Developmental Testbed Center Report
AOP 2016 Activities

1 April 2016 – 31 March 2017

1 Introduction

The Developmental Testbed Center (DTC) is a distributed facility with components at the National Center for Atmospheric Research (NCAR) and the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Global Systems Division (GSD). The purpose of the DTC is to provide a link between the research and operational communities so Numerical Weather Prediction (NWP) research can be efficiently transferred to operations. In addition, the DTC provides the research community access to the latest operational NWP code packages for research applications. The DTC meets its goals by: maintaining and supporting community code packages that represent the latest NWP technology, performing extensive testing and evaluation (T&E) of new NWP technology, developing and maintaining a state-of-the-art verification package, and connecting the NWP research and operational communities through workshops and its visitor program. Over the past year, DTC activities were organized into five focus areas: Verification, Data Assimilation (DA), Hurricanes, Regional Ensembles and Global Model Test Bed (GMTB).

Funding for the DTC is provided by NOAA's National Weather Service (NWS) and Office of Oceanic and Atmospheric Research (OAR), the Air Force (AF), NCAR, and the National Science Foundation (NSF). This report provides a description of the activities undertaken by the DTC between 1 April 2016 and 31 March 2017. These activities include those described in the DTC 2016 Annual Operating Plan (AOP), as well as a few carry-over activities from the DTC AOP 2015. The performance period for the GMTB is 1 July through 30 June. This report also provides a status update on the GMTB activities through 31 March 2017.

1.1 DTC Management

The external management structure of the DTC includes an Executive Committee (EC), a Management Board (MB), and a Science Advisory Board (SAB). Current memberships are listed below. The MB and EC are responsible for approving the DTC Annual Operating Plan (AOP), which defines the work to be undertaken by the DTC in a given year, whereas the SAB is charged with providing the DTC Director with advice on future directions of the DTC and reviewing proposals submitted to the DTC Visitor Program.

The DTC hosted its annual SAB meeting at NCAR's Foothills Campus in Boulder, CO, on 14-15 September 2016. The purpose of this meeting was to discuss strategic future directions for the DTC. Participation in this annual meeting was stronger than usual, with 16 SAB members participating in-person and the 17th member participating remotely. Day 1 of the meeting consisted of briefings from the DTC's operational partners, an overview presentation, highlights by task area and a breakout group discussion on building community. Day 2 started off with a presentation on DTC's community interactions, followed by breakout group discussions by task area. The meeting wrapped up with a briefing and discussion of SAB recommendations. General and task specific recommendations stemming from this meeting are posted on the DTC website (http://www.dtcenter.org/SAB/SAB-recommendations-Sept2016.pdf).

In March 2016, the National Centers for Environmental Prediction (NCEP) director, Bill Lapenta, charged the UCAR Community Advisory Committee for NCEP (UCACN) with conducting a review of four representative testbeds managed or co-managed by NCEP in association with one of the NCEP Centers. The DTC was one of the selected testbeds. The review committee for the DTC consisted of: Shuyi Chen (chair – University of Miami), Peter Neilley (The Weather Company), Lance Bosart (State University of New York [SUNY]-Albany) and Andy Brown (UK Met Office). As part of its review, the DTC management hosted a DTC Overview webinar on 30 September to provide background information to the external review panel, followed by an onsite visit by the review committee on 6-7 October 2016. The onsite visit consisted of presentations by DTC management and task leads and one-on-one discussions between the review committee and a cross-section of DTC staff.

DTC External Management Committees:

Executive Committee		Management Board			
Jim Hurrell	NCAR	Josh Hacker	NCAR	Mike Farrar	NOAA/NWS
Bill Lapenta	NOAA/NWS	Joe Klemp	NCAR	Fred Toepfer	NOAA/NWS
Ralph Stoffler	Air Force	Michael Gremillion	Air Force	Stan Benjamin	NOAA/OAR/ESRL
Kevin Kelleher	NOAA/OAR	Jeff Cetola	Air Force	Jian-Wen Bao	NOAA/OAR/ESRL

Science Advisory Board	
Adam Clark	National Severe Storms Laboratory (NSSL)
Robert Fovell	SUNY – Albany
Kristen Corbosiero	SUNY – Albany
Sharanya Majumdar	University of Miami
Kathy Gilbert	National Centers for Environmental Prediction (NCEP)/Weather Prediction Center
Geoff DiMego	NCEP/Environmental Modeling Center (EMC)
Jenni Evans	Pennsylvania State University
David Gochis	NCAR
S. R. Gopalakrishnan	NOAA/Atlantic Oceanographic and Meteorological Laboratory (AOML)
David Vollmer	United States Air Force (USAF) Academy
Tom Auligne	Joint Center for Satellite Data Assimilation (JCSDA)
Tim Whitcomb	Naval Research Laboratory (NRL)
Brad Colman	Climate Corporation
Zhuo Wang	University of Illinois
Kelly Mahoney	Cooperative Institute for Research in Environmental Sciences
Russ Schumacher	Colorado State University
Kayo Ide	University of Maryland

Over the past year, the DTC hosted two MB meetings: a two-hour conference call on 31 October 2016 and its annual in-person MB meeting on 18-19 January 2017 at NCAR's Foothills Campus in Boulder, CO. The focus of the October conference call was to report on recommendations from the SAB and discuss initial guidance on priorities for AOP 2017. The purpose of the January meeting was to discuss and refine the DTC's proposal for AOP 2017 and discuss nominations for SAB members to replace current members whose term expires in June 2017.

DTC management participated in two DTC EC conference calls (17 May 2016, 29 September 2016) and the annual in-person EC meeting at NWS Headquarters in Silver Spring, MD, on 3 March 2017. Recent DTC accomplishments, recommendations from the SAB, proposed activities for AOP 2017, and the future direction of the DTC were discussed at the in-person meeting. The EC also approved the DTC Director's proposal to rotate off six SAB members whose terms expire in June 2017 (Robert Fovell, Kristen Corbosiero, Sharanya Majumdar, Geoff DiMego, Jenni Evans, and Kelly Mahoney) and add six new SAB members (three-year term begins 1 July 2017). The six new SAB members are: Vincent Larson

(University of Wisconsin-Milwaukee), Xuguang Wang (University of Oklahoma), Tom Galarneau (University of Arizona), Phil Pegion (ESRL/Physical Sciences Division), Rusty Benson (Geophysical Fluid Dynamics Laboratory [GFDL]) and Vijay Tallapragada (NCEP/EMC). Quarterly reports on the progress to date were also prepared for each activity and distributed to the EC and MB members.

1.2 Community Interactions

Maintaining strong ties to both the research and operational NWP communities is critical to the DTC's ability to successfully meet its mission. Over the past year, strong ties with the operational community were maintained through the DTC's interactions with our partners at the operational centers (i.e., EMC and Air Force) both at the management level and through our team lead interactions with the appropriate team leads and/or focal points at the operational centers. The DTC also worked toward strengthening its ties to the broader research community through workshops, tutorials and the DTC Visitor Program. Information on DTC-sponsored tutorials is provided in Section 2.3. The DTC also engages the community through the distribution of its newsletter "Transitions" that serves as a forum for the research and operational communities to share information. Over the past year, the DTC distributed three issues of Transitions. All issues of Transitions can be accessed at: http://www.dtcenter.org/newsletter/.

1.2.1 Community Outreach Events

In June 2016, the DTC co-hosted with NCAR's Mesoscale and Microscale Meteorology (MMM) Laboratory the 17th Weather Research and Forecasting (WRF) Users' Workshop at NCAR's Center Green Campus in Boulder, CO. The first day consisted of lectures on WRF software and best computing practices, followed by a 3-day workshop consisting of 67 talks and nearly 80 posters. The last day consisted of four mini-tutorials on the Mesoscale Model Evaluation Testbed (MMET), Visualization and Analysis Platform for Ocean, Atmosphere and Solar Researchers (VAPOR), ensemble prediction and NCAR Command Language (NCL). The MMET instructional session was organized and conducted by DTC staff. About 190 people from 19 countries attended the workshop (http://www2.mmm.ucar.edu/wrf/users/workshops/WS2016/WorkshopPapers.php).

Also in June 2016, the DTC co-hosted with NCEP/EMC the 7th Ensemble Users' Workshop at the NOAA Center for Weather and Climate Prediction (NCWCP) in College Park, MD. This workshop attracted more than 150 participants representing a broad cross-section of expertise ranging from ensemble developers to the end users of ensemble products. The over-arching goal of the workshop was determining how to support the NWS as it moves toward a seamless operational ensemble forecast system at storm- to global-scales, from short-term to seasonal time scales, using atmosphere-only to ocean-wave and coupled ensemble prediction systems. The workshop consisted of oral and poster presentations, as well as open discussion over a 3-day period. Workshop presentations and a list of attendees are posted at http://www.dtcenter.org/events/workshops16/ensembles/. A report on the workshop is also available on the DTC website (http://www.dtcenter.org/eval/ensembles/).

1.2.2 DTC Visitor Program

The DTC Visitor Program supports visitors to work with the DTC to test new forecasting and verification techniques, models and model components for NWP. The goal is to provide the operational weather prediction centers (e.g., NCEP and Air Force) with options for near-term advances in operational weather forecasting and to provide researchers with NWP codes that represent the latest advances in technology. It also offers an opportunity for visitors to introduce new techniques that would be of particular interest to the research community into the publicly-released software systems supported by the DTC.

Over the past year, the DTC provided support for one visitor project selected in 2014 (see Table 1.2.2-1), four projects selected in 2015 (see Table 1.2.2-2) and four projects selected for funding in 2016 (see Table 1.2.2.-3). During this time period, Dr. Roebber completed his project and submitted his project report, so all projects awarded in 2014 are now complete. Mr. Otkin and colleagues also completed their project and submitted their project report. The remaining 2015 projects are nearing completion, with only the project reports remaining. For the 2016 projects, Mr. Iacono and Mr. Henderson have completed their project and submitted their project report and only the report remains for Dr. Niyogi's graduate student project by Subashini Subramanian. The other projects awarded in 2016 are well underway and the projects awarded in 2017 are either underway or anticipated to get underway in the coming months. All visitor project reports received over the past year are available on the "Visitor Program" portion of the DTC website (http://www.dtcenter.org/visitors/). In addition to project reports and relevant code deliverables, the DTC started scheduling visitor seminars during their final DTC visit that are open to local area scientists as well as remote participants. Feedback from both the visitors and local area scientists about this increased exposure of the DTC visitor projects has been overwhelmingly positive. One additional project is in the process of being awarded.

Table 1.2.2-1. 2014 Visitor Projects

PI	Institution	Project Title
Paul Roebber	University of Wisconsin-Milwaukee	Demonstration project: Development of a large member ensemble forecast system for heavy rainfall using evolutionary programming

Table 1.2.2-2. 2015 Visitor Projects

PI	Institution	Project Title
Jason Otkin	University of Wisconsin - Madison	Object based verification for the HRRR model using simulated and observed GOES infrared brightness temperatures
Gretchen Mullendore (Mariusz Starzec)	University of North Dakota	Mesoscale model intercomparison at convection-allowing resolution using MODE
Dev Niyogi (Xing Liu)	Purdue University	Improving WRF weather forecast through enhanced representation of cropland-atmosphere interactions
Joel Bedard	University of Quebec - Montreal	Implementation and validation of a geo-statistical observation operator for the assimilation of near-surface winds in GSI

Table 1.2.2-3. 2016 Visitor Projects

PI	Institution	Project Title
Michael Iacono / John Henderson	Atmospheric and Environmental Research	Testing revisions to RRTMG cloud radiative transfer and performance in HWRF
Robert Fovell	SUNY-Albany	Impact of planetary boundary layer assumptions on HWRF
Dev Niyogi (Subashini Subramanian)	Purdue University	Developing capability in idealized HWRF for assessing the impact of land surface on tropical cyclone evolution
Shaowu Bao	Coastal Carolina University	Evaluation of the microphysics scheme in HWRF 2016 version with remote-sensing data

Table 1.2.2-4. 2017 Visitor Projects

PI	Institution	Project Title
Patrick Skinner	University of Oklahoma	Quantifying the value of radar data assimilation in the Community Leveraged Unified Ensemble using object-based verification methods
Karina Apodaca	Colorado State University	R2O transition of the GOES-R GLM lightning assimilation capability in GSI for use in the NCEP GDAS
Bill Gallus	Iowa State University	Use of the CLUE to examine importance of mixed physics in ensembles
Paul Roebber	University of Wisconsin-Milwaukee	An Adaptive Bayesian Model Combination (BMC) Post Processor for the HRRR-TLE Forecast Syste
Jiang Zhu	University of Alaska-Fairbanks	Advanced Data Assimilation Techniques Applied to a Regional High Resolution Rapid Refresh Model in Alaska (HRRR-Alaska)
Ting-Chi Wu	Colorado State University	Evaluation of the Newly Developed Observation Operators for Assimilating Satellite Cloud and Precipitation Observations in GSI within the HWRF system

2 Software Systems

To serve as a bridge between operations and research, the DTC provides a framework for the two communities to collaborate in order to accelerate the transition of new scientific techniques into operational weather forecasting. This framework is based on software systems that are a shared resource with distributed development. The current operational systems are a subset of the capabilities contained in these software systems. Ongoing development of these systems is maintained under version control with mutually agreed upon software management plans. The DTC currently works with the following software systems:

- Weather Research and Forecasting (WRF) – NWP model + pre- and post-processors
- Hurricane WRF (HWRF) - set of tools for tropical storm forecasting, including a coupled atmosphere and ocean system
- Unified Post-Processor (UPP)
- Gridpoint Statistical Interpolation (GSI) data assimilation (DA) system
- Ensemble Kalman Filter (EnKF) DA System
- Modular end-to-end ensemble system
- Model Evaluation Tools (MET) – Verification package

The DTC does not generally contribute to the development of new scientific techniques for these software packages. The two exceptions are MET development and some limited physics package development for WRF to address short-comings brought to light by DTC T&E. The DTC contributes to the software management of all of these systems and user support for the publicly-released systems (WRF, HWRF, UPP, GFDL vortex tracker, GSI, EnKF and MET). All software management and user support activities are collaborative efforts with the developers, where the exact role of the DTC depends on the software package. The main developers of these packages are affiliated with EMC, ESRL, NCAR, Global Modeling and Assimilation Office (GMAO) of the National Aeronautics and Space Administration (NASA), National Environmental Satellite, Data and Information Service (NESDIS), JCSDA, GFDL, University of Rhode Island (URI) and the Hurricane Research Division (HRD) of NOAA's AOML.

The DTC is working with EMC to unify the verification systems between the two organizations through MET and METViewer, MET's accompanying database and display system. DTC staff visited EMC for a week during early May. This visit consisted of 18 meetings with approximately 50 EMC staff to discuss their current verification practices and immediate needs. Information gathered during these meetings is summarized in a requirements document that was released to all participant as well as the EMC Director on 15 September 2016. The DTC is using this information to develop a unified verification system, called MET+, through Next Generation Global Prediction System (NGGPS) funding outside of the DTC. Briefly, MET+ is a set of python wrappers to simplify setting-up and running MET to allow researchers to leverage their own unique algorithms, and systematically plot the fields and results.

For the GMTB, the DTC has been working with EMC and the physics-development community (ESRL, NCAR, NRL, and universities) to establish a Common Community Physics Package (CCPP) that will serve as a framework for efficiently developing and transitioning current and next-generation physics parameterizations into operations to meet the needs of NGGPS. Another important component of this work is establishing an Interoperable Physics Driver (IPD) that provides a framework for physical parameterization suites within the CCPP to interface with different dynamic cores. Over the past year, the requirements for the IPD and CCPP have undergone extensive review and refinement, and informed a software design. The initial design was presented to EMC and the National Unified Operational Prediction Capability (NUOPC) Physics Interoperability group, as well as several subsequent revisions that incorporated feedback from these groups. This design document, as well as the physics aliasing layer (also known as IPD v4) development performed by GFDL and made available for FV3 in March 2017, served as the foundation for the initial CCPP and IPD. GMTB's contribution to the IPD effort will enable constructing physics suites at runtime by parsing a user-friendly configuration file, allowing for running the parameterizations within the CCPP in a very flexible manner.

The GMTB intends to facilitate an environment for the CCPP, defined as a set of practices in which design, development, and deployment happen simultaneously and rapidly in the same ecosystem. To support this basic principle of modern software design, the GMTB also completed documents that describe the concept and management of the CCPP development and software, as well as a proposed governance structure to manage the evolution of the CCPP. The concept and design of the CCPP describes an ecosystem for development and transition of physics, where an EMC developer can easily develop on his own or engage external developers. Code is elevated to supported or operational status following a suite of scientific tests. Details of these tests are to be determined by a governance structure yet to be finalized. A proposal for this governance structure is included in the GMTB's CCPP Roadmap. The GMTB test harness will play a key role in making these tests accessible to all developers. A code management plan was created to support the ecosystem and meet the governance needs.

During the final two months of the reporting period (Feb-March 2017), the CCPP was created with placeholder physics – a skeleton to guide the connection to real physics. For testing purposes, the GMTB Single Column Model (SCM) was modified to call a suite composed of the CCPP placeholder physics through the IPD. This implementation can be considered the simplest "dycore cap" for the IPD – the SCM replaces actual dynamics with advective forcing, but it enables the translation of SCM state and diagnostic variables to those within the physics suite. The first CCPP release is planned for early 2018. NCAR and GSD physics developers have been engaged, and will be connecting new physics under separate funding

Building on documentation efforts from year 1, comprehensive technical web documents are now served from the DTC website, describing both the background and current function of the IPD v2 used during the NGGPS dynamical core test and the initial member of the CCPP, the 2016 operational Global Forecast System (GFS) physics suite. The content of both documents was generated roughly equally by the NCAR and GSD members, drawing on expertise from all contributors as appropriate. A review by physics experts at NOAA EMC was solicited and obtained, with feedback integrated into the final documents. The documentation for the IPD and CCPP can be accessed at the following URLs, respectively: http://www.dtcenter.org/GMTB/gmtb_ipd_doc/ and http://www.dtcenter.org/GMTB/gfs_phys_doc/.

The process for updating the documentation going forward is in flux at this time. At the request of EMC, source files containing the Doxygen-formatted comments used to generate the documentation webpages were merged into the trunk of EMC's code repository. In addition, all external figures and files necessary to generate the documentation have been placed inside a 'docs' directory next to the physics source files. As a result, documentation may be version controlled and it is possible for both DTC staff and those at NOAA EMC to generate the documentation output. Although a plan for fully integrated documentation for all components of NGGPS is not complete, the DTC has provided input to decision-makers within NOAA regarding the use of Doxygen as the appropriate tool and for defining the process for developers to update documentation as development progresses.

