Evaluation of FV3-SAR Initialized by Multiscale Hybrid EnVar Analyses for Convection Allowing Hazardous Weather Forecasting

Nicholas Gasperoni and Xuguang Wang
Multiscale data Assimilation and Predictability (MAP) Lab
University of Oklahoma, Norman, OK, USA

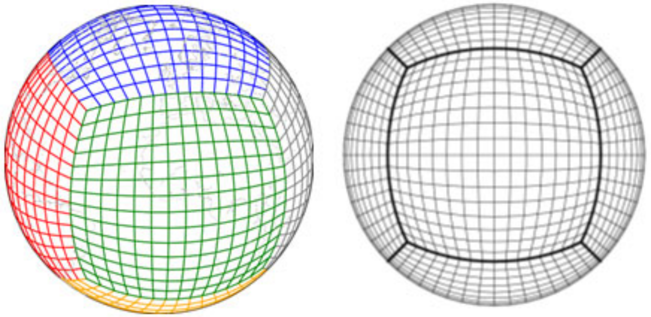
Curtis Alexander
GSD/ESRL/NOAA, Boulder, CO, USA

Jacob Carley
NOAA/NCEP/EMC, College Park, MD, USA

10th Conference on Transition of Research to Operations
100th AMS Annual Meeting, Boston, MA
Jan. 13, 2020 (Paper 1.5)

Introduction

- The Finite Volume Cubed Sphere Model (FV3) has recently replaced the GFS as the new dynamical core within NOAA's Next Generation Global Prediction System (NGGPS).
- FV3 has become the foundation for a unified modeling framework at NOAA for applications across all time and spatial scales
- A Stand Alone Regional (SAR) FV3 capability is being developed to facilitate regional applications, e.g., convective-scale, frequently-updating data assimilation cycles
- The FV3 will eventually replace the dynamical cores for regional model and analysis systems such as the North American Mesoscale Model (NAM) and the High-Resolution Rapid Refresh (HRRR) model.

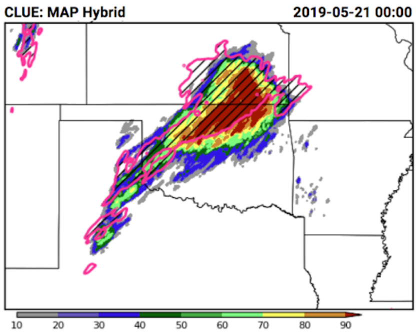


Motivation

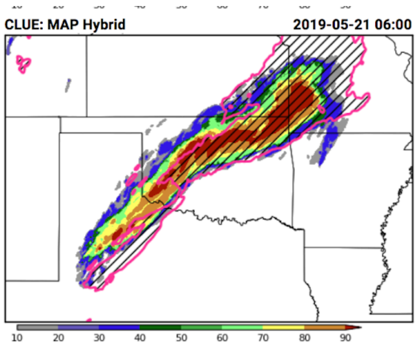
- A community workflow was created at DTC to facilitate a collaborative effort for testing FV3SAR.
- Given ongoing developments, the FV3SAR will soon be integrated with the operational convective-scale Gridpoint Statistical Interpolation (GSI)-based hybrid ensemble-variational (EnVar) data assimilation (DA) system through collaboration between NOAA EMC, GSD, and the OU MAP laboratory.
- The regional FV3SAR has not been systematically evaluated against other regional models for CAM applications.
- Here, we leverage the OU-MAP real-time EnVar analyses from the 2019 HWT Spring Forecast Experiment (SFE) to initialize FV3SAR
- **Primary objectives:**
  - **(1) Learn CAM characteristics of the FV3SAR**
  - **(2) Evaluate the performance of FV3 dynamical core against the WRF-ARW core from the SFE**
  - **(3) Tuning of relevant dynamical (and physics) namelist parameters which may lead to improvements in FV3SAR performance at CAM resolution**

OU-MAP Ensemble for HWT SFE 2019

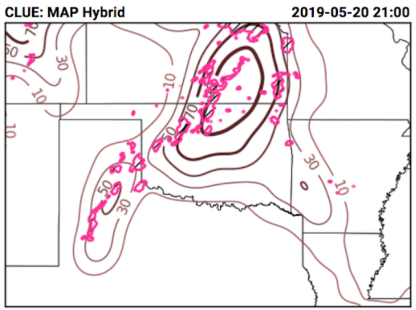
- Since 2017, OU's MAP lab has participated in the HWT SFE's Community Leveraged Unified Ensemble (CLUE), running real-time 10-member 36-hr convection-allowing forecasts over the CONUS.
- The ensemble was initialized from a full multiscale GSI-based 3DEnVar system, including radar reflectivity and conventional obs – the GSI system has been extended for convective-scale DA by OU MAP lab Johnson et al. (2015) and Wang and Wang (2017)
  - 6 hours of conventional DA prior to forecast initialization
  - 20-min cycling of radar DA for final hour
  - 40 ensemble members, GFS control analysis
  - 10-member ensemble recentered around control prior to forecast initialization
- Ensemble performed subjectively well during the 2019 experiment
  - Location and extent of precipitating systems often in line with reality, good CI prediction in many cases
  - Some underforecasting bias of UH – related to stricter temperature tendency limiter applied
  - Example from 20 May 2019 (high risk case) on right



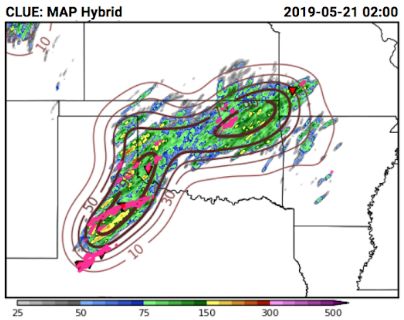
24-hr fcst 6-hr P(QPF) > 1 in. (color) and MRMS 1-in obs



36-hr fcst 6-hr P(QPF) > 1 in. (color) and MRMS 1-in obs



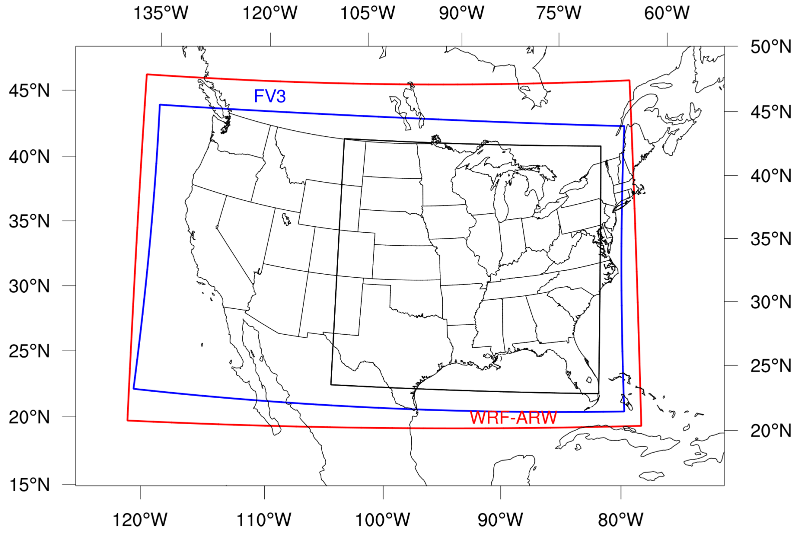
21-hr fcst NEP(CREF) > 40 dBZ (brown) with MRMS 40-dBZ CREF obs (magenta)



26-hr fcst 4-hr UH ens. max (color) with MRMS 4-hr MESH 1i n.

