Developmental Testbed Center Report
AOP 2014 Activities
1 April 2014 – 31 March 2015

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 results of research in Numerical Weather Prediction (NWP) 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. DTC activities are organized into five focus areas: Verification, Mesoscale Modeling, Data Assimilation, Hurricanes and Ensembles.

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 2014 and 31 March 2015. These activities include those described in the DTC 2014 Annual Operating Plan (AOP), as well as a few carry-over activities from the DTC AOP 2013.

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 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. For AOP 2014, DTC management implemented a new reporting process. Quarterly reports on the progress to date for each activity were prepared and distributed to the EC and MB members. Over the past year, the DTC hosted two external management meetings at NCAR's Foothills Campus in Boulder, CO: a SAB meeting on 10-12 September 2014 and a MB meeting on 20-21 January 2015. The purpose of the SAB meeting was to discuss strategic future directions for the DTC. The purpose of the MB meeting was to discuss the DTC AOP 2015 and nominations for SAB members whose term will expire in June 2015. DTC management also participated in an EC meeting at NWS Headquarters in Silver Spring, MD, on 18 February 2015. Recent DTC accomplishments, recommendations from the SAB, proposed activities for AOP 2015, and the future direction of the DTC were discussed at this meeting. The EC also approved the DTC Director's proposal to rotate off the three SAB members whose terms expire in June 2015 (Josh Hacker, Mark Stoelinga and Harold Brooks) and add five new SAB members (term begins 1 July 2015). The five new SAB members are: Russ Schumacher (Colorado State University), Brad Colman (Climate Corporation), David Gochis (NCAR), Adam Clark (University of Oklahoma) and Kayo Ide (University of Maryland).

DTC External Management Committees:

Executive Committee		Management Board			
Jim Hurrell	NCAR	Josh Hacker	NCAR	Hendrick Tolman	NOAA/NWS
Bill Lapenta	NOAA/NWS	Joe Klemp	NCAR	Fred Toepfer	NOAA/NWS
Ralph Stoffler	Air Force	Robert Swanson	Air Force	Kevin Kelleher	NOAA/OAR/ESRL
Sandy MacDonald	NOAA/OAR	John Zapotocny	Air Force	Tom Hamill	NOAA/OAR/ESRL

Science Advisory Board

Harold Brooks	National Severe Storms Laboratory (NSSL)
Robert Fovell	University of California – Los Angeles (UCLA)
Kristen Corbosiero	State University of New York – Albany
Sharanya Majumdar	University of Miami
David Novak	National Centers for Environmental Prediction (NCEP)/Weather Prediction Center
Geoff DiMego	NCEP/Environmental Modeling Center (EMC)
Jenni Evans	Pennsylvania State University (PSU)
Joshua Hacker	NCAR
S. R. Gopalakrishnan	NOAA/Atlantic Oceanographic and Meteorological Laboratory (AOML)
Evan Kuchera	Air Force
Gary Lackmann	North Carolina State University
Carolyn Reynolds	Naval Research Laboratory (NRL)
Mark Stoelinga	Vaisala
Robert J. Trapp	Purdue University
Kelly Mahoney	Cooperative Institute for Research in Environmental Sciences (CIRES)

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. During AOP 2014, the DTC distributed its Winter-Spring 2014, Summer 2014, Autumn 2014 and Winter 2015 issues of Transitions. Issues of Transitions can be accessed at: http://www.dtcenter.org/newsletter/. Staff from both nodes of the DTC contributed to the newsletter, as well as SAB and MB members and DTC visitors. In addition to these on-going efforts, the DTC worked with EMC to gather information that provided the foundation for a preliminary design for a NWP Information Technology Environment (NITE), which would facilitate the use of operational NWP systems by a broader spectrum of the research and development (R&D) community.

1.2.1 Community Outreach Events

Over the past year, the DTC co-hosted five workshops. DTC staff participated in and facilitated the 6th Ensemble User's Workshop that took place 27-29 March 2014 in College Park, Maryland. The executive summary of the workshop is available on the DTC website (http://www.dtcenter.org/eval/ensembles/).

A more extensive version of the report that involves a wider group of the workshop participants is in preparation. Once the extended report is finalized, it will also be made available on the DTC website. In May 2014, the DTC jointly organized with the Central Weather Bureau (CWB) and Taiwan Typhoon and Flood Research Institute (TTFRI) the Workshop on Numerical Prediction of Tropical Cyclones, which was held in Taipei, Taiwan. Sixty scientists from Taiwan, the United States, China, Japan, South Korean, India, Vietnam, Thailand, Philippines and Malaysia participated in discussions on cutting-edge tropical cyclone research. The workshop also provided a forum for fostering international collaborations between the research and operational communities directed at advancing the skill of tropical cyclone (TC) numerical predication. In June 2014, the DTC co-hosted with NCAR's Mesoscale and Microscale Meteorology (MMM) Division the 15th WRF Users' Workshop in Boulder, CO. The first day consisted of best-practice presentations, followed by a three-day workshop consisting of 62 talks and approximately 80 posters. The last day consisted of six mini-tutorials on MPAS, WRF-Hydro, LAPS-DA, VAPOR, NCL and IDV. Presentations are available on the workshop website at http://www2.mmm.ucar.edu/wrf/users/workshops/WS2015/WorkshopPapers.php. On 16 September 2014, the Air Force Scientific Services, Training and Standards Division (A3N) and the DTC hosted a Model for Prediction Across Scales (MPAS) workshop at the Offutt Air Force Base in Bellevue, Nebraska. The morning sessions of the workshop consisted of broad overview presentations by the MPAS development team intended for a large audience, whereas the afternoon sessions were intended for the technicians who actually run the Air Force's weather models. By bringing key members of the MPAS development team to Offutt Air Force Base to give presentations on the MPAS capabilities, the workshop provided basic background information about this new capability, as well as an opportunity for Air Force personnel with the opportunity to ask directed questions about setting up and running MPAS and how this new capability would meet the Air Force's needs. Remote access to the morning session of the workshop was also provided to the DTC's NWS partners by the Air Force. For the fifth event, the DTC worked with the NWS's Next Generation Global Prediction System (NGGPS) initiative and EMC to organize and host the "Parameterization of Moist Processes for Next-Generation Weather Prediction Models" workshop, held 27-29 January 2015 at the NOAA Center for Weather and Climate Prediction (NCWCP) in College Park, Maryland. The goals of this workshop were to inform and advise on future directions of moist process parameterization development, with a particular emphasis on NWP applications for scales and resolutions ranging from synoptic to convection-permitting, to identify the most promising ideas and inform planning for NGGPS and other existing and emerging NOAA global and regional forecast models. A report on the workshop discussions and recommendations, along with all of the presentations are available on DTC website at http://www.dtcenter.org/events/workshops15/mm_phys_15/

