Convection-Allowing Models (CAMs)

A discussion on Creating a Roadmap to a
Unified CAM-based Data-Assimilation and Forecast Ensemble

Panel Members

- Curtis Alexander (ESRL/GSD)
- Adam Clark (NSSL)
- Lucas Harris (GFDL)
- Geoff Manikin (NCEP/EMC)
- Lou Wicker (NSSL)
- Ming Xue (OU/CAPS)

A Proposed Pathway to a Unified CAM-based Ensemble

- *NMMB development frozen*
- *HREFv3 (FY19) will reflect transition to ARW, FV3 membership*
- *Eventually, single-core (FV3) system is expected (~FY21)*

FY	ARW-Framework	FV3-Framework
2018	- RAPv4/HRRRv3 Operational - RAPv5/HRRRv4 Experimental Dev: ● Storm-Scale Ensemble DA ● Larger CONUS Domain ● Improved Physics ● Testbed and objective verification	- Real-time 15-km Global + 3 km FV3 Nest - Regional Stand-Alone FV3 + DA developed, tested - Multiple physics options tested on CAM scales, including RAP/HRRR Physics Suite - Comparisons between global nests and stand-alone regional to ensure consistent behavior - Testbed and objective verification
2019	- RAPv5/HRRRv4 Experimental Testing ● Storm-Scale Ensemble DA - JEDI Observation Operators - HRRRv3 included in HREFv3 - Testbed and objective verification	- Optimization of CAM-scale DA and physics - Hourly-Updating CAM Ensemble DA - JEDI Observation Operators - FV3 CAM included in HREFv3 - Testbed and objective verification
2020	- RAPv5/HRRRv4 Implementation - ARW Development Frozen	- FV3-Rapid-Refresh Optimization ● Complete JEDI Integration ● In-Core DA
2021		Real-Time Experimental Rapid Refresh Forecast System (RRFS) ● FV3-HRRR Ready ● Merge NPS CAM Products
2022		RRFS Operational ● Unified 3-km CAM Ensemble Physics/DA for Days 1-3

Questions to keep in mind…

1) A new implementation of the HRRR model is imminent and this popular WRF-ARW-based deterministic modeling system has been optimized for over a decade, but it is slated to be subsumed by an FV3-based CAM-scale forecasting system as soon as the change can be justified by performance. In what different contexts/applications have FV3-based CAMs been tested to date and what, in your opinion do the results indicate about the capabilities of the FV3 dynamic core in a CAM framework?

2) Do we have sufficient "evidence" to reach a consensus that FV3CAM is "in the same class" as the ARW framework for CAM-scale prediction?

3) Do we have sufficient "evidence" to reach a consensus that FV3CAM is NOT "in the same class" as the ARW framework for CAM-scale prediction?

4) If additional evidence is needed to reach consensus on either question, what additional testing/metrics are required to provide sufficient evidence?

5) If the FV3 numerical method fundamentally handicaps CAM-scale prediction, can it be modified/improved?

FV3-CAM readiness: an NSSL & SPC Perspective

Presented by

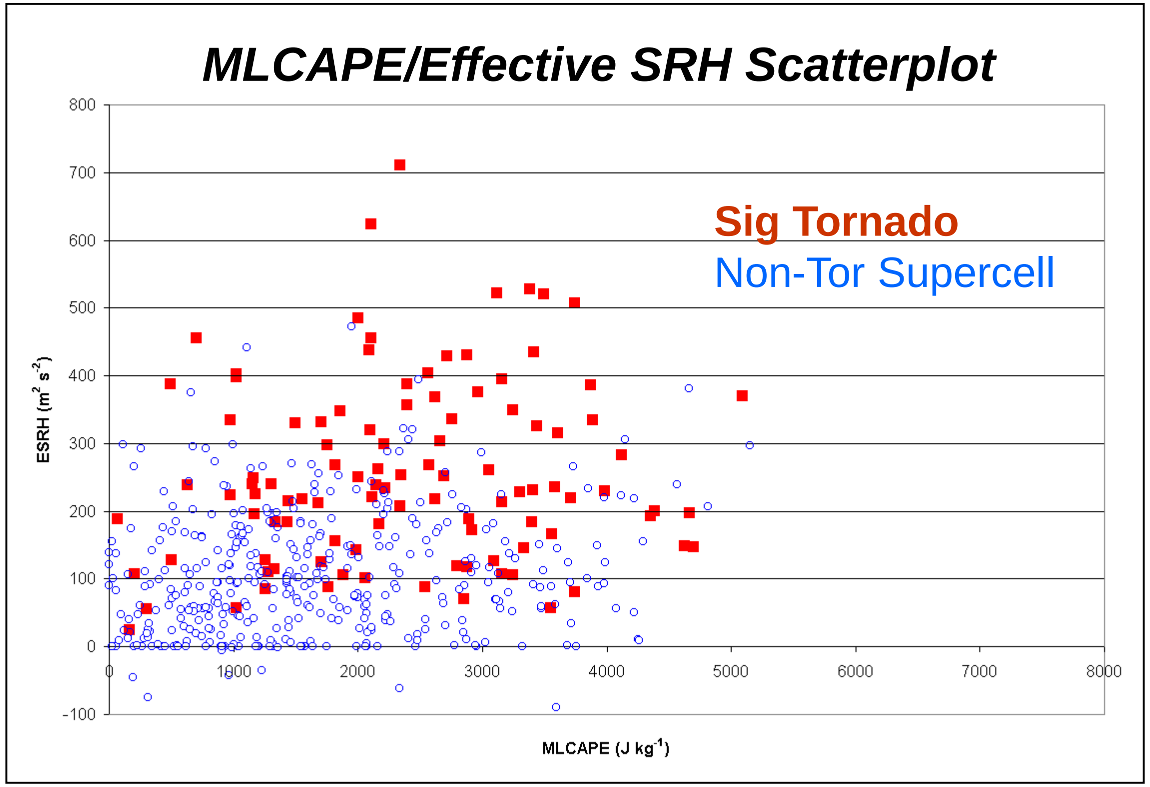
Lou Wicker and Adam Clark

with input from

NOAA's Storm Prediction Center

How/why forecasters use CAMs

- Improved predictions of the mesoscale environment are the first step to improving severe weather forecasts
  - But environment info. (CAPE/shear, etc.) may not be sufficient
    - *Similar environments can produce different convective weather*
    - *Different environments can produce similar convective weather*



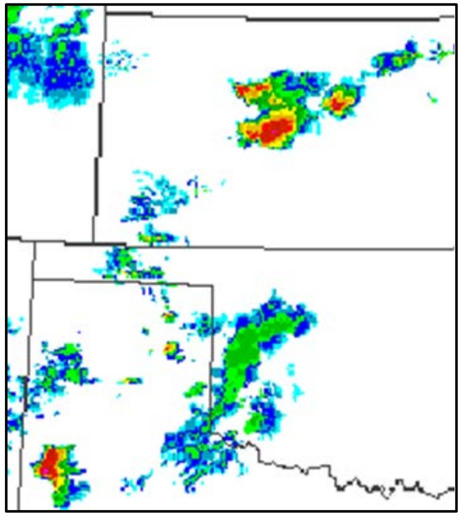
Many ***non-tornadic supercells*** share similar parts of parameter space with ***supercells producing significant tornadoes***

(From Thompson et al. 2007)

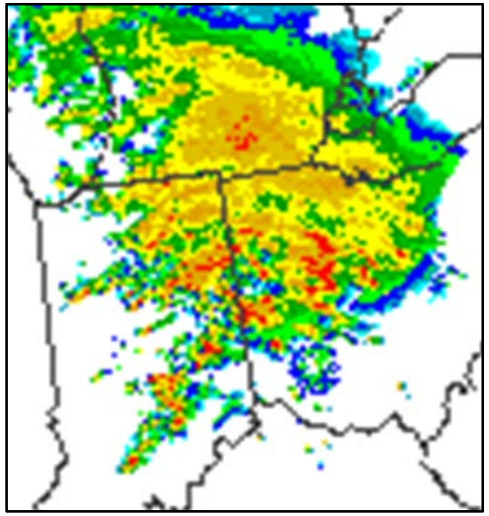
CAMs: *Convective Mode*

- Severe weather hazards (tornadoes, hail, damaging winds) are often closely related to **convective mode.**
  - *Tornadoes (discrete supercells; embedded mesovortices in QLCS).*
  - *Damaging wind (bow echoes and bowing line segments).*

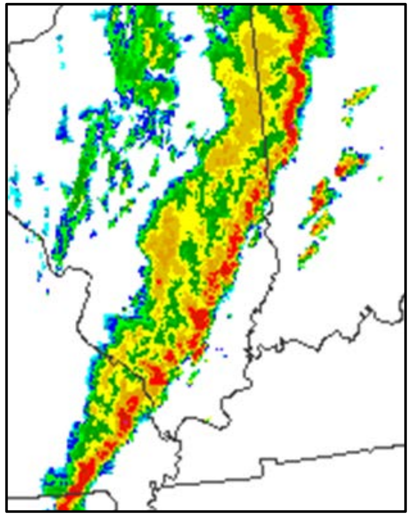
Severe weather forecasters need to accurately predict convective mode and character of storms (storm-scale details).



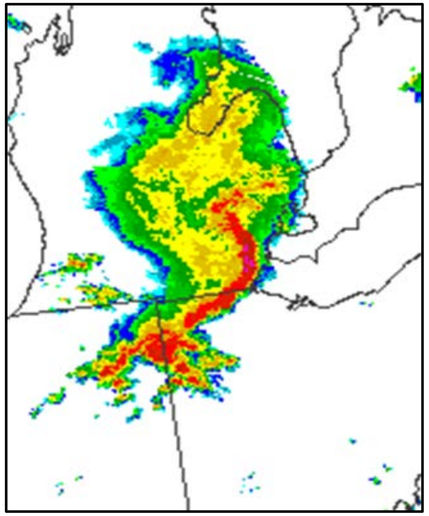
Discrete Cells



Multi-Cell Cluster



QLCS



Bow Echo

SPC/NSSL perspective…

- SPC strongly supports a unified modeling system to:
  - Consolidate and concentrate DA and modeling development efforts within NOAA and across the larger community
  - More efficiently utilize HPC resources
  - Provide forecasters with fewer, **but improved**, sources of NWP guidance.

- To maintain the current level of services, SPC needs any FV3-CAM to have similar performance characteristics (subjectively and objectively) to the ARW-CAMs.

- However, SPC recognizes the challenges…
  - If equal resources continue to be devoted to ARW and FV3 improvements, it is not clear when the FV3 will "catch up"? Going forward, how does NOAA allocate resources to develop the FV3-CAM?
  - Perhaps ARW-CAM is close to its performance maximum: logical time to shift more resources toward FV3?
  - Lots of issues to discuss…

CAM Development Process

- Last 15 years
  - Tied to the yearly severe wx cycle (and now winter wx season)
  - Tested using SPC in HWT (Norman) and (now) in the HMT (College Park).
  - Tests needing forecasters for evaluation will proceed at a much slower pace (need the weather!)

- What about accelerated development?
  - Requires large RESOURCES to reforecast 3-4 months from a year 4-6x over 12 month period?
  - Personnel needed: 10-15 FTEs? Right now in bits and pieces.

