# Evidential Deep Learning for Enhanced Winter Precipitation Prediction and Decision-Making

**Sophia Reiner**
University of Wisconsin - Madison

David John Gagne II, Charlie Becker, John Schreck, Julie Demuth, Chris Wirz, Gabrielle Gantos
***M**achine **I**ntegration and **L**earning for **E**arth **S**ystems*
**July 31, 2024**



This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the U.S. National Science Foundation under Cooperative Agreement No. 1852977.

# Overview

1. **Introduction & Motivation**
2. **Model Framework & Data**
   - ➢ quality control
3. **Model Results**
   - ➢ ECHO optimization
4. **Forecaster Guidance**
   - ➢ Interview guide
   - ➢ Analysis
5. **Conclusions**

# Motivation

Winter precipitation hazards, such as rain, snow, freezing rain, and sleet, significantly impact human safety and transportation.

# What I did

Enhanced model accuracy through hyperparameter optimization and quality control of training data.

Analyzed model performance against Numerical Weather Prediction (NWP) models and investigated failure modes

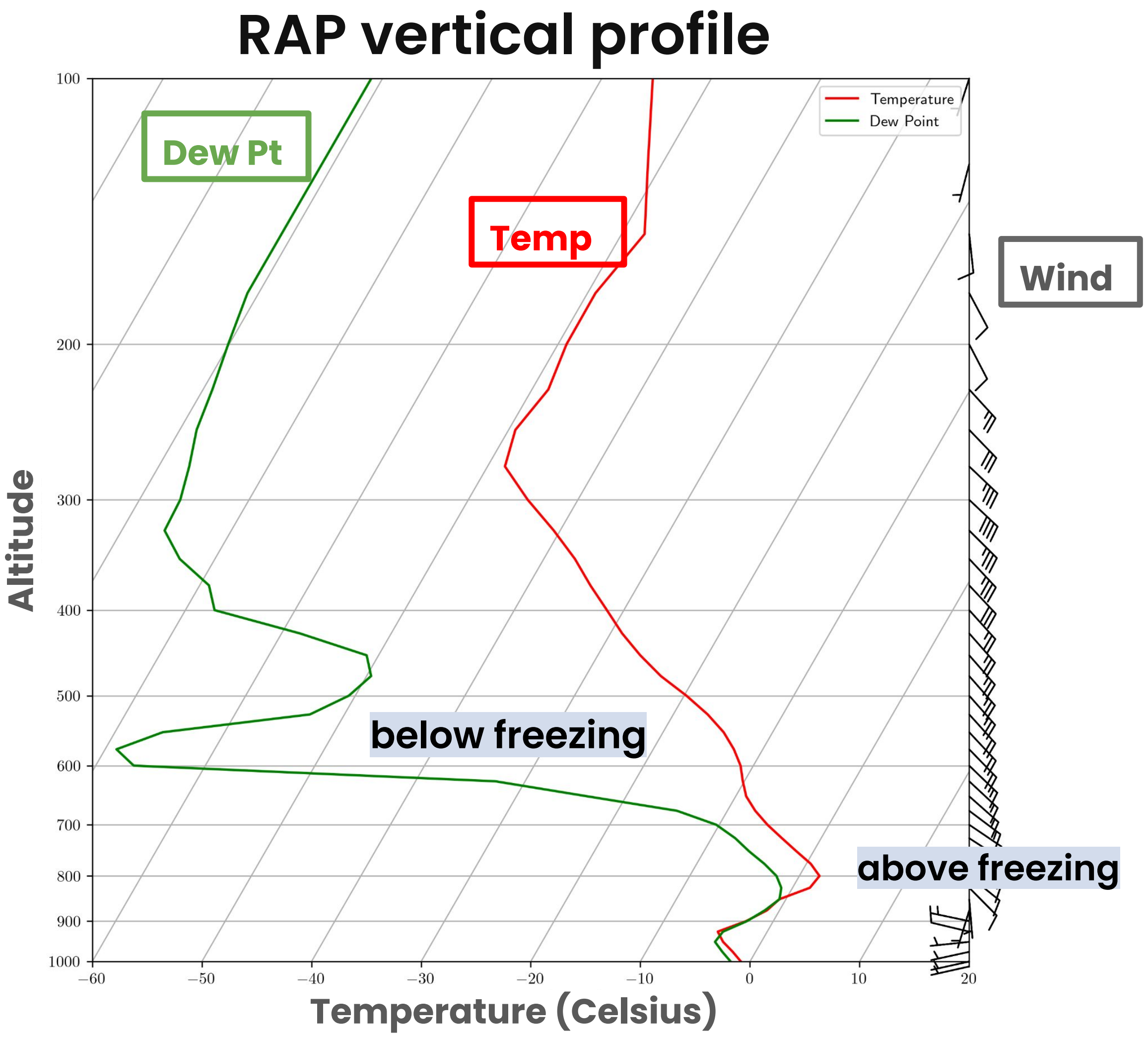
Developed visual representations of model results to ensure transparency and reliability for forecasters.

Made bug fixes in MILES / ECHO github repositories

## Input Data

**NOAA Rapid Refresh** Vertical Profile

- ➢ 0 - 5 km above surface, every 250 meters)
- ➢ Temperature, Dewpoint, U-Wind, V-Wind

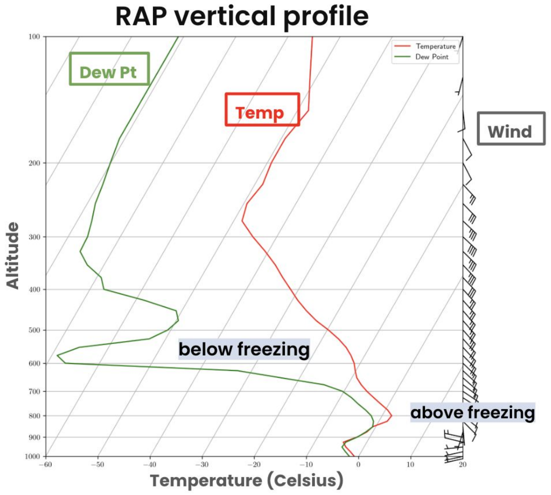


# Model Framework

## Input Data

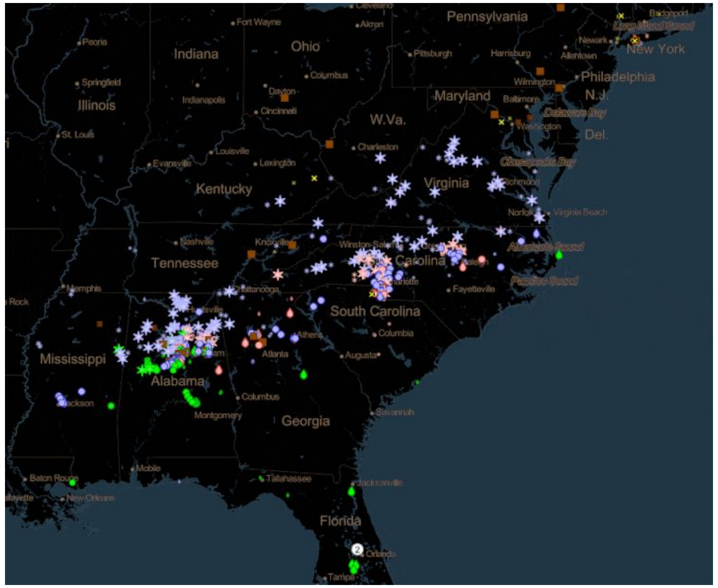
**NOAA Rapid Refresh** Vertical Profile

- ➢ 0 - 5 km above surface, every 250 meters)
- ➢ Temperature, Dewpoint, U-Wind, V-Wind




## Target

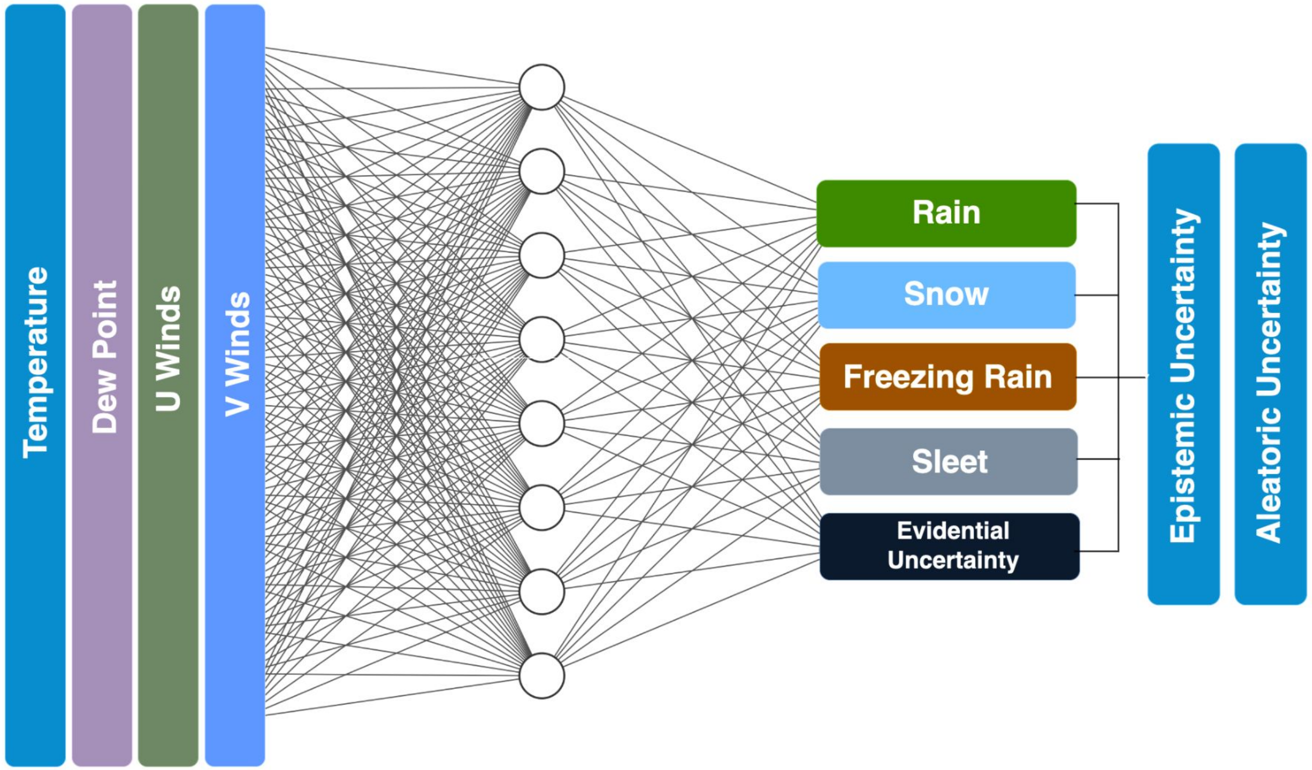
- ➢ mPING Observations of precipitation types
  - ➢ *Rain, Snow, Sleet, Freezing Rain*




mPING
crowdsourcing weather reports

# ML Methods

The model is similar to a simple dense neural network with a custom evidential loss function.



