Using Deep Learning to Improve Prediction and Understanding of High-impact Weather

- **Ryan Lagerquist**
  (ryan.lagerquist@ou.edu, @ralager_Wx)

- Dissertation defence, OU School of Meteorology

- Committee: Amy McGovern (chair), Jason Furtado, Jeff Basara, Michael Richman, Andrew Fagg, Justin Metcalf

THE UNIVERSITY OF OKLAHOMA SCHOOL OF METEOROLOGY

CIMMS

Cooperative Institute for Mesoscale Meteorological Studies

Outline

- **I have developed and tested deep-learning models for tornado prediction and front detection.**

- **Contributions to tornado prediction:**
  - My model is competitive with a current operational ML model, promising for future use
  - I use novel interpretation methods to understand physical relationships learned by models

- **Contributions to front detection:**
  - My model automates front detection over large area (North America and surrounding oceans)
  - I create and analyze 40-year climatology
  - I compare with the few previous climos that investigate ENSO influence and long-term change

- I demonstrate that **deep learning can improve prediction and understanding** of diverse high-impact weather phenomena.

Tornado Prediction: Intro

- **Skill of National Weather Service (NWS) tornado warnings has stagnated in the last decade** (Brooks and Correia 2018).

- Meanwhile, amount of data/tools available to forecasters has exploded.
  - Dual-polarization radar
  - High-resolution satellite
  - Convection-allowing models
  - …etc.

- Problem: most of these data/tools do not explicitly resolve tornadoes.

- This leaves forecasters to mentally post-process big data into tornado predictions/warnings, leading to cognitive overload (Wilson *et al.* 2017).

- **Post-processing can be automated by deep learning, which excels with big data.**

Joplin tornado damage from:
https://en.wikipedia.org/wiki/2011_Joplin_tornado#/media/File:Joplin_2011_tornado_damage.jpg

Tornado Prediction: Intro

- **I use convolutional neural nets (CNN), a deep-learning method designed to learn from gridded data.**
- In traditional ML, gridded data must be converted to scalar statistics before training model.
- This destroys spatial info that could be exploited by the model.
- CNNs see the full grid, which generally improves skill.
- **Specifically, I use CNN to forecast probability that a given storm will be tornadic in the next hour.**

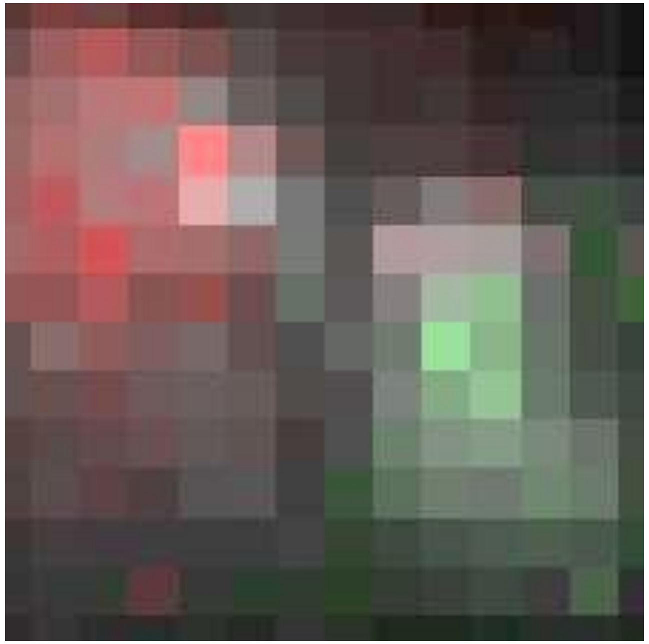


Image source: Olah *et al.* (2017)

Tornado Prediction: Input Data

- **I use the following datasets:**
  - Radar images from MYRORSS and GridRad
  - Proximity soundings from RAP weather model
  - Tornado reports

- **Details:**
  - MYRORSS = Multi-year Reanalysis of Remotely Sensed Storms (Ortega *et al.* 2012)
  - GridRad = Gridded NEXRAD WSR-88D Radar (Homeyer and Bowman 2017)
  - RAP = Rapid Refresh (Benjamin *et al.* 2016)
  - Tornado reports from Severe Weather Data Inventory (SWDI)

Tornado Prediction: Input Data

- **MYRORSS and GridRad are multi-radar datasets**, created by merging all WSR-88D radars in the continental United States.
- **Both datasets have 5-minute time steps.**
- **Datasets overlap for one year (2011), which is the testing year.**
- MYRORSS:
  - Training: 2005-08
  - Validation: 2009-10
- GridRad:
  - Training: 2012-14
  - Validation: 2015-18

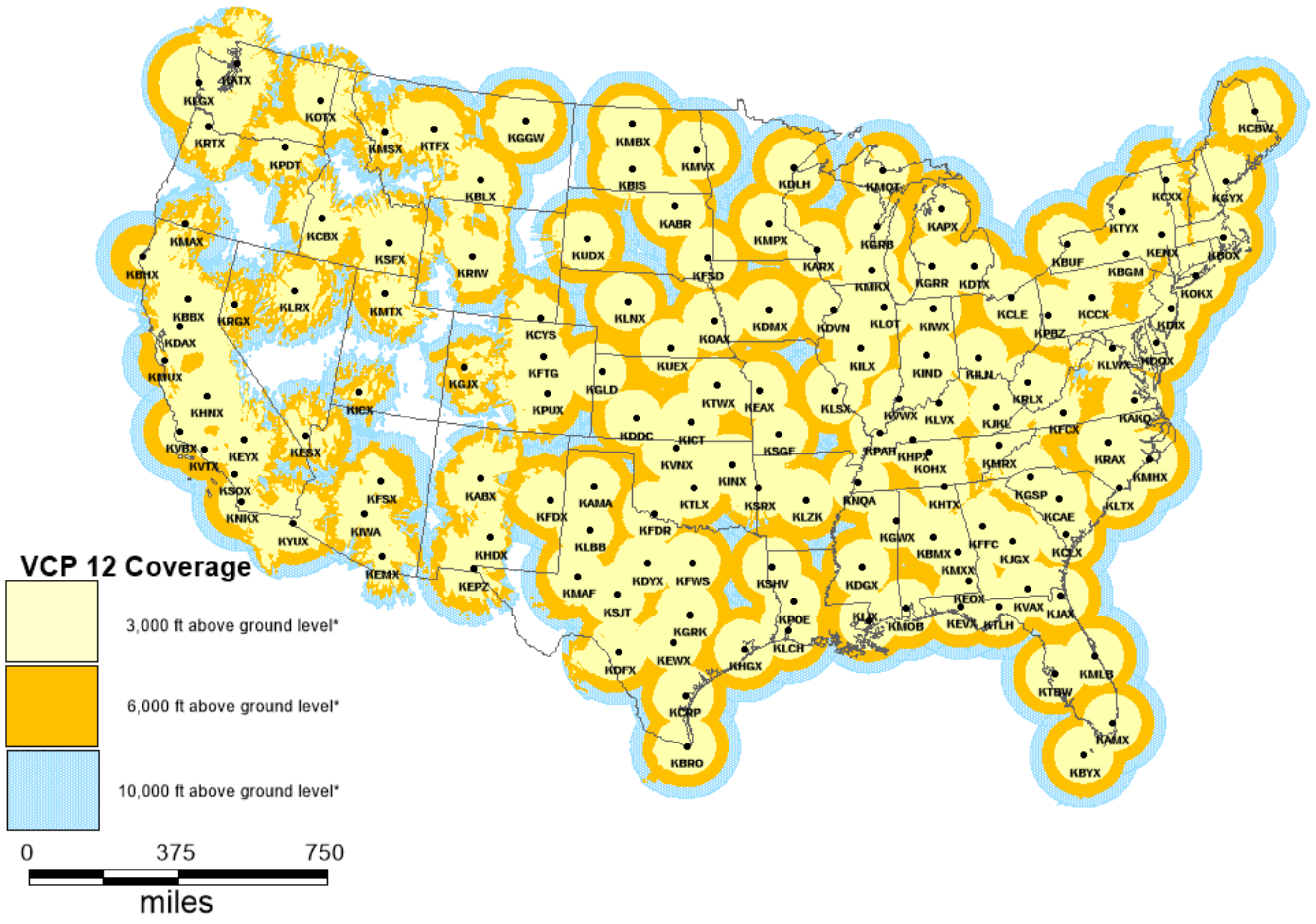
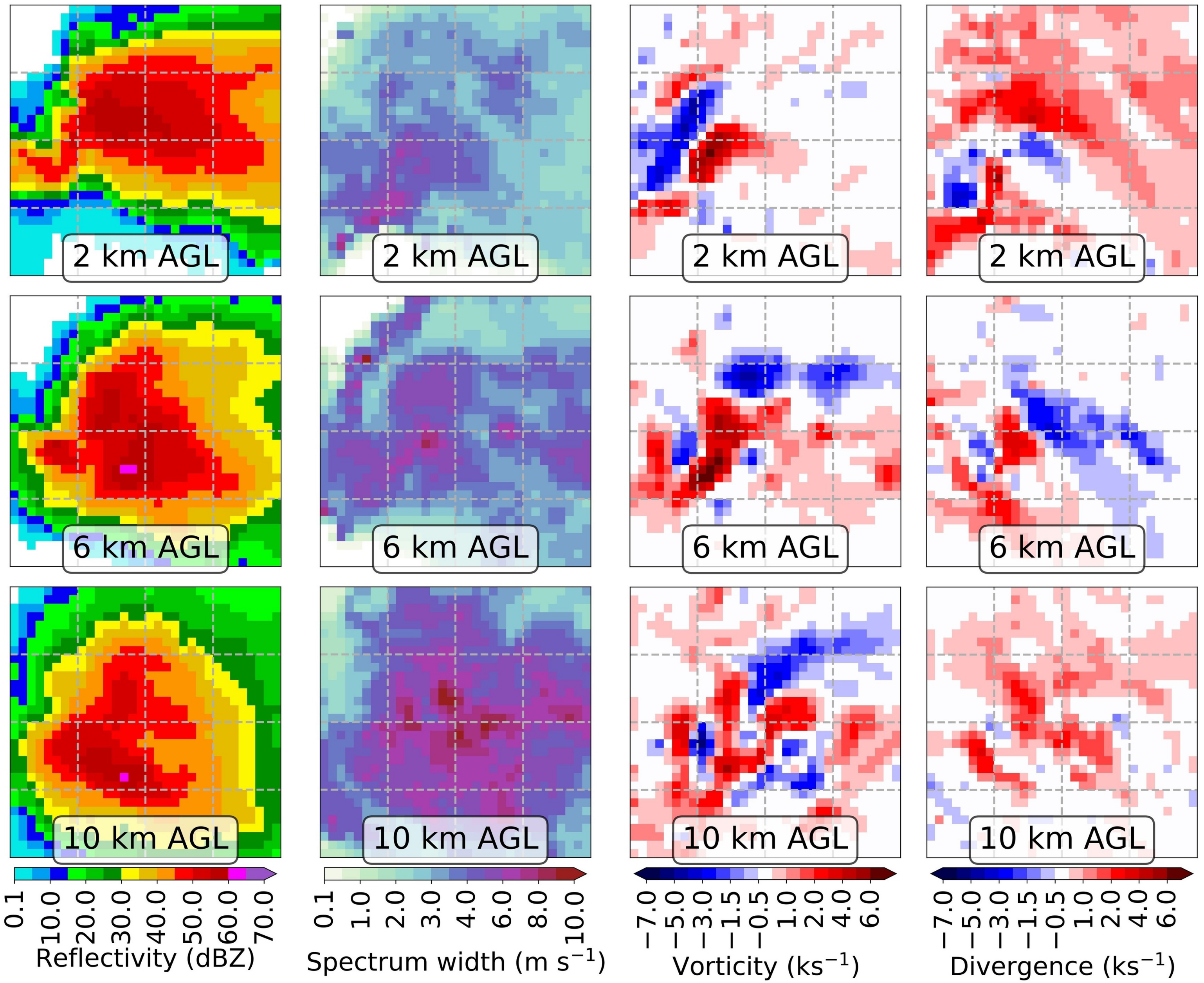


Image source: https://www.roc.noaa.gov/WSR88D/Maps.aspx

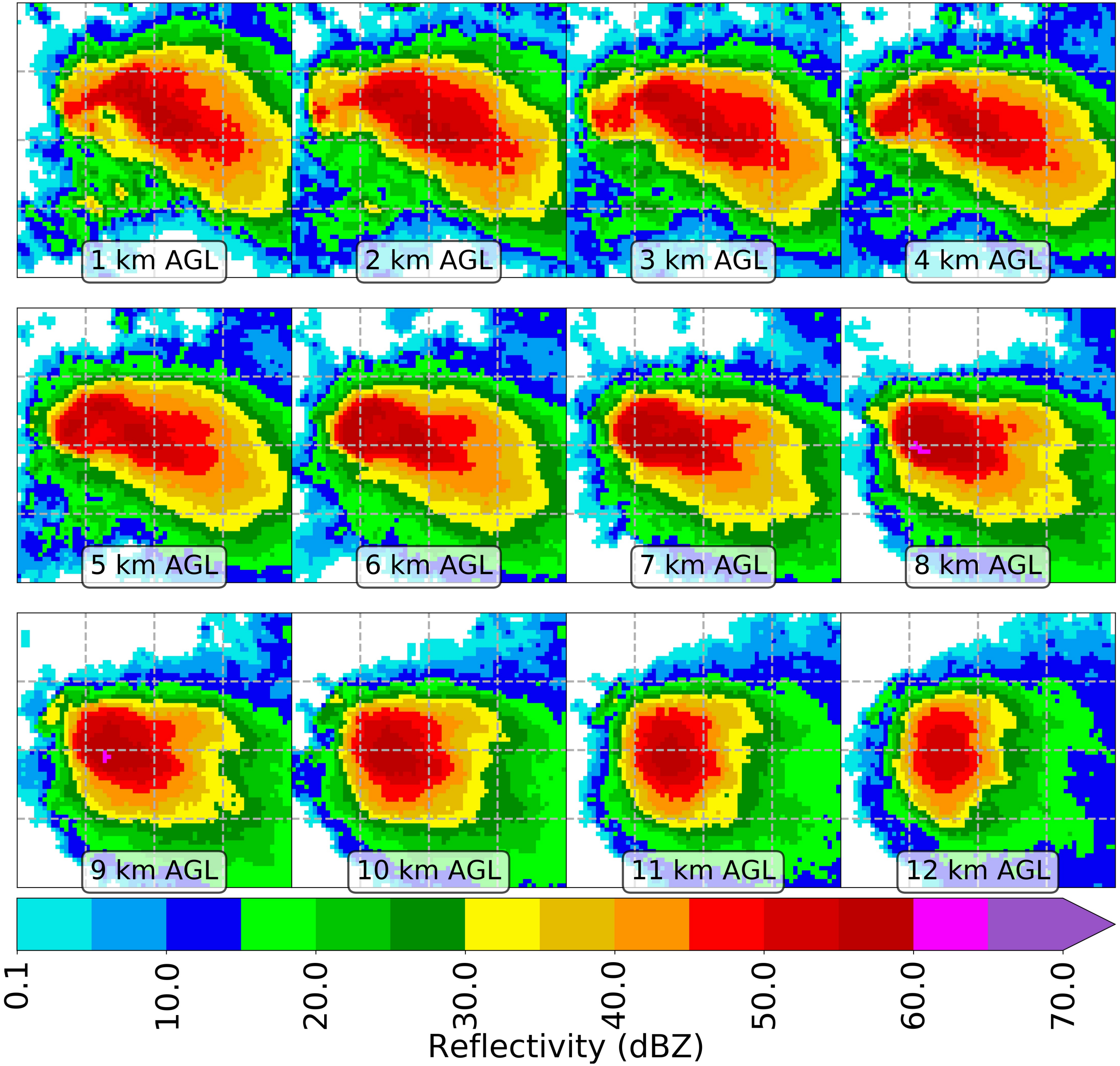
Tornado Prediction: Input Data

- GridRad has 0.0208° horizontal spacing (~2 km) and contains 3-D fields of the following variables:
  - Reflectivity
  - Velocity-spectrum width (increases with mean wind speed and turbulence)
  - Vorticity (rotational wind)
  - Divergence



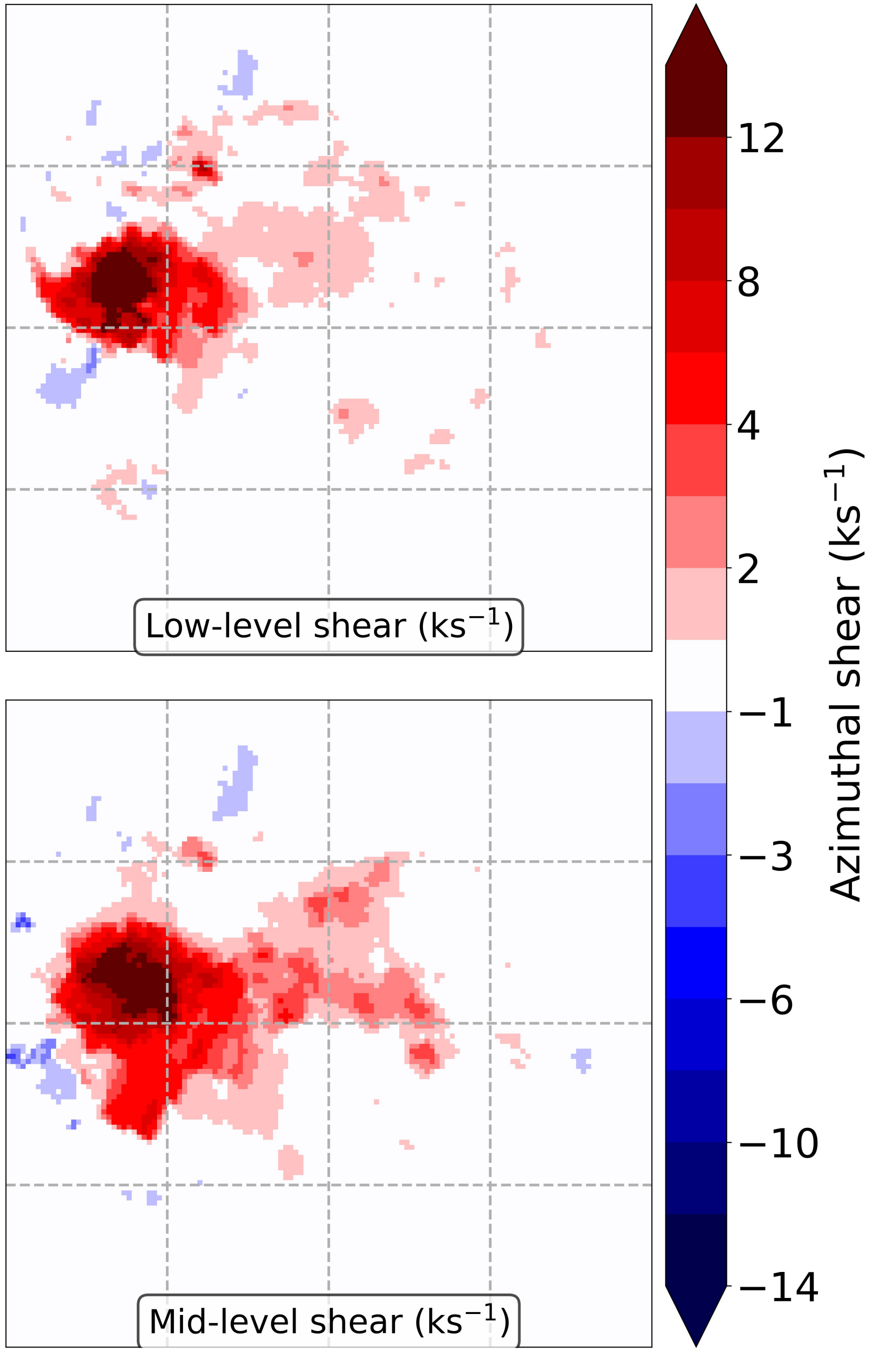
Tornado Prediction: Input Data

- MYRORSS contains the following variables:
  - Reflectivity
    (0.01° horizontal spacing, or ~1 km)
  - Azimuthal shear
    (0.005° horizontal spacing, or ~0.5 km)



Tornado Prediction: Input Data

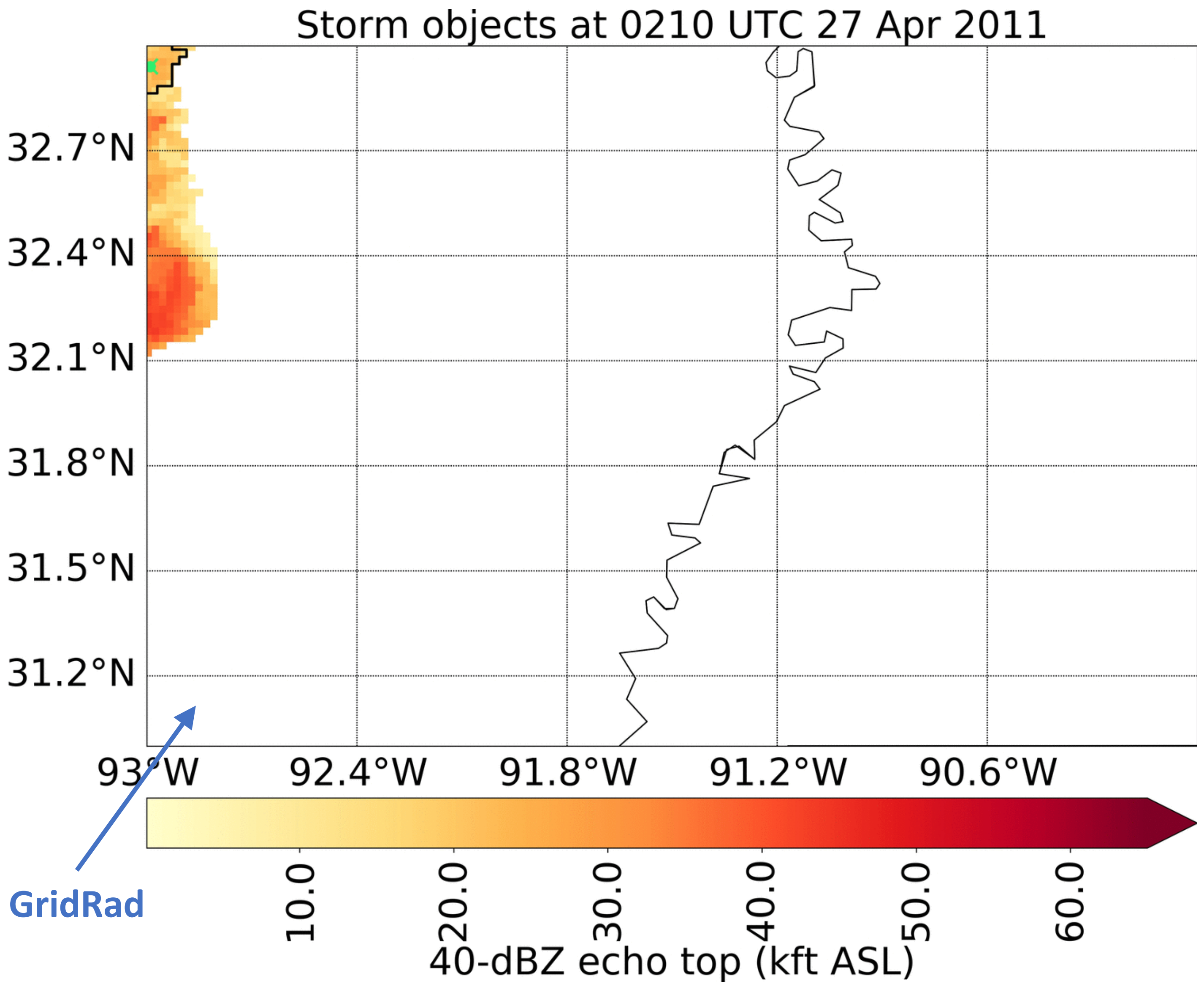
- MYRORSS contains the following variables:
  - Reflectivity
    (0.01° horizontal spacing, or ~1 km)
  - Azimuthal shear
    (0.005° horizontal spacing, or ~0.5 km)
- **Azimuthal shear = 0.5 * vorticity**
- "Low-level" = max from 0-2 km above ground (AGL)
- "Mid-level" = max from 3-6 km AGL



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

- Before training CNNs, data must be pre-processed.

