UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

WILDFIRE OCCURRENCE PREDICTION FOR CONUS WITH THE UNET3+ DEEP LEARNING MODEL

A DISSERTATION
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
DOCTOR OF PHILOSOPHY

By

BETHANY EARNEST
Norman, Oklahoma
2024

WILDFIRE OCCURRENCE PREDICTION FOR CONUS WITH THE UNET3+ DEEP LEARNING MODEL

A DISSERTATION APPROVED FOR THE
SCHOOL OF COMPUTER SCIENCE

BY THE COMMITTEE CONSISTING OF

Dr. Amy McGovern, Chair

Dr. Christopher Karstens, Co-Chair

Dr. Andrew H. Fagg

Dr. Dimitrios Diochnos

Dr. Jennifer Koch

Acknowledgements

I am deeply honored to have had the opportunity to learn from my co-advisors, Dr. Amy McGovern and Dr. Chris Karstens. Their combined wisdom has significantly broadened my understanding, guided my research, and upheld me to the highest academic standards. This dissertation would not have been possible without their unwavering support and invaluable guidance.

I extend my heartfelt thanks to my committee members, Dr. Andrew Fagg, Dr. Dimitrios Diochnos, and Dr. Jennifer Koch, for their diligent oversight and guidance. To my lab mates in the IDEA Lab, your feedback, ideas, and assistance with troubleshooting have been invaluable. I look forward to hearing about your future successes.

A special thank you to Dr. Randy Chase and Dr. Monique Shotande. During their respective tenures as postdoctoral researchers, they each played a crucial role in evolving this research, refining my code, and shaping my thinking. This work would not be what it is without their contributions.

I also wish to express my gratitude to Dr. Israel Jirak, Dr. Nick Nauslar, Dr. Matthew Elliott, and Dr. Evan Bentley for their respective roles in enhancing my understanding of the wildfire domain.

I want to express my deepest gratitude to my family: my husband, Brian, my daughter, Olivia, and my son, Elliott. You are the heart and soul behind this achievement. Your unwavering love, boundless patience, and steadfast support have been my guiding light, and I couldn't have accomplished this without you. Thank you for being my inspiration and my strength.

This dissertation was prepared with funding provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA21OAR4320204, U.S. Department of Commerce. This material is based

upon work supported by the National Science Foundation under Grant No. ICER-2019758. The statements, findings, conclusions, and recommendations are those of the author and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce.

Table of Contents

List Of Tables

List Of Figures

Abstract

Wildfire represents a risk to life and property in many areas of the United States and is of growing concern to insurance companies, legislative bodies, and the public. Accurate wildfire forecasting could allow for earlier deployment of firefighting resources resulting in less property damage and less loss of life. Accurate wildfire forecasting could lower the cost of suppressing a wildfire in progress and allow for longer lead times in communicating with the public. The purpose of this research is to explore the efficacy of applying deep-learning to the task of predicting wildfire occurrence for the contiguous United States (CONUS) in the 0-to-10-day range.

To address this challenge, I employ binary classification semantic segmentation using the UNet3+ model combined with a neighborhood loss function, Fractions Skill Score (FSS). The UNet3+ model, originally introduced for use in medical imaging, combines full scale skip connections with an encoder-decoder architecture, which allows it to capture both fine-grain detail and coarse-grain semantics simultaneously. With the neighborhood loss function, FSS, I am able to quantify model success by predictions made both in and around the location of the original fire label. I utilize two datasets as inputs to my model, first, gridMET and, second, NOAA's Global Ensemble Forecast System (GEFS), both are commonly used by fire weather forecasters. For both approaches, my fire occurrence labels are sourced from the U.S. Department of Agriculture's Fire Program Analysis fire-occurrence database (FPA-FOD), which contains spatial wildfire occurrence data for CONUS from 1992 to 2020, updated in 2022, and combines data sourced from the reporting systems of federal, state, and local organizations.

The unique contribution of this dissertation is to advance the research at the intersection of deep learning and fire occurrence prediction. To this end, I detail two proof of concept models using familiar datasets and subject matter expert informed approaches with the goal of developing a deep learning method that can outperform

current operational techniques used by forecasters for the task of fire occurrence prediction.

My first approach, described in Chapters 5 and 6, sources model inputs from gridMET, a daily, CONUS-wide, high-spatial resolution dataset of surface meteorology variables including fire danger variables. From gridMET, I source observed fire danger variables, observed weather variables, and a topography variable. I compare two models, the "All Fires" model that uses all fire occurrence instances in the label images and the "Large Lightning" model that only uses instances of fire occurrence that represent large, natural-caused fires. For the experiment, the "All Fires" model produces higher max CSI values than the "Large Lightning" model when compared on general performance and on only large lightning fire performance. The "All Fires" model also produces higher probability of wildfire when compared to both the SPC Probability of Wildfire Climatology and the "Large Lightning" model for three case studies representing the largest large lightning fires from 2018, 2019, and 2020.

In my second approach, described in Chapters 7 and 8, I source model inputs from the GEFS Reforecast dataset, a daily, 5-member ensemble numerical weather prediction model, used to produce retrospective gridded meteorological forecasts for CONUS. From GEFS, I source observed and forecast weather variables. I compare two models, the "Multi Label" model, that trains using data augmentation, and the "Pixel Label" model, that trains without using data augmentation. Both models build on the success of the previous approach by using all fire occurrences in the label images. I contextualize model performance using Max CSI and reliability calculated for three neighborhoods, 40km, 80km, 120km. For the experiment, the "Multi Label" model produces reliable results when measured at 80km and the "Pixel Label" model produces reliable results when measured at 40km. The "Multi Label" model and the "Pixel Label" model produce comparable Max CSI values for all neighborhoods for all days. Both models produce higher probability of wildfire values when compared to the SPC Probability of Wildfire Climatology on three case studies: the Camp fire, the Carr fire, and the Woolsey fire.

Chapter 1

Introduction

General Motivation Wildfires are dangerous and expensive, and are becoming larger and more frequent in the United States. According to the Climate Science Special Report (CSSR), created as part of the Forth National Climate Assessment (NCA4), "the incidence of large forest fires in the western United States and Alaska has increased since the early 1980s... and is projected to further increase in those regions as the climate warms..." (Wehner et al. 2024). According to the Federal Government's Climate Financial Risk report, to which researchers from the United States Forest Service (USFS) contributed, "lands in the National Forest System [will] experience a near doubling of the area burned by mid-century (2041-2059)" and "[a] middle-of-the-road estimate [for annual fire suppression expenditures] is a 42% increase in costs by 2050, to $3.9 billion" (Management and Budget 2024). These numbers do not account for wildfire related damages such as smoke related illness, loss of life, property, and natural resources.

Fire managers are responsible for mitigating the impacts of fire through wildfire prevention, wildfire fighting, and land management. Fire managers rely on accurate wildfire predictions to respond as quickly as possible to an active wildfire, to position crews to better effect during a wildfire season, and to communicate in a timely manner with the public prior to a wildfire (to encourage fire safe behaviour). There are currently three main methods for fire managers to receive information about fire danger or probability for the Contiguous United States (CONUS). These are the

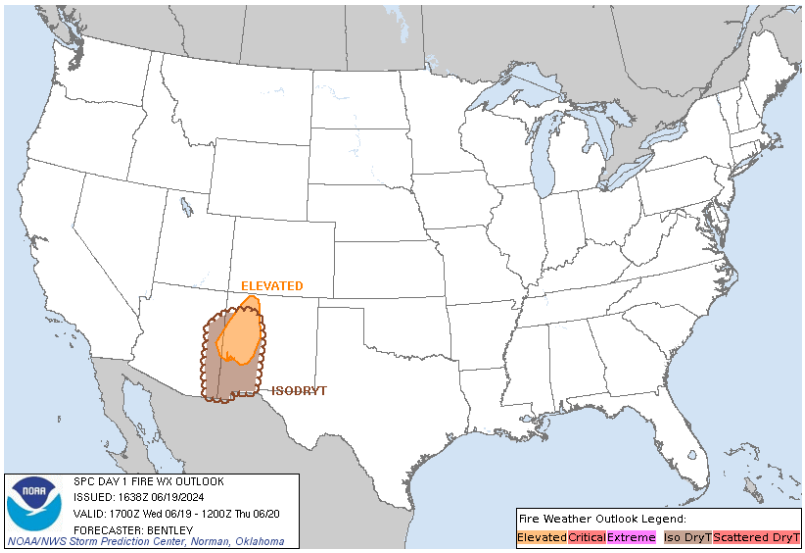


Figure 1.1: Day 1 Fire Weather Outlook for 6/19/2024

SPC Fire Weather Outlook, the SPC Probability of Wildfire Climatology, and the National Fire Danger Rating System (NFDRS).

SPC Fire Weather Outlook Fire weather forecasting describes predicting an environment in which the atmospheric elements are conducive to large-fire spread. The SPC Fire Weather Outlook, produced by human forecasters at the National Oceanic and Atmospheric Administration (NOAA) National Weather Service (NWS) Storm Prediction Center (SPC), provides fire weather forecasts for CONUS. These outlooks are represented as a combination of polygons and written text, overlaying an image of CONUS, called operational products. The polygons are created by the fire weather forecasters to delineate regions of varying fire weather potential based their assessment of weather and fuel conditions. Figure 1.1 depicts an example of a Day 1 Fire Weather Outlook.

There are three products offered as part of the SPC Fire Weather Outlook, the Day 1 Fire Weather Outlook, the Day 2 Fire Weather Outlook, and the Day 3-8

Category	*Wind*	*Relative Humidity (RH)*	*Temperature (season depending)*	*Fuels*	*Duration*
Elevated	Sustained winds >= 15 mph (10 mph Florida)	Minimum RH within 5% of regional thresholds	Temperatures > 45-55° F	Dry	Duration >= 3 hours – or critical conditions expected for a duration < 3 hours
Critical	Sustained winds >= 20 mph (15 mph Florida)	Minimum RH at or below regional thresholds	Temperatures > 50-60° F	Dry	Duration >= 3 hours
Extremely Critical	Sustained winds >= 30 mph (25 mph Florida)	RH at or below 1/3 lower than regional thresholds	Temperatures >= 60-70° F	Very Dry	Duration >= 3 hours

Table 1.1: SPC Fire Weather Outlook Criteria (SPC 2024a)

Fire Weather Outlook. Forecasts for Isolated Dry Thunderstorms and Scatterd Dry Thunderstorms are included in the Day 1 and Day 2 Fire Weather Outlooks, which you can see depicted in Figure 1.1. However, as they do not represent an analog my work, I do not describe them here.

For the Fire Weather Outlooks, polygons drawn on the CONUS image by forecasters indicate fire weather conditions that fall into one of three categories: Elevated, Critical, or Extreme as described in Table 1.1. The Day 3-8 Fire Weather Outlook contains probabilities of meeting or exceeding Critical criteria (for wind/RH conditions). "Predictability too low" is used when model uncertainty is too high. "Potential too low" is used when confidence is low for critical fire weather conditions (SPC 2024a,b).

SPC Probability of Wildfire Climatology The SPC Probability of Wildfire Climatology uses wildfire occurrence data from 1992 to 2015 to describe the probability of wildfire for CONUS using a climatological approach. The SPC Probability of Wildfire Climatology accomplishes this by taking into account the three days prior and the three days post a given date to determine likelihood of a wildfire within 80km

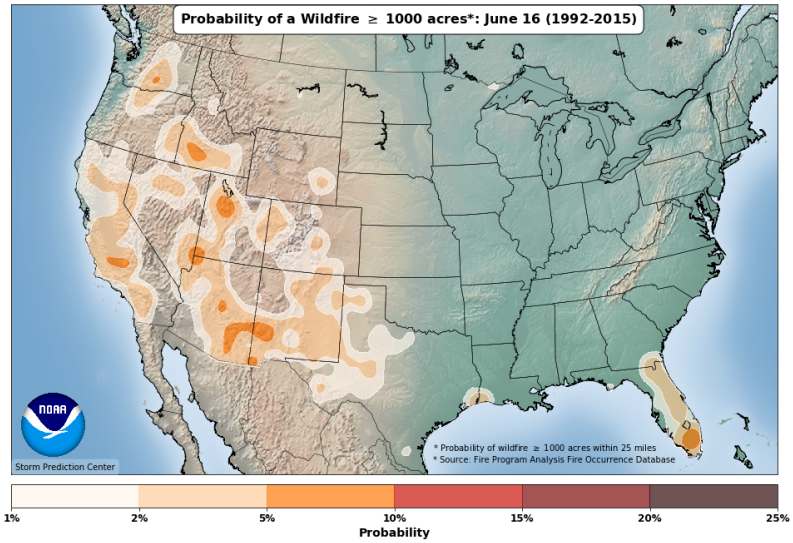


Figure 1.2: SPC Probability of Wildfire Climatology: >= 1000 Acres for 6/16/2024

of a location and does so for wildfires greater than 100 acres, 300 acres, 1000 acres, and 5000 acres. The SPC Probability of Wildfire Climatology relies on the Fire Program Analysis Fire Occurrence Database (FPA-FOD) for climatological inputs and produces a map of CONUS with areas of significance outlined to represent gradients of probabilities of wildfire. Different maps are presented for different final fire sizes. Figure 1.2 depicts an example of the SPC Probability of Wildfire Climatology. The SPC Probability of Wildfire Climatology is created by NOAA/NWS SPC (Nauslar 2020b).

National Fire Danger Rating System (NFDRS) NFDR describes the likelihood for "a fire to ignite, spread and require suppression action" (USFS 2024) and is created by the United States Department of Agriculture (USDA) and the United States Forest Service (USFS). NFDRS includes products for the current and next day and classifies fire danger into low, moderate, high, very high, and extreme. NFDRS uses fuel, topography, weather and fire risk inputs. Variables created by NFDRS,

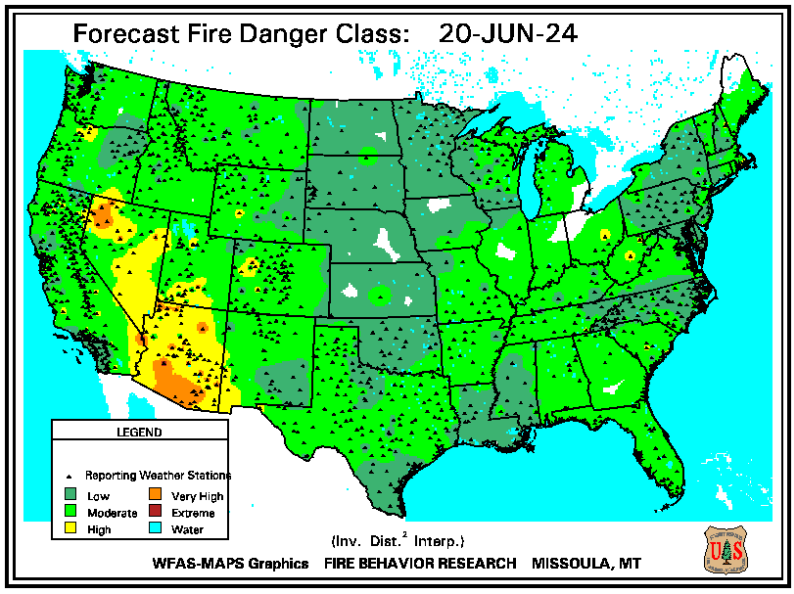


Figure 1.3: NFDRS Fire Danger Forecast for 6/20/2024

such as energy release component (ERC), burning index (BI), and dead fuel moisture (FM), seek to combine weather and fuel information into fire specific metrics. The NFDRS outputs are presented on an image of CONUS with areas colored to represent the fire danger classifications. Figure 1.3 depicts an example of the NFDRS Fire Danger Forecast as made available through the Wildland Fire Assessment System (WFAS).

Current Tooling Limitations Each of these three approaches to offering fire probability guidance to fire managers has limitations. The SPC Fire Weather Outlooks do not provide explicit predictions of fire occurrence, and instead highlight broad regions where the environment is determined to be conducive to large-fire occurrence and spread. Further, these forecasts rely on fuel assessments created from NFDRS variables, such as energy release component (ERC), burning index (BM), and fuel moisture (FM). Recent work has shown that assessments generated for ERC, BI, and FM, are highly dependent on the period used for their generation (daily, weekly,

monthly, seasonal, or annual) and different periods can produce contradictory guidance (Lin et al. 2024). The SPC Probability of Wildfire Climatology considers only historical fire data and fails to directly account for present weather or fuel conditions, or topography data. The National Fire Danger Rating System does not provide a product out to ten days and presents classifications instead of probabilities.

The work described in this dissertation is an implementation of fire occurrence prediction (FOP). FOP is described by the work of Karniadakis et al. (2021) as "[p]redictions of the number and location of fire starts in the upcoming day[s]." Fire occurrence models generally use regression methods to model the relationship between historical fire data like fire reports and environmental factors like weather or fuel. FOP modeling for CONUS is not currently being done by SPC, USDA, or USFS.

Technical Motivation The gaps in existing tooling combined with the increasing prevalence of wildfire in the United States creates a unique opportunity to present a new approach to providing wildfire probability information for CONUS.

The three methods presented here, SPC Fire Weather Outlooks, SPC Probability of Wildfire Climatology, and the NFDRS, all present wildfire probability information for fire managers using a map of CONUS with areas of interest highlighted using different colors, polygons, or gradients.

In the case of SPC Fire Weather Outlooks, to inform on fire weather, forecasters must look at multiple screens, each containing a different map of a variable of interest. In some cases, the entire map is not visible to the forecaster and they must spend further time scrolling from region to region of the United States to collect all the information they need to feed their expertise and intuition. Some maps of individual variables attempt to capture broader seasonal and annual effects using a climatological

approach and others present data from the current day, which leaves the task of contextualization up to forecasters.

Because the process of looking at multiple screens, multiple maps, and multiple areas of different maps, can be time consuming, I am motivated to create a machine learning approach to offer first guess guidance to forecasters. Because forecasters consume inputs visually and produce a visual output, I look to computer vision techniques, and semantic segmentation specifically, for my approach.

Semantic Segmentation with UNet3+ for Wildfire SPC is tasked with producing regional guidance over the CONUS domain (as opposed to predicting individual events or regions). The clear preference for fire managers is to receive wildfire probability feedback in the form of a notated CONUS image. Current methods forecast probability of an event instead of classifying the event by sub-type. Taking these three points into consideration, I decided to approach this task using binary classification semantic segmentation so I could use a deep learning model to produce a probability of wildfire field that could be projected onto a CONUS image.

Semantic segmentation is a computer vision approach by which every pixel in an image is given a prediction based on a feature space unique to that pixel. Binary classification predicts the likelihood that a pixel is part of the class of interest.

By taking this approach, I am able to provide a proof of concept for providing SPC fire weather forecasters with a first guess for fire occurrence that can help to inform their polygon creation for fire weather or stand alone as a fire occurrence product. My proof of concept considers all of the variables fire weather forecasters would previously have had to look at themselves, can simultaneously consider multiple decades of data for each variable in isolation and in combination, and can do so without the need to convert variables to climatologies.

To model this task, I chose to focus on UNet models, the Unet3+ model specifically. UNets have a distinctive architecture comprised of symmetric encoder and decoder components connected at corresponding levels by skip connections. The encoder component of the architecture captures hierarchical feature maps while downsampling the image resolution allowing it to capture context. The decoder component of the architecture uses upsampling to rebuild the feature maps to their original image resolution and captures localization with pixel level accuracy (Ronneberger et al. 2015). Skip connections help to preserve spatial information by allowing the model to recover fine-grained details that would have been lost during downsampling (Drozdzal et al. 2016).

The encoder decoder architecture ensures that information primarily flows between neighboring pixels rather than across distant regions of the input domain. As a result, the network is particularly skilled at leveraging local context when computing outputs, which is often essential for accurate spatial predictions. UNets are constructed using convolutional layers, which greatly reduces the number of parameters compared to fully connected architectures. As the parameter count in UNets does not scale directly with the input domain size, they are more computationally efficient. UNets offer multi-scale feature representation that is useful for segmenting objects that vary in size (Ibtehaz and Rahman 2020) and are able to learn from limited data (Özgün Çiçek et al. 2016).

UNets were originally created for the biomedical domain but they have been successfully generalized to other domains, including wildfire. Semantic segmentation UNets have been applied to the detection of burned area (de Bem et al. 2020; Knopp et al. 2020), active fire detection (Rashkovetsky et al. 2021; de Almeida Pereira et al. 2021), and smoke detection (Wang et al. 2022). There are currently no examples of

the application of UNets, and the UNet3+ specifically, to the task of fire occurrence prediction (FOP).

Contributions I hypothesize that we can create a binary classification, semantic segmentation, deep learning model that can contextualize both local and regional feature information representing multiple non-climatological variable inputs. In Chapter 3, I describe the non-climatological data we used to fuel this model and in Chapter 4, I describe the UNet3+ architecture I implemented to accomplish this task.

I hypothesize that this model can produce CONUS-wide wildfire occurrence probabilities with as much, if not more, accuracy than the methods described above. The model performance of the models presented in this work is compared to the SPC Probability of Wildfire Climatology in case studies that evaluate the largest natural-caused fires (of operational importance to SPC), the largest human-caused fires, and the most common occurrence days for natural-caused fires and human-caused fires respectively. We chose the SPC Probability of Wildfire Climatology to compare to, as it is the closest analog to our work within the SPC product portfolio at this time. Lessons learned derived from each attempt are then discussed in Chapters 6 and 8.

I hypothesize that we can create a performative model using data sets already familiar to fire weather forecasters and contextualize model performance using metrics already familiar to fire weather forecasters. To assess these hypotheses, this dissertation investigates how much model performance can be achieved using two different input data sets: gridMET (Chapter 5) and the Global Ensemble Forecast System (GEFS) (Chapter 7). These two datasets are both used operationally by fire weather forecasters. To measure model performance, for general performance, regional performance, and seasonal performance, I used a metric called critical success index (CSI),

which is a performance metric used to evaluate model predictions within the National Weather Service (Schaefer 1990).

The main contribution of my PhD is the development of a deep learning method that outperforms current operational techniques used by forecasters for the task of fire occurrence prediction. My research demonstrates that machine learning can be used to assist in fire occurrence forecasting. Further, I demonstrate how model performance differs by region and by season. For models that used GEFS inputs, I describe variable contribution using single pass permutation variable importance. We illustrate how two commonly used datasets, gridMET and GEFS, perform differently on the task. We propose multiple approaches to characterizing data inputs and labels and contextualize them using model performance. Chapters 5 and 6 are drawn from two published papers Earnest et al. (2024a) and Earnest et al. (2024b). Chapters 7 and 8 are in preparation to be published.

Chapter 2

Literature Review

The work in this dissertation explores the use of machine learning to predict fire occurrence for the contiguous United States. To support future discussion, this chapter lays the foundation for understanding wildfire, machine learning as it relates to fire occurrence prediction, and existing fire occurrence prediction methods.

2.1 Wildfire Prediction

2.1.1 Wildfire Overview

The term 'fire' describes the exothermic chemical process of combustion in which rapid oxidation of material releases heat and light (National Wildfire Coordinating Group 2022a). The combustion reaction requires that an oxidizer (in the case of wildfire, oxygen in the atmosphere) is combined with a flammable material (dried organic matter on, under, or above the forest floor, i.e., fuel) and a heat source (a spark generated by humans or nature) at a temperature above the flash point of the oxidizer and flammable material mixture.

These three elements (oxygen, fuel, and heat) are commonly referred to as the fire triangle or the combustion triangle (National Park Service 2016). Wildfire firefighting strategies focus on removing one or more elements of the fire triangle from the

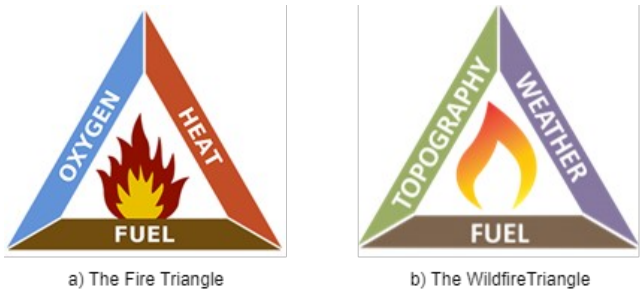


Figure 2.1: a) The Fire Triangle from National Park Service (2016) b) The Wildfire Triangle from Oregon State University (2019)

wildfire's environment to starve the fire of components necessary for the combustion reaction.

A wildfire is any fire resulting from an unplanned ignition, human or natural (National Wildfire Coordinating Group 2022c). Wildfire is governed by its own triangle, called the wildfire triangle or the fire behavior triangle, in which the key elements are fuel, weather, and topography (Oregon State University 2019). These three environmental factors influence if and how a fire will ignite, spread, and extinguish, and can work in isolation or in combination to support or suppress the risk of wildfire. Both the fire triangle and the wildfire triangle are depicted in Figure 2.1.

The fire and wildfire triangles are part of a larger ecosystem, called fire ecology which describes the effects of fire on living organisms and their environment (National Wildfire Coordinating Group 2022b). Fire managers have a part to play in fire ecology, as they work to control the time, frequency, and intensity of wildfire (Auburn University 2022).

Fire occurrence prediction (FOP) is described by the work of Jain et al. (2020) as "[p]redictions of the number and location of fire starts in the upcoming day(s)". Fire occurrence models generally use regression methods to model the relationship between historical fire data like fire reports (a formal document that captures the details of

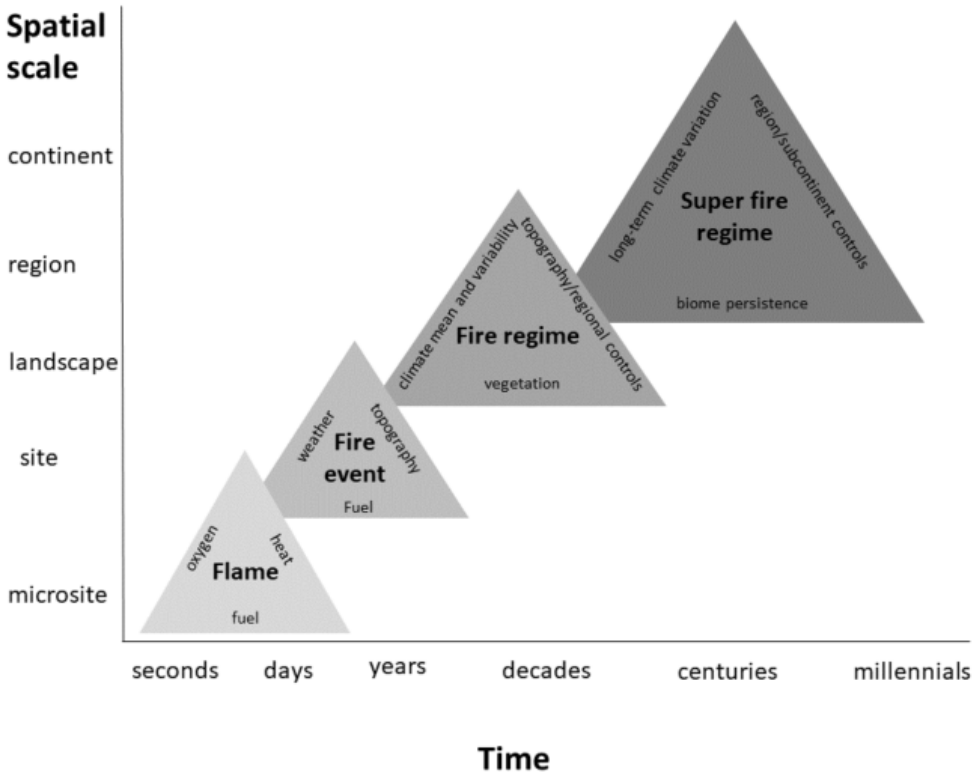


Figure 2.2: Fire Ecology. Figure from University of California Agriculture and Natural Resources (Accessed: 2024-06-22).

a fire incident) and environmental factors like weather or fuel. The motivation for building accurate FOP models is to support resource managers so they can coordinate with other fire management bodies and position resources to best manage high risk areas, to take proactive action, and to make and communicate meaningful decisions that can save lives, property, and expenses.

2.1.2 Numerical Weather Prediction Overview

One of the three datasets used for this work is GEFS, which is an example of numerical weather prediction (NWP) data. Numerical weather prediction is a process by which current weather observations are used to predict future weather variables, such as

temperature and precipitation. NWP uses initial conditions derived from observations of the current state of the atmosphere in combination with mathematical equations representing the physical laws that govern the atmosphere such as conservation of mass, momentum, and energy, to simulate the future state of the atmosphere. NWP plays a crucial role in predicting the future state of the atmosphere and has been used by governments all over the world for decades (Pu and Kalnay 2018).

A single NWP model is known as a deterministic forecast. Many NWP models are run in an ensemble format, whereby the model is run multiple times, each time with slight perturbations to the initial conditions. Ensemble approaches are used to account for uncertainty in the forecast. By using an ensemble approach, it has been demonstrated for numerous applications that better forecast skill can be achieved than is possible with a single, deterministic forecast (Kalnay et al. 1996).

There is currently no NWP model for fire occurrence prediction. I use machine learning to offer two alternative approaches. The first approach, described in Chapters 5 and 6, does not rely on NWP and uses only observed data as inputs. The second approach, described in Chapters 7 and 8, uses NWP forecasts as inputs. Using NWP data as input data for a machine learning model has a long history within the weather domain (Cho et al. 2022; Frnda et al. 2022). Some recent examples of using NWP as inputs for machine learning include the work of Hill et al. (2023) that used the GEFSv12 ensemble with random forests to predict medium range (days 4-8) severe weather and the work of Burke et al. (2020) that used machine learning to improve hail predictions from High-Resolution Ensemble Forecast system, version 2 (HREFv2). For a review of the use of AI for weather see Haupt et al. (2022).

2.1.3 Traditional Methods for Fire Probability Prediction

Information about fire probability, danger, and risk for CONUS is disseminated to the public many different ways. I will discuss three methods here: the SPC Fire Weather Outlook, the SPC Probability of Wildfire Climatology, and NFDRS. The SPC Fire Weather Outlook describes an environment in which the atmospheric elements, previous and current, are conducive to producing critical fire behavior. The SPC Probability of Wildfire Climatology depicts the probability of fire occurrence based on climatological data. The NFDRS describes the likelihood for "a fire to ignite, spread and require suppression action" (USFS 2024). All three methods seek to communicate fire probability (and, by extension, fire risk) for CONUS to the general public and to fire managers.

2.1.3.1 SPC Fire Weather Outlook

The SPC Fire Weather Outlook includes two severe weather products. The first product is the day 1 and day 2 forecast and the second product is the days 3-8 forecast. For days 1 and 2, fire risk is classified into elevated, critical, and extremely critical based on the characteristics of temperature, wind, and relative humidity, combined with the area of interest for the forecast (Florida, for example, would see elevated risk at different percentiles than the rest of the country) (SPC 2024a). For days 3-8, fire risk is characterized by percentiles. In creating outlooks, fire weather forecasters consider numerical weather prediction (NWP) data such as wind (speed and direction), temperature, and relative humidity, NFDRS variables such as energy release component (ERC), and historical fire data such as is found in the SPC Probability of Wildfire Climatology.

2.1.3.2 SPC Probability of Wildfire Climatology

The SPC Probability of Wildfire Climatology is one of the resources used by fire weather forecasters to create outlooks and relies on the Fire Program Analysis fire-occurrence database (FPA-FOD) to generate the climatology.

The SPC Probability of Wildfire Climatology is created using the following process (Nauslar 2020b). The process was created by Dr. Nick Nauslar, who, at time of writing, is the Fire Sciene and Operations Officer at the NOAA/NWS/SPC.

1) New reported wildfires for each day are put onto a grid 80 km x 80 km using the discovery date.

2) If one or more wildfires occur in a grid box three days before through three days after the valid date, that box is assigned the value one for the day. If no wildfires occur, the grid box is given a value of zero.

3) The raw frequency for each day at each grid location is found for the period (number of "1" values divided by number of years) to get a raw annual cycle.

4) The smoothed time series are then smoothed in space with a 2-D Gaussian filter (40 km in each direction).

2.1.3.3 National Fire Danger Rating System

The NFDRS, created in 1972 by the USDA and USFS, is presented here as another method for fire managers and the general public to receive information about the risk of fire (called fire danger or fire potential) in CONUS (Bradshaw et al. 1984).

NFDRS offers same day and next day products for CONUS through the Wildland Fire Assessment System (WFAS) (WFAS 2024; NWCG 2024). WFAS was removed from production on 02 August 2023 (WFAS 2024). At the time of writing, it has not been returned to production. Therefore, I did not use it as a comparison in the case

Fire Danger Level	*Description*
Low	Fuels do not ignite easily from small embers, but a more intense heat source, such as lightning, may start fires in duff or dry rotten wood. Fires in open, dry grasslands may burn easily a few hours after a rain, but most wood fires will spread slowly, creeping or smoldering. Control of fires is generally easy.
Moderate	Fires can start from most accidental causes, but the number of fire starts is usually pretty low. If a fire does start in an open, dry grassland, it will burn and spread quickly on windy days. Most wood fires will spread slowly to moderately. Average fire intensity will be moderate except in heavy concentrations of fuel, which may burn hot. Fires are still not likely to become serious and are often easy to control.
High	Fires can start easily from most causes and small fuels (such as grasses and needles) will ignite readily. Unattended campfires and brush fires are likely to escape. Fires will spread easily, with some areas of high-intensity burning on slopes or concentrated fuels. Fires can become serious and difficult to control unless they are put out while they are still small.
Very High	Fires will start easily from most causes. The fires will spread rapidly and have a quick increase in intensity, right after ignition. Small fires can quickly become large fires and exhibit extreme fire intensity, such as long-distance spotting and fire whirls. These fires can be difficult to control and will often become much larger and longer-lasting fires.
Extreme	Fires of all types start quickly and burn intensely. All fires are potentially serious and can spread very quickly with intense burning. Small fires become big fires much faster than at the "very high" level. Spot fires are probable, with long-distance spotting likely. These fires are very difficult to fight and may become very dangerous and often last for several days.

Table 2.1: NFDRS Fire Danger Level (USFS 2024)

studies described later in this work. I describe NFDRS here as it offers an analog to the work presented in this dissertation.

NFDRS ratings are posted on signs at the visitor entrance to public lands. NFDRS classifies fire risk into five categories, called Fire Danger Levels, detailed in Table 2.1, which attempt to inform on the potential for fire ignition, the potential for fire spread, and the potential for a fire to get out of control.

2.2 Machine Learning

2.2.1 Neural Networks Overview

'Neural network' is a blanket term for a collection of machine learning models that are inspired by the learning mechanisms in the human brain (Aggarwal 2018; Goodfellow

et al. 2016). Neural networks are comprised of neurons that can be interpreted as a function that takes in the outputs from the previous layer (or initial inputs if it is the input layer), multiplies it by a weight, adds a bias term, and produces a value to be passed to the next layer in a process called feed-forward propagation. Neural networks are comprised of layers that are a collection of neurons. Neural networks have three types of layers: an input layer, an output layer, and, in the case of deep learning networks, one or many hidden layers between the input and the output layer.

In this section, I will refer to an example deep learning neural network with 4 layers, an input layer, two hidden layers, each with four neurons, and an output layer. As models discussed elsewhere in this dissertation are image to image models, the example network discussed here takes in a 128 by 128 pixel image, flattened into a vector, and produces a 128 by 128 probability field, flattened into a vector, that represents the probability of an event being present in each pixel of the image.

The input layer and output layers include 16,384 neurons, one for each pixel in a 128 by 128 image. Each neuron can be interpreted as a function that takes in the outputs from the previous layer (or initial inputs if it is the input layer) and produces a value to be passed to the next layer in a process called feed-forward propagation that is described in the next section.

2.2.1.1 Feed-Forward Propagation

Feed-forward propagation is the process by which input data passes through a neural network from the input layer to the output layer, described in Equation 2.1. Feed-forward propagation relies on activation values, a, which are the value held by a neuron at any given time, weights, w, which are associated with the connections between layers, bias terms, b, which are associated with neurons, and the activation

function. In this example, I will discuss two activation functions, the *ReLU* function, which stands for 'rectified linear unit' (Nair and Hinton 2010) and the *sigmoid* function (Rumelhart et al. 1986), which I use as the final activation function in the output layer of the example neural network.

The calculation for feed forward propagation is as follows:

$$a^{(L+1)} = ReLU(wa^{(L)} + b), \tag{2.1}$$

where $a^{(L)}$ describes the column of activation values of the input layer, $a^{(L+1)}$ describes the vector of activation values of the layer following the input layer, w describes a matrix of weights for which each row represents the connections between the previous layer and the neuron associated with those connections in the following layer, and b describes the vector of bias terms that are added to the matrix vector product $wa^{(L)}$. If I define z to be a vector of pre-activation values (Aggarwal 2018), calculated by

$$z^{(L)} = wa^{(L)} + b, \tag{2.2}$$

for a given input vector $z^{(L)}$, the vector output $a^{(L+1)}$ of the *ReLU* function is described in equation 2.3.

$$a_h^{(L+1)} = \text{ReLU}(z_h^{(L)}) = \begin{cases} z_h^{(L)}, & \text{if } z_h^{(L)} > 0, and \\ 0, & \text{if } z_h^{(L)} \leq 0, \end{cases} \tag{2.3}$$

and whereby the activation function is applied to each value, h, of the pre-activation vector to produce a value of zero or greater. This calculation is repeated for all activation values in a layer, for all layers in a network, one layer at a time, proceeding from the input layer to the output layer. The process of forward propagation is depicted in Figure 2.3 for a single neuron.

The output layer, the last layer of a network, in the example network, uses a *sigmoid* final activation function that produces a value between 0 and 1 for all 16,384

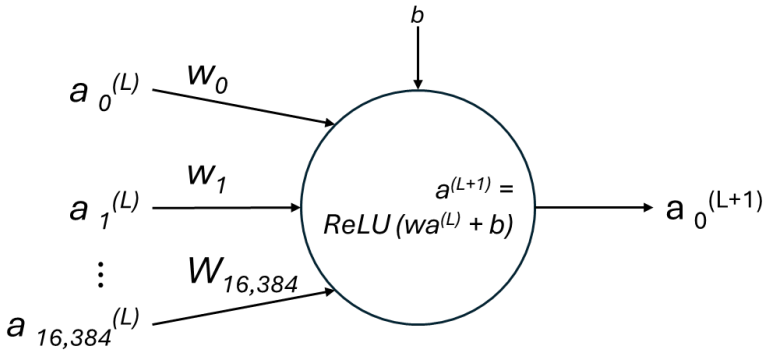


Figure 2.3: Neural Network Neuron

values in the output vector. The output of the *sigmoid* activation function can be interpreted as a probability. Given a value, $z^{(L+2)}$, which is the pre-activation vector of the layer prior to the output layer, the output vector of the *sigmoid* activation function, $\hat{y}$, which is the vector of predicted values produced by the network, is described in equation 2.4.

$$\hat{y} = \sigma(z^{(L+2)}) = \frac{1}{1 + e^{-z^{(L+2)}}} \tag{2.4}$$

2.2.1.2 Cost Function

To measure how well a network is performing on a given task, an aggregate measurement is used called a cost function, C, also called a loss function. Mean Squared Error (MSE) is a common cost function (Bishop 2006; Draper and Smith 1998) and is described in equation 2.5.

$$C = \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{2.5}$$

where n is the number of samples in the dataset being measured, y_i is the ith observed value from a vector of observed values, $\hat{y}_i$ is the ith predicted value from a vector of predicted values. The output of the MSE equation is a scalar value that describes

how well a network is performing, with higher values indicating a network that is performing poorly, and lower values indicating a network that is performing well. The range for MSE is zero to infinity. The goal of gradient descent is to minimize the cost function. This process is described in the gradient descent section and the back propagation section.

2.2.1.3 Parameter Space

The parameter space, or weight space, of a neural network is the set of elements the network can update during training to minimize the cost function. In the example network, the parameters are weights and bias terms and together, they comprise the parameter space. Therefore, the parameter space for this network is calculated

$$\text{Parameter Space} = \text{Weights} + \text{Biases} \tag{2.6}$$

where weights, w, are parameters associated with the connections between neurons in different layers and bias terms, b, are parameters associated with each neuron.

The deep learning neural network under discussion has 4 layers, an input layer, two hidden layers, each with four neurons, and an output layer. The network takes as input a 128 by 128 pixel image, flattened into a vector of 16,384 values. The network produces as output a 128 by 128 probability field, again, flattened into a vector of 16,384 values, which represents the probability of an event being present in each pixel of the image. The input layer and output layers each include 16,384 neurons, one for each pixel in a 128 by 128 image.

The total number of weights, w, in a neural network is calculated:

$$w = \sum_{L=1}^{L_{total}-1} n^{(L)} \times n^{(L+1)} \tag{2.7}$$

(Goodfellow et al. 2016), where $n^{(L)}$ is the number of neurons in layer L, $n^{(L+1)}$ is the number of neurons in layer $L+1$, and L_{total} is the total number of layers in the

network. My example network contains 131,088 weights, $131,088 = (16,384 \times 4) + (4 \times 4) + (4 \times 16,384)$.

The total number of bias terms, b, in a neural network is calculated:

$$b = \sum_{L=2}^{L_{total}} n^{(L)} \tag{2.8}$$

(Goodfellow et al. 2016), where $n^{(L)}$ is the number of neurons in layer L and L_{total} is the total number of layers in the network. My example network contains 16,392 bias terms, $16,392 = 4 + 4 + 16,384$.

The resulting parameter space contains 147,480 elements, 131,088 weights and 16,392 bias terms, which can be adjusted during the process of back propagation.

2.2.1.4 Gradient Descent

Gradient descent is the process within a neural network of finding the optimal set of parameter values to minimize the output of the cost function, $C(\mathbf{v})$, where $\mathbf{v}$ represents the vector of 147,480 weights and bias terms contained in the parameter space.

Gradient descent is an iterative optimization algorithm that can discover the local minimum of a function by calculating the gradient of the function, ∇C. The gradient of the function indicates the direction of steepest ascent for the function. By iteratively moving some amount, called the learning rate, η, in the opposite direction indicated by the gradient, $-\nabla C$, the algorithm is able to descend until it converges on a local minimum of the function (Rumelhart et al. 1986; Bishop 2006).

The behavior of gradient descent is described in the update rule (Nielsen 2015):

$$\mathbf{v}' = \mathbf{v} - \eta \nabla C, \tag{2.9}$$

where $\mathbf{v}$ represents the original vector containing the parameter space of the network, $\mathbf{v}'$ represents the updated vector containing the parameter space of the network, η

represents the learning rate, which is a scalar value representing the size of movement to be executed, and ∇C represents the gradient of the cost function, a vector.

2.2.1.5 Back-propagation

The gradient of the cost function used in the update rule of gradient descent is calculated using an algorithm called back-propagation (Rumelhart et al. 1986). Back-propagation is applied to each weight and bias in the parameter space.

