Machine Learning Uncertainty Quantification for Predicting Hazards and Impacts from Convection-Allowing LAMs

David John Gagne II
National Center for Atmospheric Research
Boulder, Colorado, USA

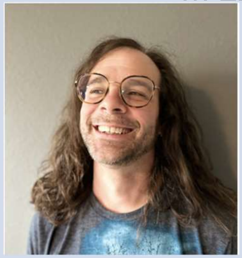
EWGLAM Workshop
25 September 2023

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

The NCAR Machine Integration and Learning for Earth Systems (MILES) Group

MILES Core

David John Gagne
ML Scientist II
CISL/RAL

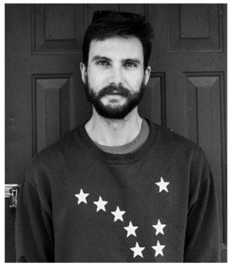


John Schreck
ML Scientist
CISL

Gabrielle Gantos
Associate Data Scientist II
CISL

Charlie Becker
Associate Data Scientist II
CISL

Kirsten Mayer
Project Scientist I
CGD



Will Chapman
Project Scientist I
CGD

Mariana Cains
Project Scientist I
MMM

Chris Wirz
Project Scientist I
MMM

Jacob Radford
Research Associate
CIRA

Maria Molina
Professor
U. Maryland

Thomas Martin
Software Engineer
Unidata

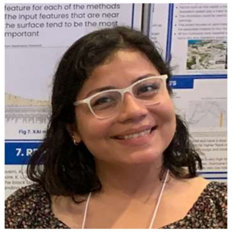
Wayne Chuang
Integration Engineer
LEAP

Julie Demuth
Project Scientist III
MMM

Jeff Anderson
Senior Scientist
CISL

Taysia Peterson
Admin
CISL

Dhamma Kimpara
Intern
CISL



Belen Saavedra Rios
Intern
CISL

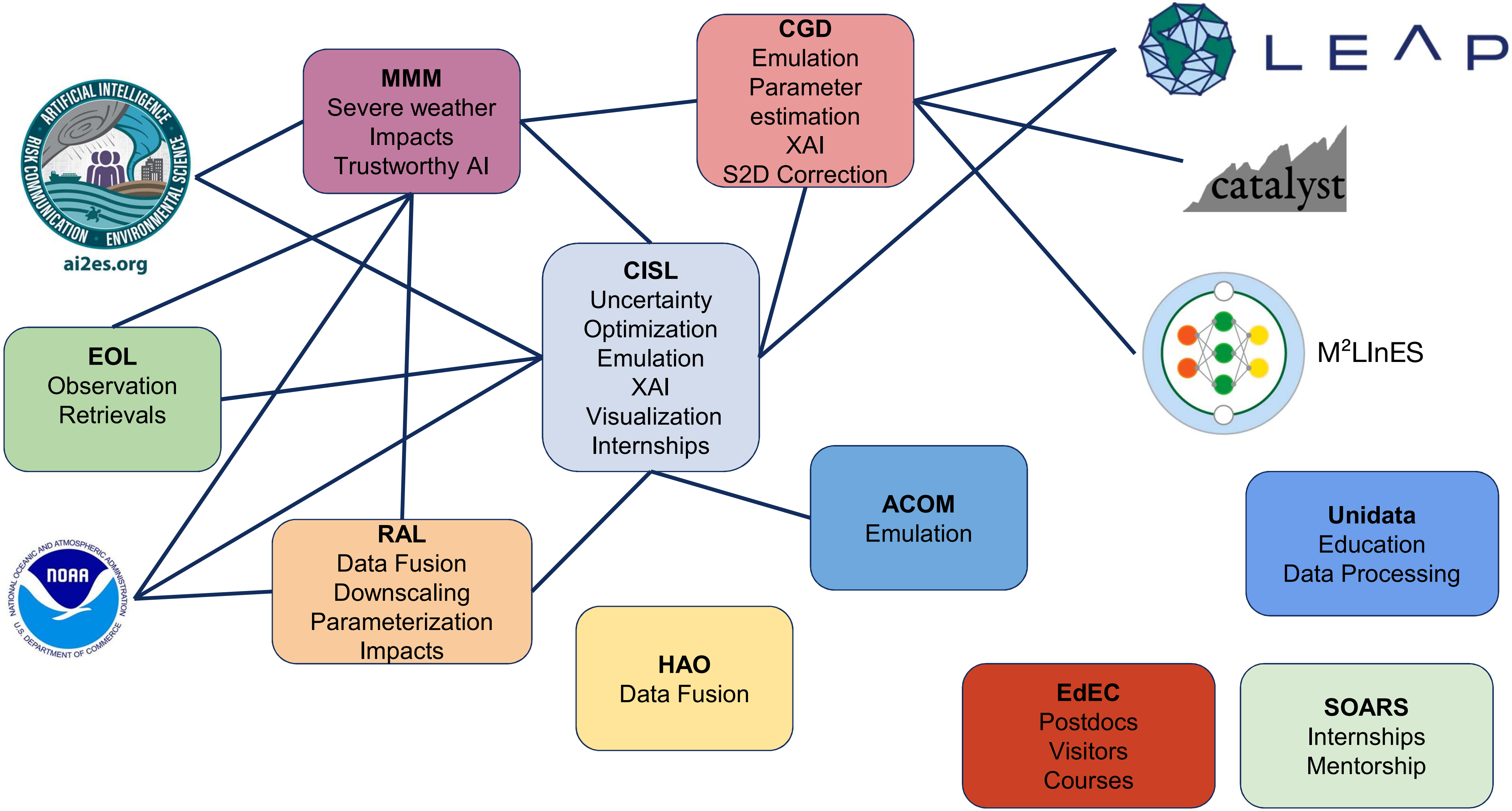
Hayden Outlaw
Intern
CISL

Da Fan
Visitor
Penn State

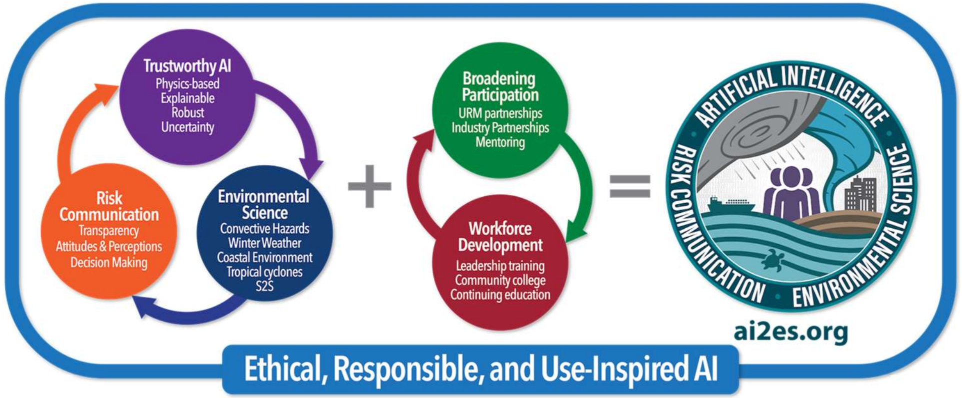
Bill Petzke
Software Engineer III,
RAL

MILES+

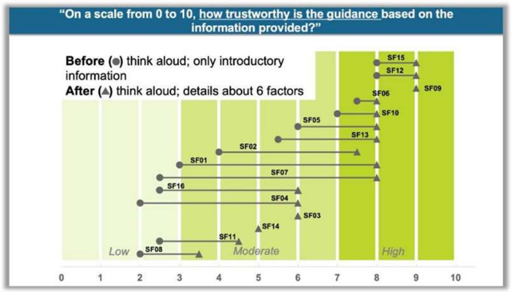
The NCAR/UCAR AI Web

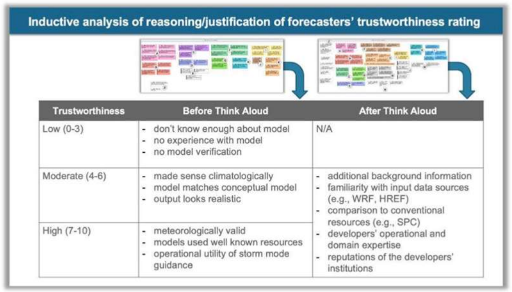


AI2ES: Developing Trustworthy AI Systems with User and Domain Expert Guidance



Assessing the Trustworthiness of AI/ML Forecast Guidance



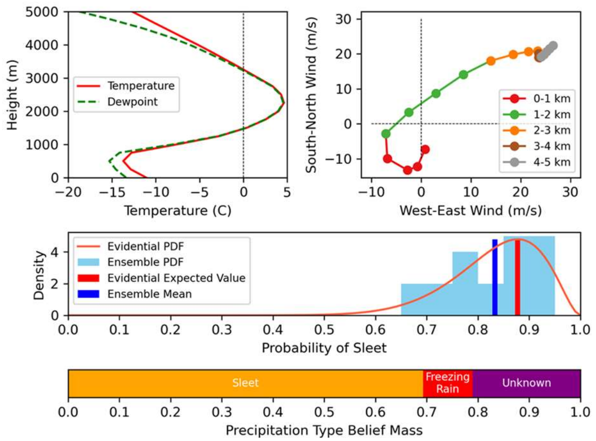


Inductive analysis of reasoning/justification of forecasters' trustworthiness rating

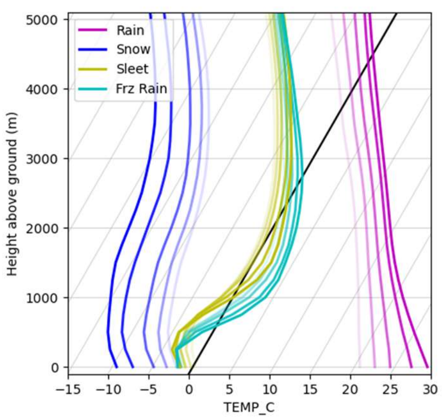
Trustworthiness	Before Think Aloud	After Think Aloud
Low (0-3)	- don't know enough about model - no experience with model - no model verification	N/A
Moderate (4-6)	- made sense climatologically - model matches conceptual model - output looks realistic	- additional background information - familiarity with input data sources (e.g., WRF, HREF) - comparison to conventional resources (e.g., SPC) - developers' operational and domain expertise - reputations of the developers' institutions
High (7-10)	- meteorologically valid - models used well known resources - operational utility of storm mode guidance	

Forecasters need to personally use a model or piece of guidance over time to build trust in it.

