

Article

Not peer-reviewed version

Probabilistic Forecasting of Lightning Strikes over Continental US and Alaska: Model Development and Verification

[Ned Nikolov](#)*, Phillip Bothwell, John Snook

Posted Date: 17 January 2024

doi: 10.20944/preprints202401.1281.v1

Keywords: lightning; model; logistic regression; forecast; prediction; wildfire; probability



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Probabilistic Forecasting of Lightning Strikes over Continental US and Alaska: Model Development and Verification

Ned Nikolov ^{1,*}, Phillip Bothwell ² and John Snook ²

¹ USDA Forest Service Rocky Mountain Research Station, Fort Collins, CO 80526;
nedialko.nikolov@usda.gov

² Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80521

* Correspondence: nedialko.nikolov@usda.gov

Abstract: Lightning is responsible for most annually burned area by wildfires in the extratropical region of the Northern Hemisphere. Hence, predicting the occurrence of wildfires requires reliable forecasting of the chance of cloud-to-ground lightning strikes during storms. Here, we describe the development and verification of a probabilistic lightning-strike algorithm designed to run on a uniform 20-km grid over the Continental USA and Alaska. The algorithm consists of a large set of logistic equations parameterized via logistic regression using long-term data records of observed lightning strikes and meteorological reanalysis fields from NOAA. Principal Component Analysis was employed to extract 13 Principal Components (strong predictors) from a list of 611 potential predictors. Our statistical analysis revealed that the occurrence of cloud-to-ground lightning strikes primarily depends on three factors: horizontal temperature distribution by pressure levels, amount of low-level atmospheric moisture, and wind vectors. These physical variables impact the vertical separation of electric charges in the lower troposphere during storms causing the voltage potential between ground and the cloud deck to increase to a level that triggers electrical discharges. Results from a forecast verification using independent data showed excellent model skill, thus making this algorithm suitable for inclusion into software systems forecasting the chance of wildfire ignitions.

Keywords: lightning; model; logistic regression; forecast; prediction; wildfire; probability

1. Introduction

Lightning has been an increasing cause for wildfires in recent decades. Although historically lightning only causes 15% - 20% of wildfire occurrences in the USA, lightning-ignited wildfires account for about 60% of the annually burned acreage. In the extratropical portion of the Northern Hemisphere, lightning is responsible for 77% of the burned wildfire area every year [1]. Hence, the capability to predict lightning 7-10 days in advance on a continental scale using operational, numerical weather forecasts as drivers is a key prerequisite for any modeling efforts aimed at quantifying the chances of wildfire ignition over large regions.

In the USA, the National Predictive Services (NPS) is an interagency group tasked with the operational delivery of 7-Day Wildfire Potential Outlooks to assist fire-management operations. NPS approached the Rocky Mountain Center (RMC) for Fire-Weather Intelligence, an applied research unit at the US Forest Service Rocky Mountain Research Station (RMRS), with a request to develop a system of statistical models to predict wildfire ignition probabilities on a 20-km national grid using numerical weather forecasts by the National Weather Service as input. To improve the forecast accuracy of wildfire ignitions, RMC devoted resources to build a set of lightning forecast equations as a first step in the project. This paper describes the methodology employed by RMC to derive a suite of regional, monthly logistic models capable of estimating the probability of cloud-to-ground (CG) lightning strikes from forecast weather fields on a 20-km grid over Continental USA (ConUS) and Alaska (AK), and the verification of these model equations against independent data.

2. Materials and Methods

The overall approach of this study followed the original methodology proposed by Bothwell [2], which was later also adopted by Buckey [3] and Richardson [4]. We used data sets of much longer duration to derive new and improved lightning prediction equations for ConUS and AK.

2.1. Georeferenced Data Sets

Table 1 lists the datasets utilized for ConUS. Two 29-year long 3-hour resolution climatological records of weather reanalysis fields and observed lightning strikes over ConUS were used to derive the lightning-forecast regression equations. The NOAA North American Regional Reanalysis (NARR) [5] containing 3-D fields from 00:00, 03:00, 06:00, 09:00, 12:00, 15:00, 18:00, and 21:00 UTC for the period from January 1990 to August 2018 were downloaded and archived in GRIB2 format. The lightning data for the lower 48 states (ConUS) were obtained from the Western Regional Climate Center at the Desert Research Institute [6] for the period from January 1990 through August 2018. Three-hour forecast gridded fields extending out to 10 days produced by the NOAA Global Forecast System (GFS) [7] were downloaded and archived in GRIB2 format (0.25 x 0.25 degree) for the fire season (April – September) of 2018 to verify the lightning forecast equations against independent data.

Table 1. Datasets utilized in the lightning-forecast model development and testing for ConUS.