Supporting Definitions The calculation of the activation value for the jth neuron in the output layer of the network, $a_j^{(L+3)}$, is:

$$a_j^{(L+3)} = \sigma(z_j^{(L+3)}), \tag{2.10}$$

where $L+3$ indicates the output layer of the four layer example network for which L is the input layer, j is the index of the neurons in the output layer of the network and $z_j^{(L+3)}$ is the pre-activation value for the jth neuron of the output layer.

The calculation of the pre-activation value for the jth neuron of the output layer is:

$$z_j^{(L+3)} = \sum_k w_{jk}^{(L+3)} a_k^{(L+2)} + b_j^{(L+3)}, \tag{2.11}$$

where $L+3$ indicates the output layer of the network, $L+2$ indicates the layer prior to the output layer of the network, j is the index of the neurons in the $L+3$ layer, k is the index of the neurons in the $L+2$ layer, jk is the index of the weights in the weight matrix, $w_{jk}^{(L+3)}$ is the weight associated with calculating the activation value of the jth neuron in the output layer, $a_k^{(L+2)}$ is the activation value of the kth neuron of the $L+2$ layer, and $b_j^{(L+3)}$ is the bias value associated with calculating the activation value of the jth neuron in the output layer.

The calculation of the cost function for an individual training sample, C_0, using MSE, is:

$$C_0 = \frac{1}{n^{(L+3)}} \sum_{j=0}^{n^{(L+3)}-1} (a_j^{(L+3)} - y_j)^2, \tag{2.12}$$

where $L+3$ indicates the output layer of the network, $n^{(L+3)}$ is the number of neurons in the output layer of the network, $a_j^{(L+3)}$ is the activation value of the jth neuron in the output layer, also described as the predicted value, $\hat{y_j}$, and y_j is the actual value associated with the jth neuron in the output layer.

Gradient for Weights The impact to the cost function for an individual training sample caused by changes to an individual weight in the parameter space is described by the derivative of C_0 with respect to $w_{jk}^{(L+3)}$ and is calculated using the chain rule:

$$\frac{\partial C_0}{\partial w_{jk}^{(L+3)}} = \frac{\partial z_j^{(L+3)}}{\partial w_{jk}^{(L+3)}} \frac{\partial a_j^{(L+3)}}{\partial z_j^{(L+3)}} \frac{\partial C_0}{\partial a_j^{(L+3)}} = a_k^{(L+2)} \sigma'(z_j^{(L+3)}) \frac{2}{n_L} (a_j^{(L+3)} - y_j). \tag{2.13}$$

The derivative of $z_j^{(L+3)}$ with respect to $w_{jk}^{(L+3)}$, $\frac{\partial z_j^{(L+3)}}{\partial w_{jk}^{(L+3)}}$, describes the impact to the jth pre-activation value of the output layer caused by changes to the $jkth$ weight in the weight matrix and is calculated:

$$\frac{\partial z_j^{(L+3)}}{\partial w_{jk}^{(L+3)}} = a_k^{(L+2)}. \tag{2.14}$$

The derivative of $a_j^{(L+3)}$ with respect to $z_j^{(L+3)}$, $\frac{\partial a_j^{(L+3)}}{\partial z_j^{(L+3)}}$, describes the impact to the jth activation value of the output layer caused by changes to the jth pre-activation value of the output layer and is calculated:

$$\frac{\partial a_j^{(L+3)}}{\partial z_j^{(L+3)}} = \sigma'(z_j^{(L+3)}). \tag{2.15}$$

The derivative of C_0 with respect to $a_j^{(L+3)}$, $\frac{\partial C_0}{\partial a_j^{(L+3)}}$, describes the impact to the cost function for a single training sample, C_0, calculated using MSE, caused by changes to the jth activation value of the output layer and is calculated:

$$\frac{2}{n_L}(a_j^{(L+3)} - y_j). \tag{2.16}$$

Gradient for Biases The impact to the cost function of an individual training sample caused by changes to an individual bias term in the parameter space is described by the derivative of C_0 with respect to $b_j^{(L+3)}$ and is calculated:

$$\frac{\partial C_0}{\partial b_j^{(L+3)}} = \frac{\partial z_j^{(L+3)}}{\partial b_j^{(L+3)}} \frac{\partial a_j^{(L+3)}}{\partial z_j^{(L+3)}} \frac{\partial C_0}{\partial a_j^{(L+3)}} = 1\sigma'(z_j^{(L+3)}) \frac{2}{n_L}(a_j^{(L+3)} - y_j). \tag{2.17}$$

The derivative of $z_j^{(L+3)}$ with respect to $b_j^{(L+3)}$, $\frac{\partial z_j^{(L+3)}}{\partial b_j^{(L+3)}}$, describes the impact to the jth pre-activation value of the output layer caused by changes in the jth bias term and is calculated:

$$\frac{\partial z_j^{(L+3)}}{\partial b_j^{(L+3)}} = 1. \tag{2.18}$$

Gradient for Previous Layer Activations The impact to the cost function of an individual training sample caused by changes to the activation of the previous layer is described by the derivative of C_0 with respect to $a_k^{(L+2)}$ and is calculated:

$$\frac{\partial C_0}{\partial a_k^{(L+2)}} = \sum_{j=0}^{n_L-1} \frac{\partial z_j^{(L+3)}}{\partial a_k^{(L+2)}} \frac{\partial a_j^{(L+3)}}{\partial z_j^{(L+3)}} \frac{\partial C_0}{\partial a_j^{(L+3)}} = \sum_{j=0}^{n_L-1} w_{jk}^{(L+3)} \sigma'(z_j^{(L+3)}) \frac{2}{n_L}(a_j^{(L+3)} - y_j). \tag{2.19}$$

The derivative of $z_j^{(L+3)}$ with respect to $a_k^{(L+2)}$, $\frac{\partial z_j^{(L+3)}}{\partial a_k^{(L+2)}}$, describes the impact to the jth pre-activation value of the output layer caused by changes in the kth activation value from the previous layer and is calculated:

$$\frac{\partial z_j^{(L+3)}}{\partial a_k^{(L+2)}} = w_{jk}^{(L+3)}. \tag{2.20}$$

Back propagation impacts activation values through the adjustment of weights and biases. The process of back propagation is performed for all layers similarly proceeding from the output layer to the input layer. For additional details, see Aggarwal (2018) and Goodfellow et al. (2016).

2.2.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are neural networks that specialize in processing data in a grid format such as images (a 2-dimensional grid of pixels). A CNN replaces matrix multiplication in the layers of its architecture with an operation called a convolution, which is involves dot products of small portions of the image, one portion at a time. In the case of a 2-dimensional grid of pixels, a convolution refers to a mathematical operation in which a function called a kernel (a 2-d array of parameters) is applied to an input (a 2-d grid of pixels) to create an output called a feature map.

Described in Goodfellow et al. (2016), the convolution of an image I with a 2D kernel K is defined as:

$$S(i,j) = (I * K)(i,j) = \sum_{m=0} \sum_{n=0} I(i+m, j+n) K(m,n). \tag{2.21}$$

where I is the input image, K is the kernel, (i,j) are the coordinates of the output location in the resulting feature map, and (m,n) are the coordinates in the kernel. This notation refers exclusively to this section and does not represent a continuation from the previous section.

Convolution describes the process by which a value is generated for each location in a feature map. The kernel moves over the input, each element of the kernel is

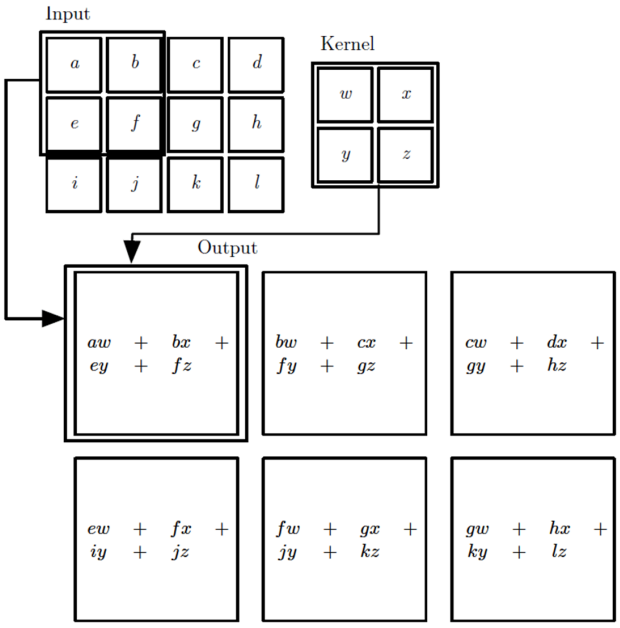


Figure 2.4: 2-D Convolution. Figure from Goodfellow et al. (2016).

multiplied by the corresponding element in the input. The products of the element-wise multiplication are summed to get the value for the location in the feature map. An example of this process is depicted in Figure 2.4.

2.2.3 UNet

The UNet is a convolutional neural network consisting of two parts, the encoder and the decoder, and is characterized by the U-shaped architecture from which it draws its name (Ronneberger et al. 2015). An example UNet architecture is depicted in Figure 2.5.

The encoder portion of the UNet architecture is composed of convolution layers with a rectified linear unit (ReLU) activation with each layer followed by a max-pooling operation. While passing through the encoder, the input image undergoes a

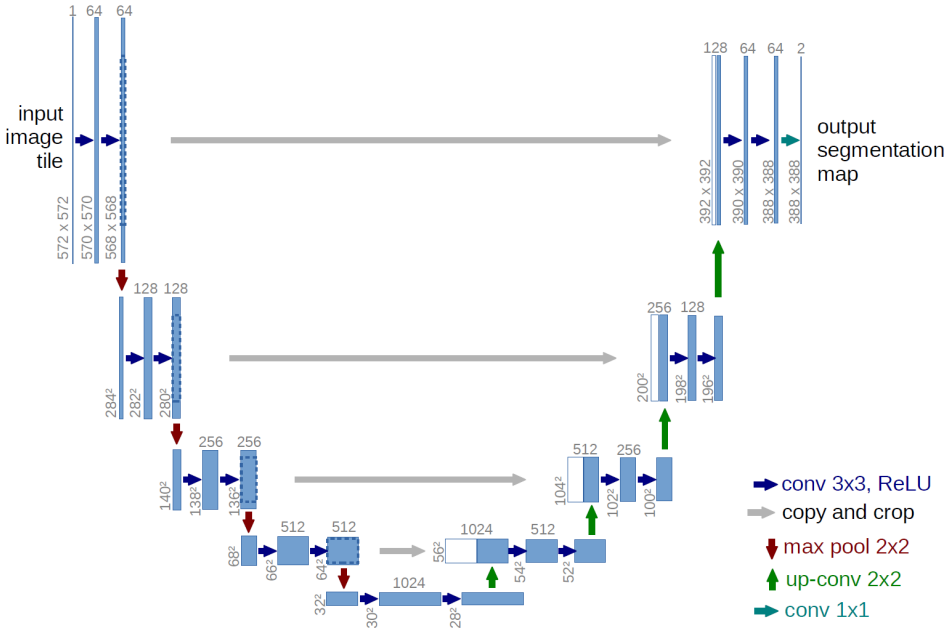


Figure 2.5: Traditional U-Net Architecture. Figure from Ronneberger et al. (2015).

process called down-sampling, whereby the spatial resolution of the image is reduced while the number of feature channels is increased.

Between the encoder and decoder, at the center of the UNet architecture, is a bottleneck layer consisting of multiple convolution layers without pooling that preserves spatial informational and retains information relevant for segmentation.

The decoder portion of the traditional UNet architecture, depicted in Figure 2.5, uses transpose convolutions, or upconvolutional layers, in a process called upsampling that increases the spatial resolution of the image while reducing the number of feature channels.

2.2.3.1 Skip Connections

Each layer in the decoder path is followed by concatenation with corresponding features maps from the encoder path, called skip connections, which help the model prevent the vanishing gradient problem, encourage feature reuse, reduce the risk of

overfitting, and, as previously stated, allow the model to recover spatial information lost during downsampling (Ronneberger et al. 2015). Skip connections are depicted in Figure 2.5 as horizontal arrows connecting the encoder portion of the architecture to the decoder portion of the architecture.

Skip connections, also called residual connections, were introduced in the work of He et al. (2015). The equation for skip connections, as described in He et al. (2015), is as follows:

$$S_1 = F(S_0, \{W_i\}) + S_0, \tag{2.22}$$

where S_0 is the input feature map, S_1 is the output feature map, and $F(S_0, \{W_i\})$ is the residual mapping to be learned.

2.2.3.2 Full-Scale Skip Connections

Skip connections play an important role in the UNet3+ model. The UNet3+ architecture's use of full-scale skip connections is one of the elements that separates it from previous UNet architectures. Full-scale skip connections include connections between the encoder and decoder at the same level, different levels, and between lower levels of the decoder and higher levels of the decoder. This allows the UNet3+ to learn low level details and high level semantics simultaneously by allowing nodes in the decoder to consume feature maps from multiple encoder levels and multiple decoder levels at once.

2.2.4 Neural Networks for Fire Occurrence Prediction

Regional Effects Fire occurrence frequency, size, and growth rates differ based on the type fuels present, which has region dependencies, and the state of the fuels present, which has seasonal dependencies (fuels are more fire prone during warm, dry

periods, such as drought). Regional effects are supported by the topography present in the region, fires tend to spread faster uphill, slopes facing the sun tend to have dryer fuels, and elevation effects the types of fuel present. Additionally, the land and fire management decisions made both to prevent fire and to manage active fire vary by region and are codified by the various land management entities - both state and federal.

The majority of prior work at the intersection of neural networks and FOP found success by limiting the overland area considered by the model to areas much smaller than CONUS, an area of over three million square miles containing many dissimilar regions. Alonso-Betanzos et al. (2002, 2003) used an artificial neural network in the Galicia region of Spain, an area of 11,419 square miles, and Vasilakos et al. (2007) used three different artificial neural networks in the Greek island of Lesvos, an area of 630 square miles. The notable exception to this trend is the work of Dutta et al. (2013), that compared ten different implementations of artificial neural networks and modeled all of Australia, an area of over 2.9 million square miles.

Fire Cause There are two main classifications of fire cause- natural and human. Natural caused fires, commonly caused by lightning in the United States, have regional components, lightning fires are more common in certain regions, environmental components, for a lightning fire to exist, lightning must exist, and situational components, lightning must exist separately from rain. These elements often differ between fire causes. Human caused fires hail from many sources such as arson, escaped campfires, and malfunctioning equipment. While the environmental elements of a lightning caused fire are, to a degree, predictable, the elements leading to an arson fire, for example, are less predictable.

For these reasons, many previous research attempts in fire occurrence address the complexity introduced by multiple fire causes by limiting their focus to a single fire cause much in the same way that previous research addressed regional effects by limiting the overland area considered by the model. Vega-Garcia et al. (1996) used an artificial neural network to predict human-caused wildfires in Alberta, Canada. Vecín-Arias et al. (2016) used a random forest model to predict lightning-caused fires in the Iberian Peninsula. In an approach that limited both cause and region, Fusco et al. (2016) used the Fire Program Analysis Fire-Occurrence Database (FPA-FOD) data set to quantify the role of humans in fire ignition in the western United States.

Chapter 3

Datasets: gridMET, GEFS, and FPA-FOD

Three datasets are used in the creation of this work: FPA-FOD, gridMET, and GEFS. All three datasets are used operationally by fire weather forecasters. In Table 3.1, I identify which datasets are used for each chapter in this work. For all models in this work, labels are sourced from FPA-FOD, which contains spatial wildfire occurrence data for CONUS and combines data sourced from the reporting systems of federal, state, and local organizations (Short 2023). The FPA-FOD observations are considered to be ground truth for the models in this work. For models presented in Chapter 5 and 6, inputs are sourced from gridMET, a daily, CONUS-wide, high-spatial resolution data set of observed surface meteorological variables, that includes weather-derived fuel variables, also referred to as fire danger variables (Abatzoglou 2013). For models presented in Chapters 7 and 8, inputs are sourced from a reforecast dataset of the GEFS ensemble, which is an ensemble of reforecast, global, gridded, weather data, both observed and forecast, created by NOAA (National Oceanic and Atmospheric Administration 2024).

Chapter	*Input Data*	*Label Data*
5 & 6	gridMET	FPA-FOD
7 & 8	GEFS	FPA-FOD

Table 3.1: Datasets by Chapter

Chapter	*Model*	*Time*	*Space*
5 & 6	Large Lightning	2000-2020	CONUS
5 & 6	All Fires	2000-2020	CONUS
7 & 8	Multi Label	2000-2019	CONUS
7 & 8	Pixel Label	2000-2019	CONUS

Table 3.2: Data Domain (time and space) for Models by Chapter

All models in this work operate over the CONUS spatial domain. Models presented in Chapters 5 and 6 use the time domain of 2000 to 2020. Models presented in Chapters 7 and 8 use the time domain of 2000 to 2019 (at the time of writing, GEFS does not include the year 2020). The data domains for the models presented in this work are described in Table 5.1. Controlled fires are not included the dataset.

3.1 FPA-FOD

FPA-FOD is a spatial database of United States wildfires created and updated by Short (2023). Data for FPA-FOD are sourced from the National Interagency Fire Management Integrated Database, the USDI Wildland Fire Management Information system, the FWS Fire Management Information System, the National Association of State Foresters database, the National Fire Incident Reporting System, the Integrated Reporting of Wildland Fire Information application, the Interagency Fire Occurrence Modules, the ICS-209 module of the SIT209 program, and from 33 U.S. states and the territory of Puerto Rico.

While Short removed redundancies and applied error checking, FPA-FOD is not without limitations. Data is not available for all years for all states. To establish a measure of completeness of the data for fire count and area burned, Short compared

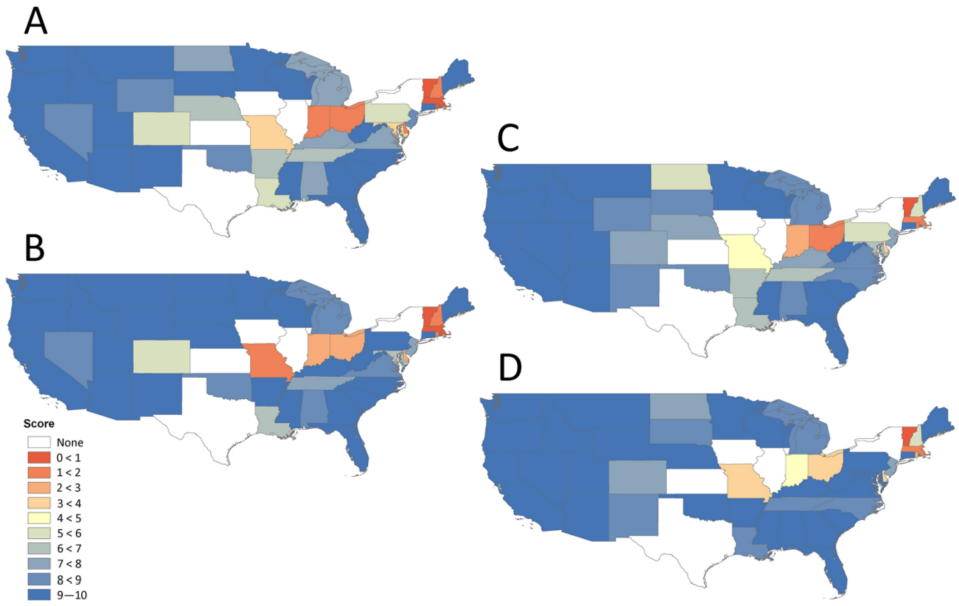


Figure 3.1: Agreement Scores: Indicating Agreement Between FPA-FOD Values and National Estimates for Wildfire Numbers and Burned Area. (A) wildfire numbers, 1992–2011, (B) wildfire numbers, 2002–2011, (C) wildfire area burned, 1992–2011, and (D) wildfire area burned 2002–2011. Figure from Short (2023).

FPA-FOD by state and year to national estimates of wildfire numbers and burned area. States were either given a score between 0 and 10 to indicate agreement between these values or no score to indicate omission due to reporting bias for the given time period, of which there were two depicted, 1992 to 2011 and 2002 to 2011. Low scores, indicating relative incompleteness of the data for that area for both time periods, were given to Missouri, Indiana, Ohio, Vermont, New Hampshire, Massachusetts, Rhode Island, and Delaware. No score, representing identified reporting bias, has been given to Iowa, Illinois, Kansas, New York, and Texas. Agreement scores are depicted in Figure 3.1.

Chapter	*Model*	*Fire Occurrence in Labels*
5 & 6	Large Lightning	Fire Occurrence with Cause = 'Natural' & Final Fire Size >= 1000 Acres
5 & 6	All Fires	All Fire Occurrence, regardless of size, cause, or region
7 & 8	Multi Label	All Fire Occurrence, regardless of size, cause, or region
7 & 8	Pixel Label	All Fire Occurrence, regardless of size, cause, or region

Table 3.3: Fire Occurrence Included in Labels for Models by Chapter

3.1.1 Selecting Fire Occurrence for Labeling

Fire occurrences are selected using the discovery date of the fire. If the discovery date falls within the range of years for the experiment, the fire occurrence is selected, with the exception of the "Large Lightning" model, which is discussed below. I recognize that some uncertainty is introduced by the choice of discovery date as the initial point of consideration for each fire as fires can smolder and not be discovered for some time post ignition but this is currently the best available information about the earliest stages of a fire.

For most models discussed in this work, all observations of fire occurrence from FPA-FOD, regardless of size, cause, or region, are selected for use in the label images with one major exception. In Chapters 5 and 6, for the "Large Lightning" model, only fire occurrence instances with a cause of "Natural" and a final fire size greater than 1000 acres are selected for use in the label images. Which fire occurrence instances are selected for use in the labels for each model can be found in Table 3.3.

3.1.2 Converting Tabular Data to Gridded Data

FPA-FOD stores instances of fire occurrence in a tabular format. To convert the fire occurrence data into an image format, I started by associating the individual fire occurrence instances with a 40km grid, selected for its compatibility with other severe weather products within SPC. For all grid cells where one or more fires are discovered

on a given day, I assigned a value of one to that gridcell for that day. I then populated all non-fire grid cells with a value of zero and did so for all days to ensure coverage for the time and space domain.

Associating Fires with a Grid Fire occurrence are selected using the discovery date, discovery latitude, and discovery longitude of each fire occurrence instance. The discovery latitude and longitude are associated with the x and y coordinates of the 40km grid which represent the center point of each grid cell. This is done using a nearest-neighbor approach. Discovery date is associated with a sample id, a key variable generated to represent all days within the time domain.

Summarizing Fires per Day per Grid Cell Fire occurrences are grouped by sample id, x, and y, to create a summary value of all fires discovered on a given day in a given grid cell. The summary value is then converted to a binary value, wherein any grid cell with a value greater than zero is given a value of one. All other grid cells retain the value of zero.

Domain Coverage As fires are not discovered in all grid cells on all days, the data set is expanded to include all days. All grid cells not containing fire discovery are given a value of zero. In this way, I am able to expand the data set to provide coverage for the entire time and space domain. Additionally, to align with the model architecture, the grid is cropped to a 128 by 128 size and centered on CONUS.

This process results in one, gridded, 128 by 128 CONUS image per day for all days within the time domain and with each grid cell given a value of one or a value of zero.

3.1.3 Addressing Imbalance

There are two types of imbalance present in the label dataset: scarcity of the event (presence of fire) among the label images and scarcity of the event within the individual label images.

Class Imbalance The first type of imbalance is a class imbalance in favor of 'no fire' (or a scarcity of the event **among** the label images). Class imbalance is present for the "Large Lightning" model presented in Chapters 5 and 6. In the case of the "Large Lightning" model, 77.26% of daily label images have no instances of fire occurrence of natural cause and final fire size greater than 1000 acres in any grid cell in CONUS. Class imbalance is not an issue either for the "All Fires" model, also presented in Chapters 5 and 6, or for the models presented in Chapters 7 and 8, as 0% of label images had no fire occurrence in any grid cell in CONUS. All three of these models use all instances of fire occurrence in their label images, regardless of size, cause, or region. As a result, the smallest number of instances of fire occurrence depicted in an image for these models is 7.

To address class imbalance, under sampling of samples with no fires or a low number of fires in the label image is performed. Any image from the training data set containing less than or equal to two fire occurrence instances within the image is dropped. Under sampling is performed for the training data set only. I selected two instances of fire as the cutoff for image exclusion by comparing the average percent of pixels per image to the percent of label images that could be retained at a given cutoff. The point at which the two lines intersected (albeit, on a multiple axis plot), as depicted in Figure 3.2.A, is then selected as the cutoff. As the "Large Lightning" model had a lower percentage of label images with instances of fire available, the cutoff for both models presented in Chapters 5 and 6 is determined by the cutoff

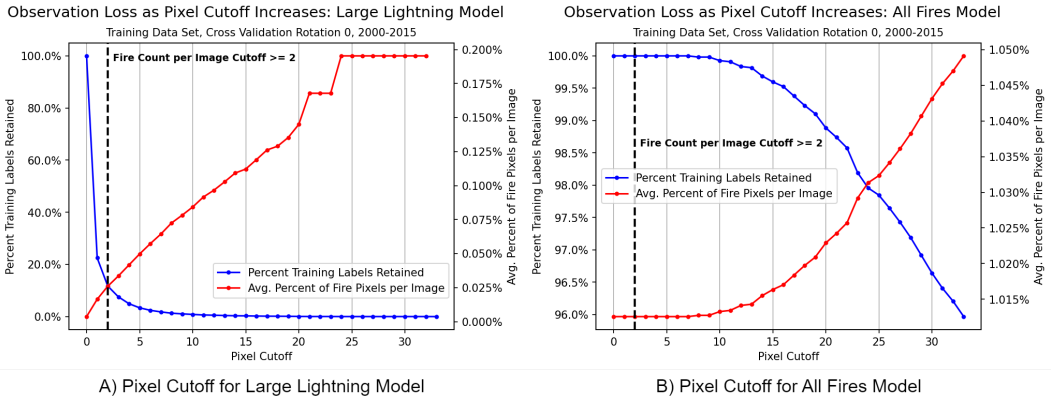


Figure 3.2: A) Pixel Cutoff for Large Lightning Model. B) Pixel Cutoff for All Fires Model. The red line describes the average percent of pixels indicating presence of fire per image. The blue line describes the percent of training labels that could be retained if samples with less than or equal to a given amount of pixels were removed. The black line indicates the cutoff (>= 2) selected for use in both models.

most appropriate for the "Large Lightning" model to ensure as many label images as possible were retained while increasing the frequency of the class of interest as much as possible. For the "Large Lightning" model, a cutoff of greater than or equal to two allowed 12% of the training images to be retained and, for the "All Fires" model, allowed 100% of the training images to be retained.

Pixel Imbalance The second type of imbalance is a pixel imbalance within the label images in favor of 'no fire' (or a scarcity of the event **within** the label images). For example, for the "All Fires" model presented in Chapters 5 and 6, the "Observation Inputs" model, and the "Forecast Inputs" model presented in Chapters 7 and 8, for which all fires were labeled regardless of size, cause, or region of discovery, the percentage of pixels labeled 'fire' fell between .04% and 3.42%.

To address the pixel imbalance within the label images, for the training dataset only, I used data augmentation. I used three methods for depicting fire occurrence

within the CONUS label images: the "Pixel" method, the "Neighborhood" method, and the "Time Smoothed Neighborhood" method. I describe these methods in detail below.

These three methods allow occurrences to be depicted more prominently in time and space and that helps to support the model's ability to learn meaningfully about the event of interest.

3.1.4 Labeling Methods

For the training dataset, three methods for depicting fire occurrence within the CONUS label images are used: the "Pixel" method, the "Neighborhood" method, and the "Time Smoothed Neighborhood" method. All three methods are used together for training. Only the "Pixel" method is used for validation and test.

Pixel The "Pixel" method provides a binary representation of fire occurrence within the CONUS label image. Each CONUS label image represents one day of CONUS and contains 16,384 pixels. Each pixel represents a single grid cell from a 40 km grid. On a given day, within a given grid cell, if one or more fire occurrences have been discovered, the grid cell is given a value of one. If no fire occurrences have been discovered, the grid cell is given a value of zero. The "Pixel" method is depicted in Figure 3.3.

Neighborhood The "Neighborhood" method built upon the "Pixel" method by increasing the size and numerical prominence of the "Pixel" label using a roughly linear decay approach. A central value of one is maintained, surrounded by values of .66, and then by values of .33, as depicted in Figure 3.4. This method of augmentation makes the binary labels more visually prominent so they are easier for the model to

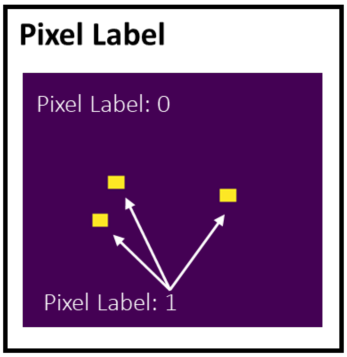


Figure 3.3: FPA-FOD Label Types: Pixel Label. Each pixel that had one or more wildfires discovered within it on a given day was given a value of one. Otherwise, the pixel was given a value of zero.

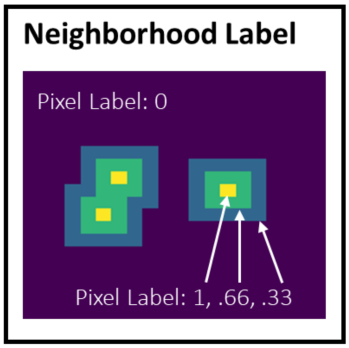


Figure 3.4: FPA-FOD Label Types: Neighborhood Label. Each pixel with a value of one is surrounded by a ring of pixels with a value of .66, that is itself surrounded by a ring of pixels with a vlue of .33.

see. As my goal is to make the labels more visually prominent, I elected not to use a Gaussian approach as a Gaussian would have expanded the spatial area of the label while uniformly dispersing the center value of one that would result in a comparable spatial area but reduced numerical prominence.

Time Smoothed Neighborhood The "Time Smoothed Neighborhood" method builds upon both the "Pixel" and "Neighborhood" methods by increasing the size of

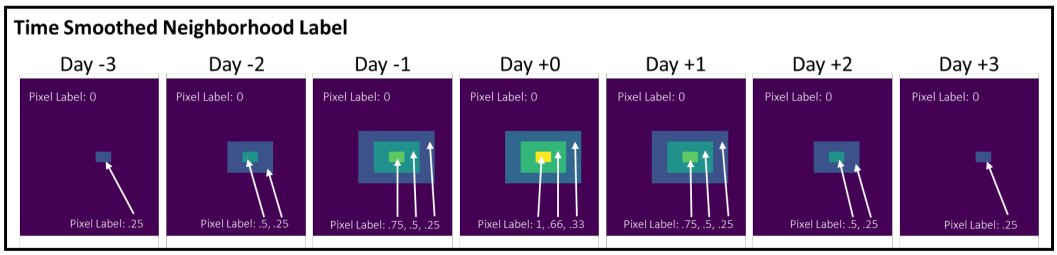


Figure 3.5: FPA-FOD Label Types: Time Smoothed Neighborhood Label. Each neighborhood label reduces in size and value as it departs in time from the date of discovery three days prior and three days post.

the label in both time and space. The "Time Smoothed Neighborhood" method takes into account the three days prior and three days post the fire occurrence discovery date. The "Neighborhood" method is used for the discovery date and then reduced both in size and numerical prominence each day as it moves away in either direction from the date of discovery in time, as depicted in Figure 3.5.

Within one day removed from the date of discovery, the center pixel is given a value of .75, surrounded by a ring of pixels with values of .5, and then surrounded by a ring of pixels with values of .25. Within two days removed from the date of discovery, the center pixel is given a value of .5 surrounded by a ring of pixels with values of .25. Within three days removed from the date of discovery, the center pixel is given a value of .25. This method is used because it mimics the method used to produce the Storm Prediction Center (SPC) Probability of Wildfire Climatology to which I compare model performance.

3.2 gridMET

gridMET combines climate data from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) with temporal attributes from regional reanalysis

Figure 3.6: Normalized gridMET Inputs as Consumed by the Model

(NLDAS-2). gridMET is validated against a network of weather stations (RAWS, AgriMet, AgWeatherNet and USHCN-2). The resulting data set is both spatially and temporally complete, providing daily coverage over the entire CONUS from 1979 onward (Abatzoglou 2013).

From gridMET, fifteen variables are sourced for the models presented in Chapters 5 and 6. These variables include four weather-derived fuel variables, or fire danger variables, one topography variable, two drought variables, and eight observed, weather variables. These variables are depicted in Figure 3.6. In addition to the sourced variables, three index variables are included- latitude, longitude, and Julian day of year.

There are three main types of variables in gridMET: the static variable, the daily variable, and the non-daily variable. For the daily variables, a new reading is available for every day in the dataset. For the non-daily variables, readings are available every 5 days. For the static variables, only one reading is available for the entire dataset.

Variable	*Type*	*Description*
Latitude	Static	Latitude of Centroid of Grid Cell
Longitude	Static	Longitude of Centroid of Grid Cell Latitude
Day of Year	Daily	Julian Day of Inputs
Dead Fuel Moisture 100 Hour	Daily	Moisture Content of Dead Organic Fuels - 1 to 3 inch diameter class
Dead Fuel Moisture 1000 Hour	Daily	Moisture Content of Dead Organic Fuels - 3 to 6 inch diameter class
Burning Index	Daily	The Contribution of Fire Behavior to the Effort of Containing a Fire
Energy Release Component	Daily	NFDRS Index Related to How Hot a Fire Could Burn
Elevation	Static	Height Above Sea Level
Palmer Drought Severity Index (and associated Category)	Non-Daily	Estimate of Relative Soil Moisture Conditions
Wind Speed	Daily	Wind Velocity at 10-meters
Wind From Direction	Daily	Wind Direction
Precipitation Amount	Daily	The Amount of Rain, Snow, Hail, etc., that has Fallen at a Given Place within a Given Period
Specific Humidity	Daily	The Weight of Water Vapor Contained in a Unit Weight of Air
Relative Humidity (Minimum)	Daily	The Amount of Atmospheric Moisture Present Relative to the Amount that would be Present if the Air Were Saturated
Surface Downswelling Shortwave Flux in Air	Daily	The Sum of Direct and Diffuse Solar Radiation Incident on the Surface
Mean Vapor Pressure Deficit	Daily	The Difference between the Amount of Moisture in the Air and How Much Moisture the Air can Hold
Potential Evapotranspiration (Daily)	Daily	The Combined Loss of Water Through the Plant's Process of Transpiration via its Vascular System and Evaporation of Water from the Earth's Surface

Table 3.4: gridMET Input Variables with Type and Description. Descriptions adapted from Ahrens (2009).

All variables sourced from gridMET, or created for the Chapters 5 and 6 models, are described in Table 3.4 along with the variable type.

3.2.1 Preprocessing

To prepare gridMET data for use in the model, the data set is regridded to 40km using linear interpolation, missing data are filled with a value of zero, and the data set is normalized to values between zero and one using linear min-max normalization. Aside these three pre-processing steps, data are left in their original form instead of being converted to a climatological form. This represents a departure from how this

data set is used by operational forecasters who consume it in a climatological form. Recent work has shown that fire danger variables specifically, such as burning index and energy release component, both included in the model dataset, are highly dependent on the range of time used to generate the climatology (daily, weekly, monthly, seasonal, or annual) and can produce contradictory guidance depending on the time range used(Lin et al. 2024). Therefore, I chose to format the model output in a non-climatological format.

Regridding Gridded Data gridMET stores data in a gridded format, gridded to 1/24th degree (approximately 4 km). Using linear interpolation, I convert gridMET's 4 km grid to 40 km grid, selected to align with operational weather products. I conducted a lengthy investigation of using a 3 km grid instead of a 40 km grid and found that it was not feasible given time and computation constraints. To align with the model architecture, the grid is cropped to a 128 by 128 size and centered on CONUS.

Missing Data While gridMET does have good coverage for the domain of interest, it did have missing values where the 128 by 128 grid overlapped with the Atlantic and Pacific oceans off the east and west coasts of CONUS. We address these missing values by filling them with a value of zero.

Data Normalization Normalizing model inputs when using neural networks helps to stabilize gradient descent by shifting all inputs into the same scale. This can lead to faster model convergence to an optimal solution. For the work presented in this dissertation, data are normalized to values between 0 and 1 using linear normalization (Bishop 2006) calculated as follows:

$$Z_i = (X_i - min(X))/(max(X) - min(X)), \tag{3.1}$$

where X represents the variable prior to normalization and Z represents the normalized variable.

Unique Processing for Non-Daily Variables For the non daily variables (for which readings are present every 5 days) an extra step to back-fill the readings to fill prior empty days is added between the re-gridding step and the fill missing data step. The reading available for day 10, for example, is cast backwards and used to populate the four prior empty days 9, 8, 7, and 6.

Unique Processing for Static Variables For the static variables (for which only one set of readings exist for the entire dataset), readings are filled forward to create daily readings (identical for all days).

3.2.2 Input Data Latency

As the goal of my work is to produce an operationalizable model, for which this work presents a proof-of-concept, input data latency is a consideration. gridMET provides observational data time stamped from midnight to midnight, western time. gridMET is updated daily, between approximately 5AM and 2:30PM (depending on the variable) with yesterday's data, resulting in about a day and a half of data latency.

Depending on what time forecasters reference gridMET on a given day, the most recently available data is either yesterday's data or data from the day before yesterday. It is generally acknowledged within SPC that, at a minimum, day-of observational data is preferable. gridMET is the preferred source for common variables used operationally by fire weather forecasters and therefore the source I used for this work even though it does not provide day-of observational data.

The impact of data latency is minimized to the extent possible by sourcing input data from the previous day for my model. The fire occurrence predictions generated by my model for today were created using the input data from yesterday. This is by design and intended to align with the update frequency of gridMET.

3.3 GEFS

The Global Ensemble Forecast System (GEFS), version 12 (GEFSv12), includes a dataset of reforecast data created by NOAA. 'Reforecast' describes retrospective forecasts produced for past time periods using current models. The reforecasts are generated for 5 ensemble members, c00, p01, p02, p03, and p04 with c00 representing the control and p01-04 representing perturbed members. Perturbed members used breeding vector and ensemble transform with rescaling (BV-ETR) cycling perturbations (Wei et al. 2008) to create initial conditions (Guan et al. 2022). Reforcasts are created using 00 UTC initial conditions for two lead times, 3 hours and 6 hours.

From GEFS, 5 variables are sourced for use in the model: precipitable water, specific humidity, temperature, v-wind (north-south wind), and u-wind (east-west wind). North-south wind and east-west wind are combined into a wind magnitude using the following calculation, with $wind_{ns}$ referring to the north-south wind component and $wind_{ew}$ referring to the east-west wind component:

$$wind_{magnitude} = sqrt(wind_{ns}^2 + wind_{ew}^2). \tag{3.2}$$

The wind magnitudes are offered to the model at two levels - 10 meter and 100 meter.

My motivation in selecting these 5 variables is to create coverage across the main weather components of fire weather that include precipitation, relative humidity, temperature, and wind. From there, I use the variables from GEFS that are easiest to

consume. I discuss later, in the variable importance section, how similar performance would have been achieved without all five variables.

3.3.1 Preprocessing

To prepare the GEFS variables for use in the model, both the temporal and spatial dimensions of the data are adjusted. I describe these adjustments in detail in the following sections.

3.3.1.1 Temporal Adjustment: Convert to Daily Readings

Each GEFS file contains both 3 hour readings (the even 'steps' indexed by 0, the odd 'steps' indexed by 1) and 6 hour readings (the odd 'steps' indexed by 0, the even 'steps' indexed by 1) for 10 days, resulting in 80 readings, or 'steps', per file. GEFS files are indexed from 1 while common computer science practice requires that I index from 0, therefore, I will refer to the steps by 0 index going forward.

Assign Hourly Readings to Days To use the 6 hour readings from each file, I reduce the steps contained in the data file to only the odd steps, which reduces the step count in each file from 80 to 40. With the remaining 40 steps, I assign them to an appropriate day, average all ensemble members together, and aggregate all readings for a given day using a selected method. Each step is described in a dedicated section below.

I assign each of the 40 steps to an appropriate day and repeat this process for all ensemble members. For example, the first two steps, representing hours 0 to 12, are assigned to day 0, resulting in 10 readings for day 0 (5 ensemble members x 2 steps). The next four steps, representing hours 12-36, are assigned to day 1, resulting in 20 readings for day 1 (5 ensemble members x 4 steps). And so on until day 10 to which

Day	*GEFS Hours*	*GEFS Step* (*Indexed by* 1)	*Step* (*Indexed by* 0)
0	0-12	2, 4	1, 3
1	12-36	6, 8, 10, 12	5, 7, 9, 11
2	36-60	14, 16, 18, 20	13, 15, 17, 19
3	60-84	22, 24, 26, 28	21, 23, 25, 27
4	84-108	30, 32, 34, 36	29, 31, 33, 35
5	108-132	38, 40, 42, 44	37, 39, 41, 43
6	132-156	46, 48, 50, 52	45, 47, 49, 51
7	156-180	54, 56, 58, 60	53, 55, 57, 59
8	180-204	62, 64, 66, 68	61, 63, 65, 67
9	204-228	70, 72, 74, 76	69, 71, 73, 75
10	228-240	78, 80	77, 79

Table 3.5: Mapping of GEFS Hours to Days

the last two steps are assigned, representing hours 228 to 240. Table 3.5 is included to clarify which steps were associated with which days.

Average Ensemble Members The steps for each day are averaged across ensemble members. Continuing the example from the previous section, this condenses the 10 readings for day 0 (5 ensemble members x 2 steps), into 2 average readings, one for each step.

Daily Aggregation All steps for a given day are aggregated using a previously decided upon method (mean, min, or max depending on the variable) as described in Table 3.6. This reduces the 2 readings for day 0, for example, down to 1 reading representing the entire day. This is done for all days 0 to 10.

Variable Name	*Daily Aggregation Method*
Temperature - 2 M	mean
Specific Humidity	min
Precipitable Water	max
Wind (east-west) - 10 M	max
Wind (east-west) - 100 M	max
Wind (north-south) - 10 M	max
Wind (north-south) - 100 M	max

Table 3.6: Daily Aggregation Method for GEFS Variables

3.3.1.2 Spacial Adjustment: Regrid & Resize

The GEFSv12 data represents a global domain with 0.25-degree grid spacing (a little less than 30 km). My work is designed to operate on a 40 km grid, to align with computational constraints, and representing only CONUS, to align with operational needs.

Negative values of longitude are commonly used for domains that are centered over CONUS. To convert the GEFS global domain to the CONUS domain, I converted the longitudinal coordinate system from 360 degrees to 180 degrees using the following calculation:

$$Longitude_{180} = (Longitude_{360} + 180)\%360 - 180. \tag{3.3}$$

I subset the dataset to include only data points that represent CONUS. Finally, I regridded the dataset from its native 30km grid to a 40km grid using linear interpolation.

