# BoltCast: Deep Learning for Medium-Range Lightning Prediction

**Brandon McClung, Major, PhD Candidate**
School of Meteorology, University of Oklahoma
Air Force Institute of Technology, Civilian Institute (AFIT/CI)

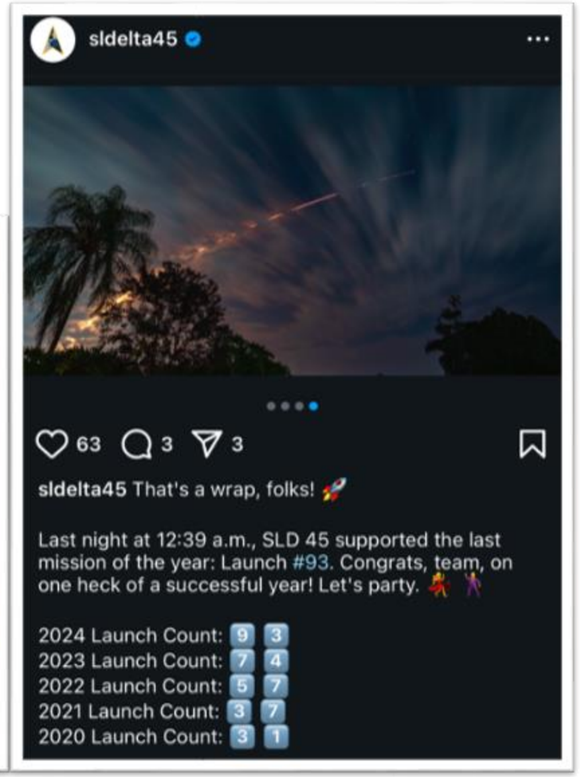
**Co-Authors:**
Dr. David Schvartzman, SoM/ARRC
Dr. Aaron Hill, SoM/AI2ES
Dr. Michael Stock, CIWRO
**Dr. Amy McGovern, SoM/AI2ES (Advisor)**

DISCLAIMER: “The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.”

“This material is based upon work supported by the National Science Foundation under Grant No. RISE-2019758.”

# Research Motivation and Goal








- Space launches continue to increase and are expected to surpass 100 launches per year along the Space Coast in the next 2-3 years, with restrictive launch criteria due to lightning
- Cloud-to-ground lightning may trigger high-impact wildfires in the western United States and across the globe
- The energy sector can be impacted by lightning strikes
- AI/NWP lightning prediction is focused on the nowcasting and storm scale, with little focus on the medium range
  - Impact mitigation/resource allocation can take 2-3 days
- **Goal:** Develop a stable, physics-based ML model to predict lightning out through 4 days to enable proper mitigation techniques and resource allocation (i.e. planning weather)

https://www.science.org/content/article/air-pollution-helps-wildfires-create-their-own-lightning
https://www.schunk-group.com/en/blog/category/energy/reliable-lightning-protection-wind-turbines~b31581

# Physics-Based Flash Rate Parameterization Schemes

| Required criteria thresholds | | | |
|---|---|---|---|
| Reflectivity (dBZ) | Temperature (°C) | Vertical velocity ($m\ s^{-1}$) | Hydrometeor ($g\ kg^{-1}$) |

Cummings summarized several flash rate parameterization schemes (FRPS) for single storms using high temporal WRF simulations, pulling from >20 different literature sources

Radar reflectivity, vertical velocity, temperature, and hydrometeor mixing ratios were common inputs into the FRPS

Provided strong justification to use similar inputs for a medium- to long-range lightning prediction tool

But what NWP and lightning data should be used, that are ML ready?

(Cummings, et al. 2024)

# The GFS and GLM Datasets

NCAR'S RDA Server, for the .25-degree GFS Archive as inputs out to 4-days at 3-hour resolution

GOES 16/17/18 GLM on AWS as predictand, merged along 100 degrees West meridian
# of radiance weighted flashes are counted in each GFS grid cell (Goodman et al. 2012)
Binary mask applied for greater than 1 flash

Model and GLM valid times: 07/01/2019 – 06/30/2024 across 00Z, 06Z, 12Z, and 18Z initializations

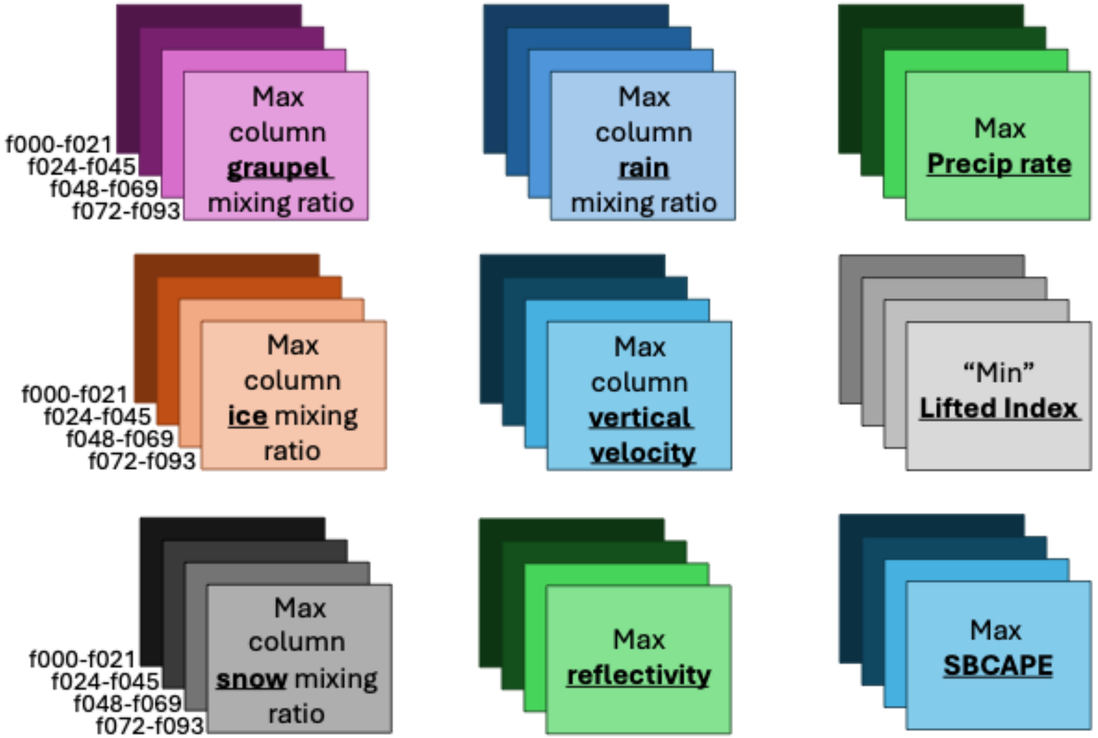
The training, validation, and test datasets are independent sliced by year, includes all seasons

