Testing of 3-km FV3 during 2017 HWT and HMT & further Development and Testing

Presented by Ming Xue
Center for Analysis and Prediction of Storms and School of Meteorology
University of Oklahoma

CAPS Team: Ming Xue, Fanyou Kong, Chunxi Zhang, Youngsun Jung, Kevin Thomas, Gang Zhao, Chengshi Liu, Keith Brewster, Nathan Snook

With inputs from Corey Potvin and Adam Clark of NSSL
Lucas Harris and SJ Lin of GFDL

January 31, 2018
SIP NGGPS Meeting at EMC

mxue@ou.edu

ARPS Simulated Tornado

CAPS Contributions to CLUE 2017

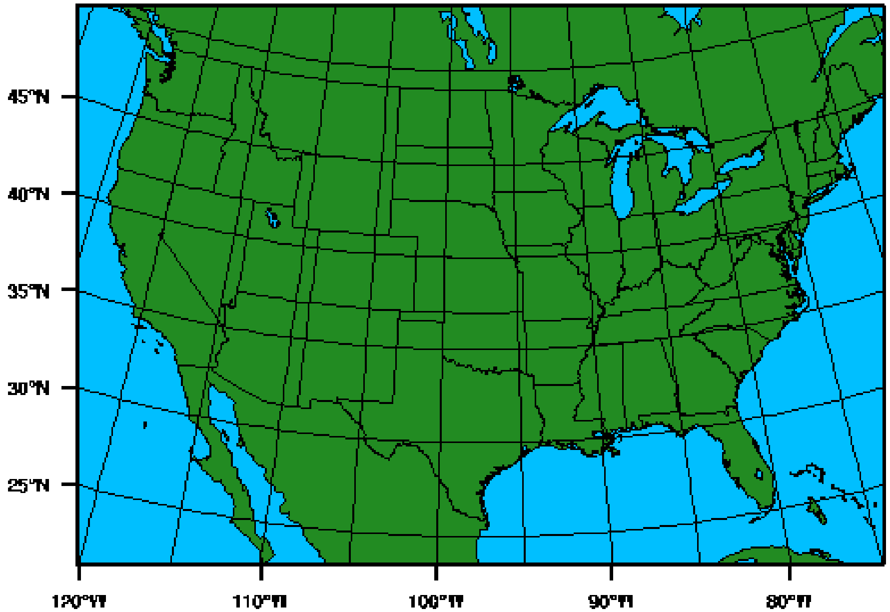
Community Leveraged Unified Ensemble (CLUE)

34 Members from CAPS (out of 79 total CLUE members)

- ***Storm-Scale Ensemble Forecast (SSEF) System using CAPS's 3DVAR/cloud analysis DA***
  - 10-member, 3-km WRF-ARW ensemble with 60-hr forecasts from 00Z
  - Multi-physics, multi-IC/LBC conditions: add SREF perturbations to NAM ICs
  - One member matches HRRR model configurations, but using CAPS 3DVAR/cloud analysis for radar DA
- ***SSEF using GSI EnKF DA plus CAPS EnKF for radar DA***
  - 10-member, 3-km WRF-ARW ensemble with 60-hr forecasts from 00Z
  - Multi-physics, 3-km 40-enemble 6-h cycled EnKF analysis ICs with radar DA
- ***Single FV3 convection-allowing forecasts (with 1-year support from NGGPS program)***
  - ~3 km over CONUS, nested within global run
  - Thompson microphysics (added by CAPS)
  - GFDL ran another version with GFDL single-moment MP – based on WSM6
- ***Single physics SSEF members with radar***
- ***Mixed microphysics SSEF with radar***

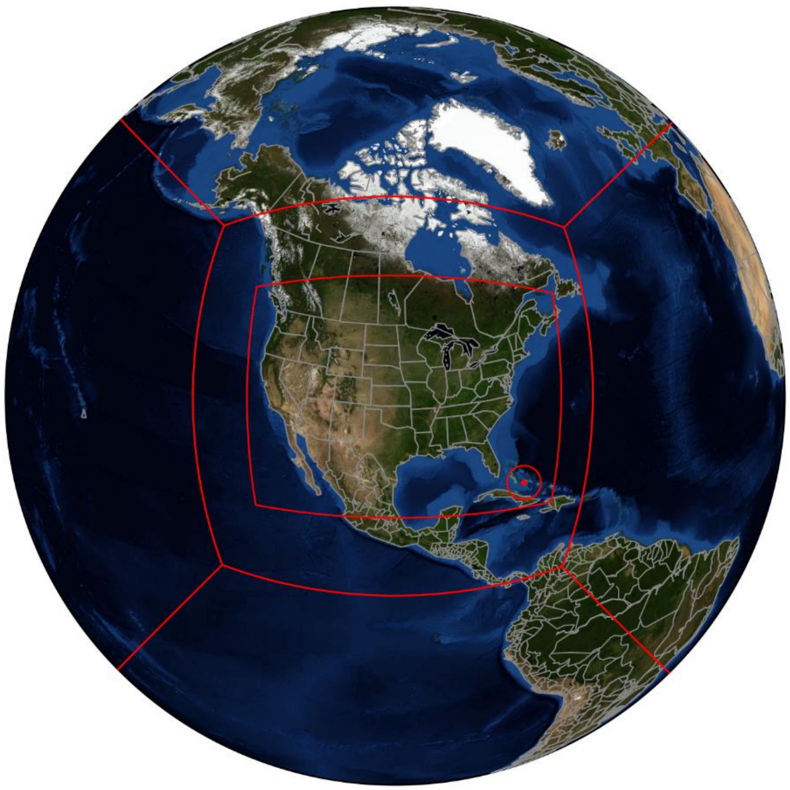
2017 CAPS 3km SSEF and FV3 Domains (run for 2017 HWT SFE and HMT FFaIR Experiments)

HWT: 5 weeks in May and early June. HMT: 4 weeks in June and July



3 km WRF Grid 1620x1120
Shared with other CLUE members

Once a day starting from 00 UTC



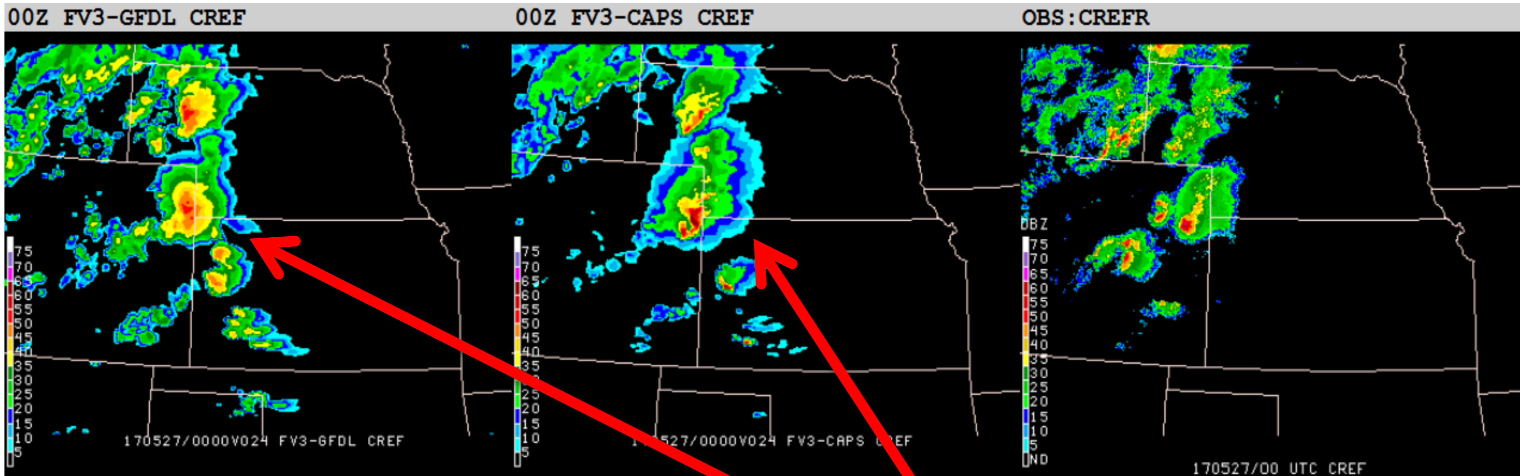
Nested FV3 Grid ~ 3 km over CONUS
~ 13 km average global grid
Coordinated with GFDL run

FV3 Configurations for 2017 HWT SFE

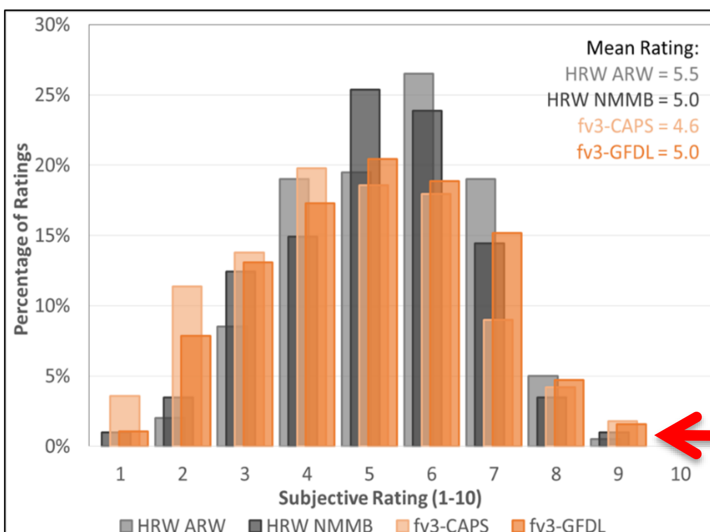
- Microphysics
  - FV3_CAPS: Thompson MP
  - FV3_GFDL: GFDL MP (based on WSM-6)
- Cumulus
  - Scale-aware SAS (Global), None (Nest)
- PBL
  - MRF (from GFS)
- Radiation
  - RRTM
- Land surface model
  - NOAH
- Initial condition
  - Cold-started from 00Z GFS T1534 analysis every day
- LBC: two-way nested within global grid

CLUE Results: FV3 (2017) – from Adam Clark

Subjective results



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



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

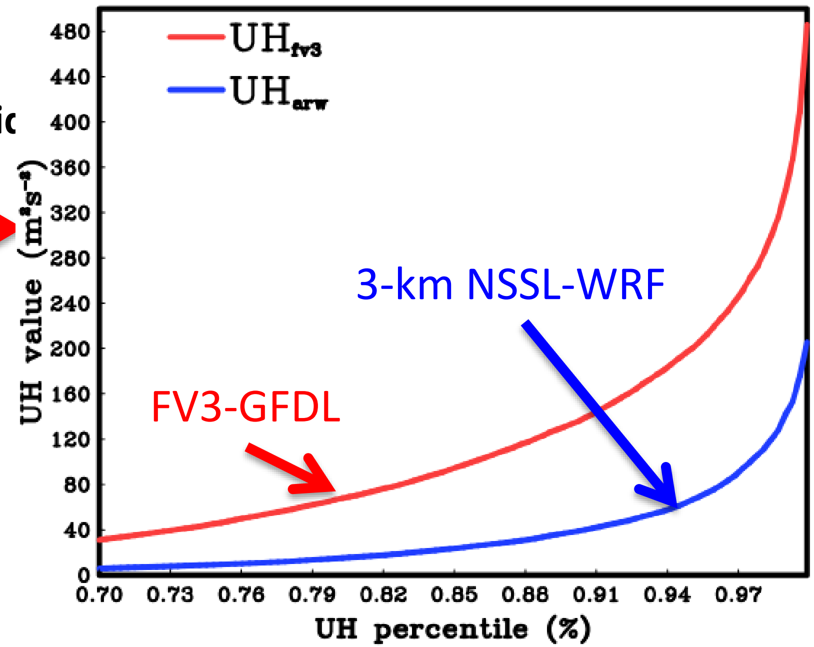
According to SFE participant ratings, FV3 is competitive with operational CAMs

FV3 obtained more 9 ratings than others

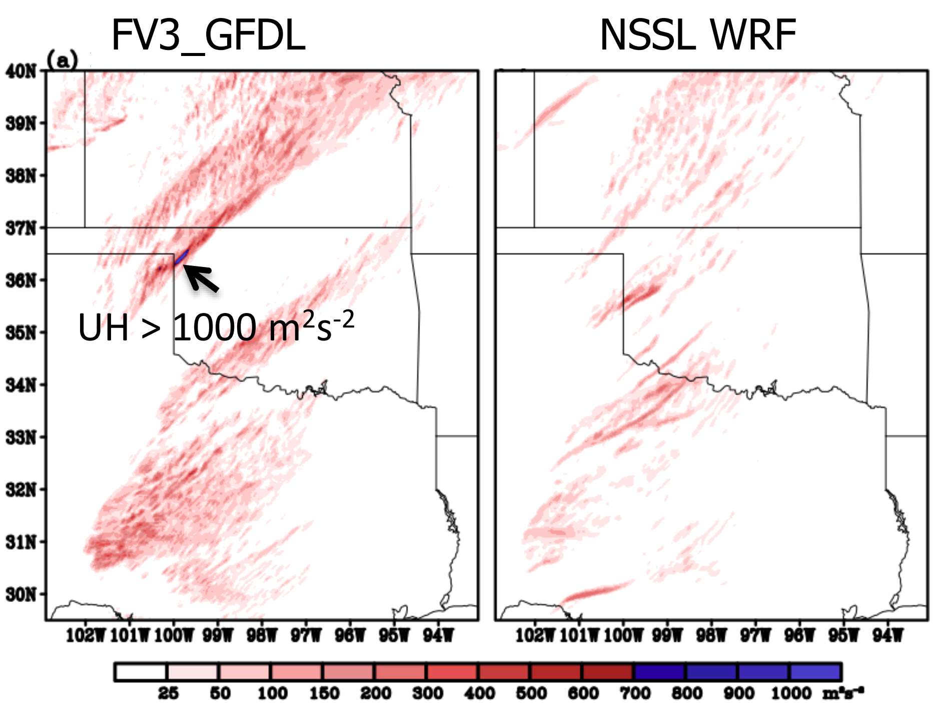
CLUE Results: FV3-GFDL (2017) – source: Adam Clark

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

- **24-h max updraft helicity (UH) remapped to 80-km grid**
- **UH distributions in terms of percentiles**
- **FV3 generally has larger UH values**



Example: 24 h maximum UH for 16 May 2017

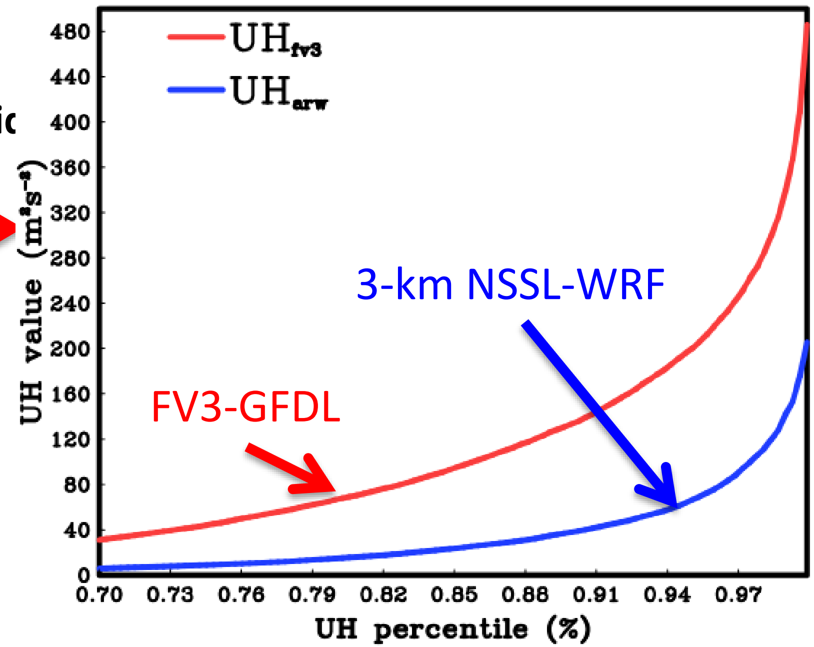


- Higher, more widespread, and noisier appearance of UH in FV3.

CLUE Results: FV3-GFDL (2017) – source: Adam Clark

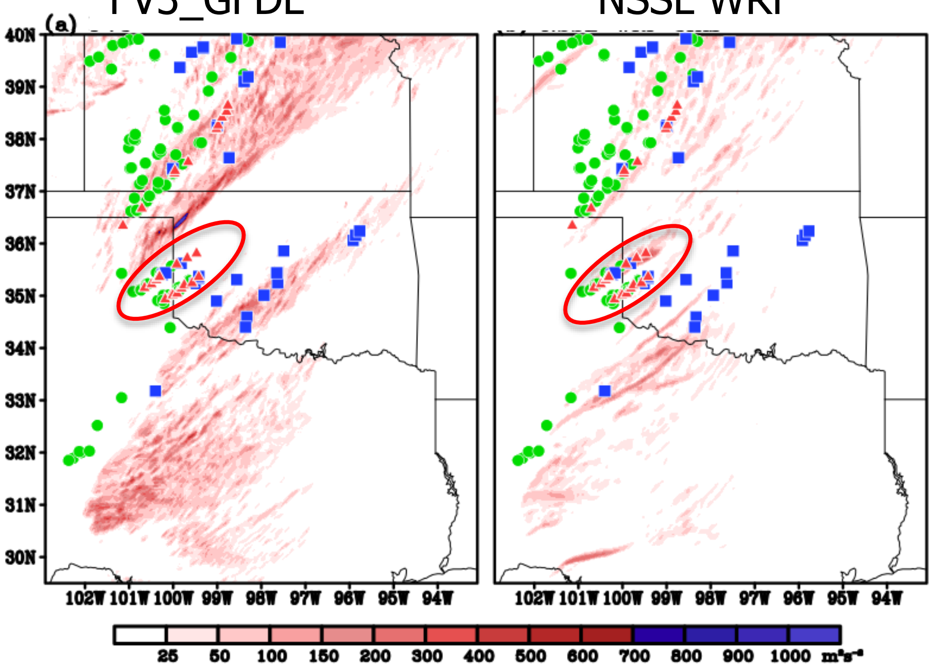
- Surrogate severe method (Sobash et al. 2011, 2016) used to compared FV3-GFDL and 3-km NSSL-WRF

- **24-h max updraft helicity (UH) remapped to 80-km grid**
- **UH distributions in terms of percentiles**
- **FV3 generally has larger UH values**



Example: 24 h maximum UH for 16 May 2017

FV3_GFDL NSSL WRF

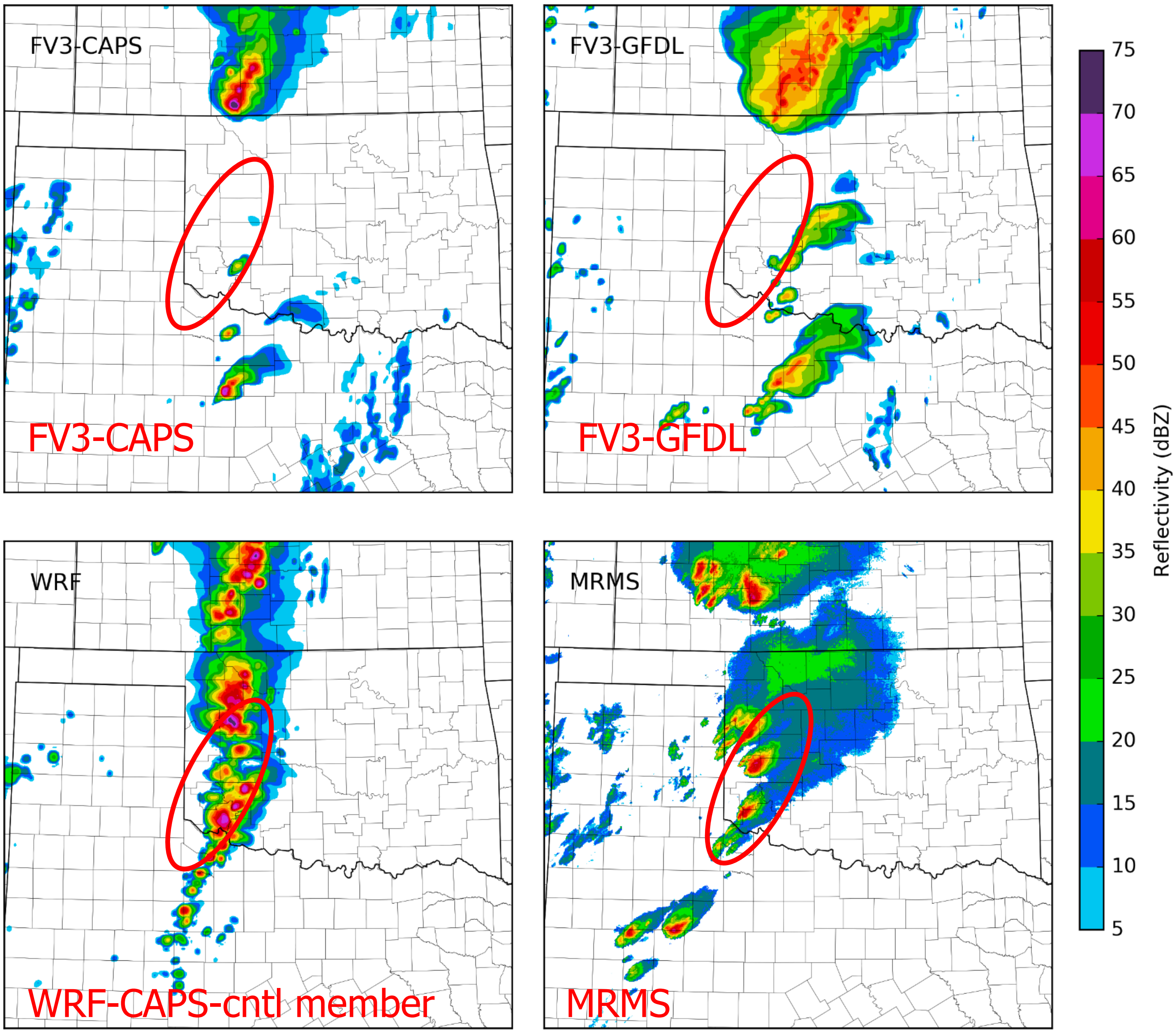


- Higher, more widespread, and noisier appearance of UH in FV3.

- Storm reports overlaid
- NSSL WRF did better with tornadic storms than FV3 on this day – this day was a challenging day for FV3!