- Challenges unique to FV3 so far (at NSSL)
  - inconsistent performance, resource issues, etc.
  - Need regional version, better documentation, more experiences.
  - Knowledge base for FV3 code still confined to a few people

- Software is still not portable enough
  - must run on wide range of systems to leverage community
  - Either NEMS + FV3 must be made workable on non-NOAA systems **OR**
  - FV3 EMC core needs to be placed in a simpler framework

Where are we today?
FV3-CAM 2017 HWT Evaluation

- CAM Performance metrics
  - Ingredients-based forecasting metrics
    - environmental soundings (not available in 2017)
    - synoptic/mesoscale forcing
  - Storm-based
    - CAM forecast of convective mode
    - UH intensity and coverage, maximum low-level winds, hail size, QPF, etc.
- Inter-model comparison project (C. Potvin NSSL)
- Surrogate severe (model evaluation and calibration)
- Goals:
  - Soundings need to represent convective potential
  - Storm-depiction needs to be relatable to radar
  - Timing and mode of CAM storms relative to observations.

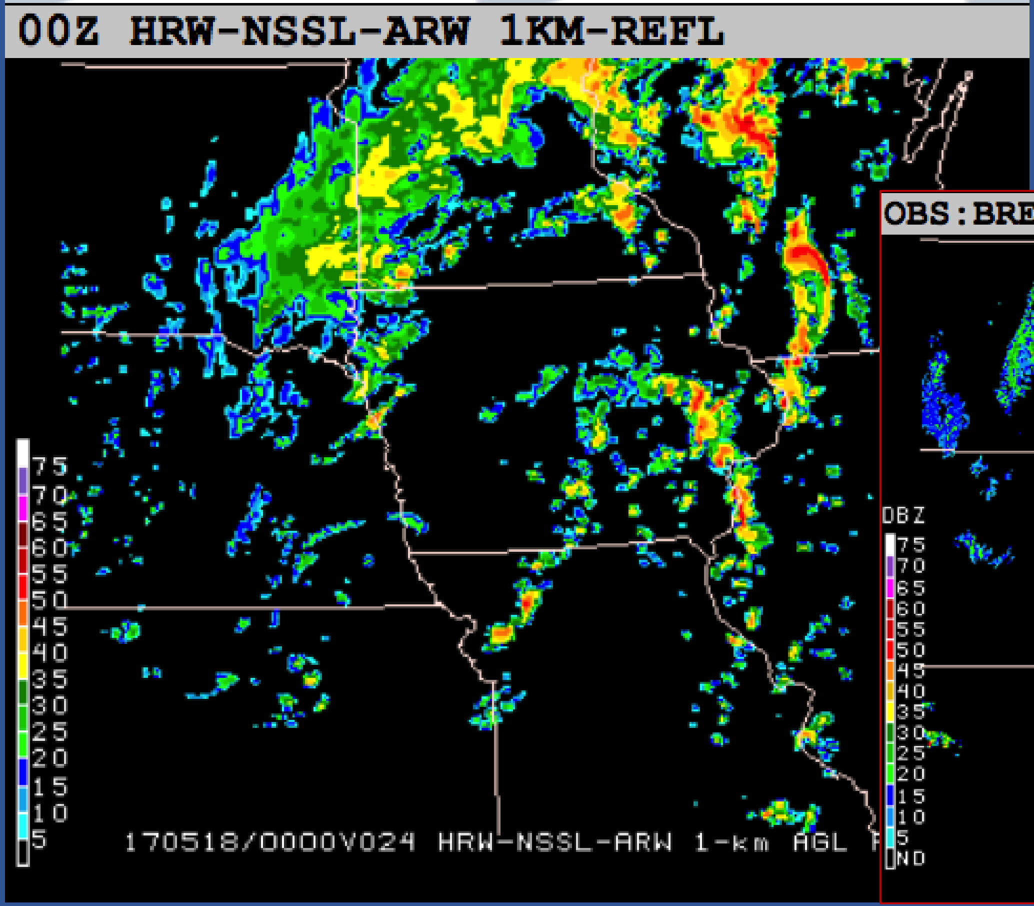
Where are we today?
FV3-CAM 2017 HWT Evaluation

- CAM Performance metrics
  - Ingredients-based forecasting metrics
    - environmental soundings (not available in 2017)
    - synoptic/mesoscale forcing
  - Storm-based
    - **CAM forecast of convective mode**
    - UH intensity and coverage, maximum low-level winds, hail size, etc.
- **Inter-model comparison project (C. Potvin NSSL)**
- **Surrogate severe (model evaluation and calibration)**
- Goals:
  - Soundings need to represent convective potential
  - Storm-depictions need to be relatable to radar
  - Timing and mode of CAM storms relative to observations.

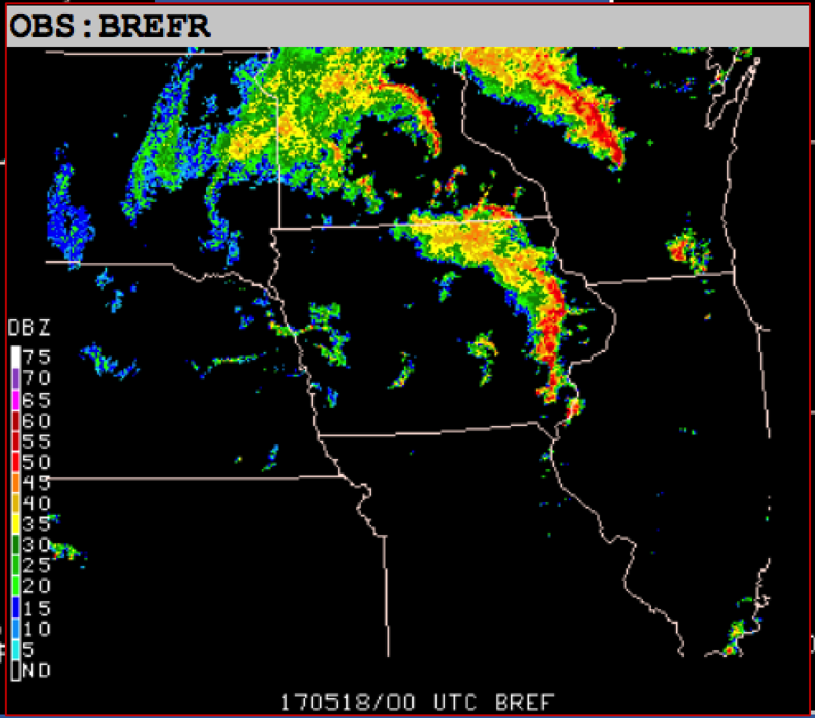
18 May IA (severe winds)