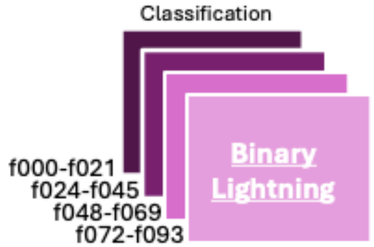
| | Training | Validation | Test |
|---|---|---|---|
| **Valid Dates (# samples)** | **July 2019 – May 2021; June 2023 – June 2024 (4365)** | **June 2021 – May 2022 (1456)** | **June 2022 – May 2023 (1452)** |

One sample is comprised of maximums of nine input variables, spanning the 4-day prediction window

For latitude and longitudes that span CONUS (128 pixels by 256 pixels)

# Model Inputs/Feature Engineering




The maximum is calculated over the 24-hour forecast window, and the column, if applicable

**Generates a five-dimensional dataset for input into the ML model data pipeline**
[number of samples, day (4), latitude (128), longitude (256), input variables (9)]

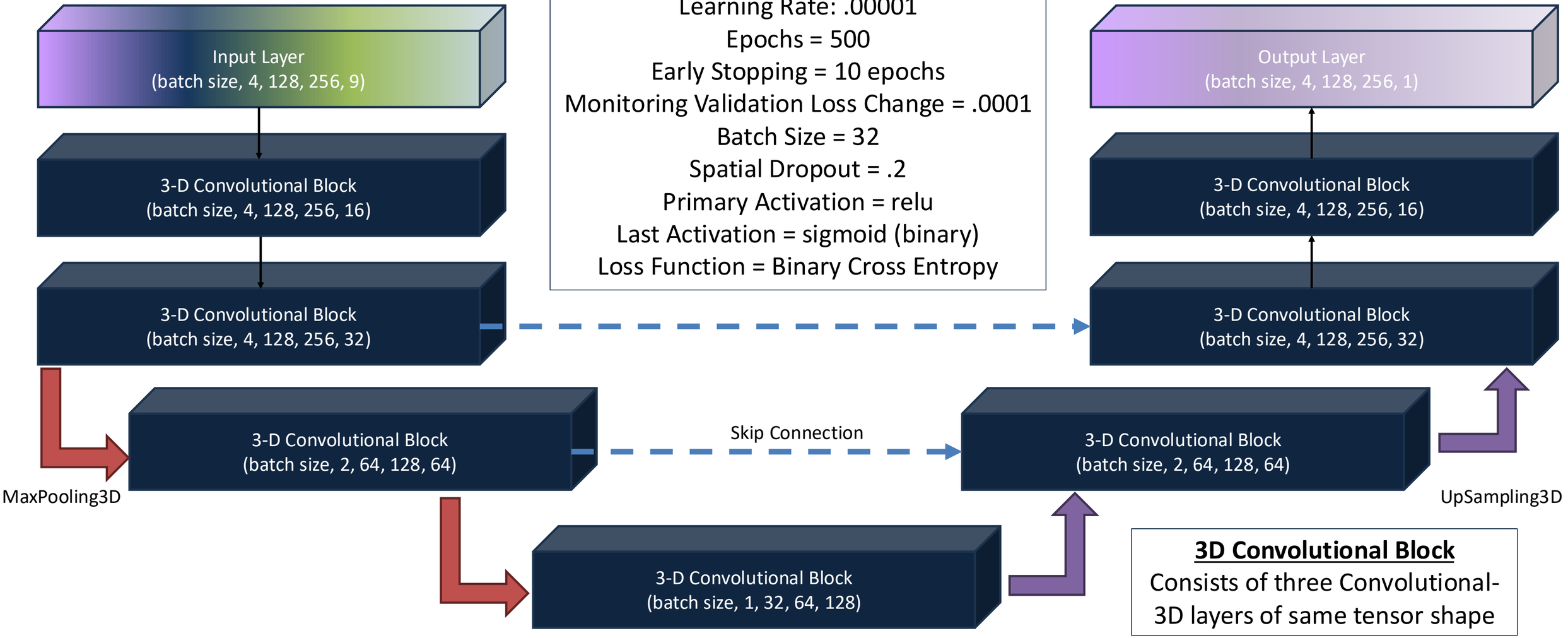
**Generates four-dimensional dataset as output of the ML model data pipeline**
[number of samples, day (4), latitude (128), longitude (256)]

**U-Net and ConvLSTM-based models and datasets were generated with Keras 3.0/Tensorflow v2.15.0**
Trained on 1x NVIDIA A100 GPU

# UNet Model

**Training/Hyper-parameters:**

Pooling Size = (2,2,2)
Convolutional Size = 4
Learning Rate: .00001
Epochs = 500
Early Stopping = 10 epochs
Monitoring Validation Loss Change = .0001
Batch Size = 32
Spatial Dropout = .2
Primary Activation = relu
Last Activation = sigmoid (binary)
Loss Function = Binary Cross Entropy



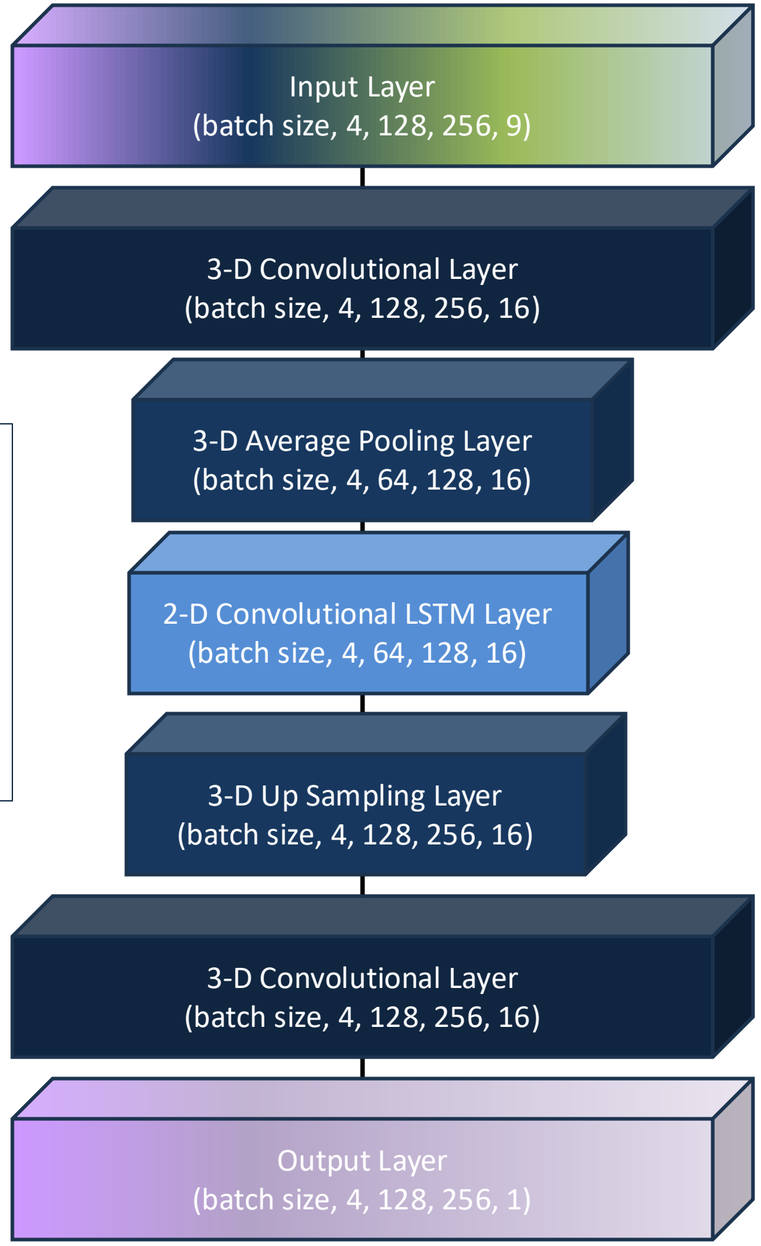

**3D Convolutional Block**

Consists of three Convolutional-3D layers of same tensor shape

# ConvLSTM Architecture

(Shi, et al. 2015) developed the ConvLSTM layers for geospatial/temporal data, focusing on precipitation forecasting