mgcains@ucar.edu



Median Soundings by Evidential Uncertainty

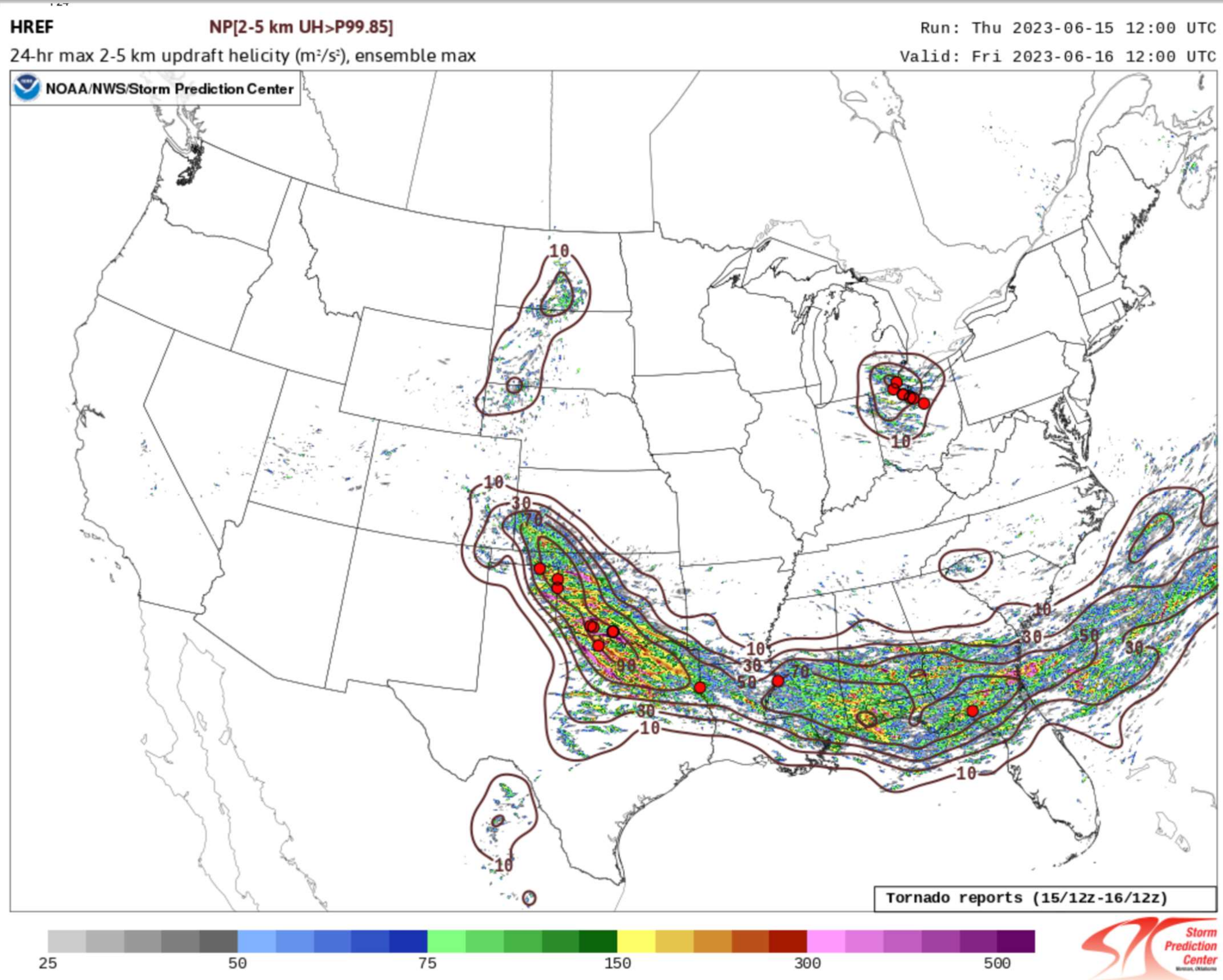


Darker is least uncertain: 10,20,40,60,80,90% least uncertain Predictions

Vision: AI2ES is developing *novel*, *physically based* AI techniques that are demonstrated to be *trustworthy*, and will directly improve *prediction, understanding, and communication* of high-impact weather and climate hazards.

CISL: David John Gagne, John Schreck, Charlie Becker, Gabrielle Gantos
MMM: Julie Demuth, Chris Wirz, Mariana Cains
RAL: Bill Petzke
Unidata: Thomas Martin

Motivation: The Limits of Convection-Allowing Models

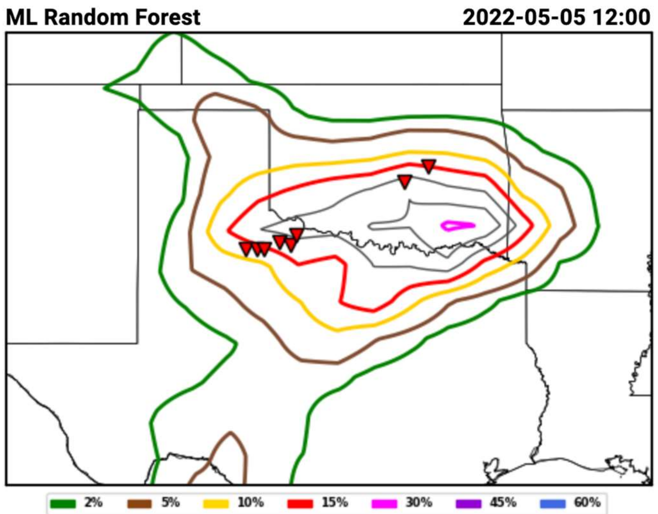


Convection-allowing NWP models (CAMs) can resolve individual storms, their modes, and their intensities.

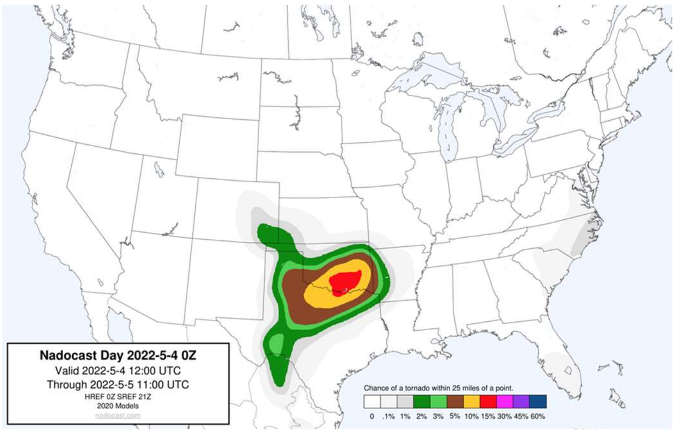
Storm hazards (tornadoes, hail, high winds, flash floods) depend on processes unresolvable for CAMS.

Motivation: Machine Learning Hazard Probabilities Derived from CAMs

Grid-Based HREF Random Forest (Loken et al.)



HREF+SREF Gradient Boosting (Hempel et al.)

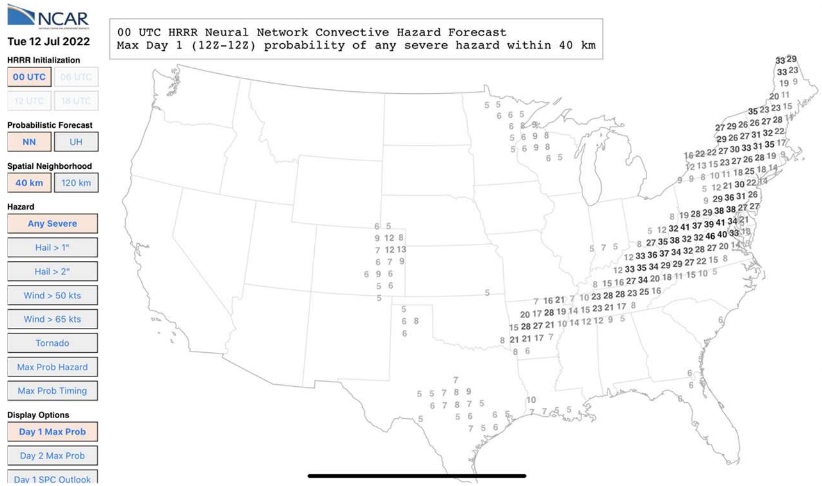


Given reports and lots of prior model output, machine learning models can estimate probabilities of severe hazards.

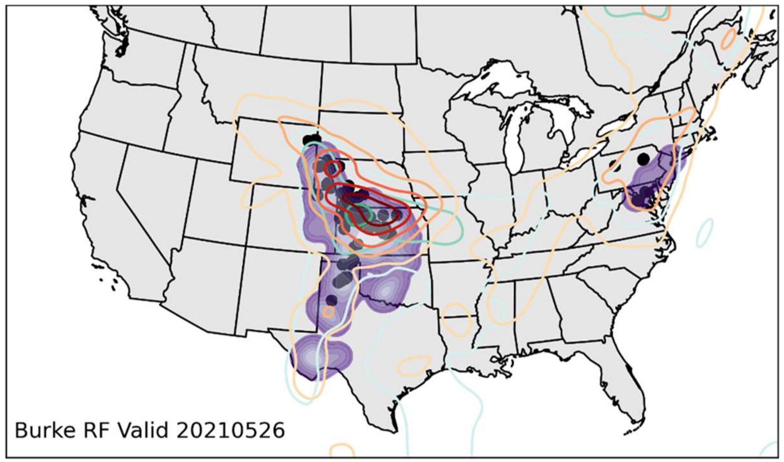
What else can machine learning infer from CAM output?

Storm Properties?
Timing?
Impacts?
Uncertainty?

Grid-based HRRR Neural Net (Sobash et al.)

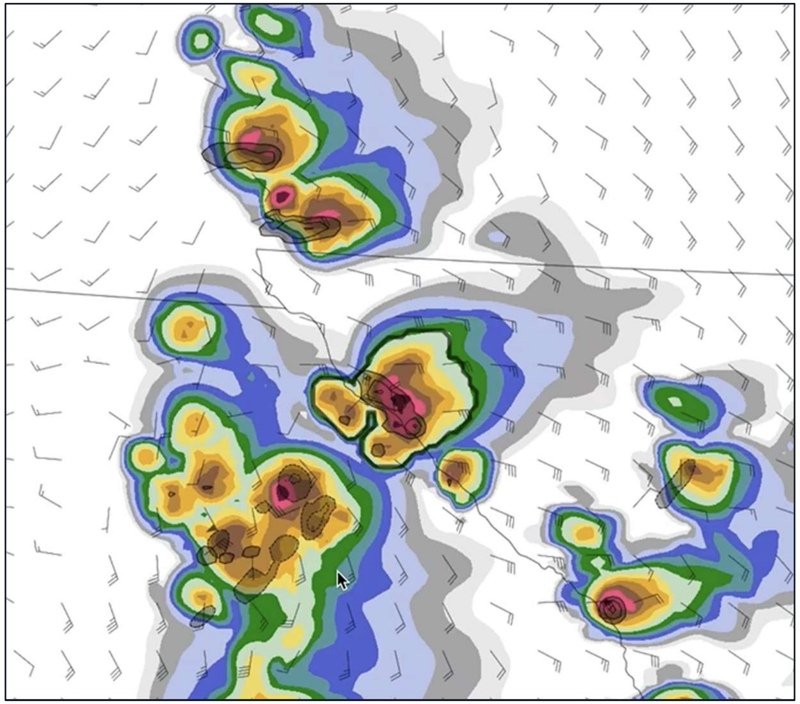


Storm-based HREF Random Forest (Burke/Gagne et al.)



Storm Mode: Motivation

Goal: Develop techniques to objectively identify convective mode in convection-allowing models (CAMs) using machine learning (ML) algorithms.



