Atmospheric Prediction Dynamics Team Report

Phase 2 Atmospheric Dynamic Core Testing and Evaluation

Jeff Whitaker and Vijay Tallapragada

Rationale for Replacing Global Spectral Model (GSM)

- Continued GFS operational performance improvements will require non-hydrostatic, convection-permitting resolutions.
- Current spectral, semi-lagrangian dycore approaching scaling limits on current HPC.
- Next-Generation computing paradigm will require scaling across potentially 100,000's processors or more

NGGPS Atm Dynamics Goals and Strategy

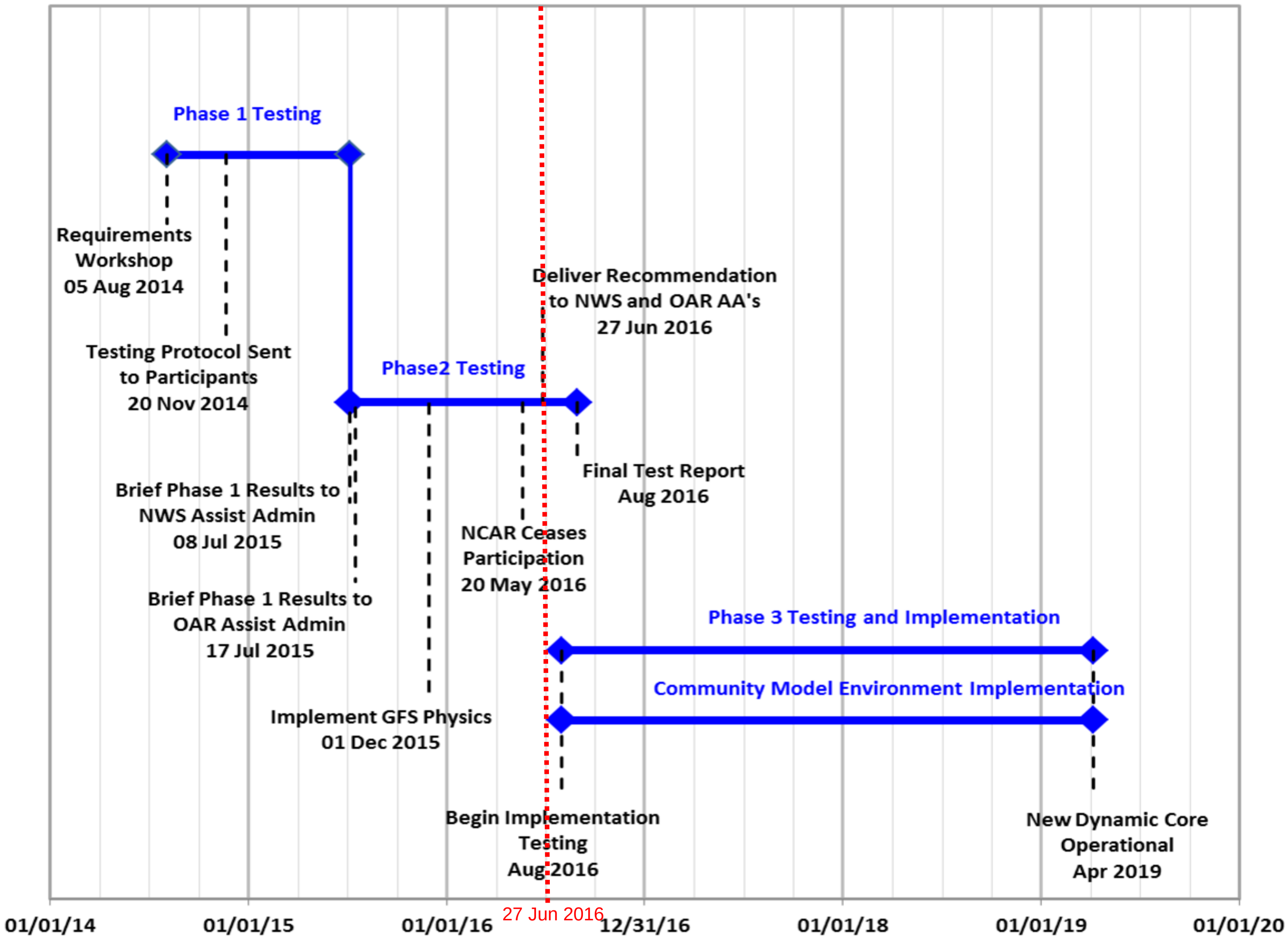
- **Goal: implement a new atmospheric dynamical core which is *non-hydrostatic* and *scalable.***

- **Strategy: Choose from among several existing U.S. development efforts rather than ‘clean-sheet of paper’ approach.**

NGGPS Atmospheric Model Phased Implementation Approach

- **Phase 1 (FY15) – Identify Qualified Dynamic Cores**
  - Evaluate technical performance
    - Scalability
    - Integration of scheme stability and characteristics
- **Phase 2 (FY16) – Select Candidate Dynamic Core**
  - Integrate with operational GFS Physics/CCPP
  - Evaluate meteorological and computational performance
- **Phase 3 (FY17-19) – Dynamic Core Integration and Implementation**
  - Implement candidate dynamic core in NEMS
  - Implement Common Community Physics Package
  - Implement data assimilation (4DEnVar with 4D incremental analysis update and stochastic physics)
  - Implement community model environment

Dynamic Core
Testing and Implementation Timeline



Testing and Evaluation Summary

NGGPS Dycore Test Group (DTG) Membership

- Ming Ji, Chair
  - Dir., Office of Sci. Tech. Integ.
- Fred Toepfer
  - NGGPS Program Manager
- Tim Schneider
  - Acting NGGPS Program Manager
- Bob Gall
  - Independent Consultant
- Ricky Rood
  - Independent Consultant
- John Thuburn
  - Independent Consultant
- Melinda Peng/Jim Doyle
  - Navy/NRL Monterey
- Ram Ramaswamy/SJ Lin
  - GFDL
- Hendrik Tolman/Vijay Tallapragada
  - NCEP/EMC
- Chris Davis/Bill Skamarock*
  - NCAR/MMM
- Kevin Kelleher/Stan Benjamin
  - ESRL/GSD
- Jeff Whitaker
  - NGGPS Test Manager
- John Michalakes
  - Chair, Advanced Computing Evaluation Committee

* Ceased participation and withdrew from DTG on 20 May 2016

New Dynamic Core Candidate Models

Phase 1 Candidate Dynamic Cores*:

* Built upon HIWPP Non-hydrostatic Model Evaluation

- Non-hydrostatic Global Spectral Model (**GSM**) - **EMC**
- Global Non-hydrostatic Mesoscale Model (**NMM** & **NMM-UJ**) - **EMC**
- Model for Prediction Across Scales (**MPAS**) - **NCAR**
- Non-hydrostatic Icosohedral Model (**NIM**) – **ESRL**
- Navy Environmental Prediction System Using the NUMA Core (**NEPTUNE**) – **Navy**
- Finite Volume Model version3 (**FV3**) – **GFDL**

- **FV3 and MPAS selected to advance to Phase 2**

Phase 2 Test Plan Development Timeline

- Test Plan Developed by DTG between June and December 2015
  – Testing Criteria Finalized by DTG at Face-to-Face Meeting in September 2015
  – Initial Test Plan Developed by November 2015 (including AVEC Test Plan)
- Test Plan Approved by DTG in January 2016

NGGPS Phase 2 Test Plan

#	Evaluation Criteria
1	Plan for relaxing shallow atmosphere approximation (deep atmosphere dynamics)*
2	**Accurate conservation of mass, tracers, entropy, and energy**
3	**Robust model solutions under a wide range of realistic atmospheric initial conditions using a common (GFS) physics package**
4	**Computational performance with GFS physics**
5	**Demonstration of variable resolution and/or nesting capabilities, including supercell tests and physically realistic simulations of convection in the high-resolution region**
6	**Stable, conservative long integrations with realistic climate statistics**
7	Code adaptable to NEMS/ESMF
8	Detailed dycore documentation, including documentation of vertical grid, numerical filters, time-integration scheme and variable resolution and/or nesting capabilities*
9	**Evaluation of performance in cycled data assimilation**
10	Implementation Plan (including costs)

Summary of Phase 2 Test Results

- Testing yielded sufficient information to evaluate both dynamic cores and produce a low risk recommendation without compromising performance or skill
- Summary of results:
  - Computationally, FV3 is more than twice as fast as MPAS with equivalent resolution
  - Full forecast experiments with GFS initial conditions and GFS physics showed significant differences between FV3 and MPAS, FV3 almost equivalent to GFS (some stability issues with MPAS forecasts)
  - FV3 performs comparable to the GFS in cycled data assimilation test (without tuning, at reduced resolution), MPAS performance inferior to GFS
  - Effective resolution for both dynamic cores is found to be similar, and higher than GFS
  - High-resolution idealized and real-data simulations show qualitatively similar results in simulations of explicit moist convection
  - Cost to implement FV3 is significantly less than MPAS in terms of manpower and computational resources

Critierion #2: Conservation Test

- DCMIP-2012 baroclinic wave idealized test, dry and moist (4.1 and 4.2) run at 13 km resolution. Simple moist physics (large-scale condensation only) included.
- Conservation of total energy, entropy and dry mass measured
- Extra advected tracer added, initialized with θ_e (difference between advected and diagnosed θ_e measured)
- ’Grid imprinting’ (signal of truncation errors at cube corners and pentagons of icosahedral grid) assessed

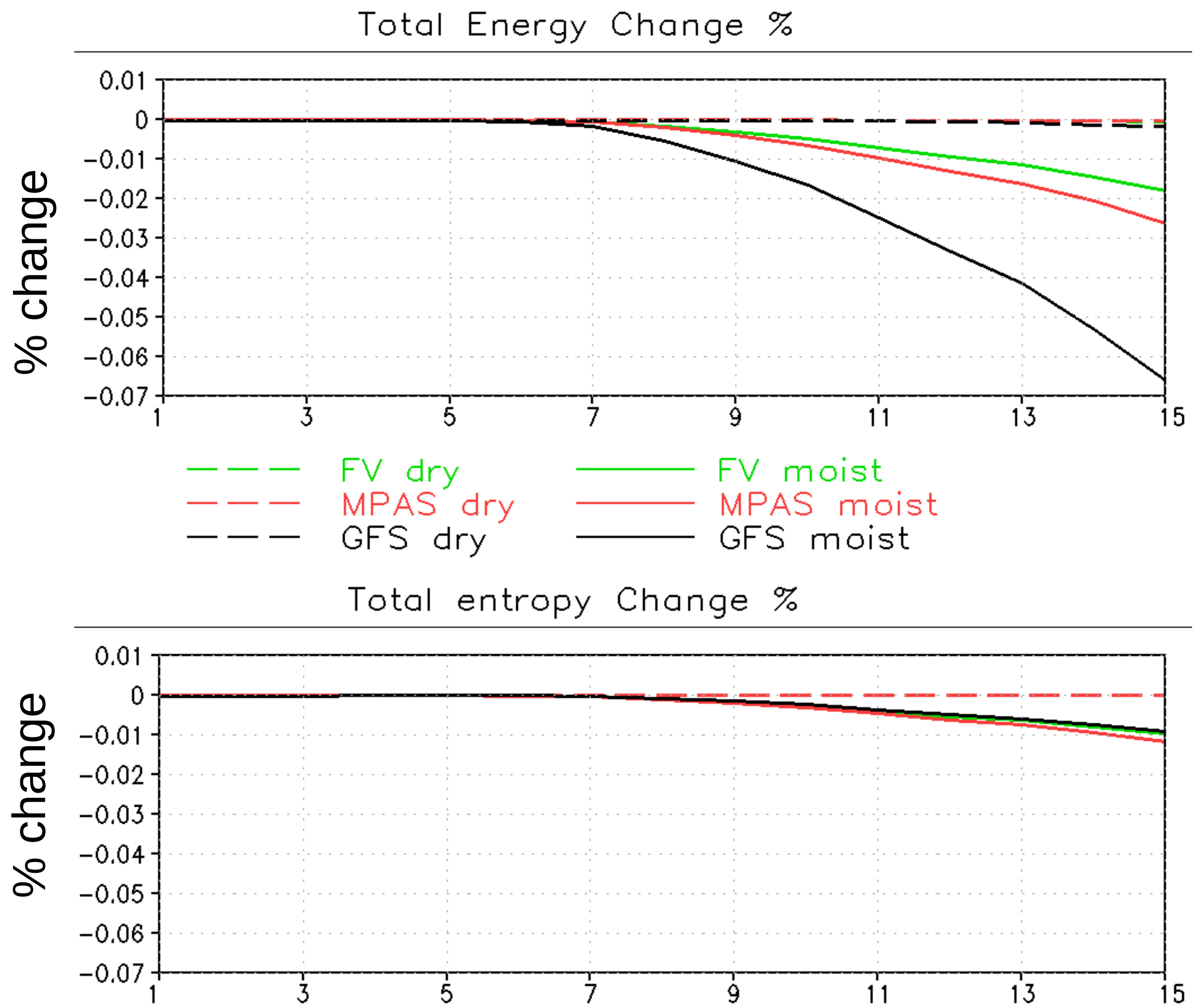
#2: Conservation Tests
Change in Total Energy and Entropy

Change in total energy (top) and entropy (bottom) as a percent change from the initial value. ***Note very tiny range on y axis.***

Energy loss nearly zero in dry case, FV3 and MPAS lose less energy than GFS in moist case.

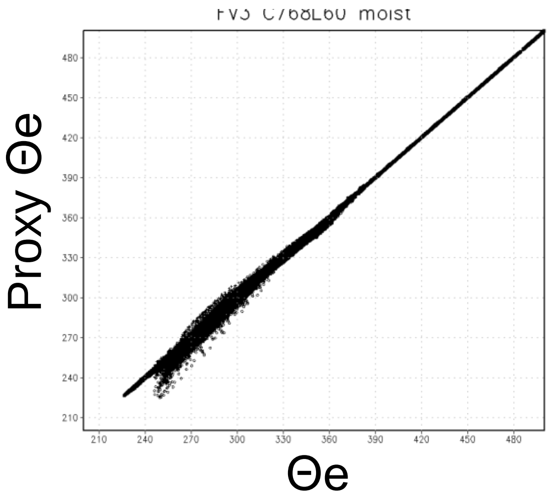
Energy loss in moist case for FV3 and MPAS is consistent with the energy removed along with condensate. Entropy changes for moist case are very small, and consistent with thermodynamic approximations made in entropy definition.

Dry mass (not shown) is conserved exactly in both FV3 and MPAS, GFS gains 0.05 hPa during integration.

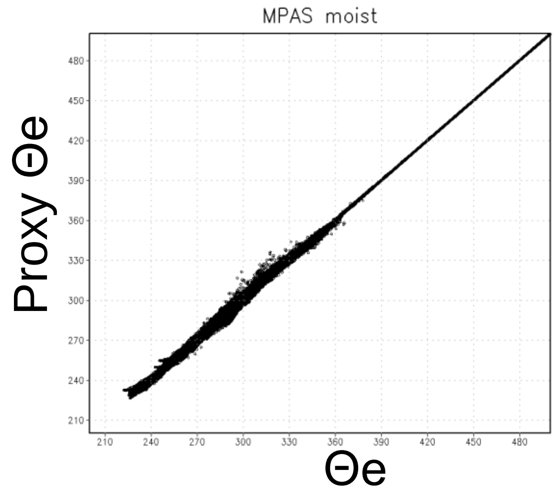


#2: Conservation Test: RMS Difference Between Advected Tracer and Dynamical Field (Day 15)

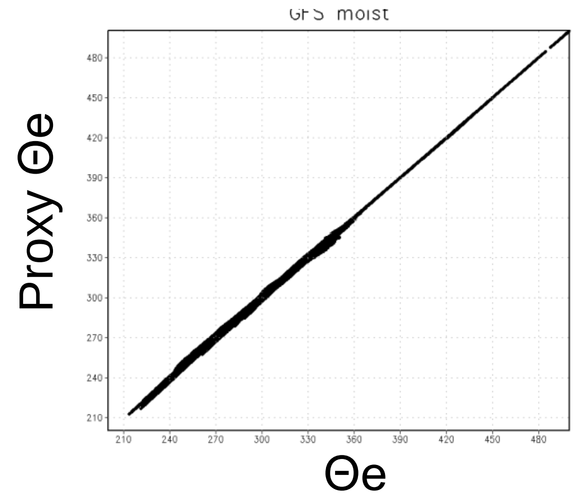
FV3 RMS=0.232



MPAS RMS=0.126



GFS RMS=0.202

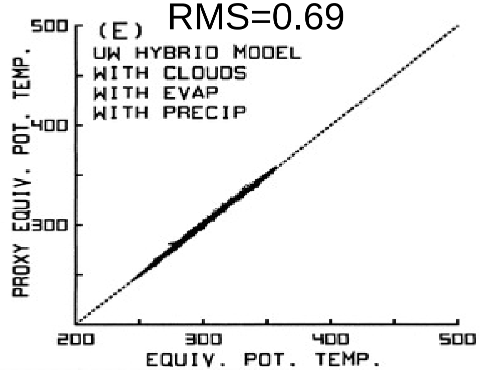


Scatterplots of Θe and proxy Θe (tracer) at day 15 for the moist baroclinic wave (DCMIP test 4.2). Compare with Figure 1 of Johnson et al. 2000.

FV3, GFS and MPAS are similar, much better than CCM3 result from Johnson et al.

