NGGPS dycore testing

Jeffrey Whitaker (NOAA/ESRL/PSD – test manager)
jeffrey.s.whitaker@noaa.gov

All the hard work done by…

Jin Lee, Gerard Ketefian, Ka Yee Wong and Man Zhang (NIM: ESRL)
Zavisa Janjic, Ratko Vasic, Tom Black, Miodrag Rancic and Vladimir Djurdjevic (NMM-B and NMM-UJ: NCEP)
Shian-Jiann Lin, Lucas Harris, Rusty Benson and Jan-Huey Chen (FV3: GFDL)
James Doyle, Sasa Gabersek, Alex Reinecke and Kevin Viner (NEPTUNE: NRL)
William Skamarock, Joseph Klemp and Sang-Hun Park (MPAS: NCAR)
Phil Pegion (analysis)
John Michalakes, Paul Madden, Mark Govett, Tom Henderson (benchmarking and performance testing)

Phase 1 testing

Status	Activities
Complete	HIWPP Idealized Tests
Complete	Computational performance and scalability testing and software evaluation by Advanced Computing Evaluation Committee (AVEC)
Complete	HIWPP 3-km, 3-Day Simulations
Complete	Phase 1 Testing Report
Complete	Dycore Test Group (DTG) assessment of Phase 1 testing results
Complete	Phase 1 testing results briefing to NCEP and NWS directors

NGGPS Phase 1 Dycore Test
Candidate Model Dynamic Cores

- FV3 (GFDL): Cubed-sphere finite-volume with flexible Lagrangian vertical coordinate (z or p base) with nesting or stretched grid capability
- MPAS (NCAR): Finite-volume C-grid staggering, icosahedral (z coordinate) with unstructured mesh refinement capability.
- NIM (ESRL): Icosahedral unstaggered A-grid mesh, finite-volume (z coordinate)
- NMM-UJ (EMC): Finite-difference, cubed-sphere version of Non-hydrostatic Mesoscale Model (p coordinate); Uniform Jacobian cubed sphere grid replaced lat/lon grid version with staggered B-grid (NMMB)
- NEPTUNE (Navy): Spectral-element (horizontal and vertical) cubed-sphere grid (z coordinate) with adaptive mesh refinement

Global Spectral Model not included – Non-hydrostatic version not available

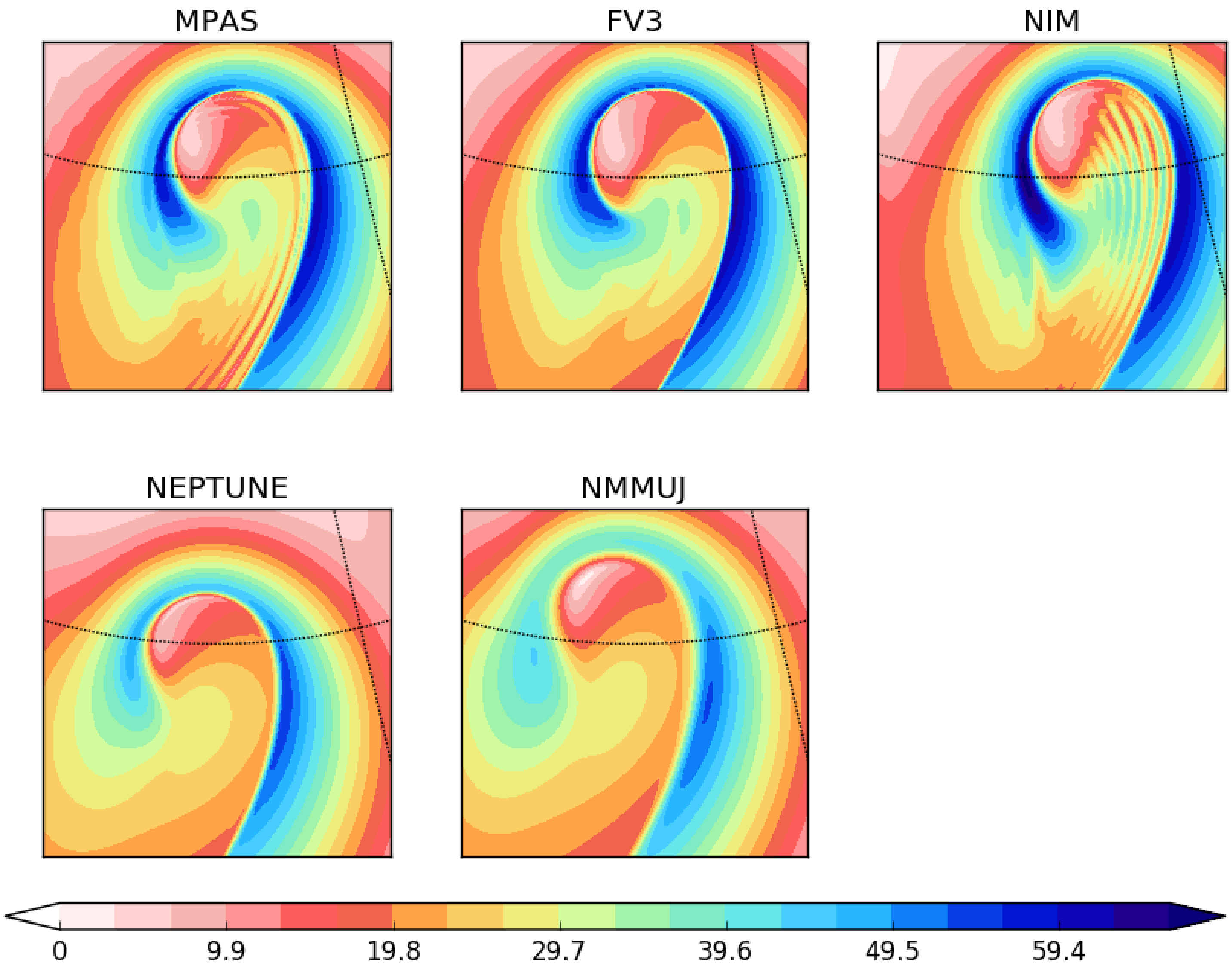
Phase 1 Dycore Testing Overview

Evaluation Criteria	How evaluation was done
Bit reproducibility for restart under identical conditions	Query model developers (AVEC)
Solution realism for dry adiabatic flows and simple moist convection	Perform series of idealized tests and evaluate solutions
High computational performance and scalability	Benchmarks run by AVEC
Extensible, well-documented software that is performance portable	Subjective evaluation of source code by AVEC
Execution and stability at high horizontal resolution (3 km or less) with realistic physics and orography	72-h forecasts with realistic physics and orography using operational GFS initial conditions (Moore tornado and Hurricane Sandy)
Lack of excessive grid imprinting	Evaluate idealized test case solutions