Dataset Name & Description	Record Length (years)	Binary Size (Gbit)
North American Regional Reanalysis (NARR) [5]: 3-D gridded historical weather dataset provided by NOAA.32-km horizontal resolution interpolated to 20-km resolution for ConUS with a 3-hour time step.	29	3,000
Lightning dataset provided the Western Regional Climate Center [6] gridded to 3-hour time steps at 20-km resolution.	29	400
NCEP GFS Forecast Fields [7]: 3-hour time steps from 0 to 240 hours/10 days; 3-D grids of 0.25 x 0.25-degree resolution interpolated to 20-km resolution used for forecast testing and verification purposes and 2019 forecasts.	1	8,000

The development of the lightning forecast model for AK used similar datasets as the ConUS model but covered a shorter time period. NARR was used as a source of observed gridded weather fields and the Alaska Lightning Detection Network (ALDN) [8] owned and maintained by the Alaska Fire Service (AFS) [9] was utilized as a source of geo-referenced lightning-strike data. The training data sets for AK spanned a period of 10 years (from 2012 through 2021). The reason for this was that, during the summer of 2012, ALDN replaced VISALA’s Impact sensors detecting lighting flashes with Time-Of-Arrival (TOA) sensors recording lightning strikes. A comparison study conducted by AFS found that the new TOA ALDN detects about 2.25 times more lightning events than the older VISALA sensors. About 60% of this increase is attributable to a difference in detecting strikes vs. flashes (a flash can contain multiple strikes) with the remaining 40% increase resulting from an improved detection efficiency, expanded spatial sensor coverage and a longer detection range. Since the TOA sensors will continue to be used in the future, we decided to exclude all lightning-flash data collected by VISALA Impact sensors prior to 2012 from our analysis.

2.2. Statistical Methods and Procedures of Model Development

Lightning strikes are binomially distributed events meaning that lightning occurrence is a variable with only two possible outcomes, 0 or 1 (i.e. absent or present). The probability of binary variables can be described by logistic equations or sigmoid functions. Our model employs logistic

regression, which is in essence a supervised machine learning algorithm that performs binary classification tasks by predicting the probability of occurrence of an event [10]. The model development proceeded in four stages.

First, lightning data and NARR weather fields were aggregated and re-projected to a 20-km resolution grid. Each month of the active fire season (spanning the period from March through November for ConUS and from May through September for AK) was represented by an average diurnal cycle of 3-hour resolution containing 8 temporal bins. Logistic equations were derived for each temporal bin of every fire-season month. Due to scarcity of lightning data during winter months over ConUS, the quarter from December through February was represented by a single diurnal set of eight 3-h long temporal bins. Since AK sees virtually no lightning events between October 01 and May 01, we did not derive forecast equations for the fall, winter and early spring months for that state. In order to improve the overall skill of the lightning forecast model, ConUS was divided into 10 climatological regions by taking into account spatial differences in topography, fuels, climate, and soil moisture (Fig. 1). A single region was used for AK.

The second step of data processing involved the creation of weather and lightning climatologies for each temporal bin in every fire-season month using 27-year long data record for ConUS and a 10-year record for AK.

Thirdly, a Principal Component Analysis (PCA) with orthogonal rotation [11] was applied to the gridded NARR and lightning climatological datasets to analyze 611 potential meteorological drivers (predictors). PCA reduced the large cohort of 3-D meteorological drivers to a smaller subset of 13 statistically significant lightning predictors.

Finally, logistic regressions were performed using the first 13 PCs along with lightning climatologies from every region and temporal bin to derive predictive equations for calculating probabilities of “one or more” and “10 or more” strikes. The logistic equations have the following general form:

$$P = \frac{1}{1 + e^{-\sum w_i \beta_i}} \quad (1)$$

where P is the probability of lightning strike, β_i is the i^{th} predictor (PC) and w_i is the estimated regression coefficient (weight) for that predictor. The “R” statistical package [12], an open-source software was utilized to perform both PCA and the logistic regressions.