2.1 Software Management

While specific software management plans differ between the various software packages, they all contain the following elements:

- Code repositories maintained under version control software.
- Protocols for proposing modifications to the software, whether the modifications are simply updates to current features, bug fixes or the addition of new features.
- Testing standards proposed software modifications must pass prior to being committed to the code repository.
- Additional testing standards used to more thoroughly check the integrity of the evolving code base.

Given all these software packages continue to evolve over time, all testing standards must be updated periodically in order to meet the maintenance requirements of the code base. Over the past year, the DTC continued to collaborate with the various developer groups on these ongoing software management activities. The DTC also continued to provide a pathway for the research community to contribute to the development of these software systems. Noteworthy events from this work over the past year are:

- ***WRF*** –Over the past year, work towards adding an option to run the Advanced Research WRF (ARW) dynamic core with a new smoothed terrain-following hybrid-vertical coordinate was completed. This code will be included in the next release of WRF and will be backwards compatible if the new option is not selected. All 2016 operational HWRF forecast system capabilities were committed to the WRF trunk, and were made available for the next community release of WRF. These capabilities include the GFS hybrid Eddy-Diffusivity Mass-Flux (EDMF) Planetary Boundary Layer (PBL) scheme (available for both ARW and Nonhydrostatic Mesoscale Model on the E grid-NMME), updates to the surface flux exchanges with the coupled ocean (NMME only), updated scale-aware Simplified Arakawa-Schubert (SAS) cumulus parameterization (ARW and NMME), a landfall option for the idealized tropical cyclone (TC) capability (NMME only), as well as miscellaneous bug fixes and tuning parameters.

- ***UPP*** – The DTC continued to work closely with EMC to manage the UPP code base through regular bi-monthly meetings. To streamline the efforts to keep the community UPP in sync with EMC's operational UPP, the community UPP source code was migrated to a branch of the EMC repository for easier code sharing and syncing. Through collaborations with operational developers and supported DTC visitors, the most recent community release of UPP included full

GRIB2 output capability, along with new microphysics-specific reflectivity output and synthetic satellite fields.

- ***HWRF*** – The DTC continued to support HWRF developers in using and adding innovations to the code repository. The DTC completed development to enable the multistorm configuration to run using the same setup as the operational configuration, with the exception of ocean coupling, and added this capability to the HWRF repository. Fixes were committed to the repository to enable backwards compatibility for the previous operational HWRF configuration. Additionally, the DTC developed a new capability to start a HWRF run from the WRF component when using wrappers to aid HWRF community developers who want to modify and run the atmospheric forecast component without gaining expertise on the full end-to-end system. The ability to simulate landfall within the idealized tropical cyclone capability, an innovation developed by a DTC visitor, was transitioned to the HWRF code repository. The HWRF repository was adapted to accommodate the transition of the WRF and WRF Pre-processing System (WPS) repositories from Subversion (SVN) to Git and the community GSI repository transition to VLab. Updated procedures and instructions were published to the HWRF developers' webpage and assistance was provided via the HWRF helpdesk. The DTC provided coordination of development activities by chairing the HWRF developers' committee bi-weekly meetings. Additionally, DTC provided enhanced support for developers contributing to the HWRF system, including Hurricane Forecast Improvement Project (HFIP)-funded principal investigators. To facilitate inter-developer collaboration, the DTC continued to host an *hwrf-contrib* repository for peer-to-peer sharing of code. The NCAR and GSD staff conducted this work jointly. The HWRF v3.7a and v3.8a Users' Guides were published as GSD technical notes:

  Biswas, M. K., L. Carson, C. Holt, L. Bernardet, 2016: Community HWRF Users Guide V3.7a. NOAA Technical Memorandum OAR GSD-46, doi:10.7289/V5SJ1HMD, 144 pp.

  Biswas, M.K., L. Carson, K. Newman, L. Bernardet, C. Holt, 2017: Community HWRF Users' Guide V3.8a, NOAA Technical Memorandum OAR GSD-47, doi:10.7289/V5/TM-OAR-GSD-47, 149 pp.

- ***GSI and EnKF*** – Over the past year, the DTC made a few critical upgrades to current code management and support efforts for the DA systems. The DTC transitioned the helpdesk to Request Tracker (RT) for better tracking of users' requests and questions. The DTC built a new community repository (svn) on NOAA's Vlab server and successfully transitioned all community repository developers to this new repository. The DTC also initiated efforts to work with EMC to improve the efficiency of code management from both sides as follows: started efforts to unify the code build tool (using cmake) for GSI, EnKF, and NCEP I/O libraries; converted the user's guides to LaTeX to share with all developers through the code repository; and transitioned community utilities (e.g., format conversion, diagnostic plotting scripts) to the EMC repository. These efforts will contribute significantly to the unification of the DTC-EMC code repositories over the coming year. On-going efforts included: support GSI/EnKF users through the new code repository and helpdesk, perform code reviews for each proposed code update and synchronize the DTC community code repository with the trunk of EMC's operational repository, coordinate the GSI/EnKF development among distributed developers by chairing the DA Review Committee and hosting three review committee meetings. A notable outcome of the review committee meetings was the addition of JCSDA as the 10th member of the committee.

For AOP 2016, the DTC scaled back its work with the NOAA Environmental Modeling System (NEMS)/ Nonhydrostatic Multiscale Model on the B grid (NMMB) due to the lack of any T&E activities utilizing this software package. While the DTC is no longer actively engaged in code management or user support for this code base, it did continue to test the portability of the NEMS software package and associated libraries through regression testing for regional applications on the NCAR supercomputer, Yellowstone. Over the past 6 months, significant infrastructure and code architecture changes over the past six months led to failed portability tests and retrofitting to a new platform is considered beyond the scope of the DTC activities at this time. Relevant feedback was provided to EMC NEMS developers.

2.2 Verification Tool Development

Over the past year, the DTC verification team completed two MET releases. MET v5.2 was released to the community on 15 August 2016 and MET v6.0 on 3 April 2017. Approximately 20 major enhancements and 50 smaller bug fixes were supported to the community in the releases, including the upgrade to reading/writing Network Common Data Form (NETCDF)4, a 25% speed-up of the Gridpoint Statistical Interpolation (GSI) diagnostic tool, and configuration file changes to make MET easier to set-up.

To support global verification, the grid-to-grid verification tools were updated to include a configuration option to apply grid box area or cosine-latitude weighting to the computation of continuous statistics, such as Anomaly Correlation, Root Mean Square Error and Bias. All tools are provided with GRIB1/2 table support for non-NCEP tables such as UK Met Office and European Center for Medium Range Forecasting (ECMWF). The ASCII2NC tool, which creates NetCDF files from point observations in ASCII format, was enhanced to include configuration options to specify the expected frequency of observations and omit output when not enough valid data are present. This feature was requested by the NEMS Global Aerosol Component (NGAC) group. A “DESC“ column (short for description) was added to MET to allow users to add a descriptor for more effective stratification of statistics in both MET and METViewer. To support NOAA’s tropical cyclone (TC) verification, the TC-pairs tool, which matches TC forecasts and observations, was updated to handle interpolated models whose model id ends in '3', read probabilistic forecasts from the “E-deck” file format, and include more flexible ways of passing the tool different filenames for the best and operational tracks.

Several measures and methods were added to MET to support verification of total cloud fraction on the global scale. New interpolation methods were added to allow interpolation of model output to be handled in the manner similar to the World Wide Merged Cloud Analysis (WWMCA) cloud fraction mapping. The tool for regridding WWMCA was enhanced to draw additional fields, such as Satellite ID (SatID) and pixel age, from the binary cloud analysis files. Also, the Gen-VX-mask tool, which creates a bitmapped masked area, was updated to compute the solar azimuth and angle values based on location and time of day to allow for the derivation of day/night mask.

The METViewer database and display system is another verification tool under development during the AOP. It was modified to support the new file formats, statistics and “line” types introduced via the METv5.2 and METv6.0 releases. The DTC modified the loading logic to handle special cases introduced by different versions of the NOAA mesoscale and ensemble verification statistic database file format (VSDB). The interface was also enhanced to compute the mean and a ratio of several curves to support the Regional Ensembles team. Basic database purging scripts were also generated. METViewer v1.9-1.12 were released, including the addition of a Taylor Diagram template, enhancements to the event equalization logic and the Method for Object-based Diagnostic Evaluation (MODE) attribute computation for EMC’s Mesoscale and Global branches, respectively. METViewer database design was interrogated, several changes were made to speed up data loading and querying. These changes included partitioning the database, not attempting to load empty files, and providing the capability to remove a single or multiple records, if needed. New database technologies (e.g. Couchbase) were also investigated and found promising. The DTC also investigated options to modify the METViewer interface to allow user logins and save user preferences.

Telecons with NOAA MET and METViewer users were held every month, starting in October 2016. Invitees include 25 engaged NOAA staff from EMC, Weather Prediction Center (WPC) and GSD. The discussions were driven by questions arising from increased use of MET and METViewer. In response to these meetings, the DTC staff answered many questions, fixed several bugs introduced with loading VSDB data into METViewer, improved the speed of loading METViewer and added new features to both MET and METViewer. METViewer code was moved into GitHub to allow both DTC nodes and external collaborators access. Additionally, planning and development of the initial python wrappers around the components, called MET+, was distributed across both DTC nodes. The GSD node contribution established a greater understanding of the MET+ components (MET and METViewer) and allowed more meaningful collaboration between the nodes.

2.3 Publicly-Released Systems

The DTC currently collaborates with developers on seven software systems that undergo a public release process: WRF, UPP, HWRF, GFDL vortex tracker, GSI, EnKF and MET. Assistance continued to be offered through email helpdesks for all packages. Information regarding the timing and version of the most recent release, along with the current number of registered users and average helpdesk tickets per month for each package are listed in Table 2.3-1. Table 2.3-2 contains a list of the web addresses for each software package's users' page.

Table 2.3-1: Code releases, number of registered users and number of helpdesk tickets per month for the publicly-released software packages supported by the DTC over the past year.

Software Package	Public Release			
	Version	Timing	Registered Users	Helpdesk tickets per month
WRF	V3.8	April 2016	~32,700	~400
	V3.8.1	August 2016		
UPP	V3.1	September 2016	~740	~10
HWRF	V3.8a	November 2016	1399	~30
GFDL Vortex Tracker	V3.5b	September 2013	617	
GSI	V3.5	August 2016	1,687	~20
EnKF	V1.1	August 2016		~2-5
MET	V5.2	August 2016	3180	~20-25

Table 2.3-2: Users page websites for publicly-released software packages.

Software Package	Users Websites
WRF	http://www.mmm.ucar.edu/wrf/users/
UPP	http://www.dtcenter.org/upp/users/
HWRF	http://www.dtcenter.org/HurrWRF/users/
GSI	http://www.dtcenter.org/com-GSI/users/
EnKF	http://www.dtcenter.org/EnKF/users/
MET	http://www.dtcenter.org/met/users/

In addition to general MET user support, the DTC verification team actively responded to requests from NOAA users regarding the use of MET and METViewer. Accomplishments over the past year include: 1) increased number of MET and MetViewer users from NOAA (from 74 to 100); 2) worked with NCEP Environmental Modeling Center (EMC), Weather Prediction Center (WPC), and Climate Prediction Center (CPC) to identify requirements necessary to unify verification between DTC and EMC; 3) white paper elucidating the requirements submitted to points of contact at EMC and EMC director; 4) started exploring database designs to handle large datasets; and 5) participated in the NGGPS Verification and Validation Team discussions.

The DTC revamped the online tutorials for both GSI and EnKF. The online tutorials were redesigned to provide more user-friendly instructions along with the latest GSI and EnKF capabilities. Test cases were carefully selected with new testing periods and domains, covering updated and additional configurations with alternative data assimilation techniques [3DVar, 3D hybrid, 4D hybrid Ensemble Variational (EnVar)], data types (conventional and satellite radiance data), and forecast models (ARW, HWRF, NMMB, WRF-chem, GFS).

2.4 DTC-supported software containers

Many times the biggest hurdle when running a new software system is getting it set up and compiled on the intended computer platform. Building complex systems that require a number of external libraries can be a large issue for users to overcome. In order to relieve some of this difficulty, a new technology referred to as a “container” has been developed that allows for complete software systems to be bundled and shipped to users. The containers include everything that is needed to run the software component, including the operating system (tools and libraries) and code - thus allowing for the user to quickly produce output without being delayed by technical issues.

Containers have been established outside of the DTC for portions of an end-to-end NWP system, including WPS, WRF, and NCL. During AOP 2016, DTC staff established containers for the UPP and MET software systems. In addition, datasets that make up two Mesoscale Model Evaluation Testbed (MMET) cases were bundled in a container. Containers are available via GitHub to run MET (https://github.com/NCAR/container-dtc-met) and to run the end-to-end system (including MET and MMET dataset (https://github.com/NCAR/container-dtc-nwp). By establishing these additional containers, the DTC is assisting the user community (especially students) with efficiently running NWP components in an effort to foster connections with future collaborators.

3 Testing and Evaluation

T&E activities undertaken by the developers of new NWP techniques from the research community are generally focused on case studies. However, in order to adequately assess these new technologies, extensive T&E must be performed to ensure they are indeed ready for operational consideration. DTC T&E generally focuses on extended retrospective time periods. The cases selected incorporate a broad range of weather regimes ranging from null, to weak and strong events. The exact periods chosen vary based on the phenomenon of focus for the test. The technique to be tested must be part of the code repositories supported by the DTC to ensure that the code has reached a certain level of maturity. The DTC’s evaluation of these retrospective forecasts includes standard verification techniques, as well as new verification techniques when appropriate. All verification statistics undergo a statistical significance (SS) assessment when appropriate. By conducting carefully controlled, rigorous testing, including the generation of objective verification statistics, the DTC is able to provide the operational community with guidance for selecting new NWP technologies with potential value for operational implementation. DTC testing also provides the research community with baselines against which the impacts of new techniques can be evaluated. The statistical results may also aid researchers in selecting model configurations to use for their projects.

3.1 Regional Ensembles

Mesoscale NWP systems are utilized in both research and operational forecasting applications and can be configured to suit a broad spectrum of weather regimes. Due to the number of approaches developed and offered by NWP systems, it is necessary to rigorously test select configurations and evaluate their performance for specific applications.

One paper associated with a past Regional Ensembles activity was published in Monthly Weather Review:

Jankov, I., J. Berner, J. Beck, H. Jiang, J.B. Olson, G. Grell, T. G. Smirnova, S. G. Benjamin, J. M. Brown, 2017: A performance comparison between multiphysics and stochastic approaches within a North American RAP ensemble. *Mon. Wea. Rev.*, **145**, 1161-79.

3.1.1 Mesoscale Model Evaluation Testbed (MMET)

The Mesoscale Model Evaluation Testbed (MMET; http://www.dtcenter.org/eval/meso_mod/mmet) provides the opportunity for the research community to conduct their own T&E of a new technique. Datasets for a number of cases deemed to be of high interest by EMC are distributed via RAMADDA, a Repository for Archiving, Managing and Accessing Diverse DAta (http://ramadda.org/). MMET datasets include a variety of initialization and observation datasets, as well as baselines for select operational configurations. Cases of interest and/or persistent operational model issues were identified throughout the year by leveraging a direct link to EMC's Model Evaluation Group (MEG) through DTC staff participation in MEG weekly telecons. A list of operational cases of interest and/or persistent model weaknesses was compiled based on these weekly discussions and publicized on the MMET webpage (http://www.dtcenter.org/eval/meso_mod/mmet/additional_cases.php).

In the past, as new versions of the WRF and NEMS code were released, MMET cases were rerun to provide current, baseline results for the user community. While the full end-to-end system is no longer updated on an annual basis, previous versions will remain available through the RAMADDA data server. Rather, this year, operational model output for several NWP systems (both deterministic and probabilistic) were evaluated and the objective verification scores provided to the research community through MMET. The operational forecast systems included the North American Mesoscale (NAM), Rapid Refresh (RAP), High-Resolution Rapid Refresh (HRRR), and Hurricane WRF (HWRF). These baselines are provided for all new and existing MMET cases for each available operational model. Two new cases were established this year: 1) a case over Alaska (20150826: Typhoon Atsani remnants affecting Alaska) and 2) a hurricane case (20160928-29: Hurricane Matthew). Another new addition this past year is the evaluation of the Storm Scale Ensemble of Opportunity (SSEO) data collected from the 2016 Hazardous Weather Testbed (HWT) Spring Experiment. This evaluation included deterministic performance results for each individual SSEO member along with probabilistic results from the ensemble as a whole for select variables; the plots for select variables, thresholds, and metrics, as well as the METViewer XMLs used to generate the plots, are available via RAMADDA.

In addition, community outreach events continue to be provided by the DTC to promote enhanced connections with future community collaborators and promote the use of MMET datasets. A 1.5 hour instructional session was offered during the 17th WRF Users' Workshop in June 2016 to raise awareness about and highlight the tools available to the community-at-large through MMET. A poster was presented on MMET at the American Meteorological Society (AMS) Annual Meeting in January 2017;

Harrold, M., J. K. Wolff, and T. Hertneky, 2017: Mesoscale Model Evaluation Testbed (MMET): Helping Connect the Research and Operational Communities. *28th Conference on Weather and Forecasting/24th Conference on Numerical Weather Prediction,* Seattle, WA, January 24-27, 2017.

and a proposal to hold a short course on using DTC-supported containers (including MMET cases) at the 2018 Annual AMS meeting is being written for submission at the end of April. Finally, a manuscript was prepared and accepted for publication in the November edition of the Bulletin of the American Meteorological Society (BAMS).

Wolff, J. K., M. Harrold, T. Hertneky, E. Aligo, J. Carley, B. Ferrier, G. DiMego, L. Nance, Y.-H. Kuo, 2016: Mesoscale Model Evaluation Testbed (MMET): A resource for transitioning NWP innovations from research to operations (R2O). *Bull. Am. Met. Soc.* DOI: http://dx.doi.org/10.1175/BAMS-D-15-00001.1

3.1.2 HRRR Enhancements

The High-Resolution Rapid Refresh (HRRR) model became operational at NCEP on 30 September 2014, and has found wide acceptance by forecasters inside and outside of the National Weather Service as guidance for a variety of weather phenomena across all seasons. The HRRR uses the ARW dynamical core, a physics package that has proved effective at capitalizing on the cloud-permitting resolution of the model, and unique initialization procedures using radar and satellite data, as well as conventional in-situ observations, together with a Rapid Refresh (RAP) forecast. In view of the importance of accurate short-term forecasts for vulnerable coastal areas, particularly along the Gulf and Atlantic coasts, the DTC investigated the value of expanding the HRRR domain toward the east and south. At the time this work was originally proposed, the operational NCEP HRRR forecasts only extended to 15 h, but ESRL is now running the HRRR experimentally to 36 h. The lateral boundaries of the HRRR domain are often within the circulation of northeast coastal snowstorms and land-falling tropical cyclones when these systems are within a 24-36-h striking distance of the US mainland. Expanding the boundaries of the current HRRR domain thus became increasingly important for these longer forecast lengths.