This results in one, gridded, CONUS image per day per variable to be used by the model. Figure 3.7 depicts a day of input images for the input variables from GEFS.

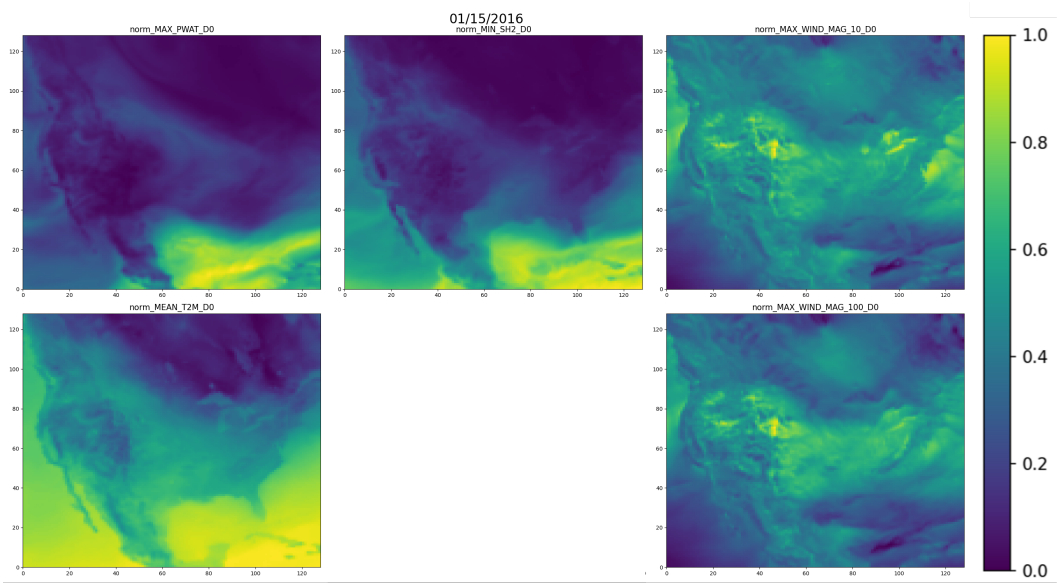


Figure 3.7: Normalized GEFS Inputs as Consumed by the Model

Chapter 4

Model, Loss, & Validation

I approach the challenge of modeling of fire occurrence as a semantic segmentation task. By doing so, I am able to consider inputs and provide predictions for the entire CONUS area while utilizing grid size to align with operationalized weather products. In support of this approach, I selected a model architecture suited to semantic segmentation, the UNet3+ model (Huang et al. 2020).

4.1 UNet3+

A UNet is a type of Convolutional Neural Network (CNN) architecture with an encoder-decoder structure, often depicted in the shape of a 'U'. The encoder half of the architecture captures feature maps while reducing the resolution and size of an image as it passes from the upper layers to the lower layers of the architecture, a process called downsampling. The decoder half of the architecture rebuilds the image to its original resolution, a process called upsampling. The encoder-decoder architecture is depicted in Figure 4.1.

The encoder-decoder architecture has a vulnerability in that information is lost as an image is reduced in resolution and size during the downsampling process. To compensate for this information loss, the UNet architecture incorporates same-scale skip connections, a connection between two layers, from the encoder to the decoder, at the same level within the architecture. Figure 4.2 depicts same-scale skip connections.

Different scale skip connections connect layers from different levels within the architecture, either from the encoder to the decoder or from lower levels of the decoder

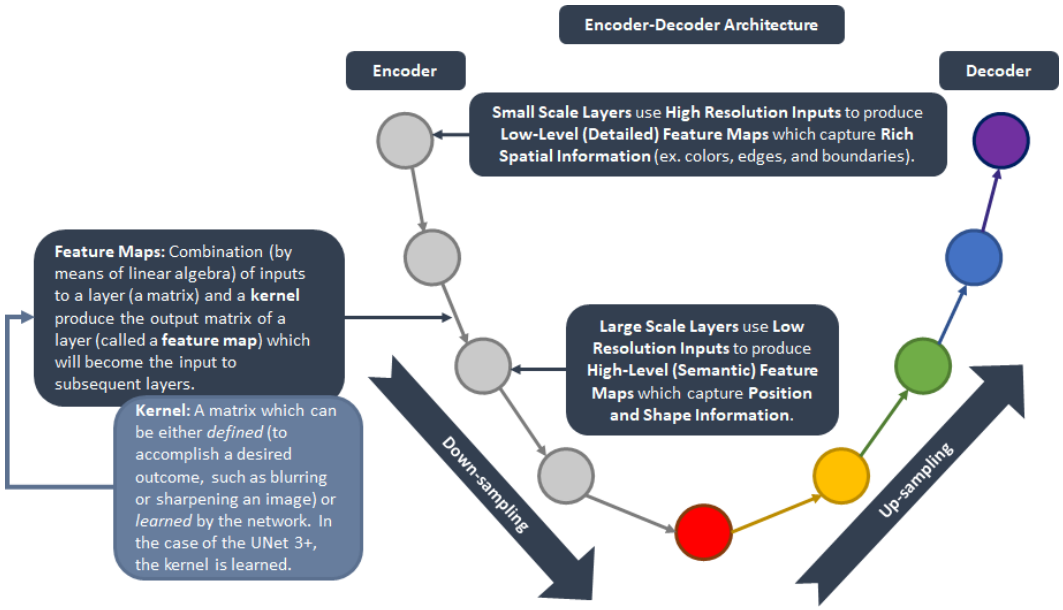


Figure 4.1: Breakdown of the UNet3+ Architecture: Layers, Scales, Feature Maps, & Kernels

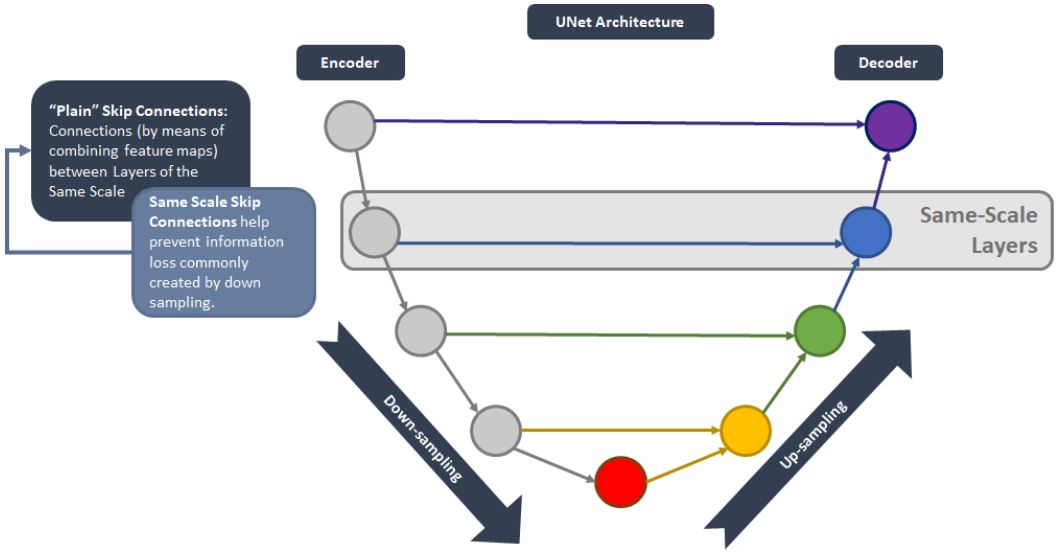


Figure 4.2: Breakdown of the UNet3+ Architecture: Same-Scale Skip Connections

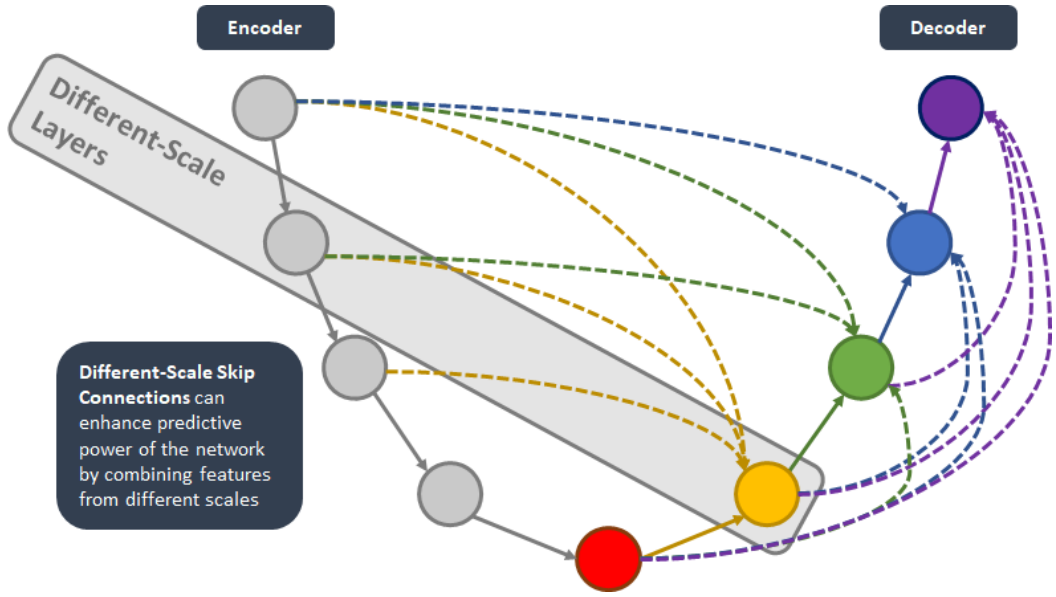


Figure 4.3: Breakdown of the UNet3+ Architecture: Different-Scale Skip Connections

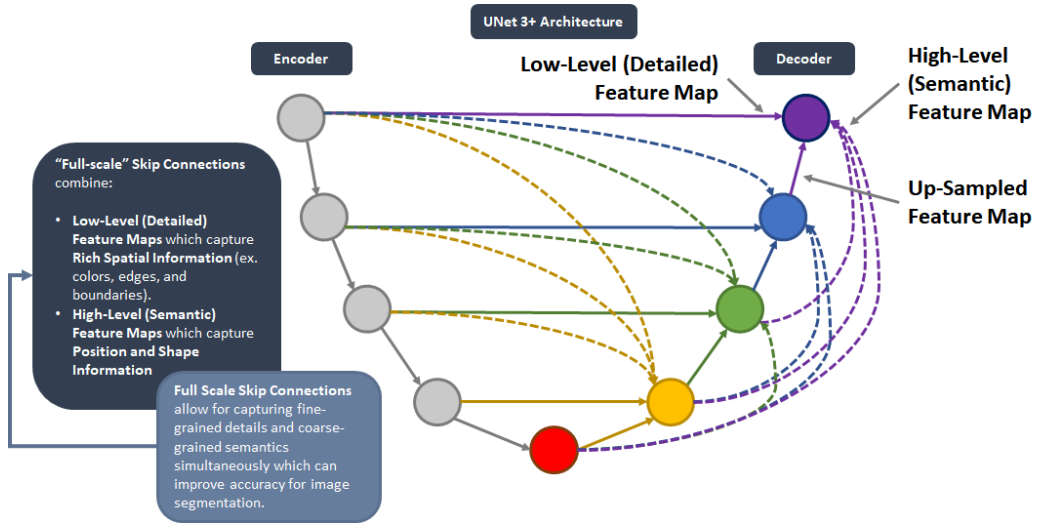


Figure 4.4: Breakdown of the UNet3+ Architecture: UNet3+ Architecture

to upper levels of the decoder. This allows the model to learn from multiple types of feature maps simultaneously. Figure 4.3 depicts same-scale skip connections.

The UNet3+ architecture is unique from previous UNet architectures (UNet and UNet++) because it uses full-scale skip connections, a combination of same-scale skip

connections and different-scale skip connections. Figure 4.4 depicts full-scale skip connections. This allows both the reduction of information loss and the learning of multiple types of feature maps (both fine-grained detail and coarse-grained semantics) simultaneously. The amount of interconnection provided by full-scale skip connections has allowed the UNet3+ to perform skillfully on semantic segmentation tasks (Huang et al. 2020).

4.1.1 Implementation Details

The UNet3+ implementation I used, originally described in the work of Huang et al. (2020), is the same for all models described in this work to offer as broad a foundation for comparison as possible. The network depth is 5, the number of filters for each down-sampling level is 32 for the first layer, 64 for the second layer, 128 for the third layer, 256 for the fourth layer, and 512 for the fifth layer. The number of filters after each full-scale skip connection is 32. The number of channels of full-scale aggregations is 160. The number of convolution layers per down-sampling level is 4. The number of convolution layers per up-sampling level, post full-scale concatenation is 4. The activation function used is ReLu and the output activation function used is sigmoid. Two dimensional max pooling is used as well as two dimensional up-sampling. Weights are pre-trained on ImageNet (Deng et al. 2009). Batch normalization is used as well as deep supervision. The network implimentation is depcted in Figure 4.5. More details concerning this implementation can be found in the work of Huang et al. (2020). Pertinent configuration details of the implementation can be found in Tables 4.1, 4.2, and 4.3.

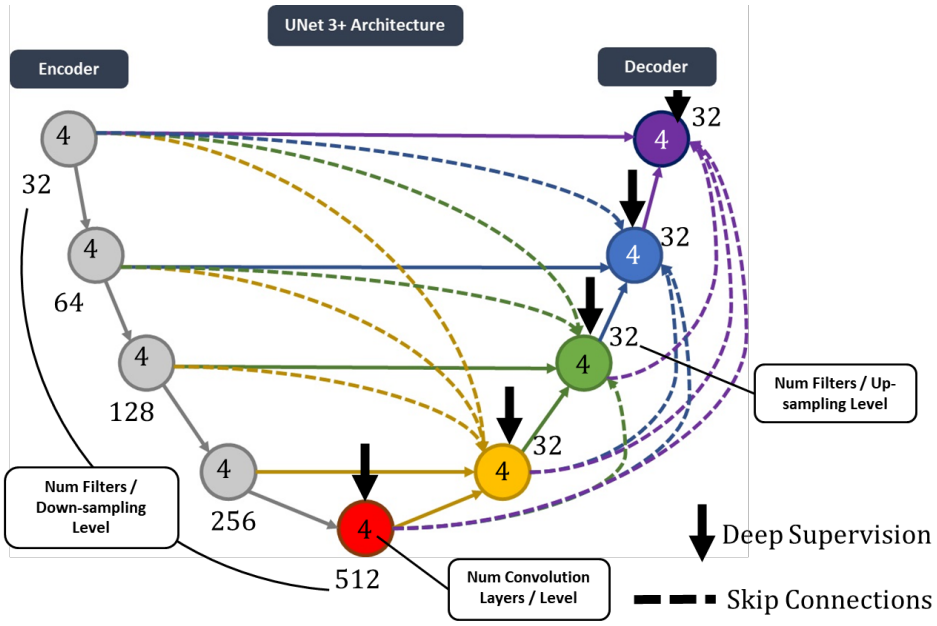


Figure 4.5: UNet3+ Network Diagram for All Models

Argument	*Setting*
filter_num_down	[32,64,128,256,512]
num_classes	1
stack_num_down	4
stack_num_up	4
filter_num_skip	auto
filter_num_aggregate	auto
activation	ReLU
output_activation	Sigmoid
batch_norm	TRUE
pool	TRUE
unpool	TRUE
weights	imagenet
deep_supervision	TRUE

Table 4.1: Unet3+ Model Configuration Details

Argument	*Setting*
learning_rate	0.0001
beta_1	0.9
beta_2	0.999
epsilon	None
decay	0
amsgrad	FALSE

Table 4.2: Adam Optimizer Configuration Details

Argument	*Setting*
fss_c	1
fss_mask_size	3

Table 4.3: Fractions Skill Score Loss Function Configuration Details

4.2 Fractions Skill Score (FSS)

During the task of semantic segmentation, a model is asked to receive inputs and produce a dense label, in which a prediction for the event of interest is generated for each pixel in an image. The challenge of this approach is that it can be position dependent. If the prediction is off by one pixel, it can be considered to be wrong depending on the loss function used to optimize the model. It was determined that prediction of the event in and around the pixel in which the event was discovered would be sufficient to meet operational needs owing to the belief that positioning fire management resources within proximity of potential fire would provide adequate coverage to take timely, mitigating action. Therefore, I selected the Fractions Skill Score (FSS) (Roberts and Lean 2008; Ebert-Uphoff et al. 2021; Justin et al. 2023, 2024) loss function.

FSS takes a neighborhood approach to model optimization, meaning that if the model prediction is close to the event but not exactly on it, the model is not penalized during the model training process. Fraction Skill Score ranges from 0 to 1, with 0 indicating no skill, 1 indicating perfect skill, and any number greater than .5 indicating reasonable skill.

FSS uses neighborhood averaging to compare the image containing prediction values to the image containing observed labels which allows it to consider both the label and its surrounding pixels when calculating the loss associated with those two images.

FSS for a neighborhood of size n is calculated using the equation:

$$FSS_{(n)} = 1 - \frac{MSE_{(n)}}{MSE_{(n)ref}} \tag{4.1}$$

.

Given a grid, observed and forecast fractions are calculated by computing the fraction of surrounding grid points for every grid point within a given neighborhood that exceed a given threshold. $MSE_{(n)}$ refers to the mean square error (MSE) for the observed and forecast fractions for a neighborhood of size n, calculated:

$$MSE_{(n)} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \left(f_{\text{forecast}}(i,j) - f_{\text{observed}}(i,j) \right)^2 , \tag{4.2}$$

where $f_{\text{forecast}}(i,j)$ is the fraction of grid cells within the neighborhood around the central point, (i,j), that exceed the threshold for the forecasts, $f_{\text{observed}}(i,j)$ is the fraction of grid cells within the neighborhood around the central point, (i,j), that exceed the threshold for the observations, and N is the total number of grid cells within the neighborhood.

$MSE_{(n)ref}$ refers to the highest MSE obtainable from the observed and forecast fractions, calculated:

$$MSE_{(n)ref} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(f_{\text{forecast}}(i,j) - \bar{f}_{\text{observed}}\right)^2, \tag{4.3}$$

where $f_{\text{forecast}}(i,j)$ is the fraction of grid cells within the neighborhood around the central point, (i,j), that exceed the threshold for the forecasts, $\bar{f}_{\text{observed}}$ is the average fraction of observed grid cells exceeding the threshold, and N is the total number of grid cells within the neighborhood.

4.3 Grouped K-Fold Cross Validation

Cross validation is a method for evaluating machine learning models whereby available model data are divided into three, independent subsets, one for training, one for validation, and one for testing. The process of defining each of the three subsets is repeated multiple times, called cross-validation rotations, resulting in different data being assigned to different subsets for each rotation. Through this process, the model's ability to generalize to unseen data can be estimated. In this work, grouped k-fold methodology (Xu and Frank 2004) was used for cross-validation. Grouped k-fold methodology is a type of k-fold cross-validation in which the folds created for the data represent set groups of samples. Grouping samples in this way allowed us to keep contiguous years together so as to avoid contaminating the validation or test data set with correlated data from the training set, known as temporal auto-correlation. Details of the cross validation approaches can be found in Tables 4.4 and 4.5.

For the work presented in chapters 5 and 6, the data set was divided into seven folds with three contiguous years each. Six cross-validation rotations were performed. The years 2018 to 2020, inclusive, the sixth fold, were held out as the testing data set for all rotations as they were the most current data in the data set and might allow the most representative view into how the model would generalize to future data.

	$Fold: 0$	$Fold: 1$	$Fold: 2$	$Fold: 3$	$Fold: 4$	$Fold: 5$	$Fold: 6$
	2000-2002	2003-2005	2006-2008	2009-2011	2012-2014	2015-2017	2018-2020
$Rotation: 0$	Train	Train	Train	Train	Train	Validation	Test
$Rotation: 1$	Train	Train	Train	Train	Validation	Train	Test
$Rotation: 2$	Train	Train	Train	Validation	Train	Train	Test
$Rotation: 3$	Train	Train	Validation	Train	Train	Train	Test
$Rotation: 4$	Train	Validation	Train	Train	Train	Train	Test
$Rotation: 5$	Validation	Train	Train	Train	Train	Train	Test

Table 4.4: Grouped K-Fold Cross-Validation for Chapters 5 & 6

	$Fold: 0$	$Fold: 1$	$Fold: 2$	$Fold: 3$	$Fold: 4$
	2000-2003	2004-2007	2008-2011	2012-2015	2016-2019
$Rotation: 0$	Train	Train	Train	Validation	Test
$Rotation: 1$	Train	Train	Validation	Train	Test
$Rotation: 2$	Train	Validation	Train	Train	Test
$Rotation: 3$	Validation	Train	Train	Train	Test

Table 4.5: Grouped K-Fold Cross-Validation for Chapters 7 & 8

The fold dedicated to the validation data set was rotated through the remaining six folds and the five folds not dedicated to testing or validation in each rotation were left for the training data set. The training dataset contained 15 years, the validation set contained 3 years, and the test set contained 3 years for all cross validation rotations. Details of the cross validation approach for chapters 5 and 6 can be found in Table 4.4.

For the work presented in Chapters 7 and 8, I switched from using gridMET as the input dataset to GEFS. This necessitated an adjustment to the time domain from 2000 to 2020, used in Chapters 5 and 6, to 2000 to 2019, as GEFS did not contain the year 2020. The cross validation approach necessarily had to be adjusted as well. The data set was divided into five folds with four contiguous years each. Four cross-validation rotations were performed. The years 2016 to 2019, inclusive, the fourth fold, were held out as testing data for all rotations as they were the most current data in the data set and might allow the most representative view into how the

model would generalize to future data. The fold dedicated to the validation dataset was rotated through the remaining four folds and the three folds not dedicated to testing or validation in each rotation were left for the training data set. The training dataset contained 12 years, the validation set contained 4 years, and the test set contained 4 years for all cross validation rotations. The cross validation approach for chapters 7 and 8 is depicted in Table 4.5.

I assume that by dividing the data between training, validation, and testing using this method that independence is maintained between the three data sets. I make this assumption with fire ecology in mind, and the understanding that wildfire trends can span not just seasons, years, and decades, but millennia (National Wildfire Coordinating Group 2022b; Auburn University 2022).

4.4 Critical Success Index (CSI)

Critical Success Index (CSI) is a performance metric used to evaluate model predictions within the National Weather Service (Schaefer 1990). By using a metric familiar to fire weather forecasters, barriers to understanding and adoption are lowered and a basis of comparison is established between this work and other works familiar to forecasters. The CSI metric relies on two sub-metrics, the Probability of Detection (POD), which is the percent of events predicted, and the False Alarm Rate (FAR), which is a ratio of false alarms to predicted events. At the heart of the CSI is the 2x2 contingency matrix, capturing hits, misses, false alarms, and correct rejections.

As the models explored in this work were trained using a neighborhood loss function, a neighborhood approach to measuring model performance was also necessary. I measured CSI for neighborhoods of size 40 km, 80 km, and 120 km, depicted in Figure 4.6, with 40 km representing a hit directly on the label pixel itself, 80 km representing

Description	*Variable*	*Event*
Hits (True Positive (TP))	X	Forecast: Yes, Observed: Yes
Misses (False Negative (FN))	Y	Forecast: No, Observed: Yes
False Alarms (False Positive (FP))	Z	Forecast: Yes, Observed: No
Correct Rejections (True Negative (TN))	W	Forecast: No, Observed: No

Table 4.6: Variables X, Y, Z, & W (Schaefer 1990)

Description	*Variable*	*Equation*
Total Number of Positive Forecasts	P	$P = X + Z$
Total Number of Events	E	$E = X + Y$
Total Number of Cases	T	$T = X + Y + Z + W$
Frequency of Event	F	$F = E/T$

Table 4.7: Variables P, E, T, & F (Schaefer 1990)

Description	*Variable*	*Equation*
Probability of Detection	POD	$POD = X/(X + Y)$ or $POD = X/E$
False Alarm Rate	FAR	$FAR = Z/(X + Z)$ or $FAR = Z/P$
Success Ratio	SR	$SR = X/(X + Z)$ or $SR = X/P$ or $SR = 1 - FAR$
Critical Success Index	CSI	$CSI = [(POD)^{-1} + (1 - FAR)^{-1} - 1]^{-1}$

Table 4.8: Variables POD, FAR, SR, & CSI (Schaefer 1990)

a hit within one neighboring pixel of the label, and 120 km representing a hit within two neighboring pixels of the label. I chose to stop at a 120 km neighborhood because it was as large as was operationally useful for my subject matter experts. Details of the CSI calculation can be explored in Tables 4.6, 4.7, and 4.8.

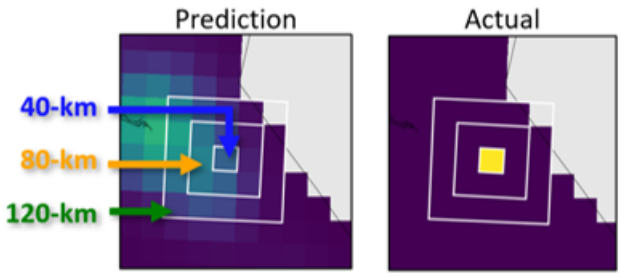


Figure 4.6: Neighborhoods for Measuring Performance

Chapter 5

Fire Occurrence Prediction using gridMET

In Chapters 5 and 6, we demonstrate performance improvement relative to climatological methods currently used operationally (the SPC Probability of Wildfire Climatology) and do so by employing minimally pre-processed, publicly available data (gridMET and FPA-FOD) in combination with an effective machine learning model (UNet3+). Barriers to adoption are addressed by working with data sets and performance metrics already familiar to both forecasters and the fire management community. We identify areas where fire occurrence is likely, based on environmental factors, without controlling for cause, size, region, or season. This model is called the 'All Fires' model and we offer it in comparison to a model limited to a focus on large lightning fires, called the 'Large Lightning' model. Model performance is analyzed by region, season, fire size, and fire cause to help identify areas where model performance can be improved and future work is proposed to address these areas. In a Chapter 6, we analyze lessons learned from applying deep learning to fire occurrence prediction.

Model	*Time*	*Space*	*Fires Included in Labels*
Large Lightning	2000-2020	CONUS	Cause='Natural', Fire Size>=1000 Acres
All Fires	2000-2020	CONUS	All Fires (regardless of size, cause, or region)

Table 5.1: Data Domain for Models by Chapter

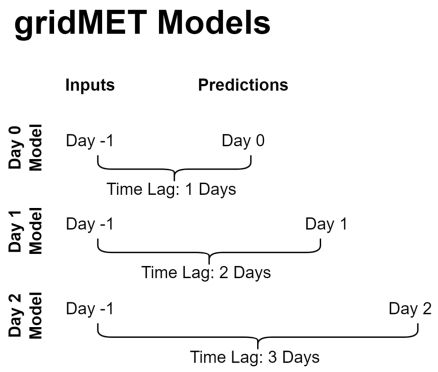


Figure 5.1: Time Lag Between Inputs and Predictions for gridMET Models

For all models, day 0 to day 10, regardless of which day is being predicted, the input data is from day -1, the day prior to the current day, to account for the gridMET data update schedule. A time lag between input data and label data results that increases as the day predicted proceeds from day 0 to day 10.

5.1 General Performance

In this work, model performance is contextualized using multiple approaches. First, general model performance is compared using CSI. CSI is generated by comparing the predictions of each model to the model's observation values for the test dataset. The test dataset uses observed 'Pixel' labels generated using either all fire occurrence instances from FPA-FOD, in the case of the 'All Fires' model, or only fire occurrence instances with a cause of 'Natural' and a final fire size greater than 1000 acres, in the case of the 'Large Lightning' model.

Second, general model performance CSI is stratified by region and season to enable us to further describe where and when each model performed best and to lay bare potential opportunities for improvement. Third, the models are compared using CSI

created with model predictions and observed values from only the areas of CONUS and days of year subject to large lightning fires to estimate how each model performs on the same subset of labels. Finally, the two models are compared on how they perform on the largest lightning fires in the test data set relative to climatological performance using the probability of fire generated from each model for the day of discovery and days proceeding discovery and those grid cells representing the 40 km, 80 km, and 120 km neighborhoods.

5.1.1 Individual Performance Plots

Individual performance plots are available in Figures 5.2 and 5.3. We see in the individual performance plots that performance differed consistently between the 40 km neighborhood, the 80 km neighborhood, and the 120 km neighborhood with the 40 km neighborhood achieving the lowest CSI values and the 120 km neighborhood achieving the highest CSI values for this experiment. This separation was consistent from day 0 to day 10 for both the 'All Fires' model and the 'Large Lightning' model. The 'All Fires' model achieved higher CSI values than the 'Large Lightning' model for all days 0 to 10, for all neighborhoods, 40 km, 80 km, and 120 km.

5.1.2 Individual Reliability Diagrams

Individual reliability diagrams are available in Figures 5.4 and 5.5. The 'All Fires' model produced maximum probability predictions between .3 and .4 and this was true for all models, day 0 to day 10. The 'Large Lightning' model produced maximum probability predictions between .1 and .2 and this was true for all models, day 0 to day 10. For the 'All Fires' model the most reliable neighborhood measurement is the 80 km neighborhood, which achieves reliability for early predictions before moving

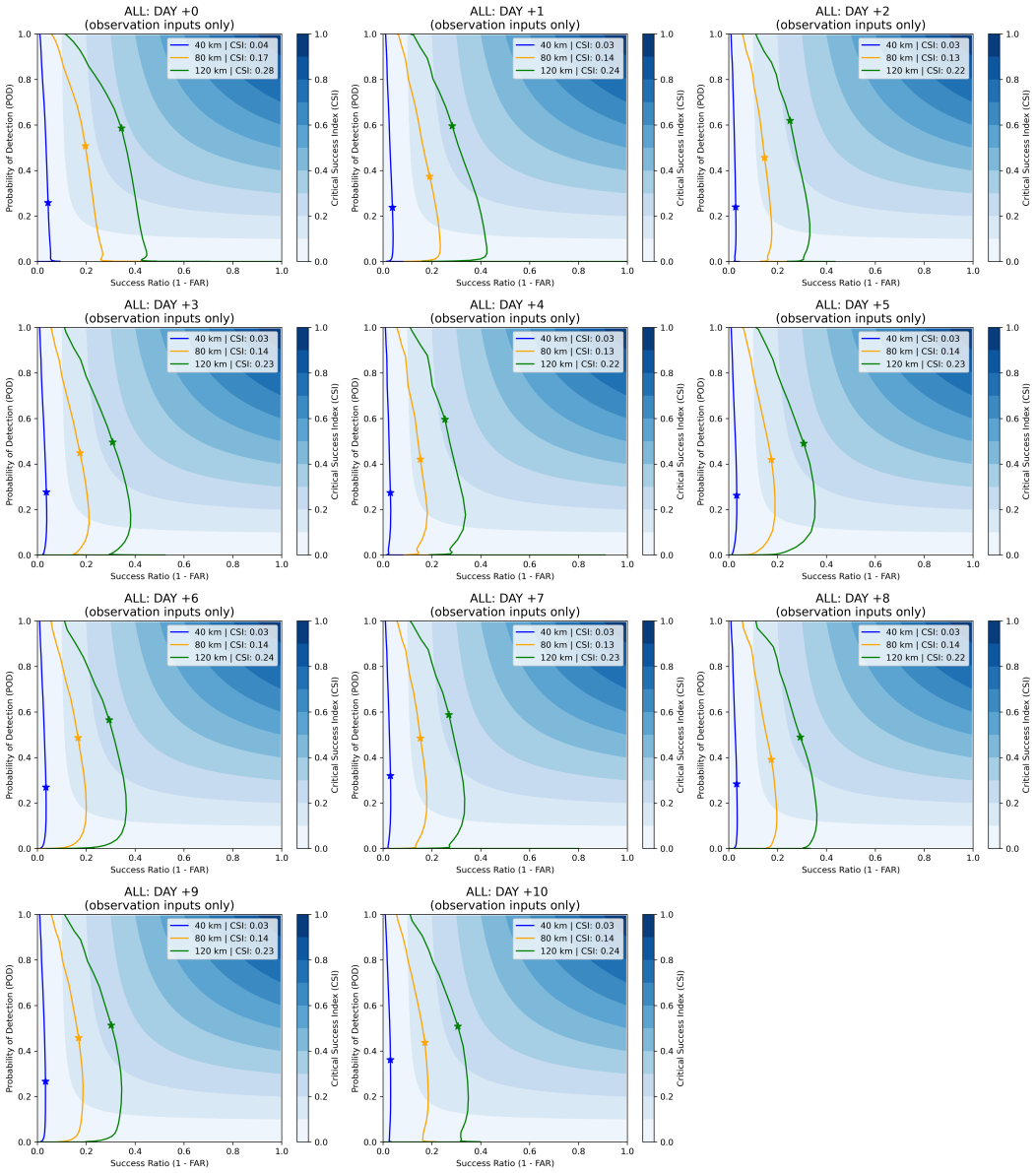


Figure 5.2: Performance Plots for Day 0 - 10 for 'All Fires' Model. For more information on how to interpret performance plots see Roebber (2009).

into over-forecasting for later predictions. Over-forecasting is a common trend for the 40 km neighborhood and under-forecasting is a common trend for the 120 km neighborhoods. This was the case for all models, day 0 to day 10.

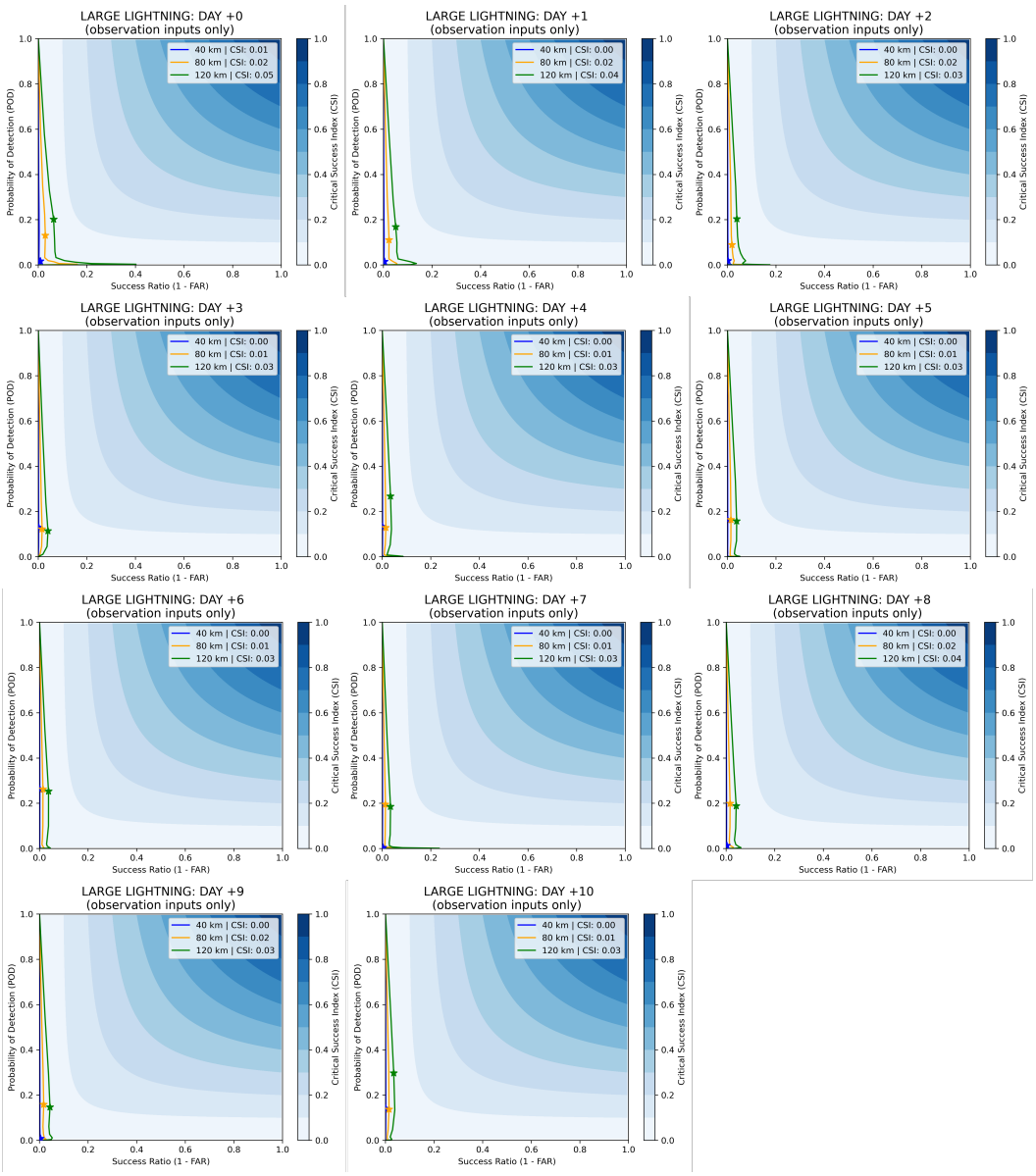


Figure 5.3: Performance Plots for Day 0 - 10 for 'Large Lightning' Model. For more information on interpreting performance plots see Roebber (2009).

For the 'Large Lightning' model the most reliable neighborhood measurement is the 80 km neighborhood for the day 0 model. After the day 0 model, the 120 km neighborhood becomes the most reliable measurement and remains the most reliable measurement out to day 10. Over-forecasting is common for the 40 km neighborhood.

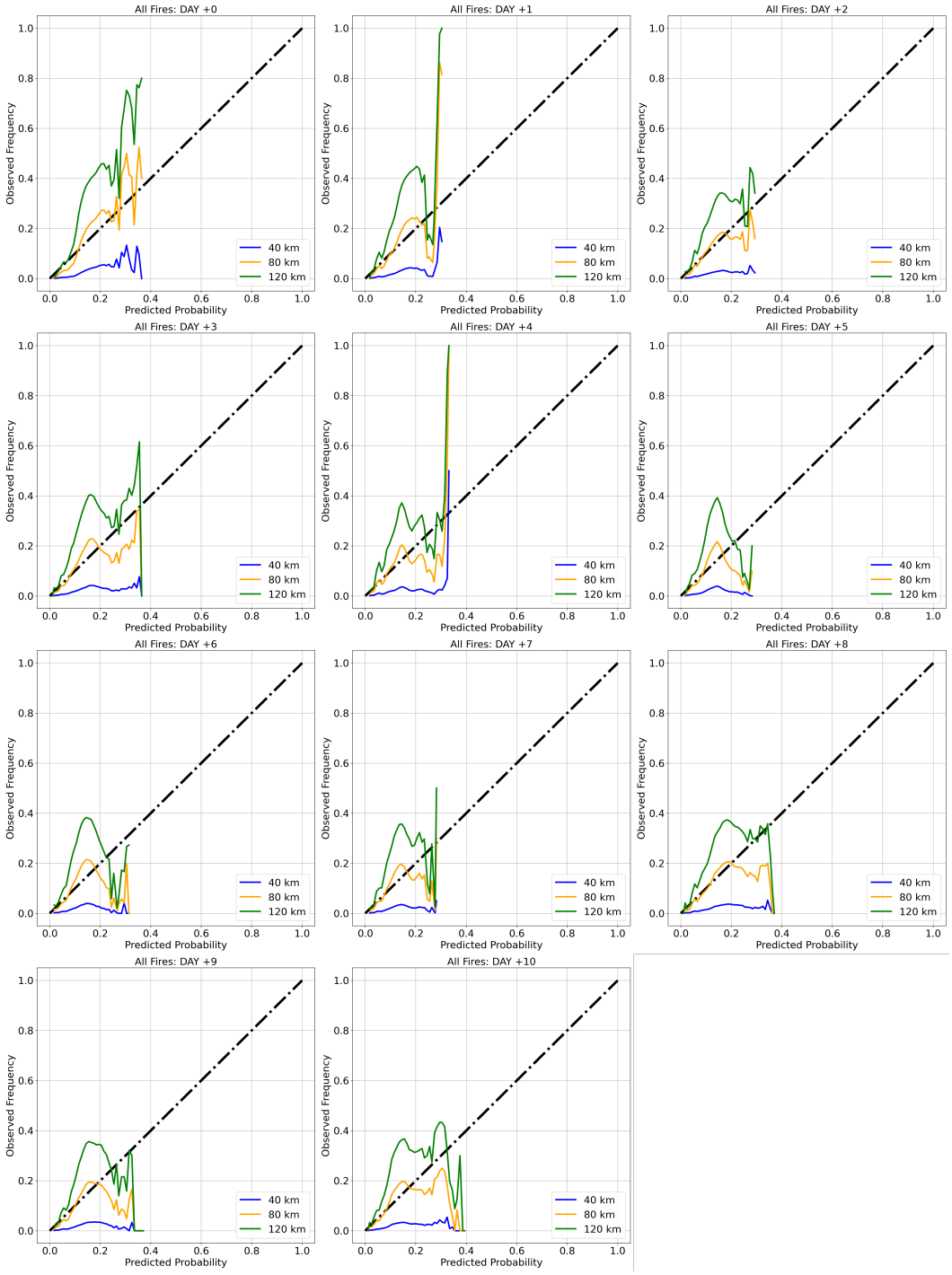


Figure 5.4: Reliability Plots for Day 0 - 10 for 'All Fires' Model. For more information on interpreting reliability plots see Bröcker and Smith (2007).

