Building Trustworthy AI for Environmental Science

Amy McGovern

Lloyd G. and Joyce Austin Presidential Professor, School of Computer Science

Professor, School of Meteorology

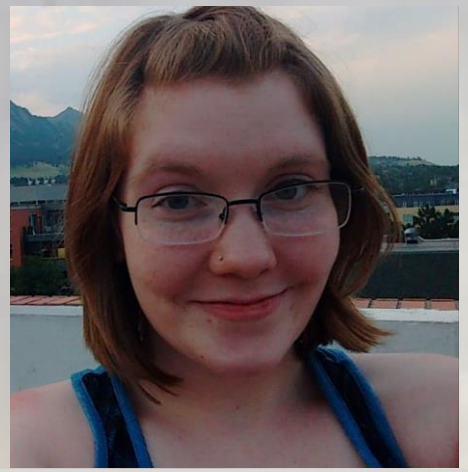
Director, IDEA Lab

University of Oklahoma

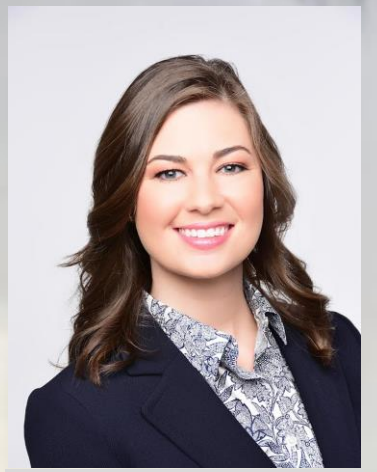
amcgovern@ou.edu

@profamymcgovern

Thank you to my students



Amanda Burke
Metr PhD student, OU



Bethany Earnest
CS PhD, OU

Grant Eckstein
CS BS, OU

David Harrison
Metr PhD student, OU

Josiah Herrington
CS MS, OU

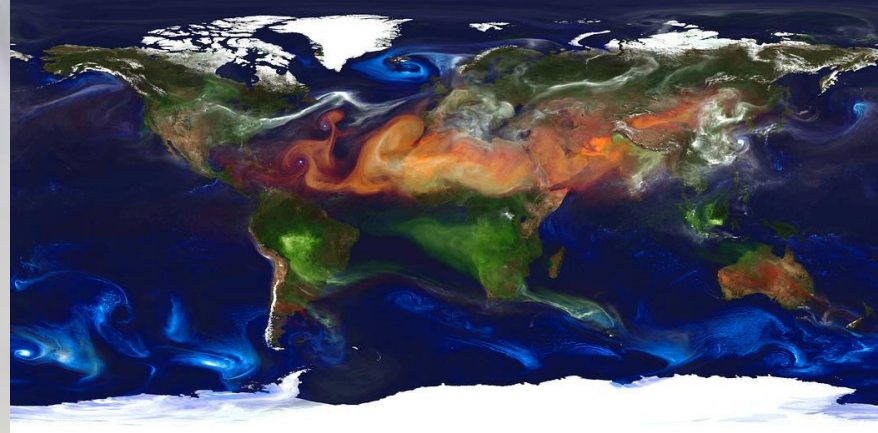
Kendall Junker
Metr MS, OU

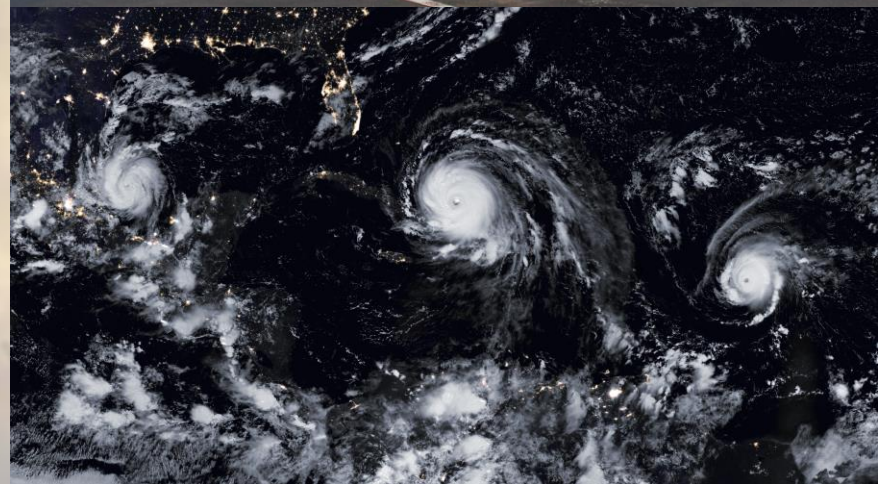
Ryan Lagerquist
OU Metr PhD, CIRA & NOAA GSL

Alyssa Woodward
Metr MS, OU

Motivation: Data Overload

- New sensing systems provide unprecedented high-resolution data
  - Satellites
  - Mobile radar
  - Crowd sourced data
  - And more!
- Sifting through data in real-time is challenging
  - Much of the data is ignored





Motivation: Critical Decisions

- Forecasters must make life-or-death decisions
  - Issue a tornado warning?
  - Evacuate a city for a hurricane?
  - Go-no-go launch
- End users make critical protective decisions
  - Close a business/school early
  - Go into a shelter
  - Get on the roads
- End user needs differ greatly
  - Severe weather: Forecasters, emergency managers, public
  - Crops: Corporate or family owned

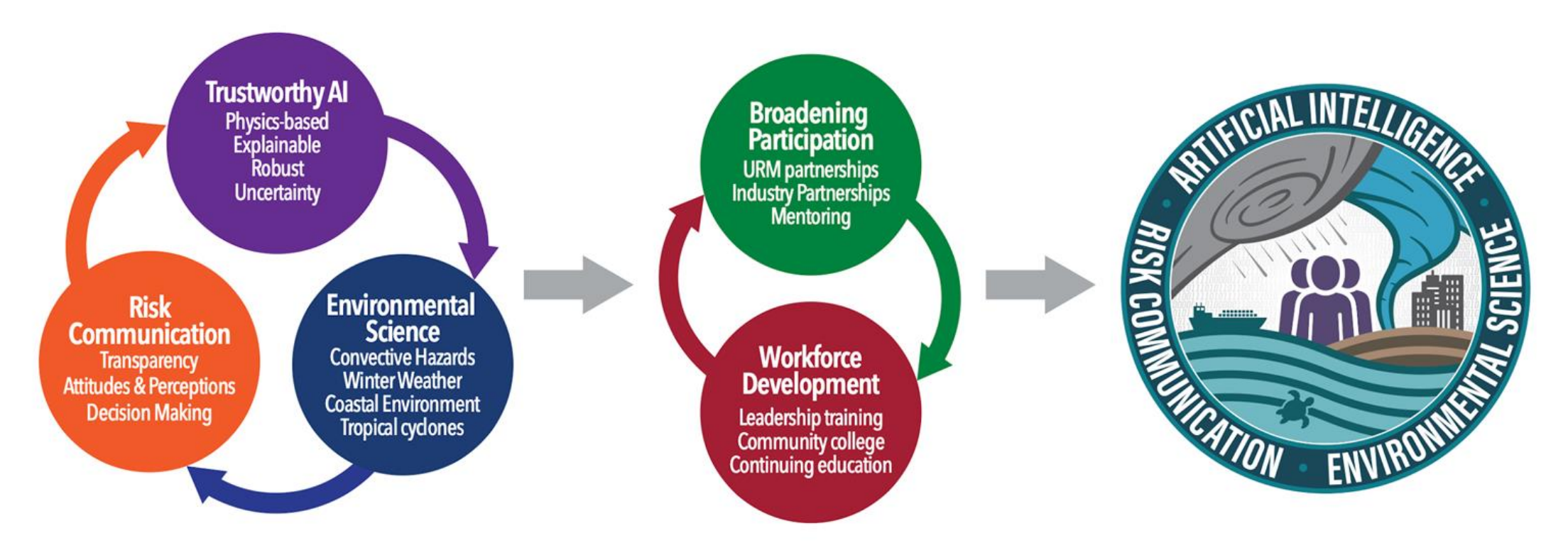
NSF AI Institute for Research on Trustworthy AI in Weather, Climate, and Coastal Oceanography

AI2ES will uniquely benefit humanity by developing novel, physically based AI techniques that are demonstrated to be trustworthy, and will directly improve prediction, understanding, and communication of high-impact environmental hazards.

@ai2enviro https://www.ai2es.org

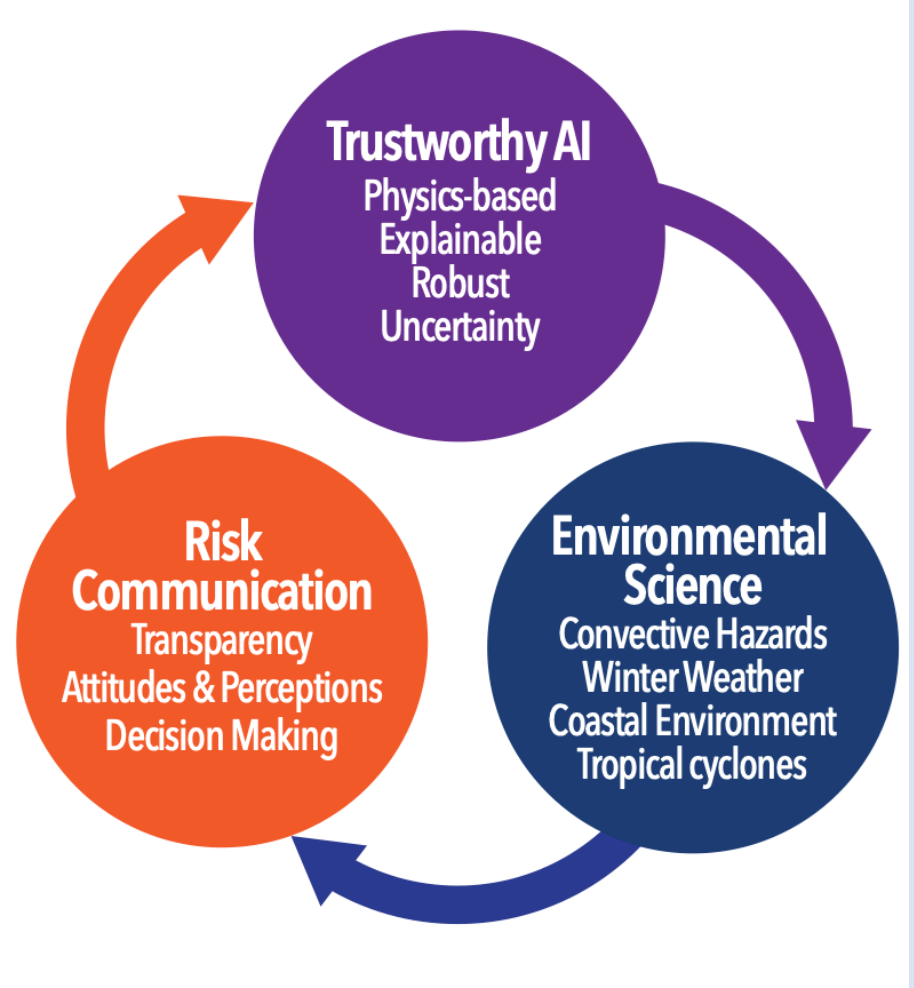
AI2ES Research

- Focus 1: Foundational research in trustworthy AI/ML
- Focus 2: Use-inspired research in ES
- Focus 3: Foundational research in AI risk communication for ES hazards
- Focus 4: AI workforce development and broadening participation



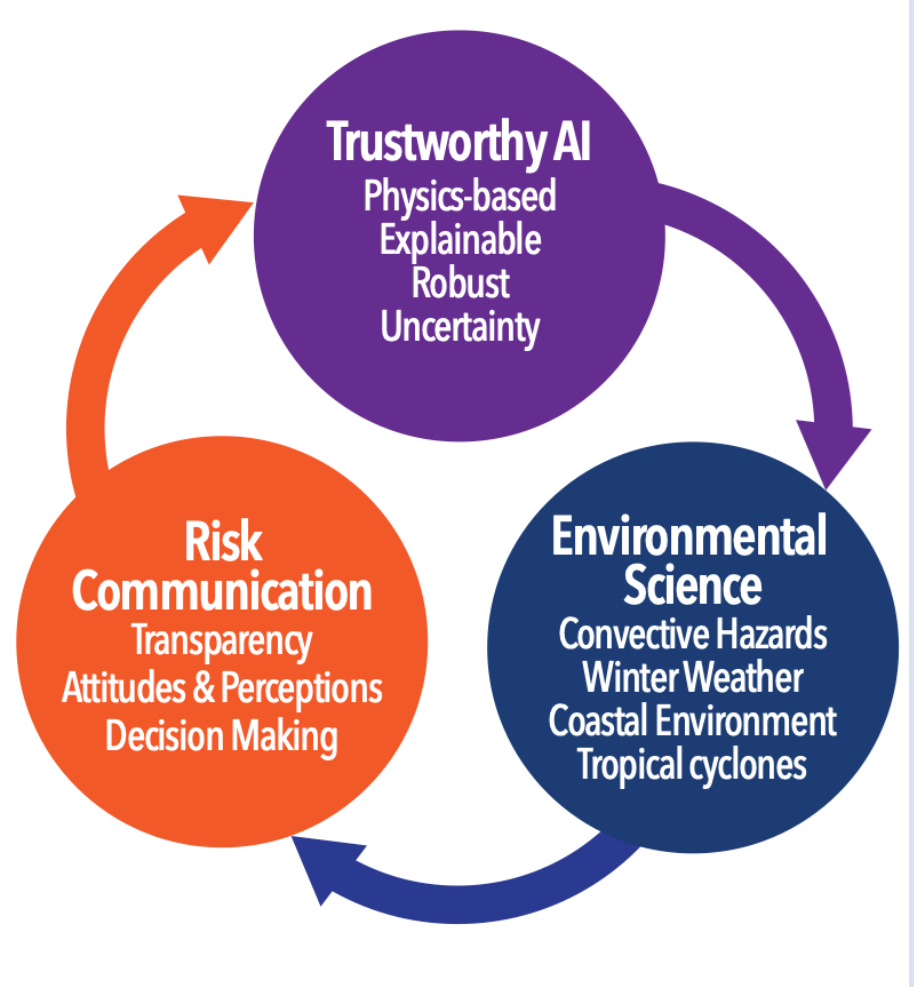
Foundational research in trustworthy AI/ML

- **Goal 1a**: Develop explainable AI methods aligned with ES domain perspectives and priorities.
- **Goal 1b**: Develop physically based AI techniques for ES domains.
- **Goal 1c**: Develop robust AI prediction techniques, and empirically and theoretically validate their performance with adversarial data (e.g., missing data or intentionally wrong data).



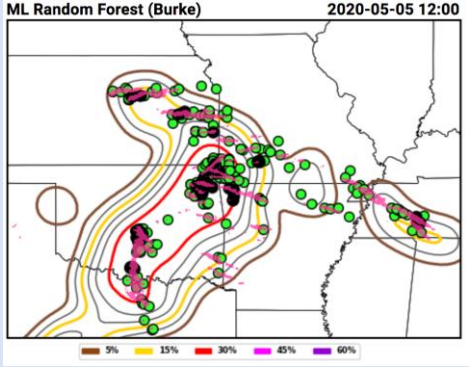
Use-inspired research in ES

- **Goal 2a**: Use trustworthy AI to provide actionable ES information to diverse users.
- **Goal 2b**: Enhance scientific and physical understanding of basic ES processes through trustworthy AI.

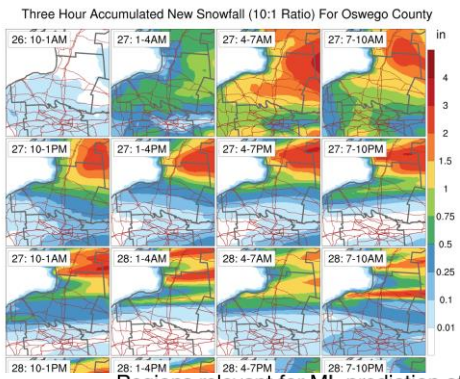


Environmental Science Applications

- Convective Weather:
  - Goal: Improve understanding of tornado and hail formation
- Winter Weather
  - Goal: Exploit underutilized winter weather data to provide tailored guidance to emergency managers and decision makers
- Tropical Cyclones
  - Goal: Improve forecasts of TC temporal evolution and rapid intensification
- Subseasonal to Seasonal (S2S) Prediction
  - Goal: Predict extreme weather 2 weeks to 2 months ahead
- Coastal Oceanography
  - Goal: Improve prediction and understanding of coastal impacts and processes

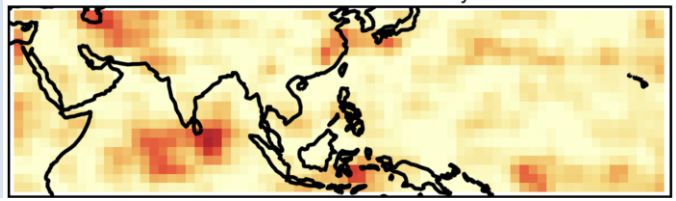


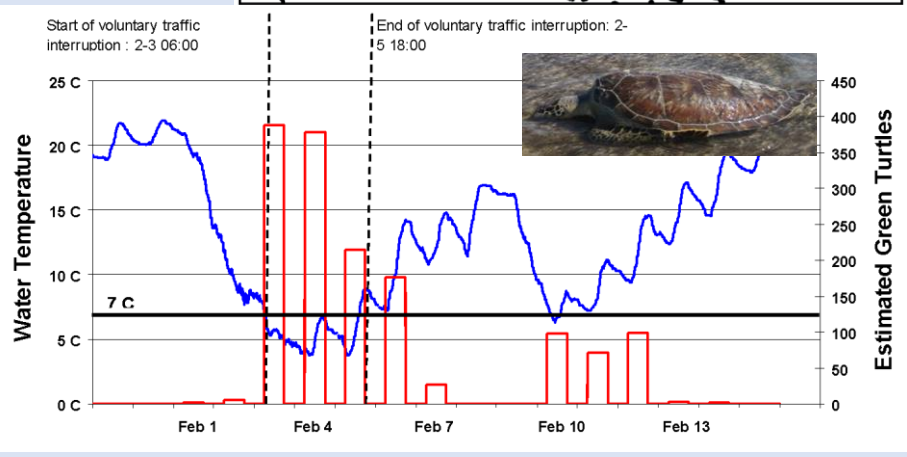
Real-time machine learning hail probability prediction with observed hail reports (green and black) From Burke et al 2020



Example of a predictive roadway weather risk tool developed for the New York State Department of Transportation, from N. Bassill

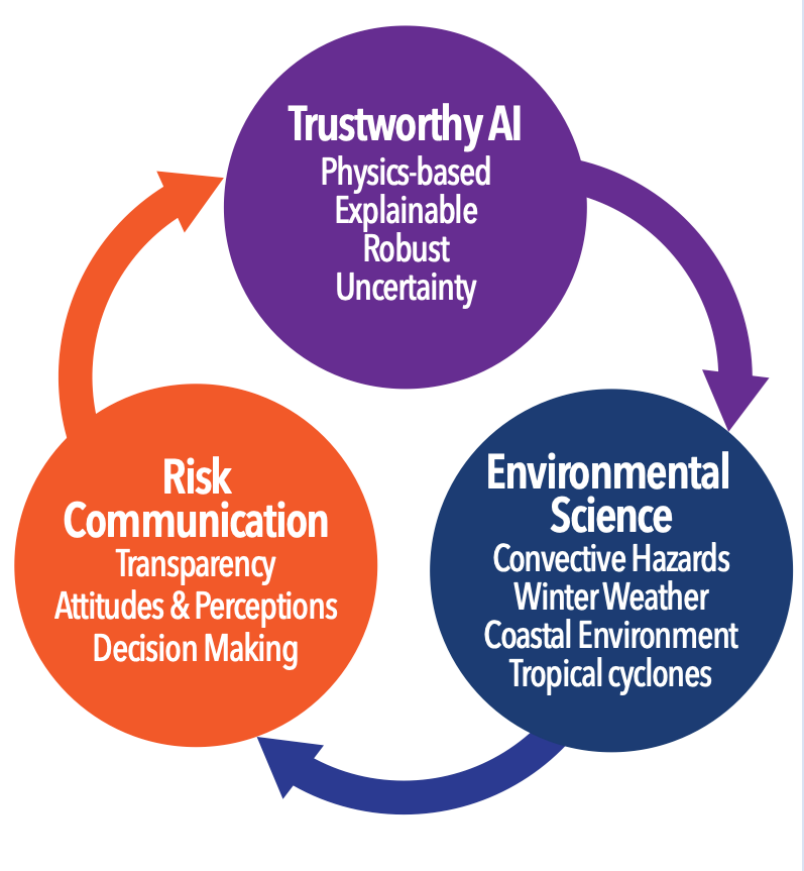
Regions relevant for ML prediction of extreme rainfall over North America 18 days later