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 continued to provide support for a visitor project selected in January 2012 (see Table 1.2.2-1) and five visitor projects selected in January 2013 (see Table 1.2.2-2). Final reports for the 2012 project and four of the 2013 projects (Clark, Zhang, Galarneau, and Mittermaier) were posted on the visitor portion of the DTC website (http://www.dtcenter.org/visitors/). In addition to his project report, the software Dr. Vigh developed to produce synthetic profiles for tropical cyclone (TC) forecasts, which provides a proto-type capability to diagnose model errors for storm size, inner core kinematic and thermodynamic structure and surface wind distribution, is in the process of being transitioned to the contributed tools available to HWRF developers and DTC staff. Other high lights from visitor projects over the past year include Dr. Fovell and his graduate student Peggy Bu DTC visit in June and July, during which they worked closely with the DTC's Hurricane Team to further analyze recent physics testing by the DTC and conduct idealized simulations to test hypotheses about the underlying reasons for performance differences. Dr. Fovell and Peggy Bu's visit culminated with an informative presentation to DTC staff, which included stimulating discussion. Dr. Mittermaier visited the DTC and EMC in August to discuss progress on her visitor project. Dr. Mittermaier briefed the DTC Verification Team on her fairly simple but enlightening approach to assessing observation uncertainty through the use of high temporal resolution observations. During her visits, Dr. Mittermaier also gave an interesting presentation to DTC and EMC staff on her application of the MODE (Method For Object-Based Diagnostic Evaluation) tool to global forecasts as part of the UKMET office implementation process. These interactions with DTC visitors provided DTC staff with an abundance of ideas for how diagnostic approaches could be incorporated into future DTC testing and evaluation activities. The only project selected in January 2013 that is still open is Dr. Fovell's, which is due in May 2015.

Table 1.2.2-1. 2012 Visitor Projects

PI	Institution	Project Title
Jonathan Vigh	NCAR	Development of an HWRF Diagnostics Module to Evaluate Intensity and Structure Using Synthetic Flight Paths Through Tropical Cyclones

Table 1.2.2-2. 2013 Visitor Projects

PI	Institution	Project Title
Adam Clark	University of Oklahoma	Object-based Time-Domain Diagnostics for High-resolution Ensemble Forecasting and Evaluation in NOAA/HWT Spring Forecasting Experiments
Robert Fovell / Peggy Bu (graduate student)	UCLA	Improving HWRF Track and Intensity Forecasts Via Model Physics Evaluation and Tuning
Thomas Galarneau	NCAR	Diagnosing Tropical Cyclone Motion Forecast Errors in HWRF
Marion Mittermaier	UK MET Office (UKMET)	Incorporating Observations Uncertainty to a Spatial Probabilistic Verification Framework for km-scale Models
Man Zhang	Colorado State University	Impact Assessment of Cloud-Affected AMSU-A Radiance Assimilation in TC inner-Core Region using Hybrid Data Assimilation Approaches

In May 2014, the DTC selected four projects for funding from the proposals submitted in response to an Announcement of Opportunity (AO) distributed in the Fall of 2013 (see Table 1.2.2-3). In addition, the DTC received three off-cycle proposals that were selected for funding (see Table 1.2.2-4). Six of these projects are already underway and progressing nicely, with the seventh scheduled to get underway over the summer. Dr. Yablonsky completed his visits, during which he worked with DTC staff to successfully transition new ocean model capabilities, as well as diagnostic tools, to the HWRF repository, and submitted a report on the ocean model portion of his project in December 2014. This report is now available on the DTC website. The wave model portion of this project will be completed through a University of Rhode Island (URI) graduate student visit to EMC in the coming months. A report on this portion of the URI project will also be posted on the DTC website.

Table 1.2.2-3. 2014 Visitor Projects

PI	Institution	Project Title
Shaowu Bao	Coastal Carolina University	Evaluation of Two HWRF Microphysics/Radiation Configurations with Remote-Sensing Data
István Geresdi	University of Péc	Towards Improving Representation of Convection and MCC Longevity in High-Resolution WRF and NEMS-NMMB Model Forecasts
Hongli Wang	Colorado State University	Estimation of Initial and Forecast Error Variances for the NCEP's Operational Short-Range Ensemble Forecast (SREF) System
Richard Yablonsky	University of Rhode Island	Developing and Supporting Global HWRF Ocean Coupling with Advanced Ocean Physics and Initialization Options and New Diagnostic Tools for Comprehensive Model Evaluation
Thomas Galarneau	NCAR	Diagnosing Tropical Cyclone Motion Forecast Errors in the 2014 HWRF Retrospective Test (H214)

Table 1.2.2-4. Off-Cycle Visitor Projects

PI	Institution	Project Title
Thomas Galarneau	NCAR	Diagnosing Tropical Cyclone Motion Forecast Errors in the 2014 HWRF Retrospective Test (H214)
Paul Roebber	University of Wisconsin-Milwaukee	Demonstration Project: Development of a Large Member Ensemble Forecast System for Heavy Rainfall using Evolutionary Programming
Jason Otkin	University of Wisconsin - Madison	Object based verification for the HRRR model using simulated and observed OES infrared brightness temperatures

