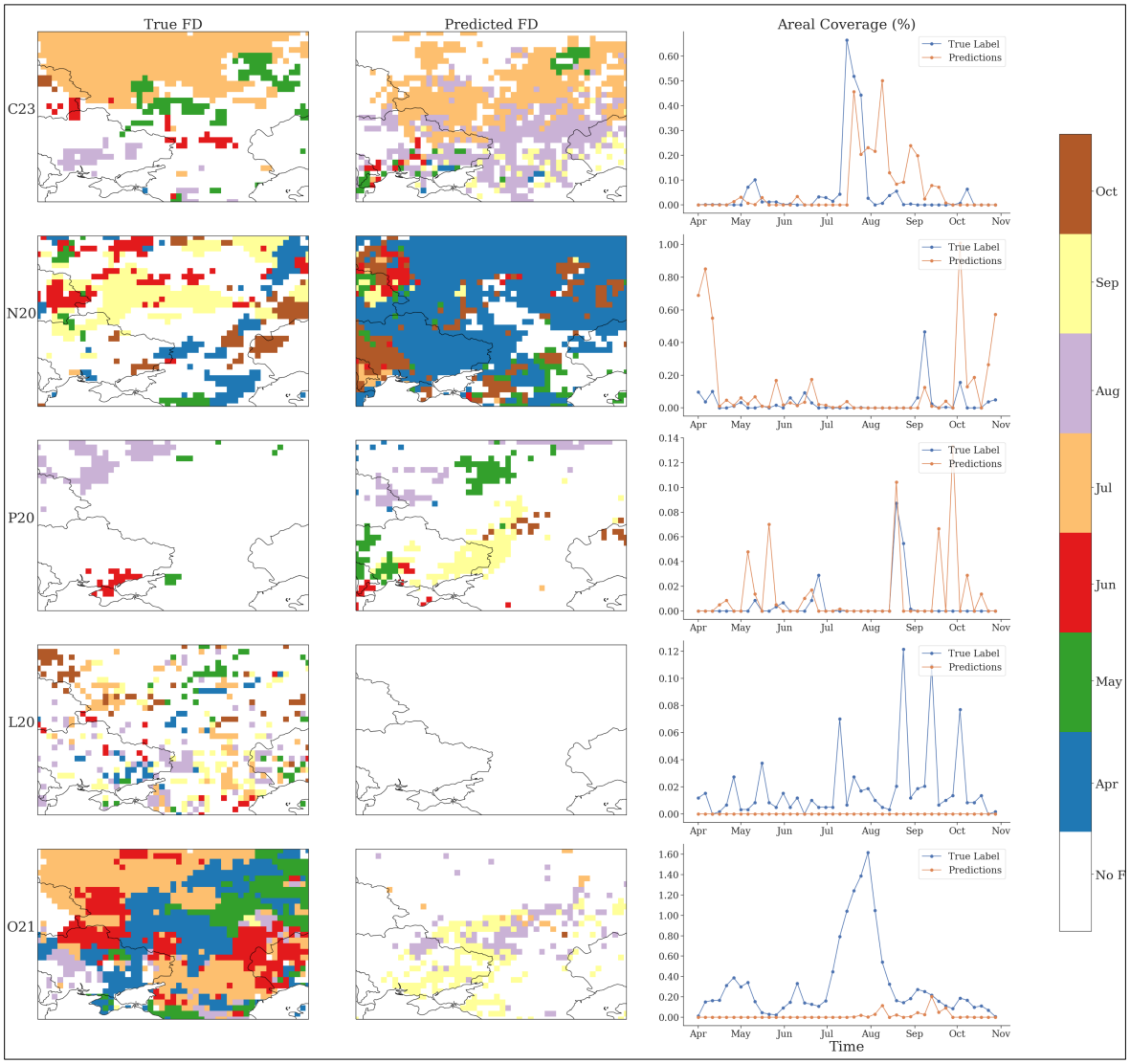


Figure 6.21: FD case study for 2010 in southern Russia for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.

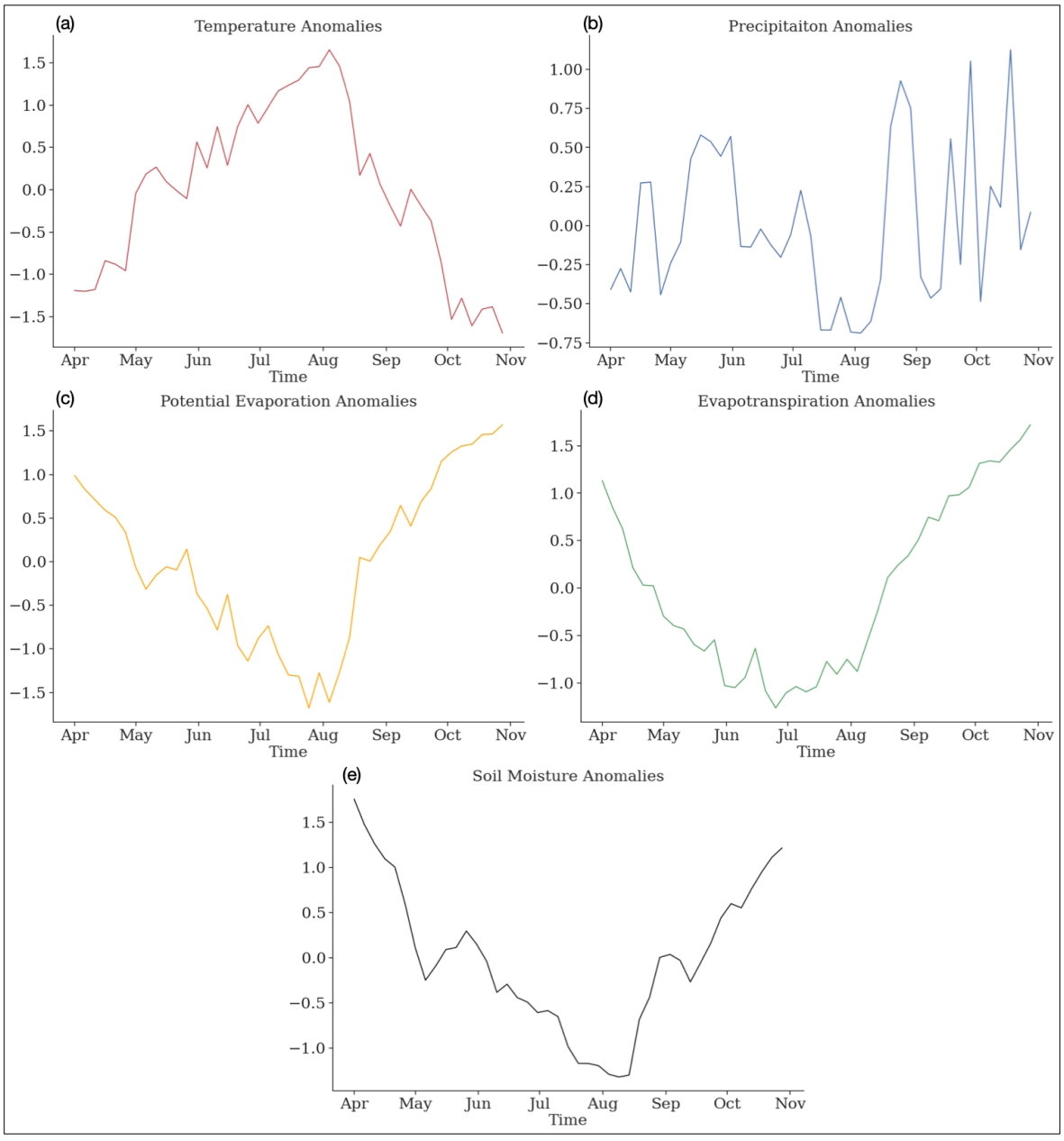


Figure 6.22: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2010 case study over southern Russia.