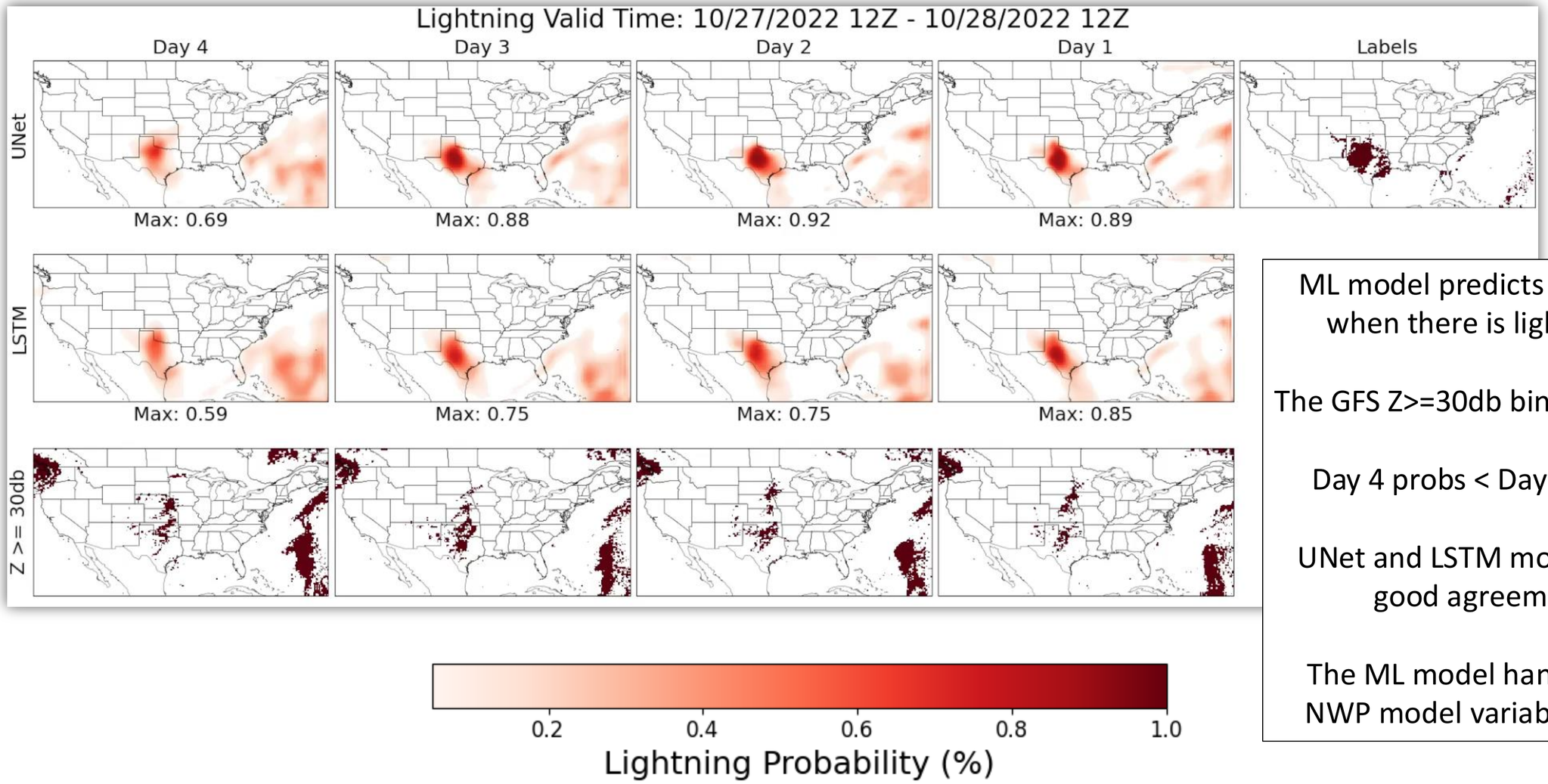
Literature suggests a mis-use of these layer types, so I wanted to understand if a simple architecture generates meaningful results




**Training/Hyper-parameters:**

Pooling Size = (2,2,2)
Convolutional Size = 4
Learning Rate: .00001
Return Sequences = True
Return State = True
Epochs = 500
Early Stopping = 10 epochs
Monitoring Validation Loss Change = .0001
Batch Size = 32
Primary Activation = relu
Last Activation = sigmoid (binary)
Loss Function = Binary Cross Entropy

# Model Output – Random Samples




ML model predicts lightning when there is lightning!