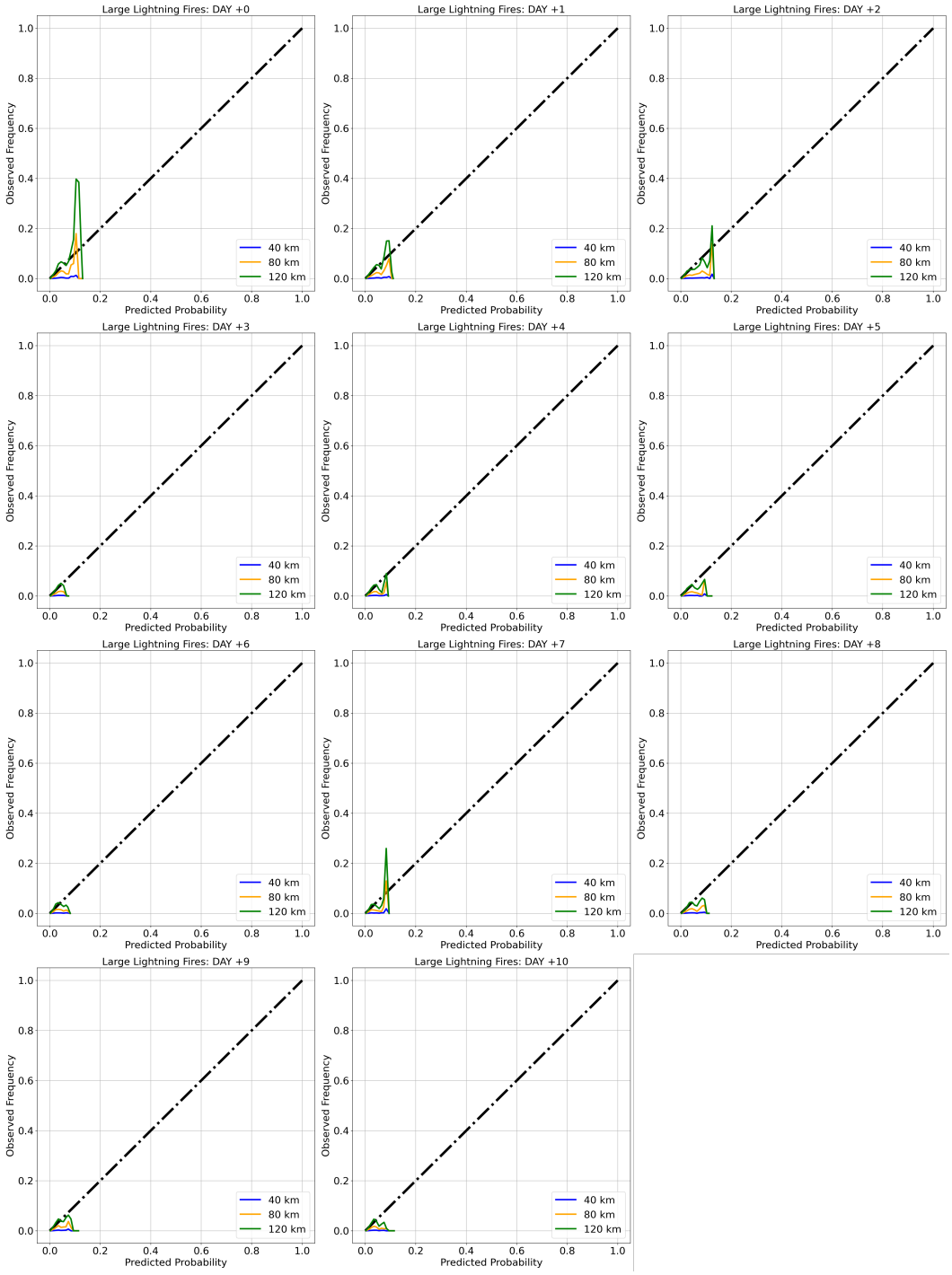


Figure 5.5: Reliability Plots for Day 0 - 10 for 'Large Lightning' Model. For more information on interpreting reliability plots see Bröcker and Smith (2007).

5.1.3 Max CSI Performance

To compare general model performance using CSI, the max CSI value is taken from the performance plots depicted in Figures 5.2 and 5.3 where it is indicated with a star and described in the legend for each neighborhood. Max CSI is plotted for all days 0 to 10 with day 0 representing the current day, day 1 representing the next day, and so on. Model predictions produced by each cross-validation rotation are averaged together to produce the predictions used to calculate CSI.

There are three important trends that can be observed in Figures 5.6 and 5.7. First, model performance did not demonstrate a dramatic downward trend as predictions moved away from the current day in time. In our modeling strategy, a model is created for each day to be predicted, day 0 to 10 (11 models). This illustrates that the predictive power of the model did not decrease as the time lag between inputs and labels increased from 0 to 10 days and suggests that each model could continue to demonstrate similar performance for days greater than 10 days, though we have not yet tested this.

Second, model performance differed by neighborhood size. The 120 km neighborhood produced higher CSI values than both the 80 km neighborhood and the 40 km neighborhood for this experiment. The relatively higher performance demonstrated by the 120 km neighborhood is likely explained by the fact that it is larger than the other two neighborhoods, and, as a result of its larger size, allows the model more opportunities to be 'right.' Neighborhood Max CSI performance is depicted in Figure 5.8.

Third, the 'All Fires' model consistently produced higher CSI values than the 'Large Lightning' model for all days predicted and for all neighborhoods calculated, as depicted in Figure 5.8. The higher CSI values observed from the 'All Fires' model may be associated with probability predictions that demonstrate less localization.

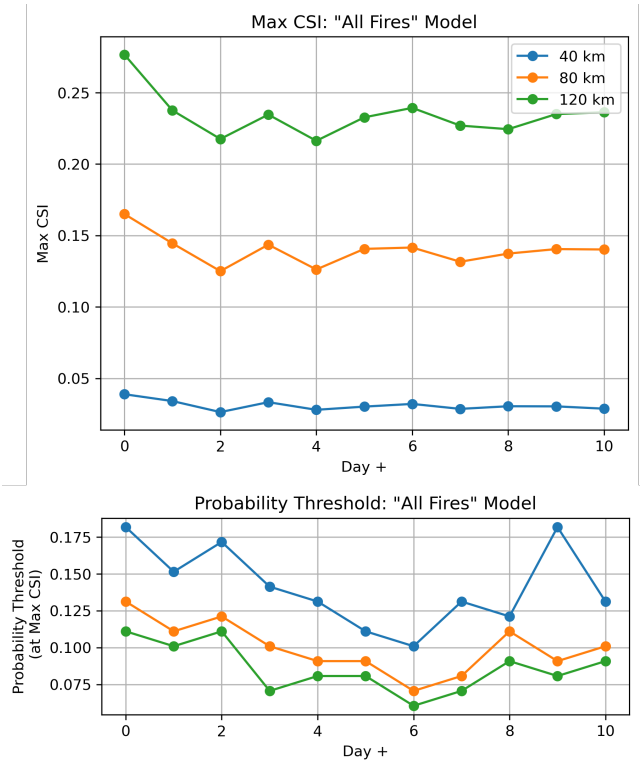


Figure 5.6: "All Fires" Model: Max CSI Values & Probability Thresholds. The blue line describes values measured at 40km, the orange line describes values measured at 80km, and the green line describes values measured at 120km. Max CSI values are drawn from the Individual Performance Plots contained in Figure 5.2. The Probability Threshold is the probability cutoff from the Individual Performance Plot at which Max CSI is achieved.

This lack of localization in predicted probabilities may stem from both the higher density of fire occurrence in more areas of CONUS (as present in the All Fires labels) and with the use of data that includes consideration for fuel and topography (both of which change at a slower rate than weather data). Non-zero probabilities covering larger areas of CONUS produced by the 'All Fires' model can be observed in all three of our case studies, Figures 5.21, 5.22, and 5.23.

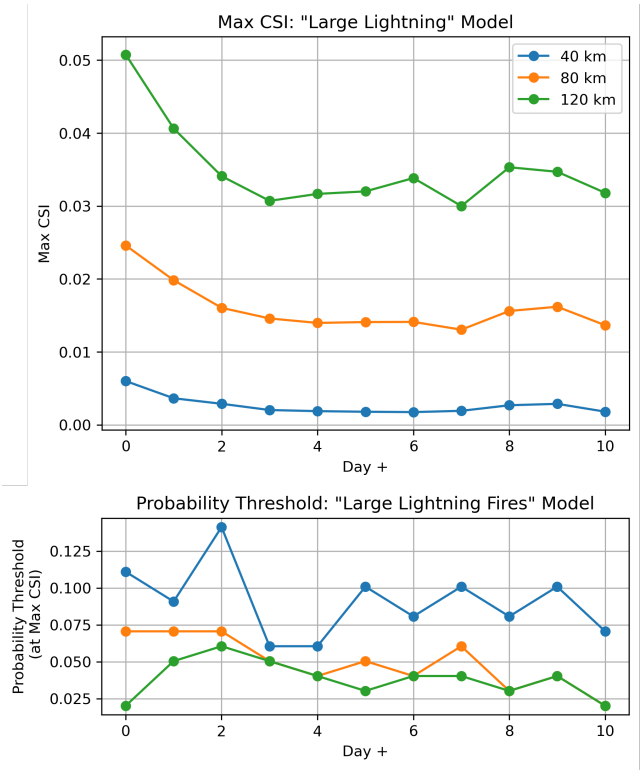


Figure 5.7: "Large Lightning" Model: Max CSI Values & Probability Thresholds. Max CSI values are drawn from the Individual Performance Plots contained in Figure 5.3. The Probability Threshold is the probability cutoff from the Individual Performance Plot at which Max CSI is achieved.

5.2 Regional Performance

To break CSI results down by region, we used the Forest Service Regions created by the United States Forest Service (United States Forest Service 2023), as depicted in Figure 5.9, to divide predicted values and observed values into regions and then calculated CSI for each region.

Figures 5.10 and 5.11 depict the regional stratification of max CSI values for each model. From it, two themes can be observed. First, max CSI performance differed by region. Some regions had higher max CSI performance than other regions, as in the

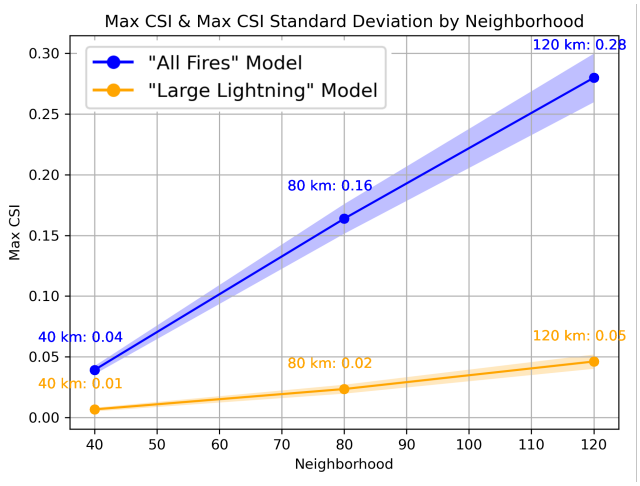


Figure 5.8: Max CSI Values by Neighborhood. The shaded region represents +/- the standard deviation produced by the cross-validation rotations.

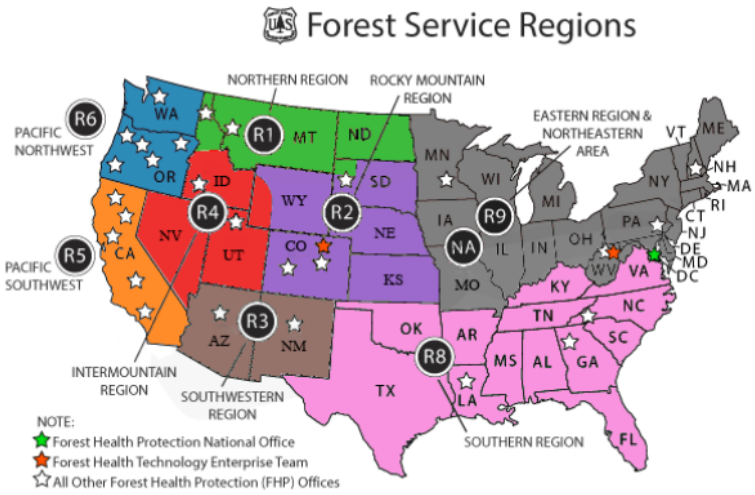


Figure 5.9: USFS Fire Regions used to stratify model performance. Figure adapted from United States Forest Service (2023).

case of the Pacific Southwest region for all neighborhoods for the 'All Fires' model and the case of the Intermountain Region and the Southwestern Region for the 'Large Lightning' model for 120 km neighborhoods. Some regions had lower max CSI performance than other regions, such as with the Eastern Region and Northeastern Area,

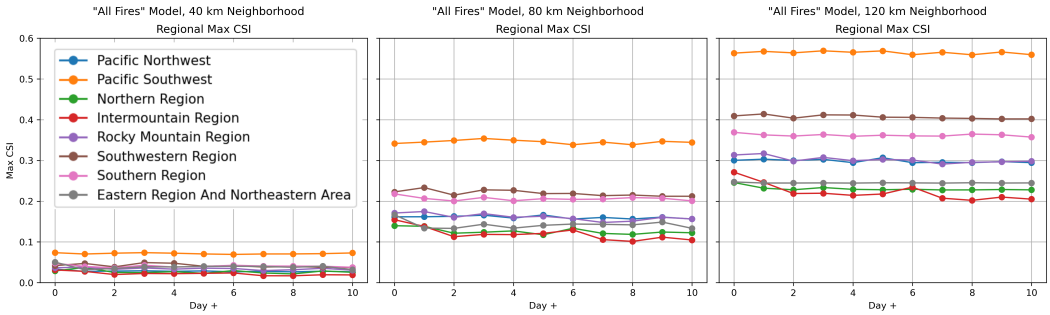


Figure 5.10: Max CSI Values by Region for the 'All Fires' Model

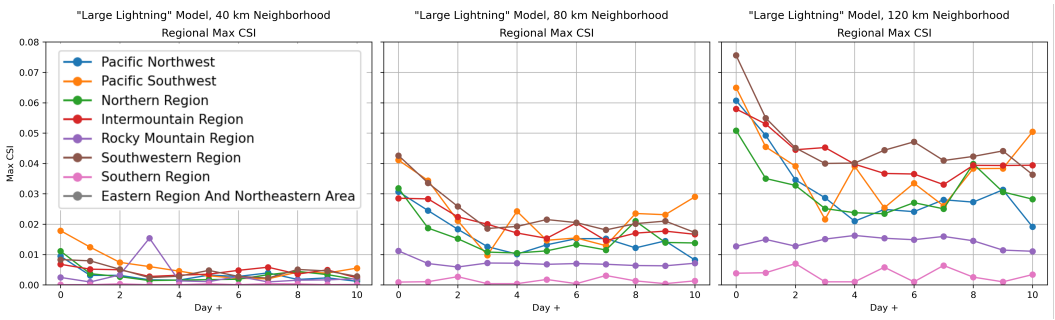


Figure 5.11: Max CSI Values by Region for the 'Large Lightning' Model

the Northern Region, and the Intermountain Region for the 120 km neighborhood for the 'All Fires' model and with the Rocky Mountain Region and Southern Region for the 'Large Lightning' model for the 80km and 120km neighborhoods.

When one region performed better than other regions, two possible interpretations are that 1) the region of interest with higher performance contained more of the fire occurrences or 2) the inputs selected for the model were more representative of fire behaviour in that region allowing the model to predict fire occurrence more accurately. Most of the cases cited herein are subject to the first interpretation. For regions where both the number of fire occurrences is high and the model's ability to predict fire occurrence is accurate, such as the Pacific Southwest, the value of POD within the CSI calculation is driven up. Since POD counts more than FAR in the calculation of CSI, these regions achieved higher values of CSI.

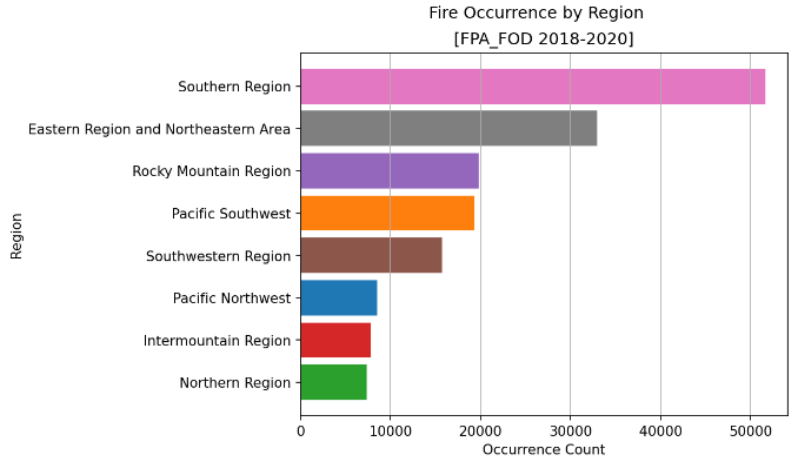


Figure 5.12: All Fires: FPA-FOD Fire Occurrence Count By Region

Fire occurrence counts by region, which applies to the 'All Fires' model, are depicted in Figure 5.12. Large lightning fire occurrence counts by region, which applies to the 'Large Lightning' model, are depicted in Figure 5.13. The 'All Fires' model included all fires, regardless of size, cause, or region. We can see in Figure 5.12 that the Southern region is home to the majority of fire occurrences, regardless of cause. We can see in Figure 5.13 that large lightning fires are more prevalent in the Intermountain and Southwestern Regions but are relatively small contributors to the larger fire occurrence counts when cause and size constraints are removed.

Second, max CSI performance is higher for all regions, for all neighborhoods, for the 'All Fires' model than for the 'Large Lightning' model for this experiment. Figure 5.14 depicts average max CSI performance for all regions for all neighborhoods for both models.

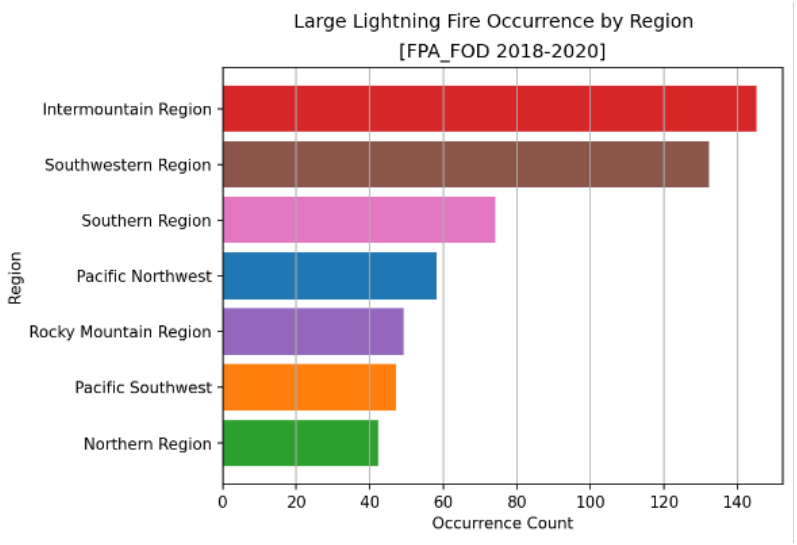


Figure 5.13: Large Lightning: FPA-FOD Fire Occurrence Count By Region

Season	*Months*
Winter	December, January, February
Spring	March, April, May
Summer	June, July, August
Fall	September, October, November

Table 5.2: Month to Season Mapping

5.3 Seasonal Performance

To break CSI performance down by season, we divided the months of the year into quarters, depicted in Table 5.2.

Figures 5.15 and 5.16 depicts max CSI performance stratified by season for each neighborhood for each model. In them we can see a repeat of themes seen in the regional stratification of performance. First, max CSI performance differed by season. Some seasons performed better than others, as with the spring and summer seasons for all neighborhoods for the 'All Fires' model and with the summer season for all

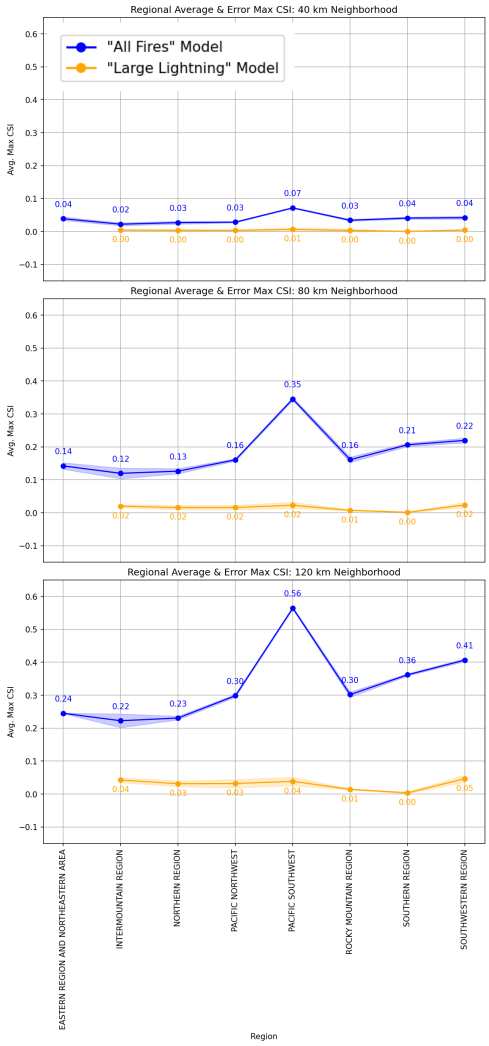


Figure 5.14: Average & Error Max CSI by Region (Aggregated across Day Dimension). The shaded region represents +/- the standard deviation produced by the cross-validation rotations.

neighborhoods for the 'Large Lightning' model. With seasonal performance, when the performance differed between seasons, the season with the higher performance contained more of the fire occurrence data, as was the case for all instances described herein.

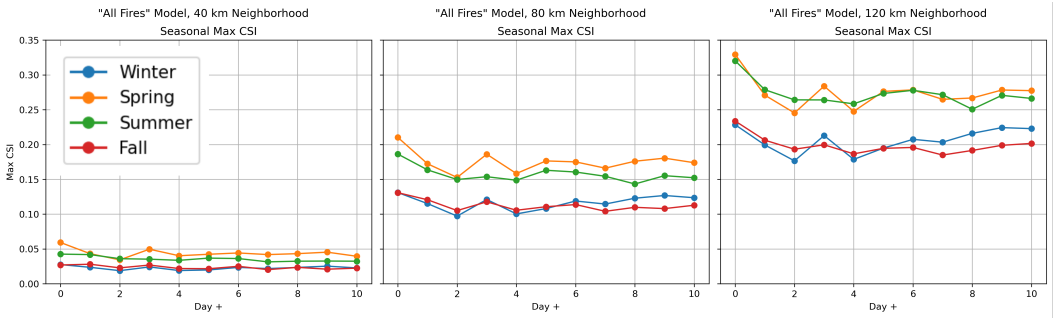


Figure 5.15: Max CSI Values by Season for the 'All Fires' Model

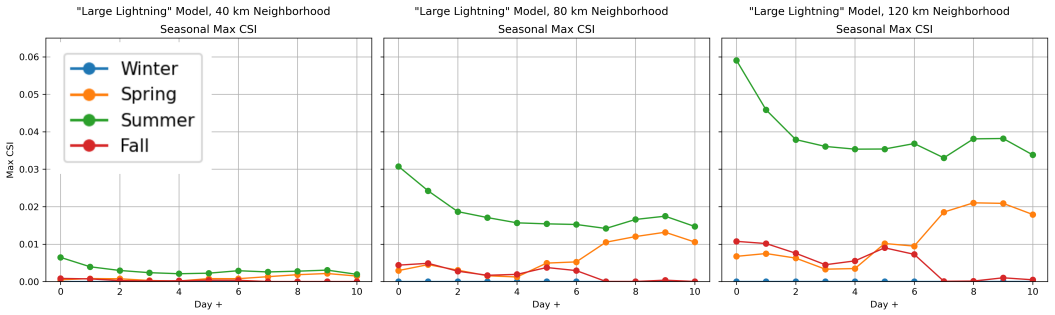


Figure 5.16: Max CSI Values by Season for the 'Large Lightning' Model

Fire occurrence counts by season, which applies to the 'All Fires' model, can be seen in Figure 5.17 and large lightning fire occurrence counts by season, which applies to the 'Large Lightning' model, can be seen in Figure 5.18.

Second, average max CSI performance was greater for all seasons for all neighborhoods for the 'All Fires' model than for the 'Large Lightning' model, as depicted in Figure 5.19.

5.4 Large Lightning Performance

To compare the 'All Fires' model to the 'Large Lightning' model more directly, we took the predictions generated by each model for only the subset of days and grid cells subject to large lightning fires and calculated CSI performance for each model using

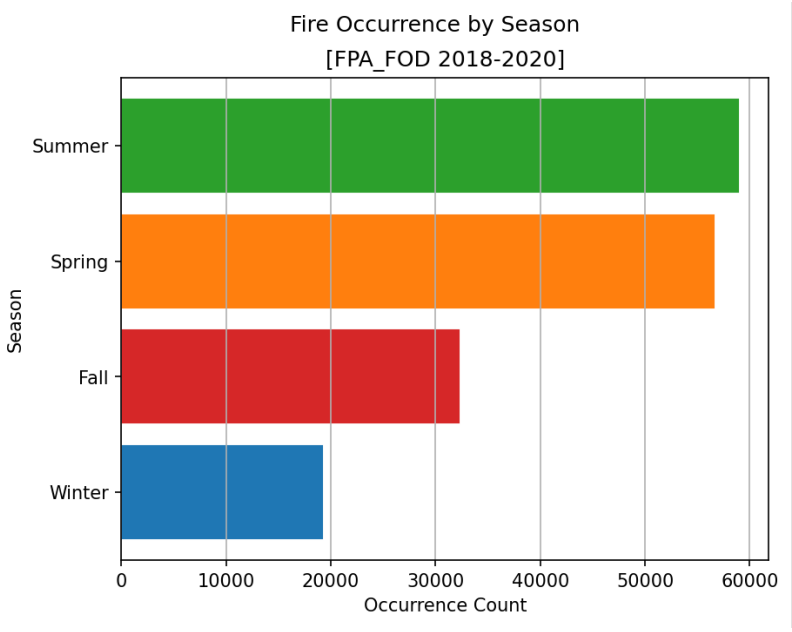


Figure 5.17: All Fires: FPA-FOD Fire Occurrence Count By Season

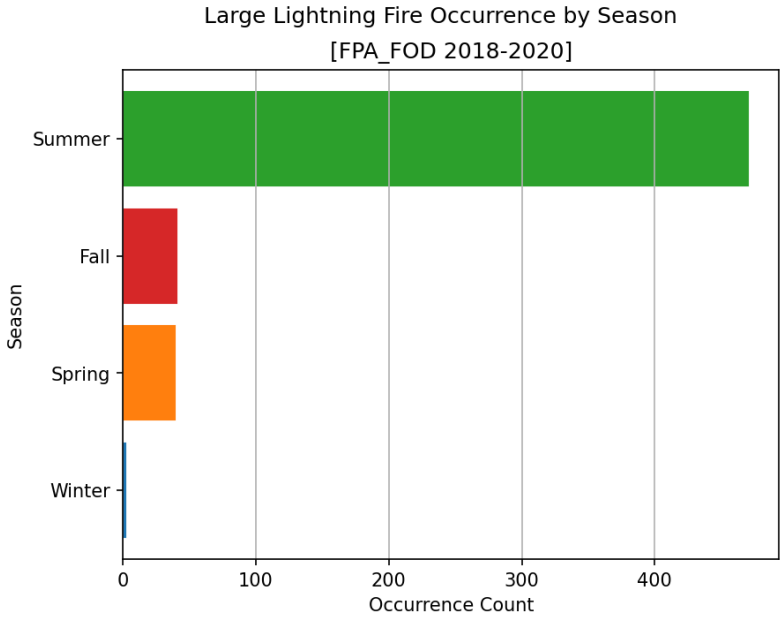


Figure 5.18: Large Lightning: FPA-FOD Fire Occurrence Count By Season

this subset of predictions. Our goal in approaching the comparison in this way was to identify the model that would have signalled the strongest for the event of interest

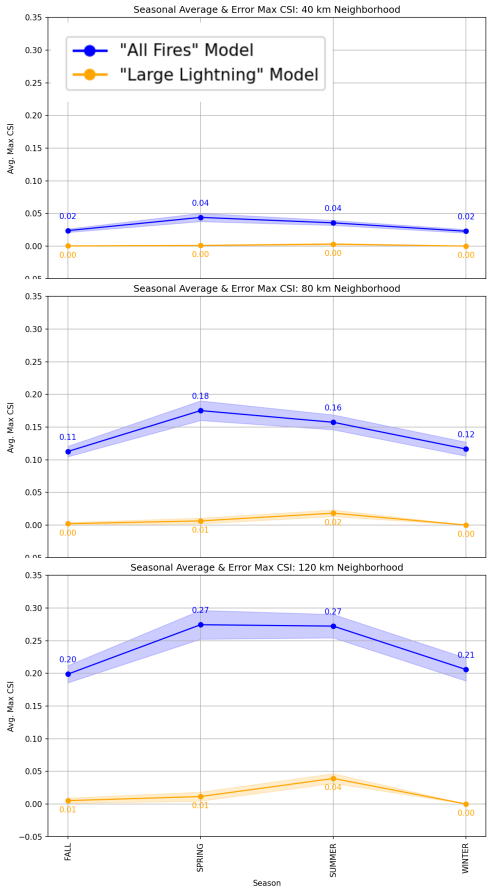


Figure 5.19: Average & Error Max CSI by Season (Aggregated across Day Dimension). The shaded region represents +/- the standard deviation produced by the cross-validation rotations.

and with what lead time. Figure 5.20 depicts the resulting Max CSI values for each model. As with general performance, model performance differed by neighborhood, with the 120 km neighborhood outperforming the 40 km and 80 km neighborhoods. Model performance was consistent from day 0 to day 10 for both models. For all days, for all neighborhoods, the 'All Fires' model produced higher CSI values the 'Large Lightning' model.

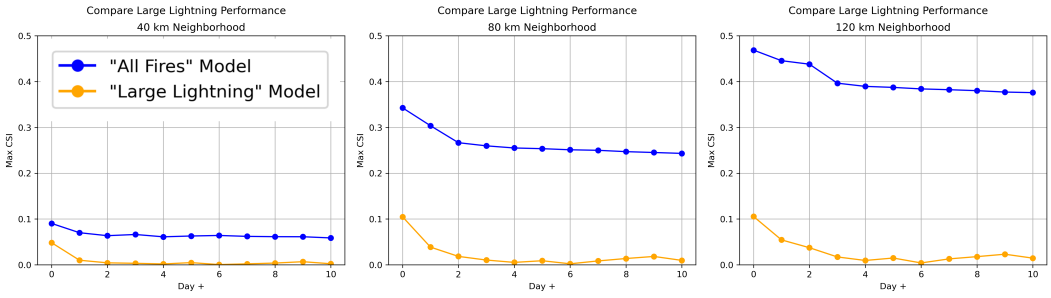


Figure 5.20: Max CSI Values for Large Lightning Labels for 'All Fires' Model and 'Large Lightning' Model

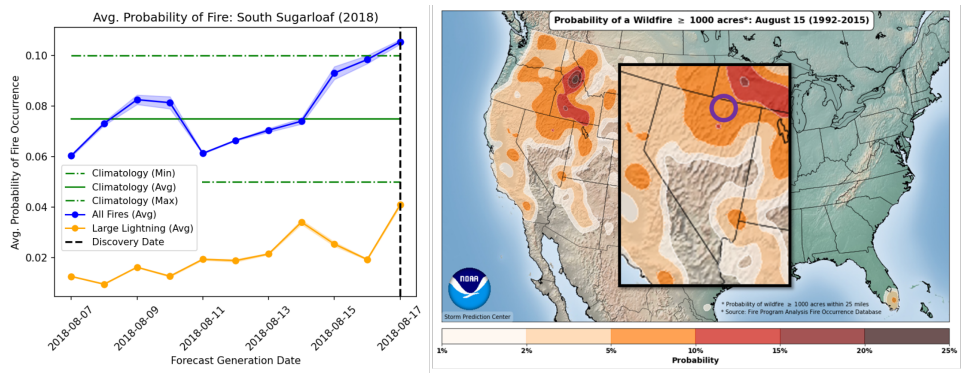
5.5 Case Studies

For case studies, we selected the largest lightning-caused fires from each year of the test data set that included 2018, 2019, and 2020. We compared the average (averaged across all cross validation rotations) probability of fire produced by each model to the probability of wildfire produced by the Storm Prediction Center (SPC) Wildfire Climatology (Nauslar 2020b). SPC's Climatology uses an 80 km grid and reports probability of wildfire within 25 miles, or approximately 40 km, and sources FPA-FOD years 1992 to 2015. SPC's climatology produces probabilities between 1% and 25%. The models in this work used a 40 km grid and reported probability of wildfire within 40 km, 80 km, and 120 km and sourced FPA-FOD years 2000 to 2020. A 'hit', when comparing the models from this work to the climatology, was anytime the model average probability outperformed the climatology average probability. For all case studies, for both models, performance was consistent across neighborhoods.

We also included case studies that described the behavior of the 'All Fires' model and the 'Large Lightning' model on the day of the year most common for human-caused fire, July 4th, and for lightning-caused fire, July 22nd.

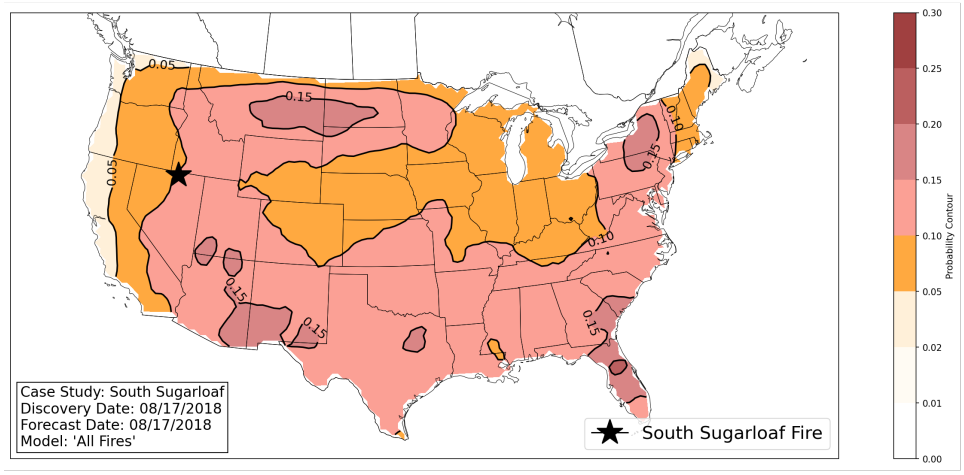
5.5.1 South Sugarloaf Fire: 2018 Largest Lightning Fire

The South Sugarloaf Fire, caused by lightning, was discovered on 17 August 2018 in the Humboldt National Forest in northern Nevada and burned approximately 232,906 acres fueled by grass, brush, and juniper (State of Nevada 2018). Figure 5.21 describes the probabilities depicted by the SPC Climatology, the 'All Fires' model, and the 'Large Lightning' model for the South Sugarloaf Fire. The SPC Climatology depicted a probability between .05 and .1, with an average value of .075, for the ten days prior to the discovery date for the location of discovery for the South Sugarloaf Fire. The 'All Fires' model predicted greater average probability of fire than the average probability produced by the climatology starting three days prior to the discovery date and greater than the max probability produced by the climatology on the discovery date for the location of discovery. The 'Large Lightning' model did not produce a probability prediction greater than the average climatology probability, with probabilities lower than the minimum climatology probability for the ten days prior to the discovery date.

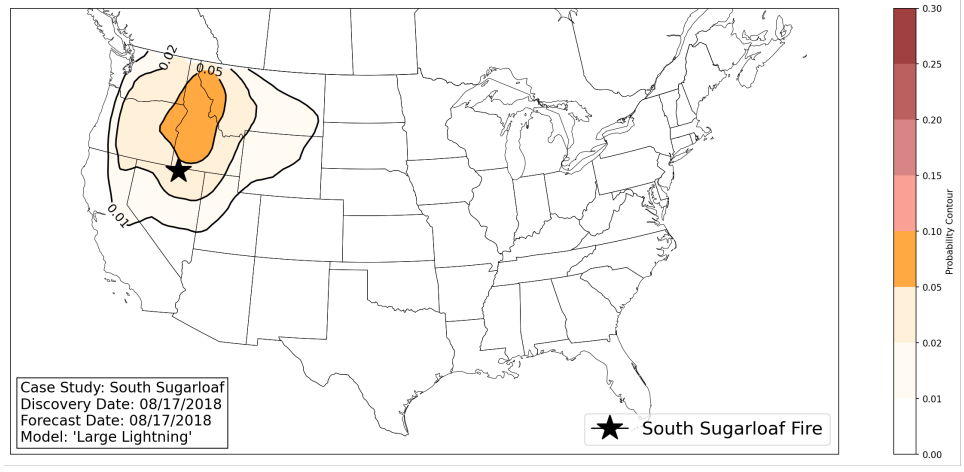


a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology

b) SPC Climatology Probability of Fire



c) "All Fires" Model Probability of Fire

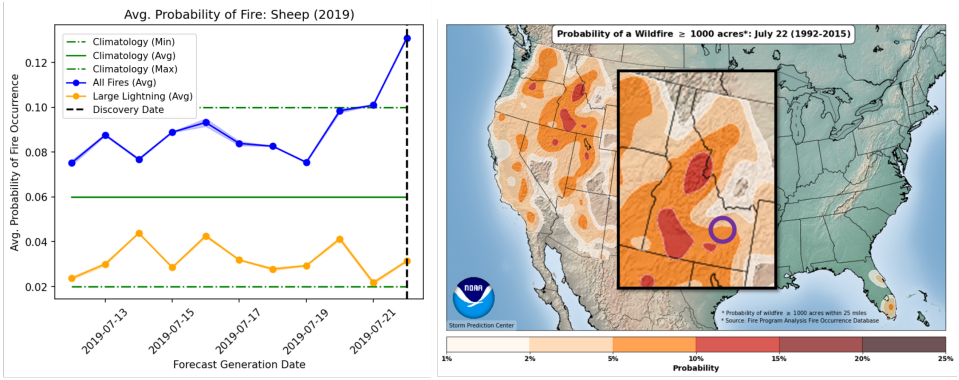


d) "Large Lightning" Model Probability of Fire

Figure 5.21: Probability of Fire: South Sugarloaf Fire, 2018 Largest Lightning Fire

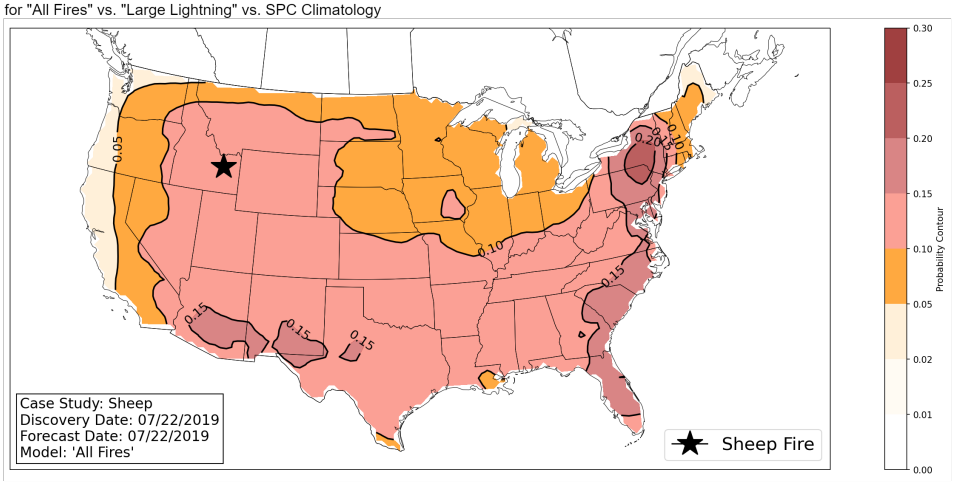
5.5.2 Sheep Fire: 2019 Largest Lightning Fire

The Sheep Fire, started by lightning, was discovered on 22 July 2019, in Idaho near Idaho Falls, and burned approximately 112,106 acres (Idaho National Laboratory 2019). Figure 5.22 describes the probabilities depicted by the SPC Climatology, the 'All Fires' model, and the 'Large Lightning' model for the Sheep Fire. SPC Climatology depicted a probability between .025 and .1, with an average of .0625, for the ten days prior to the discovery date for the location of discovery for the Sheep Fire. The 'All Fires' model predicted greater probability of wildfire than the average climatology for all ten days prior to the discovery date and greater average probability than the max climatology one day prior to the discovery date for the location of discovery. The 'Large Lightning' model, while it did not produce probabilities greater than the average climatology probability, it did produce average probabilities greater than the climatology minimum for all ten days prior to the discovery date.

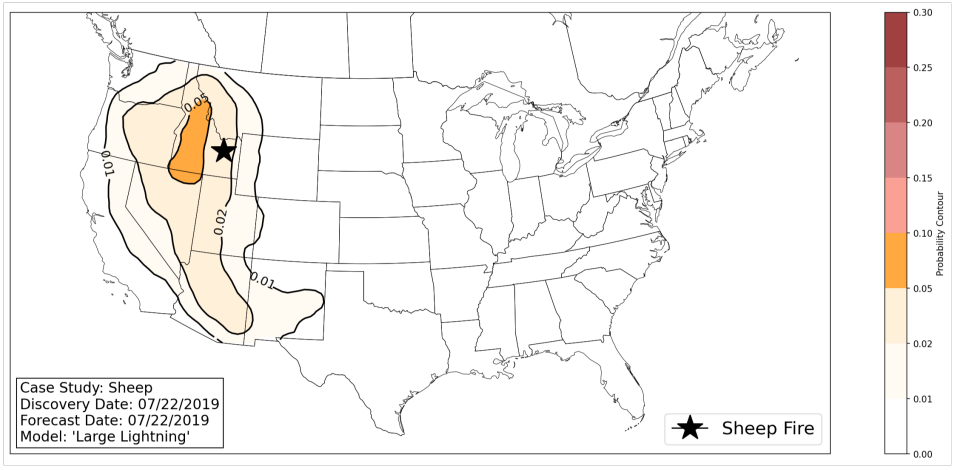


a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology

b) SPC Climatology Probability of Fire



c) "All Fires" Model Probability of Fire

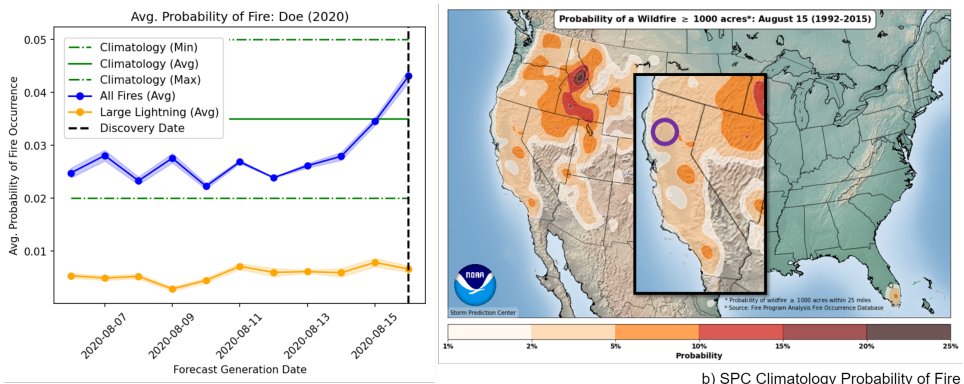


d) "Large Lightning" Model Probability of Fire

Figure 5.22: Probability of Fire: Sheep Fire, 2019 Largest Lightning Fire

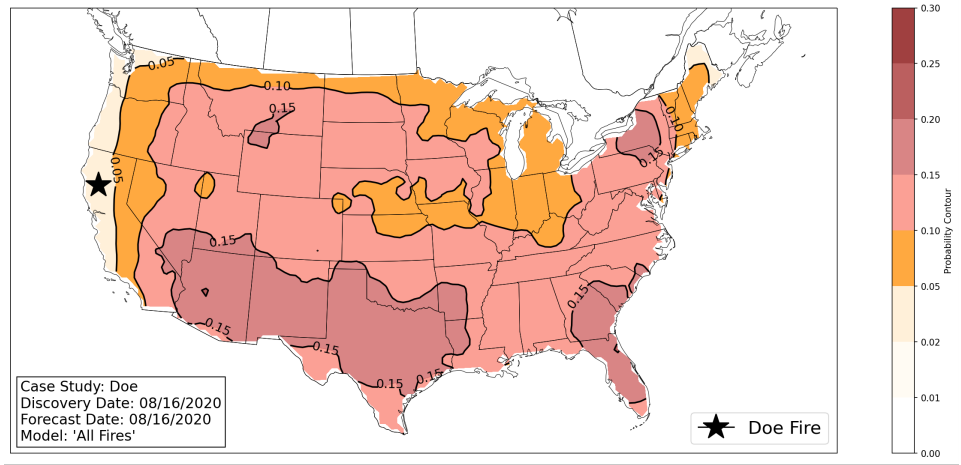
5.5.3 Doe Fire: 2020 Largest Lightning Fire

The Doe Fire, started by lightning, was discovered on 16 August 2020, in the Mendocino National Forest in California, and, as a part of the August Complex, burned approximately 1,032,648 acres (CAL FIRE 2020). Figure 5.23 describes the probabilities depicted by the SPC Climatology, the 'All Fires' model, and the 'Large Lightning' model for the Doe Fire. SPC Climatology depicted a probability between .02 and .05, with an average of .035, for the ten days prior to the discovery date for the location of discovery for the Doe Fire. The 'All Fires' model predicted greater average probability of fire than the average climatology on the date of discovery for the location of discovery. The 'Large Lightning' model did not produce probabilities greater than the average climatology probability, nor did it produce average probabilities greater than the climatology minimum, for all ten days prior to the discovery date.