Three tropical cyclone case studies were conducted with HRRR version 2 (HRRRv2), which is the version implemented operationally at NCEP in August 2016. The model initialization times for the three storms were chosen such that the center of each storm was near or slightly outside the standard domain but well inside the extended domain. The storms and initialization times of interest were: Bonnie (12 UTC 27 May 2016), Colin (00 UTC 6 Jun 2016) and Hermine (12 UTC 31 Aug 2016). Extended domain simulations were also conducted for Hurricane Hermine, with and without lightning data assimilation, to a forecast length of 36 h. Hermine was chosen based on the large number of lightning strikes occurring in an area of relatively low simulated reflectivity values at model initialization time.

Project deliverables were two-fold: (1) upgraded procedures for initialization of convection-permitting models over ocean areas prone to deep convection, mesoscale convective systems and tropical cyclones, and (2) a recommendation for domain configuration and physics suite for an expanded HRRR. The results of our experiments, particularly those for Hurricane Hermine, suggested forecast value associated with expanding the HRRR domain to the south and east. Results of sensitivity tests to lightning data assimilation prompted inclusion of a lightning reflectivity proxy algorithm within the experimental HRRRv3, for eventual implementation at NCEP in 2018. A more comprehensive summary of this work is available online at:
http://www.dtcenter.org/eval/ensembles/dtc_expanded_domain_report.docx

3.1.3 Testing and evaluation of smoothed terrain-following coordinate in WRF

Once the new smoothed, terrain-following coordinate was committed to the RAP/HRRR repository in late 2016, the GSD node of DTC began immediate testing to analyze the impacts of this new vertical coordinate. This new feature was employed in the experimental RAP and HRRR models run at GSD, and a number of cold start RAP runs were conducted as controlled tests of the new vertical coordinate. It quickly became clear that a problem existed, as numerous model crashes followed and spurious/excessive jet-level wind characteristics were found upon analysis of model results. These findings were shared with the WRF developers. A bug related to the map scale factor was found and corrected in early 2017.

Following a new version of the code being committed to the GSD repository, testing began again in March 2017 to assess the hybrid vertical coordinate. Initial qualitative tests using cold-start RAP forecasts produced stable results. Differences were found to be negligible at initialization time, but by six hours into the forecasts, 250-hPa wind speed values within the core of the upper-level jet for the hybrid coordinate were found to exceed values in forecasts using the original vertical coordinate. Outside of the major upper-level jet, a number of areas were found where winds were slower for the hybrid coordinate. Most of the differences were found to occur over, or downwind of major mountain ranges, specifically in the western areas of North America.

Further testing was undertaken with a week-long, fully-cycled RAP experiment, conducted for the period between 7-13 March 2017 and another five-day experiment for 3-7 September 2016. Results from these two retrospective runs indicated the hybrid coordinate results were sensitive to the strength of the upper-level jet. For the September 2016 experiment, Contiguous United States (CONUS)-wide, upper-level relative humidity and wind speed RMSE and bias improved by a statistically significant amount when the hybrid coordinate was used. For the March 2017 time period (Fig. 3.1.3-1), similar statistically-significant reductions in RMSE and bias were found for the CONUS as a whole; however, wind speed bias for the Western CONUS showed a statistically-significant degradation. The upper-level jet was stronger in the March 2017 retro, and therefore it is hypothesized that the smoothed, terrain-following coordinate is resulting in winds that are slightly too strong over mountainous terrain during synoptically active periods. It should be noted that verification below about 500 hPa showed no statistically significant differences between the two retrospective simulations, highlighting the impact of the hybrid vertical coordinate on jet-level winds, particularly in mountainous regions.

A final week-long simulation was conducted with the HRRR from 3-10 September 2016 to assess the impact of the smoothed, terrain-following coordinate at convective-resolving resolution. Results from this retrospective simulation showed a vastly reduced impact of the hybrid coordinate. No statistically significant differences were found between the two vertical coordinates out to the full 12 hours of each forecast, and only minor, non-statistically significant differences were found at 12 hours, limited to regions above 200 hPa. One possible explanation for these findings is that the hybrid coordinate is sensitive to resolution, with the convective-allowing model being able to better resolve mountainous terrain, thereby minimizing any differences found between the retrospective runs. Given any potential decreased sensitivity at higher resolutions, it is also possible that the retro period of September 2016 did not contain upper-level wind speed values necessary to illustrate a difference in vertical coordinates at 3-km resolution.

Overall, the hybrid vertical coordinate appears to produce the largest impact at upper levels, particularly for wind speed. In addition, these differences are amplified over mountainous terrain, where the largest displacements of the hybrid vertical coordinate from the traditional vertical coordinate are found.

Results also showed that as resolution increases, the expected differences between the two vertical coordinates decrease, likely related to improved terrain in higher-resolution models.

Given these findings and apparent sensitivity to season and resolution, the NCAR node of the DTC will investigate the hybrid vertical coordinate through a number of MMET case studies over a variety of different synoptic conditions. These results are forthcoming and a final report for this testing activity will be available on the DTC website by the end of June 2017. A manuscript highlighting findings for this activity will be submitted before the end of 2017.

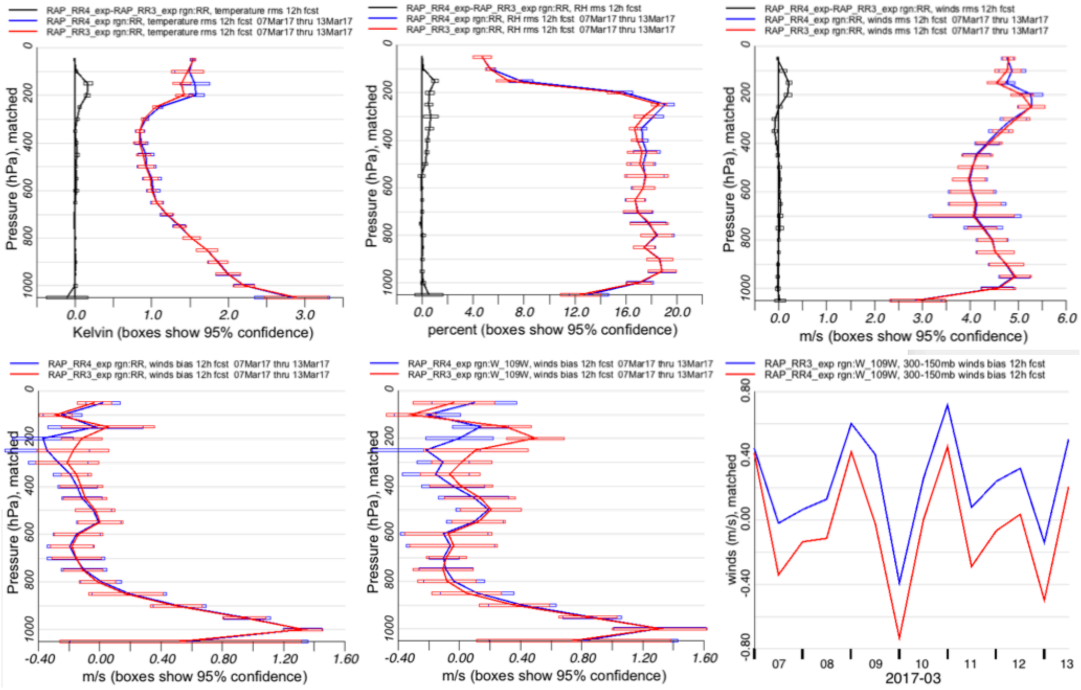


Figure 3.1.3-1. Hybrid coordinate (red) and terrain-following coordinate (blue) results from the RAP retrospective forecasts for 7-13 March 2017. CONUS RMSE is shown for temperature (upper left), relative humidity (upper center), and wind speed (upper right), and bias for wind speed (lower left). Wind speed bias for the western CONUS is shown in the lower center plot, and a time series of averaged wind speed bias for the western CONUS from 300-150 hPa is shown in the lower right.

3.1.4 Addressing uncertainty through stochastic parameter perturbations within the HRRR ensemble

In most existing regional ensemble systems, model-related uncertainty is addressed by using multiple dynamic cores, multiple physics suites, or a combination of these two approaches. While these approaches have demonstrated potential, it is time-consuming and costly to maintain such systems, especially in operations. In order to move toward a more sustainable and unified system, stochastic parameter perturbations within the HRRR physics suite were investigated with a focus on planetary boundary layer (PBL) and Land Surface Model (LSM) processes.

For AOP 2016, the Regional Ensemble team established a test harness using the Rocoto Workflow Management System to conduct functionally similar end-to-end testing of the HRRR model in both a deterministic and ensemble mode. This test harness includes MET verification tasks to evaluate the deterministic and probabilistic forecast output. The inclusion of MET in the workflow provides the opportunity to not only verify the final products, but to also iteratively adjust the ensemble design while examining how probabilistic statistics change when different approaches are utilized.

Due to the high level of complexity of running a frequently updating (hourly), high spatial resolution (3 km), large domain (Contiguous United States - CONUS) ensemble system, extensive high performance computing (HPC) resources were needed to meet this objective. A proposal was written and supercomputing resources were provided through the NCAR Strategic Capability (NSC) project support. This HPC allocation allowed for a more extensive set of tests leading to more robust results than would have otherwise been possible.

As a first step toward designing the test, numerous sensitivity tests of stochastic parameter perturbations (SPP) applied to various parameters within the PBL scheme and in combination with more commonly used stochastic approaches Stochastic Kinetic Energy Backscatter (SKEB) and Stochastic Perturbation of Physics Tendencies (SPPT), were carried out. A variety of tests were also performed involving the LSM scheme. Preliminary results highlighting initial testing of select spatial and temporal de-correlation length scales of soil moisture perturbations were presented as posters at the 2016 Fall AGU meeting

Wolff, J. K., I. Jankov, J. Beck, L. Carson, J. Frimel, M. Harrold, H. Jiang: M. Xu, 2016: Addressing model uncertainty through stochastic parameter perturbations within the High Resolution Rapid Refresh (HRRR) ensemble. Presented at *2016 Fall Meeting, AGU*, San Francisco, CA, December 12-16, 2016.and the 2017 Annual AMS meeting.

And the 2017 Annual AMS meeting

Beck, J., I. Jankov, H. Jiang, J. K. Wolff, M. Harrold, J. Frimel, and L. Carson, 2017: An evaluation of stochastic physics within the High Resolution Rapid Refresh Ensemble (HRRRE) and the impacts of High Performance Computing (HPC). *3rd Symposium on High Performance Computing for Weather, Water, and Climate,* Seattle, WA, January 24-27, 2017.

Based on the outcome of the sensitivity tests, a retrospective experiment was designed. In addition to the stochastic ensemble configuration, the plan included a control ensemble designed to include a variety of PBL and LSM schemes to represent the current state of regional ensemble configurations. The SSEO obtained from NSSL and SPC colleagues for a limited variable dataset served as a second baseline for this test. A deterministic HRRR run without perturbations was also performed to provide a baseline to make sure the SPP perturbations did not introduce an unrealistic bias. Several extended retrospective runs have been completed and the team is in the process of analyzing the results. A report on the findings will be available on the DTC website by the end of June 2017. The Regional Ensemble team will also be preparing a manuscript for submission to Monthly Weather Review, as a follow up to recently published results for testing of SPP in the RAP framework.

3.1.5 WRF testing and evaluation activity

In response to WRF configuration recommendations from NCAR's MMM division, the AF requested the DTC conduct a WRF configuration test with WRF v3.8.1 to provide critical information for a possible operational implementation. In addition to transitioning to a new version of WRF, these configuration updates included migration to the Thompson aerosol-aware microphysics scheme and the updated Rapid Radiative Transfer Model for Global Climate Models (RRTMG) radiation scheme. To address this request, DTC staff conducted an end-to-end test and evaluation activity to assess the sensitivity of replacing the Air Force's current WRF v3.5.1 operational configuration, which was previously tested by the DTC, with a proposed configuration for v3.8.1. For this test, the DTC generated retrospective WRF v3.8.1 forecasts for the same cases used for its prior WRF v3.5.1 testing activity (1 July - 30 September

2011 and 1 January - 31 March 2012) and compared these new retrospective forecasts to the archived v3.5.1 forecasts. Details of the WRF v3.5.1 and v3.8.1 configurations are summarized in Table 3.1.5-1. A project webpage is being finalized and will include pertinent information regarding the test setup and the full suite of results, along with a comprehensive final report (http://www.dtcenter.org/eval/meso_mod/afwa_test/wrf_v3.8.1/index.php).

Table 3.1.5-1: Physics suite configuration settings for version 3.5.1 and 3.8.1, along with other namelist differences between versions.

Physics Suite	v3.5.1	v3.8.1	Other namelist changes (v3.5.1→v3.8.1)
Microphysics	WSM5	Thompson aerosol-aware	• timestep: 90s → 60s • eta_levels • rh2qv_method: 1 → 2 • icloud: 1 → 3 • aer_opt: 1 → 3 • swint_opt: turned on • ysu_topdown_pblmix: turned on • use_aero_icbc: true • diff_opt: 1 → 2 • dampcoef: 0.05 → 0.2 • epssm: 0.1 → 0.5 • scalar_adv_opt: 0 → 1
Radiation (LW/SW)	RRTM/Dudhia	RRTMG/RRTMG	
Surface Layer	Monin-Obukhov similarity theory	Revised MM5 Monin-Obukhov	
Land Surface	Noah	Noah	
PBL	YSU	YSU	
Convection	Kain-Fritsch	Kain-Fritsch	

The testing methodology allowed for pair-wise differences to be computed between v3.5.1 and v3.8.1, including an assessment of both statistically significant (SS) and practically significant (PS) pair-wise differences. Consistent with the significant changes associated with the v3.8.1 configuration, a large number of SS and PS pair-wise differences were observed for both the surface and upper air metrics. Briefly, in terms of BCRMSE, a number of PS differences were seen for 2-m temperature and dew point temperature, generally favoring AFv3.8.1; very few of the differences for 10-m wind speed were PS (not shown). In terms of bias (see Fig. 3.1.5-1), a cold temperature bias at 2 m was generally observed for both configurations; however, AFv3.8.1 was generally the preferred configuration, with a few exceptions. For 2-m dew point temperature, PS differences for bias generally favored AFv3.8.1. While very few PS differences were noted for 10-m wind speed, the SS differences for bias favored AFv3.5.1. In terms of upper air verification results, upper-air temperature bias showed AFv3.8.1 as the preferred configuration over the West during the summer; results were more mixed over the West during the winter and in the East for both seasons. Upper-air dew point temperature bias showed AFv3.5.1 was favored in the West, while in the East, AFv3.8.1 was favored during the summer with more mixed results in the winter. Upper-air wind bias SS differences generally favored AFv3.5.1 in the summer with some PS differences, while winter was mixed with only one instance of PS differences.

When examining the GO Index (Fig. 3.1.5-2), a skill score developed by the AF, AFv3.5.1 was shown as the better performer for both the 00 and 12 UTC initializations during the summer and winter seasons. Based on the overall results for the individual metrics, an investigation into the cause of the AFv3.8.1 degradation in terms of the GO Index was conducted and showed that removal of RMSE for 400 hPa height from the calculation resulted in a reversal of performance, where AFv3.8.1 was preferred for all but the 00 UTC winter aggregation.

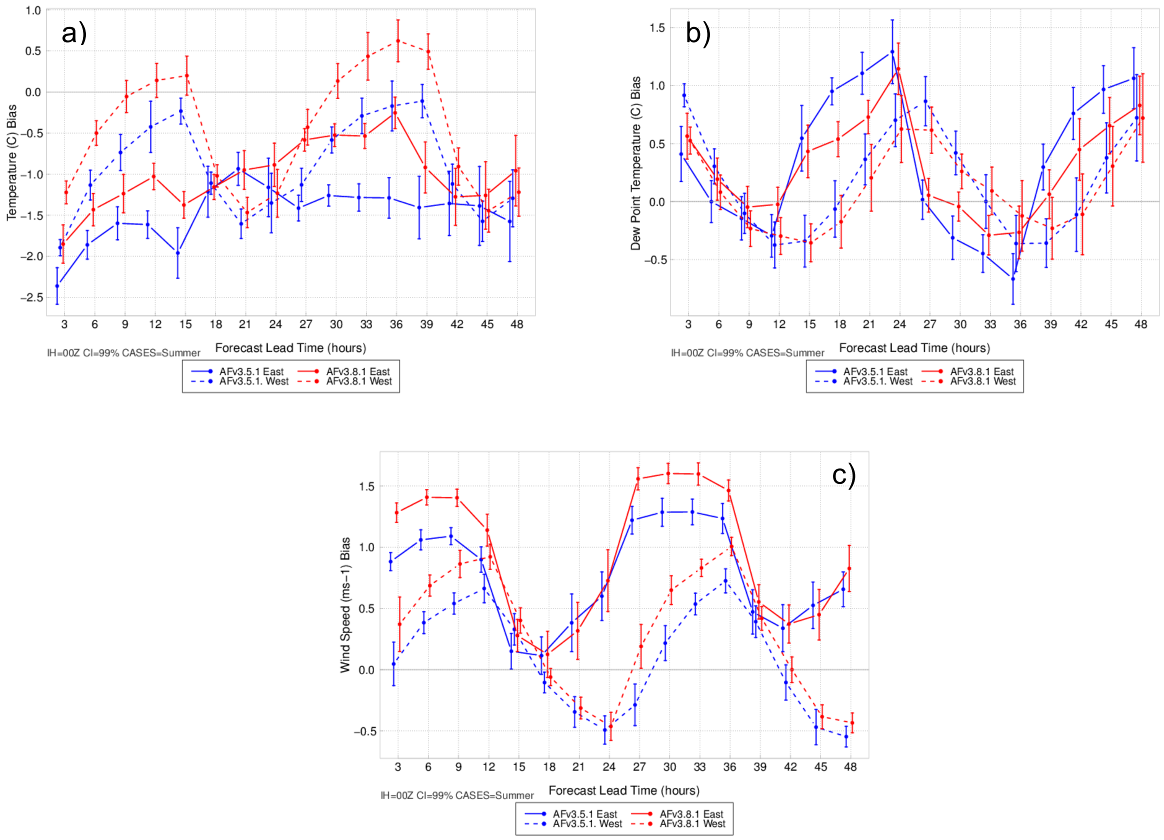


Figure 3.1.5-1. Time series plot of (a) 2-m AGL temperature (°C) (b) 2-m AGL dew point temperature (c) 10-m AGL wind speed median mean error (bias) for the 00 UTC initializations aggregated across the summer cases for the East (solid) and West (dashed) verification domains. AFv3.5.1 is in blue and AFv3.8.1 in red. The vertical bars attached to the median represent the 99% CIs.

