Atmospheric Prediction - Dynamics NGGPS Team Briefing

Jeff Whitaker and Vijay Tallapragada

NGGPS/HIWPP Phase 1 Testing Summary

- Phase 1 Testing Overview
- Phase 1 Testing Results
  - HIWPP Test Results
  - Advanced Computing Evaluation Committee (AVEC) Evaluations
- NGGPS Phase 2 Testing
  - Overview of tests
  - Schedule

NGGPS Level 1 Testing

- HIWPP Idealized Tests
- HIWPP 3-km, 3-Day Simulations
- AVEC (Benchmarks and Software Evaluation)

HIWPP/NGGPS Phase 1 Dycore Test Summary

Candidate Models

FV3 (GFDL): Cubed-sphere finite-volume with Lagrangian adaptive vertical grid (z or p coordinate) 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 (NCEP): Finite-difference, cubed-sphere version of regional NAM model (p coordinate). Replaced lat/lon grid version with B-grid staggering (*NMMB*).
NEPTUNE (NRL): Spectral-element (horizontal and vertical) cubed-sphere grid (z coordinate) with adaptive mesh refinement.

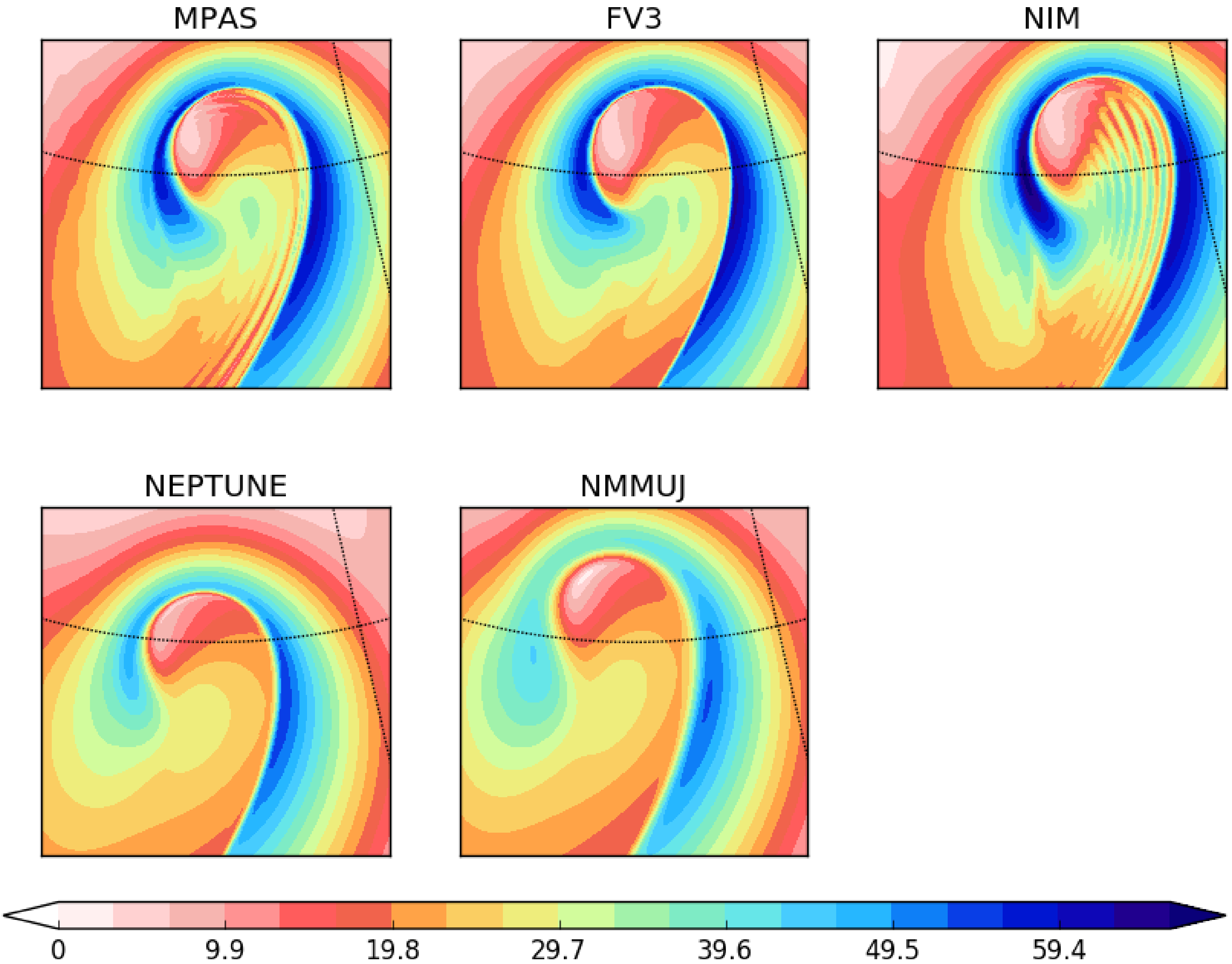
Phase 1 Dycore Testing Overview

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

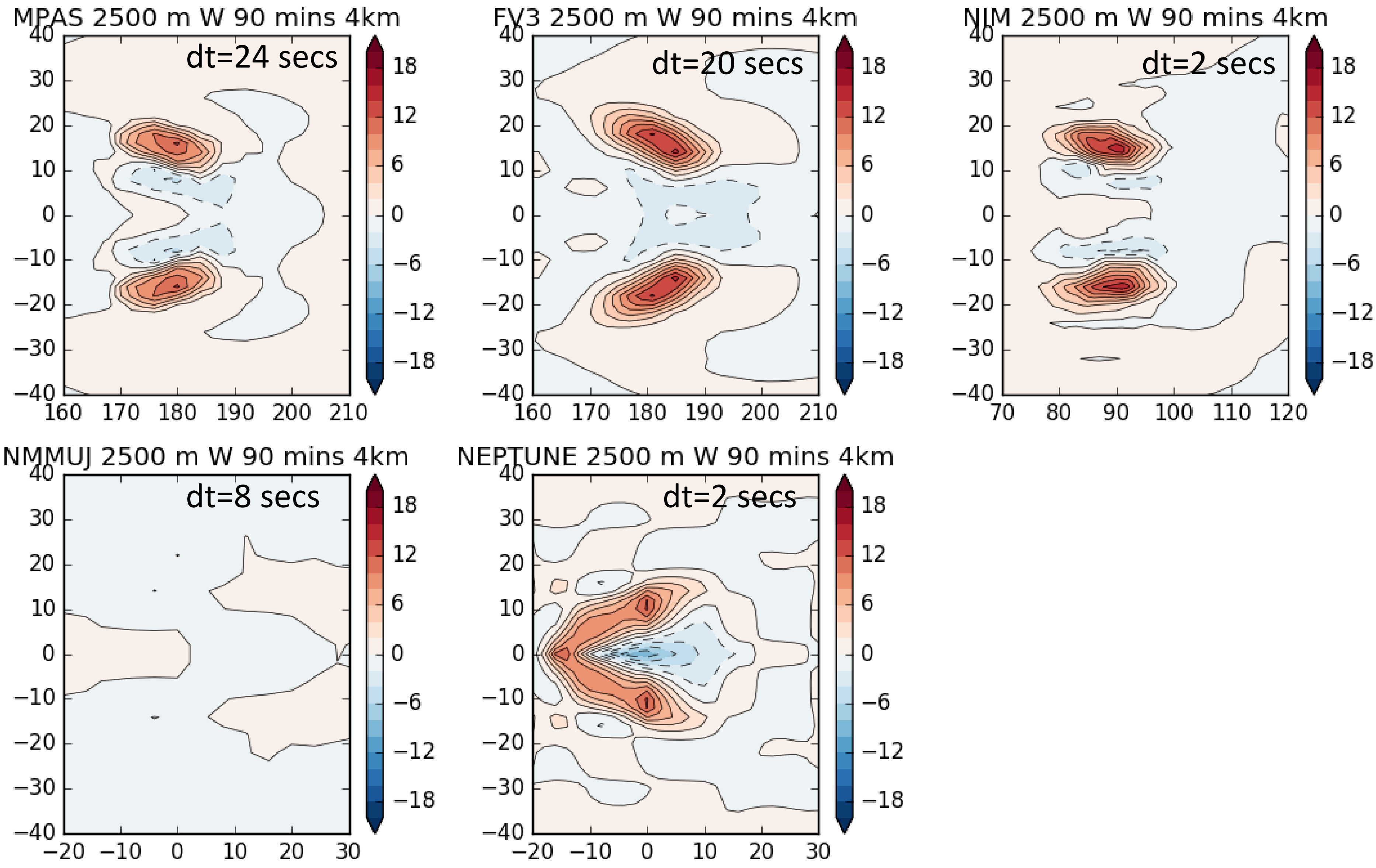
HIWPP Idealized tests

- ***Baroclinic wave test with embedded fronts*** (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 mtn 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)



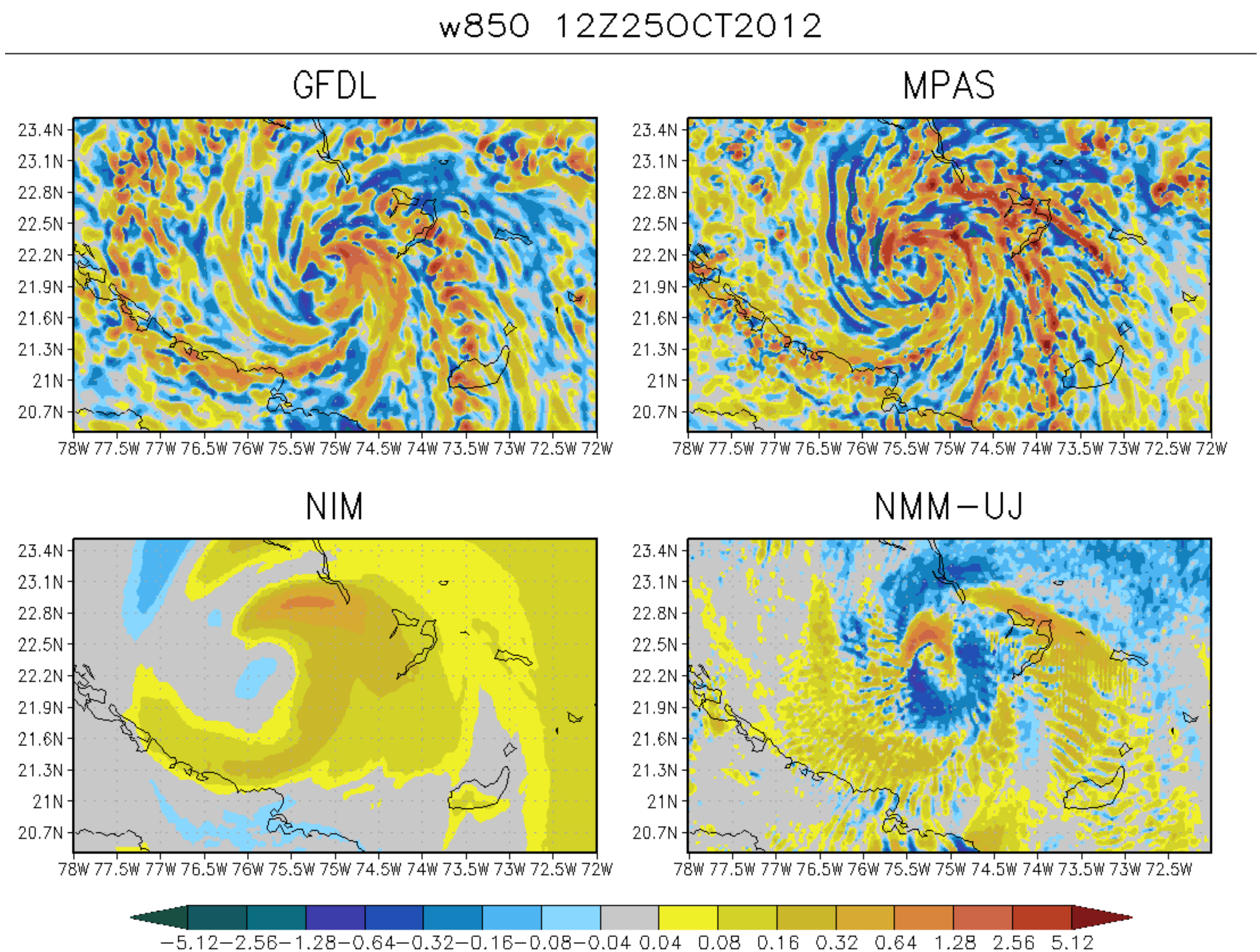
Supercell (2500-m w at 90 mins, 4-km resolution)