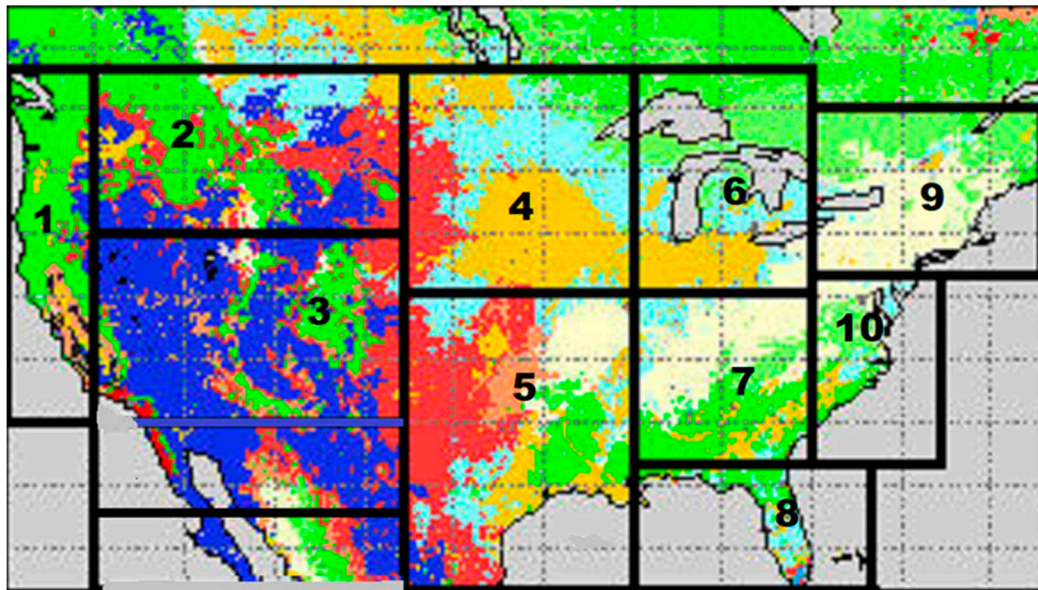


Figure 1. Climatological regions employed in the derivation of logistic lightning forecast equations over ConUS.

2.3. Model Verification Procedure

The model performance was evaluated using independent meteorological and lightning data from 2018 for ConUS and from 2022 for AK. These were data not included in the statistical parametrization of the logistic lightning equations. Meteorological drivers for the verification came from two sources: NARR reanalysis fields and GFS grids. To avoid biases introduced by forecast weather fields and allow a comparison between NARR-driven and GFS-driven lightning forecasts, we used GFS data from the 0-3 h initialization period. The skill of the lightning model was also evaluated under forecast weather conditions, where projected GFS fields out to 7 days were used.

The model verification procedure employed two primary statistical metrics: (a) Reliability Diagrams (RD) comparing model-forecast probabilities to observed relative frequencies of lightning strikes [13]; and (b) Receiver Operating Characteristic (ROC) curves quantifying the relationship between False Positive Rates (on the X axis) and True Positive Rates (on the Y axis) at different classification thresholds [14]. A key feature of the ROC diagrams is the Area Under the Curve (AUC), which provides an aggregate measure of the model performance across all possible classification thresholds. The closer is AUC to 1.0, the higher the spatial accuracy of the model forecasts. RDs and ROC curves were calculated for model predictions driven by observed weather fields from NARR and GFS (0-3 h initialization), and forecast GFS weather fields extending out to 7 days.

As a means of qualitative assessment, we generated maps overlaying forecast lightning probabilities with the observed number of lightning strikes in each 20-km grid cell. Specifically, we plotted the 24-h maximum lightning-forecast probabilities against observed lightning strikes over the same time period.

3. Results

PCA yielded 13 Principal Components (PCs) that were employed as predictors over the 10 ConUS Regions depicted in Fig. 1 and AK and for all 3-hour bins in every month. The predictors covered near-surface variables as well as vertical pressure levels from 700 mb to 100 mb at 50-mb increment (Fig. 2). The 13 PCs collectively explained nearly 61% of the total variance of CG strikes in all regions and time periods over ConUS (Fig. 2). Temperature and pressure height fields emerged as the strongest predictor among the 13 PCs accounting for 14.6% of the observed variance in CG strikes.

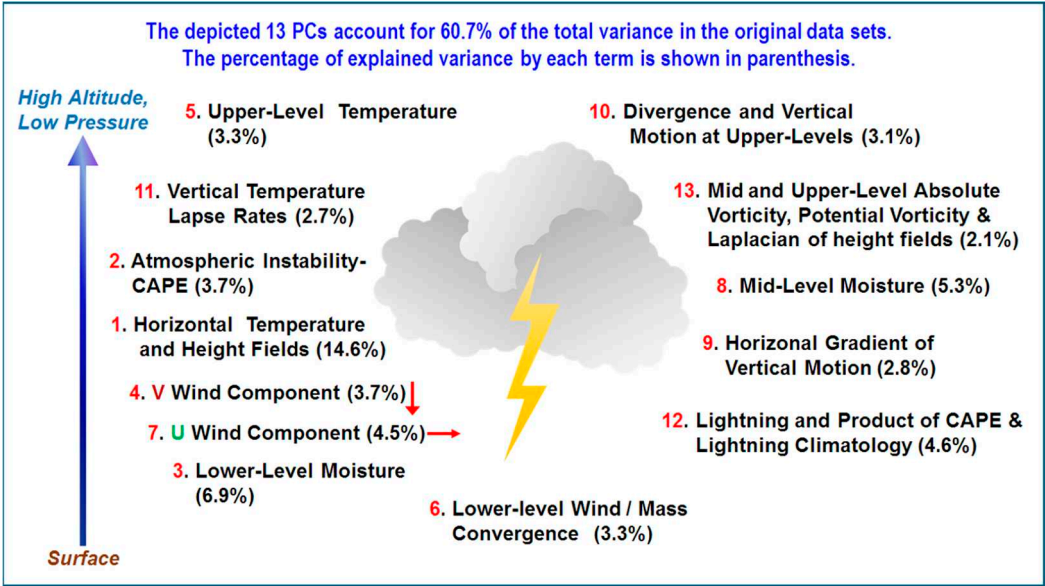


Figure 2. The 13 main lightning predictors (Principal Components) produced by PCA and the amount of variance of CG strikes explained by them over ConUS.

