30th Conference on Weather Analysis and Forecasting (WAF)/26th Conference on Numerical Weather Prediction (NWP), 12-16 January 2020, Boston, MA



# A Comparison of Multi-dynamic Core vs Single-dynamic Core Multi-Physics Ensemble Design for Convection-allowing Forecasting Initialized by the Multiscale EnVar Data Assimilation System

**Nicholas Gasperoni and Xuguang Wang**

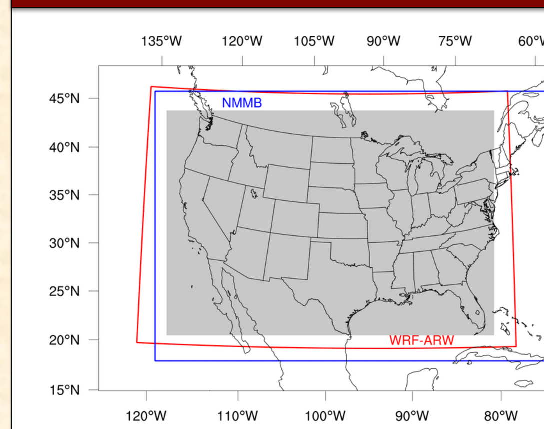
School of Meteorology, University of Oklahoma, Norman, OK, U.S.A

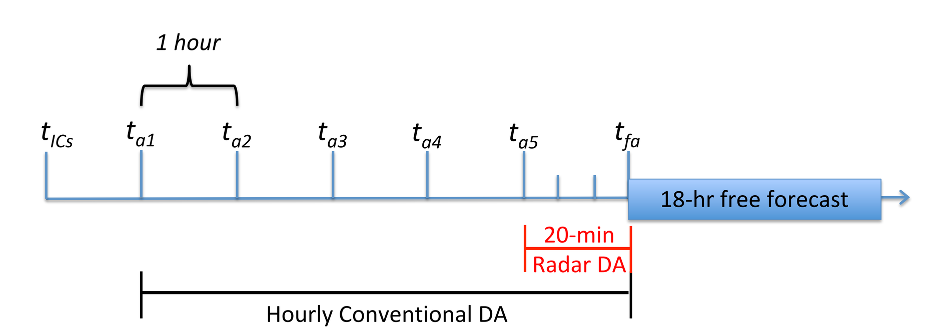
Multi-scale data Assimilation and Predictability (MAP) Laboratory, Norman, OK, U.S.A.

## Introduction/Motivation

- The Gridpoint Statistical Interpolation (GSI)-based hybrid ensemble-variational (EnVar) data assimilation (DA) scheme has been extended for convective scales including the direct assimilation of radar radial velocity and reflectivity by Johnson et al (2015) and Wang and Wang (2017).
- This GSI-EnVar system has been tested within the operational NAMRR and HRRR model contexts for the 2017 and 2018 NOAA Hazardous Weather Testbed (HWT) spring forecasting experiments by the Multiscale data Assimilation and Predictability (MAP) lab at the University of Oklahoma.
- Ten retrospective cases from 2015 and 2016 are used to facilitate development and testing of potential improvements to the convection-allowing assimilation and ensemble forecast system.
- In this study, these ten retrospective cases are used to explore the optimal design of convection allowing forecasts.
- The performance of ensemble forecasting system depends upon ability of the ensemble to represent all sources of uncertainty, including *initial condition (IC) errors* and *model errors.*
  - GSI-EnVar already samples multi-scale IC errors down to convective scale
  - ***How do we properly sample model errors? In particular, how does a multi-core ensemble compare to single-core single-physics and single-core multi-physics ensembles? Can a single-core ensemble match the spread and skill of multi-core ensemble?***

## GSI-EnVar Assimilation Setup




**Domain:**
- Resolution: 3 km
- Grid: 1621 X 1121 X 50
- Same as HWT CLUE

**Observations:**
- Conventional obs assimilated hourly
- Radar reflectivity obs assimilated every 20 minutes for last hour of DA

