A Statistical Approach To Diagnosing Storm Mode

David Stang[1], Dr. Harold Brooks[2], Dr. Michael Richman[1], and Dr. Xuguang Wang[1]

[1] School of Meteorology, University of Oklahoma
[2] NOAA/NSSL Senior Research Scientist, OU School of Meteorology

BACKGROUND

- Diagnosing storm mode (linear vs isolated) is crucial to all severe weather forecasts (Smith and Thompson et. al. 2012).

- Isolated modes are associated with greater potential for significant (EF2+) tornadoes and very large (2”+) hail, while linear modes are associated with greater potential for widespread straight-line wind damage.

- Common forecasting tools for diagnosing storm mode are Convection Allowing Models (CAMs) and High Resolution Ensemble Forecasts (HREFs).

    - Run 48 to 60 hours in the future
    - Accuracy often decreases quickly with increasing lead-time (Stratman and Coniglio et. al. 2013)
    - Subject to chaotic behavior (Schwartz and Wong et. al. 2020)

RESEARCH GOAL AND APPROACH

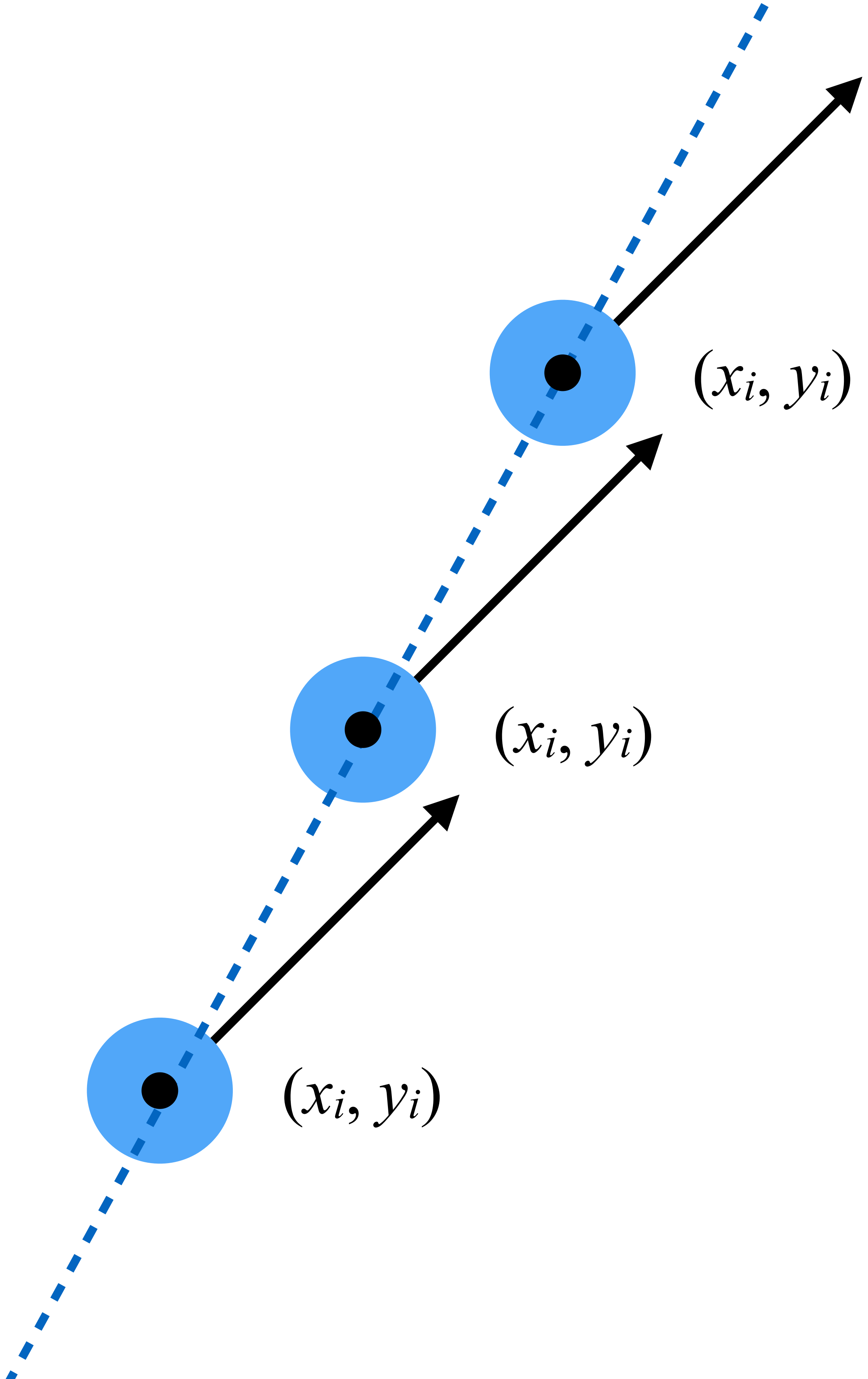
- **GOAL:** Develop a framework for predicting storm mode that only depends on synoptic scale variables.

    - Can be incorporated into models at any scale (including global models)
    - Can produce a forecast at any time range any numerical models cover
    - Can provide a “second opinion” to what CAMs and HREFs predict
    - Intended to improve medium-range forecast skill and motivate more thorough forecasting of potentially significant severe weather events

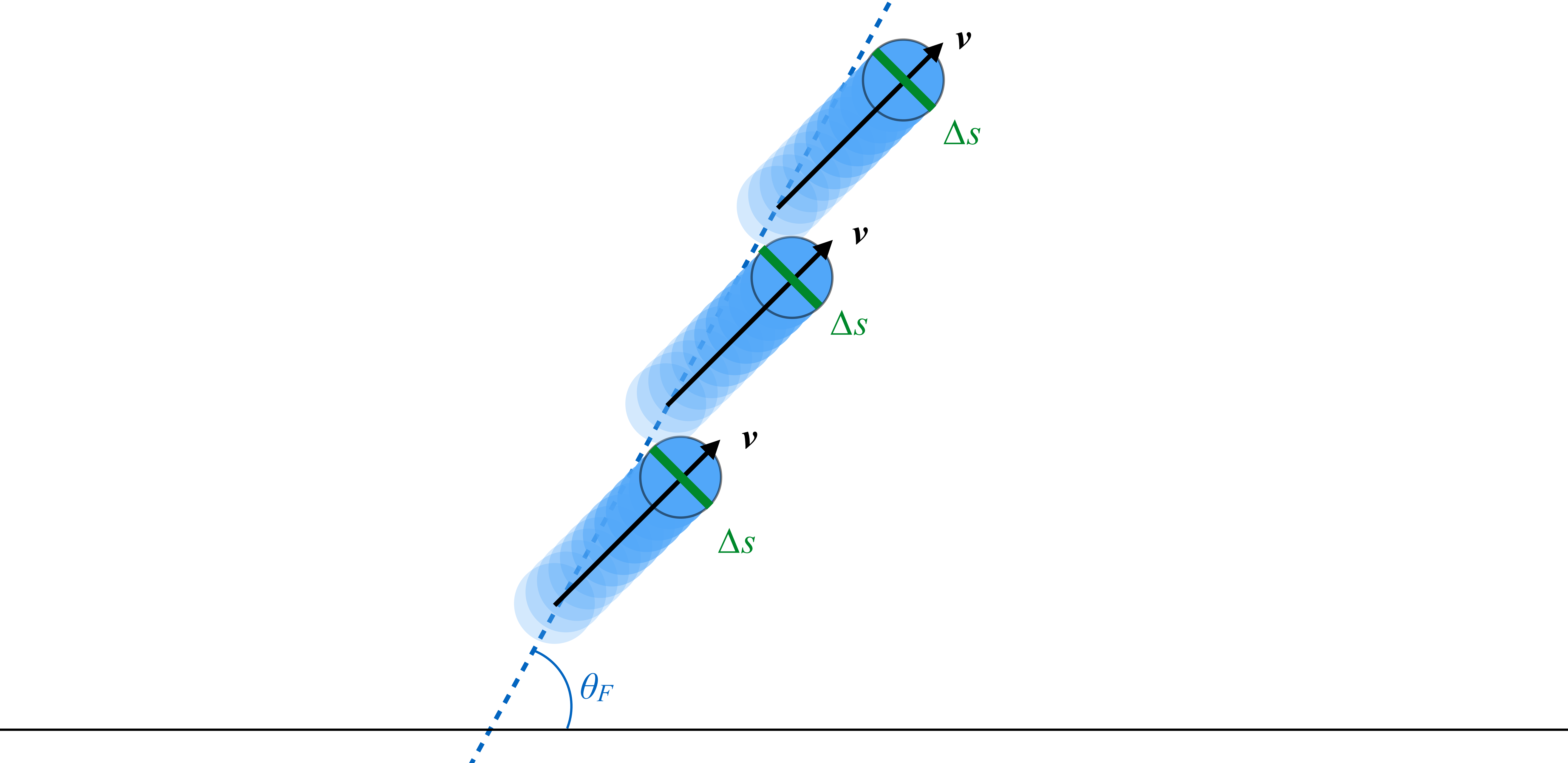
- **APPROACH:** Translate a storm-scale perspective into a formulation that is purely dependent on synoptic scale variables.

    - Situation 1: Storms that initiate along or near boundaries (theoretical)
    - Situation 2: Storms that do not initiate along boundaries (stochastic)

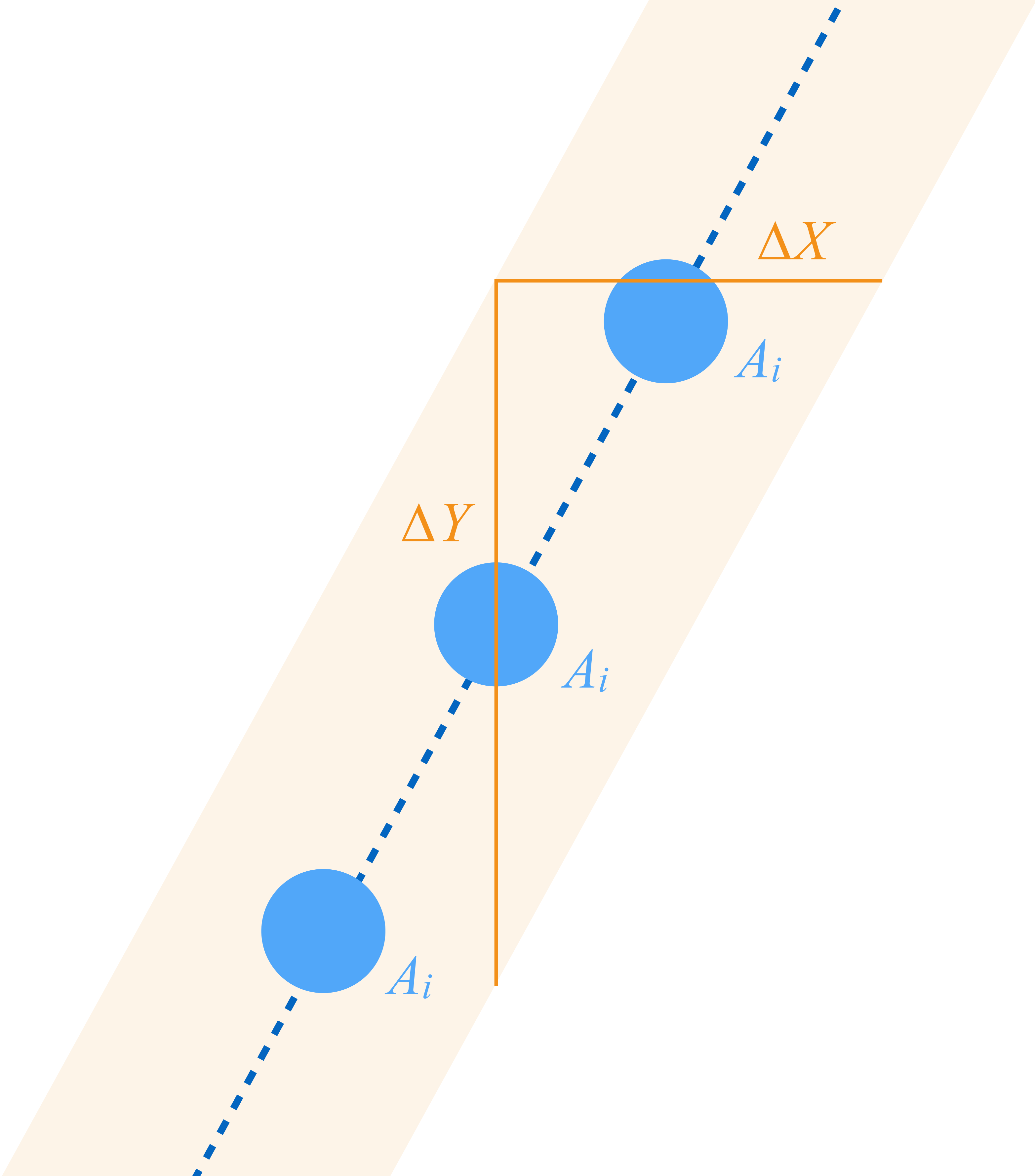
STORM-SCALE PERSPECTIVE



STORM-SCALE PERSPECTIVE

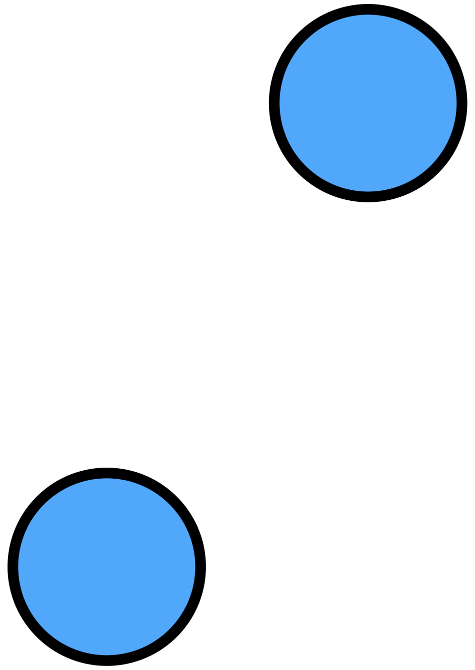


STORM-SCALE PERSPECTIVE

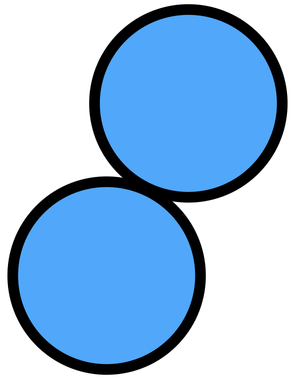


$$P_0 = \frac{n\,A_i}{A}$$

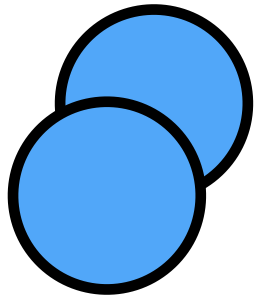
STORM-SCALE PERSPECTIVE



$D > 0$



$D = 0$



$D < 0$

PERSPECTIVE TRANSLATION

Metric	Storm Scale	Synoptic Scale
Storm Position	(x_i, y_i)	Derived from initiation zone and P_0
Storm Velocity	$\boldsymbol{v}$	LCL-EL mean wind vector
Initiation Axis	θ_F	Calculated from Theta-E field
Storm Size	Δs	Empirically estimated mean
Initiation Zone Geometry	$\Delta X, \Delta Y$	Empirically estimated mean
Storm Mode	$f(D)$	Graphical discriminant function
Initiation Probability	P_0	Estimated from instability and forcing

STORM STATISTICS

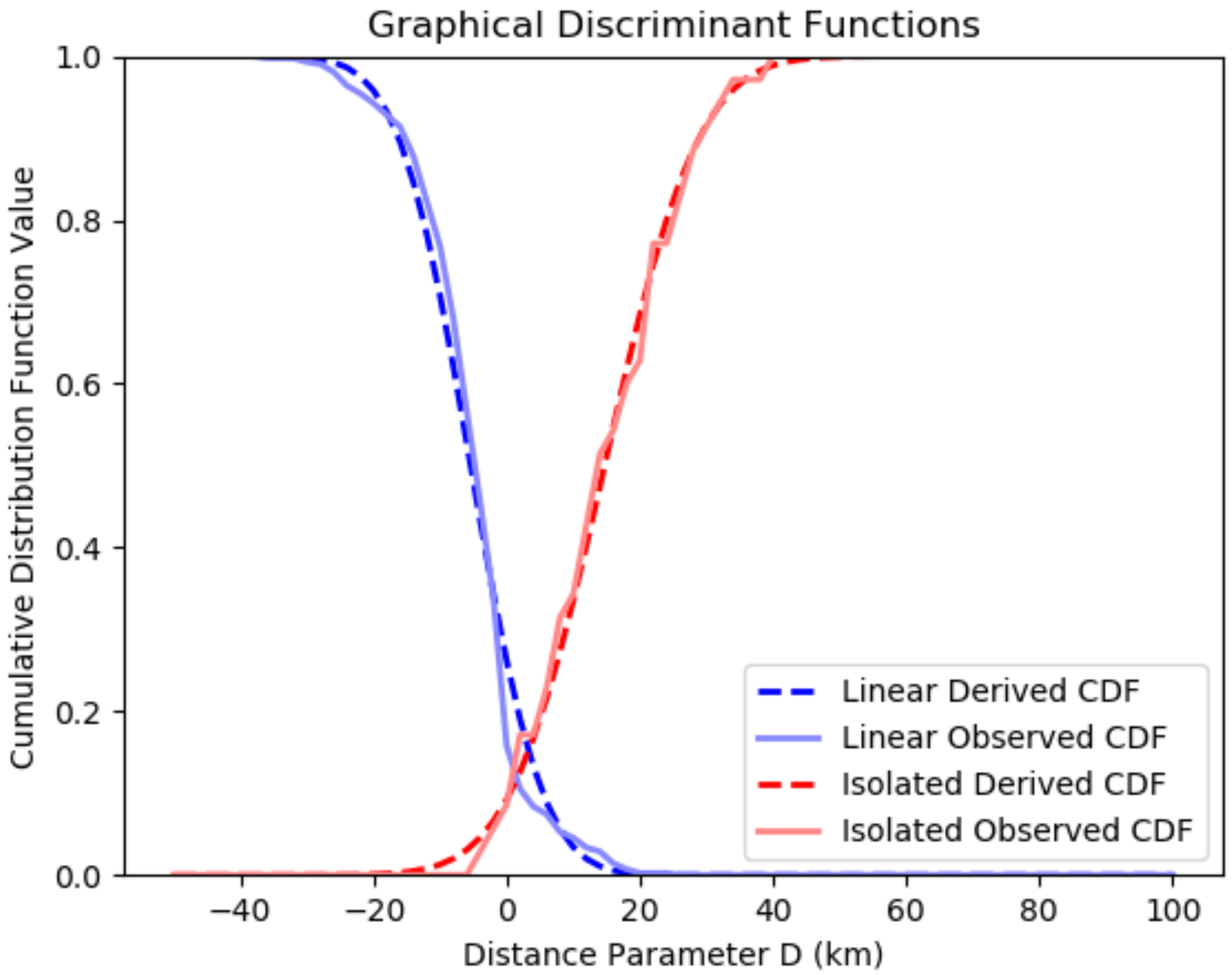
- Level II radar data from severe weather events with at least an Enhanced (Level 3) Risk from the Storm Prediction Center (SPC) were 2016-2019 obtained

- Selected base reflectivity images that showed purely isolated storms or purely linear storms (192 images in total, one image per event).

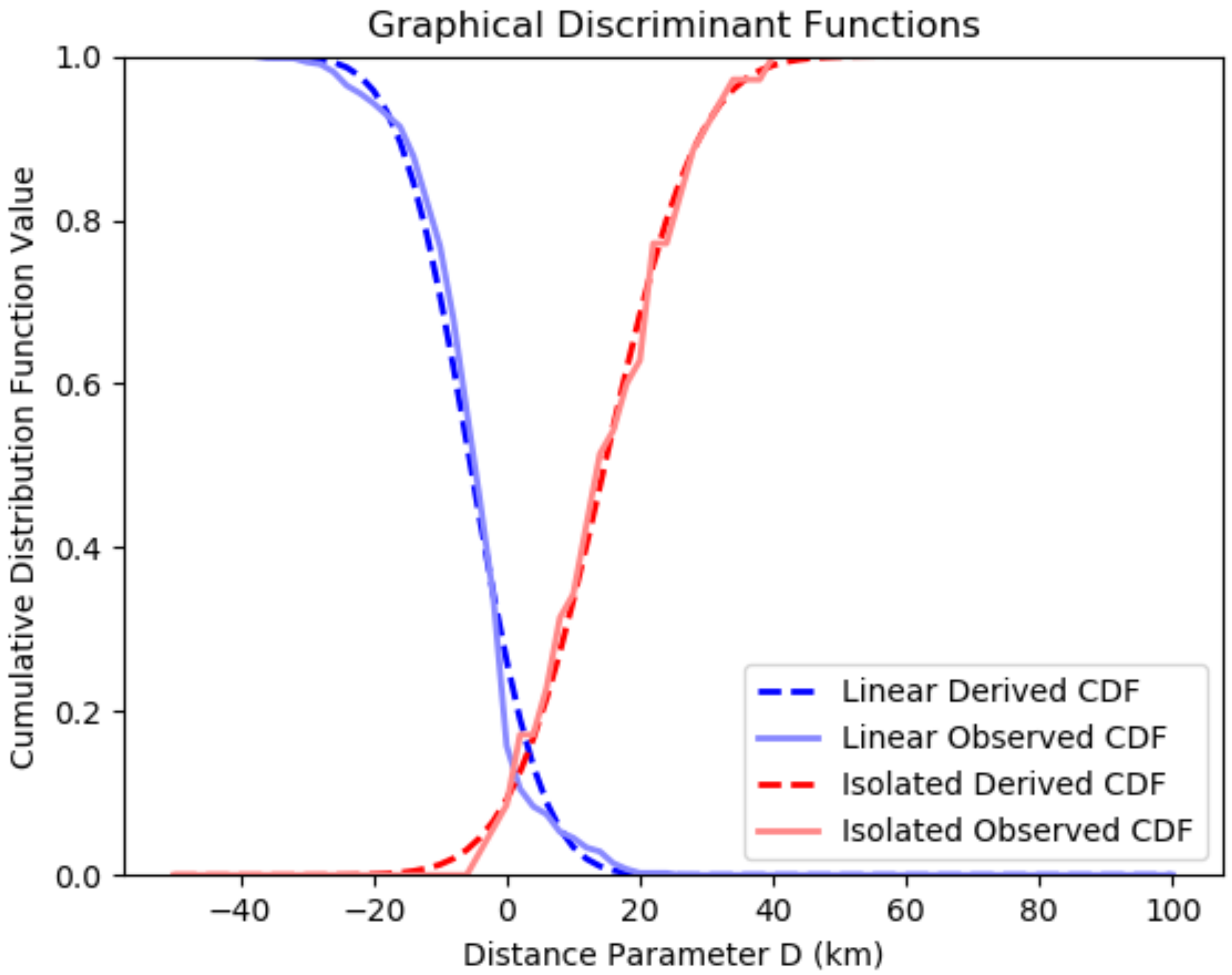
- Objective analysis identifies individual cells, estimates the largest contiguous length of reflectivity over 40 dBZ (then becomes Δs) and the central point of each cell. The distance parameter D was also calculated for each cell.

