



The NOAA Warn-on-Forecast program: Meeting the real-time requirements of thunderstorm-scale ensemble model prediction

Corey Potvin[1,2,3], Pamela Heinselman[2,*], Louis Wicker[2,], Jessica Choate[1,2], Patrick Skinner[1,2], Katie Wilson[1,2]**

[1]Cooperative Institute for Mesoscale Meteorological Studies, Norman, OK
[2]NOAA/OAR National Severe Storms Lab, Norman, OK
[3]School of Meteorology, University of Oklahoma, Norman, OK

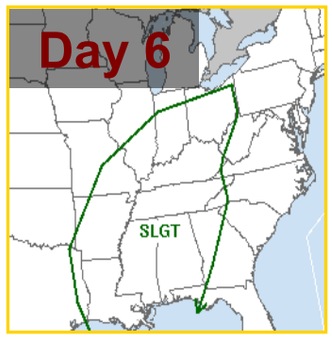
** Acting Program Manager ** Former Program Manager*

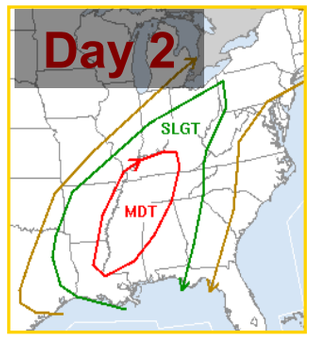
Outline

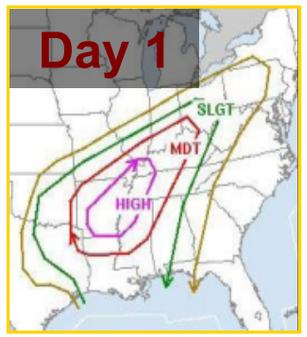
- Warn-on-Forecast (WoF) motivation and goals
- Numerical weather prediction (NWP) of thunderstorms
- WoF prototype system
- Physical science challenge: storm predictability limitations
- Computational challenge: real-time processing of ensemble and its output
- Human factors challenge: tailoring output to time-constrained forecasters

“Next-day” vs. “next-hour” severe storm prediction

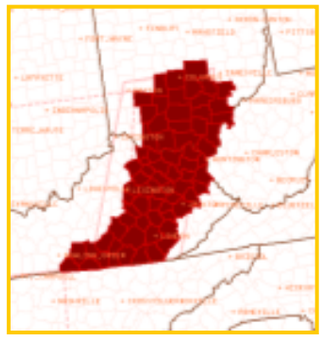
Storm Prediction Center (SPC) Outlooks



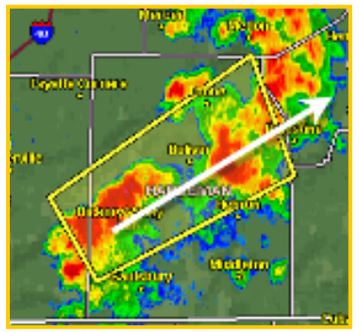




SPC Watches



WFO Warnings



Time →

12-36 hours

***Regional* storm attributes**

Where/when will storms start/end?

What storm type(s)?

What are general hail/wind/torn risks?

How widespread?

General movement?

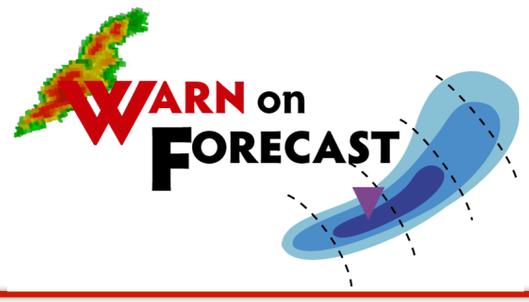
0-3 hours

***Individual* storm attributes**

Where will this storm go?

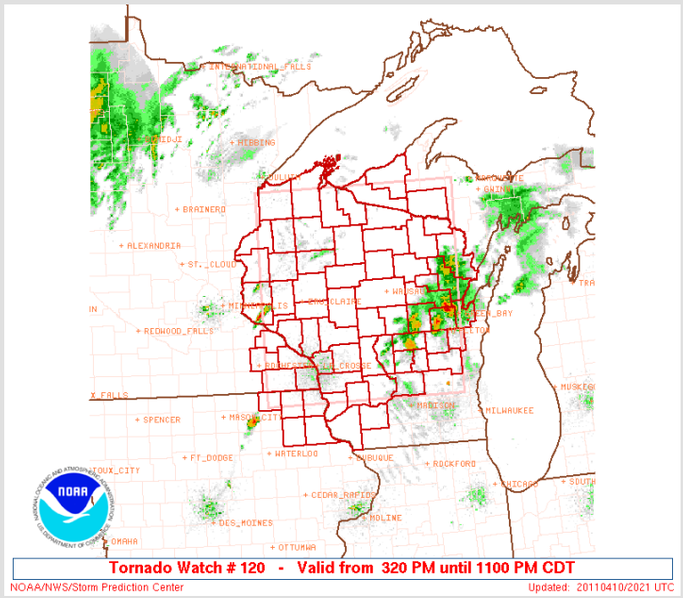
How fast will it move?

How will its risks evolve?



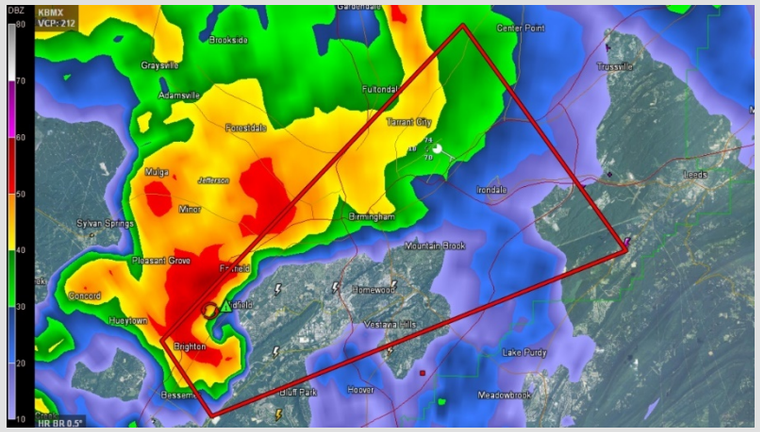
Adapted from Heather Lazrus (NCAR) and Lans Rothfusz (NSSL)

PROBLEM: There is a gap in guidance – and therefore forecast products - between the watch and warning time frames!



Watches are issued up to **8 hours** prior to severe weather occurrence

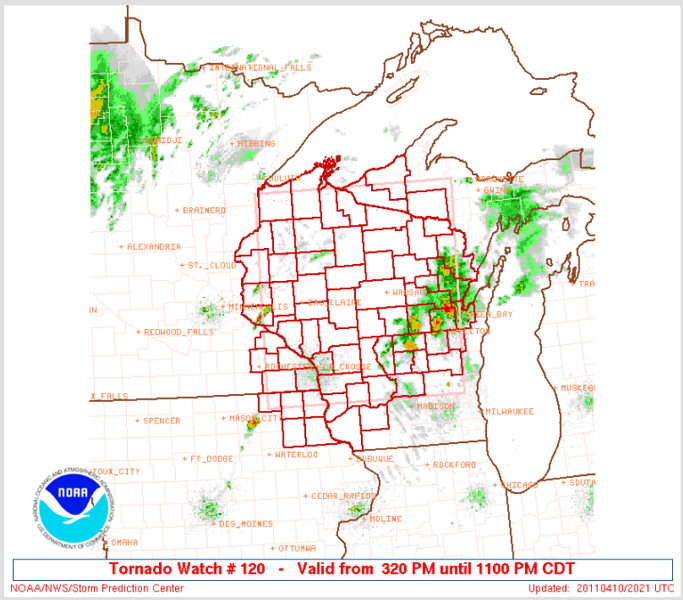
Based primarily on 3-12-h NWP model forecasts



Warnings are issued up to **30 min** prior to severe weather occurrence

Based primarily on radar observations

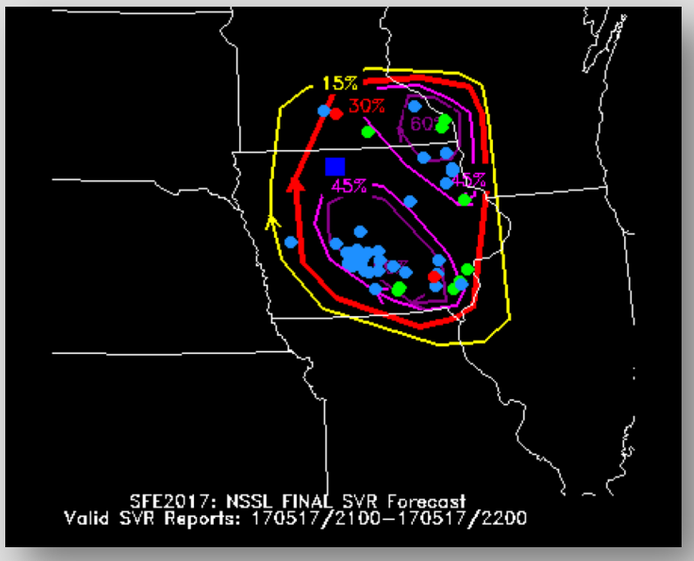
GOAL: Fill spatiotemporal gap using probabilistic guidance from Warn-on-Forecast (WoF) ensemble prediction system



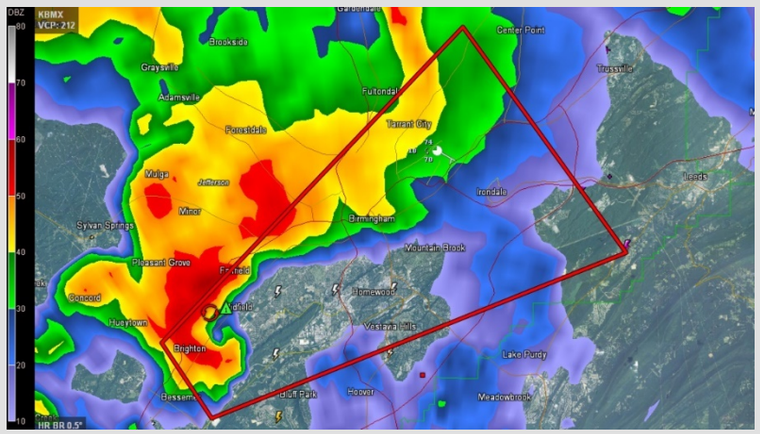
Watches are issued up to **8 hours** prior to severe weather occurrence

Based primarily on 3-12-h NWP model forecasts

1-2-hour forecast based on experimental WoF output



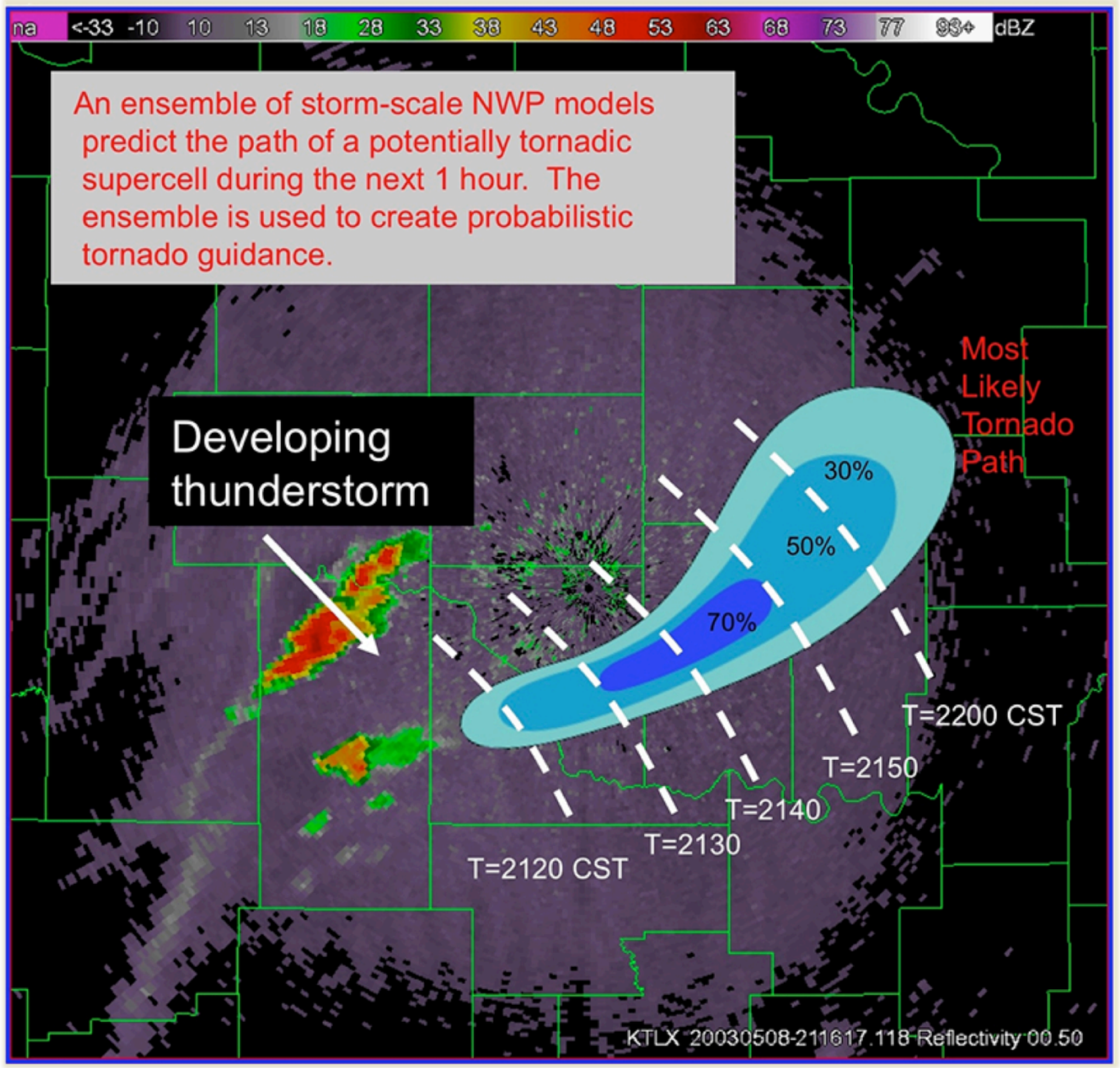
Dots = storm reports (verification)



Warnings are issued up to **30 min** prior to severe weather occurrence

Based primarily on radar observations

OUTCOME: Enable NWS to issue probabilistic warnings earlier than current (deterministic) warnings, providing advance notice to schools, hospitals, stadiums, etc.