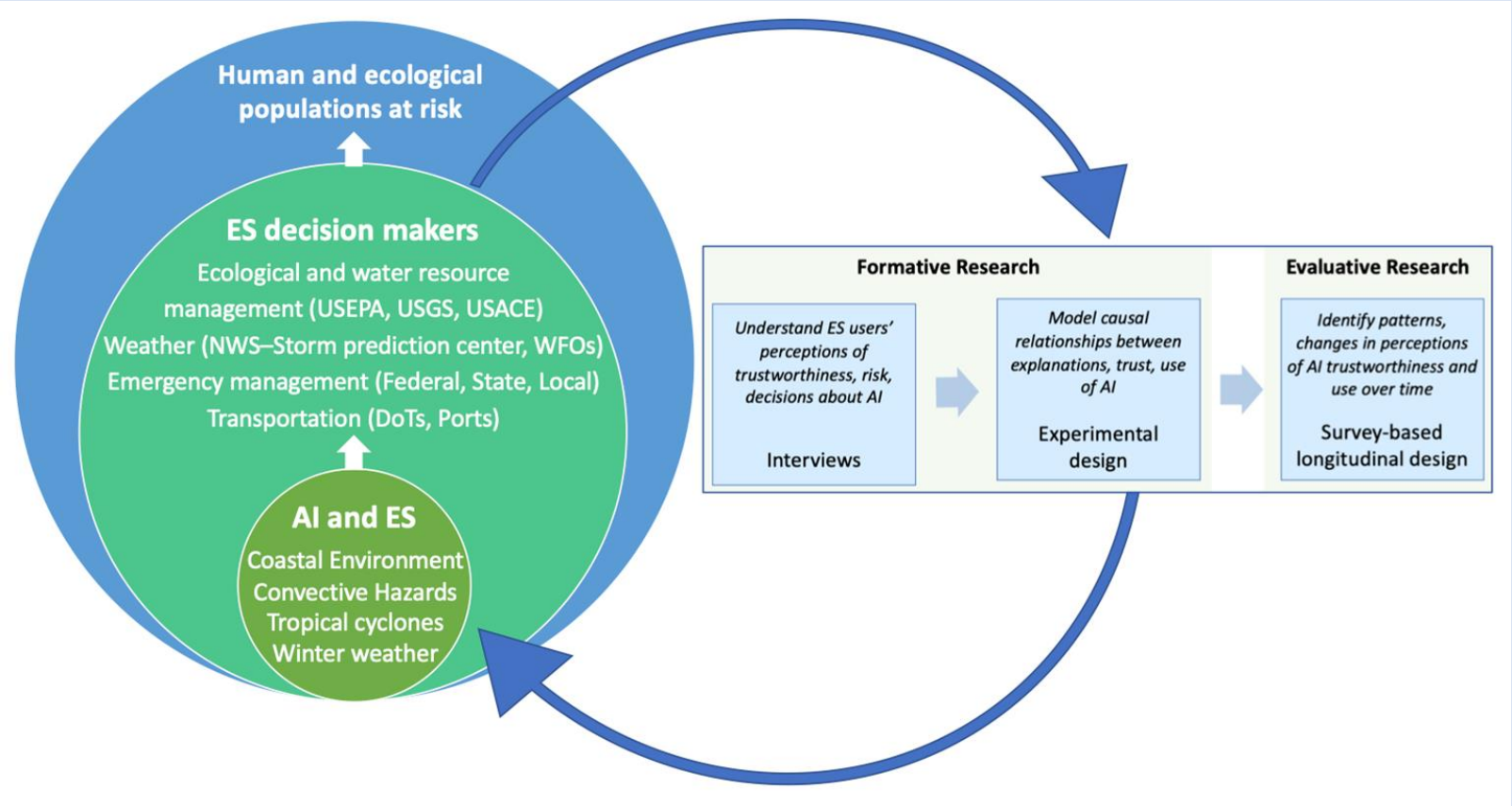
Foundational research in AI risk communication for ES hazards

- **Goal 3a**: Increase knowledge and understanding of how transparency, explanation, reproducibility and representation of uncertainty influence trust in AI for ES for influential user groups.
- **Goal 3b**: Develop models to estimate how attitudes and perception of AI trustworthiness influence risk perception and use of AI for ES.
- **Goal 3c**: Develop principled methods to inform the development of trustworthy AI approaches and the provision of AI-based information to user groups for improved ES decision making.



Foundational RC research with key, influential users

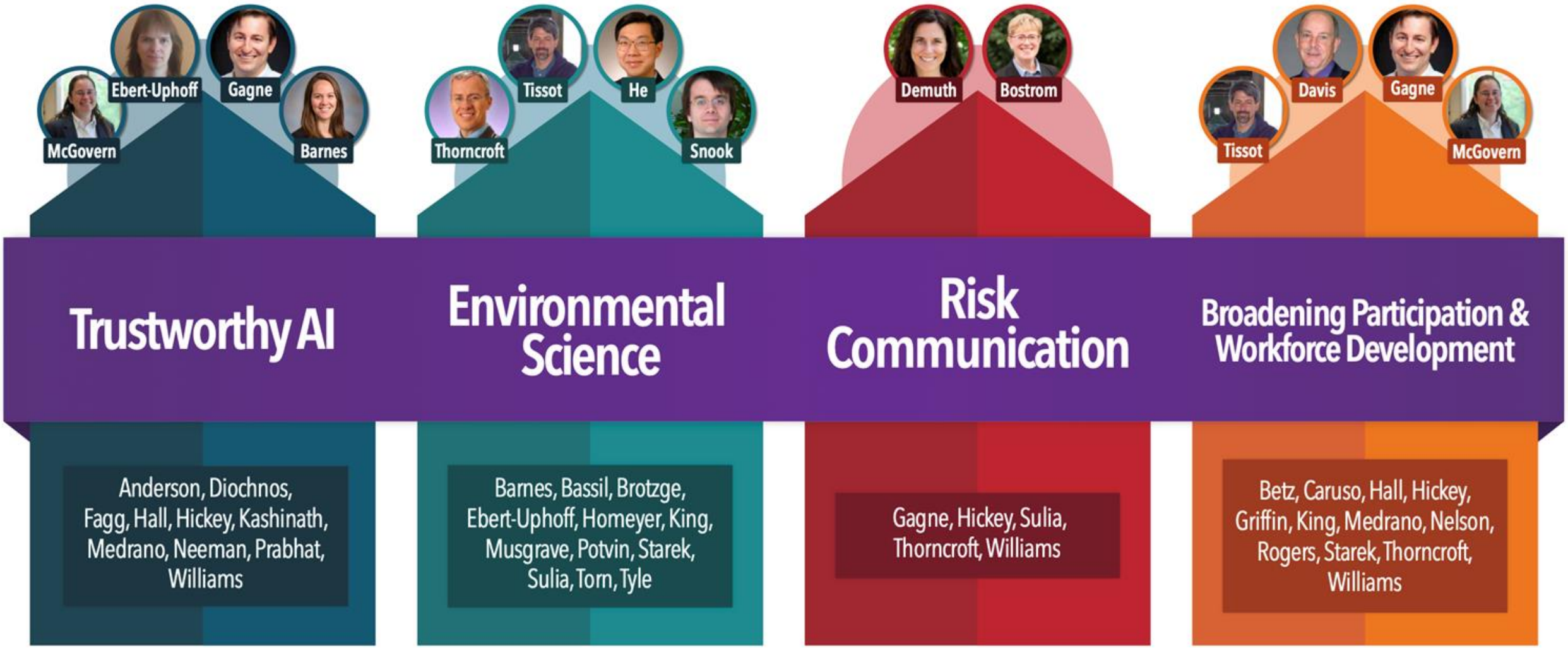
- Relationships between trust and technology acceptance are complex
- Engagement with users is essential for developing AI information that is **trustworthy**
- RC research will develop understanding of what **trustworthy** and **explainable AI** means to users, how trustworthiness and explainability influence risk perceptions and uses of AI
- RC research will inform AI innovation and evaluation across ES hazards



*"When [weather forecasters] cannot easily understand the workings of a probabilistic product or evaluate its accuracy, this reduces their **trust** in information and their willingness to use it." (Demuth et al. 2020)*

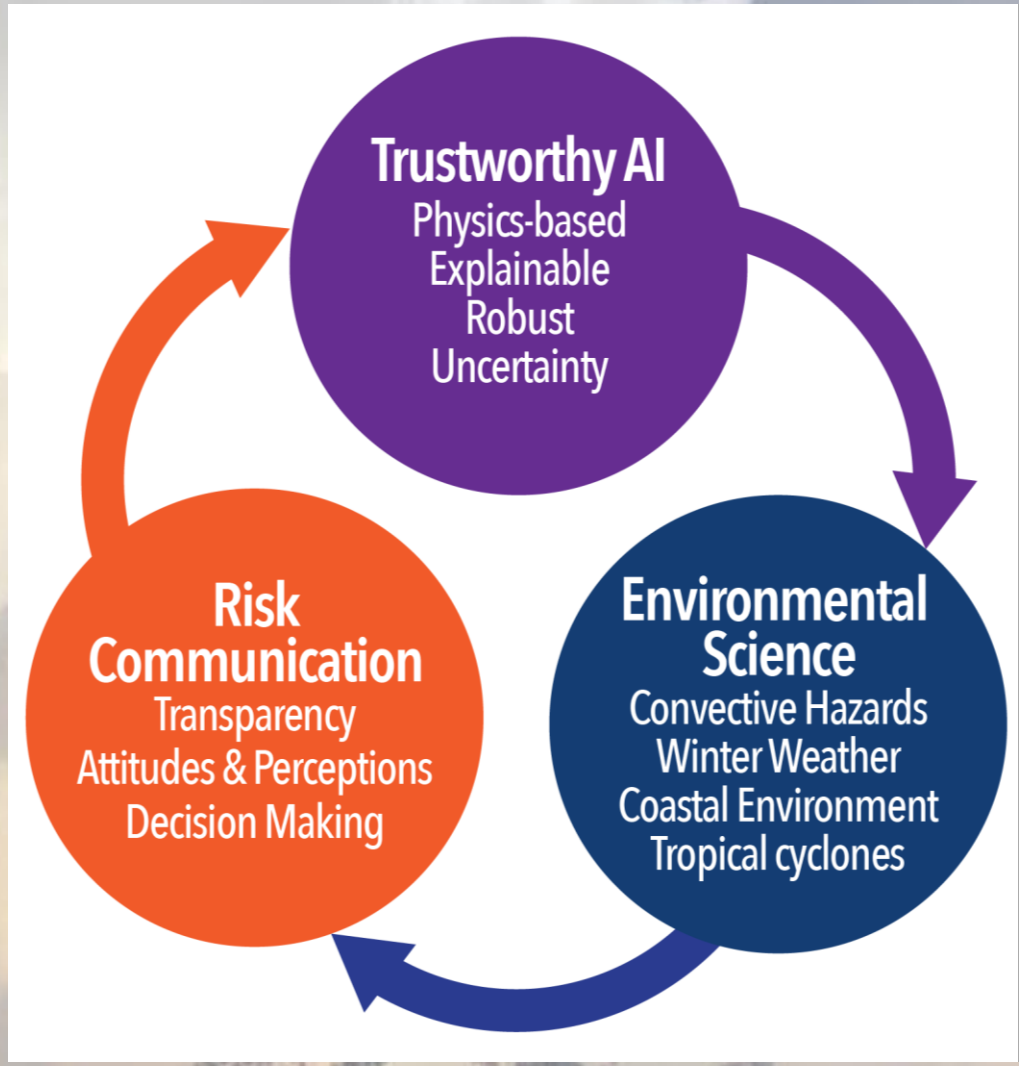
Recommendations for developing useful and usable convection-allowing model ensemble information for NWS forecasters

AI²ES Team


McGovern
Ebert-Uphoff
Gagne
Barnes
Thorncroft
Tissot
He
Snook
Demuth
Bostrom
Tissot
Davis
Gagne
McGovern
Trustworthy AI
Environmental Science
Risk Communication
Broadening Participation & Workforce Development
Anderson, Diochnos, Fagg, Hall, Hickey, Kashinath, Medrano, Neeman, Prabhat, Williams
Barnes, Bassil, Brotzge, Ebert-Uphoff, Homeyer, King, Musgrave, Potvin, Starek, Sulia, Torn, Tyle
Gagne, Hickey, Sulia, Thorncroft, Williams
Betz, Caruso, Hall, Hickey, Griffin, King, Medrano, Nelson, Rogers, Starek, Thorncroft, Williams

My AI/ML Research for Environmental Sciences

- Research questions:
  - How can we use ML to improve **prediction** and **understanding** of high-impact environmental science phenomena?
  - Can we use ML in real-time to save lives and property?
  - How can ML enable new scientific discovery?
  - When/why do different end-users use automated technology?
- My group focuses on making ML work in the real world
  - Focus extends beyond environmental sciences but today focuses on weather

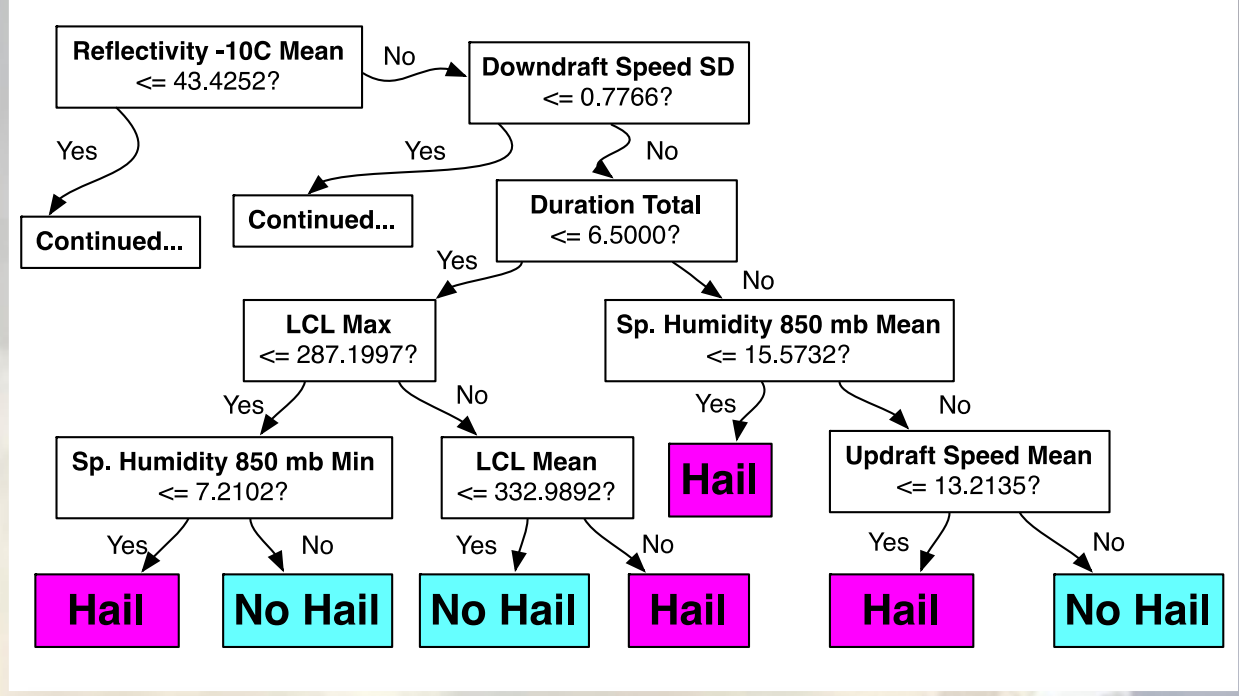


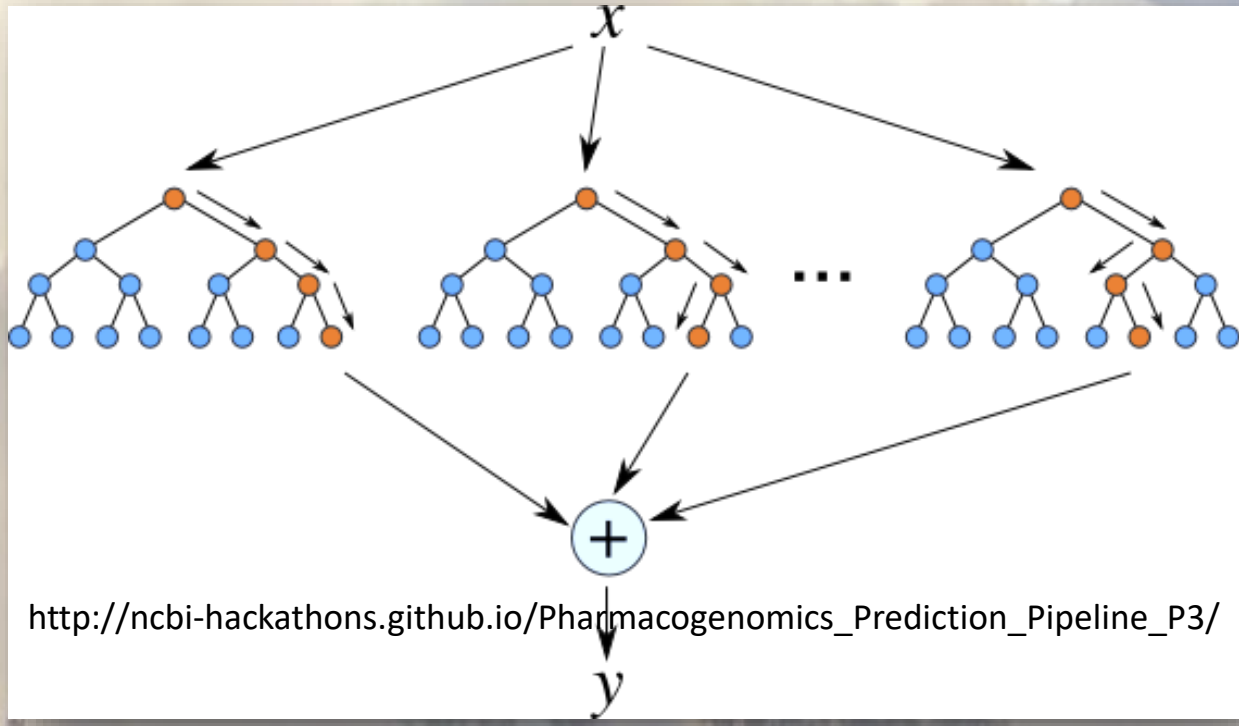
Outline

- Motivation for trustworthy AI
- Current work
  - Demonstrating ML can be used to improve prediction for multiple severe-weather hazards (this talk: hail and tornadoes)
  - Working with end-users to improve trust in ML predictions
  - Developing physically-based model interpretation and visualization techniques for environmental science
- Future work

Decision Trees and Random Forests

- Human-readable ML model
- Can predict:
  - Class labels (hail/no hail)
  - Real-values (hail size)
- Demonstrated success in meteorology
  - Selective model
- Random Forests
  - Individual trees trained on bootstrap resampled subsets of data
  - Trees use subsets of attributes at each level