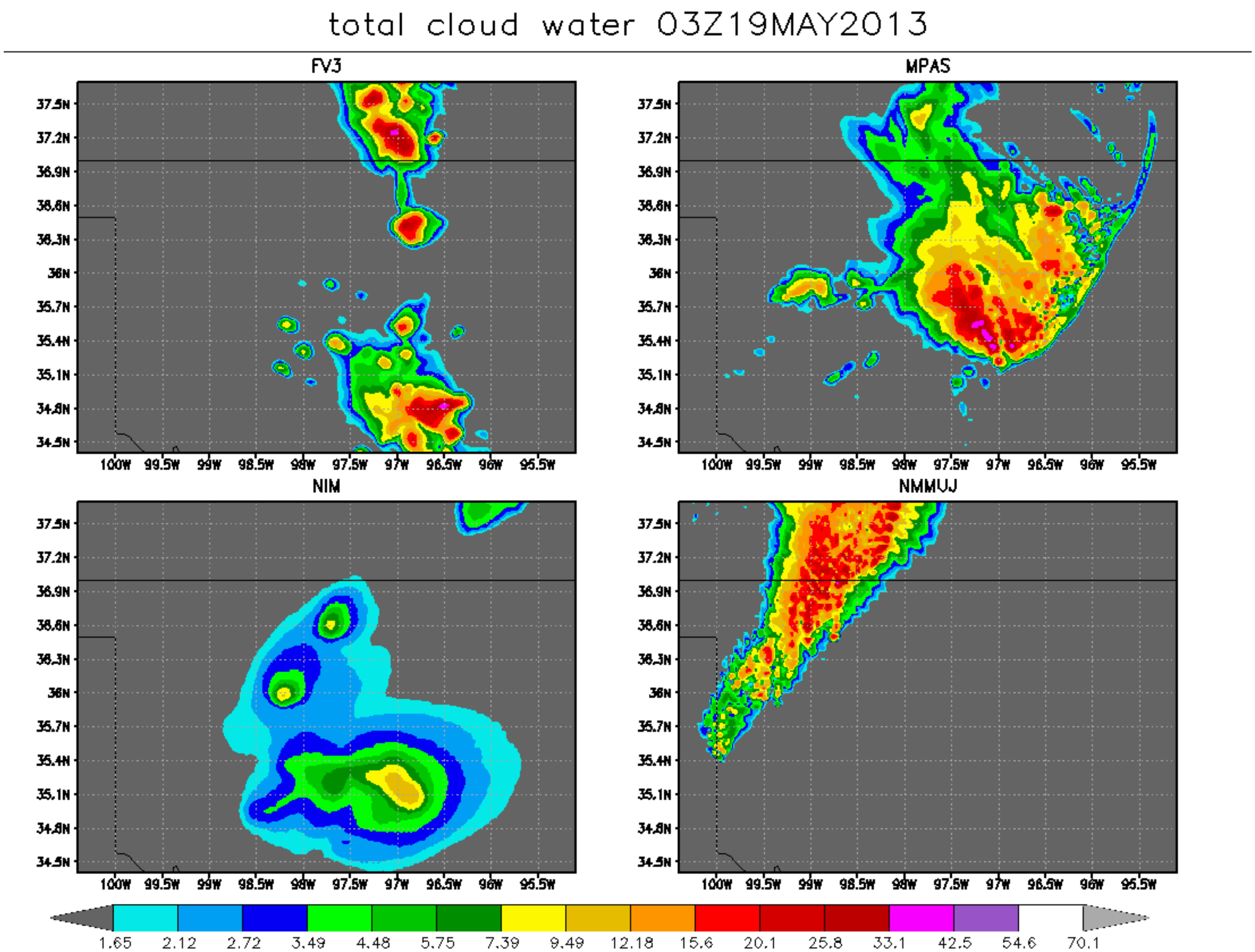
HIWPP 72-h 3-km forecast test

- 'Stress-test' dycores by running with full-physics, high-resolution orography, ICs 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 (w at 850 hPa)



Moore Tornado (total condensate)



Summary

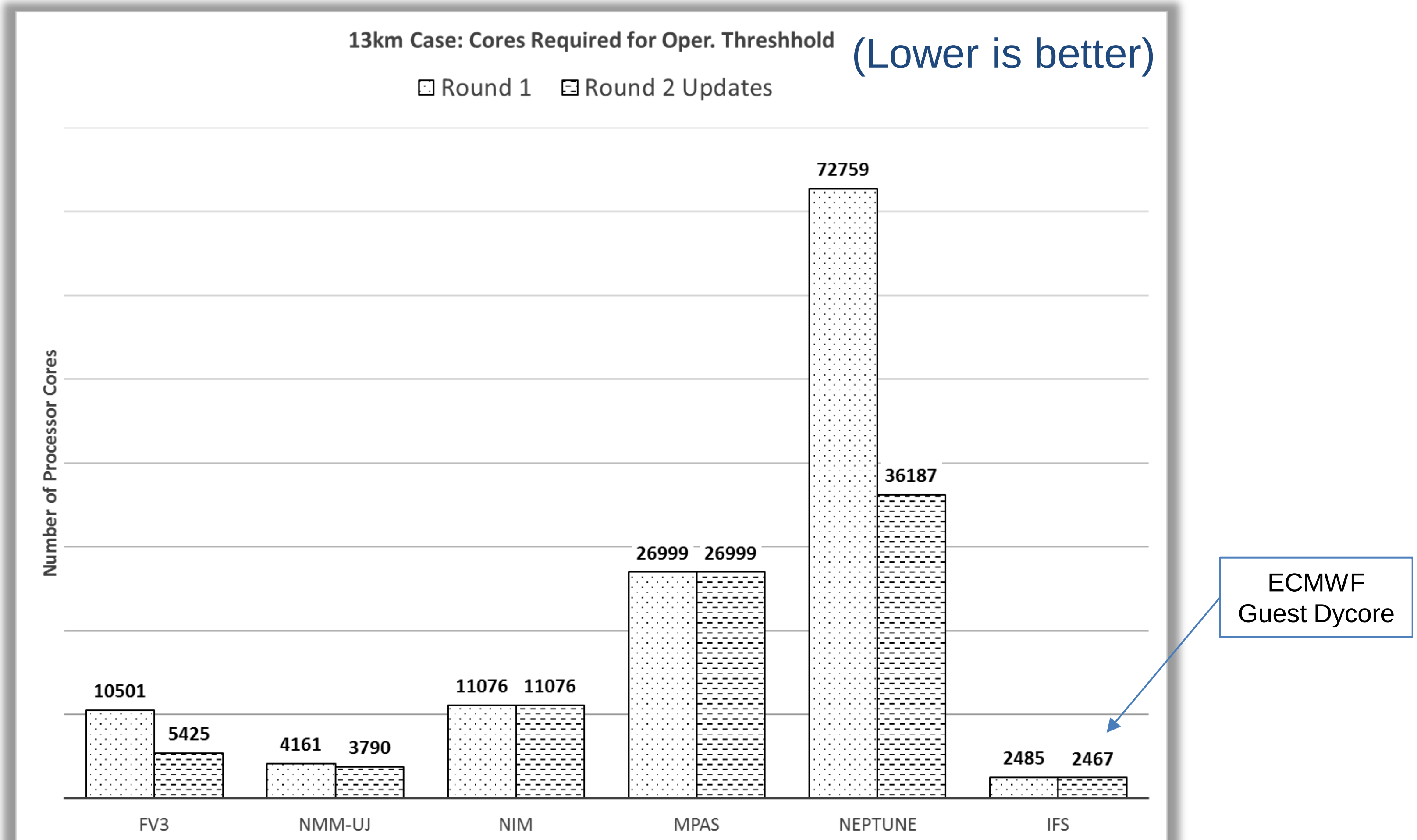
- ***FV3, MPAS*** produced highest quality solutions overall.
  - More similar to each other than other models for all tests.
- ***NIM*** produced reasonable mountain wave and supercell solutions.
  - Excessive noise near grid scale in B-wave solution.
  - Full physics forecasts excessively damped.
- ***NEPTUNE*** was not able to produce full physics 3-km forecasts.
  - B-wave too smooth, 4-km supercell not split by 90 mins.
- ***NMM-UJ*** did not produce realistic solutions for the mountain wave and supercell tests.
  - Vertical velocity fields from full physics forecasts did not show signatures expected from resolved convection.