    - 463 total storms identified (35 isolated, 418 linear)
    - Mean storm cold pool width (Δs) calculated to be ~35.8 km
    - Sample mean and standard deviation used to derive probability distribution functions (PDFs) for linear (or isolated) as a function of D

STORM STATISTICS

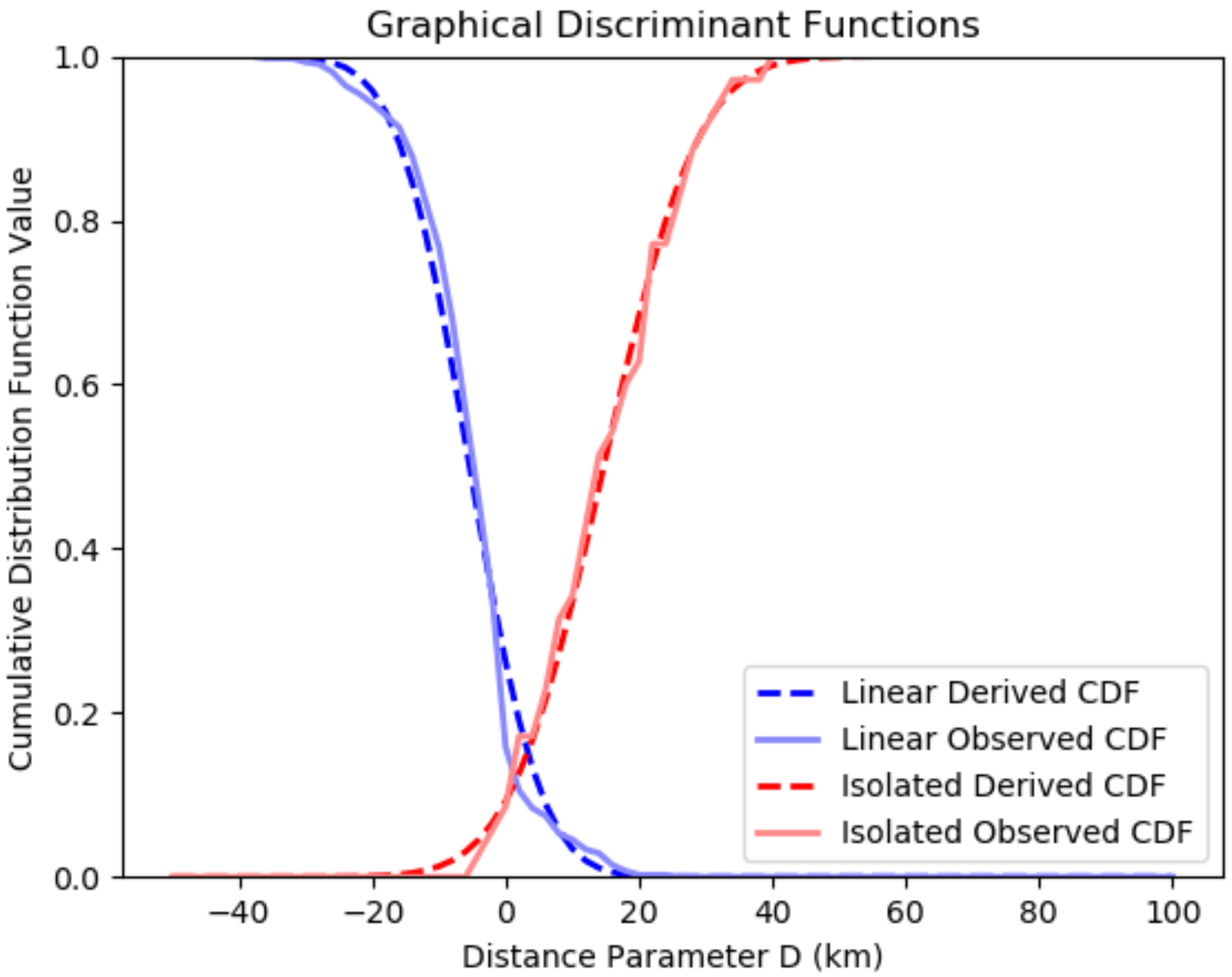


STORM STATISTICS



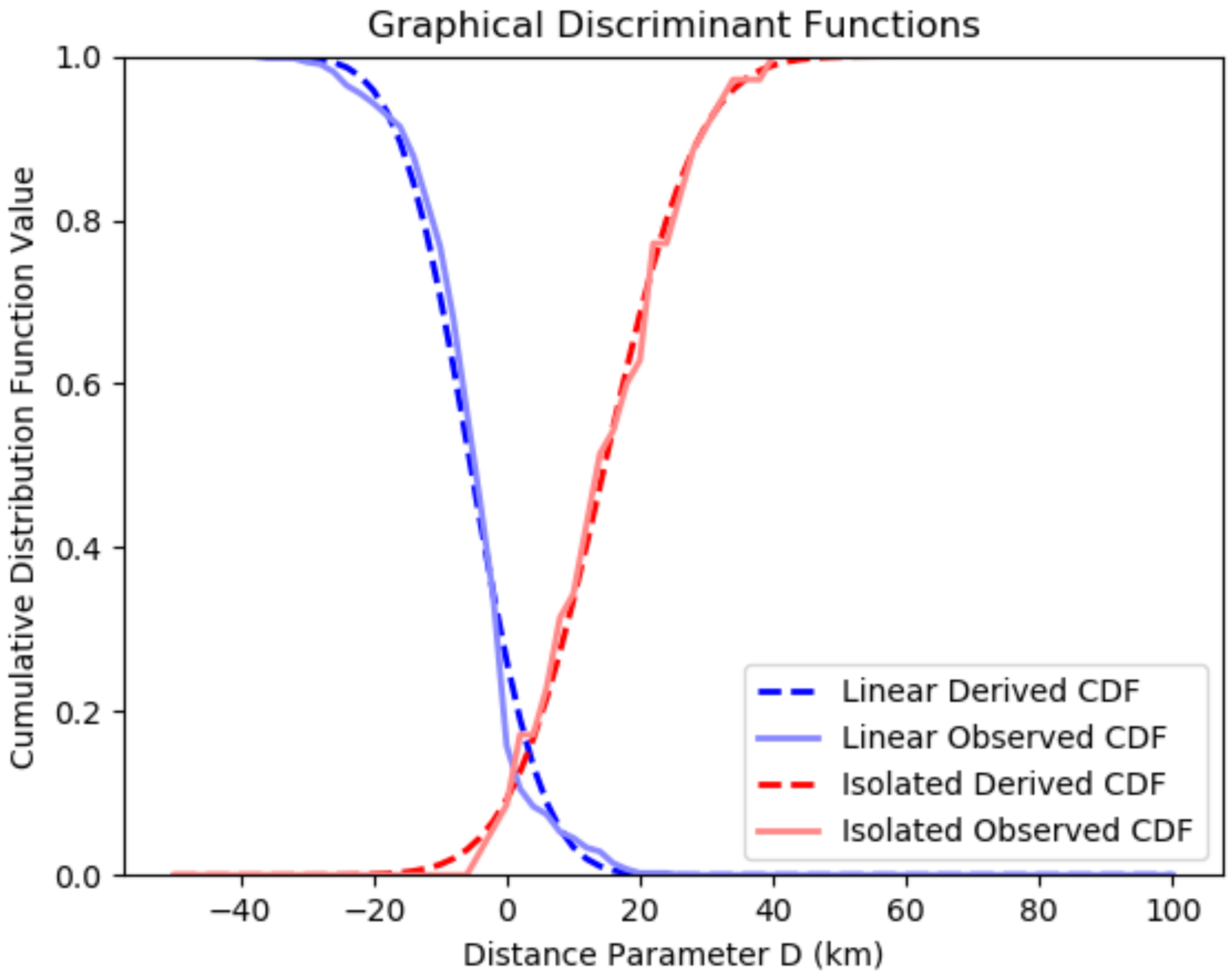
$$x_{linear} = \frac{1}{\sigma_{linear}\sqrt{2\pi}} \int_{D}^{\infty} \exp\left(- \left(\frac{D - \bar{D}_{linear}}{\sqrt{2}\ \sigma_{linear}} \right)^{2} \right) dD$$

STORM STATISTICS



$$x_{linear} = \frac{1}{\sigma_{linear}\sqrt{2\pi}} \int_{D}^{\infty} \exp\left(- \left(\frac{D - \bar{D}_{linear}}{\sqrt{2}\ \sigma_{linear}} \right)^{2} \right) dD \qquad x_{isolated} = \frac{1}{\sigma_{isolated}\sqrt{2\pi}} \int_{-\infty}^{D} \exp\left(- \left(\frac{D - \bar{D}_{isolated}}{\sqrt{2}\ \sigma_{isolated}} \right)^{2} \right) dD$$

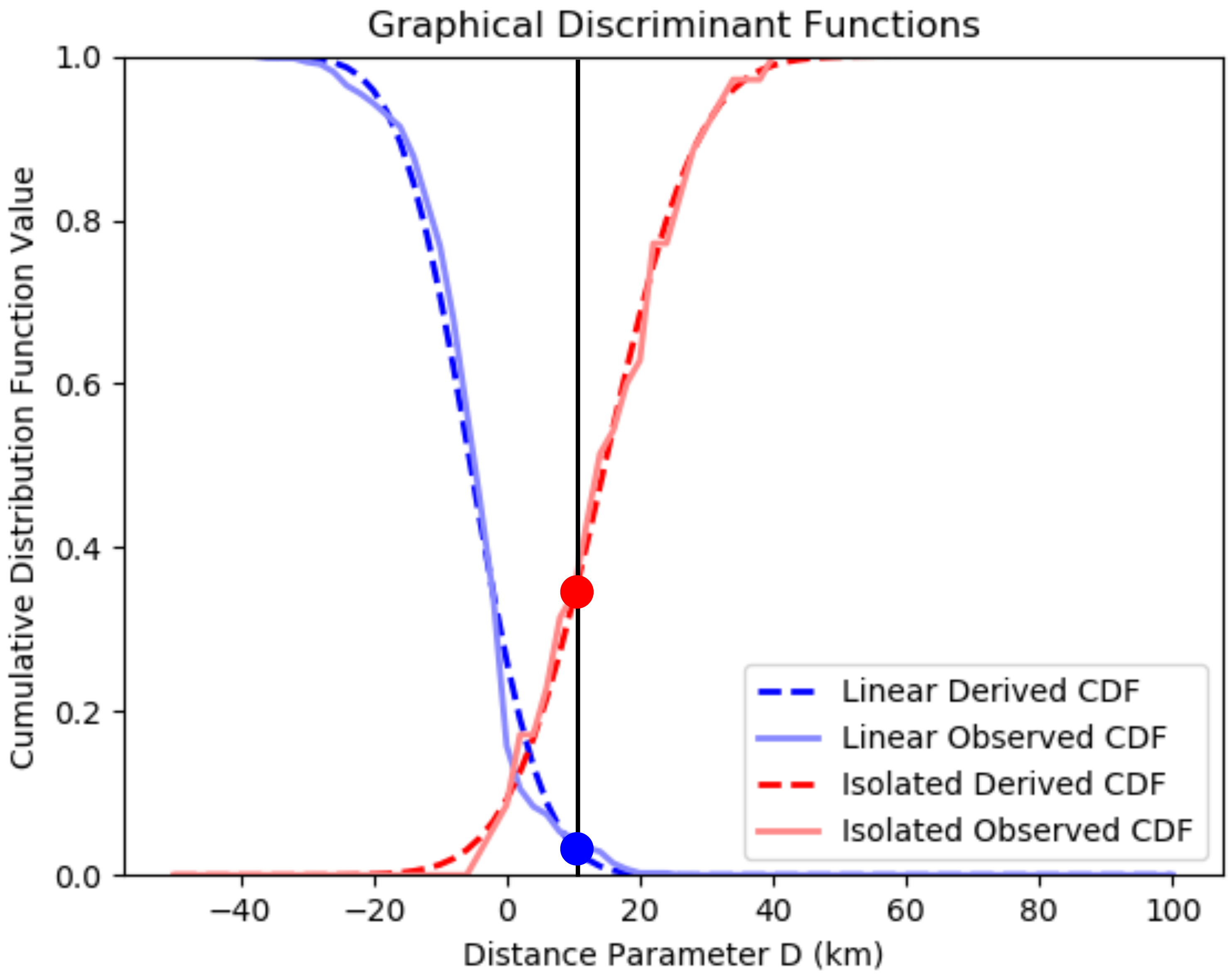
STORM STATISTICS



$$x_{linear} = \frac{1}{\sigma_{linear}\sqrt{2\pi}} \int_{D}^{\infty} \exp\left(- \left(\frac{D - \bar{D}_{linear}}{\sqrt{2}\ \sigma_{linear}} \right)^{2} \right) dD \qquad x_{isolated} = \frac{1}{\sigma_{isolated}\sqrt{2\pi}} \int_{-\infty}^{D} \exp\left(- \left(\frac{D - \bar{D}_{isolated}}{\sqrt{2}\ \sigma_{isolated}} \right)^{2} \right) dD$$

$$r_{linear} = \frac{x_{linear}}{x_{linear} + x_{isolated}} \times 100\ \% \qquad r_{isolated} = \frac{x_{isolated}}{x_{linear} + x_{isolated}} \times 100\ \%$$

STORM STATISTICS



$$x_{linear}\,(D = 10\text{ km}) = 0.03414509193910931 \qquad x_{isolated}\,(D = 10\text{ km}) = 0.33570393162851975$$

$$r_{linear} = \frac{0.034}{0.034 + 0.336} \times 100\ \% = 9.2\ \% \qquad r_{isolated} = \frac{0.336}{0.034 + 0.336} \times 100\ \% = 90.8\ \%$$

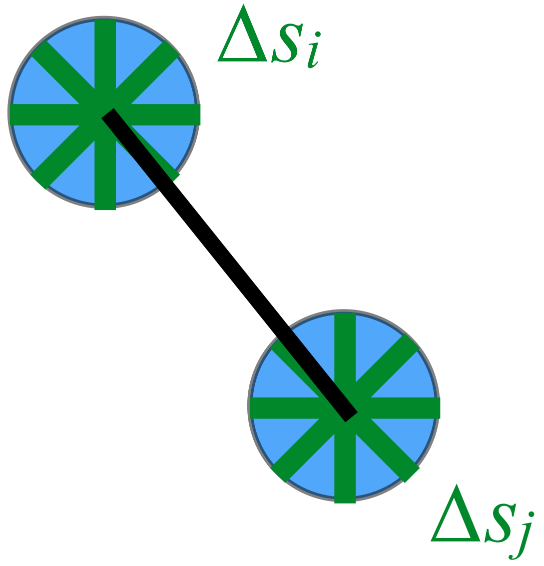
DISTANCE PARAMETER CALCULATION

- Only variable that matters is the distance parameter D

$$D(t) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} - \frac{1}{2}\Delta s_i - \frac{1}{2}\Delta s_j$$

$$x_i = x_{i_0} + v_i\, t \cos\theta_i \qquad x_j = x_{j_0} + v_j\, t \cos\theta_j$$

$$y_i = y_{i_0} + v_i\, t \sin\theta_i \qquad y_j = y_{j_0} + v_j\, t \sin\theta_j$$



- Assumptions:

  - Cold pool width and temperature is constant with time
  - All cold pools are treated as ellipses with identical major and minor axes
  - All cells have the same constant velocity (speed and direction)
  - Cells that initiate along a boundary are evenly spaced
  - Cells that do not initiate along a boundary are randomly distributed

DISTANCE PARAMETER CALCULATION

- At $t = t*$, the distance parameter D attains its minimum value $D*$, which is the minimum distance two given cold pools are from each other

$$t* = \frac{x_{i_0} - x_{j_0}}{v_j} \cos\theta_j + \frac{y_{i_0} - y_{j_0}}{v_j} \sin\theta_j$$

- This expression is obtained by the taking the first derivative of the function for D and solving for time t.

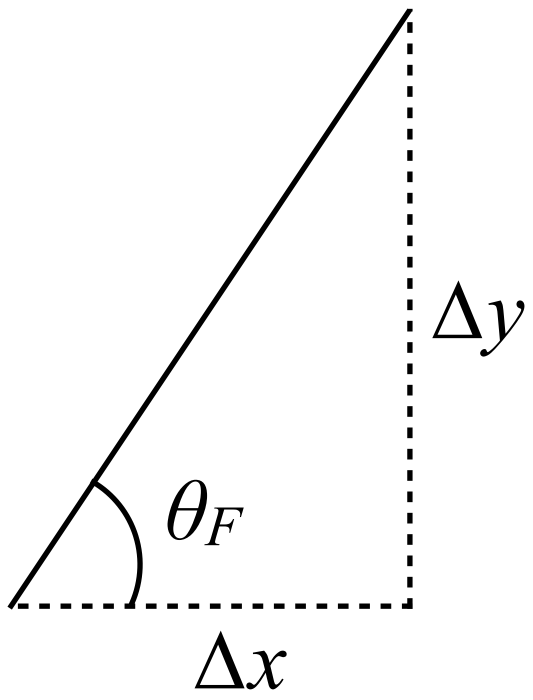
- Distance parameter D plugged in a graphical discriminant function, which probabilistically models storm mode as a function of time

BOUNDARY IDENTIFICATION

- Given a grid of 1000 mb temperature and 1000 mb moisture measurements, a derived grid of θ_e was calculated.

- From the derived grid, $|\nabla\theta_e|$ is calculated and local maxima are highlighted as “key points”.

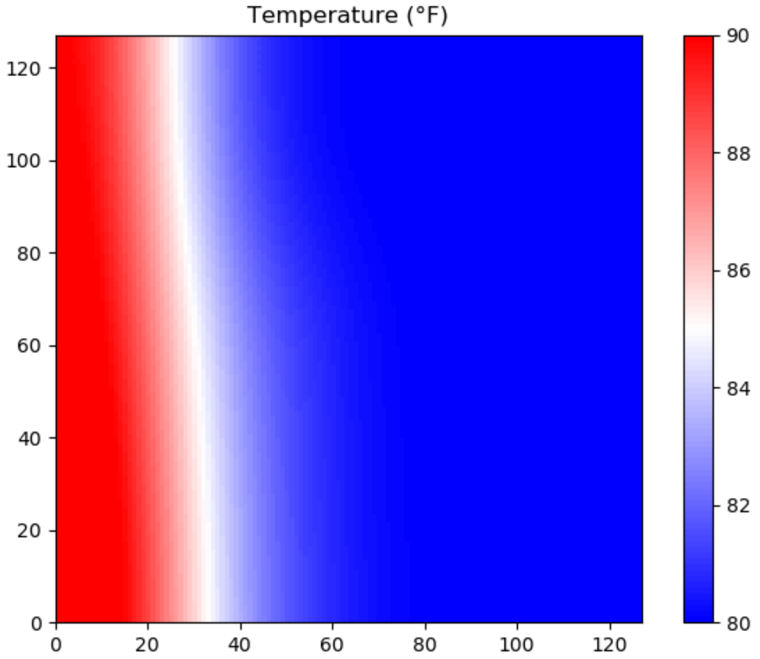
- Linear regressions of the “key points” were performed on 1000 km × 1000 km subsections, the arctangent of the best-fit slope then becomes θ_F.

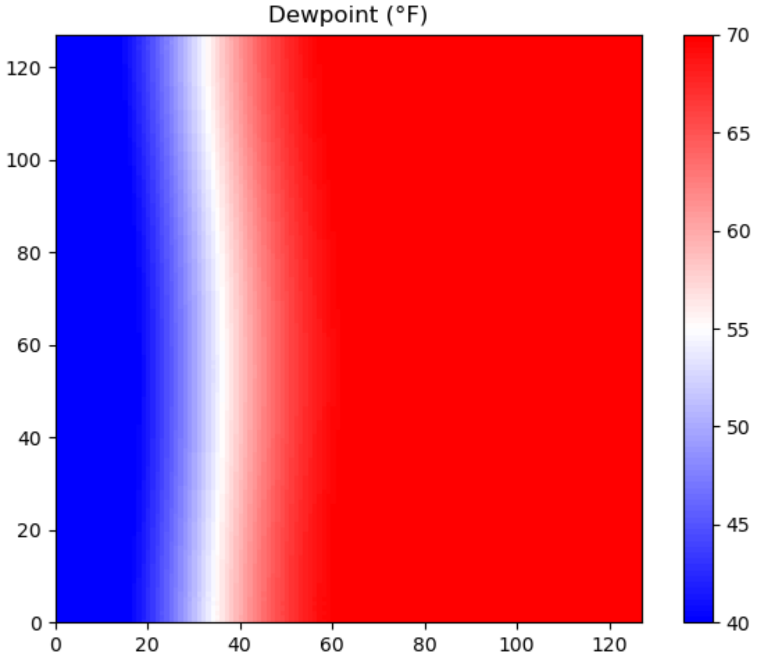


$$m = \frac{\Delta y}{\Delta x} = \tan \theta_F$$

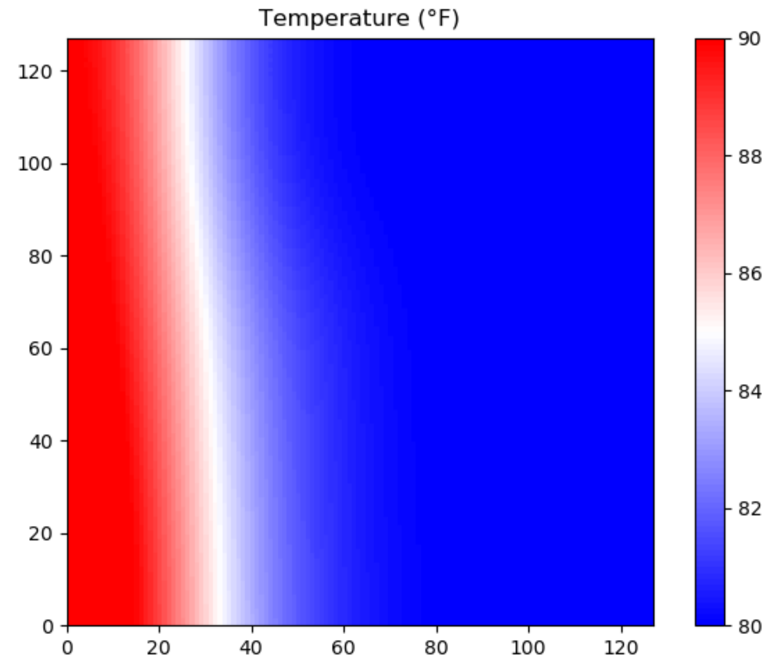
$$\theta_F = \tan^{-1}(m)$$

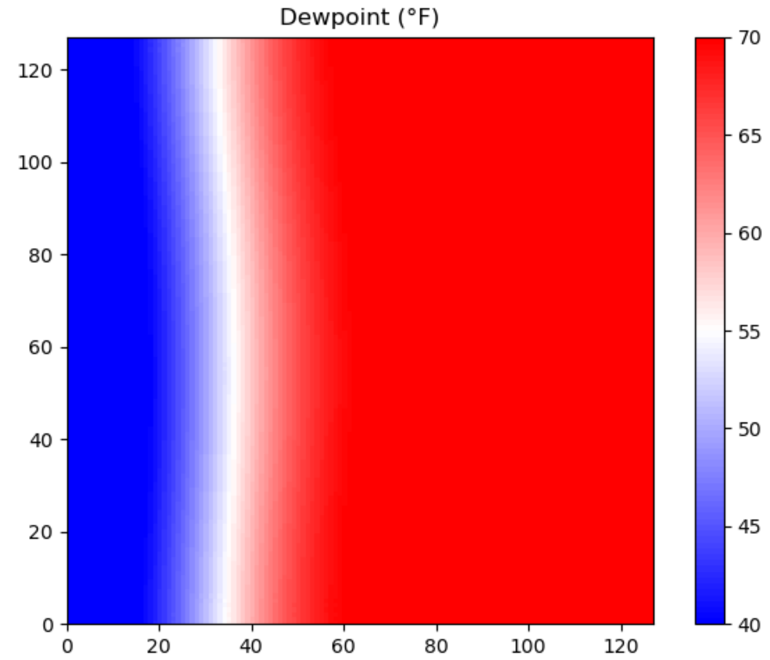
FRONT IDENTIFICATION

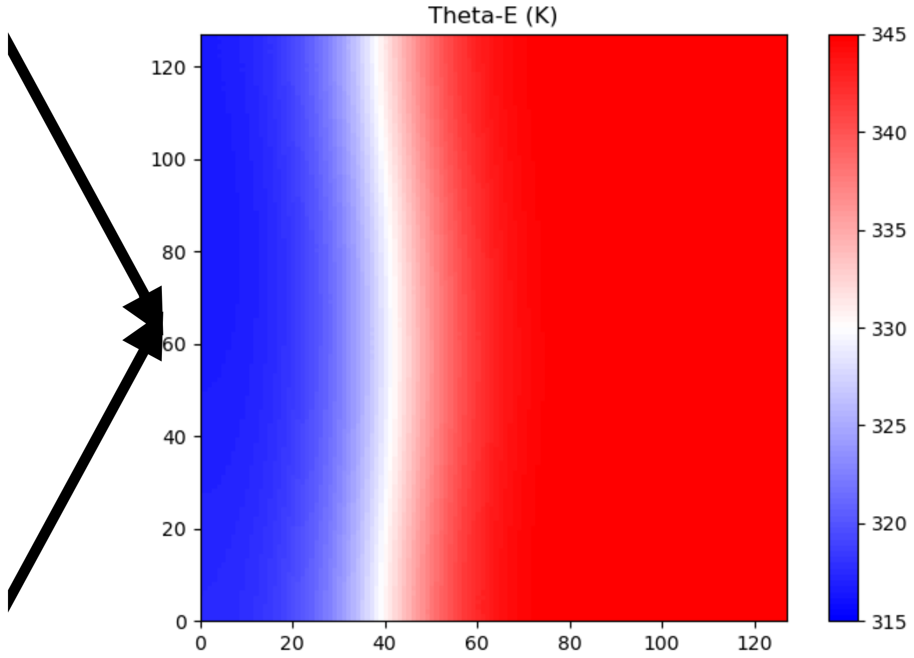

Temperature (°F)