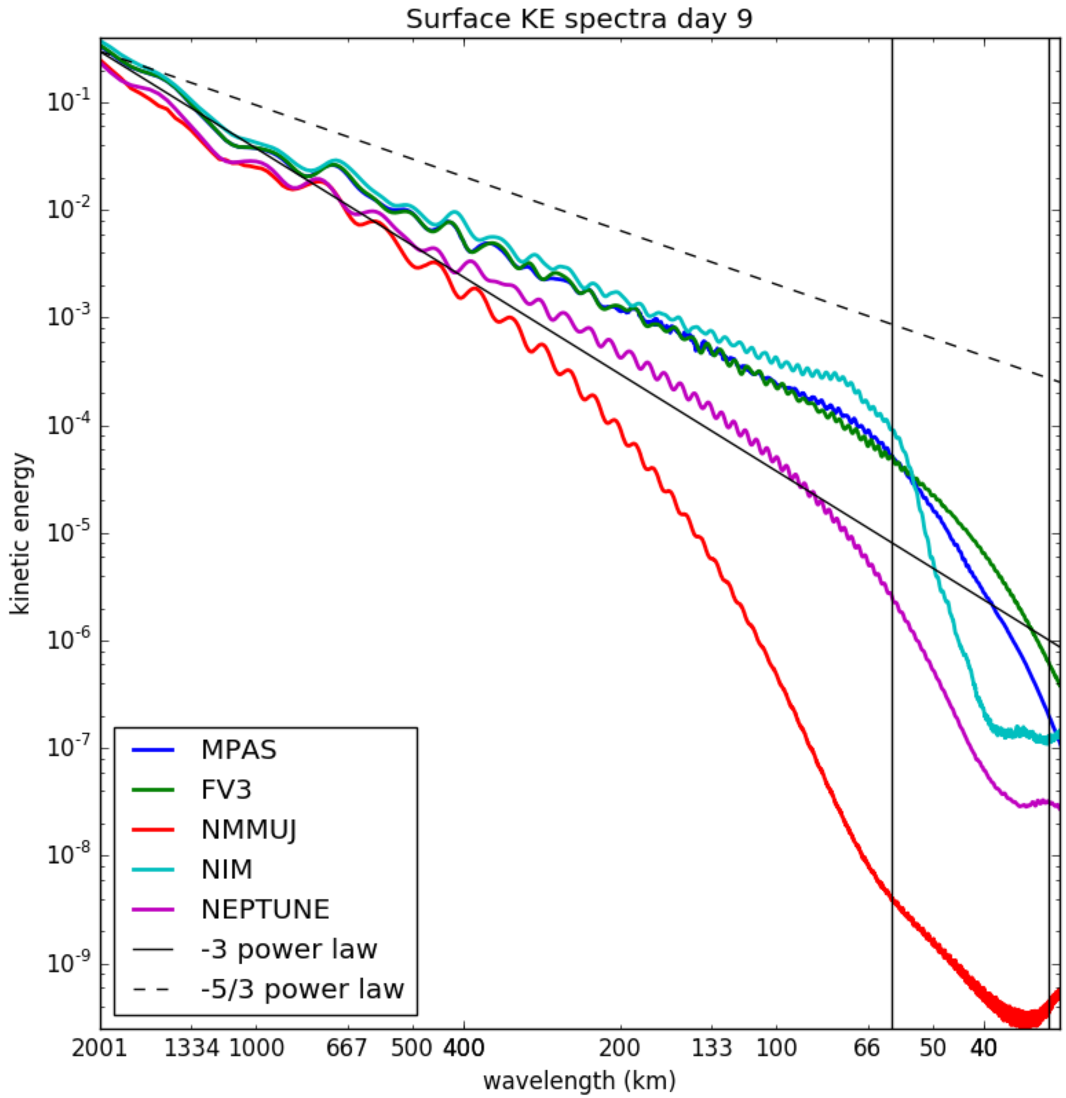
Idealized Tests

- **Baroclinic wave test with embedded fron*ts*** (DCMIP 4.1)
  - Dynamics strongly forces solution to shortest resolvable scales
  - Shows impact of truncation error near quasi-singular points on computational grid (“grid imprinting”)
  - 15/30/60/120 km horizontal resolutions with 30 and 60 vertical levels
- **Non-hydrostatic mountain waves on a reduced-radius sphere** (like DCMIP 2.1/2.2)
  - Shows ability to simulate non-hydrostatic gravity waves excited by flow over orography
  - 3 tests: M1 (uniform flow over a ridge-like mountain), M2 (uniform flow over circular mountain), M3 (vertically sheared flow over a circular mountain). Solutions are all quasi-linear
- **Idealized supercell thunderstorm on a reduced-radius sphere**
  - Convection is initiated with a warm bubble in a convectively unstable sounding in vertical shear
  - Simple Kessler warm-rain microphysics, free-slip lower boundary (no boundary layer)
  - Splitting supercell storms result after 1-2 hours of integration
  - 0.5/1/2/4 km horizontal resolutions

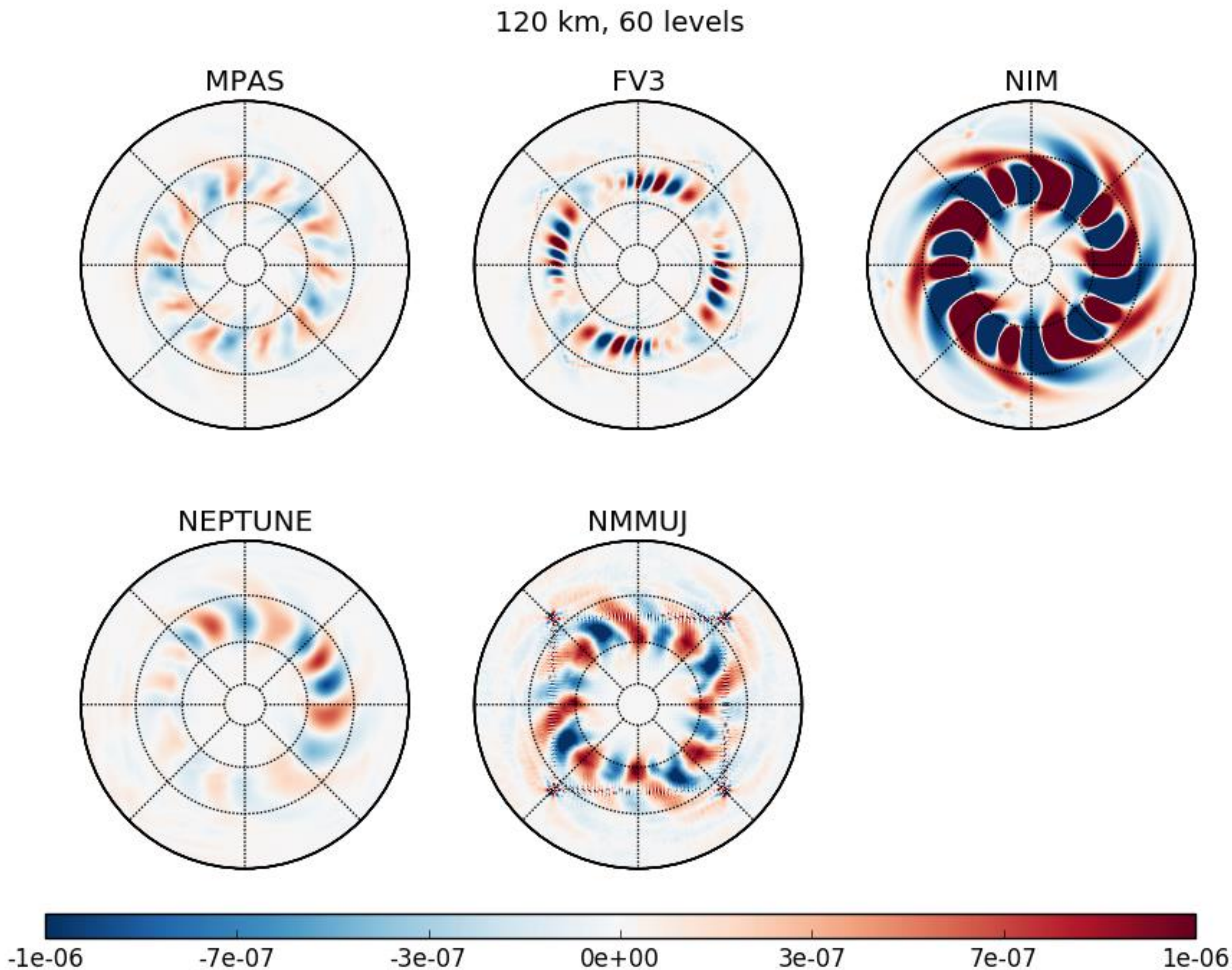
Baroclinic Wave (Sfc Wind Speed at Day 9, 15-km resolution)



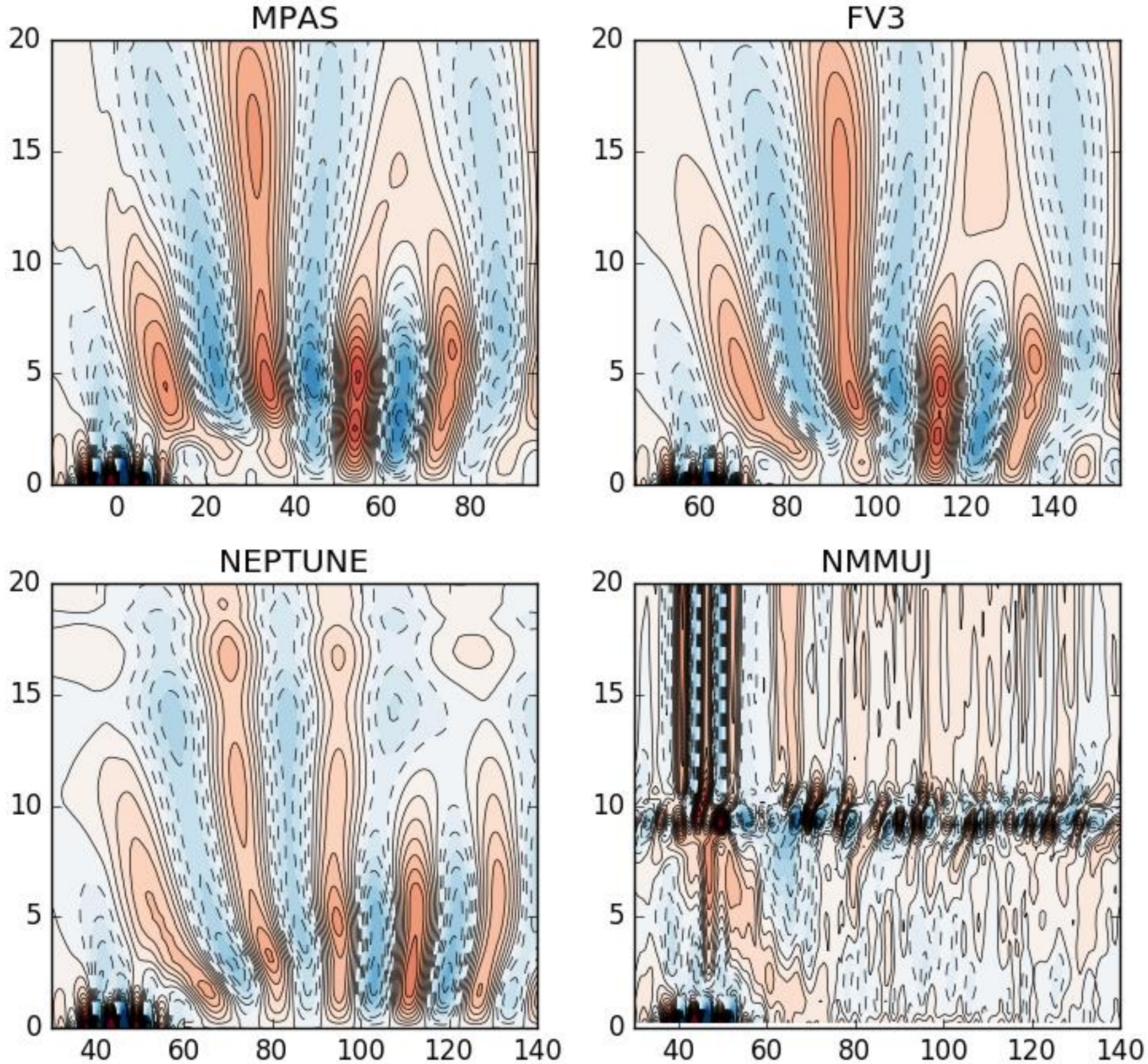
Baroclinic Wave KE Spectrum (surface, day 9, 15-km resolution)