Amazon, and the RNNs struggled to find FD for the remaining methods (Fig. 6.23). Of note with this drought event was the impacts, as it did not impact people and crops directly, but rather was harmful to the vegetative landscape, particularly the tropical trees populating the area. Particularly, the trees that died as a result of the drought greatly impacted the Amazon's ability to act as a carbon sink (Jiménez-Muñoz et al. 2016; Ribeiro et al. 2021), resulting in a reduced carbon intake for the year and yielding a more unique drought impact in the form of aiding climate change due to vegetation loss.

The last case study is one in eastern Africa during 2016 – 2017. This particular drought was unique as the area of eastern Africa is close enough to the equator to have two rainy seasons when the Inter-Tropical Convergence Zone (ITCZ) passes over it for both hemispheric summer and winter seasons (Palmer et al. 2023). However, for this drought event, both rainy seasons under performed and resulted in moisture deficits (Funk et al. 2018; Palmer et al. 2023). This event has also been generally described as a traditional long-term drought (and lasted beyond the year examined in this case study), however elements of rapid intensification, largely in Ethiopia and southern Somalia, were detected. Rapid intensification occurred in both June (C23, N20, and P20) or May (L20) and in September and October (C23, N20, L20, and O21), corresponding to rapid drying that occurred with each rainy season (the fact rapid drying occurred twice suggests there may have been some drought recovery in between them, which is matched in the increase in P and SM anomalies between rainy seasons; Fig. 6.26, 6.27, and 6.28). Though the Ada boosted trees performed well with the P20 method in this case study, and the RNNs performed relatively well with the O21 method, both ML algorithms generally struggled with the other methods for this case as well.

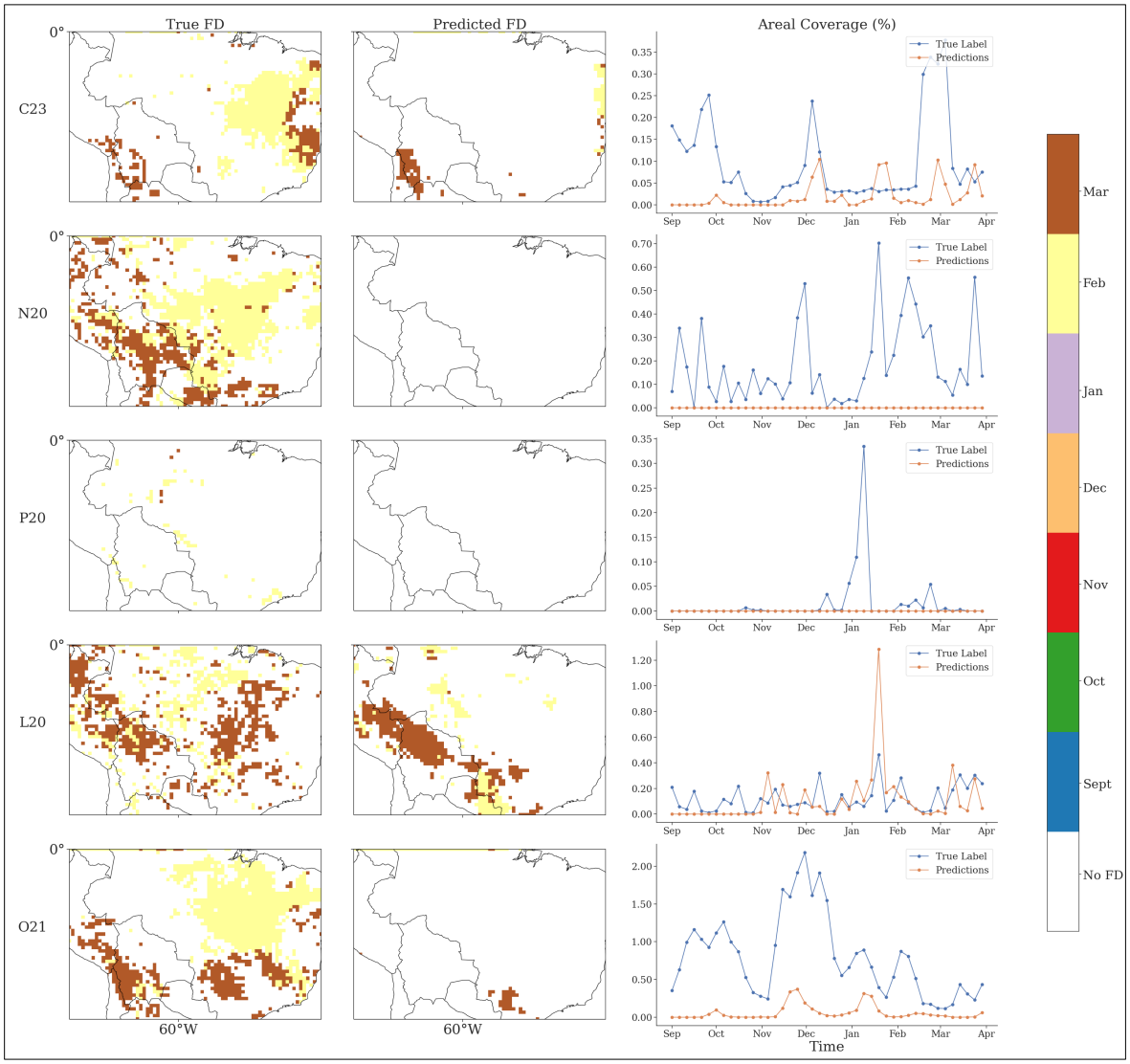


Figure 6.23: FD case study for 2015 – 2016 in the Amazon for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