24 hour FCSTS

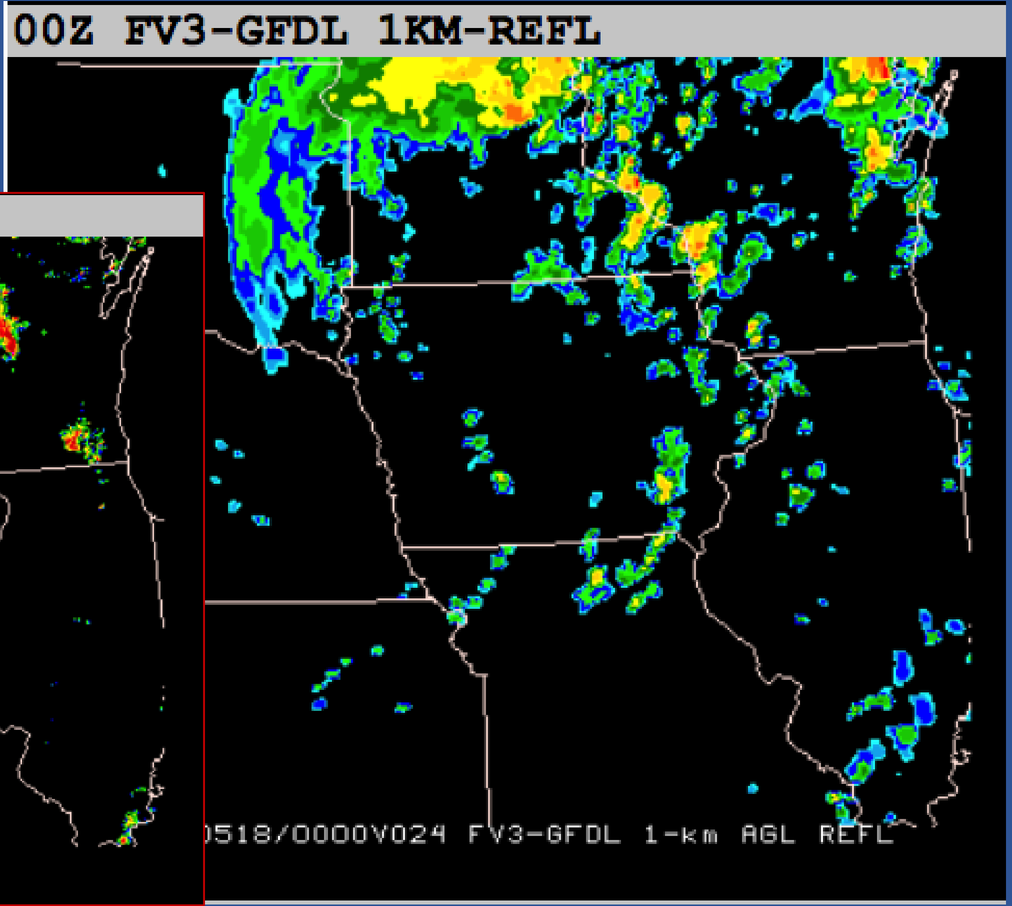
HRRR: Pretty good forecast for 24 hours

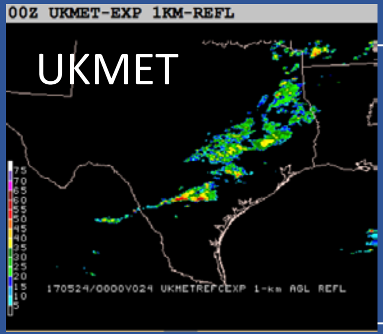


Observations



FV3-G No organized lines in E-IA or C-WI



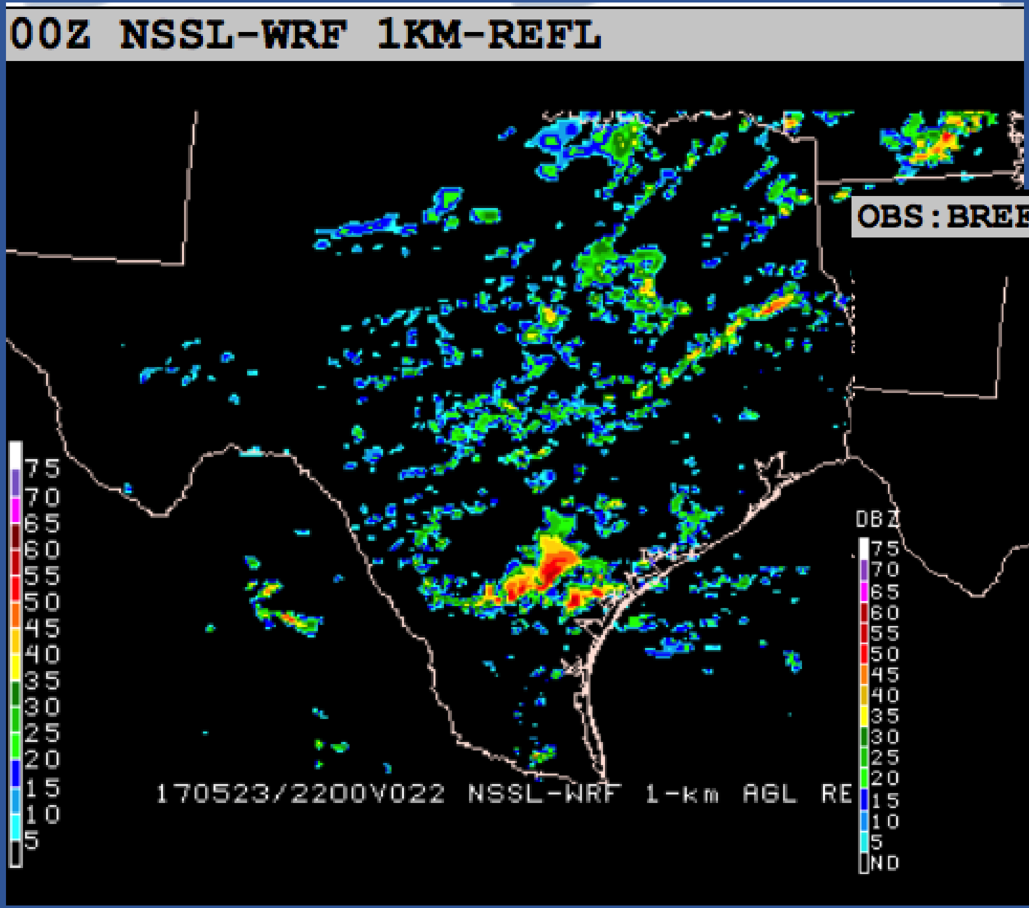


23 May 2017 TX (svr winds)

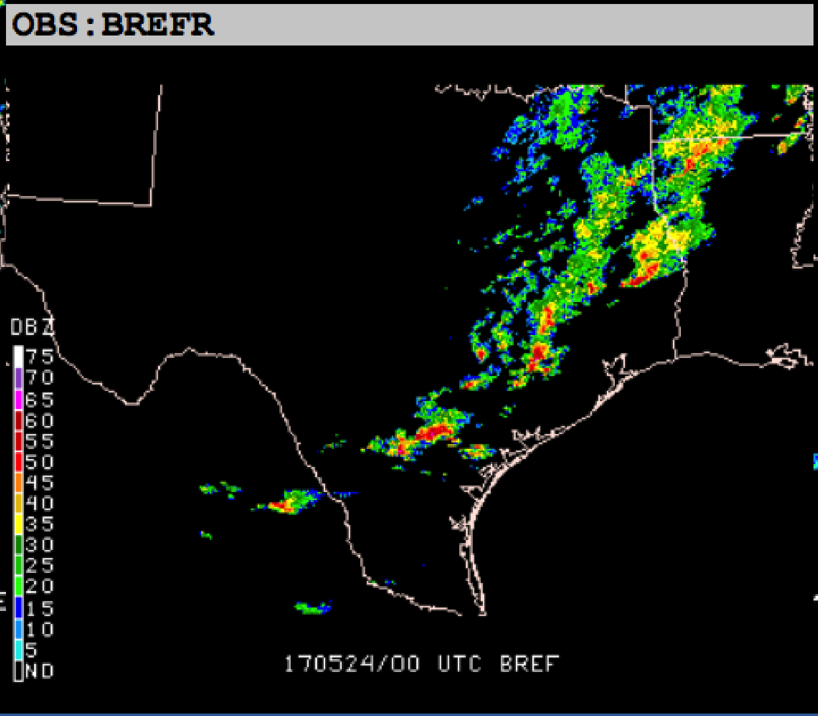
24 hour FCSTS

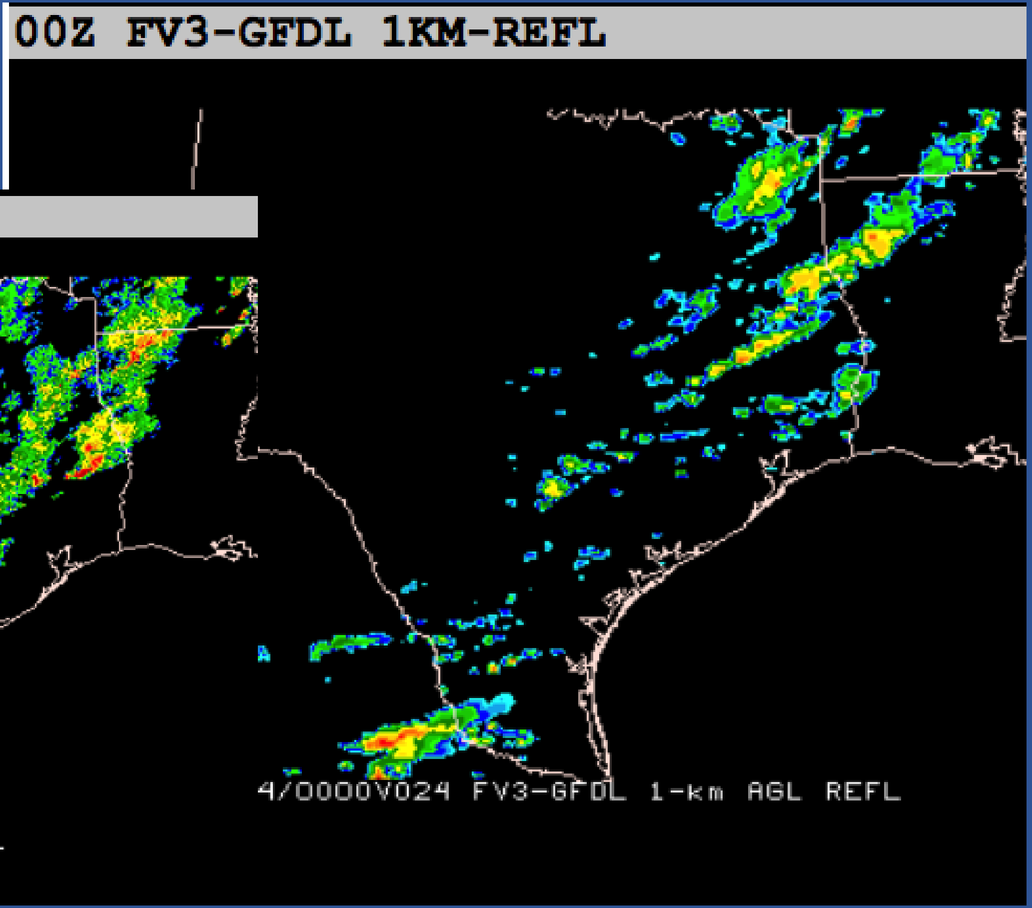
HRRR: development is 2 hours too fast (shown is 22Z forecast)