Baroclinic Wave 120-km, S. Hem. 850 vorticity (grid imprinting)

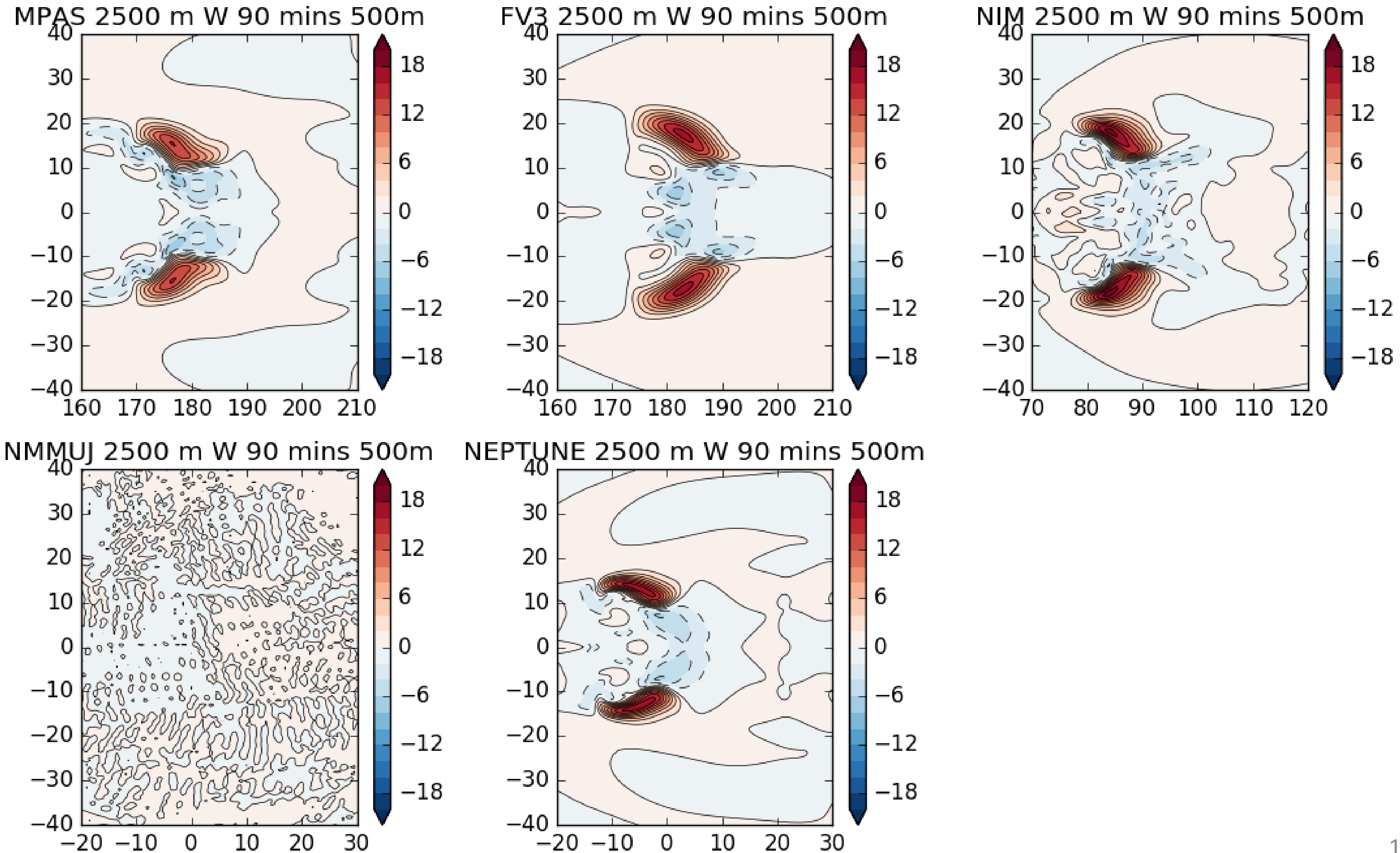


Mtn wave in shear (w cross section at equator, 1-km resolution)

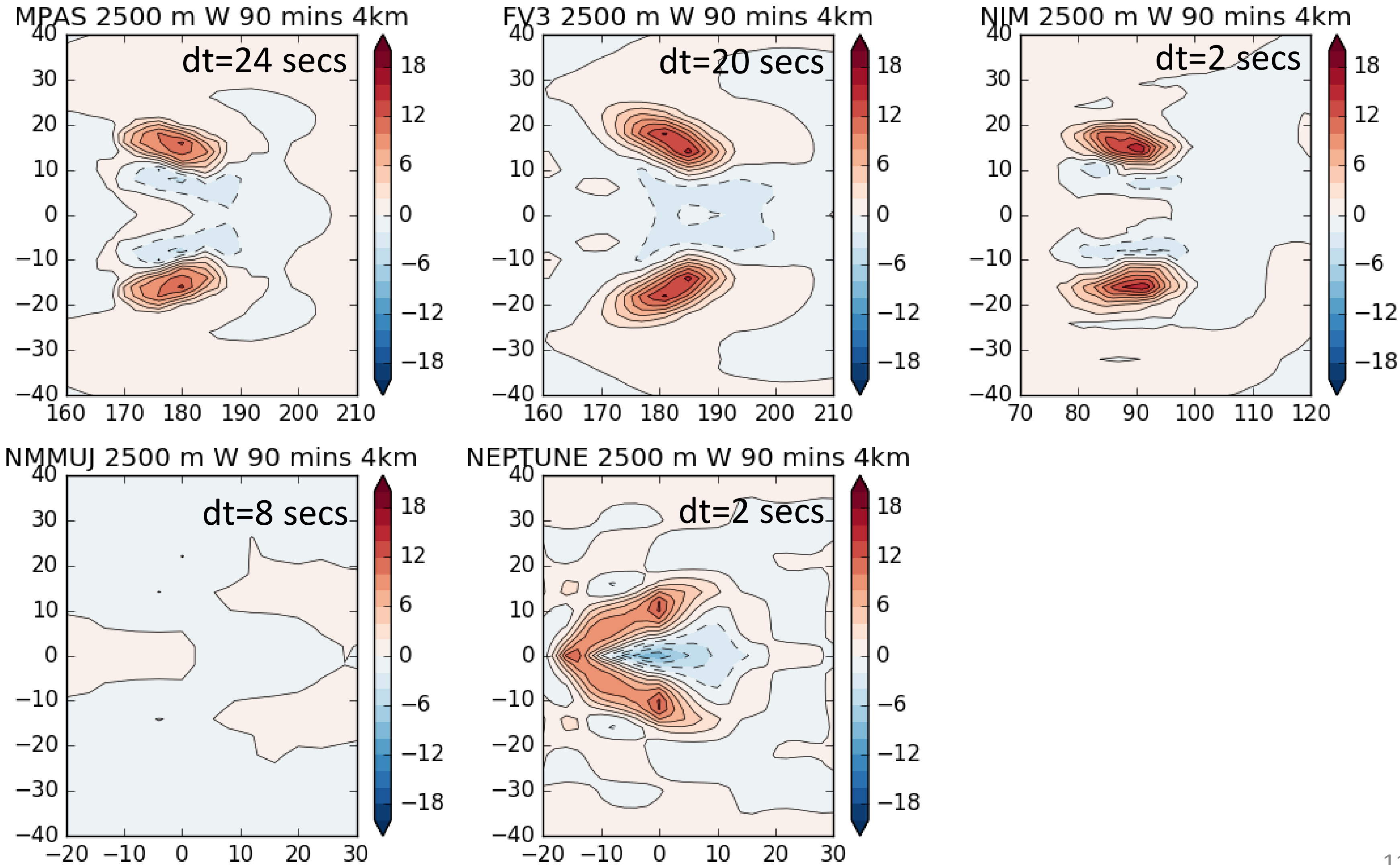


- NEPTUNE differences likely due to deep atmosphere equation set (shallow atmosphere approx matters on reduced radius sphere).
- NMMUJ produced unrealistic solutions for all mountain wave tests.