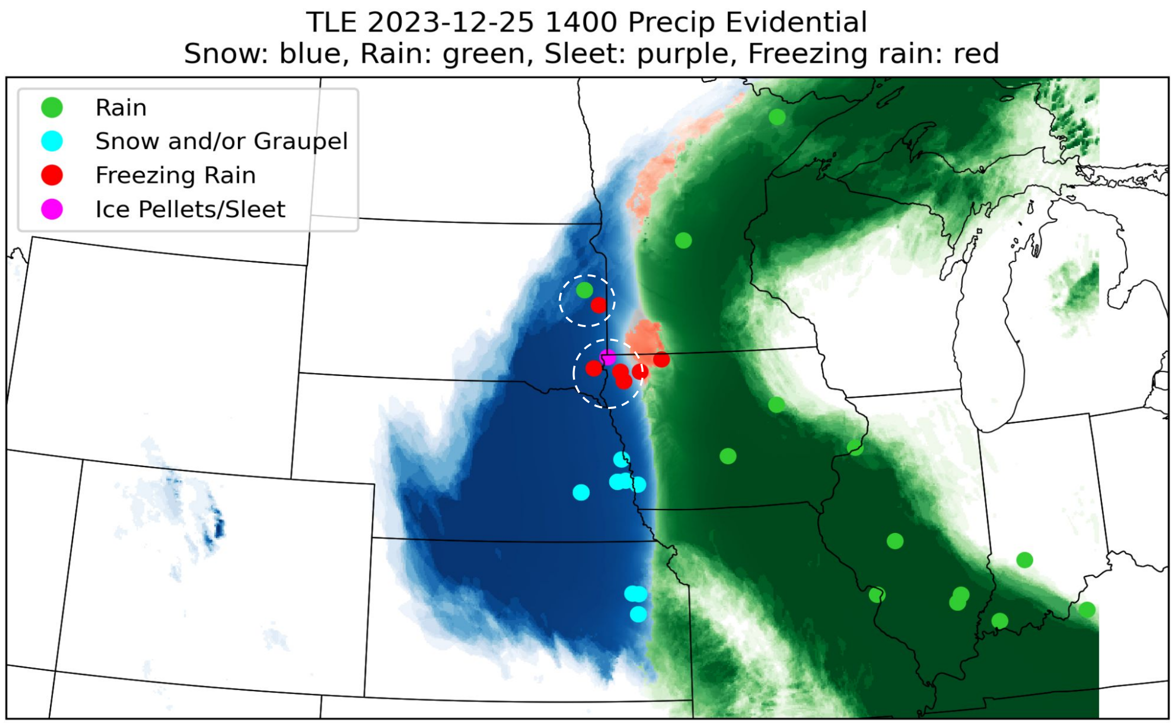

**Outputs:**

- probability of rain, snow, freezing rain, sleet
- pseudo-class of “I don’t know” representing evidential uncertainty
- uncertainties for each class are derived

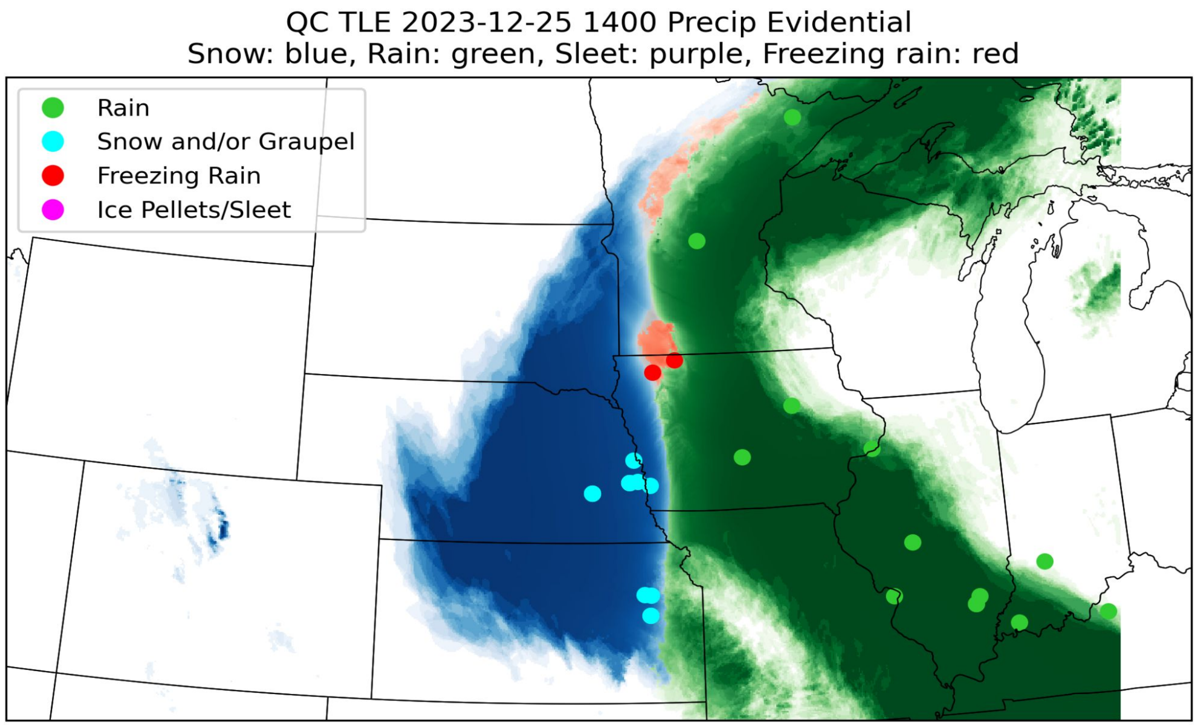
Schreck, J. et. al., 2024: Evidential Deep Learning: Enhancing Predictive Uncertainty Estimation for Earth System Science Applications

# Data & Quality Control

- ➢ **mPING** has more spatial coverage in populated regions as compared to other sources
- ➢ **Quality control** is a necessary step to reduce bias and human error



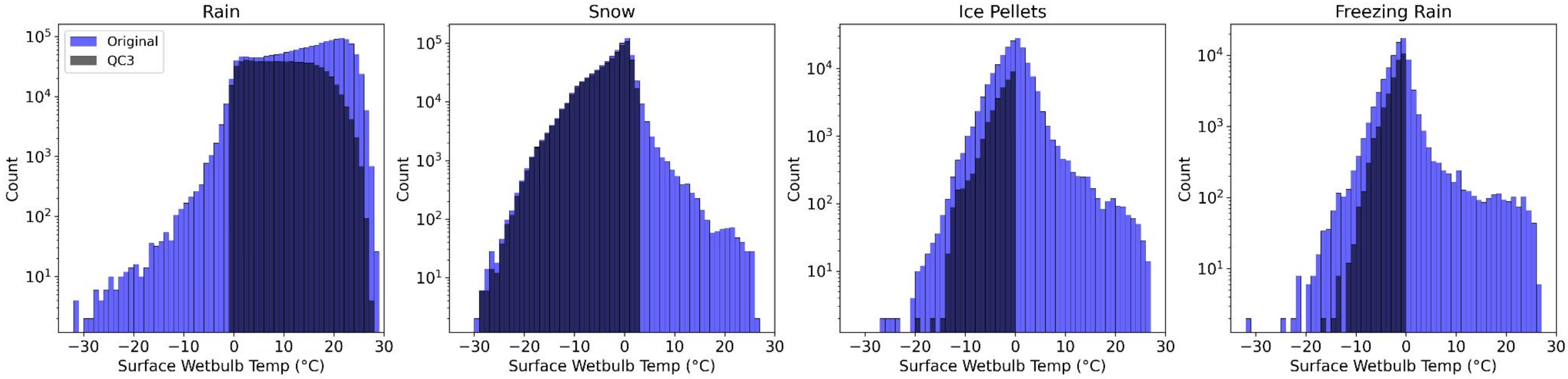

before qc



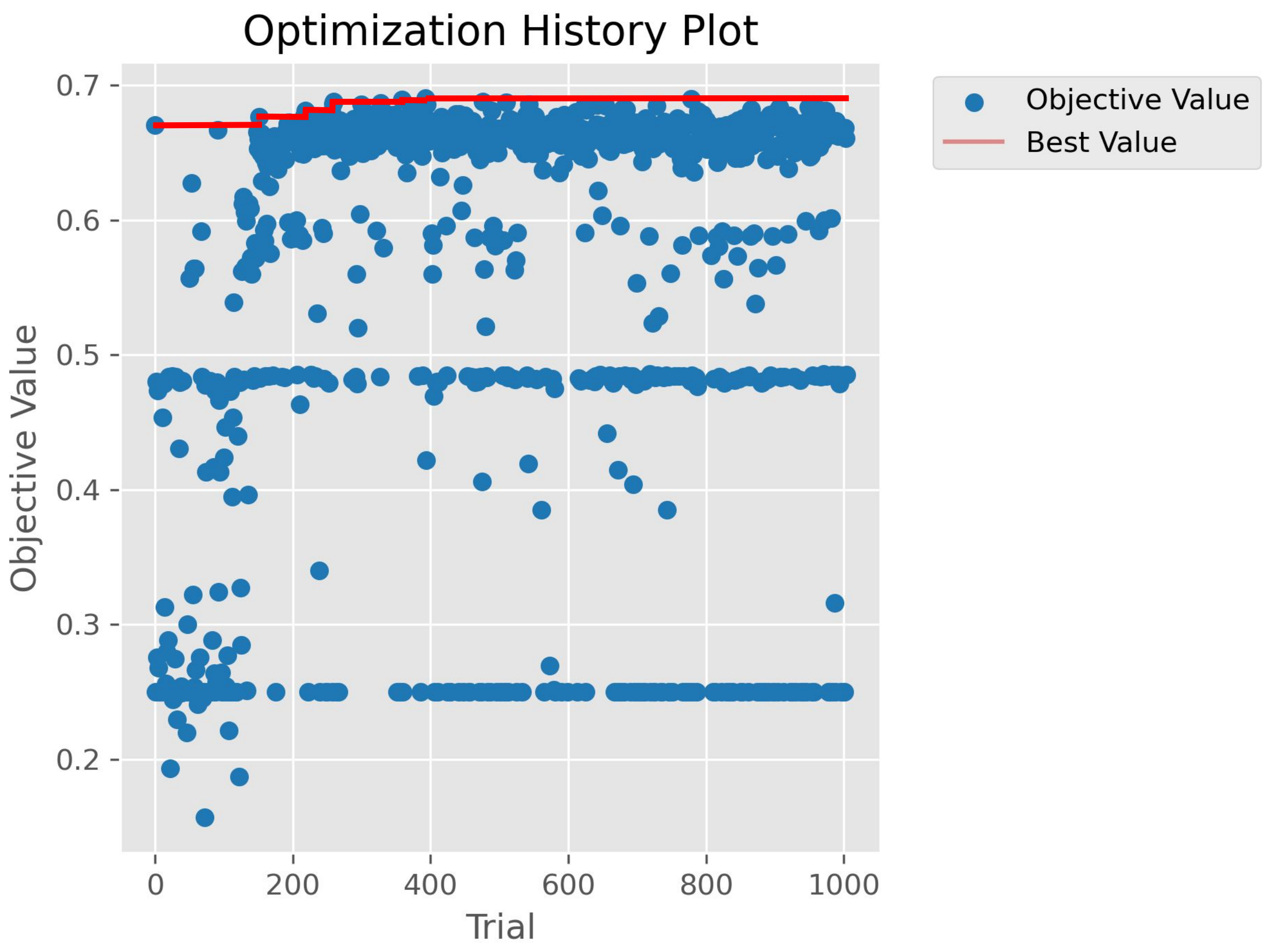

after qc

**QC procedure:**

- Locate closest grid cell to mPING coordinate
- Compute surface wet bulb temperature
- **Rain**: wet bulb temp > -1°C
- **Snow**: wet bulb temp < 3°C
- **Freezing rain & sleet (ice pellets)**: wet bulb temp < 0°C and at least 1 temp crossing from -2 to 1°C



# Hyperparameter Optimization

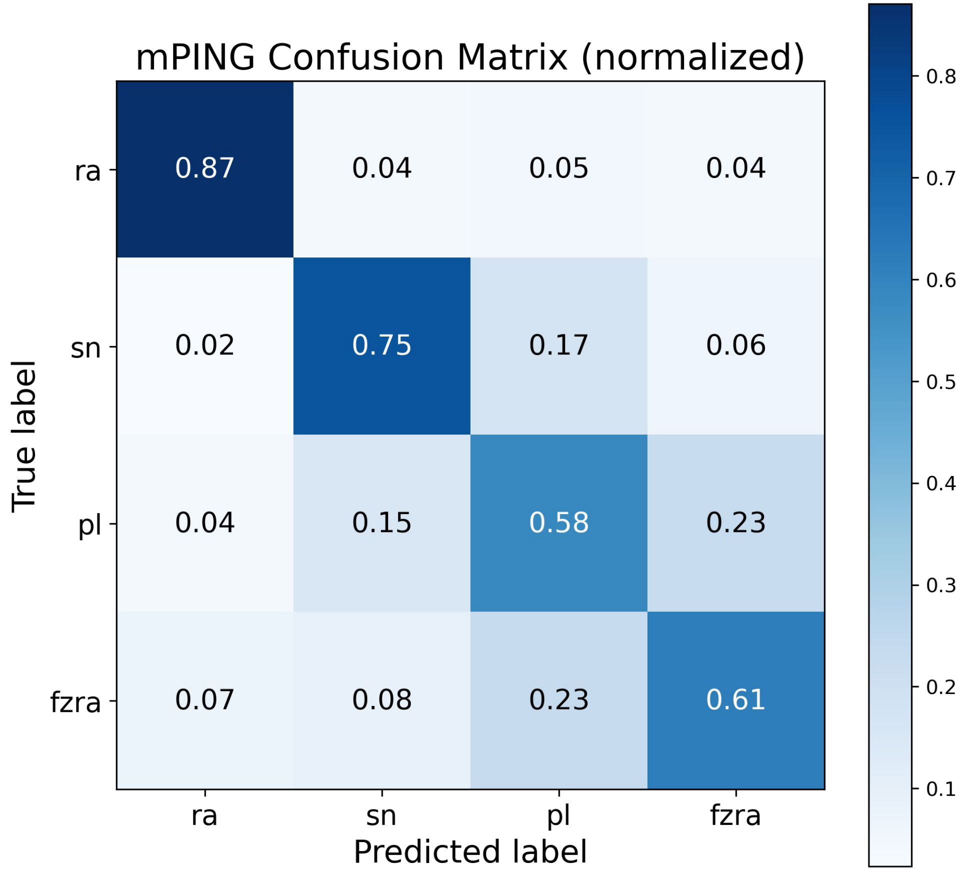



Ran hyper parameter optimization on newly QC'd data using **Earth Computing Hyperparameter Optimization** (ECHO) package

