Part I: Improving Wildfire Occurrence Prediction for CONUS Using Deep Learning and Fire Weather Variables

BETHANY L. EARNEST,[a,b,e] AMY MCGOVERN,[a,b,c,d] CHRISTOPHER KARSTENS,[e,d] AND ISRAEL JIRAK[e]

[a] *School of Computer Science, University of Oklahoma, Norman, Oklahoma*
[b] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, Norman, Oklahoma*
[c] *NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography (AI2ES), Norman, Oklahoma*
[d] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[e] *Storm Prediction Center, Norman, Oklahoma*

(Manuscript received 12 July 2023, in final form 3 April 2024, accepted 8 April 2024)

ABSTRACT: The purpose of this research is to build an operational model for predicting wildfire occurrence for the contiguous United States (CONUS) in the 1–10-day range using the U-Net 3+ machine learning model. This paper illustrates the range of model performance resulting from choices made in the modeling process, such as how labels are defined for the model and how input variables are codified for the model. By combining the capabilities of the U-Net 3+ model with a neighborhood loss function, fractions skill score (FSS), we can quantify model success by predictions made both in and around the location of the original fire occurrence label. The model is trained on weather, weather-derived fuel, and topography observational inputs and labels representing fire occurrence. Observational weather, weather-derived fuel, and topography data are sourced from the gridded surface meteorological (gridMET) dataset, a daily, CONUS-wide, high-spatial-resolution dataset of surface meteorological variables. 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, combining data sourced from the reporting systems of federal, state, and local organizations. By exploring the many aspects of the modeling process with the added context of model performance, this work builds understanding around the use of deep learning to predict fire occurrence in CONUS.

SIGNIFICANCE STATEMENT: Our work seeks to explore the limits to which deep learning can predict wildfire occurrence in CONUS with the ultimate goal of providing decision support to those allocating fire resources during high fire seasons. By exploring with what accuracy and lead time we can provide insights to these persons, we hope to reduce loss of life, reduce damage to property, and improve future event preparedness. We compare two models, one trained on all fires in the continental United States and the other on only large lightning fires. We found that a model trained on all fires produced a higher probability of fire.

KEYWORDS: Neural networks; Forest fires; Wildfires; Deep learning; Machine learning; Artificial intelligence

1. Introduction

This research explores the impact of modeling choices on model performance when applied to the task of predicting wildfire occurrence for the contiguous United States (CONUS) in the 1–10-day range through the lens of the U-Net 3+ machine learning model. The ultimate goal is to map out the extent to which fire occurrence can be predicted in advance of discovery with the aim of making these predictions available to the wildfire community to offer decision support. Fire occurrence prediction (FOP) is described in 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 historically use regression methods to model the relationship between historical fire data, such as fire reports, and environmental factors, such as weather, topography, and fuel.

Previous research has applied several machine learning models to the task of predicting fire occurrence, most commonly artificial neural networks, random forests, and support vector machines (Dutta et al. 2013; Vecín-Arias et al. 2016; Sakr et al. 2011). Many incarnations of neural networks have been used in other research on the prediction of fire occurrence using machine learning (Vega-Garcia et al. 1996; Alonso-Betanzos et al. 2002, 2003; Vasilakos et al. 2007; Dutta et al. 2013, 2016). This work applies a U-Net 3+ machine learning model (Huang et al. 2020). The U-Net 3+ model is an extension of the artificial neural network architecture which combines an encoder–decoder architecture with full-scale skip connections.

Though neural networks have been a successful machine learning model for modeling fire occurrence (Jain et al. 2020), they can be complicated to implement, computationally intensive, and difficult to interpret. These issues tend to create barriers to their use and subsequent adoption in operational settings where they are not the historically preferred model.

This paper demonstrates performance improvement relative to climatology methods currently used operationally by employing minimally preprocessed, publicly available data in combination with an effective machine learning model. Barriers to adoption are addressed by working with datasets and

Corresponding author: Bethany L. Earnest, bethany.earnest@gmail.com

DOI: 10.1175/AIES-D-23-0057.1 e230057

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 companion paper, Earnest et al. (2024), we analyze lessons learned from applying deep learning to fire occurrence prediction.

2. Related work

A key element of interdisciplinary research is working closely with subject matter experts to understand the research space. To inform our understanding of approaches and considerations employed today, we shadowed Storm Prediction Center (SPC) forecasters as they worked the fire weather forecast desk, and to ensure we proceeded always with an eye on operationalization, we collaborated directly with members of the SPC Science Support Branch. In this section, considerations shared by our subject matter experts are provided and we describe how we translated those considerations into modeling choices.

a. Regional and seasonal effects

One prevalent theme communicated to us by our subject matter experts related the variation in fire occurrence and fire behavior to the season and the region in which fire exists. For example, in regard to fire season, there are often more fires in the summer and fuels are more fire prone during warm, dry periods, such as drought. Regarding region, fire sizes and growth rates differ based on the type and state of fuels present as well as the topography present in the region. 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. All of these elements are instrumental in effective fire occurrence prediction, and we have many options as to how we account for these elements during the modeling task.

With the exclusion of Dutta et al. (2013), who compared ten different implementations of artificial neural networks and modeled all of Australia, an area of over 2.9 million square miles, many previous researchers in this space addressed regional differences by reducing the overland area addressed by their model. In 2002 and 2003, Alonso-Betanzos et al. (2002, 2003) used an artificial neural network to predict a daily fire occurrence risk index in the Galicia region of Spain, an area of 11 419 square miles. In 2007, Vasilakos et al. (2007) used three different artificial neural networks (to predict fire weather, fire hazard, and fire risk, respectively) to model the probability of fire occurrence on the Greek island of Lesvos, an area of 630 square miles. Previous attempts to model fire occurrence, independent of the machine learning method, focused on overland areas between 630 square miles and 255 541 square miles (Jain et al. 2020).

In this work, the CONUS is used, an area of over 3 million square miles containing many dissimilar regions. This is done for two reasons. First, it allows us to retain a larger amount of data for the model to learn from, a choice critical to supporting the use of deep learning to create a performative model for this task. Second, to align with the forecast domain used by the NOAA/NWS SPC, the entity tasked with national fire weather prediction within the National Weather Service.

The inclusion of many dissimilar regions does not go unaddressed. Instead of following the methods used by previous researchers and reducing the overland area of our model, we include variables which codify the differences between dissimilar regions so the model can navigate these dissimilarities effectively without intervention. Some variables that help us implicitly communicate differences between regions to the model are topography, weather, and weather-derived fuel data. We also include latitude and longitude variables, which represent the centroids of each of the grid cells (instead of a categorical variable containing only region names, for example). Location variables such as latitude and longitude both allow the model to draw out relationships between location and the other variables included in the model and stand in for confounding variables associated with location either unknown or absent from the model.

There are few examples in the literature where researchers limited the seasons included in their models, which aligns with the approach we adopted in the present work. Some variables that help implicitly communicate differences between seasons to the model are weather and weather-derived fuel data. We included a Julian day (relative to 1 January) as a variable (as opposed to a categorical variable containing only the four seasons of the year, for example). This "day of year" variable helps us communicate not only season to the model but weekends and holidays (4 July being an example of a holiday with important human-caused wildfire implications) despite not explicitly specifying which days are holidays and which days are weekends. As with latitude and longitude, time variables, such as the day of year variable, allow the model to draw out relationships between time and the other variables present in the model and stand in for any confounding variables associated with time that are either unknown or absent from the model.

Model performance is evaluated in the context of both season and region so we can identify areas where more information may be needed for the model to navigate effectively. For this purpose, "regions" are defined to be the eight fire regions defined by the U.S. Forest Service and depicted in Fig. 6 (U.S. Forest Service 2023). A "season" is defined as the division of the year into quarters by month, with March, April, and May representing spring and June, July, and August representing summer.

b. Fire size and cause

Fire size, as described by the observational data, is the final fire size. This variable is affected by circumstances in the field, for example, the presence of fire breaks such as rivers and roads, and by fire management decisions made before and

after the fire occurrence has been discovered. For example, fire managers may more aggressively contain a fire near a high population area, resulting in a smaller final fire size. While the definition of "large" varies, to align more closely with existing climatology tools, we define large as any fire with a final fire size of greater than or equal to one thousand acres.

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) dataset to quantify the role of humans in fire ignition in the western United States.

The fire occurrence dataset we used contains three fire causes: human, natural, and unknown. Large, lightning-caused fire (codified in our label dataset as "natural" fires) is an operational concern of many fire management entities. Between the years 2000 and 2020 according to FPA-FOD, there are over 1.6 million fire occurrences recorded for CONUS, resulting in a little less than 120 million acres burned. Large lightning fires account for 47.24% of the total acres burned while comprising only a little under 5000, or 0.29%, of the total fire occurrence, a low-frequency, high-impact event.

As with regional and seasonal effects, fire size and fire cause have important impacts on modeling decisions. While we addressed regional and season effects using model inputs, as described in the previous section, for fire size and cause effects, we used model labels. This allows us to measure the size of the effect fire size and cause have on model performance. We produced a model that predicted the probability of discovery of fire (any size or cause) and measured how it performed relative to a model that predicted only the probability of discovery of large, lightning fire. By increasing the constraints under which a label can be considered to contain the event of interest, we were able to explore the relationship between the density of labels available to a model and model performance without reducing the coverage of the dataset in time or space. We observed that model performance decreased as the constraints on the label increased. Model performance was then compared based on general performance, on performance on only large lightning days and grid cells, on case studies representing the largest lightning fires for each of the years in the test dataset, and on case studies representing the most fire-prone days of the year for both human-caused fire and lightning-caused fire.

3. Data

To train our model, we used two datasets, recommended by our subject matter experts and used operationally by fire weather forecasters. Inputs are sourced from gridded surface meteorological (gridMET), a daily, CONUS-wide, high-spatial-resolution dataset of observed surface meteorological variables, that includes weather-derived fuel variables (Abatzoglou 2013). Labels are sourced from FPA-FOD, which contains spatial wildfire occurrence data for CONUS, combining data sourced from the reporting systems of federal, state, and local organizations (Short 2022). We consider the observations sourced from FPA-FOD to be our ground truth.

We sourced data for all of CONUS for the years 2000–2020, inclusive, using the discovery date of the fire. We 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, at this time, this is the best available information that tells us about the earliest stages of a fire. In the case of the "All Fires model, we sourced all instances of fire occurrence regardless of size, cause, or region for inclusion in the CONUS label images. In the case of the Large Lightning model, we sourced only instances of fire occurrence with a cause of natural and a final fire size greater than 1000 acres for inclusion in the CONUS label images. Controlled fires are not included in our dataset.

a. *gridMET*

gridMET combines climate data from the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) with temporal attributes from regional reanalysis (NLDAS-2). gridMET is validated against a network of weather stations [Remote Automatic Weather Stations (RAWS), Agricultural Meteorology (AgriMet), Agricultural Weather Network (AgWeatherNet), and USHCN version 2 (USHCN-2)]. The resulting dataset is both spatially and temporally complete, providing daily coverage over the entire CONUS for the years in the dataset used for this work (Abatzoglou 2013).

Input images were sourced from gridMET representing weather variables, such as precipitation, weather-derived fuel variables, such as energy release component, and a topography variable, elevation. Input images for the model are depicted in Fig. 1, and a full list of input variables can be found in Table A1 in appendix A. Readings from the CONUS for the years 2000–20, inclusive, were used to align with future input data and current operational needs. As previously described, latitude and longitude variables were included to represent the centroid of each grid cell and a day of year variable was included to represent the Julian day.

1) PREPROCESSING

For input variable sourced from gridMET, minimal preprocessing was employed. Further discussion of our preprocessing choices can be found in Part II, Earnest et al. (2024). Using linear interpolation, data were mapped from their native 4-km grid data to a 40-km grid to align with operational weather products. We conducted a lengthy investigation of using a 3-km grid and found that it was not feasible given time and computation constraints. Data were normalized to a value between 0 and 1 (a valuable preprocessing step for neural networks). Otherwise, data were left in their original form.

FIG. 1. Normalized gridMET inputs as consumed by the model.

A linear normalization calculation is used to normalize values to values between 0 and 1:

$$Z_i = [X_i - \min(X)]/[\max(X) - \min(X)]. \quad (1)$$

2) INPUT DATA LATENCY

As the goal of our work is to produce an operationalizable model, for which this work presents a proof of concept, input data latency is a consideration. gridMET represents observational data from midnight to midnight, western time. gridMET is updated daily, between approximately 0500 and 1430 Pacific time (depending on the variable) with yesterday's data, resulting in about a day and a half of input data latency.

Depending on what time today forecasters looked at gridMET, the most recently available data would be either yesterday's data or the day before yesterday's data. While it is generally acknowledged within CIWRO and SPC that, at a minimum, day of observational data would be preferable, gridMET is the preferred source for common variables used operationally by fire weather forecasters, such as energy release component (ERC), and, therefore, the source used for this work as well.

