

UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

LIGHTNING PREDICTION WITH DEEP LEARNING AT VARYING
TEMPORAL AND SPATIAL SCALES

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

by BRANDON T MCCLUNG

Norman, Oklahoma

2026

LIGHTNING PREDICTION WITH DEEP LEARNING AT VARYING
TEMPORAL AND SPATIAL SCALES

A DISSERTATION APPROVED FOR THE
SCHOOL OF METEOROLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Amy McGovern (Chair)

Dr. Edward Mansell

Dr. Kelvin Droegemeier

Dr. Aaron Hill

Dr. David Schwartzman

Dr. Justin Metcalf

©Copyright by BRANDON T MCCLUNG 2026

All Rights Reserved.

Acknowledgments

I acknowledge first the United States Air Force for their financial and career support. The mentorship from Colonel (retired) Brandon Alexander, Ken Ferland (retired), Barb Costa (retired), and Doug Oltmer were crucial in my success of this program. The support from Brian Cizek, Jenny Stewart, and Dr. Kristin Smith in the Cape Canaveral area were instrumental in the success of LaunchCast.

Next, the mentorship, education, and support received from the Idea Lab and AI2ES was crucial to the success of this endeavor. Amy's tenacity helped navigate academic hurdles, and the lab members provided crucial insights into the development of this work. The AI2ES family, and Dr. Andrew Fagg provided the coding and technical infrastructure for model training and GPU support. Finally, the OSCER support team will no longer have to field support tickets from schooner username bmac87!

Within the National Weather Center, I'd like to acknowledge Dr. Michael Stock for his instrumental support in understanding how lightning works and ample time in his garage to build electric field mills. The same goes for the NOAA, CIWRO, and NSSL lightning researchers: Dr. Vanna Chmewliewski, Dr. Sarah Stough, Dr. Kristin Calhoun, Dr. Kevin Thiel, and Dr. Conrad Ziegler.

I thank the Triangle Group for keeping me grounded throughout this work, the pickleball crew for the weekly exercise, and Bella for providing emotional support. I wish she could have been here for graduation.

Finally, I'd like to thank my parents and family based in Seattle. Without your continued support and escapes out of Norman to the mountains, I would not have made it.

Table of Contents

Acknowledgments	iv
List of Tables	viii
List of Figures	xi
Abstract	xix
1 Introduction	1
1.1 Problem Statement	2
1.2 Novel Contributions	4
2 Literature Review	6
2.1 Lightning Overview	6
2.1.1 Physics	6
2.1.2 Impacts	8
2.1.3 Relevance to Dissertation	9
2.2 AI/ML Concepts and Meteorological Analog	10
2.2.1 Convolution Concepts	12
2.2.2 Long Short-Term Memory	15
2.2.3 Optimization and Loss Functions	18
2.2.4 Relevance to Dissertation	23
2.3 Evaluation Metrics and Explainability	23
2.3.1 Relevance to Dissertation	33
2.4 Lightning Location and Protection Instrumentation	34
2.4.1 Ground-Based Observations	34
2.4.2 Space-Based Observations	37
2.4.3 Relevance to Dissertation	38
2.5 Radar Data	38
2.5.1 MRMS Data	38
2.5.2 Horus Radar	40
2.5.3 Specific Differential Phase and Estimations	43
2.5.4 Relevance to Dissertation	45
2.6 Overview of NWP and Lightning	46

2.6.1	Operational NWP Lightning Products	46
2.6.2	NWP Based Lightning Research	48
2.6.3	Relevance to Dissertation	49
2.7	AI-Based Lightning Prediction Tools	49
2.7.1	Relevance to Dissertation	51
3	BoltCast: Medium-range and Synoptic Scale Lightning Prediction .	52
3.1	Data and Methodology	52
3.1.1	GFS Data and GLM Data	52
3.1.2	AI/ML Model Architectures	56
3.1.3	Performance Diagrams and Case Studies	62
3.1.4	Explainable AI	65
3.2	Results	67
3.2.1	Performance Diagrams and Permutation Importance	67
3.2.2	Case Studies and Partial Dependence Plots	76
4	LaunchCast: Nowcasting Lightning at the Meso-scale	85
4.1	Data and Methodology	85
4.1.1	Inputs and Binary Lightning Targets	85
4.1.2	Sample Generation	90
4.1.3	Model Architecture	92
4.1.4	Model Evaluation, Case Studies, and Explainable AI	94
4.2	Results	99
4.2.1	Objective Results	99
4.2.2	Case Studies	104
4.2.3	Explainable AI	107
5	HorusCast: Specific Differential Phase Estimation on Storm Scale .	111
5.1	Data and Methods	112
5.1.1	Horus Radar Data Inputs	112
5.1.2	Target Generation	116
5.1.3	AI/ML Model Architectures	119
5.1.4	Evaluation and Explainability	121
5.2	Results	123
5.2.1	Test Dataset Metrics and Analysis	123
5.2.2	Additional Deployments	130
6	Summary	134
6.1	Summary of Contributions	135
6.2	Model Specific Discussion, Conclusions, and Future Work	136
6.2.1	BoltCast: Medium-range and Synoptic Scale Lightning Prediction	136
6.2.2	LaunchCast: Discussion and Conclusion	139
6.2.3	HorusCast	143

6.3 The End	144
Appendix A BoltCast Supplemental Material	146
A.1 Further Architecture Details	146
Appendix B LaunchCast Supplemental Material	150
B.1 Ablation Statistics Table	150
Appendix C HorusCast Supplemental Material	156
Reference List	161

List of Tables

3.1	An example of the matching of the GFS forecast hours to the GLM observations, for 01 June 2022's GFS model runs.	54
3.2	List of input and output features. The left column describes how the feature is calculated. The right column identifies the specific GFS variable. The last row pertains to the processed GLM data.	56
3.3	Number of images per dataset type. These images generate a 5-Dimensional Tensor with batch size (a subset of the number of samples), the prediction day, latitude, longitude, and GFS variables shown in Figure 3.2. The outputs generate a 4-Dimensional Tensor with batch size, the prediction day, latitude, and longitude.	58
3.4	U-Net and LSTM model training times and the resources for each rotation listed in 3.3. Times are listed as HH:MM:SS. The number of training epochs are also provided in parentheses. Early stopping was implemented monitoring the validation loss, with a min delta of .000001 and a patience of 10 epochs, for a maximum of 500 epochs.	63
3.5	Explicit metrics highlighting the maximum CSI and corresponding threshold, and PR-AUCs for the U-Net and LSTM model, for the specified forecast day. The PR-AUC values match those in the legend in Figure 3.6. Values are calculated across model outputs from the test datasets, from all rotations listed in Table 3.3.	70
4.1	The LaunchCast model input variables and targets.** denotes variable is post-processed. Low Z_{DR} pertains to elevation angles 0.00 - 2.75 degrees at 0.25 degree resolution; mid from 3.00-9.50 degrees at 0.50 degree resolution, and high is from 10.0-18.0 degrees at 1.00 degree resolution. All observational inputs are for the previous hour. For time step examples, see Table 4.3.	87

4.2	The number of samples and lightning flashes in each year, training (2021-2022), validation (2023), and test dataset (2024).	91
4.3	Sample time slices for observational data, sub-hourly HRRR, and MERLIN target valid times in the year 2021. The [t0,t1,t2,t3] time designators are horizontally consistent with the time labels used in the input layers for line-by-line diagram provided in the GitHub repository in the data availability statement.	93
4.4	Statistical summary for each forecast/lead time of the LaunchCast model. (t0-t3) are the keys used in the code, for traceability. These statistics mirror experiment 0 in the statistics provided in the Appendix. 'p' is for a 30-minute GLM persistence forecast. 'h1' is for the HRRR f01 lightning prediction and 'h2' is for the HRRR f02 lightning prediction. The FSS to beat is provided via Eq. 4.8 for climatology, and the '-5' and '-10' designations are for the FSS 5 and 10 nautical-mile windows.	103
5.1	The inputs and target for the HorusCast model. The masks concatenated next to the model input are included, along with the units of each variable.	113
5.2	The deployment times and sample count for each dataset type on the 03 Jun 2025 deployment.	118
5.3	Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method.	127
5.4	Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method for the 19 May 2025 deployment, from 21:01:46 to 21:57:46.	131
5.5	Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method for the 28 April 2024 deployment, from 01:03:38 to 02:44:01.	133
B.1	The maximum CSI thresholds for the ablation experiment.	151
B.2	The maximum CSI values and accompanying values for the ablation experiment.	151
B.3	The AUC-PR for the ablation experiments.	152

B.4	The FSS for a 5-nautical mile window for the ablation experiments. . .	152
B.5	The FSS for a 10 nautical-mile window for the ablation experiments. . .	153
B.6	The AUC-ROC for the ablation experiments.	153
B.7	The probability of detection (POD) for the ablation experiments. . . .	154
B.8	The false alarm ratio for the ablation experiments.	154

List of Figures

2.1	An example of charge separation within a thunderstorm (from Lopez, 2016).	7
2.2	Figure 1 from (Chase et al., 2023). Schematic of (a) a perceptron, (b) a biological neuron adapted from Henley (2021), and (c) a multilayer perceptron.	11
2.3	Equations 2.1 through 2.5.	13
2.4	Figure 3 from (Chase et al., 2023), showing an example convolution on a radar image. The original hook echo radar reflectivity is located in the top-left corner. The convolution at step 0 is shown in the top row while the convolution at step 6 is shown in the bottom row. Note these images are coarsened for visualization purposes (i.e., can see the pixels). An animation of this convolution can be found in the Notebook 6 in the code repository. The result of the full convolution is shown in the top right, blues are negative, reds are positive, and they are unitless.	14
2.5	U-Net architecture taken from Ronneberger et al. (2015).	16
2.6	LSTM gates adapted from Zhang et al. (2023), Figure 10.1.4.	16
2.7	Figure 8 from (Chase et al., 2023). Schematic depicting gradient descent. (top) The blue colored lines are using a smaller learning rate (η) than the (bottom) red colored lines. Arrows depict where loss will be after a gradient step. The θ is the symbol representing the neural network current parameters (i.e., weights and biases).	19
2.8	Binary cross entropy loss plotted when the target is 0 and 1. This is the loss function used in BoltCast and LaunchCast.	21

2.9	Huber loss plotted when $\delta = 0.5$. This is the loss function used in HorusCast.	22
2.10	Adapted from Wilks (2012), this is a confusion matrix. The letters a-d are the same as in Wilks Figure 8.1a, for easy reference for the reader. .	25
2.11	An example precision and recall performance diagram.	26
2.12	An example reliability diagram.	27
2.13	Figure 1 from (Schvartzman et al., 2026). A picture of the Horus truck taken outside of the National Weather Center, Norman, OK.	41
2.14	Figure 2 from (Schvartzman et al., 2026), showcasing the various scanning strategies of the Horus radar.	42
3.1	Spatial climatology for GLM flashes. Each grid cell represents the flash extent density observed between 2021 through 2024.	55
3.2	Visual representation of the GFS-derived inputs and target output of the models. The nine input variables are shown in the first three columns, with example binary outputs on the right. GFS forecast hours are included in the bottom right.	57
3.3	U-Net Architecture diagram. Example inputs and binary outputs are included in the upper-left and upper-right. The layer names and tensor shapes are included for reference. Each 3-D Convolutional Block contains three 3-D convolutional layers. Orange arrows indicate a max pooling operation and the blue arrows indicate a 3-D UpSampling operation. Skip connections are annotated with the teal dashed arrows. Hyper- and total number of trainable-parameters are included in the center dark-gray boxes.	62
3.4	The LSTM model architecture. Example inputs and binary outputs are included in the top and bottom. The layer names and tensor dimensions are included for reference; no blocks with multiple convolutional layers are used. White boxes correspond to 3-D Convolutional Layers and the light-blue correspond to 2-D Convolutional LSTM layers. The dark-gray boxes on the right indicate the hyper- and total trainable number of parameters.	63

3.5	The top row denotes the reliability diagrams for the U-Net (a) and LSTM (b) architectures, for all 5 rotations' test datasets model outputs. Each line pertains to a specific forecast day denoted in the legend, with the Brier Score accompanying the label. The bottom row shows a histogram of the forecasted lightning probabilities for the U-Net (c) and the LSTM (d), with the y-axis on a log scale.	68
3.6	Precision-recall performance diagrams for the U-Net in Figure (a) and LSTM (b) architectures. The x-axis is the precision and the y-axis is the recall. Grey-shaded contours are for critical success index (CSI), and the dashed diagonal lines denote frequency bias. Each line pertains to one prediction day, given model predictions for the test datasets of all 5 rotations listed in Table 3.3. PR-AUC statistics are included in the legend and Table 3.5, where the maximum CSI and corresponding thresholds are also listed, for clarity. Threshold probabilities for Day 1 are included in the red text. Days 2-4 probability thresholds are omitted for clarity.	69
3.7	Plots of PR-AUC across the BoltCast domain. Diagram for the U-Net (panels a,d,g,j), and LSTM (panels b,e,h,k) and the difference (U-Net - LSTM) plotted in panels c,f,i,l. Dashed lines indicate areas where the model outputs were statistically different, with a p-value less than 1.6e-6. Similar plots for recall and precision for a threshold of 0.25 are provided in Figures 3.9 and 3.8.	71
3.8	Map of precision values for a threshold of 0.25 across the model domain.	72
3.9	Map of recall values for a threshold of 0.25 across the model domain. .	73
3.10	Grouped-permutation importance for 1000 permutations for the 0.25 threshold, ranked according to the U-Net median. The x-axis is the decrease (change) in the area under the precision-recall curve, and the y-axis lists the input variables. Box-whisker plots show the variability and medians (box-center) across all rotations and permutations. The red bars pertain to the U-Net while the blue bars pertain to the LSTM model.	74
3.11	PR-AUC bar plots for all 5 rotation's test dataset model outputs, for each season. The U-Net, is in panel (a), and the LSTM, panel (b). Red box bars pertain to Day 1, orange Day 2, light-blue Day 3, and dark-blue Day 4.	77

3.12	Bar graph of seasonal maximum CSI values for the U-Net (top) and LSTM (bottom).	78
3.13	Example model output from rotation 4's test dataset, valid between 1200 UTC 13 June 2022 - 1200 UTC June 14 2022. Panels (a) through (d) [top row] are predictions generated by the U-Net architecture. Panel (e) is the lightning labels. Panels (f) through (i) [middle row] are predictions generated by the LSTM layer. Panels (j) and (k) are generated by a 30 dBZ mask of the predicted reflectivity for Days 3 and 4; panels (l) and (m) are the 48-Hour and 24-Hour SPC HREF calibrated thunder product. The magenta dot in panels (d), (e), and (i) denote the location of Wright-Patterson AFB; used in generating the partial dependence plots in Figure 3.14	80
3.14	Partial-dependence plots for both the U-Net (solid lines) and LSTM (dotted) predictions, for Wright-Patterson AFB for lightning valid between 12 UTC 13 June 2022 and 12 UTC 14 June 2022. Each different color depiction is a different kernel-based window sized, generated on the input image for each Day 1 CAPE (left column), Reflectivity (middle column), and Precipitation Rate (right column). Rug plots along the varying x-axes denote the density of CAPE, Reflectivity, and Precipitation Rate.	81
3.15	Example model output from rotation 4, valid between 1200 UTC 9 December 2022 - 1200 UTC 10 December 2022. Panels (a) through (d) [top row] are predictions generated by the U-Net architecture. Panel (e) is the lightning labels. Panels (f) through (i) [middle row] are predictions generated by the LSTM layer. Panels (j) and (k) are generated by a 30 dBZ mask of the predicted reflectivity for Days 3 and 4; panels (l) and (m) are the 48-Hour and 24-Hour HREF calibrated thunder product. The magenta dot in panels (d), (e), and (i) denote the location of Vance AFB; the location used in generating the partial dependence plots in Figure 3.16.	82
3.16	Same as in Figure 3.14, but for Vance AFB.	83
3.17	Partial dependence plots of the GFS data not presented in Figures 3.14 or 3.16.	84

4.1	Climatology of MERLIN flashes from 2021-2024. The left column is for IC flashes and the right column is for CG flashes. The black boxes in the bottom row denote the model input domain (solid), a 64x64 image on the HRRR grid, and the target domain a 16x16 image on the HRRR grid.	89
4.2	(a) A zoomed in picture of the input domain, the target domain, and the 5-nautical mask required for target generation. Additionally, the location of the EFMs (red x's) and the MERLIN sensors (blue dots) are provided. FM07 is starred for future reference throughout the document. (b) Sample 50 HZ EFM time series 2245 - 2345 UTC 06 Jun 2024 from FM07 (inputs to case 1 in Figure 4.6) example statistics.	90
4.3	The architecture of the model built in Tensorflow. The last images represent the target domain (black box) and the location of Space Launch Complex 40 (SLC-40) for reference, accompanied by a 5 nautical mile range ring. Sample labels are provided (dark grey fill) that correspond to case study number 1 valid from 2245 - 2345 UTC 06 Jun 2024. . . .	95
4.4	Reliability diagram of the LaunchCast model. Panel (a) provides the explicit reliability for the four lead times (dark red - 00 to 15 minutes; orange/peach - 15 to 30 minutes; light blue - 30 to 45 minutes; dark blue - 45 to 60 minutes), with forecast probability on the x-axis and observed relative frequency plotted on the y-axis. Panel (b) shows the forecast histogram for the four lead times. Note the log-scale on the y-axis in panel (b).	101
4.5	Precision-recall performance diagram, highlighting model performance for the four lead times. Gray filled contours denote the CSI values, with precision or success ratio (SR) plotted on the x-axis and recall or probability of detection (POD) plotted on the y-axis. The stand-alone markers outlined in thick black denote the 30-minute GLM persistence forecast (matched to LaunchCast symbols), the pentagons correspond to the HRRR f01, and the hexagons the HRRR f02 lightning product. .	102

4.6	Example model output for 2345 UTC 06 Jun 2024 - 0045 UTC 07 Jun 2024. The grey filled pixels denote the target, while the colored contours denote the predicted probability. In panel (a), the EFM locations are filled with the maximum absolute value used as input between 2245 - 2345 UTC. The location of SLC-40 is provided for reference only. The blank fill in panel (a) is indicative of the masked EFM input values - a novelty of the model architecture. Note, the yellow fill is FM07, which had the inputs generated from the time series presented in Figure 4.2. Panel (a) is the 00-15 minute lead time, (b) is the 15-30 minute lead time, (c) is the 30-45 minute lead time, and (d) is 45-60 minute lead time.	105
4.7	Same as in Figure 4.6 but for 1950 - 2050 UTC 09 Jul 2024. EFM values are from 1850-1950 UTC.	105
4.8	Same as in Figure 4.6 but for 1835 - 1935 UTC 12 Jun 2024, a lightning cessation case. The EFM values are from 1735 - 1835 UTC.	106
4.9	Same as in Figure 4.6 but for 1550 - 1650 UTC 14 May 2024, a case where lightning began in the prediction window. The EFM values are from 1450 - 1550.	106
4.10	Plotted are the changes in AUC-PR for the 15 ablation experiments. Red bars plot changes for 00-15 minute lead times, peach/orange for 15-30 minute lead times; light blue for 30-45 minute lead times, and dark blue for 45-60 minute lead times. The experiment number corresponds to the number listed in Table B.3, in the Appendix. The data types listed below each experiment indicate the non-ablated datasets.	108
4.11	A plot of accumulated local effects for -10 °C reflectivity (a), mid-level Z_{DR} (b), forecast MSLP (c), group area (d), and FM07 mean (e) and max absolute value (f). Note the x- and y-axes vary, with panels (a-d) referencing a specific temporal input time. Panels (e) and (f) display results across the entire input hour due to the LSTM architecture. A larger subset of input variables and times is provided in Appendix B.1	109
5.1	Google Maps rendering of the location of the Horus deployment. This is the deployment location of the radar for all three-cases presented here. Storm Prediction Center storm reports (red: tornado; blue: wind; green: hail; white: Horus location) are included with an overlay of the sensing radius of the radar, 120 kilometers (1562 range gates), for the 03 June 2025 deployment. (Latitude: 35.1873300; Longitude: -97.4442620; Elevation (MSL): 342 meters; Radar Height Above Ground: 3 meters)	114

5.2	Example PPI at 7 degree elevation angle, from the Horus deployment used for training, validation, and testing.(a) is the reflectivity, (b) is the radial velocity, (c) is the spectrum width, (d) is the differential reflectivity, (e) the cross-correlation ratio, and (f) is the differential phase.	115
5.3	Histograms of the training (top row), validation (middle row), and test dataset (bottom row) data for HorusCast model.	117
5.4	HorusCast model architecture. Hyperparameters for the model are included in the upper right, along with the total number of parameters. .	120
5.5	2-D histogram depicting correlation of the predicted K_{DP} values (x-axis) and the true K_{DP} values (y-axis). The fill colors correspond to the number of range gates, on a log-scale. The statistics listed in the upper left corner correspond to equations 5.2-5.5.	124
5.6	Box-whisker diagram for the prediction error. The shaded bars denote the number of range gates indicated on the right-dual axis. The left dual axis pertains to the prediction error. The box plots are binned in .5 deg/km bins, with the bin center indicated on the x-axis.	125
5.7	Reflectivity (a) , HorusCast Output (b), NWS comparison (c), and Vulpiani target (d) for the 03 Jun 2025 deployment, valid at 22:00:07 UTC. The white dots in panel (a) and grey in (b) denote the location of the Oklahoma LMA sources between 22:00:07-22:00:09, inclusive. . .	126
5.8	Input x Gradient (IxG) attribution image of the differential phase input variable, for the 03 Jun 2025 deployment, valid at 22:00:07 UTC. . . .	128
5.9	Input x Gradient (IxG) attribution image of the range input variable, for the 03 Jun 2025 deployment, valid at 22:00:07 UTC.	129
5.10	Same as Figure 5.7, but for 21:19:46 UTC 19 May 2025, with LMA sources valid from 21:19:46 -21:19:48 UTC.	130
5.11	Same as Figure 5.7, but for 02:22:23 UTC 28 Apr 2024, with LMA sources valid from 02:22:23-02:22:25 UTC.	132
A.1	Line-by-line diagram of the encoder for the BoltCast U-Net.	147
A.2	Line-by-line diagram of the decoder for the BoltCast U-Net model. . .	148

A.3	Line-by-line diagram of the BoltCast LSTM model.	149
B.1	ALE plots for the inputs for -10C reflectivity (a-d), mid level Zdr (e-h), group area (i-l), for the different input times. Panels (m-p) showcase ALE for the HRRR MSLP predictions. The bottom row, panels (q-t) provide ALE plots for field mill 22 inputs for the entire previous hour. Scaled histograms are provided to depict a proxy of the distribution of each variable. Each line corresponds to a specific lead time, indicated by the legend in panels (a) and (q).	155
C.1	Part 1 of the HorusCast Encoder.	156
C.2	Part 2 of the HorusCast Encoder.	157
C.3	Part 3 of the HorusCast Encoder.	158
C.4	Line-by-line diagram of the HorusCast Decoder.	159
C.5	A sample ray diagram, plotted similar to one plotted in Mishler et al. (2024). This ray is valid 22:00:07 UTC 03 June 2025, at 7 degree elevation and 175.47 degree azimuth.	160

Abstract

Lightning directly causes human fatalities, sparks costly wildfires, and can be triggered during rocket launches, potentially harming high-value payloads. This threat is increasingly consequential as launch activity at Kennedy Space Center exceeds 100 launches per year. Observing the microphysical processes that trigger lightning is difficult. Recent advances in radar technology have allowed for charged regions of storms to be quantified using specific differential phase (K_{DP}), at temporal scales of seconds with a fully digital phased array radar.

I developed three deep learning models (BoltCast, LaunchCast, and HorusCast), investigating machine learning’s ability to generate lightning probabilities or its related dual-polarization variable, K_{DP} . BoltCast’s domain spans the entire CONUS with a four-day prediction, LaunchCast is near Kennedy Space Center along the east-central Florida coastline with a 1-hour prediction, and HorusCast translates all traditional and dual-polarization variables into a K_{DP} estimate. Each model was developed using custom deep neural networks. The models are evaluated with reliability or attributes diagrams, precision-recall performance diagrams, or error calculations; with varying explainability metrics applied to identify the main drivers of the model.

BoltCast was evaluated against two separate model architectures and showed high reliability out to four days, with area under the precision-recall curve (AUC-PR) values decreasing from 0.62 to 0.52, for the best architecture. Permutation importance showed that CAPE was the primary driver, with partial dependence plots showing larger changes in model output when CAPE and reflectivity are varied for a hit and miss forecast. LaunchCast transitions to a much smaller scale, making lightning predictions out to 1-hour, at 15-minute time steps, with 5 High Resolution Rapid Refresh

variables, a subset of the Multi-Radar Multi-Sensor observation suite, Geostationary Lightning Mapper group area, and electric-field-mill data as inputs. AUC-PR drops from 0.47 to 0.26 across the forecast hour. The accumulated local effect of a subset of input data revealed that the mean electric-field-mill values have influence within 5 nautical miles of the site. Lastly, HorusCast, a K_{DP} estimator, generates fairly strong K_{DP} predictions when the radar is configured to scan continuously. The model performance drops when deployed on test cases that have different scanning strategies. Input x Gradient (IxG) revealed that the differential phase Φ_{DP} had the greatest attribution. This work demonstrates the value of deep-learning’s application for lightning specific predictions and K_{DP} .

Chapter 1

Introduction

Lightning classification and prediction needs vary by customer. From 1979 - 2023 the average lightning mortality rate was 2.07 per 10 million people across the United States, with 54.3 fatalities per year (Marion et al., 2026). Throughout this time period, the average fatalities steadily decreased to 34.2 people per year from 1999-2023. But, the World Meteorological Organization defined lightning as a climate variable in 2018 (Aich et al., 2018; WMO, 2019), indicating a potential increase in lightning frequency with climate change that may offset the reduced mortality rate. Emergency managers desire lightning information several days in advance, to help aid in wildfire mitigation while also ensuring that sound infrastructure is present for personnel safety (Wanless et al., 2025). Launch weather operators, under strict launch commit criteria (Deloach and Pellicciotti, 2023) and increased pressure for increased access to space, need accurate electrification predictions within 10 minutes of launch and desired 30-minute warning lead times (Cizek, 2026; Stewart, 2026). While the mesoscale processes that pose the threat to launches, like sea-breeze convection, are well documented (Laird et al., 1995), recent phased array radar advancements can now sense the electrification of the storm at much finer temporal resolution (Palmer et al., 2023; Schwartzman et al., 2026).

With the continuity of lightning observations available, combined with archives of geo-spatial data at varying temporal scales, machine-learning techniques can be used to develop lightning prediction tools to meet emergency managers' desired multi-day lead times (Wanless et al., 2025), launch weather operators' minute to hourly predictions

(Cizek, 2026), and storm scale evolution of hydrometeor interactions at sub-minute temporal resolution (Goede et al., 2025). The next section discusses some of the current limitations these operators face, and how this dissertation can help close the gap.

1.1 Problem Statement

Forecasters are expected to translate other environmental characteristics, such as convective available potential energy and updraft velocity, to a lightning probability. With the recent launch of Geostationary Lightning Mappers (Goodman, 2012) that monitor lightning from space, lightning is continuously monitored across a large portion of the globe. The Geostationary Operational Environmental Satellite (GOES) series have monitored lightning characteristics over the Conterminous United States (CONUS) since 2019 (Goodman, 2012). These observations, concurrent with the large scale NWP models like the Global Forecast System (GFS; (NCEP, 2015)), provide an opportunity to apply AI/ML techniques to derive a lightning probability across CONUS several days in advance. I have developed a CONUS-wide, convolutional-based model that predicts lightning out to four days called BoltCast, and compared it to current operational products and a long short-term memory based model.

The recent uptick in space launches and the increase in frequency of human space flight has increased the demand for accurate lightning predictions to reduce the risk of lightning impacts to human astronauts and high-value payloads. However, BoltCast’s (McClung et al., 2026) broad spatial predictions on a daily timescale do not meet the temporal resolution requirements, described in the LLCC (Deloach and Pellicciotti, 2023). Kennedy Space Center offers a unique suite of electric field mills that are useful in lightning detection via large jumps in the observed potential gradient (Wilson, 1920; Kobayashi and Kitagawa, 1956). This specific suite of field mills has varied

success for use as inputs to 1-hour AI/ML based lightning prediction tools (Hill, 2018; Speranza, 2019), with some studies suggesting the use of field mills for prediction is completely unwarranted (Skrovan, 2020) and others showing some skill in other countries (Bao et al., 2022). These varying techniques warrant further investigation, with the aim to develop a highly skillful nowcasting model at a temporal relevance of the launch-weather community - accompanied by other geo-spatial tools like NWP data, satellite, and radar. I developed a model at 15-minute and 3-kilometer resolution, that predicts lightning with 1-hour lead time for Kennedy Space Center, and relevant military locations in the region, named LaunchCast.

The surface based field mills provide an integrated snapshot of the electric field aloft, at a specific point from the surface. Phased-array radars have shown promise to capture hydrometeor interactions aloft that generate canted ice-crystals due to their alignment with strong electric fields within a storm (Goede et al., 2025; Biggerstaff et al., 2017; Tzeng et al., 2025). Specific differential phase, or K_{DP} is a dual-polarization variable for which the radar volumes of K_{DP} exhibit strong correlation to total flash rate (e.g., number of flashes per time) (Sharma et al., 2024). While several K_{DP} estimators are in place today (Vulpiani et al., 2012; Schneebeli et al., 2014; Reimel and Kumjian, 2021b), each computes the range-based derivative of the differential phase (Φ_{DP}) slightly differently, with varied results (Reimel and Kumjian, 2021a,b). Thus, given that ML optimization is gradient based, deep-learning techniques are primed for K_{DP} estimation. To understand the electrification at the storm scale, I used Level II radar data (i.e. moment data) from a digital S-band phased-array radar, known as Horus (Schvartzman et al., 2026; Palmer et al., 2023; Palmer and Schvartzman, 2025), to develop a K_{DP} estimator with deep learning techniques. This model, named HorusCast, uses 1-D convolutional layers to estimate K_{DP} from the traditional (reflectivity, velocity, and spectrum-width) and dual-polarization (differential reflectivity, cross-correlation ratio,

and differential phase) radar variables.

1.2 Novel Contributions

My novel contributions to the scientific literature are listed below, organized by the three models I developed: BoltCast, LaunchCast, and HorusCast.

- BoltCast: (McClung et al., 2026)
 - Development of skillful AI/ML that predicts lightning across CONUS to 96 hours. Current operational products stop at 63 hours.
 - A direct comparison between a deep, high-parameter U-Net and a single LSTM-layer based, recurrent neural network.
 - The LSTM-based architecture is the first RNN-based lightning prediction tool across the entire CONUS.
- LaunchCast
 - Development of a skillful sub-hourly lightning prediction tool using re-sampled EFM and other geo-spatial data over Kennedy Space Center, outperforming persistence.
 - 30-second resampled EFM data were generated from the original 50-Hz sampled data. Previous models used minute-averaged data (Bao et al., 2022; Speranza, 2019; Hill, 2018).
 - LaunchCast provided increased operational lightning nowcasts to 15-minute temporal resolution.
 - First-ever investigation quantifying and demonstrating the power of EFM data when compared to other geo-spatial data like NWP, satellite, and radar.

- HorusCast
 - Development of a 1-D convolution-based K_{DP} estimator on a fully digital S-band phased array radar (Palmer et al., 2023; Palmer and Schwartzman, 2025; Schwartzman et al., 2026).
 - First ever documented implementation of K_{DP} estimation with deep learning techniques.

Following this chapter is a literature review, providing the scientific foundation for the subsequent results in the following chapters. Chapters 3 (BoltCast), 4 (LaunchCast), and 5 (HorusCast) all follow a similar approach: a review of the data generated for inputs and targets, followed by a thorough discussion of the ML architecture and rigorous statistical analysis of the ML models (i.e., xAI). The final chapter of this dissertation summarizes the key points from each chapter, and provides a foundation for future work.

Chapter 2

Literature Review

This chapter provides an overview of the fundamental concepts I used to develop the material in this dissertation. It opens with a brief overview of lightning and its impacts followed by applicable AI concepts. I also discuss how the models I developed are evaluated and their input sensitivity quantified. A discussion of lightning location systems used in the dissertation is provided, followed by current NWP and AI-based tools. I close each section with its relevance to the remainder of the dissertation.

2.1 Lightning Overview

2.1.1 Physics

The curiosity regarding lightning’s physics dates back to Benjamin Franklin’s kite and key experiment, and even Greek mythology with lightning being associated with punishment from mythological deities. We know today that lightning occurs in all thunderstorm types (i.e. single-cell thunderstorms, non-severe thunderstorms, severe thunderstorms, and tropical cyclones), presents significant societal impacts, and its physical nature is still not well understood. We do know, however, it requires charge separation in physical storms, not deities in Olympus.

Today, the traditional model of lightning physics revolves around the separation of charge within a storm which generates strong electric fields, as seen in Figure 2.1

(from Lopez, 2016). Here positive/negative charge is indicated with a '+'/'-' sign, with some positive charge present at the top and bottom of the cloud and negative charge in the middle of the cloud. The isothermal levels are shown to highlight the mixed-phase region of the cloud, between 0 °C and -25 °C, where initial electrification and the in-cloud flashes are triggered (Stolzenburg et al., 2015). This mixed-phase region has falling graupel particles exchange charge with ice-crystals during collisions, which drives charge separation - i.e. non-inductive charging. These types of collisions were re-created in the lab to understand this process, within an observable chamber to understand the conditions driving the charge separation (Takahashi, 1978; Saunders et al., 2006). Typically, the in-cloud flash occurs before the cloud-to-ground flashes, which require higher magnitude fields after the onset of inductive charging (Shao and Krehbiel, 1996).

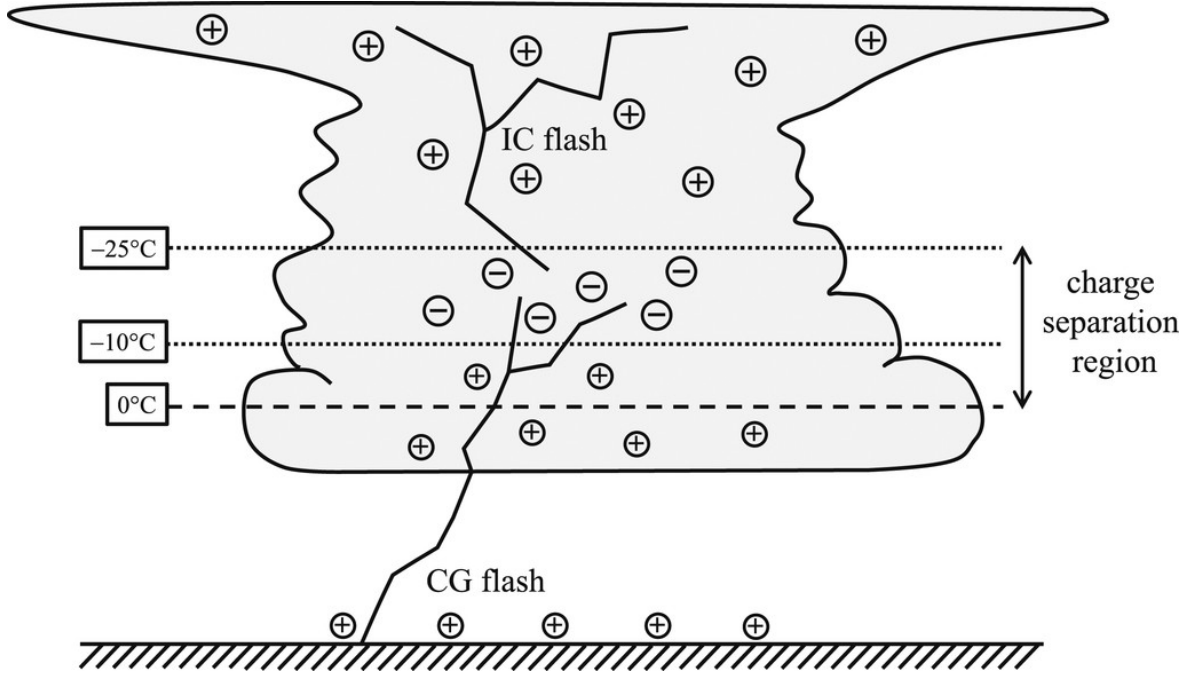


Figure 2.1: An example of charge separation within a thunderstorm (from Lopez, 2016).

Studies have correlated lightning flash properties with other storm and meteorological characteristics. In tropical cyclones and severe storms, lightning jumps are often a

precursor of an intensifying mesocyclone (Stough et al., 2017). Similarly, lightning flash rates are known to increase often prior to severe weather events such as hail and tornadoes (Schultz et al., 2009). Interestingly, the size and energy of flashes are correlated to the size of the storm (i.e. volume of electrified region) with smaller storms producing smaller, lower energy flashes, and larger storms producing bigger more energetic flashes (Bruning and MacGorman, 2013; Bruning et al., 2014). These size limitations play a role in mitigating lightning impacts for the various storm types.

2.1.2 Impacts

Lightning poses a fatal threat to humans. From 1979 - 2023 the average lightning mortality rate was 2.07 per 10 million people in the United States, with an average 54.3 people per year killed (Marion et al., 2026). Largely, these deaths are documented to occur during recreational activities, where people do not have access to proper infrastructure (Holle et al., 2005; Holle, 2016). Throughout this time period, the average fatalities steadily decreased to 34.2 people per year from 1999-2023. But, the World Meteorological Organization (WMO) defined lightning as a climate variable in 2018 (Aich et al., 2018; WMO, 2019), as a result of a potential 22 percent increase in lightning frequency with climate change that could generate more fatalities (Zeng et al., 2008; Roms, 2014). Additionally, lightning-triggered wildfires occur in hard-to-reach locations across the United States (Schultz et al., 2024), causing concern for emergency managers' preparations. The continuing current from the lightning flash is shown to trigger wildfire (Schultz et al., 2019), in high-based thunderstorms (Pérez-Invernón et al., 2021). Studies suggest 92-percent of lightning initiated wildfires were started by negative CG flashes, with 57-percent being single-stroke flashes (Schultz et al., 2019). This may compound the wildfire threat as the climate evolves.

Along the Space Coast in east-central Florida, both natural and triggered lightning

pose a threat to commercial and human space flight, with touchdown on the lunar surface forecast for 2028¹. Additionally, the commercialization of low earth orbit increases launch demand, with now over 100 launches occurring per year at Kennedy Space Center (Smith, 2025). Launch safety continues to be a priority during launch operations, with lightning being a grave concern. Apollo 12 was struck by lightning during liftoff but thankfully did not result in any fatalities. Apollo 12's impacts were followed by Centaur-67 where the vehicle was launched in electrified clouds, resulting in a triggered lightning strike due to the pointed tip of the space vehicle causing an electric corona to occur (Christian et al., 1989; Bazelyan et al., 2007). Thankfully, no loss of life occurred as launch controllers were able to terminate the rocket after launch. While fixed on the ground, the wind-energy sector is beginning to observe a similar triggered-lightning impact when lightning strikes the wind turbines (Steiger et al., 2024).

2.1.3 Relevance to Dissertation

BoltCast was designed to aid those with long desired lead times of lightning impact, primarily in the safety, wildfire, and emergency management sectors. Given the small domain and unique instrumentation, LaunchCast is designed for Kennedy Space Center related operations. Finally, HorusCast merges two state-of-the-art technologies, phased-array radars and AI/ML to help guide future studies of electrification processes and radar modernization efforts.

¹<https://www.nasa.gov/mission/artemis-iv/>

2.2 AI/ML Concepts and Meteorological Analog

AI/ML has a rich history, demonstrating its capability to both improve predictions and develop new understanding of meteorological phenomena (Haupt et al., 2022). This dissertation builds on this history of using AI/ML for meteorological applications, focusing explicitly on lightning prediction. The most relevant application is the use of AI/ML in post-processing NWP output (Haupt et al., 2020a), which both BoltCast and LaunchCast do. Early on, NWP data were translated with Model Output Statistics to provide weather specific stats between predictands and predictors (Glahn and Lowry, 1972). These statistics evolved into expert systems (Moninger et al., 1987) quickly followed by simple neural networks (i.e. deep learning) making principal component analysis calculations (Monahan, 2000). Random forests then showcased the discovery of feature importance for predictands (Breiman, 2001), paving the way for future explainability techniques (Flora et al., 2024).

Traditional ML (i.e., random forests, decision trees, etc.) showed skill predicting severe hazards like wind, heavy rain, and hail (McGovern et al., 2017; Hill et al., 2023; Clark et al., 2025). One limitation of traditional machine learning is that it considers each sample or observation as an independent sample. Meteorologically, the samples or pixels in an image are dependent on the surrounding environment. Deep learning networks help the ML model to learn those dependencies, by generating a network of neurons and perceptrons mirroring brain neurons, shown in Figure 2.2. Panel (a) in Figure 2.2 shows a list of features coincident with the neuron schematic in panel (b). Chase et al. (2023) provides this figure to highlight how learned weighted sums, combined with an activation function (a hyperparameter) over a sample artificial neural network shown in (c) can yield a prediction, $\hat{y}$, (y_{pred} in this dissertation) similar to the neurological chain in (b).

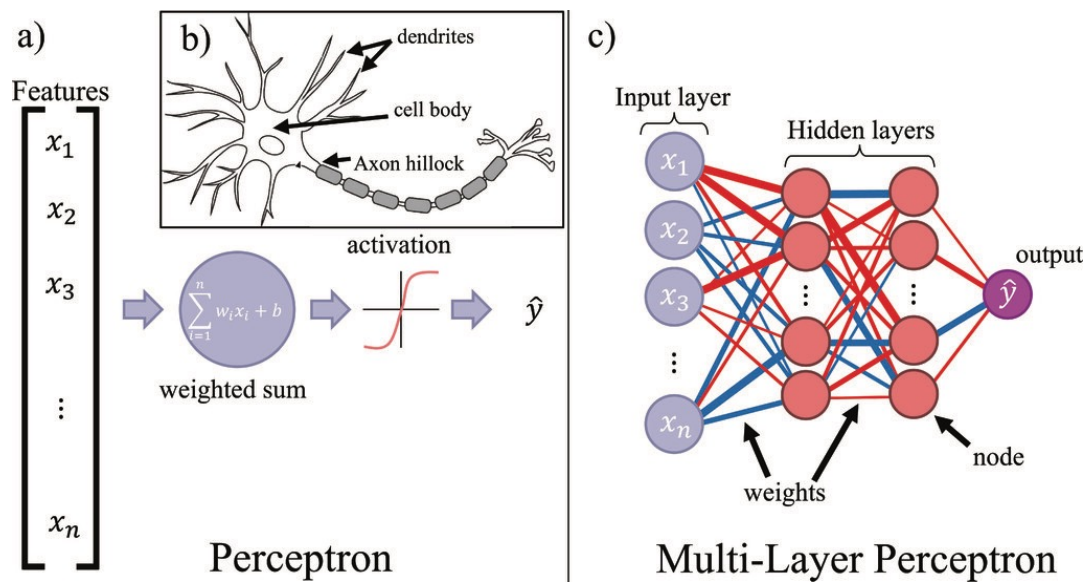


Figure 2.2: Figure 1 from (Chase et al., 2023). Schematic of (a) a perceptron, (b) a biological neuron adapted from Henley (2021), and (c) a multilayer perceptron.

I use the activation functions provided in Equations 2.1 through 2.5, throughout the dissertation. Visual representations of these functions are also provided in Figure 2.3. x in these equations is a function of the learned weights and biases for each neuron, or perceptron, p , in the model ($x = f(wp + b)$).

$$\tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.1)$$

$$\text{ReLU} = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (2.2)$$

$$\sigma = \frac{1}{1 + e^{-x}} \quad (2.3)$$

$$\text{GeLU} = x \frac{1}{2} \left[1 + \text{erf} \left(\frac{x}{\sqrt{2}} \right) \right] \quad (2.4)$$

$$f(x) = x \quad (2.5)$$

2.2.1 Convolution Concepts

Lecun et al. (1998) laid the foundation for using convolutional neural networks to classify handwritten numbers from 2-D images. The convolutional operation computes a single 2D layer of hidden units h_{ij} as from ((Prince, 2023); Eq. 10.6):

$$h_{ij} = a \left[\beta + \sum_{m=1}^c \sum_{n=1}^d w_{mn} p_{i+m-(c-1), j+n-(d-1)} \right] \quad (2.6)$$

where a represents an activation function, β the learned bias, c and d are the dimensions of the kernel dimension. w are the learned kernel weights, and p (often

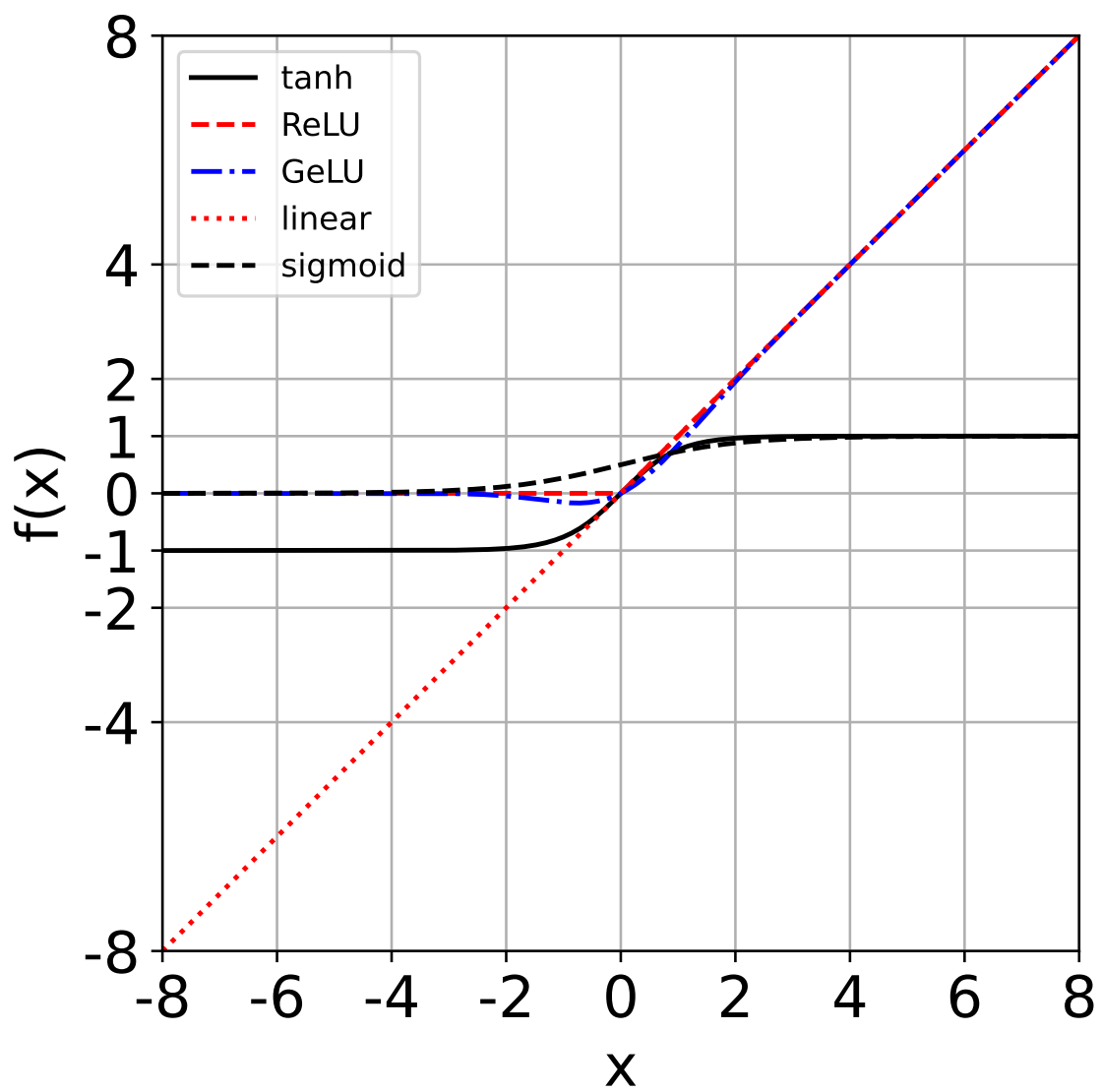


Figure 2.3: Equations 2.1 through 2.5.

labeled as x) corresponds to the gridded input (i.e. GFS, HRRR, GLM group area). This concept is visualized for radar imagery in Figure 2.4, from Chase et al. (2023). The kernel operator translates across the image at a certain stride (i.e., the number of pixels between each 2-D weighted sum) defined as a hyperparameter. This process is repeated such that each pixel in the convolved image has a value.