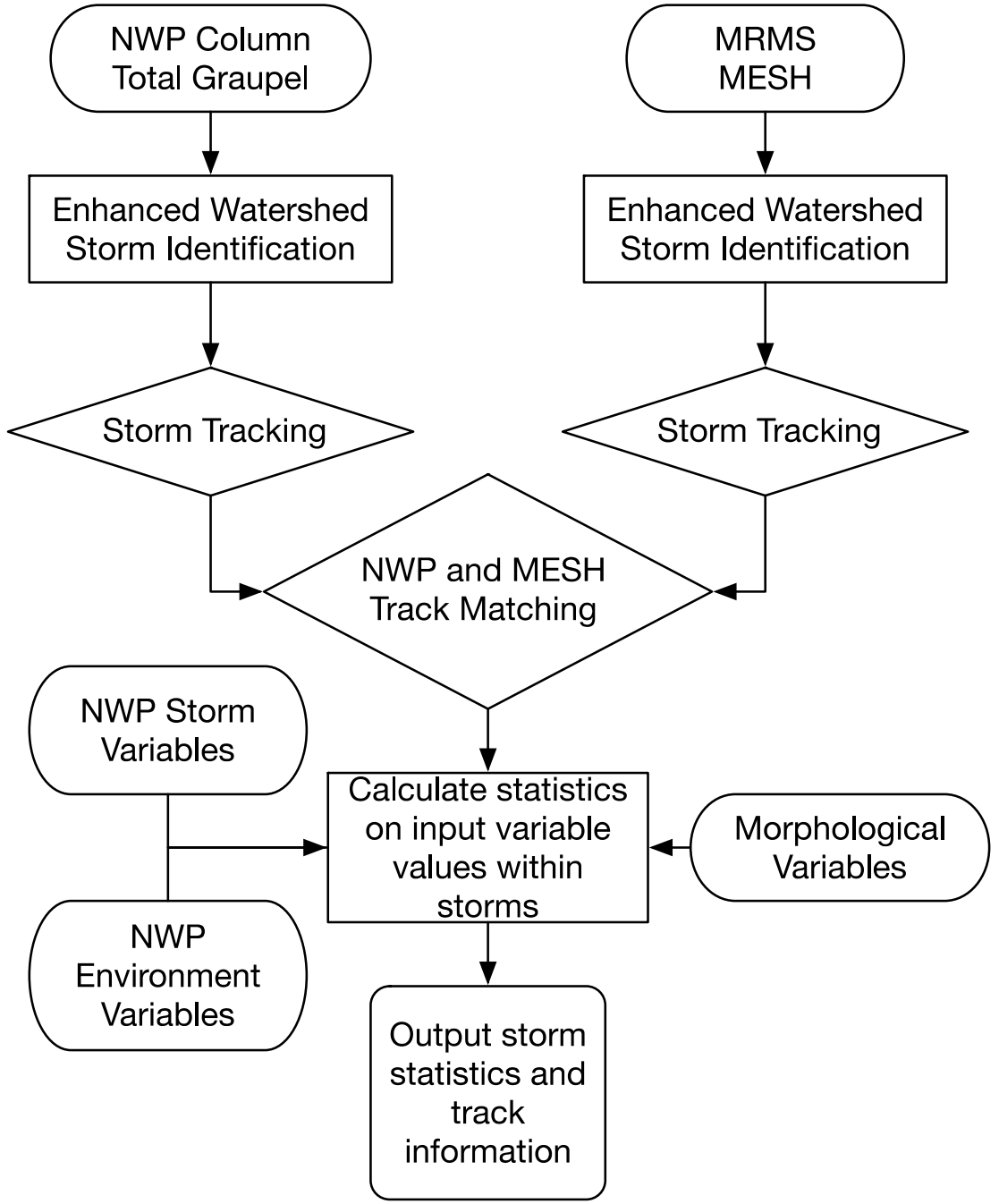




http://ncbi-hackathons.github.io/Pharmacogenomics_Prediction_Pipeline_P3/

Storm-Based Hail Forecasting

- Overall steps:
  - Extract data from NWP
  - Train ML
  - Predict hail & size
- Implemented and tested in NOAA's Hazardous Weather Testbed (multiple years)
- Details:
  - Gagne et al, WAF 2017
  - Burke et al, WAF 2020

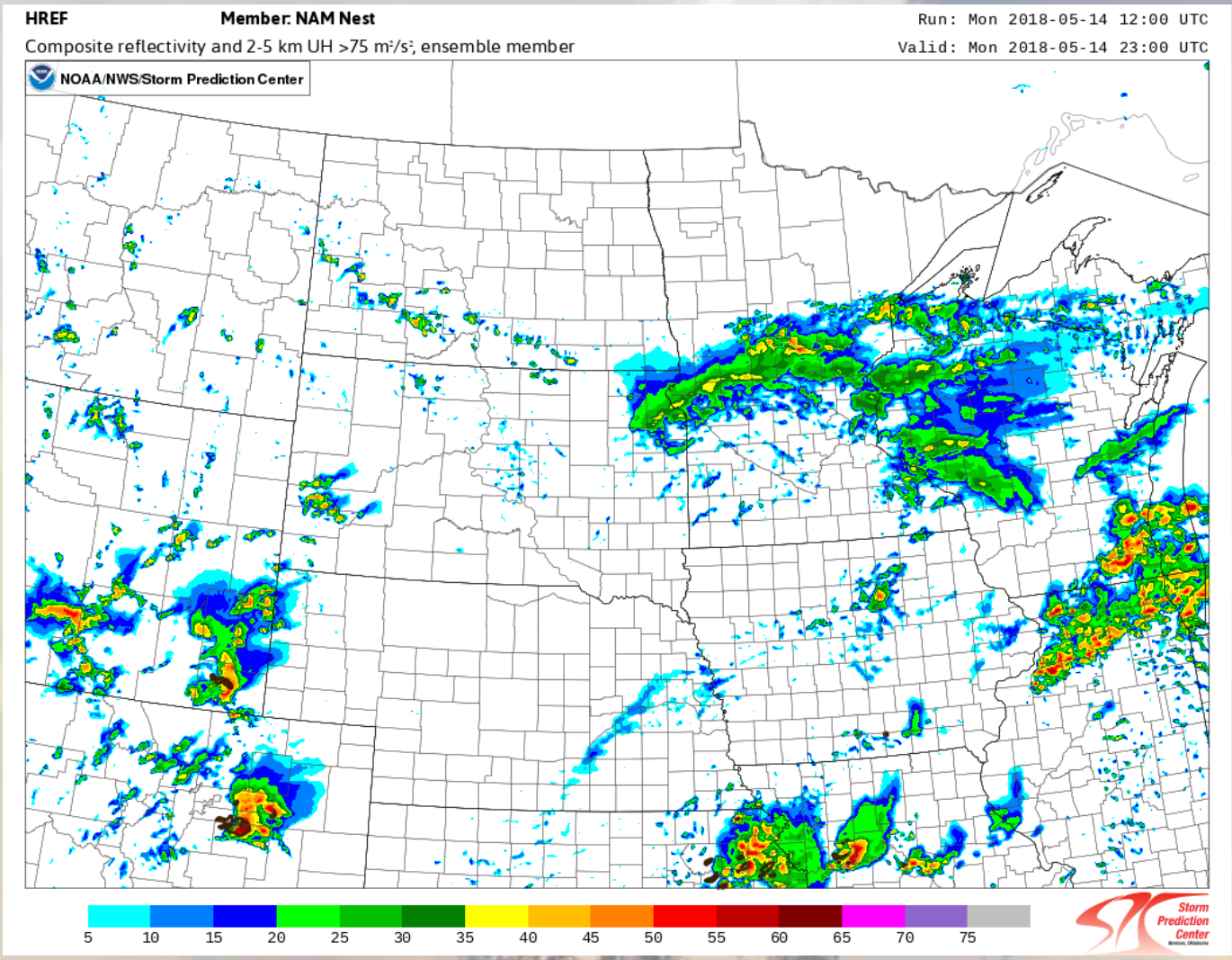


NWP Convection Allowing Models (CAMs)

- CAMs have high spatial and temporal resolution across CONUS
  - Resolution too low to resolve hazards such as hail
  - ML can predict missing hazards and correct spatial or temporal forecast errors
- Gagne et al 2017
  - CAPS Spring Experiment ensemble
  - NCAR ensemble
- Current work (Burke et al, WAF 2020)
  - High Resolution Ensemble Forecast version 2 (HREFv2)
    - Operational in Storm Prediction Center (SPC)
  - Eight member ensemble
    - Initialized 0000z and 1200z
    - Mixed model ensemble (WRF-ARW and NMMB)

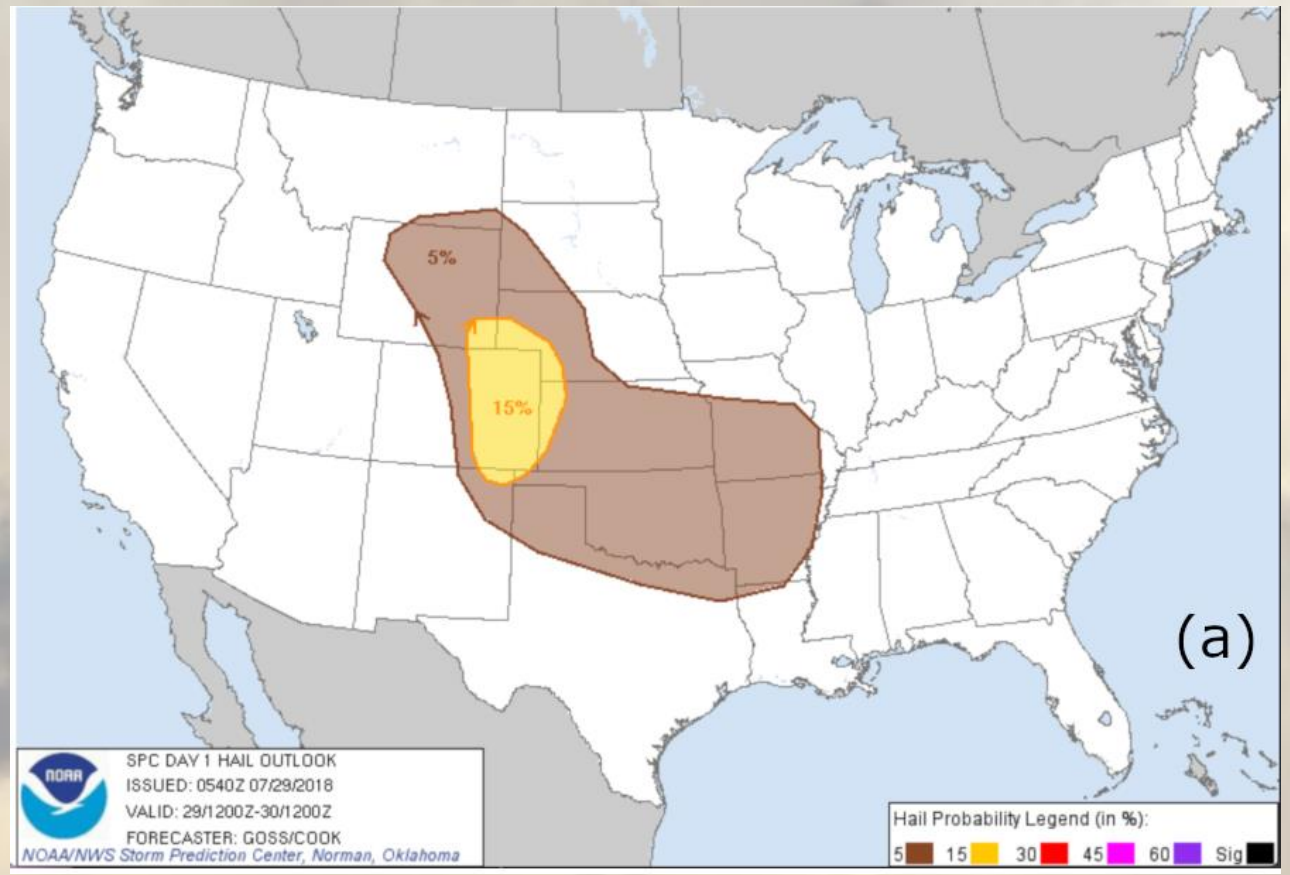
ML Training Data

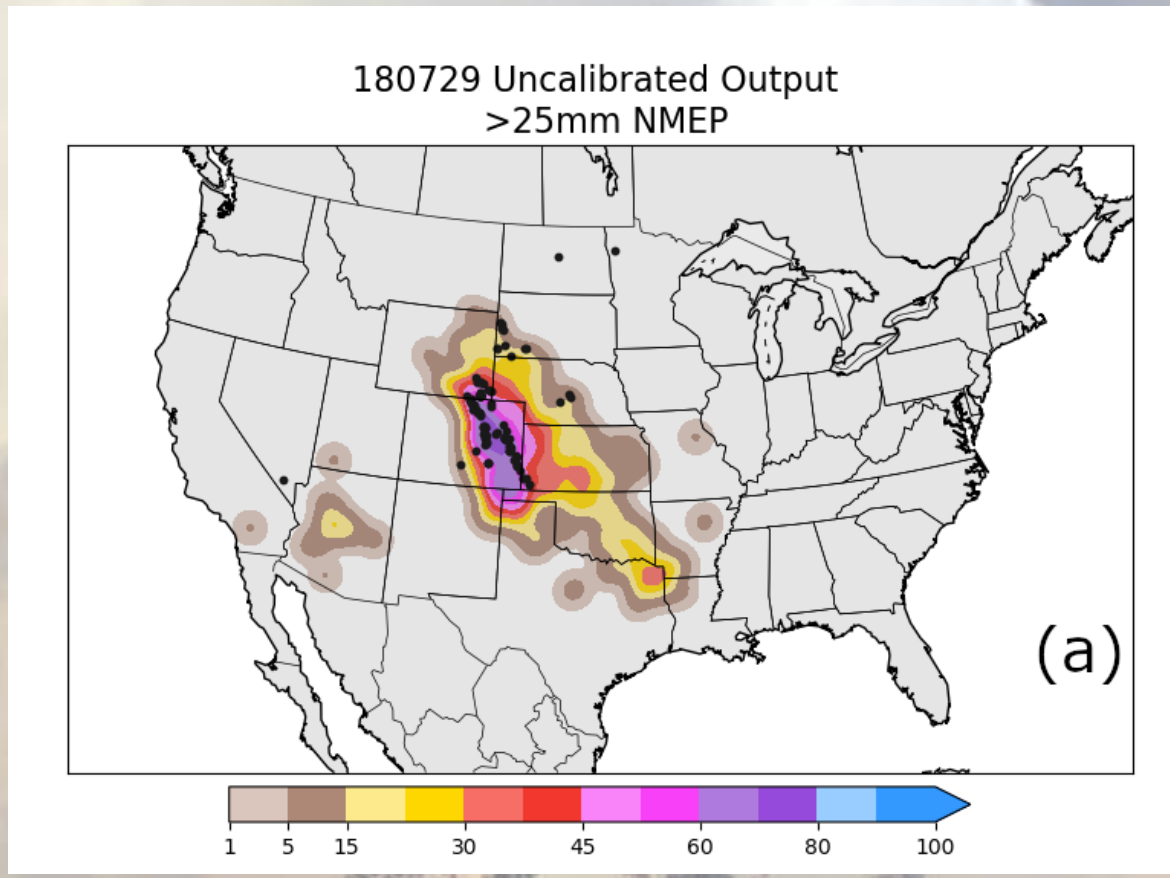
- Extract data for each storm object/track
  - Storm data: updraft/downdraft, reflectivity, precipitation, etc
  - Environment: temperature, CAPE, wind, sounding data, etc
  - Morphological: Area, shape, etc
  - Location: Forecast hour, duration, motion
- Hail labels: Maximum Estimated Size of Hail (MESH) > 19mm (3/4 in)



Initial Forecaster Evaluation: Too “hot”

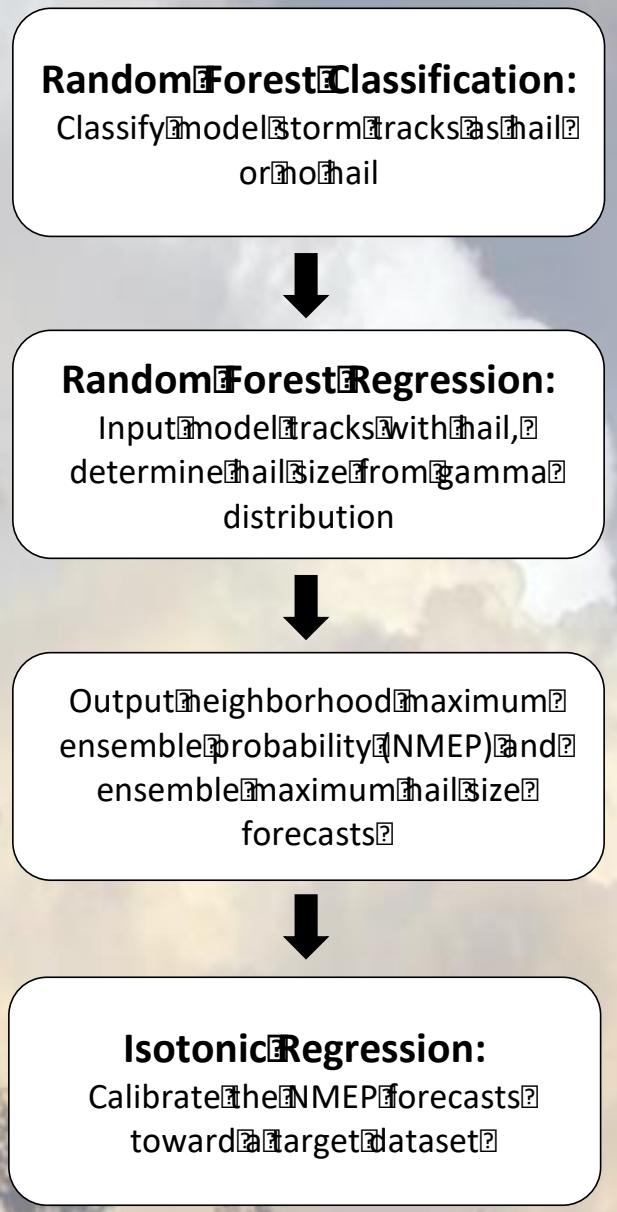
The end-user’s needs (SPC forecasters) matter