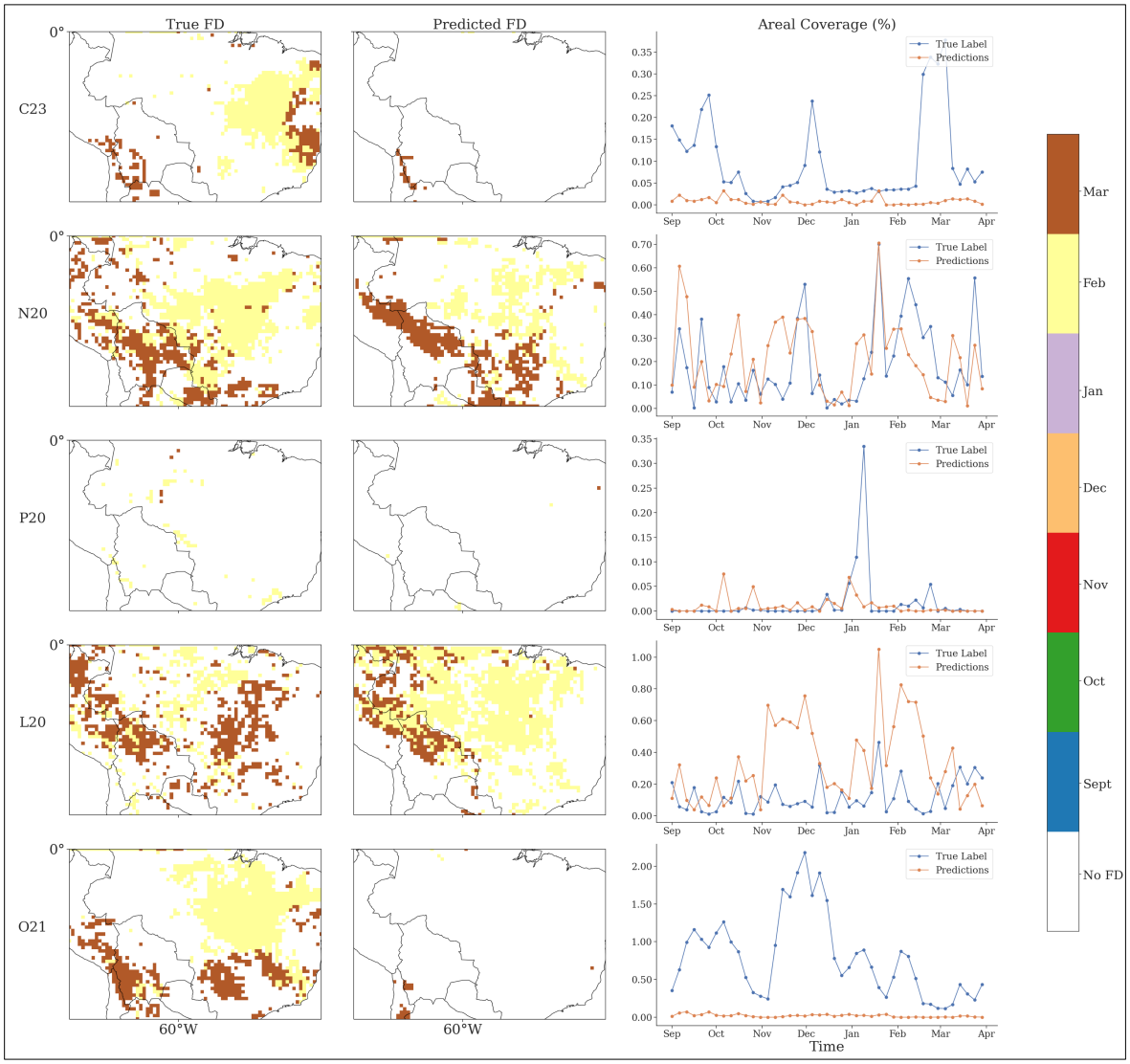


Figure 6.24: FD case study for 2015 – 2016 in the Amazon for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.



Figure 6.25: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2015 – 2016 case study over the Amazon.

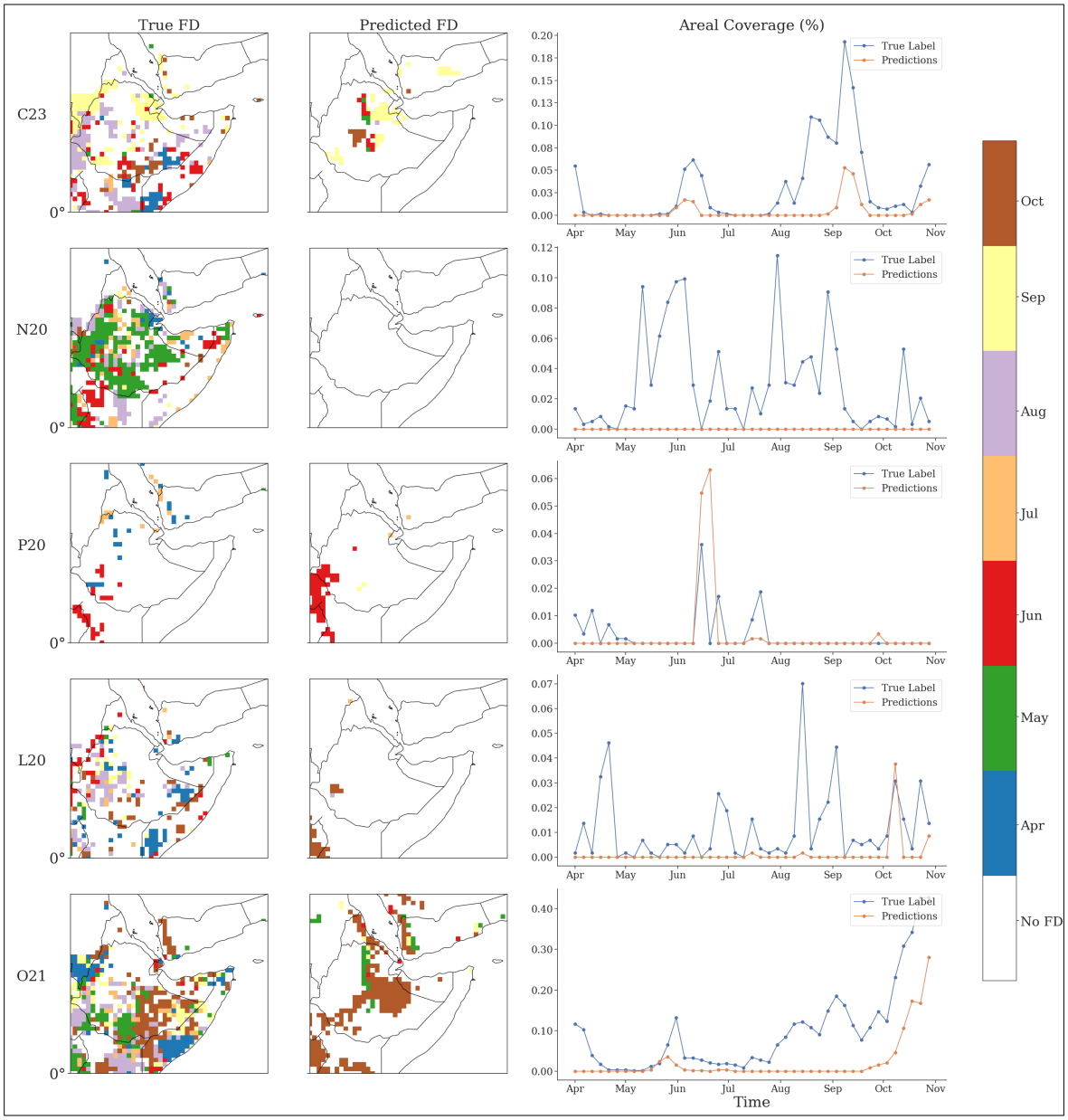


Figure 6.26: FD case study for 2016 in eastern Africa for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the RNNs.