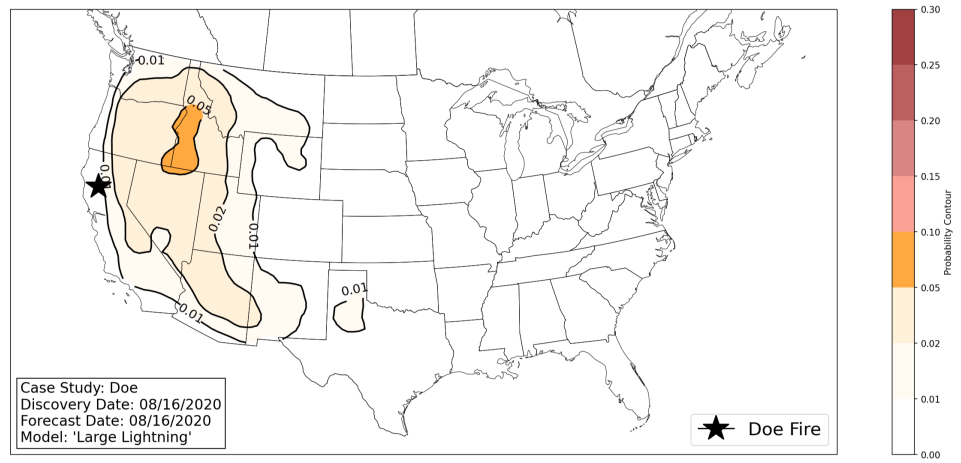


a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology

b) SPC Climatology Probability of Fire



c) "All Fires" Model Probability of Fire



d) "Large Lightning" Model Probability of Fire

Figure 5.23: Probability of Fire: Doe Fire, 2020 Largest Lightning Fire

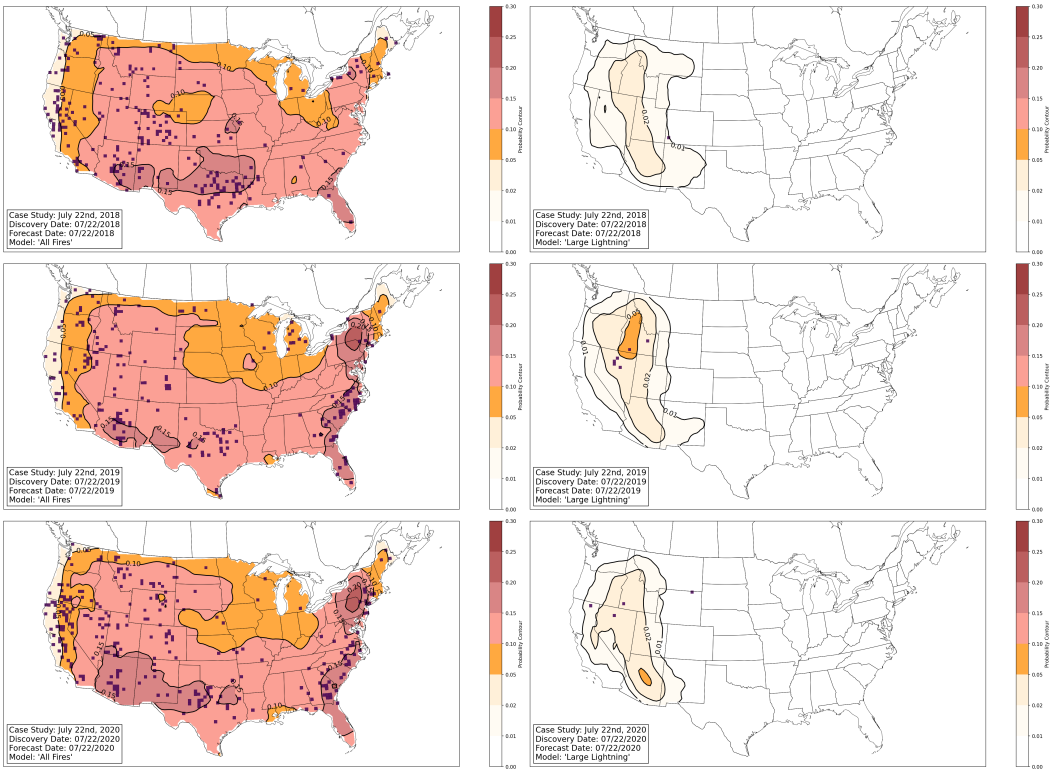


Figure 5.24: Probability of Fire: July 22nd 2018, 2019, 2020

5.5.4 Most Common Day for Lightning-Caused Fires: July 22nd

According to the work of Balch et al. (2017), across the 21 years of data from FPA-FOD that they evaluated, the most common day for lightning caused fires in the contiguous United States is July 22nd. Figure 5.24 depicts model performance for July 22nd for the years 2018, 2019, and 2020. The 'Large Lightning' model produces more localized predictions and, while the probabilities are lower, the predicted areas capture almost all of the large lightning instances discovered on that day for all three years. The 'All Fires' model produces larger areas to be fire prone, including those areas in which large lightning fires were discovered, and predicts a higher probability for these areas.

5.5.5 Most Common Day for Human-Caused Fires: July 4th

In the work of Balch et al. (2017), which used FPA-FOD, they found that '(h)uman-started wildfires accounted for 84% of all wildfires, tripled the length of the fire season, dominated an area seven times greater than that affected by lightning fires, and were responsible for nearly half of all area burned. For theses reasons, we feel it is important not to lose sight of model performance in the context of human-caused fires. Figure 5.25 depicts model performance on the most common day for human-caused wildfires in CONUS, July 4th, for the years 2018, 2019, and 2020. The 'All Fires' model predicts larger areas to be fire prone at higher probabilities while the 'Large Lightning' model predicts much smaller areas at much lower probabilities to be fire prone. While the 'All Fires' model does not catch all instances of fire (depicted in purple), California being the most notable area of misses depicted, the 'Large Lightning' model does not catch most of the large lightning fires discovered on that day.

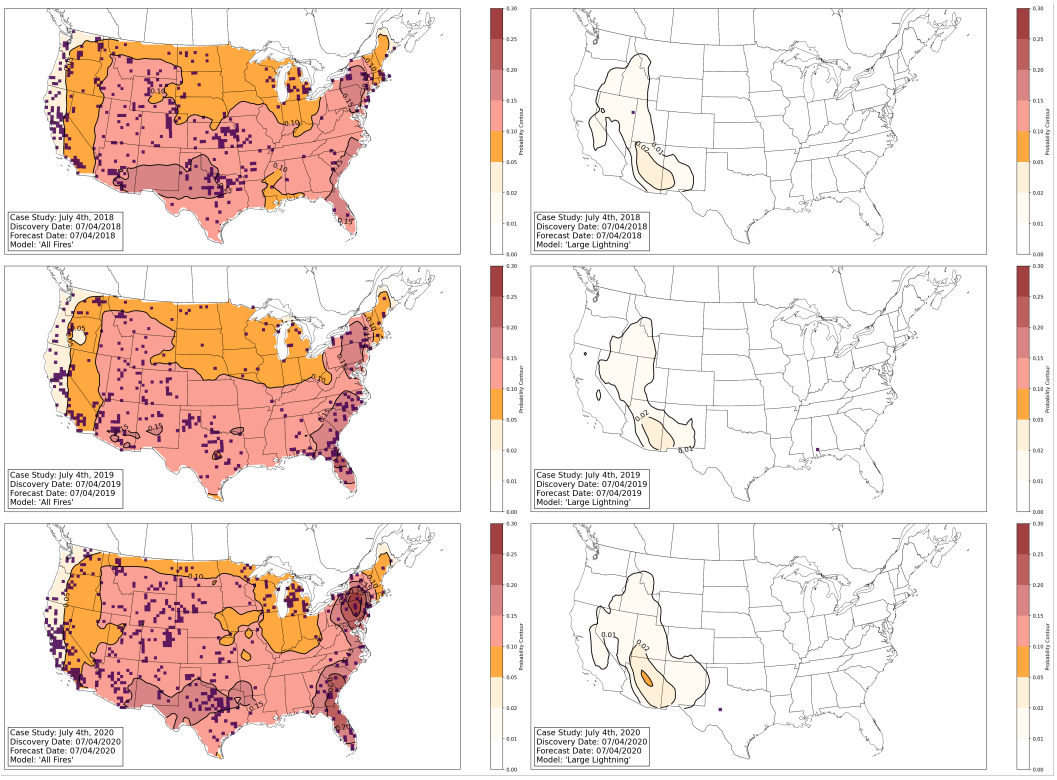


Figure 5.25: Probability of Fire: July 4th 2018, 2019, 2020

Chapter 6

Lessons Learned from Predicting Fire Occurrence using gridMET

In Chapter 5 of this work, we developed and analysed a Unet3+ deep learning model for the task of fire occurrence prediction in CONUS in the 1 to 10 day range. We developed two models, the "All Fires" model that predicted the probability of fire of any size and any cause, and the "Large Lightning" model that only predicted the probability of lightning-caused fires greater than or equal to 1000 acres in final fire size. We demonstrated that the "All Fires" model performed better than the "Large Lightning" model in general, on large lightning fires in specific, and on the largest of large lightning fires for the test data set. In Part 2 of this work, we explore the reasons why the "All Fires" model out performed the "Large Lightning" model and offer lessons learned that may be useful to other interdisciplinary researchers, specifically for those developing machine learning applications for wildfire prediction.

In the "Where is the Lightning?" section, where we discuss model performance improvements, we discuss why the models performed the way they did even in the absence of lighting inputs. In the "There are Inputs and Then There are Inputs" section, where we discuss choices that limited the model in some way, we discuss how variable selection and codification can limit model performance or remove the model's ability to offer certain insights all together.

6.1 Where is the Lightning?

One critical point worth drawing out is that neither the "All Fires" model nor the "Large Lightning" model had lightning observations as an input. This seems to be a condemning omission for models that are expected to predict large lightning fires. Exploring this question will help illustrate why performance differed between the "All Fires" model and the "Large Lightning" model.

6.1.1 Lesson 1: Higher Label Density Leads to Better Performance

The first lesson we took away from Part 1 is that more fire occurrences in the label images lead to better performance. This is a different lesson than the commonly known "more data leads to more performance" because both models were given the same number of examples (same domain in time, space, and input variables). Where the models differed is in the number of instances of fire occurrence depicted in the label images used to train the models and to measure model performance. Since the "All Fires" model included fires regardless of size or cause, more instances of fire occurrence were available for the label images offered to the model as depicted in Table 6.1. This provided the "All Fires" model with more opportunities to identify meaningful patterns between the inputs and the labels that may have allowed it to be more performative. The "All Fires" model was able to issue fire probability predictions both where large lightning fires were discovered and where other causes and sizes of fires were discovered.

Examples of this relationship can be seen in Figure 6.1, where a gap in performance can be observed between the "All Fires" model and all other models attempted, each of which sought to constrain the cause or size of the fire occurrence instances included

Cross Validation Rotation : 0	*Season*	*All Fires Label Count*	*Large Lightning Fires Label Count*
Train	Fall	216,440	269
(2000-2014)	Spring	412,083	268
	Summer	402,630	3,024
	Winter	202,069	9
Validation	Fall	49,953	51
(2015-2017)	Spring	75,108	46
	Summer	76,064	656
	Winter	38,594	0
Test	Fall	38,776	41
(2018-2020)	Spring	75,574	41
	Summer	76,821	489
	Winter	24,974	2

Table 6.1: Fire Occurrence Label Counts by Season for All Fires Model and Large Lightning Model

in the label images in some way. We validate in the following section that this increase in performance does not come at the expense of maintaining high performance on specific cases of interest, e.g. large lightning fires.

6.1.2 Lesson 2: Start with the General Case then Go Specific, Not the Other Way Around

We framed the task of predicting large lightning fires into the following three components:

Component 1: A cloud-to-ground (CG) lightning flash must meet the criteria necessary to start a fire: it must contact the ground, it must connect with something combustible, and it must have high enough amperage and low enough voltage to convert electrical charge to combustion (Pyne 2001).

Component 2: Fire mitigating weather (such as frequent, heavy rains) must not be present as it hinders the ignition and spread of fire.

Component 3: CG lightning must occur in an area that is already primed for fire.

We can see from the framing presented that the scope of each component builds on the scope of the previous component. Component 1 deals only with the ignition source. Component 2 deals with the means of mitigation for the phenomenon of interest. Component 3 deals with the larger environmental factors that can contribute to making an area suited for the phenomenon of interest. In NWP, it is common to focus first on those components that apply only to large lightning fire, components one and two, so as not to dilute the signal received by the model. This practice does not necessarily apply when using a deep learning model.

Instead, as we demonstrate in Chapter 5, by taking the opposite approach, and starting with the most general component, component three, we were able to build a performative model (the "All Fires" model), both in general and on the specific case of interest, large lightning fires. This performance stems from the "All Fires" models ability to predict where fires, any fires, are likely to be discovered in CONUS, or, said differently, which areas of CONUS are primed for fire. In contrast, the "Large Lightning" model predicts only where large lightning fires are likely to be discovered in CONUS, which provides a much narrower view into potential future fire behavior. Predicting where fires are possible, the goal of the "All Fires" model, as opposed to where large lightning fires are possible, the limitation of the "Large Lightning" model, is a key prospective shift necessary in building a model that generalizes well and has some measure of future proofing. After all, where large lightning fires have happened in the past does not dictate, exclusively, where large lightning fires, or indeed, all fires, will happen in the future.

An interdisciplinary researcher must keep in mind historical methods used to address the challenge at hand and identify areas where these historical methods can be improved. Researchers can then determine how improvements can be made using a new tool, in the case of our work, deep learning. This way, researchers can offer

the most value from a new tool while keeping the capabilities of the previous tool in mind. In our work, the historically preferred method for event forecasting in the fire weather space is to use an NWP model. Currently, there is no NWP model in place for fire weather or fire occurrence prediction though there is an entrenched preference for them and for methods that align with NWP practices. The lack of a current model allowed us the room to explore a path that diverged largely from preferred methods and we were able to offer a performative deep learning model for the task of predicting wildfire occurrence as a result. However, by creating an understanding of NWP and partnering with experts in NWP, we were able to build a foundation for knowledge transfer so that experts in the field could more readily understand how our new method works.

6.2 There are Inputs and Then There are Inputs

There are many ways to support the performance of a machine learning model. As discussed in the previous section, one way focuses on the impacts of labeling and another on how the modeler interprets the modeling space. Another way to support model performance is by selecting the best inputs for the predictions that the model will need to make. Part of this task deals with variable selection, a topic touched on in both this section and in the next section. Another piece of this task, focused on in this section, deals with how the variables, once selected, are codified (pre-processed or transformed) for use by the model.

One example of how variables can be codified for the model is the process of normalization. When using neural networks, it is important to normalize the model

inputs prior to their use. Normalizing model inputs supports the model training process by shifting all inputs into the same scale that helps stabilize gradient descent (the method by which neural networks learn). Stabilizing gradient descent can result in faster model convergence to an optimal solution. For this work, we chose to normalize all inputs to values between zero and one using min-max normalization though there are many normalization options, such as normalizing to +/- 1. Min-max normalization was an appropriate choice because our input datasets did not have an issue with outliers. Using min-max normalization with a dataset that has outliers can force all data points that are not outliers to the opposite end of the range from the outliers (if outliers are large, all non-outlier values will be cast to near zero and vice versa).

6.2.1 Lesson 3: Summary Variables Have Their Place, Just Not Necessarily in a Deep Learning Model

At the confluence of variable selection and variable codification is the question of summary variables. It is tempting in interdisciplinary work to seek out summary variables carefully created by the subject matter experts working in the space, based on years of experience and rules of thumb. This approach is not without merit and can speak to important themes valuable to the modeling task, such as what variables forecasters consider to be important. However, when machine learning is not the currently used method, it is valuable to keep in mind not only what has worked for forecasters but what will work for the machine learning model. In the case of summary variables, such as climatological variables and probabilities derived from climatology, the data components that human forecasters rely on do not necessarily provide the machine learning model with the complexity of data that it needs to produce its highest quality prediction.

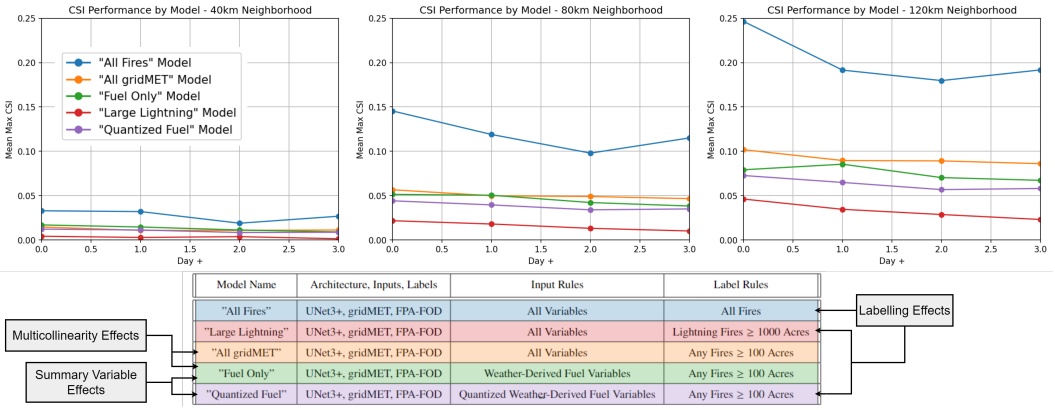


Figure 6.1: Max CSI by Model

One early approach we tried in applying machine learning to the task of fire occurrence prediction, informed by how subject matter experts use climatologies today, was to use quantized fuel variables - called the "Quantized Fuel" model in Figure 6.1. We broke weather-derived fuel variables (energy release component, burning index, and dead fuel moisture 100 hour and 1000 hour) down into quantiles (0, 10, 25, 50, 75, 90, 100) and converted those quantiles into individual variables (ex. one variable would be the 90th percentile burning index) and submitted those variables to the model as inputs, much as one would do when considering a climatology. Quantile variables were created using linear interpolation (Hyndman and Fan 1996).

The performance of the "Quantized Fuel" model, when compared to other models we developed as our work matured, was relatively low. Performance for all models is depicted in Figure 6.1 in which model performance is measured using Critical Success Index (CSI). CSI is a performance metric used to evaluate model predictions within the National Weather Service (Schaefer 1990).

Here we compare two models that used similar inputs and labeling strategies, the "Fuel Only" model and the "Quantized Fuel" model. Both models used the same input variables from the same source: burning index, energy release component, and

dead fuel moisture sourced from gridMET. The difference between the two models was the "Quantized Fuel" model relied on summary inputs (its input variables were converted to quantiles before being normalized and offered to the model) while the "Fuel Only" model relied on non-summary inputs (its input variables were only normalized before being offered to the model). The "Fuel Only" model produced higher max CSI values than the "Quantized Fuel" model, based on performance as depicted in Figure 6.1.

One interpretation for this difference in performance is that the entire distribution, as described by the non-summary input variables, carried more information than the values used to summarize the distribution, as described by the summary variables. The summary dependent model also took up more storage space and was more computationally intensive than the non-summary dependent model as we added multiple summary variables (each with the same coverage in time and space but offering less information than the original variable) in an attempt to compensate for the information lost in summarization. This resulted in a large model that trained slowly and performed relatively poorly, characteristics that are less desirable for a model bound for operationalization.

The key take-away described by these results is that the use of summary variables as model inputs when the model is capable of handling the non-summary variables results in a model starved for information. The appetite for summary variables stems largely from how the human mind works, famously described by Tversky and Kahneman (1974) as heuristics, or the shortcuts the human mind takes to make decisions in the presence of uncertainty. A deep learning model needs no such shortcut and is either biased or constrained by being provided with one. Without such a shortcut, a deep learning model is free to create its own categories and can stand in for previously necessary shortcuts. This allows decision makers to quickly connect the dots

without having to sacrifice the amount of information considered. This trait makes deep learning a great candidate for decision support.

6.2.2 Lesson 4: Multicollinearity & Variable Contribution

Multicollinearity describes a situation wherein the inputs to a model are correlated with each other. While the presence of multicollinearity does not necessarily produce a less performative model, it does impact our ability to quantify the contributions of the individual input variables for the model. When multicollinearity is present, variable importance, as measured using permutation importance, for example, is difficult to measure because it is unknown from which correlated variable the effect on the model is produced. Possible solutions for such a situation are 1) to remove all of the highly correlated variables, retaining only the uncorrelated variables of interest or 2) to apply dimensionality reduction (principal component analysis being an example of such) to reduce the amount of correlation in the data set while retaining as much information as possible.

The data sets used for this work were recommended by subject matter experts (Nauslar 2020a; Bentley 2022; Elliott 2023) because of their familiarity to fire weather forecasters with the goal of lowering barriers to adoption for the model. Between the input variables, sourced from gridMET, there is a lot of correlation, as depicted in Figure 6.2. While the rules defining which values of correlation correspond to different correlation strengths vary by situation and discipline, we have described our discretization in Figure 6.2.

Our subject matter experts would like to receive information regarding variable importance in the context of the input variables used for the model and for those input

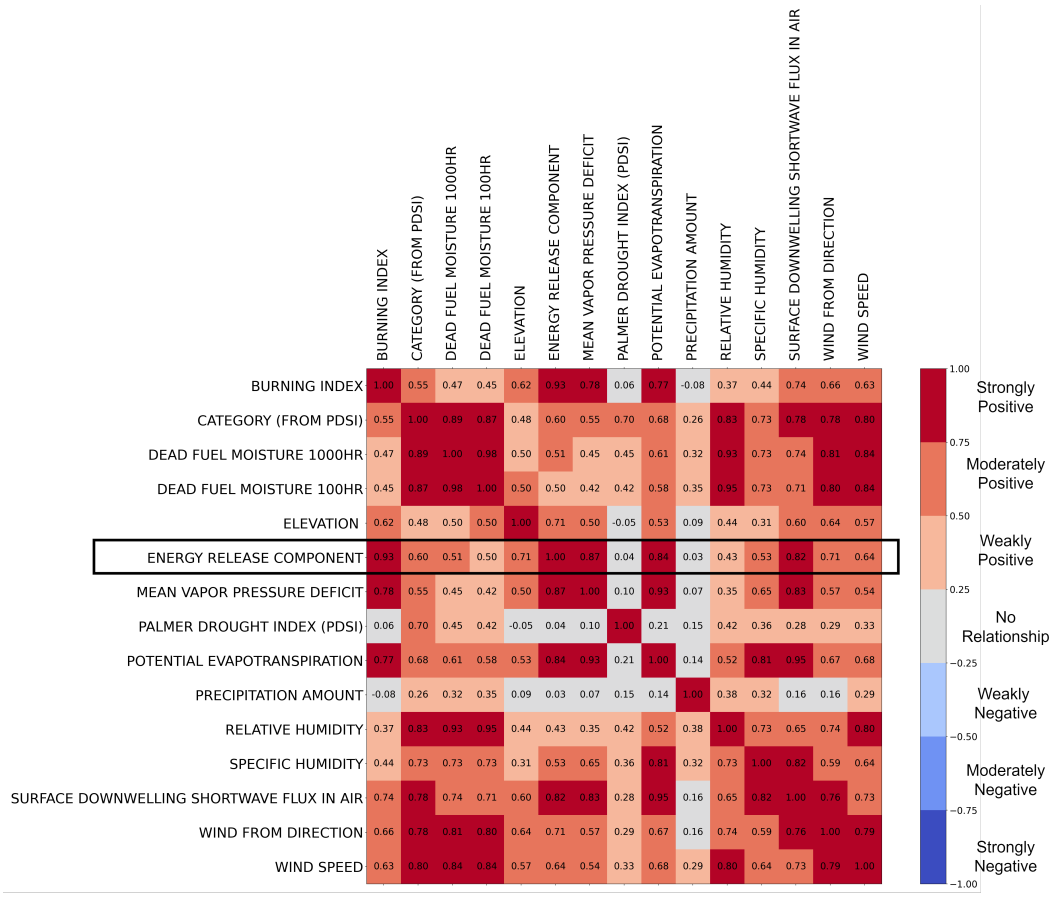


Figure 6.2: Correlation Plot of gridMET Variables

variables to be variables familiar to them. For example, energy release component (ERC) is a popular variable used in fire weather forecasting that describes the amount of available energy (BTU) per unit area (square foot) within the flaming front at the head of a fire (United States Department of Agriculture 2023). Fire weather forecasters would like to know how important ERC is to making an accurate fire weather forecast. With the "All Fires" model, perhaps we have a model offering enough accuracy for the discussion to shift to variable importance.

In Figure 6.2, all but two input variables exhibit a positive correlation with Energy Release Component (ERC). It is important to understand that ERC is calculated using burning index and 1000 hour fuel moisture, that is in turn calculated

using temperature, relative humidity, and precipitation (National Interagency Fire Center 2023). This is called structural multicollinearity, wherein correlation between variables results from using one model term to create another model term, resulting in more variables for the model but not necessarily more information for the model.

Option 1, that of removing variables highly correlated with the variable of interest, is likely not at our disposal unless we are willing to lose the majority of our input variables. Option 2, that of applying dimensionality reduction to the input variable set, as in the use of PCA, would allow us to quantify variable importance but only in terms of new, linearly uncorrelated, variables, called principal components. The resulting principal components would no longer be of a format recognized by the fire weather forecasters and would lack the transferable context desired by the fire weather forecasters.

In summation, neither option 1 nor option 2 offers the ability to quantify variable importance within the context of recognizable variable inputs using the data sets recommended for this work as familiar to fire weather forecasters. While it is still computationally possible to perform permutation variable importance on a dataset with high multicollinearity such as this, the outcome would be unlikely to offer any meaningful interpretation. As variable importance is a valuable tool for supporting model understanding, trust, and, ultimately, adoption, this creates a vulnerability for a model bound for operationalization.

Chapter 7

Fire Occurrence Prediction using GEFS

In this work, I have sought to create a model that could predict fire occurrence for CONUS accurately by predicting the areas of CONUS primed for wildfire based on environment factors. I hypothesize that such a model could capture the majority of wildfire occurrences without needing to rely on ignition specific variables such as population density and cloud-to-ground lightning. In capturing the majority of wildfire occurrences, I hypothesize that the model will also capture the wildfire occurrences most important to operational forecasters, large wildfires and large lightning caused wildfires. In previous chapters, my model could not generate variable importance information because of the multicollinearity present in the input dataset, gridMET. In Chapters 7 and 8, I describe models which use a different dataset, GEFS, with the goal of presenting a model that can continue the performance described in Chapters 5 and 6 while providing improved interpretability in the form of variable importance.

I discuss two approaches to labeling in this work. The first approach represents a return to the labeling strategy used in Chapters 5 and 6, wherein three types of labels were employed during training: pixel, neighborhood, and time-smoothed neighborhood. I call these models the "Multi Label" models. The second approach removes all additional labels and relies only on the pixel label, reducing the burden of pre-processing, while maintaining model performance when compared to the three label approach. I call these models the "Pixel Label" models.

Model	*Prediction Day*	*Input Day : Observed Inputs*	*Input Day : Forecast Inputs*
Day 0	Day 0	Day 0 Observed	Day 0 Observed
Day 1	Day 1	Day 0 Observed	Day 1 Forecast
Day 2	Day 2	Day 0 Observed	Day 2 Forecast
Day 3	Day 3	Day 0 Observed	Day 3 Forecast
Day 4	Day 4	Day 0 Observed	Day 4 Forecast
Day 5	Day 5	Day 0 Observed	Day 5 Forecast
Day 6	Day 6	Day 0 Observed	Day 6 Forecast
Day 7	Day 7	Day 0 Observed	Day 7 Forecast
Day 8	Day 8	Day 0 Observed	Day 8 Forecast
Day 9	Day 9	Day 0 Observed	Day 9 Forecast
Day 10	Day 10	Day 0 Observed	Day 10 Forecast

Table 7.1: GEFS Daily Models, Observed Vs. Forecast Input Structure

Labels		*Pixel, Neighborhood, Time Smoothed*	*Pixel*
Inputs	*Observed*	Multi Label, Observed Inputs	Pixel Label, Observed Inputs
	Forecast	Multi Label, Forecast Inputs	Pixel Label, Forecast Inputs

Table 7.2: Models: Labels and Inputs

I describe two approaches to input data in this work. The first approach, called the "observed inputs" approach, is a return to the input data approach used in Chapters 5 and 6 where in the models used observation data only. In Chapters 7 and 8, the "observed inputs" models use day 0 inputs regardless of which day the model is predicting (day 0 to day 10). This is possible because GEFS data are available for same day use. In the second approach, called the "forecast inputs" approach, models predicting for future days (beyond day 0) are given the associated forecast data for that day. The day 1 model is given day 1 forecast inputs, the day 2 model is given day 2 forecast inputs, and so on, as described in Table 7.1. Each of the four approaches described in Chapters 7 and 8 uses the UNet3+ model and a combination of one of the two labeling styles and one of the two input data styles as described in Table 7.2.

I use model performance from maximum critical success index (CSI) and reliability to establish a top model from each approach. I then compare the top models to the

SPC Probability of Wildfire Climatology using case studies of the most impactful fires of 2018: the Camp fire, the Carr fire, and the Woolsey fire.

7.1 Multi Label Model

Individual CSI performance plots for the "Multi Label" model are available in Figures 7.1 and 7.2. In Figures 7.1 and 7.2, model performance differs between neighborhood size, with the 120km neighborhood producing the highest CSI values for this experiment. CSI values are consistent from day 0 to day 10 for this experiment. These observations are true for both the forecast inputs models and the observed inputs models.

Figure 7.3 depicts the Max CSI performance for the "Multi Label" model for both the observed and forecast input approaches. The "Multi Label" model achieves a similar Max CSI regardless of whether the model is given observed inputs or forecast inputs. The three neighborhoods measured produce consistently different Max CSI values for this experiment, with the 40km neighborhood demonstrating the lowest Max CSI values and the 120km neighborhood demonstrating the highest Max CSI values. Max CSI remains relatively stable from day 0 to day 10 that suggests that the model's ability to draw predictive value from the input dataset is consistent for all models, day 0 to day 10. It would be interesting to explore in future work how far beyond 10 days performance can be maintained.

Individual reliability diagrams for the "Multi Label" model are available in Figures 7.4 and 7.5. The reliability behavior is relatively consistent between neighborhoods with the 40km, 80km, and 120km neighborhoods exhibiting similar performance regardless of which set of inputs the model is given. The 40km neighborhood measurement tends to over predict and the 120km neighborhood measurement tend to under

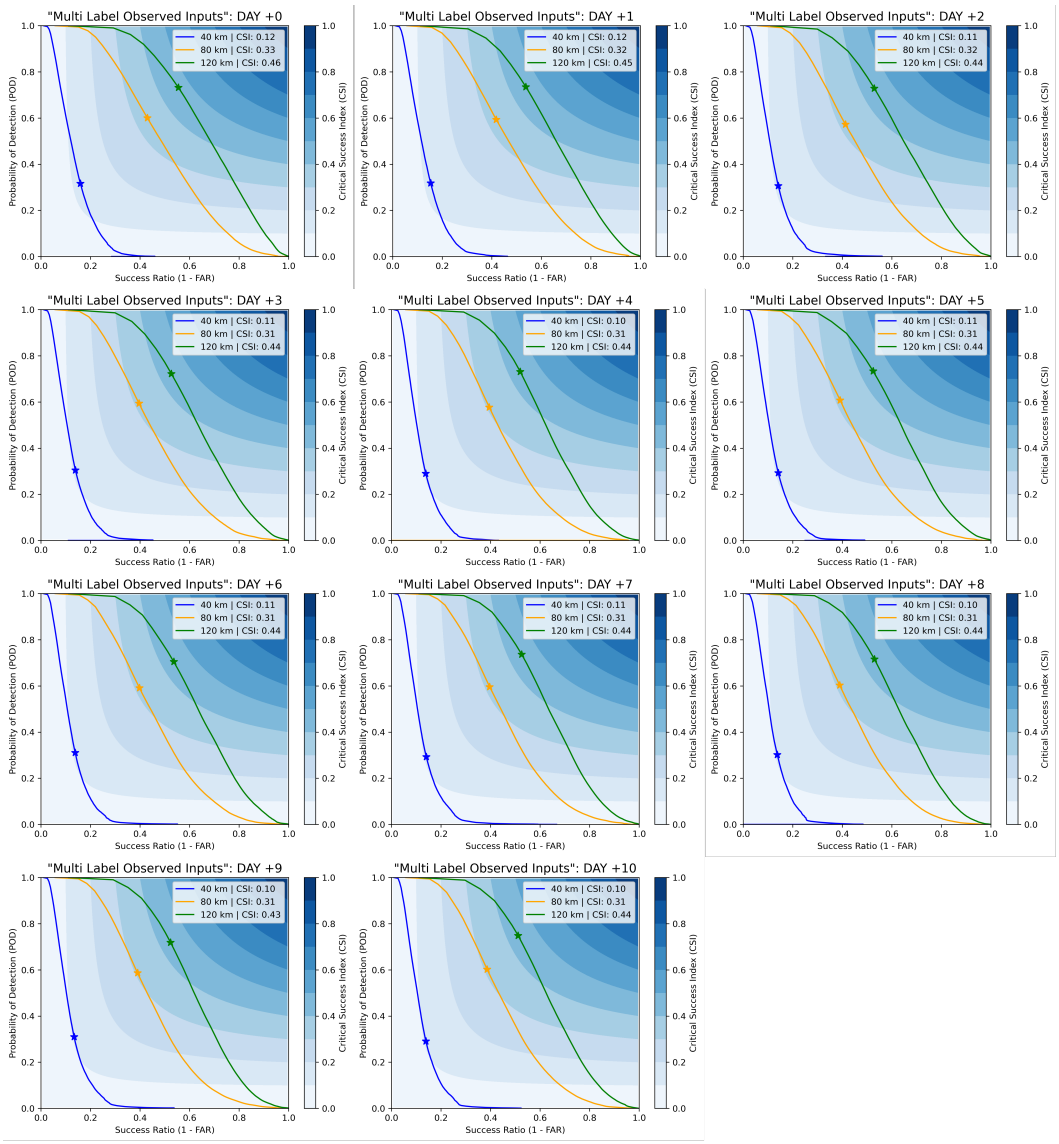


Figure 7.1: Performance Plots for Day 0 - 10 for 'Multi Label Observed Inputs' Model. For more information on interpreting performance plots see Roebber (2009).

predict. For the 80km neighborhood, the line produced by the observed frequency of the event and the predicted probability produced by the model most closely follows the 45 degree dot-dash line that indicates perfect reliability. This indicates that the 80km neighborhood measurement is the most reliable for both the forecast inputs model and the observed inputs model for this experiment.

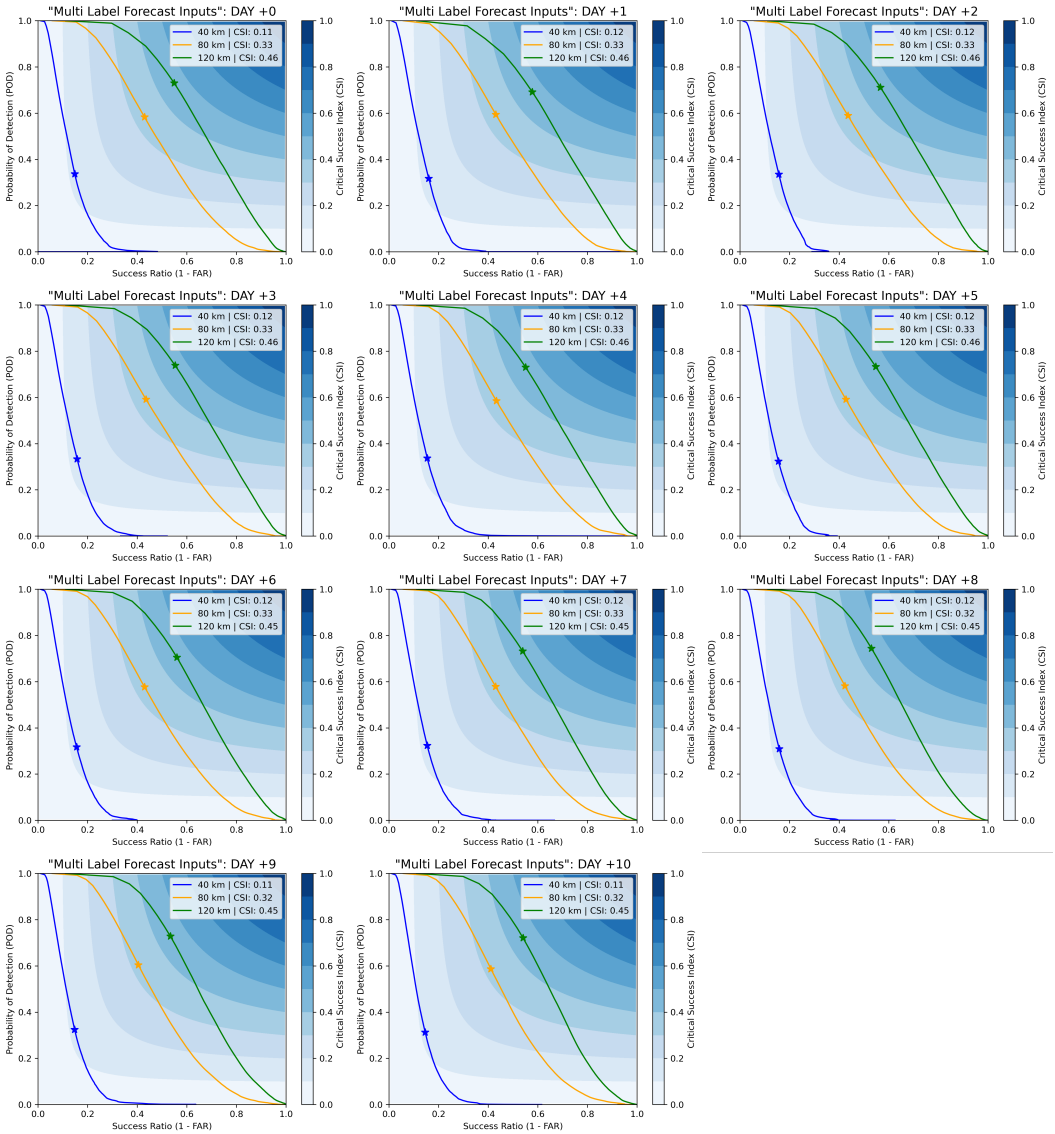


Figure 7.2: Performance Plots for Day 0 - 10 for 'Multi Label Forecast Inputs' Model. For more information on interpreting performance plots see Roebber (2009).

Figure 7.6 describes Max CSI performance for the "Multi Label" model measured at 40km, 80km, and 120km without constraining all neighborhoods to the same range of values for the y-axis. This allows us to observe the performance differences between the forecast inputs model and the observed inputs model more effectively as well as model performance from day 0 to day 10. Figure 7.6 also describes the probability

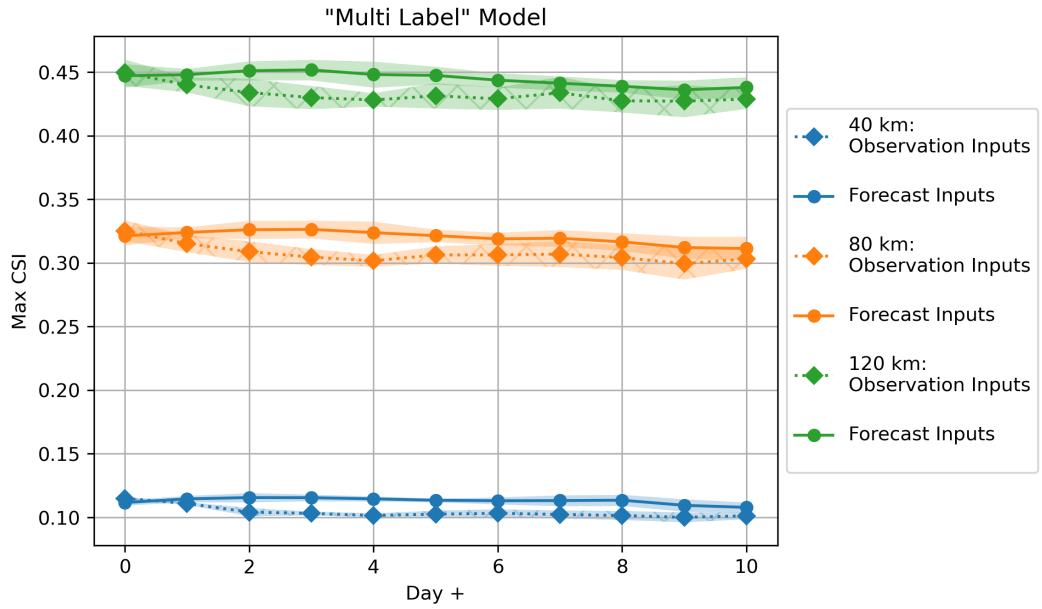


Figure 7.3: Max CSI Values for the "Multi Label" Model

threshold at which the Max CSI was achieved. Looking at the neighborhood performance in Figure 7.6, I can see two additional themes. First, in Figure 7.3 the decrease in performance from day 0 to day 10 is not large. In Figure 7.6, however, a slight performance reduction from day 0 to day 10 for this experiment is observable. For all three neighborhoods, day 0 Max CSI is higher than day 10 Max CSI. The forecast models produced higher Max CSI values for this experiment than the observed models for all three neighborhoods, for all days post day 0. Overlap between the performance demonstrated by the forecast input model and the observed input model appears to increase as the neighborhood size increases. When observing the probability thresholds associated with the Max CSI achieved for each day, in Figure 7.6, there is a large amount of overlap between the two input approaches for all three neighborhoods and, from day 0 to day 10, the trend of the probability thresholds stays relatively flat.