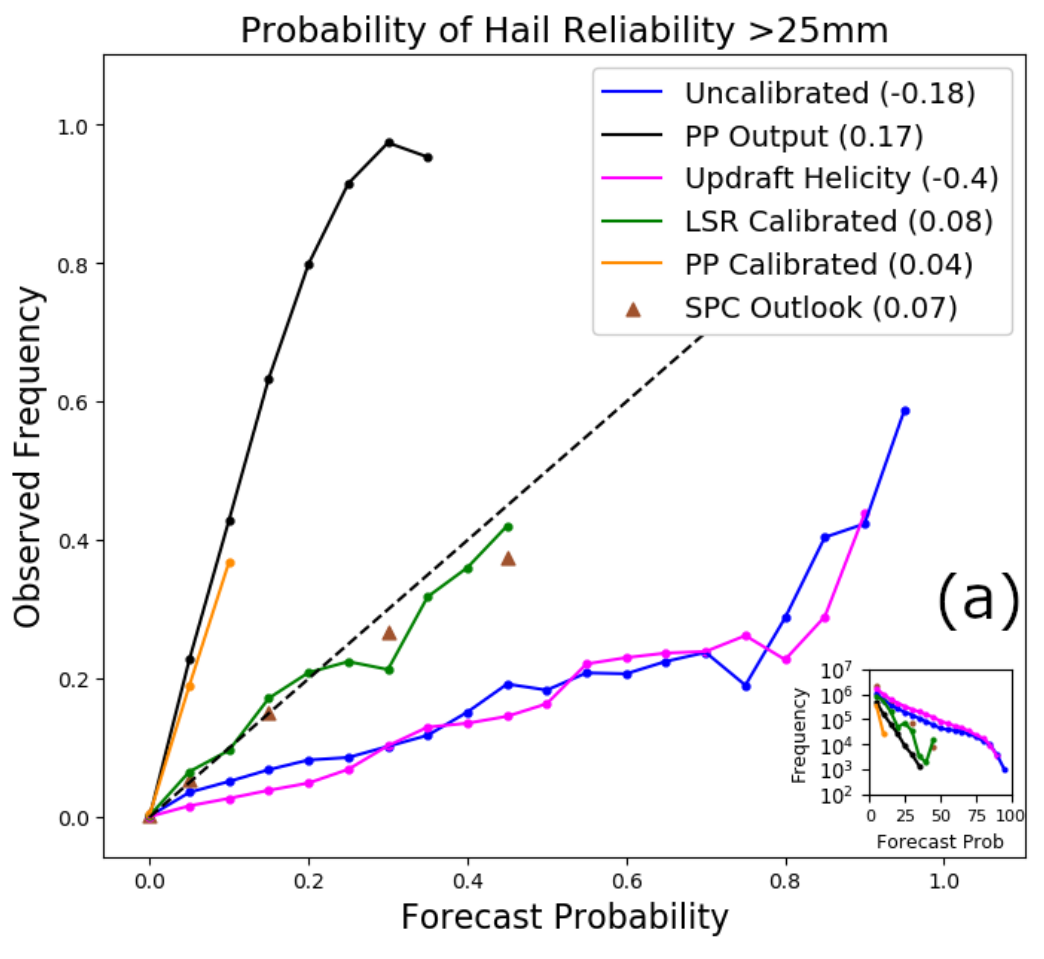


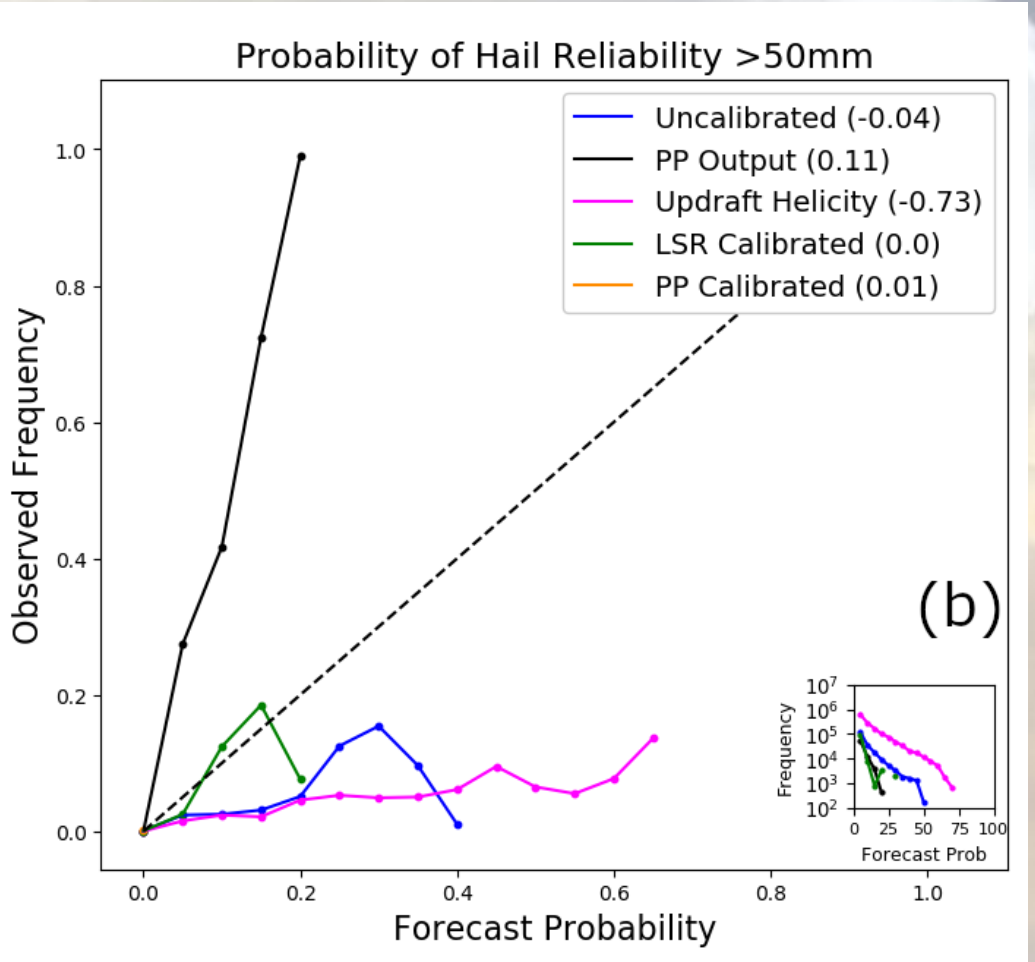
Calibration: Forecaster Trust

- New addition to this work focuses on trustworthy AI
- Human analysts did not “trust” the output of the model because it was “too hot”
- Probability calibration
  - Calibrated to local storm reports, practically perfect (SPC verification metric), and MESH



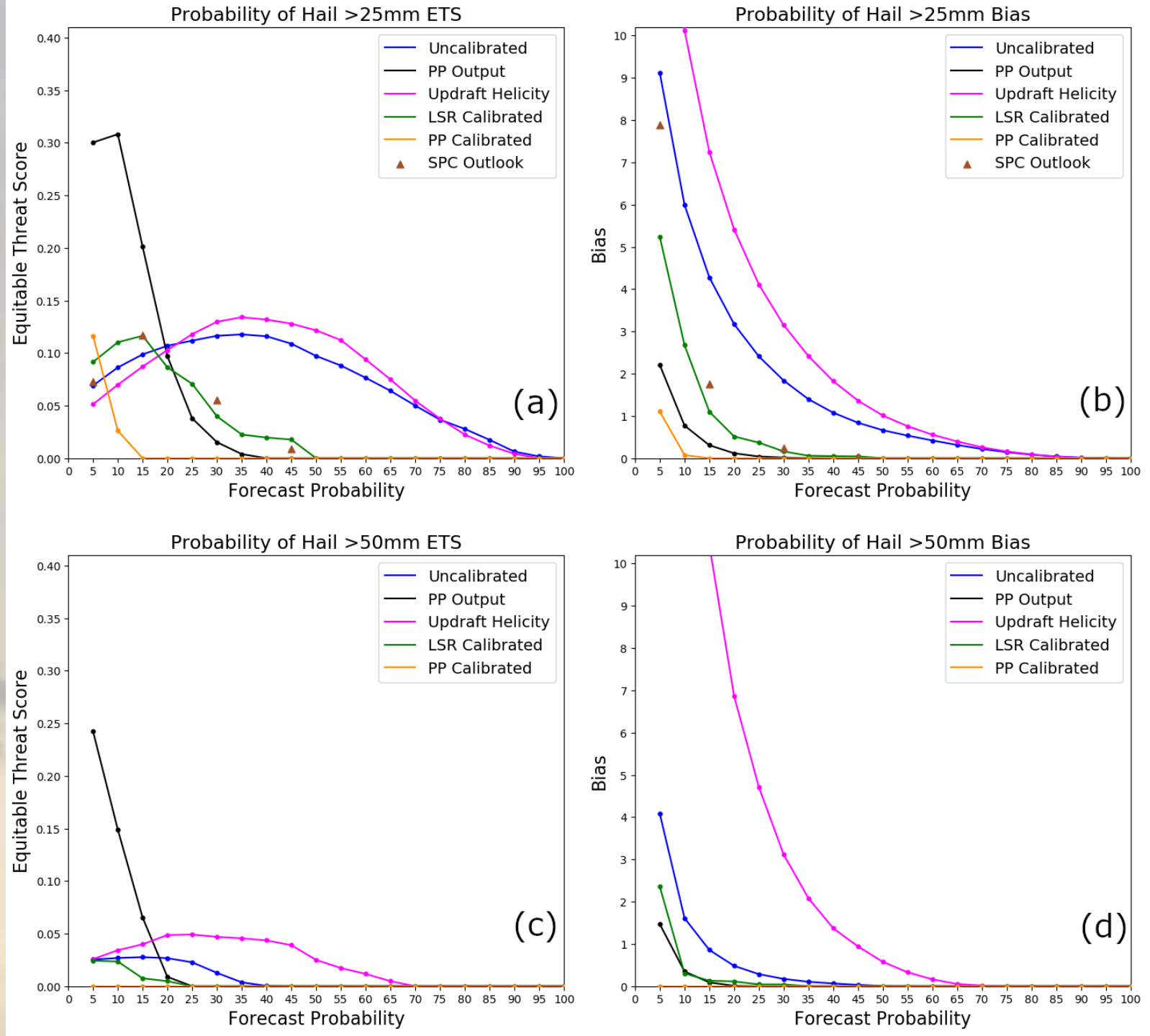
Objective verification: Reliability





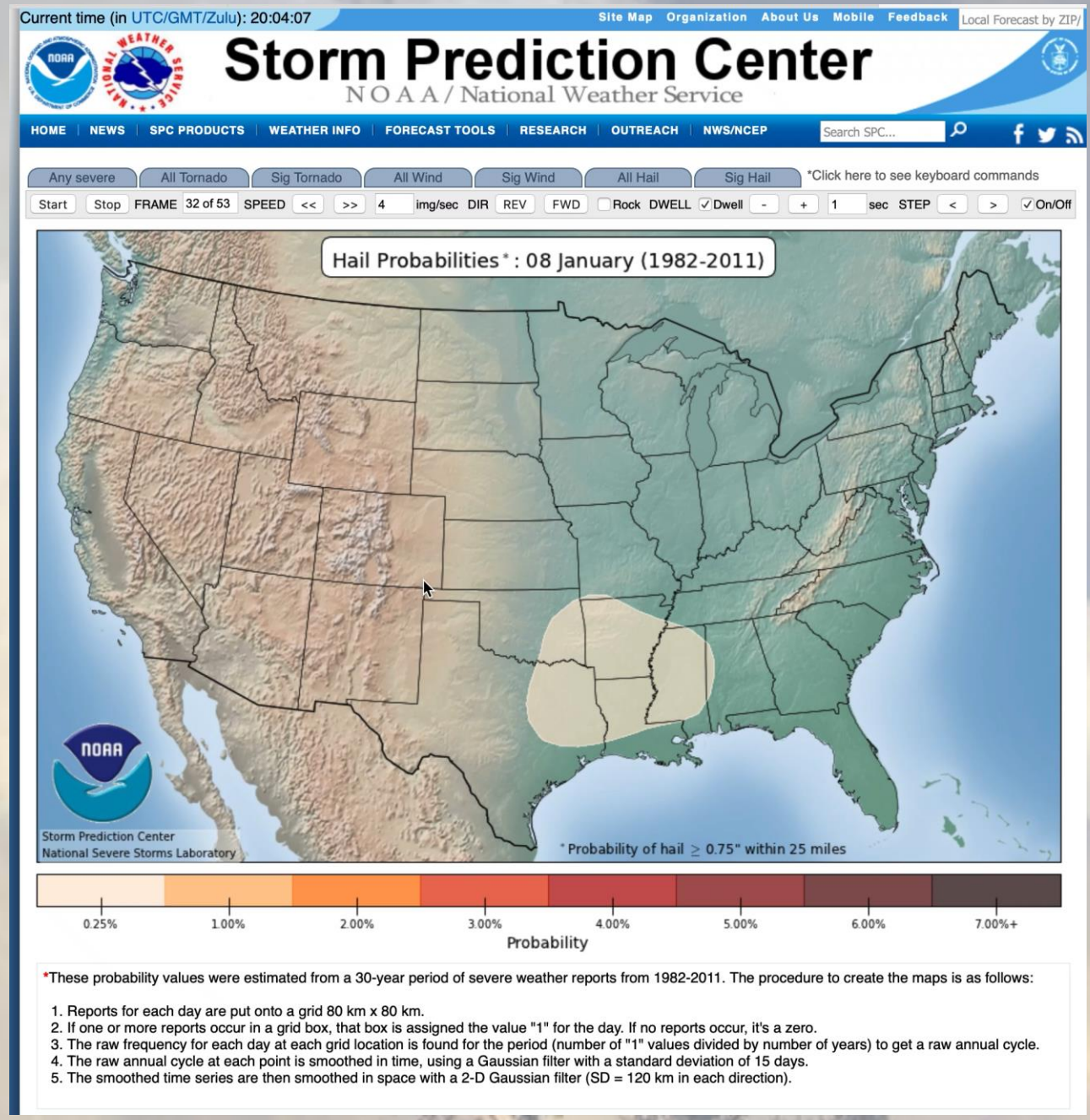
Objective verification

- Calibrated forecasts have high ETS and lower bias

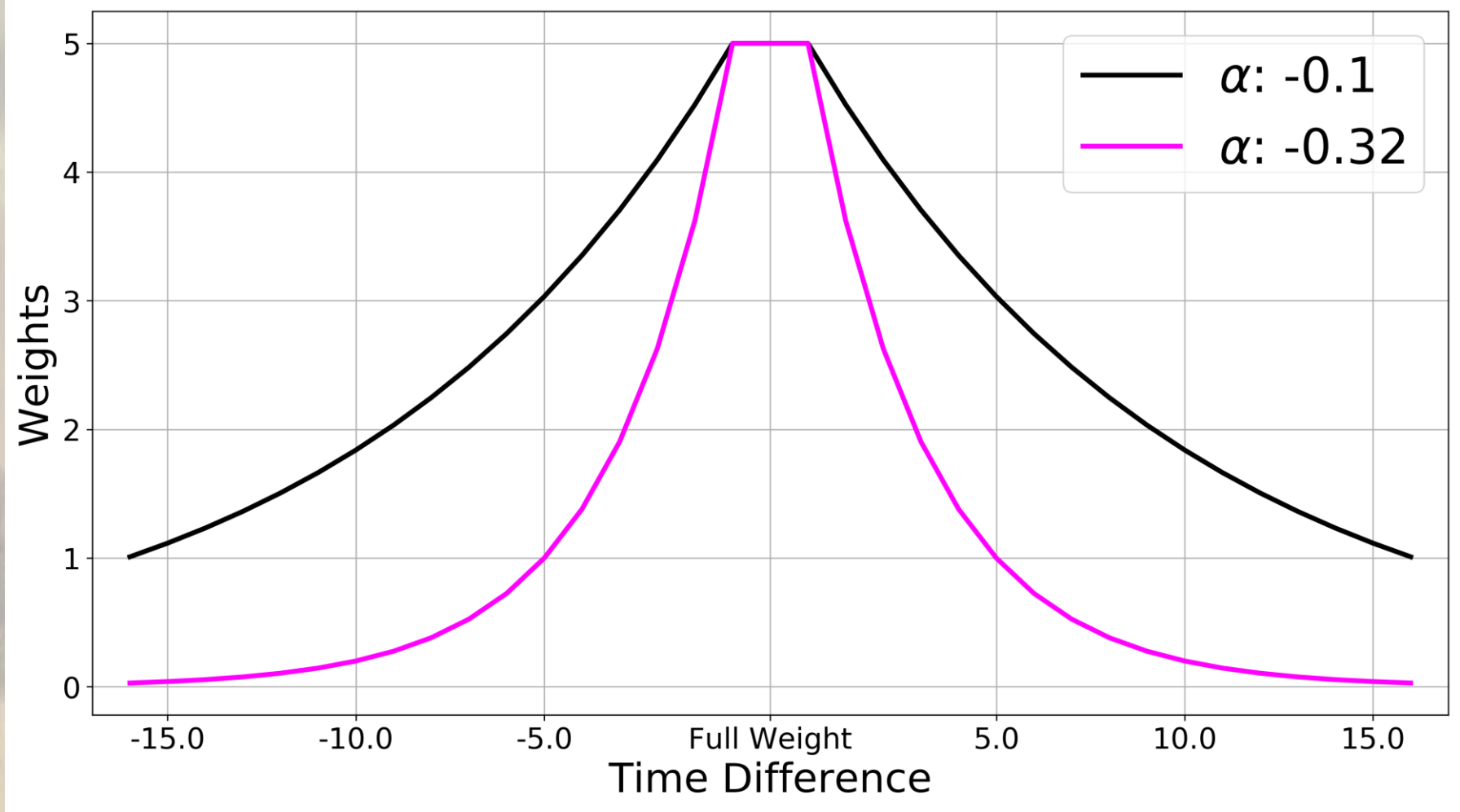


Continuing to Improve Forecaster Trust

- Observations:
  - Hail production differs by season
  - Hail production differs by region
  - Training data is limited
    - Need both hail observations and NWP data to train
- Research question: can we weigh the training data to maximize training power and observe regional hail differences to improve trust?