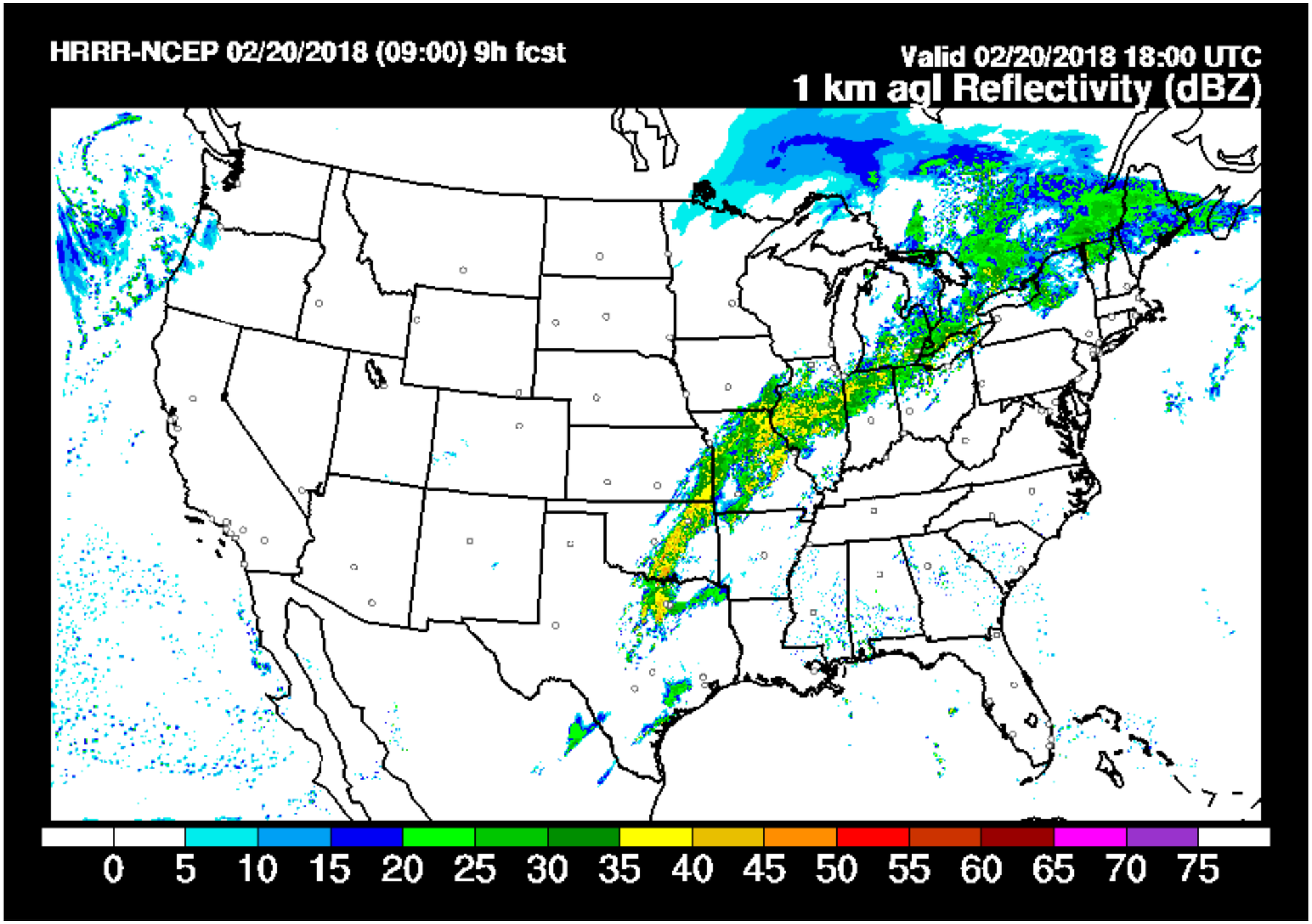


Example of a probabilistic warning (Stensrud et al. 2009)

Convection-allowing models (CAMs)

- Numerical weather prediction (NWP) models with horizontal grid spacing $\Delta x \leq 4$ km
  - Sufficient resolution to simulate/predict storms reasonably well
- Many physical processes must still be parameterized (e.g., cloud microphysics, sub-grid-scale mixing)
- State-of-the-art operational CAM: High-Resolution Rapid Refresh (HRRR)
  - Δx=3 km; run hourly

Sample HRRR output



Problem: CAM forecasts have large errors

- Storms have inherently limited predictability (i.e., perturbations grow non-linearly)
- Forecast initial conditions (ICs) suffer large uncertainty and bias
  - Only small portion of atmospheric state is observed
  - Observations have errors
- Model errors are often severe
  - Computational constraints on model resolution, numerics, and physics parameterizations
  - Limited knowledge of difficult-to-observe atmospheric processes (e.g., interactions between different types of cloud & precipitation particles)

(Imperfect) Solution: CAM ensembles

- Groups of CAM forecasts with different ICs and, optionally, different physics parameterizations and even dynamical cores (i.e., governing equations, model numerics, grids)
- Accounts for IC and model uncertainty
- Ensemble forecast ideally provides representative sampling of probability distribution function of future atmospheric state

CAM ensemble initialization

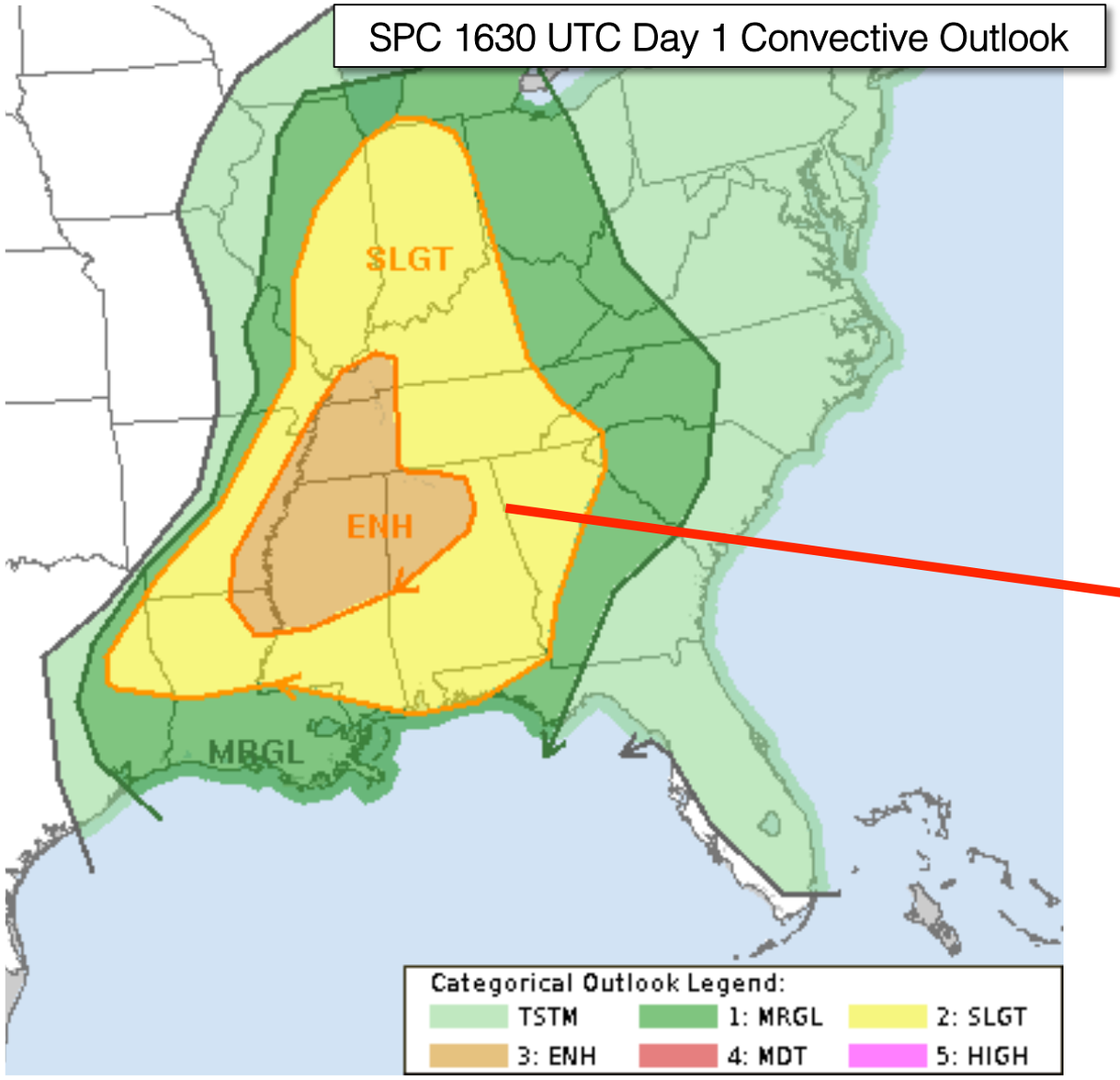
- Data assimilation (DA): optimally blend latest observations with background fields provided by model forecast(s)
  - Surface, upper-air (e.g., balloon-borne soundings), optionally radar and satellite
  - Requires background error covariance estimates (BECs) to relate observations to model state variables
- Variational DA: deterministic; constant BECs
- Ensemble Kalman filter (EnKF) DA: *flow-dependent* BECs provided by ensemble of forecasts
  - CAM DA systems use either EnKF or hybrid EnKF-variational framework

NSSL **E**xperimental **W**oF **S**ystem for **e**nsembles (**NEWS-e**)

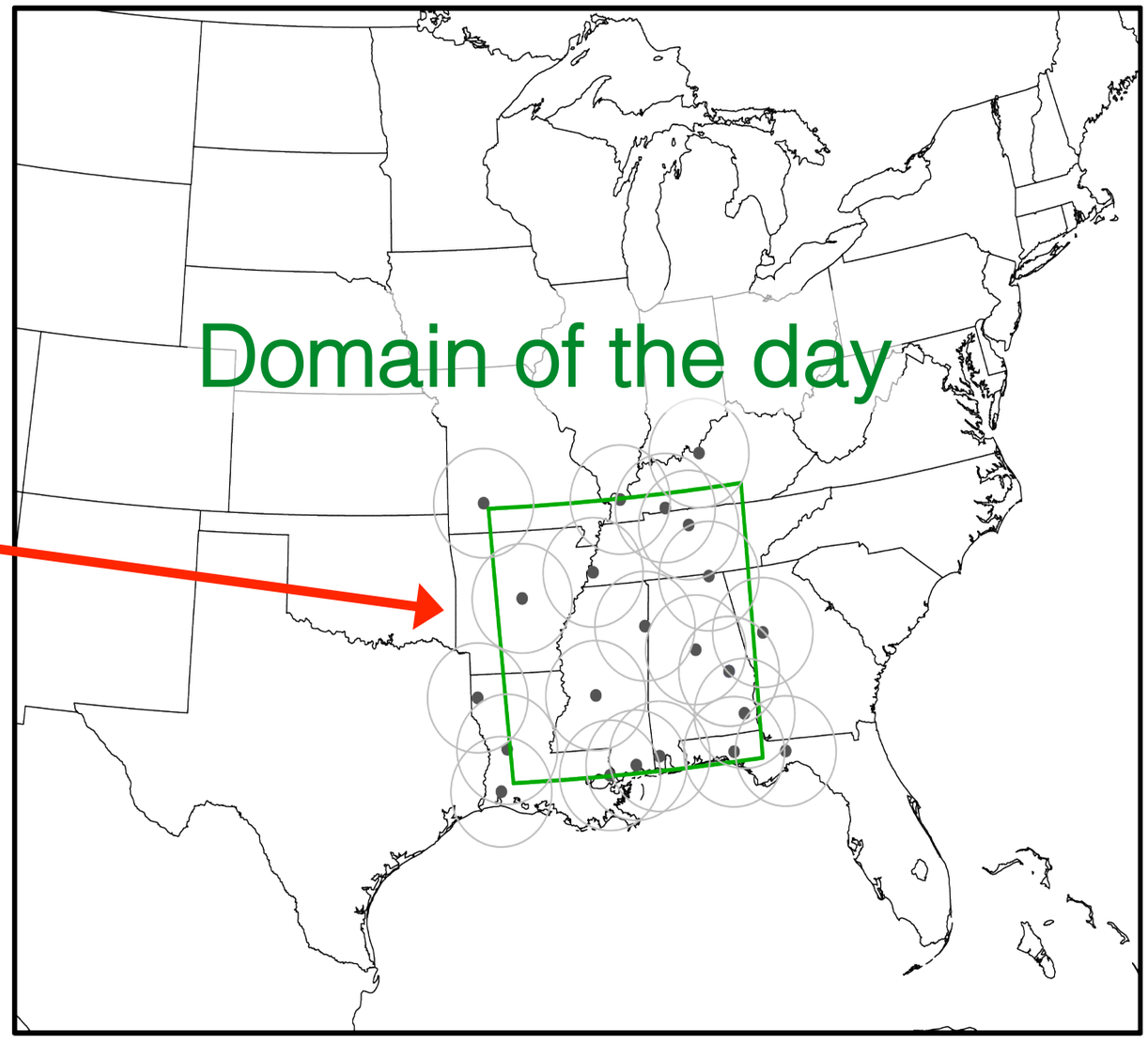
- Weather-adaptive, real-time CAM ensemble
  - Daily targets region of greatest severe weather risk
- Frequently-updating
  - DA every 15 min, forecast every 30 min
- 36 members; 18 used for forecasts
- Physics diversity
- Evaluated every spring in NOAA Hazardous Weather Testbed (HWT) by scientists and forecasters

Weather-adaptive and On-demand

based on Storm Prediction Center Day-1 Outlook



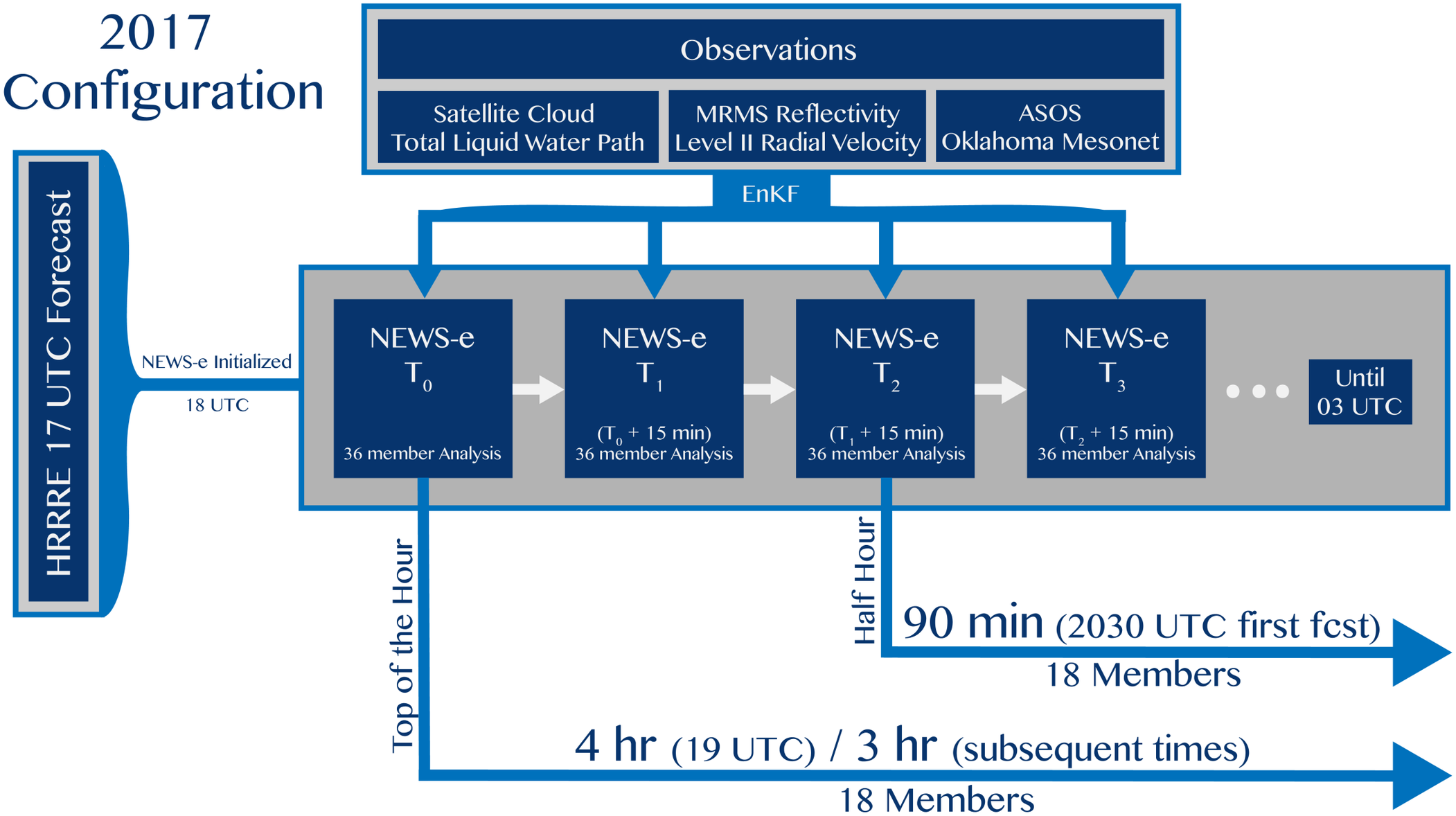
3-km HRRRE background and nested NEWS-e grid