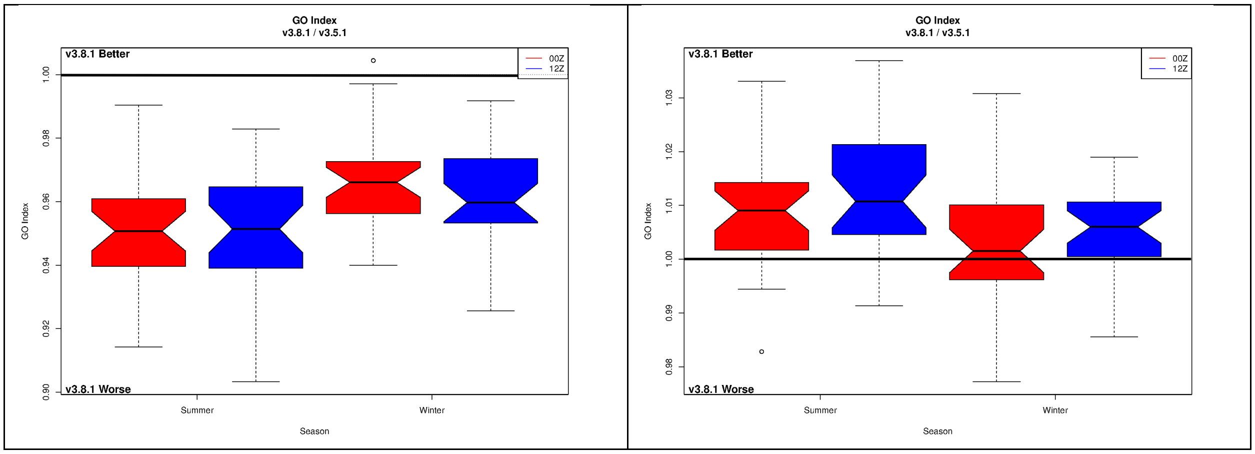


Figure 3.1.5-2. Boxplots of GO Index values aggregated across the summer and winter season, stratified by initialization time, where 00 UTC is in red and 12 UTC is in blue for (a) the standard GO Index calculation and (b) the GO Index calculation without 400hPa height. The median value is the thick black line located at the vertex of the notches, the notches around the median are an approximation of the 95% confidence about the median, the whiskers, denoted by the black, dashed lines, denote the largest values that are not outliers, and the circles represent the outliers.

3.2 Hurricanes

3.2.1 Impact of Thompson microphysics in HWRF

2013 T&E activities revealed that Thompson microphysics in the 2013 version of HWRF produced improvements in track for the Atlantic (AL) basin, but degraded the track forecasts for the Eastern North Pacific (EP) basin. Given the significant upgrades to the operational HWRF system after the 2013 version, performance when using the Thompson microphysics scheme within the 2015 HWRF was re-evaluated. This T&E activity was designed in close collaboration with the EMC HWRF team to inform 2016 pre-implementation testing, where the Thompson and advected Ferrier-Aligo microphysics schemes were both candidates for replacement of the operational Ferrier-Aligo microphysics scheme. The focus of the DTC's evaluation was to determine the impact of replacing the operational Ferrier-Aligo microphysics scheme with the Thompson microphysics scheme. The operational Ferrier-Aligo scheme advects total condensate only, whereas the Thompson scheme advects individual species. The experiments included five storms from the AL basin and eleven storms in the EP basin that occurred during the 2014 and 2015 seasons. Particular emphasis was placed on EP basin storms in response to the 2013 T&E results. Prior to conducting the retrospective test, both the Thompson scheme and the partial cloudiness (PC) scheme within the Rapid Radiative Transfer Model for Global Climate Models (RRTMG) parameterization were modified in an effort to understand and address the cause of the increased track error in the EP basin. These modifications included fall speed changes within the Thompson microphysics scheme and alterations to the RRTMG partial cloudiness scheme to implement a bug fix and change the lower limit of the snow and ice particle size. The majority of the work on this T&E activity was completed prior to this reporting period and was described fully in the AOP 2015 report. Results revealed that the experimental configuration produced improved track and intensity forecasts in the AL basin. However, in the EP basin, the experimental configuration improved the spatial distribution of clouds, but these improvements did not translate into improvements in track and intensity forecasts.

The full report for this Thompson microphysics evaluation is now available on the DTC webpage: http://www.dtcenter.org/eval/hwrf_thomp2016/.

The DTC presented results from this work during the 32nd AMS Hurricane and Tropical Meteorology Conference (April 2016) and the 17th Annual WRF Users' Workshop (June 2016):

Holt, C., M. Biswas, Z. Zhang, S. Trahan, L. Bernardet, G. Thompson, K. Newman: An evaluation of alternative species- advecting microphysics schemes in Hurricane WRF, 32nd Conference on Hurricanes and Tropical Meteorology, 18-22 April 2016, San Juan, PR.

Thompson, G., L. Bernardet, K. Newman, M. Biswas, and C. Holt: Towards improving explicitly resolved and sub-grid-scale clouds in Hurricane WRF. 32nd Conference on Hurricanes and Tropical Meteorology, 18-22 April 2016, San Juan, PR.

Holt, C. M. K. Biswas, Z. Zhang, S. Trahan, L. R. Bernardet, G. Thompson, K. M. Newman: An evaluation of alternative species-advecting microphysics schemes in Hurricane WRF. WRF Users' Workshop, 27-30 June 2016, Boulder, CO.

3.2.2 HWRF physics advancement

For AOP 2016, the DTC partnered with DTC Visitor Program Principle Investigators and subject area experts to help coordinate and test the performance of alternate physics schemes and innovations relative to the current parameterizations within the HWRF physics suite. Physics advancements considered for testing covered radiation, planetary boundary layer (PBL), microphysics and cumulus parameterizations, summarized in Table 3.2.2-1. In addition to coordination and support for HWRF physics developers, the DTC evaluated the code readiness of candidate physics advancements and consulted with the EMC hurricane team on top priorities for HWRF 2017 pre-implementation testing.

Table 3.2.2-1. Candidate physics advancements for testing and evaluation. Innovations selected for testing are indicated in bold.

Physics Developer	Institution	Scheme	Description
M. Iacono, J. Henderson	**AER – DTC Visitor Program PI**	**RRTMG radiation**	**Cloud overlap methodology**
G. Thompson	**NCAR/RAL and DTC**	**RRTMG radiation**	**Modified partial cloudiness scheme**
G. Thompson	NCAR/RAL and DTC	Thompson microphysics	Enhanced scheme based on AOP 2015 T&E results
S. Bao	CCU – DTC Visitor Program PI	advected Ferrier-Aligo microphysics	Investigation of advected microphysics within HWRF
R. Fovell	U. Albany – DTC Visitor Program PI	Yonsei University PBL	HWRF sensitivity to alternative PBL/surface layer schemes
G. Grell, E. Grell	**NOAA/ESRL – NGGPS PI**	**Grell-Freitas cumulus**	**Replacement scheme for scale-aware Simplified Arakawa-Schubert scheme**

Retrospective cases were run for four storms in the AL basin (Edouard, Gonzalo, Matthew, Fiona) and two storms in the EP basin (Patricia, Dolores) that occurred during the 2014-2016 hurricane seasons. A limited number of storms were run due to computational constraints. In an effort to increase storm diversity with limited resources, 126 hour forecasts were run every 18 hours, with 12 hour forecasts for all initializations for cycling purposes. This restriction limits the sample size for forecast lead times beyond 12 hours, decreasing the likelihood of obtaining statistically significant differences. Four parallel experiments were run to test the sensitivity of the three experimental physics configurations. The control (CL) was run using the 2016 operational HWRF default settings. Two RRTMG cloud-radiation experiments were conducted to test the sensitivity of an alternate cloud overlap (CO) methodology and the impact of a modified partial cloudiness scheme. Additionally, a cumulus parameterization replacement test (GF) was run to investigate the impact of the Grell-Freitas (GF) scheme compared to the operational scale-aware simplified Arakawa-Schubert (SAS).

A new cloud overlap technique for the RRTMG radiation parameterization, exponential-random (ER), was tested as a replacement for the default maximum-random assumption. The ER technique alters the overlap of continuous cloud layers to allow for an exponential transition from maximum to random. Other applications have shown this method to be more realistic relative to radar measurements within vertically deep clouds, adding motivation to test within HWRF. The track and intensity results in the AL basin suggested modest reductions in track error, particularly beyond 2 days (Figure 3.2.2-1). However, these differences are not statistically significant (SS). Absolute intensity errors indicate smaller non-SS mean errors out to 30 hours, with mixed impact throughout the intermediate and longer lead times (Figure 3.2.2-1). Both the experimental and control configurations exhibited a negative intensity bias (not shown), with reduced non-SS mean biases for CO beyond 3 days. Due to limited cases in the EP basin, results are pending the addition of storms to the sample.

A partial cloudiness scheme was implemented within the RRTMG radiation scheme for the 2015 operational HWRF system to address excessive short-wave radiation reaching the surface due to transparency of SAS clouds to RRTMG and a lack of stratus representation. Adjustments were made to the relative humidity threshold methodology to further address solar radiation biases. Prior tests of these modifications using WRF-ARW over CONUS resulted in reduced solar radiation biases. These updates were tested to assess whether scheme improvements would translate to the HWRF system. Similar to the CO configuration, track errors in the AL basin had a tendency for slightly reduced mean errors for the PC configuration relative to CL at the longest lead times. Again, these differences were not SS (Figure 3.2.2-1). Absolute intensity errors were smaller for the PC configuration beyond 48 hours, with SS differences favoring PC at the 84-, 90-, and 108-hour lead times. The PC configuration exhibited a negative intensity bias of -5 to -10 kts (not shown), similar to the CL configuration. Mean intensity differences with respect to lead time were mixed and non-SS. As with the CO configuration, results from the EP basin are pending increased sample size.

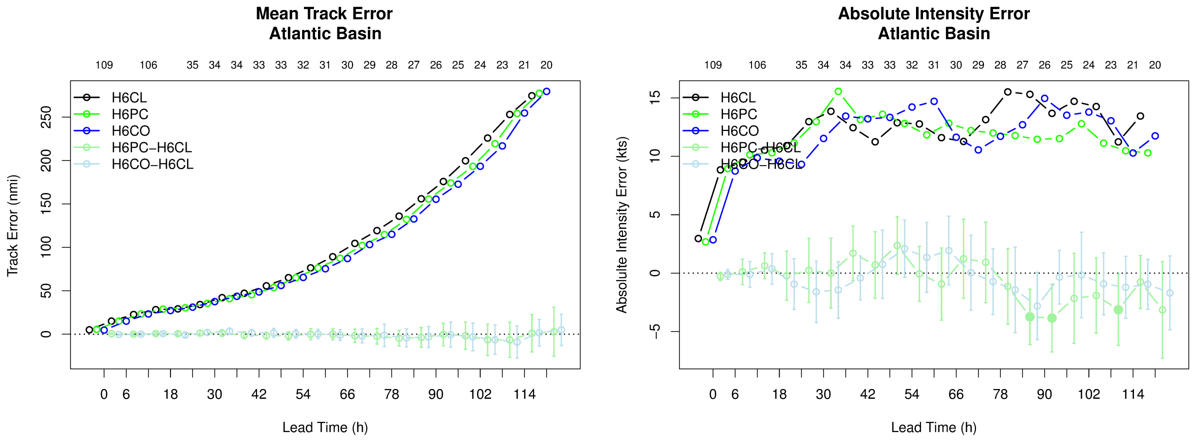


Figure 3.2.2-1. Mean track errors (left) and absolute intensity errors (right) in the AL basin with respect to lead time. The CL (operational) is in black, PC in green, and CO in blue. Pair-wise differences (experiment minus control) are shown in light shades with 95% confidence intervals.

The GF scheme employs an ensemble approach to represent convection, using a collection of parameters and algorithms to represent convective triggers, vertical mass flux, and closures. Additionally, the scheme is scale-aware, making it suitable for HWRF's nested grid configuration. Track errors indicate statistically significant (SS) differences between the GF and CL for some of the early lead times (18-h improvement and 48-h, 54-h degradation), whereas mean differences show non-SS smaller track errors for GF beyond 84-hours (Figure 3.2.2-2). The mean intensity bias for GF appears to be smaller than that for the CL, but the differences are not SS (Figure 3.2.2-2). The operational HWRF is known to under-predict intensity for strong AL storms, which is heavily represented with this sample. The GF configuration tends to reduce this dominate tendency; however, error bars suggest the GF configuration may tend toward over-prediction. In the EP basin, intensity traces suggest a tendency for the GF configuration to better represent rapid intensification (RI) for specific initialization times. Additional RI cases in the EP basin are under investigation.

Additional evaluation is underway to verify HWRF-simulated brightness temperatures (BT) against Geostationary Operational Environmental Satellite (GOES-13, channel 4) BTs. Currently, Hurricane Matthew has been verified on the parent domain (d01) and innermost nest (d03). In addition to the four aforementioned configurations, a supplemental test (PC+CO) was run to incorporate the combined impact of the PC and CO innovations. Figure 3.2.2-3 demonstrates the different attributes of each configuration and domain, shown by probability density functions (PDFs) of the brightness temperature. For the parent domain, the observed BT frequency increases steadily up to 280 K, followed by a sharp increase peaking at 290 K. The GF and CO configurations peak at cooler temperatures, whereas the PC and PC+CO experiments demonstrate an improvement as they shift the PDF towards warmer BTs. For the inner nest, the observational distribution is approximately uniform. Conversely, the model-simulated BT PDFs are clearly bi-modal. Notably, on both domains the model-simulated BTs all underestimate the observed frequency of BT from 235-275 K.

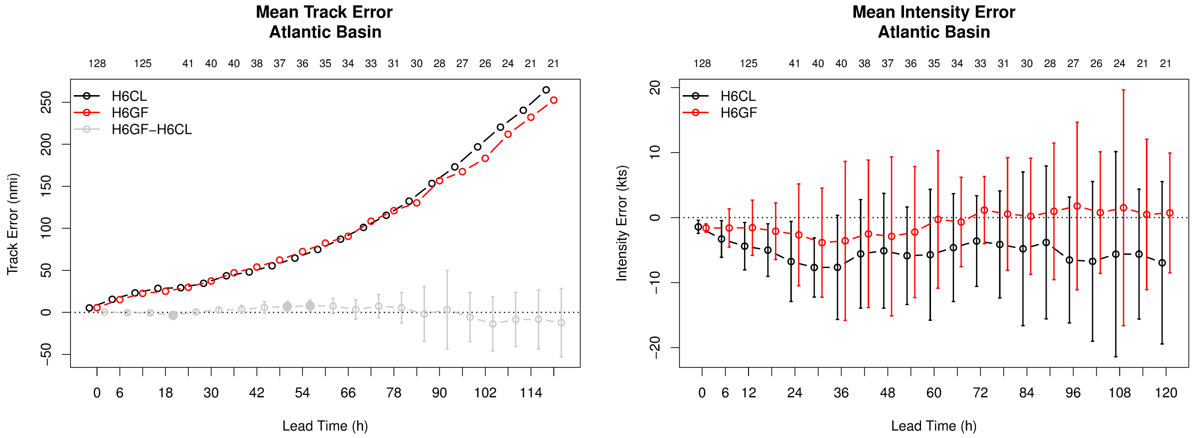


Figure 3.2.2-2. Mean track errors (left) and mean intensity errors (right) in the AL basin with respect to lead time. CL (operational) is in black, GF in red, pair-wise differences (GF-CL) are shown in grey with 95% confidence intervals.

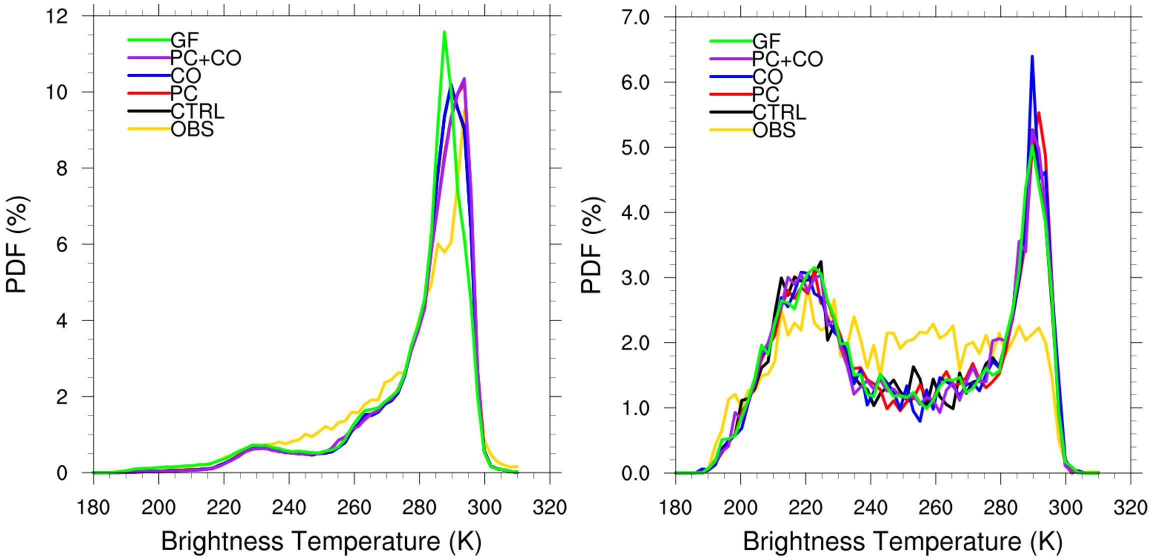


Fig. 3.2.2-3. Probability density functions (PDFs) of the observed (yellow line) and model-simulated brightness temperatures for the GF (green), PC+CO (purple), CO (blue), PC (red), and CL (black) simulations at forecast hour 24 on d01 (left) and d03 (right).

Fractions skill score (FSS) was also computed to show the skill of each configuration (not shown). Results indicate the GF configuration performs the best for all BT thresholds except the warmest (greater than 290 K) on the parent domain, whereas the PC configuration demonstrated the best skill on the inner-nest for BT thresholds greater than 250 K. Notably, none of the configurations worsened HWRF's ability to reproduce the observed BT PDF or substantially degraded the FSS relative to the CL.