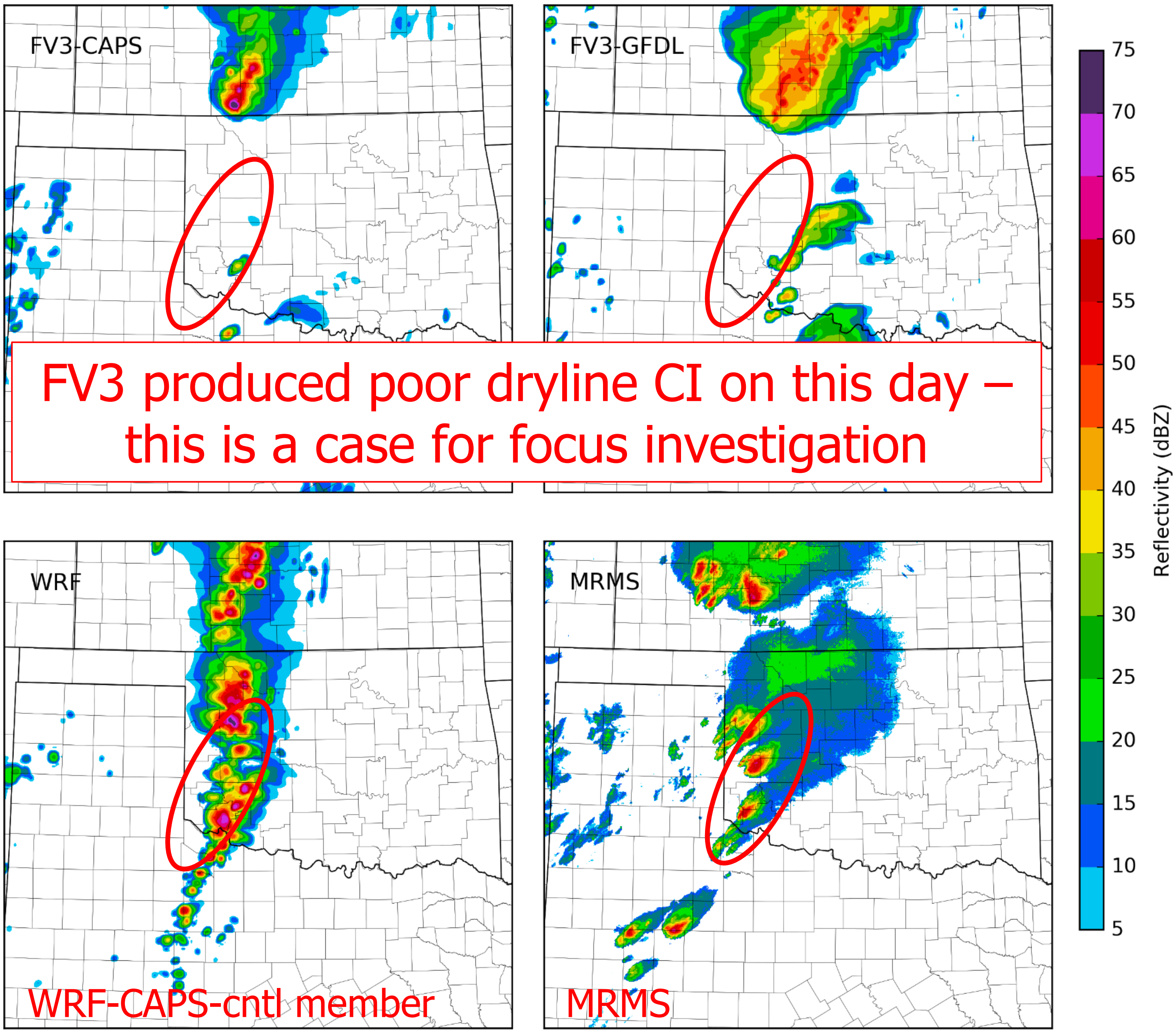
16 May 2017 Composite Reflectivity (00Z 17 May)

t = 24h



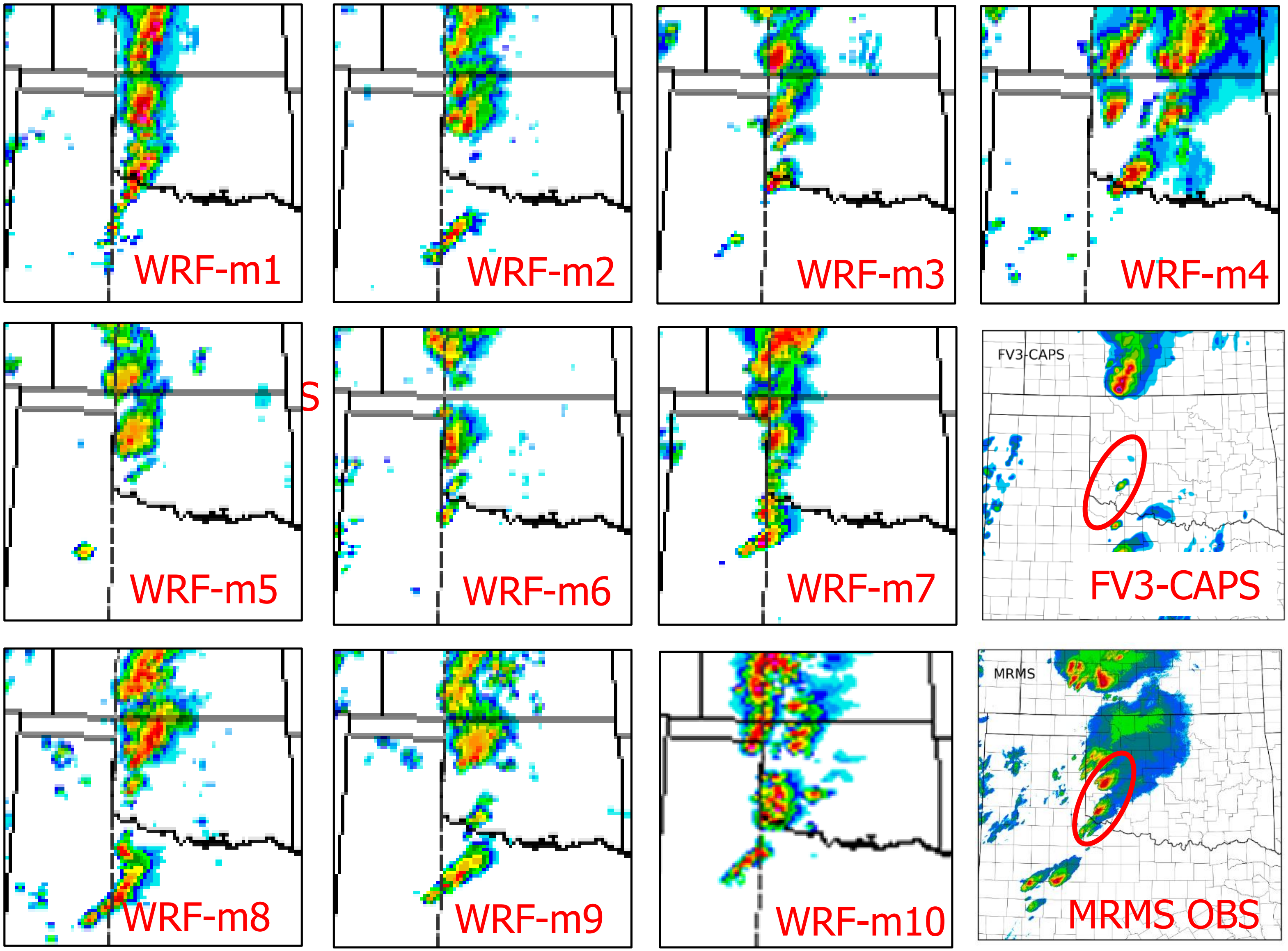
16 May 2017 Composite Reflectivity (00Z 17 May)

t = 24h



FV3 produced poor dryline CI on this day – this is a case for focus investigation

16 May 2017 Composite Reflectivity (00Z 17 May)

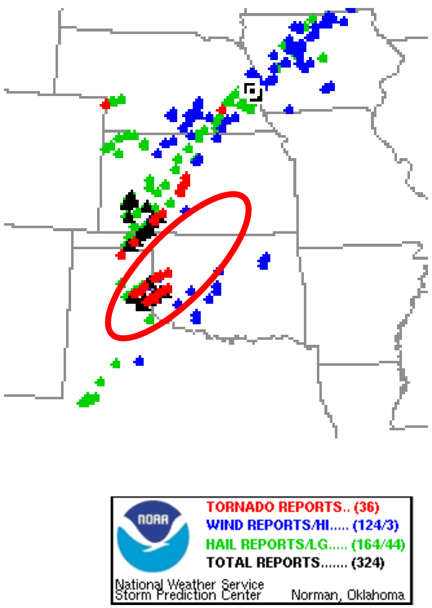


t = 24h

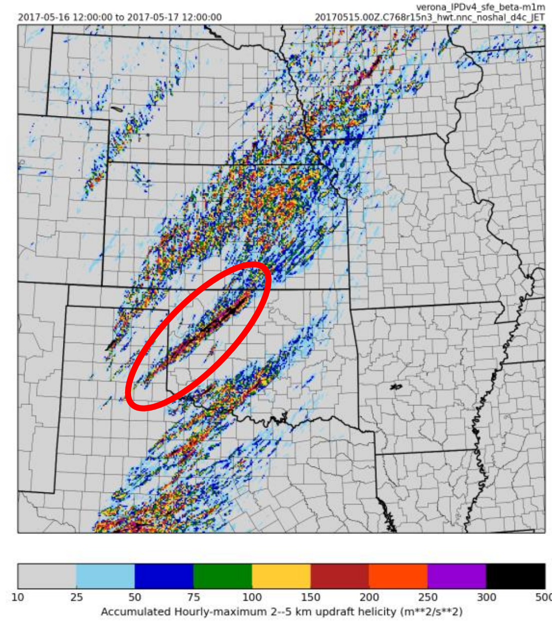
Large Variability among WRF multi-physics ensemble forecasts of CAPS

Maximum 2–5 km UH
12Z 16 May–12Z 17 May

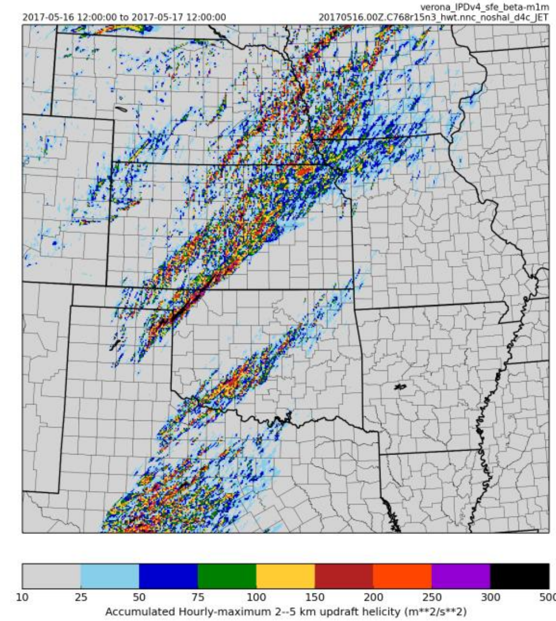
Observations



fvGFS Init 00Z 15 May



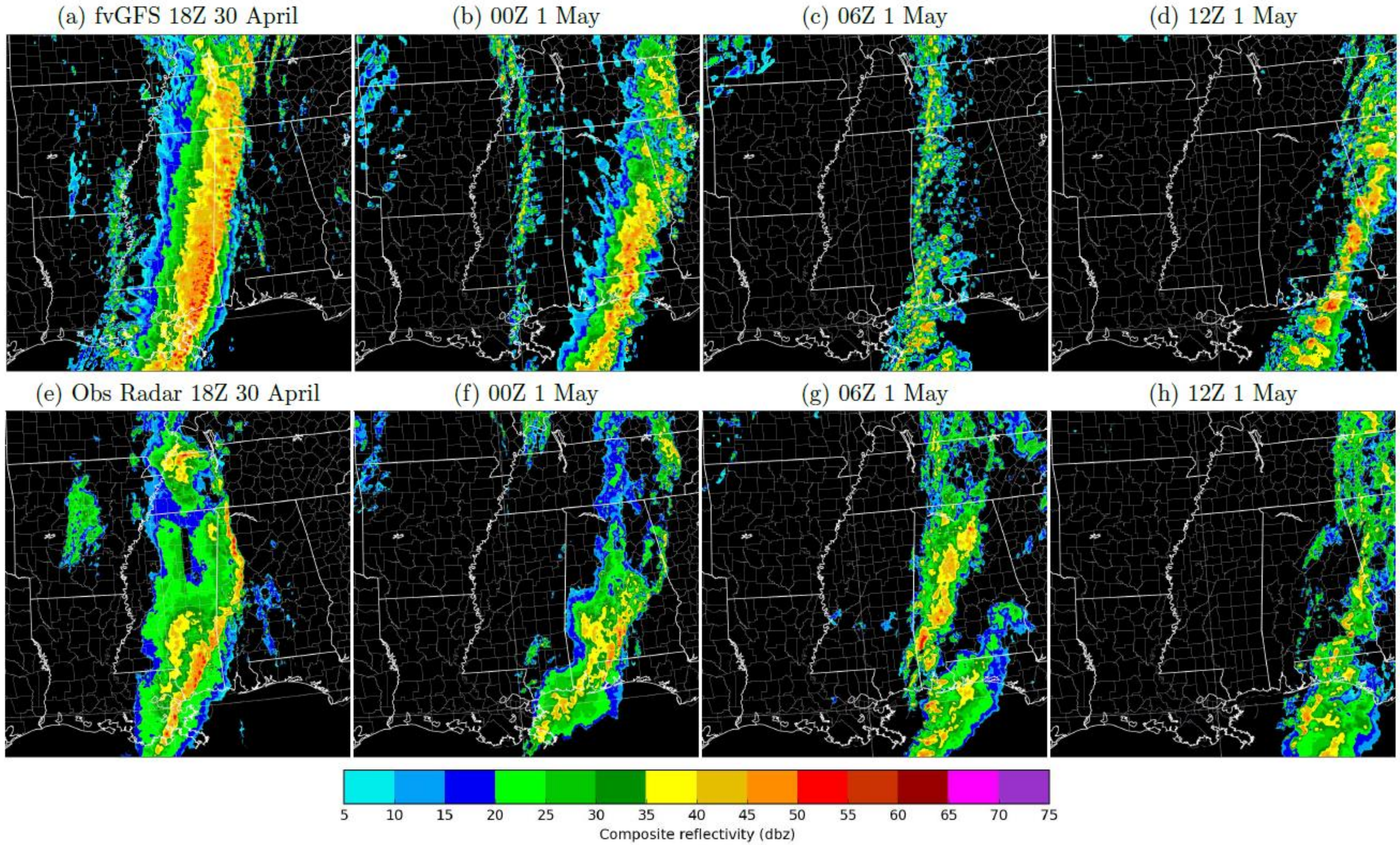
fvGFS Init 00Z 16 May



GFDL 15 May forecast had substantially more intense UH tracks in west OK and North/NW TX

Soil moisture condition?

FV3_GFDL forecast (upper) and observed composite reflectivity from 00 Z April 30



t = 18 h t = 24 h t = 30h t = 36 h

Figure 12. Composite reflectivity from fvGFS forecast initialized 00Z 30 April 2017 (top row, a–d) and observations (bottom row, e–h). Henceforth, all 3-km model output depicts shaded unsmoothed native nested-grid cells, unless otherwise stated.

From Harris et al. (2018 JGR Submitted)

FV3_GFDL forecast (upper) and observed composite reflectivity from 00 Z May 27
A severe derecho case

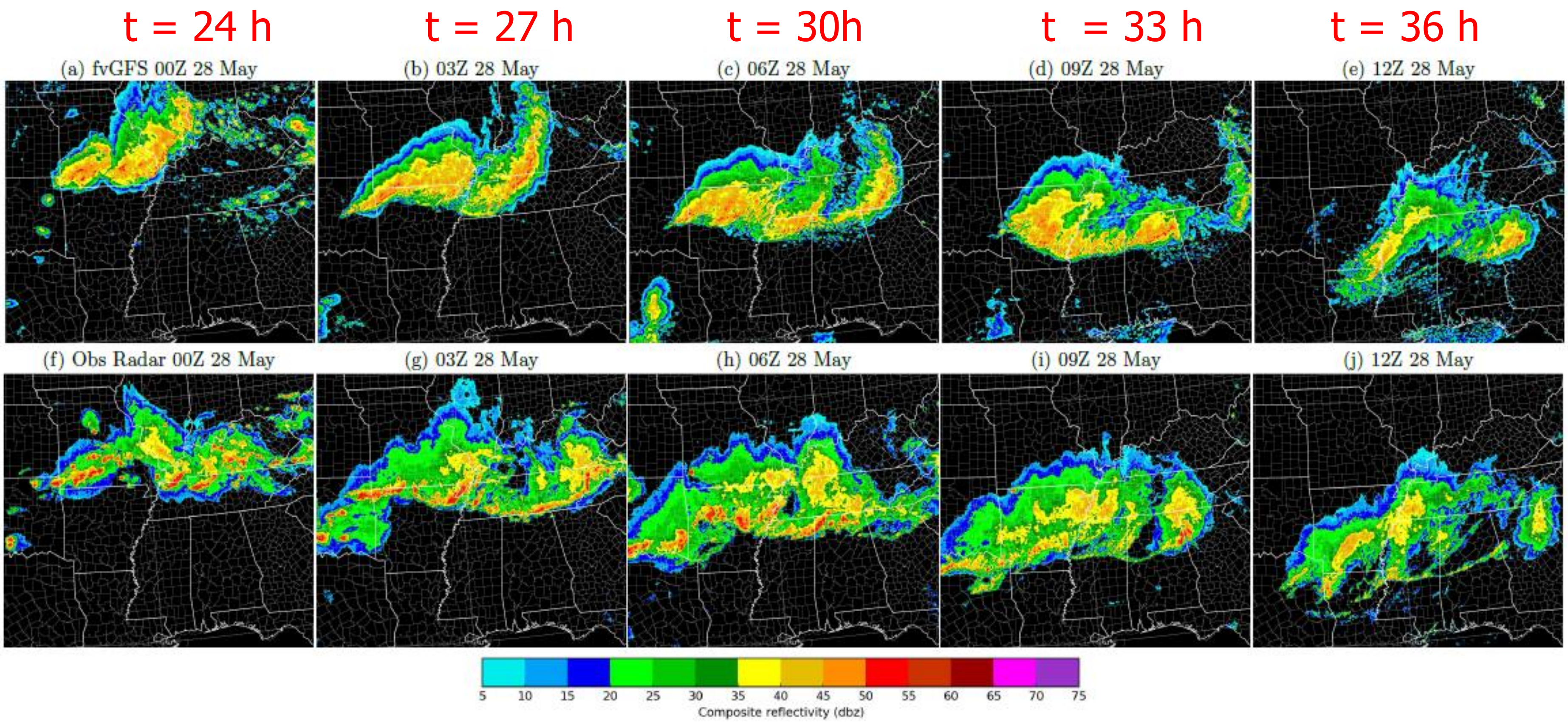


Figure 19. As in Figure 12, but for the forecast initialized 00Z 27 May 17.

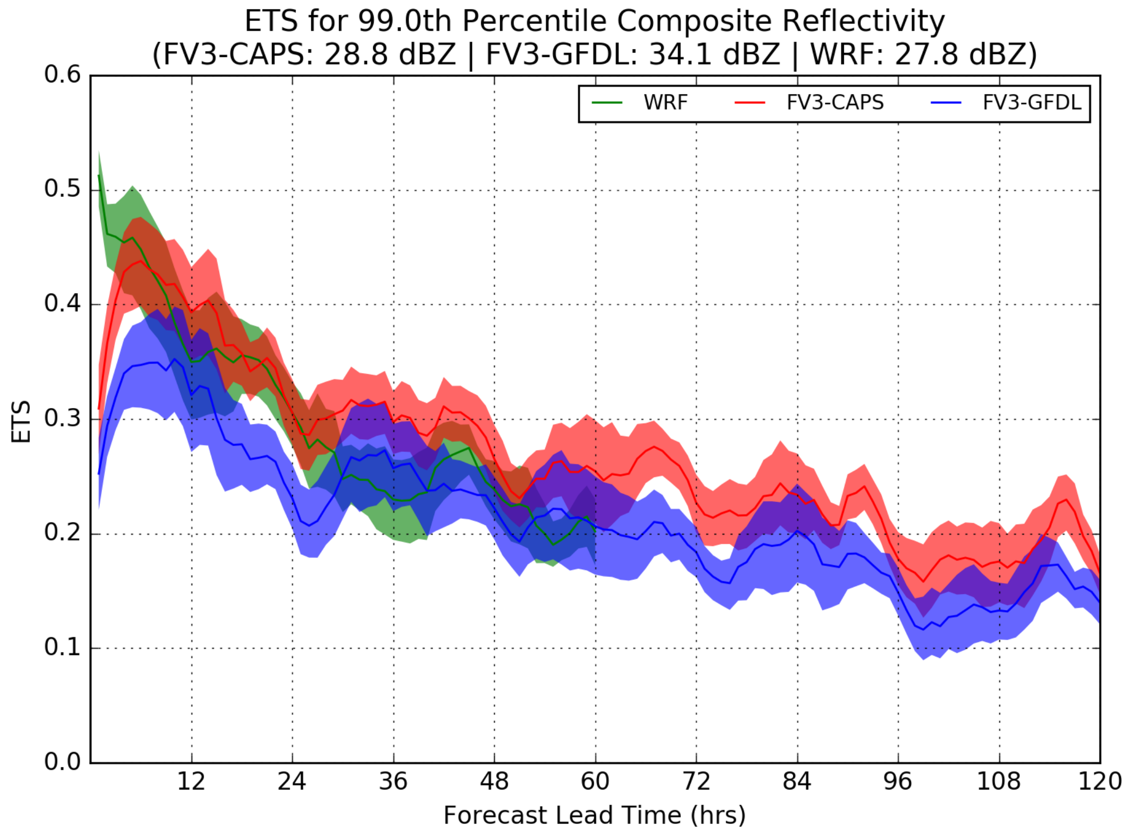
From Harris et al. (2018 JGR Submitted)

CAPS Storm-Scale Ensemble Control Member for 2017 HWT SFE

- WRF-ARW
- MYJ PBL
- Thompson Microphysics
- RRTMG Radiation (SW & LW)
- Noah land-surface model
- No cumulus parameterization
- IC: 00 UTC NAM analysis background + ARPS 3DVAR/cloud analysis with radar data
- LBC: NAM 00 UTC forecasts

Neighborhood ETSs of Composite Reflectivity for 99 Percentile
23 days, FV3_CAPS, FV3_GFDL and WRF_cn for HWT

R = 30 km

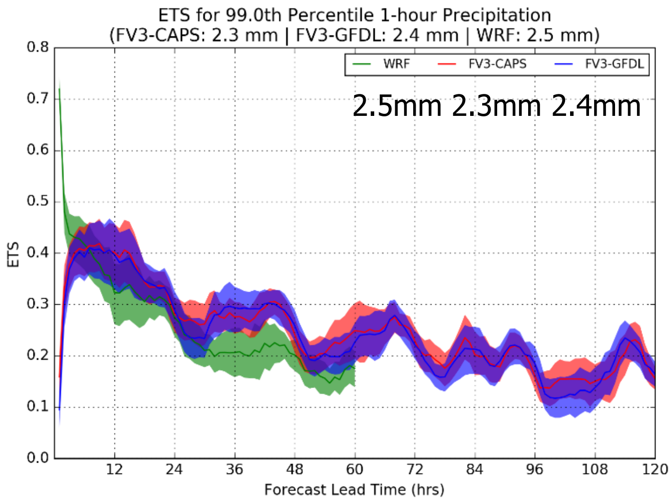


We see much larger difference between FV3_CAPS and FV3_GFDL in terms of composite reflectivity

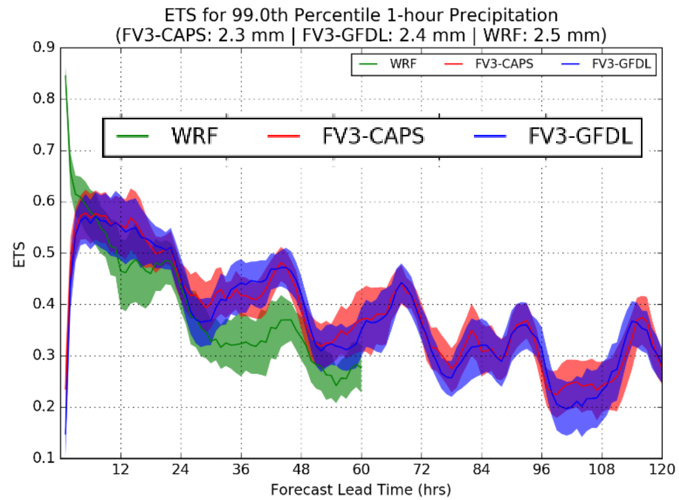
Neighborhood ETSs of hourly precipitation
23 days, FV3_CAPS, FV3_GFDL and WRF_cn for HWT

99th Percentile
Intense
(~2.4 mm/s)
precipitation

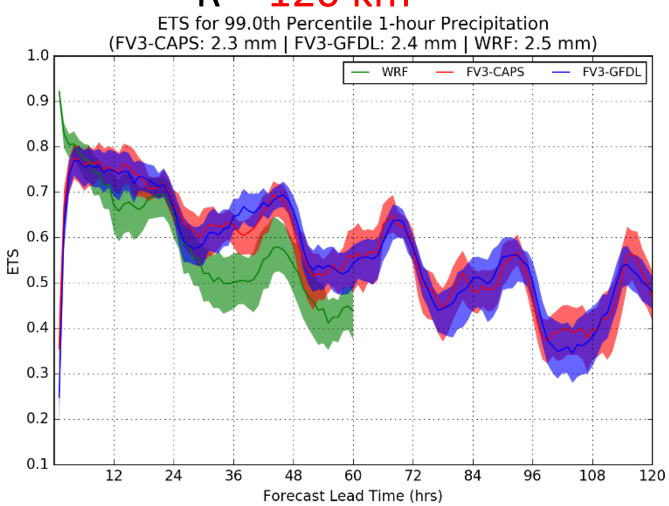
R = 30 km



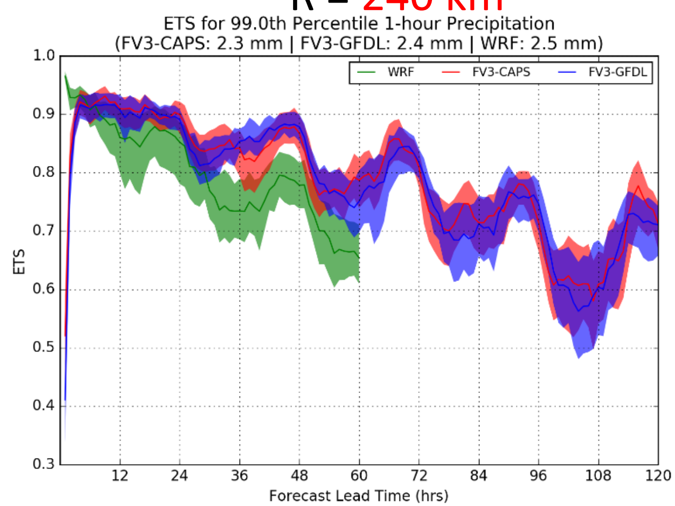
R = 60 km



R = 120 km



R = 240 km

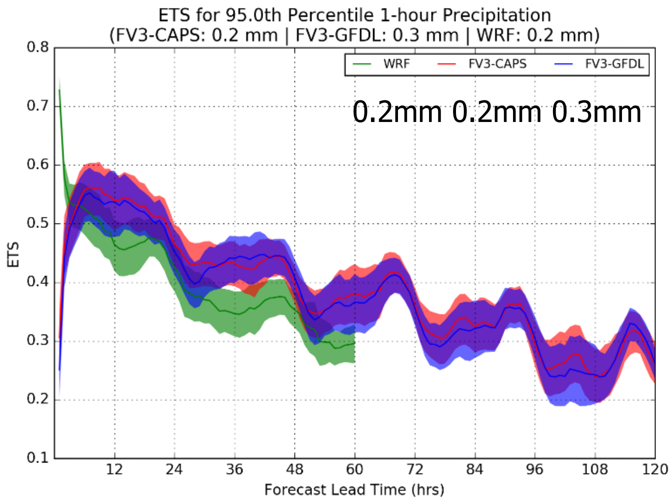


Shading is 90% confidence interval on bootstrap resampled (10000 times) set of forecasts