Experiment Design: Ensemble FV3 Forecast

- Goal: objective, fair comparison of FV3 and ARW forecasts
  - want to attribute differences primarily to dynamical cores as much as possible
- Choose 10 active severe weather cases from 2019 SFE.
- Ran with HRRRv4 physics implementation into the FV3 (via Common Community Physic Package, CCPP)
- ICs: Recentered final analysis from EnVar (HWT) – interpolate 10 members (via chgres_cube2) to the FV3-SAR grid for ICs, surface data, and LBC at hour 0.
  - Pro: Ability to study how FV3 performs with ICs from fully-sampled multiscale ensemble analysis
  - Con: The final analysis taken from EnVar cycling with WRF-HRRR model core, inconsistency of ICs with model core (FV3)
- LBCs: Use EnVar-initialized 10-mem WRF forecasts from SFEs
  - Pro: Consistency with EnVar analysis (as opposed to using LBCs from another source)
  - Con: Inconsistency of dynamical core, and consistency with EnVar may not be as important for longer forecast hours
  - Applied 10 rows of model blending (*thanks to Tom Black for providing us early access*!)



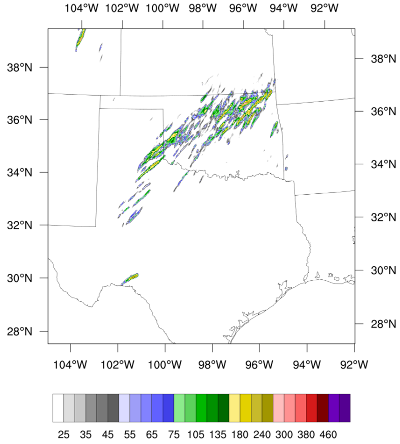
- Verification done over eastern CONUS (black box) to avoid complex terrain

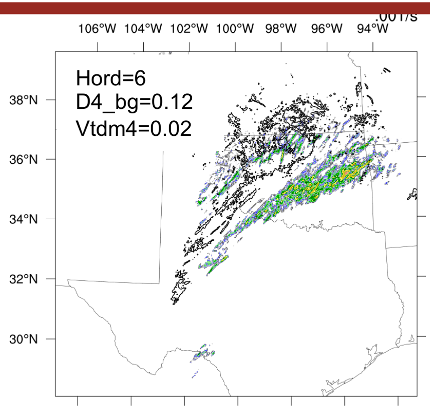
Choosing a Baseline FV3-SAR Configuration

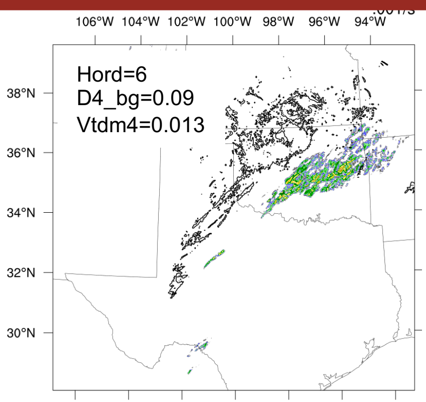
- Ran sensitivity forecasts on May 20, 2019, testing values chosen for dynamics parameters:
  - hord: horizontal advection method – 5 (fastest, more accurate; less diffusive, but more unstable) vs. 6 (monotonicity constraint; more diffusive, stable)
  - d4_bg and vtdm4: Coefficients for divergence and vorticity damping
- Evaluated composite reflectivity, 2-5km max 4-hourly updraft helicity, 4-hourly max updraft velocity
- Compare subjectively to “baseline” WRF-HRRRv4
  - WRF is well-tuned and well-research model

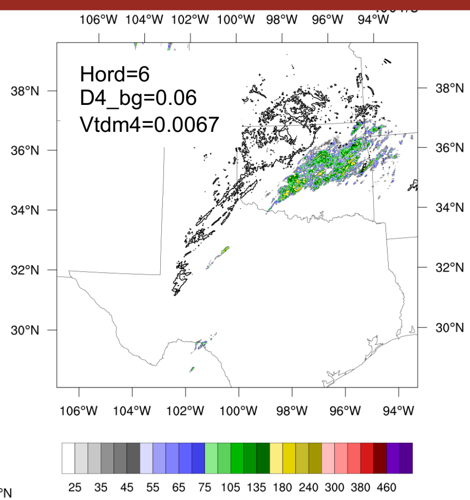
4-hr-mx Updraft Helicity (m2/s2)
Valid 23Z May 20, 2019 (fh23)