The results of the pre-implementation testing were to adopt the PC innovations for the 2017 operational HWRF configuration. The GF configuration continued to demonstrate promise with larger pre-implementation tests performed by EMC, however reproducibility issues when running with a different number of processors caused delays, resulting in a deferred decision on the GF configuration for the 2018 HWRF implementation. Finally, inclusion of the CO innovation into HWRF was tabled until developers implement a namelist option for the new overlap method. Further analysis of all three configurations are underway for inclusion in the project final report, which will be posted on the DTC website (http://www.dtcenter.org/eval/hwrf_GF_PC_CO/). Additional cases are being added to increase the sample, particularly in the EP basin. Furthermore, case studies focusing on particular storms are ongoing to better understand the behavior of each physics innovation.

3.3 Data Assimilation

One paper associated with past DA activities was published in *BAMS*:

H. Shao, J. Derber, X.-Y. Huang, M. Hu, K. Newman, D. Stark, M. Lueken, C. Zhou, L. Nance, Y.-H. Kuo, and B. Brown, 2016: Bridging Research to Operations Transitions: Status and Plans of Community GSI. *Bull. Amer. Meteor. Soc.,* **97**, 1427–1440, doi: 10.1175/BAMS-D-13-00245.1.

3.3.1 Regional Ensemble Based DA T&E

Under AOP 2015 funding, the DTC built an experimental 4D hybrid EnVar data assimilation system in the context of RAP, evaluated the code readiness for regional applications, and provided feedback to developers on its initial assessment of forecast impacts of 4D hybrid EnVar data assimilation. Throughout the process of setting up and testing this capability, the DTC reported bug fixes and missing capabilities to the GSI developers. Replacing the 3D hybrid EnVar data assimilation step in the RAP workflow, 4D hybrid EnVar data assimilation generated minimal impacts on the analyses and forecasts. In addition to reporting on the outcome of its test, the DTC identified areas that need more work to improve the current 4D EnVar capabilities. The final report, which is available on the DTC webpage at http://www.dtcenter.org/eval/data_assim/4denvar/rap_15km/, provides a detailed description of the experiments and discussion of the results.

The DTC also presented the results at two workshops:

H. Shao, M. Hu, C. Zhou, K. Newman, X. Zhang, and C. Holt, 2016: Testing and evaluation of four-dimensional ensemble variational data assimilation for regional weather forecasts. The 7th NOAA Testbed and Providing Grounds Workshop, College Park, Maryland.

C. Zhou, M. Hu, K. Newman, H. Shao, and X. Zhang, *2016:* Initial assessment of the GSI-based 4D hybrid ensemble-variational data assimilation and its application for regional forecasts. The 17th Annual WRF Users' Workshop, Boulder, Colorado.

K. Newman, M. Hu, C. Zhou, and H. Shao, *2016:* Investigating the capability of GSI four-dimensional ensemble variational data assimilation for WRF-ARW applications. The 17th Annual WRF Users' Workshop, Boulder, Colorado.

3.3.2 High Resolution (3km) EnVar Testing and Evaluation

The DA T&E activity for AOP 2016 focused on 4D hybrid EnVar capabilities for high-resolution regional data assimilation in context of the HRRR system. HRRR currently uses a GSI-based 3D hybrid EnVar DA system, which uses the global ensemble (~30 km) for the ensemble background error calculation and one-hour ARW (3-km) forecasts initialized with RAP (13 km) analyses from the previous cycle (so called "pre-forecast") as the DA background. This T&E activity was divided into two focus areas: 1) demonstration of 4D hybrid EnVar system for HRRR, and 2) feasibility and impact assessment for fast cycling of 4D hybrid EnVar.

3.3.2.1 Demonstration of 4D hybrid EnVar system for HRRR

The DTC set up two workflows: 1) a workflow based on operational HRRR system and 2) a workflow that replaced the 3D hybrid EnVar DA system with an experimental 4D hybrid EnVar system. Similar to the 3D hybrid EnVar, the 4D hybrid EnVar was configured to use ensemble input from the global ensemble to compute the flow-dependent error covariance. Note that each 4D analysis requires multiple-time levels (3 time levels for this test) of ensemble and background input files, in contrast to the one ensemble and one background file required for each 3D analysis). The DTC used these workflows to conduct hourly update experiments. Due to computing constraints, the DTC selected a reduced HRRR domain for the test. This test focused on the time period 3-10 September 2016, which included a fast-evolving convective scale event. Figure 3.3.2.1-1 shows the test domain with the analyzed reflectivity from the 4D hybrid EnVar experiment at 1200 UTC on 8 September 2016 (left panel) and the observed radar reflectivity at 1155 UTC on 8 September 2016 (right panel).

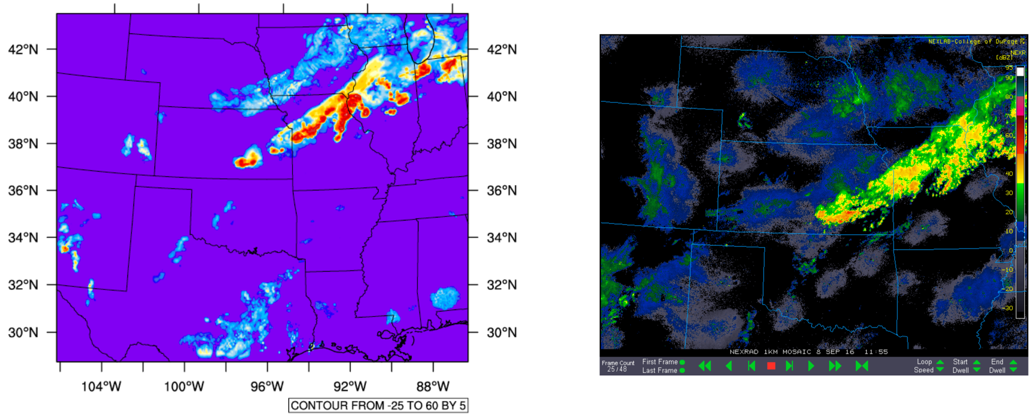


Figure 3.3.2.1-1. The test domain and 4D hybrid EnVar analyzed reflectivity at 1200 UTC on 8 September 2016 (left panel) and the observed radar reflectivity at 1155 UTC on 8 September 2016 (right panel).

Figure 3.3.2.1-2 shows the domain-averaged RMSE for the wind background and analyses from 3D and 4D runs. The results indicate the 4D hybrid EnVar technique improves the fit of both background and analyses to observations. Similar results were also found for humidity and temperature at most of vertical levels. For the forecasts, the impacts of the 4D technique became larger for longer forecast range and resulted in remarkable differences from the 3D results. Figure 3.3.2.1-3 shows the difference between the simulated reflectivity forecasts from the 4D and 3D runs at different lead times. The pattern in the difference fields points to the impacts of the 4D technique on the rain band locations and magnitude at convective scales. At forecast hour 6, the magnitude of the difference is larger such that the reflectivity differences are of the same order of magnitude as the observed reflectivity. The DTC is currently performing more in-depth diagnostics and analysis of verification results from this testing activity. A detailed report is expected to available be the end of May and will be posted on the DTC webpage.

In addition, the DTC examined the impacts of replacing the GFS ensemble with a high-resolution ARW ensemble in the hybrid runs. A case study shows that a 3D hybrid run using the ARW ensemble (not shown here) can result in large differences in 6-h forecasts compared with 3D hybrid runs using the GFS ensemble. This outcome indicates the ensemble representation at convective scales can lead to impacts similar to replacing 3D with 4D data assimilation.

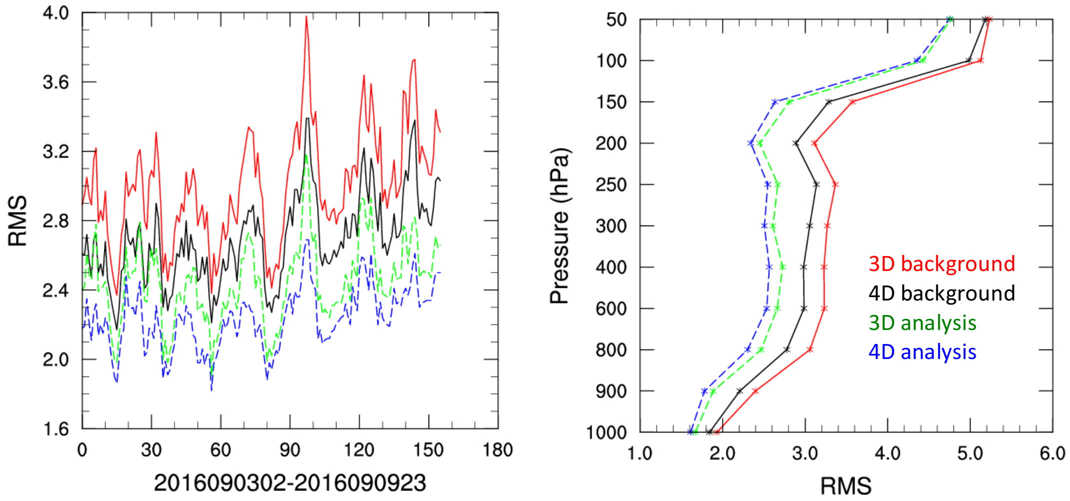


Figure 3.3.2.1-2. The time series (left panel) and vertical profiles of domain averaged (right panel) RMSEs for the wind background and analyses generated from the 3D (red lines for background and green line for analyses) and 4D (black lines for background and blue lines for analyses) hybrid EnVar experiments.

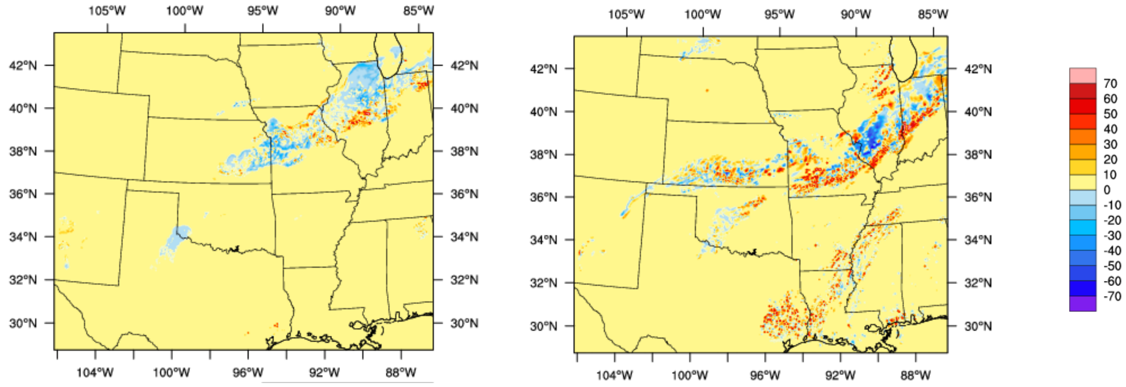


Figure 3.3.2.1-3. Difference between 1-h forecasts (left panel) and 6-h forecasts (right panel) of model simulated reflectivity produced by the 4D and 3D hybrid EnVar experiments.

3.3.2.2 Feasibility and impact assessment for fast cycling of 4D hybrid EnVar

Based on the outcome from the first focus area, the DTC modified the workflow to increase the analysis update frequency and cycled the hybrid data assimilation to every 15 minutes throughout the pre-forecast hour. The DTC also added the capability to GSI to perform sub-hourly (in minutes) analysis updates. The new workflows for the 3D and 4D sub-hourly cycling data assimilation system for HRRR are shown in Figure 3.3.2.2-1. Due to the substantial increase in computational resources associate with going to sub-hourly, the DTC only performed tests for a subset of testing period (8-10 September 2016) used for the first focus area. The results showed that both the 3D and 4D sub-hourly cycling analysis produced a better fit to observations than that of the 3D hourly cycling, followed by neutral to slightly negative impacts at the following hours (figures not shown). The sub-hourly 3D and 4D EnVar configurations may need to be tuned in terms of observation time windows for assimilation, or observation error may need to be adjusted for high-frequency cycling. It is also important to keep in mind that these sub-hourly experiments used a GFS ensemble, which may lead to representation issues for background errors at convective scales. The DTC is performing further diagnostics to compare the 3D and 4D results. A detailed report is expected to available by the end of May and will be posted to the DTC webpage.

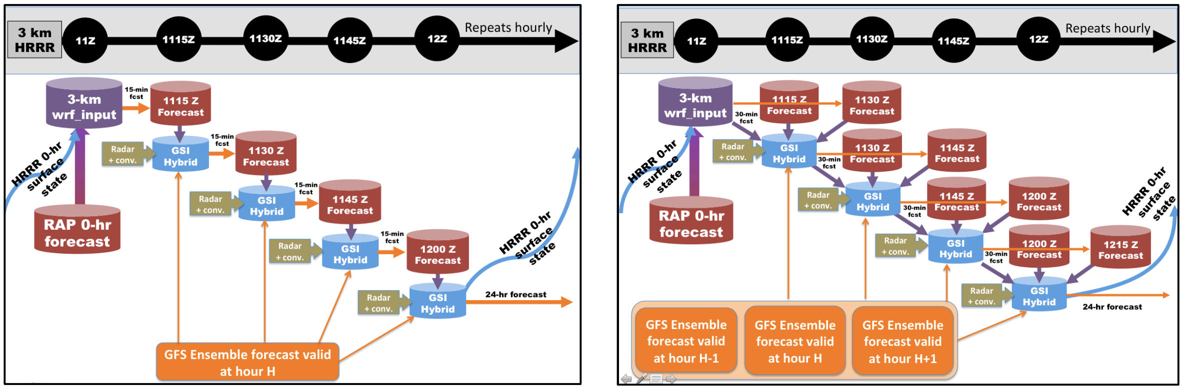


Figure 3.3.2.2-1. Sub-hourly 3D (left panel) and 4D (right panel) hybrid EnVar workflow diagrams showing the evolution of the 15-minute assimilation cycles prior to the 24-hour forecast issued on the hour. Note that multiple GFS ensembles and 30-min HRRR forecasts at each cycle are needed for 4D hybrid EnVar assimilation.

3.4 Global Model Test Bed (GMTB)

3.4.1 Physics Testbed

To facilitate the development of an advanced physics suite for NWS's NGGPS, the DTC is developing a uniform 'test harness' to enable in-depth investigation of various physical parameterizations. This test harness is currently being used by the GMTB for its T&E activities, and has been made available to community scientists working with the GMTB. As an example, developers of the Grell-Freitas cumulus parameterization ran preliminary tests using the GMTB test harness to prepare code for a more comprehensive test, which will be conducted by the GMTB staff. The test harness mimics the logical progression for testing newly developed parameterizations that typically takes place within the scientific community. Components are gradually added as one moves through the hierarchy until the full forecast model complexity is reached. It is designed to complement both the existing testing protocol at EMC and independent testing typically performed by parameterization developers. Figure 3.4.1-1 illustrates the hierarchical tiers of the test harness, represents how the DTC envisions the division of effort (GMTB's likely role denoted by blue) and shows how the harness fits within EMC's existing testing framework.

3.4.1.1 Single Column Model

As part of the GMTB physics test harness, a Single Column Model (SCM) that makes use of the IPD has been developed and lightly tested. The SCM is driven by specifying an initial profile representing the thermodynamic state and accompanying horizontal winds. Changes to the profile caused by large-scale advection are applied through "forcing" terms, which effectively replace the dynamics of a three-dimensional model. A physics suite calculates sub-grid scale processes and changes the profile in concert with the applied forcing.

Design of the SCM focused on community-friendliness by minimizing external dependencies and using community-sanctioned coding practices. Using the SCM only requires the cmake utility for building and the Fortran netCDF library (I/O) to be installed and accessible. Python-based scripts for plotting and analysis, easily set up with an editable configuration file, are included in the distribution. As of now, the GFS physics source code is bundled with the SCM source code as a separate repository. For testing purposes, this code is updated occasionally to work with the top-of-trunk GFS physics code, although it will likely only support specific, tagged versions of this code in the future. The code has been tested on a late model Macintosh, as well as on NOAA's Research and Development (R&D) machine (Theia) and NCAR's Yellowstone. In addition, a GMTB SCM User's Guide and technical documentation was developed with Doxygen and is available on the DTC's website: http://www.dtcenter.org/GMTB/gmtb_scm_doc/.

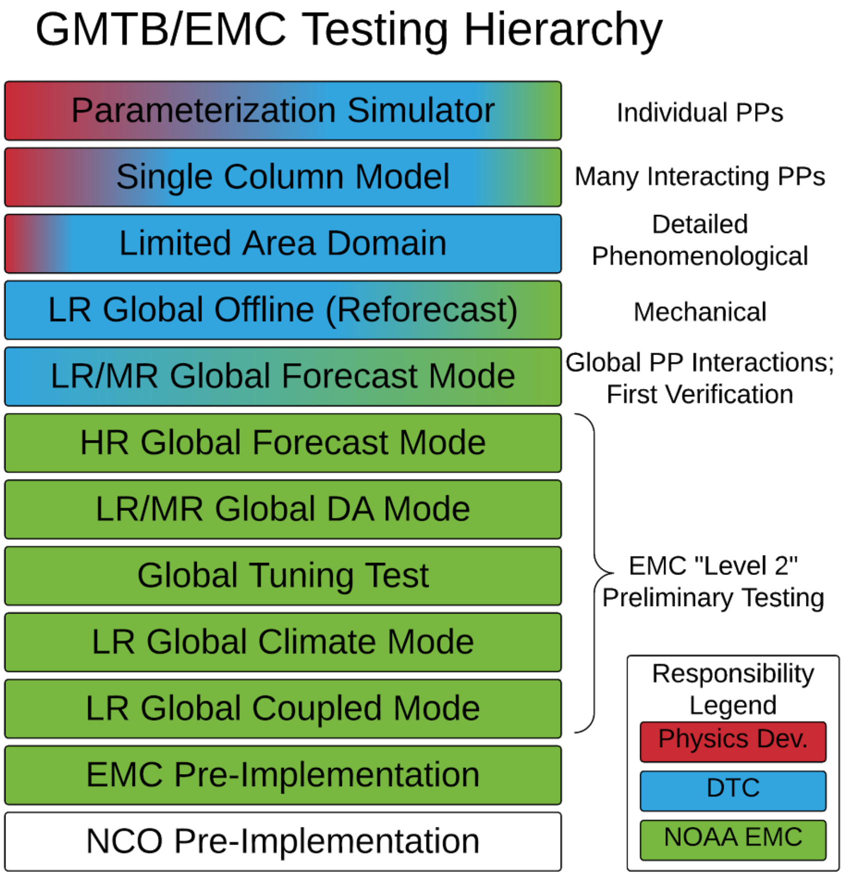


Figure 3.4.1-1. Diagram illustrating the testing hierarchy plan to support physics development for NGGPS. LR indicates low resolution, MR medium resolution, and HR high resolution. Color shading indicates where the different groups are anticipated to focus their efforts (red – physics developers, blue – GMTB task within the DTC, and green – EMC). PP stands for physics parameterization.