1.2.3 NWP Information Technology Environment (NITE)

For scientists outside of the NWS to contribute relevant R&D to NCEP's modeling suite, it is important that they work with the current operational codes, workflows, and input datasets. However, obtaining such codes and inputs, and configuring the system to run with data assimilation and cycling workflows identical to those used in operations, can be a daunting task for the research community. To facilitate the use of operational NWP systems by the broader R&D community, the DTC undertook the task of assembling a preliminary design for a framework referred to as the NWP Information Technology Environment or NITE that would facilitate preparing and running research experiments using NCEP's modeling systems. For this activity, the DTC pursued a two-pronged approach: 1) gathering information on current tools and capabilities for running NWP systems that are available both nationally and internationally, and 2) gathering information on requirements for NITE to be useful to both developers at EMC and the broader research community. The NITE team gathered information on available frameworks for running both NWP and climate modeling systems. Central to this information gathering activity was a visit to ECMWF and the UKMET office by DTC staff and an EMC representative in December 2014. To gather information on the requirements for NITE, a questionnaire was formulated in close collaboration with EMC and distributed to scientists at EMC, the DTC and the broader community. A report on their findings and a preliminary design can be accessed at: http:/www/dtcenter.org/eval/NITE/. Numerous presentations on this concept have been given or are planned for workshops and conferences to publicize this work and collect feedback from the community.

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)
- NOAA Environmental Modeling System (NEMS)
- Gridpoint Statistical Interpolation (GSI) and Ensemble Kalman Filter data assimilation (DA) systems
- Modular end-to-end ensemble system
- Model Evaluation Tools (MET) – Verification package

With the exception of MET, the DTC does not, for the most part, contribute to the development of new scientific techniques for these software packages. Rather, the DTC contributes to the software management of all of these systems and user support for the publicly-released systems (WRF, HWRF, GSI 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), Geophysical Fluid Dynamics Laboratory (GFDL), URI and the Hurricane Research Division (HRD) of NOAA's AOML. DTC activities are currently focused on the regional application of these software systems. In addition to working with these individual software systems, the DTC is involved in efforts to apply the GSI-Hybrid DA technique to operational forecasts, which requires linking the GSI DA system with both an ensemble system and the operational model.

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 – The WRF code repository trunk and the main HWRF development branch were synchronized to assure the WRF code used by NCEP and the community stays unified. In preparation for the next WRF code release, new developments and improvements were collected from the community and merged into the WRF code repository and test scripts were updated to accommodate the new code. New physics options that will be included in the May 2015 release are: an updated Tiedtke convection scheme, a scale-aware version of the Kain-Fritsch convection scheme, a fast version of the Rapid Radiative Transfer Model for Global Climate Models (RRTMG), a scale-aware Yonsei University (YSU) planetary boundary layer (PBL) scheme, and a scale-aware relative humidity based partial cloudiness scheme. The partial cloudiness scheme, which was developed through DTC funding, will be included in the 2015 HWRF implementation.

Unified Post-Processor (UPP) – The DTC continued to work closely with EMC to manage the UPP code base through regular bi-monthly meetings. Efforts to keep the community UPP repository in sync with EMC's operational UPP repository are ongoing. Work is underway, through collaboration with EMC staff, to combine the automated regression tests for the operational and community codes to create a multi-platform, cross-organizational test suite. Interactions with community users led to the inclusion of several bug fixes and the addition of new capabilities, including new microphysics-specific reflectivity output fields and new output fields for synthetic satellite imagery.

NEMS – Work continued to enhance the portability of the software package and associated libraries. Incorporating the Thompson microphysics scheme into the NEMS framework and coupling it to the Rapid Radiative Transfer Model (RRTM) for use with the Nonhydrostatic Multiscale Model on the B grid (NMMB) was completed and the updated software was officially checked into the NEMS repository at EMC. The tag associated with this check-in is the code base for the extensive testing and evaluation efforts undertaken during AOP 2014. This code was also made available as a friendly user release. The DTC also developed utilities to dump, view, and convert NEMS intermediate output files to Network Common Data Form (NetCDF). Work was also undertaken to implement the RUC LSM in NMMB. This new capability will be added to the NEMS repository during AOP 2015.

HWRF – The HWRF system moved from using *ksh* scripts to *python* scripts developed through a close collaboration between EMC and DTC. This transition improves system compatibility and ensures that EMC, NCEP Central Operations (NCO), the DTC and the research community work with functionally equivalent scripts and codes. Development of enhanced HWRF scripting and automation to run a multi-storm configuration is near completion and will be checked into the HWRF repository during AOP 2015. More information about the new features and capabilities can be found at http://www.dtcenter.org/HurrWRF/users/. In addition to these on-going software maintenance activities, the DTC developed tools for enhanced support to HFIP-funded principal investigators contributing to HWRF development. To help facilitate the transfer of new innovations from new and existing developers to the HWRF repository, the DTC created a website http://www.dtcenter.org/HurrWRF/developers/ that documents the code management procedures for developers who want to contribute to the HWRF system. It lays out instructions on how to use the repository, check out and commit codes to alleviate problems with transferring codes between multiple repositories and protect integrity of the code. Support for HWRF developers included assisting HRD developers with the application of the restart utility in the HWRF Ensemble Data Assimilation System (HEDAS). The DTC also collaborated with URI to implement ocean coupling for all ocean basins using Message Passing Interface Princeton Ocean Model for Tropical Cyclones (MPIPOM-TC), along with an option to initialize the ocean model from alternate datasets, such as the Navy's Coupled Data Assimilation (NCODA) analysis. The DTC also tested and provided feedback on diagnostic tools for the ocean model provided by URI.

Holt, C., L. Bernardet, T. Brown, and R. Yablonsky, 2014: *Community HWRF Users Guide V3.6a. NOAA Tech. Memo. OAR GSD-45*, doi:10.7289/V5BC3WG5, 128 pp, [Available at http://docs.lib.noaa.gov/noaa_documents/OAR/GSD/NOAA_TM_GSD-45.pdf.]