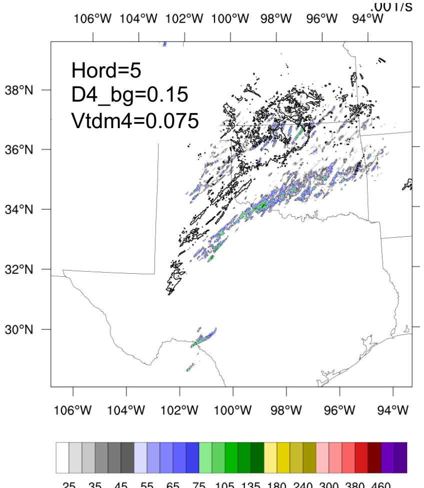
WRF HRRRv4

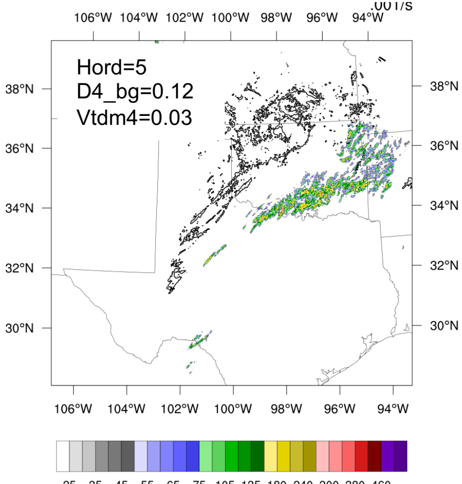








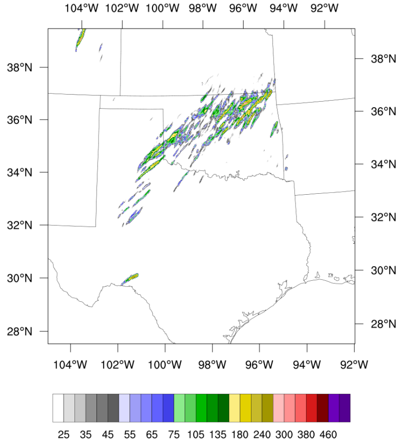


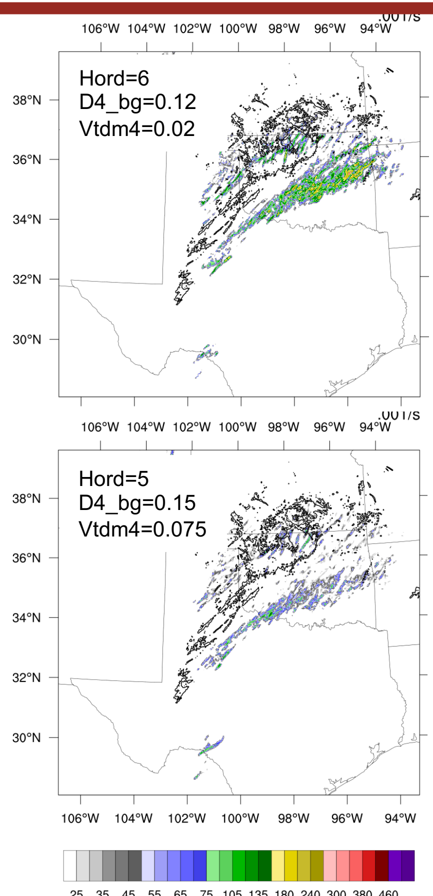


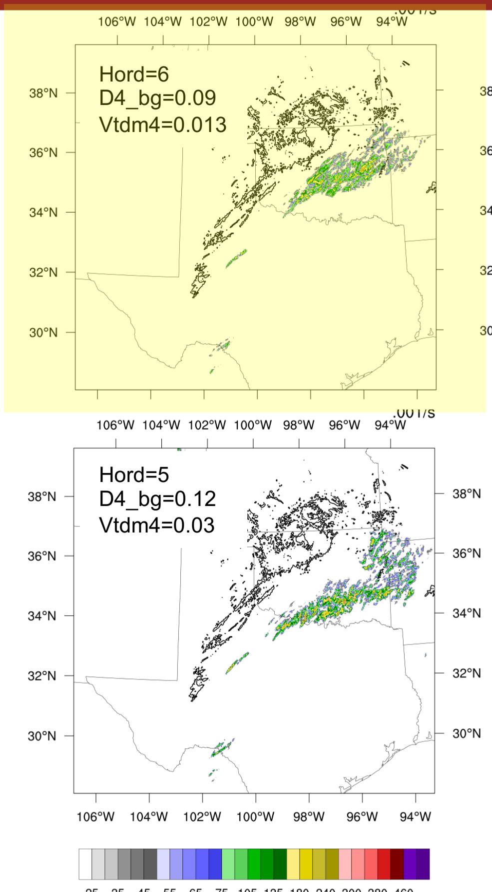
- More diffusive hord=6 has good structure, magnitudes comparable with HRRRv4
- Decreasing damping does not show notable impacts on magnitudes
- Hord=5 has notable forward-propagation error of the system (more than hord=6)
  - Similar finding noted by EMC at EnKF Workshop
- Decreased damping leads to more comparable storm structure, but further propagation error

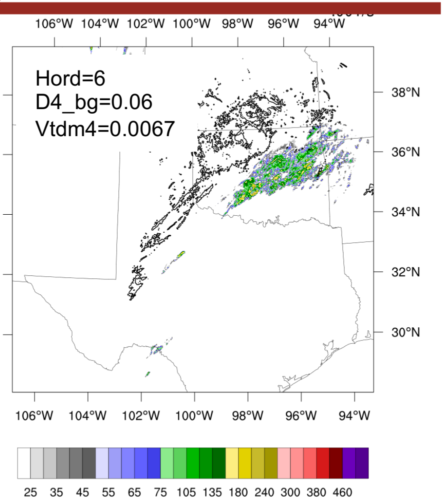
4-hr-mx Updraft Helicity (m2/s2)
Valid 23Z May 20, 2019 (fh23)

WRF HRRRv4



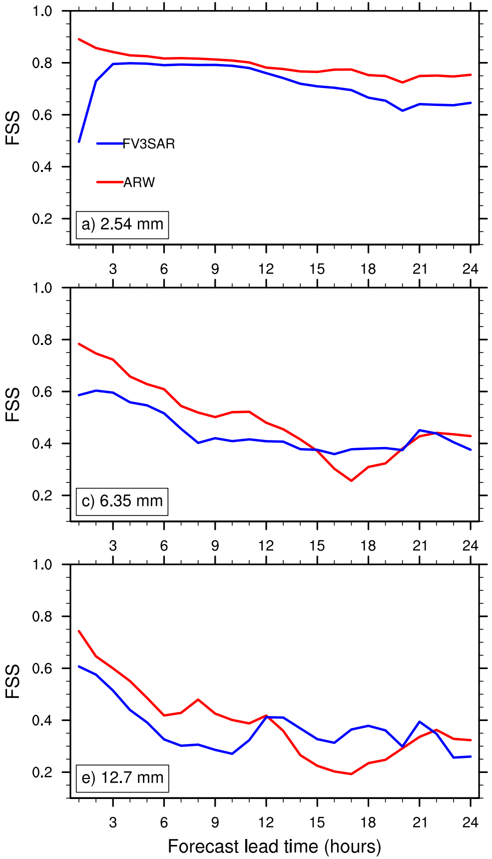




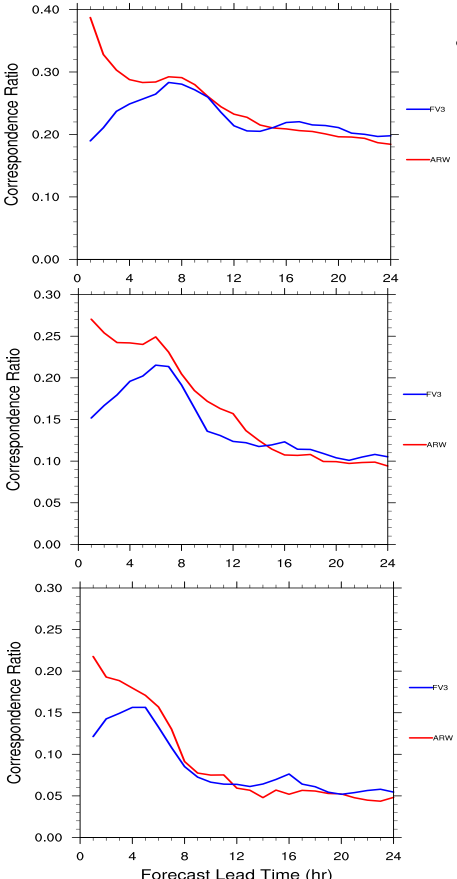


- Given these results (and other subjective considerations not shown, e.g. composite reflectivity), we chose the initial baseline configuration to be hord=6 (which has built-in diffusion), and coefficients d4_bg=0.09 and vtdm4=0.013