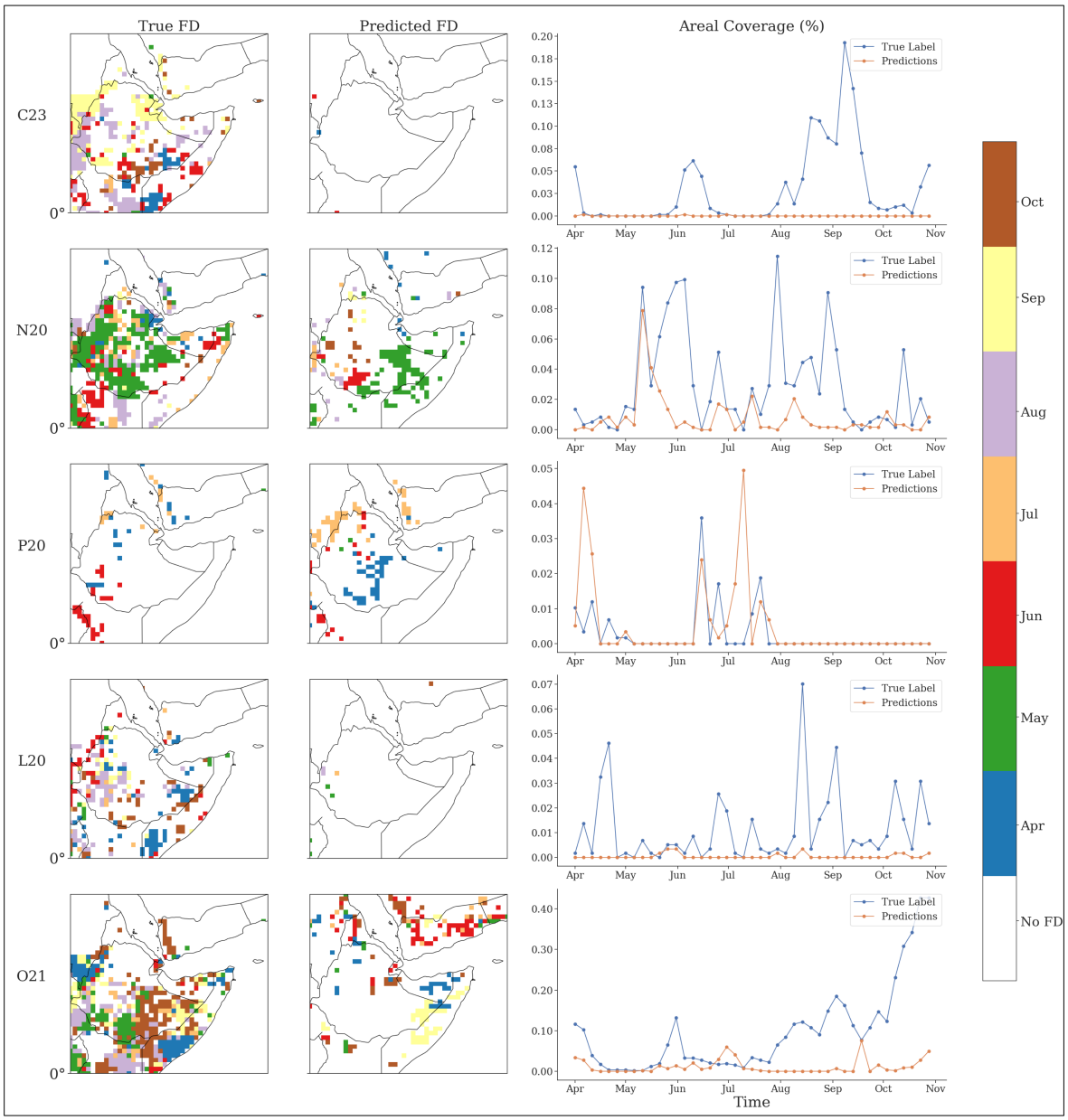


Figure 6.27: FD case study for 2016 in eastern Africa for each FD identification method. (left column) True labels, (center) predicted labels from the test dataset, (right) true and predicted FD coverage over the domain. FD predictions were made by the Ada boosted trees.



Figure 6.28: Spatially domain averaged standardized anomalies of (a) temperature, (b) precipitation, (c) potential evaporation, (d) evaporation, and (e) soil moisture for the 2016 case study over eastern Africa.

The impacts of this event was more notable (to the local population and society) and extreme than the previous case studies. This is because a vast majority of the area's popular works in and relies on local agriculture for their food supply (Palmer et al. 2023). Given much of the local agriculture was deeply impacted by the drought, the region experienced famine and deaths as a direct consequence of this drought event (Funk et al. 2018; Palmer et al. 2023). This particular event also showed that while there are ways people could adapt to drought, those adaptations are not evenly or equally distributed across the globe, and some areas are more vulnerable to drought and FD, and their impacts, than others. This also suggests there is a level of societal interaction to determine how impactful a drought and FD is (e.g., what the society does to adapt to the drought, such as limiting water use, any aid given or withheld from local governments or other from other societies), a feature that is not often discussed in the literature.

Chapter 7

Conclusions

This dissertation's contributions included an initial investigation of FD identification and prediction using six different ML algorithms; three standard ML algorithms (SVMs, RFs, and Ada boosted trees) and three deep learning algorithms (ANNs, RNNs, and convolutional U-nets). The results showed found that the most skillful of the standard ML algorithms (Ada boosted trees) showed promise in being able to represent FD events across multiple identification methods, while also remaining competitive compared to DL algorithms. The standard 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 the spatial coverage of FD events was often (though not always) well represented. However, feature importance showed that the standard ML algorithms were over relying on select variables (SM, P, and PET), and the ML algorithms may have learned climatologies and patterns in those variables as opposed to synthesizing the information to identify FD. Predictions in FD case studies could often be understood using these select variables. As a result, the standard 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 overemphasized hotspot regions, creating numerous false positives in those areas.

In comparison, the standard ANN structure had a surprisingly difficult time identifying FD. U-nets improved on this but were not able to outperform Ada boosted

trees. However the RNNs showed some improvement in identifying FD over the Ada boosted trees (except for the O21 method), though the skill score of the Ada boosted trees was still competitive. Specifically, the RNNs reduced the rate of over prediction of FD in hotspot regions (fewer false positives), while also giving 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 Ada boosted 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 on to two months later than verification, resulting in low skill scores. Interestingly, ML algorithms predicted the C23 and O21 methods better than others, suggesting they may have had an easier time recognizing patterns similar to or correlated with SM patterns and climatologies.

For the global performance, the RNNs and Ada boosted trees struggled to capture global FD patterns, preferring to focus on a few regions (certain latitudes) at the expense of others (e.g., ML models for the C23 method focused on getting FD identification in the Mid-Latitudes, but struggled to identify FD in the tropics, while they focused on the tropics for the L20 method at the expense of ignoring the rest of the globe, and the Ada boosted trees for the P20 and N20 methods ocused more on the High Latitudes). At the global level, both ML models struggled to learn temporal struggled to learn temporal and seasonal patterns of FD, and with individual FD events While these results suggested ML models for FD should be localized to specific regions for better results, they also suggested where the ML models might perform better, such as in the tropics for the L20 method, Mid-Latitudes for the C23 method, and so on.