GSI and EnKF – An EnKF code management framework was set up following the protocol used by the GSI effort and the EnKF and GSI repositories were merged. The GSI and EnKF codes are now being reviewed by a joint DA Review Committee (DRC). Work related to the EnKF code base also included developing regression tests and composing a user's guide. The DTC continued to facilitate the joint GSI/EnKF Data Assimilation Review Committee (DRC) code review process by performing code testing and reviews for each proposed code update and synchronizing the DTC trunk with the latest EMC trunk. A multi-platform compilation tool was developed using autotools, replacing the current community build system. The tool includes compilation of GSI, EnKF, and their supplemental libraries, as well as the NCEP compilation options for NOAA research and production computers. The tool was designed to incorporate the existing capabilities at the DTC and EMC. The first version of the tool has been ported to EMC for feedback.

2.2 Verification Tool Development

MET v5.0, which was released on 8 September 2014, included bug fixes, an auto-configuration capability to make compiling easier, calculations that more consistent with NCEP methods, updated GRIB1 and GRIB2 tables to be consistent with NCEP's usage, twelve additional continuous, categorical and neighborhood statistics, and expanded MET-TC capabilities. MET community support also included 90 bug fixes and resolved development tasks and 155 help desk tickets that were addressed and closed. There were 370 total downloads since release. A MET Tutorial was given on 2-3 February 2015 in conjunction with the WRF tutorial the previous week. The total number of participants was 24 including many international participants and three drop-ins. Feedback from participants was positive.

During AOP 2014, MET-TC capabilities were expanded in four basic areas in response to user needs: 1) discrimination between TCs over land or water, 2) methodology for aggregating interpolated forecasts, 3) controls on the verification sample with respect homogeneity across lead times, and 4) specifying thresholds for RI/RW. The capability to discriminate between TCs over land or water was initially limited to the Atlantic and eastern North Pacific basins. With the addition of a global 1/10th degree distance to land data file, this capability is now available in all basins. NHC forecasters generally use guidance from dynamical models that have been converted to "early" model guidance by applying either a 6-h or 12-h interpolation. At the request of NHC, the TC-Pairs tool was enhanced to allow two options for aggregating interpolated forecasts based on a configuration flag. The sample for option 1 is composed of 6-h interpolations when they are available and 12-h interpolations only when the 6-h is not available. The sample for option 2 is composed of 12-h interpolations when they are available and 6-h interpolation only when the 12-h is not available. The sample used to compute track and intensity errors generally varies with lead time, with the largest sample associated with shorter lead times and the smallest sample with the longest lead times. Variations in relative performance between two forecast models or model configurations across lead times may stem from the fact that the sample is not uniform across lead times. To address this consideration, the TC-Stat tool was enhanced by adding a configuration option to allow the user to specify lead times that must be present for a case to be included in the event equalization. Finally, TC-Stat was further refined by adding more flexibility to the

RI/RW criteria. Users now have the ability to identify RI/RW events based on the standard definition of RI/RW or user defined specifications of thresholds for magnitude of intensity change and/or the time period over which the change occurs. The enhancements to the MET-TC software package were an HFIP AOP 2013 carryover activity. All of these enhancements to MET-TC are included in MET v5.0.

During AOP 2014, the DTC verification team worked with the teams for the other focus areas to maintain the MET and METViewer packages for use in DTC T&E activities, including minor updates, bug fixes, and technical support. The team also worked with each focus area to augment their verification capability based on their evaluation needs. The additions and enhancements to MET stemming from these interactions, which will be included in the summer 2015 MET release, are described below.

For the Hurricanes focus area, the definition of rapid intensification (RI) / rapid weakening (RW) events within MET-TC was enhanced to provide the ability to modify the window over which forecast and Best Track RI/RW events are matched to provide a more lenient definition of the contingency table that allows for timing errors. Additionally, the masking capability within MET was augmented to allow users to define more complicated regions through the selection of the union, intersection, or symmetric difference of multiple masking regions. This capability allowed for easy generation of the "mega domain" and storm relative masks discussed in Section 3.

The Mesoscale Modeling focus area requested an advancement in MET to provide regridding automatically within the statistical tools. This work was conducted in support of the Mesoscale Modeling Evaluation Testbed (MMET) to provide end-users with an easy way to regrid the provided observation data to their particular application. In the current METv5.0 release, users are required to put both forecast and observed data on the same grid prior to running MET tools using a tool such as copygb within UPP. This new MET capability allows the user to pass in two grids that are not on the same projection and define the grid on which the verification is performed, including: interpolate observed field to forecast grid; interpolate forecast field to observed grid; or interpolate both grids to a third grid that is either a standard in MET, a user generated grid, or a text string that defines the grid (similar to copygb). Figure 2.2-1 shows an example of the regridding capability. Interpolation methods that are supported include: unweighted and distance-weighted mean, minimum, maximum, median, least squares, bilinear and budget (to conserve mass). Automated regridding is now available within all MET tools. Additionally, a standalone tool, called regrid_data_plane, has been added to allow for regridding outside of the MET tools. Finally, the regridded fields may be written out to NetCDF files for use at a later date or by other packages. This advancement of MET provides increased utility of the package by decreasing the number of steps needed to perform verification, as well as decreasing the amount of data that must be stored in regridded formats.

For the Data Assimilation focus area, MET was enhanced to read the output of the GSI diagnostic file, including the background error and innovations, and write the information out in the matched pair line type. The MET Stat-Analysis tool may then be used to calculate traditional statistics (e.g. Root Mean Square Error and Probability of Detection) and aggregate over the entire grid, regions or at individual observing locations. Figure 2.2.2 provides an example of GSI innovations aggregated by observation station using Stat-Analysis and plotted using NCAR Command Language (NCL). This capability is currently being extended to read GSI EnKF diagnostic files and calculate ensemble-based statistics.

The verification team's contributions to the Ensembles focus area during AOP 2014, which were funded by AOP 2013 carryover, focused on adding support within METViewer for the NCEP/EMC ensemble Verification Statistic DataBase (VSDB) file format. Several additions to database tables were made and a specialized database loader was developed to place the VSDB data into the METViewer structure. The Ensembles team has uploaded data from recent T&E activities and is currently testing the functionality. Table 2.2-1 provides the status for the statistics generated by the NCEP/EMC aggregation script that use

VSDB files. Available connotes the statistic is now available for aggregation and plotting using METViewer v1.1. Addition of the remaining fields will be completed during AOP 2015 using MET development funds.