Preliminary 10 Case Verification of 1-h QPF FSS (left) and Correspondence Ratio (right)

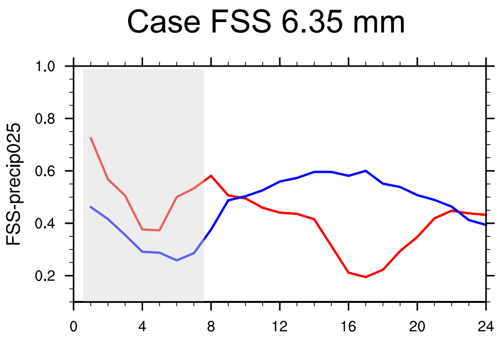


- ARW universally more skillful at light (2.54 mm) threshold
  - spinup from IC can be seen for *3-h*
- At heavier thresholds, ARW more skillful for forecast hours 1-12, then FV3 tends to be more skillful at later hours ~13-21

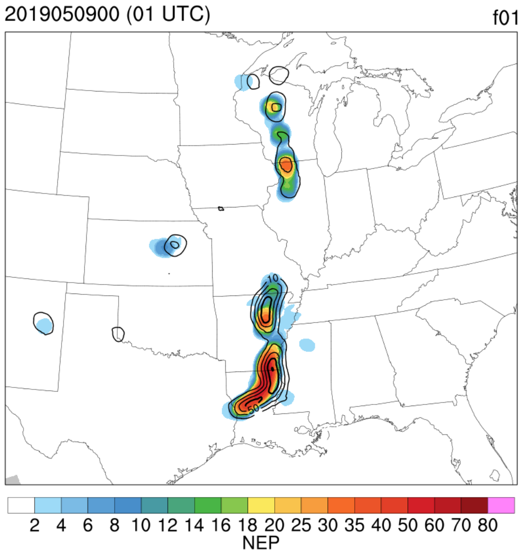


- Correspondence Ratio (CR) – simple measure of spread of precipitation systems
  - Ratio of ensemble area intersection to union at each threshold
  - *Lower values indicate less agreement – more spread*
- Spinup effects seem apparent up to *6-h* in terms of CR
- FV3 has more apparent spread than ARW early on, then slightly less at later hours
  - But with lower skill, this indicates too much spread, overdispersive
    - Indicative of spurious precip?

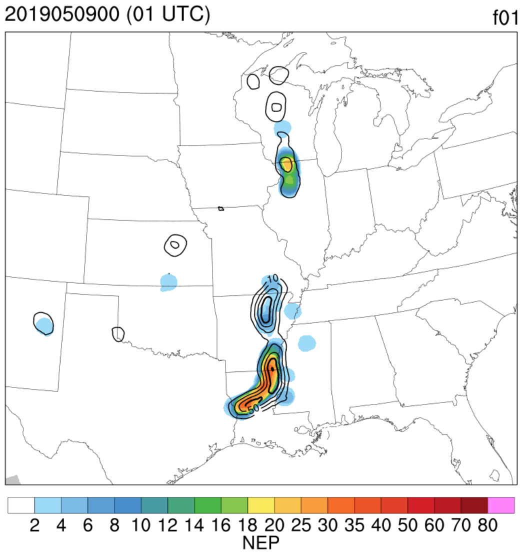
Case Example – 9 May 2019
Hours 1-7, NEP QPF > 6.35 mm



ARW



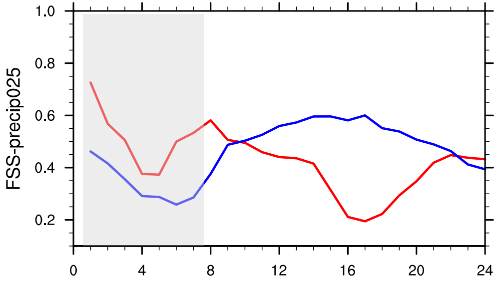
FV3SAR



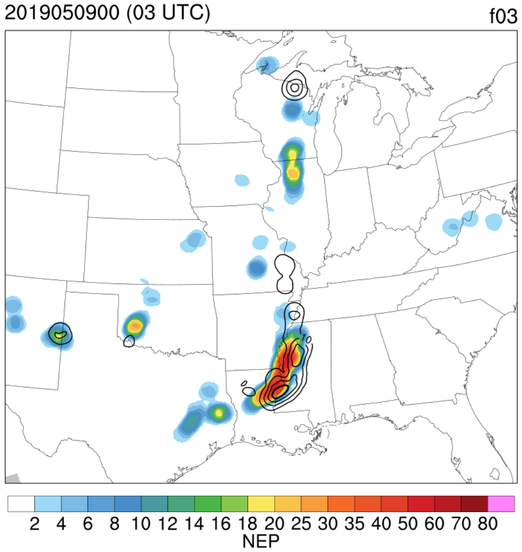
Weaker probabilities,
west bias

Case Example – 9 May 2019
Hours 1-7, NEP QPF > 6.35 mm

Case FSS 6.35 mm

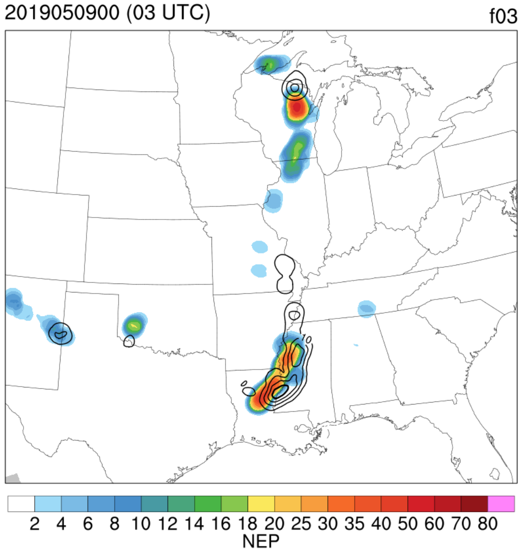


ARW



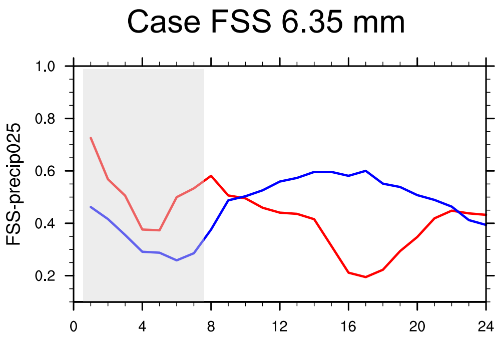
Early (1-2 h) SE TX development, but accurate location