We are insulated from the impact of this latency because our model sources input data from the previous day. The fire occurrence predictions generated by our 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.

b. FPA-FOD

FPA-FOD is a spatial database of U.S. wildfires created and updated by Short (2022). Data for FPA-FOD are sourced from the National Interagency Fire Management Integrated Database, the U.S. Department of the Interior (USDI) Wildland Fire Management Information System, the Fish and Wildlife Service (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 SIT-209 Program, and 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 were 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 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, 1992–2011 and 2002–2011. 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, was given to Iowa, Illinois, Kansas, New York, and Texas.

As with gridMET, the years 2000–2020, inclusive, and only fire occurrence instances from the CONUS were sourced from FPA-FOD. Which instances were sourced from FPA-FOD for inclusion in the label images was determined by the model. For the All Fires model, all fire occurrence instances were sourced, regardless of fire size, fire cause, or region of discovery. For the Large Lightning model, only fire occurrence instances with a cause of natural and a final fire size greater than 1000 acres were sourced. The discovery latitude and longitude of each fire occurrence instance were mapped to a 40-km grid,

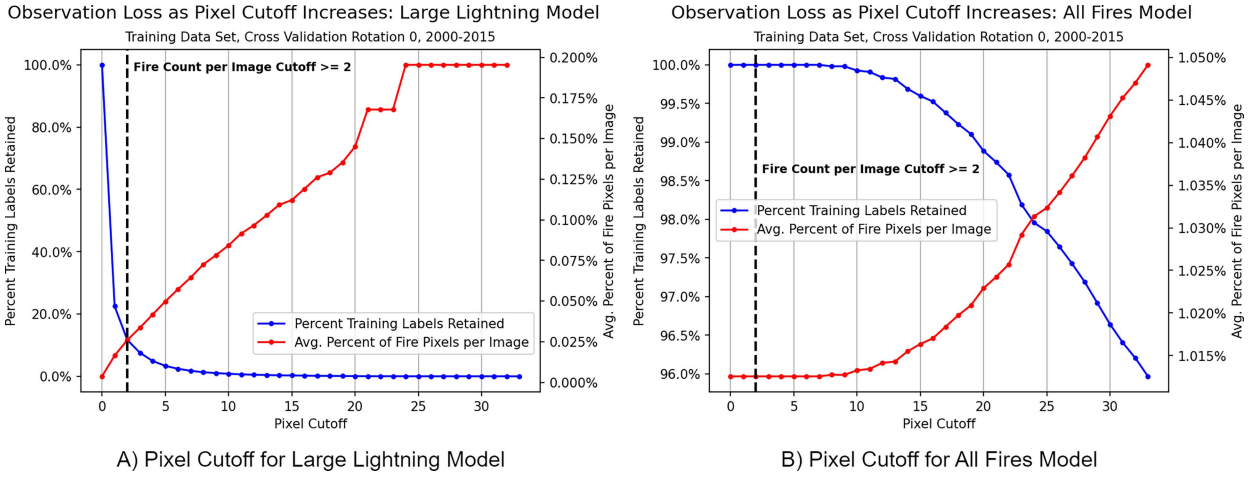


FIG. 2. Observation loss vs percent fire pixel increase by the cutoff.

resulting in 7671 CONUS-wide label images for each model (7661 label images after removing the last 10 days of the dataset, which could not be modeled for all future days).

Of the 7661 label images, for the Large Lightning model, only 22.74% of label images had one or more fire occurrences represented in all of CONUS (with 77.26% of label images containing no fire occurrence anywhere in CONUS for that day). Of the 16 384 (128 × 128) pixels present in each image, between 0.00% and 0.20% of pixels were labeled as containing a fire occurrence. For the All Fires model, fully 100% of label images had at least one fire occurrence represented somewhere in CONUS and between 0.04% and 3.42% of pixels were labeled as containing a fire occurrence.

1) ADDRESSING IMBALANCE

There were two types of imbalances present in the label dataset. The first type of imbalance is a class imbalance in favor of "no fire" (or a scarcity of the event *among* the label images). For example, for the Large Lightning model, nearly 80% of label images had no fire occurrence anywhere in CONUS. 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, at best, less than 4% of pixels in a label image were labeled with a value of 1 to indicate "fire."

To address the class imbalance, undersampling of the class of interest, that of fire, was performed by dropping any image from the training dataset containing less than or equal to two fire occurrence instances within the image. Undersampling was performed for the training dataset only. Two instances of fire were selected 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 Fig. 2a, was 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 was determined by the cutoff 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 2 allowed 12% of the training images to be retained and, for the All Fires model, allowed 100% of the training images to be retained.

To address the pixel imbalance within the label images, for the label images which remained after class imbalance had been addressed, two additional methods were used when depicting fire occurrence instances within the CONUS label image. The two methods, described below as the "neighborhood" method and the "time smoothed neighborhood" method, allowed fire occurrence instances to be depicted more prominently in time and space. These additional methods were applied to the training dataset only, which allowed for data augmentation and helped to ensure the model was able to learn meaningfully about the event of interest.

2) LABELING METHODS

For the training dataset, three methods for depicting fire occurrence within the CONUS label images were used: the "pixel" method, the neighborhood method, and the time smoothed neighborhood method. All three methods were used together for training. Only the pixel method was used for validation and test. These methods are depicted in Fig. 3.

The pixel method provided a binary representation of fire occurrence within the CONUS label image. Each CONUS label image represented 1 day for all of CONUS and contained 16 384 pixels. Each pixel represented a single grid cell from a 40-km grid. On a given day, within a given grid cell, if one or more fire occurrences were discovered, the grid cell was given a value of 1. If no fire occurrences were discovered, the grid cell was given a value of 0.

The neighborhood method builds upon the pixel method by increasing the size and numerical prominence of the pixel label using a linear decay approach. A central value of 1 was maintained, surrounded by values of 0.66, and then by values of 0.33, as depicted in Fig. 3. This method was chosen to make the binary labels more visually prominent so they would be easier for the model to see.

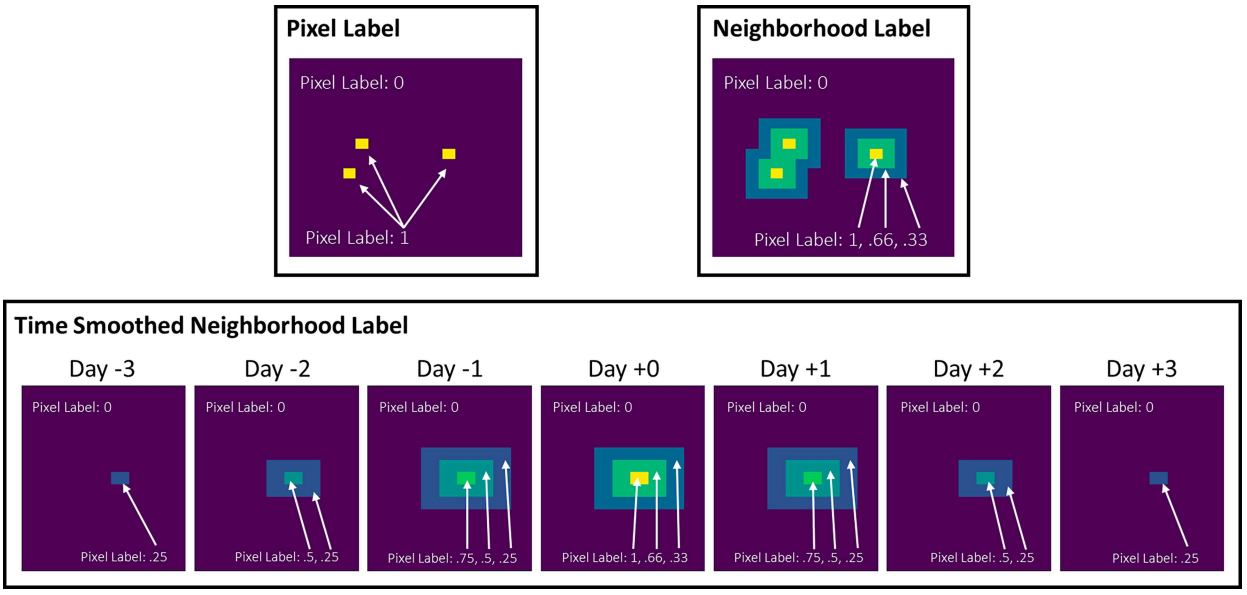


FIG. 3. Values for FPA-FOD labels: pixel, neighborhood, and time smoothed neighborhood.

The time smoothed neighborhood method builds upon both the pixel and neighborhood methods by increasing the size of the label in both time and space. It took into account the 3 days prior and 3 days after the fire occurrence discovery date. The neighborhood method is used for the discovery date and then reduced in both size and numerical prominence each day as it moves away in either direction from the date of discovery in time, as depicted in Fig. 3. This method was chosen because it mimics the method used to produce the SPC fire climatology to which we compare model performance.

4. Methods

In this work, the modeling of fire occurrence is approached as an image segmentation task. By doing so, we were able to consider inputs and provide predictions for the entire CONUS area while adjusting grid size to align with operationalized weather products. In support of this approach, a model architecture suited to image segmentation, the U-Net 3+ model (Huang et al. 2020), was selected.

Both models were trained on the same domain of input data—same days, same grid cells, and same input variables. The difference between the two models is the instances of fire depicted in the label images. The All Fires model was given CONUS images depicting all fires from the time period, regardless of size, cause, or region. The Large Lightning model was given CONUS images depicting only those instances of fire caused by lightning with a final fire size greater than or equal to 1000 acres. Described in Table A1 in appendix A are details on how data were divided into training, validation, and test.

a. *U-Net 3+*

A U-Net 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, 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 Fig. 4.

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 U-Net architecture incorporates same-scale skip connections, a connection between two layers, from the encoder to the decoder, at the same level within the architecture. Different-scale skip connections allow feature maps from different layers to be propagated both from the encoder to the decoder and from lower layers of the decoder to higher layers of the decoder and holds the added benefit of allowing the model to learn from multiple types of feature maps simultaneously.

The U-Net 3+ architecture is unique from previous U-Net architectures (U-Net and U-Net++) because it uses full-scale skip connections, a combination of same-scale skip connections and different-scale skip connections. This allows both the reduction in 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 U-Net 3+ to perform skillfully on image segmentation tasks (Huang et al. 2020), which is recommended for use in this work, as we used an image segmentation approach. Configuration details of our U-Net 3+ implementation can be found in Tables A2–A4 in appendix A.

b. *FSS*

During the task of image 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

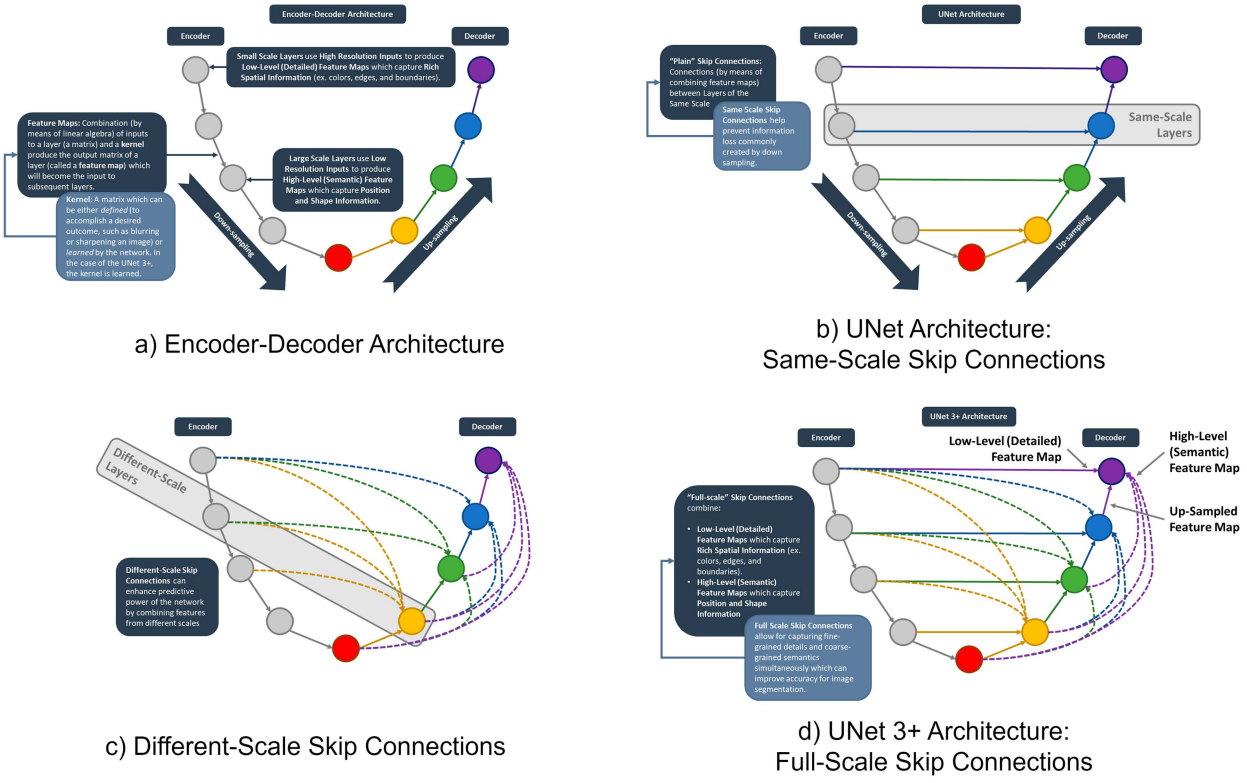


FIG. 4. Breakdown of the U-Net 3+ architecture: (a) encoder–decoder architecture, and kernels; (b) same-scale skip connections; (c) different-scale skip connections; (d) full-scale skip connections.

dependent. If the prediction is off by 1 pixel, it can be considered to be wrong depending on the loss function used to optimize the model. After discussing with the subject matter experts, 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. In support of this, we selected a loss function that 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. The FSS (Roberts and Lean 2008) loss function was selected. It 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:

$$\mathrm{FSS}_{(n)} = 1 - \mathrm{MSE}_{(n)}/\mathrm{MSE}_{(n)\mathrm{ref}}. \quad (2)$$

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 exceeds a given threshold. $\mathrm{MSE}_{(n)}$ refers to the mean square error (MSE) for the observed and forecast fractions for a neighborhood of size n, and $\mathrm{MSE}_{(n)\mathrm{ref}}$ refers to the highest MSE obtainable from the observed and forecast fractions.

c. *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, a grouped k-fold methodology 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 the 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 dataset with correlated data from the training set, known as temporal autocorrelation.