Dewpoint (°F)

FRONT IDENTIFICATION

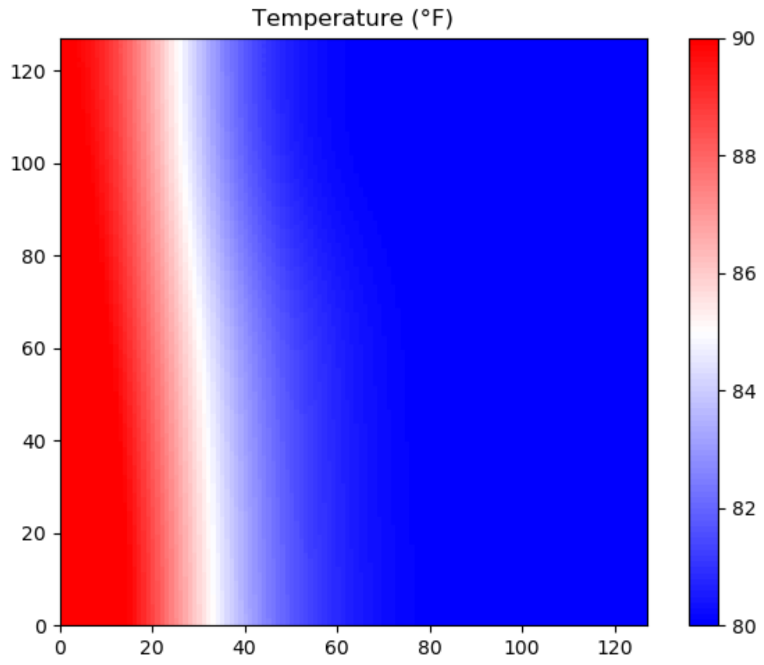


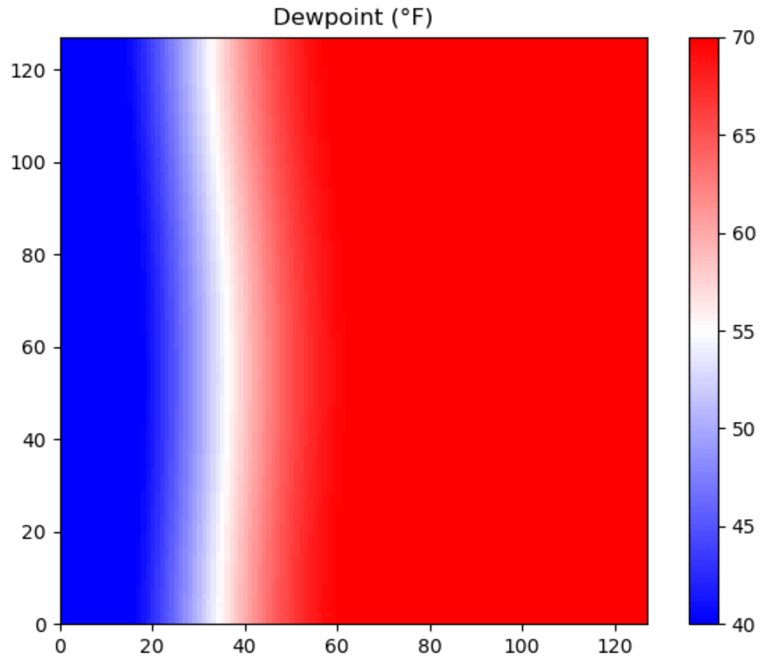


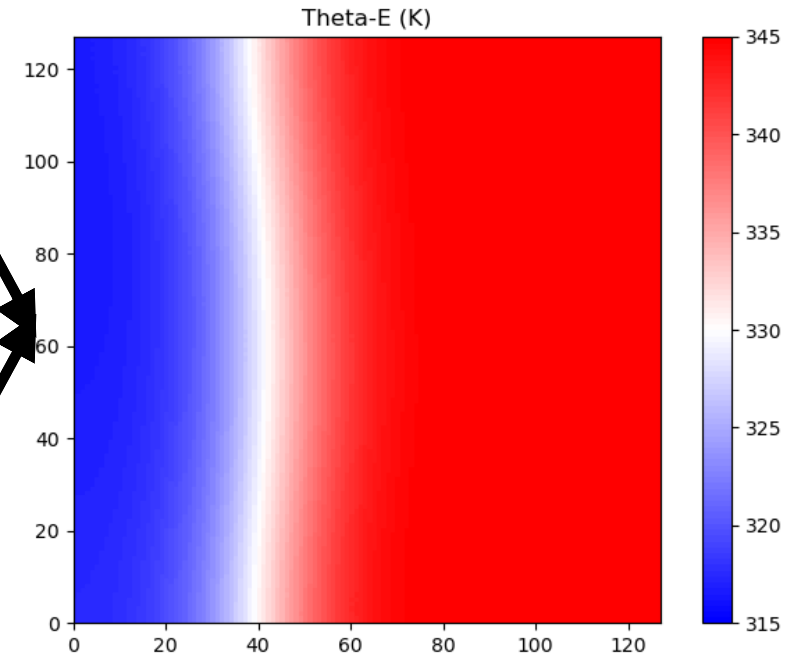


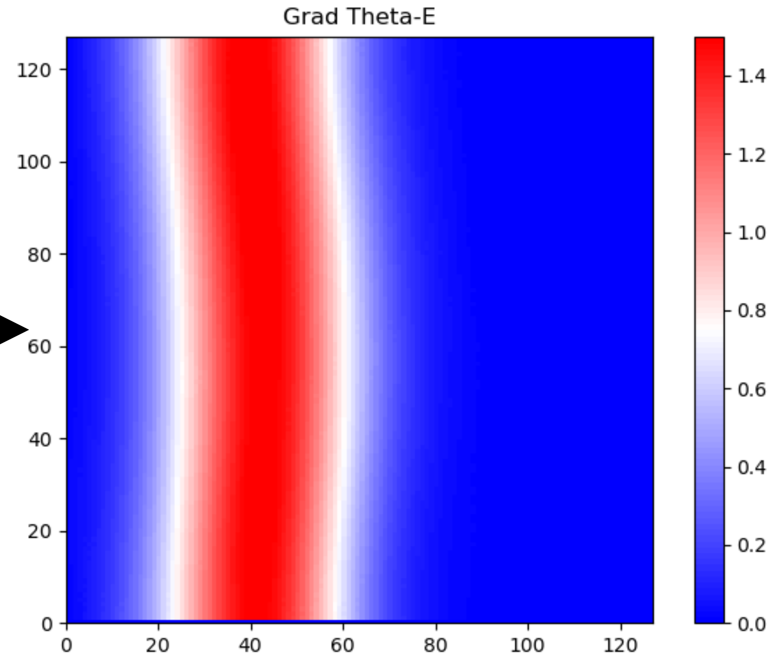
Combine temperature and dewpoint data into a grid of θ_e

FRONT IDENTIFICATION



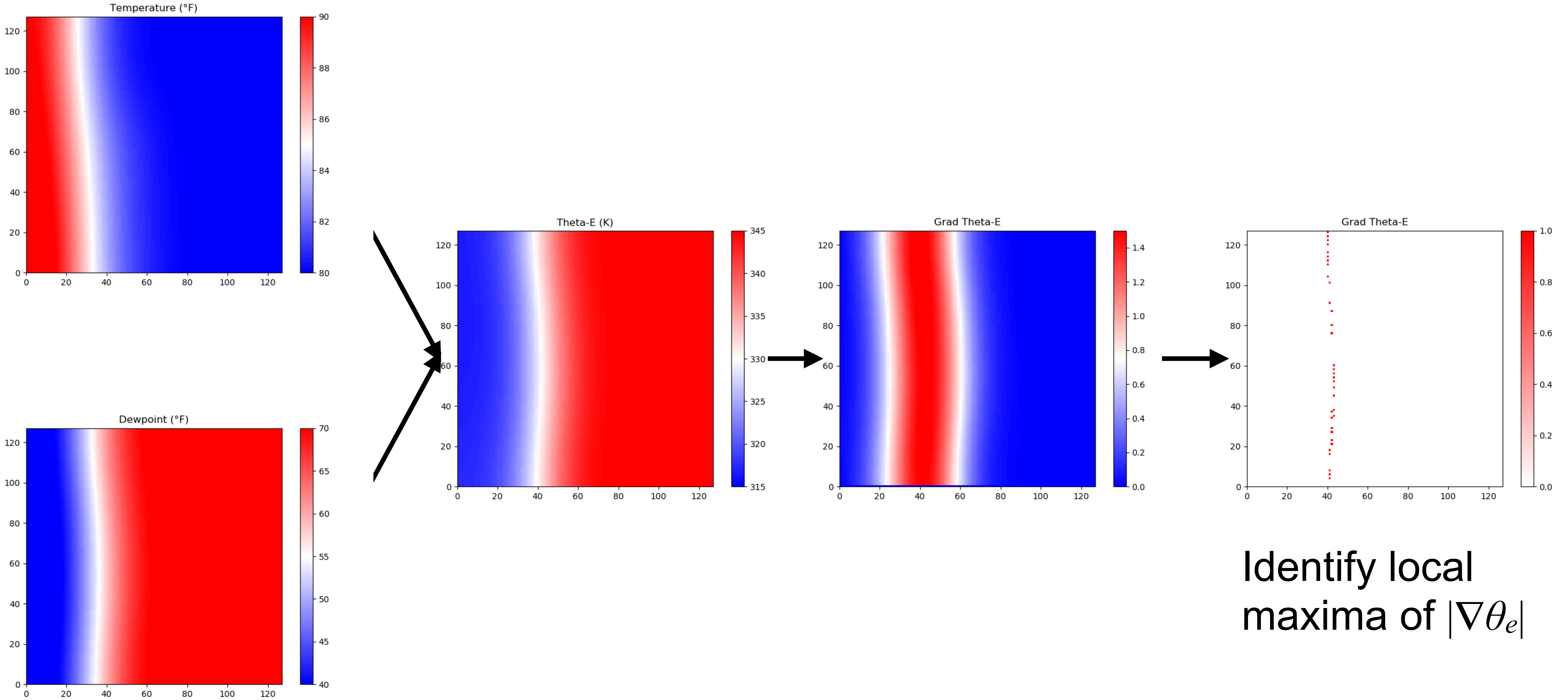






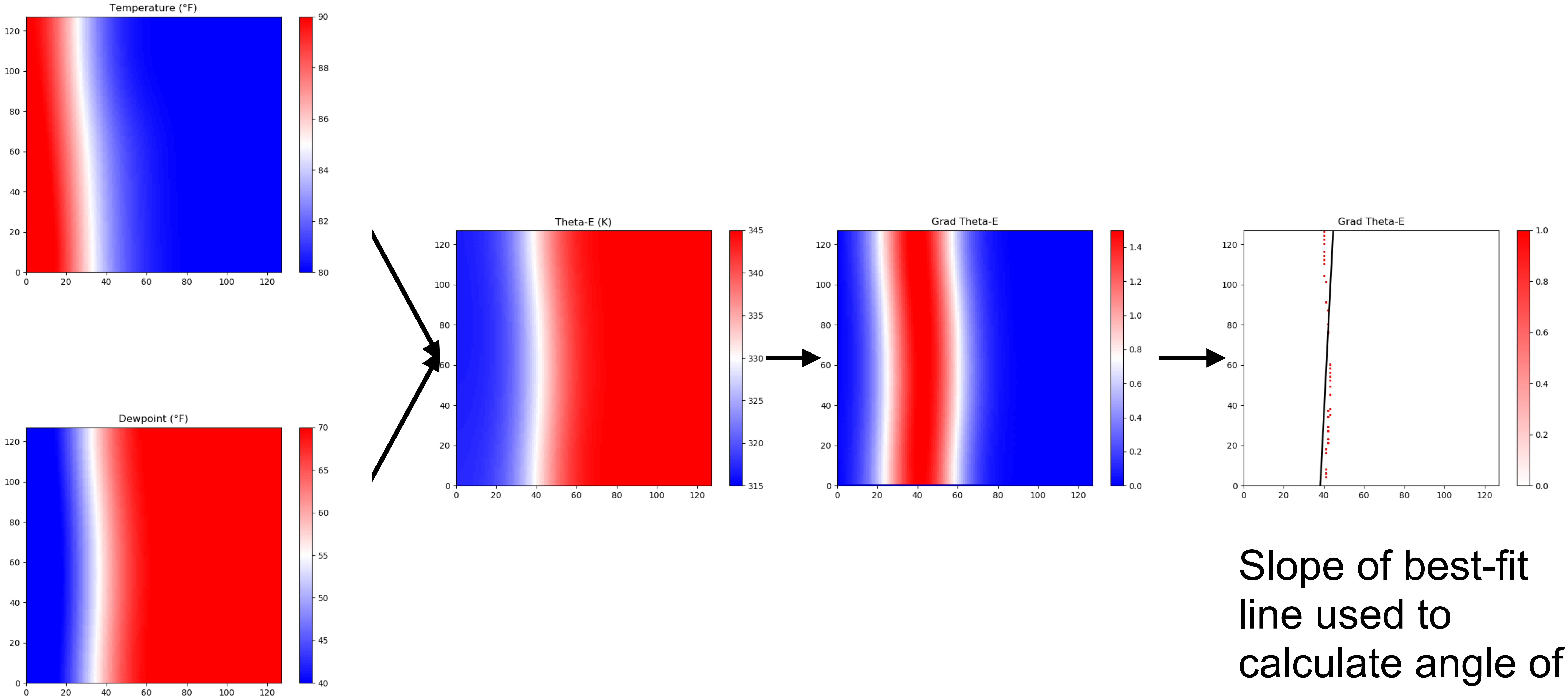
Calculate the magnitude of $\nabla\theta_e$

FRONT IDENTIFICATION



Identify local maxima of $|\nabla\theta_e|$

FRONT IDENTIFICATION



Slope of best-fit line used to calculate angle of boundary

INITIATION ZONE IDENTIFICATION

- Mean values for ΔX and ΔY obtained by examining 7204 radar base reflectivity images taken from events with at least an Enhanced (Level 3) Risk from SPC

- ΔY for each radar image is calculated by the following algorithm:

1) Estimate the coordinates (x_i, y_i) and width (Δs) of each storm depicted
2) Perform a linear regression on the points from (1)
3) Calculate “errors” (ε) between each individual storm and the line (2)
4) Calculate z-score for each “error”
5) Examine all storms that are within two standard deviations of the line (2)
6) ΔY is the maximum value of $2 \cdot \varepsilon + 2 \cdot \Delta s$
7) ΔX is the value of ΔY divided by the tangent of the best fit line’s angle

- The means of ΔX and ΔY are used in the forecast tool

INITIATION PROBABILITY (P_0)

- Estimated initiation probability based on forecast soundings and vertical wind w
- Determine critical values of $\hat{T}_v$ and $\hat{w}$:
  - $\hat{T}_v$: Minimum virtual temperature the 100mb mixed layer air parcel must have to reach the tropopause for the predicted vertical velocity
  - $\hat{w}$: Minimum vertical velocity the 100mb mixed layer air parcel must have to reach the tropopause for the predicted virtual temperature

$$P_T = \frac{1}{\sigma_T\sqrt{2\pi}} \int_{-\infty}^{T_v} \exp\left(-\left(\frac{T_v - \hat{T}_v}{\sqrt{2}\ \sigma_T} \right)^2 \right) dT_v \qquad P_w = \frac{1}{\sigma_w\sqrt{2\pi}} \int_{-\infty}^{w} \exp\left(-\left(\frac{w - \hat{w}}{\sqrt{2}\ \sigma_w} \right)^2 \right) dw$$

$$P_0 = \frac{1}{2}\left(P_T + P_w \right)$$

INITIATION PROBABILITY (P_0)

- Air parcel traces conducted using the following algorithm:

1) Calculate ambient temperature at parcel height (linear interpolation)
2) Calculate ambient pressure at parcel height (linear interpolation)
3) Calculate change in acceleration (buoyant force)
4) Calculate change in vertical speed using acceleration from (3)
5) Calculate change in parcel height from (4)
6) Calculate temperature decrease from (5)
7) If air parcel is unsaturated:
    - Calculate new value of r_s using (6)
    - Check if $r > r_s$ to determine whether the air parcel is saturated
8) Repeat steps (1) - (7) until the air parcel is stable or reaches the tropopause

INITIATION PROBABILITY (P_0)

- Air parcel traces conducted using the following algorithm:

1) $$\bar{T} = \left(\frac{T_j - T_i}{z_j - z_i}\right)(z - z_i) + T_i$$

2) $$P = \left(\frac{P_j - P_i}{z_j - z_i}\right)(z - z_i) + P_i$$

3) $$a_{n+1} = a_n + g_0\left(\frac{T'}{\bar{T}} - 1\right)\Delta t$$

4) $$w_{n+1} = w_n + a_{n+1}\,\Delta t$$

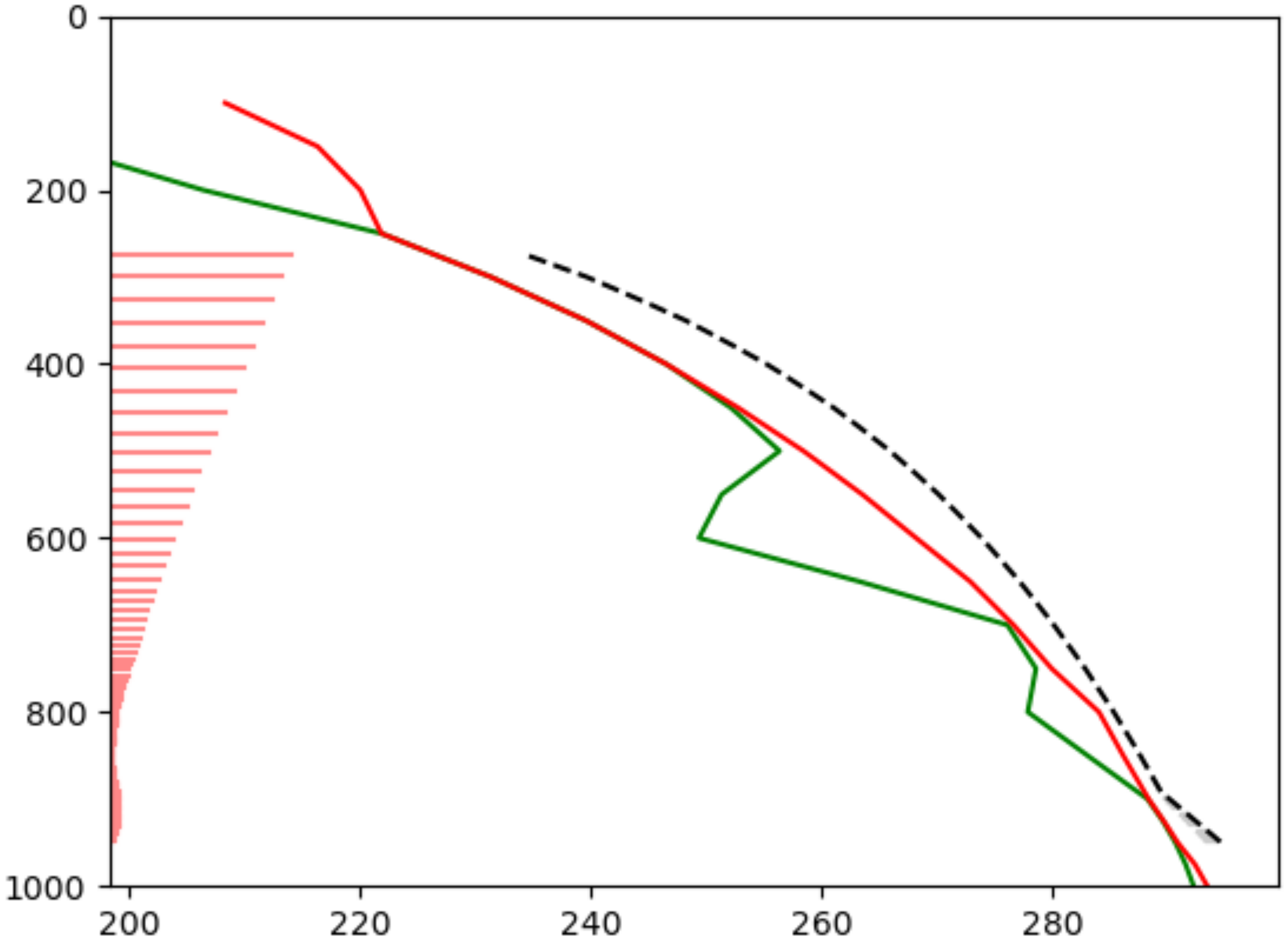
5) $$z_{n+1} = z_n + w_{n+1}\,\Delta t$$

6) $$T_{n+1} = T_n - \gamma\left(z_{n+1} - z_n\right)$$

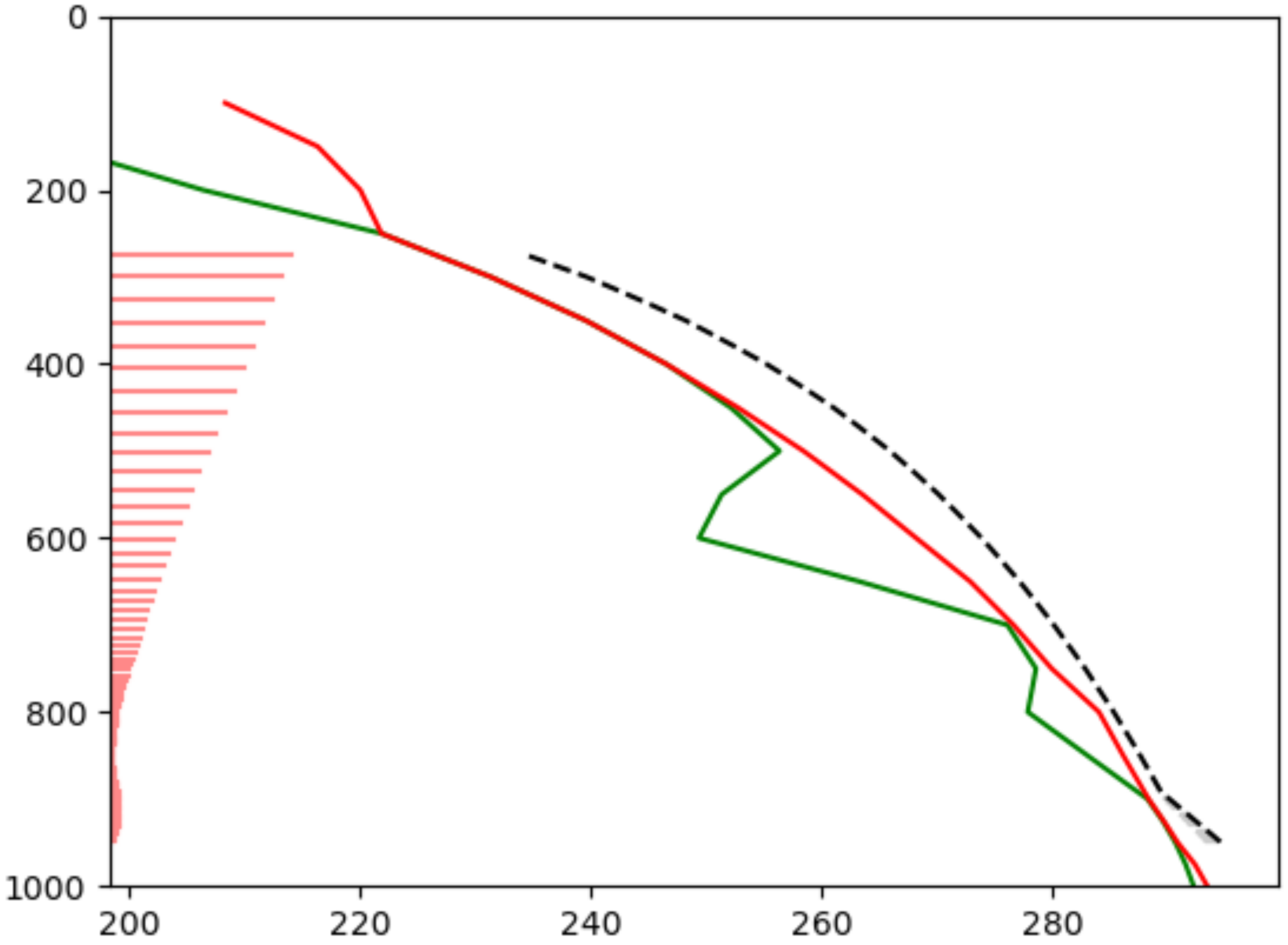
T_j = Temperature measurement just above z
T_i = Temperature measurement just below z
z_j = Height measurement just above z
z_i = Height measurement just below z
P_j = Pressure measurement just above z
P_i = Pressure measurement just below z

if the air parcel is saturated: $\gamma = \Gamma_m(T', P)$
if the air parcel is unsaturated: $\gamma = \Gamma_d$