AVEC Phase 1 Evaluations

- Advanced Computing Evaluation Committee formed August 2014 to evaluate and report on performance, scalability and software readiness of five NGGPS candidate dycores
- Reports
  - NGGPS Level 1 Benchmarks – April 30, 2015
  - NGGPS Level 1 Software Evaluation (addendum to above) – May 28, 2015
- Benchmarks on 130-thousand core HPC system at DOE: “Edison”
  - 13-km and 3-km workloads based on HIWPP non-hydrostatic test case
  - Model groups agreed on each others’ configurations
  - Time step and other configuration options were “best guesses”
  - Groups that changed codes or configurations to improve benchmark performance were required to resubmit results for HIWPP test case

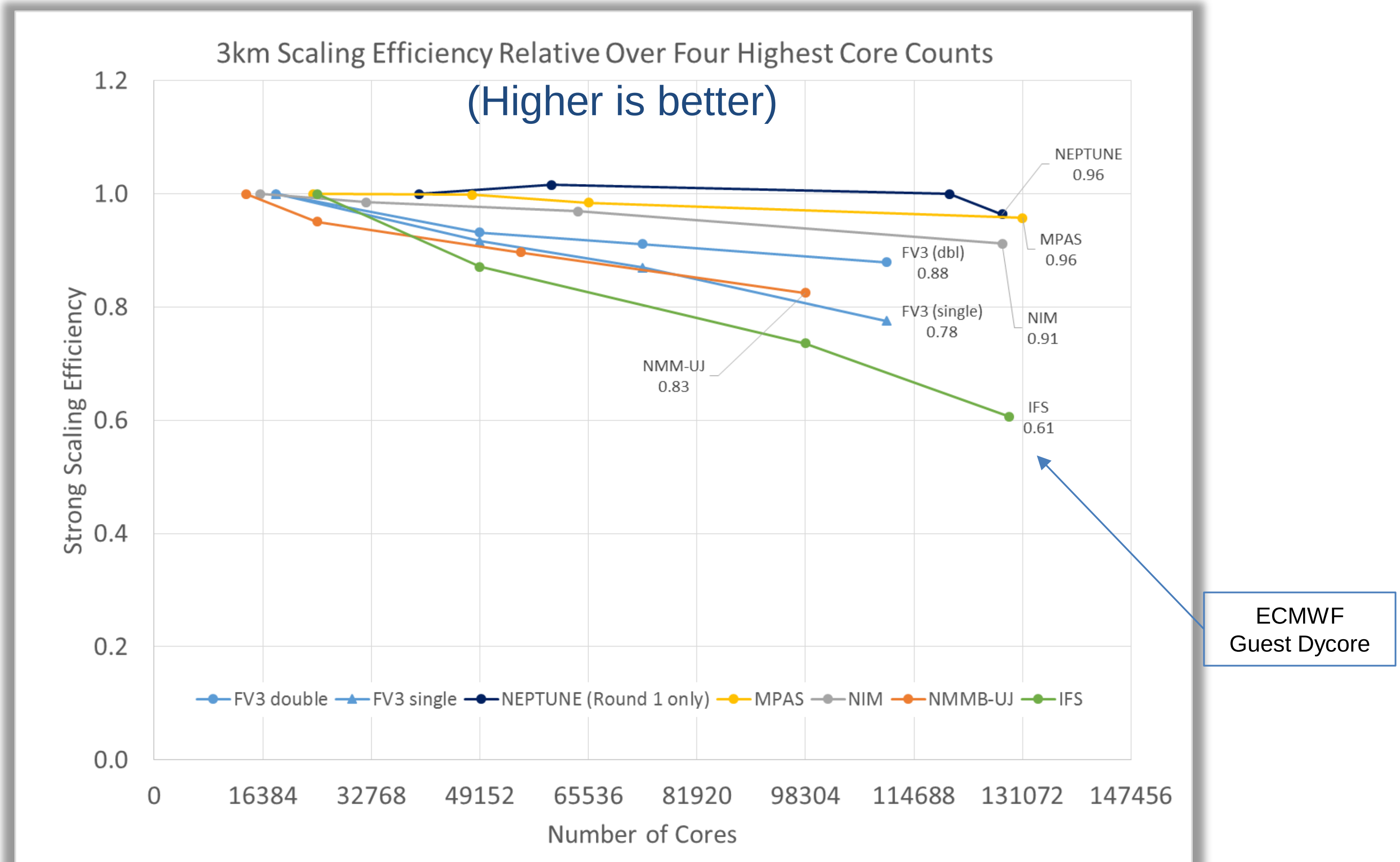
AVEC Phase 1 Evaluations: Performance

- 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 Phase 1 Evaluations: Scalability

- Scalability: ability to efficiently use large numbers of processor cores
  - All codes showed good scaling.
  - Rankings (most to least scalable): NEPTUNE, MPAS, NIM, FV3, NMM-UJ



AVEC Level 1 Evaluations: Software