The GFS Z>=30db binary is poor

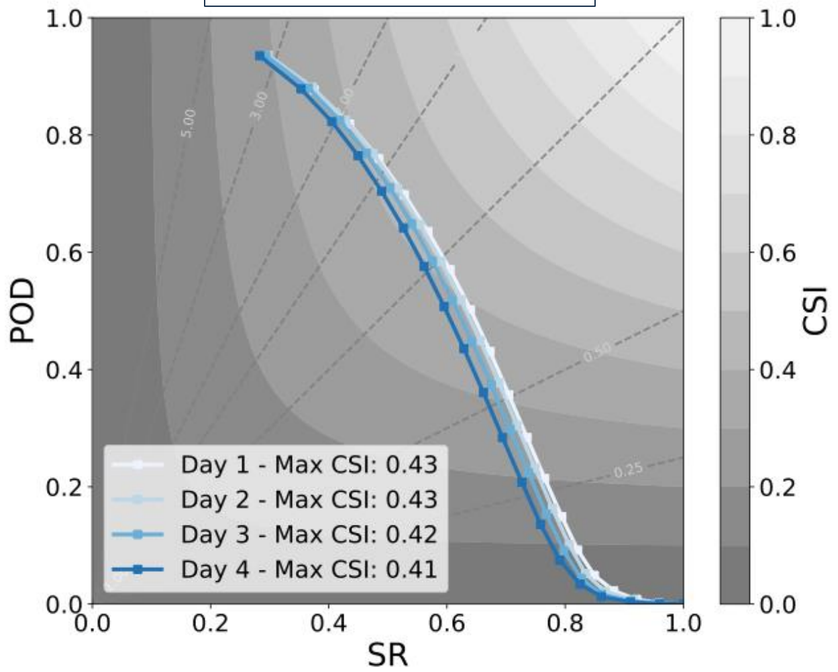
Day 4 probs < Day 1 probs

UNet and LSTM model are in good agreement

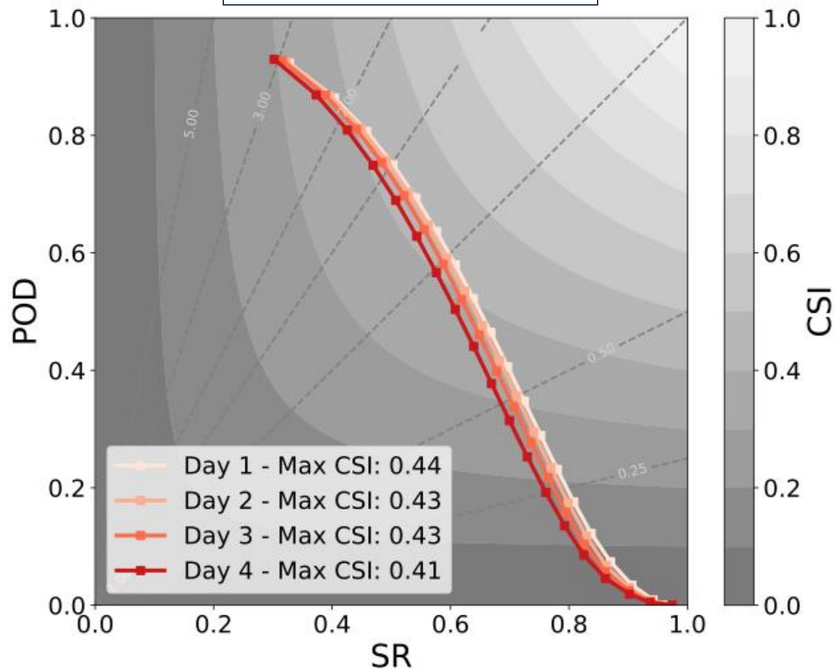
The ML model handles the NWP model variability well

# Precision Recall Curves

LSTM: AUC - .573




UNet: AUC - .593




$$POD = \frac{TP}{TP+FN}$$

$$SR = \frac{TP}{TP+FP}$$

$$CSI = \frac{TP}{TP+FN+FP}$$

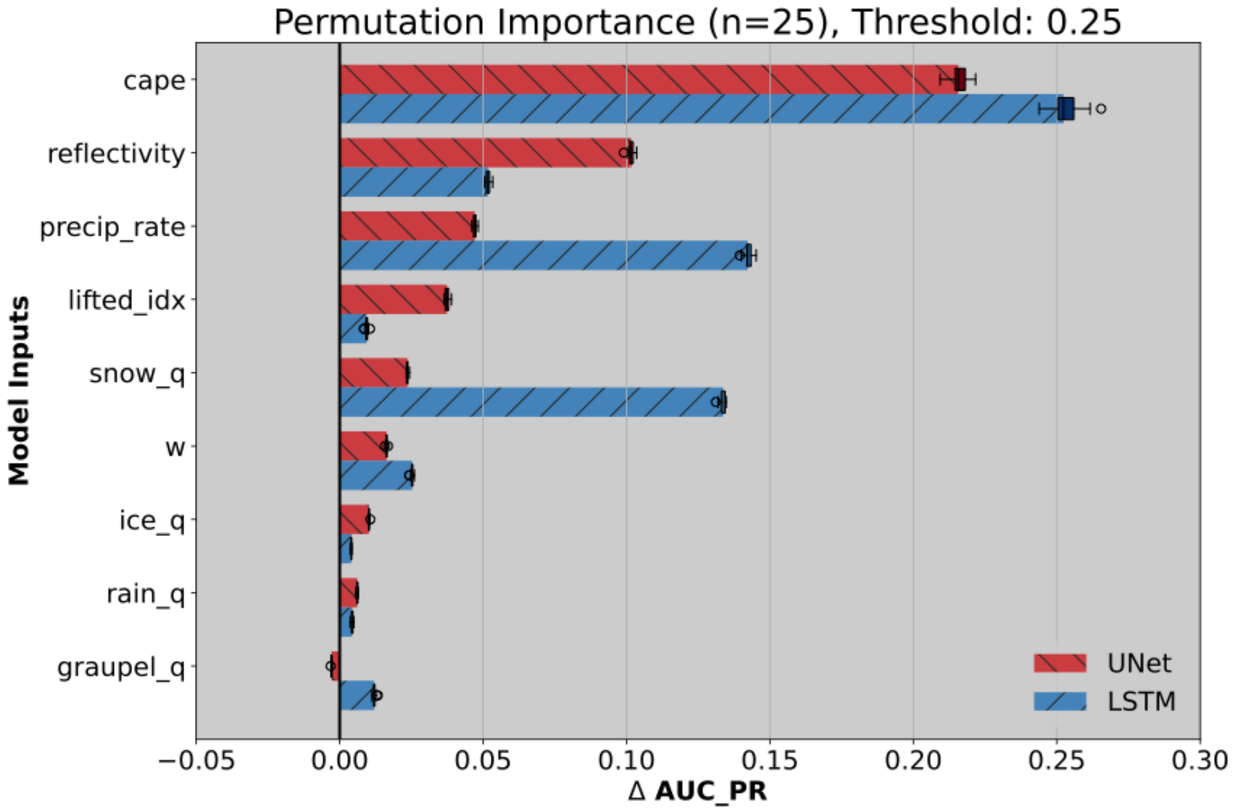
- **TP: True Positive**: Forecasted lightning & occurred
- **FP: False Positive**: Forecasted lightning & didn't occur
- **FN: False Negative:** Lightning occurred but wasn't forecast

Evaluate based on forecasted probability 5% bins: .05, .1, etc. w/in TensorFlow Metrics API

Consistent decrease in CSI scores as the forecast lead time increases
Day 1 slightly outperforms Day 2, but not by much

LightNet+ boasts a .21 threat score out to 6 hours at 4-km resolution (Geng et al. 2021)
HREF Probabilistic Thunder Guidance generates .44 avg CSI thru Day 1 (Harrison et al. 2022)

# xAI: Single Input Permutation Importance




Permute each variable in **<u>test dataset</u>**, along sample space dimension, then make prediction to calculate model performance delta

CAPE is most important for UNet and LSTM model
CAPE, precipitation rate, snow mixing ratio, vertical velocity, and graupel had higher importance in the LSTM model
Graupel mixing ratio is relatively low for both models

Little variability (small box/whisker) among each feature

# Discussion and Conclusion

- BoltCast fills the medium-range AI-based lightning prediction gap, with moderate skill, using a U-Net and ConvLSTM architecture 9 GFS input variables as predictors and GOES 16/17/18 GLM flashes as predictand

- BoltCast prediction metrics decrease slightly as the forecast lead time increases from .44 - .41, but beat or maintain predictive skill beyond lead-times of LightNet+ and HREF probabilistic forecasts

- Increasingly complex ConvLSTM model architectures are not necessarily needed, if the data is staged smart

- Permuting CAPE had the biggest impact on UNet and LSTM model performance, while graupel mixing ratio never made it in the top 5 for either model

- **Future work:**
  - Seasonality metrics
  - Creating group permutations for highly correlated variables
  - Submitting to *Artificial Intelligence for Environment Science* AMS Journal
  - What about flash rates?