Details of our cross-validation approach can be found in Fig. A1 in appendix A. The dataset was divided into seven folds with three contiguous years each. Six cross-validation rotations were performed. The years 2018–2020, inclusive, were held out as the testing dataset for all rotations as they were the most current data in the dataset and might allow the most representative view into how the model would generalize to future data. The fold dedicated to the validation dataset was

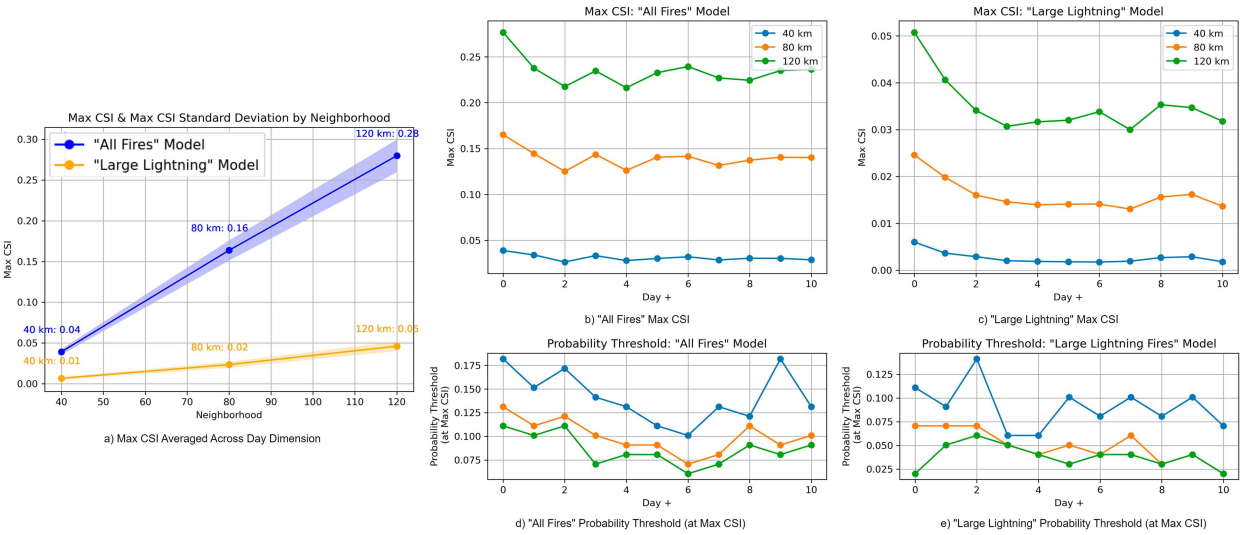


FIG. 5. Max CSI values and probability thresholds.

rotated through the remaining six folds, and the five folds not dedicated to testing or validation in each rotation was left for the training dataset. We assume that by dividing the data between training, validation, and testing using this method, independence is maintained between the three datasets. We 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 2022).

d. 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 submetrics, the probability of detection (POD), which is the percent of events predicted, and the false alarm rate (FAR), which is the ratio of false alarms to predicted events. At the heart of the CSI is the 2×2 contingency matrix, capturing hits, misses, false alarms, and correct rejections.

$$\mathrm{CSI} = [(\mathrm{POD})^{-1} + (1 - \mathrm{FAR})^{-1} - 1]^{-1}. \qquad (3)$$

As the models explored in this work were trained using a neighborhood loss function, a neighborhood approach to measuring model performance was also necessary. We measured CSI for neighborhoods of size 40, 80, and 120 km, with 40 km representing a hit directly on the label pixel itself, 80 km representing a hit within one neighboring pixel of the label, and 120 km representing a hit within two neighboring pixels of the label. We chose to stop at a 120-km neighborhood because it was as large as was operationally useful for our subject matter experts.

5. Results

In this work, model performance was explored using multiple approaches. First, general model performance was compared using CSI. CSI was generated by comparing the predictions of each model to the model's observation values for the test dataset. The All Fires model used observed pixel labels generated using all fire occurrence instances from FPA-FOD. The Large Lightning model used observed pixel labels generated using only fire occurrence instances with a cause of natural and a final fire size greater than 1000 acres from FPA-FOD. Additional general performance information, using the same CSI calculations just described, are available in the appendix in the form of performance plots and reliability diagrams for all 10 days of prediction and both models in Figs. B1–B4 in appendix B.

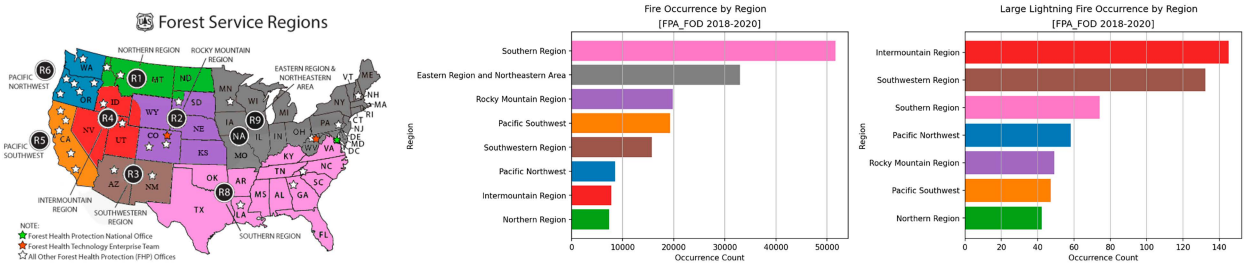
Second, general model performance CSI was stratified by region and season to enable us to further explore where and when each model performed best and to lay bare potential opportunities for improvement.

Third, the models were 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 performed on the same general subset of labels.

Finally, the two models were compared on how they performed on the largest lightning fires in the test dataset relative to climatology performance using the probability of fire generated from each model for the day of discovery and days preceding discovery and those grid cells representing the 40-, 80-, and 120-km neighborhoods.

a. General performance

To compare general model performance using CSI, the maximum CSI value was taken from the performance plots depicted in Figs. B1 and B2 in appendix B and plotted for all days 0–10 with day 0 representing the current day, day 1 representing the next day, and so on. All days were given inputs from day −1, the day prior to the current day. Model predictions produced



a) USFS Fire Regions b) Fire Occurrence by Region c) Large Lightning Fire Occurrence by Region

FIG. 6. (a)–(c) Fire regions; (a) is adapted from the U.S. Forest Service (2023).

by each cross-validation rotation were averaged together to produce the predictions used to calculate CSI. There are three important trends that can be observed in Fig. 5.

First, model performance did not demonstrate a dramatic downward trend across corresponding thresholds as predictions moved away from the current day in time. 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 consistently outperformed both the 80-km neighborhood and the 40-km neighborhood. 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 explored and, as a result of its larger size, allows the model more opportunities to be "right." This can be seen in Fig. 5 and Figs. B1 and B2 in appendix B. Seen for the All Fires model in the reliability plots, depicted in Figs. B3 and B4 in appendix B, overforecasting is a common trend for the 40-km neighborhood, while underforecasting is a common trend for the 120-km neighborhoods. The 80-km neighborhood achieves some level of reliability for early predictions before moving into overforecasting. For the Large Lightning model, overforecasting is common for all neighborhoods.

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 Fig. 5. The higher CSI values observed from the All Fires model may be associated with probability predictions which demonstrate less localization. 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 the use of data which includes consideration for fuel and topography (both of which change at a slower rate than weather data). Nonzero probabilities covering larger areas of CONUS produced by the All Fires model can be observed in all three of our case studies (Figs. 11–13).

b. *Regional performance*

To break CSI results down by region, we used the Forest Service Regions created by the U.S. Forest Service (2023), as depicted in Fig. 6. Figure 7 depicts the regional stratification of maximum CSI values for both models, and from it, two themes can be observed.

First, maximum CSI performance differed by region as can be observed in Fig. 7. Some regions had higher maximum CSI performance than other regions, as in the 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 the 120-km neighborhood. Some regions had lower maximum CSI performance than other regions, such as the eastern region and northeastern area, 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 80- and 120-km neighborhoods. When a region performed better than other regions, two possible interpretations were that 1) the region of interest with higher performance contained more of the data or 2) the inputs selected for the model were more representative of fire behavior in that region. Most of the cases cited herein were subject to the first interpretation, save the case of the Pacific Southwest region for all neighborhoods for the All Fires model, which was likely subject to the second interpretation.

Fire occurrence counts by region (which apply to the All Fires model) and large lightning fire occurrence counts by region (which apply to the Large Lightning model) can be observed in Fig. 6. It is important to remember, when observing Fig. 6, that the All Fires model included all fires, regardless of size, cause, or region. We can see in Fig. 6b that the southern region is home to the majority of fire occurrences, regardless of cause. 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, the average maximum CSI performance was higher for all regions, for all neighborhoods, and for the All Fires model than for the Large Lightning model. Figure B5 in appendix B depicts the average maximum CSI performance for all regions for all neighborhoods for both models.

c. *Seasonal performance*

To break CSI performance down by season, we divided the year into quarters by month. Figure 8 describes which months were associated with each season. Figure 9 depicts the maximum CSI performance stratified by season for each neighborhood for both models and, in it, can be seen a repeat of themes seen in the regional stratification of performance.

First, maximum CSI performance differed by season, as depicted in Fig. 9. Some seasons performed better than others, as

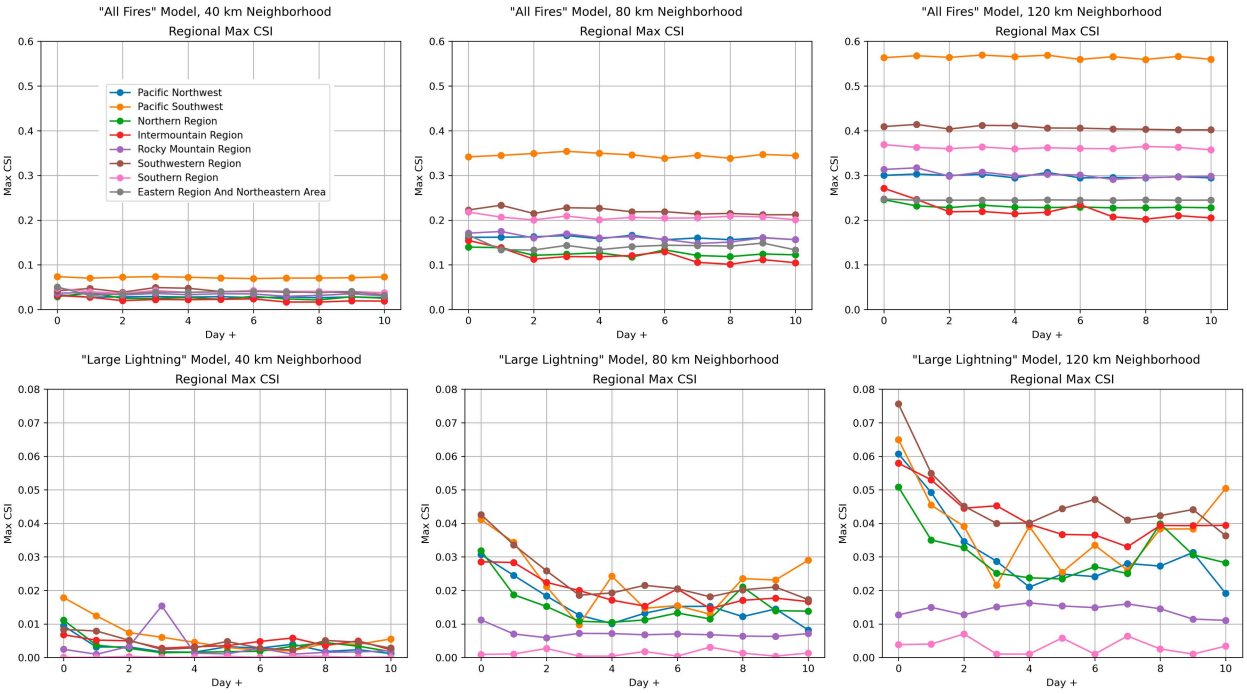


FIG. 7. Max CSI values by region for the All Fires model and the Large Lightning model.

with the spring and summer seasons for all neighborhoods and the All Fires model and with the summer season for all 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. Fire occurrence counts by season (which applies to the All Fires model) and large lightning fire occurrence counts by season (which applies to the Large Lightning model) can be observed in Fig. 8.