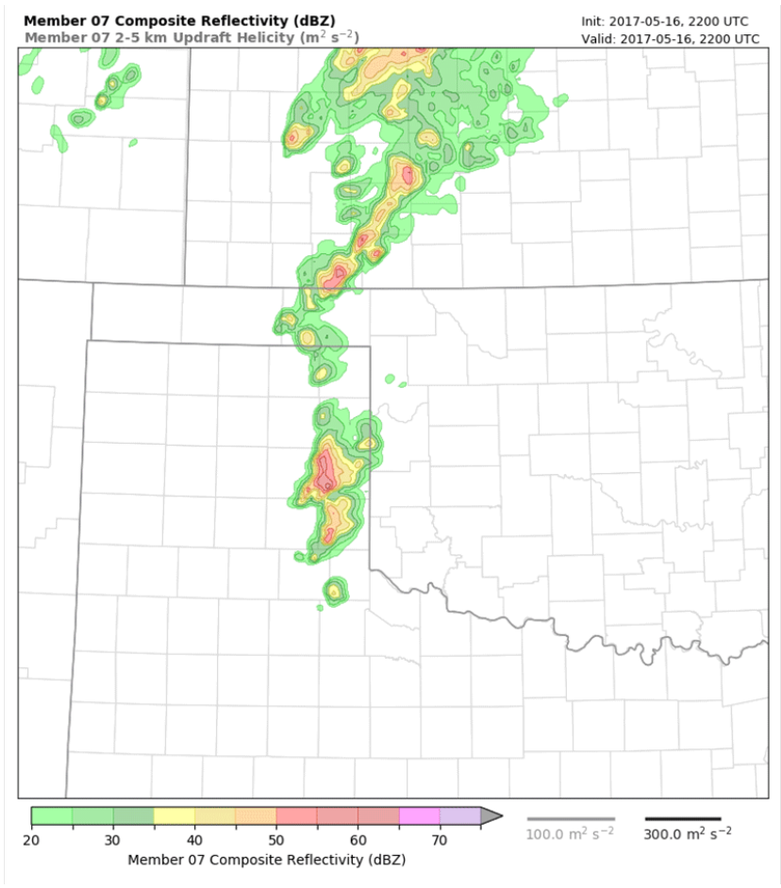
Radar locations within NEWS-e grid shown as blue dots with 150-km range rings

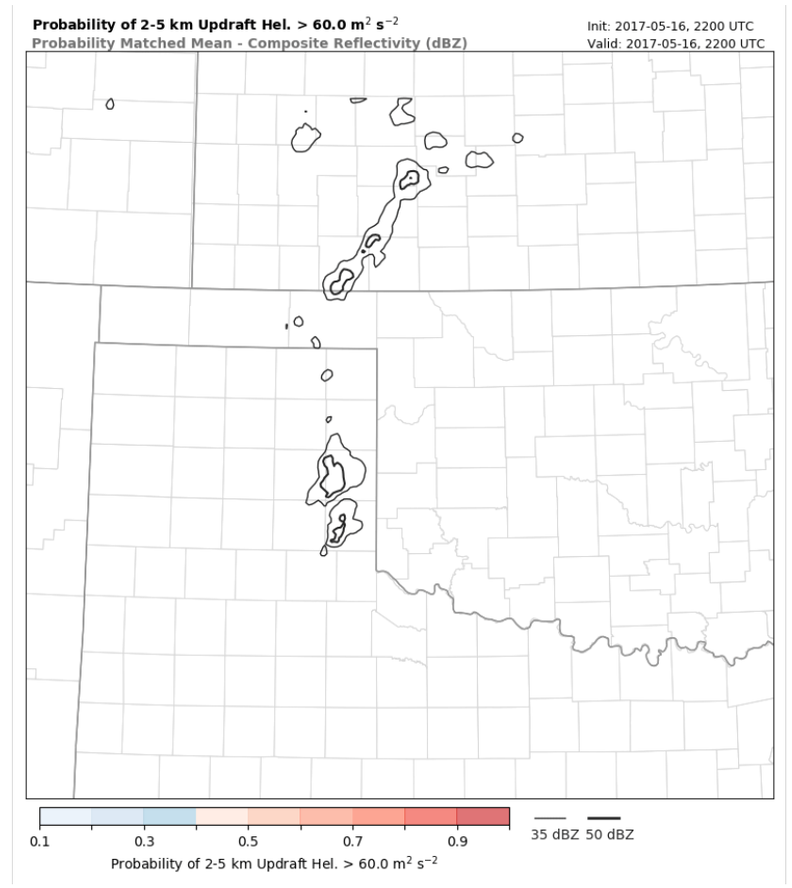
NEWS-e daily operations

2017 Configuration



Sample NEWS-e 3-h forecast





Single member
reflectivity, rotation

All members
rotation

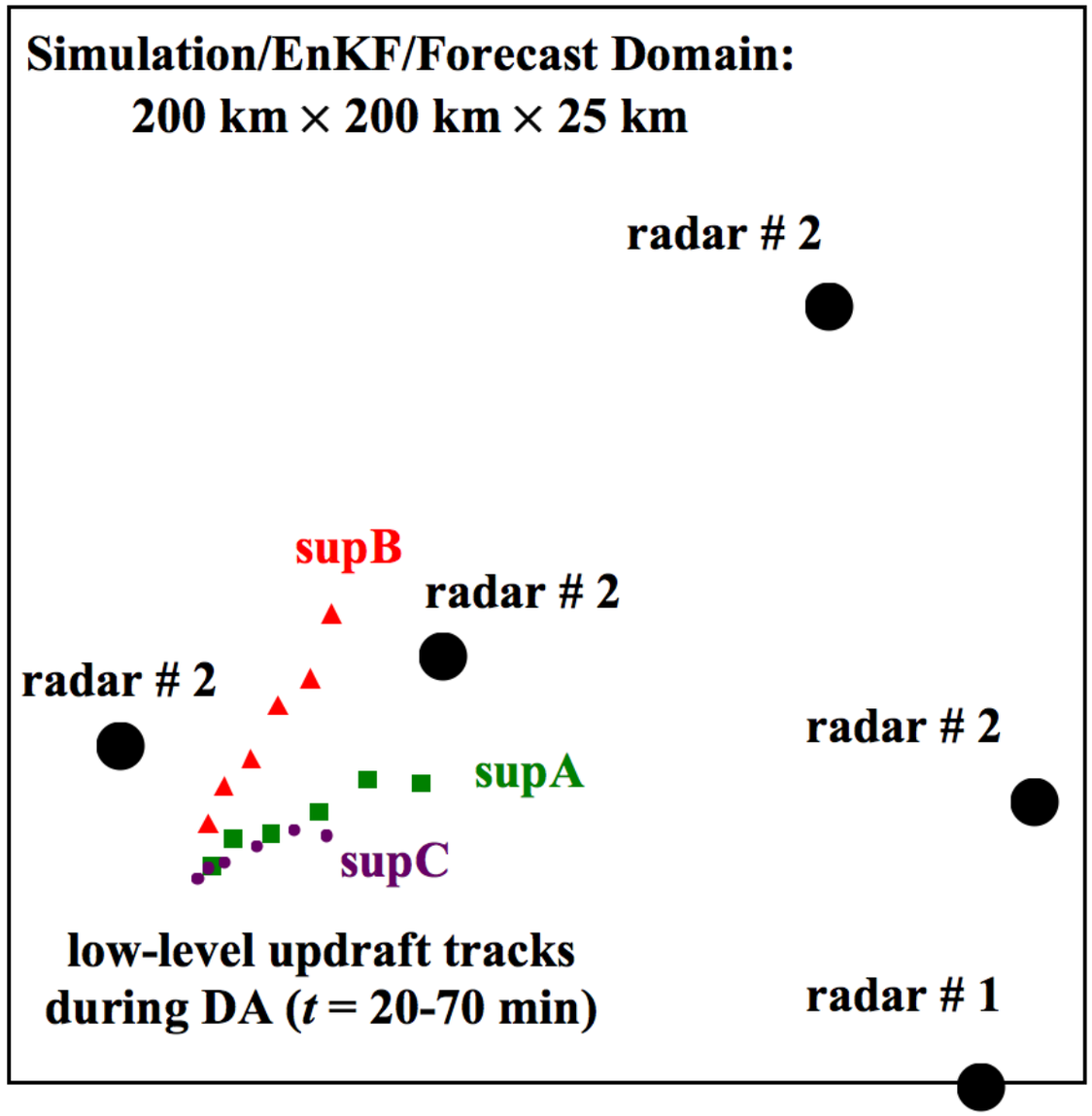
Probabilities
rotation

Exploring storm-scale predictability

- Optimizing ensemble design requires knowledge of *impacts from different forecast error sources*
  - E.g., sacrifice ensemble size for resolution?
- Construct idealized sensitivity experiments that allow isolation and systematic exploration of errors
  - But difficult to represent real-world model errors!
- Compare real-world (e.g., NEWS-e) forecasts generated using different configurations
  - But difficult to verify and to isolate impacts of individual error sources!
- Critical to perform both experiment types and hybrids thereof
- Use machine learning to identify forecast biases & sensitivities in real-world output

Sensitivity to radar-to-storm geometry
(Potvin and Wicker 2013)

- Idealized Observing System Simulation Experiments (OSSEs)
- 3 simulated supercells
- Assimilate pseudo-radar data with EnKF, then ensemble forecast
- Radar #1 > 100 km away
- Radar #2 repositioned to vary radar-storm distance, cross-beam angles (CBAs)
- Will frequently poor radar-storm geometry fundamentally limit WoF?

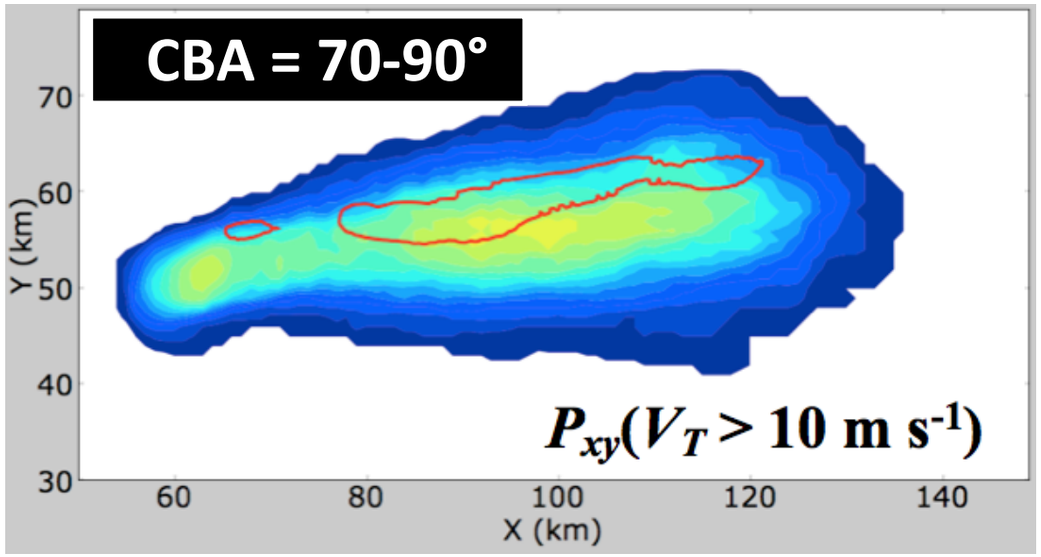


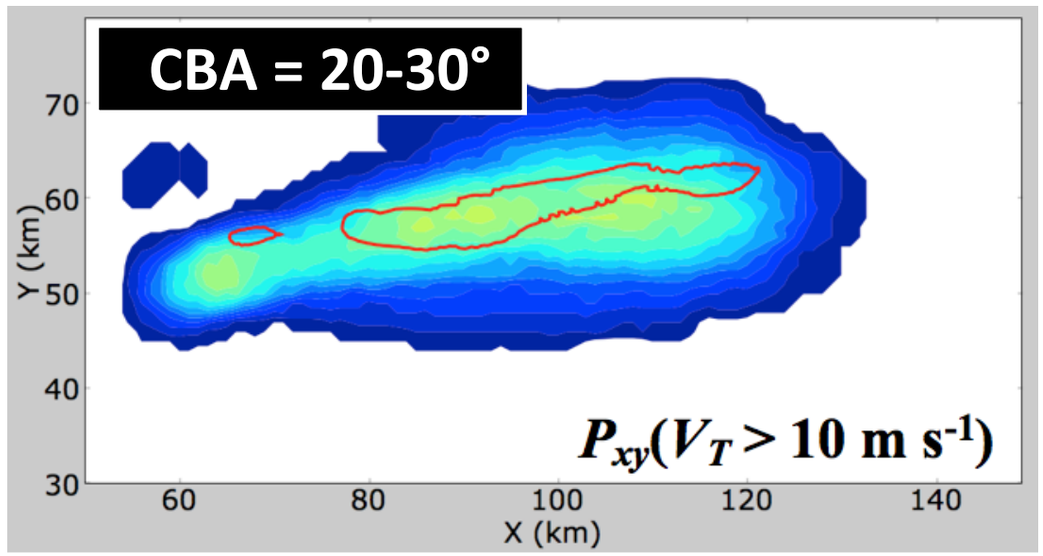
Poor radar-to-storm geometry does not unduly degrade low-level rotation forecasts