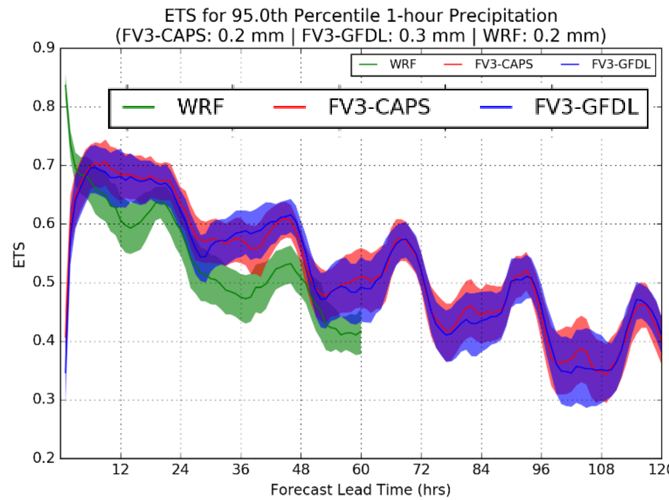
Neighborhood ETSs of hourly precipitation
23 days, FV3_CAPS, FV3_GFDL and WRF_cn for HWT

95th Percentile Light (~0.2mm/h) precipitation

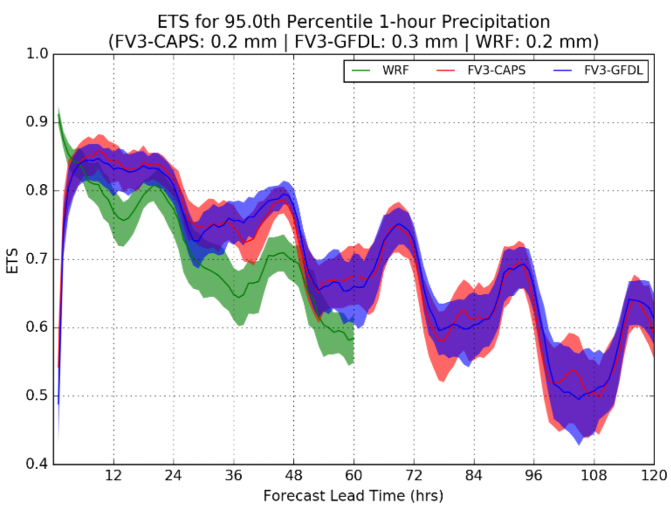
R = 30 km



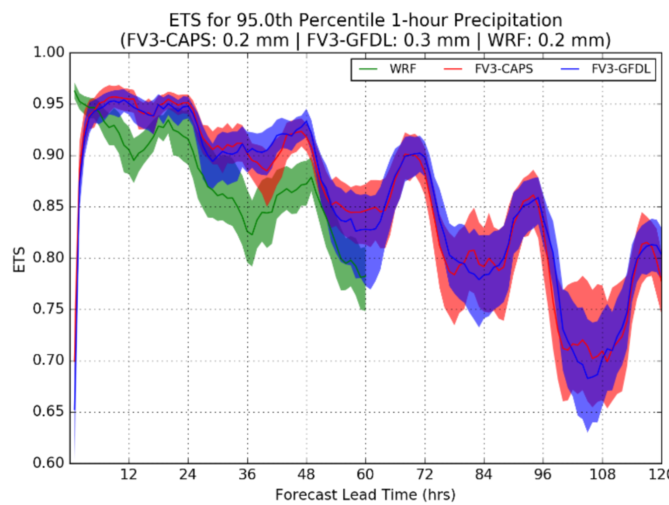
R = 60 km



R = 120 km

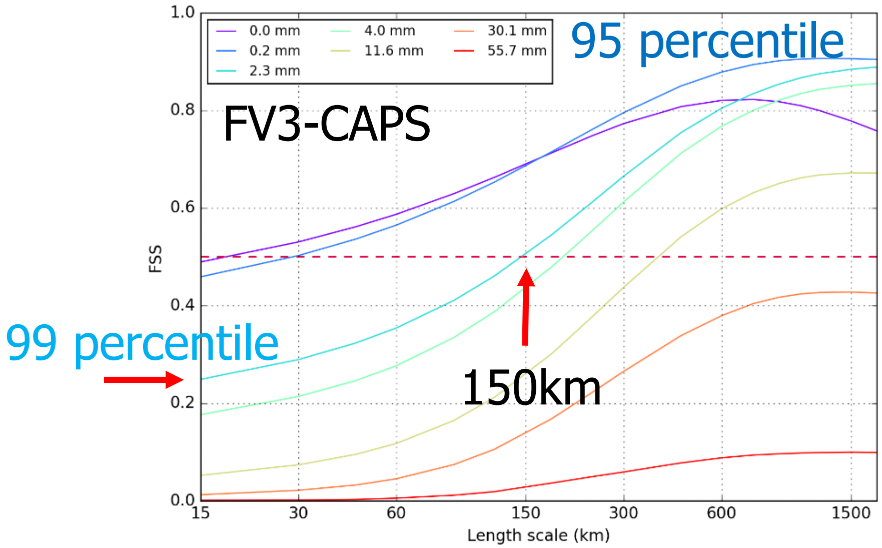


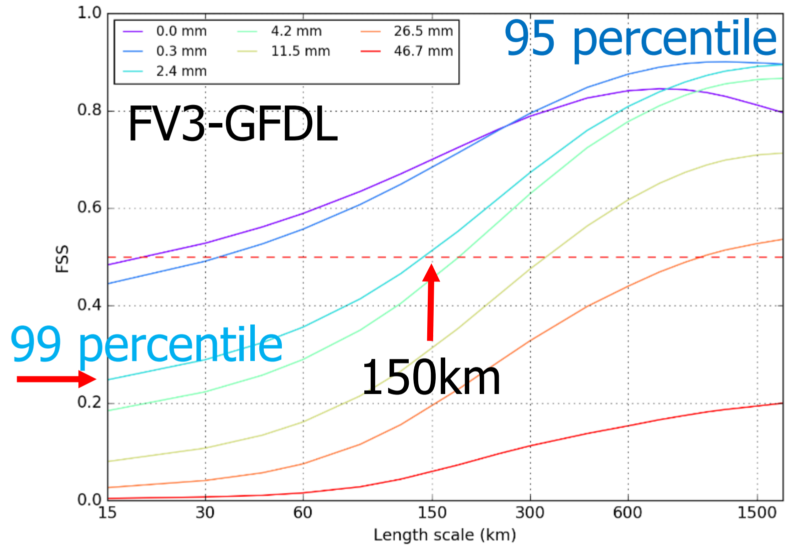
R = 240 km

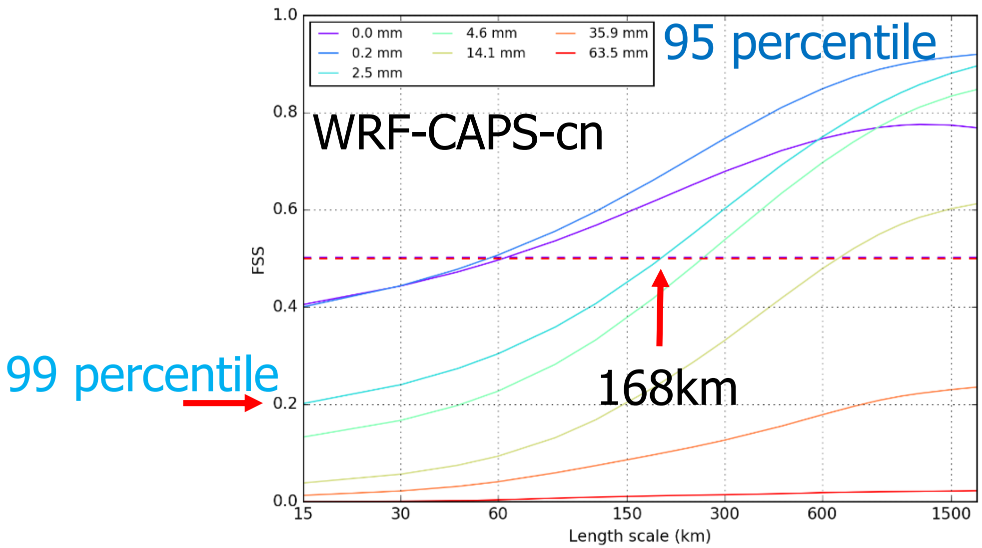


Shading is 90% confidence interval on bootstrap resampled (10000 times) set of forecasts

Fractional Skill Scores of hourly precipitation for different scales and thresholds Aggregated over all forecasts from 6 to 84 hours



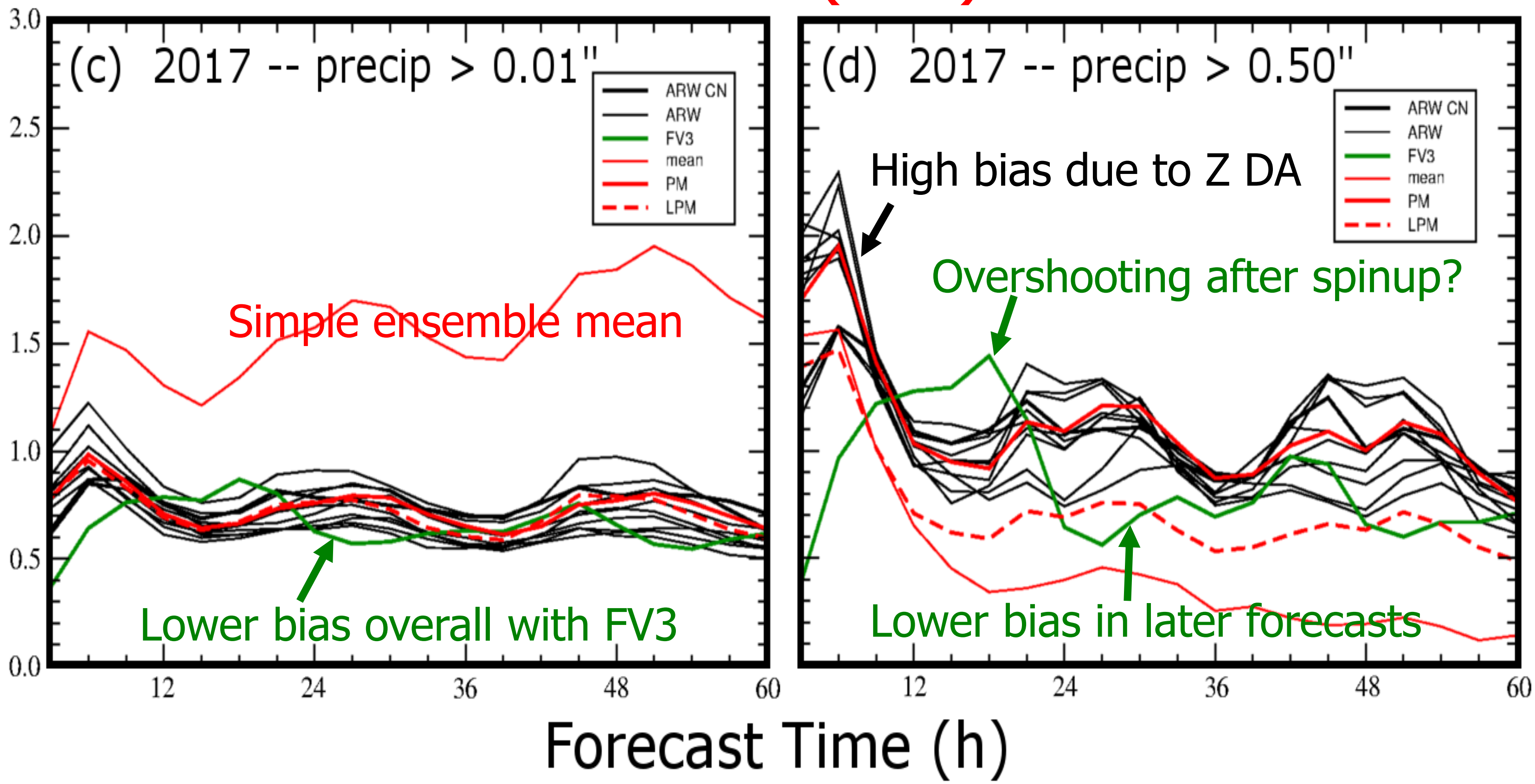




FV3 forecasts appear more skillful in terms of

CAPS Forecasts for HMT FFaIR Experiment

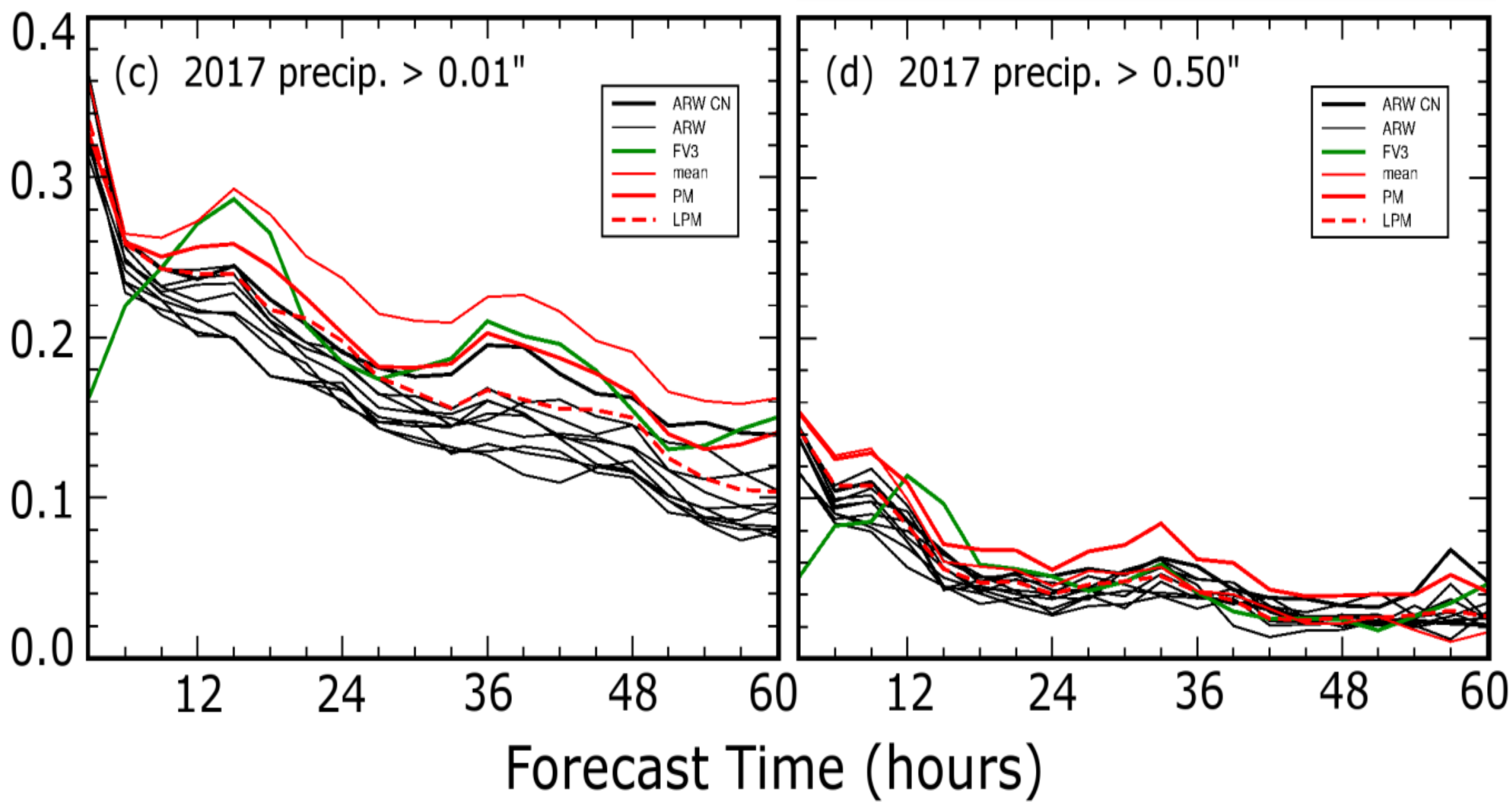
Bias error (mm)



Bias (in mm) of 3-h accumulated precipitation, verified against observed MRMS precipitation data, for all operational days of the CAPS HMT ensemble forecasts

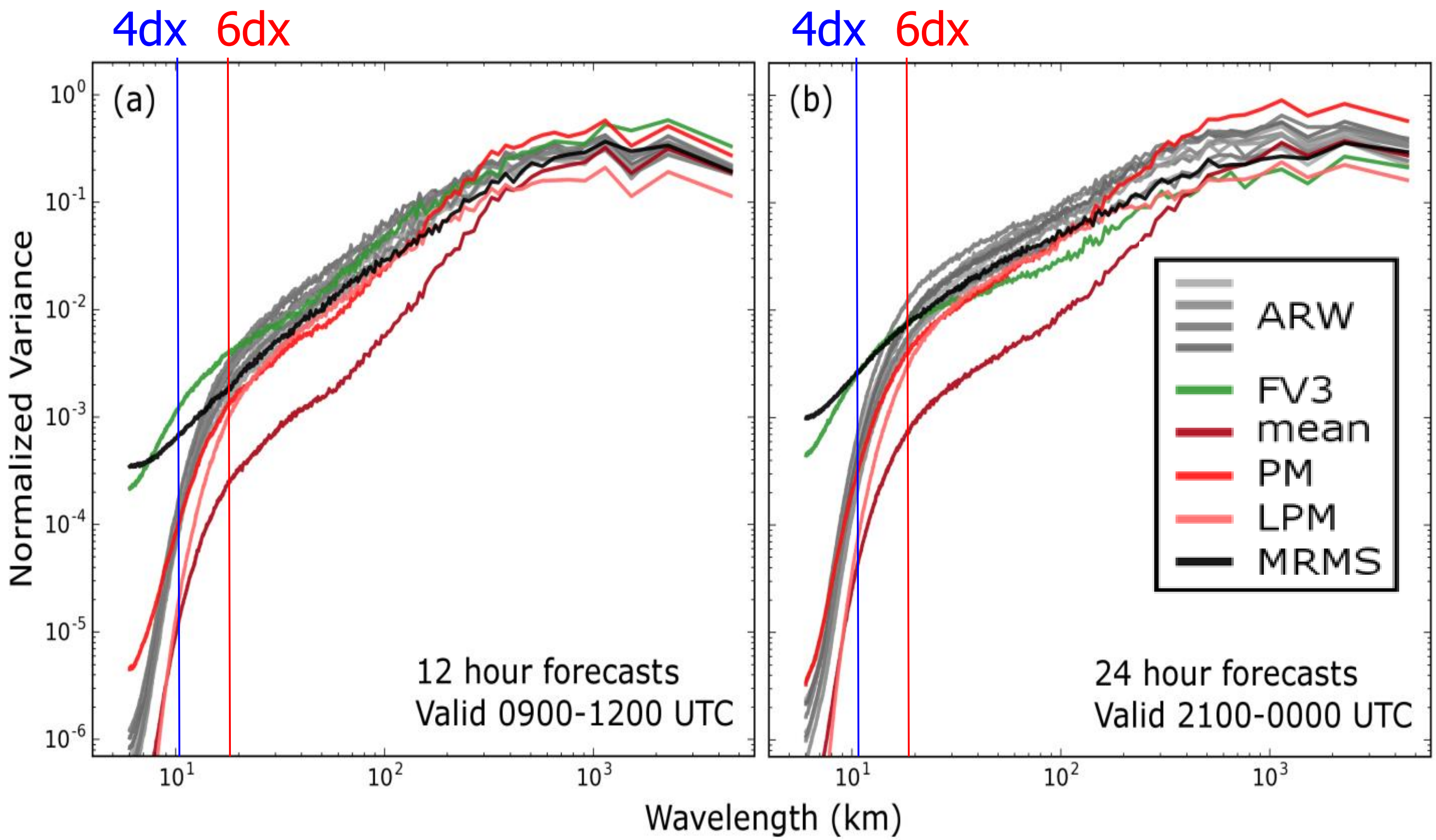
CAPS Forecasts for HMT FFaIR Experiment

ETS of 3 hourly precipitation



- FV3 ETS scores higher than all ARW members for light precipitation
- FV3 ETS scores about average among the ARW members for heavier precipitation
- No biases correction with these scores

Normalized variance spectra of 3-hourly precipitation from CAPS WRF ensemble and FV3 forecasts for 2017 HMT FFaIR Experiment



FV3 is able to retain more energy in the sub-10km scales than the WRF-ARW forecasts, resulting in a spectrum slope that is much closer to that of observations (black)!!

Nested Grid 250 hPa KE Spectra

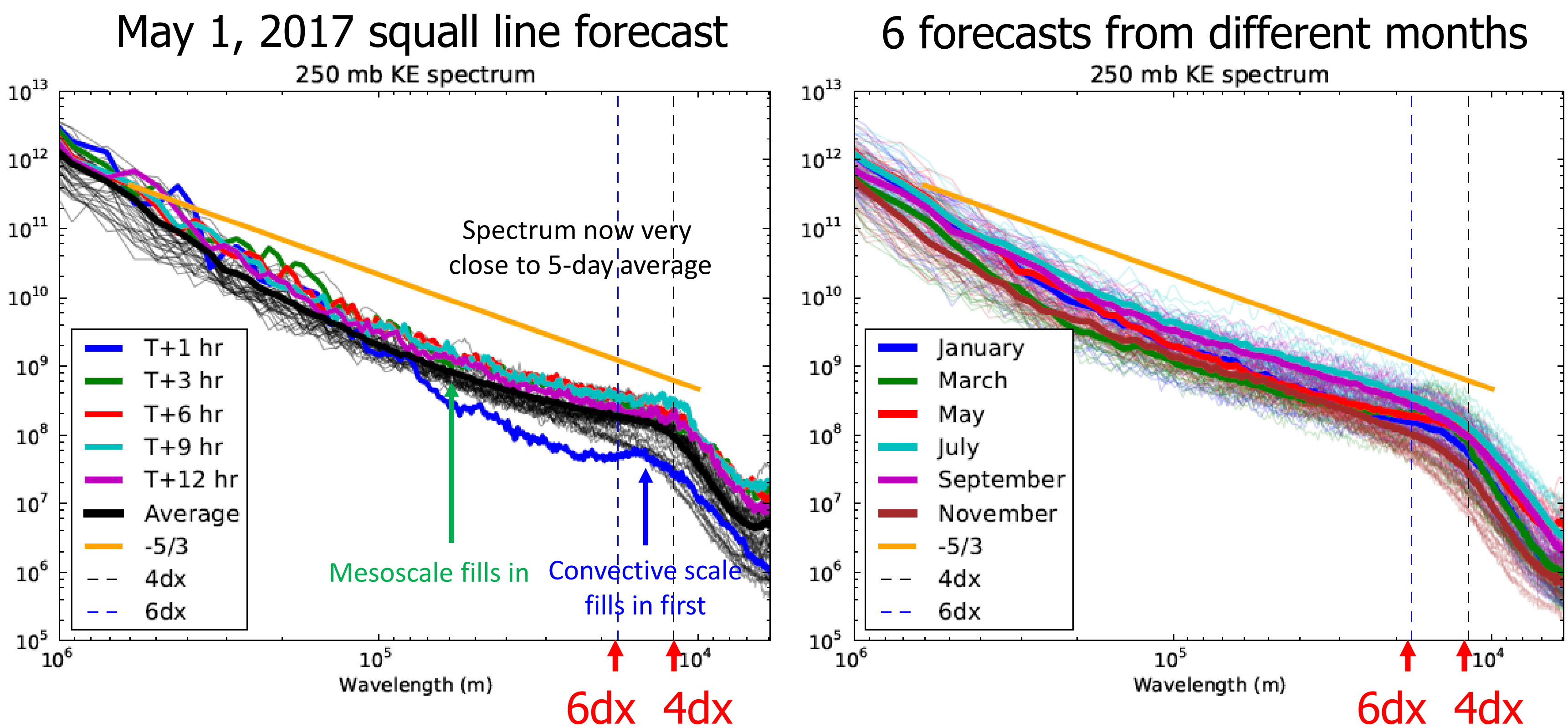


Figure 14. Nested-grid 250 mb kinetic energy spectra (m^3 s^{-2}). Left: forecast initialized 01 May 2017. Light gray lines are plotted every three hours starting at 15Z on 01 May; the average of these times is shown as a heavy black line. Right: Time-averaged 250-mb kinetic energy spectra for six different forecasts at different times of the year.