Logistic equations quantifying probabilities of CG strikes were derived using multi-year climatologies from each 3-h diurnal bin of every month and region employing a 27-year long record of NARR and lightning grids for ConUS and 10-year long data records for AK. Each month of year was described by 8 logistic equations per region, i.e. one equation for every 3-hour bin of the average

24-hour diurnal cycle for that month. This produced 728 equations for ConUS and 40 equations for AK.

Output files from the “R” software containing the lightning-prediction equations were combined with 3D meteorological fields generated by GFS to produce operational lightning probability forecasts at 20-km resolution out to 7 days for ConUS and AK. The forecasts were delivered in a graphic format to field users through a customized website [15].

3.1. Model Verification

Figure 3 compares side-by-side Reliability Diagrams of NARR- and GFS-driven 3-h forecasts for June, July and August of 2018 over ConUS employing independent meteorological and lightning data. Figure 4 shows the ROC curves and corresponding AUCs for the same data. The AUC values being close to or slightly above 0.9 indicates excellent-to-outstanding model accuracy in predicting probabilities of one or more CG lightning strikes over ConUS [16]. Surprisingly, the GFS-driven forecasts (using the initial 3 hours) performed slightly better than those driven by NARR data. This might be a result of recent code upgrades made to the GFS model by NOAA.

Figure 5 depicts RDs of 24-h maximum lightning probabilities for forecast days 1, 3 and 7 in August of 2018 over ConUS. It is important to note that forecast probabilities greater than 0.7 only occur in less than 0.1% of grid points and that 99.9% of the ConUS area exhibits a chance of lightning strikes less than 0.7 (70%). Thus, the larger deviations of the red curves from the 1:1 line for probabilities greater than 0.7 visible in Figures 3 and 5 are essentially immaterial in regard to the actual model skill. Figure 5 illustrates that the model also yielded reliable predictions up to 7 days in advance using forecast GFS weather fields as drivers.

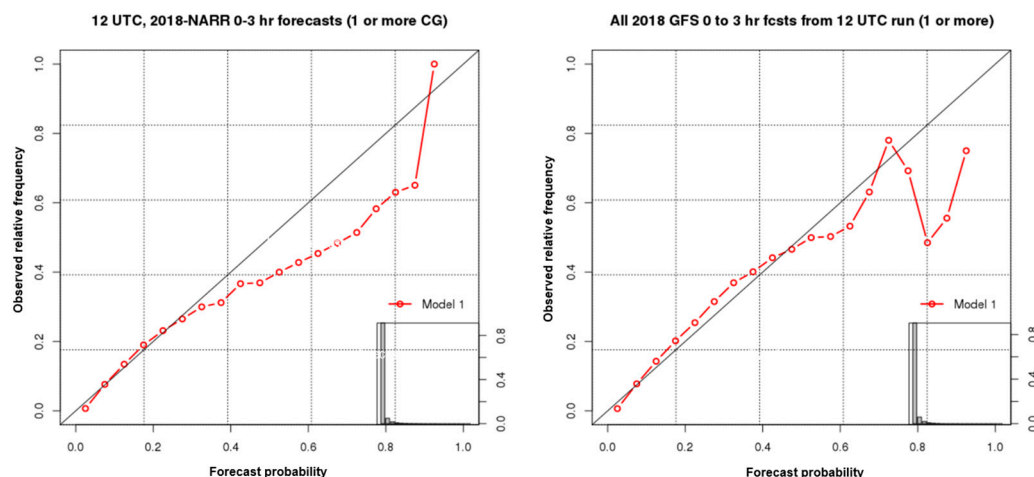


Figure 3. Reliability diagrams of NARR-based (left panel) and GFS-based (right panel) lightning forecasts for June, July and August of 2018 (3-hour forecasts based on 12 UTC data) over ConUS.