# References

(2015). NCEP GFS 0.25 Degree Global Forecast Grids Historical Archive. Boulder, CO, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory.

Cummings, K. A., et al. (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).

Geng, Y. a., et al. (2021). "A deep learning framework for lightning forecasting with multi-source spatiotemporal data." Quarterly Journal of the Royal Meteorological Society **147**(741): 4048-4062.

Goodman, S. (2012). Algorithm Theoretical Basis Document: GLM Lightning Cluster Filter Algorithm. N. N. C. f. S. A. a. Research.

Harrison, D. R., et al. (2022). "Utilizing the High-Resolution Ensemble Forecast System to Produce Calibrated Probabilistic Thunderstorm Guidance." Weather and forecasting **37**(7): 1103-1115.

Shi, X., et al. (2015). "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting." Arxiv.

Tippett, M. K. and W. J. Koshak (2018). "A Baseline for the Predictability of U.S. Cloud-to-Ground Lightning." Geophysical research letters **45**(19).

# Questions?

- BoltCast fills the medium-range AI-based lightning prediction gap, with moderate skill, using a U-Net and ConvLSTM architecture 9 GFS input variables as predictors and GOES 16/17/18 GLM flashes as predictand
- BoltCast prediction metrics decrease slightly as the forecast lead time increases from .44 - .41, but beat or maintain predictive skill beyond lead-times of LightNet+ and HREF probabilistic forecasts
- Increasingly complex ConvLSTM model architectures are not necessarily needed, if the data is staged smart
- Permuting CAPE had the biggest impact on UNet and LSTM model performance, while graupel mixing ratio never made it in the top 5 for either model

https://www.linkedin.com/in/brandon-mcclung/

bmac7167@ou.edu

https://github.com/bmccl87/BoltCast

# BACKUP SLIDES

# Metrics

- POD: Probability of Detection
  - 1 is best
  - 0 is worst
- SR: Success Ratio
  - 1 is best
  - 0 is worst
- CSI: Critical Success Index
  - 1 is best
  - 0 is worst

$$POD = \frac{TP}{TP+FN}$$

$$SR = \frac{TP}{TP+FP}$$

$$CSI = \frac{TP}{TP+FN+FP}$$

- TP: True Positive
  - Forecasted lightning and lightning occurred
- FP: False Positive
  - Lightning was forecast but didn't occur
- FN: False Negative
  - Lightning occurred but wasn't forecast

We're evaluating based on forecasted probability 5% bins: .05, .1, .15, etc. within TensorFlow Keras API

# The GFS and GLM Datasets

NCAR'S RDA Server, for the .25 degree Global Forecast System Archive as inputs out to 4-days at 3-hour resolution

GOES 16/17/18 Global Lightning Mapper as outputs, merged along 100 degrees West meridian

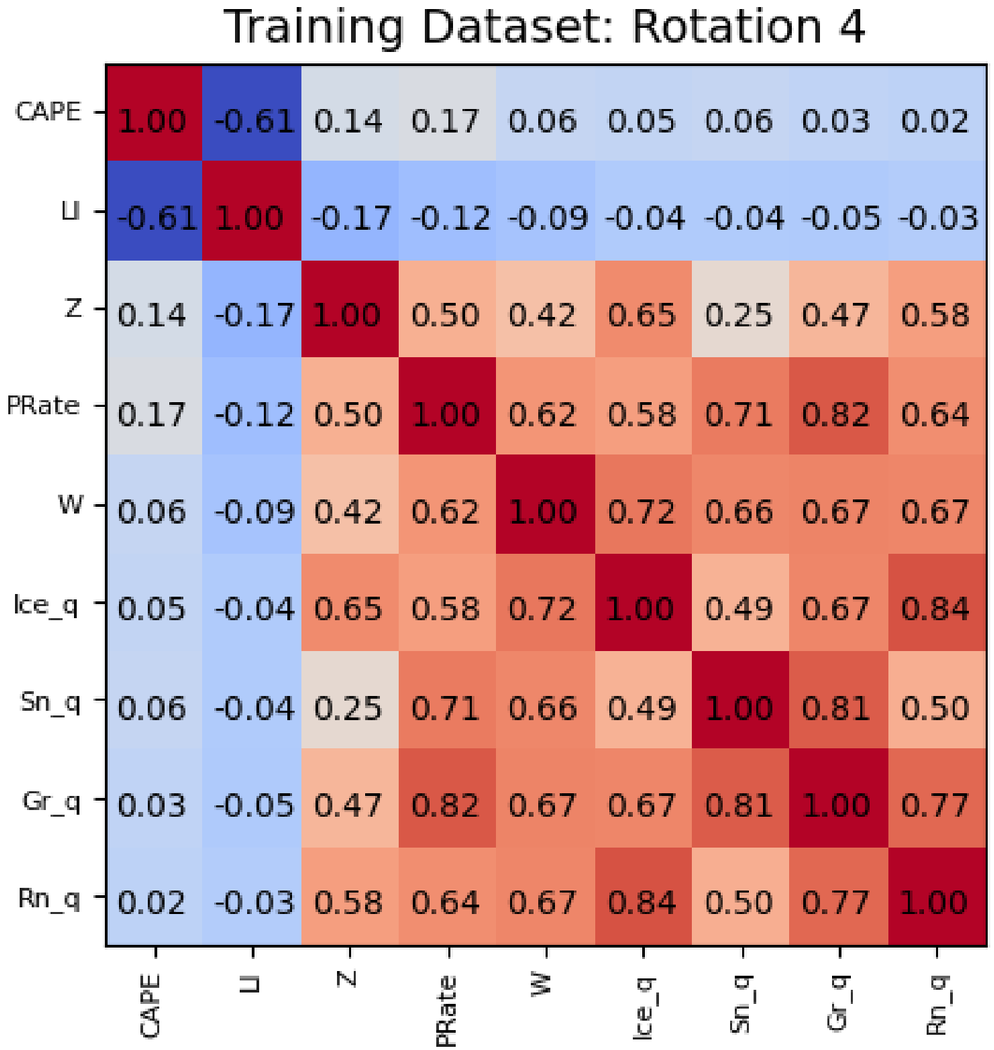
Model valid times: 07/01/2019 – 06/30/2024 across 00Z, 06Z, 12Z, and 18Z initializations

K-folds cross validation generates a different training dataset to increase model robustness/reliability