FV3SAR

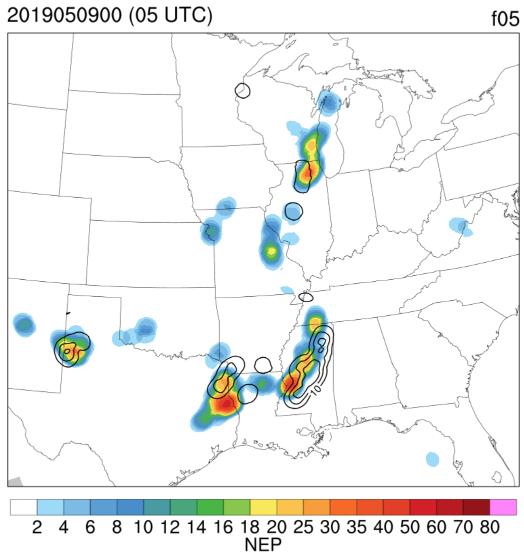


Weaker probabilities, west bias

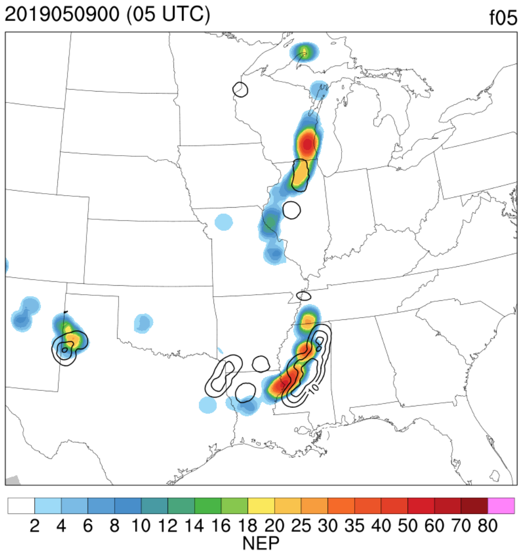
Case Example – 9 May 2019
Hours 1-7, NEP QPF > 6.35 mm



ARW

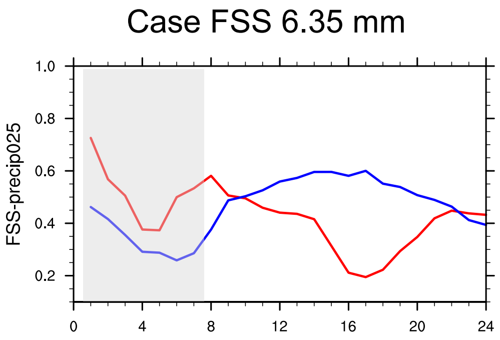


FV3SAR

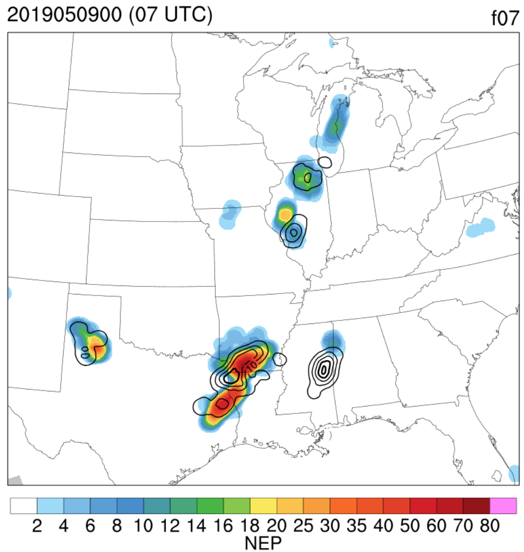


No convection in E TX

Case Example – 9 May 2019
Hours 1-7, NEP QPF > 6.35 mm

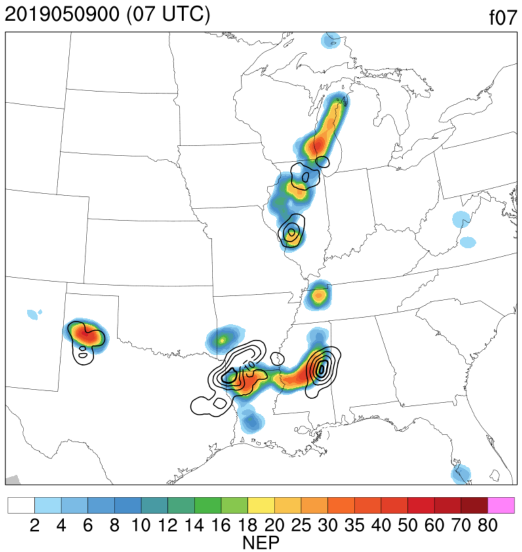


ARW



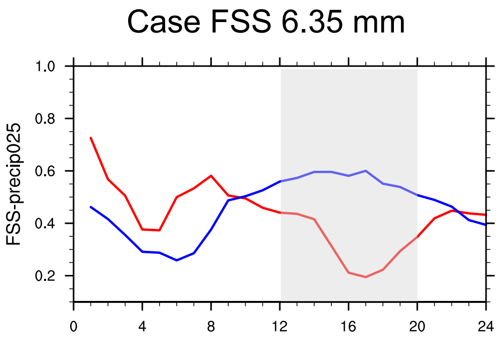
-Great storm location TX/LA
-Accurate (slightly early) decay of MCS in MS