- Software evaluations intended to highlight strengths and weaknesses of codes to be ready for NGGPS
  - Note: snapshot in time, all codes under active development
- Preliminary results based on self-reports from AVEC questionnaire*:
  - Software maturity: FV3, NIM, MPAS, NEPTUNE, NMM-UJ
  - Nesting or mesh refinement: FV3, MPAS, NEPTUNE, NMM-UJ, NIM
  - Support for thread parallelism: FV3, NIM, NMM-UJ, MPAS, NEPTUNE
  - Reproducibility: FV3, NIM, NMM-UJ, MPAS, NEPTUNE
  - Advanced architectures: NIM, FV3; NMM-UJ, MPAS, NEPTUNE
- Additional evaluation including detailed code inspection and review of documentation will continue into Level 2 testing

*Stoplight color coding by AVEC Chair, John Michalakes (not full AVEC)

NGGPS Phase 1 Testing Summary Assessment

	Idealized Tests	3-km, 3-day forecasts	Performance	Scalability	Nesting or Mesh Refinement	Software Maturity
FV3	🟢	🟢	🟢	🟢	🟢	🟢
MPAS	🟢	🟢	🟡	🟢	🟢	🟢
NIM	🟡	🔴	🟢	🟢	🔴	🟢
NMMUJ	🔴	🟡	🟢	🟢	🟡	🔴
NEPTUNE	🟡	🔴	🔴	🟢	🟢	🔴

🟢 Meets or exceeds readiness for capability
🟡 Some capability but effort required for readiness
🔴 Capability in planning only or otherwise insufficiently ready