The training, validation, and test datasets are independent sliced by year, includes all seasons

| Rotations | | Training | Validation | Test |
|---|---|---|---|---|
| 0 | Valid Dates (# samples) | July 2019 – May 2022; June 2024 (4361) | June 2022 – May 2023 (1452) | June 2023 – May 2024 (1460) |
| 1 | Valid Dates (# samples) | June 2020 – May 2022; June 2024 (4470) | June 2023 – May 2024 (1460) | July 2019 – May 2020 (1343) |
| 2 | Valid Dates (# samples) | June 2021 – June 2024 (4471) | July 2019 – May 2020 (1343) | June 2020 – May 2021 (1459) |
| 3 | Valid Dates (# samples) | July 2019 – May 2020; June 2022 – June 2024 (4358) | June 2020 – May 2021 (1459) | June 2021 – May 2022 (1456) |
| 4 | Valid Dates (# samples) | July 2019 – May 2021; June 2023 – June 2024 (4365) | June 2021 – May 2022 (1456) | June 2022 – May 2023 (1452) |

# Correlation Matrices




Data are correlated

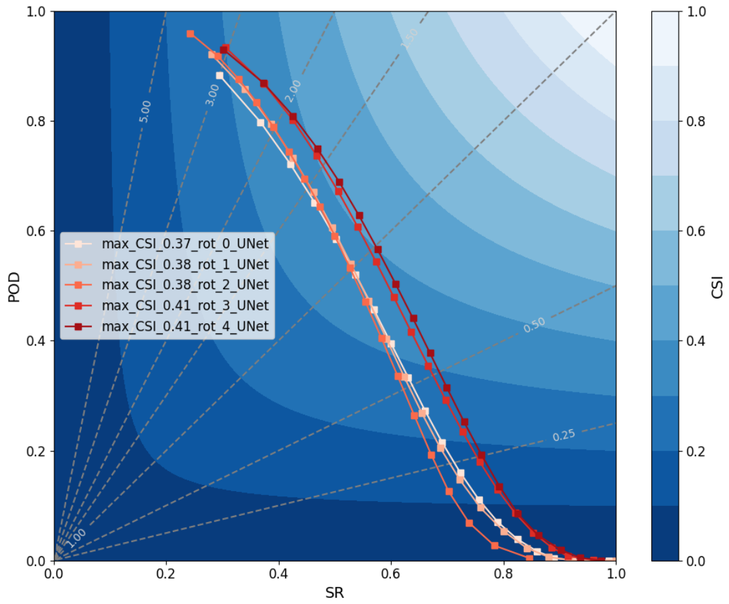
Want to see the pale blue color throughout the matrix

Un-correlated data allow for the model to generalize better

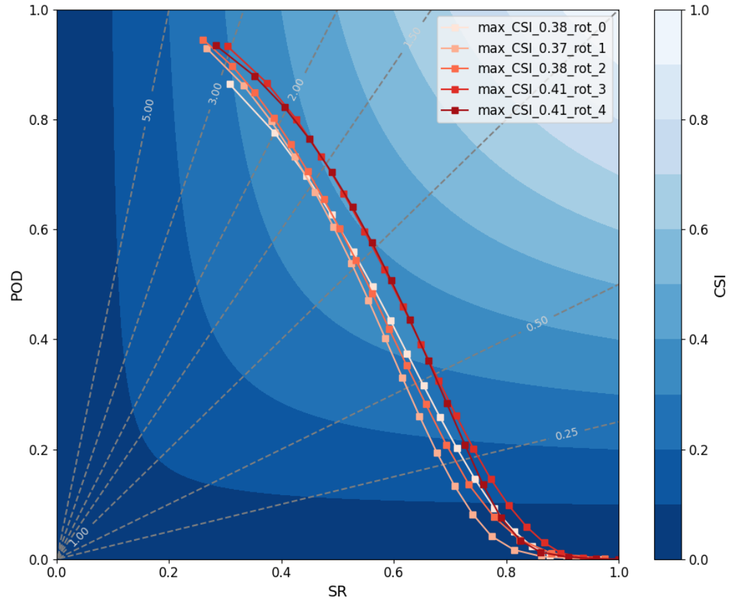
No terrain, lat, lon to prevent overfitting

# Precision Recall Curves: All Forecast Days

UNet




1-Layer LSTM




Rotation 4 achieves the best performance for the UNet and LSTM

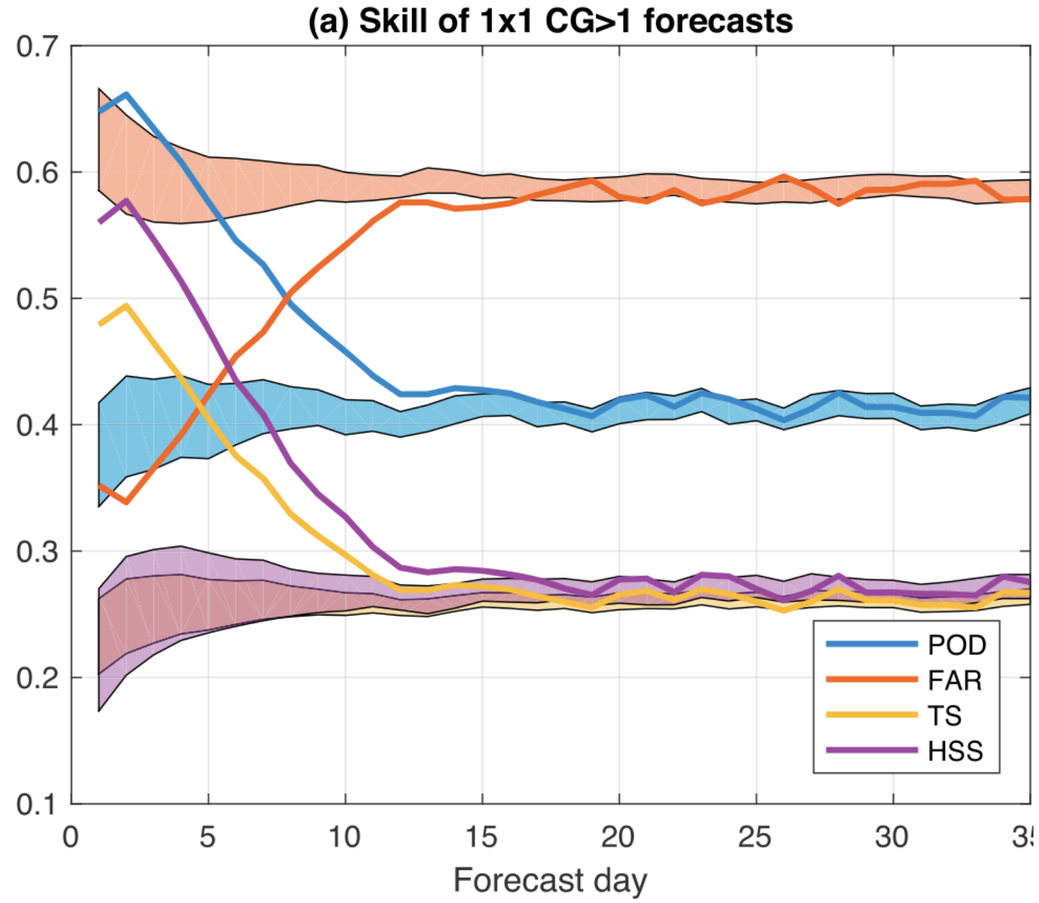
Maximum CSIs occur between 20-40% thresholds