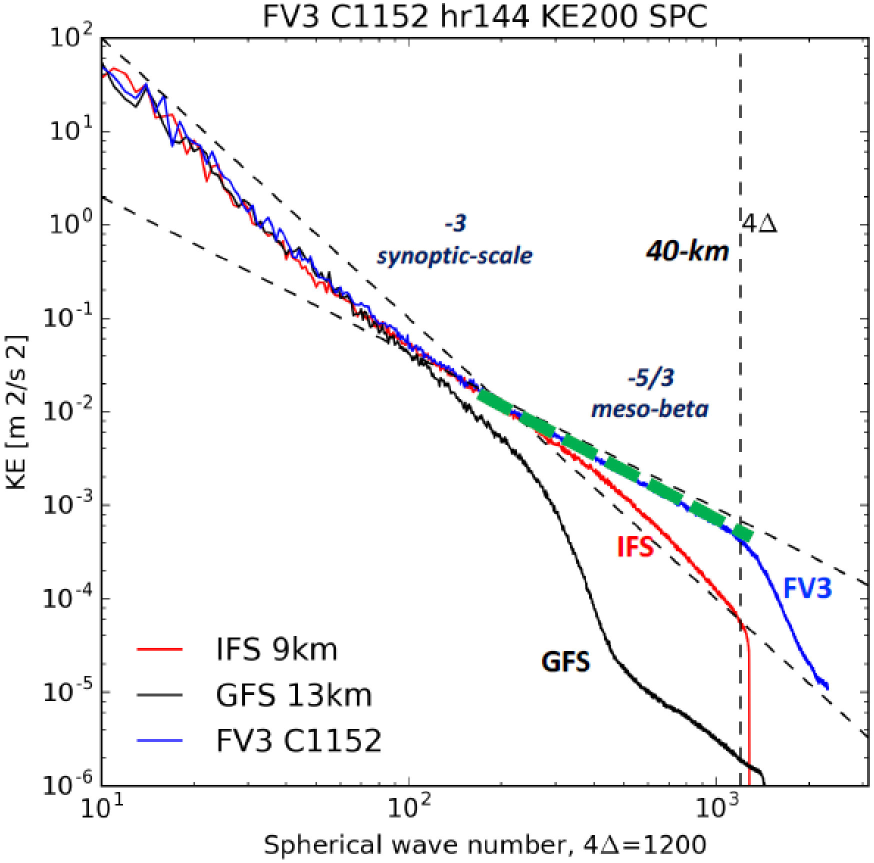
FV3 seems to be resolving spectra up to 4 dx

From Lucas Harris

Behavior consistent with FV3 global runs

Most operational global NWP models should be "mesoscale resolving" with dx ~10 km
How well do ECMWF-IFS, NCEP-GFS, and FV3-GFS actually resolve the meso-scale?

200-mb KE spectra



***Total diffusion = implicit + explicit diffusion**

- FV3 at C1152 (9km, roughly the same as IFS resolves the"-5/3" meso-beta (20-200 km) spectrum
- The IFS has lower energy in the meso-scale; but it does follow "-3" spectrum (synoptic scale) well
- The GFS has the least amount of energy in the meso-scale (3 orders of magnitude smaller than FV3 and the theoretical value)

From SJ Lin (Global Forecasts)

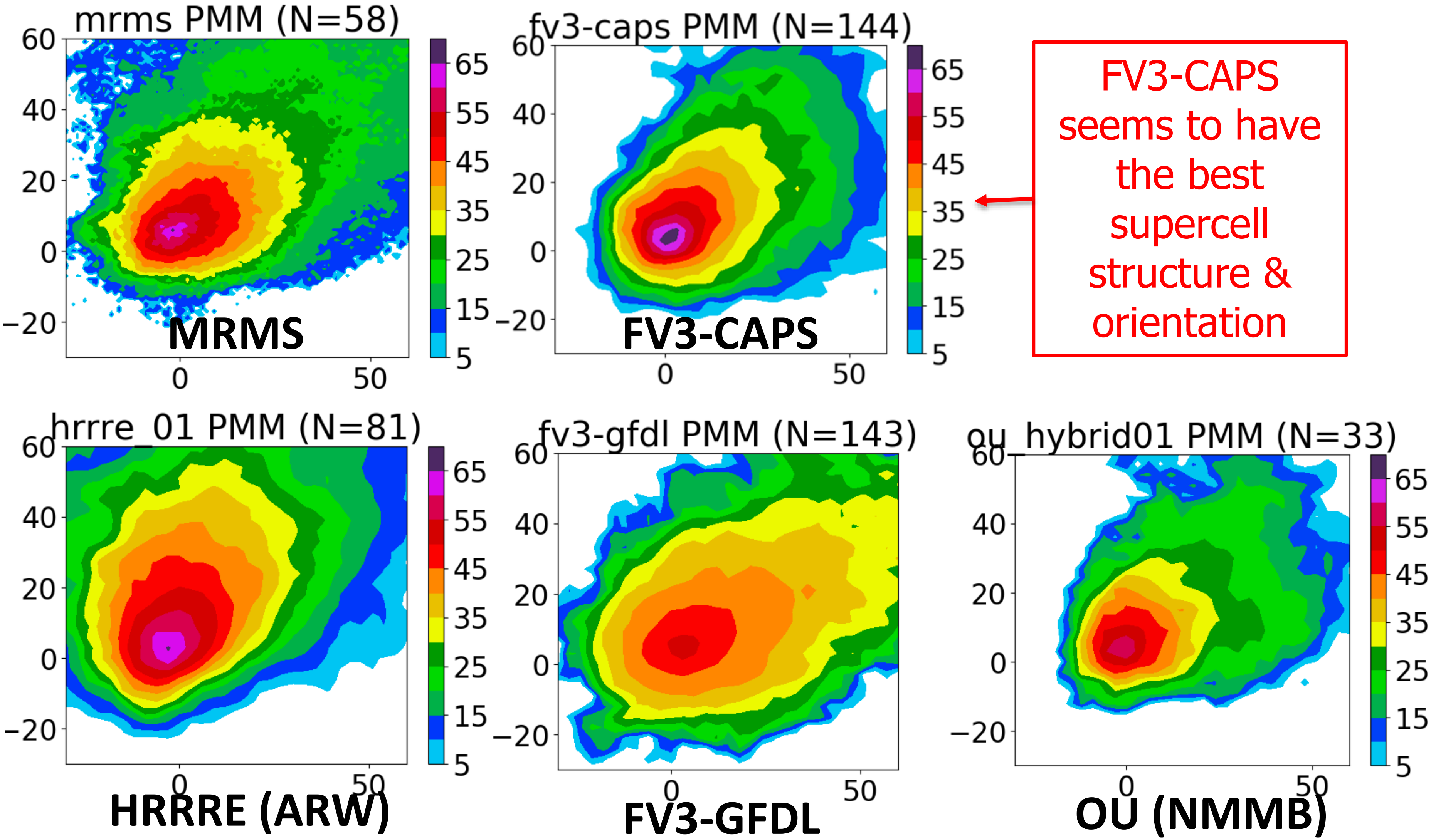
Supercell composites

From Corey Potvin/NSSL

- How well does model supercell structure match observations?
- Identify supercells using DBZCOMP, UH percentile objects
- Compute probability-matched median (PMM) of selected field(s)

PMM Composite dBZ (preliminary)

break down well by dynamical core



Orientation too northward

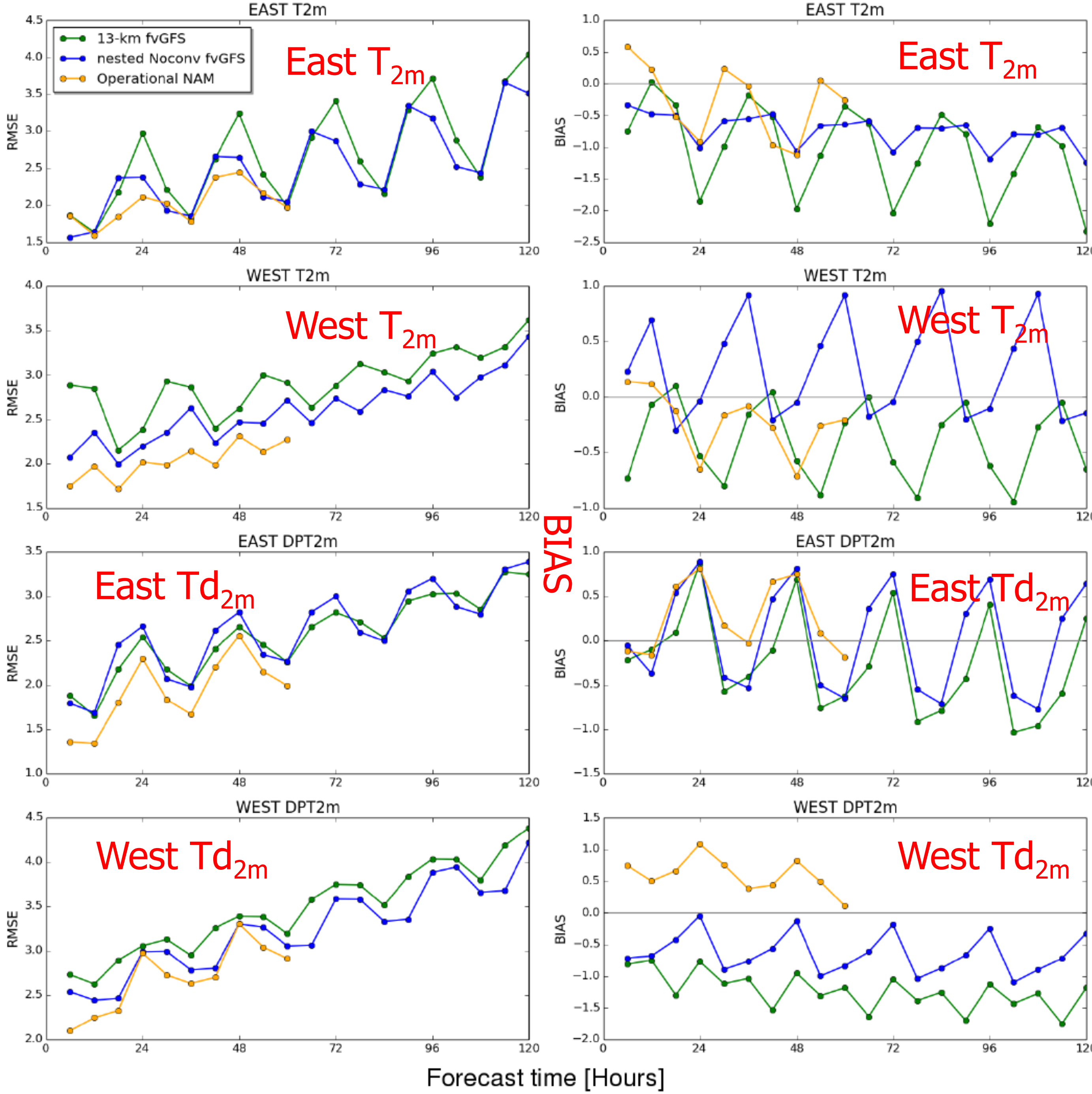
From Corey Potvin/NSSL

GFDL FV3 forecasts during 2017 HWT Spring Experiment

13-km grid
3-km grid
3-km NAM

Larger sfc T and Td forecast errors that 3-km NAM

Likely due to poorer soil model initialization



From Lucas Ferris

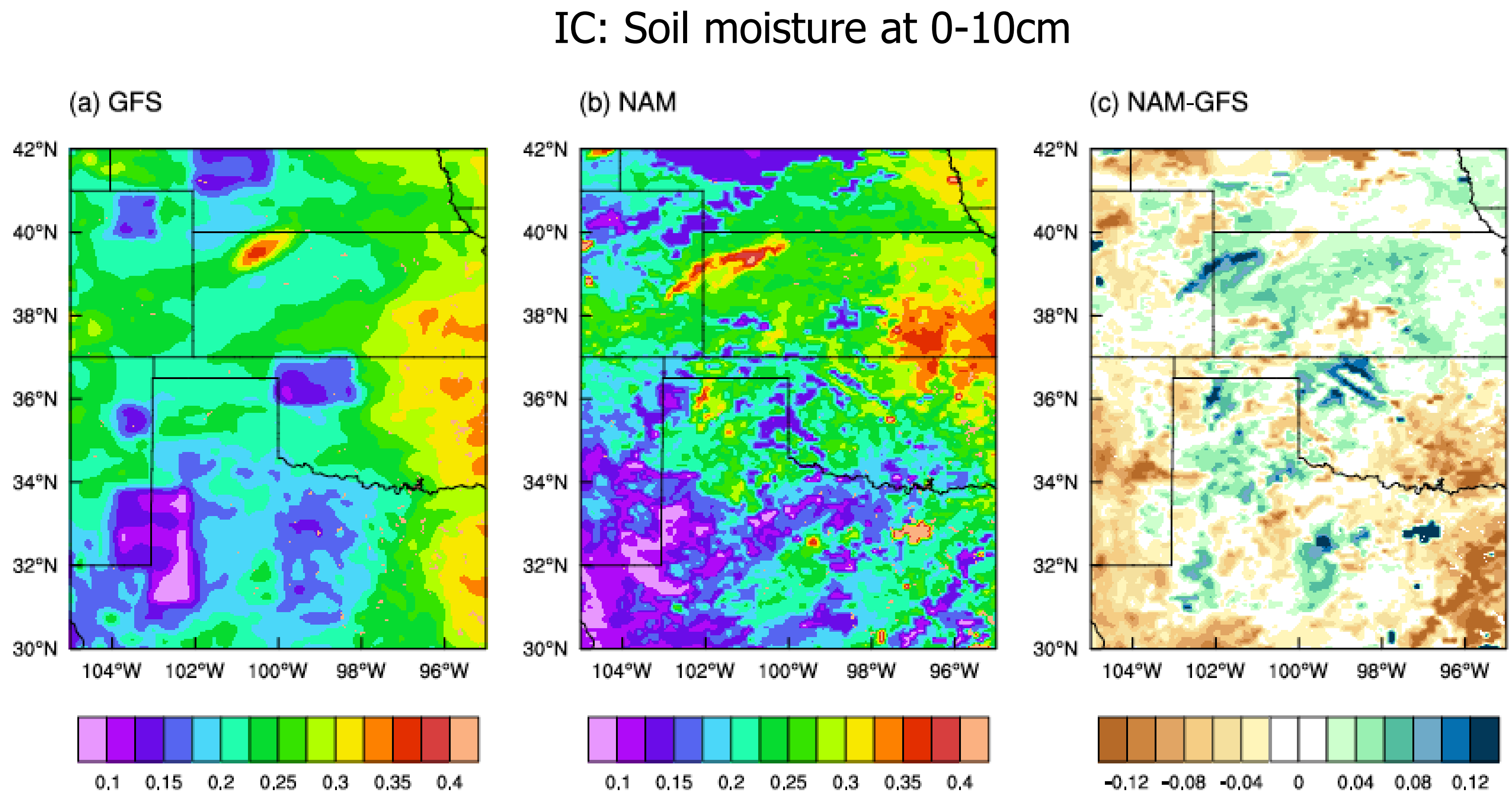
Summary and Thoughts (Tentative)

- 3-km nested grid FV3 forecasts during HWT SFE and HMT FFaIR appear at least comparable with WRF-based forecasts, and some aspects are better than CAPS WRF control and ensemble members;
    - The difference between FV3 and WRF forecasts appear within the variability/uncertainty of WRF ensemble members;
    - FV3 reflectivity using GFDL microphysics was too smooth/has low-biases, while CAPS forecasts with Thompson scheme were more inline with other WRF forecasts and with MRMS data.
    - GFDL hourly precipitation forecasts do not suffer low-bias much.
- Neighborhood ETS scores and FSSs of FV3 are higher than CAPS WRF control member (even without the benefit of radar data), and the differences are more significant beyond day one.
    - The CAPS WRF control is a better performing one among CAPS ensemble members during HMT FFaIR.

Summary and Thoughts (Tentative)

- FV3 is able to resolve spectra up to 4 dx wavelength while WRF stops at 6 dx
  - FV3 has higher UH values than WRF ARW, no involving any averaging in its vorticity and w calculations, and perhaps also due to higher effective resolution;
- FV3_CAPS seems to produce the best composite supercell structures and orientation
- FV3's surface T and Td forecast errors are ~0.5 to 1 K larger than those of 3-km NAM
  - Need better initialization of soil conditions? Better PBL scheme?
- FV3 versions run in 2017 used a relatively poor PBL scheme and from coarser resolution GFS analyses
  - How much difference do more sophisticated PBL and MP schemes make?
  - How much difference does the IC make? Importance of DA
  - How much difference does the radar data assimilation make?
- Is FV3 ready to for CAM forecasting?
  - Maybe. No major show stopper
  - But there is still much room for improvement (e.g., physics, DA, soil IC/LULC, consistency tuning) – that is actually a good thing.
- Many more aspects of forecasts need to be evaluated.

Soil moisture in GFS and NAM analyses at 00 UTC May 16, 2017



Low-resolution LULC from GFS

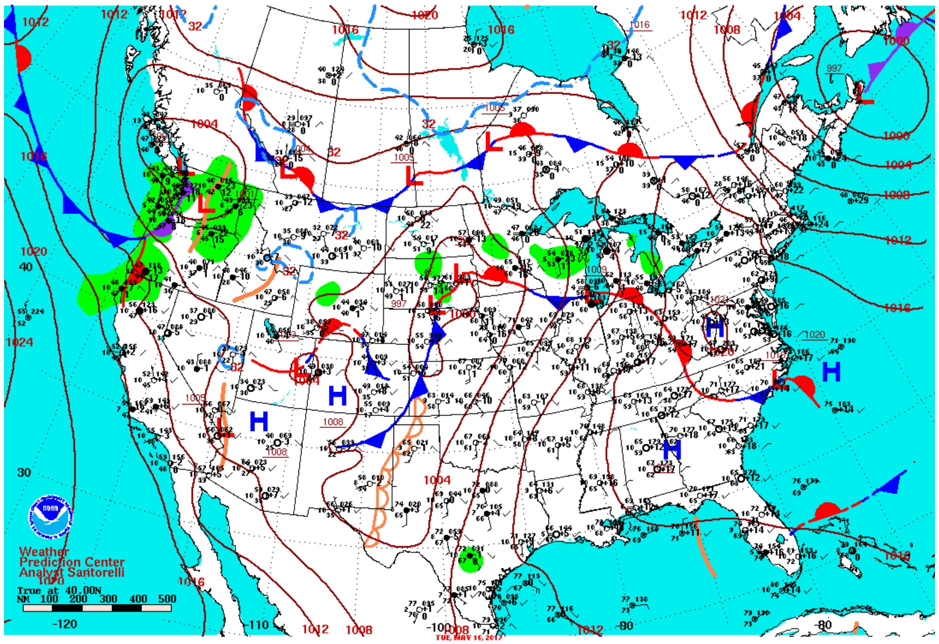
Physics schemes CAPS implemented in FV3

(based on the most recent version of FV3 from GFDL)

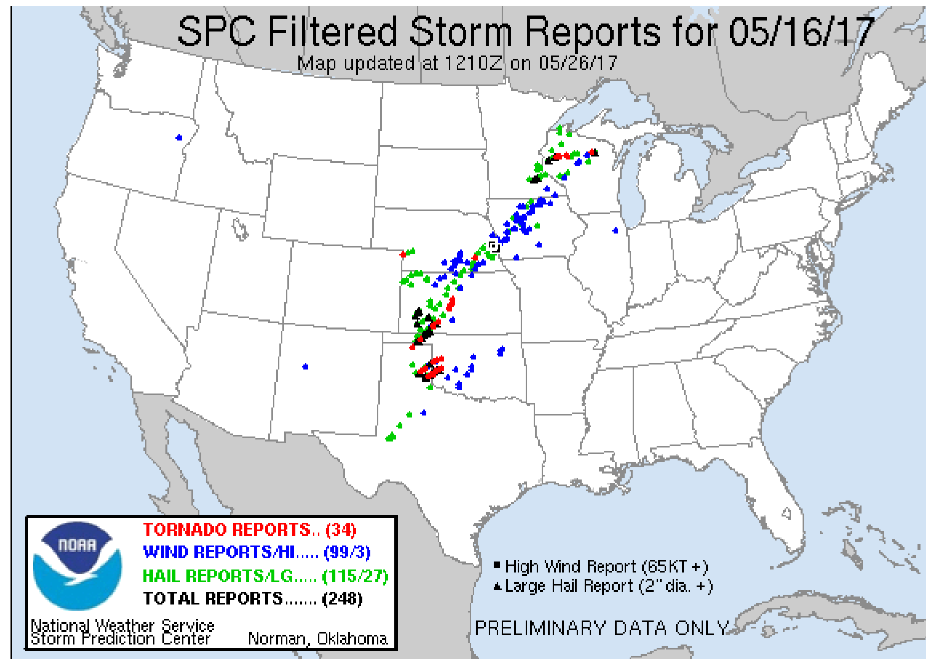
- **Microphysics Scheme**: Thompson, NSSL, MY, and Morrison
- **PBL Schemes**: YSU, Shinhong Scale-aware YSU, MYJ, MYNN (with scale-awareness), and E-epsilon (still under testing) - *all PBL schemes share the same surface layer scheme*
- **Cumulus Scheme**: New Tiedtke (implemented by Chunxi Zhang and available in WRF) with both deep and shallow convection.

May 16-17 Dry Line Case

(The day of poor FV3 forecast of dryline initiation of tornadic supercell storms



Surface Weather Map and Station Weather at 7:00 A.M. E.S.T.