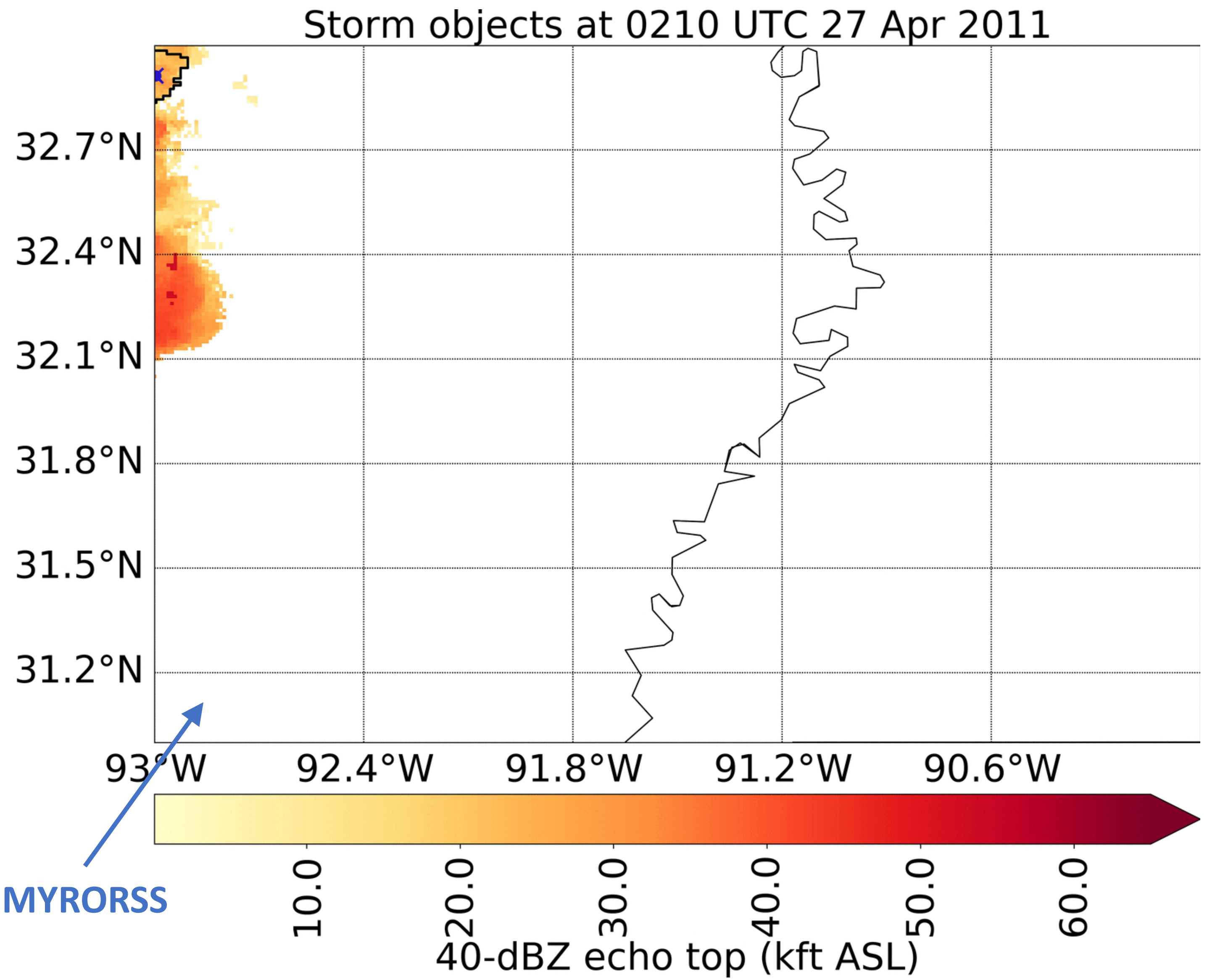
- **One CNN input = one storm object (one storm at one time).**

- 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



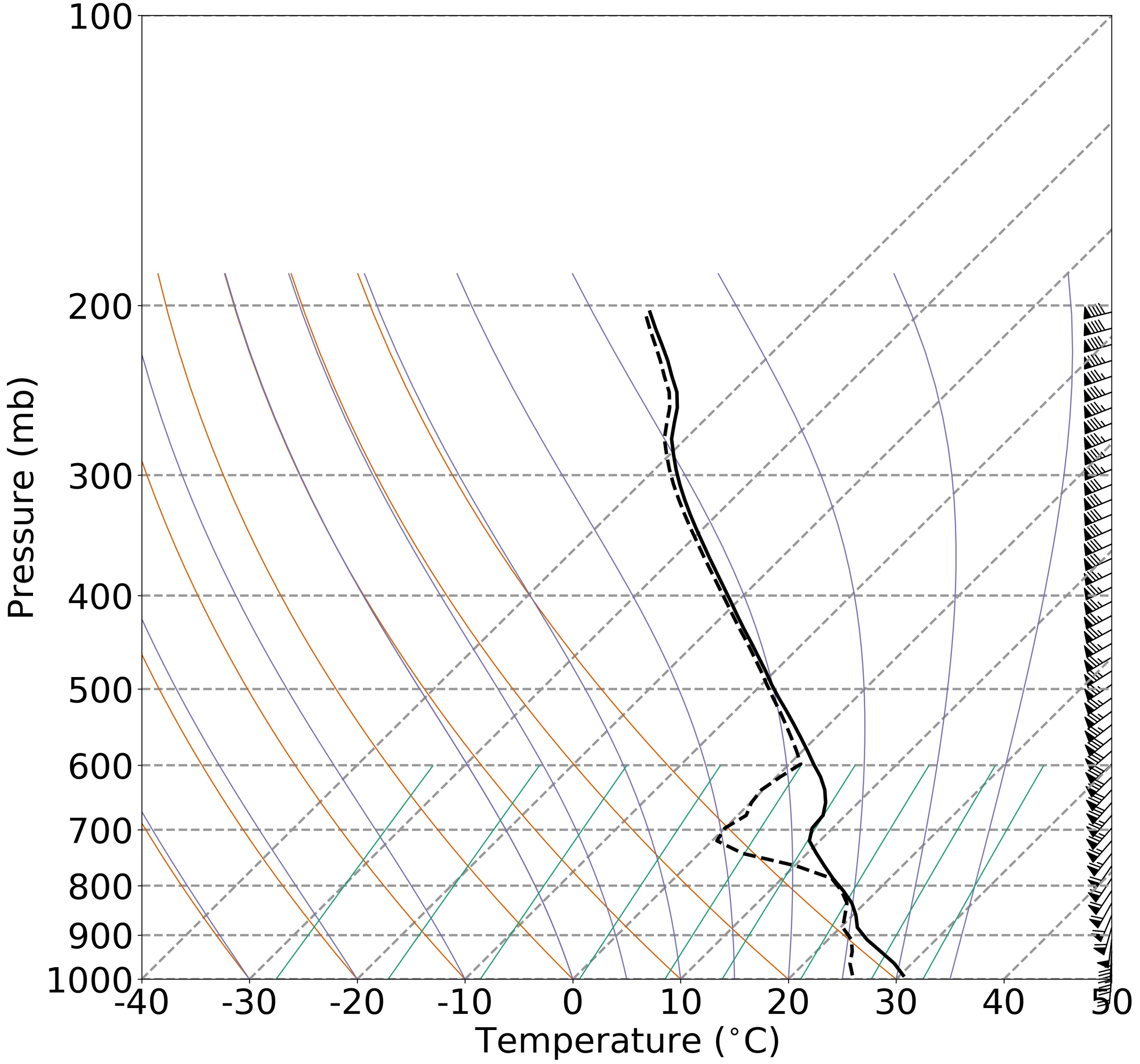
Tornado Prediction: Input Data

- 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”



Convolutional Neural Networks (CNN)

- CNNs have three main components:

1. **Convolutional layers**
    - Made up of convolutional filters that detect spatial features.
    - Convolutional filters have been used in image-processing for decades for blurring, sharpening, edge detection, etc.
    - In traditional applications the filter weights are fixed; in a CNN the weights are learned.

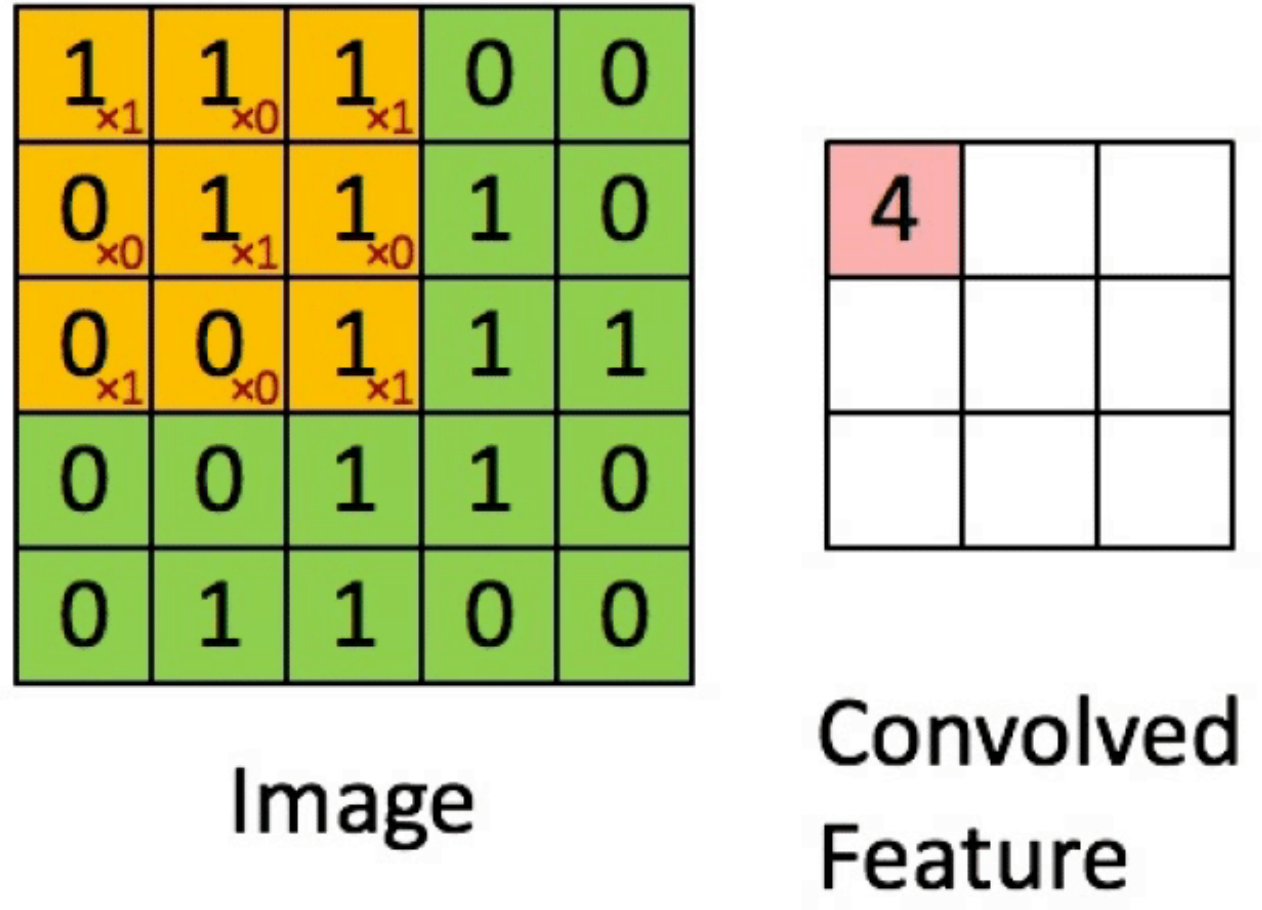


Image source:
http://i-systems.github.io/HSE545/machine%20learning%20all/Workshop/180208_COSEIK/06_CNN.html

Convolutional Neural Networks (CNN)

- CNNs have three main components:

2. Pooling layers

- Downsample the grid to lower resolution.
- Shallow conv layers (before much pooling) learn small-scale features, while deep conv layers learn large-scale features.
- Multiple scales often important for weather prediction.

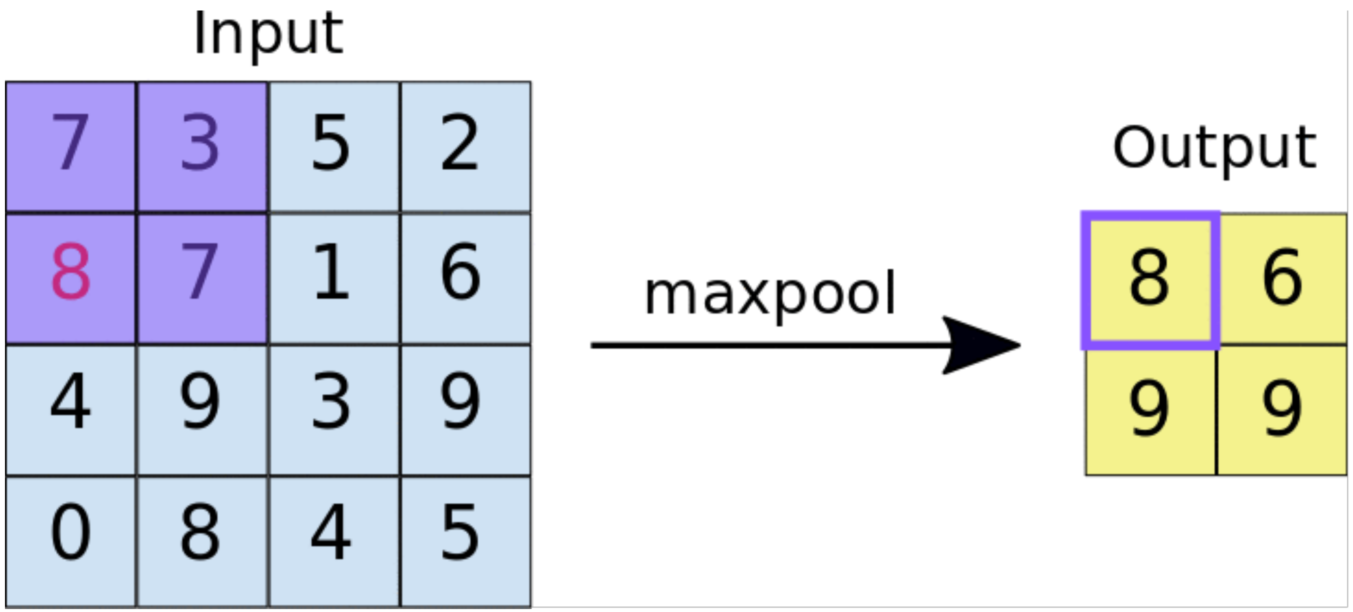


Image source: https://developers.google.com/machine-learning/practica/image-classification/convolutional-neural-networks

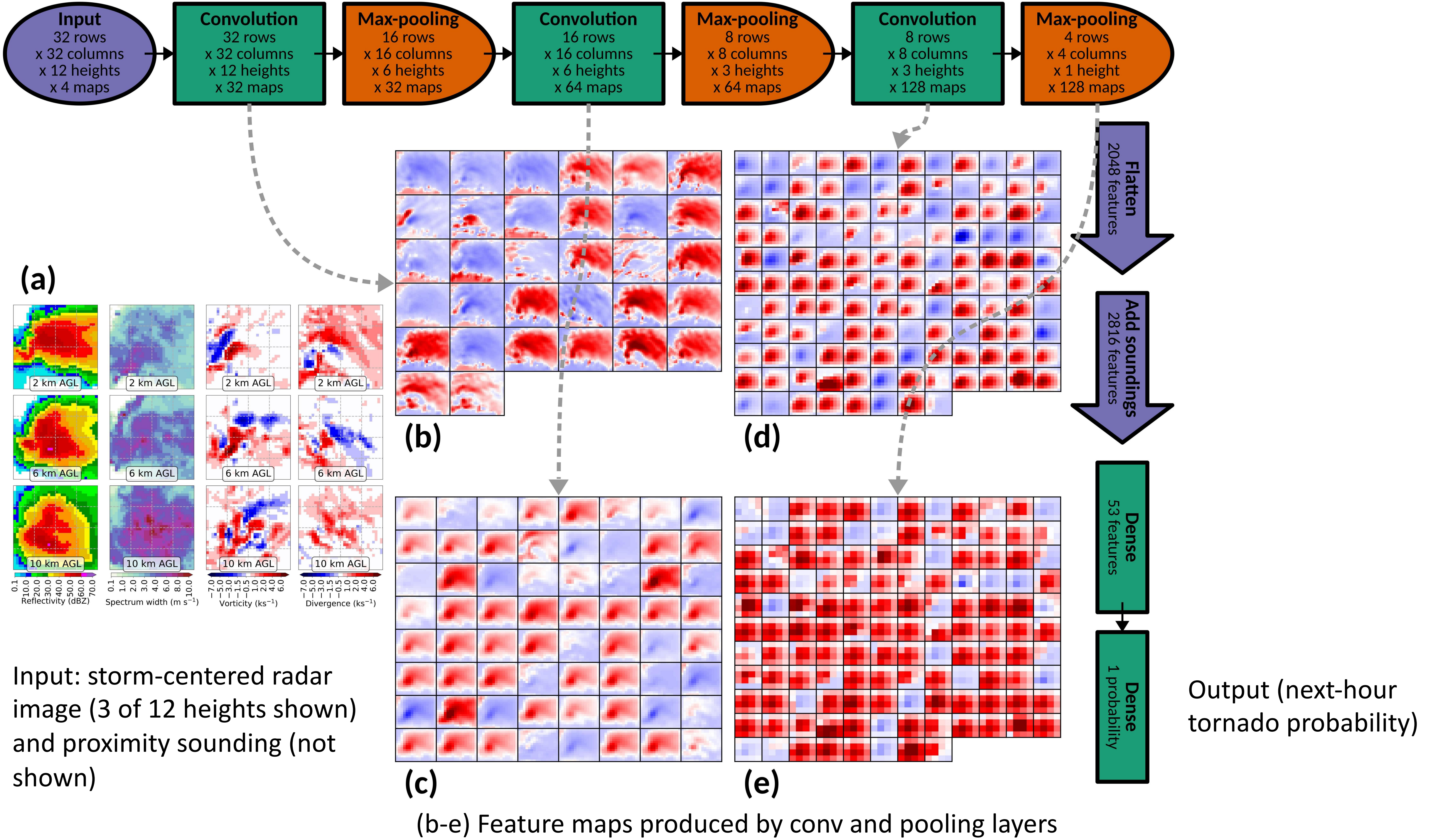
Convolutional Neural Networks (CNN)

- CNNs have three main components:

3. **Dense (fully connected) layers**
   - Spatially agnostic layers from traditional neural nets.
   - These transform features created by conv and pooling layers into final prediction.

- **To summarize:**
  - Conv and pooling layers transform gridded data into features.
  - Dense layers transform features into predictions.
  - CNN learns both transformations simultaneously.

- CNN architecture used for GridRad data shown on next page.



Input
32 rows
x 32 columns
x 12 heights
x 4 maps
Convolution
32 rows
x 32 columns
x 12 heights
x 32 maps
Max-pooling
16 rows
x 16 columns
x 6 heights
x 32 maps
Convolution
16 rows
x 16 columns
x 6 heights
x 64 maps
Max-pooling
8 rows
x 8 columns
x 3 heights
x 64 maps
Convolution
8 rows
x 8 columns
x 3 heights
x 128 maps
Max-pooling
4 rows
x 4 columns
x 1 height
x 128 maps
Flatten
2048 features
Add soundings
2816 features
Dense
53 features
Dense
1 probability
Output (next-hour tornado probability)
(a)
2 km AGL
6 km AGL
10 km AGL
Reflectivity (dBZ)
Spectrum width (m s⁻¹)
Vorticity (ks⁻¹)
Divergence (ks⁻¹)
Input: storm-centered radar image (3 of 12 heights shown) and proximity sounding (not shown)
(b)
(c)
(d)
(e)
(b-e) Feature maps produced by conv and pooling layers

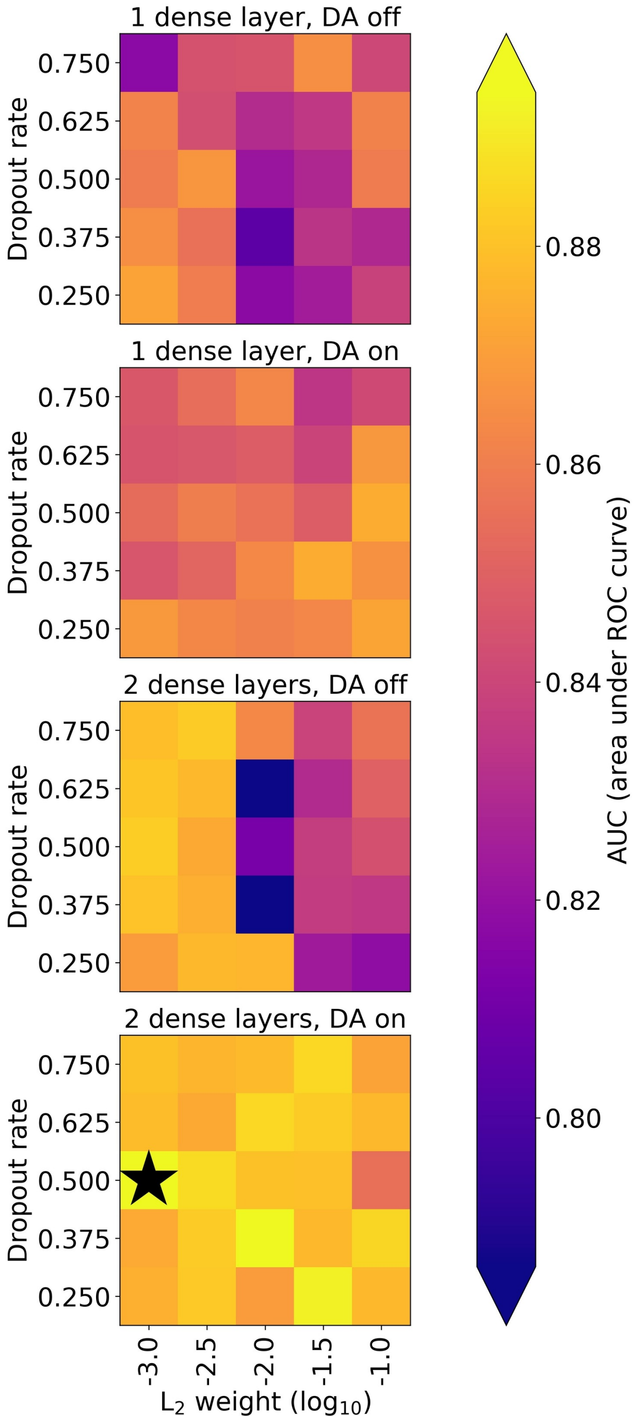
Tornado Prediction: Hyperparameter Experiment

- **“Hyperparameter”** = characteristic of model itself *(e.g.*, number of layers) that must be chosen *a priori*.
  - Model weights are fit to training data; hyperparameters are fit to validation data.

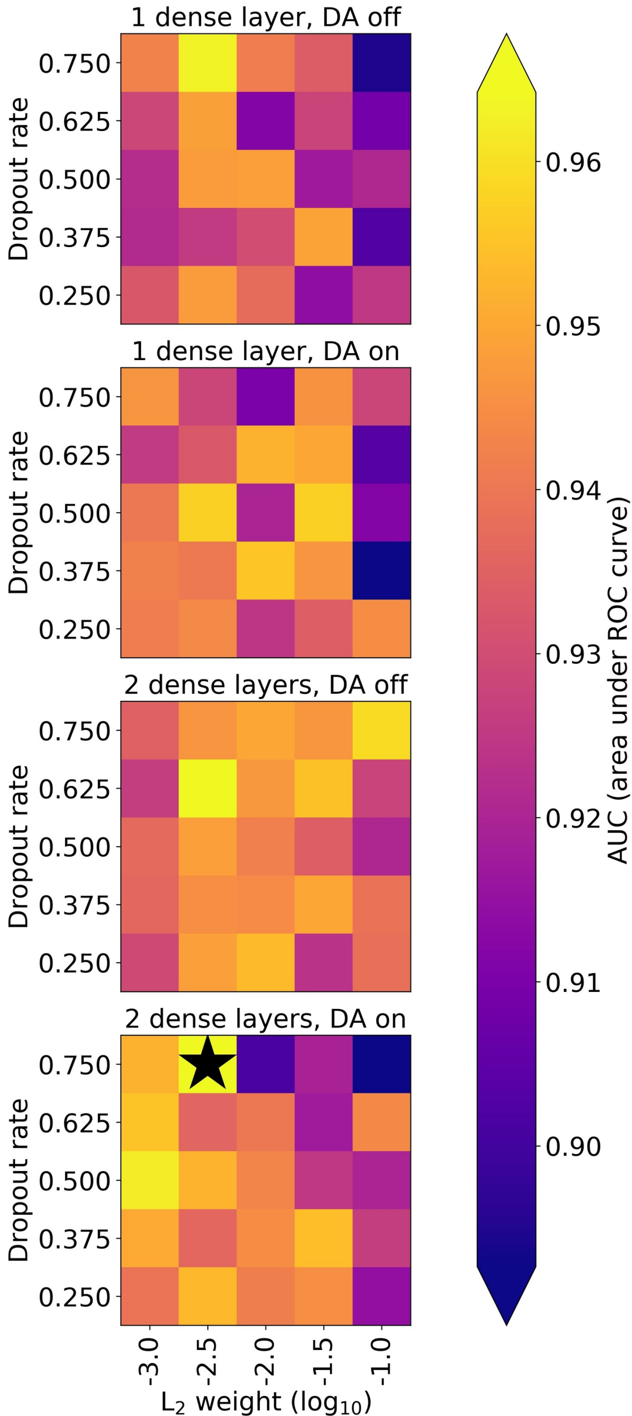
- **I perform a grid search over 4 hyperparameters, which mainly control overfitting:**
  - Weight for L_2 regularization
  - Rate for dropout regularization
  - Number of dense layers
  - Data augmentation (on/off)

- For both MYRORSS and GridRad, I choose model with highest AUC (area under ROC curve) on validation data.

- Models generally perform best with data augmentation and 2 dense layers (instead of 1).



GridRad



MYRORSS

Tornado Prediction: Hyperparameter Experiment

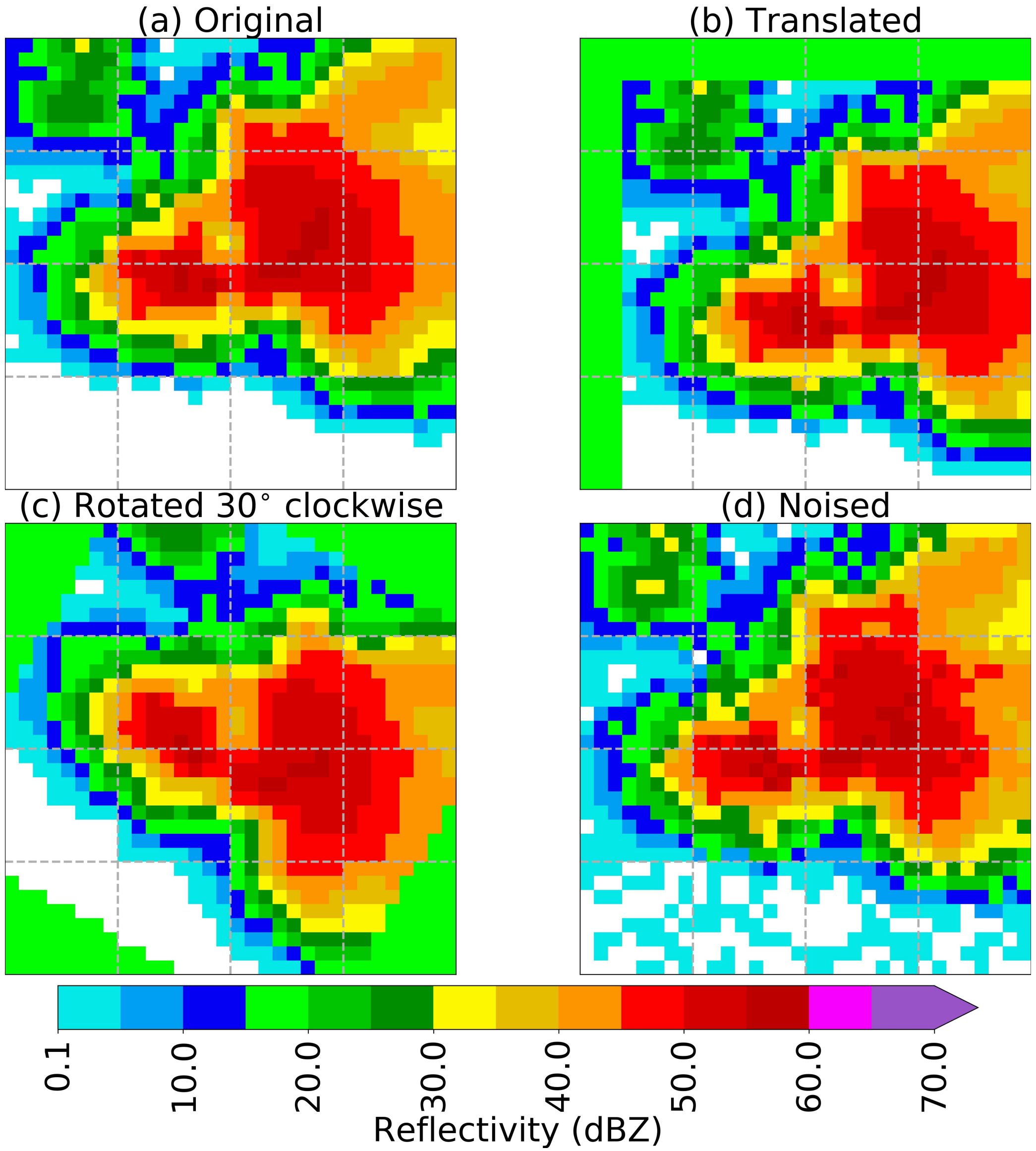
- **Data augmentation, used during training, allows the CNN to generalize better (overfit less).**
  - Apply small perturbations to predictors and assume that the label (tornadic or non-tornadic) stays the same.