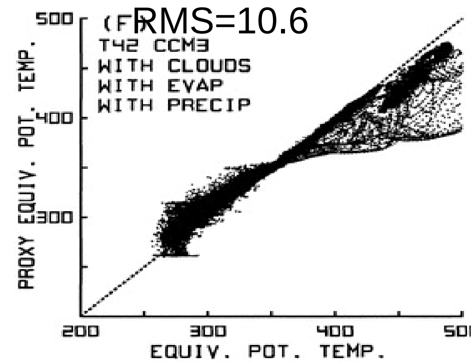
UW

RMS=0.69



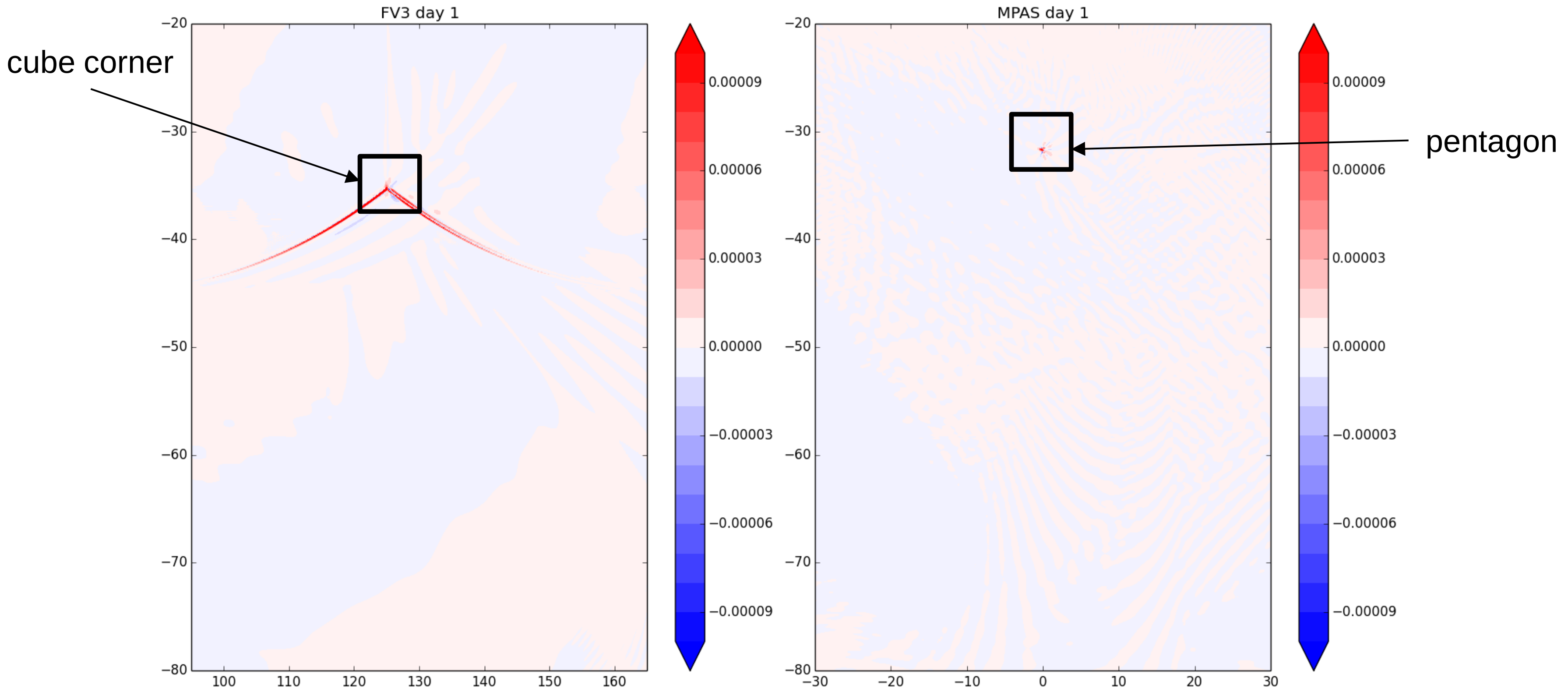
CCM3

RMS=10.6

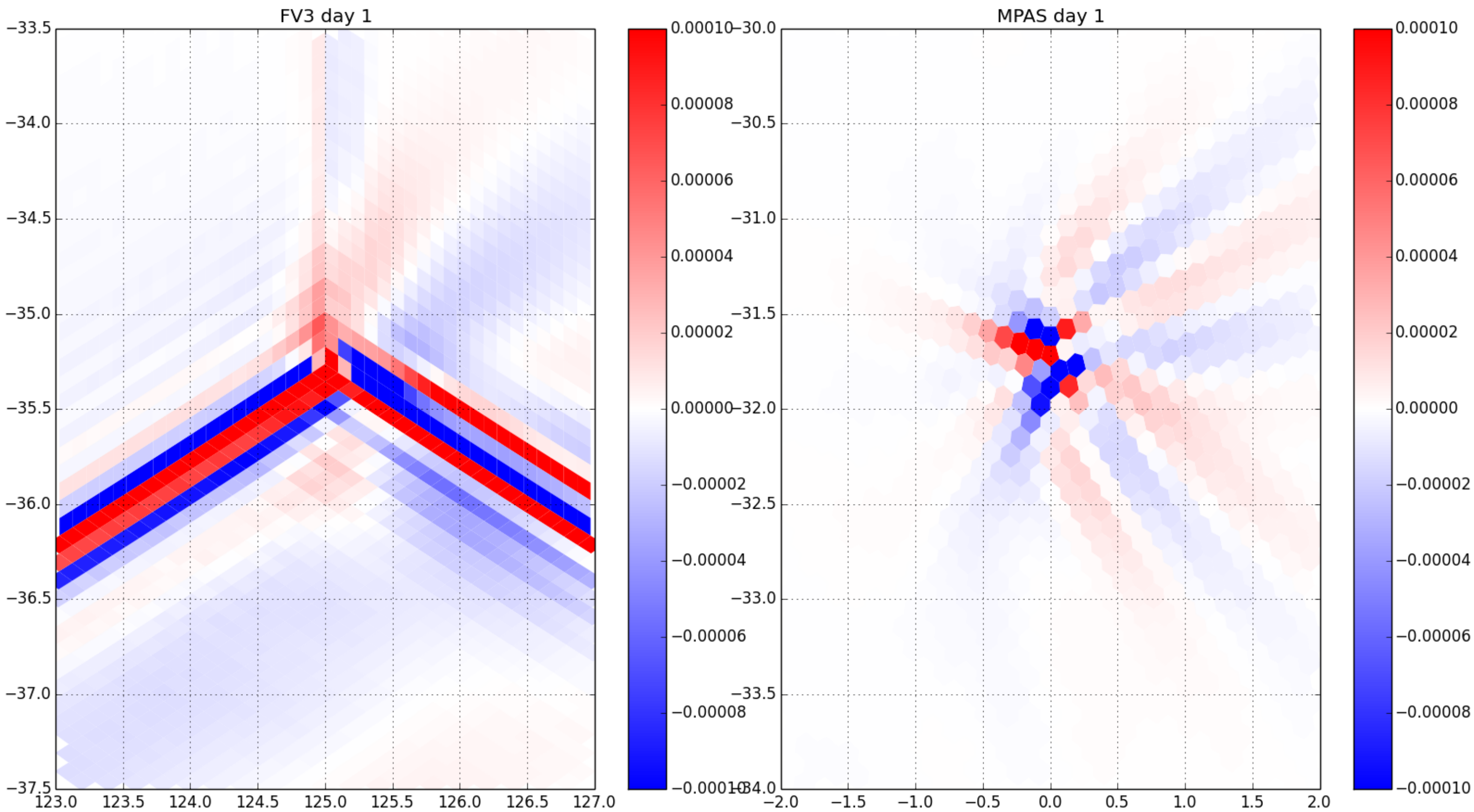


Day-10 scatter plots from Johnson et al. 2000

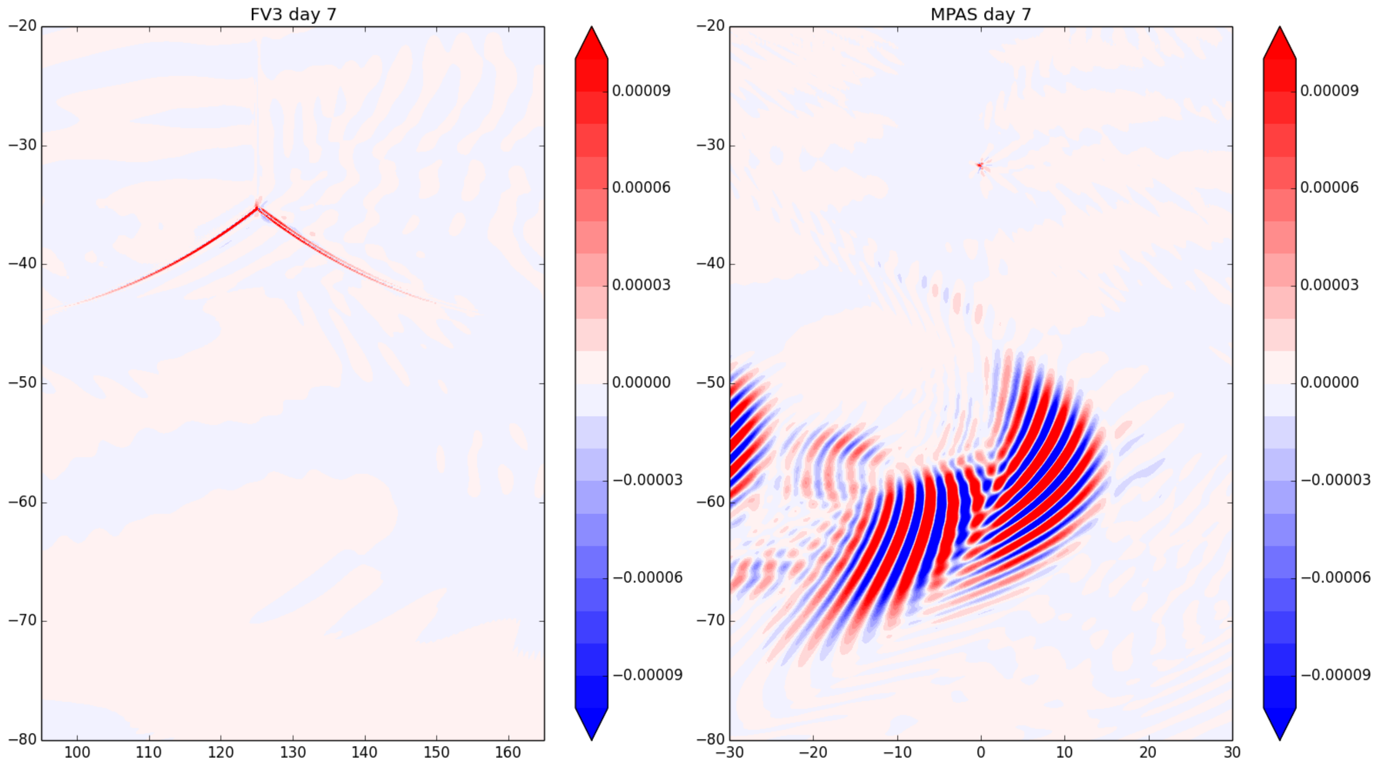
#2: Conservation Test Case (Grid Imprint Assessment): *Dry Case* (Southern Hem) Vertical Velocity at Lowest Level, Day 1 (Zonal Mean Removed)



#2: Conservation Test (Grid Imprinting Assessment): Zoom-in on Cube Corner, Pentagon (Level 1 w)



#2: Conservation Test (Grid Imprinting Assessment): Level 1 w at day 7

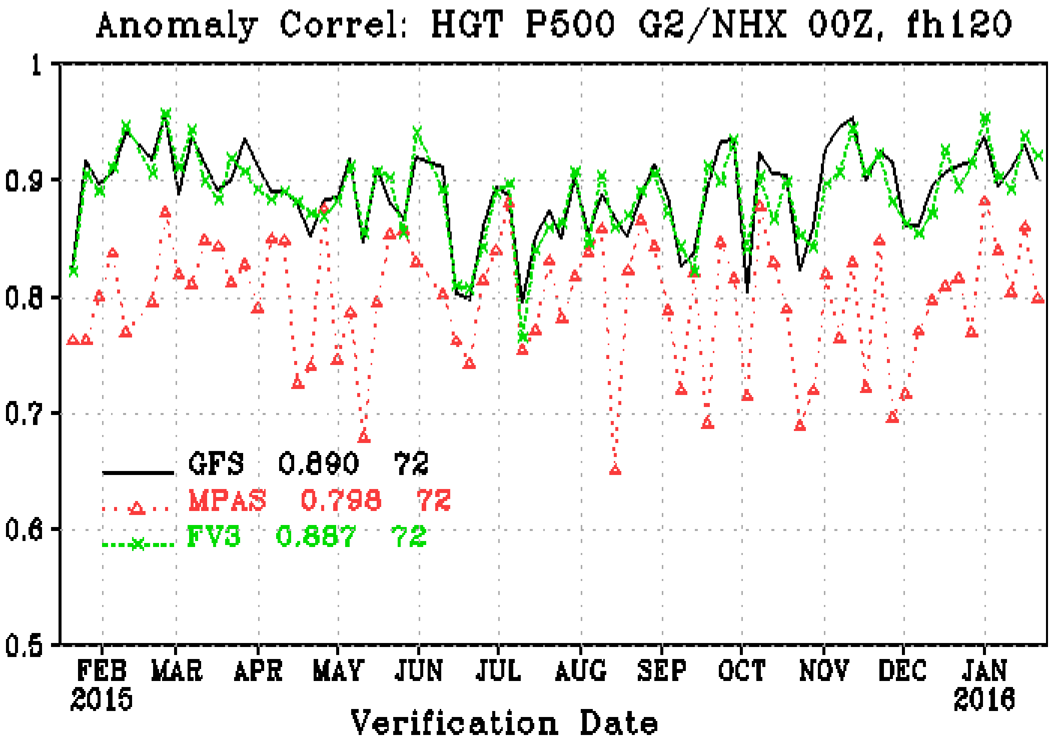


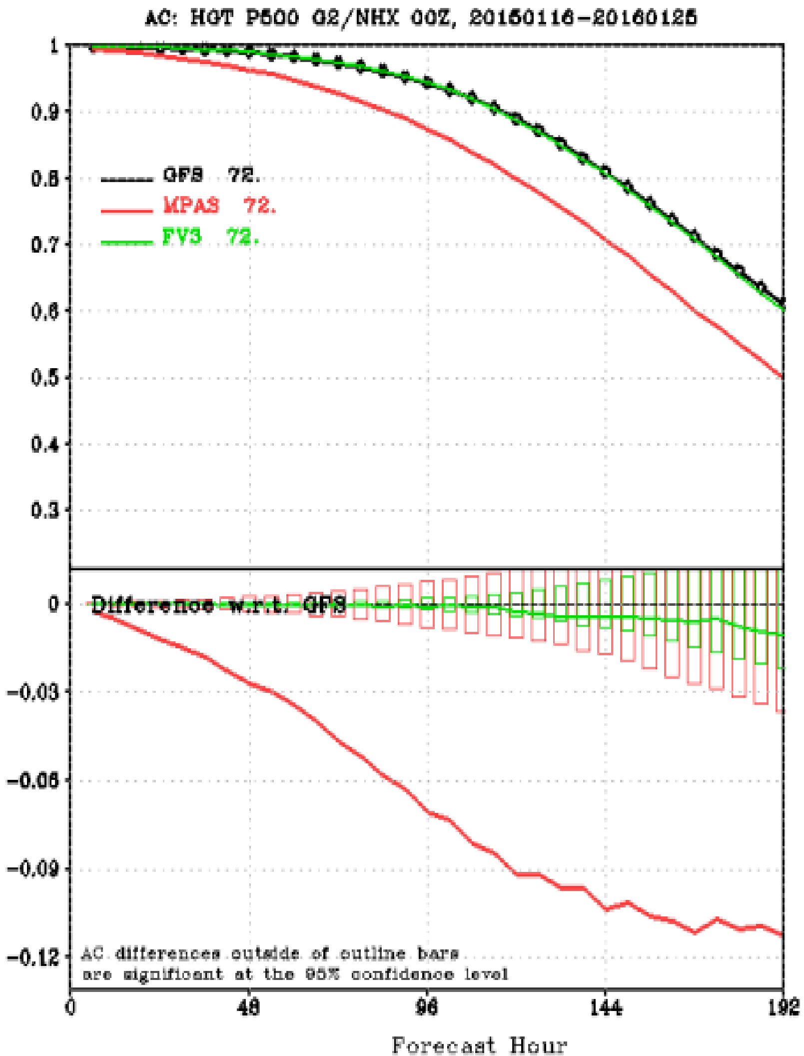
#3: Retrospective 13 km 10-d Forecasts with GFS physics

- GFS physics package (provided by EMC) implemented in both models by FV3 and MPAS development teams
- 74 retrospective 10-d forecasts run at 13 km resolution with 64 vertical levels, initialized from GFS analyses every 5th day for calendar year 2015
- Validated using NCEP verification suite, compared to operational GFS forecasts; statistics available at:

  http://www.emc.ncep.noaa.gov/gmb/wx24fy/nggps/web/
- Goals:
  - Assess ‘robustness’ over a wide-range of atmospheric flow conditions
  - Assess work required to replace spectral dycore in operational GFS