FV3-G Never develops SW storms



Observations



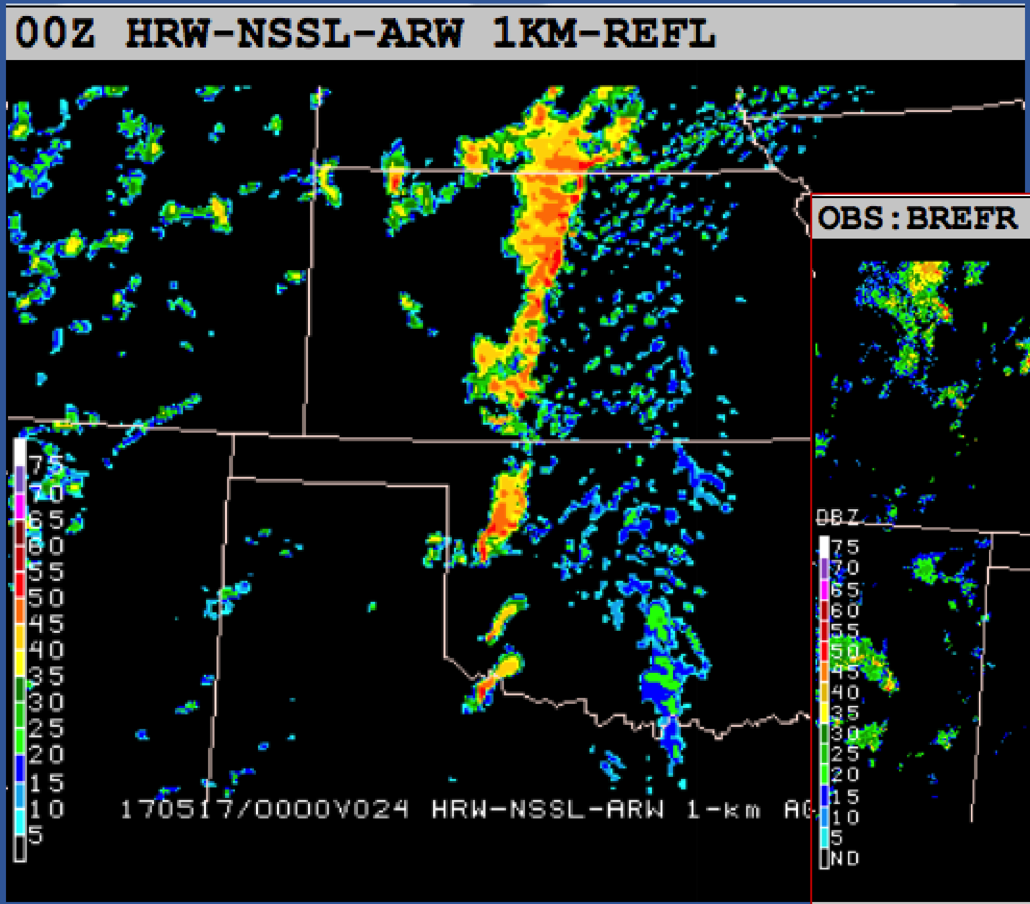


16 May OK (tornadic supercells)

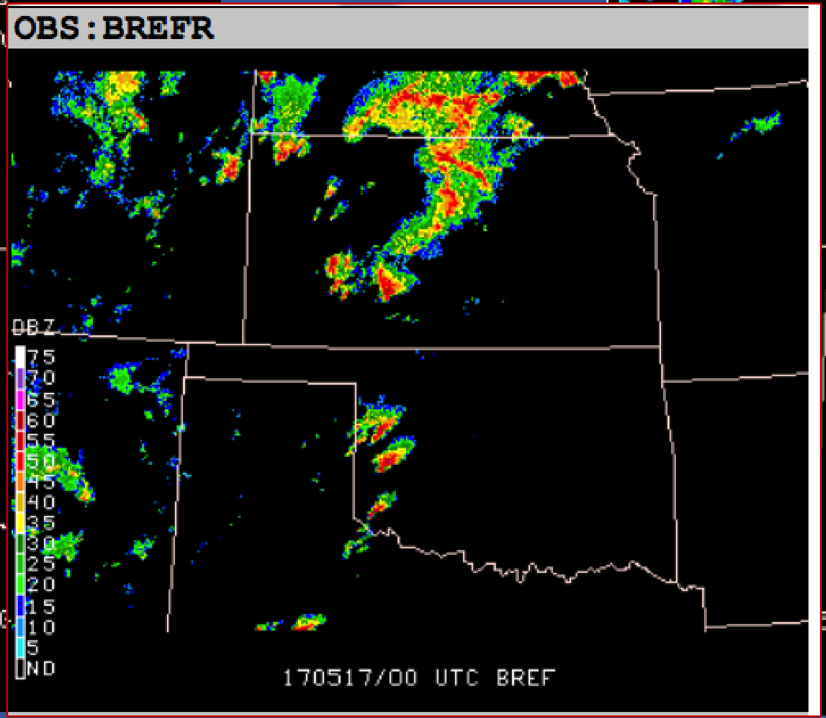
24 hour FCSTS

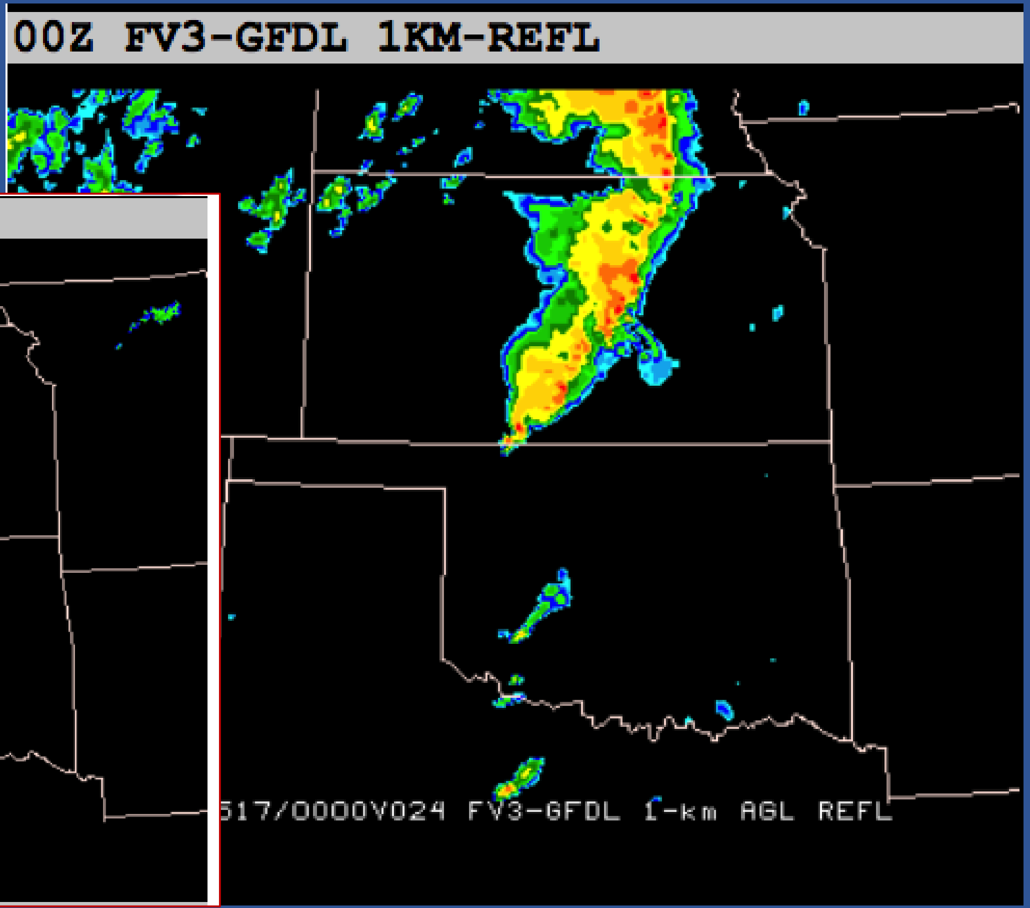
HRRR: Good fcst, too much convection in S. KS/NW OK

FV3-G No convection develops along the dryline

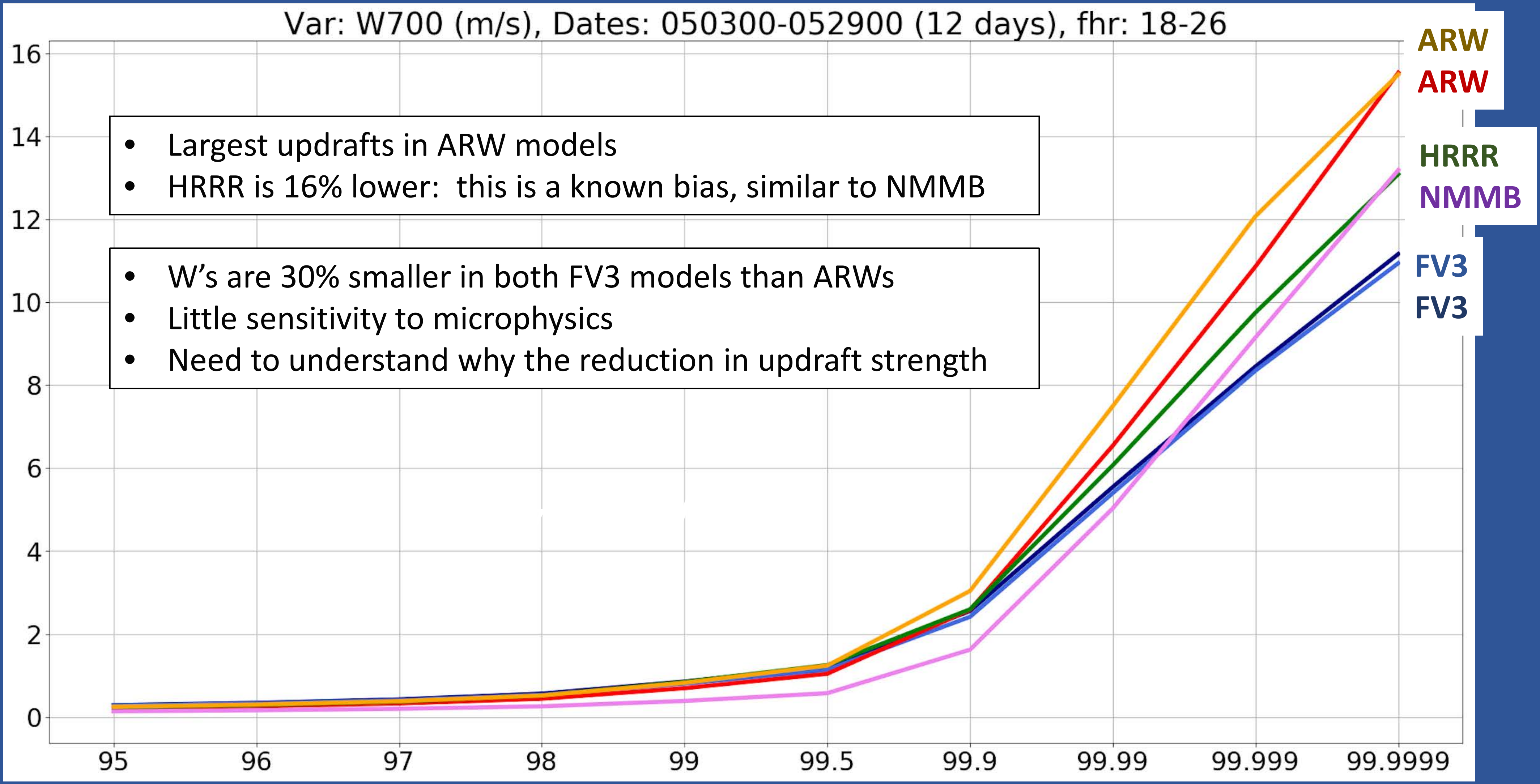


Observations





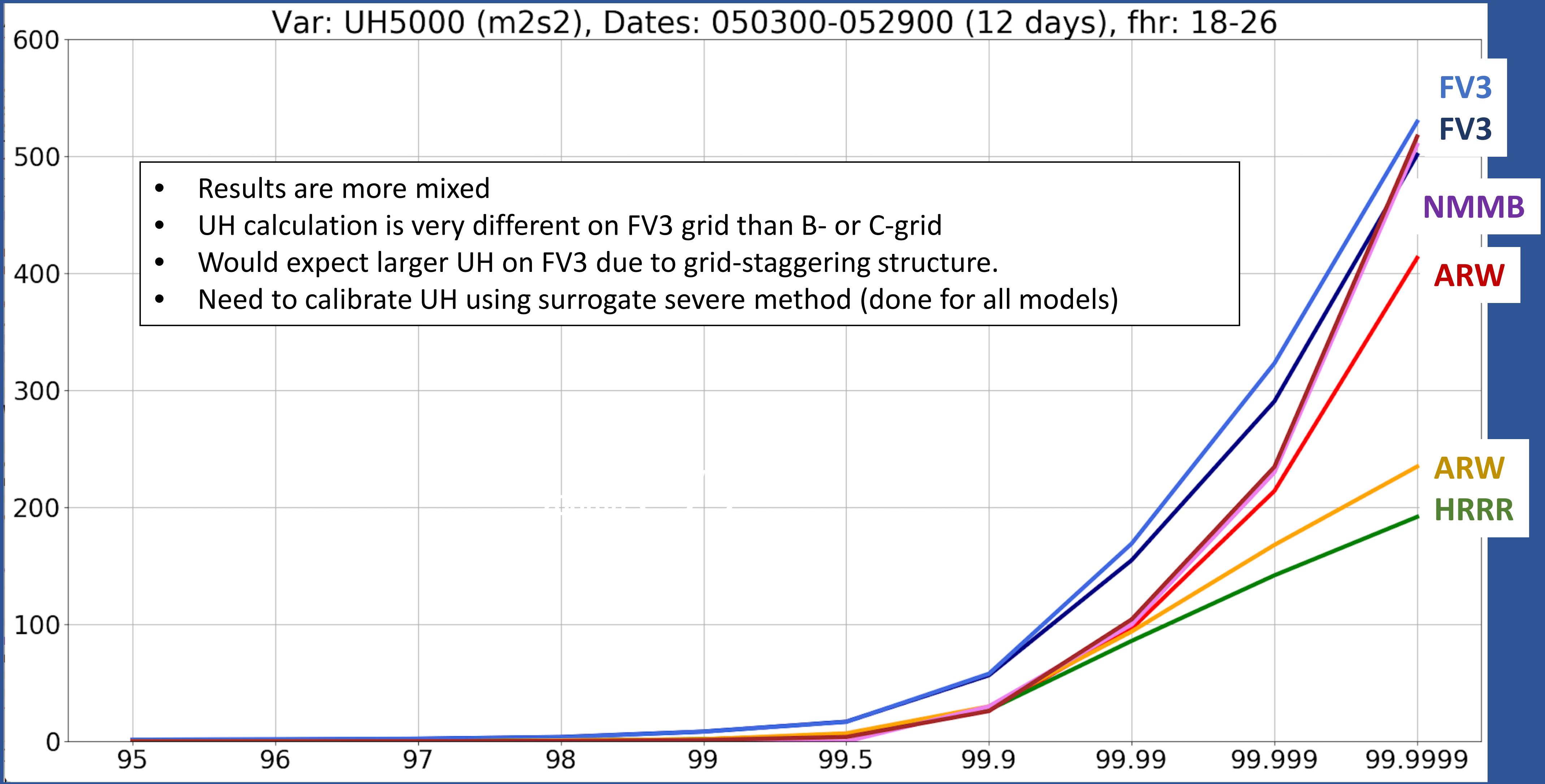
Inter-Model Comparisons: W_{700mb} Percentiles



- Largest updrafts in ARW models
- HRRR is 16% lower: this is a known bias, similar to NMMB

- W's are 30% smaller in both FV3 models than ARWs
- Little sensitivity to microphysics
- Need to understand why the reduction in updraft strength