FV3SAR

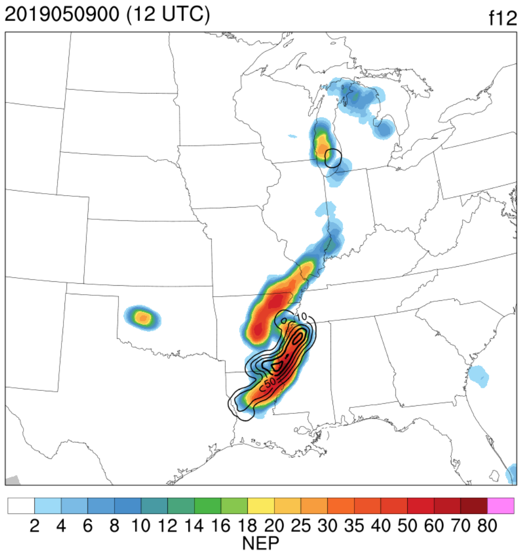


-Delayed development by 3-4 h
-MCS in MS late to decay

Case Example – 9 May 2019
Hours 12-20, NEP QPF > 6.35 mm

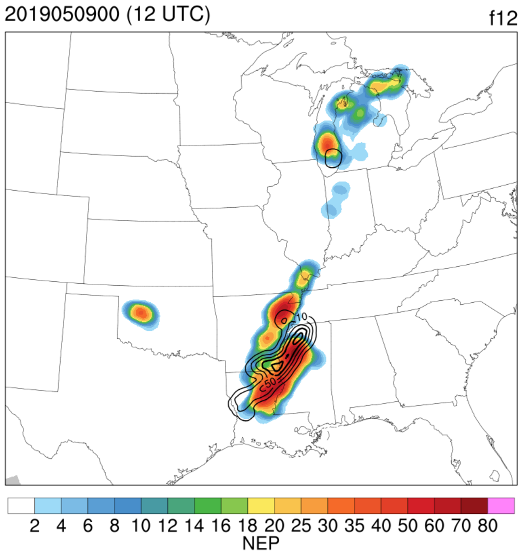


ARW



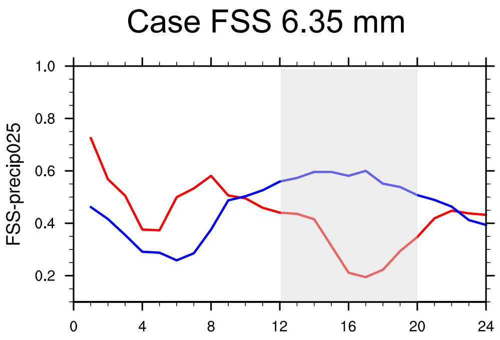
Spurious precip system in Arkansas-Missouri

FV3SAR

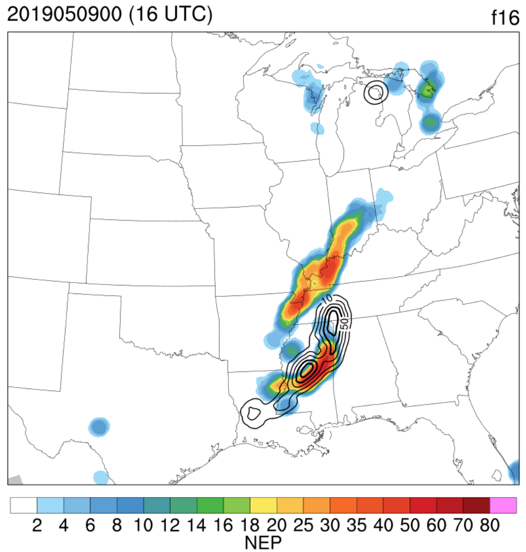


Not as much spurious precip

Case Example – 9 May 2019
Hours 12-20, NEP QPF > 6.35 mm

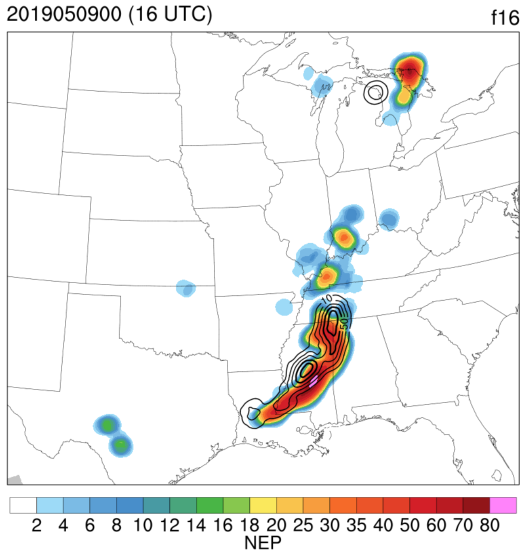


ARW



-Broad spurious precipitating system
-Extent of main line limited

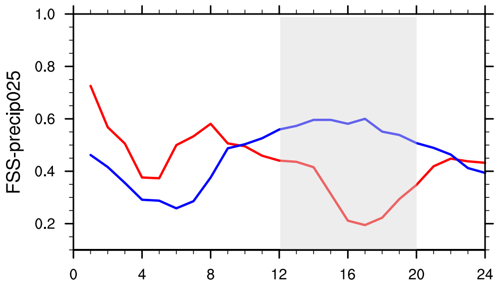
FV3SAR



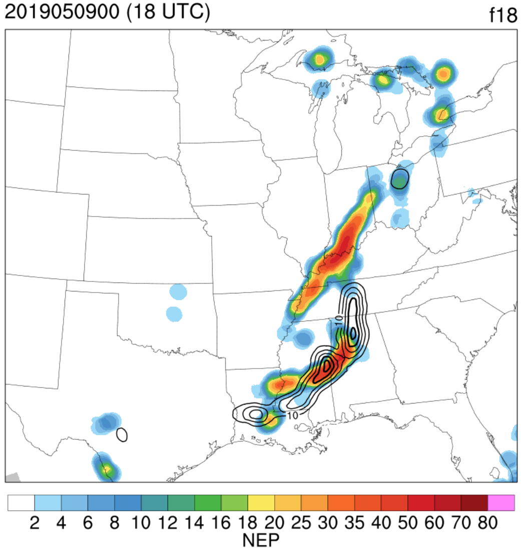
-Limited spurious precip to north
-Convective line matches length, location of reality with strong probabilities

Case Example – 9 May 2019
Hours 12-20, NEP QPF > 6.35 mm

Case FSS 6.35 mm

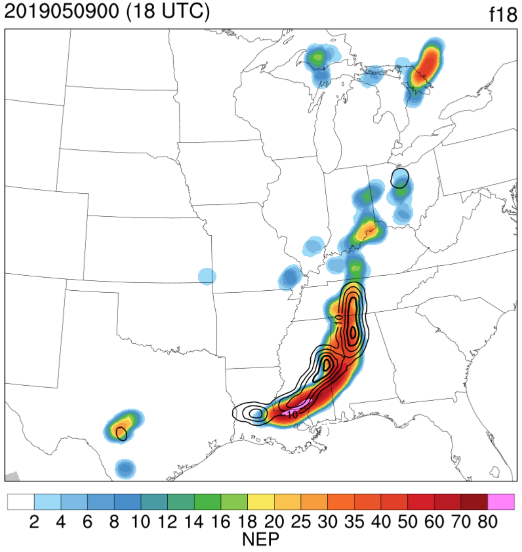


ARW



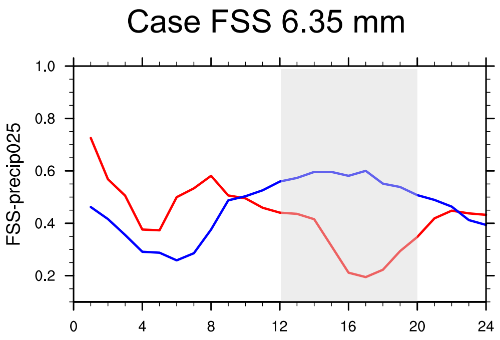
-Broad spurious precipitating system
-Extent of main line limited