Supercell (2500-m w at 90 mins, 500-m resolution)



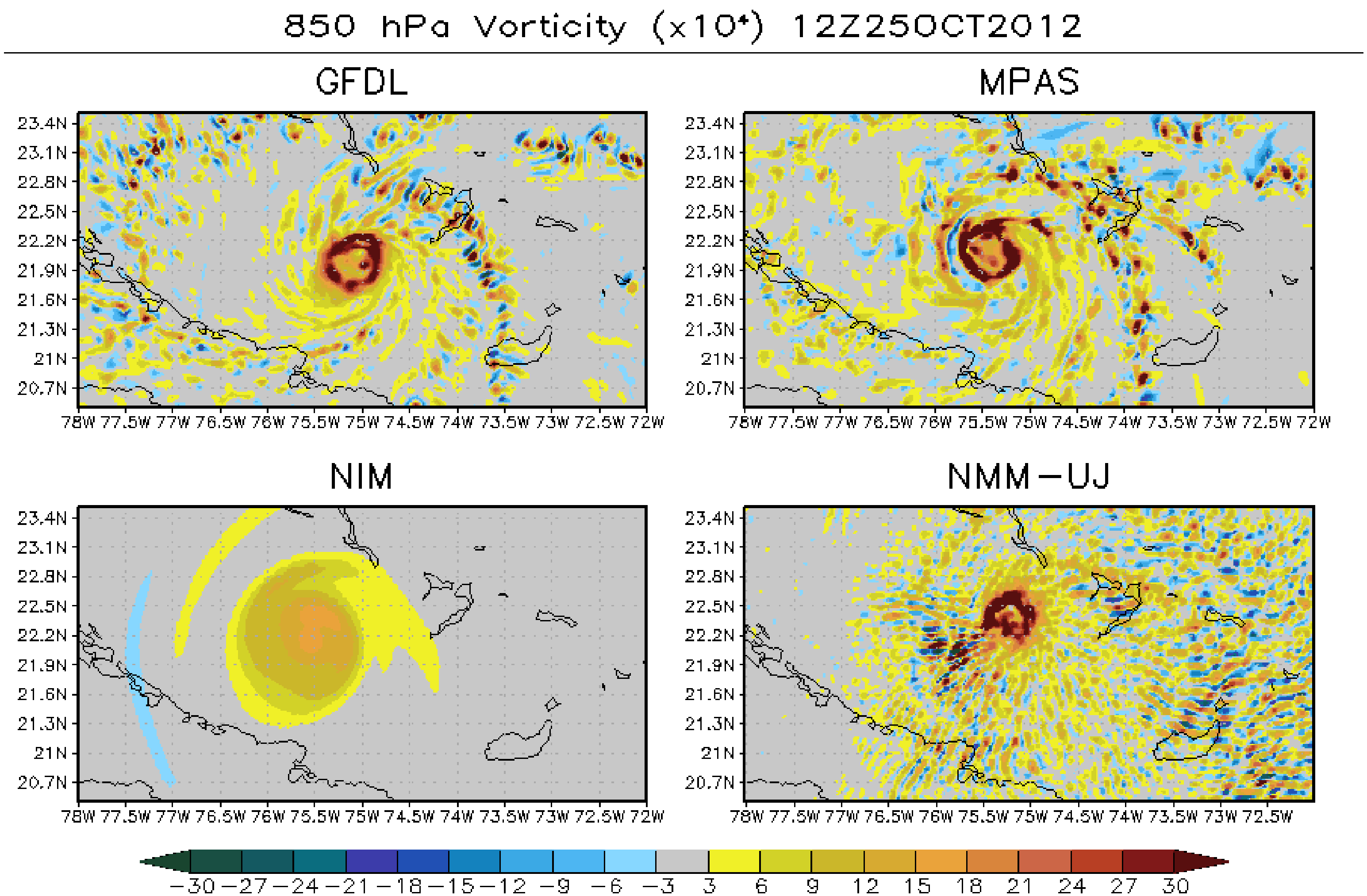
Supercell (2500-m w at 90 mins, 4-km resolution)



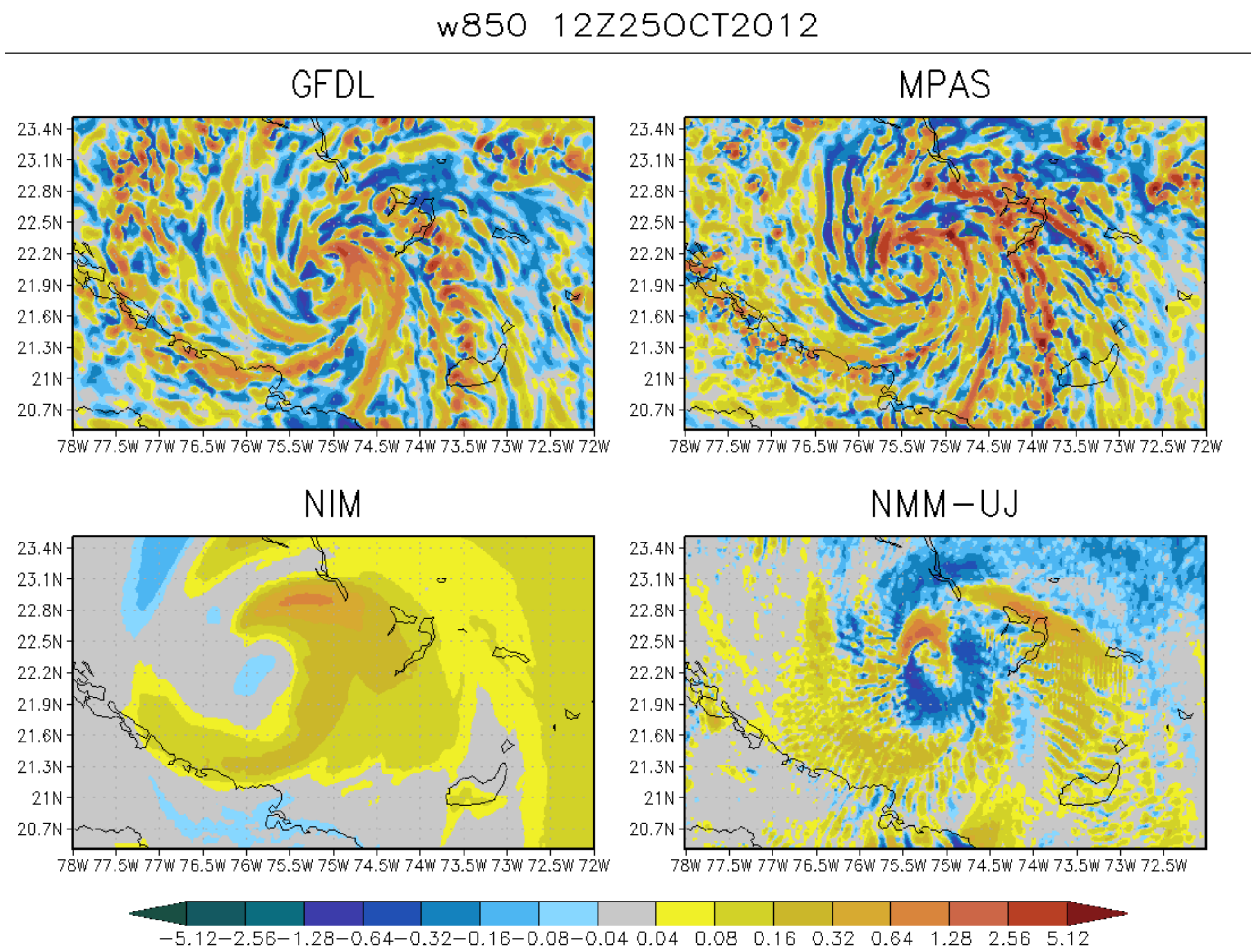
72-h 3-km Forecast Test

- ‘Stress-test’ dycores by running with full-physics, high-resolution orography, initial conditions from operational NWP system
  - Different physics suites used in each model
- Two cases chosen:
  - Hurricane Sandy 2012102418 (also includes WPAC typhoon)
  - Great Plains tornado outbreak (3-day period beginning 2013051800). Includes Moore OK EF5 tornado around 00UTC May 19
- Focus not on forecast skill, but on ability of dycores to run stably and produce reasonable detail in tropical cyclones and severe convection
  - Also look at global quantities like KE spectra, total integrated precipitation/water vapor/dry mass