**IC and LBC ensemble** are provided by re-centering GEFS (20) and SREF (20) perturbations to GFS-ctl

## Forecast Ensemble Configuration

| Experiment Name | Model Core | Member # | Microphysics Scheme | PBL Scheme | LSM Scheme |
|---|---|---|---|---|---|
| **NMMB** | NMMB | M0-M9 | Ferrier-Aligo | MYJ | Noah |
| **ARW-SP** (Single Physics) | ARW | M0-M9 | Thompson | MYJ | RUC |
| **MM** (Multi-Model) | NMMB + ARW | M0-M9 randomly split and taken from NMMB and ARW-SP experiments | | | |
| **ARW-MP** | ARW | M0 (control) | Thompson | MYJ | RUC |
| (Multi-Physics) | | M1 | Thompson | MYJ | Noah |
| | | M2 | NSSL | YSU | Noah |
| | | M3 | NSSL | MYNN | Noah |
| | | M4 | Morrison | MYJ | Noah |
| | | M5 | P3 | YSU | Noah |
| | | M6 | NSSL | MYJ | Noah |
| | | M7 | Morrison | YSU | Noah |
| | | M8 | P3 | MYNN | Noah |
| | | M9 | Thompson | MYNN | Noah |
| **ARW-MPSKEB** | ARW | As in ARW-MP but including application of SKEB during forecast | | | |

- Each ensemble free forecast is initialized from a **10-member subset** of the corresponding final GSI-EnVar analysis (NMMB or ARW), with the first member being the control member
- **The MM ICs are a random sampling of the 10 final analyses from the NMMB and ARW-SP experiments (5 from each experiment)**
- **ARW-MP and ARW-MPSKEB are initialized from the final analysis from the ARW experiments (ARW-SP)**

| Case name | Final Analysis time | Synoptic forcing |
|---|---|---|
| May25_2015 | 1300 UTC 25 May 2015 | Strongly |
| June17_2016 | 2000 UTC 17 June 2016 | Weakly |
| May16_2015 | 2300 UTC 16 May 2015 | Strongly |
| July06_2016 | 0100 UTC 05 July 2016 | Weakly |
| July07_2016 | 0000 UTC 07 July 2016 | Weakly |
| June26_2015 | 0400 UTC 26 June 2015 | Weakly |
| July10_2016 | 0400 UTC 10 July 2016 | Weakly |
| May23_2016 | 2300 UTC 23 May 2016 | Moderately |
| Sept11_2015 | 0100 UTC 11 September 2015 | Moderately |
| July14_2015 | 1900 UTC 14 July 2015 | Strongly |

- The cases include many examples of both **discrete isolated storms** (e.g., May 23, 2016 case) and **organized MCSs** (e.g., long-lived squall lines in May 16, 2015 and May 25, 2015 cases).
- Diverse synoptic forcing and organizing mechanisms are also included in these cases, such as strong upper level trough and surface cold fronts, slow moving or stationary frontal zones with multiple clusters growing upscale.

## Neighborhood Verification Results

- 1-hour Quantitative Precipitation Estimate (QPE) and Composite Reflectivity (CREF) forecasts are verified against Multi-Radar Multi-Sensor (MRMS) gridded observations.
- 48-km neighborhood radius applied for QPE and CREF ensemble verification