Inter-Model Comparisons: Updraft Helicity



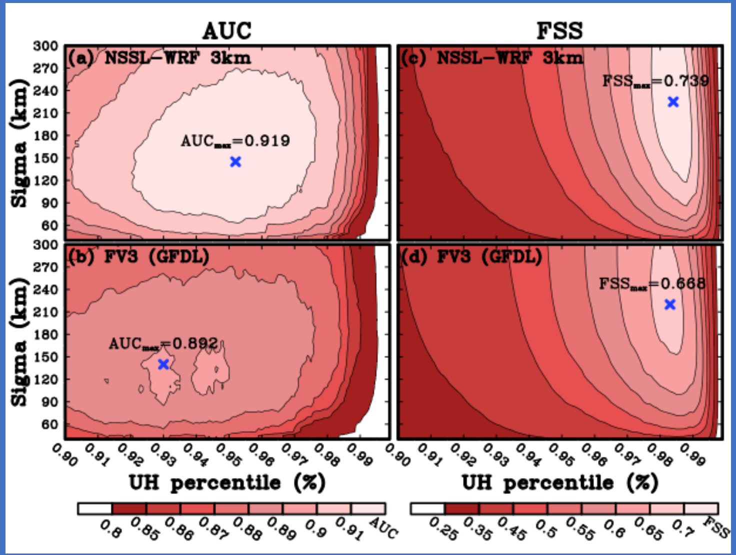
- Results are more mixed
- UH calculation is very different on FV3 grid than B- or C-grid
- Would expect larger UH on FV3 due to grid-staggering structure.
- Need to calibrate UH using surrogate severe method (done for all models)

UH Surrogate Severe: ARW vs FV3-G results for HWT 2017

Sobash Surrogate Severe Method

- UH percentiles examined for fair comparisons
- Used to evaluate and “calibrate” each model’s UH climatology with observed severe weather reports each day.
- **NSSL-WRF 3km is still better**
  - **higher FSS and AUC**
  - **differences are significant**

	MAX AUC	MAX FSS
FV3	0.89	0.67
NSSL-WRF	***0.92***	***0.74***



Discuss…

Questions to keep in mind…

1) A new implementation of the HRRR model is imminent and this popular WRF-ARW-based deterministic modeling system has been optimized for over a decade, but it is slated to be subsumed by an FV3-based CAM-scale forecasting system as soon as the change can be justified by performance. In what different contexts/applications have FV3-based CAMs been tested to date and what, in your opinion do the results indicate about the capabilities of the FV3 dynamic core in a CAM framework?

2) Do we have sufficient "evidence" to reach a consensus that FV3CAM is "in the same class" as the ARW framework for CAM-scale prediction?

3) Do we have sufficient "evidence" to reach a consensus that FV3CAM is NOT "in the same class" as the ARW framework for CAM-scale prediction?

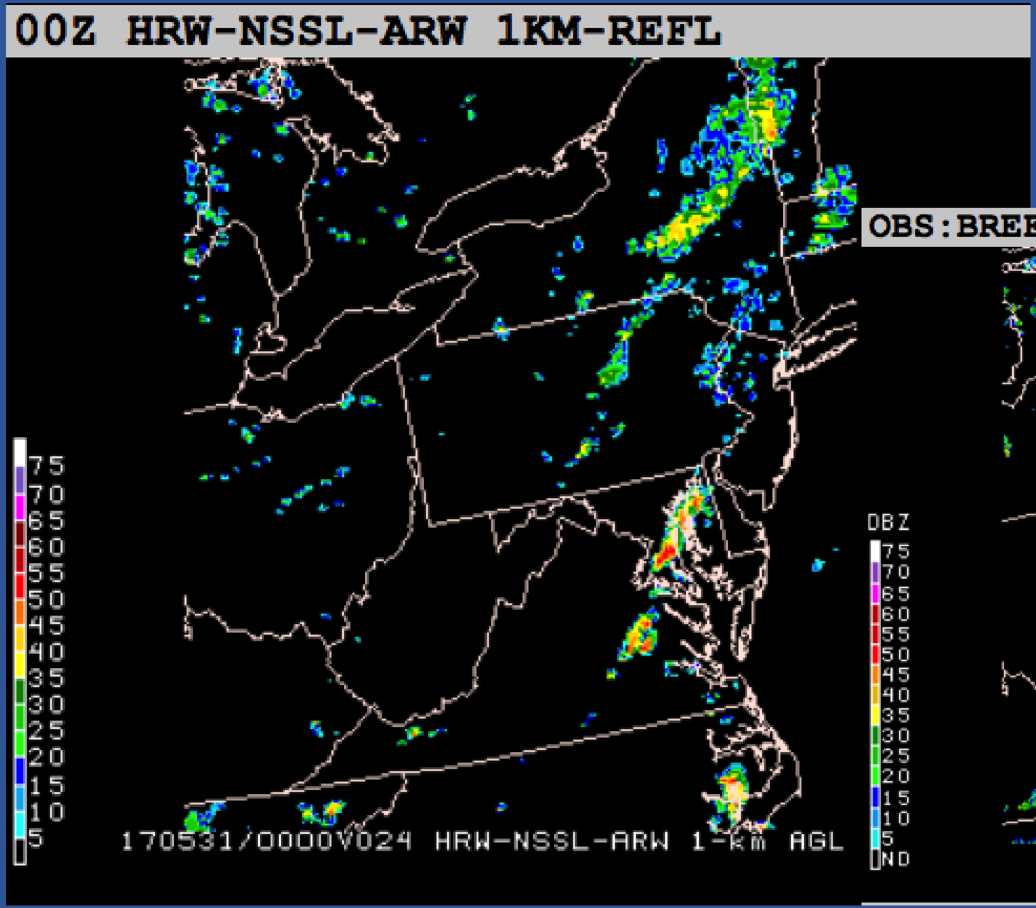
4) If additional evidence is needed to reach consensus on either question, what additional testing/metrics are required to provide sufficient evidence?

5) If the FV3 numerical method fundamentally handicaps CAM-scale prediction, can it be modified/improved?