In total, the contributions of this dissertation showed the ML algorithms have promise in representing FD events as a whole, with Ada boosted trees and RNNs (using LSTM layers) showing the most skill and the most skillful models stem from training on the regional scale. 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 patterns and timing of FD for some identification methods (i.e., P20 and O21). This is possibly due to the RNNs failing to properly synthesize surface variables and learn surface interactions (e.g., predicted RNN frequency climatology for the O21 method strongly resembled PET and T climatologies). 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 a FD indicator is predicted (e.g., in Foroumandi et al. 2024). Further, investigations into this problem using regression models is also recommended to give a different perspective on the problem and to investigate whether the ML performance improves in learning a continuous variables over a complex event that depends on multiple variables (like FD).

In addition, this dissertation also contributed a unique opportunity to examine the climatologies and seasonalities of multiple FD identification methods, applied to CONUS and to the globe, and how they represent different case study events. And in general, they showed similar patterns, though variations were still present between each identification method. For example, C23 focused more heavily on agricultural regions and moisture transition regions, where the evaporative stress from vegetation is higher compared to native vegetation. In contrast, the N20 method focused more on the western United States and earlier summer due to its emphasis on atmospheric water deficit (i.e., use of SPEI). The P20 method also focused more on mid to early summer FDs due to its emphasis on atmospheric demand (using EDDI to identify

FD). Similarly, the L20 and O21 methods showed climatological patterns similar to SM, identifying hotspots more closely to the C23 method, and identifying FD more in the late growing season (August – October in the NH, and September – March in the SH), when the SM is lower (Illston et al. 2004) and in early spring.

Globally, the different identification methods showed similar hotspot regions, such as in the eastern Amazon, western Europe, central United States, the southern Congo, northern Australia, and eastern and southern Asia. The strength of those hotspots, however, varied from one FD identification method to the next, and some FD identification methods may have additional hotspots. For example, most identification methods highlighted the Sahel region, however the P20 method does not due to its focus on PET, which would not increase as rapidly due to consistently high temperatures in the wet and dry seasons (however, transitions from wet to dry seasons would result in more detectable FD using P, SM, and ET). Some identification methods also showed other patterns, such as a heavy emphasis on the tropics for the L20 method, potentially due to how easily the leeched soils experience moisture decreases, while the N20 method generally has a higher frequency of FD and identified more FD on mountain slopes(showing an interestingly high rate of rapid precipitation decline/atmospheric demand increase). Despite differences in spatial patterns, the identification methods showed similar spatial trends across each spatial trends across each hotspot/FD vulnerable region identified in Christian et al. (2021). Each method showed similar seasonalities, and similar temporal trends (with the strongest increasing trend in FD coverage occurring in Asia Minor, the Iberian Peninsula, western Russia, and the trends were closer to neutral in other regions).

In general, each FD identification method gives its own perspective and flavor on FD, and examining and comparing their climatologies on how they handle different FD events can provide useful insights and is recommended for future studies. Given

the multiplicity 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 opinion, similar to what the USDM uses.

Another contribution from this dissertation was the opportunity to examined a wider variety of FD events in regions across the world that are infrequently discussed, and to examine differing levels and types of impacts. This was performed via global case studies, which focused on impactful traditional droughts, that also had elements of rapid intensification. These case studies showed that the impacts of drought and FD were not uniformly distributed across the globe. For example, the 2001 FD in India primarily impacted agriculture and food supply (Mahto and Mishra 2020), whereas the 2010 Russian FD was coupled with a strong heatwave (Christian et al. 2020). The 2015/16 drought in the Amazon resulted in high tree mortality that contributed to fires and had notable contributions towards climate change (Jiménez-Muñoz et al. 2016; Ribeiro et al. 2021). And the 2016/17 drought in eastern Africa showed mitigation strategies for drought are not evenly distributed across the globe, as that drought severely impacted east Africa's agriculture and food supply and directly resulted in famine and numerous deaths (Funk et al. 2018; Palmer et al. 2023). These also showed that there is a societal response that is important in determining how impactful droughts and FDs are. For example, many of the fires in the Amazon may not be naturally caused, and many of those may not be intentionally caused (Ribeiro et al. 2021), or in eastern Africa there were cases where the food relief may or may not be received by the populace (Palmer et al. 2023).

Future studies should examine some of the select ML algorithms (namely the Ada boosted trees and RNN structures) more deeply, but also more complex models such

as generative adversarial networks, to examine how the ML algorithms make their predictions and how to improve them. 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 and to focus on specific regions to improve ML performance. Separate from ML, future investigations into FD are recommended to examine global FD for other methods (not all identification methods were investigated here), as well as include multiple identification methods when investigating FD in general in order to gain multiple perspectives on FD and its developments, and thus deliver a more complete picture of FD events. Future works are also recommended to investigate different forms of drought and FD impact to better understand those impacts, their potential contributions towards climate change (especially for regions where droughts and FDs are expected to become more common), and the interaction between society and drought to better understand how drought impacts are modified by societal reactions and to implement methods of drought adaptations so that few to no locations experience famine as a result of drought and FD events.

Reference List

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
URL `https://www.tensorflow.org/`

AghaKouchak, A., B. Pan, O. Mazdiyasni, M. Sadegh, S. Jiwa, W. Zhang, C. A. Love, S. Madadgar, S. M. Papalexiou, S. J. Davis, K. Hsu, and S. Sorooshian, 2022: Status and prospects for drought forecasting: opportunities in artificial intelligence and hybrid physical–statistical forecasting. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **380**.
URL

Alencar, P. H. L. and E. N. Paton, 2022: How do we identify flash droughts? a case study in central european croplands. *Hydrology Research*.
URL

Alley, W. M., 1984: The palmer drought severity index: Limitations and assumptions. *Journal of Climate and Applied Meteorology*, **23**, 1100–1109.
URL

Anderson, M. C., C. Hain, J. Otkin, X. Zhan, K. Mo, M. Svoboda, B. Wardlow, and A. Pimstein, 2013: An intercomparison of drought indicators based on thermal remote sensing and nldas-2 simulations with u.s. drought monitor classifications. *Journal of Hydrometeorology*, **14**, 1035–1056.
URL `http://dx.doi.org/10.1175/JHM-D-12-0140.1`

Au, Q., J. Herbinger, C. Stachl, B. Bischl, and G. Casalicchio, 2022: Grouped feature importance and combined features effect plot. *Data Mining and Knowledge Discovery*, **36**, 1401–1450.
URL `http://dx.doi.org/10.1007/s10618-022-00840-5`

Babaeian, E., S. Paheding, N. Siddique, V. K. Devabhaktuni, and M. Tuller, 2022: Short- and mid-term forecasts of actual evapotranspiration with deep learning. *Journal of Hydrology*, **612**, 128078.
URL

Baek, S. H., J. E. Smerdon, R. Seager, A. P. Williams, and B. I. Cook, 2019: Pacific ocean forcing and atmospheric variability are the dominant causes of spatially

widespread droughts in the contiguous united states. *Journal of Geophysical Research: Atmospheres*, **124**, 2507–2524.
URL

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. *Environmental Research Letters*, **14**, 084025.
URL http://dx.doi.org/10.1088/1748-9326/ab2cc0

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. *Environmental Data Science*, **1**.
URL

Buitinck, L., G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, 2013: API design for machine learning software: experiences from the scikit-learn project. *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108–122.