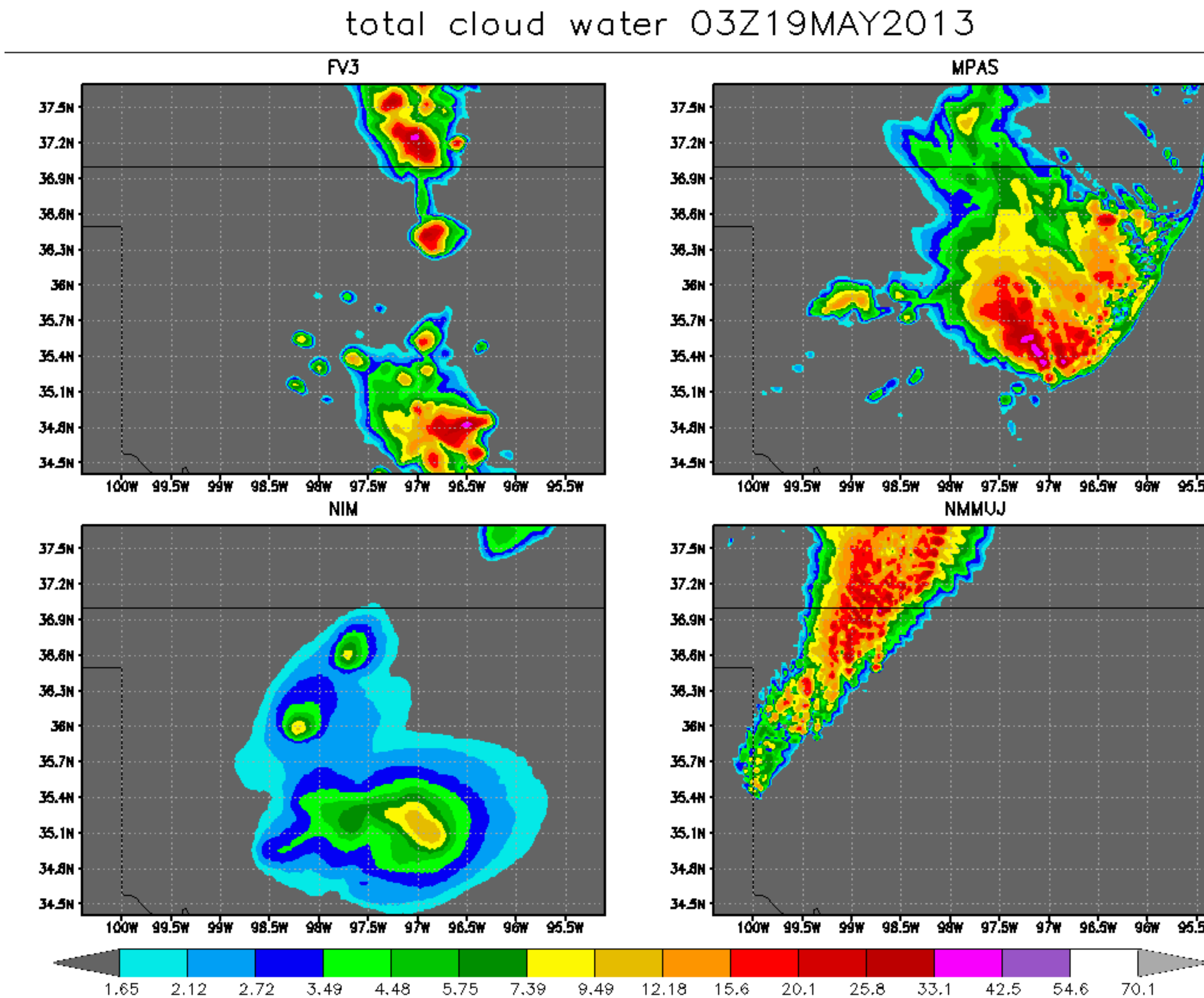
Hurricane Sandy (vorticity at 850 hPa)



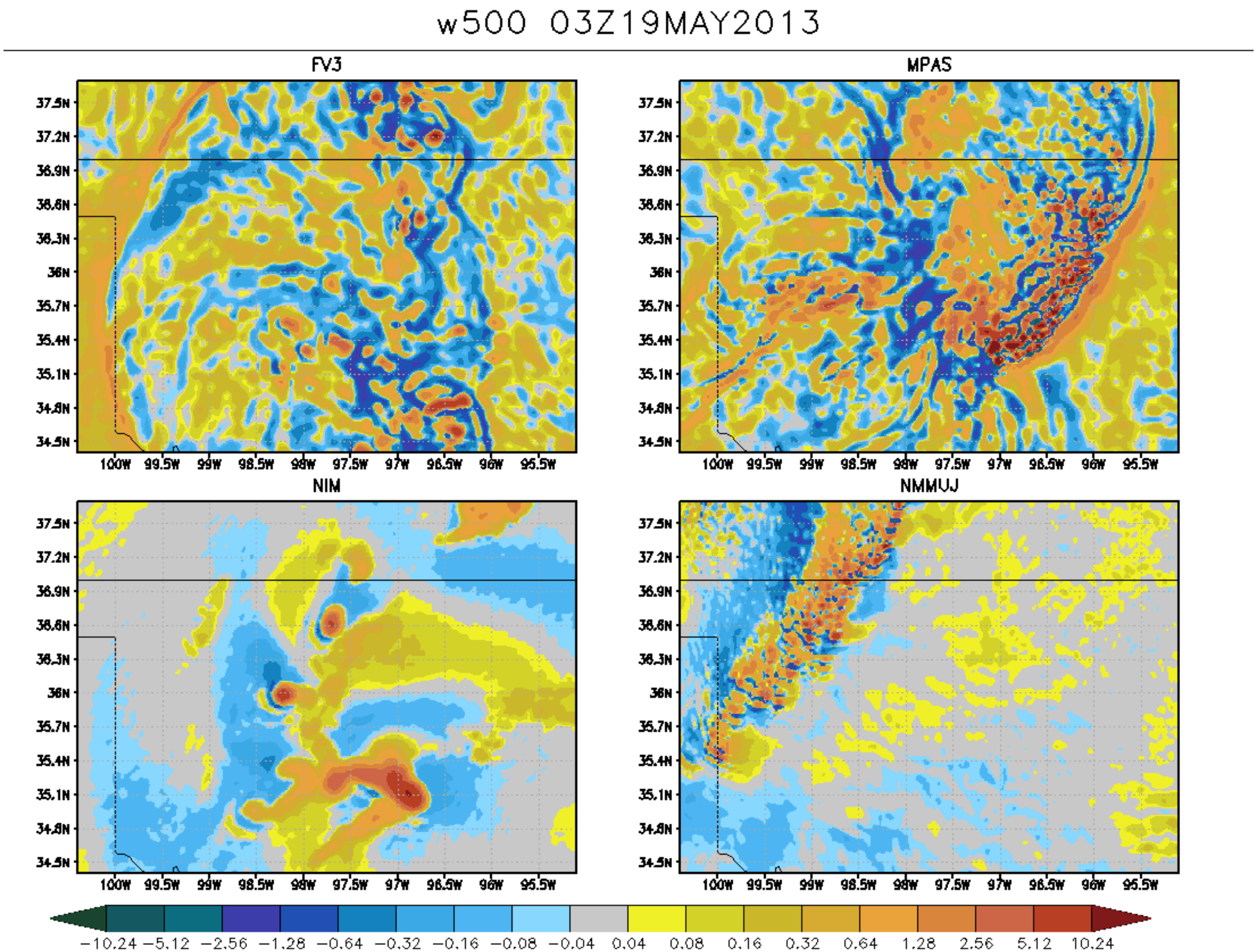
Hurricane Sandy (w at 850 hPa)



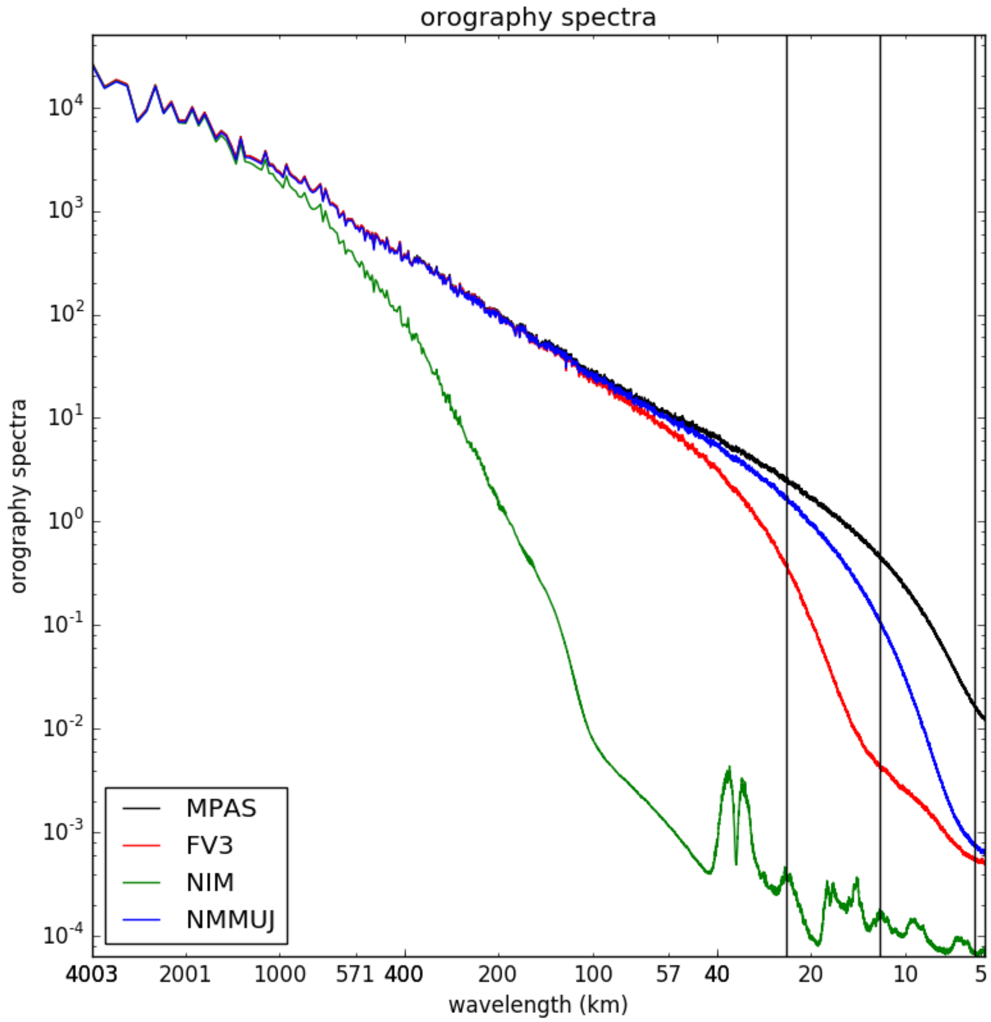
Moore Tornado (total condensate)