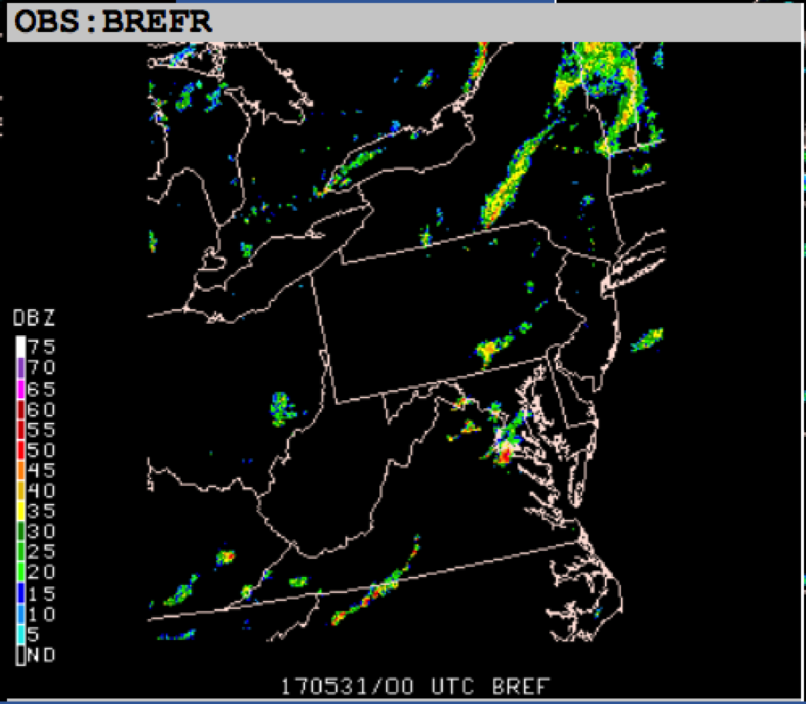
30 May DC storms

24 hour FCSTS

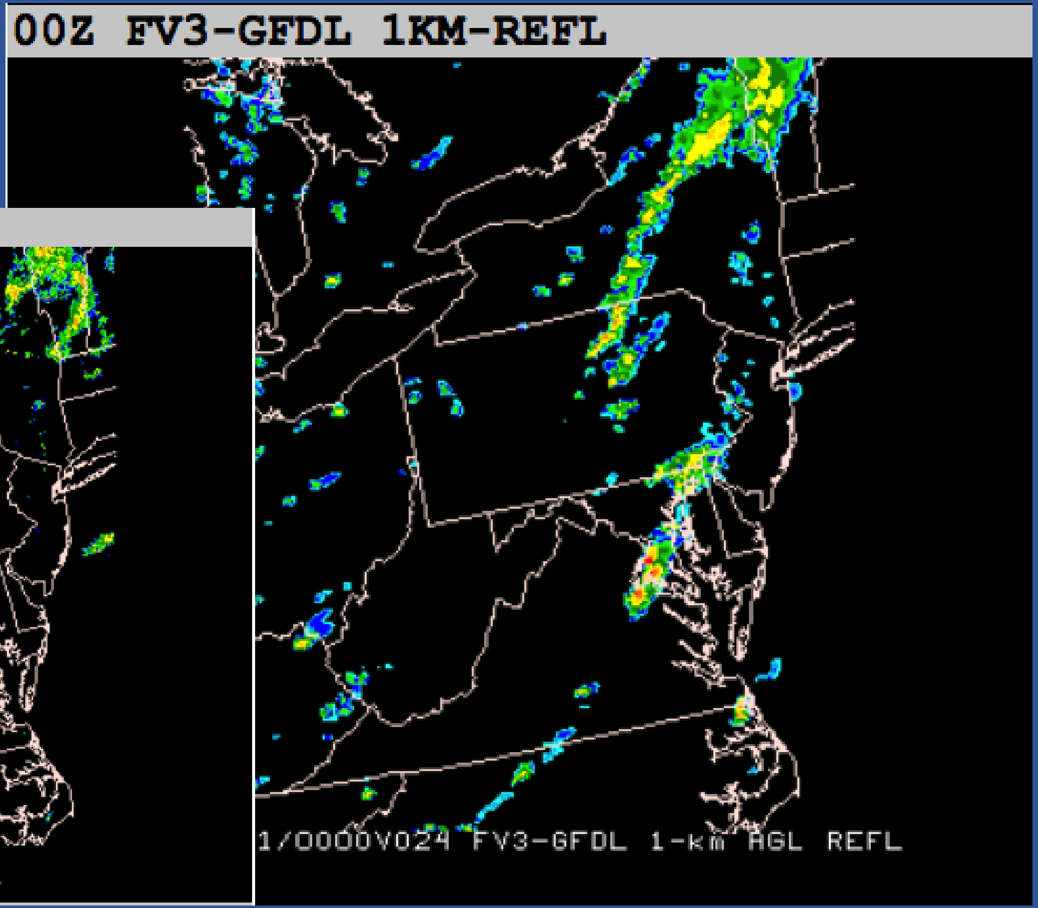
HRRR: Pretty good forecast



Observations

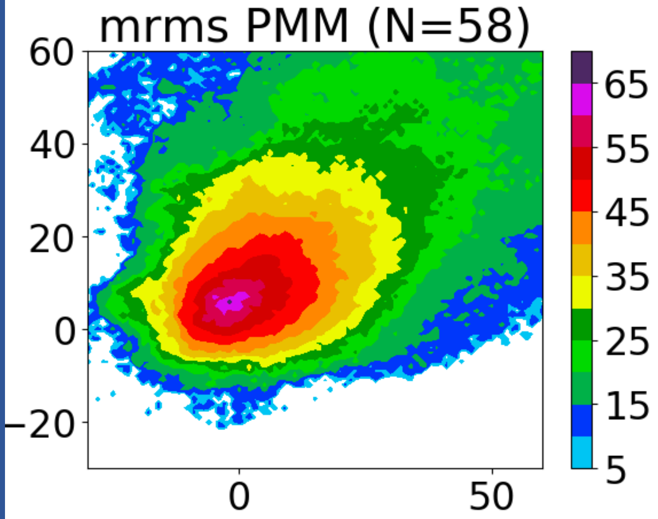


FV3-G Pretty good forecast

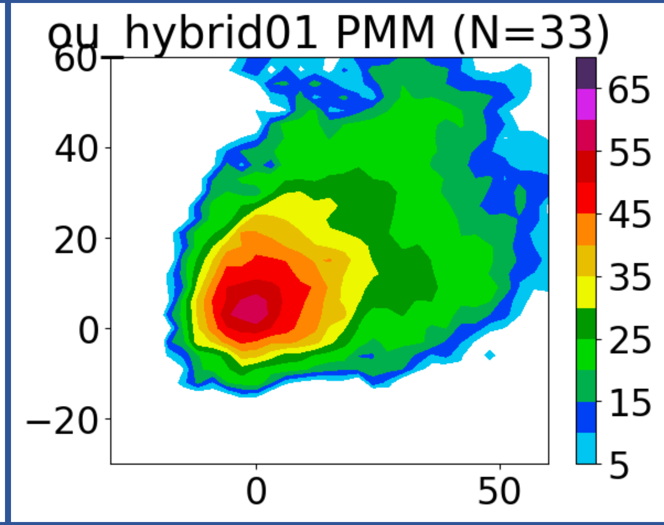


Prob Matched Mean composite dBZ (preliminary)

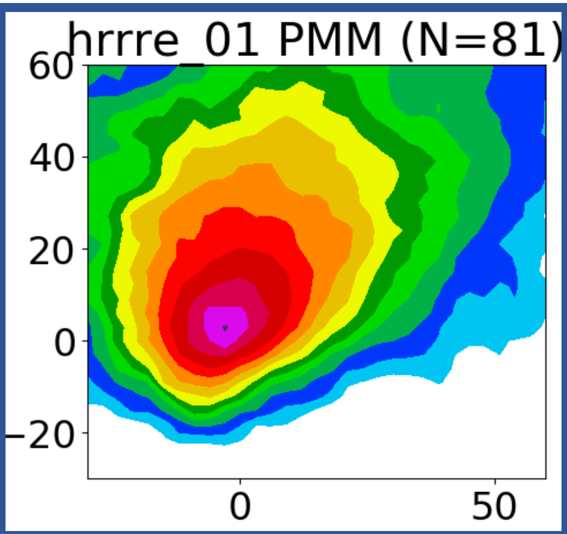
MRMS



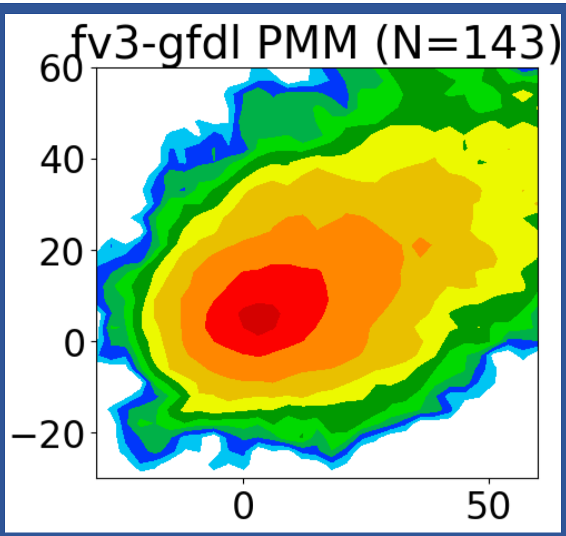
OU (NMMB)



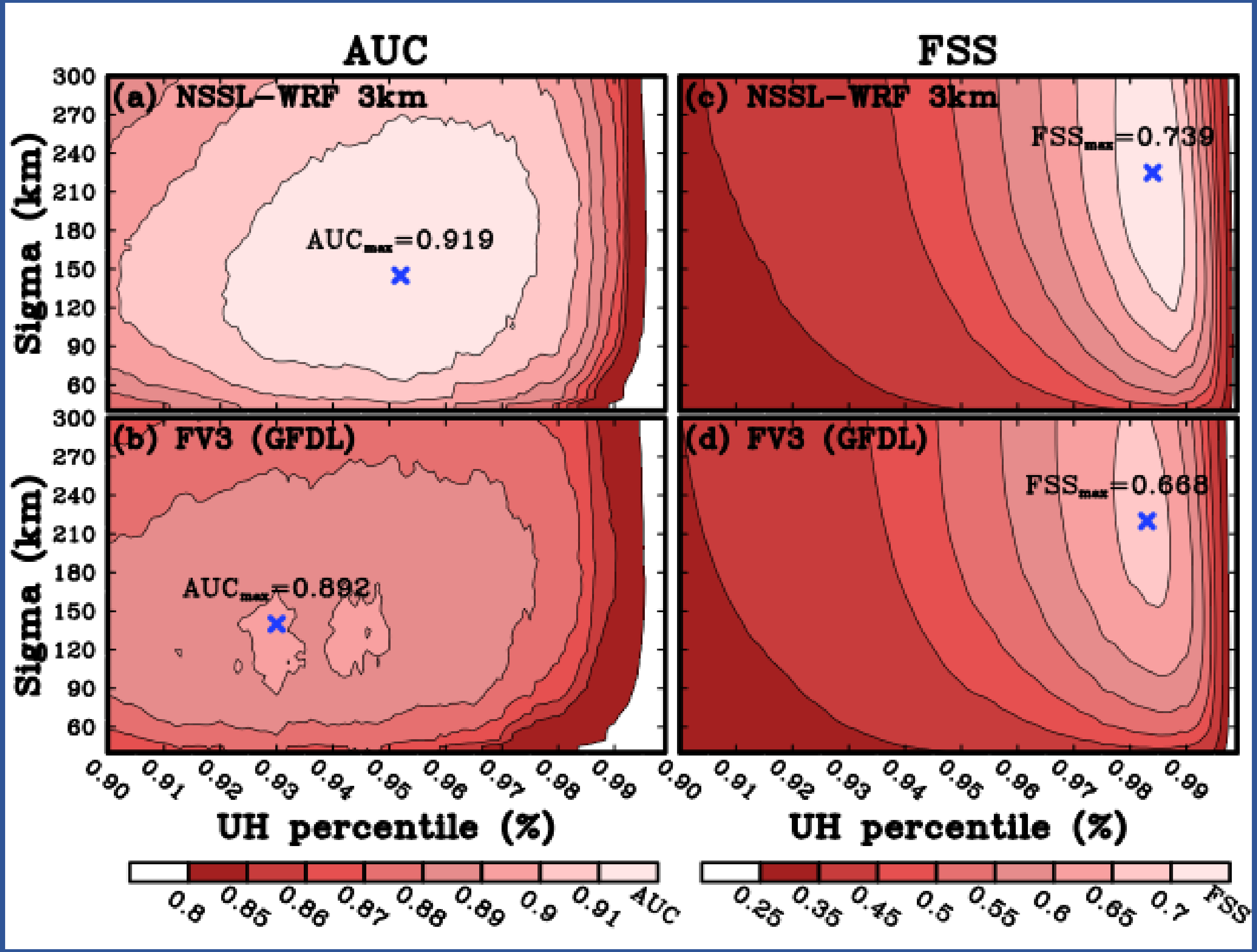
HRRRE (ARW)



FV3-GFDL



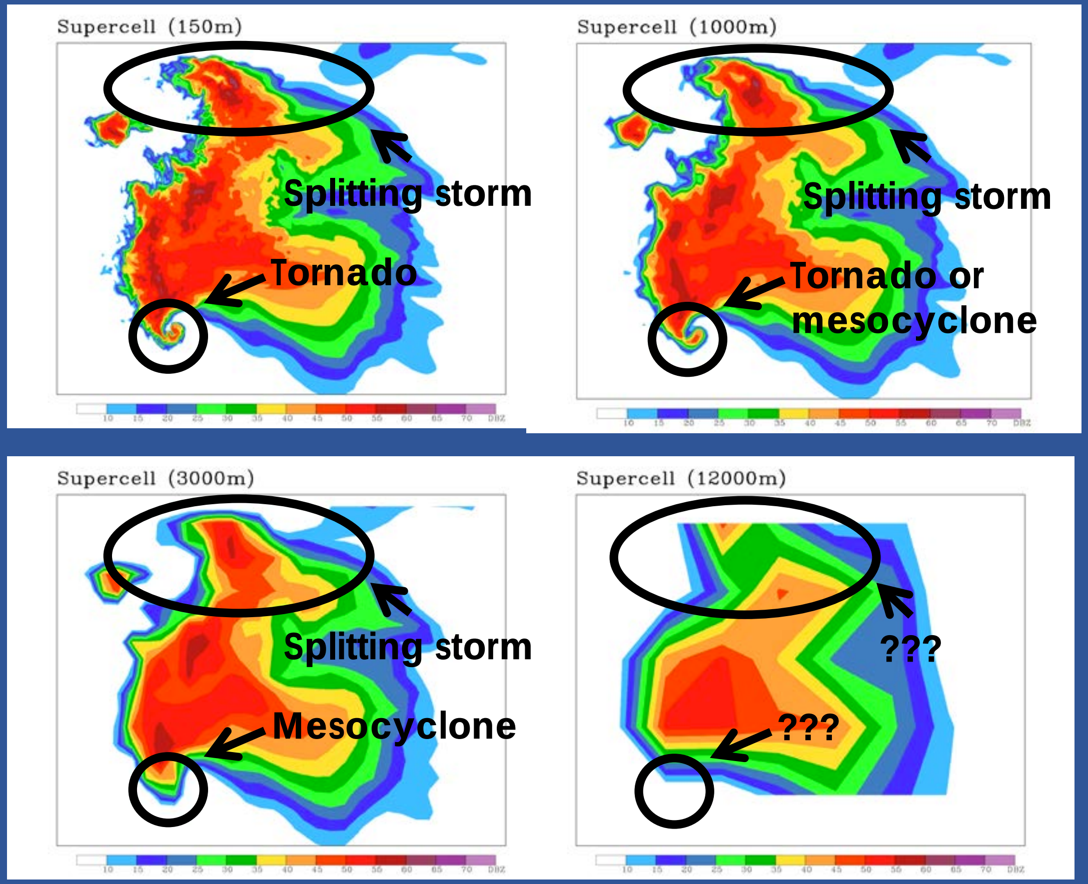
Surrogate Severe: ARW vs FV3-G results for HWT 2017



<u>Surrogate Severe Method:</u>

- Methods follow Sobash et al. papers
- UH percentiles examined for fair comparisons
- Used to "calibrate" model's UH climatology with observed severe weather reports each day.
- **NSSL-WRF 3km is clear winner**
  - **higher FSS and AUC**
  - **differences are significant**