So far, the SCM is set up to run individual cases like those supplied by the Global Energy and Water cycle EXchanges (GEWEX) Global Atmospheric System Studies (GASS) program. These cases often derive initial conditions and advective forcing from field campaigns, and are intended to study specific physical phenomena and how they are represented by physics suites. The "catalog" of cases to use with the SCM is a work-in-progress, with one shallow convective case based on the transition from stratocumulus-to-cumulus as observed during the Atlantic Stratocumulus to cumulus Transition EXperiment (ASTEX) field campaign and one deep convective case as observed during the Tropical Warm Pool – International Cloud Experiment (TWP-ICE) field campaign. Both cases are initialized and forced based on observations made during their respective field campaigns. Although both cases use horizontal advective tendencies with prescribed vertical motion, it is possible to configure the SCM to use total advective tendencies and relaxation forcing as described in Randall and Cripe (1999). Going forward, the GMTB will add more cases to the catalog, including cases that will require changes to the underlying GFS physics code (e.g., ability to turn off specified physics schemes within the suite). The cases added by the GMTB can also serve as an example for community members to add cases of interest. In addition, the SCM is set up to easily run using forcing ensembles that can be used to understand a physics suite's response to uncertainty in the forcing.

3.4.1.2 Workflow for Low/Medium Resolution Global Forecast Tests

Building on previous progress, the GMTB successfully established an end-to-end workflow system for running NEMS/Global Spectral Model (GSM) and UPP (strongly leveraging EMC capabilities), as well as running DTC-contributed components. The end-to-end-workflow system reached a mature state that allowed for running a test of the Grell-Freitas (GF) convective parameterization (see description in section 3.4.2).
The DTC-contributed workflow components for creating Python-based forecast plots (e.g. temperature, moisture, convective vs. non-convective precipitation) and verification results (e.g., near-surface, upper-air, and precipitation verification) continued to be upgraded to include additional features and flexibility. A script to plot tropical cyclone tracks for each model initialization was created and added to the automated workflow a number of configuration files for METViewer (a user interface for plotting MET output) were modified and upgraded to generate improved verification plots. This work included adding the 'scorecard' capability to the verification arsenal; the 'scorecard' is a way to summarize patterns in the performance differences between two configurations, including level of significance, for specified metrics, variables, levels, regions, and times. The 'scorecard' was developed by the NCAR Verification team with NGGPS funding and was made available to GMTB for beta-testing ahead of its release to EMC and others.

Work is also underway to expand the testbed capabilities to equip physics developers with a wide range of tools to assess strengths and deficiencies of physics. The capability to produce bias information from GSI diagnostic files, which provide O-B (observation – background) information will soon be available. In addition, the GMTB is collaborating with NGGPS PI Jason Otkin to include synthetic satellite output from UPP to help with evaluating the model's ability to accurately simulate clouds and moisture. The GMTB has been iterating with J. Otkin and his team to use raw model output from the GF test to run through UPP in order to test updates to the radiative transfer model employed by UPP. Tropical cyclogenesis verification is in the process of being implemented. In addition, the capability to perform 6-h global precipitation verification is under development.

The GMTB revitalized its collaborative dialogue with EMC's global team with respect to migrating from their current scripting architecture to a unified global workflow, which includes attending bi-weekly meetings hosted by EMC. The GMTB is actively testing and running the new Rocoto-based workflow (v3.0.0) in preparation for the next testing effort, which will include cycled DA.

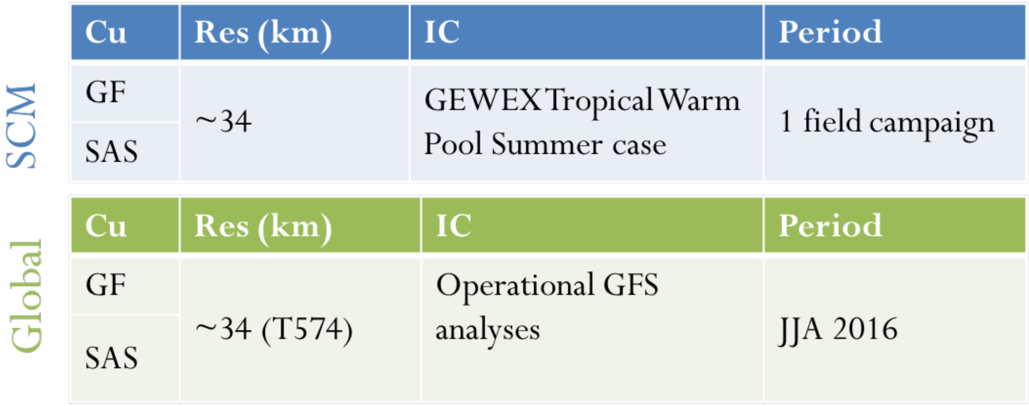
In addition, the team has continued to manage the scripts, and configuration files used in the GMTB workflow through a Git repository on VLab. A substantial effort has also been put forth to document the GSM/UPP workflow, as well as the diagnostic and verification workflow. As the GMTB transitions to EMC's unified global workflow, the documentation will continue to be updated accordingly. Similar to other GMTB documentation, the workflow documentation uses Doxygen.

3.4.2 Grell-Freitas convective parameterization test

The GMTB conducted a test of the Grell-Freitas (Grell and Freitas 2014) convective parameterization to provide input on the establishment and development of an advanced physics suite for NOAA's GFS. This parameterization was selected for testing, through consultation with EMC and the NGGPS Program Office and Physics Team, because of its potential for improving forecasts. It is a state-of-the-art scheme that includes a scale-aware feature, which makes the scheme suitable for use across a wide range of model resolutions. It incorporates an ensemble approach in its representation of convection, which allows perturbation by stochastic fields for deterministic forecasting, as well as ensemble data assimilation. Flux-form vertical tracer transport, wet scavenging, and aerosol awareness are also options in this scheme. The scheme's maturity, its history of operational use at NCEP in the RAP, and the fact that its development is funded by NGGPS also led to choosing the GF scheme for testing.

This test was conducted using GMTB's hierarchical testbed, which currently consists of a SCM and a workflow for running the GFS. In both cases, a control using the GFS operational Simplified Arakawa Schubert (SAS) convective scheme was created (GFS-SAS) and compared against the experimental configuration (GFS-GF) whose physics suite was the same as GFS-SAS except the GF deep and shallow convection schemes were used in place of SAS (Table 3.4.2-1). The entire test results, as well as the final report, are posted on the DTC [website](). Select results are presented below to highlight the key findings.

Table 3.4.2-1. Table summarizing the basic elements of Grell-Freitas convective parameterization test.



	Cu	Res (km)	IC	Period
SCM	GF	~34	GEWEX Tropical Warm Pool Summer case	1 field campaign
	SAS			
	Cu	**Res (km)**	**IC**	**Period**
Global	GF	~34 (T574)	Operational GFS analyses	JJA 2016
	SAS			

For the SCM, a single case based on a deep convection-focused field campaign was used to provide insight into how GFS-SAS performs compared to GFS-GF. The testing paradigm follows the one described in Randall et al. (2003) and Zhang et al. (2016), namely initial conditions and column forcing are derived from observations obtained during Intense Observation Periods. The atmospheric physics suite that makes up the SCM is allowed to respond to the forcing by generating parameterized clouds and precipitation, radiative heating, vertical mixing, etc. Given identical forcing, the GFS-GF suite produced smaller convective tendencies and a much lower convective precipitation ratio than the GFS-SAS suite (Fig. 3.4.2-1). GFS-GF reduced the dry bias in the boundary layer and generally produced a higher cloud fraction during the deep convective period compared to GFS-SAS.

The global forecasts were run at a relatively low resolution (~34 km), in free-forecast mode (no data assimilation or cycling) and without tuning of the physics suite. The operational GFS analyses were used to initialize the retrospective cold start forecasts covering the time period June – August 2016 (JJA). The global model forecasts displayed behavior similar to the SCM, in the sense that GFS-GF had lower convective precipitation (Fig. 3.4.2-2). Differences between precipitation characteristics of GFS-SAS and GFS-GF were noticeable over the CONUS domain. While the 6-h precipitation frequency biases showed a prominent diurnal signal for both configurations at the 0.01" threshold (Fig. 3.4.2-3), as well as 0.1" and 0.25" thresholds (not shown), GFS-GF produced a larger diurnal signal in the frequency bias, with similar magnitudes to GFS-SAS at 18-UTC, but lower bias at 00 UTC.

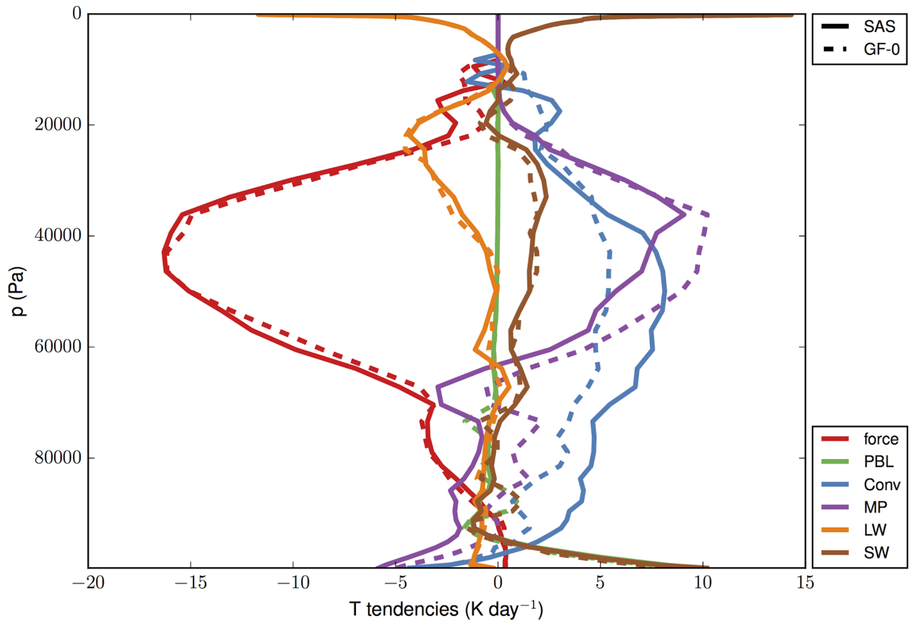


Figure 3.4.2-1: Mean profiles of temperature tendencies (K day^{-1}) for the active phase of the TWP-ICE case. Colors denote forcing (red), PBL scheme (green), convective schemes (deep + shallow, blue), and microphysics scheme (purple). Line types denote the physics suite: GFS-SAS (solid) and GFS-GF (dashed). Tendencies due to longwave and shortwave radiation are in orange and brown, respectively.

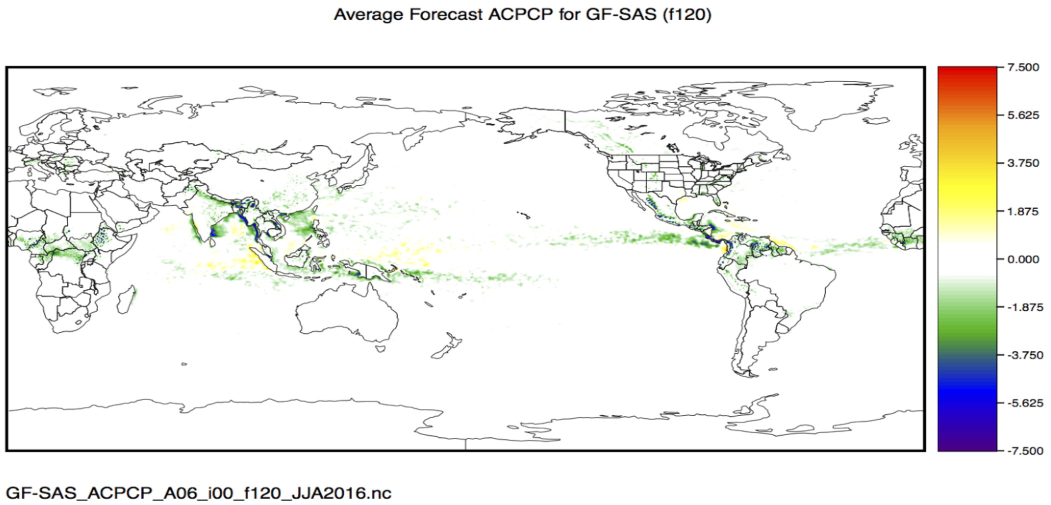


Figure 3.4.2-2: Average 6-h accumulated convective precipitation (mm) over the three-month test period (JJA 2016) at the 120-h forecast lead time for GFS-GF - GFS-SAS.

The global forecasts were run at a relatively low resolution (~34 km), in free-forecast mode (no data assimilation or cycling) and without tuning of the physics suite. The operational GFS analyses were used to initialize the retrospective cold start forecasts covering the time period June – August 2016 (JJA). The global model forecasts displayed behavior similar to the SCM, in the sense that GFS-GF had lower convective precipitation (Fig. 3.4.2-2). Differences between precipitation characteristics of GFS-SAS and GFS-GF were noticeable over the CONUS domain. While the 6-h precipitation frequency biases showed a prominent diurnal signal for both configurations at the 0.01” threshold (Fig. 3.4.2-3), as well as 0.1” and

0.25” thresholds (not shown), GFS-GF produced a larger diurnal signal in the frequency bias, with similar magnitudes to GFS-SAS at 18-UTC, but lower bias at 00 UTC.

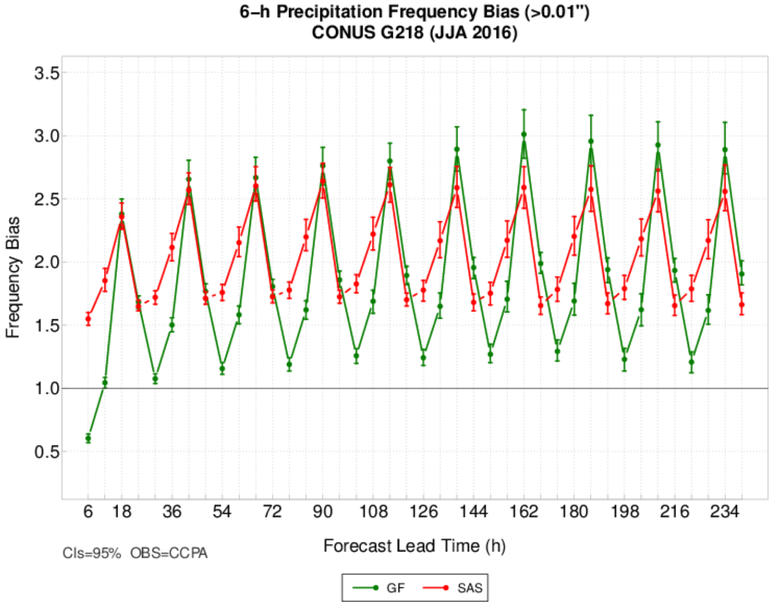
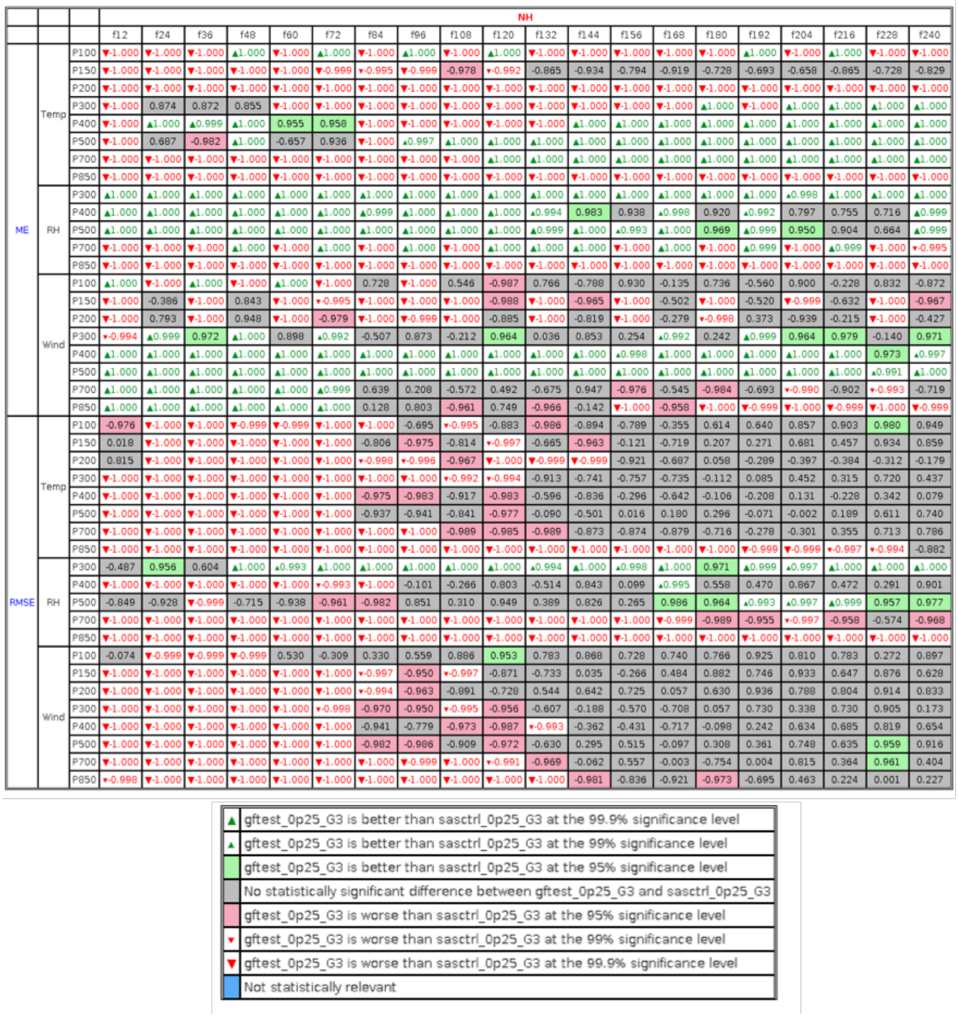


Figure 3.4.2-3: Frequency bias of 6-h accumulated precipitation (in) for GFS-SAS (red) and GFS-GF (green) aggregated over the CONUS domain for the 0.01” threshold as a function of forecast lead time (h) for JJA 2016. The vertical bars surrounding the aggregate value represent the 95% CIs.