#3: Retrospective 13 km Forecast Skill



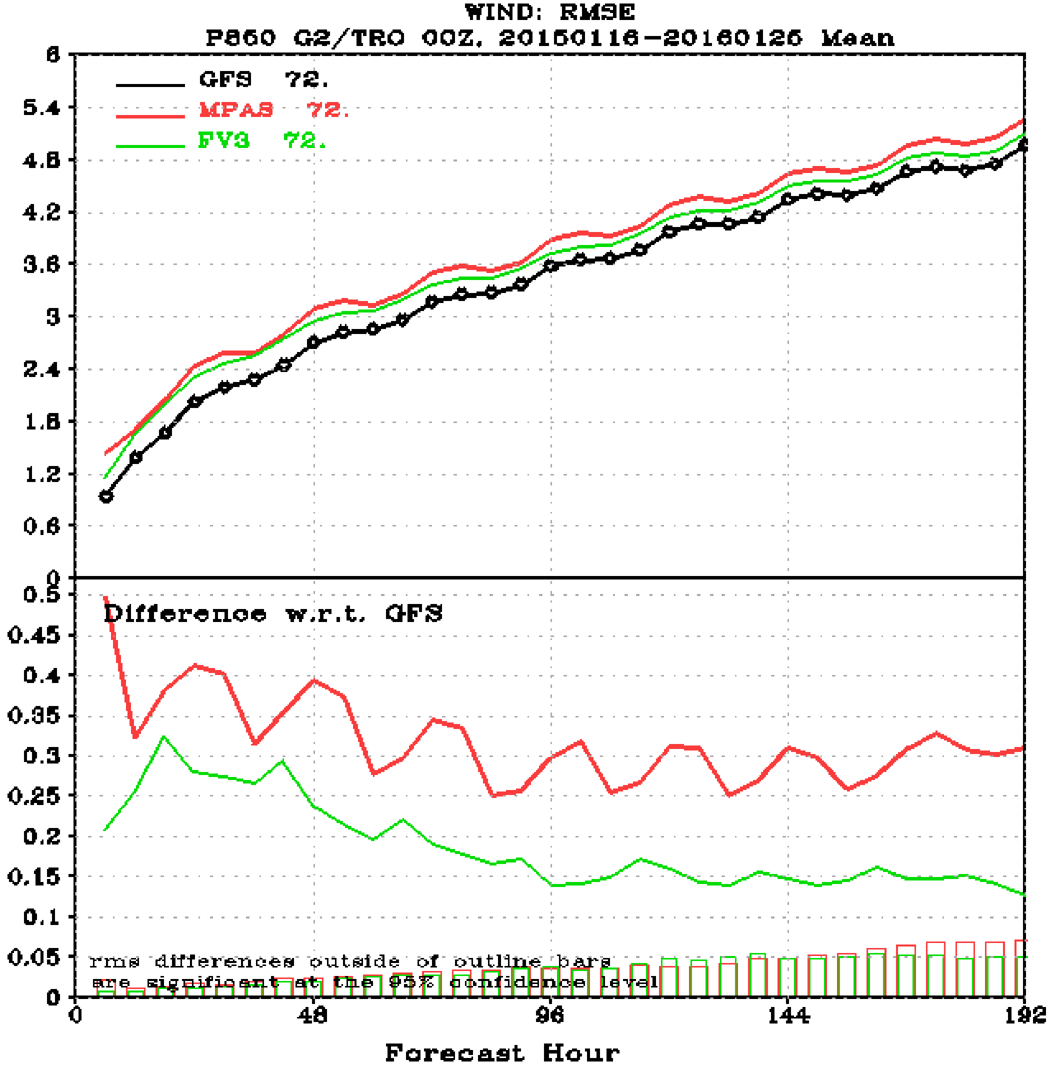


FV3 forecast skill matches the GFS using GFS ICs and GFS Physics

#3: Retrospective 13 km Forecast Skill

GFS outperforms FV3 in tropics, when GFS analysis is used for verification.

FV3 outperforms MPAS in tropics, but not by as much as in mid-lats.



#4: Performance Benchmark: Methodology

- GFS physics runs with double (64b) floating point precision
- Configurations same as for retro forecasts
- 3 nominal resolutions: 15 km, 13 km, 11 km; 63 levels (so differences in effective resolution could be accounted for). Benchmark parameters agreed to by NCAR and GFDL
- Dedicated access to Cori system at NERSC (similar to Luna/Surge); runs conducted on otherwise empty machine
- Metric: Number of processors required to achieve 8.5 minutes per day simulation rate
- Multiple runs varying numbers of processors to straddle 8.5 min/day simulation rate
- Also tested were:
  - Efficiency of mesh refinement strategies (using configuration for criteria #5)
  - Performance with 15 and 30 extra tracers

#4: Performance Benchmark Results: Configurations

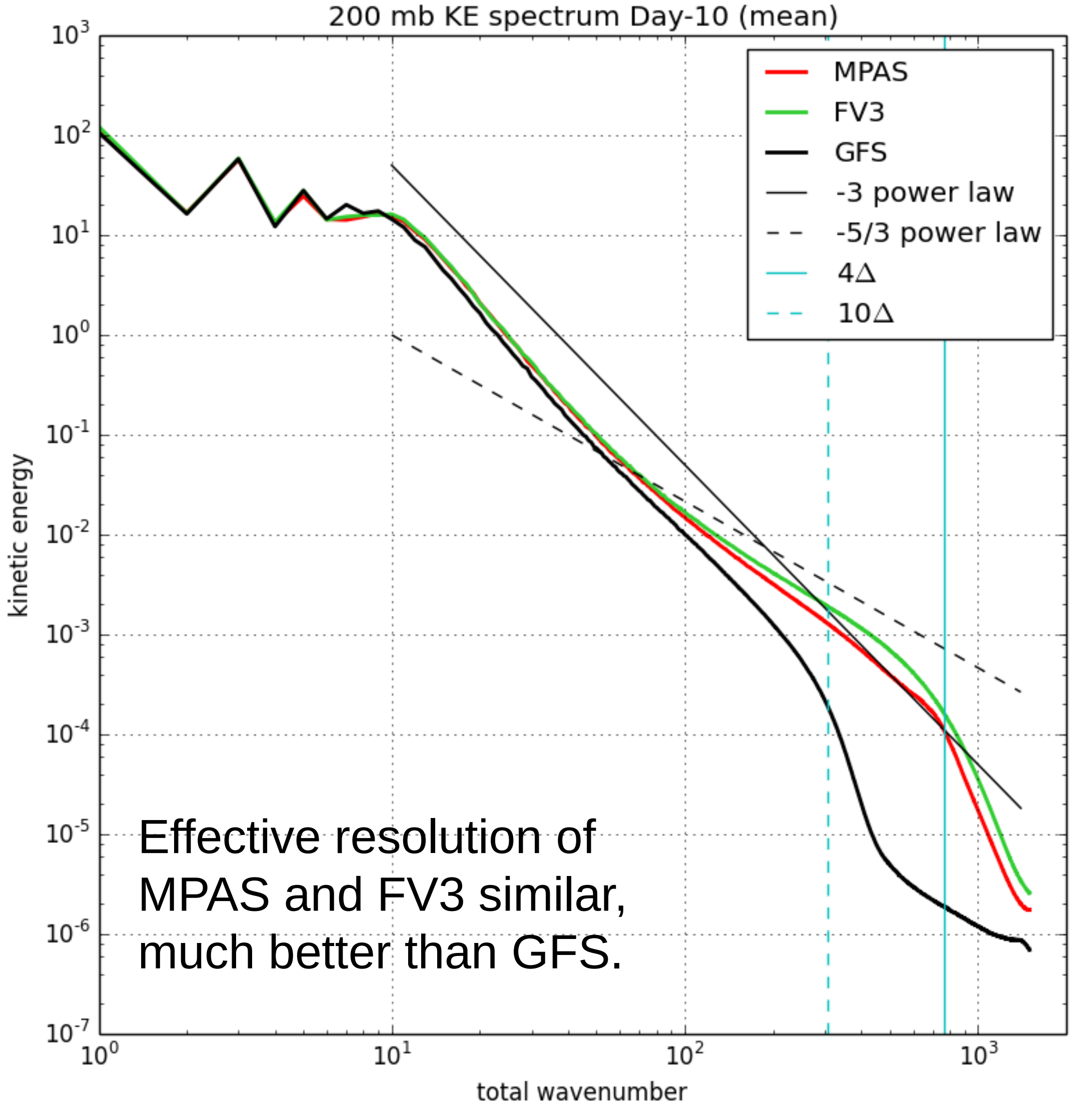
Eval. Criterion #4 -- Performance with GFS Physics

	FV-3	MPAS
Nominal resolution (km)	13.03 (equat.), 12.05 (avg.)	13
Grid Points	3,538,944	3,504,642
Vertical Layers	63	63
Time Step (sim. sec)	112.5 (dyn.), 18.75 (acous.)	75 (transport), 37.5 (dynamics), 18.75 (acoustic)
Radiation Time Step	3600	3600
Physics (other) Time Step	225	225
Tracers	3	3

Coarser than nominal resolution (km)	15.64 (equat.), 14.46 (avg.)	15
Grid Points	2,547,600	2,621,442
Vertical Layers	63	63
Time Step	225 (dyn.), 22.5 (acous.)	90 (transport), 45 (dynamics), 22.5 (acoustic)
Radiation Time Step	3600	3600
Physics Time Step	225	180

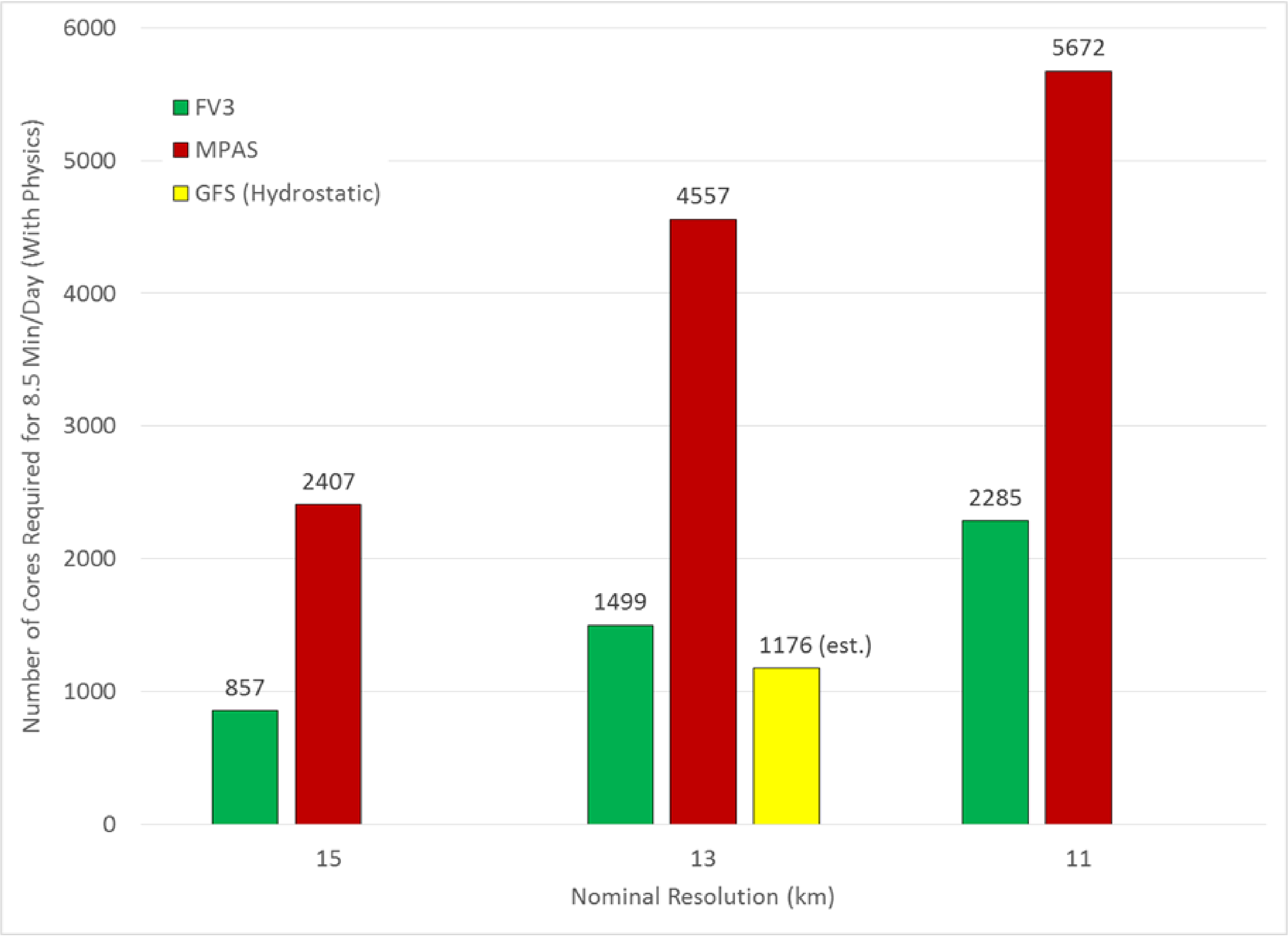
Finer than nominal resolution (km)	11.72 (equat.), 10.34 (avg.)	11
Grid Points	4,816,896	4,858,092
Vertical Layers	63	63
Time Step	112.5 (dyn.), 16.07 (acous.)	60 (transport), 30 (dynamics), 15 (acoustic)
Radiation Time Step	3600	3600
Physics Time Step	225	180

#4 Performance Benchmark: KE Spectra (Effective Resolution)



FV3, MPAS and GFS 10-d forecasts at 13km nominal resolution.

#4: Performance Benchmark Results (J. Michalakes)



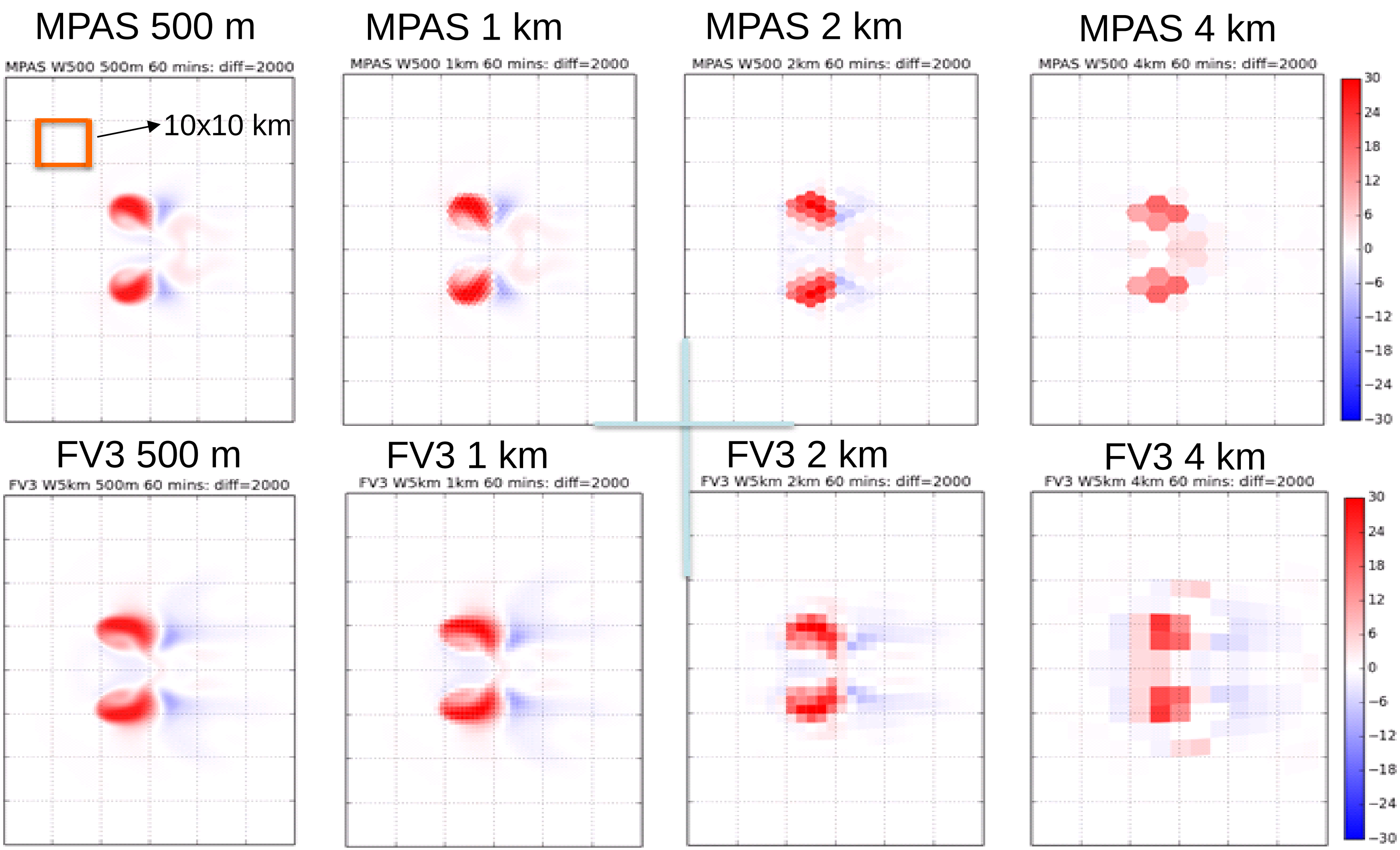
#5: Demonstration of Variable Resolution and Nesting Capabilities

Includes simulations of convection in the high-resolution region, and includes supercell and tropical cyclone (TC) idealized tests

- Real-data forecasts:
  - Mesh varies from 13 km to 3 km over CONUS
  - GFS physics with deep convection disabled
  - Initial conditions for 2013051800 (Moore tornado) and 2012102418 (Hurricane Sandy), forecasts run to 10 days
  - MPAS used a non-uniform mesh, FV3 used a combination of a global stretched grid and a nest
- Idealized tests:
  - Since cases chosen involve severe convection and tropical cyclones, companion idealized tests used to isolate impact of dynamical core on simulations of these phenomena (with highly idealized physics and no mesh refinement)
  - Supercell test (DCMIP-2016, reduced sphere 0.51/2/4 km) also run in Phase I, but not with identical diffusion settings
  - TC test from DCMIP-2012 (full sphere, 13 km)