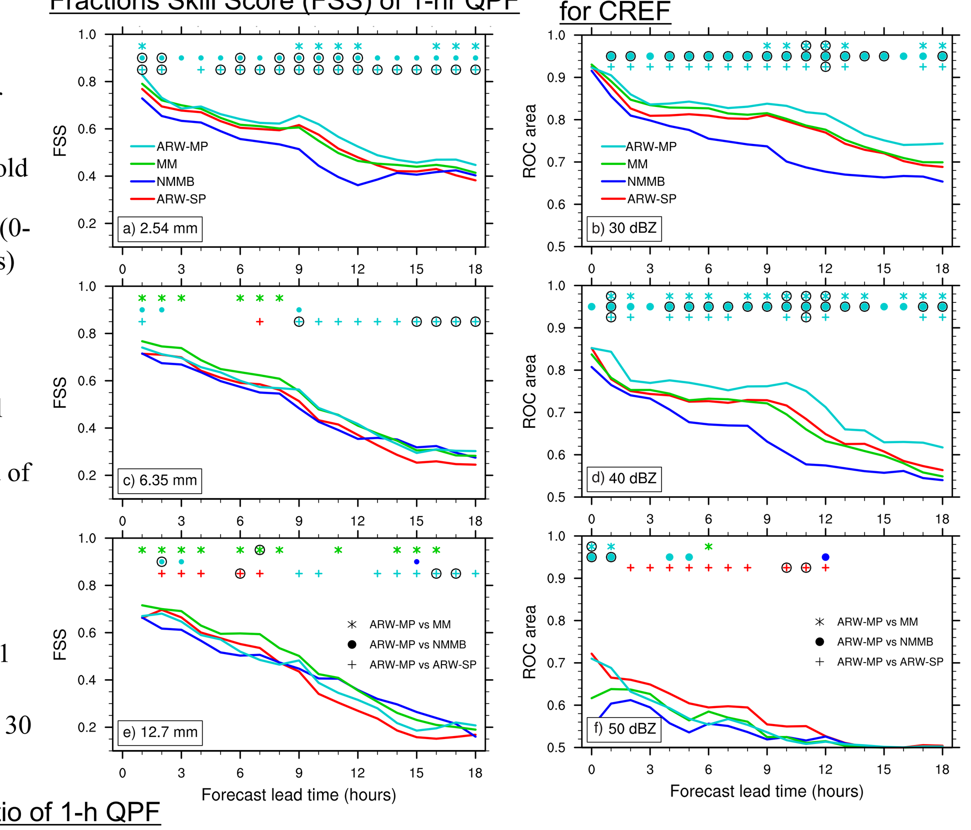
FSS Notes
- ARW-SP outperforms NMMB at 0.1 in. threshold.
- MM competitive with best single-model single-physics run for 0.1 in threshold
- ARW-MP shows benefit at later forecast times at 0.1 in threshold

---------------

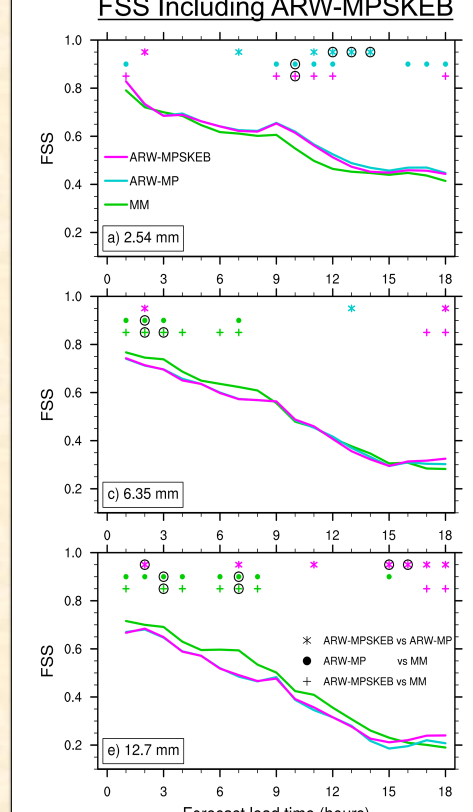
- At higher thresholds, ARW-SP > NMMB for early lead times (0-9 hrs), then NMMB > ARW-SP for later lead times (12+ hours)
- MM best experiment from hours 0-9
- ARW-MP generally in the middle, or tied for best skill in 9-18 hours at 0.25 in.
- ARW-MPSKEB slightly improved FSS over ARW-MP in final ~4 hours of heavy precipitation, but large increases in spread (shown by correspondence ratio below), matching MM by end of forecast