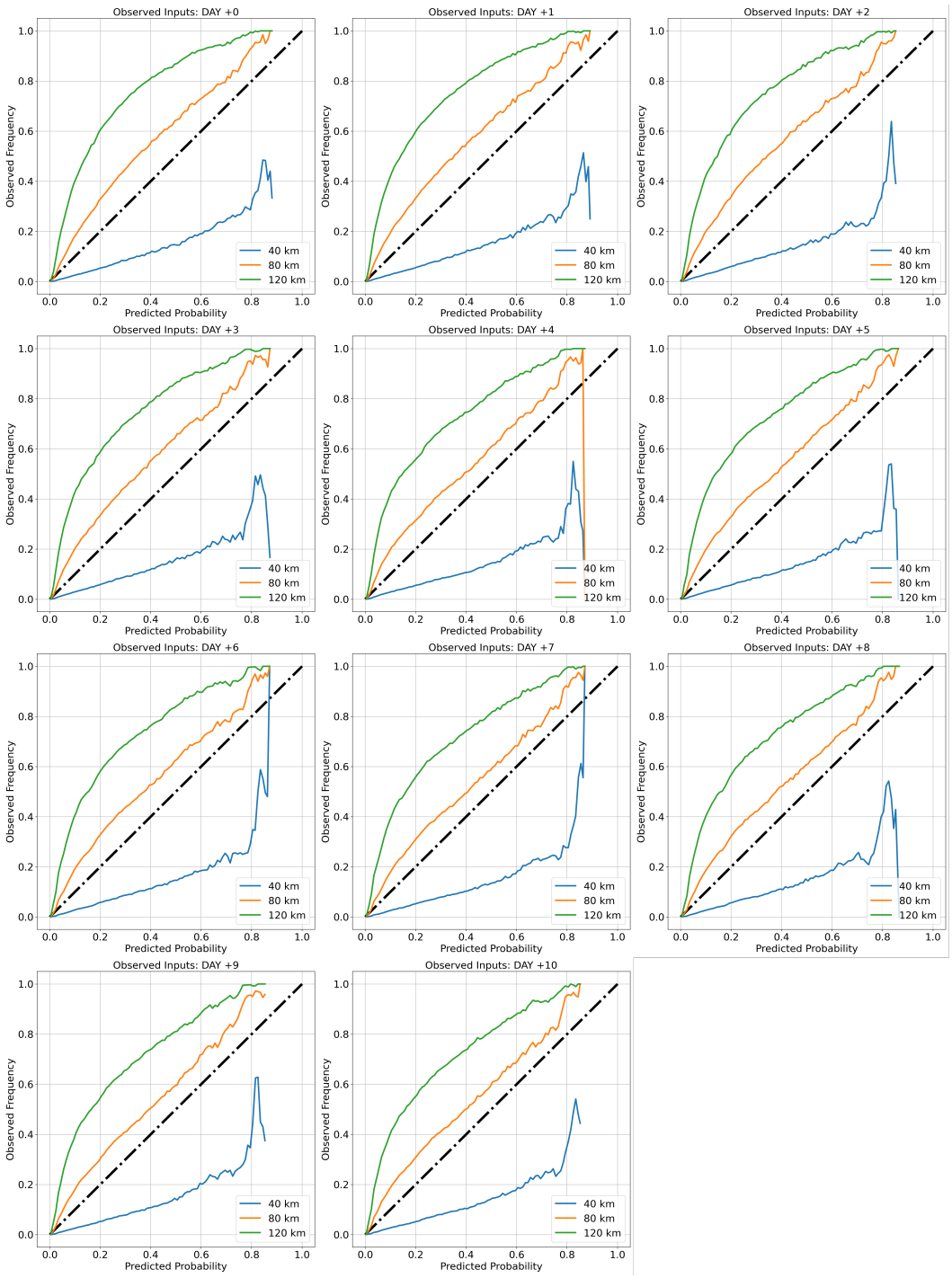


Figure 7.4: Reliability Plots for Day 0 - 10 for 'Multi Label Observed Inputs' Model. For more information on interpreting reliability plots see Bröcker and Smith (2007).

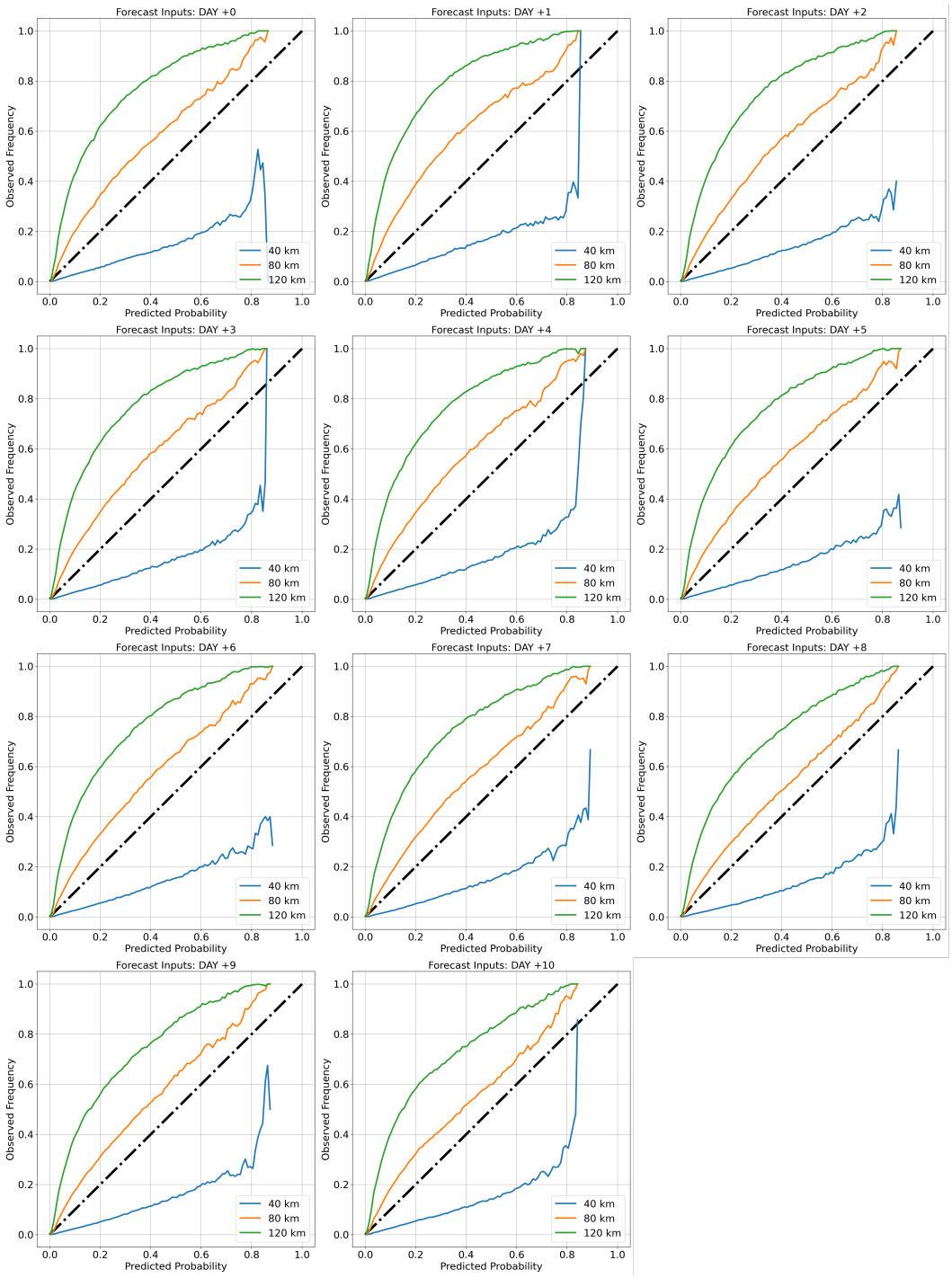


Figure 7.5: Reliability Plots for Day 0 - 10 for 'Multi Label Forecast Inputs' Model. For more information on interpreting reliability plots see Bröcker and Smith (2007).

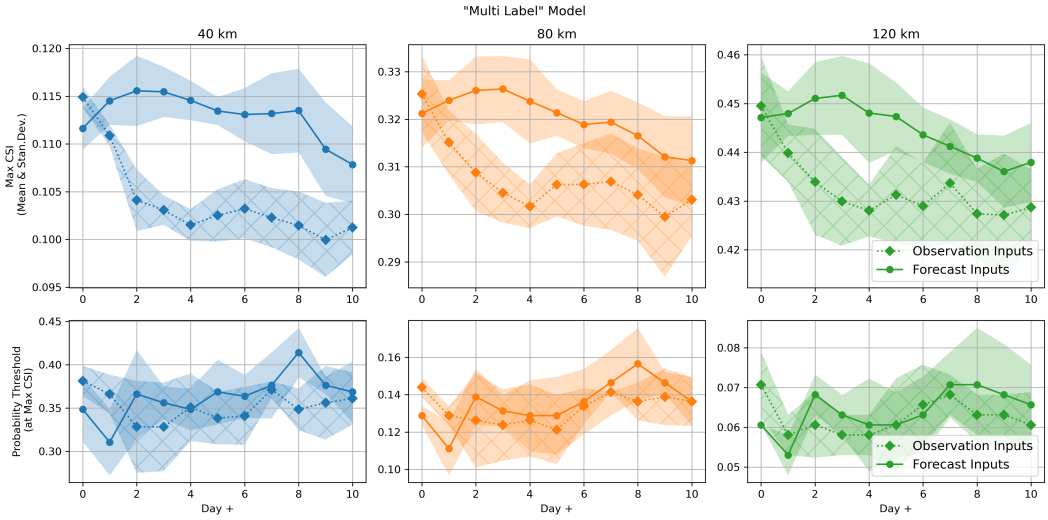


Figure 7.6: Max CSI Values & Probability Thresholds for the “Multi Label” Model. Max CSI values are drawn from the Individual Performance Plots contained in Figures 7.1 and 7.2. The Probability Threshold is the probability cutoff from the Individual Performance Plot at which Max CSI is achieved.

I select the ”Multi Label” forecast inputs model measured at 80km as the model to use in the case study comparisons because the 80km measurement of reliability has the most reliability and the forecast inputs model produces slightly higher CSI values than the observed inputs model for this experiment.

7.2 Pixel Label Model

Individual CSI performance plots for the “Pixel Label” model are available in Figures 7.7 and 7.8. In Figures 7.7 and 7.8, model performance differs between neighborhood size, with the 120km neighborhood producing the highest CSI values for this experiment. CSI values are consistent from day 0 to day 10 for this experiment. These observations are true for both the forecast inputs models and the observed inputs models. Max CSI values are taken from the individual performance plots depicted in

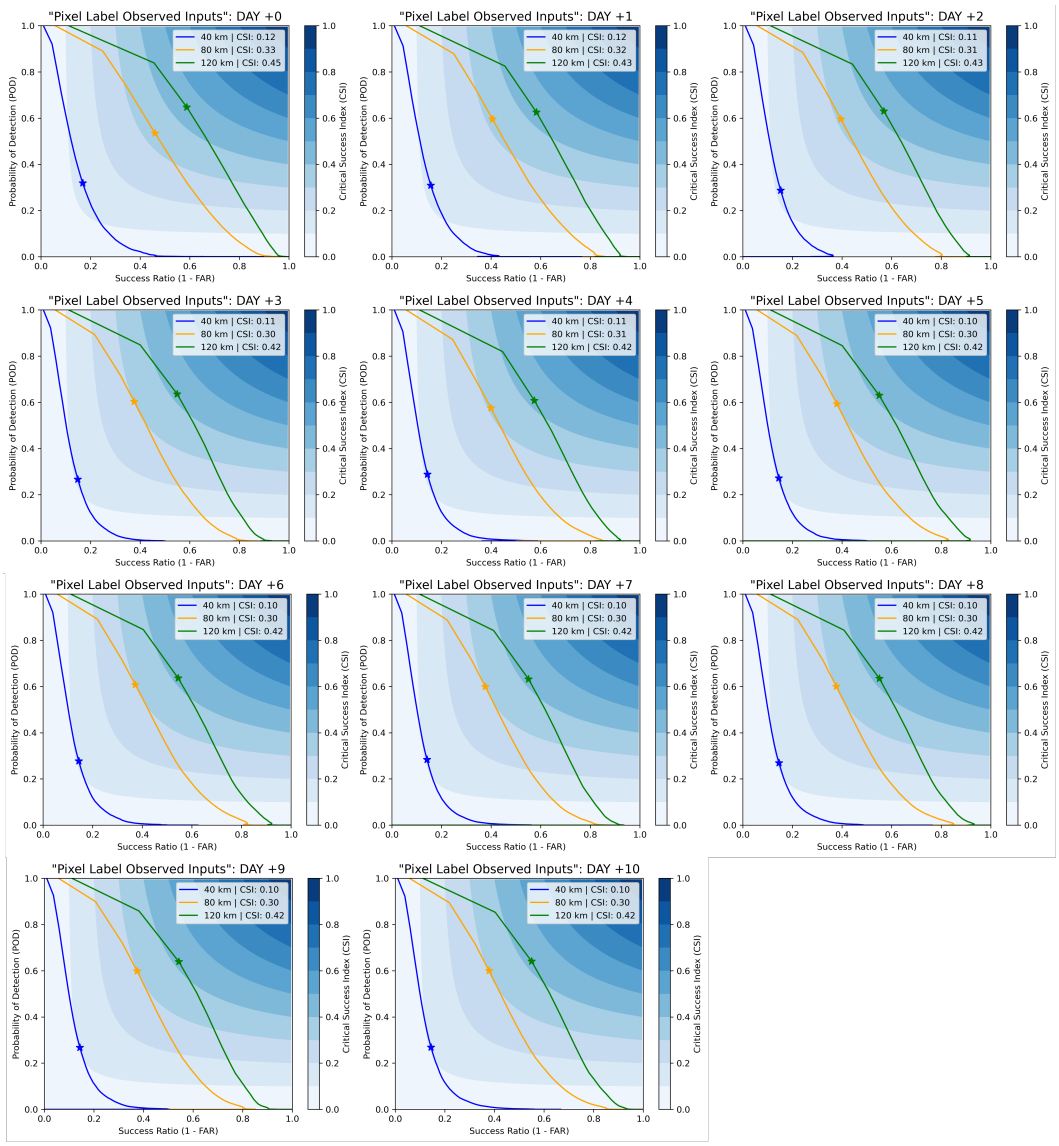


Figure 7.7: Performance Plots for Day 0 - 10 for 'Pixel Label Observed Inputs' Model.

Figures 7.7 and 7.8. The Max CSI value for each neighborhood is indicated with a star and described in the legend for each plot.

Figure 7.9 depicts the max CSI performance for the "Pixel Label" model for both the observed and forecast input approaches. In Figure 7.9, I can see all three themes from the "Multi Label" model repeated in the "Pixel Label" model. I can see that there is not a large difference in CSI between the observed inputs and the forecast

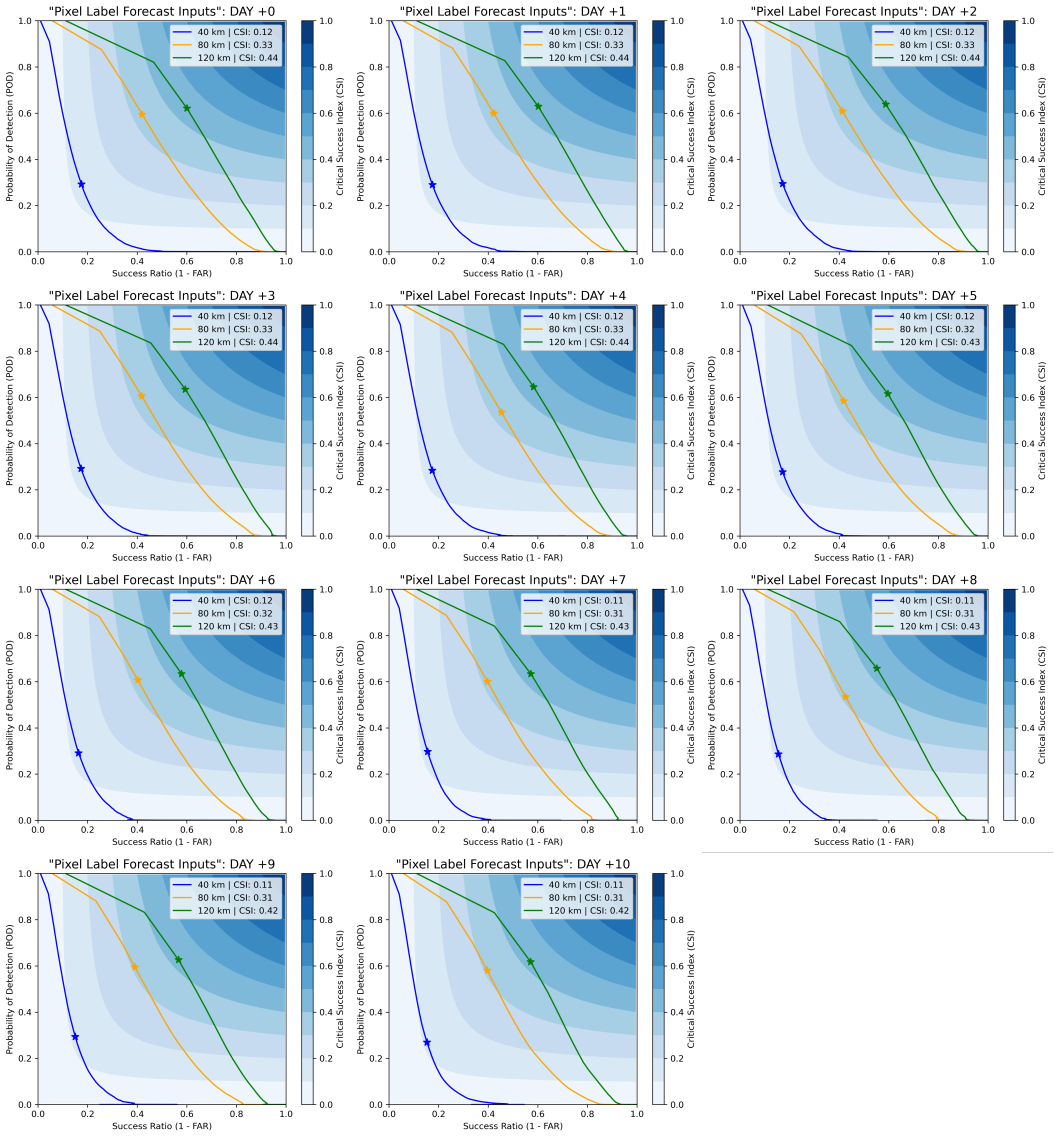


Figure 7.8: Performance Plots for Day 0 - 10 for 'Pixel Label Forecast Inputs' Model.

inputs for this experiment. I see that neighborhoods behaved differently, with 40km having the lowest CSI and 120km having the highest CSI for this experiment. And I see that performance remains relatively stable out to 10 days for this experiment. Additionally, comparing the "Multi Label" model and the "Pixel Label" model using the Max CSI values depicted in Figures 7.3 and 7.9, I can see that both models had fairly similar performance for each neighborhood.

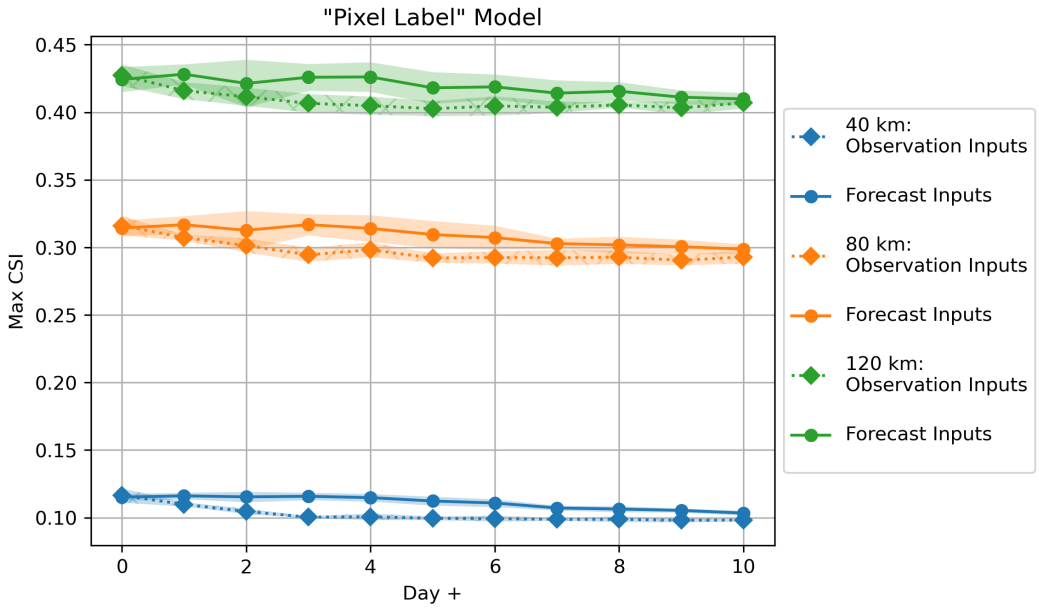


Figure 7.9: Max CSI Values for the "Pixel Label" Model

Individual reliability diagrams are available in Figures 7.10 and 7.11. In Figures 7.10 and 7.11, comparing the observed inputs approach on the left with the forecast inputs approach on the right, as was the case with the "Multi Label" model, reliability behavior is relatively consistent for each neighborhood. I see in the "Multi Label" model that the most reliable neighborhood was the 80 km neighborhood. For the "Pixel Label" model, I see that the 40km neighborhood is now the most reliable for both the forecast inputs model and the observed inputs model for this experiment.

Figure 7.12 describes Max CSI performance for the "Pixel Label" model without constraining all neighborhoods to the same y-axis. In it, I can see the themes from the "Multi Label" model repeated. First, I can see that CSI decreases as predictions move from day 0 to day 10. Secondly, I see that the forecast inputs model produced higher CSI values than the observed inputs model with the appearance of some convergence as the neighborhood size increases. When observing the probability thresholds associated with the Max CSI achieved for each day, as was the case with the "Multi Label"

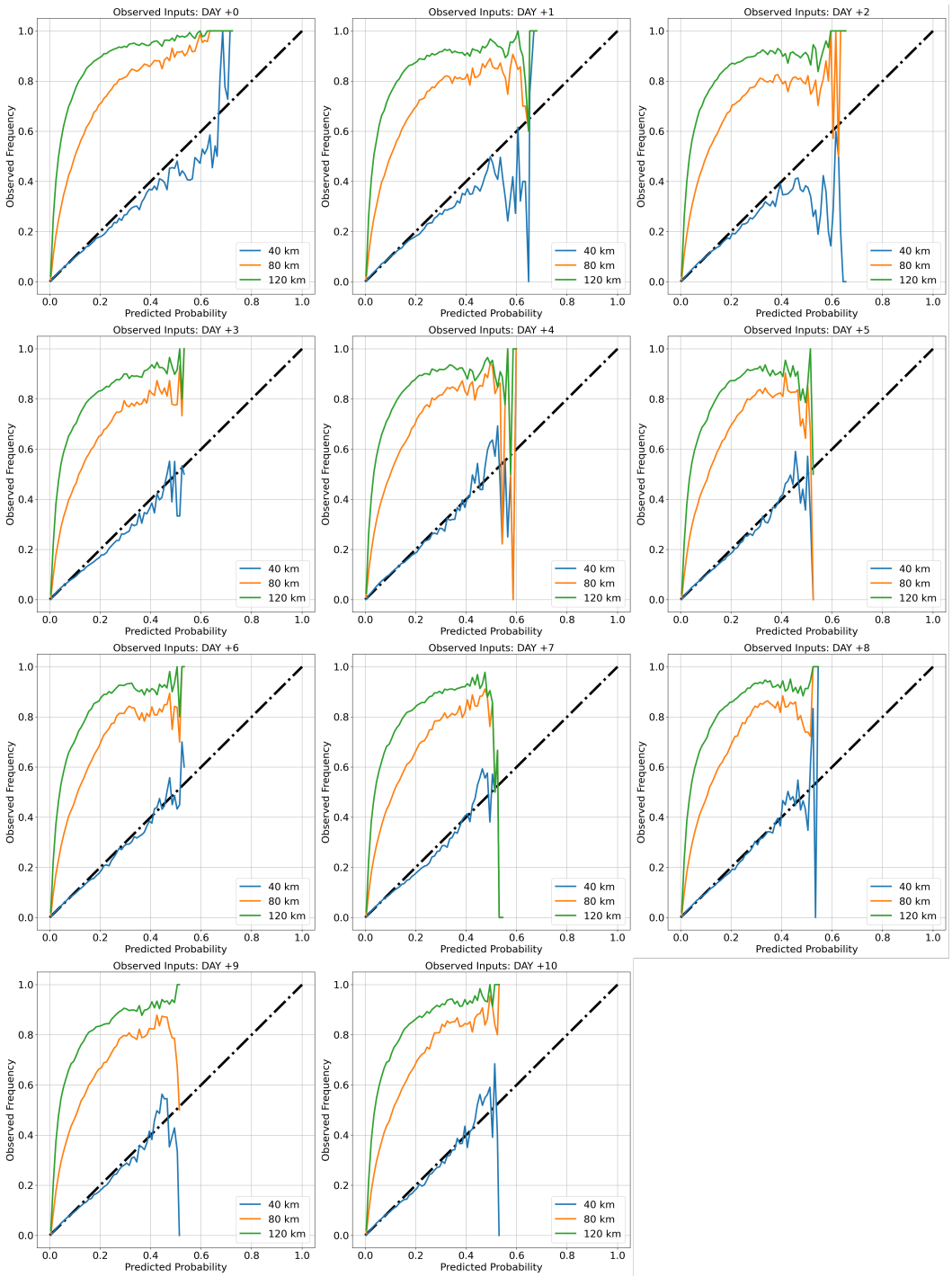


Figure 7.10: Reliability Plots for Day 0 - 10 for 'Pixel Label Observed Inputs' Model.

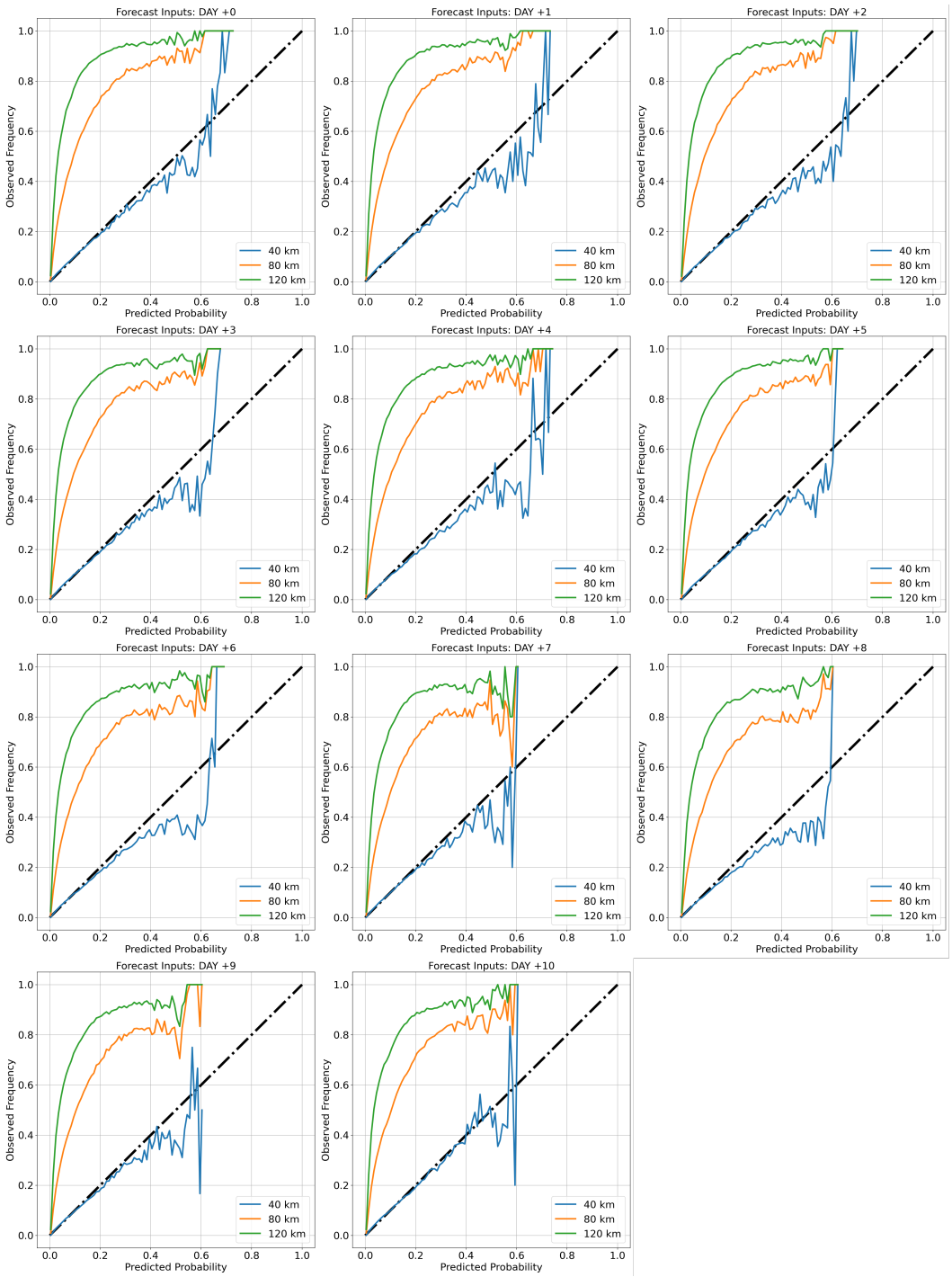


Figure 7.11: Reliability Plots for Day 0 - 10 for 'Pixel Label Forecast Inputs' Model.

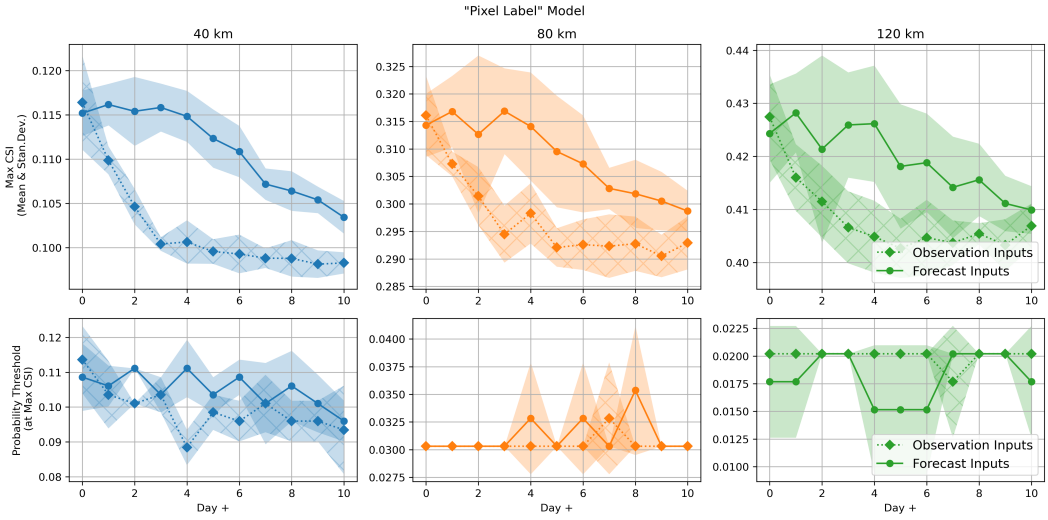


Figure 7.12: Max CSI Values & Probability Thresholds for the "Pixel Label" Model. Max CSI values are drawn from the Individual Performance Plots contained in Figures 7.7 and 7.8. The Probability Threshold is the probability cutoff from the Individual Performance Plot at which Max CSI is achieved.

model, there is a large amount of overlap between the two input approaches for all three neighborhoods. From day 0 to day 10, the trend of the probability thresholds depicted in Figure 7.9 stays relatively flat for the 80km and 120km neighborhoods and demonstrates some decreasing trend for the 40km neighborhood.

I chose to use the "Pixel Label" model with forecast inputs measured at 40km for use in the case studies because the forecast inputs model produced slightly higher CSI values than the observed inputs model and because the 40km measurement of reliability demonstrated the most reliability for this experiment.

7.3 Top Model Performance

Figure 7.13 describes the reliability achieved for each top model as well as the Max CSI achieved and the probability thresholds at which the Max CSI was achieved for

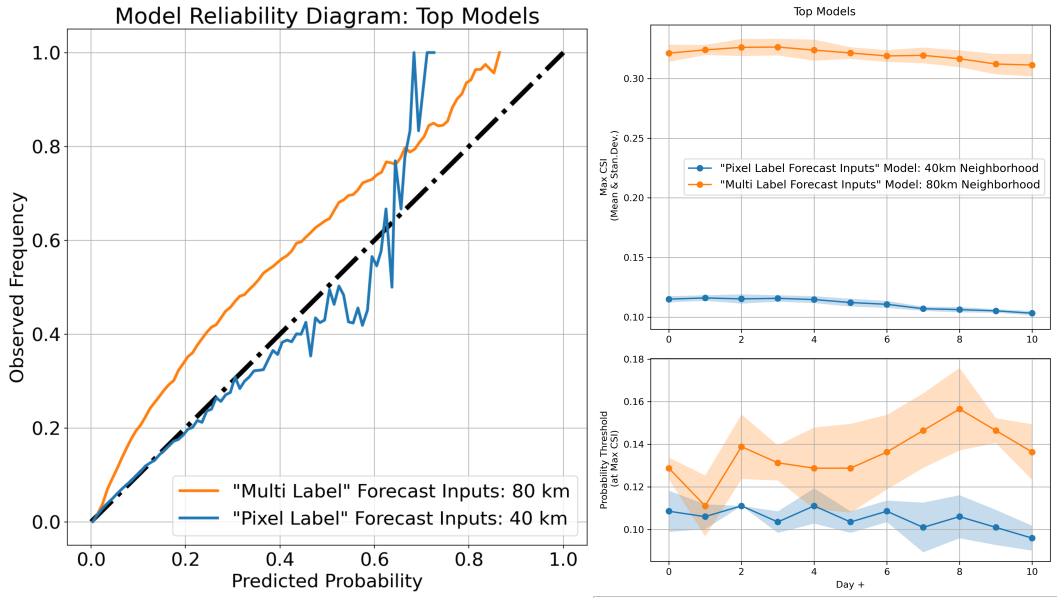


Figure 7.13: Comparing Top Models

this experiment. The UNet3+ is capable of offering reliability at two different levels of precision. If the prediction must be right on the pixel of interest, users can expect a CSI between .1 and .15 and should refer to the "Pixel Label" model results in the following case studies. If the prediction can fall within one pixel of the pixel of interest, users can expect a CSI around .3 and should refer to the "Multi Label" model results in the following case studies.

7.4 Case Studies

In this section, I describe three case studies of impactful fires. I compare model performance to the SPC Probability of Wildfire Climatology. While the climatology represents only a small portion of the data a fire weather forecaster would consider in creating a forecast, it does represent the closest analog to my work currently in operational use by fire weather forecasters. Please note that the color scheme for the 1000 acre product is different than the 5000 acre product. The model results color

scheme is designed to mimic the color scheme used for the 1000 acre product. Also note that the climatology is not available for all days so, where necessary, I select the closest date prior to the date of interest.

My case study fires are sourced from the 2018 California fire season and are the Camp Fire, the Carr Fire, and the Woolsey Fire. Described in Nauslar (2019), the Camp and Carr fires represent the deadliest 2018 wildfires and two of the largest wildfires in California history. The Camp, Carr, and Woolsey fires are described as "three of the eight most destructive wildfires" and together cost an estimated "$27 billion (U.S. dollars) in insured losses, suppression, and cleanup" (Nauslar 2019).

It is important to note that the 2018 year is part of the test set for my models and, therefore, represents data unseen by the model during training or validation. This makes these three fires a good test of my models' ability to generalize to future data. For the case studies, I showcase model performance for only the forecast input models as they demonstrated slightly higher CSI values than the observed inputs models, as described in the previous section.

7.4.1 Case Study: The Camp Fire

The Camp Fire was discovered on the morning of November 8, 2018 in Butte County in Northern California. The Camp Fire was caused by a downed electrical transmission line operated by the Pacific Gas and Electric Company (PG&E) and burned a total of 153,336 acres, including the town of Paradise, California, before it was contained on November 25, 2018.

Figure 7.14 depicts the SPC Probability of Wildfire Climatology for fires greater than 1000 acres and greater than 5000 acres for November 7th, the closest available date to the discovery date of the Camp Fire, and for October 29th, the date 10 days prior to the discovery date of the Camp Fire.

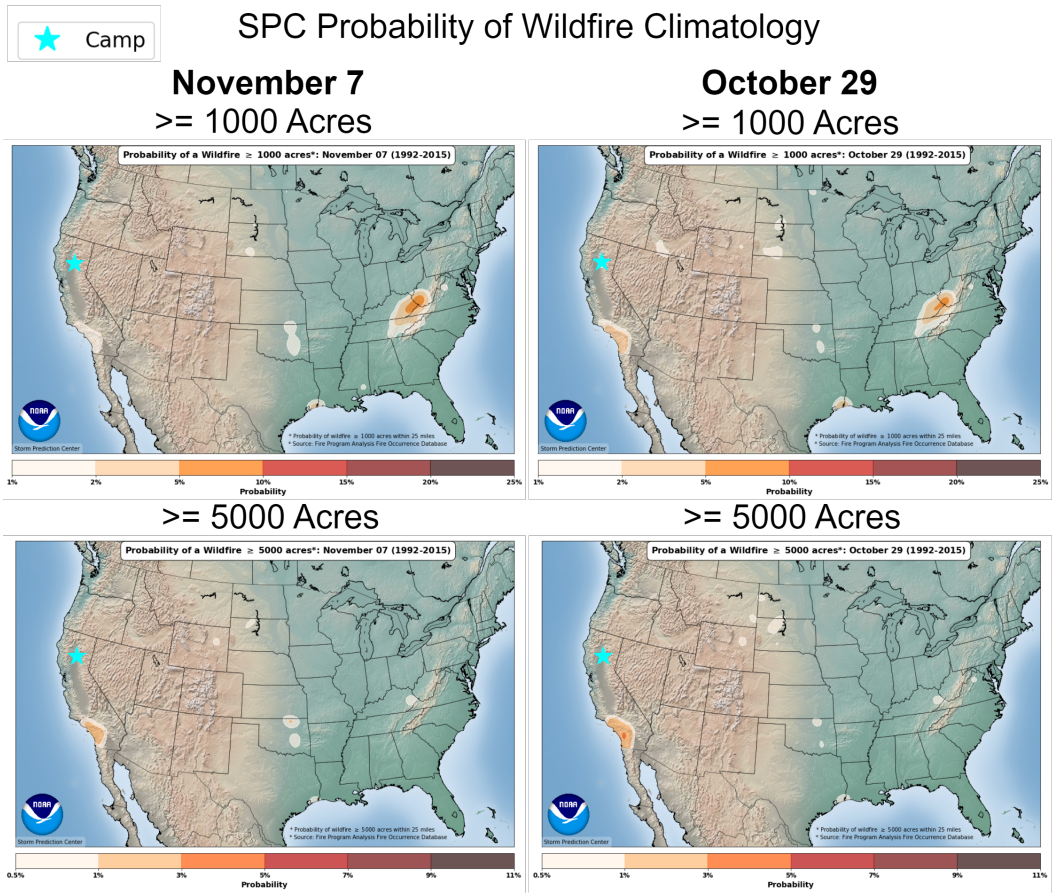


Figure 7.14: Camp Fire: SPC Probability of Wildfire Climotology

By presenting these two dates, I seek to answer two questions:

1) What can the climatology tell us about the date of interest ***on*** the date of interest (does the climatology recognize same day fire probability)?

2) What could the climatology have told us about the date of interest ***ten days prior*** to the date of interest or how much lead time could the climatology have offered (within the extent of this experiment)?

Visual inspection shows us that the climatology depicts no probability of wildfire for the area of discovery of the Camp fire, marked with an aqua star, either on the date of interest or ten days prior to the date of interest.

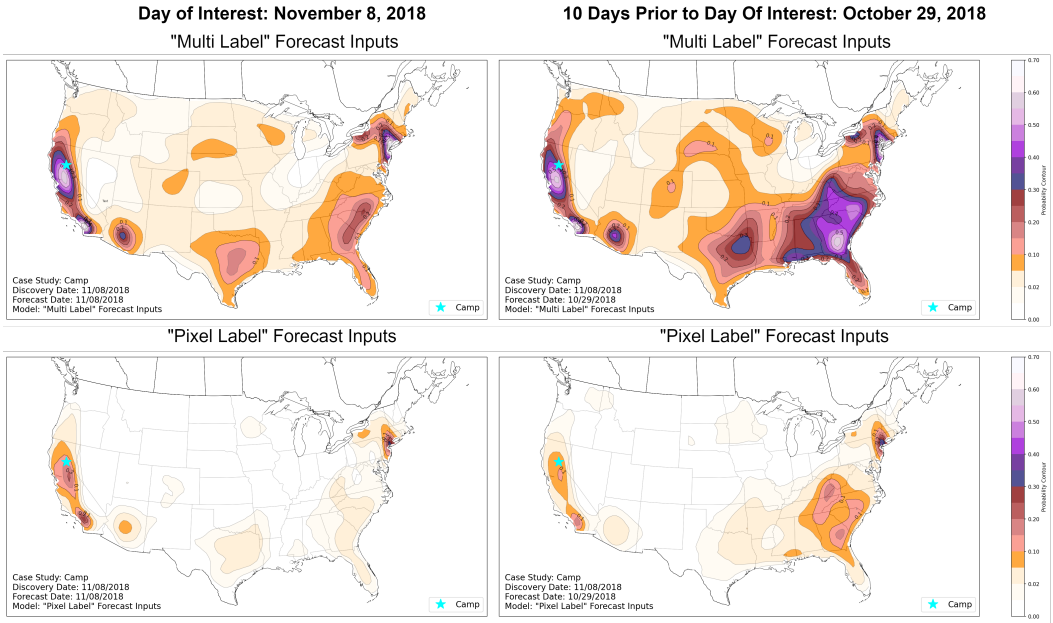


Figure 7.15: Camp Fire: "Multi Label" Forecast Inputs Model and "Pixel Label" Forecast Inputs Model

Figure 7.15 depicts the predictions from the "Multi Label" forecast inputs model and the "Pixel Label" forecast inputs model. Figure 7.15 depicts the model results created on November 8, 2018 (the date of discovery) and on October 29, 2018 (ten days prior to the date of discovery). The "Multi Label" model produced a probability of wildfire above .3 for both the date of interest and ten days prior to the date of interest. The "Pixel Label" model produced a probability of wildfire above .1 for both the date of interest and ten days prior to the date of interest.