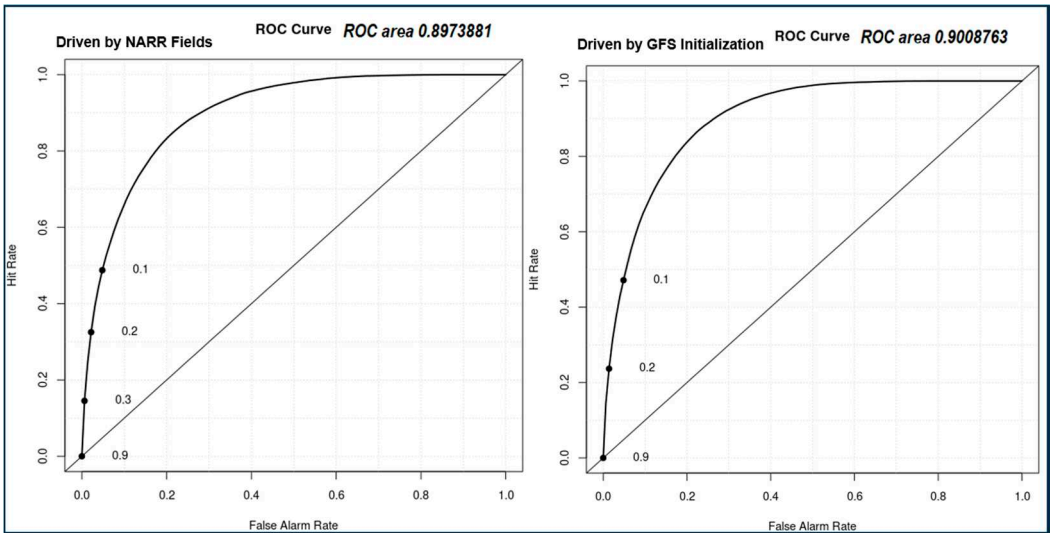


Figure 4. ROC curves with AUC shown in the upper right corner for lightning probability forecasts driven by NARR data (left panel) and GFS initialization fields (right panel) covering the same time period as in Fig. 3.

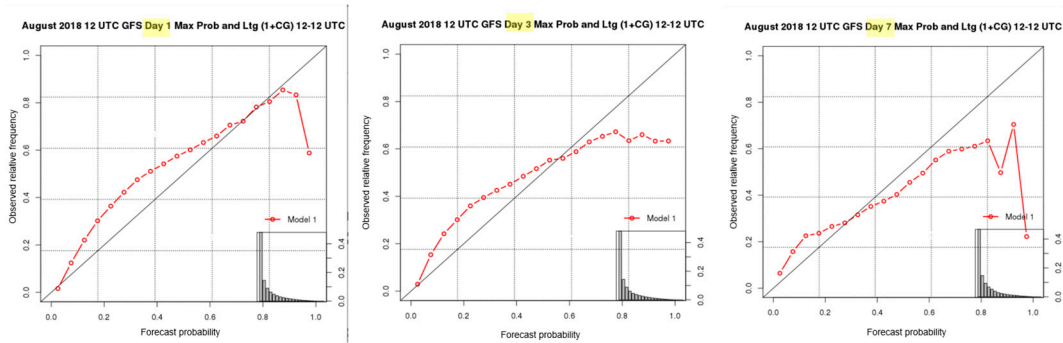


Figure 5. Reliability diagrams of GFS-driven daily maximum-probability lightning forecasts for days 1, 3, and 7 over ConUS.

Figure 6 depicts a Reliability Diagram of NARR-based lightning forecasts for Alaska covering the complete fire season (May 01 - Aug. 31) of 2022. Once again, the red curve is very close to a perfect model-data match (the 1:1 line) for all probabilities that are meaningfully represented over the AK domain.

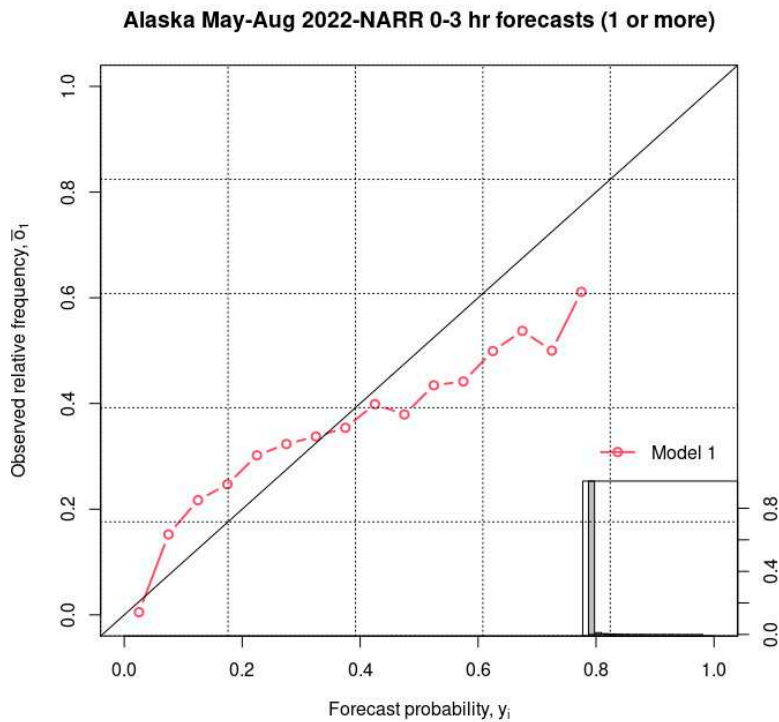


Figure 6. Reliability Diagram of modeled probabilities of one or more lightning strikes over AK for all 3-hour periods from May 01 through Aug 31, 2022. The lightning model was driven by NARR weather fields.