ROC Areas Notes
- Measure of forecast discrimination (between event and non-event occurrences)
- For precip, AUC for multi-model and ARW-MP are best at 0.1 in, but indistinguishable from one another (not shown)
- However, in terms of CREF, ARW-MP shows clear benefit at 30 and 40 dbz

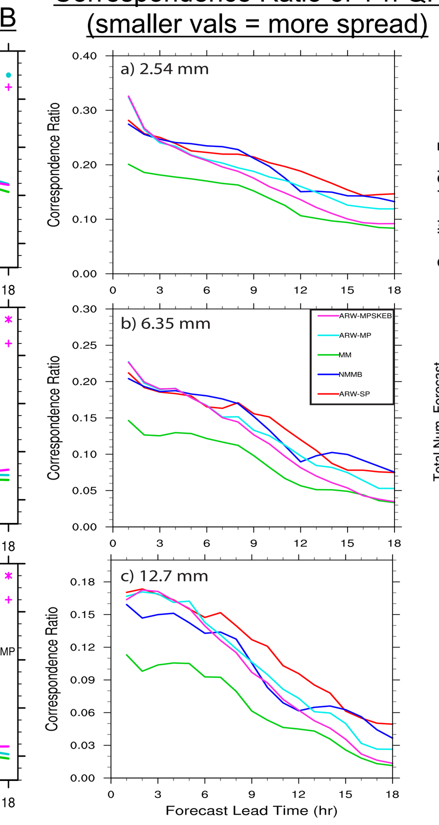
Fractions Skill Score (FSS) of 1-hr QPF | Relative Operating Characteristics (ROC) Area Under the Curve for CREF



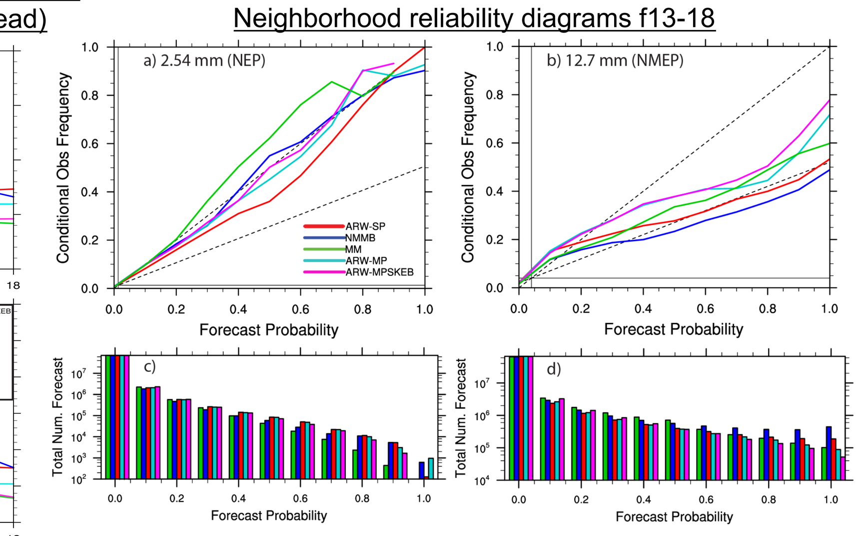

FSS Including ARW-MPSKEB




Correspondence Ratio of 1-h QPF (smaller vals = more spread)