UNet Max CSI: .41, LSTM Max CSI: .41

When the number of LSTM layers are varied, the CSI scores do not improve (not shown)

Implies complex LSTM model architectures do not add much value

# Long-Term Study




(Tippet, 2018)

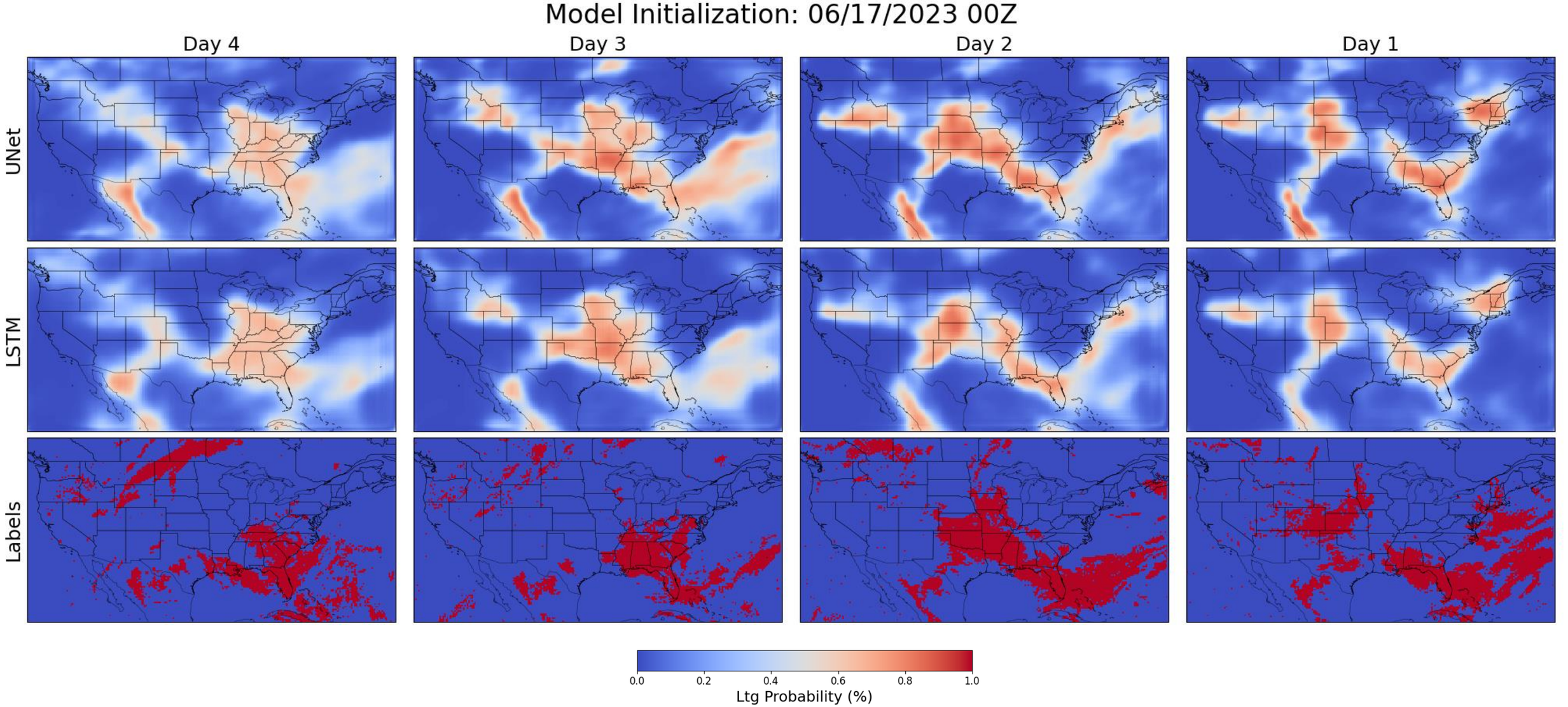
Examined correlation CAPE and precipitation rate product to a CONUS-Wide NLDN CG classification

Used the GEFS – Reforecast Dataset to examine metrics of CG flashes and how well the produce of CAPE and precipitation correlates

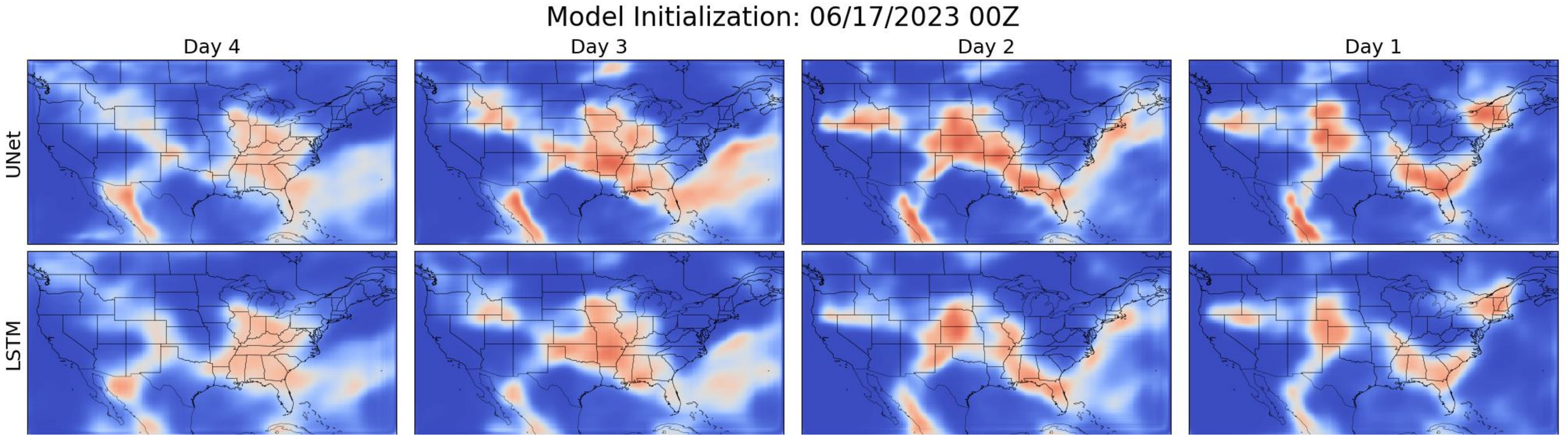
Provided thorough baseline for AI/ML performance target

Limitation of this work is that the code is not available to understand how the **lightning anomalies** are calculated for the correlations and only applies to 00Z reanalysis data

# Single Model Run/Variable Label Comparison



# Single Model Run/Variable Label Comparison




Overall, the 1-layer LSTM model (bottom) is in spatial agreement with the 3-D Convolutional UNet (top)

Model discrepancies do exist when compared to the labels

The 1-layer LSTM model trains a lot faster than the 3-D Convolutional UNet model

Day 2, qualitatively seems to be generating the higher lightning probabilities and aligned best with the labels