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. *Weather and Forecasting*, **37**, 1509–1529.
URL http://dx.doi.org/10.1175/WAF-D-22-0070.1

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. *Weather and Forecasting*, **38**, 1271–1293.
URL http://dx.doi.org/10.1175/WAF-D-22-0187.1

Chen, L. G., J. Gottschalck, A. Hartman, D. Miskus, R. Tinker, and A. Artusa, 2019: Flash drought characteristics based on u.s. drought monitor. *Atmosphere*, **10**, 498.
URL http://dx.doi.org/10.3390/atmos10090498

Cho, K., B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, 2014: Learning phrase representations using rnn encoder-decoder for statistical machine translation.
URL https://arxiv.org/abs/1406.1078

Chollet, F. et al., 2015: Keras. https://keras.io.

Christian, J. I., J. B. Basara, 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. *Nature Communications*, **12**.
URL

Christian, J. I., J. B. Basara, E. D. Hunt, J. A. Otkin, and X. Xiao, 2020: Flash drought development and cascading impacts associated with the 2010 russian heatwave. *Environmental Research Letters*, **15**, 094078.
URL http://dx.doi.org/10.1088/1748-9326/ab9faf

Christian, J. I., J. B. Basara, L. E. Lowman, X. Xiao, D. Mesheske, and Y. Zhou, 2022: Flash drought identification from satellite-based land surface water index. *Remote Sensing Applications: Society and Environment*, **26**, 100770.
URL

Christian, J. I., J. B. Basara, J. A. Otkin, and E. D. Hunt, 2019a: Regional characteristics of flash droughts across the united states. *Environmental Research Communications*, **1**, 125004.
URL http://dx.doi.org/10.1088/2515-7620/ab50ca

Christian, J. I., J. B. Basara, J. A. Otkin, E. D. Hunt, 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. *Journal of Hydrometeorology*, **20**, 833–846.
URL http://dx.doi.org/10.1175/JHM-D-18-0198.1

Christian, J. I., M. Hobbins, A. Hoell, J. A. Otkin, T. W. Ford, A. E. Cravens, K. A. Powlen, H. Wang, and V. Mishra, 2024: Flash drought: A state of the science review. *WIREs Water*, **11**.
URL http://dx.doi.org/10.1002/wat2.1714

Christian, J. I., E. R. Martin, J. B. Basara, J. C. Furtado, J. A. Otkin, L. E. L. Lowman, E. D. Hunt, V. Mishra, and X. Xiao, 2023: Global projections of flash drought show increased risk in a warming climate. *Communications Earth amp; Environment*, **4**.
URL http://dx.doi.org/10.1038/s43247-023-00826-1

Cook, B. I., E. R. Cook, J. E. Smerdon, R. Seager, A. P. Williams, S. Coats, D. W. Stahle, and J. V. Díaz, 2016: North american megadroughts in the common era: reconstructions and simulations. *WIREs Climate Change*, **7**, 411–432.
URL

Cook, B. I., J. S. Mankin, and K. J. Anchukaitis, 2018: Climate change and drought: From past to future. *Current Climate Change Reports*, **4**, 164–179.
URL

Dikshit, A. and B. Pradhan, 2021: Interpretable and explainable AI (XAI) model for spatial drought prediction. *Science of The Total Environment*, **801**, 149797.
URL

Dikshit, A., B. Pradhan, A. Huete, and H.-J. Park, 2022a: Spatial based drought assessment: Where are we heading? a review on the current status and future.

Science of The Total Environment, **844**, 157239.
URL

Dikshit, A., B. Pradhan, and M. Santosh, 2022b: Artificial neural networks in drought prediction in the 21st century–a scientometric analysis. *Applied Soft Computing*, **114**, 108080.
URL

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. *Agricultural and Forest Meteorology*, **330**, 109288.
URL http://dx.doi.org/10.1016/j.agrformet.2022.109288

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. *Journal of Geophysical Research: Atmospheres*, **108**.
URL http://dx.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. *Artificial Intelligence for the Earth Systems*, **3**.
URL http://dx.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. *Agricultural and Forest Meteorology*, **247**, 414–423.
URL

Ford, T. W., 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. *Geophysical Research Letters*, **42**, 9790–9798.
URL http://dx.doi.org/10.1002/2015GL066600

Foroumandi, E., K. Gavahi, and H. Moradkhani, 2024: Generative adversarial network for real-time flash drought monitoring: A deep learning study. *Water Resources Research*, **60**.
URL http://dx.doi.org/10.1029/2023WR035600

Funk, C., L. Harrison, S. Shukla, C. Pomposi, G. Galu, D. Korecha, G. Husak, T. Magadzire, F. Davenport, C. Hillbruner, G. Eilerts, B. Zaitchik, and J. Verdin, 2018: Examining the role of unusually warm indo-pacific sea-surface temperatures in recent african droughts. *Quarterly Journal of the Royal Meteorological Society*, **144**, 360–383.
URL http://dx.doi.org/10.1002/qj.3266

Ganguli, P. and M. J. Reddy, 2013: Ensemble prediction of regional droughts using climate inputs and the SVM-copula approach. *Hydrological Processes*, **28**, 4989–5009.
URL

Gavahi, K., P. Abbaszadeh, and H. Moradkhani, 2022: How does precipitation data influence the land surface data assimilation for drought monitoring? *Science of The Total Environment*, **831**, 154916.
URL

Guttman, N. B., 1999: Accepting the standardized precipitation index: A calculation algorithm. *Journal of the American Water Resources Association*, **35**, 311–322.

Hao, Z., V. P. Singh, and Y. Xia, 2018: Seasonal drought prediction: Advances, challenges, and future prospects. *Reviews of Geophysics*, **56**, 108–141.
URL

Hastie, T., J. Friedmand, and R. Tibshirani, 2009: *The Elements of Statistical Learning*. Springer Series in Statistics, Springer New York, NY, 2 edition.