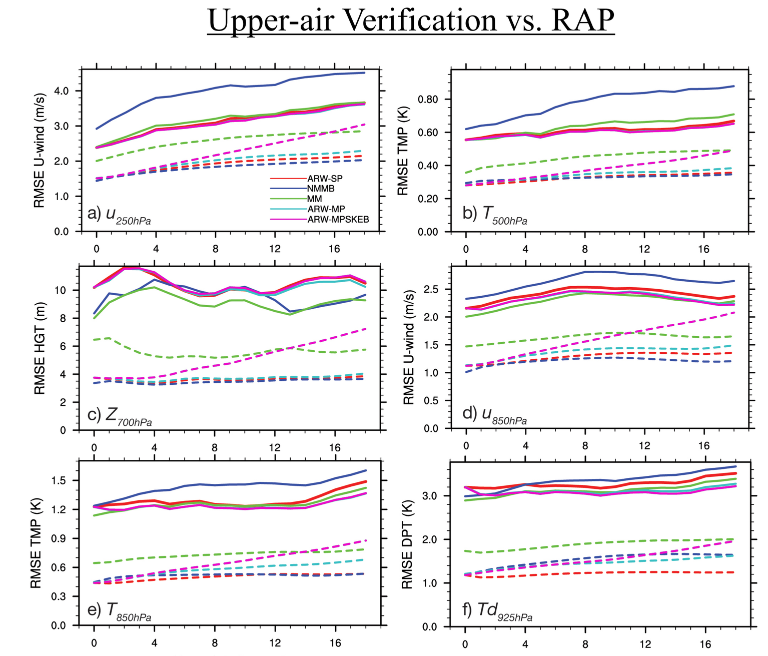

Neighborhood reliability diagrams f13-18



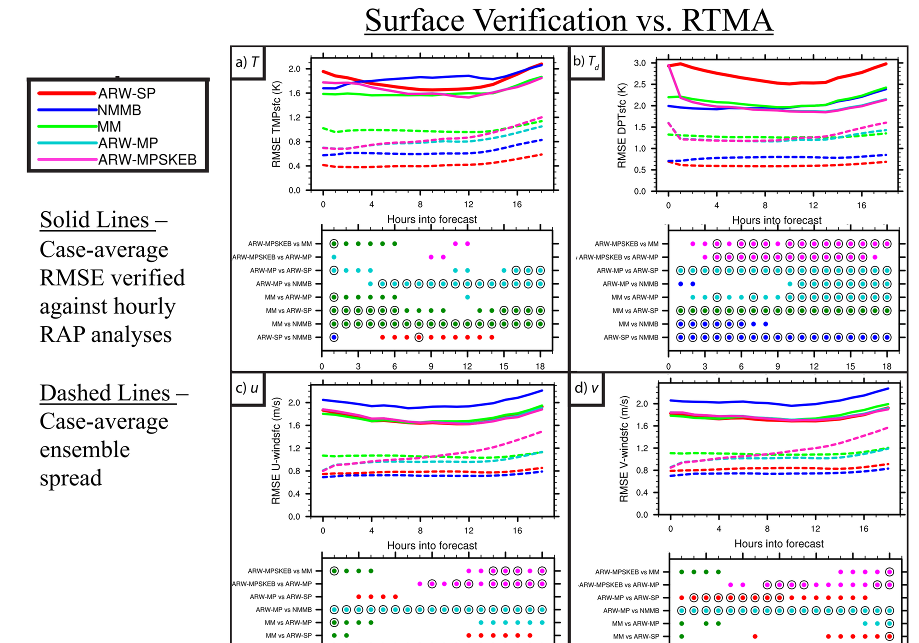

- At the lighter 2.54 mm threshold, there is a noticeable improvement in reliability of the mid 50-70% probability ranges comparing ARW-MPSKEB to ARW-MP.
- Model error experiments more reliable than SMSP at 12.7mm, with ARW-MPSKEB having the best reliability of higher probabilities.