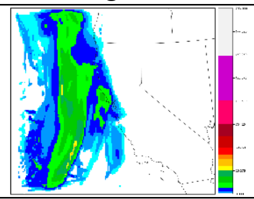
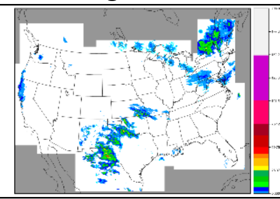
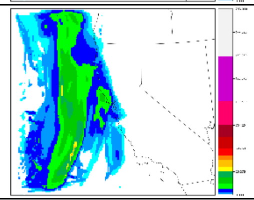
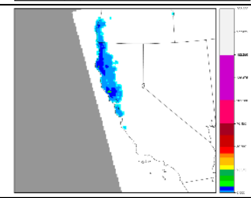
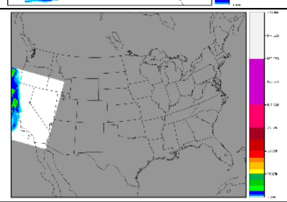
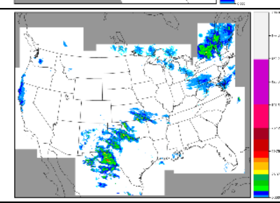
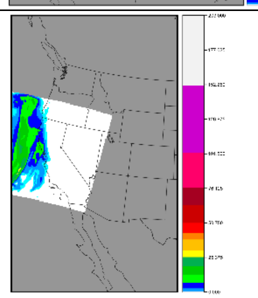
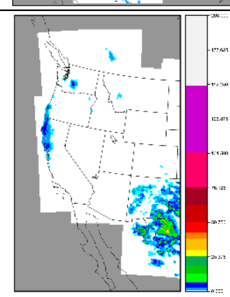
	9 km Regional Forecast	4km Stage IV Observations
Original Grids		
Verify on Forecast Grid		
Verify on Observation Grid		
Verify on a different Grid (e.g. DTC 165)		

Figure 2.2-1. Example of automated regridding tool in MET. Forecast field is 2012 Hydrometeorology Testbed – West ensemble mean precipitation forecast. Observation field is Stage IV precipitation estimate. Top row: Original grids. Second row: Fields interpolated to forecast grid. Third row: Fields interpolated to observation grid. Bottom Row: Fields interpolated to a third grid, the DTC 165 over the western U.S.

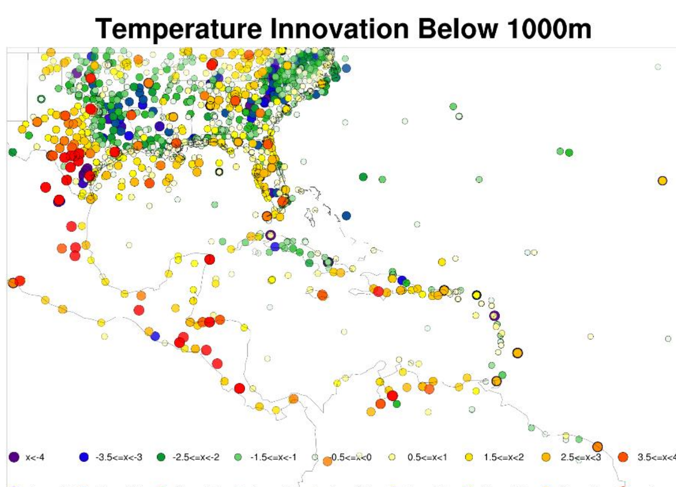


Figure 2.2.2. Example of GSI innovations aggregated using the MET Stat-Analysis tool.

Table 2.2-1 VSDB Statistics Available in METViewer. *N/A* indicates statistic is still being implemented.

Statistics	**VSDB Name**	**METViewer Name**
Root Mean Square Error (RMSE)	RMSE	RMSE
Ensemble Spread (or Standard Deviation)	Spread	ESTDEV
Ratio: Spread/RMSE	Spread/RMSE	*N/A*
Mean Absolute Error (MAE)	ABSerr	MAE
Mean Error (or Bias)	MeDerr	ME
Pattern Anomaly Correlation (PAC)	PAC	PAC
Continuous Ranked Probability Score (CRPS)	CRPS	CRPS
Continuous Ranked Probability Skill Score (CRPSS)	CRPSS	CRPSS
Brier Skill (BS)	Brier	BS
Brier Skill Score (BSS)	BSS	BSS
Ensemble Rank Histogram	Histogram	Rank Histogram
Ensemble Relative Position	Reli-position	*N/A*
Points for Reliability Diagram	Reliability	*N/A*
Points for Receiver Operator Characteristic Curve (ROC)	ROC	*N/A*
Economic Value	Economic Value	*N/A*

2.3 Publicly-Released Systems

The DTC currently collaborates with developers on six software systems that undergo a public release process: WRF, UPP, HWRF, GFDL vortex tracker, GSI and MET. Assistance continued to be offered through email helpdesks for all packages. Statistics regarding the timing and version of the most recent release, 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.

In addition to this on-going work, the DTC made great strides over the past years towards adding two more software packages to the list of supported publicly-released packages. A Friendly User Release (FUR v0.9) for NEMS was packaged for distribution that included the source code for NEMS and the NMMB Preprocessing System (NPS), 3 general test cases (single domain, nested domain, and Thompson/RRTM example), and instructions for building and running NEMS-NMMB on the NCAR supercomputer (Yellowstone) and two NOAA research computing platforms (zeus and jet). The release announcement was distributed on 18 February 2015 to a target audience by email and made publicly available on the newly established NEMS-NMMB webpage (http://www.dtcenter.org/nems-nmmb/users/). The NEMS FUR served as the code base for the NMMB Tutorial the DTC offered on 1-2 April 2015. In addition to NEMS, the DTC also completed work toward a beta release of the EnKF DA system on 31 January 2015, including an EnKF Users' Guide.