In summary, as described in Table 7.3, in the Camp Fire case study, both the "Pixel Label" model and the "Multi Label" model produced higher probabilities of wildfire than the SPC Probability of Wildfire Climatology and were able to do so ten days in advance of the discovery date of the Camp Fire. Additionally, all four methods saw a constant average probability of wildfire from ten days prior to the day of interest.

Case Study :	*Camp Fire*		*Probability of Wildfire*
Method	*Approach*	*Discovery Date*	*Ten Days Prior*
SPC Climatology	1000 Acre	0	0
SPC Climatology	5000 Acre	0	0
UNet3+	"Multi Label" Forecast Inputs	.3-.4	.3-.35
UNet3+	"Pixel Label" Forecast Inputs	.1-.15	.05-.15

Table 7.3: Camp Fire Probability of Wildfire

7.4.2 Case Study: The Carr Fire

The Carr Fire was reported the afternoon of July 23, 2018 and was caused by the rim of a flat tire creating sparks against asphalt. The Carr Fire burned 229,651 acres in Northern California before being contained on August 30, 2018.

Figure 7.16 depicts the SPC Probability of Wildfire Climatology for July 22nd, the closest available date to the discovery date of the Carr Fire, and for July 13th, the date ten days prior to the discovery date of the Carr Fire.

For the 1000 acre product, ten days prior to the Carr Fire discovery date, a probability of wildfire between .01 and .05 was indicated. For the day of discovery for the Carr Fire, a probability of wildfire between .02 and .05 was indicated. As I move from ten days prior to the day of the event, the average probability of wildfire indicated by the 1000 acre product increased from $.03 = (.01 + .05)/2$ to $.035 = (.02 + .05)/2$.

For the 5000 acre product, ten days prior to the date of interest, a probability of wildfire between .005 and .03 is indicated. For the date of interest, a probability of wildfire between .01 and .03 is indicated. The average probability of wildfire indicated by the 5000 acre product increased from .0175 to .02.

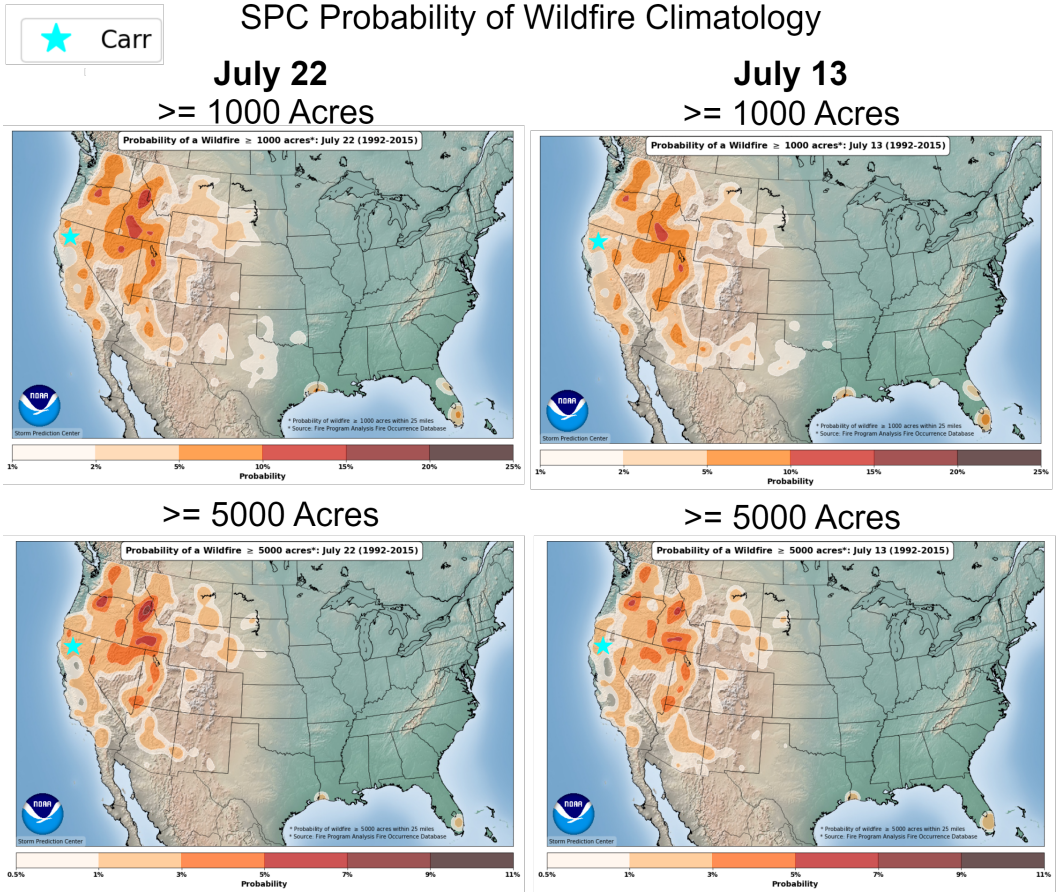


Figure 7.16: Carr Fire: SPC Probability of Wildfire Climotology

Figure 7.17 depicts the predictions from the "Multi Label" forecast inputs model and the "Pixel Label" forecast inputs model for the date of discovery of the Carr Fire and the date ten days prior to the date of discovery.

The "Multi Label" model produced a probability of wildfire between .55 and .65 for both the date of interest and ten days prior to the date of interest. As I move from ten days prior to the day of interest, the average probability of wildfire remained constant at .5.

The "Pixel Label" model, ten days prior to the date of interest, produced a probability of wildfire between .15 and .2. For the date of interest, the "Pixel Label" model

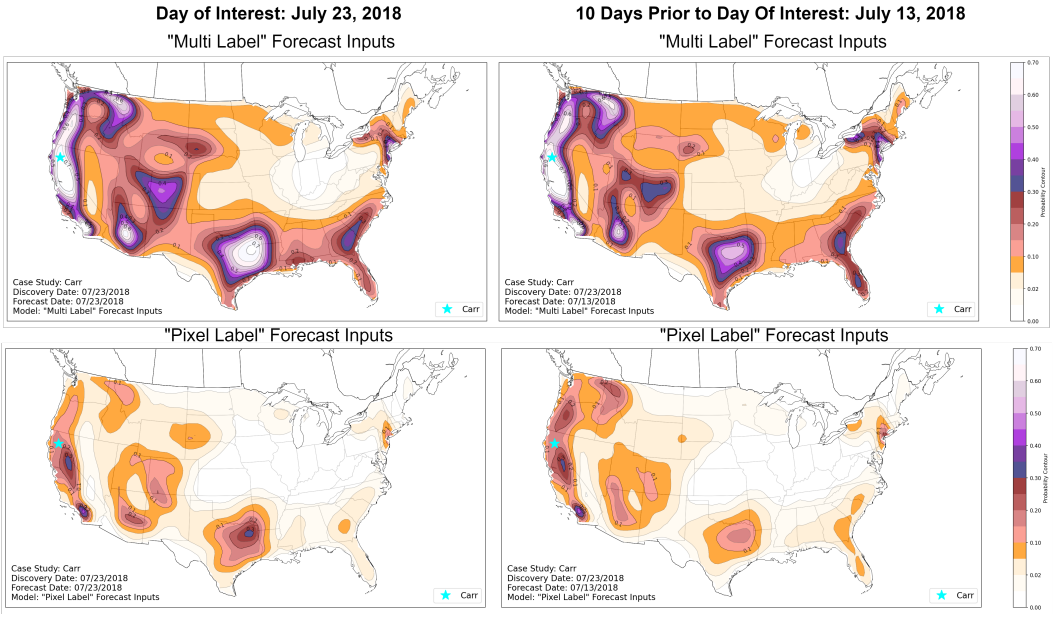


Figure 7.17: Carr Fire: "Multi Label" Forecast Inputs Model and "Pixel Label" Forecast Inputs Model

Case Study :	*Carr Fire*		*Probability of Wildfire*
Method	*Approach*	*Discovery Date*	*Ten Days Prior*
SPC Climatology	1000 Acre	.02-.05	.01-.05
SPC Climatology	5000 Acre	.01-.03	.005-.03
UNet3+	"Multi Label" Forecast Inputs	.55-.65	.55-.65
UNet3+	"Pixel Label" Forecast Inputs	.2-.25	.15-.2

Table 7.4: Carr Fire Probability of Wildfire

produced a probability of wildfire between .2 and .25. As I move from ten days prior to the day of interest, the average probability of wildfire increased from .175 to .225.

In summary, as described in Table 7.4, in the Carr Fire case study, both the "Pixel Label" model and the "Multi Label" model produced higher probabilities of wildfire than the SPC Probability of Wildfire Climatology and were able to do so ten days in advance of the discovery date of the Carr Fire. Additionally, both climatology

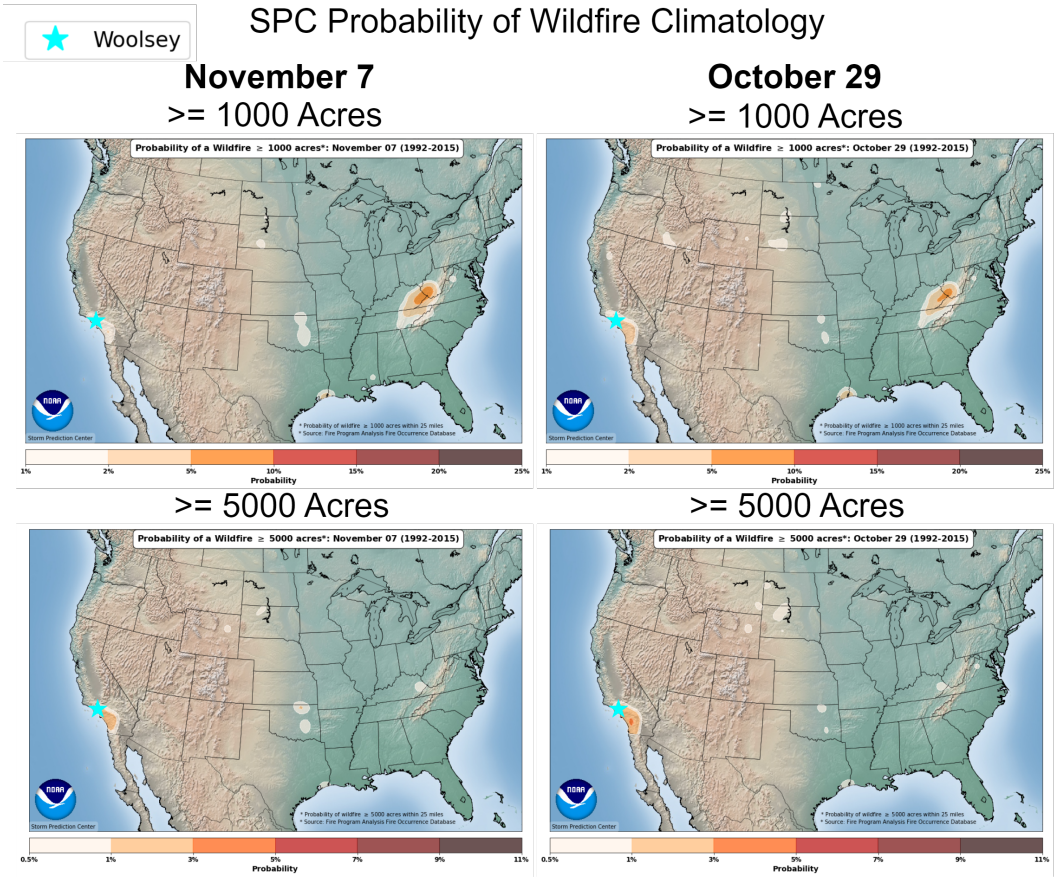


Figure 7.18: Woolsey Fire: SPC Probability of Wildfire Climotology

methods and the "Pixel Label" model saw the average probability of wildfire increase from ten days prior to the day of interest.

7.4.3 Case Study: The Woolsey Fire

The Woolsey Fire was discovered on November 8, 2018 in the Woolsey Canyon in Los Angeles, California. The Woolsey Fire was caused by electrical equipment owned and operated by Southern California Edison (SCE). While it only burned 96,949 acres before containment, hundreds of houses in Malibu and state and national parklands were damaged by the fire.

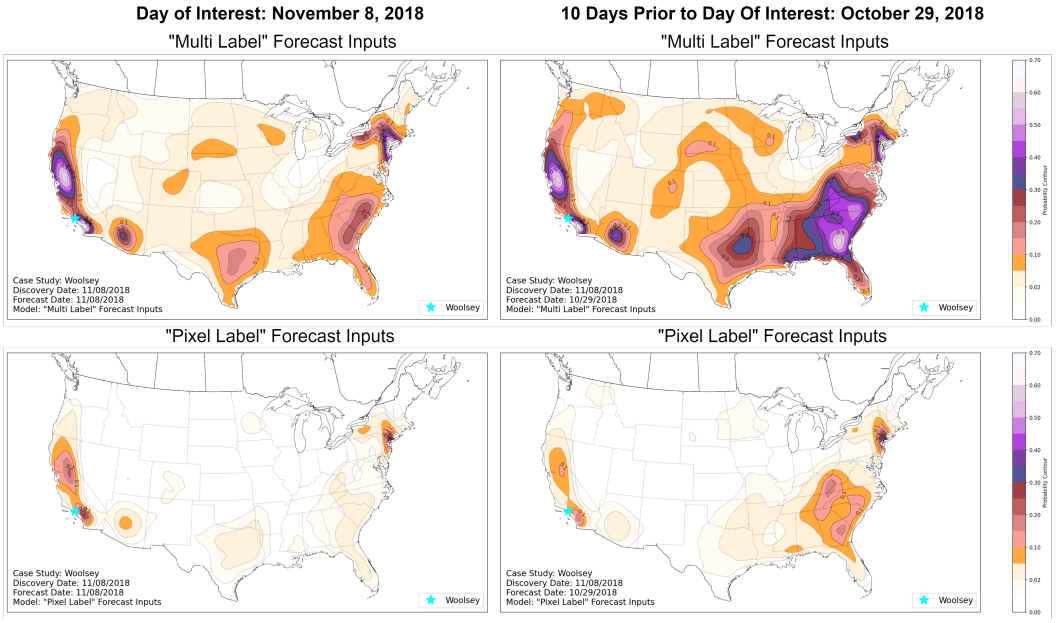


Figure 7.19: Woolsey Fire: "Multi Label" Forecast Inputs Model and "Pixel Label" Forecast Inputs Model

Figure 7.18 depicts the SPC Probability of Wildfire Climatology for November 7th, the closest available date to the discovery date of the Woolsey Fire, and for October 29th, the date ten days prior to the discovery date of the Woolsey fire. Please note, the Woolsey fire shares its discovery date with the Camp fire discussed previously in this section.

For the 1000 acre product, the average probability of wildfire decreased from .03 to .015 as I move from ten days prior to the date of interest to the date of interest. For the 5000 acre product, the average probability of wildfire decreased from .02 to .0175.

Figure 7.19 depicts the predictions from the "Multi Label" forecast inputs model and the "Pixel Label" forecast inputs model for the date of discovery of the Woolsey fire and the date ten days prior to the date of discovery. As I move from ten days prior to the date of interest, the probability of wildfire produced by the "Multi Label"

Case Study :	*Woolsey Fire*		*Probability of Wildfire*
Method	*Approach*	*Discovery Date*	*Ten Days Prior*
SPC Climatology	1000 Acre	.01-.02	.01-.05
SPC Climatology	5000 Acre	.005-.03	.01-.03
UNet3+	"Multi Label" Forecast Inputs	.2-.3	.3-.4
UNet3+	"Pixel Label" Forecast Inputs	.05-.1	.05-.1

Table 7.5: Woolsey Fire Probability of Wildfire

model decreased from .035 to .03. The probability of wildfire produced by the "Pixel Label" model stayed constant at .06.

In summary, as described in Table 7.5, in the Woolsey Fire case study, both the "Pixel Label" model and the "Multi Label" model produced higher probabilities of wildfire than the SPC Probability of Wildfire Climatology and were able to do so ten days in advance of the discovery date of the Woolsey Fire. Unique to the Woolsey fire, for both climatology methods and the "Multi Label" model, the average probability of wildfire decreased as I move from ten days prior to the day of interest.

Chapter 8

Lessons Learned from Predicting Fire Occurrence using using GEFS

In this section, I discuss some of the important lessons I took away from modeling fire occurrence using GEFS data. First, I describe multicollinearity within the GEFS dataset and the results of the permutation variable importance studies. I then compare the four models focused on in this work based on two case studies: the largest natural-cased fire of the 2018 fire season, the South Sugarloaf fire (revisited from Chapter 5), and the largest human-caused fire of the 2018 fire season, the Camp fire (revisited from Chapter 7). I describe how each model performs on the case study fire in isolation, the largest fires discovered on that day, and all fires discovered on that day regardless of size.

8.1 Multicollinearity within the GEFS Dataset

As described in Chapter 6, the ability of a model to provide feedback on variable importance depends on the correlation present between the input variables for the model. In Chapter 6, I found that the gridMET dataset, while commonly used operationally, would not be able to provide variable importance feedback due to the presence of structural multicollinearity, whereby correlation between variables arises from using one model term to create another model term. This finding motivated my

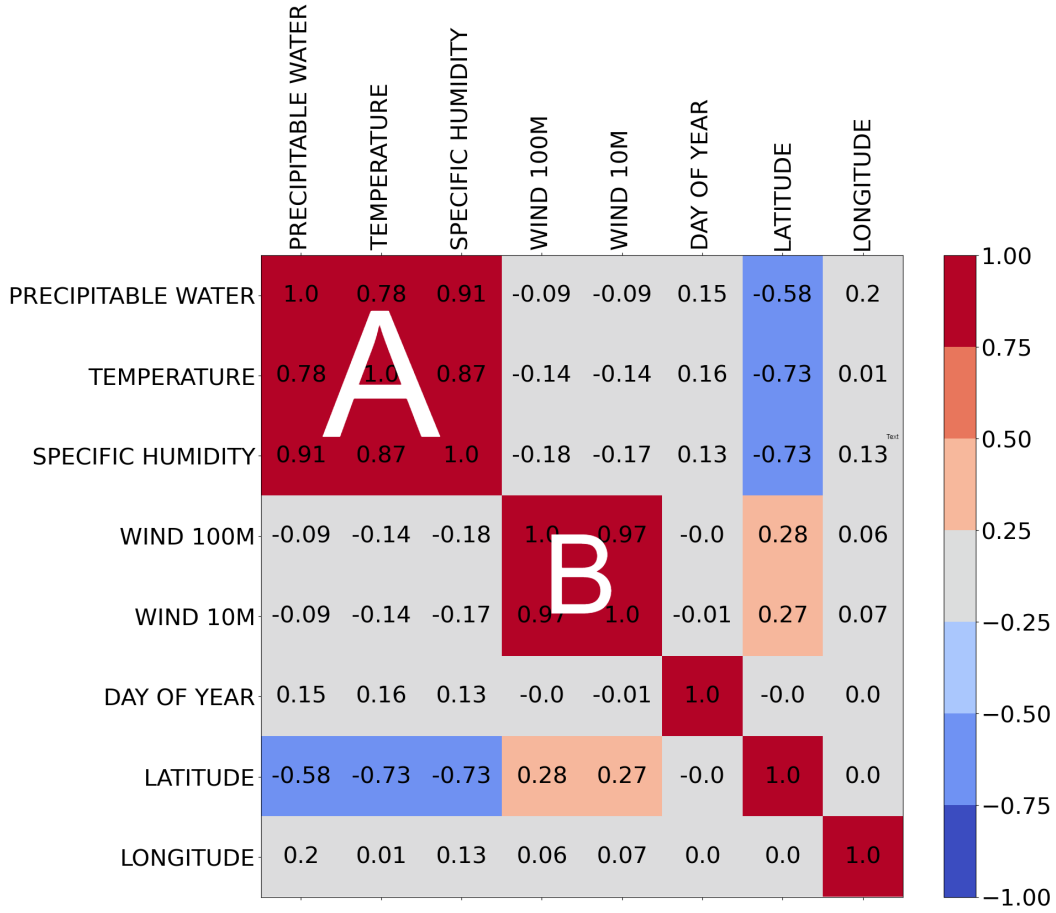


Figure 8.1: GEFS Input Data Correlation

choice to move from using gridMET as the input dataset to using GEFS as the input dataset. I hypothesize that using observed and forecast weather variables can offer less correlation between variables as compared to fire weather variables and can do so while enabling comparable or improved model performance when compared to the gridMET models.

Figure 8.1 depicts the correlation present in the GEFS dataset. There is less correlation than in the gridMET dataset and the correlation present is localized to distinguishable sets of variables. Conversely, in the gridMET dataset most variables were highly correlated with each other and isolating one variable from another was not

possible. In the GEFS dataset, there are two clear groups of positive intercorrelated variables. Group A is comprised of precipitable water, temperature, and specific humidity and Group B is comprised of wind magnitude measured at 10 meters and 100 meters. There is also a strong negative correlation between latitude and the Group A variables.

Because all three variables from Group A, for example, are highly correlated with each other, two things are true. First, variable importance cannot speak to how one variable from Group A contributes to model performance relative to another. Instead, all three variables from Group A must be removed together to speak to how they contributed to model performance as a group. Second, because of the intercorrelated nature of the Group A variables, I could choose to retain only one of the three variables and maintain comparable model performance. These two points are also true for Group B. This means, theoretically, instead of using eight variables, I could select one variable from Group A and one variable from Group B and, in addition to the Latitude, Longitude, and Day of Year variables, achieve comparable model performance using four variables to the performance described for the GEFS models in Chapter 7.

8.2 Permutation Variable Importance (PVI)

Now that the correlation within the input dataset can be isolated, I can perform permutation variable importance studies. PVI is a method to estimate how much model performance depends on individual features used as inputs to the model (Breiman 2001; Fisher et al. 2019). To estimate variable importance, I use a series of permutation importance studies for the day one "Multi Label" model. I describe only the

Forecast Inputs results here as a continuation of the work described in Chapter 7. PVI is assessed for the test dataset for each experiment.

The process for running a permutation importance study is as follows:

1. The model performance is first evaluated on a test set using a chosen metric, in this case, Max CSI, to establish a baseline for model performance.

2. Each variable, or in this case, a group of variables, is randomly permuted across all samples to break the relationship between the input variable and the label. In this study, I randomized along both the time and space dimensions by randomly all the pixels within an image after randomly shuffling the images.

3. The model's performance is then measured again on the test set using the previous model and the permuted data.

4. The difference between the baseline model performance and the permuted model performance indicates the dependence of that model's performance on that input variable, or group of variables, as in this case.

As discussed in the previous section, the variables contained in Group A are highly correlated with each other, as are the variables in Group B, so they can not be assessed individually and must be removed together to measure variable contribution. Described in in Figure 8.2 is the impact to model performance resulting from the randomization of all variables in Group A and then, separately, the randomization of all variables in Group B. The model was more impacted by the randomization of Group A than by the randomization of Group B. This suggests that variables that have a direct impact on the state of the fuels, which precipitation and temperature do, have a larger impact on the model's ability to predict wildfire than variables that have a more direct relationship with fire spread, which wind does.

Figure 8.3 compares two permutation experiments, one that focuses on time and another that focuses on space. For the time experiment, the Day of Year variable

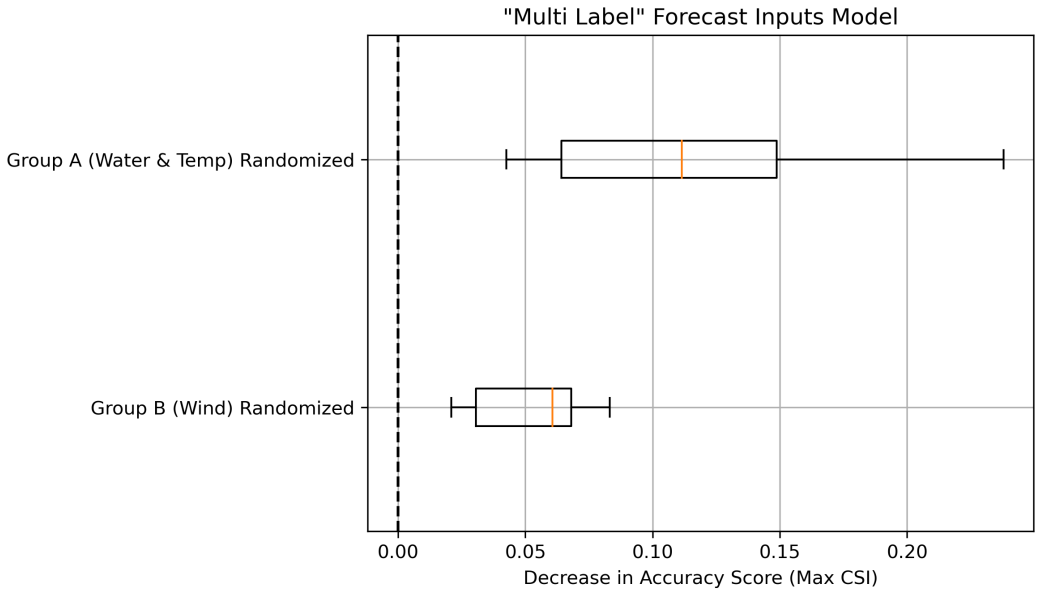


Figure 8.2: Permutation Variable Importance: Group A VS. Group B

is randomized. For the space experiment, the Latitude and Longitude variables, which represent the centroids of the grid cells, are randomized. The impact of the randomization of the Latitude and Longitude variables may speak to regional effects not yet accounted for within the model. The addition of fuel input variables, not in the form of fire danger variables, may help to address this gap. This must be reserved for future work.

The impact of randomization of the Day of Year variable may speak to season effects not yet account for within the model. It may also be tied to the Day of Year encoding that uses Julian Day. Julian day is subject to wrap-around effects in which values near the start of the year and values near the end of the previous year are temporally close but numerically distant. This can distort continuous relationships and is particularly impactful for trends that occur near the boundary between years. The fire season is traditionally constrained to warmer months and this perhaps insulates my work from the larger impacts of wrap-around effects. One option for

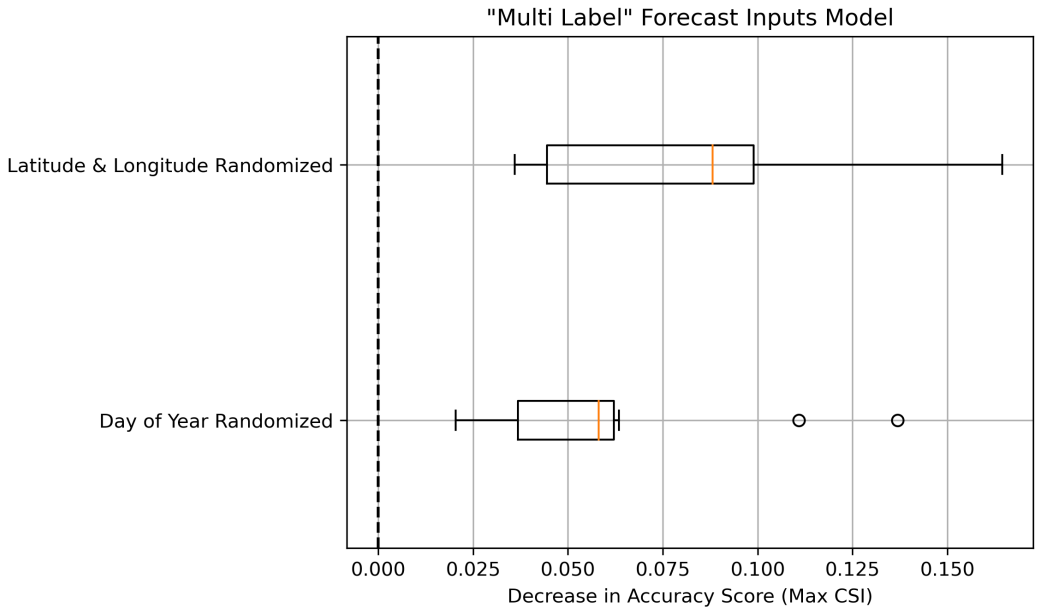


Figure 8.3: Permutation Variable Importance: Time VS. Space

future work would be to explore solutions to wrap-around effects such as using cyclic transformations or an annual indicator in addition to the Day of Year variable.

8.3 Case Studies

Over the course of this work, I attempted multiple approaches to modeling fire occurrence including varying the input dataset used, varying how the input data was structured, varying how the labels were structured, and varying what fire occurrences were included in the labels. Table 8.1 describes the four main approaches I took that will be compared in the following case studies. Although these are not the only approaches I attempted, they are the closest to what I believe an operational model will resemble.

For the sake of comparison, all models discussed in this section use observation inputs. I acknowledge that this is a departure from Chapter 7, where I focused on

Model	*Dataset*	*Input Type*	*Label Type*	*Fires Labeled*
"Large Lightning" Model	gridMET	Observed	Multi Label	**Only Large Lightning Fires**
"All Fires" Model	**gridMET**	Observed	Multi Label	All Fires
"Multi Label" Observed Inputs Model	**GEFS**	Observed	Multi Label	All Fires
"Pixel Label" Observed Inputs Model	GEFS	Observed	**Pixel Label**	All Fires

Table 8.1: Modeling Approaches

the GEFS models which used forecast inputs but the gridMET models only had access to observation inputs. I believe that limiting the GEFS models to observation inputs as well will broaden the foundation for comparison. All models use the space domain of CONUS. The gridMET models use a time domain of 2000 to 2020 and the GEFS models use a time domain of 2000 to 2019. I assume that the single year difference between the two time domains does not have a significant impact on the analysis presented in the case studies. For this section, I use the South Sugarloaf fire, the largest natural-caused fire from the 2018 fire season, and the Camp fire, the largest human-caused fire from the 2018 fire season, to compare model performance. The case studies focus on the year 2018, which was part of the test set for both the gridMET models and the GEFS models. In the following case studies, I depict the model predictions as images of CONUS with areas of probability identified to simulate what forecasters would see should they use the predictions of my models to help inform their intuition.

The comparisons drawn in the case studies below rely on visual analysis of the CONUS maps, the probability fields produced by the models, and the observed fire occurrences. I contextualize the case study evaluations using two concepts: a subjective evaluation of localization, intended to speak to the proximity of the predictions to each other, and a subjective evaluation of probability of wildfire, intended to speak to the proximity of the predictions to the fire occurrence(s) of interest. Both localization and probability will be ranked for each model to aid in establishing a model

Probability

Classification	*Description*
Bad	Areas of highest probability are not near the Element(s) of Interest.
Medium	Areas of highest probability are near the Element(s) of Interest but also elsewhere in CONUS.
Good	Areas of highest probability are near the Element(s) of Interest.

Table 8.2: Probability

Localization

Classification	*Description*
Bad	Predictions are widely dispersed.
Medium	Predictions fall between widely dispersed and densely concentrated.
Good	Predictions are densely concentrated.

Table 8.3: Localization

with the highest relative performance. Context for rankings are defined in Tables 8.2 and 8.3.

8.3.1 South Sugarloaf Fire

The South Sugarloaf Fire was the largest natural-caused fire of the 2018 fire season. The South Sugarloaf Fire was discovered on August 17, 2018 in the Humboldt National Forest in northern Nevada and burned approximately 232,906 acres before it was contained (State of Nevada 2018). Figure 8.4 depicts model predictions for the day of discovery, August 17, 2018, of the South Sugarloaf Fire. The discovery location of the South Sugarloaf Fire is indicated with an aqua-colored star.

The "All Fires" model, shown in the upper left hand corner of Figure 8.4, depicts a widely dispersed probability field relative to the other models depicted. While the probability of wildfire predicted by the model for the location of discovery of the South Sugarloaf Fire is on the medium to high side of the range of probability values the "All Fires" model is able to predict for the day of discovery, the same probability value is predicted for large portions of CONUS. The area of highest probability predicted by

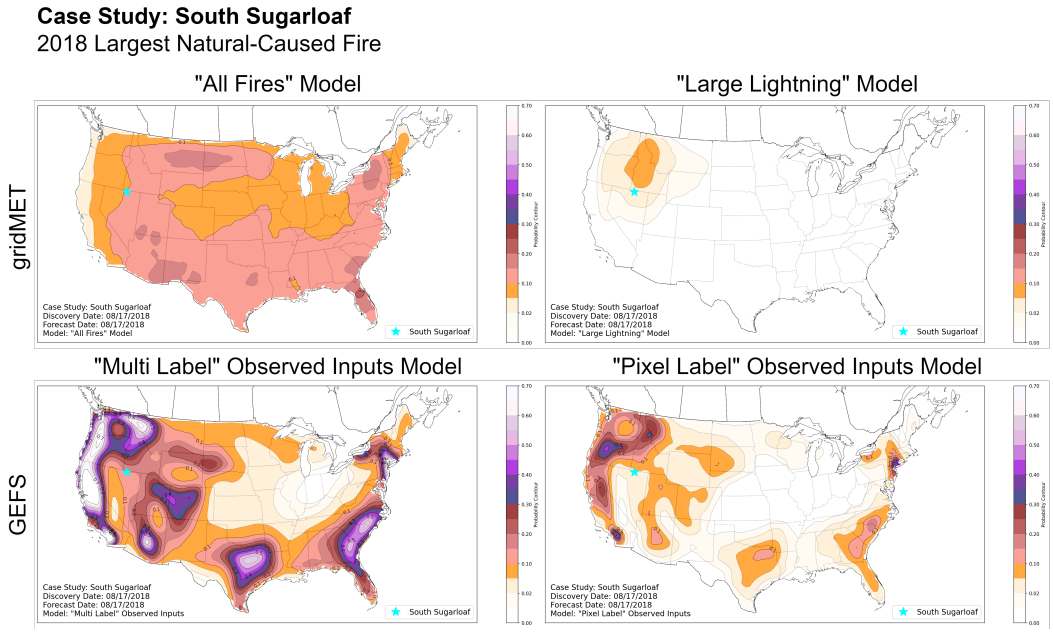


Figure 8.4: South Sugarloaf Individual Fire Case Study

the model is located in Florida on this day, which is not near the location of discovery of the South Sugarloaf Fire. The lack of agreement between the highest probabilities and the element of interest make the "All Fires" model less useful for this case study and the widely dispersed probabilities suggest low utility in general.

The "Large Lightning" model, shown in the upper right hand corner of Figure 8.4, depicts a densely concentrated probability field relative to the other models depicted. The probability of wildfire predicted by the model for the location of discovery of the South Sugarloaf Fire is on the medium to high side of the range of probabilities predicted by the "Large Lightning" model for the day of discovery. The highest probabilities predicted by the model are close to the location of the discovery of the South Sugarloaf Fire. The concentrated probabilities combined with the agreement between the highest probabilities and the element of interest make the "Large Lightning" model a useful model to reference for this case study. This is to be expected,

given that the South Sugarloaf Fire is a large, natural-caused fire and the "Large Lightning" model was designed to predict large, natural-caused fires.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.4, depicts a varied probability field in which high probability areas are more densely concentrated and low probability areas are more widely dispersed with almost all areas of CONUS having been assigned some probability of wildfire. The probability of wildfire predicted by the model for the location of discovery of the South Sugarloaf Fire is on the low to medium side of the range of probability predicted by the "Multi Label" model for the day of discovery. Some of the highest probabilities areas predicted by the model are close to the location of discovery of the South Sugarloaf fire but many are located in other areas of CONUS not near the location of discovery. The lack of strong signal at the location of discovery suggests that this is not a useful model for this case study. However, the strong signals elsewhere in CONUS suggest that this model might be capturing other valuable information.

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.4, also depicts a varied probability field in which high probability areas are more densely concentrated and low probability areas are more widely dispersed, however, not all areas of CONUS are assigned a probability of wildfire by this model. The probability of wildfire predicted by the "Pixel Label" model for the location of discovery of the South Sugarloaf Fire is on the low side of the range of probabilities predicted by the model for the day of discovery. Some of the highest probabilities areas predicted by the model are close to the location of discovery of the South Sugarloaf fire but many are located in other areas of CONUS not near the location of discovery. The lack of strong signal at the location of discovery suggests that this is not a useful model for this case study. However, the strong signals elsewhere in CONUS suggest that this model might be capturing other valuable information. Additionally, the fact that the

South Sugarloaf Fire

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Large Lightning	**2**	**1**	**1**
Multi Label	5	2	3
Pixel Label	6	4	2
All Fires	7	3	4

Table 8.4: South Sugarloaf Probability and Localization Rank

"Pixel Label" model did not predict probability of wildfire for all areas of CONUS may suggest that the "Pixel Label" model is more discerning than the "Multi Label" model, which predicted some level of probability of wildfire for all of CONUS.

For each of the four models, I assigned a force rank based on visual analysis of the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank). The sum resulting from the two rankings indicates which model best captures the element(s) of interest for the case study with lower rankings indicating a model that is capturing the element(s) well and higher rankings indicating a model that is capturing the element(s) poorly. Rankings are described in Table 8.4.

8.3.1.1 The Largest Fires Discovered on August 17, 2018

Figure 8.5 depicts model predictions for the day of discovery of the South Sugarloaf Fire, August 17, 2018. The location of discovery of the South Sugarloaf Fire is indicated with an aqua-colored star. The probability fields depicted in this section are the same probability fields presented in the previous section. For this section, instead of depicting only the location of discovery of the South Sugarloaf Fire, all fire occurrence discovered on August 17, 2018 are depicted on the map. The size

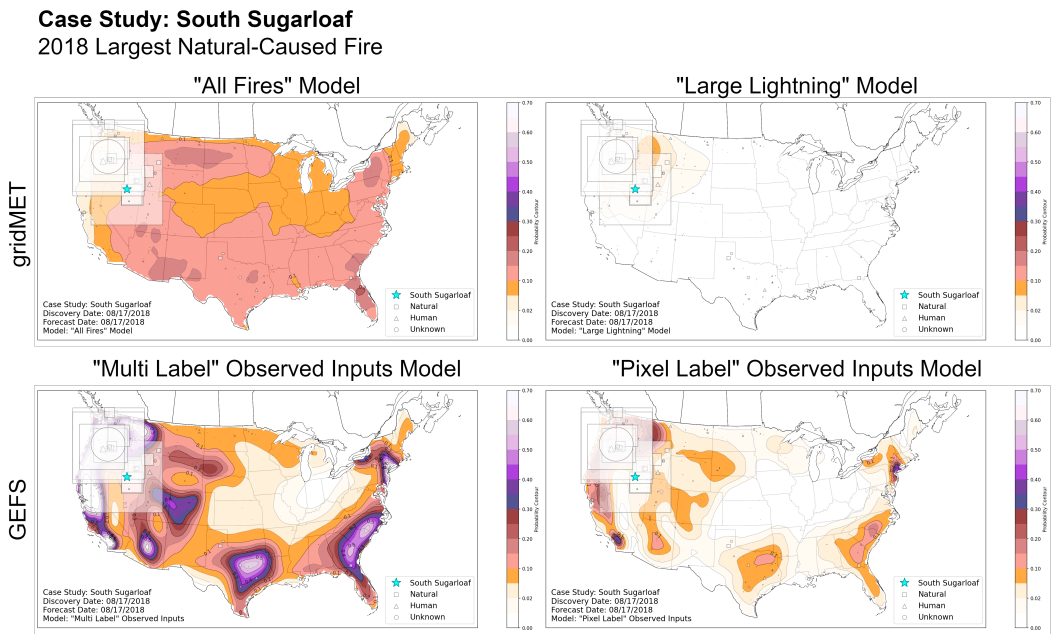


Figure 8.5: South Sugarloaf Large Fires Discovered Case Study

of the marker indicates the relative final fire size of the wildfire. The shape of the marker indicates the cause of the wildfire with squares representing natural caused fires, triangles representing human caused fires, and circles representing fires with unknown cause. The majority of large fires discovered on August 17, 2018 are large natural caused fires located in the northwestern region of CONUS.

For the largest fires discovered on the day of discovery, the "All Fires" model, shown in the upper left hand corner of Figure 8.5, predicts probabilities from low to medium within the range of probabilities predicted by the model for that day. The highest probabilities predicted for the day of discovery are not located near the large fires. As with the previous section, areas of low to medium probability cover most of CONUS and the highest probabilities predicted by the model are not near the large fires. The lack of agreement between the highest probabilities and the elements of interest indicate that this model is not useful for this case study.

The "Large Lightning" model, shown in the upper right hand corner of Figure 8.5, predicts medium to high probabilities for the largest fires and the highest probabilities predicted by the model are close to the largest fires. These two elements being in agreement indicates that the "Large Lightning" model is a useful model for this case study. Once again, this is to be expected, given that the largest fires discovered on this day are primarily natural-caused fires and the "Large Lightning" model was designed to predict large, natural-caused fires.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.5, predicts medium to high probabilities for the largest fires and some of the areas of highest probability are collocated with the largest fires. However, there are other areas of high probability that are not near the largest fires. The high probability near the largest fires indicates that the "Multi Label" model can be useful for this case study. The high probability values elsewhere in CONUS may suggest that this model has additional utility not highlighted by this case study.

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.5, predicts medium to high probabilities for the largest fires and most of the areas of highest probability are collocated with the largest fires. There are some areas of high probability not near the largest fires. The high probability near the largest fires indicates that the "Pixel Label" model can be useful for this case study. The high probability values elsewhere in CONUS may suggest that this model has additional utility not highlighted by this case study.

My subjective rankings for the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank) based on visual analysis are described in Table 8.5.