Figure 7 compares Reliability Diagrams of lightning predictions driven by either NARR data (left panel) or GFS initialization fields (right panel) during the most active portion of the 2022 fire season (Jun. 01 – Aug. 31). The NARR-based lightning forecasts display better accuracy than those based on GFS initialization fields.

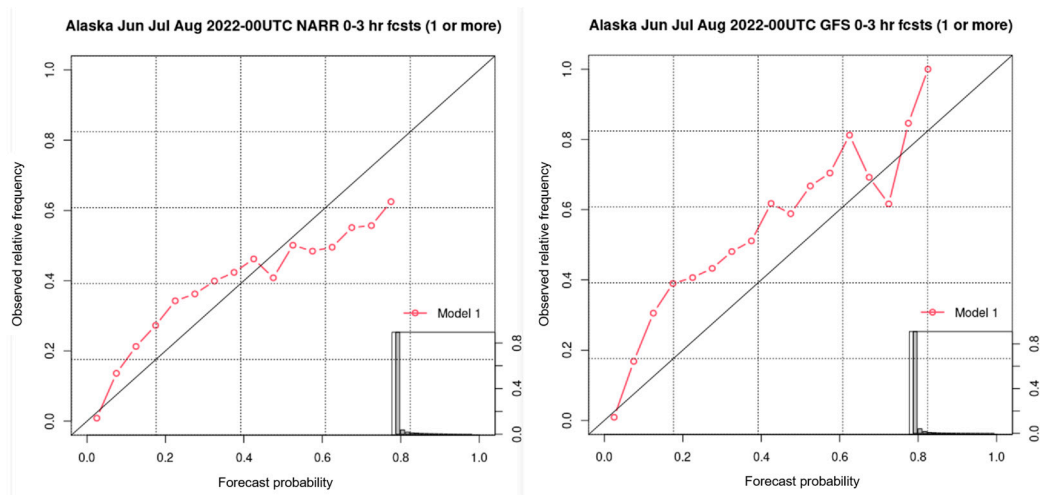


Figure 7. Reliability Diagram of modeled probabilities of 1 or more lightning strikes over AK for all 3-hour periods between June 01 and Aug 31, 2022. Left panel: model driven by NARR data (ROC AUC = 0.9398); Right panel: model driven by GFS initialization fields (ROC AUC = 0.9444).

Figure 8 shows RDs and ROC curves for lightning predictions over AK out to 168 hours driven by GFS forecast weather fields in June and July of 2022. As expected, the accuracy of lightning predictions deteriorates with uncertainty of the weather forecasts as it progresses through time. Still,

the ROC AUCs indicate an outstanding model accuracy up to forecast hour 99, and excellent accuracy up to hour 147

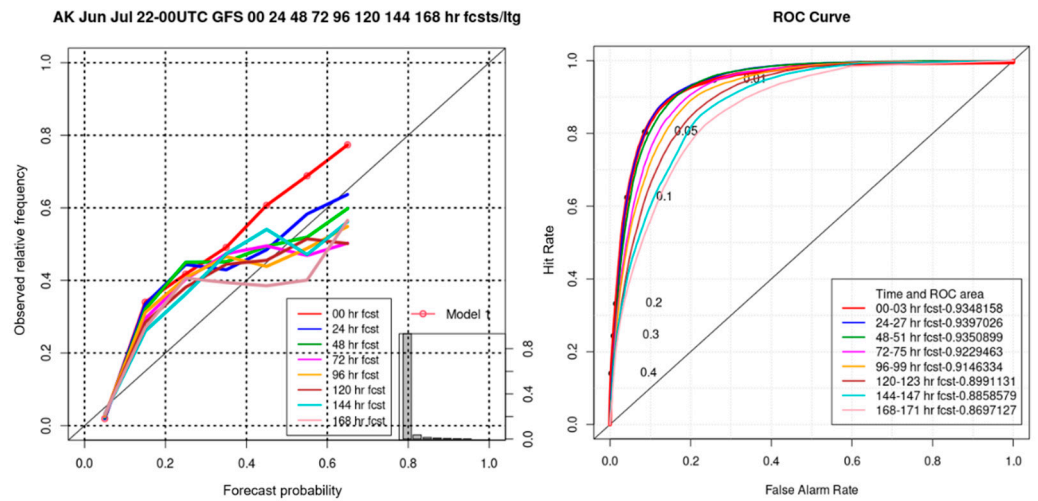


Figure 8. Reliability Diagrams (left panel) and ROC curves (right panel) of predicted probabilities of 1 or more lightning strikes over Alaska for all 3-hour periods between June 01 and July 31, 2022 based on forecast weather fields provided by GFS The AUC values are shown within the color legend of right panel.

Figure 9 displays observed lightning data overlaid on predicted 24-h maximum lightning probabilities over ConUS for forecast days 1, 3, 5 and 7 beginning on July 15, 2018. Figure 10 shows similar overlays for AK during two high-intensity lightning events that occurred on June 4 and June 20 in 2022. These overlay maps visually illustrate the ability of the lightning prediction model to successfully capture regions of intense lightning.