1. **Supervised learning** using hand-labeled CAM storm objects and ML algorithms

1. **Unsupervised/semi-supervised learning** using CAM storm objects, ML, and clustering algorithms

Diagnosing Storm Mode with Deep Learning in Convection-Allowing Models

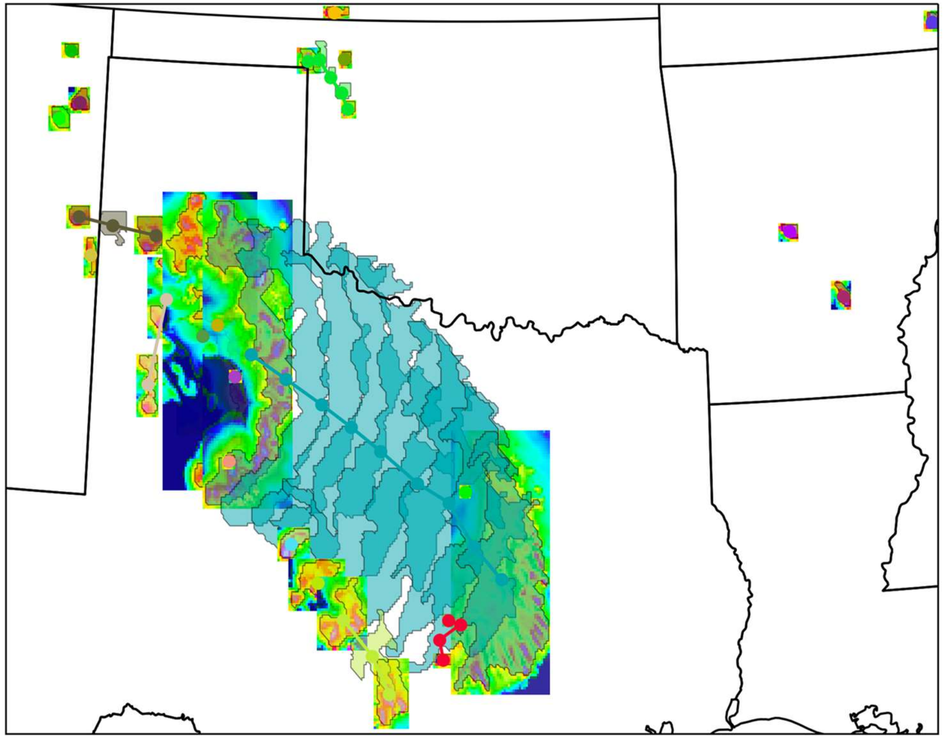
Ryan A. Sobash, David John Gagne II, Charlie L. Becker, David Ahijevych, Gabrielle N. Gantos, and Craig S. Schwartz

Online Publication: 05 May 2023

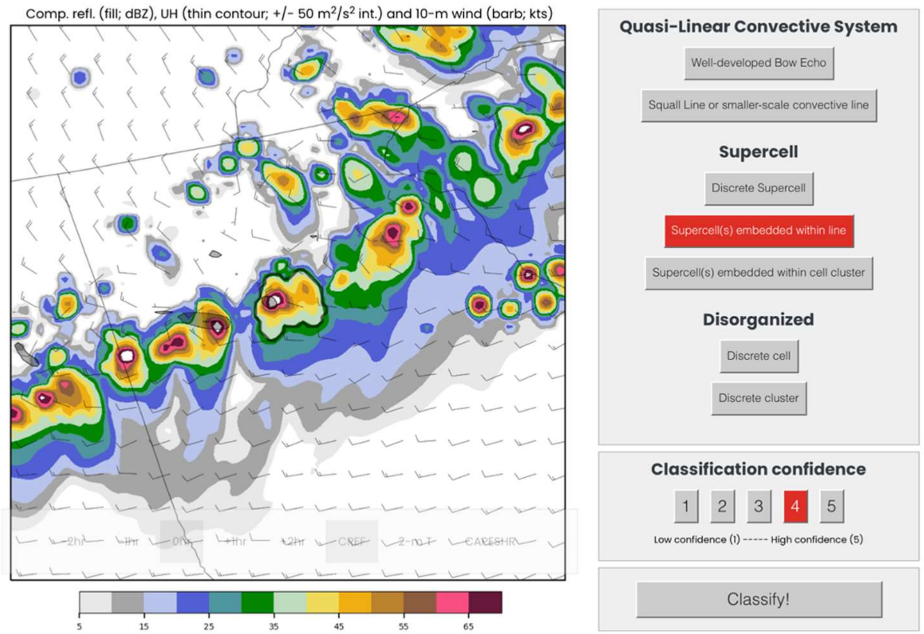
DOI: https://doi.org/10.1175/MWR-D-22-0342.1

Storm Mode: Segmentation, Tracking and Labeling

Hagelslag: segment storms and track with image processing techniques



Hand Labeling: web interface



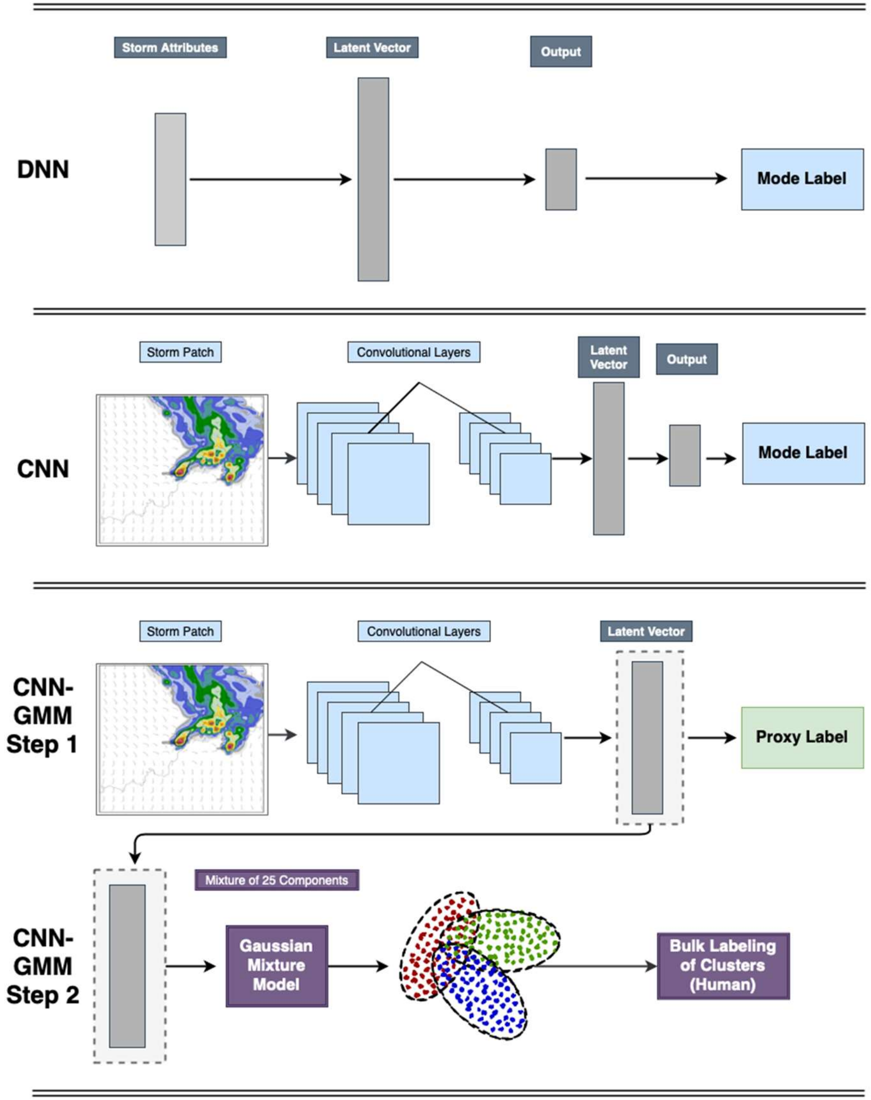
Challenge: segment storms with different spatial scales

Solution: define storms based on two radar reflectivity thresholds and merge adjacent objects

Challenge: disagreements about storm mode among labelers

Solutions: hierarchical classification, confidence ratings, multiple raters for each storm.

Model Architectures



Challenge: Storm does not fit entirely within image patch

Solutions: use summary metrics based on full storm extent, expand patch, use a grid-based segmentation

Challenge: Pre-processing pipeline changed during project

Solutions: relabel quickly using proxy labels and bulk labeling of storms based on clusters

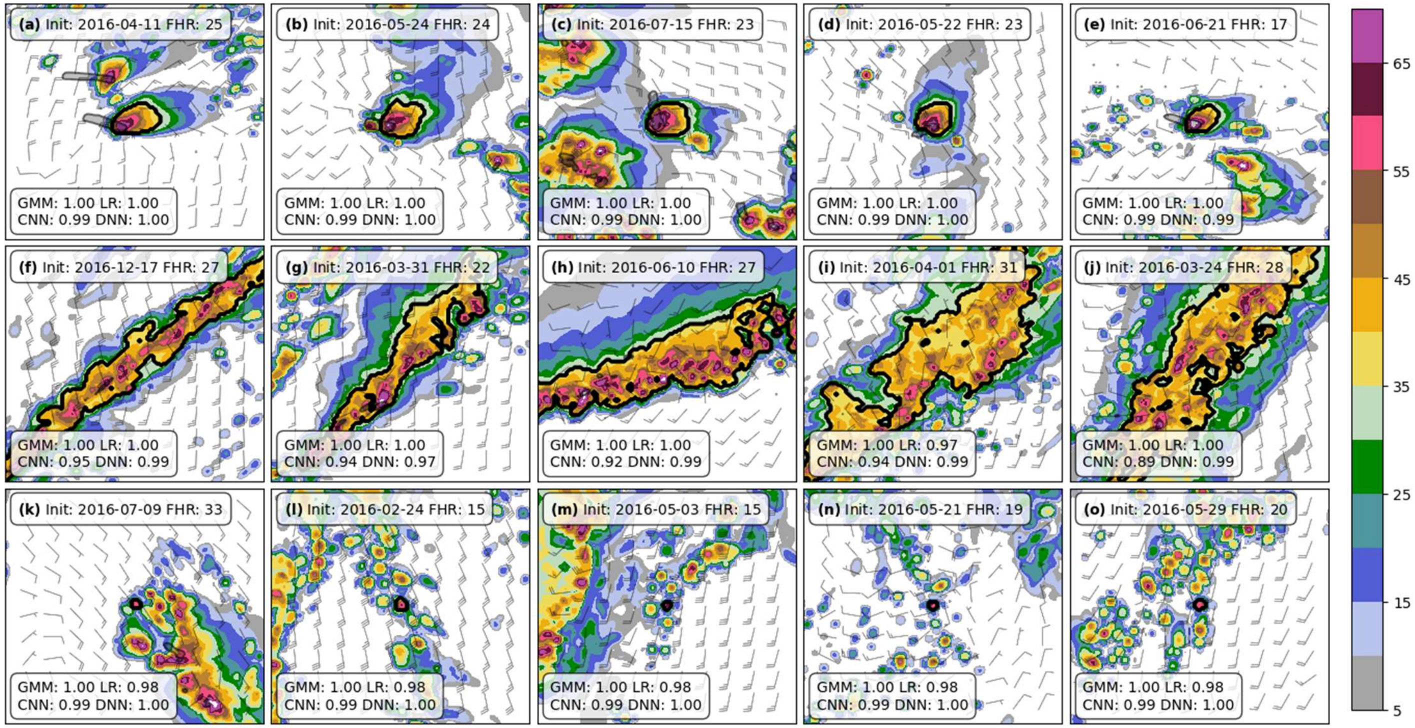
Model Performance

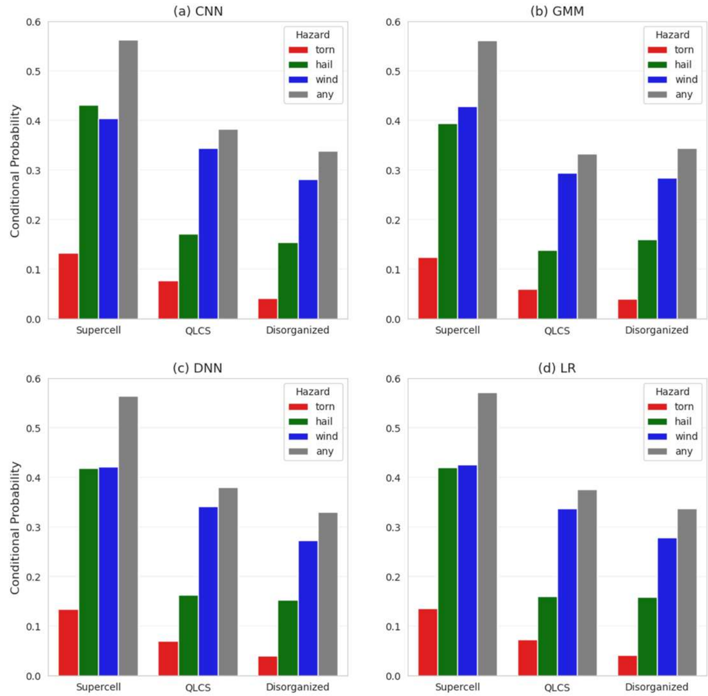
QLCS 114 storms	ROCA	BSS
CNN	0.879	0.286
GMM	0.895	0.005
DNN	0.918	0.412
LR	0.895	0.359
CNN-DNN avg	0.921	0.392