Largest Fires Discovered

August 17, 2018

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Large Lightning	**2**	**1**	**1**
Pixel Label	5	3	2
Multi Label	5	2	3
All Fires	8	4	4

Table 8.5: Largest Fires Discovered on August 17, 2024 Probability and Localization Rank

8.3.1.2 All Fires Discovered on August 17, 2018

Figure 8.6 depicts model predictions for the day of discovery of the South Sugarloaf Fire, August 17, 2018. The location of discovery of the South Sugarloaf Fire is indicated with an aqua-colored star. The probability fields depicted in this section are the same probability fields presented in the previous two sections. For this section, all fire occurrences discovered on August 17, 2018 are once again depicted on the map. Unlike the previous section, markers are no longer sized by the relative final fire size of the wildfire. Each fire occurrence marker is given the same size so as to promote equal visual prominence for all fire occurrences discovered on this day regardless of size or cause. The shape of the marker indicates the cause of the wildfire with squares representing natural caused fires, triangles representing human caused fires, and circles representing fires with unknown cause.

The "All Fires" model, shown in the upper left hand corner of Figure 8.6, predicts some probability of wildfire for all instances of fire occurrence discovered on this day. However, the diffuse nature of the majority of probability predictions made by the model make it difficult to know which parts of CONUS are the most important.

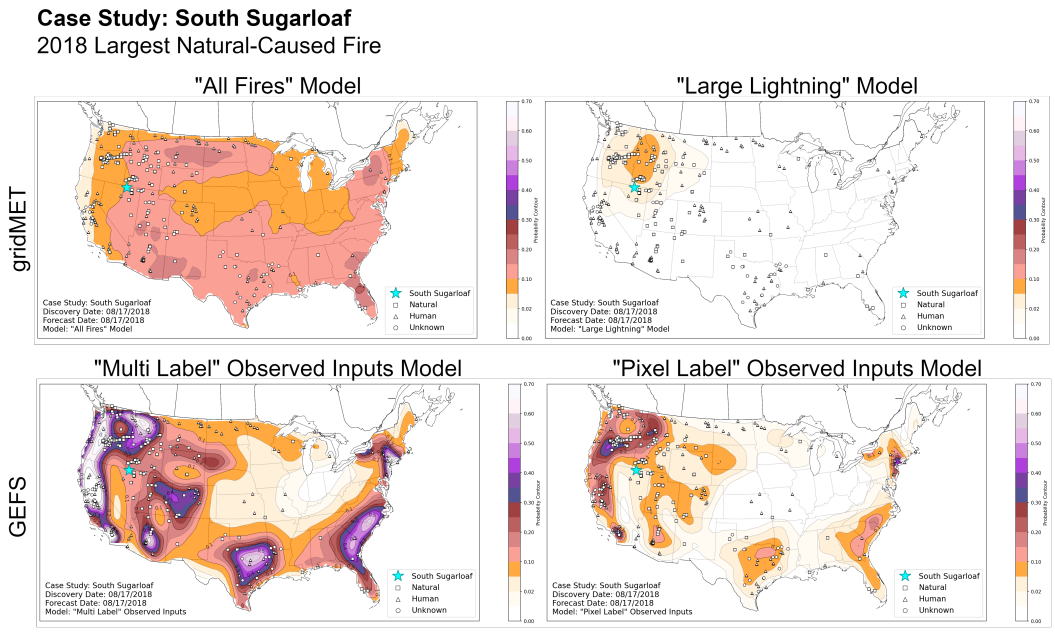


Figure 8.6: South Sugarloaf All Fires Discovered Case Study

While the "Large Lightning" model, shown in the upper right hand corner of Figure 8.6, predicts large natural-caused fires well, as demonstrated by the individual fire case study and the largest fires discovered on the day of discovery case study, it does miss fires that are either not large or not not natural-caused. This is to be expected, but it does present a notable blind spot that will continue to be apparent in the Camp Fire case studies.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.6, predicts medium to high probability for the majority of fire occurrences discovered on this day. The high probability areas are collocated with the majority of fire occurrences, with the two exceptions. There are six fires in the center of CONUS, five human caused fires and one natural caused fire, that the model predicted low probability for and the southeastern region of CONUS that has a medium to high probability area but not a large number of fires.

All Fires Discovered

August 17, 2018

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Pixel Label	**3**	**1**	**2**
Multi Label	5	2	3
Large Lightning	5	4	1
All Fires	7	3	4

Table 8.6: All Fires Discovered on August 17, 2024 Probability and Localization Rank

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.6, predicts medium to high probability for the majority of fire occurrences discovered on this day. The high probability areas are collocated with the majority of fire occurrences. The same six fires in the center of CONUS discussed for the "Multi Label" model were predicted poorly by this model which predicted no to low probability of wildfire. For the southeastern region of CONUS, where there are not many fires for this day, unlike the the "Multi Label" model, the "Pixel Label" model model predicted low to medium probability which perhaps makes it less incorrect.

My subjective rankings for the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank) based on visual analysis are described in Table 8.6

8.3.2 Camp Fire

The Camp Fire, discovered on November 8, 2018, in Butte County, Northern California, was ignited by a downed electrical transmission line operated by Pacific Gas and Electric Company (PG&E). Figure 8.7 depicts model predictions for the day of discovery, November 8, 2018, of the Camp Fire. The discovery location of the Camp

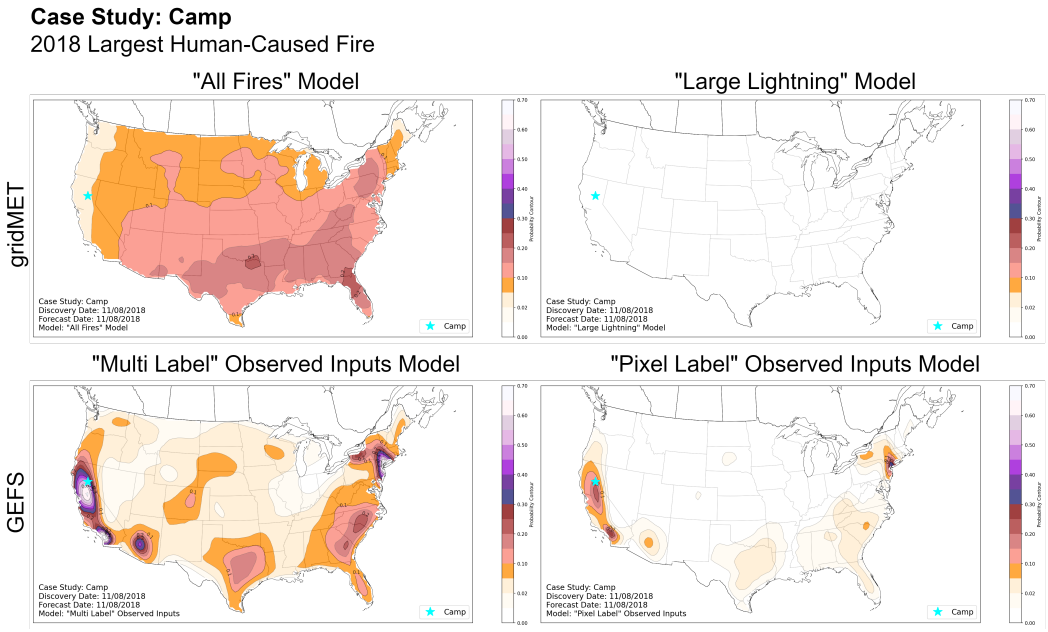


Figure 8.7: Camp Individual Fire Case Study

Fire is indicated with an aqua-colored star. The Camp Fire is a particularly interesting example because it was discovered in a month outside of the traditional fire season.

The "All Fires" model, shown in the upper left hand corner of Figure 8.7, continues to depict a broadly diffuse probability field. The model predicts a low probability of wildfire for the location of discovery of the Camp Fire and the highest areas of probability are not located near the location of discovery of the Camp Fire. The low probability prediction combined with the lack of agreement between the high probability areas and the element of interest suggest that the "All Fires" model is not useful for this case study.

The "Large Lightning" model, shown in the upper right hand corner of Figure 8.7, cannot be considered for this case study as it is unable to produce a signal. This is to be expected because the "Large Lightning" model is designed to predict large

Camp Fire

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Pixel Label	**3**	**2**	**1**
Multi Label	**3**	**1**	**2**
All Fires	6	3	3
Large Lightning	-	-	-

Table 8.7: Individual Fires Case Study: Camp Probability and Localization Rank

natural-caused fires only. In the case of the Camp Fire, which is a large human-caused fire, the "Large Lightning" model has no utility.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.7, depicts densely concentrated high probability predictions collocated with the location of the Camp Fire. The "Multi Label" model also predicts two small areas of similar probability, one in southern California, and the other in the northeastern region of CONUS. Low probability predictions from this model continue to be broadly diffuse.

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.7, like the "Multi Label" model, depicts densely concentrated probability predictions. Unlike the "Multi Label" model, the low probability areas predicted by the "Pixel Label" model are also spatially concentrated. For the location of discovery of the Camp Fire, the "Pixel Label" model predicted a medium to high probability. It predicted a similar level of probability for two other areas in CONUS, the same areas signalled by the "Multi Label" model, in southern California and in the northeastern region.

My subjective rankings for the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank) based on visual analysis are described in Table 8.7.

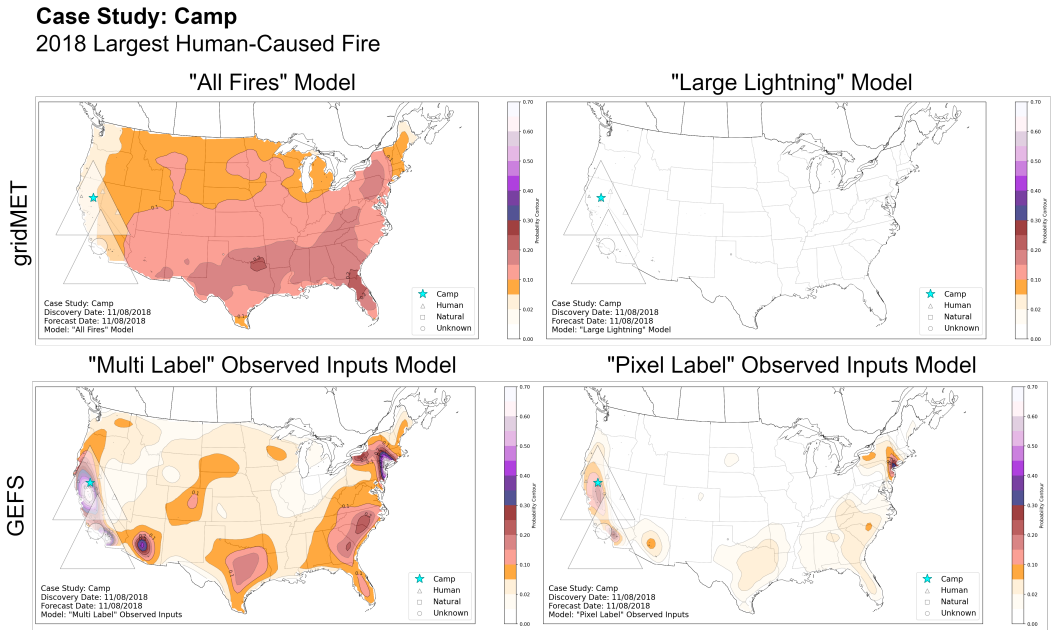


Figure 8.8: Camp Large Fires Discovered Case Study

8.3.2.1 The Largest Fires Discovered on November 8, 2018

Figure 8.8 depicts model predictions for the day of discovery of the Camp Fire, November 8, 2018. The location of discovery of the Camp Fire is indicated with an aqua-colored star. The probability fields depicted in this section are the same probability fields presented in the previous section. For this section, instead of depicting only the location of discovery of the Camp Fire, all fire occurrence discovered on November 8, 2018 are depicted on the map. The size of the marker indicates the relative final fire size of the wildfire. The shape of the marker indicates the cause of the wildfire with squares representing natural caused fires, triangles representing human caused fires, and circles representing fires with unknown cause. The majority of large fires discovered on November 8, 2018 are large human caused fires located in California.

The "All Fires" model, shown in the upper left hand corner of Figure 8.8, predicts low to medium probability for the largest fire occurrences discovered on this day.

The highest probabilities predicted by this model for this day are not near the largest fire occurrences. The lower probability predicted for the largest fires and the lack of agreement between the highest probability areas and the largest fires suggests that this model is not useful for this case study.

The "Large Lightning" model, shown in the upper right hand corner of Figure 8.8, is excluded from this case study but is presented here to in an attempt to fully illustrate the capabilities of the model.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.8, predicts high densely concentrated probability areas collocated with the largest fires discovered on this day. It also predict a small area of similar probability in the northeastern region and areas of medium probability in the southern and southeastern regions of CONUS which are not near the largest fires. The "Multi Label" model's low probability predictions continue to be broadly diffuse.

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.8, predicts medium to high densely concentrated probability areas for the largest fires discovered on this day. It predicts similar probability for one small area in the northeastern region and areas of low probability in the southern and southeastern regions of CONUS which are not near the largest fires. The low probability areas predicted by the "Pixel Label" model continue to be densely concentrated, unlike the predictions of the "Multi Label" model.

My subjective rankings for the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank) based on visual analysis are described in Table 8.8.

Largest Fires Discovered

November 8, 2018

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Pixel Label	**3**	**2**	**1**
Multi Label	**3**	**1**	**2**
All Fires	6	3	3
Large Lightning	-	-	-

Table 8.8: Largest Fires Case Study: Camp Probability and Localization Rank

8.3.2.2 All Fires Discovered on November 8, 2018

Figure 8.9 depicts model predictions for the day of discovery of the Camp Fire, November 8, 2018. The location of discovery of the Camp Fire is indicated with an aqua-colored star. The probability fields depicted in this section are the same probability fields presented in the previous two sections. For this section, all fire occurrences discovered on November 8, 2018 are once again depicted on the map. Unlike the previous section, markers are no longer sized by the relative final fire size of the wildfire. Each fire occurrence marker is given the same size so as to promote equal visual prominence for all fire occurrences discovered on this day regardless of size or cause. The shape of the marker indicates the cause of the wildfire with squares representing natural caused fires, triangles representing human caused fires, and circles representing fires with unknown cause.

The "All Fires" model, shown in the upper left hand corner of Figure 8.9, predicts low to medium probability for the majority of fire occurrences discovered on this day. The highest probability areas predicted by the model for this day are not near the majority of fire occurrences discovered on this day. The lack of agreement between the high probability areas and the elements of interest combined with the spatially

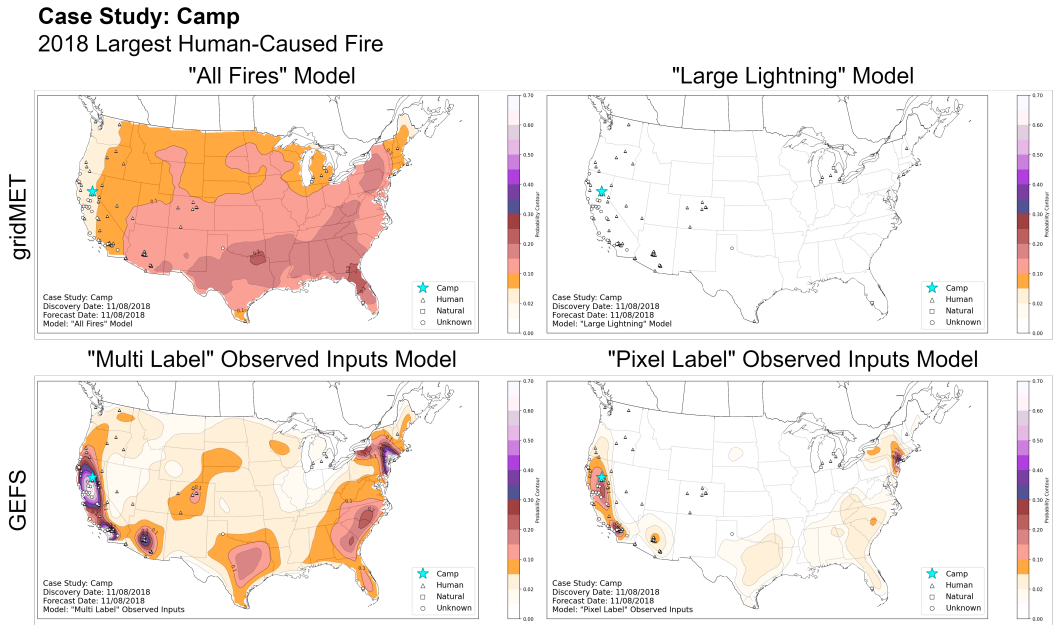


Figure 8.9: Camp All Fires Discovered Case Study

diffuse nature of the predictions and the low to medium probabilities predicted for the elements of interest suggest that this model is not useful for this case study.

The "Large Lightning" model, shown in the upper right hand corner of Figure 8.9, is excluded from this case study but is presented here to in an attempt to fully illustrate the capabilities of the model.

The "Multi Label" model, shown in the lower left hand corner of Figure 8.9, depicts densely concentrated areas of high probability collocated with the majority of fire occurrences discovered on this day. The model predicated an area of similar probability in the northeastern region as well as areas of medium probability in the southern and southeastern regions of CONUS which are not near the fire occurrences discovered on this day.

The "Pixel Label" model, shown in the lower right hand corner of Figure 8.9, depicts densely concentrated areas of medium to high probability collocated with the majority of fire occurrences discovered on this day with two exclusions. First,

All Fires Discovered

November 8, 2018

Model	*Rank Sum*	*Probability Rank*	*Localization Rank*
Pixel Label	**3**	**2**	**1**
Multi Label	**3**	**1**	**2**
All Fires	6	3	3
Large Lightning	-	-	-

Table 8.9: All Fires Case Study: Camp Probability and Localization Rank

it predicted no probability of fire for a collection of fires in Colorado and, second, it predicted a small area of high probability in the northeastern region where there were no fire occurrences.

My subjective rankings for the proximity of predictions to each other (localization rank) and the proximity of predictions to the element of interest (probability rank) based on visual analysis are described in Table 8.9.

8.4 Localization: A Deeper Look

Figure 8.10 summarizes the rank sums of the six case studies, two wildfires with three examples per wildfire, described in this section. The "Pixel Label" model was the top model for four of the studies, the "Large Lightning" model was the top model for two of the studies, the "Multi Label" model was the top model for three of the studies (alongside the "Pixel Label" model), and the "All Fires" model was the top model for none of the studies.

The "All Fires" model is held back by the spatially diffuse property of its predictions which make it difficult to know which areas of CONUS are important. Following

Sum Total of Rank

Case Study	*Fire*	All Fires	Large Lightning	Multi Label	Pixel Label
Individual Fires	*South Sugarloaf*	7	2	5	6
	Camp	6	-	3	3
Large Fires	*South Sugarloaf*	8	2	5	5
	Camp	6	-	3	3
All Fires	*South Sugarloaf*	7	5	5	3
	Camp	6	-	3	3
	TOTAL	40	9	24	23

Table 8.10: Case Study Summary of Rank: Lower Numbers Indicating More Useful Models and Higher Numbers Indicating Less Useful Models.

this thread, the "Pixel Label" model had the least spatially diffuse predictions (setting aside the "Large Lightning" model for now) and the "Multi Label" model fell somewhere in between. The only difference between the "Pixel Label" model and the "Multi Label" model is that the labels used for the "Pixel Label" model are not expanded in time and space. The only difference between the "Multi Label" model and the "All Fires" model is that the "All Fires" model uses gridMET data and the "Multi Label" model uses GEFS data. Here are two candidates for what is contributing to the difference in localization: expanded labels and the dataset used, with GEFS offering more densely concentrated predictions than gridMET (again, setting the "Large Lightning" model aside for now). Why might this be?

Thinking first about the two datasets, GEFS contains weather variables and gridMET contains both weather and fire danger variables. Fire danger variables attempt to capture both fuel and weather effects in the same variable. Fuel changes at a much slower rate than weather. By combining weather and fuel together into the same variable, the localization made possible by the high granularity, rapidly changing weather data is diluted by the lower granularity, more slowly changing fuel data. This is not to say that fuel data cannot be included in the model. If the fuel data is included in the model as variables that are separate from the weather variables, the model is able

to establish unique relationships with each variable. This, theoretically, would reduce or entirely remove the impact of the dilution caused by the fire danger variables. This question must be left to future work to answer.

Circling back to the "Large Lightning" model which used gridMET data but still achieved a high level of localization, I look to the "All Fires" model for answers. The only difference between the "All Fires" model, which had low localization, and the "Large Lightning" model, which had high localization, was that the "Large Lightning" model used a small subset of fire occurrences in its label images. This resulted in a high amount of sparsity of the event of interest within the label images. I suggest that it is this sparsity that allowed the model to create more densely concentrated predictions. The consequences of supporting localization by limiting the number of fire occurrences in the label images to a subset representing a narrower definition of the phenomenon is two-fold. First, the probabilities predicted by the "Large Lightning" model are the lowest of all of the models attempted in the course of this work. Second, as can be seen in Table 8.10, the "Large Lightning" model is only able to perform in use cases that focused on large, natural-caused fires. However, this is not to lose sight of the upside of such a constraint. When the use case is a single or a set of large, natural-caused fires, the "Large Lightning" model performs the best relative to the other models depicted.

Shifting gears, I move on to discuss expanded labels which are another potential contributing factor to producing more diffuse probability areas. By expanding labels in time and space for the training data, it is possible that the the predictions produced by the model expanded along the time and space dimensions as well. Expansion along the space dimension may have slightly reduced the spatial concentration of the predictions produced by the model. Expansion along the space and time dimensions may have increased the probabilities the model is able to produce as there are an

increased number of feature spaces associated with a fire outcome. The "Multi Label" model produces the highest probabilities of any model created in this work and does so without complete loss of localization (as is the case for the "All Fires" model) and without reduced efficacy on different types of fires (as is the case with the "Large Lightning" model).

In summary, both the "Multi Label" model and the "Pixel Label" model are able to produce densely concentrated, accurate, high probability predictions making them useful models for the six case studies depicted in this section.

Chapter 9

Conclusions

Wildfires in the U.S. are becoming larger, more frequent, and increasingly costly. Reports, such as the Climate Science Special Report and the Federal Government's Climate Financial Risk report, highlight the rising incidence of large forest fires due to climate change, with projections indicating a near doubling of burned areas in national forests by mid-century and significant increases in fire suppression costs. In response to these growing challenges, this work introduces a new approach to predict wildfire occurrence across the contiguous United States. The approach focuses on environmental factors rather than ignition-specific variables like population density or lightning strikes. The goal is to predict areas at risk of wildfire and capture the most operationally significant events, such as large wildfires.

To achieve this end, I have created a selection of proof-of-concept models using a binary classification, semantic segmentation, deep learning architecture. These models can predict fire occurrence for CONUS out to 10 days and can do so with higher probability of wildfire than the SPC Probability of Wildfire Climatology, the tool relied on operationally by fire weather forecasters.

Chapters 5 and 6 present a comparison between two fire occurrence prediction models using gridMET data: the "All Fires" model, which includes all fire occurrences, and the "Large Lightning" model, which focuses on large, naturally-caused fires. The "All Fires" model consistently produces higher values than the "Large

Lightning" model for both CSI and probability of wildfire for the experiments presented and higher than both the "Large Lightning" model and the SPC Probability of Wildfire for three case studies of the largest, naturally-caused wildfires for the years 2018, 2019, and 2020: the South Sugarloaf Fire, the Sheep Fire, and the Doe Fire, respectively. The findings highlight the value of incorporating more fire occurrence instances in labels rather than fewer, informing future work aimed at enhancing model interpretation while maintaining or improving model performance. Lessons learned from creating the gridMET models include the importance of retaining full variable distributions for input variables instead of using summary variables and the challenges posed by multicollinearity in input datasets. To address these issues, the work described in Chapters 7 and 8 uses non-summary input variables and a new input dataset with less multicollineary to improve model interpretability and operational effectiveness.

In Chapters 7 and 8, I present a set of models which use GEFS input data, building on previous work from Chapters 5 and 6. Two labeling approaches are compared: the "Multi Label" model, which uses spatial and temporal label expansion data augmentation, and the "Pixel Label" model, which uses binary labels. Both models achieve similar Max CSI values across different neighborhood sizes, with performance improving slightly when using forecast inputs rather than observed inputs. The "Multi Label" model with forecast inputs performed best when measured at 80 km, while the "Pixel Label" model performed best when measured at 40 km. When compared to the SPC Probability of Wildfire Climatology on three case study fires from the 2018 fire season, the Camp Fire, the Carr Fire, and the Woolsey Fire, the machine learning models provided higher wildfire probability predictions, even ten days prior to the discovery date of the wildfires. Additionally, I explored permutation variable importance, a task which became possible due to the reduced multicollinearity in the

GEFS dataset compared to the gridMET dataset. The study found that variables related to the state of the fuels (precipitable water, temperature, and specific humidity) had a greater impact on the model's ability to predict fire occurrence than wind magnitude variables which are more commonly associated with fire spread. Chapter 8 concluded with a visual comparison of the gridMET and GEFS (observation input) models for two major fires: the South Sugarloaf Fire, the largest natural-caused fire of the 2018 fire season, and the Camp Fire, the largest human-caused fire of the 2018 fire season. The GEFS models consistently offered higher probability of wildfire than the gridMET models, with the "Pixel Label" model and the "Multi Label" model vying for primacy. This analysis underscores the superiority of GEFS data for fire prediction relative to the gridMET dataset, offering enhanced accuracy, precision, and performance, especially for large and impactful wildfires.

Future work could explore several key areas for model improvement. Extending predictions beyond day 10 would provide insight into how far into the future the model can maintain its performance. Additionally, comparing the model's predictions to a baseline of random performance could offer a clearer perspective on its accuracy. Addressing wrap-around effects from the Day of Year variable is another important consideration; solutions could include cyclic transformations or the introduction of an annual indicator. Finally, incorporating fuel input variables, not simply as fire danger indices, may help address current gaps in the model's performance and should be pursued in future research.

In conclusion, the key contribution of my PhD research was the development of a deep learning approach that outperformed current operational techniques used by forecasters for predicting fire occurrence. My models were evaluated using Critical Success Index (CSI), wildfire probability, and visual analysis, and were bench-marked against the SPC Probability of Wildfire Climatology. In case studies of the largest

human-caused and natural-caused fires, the machine learning models consistently delivered higher wildfire probability values, with the GEFS-based models excelling in spatial accuracy. Importantly, this success was achieved using datasets already employed by operational forecasters, and I established a robust foundation for comparison with traditional methods by utilizing metrics widely recognized and trusted within the fire weather forecasting community.

The broader implications of this work lie in addressing a challenging prediction problem that has historically lacked effective methods. Traditional techniques have not been well-suited to accurately predict fire occurrence, leaving much of the task to human intuition based on years of experience. With the advent of deep learning, we now have the ability to incorporate the complexity and subtle relationships that a human might learn over time and codify them into a model. This advancement has the potential to transform fire prediction into a more systematic and reliable process, ultimately leading to outputs that can greatly benefit operational decision-making.

Reference List

Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, **33**, 121–131.
URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.3413

Aggarwal, C. C., 2018: *Neural Networks and Deep Learning*. Springer Cham, 1 edition, XXIII, 497 pp.
URL https://doi.org/10.1007/978-3-319-94463-0

Ahrens, C., 2009: *Meteorology Today: An Introduction to Weather, Climate, and the Environment*. International student edition, Brooks/Cole, CengageLearning.
URL https://books.google.com/books?id=SrkRAQAAIAAJ

Alonso-Betanzos, A., O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, M. I. P. Andrade, and E. Jiménez, 2003: An intelligent system for forest fire risk prediction and firefighting management in galicia. *Exp. Syst. Appl*, **25(4) : 545–554**.

Alonso-Betanzos, A., O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, J. Canda, E. Jimenez, et al., 2002: A neural network approach for forestal fire risk estimation.

Auburn University, 2022: The Ecology of Fire. Accessed 10 February 2022, http://www.auburn.edu/academic/forestry_wildlife/fire/.

Balch, J. K., B. A. Bradley, J. T. Abatzoglou, R. C. Nagy, E. J. Fusco, and A. L. Mahood, 2017: Human-started wildfires expand the fire niche across the United States. *Proceedings of the National Academy of Sciences*, **114**, 2946–2951.
URL https://www.pnas.org/doi/abs/10.1073/pnas.1617394114

Bentley, E., 2022: Personal conversation with Evan Bentley. Personal communication, october 17, 2022.

Bishop, C. M., 2006: *Pattern Recognition and Machine Learning*. Springer, New York.

Bradshaw, L. S., J. E. Deeming, R. E. Burgan, and J. D. Cohen, 1984: The 1978 national fire-danger rating system: technical documentation. *General Technical Report INT-169*.
URL http://dx.doi.org/10.2737/INT-GTR-169

Breiman, L., 2001: Random forests. *Machine learning*, **45**, 5–32.

Bröcker, J. and L. A. Smith, 2007: Increasing the reliability of reliability diagrams. *Weather and Forecasting*, **22**, 651 – 661.
URL https://journals.ametsoc.org/view/journals/wefo/22/3/waf993_1.xml

Burke, A., N. Snook, D. J. G. II, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Weather and Forecasting*, **35**, 149–168.

CAL FIRE, 2020: August Complex (includes Doe Fire). https://www.fire.ca.gov/incidents/2020/8/16/august-complex-includes-doe-fire/, Accessed: 2023-05-26.

Cho, D., C. Yoo, B. Son, J. Im, D. Yoon, and D.-H. Cha, 2022: A novel ensemble learning for post-processing of nwp model's next-day maximum air temperature forecast in summer using deep learning and statistical approaches. *Weather and Climate Extremes*, **35**, 100410.
URL https://www.sciencedirect.com/science/article/pii/S2212094722000044

de Almeida Pereira, G. H., A. M. Fusioka, B. T. Nassu, and R. Minetto, 2021: Active fire detection in landsat-8 imagery: A large-scale dataset and a deep-learning study. *ISPRS Journal of Photogrammetry and Remote Sensing*, **178**, 171–186.
URL https://www.sciencedirect.com/science/article/pii/S092427162100160X

de Bem, P. P., O. A. de Carvalho Júnior, O. L. F. de Carvalho, R. A. T. Gomes, and R. Fontes Guimarães, 2020: Performance analysis of deep convolutional autoencoders with different patch sizes for change detection from burnt areas. *Remote Sensing*, **12**.
URL https://www.mdpi.com/2072-4292/12/16/2576

Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, 2009: Imagenet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 248–255.

Draper, N. R. and H. Smith, 1998: Applied regression analysis. *Wiley Series in Probability and Statistics*.

Drozdzal, M., E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, 2016: The importance of skip connections in biomedical image segmentation.
URL https://arxiv.org/abs/1608.04117

Dutta, R., J. Aryal, A. Das, and J. B. Kirkpatrick, 2013: Deep cognitive imaging systems enable estimation of continental-scale fire incidence from climate data. *Sci. Rep*, **3(1): 3188**.

Earnest, B. L., A. McGovern, C. Karstens, and I. Jirak, 2024a: Part I: Improving Wildfire Occurrence Prediction for CONUS Using Deep Learning and Fire Weather Variables. *Artificial Intelligence for the Earth Systems*, **3**, e230057.
URL https://journals.ametsoc.org/view/journals/aies/3/3/AIES-D-23-0057.1.xml

—, 2024b: Part II: Lessons Learned from Predicting Wildfire Occurrence for CONUS Using Deep Learning and Fire Weather Variables. *Artificial Intelligence for the Earth Systems*, **3**, e230058.
URL https://journals.ametsoc.org/view/journals/aies/3/3/AIES-D-23-0058.1.xml

Ebert-Uphoff, I., R. Lagerquist, K. Hilburn, Y. Lee, K. Haynes, J. Stock, C. Kumler, and J. Q. Stewart, 2021: CIRA guide to custom loss functions for neural networks in environmental sciences - version 1. *CoRR*, **abs/2106.09757**.
URL https://arxiv.org/abs/2106.09757

Elliott, M., 2023: Personal conversation with Matt Elliott. Personal communication, october 26, 2023.

Fisher, A., C. Rudin, and F. Dominici, 2019: All models are wrong, but many are useful: Variable importance for black-box, proprietary, or misspecified prediction models, using model class reliance. *arXiv preprint arXiv:1801.01489*.
URL https://arxiv.org/abs/1801.01489

Frnda, J., M. Durica, J. Rozhon, et al., 2022: ECMWF short-term prediction accuracy improvement by deep learning. *Scientific Reports*, **12**, 7898.

Fusco, E., J. Abatzoglou, J. Balch, J. Finn, and B. Bradley, 2016: Quantifying the human influence on fire ignition across the western usa. *Ecological Applications*, **26**.

Goodfellow, I., Y. Bengio, and A. Courville, 2016: *Deep Learning*. MIT Press.
URL http://www.deeplearningbook.org

Guan, H., Y. Zhu, E. Sinsky, B. Fu, W. Li, X. Zhou, X. Xue, D. Hou, J. Peng, M. M. Nageswararao, V. Tallapragada, T. M. Hamill, J. S. Whitaker, G. Bates, P. Pegion, S. Frederick, M. Rosencrans, and A. Kumar, 2022: Gefsv12 reforecast dataset for supporting subseasonal and hydrometeorological applications. *Monthly Weather Review*, **150**, 647 – 665.
URL https://journals.ametsoc.org/view/journals/mwre/150/3/MWR-D-21-0245.1.xml

Haupt, S. E., D. J. Gagne, W. W. Hsieh, V. Krasnopolsky, A. McGovern, C. Marzban, W. Moninger, V. Lakshmanan, P. Tissot, and J. K. Williams, 2022: The history

and practice of ai in the environmental sciences. *Bulletin of the American Meteorological Society*, **103**, E1351 – E1370.
URL https://journals.ametsoc.org/view/journals/bams/103/5/BAMS-D-20-0234.1.xml

He, K., X. Zhang, S. Ren, and J. Sun, 2015: Deep residual learning for image recognition. *CoRR*, **abs/1512.03385**.
URL http://arxiv.org/abs/1512.03385

Hill, A. J., R. S. Schumacher, and I. L. Jirak, 2023: A new paradigm for medium-range severe weather forecasts: Probabilistic random forest–based predictions. *Weather and Forecasting*, **38**, 251 – 272.
URL https://journals.ametsoc.org/view/journals/wefo/38/2/WAF-D-22-0143.1.xml

Huang, H., L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, 2020: Unet 3+: A full-scale connected unet for medical image segmentation. *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, IEEE, https://arxiv.org/abs/2004.08790.

Hyndman, R. and Y. Fan, 1996: Sample quantiles in statistical packages. *The American Statistician*, **50**, 361–365.

Ibtehaz, N. and M. S. Rahman, 2020: Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, **121**, 74–87.
URL http://dx.doi.org/10.1016/j.neunet.2019.08.025

Idaho National Laboratory, 2019: One Year After Historic Sheep Fire, Idaho Site Contractors, Agencies Eye Coming Fire Season. https://inl.gov/article/after-sheep-fire/, accessed 26 May 2023.

Jain, P., S. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, 2020: A review of machine learning applications in wildfire science and management. *Environmental Reviews*, **28**, 488.

Justin, A. D., A. McGovern, and J. T. Allen, 2024: Efficient identification of frontal boundaries over NOAA's Unified Surface Analysis domain using AI. *Artificial Intelligence for the Earth Systems*, in review.

Justin, A. D., C. Willingham, A. McGovern, and J. T. Allen, 2023: Toward operational real-time identification of frontal boundaries using machine learning. *Artificial Intelligence for the Earth Systems*, **2**, e220052.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, ..., and D. Joseph, 1996: The ncep/ncar 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437–471.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021: Physics-informed machine learning. *Nature*, **3**, 422–440.

Knopp, L., M. Wieland, M. Rättich, and S. Martinis, 2020: A deep learning approach for burned area segmentation with sentinel-2 data. *Remote Sensing*, **12**.
URL https://www.mdpi.com/2072-4292/12/15/2422

Lin, P., M. Elliott, D. Harrison, E. Bentley, I. Jirak, J. Vancil, K. Halbert, and P. Marsh, 2024: Developing dry thunderstorm verification tools to improve fire weather forecasting at NOAA's storm prediction center, american Meteorology Society 104th Annual Meeting.
URL https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Paper/429328

Management and O. o. Budget, 2024: Climate financial risk: The federal government's budget exposure to financial risk due to climate change.
URL https://www.whitehouse.gov/wp-content/uploads/2024/03/climate_budget_exposure_fy2025.pdf

Nair, V. and G. E. Hinton, 2010: Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 807–814.

National Interagency Fire Center, 2023: Energy Release Component. Accessed 2 June 2023, https://gacc.nifc.gov/swcc/dc/azpdc/predictive-services/documents/fuels-fire-danger/ERC.pdf.

National Oceanic and Atmospheric Administration, 2024: NOAA global ensemble forecast system (gefs) re-forecast. Accessed: 6-7-2024.
URL https://registry.opendata.aws/noaa-gefs-reforecast

National Park Service, 2016: Wildland Fire Facts: There Must Be All Three. Accessed 10 February 2022, https://www.nps.gov/articles/wildlandfire-facts-fuel-heat-oxygen.htm.

National Wildfire Coordinating Group, 2022a: Fire. Accessed 10 February 2022, https://www.nwcg.gov/term/glossary/fire.

—, 2022b: Fire Ecology. Accessed 10 February 2022, https://www.nwcg.gov/term/glossary/fire-ecology.

—, 2022c: Wildfire. Accessed 10 February 2022, https://www.nwcg.gov/term/glossary/wildfire.

Nauslar, N. J., 2019: Wildfire and weather: Understanding and predicting fire behavior. Accessed: 2024-06-08.
URL https://www.spc.noaa.gov/publications/nauslar/sidebar.pdf

—, 2020a: Personal conversation with Nick Nauslar. Personal communication, april 1, 2020.

—, 2020b: Storm Prediction Center Probability of Wildfire. Accessed 25 May 2023, https://www.spc.noaa.gov/new/FWclimo/climo.php?parm=1000ac.

Nielsen, M. A., 2015: *Neural Networks and Deep Learning*. Neural Networks and Deep Learning.
URL http://neuralnetworksanddeeplearning.com/

NWCG, 2024: https://www.nwcg.gov/publications/pms437/fire-danger/nfdrs-system-inputs-outputs#TOC-Observations-and-Forecasts.

Oregon State University, 2019: Fire FAQs—What is forest fuel, and what are fuel treatments? Accessed 10 February 2022, https://catalog.extension.oregonstate.edu/em9230/html.

Pu, Z. and E. Kalnay, 2018: Nwp basics. https://www.inscc.utah.edu/~pu/6500_sp12/Pu-Kalnay2018_NWP_basics.pdf, accessed: 2024-06-22.

Pyne, S. J., 2001: *Fire: A Brief History*. University of Washington Press.
URL http://www.jstor.org/stable/j.ctvcwnf8f

Rashkovetsky, D., F. Mauracher, M. Langer, and M. Schmitt, 2021: Wildfire detection from multisensor satellite imagery using deep semantic segmentation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **14**, 7001–7016.

Roberts, N. M. and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136**, 78–97.
URL https://journals.ametsoc.org/view/journals/mwre/136/1/2007mwr2123.1.xml

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Weather and Forecasting*, **24**, 601–608.

Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. *CoRR*, **abs/1505.04597**.
URL http://arxiv.org/abs/1505.04597

Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323**, 533–536.
URL https://doi.org/10.1038/323533a0

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Weather and Forecasting*, **5**, 570–575.
URL https://journals.ametsoc.org/view/journals/wefo/5/4/1520-0434_1990_005_0570_tcsiaa_2_0_co_2.xml

Short, K. C., 2023: Spatial wildfire occurrence data for the united states, 1992-2020 [fpa_fod_20221014] 6th edition.
URL https://doi.org/10.2737/RDS-2013-0009.6

SPC, 2024a: Critical criteria for severe thunderstorm and tornado watches. Available online from the Storm Prediction Center (SPC), National Weather Service, NOAA.
URL https://www.spc.noaa.gov/misc/Critical_Criteria_for_web.pdf

—, 2024b: Fire weather outlooks. Available online from the Storm Prediction Center (SPC), National Weather Service, NOAA.
URL https://www.spc.noaa.gov/products/fire_wx/fwdy1.html

State of Nevada, 2018: South Sugarloaf Fire. Accessed 26 May 2023, https://sagebrusheco.nv.gov/uploadedFiles/sagebrusheconvgov/content/Meetings/2018/18_0829\%20South\%20Sugarloaf\%20Fire\%20PowerPoint.pdf.

Tversky, A. and D. Kahneman, 1974: Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.
URL https://www.ncbi.nlm.nih.gov/pubmed/17835457

United States Department of Agriculture, 2023: Energy Release Component (ERC) Fact Sheet. Accessed 2 June 2023, https://www.fs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb5339121.pdf.

United States Forest Service, 2023: USFS Fire Regions. https://www.fs.usda.gov/foresthealth/contact-us/regional-contacts.shtml, accessed 12 May 2023.

University of California Agriculture and Natural Resources, Accessed: 2024-06-22: Fire ecology. University of California Agriculture and Natural Resources, https://cemariposa.ucanr.edu/Fire_Information/Fire_Ecology/.

USFS, 2024: National fire danger rating system. Accessed 3 April 2024, https://www.fs.usda.gov/detail/cibola/landmanagement/resourcemanagement/?cid=stelprdb5368839.

Vasilakos, C., K. Kalabokidis, J. Hatzopoulos, G. Kallos, and Y. Matsinos, 2007: Integrating new methods and tools in fire danger rating. *Int. J. Wildl. Fire*, **16(3): 306–316**.

Vecín-Arias, D., F. Castedo-Dorado, C. Ordóñez, and J. R. Rodríguez-Pérez, 2016: *Biophysical and lightning characteristics drive lightning-induced fire occurrence in the central plateau of the Iberian Peninsula*, volume 225:36–47. Agric. For. Meteorol.

Vega-Garcia, C., B. S. Lee, P. M. Woodard, and S. J. Titus, 1996: *Applying neural network technology to human-caused wildfire occurrence prediction*, volume 10:9–18. AI Appl.
URL https://cfs.nrcan.gc.ca/publications?id=18949

Wang, Z., P. Yang, H. Liang, C. Zheng, J. Yin, Y. Tian, and W. Cui, 2022: Semantic segmentation and analysis on sensitive parameters of forest fire smoke using smoke-unet and landsat-8 imagery. *Remote Sensing*, **14**.
URL https://www.mdpi.com/2072-4292/14/1/45

Wehner, M., J. Arnold, T. Knutson, K. Kunkel, and A. LeGrande, 2024: Climate science special report: Fourth national climate assessment, volume i u.s. global change research program, washington, dc, usa.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (et) technique in the ncep global operational forecast system. *Tellus*, **60A**, 62–79.

WFAS, 2024: Wildland fire assessment system. https://www.wfas.net/index.php?option=com_content&view=article&id=75&Itemid=478.

Xu, X. and E. Frank, 2004: Logistic regression and adaptive boosting for labeled bags of instances. *Pacific-Asia conference on knowledge discovery and data mining*, Springer, 272–281.

Özgün Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, 2016: 3d u-net: Learning dense volumetric segmentation from sparse annotation.
URL https://arxiv.org/abs/1606.06650