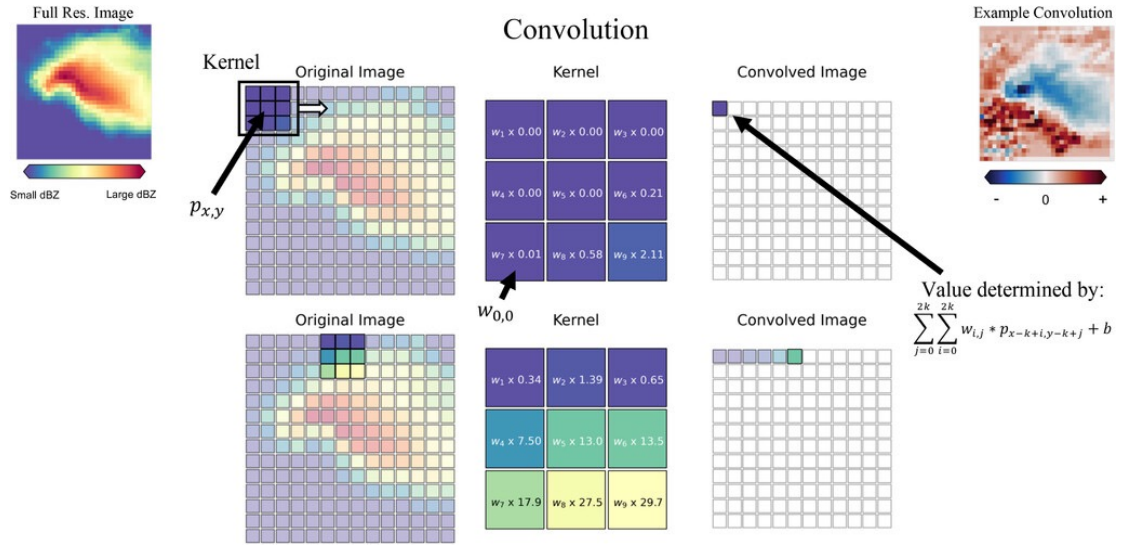


Figure 2.4: Figure 3 from (Chase et al., 2023), showing an example convolution on a radar image. The original hook echo radar reflectivity is located in the top-left corner. The convolution at step 0 is shown in the top row while the convolution at step 6 is shown in the bottom row. Note these images are coarsened for visualization purposes (i.e., can see the pixels). An animation of this convolution can be found in the Notebook 6 in the code repository. The result of the full convolution is shown in the top right, blues are negative, reds are positive, and they are unitless.

A series of these convolution operation or layers can be employed in an architecture

called a U-Net, presented in Figure 2.5 (from Ronneberger et al., 2015). Here each vertical box is a convoluted image, feeding into the next one. The width of the box is correlated to the number of filters or features in each convoluted image, indicated at the top of each box. The 2-D image size is stated as a squared value on the bottom left. The encoder utilizes max pooling operations to halve the image size between each encoder layer, ensuring the maximum value across the pooling kernel is fed into the next layer. At the bottom, a large latent space is provided where non-linear relationships may be learned. The decoder employs up-sampling between each layer to recreate the desired predicted image resolution. (Ronneberger et al., 2015) cropped skip connections carrying information from each encoder layer directly to the decoder. This encoder and decoder structure is the basis for the model architectures I present in Figures 3.3 (McClung et al., 2026) and 4.3, where the 2-D weighted sum, was expanded to a 3-D weighted sum, to incorporate the temporal dimension. Recurrent neural networks are often used to learn weights across time, and are presented in the next section.

2.2.2 Long Short-Term Memory

Initial development of models to learn features across time-series data began with recurrent neural networks. These networks introduce the idea of temporal memory (Hopfield, 1982), that allow the model to learn temporal patterns with backpropagation over time with the introduction of different gates to learn the temporal sequence (Rumelhart et al., 1986). With these architectures plagued by exploding gradients during optimization, the long short-term memory (LSTM) unit was introduced, with a three-gate architecture to help aid in reducing the exploding gradients (Hochreiter and Schmidhuber, 1997). The input, forget, and output gate are all visualized in Figure 2.6, and mathematically described in Equations 2.7, 2.8, and 2.9.

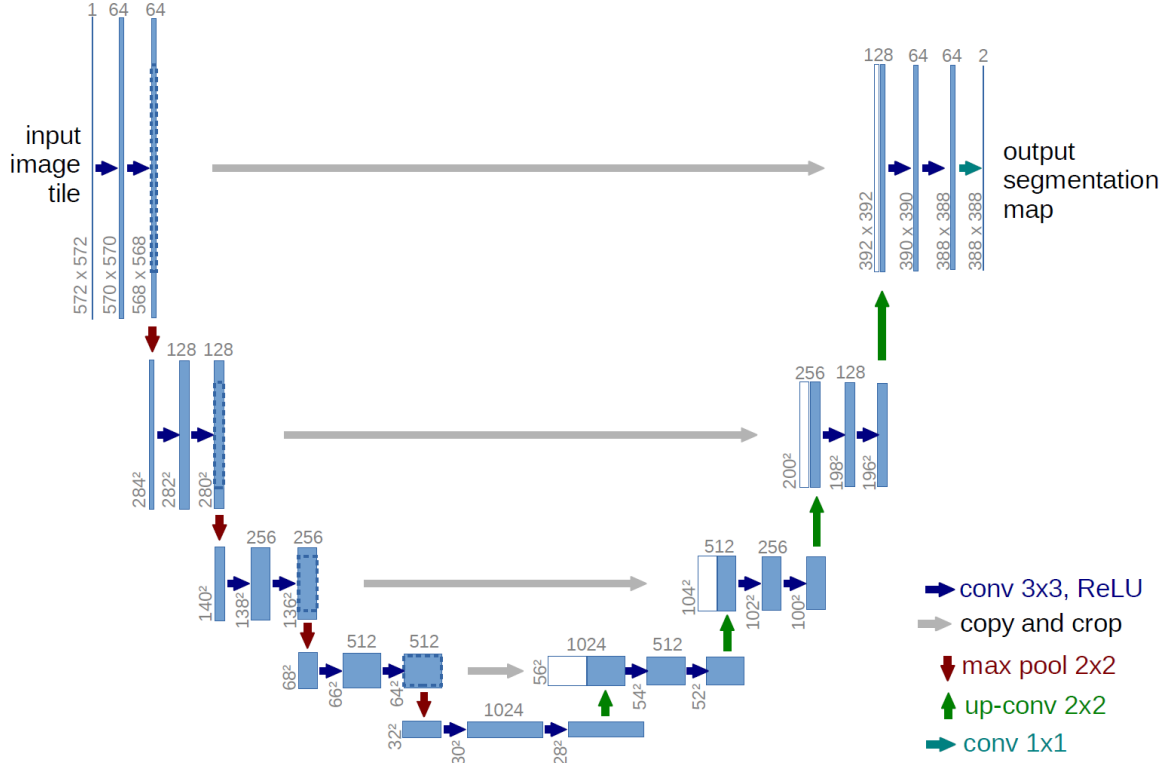


Figure 2.5: U-Net architecture taken from Ronneberger et al. (2015).

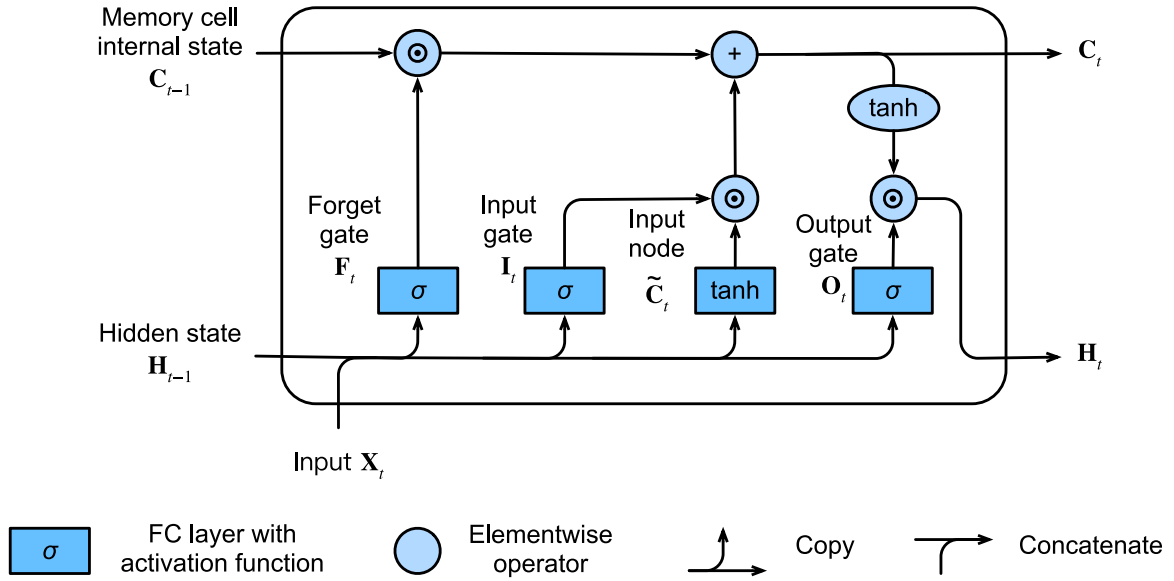


Figure 2.6: LSTM gates adapted from Zhang et al. (2023), Figure 10.1.4.

$$\mathbf{I}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xi} + \mathbf{H}_{t-1} \mathbf{W}_{hi} + \mathbf{b}_i) \quad (2.7)$$

$$\mathbf{F}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xf} + \mathbf{H}_{t-1} \mathbf{W}_{hf} + \mathbf{b}_f) \quad (2.8)$$

$$\mathbf{O}_t = \sigma(\mathbf{X}_t \mathbf{W}_{xo} + \mathbf{H}_{t-1} \mathbf{W}_{ho} + \mathbf{b}_o) \quad (2.9)$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{X}_t \mathbf{W}_{xc} + \mathbf{H}_{t-1} \mathbf{W}_{hc} + \mathbf{b}_c) \quad (2.10)$$

$$\mathbf{C}_t = \mathbf{F}_t \odot \mathbf{C}_{t-1} + \mathbf{I}_t \odot \tilde{\mathbf{C}}_t \quad (2.11)$$

$$\mathbf{H}_t = \mathbf{O}_t \odot \tanh(\mathbf{C}_t) \quad (2.12)$$

Here $\mathbf{X}$ pertains to the input sequence, $\mathbf{W}$ the weights learned to either the input i , output o , or forget f , gates for a specific time, t . The $\odot$ symbol in Equations 2.11 and 2.12 represent element-wise matrix multiplication. Figure 2.6 suggests a series of summations and multiplications weighted by the sigmoid function ranging from 0 to 1, or the tanh function -1 to 1. The forget gate determines how much of the input to forget then influences the memory cell. This memory is then propagated to the input gate and node, where the new input data is then added to the memory cell state. This updated memory state is then output, followed by a product of the output gate, to generate the state at the next time step.

Shi et al. (2015) replaces the matrix multiplication in equations 2.8 to 2.12 with a convolution operation to allow LSTM's to accommodate time series data that are 2D

and 3D, like radar data or NWP output.

2.2.3 Optimization and Loss Functions

The weights and biases, w and b in Figures 2.4 and 2.2, and $\mathbf{W}$ and $\mathbf{b}$ in Equations 2.8 to 2.12 are learned via gradient descent, or optimization. This concept is presented in Equations 2.13 through 2.15, adapted from Prince (2023). Here, ϕ is the set of weights or parameters, and $\hat{\phi}$ is the set of parameters such that the loss function, L is minimized.

$$\hat{\phi} = \operatorname{argmin} [L[\phi]] \quad (2.13)$$

To find the minimum of the loss function, the derivative of the loss function is computed with respect to each parameter:

$$\frac{\partial L}{\partial \phi} = \begin{bmatrix} \frac{\partial L}{\partial \phi_0} \\ \frac{\partial L}{\partial \phi_1} \\ \vdots \\ \frac{\partial L}{\partial \phi_n} \end{bmatrix} \quad (2.14)$$

with n being the total number of parameters in the model. ϕ gets updated using:

$$\phi \leftarrow \phi - \alpha \cdot \frac{\partial L}{\partial \phi} \quad (2.15)$$

In Equation 2.15, α is annotated as the learning rate (here, $\alpha = \eta$ in Figure 2.7). Chase et al. (2023)'s visualization of the learning rate showcases how selecting the learning rate matters how aggressively the learned model parameters achieve the minimum parabolic loss function, in this case.

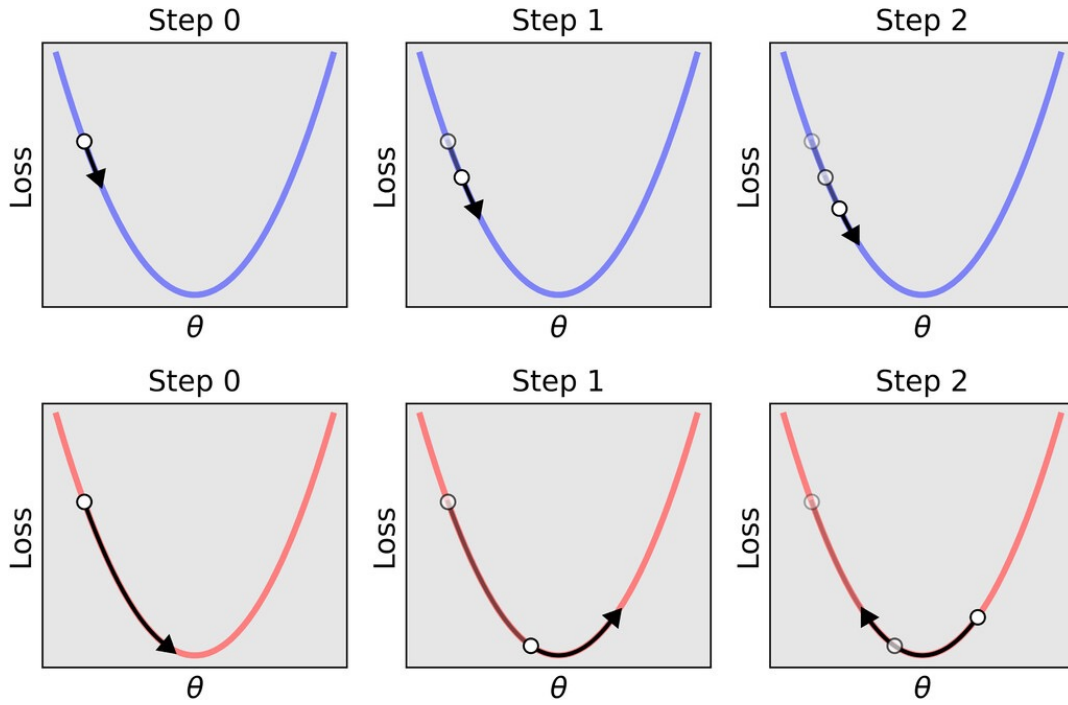


Figure 2.7: Figure 8 from (Chase et al., 2023). Schematic depicting gradient descent. (top) The blue colored lines are using a smaller learning rate (η) than the (bottom) red colored lines. Arrows depict where loss will be after a gradient step. The θ is the symbol representing the neural network current parameters (i.e., weights and biases).

I used the binary cross entropy loss function from Equation 2.16 in the BoltCast and LaunchCast models, to derive lightning probabilities y_{pred} from binary lightning targets y_{true} . This loss function is particularly useful for binary classification, due to its behavior shown in Figure 2.8. When the $y_{true} = 1$ and the predicted probability is near 0, the loss becomes large, indicating the model will be penalized severely. Same for when $y_{true} = 0$ and the y_{pred} is near 1. N in equation 2.16 refers to the total number of pixels on the prediction grid. Thus, during training, the loss is averaged over the entire prediction domain.

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_{true_i} \log(y_{pred_i}) + (1 - y_{true_i}) \log(1 - y_{pred_i})] \quad (2.16)$$

HorusCast is a regression problem, predicting the true value of K_{DP} . Thus an error, e in 2.17, based loss function is required. The huber loss function linearly penalizes the model when the error grows beyond a specified value (.5 deg/km) while squaring smaller error values, penalizing the model less.

$$\epsilon = y_{true} - y_{pred} \quad (2.17)$$

$$L_{\delta}(\epsilon) = \begin{cases} \frac{1}{2}\epsilon^2, & |\epsilon| \leq \delta, \\ \delta \left(|\epsilon| - \frac{1}{2}\delta \right), & |\epsilon| > \delta. \end{cases} \quad (2.18)$$

$$\mathcal{L}_{\text{Huber}} = \frac{1}{N} \sum_{i=1}^N L_{\delta}(y_{true}, y_{pred}). \quad (2.19)$$

I determined the loss function, learning rate, activation function, and number of steps or epochs by conducting a hyper-parameter search. Throughout this document, I refer to hyper-parameters as parameters that are fixed by me, and not learned by the

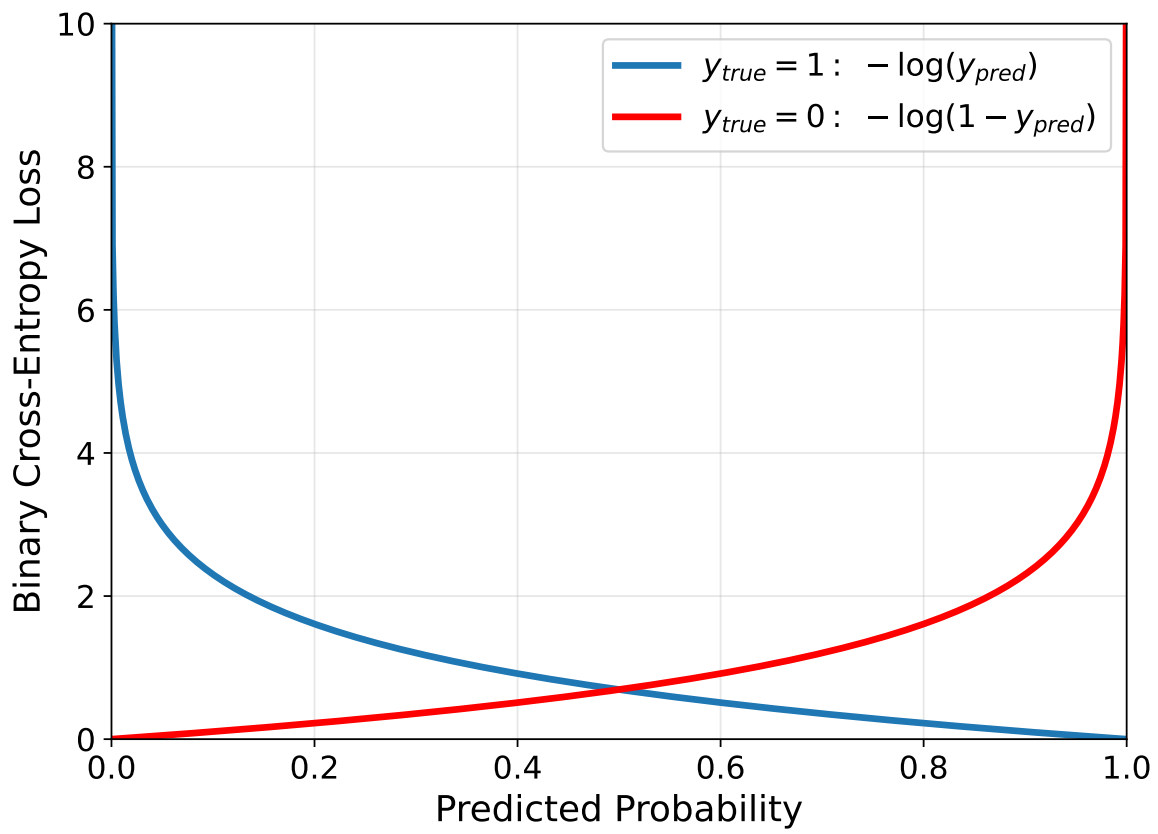


Figure 2.8: Binary cross entropy loss plotted when the target is 0 and 1. This is the loss function used in BoltCast and LaunchCast.

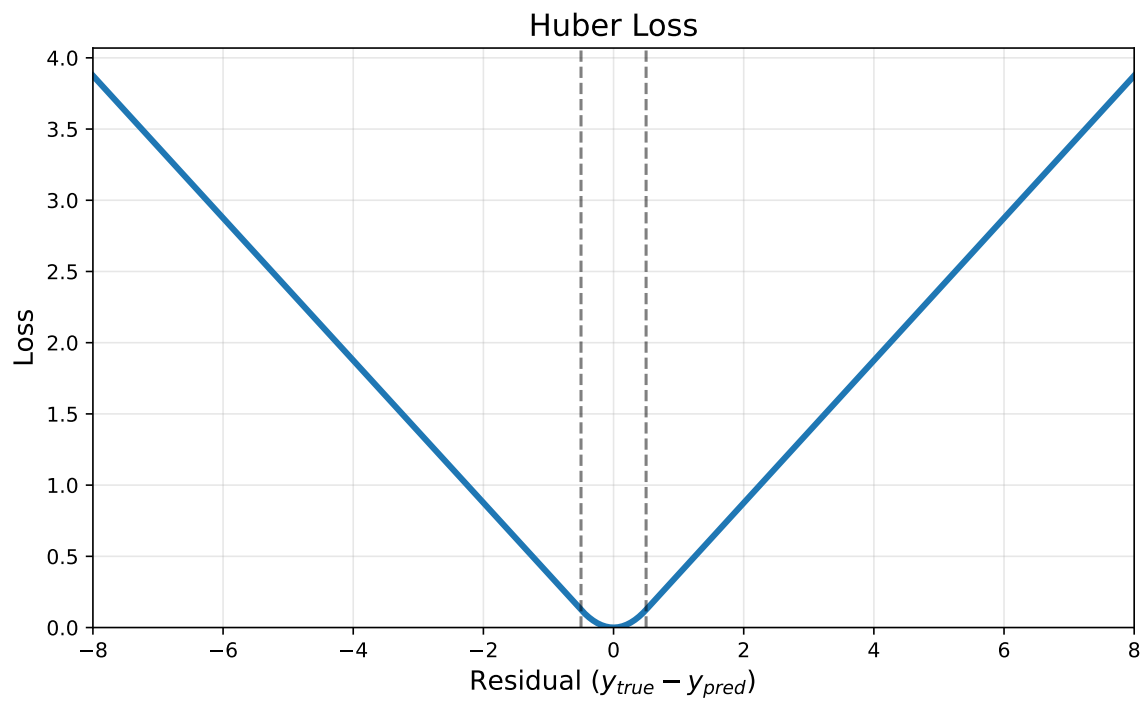


Figure 2.9: Huber loss plotted when $\delta = 0.5$. This is the loss function used in Horus-Cast.

model.

2.2.4 Relevance to Dissertation

In the following chapters, each chapter uses a subset of the tools provided above. BoltCast compares both a U-Net and LSTM model architecture, with relu and sigmoid activation functions. LaunchCast implemented a similar encoder/decoder architecture, with gelu and sigmoid activation functions. Both BoltCast and LaunchCast used the binary cross entropy loss function. The HorusCast model used the Huber loss, with the gelu and linear activation functions.

2.3 Evaluation Metrics and Explainability

The foundation for the evaluation metrics used in this dissertation is the confusion matrix, consisting of:

- True Positive (TP; a hit) when the model predicted the event correctly;
- True Negative (TN) when the model correctly predicted no event
- False Positive (FP) when the model predicted the event, but the event didn't occur
- False Negative (FN; a miss) when the model did not predict the event, but the event did occur

also shown in Figure 2.10, adapted from Wilks (2012). The confusion matrix is the foundation for calculating the precision (Eq. 2.20), recall (Eq. 2.21), and critical success index (CSI; Eq. 2.22). Precision, recall, and CSI are all maximized at 1 (strong model performance), and minimized at 0 (poor model performance). Switching the *FP*

in the denominator of precision for the FN in recall introduces a trade-off. A highly precise forecast can have low recall if the model only predicts lightning a small fraction of the time (low FN s). Whereas recall can be high, while the model over predicts the frequency of lightning (high FP s). The CSI accounts for this trade-off, by incorporating the total number of occasions on which that event was forecast and/or observed in the denominator (Wilks, 2012). Precision is also referred to as success ratio, and recall as probability of detection. For a model that generates probabilistic output, these metrics are calculated at thresholded probabilities then reported in Chapters 3 and 4.

$$Precision = \frac{TP}{TP + FP} \quad (2.20)$$

$$Recall = \frac{TP}{TP + FN} \quad (2.21)$$

$$CSI = \frac{TP}{TP + FN + FP} \quad (2.22)$$

A precision-recall performance diagram can be used to help visualize the tradeoff between the two metrics. An example is shown in Figure 2.11, adapted from Roebber (2009). A high-skilled model will have all values in the upper right corner of the plot, where a low-skilled model has values in the lower left corner of the plot. The area under the curve can be calculated to determine a skill across all thresholds, with a value of 1 indicating the maximum performance. This metric is documented as AUC-PR throughout the remainder of this document, pertaining to the area under the

		<u>Observed</u>	
		<u>Yes</u>	<u>No</u>
<u>Forecast</u>	<u>Yes</u>	True Positive (TP; a) "Hit"	False Positive (FP; b) "False Alarm"
	<u>No</u>	False Negative (FN; c) "Miss"	True Negative (TN; d) "Correct Negative"

Figure 2.10: Adapted from Wilks (2012), this is a confusion matrix. The letters a-d are the same as in Wilks Figure 8.1a, for easy reference for the reader.

precision recall curve.

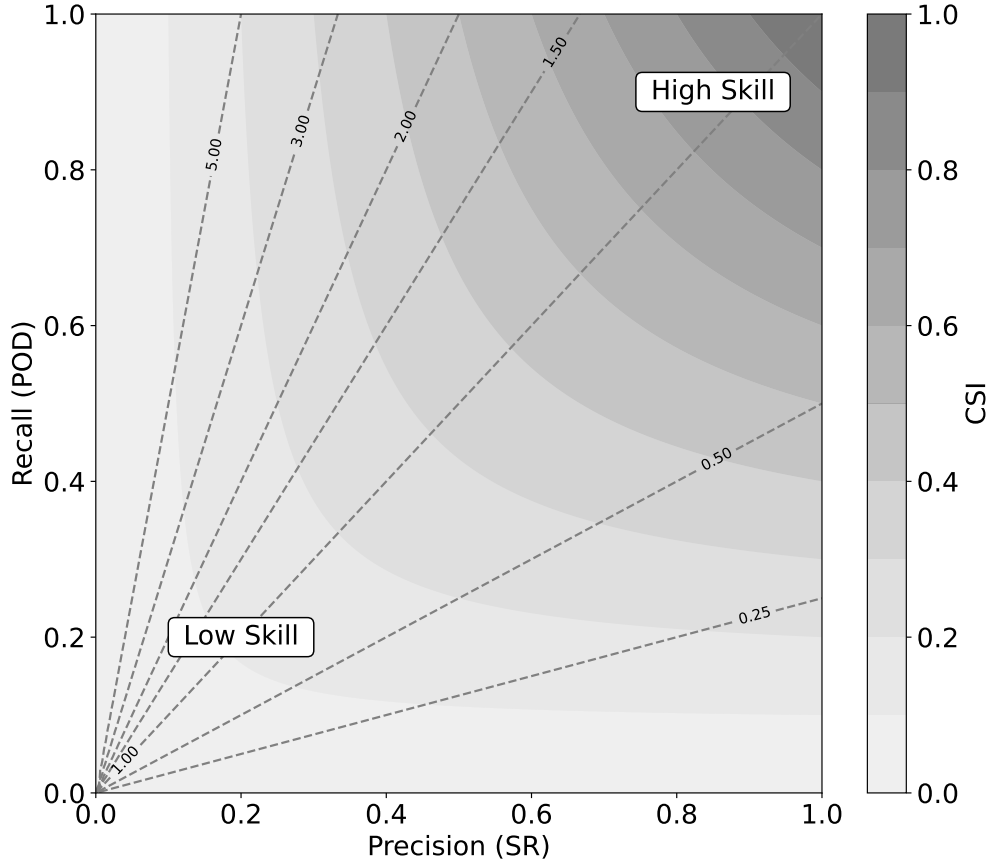


Figure 2.11: An example precision and recall performance diagram.

An additional way to evaluate probabilistic output is via an attributes and reliability diagram, shown in Figure 2.12. This diagram largely shows the correlation of the predicted probabilities when compared to observed frequency (Hsu and Murphy, 1986). To do so, the predictions are assigned a forecasted probability bin from 0 to 1, in increments of 0.05. Then the mean predicted probability is calculated of all predictions within each bin. That generates the values on the x-axis. The y-axis values are calculated by taking the mean of the targets in each forecasted probability bin. It is

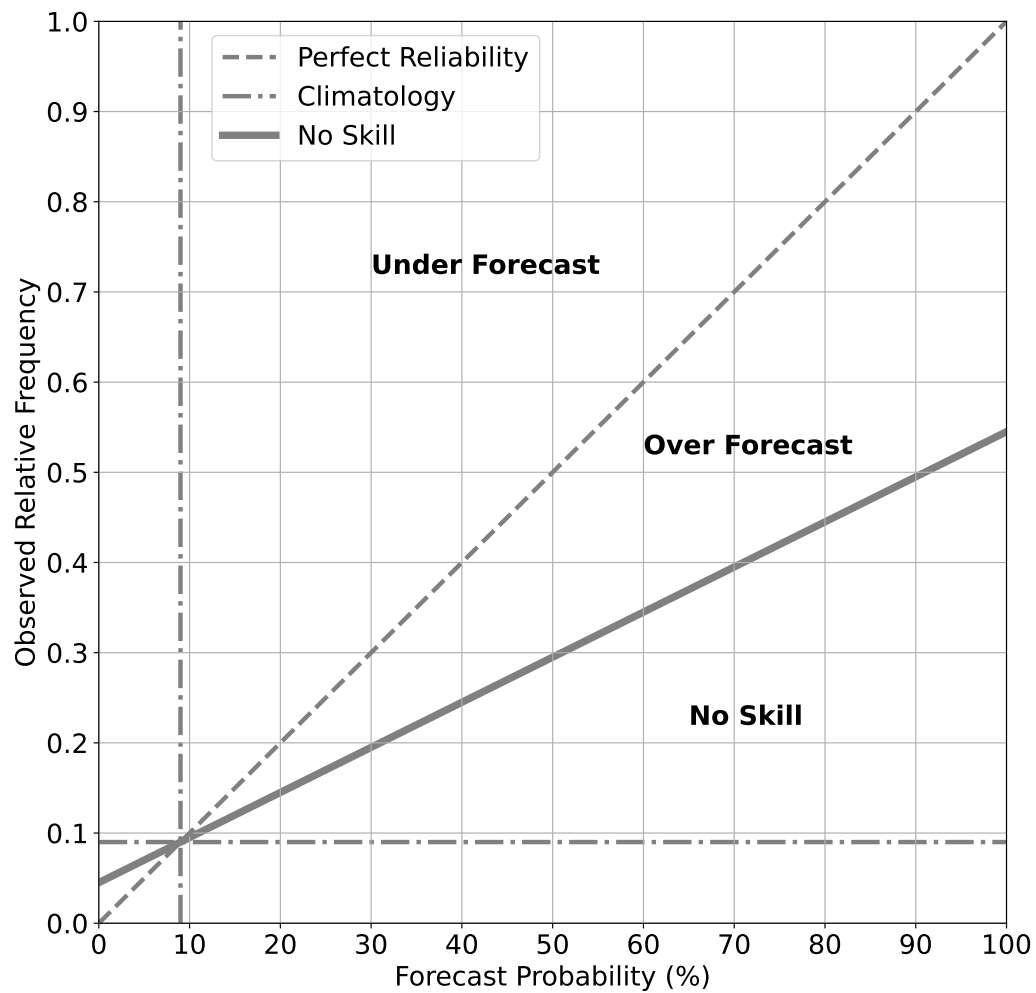


Figure 2.12: An example reliability diagram.

often compared to climatology (dashed-dotted line), a perfectly reliable (dashed line), and no-skill (solid line) forecast to visualize which forecast probabilities contribute skill. An ideal, perfectly reliable forecast would follow the dashed line. This diagram also informs whether the model over forecasts (under the perfect reliability line) or under forecasts (over the perfect reliability), as annotated in Figure 2.12. I also present the brier score, or mean squared error, shown in Equation 2.23 in legends of the reliability diagrams presented in Chapters 3 and 4.

$$BS = \frac{\sum_{j=1}^N (y_{pred} - y_{true})_j^2}{N} \quad (2.23)$$

$$FBS = \frac{\sum_{k=1}^M (f_k - o_k)^2}{M} \quad (2.24)$$

$$FBS_{ref} = \frac{\sum_{k=1}^M (f_k^2) + \sum_{k=1}^M (o_k)^2}{M} \quad (2.25)$$

$$FSS = 1 - \frac{FBS}{FBS_{ref}} \quad (2.26)$$

$$FSS_{baseline} = .5 + \frac{f0}{2} \quad (2.27)$$

One drawback of the previous evaluation approaches is that they do not account for near misses and hits, or the double penalty in point-wise scoring. For this, I use the fractions skill score (FSS) which calculates the skill based on the fractions of hits or misses within a specified spatial neighborhood. Here, M in Equations 2.24 and 2.25 indicate the number of pixels within the spatial window. The reference fractions brier score, FBS_{ref} , pertains to the worst case error that could be encountered in the window, and the fractions brier score, FBS , pertains to the mean squared error within the window. Equation 2.26 is the final formulation, as presented in Lean and Roberts (2008), where the ideal value is one. FSS is often reported in reference to a climatological baseline, indicated in Equation 2.27, (Lean and Roberts, 2008), where f_0 is the climatological frequency. This is also called the FSS to beat.

Unfortunately, these performance metrics are not enough to determine which inputs drive the ML model to make its prediction. ML models are often critiqued as a black box, without the ability to understand or explain model inputs' impact on performance or output (McGovern et al., 2019). I chose to vary the methods across the three models to show diversity in the application of explainable AI techniques. Each technique presented here is model-agnostic, and could be applied to any other model since these techniques focus on perturbing the model inputs and measuring the impact on either model performance or output. While there is debate whether or not to use the training or the test dataset for explainability (Flora et al., 2024), I use the test dataset to understand the drivers of the model on inputs that mimic operational data feeds. The techniques used throughout this dissertation with a brief description are provided below:

- Single variable permutation importance randomizes each input along the sample dimension, then measures the change in model performance (i.e., AUC-PR) for a set number of randomizations. Feature importance is ranked based upon the

randomized input that drove the greatest change in model performance. This is represented in Equation 2.28.

$$PI = \frac{1}{N} \sum_{p=1}^N (AUC_{Reference} - AUC_{p,j}) \quad (2.28)$$

Here the $AUC_{Reference}$ is the un-permuted AUC-PR. $AUC_{p,j}$ is the AUC when input feature j is shuffled along the sample dimension, for permutation number p . Once N (1000 for BoltCast) permutations are completed for each feature j , the mean (or median) permutation importance (PI) can be calculated, and the spread across N permutations are presented as box-whisker plots in Chapter 3. Inputs with positive, high magnitude changes in model performance, indicate that specific input contributes to the performance. Low amplitude positive changes indicate that input does not influence performance. Values can be negative, indicating the permutations increased model performance, but were not observed in this work.

- Partial dependence plots monitor the change in the model output (i.e., lightning probability), when the model inputs are changed based on the values within test dataset distribution (Friedman, 2001; Trok et al., 2023). The partial dependence function described in Molnar (2022) is:

$$\hat{f}_s = E_{X_c} \left[\hat{f}(x_s, X_c) \right] = \frac{1}{n} \sum_{i=1}^n \hat{f}(x_s, x_c^{(i)}) \quad (2.29)$$

Partial dependence (PD) examines the change in the model output for the features in a specific set s . For BoltCast, this refers to a specific NWP input variable, like CAPE, which is varied across the entire distribution of the dataset (Molnar, 2022). Here, X_c denotes the entire input feature set, and $\hat{f}$ denotes the trained

model. $x_C^{(i)}$ are the actual feature values from the dataset (Molnar, 2022). The PD value results should remain within the physical realm of model outputs (i.e. 0 to 100 percent probability). Values outside of that realm indicate model architecture issues. Partial dependence plots showcase the distribution density of the test dataset beneath the corresponding model output, allowing interpretability for the frequency of the value as well. One key assumption is that the features in set s are not correlated, which is a drawback for using this approach for geospatial datasets. One advantage is that the PD function is easy to implement and has a causal interpretation (Molnar, 2022). Meaning, for the specific features and the model only, the interpretation is causal since the outcome, or lightning probability, is modeled for that specific task. The causal nature does not transfer to the real world (Molnar, 2022).

- Ablation studies showcase the change in model performance, when input data are not included into the model (Meyes et al., 2019). This is similar to a data denial experiment in data assimilation, used to determine relative impact of observations. This technique is chosen to mimic operational data outages, that occur at operational units. In Chapter 4, ablated values are determined based on the input variable to avoid ablating to values that have physical meaning.
- Accumulated local effect (ALE) method calculates the differences in predictions, where each feature, j is replaced with upper (z_k) and lower (z_{k-1}) bounds of a set number of quantiles, k (Molnar, 2022; Apley and Zhu, 2020). This concept is presented in Equation 2.30 and 2.31.

$$\hat{f}_{j,ALE}(x) = \sum_{k=1}^{k_j(x)} \frac{1}{n_j(k)} \sum_{i: x_j^{(i)} \in N_j(k)} \left[\hat{f}(z_{k,j}, x_j^{(i)}) - \hat{f}(z_{k-1,j}, x_j^{(i)}) \right] \quad (2.30)$$

$$\hat{f}_{j,ALE} = \hat{f}_{j,ALE}(x) - \frac{1}{n} \sum_{i=1}^n \hat{f}_{j,ALE}(x_j^{(i)}) \quad (2.31)$$

The difference in prediction is the effect the feature has for an instance in a certain interval (Molnar, 2022). The right-hand sum in Equation 2.30 adds up the effects of all instances within an interval, which appears across the entirety of the model domain (Molnar, 2022). Then, divide by the number of instances in this interval to get the average difference of the predictions for this interval - referencing local in the title (Molnar, 2022). The left hand sum in Equation 2.30 references accumulated in the title, meaning the un-centered ALE for the fourth interval is the sum of the effects of intervals one through four (Molnar, 2022). The mean centered ALE is provided in 2.31, where the average of the un-centered ALE, across n intervals is subtracted such that ALE is reported in reference to the average prediction.

While the implementation of ALE is more complex than PD, its computational cost is far less, assuming the number of intervals chosen for a feature is less than the number of unique feature values. ALE is also preferred for correlated datasets. One dis-advantage is if the intervals of the features are not chosen correctly, then small scale effects of the model cannot be captured. Thus, quantile-based intervals are recommended and implemented in LaunchCast.

- Input Times Gradient (IxG) provides a first-order, linear approximation of the attribution of the model output to its input (Ancona et al., 2017; Mamalakis et al., 2022).

$$IxG = x_f \frac{\partial f}{\partial x_f} \quad (2.32)$$

This method simply multiplies the original input value times the derivative of the model output with respect to that variable. IxG accounts for the sensitivity of the model to small changes in input variables, while also accounting for the amplitude of the input feature itself. Thus if a feature with a small amplitude has high sensitivity, its attribution is quantified similarly to a high amplitude feature with low sensitivity. This method is the only method that requires backpropagation through the model. All other methods used herein, utilize forward passes of the model.

2.3.1 Relevance to Dissertation

BoltCast and LaunchCast both utilize reliability diagrams, AUC-PR diagrams, and CSI to evaluate the model. LaunchCast provides additional metrics with FSS, AUC-ROC, and False Alarm Ratio for transparency to launch weather operators. Since HorusCast is a regression model, the main evaluation metrics are root mean squared error, mean absolute error, bias, correlation coefficient, and spearman ranked correlation coefficient. All models are accompanied by example output, or case studies and explainability metrics. BoltCast implemented permutation importance and partial dependence plots. LaunchCast used ablation and accumulated local effect, concluding with HorusCast's use of input times gradient to depict the attribution of traditional and dual-polarization variables. All explainability metrics are documented for the test dataset, mimicking an operationally deployed model, where the data have been unseen during training.

2.4 Lightning Location and Protection Instrumentation

This section provides an overview of the ways lightning has been observed, both from ground and space, with a focus on how specific instrumentation was used throughout the dissertation.

2.4.1 Ground-Based Observations

2.4.1.1 Lightning Location

Lightning location systems vary based on the lightning feature in which they are tuned to observe (e.g. breakdown from channel propagation versus return strokes), or radio waves to listen for. In the extreme low frequency (ELF) the electrified ionosphere helps provide a waveguide for the radio waves emitted by lightning to be observed in other hemispheres. This provides the basis for instrumentation like the Worldwide Lightning Location Network (WWLLN) (Lay et al., 2004) and Vaisala’s Global Lightning Detection 360 system (GLD360²). These worldwide lightning detection systems are often plagued with low detection efficiency (due to the large distance between receivers) when compared to other shorter-range detection systems, like the National Lightning Detection Network (NLDN) or Earth-Networks Total Lightning Network, where density of stations over land is much greater. These systems are a bit higher frequency in the (very-low frequency; VLF) range and have significantly better detection efficiency when compared to WWLLN. These data provide a specific location of a flash, often accompanied by a power or polarity. These data are worth mentioning, as they can aid in generating global lightning climatologies and climate related studies (Wang et al.,

²<https://www.vaisala.com/en/products/systems/lightning/gld360>

2023, 2024). However, these global datasets were not used in this dissertation.

Increasing the listening frequency band into the very high frequency (VHF) range, lightning sources and charge can be inferred from lightning mapping arrays (Rison et al., 1999; Koshak et al., 2004). These lightning mapping arrays are in place across Oklahoma (Chmielewski et al., 2022), Northern Alabama/New Mexico, Washington DC (Quick et al., 2020), and Colorado³. They provide lightning locations in Euclidean, 3D space, but fail to recognize whether lightning flashes reach the ground due to the curvature of the Earth, time difference calculation of sources in post-processing, and its inability to detect return strokes. Recently, the Oklahoma LMA sources have been overlaid on radar data to investigate radar echoes impacted by lightning flashes (Goede et al., 2025).

KSC recognized VHF’s limitations to sense cloud-to-ground strikes on their original lightning detection and ranging system (Boccippio et al., 2001) and installed a separate system to detect cloud-to-ground flashes called the 4D lightning surveillance system (Roeder and Saul, 2012, 4DLSS). This system has been used to study lightning’s relation with dual-polarization variables at KSC (Travis, 2015; Thurmond, 2014; Woodard et al., 2012) and attempted to optimize lightning warning radii at KSC and its surrounding military complexes (Gutman et al., 2019). But, throughout these systems’ life cycles detection efficiency (defined as the number of flashes observed compared to the number of flashes that occurred) concerns still warranted a pursuit of higher detection efficiency with more spatial accuracy.

The Mesoscale Eastern Range Lightning Information Network (MERLIN) was installed in 2017 to help address the detection efficiency issue. MERLIN comprises of 10 Vaisala TLS200 sensors, that detect in-cloud and cloud-to-ground sources on the same system (Roeder, 2017). This system has demonstrated a 10-percent increase in

³<http://lightning.nmt.edu/colma/>

detection efficiency, when compared to 4DLSS (Roeder, 2017). MERLIN also increased location accuracy to below 58-meters, as compared to the 350-meter accuracy of the 4DLSS. Holland (2021) used the new location-accuracy of MERLIN to suggest a 1-nautical mile reduction in the warning radii. I use these MERLIN data to generate the lightning targets for LaunchCast; the first time they are used in a ML application.

2.4.1.2 Electric Field Mill Observations

While electric field mills (EFMs) do not explicitly locate lightning strikes they do capture the abrupt discharge of the lightning strike and the associated change in electric field (Wilson, 1920; Kobayashi and Kitagawa, 1956). The instrumentation is set up such that a positive value observed on the instrumentation coincides with a net positive charge aloft; a negative value coincides with a net negative charge aloft (Koshak, 1990). The EFM measures the negative vertical electric field vector, provided the horizontal components of the electric field are zero at the surface.

The unique nature of these observations have limited operational uses, except at Kennedy Space Center where the lightning protection and lightning warning system (LPLWS) installed 31 static EFMs in support of space launch observations (Jacobson and Krider, 1976). In a research capacity, balloon-borne EFMs have been launched to correlate electric field values aloft to radar observations and understand the charge structures of storms overhead (Bruning et al., 2007; MacGorman et al., 2008).

The LPLWS at Kennedy Space Center informs the Lightning Launch Commit Criteria used to prevent triggered lightning (Deloach and Pellicciotti, 2023). Specifically regarding the EFM suite, Rule 7 requires there to be observations less than 1000 V/m within 5-nautical miles of the path of launch for at least 15 minutes (Deloach and Pellicciotti, 2023), such that a triggered strike does not occur. While these observations help launch satellites into space, they do not help sense lightning from space as the

observations discussed in the next section.

2.4.2 Space-Based Observations

Lightning observations from space began with the polar orbiting lightning imaging sensor (LIS) on the Total Rainfall Measurements and Monitoring satellite mission (Christian et al., 1998). These initial space-based observations provided strong correlations between observed precipitation and lightning flash rates on a global scale. Additionally, they provided a basis for a global lightning climatology (Cecil et al., 2014) and the understanding of flash rate differences over land and the ocean. These initial useful demonstrations informed the development of the geo-stationary lightning mappers (GLMs).

GLMs allow for continuous monitoring of total lightning’s (i.e., cloud-to-ground and in-cloud; CG/IC) optical energy to derive flash rates, flash size, and flash energy (or the optical output) from space (Rudlosky et al., 2019; Goodman, 2012; Bruning et al., 2019). The Geostationary Operational Environmental Satellite (GOES) GLM’s limitations include not being able to discriminate between IC and CG flashes, while also having a detection efficiency of 20 to 40 percent for short-duration and small flashes lower in the storm (Zhang and Cummins, 2020). These detection efficiencies are compounded during the day, when the difference between the background optical energy and the flash optical energy is reduced due to the sun (Rutledge et al., 2020). Other limiting factors include false flashes generated eastward of the Bahamas, with an artifact known as the ”Bahama Bar” (Fairman, 2020; Trabing et al., 2024).