Major breakthrough in HWT: Use of "storm-scale" models to forecast severe weather (2003-04)



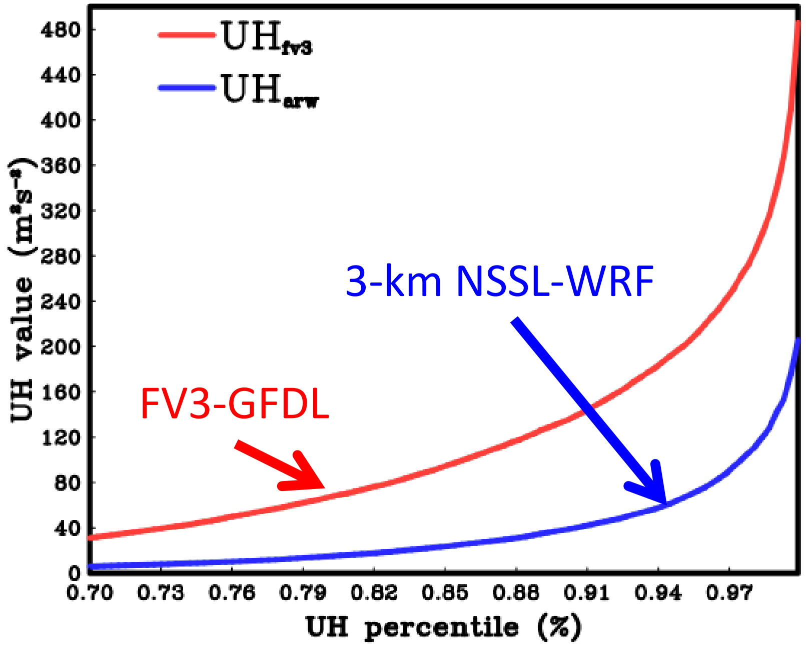
Numerical supercell simulations

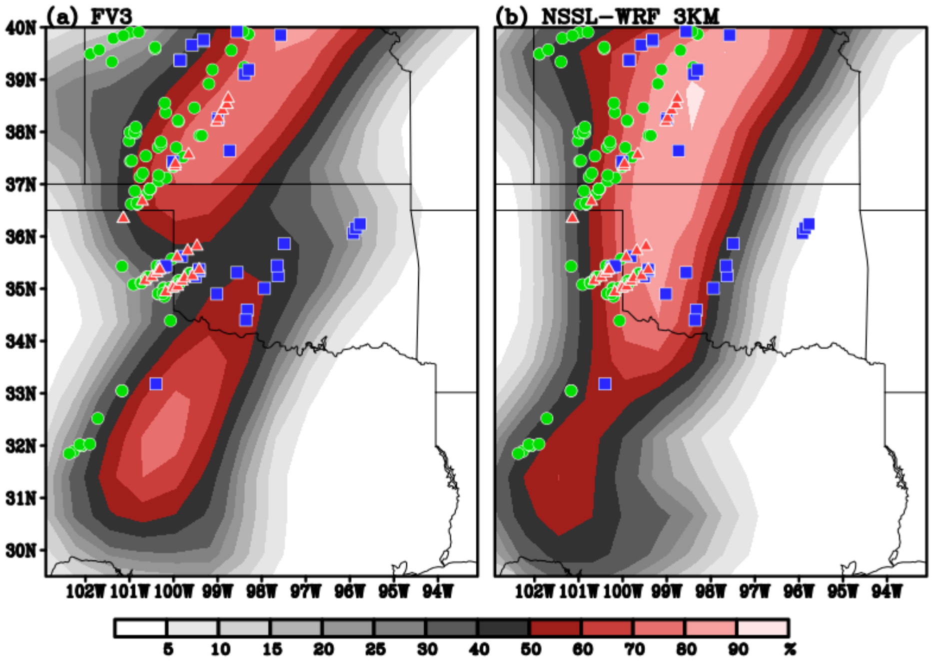
- 150m: Only possible in research setting
- 1000m: Operational models in 5-10 years
- 3000m: Current short-term forecast model
- 12000m: Highest resolution ~ five years ago.
- **Storm scale models are a major shift from "ingredients-based" forecasting.**

CLUE Results: FV3-GFDL (2017)

- Surrogate severe method (Sobash et al. 2011, 2016) used to compared FV3-GFDL and 3-km NSSL-WRF

- UH remapped to 80-km grid (24-h max in 80-km box)
- A range of UH percentiles tested as severe weather "surrogates. Percentiles must be used because UH climatologies are very different!
- To generate severe weather probabilities, a Gaussian with several σ tested.





- Example: 24 h maximum UH for 16 May 2017 (FV3 left, NSSL-WRF right) with reports overlaid.
- Higher, more widespread, and noisier appearance of UH in FV3.

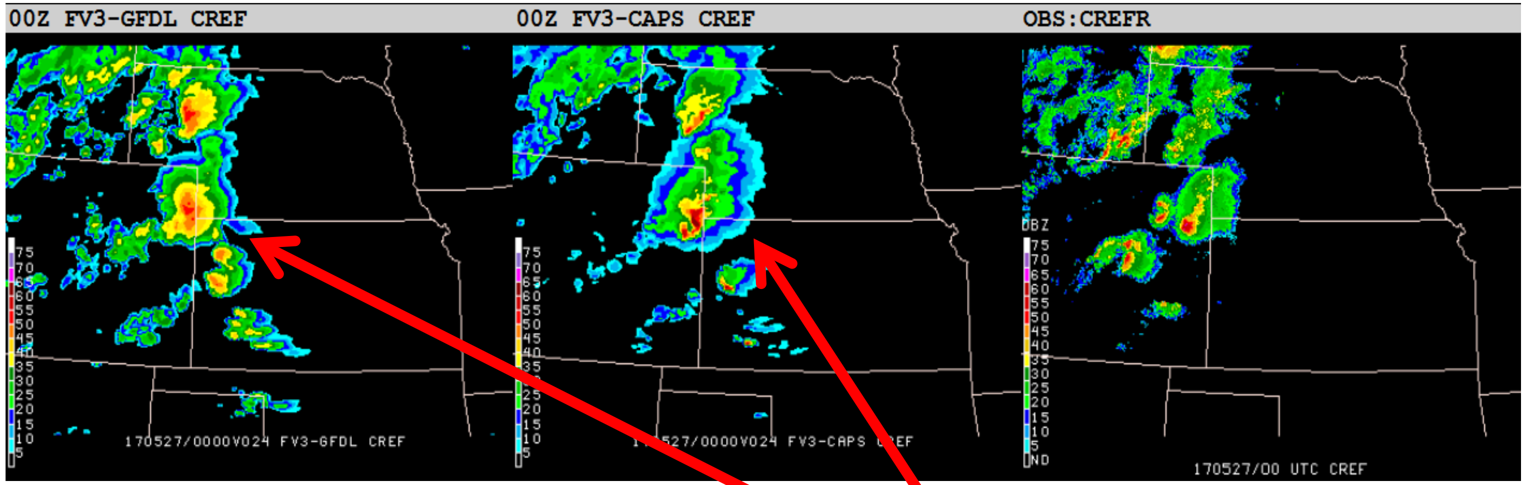
- Severe weather probs (p=95%, σ=75km), severe weather probs (p=95%, σ=75km) with reports overlaid.

Aggregate Surrogate Severe results

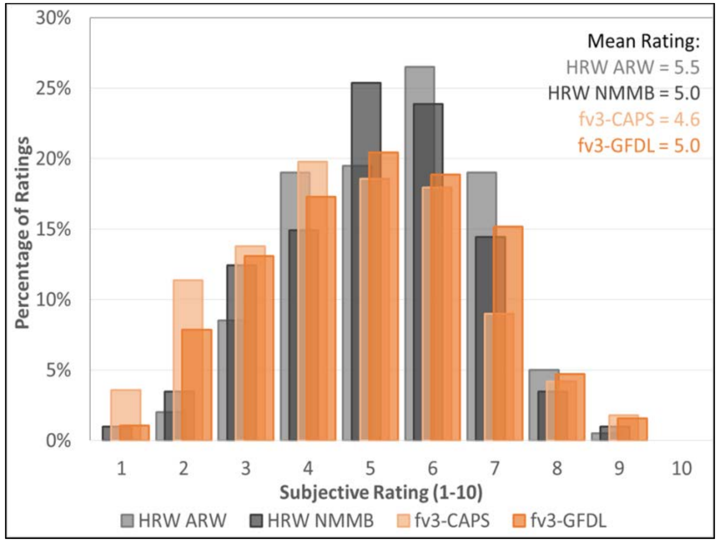
	MAX AUC	MAX FSS
FV3	0.89	0.67
NSSL-WRF	***0.92***	***0.74***

CLUE Results: FV3 (2017)

Subjective results



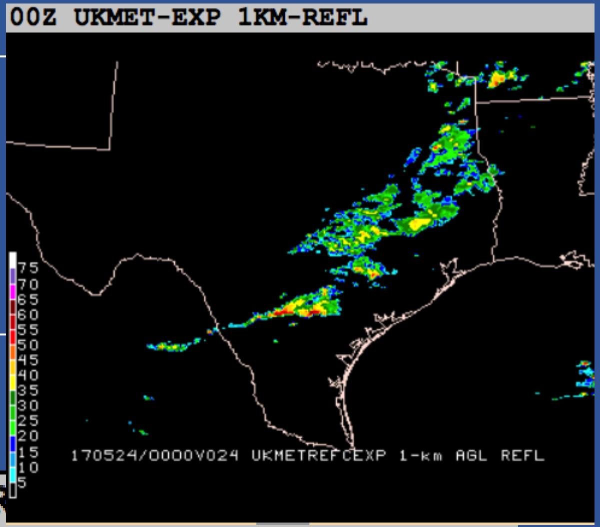
Example of subjective comparison plots used for rating CAM performance at convective scales. The left panel shows 24-h forecast of composite reflectivity of the FV3-GFDL, the middle panel shows the 24-h forecast of composite reflectivity of the FV3-CAPS, and the right panel shows the observed composite reflectivity at 0000 UTC on 27 May 2017.