- **This allows the CNN to generalize better (overfit less).**

- Specifically, I apply 17 perturbations to each storm-centered radar image:
  - Horizontal rotation (-15°, +15°, -30°, +30°)
  - Horizontal translation (move three grid cells in eight directions spaced equally from 0-315°)
  - Add Gaussian noise five times (variance of noise = 0.1 * variance of radar variable)

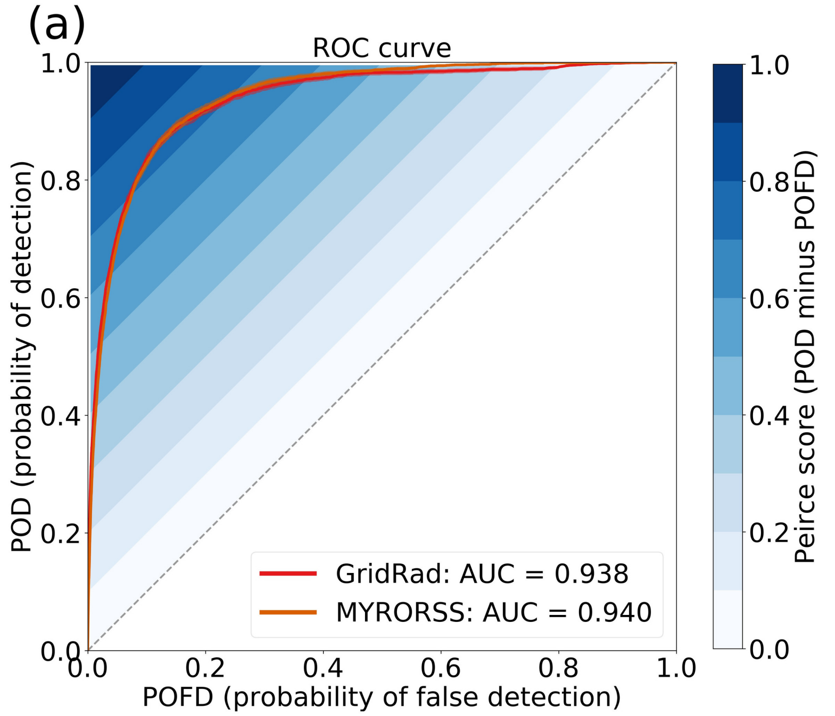
Data Augmentation

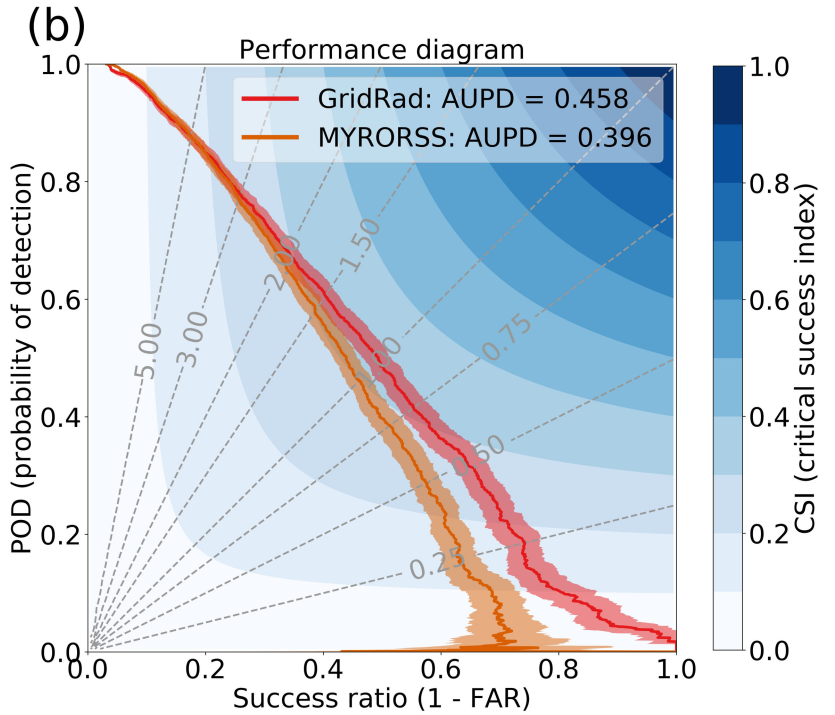
- **Right: three perturbations for reflectivity at 3 km AGL.**
- Same perturbations applied in tandem to all variables at all heights.



Tornado Prediction: Model Evaluation

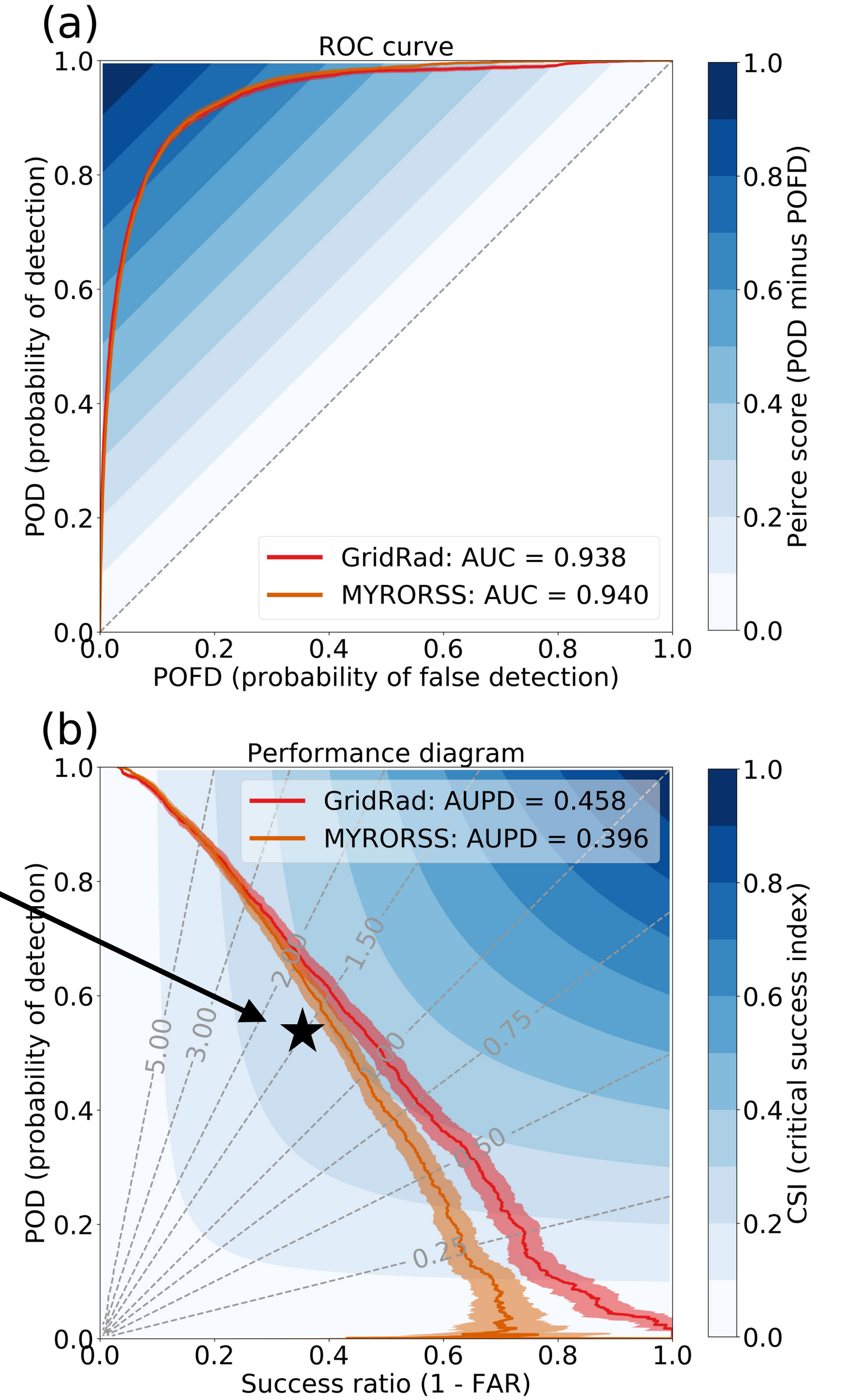
- **Right: results on testing data**
- Testing sets for MYRORSS and GridRad contain the same storm objects (ensured by matching technique)
- 116 629 storm objects, 3.19% tornadic in next hour
- AUC > 0.9 for both models, generally considered “excellent” performance
- However, maximum CSI is low (~0.3)
- Low CSI is typical for rare events, because high CSI requires high POD and low FAR





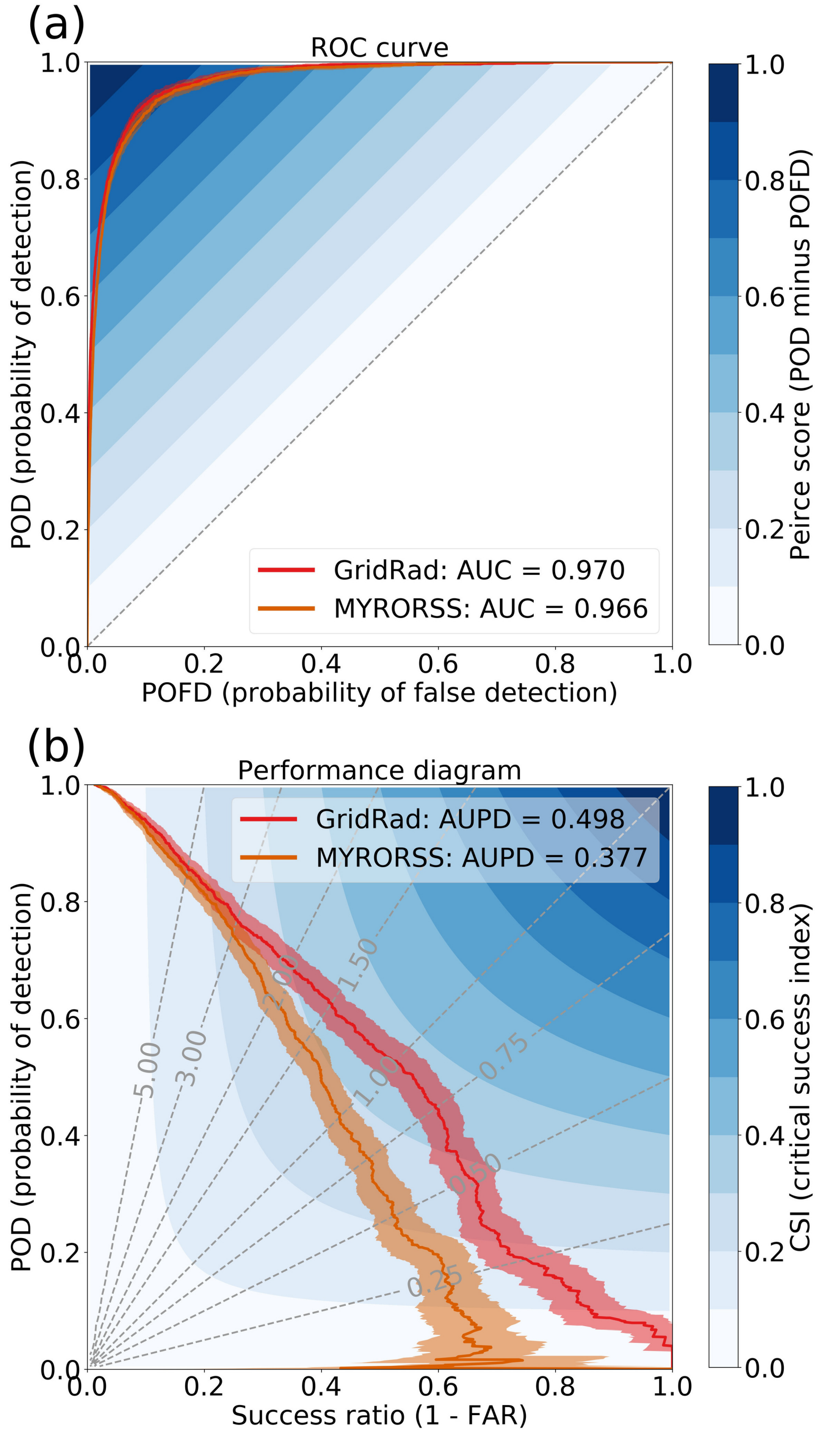
Tornado Prediction: Model Evaluation

- **Results comparable to ProbSevere** (Cintineo *et al.* 2018), an ML model currently used in operations.
- ProbSevere achieves lower CSI (0.27) with higher event frequency (4.94%).
- **However, comparison is not apples-to-apples.**
  - ProbSevere uses real-time version of MYRORSS data
  - ProbSevere predicts *all* severe weather (tornado or hail or damaging wind)
- **Nonetheless, comparison suggests my CNNs would be useful in operations.**



Tornado Prediction: Model Evaluation

- **Right: same but excluding weak (EF-0 and EF-1) tornadoes**
- Weak tornadoes are often not reported, especially in remote areas and at night
- 114 427 storm objects, 1.33% tornadic in next hour
- If skill were independent of tornado strength, would except same AUC and decrease in CSI
- However, both AUC and CSI increase (models are better for strong tornadoes)



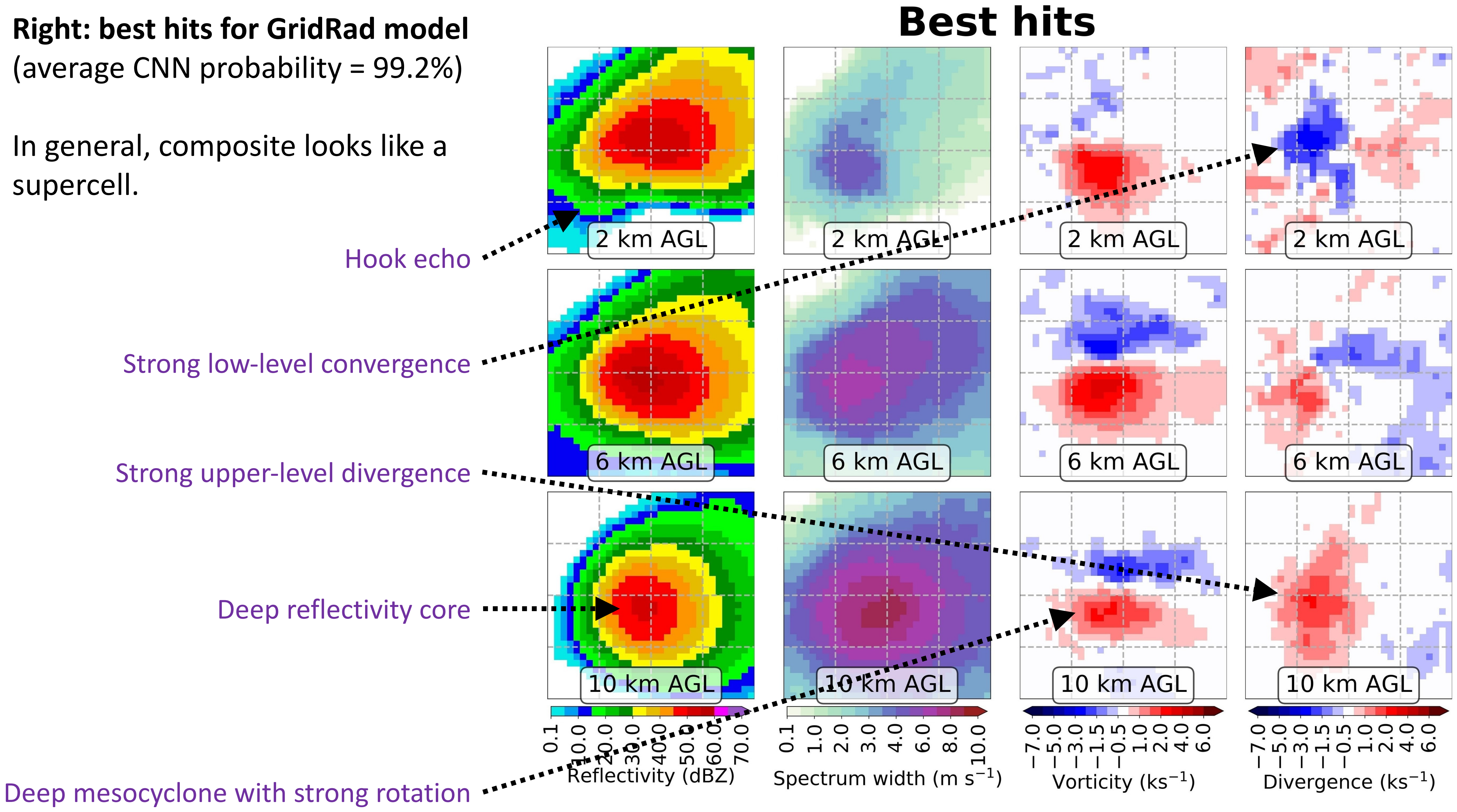
Tornado Prediction: Model Evaluation

- **The next few slides will show extreme cases:**
  - 100 best hits (tornadic storms with high CNN probability)
  - 100 worst false alarms (non-tornadic storms with high probability)
  - 100 worst misses (tornadic storms with low probability)
  - 100 best correct nulls (non-tornadic storms with low probability)

- Storm objects in each set are composited by probability-matched means (PMM; Ebert 2001).

- PMM preserves spatial structure better than computing mean grid point by grid point.

- **Right: best hits for GridRad model** (average CNN probability = 99.2%)
- In general, composite looks like a supercell.



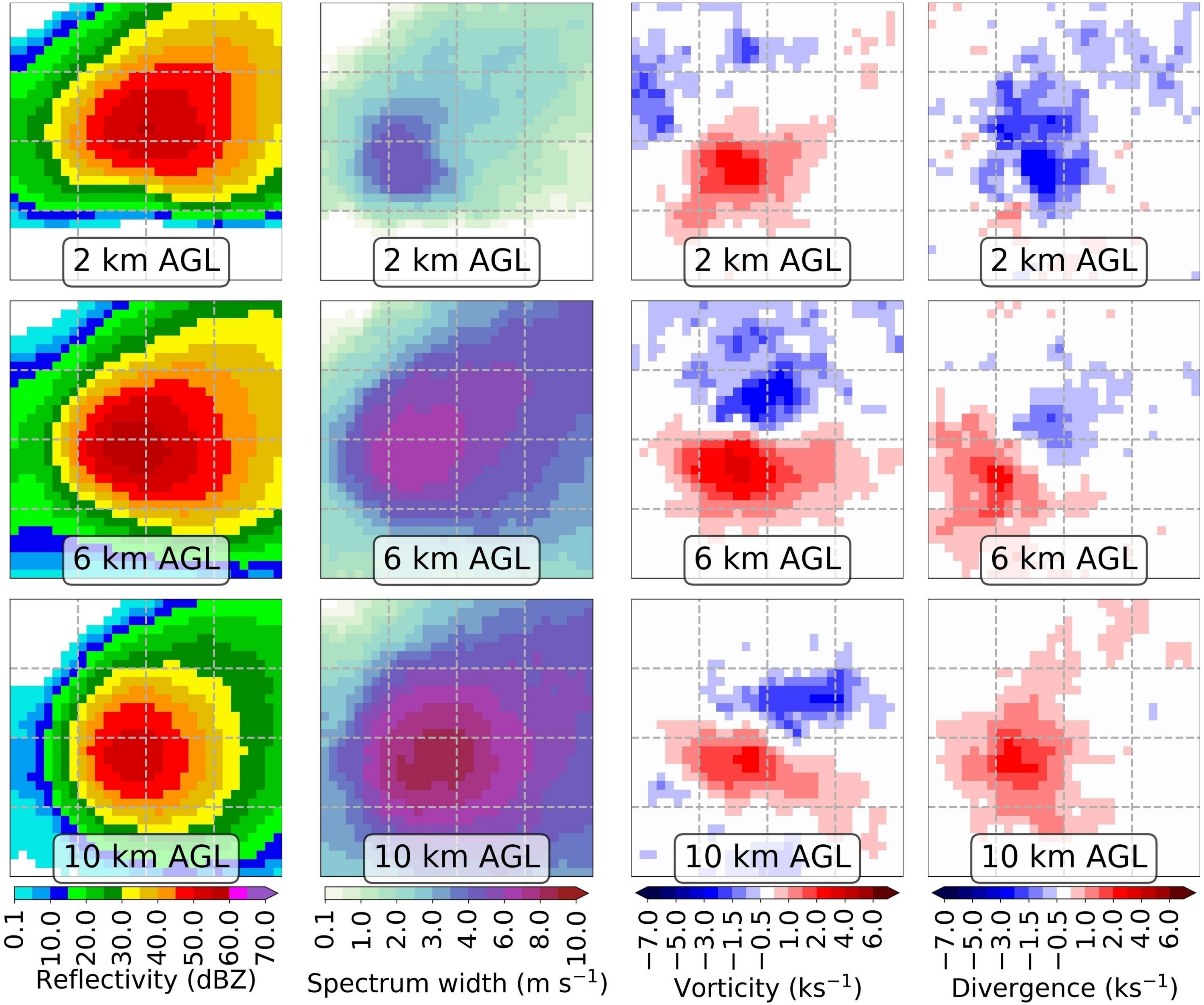
- **Right: worst false alarms for GridRad model** (average CNN probability = 98.8%)

- Worst false alarms look very similar to best hits.

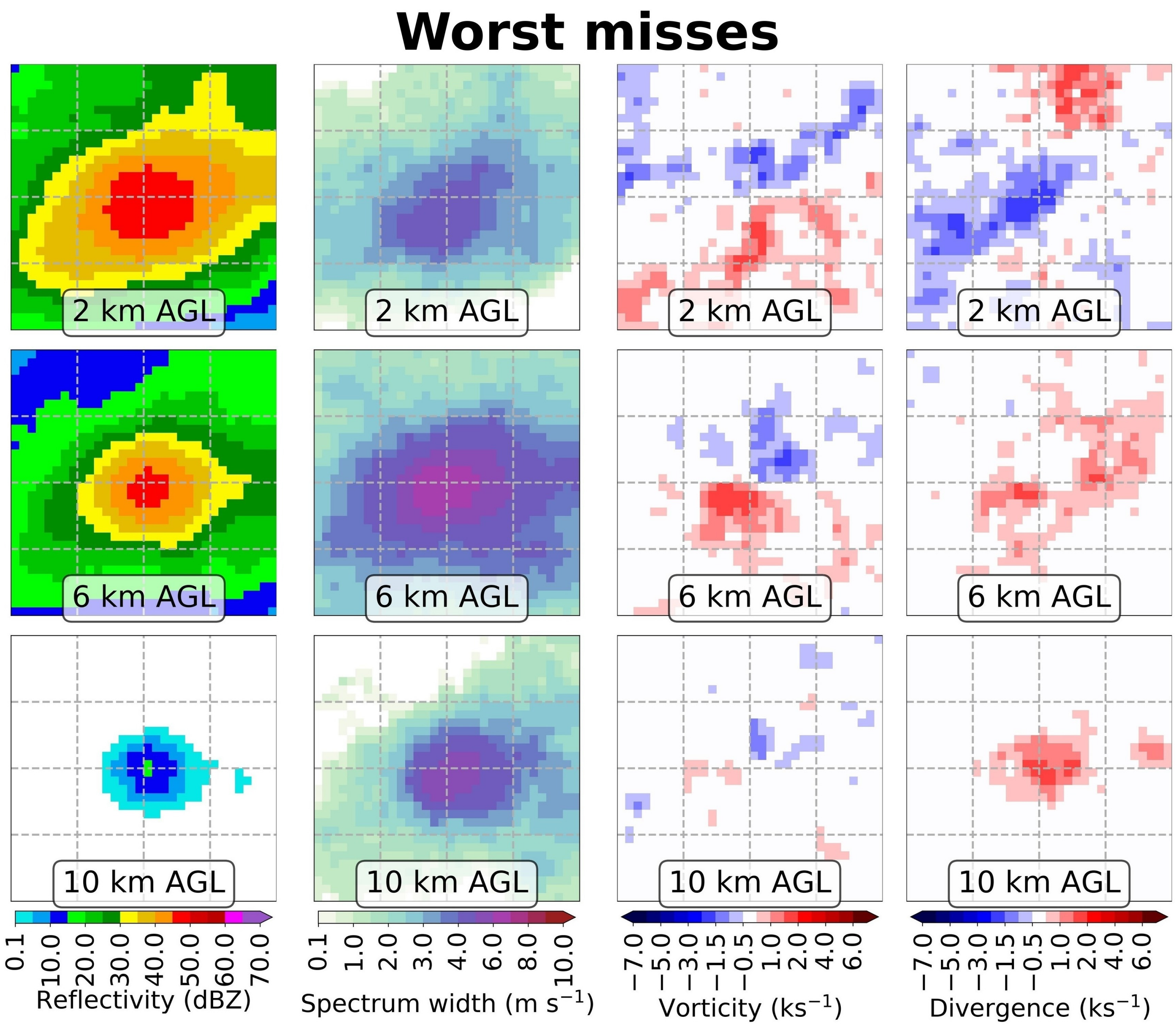
- **76 of the 100 storms have an NWS tornado warning, so they are false alarms for humans as well.**

- Similarity between best hits and false alarms caused by dichotomous labeling:
  - Funnel cloud that almost touches down = “no”
  - Weak tornado that briefly touches down = “yes”

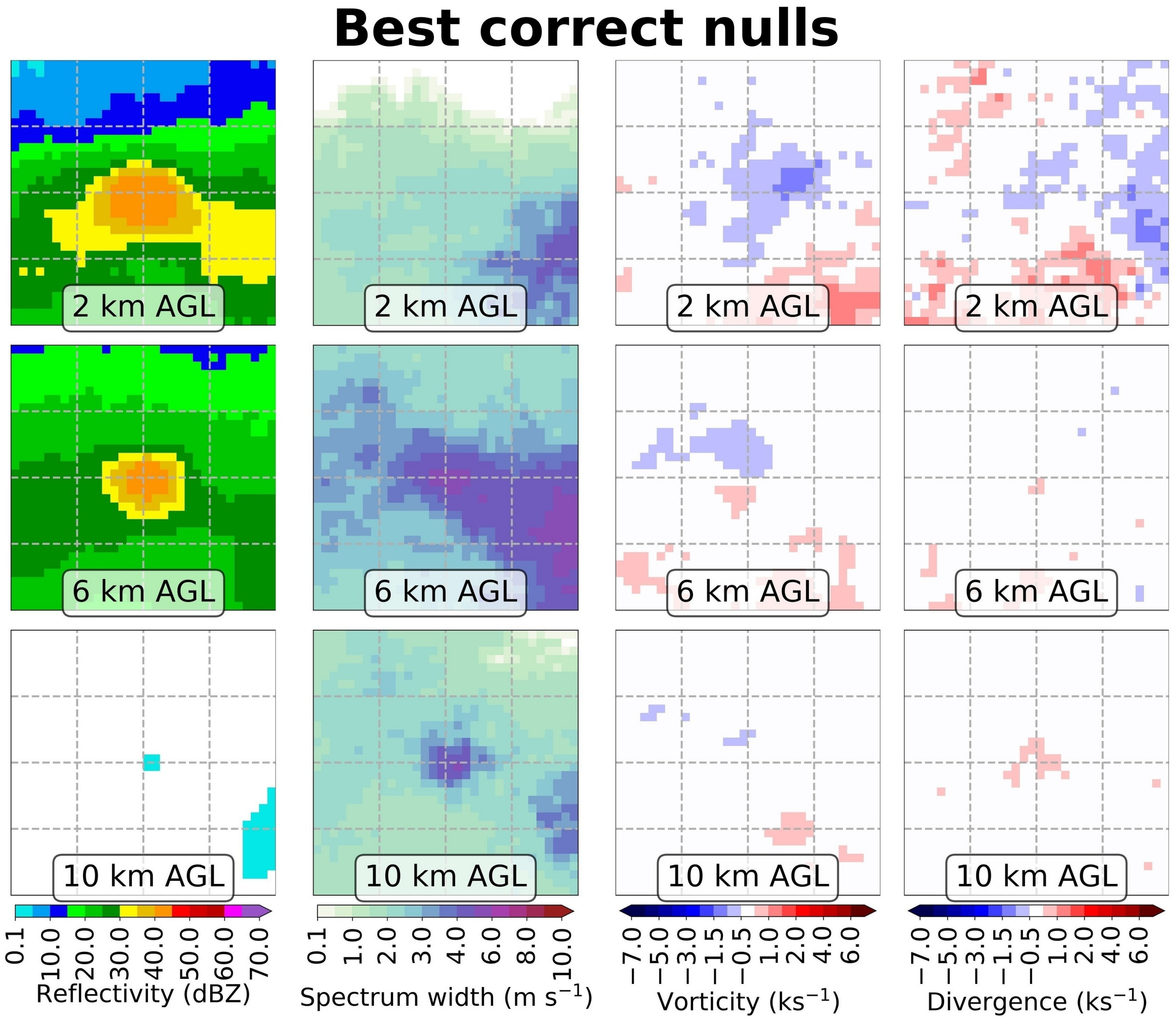
Worst false alarms



- **Right: worst misses for GridRad model** (average CNN probability = 8.6%)
- Shallow elongated reflectivity core with weak rotation.
- By inspection, 67 of the 100 storms are part of quasi-linear convective systems (QLCS).
- QLCS tornadoes are a common failure mode for humans and other forecasting methods (Brotzge *et al.* 2013; Anderson-Frey *et al.* 2016).