Second, average maximum CSI performance was greater for all seasons for all neighborhoods for the All Fires model than for the Large Lightning model, as depicted in Fig. B6 in appendix B.

d. *Large lightning performance*

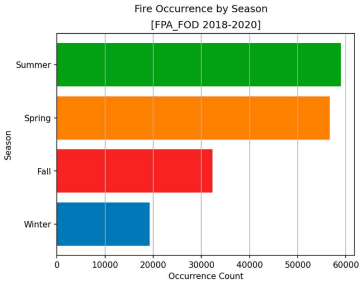
To compare the All Fires model to the Large Lightning model more directly, we took the predictions generated by each model for only a subset of days and grid cells subject to large lightning fires and calculated CSI performance for each model using this subset of predictions. Figure 10 depicts the resulting maximum CSI values for each model. As with general performance, model performance differed by neighborhood, with the 120-km neighborhood outperforming the 40- 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 outperformed the Large Lightning model.

e. *Case studies*

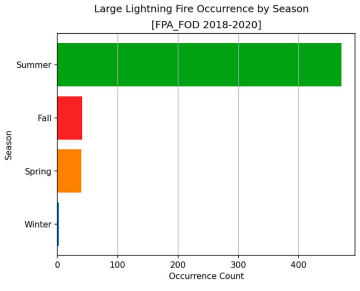
For case studies, we selected the largest lightning-caused fires from each year of the test dataset, which included 2018, 2019, and 2020. We compared the average (averaged across all cross-validation rotations) probability of fire produced by each model to the probability of wildfire produced by the SPC

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

a) Season by Month



b) Fire Occurrence by Season



c) Large Lightning Fire Occurrence by Season

FIG. 8. Fire seasons.

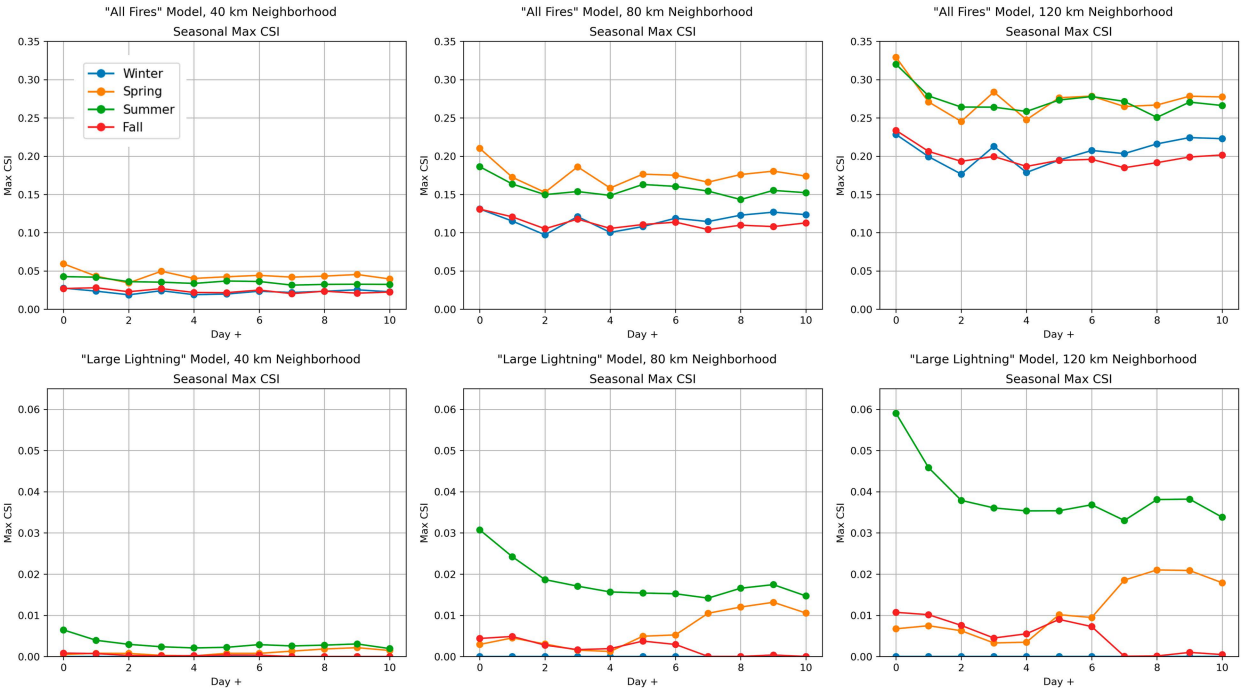


FIG. 9. Max CSI values by season for the All Fires model and the Large Lightning model.

wildfire climatology (N. Nauslar 2020). SPC's climatology uses an 80-km grid and reports the probability of wildfire within 25 miles, or approximately 40 km, and sources FPA-FOD years 1992–2015. SPC's climatology produces probabilities between 1% and 25%. The models in this work used a 40-km grid and reported the probability of wildfire within 40, 80, and 120 km and sourced FPA-FOD years 2000–2020. 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, 4 July, and for lightning-caused fire, 22 July.

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 11 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 0.05 and 0.1, with an average value of 0.075, for the 10 days prior to the discovery date for the location of discovery for the South Sugarloaf Fire. The All Fires model predicted a greater average probability of fire than the average probability produced by the climatology starting 3 days prior to the discovery date and greater than the maximum probability produced by the climatology on to the discovery date for the

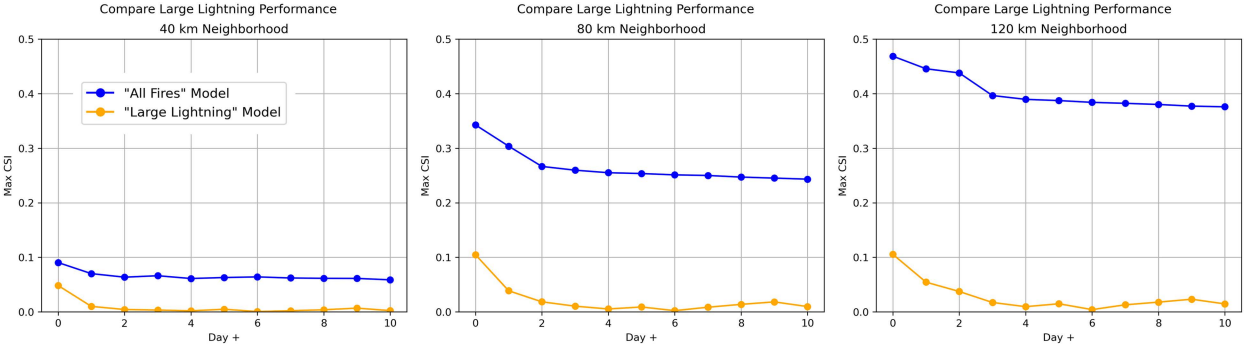
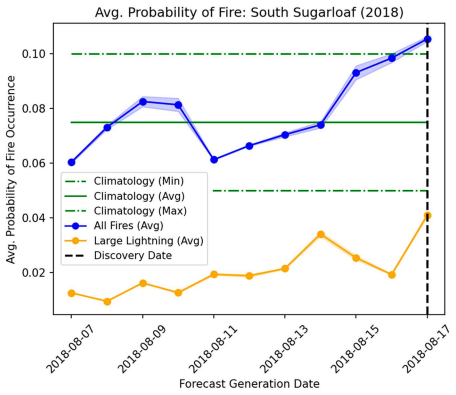
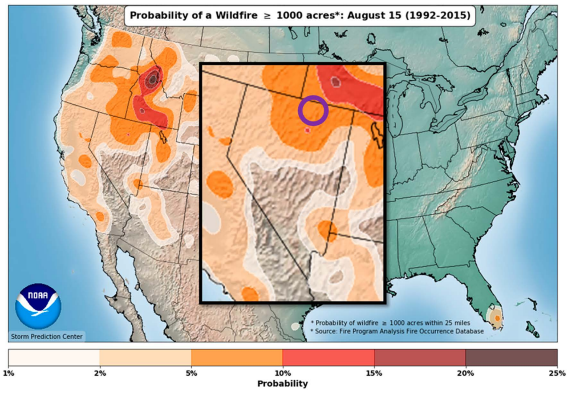


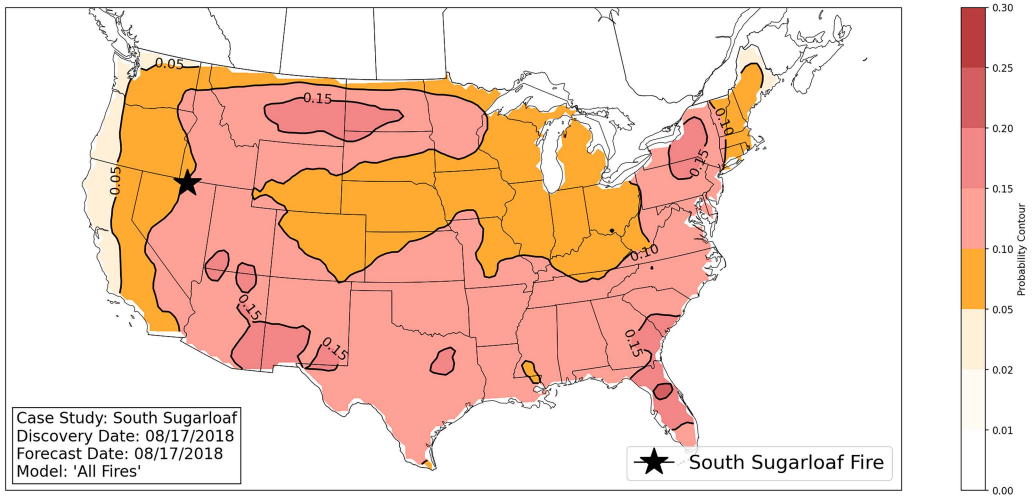
FIG. 10. Max CSI values for Large Lightning labels for the All Fires model and the Large Lightning model.



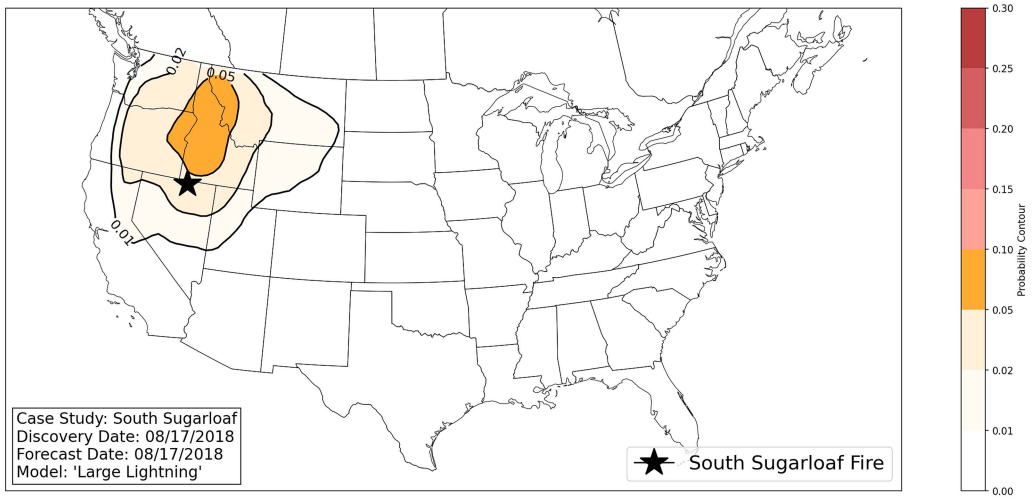
a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology



b) SPC Climatology Probability of Fire

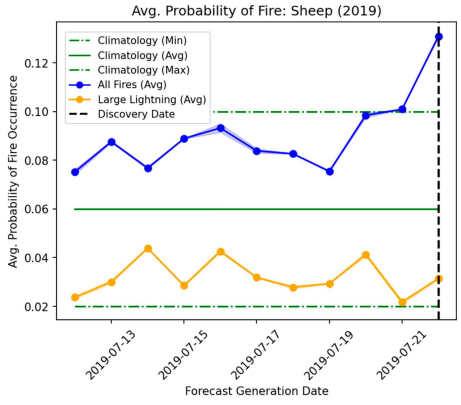


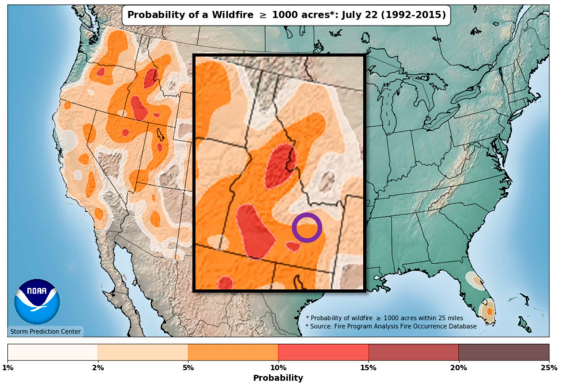
c) "All Fires" Model Probability of Fire