The DTC hosted its first international HWRF tutorial on 22-23 May 2014 at the TTFRI National Applied Research Laboratories in Taipei, Taiwan. The tutorial lectures were provided by members of EMC's HWRF development team and DTC staff, with opening remarks provided by TTFRI's director, C.S. Lee, and HFIP's development manager, Robert Gall. This tutorial, which had 25 participants, provided a great opportunity to foster international collaboration on the continued development of HWRF. All slides for the tutorial are available to the community at:
http://www.dtcenter.org/HurrWRF/users/tutorial/2014_Taiwan_tutorial/tutorial2014.php.

The DTC co-hosted with NCAR's Mesoscale and Microscale Meteorology (MMM) Division two Basic WRF tutorials at NCAR's Foothills Laboratory in Boulder, Colorado (July 2014 and January 2015). Both tutorials included lectures given by system experts and one-on-one assistance during practical sessions. Two new lectures were added to the July tutorial, with a focus on the fundamentals of atmospheric modeling and physics. These new lectures were given by a guest speaker, Dr. Song-Hou Hong of the Korea Institute of Atmospheric Prediction Systems. Slides from the July and January tutorials are available at http://www2.mmm.ucar.edu/wrf/users/tutorial/tutorial_presentation_summer_2014.htm and http://www2.mmm.ucar.edu/wrf/users/tutorial/tutorial_presentation_winter_2015.htm, respectively.

The DTC also hosted its annual GSI tutorial on 14-16 July 2014 at NCAR's Foothills Laboratory in Boulder, Colorado, with 41 students (maximum capacity) participating. The guest university speaker this year was Dr. Milija Zupanski from Colorado State University. Other speakers were invited from NCEP, NOAA, NASA, NCAR, and the DTC. All slides are available to the community at: http://www.dtcenter.org/com-GSI/users/docs/index.php. In addition, the DTC hosted a complementary EnKF instructional session to friendly users on 17 February 2015. Speakers included NOAA's code development team, as well as DTC staff. Three hours of lectures were given covering various aspects of the EnKF system, including theory, compilation, configuration, running, and diagnostics. The instructional session reached its maximum capacity (total 40 participants, 26 onsite and 14 remote).

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 six months.

Software Package	Public Release			
	Version	Timing	Registered Users	Helpdesk tickets per month
WRF	V3.6.1	August 2014	~25,900	~400
UPP	V2.2	April 2014	-	
NEMS	FUR v0.9	February 2015	21	-
HWRF	V3.6a	September 2014	970	~25
GFDL Vortex Tracker	V3.5a	September 2013	450	
GSI	V3.3	June 2014	1,272	~30
EnKF	Beta v1.0	January 2015	-	~10
MET	V5.0	September 2014	2,710	~30

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

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

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 and includes both model inter-comparison and stand-alone (Reference Configuration (RC)) T&E. 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 Mesoscale Modeling

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.

3.1.1 Testing Protocol and Mesoscale Model Evaluation Testbed (MMET)

The Mesoscale Model Evaluation Testbed (MMET; http://www.dtcenter.org/eval/mmet), which was established to assist the research community with the initial stage of testing in the research to operations (R20) process, continues to be maintained and enhanced over time. MMET provides the opportunity for the research community to conduct their own T&E of a new technique. Datasets for thirteen cases, deemed to be of high interest by EMC, have been collected and are being distributed via RAMADDA, a **R**epository for **A**rchiving, **M**anaging and **A**ccessing **D**iverse **DA**ta, originally developed by Unidata that is now a third party package developed by an open source project (http://ramadda.org/). MMET data sets include a variety of initialization and observation data sets that have been gathered and packaged for distribution. Baselines for select operational configurations established by the DTC utilizing the MMET data sets are also available through RAMADDA. The baseline results were updated for each of the thirteen cases with a more current version of the model code, including: WRFv3.6.1 and a NEMS tag from Sept 2014. Presentations describing MMET were given at the WRF-GRAPES workshop (Oct 2015) and NCAR/RAL retreat (Dec 2015) to solicit prospective collaborators. In addition, DTC staff mentored a Significant Opportunity in Atmospheric Research and Science (SOARS) protégé, Anthony Torres, from the University of Michigan, who used MMET to run the 29 June 2012 derecho case with different physical parameterization schemes and horizontal grid spacing to investigate the impact. The protégé presented his work at the 2014 Society of Hispanic Professional Engineers Meeting (Nov 2014) and the 95th AMS Annual Meeting (Jan 2015).

In addition to on-going maintenance, DTC staff collaborated with scientists from NOAA ESRL's Physical Sciences Division (PSD) to identify an atmospheric river (AR) case (14-17 February 2011) from the Hydrometeorology Testbed (HMT) to include in MMET. The DTC's contribution to this effort was funded under its carryover USWRP funding from AOP 2013. This particular case will be run for a 15-km CONUS parent domain and a 5-km nest centered over California (Fig. 3.1.1-1). Baselines for two configurations will be produced by the DTC; one Advanced Research WRF (ARW) configuration that uses a physics suite defined and tuned by PSD staff and one NEMS NMMB configuration. In addition to the traditional data provided for the standard cases, the AR case will include plots of integrated water vapor, as well as the scripts used to produce them.