- **Right: best correct nulls for GridRad model** (average CNN probability = 0.004%)
- Weak reflectivity and rotation at all heights.
- By inspection, these storms are mostly short-lived cells in mesoscale convective systems (MCS).



Tornado Prediction: Model Interpretation

- I use several interpretation methods to understand physical relationships learned by the CNNs.

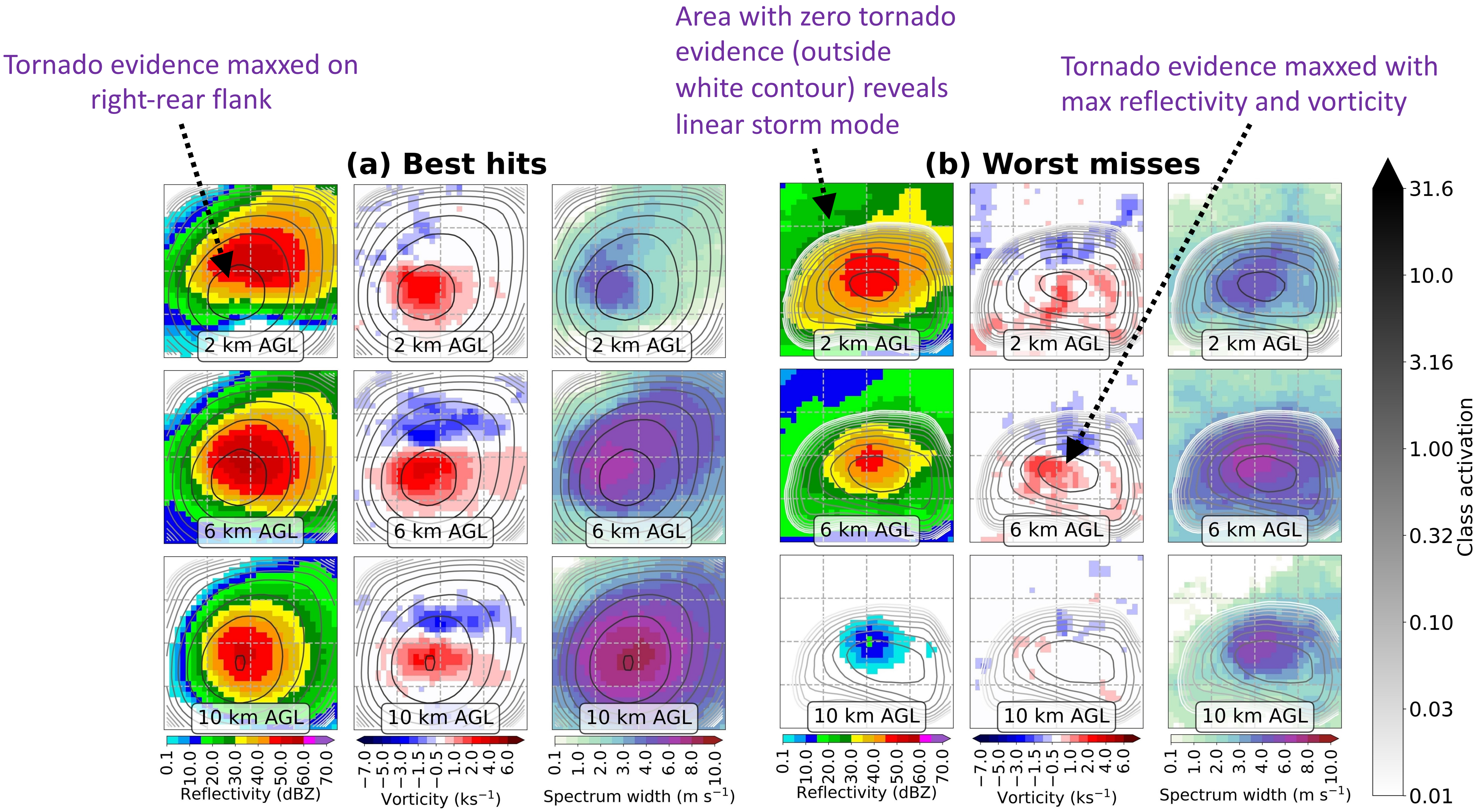
- I will show just a few results here (for more details, see McGovern *et al.* 2019 and 2020).

- **Based on literature, expected the following features to be conducive to tornadoes:**
    - Deep reflectivity core
    - Strongly rotating, compact low-level mesocyclone
    - Discrete storm (isolated from other storms)
    - Strong low-level wind shear, relative humidity, instability
    - Weak reflectivity in rear-flank downdraft (RFD)
        - Strong reflectivity suggests a lot of evaporative cooling and negative buoyancy, which could prevent tornadogenesis (Markowski *et al.* 2002; Markowski and Richardson 2009)

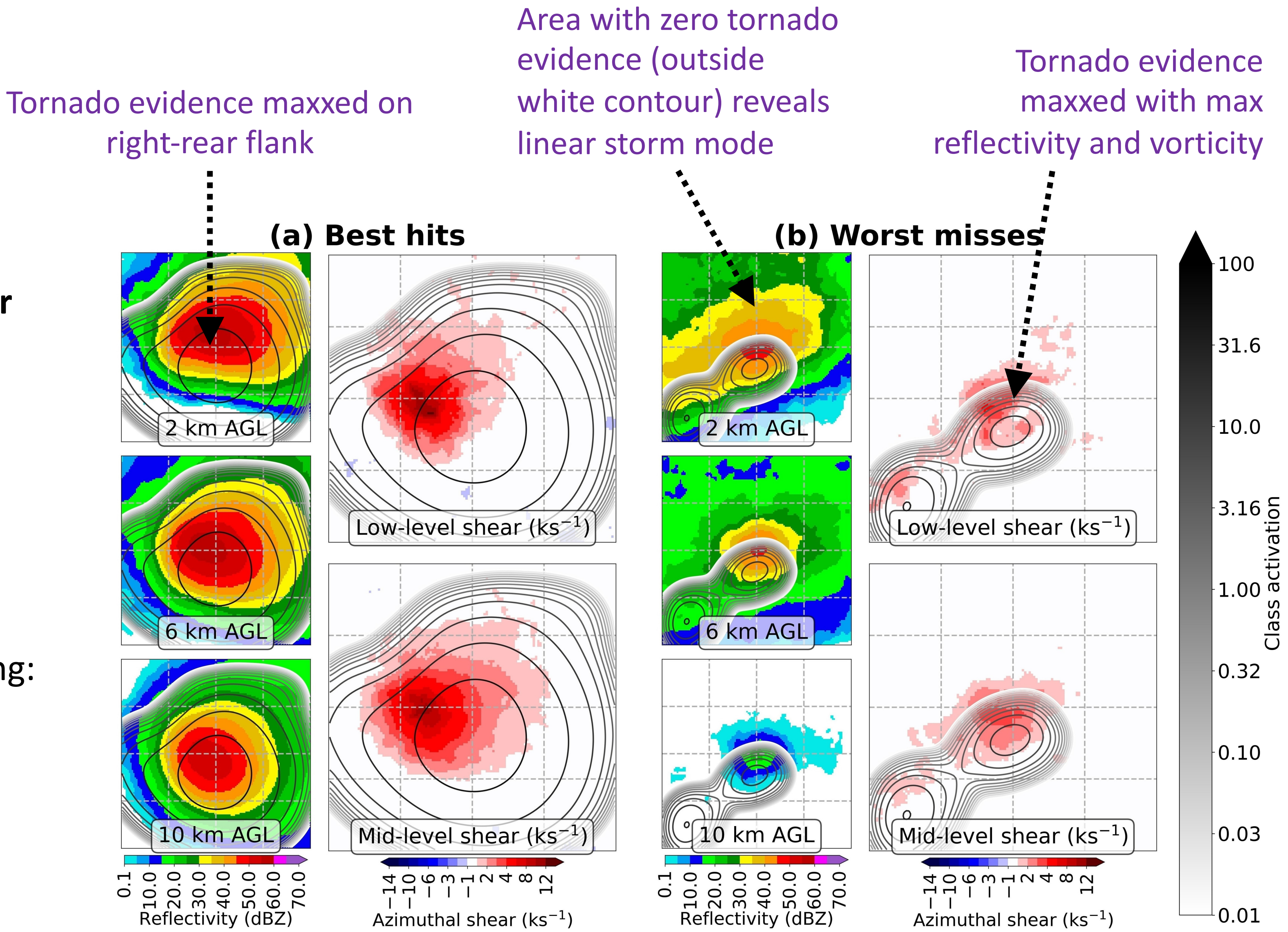
Class-activation Maps (CAM)

- Class activation (Zhou *et al.* 2016) is amount of **evidence for the positive class (tornado in next hour).**
- Class activation is defined at each grid point, so can be viewed as a map.
- I will use “class activation” and “tornado evidence” interchangeably.

Below: composited CAMs for GridRad model



- **Right: composited CAMs for MYRORSS model**
- Results are similar overall.
- Encouraging sign for generalizability, since MYRORSS and GridRad models differ in the following:
  - Architecture
  - Input dataset
  - Training period



Saliency Maps

- Saliency (Simonyan *et al.* 2014), also called sensitivity, is defined as follows.

$$\text{saliency} = \left.\frac{\partial p}{\partial x}\right|_{x=x_0}$$

- p = tornado probability
- x = input value (one predictor at one grid point)
- x_0 = value of x in actual storm

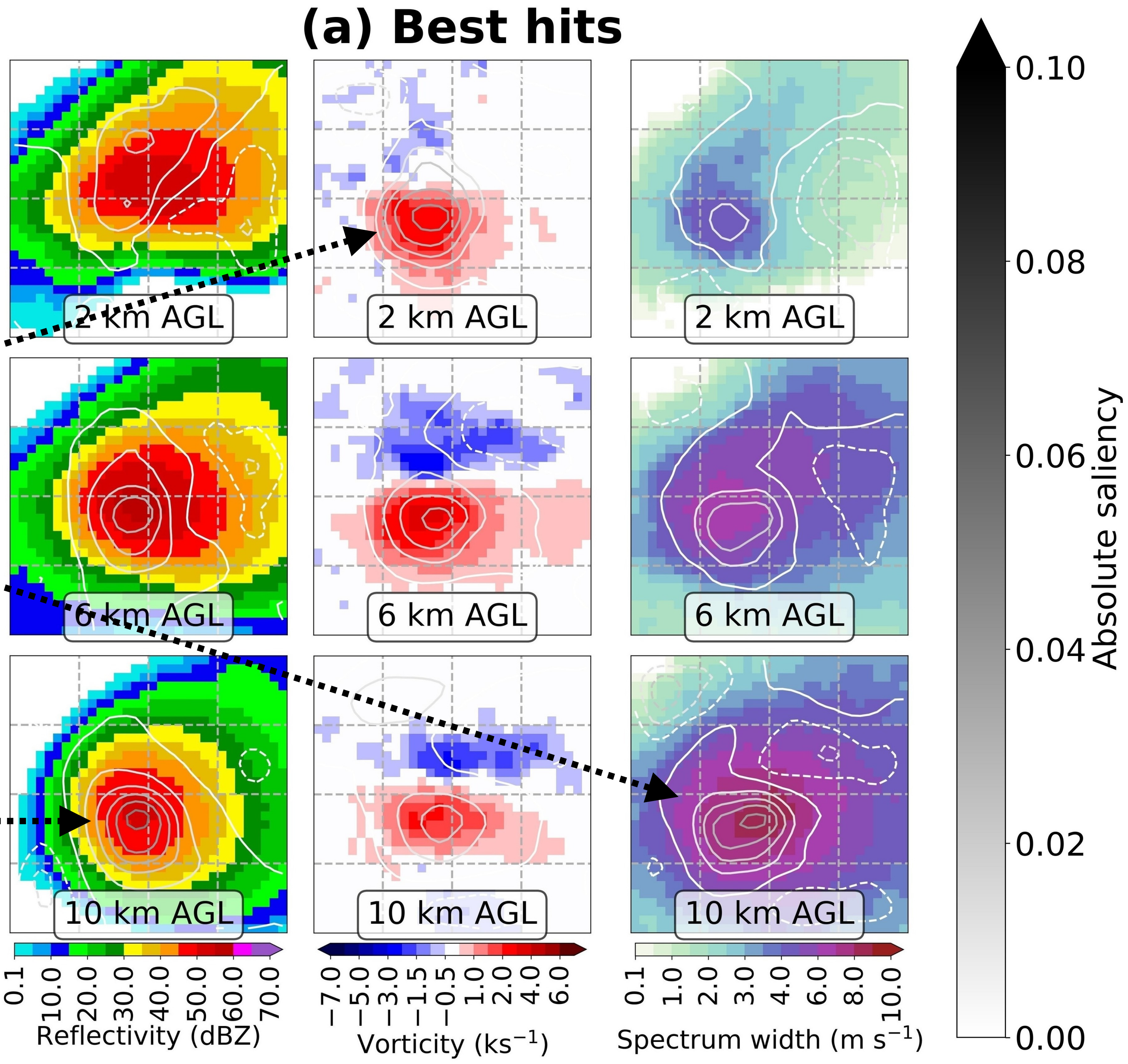
- **Thus, saliency is a linear approximation to $\frac{\partial p}{\partial x}$ around the point $x = x_0$.**

- **Right: saliency map for best hits in GridRad model**
- Solid (dashed) contours for positive (negative) saliency

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

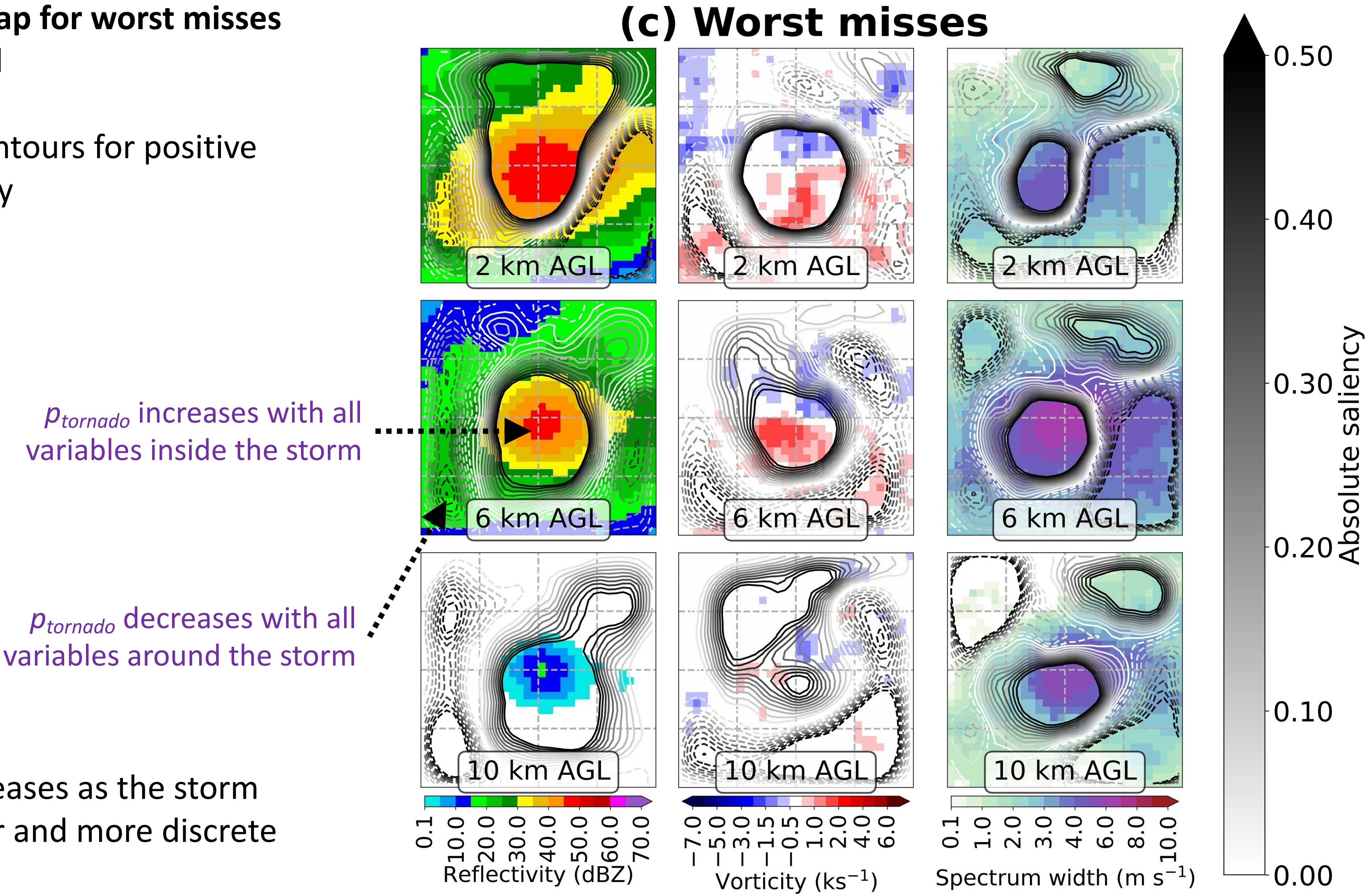
$p_{tornado}$ increases with spectrum width

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



- **Right: saliency map for worst misses in GridRad model**

- Solid (dashed) contours for positive (negative) saliency



- Thus, $p_{tornado}$ increases as the storm becomes stronger and more discrete

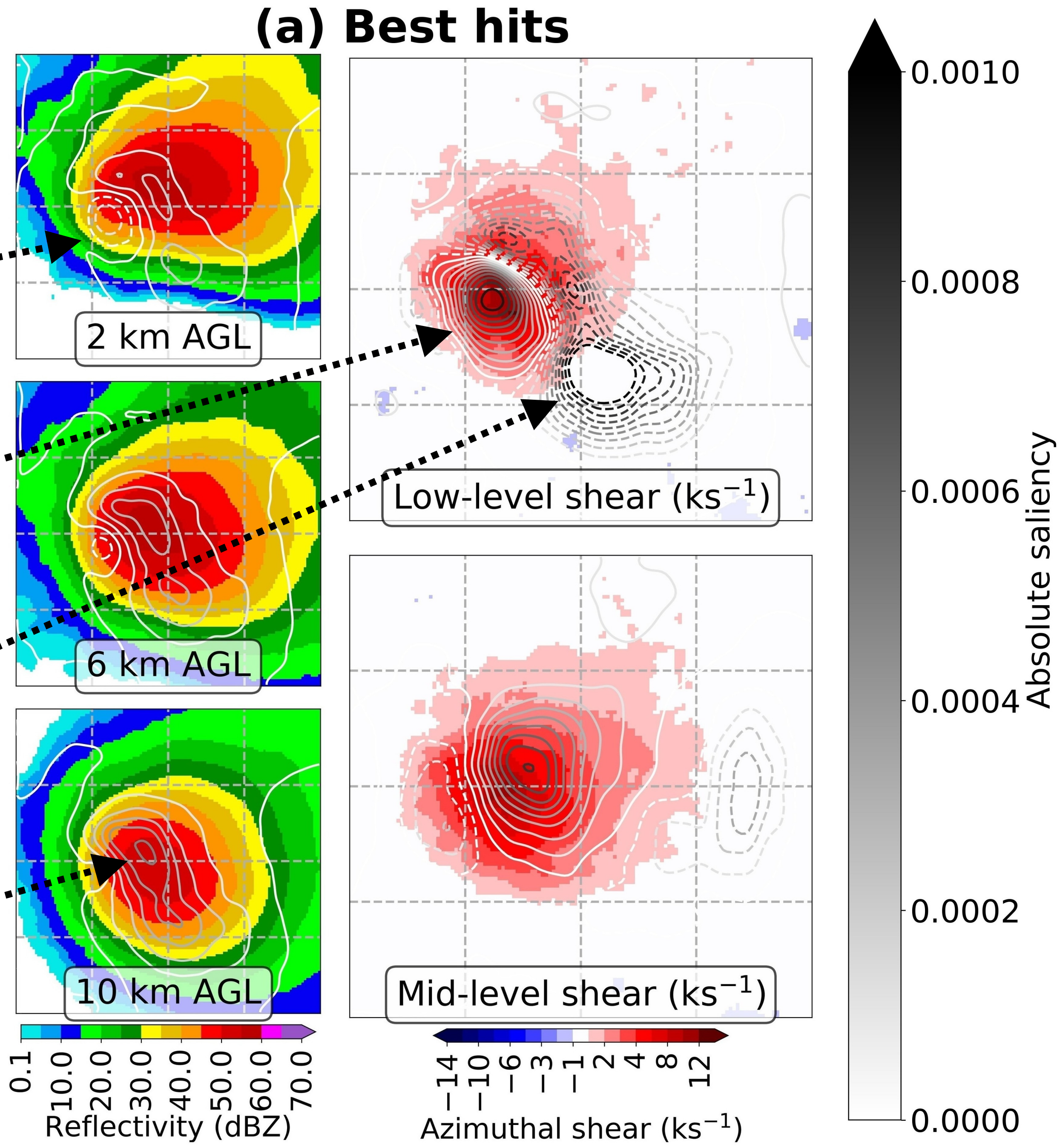
- **Right: saliency map for best hits in MYRORSS model**
- Solid (dashed) contours for positive (negative) saliency

$p_{tornado}$ decreases with reflectivity in RFD, especially at low levels

$p_{tornado}$ increases with rotation in low-level mesocyclone

$p_{tornado}$ increases as low-level mesocyclone contracts

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



Backward Optimization (BWO)

- Backward optimization (BWO; Erhan *et al.* 2009) creates synthetic input to minimize or maximize CNN prediction (tornado probability).

- **I use BWO to decrease tornado probability for best hits in MYRORSS model.**

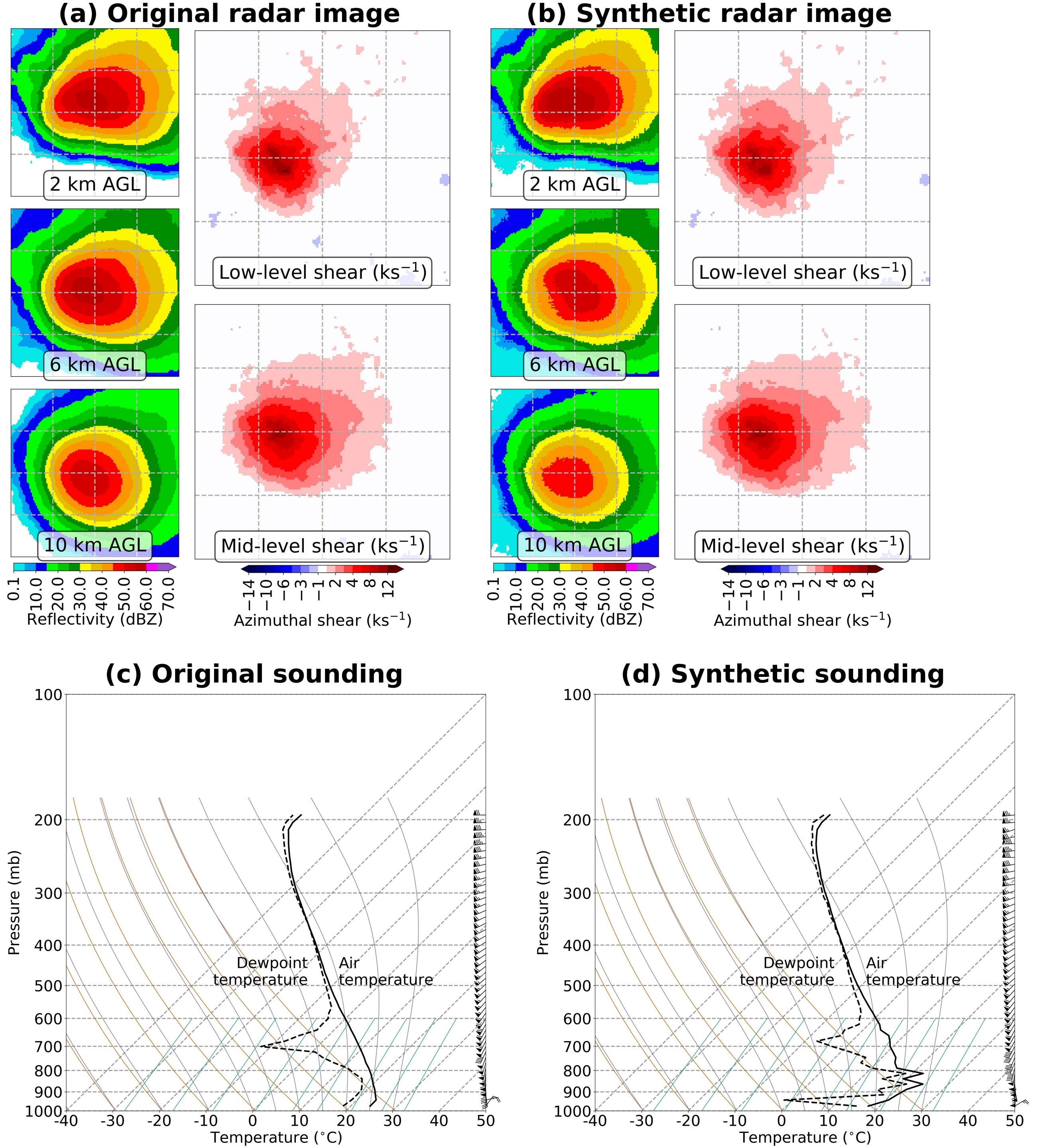
- On average for the 100 storms, decreases probability from 99.6% to 9.7%.