INITIATION PROBABILITY (P_0)

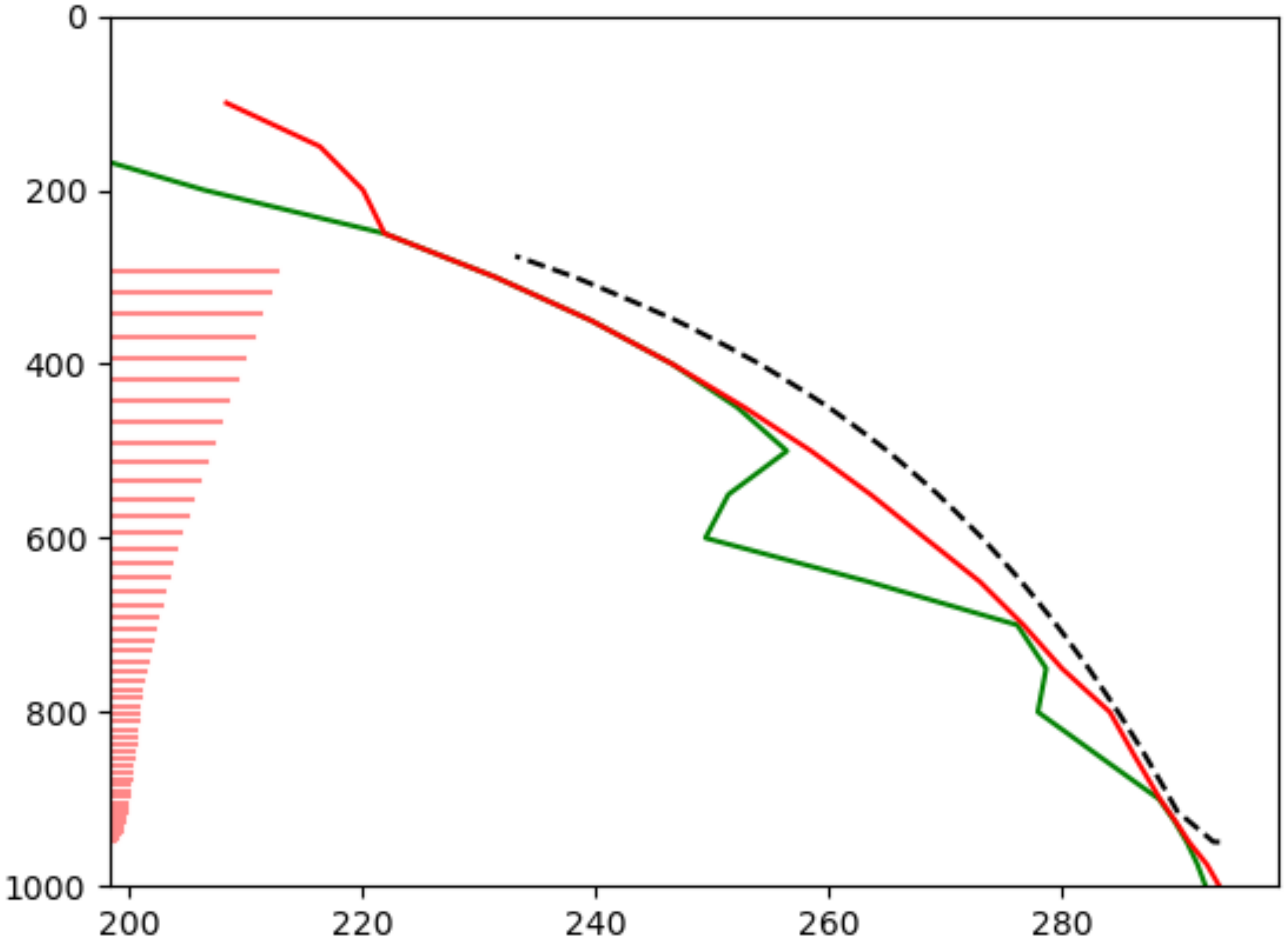


INITIATION PROBABILITY (P_0)



$\hat{T}_v = 294.6$ K, $P_T = 0.072$

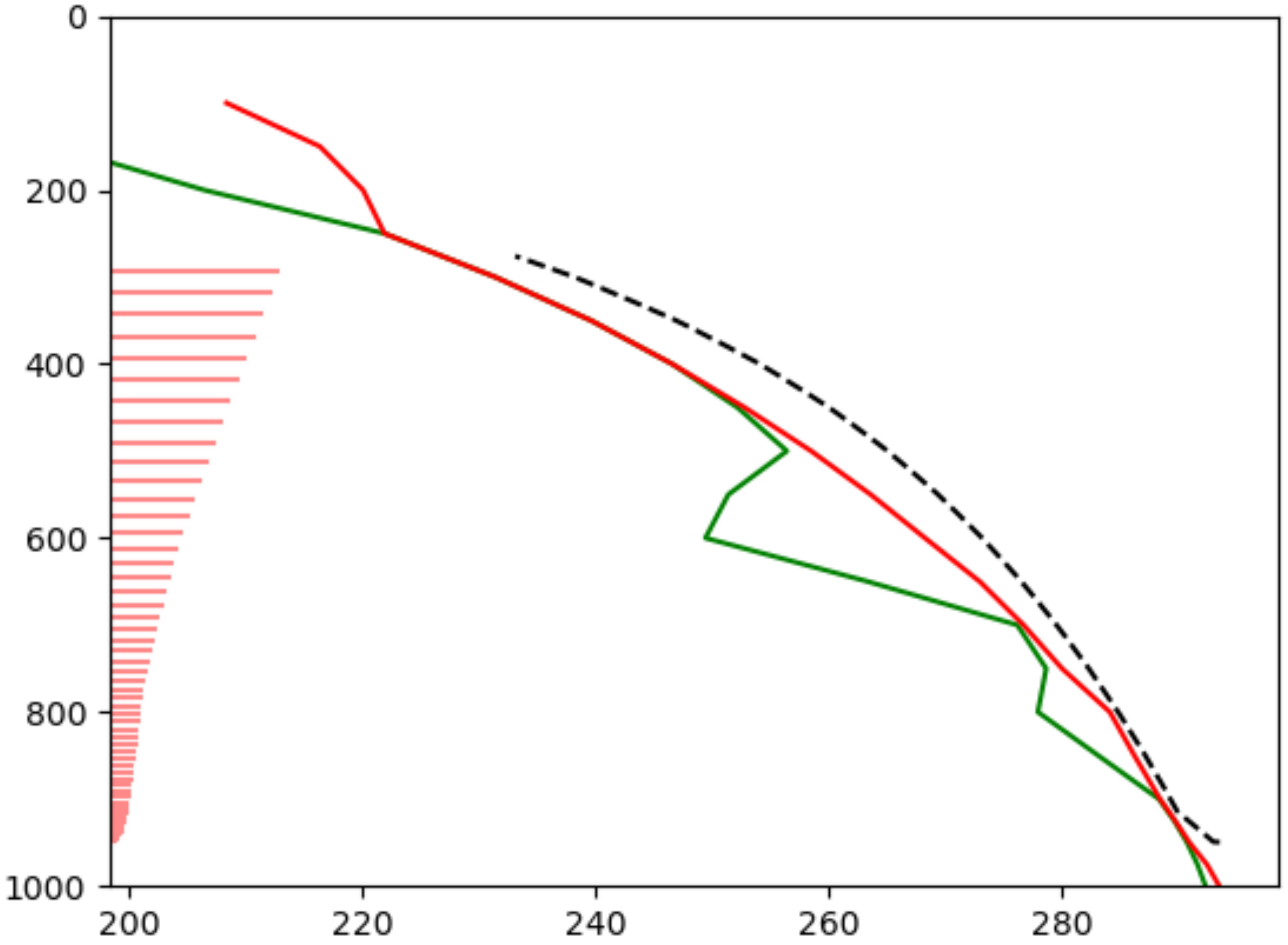
INITIATION PROBABILITY (P_0)



$$\hat{T}_v = 294.6 \text{ K}, P_T = 0.072$$

$$\hat{w} = 7.14 \text{ m s}^{-1}, P_w \approx 0.000$$

INITIATION PROBABILITY (P_0)



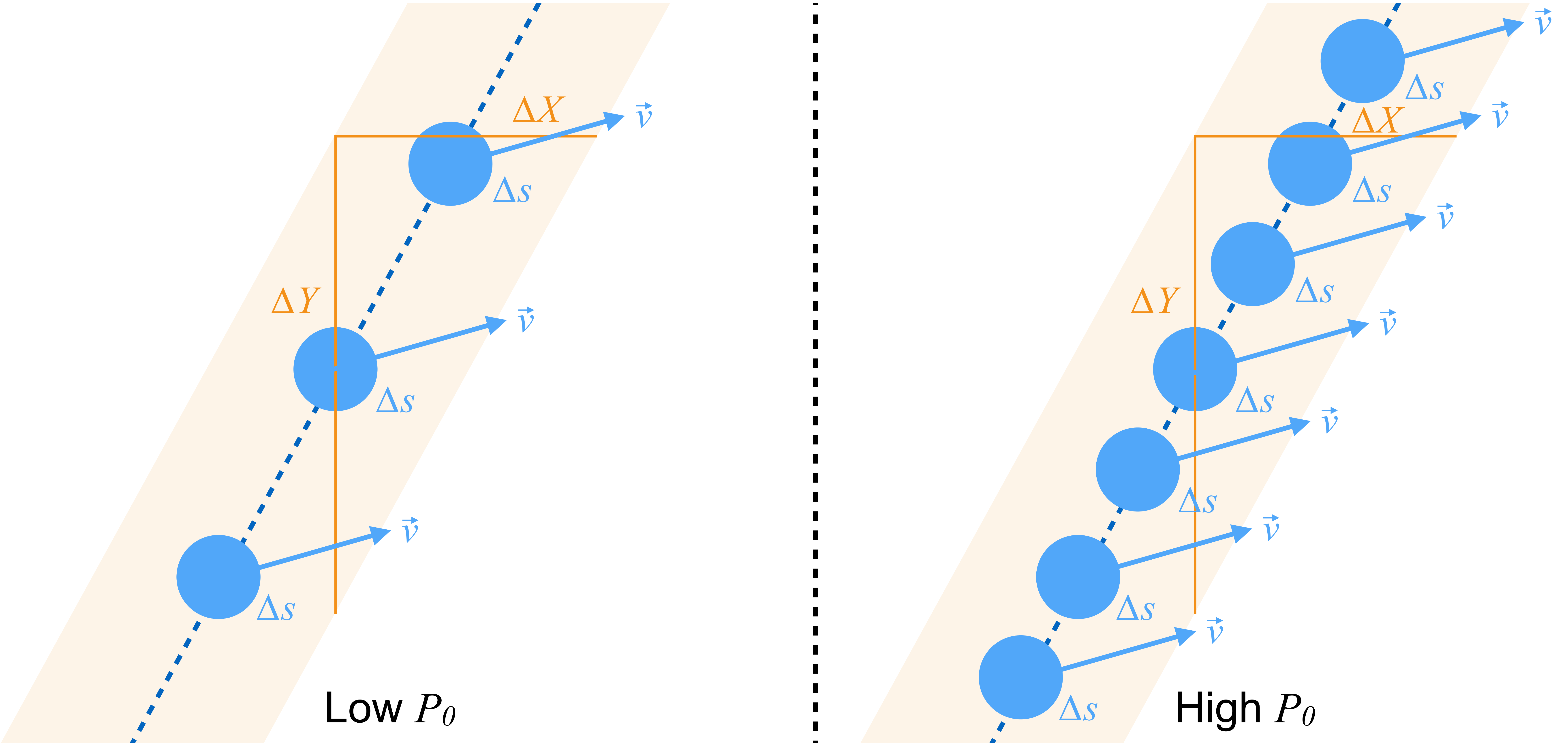
$$\hat{T}_v = 294.6\ \mathrm{K},\ P_T = 0.072$$

$$\hat{w} = 7.14\ \mathrm{m\ s^{-1}},\ P_w \approx 0.000$$

$$P_0 = 0.036$$

MODEL ILLUSTRATION

Frontal Convection

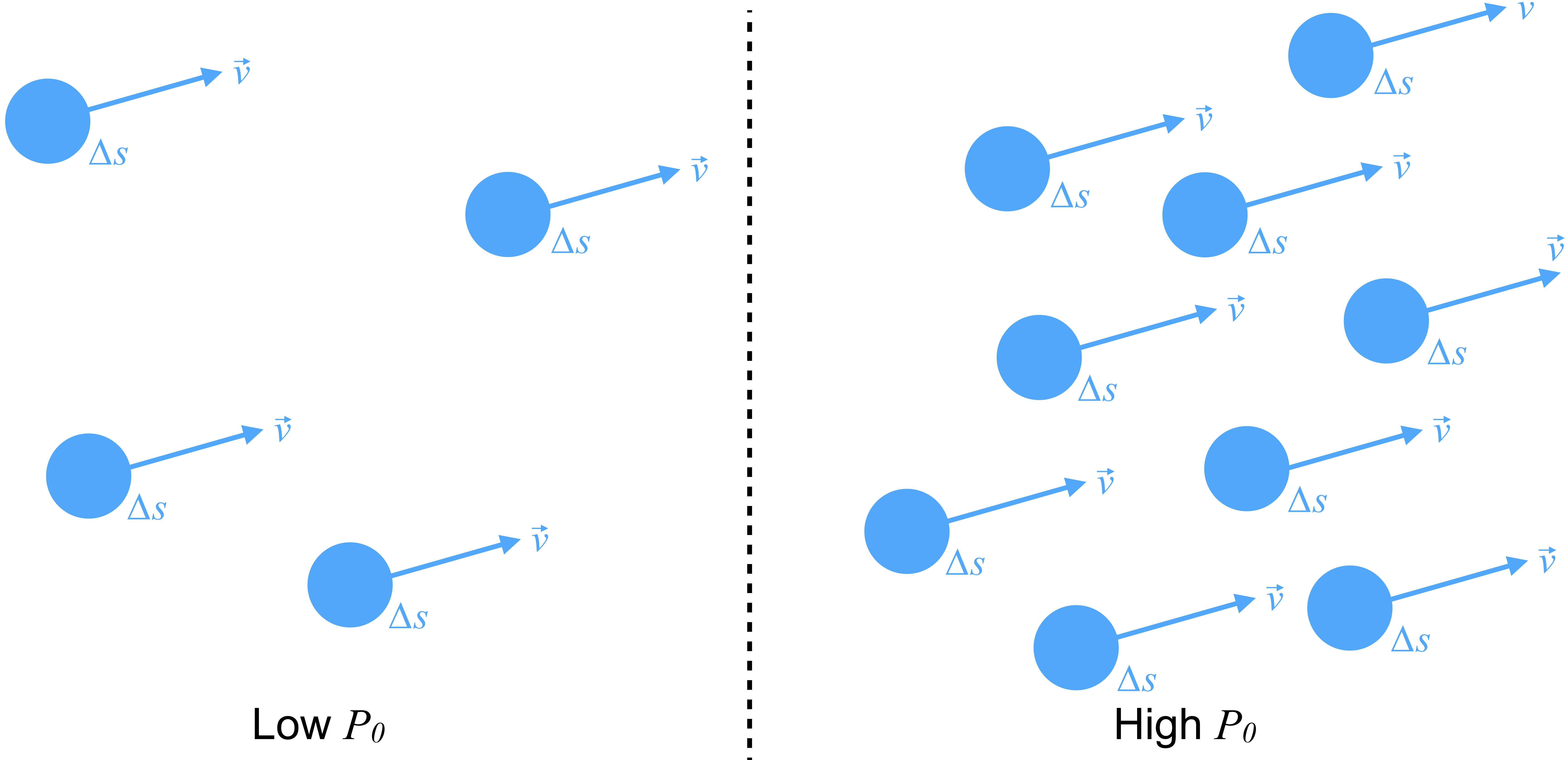


Low P_0

High P_0

MODEL ILLUSTRATION

Non-Frontal Convection



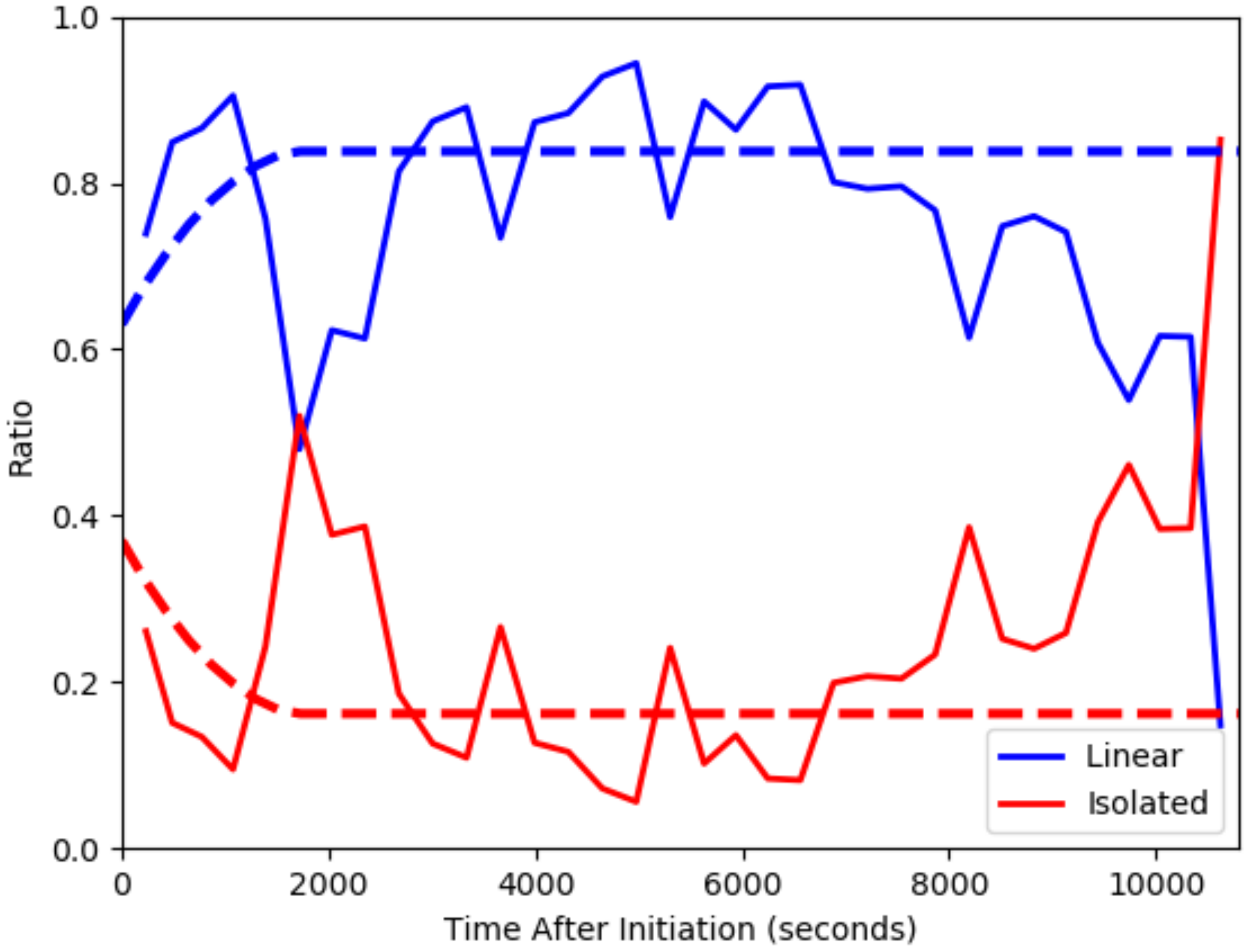
VERIFICATION – RADAR REANALYSIS

- Accuracy of formulation assessed by two methods (1/2):

  1) Re-analysis of radar data (post-event analysis)

     - 3-hour timeframes of radar base reflectivity data obtained from events with at least an Enhanced (Level 3) Risk from SPC
     - Used individual radar data (instead of a composite, this is to avoid inconsistencies that can result from compositing radar data)
     - P_0 calculated by dividing total storm coverage by initiation zone area
     - Mean wind vector obtained from nearest observed RAOB