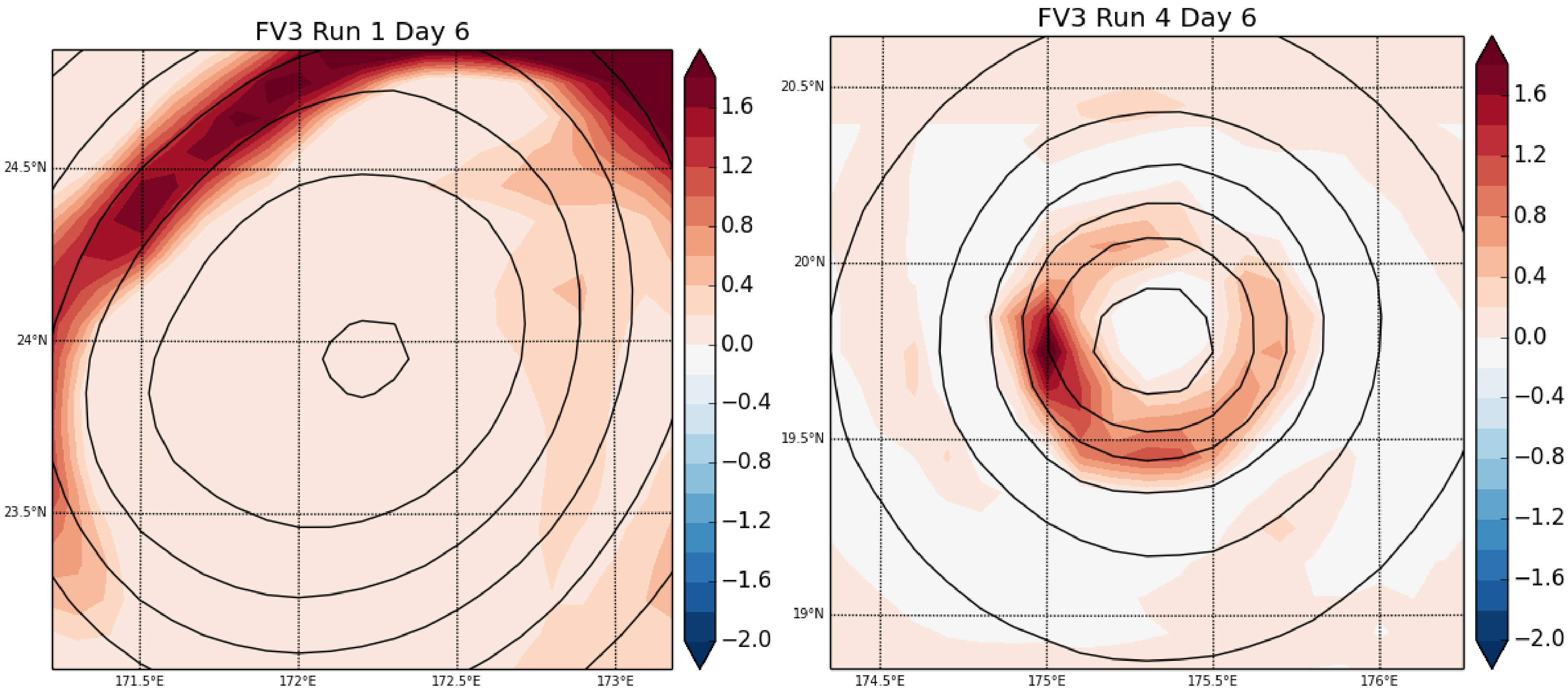
#5: Idealized Supercell Test

500 hPa Vertical Velocity (m/s), All Resolutions



#5: Idealized TC Test

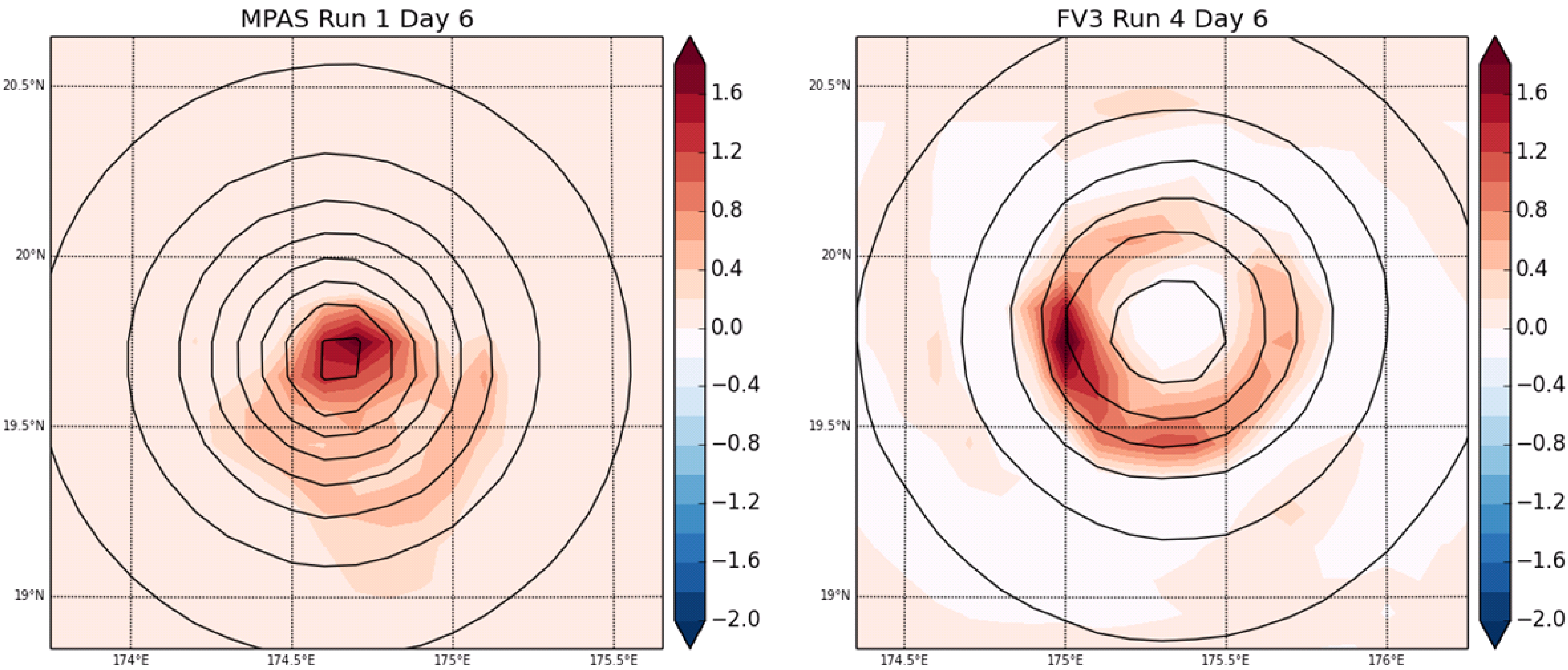
MSLP (Black Lines), 500hPa Vertical Velocity (color, m/s)



FV3 as originally configured has a huge eye (left); removing the vertical 2dz filter produced a much smaller, more realistic storm structure (right).

#5: Idealized TC Test

MSLP (black lines), 500hPa Vertical Velocity (color, m/s)

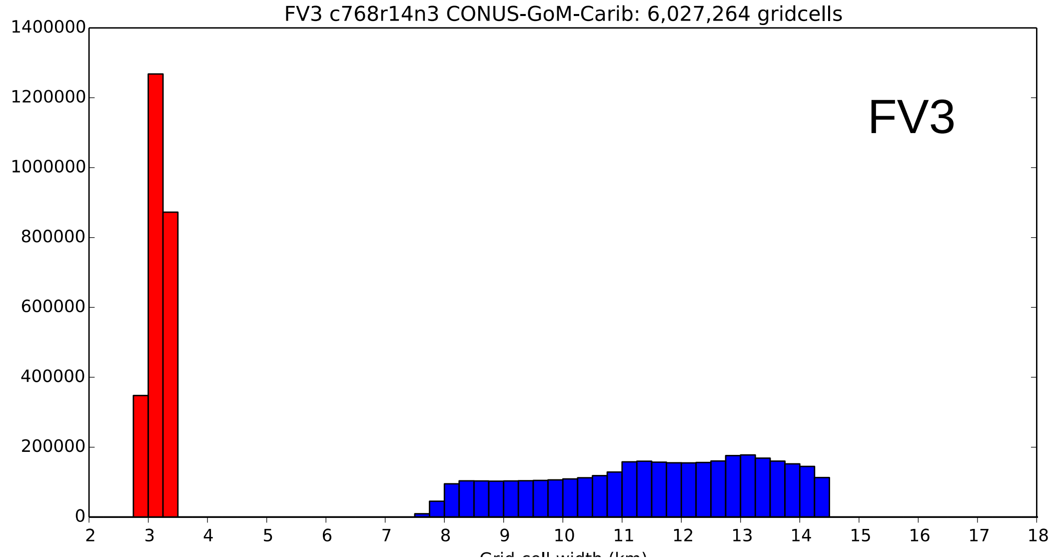


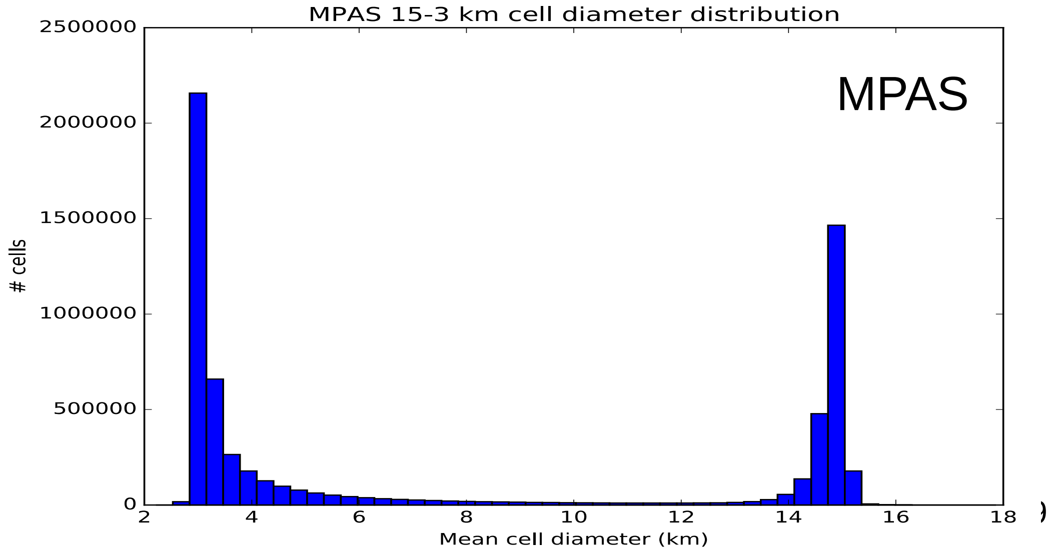
MPAS updraft is maximum in center of storm – no local minimum in eye.
FV3 updraft is still concentric, with subsidence in eye.

***MPAS real-data TC simulations did not have this structure.**

#5: Variable Resoluton Configurations

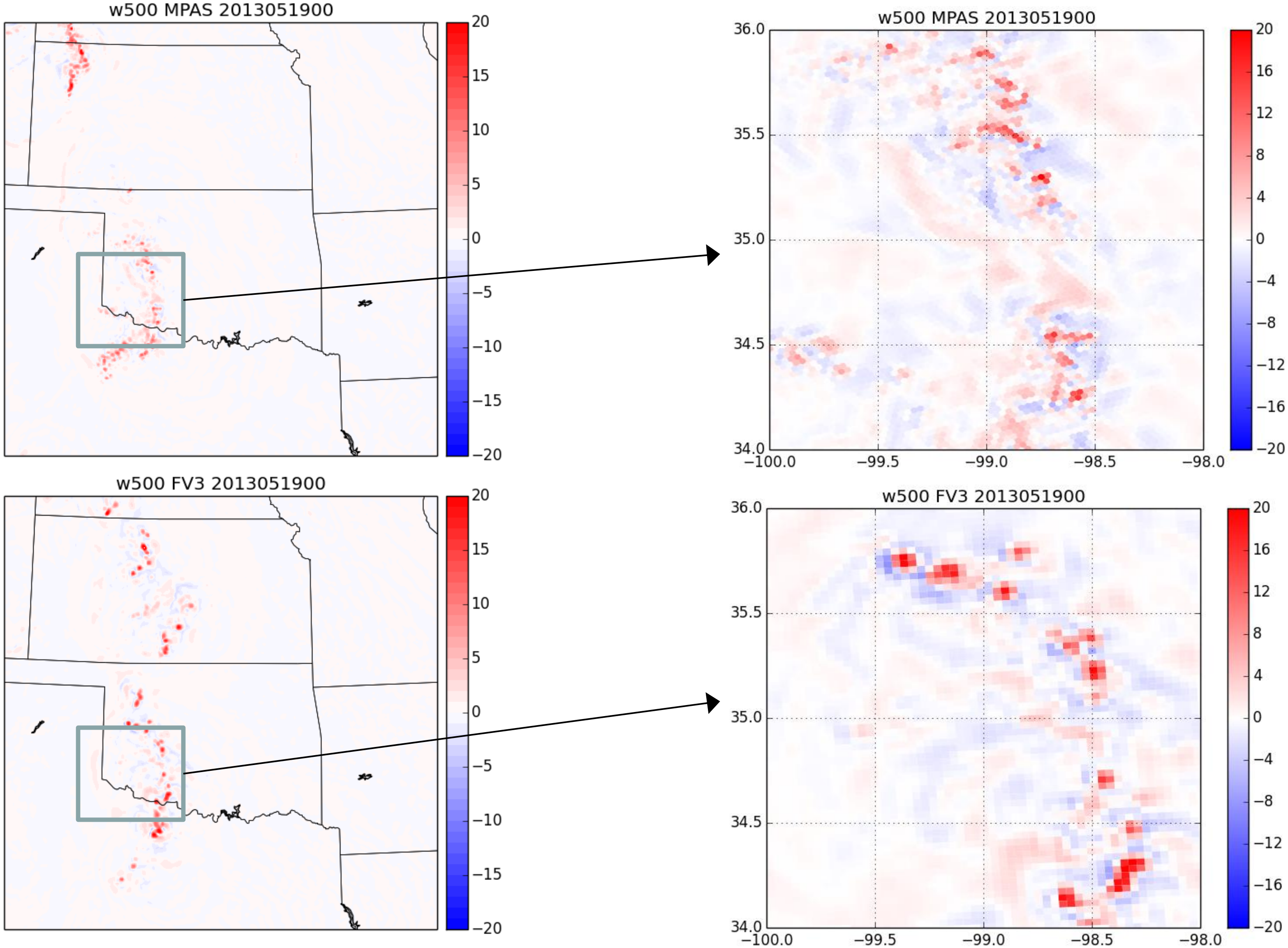
Histograms of grid cell size





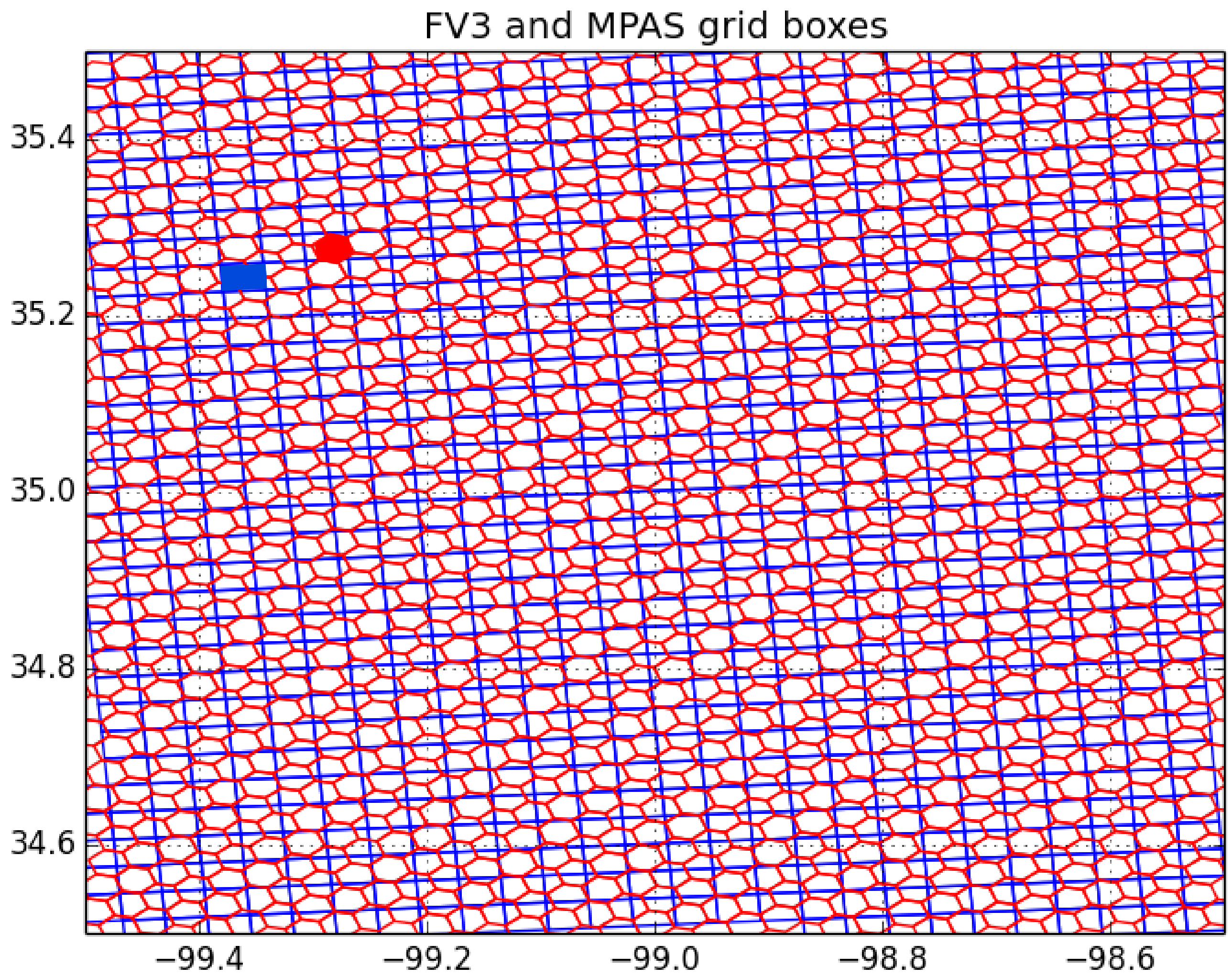
#5: Variable Resolution Tests

Moore Tornado Case – 24h Fcst Valid 00UTC May 19
500hPa Vertical Velocity (m/s)