- **BWO has little effect, except in the sounding below 700 mb:**
  - Creates deep temperature inversion, reducing CAPE to zero
  - Decreases low-level wind speed and thus shear

- However, synthetic sounding does not look very realistic (has the “jaggies”).

- I use several physical constraints to alleviate this problem (looked much worse without).

- Nonetheless, more work needed if we want to use ML to create realistic weather data.



Front Detection: Intro

- Synoptic-scale fronts (henceforth just “fronts”) often trigger extreme weather, including heavy precipitation and severe thunderstorms.
- A front is a transition zone between two air masses with different thermal properties.
- Typically defined by (potential) temperature, wet-bulb (potential) temperature, or equivalent (potential) temperature.

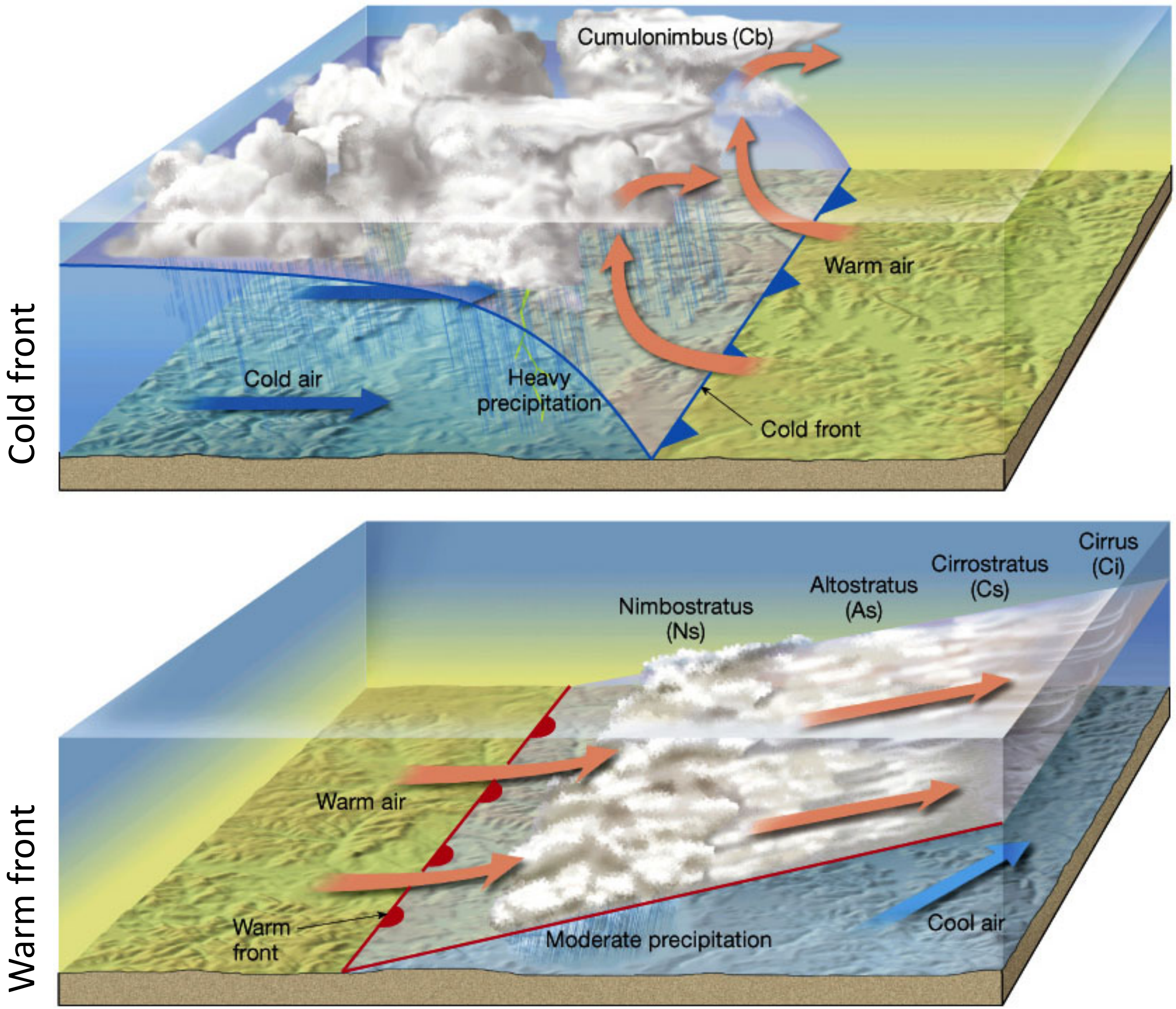


Image source: Figure 9.6 of Lutgens and Tarbuck (2000)

Front Detection: Intro

- **Front detection is usually done by hand or by numerical frontal analysis** (NFA; Hewson 1998).

- **Both have major disadvantages:**
  - Hand analysis is time-consuming
  - NFA typically produces noisy results and captures only specific types of fronts
    - Example: Schemm *et al.* (2015) found that commonly used method rarely detects warm fronts

- **This has spurred recent efforts to use deep learning** (Liu *et al.* 2016; Racah *et al.* 2017; Kurth *et al.* 2018; Kunkel *et al.* 2018; Lagerquist *et al.* 2019).

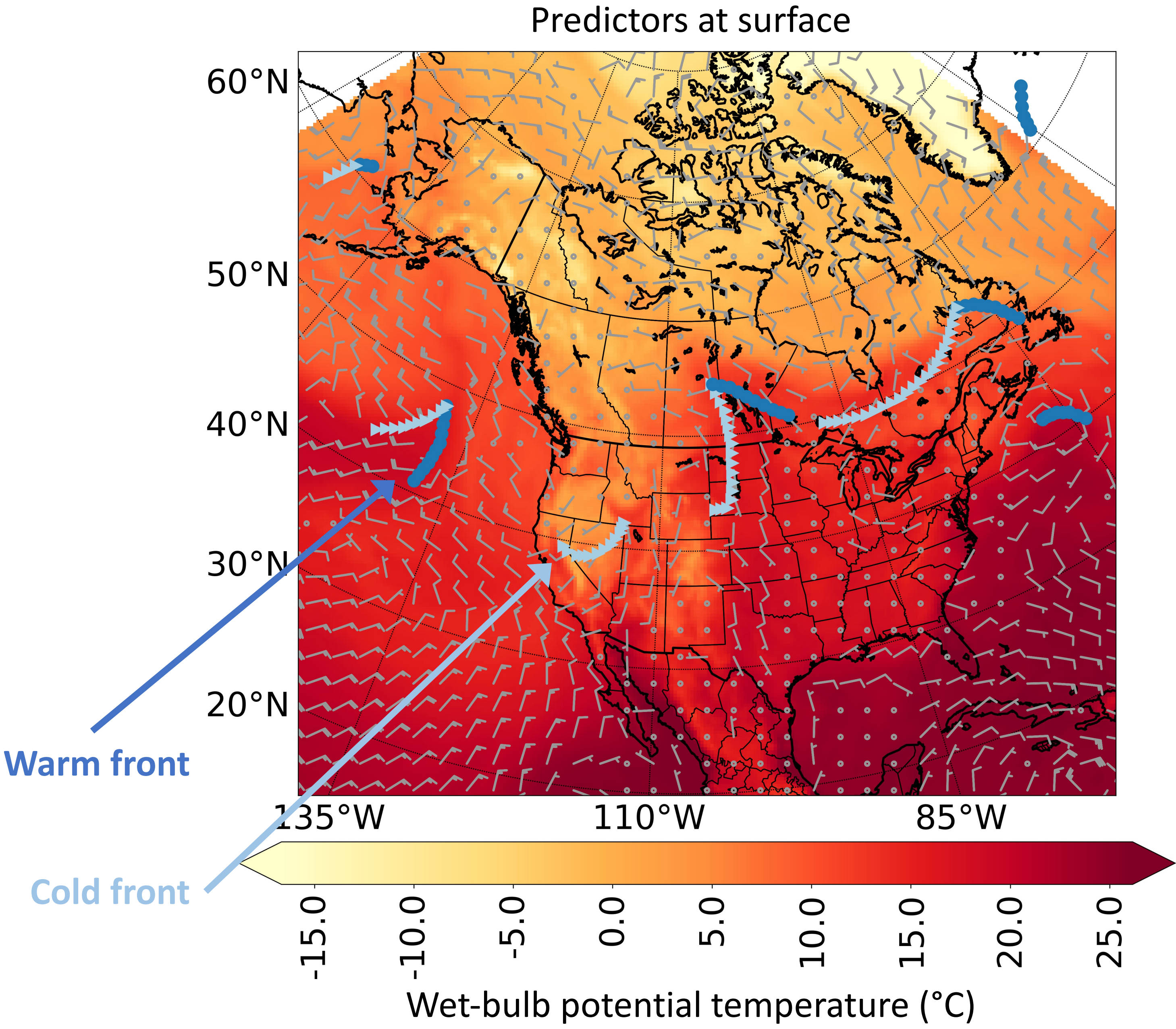
- CNNs are well suited for front detection, because they can directly process spatial grids.

Front Detection: Input Data

- **I use two datasets with 3-hour time steps:**
  - ERA5 reanalysis (Hersbach and Dee 2016) for predictors
  - Weather Prediction Center (WPC) surface fronts for labels

- **I use the following ERA5 variables at both the surface and 850 mb:**
  - Temperature
  - Specific humidity
  - Wind (*u* and *v*)

- Training: 2008-14
- Validation: 2015-16
- Testing: 2017

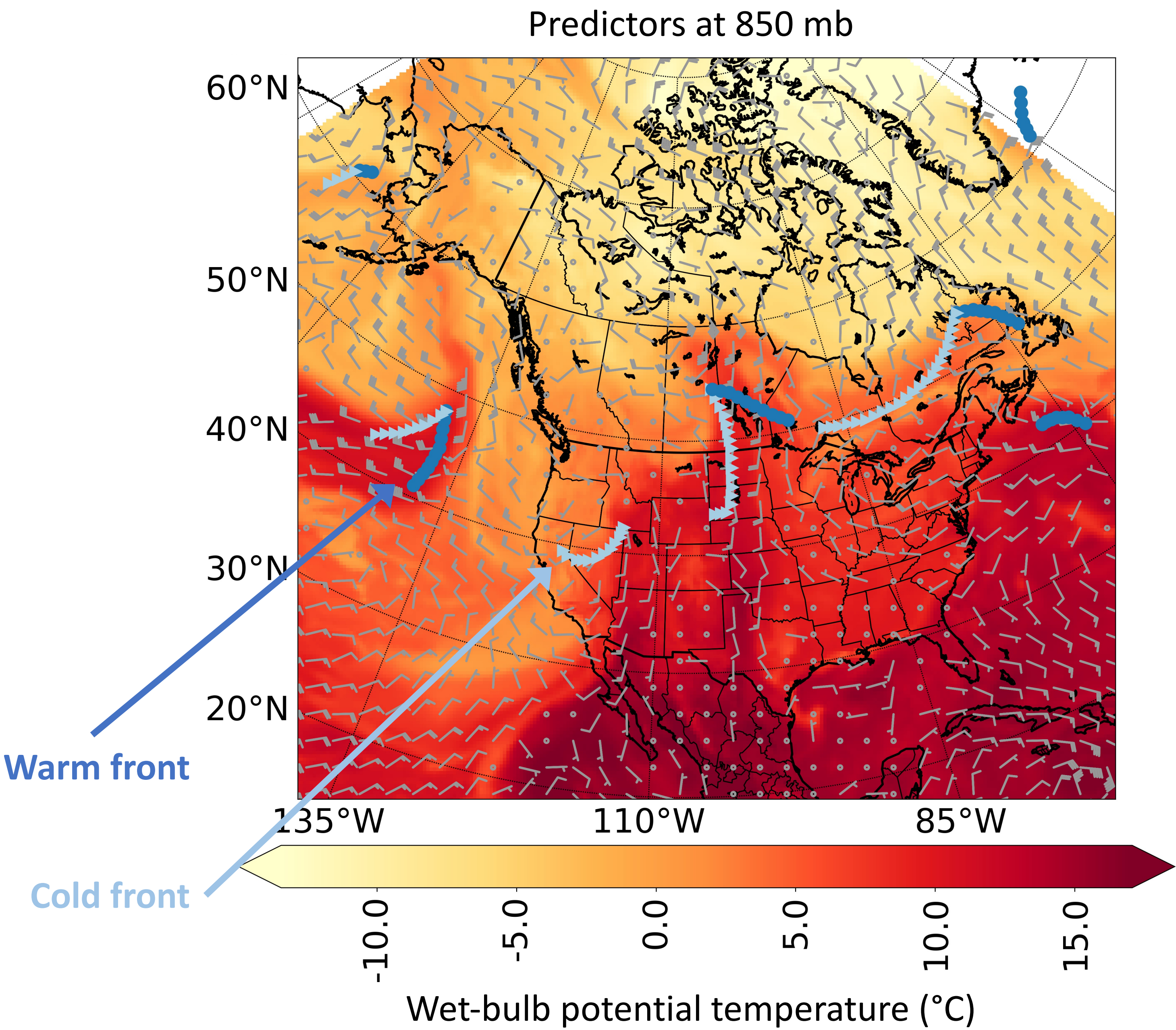


Front Detection: Input Data

- **I use two datasets with 3-hour time steps:**
  - ERA5 reanalysis (Hersbach and Dee 2016) for predictors
  - Weather Prediction Center (WPC) surface fronts for labels

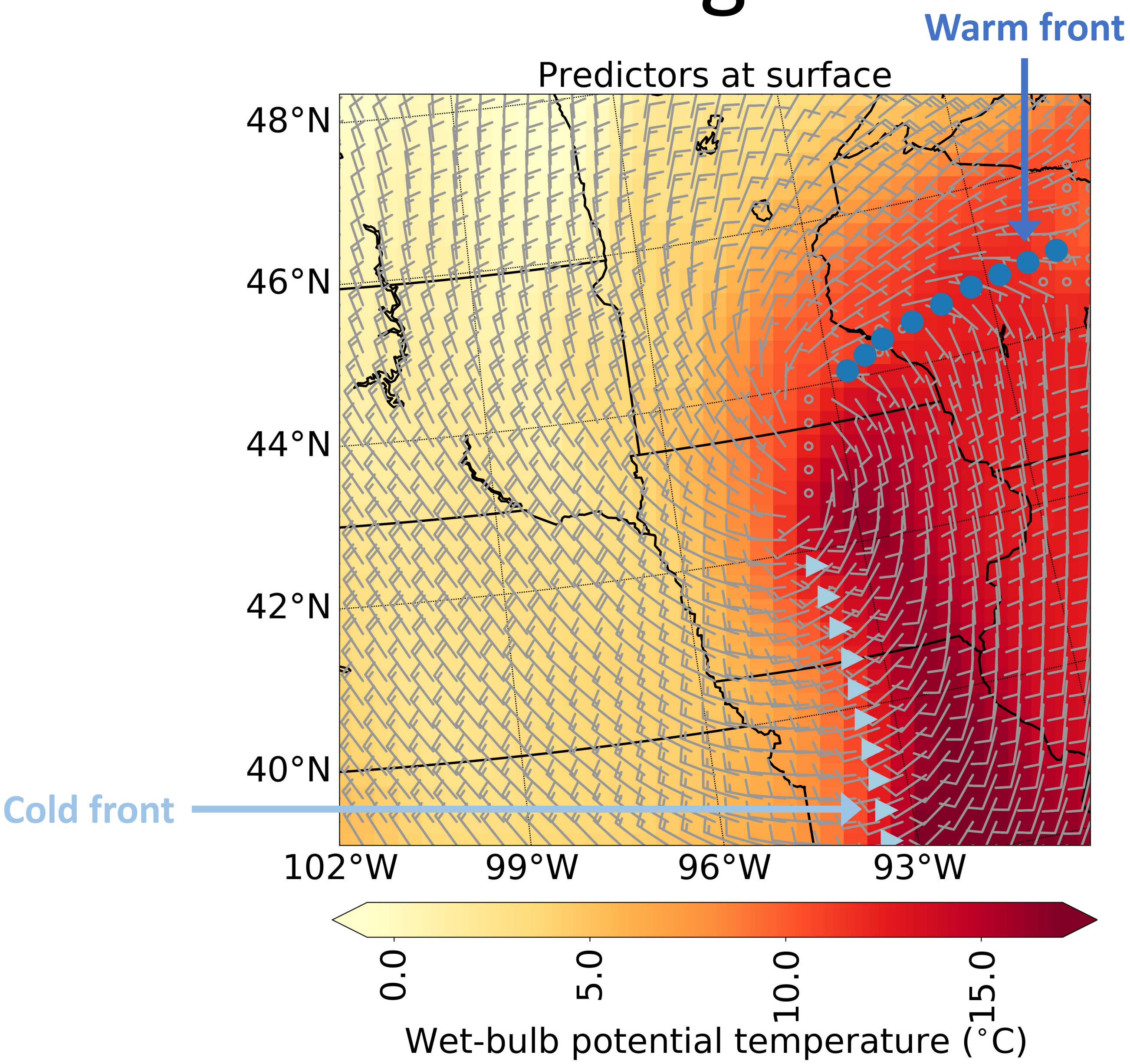
- **I use the following ERA5 variables at both the surface and 850 mb:**
  - Temperature
  - Specific humidity
  - Wind (*u* and *v*)

- Training: 2008-14
- Validation: 2015-16
- Testing: 2017



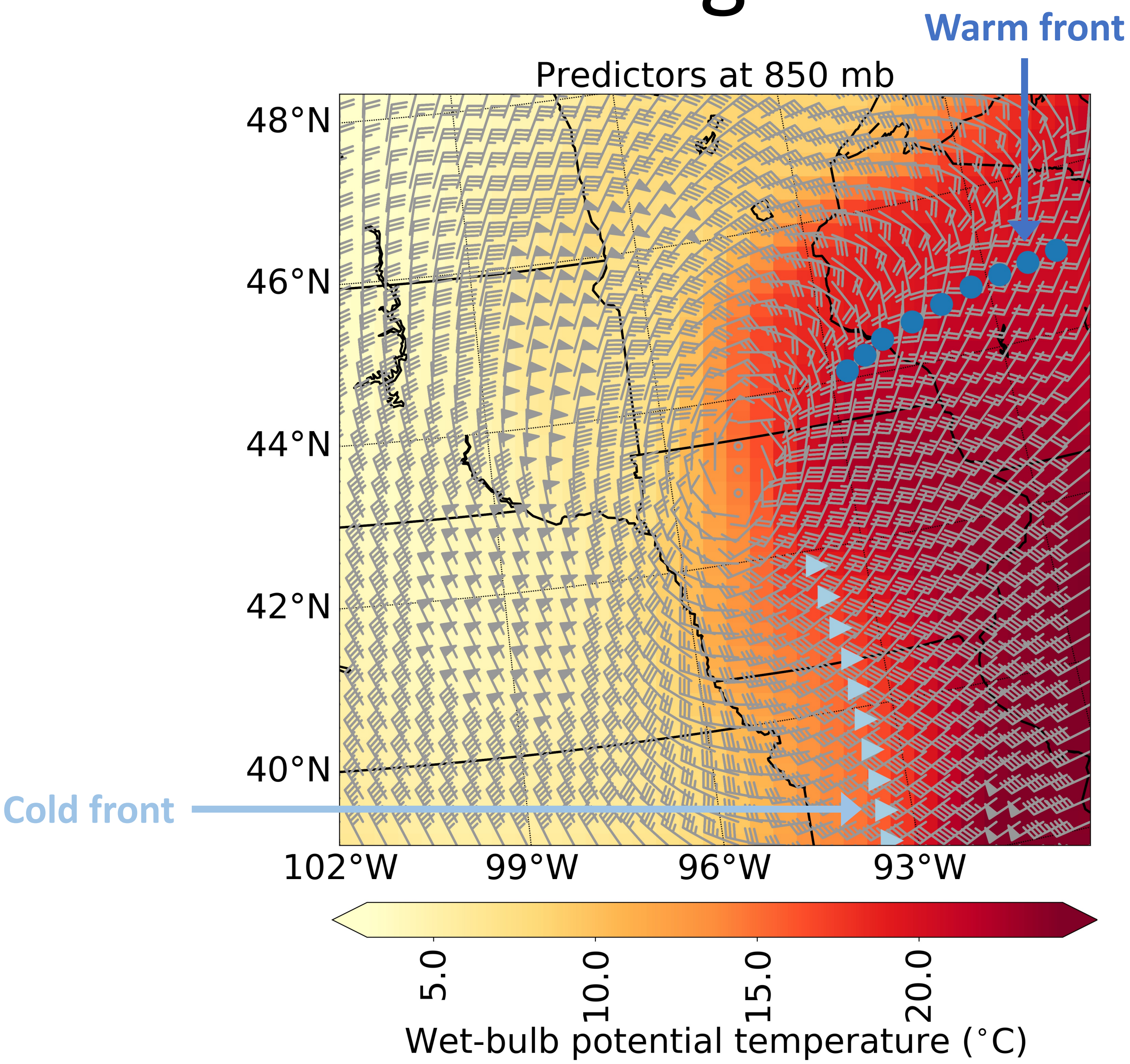
Front Detection: Machine Learning

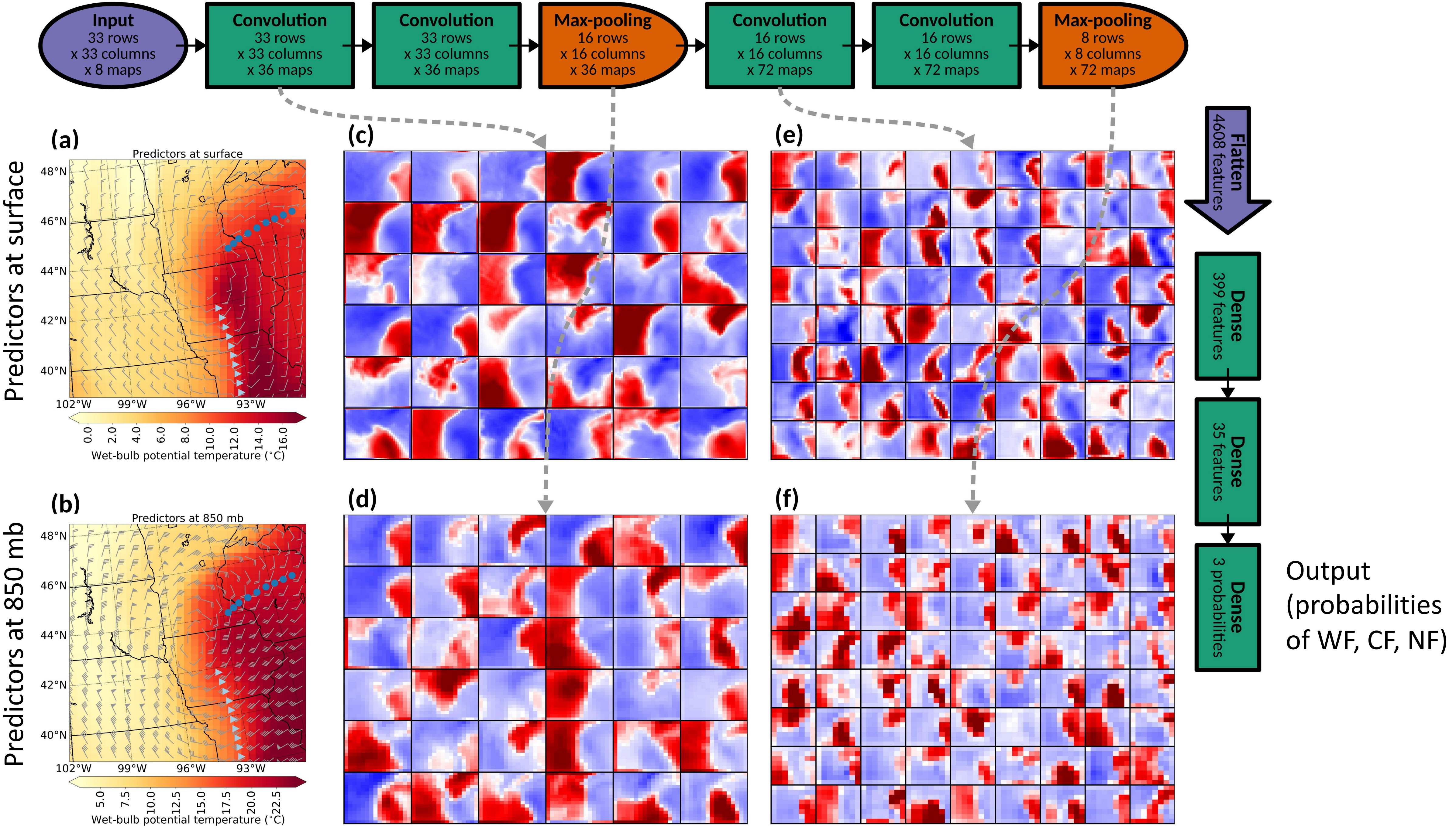
- One CNN input = small “patch” at one time step.
- Patch is 33 x 33 grid cells (1056 x 1056 km).
- **Label is based on type of front (if any) passing through center grid cell:**
  - Warm front (WF)
  - Cold front (CF)
  - Neither (NF)
- **I use grid search to optimize the following hyperparameters:**
  - Predictors (tried *u, v, T, q,* θ_w, *Z*)
  - Vertical levels (tried surface, 1000 mb, 950 mb, 900 mb, 850 mb)
  - Number of conv layers (tried values from 2-12)
- I do not use data augmentation for fronts (makes validation performance worse).



Front Detection: Machine Learning

- One CNN input = small "patch" at one time step.
- Patch is 33 x 33 grid cells (1056 x 1056 km).
- **Label is based on type of front (if any) passing through center grid cell:**
  - Warm front (WF)
  - Cold front (CF)
  - Neither (NF)
- **I use grid search to optimize the following hyperparameters:**
  - Predictors (tried *u, v, T, q,* θ_w, *Z*)
  - Vertical levels (tried surface, 1000 mb, 950 mb, 900 mb, 850 mb)
  - Number of conv layers (tried values from 2-12)
- I do not use data augmentation for fronts (makes validation performance worse).