FV3SAR

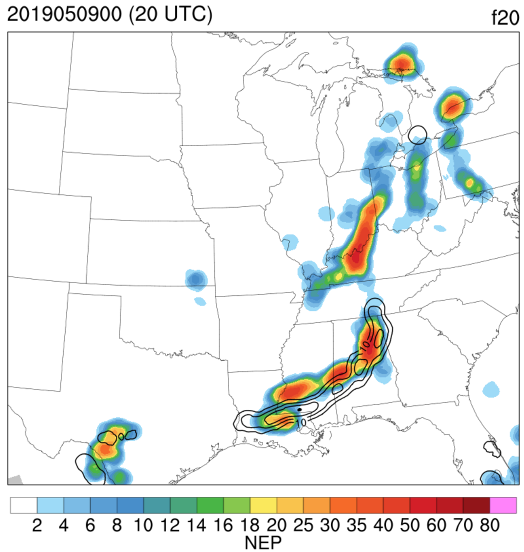


-Limited spurious precip to north
-Convective line matches length, location of reality with strong probabilities

Case Example – 9 May 2019
Hours 12-20, NEP QPF > 6.35 mm

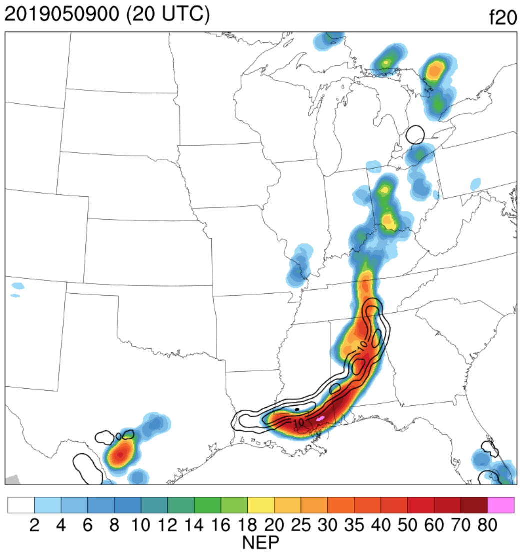


ARW



-Broad spurious precipitating system
-Delay in development of line

FV3SAR



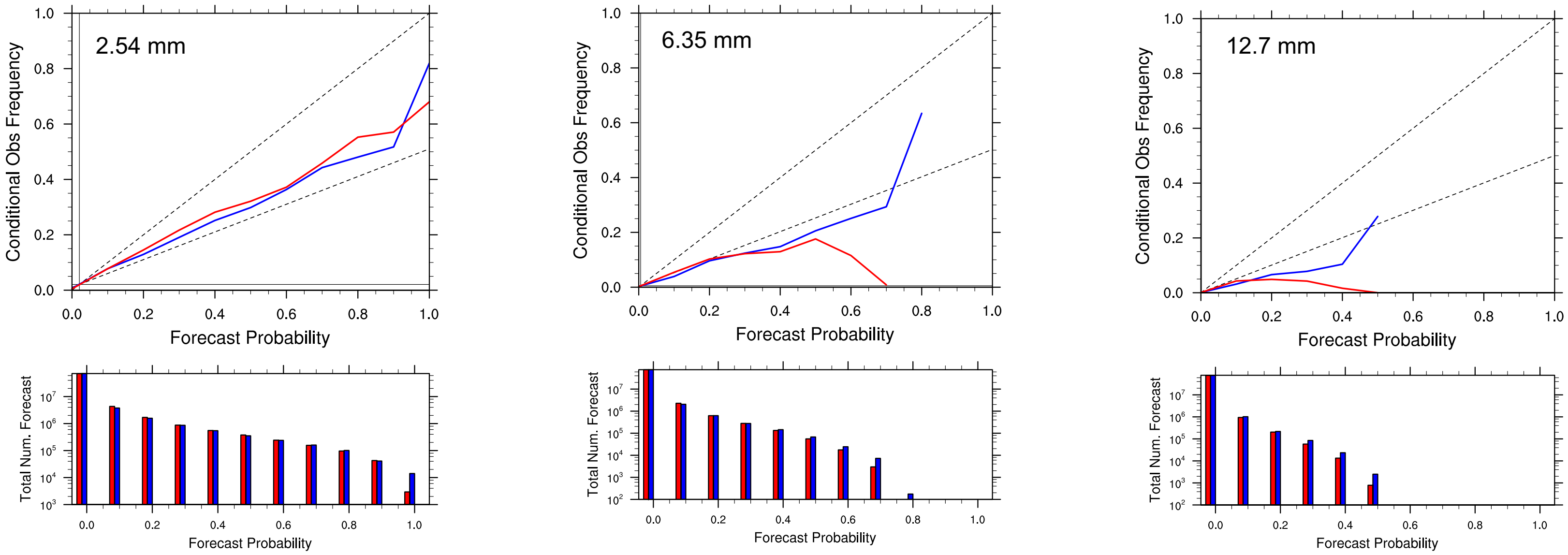
-Excellent development of line
-Slight forward propagation error

Subjective Summary – May 9, 2019

- ARW outperforms FV3 at early hours due to better matching location, strength of early convective systems from IC time.
  - Subsequently, more accurate development of convection in SE Texas (though slightly early), leads to accurate short-term forecast
- FV3, on the other hand, was delayed in convective development after ICs, with weaker probabilities and 3-4 hour delay in development over E TX.
  - This may indicate significant spinup issues, possibly related to different model used during DA
  - FV3 inherently disadvantaged by different model used during EnVar cycling

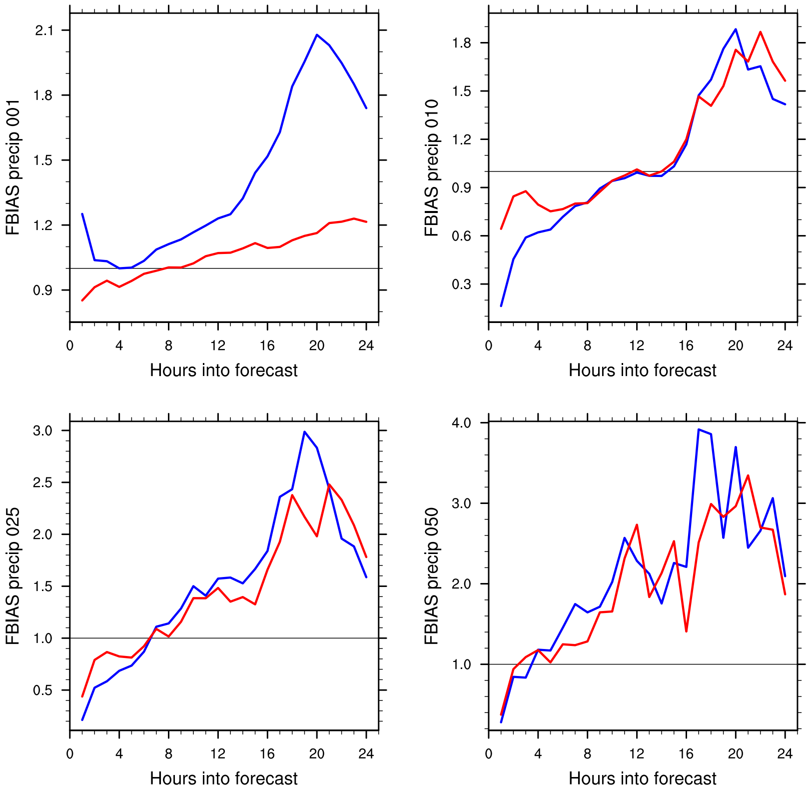
- In the latter half of the forecast, the ARW simulation had excessive spurious precipitation
  - FV3, on the other hand, had substantially less spurious precip
- The FV3 had excellent development of a bowing line of convection, matching location and extent with reality
  - ARW had smaller, weaker line