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. *Environmental Research Letters*, **14**, 074019.
URL `http://dx.doi.org/10.1088/1748-9326/ab22c3`

Hersbach, B. B. B. P. B. G. H. A. M. S. J. N. J. P. C. R. R. R. I. S. D. S. A. S. C. D. D. T. J.-N., H., 2023: Era5 hourly data on single levels from 1940 to present.
URL `https://cds.climate.copernicus.eu/doi/10.24381/cds.adbb2d47`

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. *Journal of Hydrometeorology*, **17**, 1745–1761.
URL `http://dx.doi.org/10.1175/JHM-D-15-0121.1`

Hochreiter, S. and J. Schmidhuber, 1997: Long short-term memory. *Neural Computation*, **9**, 1735–1780.
URL

Hoffmann, D., A. J. E. Gallant, and M. Hobbins, 2021: Flash drought in cmip5 models. *Journal of Hydrometeorology*, 1439 – 1454.
URL

Hsieh, W. W., 2022: Evolution of machine learning in environmental science—a perspective. *Environmental Data Science*, **1**.
URL

Hunt, E. D., K. G. Hubbard, D. A. Wilhite, T. J. Arkebauer, and A. L. Dutcher, 2009: The development and evaluation of a soil moisture index. *International Journal of Climatology*, **29**, 747–759.
URL http://dx.doi.org/10.1002/joc.1749

Iglesias, V., W. R. Travis, and J. K. Balch, 2022: Recent droughts in the united states are among the fastest-developing of the last seven decades. *Weather and Climate Extremes*, **37**, 100491.
URL

Illston, B. G., J. B. Basara, and K. C. Crawford, 2004: Seasonal to interannual variations of soil moisture measured in oklahoma. *International Journal of Climatology*, **24**, 1883–1896.
URL http://dx.doi.org/10.1002/joc.1077

Jiménez-Muñoz, J. C., C. Mattar, J. Barichivich, A. Santamaría-Artigas, K. Takahashi, Y. Malhi, J. A. Sobrino, and G. v. d. Schrier, 2016: Record-breaking warming and extreme drought in the amazon rainforest during the course of el niño 2015–2016. *Scientific Reports*, **6**.
URL http://dx.doi.org/10.1038/srep33130

Kim, D., W. 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 Resources Research*, **55**, 6244–6267.
URL http://dx.doi.org/10.1029/2019WR024991

Kim, D. and J. Rhee, 2016: A drought index based on actual evapotranspiration from the bouchet hypothesis. *Geophysical Research Letters*, **43**, 10,277–10,285.
URL http://dx.doi.org/10.1002/2016GL070302

Kim, W. M. and C. C. Raible, 2021: Dynamics of the mediterranean droughts from 850 to 2099 CE in the community earth system model. *Climate of the Past*, **17**, 887–911.
URL

Kim, Y., S. Kim, H. Jeong, and H. An, 2022: Quantitatively defining megadrought based on drought events in central chile. *Geomatics, Natural Hazards and Risk*, **13**, 975–992.
URL

Krishnamurthy R, P. K., J. B. Fisher, R. J. Choularton, and P. M. Kareiva, 2022: Anticipating drought-related food security changes. *Nature Sustainability*, **5**, 956–964.
URL

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. *Nature Reviews Earth; Environment*, **3**, 872–889.
URL

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.
URL

Li, W., M. Migliavacca, M. Forkel, J. M. C. Denissen, M. Reichstein, H. Yang, G. Duveiller, U. Weber, and R. Orth, 2022: Widespread increasing vegetation sensitivity to soil moisture. *Nature Communications*, **13**.
URL

Lisonbee, J., M. Woloszyn, and M. Skumanich, 2021: Making sense of flash drought: definitions, indicators, and where we go from here. *Journal of Applied and Service Climatology*, **2021**, 1–19.
URL http://dx.doi.org/10.46275/JOASC.2021.02.001

Liu, J., F. Rahmani, K. Lawson, and C. Shen, 2022: A multiscale deep learning model for soil moisture integrating satellite and in situ data. *Geophysical Research Letters*, **49**.
URL

Liu, X., X. Zhu, Q. Zhang, T. Yang, Y. Pan, and P. Sun, 2020a: A remote sensing and artificial neural network-based integrated agricultural drought index: Index development and applications. *CATENA*, **186**, 104394.
URL

Liu, Y., E. Racah, Prabhat, J. Correa, A. Khosrowshahi, D. Lavers, K. Kunkel, M. Wehner, and W. Collins, 2016: Application of deep convolutional neural networks for detecting extreme weather in climate datasets.
URL https://arxiv.org/abs/1605.01156

Liu, Y., Y. Zhu, L. Zhang, L. Ren, F. Yuan, X. Yang, and S. Jiang, 2020b: Flash droughts characterization over china: From a perspective of the rapid intensification rate. *Science of The Total Environment*, **704**, 135373.
URL

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. *Journal of Hydrometeorology*, **24**, 1395–1415.
URL http://dx.doi.org/10.1175/JHM-D-22-0158.1

Lundberg, S. M. and S.-I. Lee, 2017: A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, eds., Curran Associates, Inc., volume 30.

Mahto, S. S. and V. Mishra, 2020: Dominance of summer monsoon flash droughts in india. *Environmental Research Letters*, **15**, 104061.
URL

—, 2023: Increasing risk of simultaneous occurrence of flash drought in major global croplands. *Environmental Research Letters*, **18**, 044044.
URL http://dx.doi.org/10.1088/1748-9326/acc8ed

Mamalakis, A., I. Ebert-Uphoff, and E. A. Barnes, 2022: Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, **1**.
URL

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. *Environmental Data Science*, **1**.
URL

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. *Journal of Hydrometeorology*, **17**, 1763–1779.
URL http://dx.doi.org/10.1175/JHM-D-15-0122.1

McGovern, A., I. Ebert-Uphoff, D. J. Gagne, and A. Bostrom, 2022: Why we need to focus on developing ethical, responsible, and trustworthy artificial intelligence approaches for environmental science. *Environmental Data Science*, **1**.
URL

McGovern, A., R. Lagerquist, D. John Gagne, 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. *Bulletin of the American Meteorological Society*, **100**, 2175–2199.
URL http://dx.doi.org/10.1175/BAMS-D-18-0195.1

Mehr, A. D., A. R. Ghiasi, Z. M. Yaseen, A. U. Sorman, and L. Abualigah, 2022: A novel intelligent deep learning predictive model for meteorological drought forecasting. *Journal of Ambient Intelligence and Humanized Computing*.
URL

Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jović, J. Woollen, E. Rogers, E. H. Berbery, and et al., 2006: North american regional reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360.
URL http://dx.doi.org/10.1175/BAMS-87-3-343

Mohamadi, S., S. S. Sammen, F. Panahi, M. Ehteram, O. Kisi, A. Mosavi, A. N. Ahmed, A. El-Shafie, and N. Al-Ansari, 2020: Zoning map for drought prediction using integrated machine learning models with a nomadic people optimization algorithm. *Natural Hazards*, **104**, 537–579.
URL

Noguera, I., F. Domínguez-Castro, and S. M. Vicente-Serrano, 2020: Characteristics and trends of flash droughts in spain, 1961–2018. *Annals of the New York Academy of Sciences*, **1472**, 155–172.
URL http://dx.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. *Hydrology and Earth System Sciences*, **25**, 565–581.
URL

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. *Journal of Hydrometeorology*, **14**, 1057–1074.
URL http://dx.doi.org/10.1175/JHM-D-12-0144.1

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. *Journal of Hydrometeorology*, **15**, 938–956.
URL http://dx.doi.org/10.1175/JHM-D-13-0110.1

Otkin, J. A., M. C. Anderson, C. Hain, M. Svoboda, D. Johnson, R. Mueller, T. Tadesse, B. Wardlow, and J. Brown, 2016: Assessing the evolution of soil moisture and vegetation conditions during the 2012 united states flash drought. *Agricultural and Forest Meteorology*, **218-219**, 230–242.
URL http://dx.doi.org/10.1016/j.agrformet.2015.12.065

Otkin, J. A., 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. *Bulletin of the American Meteorological Society*, **99**, 911–919.
URL http://dx.doi.org/10.1175/BAMS-D-17-0149.1

Otkin, J. A., Y. Zhong, E. D. Hunt, J. I. Christian, J. B. Basara, H. Nguyen, M. C. Wheeler, T. W. Ford, A. Hoell, M. Svoboda, and et al., 2021: Development of a flash drought intensity index. *Atmosphere*, **12**, 741.
URL http://dx.doi.org/10.3390/atmos12060741

Palmer, P. I., C. M. Wainwright, B. Dong, R. I. Maidment, K. G. Wheeler, N. Gedney, J. E. Hickman, N. Madani, S. S. Folwell, G. Abdo, R. P. Allan, E. C. L.