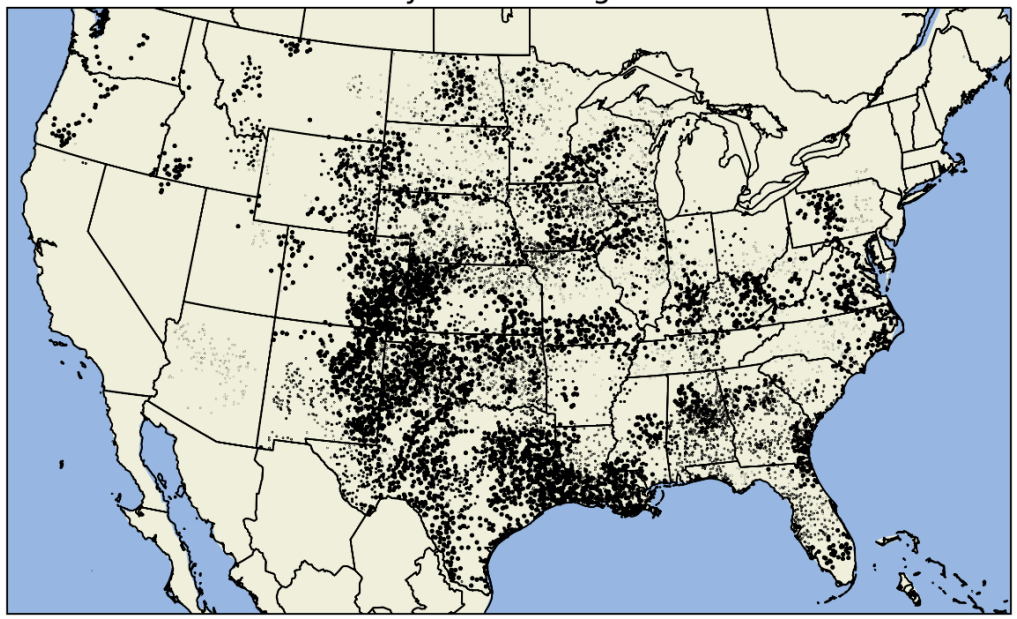


Monthly Storm Weighting

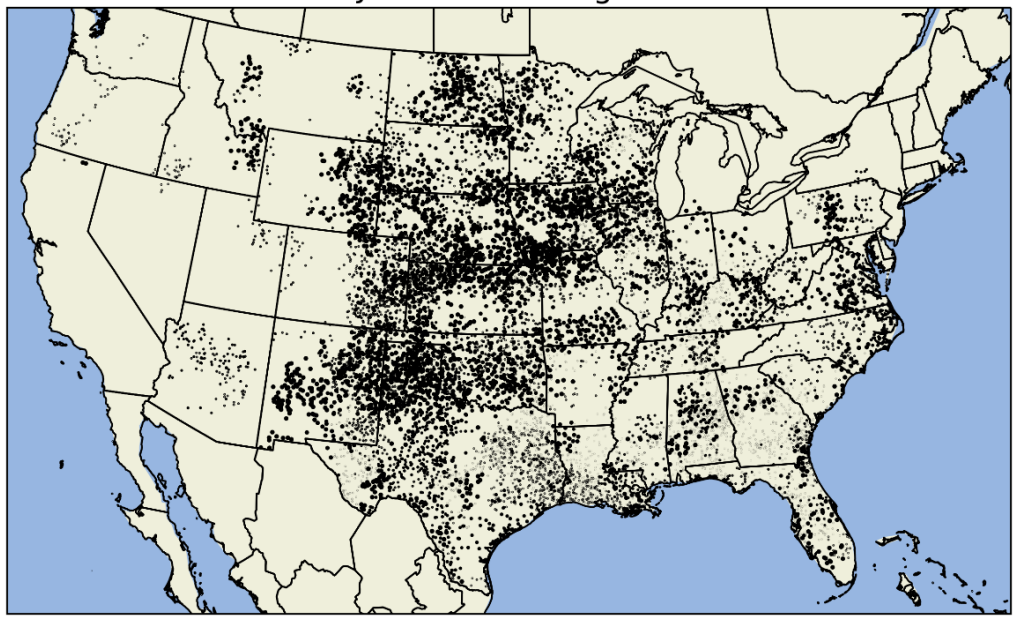


Monthly Storm Weighting

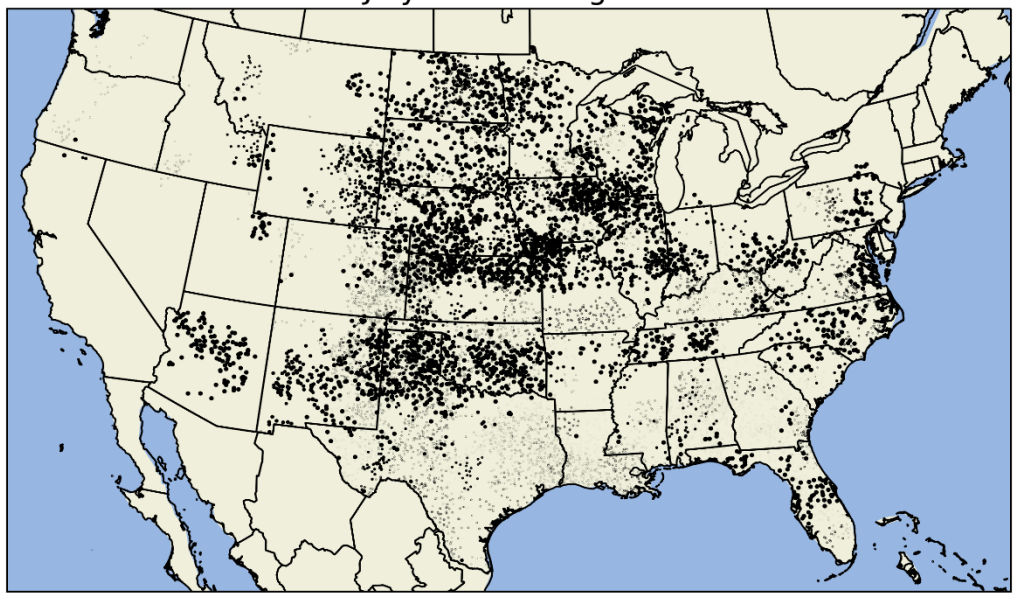
May Storm Weights



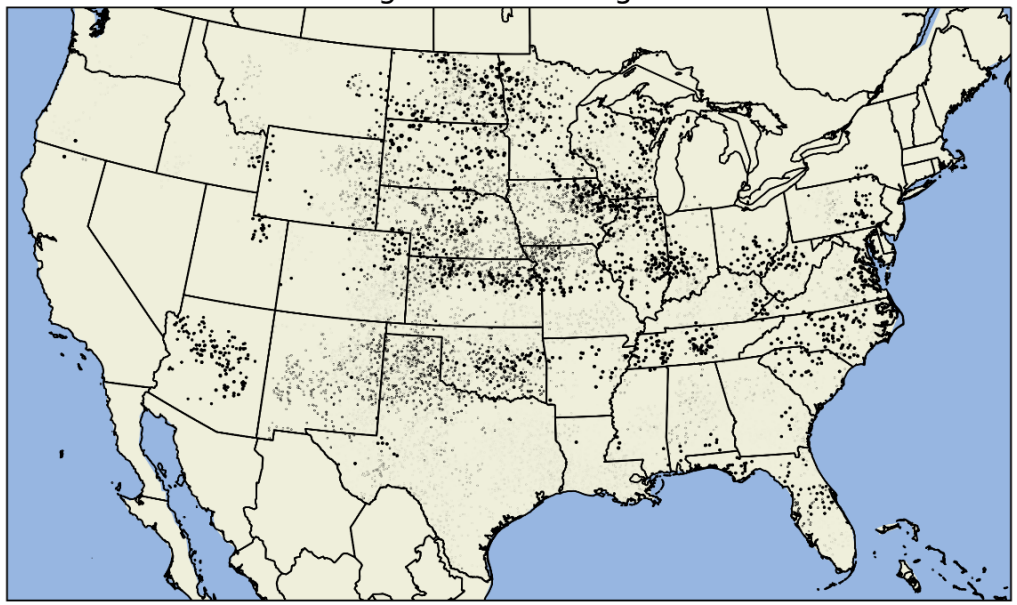
June Storm Weights



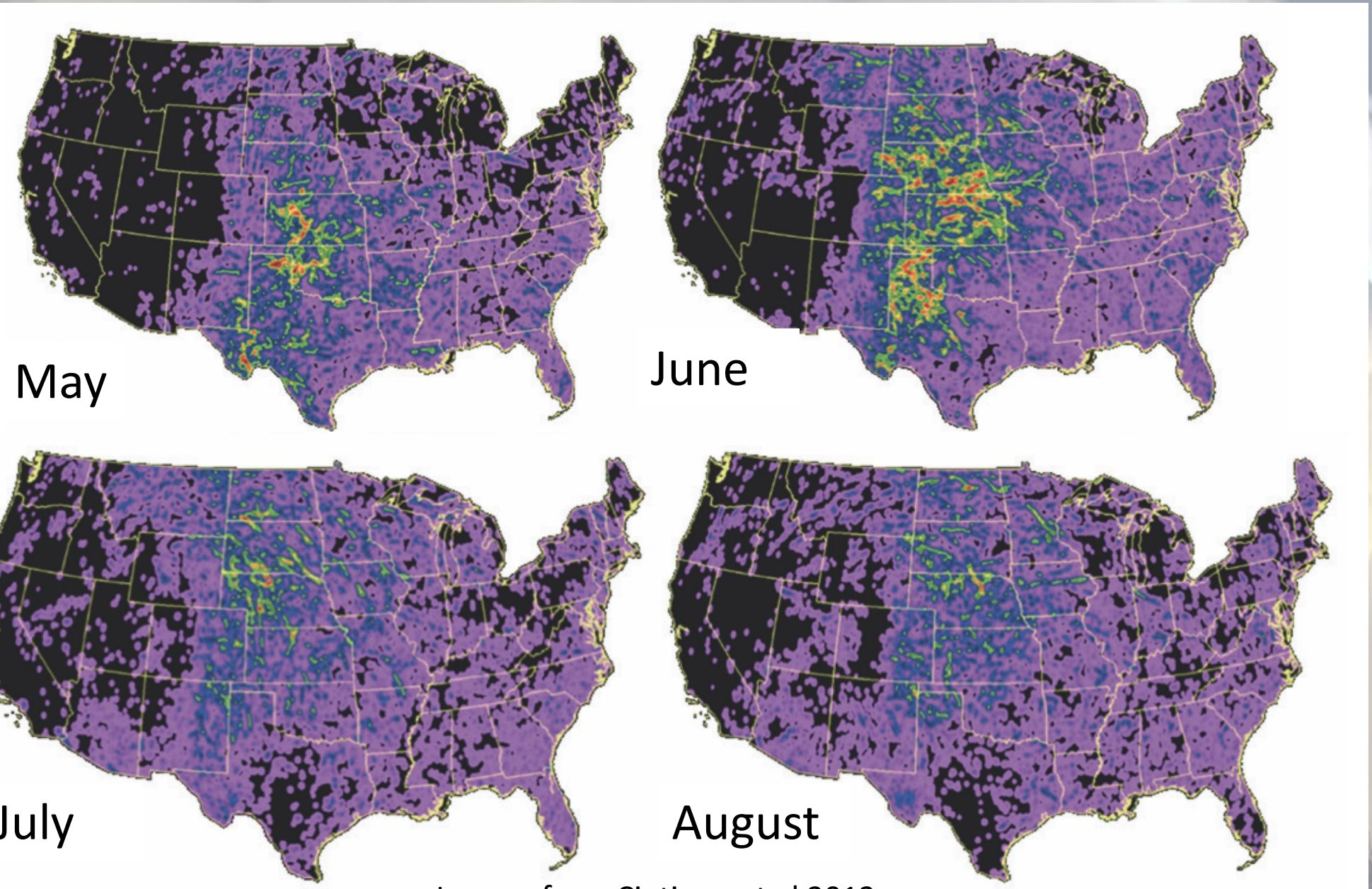
July Storm Weights



August Storm Weights



Monthly Storm Climatology



Images from Cintineo et al 2012

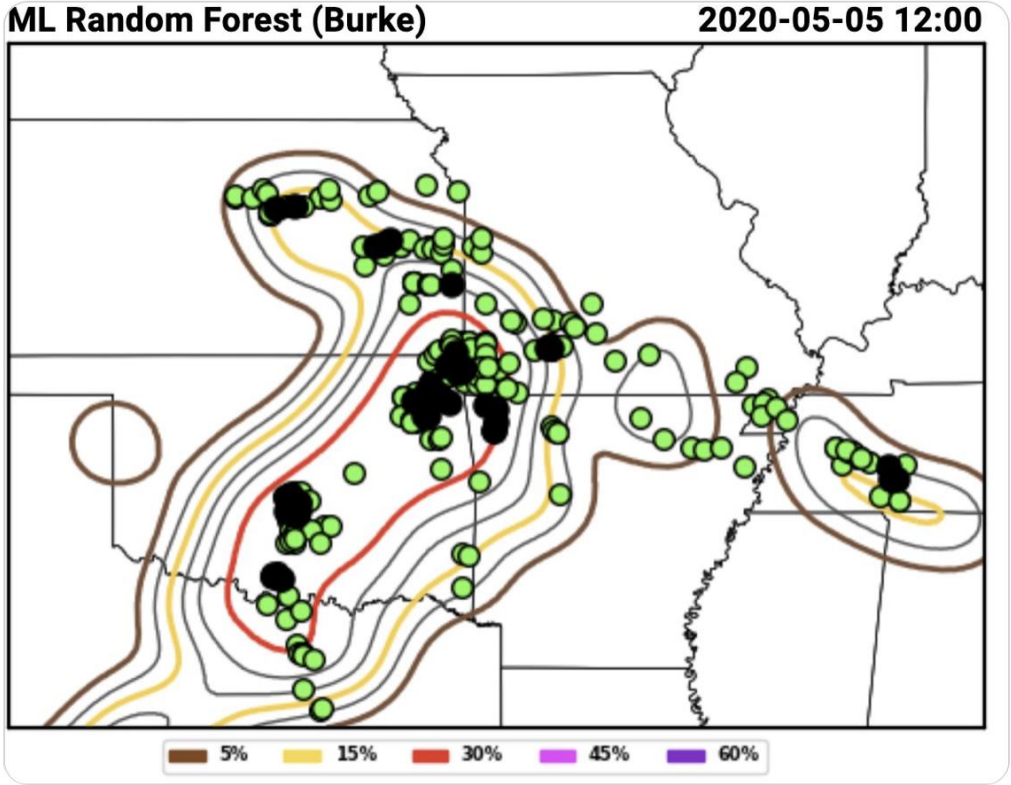
Evaluation of Forecaster Trust: HWT

- New spatial weighted tested in NOAA's Hazardous Weather Testbed in Spring 2020 (all virtual)
- "AI could be a game changer" – Adam Clark of the Storm Prediction Center
- Additional evaluations coming soon (paper in preparation)

Tweet

Adam Clark
@AdamClarkWx

CAMs don't directly simulate severe wx, so computing model-derived tor/hail/wind probs is hard. Storm "proxies" (e.g., UH) are useful, but #AI could be a game changer. #SFE2020 features 4 AI-based severe wx projects. Good ex. of skillful hail prediction by @AmandaLeo_wx here.

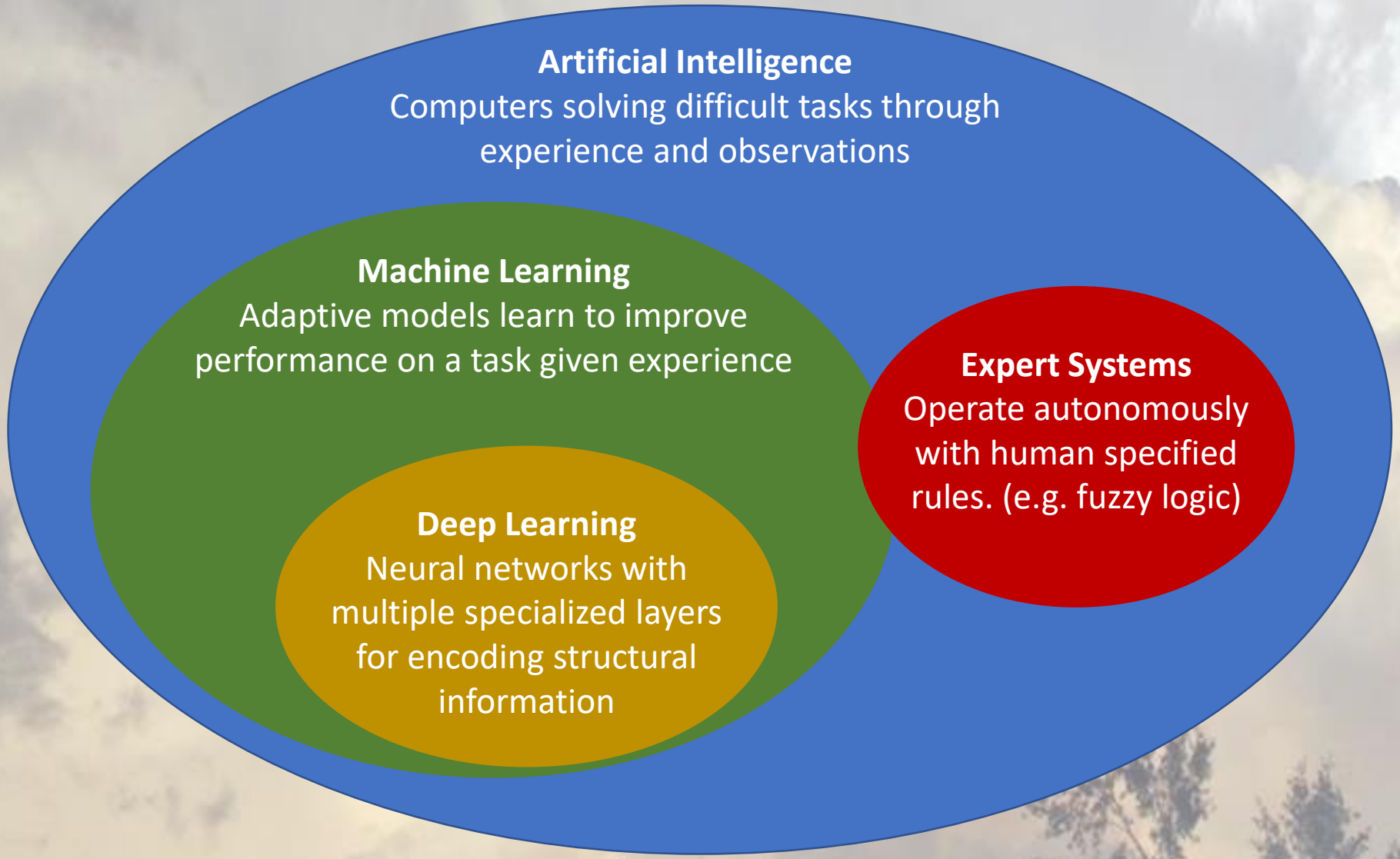


11:45 AM · May 9, 2020 · Twitter Web App

Outline

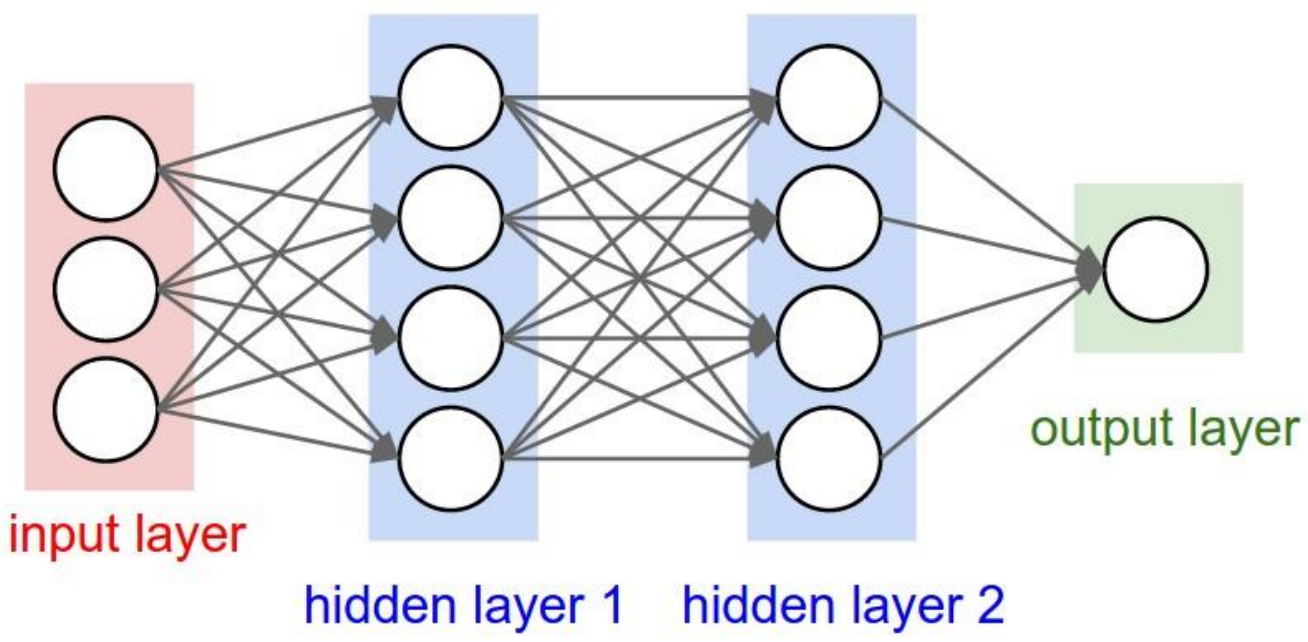
- Motivation for trustworthy AI
- Current work
  - Demonstrating ML can be used to improve prediction for multiple severe-weather hazards (this talk: hail and tornadoes)
  - Working with end-users to improve trust in ML predictions
  - Developing physically-based model interpretation and visualization techniques for environmental science
- Future work

What Is Deep Learning?

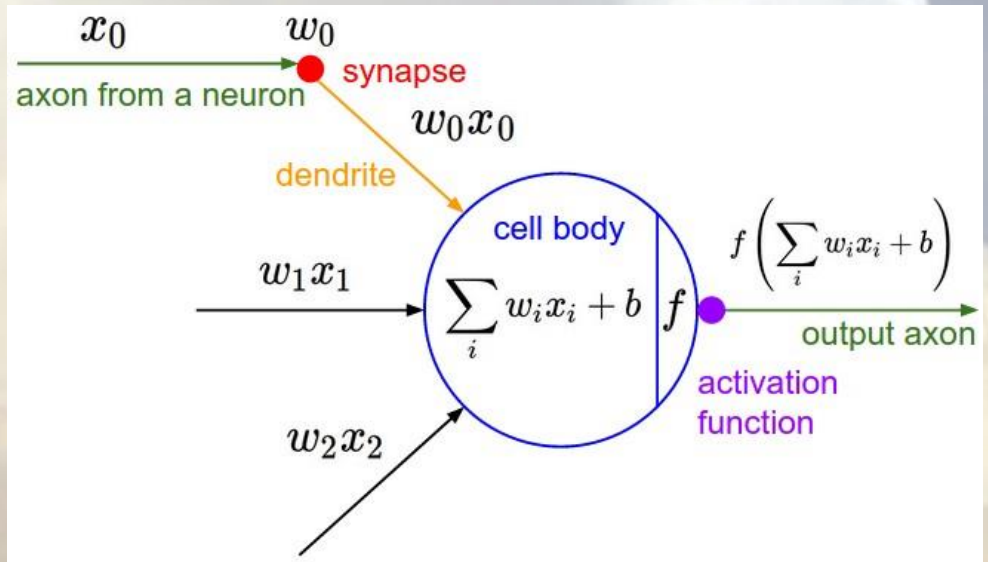


Neural Network Basics

Artificial Neural Network Structure



Perceptron (artificial neuron)



Training Procedure

1. Send batch of training examples through network
2. Calculate prediction error
3. Calculate error gradients back through layers and update weights
4. Repeat over all training examples until errors are satisfactory