The benefits of continuous space based lightning observations expanded beyond the GOES domain with other countries and continents began launching geo-stationary lightning mappers, to include China on their Feng-Yun series (Hui et al., 2020) and Europe on their EUMETSAT MTG-LI mission (Grandell et al., 2010).

2.4.3 Relevance to Dissertation

The BoltCast model utilized space-based GLM observations to generate the training targets. LaunchCast utilized the electric field mill observations as inputs into the model, accompanied by GLM group area, to a target generated with MERLIN data. Finally, the analysis section of the HorusCast model utilized LMA data to qualitatively correlate the location of the predicted specific differential phase to LMA source location.

2.5 Radar Data

In this section I describe the observational radar data used in LaunchCast and HorusCast. The discussion opens with a review of the MRMS products used in LaunchCast, then goes into detail of the Horus radar data used for model training in HorusCast. A discussion on how traditional and polarimetric radar variables are correlated to lightning is integrated throughout the section.

2.5.1 MRMS Data

The Multi-Radar Multi-Sensor (MRMS) data suite provides radar derived products across the Conterminous United States CONUS (Lakshmanan et al., 2016). The data are provided on a 1-kilometer grid at a 2-minute update, with derived values ranging from NWP derived fields to isothermal reflectivity levels and differential reflectivity (Lakshmanan et al., 2016). The WSR-88D radar data are merged using an interpolation scheme, that provides continuous radar fields over large areas (Lakshmanan et al., 2006). A major goal of the MRMS project is to quantify precipitation estimates, across CONUS (Zhang et al., 2016).

However, the quality controlled and gridded radar data have been used extensively

in ML efforts, focused on improving the quantitative precipitation estimates with convolutional neural networks and diffusion (Osborne et al., 2026; Schreck et al., 2025; Vicens-Miquel et al., 2025; Han et al., 2022). The MRMS data have also been the target used to emulate radar data from satellite data (Hilburn et al., 2021; Veillette et al., 2020). Additionally, it is a lightning target proxy based on the binary classification of reflectivity values greater than 30 dBZ at the -10°C isotherm (Ortland et al., 2023). LightningCast has also explored data fusion techniques by incorporating MRMS data in addition to satellite data (Cintineo et al., 2026). All of these applications indicate MRMS is well suited for ML model inputs or targets, for non-lightning and lightning applications.

This dissertation utilizes the isothermal reflectivities, differential reflectivity (Z_{DR}), and vertically integrated ice (VII) variables from MRMS as inputs into the LaunchCast model as they are correlated with lightning occurrence. Reflectivity values observed at heights between 0°C and -20°C (within the mixed-phase region) show strong indication of lightning potential due (Liu et al., 2012; Orville et al., 2011) to the non-inductive charging observed in Takahashi (1978). Z_{DR} columns indicate an updraft with precipitation, that may extend beyond 0°C which is a positive indication of lightning (Kumjian et al., 2014; Woodard et al., 2012), but not necessarily directly co-located with flash location (Sharma et al., 2024). The Z_{DR} variable’s relation with observed charge regions in storms was studied with EFMs launched on balloons (Bruning et al., 2007), indicative that EFM and Z_{DR} data are useful for lightning parameterizations given Z_{DR} ’s ability to detect rimed particles (when Z_{DR} is near zero or negative) in mixed-phase regions. VII is also a strong predictor of lightning, indicating the ice is in the same mixed phase region, indicated between the 0°C and -20°C isotherms (Gremillion and Orville, 1998; Orville et al., 2011).

This dissertation does not focus on radar quality control or re-gridding processes

that informed the development of the MRMS dataset (Lakshmanan et al., 2006, 2016), and they are beyond the scope of this work.

2.5.2 Horus Radar

The radar termed Horus, has its name rooted in ancient Egyptian mythology, deemed the 'all-seeing' God. This radar is largely an experimental and research phased array radar (PAR), based on a commercial truck chassis, as shown in Figure 2.13 (Palmer et al., 2023). The Advanced Radar Research Center, in conjunction with the National Severe Storms Lab, developed this radar to combat the limitations of a fixed, static, and planar mounted phased array system that can only scan across elevation angles, with a small fixed azimuthal range (Palmer et al., 2023). Horus is allowed to mechanically rotate (Schvartzman et al., 2026), improving upon these limitations. The radar was developed to meet National Oceanic and Atmospheric Administration (NOAA) specific bias requirements for the polarimetric variables (Pehl et al., 2024), and is a demonstrator of the next generation of radars.

The digital nature of the radar allows for different scanning modes to be implemented. Schvartzman et al. (2026) documents the scanning strategies shown in Figure 2.14. Horus can provide traditional plan position indicators (PPI) at a specific elevation angle, or provide a vertical slice of the storm in a range height indicator. Panels (c) and (d) depict more complex transmission beams with pencil beams or one dimensional imaging that require mechanical rotation of the radar. Panel (e) shows two-dimensional imaging, where the radar plane is fixed and the digital beam steering allows for a higher temporal and spatial sampling of the storm.



Figure 2.13: Figure 1 from (Schvartzman et al., 2026). A picture of the Horus truck taken outside of the National Weather Center, Norman, OK.

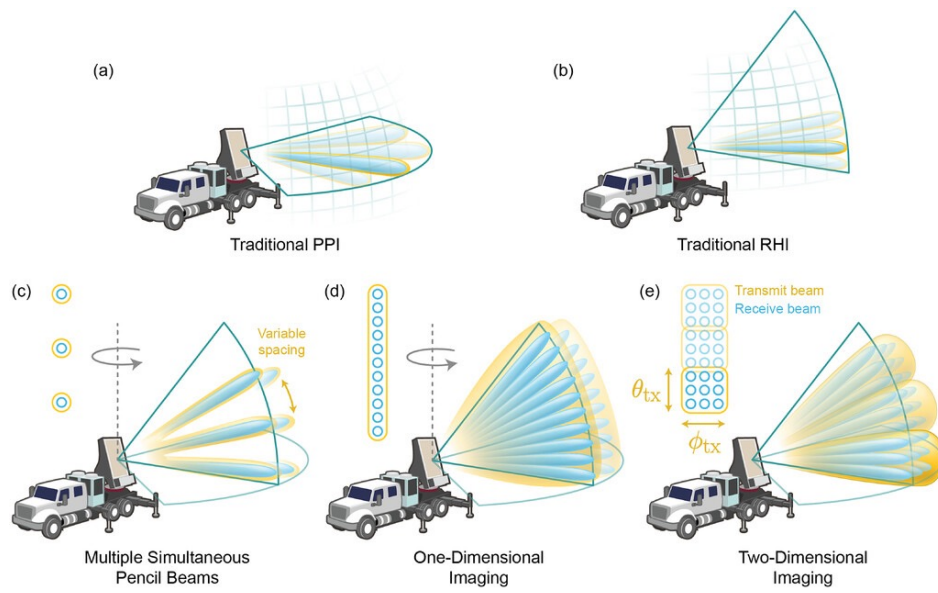


Figure 2.14: Figure 2 from (Schvartzman et al., 2026), showcasing the various scanning strategies of the Horus radar.

2.5.3 Specific Differential Phase and Estimations

Specific differential phase, K_{DP} , is dependent on the polarimetric observation of total differential phase, $\hat{\Phi}_{DP}$.

$$\hat{\Phi}_{DP} = \phi_{DP} + \delta + \epsilon = \phi_h - \phi_v \quad (2.33)$$

The total differential phase, $\hat{\Phi}_{DP}$ is cumulative along the radar beam and increases monotonically as radar range increases (Mishler et al., 2024). It measures the difference in phase shift of the horizontal (ϕ_h) and vertical (ϕ_v) propagating radar waves that are received at the radar. Positive values indicate more horizontally oriented hydrometeors and negative values indicate the presence of more vertically oriented hydrometeors. Equation 2.33 decomposes $\hat{\Phi}_{DP}$ into the differential propagation phase ϕ_{DP} , differential back-scattering phase δ , and the measurement error ϵ . ϕ_{DP} , the differential propagation phase is only measured when the radar beam interacts with precipitation, and the measurement error is often negligible. ϕ_{DP} cannot be sensed directly, as the scatter radar beam must travel back to the radar. When the scatterers, like rain, are less than the wavelength of the radar (S-band is approximately 10-cm) the Rayleigh scattering regime is present, and ϕ_{DP} is dominated by the horizontal propagation phase, and δ remains low. When hail or tornado debris is introduced, the Rayleigh scattering regime is no longer applicable, and geometric scattering is introduced, introducing more scattering and increasing δ (Mishler et al., 2024).

Equation 2.34 is the physical representation of specific differential phase. Unfortunately, due to sensing limitations, a continuous ϕ_{DP} field intrinsic to the storm cannot be observed directly, only estimated at the radar. Thus, K_{DP} estimations are implemented with varying discretizations of the derivative and accounting of the back-scattering δ to account for the hydrometeor type.

$$K_{DP} = \frac{1}{2} \frac{\partial \phi}{\partial r} \quad (2.34)$$

The two K_{DP} estimations used in Chapter 5 are presented in Equation 2.35, (Reimel and Kumjian, 2021b; Vulpiani et al., 2012). Equation 2.36 was adapted from the National Severe Storms Laboratory (Carlin, 2026).

$$K'_{DP} \approx 0.5 \left[\Phi_{DP} \left(r_k + \frac{L}{2} \right) - \Phi_{DP} \left(r_k - \frac{L}{2} \right) \right] \quad (2.35)$$

The National Weather Service K_{DP} estimator is implemented in the Operational Radar Product Generator (ORPG). The estimate used is based on the least squares method discussed in (Doviak and Zrnica, 1993), and shown in Equation 2.36.

$$K'_{DP_{NWS}} \approx \frac{\sum_{i=1}^n \left(\left[\hat{\Phi}_{DP}(i) - \bar{\hat{\Phi}}_{DP} \right] [r(i) - \bar{r}] \right)}{2 \sum_{i=1}^n [r(i) - \bar{r}]^2} \quad (2.36)$$

This method specifies a fixed window width along the ray, to make the K_{DP} calculation at the center point, the terms with the bar in Equation 2.36. A 9 gate (2,250 meter) window is used when the center point has a reflectivity greater than or equal to 40 dBZ, and a 25 (6,250 meters) gate window is used for all other reflectivity values. n in Equation 2.36 refers to the number of gates. This methodology is based on the fixed, WSR-88D network deployed across the CONUS. I use a 2,250 meter window in the NWS requirement during high reflectivities to derive the Horus based window size of 29 gates for the Vulpiani K_{DP} target, calculated from Equation 2.35. I also use the same window size to calculate an NWS comparison, to show the skill of HorusCast relative to the ORPG product.

The variability of these K_{DP} estimates is discussed in Reimel and Kumjian (2021a). They compare other methods like the Schneebeli et al. (2014) and Measaka et al. (2012).

These methods are available in Py-ART (Helmus and Collis, 2016), but generate a noisier K_{DP} field. Thus, the Vulpiani method was chosen for its smooth and more continuous K_{DP} signal which is favored by ML gradient descent algorithms.

Negative K_{DP} implies an electrified area of the storm (Ryzhkov and Zrnica, 2007), that may induce lightning. Observations indicate that ice-crystals align parallel to the vertically oriented electric field, in thunderstorms (Caylor and Chandrasekar, 1996). Physically, this generates an average projection of ice crystals onto the vertical direction causing the differential phase to decrease with range, generating a negative K_{DP} (Ryzhkov and Zrnica, 2007). Melnikov et al. (2019) used the negative K_{DP} region to compare different transmit and receive modes of the radar, indicating that alternating transmit and receive modes generates larger electrification regions when compared to a simultaneous transmit and receive mode; the Horus configuration for data used in this dissertation is in simultaneous transmit and receive. In Chapter 5, the error of the negative K_{DP} is presented.

2.5.4 Relevance to Dissertation

While loosely connected, BoltCast utilized forecast-simulated reflectivity both as inputs and as a crude comparison to convection capable of producing lightning. LaunchCast transitioned to the observational and quality controlled isothermal reflectivities, vertically integrated ice, and differential reflectivity products in the MRMS product suite (Howard et al., 2016; Lakshmanan et al., 2016). Finally, HorusCast utilized state-of-the-art digital phased array radar traditional and dual polarization variables to estimate specific differential phase.

I want to emphasize that this dissertation is not aimed to influence radar theory. This dissertation is an application of the post-processed and quality controlled radar data only, based on the observations presented in this section. For HorusCast, specifi-

cally, I was limited to the post-processed data hosted on a server several months after each deployment.

2.6 Overview of NWP and Lightning

This section provides an overview of the NWP products used in the models I developed. It also provides further background of some research based methods for lightning parameterization with traditional NWP.

2.6.1 Operational NWP Lightning Products

Harrison et al. (2022) developed a statistically calibrated thunder product, based on a 40-dBZ reflectivity, lifted index, and quantitative precipitation forecast derived from the ensemble members within Storm Prediction Center’s High Resolution Ensemble Forecast System (Roberts et al., 2019). This thunder product was designed to predict only cloud-to-ground flashes, and not capture the impacts of in-cloud flashes. This product provides forecasts twice daily, initialized at 00Z and 12Z, at 1-hour time-steps out to 48 hours (Harrison et al., 2022).

The Short Range Ensemble Forecast (SREF) System is scheduled for decommission in late 2026. The SREF lightning product (i.e., cloud-physics thunder parameter) is a function of equilibrium level temperature and convective available potential energy (CAPE), that provides lightning forecasts out to 63 hours (Bright et al., 2005). The HREF product shows improved skill when compared to the legacy SREF calibrated thunder product. Harrison et al. (2022) document a 12Z HREF maximum CSI score of .32 that decreases to .2, and the SREF struggles to exceed .225 that decreases to .15 throughout the forecast window. The SREF CSI values never exceed the HREF throughout the 48-hour forecast window.

The HREF includes the High Resolution Rapid Refresh (HRRR), which is a convection-allowing model that updates every hour (Dowell et al., 2022). A model forecast is initialized every hour. When initialized at 00, 06, 12, and 18 UTC, a 48-hour forecast is provided. Otherwise, an 18-hour forecast is generated. Forecasts 1-hour and beyond provide a lightning product, parameterized by the product of the graupel mixing ratio and the updraft velocity (Blaylock and Horel, 2020; McCaul et al., 2009). Additionally, a sub-hourly HRRR is developed on a smaller subset of variables, with 15-minute temporal resolution (Dowell et al., 2022)⁴, although without the lightning parameterization. This rapidly updating system is valuable at short-range scales (1-6 hours), due to its rapid updates. However, the skill of a binary lightning mask generated from a flash extent density greater than 0.04 flashes per square kilometer is poor; documented to only generate skill when aggregated to 240-km, roughly 80 times larger than the native resolution of the model (Blaylock and Horel, 2020). LaunchCast performance metrics are compared to the HRRR lightning parameterization in Chapter 5.

At the medium-scale, or several day forecasts, forecasters are forced to translate convection related parameters from large global models. Global models depict large areas of instability (i.e., CAPE, lifted index), coincident with dynamical forcing (i.e., vertical velocities), requiring a human forecaster to discern whether or not lightning will occur. Specifically, the Global Forecast System (NCEP, 2004) provides global forecasts out to 384 hours. This specific model is archived at .25 degree resolution (NCEP, 2015). After July 2018, the GFS included the hydrometeor mixing ratios as model output, with the update to the Zhou et al. (2019) microphysics scheme. In the next section, I discuss how the hydrometeor mixing ratios influenced NWP flash rate parameterizations. Unfortunately, the forecasted presence of graupel, ice, or snow does not solely determine whether lightning will occur. But, when used as a ML input with

⁴https://rapidrefresh.noaa.gov/hrrr/HRRRsubh/Welcome.cgi?dsKey=hrrr_subh_ncep_jet

a GLM-based target, the mixing ratios may add skill to lightning prediction out to four days, and are included in the BoltCast model.

2.6.2 NWP Based Lightning Research

I review these techniques to show that NWP models are able to parameterize lightning and charge on the storm-scale. The AI/ML approaches I use enable direct learning of statistical relationships between operational-scale NWP predictors and observed lightning, without requiring explicit storm-scale intermediaries. The AI/ML models I developed do not account for storm scale electrification or the conservation of charge, a limitation of this work. Thus, no insights can be gained on the relation of my models' inputs to electrification processes, that were first discretized in a lab setting.

Lab experiments were transitioned to a numerical model by varying ice and cloud nuclei number concentrations, and exploring charge distributions within a parameterized storm (Takahashi, 1984). These particular integrations matched the lab observations well, and set the stage for future numerical studies with more advanced models. Lightning flashes were able to be stochastically propagated through storms, ensuring the conservation of charge occurred (Mansell et al., 2005). Identifying a flash rate, however, still posed a challenging problem until Dahl et al. (2011) developed a simple model based on a plate capacitor, to determine the flash rate of isolated thunderstorms and mesoscale convective systems; with the primary limitation still not accounting for charge separation. Additional impacts of including the electric potential in models were studied in hurricanes, severe thunderstorms, and winter storms by including a Poisson solver of the electric potential, charging rate, and electric field magnitude studying the charge structures (Fierro et al., 2013). Due to the large computational time of these discretizations, flash rate simplifications utilizing only hydrometeor mixing ratios and CAPE were developed that match flash rates observed from space (Lopez, 2016). Re-

cent reviews of the flash rate parameterization schemes noted some commonalities, where radar reflectivity, updraft velocity, temperature, and hydrometeor mixing ratio are used across 18 different schemes and techniques (Cummings et al., 2024).

2.6.3 Relevance to Dissertation

BoltCast translated traditional GFS convection related outputs into a specific lightning probability, 4-days in advance. BoltCast provided a direct, but qualitative, comparison to the HREF calibrated thunder guidance product. LaunchCast utilized sub-hourly HRRR near-surface parameters as inputs, providing forecasts of mean sea level pressure (MSLP), divergence, precipitation rate, temperature, and dew point; mimicking forecast surface observations. A comparison to the HRRR 1- and 2-hour lightning products were made, showing higher skill in that comparison. All of these traditional NWP input parameters were rigorously evaluated with explainable AI techniques, showcasing the importance of CAPE and MSLP. HorusCast is solely based on observations, and does not rely on any NWP data.

2.7 AI-Based Lightning Prediction Tools

LightningCast (Cintineo et al., 2022), a ML-based lightning product, utilized deep learning (i.e., U-Net) for lightning prediction. LightningCast uses infrared and visible GOES satellite imagery to make a binary prediction of GLM lightning up to 1-hour. LightningCast generates the target by accumulating the flashes over a 1-hour period, then generating a binary mask for the training target. Follow-on experiments switched the LightningCast target from GLM to a 30-dBZ binary target at -10°C with the same architecture (Ortland et al., 2023). More recently, LightningCast included the -10°C isothermal reflectivity into an input while also increasing the training dataset

size (Cintineo et al., 2026). LightningCast was initially trained over CONUS, but has recently been deployed where similar lightning imagers exist using Himawari-8 data⁵. The ML architecture basis for all of these models is the 2-D convolutional U-Net, described in Ronneberger et al. (2015). While these U-Nets translate observations into a lightning or thunderstorm probability, NWP data can be integrated into deep-learning architectures to assist in lightning prediction and extend predictions beyond one-hour.

Alternative architectures to U-Nets have been explored for lightning applications. Recurrent neural networks (RNN) have been used previously in translating NWP data into lightning probabilities (Wang et al., 2023). The spatial convolutions in RNNs combined with hidden- and cell-states allow for multi-scale spatiotemporal dependencies to be learned (Shi et al., 2015). LightNet+ boasts a skillful lightning forecast with 6-hour lead time over China using six convolutional long short-term memory (LSTM) layers (Geng et al., 2021). In Switzerland, high- and low-spatial resolution NWP data are merged into 12 LSTM layers to provide a 1-hour lead time prediction of lightning (Leinonen et al., 2022). RNN architectures for lightning prediction across Africa have also been implemented, with only 2-3 LSTM layers for 1-hour lead time (Essa et al., 2020). NASA’s SPoRT AI tool also has a documented blend of LSTM and CNN that predicts GLM lightning out to 30-minutes using satellite and isothermal reflectivity as inputs⁶. Pre-processing an ensemble of WRF microphysics schemes can provide a potential increase in lightning lead times, for very small domains (Thomas and Noble, 2024), but sufficient training data must be provided. Sobash and Ahijevych (2025) uses re-gridded HRRR data as inputs to a CNN, showcasing the utility of CNNs’ predictive skill with the HRRR data at 40- and 80-km resolution. Finally, traditional, surface observation based approaches have been used in Switzerland to predict lightning within

⁵https://cimss.ssec.wisc.edu/cspgeo/lightningcast_v1.0b1.html

⁶<https://weather.ndc.nasa.gov/sport/lightning-viewer/>

specified radii of the observation station (Mostajabi et al., 2019).

2.7.1 Relevance to Dissertation

The previous AI-based lightning prediction tools provided a strong foundation to build upon. They included unique ways to evaluate the lightning based tools, and different perspectives to consider when building BoltCast, LaunchCast, and HorusCast. The tools mentioned above do not, however, provide multi-day lead times as BoltCast does, generate skillful nowcasts with EFM data as LaunchCast does, or aid in K_{DP} estimation as HorusCast does. The following chapters provide details and analysis discussing my three novel approaches to lightning prediction and K_{DP} estimation with AI.

Chapter 3

BoltCast: Medium-range and Synoptic Scale Lightning Prediction

This chapter reviews the development of a new model that generates skillful lightning predictions out to 96 hours. It describes the model data used as inputs, the lightning target generation, two model architectures, and the evaluation metrics and explainable AI techniques.

3.1 Data and Methodology

3.1.1 GFS Data and GLM Data

Output from the Global Forecast System (GFS) from July 2019 through June 2024 is used due to its common availability to forecasters, routine initializations (00Z, 06Z, 12Z, and 18Z), 3 hourly temporal resolution to 96 hours, and 0.25 degree archive (NCEP, 2015). These characteristics of the GFS output make it particularly advantageous to generate CONUS-wide lightning forecasts. GFS data over the CONUS, between 21.25 and 52.50 degrees N, and 234.00 and 297.75 degrees are used, resulting in a 128 by 256 image size used by the deep learning models.

The GFS upgrade in June 2019 included the Geophysical Fluid Dynamics Laboratory microphysics scheme, allowing outputs of hydrometeor types (snow, ice, rain,

graupel) (Zhou et al., 2022). These mixing ratios and the upward velocity are included as model inputs, listed in Table 3.2, due to their use in other flash rate parameterization schemes (Cummings et al., 2024). Maximum values throughout the atmospheric column (1000- to 250- millibars where applicable) and across prediction days (Day 1: Forecast hours 00-21; Day 2: Forecast hours 24-45; Day 3: Forecast hours 48-69; Day 4: Forecast hours 72-93) are calculated for the hydrometeor variables and the vertical velocity. Minimum Lifted Index (LI) is computed across the prediction days due to negative values indicating higher chances of convection with lightning, similar to high values of surface-based CAPE. Additionally, daily maximums of instantaneous precipitation rates and reflectivity values are also indicative of convection capable of producing lightning. The environmental input parameters are helpful in constraining the GFS simulated environment to areas most conducive to lightning-producing storms. The input features are normalized using min-max scaling, with minimums and maximums calculated across all available data spanning the timeframes listed in Table 3.3. These data are stored in a 5-D TensorFlow tensor, with dimensions of (no. samples, prediction day (4), latitude (128), longitude (256), parameter (9)). These inputs are represented visually in Figure 3.2. Further, terrain, latitude, and longitude are not included as inputs as they have shown little impact on ML model performance, may cause the ML model to overfit (Leinonen et al., 2022), and do not allow the model to be responsibly used outside of CONUS.

Binary lightning classifications are derived from GOES Level 2 GLM data, also shown in Figure 3.2. For reference, a climatology of GLM flash extent density is provided in Figure 3.1. Latitudes and longitudes from the flashes are counted in each GFS grid cell. A binary mask is created where grid cells with one or more flashes are masked with a 1, and the other cells without a flash are masked to zero. This procedure generates a binary grid of labels for training. GOES-16 data are used east

of 260 degrees and GOES-17 (before January 1, 2023) and GOES-18 (after January 1, 2023) west of 260 degrees. Daily classification images are generated for each 24-hour prediction day on a six-hour sliding window based on the 00 UTC, 06 UTC, 12 UTC, and 18 UTC model runs. These classifications are then matched to the corresponding GFS prediction day, indicated in the 01 June 2022 example in Table 3.1. This matching scheme ensures GFS data are not used twice (e.g, f024 is only used in Day 2, and not inclusive for Day 1) within a sample and ensures the lightning prediction is for 3-hours beyond f021, f045, etc.

GFS Model Run	GLM Valid Times
0000 UTC 1 June 2022	—
Day 1: f000 - f021	0000 UTC 1 Jun 2022 - 0000 UTC 2 Jun 2022
Day 2: f024 - f045	0000 UTC 2 Jun 2022 - 0000 UTC 3 Jun 2022
Day 3: f048 - f069	0000 UTC 3 Jun 2022 - 0000 UTC 4 Jun 2022
Day 4: f072 - f093	0000 UTC 4 Jun 2022 - 0000 UTC 5 Jun 2022
0600 UTC 1 Jun 2022	—
Day 1: f000 - f021	0600 UTC 1 Jun 2022 - 0600 UTC 2 Jun 2022
Day 2: f024 - f045	0600 UTC 2 Jun 2022 - 0600 UTC 3 Jun 2022
Day 3: f048 - f069	0600 UTC 3 Jun 2022 - 0600 UTC 4 Jun 2022
Day 4: f072 - f093	0600 UTC 4 Jun 2022 - 0600 UTC 5 Jun 2022
1200 UTC 1 Jun 2022	—
Day 1: f000 - f021	1200 UTC 1 Jun 2022 - 1200 UTC 2 Jun 2022
Day 2: f024 - f045	1200 UTC 2 Jun 2022 - 1200 UTC 3 Jun 2022
Day 3: f048 - f069	1200 UTC 3 Jun 2022 - 1200 UTC 4 Jun 2022
Day 4: f072 - f093	1200 UTC 4 Jun 2022 - 1200 UTC 5 Jun 2022
1800 UTC 1 Jun 2022	—
Day 1: f000 - f021	1800 UTC 1 Jun 2022 - 1800 UTC 2 Jun 2022
Day 2: f024 - f045	1800 UTC 2 Jun 2022 - 1800 UTC 3 Jun 2022
Day 3: f048 - f069	1800 UTC 3 Jun 2022 - 1800 UTC 4 Jun 2022
Day 4: f072 - f093	1800 UTC 4 Jun 2022 - 1800 UTC 5 Jun 2022

Table 3.1: An example of the matching of the GFS forecast hours to the GLM observations, for 01 June 2022’s GFS model runs.

These data are built into five different rotations for cross-validation listed in Table 3.3. Each rotation pertains to a separate trained ML model, thus this paper provides

GLM Climatology

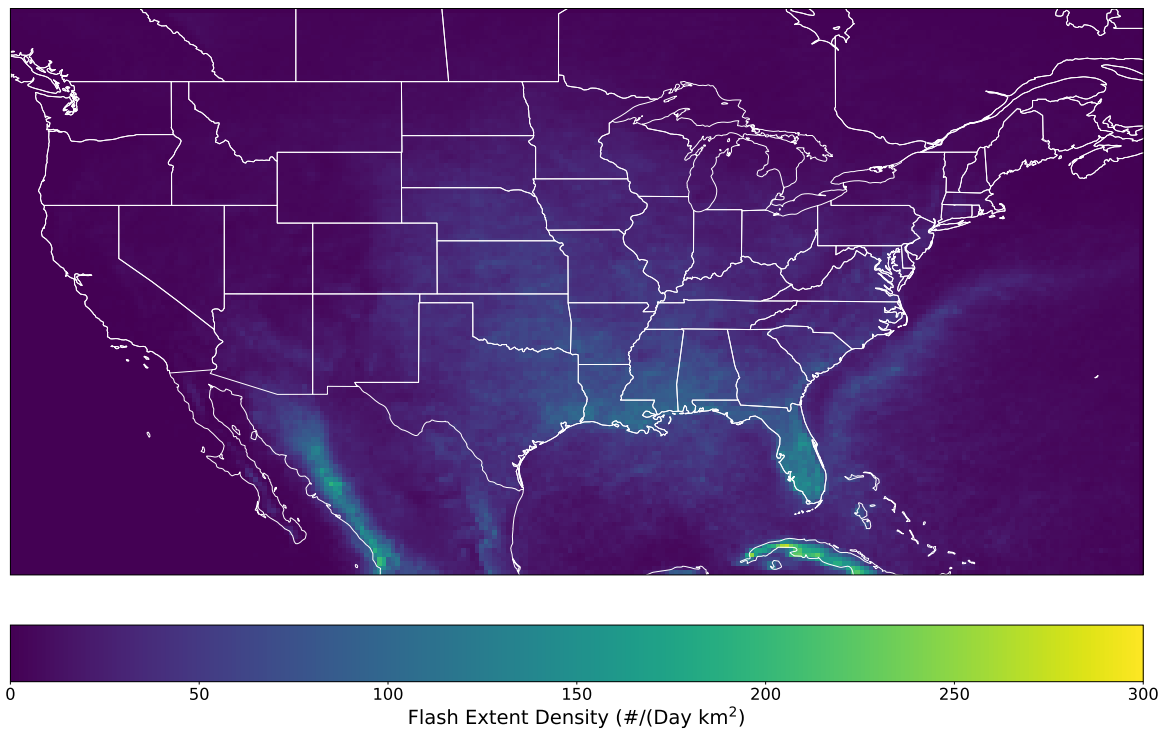


Figure 3.1: Spatial climatology for GLM flashes. Each grid cell represents the flash extent density observed between 2021 through 2024.

results and accumulated statistics for 10 different ML models: 5 for the U-Net and 5 for the LSTM architecture. The statistics presented in the results section are for the test datasets, and are calculated by concatenating the outputs of the 5 rotations for the U-Net and LSTM architectures. For reproducibility, the number of GLM flashes from GOES-16,-17, and -18 are included for each training, validation, and test dataset.

Calculation	Parameter
Maximum between 1000–250 mb over a 24-hour prediction day	Graupel Mixing Ratio [kg/kg] Ice Mixing Ratio [kg/kg] Rain Mixing Ratio [kg/kg] Snow Mixing Ratio [kg/kg] Vertical Velocity [m/s]
Maximum GFS Model Derived Composites over a 24-hour prediction day	Composite Reflectivity [dBZ] Surface Based CAPE [J/kg] Lifted Index [K] (negative) Precipitation Rate [$kg/(m^2 s)$]
Output	Binary Lightning Classification

Table 3.2: List of input and output features. The left column describes how the feature is calculated. The right column identifies the specific GFS variable. The last row pertains to the processed GLM data.

3.1.2 AI/ML Model Architectures

Both U-Net-like and LSTM-based ML architectures can skillfully predict lightning (Cintineo et al., 2022; Geng et al., 2021; Leinonen et al., 2022). However, a direct comparison has never been documented to determine their relative skill for binary lightning classification. It is hypothesized that one LSTM layer can generate similar model performance as a robust U-Net architecture because the recurrent nature of an LSTM model allows learning to occur across the temporal dimension. A 3-D convolutional U-Net can learn the same relationships if the convolutions are calculated across the temporal, spatial, and feature dimensions. Both model architectures presented here are

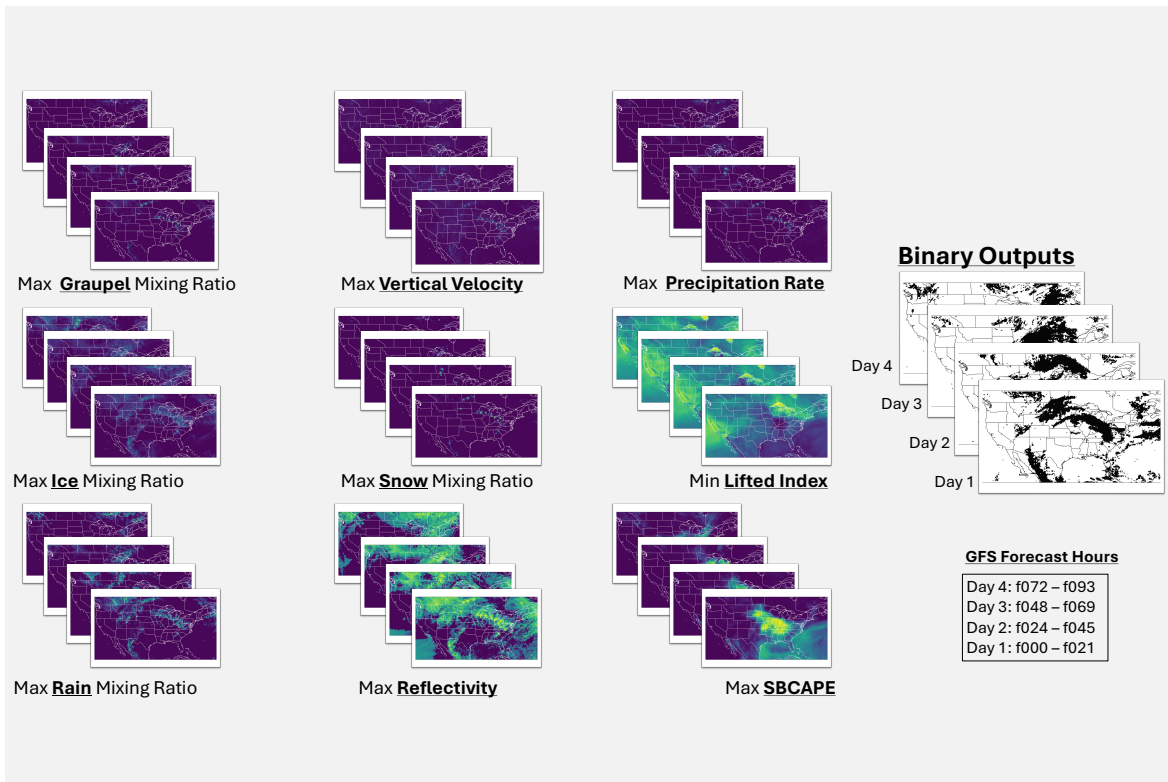


Figure 3.2: Visual representation of the GFS-derived inputs and target output of the models. The nine input variables are shown in the first three columns, with example binary outputs on the right. GFS forecast hours are included in the bottom right.

Rot		Training	Validation	Test
1	Valid Times	July 2019 – May 2022; June 2024	June 2022 – May 2023	June 2023 – May 2024
	Samples	4,361	1,452	1,460
	G16 Flash Ct	743,524,069	259,731,569	318,121,780
	G17/18 Flash Ct	223,673,177	95,347,152	102,155,527
2	Valid Times	June 2020 – May 2022; June 2024	June 2023 – May 2024	July 2019 – May 2020
	Samples	4,470	1,460	1,343
	G16 Flash Ct	782,033,358	318,121,780	221,222,280
	G17/18 Flash Ct	243,740,124	102,155,527	75,280,205
3	Valid Times	June 2021 – June 2024	July 2019 – May 2020	June 2020 – May 2021
	Samples	4,471	1,343	1,459
	G16 Flash Ct	870,691,564	221,222,280	229,463,574
	G17/18 Flash Ct	281,987,937	75,280,205	63,907,714
4	Valid Times	July 2019 – May 2020; June 2022 – June 2024	June 2020 – May 2021	June 2021 – May 2022
	Samples	4,358	1,459	1,456
	G16 Flash Ct	830,657,242	229,463,574	261,256,602
	G17/18 Flash Ct	285,065,356	63,907,714	72,202,786
5	Valid Times	July 2019 – May 2021; June 2023 – June 2024	June 2021 – May 2022	June 2022 – May 2023
	Samples	4,365	1,456	1,452
	G16 Flash Ct	800,389,247	261,256,602	259,731,569
	G17/18 Flash Ct	253,625,918	72,202,786	95,347,152

Table 3.3: Number of images per dataset type. These images generate a 5-Dimensional Tensor with batch size (a subset of the number of samples), the prediction day, latitude, longitude, and GFS variables shown in Figure 3.2. The outputs generate a 4-Dimensional Tensor with batch size, the prediction day, latitude, and longitude.

generated with the Adam optimizer (Kingma and Ba, 2014) and evaluated in Keras 3.0 and Tensorflow version 2.15.0 (Abadi et al., 2015). These two models are widely used in prediction and classification tasks today, with great versatility. With the primary contribution focused on extending the lead times of lightning prediction, documenting the performance of these convolutional based models provides a baseline for future development with more complex deep-learning architectures like graph neural networks, vision transformers, or diffusion models.

3.1.2.1 3-D Convolutional Neural Network: U-Net

Ronneberger et al. (2015) developed an encoder-decoder architecture (in the shape of a U) to conduct image classification tasks for medical images. These types of models have been applied to classify and predict meteorological phenomenon like fronts, wildfires, and lightning (Justin et al., 2023, 2025; Earnest et al., 2024a,b; Cintineo et al., 2022).

Lightning- and Thunder-cast (Cintineo et al., 2022; Ortland et al., 2023) use the same infrared and visible satellite-imagery inputs into a series of 2-D convolutions to classify lightning or reflectivity greater than or equal to 30 dBZ at the -10°C isotherm, respectively. The 2-D convolutional layers allow for multiple observation types (i.e., numerous GOES ABI channels) in the filter dimension, at one valid time, to make predictions one hour later. However, neither model includes additional temporal information beyond 1-hour during model training.

One way to include additional temporal information is to expand the convolution operation to three dimensions. Then, when performing a 3-D convolution, information across multiple time-steps can aid in the prediction. Here, I use the 3-D convolutional U-Net shown in Figure 3.3, ensuring learning across the temporal dimension.

The input layer is generated by concatenating the 9 input images (see Figure 3.2) across the four prediction days. These inputs are put into an encoder series of four

convolutional blocks, and each block contains three 3-D convolution layers, with a $t \times 4 \times 4$ kernel filter, where t is either 4, 2, or 1 depending on the depth of the U-Net. Between each convolutional block, a max pooling operation is conducted until a single-time vector is generated at the bottom of the U-Net. The number of convolutional blocks and layers within each block were considered a hyperparameter, and varied to minimize the validation loss. A symmetric decoder is used to predict the lightning classification labels at the same spatial resolution of the input images. To restore the original image shape, a 3-D UpSampling operation is performed.

A hyperparameter search was conducted monitoring the validation loss - binary cross entropy. The learning rate was established by monitoring continuous validation loss curves, resulting in 0.00001 for both the U-Net and LSTM. With the learning rate established, for both the U-Net and LSTM, a hyperband search (Li et al., 2018) was conducted on each model varying the 2D convolutional sizes, convolutional activation functions, 3D spatial dropout, and L2 regularization. The hyperband search was run for maximum epochs of 50, a factor of 3, and 2 hyperband iterations. The search resulted in the same hyper-parameters chosen for both the UNet and LSTM: a 2-D convolution size of 4, the relu activation function (Agarap, 2019), no spatial dropout, and no L2 regularization. Although the optimal hyperparameters were identical for the two architectures, they were obtained through independent searches, indicating that this outcome reflects the shared data characteristics rather than imposed architectural constraints.

3.1.2.2 One-layer LSTM

An early adaptation of two-dimensional convolutional LSTMs (termed ConvLSTM2D in TensorFlow) in meteorology was for precipitation nowcasting (Shi et al., 2015). The LSTM feature allows for temporal information to flow from one NWP model time

step to the next, while the convolutional feature allows for spatial relationships to be learned. LightNet+ used this layer type with WRF model inputs to predict lightning out to 6 hours (Geng et al., 2021). The LightNet+ deep-learning architecture informed initial developments of the one-layer LSTM deep-learning architecture presented in Figure 3.4.

In this study, the time steps (i.e., prediction day) are concatenated in the first non-batched dimension and input into one ConvLSTM2D layer, to allow learning across the forecast days. Here, short- and long-term (i.e., hidden and cell) memory states are generated across the 4 forecast days. These short-term hidden states are passed from Day 1 to 2, Day 2 to 3, and Days 3 to 4, within the ConvLSTM2D layer. The sequence of hidden states is returned as output from the ConvLSTM2D layer; but the long-term cell state is not. Geng’s LightNet+ model performs convolutions on the hidden and cell states themselves, breaking away from the ConvLSTM2D layers’ suggested use in Shi’s paper (Shi et al., 2015). This may contribute to LightNet+’s low CSI values of 0.30 for a 6-hour lead time. For brevity, throughout the remainder of this paper the model presented in Figure 3.4 will be referred to as the LSTM model.

The recurrent nature of the LSTM layers requires the hyperbolic tangent activation function to be used with other hyperparameters (see Fig. 3.4). The ReLU activation function is used only in the 3-D convolutional layers (Agarap, 2019). The training time and resources for the presented U-Net and LSTM architectures are presented in Table 3.4. The Weights and Biases software, a web-based¹ platform was used to document model training times and monitor HPC resource allocations. The total model training time, including loading the pre-processed TensorFlow datasets with a batch size of 32, are listed in Table 3.4. Data downloading, processing and staging time is not considered in the timing computations. It is noted the LSTM model trained faster than the U-Net

¹<https://wandb.ai/>

under identical early-stopping and batch settings on $1 \times A100$ GPU.

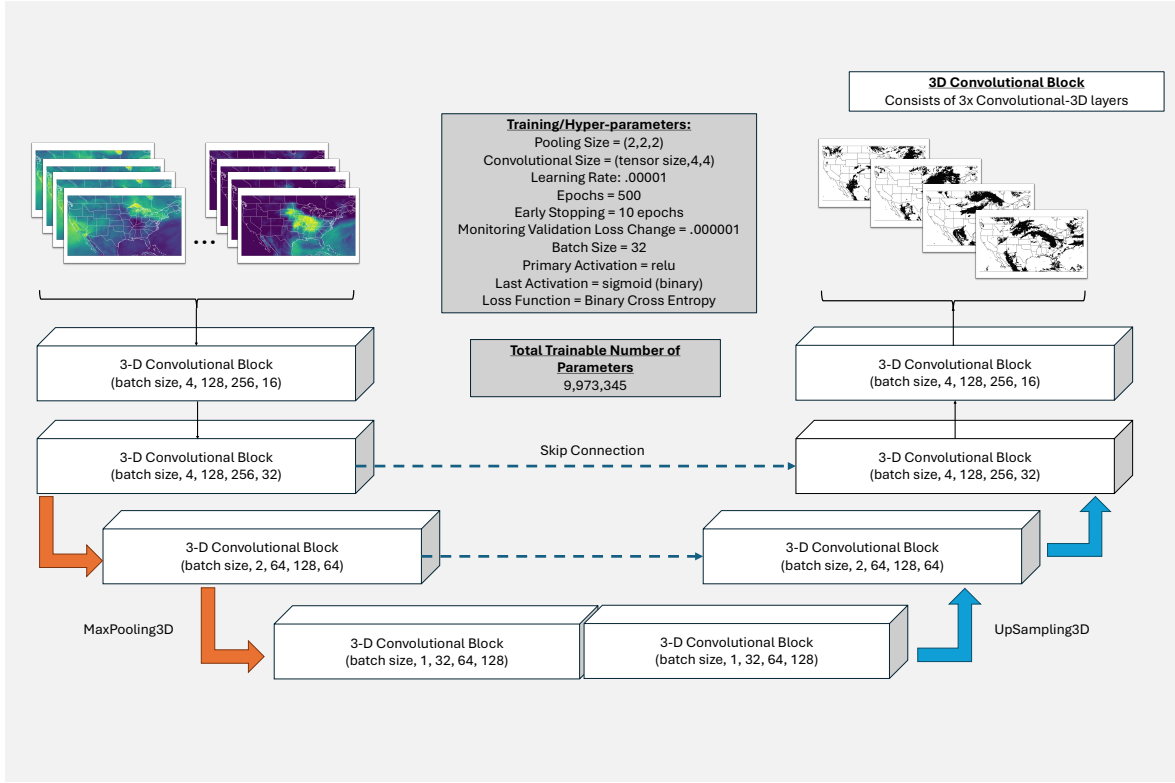


Figure 3.3: U-Net Architecture diagram. Example inputs and binary outputs are included in the upper-left and upper-right. The layer names and tensor shapes are included for reference. Each 3-D Convolutional Block contains three 3-D convolutional layers. Orange arrows indicate a max pooling operation and the blue arrows indicate a 3-D UpSampling operation. Skip connections are annotated with the teal dashed arrows. Hyper- and total number of trainable-parameters are included in the center dark-gray boxes.

3.1.3 Performance Diagrams and Case Studies

To compare the two model architectures and their corresponding outputs across all 5 rotations, attributes and precision-recall diagrams are employed to objectively evaluate model performance (Hsu and Murphy, 1986; Roebber, 2009). The attributes diagram documents an under- or over-forecast bias when compared to a perfect forecast and no-skill forecast, by plotting the observed probability versus forecast probability (Hsu

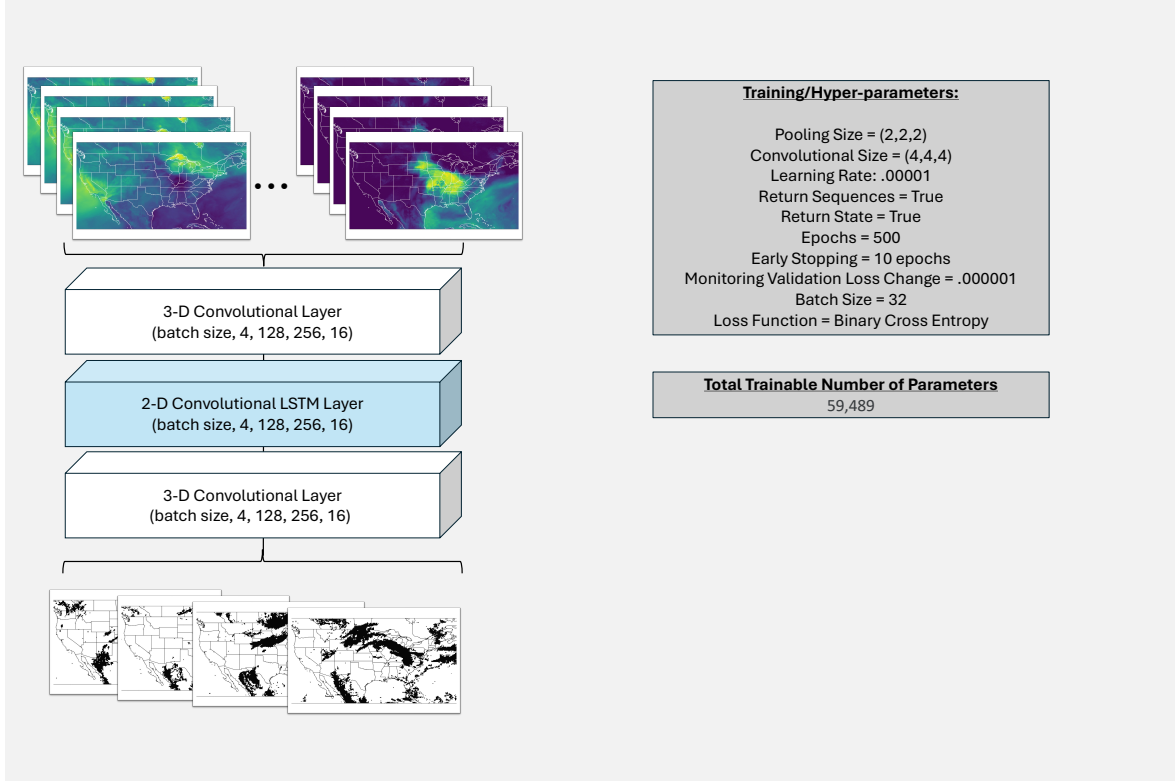


Figure 3.4: The LSTM model architecture. Example inputs and binary outputs are included in the top and bottom. The layer names and tensor dimensions are included for reference; no blocks with multiple convolutional layers are used. White boxes correspond to 3-D Convolutional Layers and the light-blue correspond to 2-D Convolutional LSTM layers. The dark-gray boxes on the right indicate the hyper- and total trainable number of parameters.