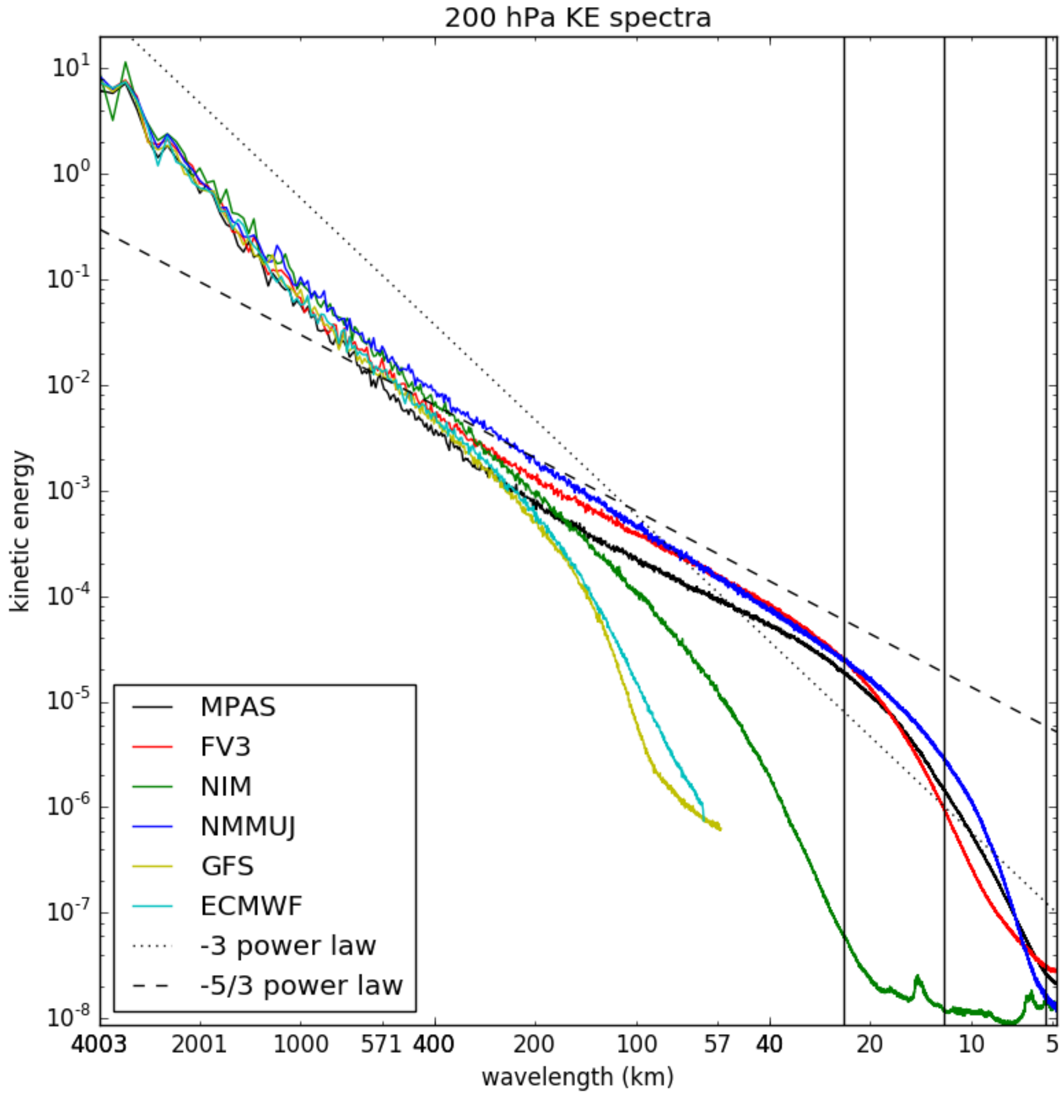
Moore Tornado (w at 500 hPa)



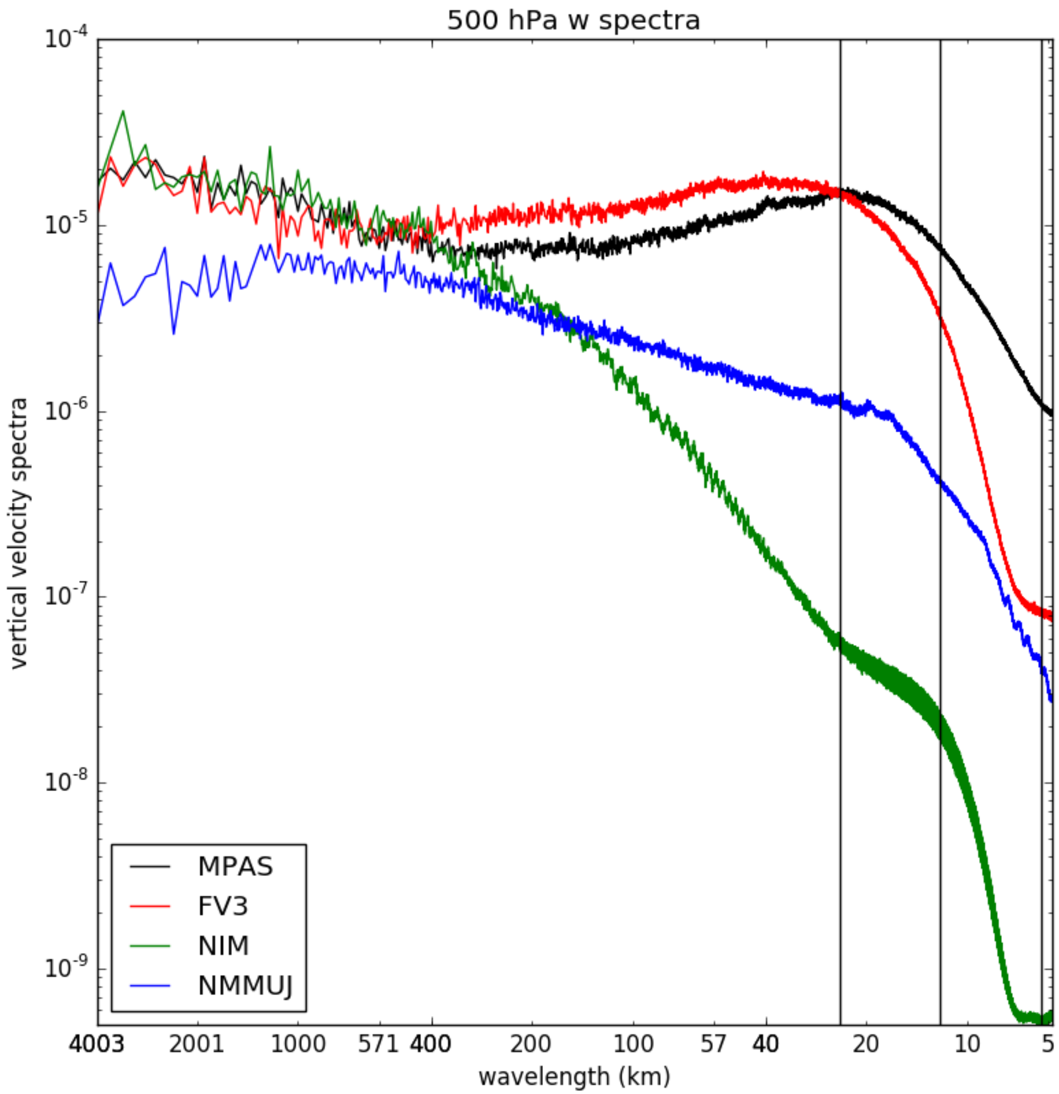
Orography spectra



200 hPa KE Spectrum (3-km 72-h forecasts, Hurricane Sandy)



500 hPa w Spectrum (3-km 72-h forecasts, Moore Tornado)



Advanced Computing Evaluation Committee

AVEC formed August 2014 to evaluate and report on performance, scalability and software readiness of five NGGPS candidate dycores

Model	Organization	Numeric Method	Grid
NIM	NOAA/ESRL	Finite Volume	Icosahedral
MPAS	NCAR/LANL	Finite Volume	Icosahedral/Unstructured
NEPTUNE	Navy/NRL	Spectral Element	Cubed-Sphere with AMR
HIRAM/FV-3	NOAA/GFDL	Finite Volume	Cubed-Sphere, nested
NMMB	NOAA/EMC	Finite difference/Polar Filters	Cartesian, Lat-Lon
~~GFS-NH *~~	~~NOAA/EMC~~	~~Semi-Lagrangian/Spectral~~	~~Reduced Gaussian~~
IFS (RAPS13)**	ECMWF	Semi-Lagrangian/Spectral	Reduced Gaussian

* Current operational baseline, non-hydrostatic option under development, No version of GFS was available for AVEC tests