## Spread-RMSE - Verified against RAP & RTMA analyses

Upper-air Verification vs. RAP




Surface Verification vs. RTMA



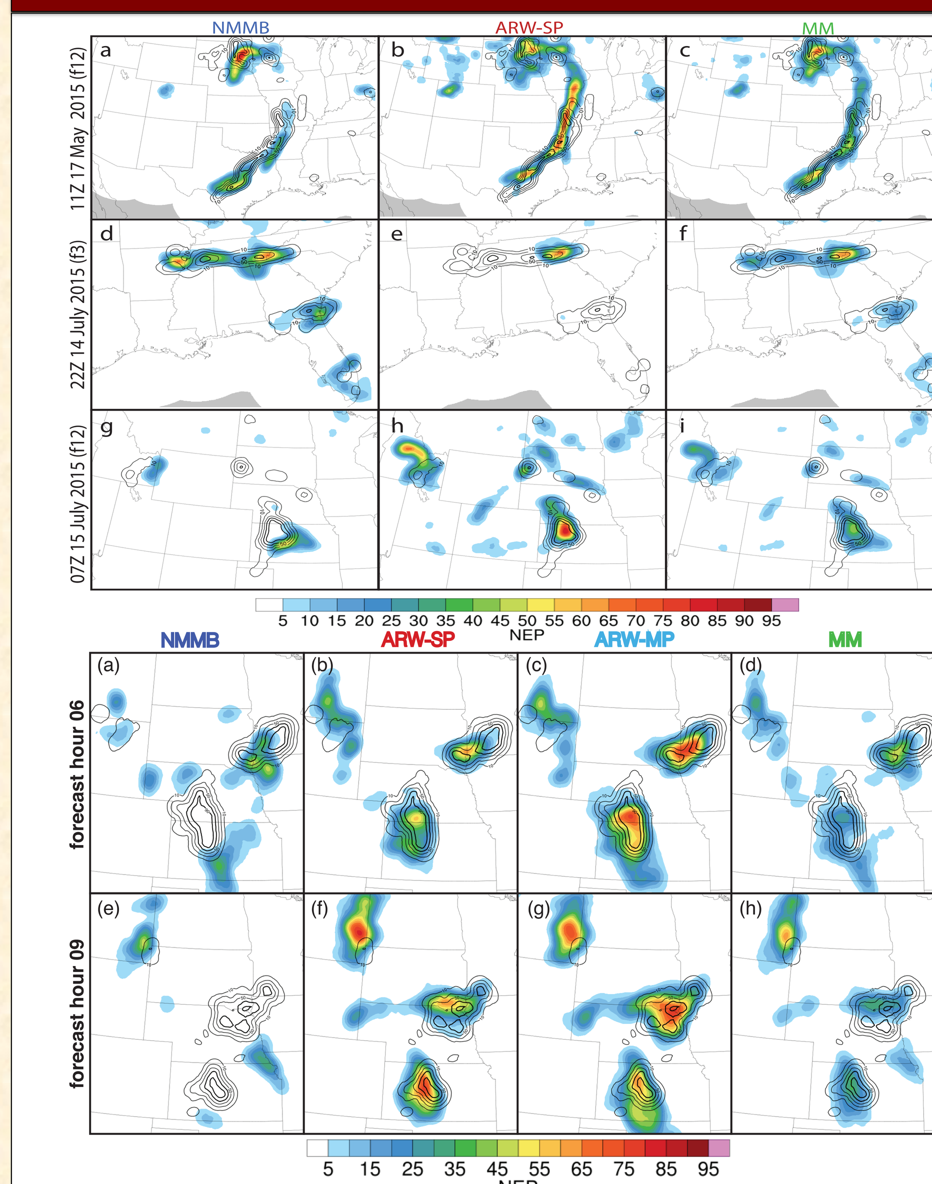

Solid Lines – Case-average RMSE verified against hourly RAP analyses

Dashed Lines – Case-average ensemble spread

- NMMB generally highest in error at most upper levels in wind, temp, dewpoint, with each of the other experiments clustered tightly together for fields at or above 500 hPa
- Below 700 hPa, larger differences in error with ARW-MP and ARW-MPSKEB having lowest error in thermodynamic fields, followed by MM
- MM adds significant spread over single-model consistently at all levels, variables.
- ARW-MP increases in spread with increasing forecast time, particularly for near-surface variables
- ARW-MPSKEB substantially increases spread over the course of the forecast, most notably in wind and geopotential height
  --Spread approaches MM spread by fhr 18