Supercell 113 storms	ROCA	BSS
CNN	0.891	0.427
GMM	0.900	0.302
DNN	0.901	0.448
LR	0.866	0.373
CNN-DNN avg	0.911	0.480

Disorganized 529 storms	ROCA	BSS
CNN	0.855	0.335
GMM	0.869	0.186
DNN	0.891	0.452
LR	0.869	0.409
CNN-DNN avg	0.890	0.439

Storms with Predicted Mode Agreement

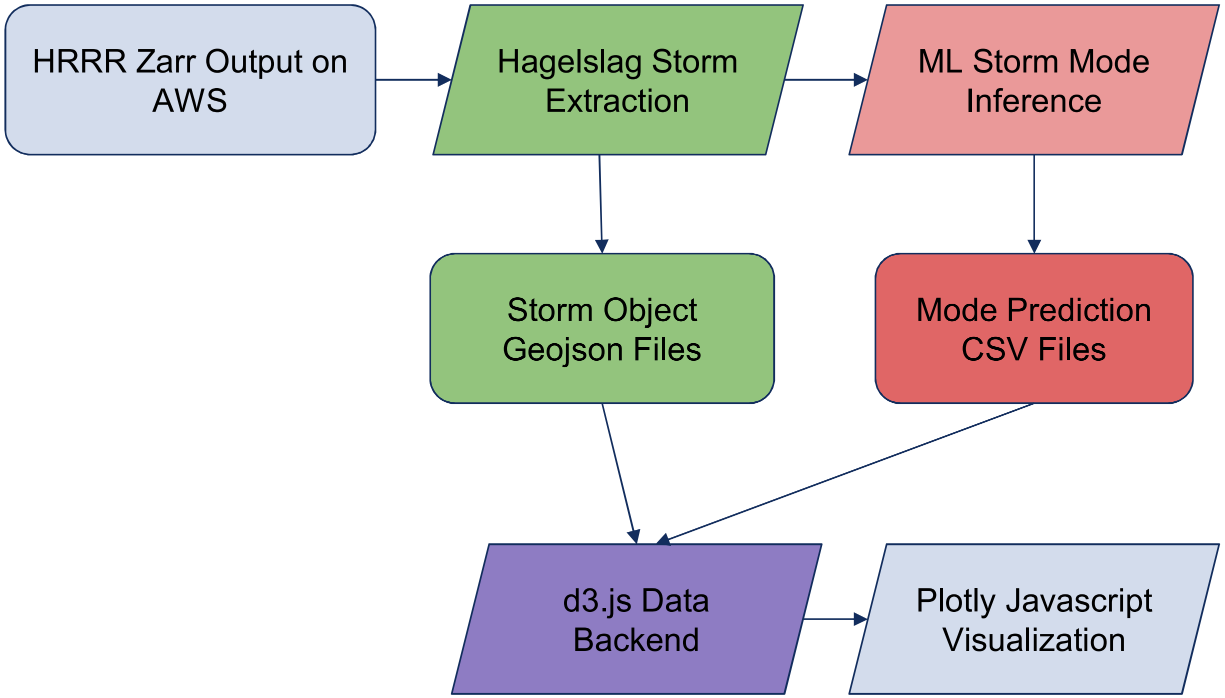


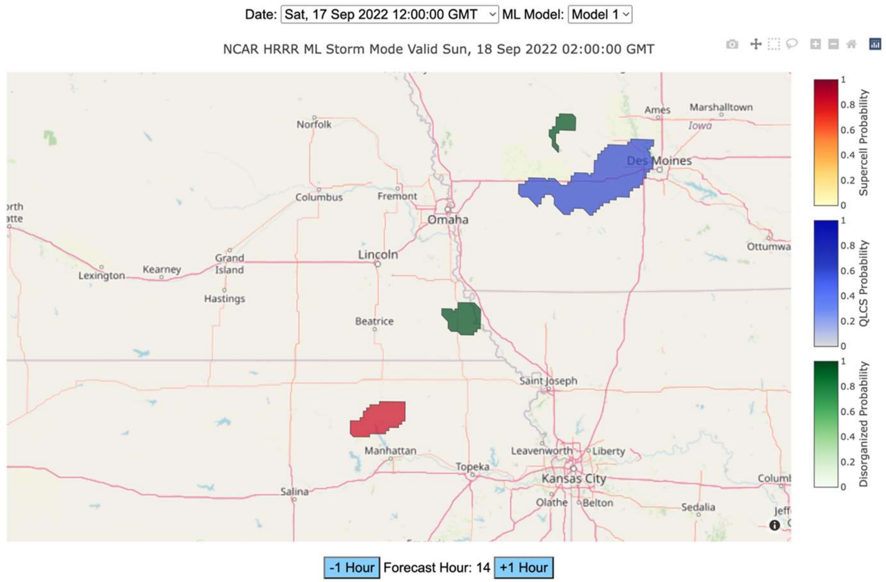


Challenge: evaluate storm mode models without a massive hand labeling effort

Solution: examine consistency among storm labels and conditional probability of different severe hazards given mode

Storm Mode Visualization Pipeline



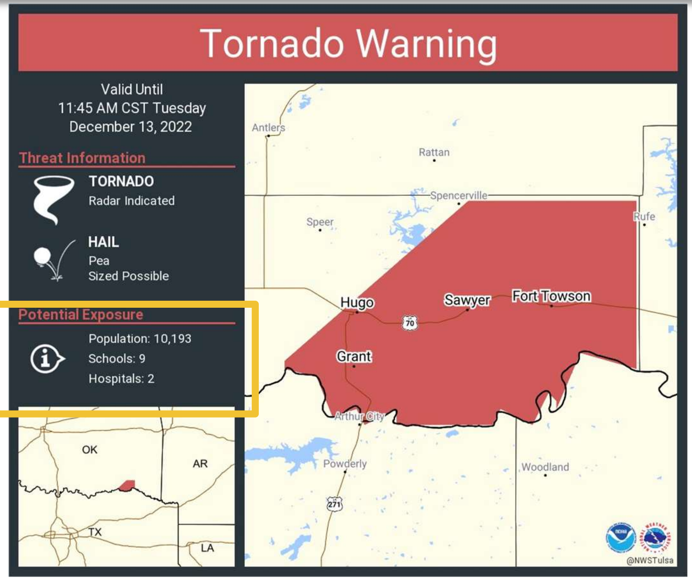


ncar.github.io/modeview

CRISIS

- Tornado risks and impacts are influenced by intersecting meteorological and societal factors
- The goal of integrated tornado risk modeling (e.g., Strader et al. 2016) is to understand and reduce impacts on people
- Modeling of population mobility in tornado risk models is limited
  - Many studies assume a static population (e.g., Hatzis et al. 2020)
  - Some recent work accounts for day-night population differences (Strader et al. 2022)
  - Limited understanding of uncertainties in population estimates and how those intersect with dynamic meteorological data
- How does one integrate spatio-temporally varying meteorological and societal data into a convergent risk model?
- What can we learn from this process about integrated tornado risks and associated uncertainties?

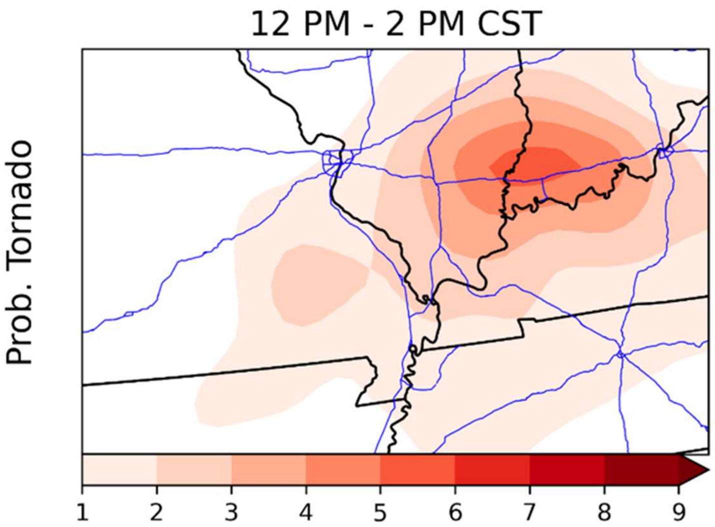
Leads: David John Gagne (CISL), Olga Wilhelmi (RAL), Rebecca Morss (MMM)



What is a more robust estimate of the exposure to a tornado? What is the uncertainty?

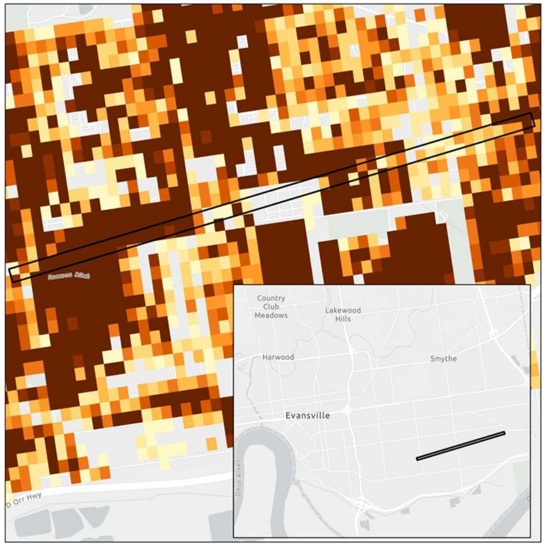
CRISIS: Integration

Model Hazard



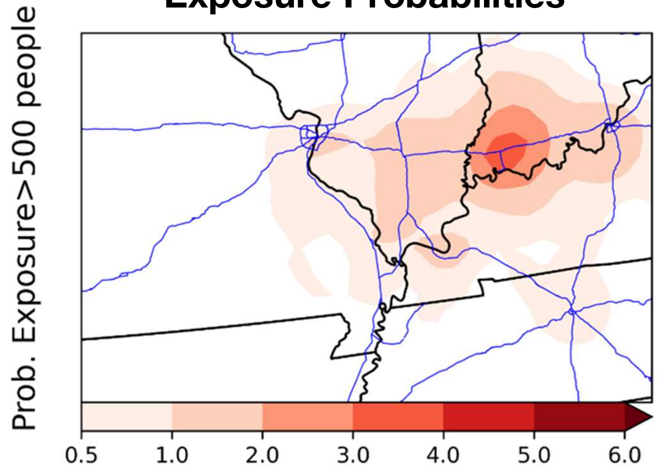
Generate probabilities of tornado, probability distributions of how many tornadoes are expected to occur, and 10,000 synthetic tracks.