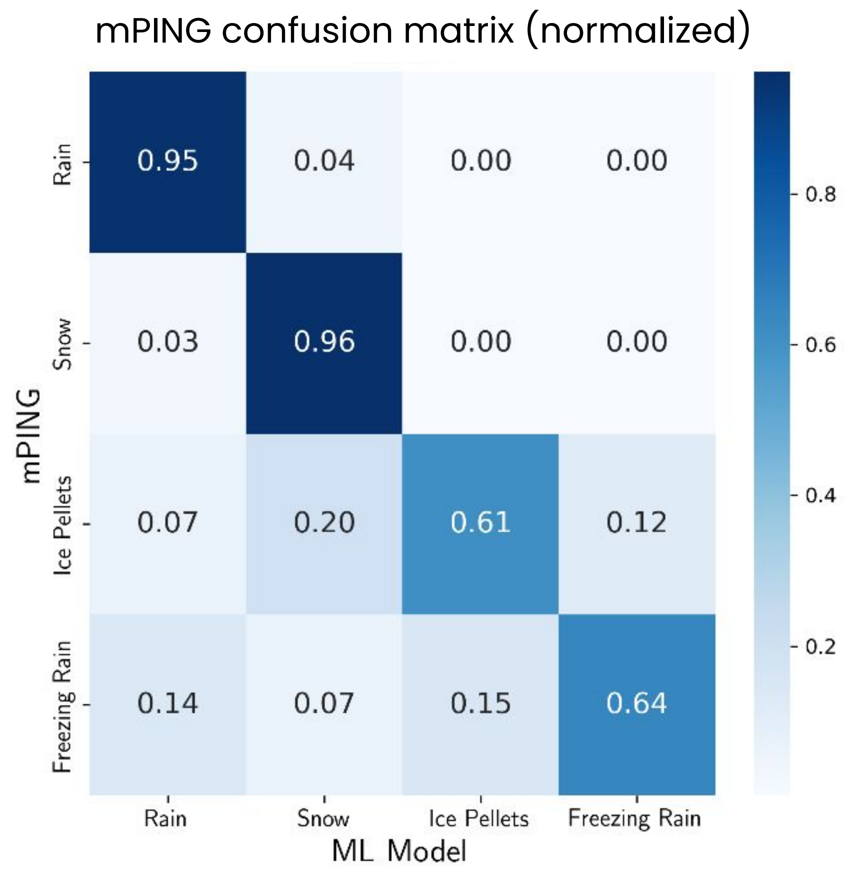
- 1000 trials
- **maximize average validation accuracy**

# Hyperparameter Optimization Results

## Previous qc



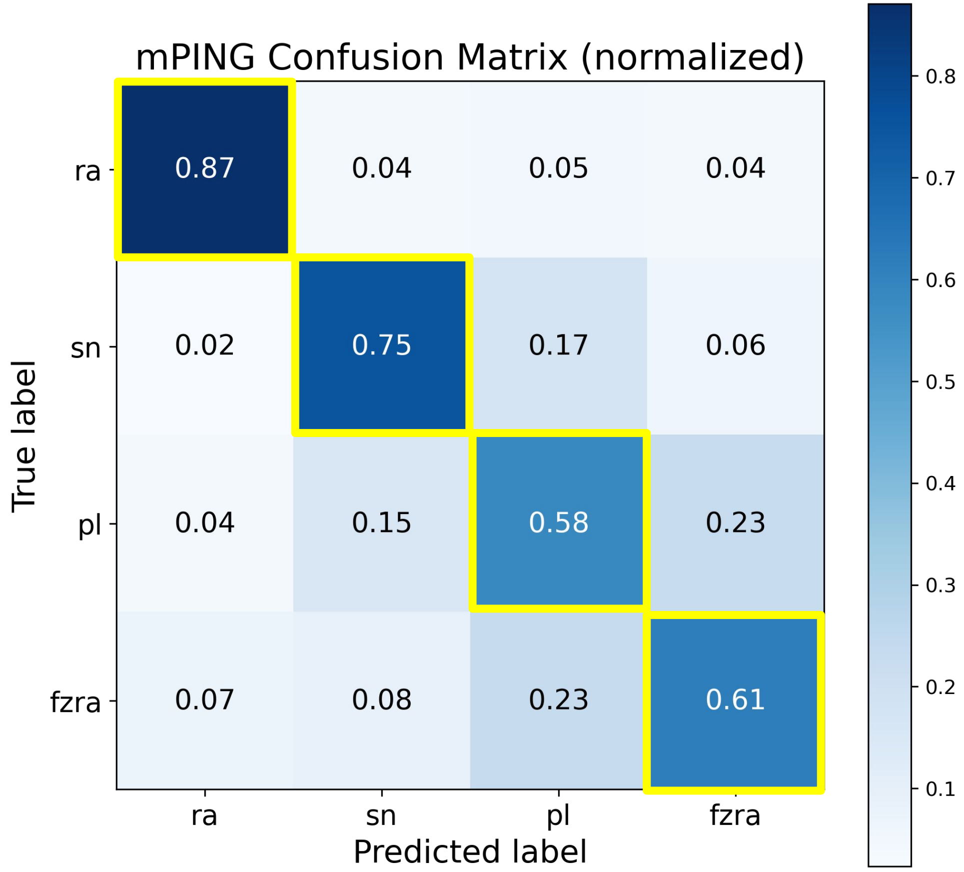

## New qc



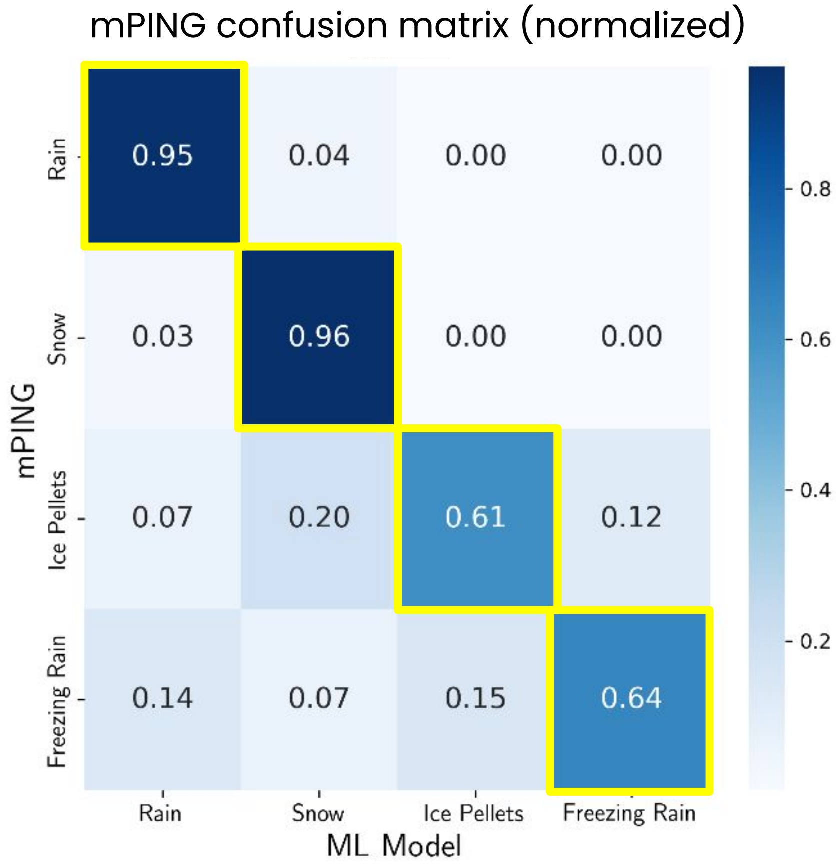

* normalized by truth

# Hyperparameter Optimization Results

## Previous qc



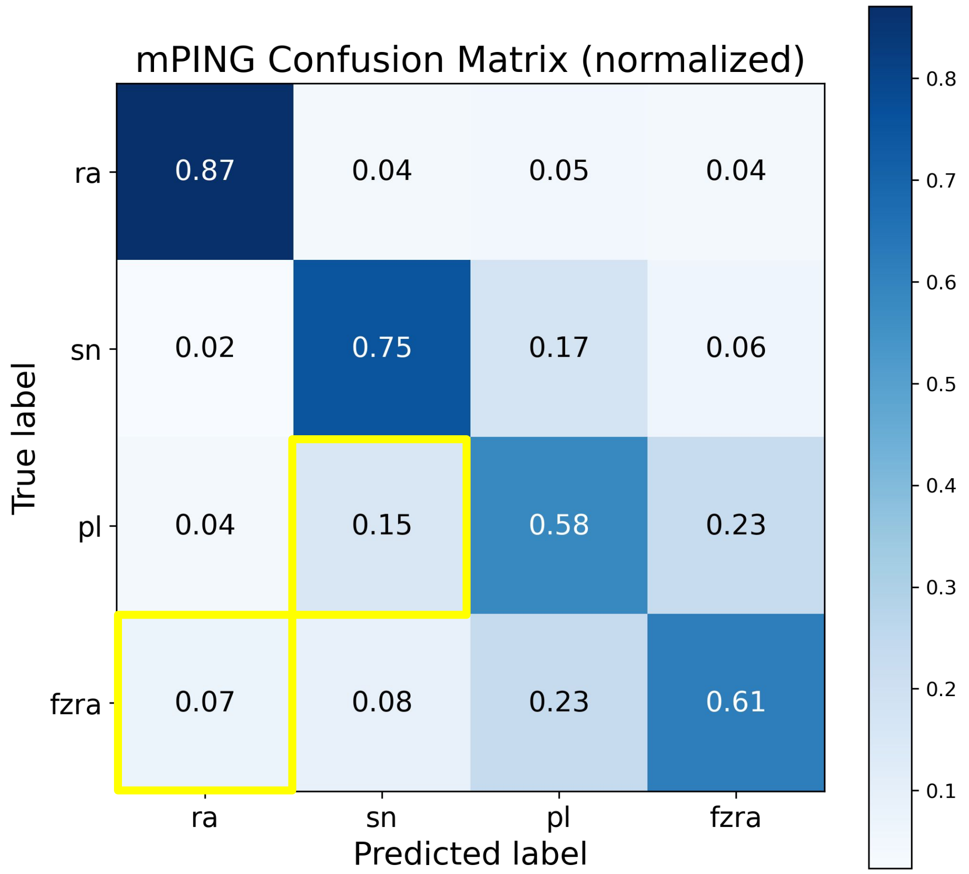

## New qc



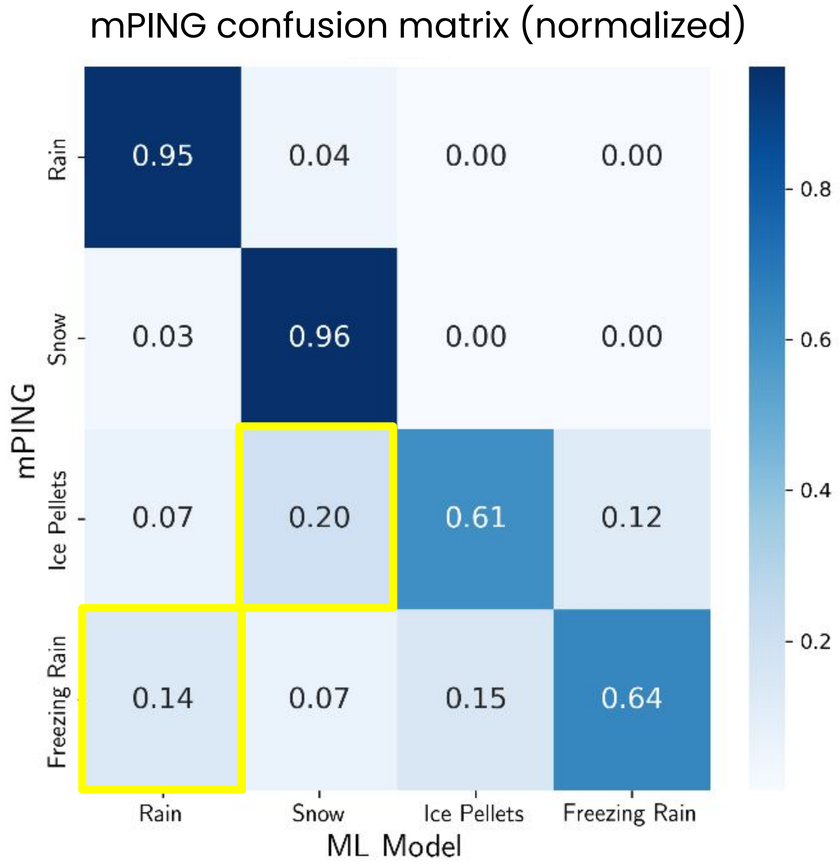

* normalized by truth

# Hyperparameter Optimization Results

## Previous qc




## New qc




* normalized by truth

# Trustworthy AI & Forecaster Guidance

## Interdisciplinary Research Team

**Risk Communication Scientists**
**Environmental & Atmospheric Scientists**
**AI/ML Scientists and Developers**

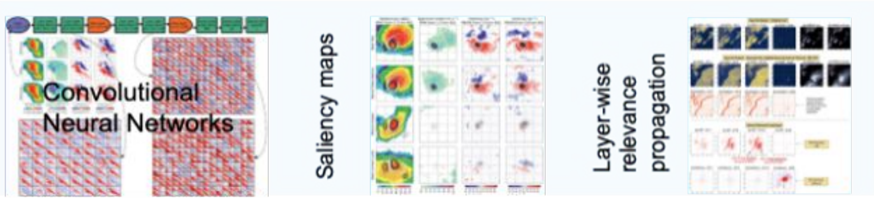
## Social Science Data Collection Methods

Formative Research

&

Evaluative Research

## AI/ML & XAI Techniques




## Expert End-users & Decision Makers

EMERGENCY MANAGEMENT
PREPAREDNESS
RESPONSE
RECOVERY
MITIGATION

US Army Corps of Engineers.