With the copious amount of verification results produced for this test, a “scorecard” was a straightforward way to identify patterns in the performance differences between the two configurations, including level of significance, for specified metrics, variables, levels, regions, and times. The scorecard for global sub-regions helped identify that upper-air wind speed had the fewest statistically significant (SS) differences compared to other variables, as shown in Fig. 3.4.2-4 for Northern Hemisphere (NH). The NH clearly signaled GFS-SAS as performing better for the earlier part of the forecast period. However, GFS-GF was the better performer for temperature bias later in the forecast period. Follow-up diagnostics indicated the improved performance of GFS-GF for this metric was related to the GFS-SAS warming up progressively over the NH throughout the forecast period.

Tropical cyclone track errors averaged over all the basins were similar for both model configurations (Fig. 3.4.2-5a). While accuracy in TC intensity forecasts (Fig. 3.4.2-5b) is not expected of a model run at this coarse resolution, it is interesting to note that storms produced by GFS-SAS are more intense (not shown) and have less absolute intensity error than those produced by GFS-GF.

The GMTB’s testing and evaluation of the GF cumulus parameterization illustrated the complexity -- yet scientific usefulness -- of connecting a new scheme to the GSM. The success of this test was heavily dependent on interactions among and investment by the GMTB, the physics developer, and EMC’s Global Team. The close collaboration and iteration with the developer helped ensure the GMTB properly connected the GF parameterization within the GSM code. In addition, the collaboration with the EMC Global Team was essential to the GF test. Testing with the SCM resulted in several key findings, with one finding aligning with results from the 3-D global forecasts. While not all results from the SCM could be translated to the full global forecasts, this test highlighted the utility and process of the hierarchical testing. Moving forward, it will be necessary to further engage with EMC and continue to get feedback regarding desired verification methods and displays, which would then be prioritized for future implementation in the testbed.



			NH																			
			f12	f24	f36	f48	f60	f72	f84	f96	f108	f120	f132	f144	f156	f168	f180	f192	f204	f216	f228	f240
ME	Temp	P100	▼-1.000	▼-1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▼-1.000
		P150	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-0.999	▾-0.995	▼-0.999	-0.978	▾-0.992	-0.865	-0.934	-0.794	-0.919	-0.728	-0.693	-0.658	-0.865	-0.728	-0.829
		P200	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000
		P300	▼-1.000	0.874	0.872	0.855	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▲1.000	▲1.000	▲1.000
		P400	▼-1.000	▲1.000	▲0.999	▲1.000	0.955	0.958	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000
		P500	▼-1.000	0.687	-0.982	▲1.000	-0.657	0.936	▼-1.000	▴0.997	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000
		P700	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000
		P850	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000
	RH	P300	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▴0.998	▲1.000	▲1.000	▲1.000
		P400	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲0.999	▲1.000	▲1.000	▲1.000	▴0.994	0.983	0.938	▴0.998	0.920	▴0.992	0.797	0.755	0.716	▲0.999
		P500	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲0.999	▲1.000	▴0.993	▲1.000	0.969	▲0.999	0.950	0.904	0.664	▲0.999
		P700	▼-1.000	▼-1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▲1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	▲0.999	▼-1.000	▲0.999	▼-1.000	▾-0.995
		P850	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000
	Wind	P100	▲1.000	▼-1.000	▲1.000	▼-1.000	▲1.000	▼-1.000	0.728	▼-1.000	0.546	-0.987	0.766	-0.788	0.930	-0.135	0.736	-0.560	0.900	-0.228	0.832	-0.872
		P150	▼-1.000	-0.386	▼-1.000	0.843	▼-1.000	▾-0.995	▼-1.000	▼-1.000	▼-1.000	-0.988	▼-1.000	-0.965	▼-1.000	-0.502	▼-1.000	-0.520	▼-0.999	-0.632	▼-1.000	-0.967
		P200	▼-1.000	0.793	▼-1.000	0.948	▼-1.000	-0.979	▼-1.000	▼-0.999	▼-1.000	-0.885	▼-1.000	-0.819	▼-1.000	-0.279	▾-0.998	0.373	-0.939	-0.215	▼-1.000	-0.427
		P300	▾-0.994	▲0.999	0.972	▲1.000	0.898	▴0.992	-0.507	0.873	-0.212	0.964	0.036	0.853	0.254	▴0.992	0.242	▲0.999	0.964	0.979	-0.140	0.971
		P400	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▴0.998	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	0.973	▴0.997
		P500	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▴0.991	▲1.000
		P700	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲0.999	0.639	0.208	-0.572	0.492	-0.675	0.947	-0.976	-0.545	-0.984	-0.693	▾-0.990	-0.902	▾-0.993	-0.719
		P850	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	0.128	0.803	-0.961	0.749	-0.966	-0.142	▼-1.000	-0.958	▼-1.000	▼-0.999	▼-1.000	▼-0.999	▼-1.000	▼-0.999
RMSE	Temp	P100	-0.976	▼-1.000	▼-1.000	▼-0.999	▼-0.999	▼-1.000	▼-1.000	-0.695	▾-0.995	-0.883	-0.986	-0.894	-0.789	-0.355	0.614	0.640	0.857	0.903	0.980	0.949
		P150	0.018	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.806	-0.975	-0.814	▾-0.997	-0.665	-0.963	-0.121	-0.719	0.207	0.271	0.681	0.457	0.934	0.859
		P200	0.815	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.998	▾-0.996	-0.967	▼-1.000	▼-0.999	▼-0.999	-0.921	-0.687	0.058	-0.289	-0.397	-0.384	-0.312	-0.179
		P300	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.992	▾-0.994	-0.913	-0.741	-0.757	-0.735	-0.112	0.085	0.452	0.315	0.720	0.437
		P400	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.975	-0.983	-0.917	-0.983	-0.596	-0.836	-0.296	-0.642	-0.106	-0.208	0.131	-0.228	0.342	0.079
		P500	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.937	-0.941	-0.841	-0.977	-0.090	-0.501	0.016	0.180	0.296	-0.071	-0.002	0.189	0.611	0.740
		P700	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.989	-0.985	-0.989	-0.873	-0.874	-0.879	-0.716	-0.278	-0.301	0.355	0.713	0.786
		P850	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-0.999	▼-0.999	▾-0.997	▾-0.994	-0.882
	RH	P300	-0.487	0.956	0.604	▲1.000	▴0.993	▲1.000	▲1.000	▲1.000	▲1.000	▲1.000	▴0.994	▲1.000	▴0.998	▲1.000	0.971	▲0.999	▴0.997	▲1.000	▲1.000	▲1.000
		P400	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.993	▼-1.000	-0.101	-0.266	0.803	-0.514	0.843	0.099	▴0.995	0.558	0.470	0.867	0.472	0.291	0.901
		P500	-0.849	-0.928	▼-0.999	-0.715	-0.938	-0.961	-0.982	0.851	0.310	0.949	0.389	0.826	0.265	0.986	0.964	▴0.993	▴0.997	▲0.999	0.957	0.977
		P700	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-0.999	-0.989	-0.955	▾-0.997	-0.958	-0.574	-0.968
		P850	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000
	Wind	P100	-0.074	▼-0.999	▼-0.999	▼-0.999	0.530	-0.309	0.330	0.559	0.886	0.953	0.783	0.868	0.728	0.740	0.766	0.925	0.810	0.783	0.272	0.897
		P150	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.997	-0.950	▾-0.997	-0.871	-0.733	0.035	-0.266	0.484	0.882	0.746	0.933	0.647	0.876	0.628
		P200	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.994	-0.963	-0.891	-0.728	0.544	0.642	0.725	0.057	0.630	0.936	0.788	0.804	0.914	0.833
		P300	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▾-0.998	-0.970	-0.950	▾-0.995	-0.956	-0.607	-0.188	-0.570	-0.708	0.057	0.730	0.338	0.730	0.905	0.173
		P400	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.941	-0.779	-0.973	-0.987	▾-0.993	-0.362	-0.431	-0.717	-0.098	0.242	0.634	0.685	0.819	0.654
		P500	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.982	-0.986	-0.909	-0.972	-0.630	0.295	0.515	-0.097	0.308	0.361	0.748	0.635	0.959	0.916
		P700	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-0.999	▼-1.000	▾-0.991	-0.969	-0.062	0.557	-0.003	-0.754	0.004	0.815	0.364	0.961	0.404
		P850	▾-0.998	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	▼-1.000	-0.981	-0.836	-0.921	-0.973	-0.695	0.463	0.224	0.001	0.227

▲	gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99.9% significance level
▴	gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 99% significance level
🟩	gftest_0p25_G3 is better than sasctrl_0p25_G3 at the 95% significance level
⬜	No statistically significant difference between gftest_0p25_G3 and sasctrl_0p25_G3
🟥	gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 95% significance level
▾	gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99% significance level
▼	gftest_0p25_G3 is worse than sasctrl_0p25_G3 at the 99.9% significance level
🟦	Not statistically relevant

Figure 3.4.2-4: Scorecard documenting the relative performance of GFS-SAS and GFS-GF over the NH for mean bias and RMSE for temperature, relative humidity, and wind speed by forecast lead time and vertical level for JJA 2016. Green (red) shading indicates GFS-GF (GFS-SAS) was better than GFS-SAS (GFS-GF) at the 95% significance level. Small green (red) arrows indicate GFS-GF (GFS-SAS) was better than GFS-SAS (GFS-GF) at the 99% significance level. Large green (red) arrows indicate GFS-GF (GFS-SAS) was better than GFS-SAS (GFS-GF) at the 99.9% significance level. Grey shading indicates no statistically significant differences between GFS-SAS and GFS-GF.

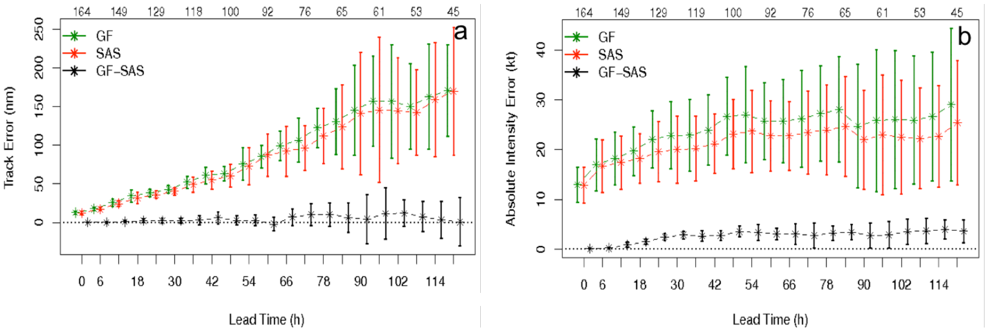


Figure 3.4-6: (a) Mean track errors (nm), and (b) mean absolute intensity errors (kt) with 95% confidence intervals with respect to lead time (h) for GFS-SAS (red), GFS-GF (green) and their pairwise differences (black) in all basins for JJA 2016.

3.4.3 CICE test

Over the past year, the GMTB completed a test of The Los Alamos sea ice model (CICE), which stemmed from the NGGPS Sea Ice Modeling Workshop organized by GMTB in February 2016. As described in the workshop's final report (link), participants recommended the tentative adoption of CICE, pending follow-up testing, and addressing concerns raised regarding model governance and differences in staggering between the grids used in NCEP's Unified Global Coupled System (UGCS) ocean models and CICE.

The test plan, devised jointly by GMTB, EMC, the workshop committee and interested workshop participants, included experiments over a one-year period, with CICE being run in a standalone framework, forced by atmospheric and oceanic fields from the NCEP operational Climate Forecast System version 2 (CFSv2). GMTB conducted the 30-day forecast runs, and evaluation was performed jointly by GMTB, EMC, and ESRL.

The experiment had three phases, as described in Table 3.4.3-1, with varying model resolution and approaches to initializing the atmosphere and ocean fields, as well as constraining the Sea Surface Temperature (SST).

Table 3.4.3-1. CICE model resolution at the pole (km), dataset for atmospheric initialization and forcing, dataset for ocean initialization, and method/dataset for ocean forcing.

		CICE	**Atmos Init and Forcing**	**Ocean Init**	**Ocean Forcing**
Phase	**1**	30 km	CFSv2 1.0^{0}	CFSv2 1.0^{0}	CFSv2 1^{0} 6-hourly forcing
	2	15 km	CFSv2 0.2^{0}	CFSv2 0.5^{0}	CFSv2 0.5^{0} 6-hourly forcing
	3	15 km	CFSv2 0.2^{0}	CFSv2 0.5^{0}	Freely evolving

Generally speaking, the Phase-1 and Phase-2 forecasts at the end of the month-long integrations are in good agreement with the CFSv2 initial conditions at the beginning of the next month, indicating a very good forecast for the end of the month. The major exceptions are in the summer seasons of both hemispheres, where excessive melting occurs (Fig. 3.4.3-1). Follow-up diagnostics indicated that basal melting, caused by warm SSTs, was the cause of the excessive melting.

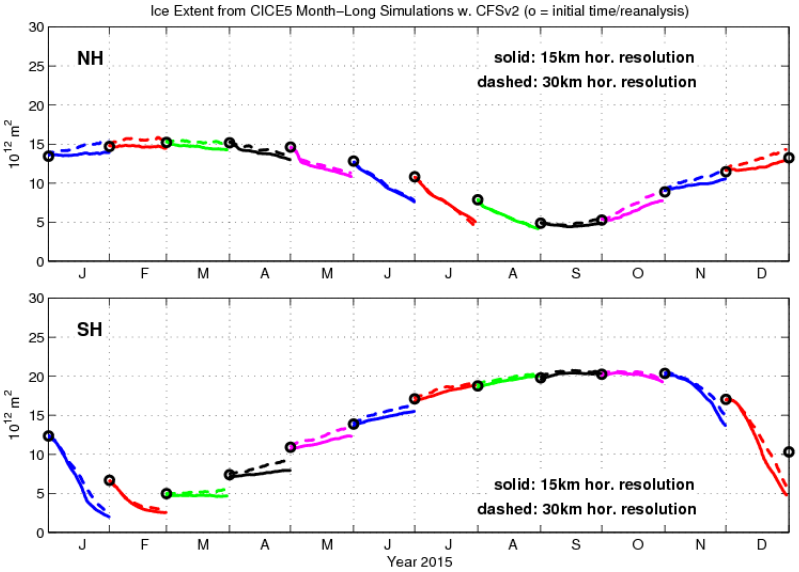


Figure 3.4.3-1. Ice extent (10^{12} m^2) during month-long integrations for Phase 1 (dashed) and 2 (solid) in the Northern (top) and Southern hemisphere (bottom). The circles indicate the initial conditions from CFSv2 at the beginning of each month.

When the SST from CFSv2 forcing was replaced by freely-evolving SST from the CICE internal mixed-layer ocean model, basal melting was reduced and prediction of ice extent at the end of summer was substantially improved (Fig. 3.4.3-2). Comprehensive results from this test can be found in the report (link) and on the test website (link).

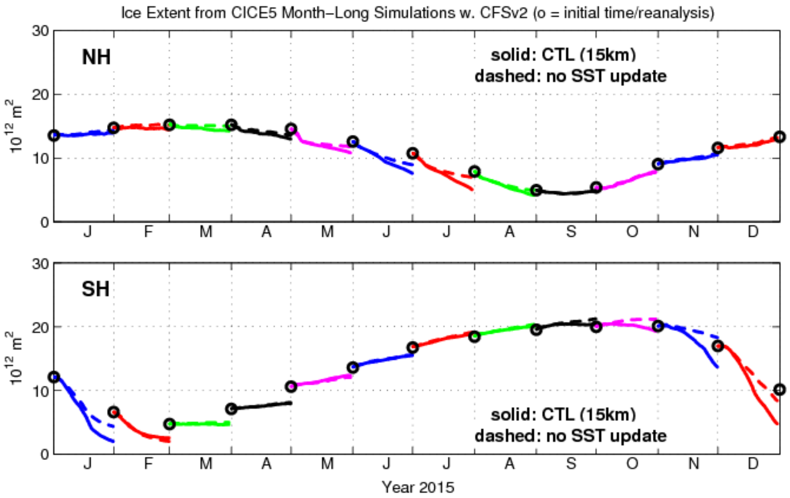


Figure 3.4.3-2. Same as Fig. 3.4.3-1, except for Phase 2 (solid) and Phase 3 (dashed).

3.5 Cloud Verification

At the request of the AF, the DTC investigated approaches for evaluating cloud predictions from NWP models and statistical predictions of cloud properties. This evaluation included numerous NWP model forecasts utilized by the AF. Raw model output from the ARW limited area model, over the Northern Hemisphere, and the AF's Global Air Land Weather Exploitation Model (GALWEM – referred to as UM in figures) global model forecasts were provided by the AF. The DTC collected Global Forecast System (GFS) 0.5-degree model forecasts via NCEP archives (www.nomads.ncdc.noaa.gov). In addition to raw model output (RAW), the AF provided Diagnostic Cloud Forecast (DCF) model output from each of the aforementioned models. The DCF system is a statistical post-processing model that can be appended to any NWP model, producing three-dimensional cloud forecasts for global and regional domains. In the discussion below, evaluations of the raw model output will refer to the model identifier combined with "RAW" (i.e., UMRAW, GFSRAW) and evaluations of DCF products based on these models will refer to the model identifier combined with "DCF" (i.e., UMDCF, GFSDCF). Additionally, the AF provided Advect Cloud (ADVCLD) model output; this system provides cloud cover predictions out to nine hours. All forecasts were collected to maximize the number of concurrent predictions.

The World-Wide Merged Cloud Analysis (WWMCA) product is a satellite-based global analysis developed at the AF. These analyses are run every 30 min. The WWMCA-R dataset, generated every hour, is a reanalysis of the WWMCA product created in post-processing that includes more input data; it is designed to address data latency limitations of the real-time WWMCA product. The WWMCA and WWMCA-R datasets were delivered for both the Northern and Southern Hemispheres, for each initialization time for the four weeks of forecast data delivered.

The U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program has several sites with highly instrumented ground stations directed at cloud and radiation research. Data were collected from the Southern Great Plains (SGP) facility in Lamont, Oklahoma, in order to obtain active sensor cloud data. Value Added Products (VAP) were identified as an ideal dataset due to their quality controlled and post processed nature. The ARM datasets include ceilometer, lidar, total sky imager, and radiative flux analysis products for establishing cloud fraction. Limited data availability for the provided forecast periods precluded in-depth examination of this data.

The DTC's MET (http://www.dtcenter.org/met/users/) software package was used for the verification analyses in this study. MET includes a wide variety of tools for verification and many options for interpolation, matching, and re-gridding of forecasts and observation fields. The MET verification tools comprise both traditional and spatial methods, including traditional approaches for categorical, continuous and probabilistic forecasts, and spatial methods such as neighborhood methods and the Method for Object-based Diagnostic Evaluation (MODE).