Intersect with Population



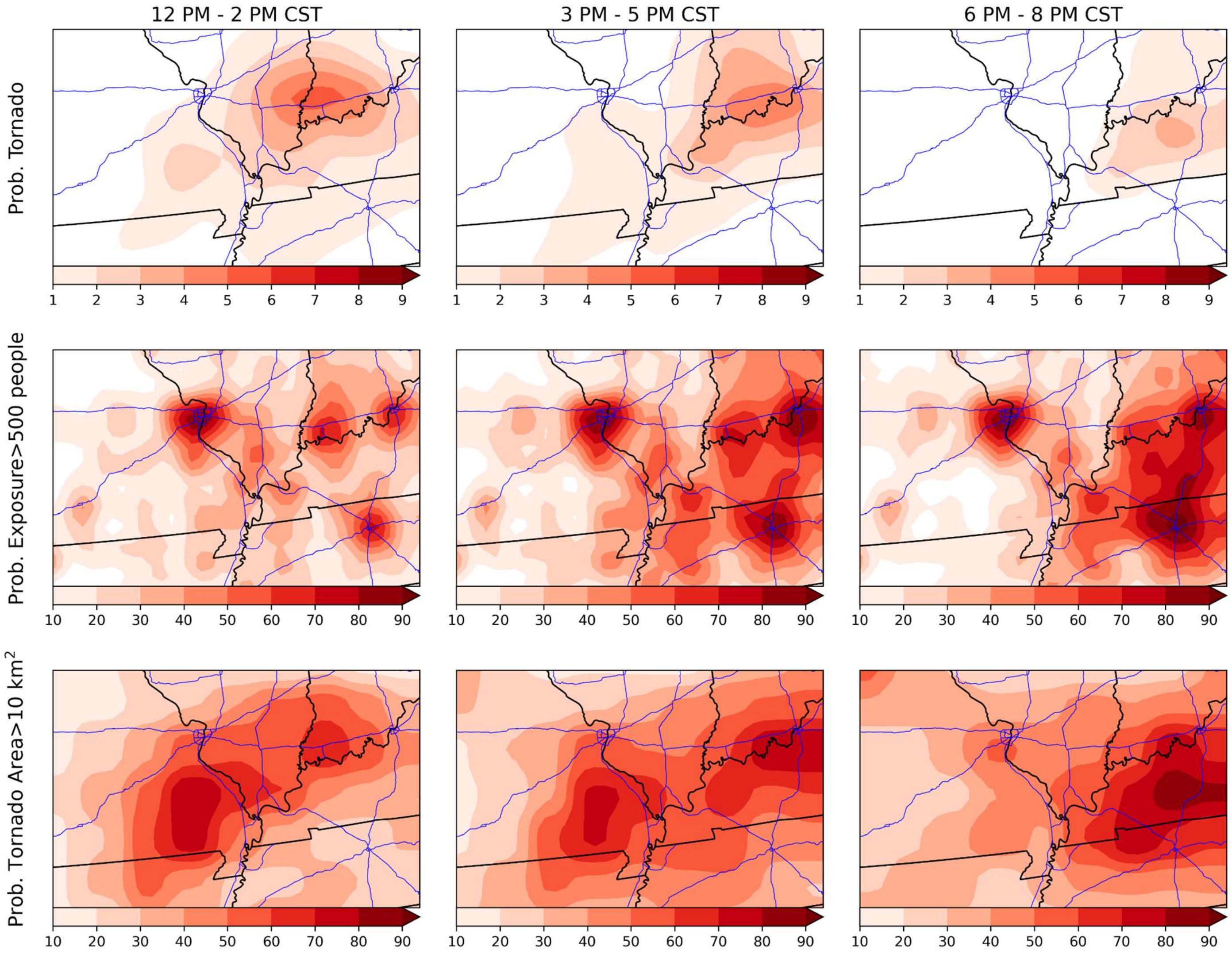
Intersect tornado tracks with a time-interpolated 100 m grid of population density data and sum the number of people exposed within each track.

Combine Tornado and Exposure Probabilities

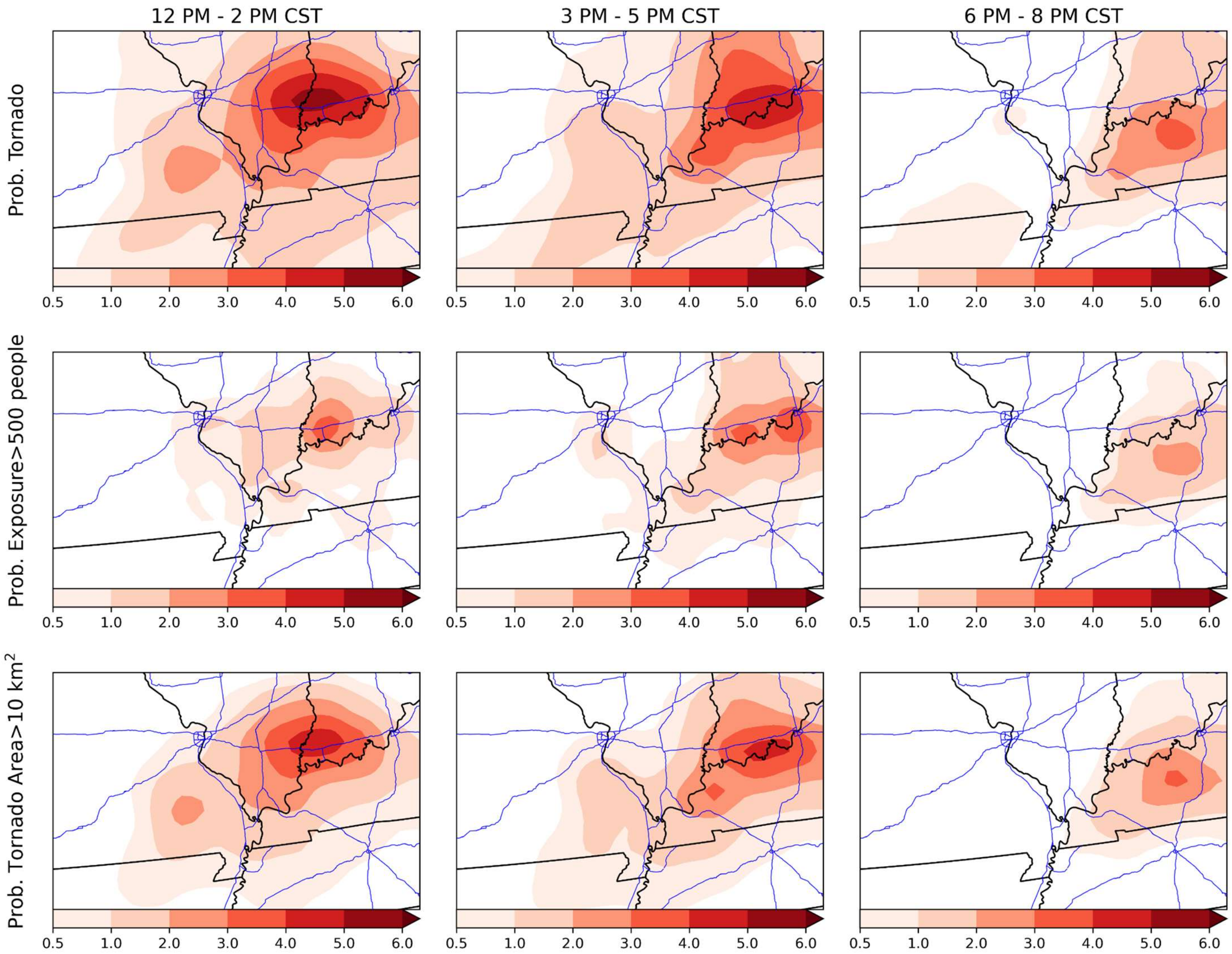


Determine probability of exposure from number of tracks impacting population and aggregate with tornado and count probabilities to produce unconditional exposure estimate.