Rotation	U-Net Training Time (Epochs)	LSTM Training Time (Epochs)	GPU Resources
1	3:39:22 (103)	2:11:39 (222)	1xNVIDIA-A100
2	4:58:25 (136)	2:32:20 (255)	1xNVIDIA-A100
3	3:58:21 (116)	2:05:33 (202)	1xNVIDIA-A100
4	3:21:09 (099)	3:06:20 (319)	1xNVIDIA-A100
5	4:10:33 (124)	4:08:31 (415)	1xNVIDIA-A100

Table 3.4: U-Net and LSTM model training times and the resources for each rotation listed in 3.3. Times are listed as HH:MM:SS. The number of training epochs are also provided in parentheses. Early stopping was implemented monitoring the validation loss, with a min delta of .000001 and a patience of 10 epochs, for a maximum of 500 epochs.

and Murphy, 1986). These diagrams depict if forecast have skill relative to a baseline climatology. This paper highlights the U-Net and LSTM attributes for the 4 forecast days, and provides the Brier Score (i.e., mean squared error; Equation 3.1) for each forecast day in the legend of Figure 3.5. Here y_{true} is the observed lightning classifications, y_{pred} is the predicted lightning probabilities, and N is the number of predictions across space. In addition, histograms of the predicted probabilities are shown to document any biases in the predictions. For further evaluation, precision-recall curves are provided to quantify model performance across all probabilistic prediction thresholds. These curves are dependent on the precision or success ratio (SR), recall or probability of detection (POD), and critical success index (CSI) shown in equations 3.2, 3.3, and 3.4. Equations 3.2, 3.3, and 3.4 are a function of true positives (TP ; hit), false negatives (FN ; miss), or false positive (FP), where the model predicted an event that was not observed (Wilks, 2012). CSI, precision, and recall are calculated across thresholds varying from 0 to 1 in increments of .05, using the Tensorflow Keras Metrics API. All of these metrics are maximized at a value of 1.

These metrics allow for the area under each precision-recall curve (PR-AUC) to be calculated, encompassing model performance at all thresholds. The PR-AUC is presented for individual seasons to examine whether or not the model has biases towards one season. The seasons were used from each rotation’s test dataset. Here meteorological seasons are used: Summer: June, July, August; Fall: September, October, November; Winter: December, January, February; Spring: March, April, May.

$$BS = \frac{\sum_{j=1}^N (y_{pred} - y_{true})_j^2}{N} \quad (3.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

$$CSI = \frac{TP}{TP + FN + FP} \quad (3.4)$$

To build trust in the model outputs, I provide contrasting (a well-forecast and a miss-forecast) lightning forecasts for 12 UTC 13-14 June 2022 and 12 UTC 09-10 December 2022.

3.1.4 Explainable AI

Explainable AI (XAI) techniques allow users to understand how the inputs impact ML model performance (McGovern et al., 2019; Flora et al., 2024). Using XAI helps identify whether the atmospheric, derived 2-D parameters like CAPE and reflectivity impact model performance more than the hydrometeor mixing ratios and vertical velocity, for example. Single-variable permutation importance is computed, where importance is quantified as the change in the PR-AUC. A positive importance value implies the permuted inputs had a smaller PR-AUC, and a negative value implies the permuted inputs generated a larger PR-AUC. This examination of permutation importance is not meant to infer any physical implications regarding the electrification processes outlined in Byers or Takahashi (1949; 1978), but rather quantify if components of flash

rate parameterizations listed in Cummings (2024) having meaningful impact for the U-Net and LSTM models.

To compute the single variable importance, each input variable was randomized along the number of samples dimension for the test dataset only. Meaning a physically coherent 2-D variable field is used at a mismatched time in the model predictions. The permuted (randomized) input variable with the greatest change in the area under the precision recall curve is deemed to have the greatest importance. A limitation of permutation importance is that the input variables are slightly correlated, making it difficult for single-variable permutation to explicitly predict uni-variate importance. Thus, a second form of interpretability is introduced, via partial dependence plots (PDPs) for the hit- and missed-forecast case studies.

The partial dependence plots vary the model inputs based on the values seen in the test dataset, while monitoring the model output (the probability of lightning), at specific locations - Wright-Patterson AFB on 12 UTC 13-14 June 2022 (a forecast hit) and Vance AFB on 12 UTC 09-10 December 2022 (a forecast miss). These plots were adopted from Friedman (2001) and Trok (2023) to accommodate the higher-dimensional datasets used herein. Given both the U-Net and LSTM model are convolution-based, the spatial variability is explored by varying the inputs in the pixels surrounding Vance and Wright-Patterson, in multiples of the convolution size. For the 1-kernel example, 4 grid cells to the north, south, east, and west of the grid point of interest were varied - 8 grid cells for 2 kernels, and 12 grid cells for 3 kernels. The lightning probability is monitored only for the Day 1 prediction.

3.2 Results

3.2.1 Performance Diagrams and Permutation Importance

Figure 3.5 shows the reliability diagrams for the U-Net and LSTM (panels a and b) in the top row, accompanied with histograms of the U-Net and LSTM (panels c and d) predictions in the bottom row. Note, the y-axis is on a log-scale. A model is deemed reliable when the lines follow the perfect forecast line, over-forecast when the lines fall below the perfect forecast line, and over-forecast when the lines fall above the perfect forecast line. Both the U-Net and LSTM have high reliability and skill, given the reliability lies close to the perfect forecast line. The U-Net boasts slightly better Brier Scores of 0.051 on Day 1 increasing to 0.058 on Day 4, while the LSTM model has a similar trend increasing from 0.052 on Day 1 to 0.060 of Day 4. All forecast days trend along the perfect forecast for predicted probabilities (x-axis) less than 50. Above 50 percent, a trend towards under-forecasting is observed as all forecast days for the U-Net and LSTM dip below the perfect forecast line. The LSTM model dips below a skillful forecast for probabilities greater than 90 percent, on Day 4. This reliability skew is also shown in the histograms in panels (c) and (d), where there are more probabilities generated less than 50 percent, than above. The solid grey line is the no-skill line, and we see that each point above 10 percent probability is above the no-skill line for the U-Net. The LSTM model has reliable skill through day 3, for all forecast probabilities, and up to 90 percent forecast probability for Day 4. Beyond Day 4's 90 percent forecast probability, no skill is added.

The precision-recall curves are provided in Figure 3.6 - the panels (a) pertaining to the U-Net and (b) the LSTM models, for all rotations and seasons. CSI values of 1 are ideal, and all points should strive to be in the upper right hand corner, in the

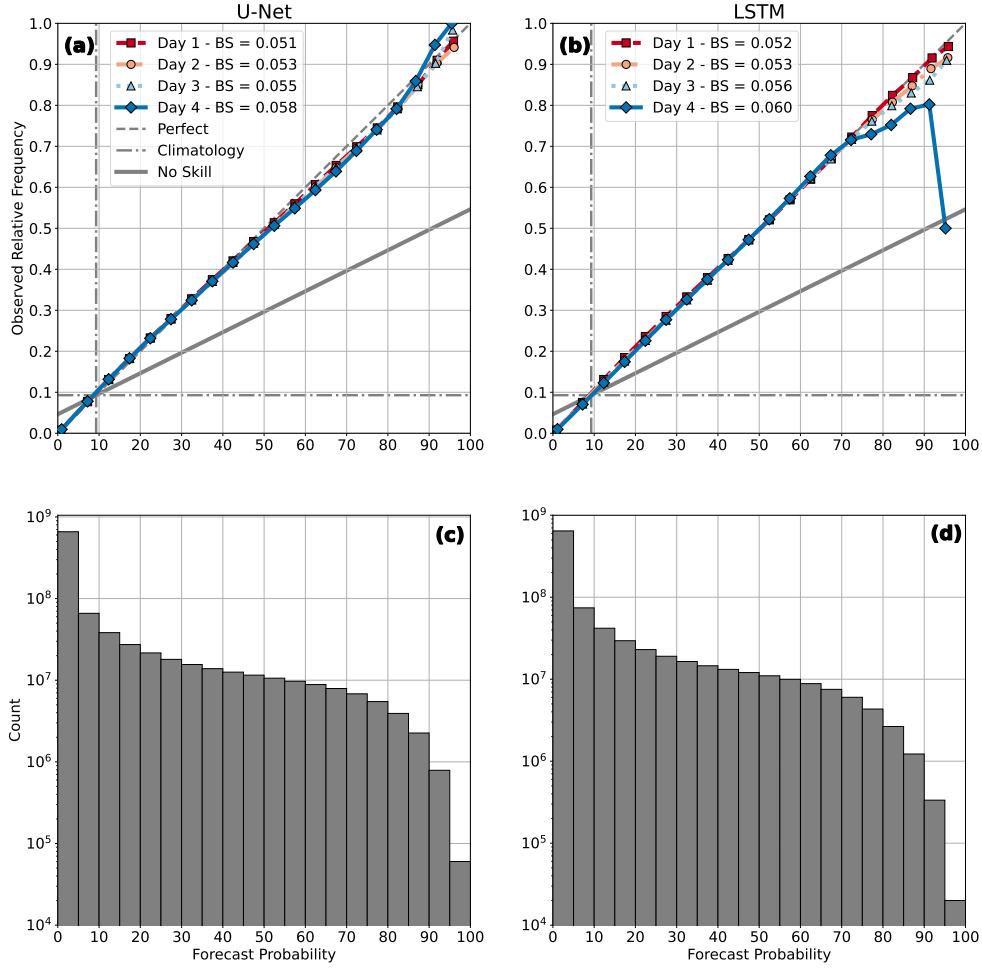


Figure 3.5: The top row denotes the reliability diagrams for the U-Net (a) and LSTM (b) architectures, for all 5 rotations' test datasets model outputs. Each line pertains to a specific forecast day denoted in the legend, with the Brier Score accompanying the label. The bottom row shows a histogram of the forecasted lightning probabilities for the U-Net (c) and the LSTM (d), with the y-axis on a log scale.

dark grey shading. For clarity, the maximum CSI, its accompanying threshold, and the PR-AUC are provided in Table 3.5.

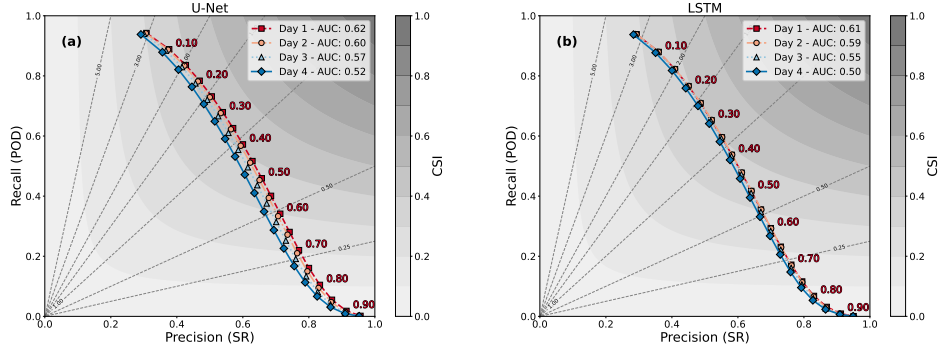


Figure 3.6: Precision-recall performance diagrams for the U-Net in Figure (a) and LSTM (b) architectures. The x-axis is the precision and the y-axis is the recall. Grey-shaded contours are for critical success index (CSI), and the dashed diagonal lines denote frequency bias. Each line pertains to one prediction day, given model predictions for the test datasets of all 5 rotations listed in Table 3.3. PR-AUC statistics are included in the legend and Table 3.5, where the maximum CSI and corresponding thresholds are also listed, for clarity. Threshold probabilities for Day 1 are included in the red text. Days 2-4 probability thresholds are omitted for clarity.

A downward and leftward shift in the curves of Days 2-4 in panels (a) and (b) indicates decreased model performance with increasing NWP model lead time at all thresholds. Quantitatively, the U-Net model PR-AUCs decreases from 0.62 at Day 1 to 0.52 at Day 4. Similarly, the LSTM model PR-AUCs decreases from 0.61 to 0.50 from Day 1 to Day 4. The maximum CSI's all reside at the 0.30 and 0.25 threshold for the U-Net and LSTM respectively, as indicated in Table 3.5, with similar decreases in values: 0.43 (Day 1) to 0.40 (Day 4) for the U-Net and 0.43 (Day 1) and 0.39 (Day 4) for the LSTM. The attribute and performance diagrams provide metrics for the model as a whole, but do not provide any spatial or regional details.

To understand regional skill variation, the PR-AUC metric was calculated for each GFS grid point included in model training, to examine regional dependencies across CONUS (Figure 3.7). When visually examining the U-Net and LSTM columns, for

Model	Lead Time	Max CSI	PR-AUC
U-Net	Day 1	0.43	0.62
	Day 2	0.42	0.60
	Day 3	0.42	0.57
	Day 4	0.40	0.52
LSTM	Day 1	0.43	0.61
	Day 2	0.42	0.59
	Day 3	0.41	0.55
	Day 4	0.39	0.50

Table 3.5: Explicit metrics highlighting the maximum CSI and corresponding threshold, and PR-AUCs for the U-Net and LSTM model, for the specified forecast day. The PR-AUC values match those in the legend in Figure 3.6. Values are calculated across model outputs from the test datasets, from all rotations listed in Table 3.3.

the different forecast days (rows), few meaningful differences are discernible. Terrain features seem to indicate higher values of PR-AUC in both models, indicated by the bright yellows in the southern Rocky Mountains and western Mexico. The Florida peninsula shows higher PR-AUC values, than the surrounding southeastern US. Low PR-AUC is observed over the eastern Pacific and along the western United States coast.

To understand the differences in the metrics a difference plot is provided in the right column, with hatching denoting where the model outputs are statistically different (via t-test), with p-values less than $1.6\text{e-}6$ using the significance test threshold outlined by Wilks (2016), $\alpha = 1 - (1 - \alpha_0)^{1/N_0}$. Here α_0 is set to 0.05 and N_0 is set the number of spatial pixels, 32768; this performs a t-test for each pixel in the image. The hatching is present across the west coast, California, Nevada, Mexico; a small area along the Colorado Front Range and upper Great Lakes; followed by a continuous swath along the Louisiana, Mississippi, Alabama, and Florida coastlines indicating the model output differences between the U-Net and LSTM are statistically significant in those regions. The Appalachian region through New England have larger swaths of blue color for all forecast days, indicating a greater LSTM PR-AUC in these regions. In contrast,

the mid-west and western United States have more warm colors near 0.05 indicating slightly higher PR-AUC values for the U-Net model. Overall, the differences between the U-Net and LSTM are minimal throughout CONUS. Similarly, maps of precision and recall for a 0.25 threshold are included in Figures 3.8 and 3.9. Precision and recall show similar spatial patterns and statistical significance as the PR-AUC, with most significant differences hugging the Sierra Nevada mountain range in California, near the Great Lakes, and the oceans. Otherwise, across the Great Plains and northeastern United States, there were no significant differences between the U-Net and LSTM models.

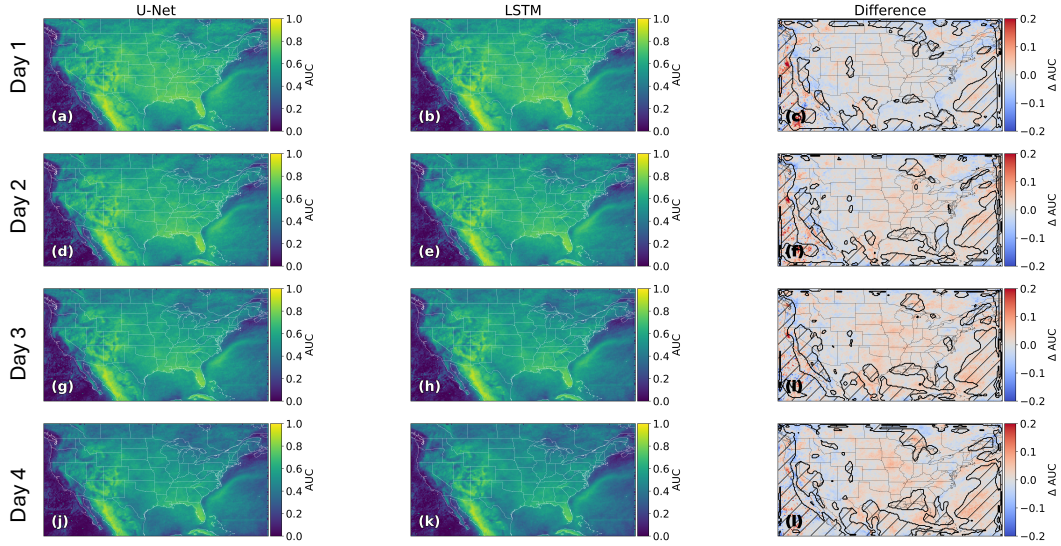


Figure 3.7: Plots of PR-AUC across the BoltCast domain. Diagram for the U-Net (panels a,d,g,j), and LSTM (panels b,e,h,k) and the difference (U-Net - LSTM) plotted in panels c,f,i,l. Dashed lines indicate areas where the model outputs were statistically different, with a p-value less than $1.6e-6$. Similar plots for recall and precision for a threshold of 0.25 are provided in Figures 3.9 and 3.8.

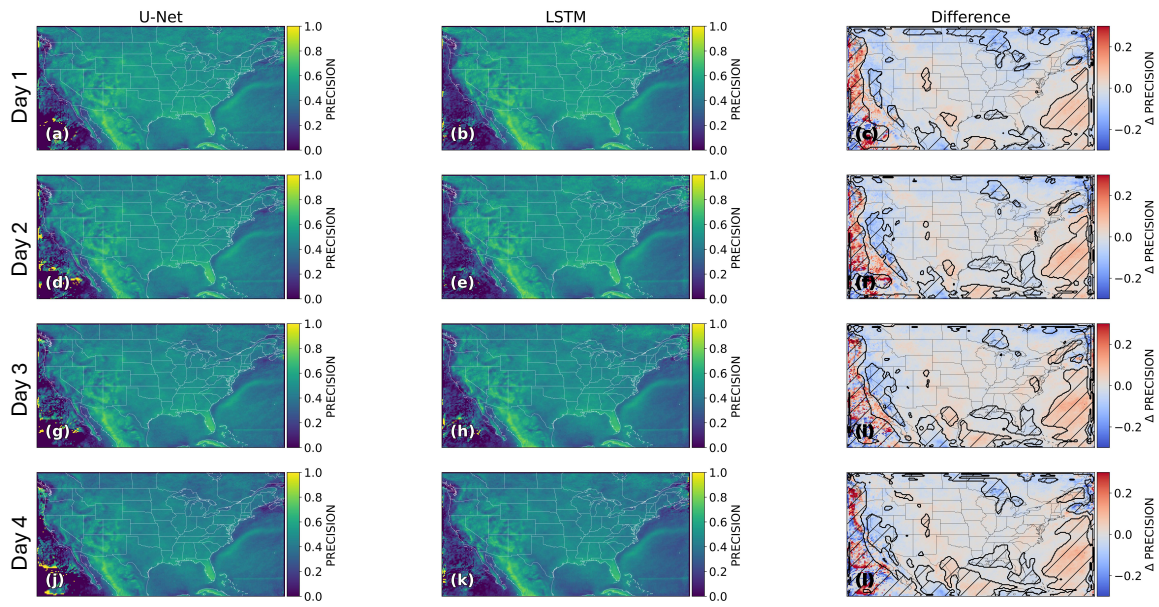


Figure 3.8: Map of precision values for a threshold of 0.25 across the model domain.

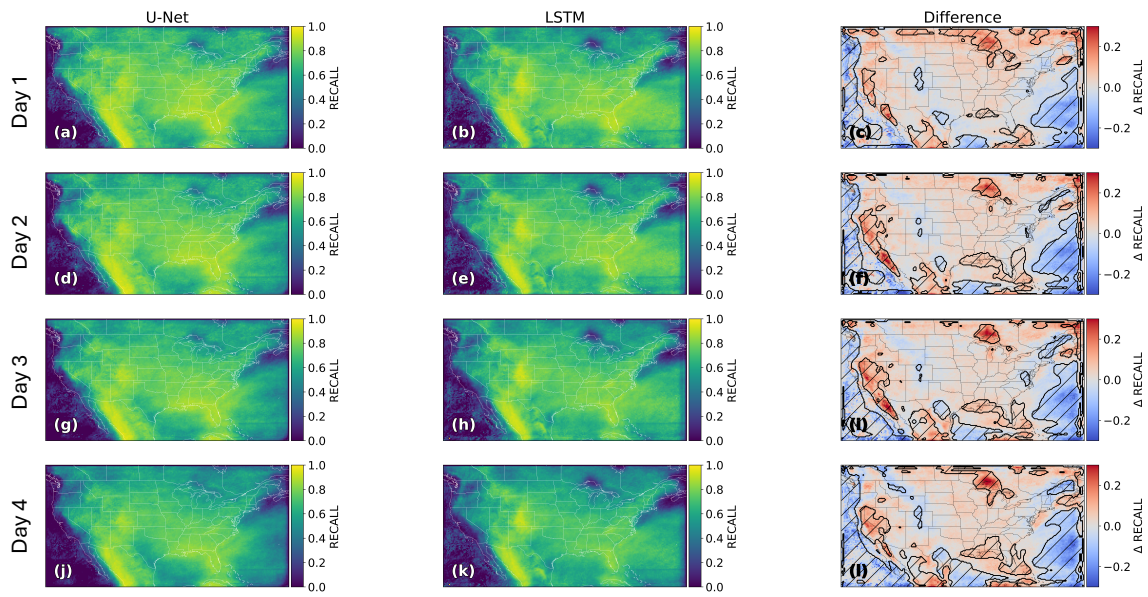


Figure 3.9: Map of recall values for a threshold of 0.25 across the model domain.

Before examining the seasonality and case studies of model performance, I examined the results of permutation importance to better understand which input features matter most to the forecast. Figure 3.10 depicts variability of the area under the precision-recall curve (PR-AUC) for 1000 different permutations, for each of the nine input variables, across all five rotations. Here, the change in the PR-AUC is termed the importance of the variable, to be consistent with AI/ML terminology presented in other XAI literature (Flora et al., 2024; McGovern et al., 2019).

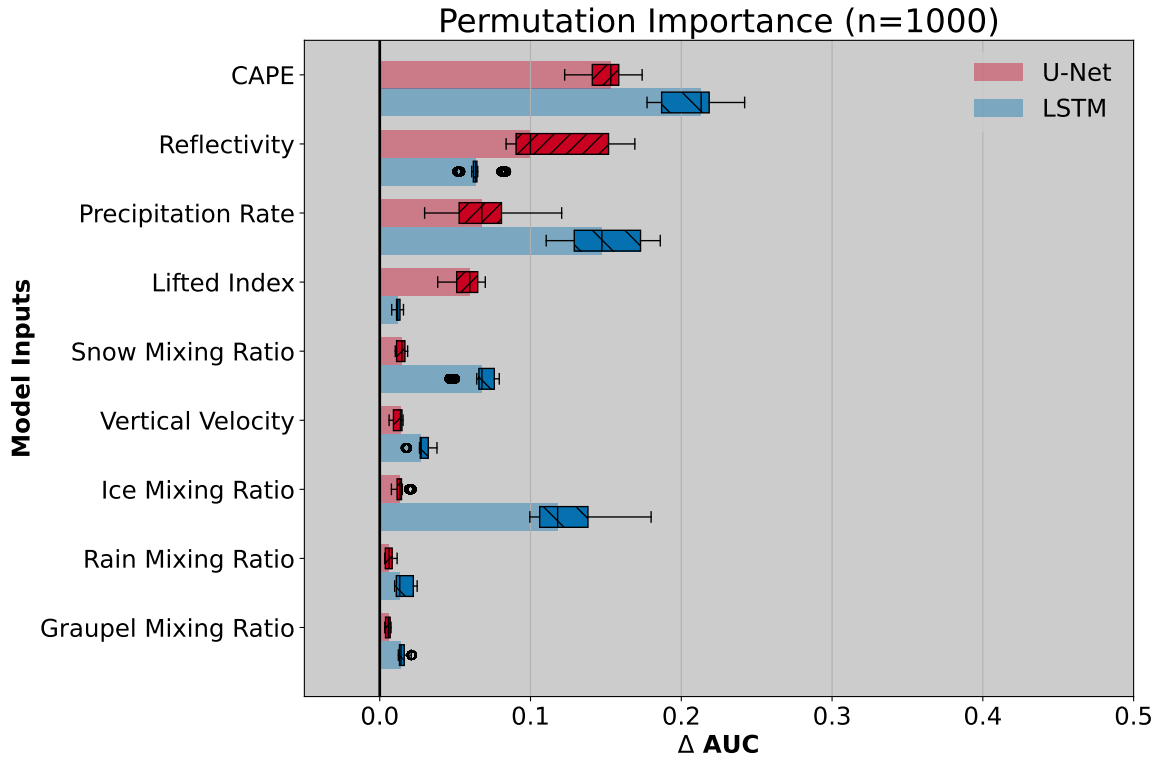


Figure 3.10: Grouped-permutation importance for 1000 permutations for the 0.25 threshold, ranked according to the U-Net median. The x-axis is the decrease (change) in the area under the precision-recall curve, and the y-axis lists the input variables. Box-whisker plots show the variability and medians (box-center) across all rotations and permutations. The red bars pertain to the U-Net while the blue bars pertain to the LSTM model.

The greatest changes in PR-AUC accompany the CAPE, reflectivity, and precipitation rate for the U-Net. CAPE, precipitation rate, and ice mixing ratio accompanied

the greatest change in PR-AUC for the LSTM model. Figure 3.10 also depicts higher importance for all hydro-meteor mixing ratios for the LSTM model. The coincidence of CAPE and precipitation rate aligns well with meteorological intuition, that suggests lightning and thunderstorms require instability, which in turn generates high reflectivities and precipitation rates. The maximum hydrometeor mixing ratios did not generate large changes in PR-AUC when permuted for the U-Net, implying the U-Net favors instability and precipitation over specific hydrometeor mixing ratios. More modest changes in PR-AUC for the LSTM model, imply the LSTM model favors the mixing ratios more than the U-Net. These results could also be a by-product of only a binary classification, vice flash rates, where the hydrometeor mixing ratios have been used in flash rate parameterization schemes prior to this work (Cummings et al., 2024). Permutation importance can also be biased with correlated inputs, hence partial dependence plots are provided in the next section. The results up to this point are conducted over all seasons; results for specific seasons are presented next.

The PR-AUC is computed for the U-Net and LSTM architectures subset on Summer, Fall, Winter, and Spring seasons, shown in Figure 3.11. Panel (a) depicts the PR-AUC for the U-Net model and panel (b) for the LSTM model. Summer sees the greatest model performance with higher PR-AUCs than any other season. For the U-Net, the PR-AUCs decrease from 0.66 to 0.56 in the summer, 0.52 to 0.43 in the fall, 0.49 to 0.39 in the winter, and 0.62 to 0.50 in the spring. The LSTM PR-AUCs decreased from 0.66 to 0.56 in the summer, 0.52 to 0.43 in the fall, 0.49 to 0.39 (increases in Day 2 and 3) in the winter, and 0.62 to 0.50 in the spring. These results are promising, given the higher PR-AUC scores during the summer for both the U-Net and LSTM that follows the general trend of convective activity across the CONUS from season to season. A similar graph for maximum CSI values is provided in Appendix 3.12. With a thorough understanding of the PR-AUC scores spatially and seasonally,

a few case studies are presented in the next section to demonstrate how the model handles the GFS data for cases where lightning was a forecast hit and miss.

3.2.2 Case Studies and Partial Dependence Plots

Figures 3.13 and 3.15 show example model output for cases where lightning was forecast as a hit and miss, respectively. I also include a comparison with the Storm Prediction Center’s 24-hr HREF calibrated thunder guidance (a cloud-to-ground lightning forecast tool) for the Day 1 and Day 2 predictions (Harrison et al., 2022) and a simple convective parameterization mask for the GFS maximum composite reflectivities in excess of 30 dBZ at Days 3 and 4, a documented convection initiation threshold (Wilson and Schreiber, 1986). A threshold of 30 dBZ reflectivity has been documented to outperform lightning flash rate schemes (Xu et al., 2010), but with a potential over-forecast bias (Huang et al., 2024). Additionally, 30 dBZ aloft, at both -10 degrees Celsius and -20 degrees Celsius, is used extensively to predict lightning occurrence (Ortland et al., 2023); thus a 30 dBZ mask is a justified tool for predicting lightning from the GFS data. The HREF thunder guidance was chosen due to data availability (Harrison, personal communication), while SREF data was not publicly available.

The hit-forecast case (Fig. 3.13) is valid between 12 UTC 13 June 2022 and 12 UTC 14 June 2022. At the beginning of the forecast window, the day-4 PR-AUCs are 0.44 and 0.47 for the U-Net and LSTM models, respectively. The U-Net PR-AUCs increase to 0.66 for the day 1 forecasts, with a slight dip in the PR-AUC observed from 0.59 in day 3 to 0.57 in day 2. The LSTM model shows a similar dip from 0.55 in day 3 to 0.51 in day 2, finally climbing to 0.64 in day 1. The high probabilities, greater than 50 percent, do show spatial variability across the northern tier states through the Great Lakes regions for both models. The higher probabilities in western Mexico (coincident with the high PR-AUCs in Figure 3.7) are fairly steady state, with little

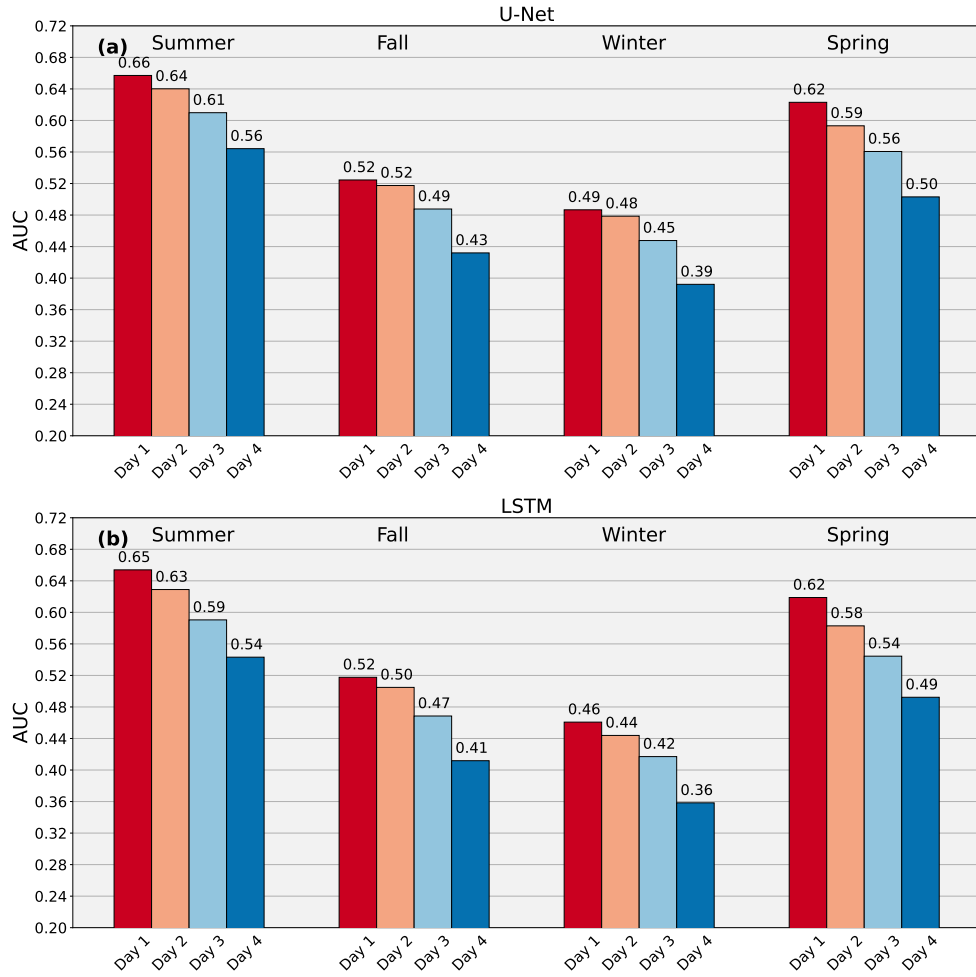


Figure 3.11: PR-AUC bar plots for all 5 rotation's test dataset model outputs, for each season. The U-Net, is in panel (a), and the LSTM, panel (b). Red box bars pertain to Day 1, orange Day 2, light-blue Day 3, and dark-blue Day 4.

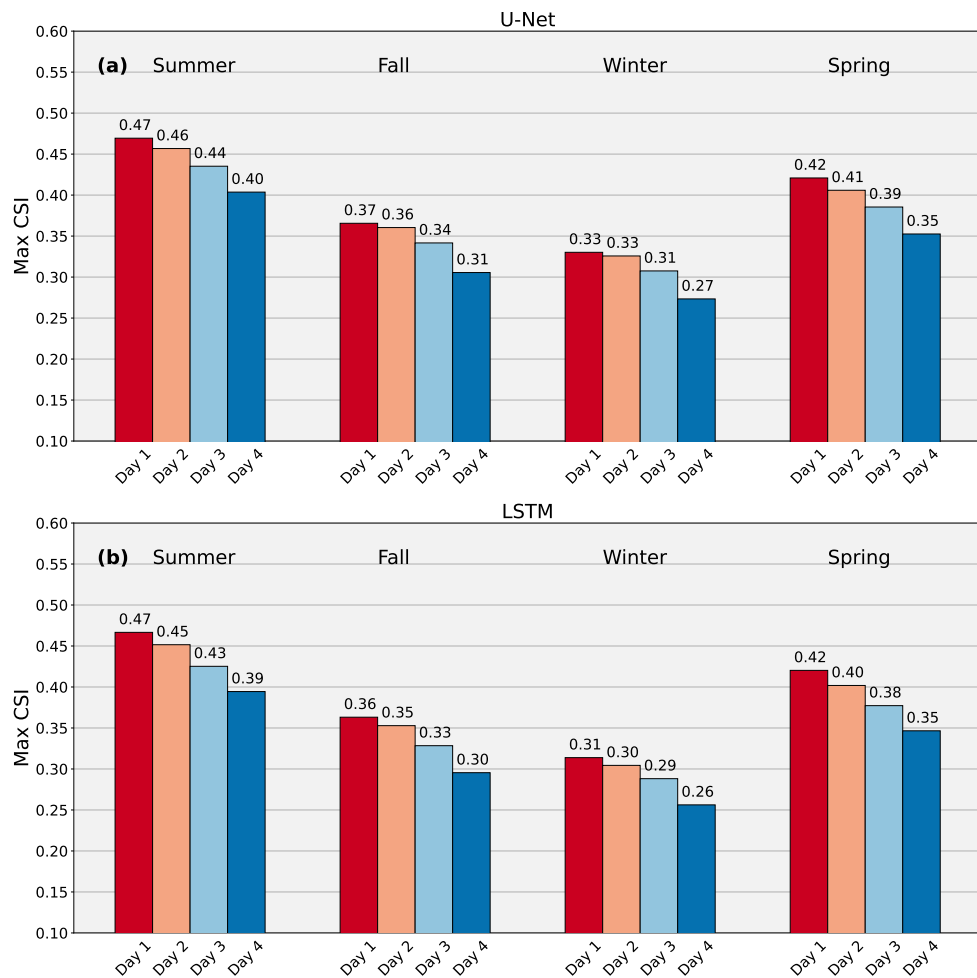


Figure 3.12: Bar graph of seasonal maximum CSI values for the U-Net (top) and LSTM (bottom).

variability across lead times.

The Day 3 and 4 predictions for the U-Net and LSTM models eliminate the convection potential in the Pacific Northwest indicated in the reflectivity mask which was not observed in the labels. The Day 1 and Day 2 HREF (CG-only) comparisons in the bottom row, show smaller spatial variability for probabilities greater than 50 percent. Given good U-Net and LSTM agreement with the labels across CONUS, let's try to understand what drove the high lightning probabilities over WPAFB (located at the magenta dot), by varying the model inputs at and around that location and monitoring the response of the model output.

The partial dependence plots for the U-Net's three most important variables determined by permutation importance (CAPE; panel (a), Reflectivity; panel (b), and Precipitation Rate; panel (c)) are shown in Figure 3.14. Reminder, that only one GFS data input is perturbed at a time while others retain observed values within the convolutional neighborhood. At WPAFB the lower CAPE values produce lower lightning probabilities for both the U-Net (solid) and the LSTM (dotted). LSTM probabilities drop to near 60 percent when CAPE values approach zero for the largest 3-kernel/12-pixel spatial modification, while the U-Net probabilities remain above 80 percent for all variations. This indicates for the lower CAPE values that the LSTM model is more sensitive to the surrounding CAPE values than the U-Net. As CAPE values increase all spatial modifications converge to near 90 percent, with minimal difference. A similar trend is observed when larger spatial modifications tend to decrease the lightning probabilities for decreasing reflectivity values approaching -20 dBZ (a GFS model artifact) and precipitation rate approaching zero. However, the LSTM model showed less spatial variability across all of the kernel sizes with lightning probabilities never dropping below 75 percent for all reflectivity and precipitation rates. The U-Net variability dipped to 90 percent for lower reflectivities and precipitation rates for 3-kernel

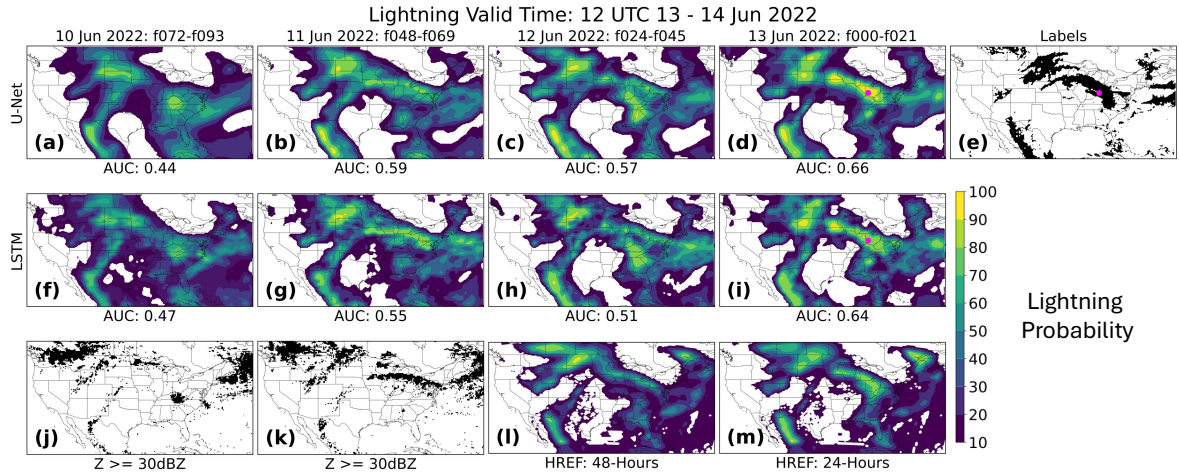


Figure 3.13: Example model output from rotation 4's test dataset, valid between 1200 UTC 13 June 2022 - 1200 UTC June 14 2022. Panels (a) through (d) [top row] are predictions generated by the U-Net architecture. Panel (e) is the lightning labels. Panels (f) through (i) [middle row] are predictions generated by the LSTM layer. Panels (j) and (k) are generated by a 30 dBZ mask of the predicted reflectivity for Days 3 and 4; panels (l) and (m) are the 48-Hour and 24-Hour SPC HREF calibrated thunder product. The magenta dot in panels (d), (e), and (i) denote the location of Wright-Patterson AFB; used in generating the partial dependence plots in Figure 3.14

spatial modification. It is important to note that input variations are only computed for one variable at a time. Thus, the values of the eight non-permuted variables are contributing to the higher lightning probabilities. Additional partial dependence plots are provided in Supplementary Materials (e.g., Figure 3.17).

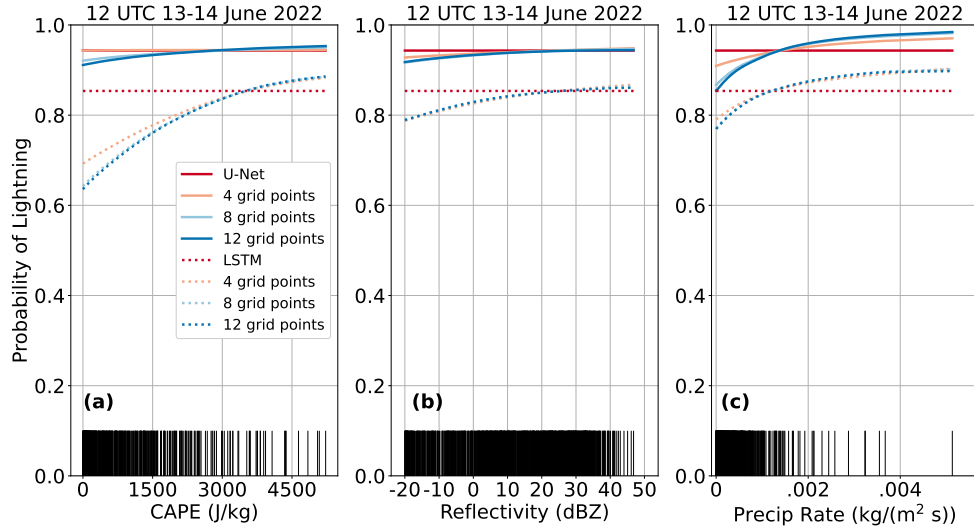


Figure 3.14: Partial-dependence plots for both the U-Net (solid lines) and LSTM (dotted) predictions, for Wright-Patterson AFB for lightning valid between 12 UTC 13 June 2022 and 12 UTC 14 June 2022. Each different color depiction is a different kernel-based window sized, generated on the input image for each Day 1 CAPE (left column), Reflectivity (middle column), and Precipitation Rate (right column). Rug plots along the varying x-axes denote the density of CAPE, Reflectivity, and Precipitation Rate.

The missed-forecast case in Figure 3.15 is valid between 12 UTC 09 December 2022 and 12 UTC 10 December 2022. This date is chosen to show a case when both the U-Net and LSTM are poorly performing, for transparency. Quantitatively, the U-Net and LSTM achieved a maximum PR-AUC of 0.22 and 0.27 on day 2. The U-Net and LSTM hint that lightning is possible with lower forecast probabilities throughout the forecast window, but high probabilities never emerge. Whereas, the cloud-to-ground calibrated HREF guidance generates high probabilities over central Oklahoma, near 73 percent, across both days. Additionally, the reflectivity mask for Days 3 and 4 suggests

convection and lightning potential over the central US. However, the reflectivity mask performs poorly over the Great Lakes, generating lightning where no lightning was observed. The partial dependence plots for Vance AFB, the magenta dot in Figure 3.15 where lightning was missed, is presented in Figure 3.16 to try to understand why the lightning was missed.

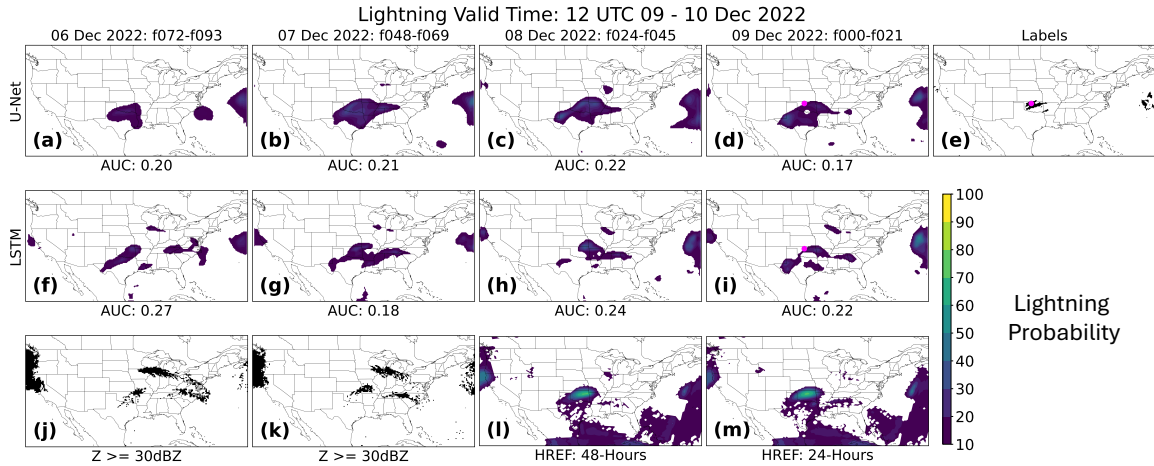


Figure 3.15: Example model output from rotation 4, valid between 1200 UTC 9 December 2022 - 1200 UTC 10 December 2022. Panels (a) through (d) [top row] are predictions generated by the U-Net architecture. Panel (e) is the lightning labels. Panels (f) through (i) [middle row] are predictions generated by the LSTM layer. Panels (j) and (k) are generated by a 30 dBZ mask of the predicted reflectivity for Days 3 and 4; panels (l) and (m) are the 48-Hour and 24-Hour HREF calibrated thunder product. The magenta dot in panels (d), (e), and (i) denote the location of Vance AFB; the location used in generating the partial dependence plots in Figure 3.16.

Greater variability is seen at VAFB than at WPAFB for CAPE and precipitation rate (Fig. 3.16a,c). Lightning probability remains unchanged and near zero (dark red horizontal curve) for the single pixel modifications, but there are noticeable increases

in predicted probabilities when the spatial area is increased, for both the U-Net and the LSTM. The lower probability bias discovered in the WPAFB case is also present in the VAFB case. The predicted probabilities remain near zero for all variations of reflectivity in Fig. 3.16b, indicating that this specific case is not dependent on reflectivity. This may be due to the nature of winter-time lightning, occurring within shallower convective cells, not able to generate higher reflectivity rates.

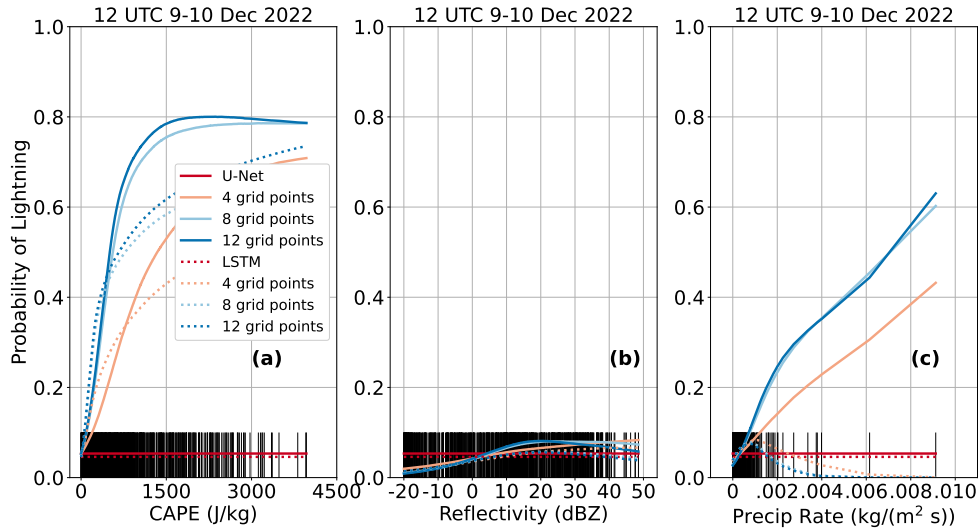


Figure 3.16: Same as in Figure 3.14, but for Vance AFB.