Initial investigation of cloud verification metrics was undertaken during AOP 2015 (see http://www.dtcenter.org/verification/reports/DTC_AF_Cloud_Verification_Report_FY15.pdf and http://www.dtcenter.org/verification/reports/Report_on_New_Cloud_Verification_Methods_Final_Aug16.pdf). For AOP 2016, the DTC continued to explore the total cloud fraction (TCDC) field and expanded the demonstration of several methods to the global scale, exploring several thresholds and convolution radii to identify the best configuration for MODE, obtaining additional summer data and adding measures to MET. New methods added to MET as part of this project (day/night mask, land/sea mask, satellite grouping masks and regions defined by latitude bands) were used during the evaluation. As part of defining an optimal MODE configuration, multiple MODE convolution radii and threshold pairs were investigated using the new "quilt" configuration method introduced in METv5.2. Scores were computed for cloudy (TCDC > 80%) and clear (TCDC < 80%) conditions.

Box plots shown in Figure 3.4-1 indicate the skill of all TCDC forecasts (ADVCLD, GFSDCF, GFSRAW, UMDCF, UMRAW) drop off markedly within the first six hours and then generally plateau out to 78 hours (beyond the range of these plots). The exception is the continued reduction in skill for the ADVCLD (red) for clear conditions (right panel). Additionally, the RAW fields tend to have a very low Probability of Detection - Yes (PODY) for clear conditions. Also included in Fig. 3.4-1 is a comparison between WWMCA and WWMCA-R. Interestingly, the score of approximately 0.9 for WWMCA (purple) when compared to WWMCA-R indicates there is approximately a 10-15% change in the cloud field once the re-analysis is performed. Finally, even though there are only four weeks in the sample, the height of the boxes and lack of outlies (circles) indicate the distribution of scores are fairly small. This narrow distribution is due to the large quantity of grid points over the global domain.

Performance diagrams provide information about a set of related categorical scores to provide a complete picture of model performance in one diagram (Roebber, 2009). An example of a performance diagram is shown in Figure 3.4-2. In this diagram, the upper right corner indicates a perfect forecast, whereas the lower left corner indicates a forecast with no skill. Performance diagrams plot PODY versus success ratio (1- False Alarm Ratio). The straight lines emanating from the origin are Frequency Bias (1 is perfect, while values greater than 1 indicate over-forecasting and values less than 1 indicate under-forecasting) and the curved lines are the Critical Success Index (CSI) values. Figure 3.4-2 indicates the scores for clear conditions from the RAW fields tend to be lower than those for their companion DCF fields. Scores for high latitude regions (above 50 degrees) are shown as circles; for mid-latitudes (30-50 degrees) are triangles; and tropical regions (0-30 degrees) are squares. In the Northern Hemisphere (left), the scores increase from poles to tropics. Interestingly, in the Southern Hemisphere (right), scores are lowest in the mid-latitudes. WWMCA is an analysis and ADVCLD only provides forecasts to 9 hours. Hence, neither are included in these performance diagrams for the 24 hour forecasts.

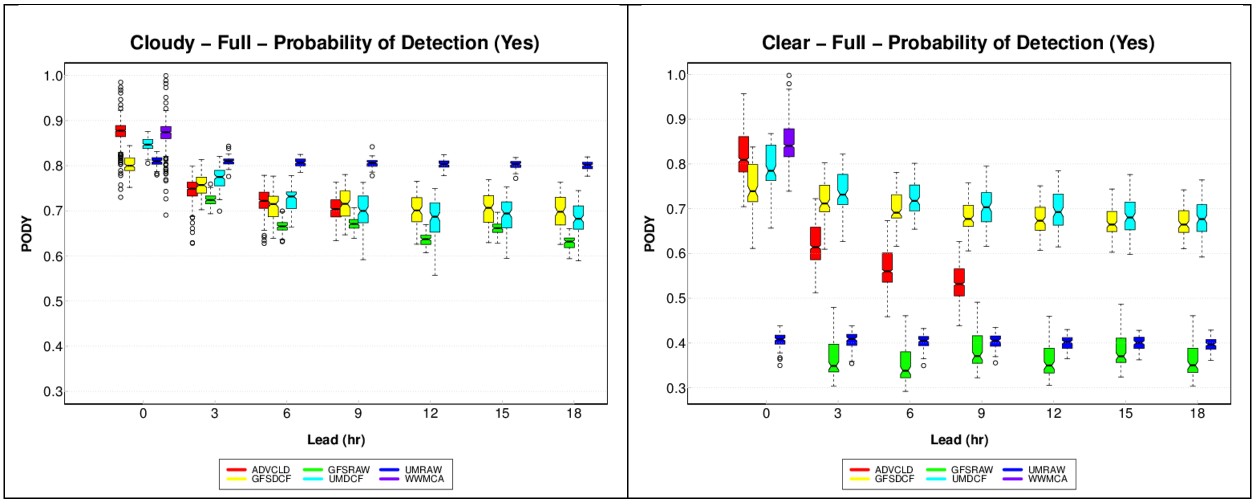


Figure 3.5-1. Box plots of Probability of Detection (PODY) computed against WWMCA-R for ADVCLD (red), GFSDCF (yellow), GFSRAW (green), UMDCF (light blue), UMRAW (dark blue) and WWMCA operational analysis (purple). Scores shown are for cloudy (left) and clear (right) conditions.

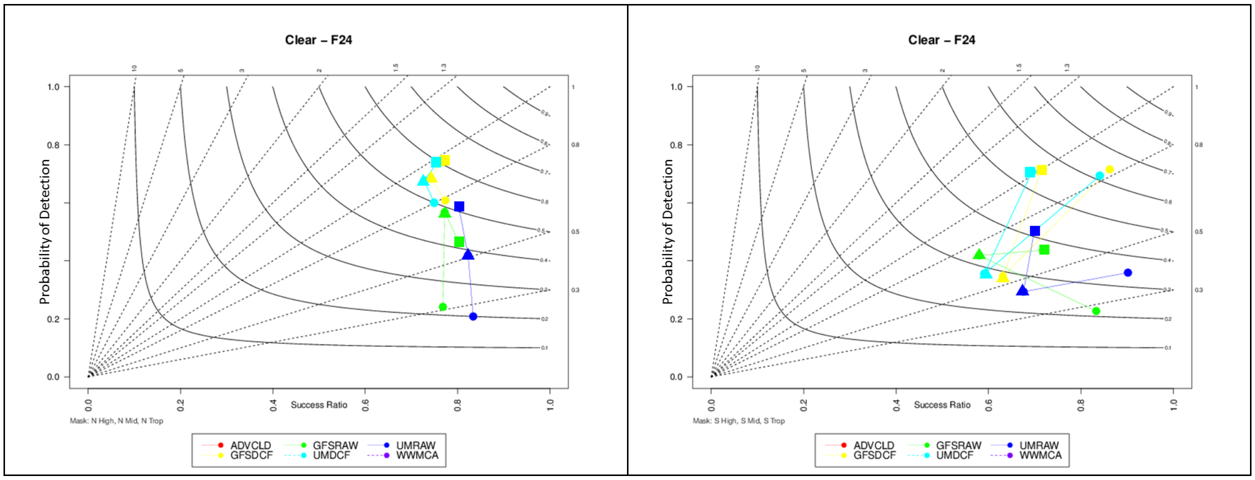


Figure 3.5-2. Performance diagrams for categorical results stratified by regions in the Northern (left) and Southern (right) hemisphere. High latitudes are shown as circles, mid-latitudes as triangles and tropics as squares. Diagrams are for clear conditions. Color scheme is same as that in Fig. 3.5-1.

Several MODE configurations were tested on the North American G212 domain during AOP 2015. The resulting configurations were the starting point for the MODE testing performed during AOP 2016. It was found that the regional configurations were not optimal for use over the global domain for several reasons, including too fine a convolution radius resulted in objects that were more structured than necessary. Also, the centroid distance and area ratio settings needed to be modified to decrease their influence on matching and merging. Figure 3.4-3 shows the forecast (upper right) and analysis (upper left) for one case, along with the resulting objects for a cloudy (lower left) and clear (lower right) configuration. The configuration using a convolution radii of 30 gridpoints and thresholds of >=80 and <=20, for cloudy and clear respectively, resulted in the object representation to most closely match subject impressions of what defines the cloudy or clear areas. This configuration was applied to all 4 seasonal cases and will be summarized in the final report, which will be posted on the DTC website.

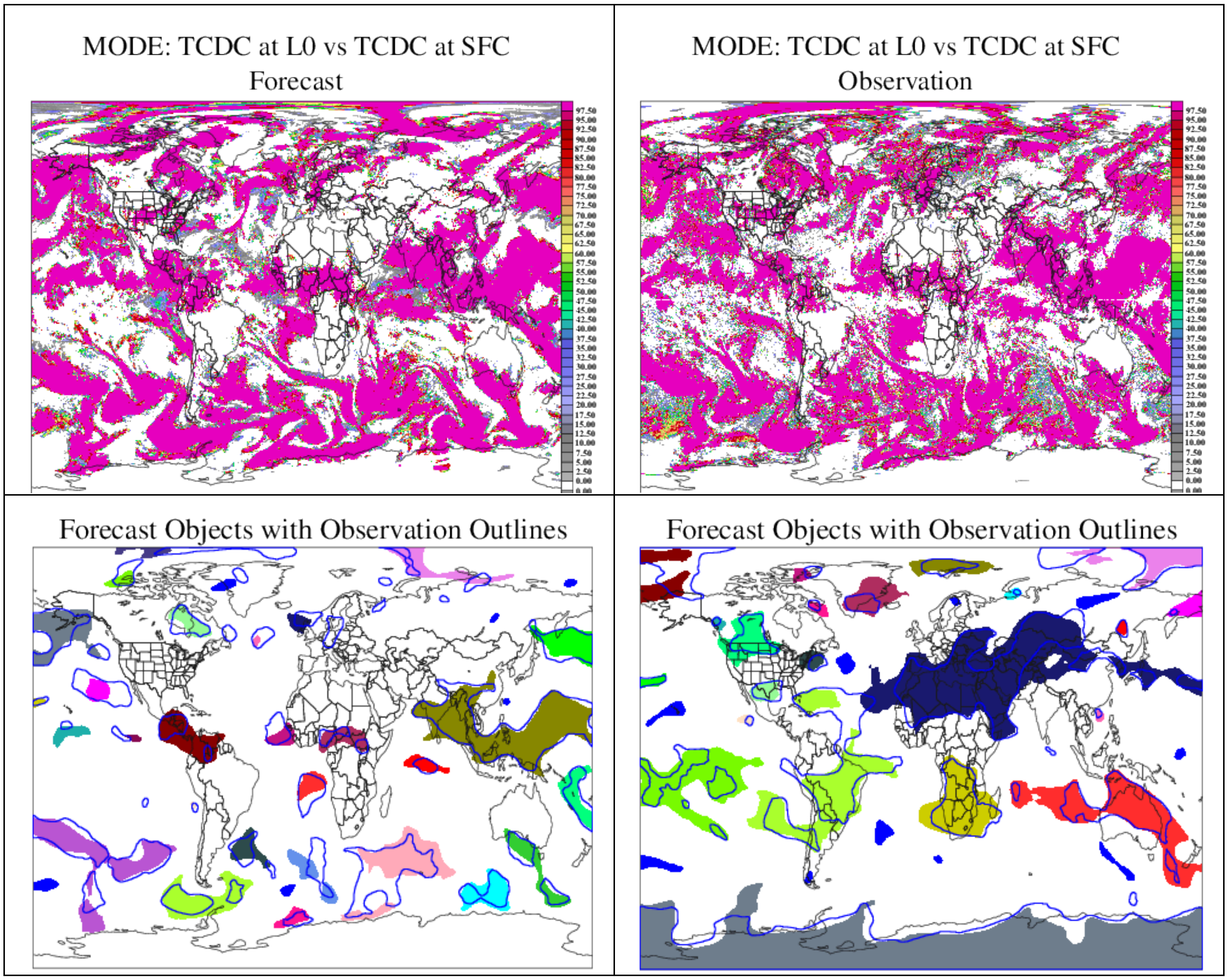


Figure 3.5-3. 6 hour forecast (upper right) and analysis (upper left) fields for 5 August 2016 valid at 12 UTC. Resulting MODE objects for a cloudy (lower left) and clear (lower right) configurations. Thresholds of >=80 and <=20 were applied for cloudy and clear, respectively. Forecast objects are shaded and analysis objects are outlines.

4 References

Hunke, E., and W. Lipscomb, 2008: CICE: The Los Alamos Sea Ice Model documentation and software user's manual version 4.0, Tech. Rep. LA-CC-06-012, Los Alamos Natl. Lab., Los Alamos, N. M.

Randall, D. A., and Cripe, D. G (1999). Alternative methods for specification of observed forcing in single-column models and cloud system models. Journal of Geophysical Research: Atmospheres 104, D20, 24527–24545.

Randall, D., S. Krueger, C. Bretherton, J. Curry, P. Duynkerke, M. Moncrieff, B. Ryan, D. Starr, M. Miller, W. Rossow, G. Tselioudis, and B. Wielicki, 2003: Confronting models with data - the GEWEX cloud systems study. Bull. Amer. Meteor. Soc., 84, 455–469. DOI: http://dx.doi.org/10.1175/BAMS-84-4-455.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601-608.

Zhang, M., R. C. J. Somerville, and S. Xie, 2016: The SCM concept and creation of ARM forcing datasets. *Meteor. Monogr.* **57**, 24.1–24.12. DOI: http://dx.doi.org/10.1175/AMSMONOGRAPHS-D-15-0040.1.

5 Acronyms and Abbreviations

AER	Atmospheric and Environmental Research
ADVCLD	Advect Cloud
AF	Air Force
AL	Atlantic
AMS	American Meteorological Society
AOML	Atlantic Oceanographic and Meteorological Laboratory
AOP	Annual Operating Plan
ARM	Atmospheric Radiation Measurement
ARW	Advanced Research WRF
ASTEX	Atlantic Stratocumulus to cumulus Transition EXperiment
BAMS	Bulletin of the American Meteorology Society
BT	Brightness Temperature
CCPP	Common Community Physics Package
CFSv2	Climate Forecast System version 2
CL	Control
CO	Cloud Overlap
CONUS	Contiguous United States
CPC	Climate Prediction Center
CSI	Critical Success Index
DA	Data Assimilation
DCF	Diagnostic Cloud Forecast
DOE	U.S. Department of Energy
DTC	Developmental Testbed Center
EC	Executive Committee
ECMWF	European Center for Medium Range Forecasting
EDMF	Eddy-Diffusivity Mass-Flux
EMC	Environmental Modeling Center
EnKF	Ensemble Kalman Filter
EnVar	Ensemble-Variational
EP	Eastern North Pacific
ER	Exponential Random
ESRL	Earth System Research Laboratory
FSS	Fractions Skill Score
GALWEM	Global Air Land Weather Exploitation Model
GASS	Global Atmospheric System Studies
GEWEX	Global Energy and Water cycle EXchanges
GF	Grell-Freitas
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecasting System
GMAO	Global Modeling and Assimilation Office
GMTB	Global Model Test Bed
GOES	Geostationary Operational Environmental Satellite

GSD	Global Systems Division
GSI	Gridpoint Statistical Interpolation
GSM	Global Spectral Model
HFIP	Hurricane Forecast Improvement Project
hPa	Hectopascals
HPC	High Performance Computing
HRD	Hurricane Research Division
HRRR	High Resolution Rapid Refresh
HWRF	Hurricane WRF
HWT	Hazardous Weather Testbed
IPD	Interoperable Physics Driver
JCSDA	Joint Center for Satellite Data Assimilation
LSM	Land Surface Model
MB	Management Board
MEG	Model Evaluation Group
MET	Model Evaluation Tools
MMET	Mesoscale Model Evaluation Testbed
MMM	Mesoscale and Microscale Meteorology (Laboratory at NCAR)
MODE	Method for Object-based Diagnostic Evaluation
NAM	North American Mesoscale
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCL	NCAR Command Language
NCWCP	NOAA Center for Weather and Climate Prediction
NEMS	NOAA Environmental Modeling System
NESDIS	National Environmental Satellite, Data and Information Service
NetCDF	Network Common Data Form
NGAC	NEMS Global Aerosol Component
NGGPS	Next Generation Global Prediction System
NH	Northern Hemisphere
NMMB	Nonhydrostatic Multiscale Model on the B grid
NOAA	National Oceanic and Atmospheric Administration
NRL	Naval Research Laboratory
NSC	NCAR Strategic Capability
NSF	National Science Foundation
NSSL	National Severe Storm Laboratory
NUOPC	National Unified Operational Prediction Capability
NWP	Numerical Weather Prediction
NWS	National Weather Service
OAR	Office of Oceanic and Atmospheric Research
PBL	Planetary Boundary Layer
PC	Partial Cloudiness
PDF	Probability Density Function
PODY	Probability of Detection – Yes
PS	Practically Significant
R2O	Research to Operations
R&D	Research and Development

RAMADDA	Repository for Archiving, Managing and Accessing Diverse DAta
RAP	Rapid Refresh
RI	Rapid Intensification
RMSE	Root Mean Square Error
RRTMG	Rapid Radiative Transfer Model for Global Climate Models
RT	Request Tracker
SAB	Science Advisory Board
SAS	Simplified Arakawa-Schubert
SCM	Single Colum Model
SGP	Southern Great Plains
SKEB	Stochastic Kinetic Energy Backscatter
SPP	Stochastic Parameter Perturbations
SPPT	Stochastic Perturbation of Physics Tendencies
SSEO	Storm Scale Ensemble of Opportunity
SS	Statistical Significance
SST	Sea Surface Temperature
SVN	Subversion
SUNY	State University of New York
T&E	Testing and Evaluation
TC	Tropical Cyclone
TCDC	Total Cloud Fraction
TWP-ICE	Tropical Warm Pool - International Cloud Experiment
UCACN	UCAR Community Advisory Committee for NCEP
UGCS	Unified Global Coupled System
UPP	Unified Post-Processor
URI	University of Rhode Island
USAF	United States Air Force
UTC	Coordinated Universal Time
VAP	Value Added Products
VAPOR	Visualization and Analysis Platform for Ocean, Atmosphere and Solar Researchers
WPC	Weather Prediction Center
WPS	WRF Preprocessing System
WRF	Weather Research and Forecasting
WWMCA	World Wide Merged Cloud Analysis
WWMCA-R	World Wide Merged Cloud Analysis, Reanalysis