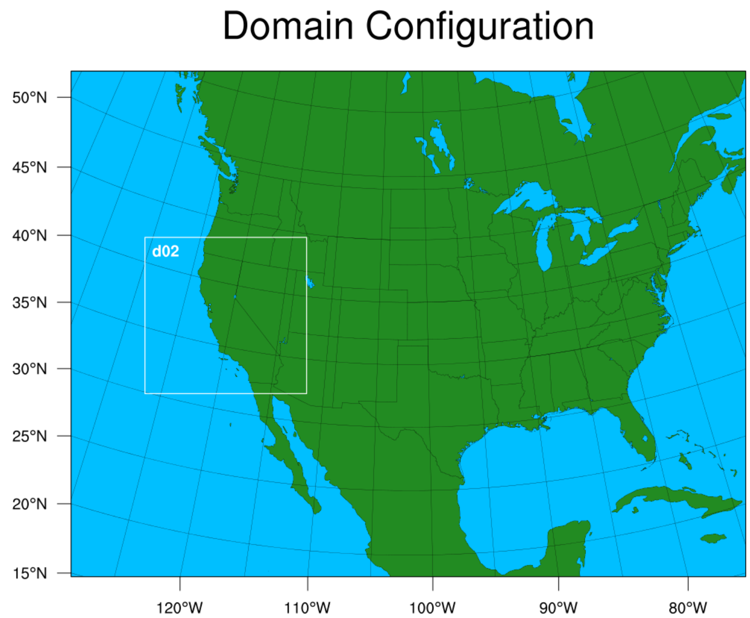


Figure 3.1.1-1. Computational domains used for the 14-17 February 2011 AR case to be included in MMET. The outer box defines the 15-km CONUS domain (d01), while the inner white box defines the 5-km nest (d02).

3.1.2 NEMS

3.1.2.1 Impact of Thompson/RRTM on NAM physics suite

A comprehensive inter-comparison T&E activity employing the NEMS-NMMB model infrastructure was conducted and extensive analysis is underway. The focus for this testing is an assessment of the impact of replacing the Ferrier-hires microphysics scheme with the Thompson microphysics scheme within NMMB for the North American Mesoscale (NAM) application (Table 3.1.2.1-1). This T&E activity, which is a carryover activity from AOP 2013, employed a parent domain at 12-km grid spacing and two nests at 3-km grid spacing: CONUS and Alaska (Fig. 3.1.2.1-1). The end-to-end system included the NPS, NMMB, UPP, and MET software packages. The NEMS-NMMB code used for this T&E activity corresponded to the FUR v0.9. The testing period included one month per season in 2013-2014 (Table 3.1.2.1-2) with cases initialized every 36 hours and run out to 48 hours (for a total of 94 potential cases), providing a distribution of both 00 and 12 UTC initializations.

Table 3.1.2.1-1. Physics suites for the NMMB T&E activity.

Physics Scheme	Baseline Configuration	Replacement Configuration
Microphysics	Ferrier-hires	Thompson
Radiation Shortwave (SW) and Longwave (LW)	Rapid Radiative Transfer Model	Rapid Radiative Transfer Model
Surface Layer	Mellor-Yamada-Janjic	Mellor-Yamada-Janjic
Land Surface Model	Noah	Noah
Planetary Boundary Layer	Mellor-Yamada-Janjic	Mellor-Yamada-Janjic
Convection	Betts-Miller-Janjic (parent only)	Betts-Miller-Janjic (parent only)

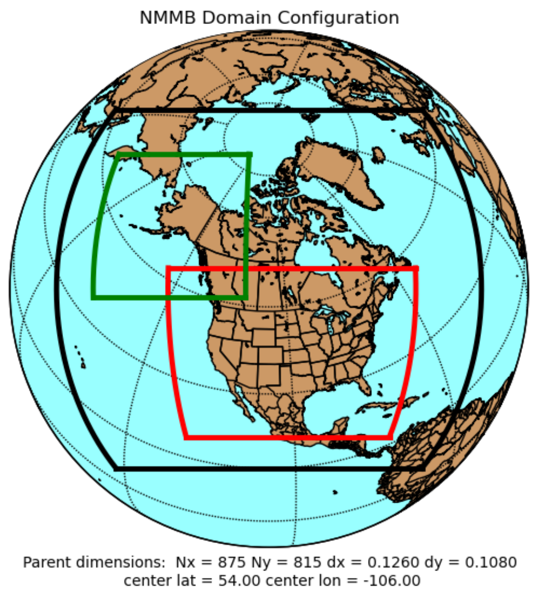


Figure 3.1.2.1-1. Computation domains used for the NMMB T&E activity. The black box defines the 12-km parent domain, the red box defines the 3-km CONUS nest, and the green box defines the 3-km Alaska nest.

Table 3.1.2.1-2. Dates for the NMMB T&E activity.

Season	Dates
Fall	12 Oct – 15 Nov 2013
Winter	16 Jan – 19 Feb 2014
Spring	16 Apr – 17 May 2014
Summer	6 Jul – 9 Aug 2014
Fall	12 Oct – 15 Nov 2013
Winter	16 Jan – 19 Feb 2014

The full evaluation included an assessment of several different variables. In terms of traditional verification approaches, the surface and upper-air temperature, dew point temperature, and wind speed were evaluated using bias-corrected root mean squared error and bias and the precipitation accumulation and composite reflectivity variables were evaluated using Gilbert skill score (GSS) and frequency bias for 3- and 24-h accumulations. For each of the evaluated parameters, confidence intervals at the 99% level were applied to objectively assess statistical and practical significance. Further analysis is underway to continue to compare model output fields from each configuration such as shortwave and longwave radiation, surface fluxes (sensible, latent and ground heat), and PBL height.

Preliminary results of the T&E inter-comparison for near-surface variables indicate that the 2-m temperature for both configurations exhibited warm biases in the summer and neutral-to-cold biases in the winter (not shown). When SS pair-wise differences were present (see Table 3.1.2.1-3), the Thompson configuration typically had lower bias values, leading to better performance during the summer (when there was a warm bias) and worse performance in the winter (when there was a cold bias). For 2-m dew point temperature, both configurations produced a dry bias in summer and a moist bias in winter (not shown); overall, the Thompson configuration generally performs better for 2-m dew point temperature with a few exceptions in the winter across the Eastern CONUS. Finally, for 10-m wind speed, there was a consistent high bias across the Eastern CONUS regardless of season, while for winter across the west there was a low bias (not shown). None of the pair-wise differences between the configurations were practically significant. A description of the test set-up, model forecast graphics, the full suite of verification results, the final report, and other supplementary information will be available later this year at http://www.dtcenter.org/eval/meso_mod/nmmb_test/nems_v0.9/index.php.