Generally, for both the WPAFB and VAFB cases, the larger spatial variations drove a higher amplitude response in the lightning probabilities than the smaller spatial variations. Additional partial dependence plots are generated and provided in the Appendix (Figure 3.17) for completeness. These partial dependence plots also continue to present a low-probability prediction bias in the LSTM model, meaning the LSTM model struggles to generate predicted probabilities over 90 percent, with the exception of snow mixing ratio. The next section provides an overview of these results and provides further discussion of the U-Net and LSTM architectures.

This chapter demonstrated that both the U-Net and LSTM models generate skillful

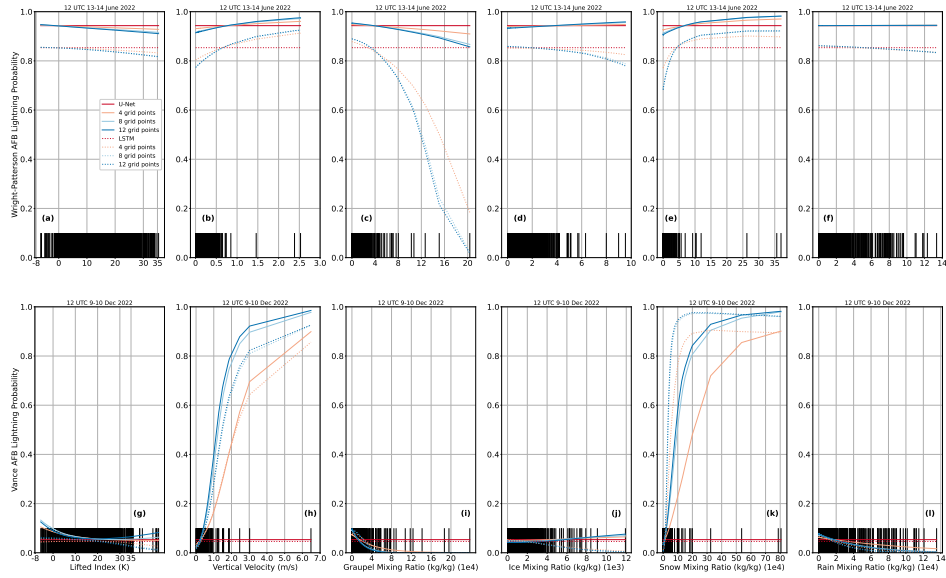


Figure 3.17: Partial dependence plots of the GFS data not presented in Figures 3.14 or 3.16.

predictions of lightning out to four days across CONUS. The next chapter transitions into a nowcasting role, presenting the skill for lightning prediction at 15-minute intervals out to 1-hour, over Kennedy Space Center with a unique suite of EFMs.

Chapter 4

LaunchCast: Nowcasting Lightning at the Meso-scale

This chapter reviews the development of a lightning nowcasting model over Kennedy Space Center and the adjacent military complexes. It describes the model data used as inputs, the lightning target generation, and the relevant metrics presented in Chapter 2. It concludes with an explainability analysis using accumulated local effects.

4.1 Data and Methodology

This section provides an overview of the data and methods applied to generate the results in the following section. Data input and target generation are discussed, followed by explicit details describing how each data sample is generated. This section concludes with a discussion of the model architecture and metrics used for evaluation and explainability.

4.1.1 Inputs and Binary Lightning Targets

Both observational and model data are used as inputs into LaunchCast (LC) (Table 4.1). The Multi-Radar Multi Sensor (MRMS) (Lakshmanan et al., 2016; Howard et al., 2016) quality-controlled isothermal reflectivity, vertically integrated ice (VII),

and differential reflectivity Z_{DR} are used as inputs. Isothermal reflectivities are known to have predictive capability for total lightning (Gremillion and Orville, 1998; Schumacher et al., 2012). Thus, $0^{\circ}C$, $-10^{\circ}C$, and $-20^{\circ}C$ isothermal reflectivities are used. $-5^{\circ}C$ and $-15^{\circ}C$ isothermal reflectivities are not used to decrease correlation among the reflectivity inputs. High values of vertically integrated ice (VII) coincident with regions of strong reflectivity indicate mixed-phase hydrometeors are present, and conducive to electric charging (Gremillion and Orville, 1998). Finally, Z_{DR} is included as columns of positive Z_{DR} are correlated with lightning occurrence (Zhao et al., 2024). All MRMS variables were re-gridded to the coarser 3-km, lambert-conformal HRRR grid.

GLM group area and energy are known to help discriminate IC and CG flashes detected on ground-based systems (Ringhausen et al., 2021). GLM group area is included as an input into the model to depict where lightning occurred during the previous hour. The groups were gridded to the 3-km HRRR grid, where the maximum group area was taken if multiple groups existed in the same grid cell. Group energy is not included due to its correlation with group area.

The vertical component of the electric field (potential gradient) is recorded at 50 Hz (0.02 sec). 26 of the 31 static EFMs are used in this study, as 5 EFMs were either listed as inactive, or had more than 1-year of data missing. Their locations are shown in Figure 4.2a, with sample 50 Hz data shown in 4.2b. These data were downloaded from the KSC Weather Archive and resampled at a 30-second interval (notionally, 1500 observations). 6 different statistics are calculated across the 30-second re-sampled interval and used as feature inputs: mean, standard deviation, variance, number of observations where the electric field magnitude was greater than 1000 V/m (informed by LLCC Rule 7), root mean square, and maximum absolute value. If fewer than 500 of the 1500 nominal observations are available, then a mask input value is assigned

Table 4.1: The LaunchCast model input variables and targets.** denotes variable is post-processed. Low Z_{DR} pertains to elevation angles 0.00 - 2.75 degrees at 0.25 degree resolution; mid from 3.00-9.50 degrees at 0.50 degree resolution, and high is from 10.0-18.0 degrees at 1.00 degree resolution. All observational inputs are for the previous hour. For time step examples, see Table 4.3.

Observation Inputs	Parameter	Units
MRMS	0°C Isothermal Z	dBZ
	-10°C Isothermal Z	dBZ
	-20°C Isothermal Z	dBZ
	VII	kgm^{-2}
	Low Z_{DR}	dB
	Mid Z_{DR}	dB
	High Z_{DR}	dB
GLM	Group Area	m^2
50 Hz Potential Gradient at 26 static sites Resampled at 30 seconds	Mean	Vm^{-1}
	Standard Deviation	Vm^{-1}
	Variance	$V^{1/2}m^{-1/2}$
	No. $ E > 1000 V m^{-1}$	Unitless
	Root Mean Square	Vm^{-1}
	Maximum Absolute Value	Vm^{-1}
	Station Weighted Distances	<i>Unitless</i>
Model Inputs	Parameter	Units
Sub-hourly HRRR	Divergence**	s^{-1}
	2m Temperature	K
	2m Dew-point	K
	Mean Sea Level Pressure	hPa
	Precipitation Rate	$kgm^{-2}s^{-1}$
Target	Parameter	Type
MERLIN and MRMS	Total Lightning with $Z \geq 30 dBZ$ at 0°C	Binary Classification

to each input feature/statistic - a novelty of this work. An additional novelty is the exponentially weighted station horizontal proximity proxy for each of the 26 static EFMs. The distance was calculated to each target grid cell from each site, followed by an exponential decay of the distances with a decay factor set to 5-nautical miles, informed by the LLCC Rule 7. The horizontal proximity is a limitation provided the electric field is a function of the Euclidean distance which accounts for the charge altitude.

The final input suite is the sub-hourly HRRR (Dowell et al., 2022) features listed in Table 4.1. Convergence is a key parameter KSC forecasters use in their workflow (Smith, 2025) and was used in early explorations of CNNs’ applicability to predict CG lightning strikes (Wilson and Schreiber, 1986; Barnes and Frankel, 1990; Barnes, 1991). It is calculated from the 10-meter u- and v-wind components. Two meter temperature/dew point, mean sea level pressure, and precipitation rate capture the sea-breeze circulation and any predicted precipitation, generated by warmer surface temperatures observed over land. These data were chosen to mimic gridded wind tower observations that are present across KSC.

The MERLIN lightning location system is based on 10 Vaisala TLS200 sensors (Roeder, 2017). It discriminates between in-cloud (IC) and cloud-to-ground (CG) sources. Sources are converted to flashes if the sources occur within 300ms of each other. The latitudes and longitudes of the sources are averaged to provide the location of the flash. The flash climatology of IC and CG flashes is provided in Figure 4.1. I observe a peak of lightning flashes in both IC and CG flashes at 21Z, and in the month of August. Minimal flashes occur during the winter months of December, January, and February. These months are not included in LC model training, analysis, or evaluation. The bottom row depicts the spatial distribution of all flashes observed on the MERLIN system. A distinct bias is observed with more flashes occurring over land than over the

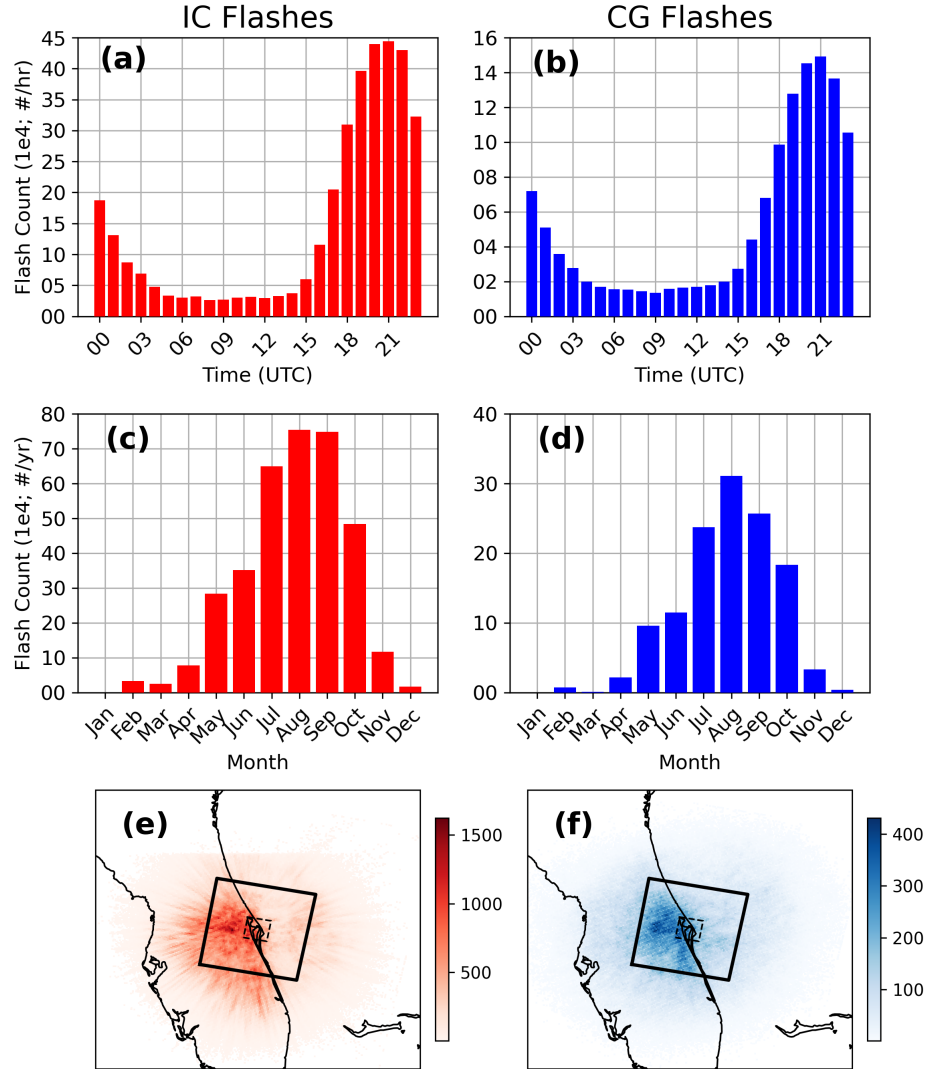


Figure 4.1: Climatology of MERLIN flashes from 2021-2024. The left column is for IC flashes and the right column is for CG flashes. The black boxes in the bottom row denote the model input domain (solid), a 64x64 image on the HRRR grid, and the target domain a 16x16 image on the HRRR grid.

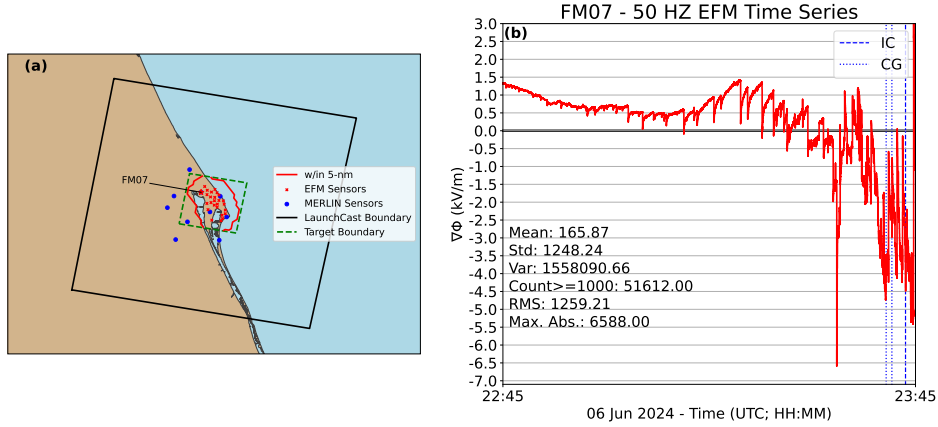


Figure 4.2: (a) A zoomed in picture of the input domain, the target domain, and the 5-nautical mask required for target generation. Additionally, the location of the EFMs (red x's) and the MERLIN sensors (blue dots) are provided. FM07 is starred for future reference throughout the document. (b) Sample 50 HZ EFM time series 2245 - 2345 UTC 06 Jun 2024 from FM07 (inputs to case 1 in Figure 4.6) example statistics.

water. The black box (solid) denotes the LC model input domain, ensuring adequate coverage over the water to capture the land-sea contrast, sea-breeze circulation, and of the local IC and CG flash maxima. The dashed box indicates the model target domain, a 16x16 image. The larger input domain and the target domain are both centered on the mean latitude and longitude of the 26 static EFM sites used as inputs for this model. A zoomed in layout of the MERLIN sensors, the EFM locations, and the input and target domains are provided in Figure 4.2.

4.1.2 Sample Generation

Training data are generated by sampling from lightning onset, continuation, and ending within the inner target model domain. Sample times are filtered when there are 5 or more flashes or groups on the MERLIN target or GLM input domain. Then the samples are selected for targets to be within 5 nautical miles of the EFM suite informed by LLCC Rule 7. Table 4.2 shows the target flash count and type that are used to generate

the samples, with the subsets of the onset (no GLM groups present in the previous hour with lightning occurring within 5-nautical miles of the field mill suite) and end of lightning (1-3 continuous target forecast windows had zero lightning) documented. Note, this model is biased towards predicting the continuation of lightning, provided the small sample counts listed in Table 4.2 for the onset and end of lightning. This bias is due to convective pre-conditioning with GLM group area as inputs. Other sample generation schemes were examined (with higher model performance), but they failed to include the hard operational problem of the onset and decay of lightning which can shorten launch windows or cause misses in observed lightning warnings. Two qualitative case studies are provided highlighting the model performance during the onset and end of lightning. Additional initiation and cessation evaluations are not developed further, given that they only account for a small subset of the data.

Year	2021	2022	2023	2024
Type	(Train)	(Train)	(Val.)	(Test)
MERLIN IC Flashes	364,455	464,023	708,961	433,502
MERLIN CG Flashes	146,950	142,561	292,964	153,302
Number of Samples	4,380	4,286	5,084	4,022
Subset - Onset	80	141	114	208
Subset - Ended	132	171	158	159

Table 4.2: The number of samples and lightning flashes in each year, training (2021-2022), validation (2023), and test dataset (2024).

To mimic the data available to operational forecasters, a new training sample is generated every five minutes. Two sample observational input valid times, HRRR initialization times, and target valid times are shown in Table 4.3. This resampling method assumes zero latency in the sub-hourly HRRR data availability from its initialization time. Additionally, the 1-hour blocks of gridded observational data inputs and the MERLIN targets are broken into a time-series of 4 15-minute increments (i.e. the sub-hourly HRRR temporal resolution and LLCC rule 7 desired time). The EFM time

series are input into the model in 1-hour time series with length 120. To synchronize the isothermal reflectivities and VII (2-minute resolution) to the rolling observational windows shown in Table 4.3 the maximum values across the 15-minute input windows are taken. Additionally, the Z_{DR} data were at 5-minute resolution, for 31 different elevation angles. These data were binned into low elevations (0-2.75 degrees), mid elevations (3-9.5 degrees), and high elevations (10-18 degrees), and input as 3 separate features. The max Z_{DR} was calculated for each grid cell for the corresponding elevation angles, then the maximum of the available scans was taken across each 15-minute time window. All data were normalized using z-score normalization, where the mean was subtracted from the original input values and the result was divided by the standard deviation. Here, the mean and standard deviation were calculated across all samples from 2021-2024.

4.1.3 Model Architecture

The model architecture was implemented using Tensorflow v2.15.0 (Abadi and Agarwal, 2015). Figure 4.3 shows the model architecture diagram for all model inputs. Separate encoder branches are used for the NWP and observational data (dark blue blocks) that mimic the U-Net encoder used by McClung et al. (2026). Each block consists of two 3-D convolutional layers, followed by a pooling layer. Once the spatial and time dimensions are reduced to one, the NWP and observation encoders are concatenated to the EFM encoder. The EFM data are input into a shared LSTM layer (Hochreiter and Schmidhuber, 1997) where each EFM site is input as a separate time-series to minimize data dropouts due to quality control. Once the previous hours' EFM data is through the LSTM layer, the layer is expanded and multiplied by the exponentially weighted distances of each field mill to the target grid, to create a weighted image similar to Bao et al. (2022). This image then undergoes one 2-D convolutional block, before being

Table 4.3: Sample time slices for observational data, sub-hourly HRRR, and MERLIN target valid times in the year 2021. The [t0,t1,t2,t3] time designators are horizontally consistent with the time labels used in the input layers for line-by-line diagram provided in the GitHub repository in the data availability statement.

No.	Observational Data	HRRR Model
1	2105 - 2205 UTC 10 Apr 2021	2200 UTC 10 Apr 2021
Target	2205 - 2305 UTC 10 Apr 2021	
Input-t0	2105 - 2120 UTC 10 Apr 2021	2200 UTC 10 Apr 2021
Target-t0	2205 - 2220 UTC 10 Apr 2021	–
Input-t1	2120 - 2135 UTC 10 Apr 2021	2215 UTC 10 Apr 2021
Target-t1	2220 - 2235 UTC 10 Apr 2021	–
Input-t2	2135 - 2150 UTC 10 Apr 2021	2230 UTC 10 Apr 2021
Target-t2	2235 - 2250 UTC 10 Apr 2021	–
Input-t3	2150 - 2205 UTC 10 Apr 2021	2245 UTC 10 Apr 2021
Target-t3	2250 - 2305 UTC 10 Apr 2021	–
2	2110 - 2210 UTC 10 Apr 2021	2200 UTC 10 Apr 2021
Target	2210 - 2310 UTC 10 Apr 2021	–
Input-t0	2110 - 2125 UTC 10 Apr 2021	2200 UTC 10 Apr 2021
Target-t0	2210 - 2225 UTC 10 Apr 2021	–
Input-t1	2125 - 2140 UTC 10 Apr 2021	2215 UTC 10 Apr 2021
Target-t1	2225 - 2240 UTC 10 Apr 2021	–
Input-t2	2140 - 2155 UTC 10 Apr 2021	2230 UTC 10 Apr 2021
Target-t2	2240 - 2255 UTC 10 Apr 2021	–
Input-t3	2155 - 2210 UTC 10 Apr 2021	2245 UTC 10 Apr 2021
Target-t3	2255 - 2310 UTC 10 Apr 2021	–
...	...	...

concatenated to the NWP and observation encoders.

The bottleneck, after all three encoders are concatenated, uses a spatial convolution size dropped to (1,1) to ensure the weights are learned from all features, at each grid point. The decoder expands the time dimension into a 1-hour lightning prediction at 15-minute temporal resolution (informed again by LLCC Rule 7). The last layer uses a sigmoid activation function to generate the lightning probabilities. A grid and hyperband search was used to vary the learning rate, convolutional activation functions, spatial dropout, and L2 regularization (Li et al., 2018). The AUC-PR was maximized on the validation dataset, for the hyperband search. The final hyperparameters used are included in Figure 4.3. All training and tuning were performed using the Adam optimizer (Kingma and Ba, 2014), for all model inputs. These hyperparameters remain fixed throughout the results presentation, where model inputs are varied to determine model sensitivity.

4.1.4 Model Evaluation, Case Studies, and Explainable AI

Objective model evaluation is provided via attributes and precision-recall diagrams (Hsu and Murphy, 1986; Roebber, 2009). The attributes diagram documents under- and over-forecast biases by comparing observed probability to forecast probability, climatology, and no-skill (Hsu and Murphy, 1986). This paper highlights the Launch-Cast model attributes, depicting the four lead times separately, and provides the Brier Score, or mean-squared error (Equation 4.1) for each lead time. Here y_{true} is the binary label, and y_{pred} is the predicted probability, with N representing the number of predictions across space. Additionally, histograms of the predicted probabilities at the four-lead times are provided to document biases in the predictions. Further evaluation via precision-recall curves is provided to quantify model performance across probabilistic prediction thresholds. These curves are based on equations 4.2, 4.3, and 4.4 for

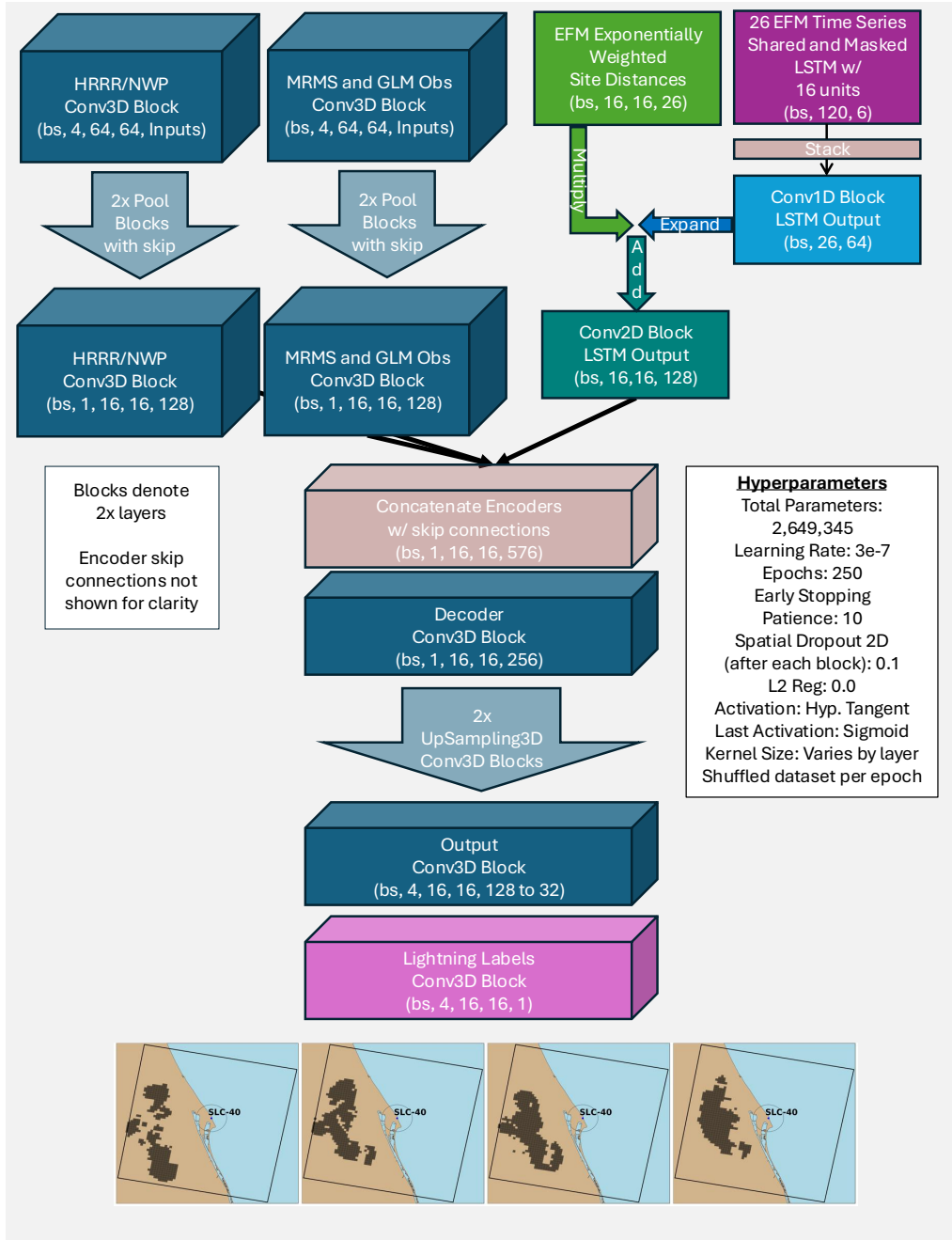


Figure 4.3: The architecture of the model built in Tensorflow. The last images represent the target domain (black box) and the location of Space Launch Complex 40 (SLC-40) for reference, accompanied by a 5 nautical mile range ring. Sample labels are provided (dark grey fill) that correspond to case study number 1 valid from 2245 - 2345 UTC 06 Jun 2024.

quantifying success ratio (SR), recall or probability of detection (POD), and critical success index (CSI), respectively. These metrics are dependent on the true positives (TP; hit), false negatives (FN; miss), and false positive (FP), where the model predicted an event that was not observed. These metrics were calculated for thresholds across 0 to 1 in increments of 0.05, using the Tensorflow Keras Metrics API (Abadi and Agarwal, 2015).

$$BS = \frac{\sum_{j=1}^N (y_{pred} - y_{true})_j^2}{N} \quad (4.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (4.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (4.3)$$

$$CSI = \frac{TP}{TP + FN + FP} \quad (4.4)$$

For each lead time, the maximum CSI and accompanying threshold are calculated and documented. One drawback from this approach is that these evaluations are pixel, or grid-point, based and do not capture near misses or spatially displaced hits. The fraction skill score (FSS) allows for near misses and hits to be accounted for by evaluating model performance given a pre-defined window size (Lean and Roberts, 2008).

Equations 4.5, 4.6, 4.7, and 4.8 describe the evolution of the FSS calculation where M is the number of grid points in the evaluation window, f_k is the binary threshold-ed forecast fraction, o_k the observed binary fraction, and f_0 the climatological mean. A baseline FSS is used to document the climatological skill in the FSS framework. In this paper, two evaluation windows are used of 10 nautical miles (LLCC limitation) and 5 nautical miles (Air Force weather lightning warning distance), with the FSS calculated at each lead time for the threshold associated with the maximum CSI. An additional evaluation is discussed, comparing to a 30-minute (informed by LLCC rule 6) binary lightning persistence forecast, derived from the GLM group area inputs. This persistence forecast is derived by generating a binary prediction across the target domain and predicted timesteps, where the GLM group area was greater than zero 30-minutes prior. An NWP comparison is also provided, by utilizing the HRRR forecast hour 1 and 2 (f01 and f02) lightning output. The binary mask created by Blaylock (2020) is applied and compared to the LaunchCast model output. A direct comparison to current operational human-derived forecasts is not possible due to operational security reasons. However, launch weather officers were consulted for lightning impacts, a better understanding of how radar and EFMs are used operationally, and their input to drive the development of the model (Cizek, 2026; Stewart, 2026).

$$FBS = \frac{\sum_{k=1}^M (f_k - o_k)^2}{M} \quad (4.5)$$

$$FBS_{ref} = \frac{\sum_{k=1}^M (f_k^2) + \sum_{k=1}^M (o_k)^2}{M} \quad (4.6)$$

$$FSS = 1 - \frac{FBS}{FBS_{ref}} \quad (4.7)$$

$$FSS_{baseline} = .5 + \frac{f0}{2} \quad (4.8)$$

Case studies are presented showcasing a strong model prediction for the continuation of lightning across the prediction window, a case where the model did not do well, then a lightning onset and ending case. The maximum absolute value of the potential gradient is provided, showcasing the spatial change in potential gradient across the suite, with some higher values seen closer to the onset of lightning.

These maps are then followed by an ablation study across the entire model domain. Ablation is used to document the operational impact of specific data outages, informing forecasters what to expect for the specific outages. 15 different experiments are presented. Ablating each dataset to zero would generate non-physical inputs. Thus, the HRRR data are ablated to the mean value of the training, validation, and test samples. Isothermal reflectivities, GLM group energy, EFM count greater than 1000, variance, and standard deviation are set to zero. The EFM mean, root mean square, and maximum absolute value are set to the mean of the training, validation, and test samples to replicate the fair weather electric field. Differential reflectivities are set to -5 dB, the normalized mask value. Zero cannot be used for differential reflectivity as it has physical meaning. This ablation technique allows the ablated data to be treated as valid data to examine model response, while also not requiring re-training of the model by eliminating observation specific encoder branches. A complete table of the statistics is provided in the Appendix (Tables B.2 - B.5), but the primary metric chosen

to examine in the results section is the AUC-PR.

Finally, the accumulated local effect ALE (Apley and Zhu, 2020) is calculated across the entire 2024 test dataset and target model domain. A unique subset of 20 quantiles are calculated from 0 to 1, to quantify changes in lightning probability. ALE is chosen to capture the impact on the model output given some inputs are correlated with each other. ALE for -10C isothermal reflectivity, mid-Zdr, GLM group area, MSLP, and one field mill (FM07) are presented in the results section. The other field mills and model inputs were examined and showed similar results, but are left out for brevity.

4.2 Results

This section provides a detailed analysis of the model via attributes and AUC-PR performance diagrams, case studies, and concludes with an explainability analysis.

4.2.1 Objective Results

The attributes diagram presents in Figure 4.4 the reliability of model. Figure 4.4a depicts strong divergence in the reliability of the lightning forecast above 50 percent probability, beyond the 00-15 minute lead time. The 00-15 minute lead time follows the perfect forecast line out to probabilities near 95 percent, with a slight over forecast bias above 60 percent. 15-30 minute lead time does continue to add skill beyond 50 percent, but the over-forecast bias increases. Model reliability drops drastically beyond 30 minutes, with some high forecast probabilities not adding any skill. While the goal of the model is to add skill for all nowcasting lead times, a skillful forecast out to 30 minutes is still operationally relevant for launch weather officers and resource protection. The Brier Score (BS) increases from 0.060 to 0.076 in the nowcasting window. In the bottom panel (Figure 4.4b) the histograms of forecast probability are

provided for each lead time; note the log-scale on the y-axis. High forecast probabilities are observed out to 95 percent for all lead times. A more quantitative investigation into the model performance throughout the 1-hour forecast window is presented in Figure 4.5 with the AUC-PR diagram.

Figure 4.5 presents the precision-recall performance diagrams, for each specific lead time, accompanied with maximum CSI scores and the AUC in the text box. The thresholds accompanying each maximum CSI score are documented in Table B.2, in the Appendix. The 00-15 minute lead time has the highest AUC of 0.47, with a max CSI of 0.34. These values decrease to a max CSI of 0.22 and AUC of 0.26 for the 45-60 minute lead times. Qualitatively, a leftward and downward shift is observed for each lead time, with the 45-60 minute lead time observing a significant drop-off beyond the 45 percent threshold. The AUC-PR diagram demonstrates that the metrics favor the shorter lead times out to 30-minutes, rather than the latter half of the 30-60 minute forecast window. The locations of the persistence and HRRR-f01/f02 forecast metrics are documented with the markers indicated in the legend. These points are shifted significantly toward the bottom left corner, from the LaunchCast lead time curves. Table 4.4 shows specific components of the AUC-PR curve, along with FSS. The FSS for a 5-nautical mile window decreases from 0.67 to 0.48, and 0.72 to 0.54 for a 10-nautical mile window, verifying the trend observed in AUC-PR. The 45-60 minute prediction falls below the climatologically informed FSS value. However, these FSS values exceed the persistence FSS values (FSSp) at all lead times. This is a novel contribution in lightning nowcasting for KSC, given current operational models like LightningCast do not document statistics at 15-minute temporal resolution. With reliability and AUC-PR/CSI metrics documented, a few case studies are presented next to highlight model output.

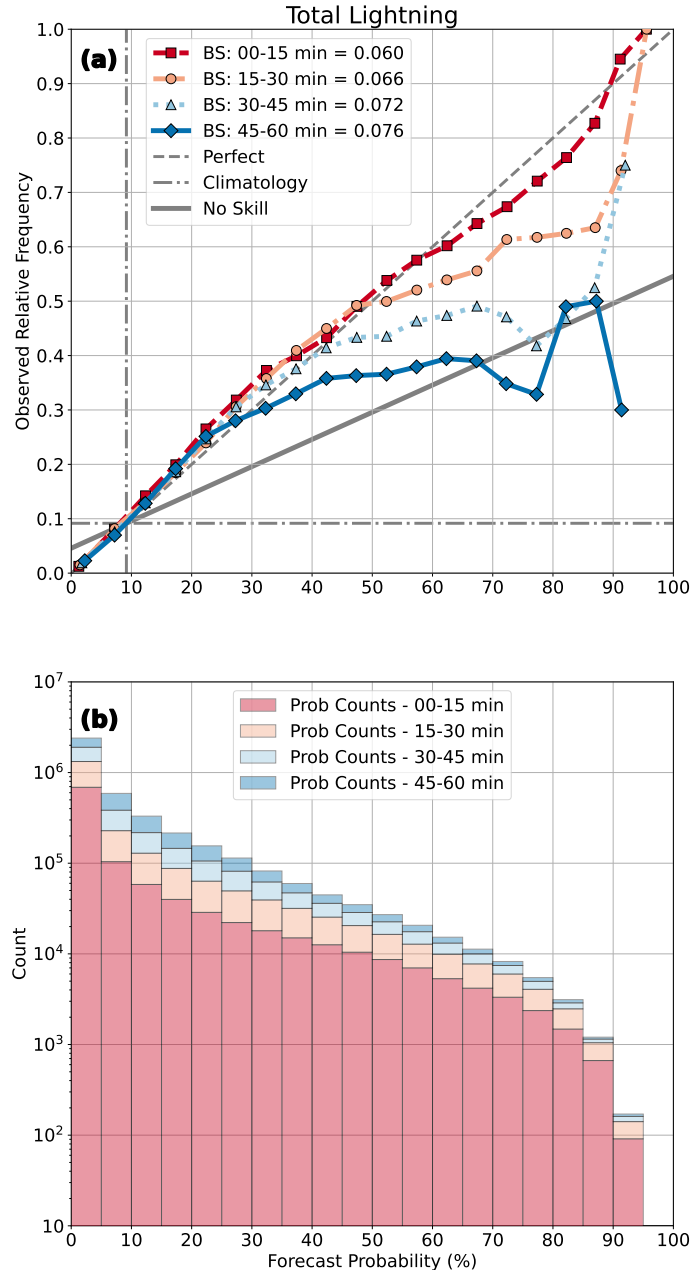


Figure 4.4: Reliability diagram of the LaunchCast model. Panel (a) provides the explicit reliability for the four lead times (dark red - 00 to 15 minutes; orange/peach - 15 to 30 minutes; light blue - 30 to 45 minutes; dark blue - 45 to 60 minutes), with forecast probability on the x-axis and observed relative frequency plotted on the y-axis. Panel (b) shows the forecast histogram for the four lead times. Note the log-scale on the y-axis in panel (b).

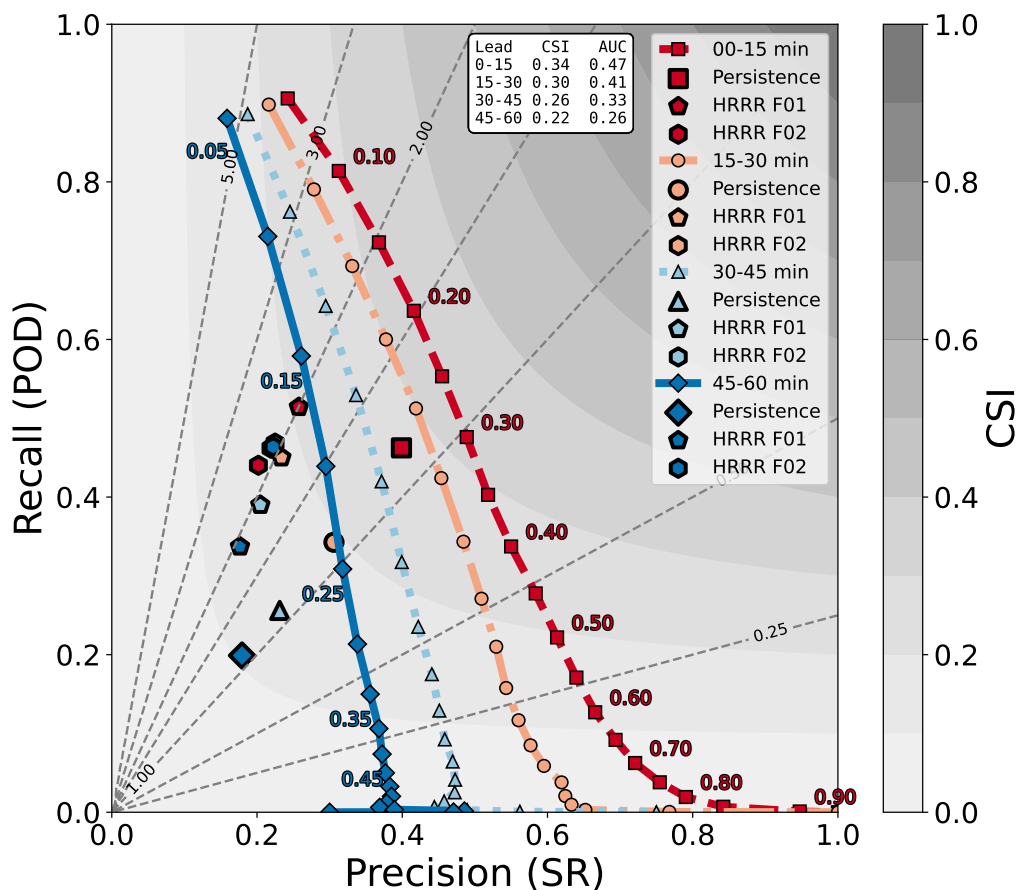


Figure 4.5: Precision-recall performance diagram, highlighting model performance for the four lead times. Gray filled contours denote the CSI values, with precision or success ratio (SR) plotted on the x-axis and recall or probability of detection (POD) plotted on the y-axis. The stand-alone markers outlined in thick black denote the 30-minute GLM persistence forecast (matched to LaunchCast symbols), the pentagons correspond to the HRRR f01, and the hexagons the HRRR f02 lightning product.

Table 4.4: Statistical summary for each forecast/lead time of the LaunchCast model. (t0-t3) are the keys used in the code, for traceability. These statistics mirror experiment 0 in the statistics provided in the Appendix. 'p' is for a 30-minute GLM persistence forecast. 'h1' is for the HRRR f01 lightning prediction and 'h2' is for the HRRR f02 lightning prediction. The FSS to beat is provided via Eq. 4.8 for climatology, and the '-5' and '-10' designations are for the FSS 5 and 10 nautical-mile windows.

Statistic	00-15 min Forecast	15-30 min Forecast	30-45 min Forecast	45-60 min Forecast	App. Tab.
Max CSI Threshold	0.20	0.20	0.20	0.15	B.2
Max CSI	0.34	0.30	0.26	0.22	B.2
AUC-PR	0.47	0.40	0.33	0.26	B.3
AUC-PR-p	0.23	0.17	0.13	0.11	—
AUC-PR-h1	0.18	0.16	0.14	0.12	—
AUC-PR-h2	0.14	0.15	0.15	0.15	—
FSS-5	0.67	0.62	0.56	0.48	B.4
FSS-5-p	0.65	0.52	0.41	0.32	—
FSS-5-h1	0.48	0.44	0.38	0.33	—
FSS-5-h2	0.39	0.42	0.42	0.42	—
FSS-10	0.72	0.67	0.62	0.54	B.5
FSS-10-p	0.71	0.59	0.49	0.39	—
FSS-10-h1	0.53	0.49	0.43	0.39	—
FSS-10-h2	0.45	0.48	0.48	0.48	—
FSS to beat	0.54	0.54	0.54	0.54	—
AUC-ROC	0.88	0.85	0.82	0.78	B.6
AUC-ROC-p	0.69	0.63	0.58	0.55	—
AUC-ROC-h1	0.68	0.65	0.62	0.59	—
AUC-ROC-h2	0.64	0.65	0.65	0.65	—
POD/Recall	0.64	0.60	0.53	0.58	B.7
POD-p	0.46	0.34	0.26	0.19	—
POD-h1	0.51	0.45	0.39	0.34	—
POD-h2	0.44	0.46	0.47	0.46	—
FAR	0.59	0.62	0.66	0.74	B.8
FAR-p	0.60	0.69	0.77	0.82	—
FAR-h1	0.74	0.77	0.80	0.82	—
FAR-h2	0.80	0.78	0.78	0.78	—

4.2.2 Case Studies

In this section several case studies are presented capturing model output with high probability, marginal probability, the onset of lightning, and the cessation of lightning.

Figure 4.6 depicts a case, valid from 2345 - 0045 UTC 06 - 07 Jun 2024 where the model performed exceptionally well. Panel (a)-(d) depict maximum probabilities ranging from 93.19 to 96.04 percent, with the maximum occurring in panel (b), in the 15-30 minute prediction window. AUC-PR increased out to 45 minutes lead time in panel (c), from 0.81 to 0.89. This value drops significantly to 0.46 in panel (d). Qualitatively, these high probabilities align well with the targets (dark gray fill), with no targets falling outside of a predicted probability. Panel (a) includes the observed EFM maximum absolute value that was input from 2145 - 2245 UTC - the previous hour. A spatial gradient is seen with values near 6000 V/m occurring near the 80-90 percent probability contour, falling off to below 1000 V/m to the southeast of the model domain. This is similar to the result in Handel 2022, where a gradient of field mill values was observed across the suite. For transparency, a similar case with weaker predicted probabilities is shown in Figure 4.7.

In Figure 4.7, valid times are from 1950-2050 UTC 09 Jul 2024. Maximum predicted probabilities drop from 60.06 to 36.34 in panels (a) through (d), where the high 0.97 AUC-PR drops to 0.24. Again, most lightning targets are within non-zero predicted probability contours, and a similar spatial gradient across the field mill suite is observed with values near 6000 V/m occurring at the edge of the lightning target in panel (a) dropping to close to 1000 V/m near the coast.

For further transparency, lightning cessation and onset cases are presented in Figures 4.8 and 4.9. In both cases lightning targets occur outside of any small predicted probabilities. AUC values never reach above 0.20 in either case. In Figure 4.8a high

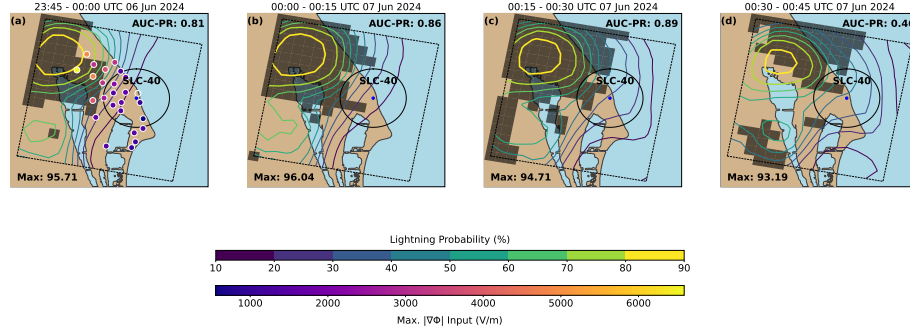


Figure 4.6: Example model output for 2345 UTC 06 Jun 2024 - 0045 UTC 07 Jun 2024. The grey filled pixels denote the target, while the colored contours denote the predicted probability. In panel (a), the EFM locations are filled with the maximum absolute value used as input between 2245 - 2345 UTC. The location of SLC-40 is provided for reference only. The blank fill in panel (a) is indicative of the masked EFM input values - a novelty of the model architecture. Note, the yellow fill is FM07, which had the inputs generated from the time series presented in Figure 4.2. Panel (a) is the 00-15 minute lead time, (b) is the 15-30 minute lead time, (c) is the 30-45 minute lead time, and (d) is 45-60 minute lead time.

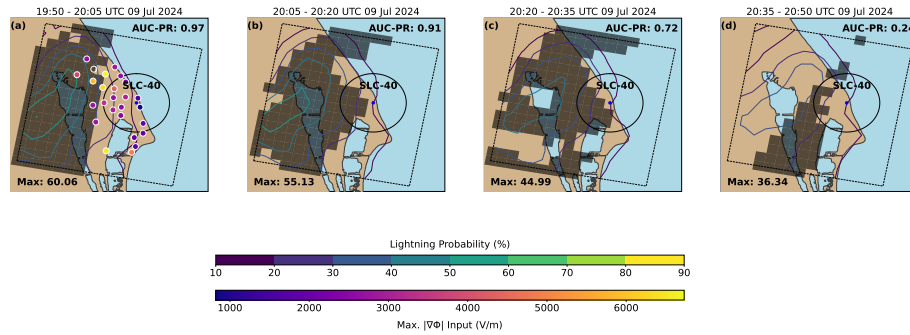


Figure 4.7: Same as in Figure 4.6 but for 1950 - 2050 UTC 09 Jul 2024. EFM values are from 1850-1950 UTC.

EFM values are still seen under target, approaching 7000 (V/m) with fair-weather values observed, never exceeding 300 (V/m) in Figure 4.9. These cases are primarily presented to show the onset and end of lightning are included in the datasets. Given there are weak quantitative metrics in these case studies, these subsets are not evaluated further on their own, but are included in the explainability metrics across the entire test dataset presented in the next section.