d) "Large Lightning" Model Probability of Fire

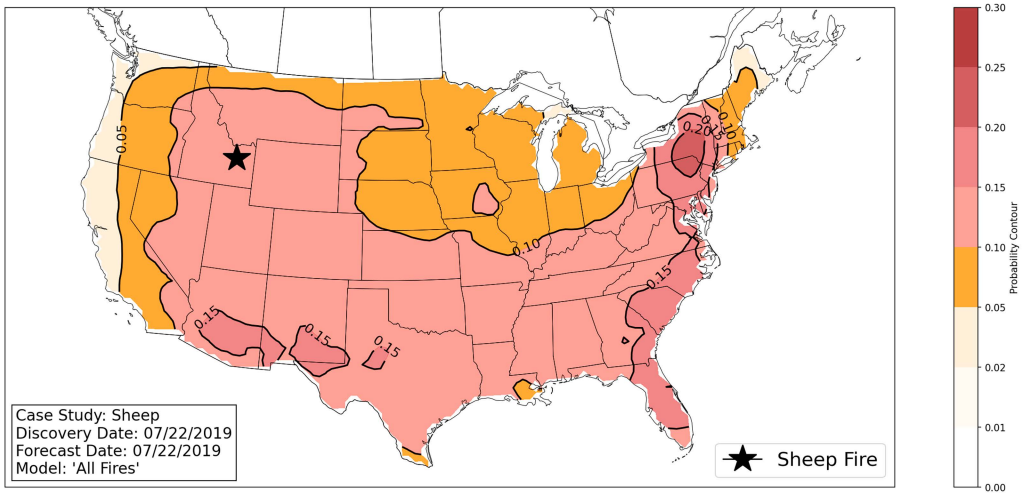
FIG. 11. Probability of fire: South Sugarloaf Fire, 2018 largest lightning fire.



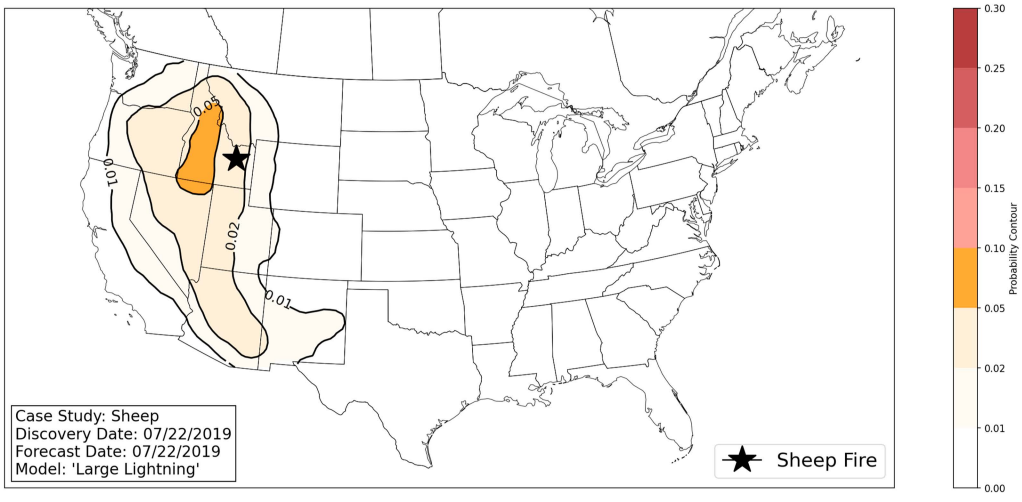


a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology

b) SPC Climatology Probability of Fire

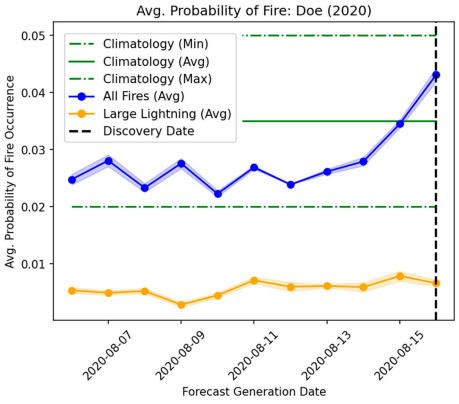


c) "All Fires" Model Probability of Fire

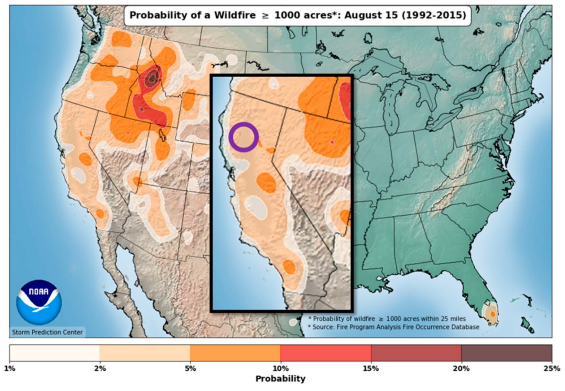


d) "Large Lightning" Model Probability of Fire

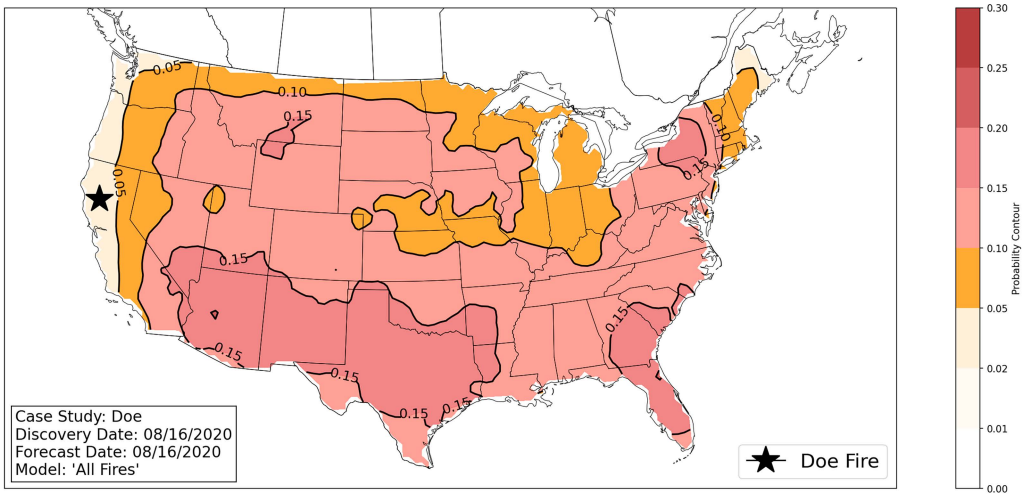
FIG. 12. Probability of fire: Sheep Fire, 2019 largest lightning fire.



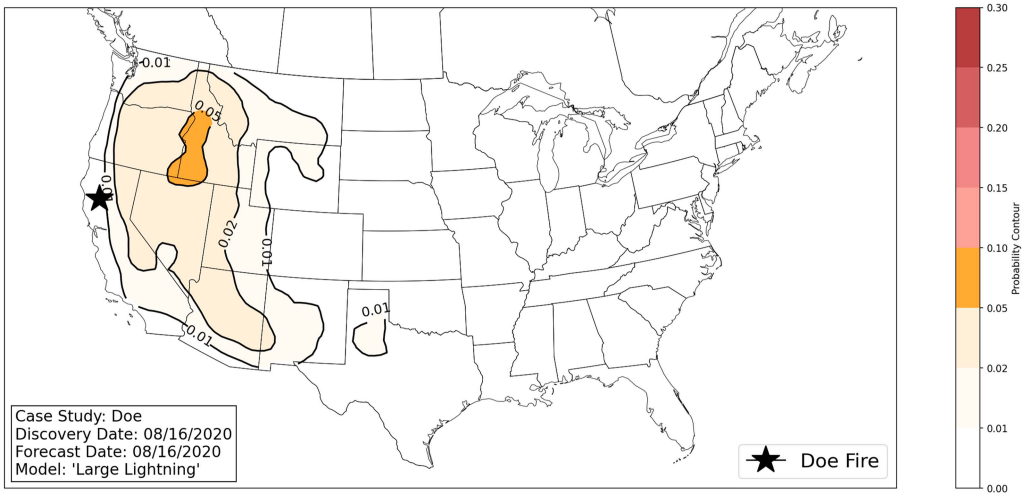
a) Average & Error Probability of Fire
(Aggregated Across Neighborhood)
for "All Fires" vs. "Large Lightning" vs. SPC Climatology



b) SPC Climatology Probability of Fire



c) "All Fires" Model Probability of Fire



d) "Large Lightning" Model Probability of Fire

FIG. 13. Probability of fire: Doe Fire, 2020 largest lightning fire.

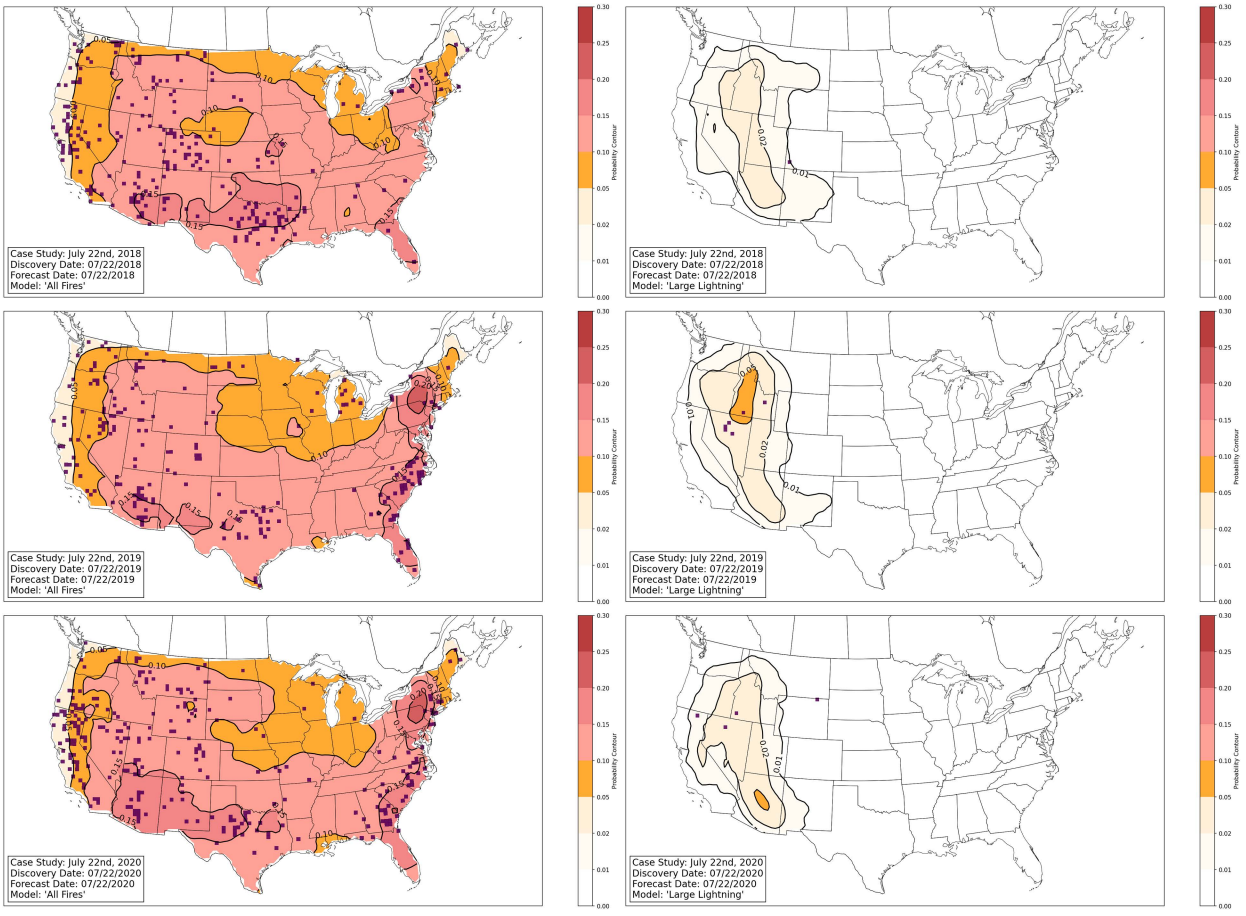


FIG. 14. Probability of fire: 22 Jul 2018, 2019, and 2020.

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 10 days prior to the discovery date.

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 12 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 0.025 and 0.1, with an average of 0.0625, for the 10 days prior to the discovery date for the location of discovery for the Sheep Fire. The All Fires model predicted a greater average probability of fire than the average climatology for all 10 days prior to the discovery date and a greater average probability than the maximum climatology 1 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, did produce average probabilities greater than the climatology minimum for all 10 days prior to the discovery date.

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 13 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 0.02 and 0.05, with an average of 0.035, for the 10 days prior to the discovery date for the location of discovery for the Doe Fire. The All Fires model predicted a 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 10 days prior to the discovery date.