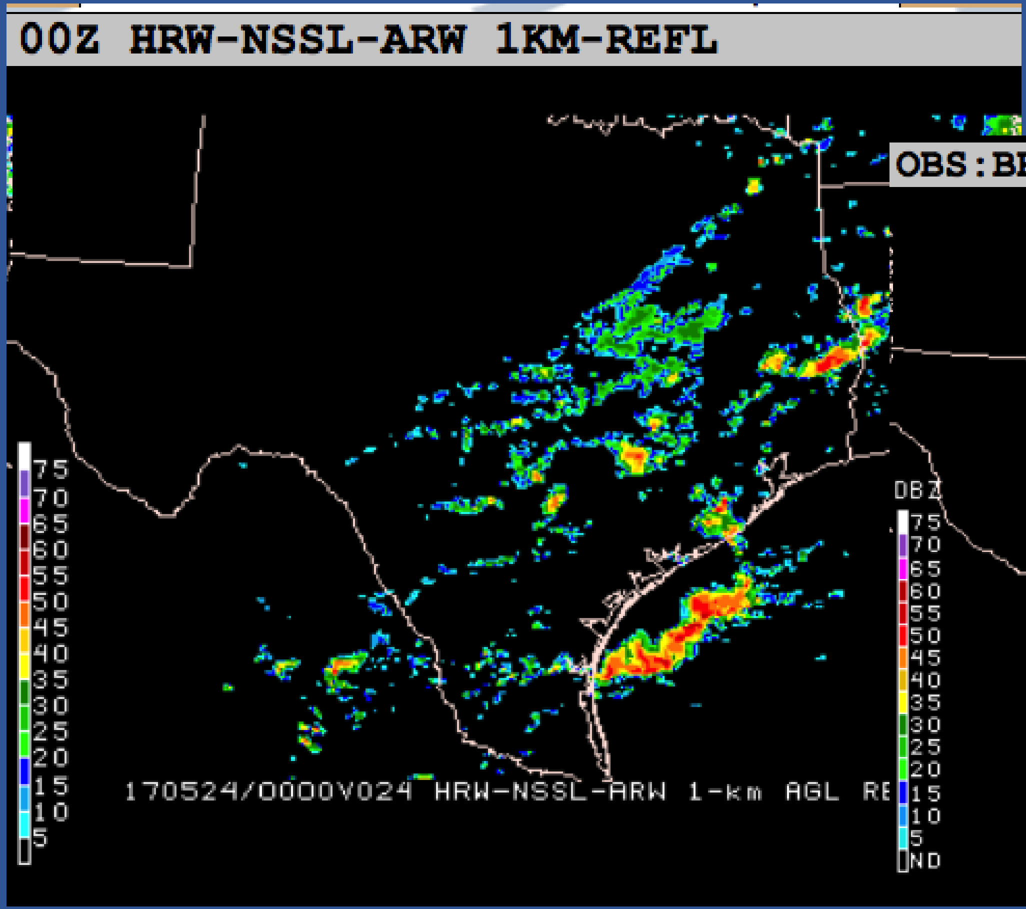


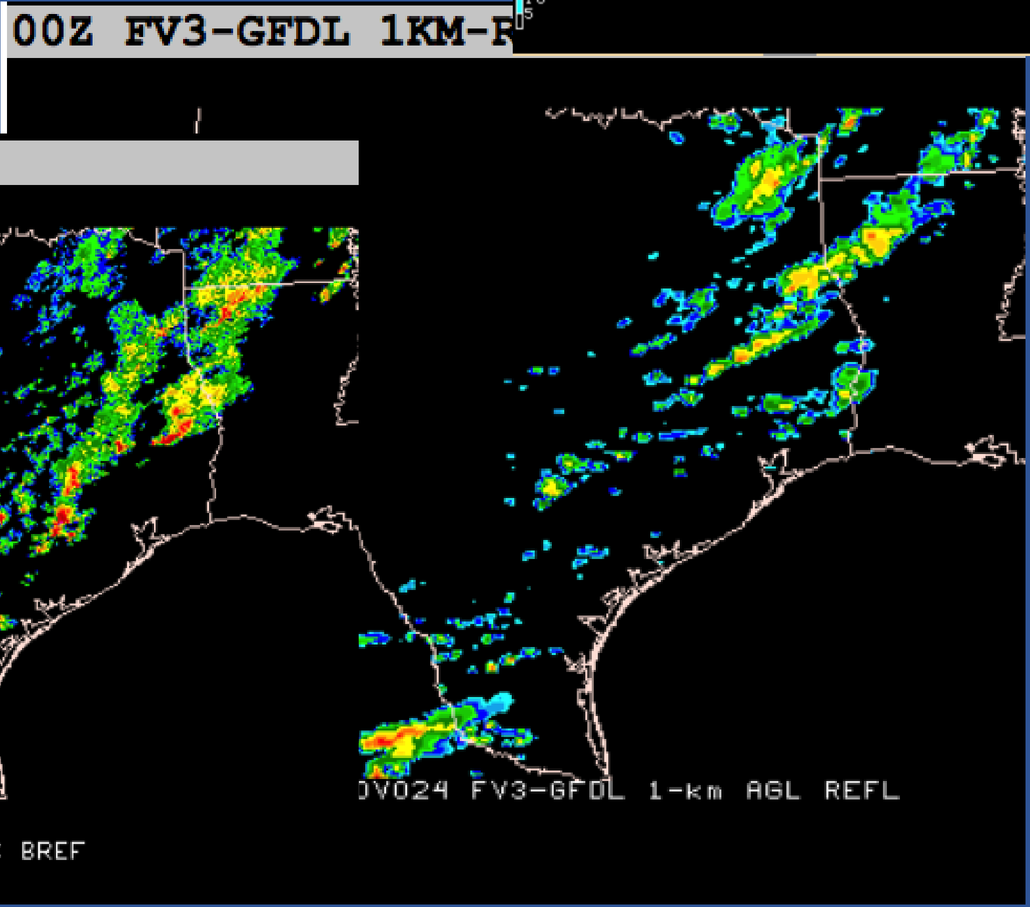
Note the different character in simulated reflectivity – left uses GFDL microphysics, right uses Thompson.

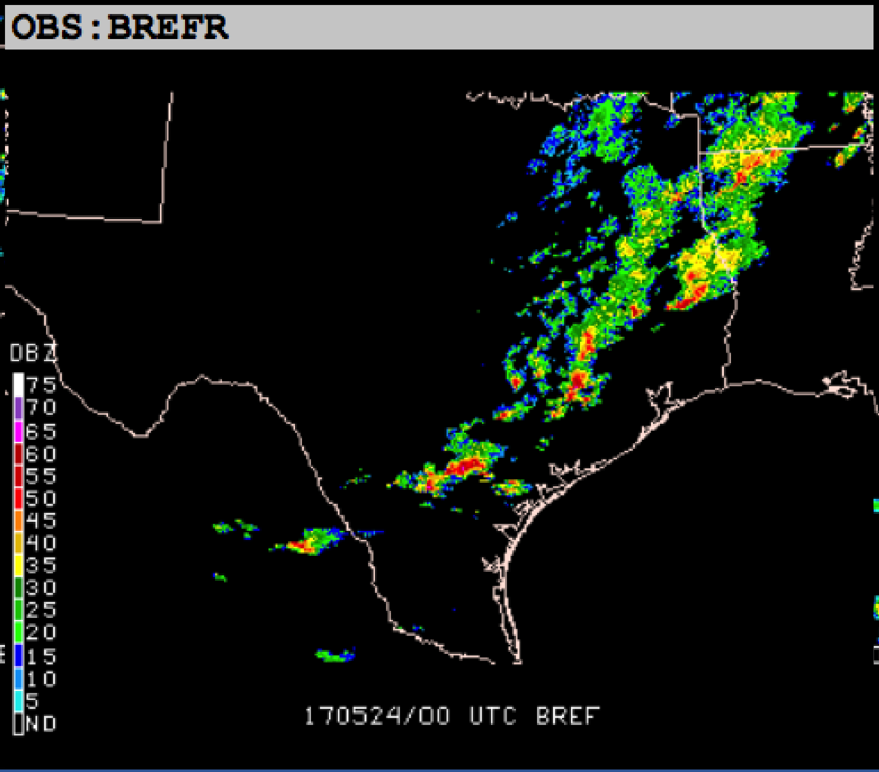
According to SFE participant ratings, FV3 is competitive with operational CAMs

23 May 2017 TX Supercells (svr winds)









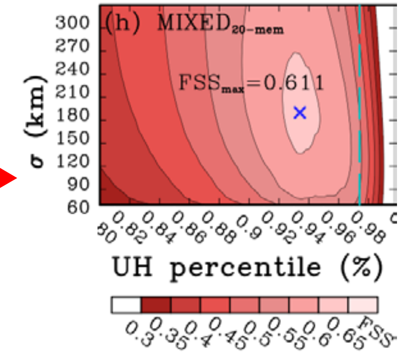
Challenges with FV3: NSSL Experiences

- Our senior computational scientist experience…
- OAR funded project to be running "NSSL-FV3" Global+CAM starting 1 Dec 2017.
- Many challenges…inconsistent performance, resource issues, etc.
  - GFDL package took about 4080 seconds on Jet;
  - EMC package took about **35349** seconds on Jet;
  - **Hot off the PRESS: NSSL machine: EMC version: 4243 secs GFDL version: 3664 secs** 
- Knowledge base still confined to a few people
- Software not portable enough: must run on wide range of systems to leverage community
  - Either NEMS + FV3 must be made workable on non-NOAA systems or
  - FV3 EMC core needs to be placed in a simpler framework
  - *Where does the global versus regional FV3 version fit in all of this?*

CLUE Results: Multi-core vs. Single core (2016)

Surrogate severe (Sobash et al. 2011, 2016) used for severe weather verification

- UH remapped to 80-km grid
- Severe weather probabilities derived using different UH %iles and smoothing.
- 100 %iles and 53 σs tested (5300 sets of probs)
- For QPF, find obs rainfall threshold that gives same frequency as severe weather (P = 2.71 inches).



Surrogate severe – max scores

	AUC	FSS
ARW	.904	.595
NMMB	.901	.600
Mixed-10mem	***.908***	***.615***
Mixed-20mem	***.908***	.611

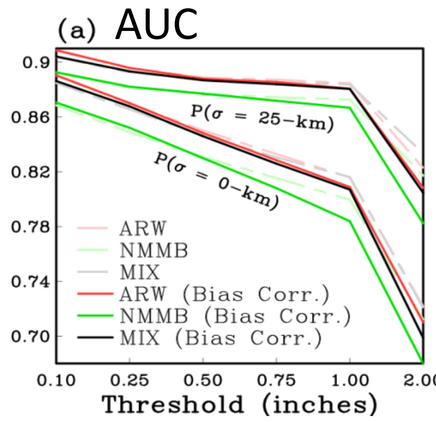
Surrogate QPF – max scores

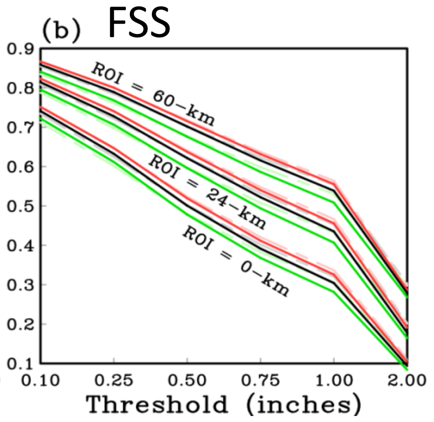
	AUC	FSS
ARW	.931	.677
NMMB	.922	.643
Mixed-10mem	***.935***	***.686***
Mixed-20mem	.932	.685

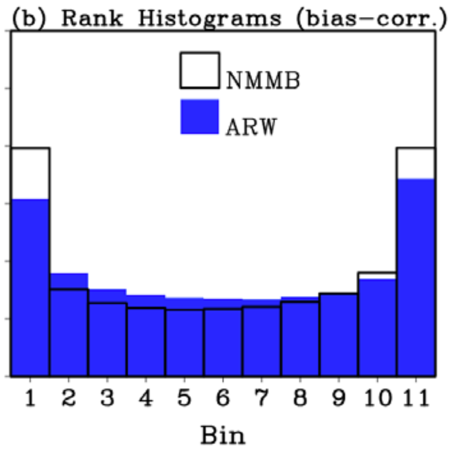
Main results – (will appear in a 2018 BAMS article)

- Differences among subsets not significant
- Multi-core slightly better than single core
- 20 mems gives no added benefit

QPF results







Main QPF results

- Single core equal or better than multi-core.
- ARW significantly better than NMMB.
- ARW rank histograms (bias-corrected) flatter than NMMB (better spread).