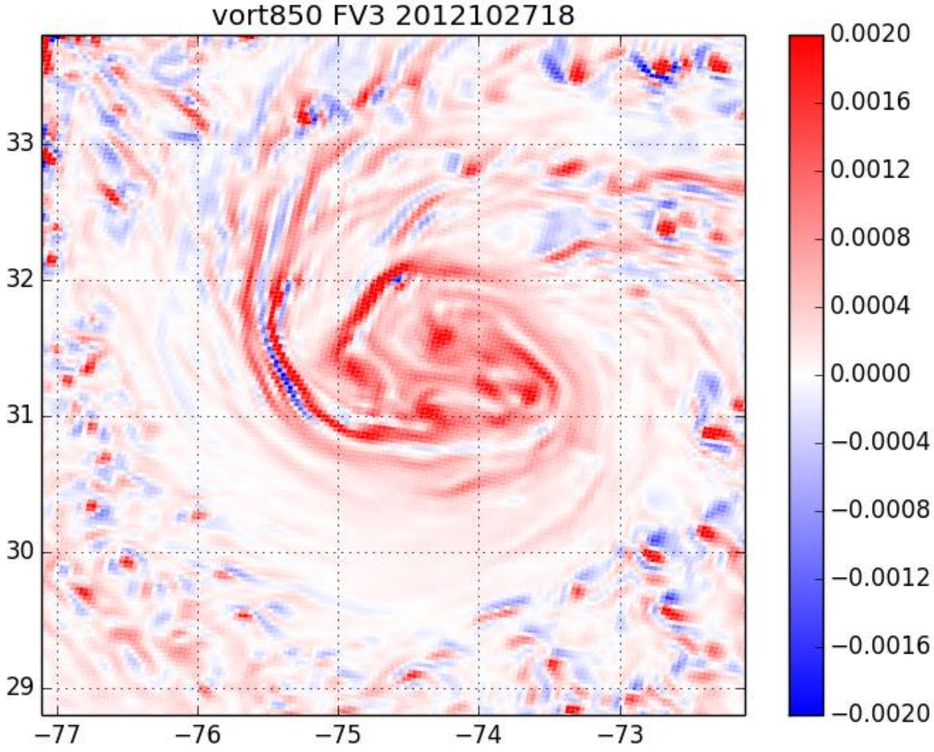
#5: Variable Resolution Tests: Grid Structure in Region of Interest

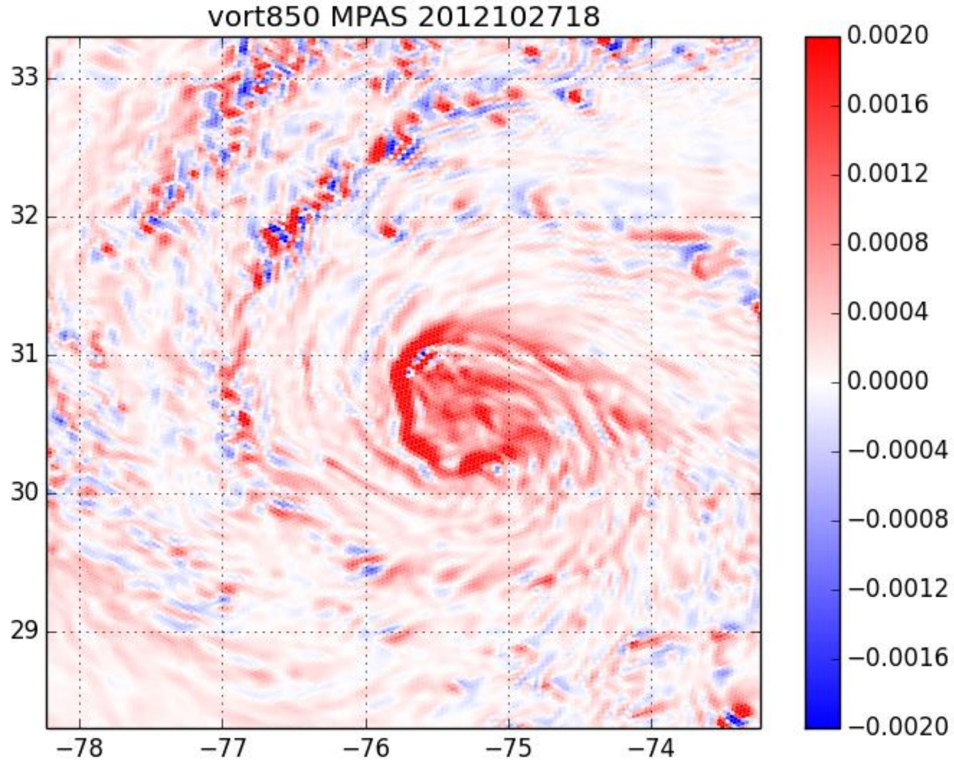
MPAS grid cells (red) are smaller in the region of interest



#5: Variable Resolution Tests:

Hurricane Sandy Case: 72h Fcst Valid 18 UTC Oct 27
850 hPa Vertical Vorticity (s^{-1})

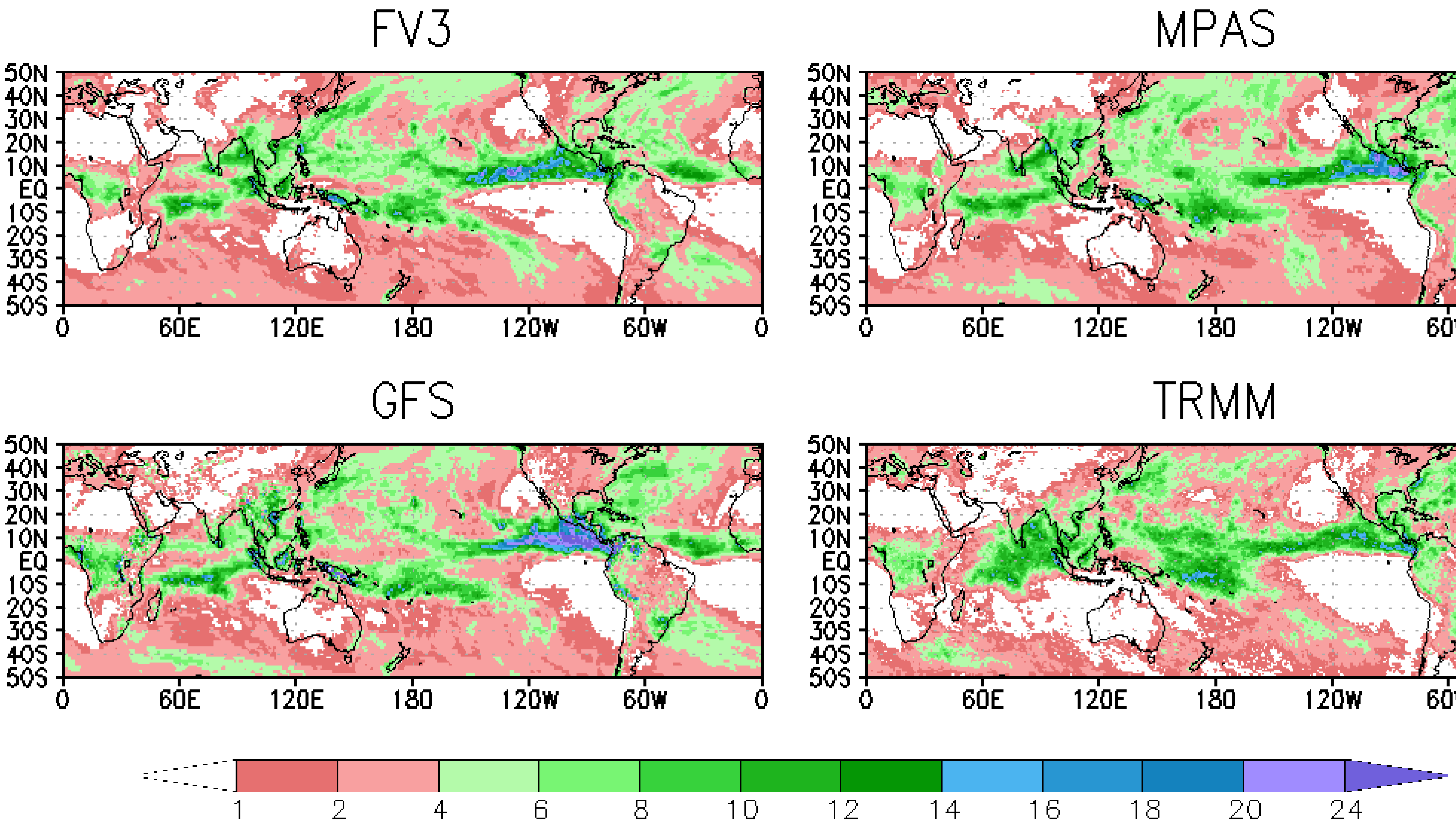




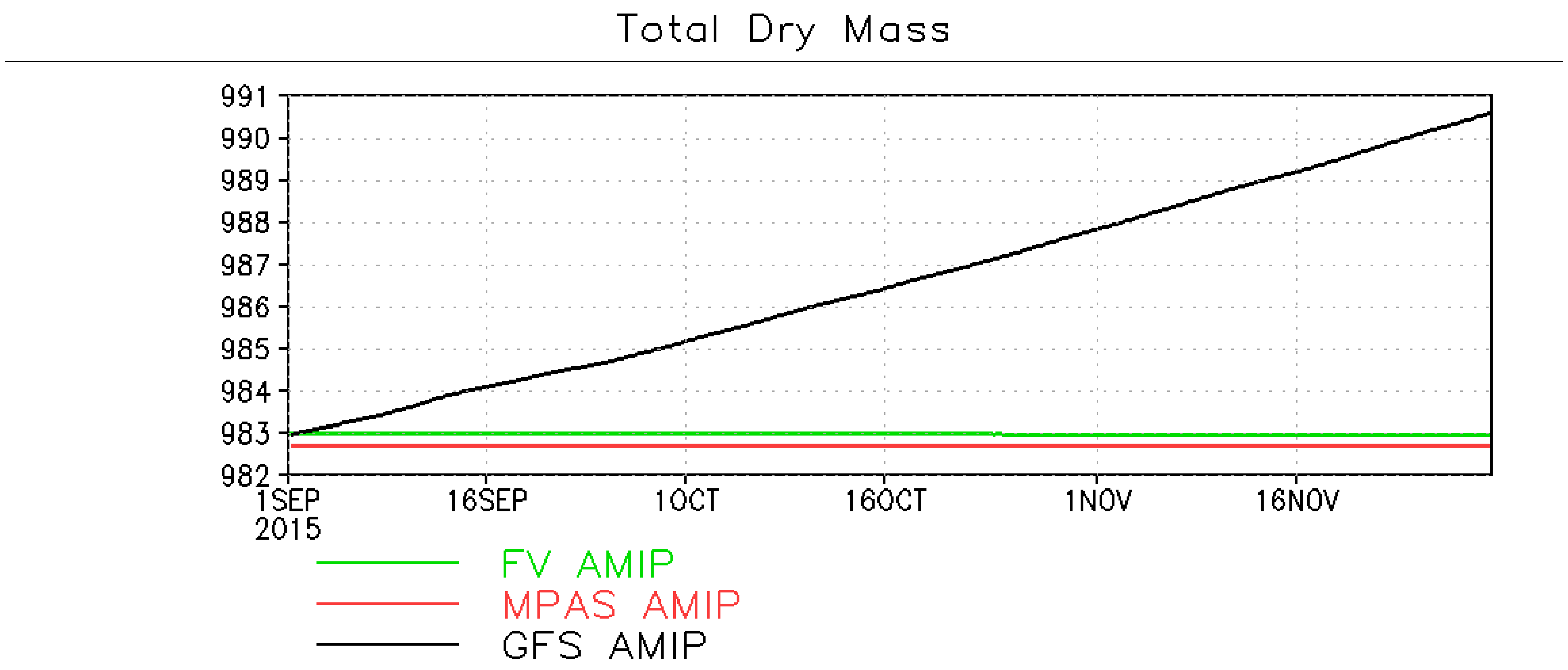
#6: Stable, Conservative Long Integrations with Realistic Climate Statistics

- 90 days runs at reduced resolution (~50 km), from GFS 00UTC Sep 1 2015 analysis, with surface conditions updated every 6 hours
- Assessment will include:
  - 90-day mean statistics
  - Time series of dry mass, energy
  - Detection of ‘grid imprint’

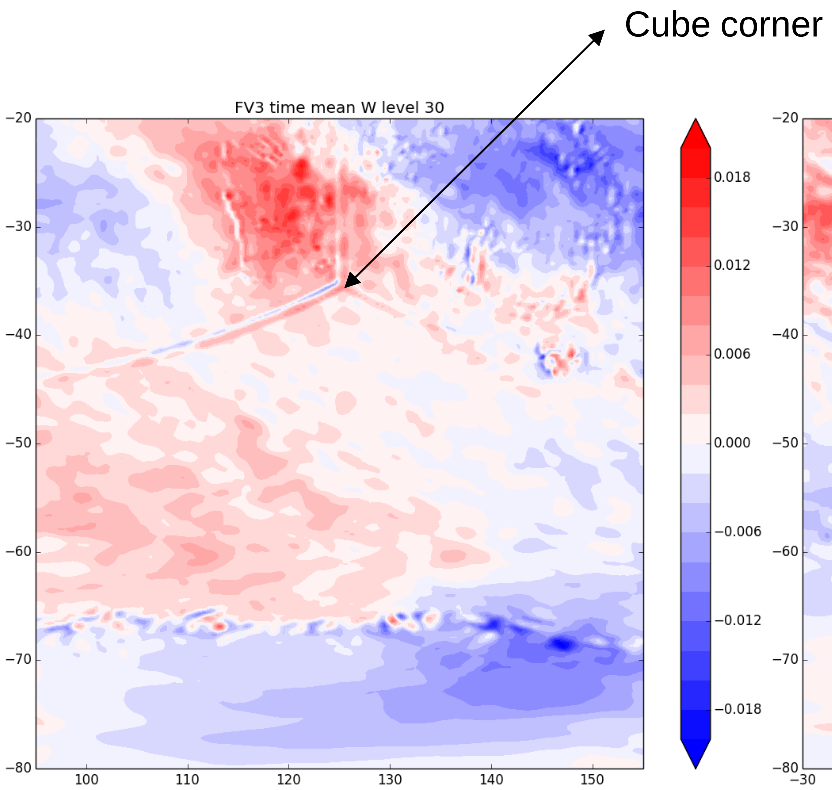
#6: Stable, Conservative Long Integrations with Realistic Climate Statistics
Day 0-90 Mean, IC 2015090100, ~50 km Resolution

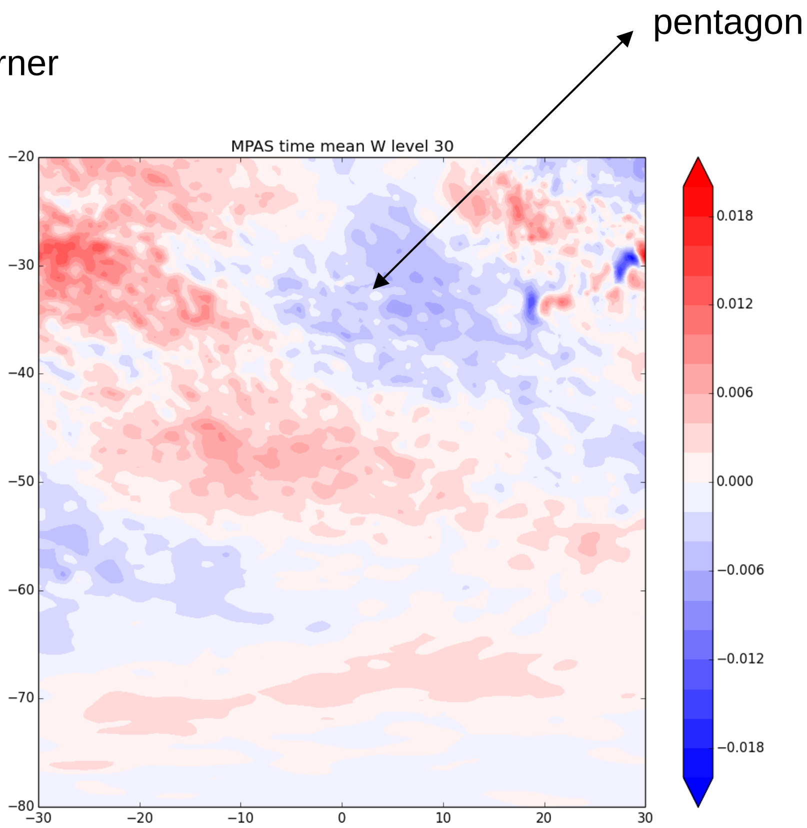


#6: Stable, Conservative Long Integrations with Realistic Climate Statistics
Day 0-90 Mean, IC 2015090100, ~50 km Resolution