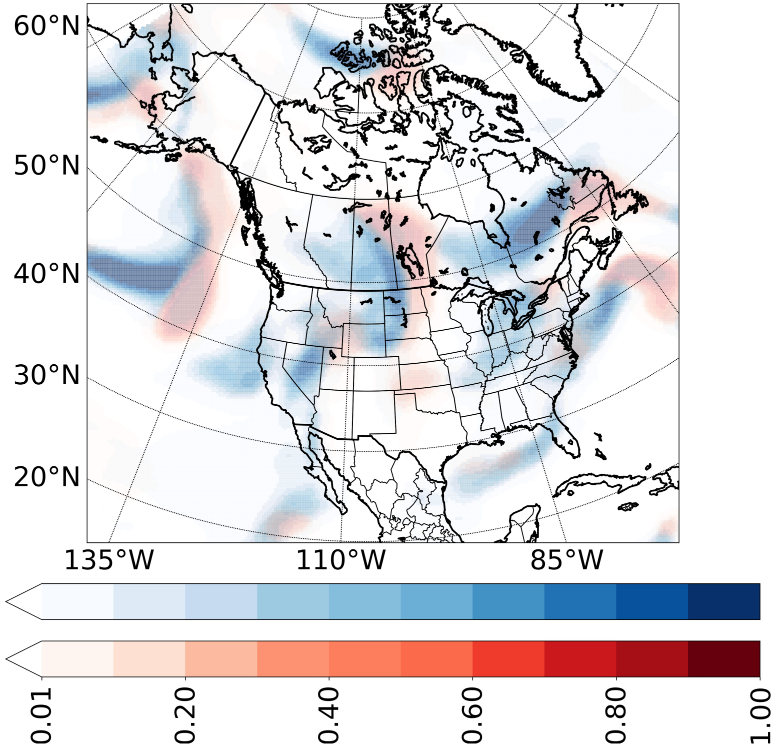


(c-f) Feature maps produced by conv and pooling layers

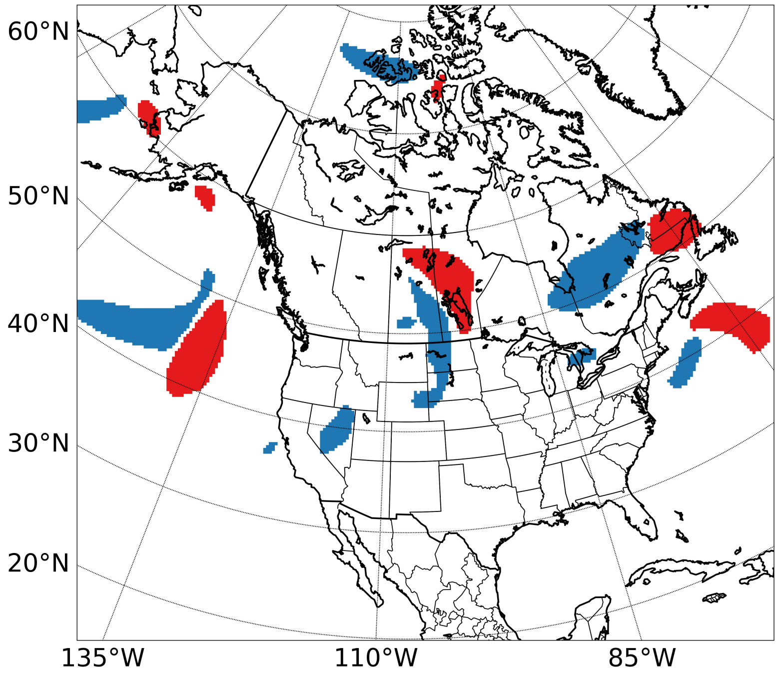
Front Detection: Machine Learning

- To apply trained CNN to full grid, slide 33 x 33 window around, centering on every grid cell.
- Before creating climatology, I convert probability fields to frontal zones, using method shown below.

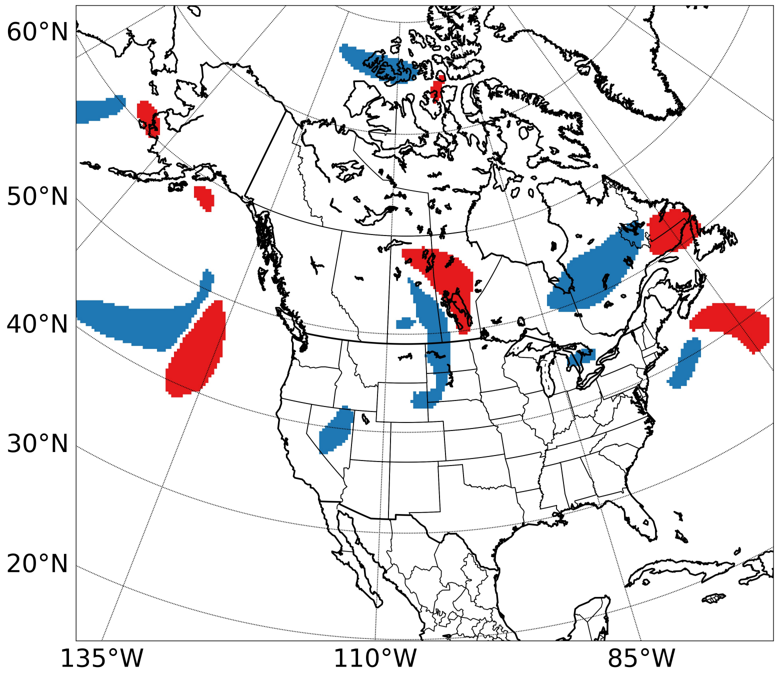
Probability fields (warm and cold)

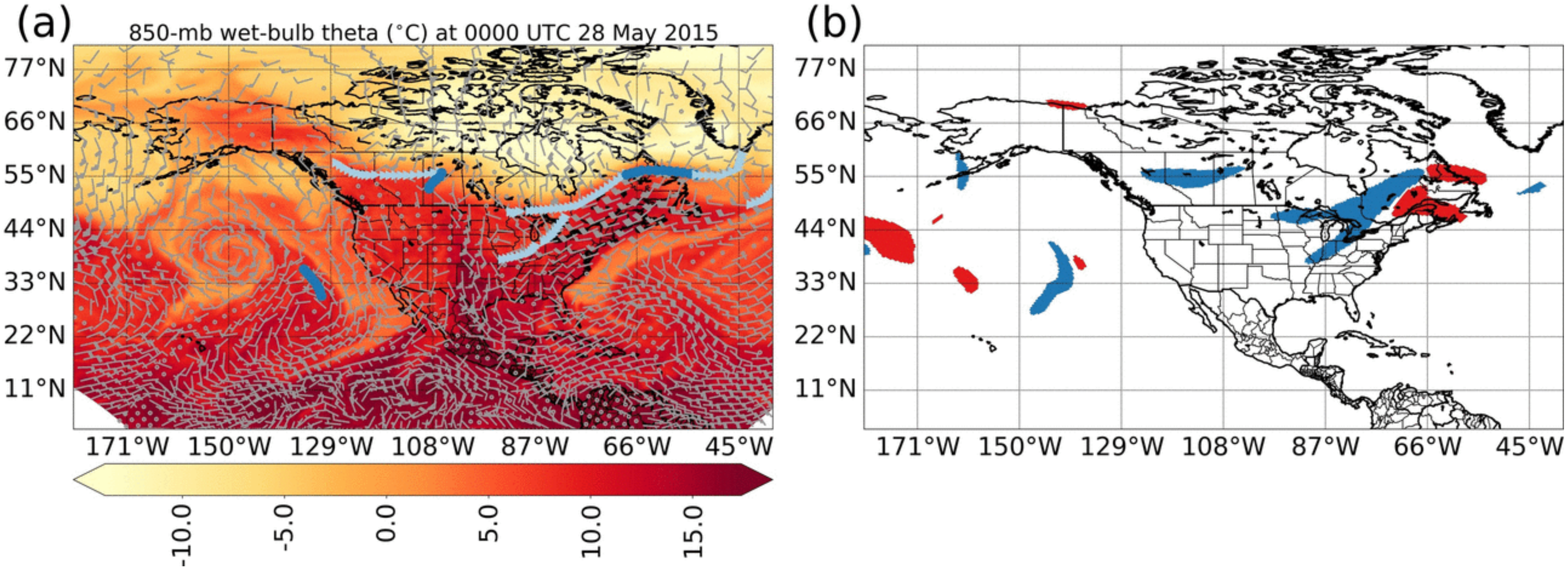


Apply probability threshold of 65%



Apply length threshold of 200 km




(a)
850-mb wet-bulb theta (°C) at 0000 UTC 28 May 2015
77°N
66°N
55°N
44°N
33°N
22°N
11°N
171°W
150°W
129°W
108°W
87°W
66°W
45°W
-10.0
-5.0
0.0
5.0
10.0
15.0
(b)

Front Detection: Climatology

- The climatology spans 40 years (1979 to 2018).

- **I will show the following analyses for both WF and CF frequency:**
  - Averages over the 40 years
  - Variability relative to the El Niño – Southern Oscillation (ENSO)
  - Trends over the 40 years

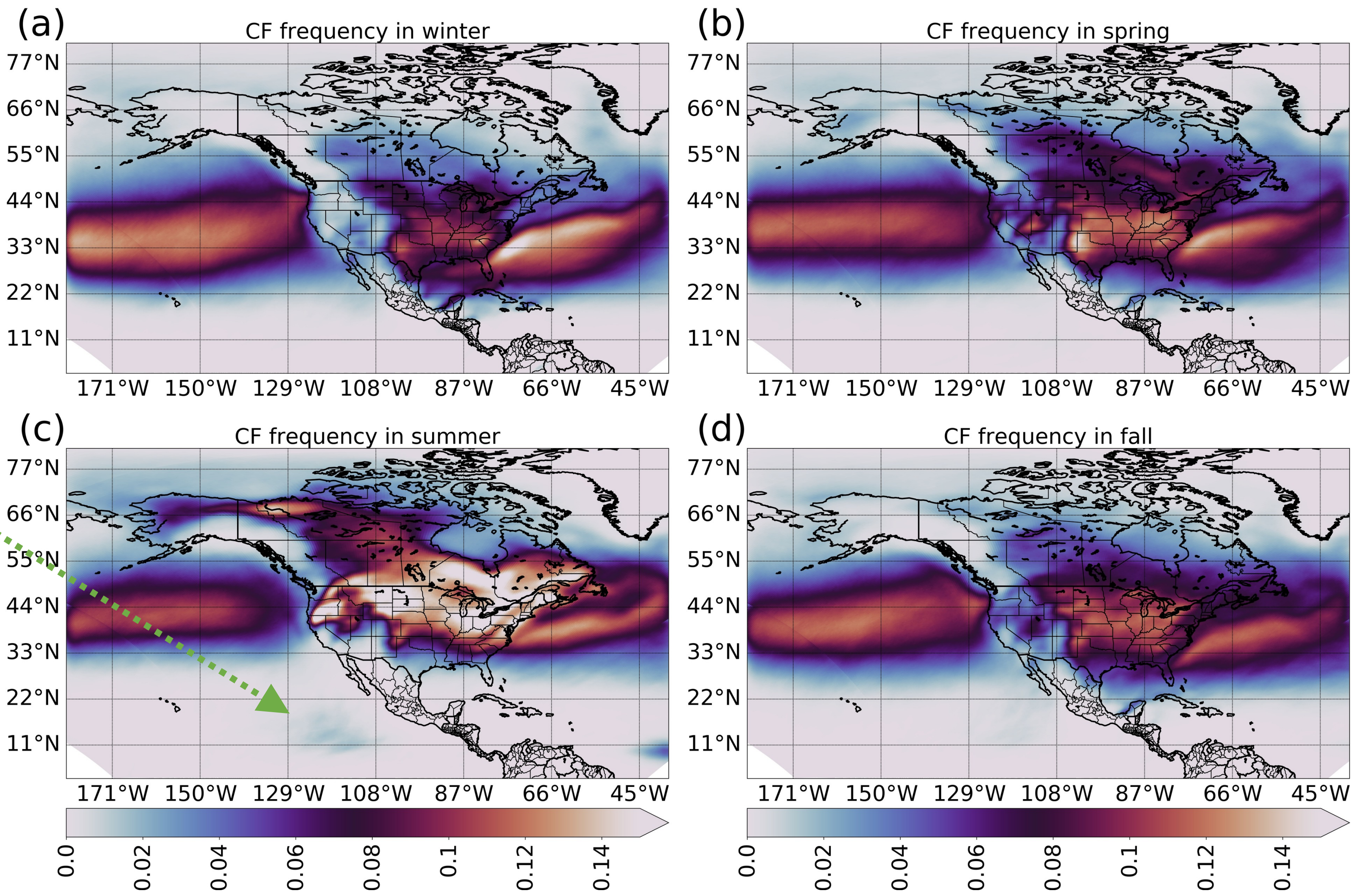
- ”Frequency” = percentage of time steps with a warm or cold front.

Average CF frequency by season

- **Cold fronts are most common in mid-latitude cyclone track**, especially over Pacific and Atlantic.

- **Mid-latitude cyclone track moves ~10° poleward from winter to summer**, due to global annual heating cycle.

- Summer cold fronts in tropical eastern Pacific due almost entirely due to moisture gradients (invasion of dry subtropical air).
  - Berry *et al.* (2011b) found similar max in eastern tropical Pacific and made the same conclusion.



- **Warm fronts are also most common in mid-latitude cyclone track.**

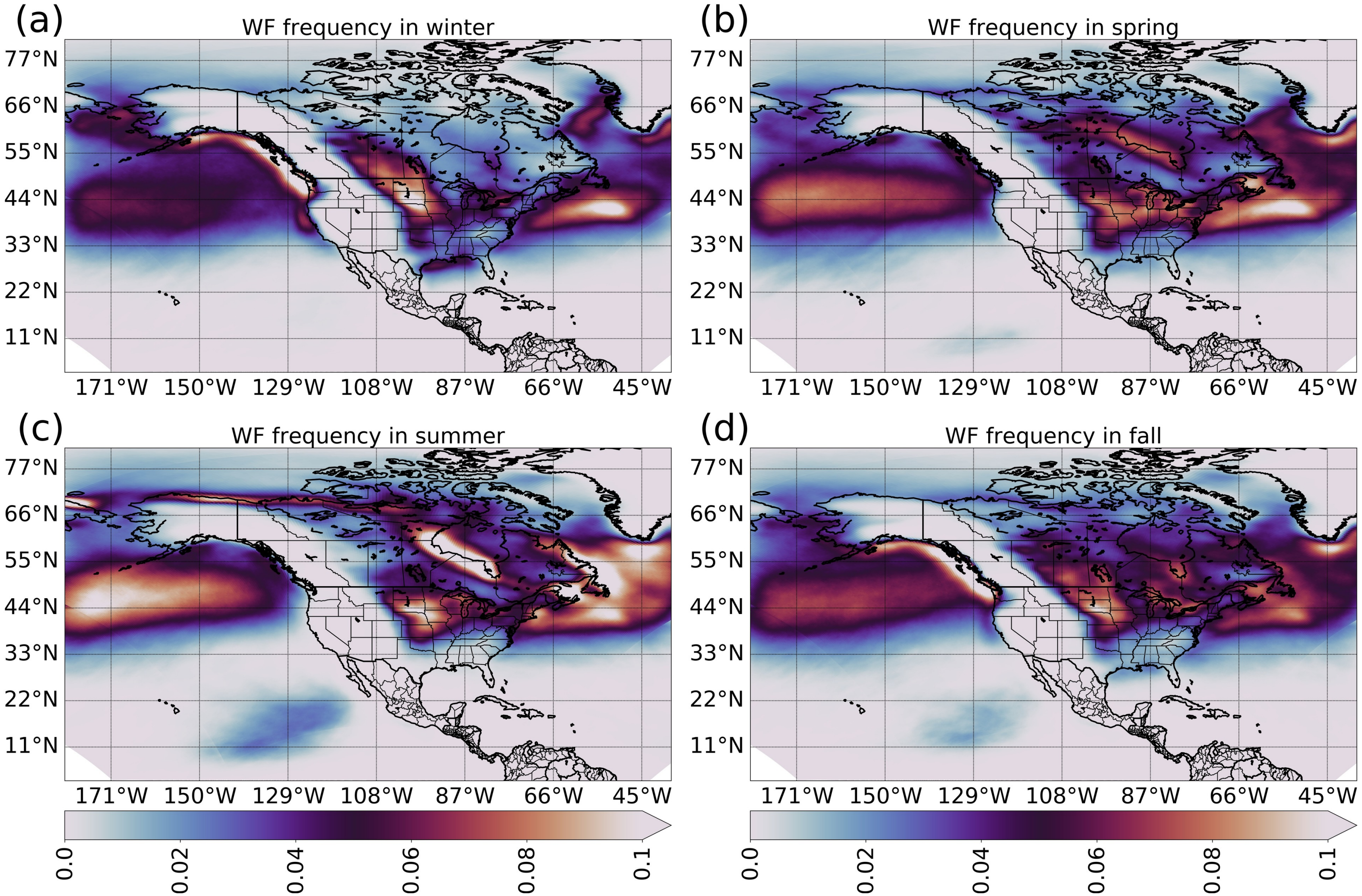
- **Large-scale WF maxima occur ~10° north of large-scale CF maxima,** due to mean frontal positions relative to parent cyclones.

- **Warm fronts occur at land-sea boundaries more often than cold fronts.**

- These occur only when there is warm advection across the boundary, so the CNN is **not** mistaking stationary fronts as warm fronts.

- The CNN has learned that cold fronts are typically stronger, so advection across land-sea boundary reaches WF threshold more often than CF threshold.

Average WF frequency by season



Front Climatology: ENSO-relative Variability

- ENSO is an irregular periodic variation in sea-surface temperature (SST) across the equatorial Pacific.
- The two phases are El Niño (warm eastern Pacific) and La Niña (cool eastern Pacific).

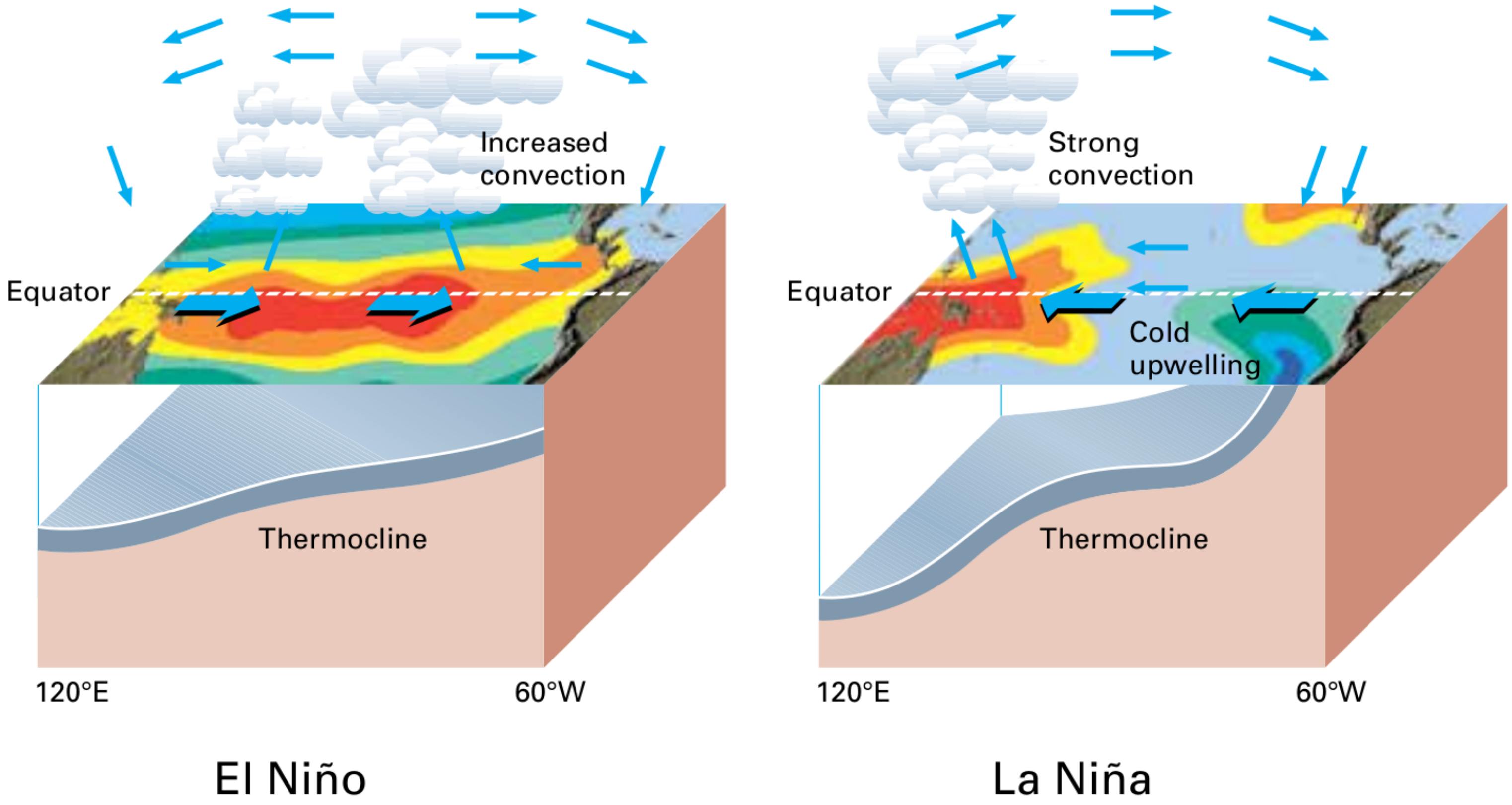


Image source: World Meteorological Organization (2014)

Front Climatology: ENSO-relative Variability

- **El Niño effects:**

⟹ More convection in eastern Pacific
⟹ Hadley cell strengthens and contracts
⟹ Subtropical jet shifts southward
⟹ Mid-latitude cyclone track shifts southward

- **La Niña effects:**

⟹ Roughly opposite (northward shift)

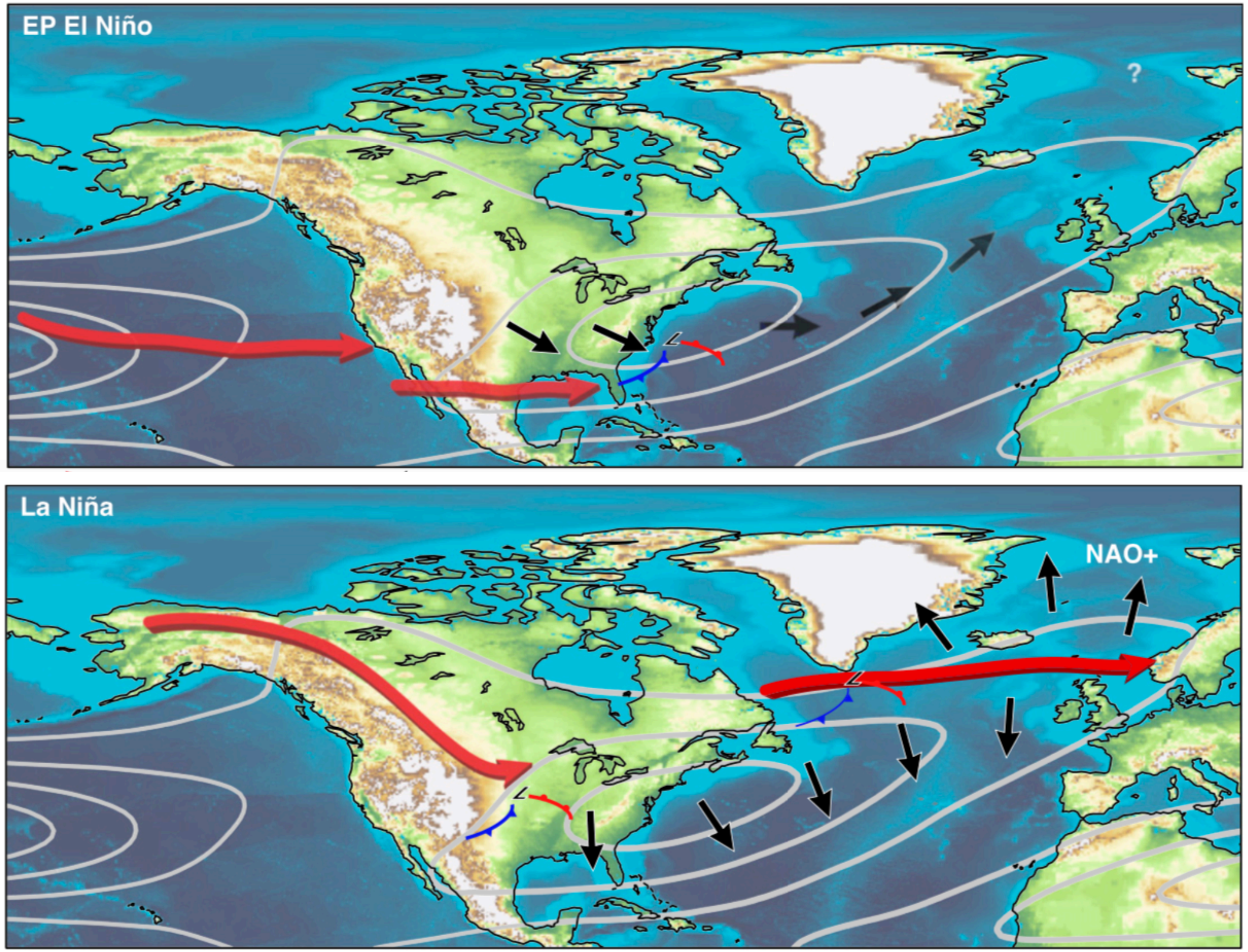


Image source: Figure 13 of Schemm *et al.* (2018)

Front Climatology: ENSO-relative Variability

- **I define ENSO phase by standardized anomaly of Niño 3.4 index** (*z*):
  - Neutral: $-0.5 < z < 0.5$
  - Strong El Niño: $z \geq 1$
  - Strong La Niña: $z \leq -1$

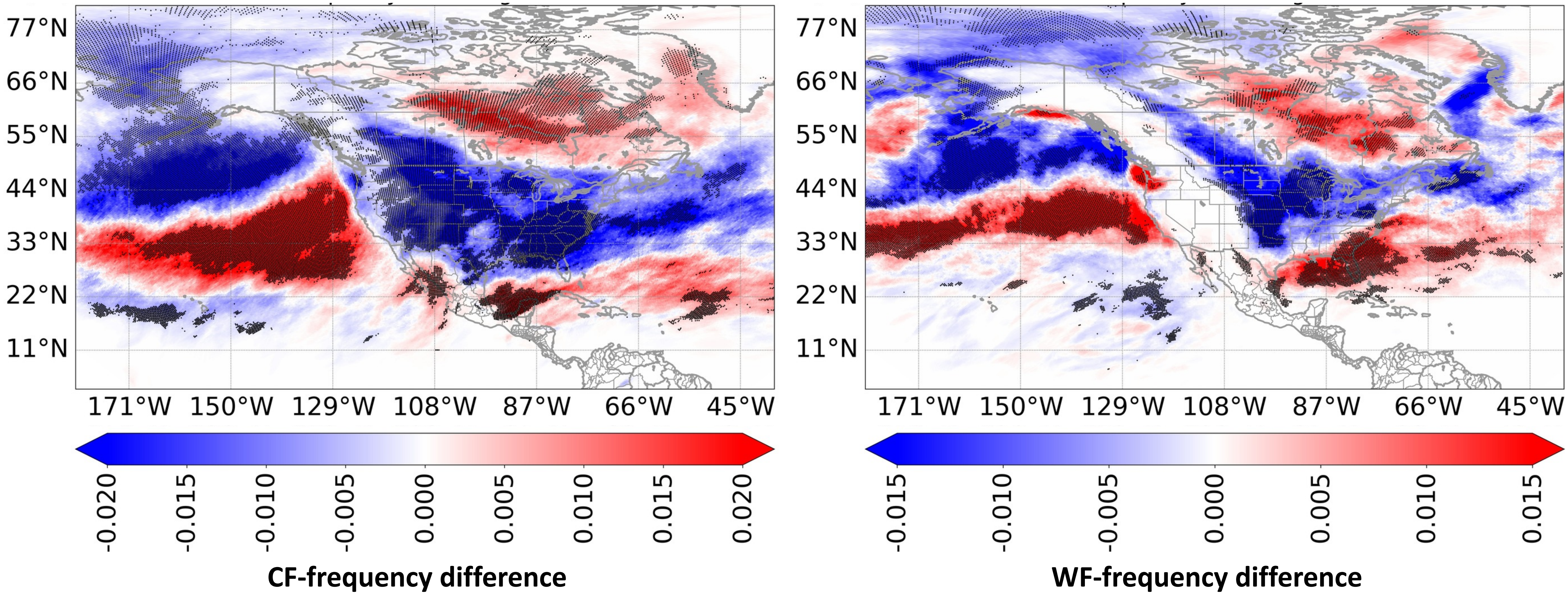
- **I will show the following differences for both WF and CF frequency:**
  - Strong El Niño minus in neutral phase
  - Strong La Niña minus in neutral phase

- **I use Monte Carlo test to find significant grid points.**
  - Two-tailed test, 20 000 shuffling iterations, 95% confidence level
  - I shuffle entire spatial maps together to control false-discovery rate

- I will focus on winter and spring, when ENSO teleconnections are strongest.