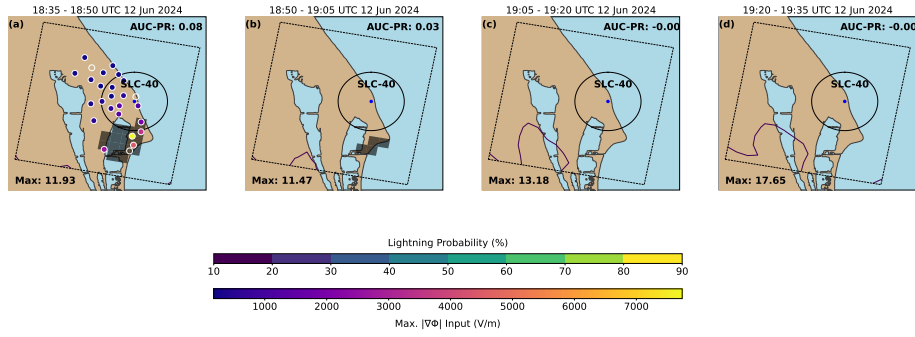


Figure 4.8: Same as in Figure 4.6 but for 1835 - 1935 UTC 12 Jun 2024, a lightning cessation case. The EFM values are from 1735 - 1835 UTC.

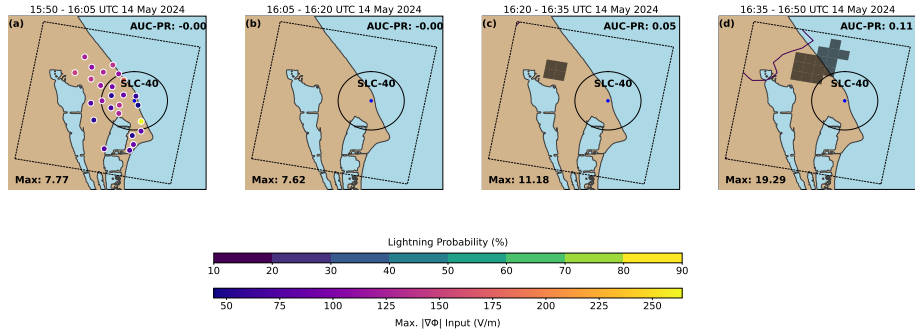


Figure 4.9: Same as in Figure 4.6 but for 1550 - 1650 UTC 14 May 2024, a case where lightning began in the prediction window. The EFM values are from 1450 - 1550.

4.2.3 Explainable AI

A grouped ablation study is presented in Figure 4.10. Ablation was used to mimic operational data outages and document model degradations when data are not present. In Figure 4.10 the y-axis denotes the percent change in AUC-PR, where smaller values indicate minimal model degradation, and high values indicate significant model degradation.

Figure 4.10 documents the different combinations of input datasets. Experiment 4 has the smallest change in AUC-PR, indicating the GLM group area dataset had minimal impact on the model performance, provided MRMS, HRRR, and the EFM data were unchanged. Experiment 1 had the second lowest reduction in AUC-PR, where MRMS, GLM, and EFM were unchanged, followed by experiment 5 where only the MRMS and EFM data were left unchanged, or non-ablated. Experiments 1, 4, and 5 all include the 26 EFM time series data, indicating the suite of EFMs have significant influence over the model statistics. The trend continues when examining the other 12 experiments, for instance, in experiments 8, 9, and 12 where the EFMs are not ablated the reduction in AUC-PR is smaller than when the EFM data are ablated. This is a novel discovery, given this first effort to employ explainability metrics on this non-traditional EFM dataset. When only one set of inputs is left unablated, as shown in experiments 7, 11, 13, and 14, very large reductions in the AUC-PR are observed. This is indicative that the model has not favored one particular dataset, and all dataset inputs play a role in the model predictions.

From this ablation study, it is gleaned that the MRMS and EFM data play the strongest role in generating the lightning predictions, with the GLM data playing the weakest role. This result is investigated further with ALE plots for select variables from each input type.

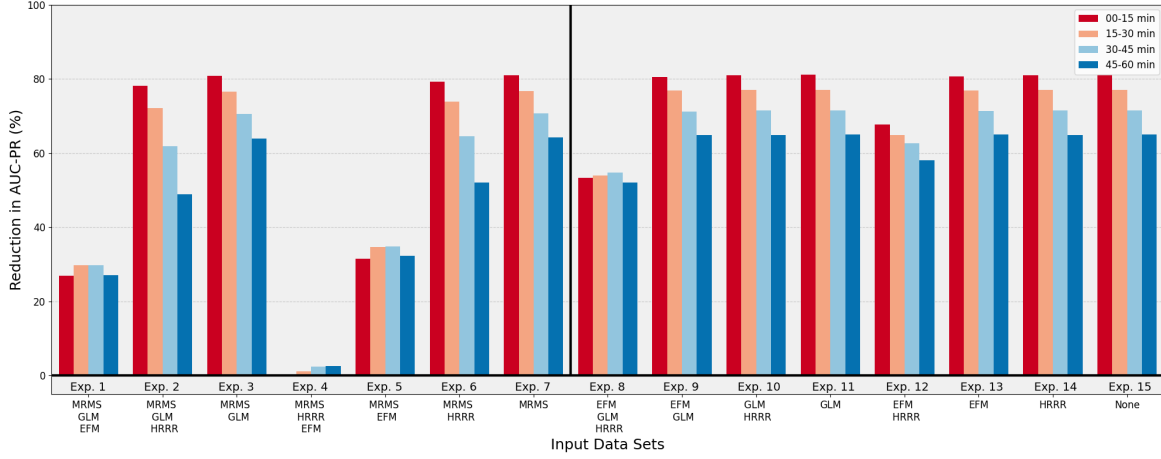


Figure 4.10: Plotted are the changes in AUC-PR for the 15 ablation experiments. Red bars plot changes for 00-15 minute lead times, peach/orange for 15-30 minute lead times; light blue for 30-45 minute lead times, and dark blue for 45-60 minute lead times. The experiment number corresponds to the number listed in Table B.3, in the Appendix. The data types listed below each experiment indicate the non-ablated datasets.

Figure 4.11 shows the ALE for a subset of input variables. The first two rows, correspond to a specific observation input window or HRRR forecast time. The last row, given the EFM time series, encapsulates the entire previous hour of EFM observations. Scaled histograms are provided as a proxy for observation density across the values used in the test dataset.

For the MRMS data in the top row larger values of reflectivity and Z_{DR} are shown to increase the forecast probability from the average prediction. Smaller values of reflectivity are seen to have the opposite affect, while the negative Z_{DR} values begin to generate positive ALE values, albeit small and less than .05. The magnitude change across the spectrum of reflectivity is almost four times greater than the change in magnitude seen for Z_{DR} , indicating a large local effect for reflectivity inputs. Other temporal inputs are presented in Appendix Figure B.1, noting the reversal in the ALE signal for the older observational inputs. MSLP for the last HRRR forecast time step used indicates an increase in probability for .2 percent for 30-60 minutes, while still

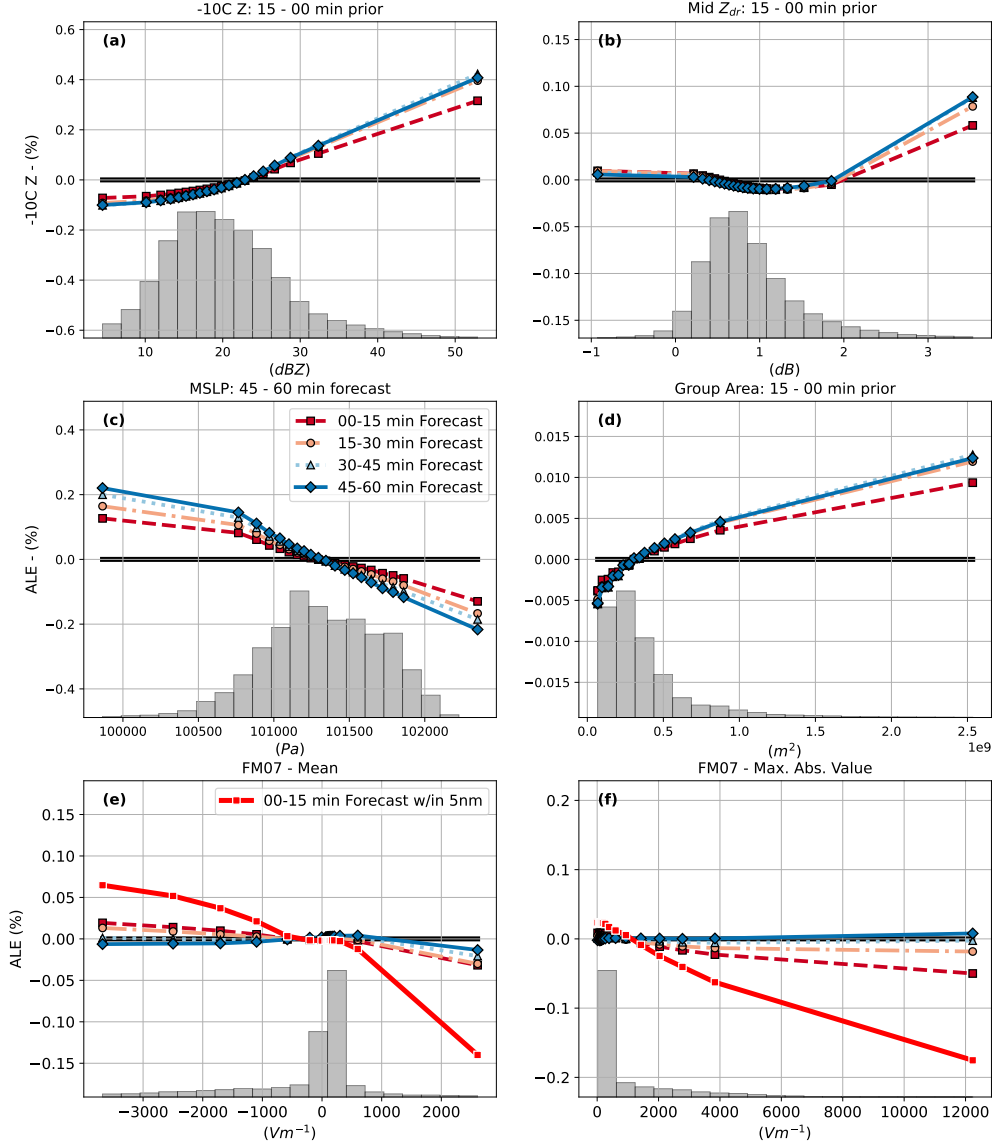


Figure 4.11: A plot of accumulated local effects for -10 °C reflectivity (a), mid-level Z_{DR} (b), forecast MSLP (c), group area (d), and FM07 mean (e) and max absolute value (f). Note the x- and y-axes vary, with panels (a-d) referencing a specific temporal input time. Panels (e) and (f) display results across the entire input hour due to the LSTM architecture. A larger subset of input variables and times is provided in Appendix B.1

influencing the 00-30 minutes. This influence changes as MSLP values increase. In panel (d) GLM group area is shown to increase slightly, with larger flashes, albeit the scale of the influence is significantly less than the MSLP, reflectivity, and Z_{DR} .

Finally, ALE plots are provided for Field Mill 07 in the bottom row. The brighter, solid red line includes the ALE for within 5-nautical miles of the field mill site for the first predictive time step. Note again, panels (e and f) are not input or forecast time windows, but two different EFM statistics: mean and maximum absolute values. A larger response in ALE within 5-nautical miles of the field mill is observed, indicating an influence from the distance exponential weighting in the model. Strong negative mean field values in panel (q) show an increase in lightning probabilities, while strong positive field values show a decrease in lightning probabilities. This may be attributed to the electric field reversal often observed when electrified storms are far away from the field mill, due to the Euclidean distance of the negative charge aloft generating a positive value at the field mill site. A strong bias towards fair weather field values is seen in both the mean and maximum absolute values as well.

Chapter 5

HorusCast: Specific Differential Phase Estimation on Storm Scale

This chapter presents the results from the HorusCast model. HorusCast estimates specific differential phase, K_{DP} , from the other polarimetric variables. K_{DP} can be used to describe severe hail, tornado-genesis, heavy precipitation, and lightning. Current estimators vary in their numerical discretizations of the range derivative of the differential phase. The aim of this chapter is to develop a machine learning (ML) model that can accurately estimate the K_{DP} with a fully digital, S-band phased array radar, with a Horus-specific Vulpiani K_{DP} estimate as the target (Vulpiani et al., 2012). I chose the Horus radar to prove the technique works on a potential future replacement of the WSR-88D network (Schvartzman et al., 2026). This is the first time Horus data have been used in a ML architecture. The high spatial and temporal resolution allowed for a large number of samples to be generated from a single storm, a limitation of this work. This chapter discusses how the training, validation, and test datasets are developed, the 1-D convolutional neural network (CNN) developed for predictions, and concludes with a presentation of results.

5.1 Data and Methods

5.1.1 Horus Radar Data Inputs

I use the 03 June 2025 central Oklahoma Horus deployment to generate the training, validation, and test datasets for the HorusCast model. Due to the irregular deployment schedule and varying scanning strategies, one deployment with a consistent scanning strategy was used for the model training and evaluation. The limitations of only using one deployment is discussed later in this section.

Radar scans began at 21:12:19 UTC and ended at 22:13:55 UTC. The deployment consisted of sampling a severe line of convection approaching from the west-northwest that provided a history of severe storm reports, shown in Figure 5.1. There are numerous reports of severe wind in central Oklahoma, with one tornado reported at 2204 UTC, 4 miles southeast of Newcastle, Oklahoma. The life-cycle of the tornado was included in the radar sampling, and is included in the test dataset metrics presented later.

For this deployment, Horus was set to a mechanically rotated 1-D spoiled beam, resulting in a 360-degree volume scan every 24 seconds and a range resolution of 77-meters (Palmer and Schwartzman, 2025; Schwartzman et al., 2026). The location of the radar is provided in Figure 5.1. A plan position indicator (PPI) valid at 21:47:07 UTC, for 7 degrees elevation angle, is also shown in Figure 5.2. Figure 5.2a indicates strong convection with reflectivity values greater than 55 dBZ, with Z_{DR} greater than 3 dB in pockets of the stronger convective cores shown in 5.2d. The velocity field in 5.2b shows a strong couplet between the -10 and -20 azimuthal distance lines, indicative of the strong velocities preceding the tornado report. The cross-correlation ratio in 5.2e shows most values above .8. Spectrum width data in 5.2c show a larger spread of the

velocity coincident with the higher reflectivities in the strong line of convection to the west. Finally, the Φ_{DP} field in the bottom right corner shows increasing Φ_{DP} with range, as the radar ray encounters more back-scattering along the beam. These six traditional and dual-polarization variables, accompanied by range and signal-to-noise ratio, were accompanied with a binary mask for input into the model. Previous K_{DP} estimators use valid reflectivity masks to dictate the window width of the numerical derivation of the Φ_{DP} field (Reimel and Kumjian, 2021b). Additionally, they use ρ_{hv} to filter out bad range gates. Early goals of this project were to generate a K_{DP} prediction in real-time, and it was hypothesized that Z_{DR} and V_r would help in that temporal prediction; thus, they are included as inputs into this estimator. These features and their corresponding masks are listed in Table 5.1. The model is designed such that the K_{DP} estimation can be employed directly on a Py-ART (Helmus and Collis, 2016) extraction, so long as the range-gate dimension is maintained.

Input Variable	Mask	Units
Reflectivity (Z)	$Z > 0$	dBZ
Differential Reflectivity (Z_{DR})	$-2 \leq Z_{DR} \leq 7$	dB
Velocity (V)	$-60 \leq V \leq 60$	m/s
Spectrum Width (σ)	$\sigma \geq 0.0$	m/s
Cross-correlation Ratio (ρ_{hv})	$\rho_{hv} \geq 0.8$	<i>None</i>
Differential Phase (Φ_{DP})	Is Real	Degrees
Signal-to-Noise Ratio (SNR)	$SNR \geq 10$	<i>None</i>
Range (R)	$R \geq 0.0$	m
Target Variable	Mask	Units
Specific Differential Phase (K_{DP})	$Z + SNR + \rho_{hv}$	deg/km

Table 5.1: The inputs and target for the HorusCast model. The masks concatenated next to the model input are included, along with the units of each variable.

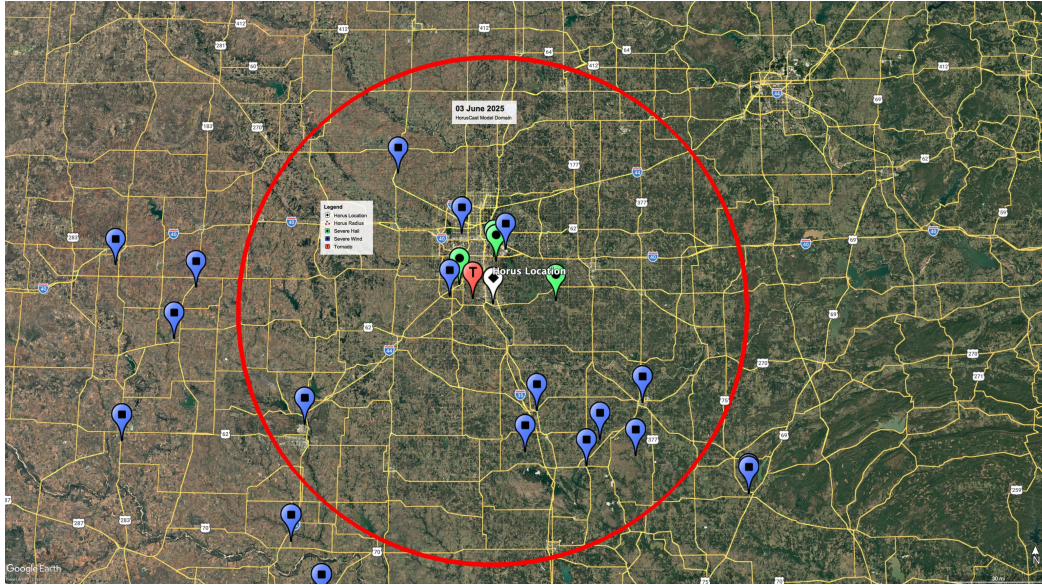


Figure 5.1: Google Maps rendering of the location of the Horus deployment. This is the deployment location of the radar for all three-cases presented here. Storm Prediction Center storm reports (red: tornado; blue: wind; green: hail; white: Horus location) are included with an overlay of the sensing radius of the radar, 120 kilometers (1562 range gates), for the 03 June 2025 deployment. (Latitude: 35.1873300; Longitude: -97.4442620; Elevation (MSL): 342 meters; Radar Height Above Ground: 3 meters)

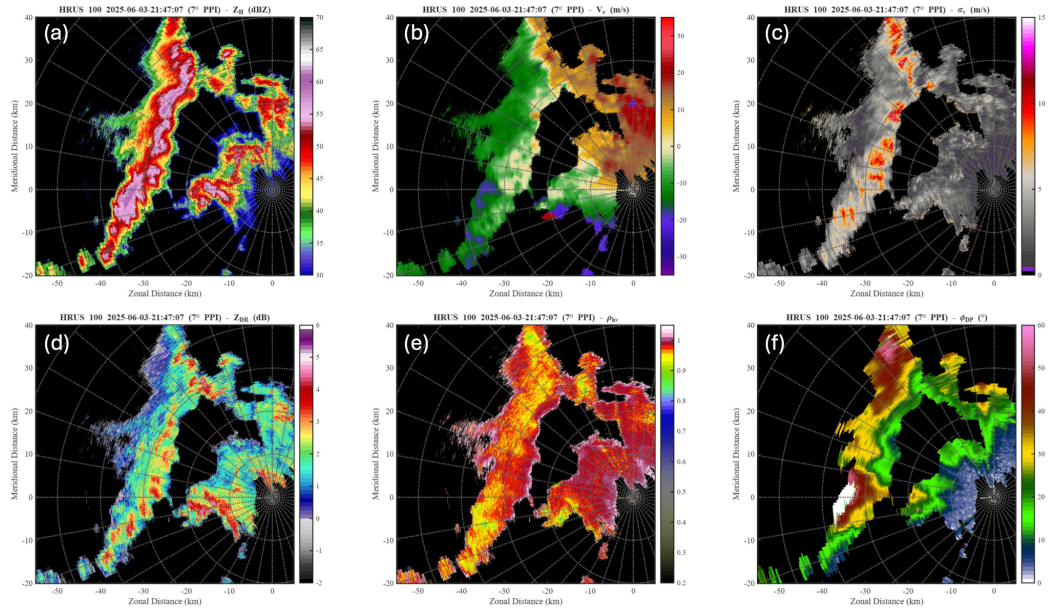


Figure 5.2: Example PPI at 7 degree elevation angle, from the Horus deployment used for training, validation, and testing. (a) is the reflectivity, (b) is the radial velocity, (c) is the spectrum width, (d) is the differential reflectivity, (e) the cross-correlation ratio, and (f) is the differential phase.

5.1.2 Target Generation

There is no Horus-specific approach to generate K_{DP} . Thus, methods from previous K_{DP} estimators from traditional dish antennas, with slow volume scan times, and much coarser spatial resolution are adapted to build the HorusCast K_{DP} target. Developing an ML-based K_{DP} estimation on a fully digital phased array system enables high resolution K_{DP} estimates to be generated, which can improve forecast operations. The higher frequency volume scans provided by Horus also enable a study of K_{DP} predictability, specifically identifying how many volume scans in advance negative K_{DP} can be detected. The digital nature of Horus allows for the beam to be steered almost instantaneously, and a K_{DP} prediction with a few minutes lead time could allow for the lightning strikes that generate from negative K_{DP} regions to be studied deeper. Generating K_{DP} with lead time was beyond the scope of this work, but needed model architecture changes and sample generation are discussed further in Chapter 6.

To generate the K_{DP} target, the Vulpiani et al. (2012) method is used to ensure a smoother K_{DP} signal, with a range window size of 2,250 meters described by the NWS operational K_{DP} algorithm for $Z \geq 40$ dBZ (Reimel and Kumjian, 2021b). The Vulpiani method is widely accepted as the state-of-the-art K_{DP} estimator. When applied to Horus, this generates a window size of 29-range gates for Horus' 77-meter range resolution. The convective window size is used since the radar is deployed only during severe convection; noting that some stratiform precipitation ($Z < 40$ dBZ) is present in the deployments. Merging the stratiform and non-stratiform K_{DP} targets introduced too much noise in the target. The operational NWS K_{DP} algorithm directs the use of a broader window around 6,250 meters (or 25-WSR-88D range gates). When a 6,250 meter window size was used, with 39-Horus range gates for the Vulpiani window, smaller amplitude K_{DP} values were observed. When merged with the convective target,

discontinuities along each ray were generated. **Thus, the target only utilizes the Vulpiani convective K_{DP} target, calculated across 29-Horus range gates.** It is worth exploring smoothing functions or non-Vulpiani based K_{DP} estimates in future work.

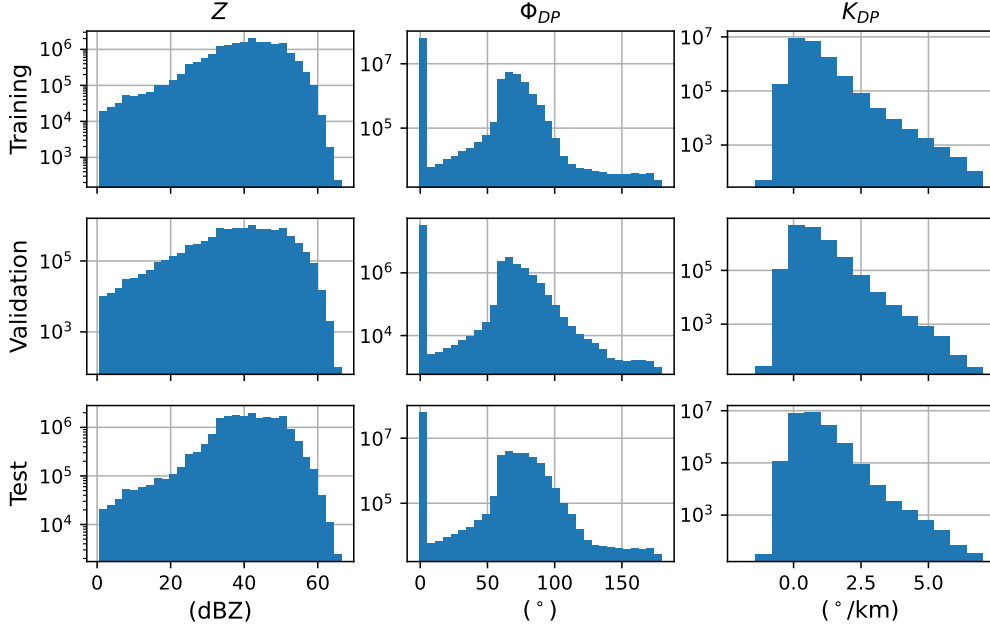


Figure 5.3: Histograms of the training (top row), validation (middle row), and test dataset (bottom row) data for HorusCast model.

Previous K_{DP} estimations are computed at specific elevations and azimuths, or along one ray (the terminology used throughout the rest of this document). This drives the reasoning to use individual rays in this research. Table 5.2 documents the radar scan times used to generate the training, validation, and test datasets. The datasets are generated such that there is no cross-contamination of rays from one-scan to the next across the different dataset types. However, the same meteorological event does occur across the training, validation, and test datasets, potentially overfitting HorusCast to this one storm. Given the 360-degree mechanical rotation of the radar during this deployment, there are a lot of rays that do not sense precipitation (i.e.

the southwest corner of Figure 5.2). Thus, a requirement for training and validation is the ray must contain 200 radar gates that have a non-zero reflectivity value. This ensures 1) the beam was not blocked, and 2) that there is a storm present that can generate enough back-scattering for a coherent Φ_{DP} field. This was implemented only for model training and validation; during model inference for the test dataset, this restriction is relaxed. Finally, it was imperative to keep the radar data at its original range resolution, to ensure the Φ_{DP} was as continuous as possible. Regridding to a coarser resolution would artificially inject discontinuities in the Φ_{DP} field, along the range dimension limiting the ML model’s ability to parameterize $d\Phi_{dp}/dr$.

Table 5.2: The deployment times and sample count for each dataset type on the 03 Jun 2025 deployment.

Dataset	Time (UTC)	Total Rays	ML Rays
Training	21:00:07 - 21:49:43	604,409	51,755
Validation	21:50:07 - 21:59:43	112,099	27,559
Test	22:00:07 - 22:14:31	180,362	51,624

Figure 5.3 shows the histograms for a subset of the inputs: reflectivity (Z), differential phase (Φ_{DP}), and the target K_{DP} . These show that the data across the training, validation, and test datasets have a similar distribution - an ideal ML setup. These input datasets are normalized using z-score normalization as inputs, with the mean and standard deviation calculated on the training data only. The K_{DP} target is not normalized. An individual ray example is provided for transparency of how the model sees one ray, in Appendix Figure C.5.

Horus’ deployments vary with irregular volume scans and scanning strategies. Initially integrating these deployments into the training, validation, and test datasets was explored. However, the model struggled to converge on an acceptable solution. For instance, the 2-D sector scanning strategy (Schvartzman et al., 2026) implemented on 28 April 2024 generated a 3-second volume update with a 20-meter range resolution

and dimension size of 1822. The training data for the 03 June 2025 case generates a range resolution of 77-meter with a dimension size of 1562. In order to accommodate the varying range dimension sizes padded batches in TensorFlow are required, introducing artificial data points to the model. Also, the ray-based samples generate an order of magnitude greater number of samples for the 28 April 2024 case, due to the 3-second volume sector scans, significantly biasing the training data to that case. Lastly, I was not able to influence the deployment in any way, and was limited to the use of the Level-II/moment products and metadata provided several months after the initial deployment. This prevented me from requesting specific volume-update times or scanning strategies. For these reasons, the 03 June 2025 case is used for training, validation, and testing where consistent and continuous sampling occurred. However, to accommodate the varying range dimension sizes, the model architecture allows a varying range gate dimension discussed below. A sample from the 28 April 2024 case and 19 May 2025 case are provided, showcasing this capability. An additional limitation is the same meteorological event occurs throughout the training, validation, and test dataset. Thus, large inferences about ML’s ability to generate K_{DP} estimates beyond severe-convection, cannot be made at this time.

5.1.3 AI/ML Model Architectures

I present the model architecture in Figure 5.4. The number of range gates is not specified allowing the model to be used on other deployments where the range differed from 03 June 2025. The 1-D convolutional neural network (CNN) I developed has a unique input layer for each input variable, to allow for easier explainability calculations on a GPU. Each observed variable is concatenated to its respective mask in Table 5.1 then all observed variables are concatenated before entering the encoder. This model then employs a 3-layer encoder and symmetric decoder to estimate the K_{DP} . Each

layer increases the number of filters from 128 to 512 in the encoder, and vice versa in the decoder. Other U-Net-like model architectures were explored, but lost the spatial information of the storm when pooled in the encoder.

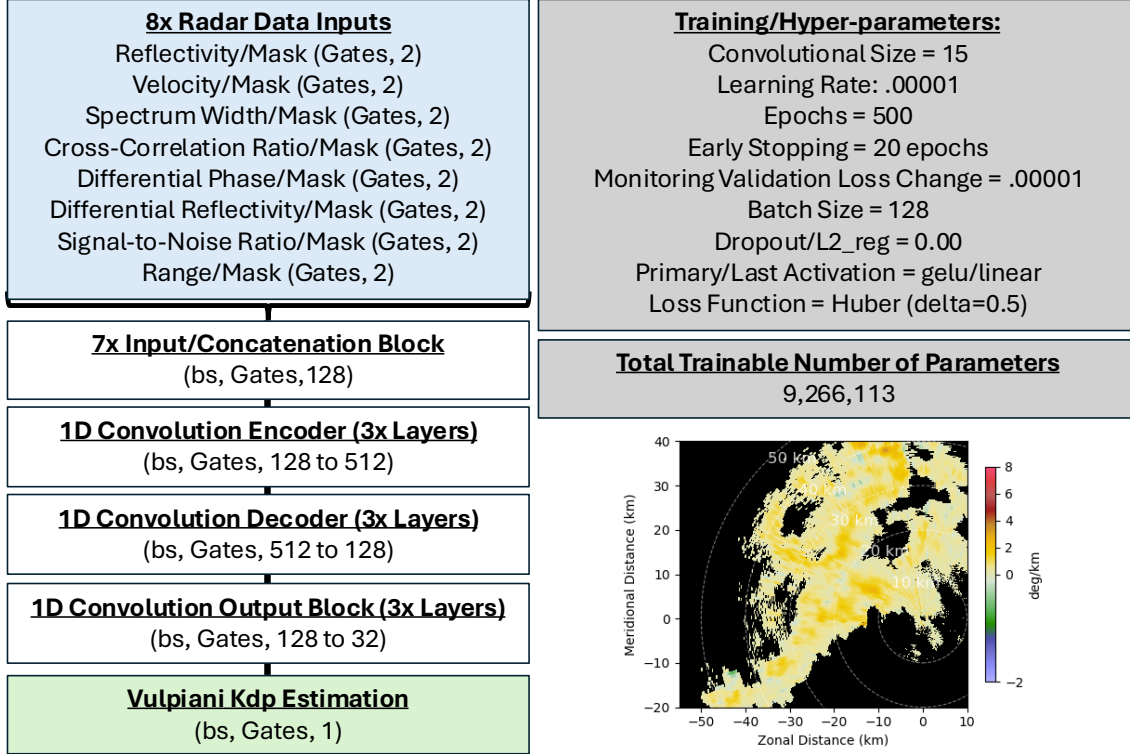


Figure 5.4: HorusCast model architecture. Hyperparameters for the model are included in the upper right, along with the total number of parameters.

The model was developed and trained using TensorFlow v2.15.0 (Abadi et al., 2015) using the Adam optimizer (Kingma and Ba, 2014). A simple hyperparameter search was conducted via grid search varying the learning rate, convolution size, activation function, and dropout. The final hyperparameters used for this model are included in Figure 5.4, and remain fixed throughout the evaluation section. The dataset here is generated for input into TensorFlow (Abadi et al., 2015), using a TensorFlow generator so sample weights can be applied to only calculate the Huber loss where there is reflectivity greater than zero, cross-correlation coefficient greater than 0.8, and the

signal-to-noise ratio greater than 10, as indicated in Table 5.1. One advantage of sample weights is I can weight the samples where large K_{DP} values are in the target, penalizing the model more when large K_{DP} is present.

$$sw = \begin{cases} 1, & |K_{DP}| < 1, \\ 2, & 1 \leq |K_{DP}| < 2, \\ 4, & 2 \leq |K_{DP}| < 4, \\ 8, & 4 \leq |K_{DP}| \end{cases} \quad (5.1)$$

Thus, K_{DP} values between 1 and 2 were weighted with a 2, 2 to 4 a 4, and greater than 4 an 8 as indicated in Equation 5.1. The values were chosen by monitoring the spread of the error in K_{DP} values greater than 2 deg/km, and varying the values with a small grid hyper-parameter search. I discovered this reduced the median error for large magnitude K_{DP} values to within one degree per kilometer. Sample weighting also reduces the need to generate custom loss functions, for easier reproducibility and transparency. To help ensure generalization, the first 50,000 samples of the training dataset are shuffled after each training epoch is completed.

5.1.4 Evaluation and Explainability

The primary mechanism for evaluating the HorusCast model is through root mean squared error ($RMSE$, Eq 5.2), mean absolute error (MAE , Eq 5.3), Pearson correlation coefficient (r ; Eq 5.4), and the ranked Spearman correlation coefficient (ρ ; Eq 5.5). In equations 5.2 through 5.5 N here is the total number of range gates, y_{pred} is the predicted K_{DP} , y_{true} is the target K_{DP} , and D_i is the difference in the rank of each sorted prediction pair. These values are provided accompanied by 2-D histograms and box and whisker plots to show the distribution of error across varying K_{DP} bins. An

error comparison to the NWS operational K_{DP} , from Equation 2.36, product is also provided, where the minimum system K_{DP} is set to 20 degrees (Carlin, 2026).

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{true_i} - y_{pred_i})^2} \quad (5.2)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_{true_i} - y_{pred_i}| \quad (5.3)$$

$$r = \frac{\text{cov}(y_{true}, y_{pred})}{\sigma_{y_{true}} \sigma_{y_{pred}}} \quad (5.4)$$

$$\rho = 1 - \frac{6 \sum_{i=1}^N D_i^2}{N(N^2 - 1)} \quad (5.5)$$

For explainability, feature attribution is provided via input times gradient, within the TensorFlow framework. Attribution methods aim to describe the marginal contribution of each input feature to model output (Mamalakos et al., 2022). Here the derivative of the model output is taken, then multiplied by the original input value, shown in Equation 2.32. One drawback here is that a small magnitude input with a high sensitivity can generate similar attribution as high magnitude input with a low sensitivity.

Additionally, Oklahoma Lightning Mapping Array (Rison et al., 1999) data are also overlaid on case studies to help qualitatively correlate regions where lightning occurred with higher K_{DP} values. The LMA data were acquired from the NOAA server (<https://data.nssl.noaa.gov/thredds/catalog/WRDD/OKLMA/catalog.html>).

5.2 Results

I present quantitative metrics with a 2-D histogram and a binned error analysis for the HorusCast model first, followed by case studies. The case studies are presented first for a case in the test dataset with explainability metrics. This is followed by two cases on different Horus deployments: 28 Apr 2024 and 19 May 2025. This is to show the transferability of the model, and discuss how the different scanning strategies impact the model estimates.

5.2.1 Test Dataset Metrics and Analysis

Figure 5.5 presents a 2-D histogram of the error distribution for the HorusCast model. With an r value of 0.97 and ρ of 0.96, this indicates ML is a valid tool for K_{DP} estimation for the 03 June 2025 case where severe convection was sampled. The mean absolute error (MAE) is at 0.07 degrees per kilometer, while the root-mean squared error is calculated at 0.11 degrees per kilometer. Additionally, the bias is calculated to be near 0.00. These statistics are calculated across the entirety of the test dataset described in Table 5.2.

A binned error analysis is presented in Figure 5.6. Note the dual y-axes, with the prediction error on the left for the whisker diagrams, and the logarithmic range gate count on the right for the background histogram. The errors are calculated across 0.50 degree per kilometer bins, with the center of the bin values presented on the x-axis. Here the histogram peaks at 0.75 degrees per kilometer with $10^{6.8}$ range gates in that bin. The range gate count drops as you approach higher positive K_{DP} values, where there are less than 100 range gates counting toward the spread in the error for K_{DP} values greater than 6 degrees per kilometer. The spread among the box and whisker plots becomes larger at 3.25 degrees per kilometer. The blue line within the box

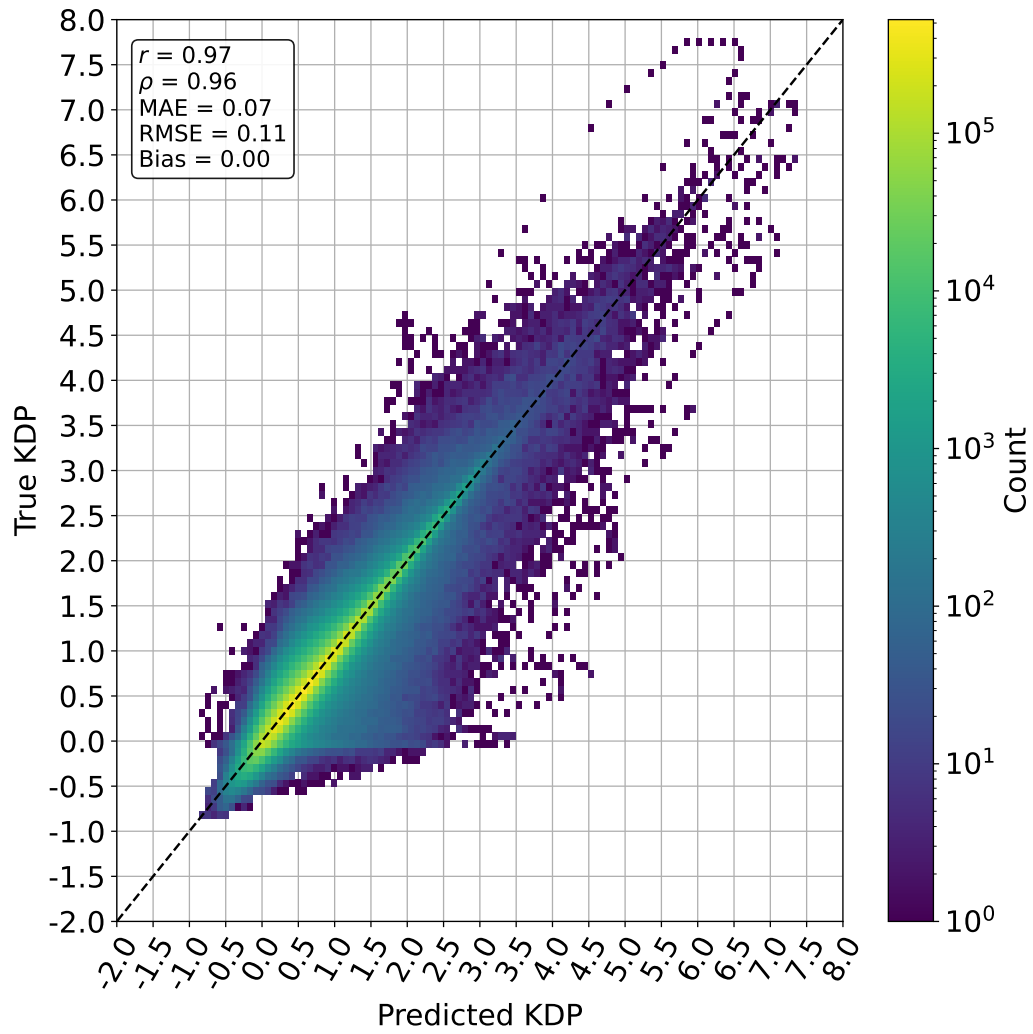


Figure 5.5: 2-D histogram depicting correlation of the predicted K_{DP} values (x-axis) and the true K_{DP} values (y-axis). The fill colors correspond to the number of range gates, on a log-scale. The statistics listed in the upper left corner correspond to equations 5.2-5.5.

indicate the median error value, with the bottom and top representing the 25th and 75th percentiles. Whiskers extend to 1.5 times the interquartile range. Large variability is seen between 3 and 7.5 degrees per kilometer, where the sample count is low. It is hypothesized the tornado occurring during the test dataset may have generated this large variability. The median error value does hover near -0.1, indicating an under-prediction of K_{DP} for large, positive K_{DP} values. From 0-3 degrees per kilometer the error spread is very low, with median values portrayed very close to 0.0. Shifting to negative K_{DP} , the model still struggles to capture the magnitude of the negative K_{DP} values. Thus, when comparing to the magnitude, a slight underestimation bias is in place for negative K_{DP} values.

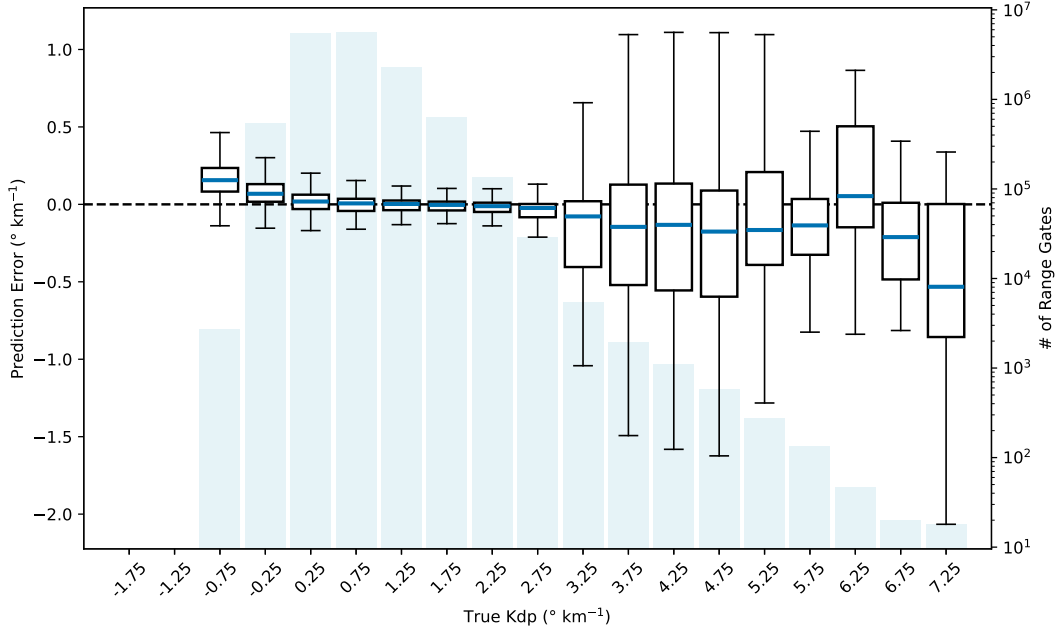


Figure 5.6: Box-whisker diagram for the prediction error. The shaded bars denote the number of range gates indicated on the right-dual axis. The left dual axis pertains to the prediction error. The box plots are binned in .5 deg/km bins, with the bin center indicated on the x-axis.

The first scan of the test dataset is shown in Figure 5.7. The reflectivity values

show a strong line of convection approaching the radar from the west-northwest, with a majority of values greater than 40-dBZ. The HorusCast output is shown in Figure 5.7b, with the LMA sources overlaid. The HorusCast output aligns qualitatively very well with the Vulpiani target, with the negative K_{DP} values denoted in the blue tinted pixels in the same locations as the Vulpiani target. The LMA sources are observed within the storm, however, displaced from the negative K_{DP} regions, similar to Melnikov et al. (2019) and Ryzhkov and Zrnica (2007). The NWS comparison in panel (c) portrays a completely different picture of the K_{DP} field as compared to the Vulpiani target. The higher magnitude and discontinuous NWS K_{DP} values are observed close to the radar, within 30-kilometers, with more spatially continuous features beyond. Overall, the model output in panel (b) aligns really well with the Vulpiani target in panel (d).

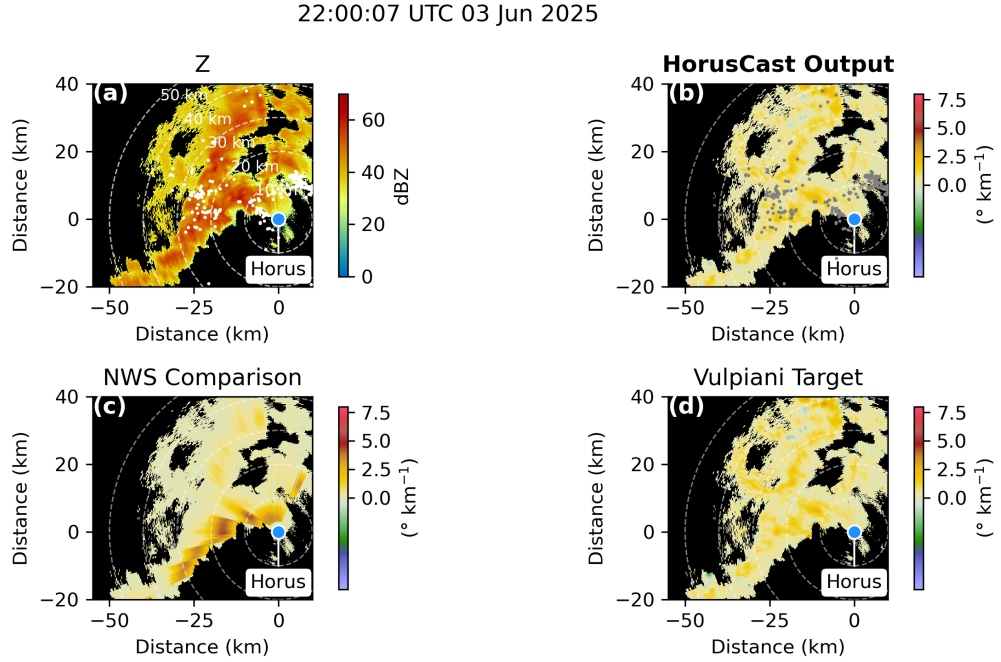


Figure 5.7: Reflectivity (a) , HorusCast Output (b), NWS comparison (c), and Vulpiani target (d) for the 03 Jun 2025 deployment, valid at 22:00:07 UTC. The white dots in panel (a) and grey in (b) denote the location of the Oklahoma LMA sources between 22:00:07-22:00:09, inclusive.

A further comparison of the statistics of HorusCast to the NWS and Vulpiani targets are captured in Table 5.3 across the entire test dataset. The values in the first row are the same values presented in Figure 5.5. The second row is the error of the NWS method and Vulpiani, and the last row the error of the HorusCast model when compared to the NWS K_{DP} calculation. It is seen that the evaluation statistics are comparable for the NWS - Vulpiani and HorusCast - NWS comparisons, with the bias changing from 0.51 to -0.50.

Table 5.3: Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method.

	r	ρ	MAE	RMSE	Bias
HorusCast - Vulpiani	0.97	0.96	0.07	0.11	0.00
NWS - Vulpiani	0.26	0.34	0.95	1.39	0.51
HorusCast - NWS	0.27	0.35	0.95	1.39	-0.50

I show the 2-D input times gradient (IxG) images for Φ_{DP} in Figure 5.8 and range (r) in Figure 5.9. These two features are shown due to K_{DP} 's dependence on these variables, provided in Equation 2.34. A clear IxG signal is seen on the boundaries of the storm, in both figures. IxG values for Φ_{DP} are two orders of magnitude greater than the range. Specifically, in Figure 5.8 negative IxG values (i.e., suppression of K_{DP}) are observed on the leading edge of the storm, transitioning to positive IxG values on the far edge of the storm (i.e., enhancing the K_{DP}), with fairly constant values in the center of the storm. The signs are reversed on the range IxG image in Figure 5.9, with still constant near zero IxG in the center of the storm. This indicates the model learned where the storm is and when the K_{DP} prediction was influenced by the presence of scatters, or where Φ_{DP} changed rapidly - in line with physical intuition.