Definitions

Batch: subset of training examples used to update weights

Epoch: One pass through all examples in training set

Images from http://cs231n.github.io/convolutional-networks/

Convolutional Neural Net

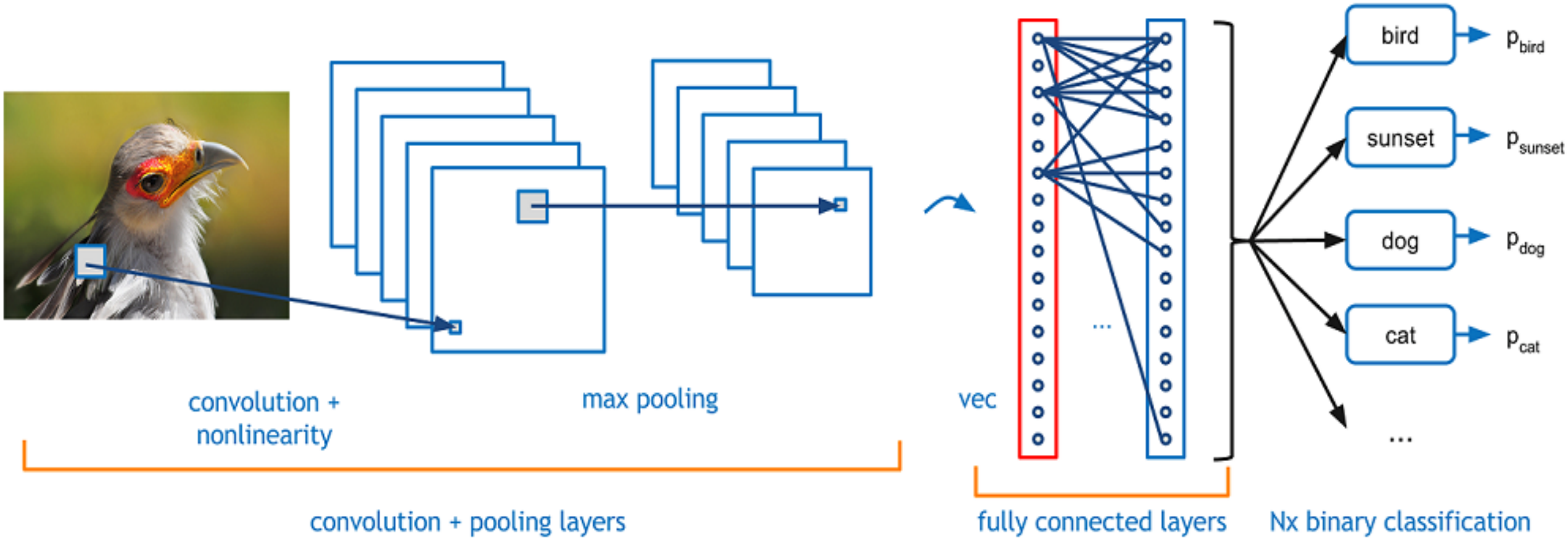
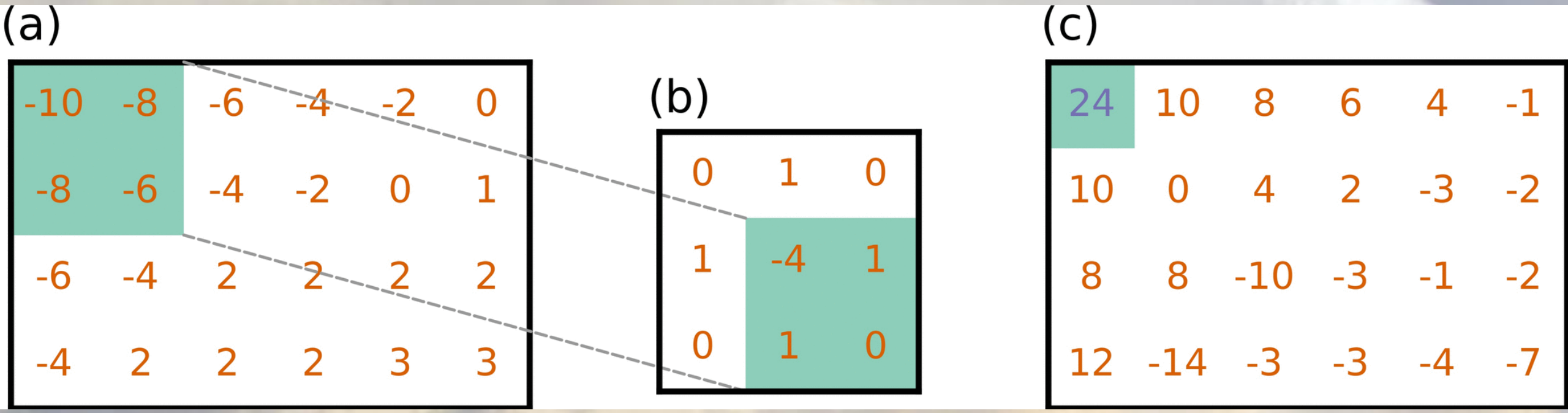
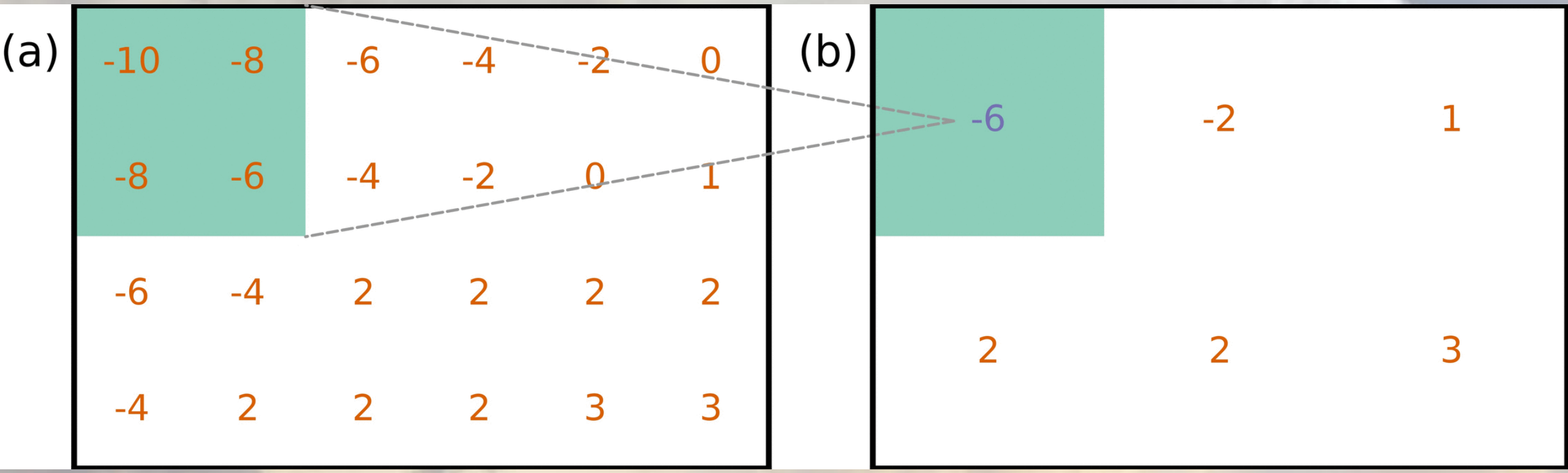


Image source: adeshpande3.github.io/A-Beginner%27s-Guide-To-Understanding-Convolutional-Neural-Networks/

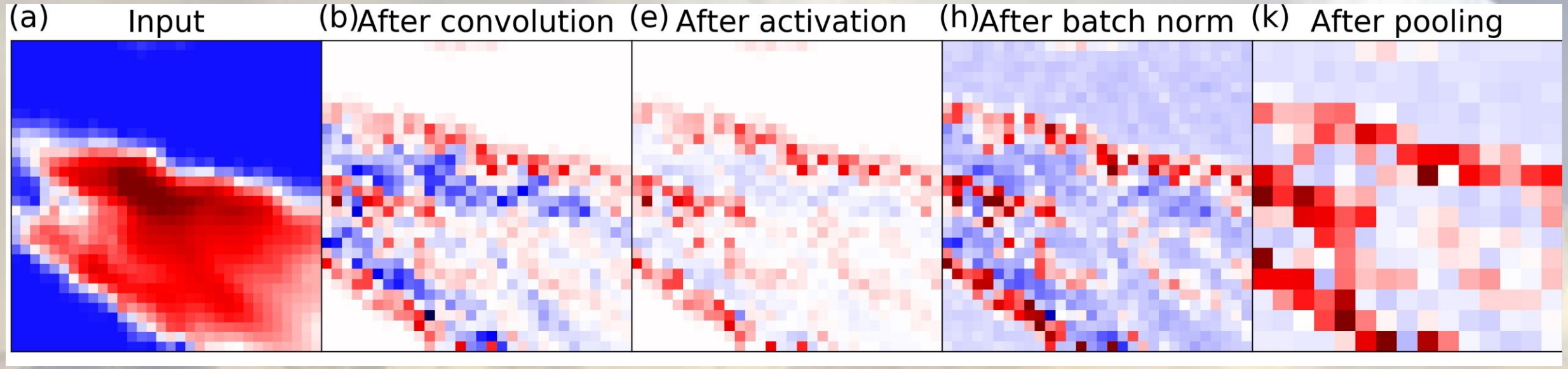
Convolutional Layers

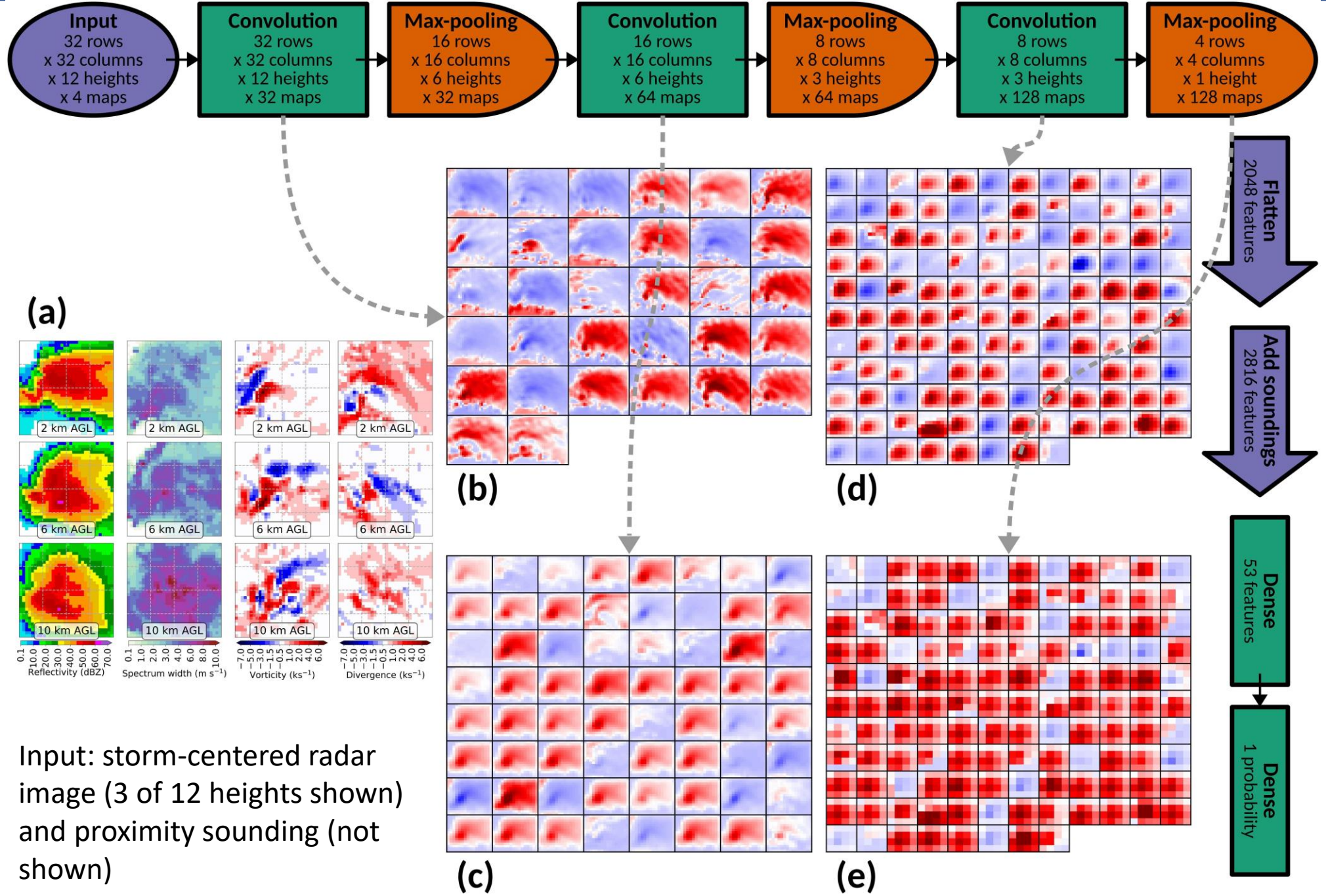


Pooling Layers



CNN Example


(a) Input
(b) After convolution
(e) After activation
(h) After batch norm
(k) After pooling



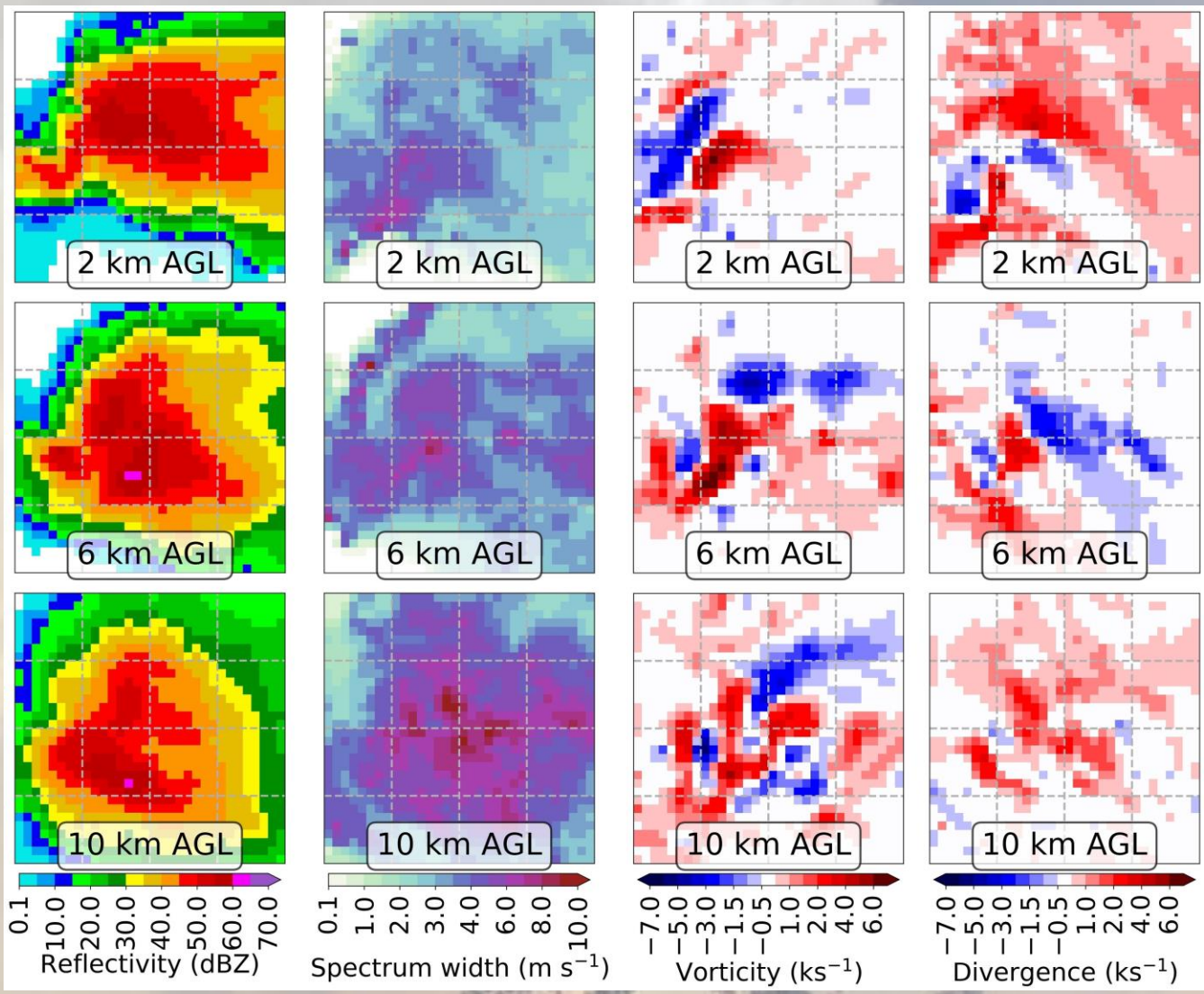
Input: storm-centered radar image (3 of 12 heights shown) and proximity sounding (not shown)

Output (next-hour tornado probability)

(b-e) Feature maps produced by conv and pooling layers

Training Data

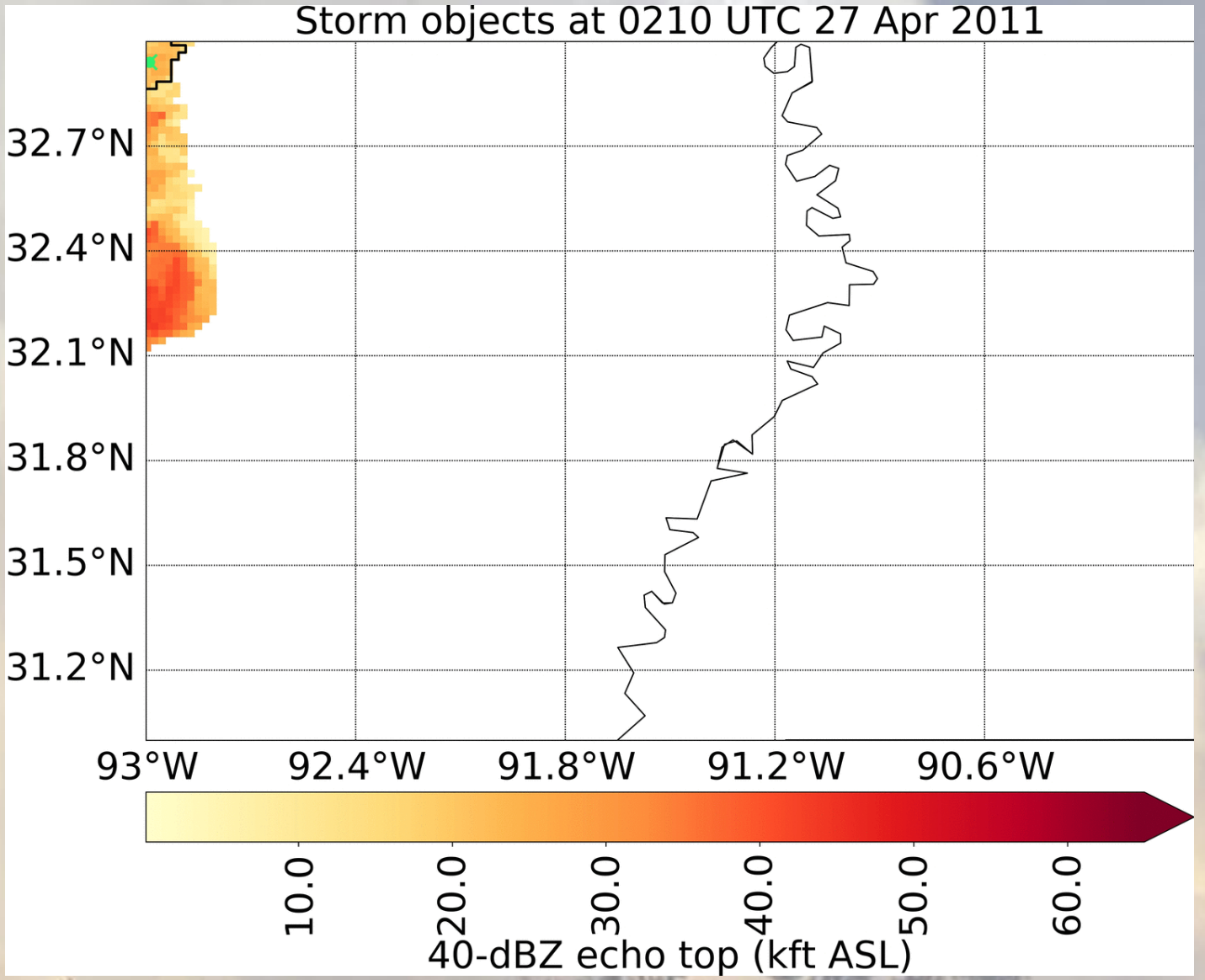
- Merged radar data from GridRad
  - http://gridrad.org
  - Resolution: 0.02° x 0.02° x 1 km
- GridRad fields for Tornado:
  - Reflectivity (ZH)
  - Velocity-spectrum width (increases with mean wind speed and turbulence)
  - Vorticity (rotational wind)
  - Divergence



Tornado Prediction: Input Data

- Before training CNNs, data must be pre-processed.
- **One CNN input = one storm object (one storm at one time).**
- Pre-processing steps are as follows:

1. **Outline storm cells at each time step**
2. **Track storm cells over time**
3. **Create storm-centered radar images**
   - One per storm object
   - On equidistant grid with storm motion towards the right