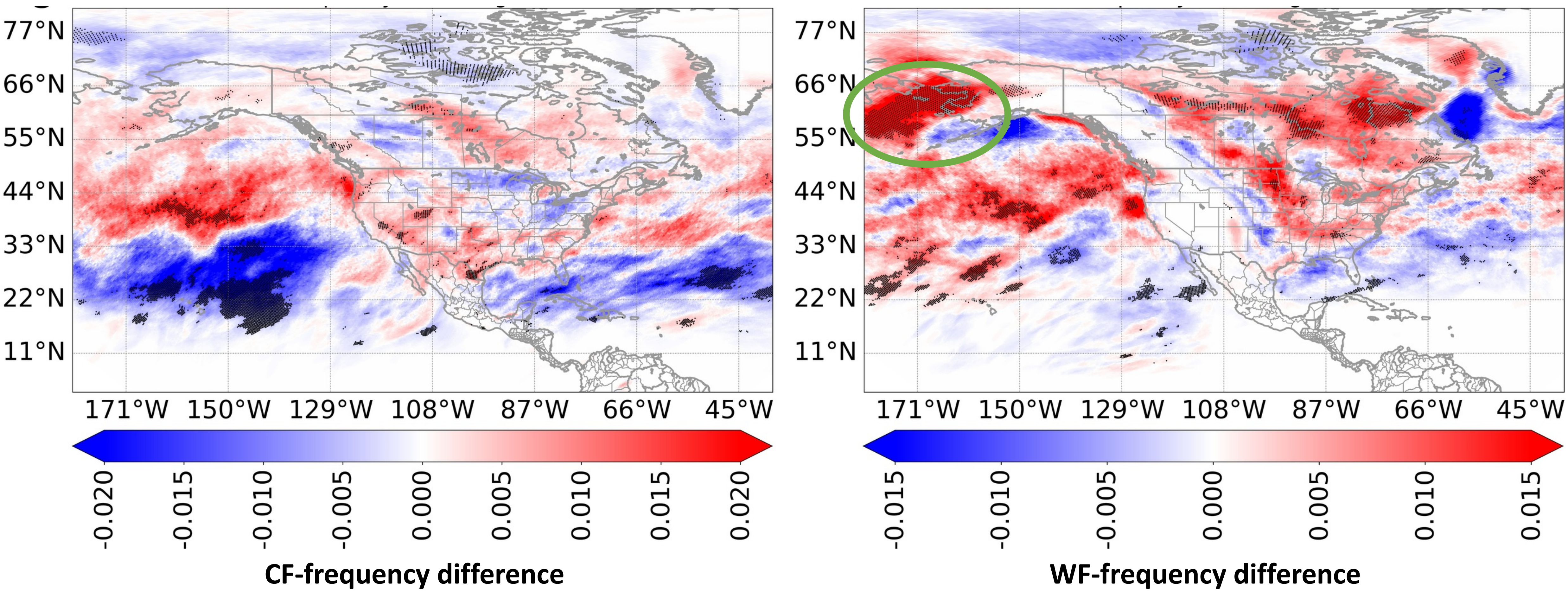
- **Southward shift in activity is consistent with southward shift in subtropical jet and cyclone track.**
- **Increased WF and CF frequency over Gulf of Mexico** are consistent with eastward extension of subtropical jet.
- Hardy and Henderson (2003) found similar pattern but without significance.
- **Increased WF and CF frequency over Hudson Bay** maybe due to anomalous polar jet stream during El Niño.

Strong El Niño in winter



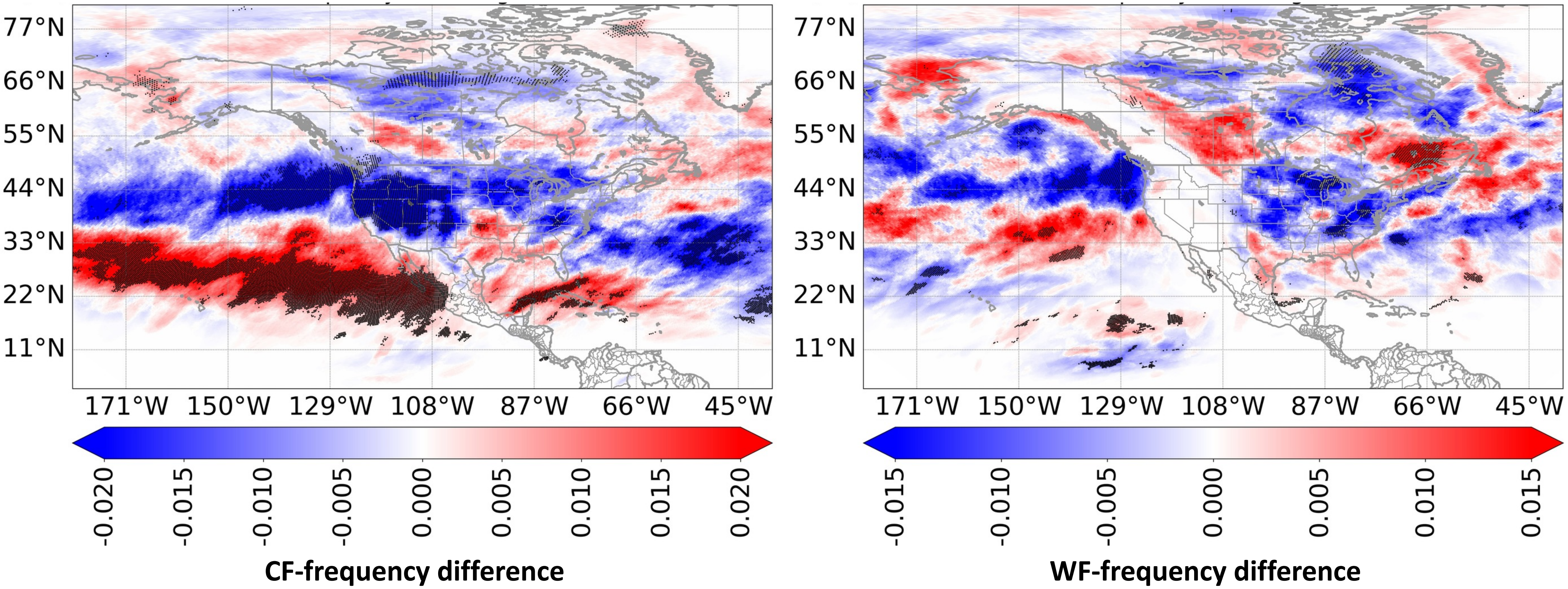
- **Northward shift in activity is consistent with northward shift in subtropical jet and cyclone track.**
- **Increased WF frequency over Bering Sea** (with decrease along south coast of Alaska) could be due to La Niña shifting position of Aleutian low westward (Niebauer 1998).
- **La Niña results weaker and less significant than El Niño**, because La Niña is less common and has weaker teleconnections.
- **Results for winter El Niño and La Niña are broadly consistent with previous climatology** (Rudeva and Simmonds 2015).

Strong La Niña in winter



- **Overall, winter and spring responses to El Niño are similar (southward shift).**
- **Exceptions: spring response is weaker, and significant grid pts cover smaller range of latitudes.**
- This is because spring has:
  - Weaker SST anomalies
  - Weaker westerly wind connecting mid-latitudes to tropical heat source

Strong El Niño in spring



Front Climatology: Long-term Trends

- **Expected effects of global warming:**
  - Poleward expansion of Hadley cell (Davis and Rosenlof 2012; Lucas *et al.* 2014; Schmidt and Grise 2017)

  $\Longrightarrow$ Poleward shift of subtropical jet and mid-latitude cyclone track

  $\Longrightarrow$ **Poleward shift of front activity**

  - Arctic amplification (Serreze and Barry 2011)

  $\Longrightarrow$ Weaker temperature gradient at high latitudes

  $\Longrightarrow$ **Fewer fronts at high latitudes**

- **For each season I compute linear trend in WF and CF frequency.**

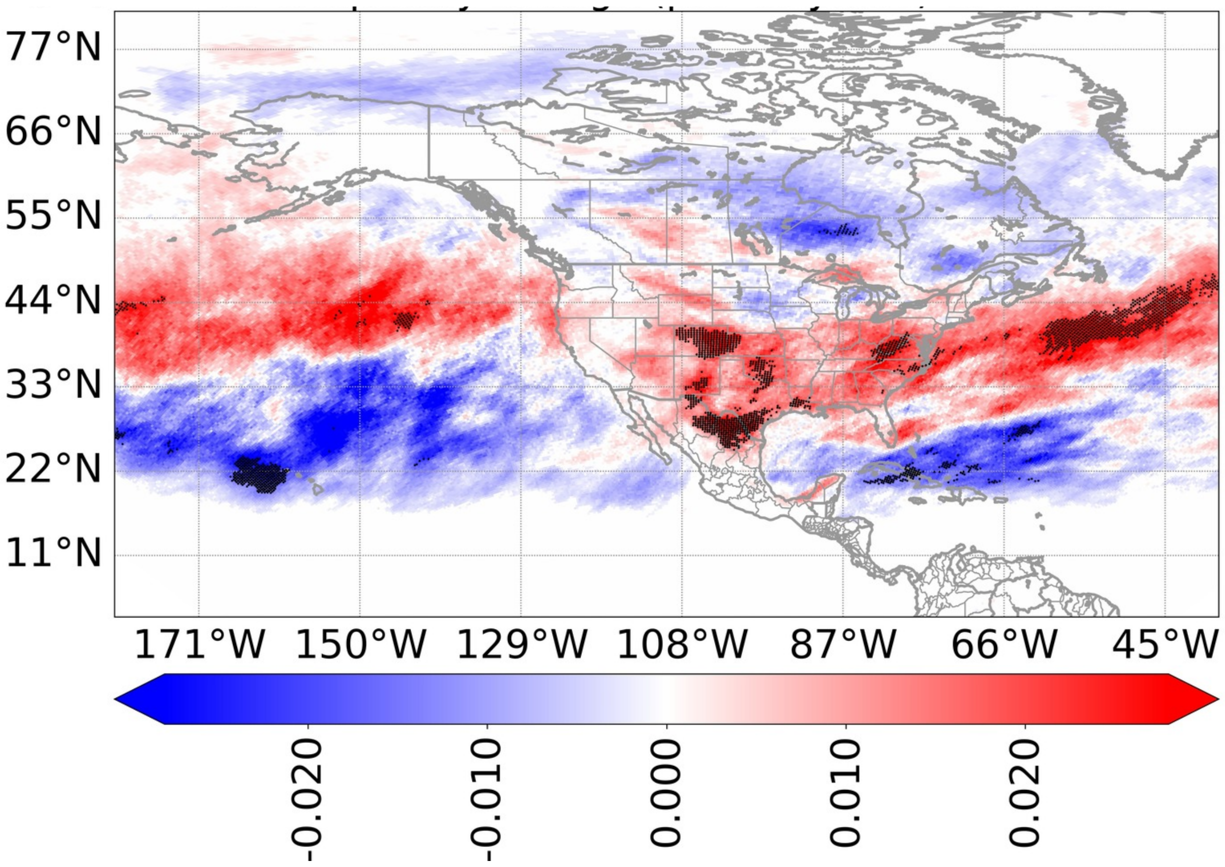
- **I use Mann-Kendall test to find significant grid points.**
  - I use Equation 3 of Wilks (2016) to keep false-discovery rate below 10%
  - However, this method is overly conservative, leading to p-value threshold < 0.015

- Due to lack of significance in other seasons, I will show winter only.

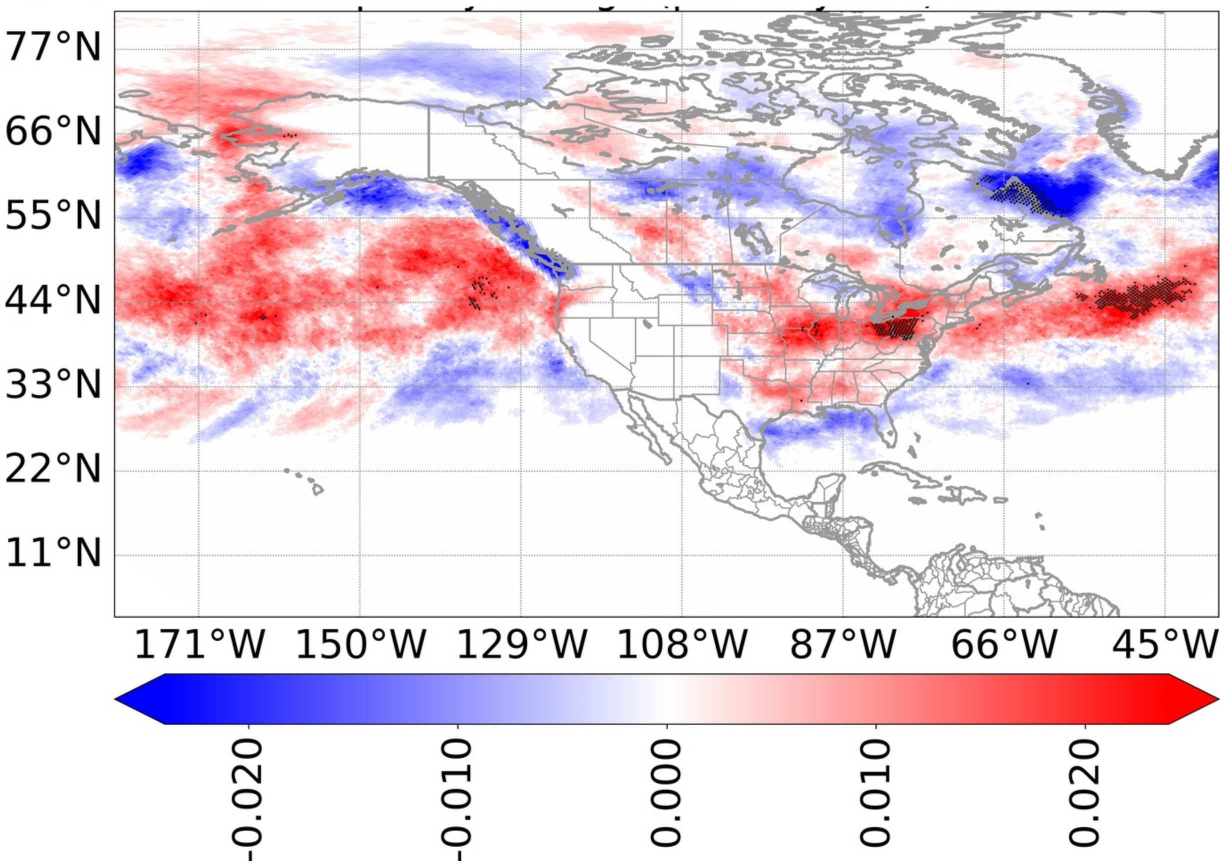
- **Northward shift in activity is consistent with northward shift in subtropical jet and cyclone track.**
- **Decreased WF and CF frequency over Arctic are consistent with loss of baroclinicity due to Arctic amplification.**

- Two previous climos (Rudeva and Simmonds 2015; Berry *et al.* 2011a) found the same patterns but with more significance.
- However, Berry *et al.* (2011a) found general decrease over Atlantic, rather than northward shift.
- Could learn more by applying CNN to climate-model output.

Frequency trend (per 40 years) in winter



Cold fronts



Warm fronts

Summary and Future Work

- I developed and tested CNNs for two tasks: next-hour tornado prediction and front detection.

- **Tornado models perform competitively with operational model.**

- Failure modes are non-tornadic supercells and tornadic QLCS cells (difficult for humans as well).

- **CNN-interpretation methods highlight physical relationships involving:**
  - Depth of reflectivity core
  - Strength and compactness of low-level mesocyclone
  - Discreteness of storm
  - Reflectivity in rear-flank downdraft

- **Future work:**
  - Operationalizing for Hazardous Weather Testbed
  - Comparing human vs. CNN interpretations
  - Improving performance for QLCS tornadoes
  - Using interpretation methods to guide discovery of new knowledge (like Wagstaff and Lee 2018 for Mars rovers)

- **Papers:** McGovern *et al.* (2019); McGovern *et al.* (2020); Lagerquist *et al.* (2020b, conditionally accepted)

Summary and Future Work

- Front detection: **trained CNN to draw warm and cold fronts in reanalysis data.**

- **Created and analyzed 40-year climatology over North America:**
  - Fronts most common in mid-latitude cyclone track
  - These fronts shift equatorward with El Niño, poleward with La Niña, may be shifting poleward over long term
  - Results generally consistent with previous climos that investigate ENSO and long-term change (Berry *et al.* 2011a,b; Rudeva and Simmonds 2015)
  - Some results need more investigation (*e.g.,* long-term trend in Atlantic)

- **Future work:**
  - Operationalize for use by forecasters
  - Investigate front activity in future climate
  - Investigate climatology of front-related extreme weather

- **Papers:** Lagerquist *et al.* (2019); Lagerquist *et al.* (2020a, conditionally accepted)

Acknowledgements

- Amy McGovern
- David John Gagne II
- John Allen
- Cameron Homeyer
- Brian Williams
- Friends and family
- WxTwitter and AcademicTwitter communities
- All of you!

Hypatia

Eratosthenes and Enheduanna

References

Adrianto, I., T. Trafalis, and V. Lakshmanan, 2009: "Support vector machines for spatiotemporal tornado prediction." *International Journal of General Systems*, **38 (7)**, 759–776, URL https://doi.org/10.1080/03081070601068629.

Anderson-Frey, A., Y. Richardson, A. Dean, R. Thompson, and B. Smith, 2016: "Investigation of near-storm environments for tornado events and warnings." *Weather and Forecasting*, **31 (6)**, 1771–1790, URL https://doi.org/10.1175/WAF-D-16-0046.1.

Benjamin, S., and Coauthors, 2016: "A North American hourly assimilation and model forecast cycle: The Rapid Refresh." *Monthly Weather Review*, **144 (4)**, 1669–1694, URL https://doi.org/10.1175/MWR-D-15-0242.1.

Berry, G., C. Jakob, and M. Reeder, 2011a: "Recent global trends in atmospheric fronts." *Geophysical Research Letters*, **38 (21)**, URL https://doi.org/10.1029/2011GL049481.

Berry, G., M. Reeder, and C. Jakob, 2011b: "A global climatology of atmospheric fronts." *Geophysical Research Letters*, **38 (4)**, URL https://doi.org/10.1029/2010GL046451.

Brooks, H., and J. Correia, 2018: "Long-term performance metrics for National Weather Service tornado warnings." *Weather and Forecasting*, **33 (6)**, 1501–1511, URL https://doi.org/10.1175/WAF-D-18-0120.1.

Brotzge, J., S. Nelson, R. Thompson, and B. Smith, 2013: "Tornado probability of detection and lead time as a function of convective mode and environmental parameters." *Weather and Forecasting*, **28 (5)**, 1261–1276, URL https://doi.org/10.1175/WAF-D-12-00119.1.

Cintineo, J., M. Pavolonis, J. Sieglaff, and D. Lindsey, 2014: "An empirical model for assessing the severe weather potential of developing convection." *Weather and Forecasting*, **29 (3)**, 639–653, URL https://doi.org/10.1175/WAF-D-13-00113.1.

Cintineo, J., and Coauthors, 2018: "The NOAA/CIMSS ProbSevere Model: Incorporation of total lightning and validation." *Weather and Forecasting*, **33 (1)**, 331–345, URL https://doi.org/10.1175/WAF-D-17-0099.1.

Davis, S., and K. Rosenlof, 2012: "A multidiagnostic intercomparison of tropical-width time series using reanalyses and satellite observations." *Journal of Climate*, **25 (4)**, 1061–1078, URL https://doi.org/10.1175/JCLI-D-11-00127.1.

References

Dieleman, S., K. Willett, and J. Dambre, 2015: “Rotation-invariant convolutional neural networks for galaxy morphology prediction.” *Nature*, **450 (2)**, 1441–1459, URL https://doi.org/10.1093/mnras/stv632.

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

Erhan, D., Y. Bengio, A. Courville, and P. Vincent, 2009: “Visualizing higher-layer features of a deep network.” Tech. rep., link.

Gagne, D., A. McGovern, J. Basara, and R. Brown, 2012: “Tornadic supercell environments analyzed using surface and reanalysis data: A spatiotemporal relational data-mining approach.” *Journal of Applied Meteorology and Climatology*, **51 (12)**, 2203–2217, URL https://doi.org/10.1175/JAMC-D-11-060.1.

Gagne, D., S. Haupt, and D. Nychka, 2019: “Interpretable deep learning for spatial analysis of severe hailstorms.” *Monthly Weather Review*, **147 (8)**, 2827–2845, URL https://doi.org/10.1175/MWR-D-18-0316.1.

Gil, Y., and Coauthors, 2019: “Intelligent systems for geosciences: An essential research agenda.” *Communications of the Association for Computing Machinery*, **62 (1)**, 76–84, URL https://dl.acm.org/doi/10.1145/3192335.

Hardy, J., and K. Henderson, 2003: “Cold front variability in the southern United States and the influence of atmospheric teleconnection patterns.” *Physical Geography*, **24 (2)**, 120–137, URL https://doi.org/10.2747/0272-3646.24.2.120.

Hersbach, H., and D. Dee, 2016: “ERA5 reanalysis is in production.” *ECMWF Newsletter*, **147 (7)**, 5–6, URL https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production.

Hewson, T., 1998: “Objective fronts.” *Meteorological Applications*, **5 (1)**, 37–65, URL https://doi.org/10.1017/S1350482798000553.

Homeyer, C., and K. Bowman, 2017: “Algorithm Description Document for Version 3.1 of the Three-Dimensional Gridded NEXRAD WSR-88D Radar (GridRad) Dataset.” Tech. rep., University of Oklahoma. URL http://gridrad.org/pdf/GridRad-v3.1-Algorithm-Description.pdf.

References

Insurance Information Institute, 2019: “Facts + Statistics: Tornadoes and Thunderstorms.” URL https://www.iii.org/fact-statistic/facts-statistics-tornadoes-and-thunderstorms.

Karras, T., T. Aila, S. Laine, and J. Lehtinen, 2018: “Progressive growing of GANs for improved quality, stability, and variation.” *arXiv pre-prints*, **1710 (10196v3)**, URL https://arxiv.org/abs/1710.10196.

Kitzmiller, D., W. McGovern, and R. Saffle, 1995: “The WSR-88D severe weather potential algorithm.” *Weather and Forecasting*, **10 (1)**, 141–159, link.

Krizhevsky, A., I. Sutskever, and G. Hinton, 2012: “Imagenet classification with deep convolutional neural networks.” *Advances in Neural Information Processing Systems*, 1097–1105, URL http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networ.

Kunkel, K., J. Biard, and E. Racah, 2018: “Automated detection of fronts using a deep learning algorithm.” *Conference on Artificial and Computational Intelligence and its Applications to the Environmental Sciences*, Austin, Texas, American Meteorological Society, URL https://ams.confex.com/ams/98Annual/meetingapp.cgi/Paper/333480.

Kurth, T., and Coauthors, 2018: “Exascale deep learning for climate analytics.” *International Conference for High Performance Computing, Networking, Storage, and Analysis*, Dallas, Texas, Institute of Electrical and Electronics Engineers (IEEE), URL https://doi.org/10.1109/SC.2018.00054.

Lagerquist, R., A. McGovern, and D. Gagne, 2019: “Deep learning for spatially explicit prediction of synoptic-scale fronts.” *Weather and Forecasting*, **34 (4)**, 1137–1160, URL https://doi.org/10.1175/WAF-D-18-0183.1.

Lagerquist, R., J. Allen, and A. McGovern, 2020a: “Climatology and variability of warm and cold fronts over North America from 1979-2018.” *Journal of Climate*, **conditionally accepted**.

Lagerquist, R., A. McGovern, C. Homeyer, D. Gagne, and T. Smith, 2020b: “Deep learning on three-dimensional multiscale data for next-hour tornado prediction.” *Monthly Weather Review*, **conditionally accepted.**

Lakshmanan, V., I. Adrianto, T. Smith, and G. Stumpf, 2005: “A spatiotemporal approach to tornado prediction.” *IEEE International Joint Conference on Neural Networks*, **3**, 1642–1647, URL https://doi.org/10.1109/IJCNN.2005.1556125.

References

Liu, Y., and Coauthors, 2016: “Application of deep convolutional neural networks for detecting extreme weather in climate datasets.” *arXiv pre-prints*, **1605 (01156)**, URL https://arxiv.org/abs/1605.01156.

Lucas, C., B. Timbal, and H. Nguyen, 2014: “The expanding tropics: A critical assessment of the observational and modeling studies.” *Wiley Interdisciplinary Reviews: Climate Change*, **5 (1)**, 89–112, URL https://doi.org/10.1002/wcc.251.

Lutgens, F., and E. Tarbuck, 2000: “The Atmosphere: An Introduction to Meteorology,” Vol. 8, Prentice Hall.

Markowski, P., J. Straka, and E. Rasmussen, 2002: “Direct surface thermodynamic observations within the rear-flank downdrafts of nontornadic and tornadic supercells.” *Monthly Weather Review*, **130 (7)**, 1692–1721, link.

Markowski, P., and Y. Richardson, 2009: “Tornadogenesis: Our current understanding, forecasting considerations, and questions to guide future research.” *Atmospheric Research*, **93 (1-3)**, 3–10, URL https://doi.org/10.1016/j.atmosres.2008.09.015.

Marzban, C., and G. Stumpf, 1996: “A neural network for tornado prediction based on Doppler radar-derived attributes.” *Journal of Applied Meteorology*, **35 (5)**, 617–626, link.

McGovern, A., R. Lagerquist, D. Gagne, G. Jergensen, K. Elmore, C. 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, URL https://doi.org/10.1175/BAMS-D-18-0195.1.

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*, **accepted**.