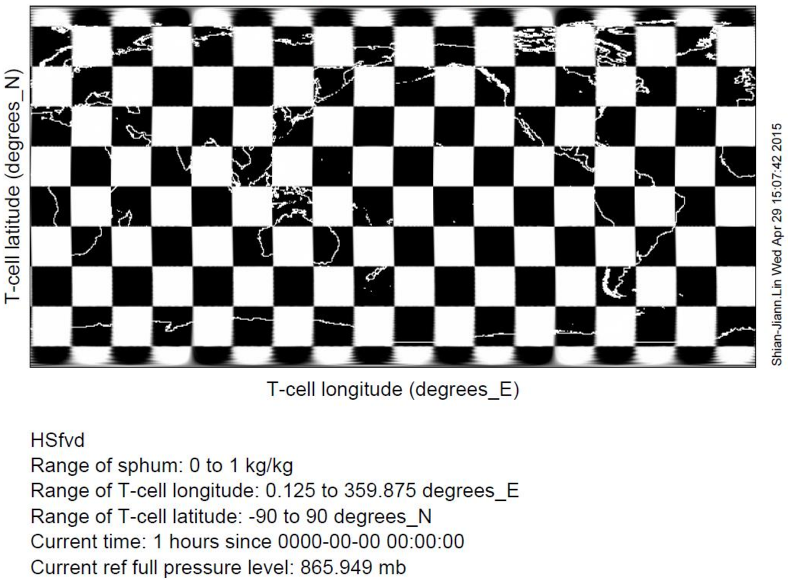
** Guest dycore, <u>hydrostatic</u>, GFS proxy

Caveats

- The performance and scaling results in this report are a snapshot in time of NWP software that is under active development. The test workloads are based on an idealized atmospheric case that does not include physics.
- The choice of time step for the idealized benchmark runs was best-guess of what would be needed for full-physics real-data forecasts on the part of the modeling groups. In adjusting benchmarking results to the operational speed requirement, we also assumed that dynamics represents half the run time of a full-physics model.
- Benchmarks were compute-only. AVEC did not evaluate important aspects of performance such as I/O, initialization costs, or other factors that would not represent full physics realizations of the models. Such testing will occur in future Level-2 evaluations under the NGGPS test plan.
- AVEC evaluated model performance with no consideration for solution quality. Each candidate model's benchmarks were conducted with the same formulation and configuration used to run the test cases just presented.

Workloads

- Baroclinic wave case from HIWPP non-hydrostatic dycore testing (DCMIP 4.1)
  - Added 10 artificial 3D tracer fields to simulate cost of advection
  - Initialized to checkerboard pattern to trigger cost of monotonic limiters
- 13 km workload
  - Represent current and near-term global NWP domains
  - Measure performance of the code with respect to operational time-to-solution requirement (8.5 minutes/forecast day)
- 3 km workload
  - Represent workloads that might be in operations within lifetime of NGGPS
  - Measure ability to scale to efficiently utilize many times greater computational resources



Checkerboard tracer initialization pattern after one hour FV3 integration. Image provided by S. J. Lin, NOAA/GFDL

Benchmark Configurations

		NH-GFS (Baseline) *	FV3	MPAS	NIM	NMMB-UJ	NEPTUNE	IFS (RAPS13) *
Nominally 13km	**Resolution**	13 km (TL1534)	~12 km (C768)*	12km *	13.4 *	13 km	12.71 km *	12.5 km (Tc799)
	Grid Points	3072x1536 (unreduced) 3,126,128 (reduced)	6x768x768 3,538,944	4,096,002 **	3,317,762	6x768x768 3,538,944 *	3,110,402 **	3,336,946 (reduced)
	Vertical Layers *	128	127 **	127 ***	128	128	127 ***	137
	Time Step	TBD	600s (slow phys) 150s (vertical, fast phys) 150/10 (horiz. acoustic)	72 s (RK3 dynamics) 12 s (acoustic) 72 s (RK3 scalar transport)	72 s	24 s **	75 s (advective), 15 s (sound) ****	450
Nominally 3km	**Resolution**	3 km (TL6718)	~3 km (C3072) *	3km	3.3 km **	3 km	3.13 km *	3.125 km (Tc3199)
	Grid Points	13440x6720 (unred.) 59,609,088 (reduced) **	6x3072x3072 56,623,104	65,536,002	53,084,162	6x3072x3072 56,623,104 *	61,440,000 **	51,572,436 (reduced)
	Vertical Layers *	128	127 **	127 ***	128	128	128	137
	Time Step	TBD	150 s (slow phys) 37.5 s (vertical, fast phys) 37.5/10 s (horiz. acoustic)	18 s (RK3 dynamics) 3 s (acoustic) 18 s (RK3 scalar transport)	18 s	6 s **	15 s (slow RK3 dyn.) 2.5 s (fast dyn.)	120

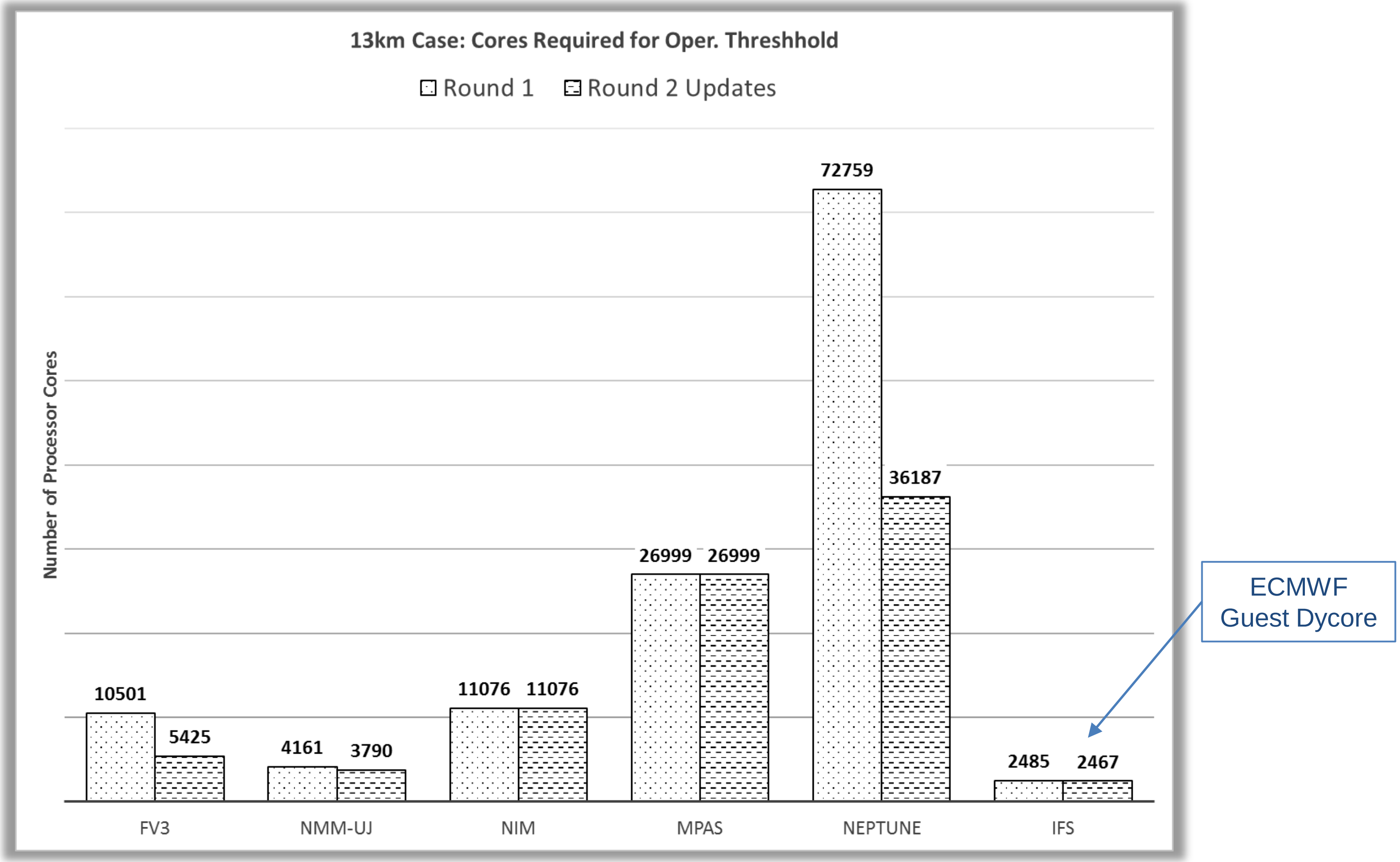
Table A3-1. Model-specific Benchmark Configurations

Computational Resources

- Edison: National Energy Research Scientific Computing Center (DOE/NERSC)
  - 4M core hours in two sessions totaling 12 hours of dedicated machine access
  - 133,824 cores in 5,576 dual Intel Xeon Ivy Bridge nodes (24 cores per node)
  - Cray Aries with Dragonfly network topology
  - https://www.nersc.gov/users/computational-systems/edison/configuration
- Stampede: Texas Advanced Computing Center
- Pleiades: NASA Ames Research Center