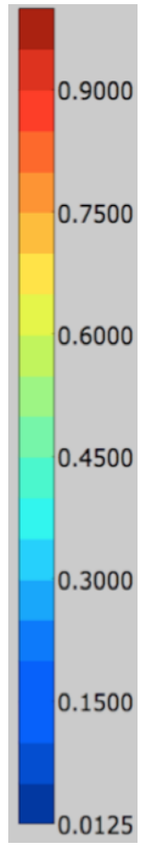
Excellent Cross-Beam Angles

Poor Cross-Beam Angles

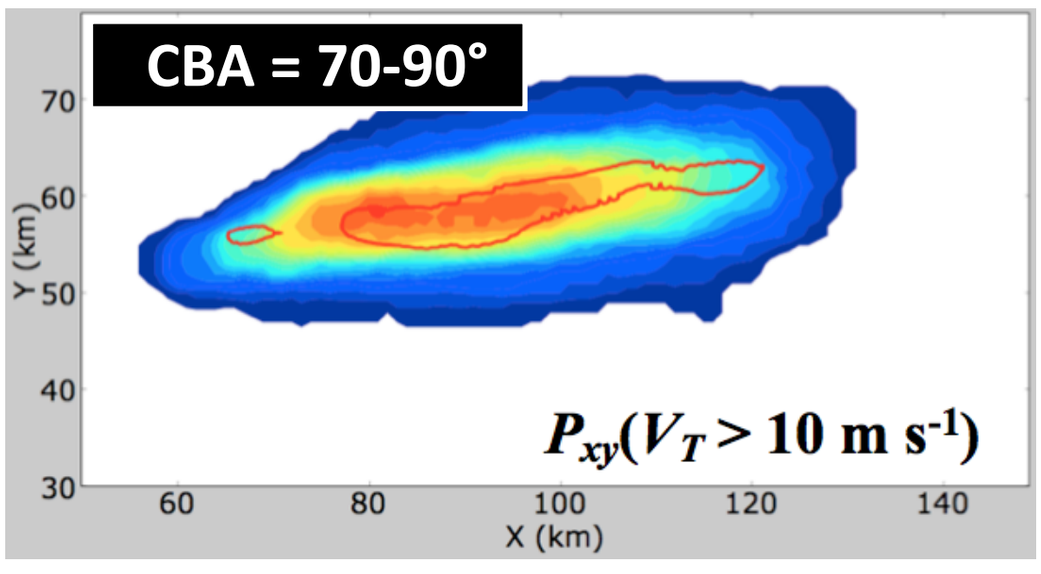
Both radars > 100 km away

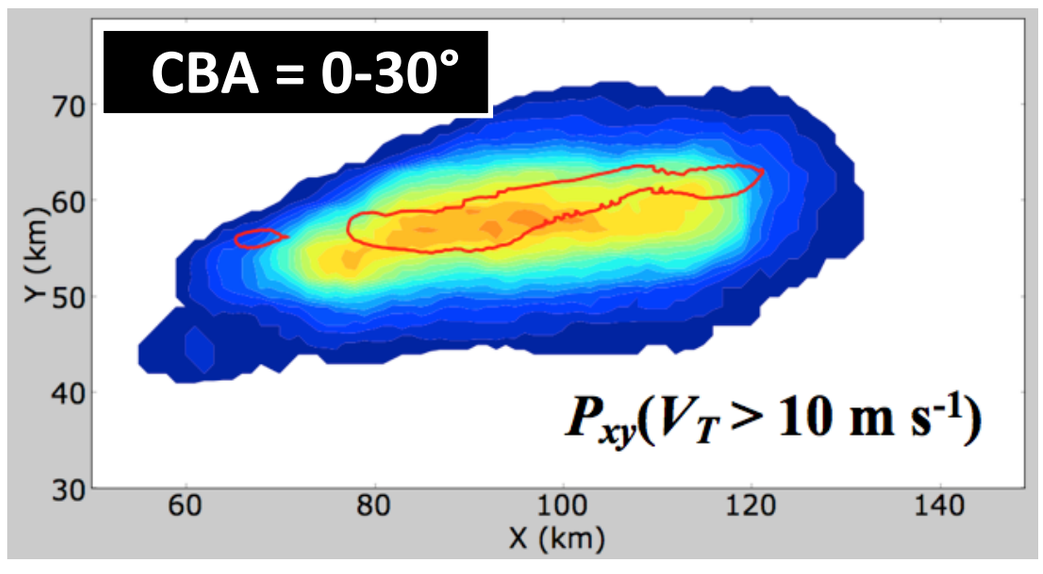






Radar #2 much closer



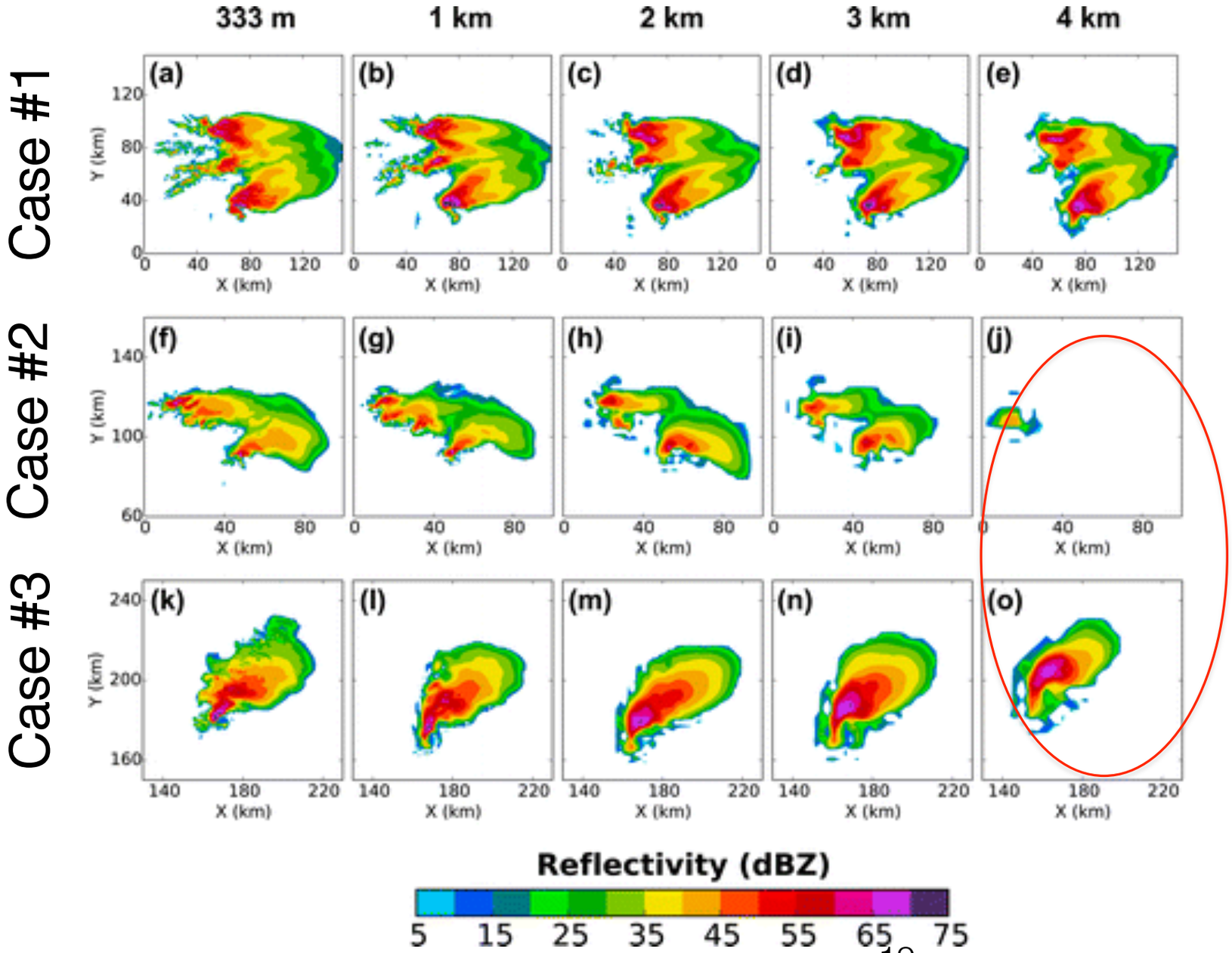


Neighborhood ensemble probability of strong low-level rotation; Red = TRUTH

Sensitivity to forecast grid Δx

Potvin and Flora (2015)

- Idealized simulations with Δx = 333 m (TRUTH), 1-4 km



Reflectivity at
***t* = 2 h, *z* = 1 km AGL**

4-km Δx too coarse: storms often develop late or die early

3-km Δx much better

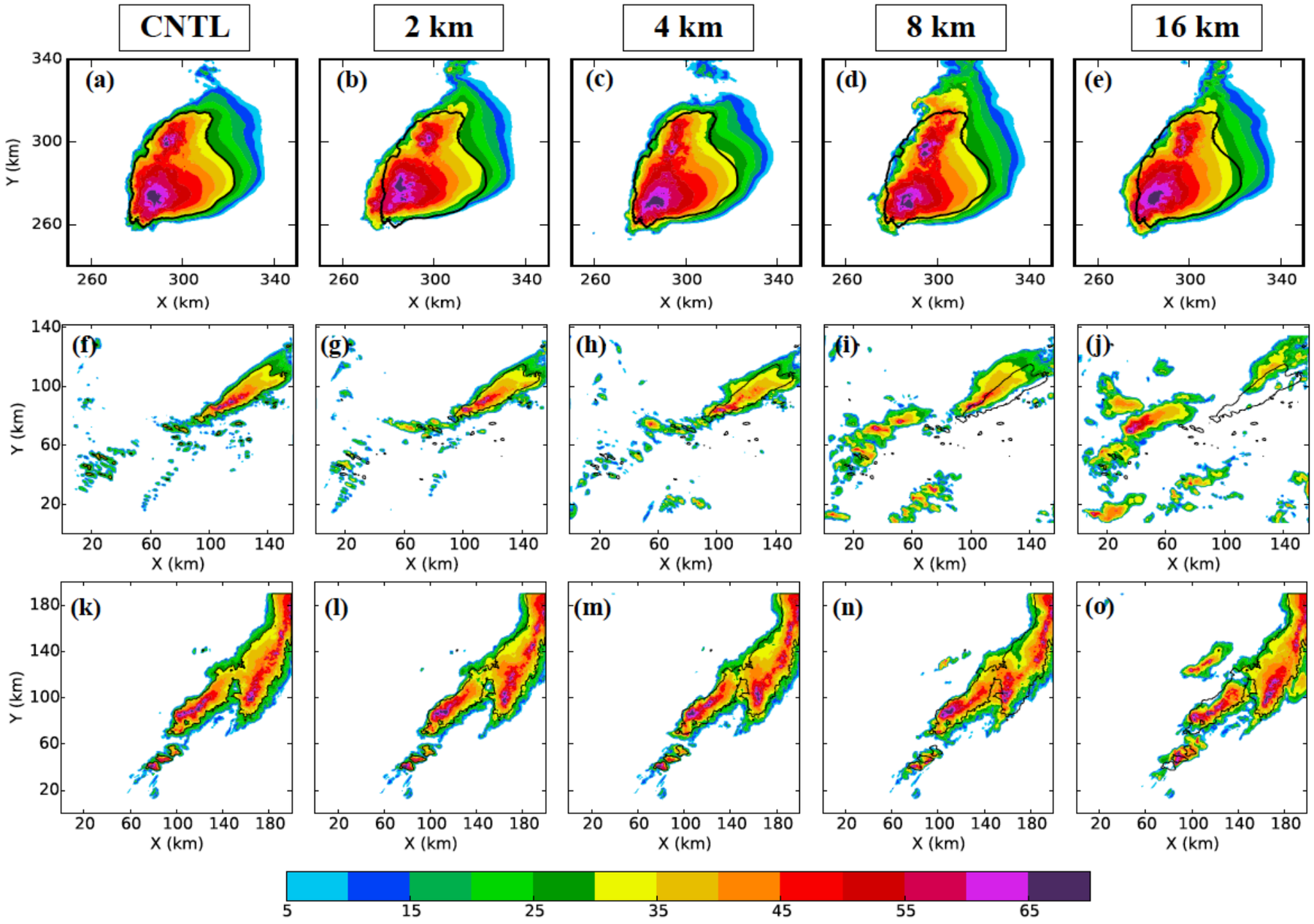
Sensitivity to IC resolution

Potvin et al. (2017)

- What are Δx and observational requirements for DA?
- Select single-member NEWS-e analysis of a *real* supercell and downscale to Δx ≈ 300 m
- Generate spatially filtered ICs
  - Cutoff wavelengths = 2, 4, 8, or 16 km
- Add noise to generate ensemble ICs
  - Critical for identifying *systematic* impacts of IC resolution
- Integrate ensembles for 2 hours
- Leverages strengths of both idealized, real-data frameworks

Surprising insensitivity to IC resolution!

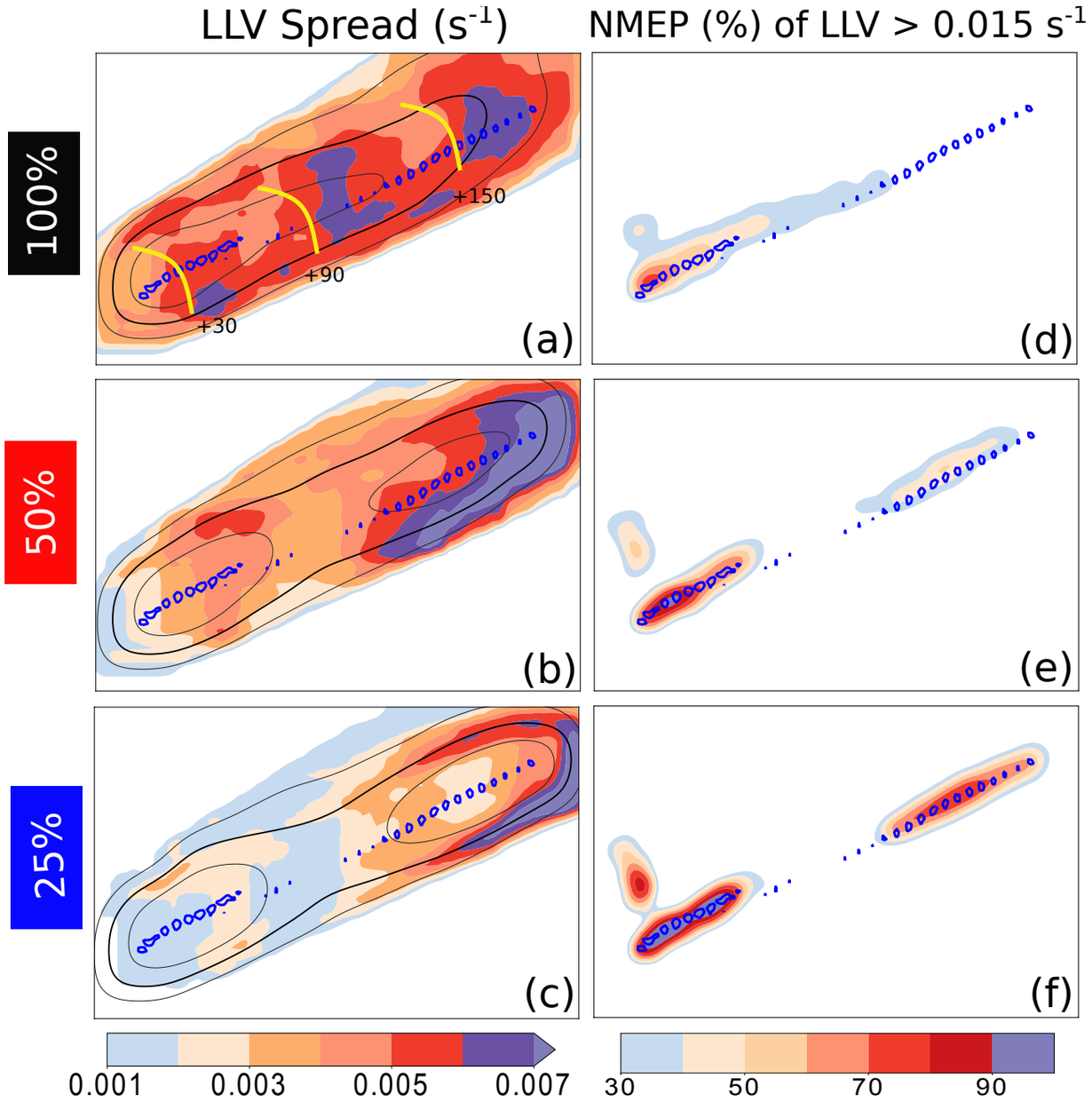
Probability-matched ensemble mean dBZ at *t* = 2 h, *z* = 2 km AGL



Sensitivity to IC spread
Flora et al. (2018)

- Assess how much forecast spread can be reduced by decreasing IC uncertainty (through, e.g., more observations or better NWP models)

- Figure: Probability of intense low-level rotation given 100%, 50%, 25% of contemporary IC uncertainty (blue=verification)



Using *machine learning* to post-process WoF output

- Facilitates evaluation of system performance (next slide)
- Provides an *automated* and (assuming a suitable training dataset) *reliable* way to correct model biases in ensemble statistics

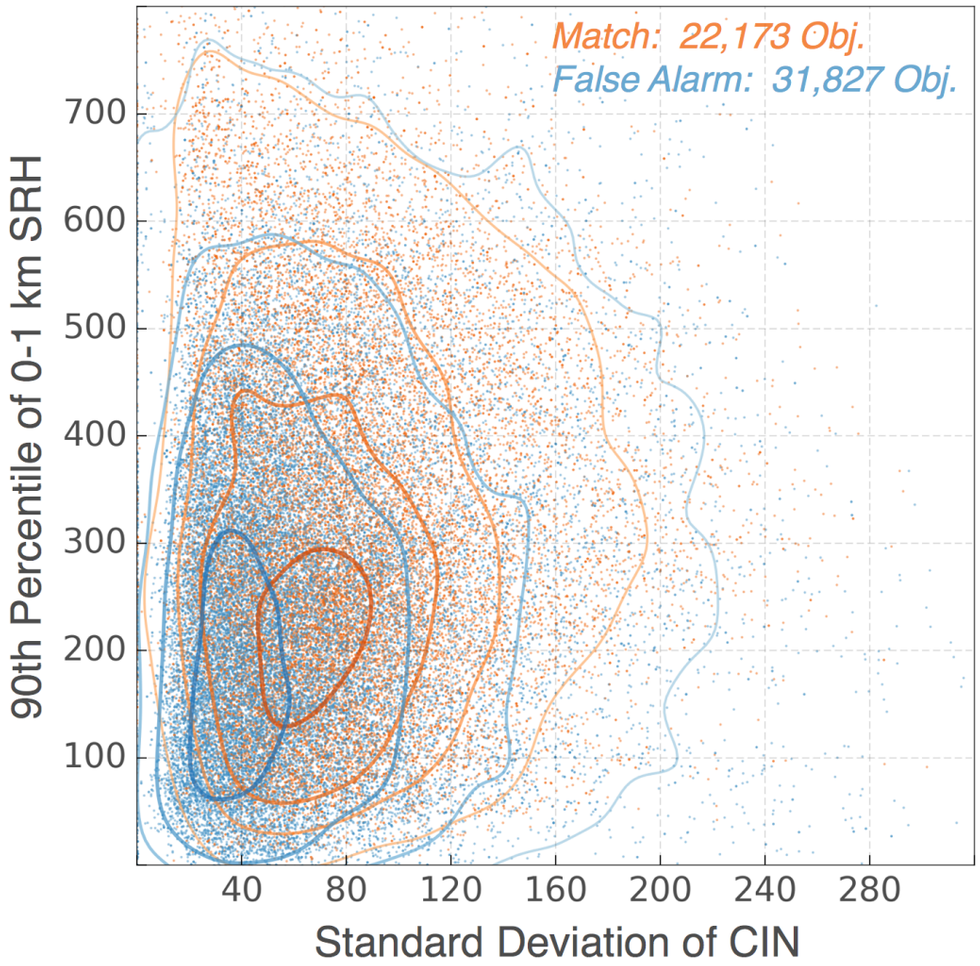
Feature Importance for NEWS-e: Matches vs. False Alarms

Matches, false alarms determined using rotation track objects and scales typical of NWS warnings