4) MOST COMMON DAY FOR LIGHTNING-CAUSED FIRES: 22 JULY

According to the work of Balch et al. (2017), across the 21 years of data from FPA-FOD which they evaluated, the most common day for lightning-caused fires in the contiguous United States is 22 July. Figure 14 depicts model performance

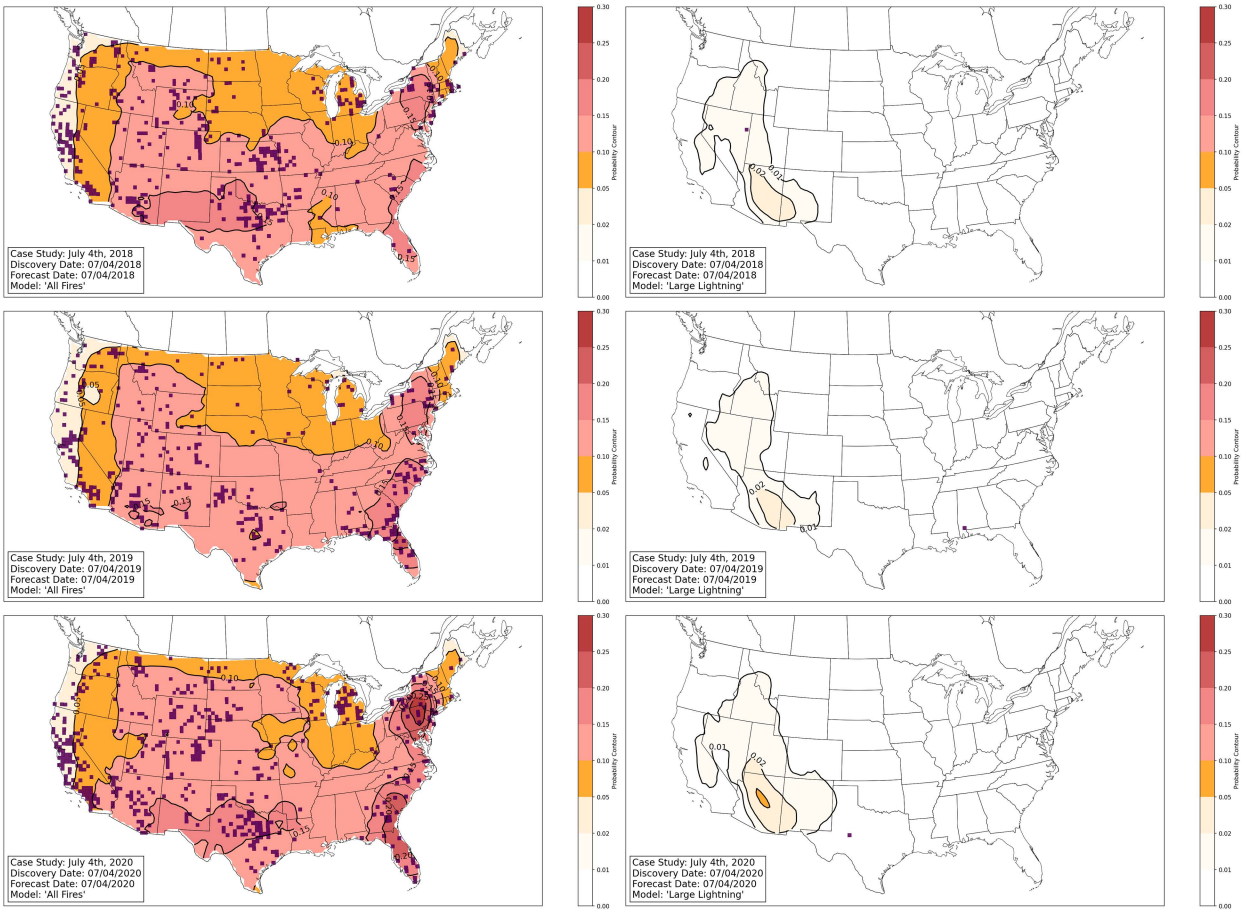


FIG. 15. Probability of fire: 4 Jul 2018, 2019, and 2020.

for 22 July 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 3 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) MOST COMMON DAY FOR HUMAN-CAUSED FIRES: 4 JULY

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 these reasons, we feel it is important not to lose sight of model performance in the context of human-caused fires. Figure 15 depicts model performance on the most common day for human-caused wildfires in CONUS, 4 July, 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.

6. Discussion and future work

In this work, we presented a proof of concept for an operationalizable fire occurrence prediction model for CONUS. We explored the comparison of the All Fires model to the Large Lightning model. The All Fires model outperformed the Large Lightning model in all three methods of comparison: general performance, large lightning performance, and case study performance.

First, the All Fires model produced higher CSI than the Large Lightning model on all three neighborhoods (40, 80, and 120 km). Second, the All Fires model and the Large Lightning model were compared based on how they performed on the subset of days and grid cells subject to large lightning fires. Here again, the All Fires model produced a higher CSI than the Large Lightning model for all neighborhoods calculated. Third, the All Fires model and the Large Lightning model were compared based on how each model

performed on the largest of large lightning fires for the 3 years contained within the test dataset: the 2018 South Sugarloaf Fire, the 2019 Sheep Fire, and the 2020 Doe Fire (as part of the August Complex). In all three case studies, the All Fires model produced higher CSI values than the Large Lightning model for all 10 days leading up to the date of discovery for each fire. In all cases, the All Fires model was able to outperform the SPC wildfire climatology, at least on the discovery date of the fire, as in the case of the 2020 Doe Fire, and, at most, 10 days prior to the discovery date of the fire, as in the case of the 2019 Sheep Fire. In none of the three cases, on none of the days calculated, was the Large Lightning model able to produce higher CSI values than the SPC wildfire climatology or the All Fires model.

In summary, the All Fires model produced higher CSI values than the Large Lightning model in general, on large lightning fires, and on the largest of large lightning fires present in the most current data in the dataset. The All Fires model resulted in a model simultaneously more performative overall and on the phenomenon of interest even when applied to extreme examples of the phenomenon. We explore the reasons for these performance improvements in Earnest et al. (2024). We also saw performance improvement from both models as the size of the neighborhood increased from 40 to 120 km, with 120 km offering the best performance.

From the comparison of the All Fires model to the Large Lightning model, we observed the value of including more fire occurrence labels rather than less and will carry this learning into our future work. Our future work will focus on refining the predictions made by the All Fires model so they produce a prediction with the performance of the All Fires model and the localization of the Large Lightning model. We will explore two methods for achieving this end. The first method will be to include phenomenon-specific variables to our All Fires model, specifically lightning, dry thunderstorms, and population density to boost the signal for specific cases of interest. The second method will be to explore the impact of optimizing the All Fires model to both All Fires labels and Large Lightning labels.

From our exploration of different methods for codifying model inputs, we learned the value of retaining the entire variable distribution and will carry this learning forward into our future work as well. From our inability to quantify variable importance in the context of model input variables, described in detail in Earnest et al. (2024), we found that multicollinearity between input variables was a vulnerability of our current input dataset. To address this limitation, we will explore datasets and variables with less correlation in our future work. Datasets we plan to explore are LANDFIRE (LF), which includes vegetation and fuel national geospatial data, and observed and forecast weather data from the Global Ensemble Forecast System (GEFS).

Acknowledgments. This material was prepared by Bethany Earnest with funding provided by NOAA/Office of Oceanic and Atmospheric Research under the NOAA–University of Oklahoma Cooperative Agreement NA21OAR4320204, U.S. Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NOAA or the U.S. Department of Commerce. This material is based on work supported by the National Science Foundation under Grant ICER-2019758. Thank you to William McGovern-Fagg for his adaptation of the color scheme for Fig. 6. Thank you to Dr. Randy Chase and to Dr. Monique Shotande. During their respective tenures as postdoctoral researchers, they each helped to evolve this research, my code, and my thinking. Without them, this work would not be what it is. It was an honor to learn from you; thank you both.

Data availability statement. Data analyzed in this study were a reanalysis of existing data, which are openly available at locations cited in the reference section. Further documentation about data processing will be available prior to publication.

APPENDIX A

Appendix Title

Figure A1 and Tables A1–A4 describe which input variables we used, how our cross validation was done, and how our UNet3+ and loss function were configured.

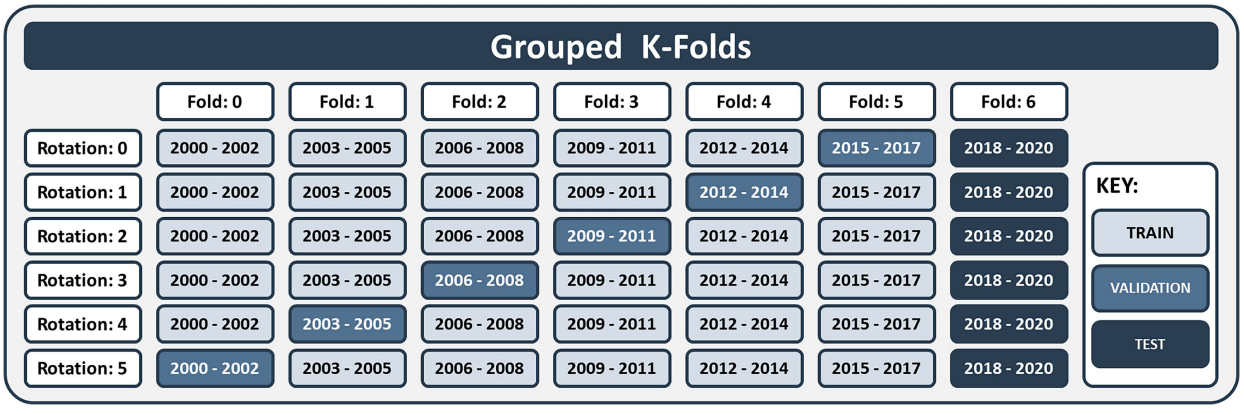


FIG. A1. Grouped *k*-fold cross validation.

TABLE A1. gridMET input variables.

Variable	Description
Latitude	Latitude of centroid of grid cell
Longitude	Longitude of centroid of gridcell latitude
Day of year	Julian day of inputs
Dead fuel moisture 100 h	Moisture content of dead organic fuels—1–3-in. diameter class
Dead fuel moisture 1000 h	Moisture content of dead organic fuels—3–6-in. diameter class
Burning index	The contribution of fire behavior to the effort of containing a fire
ERC	National Fire Danger Rating System index related to how hot a fire could burn
Elevation	Height above sea level
Palmer drought severity index (and associated category)	Estimate of relative soil moisture conditions
Wind speed	Wind velocity at 10 m
Wind from direction	Wind direction
Precipitation amount	The amount of rain, snow, hail, etc., that has fallen at a given place within a given period
Specific humidity	The weight of water vapor contained in a unit weight of air
Relative humidity (minimum)	The amount of atmospheric moisture present relative to the amount that would be present if the air were saturated
Surface downwelling shortwave flux in air	The sum of direct and diffuse solar radiation incident on the surface
Mean vapor pressure deficit	The difference between the amount of moisture in the air and how much moisture the air can hold
Potential evapotranspiration (daily)	The combined loss of water through the plant's process of transpiration via its vascular system and evaporation of water from Earth's surface

TABLE A2. U-Net 3+ model configuration details.

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 A3. Adam optimizer configuration details.

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

TABLE A4. FSS loss function configuration details.

Argument	Setting
fss_c	1
fss_mask_size	3

APPENDIX B

Appendix Title

Figures B1–B4 describe the individual model performance plots, one for each day, for each model, and the individual model reliability plots, again, one for each day, for each model. Figures B5–B6 summarize model performance by region and season.