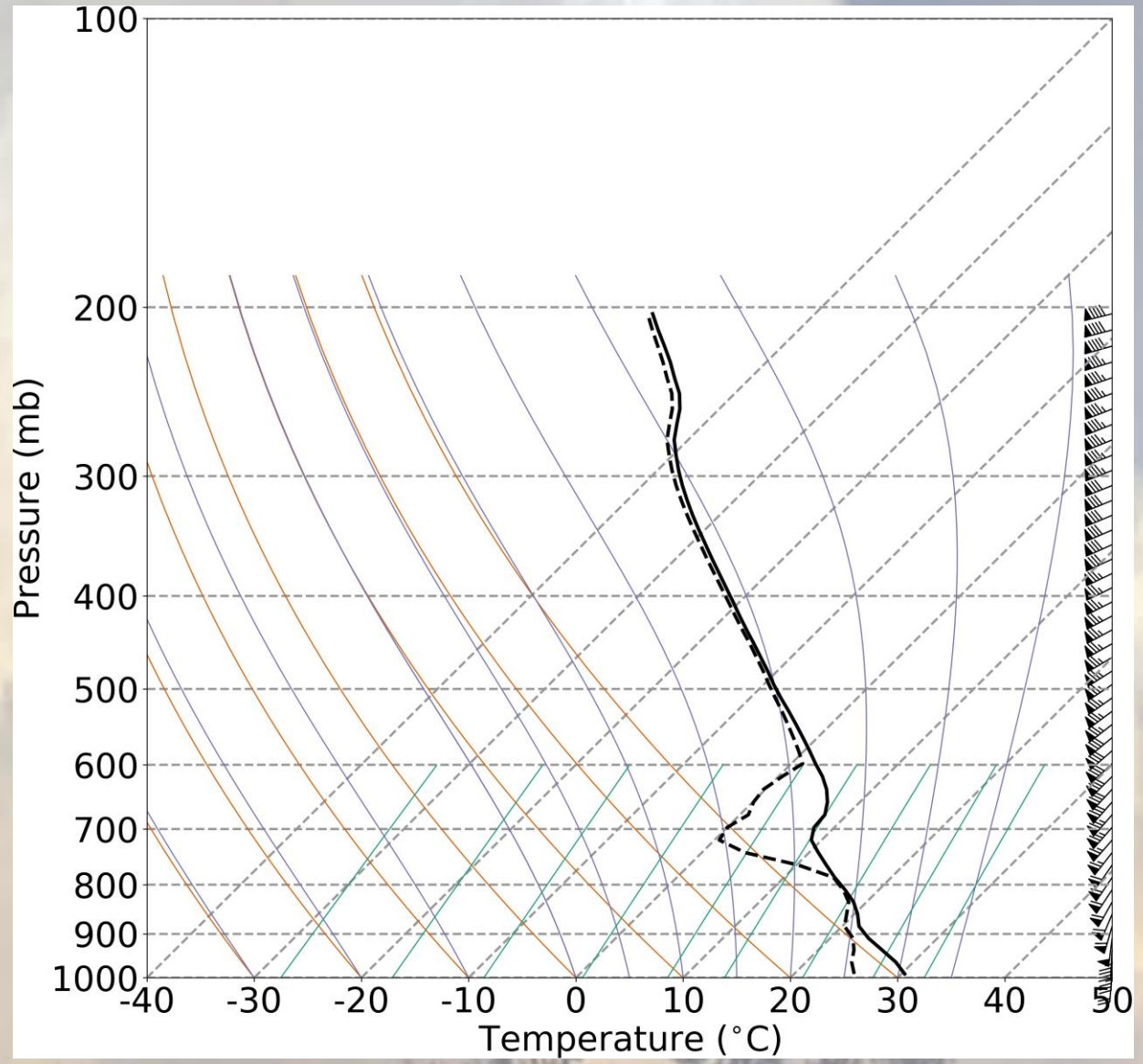
Tornado Prediction: Input Data

- Pre-processing steps are as follows:

4. **Create proximity soundings**
   - One per storm object
   - Represents near-storm environment

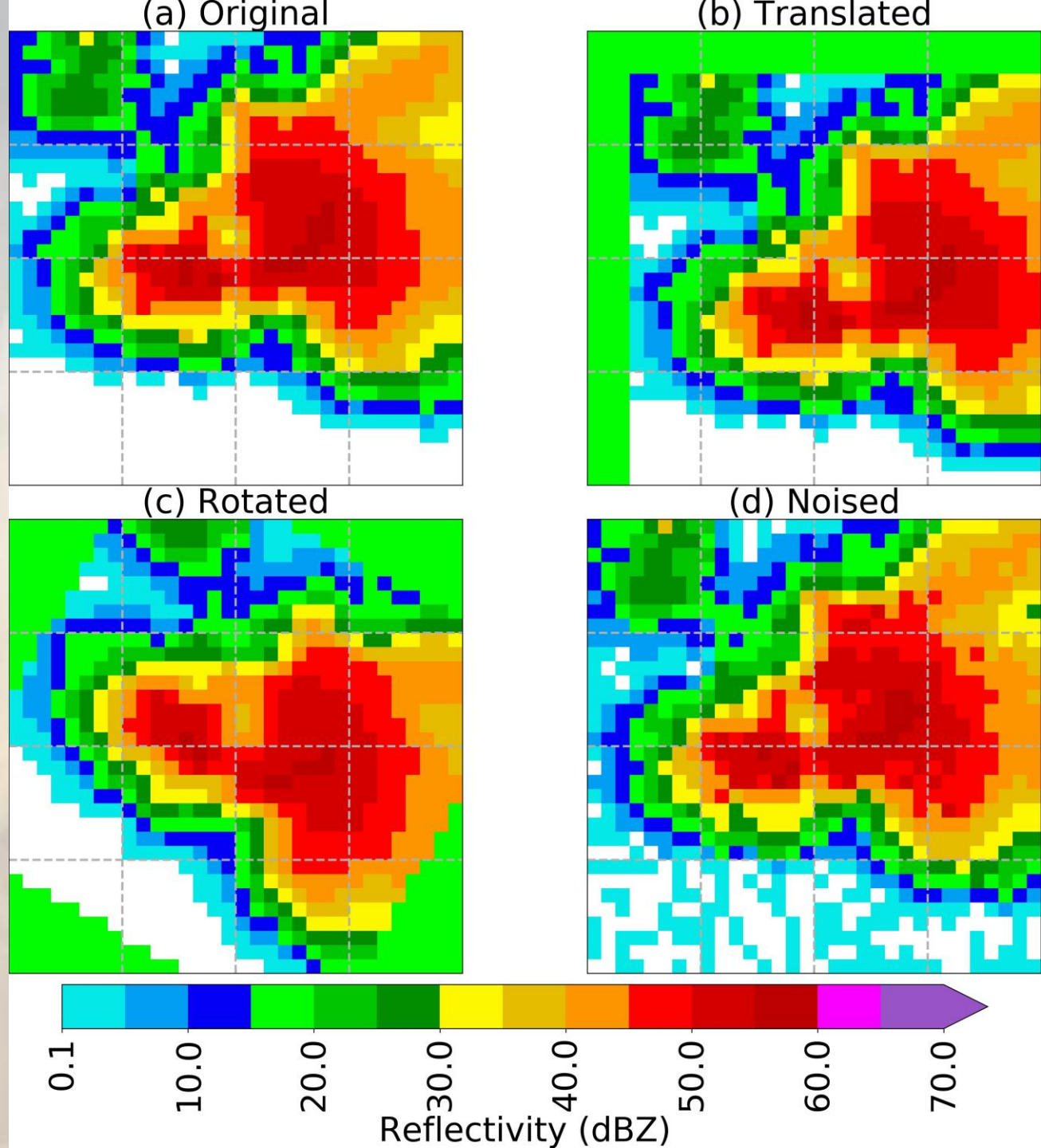
5. **Link tornado reports to storms**

6. **Create labels**
   - One per storm object
   - “Yes” if tornadic in next hour, else “no”

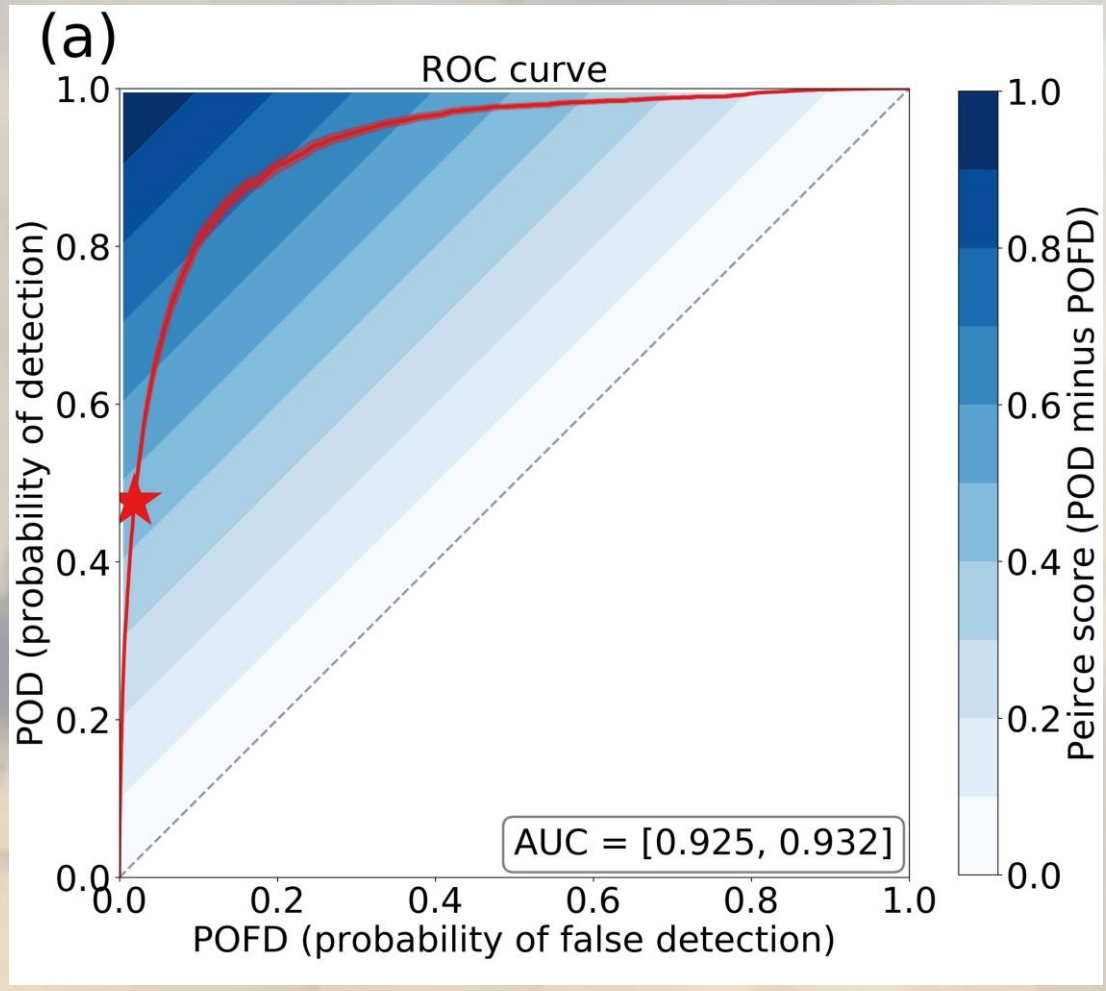


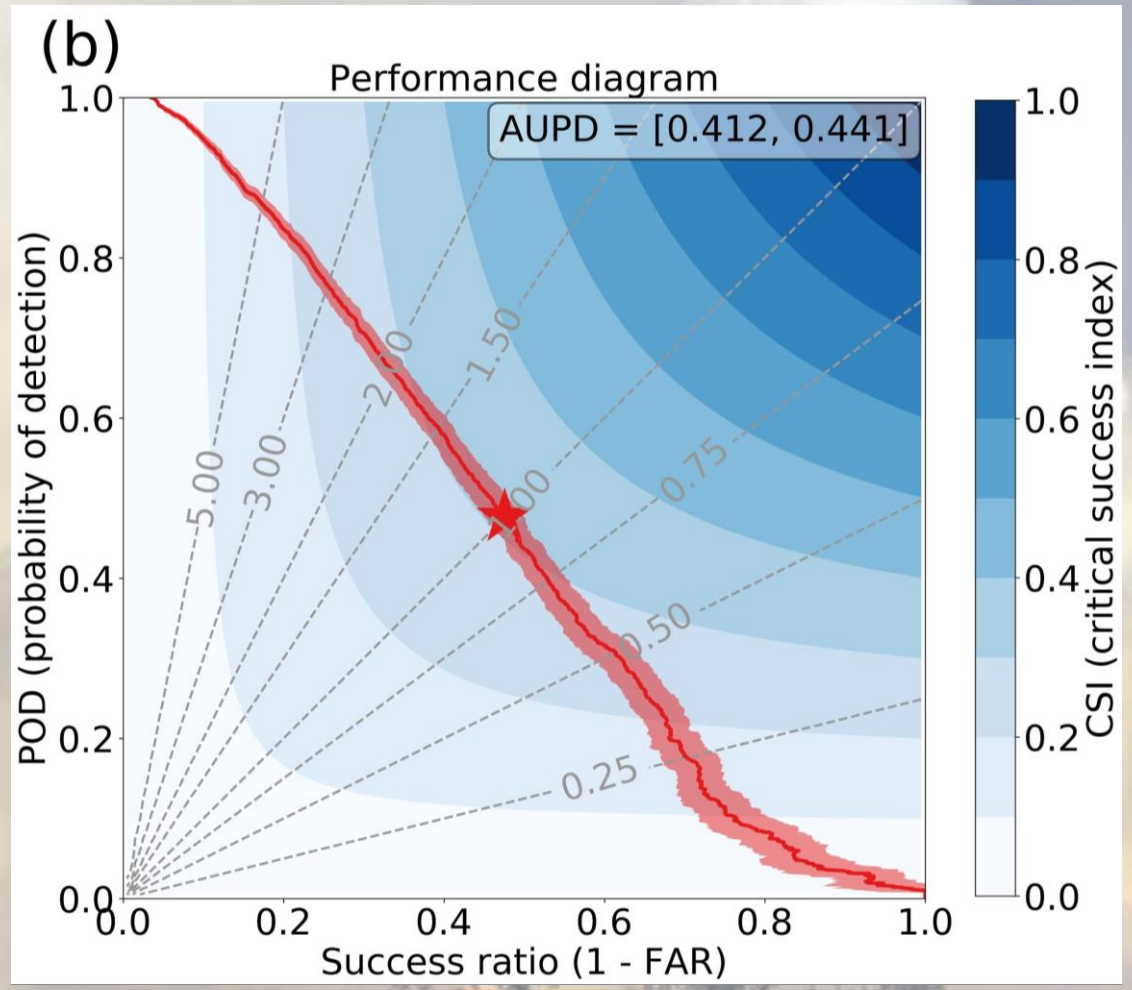
Training Data

- Training (2012-2014)
  - 40,903 total
  - 3575 tornadic
- Validation (2015-2018)
  - 96,868 total
  - 2884 tornadic
- Testing (2011)
  - 130,955 total
  - 4611 tornadic
- Data augmentation:
  - 16 different augmentations
  - Improves model robustness



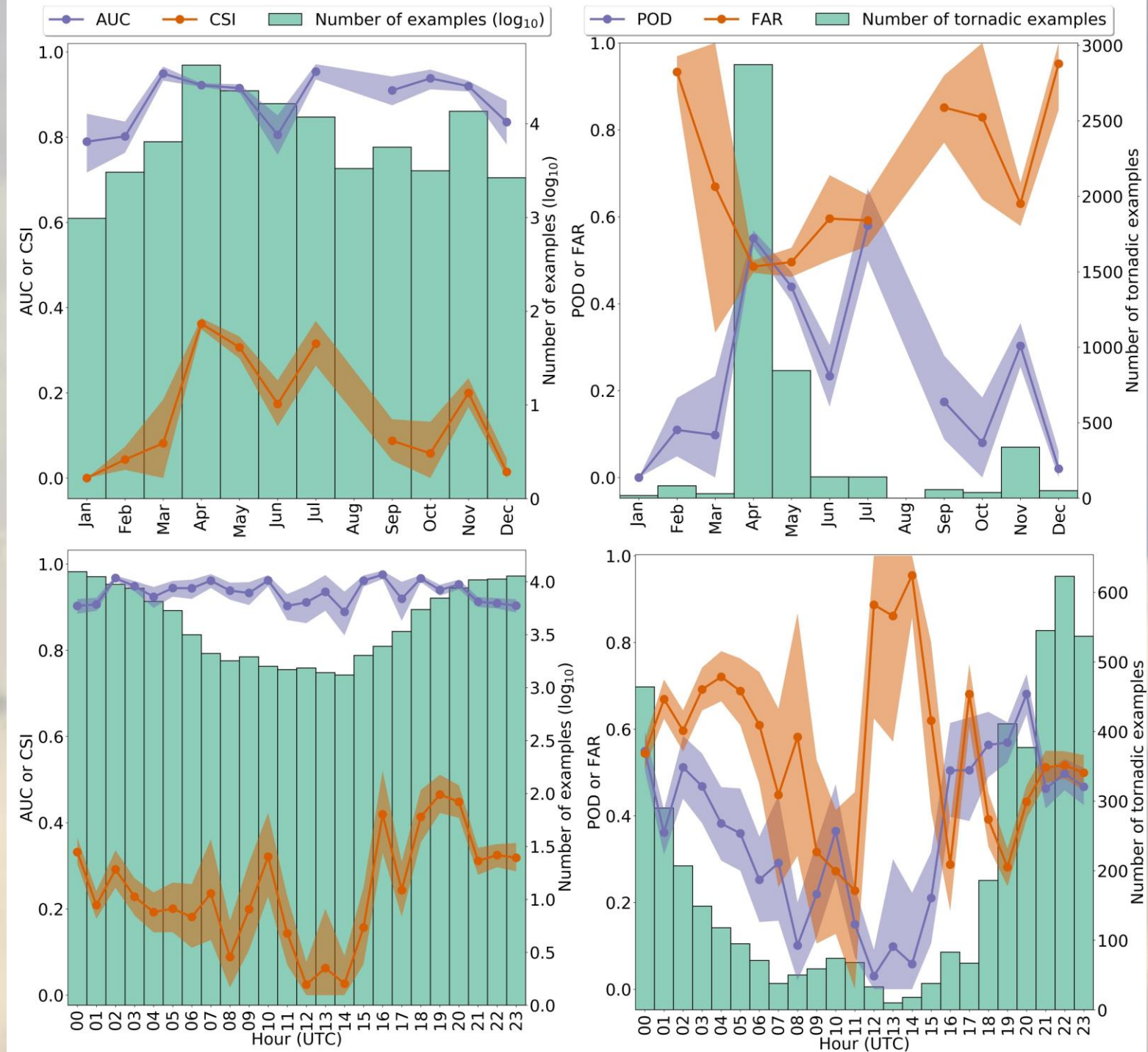
Objective Evaluation





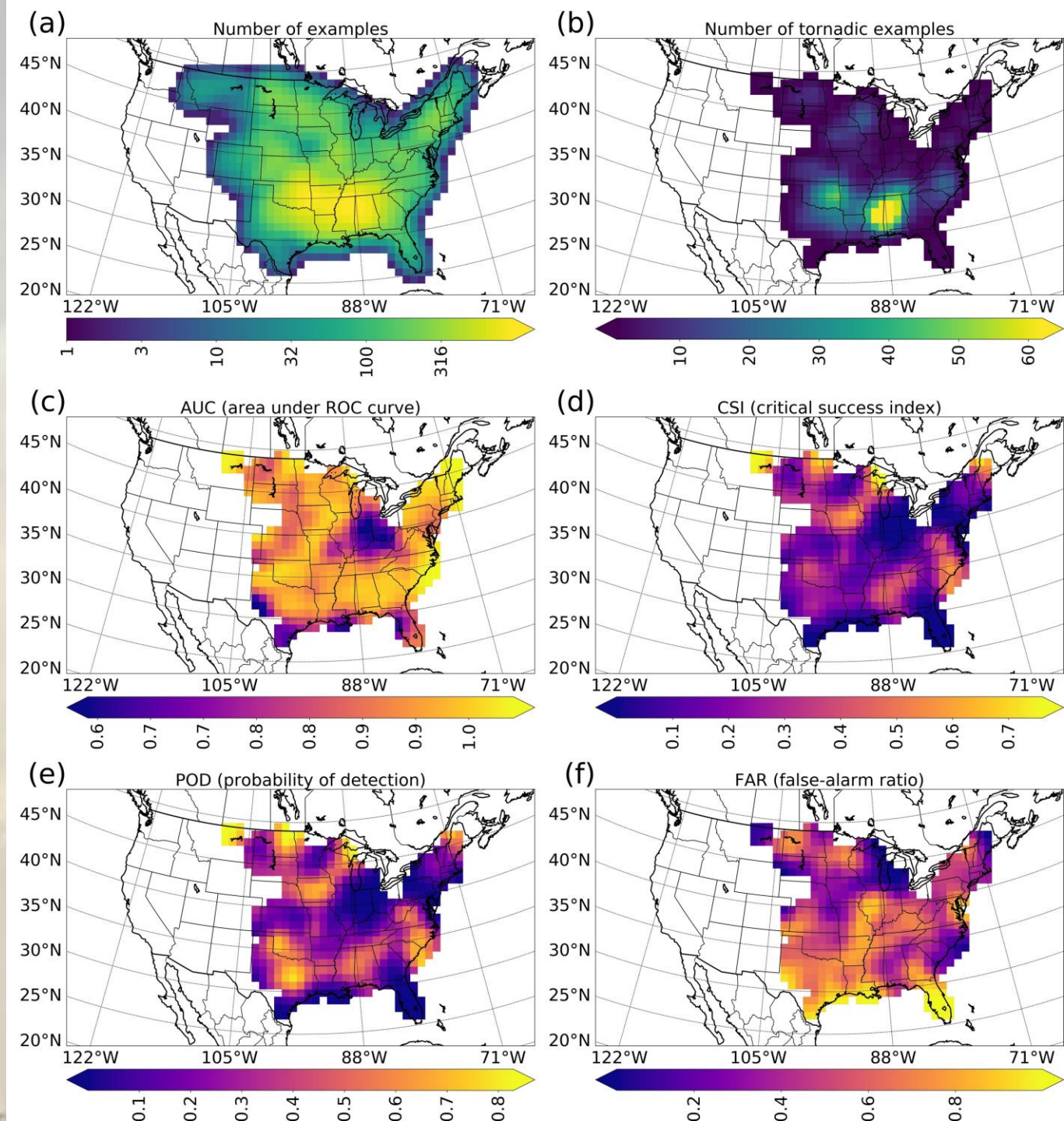
Hourly Evaluation

- AUC fairly consistent across months
- POD peaks in tornadic months
- FAR is lower in tornadic months