Cross Validation Score: 0.67

1) Area, 2) CIN, 3) 0-1 SRH

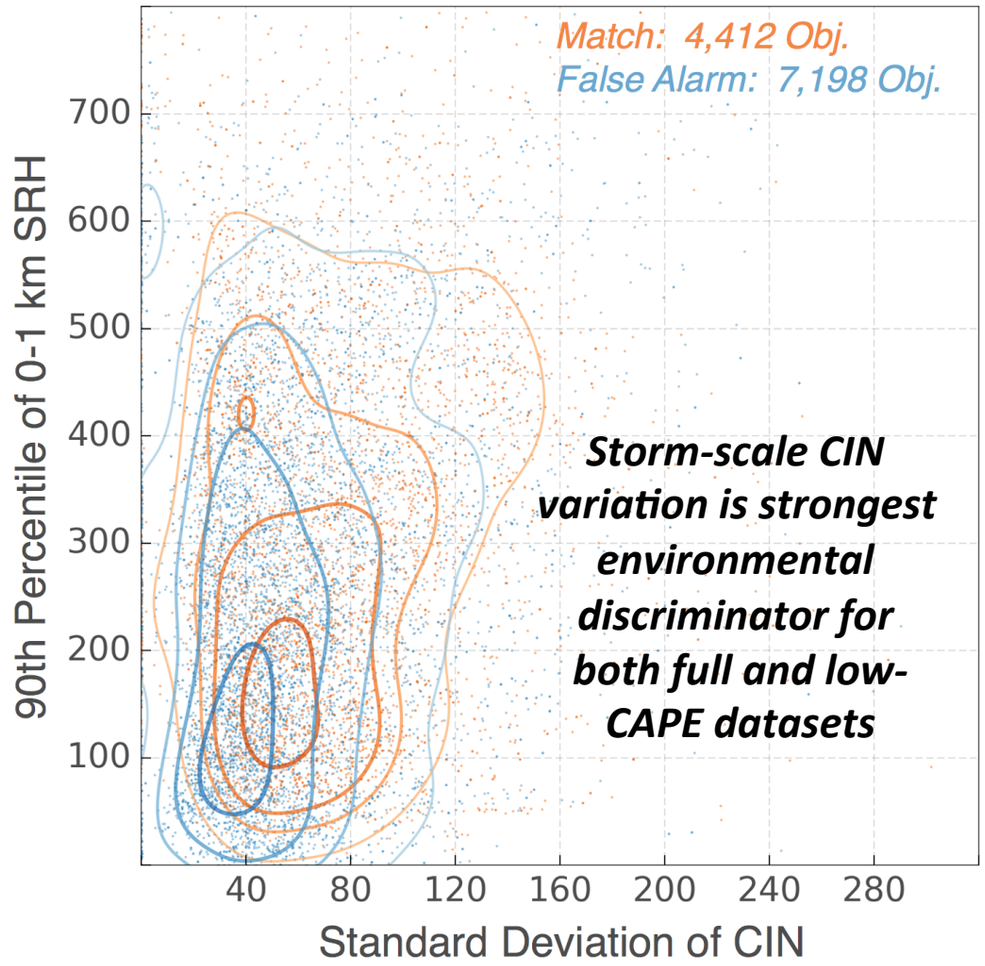
All 60-Minute Forecast Objects



Cross Validation Score: 0.74

1) Area, 2) CIN, 3) 0-1 SRH

'Low-CAPE' 60-Minute Forecast Objects

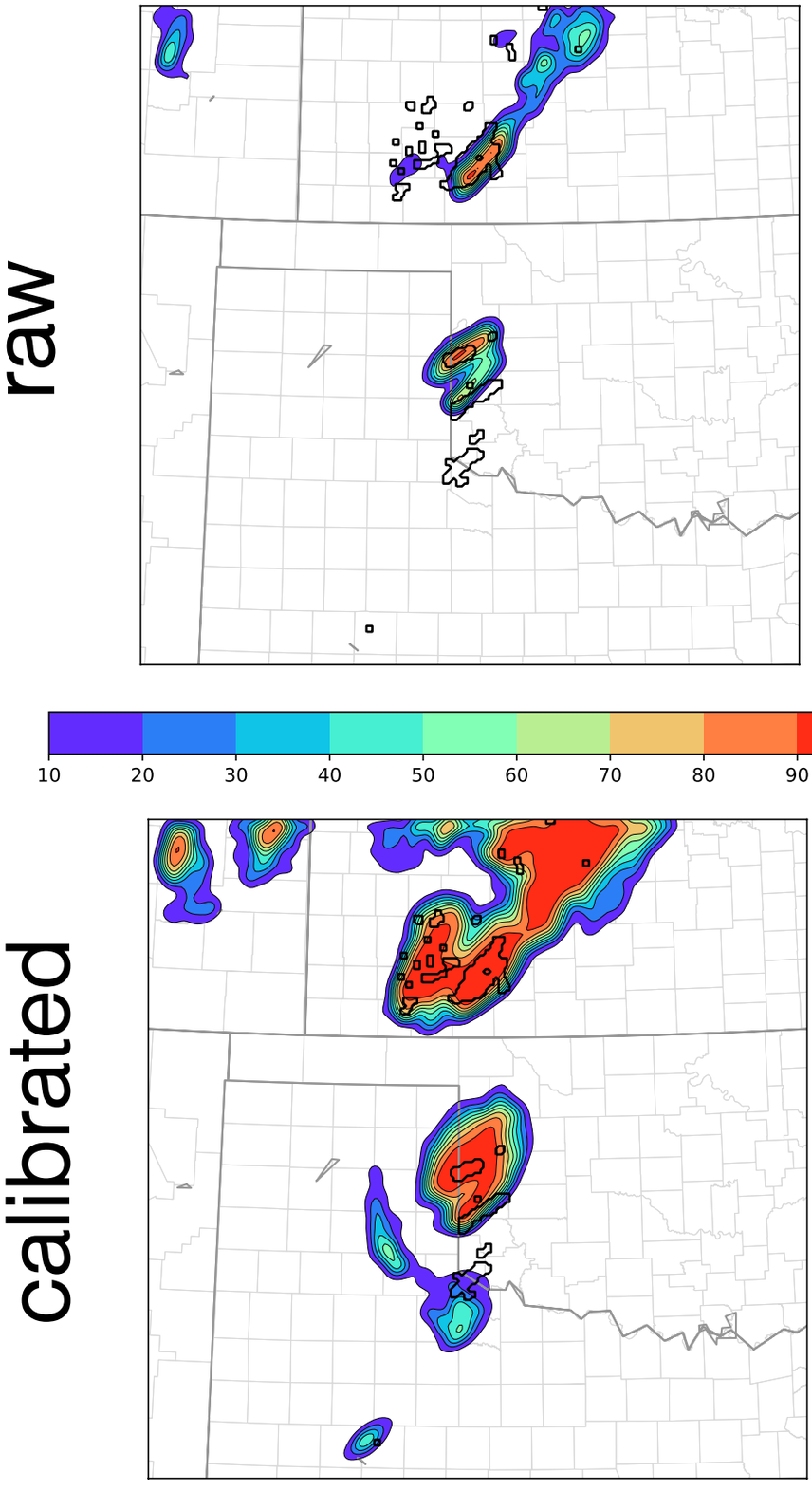


Calibrating WoF output

- Apply machine learning model to ensemble output to correct forecast biases

- Figure: probabilistic forecasts of strong low-level rotation from (top) raw NEWS-e output and (bottom) random forest trained on many NEWS-e cases and applied to this case. **Black contours = verification**.

raw



calibrated

Real-time computational challenges

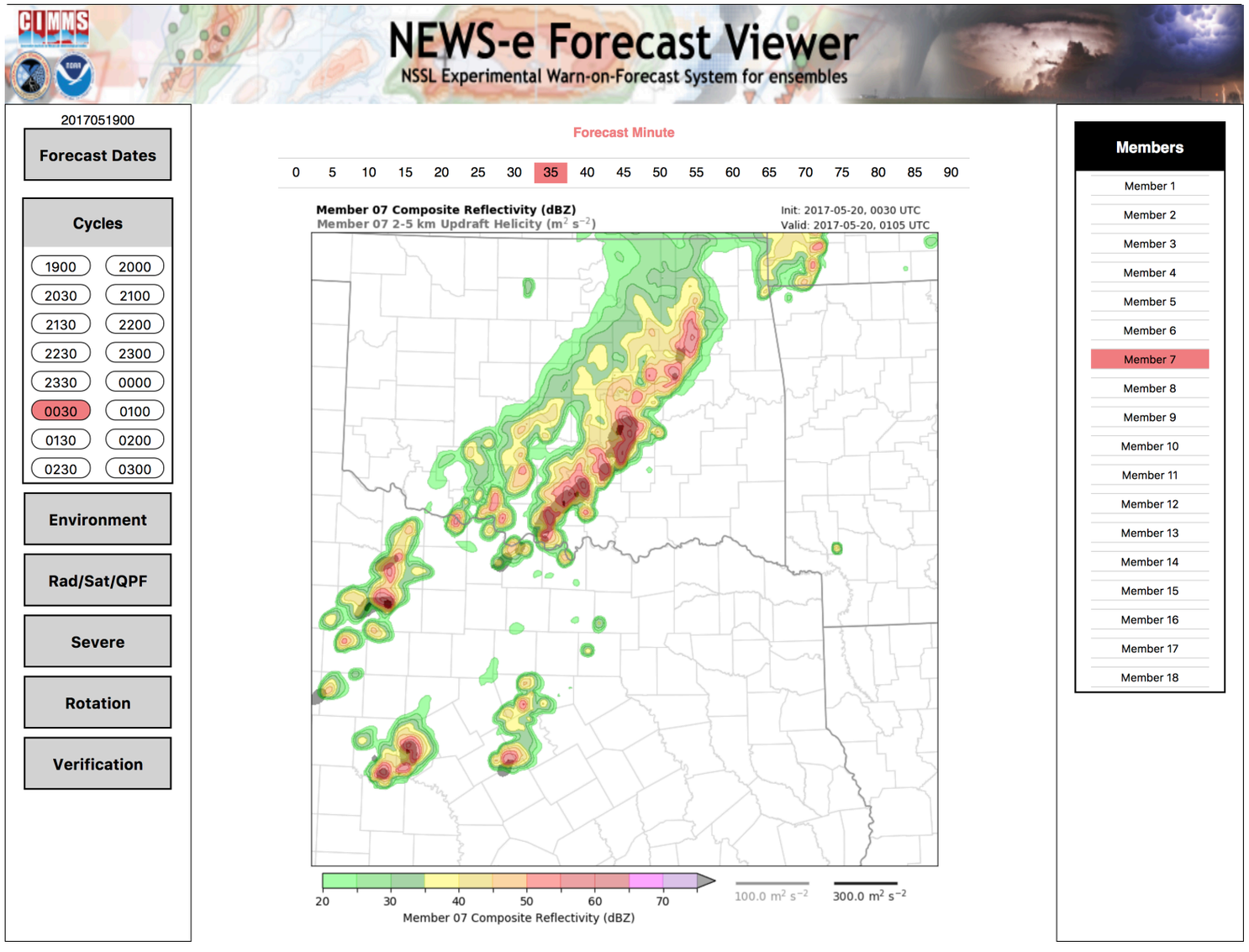
- Every 15 min, 36 members updated using O(10,000) obs then integrated to next cycle
- Every 30 min, 18 members with 3.5 million grid points integrated 1.5 or 3 h (18 or 36 output times)
- Each forecast must be post-processed into >100 products and uploaded to website within 30 min of initialization
  - 7500 images generated per forecast
  - Products should convey useful ensemble information for the entire forecast period to forecasters in < 30 s
- Total ensemble output = 1.2 TB/day
- I/O is major bottleneck – suggestions??

Real-time computation

- Cray with >4000 Ivy Bridge cores; 2880 used by NEWS-e
- Parallelization
  - Ensemble member forecasts run simultaneously
  - MPI (domain decomposition)
  - Lustre parallel file system
- Storage reduction: netCDF4 compression, lossy compression (post hoc)

NEWS-e website

- Developed based on feedback from NWS forecasters, HWT participants
- Being in same building as Norman WFO *very* helpful
- Being revamped by computer scientist
- Video tutorials, popup help images/ notes in development



NEWS-e website will provide real-time guidance during the spring of 2018 and all cases from 2016 and 2017 are available for viewing:

www.nssl.noaa.gov/projects/wof/news-e/images.php

Questions: *patrick.skinner@noaa.gov*

NOAA Hazardous Weather Testbed (HWT)

https://hwt.nssl.noaa.gov/

- A space to bring the research and operational communities together
- Important for achieving the FACETs vision (Forecasting a Continuum of Environmental Threats)
- Research consisting of exploratory and experimental work

Interdisciplinary Research

- Simulated Real-Time Experiments
- Performance Analysis
- Cognitive Task Analysis
- Eye Tracking
- Confidence and Workload Assessments
- Focus Groups
- Integrated Warning Teams
- Surveys, Evaluations, and Observations

NEWS-e testing in HWT Spring Forecast Experiment

Multi-method approach provides holistic assessment of forecasters' use of data!

1) Pseudo-Operational Testing

Retrospective evaluation of NEWS-e output and derived experimental forecasts

2) Real-time Operational Testing

Working with forecasters to use NEWS-e output in forecast/warning decisions

3) Survey Work

Gain understanding of meteorologists' interpretations of probability concepts used in WoF products

1) Pseudo-Operational Testing

- Explored how short-term ensemble forecast guidance from NEWS-e could be used by an expert forecaster and SFE participants to produce 1-hour severe weather probabilistic outlooks
- Observed how the forecaster's understanding, use, and attitudes about NEWS-e guidance evolved during the experiment
  - Screencasts, observer notes, and forecaster questionnaires

Daily participant survey

Daily Wrap-up

Date: 5/9/2017

What type of event would you classify today's event as? (Check one)

- ⊘ Cellular
- ○ Linear
- ○ Mixed

What severe weather threats do you believe are most likely and during what hour(s)?

Hail Both periods

What type of forecast would you classify today's event as? (Check one)

- ○ Low probability/high consequence
- ○ Low probability/low consequence
- ○ High probability/high consequence
- ⊘ High probability/low consequence

How difficult was today's forecast? (Circle one)

0 1 2 (3) 4 5 6 7 8 9 10

Not at all difficult *Normal difficulty* *Extremely difficult*

How useful was the NEWS-e forecast?

0 1 2 3 4 5 (6) 7 8 9 10

Not at all useful *Normal usefulness* *Extremely useful*

Provided good info for a cell in NM / NOT SURE about development far WSW panhandle

How much trust do you hold in today's NEWS-e products?

0 1 2 3 4 (5) 6 7 8 9 10

No trust *Normal amount of trust* *Large amount of trust*

How confident are you in today's outlooks?

0 1 2 3 4 5 6 7 (8) 9 10

No confidence *Normal amount of confidence* *Extremely confident*

Is your confidence level affected more by the forecast difficulty or trust in the NEWS-e products? Why?

Confidence is based on satellite trends / atmospheric conditions

What products were most helpful in today's forecast? Why?

Composite Reflectivity ENSemble

Which product group did you prefer during your forecast? Check all that apply.

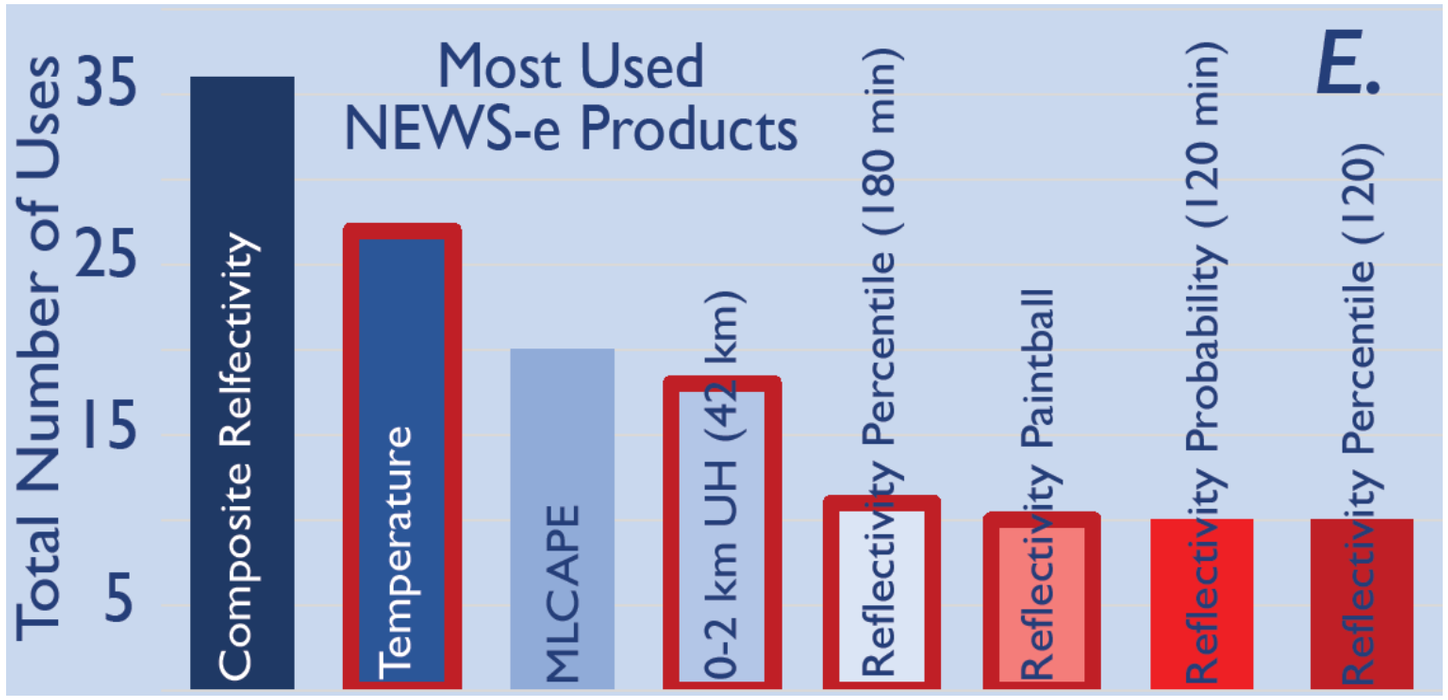
- ○ Ensemble forecast products
- ⊘ Ensemble distribution products – SEE all of the model iterations – I really like No Viewer configuration

Looking out to the rest of the afternoon, what are your major concerns with today's forecasts?