Black, L. Feng, M. Gudoshava, K. Haines, C. Huntingford, M. Kilavi, M. F. Lunt, A. Shaaban, and A. G. Turner, 2023: Drivers and impacts of eastern african rainfall variability. *Nature Reviews Earth amp; Environment*, **4**, 254–270.
URL http://dx.doi.org/10.1038/s43017-023-00397-x

Palmer, W., 1965: *Meteorological Drought*. Number v. 30 in Meteorological Drought, U.S. Department of Commerce, Weather Bureau.
URL 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. *Agricultural and Forest Meteorology*, **216**, 157–169.
URL

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and Édouard Duchesnay, 2011: Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, **12**, 2825–2830.
URL http://jmlr.org/papers/v12/pedregosa11a.html

Pendergrass, A. G., G. A. Meehl, R. Pulwarty, M. Hobbins, A. Hoell, A. AghaKouchak, C. J. W. Bonfils, A. J. E. Gallant, M. Hoerling, D. Hoffmann, and et al., 2020: Flash droughts present a new challenge for subseasonal-to-seasonal prediction. *Nature Climate Change*, **10**, 191–199.
URL http://dx.doi.org/10.1038/s41558-020-0709-0

Prodhan, F. A., J. Zhang, F. Yao, L. Shi, T. P. P. Sharma, D. Zhang, D. Cao, M. Zheng, N. Ahmed, and H. P. Mohana, 2021: Deep learning for monitoring agricultural drought in south asia using remote sensing data. *Remote Sensing*, **13**, 1715.
URL

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. *Nature Communications*, **13**.
URL http://dx.doi.org/10.1038/s41467-022-28752-4

Rahmati, O., F. Falah, K. S. Dayal, R. C. Deo, F. Mohammadi, T. Biggs, D. D. Moghaddam, S. A. Naghibi, and D. T. Bui, 2020: Machine learning approaches for spatial modeling of agricultural droughts in the south-east region of queensland australia. *Science of The Total Environment*, **699**, 134230.
URL

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. *Agricultural and Forest Meteorology*, **237-238**, 105–122.
URL

Ribeiro, G. G., L. O. Anderson, N. J. C. Barretos, R. Abreu, L. Alves, B. Dong, F. C. Lott, and S. F. B. Tett, 2021: Attributing the 2015/2016 amazon basin drought to anthropogenic influence. *Climate Resilience and Sustainability*, **1**.
URL http://dx.doi.org/10.1002/cli2.25

Runde, I., Z. Zobel, and C. Schwalm, 2022: Human and natural resource exposure to extreme drought at 1.0 °c–4.0 °c warming levels. *Environmental Research Letters*, **17**, 064005.
URL

Sak, H., A. Senior, and F. Beaufays, 2014: Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition.
URL https://arxiv.org/abs/1402.1128

Shah, J., V. Hari, O. Rakovec, Y. Markonis, L. Samaniego, V. Mishra, M. Hanel, C. Hinz, and R. Kumar, 2022: Increasing footprint of climate warming on flash droughts occurrence in europe. *Environmental Research Letters*, **17**, 064017.
URL

Singh, S., M. Kaushik, A. Gupta, and A. K. Malviya, 2019: Weather forecasting using machine learning techniques. *Proceedings of 2nd International Conference on Advanced Computing and Software Engineering*.

Stager, J. C., B. Wiltse, B. F. Cumming, T. C. Messner, J. Robtoy, and S. Cushing, 2021: Hydroclimatic and cultural instability in northeastern north america during the last millennium. *PLOS ONE*, **16**, e0248060.
URL

Stevenson, S., J. T. Overpeck, J. Fasullo, S. Coats, L. Parsons, B. Otto-Bliesner, T. Ault, G. Loope, and J. Cole, 2018: Climate variability, volcanic forcing, and last millennium hydroclimate extremes. *Journal of Climate*, **31**, 4309–4327.
URL

Svoboda, M., D. LeComte, M. Hayes, R. Heim, K. Gleason, J. Angel, B. Rippey, R. Tinker, M. Palecki, D. Stooksbury, and et al., 2002: The drought monitor. *Bulletin of the American Meteorological Society*, **83**, 1181–1190.
URL http://dx.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*.
URL

Vicente-Serrano, S. M., S. Beguería, and J. I. López-Moreno, 2010: A multiscalar drought index sensitive to global warming: The standardized precipitation evapotranspiration index. *Journal of Climate*, **23**, 1696–1718.
URL

Vicente-Serrano, S. M., D. G. Miralles, F. Domínguez-Castro, C. Azorin-Molina, A. El Kenawy, T. R. McVicar, M. Tomás-Burguera, S. Beguería, M. Maneta, and M. Peña-Gallardo, 2018: Global assessment of the standardized evapotranspiration deficit index (sedi) for drought analysis and monitoring. *Journal of Climate*, **31**, 5371–5393.
URL http://dx.doi.org/10.1175/JCLI-D-17-0775.1

Watson, P. A. G., 2022: Machine learning applications for weather and climate need greater focus on extremes. *Environmental Research Letters*, **17**, 111004.
URL https://dx.doi.org/10.1088/1748-9326/ac9d4e

Wells, N., S. Goddard, and M. J. Hayes, 2004: A self-calibrating palmer drought severity index. *Journal of Climate*, **17**, 2335–2351.
URL

Woodcock, F., 1976: The evaluation of yes/no forecasts for scientific and administrative purposes. *Monthly Weather Review*, **104**, 1209–1214.
URL http://dx.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. *Nature Communications*, **10**.
URL

Yuan, X., 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.
URL http://dx.doi.org/10.1126/science.abn6301

Zaremba, W., I. Sutskever, and O. Vinyals, 2014: Recurrent neural network regularization.
URL https://arxiv.org/abs/1409.2329

Zargar, A., R. Sadiq, B. Naser, and F. I. Khan, 2011: A review of drought indices. *Environmental Reviews*, **19**, 333–349.
URL http://dx.doi.org/10.1139/a11-013

Zhang, L., Y. Liu, L. Ren, A. J. Teuling, Y. Zhu, L. Wei, L. Zhang, S. Jiang, X. Yang, X. Fang, and H. Yin, 2022: Analysis of flash droughts in china using machine learning. *Hydrology and Earth System Sciences*, **26**, 3241–3261.
URL

Zomer, R. J., J. Xu, and A. Trabucco, 2022: Version 3 of the global aridity index and potential evapotranspiration database. *Scientific Data*, **9**.
URL http://dx.doi.org/10.1038/s41597-022-01493-1