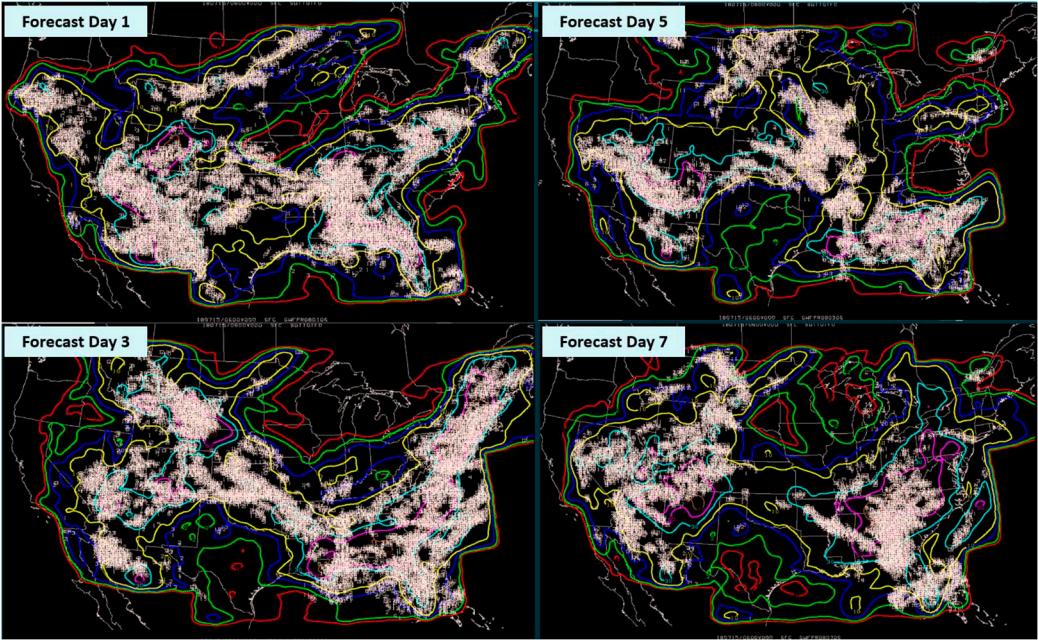


Figure 9. Maximum probability of lightning (color contours) predicted by the model over 24-h periods (12:00 to 12:00 UTC) for forecast days 1, 3, 5, and 7 beginning on July 15, 2018 overlaid on observed lightning strikes (numbers in white). Contours delineate: 1% red, 2% green, 5% blue, 10% yellow, 30% cyan, 50% magenta 50%, 70% brown.

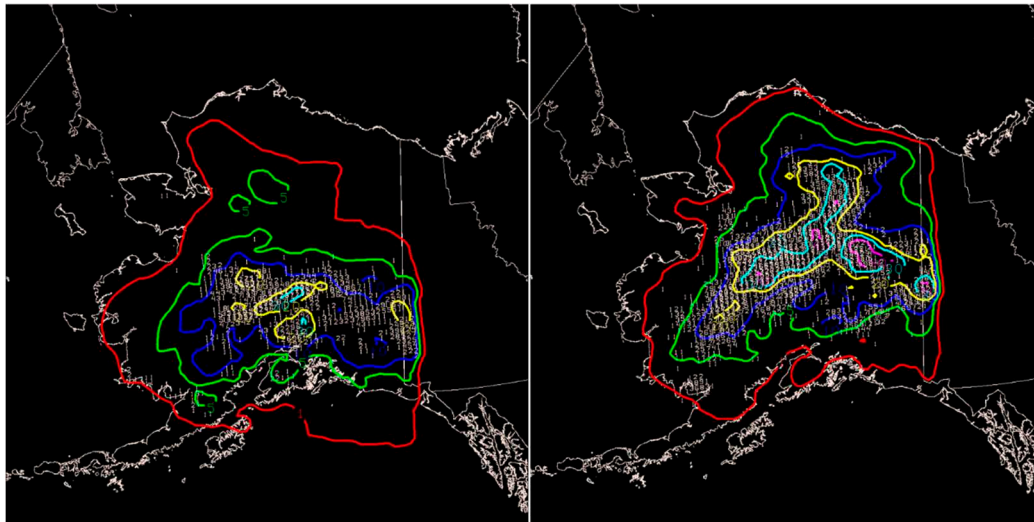


Figure 10. Lightning observations (number of strikes per grid cell in white) overlaid on the first 24-hour maximum probability forecast over AK shown as color contours of strike probabilities for June 4 (left panel) and June 20 (right panel), 2022: 1% (red); 5% (green); 10% (blue); 20% (yellow); 30% (cyan).