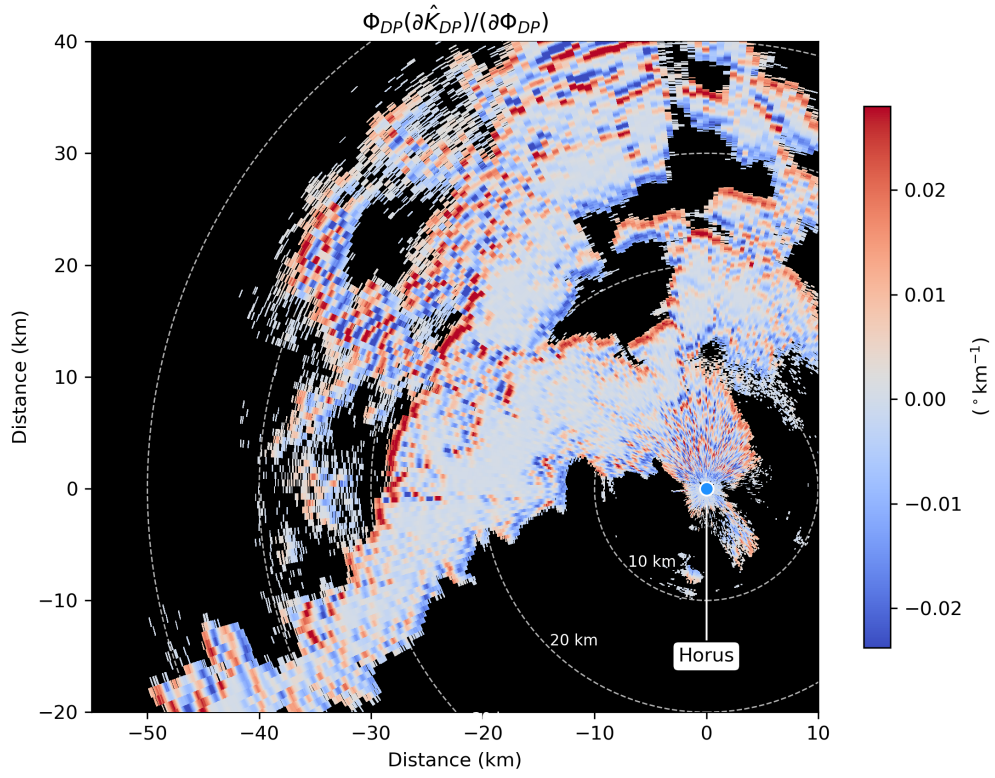


Figure 5.8: Input x Gradient (IxG) attribution image of the differential phase input variable, for the 03 Jun 2025 deployment, valid at 22:00:07 UTC.

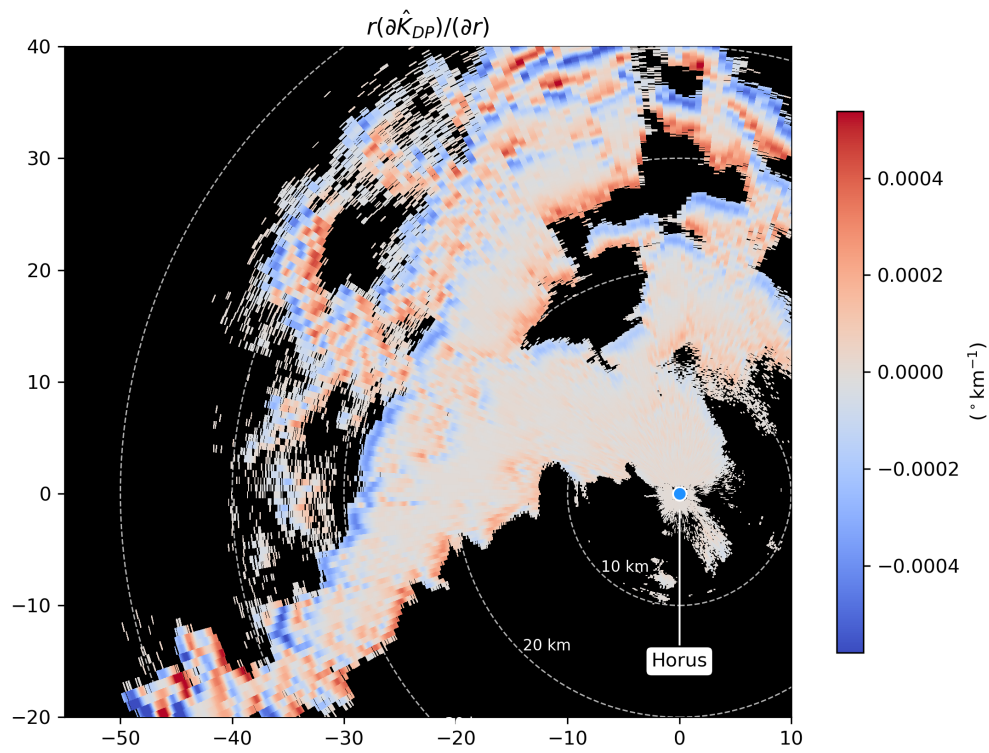


Figure 5.9: Input x Gradient (IxG) attribution image of the range input variable, for the 03 Jun 2025 deployment, valid at 22:00:07 UTC.

5.2.2 Additional Deployments

For transparency, it is important to understand how the model performs when the model is deployed on a different storm, and when the radar scanning strategy is varied. Figure 5.10 shows a scan at 21:19:46 UTC from the 19 May 2025 deployment, where the radar scanning strategy was the same as for 03 June 2025 case presented above.

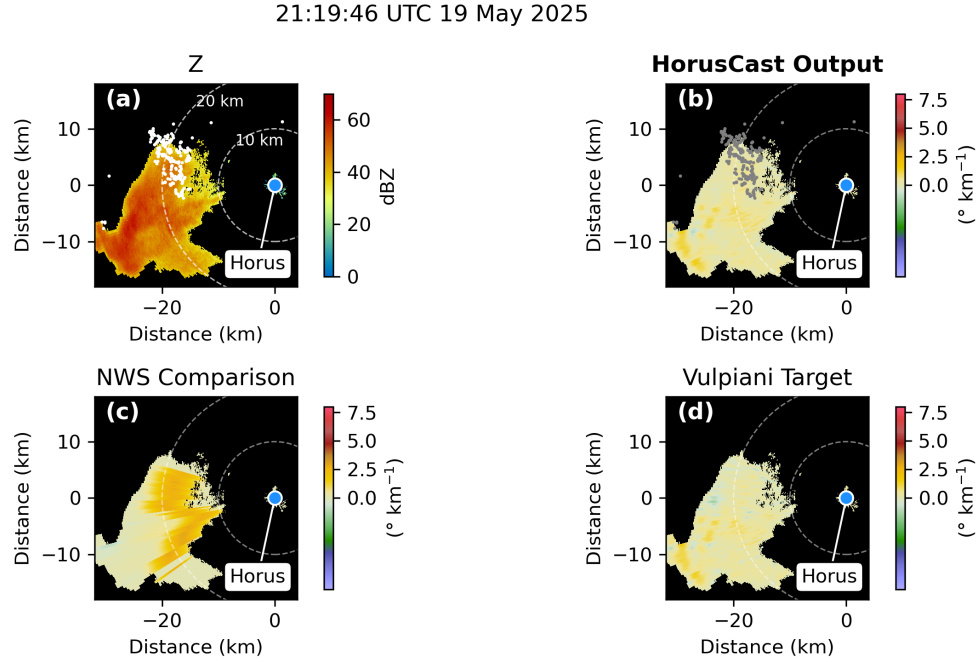


Figure 5.10: Same as Figure 5.7, but for 21:19:46 UTC 19 May 2025, with LMA sources valid from 21:19:46 -21:19:48 UTC.

This deployment resulted in scanning a single storm, for 56 minutes from 21:01:46 - 21:57:46 UTC, at the same location described in Figure 5.1. Similar to the 03 June case, there is good agreement in the HorusCast model output in panel (b) and the Vulpiani target in panel (d). The NWS comparison shows a similar, irregular K_{DP} pattern that is seen in Figure 5.7c, not in good agreement with the Vulpiani target. The LMA sources are present over the forward flank of the storm. The statistics over the entirety of the deployment are presented in Table 5.4. Poor correlations are still

observed between the NWS and the HorusCast output and Vulpiani target. I also note that the correlation coefficients are slightly less here (0.84 and 0.82), than the test dataset case presented earlier where (0.97 and 0.94) were presented. The error values also increase when compared to the test dataset case. This indicates the model is capable of being deployed on a similar radar scanning strategy and still maintain some skill, but might be overfit to the 03 June 2025 case given the drop in correlation and increase in error.

Table 5.4: Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method for the 19 May 2025 deployment, from 21:01:46 to 21:57:46.

	r	ρ	MAE	RMSE	Bias
HorusCast - Vulpiani	0.84	0.82	0.17	0.27	0.08
NWS - Vulpiani	0.13	0.21	0.86	1.22	0.45
HorusCast - NWS	0.13	0.22	0.88	1.19	-0.38

I next show HorusCast deployed on a 2-D sector scan in Figure 5.11, for 02:22:23 UTC 28 Apr 2024. Panel (a) shows a fully saturated scan with increasing reflectivity values as you get closer to the radar. There is some lightning present in the sector scan, but most of the sources reside outside of the scan. Moving to panel (b) a fairly homogeneous K_{DP} field is presented with ripples throughout the scan, oscillating between positive and negative values. This is not in good agreement with the K_{DP} target presented in panel (d) or the NWS field in panel (c). Note, however, beyond a few kilometers from the radar, the NWS comparison in (c) is in better spatial agreement with the target than the HorusCast output. This is represented in the correlation error values, presented in Table 5.5. Across the entire deployment, the NWS product has higher correlation values, and more agreeable error values, as opposed to the HorusCast output. I suspect this is due to the proximity of the storm to the radar, and the model cannot detect the storm boundaries, informed by the IxG analysis presented earlier.

This is discussed further in Chapter 6.

02:22:23 UTC 28 Apr 2024

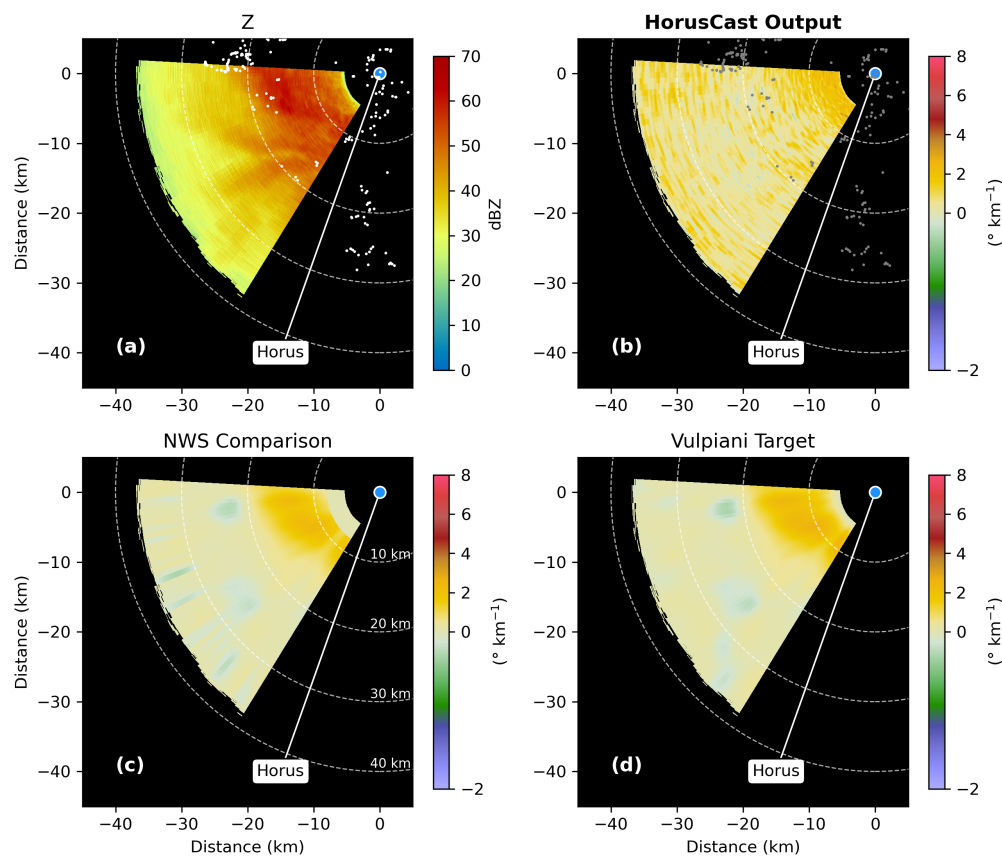


Figure 5.11: Same as Figure 5.7, but for 02:22:23 UTC 28 Apr 2024, with LMA sources valid from 02:22:23-02:22:25 UTC.

Table 5.5: Statistical comparison of the HorusCast model output, the Vulpiani target, and the NWS least squares method for the 28 April 2024 deployment, from 01:03:38 to 02:44:01.

	r	ρ	MAE	RMSE	Bias
HorusCast - Vulpiani	-0.05	-0.05	1.48	1.71	1.31
NWS - Vulpiani	0.43	0.63	0.43	1.08	-0.05
HorusCast - NWS	-0.43	-0.37	1.84	2.17	1.35

Chapter 6

Summary

I first discuss the overall contributions and conclusions of the dissertation before discussing the specific details and conclusions of each chapter. Given the varied objectives of each model, I discuss common themes and conclusions that span the dissertation.

First, I implemented **only one** LSTM layer to aid ML-based lightning predictions. BoltCast used one convolutional based long-short term memory (ConvLSTM) and LaunchCast used only one shared LSTM layer for the electric field mill (EFM) encoding of all 26 sites. Model architectures like LightNet+ (Zhou et al., 2020) explicitly extract the hidden and cell states from the LSTM layers and perform convolutions over their respective sequences, injecting activation functions beyond the original intent of the architecture. I demonstrated one LSTM-layer’s versatility, founded on the principles presented in Figure 2.6, to generate skillful predictions for either multi-day lightning outlooks or 1-hour lightning nowcasts with non-traditional EFM data.

Second, I demonstrated the continued versatility of lightning location data in ML models, either as inputs, targets, or evaluation aids. LaunchCast solidified this theme utilizing two ground-based systems (MERLIN and the EFMs) and space-based instrumentation (GLM) in one model. BoltCast translated GFS data into the probability of a GLM flash, while HorusCast utilized LMA data as an evaluation aid. A sub-theme here is the particular value added of space-based lightning detection, as its operational status over the United States (all of the domains used here) beyond 2030 remains uncertain. Accurate, efficient, and continuous lightning monitoring across the United

States allow for great statistical inferences to be made, foregoing the need for costly numerical discretizations.

Lastly, explainable AI techniques that are based on adjusting model inputs to rank data importance can be applied across varying meteorological scales. Five different model-agnostic approaches were implemented showcasing the drivers of each model. This provides data architects with an explicit prioritization of which data to keep and discard as data volumes increase in the future.

I repeat the novel, and itemized contributions from the introduction in the following section. The dissertation closes with detailed discussions of BoltCast, LaunchCast, and HorusCast along with outlines of future work for each model.

6.1 Summary of Contributions

My novel contributions to the scientific literature are listed below, organized by the three models I developed: BoltCast, LaunchCast, and HorusCast.

- BoltCast: (McClung et al., 2026)
 - Development of skillful AI/ML that predicts lightning across CONUS to 96 hours. Current operational products stop at 63 hours.
 - A direct comparison between a deep, high-parameter U-Net and a single LSTM-layer based, recurrent neural network.
 - My LSTM-based architecture is the first RNN based lightning prediction tool across the entire CONUS.
- LaunchCast
 - Development of a skillful sub-hourly lightning prediction tool using re-sampled

EFM and other geo-spatial data over Kennedy Space Center, outperforming persistence.

- 30-second resampled EFM data were generated from the original 50-Hz sampled data. Previous models used minute-averaged data (Bao et al., 2022; Speranza, 2019; Hill, 2018).
 - LaunchCast provided increased operational lightning nowcasts to 15-minute temporal resolution.
 - First-ever investigation quantifying and demonstrating the power of EFM data when compared to other geo-spatial data like NWP, satellite, and radar.
- HorusCast
 - Development of a 1-D convolutional based K_{DP} estimator on a fully digital S-band phased array radar (Palmer et al., 2023; Palmer and Schwartzman, 2025; Schwartzman et al., 2026).
 - First ever ML-based K_{DP} estimation with deep learning techniques.

6.2 Model Specific Discussion, Conclusions, and Future Work

6.2.1 BoltCast: Medium-range and Synoptic Scale Lightning Prediction

The 3-D convolutional neural network allowed the ML model to learn weights across the temporal NWP output dimension, while the LSTM, a recurrent neural network, model shared temporal information through the hidden and cell states passed along the NWP output dimension.

An attributes diagram revealed high-reliability by both the U-Net and LSTM models, with a slight under-forecast bias for higher lightning probabilities. Precision-recall curve evaluations showed U-Net PR-AUC metric scores decreasing with lead time from 0.62 at day 1 to 0.52 at day 4; LSTM PR-AUC metrics decreased from 0.61 at day 1 to 0.50 at day 4. When the PR-AUC was calculated at each GFS model grid point, investigating regional skill, high PR-AUC values are seen over the mountainous terrain in the western CONUS, and throughout the Florida peninsula. A t-test indicated statistically significant differences in model output between the U-Net and the LSTM model along the coastal regions of the western and southeast United States, and upper Great Lakes region. Permutation importance plots provided insights into how the model inputs affect the model output when changed. Here, the importance, or change in PR-AUC, was deemed greatest for the CAPE, reflectivity, and precipitation rates for the U-Net and CAPE, ice mixing ratio, and precipitation rate for the LSTM. These results apply across all seasons. When evaluating for particular seasons, summer had the higher PR-AUC values across all forecast days, with winter having the lowest PR-AUC values.

A hit-forecast and miss-forecast provided insights into predicted probabilities. The hit-forecast over WPAFB showed increasing PR-AUC by nearly 20 percent from day 4 to day 1. The coincident partial dependence plot for WPAFB showed little variability when the spatial extent of the single-variable modifications were implemented. The U-Net partial-dependence plots do reveal very high lightning probabilities across all modifications, near 100 percent. For the miss-forecast case over VAFB, the PR-AUC scores maxed out at Day 2 for the U-Net and Day 4 for the LSTM, then deteriorated in the shorter lead times. The spatial variability from the partial-dependence plots is far greater, showing dramatic increases in probability with the higher values for CAPE and precipitation rates. Reflectivity did not show the same variability, with predicted

probabilities remaining fairly low for all observed values. The partial-dependence plots also revealed a clear bias within the LSTM model architecture, to struggle to predict lightning probabilities over 90 percent, while the U-Net model can predict lightning probabilities close to 100 percent.

Given the high PR-AUC scores shown in Figures 3.6 and 3.11, the ability to generate lightning probabilities greater than 90 percent, the U-Net is a more reliable choice. This particular implementation of an LSTM model still holds merit, however. This LSTM model was able to generate reliable predictions, provided it under-forecasts higher lightning probabilities. Nonetheless, this LSTM model is the first RNN predicting lightning over the entire CONUS. A higher parameter LSTM based model may be able to overcome these limitations, but this work sought to show the merit of one-layer.

These two models are best suited for the summer months, when large areas of lightning are present. They are trained over CONUS, where predominant storm motions are from west to east, thus the models should be used with caution outside of CONUS. Finally, if the models are transitioned to operations, they are suited for low-risk, low-impact, non-time sensitive operations. Human forecasters could use this model to supplement their current workflows and aid in human-in/over-the-loop product generation. Additional applications include integrating a lightning product into the GFS post-processing, using these models.

A continuation of this work includes uncertainty quantification and investigating further explainability techniques like shapley additive explanations and layer-wise relevance propagation (Toms et al., 2020; Lundberg and Lee, 2017). Additionally, it would be interesting to study the utility of AI/ML to predict the flash rates of lightning, versus a binary classification. In doing so, specific lightning loss functions and metrics can be developed to account for the sparse nature of lightning across the CONUS. The prediction window could be also expanded out to 7 days or shortened so data from convection

allowing models, like the High Resolution Rapid Refresh, could be included. Spatial expansion could be useful as well, examining how similar global NWP inputs could predict lightning over Europe, Asia, and Africa using locally available lightning imagers or ground-based systems. Finally, using an operational ground-based lightning detection system like Earth Networks Total Lightning Network (ENTLN), Vaisala’s National Lightning Detection Network (NLDN), or World Wide Lightning Network (WWLN) may provide another useful comparison as lightning detection efficiencies vary across different platforms, and GLM does not discriminate between in-cloud and cloud-to-ground lightning while some ground networks do. Additional GLM limitations include lower detection efficiencies during the day and for smaller flashes. Additionally, the GFS data used as inputs in this work are slightly correlated and rely on daily maximum and minimum calculations.

6.2.2 LaunchCast: Discussion and Conclusion

The demand for easy access to space continues as space traffic increases at Kennedy Space Center. Kennedy Space Center hosts a unique suite of static electric field mills that monitor the cumulative vertical electric field over each sensor (Wilson, 1920). Historically, the EFM suite has been used to study the electrification of thunderstorms, correlating the EFM observations with radar observations (Cummins et al., 2025; Handel et al., 2022). These field mill observations, combined with radar data, space-based lightning observations, and high-temporal NWP data were used to create a lightning nowcasting tool with 1-hour lead time, with moderate skill in the first half of the prediction window. These results present a first skillful (and successful) attempt to blend an EFM dataset with other geo-spatial data.

Two CNN encoders encode the NWP and radar and satellite based data to provide a larger atmospheric context for the model, while one shared LSTM layer is used to

encode the EFM time series data predicting lightning on a smaller, EFM suite centered target domain. Reliability of the model is presented, capturing high reliability when the forecasted probabilities are less than 50 percent. Above 50 percent, there is larger divergence in the model reliability at higher probabilities, especially from 30 minutes to 60 minutes. Transitioning to an AUC-PR diagram, AUC-PR declines from 0.47 to 0.27 across the prediction window, with maximum CSI values doing the same decreasing from 0.34 to 0.22. AUC-PR alone beats a 30-minute persistence forecast derived from GLM group area, at all lead times. Additionally, the FSS scores presented in Table 4.4 for a 5- and 10-nautical mile window beat the 30-minute persistence forecast at all lead times.

Case studies revealed that when lightning is forecast across all four predictive windows, the model performs fairly well. Maximum predicted probabilities were higher than 95 percent with a spatial gradient in the absolute value across the EFM suite. When transitioning to cases where lightning began during the prediction window or ended, the model struggles to make those predictions, noting fair weather electric field values are observed in the previous hour. This indicates that the convective preconditioning of the model with either the MRMS or GLM data play a critical role in generating lightning probabilities. It also warrants caution when using this model as a first flash tool. We recommend using this tool as an additional insight for lightning prediction, and when lower probabilities are predicted to investigate further for a potential lightning watch or warning.

To understand the drivers of the model, 15 different ablation experiments were presented investigating how the model performs when the model inputs are not available. It was revealed here that the EFM suite and the MRMS data played a large role in generating the AUC-PR, with the HRRR and the GLM data not having as strong of influence. This is a unique result in that it quantifies and qualifies the EFM suite

for use as a lightning predictor, in direct conflict with other AI-based EFM predictor studies employed over Kennedy Space Center (Skrovan, 2020).

Additionally, ALE was calculated for one site to demonstrate the local effect on one field mill site. This accommodated each field mill’s separate input into the model, with its own normalization statistics. This results in different quantiles for each EFM site’s ALE calculations. Thus, if a cumulative ALE effect for the entire EFM suite is desired, interpolations between quantiles for each site would need to occur, potentially diluting the ALE contributed at each bin. ALE maps can be generated, but might be harder to interpret given the small target domain, low signal magnitude, and cumulative nature of the ALE toward for the larger bins.

The main contribution of this work is integrating the EFM data with other geo-spatial data in a skillful AI model. Here, the EFM data were re-sampled at a 30-second temporal resolution, twice the temporal resolution than other studies using EFM data (Bao et al., 2022; Hill, 2018; Speranza, 2019). Previous work was limited to using 1-minute averaged EFM data only. Additionally, the EFM model inputs were kept as time-series data, and not pre-processed into an image, allowing more EFM data to be used and not dropped as a result of quality control. Other EFM-based ML models only used the EFM data as inputs to predict the binary classification of lightning at a specific location. The ablation study captured the relative importance of the model inputs proving the MRMS and EFM inputs have the greatest impact on model performance. This ablation study provides valuable statistics to operators, indicating what happens to the model when specific data are not included - mimicking operational data outages. Finally, the largest contribution was creating a 15-minute temporal resolution lightning prediction tool that has comparable skill out to 30 minutes, when compared to LightningCast (Cintineo et al., 2022). LightningCast’s operational prowess is great, but its temporal resolution does not meet the temporal resolution dictated by LLCC

rule 6, and can generate partial to no lead time for resource protection at space ports and airfields, alike.

Future work includes adjusting the model architecture to capture the spatial dependencies better. Given the spatial disparities between the gridded radar data and the single-site EFM data, we recommend pursuing a more complex graph neural network or transformers to accommodate. This graph neural network may allow for the model to discriminate between in-cloud and cloud-to-ground lightning presented on the MERLIN system. This discrimination was an initial goal and explored heavily, however, the target images became sparse, and mimicked noise and became hard for the model to converge on a solution during training. Pursuing a graph neural network may also allow other disparate surface observations from wind towers and balloons to be included as inputs, allowing geo-spatial relationships to be established between each sensor type. Additionally, given the suite of EFMs that consistently measure the cumulative electric field overhead (a proxy for charge), LaunchCast domain may be used to establish a loss function that penalizes the model for not conserving charge, a current limitation of this work. Finally, with the local effect documented here by one field mill site, it is worth exploring a model with only an EFM encoder, then aggressively investigating the negative ALE generated by the large amplitude positive fields. This could be due to the anomalous nature of the fields, the electric field reversal due to the approach and departure of the storms, or potential positive cloud-to-ground lightning occurrence. A robust EFM encoder mimicking the one presented here, with more aggressive explainability could potentially tease this out, and potentially create the demand for EFM data to be assimilated into traditional NWP.

6.2.3 HorusCast

HorusCast demonstrated the ability to merge two state-of-the-art technologies, AI/ML and a fully digital phased array radar, to skillfully estimate K_{DP} . HorusCast provides a skillful estimate of K_{DP} when compared to the Vulpiani method, for Horus' 1-D mechanically rotated scanning strategy. When the model is transitioned to the 2-D sector scanning strategy, with finer temporal and range resolution, the model struggles to regenerate the Vulpiani target. A gradient-based method indicated the model was able to learn where the storm was located, with higher magnitude IxG values along the storm boundaries for Φ_{DP} and range. Within the storm boundaries, these values were reduced to near zero.

GPU-based calculations are generally known to be faster when compared to the same calculation on a CPU. Comparing the GPU-based HorusCast model in its current state with the 8 masks and 8 radar variable inputs isn't a fair comparison to a CPU-only Py-ART Vulpiani calculation, which only operates on the singular Φ_{DP} field. I recommend re-training the model with only the Φ_{DP} and range fields, to generate a fair computational time comparison. The other variables were used to provide the baseline, for when a temporal prediction is explored.

I originally hypothesized that the higher scanning frequency of Horus could generate a K_{DP} prediction with 1-3 minute (2-6 radar scans) lead time. This may still be possible, even with the current HorusCast dataset. The model architecture in Figure 5.4 would need to transition to a singular ConvLSTM1D layer in TensorFlow, also modifying what is declared a training sample. Several 1-D rays in time need to be stacked into one sample, then trained to the Vulpiani target a set number of radar scans ahead. This is why the other traditional and polarimetric variables in addition to Φ_{DP} are included, as it is hypothesized they may aid in the prediction. Note, generating

the samples in this manner may decrease the total number of training samples. This is a trade-off of increasing the model complexity for a prediction in time.

Another ML consideration is the use of a 2-D CNN, rather than a 1-D model. The primary limitation of the 2-D CNN is in the polar geometry, where unraveling the data in azimuth could present challenges. A CoordConv layer could be used, where the radar-centered Cartesian grids are also treated as model inputs, to allow the sharing of information across the azimuthal boundary. But, again, training sample size could decrease by one to two orders of magnitude.

Long-term, I think it is important to expand the dataset beyond the intermittent Horus deployments to the WSR-88D network. The data archive at the National Center for Environmental Information can be leveraged to do so, in varying storm types and conditions, to increase the generalization of the model beyond severe convection - a bias of the existing model.

Additionally, the research-to-operations pathway is uncertain for HorusCast. It is currently unknown where in the data generation pipeline the model could be deployed - is it on the truck or in post-processing? If a HorusCast-like model converges on WSR-88D data, can the Operational Radar Product Generator handle the ML model inference? While personal computers can deploy and analyze the HorusCast model, some HorusCast users may not have the required compute.

6.3 The End

I thank the reader for considering the work presented in this dissertation. I've demonstrated the skill and performance of ML models' capability to predict and locate lightning strikes and generate specific differential phase in storms. Overall, ML is very capable in discerning lightning's potential given either observational, model, or remotely

sensed data, and should continue to be explored aggressively.

Appendix A

BoltCast Supplemental Material

Line by line model architectures and simpler experiments or side projects that might be worth discussing.

A.1 Further Architecture Details

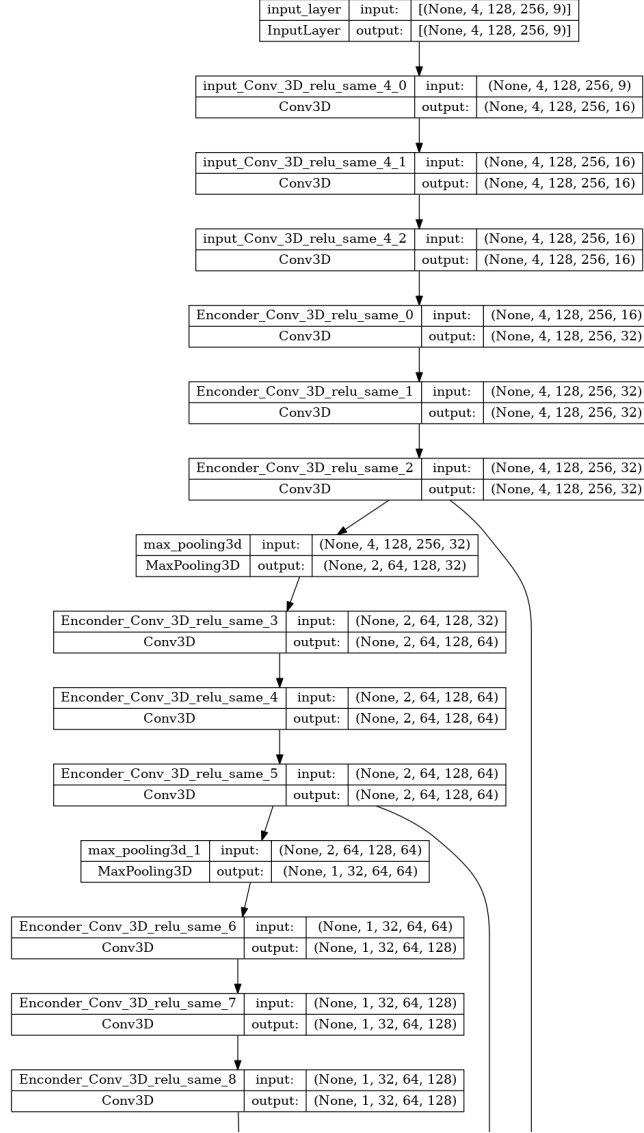


Figure A.1: Line-by-line diagram of the encoder for the BoltCast U-Net.

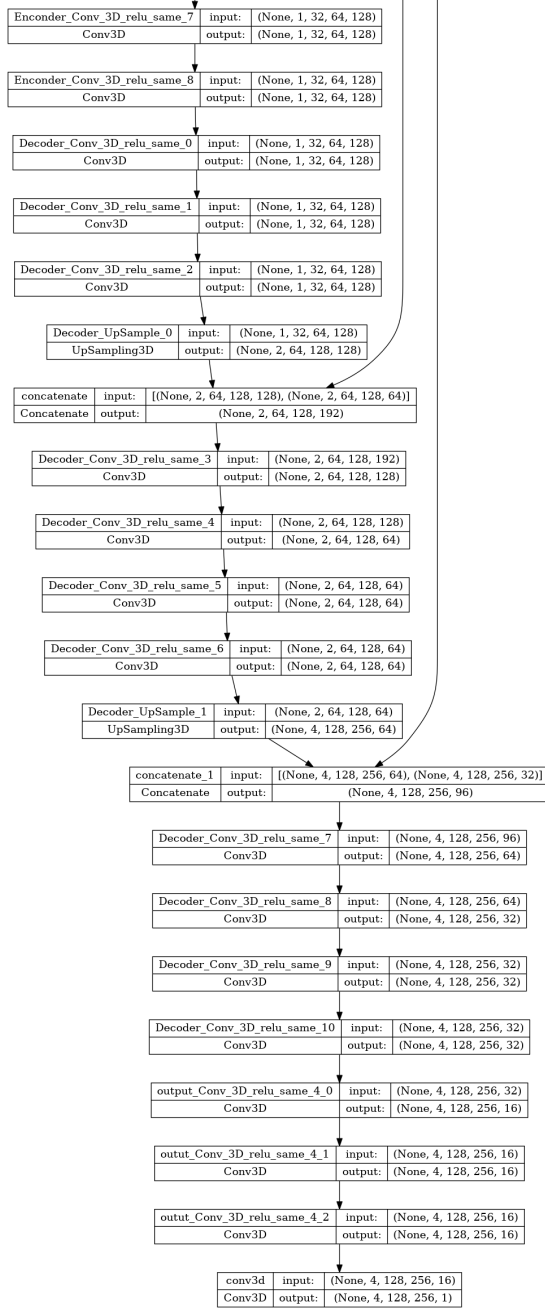


Figure A.2: Line-by-line diagram of the decoder for the BoltCast U-Net model.

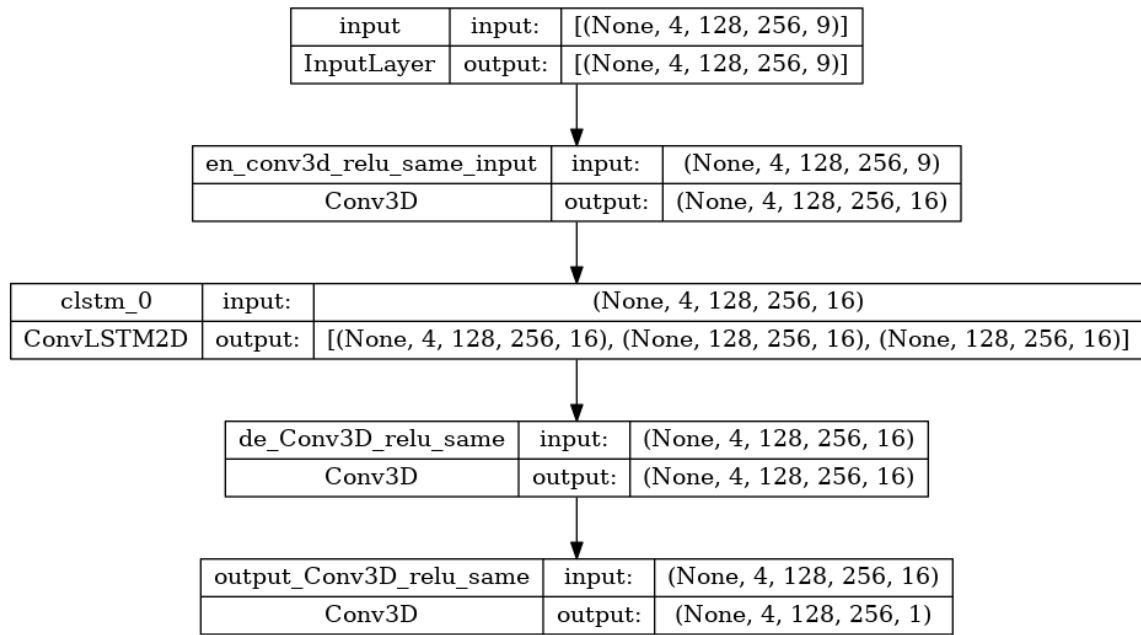


Figure A.3: Line-by-line diagram of the BoltCast LSTM model.

Appendix B

LaunchCast Supplemental Material

B.1 Ablation Statistics Table

This section highlights the specific ablation statistics that were calculated for the 15 different ablation studies. The AUC-PR values presented in Table B.3 are presented in Figure 4.10. Experiment 0 is when all data are put into the model. An excel document resides in the GitHub repository for transparency with a higher number of significant figures, and easier interrogation for the reader.

Table B.1: The maximum CSI thresholds for the ablation experiment.

Exp. No.	Thresh (t0)	Thresh (t1)	Thresh (t2)	Thresh (t3)
0	0.20	0.20	0.20	0.15
1	0.05	0.05	0.05	0.05
2	0.05	0.05	0.05	0.05
3	0.05	0.05	0.05	0.05
4	0.20	0.20	0.20	0.15
5	0.05	0.05	0.05	0.05
6	0.05	0.05	0.05	0.05
7	0.05	0.05	0.05	0.05
8	0.05	0.05	0.05	0.05
9	0.05	0.05	0.05	0.05
10	0.05	0.05	0.05	0.05
11	0.05	0.05	0.05	0.05
12	0.05	0.05	0.05	0.05
13	0.05	0.05	0.05	0.05
14	0.05	0.05	0.05	0.05
15	0.05	0.05	0.05	0.05

Table B.2: The maximum CSI values and accompanying values for the ablation experiment.

Exp. No.	Max CSI (t0)	Max CSI (t1)	Max CSI (t2)	Max CSI (t3)
0	0.34	0.30	0.26	0.22
1	0.29	0.25	0.21	0.19
2	0.02	0.03	0.06	0.10
3	0.01	0.01	0.01	0.02
4	0.33	0.30	0.25	0.22
5	0.26	0.23	0.20	0.17
6	0.02	0.02	0.05	0.09
7	0.01	0.01	0.01	0.02
8	0.17	0.14	0.11	0.11
9	0.01	0.01	0.01	0.01
10	0.01	0.01	0.01	0.01
11	0.01	0.00	0.00	0.01
12	0.10	0.08	0.07	0.08
13	0.01	0.00	0.01	0.01
14	0.01	0.01	0.01	0.01
15	0.01	0.00	0.00	0.01

Table B.3: The AUC-PR for the ablation experiments.

Exp. No.	AUC-PR (t0)	AUC-PR (t1)	AUC-PR (t2)	AUC-PR (t3)
0	0.47	0.40	0.33	0.26
1	0.34	0.28	0.23	0.19
2	0.10	0.11	0.12	0.14
3	0.09	0.09	0.10	0.10
4	0.47	0.40	0.32	0.26
5	0.32	0.26	0.21	0.18
6	0.10	0.10	0.12	0.13
7	0.09	0.09	0.09	0.09
8	0.22	0.18	0.15	0.13
9	0.09	0.09	0.09	0.09
10	0.09	0.09	0.09	0.09
11	0.09	0.09	0.09	0.09
12	0.15	0.14	0.12	0.11
13	0.09	0.09	0.09	0.09
14	0.09	0.09	0.09	0.09
15	0.09	0.09	0.09	0.09

Table B.4: The FSS for a 5-nautical mile window for the ablation experiments.

Exp. No.	FSS-5 (t0)	FSS-5 (t1)	FSS-5 (t2)	FSS-5 (t3)
0	0.67	0.63	0.56	0.49
1	0.34	0.24	0.16	0.19
2	0.01	0.01	0.01	0.04
3	0.00	0.00	0.00	0.00
4	0.68	0.63	0.56	0.48
5	0.26	0.19	0.13	0.15
6	0.01	0.00	0.00	0.02
7	0.00	0.00	0.00	0.00
8	0.08	0.04	0.01	0.01
9	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00
12	0.02	0.01	0.00	0.00
13	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00

Table B.5: The FSS for a 10 nautical-mile window for the ablation experiments.

Exp. No.	FSS-10 (t0)	FSS-10 (t1)	FSS-10 (t2)	FSS-10 (t3)
0	0.72	0.67	0.62	0.54
1	0.38	0.27	0.20	0.23
2	0.01	0.01	0.01	0.04
3	0.00	0.00	0.00	0.00
4	0.73	0.68	0.62	0.54
5	0.29	0.22	0.16	0.19
6	0.01	0.00	0.00	0.02
7	0.00	0.00	0.00	0.00
8	0.08	0.04	0.01	0.01
9	0.00	0.00	0.00	0.00
10	0.00	0.00	0.00	0.00
11	0.00	0.00	0.00	0.00
12	0.02	0.01	0.00	0.01
13	0.00	0.00	0.00	0.00
14	0.00	0.00	0.00	0.00
15	0.00	0.00	0.00	0.00

Table B.6: The AUC-ROC for the ablation experiments.

Exp. No.	AUC-ROC (t0)	AUC-ROC (t1)	AUC-ROC (t2)	AUC-ROC (t3)
0	0.88	0.85	0.82	0.78
1	0.69	0.66	0.64	0.63
2	0.51	0.51	0.53	0.55
3	0.50	0.50	0.50	0.50
4	0.87	0.85	0.81	0.78
5	0.67	0.64	0.62	0.62
6	0.51	0.51	0.52	0.54
7	0.50	0.50	0.50	0.50
8	0.60	0.57	0.56	0.56
9	0.50	0.50	0.50	0.50
10	0.50	0.50	0.50	0.50
11	0.50	0.50	0.50	0.50
12	0.55	0.54	0.53	0.53
13	0.50	0.50	0.50	0.50
14	0.50	0.50	0.50	0.50
15	0.50	0.50	0.50	0.50

Table B.7: The probability of detection (POD) for the ablation experiments.

Exp. No.	POD (t0)	POD (t1)	POD (t2)	POD (t3)
0	0.64	0.60	0.53	0.58
1	0.43	0.37	0.34	0.35
2	0.03	0.03	0.07	0.14
3	0.01	0.00	0.01	0.02
4	0.59	0.56	0.49	0.55
5	0.37	0.31	0.29	0.31
6	0.02	0.03	0.06	0.12
7	0.01	0.01	0.01	0.02
8	0.22	0.17	0.15	0.20
9	0.01	0.01	0.01	0.02
10	0.01	0.01	0.01	0.02
11	0.01	0.00	0.00	0.01
12	0.12	0.10	0.09	0.14
13	0.01	0.00	0.01	0.02
14	0.01	0.01	0.01	0.02
15	0.01	0.00	0.00	0.01

Table B.8: The false alarm ratio for the ablation experiments.

Exp. No.	F.A.Ratio (t0)	F.A.Ratio (t1)	F.A.Ratio (t2)	F.A.Ratio (t3)
0	0.58	0.62	0.66	0.74
1	0.54	0.56	0.63	0.72
2	0.75	0.66	0.69	0.76
3	0.91	0.62	0.74	0.89
4	0.57	0.61	0.66	0.74
5	0.53	0.55	0.63	0.72
6	0.79	0.70	0.71	0.77
7	0.93	0.63	0.73	0.89
8	0.60	0.63	0.72	0.81
9	0.77	0.78	0.86	0.91
10	0.92	0.89	0.90	0.91
11	0.94	—	0.88	0.92
12	0.65	0.66	0.75	0.84
13	0.79	0.81	0.87	0.91
14	0.93	0.89	0.91	0.91
15	0.94	—	—	0.92

ALE

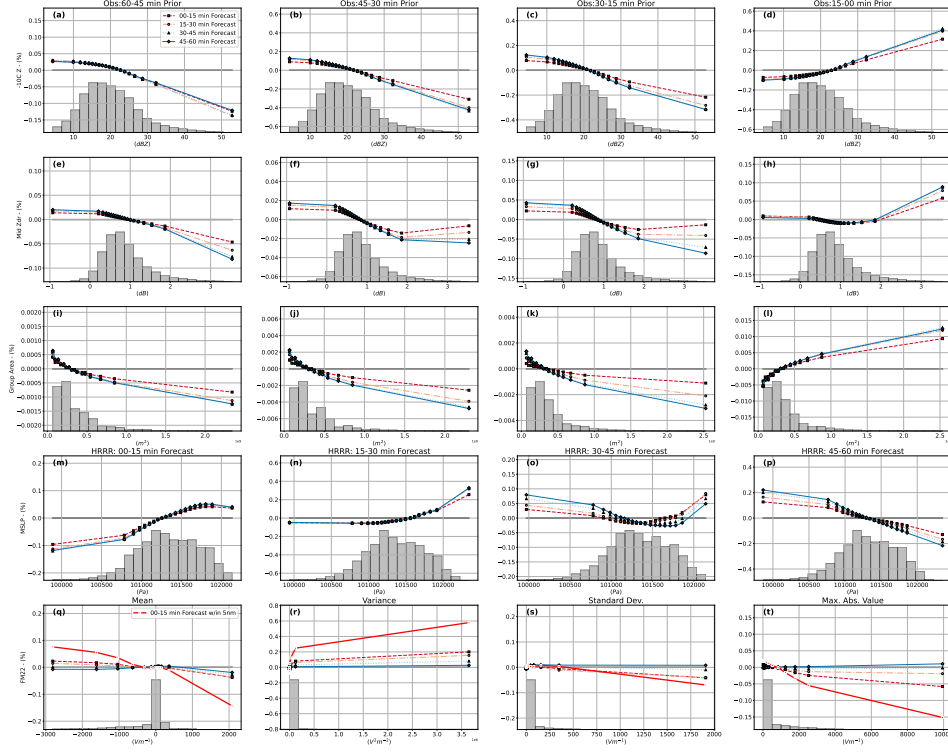


Figure B.1: ALE plots for the inputs for -10C reflectivity (a-d), mid level Zdr (e-h), group area (i-l), for the different input times. Panels (m-p) showcase ALE for the HRRR MSLP predictions. The bottom row, panels (q-t) provide ALE plots for field mill 22 inputs for the entire previous hour. Scaled histograms are provided to depict a proxy of the distribution of each variable. Each line corresponds to a specific lead time, indicated by the legend in panels (a) and (q).

Appendix C

HorusCast Supplemental Material



Figure C.1: Part 1 of the HorusCast Encoder.



Figure C.2: Part 2 of the HorusCast Encoder.