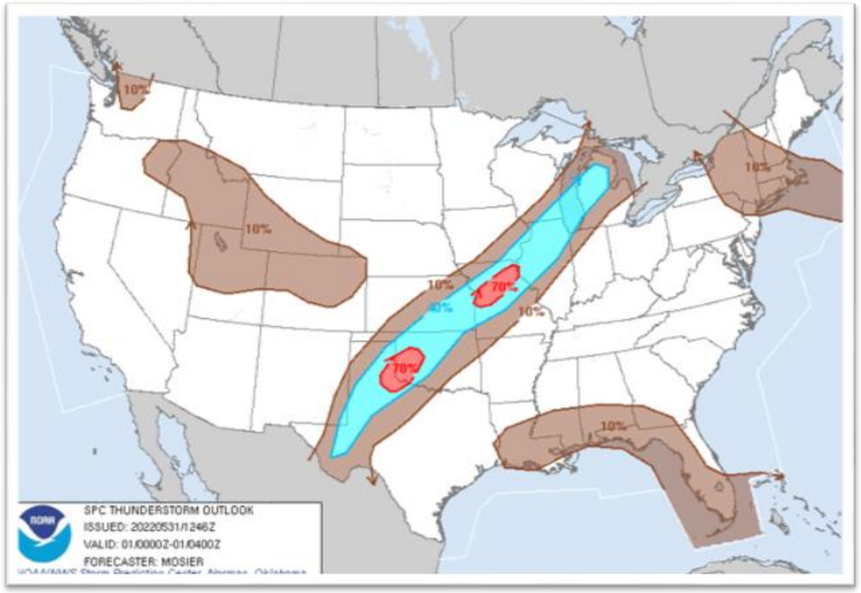
Ltg Probability (%)

# Research Gap and Question

**Research Gap/Question:** AI-based medium range lightning prediction out to 4 days. Both AI and physics based nowcasting or climate schemes exist. Operationally relevant for resource allocation and protection. Can AI be used responsibly to predict binary lightning probabilities?

**Hypothesis:** If a model can be trained/stabilized, then the critical success index (CSI) will decrease with increasing lead time.

**Model Design:** Model is designed for the operational forecaster, considering the data that are available to these forecasters. Re-analysis data are generally not available to forecasters for operational forecasts. Not intended to replace product. Trained on all-seasons, embodying responsible AI principles.



# Physics-Based Flash Rate Parameterization Schemes

**Table 1**
*Flash Rate Parameterization Schemes and Associated Required Criteria Thresholds Tested in the WRF-Chem Cloud-Resolved Simulation of 29–30 May 2012 Oklahoma Case Study*

| | | Required criteria thresholds | | | |
|---|---|---|---|---|---|
| Flash rate parameterization scheme | Flash rate equation (reference) | Reflectivity (dBZ) | Temperature (°C) | Vertical velocity (m $s^{-1}$) | Hydrometeor (g $kg^{-1}$) |
| Maximum vertical velocity ($w_{max}$) | $f = 5 \times 10^{-6} \times (w_{max})^{4.5}$ (Price & Rind, 1992) | | | Maximum | |
| Cloud top height (CTH)[a] | $f = 3.44 \times 10^{-5} \times CTH^{4.9}$ (Price & Rind, 1992) | >20 | <0 | | |
| Updraft volume (UV) | $f = 6.75 \times 10^{-11} \times UV_5$ - 13.9 (Deierling & Petersen, 2008) | | <−5 | >5 | |
| Ice water path (IWP) | FD = 33.33 × IWP - 0.17 (Petersen et al., 2005) | >25 | <−10 | | |
| Precipitation ice mass (PIM)[b] | $f = 3.4 \times 10^{-8} \times PIM$ - 18.1 (Deierling et al., 2008) | | <−5 | w < vts | snow > 0.25<br>graupel > 0.5 |
| Ice mass flux product (IMFP)[b] | $f = 9.0 \times 10^{-15} \times (f_p \times f_{np}) + 13.4$ (Deierling, 2006; Deierling et al., 2008) | | <−5 | Non-precipitation: w > vts;<br>Precipitation: w < vts | ice > 0<br>snow > 0.25<br>graupel > 0.5 |
| CSU graupel echo volume (CSU GEV) | $f = 7.0 \times 10^{-2} \times GEV$ (Basarab et al., 2015) | >35 | −40 < T < −5 | | graupel > 0.5 |
| CSU 35dBZ echo volume (CSU 35EV) | $f = 7.2 \times 10^{-2} \times 35EV$ (Basarab et al., 2015) | >35 | −40 < T < −5 | | |
| CSU precipitation ice mass (CSU PIM) | $f = 1.2 \times 10^{-7} \times PIM$ (Basarab et al., 2015) | >35 | −40 < T < −5 | | |
| UAH graupel echo volume (UAH GEV10) | f = 0.0534 × GEV10 (Carey et al., 2015) | | −40 < T < −10 | | graupel > 0.5 |
| UAH graupel echo volume (UAH GEV5) | f = 0.0430 × GEV5 | | −40 < T < −5 | | graupel > 0.5 |
| UAH updraft volume (UAH UV) | f = 0.0337 × UV | | −40 < T < −10 | >5 | |
| UAH graupel echo volume for supercells (UAH GEV10S) | f = 0.0674 × GEV10S | | −40 < T < −10 | | graupel > 0.5 |
| UAH graupel echo volume for supercells (UAH GEV5S) | f = 0.0569 × GEV5S | | −40 < T < −5 | | graupel > 0.5 |
| UAH updraft volume for supercells (UAH UVS) | f = 0.0403 × UVS | | −40 < T < −10 | >5 | |
| Upward cloud ice flux (ICEFLUX) | $FD_I = 6.58 \times 10^{-7} \times ICEFLUX$ (Finney et al., 2014) | | | w > vts | ice > 0<br>snow > 0.25 |
| OK 35dBZ echo volume (OK 35EV) | f = 0.0111 × 35EV + 1.6619 | >35 | −40 < T < −5 | | |
| Blended lightning threat (LTG3) | $FD_{LTG3}$ = (0.95 × graupel flux) + (0.05 × VII) (McCaul et al., 2009, 2020) | | T = −15 | | graupel > 0.5 |

*Note*. f is total flash rate in flashes $min^{-1}$, FD is flash density in flashes $km^{-2}$ $day^{-1}$, $FD_I$ is flash density over land in flashes $m_{grid\ cell}^{-2}$ $s^{-1}$, and $FD_{LTG3}$ is flash density in flashes $km^{-2}$ 5-$min^{-1}$. FRPS units and the storm parameters used within each flash rate equation are further defined in Section 2. [a]Indicates the highest altitude where both criteria are met. [b]Indicates the storm vertical velocity (w) and snow fall speed (vts) are both required.

(Cummings, et al. 2024)

Cummings summarized several FRPS in his most recent work

These are mostly for WRF simulations and storm specific, and can be calculated on the internal time steps of the model

Radar reflectivity, vertical velocity, temperature, and hydrometeor mixing ratios were common inputs into the FRPS

Provided strong foundation to use similar inputs for BoltCast deep learning model

# AI/ML Publication Rate