4. Discussion

We presented here the first and only gridded lightning-probability forecast algorithm for North America derived from 29-year long data records running on a uniform 20-km continental grid. The statistical procedure of model derivation revealed that the occurrence of lightning strikes primarily depends on 3 factors: the horizontal temperature distribution by pressure levels, the amount of low-level atmospheric moisture, and the wind vectors. These physical variables apparently aid in controlling the vertical separation of electric charges in the lower troposphere during storms, increasing the voltage potential between the cloud deck and the ground to a level that triggers electrical discharge.

Results obtained from the verification discussed in Section 3.1. employing independent data indicate that probabilistic predictions of CG lightning strikes are feasible and can be statistically reliable up to 7 days in advance when driven by output from operational, numerical weather forecast models such as GFS. Hence, the new lightning prediction algorithm can be incorporated into future forecasting systems of fire-ignition probabilities to quantify the risk of naturally occurring wildfires on a national level.

Author Contributions: Conceptualization, Ned Nikolov and Phillip Bothwell; Methodology, Phillip Bothwell; Software, Phillip Bothwell and John Snook; Validation, Phillip Bothwell and Ned Nikolov; Formal analysis, Phillip Bothwell; Investigation, Ned Nikolov; Resources, John Snook; Data curation, Phillip Bothwell and John Snook; Writing—original draft preparation, Ned Nikolov; Writing—review and editing, Phillip Bothwell and John Snook; Visualization, Phillip Bothwell and Ned Nikolov; Supervision, Ned Nikolov; Project administration, Ned Nikolov; Funding acquisition, Ned Nikolov. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by a Joint Venture Agreement between USFS Rocky Mountain Research Station and the Cooperative Institute for Research in the Atmosphere (CIRA) at Colorado State University (CSU) in Fort Collins, CO. Early stages of this project were also partially supported by the US Bureau of Land Management.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: NARR (<https://rda.ucar.edu/datasets/ds608-0/>); Lightning observations over ConUS (<https://wrcc.dri.edu/>); Lightning observations over Alaska (<https://fire.ak.blm.gov/predsvcs/maps.php>).

Acknowledgments: The authors thank Edward D. Delgado, a former Program Manager of the National Predictive Services in Boise ID for his vision to launch this project and his ability to secure partial, initial interagency funding for it. The authors acknowledge the critical administrative role played in this project by Dr.

Kyle Hilburn at CSU CIRA and the indispensable IT assistance provided to the RMC team by Steven Finley also at CSU CIRA.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Janssen, T.A.J.; Jones, M.W.; Finney, D. et al. Extratropical forests increasingly at risk due to lightning fires. *Nat. Geosci.* 2023, 16, 1136–1144. <https://doi.org/10.1038/s41561-023-01322-z>
2. Bothwell, P.D. Prediction of Cloud-to-Ground Lightning in the Western United States, Ph.D. Dissertation, University of Oklahoma. 2002, 178 pp.
3. Buckley, D. Using the Perfect Prognosis Method to Forecast Cloud to Ground Lightning in Alaska. M.S. Thesis. 2009, Department of Meteorology, University of Oklahoma, 82 pp.
4. Richardson, L.M. A perfect prog approach to forecasting dry thunderstorms over the CONUS and Alaska. M.S. Thesis. 2013, School of Meteorology, University of Oklahoma, 85 pp.
5. North American Regional Reanalysis (NARR): <https://rda.ucar.edu/datasets/ds608-0/>
6. Western Regional Climate Center at the Desert Research Institute: <https://wrcc.dri.edu/>
7. NOAA Global Forecast System (GFS): <https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast>
8. Alaska Lightning Detection Network (ALDN): <https://fire.ak.blm.gov/predsvcs/maps.php>
9. Alaska Fire Service (AFS): <https://www.blm.gov/programs/fire-and-aviation/regional-info/alaska-fire-service>
10. Rymarczyk, T.; Kozłowski, E.; Kłosowski, G.; Niderla, K. Logistic regression for machine learning in process tomography. *MDPI Sensors* 2019, 19(15), 3400. <https://doi.org/10.3390/s19153400>
11. Bloomfield, P.; Davis, J. Orthogonal rotation of complex principal components. *Int. J. Climatology*. 1994, 4(7), 759–775. <https://doi.org/10.1002/joc.3370140706>
12. The R Project for Statistical Computing: <https://www.r-project.org/>
13. Bröcker, J.; Smith, L.A. Increasing the reliability of Reliability Diagrams. *Weather & Forecasting*. 2007, 22, 651–661. <https://doi.org/10.1175/WAF993.1>
14. Fawcett, T. An introduction to ROC analysis. *Pattern Recog. Lett.* 2006, 27, 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
15. USFS Rocky Mountain Center website: https://fireweather.cira.colostate.edu/index8.php?usr=wfdss_gusr&page=rmc
16. Mandrekar, J.N. Receiver Operating Characteristic curve in diagnostic test assessment. *J. Thoracic Oncol.* 2010, 5(9), 1315–1316. <https://doi.org/10.1097/JTO.0b013e3181ec173d>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.