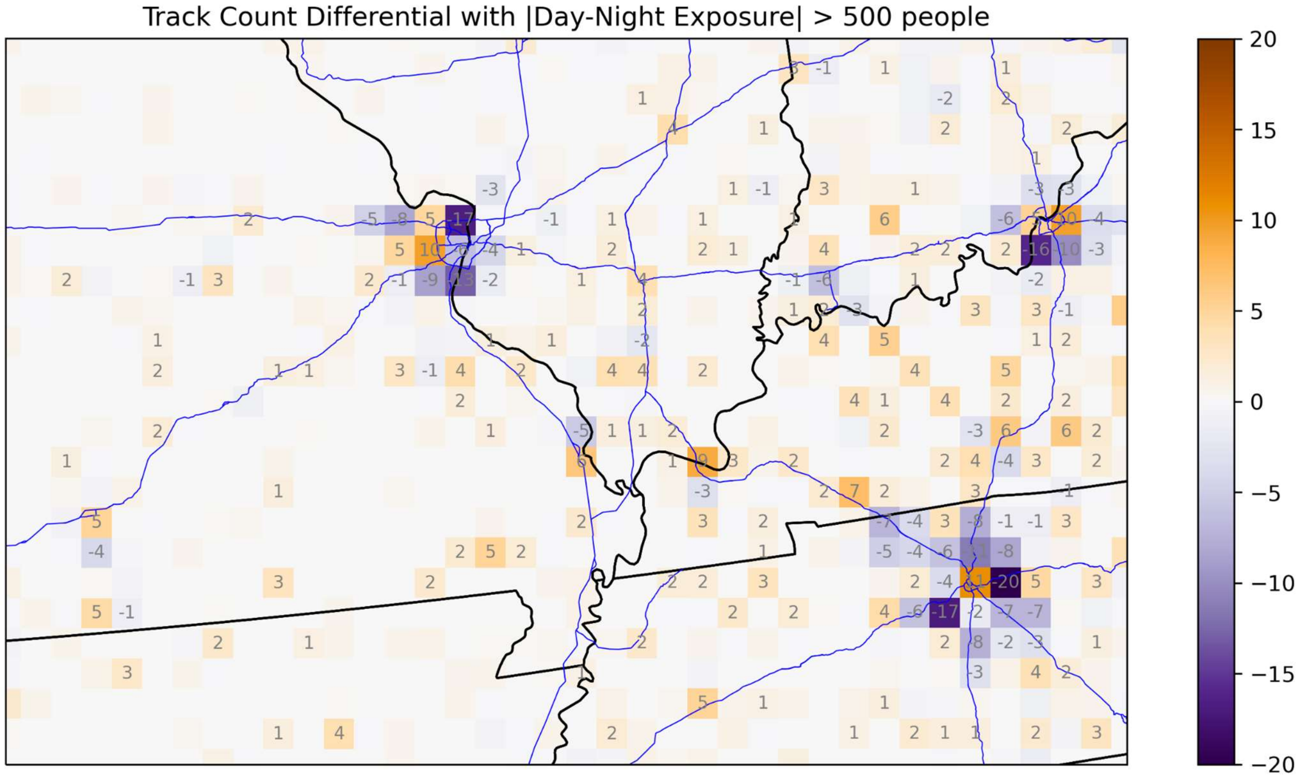
CRISIS: Conditional Probabilities

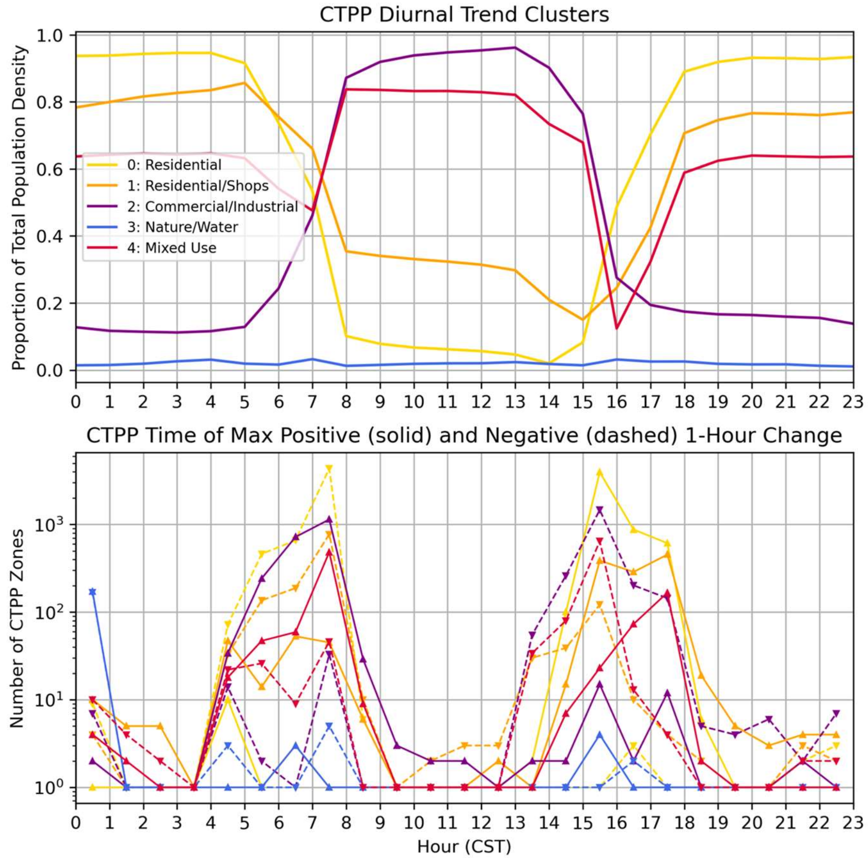


CRISIS: Integrated Probabilities

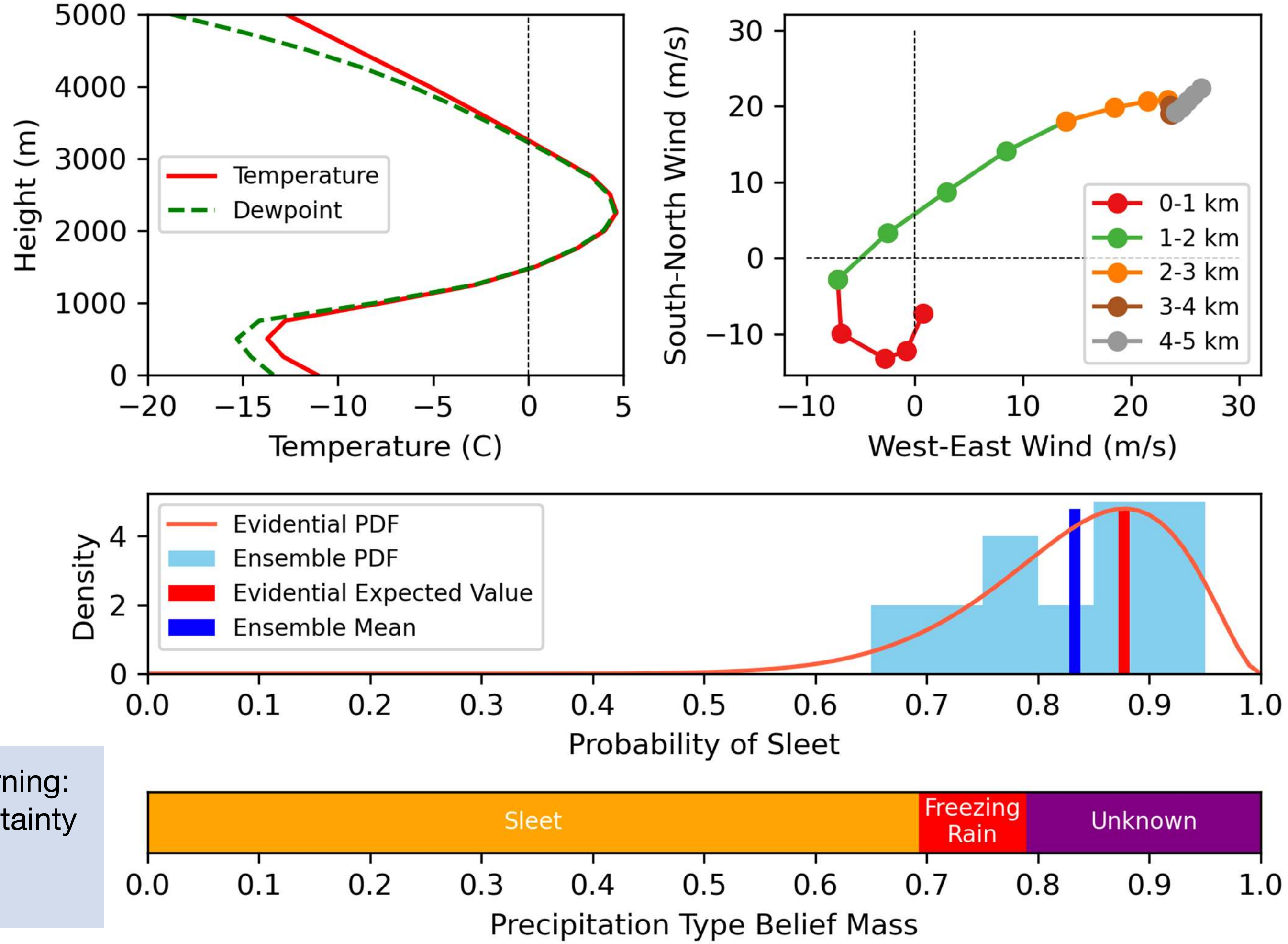


CRISIS: Population Trends





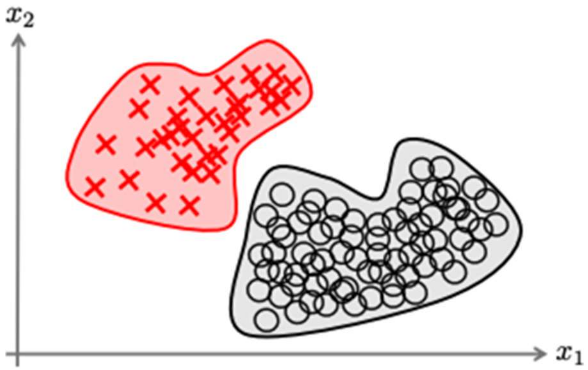
Evidential Deep Learning



Paper: Evidential Deep Learning: Enhancing Predictive Uncertainty Estimation for Earth System Science Applications

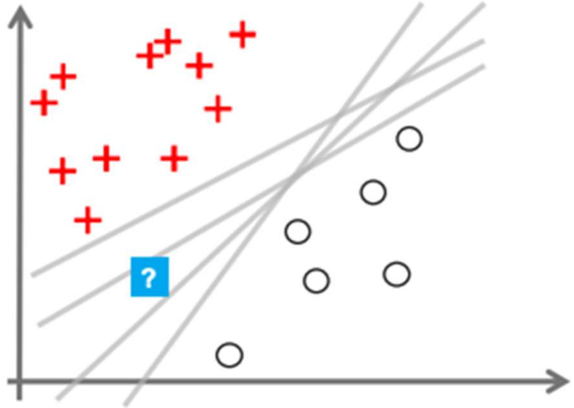
Decomposition of Uncertainty

Aleatoric Uncertainty



Definition: Uncertainty from variation in data.
Estimated by: Single probabilistic AI model.
Reduce by: Gather more informative inputs

Epistemic Uncertainty



Definition: Uncertainty from variation in models.
Estimated by: Ensemble of deterministic AI models.
Reduce by: Gather more examples or use simpler models.

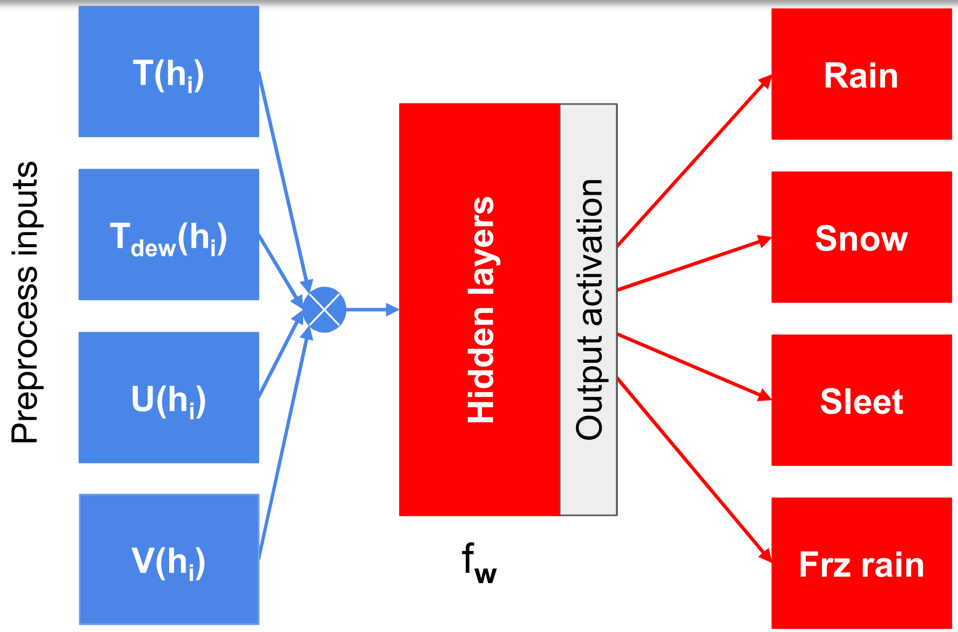
Total Uncertainty

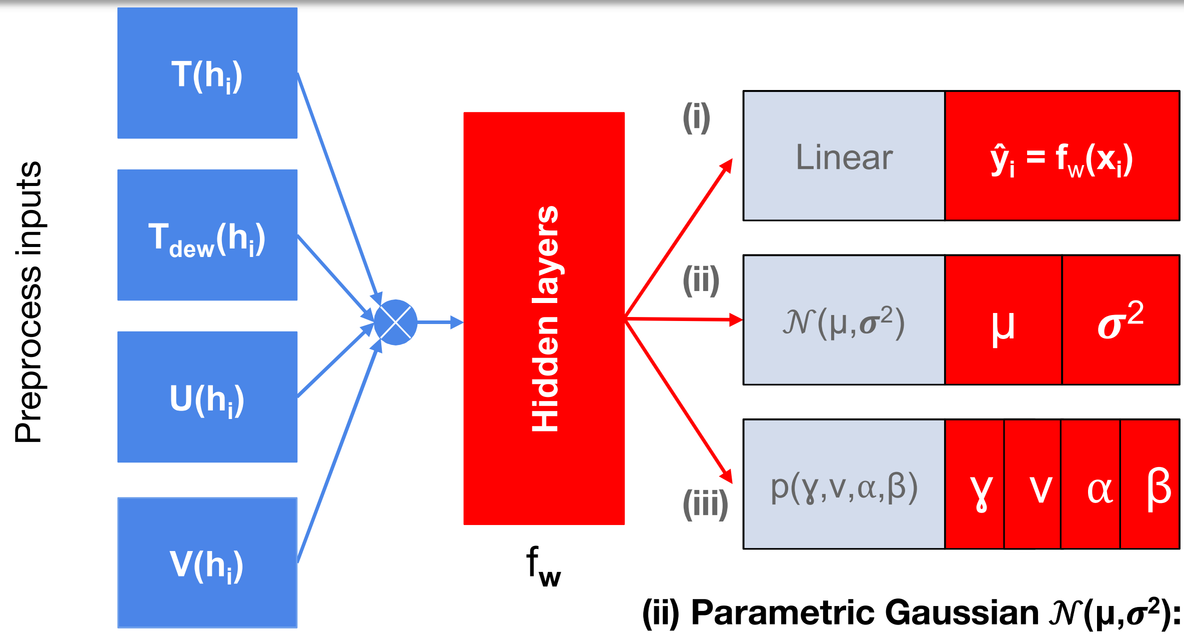
Definition: Combined aleatoric and epistemic uncertainty.
Estimated by:

1) Ensemble of probabilistic AI models
2) Single "evidential" (higher-order probabilistic) AI model
3) Bayesian AI models

Collaborators
John Schreck, Charlie Becker, Gabrielle Gantos, Julie Demuth, Chris Wirz, Jacob Radford, Nick Bassil, Kara Sulia, Chris Thorncroft, Amy McGovern, Eliot Kim, Justin Willson, Kim Elmore, Maria Molina

Evidential Model Architectures





(i) Deterministic:

Predict probabilities for classes

Loss = Cross-entropy

p_k = Softmax($f_w(T,T_{dew},U,V))_k$

(ii) Evidential:

(Sensoy et al. 2018)
Predict evidence for classes

Loss = Evidential

e_k = ReLU($f_w(T,T_{dew},U,V))_k$

$\boldsymbol{\alpha}_k = e_k + 1$

Compute S, evidential u, and the probabilities p_k

(i) Deterministic:

Predict values for the defined tasks

Loss = RMSE/MAE/etc

Number of outputs = 1

(ii) Parametric Gaussian $\mathcal{N}(\mu,\sigma^2)$:

Predict the mean and variance for each task

Loss = NLL

Number of outputs = 2

(iii) Parametric Normal-Inverse Gamma p(ɣ,v,α,β):
(Amini et al. 2020)
Predict evidence for parameters for each task
Loss = Evidential, Number of outputs = 4
Post-prediction: Compute mean, aleatoric, and epistemic uncertainties

Probabilistic Forecast Example: Classifying Winter Precipitation Type

Data

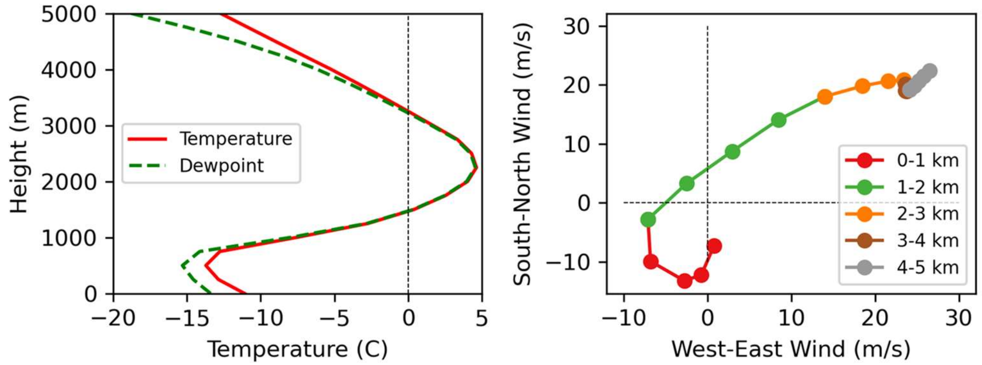
- ➢ **NOAA Rapid Refresh** Vertical Profiles
- ➢ Interpolate from pressure to height coords

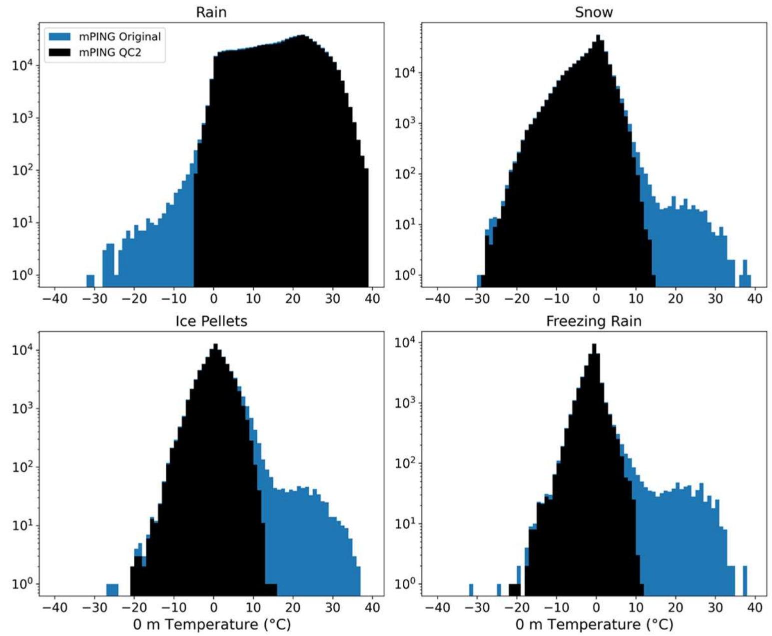
Input (0 - 5 km above surface, every 250 meters)

- ➢ Temperature, Dewpoint, U-Wind, V-Wind

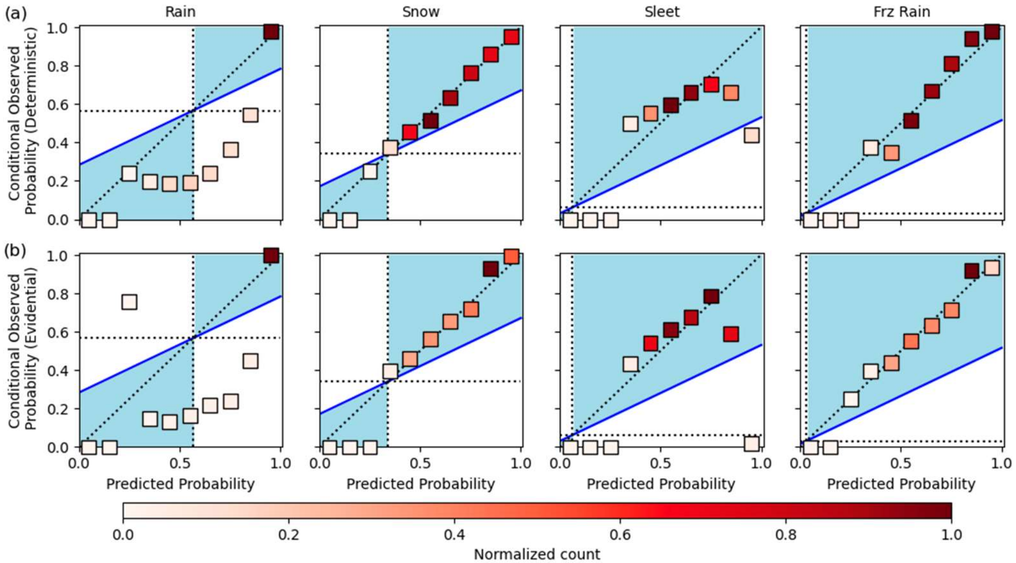
Target

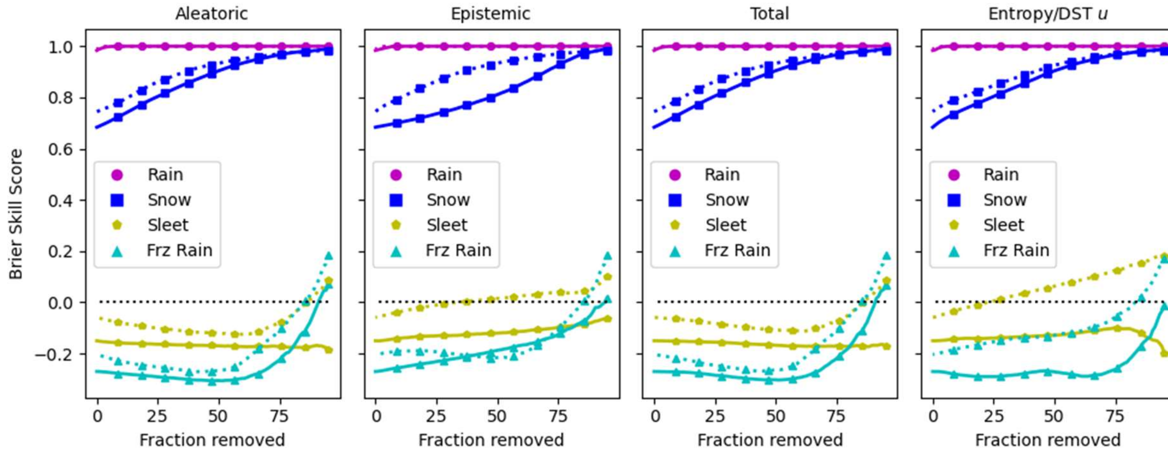
- ➢ mPING Crowd-sourced reports of winter precipitation types
  - ➢ *Rain, Snow, Sleet, Freezing Rain*