AAPA

USGS science for a changing world

## Trustworthy & User-Driven AI/ML-derived Weather Information and Products

**Figure courtesy of Mariana Cains**

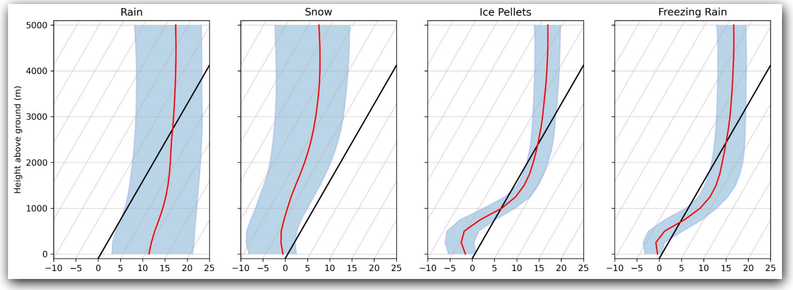
## Considerations when selecting new forecast guidance:

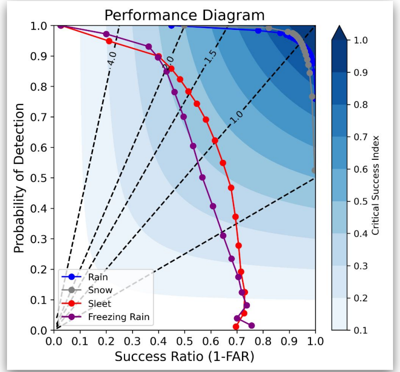
How new guidance verifies compared to existing guidance

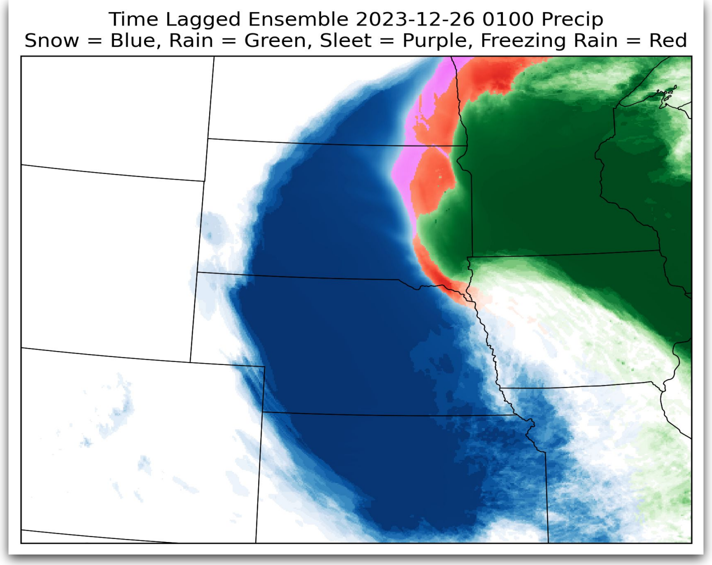
Understanding failure modes

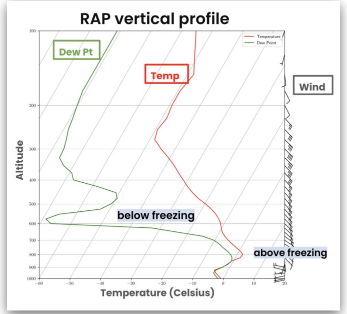
Ability to examine guidance predictions for archived cases

Understanding inputs



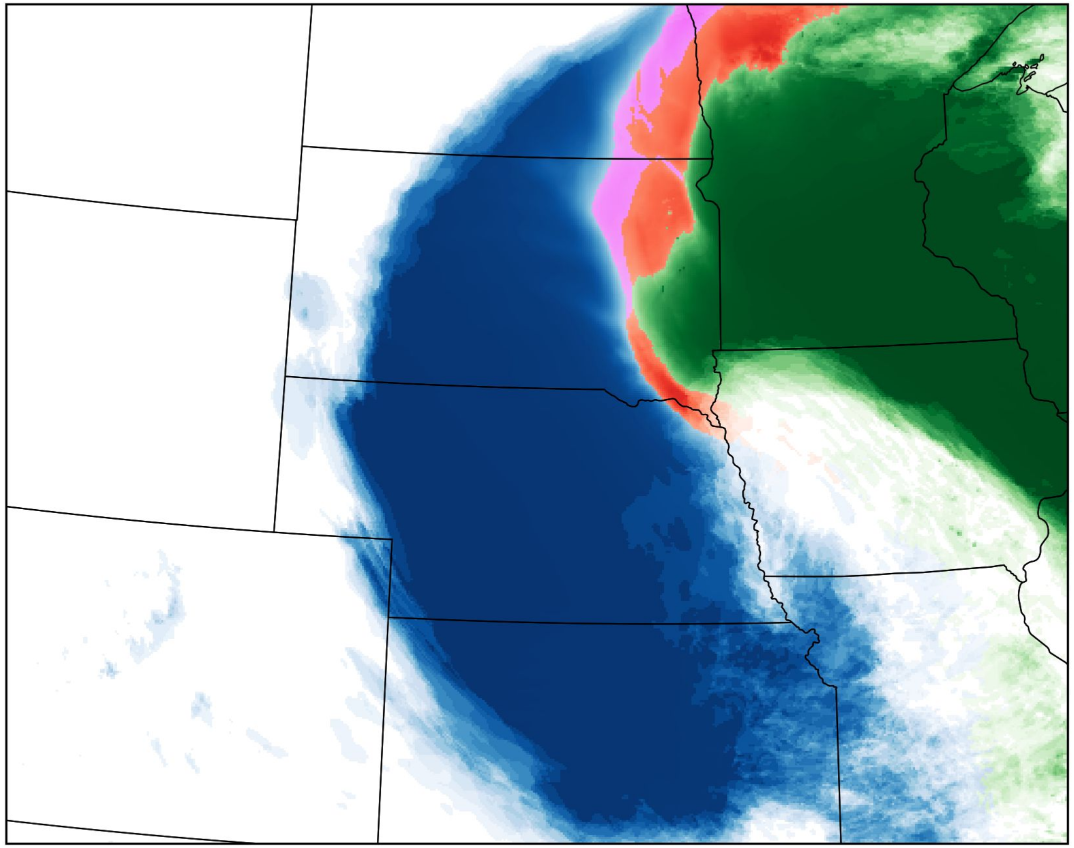



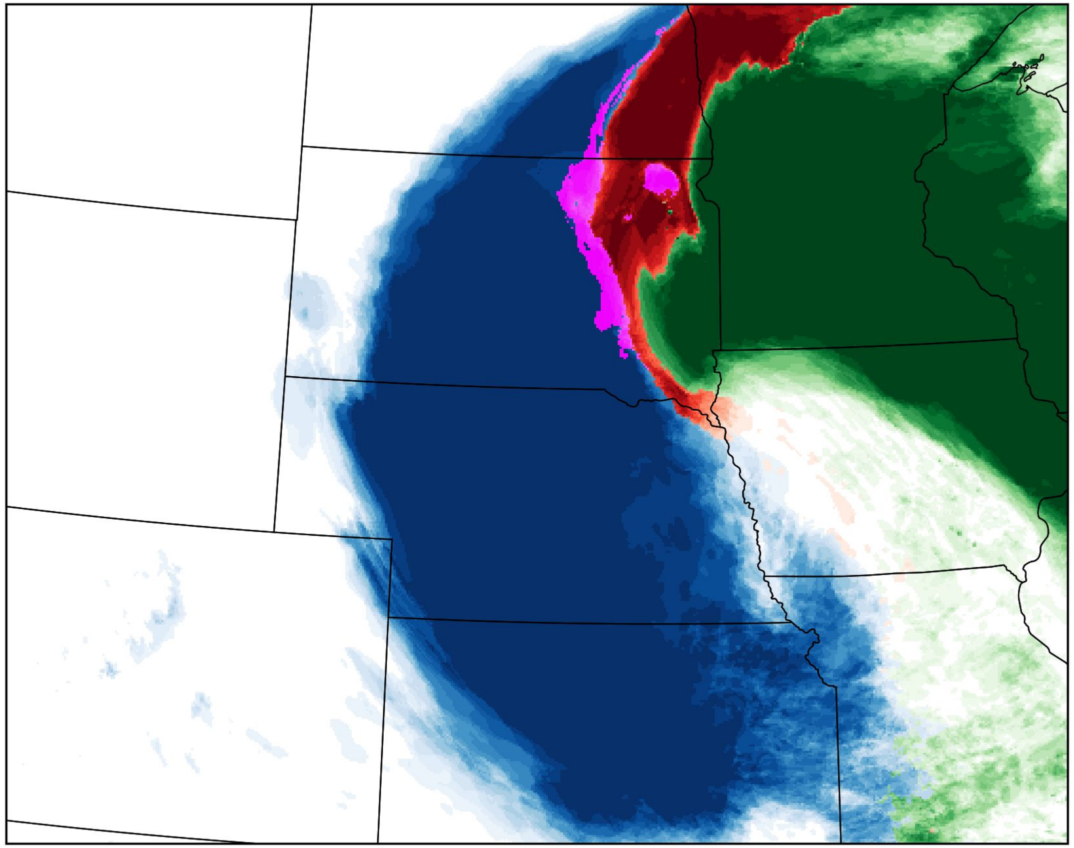

(Cains et. al 2024)