VERIFICATION – RADAR REANALYSIS

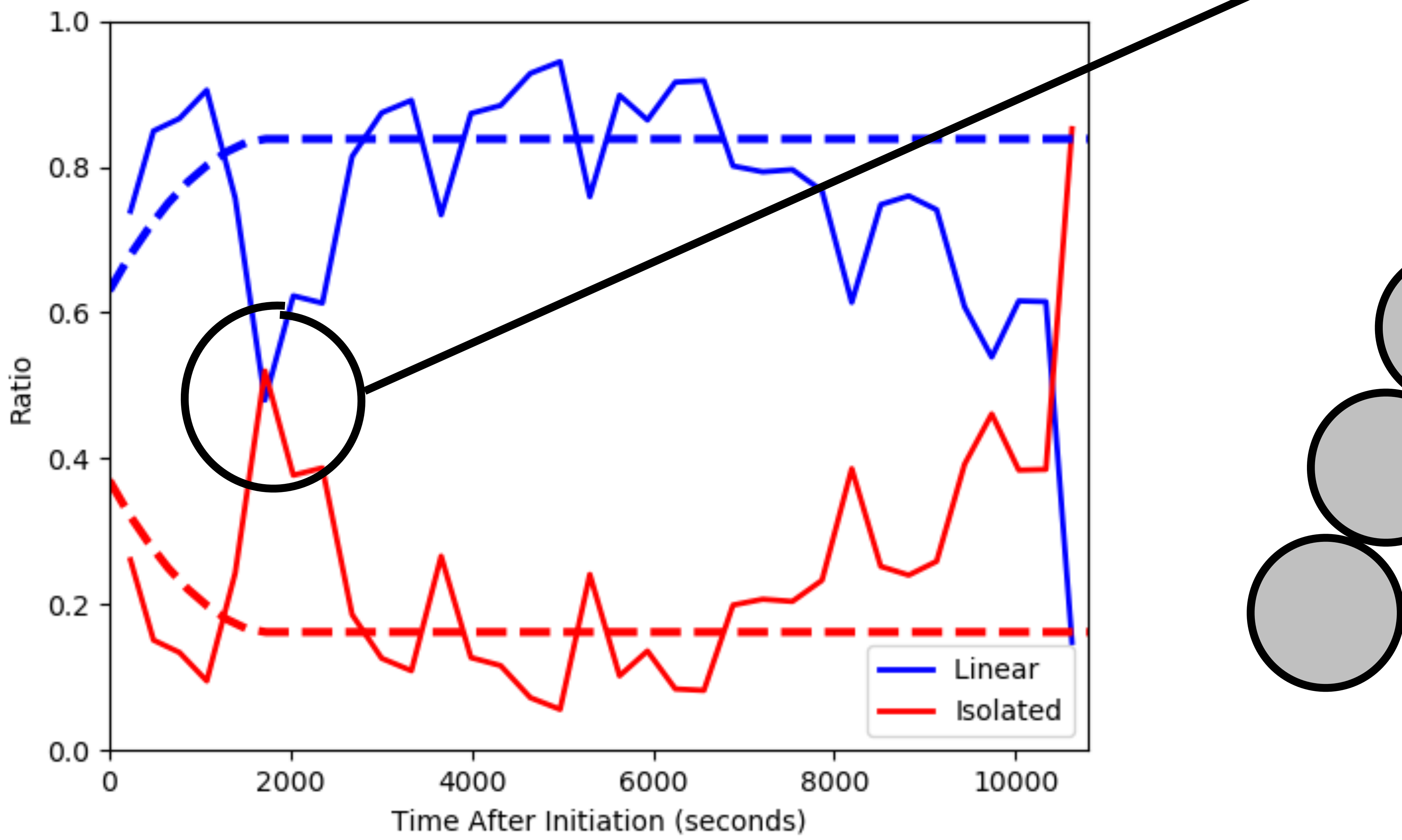


- - - - - - **Predicted**

——— **Observed**

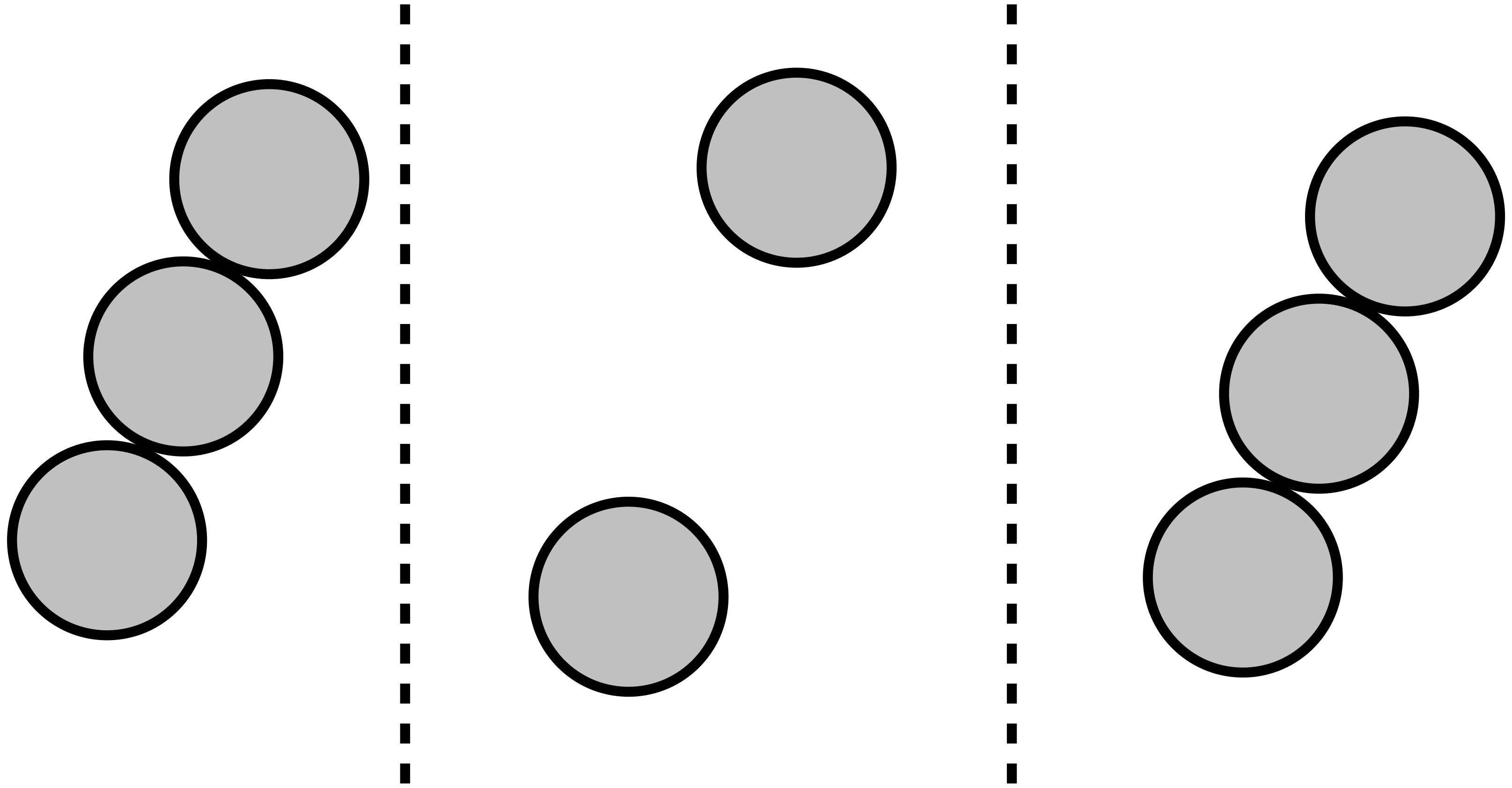
VERIFICATION – RADAR REANALYSIS

Weakening embedded cells not being identified



- - - - - - **Predicted**

—— **Observed**



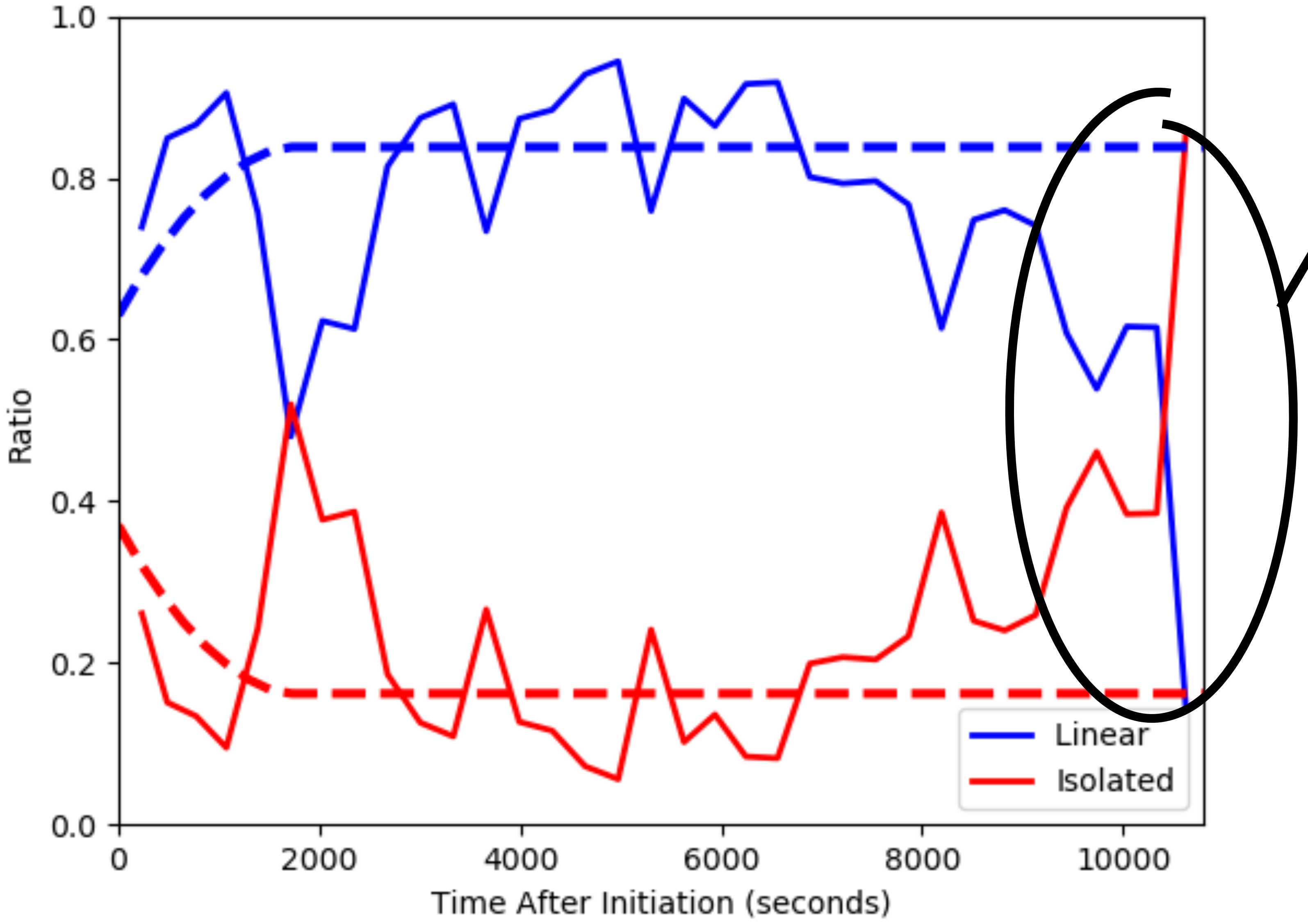
t_0 **Linear**

$t_0 + \Delta t$ **Isolated**

$t_0 + 2 \cdot \Delta t$ **Linear**

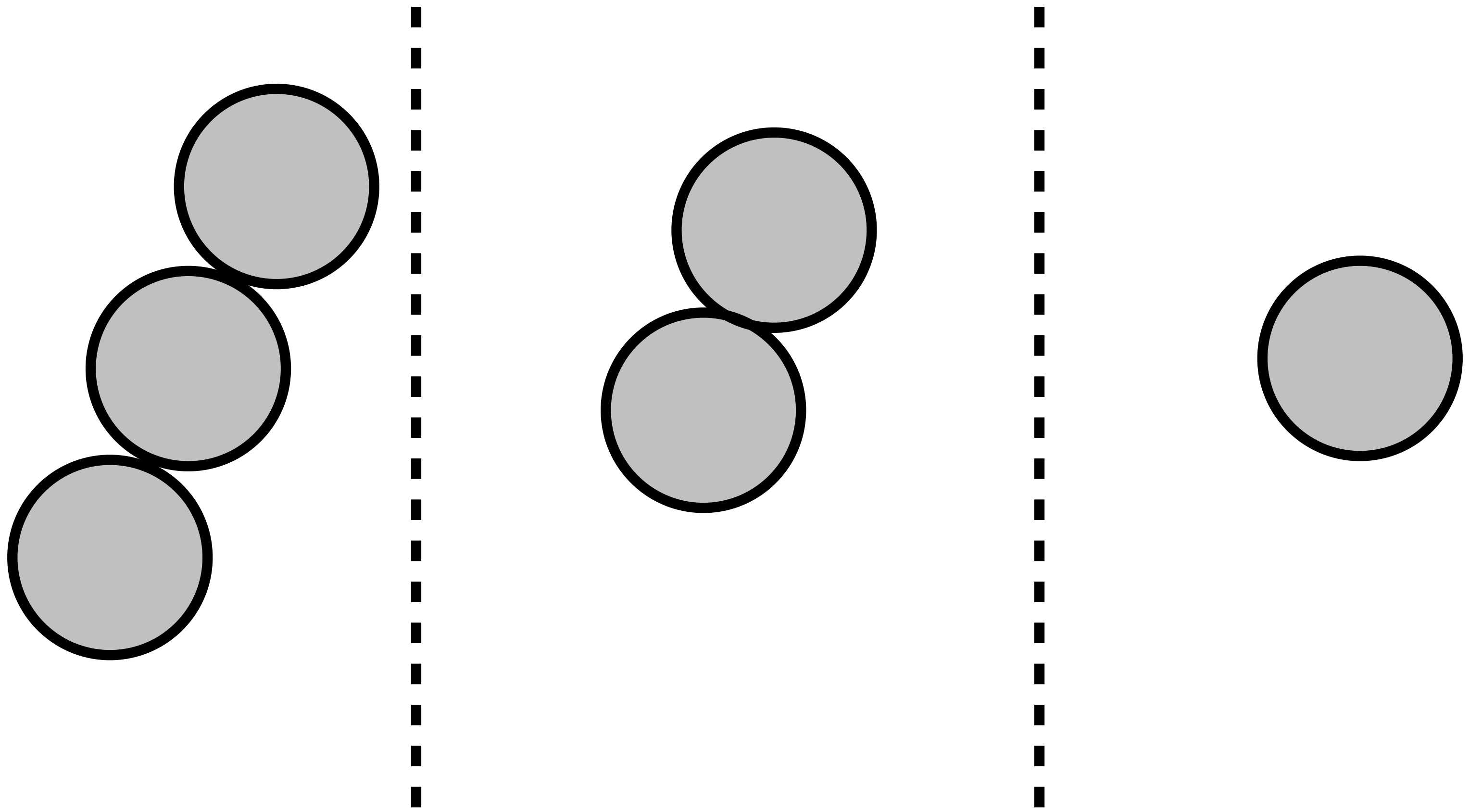
VERIFICATION – RADAR REANALYSIS

Storms moving out of radar range



╌╌╌╌ **Predicted**

──── **Observed**

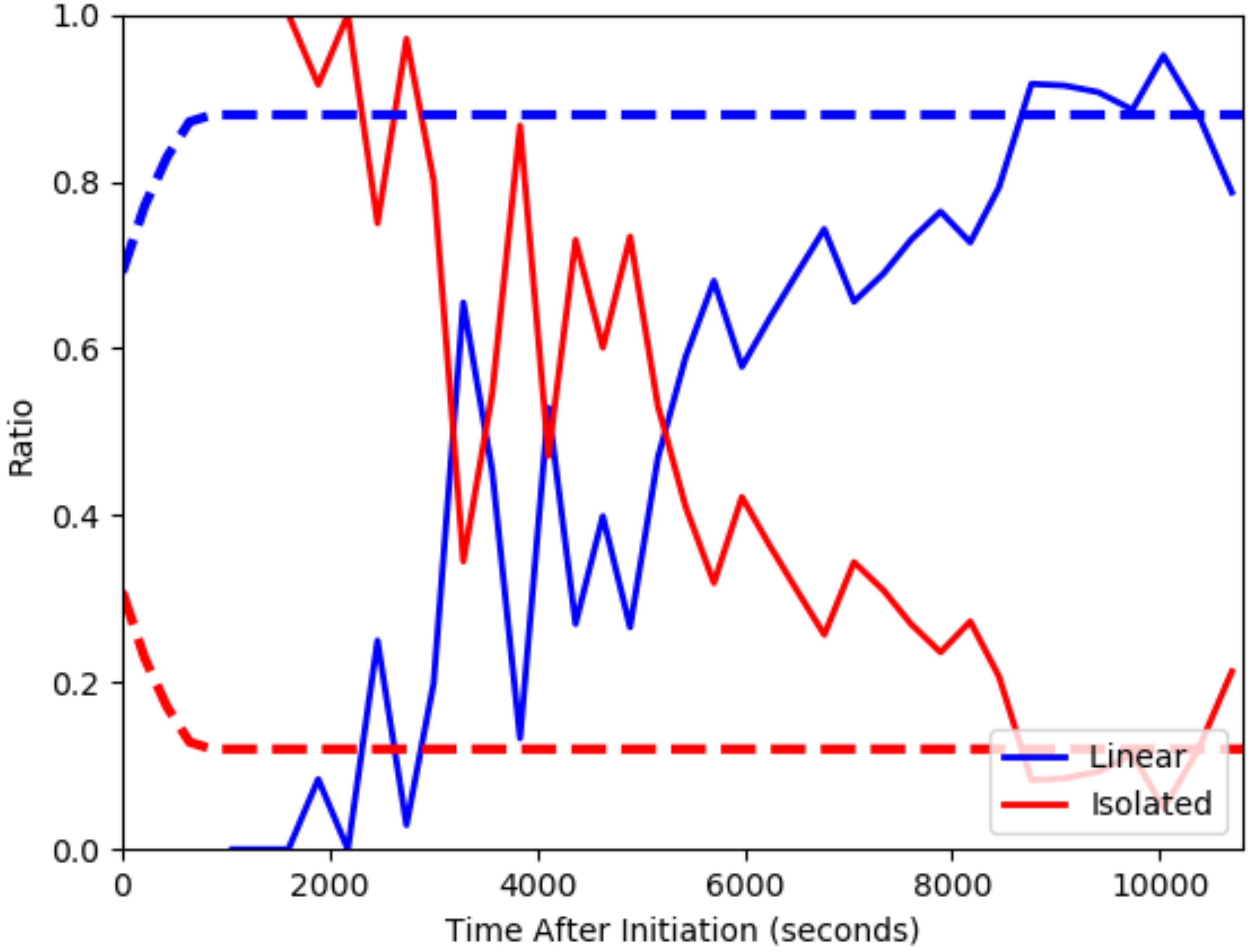


t_0
Linear

$t_0 + \Delta t$
Linear

$t_0 + 2 \cdot \Delta t$
Isolated

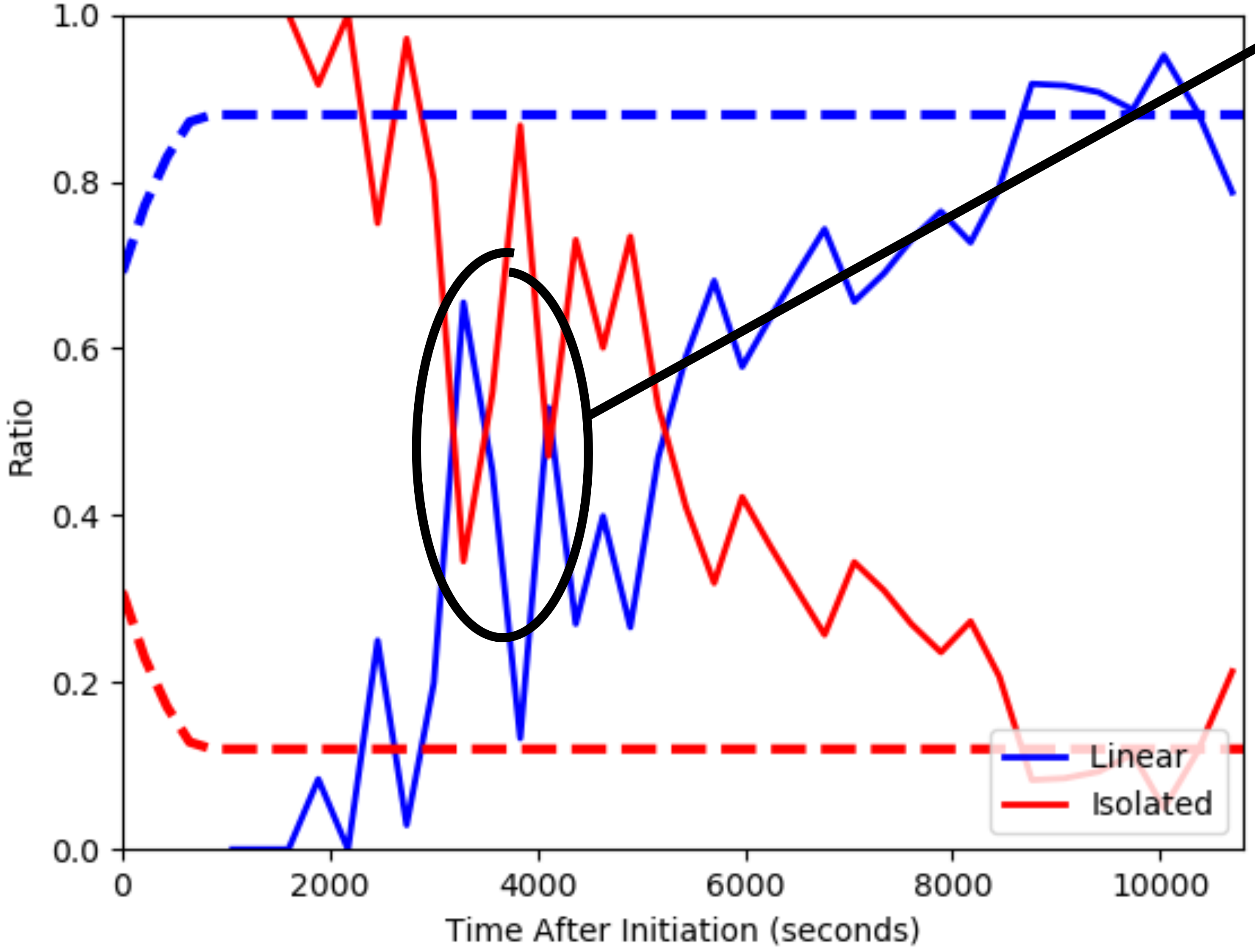
VERIFICATION – RADAR REANALYSIS



- - - - - - **Predicted**

—— **Observed**

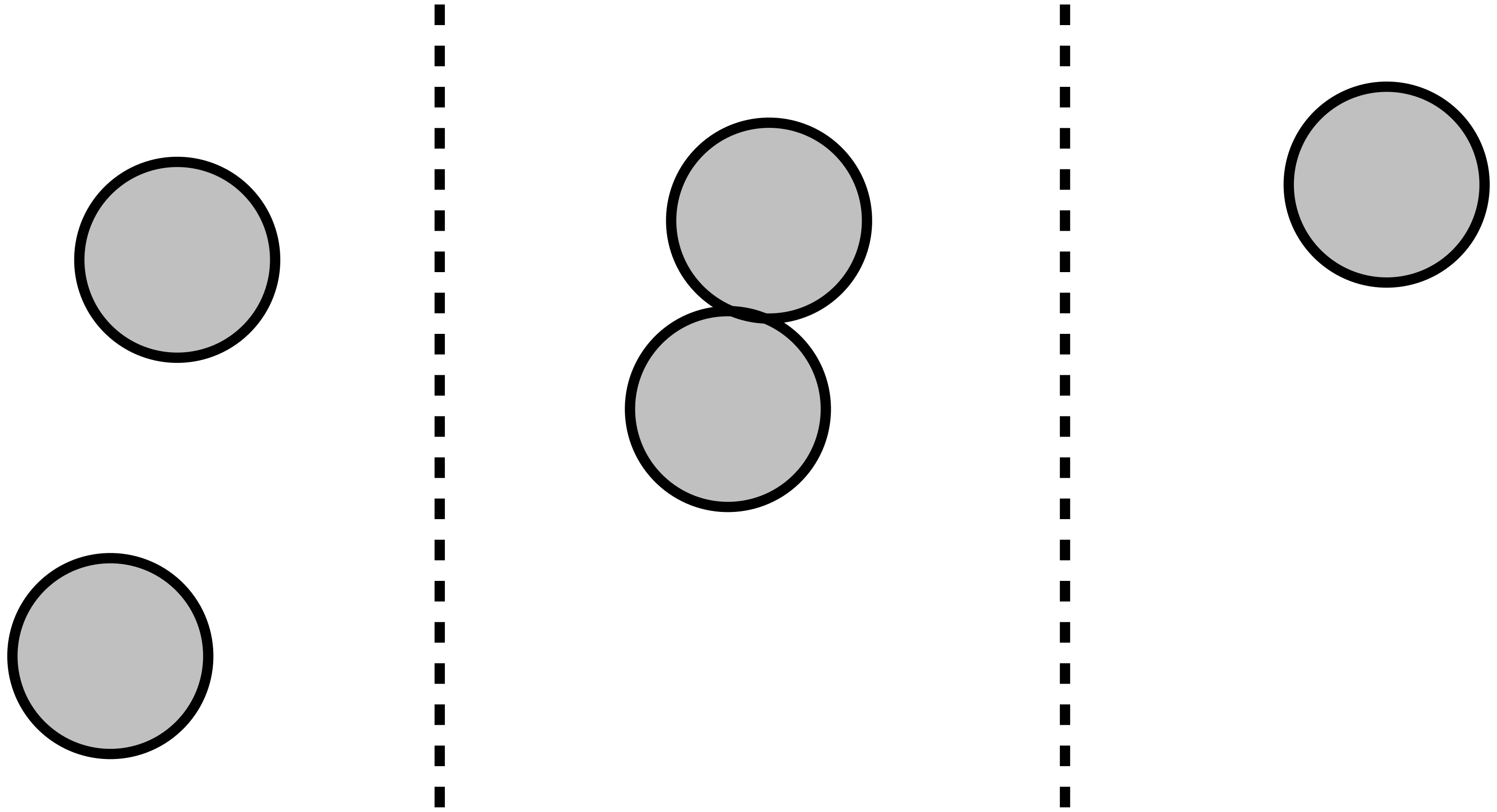
VERIFICATION – RADAR REANALYSIS



Cell Mergers

- - - - - - **Predicted**

━━━━ **Observed**

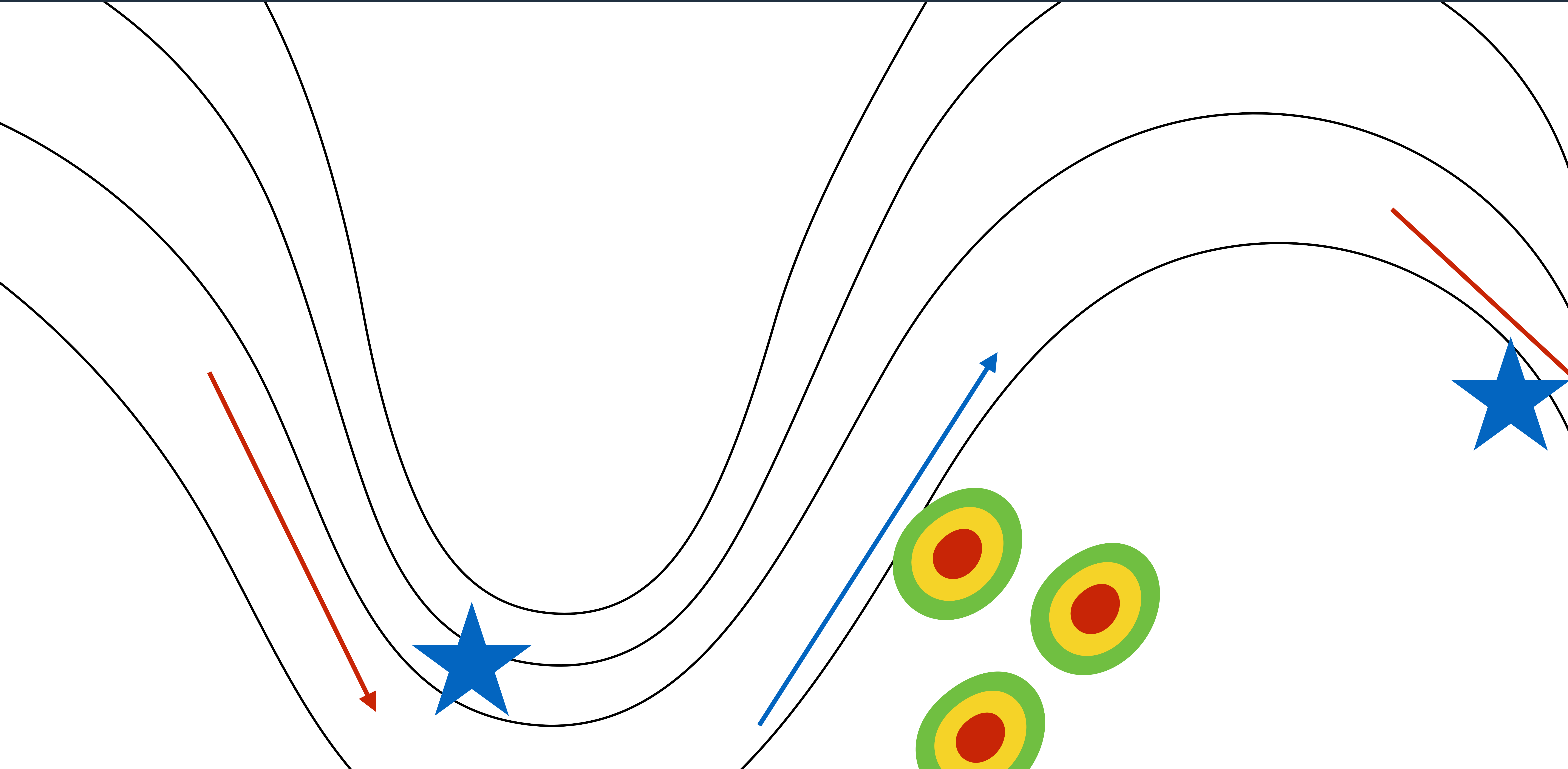


t_0 **Isolated**

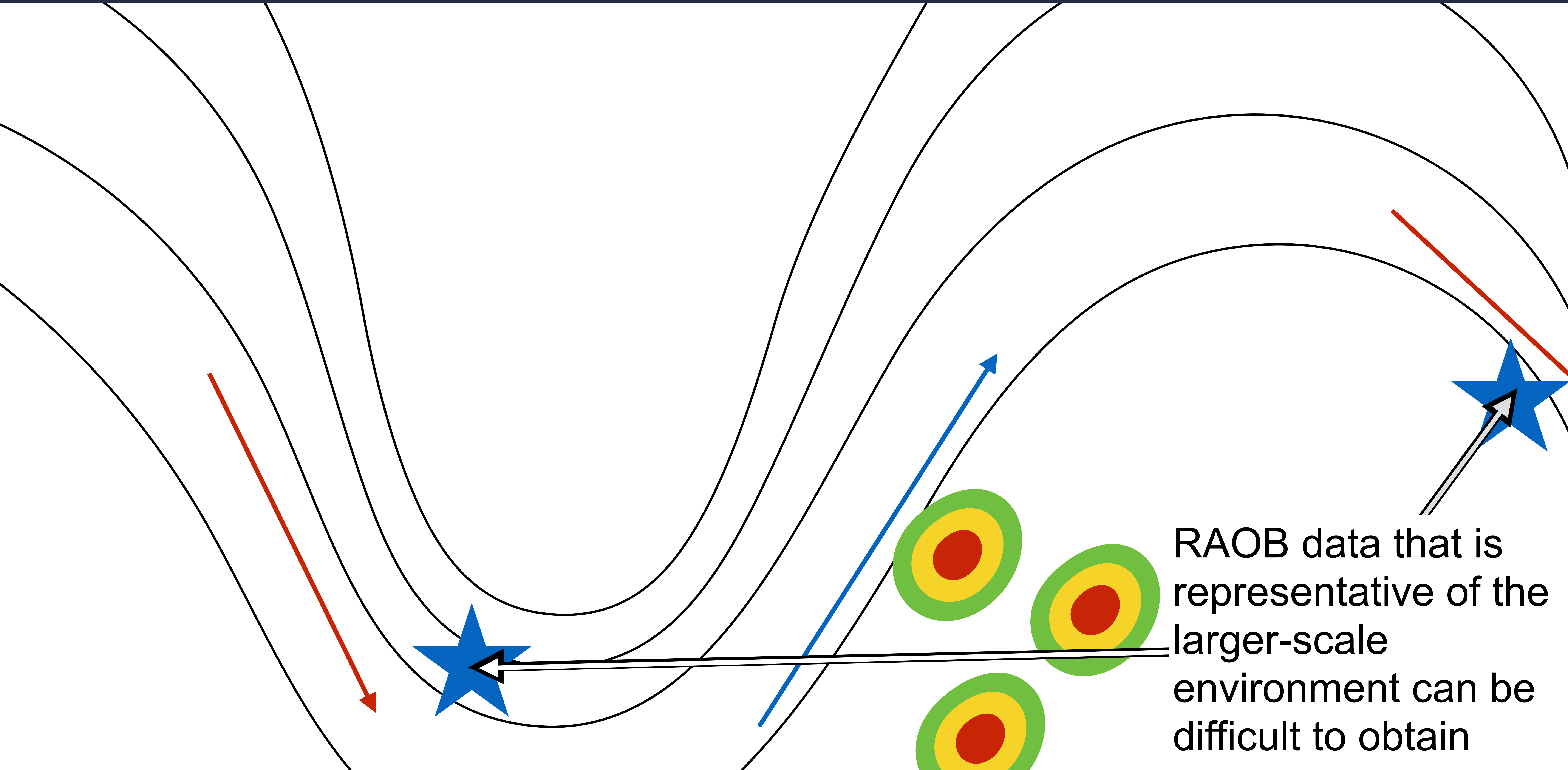
$t_0 + \Delta t$ **Linear**

$t_0 + 2 \cdot \Delta t$ **Isolated**

VERIFICATION – RADAR REANALYSIS



VERIFICATION – RADAR REANALYSIS



RAOB data that is representative of the larger-scale environment can be difficult to obtain

RADAR REANALYSIS RESULTS

- Key evaluation statistics:

Mean Error:	21.1 %
Median Error:	8.0 %
Bias:	+ 1.9 % Linear

- Large mean error primarily caused by cases that involved a prediction for purely isolated modes but actually involved largely linear modes (caused by observational error?)