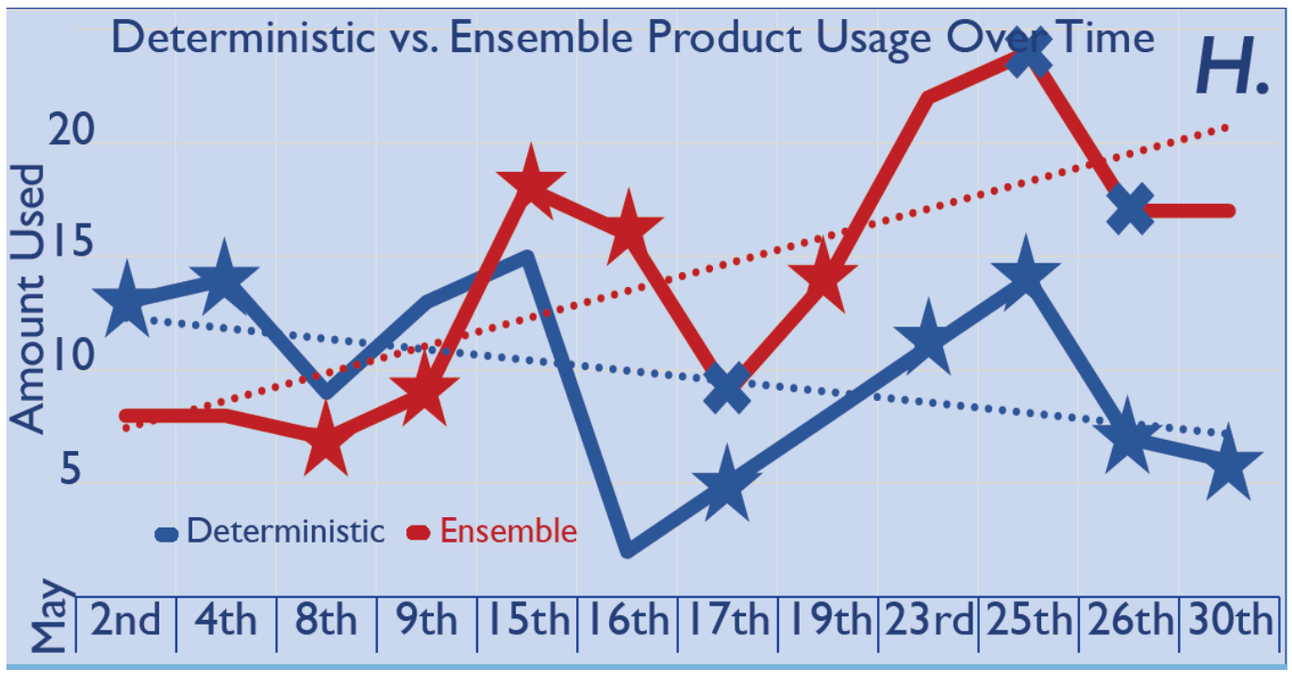
STORMS with very large hail & TORNADOES

Questions or comments?

THANKS for helping to make the Viewer easier to use.



Expert forecaster's product usage as determined by screen recordings



Evolution of expert forecaster's use of deterministic vs. ensemble guidance

Expert Forecaster Comments:

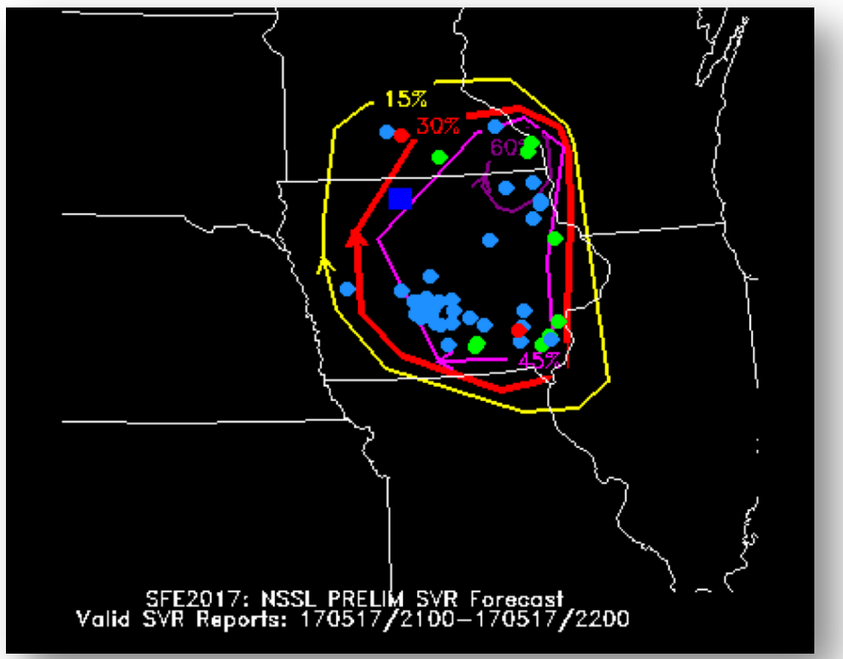
"On most days, NEWS-E was exceptional at identifying which storms had the greatest potential to become severe."

"NEWS-e is a big help in forecasting storm evolution!"

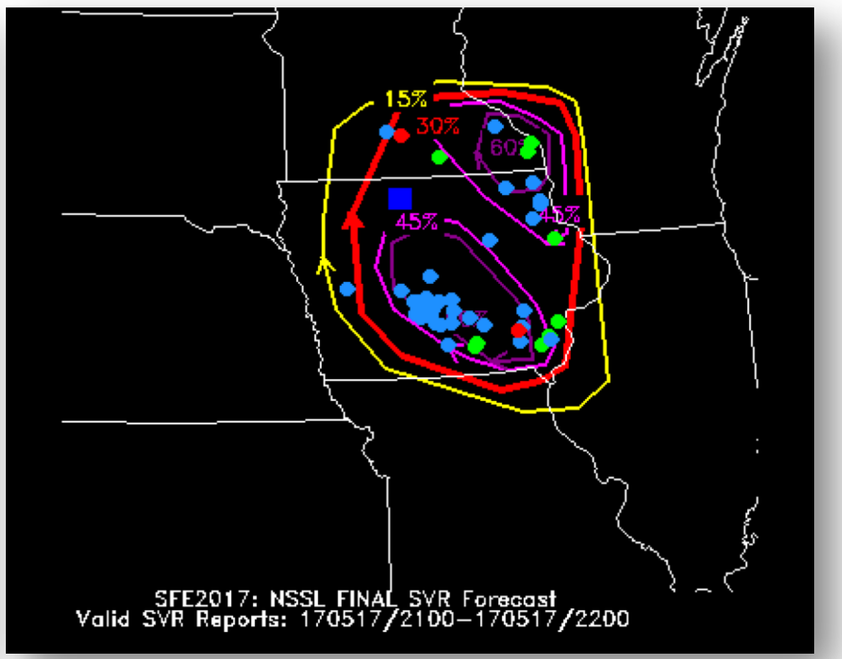
Expert forecaster used NEWS-e and observations to issue 1-hour Severe Weather Outlooks

17 May 2017 Straight-line wind forecast

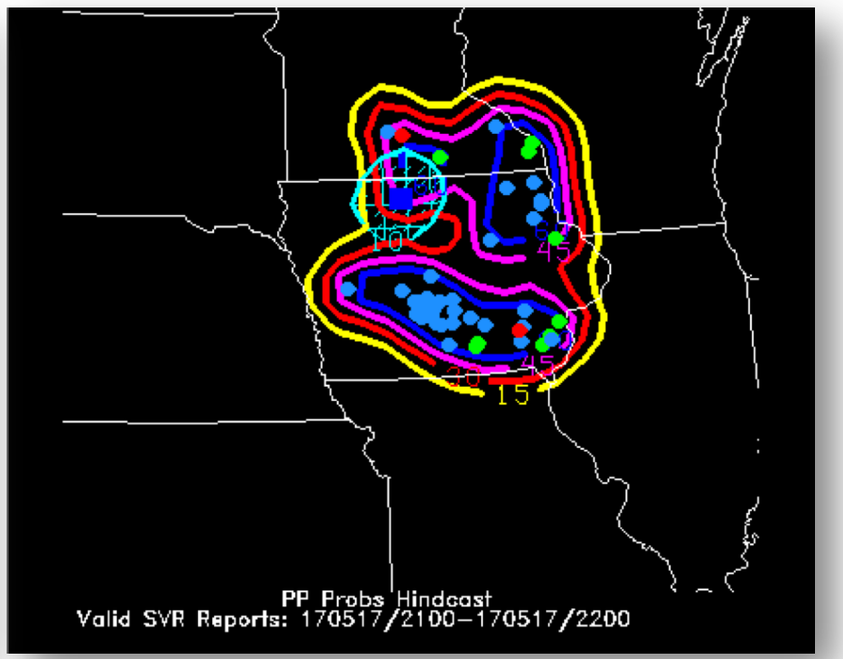
2100-2200 UTC Outlook issued from 1900 WoF Forecast



2100-2200 UTC Outlook issued from 2000 WoF Forecast



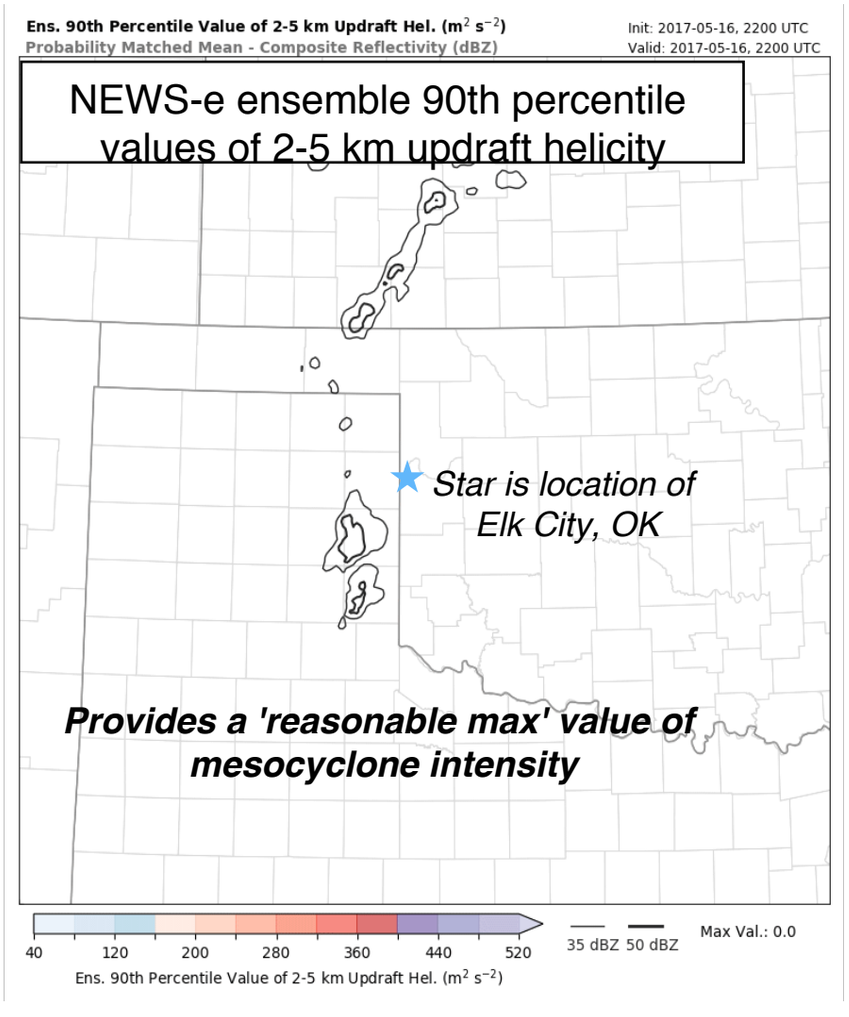
Practically perfect hindcast (verification)



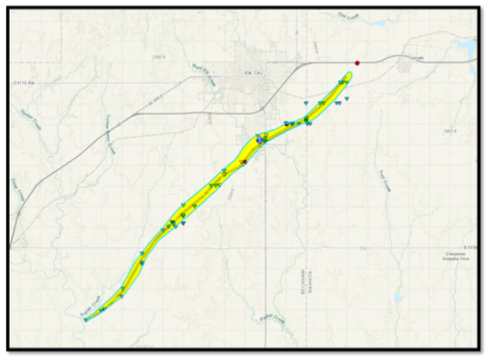
2-3-h forecasts had substantial skill, and 1-2-h forecasts were even better

2) Real-time Operational Testing: How does Norman WFO use NEWS-e guidance?

Success story: Elk City, OK EF-2 Tornado



- 6:46 to 7:12 CDT
- 15 miles long, 1000 yards wide
- $25 Million, 1 dead, 10 injured



Images courtesy Todd Lindley

2) Real-time Operational Testing: How does Norman WFO use NEWS-e guidance?

Collaborative interpretation of NEWS-e forecasts between Warn-on-Forecast group and Norman NWS on May 16th

Continued use of NEWS-e guidance by NWS throughout the day

```
...SIGNIFICANT WEATHER ADVISORY FOR northwestern Harmon...
southwestern Roger Mills...western Beckham and northwestern Greer
Counties Until 545 PM CDT...

Storms capable of producing tornadoes were located in the Texas
panhandle. One storm was located southwest of Wheeler and the other
located northwest of Wellington at 515 pm. The storms were moving
northeast at 35 MPH. These storms will move into western Oklahoma
before 6 PM. Severe weather is likely with these storms as they move
into Oklahoma and there is a high probability that tornado warnings
will be issued.

PRECAUTIONARY/PREPAREDNESS ACTIONS...

Monitor the situation closely. Be ready to act quickly if a warning
is issued or if storms threaten you.
```

*Resulted in inclusion of wording indicating a high likelihood of tornado warnings in a significant weather advisory - **Represents the first operational usage of Warn-on-Forecast***