Preliminary 10 Case Verification of 1-h QPF Reliability (fhr 13-24 aggregate)



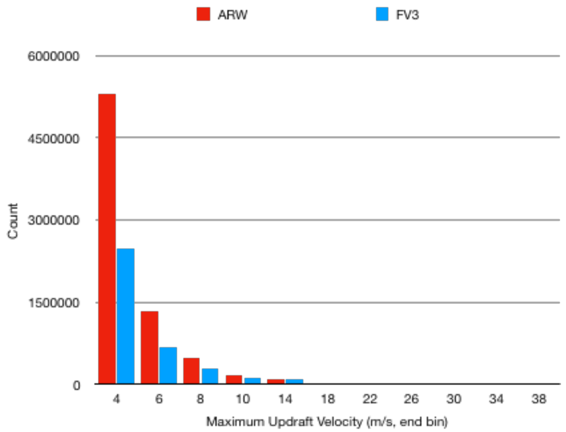
- FV3 and ARW have *comparable reliability* (ARW slightly more reliable at 2.54mm threshold)
- FV3 has increased reliability and sharpness for higher probabilities, particularly for heavier precipitation thresholds – more ensemble agreement (less spread)

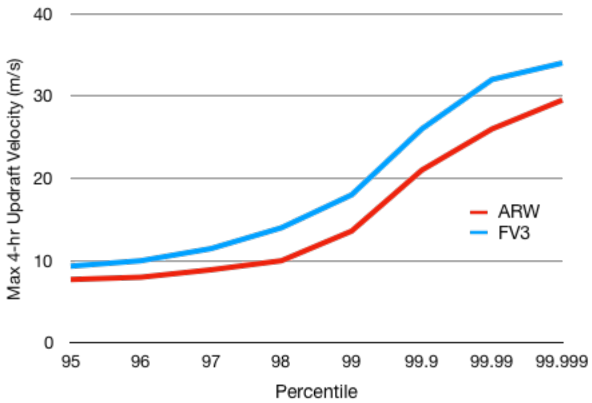
Preliminary 10 case Verification of 1-h QPF Frequency Bias



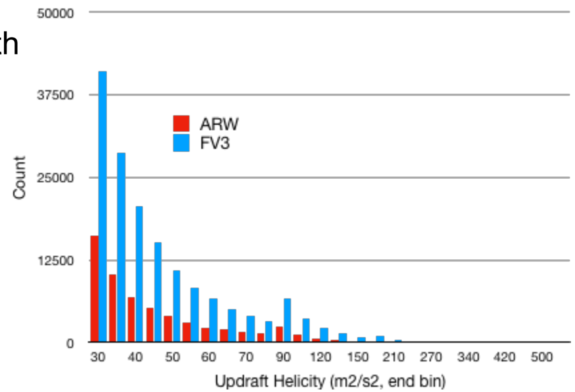
- FV3 has many more points with precipitation, shown by lightest threshold
- With the exception of lightest threshold (0.254 mm, marking precip or no precip), frequency bias of ARW and FV3 are comparable
  - FV3 has slightly larger frequency bias for later hours

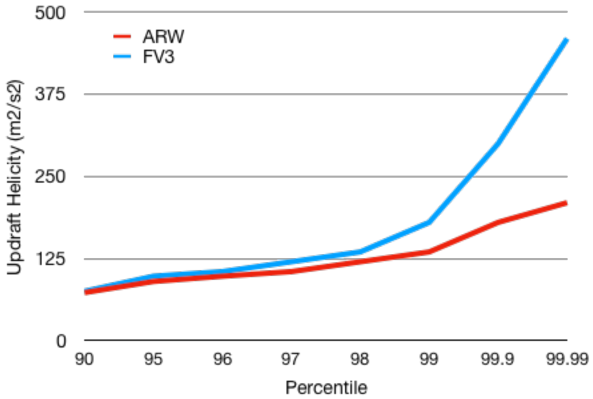
Histograms of 4-hr Ens. Max Updraft Velocity (left) and Helicity (right)





- ARW has many more grid points with smaller updraft velocity – effect of temperature tendency limiter in ARW?
- FV3 has longer tail in distribution – strongest storms have larger updrafts than ARW





- FV3 and ARW have similar shape, but FV3 has many more grid points with UH > 25 m2/s2
- FV3 capable of extrema UH values much larger than ARW
- ARW may be limited by temperature tendency – excessive damping causing less extreme structure

Conclusions

- Objective – evaluate FV3 dynamical core performance compared to ARW dynamical core with all else being (mostly) equal for convective-scale forecasting, initializing from multiscale EnVar ensemble analyses from OU-MAP HWT SFE 2019
- Initial dynamical tunings done subjectively with May 20, 2019 case
- Preliminary 10-case 1-h QPF verification and subjective analysis shows ARW to be largely more skillful than FV3SAR; however, the FV3SAR did have periods of better skill and reliability (forecast hours 13-24, heavier precipitation thresholds)
    - Case study (9 May 2019) showed delay in developing convective systems from ICs
- FV3 had more extreme values in UH, VV than ARW
    - ARW likely hindered by temperature tendency limiter
- Despite best efforts, experiment design is inherently unfair to FV3SAR, since analyses were produced using a different model during cycling (ARW-HRRR config).
    - Spinup effects last 3-6 hours
    - Likely explains relatively poor performance for first half of the forecast

Future Work

- Systematic tuning of key dynamics and physics parameters – examine objectively how sensitive resulting verifications of 1-h QPF, composite reflectivity, etc. are to each parameters
  - Question to consider: *How should temperature tendency limiter (ttendlim), a physics parameter for Thompson, be tuned relative to dynamics options? Should ttendlim be smaller than that used in ARW, given differences in dynamical core? Previous results show strongest sensitivity of storm structure (UH, VV) to ttendlim, but properly tuned value may depend on dynamics options chosen.*

- For fairer comparison of dynamical cores initialized from multiscale EnVar analysis, FV3 should run its own multiscale EnVar analysis rather than initialize from analyses produced during HWT (using ARW)
  - Efforts are ongoing in collaboration with EMC on running FV3 SAR's own cycled EnVar
  - Further develop cycled FV3SAR EnVar with JEDI

Thank you!