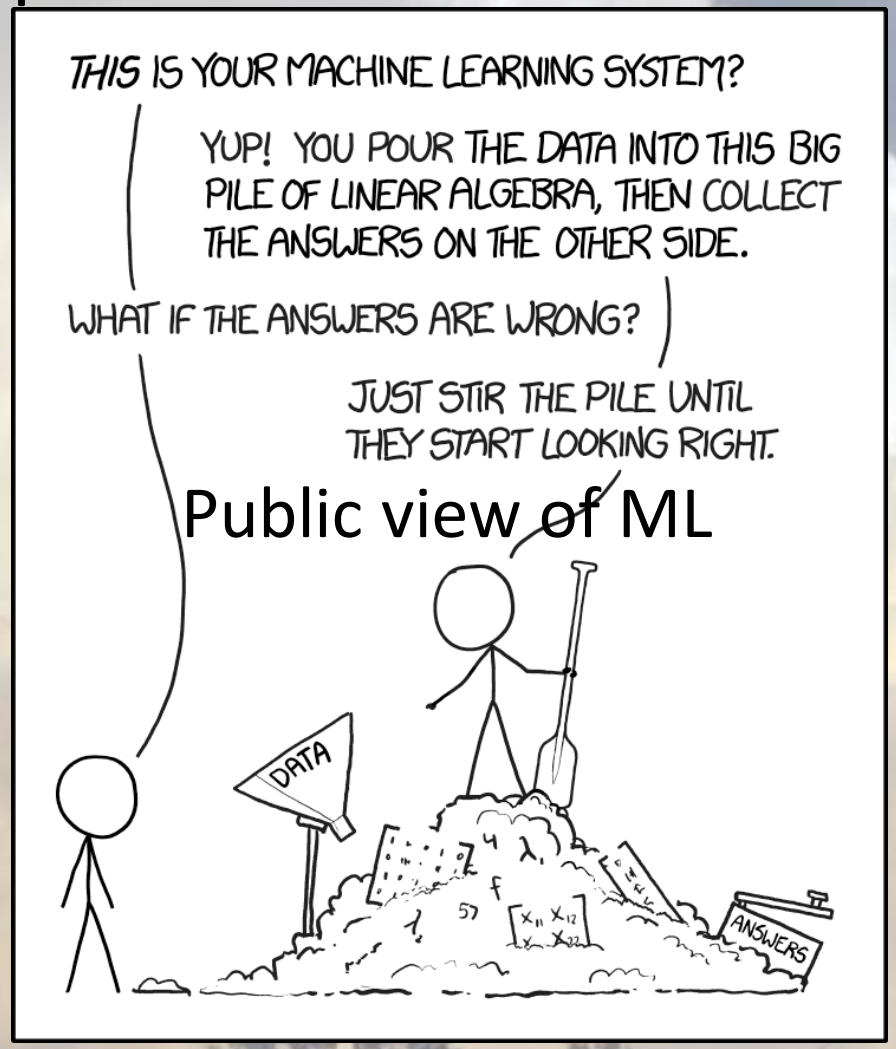
Spatial Evaluation

- Overall performance is best in areas with more tornadoes
- Need more examples in the harder areas



Trustworthy AI: Model Interpretation

- Our goal: demystify ML and deep learning models for environmental scientists by demonstrating benefits and drawbacks of model interpretation and visualization (MIV)



https://xkcd.com/1838/

Saliency Maps

- Saliency = gradient of model activation with respect to input value (Simonyan et al. 2014)

- Mathematically: $$\text{saliency} = \left.\frac{\partial a}{\partial x}\right|_{x=x_0}$$

  - a = activation of some model component
  - x = predictor (one variable at one pixel)
  - x_0 = actual value (in dataset example)

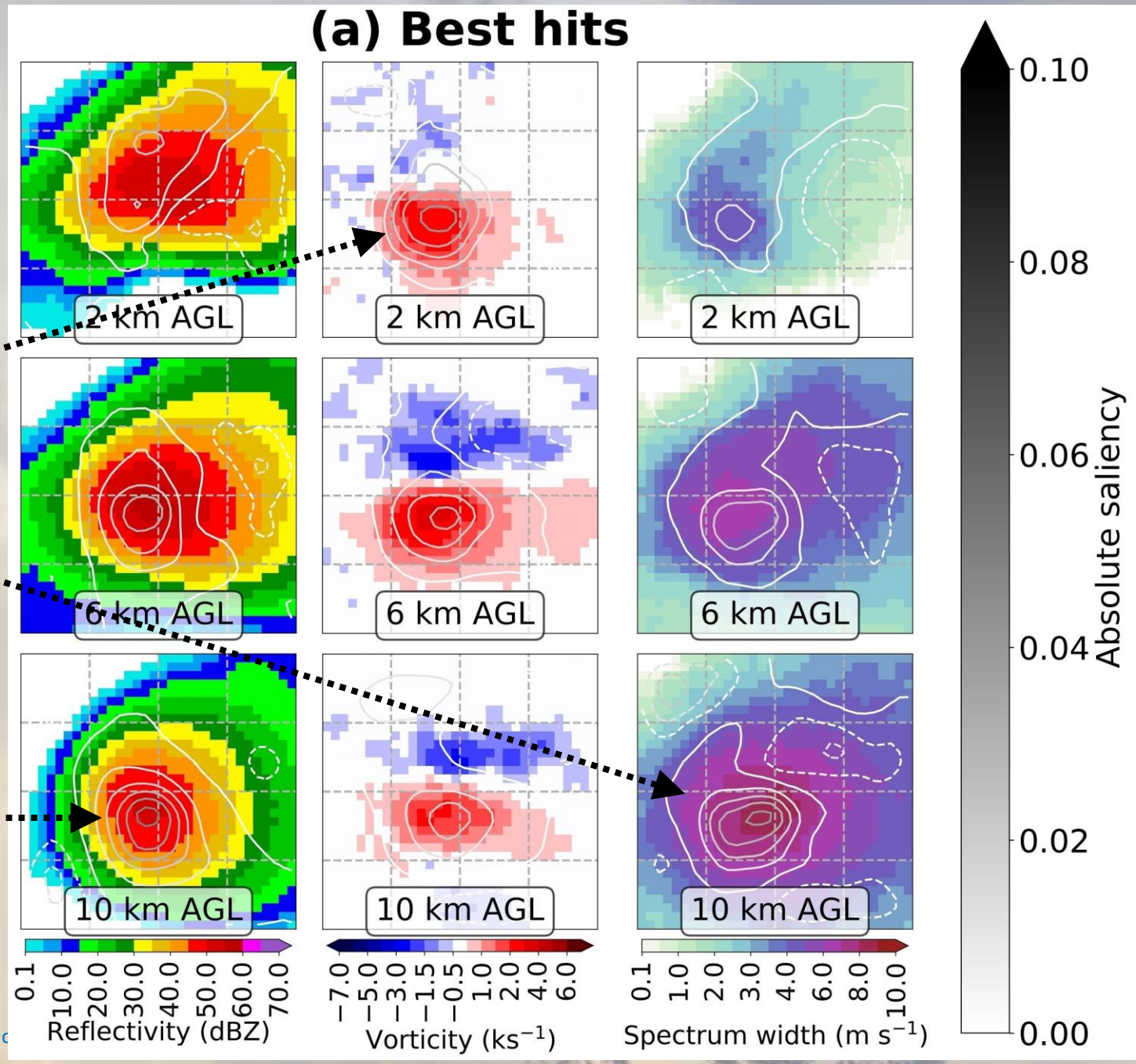
- Linear approximation to $\frac{\partial a}{\partial x}$ about $x = x_0$.
- In other words, saliency tells us how model reacts when x is perturbed from x_0

Saliency Maps on Tornado Data

$p_{tornado}$ increases with vorticity in mesocyclone, especially at lower levels

$p_{tornado}$ increases with spectrum width

$p_{tornado}$ increases with reflectivity in core, especially at upper levels



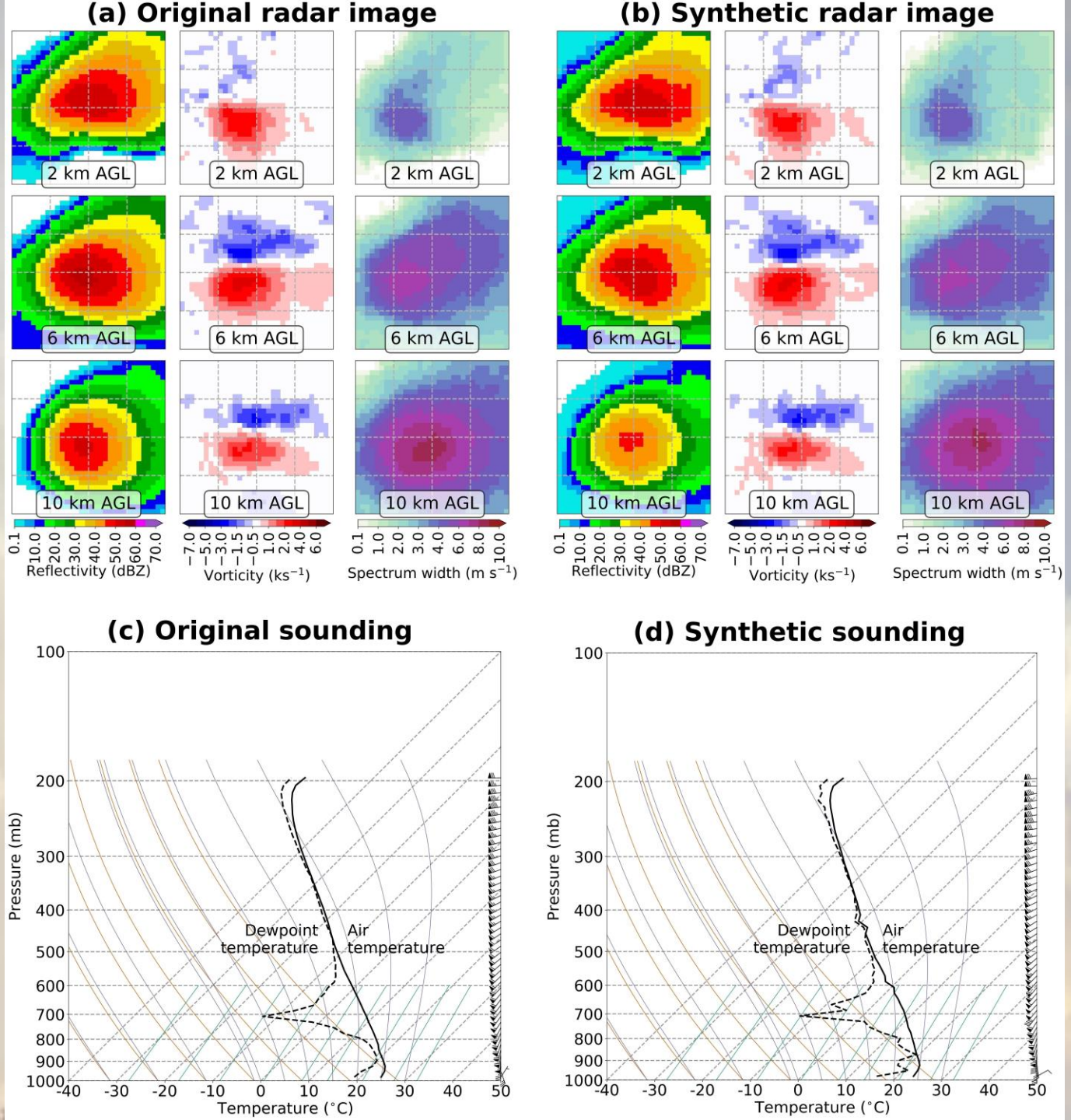
- **Average (PMM) saliency map for 100 best hits** (tornadic storms with average probability of 99.6%).

Ebert, E., 2001: "Ability of a poor man's ensemble to predict the probability and distribution o precipitation." *Monthly Weather Review*, **129 (10)**, 2461–2480.

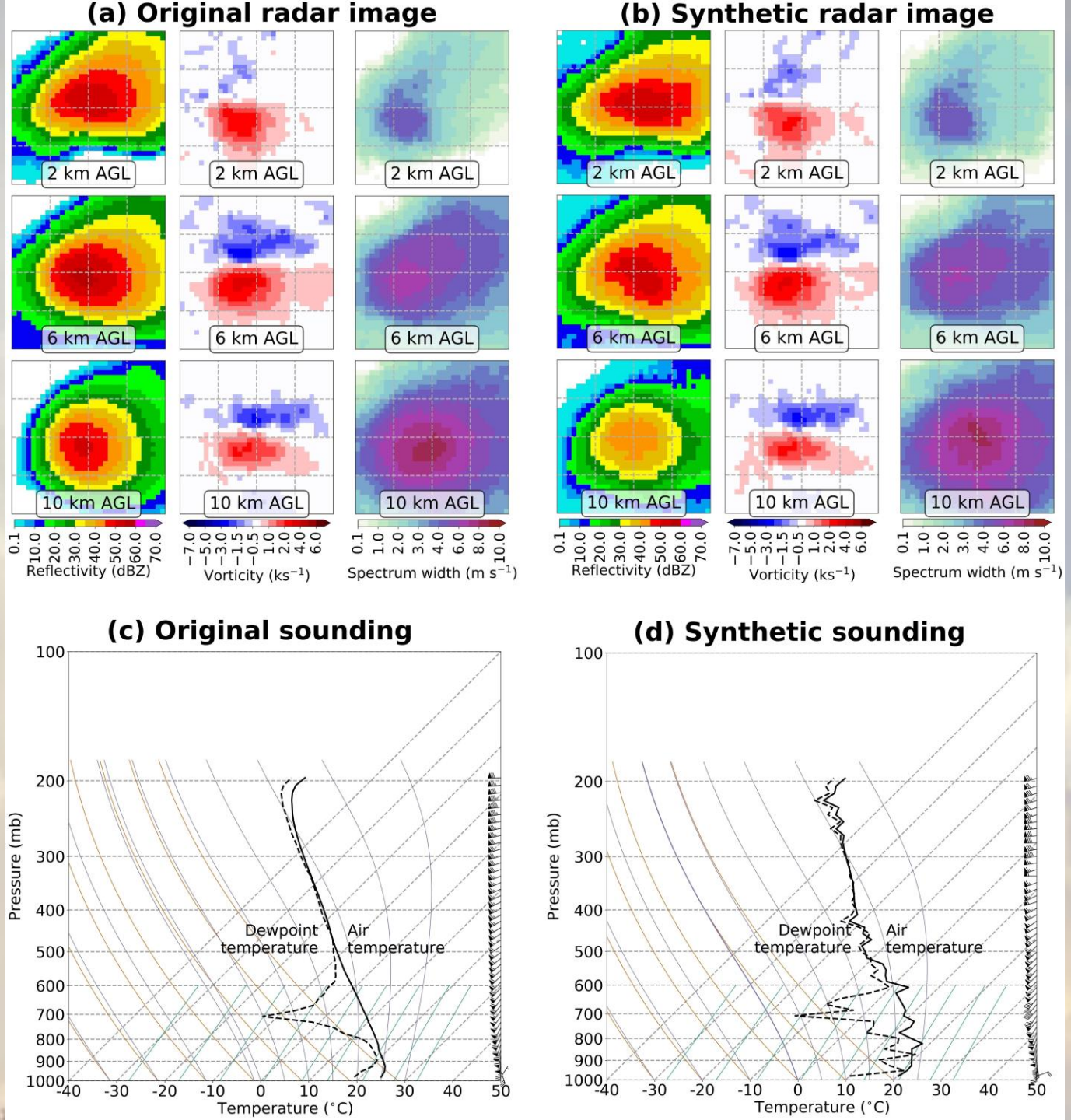
Backwards Optimization

- Also called “feature optimization” (Olah et al. 2017)
- Goal: create synthetic input example that maximizes activation of some model component
- “Some model component” might be:
  - Warm-front probability (activation of 2nd output neuron)
  - Cold-front probability (activation of 3rd output neuron)
  - Channels in final convolution layer (just before fully connected layer)
- Procedure involves gradient descent, which requires initial seed
- Initial seed might be:
  - Uniform image (e.g., all zeros)
  - Random image
  - **Dataset example**

- **Right: we use BWO to decrease tornado probability for best hits.**
- On average for the 100 storms, decreases probability from 99.2% to 6.9%.
- **Effects of BWO are small, except:**
  - Decreases depth of reflectivity core (see 10 km AGL)
  - Removes moisture near surface (see dewpoint in sounding)
  - Decreases low-level wind speed and thus shear (see sounding)
- **However, synthetic sounding looks a bit unrealistic** (has the "jaggies").
- **This looks much worse for BWO without physical constraints (next slide).**
- Nonetheless, more work needed if we want to use ML to create realistic weather data.



- **Right: we use BWO to decrease tornado probability for best hits.**
- On average for the 100 storms, decreases probability from 99.2% to 6.9%.
- **Effects of BWO are small, except:**
  - Decreases depth of reflectivity core (see 10 km AGL)
  - Removes moisture near surface (see dewpoint in sounding)
  - Decreases low-level wind speed and thus shear (see sounding)
- **However, synthetic sounding looks a bit unrealistic** (has the "jaggies").
- **This looks much worse for BWO without physical constraints (next slide).**
- Nonetheless, more work needed if we want to use ML to create realistic weather data.



Acknowledgments

- Some of the computing for this project was performed at the OU Supercomputing Center for Education & Research (OSCER) at the University of Oklahoma (OU).
- This material is based upon work supported by the National Science Foundation under Grant Numbers EF-1340921, DGE-1545261, AGS-1802627, ICER-2019758 and NOAA JTTI Grant No. NA16OAR4590239.
- Funding was provided by NOAA/Office of Oceanic and Atmospheric Research under NOAA-University of Oklahoma Cooperative Agreement #NA16OAR4320115, U.S. Department of Commerce.

References

- Burke, A.; Snook, N.; Gagne II, D.J.; McCorkle, S.; and McGovern, A. (2020) *Calibration of Machine Learning-Based Probabilistic Hail Predictions for Operational Forecasting*. Weather and Forecasting, 35, 149-168
- McGovern, A., D.J. Gagne II, R. Lagerquist, K. Elmore, and G.E. Jergensen, (2019) *Making the black box more transparent: Understanding the physical implications of machine learning*. Bulletin of the American Meteorological Society, Volume 100, Number 11, Pages 2175-2199
- Lagerquist, R. “Using deep learning to improve prediction and understanding of weather phenomena”, PhD Dissertation, May 2020, University of Oklahoma School of Meteorology
- Lagerquist, Ryan; McGovern, Amy; Homeyer, Cameron; Gagne II, David John; Smith, Travis. (2020) *Deep Learning on Three-dimensional Multiscale Data for Next-hour Tornado Prediction*. To appear in Monthly Weather Review.
- McGovern, A.; R. Lagerquist; and D. Gagne (2020) *Using machine learning and model interpretation and visualization techniques to gain physical insights in atmospheric science*. Proceedings of the International Conference on Learning Representations.
- Gagne II, David John; McGovern, Amy; Haupt, Sue Ellen; Sobash, Ryan; Williams, John K. and Xue, Ming. (2017) Storm-Based Probabilistic Hail Forecasting with Machine Learning Applied to Convection-Allowing Ensembles. Weather and Forecasting, 32, 1819-1840.