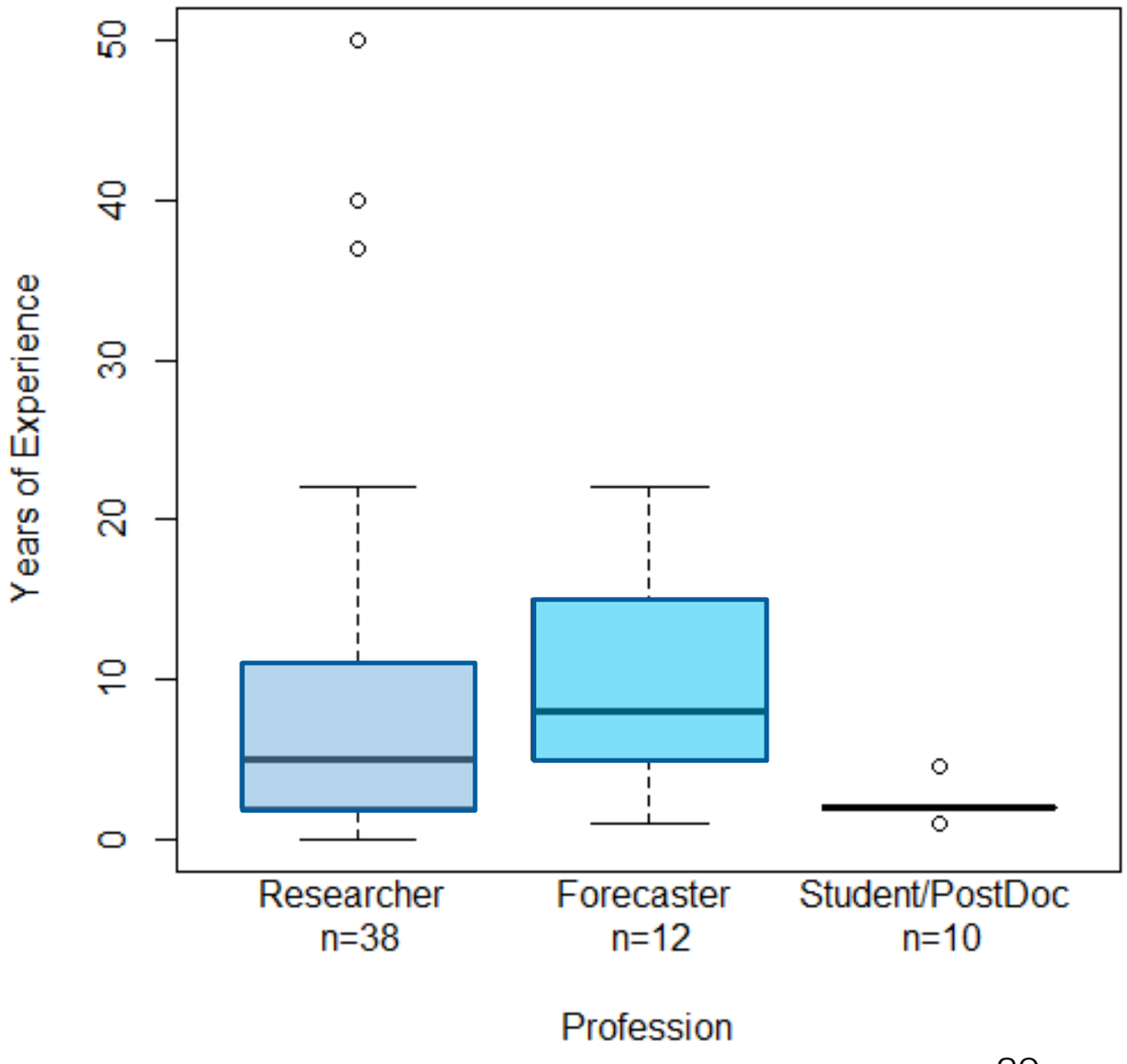
"We had a picture of the storms and their evolution before they became life-threatening. We used this model guidance to forecast with greater lead time and greater confidence" - Todd Lindley, NWS Norman

"Based on the information from the NWS, we were able to activate outdoor warning sirens about 30 minutes ahead of the tornado" - Lonnie Risenhoover, Elk City EM

3) Survey: Interpretations of NEWS-e Products

Goal: To gain understanding of meteorologists' interpretations of probability concepts used in WoF products, including ensemble probabilities and percentiles.

Demographic Information (60 of 62 Respondents)



Gender/ Age/ Education

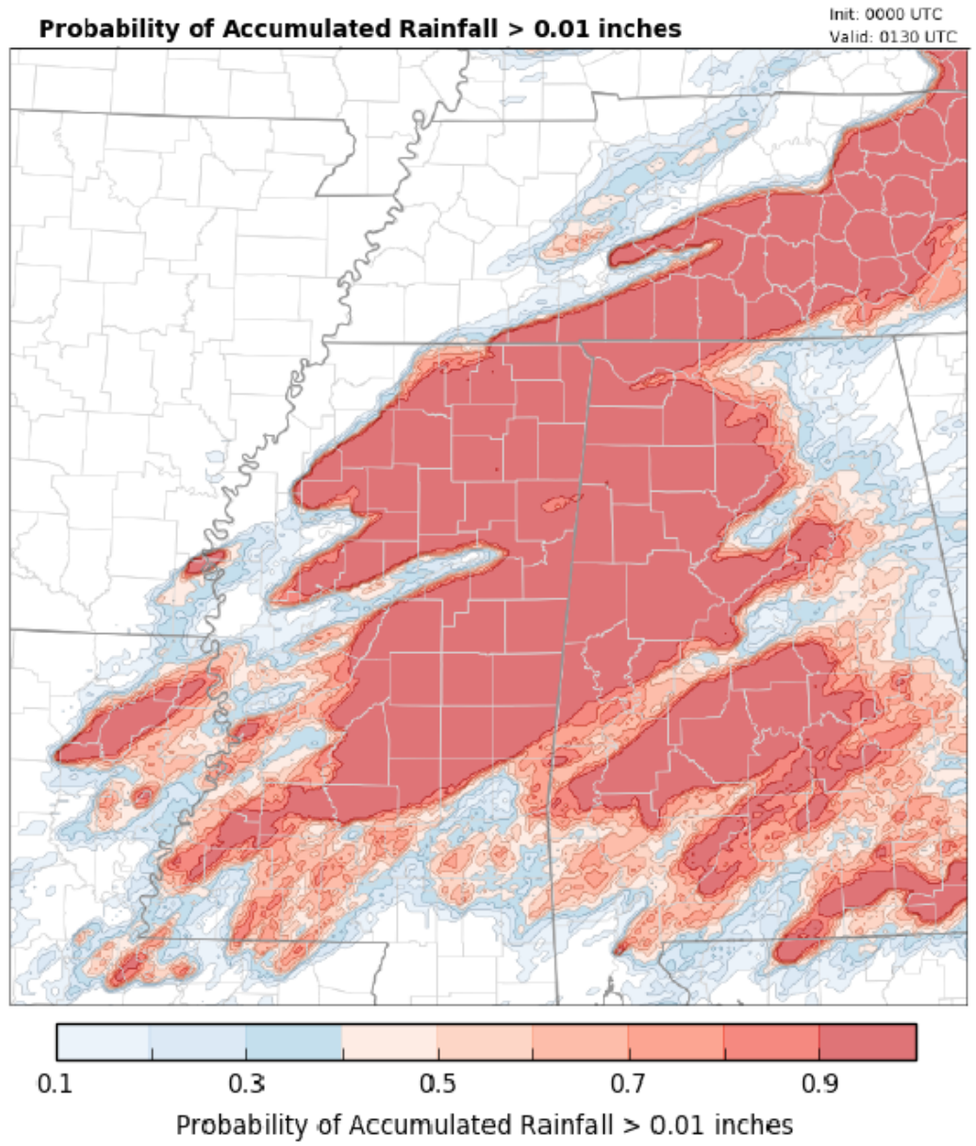
Female, n=8
Male, n=52

18-29, n=20
30-49, n=34
50-64, n=6

College graduate, n=7
Post graduate, n=53

Q: In 1-3 sentences, please describe what kind of event is depicted by this graphic.

Intended interpretation: Widespread and some isolated areas have > 90% probability of exceeding 0.01” of rainfall between 0000 and 0130 UTC.



Meteorologist interpretations on rainfall amount (n=54):

- Light rain or low magnitude (n=9);
  *4 with high certainty; 1 with low certainty

- Expressed using an inequality (n=19)

- Generic rainfall event, no amount specified (n=14)

“A widespread **rain event**.”

“A highly certain widespread rainfall event. No idea how large of magnitude of rainfall is expected.”

- Heavy rainfall (n=5)

“Heavy rain event.”

“There could be excessive rainfall but we can't know since the probability is for rainfall over 0.01 inch.”

“Flash flood.”

Q4. In an ensemble-based probabilistic forecast, what do you think the 70th percentile value of accumulated rainfall represents?

Of 60 participants, 55% (n=33) demonstrated a clear understanding

70% of members have a value less than this/ 70th percentile (n=29)	30% (or, at least a minority) of members have a value more than this (n=10)	High-end possibility/ showing something akin to the max (n=8)	Probability distribution function concept (n=12)
"It means that 70% of ensemble members have this much precipitation or less over the accumulated time period."	"70% of the ensemble members have less than the amount shown / 30% of the members have more."	"… Getting towards a "maximum reasonable" precipitation estimate from the raw ensemble distribution."	"If there are 10 ensemble members, the QPF from the 7th-highest member at any given point."

Q4. In an ensemble-based probabilistic forecast, what do you think the 70th percentile value of accumulated rainfall represents?

Of 60 participants, 45% (n=27) demonstrated misunderstanding or ambiguity

"The accumulated rainfall amount that 70% of the members are producing."	"Accumulated rainfall that at least 70% of the members agree on."	"The regions shaded represent the union of at least 70% of ensemble members."

Quick demographic analysis (26 of 27 of the respondents)

- N=16 (44.4%), Research scientist/professor (experience ranging 0.8–37 years)
- N=4 (33.3%), Forecaster (experience ranging 4–15 years)
- N=6 (60%), Graduate student (experience ranging 1–4.5 years)

This lack of understanding spread across academic, research, and operational participants with varied experience

Summary

- The NOAA WoF program is developing NWP ensemble prediction systems that will ultimately enable forecasters to issue tornado, severe thunderstorm, and flash flood warnings further in advance

- A prototype real-time WoF system, NEWS-e, is being evaluated every spring by operational forecasters and research scientists

- NEWS-e produces valuable probabilistic forecast guidance for severe thunderstorms at 0-3-hour lead times

- The limited predictability of storms and massive data volume of storm-scale ensembles pose major scientific, computational, and human factors challenges

Ongoing challenges

- Refine knowledge of impacts of different forecast error sources in order to better prioritize improvements to model and ensemble configurations
- Refine knowledge of how forecasters use ensemble output in order to better design WoF products and forecaster training
- Improve ability to anticipate forecast skill for individual events and storms
  - For example, need better understanding of relationship between storm environment and forecast skill
  - Develop machine learning techniques for calibrating forecast probabilities
- Scale ensembles to O(10,000) CPU cores to achieve model $\Delta x \leq 1$ km (especially important for tornado prediction)
  - Real-time post-processing will also need to accelerate
- Improve physics parameterizations, errors in which greatly limit storm-scale predictability
- Verification – a major challenge at these scales!!
  - Lack of observations
  - Design error metrics that are consistent with subjective impressions

The rest of the Warn-on-Forecast team:

Gerry Creager, Monte Flora, Jidong Gao, Junjun Hu, Thomas Jones, Kent Knopfmeier, Swapan Mallick, Derek Stratman, Yunheng Wang, Nusrat Yussouf

References

Potvin, C. K., E. M. Murillo*, M. L. Flora*, and D. M. Wheatley, 2017: Sensitivity of supercell simulations to initial-condition resolution. *J. Atmos. Sci.*, **74**, 5-26.

Potvin, C. K., and M. L. Flora*, 2015: Sensitivity of idealized supercell simulations to horizontal grid spacing: Implications for Warn-On-Forecast. *Mon. Wea. Rev.*, **143,** 2998-3024.

Potvin, C. K., and L. J. Wicker, 2013: Assessing ensemble forecasts of low-level supercell rotation within an OSSE framework. *Wea. and Forecasting*, **28**, 940-960.

Stensrud, D. J., and Coauthors, 2009: Convective-scale warn-on-forecast system: A vision for 2020. *Bull. Amer. Meteor. Soc.*, **90**, 1487–1499.

Stensrud, D. J., and Coauthors, 2013: Progress and challenges with Warn-on-Forecast. *Atmos. Res*., **123**, 2-16.

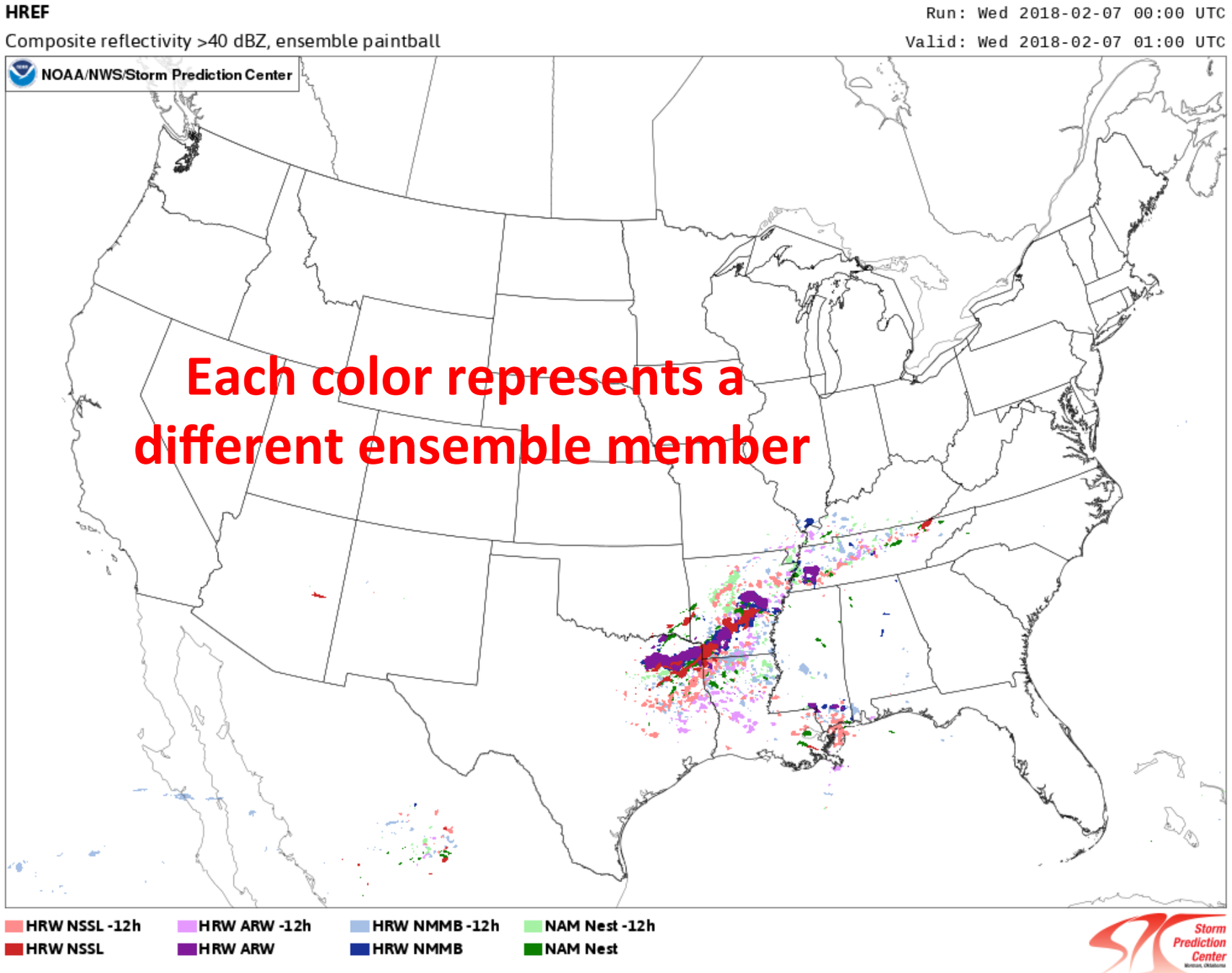
Wheatley, D. M., K. H. Knopfmeier, T. A. Jones, and G. J. Creager, 2015: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part I: Radar data experiments, Wea. Forecasting, 30, 1795-1817.

Jones, T. A., K. Knopfmeier, D. Wheatley, G. Creager, P. Minnis, and R. Palikonda, 2016: Storm-scale data assimilation and ensemble forecasting with the NSSL Experimental Warn-on-Forecast System. Part II: Combined radar and satellite data experiments. Wea. Forecasting, 31, 297-327.