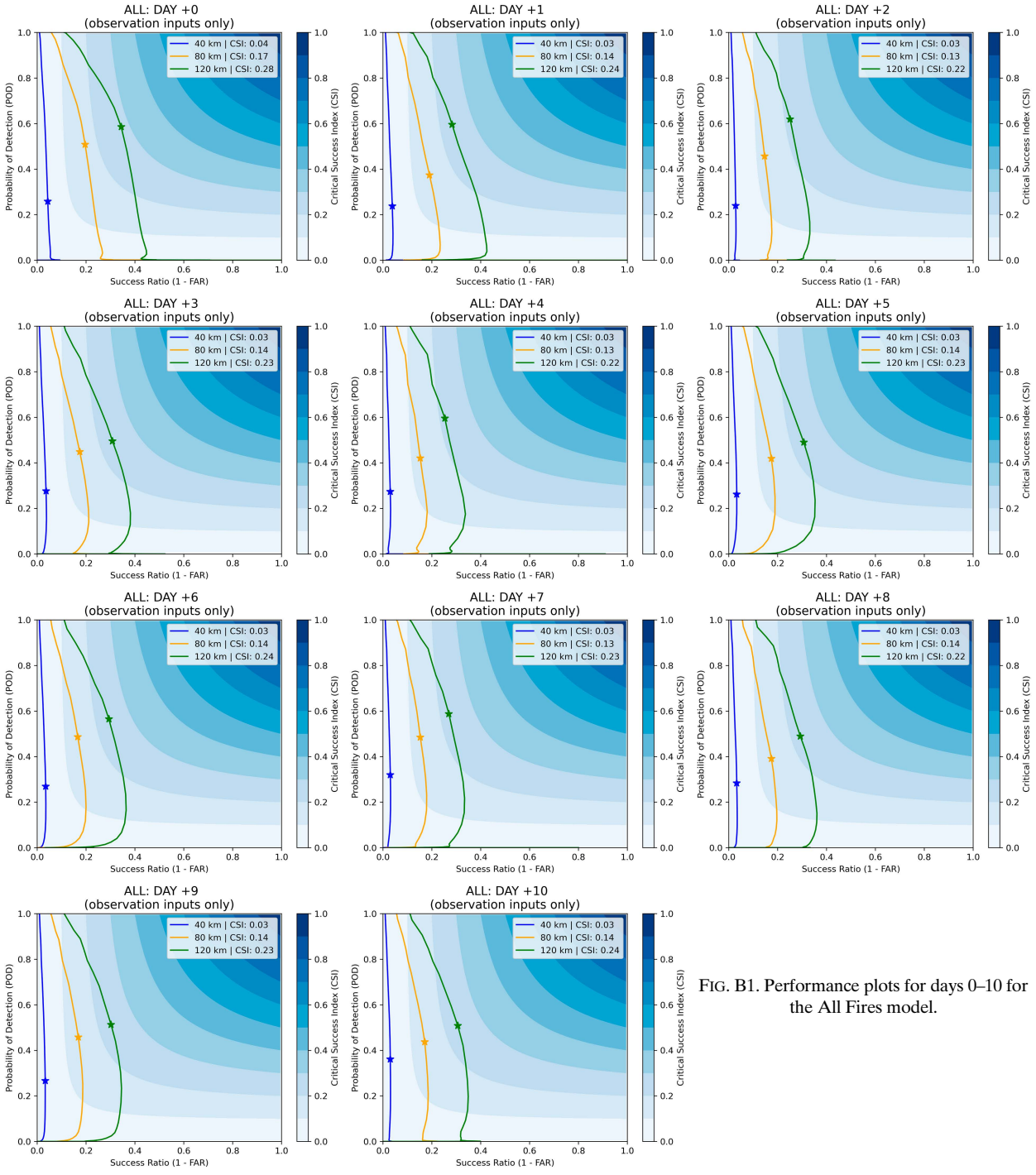


FIG. B1. Performance plots for days 0–10 for the All Fires model.

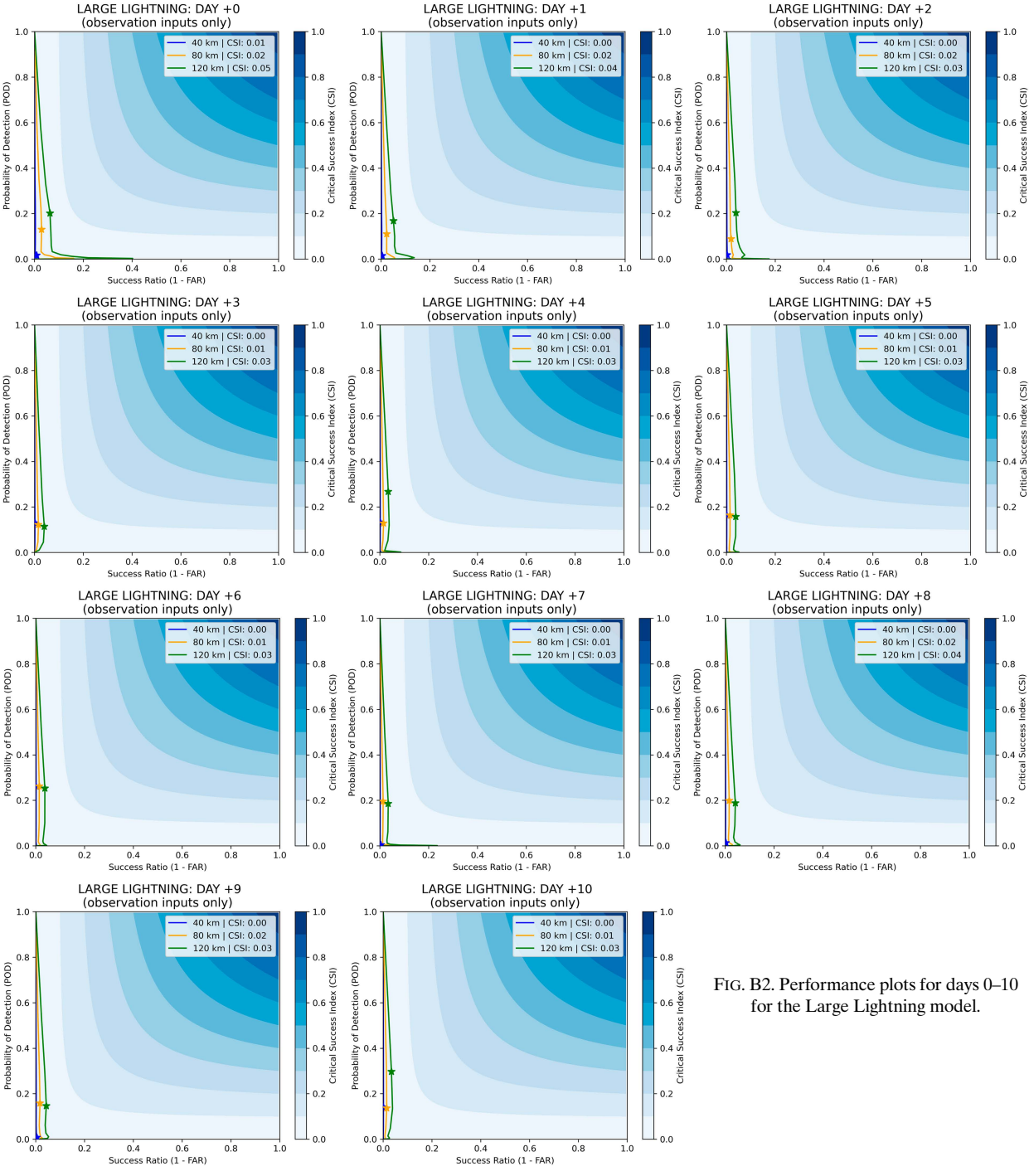


FIG. B2. Performance plots for days 0–10 for the Large Lightning model.

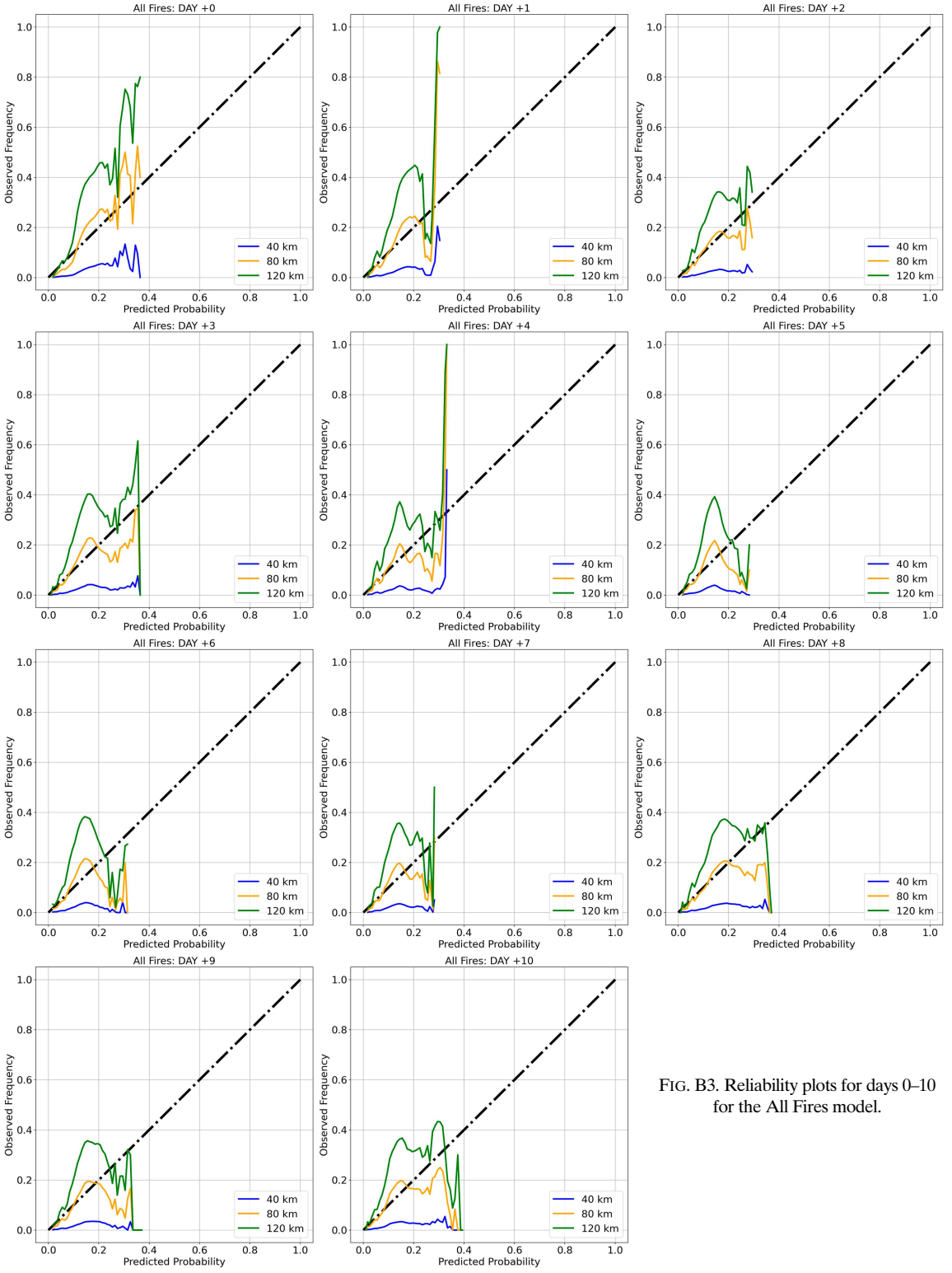


FIG. B3. Reliability plots for days 0–10 for the All Fires model.

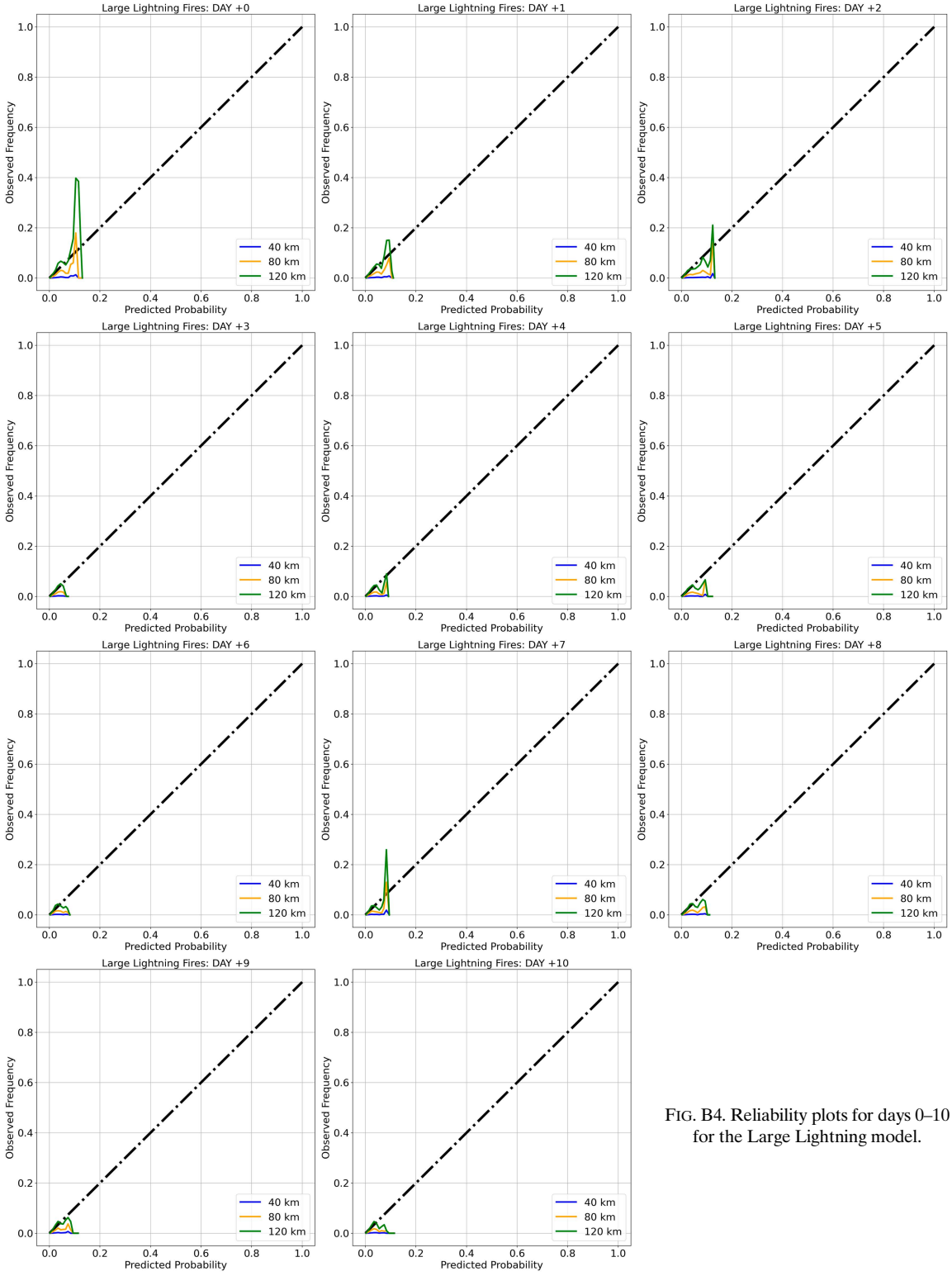


FIG. B4. Reliability plots for days 0–10 for the Large Lightning model.