Niebauer, H., 1998: “Variability in Bering Sea ice cover as affected by a regime shift in the North Pacific in the period 1947–1996.” *Journal of Geophysical Research: Oceans*, **103 (C12)**, 27717–27737, URL https://doi.org/10.1029/98JC02499.

Olah, C., A. Mordvintsev, and L. Schubert, 2017: “Feature visualization.” *Distill*, URL https://distill.pub/2017/feature-visualization.

References

Ortega, K., T. Smith, J. Zhang, C. Langston, Y. Qi, S. Stevens, and J. Tate, 2012: "The Multi-year Reanalysis of Remotely Sensed Storms (MYRORSS) project." *Conference on Severe Local Storms*, Nashville, Tennessee, American Meteorological Society, URL https://ams.confex.com/ams/26SLS/webprogram/Handout/Paper211413/p4_74_ortegaetal_myrorss.pdf.

Racah, E., C. Beckham, T. Maharaj, S. Kahou, Prabhat, and C. Pal, 2017: "ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events." *Advances in Neural Information Processing Systems*, Long Beach, California, Neural Information Processing Systems, URL http://papers.nips.cc/paper/6932-extremeweather-a-large-scale-climate-dataset-for-semi-supervised-de.

Rakhlin, A., A. Shvets, V. Iglovikov, and A. Kalinin, 2018: "Deep convolutional neural networks for breast cancer histology image analysis." *arXiv pre-prints*, **1802**, URL https://doi.org/10.1007/978-3-319-93000-8_83.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat, 2019: "Deep learning and process understanding for data-driven Earth system science." *Nature*, **566**, 195–204, URL https://doi.org/10.1038/s41586-019-0912-1.

Rudeva, I., and I. Simmonds, 2015: "Variability and trends of global atmospheric frontal activity and links with large-scale modes of variability." *Journal of Climate*, **28 (8)**, 3311–3330, URL https://doi.org/10.1175/JCLI-D-14-00458.1.

Schemm, S., I. Rudeva, and I. Simmonds, 2015: "Extratropical fronts in the lower troposphere – global perspectives obtained from two automated methods." *Quarterly Journal of the Royal Meteorological Society*, **141 (690)**, 1686–1698, URL https://doi.org/10.1002/qj.2471.

Schemm, S., G. Rivière, L. Ciasto, and C. Li, 2018: "Extratropical cyclogenesis changes in connection with tropospheric ENSO teleconnections to the north Atlantic: Role of stationary and transient waves." *Journal of the Atmospheric Sciences*, **75 (11)**, 3943– 3964, URL https://doi.org/10.1175/JAS-D-17-0340.1.

Schmidt, D., and K. Grise, 2017: "The response of local precipitation and sea level pressure to Hadley cell expansion." *Geophysical Research Letters*, **44 (20)**, 10573–10582, URL https://doi.org/10.1002/2017GL075380.

Serreze, M., and R. Barry, 2011: "Processes and impacts of Arctic amplification: A research synthesis." *Global and Planetary Change*, **77 (1-2)**, 85–96, URL https://doi.org/10.1016/j.gloplacha.2011.03.004.

References

Silver, D., and Coauthors, 2016: “Mastering the game of Go with deep neural networks and tree search.” *Nature*, **529 (7587)**, 484–489, URL https://doi.org/10.1038/nature16961.

Silver, D., and Coauthors, 2017: “Mastering the game of Go without human knowledge.” *Nature*, **550 (7676)**, 354–359, URL https://doi.org/10.1038/nature24270.

Simonyan, K., A. Vedaldi, and A. Zisserman, 2014: “Deep inside convolutional networks: Visualizing image classification models and saliency maps.” *arXiv pre-prints*, **1312**, URL https://arxiv.org/abs/1312.6034.

Suwajanakorn, S., S. Seitz, and I. Kemelmacher-Shlizerman, 2017: “Synthesizing Obama: Learning lip sync from audio.” *ACM Transactions on Graphics*, **36 (4)**, URL https://doi.org/10.1145/3072959.3073640.

Wagstaff, K., and J. Lee, 2018: “Interpretable discovery in large image data sets.” *arXiv pre-prints*, **1806**, URL https://arxiv.org/abs/1806.08340.

Wilson, K., P. Heinselman, C. Kuster, D. Kingfield, and Z. Kang, 2017: “Forecaster performance and workload: Does radar update time matter?” *Weather and Forecasting*, **32 (1)**, 253–274, URL https://doi.org/10.1175/WAF-D-16-0157.1.

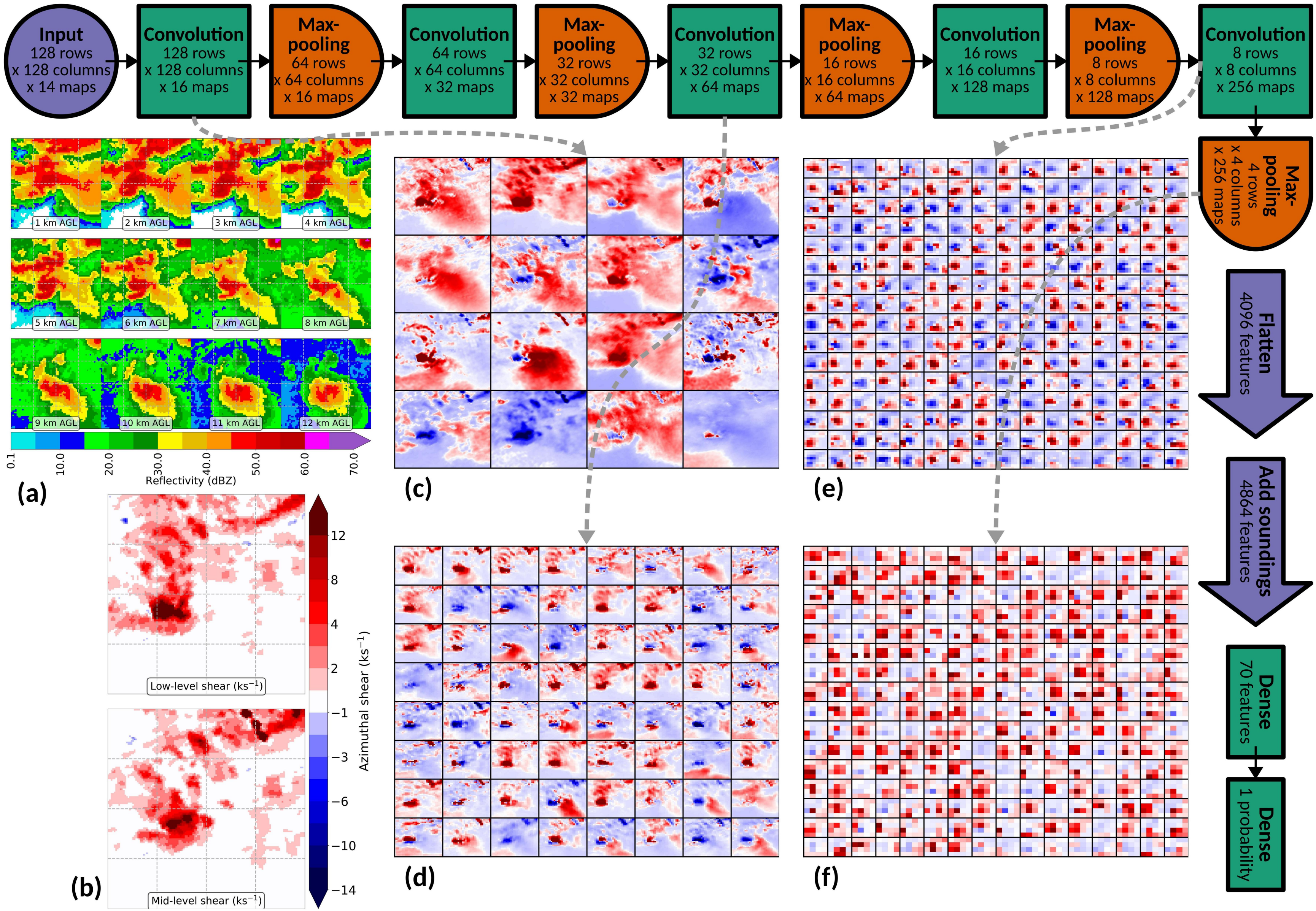
Wimmers, A., C. Velden, and J. Cossuth, 2019: “Using deep learning to estimate tropical cyclone intensity from satellite passive microwave imagery.” *Monthly Weather Review*, **147 (6)**, 2261–2282, URL https://doi.org/10.1175/MWR-D-18-0391.1.

World Meteorological Organization, 2014: “El Niño / Southern Oscillation.” Tech. rep. Link.

Zhao, M., T. Li, M. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi, 2018: “Through-wall human pose estimation using radio signals.” *Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, IEEE, URL http://openaccess.thecvf.com/content_cvpr_2018/html/Zhao_Through-Wall_Human_Pose_CVPR_2018_paper.html.

Zhou, B., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, 2016: “Learning deep features for discriminative localization.” *Conference on Computer Vision and Pattern Recognition*, Las Vegas, Nevada, IEEE, URL https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Zhou_Learning_Deep_Features_CVPR_2016_paper.html.

- **Right: architecture for CNN trained with MYRORSS data.**
  - (a) Storm-centered reflectivity at 1, 2, …, 12 km AGL
  - (b) Storm-centered low-level and mid-level azimuthal shear
  - (c-f) Feature maps created by conv and pooling layers
- Another branch of the CNN does conv and pooling over the proximity sounding (not shown).
- Both sounding-derived and radar-derived features are sent to dense layers.
- Pooling layers double horizontal grid spacing of radar image from 0.375 km (original) to 0.75, 1.5, 3, 6, then 12 km.
- Thus, shallow conv layers (near the left) learn small-scale features, while deep conv layers (near the right) learn large-scale features.



ML for Tornado Prediction: Gory Details

1. Each conv layer uses the leaky-ReLU activation function with slope = 0.2, followed by batch normalization.

2. Same for each dense layer except the last.

3. The last dense layer uses the sigmoid activation function, which forces its output (next-hour tornado probability) to range from 0…1.

4. I use L_2 regularization for conv layers (strength of 10^{-3} for GridRad model, $10^{-2.5}$ for MYRORSS model).

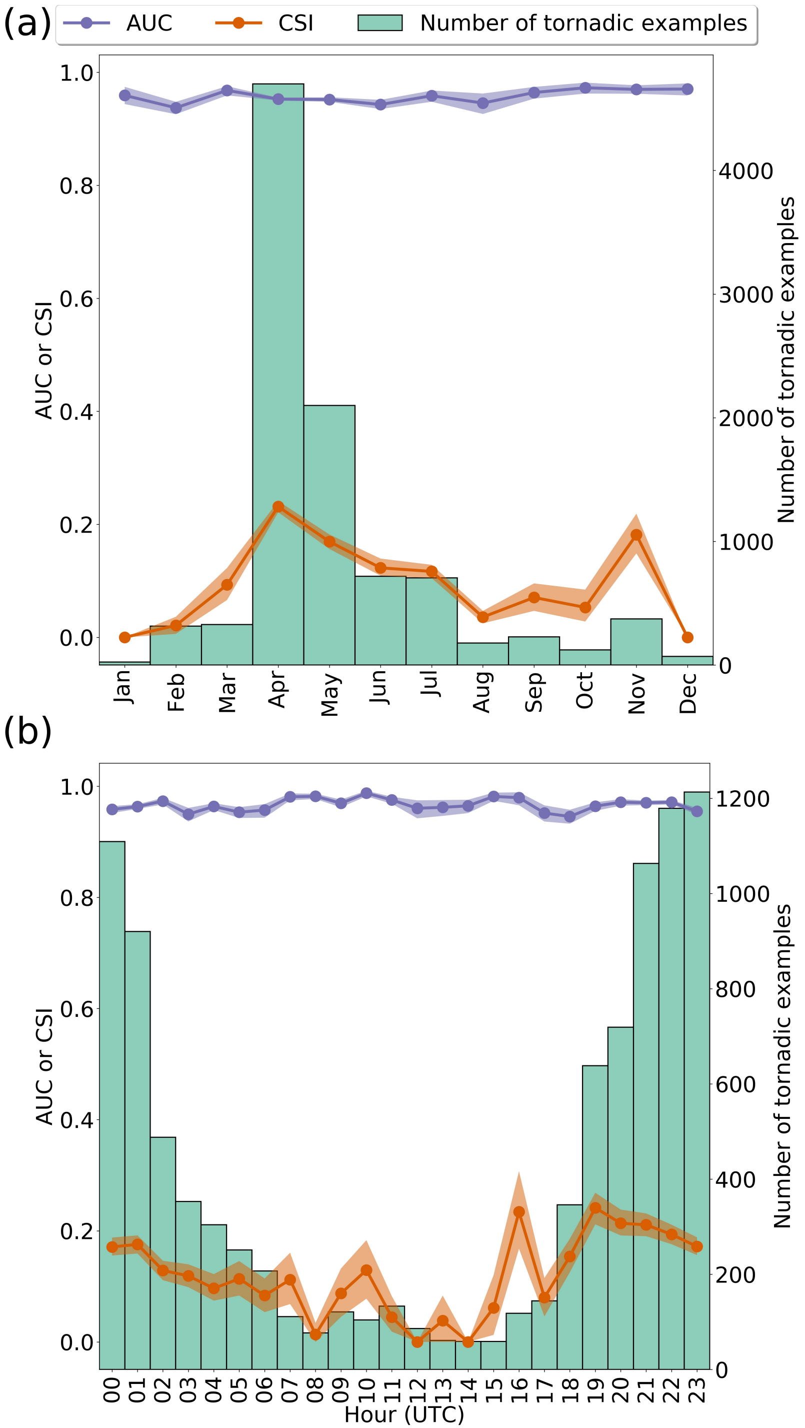
5. I use dropout regularization for all dense layers except the last (dropout rate of 0.5 for GridRad model, 0.75 for MYRORS model).

6. To handle class imbalance, I resample training data to 50% positive examples and 50% negative ("positive example" = storm that is tornadic in the next hour).
   - Resampling is used only for training.
   - Results on validation and testing data are based on full distribution, where tornadoes are a rare event.

7. I use data augmentation during training (see earlier slide).

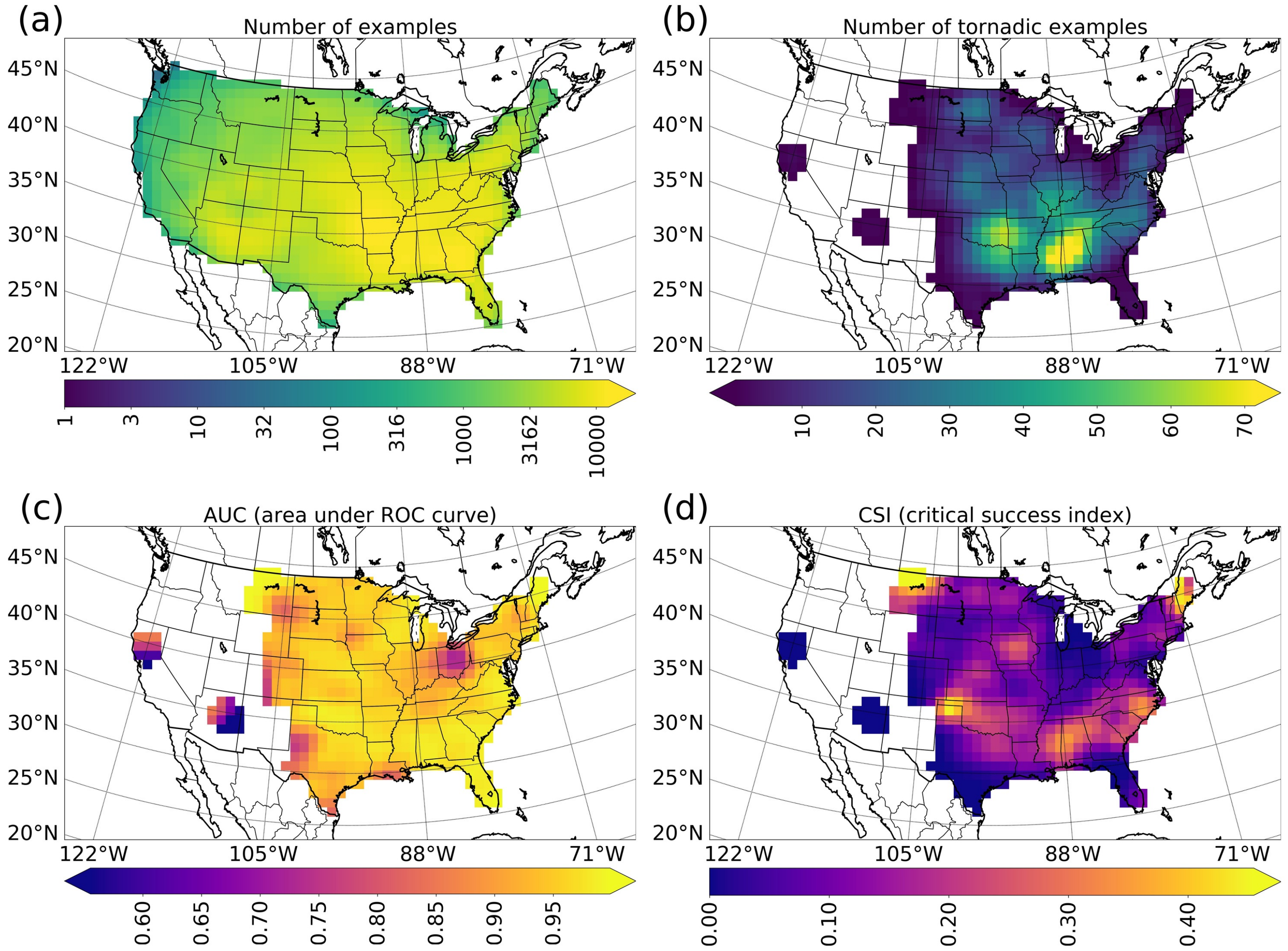
Tornado Prediction: Model Evaluation

- **Right: monthly and hourly performance of MYRORSS model on testing data.**
- AUC does not vary much with time.
- However, CSI varies a lot (sensitive to event frequency).
- CSI is best in afternoon and evening (18-05 UTC) and spring, when tornadoes are most common.



Tornado Prediction: Model Evaluation

- **Right: regional performance of MYRORSS model on testing data**
- AUC does not vary much regionally (insensitive to event frequency).
- CSI varies a lot (increases with event frequency).
- CSI is best from southern Plains to southeast, where tornadoes are most common.



Front Detection: Input Data

- Before training CNNs, data must be pre-processed:

1. Interpolate ERA5 data from lat-long grid (0.281) to equidistant grid (32 km)
   - Prevents issues that arise from unequal grid spacing (fronts overdetected near equator, underdetected near pole)

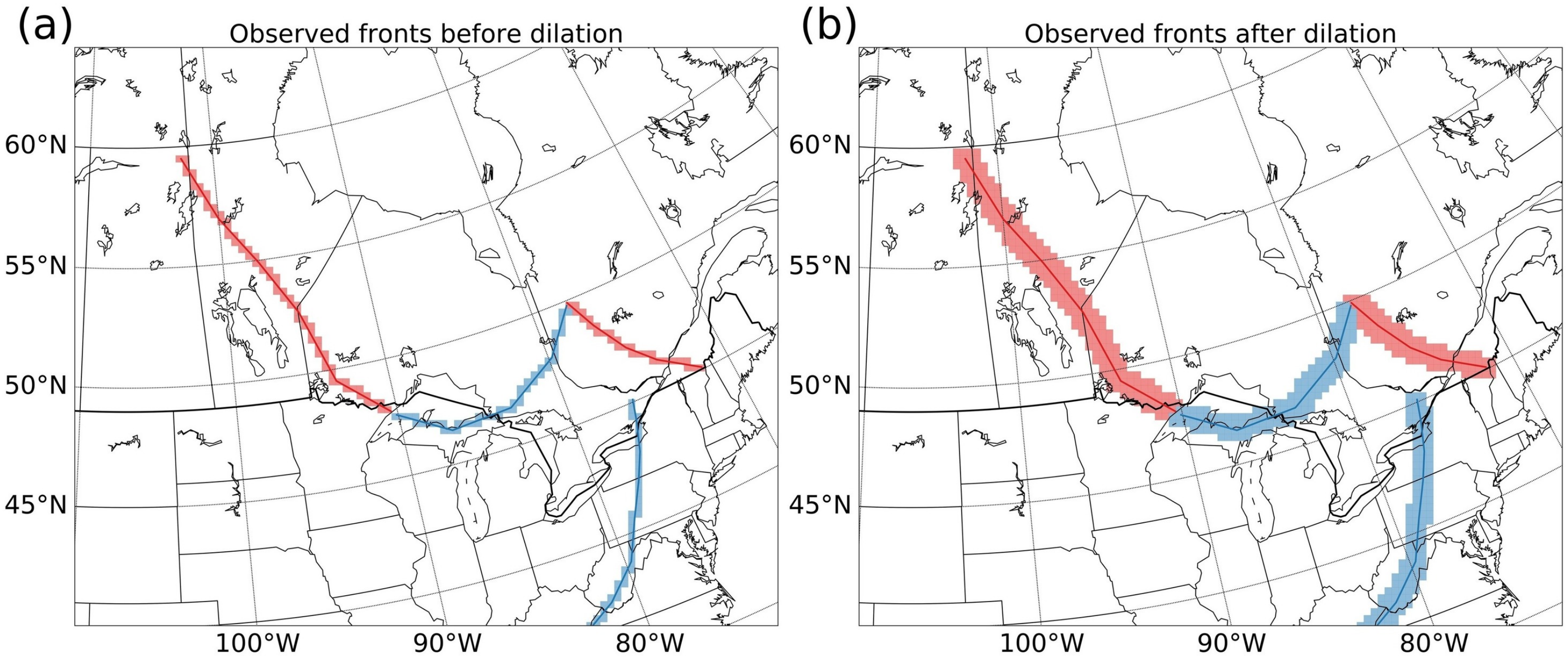
2. Rotate ERA5 winds from Earth-relative to grid-relative coordinates
   - Puts temperature gradient ($\vec{\nabla} T$), moisture gradient ($\vec{\nabla} q$), and wind vector ($\vec{v}$) in the same coordinates
   - Makes it easier for CNN to represent quantities like advection ($-\vec{v} \cdot \vec{\nabla} T$ and $-\vec{v} \cdot \vec{\nabla} q$)

3. Convert WPC fronts to gridded masks (on the same 32-km grid as predictors)

Front Detection: Input Data

- Before training CNNs, data must be pre-processed:

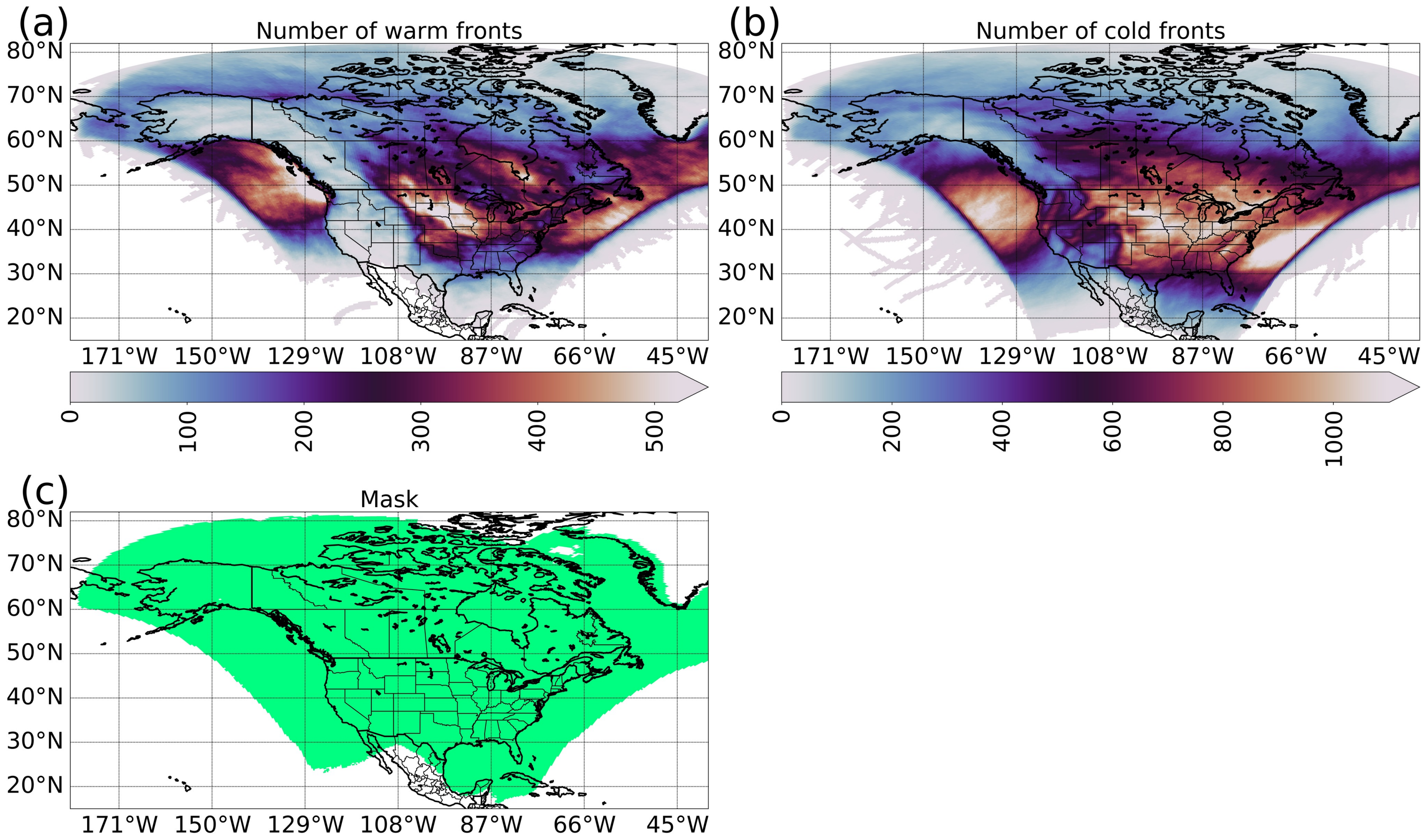
4. Dilate WPC fronts
   - Replace each frontal grid cell with 3 x 3 neighbourhood
   - Turns fronts from 1-D lines into 2-D regions (more physically realistic)
   - Also accounts for representativity error due to grid spacing



Front Detection: Input Data

- Before training CNNs, data must be pre-processed:

5. Mask out grid cells where WPC does not typically label fronts
   - Specifically, mask out grid cells with < 100 fronts in the dataset
   - These grid cells are not used for model development (training, validation, and testing)
   - These grid cells **are** used to create the climatology, because at this point correct answers are not needed (CNN has already been trained)



ML for Front Detection: Gory Details

1. Each conv layer uses the leaky-ReLU activation function with slope = 0.2, followed by batch normalization.

2. Same for each dense layer except the last.

3. The last dense layer uses the softmax activation function, which forces its outputs (three probabilities) to be positive and sum to 1.0.

4. I use L_2 regularization for conv layers (strength of 10^{-3}).

5. I use dropout regularization for all dense layers except the last (dropout rate of 0.5).

6. To handle class imbalance, I resample training data to 50% NF patches, 25% WF patches, 25% CF patches.
   - Resampling is used only for training.