# Comparing New Guidance to Existing Guidance

Comparing HRRR precip categorizations to Evidential categorizations



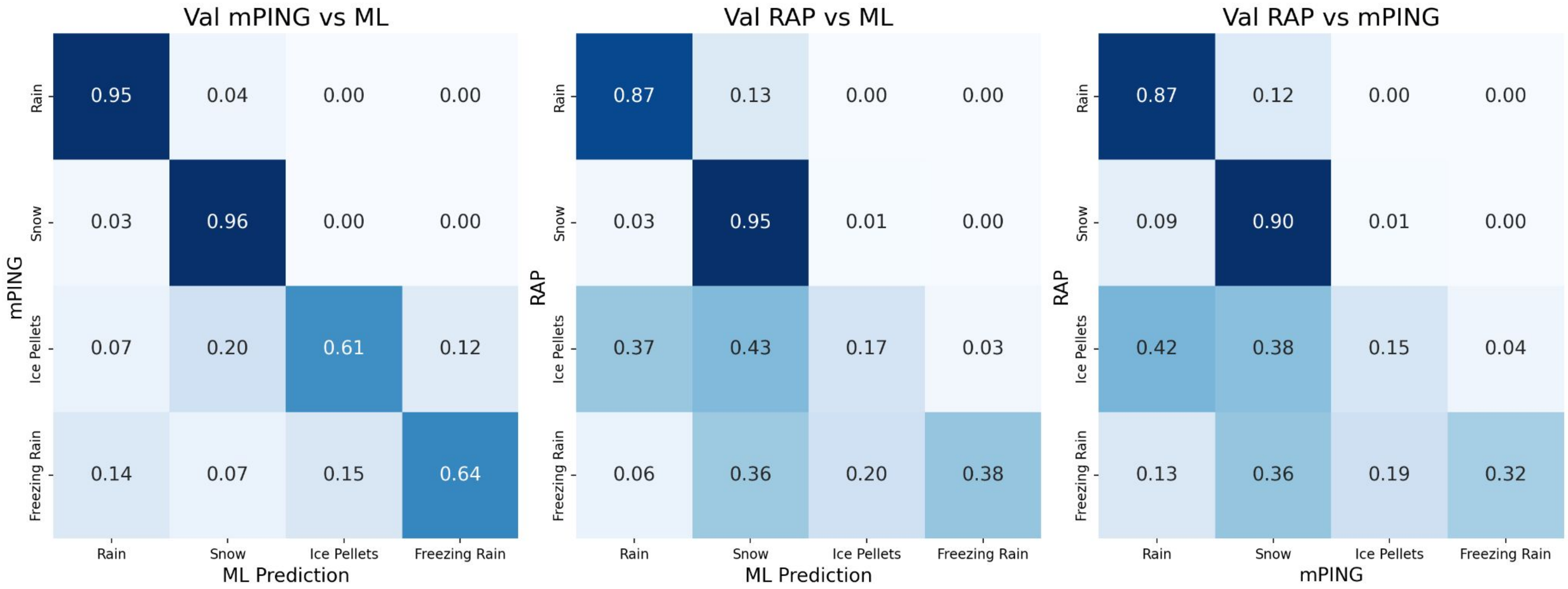

Evidential




HRRR

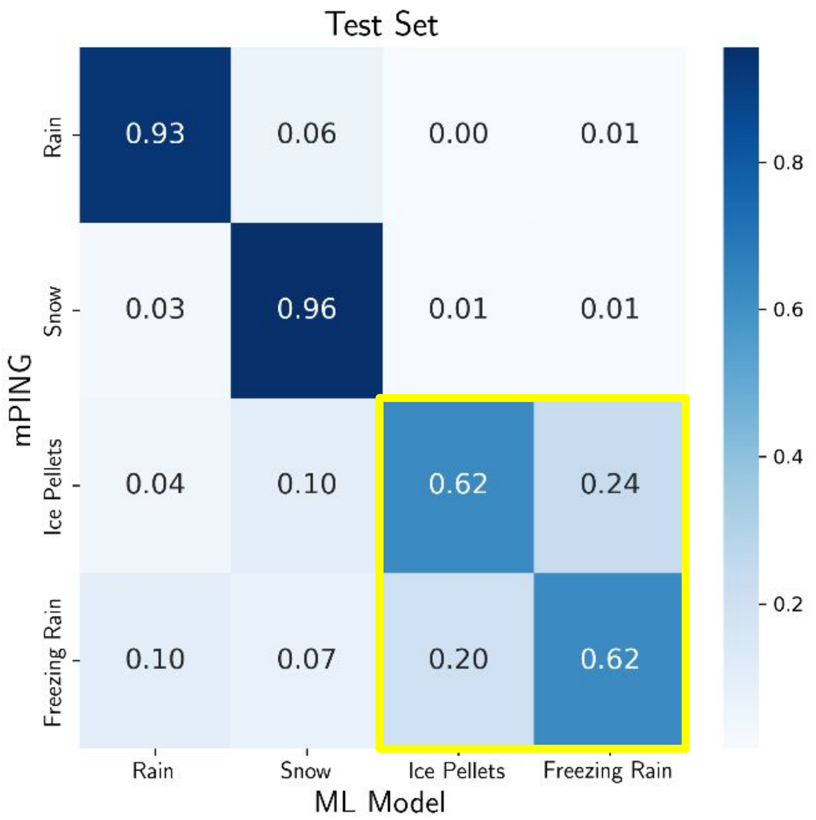
# Comparing New Guidance to Existing Guidance

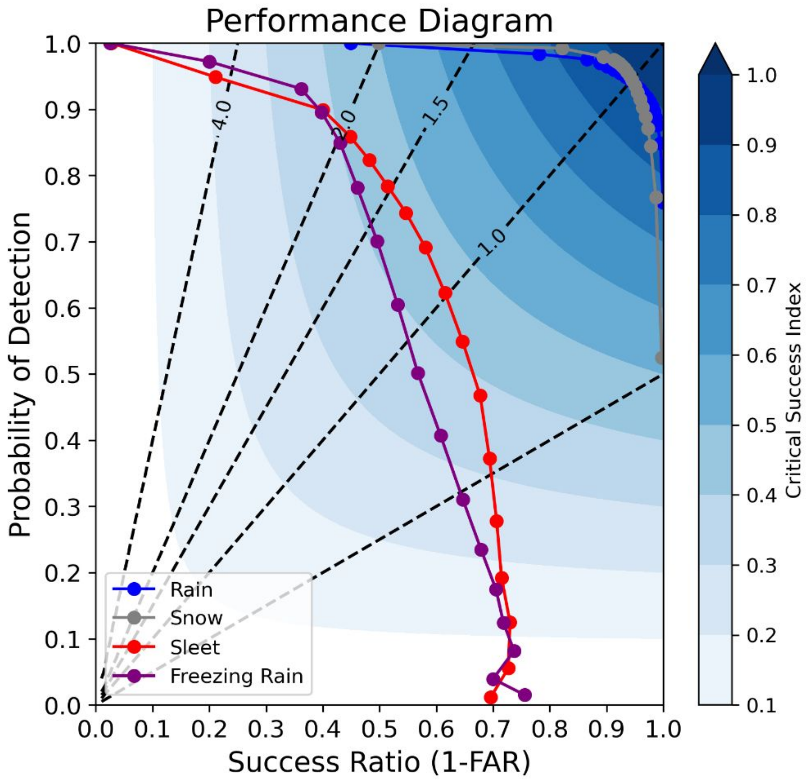
Comparing HRRR precip categorizations to Evidential categorizations



# Understanding Failure Modes

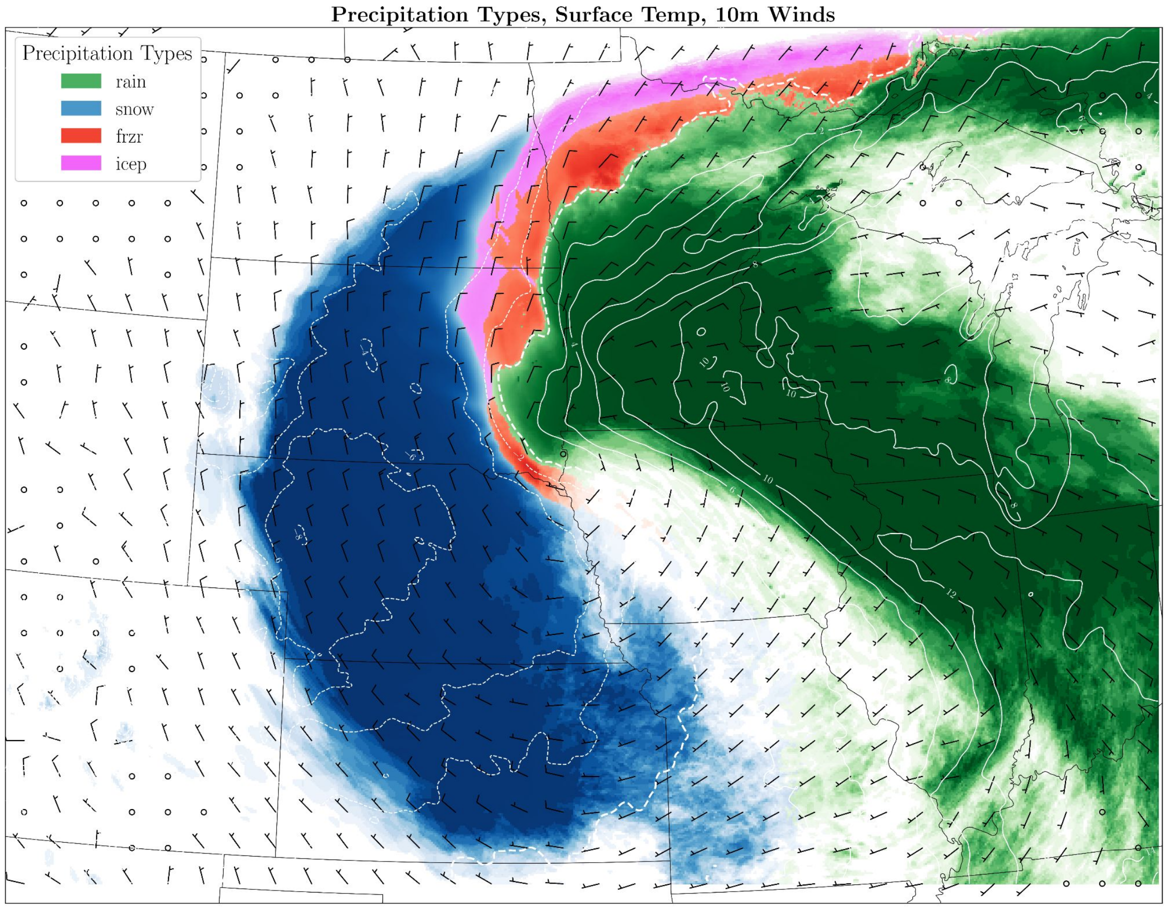
The evidential model is more likely to incorrectly classify freezing rain and sleet.