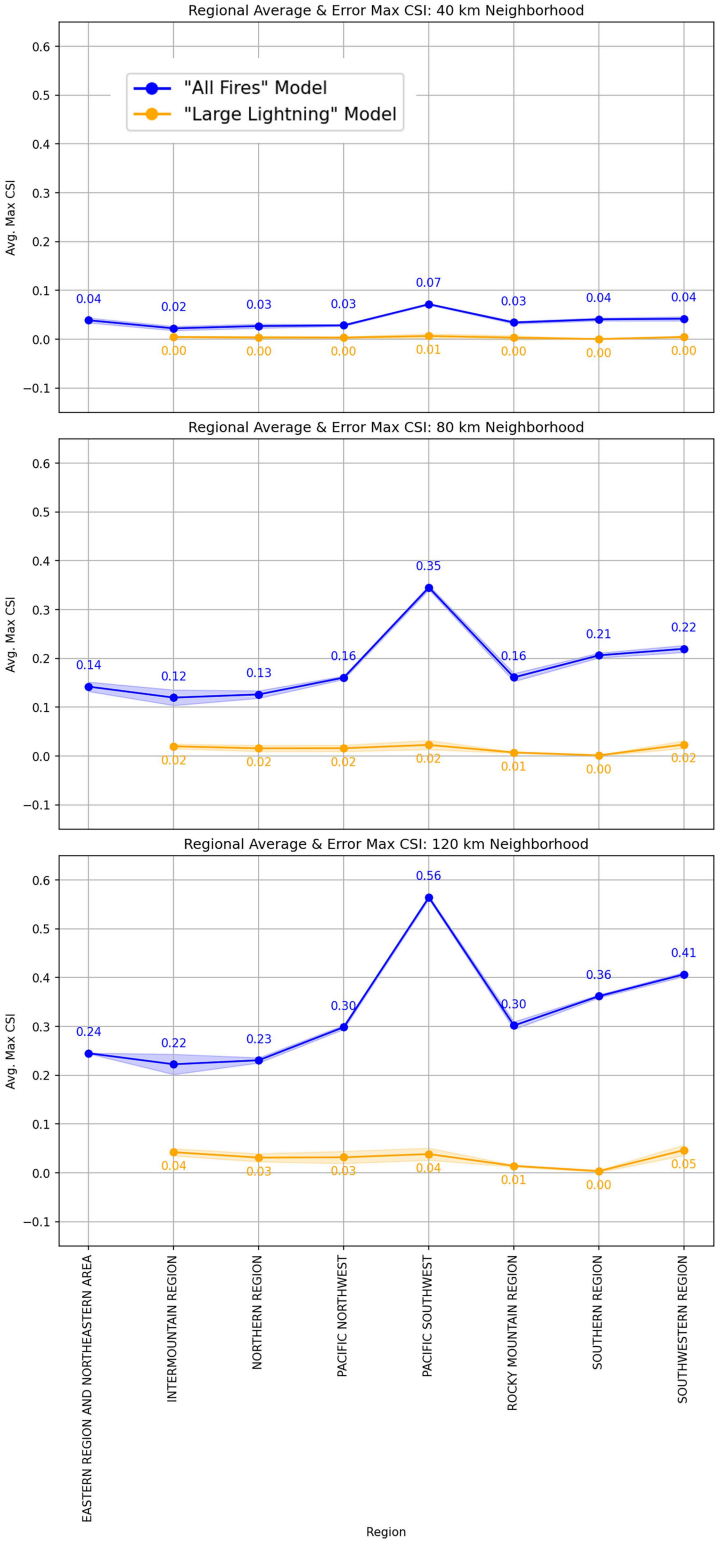


FIG. B5. Average and error max CSI by region (aggregated across day dimension).

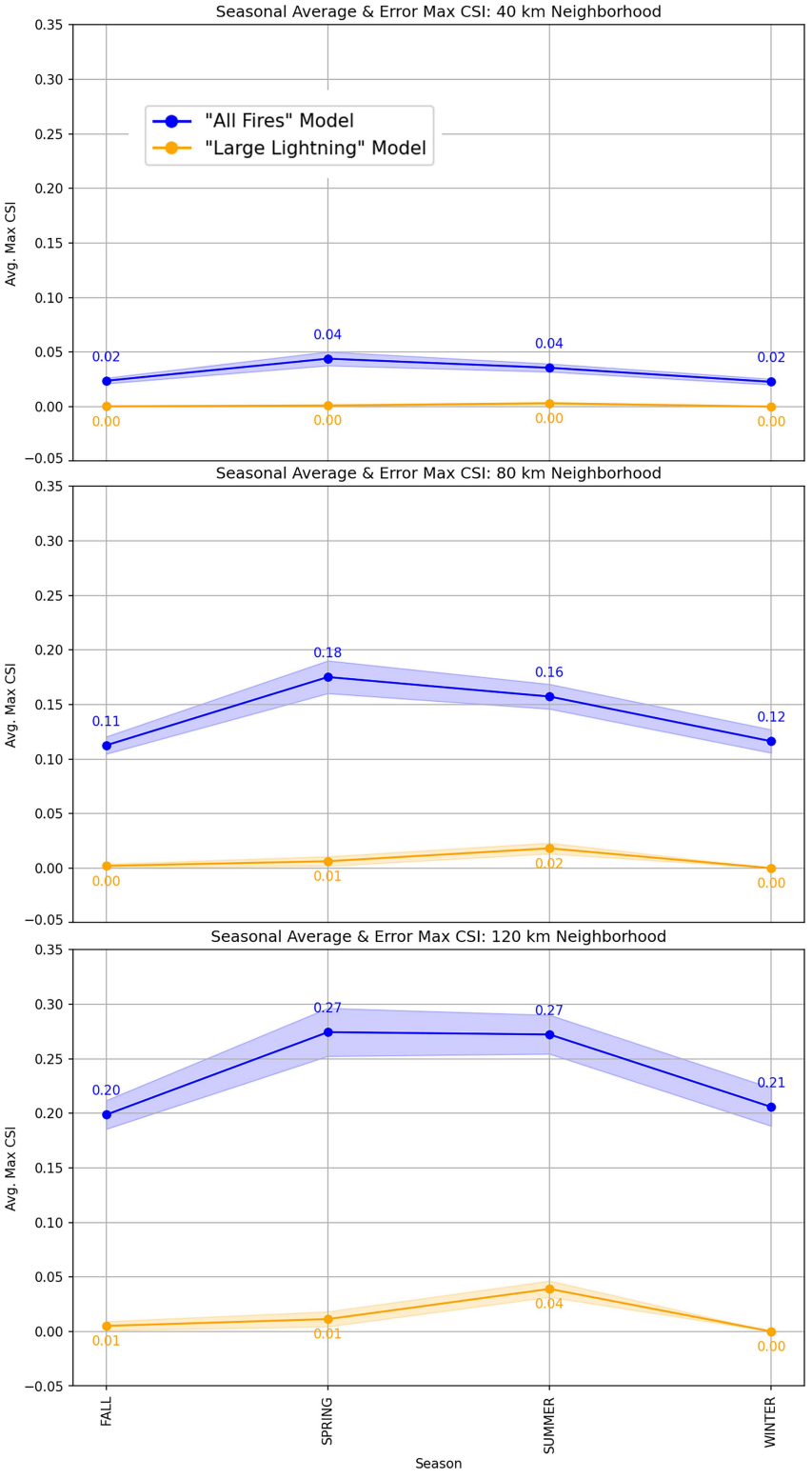


FIG. B6. Average and error max CSI by season (aggregated across day dimension).

REFERENCES

Abatzoglou, J. T., 2013: Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.*, **33**, 121–131, https://doi.org/10.1002/joc.3413.

Alonso-Betanzos, A., and Coauthors, 2002: A neural network approach for forestal fire risk estimation. *ECAI'02: Proceedings of the 15th European Conference on Artificial Intelligence*, IOS Press, 643–647.

——, O. Fontenla-Romero, B. Guijarro-Berdiñas, E. Hernández-Pereira, M. I. P. Andrade, E. Jiménez, J. L. L. Soto, and T. Carballas, 2003: An intelligent system for forest fire risk prediction and firefighting management in Galicia. *Expert Syst. Appl.*, **25**, 545–554, https://doi.org/10.1016/S0957-4174(03)00095-2.

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. *Proc. Natl. Acad. Sci. USA*, **114**, 2946–2951, https://doi.org/10.1073/pnas.1617394114.

Cal Fire, 2020: August complex (includes Doe fire). Mendocino National Forest, accessed 26 May 2023, https://www.fire.ca.gov/incidents/2020/8/16/august-complex-includes-doe-fire/.

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**, 3188, https://doi.org/10.1038/srep03188.

——, A. Das, and J. Aryal, 2016: Big data integration shows Australian bush-fire frequency is increasing significantly. *Roy. Soc. Open Sci.*, **3**, 150241, https://doi.org/10.1098/rsos.150241.

Earnest, B. L., A. McGovern, C. Karstens, and I. Jirak, 2024: Part II: Lessons learned from predicting wildfire occurrence for CONUS using deep learning and fire weather variables. *Artif. Intell. Earth Syst.*, **3**, e230058, https://doi.org/10.1175/AIES-D-23-0058.1.

Fusco, E., J. Abatzoglou, J. K. Balch, J. T. Finn, and B. A. Bradley, 2016: Quantifying the human influence on fire ignition across the western USA. *Ecol. Appl.*, **26**, 2390–2401, https://doi.org/10.1002/eap.1395.

Huang, H., and Coauthors, 2020: UNet 3+: A full-scale connected UNet for medical image segmentation. arXiv, 2004.08790v1, https://doi.org/10.48550/arXiv.2004.08790.

Idaho National Laboratory, 2019: One year after historic sheep fire, Idaho site contractors, agencies eye coming fire season. Idaho National Laboratory, accessed 26 May 2023, https://inl.gov/article/after-sheep-fire/.

Jain, P., S. C. P. Coogan, S. G. Subramanian, M. Crowley, S. Taylor, and M. D. Flannigan, 2020: A review of machine learning applications in wildfire science and management. *Environ. Rev.*, **28**, 478–505, https://doi.org/10.1139/er-2020-0019.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021: Physics-informed machine learning. *Nat. Rev. Phys.*, **3**, 422–440, https://doi.org/10.1038/s42254-021-00314-5.

National Wildfire Coordinating Group, 2022: Fire ecology. NWCG Glossary of Wildland Fire, accessed 10 February 2022, https://www.nwcg.gov/publications/pms205/nwcg-glossary-of-wildland-fire-pms-205.

Nauslar, N., 2020: Storm prediction center probability of wildfire. NOAA NWS, accessed 25 May 2023, https://www.spc.noaa.gov/new/FWclimo/climo.php?parm=1000ac.

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, **136**, 78–97, https://doi.org/10.1175/2007MWR2123.1.

Sakr, G. E., I. H. Elhajj, and G. Mitri, 2011: Efficient forest fire occurrence prediction for developing countries using two weather parameters. *Eng. Appl. Artif. Intell.*, **24**, 888–894, https://doi.org/10.1016/j.engappai.2011.02.017.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575, https://doi.org/10.1175/1520-0434(1990)005<0570:TCSIAA>2.0.CO;2.

Short, K. C., 2022: Spatial wildfire occurrence data for the United States, 1992-2020 [fpa_fod_20221014] 6th edition. Forest Service Research Data Archive, accessed 31 May 2024, https://doi.org/10.2737/RDS-2013-0009.6.

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.

U.S. Forest Service, 2023: USFS fire regions. USFS USDA, accessed 12 May 2023, https://www.fs.usda.gov/foresthealth/contact-us/regional-contacts.shtml.

Vasilakos, C., K. Kalabokidis, J. Hatzopoulos, G. Kallos, and Y. Matsinos, 2007: Integrating new methods and tools in fire danger rating. *Int. J. Wildland Fire*, **16**, 306–316, https://doi.org/10.1071/WF05091.

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. *Agric. For. Meteor.*, **225**, 36–47, https://doi.org/10.1016/j.agrformet.2016.05.003.

Vega-Garcia, C., B. S. Lee, P. M. Woodard, and S. J. Titus, 1996: Applying neural network technology to human-caused wildfire occurrence prediction. *AI Appl.*, **10**, 9–18.