FV3 Tests using HWT Grid

Thompson microphysics + different PBL schemes

Initial condition: GFS Analysis

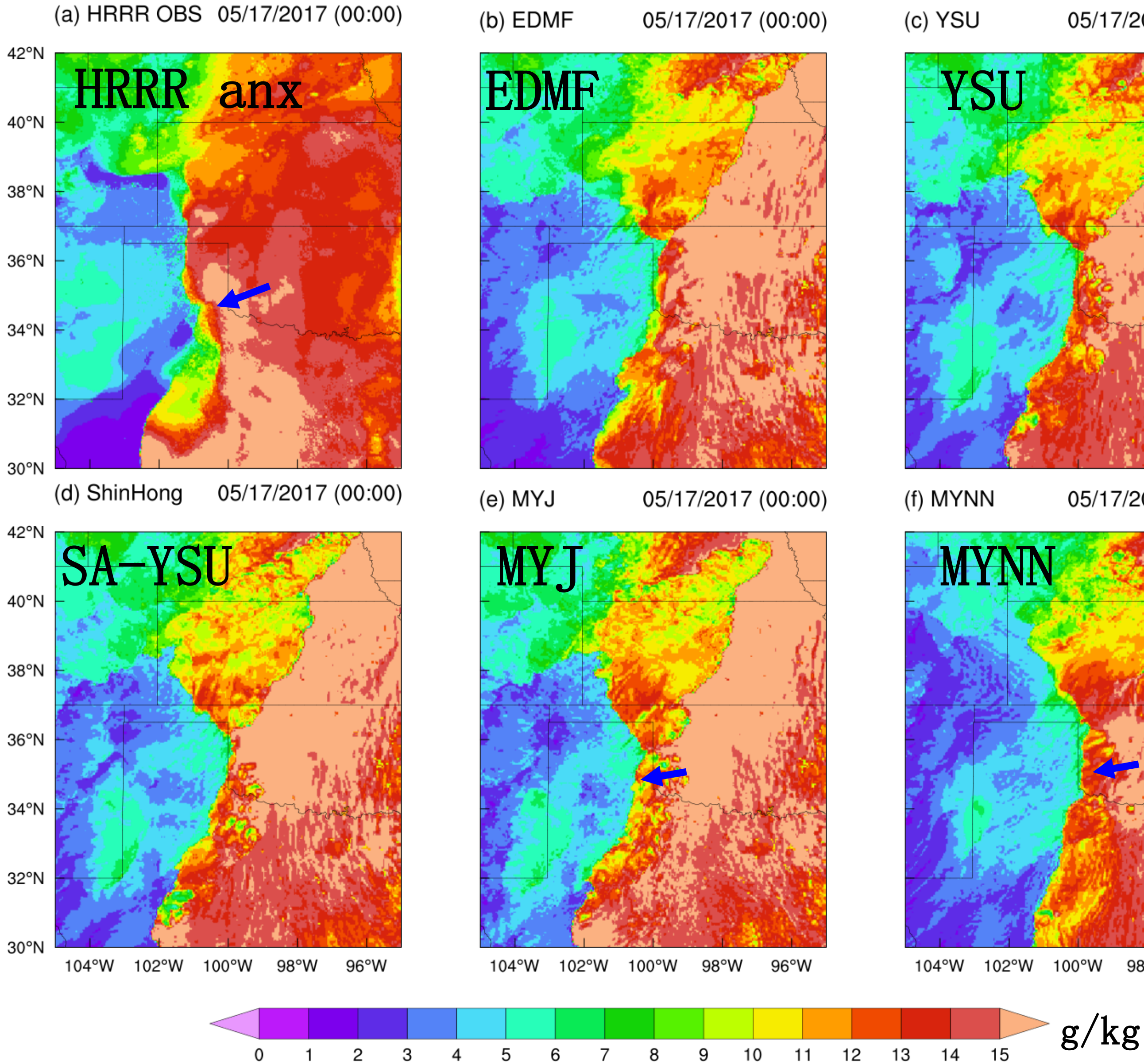
Initial time: 00 UTC 16 May 2017

Nested domain: ~3 km horizontal resolution

PBL schemes
EDMF (GFS operational scheme)
YSU
Shinhong (scale-aware YSU)
MYJ
MYNN (scale-aware)

Specific Humidity at 2 m Height (t = 24 h)

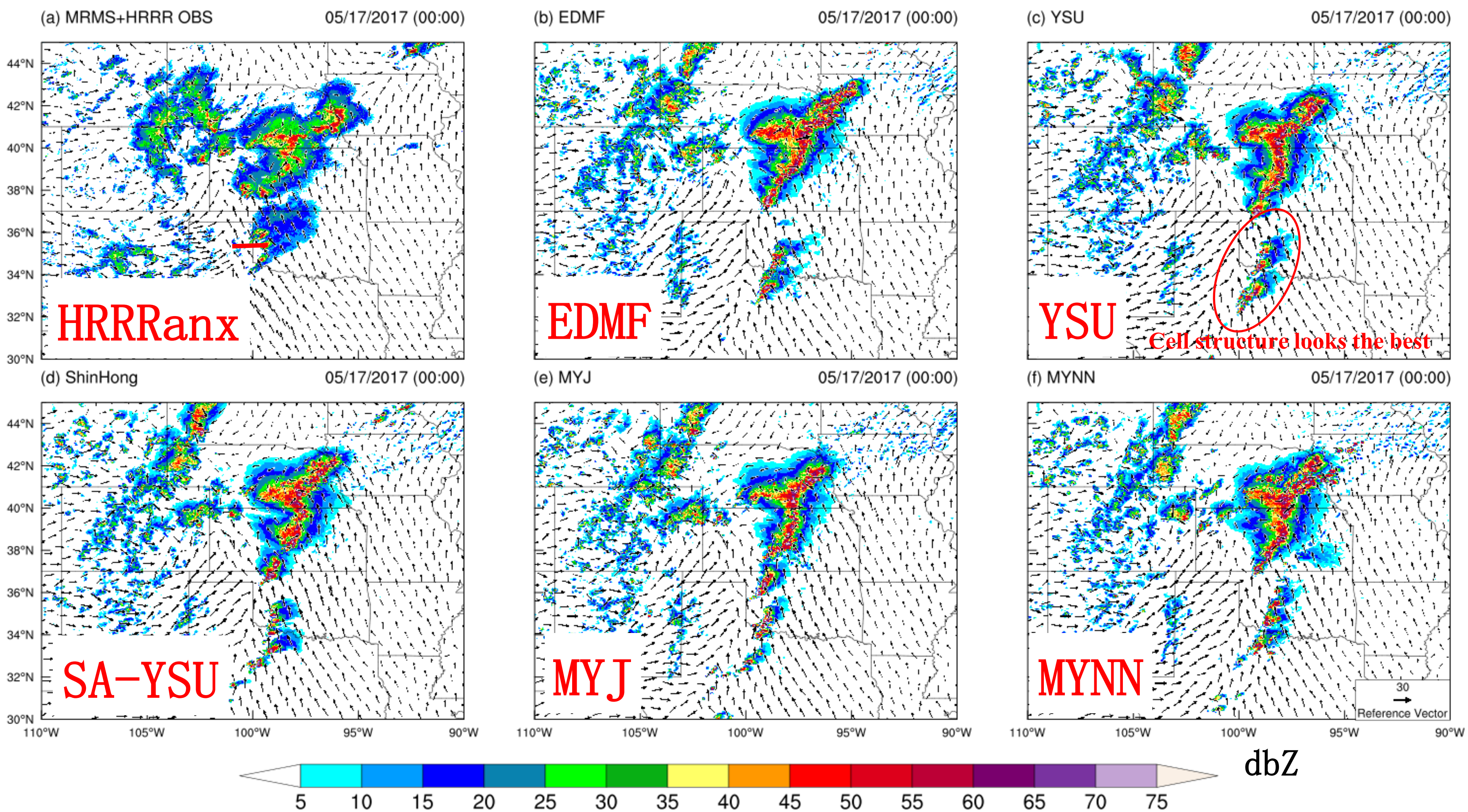
t=24h



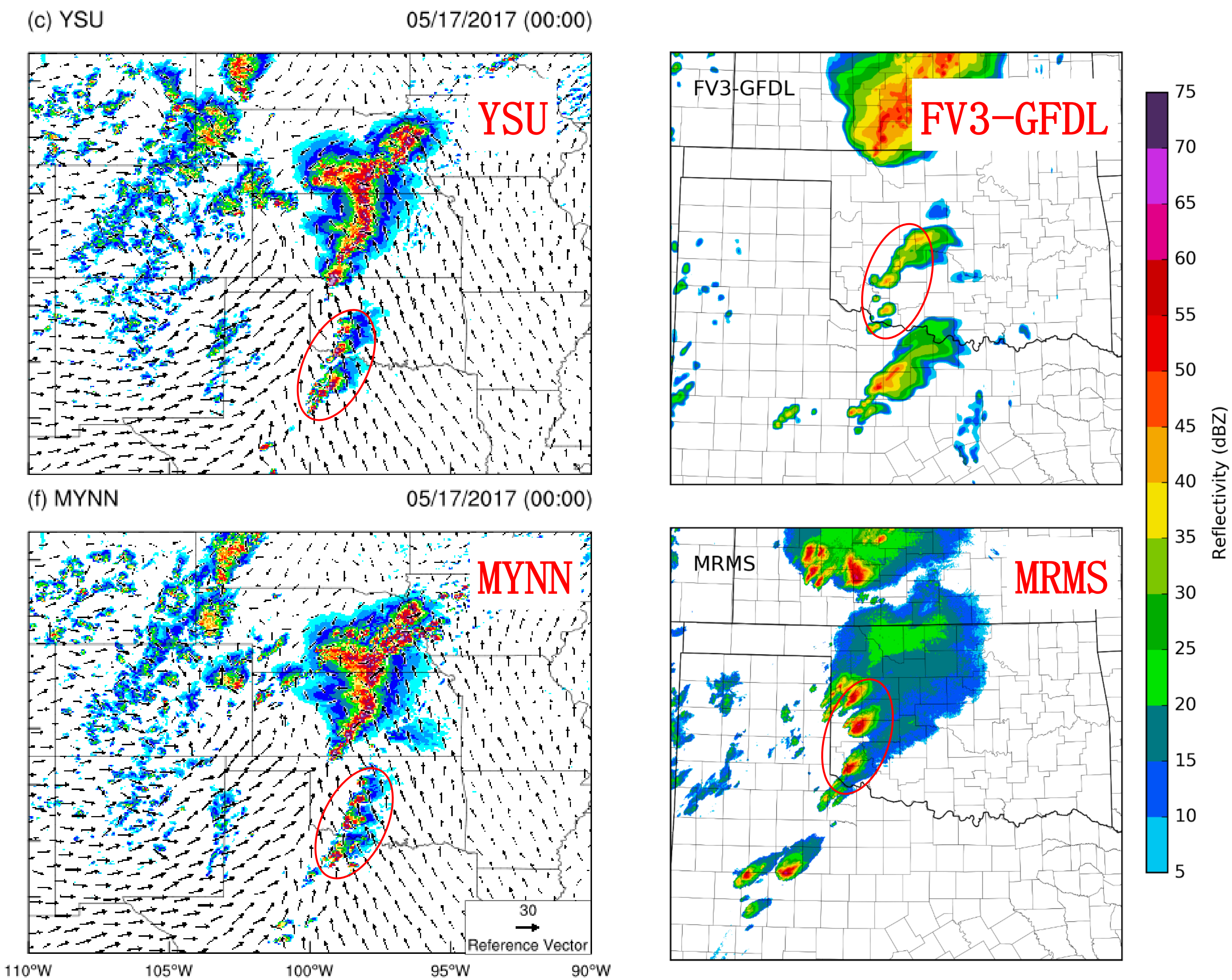
Not too different dryline locations in all PBL runs, with MYJ position west most

Composite Reflectivity & Surface Wind Vectors (HRRR anx+MRMS and FV3 Forecasts)

t=24h



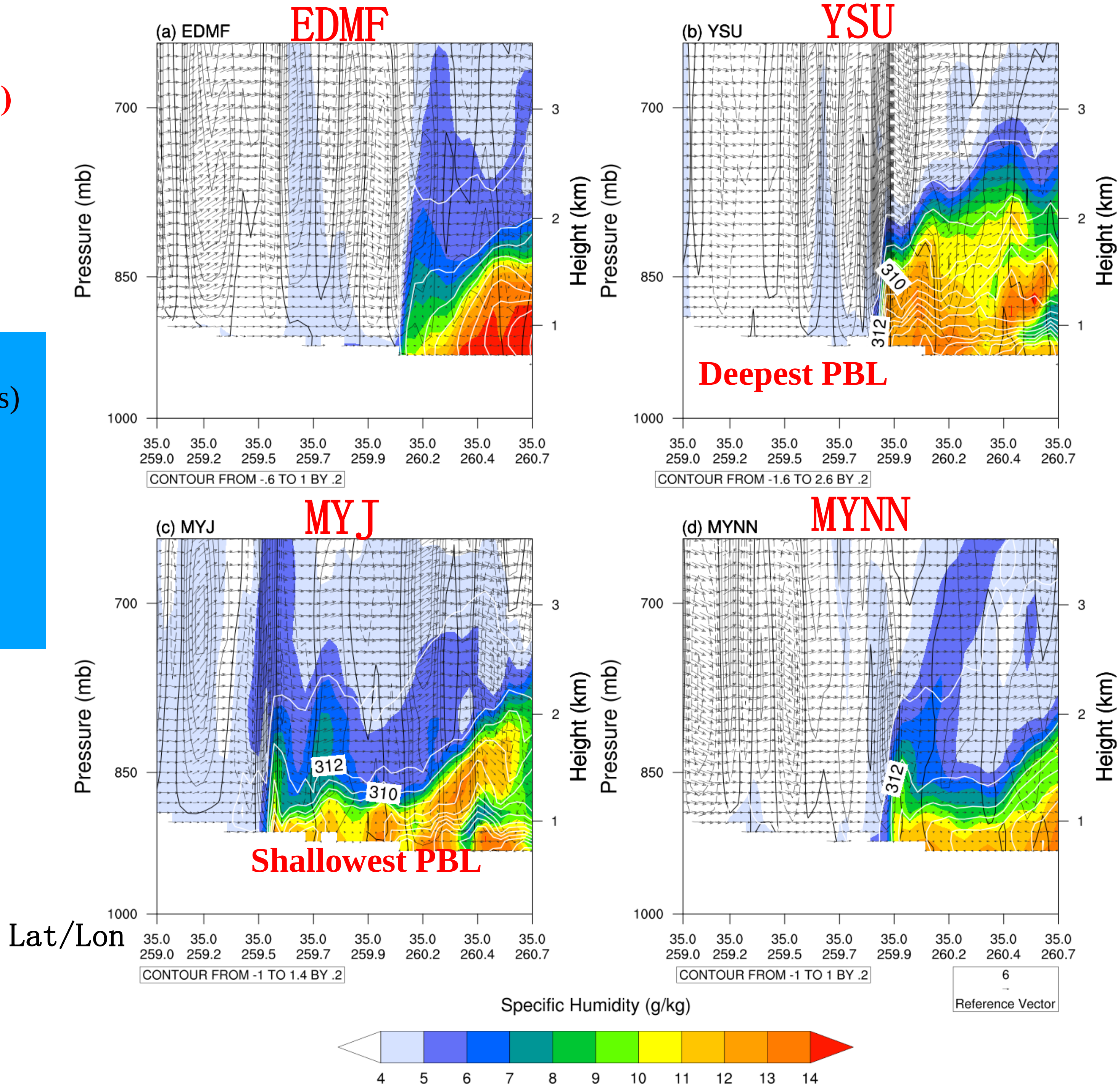
YSU and MYNN produced better supercell forecasts than others but still positioned too far east



(c) YSU
05/17/2017 (00:00)
YSU
(f) MYNN
05/17/2017 (00:00)
MYNN
30
Reference Vector
110°W
105°W
100°W
95°W
90°W
FV3-GFDL
FV3-GFDL
MRMS
MRMS
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
Reflectivity (dBZ)

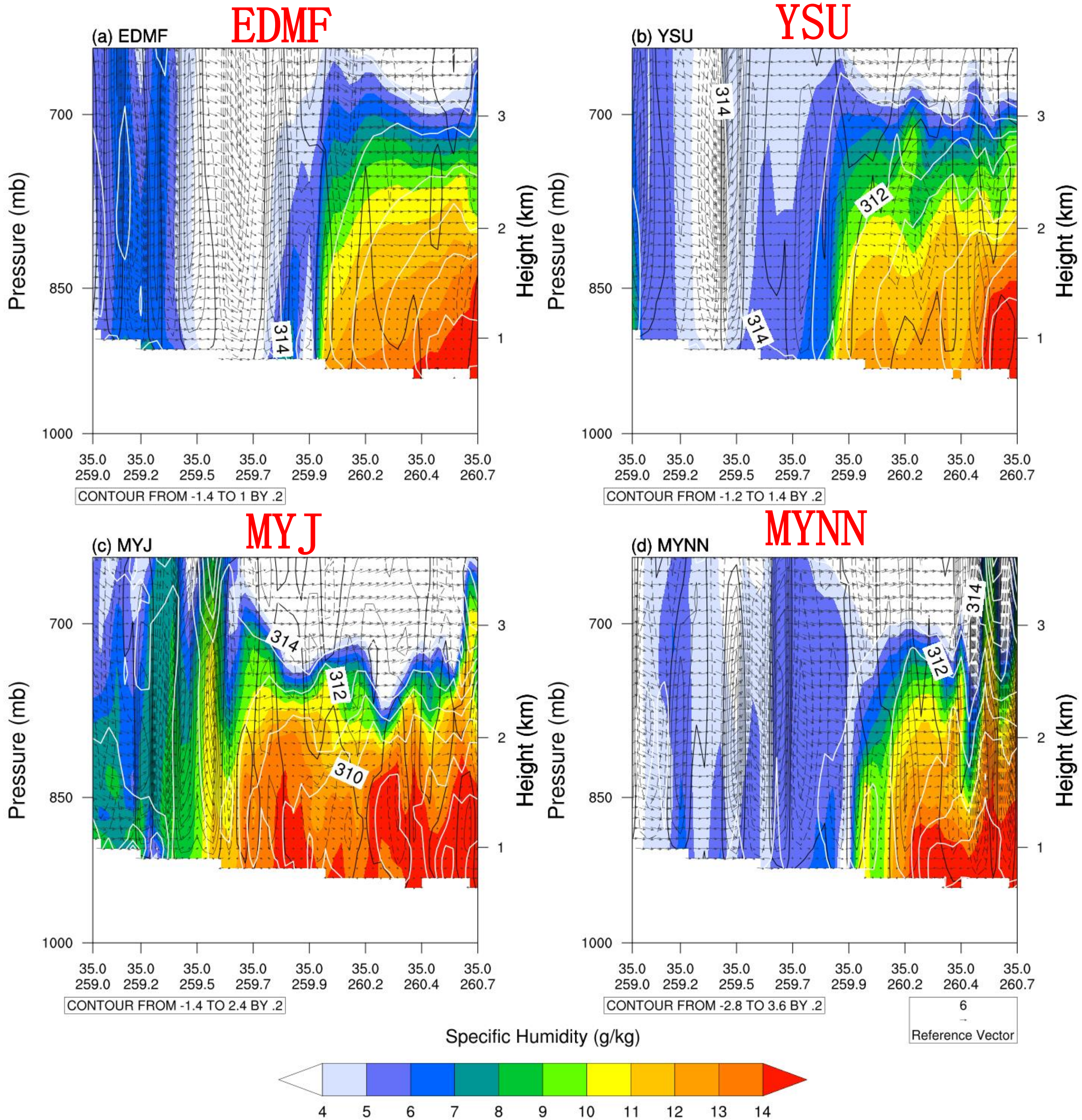
00 UTC 17 May
(7 pm Local Time)

Black contour:
Vertical velocity (m/s)
White contour:
Virtual potential
temperature (K)
Vector (U;W, m/s)
Shading:
Specific humidity

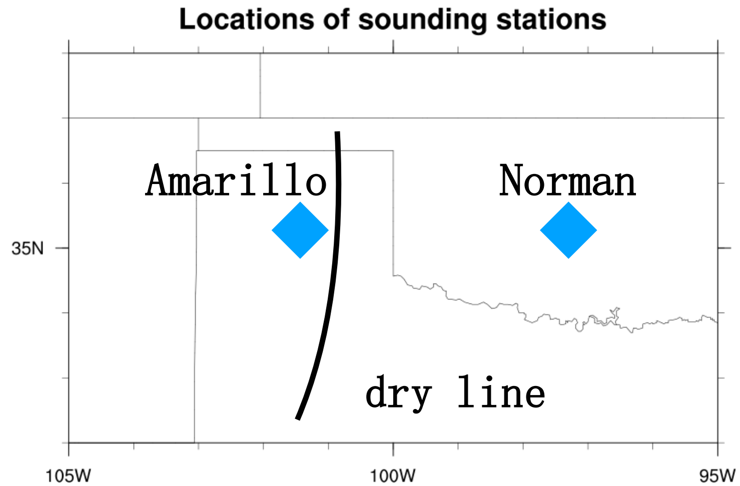


21 UTC 16 May (4 pm Local Time)

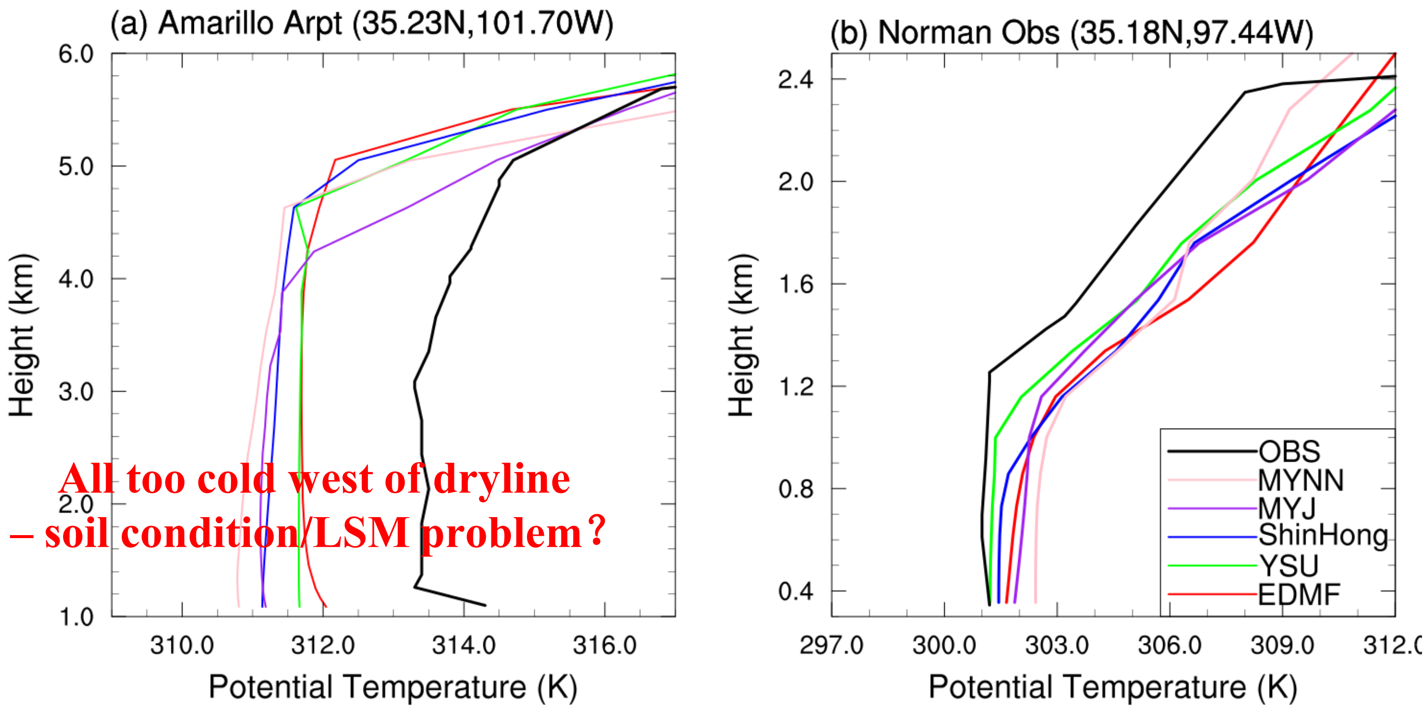
Black contour: Vertical velocity (m/s)
White contour: Virtual potential temperature (K)
Vector (U;W, m/s)
Shading: Specific humidity