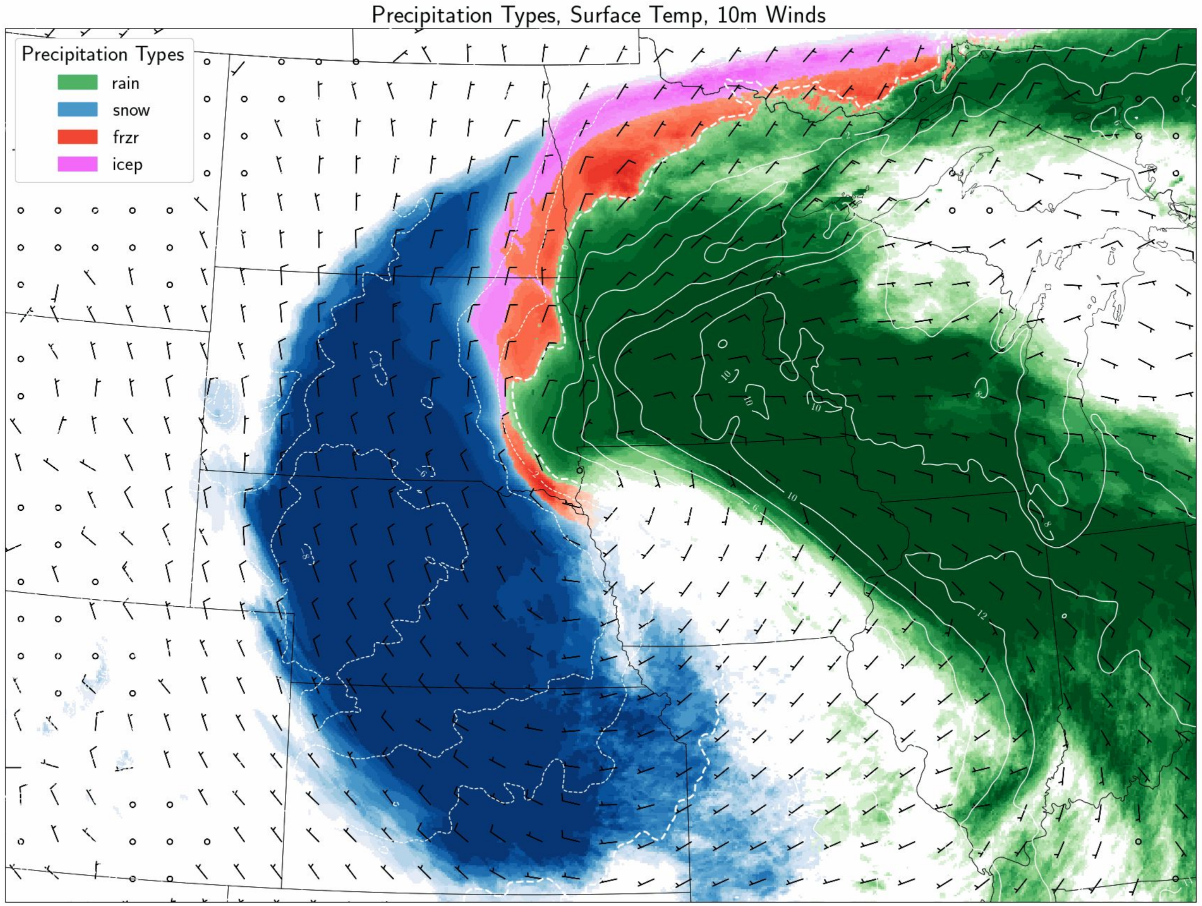


# Examining Predictions of Archived Cases

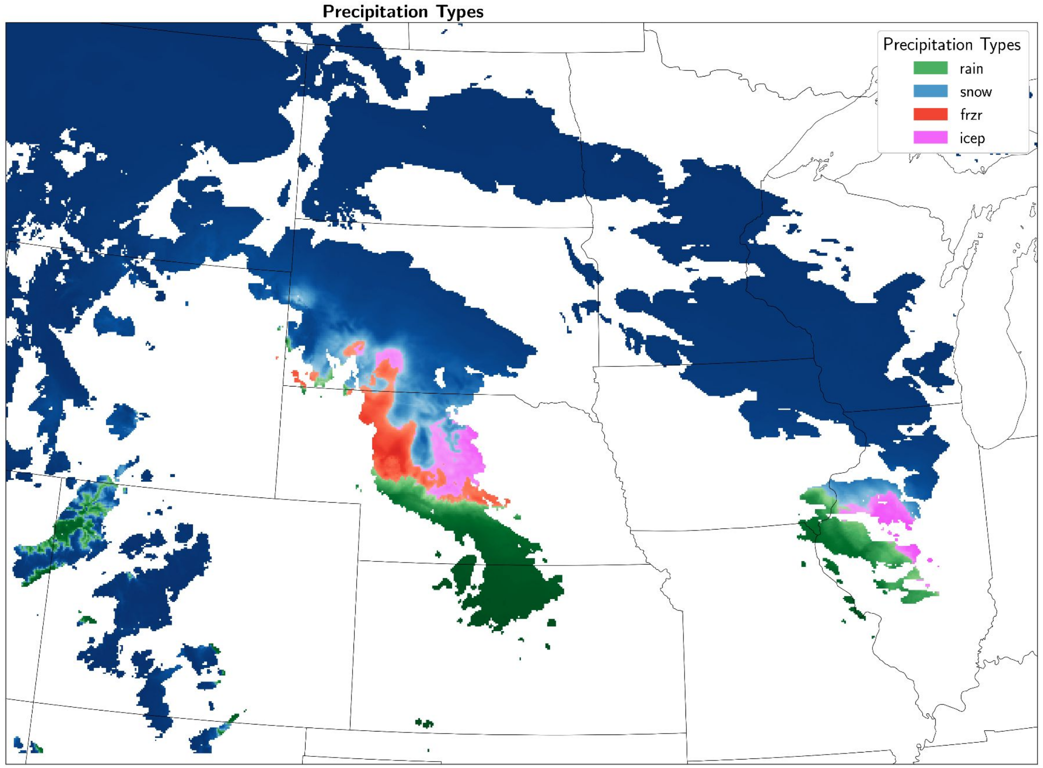
## Case Study 1: 2023-12-26 0100

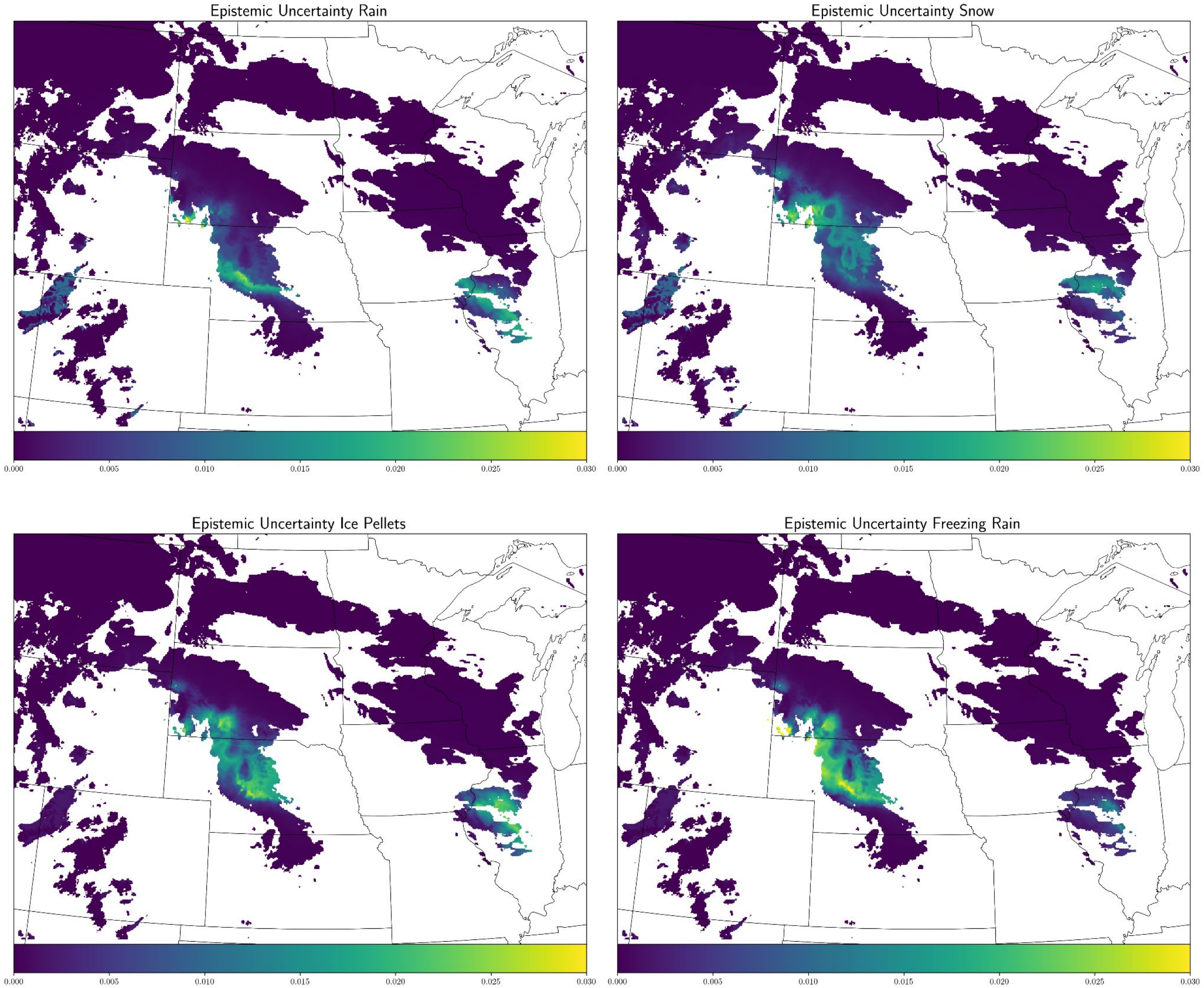


# Examining Predictions of Archived Cases

Case Study 2: 2024-03-24 0800



# Conclusion

Quality control procedures combined with ECHO optimization improves model performance

P-Type model performs well with rain and snow and under-performs with sleet and freezing rain

Case studies show evidence of model performance and help forecasters understand the basis for its predictions.

# Acknowledgements


sreiner@ucar.edu
sophiaanr

**Thanks to**

→ **MILES group:**
David John Gagne II, Charlie Becker, John Schreck, Julie Demuth, Chris Wirz, Gabrielle Gantos

→ **SIParCS:**
Virginia Do,  Jessica Wang, Jerry Cyccone, and the intern cohort

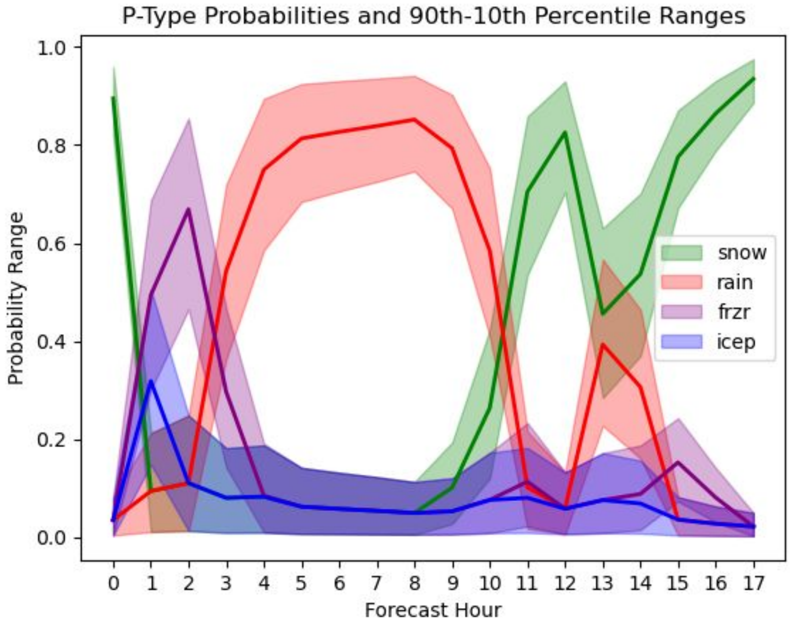
# References

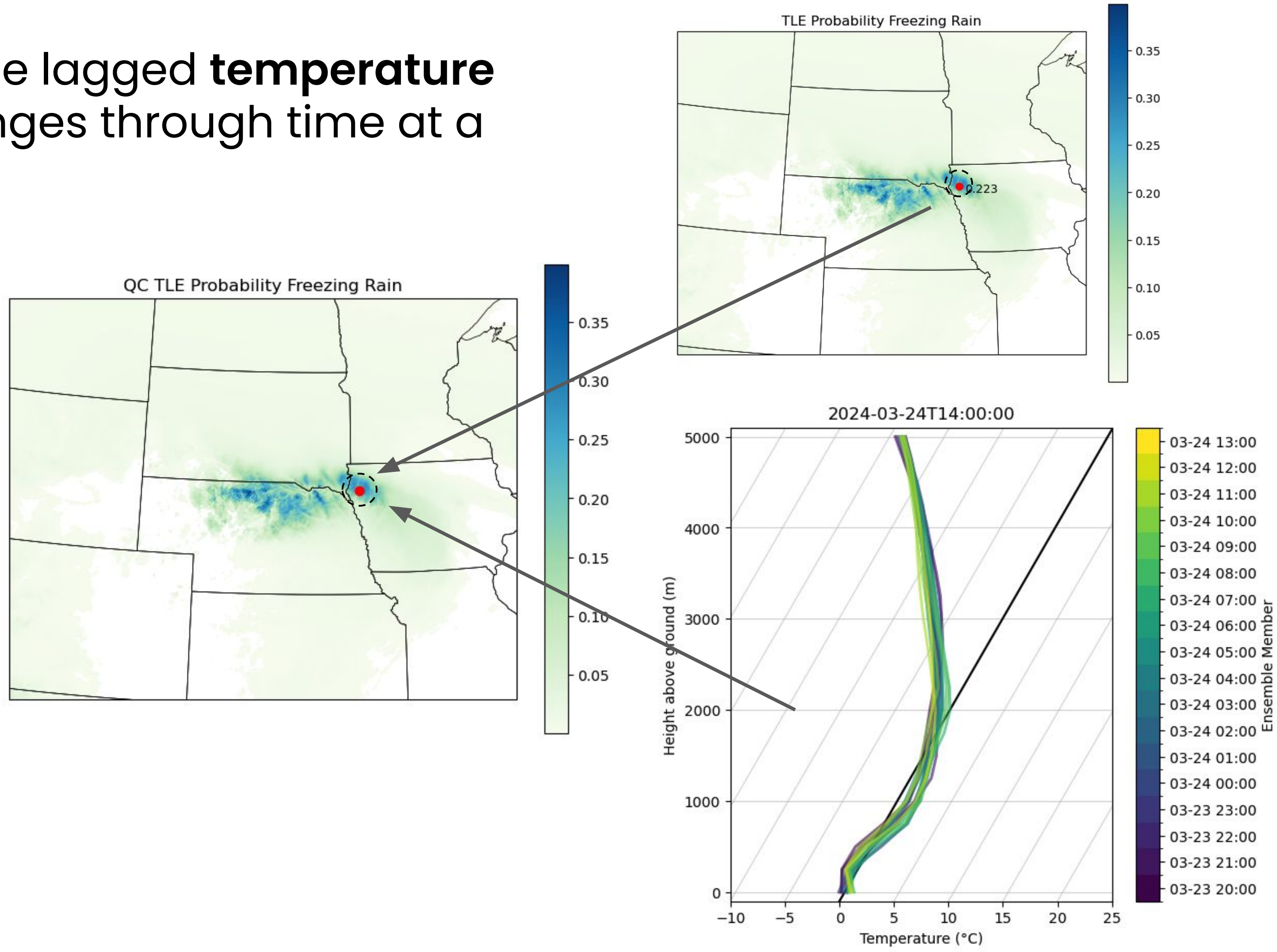
Schreck, J. S., Gagne II, D. J., Becker, C., Chapman, W. E., Elmore, K., Fan, D., Gantos, G., Kim, E., Kimpara, D., Martin, T., Molina, M. J., Pryzbylo, V. M., Radford, J., Saavedra, B., Willson, J., & Wirz, C. (2024). Evidential Deep Learning: Enhancing Predictive Uncertainty Estimation for Earth System Science Applications. arXiv. https://arxiv.org/abs/2309.13207

Cains, M. G., Wirz, C. D., Demuth, J. L., Bostrom, A., Gagne II, D. J., McGovern, A., Sobash, R. A., & Madlambayan, D. (2024). Exploring NWS Forecasters' Assessment of AI Guidance Trustworthiness. Weather and Forecasting. https://doi.org/10.1175/WAF-D-23-0180.1

# Discussion

**Probability time series** and time lagged **temperature profiles** can also visualize changes through time at a specific location



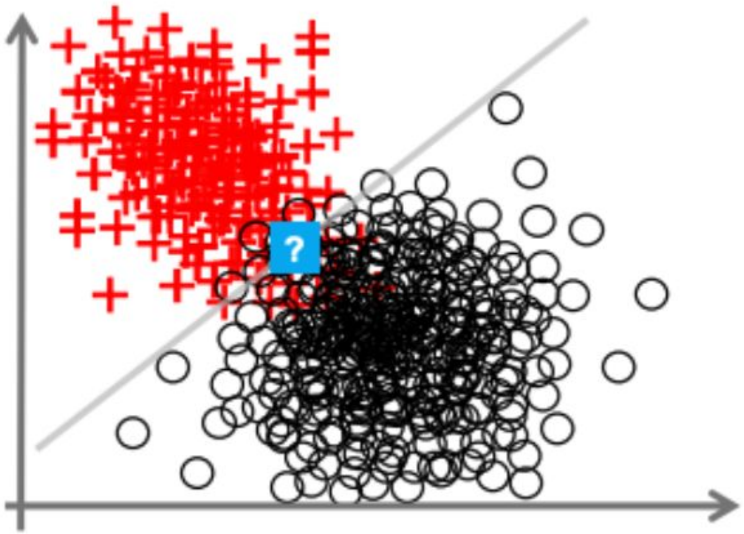
# Why distinguish between types of uncertainty?