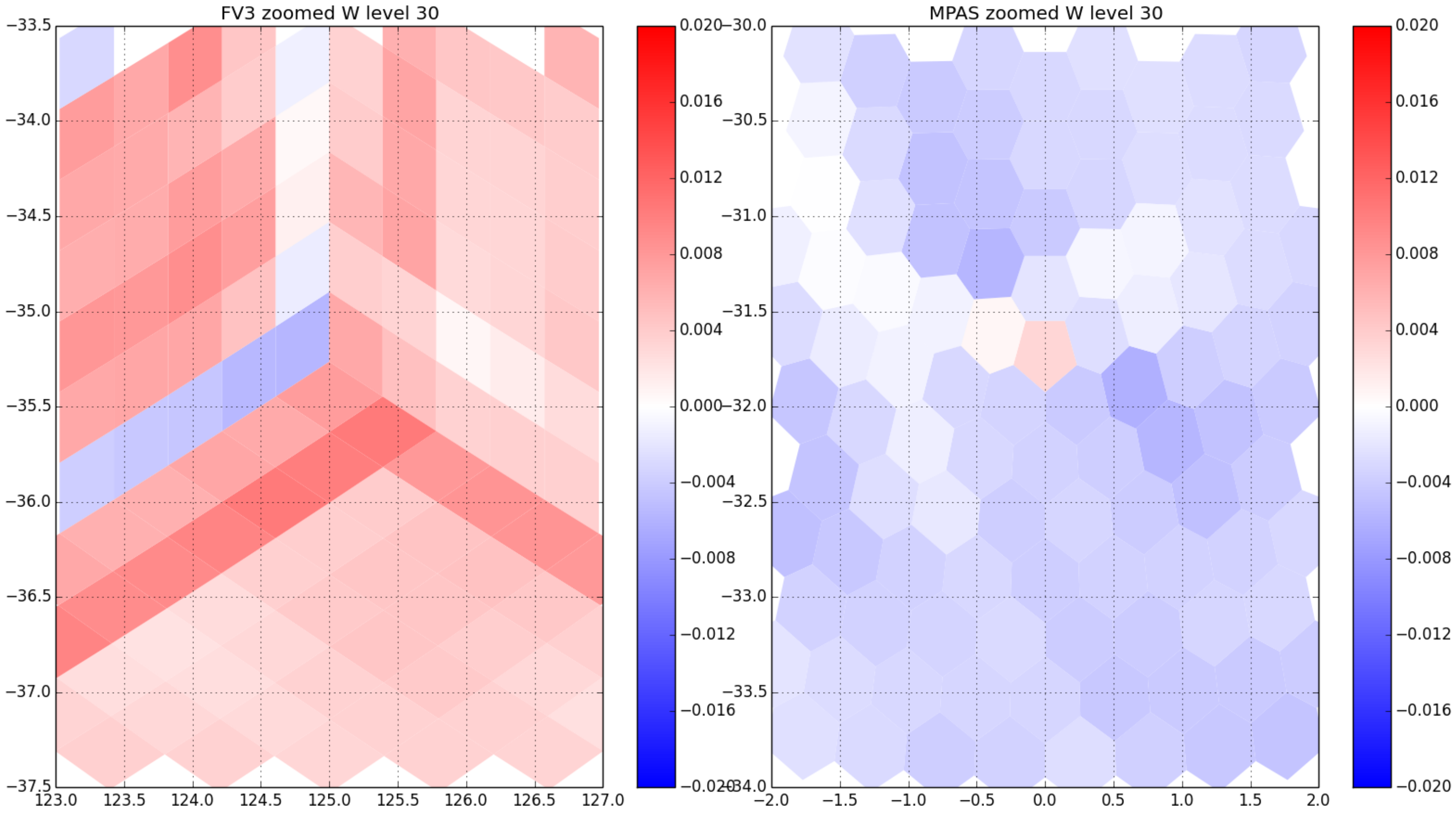


#6: Long integrations: grid imprinting 90 day mean w at level 30





#6: Long integrations: grid imprinting 90 day mean w at level 30 (zoom in)

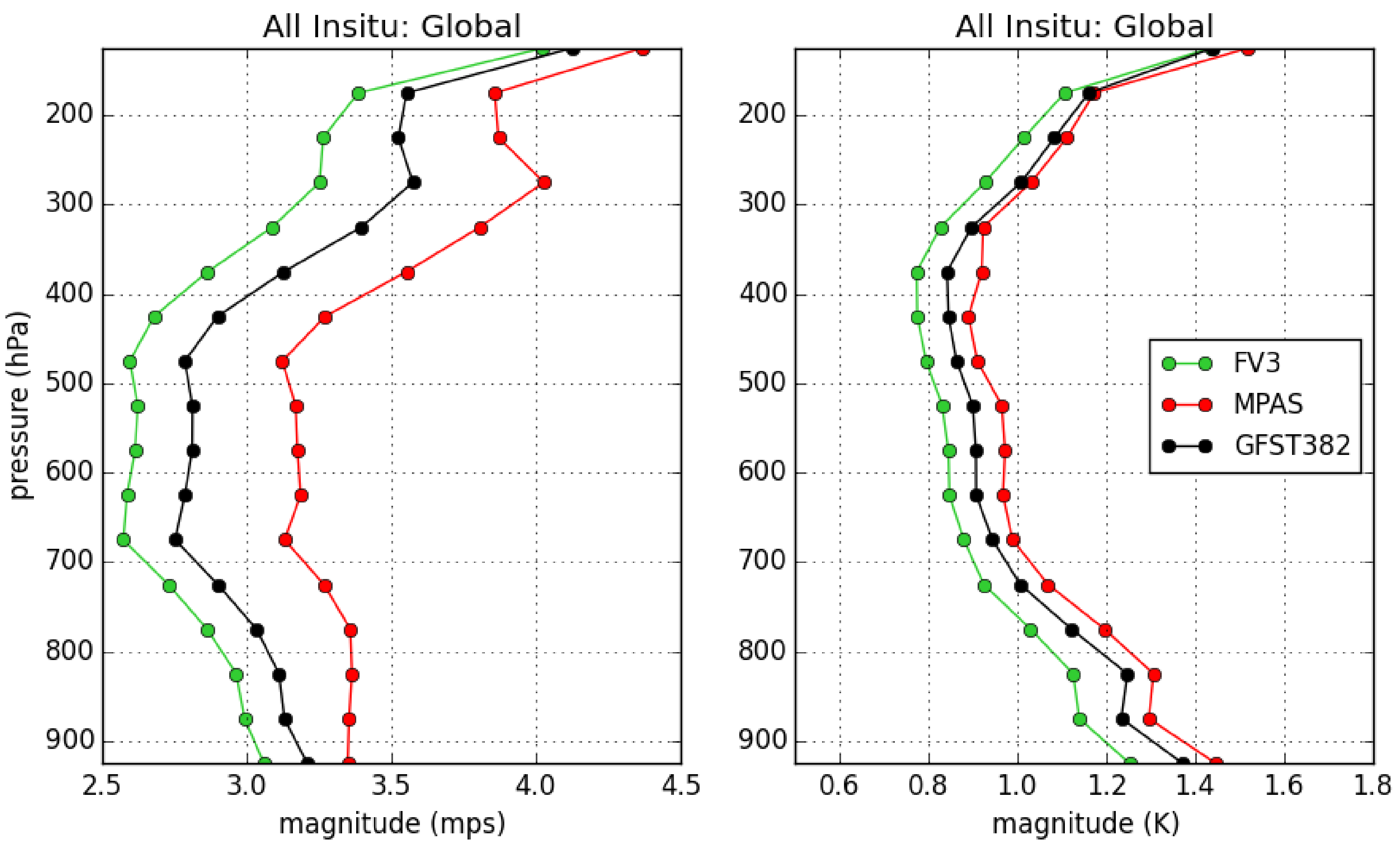


#9: Evaluation of Performance in Cycled Data Assimilation (DA)

- Both models interfaced to operational 4D ensemble-variational DA system
- Due to time and HPC constraints, tests run at reduced resolution (~50 km)
- 80 member ensemble, cycle started at 2015090100
- Differences with operational configuration:
  - No high-resolution control analysis
  - No static background error component (full ensemble used to maximize feedback between dycore and DA)
  - No digital filter or tangent-linear balance constraint
  - No stochastic physics in ensemble (multiplicative inflation increased to compensate)
- Baseline GFS experiment at T382 resolution for reference
- Assessing:
  - Work required to replace spectral dycore in GDAS
  - Whether issues arise that may not be evident when models initialized from ‘foreign’ analysis

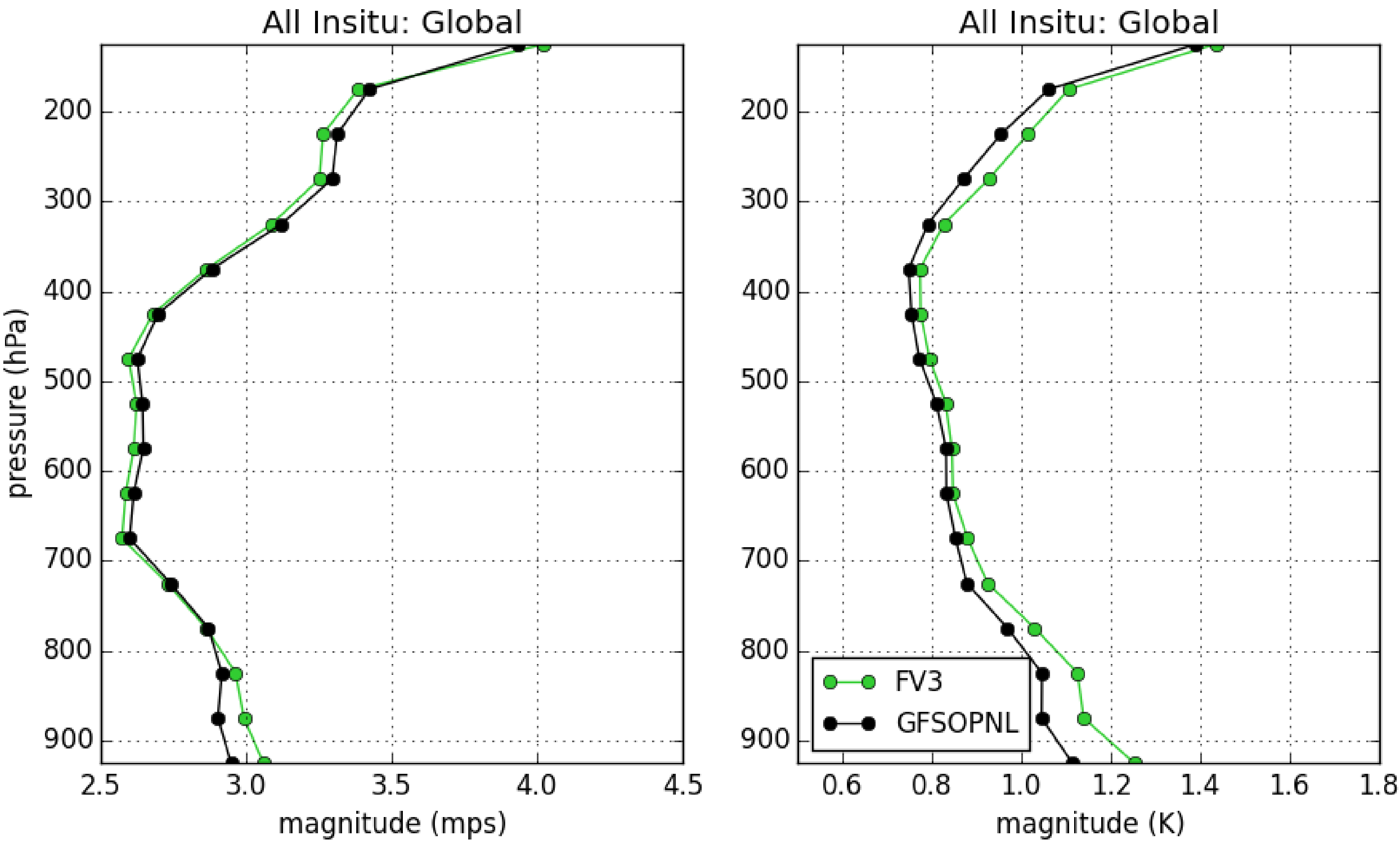
#9: DA Cycling: RMS Fit of First-Guess to All In-situ Observations

Vector Wind (left) and Temp (right) O-F (2015090500-2015092900)



#9: DA Cycling: RMS Fit of First-Guess to All In-situ Observations

Vector Wind (left) and Temp (right) O-F (2015090500-2015092900)



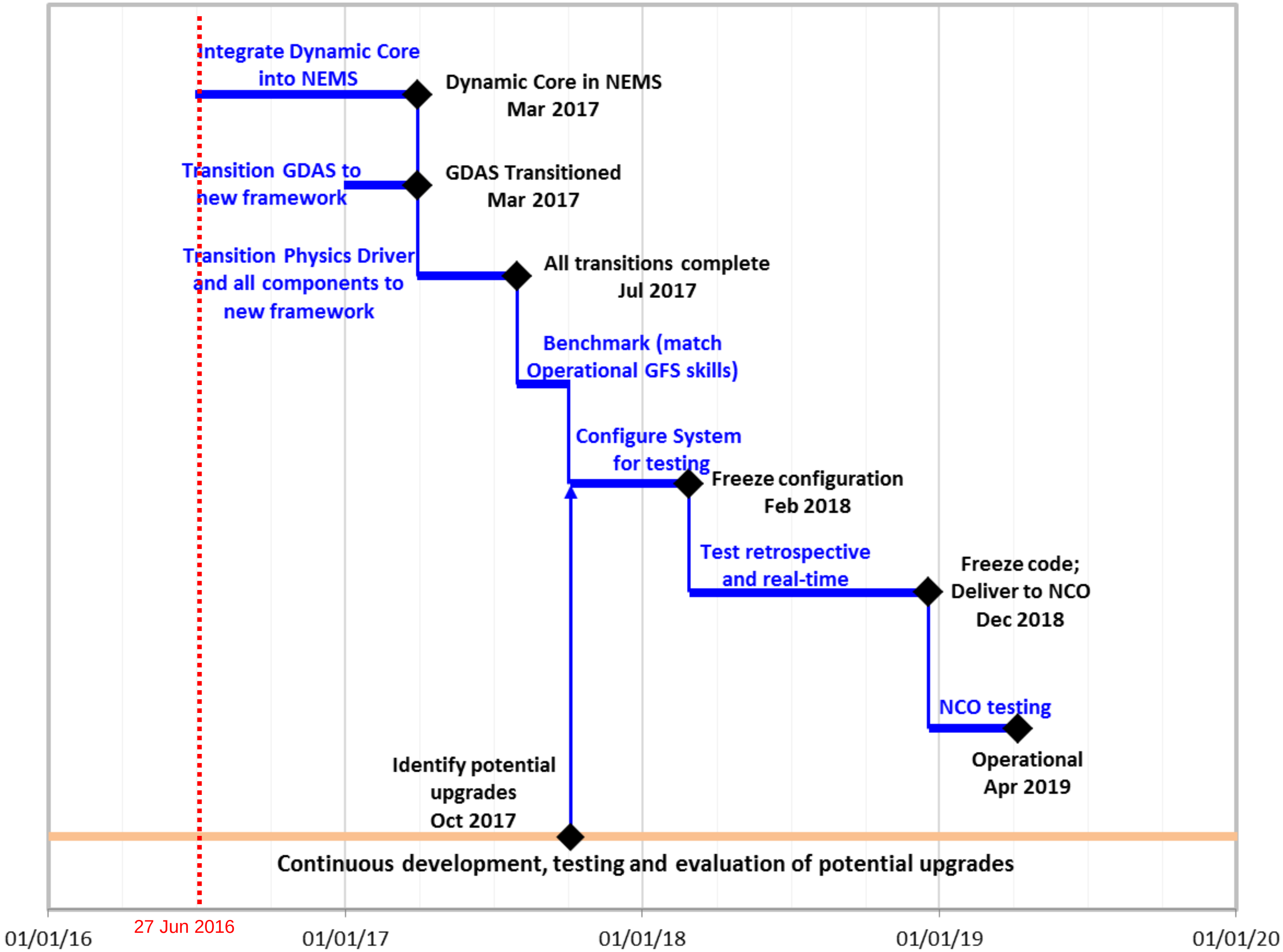
Overall Assessment and NGGPS Program Manager Recommendation

The FV3 core represents the lowest risk, lowest cost alternative for the new NGGPS atmospheric model

- Compared to the MPAS, FV3:
  - Meets all technical needs
  - Less expensive to implement
  - Higher readiness for implementation
  - Significantly better technical and computational performance
  - Lower risk
- NGGPS strategy has always been to find and implement the best global model (not the best convective scale model, although nothing in results precludes eventual global/convective-scale unification based on FV3)

Recommendation: Select GFDL FV3 and proceed to NGGPS Phase 3 dynamic core integration and implementation

Phase 3 Implementation Detail



Strategy to Implement Community Model Environment

- **Q1FY17: Hold a workshop to collect input on how to structure the community model environment, including:**
  - Code hosting environment (e.g. github)
  - Processes for O2R and R2O
  - Governance
  - How will support be provided?
  - What models will be supported (atmosphere dycore, ocean, land…)?
- **Develop detailed documentation, include users guide**
- **Q1FY18: Code released, with documentation**
- **Q1FY19: First users workshop/tutorial**

Questions?