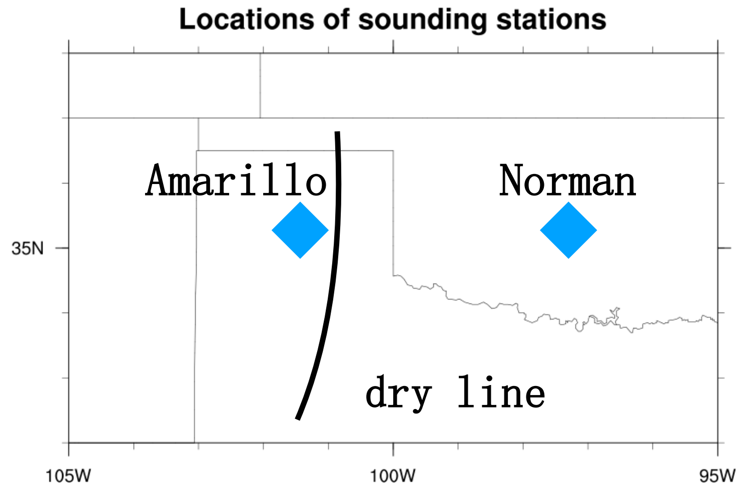
00 UTC Soundings



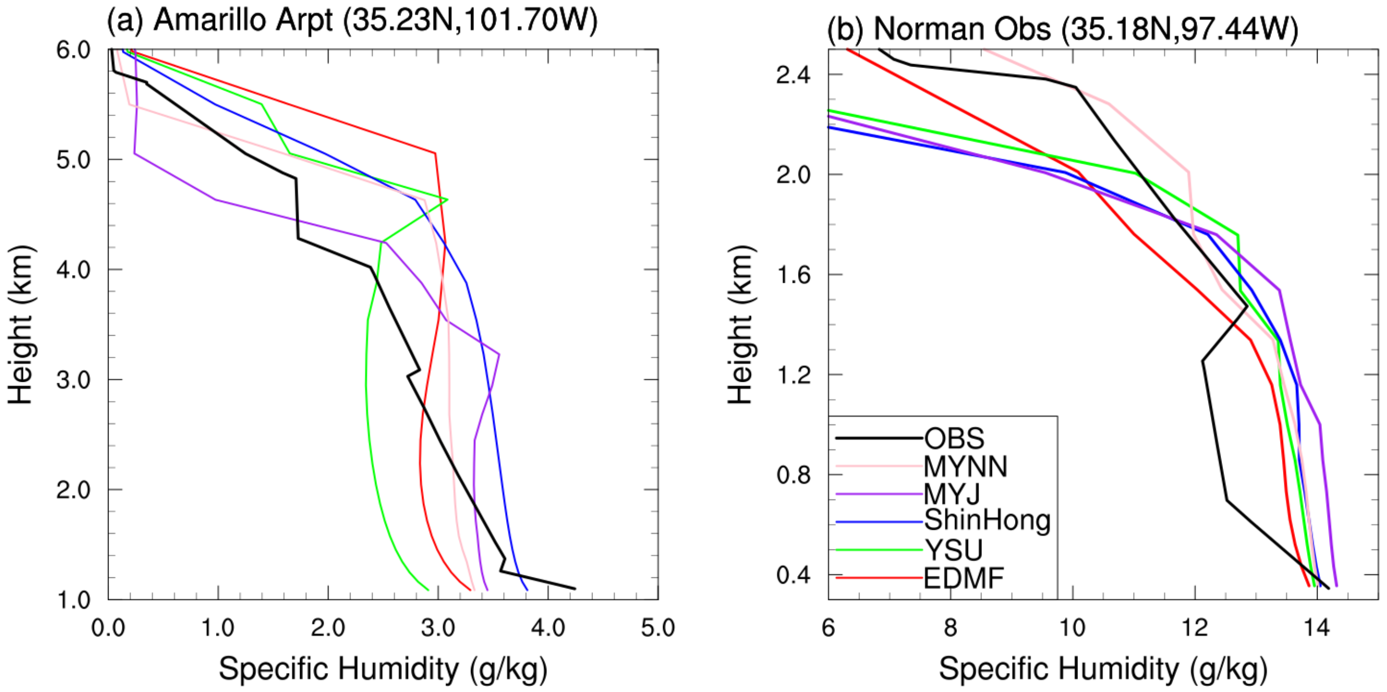
Potential Temperature



00 UTC Soundings

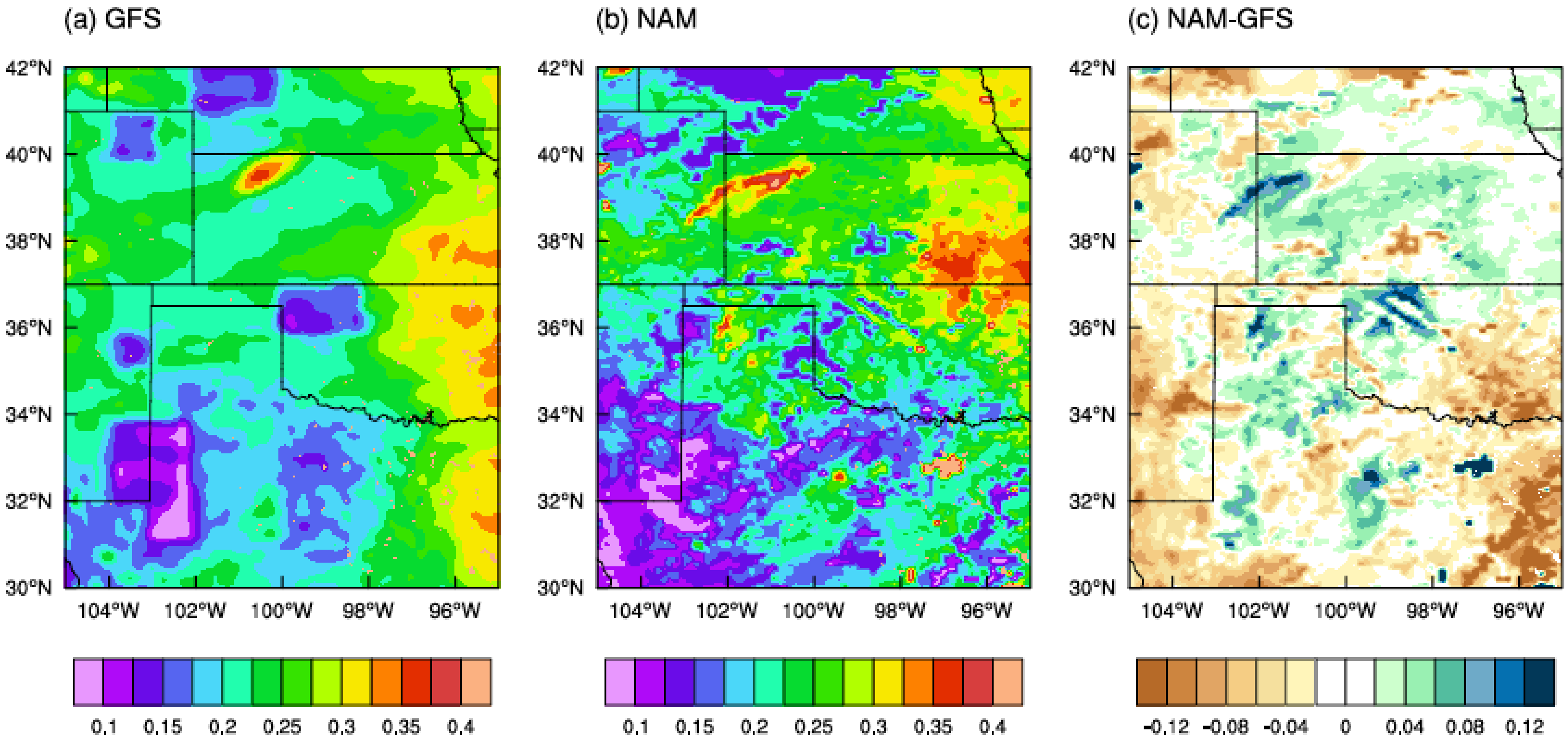


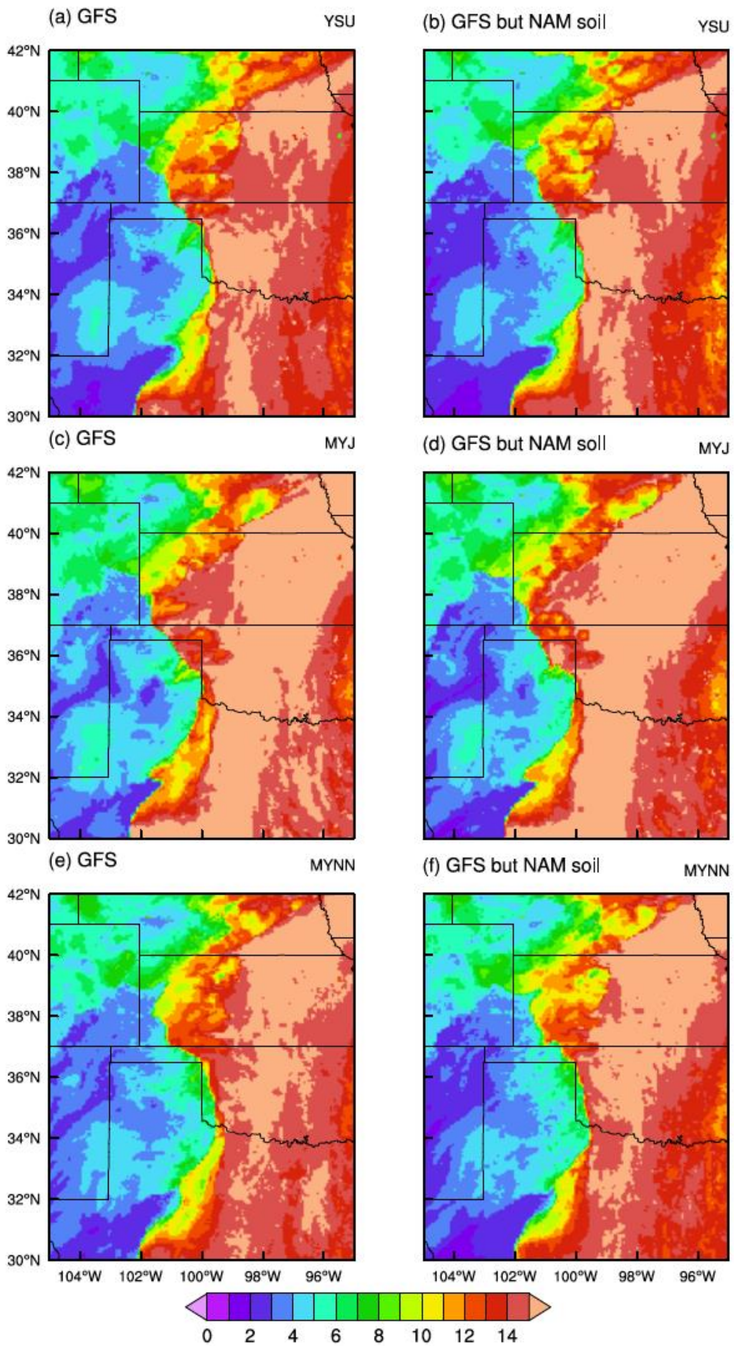
Specific humidity



Impact of soil moisture

IC: Soil moisture at 0-10cm





WRF forecasts using GFS IC but NAM soil states

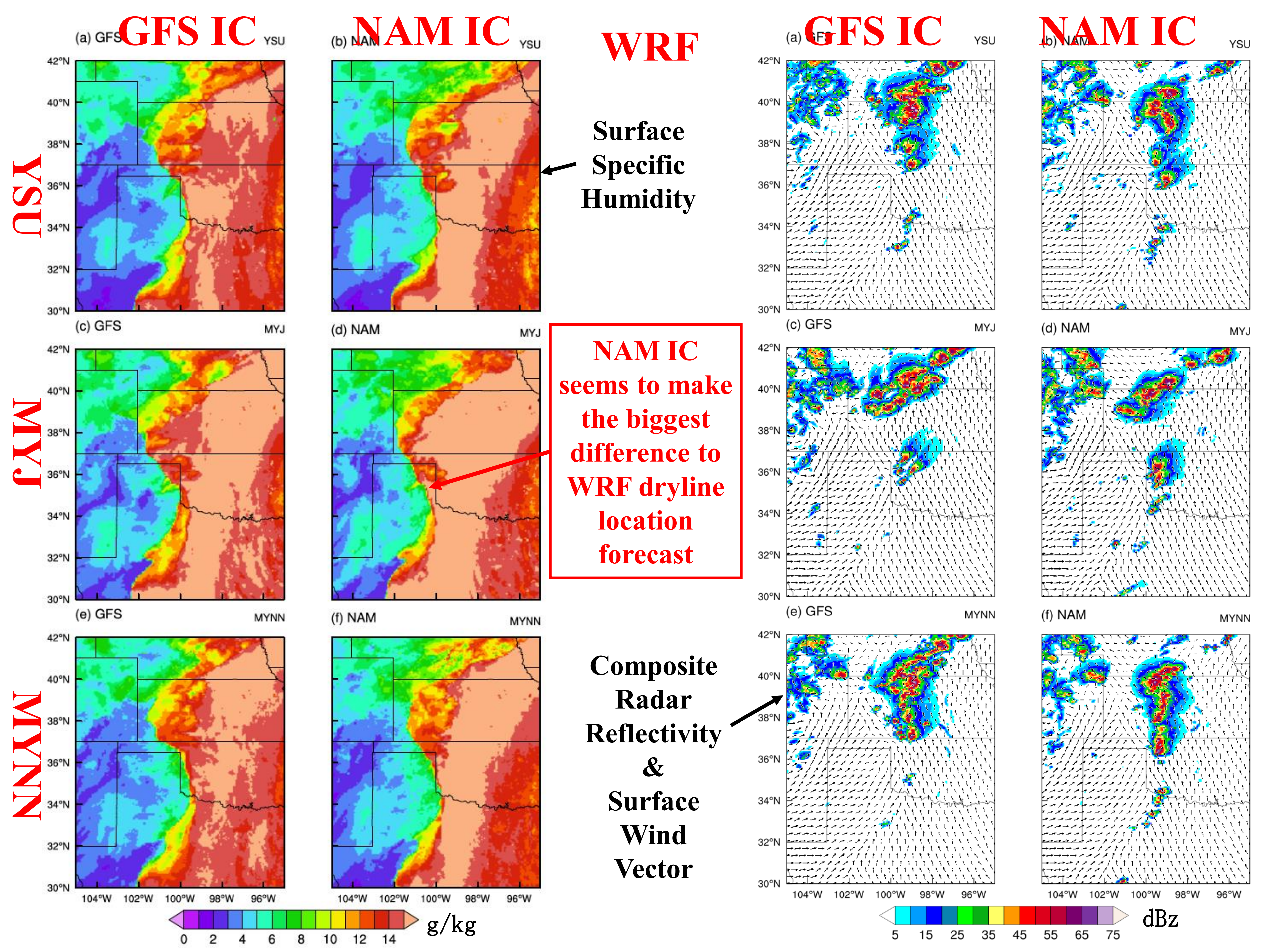
Specific humidity at 2 m height

The impact of soil moisture on dyline location is small

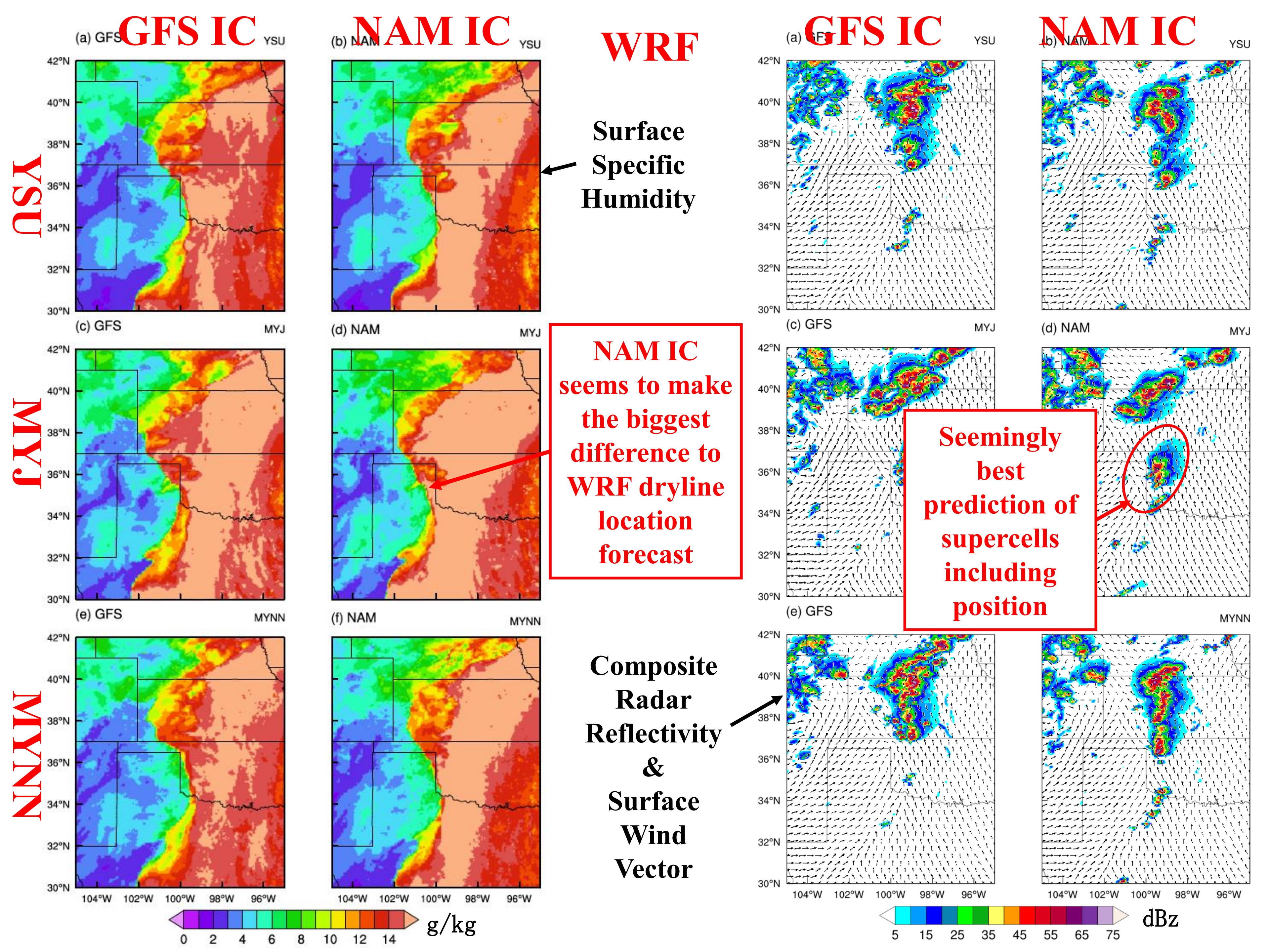
WRF-ARW Tests using HWT Grid
Thompson microphysics + different PBL schemes
Initial condition: GFS v.s. NAM Analysis

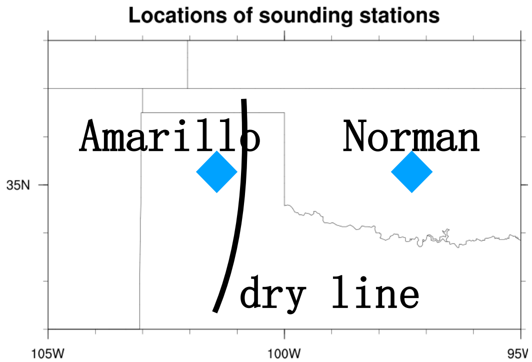
Single domain: 3 km horizontal resolution

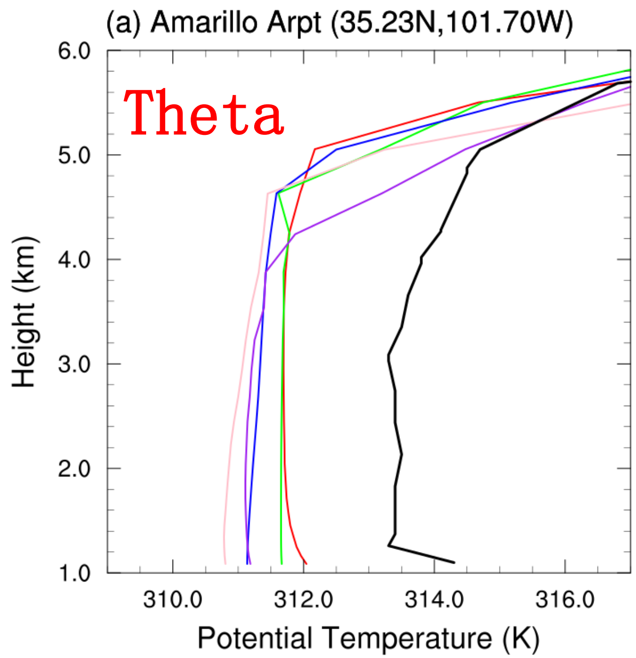
PBL schemes
YSU
MYJ
MYNN (scale-aware)

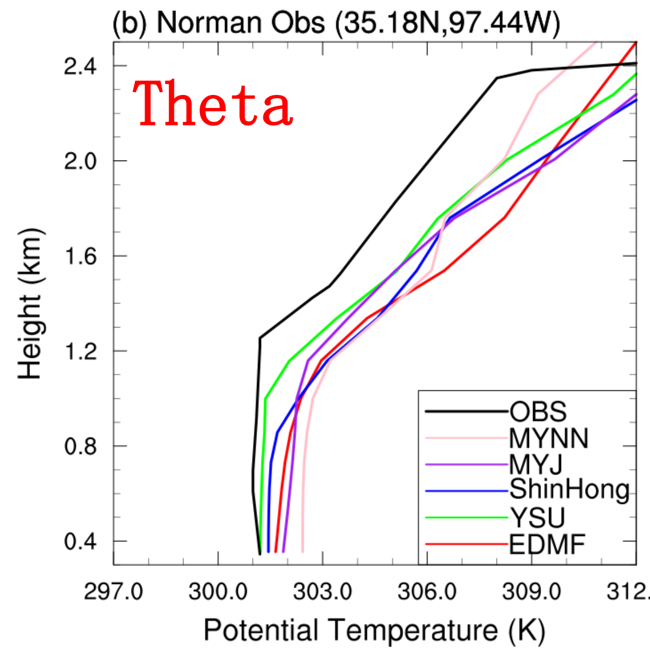


GFS IC
NAM IC
WRF
GFS IC
NAM IC
YSU
MYJ
MYNN
(a) GFS YSU
(b) NAM YSU
(c) GFS MYJ
(d) NAM MYJ
(e) GFS MYNN
(f) NAM MYNN
Surface Specific Humidity
NAM IC seems to make the biggest difference to WRF dryline location forecast
Composite Radar Reflectivity & Surface Wind Vector
42°N
40°N
38°N
36°N
34°N
32°N
30°N
104°W
102°W
100°W
98°W
96°W
0 2 4 6 8 10 12 14
g/kg
5 15 25 35 45 55 65 75
dBz


GFS IC
NAM IC
WRF
GFS IC
NAM IC
YSU
MYJ
MYNN
(a) GFS YSU
(b) NAM YSU
(c) GFS MYJ
(d) NAM MYJ
(e) GFS MYNN
(f) NAM MYNN
42°N
40°N
38°N
36°N
34°N
32°N
30°N
104°W 102°W 100°W 98°W 96°W
0 2 4 6 8 10 12 14
g/kg
Surface Specific Humidity
NAM IC seems to make the biggest difference to WRF dryline location forecast
Seemingly best prediction of supercells including position
Composite Radar Reflectivity & Surface Wind Vector
5 15 25 35 45 55 65 75
dBz