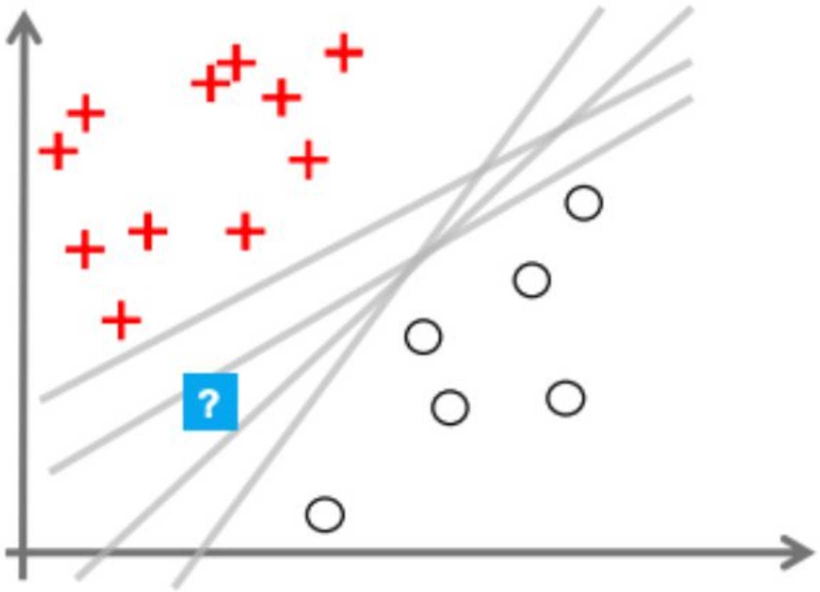
**Aleatoric Uncertainty**

- **Irreducible**: more examples do not help; more relevant features would be needed