NGGPS Website:

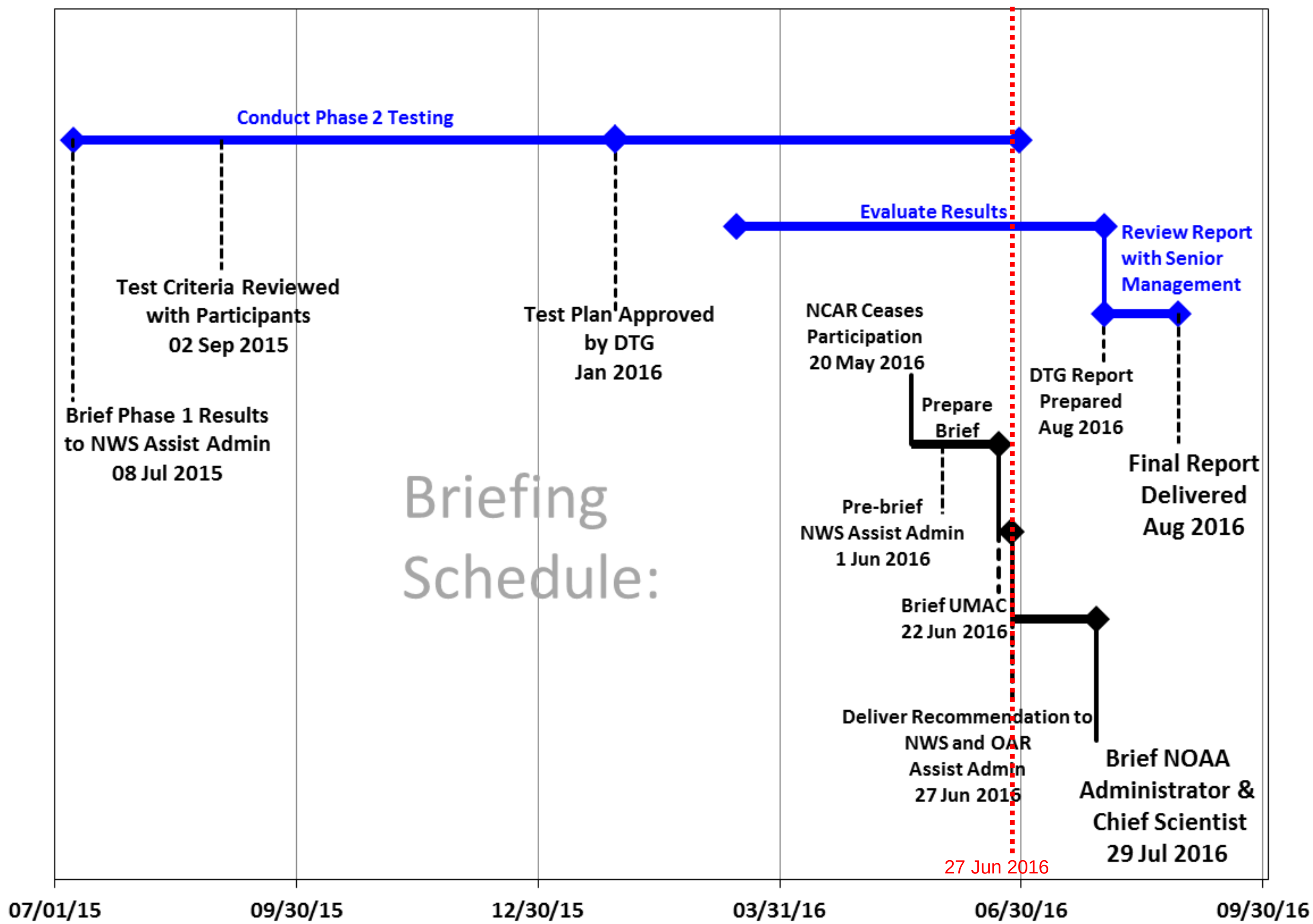
http://www.weather.gov/sti/stimodeling_nggps

Information on NGGPS dycore testing is available at:

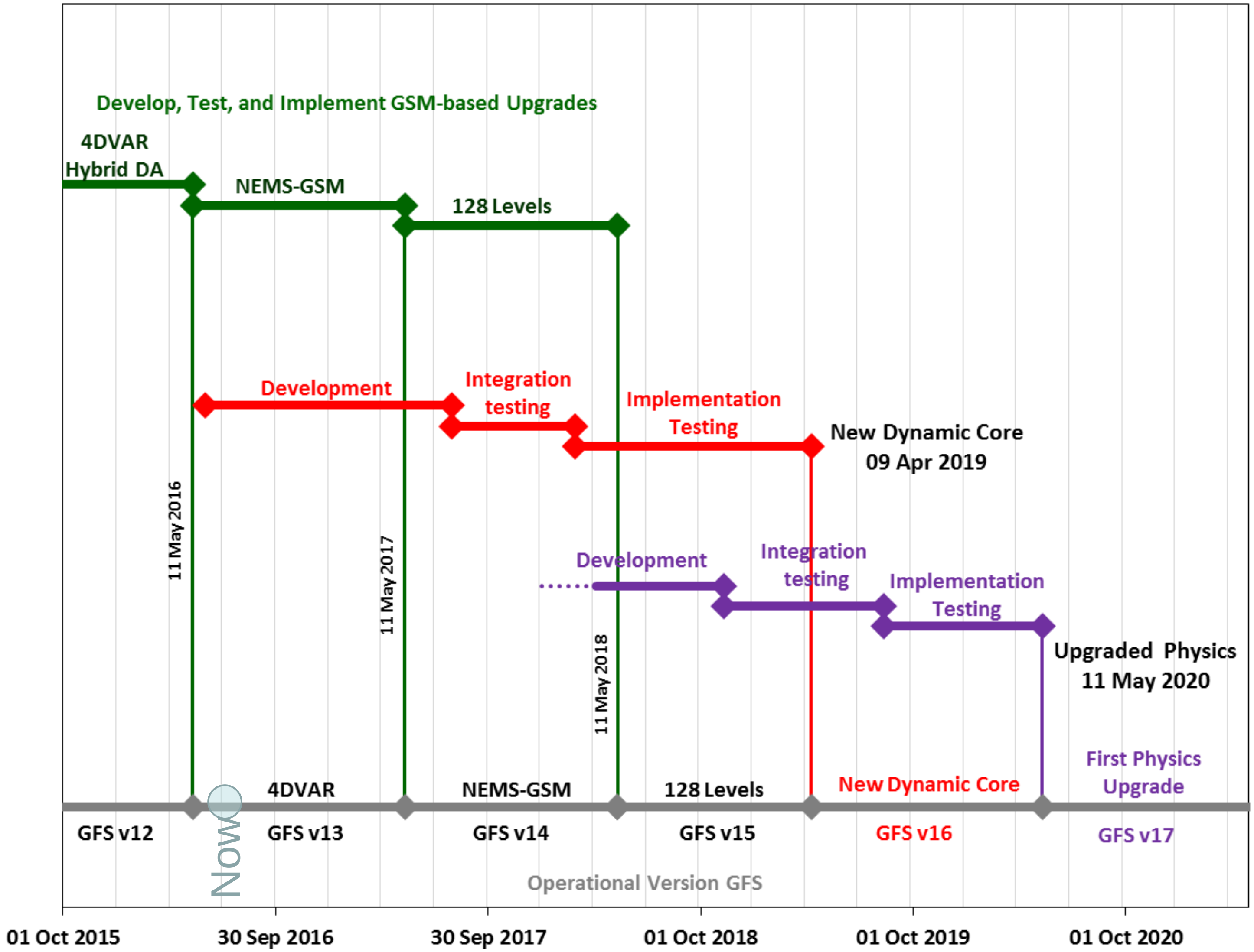
http://www.weather.gov/sti/stimodeling_nggps_implementation_atmdynamics

Back-Up Slides

Phase 2 Testing, Evaluation and Reporting Schedule



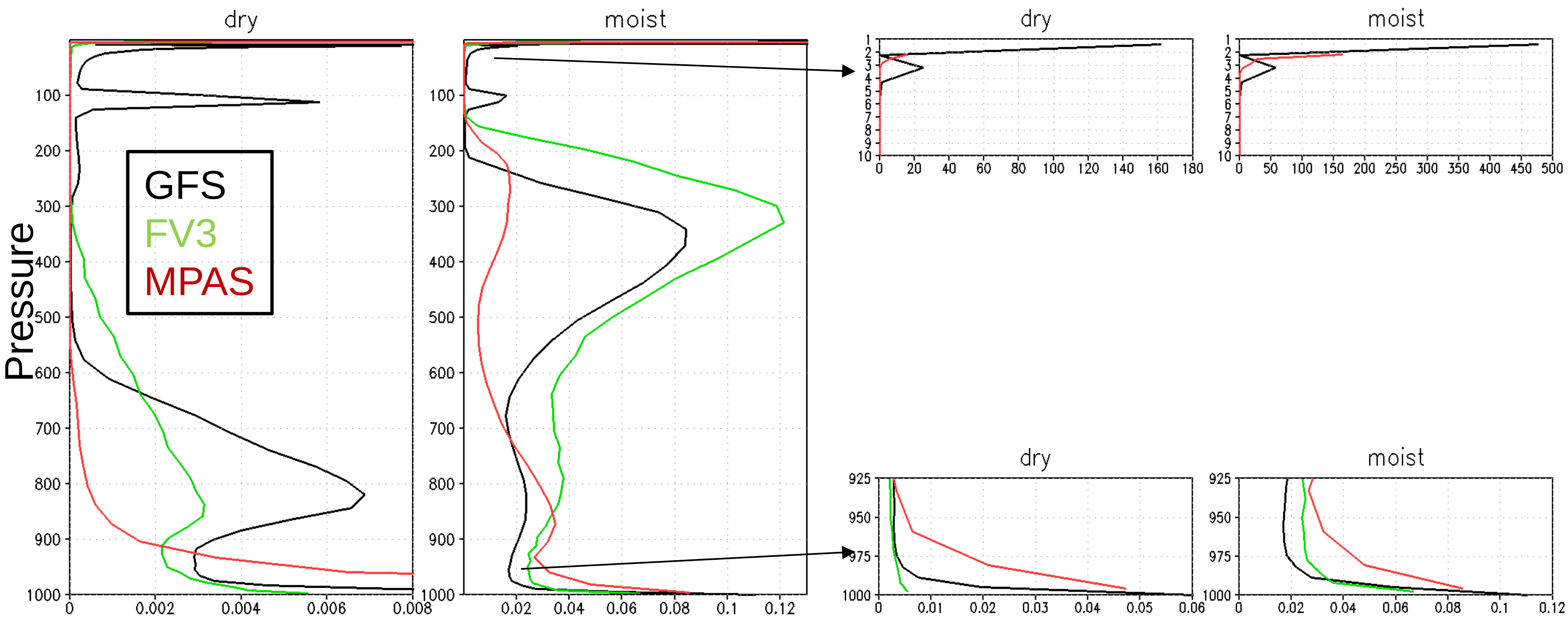
GFS Development and Operational Upgrade Plan



#1: Whole Atmosphere Model (WAM) Suitability

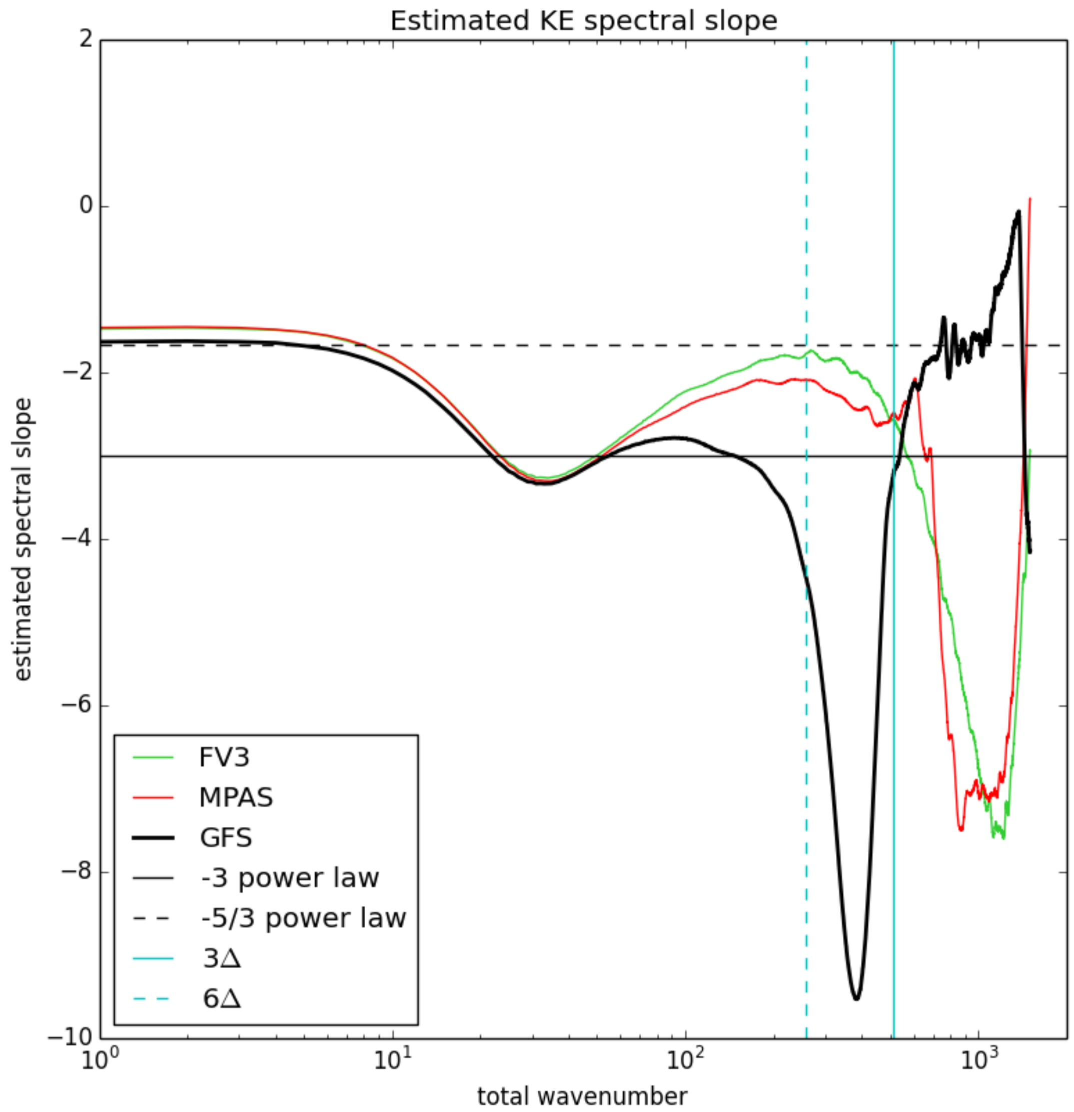
- SWPC WAM development team considered approaches by MPAS and FV3 to SWx requirements:
  - Both dycore teams have adequate plans in place for addressing SWx requirements for the next generation WAM and no preference was given to either dycore
  - Some requirements are not fully addressed by either dycore such as the approach to thermodynamics in a whole atmosphere
  - Significant effort still remains to adapt both dycores to the full atmosphere altitude/pressure domain currently covered by WAM

#2: Conservation Test: RMS Difference Between Advected Tracer and Dynamical Field (Day 15)



Global average RMS difference between prognostic equivalent potential temperature and tracer equivalent potential temperature calculated for each model level. Insets on right show detail at lower and upper levels of model, note that x-axes scales are much larger in insets.

#4: Performance Benchmark Results: Estimated Spectral Slope



#4: Performance Benchmark Results: Tracer advection performance

- Tracer advection benchmarks on Cori
  - Measure cost as a function of number of 3D tracer fields
    - Workloads and configuration:
      - 13 km case on number of cores needed for 8-8.5 min/day
    - Baseline: 3 tracer fields
    - Add 15 and 30 artificial tracers
  - Result: cost for full tracer load increased by factor of 2.5 for MPAS versus 1.53 for FV3 compared to baseline.

	Cores	Number of tracers / Minutes			Factor (lowest to highest)
MPAS	4800	3 / 8	18 / 14.6	33 / 19.8	2.5
FV3	1536	3 / 8.14	15 / 9.8	30 / 12.0	1.5 (1.53 adjusted)

Adjustment for FV3 workloads using 15 and 30 tracers ***total*** *instead of 15 and 30* ***additional*** *tracers per Test Plan.*

#4: Performance Benchmark Results: Refinement Efficiency

- Part of Criterion #5 evaluation
- How efficient is non-uniform at saving cost compared with uniform 3 km resolution on same number of processors?
- Benchmark and adjust for differences in resolution and area of refinement
- FV3's nesting scheme was more efficient than MPAS's in-place mesh refinement

Definition of nesting efficiency E:

a_g = area of domain (5.101e14 m^2)

a_h = area of refinement (FV3: 2.52e13 m^2 ; MPAS: 2.82e13 m^2)

$r = a_h / a_g$ fraction of domain at high resolution (for uniform res. Domain, r = 1)

dx_L = lowest resolution

dx_H = highest resolution

$C = r\,(dx_L / dx_H)^3\, C_{cellstep} + (1-r)\, C_{cellstep}$ *(C is "cost")*

$$S_{ideal} = \frac{(dx_L / dx_H)^3}{r\,(dx_L / dx_H)^3 + 1 - r} \quad \begin{matrix} \leftarrow C_{uniform} \\ \leftarrow C_{refined} \end{matrix}$$

(Note: $C_{cellstep}$ factors out)

$$S_{measured} = \frac{T_{uniform}}{T_{refined}}$$

$$E = S_{measured} / S_{ideal}$$

#4: Performance Benchmark Results: Refinement Efficiency (continued)

- Part of Criterion #5 evaluation
- How efficient is non-uniform at saving cost compared with uniform 3 km resolution on same number of processors?
- Benchmark and adjust for differences in resolution and area of refinement
- FV3's nesting scheme was more efficient than MPAS's in-place mesh refinement