Flora, M. L.*, C. K. Potvin, and L. J. Wicker: Supercell predictability: Exploring ensemble forecast sensitivity to initial condition spread. Conditionally accepted, *Mon. Wea. Rev.*

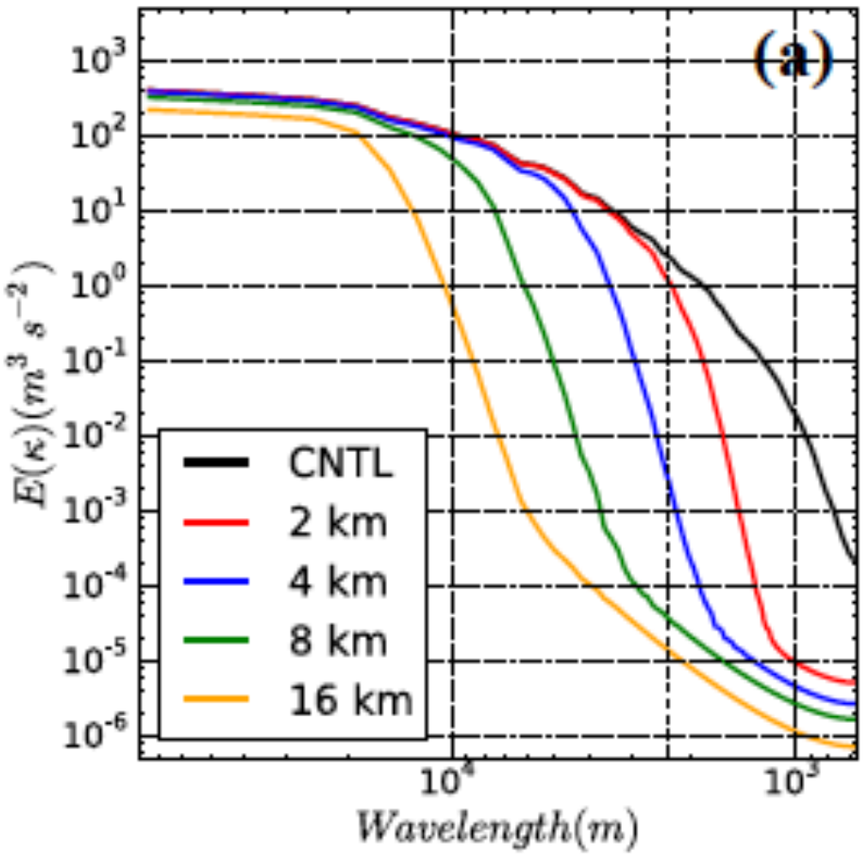
Extra slides

Sample HREF output

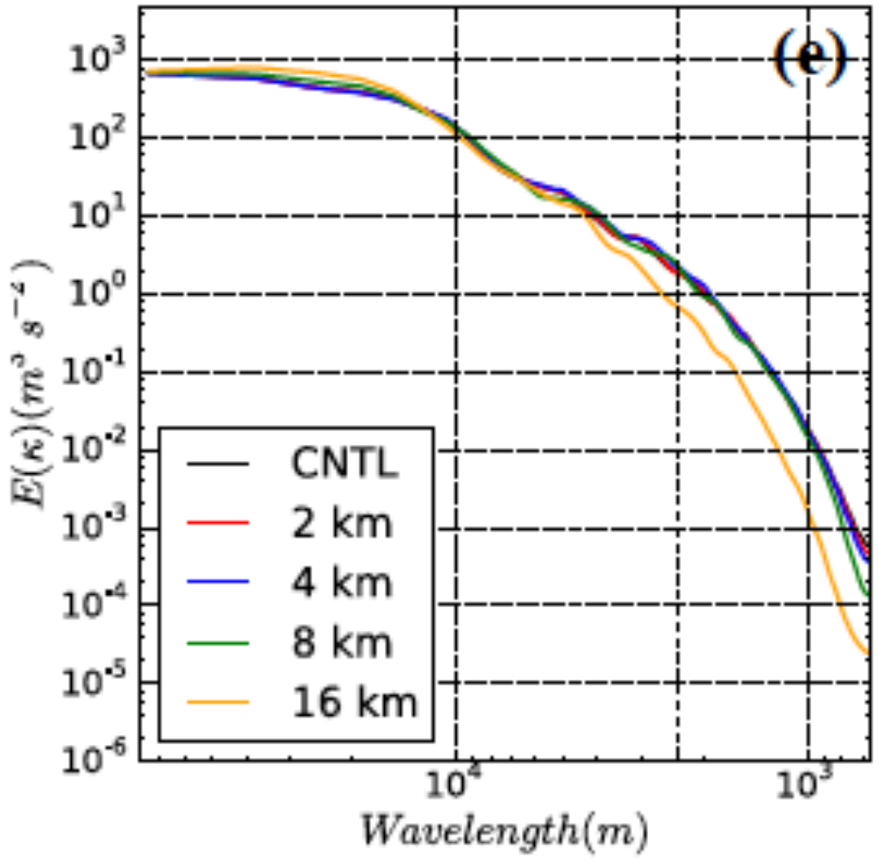


Largely since missing scales regenerate within 5-10 min of forecast

_w_ spectra at _t_ = 0 min



_w_ spectra at _t_ = 5 min



- Similar results for other cases & variables
- In organized convection, larger scales (intra-storm and environment) strongly determine evolution

Verification of NEWS-e Forecasts

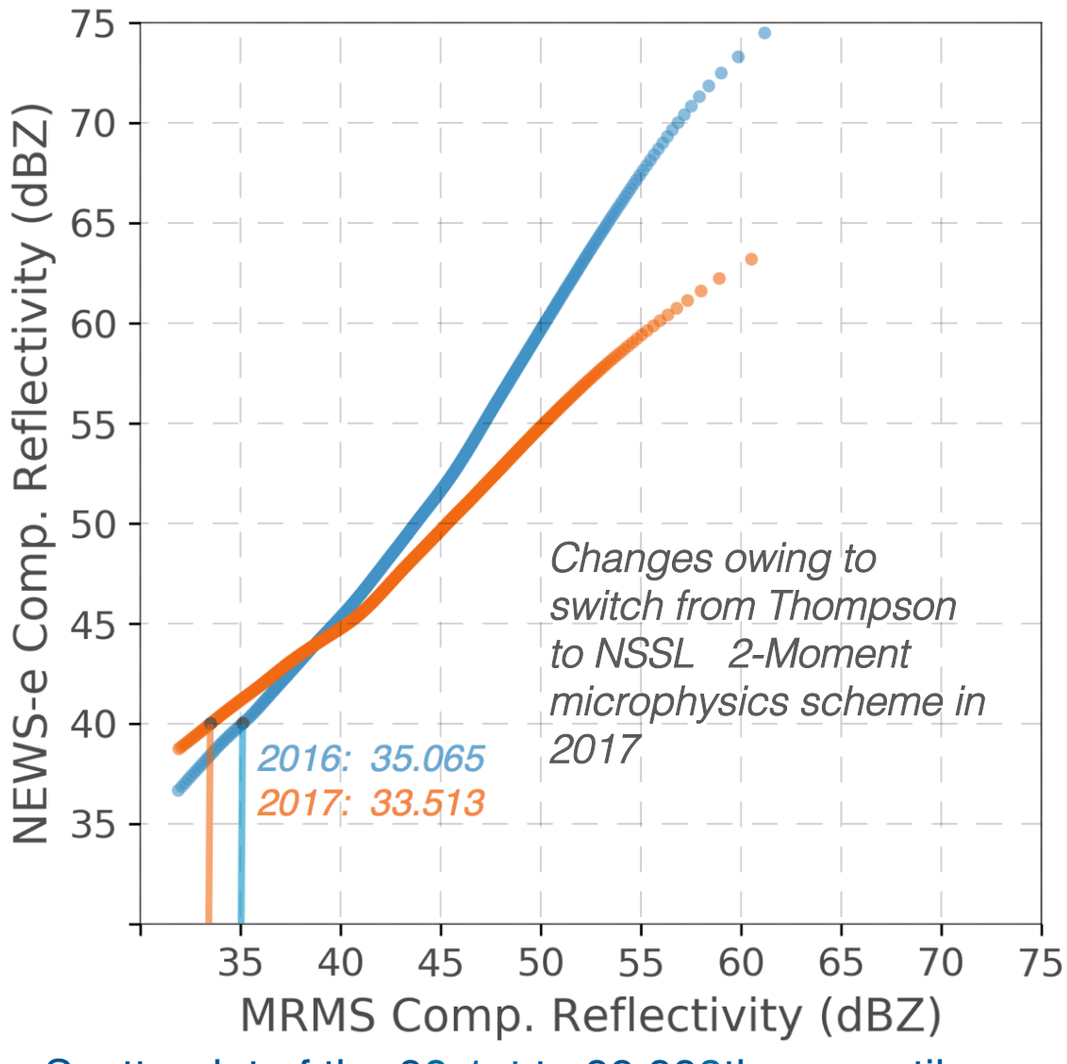
Challenge: Need to verify forecasts of sparse, transient phenomena, such as mesocyclones, that are not fully observed by conventional observations

Object-based verification (e.g. Davis et al. 2006a, b) allows forecasted thunderstorms or mesocyclones to be matched to observational proxies.

1. Climatologies of forecast and observed quantities allow corresponding values of extreme percentiles in the cumulative distribution function of each dataset to be matched

2. These *percentile thresholds* are used for identifying objects in forecast and verification fields

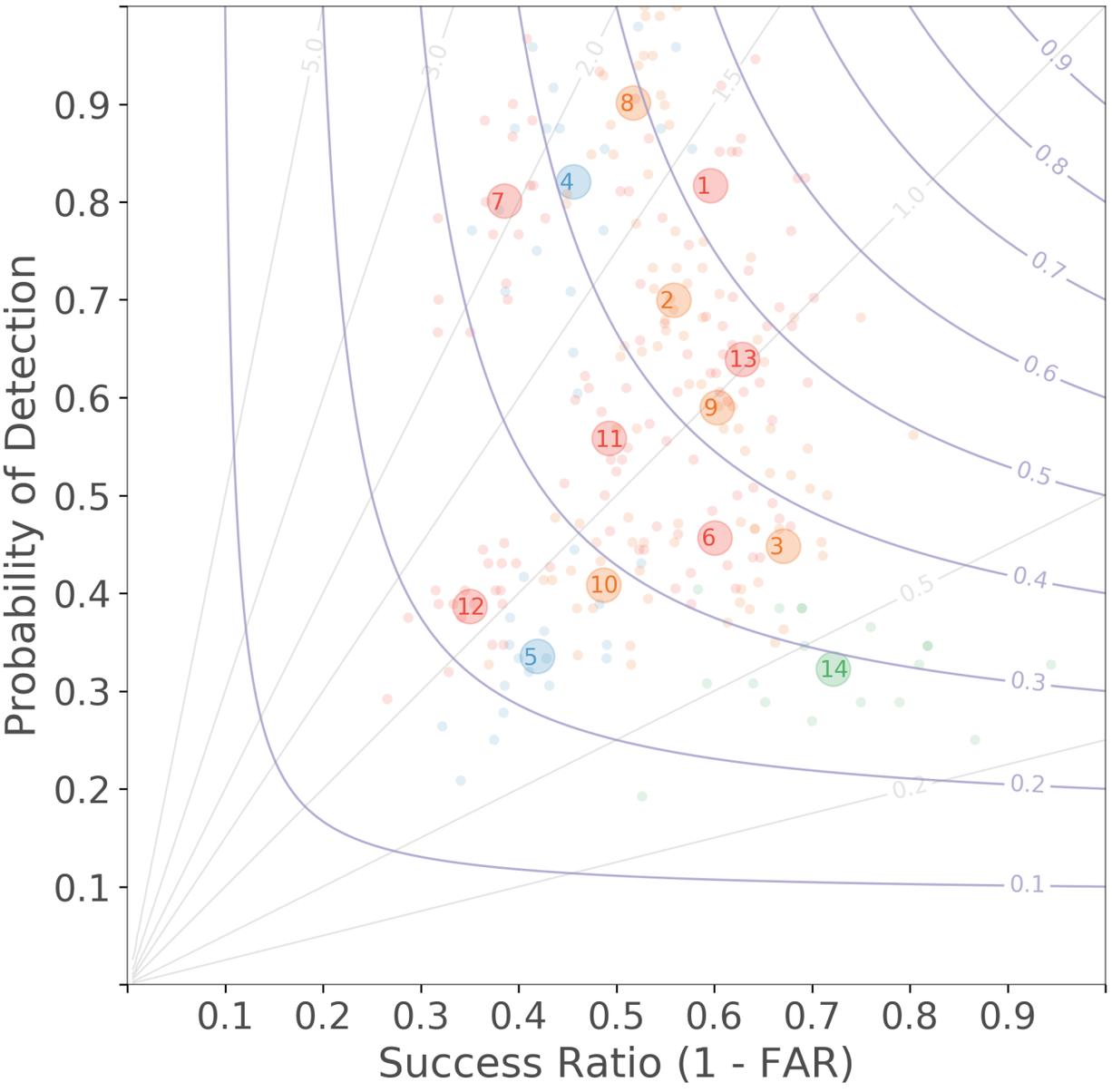
3. Forecast objects can be matched to observed objects in space and time, allowing contingency-table based verification scores to be calculated



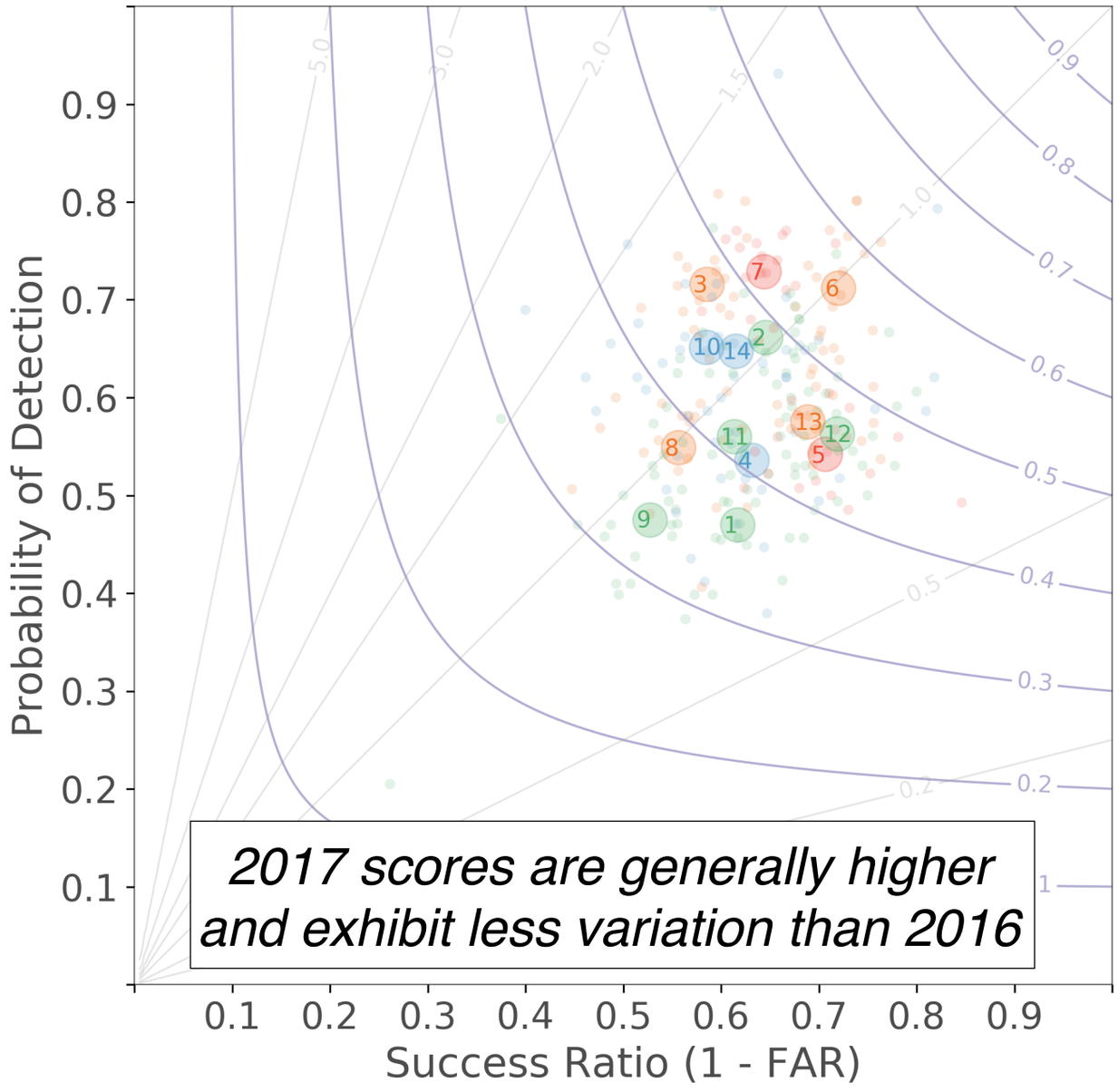
Scatterplot of the 99.1st to 99.998th percentile values of NEWS-e forecast and MRMS observed composite reflectivity. matching values are used to identify thresholds for object identification

NEWS-e Composite Reflectivity Verification

2016 NEWS-e at 1-hr lead time



2017 NEWS-e at 1-hr lead time



2017 scores are generally higher and exhibit less variation than 2016

Red: >=Enhanced risk, cellular; Orange: >=Enhanced risk, mixed or linear
Green: <Enhanced risk, cellular; Blue: <Enhanced risk, mixed or linear

Large circles represent ensemble mean scores for each case during the year (14 cases for each year) and small circles represent scores for individual ensemble members. Color coding is by storm mode and maximum SPC risk in NEWS-e domain.

NEWS-e Composite Reflectivity Verification

Improved reflectivity forecasts largely attributable to switch to NSSL 2-moment microphysical parameterization in 2017