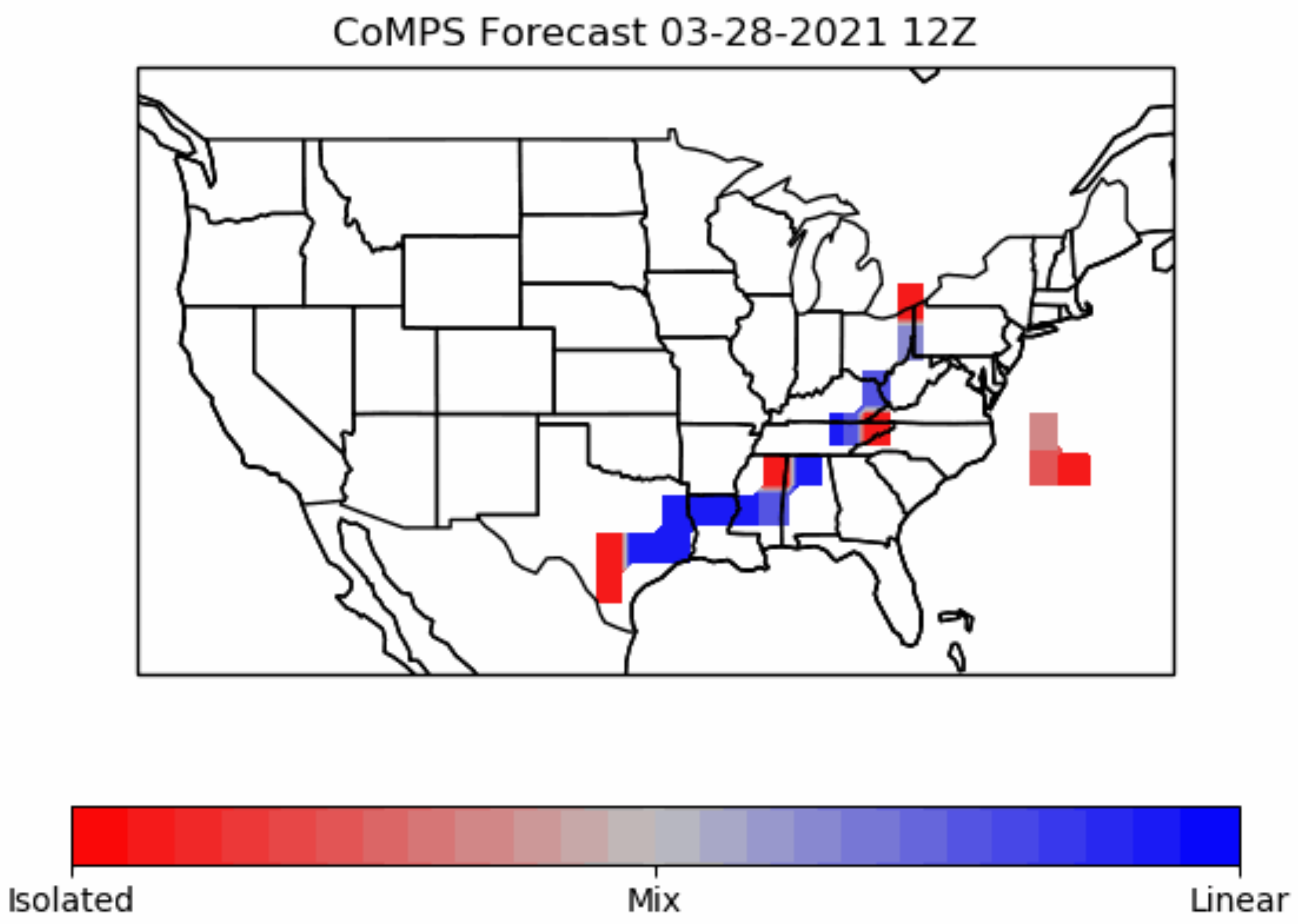
VERIFICATION – MODEL ANALYSIS

- Accuracy of formulation assessed by two methods (2/2):

  2) Analysis of model output from 0.5° Global Forecast System (GFS) and 0.5° Global Ensemble Forecast System (GEFS)

     - When SPC issues a Day 4+ outlook area (15% or 30%), a convective mode forecast is produced using the 00Z model suite (last suite a forecaster would have seen just prior to issuing the Day 4+ area)
     - P_0 estimated from temperature and vertical velocity fields
     - Point forecast soundings used to determine mean wind vector
     - Data assimilation scheme from Wang et. al. 2004 (ensemble member dressing) used to more accurately estimate forecast model error
     - Plot storm mode prediction for minimum distance parameter value $D*$
     - **Only determines storm mode; not severity or structure**

VERIFICATION – MODEL ANALYSIS

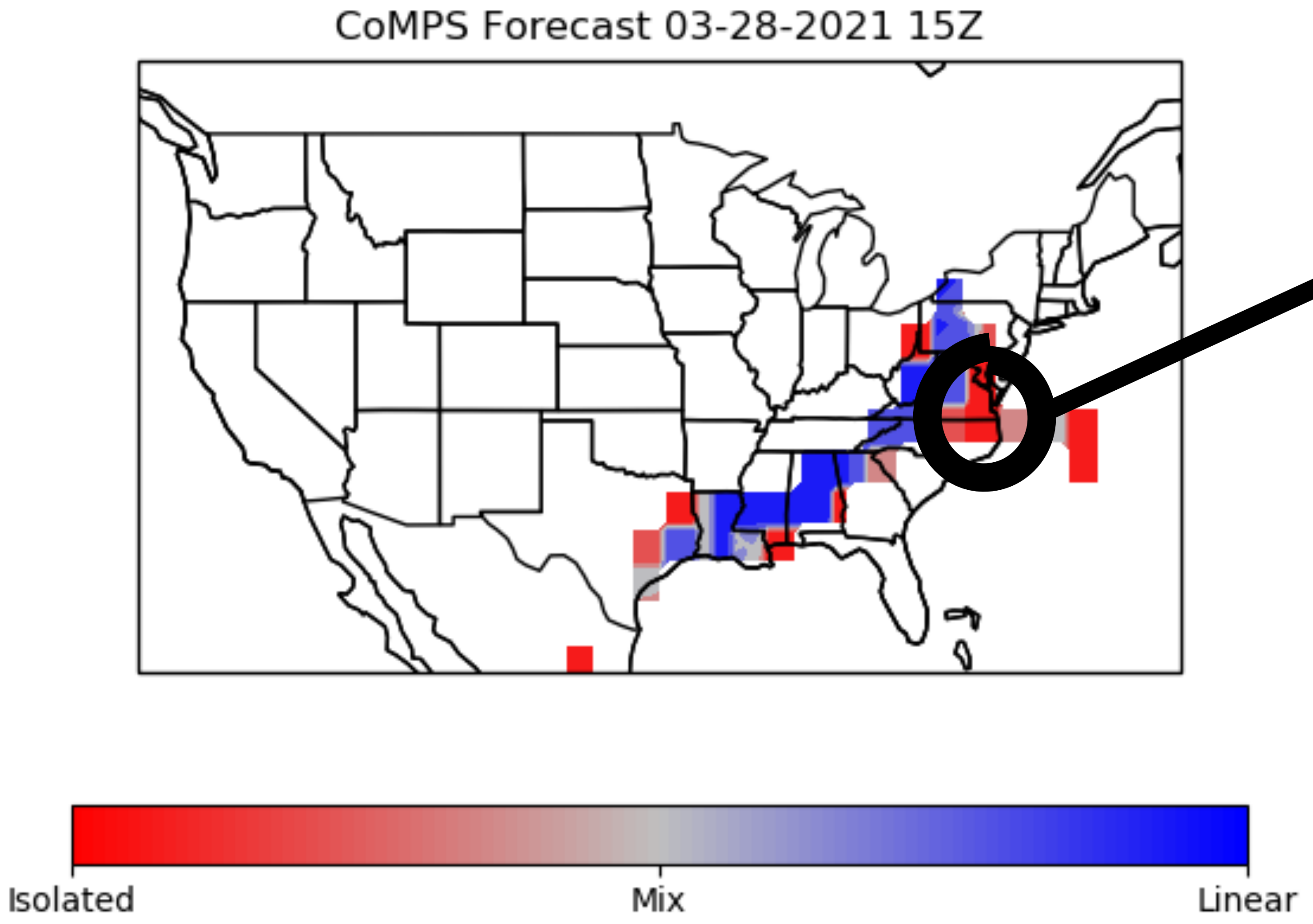


- Convective mode prediction made for 12Z 3-28-2021 to 12Z 3-29-2021 using GFS and GEFS data from 03-25-2021 00Z run

- QLCS event occurred along a cold front

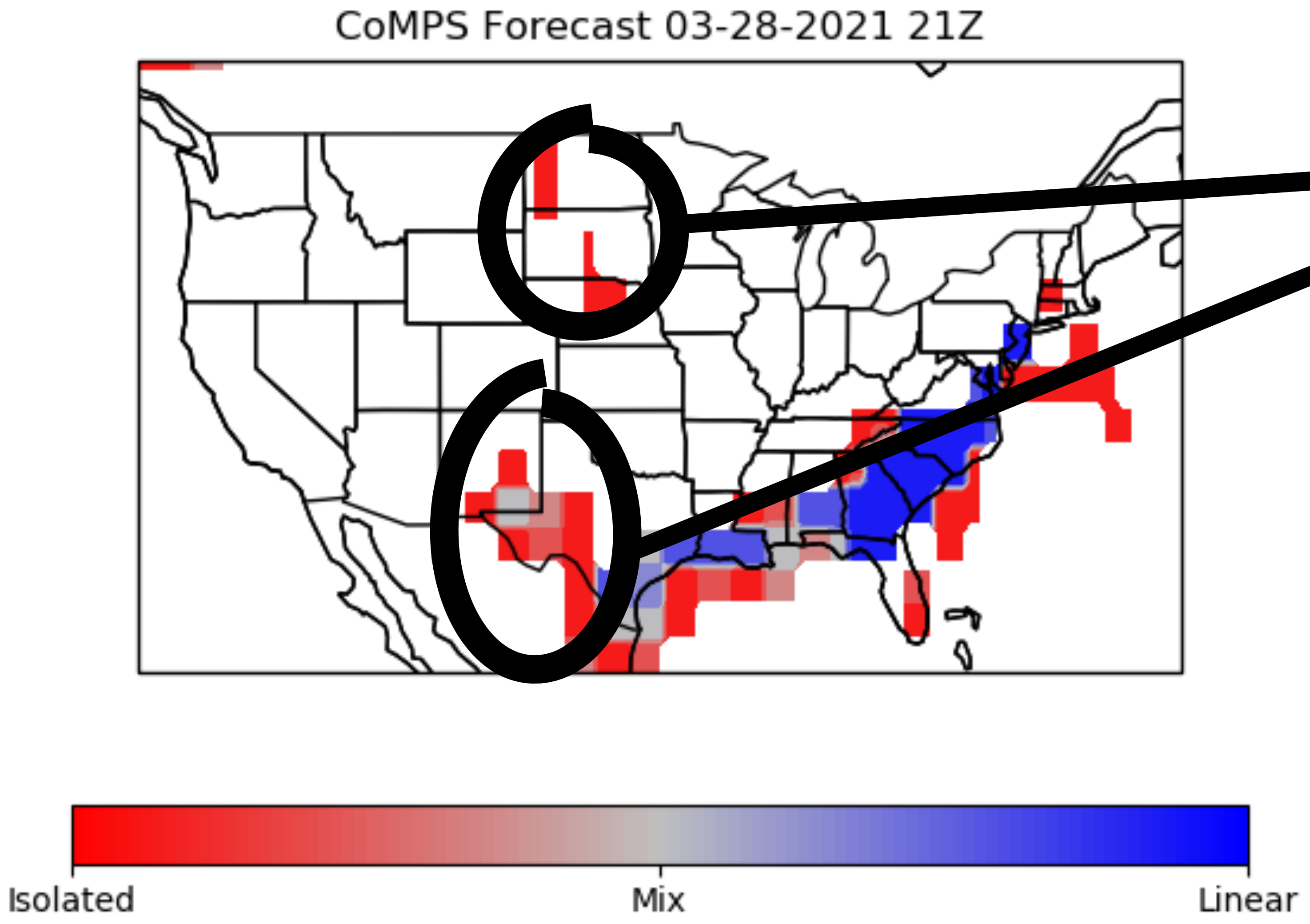
- A few prefrontal cells formed ahead of the cold front in North Carolina, but these cells were short-lived and non-severe

VERIFICATION – MODEL ANALYSIS



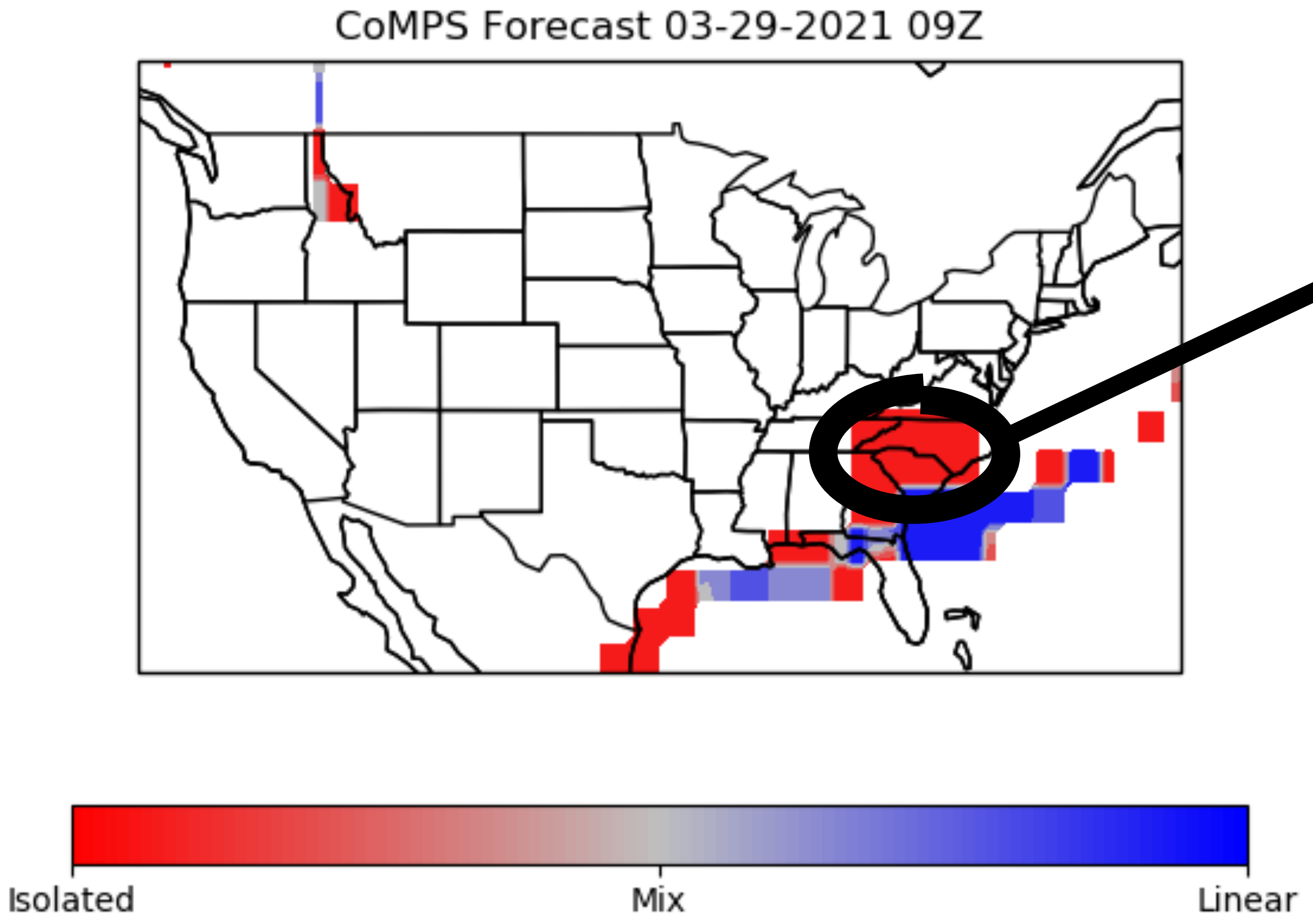
System predicting isolated convection ahead of the cold front

VERIFICATION – MODEL ANALYSIS



Orographic lift or other terrain influences improperly resolved?