FV3

GFS IC

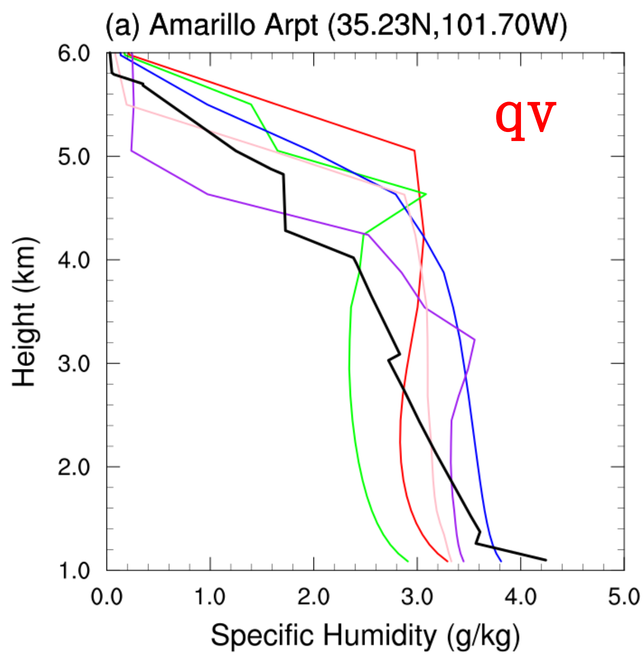
Clear difference

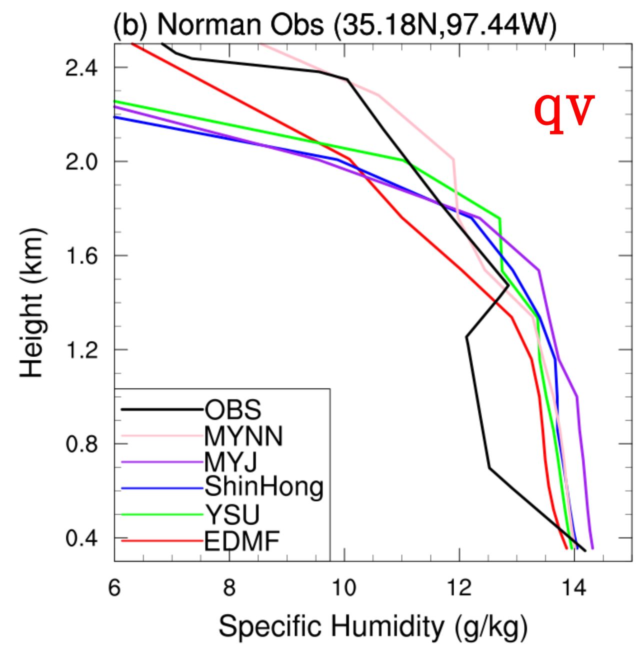
In FV3 an WRF

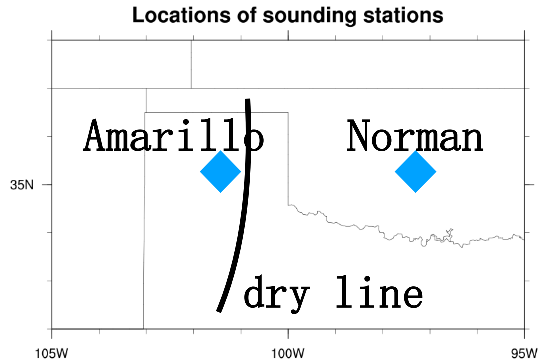
BL qv and q

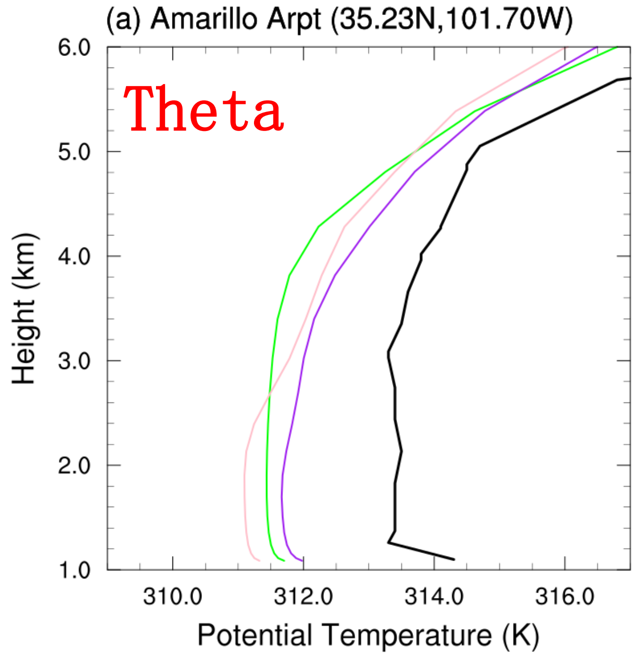
FV3 errors

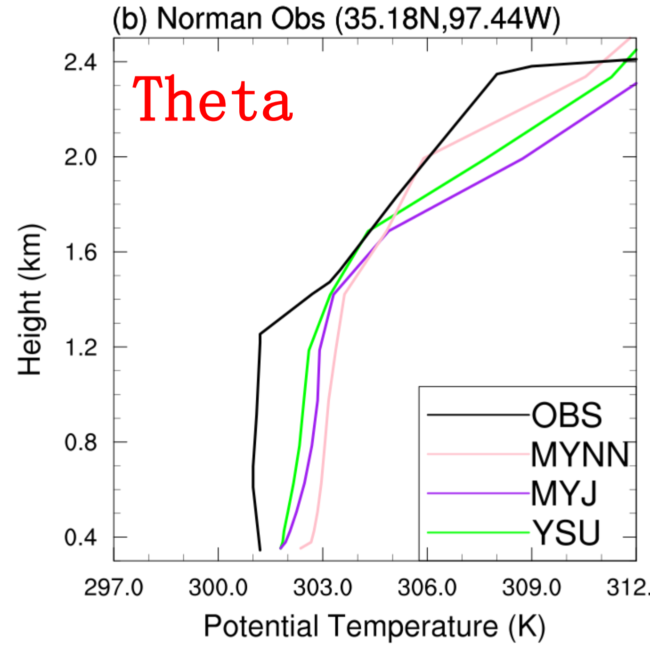
a little smaller





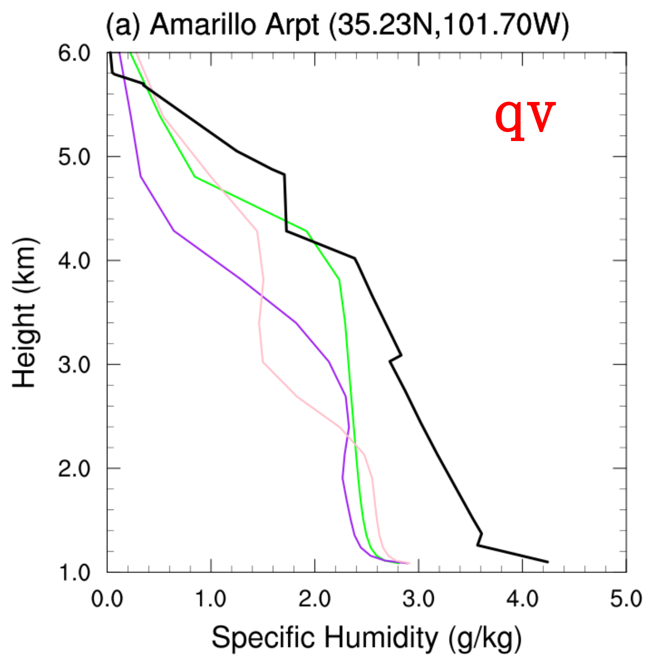


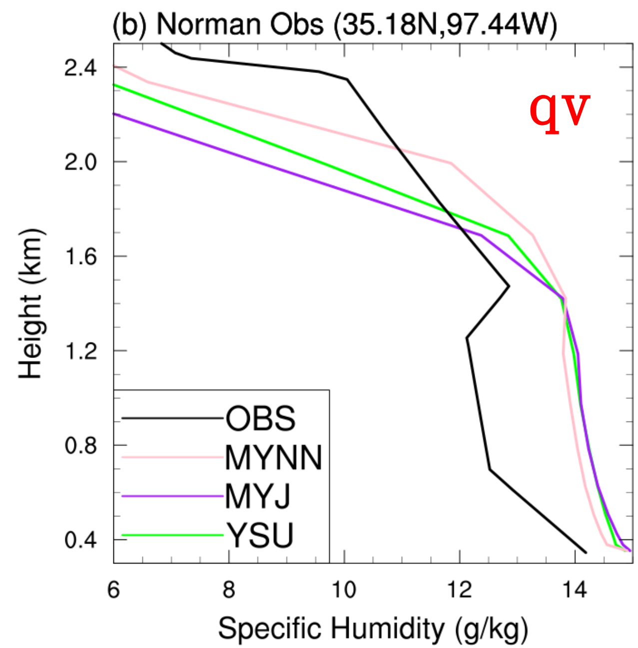




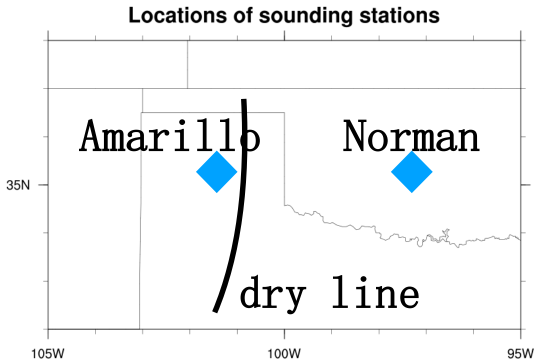
WRF

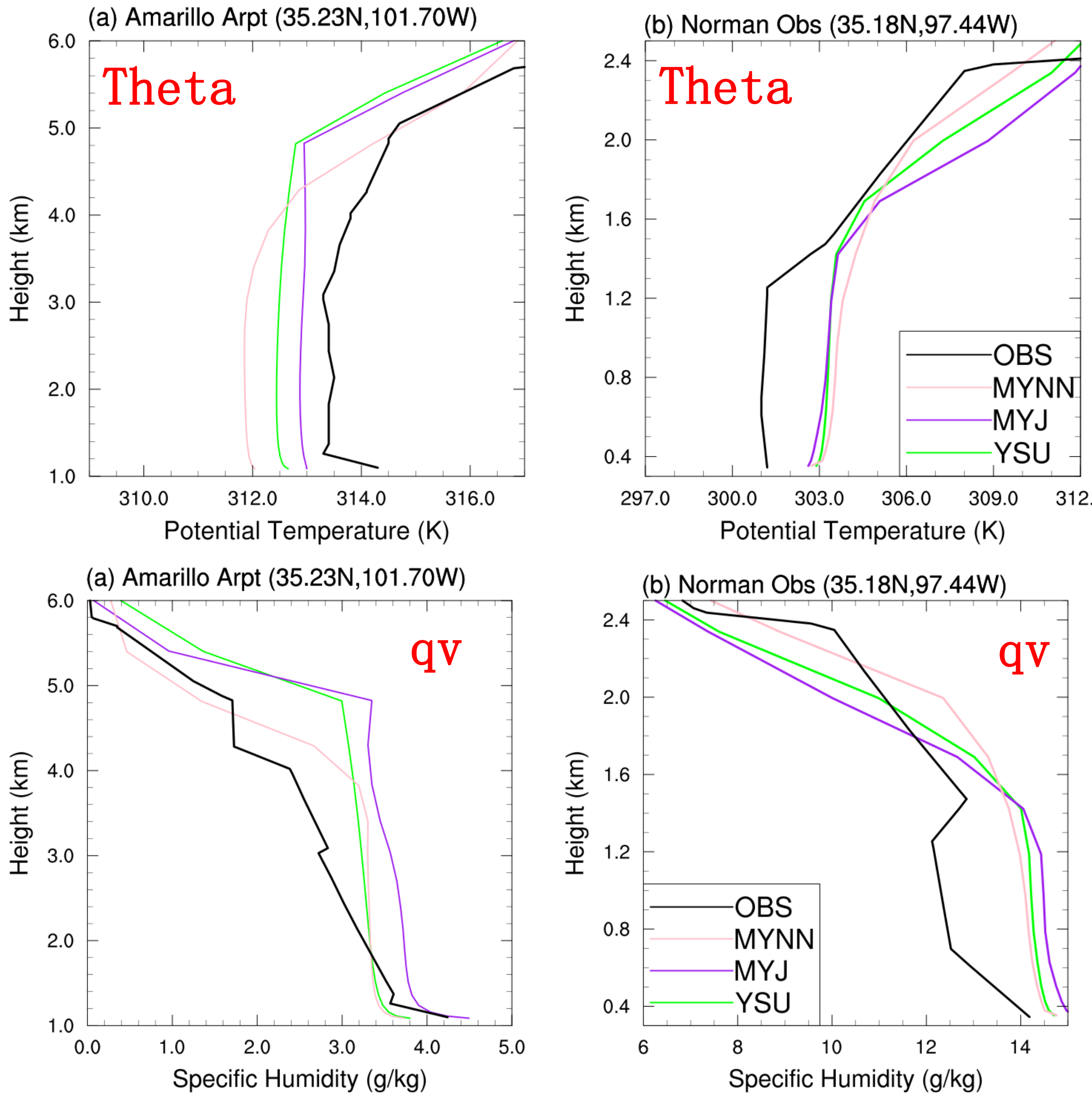
GFS IC





Clear difference
In FV3 an WRF
BL qv and q
FV3 errors
a little smaller





WRF

NAM IC

Warmer, wetter boundary layer with NAM IC

FV3 Tests using HWT Grid

Different MP or PBL schemes

Initial condition: GFS Analysis

Five active cases (3 from 2017 HWT, 2 from 2017 HMT)

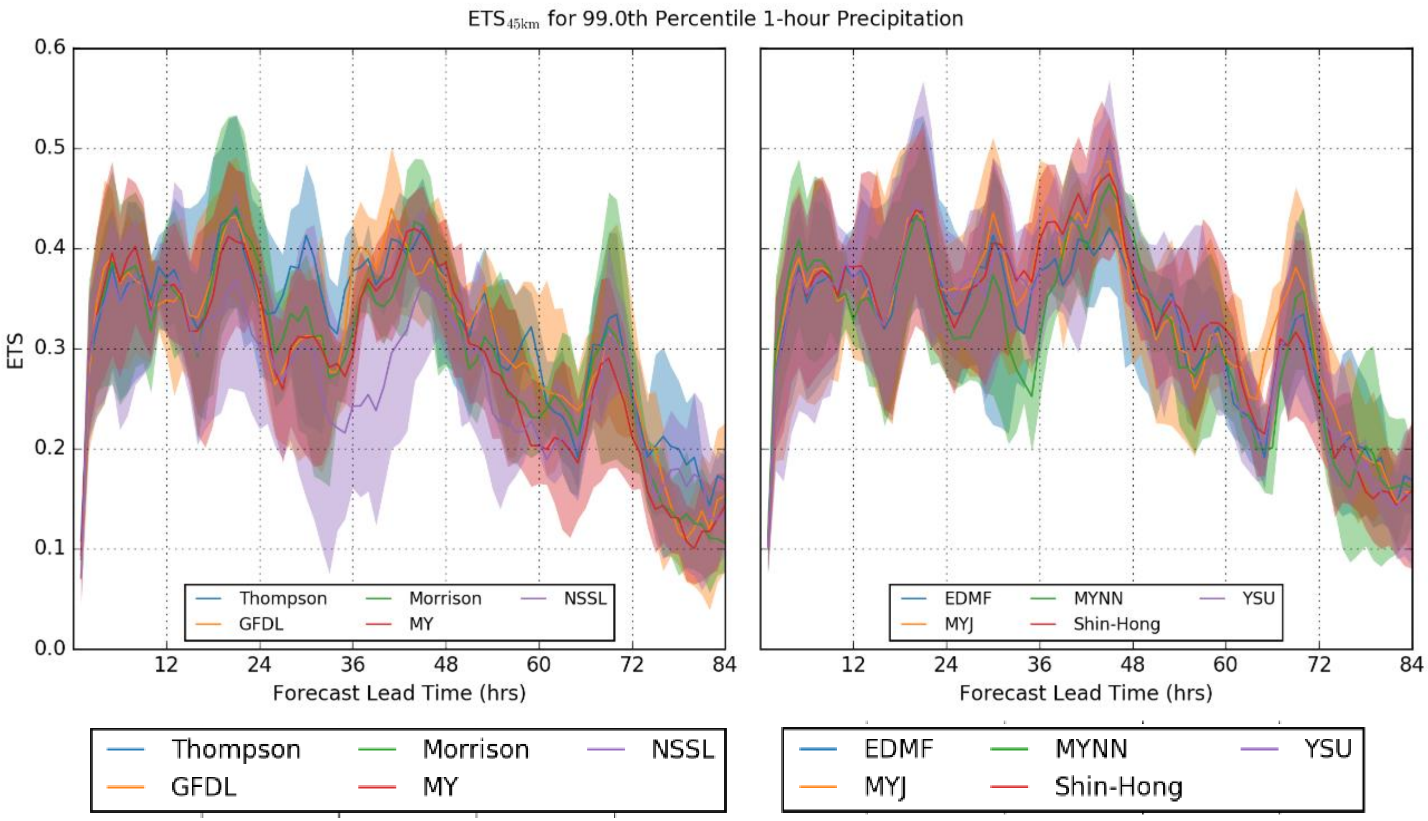
84 hours forecasts

May 12, 16, 19 and July 14, 19 2017

Nested domain: ~3 km horizontal resolution

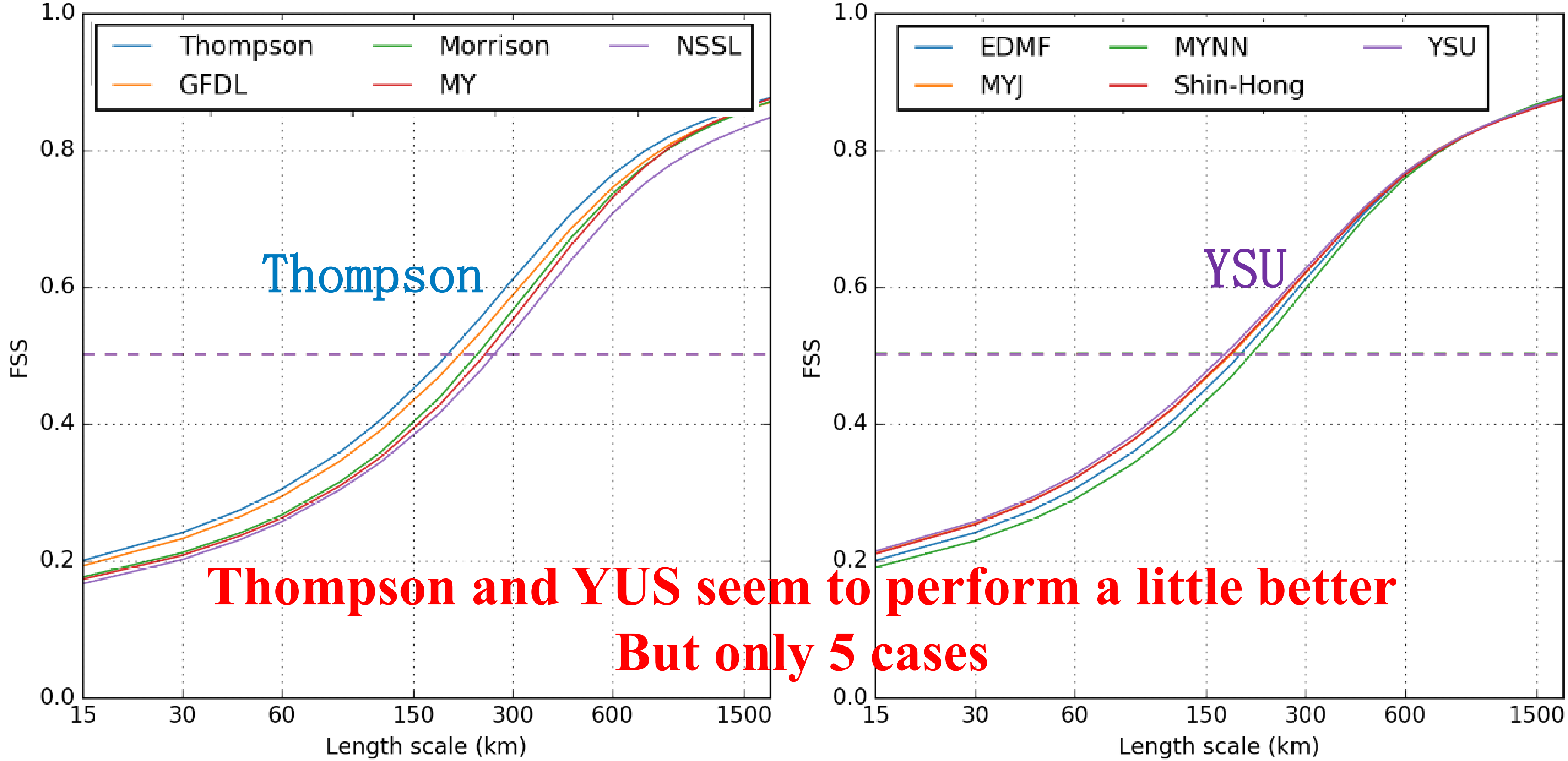
PBL schemes with Thompson MP	MP schemes with EDMF PBL
EDMF (GFS operational scheme)	GFDL 1-moment
YSU	Thompson partially 2-moment
Shinhong (scale-aware YSU)	NSSL 2-moment
MYJ	MY 2-moment
MYNN (scale-aware)	Morrison 2-moment

FV3 Physics Tests – 5 cases



- Neighborhood ETS (45 km)
- 99th percentile for each member (~2 mm)
- Little separation among physics members (NSSL somewhat worse but only 5 cases)

FV3 Physics Tests – Fractional Skill Scores



Thompson and YUS seem to perform a little better
But only 5 cases

- Fractions Skill Score (Roberts and Lean 2008)
  - Dashed line is “Minimum skillful forecast”
- Average over all forecast hours (1 – 84 h)
- 99^{th} percentile for each member

WRF Tests using HWT Grid

Different MP or PBL schemes

Initial/boundary conditions: GFS Analysis

Same 5 cases as FV3

84 hour forecasts

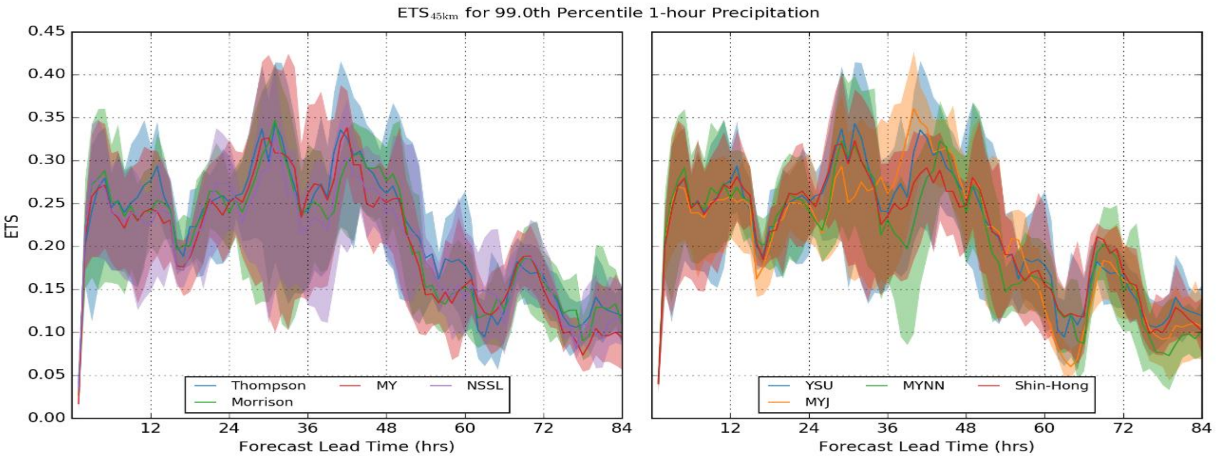
3 km horizontal resolution

PBL schemes with Thompson MP	MP schemes with MYJ PBL
YSU	Thompson
Shinhong (scale-aware YSU)	NSSL 2-moment
MYJ	MY 2-moment
MYNN (scale-aware)	Morrison 2-moment

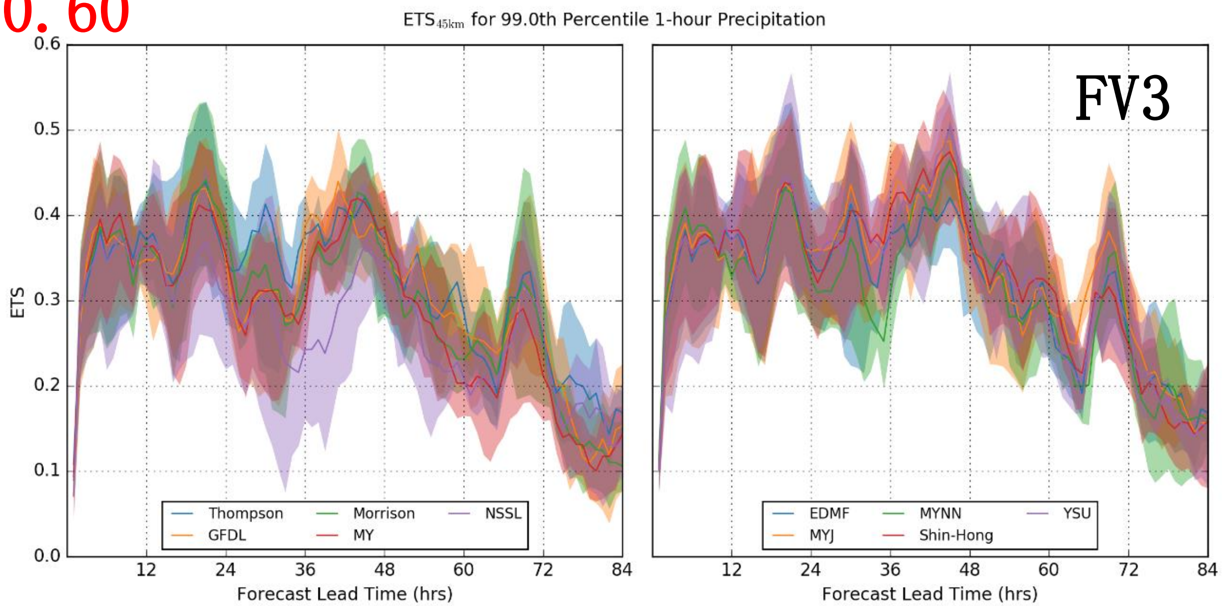
WRF and FV3 Physics Tests with GFS IC

0. 45

WRF



0. 60



- 99 percentile, R=45km
- Also little separation between physics schemes in WRF
- FV3 generally has higher scores than WRF (using same GFS IC)