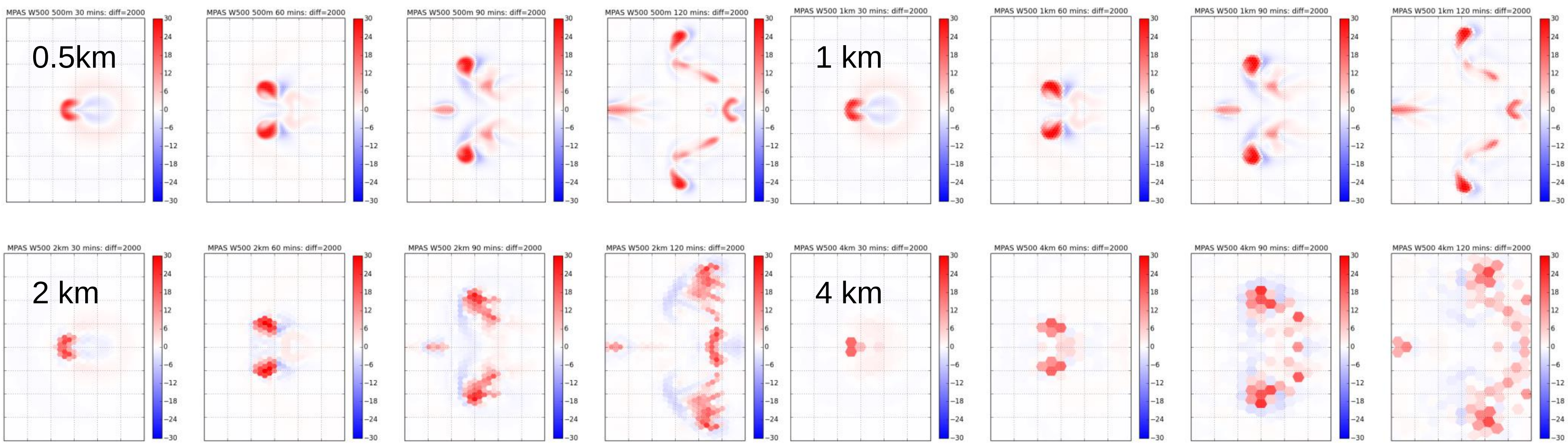
	FV3	MPAS
ag (global domain area m^2)	5.101E+14	5.101E+14
ah (high res area m^2)	2.52E+13	2.82E+13
percent of domain in high res r = ah/ag	4.94E-02	5.53E-02
dx low	14	15
dx high	3	3
dx l / dx h	4.67	5.00
(dx l / dx h) ^ 3	101.63	125.00
T-uniform (ideal)	101.63	125.00
T-reduced (ideal)	5.97	7.86
ideal speedup from refinement	17.02	15.91
T_uniform (measured)	345.93	344.65
T_refined (measured)	20.98	34.10
observed speedup from refinement	16.49	10.11
Efficiency	**96.9%**	**63.5%**

#5: Modifications to Phase 1 Supercell Test Case Configuration

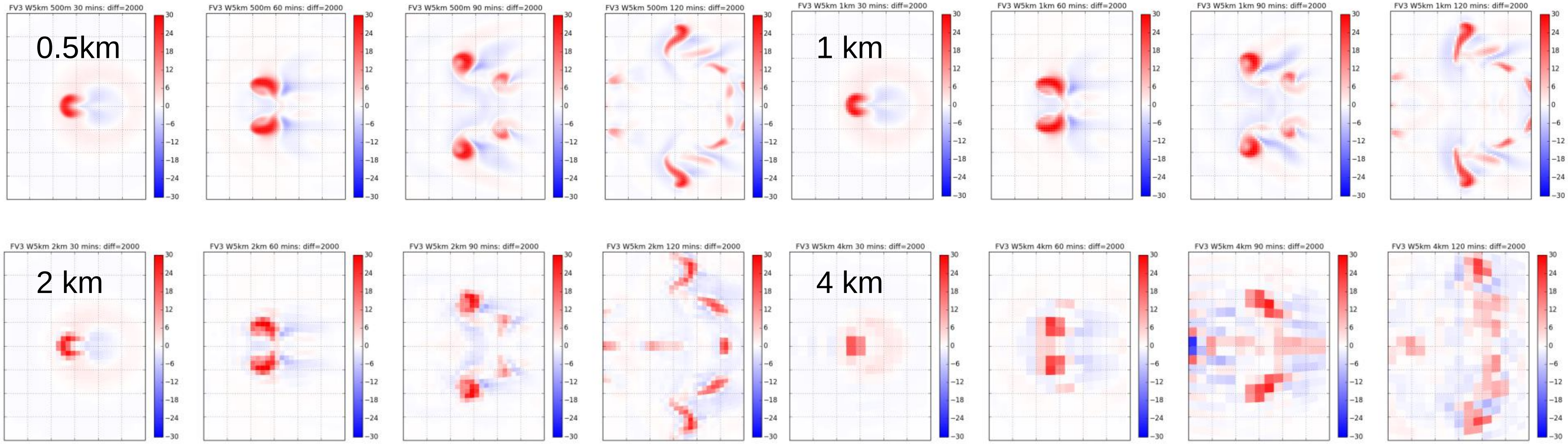
- MPAS
  - Disable vertical diffusion
  - Set Prandtl number to 1 (so that horizontal diffusion coefficient is same for all variables)
  - Physics timestep same as large RK step
  - Large RK step set to 3,6,12,24 seconds for 500m,1km,2km,4km resolutions
  - Number of acoustic timesteps per large RK step set to 6 in all cases
- FV3
  - Disable Smagorinsky diffusion by setting dddmp=0
  - Disable monotonic horizontal transport
  - Turn on 2nd order horizontal diffusion of tracers (using inline_q=.T. to ensure that tracers are integrated on the same time step as other prognostic variables)
  - Physics timestep set to 20,20,20,25 secs for 500m,1km,2km,4km resolutions
  - Number of vertical remaps per physics timestep (k_split) set to 8,5,2,1 for 500m,1km,2km,4km resolutions
  - Number of acoustic time steps per vertical remap (n_split) set to 5 in all cases

- With these mods, both models use constant 2nd order horizontal diffusion for all variables, no vertical diffusion. A horizontal diffusion coefficient of 2000 m^2/s is used, since it appears to produce a converged solution at 500 m for both models.

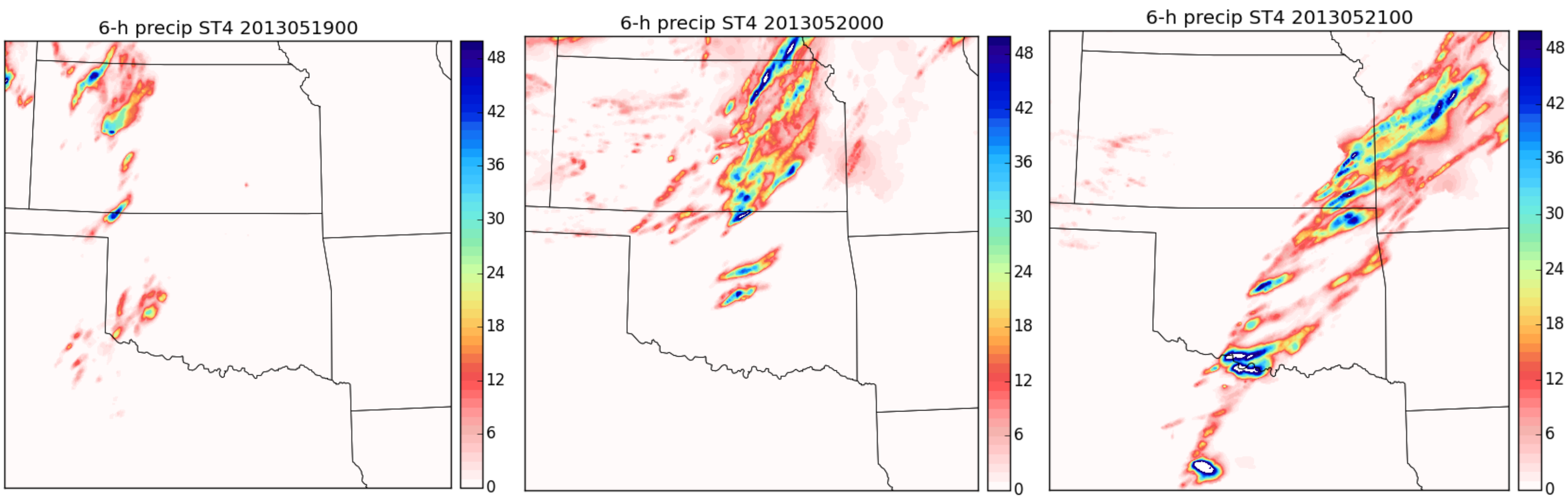
#5 Supercell Test: MPAS 500 hPa w



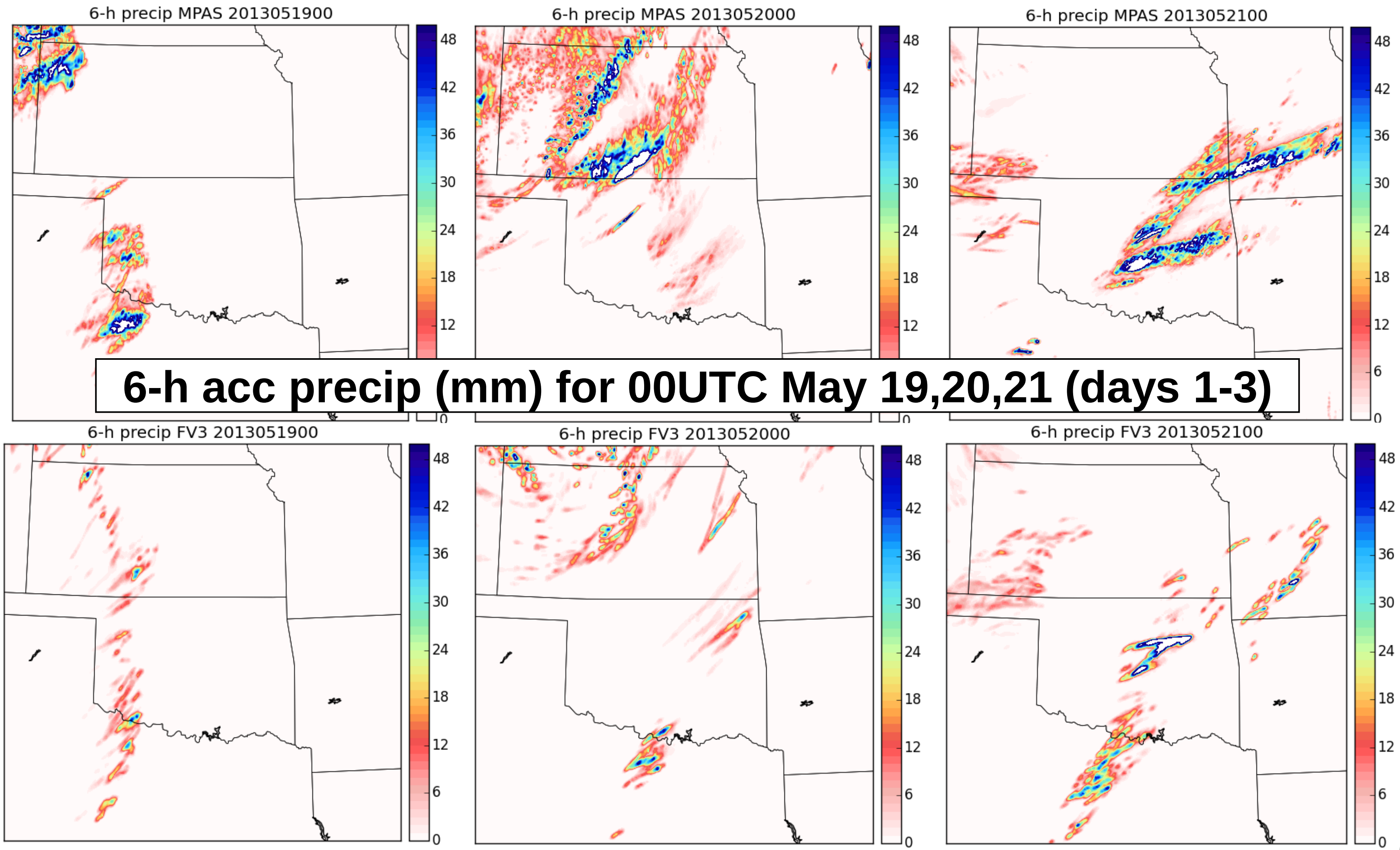
#5: Supercell Test: FV3 500 hPa w



#5: Moore Tornado Case: Stage IV Precipitation Analyses

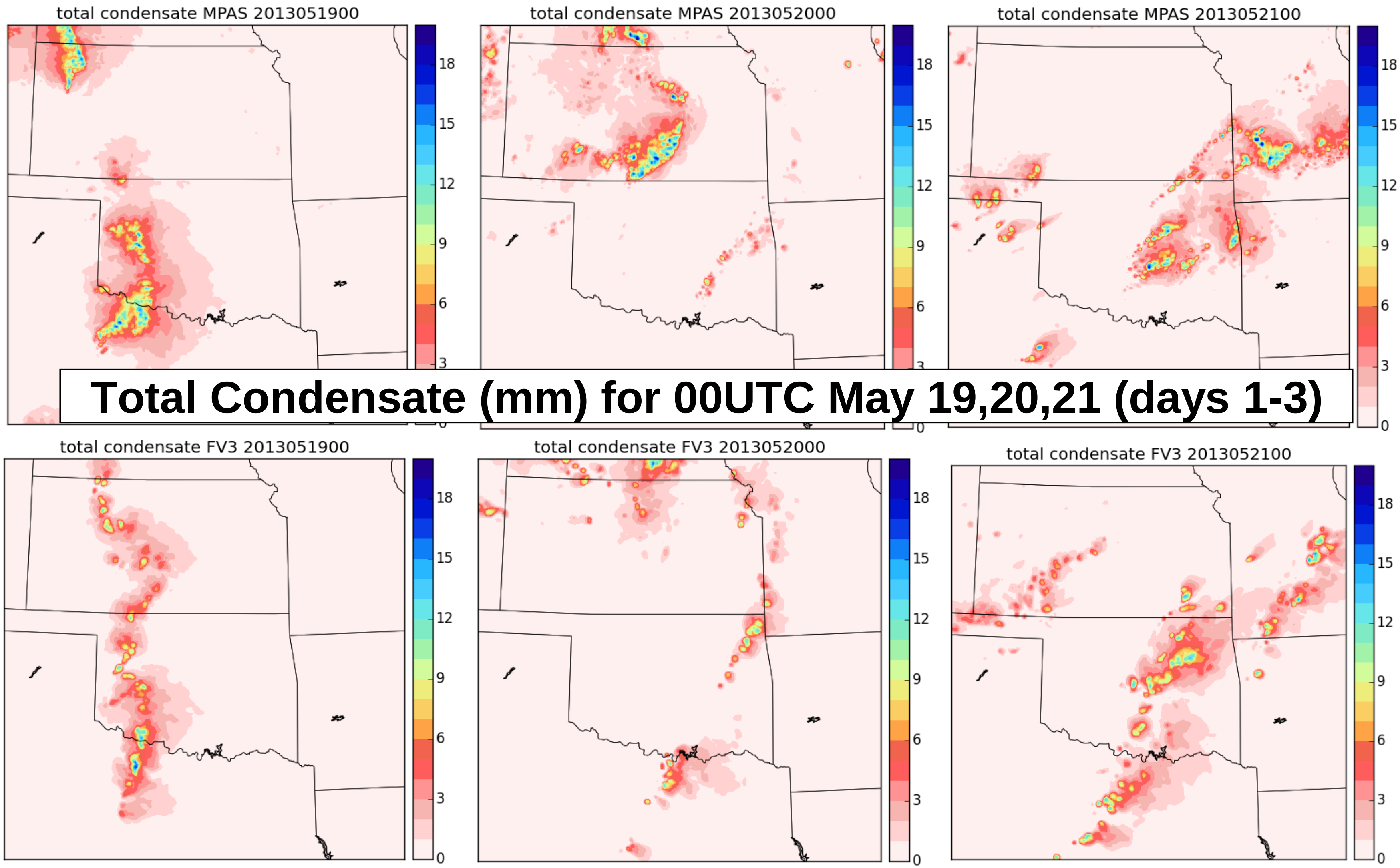


#5: Moore Tornado Case: Simulated Precipitation



6-h acc precip (mm) for 00UTC May 19,20,21 (days 1-3)

#5: Moore Tornado Case: Simulated Total Cloud Condensate



Total Condensate (mm) for 00UTC May 19,20,21 (days 1-3)

Criteria #7 and #8

- **#7: Code adaptable to NEMS/ESMF**
  - Self-reporting on questionnaire from EMC. GFDL completed (no issues) / NCAR incomplete
- **#8: Detailed dycore documentation**
  - Complete – Both dycores sufficiently documented for Phase 2 evaluation (but more will be needed for community model environment)

#10: Implementation Plan - Costs

Initial Implementation (transition to operations) Cost in FTEs (in addition to existing personnel managing O&M for operational GFS)

Activity	FY17		FY18		FY19		FY20		Total	
	MPAS	FV3	MPAS	FV3	MPAS	FV3	MPAS	FV3	MPAS	FV3
Dycore integration into NEMS	3	3	2	2	2	2	2	0	9	7
Physics implementation	2	1	2	1	1	1	1	0	6	3
Physics Driver implementation	1	1	2	1	1	1	1	0	5	3
DA integration	4	2	3	2	3	2	2	0	12	6
Pre/Post	2	2	2	2	1	1	1	0	6	5
Benchmarking	0	0	4	3	4	4	5	0	13	7
Code Management	2	2	2	2	2	2	2	2	8	8
Computational efficiency	2	1	2	1	2	1	2	0	8	3
Transition to operations	0	0	0	0	0	3	3	0	3	3
Total	**16**	**12**	**19**	**14**	**16**	**17**	**19**	**2**	**70**	**45**

Computer Resource Requirements for Initial Implementation (FY17-FY19 for FV3 and FY17-FY20 for MPAS)

	CPU*	CPU Hours**	Disk	Period	% change w.r.t. GFS
GFS	5,150,880	399,840	10 PB	FY17-FY18	0
FV3	6,565,620	509,660	30 PB (2 streams)	FY17-FY19	28%
MPAS	19,959,660	1,549,380	45 PB (3 streams)	FY17-FY20	288%

***CPU* = Y x 4 cycles x 365 days x 3 years, Y is number of cores required for 8.5 min/day**

Y = 1176 (GFS), 1499 (FV3), 4557 (MPAS) based on current operational resolution (~13 km). 1176 1499 4557

Computational requirements for intended implementation configuration TBD

****CPU hours = Y x 8.5 min/day x 10 days x 4 cycles**

HPC resources for Data Assimilation is not included

Availability of computational resources will require development/testing of FV3 in two parallel streams while MPAS would require three parallel streams

Summary	**Implementation Costs (Human Resources) for MPAS are 55% more compared to FV3** **Implementation Costs (computational resources) for MPAS are 204% more compared to FV3**