- ARW-SP lower in error by 0.1-0.4 m/s and K for wind and temperature than NMMB; however NMMB lower in dewpoint error by about 1 K.
- Differences in MM and ARW-MP are generally small in wind.
- MM significantly lower (0.1-0.2 K) than ARW-MP for hours 1-6 temperature, while ARW-MP significantly lower by 0.1-0.3 K in dewpoint at hours 2-18.
- MM has consistently highest spread, particularly early lead times
- Combination of SKEB and MP shows most spread in near-surface wind in final 12 hours of forecast, and matches spread of MM in thermodynamic variables

## Case Examples of MM and MP Benefits

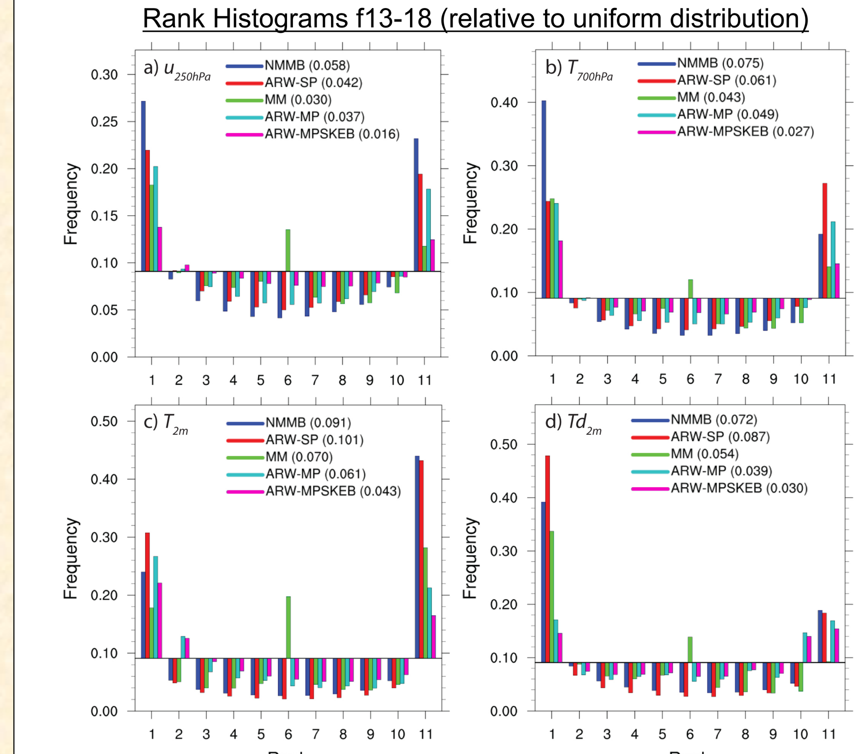



- After subjectively analyzing on a case-by-case basis, there are two main reasons why MM shows benefit over NMMB and ARW-SP
  1. Better location and storm coverage in some cases – i.e. errors of NMMB and ARW-SP are "opposite" (top row at left)
  2. Avoidance of single-model skill drop-offs in some cases – especially important when both models are similar in skill overall but on a case-by-case or hour-by-hour basis the best SMSP experiment differs (e.g. middle row, NMMB > ARW_SP; however, bottom row, ARW_SP > NMMB. MM maintains storm coverage at both times/cases)