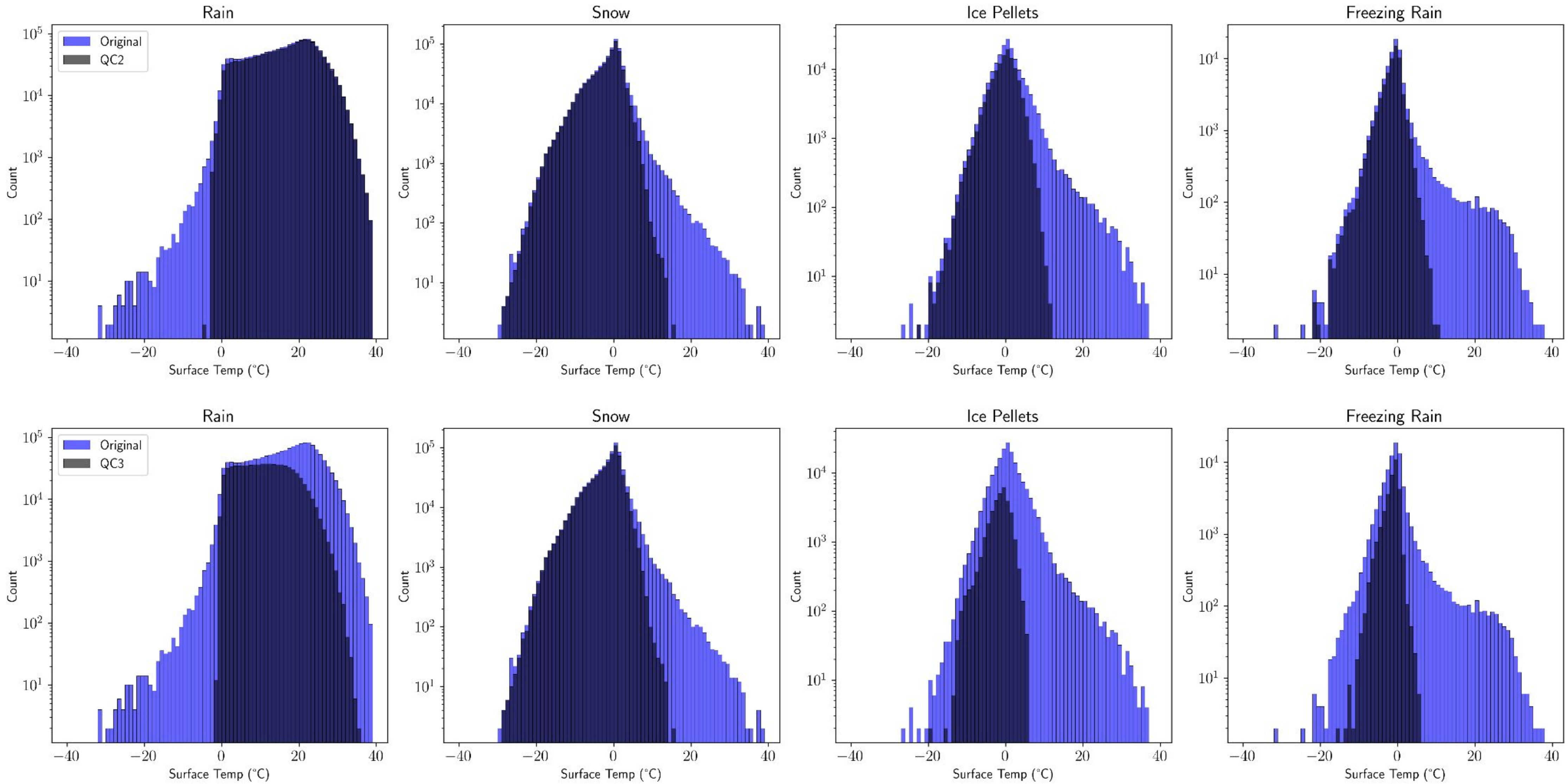
**Epistemic Uncertainty**

- ***Reducible***: more data examples can reduce this uncertainty

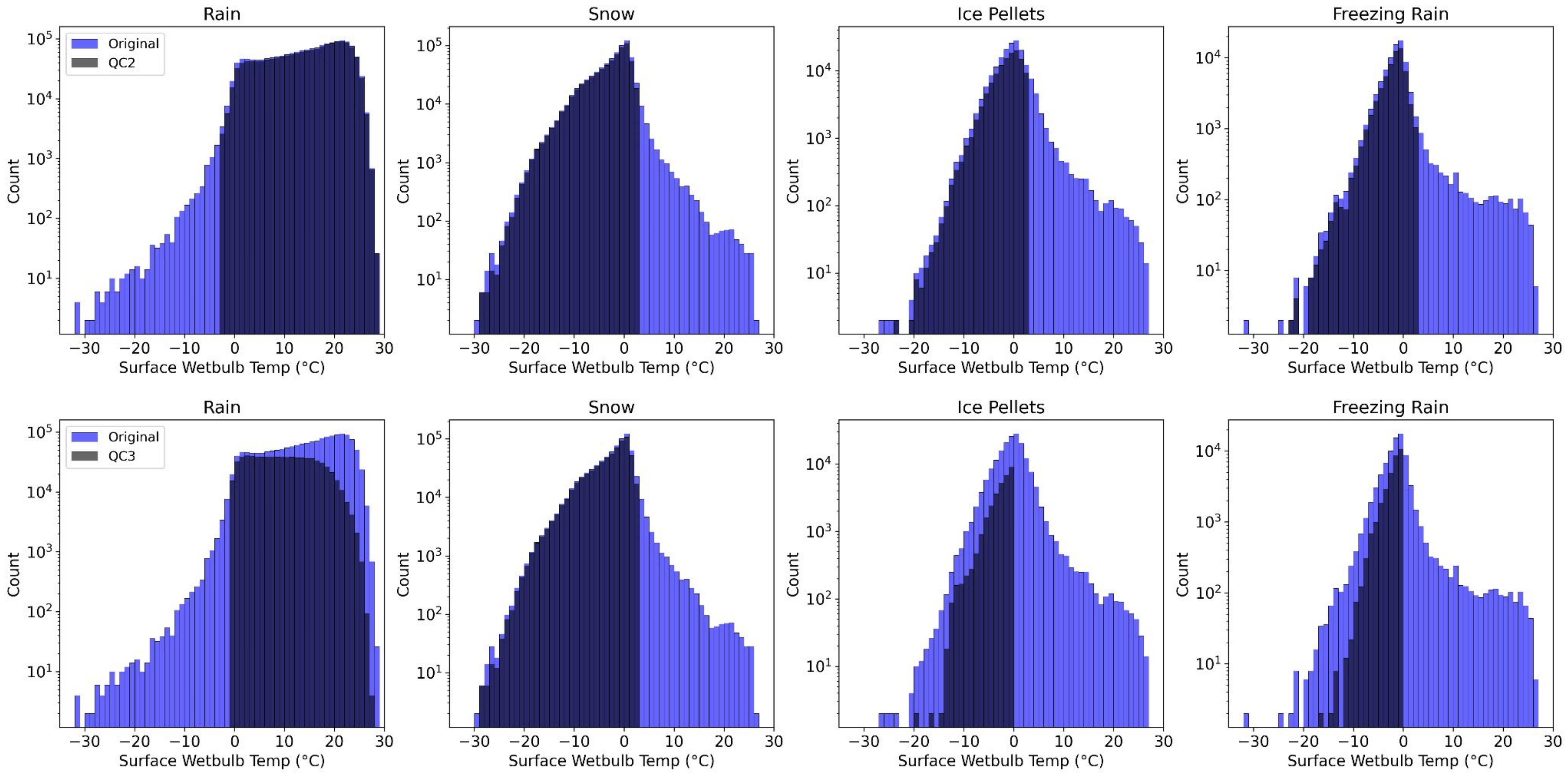


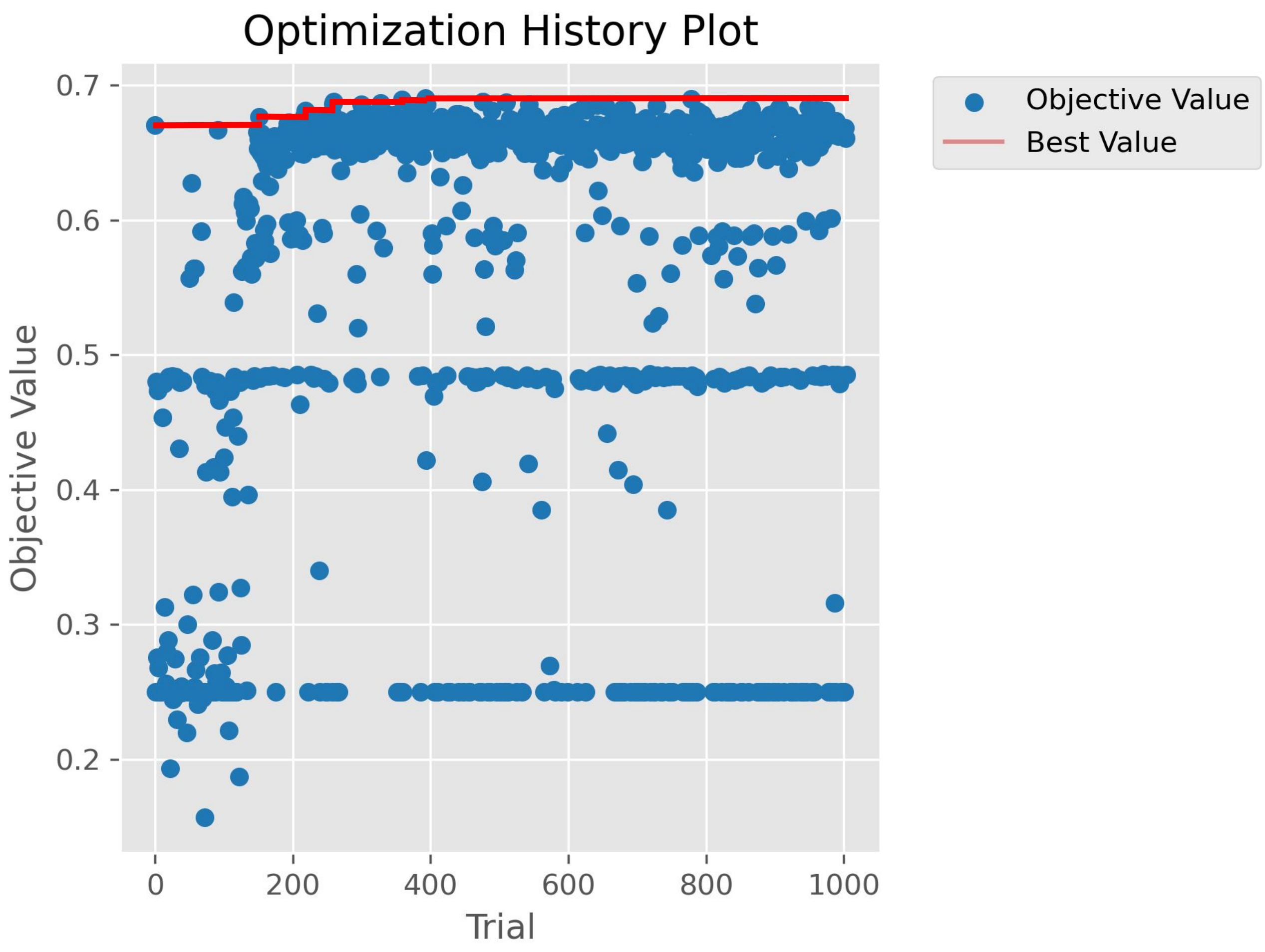
# Quality Control



# Quality Control



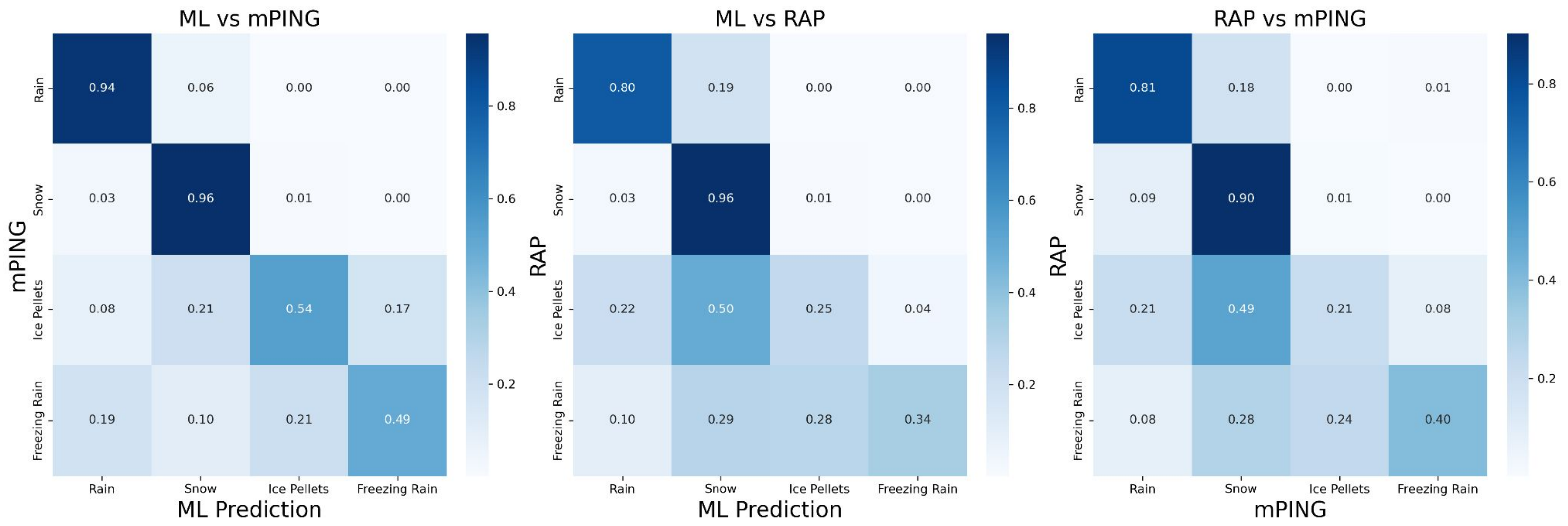
# Hyperparameter Optimization



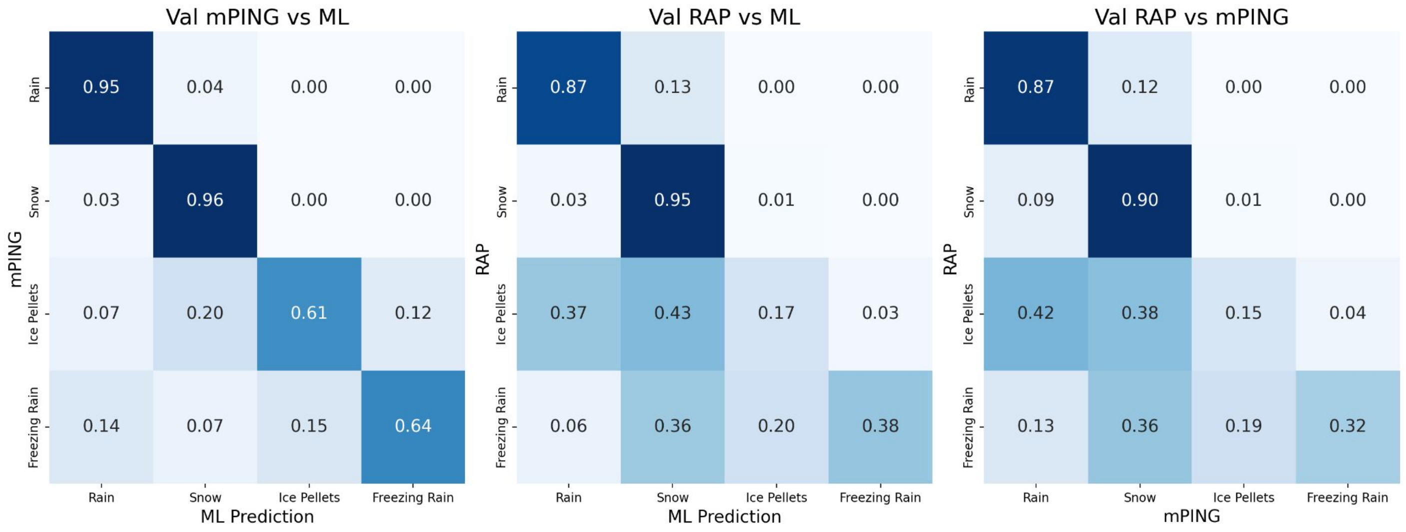

```
metric: val_ave_acc

model:
  activation: leaky_relu
  annealing_coeff: 34
  batch_size: 1130
  dropout_alpha: 0.11676011477923032
  epochs: 100
  evidential: true
  hidden_layers: 4
  hidden_neurons: 212
  l2_weight: 0.000881889591229087
  loss: evidential
  lr: 0.004800502096767794
  optimizer: adam
  output_activation: linear
  use_dropout: 1
```

# Results



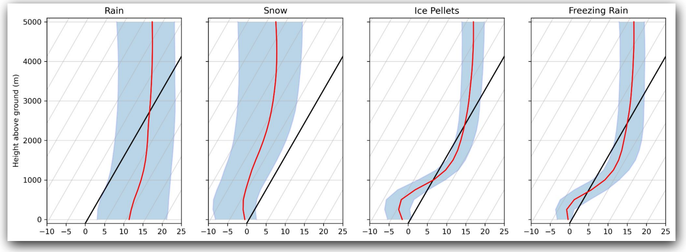

non echo true norm



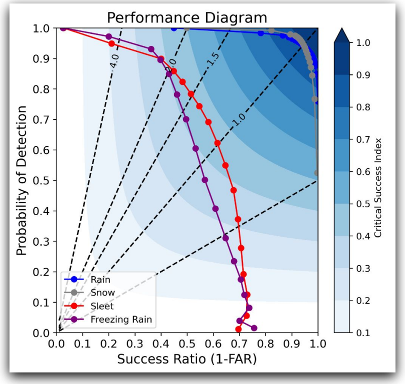

echo true norm

# Trustworthy AI & Forecaster Guidance

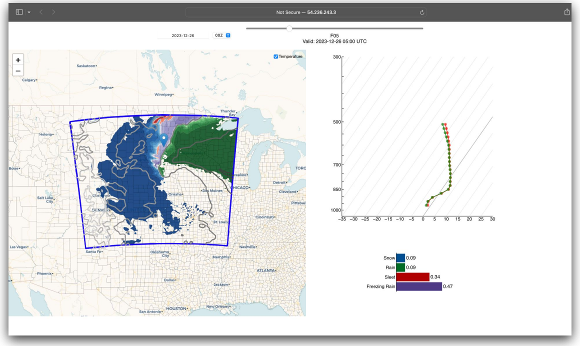
Ptype Model



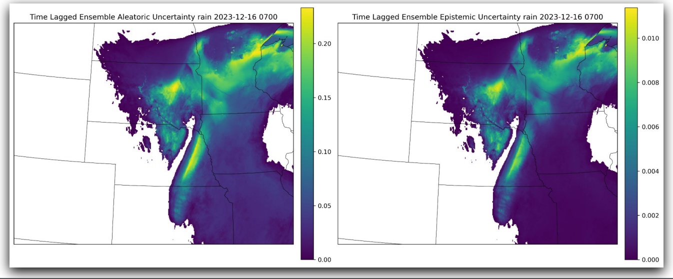

Performance



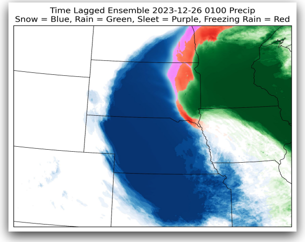

Web access



Comparison




Case Study




Considerations when selecting new guidance:

- How new guidance verifies compared to existing guidance
- Understanding failure modes
- Ability to examine guidance predictions for archived cases
- Understanding inputs

my focus

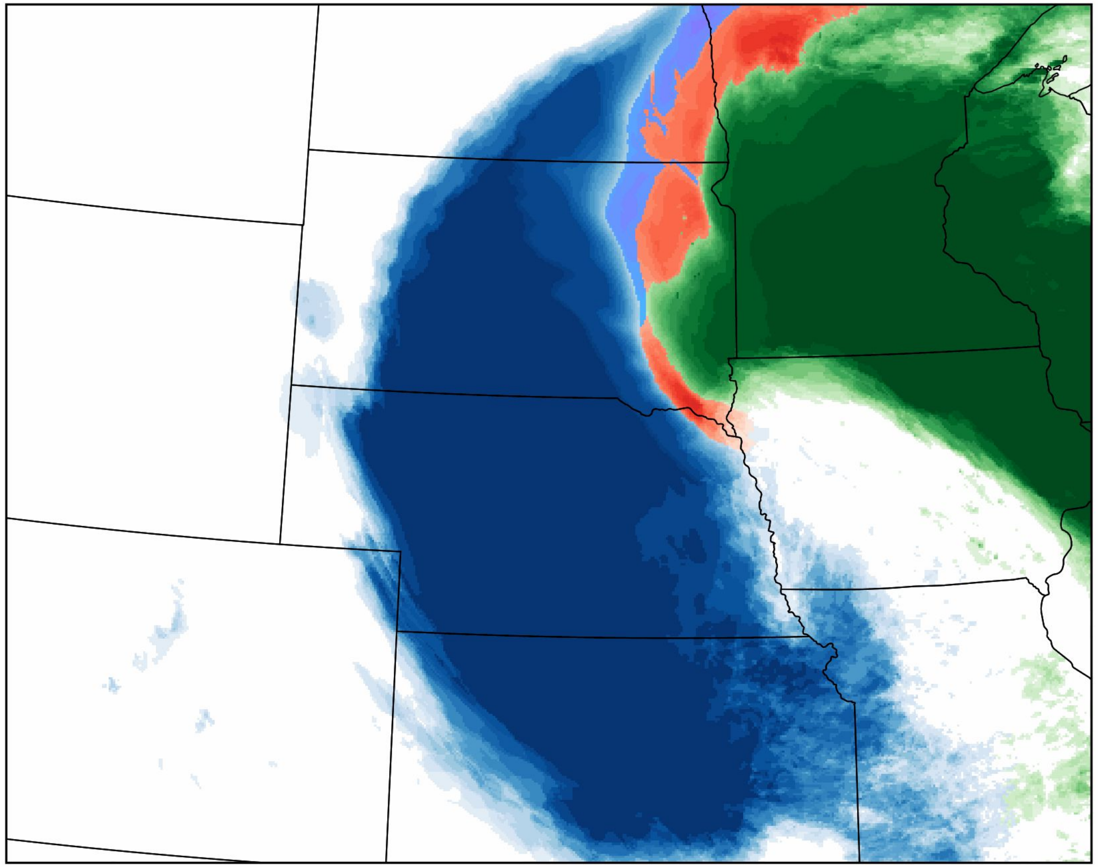
- Ability to sample the guidance via web-based tools
- Comparing guidance to observational data
- Understanding how output is derived

(Cains et. al 2024)

# Discussion

Analyzing changes in precipitation type across **time** and **space** is particularly important.

Time Lagged Ensemble 2023-12-26 0100 Precip
Snow = Blue, Rain = Green, Sleet = Purple, Freezing Rain = Red



Animations are a useful way to visualize data across both space and time