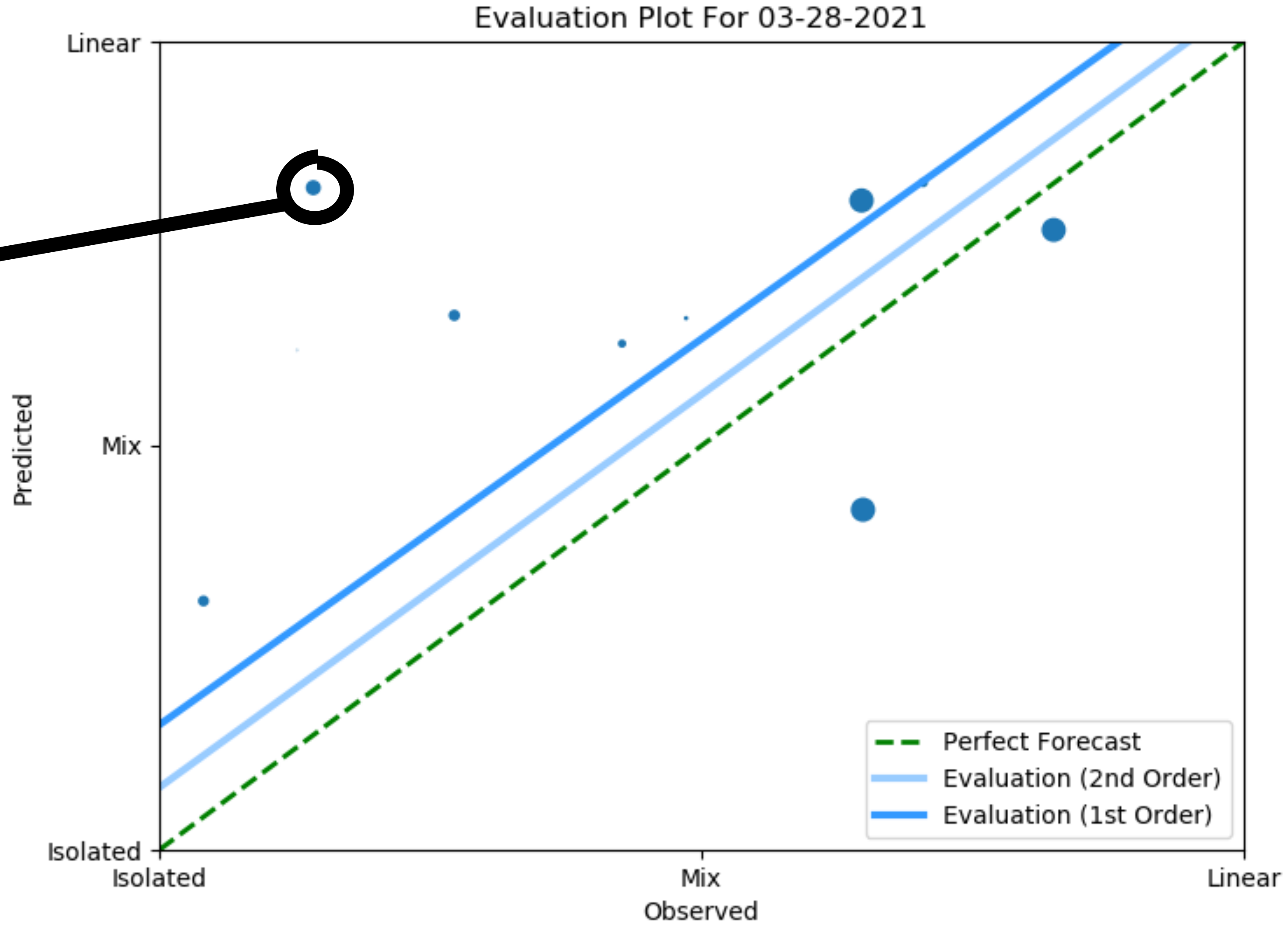
VERIFICATION – MODEL ANALYSIS



Prediction for anafrontal (post-front) convection?

MODEL ANALYSIS RESULTS

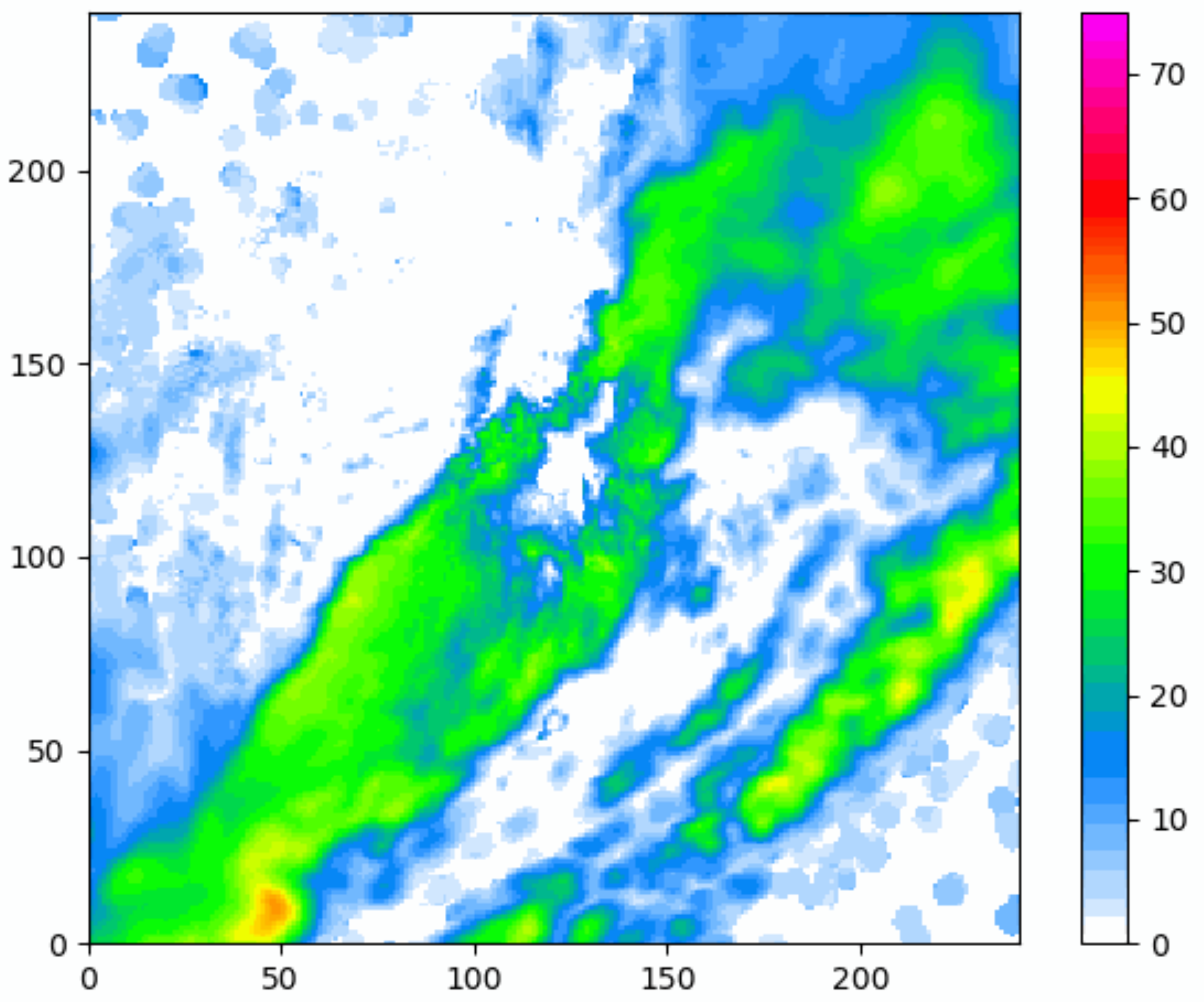
- 3-hour time slots of radar base reflectivity data used as verifying observation
- Objective analysis of radar data determines percentage of isolated modes and linear modes that were observed



Outlier due to poor objective analysis

MODEL ANALYSIS RESULTS

- Objective analysis only identified embedded heavier cores of precipitation
- Misidentified as isolated modes



MODEL ANALYSIS RESULTS

Linear	Observed	Not Observed
Forecasted	*a*	*b*
Not Forecasted	*c*	*d*

Isolated	Observed	Not Observed
Forecasted	*a*	*b*
Not Forecasted	*c*	*d*

Linear	Observed	Not Observed
Forecasted	*a*	*b*
Not Forecasted	*c*	*d*

Isolated	Observed	Not Observed
Forecasted	*a*	*b*
Not Forecasted	*c*	*d*

$$A = \frac{a_L + d_L}{a_L + b_L + c_L + d_L} = \frac{a_I + d_I}{a_I + b_I + c_I + d_I}$$

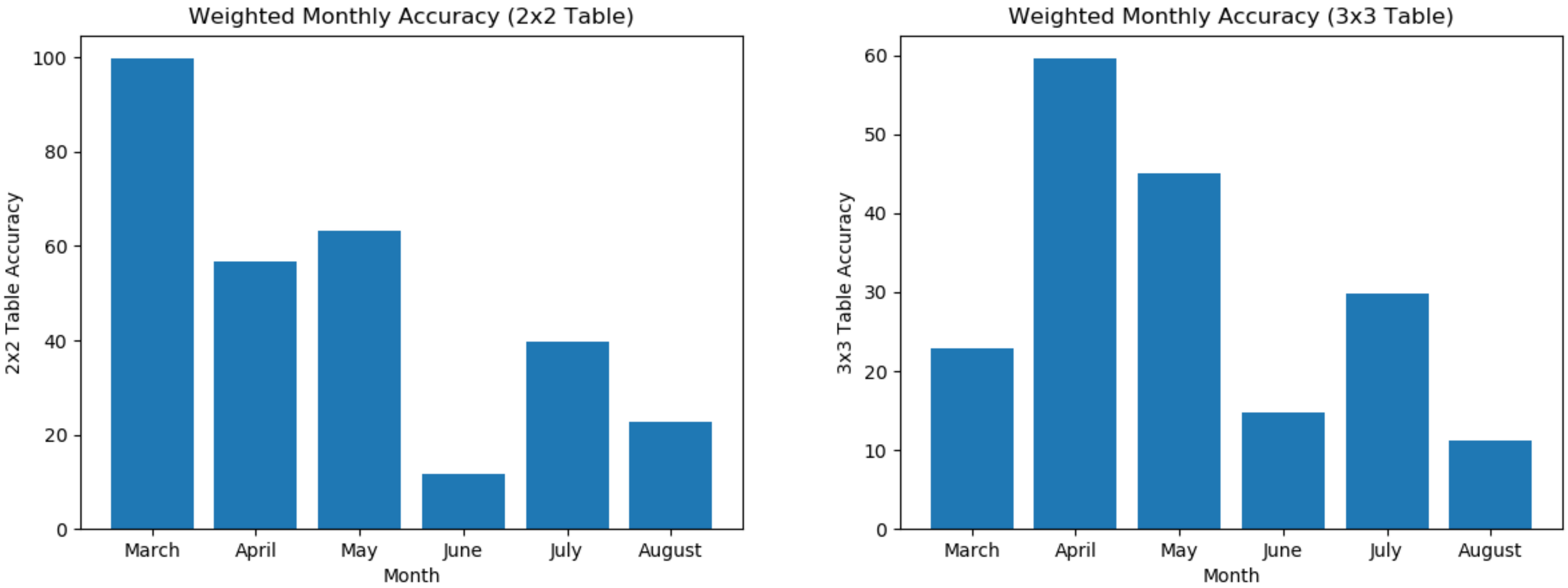
$$A_W = \frac{a_{LW} + d_{LW}}{a_{LW} + b_{LW} + c_{LW} + d_{LW}} = \frac{a_{IW} + d_{IW}}{a_{IW} + b_{IW} + c_{IW} + d_{IW}}$$

- Unweighted accuracy: 49 %
- Weighted accuracy: 54 %

+1.16 % bias towards isolated modes

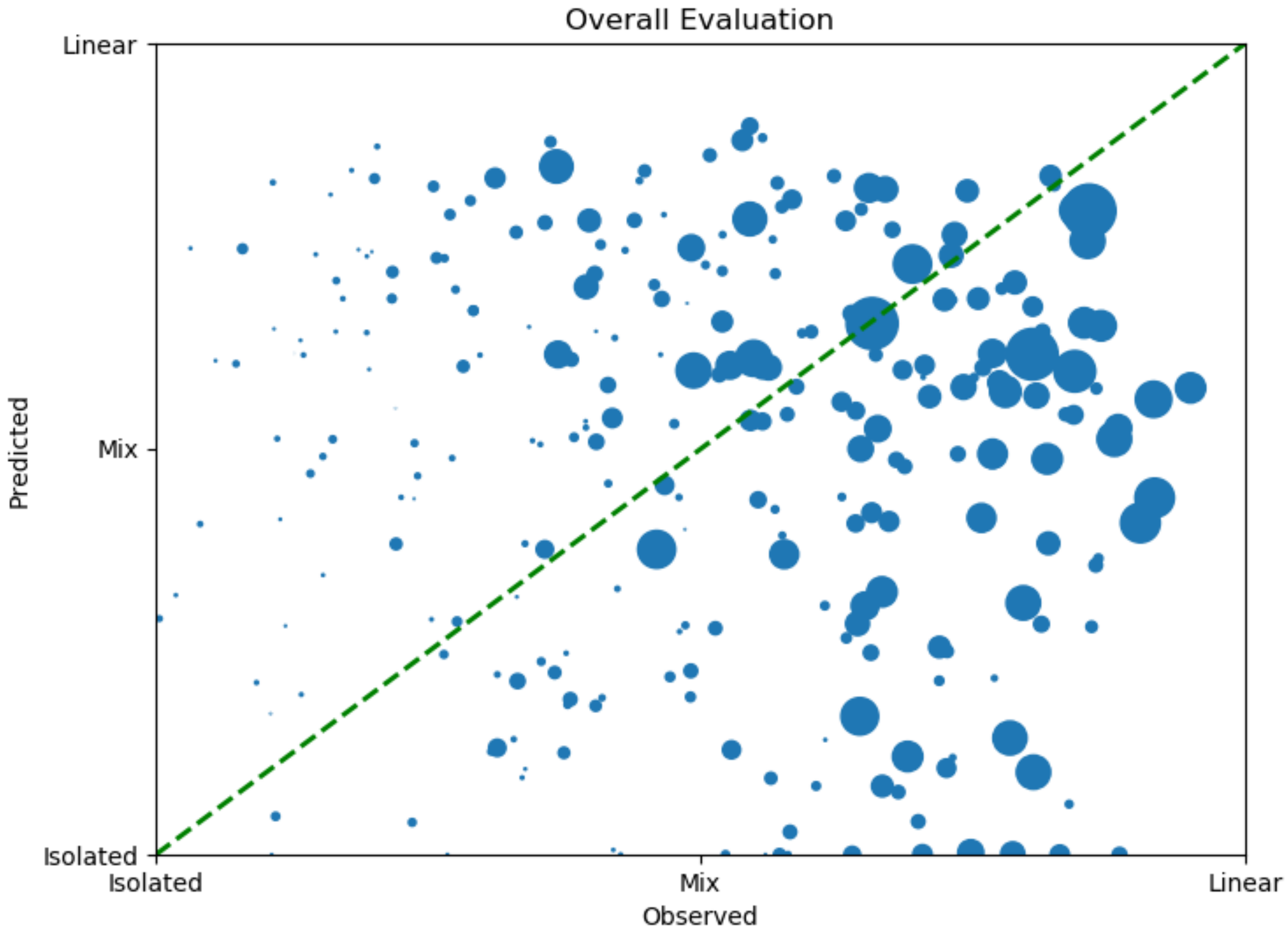
Median Error: 24.39 %

MODEL ANALYSIS RESULTS



- Overall forecast accuracy was highest during the spring months
- Poor accuracy in the summer likely attributable to the chaotic nature of summer events, which are heavily driven by mesoscale and microscale features

MODEL ANALYSIS RESULTS



STRENGTHS

- The forecasting tool performed well in the following situations:

  - Synoptically-driven events where mesoscale and/or microscale features are either insignificant or non-existent
  - Dry line setups in the Great Plains
  - Cold front setups in the Deep South, cold fronts in the Plains were poorly modeled by the GFS/GEFS
  - Non-frontal convection where observed storm width is comparable to Δs

- Cold fronts have a tendency to trend slower on subsequent model runs, which can allow a dry line to form in the Plains before the cold front hits the warm sector

- Dry lines are very rare in the Deep South, effectively eliminating this errant trend

WEAKNESSES

- The forecasting tool performed poorly when storm-scale processes significantly affected the larger-scale environment. Examples include:

  - Outflow boundaries
  - Localized weakening of inversions
  - Localized destabilization of the boundary layer (typically an overnight phenomenon)
  - Remnant mesoscale convective systems (MCSs)
  - Mesoscale convective vortices (MCVs)

- Global models, at their current grid resolution, simply cannot resolve these processes with any degree of consistency or reliability

- Significant errors in GFS/GEFS output also lead to poor forecast accuracy

SUMMARY

- Primary goal was to establish proof of concept, which has arguably been fulfilled

- Forecast tool performed well in some situations and performed poorly in some situations, though the performance was generally positive

- Gained insight into potential weaknesses and shortcomings, which are important for forecasters to know

- Established a theoretical framework that can be fine-tuned in future studies

FUTURE WORK

- Attempt to relate storm geometry to ambient conditions (assumed an ellipse with roughly equivalent major and minor axes)

- Better modeling for storm cold pools (accounting for the rate of expansion of strong heating and weakening in the presence)

- Time-dependent modeling for right-moving and left-moving supercells that deviate from mean wind vector (essentially factoring in hodograph shape)

- Alternative classification schemes for linear and isolated (used graphical discriminant function here, but other potentially relevant techniques exist)

- Research to more directly relate P_0 and storm size Δs to ambient environment

FUTURE WORK

- Compare this theoretical and statistical approach to a pure machine learning approach

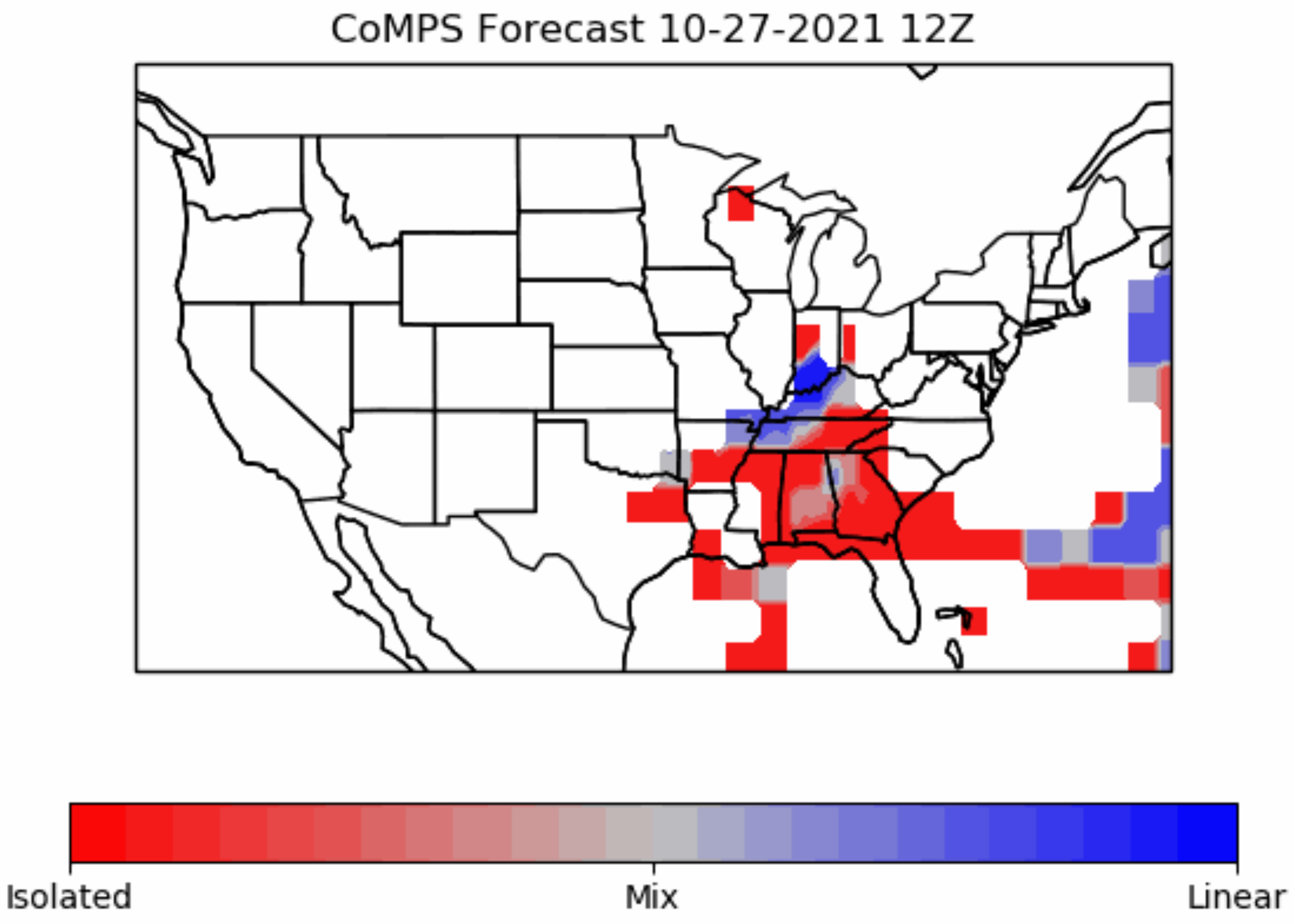
- Evaluate forecasts for different global models (e.g. ECMWF, GDPS, UKMET)

- Alternate schemes for parcel tracing algorithm (e.g. accounting for drag, mixing, pressure perturbations, wind shear destroying/enhancing updrafts)