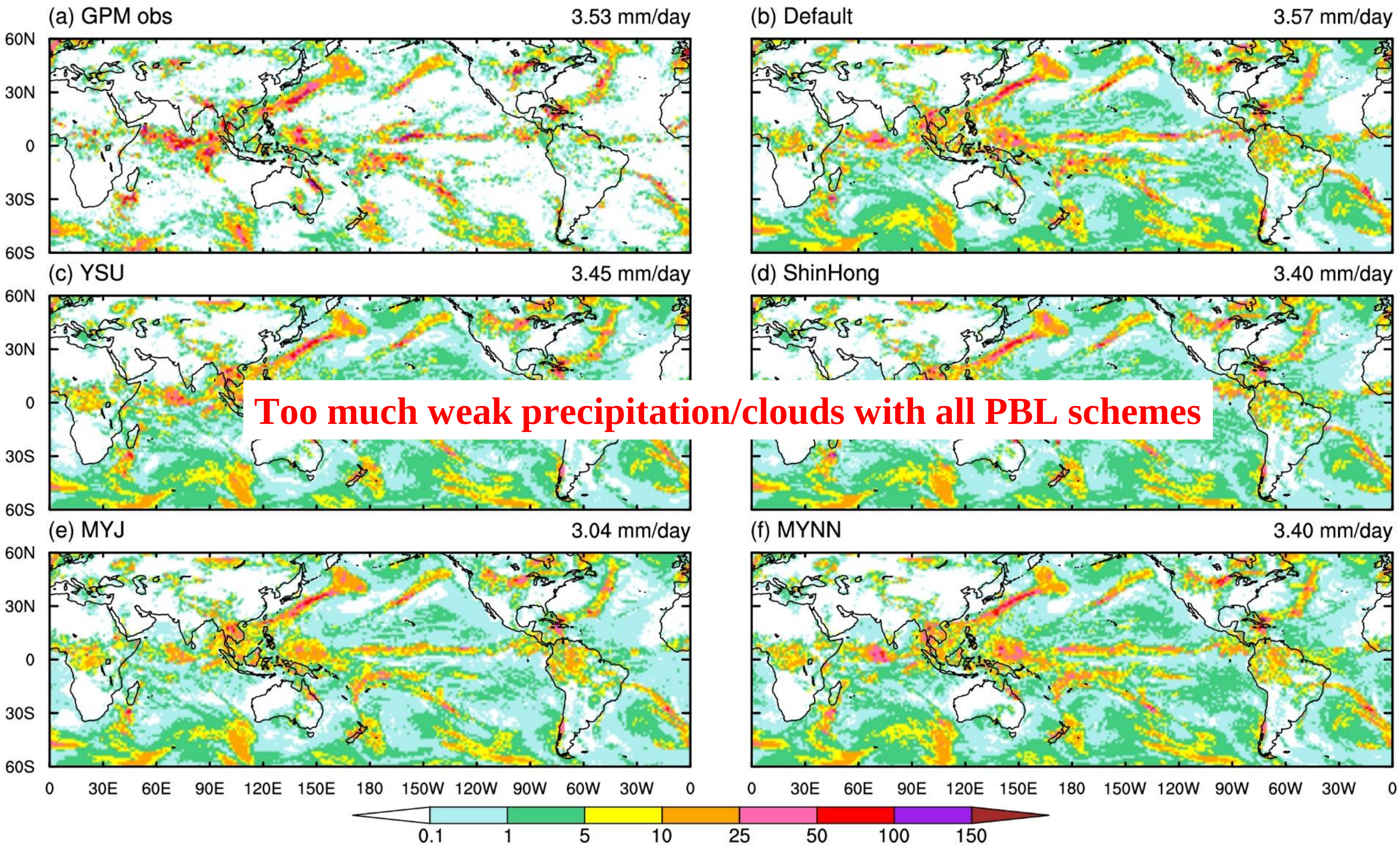
Summary on Physics Testing

- The simulated dryline location, PBL structure and convective initiation in FV3 are sensitive to PBL scheme for the May 16-17 tornadic supercell storm case;
    - YSU and MYNN produced much improved forecasts of supercells but the positions were still too far east;
    - The dryline position in WRF forecasts was improved more by the use of NAM IC
- No clear separation of hourly precipitation forecasts up to 84 hours using different PBL or MP schemes with either FV3 or WRF for 5 active cases.
- FV3 using GFS IC more skillful than WRF with GFS IC
    - Importance of regional DA, including land DA
- Should develop optimized/compatible physics suites
    - If not clear winners emerge, multi-physics ensemble makes sense
- New Tiedtke cumulus scheme in FV3 seems to improve global precipitation forecast

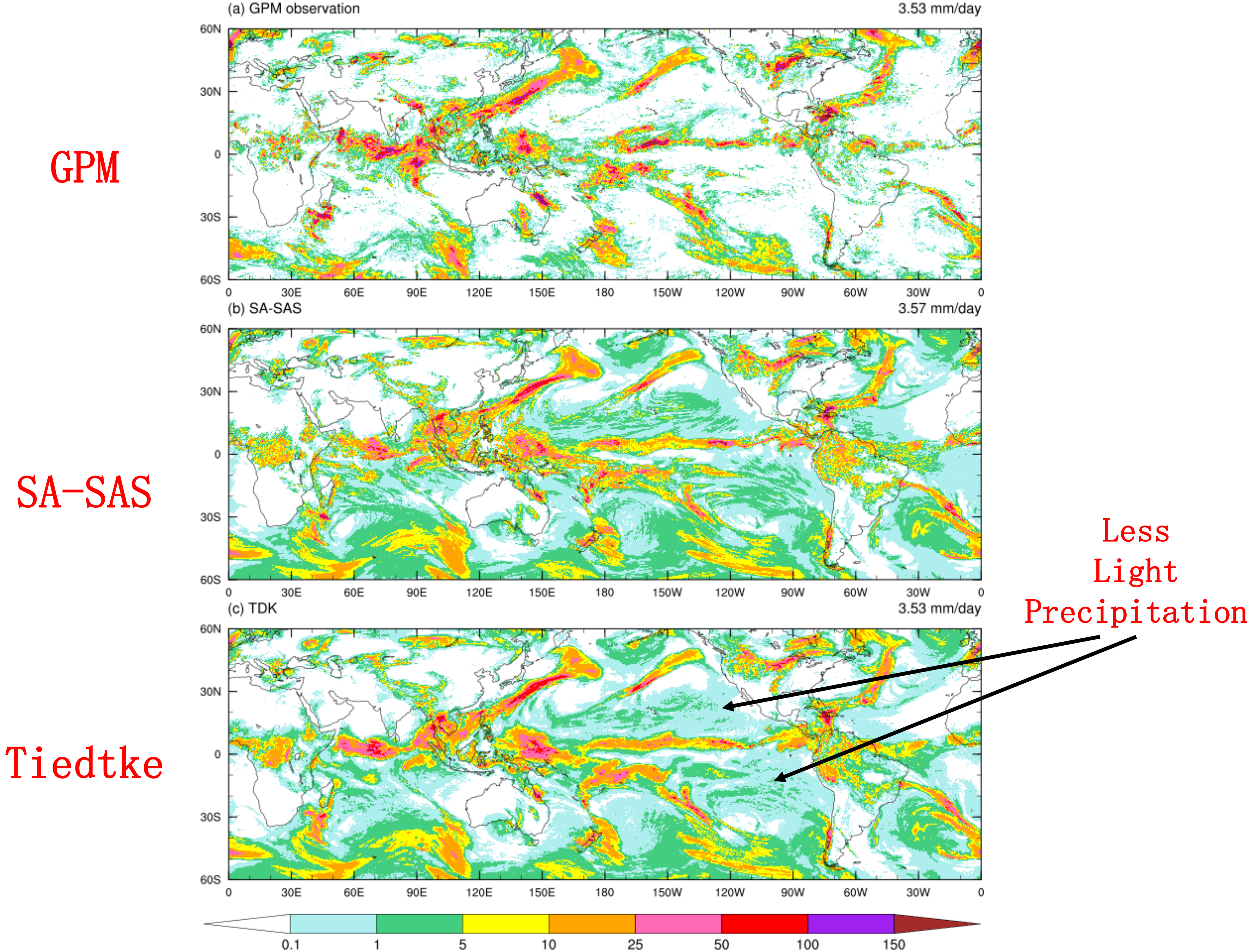
Precipitation and Clouds in single 13-km FV3 Global Domain

- FV3 version 1.3 from GFDL (Mid-September 2017)
- Hybrid eddy-diffusivity mass-flux (EDMF) PBL scheme
- GFDL microphysics scheme
- Noah Land surface model
- Cumulus scheme
  - Scale-aware SAS for both deep and shallow cumulus
  - New Tiedtke cumulus scheme tested

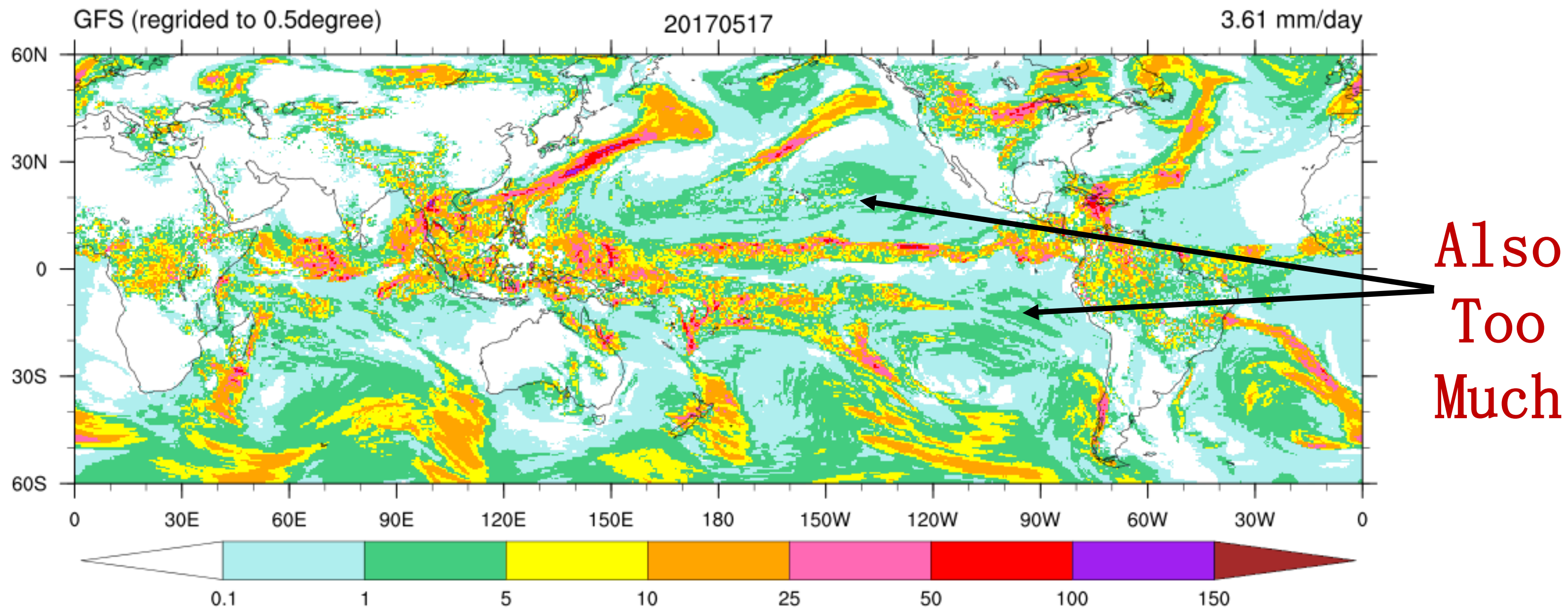
Model initialized at 00 UTC 16 May 2017.
Accumulated precipitation GPM estimate and forecasts (24-48 h) using SA-SAS



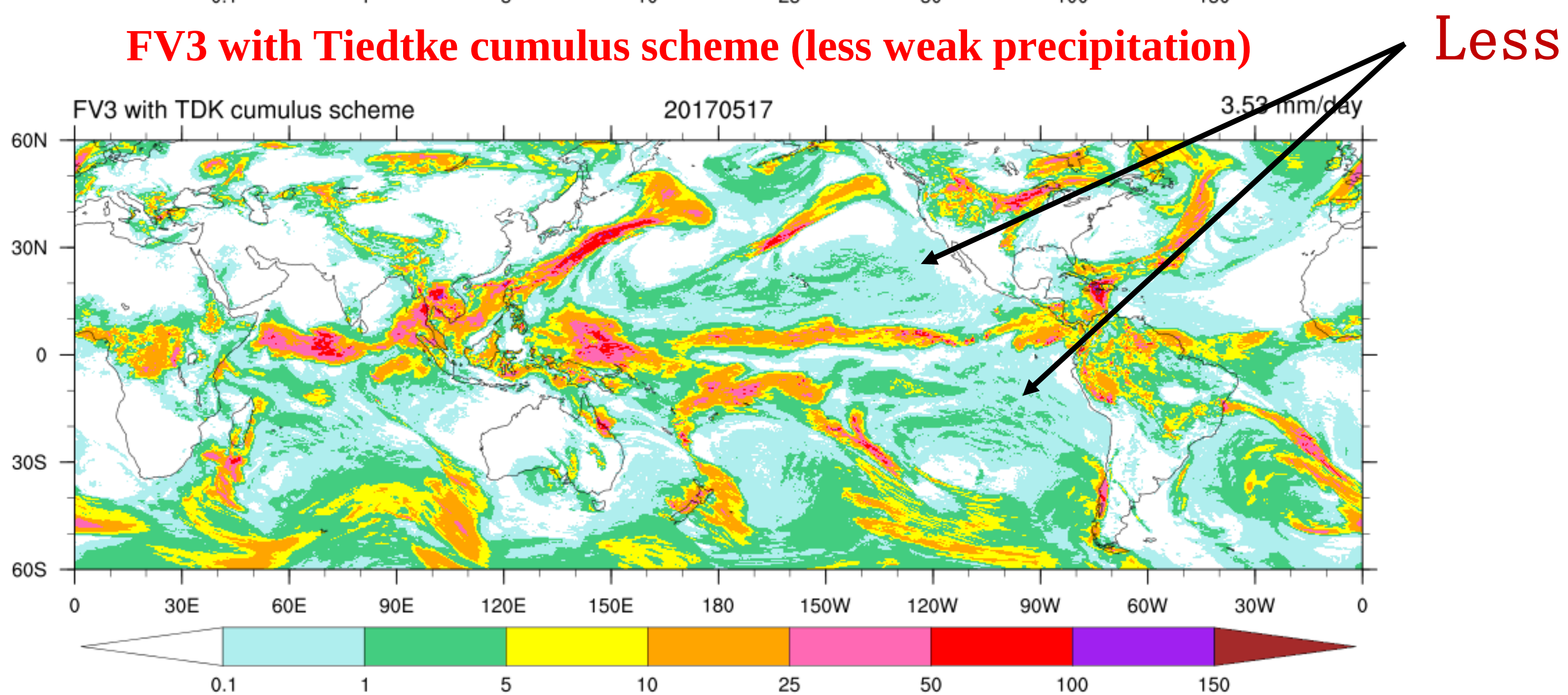
Global precipitation amount is sensitive to PBL schemes!


GPM
SA-SAS
Tiedtke
Less
Light
Precipitation
(a) GPM observation
3.53 mm/day
(b) SA-SAS
3.57 mm/day
(c) TDK
3.53 mm/day
60N
30N
0
30S
60S
0
30E
60E
90E
120E
150E
180
150W
120W
90W
60W
30W
0.1
1
5
10
25
50
100
150

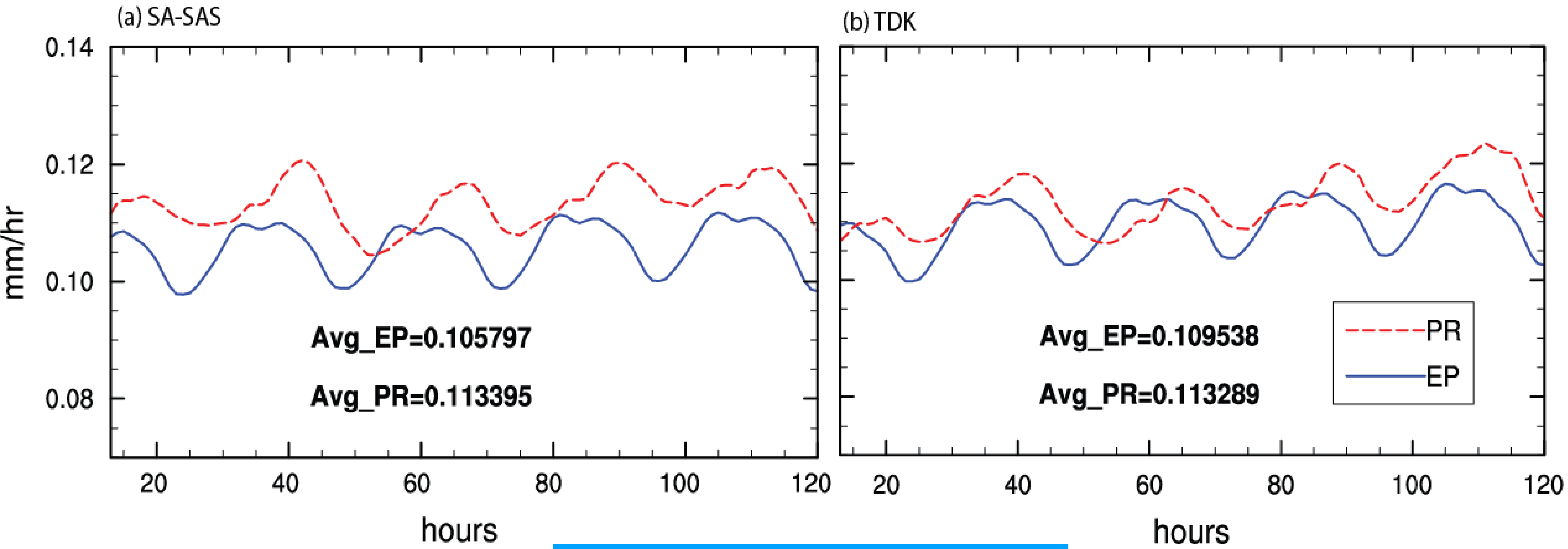
Operational GFS Forecast (similar problem)



FV3 with Tiedtke cumulus scheme (less weak precipitation)



Domain total precipitation
versus
Total surface moisture flux

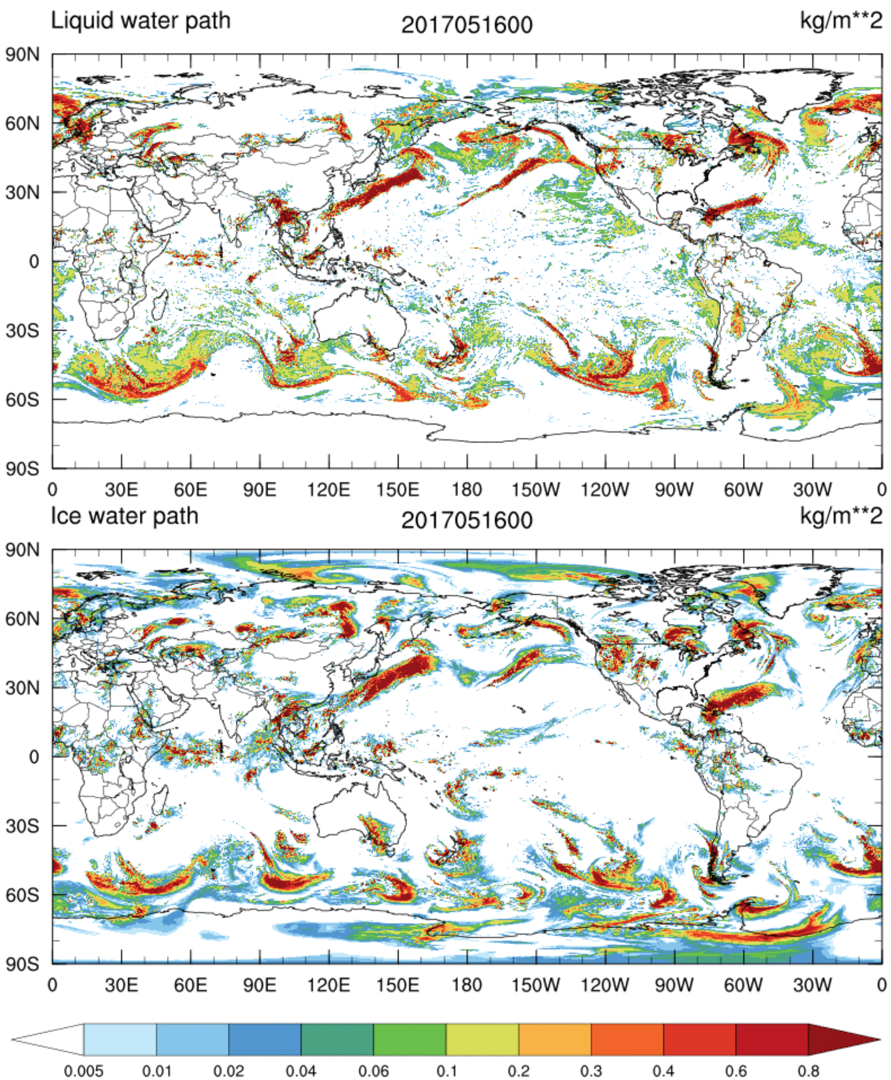


PR: precipitation
EP: evapotranspiration

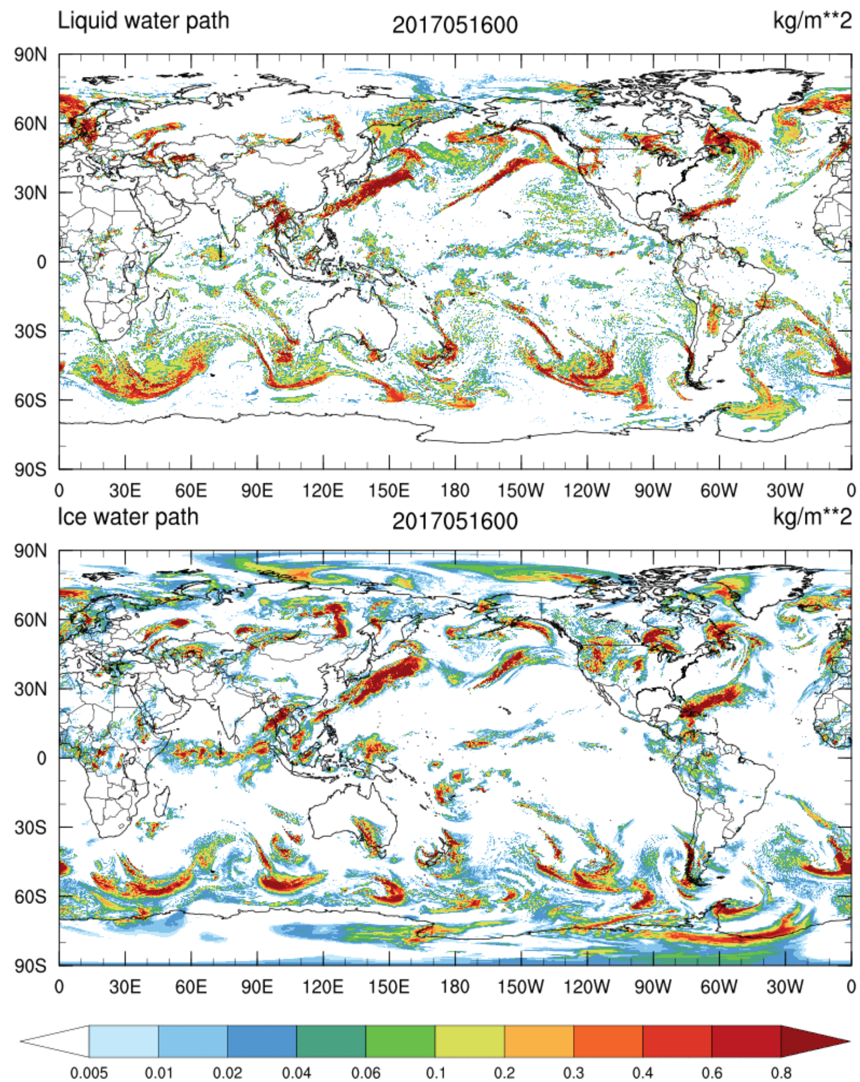
Current GFS cumulus scheme in FV3 over-predicts precipitation

TDK has smaller difference

FV3 with SA-SAS (of GFS)



FV3 with Tiedtke



Timing Statistics for Every Hour of Forecast

FV3 single 13 km grid 16 nodes, 768 cores		FV3 nested 13 and 3 km grids 24 nodes, 1152 cores	WRF 3 km grid 16 modes, 768 cores
GFDL	1.1 min	3.7 min	
Thompson	1.4 min	4.8 min	4.1 min
NSSL	1.4 min	4.7 min	5.2 min
MY	1.4 min	4.6 min	4.1 min
Morrison	1.6 min	5.4 min	3.9 min

On TACC Stampede 2: Skylake nodes

The differences across PBL schemes are small

FV3 forecasts with 13-3 km nested grid cost about the same as a single 3km WRF using 1.5 times the number of cores