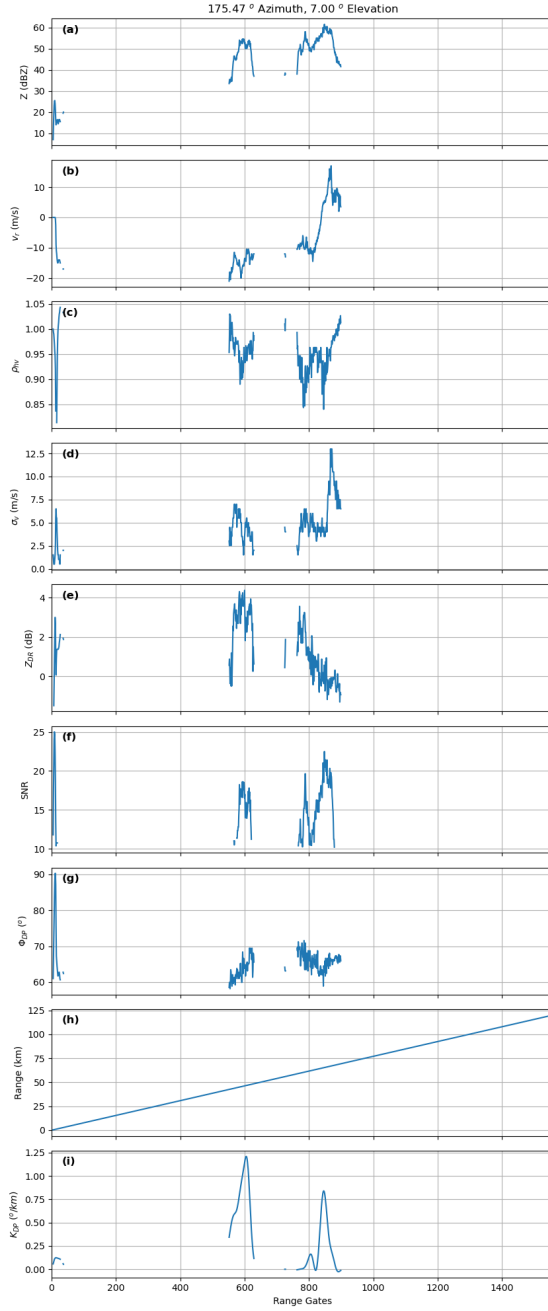


Figure C.5: A sample ray diagram, plotted similar to one plotted in Mishler et al. (2024). This ray is valid 22:00:07 UTC 03 June 2025, at 7 degree elevation and 175.47 degree azimuth.

Reference List

- Abadi, M., and A. Agarwal, 2015: Tensorflow: Large-scale machine learning on heterogeneous distributed systems, technical Manual.
- Abadi, M., and Coauthors, 2015: TensorFlow: Large-scale machine learning on heterogeneous systems. Keras, URL <https://www.tensorflow.org/>, software available from tensorflow.org.
- Agarap, A. F. M., 2019: Deep learning using rectified linear units (relu), arxiv.
- Aich, V., R. Holzworth, S. Goodman, Y. Kuleshov, C. Price, and E. Williams, 2018: Lightning: A new essential climate variable. *Eos (Washington, D.C.)*, **99**, <https://doi.org/10.1029/2018EO104583>.
- Ancona, M., E. Ceolini, C. Öztireli, and M. H. Gross, 2017: Towards better understanding of gradient-based attribution methods for deep neural networks. *International Conference on Learning Representations*, URL <https://api.semanticscholar.org/CorpusID:3728967>.
- Apley, D. W., and J. Zhu, 2020: Visualizing the effects of predictor variables in black box supervised learning models. *Journal of Royal Statistical Society*, **88**, 1059–1086.
- Bao, R., Y. Zhang, B. J. Ma, Z. Zhang, and Z. He, 2022: An artificial neural network for lightning prediction based on atmospheric electric field observations. *Remote Sensing*, **14** (17), <https://doi.org/10.3390/rs14174131>.
- Barnes, A. A., 1991: Predicting cloud-to-ground lightning with neural networks. NASA.
- Barnes, J. A., and D. Frankel, 1990: Investigation of the prediction of lightning strikes using neural networks. NASA, <https://doi.org/10.2514/6.1990-484>.
- Bazelyan, E. M., N. L. Aleksandrov, Y. P. Raizer, and A. M. Konchakov, 2007: The effect of air density on atmospheric electric fields required for lightning initiation from a long airborne object. *Atmospheric Research*, **86** (2), 126–138, <https://doi.org/10.1016/j.atmosres.2007.04.001>.
- Biggerstaff, M. I., Z. Zounes, A. Addison Alford, G. D. Carrie, J. T. Pilkey, M. A. Uman, and D. M. Jordan, 2017: Flash propagation and inferred charge structure relative to radar-observed ice alignment signatures in a small Florida mesoscale con-

- vective system. *Geophysical Research Letters*, **44** (15), 8027–8036, <https://doi.org/10.1002/2017gl074610>.
- Blaylock, B. K., and J. Horel, 2020: Comparison of lightning forecasts from the high-resolution rapid refresh model to geostationary lightning mapper observations. *Weather and Forecasting*, **35**, 401–416, <https://doi.org/10.1175/WAF-D-19-0141.1>.
- Boccippio, D. J., S. Heckman, and S. J. Goodman, 2001: A diagnostic analysis of the kennedy space center LDAR network: 1. data characteristics. *Journal of Geophysical Research: Atmospheres*, **106** (D5), 4769–4786, <https://doi.org/https://doi.org/10.1029/2000JD900687>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900687>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2000JD900687>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32.
- Bright, D. R., M. S. Wandishin, R. E. Jewell, and S. J. Weiss, 2005: A physically based parameter for lightning prediction and its calibration in ensemble forecasts. NOAA.
- Bruning, E. C., and D. R. MacGorman, 2013: Theory and observations of controls on lightning flash size spectra. *Journal of the Atmospheric Sciences*, **70** (12), 4012–4029, <https://doi.org/10.1175/jas-d-12-0289.1>.
- Bruning, E. C., W. D. Rust, T. J. Schuur, D. R. MacGorman, P. R. Krehbiel, and W. Rison, 2007: Electrical and polarimetric radar observations of a multicell storm in tex. *Monthly Weather Review*, **135** (7), 2525–2544, <https://doi.org/10.1175/mwr3421.1>.
- Bruning, E. C., S. A. Weiss, and K. M. Calhoun, 2014: Continuous variability in thunderstorm primary electrification and an evaluation of inverted-polarity terminology. *Atmospheric Research*, **135**, 274–284, <https://doi.org/10.1016/j.atmosres.2012.10.009>, URL <https://www.researchgate.net/publication/266100030203600023>, [bruning, Eric C. Weiss, Stephanie A. Calhoun, Kristin M. Bruning, Eric/0000-0003-1959-442X; Calhoun, Kristin/0000-0003-2858-256X 1873-2895](https://www.researchgate.net/publication/266100030203600023).
- Bruning, E. C., and Coauthors, 2019: Meteorological imagery for the geostationary lightning mapper. *Journal of Geophysical Research: Atmospheres*, **124** (24), 14 285–14 309, <https://doi.org/10.1029/2019jd030874>.
- Byers, H. R., and R. R. Braham, 1949: The thunderstorm: report of the thunderstorm project. US Weather Bureau.
- Carlin, J., 2026: NWS operational KDP calculation. National Severe Storms Lab.

- Caylor, I. J., and V. Chandrasekar, 1996: Time-varying ice crystal orientation in thunderstorms observed with multiparameter radar. *IEEE Trans. Geosci. Remote. Sens.*, <https://doi.org/10.1109/36.508402>, URL <https://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=508402&ref=>.
- Cecil, D. J., D. E. Buechler, and R. J. Blakeslee, 2014: Gridded lightning climatology from TRMM-LIS and OTD: Dataset description. *Atmospheric Research*, **135–136**, 404–414, <https://doi.org/10.1016/j.atmosres.2012.06.028>.
- Chase, R. J., D. R. Harrison, G. M. Lackmann, and A. McGovern, 2023: A machine learning tutorial for operational meteorology. Part II: Neural networks and deep learning. *Weather and Forecasting*, **38** (8), 1271–1293, <https://doi.org/https://doi.org/10.1175/WAF-D-22-0187.1>, URL <https://journals.ametsoc.org/view/journals/wefo/38/8/WAF-D-22-0187.1.xmlhttps://journals.ametsoc.org/downloadpdf/journals/wefo/38/8/WAF-D-22-0187.1.pdf>.
- Chmielewski, V. C., J. Blair, D. Kennedy, D. MacGorman, and K. M. Calhoun, 2022: A comparison of processing methods for the Oklahoma lightning mapping array. *Earth and Space Science*, **9** (7), e2021EA002081, <https://doi.org/https://doi.org/10.1029/2021EA002081>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EA002081>, e2021EA002081 2021EA002081, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021EA002081>.
- Christian, H., R. Blakeslee, S. Goodman, W. Koshak, and D. Mach, 1998: Science data validation plan for the lightning imaging sensor (LIS). Report, Earth Science Department, Science Directorate NASA/Marshall Space Flight Center.
- Christian, H. J., V. Mazur, B. D. Fisher, L. H. Ruhnke, and K. Crouch, 1989: The Atlas/Centaur lightning strike incident. *Journal of Geophysical Research*, **94** (D11), 13 169–13 177, <https://doi.org/10.1029/JD094iD11p13169>.
- Cintineo, J. L., M. J. Pavolonis, L. Heuscher, and J. M. Sieglaff, 2026: The impact of radar reflectivity data in a satellite-based lightning nowcasting model. *Weather and Forecasting*, **41** (1), 149 – 168, <https://doi.org/10.1175/WAF-D-25-0067.1>, URL <https://journals.ametsoc.org/view/journals/wefo/41/1/WAF-D-25-0067.1.xml>.
- Cintineo, J. L., M. J. Pavolonis, and J. M. Sieglaff, 2022: Probsevere lightningcast: A deep-learning model for satellite-based lightning nowcasting. *Weather and Forecasting*, **37** (7), 1239–1257, <https://doi.org/https://doi.org/10.1175/WAF-D-22-0019.1>, URL <https://journals.ametsoc.org/view/journals/wefo/37/7/WAF-D-22-0019.1.xmlhttps://journals.ametsoc.org/downloadpdf/journals/wefo/37/7/WAF-D-22-0019.1.pdf>.

- Cizek, B., 2026: Personal communication with 45th weather squadron. USSF.
- Clark, A. J., K. A. Hoogewind, A. J. Hill, E. D. Loken, and M. J. Hosek, 2025: Extended range machine learning severe weather guidance based on the operational GEFS. *Weather and Forecasting*, **40**, 1325 – 1343, <https://doi.org/10.1175/WAF-D-24-0116.1>, URL <https://journals.ametsoc.org/view/journals/wefo/40/8/WAF-D-24-0116.1.xml>.
- Cummings, K. A., and Coauthors, 2024: Evaluation of lightning flash rate parameterizations in a cloud-resolved WRF-Chem simulation of the 29–30 may 2012 oklahoma severe supercell system observed during dc3. *Journal of Geophysical Research: Atmospheres*, **129** (11), <https://doi.org/10.1029/2023jd039492>.
- Cummins, K. L., C. J. Biagi, A. Nag, M. N. Plaisir, A. Tempert, and H. K. Rassoul, 2025: On the benefits of tower-mounted measurement of near-surface quasi-static electric fields. *Journal of Atmospheric and Oceanic Technology*, **42** (10), 1281–1297, <https://doi.org/10.1175/jtech-d-25-0010.1>.
- Dahl, J. M. L., H. Höller, and U. Schumann, 2011: Modeling the flash rate of thunderstorms. Part I: Framework. *Monthly Weather Review*, **139**, 3091–3111, <https://doi.org/10.1175/MWR-D-10-05031.1>.
- Deloach, W. R., and J. W. Pellicciotti, 2023: NASA standard for lightning launch commit criteria. NASA.
- Doviak, R., and D. Zrnic, 1993: *Doppler Radar and Weather Observations*. 2nd ed., Dover Publications, Inc.
- Dowell, D. C., and Coauthors, 2022: The high-resolution rapid refresh (hrrr): An hourly updating convection-allowing forecast model. Part I: Motivation and system description. *Weather and Forecasting*, **37** (8), 1371–1395, <https://doi.org/10.1175/waf-d-21-0151.1>.
- Earnest, B. L., A. McGovern, C. Karstens, and I. Jirak, 2024a: Part I: Improving wildfire occurrence prediction for conus using deep learning and fire weather variables. *Artificial Intelligence for the Earth Systems*, **3** (3), <https://doi.org/10.1175/aies-d-23-0057.1>.
- Earnest, B. L., A. McGovern, C. Karstens, and I. Jirak, 2024b: Part II: Lessons learned from predicting wildfire occurrence for conus using deep learning and fire weather variables. *Artificial Intelligence for the Earth Systems*, **3** (3), <https://doi.org/10.1175/aies-d-23-0058.1>.

- Essa, Y., R. Ajoodha, and H. G. Hunt, 2020: A LSTM recurrent neural network for lightning flash prediction within southern africa using historical time-series data. *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, 1–6, <https://doi.org/10.1109/CSDE50874.2020.9411544>.
- Fairman, S., 2020: The detection of continuing current in lightning using the geostationary lightning mapper and exploring its relationship to lightning-initiated wildfires. Thesis, University of Alabama in Huntsville.
- Fierro, A. O., E. R. Mansell, D. R. MacGorman, and C. L. Ziegler, 2013: The implementation of an explicit charging and discharge lightning scheme within the WRF-ARW model: Benchmark simulations of a continental squall line, a tropical cyclone, and a winter storm. *Monthly Weather Review*, **141** (7), 2390–2415, <https://doi.org/10.1175/mwr-d-12-00278.1>.
- Flora, M. L., C. K. Potvin, A. McGovern, and S. Handler, 2024: A machine learning explainability tutorial for atmospheric sciences. *Artificial Intelligence for the Earth Systems*, **3** (1), <https://doi.org/10.1175/aies-d-23-0018.1>.
- Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, **29**, 1189–1232.
- Geng, Y., and Coauthors, 2021: A deep learning framework for lightning forecasting with multi-source spatiotemporal data. *Quarterly Journal of the Royal Meteorological Society*, **147** (741), 4048–4062, <https://doi.org/10.1002/qj.4167>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology and Climatology*, **11** (8), 1203 – 1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2), URL https://journals.ametsoc.org/view/journals/apme/11/8/1520-0450_1972_011_1203_tuomos_2_0_co_2.xml.
- Goede, V., D. Schwartzman, D. Bodine, V. C. Chmielewski, T. Yu, E. Bruning, M. Tzeng, and M. Stock, 2025: Rapid sampling and polarimetric statistics of lightning with a phased array radar. *Geophysical Research Letters*, **52** (2), <https://doi.org/10.1029/2024gl112193>.
- Goodman, S., 2012: Algorithm theoretical basis document: GLM lightning cluster filter algorithm. NOAA.
- Grandell, J., and Coauthors, 2010: EUMETSAT meteosat third generation (MTG) lightning imager: From mission requirements to product development. *AGU 2010 Annual Meeting*, Vol. -1.

- Gremillion, M. S., and R. E. Orville, 1998: Thunderstorm characteristics of cloud-to-ground lightning at the Kennedy Space Center, Florida: A study of lightning initiation signatures as indicated by the wsr-88d. *Weather and forecasting*, **14**, 640–649.
- Gutman, A. J., W. P. Roeder, A. J. Geyer, E. D. White, and D. L. Sanderson, 2019: Optimizing the lightning warning radii at spaceport Florida. *Weather and Forecasting*, **35** (2), 523–536, <https://doi.org/10.1175/waf-d-19-0129.1>.
- Han, L., H. Liang, H. Chen, W. Zhang, and Y. Ge, 2022: Convective precipitation nowcasting using u-net model. *IEEE Transactions on Geoscience and Remote Sensing*, **60**, 1–8, <https://doi.org/10.1109/TGRS.2021.3100847>.
- Handel, S. C., K. L. Cummins, and E. P. Krider, 2022: Surface potential gradients and nexrad radar reflectivities before the onset of lightning at the ksc-er. *Journal of Geophysical Research: Atmospheres*, **127** (18), <https://doi.org/10.1029/2022jd036681>.
- Harrison, D. R., M. S. Elliott, I. L. Jirak, and P. T. Marsh, 2022: Utilizing the high-resolution ensemble forecast system to produce calibrated probabilistic thunderstorm guidance. *Weather and Forecasting*, **37** (7), 1103–1115, <https://doi.org/10.1175/waf-d-22-0001.1>.
- Haupt, S. E., W. Chapman, S. V. Adams, C. Kirkwood, J. S. Hosking, N. H. Robinson, S. Lerch, and A. C. Subramanian, 2020a: Towards implementing ai post-processing in weather and climate: Proposed actions from the oxford 2019 workshop. *Philos. Trans. Roy. Meteor. Soc.*, **A379**, 20200 091.
- Haupt, S. E., and Coauthors, 2022: The history and practice of ai in the environmental sciences. *Bulletin of the American Meteorological Society*, **103** (5), E1351 – E1370, <https://doi.org/10.1175/BAMS-D-20-0234.1>, URL <https://journals.ametsoc.org/view/journals/bams/103/5/BAMS-D-20-0234.1.xml>.
- Helmus, J. J., and S. M. Collis, 2016: The python arm radar toolkit (py-art), a library for working with weather radar data in the python programming language. *Journal of Open Research Software*, <https://doi.org/10.5334/jors.119>.
- Hilburn, K. A., I. Ebert-Uphoff, and S. D. Miller, 2021: Development and interpretation of a neural-network-based synthetic radar reflectivity estimator using goes-r satellite observations. *Journal of Applied Meteorology and Climatology*, **60** (1), 3 – 21, <https://doi.org/10.1175/JAMC-D-20-0084.1>, URL <https://journals.ametsoc.org/view/journals/apme/60/1/jamc-d-20-0084.1.xml>.
- Hill, A. J., R. S. Schumacher, and I. L. Jirak, 2023: A new paradigm for medium-range severe weather forecasts: Probabilistic random forest-based predictions. *Weather*

- and Forecasting, **38**, 251 – 272, <https://doi.org/10.1175/WAF-D-22-0143.1>, URL <https://journals.ametsoc.org/view/journals/wefo/38/2/WAF-D-22-0143.1.xml>.
- Hill, D., 2018: Evaluating the efficacy of electrical field mills to predict lightning events. *25th International Lightning Detection Conference and 7th International Lightning Meteorology Conference*.
- Hochreiter, S., and J. Schmidhuber, 1997: Long-short term memory. *Neural Computation*, **9**, 1735–1780.
- Holland, K., 2021: Ccsfs/ksc total lightning warning radii optimization for merlin using pre-existing lightning aress. Thesis, Engineering Physics.
- Holle, R. L., 2016: A summary of recent national-scale lightning fatality studies. *Weather, climate, and society*, **8** (1), 35–42, <https://doi.org/10.1175/wcas-d-15-0032.1>.
- Holle, R. L., R. E. López, and B. C. Navarro, 2005: Deaths, injuries, and damages from lightning in the United States in the 1890s in comparison with the 1990s. *Journal of applied meteorology (1988)*, **44** (10), 1563–1573, <https://doi.org/10.1175/JAM2287.1>.
- Hopfield, J. J., 1982: Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, **79** (8), 2554–2558, <https://doi.org/10.1073/pnas.79.8.2554>, URL <https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554>, <https://www.pnas.org/doi/pdf/10.1073/pnas.79.8.2554>.
- Howard, K., and Coauthors, 2016: Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97** (4), 621–638, <https://doi.org/10.1175/bams-d-14-00174.1>.
- Hsu, W.-r., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *International Journal of Forecasting*, **2**, 285–293.
- Huang, Y., J. Wang, L. Cai, Y. Fan, H. Wang, and T. Zhang, 2024: Relationship between radar reflectivity thresholds and very low frequency/low frequency total lightning for thunderstorm identification. *High Voltage*, **9** (5), 1068–1080, <https://doi.org/10.1049/hve2.12441>.
- Hui, W., W. Zhang, W. Lyu, and P. Li, 2020: Preliminary observations from the china fengyun-4a lightning mapping imager and its optical radiation characteristics.

- Remote Sensing*, **12** (16), <https://doi.org/10.3390/rs12162622>, URL <https://www.mdpi.com/2072-4292/12/16/2622>.
- Jacobson, E. A., and E. Krider, 1976: Electrostatic field changes produced by Florida lightning. *Journal of the Atmospheric Sciences*, **33**, 103–116.
- Justin, A. D., A. McGovern, and J. T. Allen, 2025: Frontfinder ai: Efficient identification of frontal boundaries over the continental United States and noaa’s unified surface analysis domain using the unet3+ model architecture. *Artificial Intelligence for the Earth Systems*, **4** (1), <https://doi.org/10.1175/aies-d-24-0043.1>.
- Justin, A. D., C. Willingham, A. McGovern, and J. T. Allen, 2023: Toward operational real-time identification of frontal boundaries using machine learning. *Artificial Intelligence for the Earth Systems*, **2** (3), <https://doi.org/10.1175/aies-d-22-0052.1>.
- Kingma, D. P., and J. Ba, 2014: Adam: A method for stochastic optimization. *ArXiv*.
- Kobayashi, M., and N. Kitagawa, 1956: Recording apparatuses for use in studies of field-change due to lightning discharge. Meteorological Research Institute.
- Koshak, W. J., 1990: Analysis of lightning field changes produced by Florida thunderstorms. *NASA Tech Manual 103539*.
- Koshak, W. J., and Coauthors, 2004: North alabama lightning mapping array (lma): Vhf source retrieval algorithm and error analyses. *Journal of Atmospheric and Oceanic Technology*, **21** (4), 543 – 558, [https://doi.org/10.1175/1520-0426\(2004\)021<0543:NALMAL>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0543:NALMAL>2.0.CO;2), URL https://journals.ametsoc.org/view/journals/atot/21/4/1520-0426_2004_021_0543_nalmal_2_0_co_2.xml.
- Kumjian, M. R., A. P. Khain, N. Benmoshe, E. Ilotoviz, A. V. Ryzhkov, and V. T. J. Phillips, 2014: The anatomy and physics of zdr columns: Investigating a polarimetric radar signature with a spectral bin microphysical model. *Journal of Applied Meteorology and Climatology*, **53** (7), 1820–1843, <https://doi.org/10.1175/jamc-d-13-0354.1>.
- Laird, N. F., D. A. Kristovich, R. M. Rubaer, H. T. Ochs, and L. J. Miller, 1995: The cape canaveral sea and river breezes: Kinematic structure and convective initiation. *Monthly Weather Review*, **123**.
- Lakshmanan, V., T. Smith, K. Hondl, G. J. Stumpf, and A. Witt, 2006: A real-time, three-dimensional, rapidly updating, heterogeneous radar merger technique for reflectivity, velocity, and derived products. *Weather and Forecasting*, **21** (5), 802 – 823, <https://doi.org/10.1175/WAF942.1>, URL https://journals.ametsoc.org/view/journals/wefo/21/5/waf942_1.xml.

- Lakshmanan, V., and Coauthors, 2016: Multi-radar multi-sensor (mrms) severe weather and aviation products: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97** (9), 1617–1630, <https://doi.org/10.1175/bams-d-14-00173.1>.
- Lay, E. H., R. H. Holzworth, C. J. Rodger, J. N. Thomas, O. Pinto Jr., and R. L. Dowden, 2004: Wwll global lightning detection system: Regional validation study in brazil. *Geophysical Research Letters*, **31** (3), <https://doi.org/10.1029/2003GL018882>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2003GL018882>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2003GL018882>.
- Lean, H. W., and N. M. Roberts, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, **136** (1), 78–97, <https://doi.org/10.1175/2007mwr2123.1>.
- Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998: Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86** (11), 2278–2324.
- Leinonen, J., U. Hamann, and U. Germann, 2022: Seamless lightning nowcasting with recurrent-convolutional deep learning. *Artificial Intelligence for the Earth Systems*, **1** (4), <https://doi.org/10.1175/aies-d-22-0043.1>.
- Li, L., K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, 2018: Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, **18**, 1–52.
- Liu, C., D. J. Cecil, E. J. Zipser, K. Kronfeld, and R. Robertson, 2012: Relationships between lightning flash rates and radar reflectivity vertical structures in thunderstorms over the tropics and subtropics. *Journal of Geophysical Research: Atmospheres*, **117** (D6), <https://doi.org/10.1029/2011jd017123>.
- Lopez, P., 2016: A lightning parameterization for the ecmwf integrated forecasting system. *Monthly Weather Review*, **144** (9), 3057–3075, <https://doi.org/10.1175/mwr-d-16-0026.1>.
- Lundberg, S., and S.-I. Lee, 2017: A unified approach to interpreting model predictions. ArXiv.
- MacGorman, D. R., and Coauthors, 2008: Telex the thunderstorm electrification and lightning experiment. *Bulletin of the American Meteorological Society*, **89** (7), 997–1013, <https://doi.org/10.1175/2007bams2352.1>, URL [GotoISI://WOS:000265614700006](https://doi.org/10.1175/2007bams2352.1), macGorman, Donald R. Rust, W. David Schuur, Terry

- J. Biggerstaff, Michael I. Straka, Jerry M. Ziegler, Conrad L. Mansell, Edward R. Bruning, Eric C. Kuhlman, Kristin M. Lund, Nicole R. Biermann, Nicholas S. Payne, Clark Carey, Larry D. Krehbiel, Paul R. Rison, William Eack, Kenneth B. Beasley, William H. MacGorman, DONALD/AAH-1726-2020; Biggerstaff, Michael I./N-4728-2019 MacGorman, DONALD/0000-0002-2395-8196; Biggerstaff, Michael I./0000-0002-6690-784X; Bruning, Eric/0000-0003-1959-442X; Calhoun, Kristin/0000-0003-2858-256X; Eack, Kenneth/0000-0001-7231-5990; Carey, Lawrence/0000-0003-2255-877X.
- Mamalakis, A., I. Ebert-Uphoff, and E. A. Barnes, 2022: Neural network attribution methods for problems in geoscience: A novel synthetic benchmark dataset. *Environmental Data Science*, **1**, <https://doi.org/10.1017/eds.2022.7>.
- Mansell, E. R., D. R. MacGorman, C. L. Ziegler, and J. M. Straka, 2005: Charge structure and lightning sensitivity in a simulated multicell thunderstorm. *Journal of Geophysical Research: Atmospheres*, **110** (D12), <https://doi.org/10.1029/2004JD005287>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005287https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2004JD005287?download=true>.
- Marion, J. W., E. Lanning, and D. G. Brown, 2026: Assessment of lightning mortality in the United States from 1979 to 2023. *Natural Hazards*, **122** (6), <https://doi.org/10.1007/s11069-025-07899-5>, URL <https://link.springer.com/content/pdf/10.1007/s11069-025-07899-5.pdf>.
- McCaul, E. W., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting*, **24** (3), 709–729, <https://doi.org/10.1175/2008waf2222152.1>.
- McClung, B., A. McGovern, A. J. Hill, D. Schwartzman, and M. Stock, 2026: Boltcast: Medium-range lightning prediction with neural and long short-term memory networks. *Artificial Intelligence for the Earth Systems*, **5** (2), <https://doi.org/10.1175/aies-d-25-0044.1>.
- McGovern, A., K. L. Elmore, D. J. Gagne, S. E. Haupt, C. D. Karstens, R. Lagerquist, T. Smith, and J. K. Williams, 2017: Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society*, **98** (10), 2073–2090, <https://doi.org/10.1175/bams-d-16-0123.1>.
- McGovern, A., R. Lagerquist, D. John Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, **100** (11), 2175–2199, <https://doi.org/10.1175/bams-d-18-0195.1>.

- Measaka, T., K. Iwanami, and M. Maki, 2012: Non-negative kdp estimation by monotone increasing phi dp assumption below melting layer. ERAD 2012.
- Melnikov, V., D. S. Zrnić, M. E. Weber, A. O. Fierro, and D. R. MacGorman, 2019: Electrified cloud areas observed in the shv and ldr radar modes. *Journal of Atmospheric and Oceanic Technology*, **36** (1), 151–159, <https://doi.org/10.1175/jtech-d-18-0022.1>.
- Meyes, R., M. Lu, C. Waubert de Puiseau, and T. Meisen, 2019: Ablation studies in artificial neural networks. arXiv, <https://doi.org/10.48550/arXiv.1901.08644>.
- Mishler, R., G. Zhang, and V. N. Mahale, 2024: Improving estimation of specific differential phase and assessing differential backscattering phase. *Journal of Applied Meteorology and Climatology*, **63** (7), 803–819, <https://doi.org/10.1175/jamc-d-23-0204.1>.
- Molnar, C., 2022: *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Creative Commons.
- Monahan, A. H., 2000: Nonlinear principal component analysis by neural networks: Theory and application to the lorenz system. *J. Climate*, **13**, 821–835.
- Moninger, W. R., J. Davis, R. Dyer, R. Kittredge, R. McArthur, A. H. Murphy, and I. R. Racer, 1987: Summary of the first conference on artificial intelligence research in environmental sciences (airies). *Bulletin of the American Meteorological Society*, **68** (7), 793–800, URL <http://www.jstor.org/stable/26225427>.
- Mostajabi, A., D. L. Finney, M. Rubinstein, and F. Rachidi, 2019: Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. *npj Climate and Atmospheric Science*, **2** (1), <https://doi.org/10.1038/s41612-019-0098-0>.
- NCEP, 2004: The gfs atmospheric model. *NOAA Institutional Repository*, URL <https://repository.library.noaa.gov/view/noaa/11406>, (U.S.).
- NCEP, 2015: Ncep gfs 0.25 degree global forecast grids historical archive. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, URL <https://doi.org/10.5065/D65D8PWK>, <https://doi.org/10.5065/D65D8PWK>.
- Ortland, S. M., M. J. Pavolonis, and J. L. Cintineo, 2023: The development and initial capabilities of thundercast, a deep-learning model for thunderstorm nowcasting in the United States. *Artificial Intelligence for the Earth Systems*, <https://doi.org/10.1029/2023AI200001>.

1175/aies-d-23-0044.1.

- Orville, R. E., C. Schumacher, R. M. Mosier, and L. D. Carey, 2011: Radar nowcasting of cloud-to-ground lightning over houston, texas. *Weather and Forecasting*, **26** (2), 199–212, <https://doi.org/10.1175/2010waf2222431.1>.
- Osborne, A. P., J. Zhang, R. A. Clark, and K. W. Howard, 2026: Optimization of a machine learning multi-radar multi-sensor (MRMS) precipitation estimation approach over the western United States. *Artificial Intelligence for the Earth Systems*, **5** (2), 240 013, <https://doi.org/10.1175/AIES-D-24-0013.1>, URL <https://journals.ametsoc.org/view/journals/aies/5/2/AIES-D-24-0013.1.xml>.
- Palmer, R. D., and D. Schwartzman, 2025: Emerging trends in radar: Phased arrays for weather observations. *IEEE Aerospace and Electronic Systems Magazine*, **40** (6), 60–64, <https://doi.org/10.1109/maes.2024.3524937>.
- Palmer, R. D., and Coauthors, 2023: Horus—a fully digital polarimetric phased array radar for next-generation weather observations. *IEEE Transactions on Radar Systems*, **1**, 96–117, <https://doi.org/10.1109/trs.2023.3280033>.
- Pehl, J., D. Schwartzman, D. Zrnić, A. Ryzhkov, M. Herndon, and R. D. Palmer, 2024: Evaluation of polarimetric calibration of the horus radar using weather returns. *IEEE*, 1–7 pp., <https://doi.org/10.1109/array58370.2024.10880343>.
- Pérez-Invernón, F. J., and Coauthors, 2021: Lightning-ignited wildfires and long continuing current lightning in the mediterranean basin: preferential meteorological conditions. *Atmospheric Chemistry and Physics*, **21** (23), 17 529–17 557, <https://doi.org/10.5194/acp-21-17529-2021>, URL <https://acp.copernicus.org/articles/21/17529/2021/>.
- Prince, S., 2023: *Understanding Deep Learning*. The MIT Press.
- Quick, M. G., and Coauthors, 2020: Mid-Atlantic Lightning Mapping Array (MALMA) multi-band network performance. *AGU Fall Meeting Abstracts*, AGU Fall Meeting Abstracts, Vol. 2020, AE008–0003.
- Reimel, K. J., and M. Kumjian, 2021a: Evaluation of kdp estimation algorithm performance in rain using a known-truth framework. *Journal of Atmospheric and Oceanic Technology*, **38** (3), 587–605, <https://doi.org/10.1175/jtech-d-20-0060.1>.
- Reimel, K. J., and M. R. Kumjian, 2021b: Supplemental material: Kdp algorithm descriptions. *Journal of Atmospheric and Oceanic Technology*.

- Ringhausen, J., P. Bitzer, W. Koshak, and J. Mecikalski, 2021: Classification of GLM flashes using random forests. *EARTH AND SPACE SCIENCE*, **8** (11), <https://doi.org/10.1029/2021EA001861>, URL <https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2021EA001861?download=true>.
- Rison, W., R. J. Thomas, P. R. Krehbiel, T. Hamlin, and J. Harlin, 1999: A gps-based three-dimensional lightning mapping system: Initial observations in central new mexico. *Geophysical Research Letters*, **26** (23), 3573–3576, <https://doi.org/10.1029/1999GL010856>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/1999GL010856>, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/1999GL010856>.
- Roberts, B., B. T. Gallo, I. L. Jirak, and A. J. Clark, 2019: The high resolution ensemble forecast (HREF) system: Applications and performance for forecasting convective storms. *American Geophysical Union*, AGU, <https://doi.org/10.1002/essoar.10501462.1>.
- Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Weather and Forecasting*, **24** (2), 601–608, <https://doi.org/10.1175/2008waf2222159.1>.
- Roeder, J., William; Saul, 2017: The new mesoscale eastern range lightning information system. *American Meteorological Society Annual Conference 2017*.
- Roeder, W., and J. M. Saul, 2012: Four dimensional lightning surveillance system: Status and plans. United States Air Force.
- Romps, D. M., 2014: Projected increase in lightning strikes in the United States due to global warming. *Science*.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-net: Convolutional networks for biomedical image segmentation. Springer.
- Rudlosky, S. D., S. J. Goodman, K. S. Virts, and E. C. Bruning, 2019: Initial geostationary lightning mapper observations. *Geophysical Research Letters*, **46** (2), 1097–1104, <https://doi.org/10.1029/2018gl081052>.
- Rumelhart, D. E., G. E. Hinton, and R. J. Williams, 1986: Learning representations by back-propagating errors. *Nature*, **323** (6088), 533–536, <https://doi.org/10.1038/323533a0>, URL <https://doi.org/10.1038/323533a0>.
- Rutledge, S. A., K. A. Hilburn, A. Clayton, B. Fuchs, and S. D. Miller, 2020: Evaluating geostationary lightning mapper flash rates within intense convective storms. *Journal of Geophysical Research: Atmospheres*, **125** (14), e2020JD032827, <https://doi.org/>

- <https://doi.org/10.1029/2020JD032827>, URL <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020JD032827>, e2020JD032827 2020JD032827, <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020JD032827>.
- Ryzhkov, A., and D. Zrnic, 2007: Depolarization in ice crystals and its effect on radar polarimetric measurements. *Journal of Atmospheric and Oceanic Technology*, <https://doi.org/10.1175/JTECH2034.1>.
- Saunders, C. P. R., H. Bax-norman, C. Emersic, E. E. Avila, and N. E. Castellano, 2006: Laboratory studies of the effect of cloud conditions on graupel/crystal charge transfer in thunderstorm electrification. *Quarterly Journal of the Royal Meteorological Society*, **132** (621), 2653–2673, <https://doi.org/10.1256/qj.05.218>.
- Schneebeli, M., J. Grazioli, and A. Berne, 2014: Improved estimation of the specific differential phase shift using a compilation of kalman filter ensembles. *IEEE Transactions on Geoscience and Remote Sensing*, **52** (8), 5137–5149, <https://doi.org/10.1109/tgrs.2013.2287017>.
- Schreck, J. S., and Coauthors, 2025: Community research earth digital intelligence twin: a scalable framework for ai-driven earth system modeling. *npj Climate and Atmospheric Science*, **8** (1), 239, <https://doi.org/10.1038/s41612-025-01125-6>, URL <https://doi.org/10.1038/s41612-025-01125-6>.
- Schultz, C. J., P. M. Bitzer, M. Antia, J. L. Case, and C. R. Hain, 2024: Assessing flash characteristics in lightning-initiated wildfire events between 1995 and 2020 within the contiguous United States. *Journal of Applied Meteorology and Climatology*, **63** (4), 543–551, <https://doi.org/10.1175/jamc-d-23-0166.1>.
- Schultz, C. J., N. J. Nauslar, J. B. Wachter, C. R. Hain, and J. R. Bell, 2019: Spatial, temporal and electrical characteristics of lightning in reported lightning-initiated wildfire events. *Fire*, **2** (2), <https://doi.org/10.3390/fire2020018>, URL <https://www.mdpi.com/2571-6255/2/2/18>.
- Schultz, C. J., W. A. Petersen, and L. D. Carey, 2009: Preliminary development and evaluation of lightning jump algorithms for the real-time detection of severe weather. *Journal of Applied Meteorology and Climatology*, **48** (12), 2543 – 2563, <https://doi.org/10.1175/2009JAMC2237.1>, URL <https://journals.ametsoc.org/view/journals/apme/48/12/2009jamc2237.1.xml>.
- Schumacher, C., R. E. Orville, and G. N. Seroka, 2012: Radar nowcasting of total lightning over the kennedy space center. *Weather and Forecasting*, **27** (1), 189–204, <https://doi.org/10.1175/waf-d-11-00035.1>.

- Schvartzman, D., and Coauthors, 2026: Advanced weather surveillance capabilities of the fully digital horus phased array radar. *Bulletin of the American Meteorological Society*, **107** (1), E59–E78, <https://doi.org/10.1175/bams-d-24-0296.1>.
- Shao, X. M., and P. R. Krehbiel, 1996: The spatial and temporal development of intracloud lightning. *Journal of Geophysical Research: Atmospheres*, **101** (D21), 26 641–26 668, <https://doi.org/10.1029/96jd01803>.
- Sharma, M., R. L. Tanamachi, and E. C. Bruning, 2024: Investigating temporal characteristics of polarimetric and electrical signatures in three severe storms: Insights from the vortex-southeast field campaign. *Monthly Weather Review*, **152** (7), 1511–1536, <https://doi.org/10.1175/mwr-d-23-0144.1>.
- Shi, X., Z. Chen, H. Wang, and D.-Y. Yeung, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Arxiv*.
- Skrovan, C. A., 2020: 45 ws electric field mill lightning prediction threshold analysis. Thesis, Operations Research.
- Smith, K., 2025: Personal communication with nasa. NASA.
- Sobash, R. A., and D. A. Ahijevych, 2025: Evaluating machine learning–based probabilistic lightning forecasts using the hrrr: A comparison to three forecast baselines. *Weather and Forecasting*, **40** (5), 759 – 774, <https://doi.org/10.1175/WAF-D-24-0140.1>, URL <https://journals.ametsoc.org/view/journals/wefo/40/5/WAF-D-24-0140.1.xml>.
- Speranza, D., 2019: Lightning prediction using recurrent neural networks. Thesis, Engineering Physics.
- Steiger, S. M., and Coauthors, 2024: Winter lightning to the lee of lake ontario: The lake-effect electrification (lee) field campaign. *Bulletin of the American Meteorological Society*, **105** (11), E2026 – E2046, <https://doi.org/10.1175/BAMS-D-23-0176.1>, URL <https://journals.ametsoc.org/view/journals/bams/105/11/BAMS-D-23-0176.1.xml>.
- Stewart, J., 2026: Personal communication with 45th weather squadron. USSF.
- Stolzenburg, M., T. C. Marshall, and P. R. Krehbiel, 2015: Initial electrification to the first lightning flash in new mexico thunderstorms. *Journal of Geophysical Research: Atmospheres*, **120** (21), <https://doi.org/10.1002/2015jd023988>.
- Stough, S. M., L. D. Carey, C. J. Schultz, and P. M. Bitzer, 2017: Investi-

- gating the relationship between lightning and mesocyclonic rotation in supercell thunderstorms. *Weather and Forecasting*, **32** (6), 2237 – 2259, <https://doi.org/10.1175/WAF-D-17-0025.1>, URL https://journals.ametsoc.org/view/journals/wefo/32/6/waf-d-17-0025_1.xml.
- Takahashi, T., 1978: Riming electrification as a charge generation mechanism in thunderstorms. *Journal of the Atmospheric Sciences*, **35**.
- Takahashi, T., 1984: Thunderstorm electrification - a numerical study. *Journal of the Atmospheric Sciences*, **41**, 2541–2558.
- Thomas, A. M., and S. Noble, 2024: A physics-based ensemble machine-learning approach to identifying a relationship between lightning indices and binary lightning hazard. *Frontiers in Earth Science*, **Volume 12 - 2024**, <https://doi.org/10.3389/feart.2024.1376605>, URL <https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2024.1376605>.
- Thurmond, K. R., 2014: Operational cloud-to-ground lightning initiation forecasting utilizing s-band dual polarization radar. Thesis, Engineering Physics.
- Toms, B. A., E. A. Barnes, and I. Ebert-Uphoff, 2020: Physically interpretable neural networks for the geosciences: Applications to earth system variability. *Journal of Advances in Modeling Earth Systems*, **12** (9), <https://doi.org/10.1029/2019ms002002>.
- Trabing, B. C., K. Hilburn, S. Stevenson, K. D. Musgrave, and M. DeMaria, 2024: Quality control of geostationary lightning mapper observations for tropical cyclone applications. *Journal of Atmospheric and Oceanic Technology*, **41** (9), 889–901, <https://doi.org/10.1175/jtech-d-23-0138.1>.
- Travis, A. J., 2015: Utilizing four dimensional lightning and dual-polarization radar to develop lightning initiation forecast guidance. Thesis, Engineering Physics.
- Trok, J. T., F. V. Davenport, E. A. Barnes, and N. S. Diffenbaugh, 2023: Using machine learning with partial dependence analysis to investigate coupling between soil moisture and near-surface temperature. *Journal of Geophysical Research: Atmospheres*, **128** (12), <https://doi.org/10.1029/2022jd038365>.
- Tzeng, M.-D., T.-Y. Yu, D. Schwartzman, D. Bodine, J. C. Hubbert, V. Goede, and V. Chmielewski, 2025: The impact of antenna cross-pol contamination on polarimetric spectra from ice alignment signatures. *IEEE Transactions on Geoscience and Remote Sensing*, **63**, 1–19, <https://doi.org/10.1109/tgrs.2025.3597189>.
- Veillette, M., S. Samsi, and C. Mattioli, 2020: Sevir : A storm event im-

- agery dataset for deep learning applications in radar and satellite meteorology. *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., Curran Associates, Inc., Vol. 33, 22 009–22 019, URL https://proceedings.neurips.cc/paper_files/paper/2020/file/fa78a16157fed00d7a80515818432169-Paper.pdf.
- Vicens-Miquel, M., A. McGovern, A. J. Hill, E. Foufoula-Georgiou, C. Guilloteau, and S. S. P. Shen, 2025: A diffusion-based framework for high-resolution precipitation forecasting over conus. ArXiv, URL <https://arxiv.org/abs/2512.09059>, 2512.09059.
- Vulpiani, G., M. Montopoli, L. D. Passeri, A. G. Gioia, P. Giordano, and F. S. Marzano, 2012: On the use of dual-polarized c-band radar for operational rainfall retrieval in mountainous areas. *Journal of Applied Meteorology and Climatology*, **51** (2), 405–425, <https://doi.org/10.1175/jamc-d-10-05024.1>.
- Wang, S., Y. Wada, S. Hayashi, T. Ushio, and V. Chandrasekar, 2024: Observation of electrical alignment signatures in an isolated thunderstorm by dual-polarized phased array weather radar and the relationship with intracloud lightning flash rate. *Journal of Geophysical Research: Atmospheres*, **129** (19), <https://doi.org/10.1029/2023jd040681>.
- Wang, X., K. Hu, Y. Wu, and W. Zhou, 2023: A survey of deep learning-based lightning prediction. *Atmosphere*, **14** (11), <https://doi.org/10.3390/atmos14111698>.
- Wanless, A., E. Meister, S. Stormer, D. Hogg, M. Krocak, J. Ripberger, and H. Jenkins-Smith, 2025: The Extreme Weather and Emergency Management Survey, Wave 8: Emergency Manager Workflows. Harvard Dataverse, URL <https://doi.org/10.7910/DVN/ZR3WUE>, <https://doi.org/10.7910/DVN/ZR3WUE>.
- Wilks, D. S., 2012: *Statistical Methods in the Atmospheric Sciences*, Vol. 100. Elsevier.
- Wilks, D. S., 2016: “the stippling shows statistically significant grid points”: How research results are routinely overstated and overinterpreted, and what to do about it. *Bulletin of the American Meteorological Society*, **97** (12), 2263–2273, <https://doi.org/10.1175/bams-d-15-00267.1>.
- Wilson, C., 1920: Iii. investigations on lightning discharges and on the electric field of thunderstorms. Royal Society Publishing.
- Wilson, J. W., and W. E. Schreiber, 1986: Initiation of convective storms at radar-observed boundary layer convergence lines. *Monthly Weather Review*, **114**, 2516–2536.

- WMO, 2019: Lightning for climate: A study by the task team on lightning observation for climate observations atmospheric observation panel for climate (aopc). Report, World Meteorological Organization.
- Woodard, C. J., L. D. Carey, W. A. Petersen, and W. P. Roeder, 2012: Operational utility of dual polarization variables in lightning initiation forecasting. *Electronic Journal of Operational Meteorology*, **13**, 79–102.
- Xu, W., E. J. Zipser, C. Liu, and H. Jiang, 2010: On the relationships between lightning frequency and thundercloud parameters of regional precipitation systems. *Journal of Geophysical Research: Atmospheres*, **115** (D12), <https://doi.org/10.1029/2009jd013385>.
- Zeng, G., J. A. Pyle, and Y. P. J., 2008: Impacts of climate change on tropospheric ozone and its global budgets. *Atmospheric Chemistry and Physics*, **8**, 369–387.
- Zhang, A., Z. Lipton, M. Li, and A. Smola, 2023: *Dive Into Deep Learning*. Cambridge Univeristy Press.
- Zhang, D. L., and K. L. Cummins, 2020: Time evolution of satellite-based optical properties in lightning flashes, and its impact on GLM flash detection. *Journal of Geophysical Research-Atmospheres*, **125** (6), <https://doi.org/10.1029/2019jd032024>, URL <https://doi.org/10.1029/2019JD032024?download=true>, zhang, Daile Cummins, Kenneth L. Zhang, Daile/0000-0003-2571-6104; Cummins, Kenneth/0000-0001-9871-691X 2169-8996.
- Zhang, J., and Coauthors, 2016: Multi-radar multi-sensor (mrms) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, **97** (4), 621 – 638, <https://doi.org/10.1175/BAMS-D-14-00174.1>, URL <https://journals.ametsoc.org/view/journals/bams/97/4/bams-d-14-00174.1.xml>.
- Zhao, C., and Coauthors, 2024: Technical note: On the ice microphysics of isolated thunderstorm and non-thunderstorms in southern china: A radar polarimetric perspective. *Egusphere*, <https://doi.org/10.5194/egusphere-2024-907>.
- Zhou, K. H., Y. G. Zheng, W. S. Dong, and T. B. Wang, 2020: A deep learning network for cloud-to-ground lightning nowcasting with multisource data. *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY*, **37** (5), 927–942, <https://doi.org/10.1175/JTECH-D-19-0146.1>, URL <https://journals.ametsoc.org/downloadpdf/journals/atot/37/5/jtech-d-19-0146.1.pdf>.
- Zhou, L., L. Harris, and J.-H. Chen, 2022: The gfdl cloud microphysics parameteriza-

tion. NOAA.

Zhou, L., S.-J. Lin, J.-H. Chen, L. M. Harris, X. Chen, and S. L. Rees, 2019: Toward convective-scale prediction within the next generation global prediction system. *Bulletin of the American Meteorological Society*, **100** (7), 1225 – 1243, <https://doi.org/10.1175/BAMS-D-17-0246.1>.