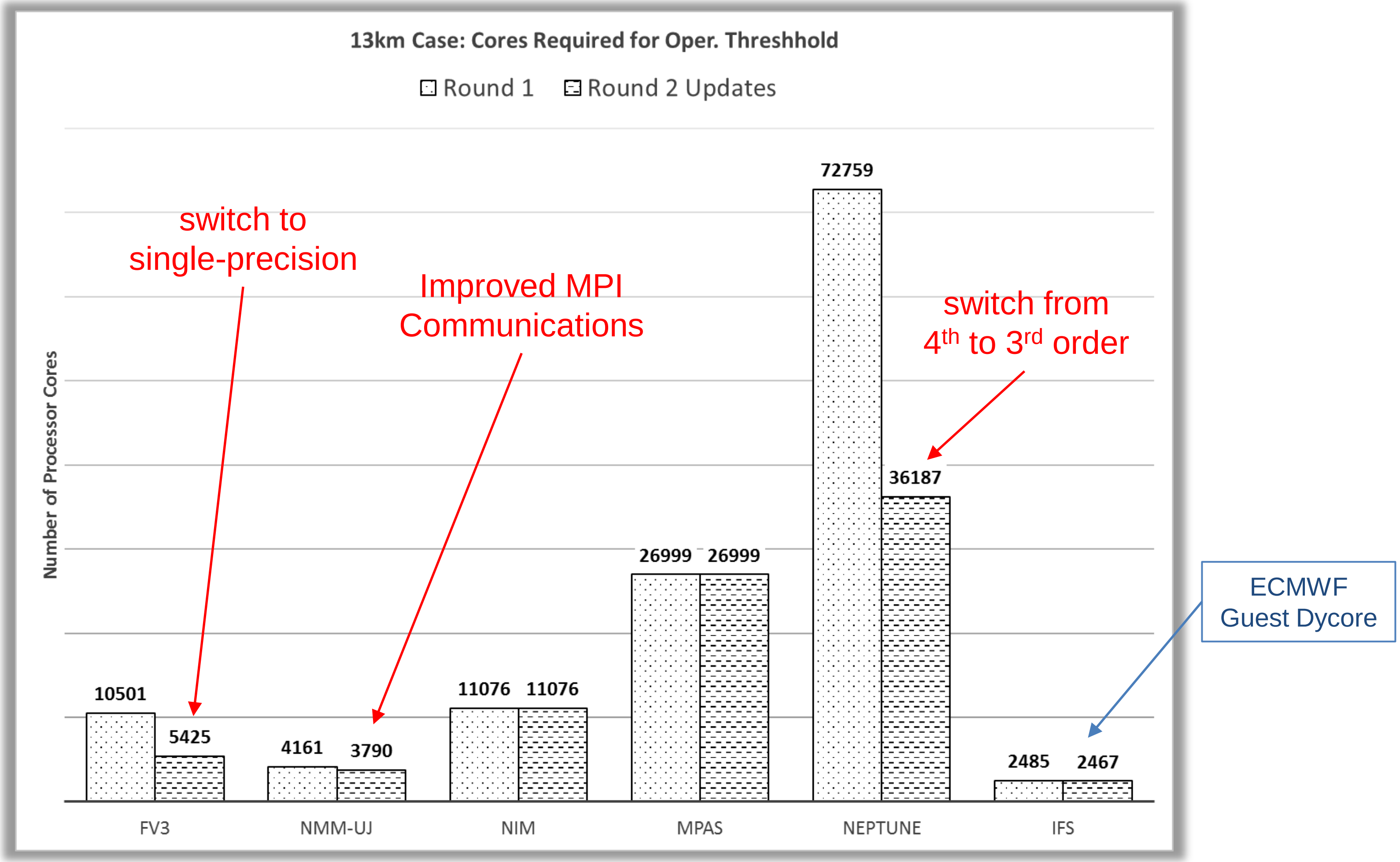
AVEC Level-1 Evaluations: Performance

- Performance:
  - Number of processor cores needed to meet operational speed requirement with 13-km workload
  - Rankings (fastest to slowest): NMM-UJ, FV3, NIM, MPAS, NEPTUNE



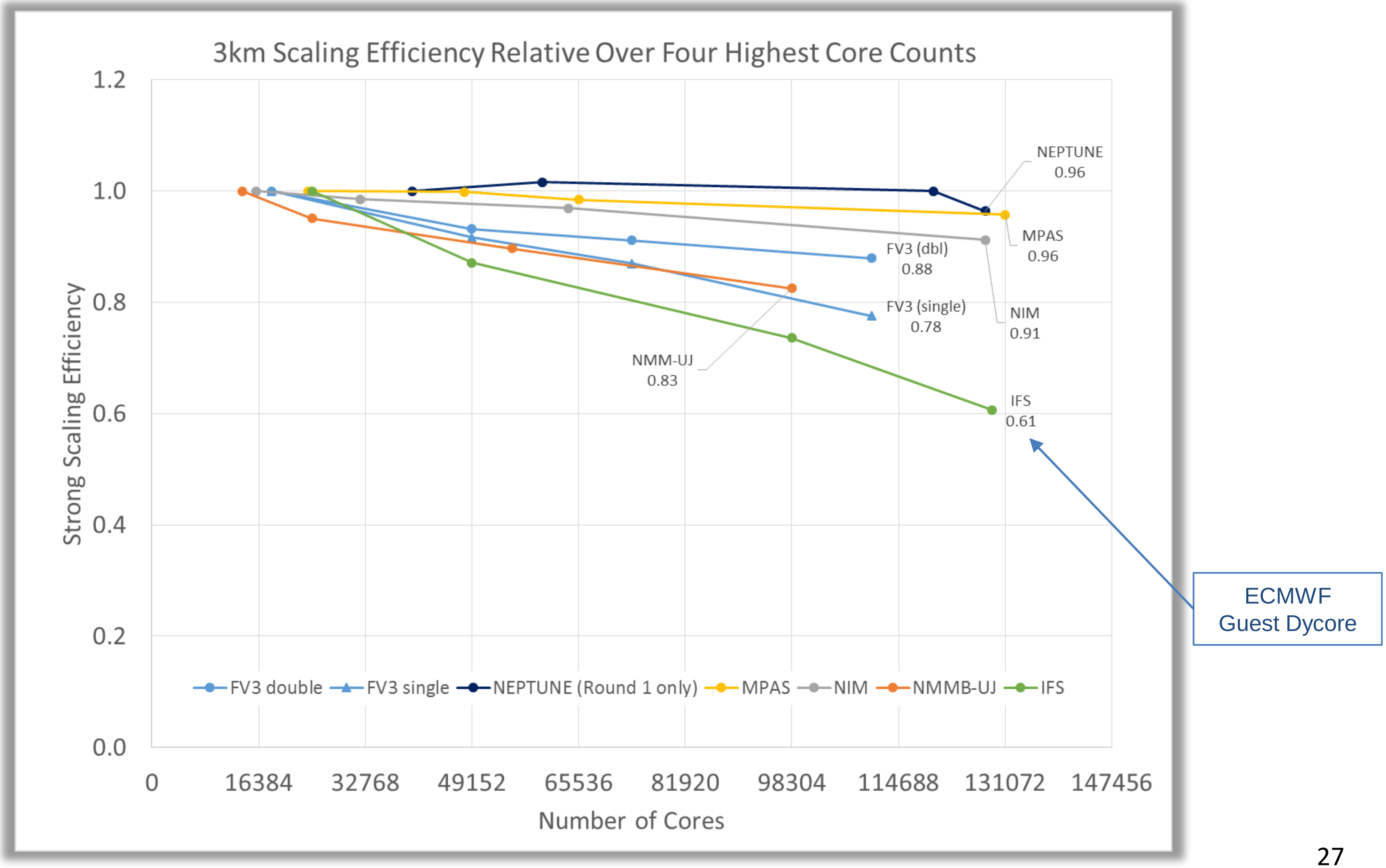
AVEC Level-1 Evaluations: Performance

- Performance:
  - Number of processor cores needed to meet operational speed requirement with 13-km workload
  - Rankings (fastest to slowest): NMM-UJ, FV3, NIM, MPAS, NEPTUNE



AVEC Level-1 Evaluations: Scalability

- Scalability: ability to efficiently use large numbers of processor cores
  - IFS excepted, all codes showed good scaling (no global comms).
  - Rankings (most to least scalable): NEPTUNE, MPAS, NIM, FV3, NMM-UJ



NGGPS Phase 1 Testing
Project Summary Assessment

	Idealized Tests	3-km, 3-day forecasts	Performance	Scalability	Nesting or Mesh Refinement	Software Maturity
FV3	Green	Green	Green	Green	Green	Green
MPAS	Green	Green	Yellow	Green	Green	Green
NIM	Yellow	Red	Green	Green	Red	Green
NMM-UJ	Red	Yellow	Green	Green	Yellow	Red
NEPTUNE	Yellow	Red	Red	Green	Green	Red

- Green: Meets or exceeds readiness for needed capability
- Yellow: Some capability but effort required for readiness
- Red: Capability in planning only or otherwise insufficiently ready

Recommendation (strongly endorsed by NWS) is to perform phase 2 testing with only MPAS and FV3

Phase 2 testing (MPAS and FV3)

Under development and review by NGGPS DTG

test	what is being tested
1	Option to relax the shallow atmosphere approximation (deep atmosphere dynamics)
2	Accurate conservation of mass, tracers and entropy
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 (further evaluation of performance at cloud-permitting resolutions)
6	Stable, conservative long integrations with realistic climate statistics under idealized forcing with simple physics
7	Suitability of code for integration into NEMS (init/run/finalize structure)

To be completed next year at this time – followed by final NGGPS dycore selection