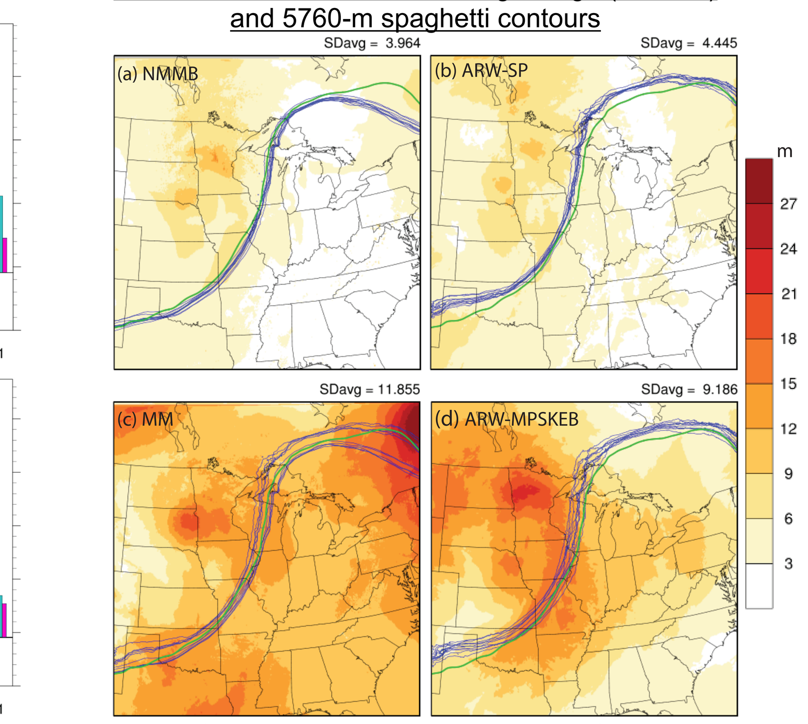
- In terms of MP, we found benefits in cases where NMMB was inferior to ARW core.
- The most notable improvement of ARW_MP over ARW_SP tended to be in cases with weak or moderate synoptic forcing
  - Enhanced sensitivity to physics scheme, increased importance of forecast ensemble needing to account for model physics uncertainties.

## Ensemble Spread Diagnostics

Rank Histograms f13-18 (relative to uniform distribution)




Ensemble Std. Dev. of 500-hPa geo. hgt. (color fill) and 5760-m spaghetti contours




16 May 2015 fhr 18 shown. Blue – ensemble members, Green – RAP analysis

## Discussion

- Though many previous studies have examined ensemble design in multi-core or multi-physics contexts, this study examines them in the context of an optimal set of IC perturbations created by a multiscale hybrid DA system
- Among SMSP experiments, ARW-SP had superior performance to the NMMB for lighter precipitation fields and earlier forecast times, as well as much of the mean RMSE verification. The NMMB decayed MCSs too early in cases where decaying MCSs occurred in reality
- Each of the model error experiments MM, ARW-MP, and ARW-MPSKEB compared favorably to NMMB and ARW-SP in many of objective verification scores and all of the ensemble spread diagnostics.
- The MM experiment had the highest FSS for heavy precipitation thresholds, as well as the most amount of spread added consistently throughout the entire forecast. However, this added spread comes at the cost of undesirable ensemble clustering seen in spread diagnostics.
- The ARW-MP experiment significantly improved upon FSS precipitation verification over ARW-SP for lighter precipitation thresholds and the final 9-12 hours of heavier precipitation thresholds. This was accompanied with large increases in forecast discrimination and reliability of precipitation, as well as forecast spread of near-surface fields. However, the increase in spread was limited mainly to lower level fields below 850 hPa and took a "spin up" period of at least 6 hours for noticeable increases to appear.
- Adding SKEB on top of ARW-MP showed small but significant improvements to precipitation verification in the latter half of the forecast for heavy (6.35 and 12.7 mm) precipitation. Additionally, there were substantial increases in ensemble spread over time for both precipitation systems and for upper level fields (in particular wind and geopotential height).
- The comparison of MM to ARW-MP and ARW-MPSKEB led to some mixed results and depends upon what aspect of the forecast is examined. For early forecast lead times, MM is favored. However, ensemble clustering was found, and as SKEB and MP effects spun up the ARW-MPSKEB experiment showed much more favorable ensemble distributions with similar or lower skill to MM within the final 6-9 hours of the forecast.

Results documented in recently accepted publication:

Gasperoni, N.A., X. Wang, and Y. Wang, 2019: A comparison of methods to sample model errors for convection-allowing ensemble forecasts in the setting of multiscale initial conditions produced by the GSI-based EnVar assimilation system. *Mon. Wea. Rev.,* accepted.