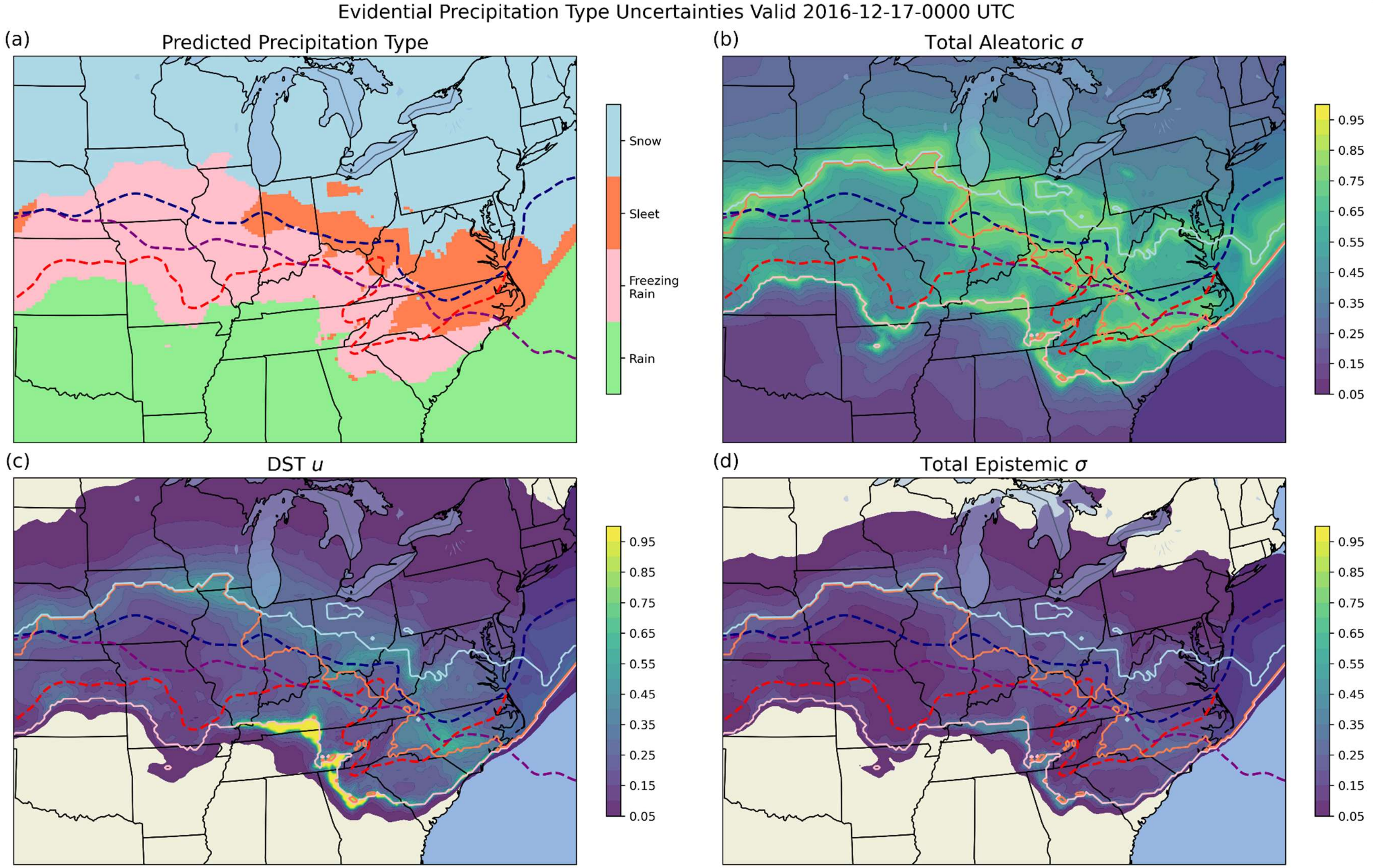
Precipitation-type Validation



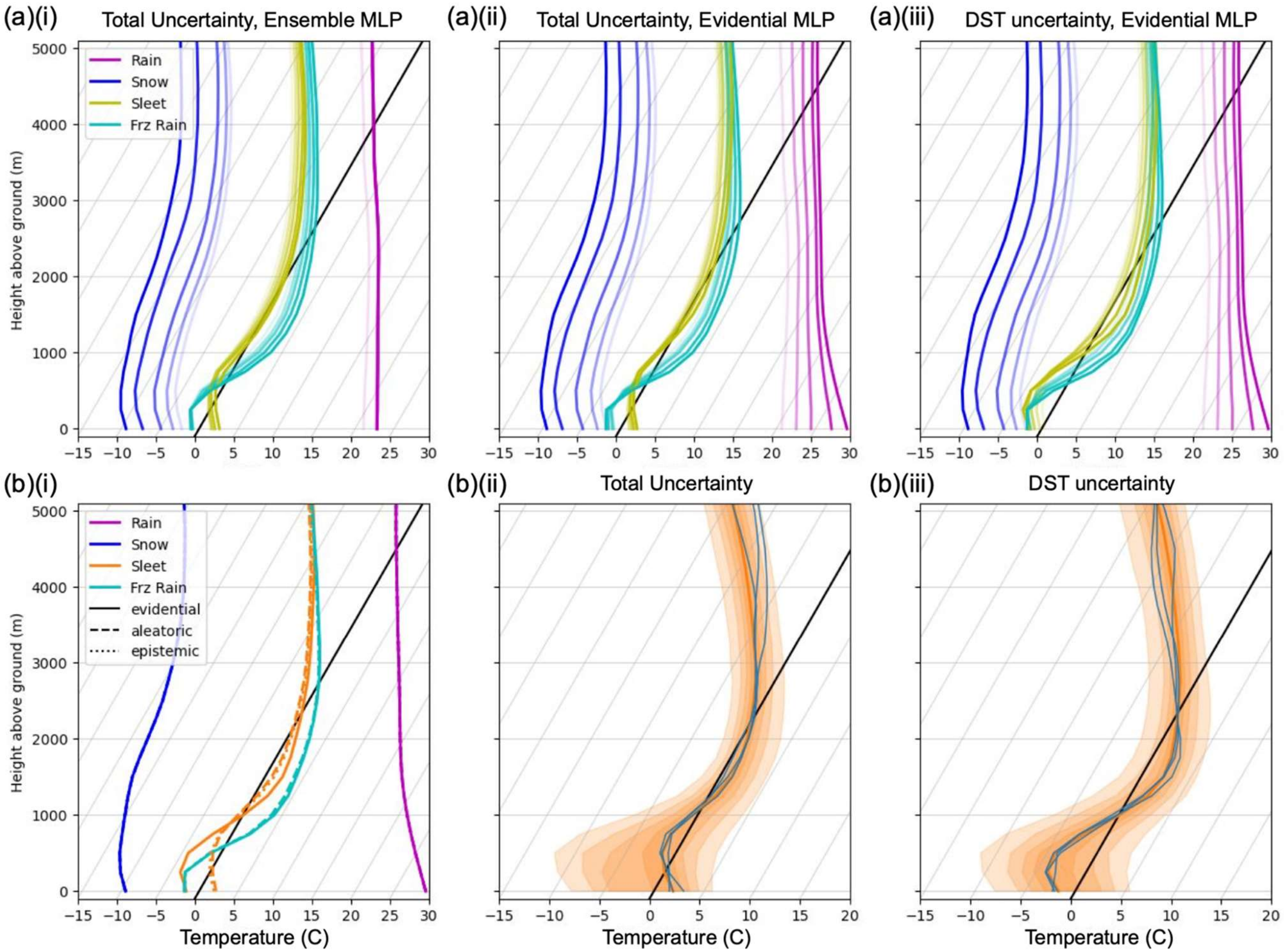


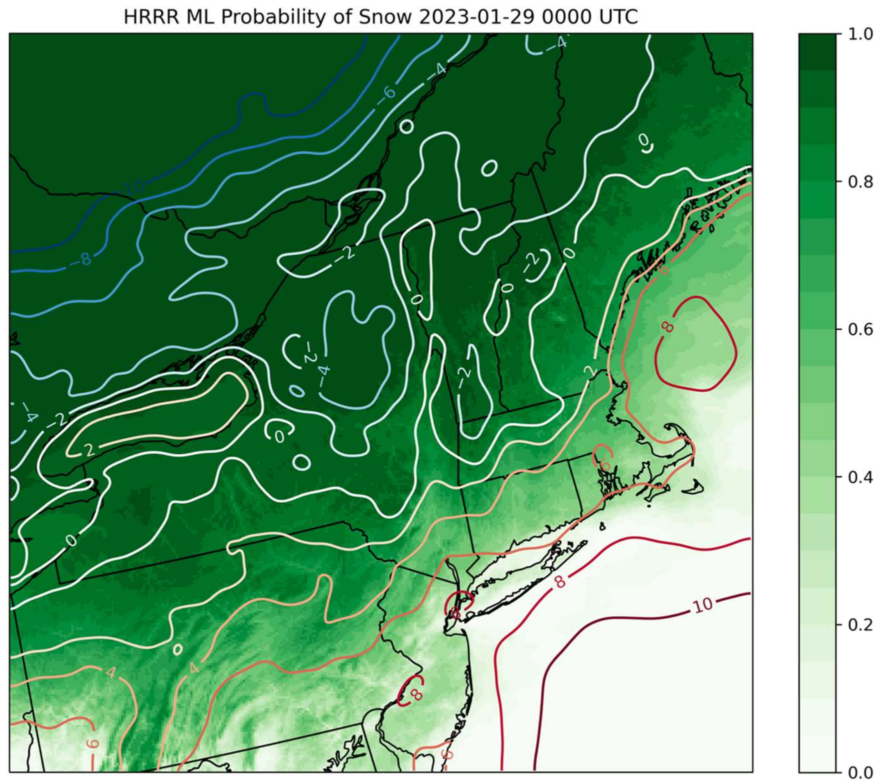
How well does each type of uncertainty discriminate between easier and harder to classify events?

Regional Case Study



Uncertainty Drivers

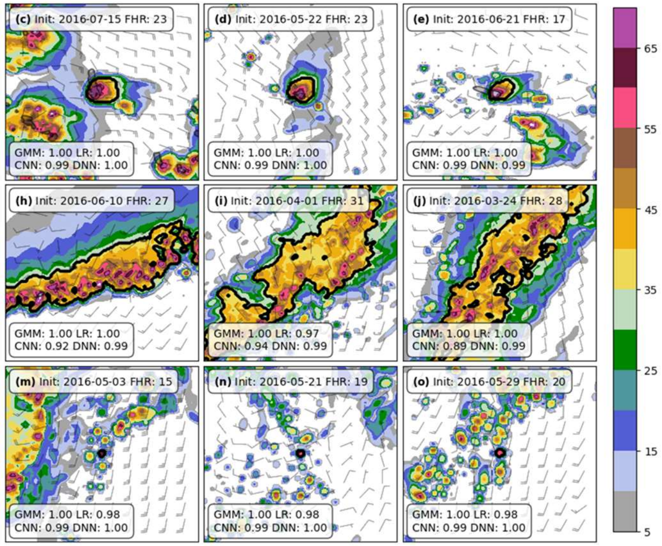


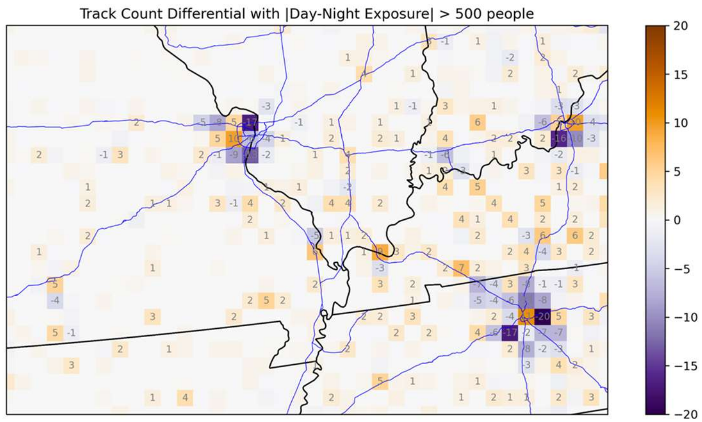


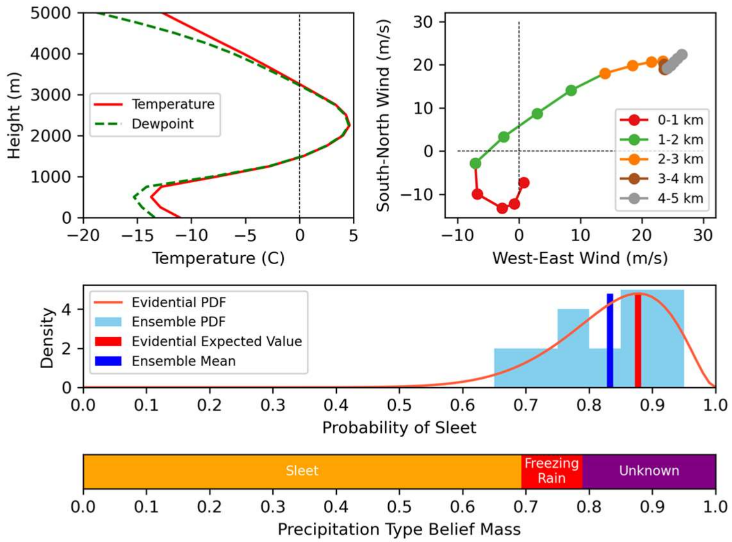
MILES Group Python Packages

- **miles-guess** (github.com/ai2es/miles-guess):
  - Implementations of evidential neural networks, deep ensembles, and Monte Carlo dropout
- **echo-opt** (github.com/NCAR/echo-opt):
  - Distributed hyperparameter optimization on HPC systems
  - Supports GPU allocation, XAI visualization for hyperparameter settings
- **hagelslag** (github.com/djgagne/hagelslag):
  - Object segmentation, tracking, and data extraction for convection-allowing model output
  - verification scores and plots
- **bridgescaler** (github.com/NCAR/bridgescaler):
  - Reproducible saving/loading of sklearn preprocessing scalers and transforms
  - Custom scalers for groups of variables and image patches

Summary







Diagnosing convective mode helps forecasters adjust their mental models of timing and hazards.

Intersecting hazard and population data reveals particularly sensitive areas.

Evidential deep learning decomposes uncertainty to provide more physical insight and guide model updates.

Email: dgagne@ucar.edu