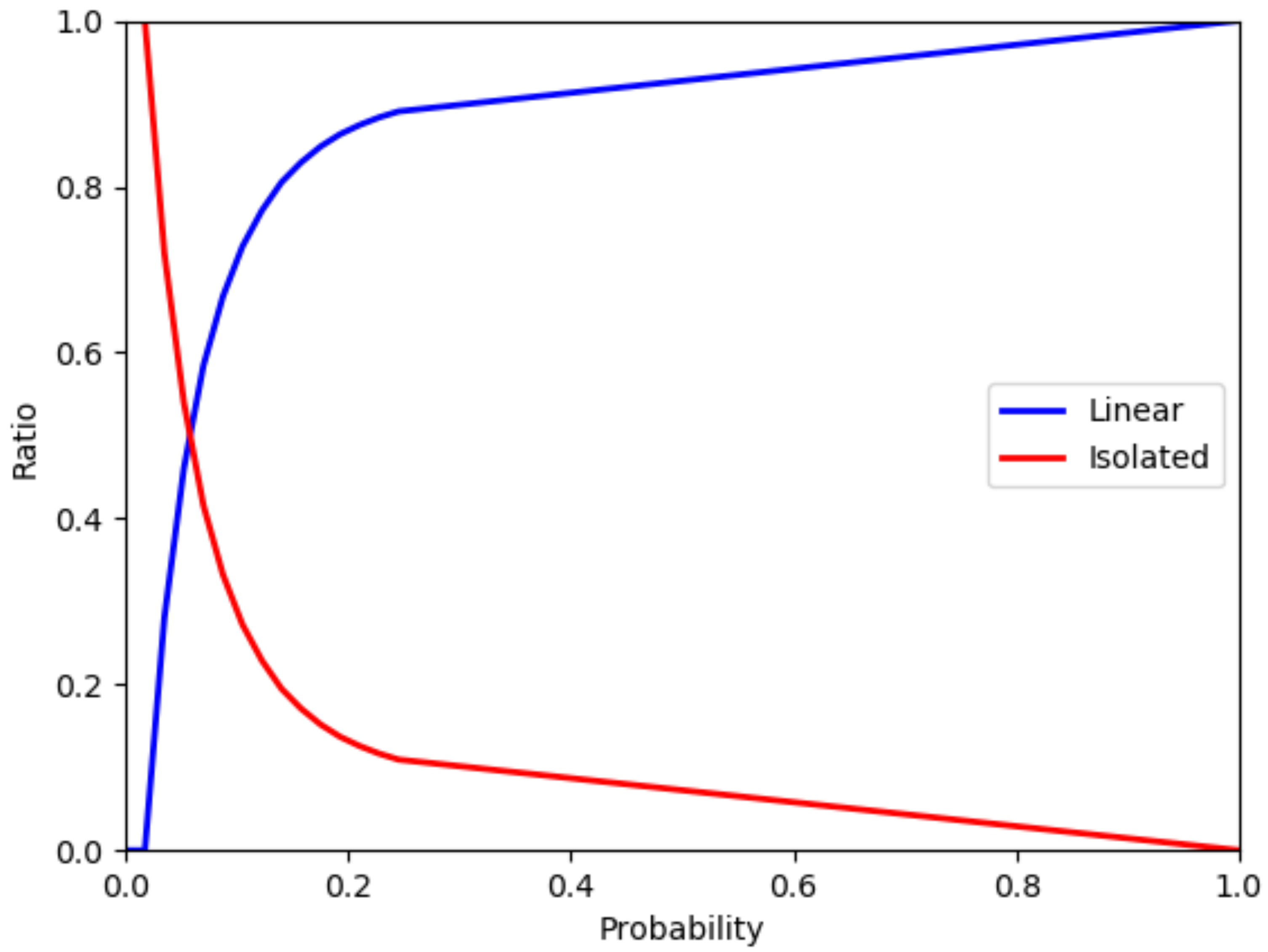
- Accounting for elevated convection, using a 100mb mixed layer parcel can underrepresent or entirely disregard the potential for elevated cells

- An evaluation of how much skill is added (or lost) when forecasters use this tool

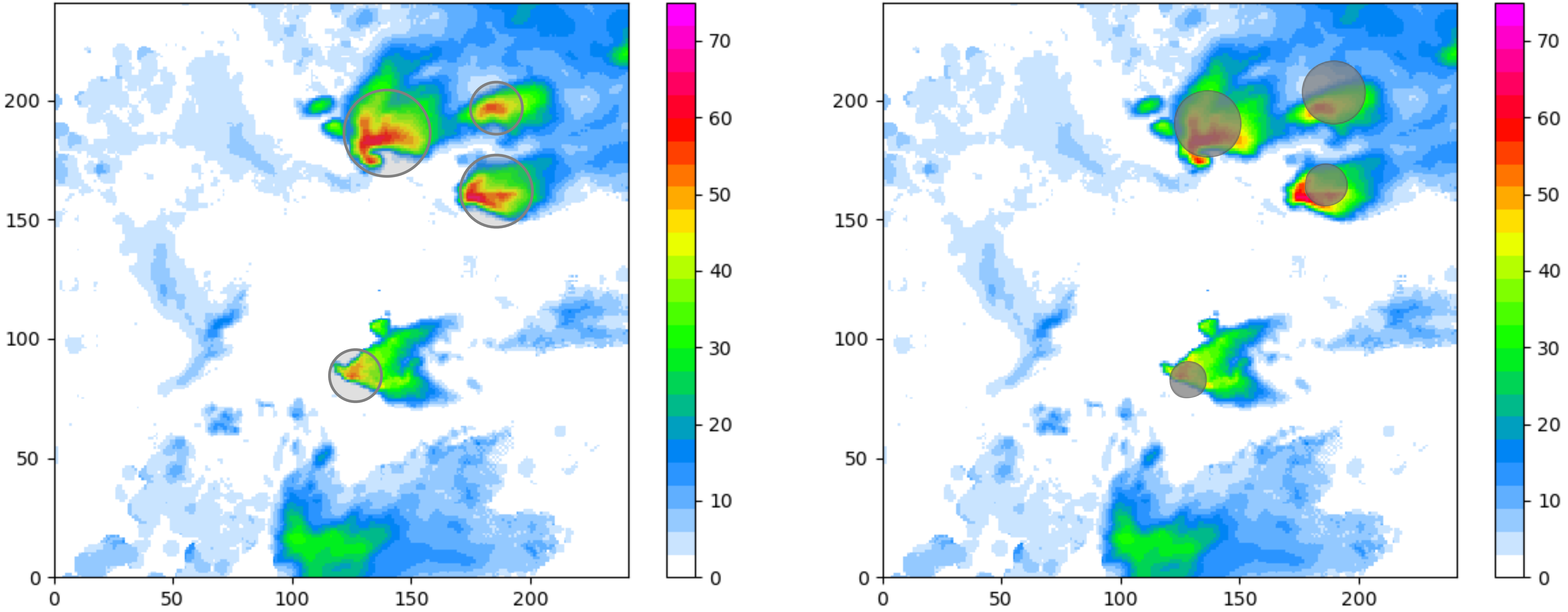
QUESTIONS



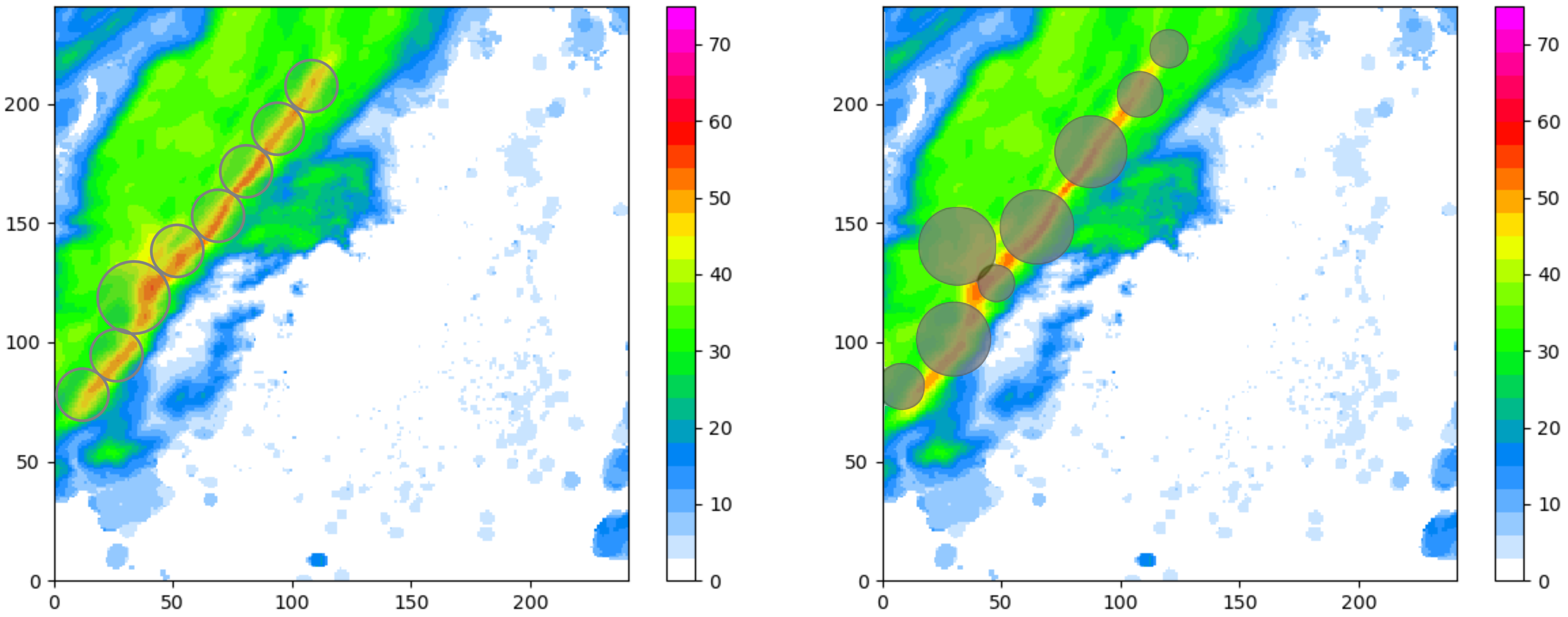
STOCHASTIC MODEL



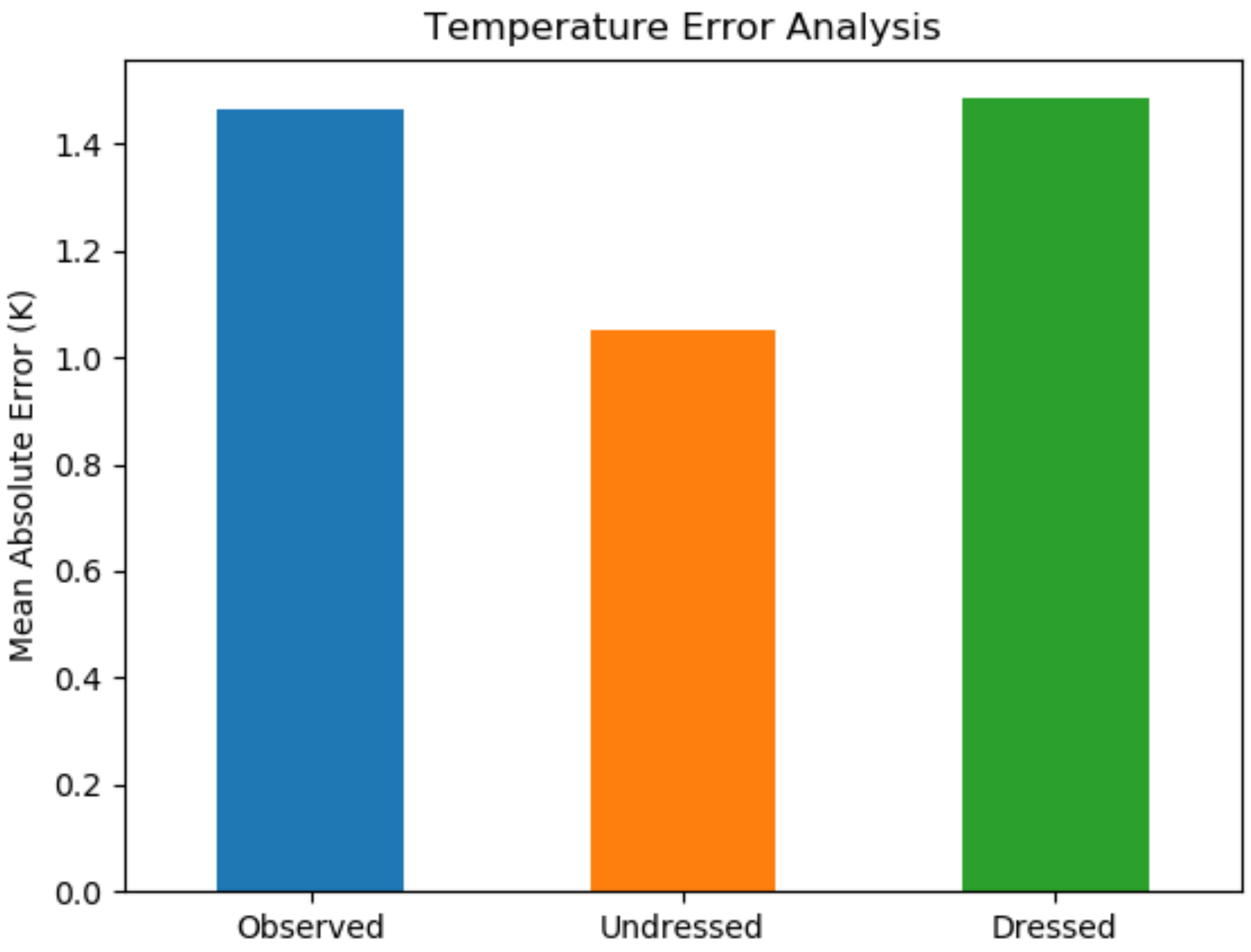
OBJECTIVE ANALYSIS (ISOLATED)

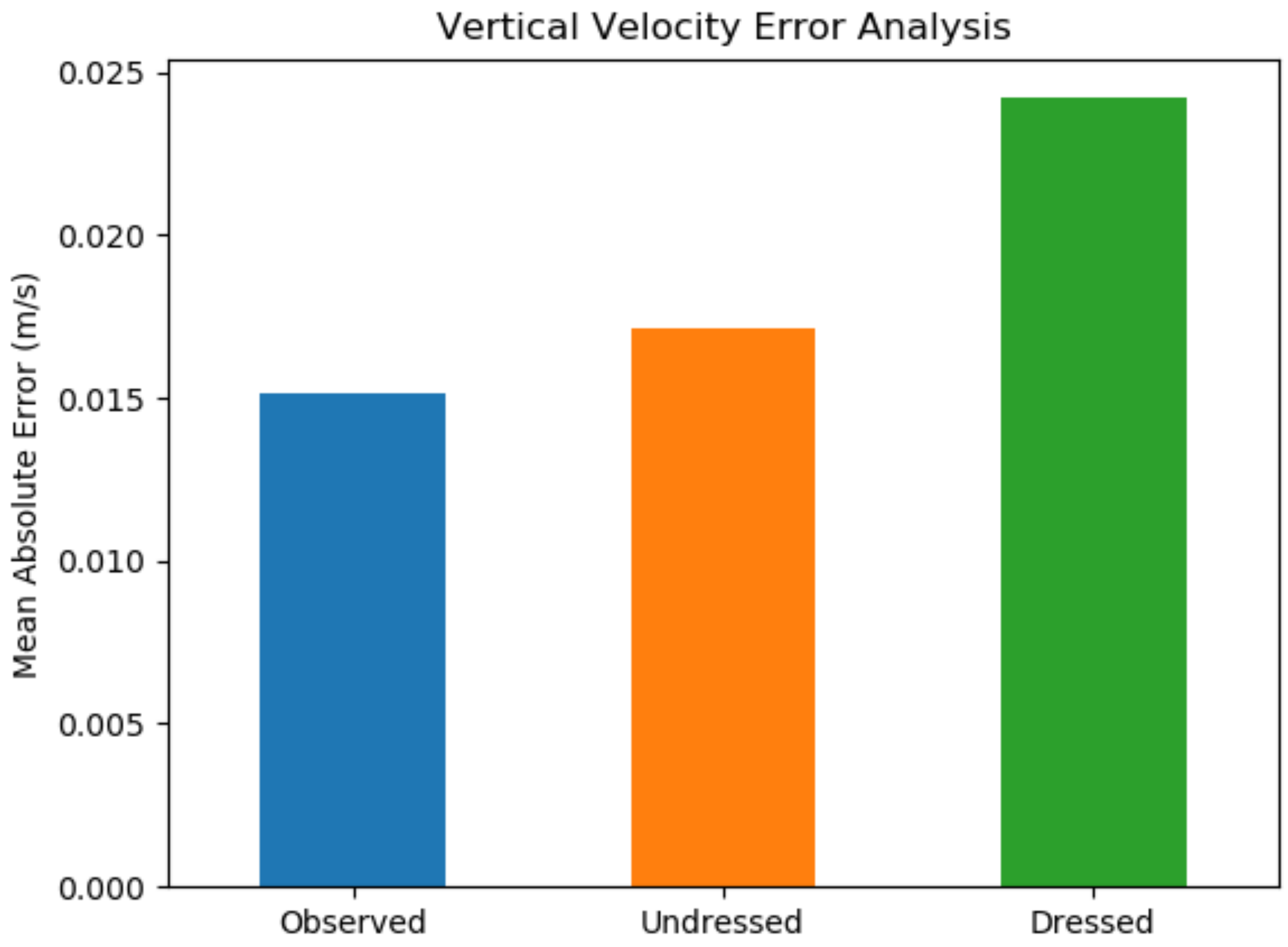


OBJECTIVE ANALYSIS (LINEAR)

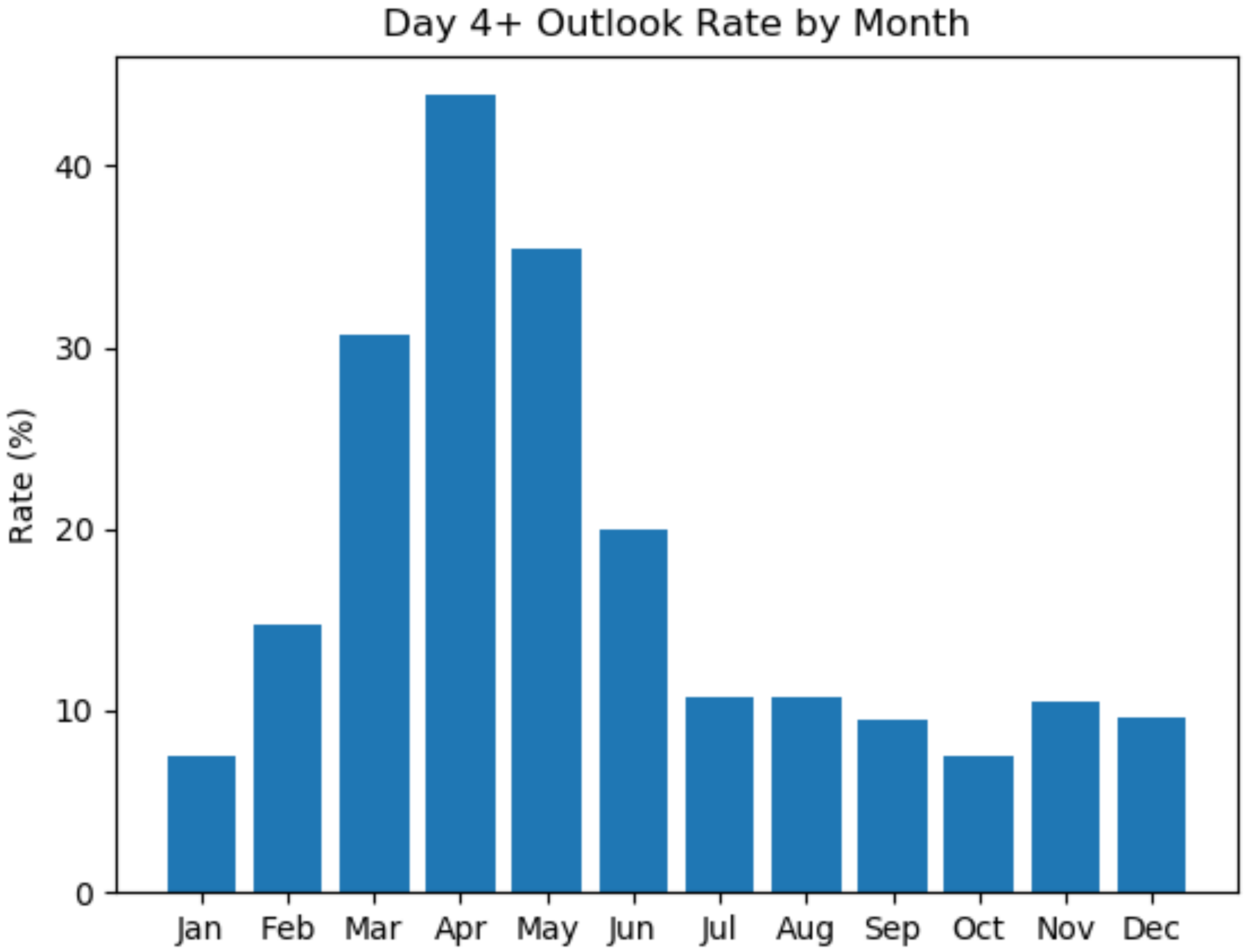


MODEL ERROR ANALYSIS

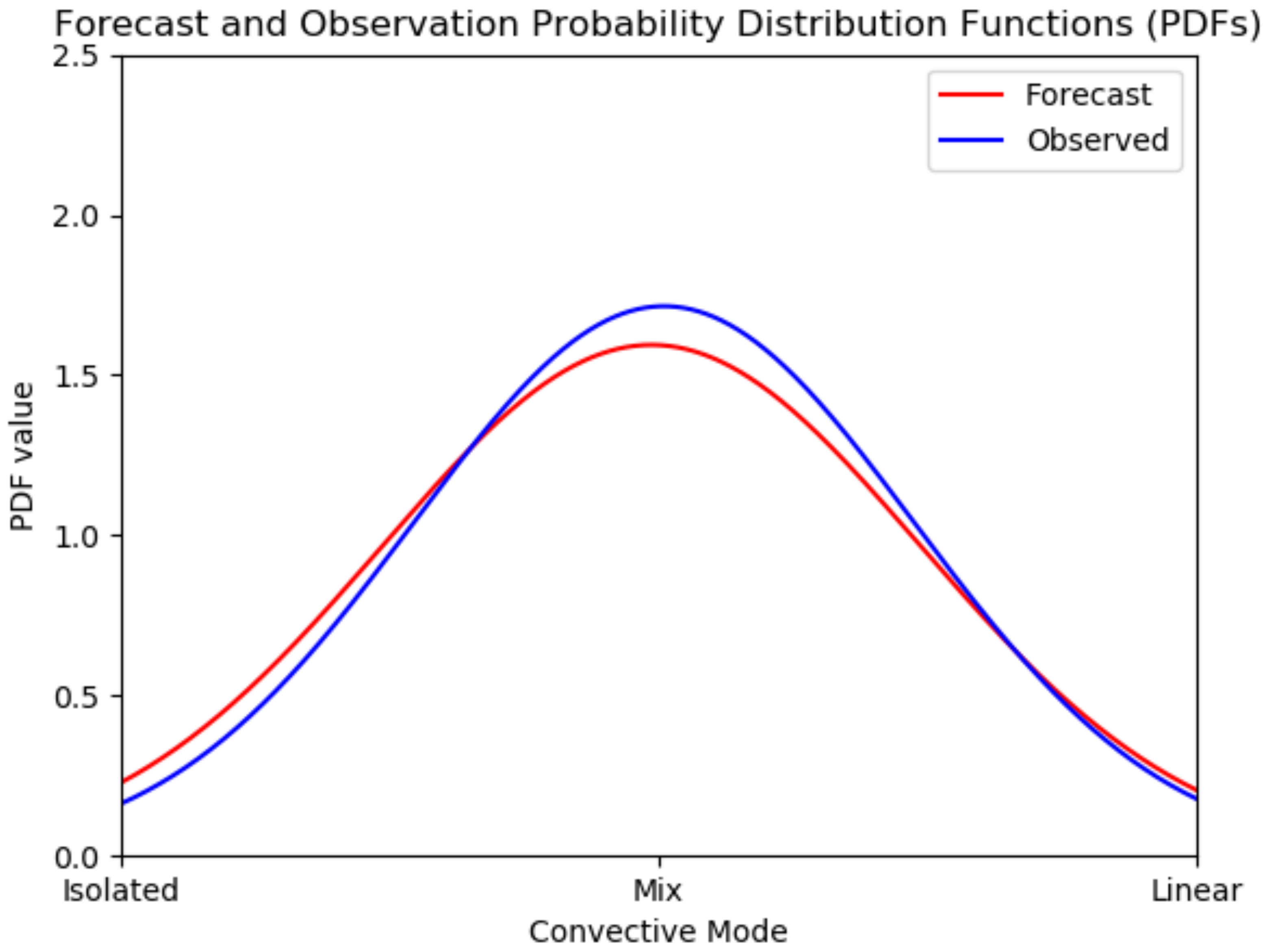




EXTENDED OUTLOOK DATA



LINEAR ERROR PROBABILITY SPACE



LINEAR ERROR PROBABILITY SPACE