Dycore Test Group (DTG) Membership

- Chair: Ming Ji (Director, Office of Science and Technology Integration)
- Fred Toepfer (NGGPS Program Manager)
- Test Manager: Jeff Whitaker (ESRL)
- Bob Gall, Ricky Rood, John Thuburn – independent consultants
- Melinda Peng (Navy – NEPTUNE)
- V. Ramaswamy (GFDL - FV3)
- Hendrik Tolman (EMC – NMM-UJ)
- Chris Davis (NCAR – MPAS)
- Kevin Kelleher (ESRL – NIM)

NGGPS Phase 1 Testing Assessment

- DTG assessed results:
  - Sufficient information is available to proceed with fewer dycores to Level 2 testing
  - Several dycores consistently produced solutions of higher quality and/or were more mature than other dycores
- Final reports (along with DTG charter) available at *http://www.nws.noaa.gov/ost/nggps/dycoretesting.html*

Majority agreed only FV3 and MPAS were ready for Level 2 testing

Phase 2 Tests (FV3 and MPAS)

Phase 2 Eval #	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 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)

Conservation test *Done*

- HIWPP baroclinic wave test case with and without large-scale condensation
  - Measure conservation of entropy/energy/mass
  - Advect extra tracer initialized with equivalent potential temp field, look at differences between tracer and theta_e field derived from model prognostic variables.
  - Diagnose spurious cross-isentropic mass transports.

Robust model solutions under a wide range of realistic atmospheric initial conditions using a common (GFS) physics package *Underway*

- Implement and validate a common physics package in each model (GFS physics).
- Perform 10-day retrospective forecasts initialized from GFS analyses every 5-days for 1 calendar year (2015).
- Resolution (horizontal and vertical) same as current operational GFS (13km, 63 levels, top at 0.6 hPa).
- ESRL will generate ”GFS-lookalike” files from native model output, NCEP will run GFS verification suite.

Computational Performace with GFS physics (AVEC)

- 2 series of benchmarks
  - 1st will measure model to determine the computational resources required to meet an operational speed requirement of 8.5 minutes per forecast day.
  - 2nd will measure the effect on performance of varying the number of tracers being advected.
  - Only performance, not scalability will be measured (Phase 1 tests show both models scale very well).
- 3 horizontal resolutions (11, 13, 15 km)
  - Fit a function that predicts performance given resolution.
  - If 'effective resolution' of each model can be estimated for specific phenomena, this performance function can be used to make an 'apples to apples' comparison of cost.
- AVEC will also evaluate the efficiency of mesh refinement/nesting schemes as well as the cost and complexity of setting up the refined mesh/nest.

Demonstration of variable resolution and/or nesting capabilities, including physically realistic simulations of convection in the high-resolution region

- Real-data (and idealized if time permits) tests will be performed using a variable resolution with 15 km or coarser mesh telescoping to ~3 km over a CONUS-sized domain.
- Common (GFS) physics will be used, with modifications as needed to accommodate a convection permitting domain.
- Simulations of moist convection and hurricanes in the 3 km domain will be evaluated by DTG, supplemented by subject matter experts.

Stable, conservative long integrations with realistic climate statistics

- "AMIP" runs with observed SSTs and sea-ice conditions at reduced resolution (with GFS physics).
- Multi-year integrations will be evaluated for
  - Conservation
  - Grid imprinting
  - Climate statistics

Performance in cycled DA

- Each model will be integrated into the operational ensemble-variational global DA system.
  - By creating ‘GFS-lookalike’ first-guess files, interpolated increments back to native grid.
  - ESRL will develop workflow, work with NCEP to run experiments, evaluate results.
- Want to expose issues that can arise when models are run in a cycled data assimilation system that might not be evident when they are ‘cold-started’ from a ‘foreign’ analysis.

Tasks involving self-reporting

- Plan to implement deep-atmosphere equation set and other features needed for space-weather applications.
- Plan to integrate into the NEMS/ESMF framework.
- Detailed documentation.
- Plan to implement in NCEP operations.

Timeline

- Complete tests: March-April 2016
- Evaluate results, prepare report, make preliminary recommendation: by end of May 2016
- Final decision by NWS management June 30, 2016