Table 3.1.2.1-3. Statistically significant (light shading) and practically significant (dark shading) differences for 2-m temperature, 2-m dew point temperature, and 10-m wind speed bias by season, region, and forecast lead time.

			f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
2-m Temperature	Summer	East	ThompMP	NAMOC	NAMOC	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *
		West	--	NAMOC	NAMOC	NAMOC	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *
	Winter	East	ThompMP *	ThompMP *	ThompMP *	--	NAMOC *	NAMOC *	NAMOC *	NAMOC *	NAMOC *	NAMOC *	ThompMP *	--	NAMOC *	NAMOC *	NAMOC *	NAMOC *
		West	--	--	--	--	--	NAMOC *	NAMOC *	NAMOC *	NAMOC *	NAMOC *	ThompMP *	--	NAMOC *	NAMOC *	NAMOC *	NAMOC *
2-m Dew Point	Summer	East	NAMOC	NAMOC	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *	ThompMP *
		West	NAMOC	--	--	--	--	ThompMP *	--	ThompMP *	--	--	ThompMP *	ThompMP *	ThompMP *	ThompMP *	--	--
	Winter	East	NAMOC	NAMOC *	NAMOC *	NAMOC *	--	--	--	--	--	NAMOC *	NAMOC *	--	--	ThompMP *	ThompMP *	--
		West	NAMOC	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
10-m Wind Speed	Summer	East	ThompMP	--	--	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP	--	--	ThompMP	ThompMP	ThompMP	ThompMP	ThompMP
		West	NAMOC	--	NAMOC	--	--	NAMOC	NAMOC	ThompMP	NAMOC	NAMOC	--	--	ThompMP	ThompMP	--	NAMOC
	Winter	East	--	NAMOC	--	--	ThompMP	NAMOC	NAMOC	--	--	--	--	--	--	NAMOC	NAMOC	--
		West	--	--	--	--	--	--	NAMOC	NAMOC	NAMOC	NAMOC	--	--	--	NAMOC	NAMOC	NAMOC

3.1.2.2 *NAM-RR functionally similar operational environment*

Through collaborations with EMC, work was conducted to establish a functionally similar operational environment (FSOE) for the NAM Rapid Refresh (NAM-RR) system (Fig. 3.1.2.2-1) on the NCAR supercomputer, Yellowstone, using the Rocoto Workflow Management System. Code for the hourly updating NAM-RR system was checked out of the EMC repository and ported to Yellowstone. Using sample initialization and observation data pulled from EMC, preliminary hourly end-to-end system testing using the NAM operational physics suite was successfully completed for the 12-km North American parent and 3-km CONUS nest domain. The next phase of the project will focus on adding the UPP and MET components to the workflow for evaluating the forecast system performance.

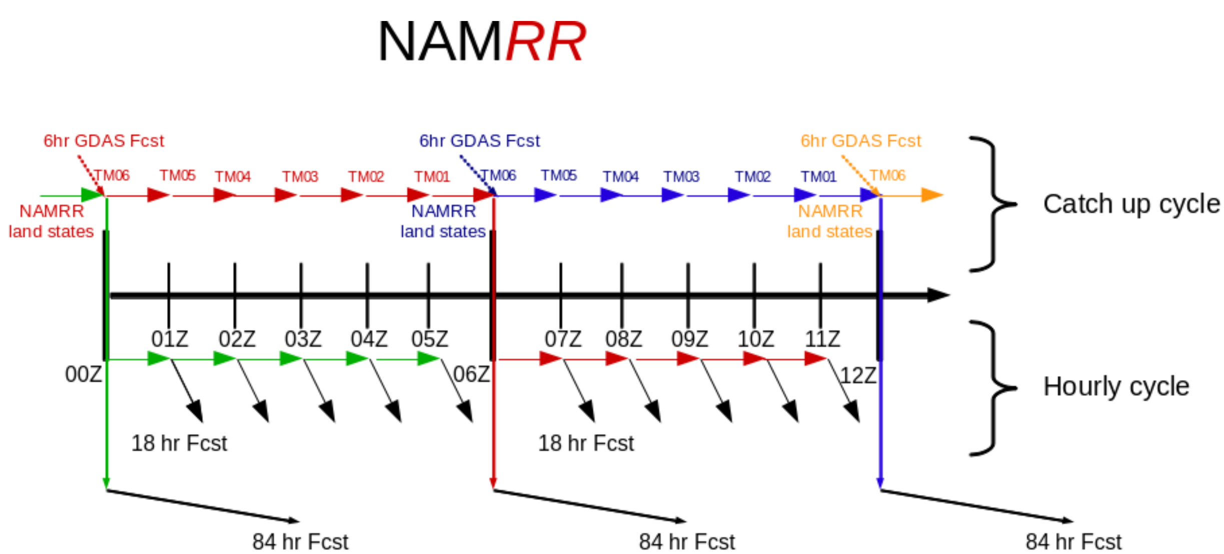


Figure 3.1.2.2-1. Schematic of the NAM-RR flow diagram. Courtesy of Jacob Carley, EMC.

3.1.3 WRF

3.1.3.1 WRF version T&E

To provide the Air Force with information regarding the progression of WRF code through time, the DTC tested the Advanced Research WRF (ARW) dynamic core with the two most recent releases of WRF (v3.5.1 and v3.6) using the Air Force operational configuration, which includes WSM5 (microphysics), Dudhia/RRMT (shortwave/longwave radiation), M-O (surface layer), Noah (land surface model), YSU (planetary boundary layer), and KF (cumulus). The test was set up and run in the same manner as testing conducted last year with three prior versions (v3.4, v3.4.1, and v3.5), and complements previous results. For each test, the end-to-end modeling system components were the same: WPS, WRF, the Unified Post Processor (UPP) and the Model Evaluation Tools (MET). Testing was conducted over two three-month periods (a warm season during July-September 2011 and a cool season during January-March 2012), effectively capturing model performance over a variety of weather regimes. To isolate the impacts of the WRF model code itself, 48-h cold start forecasts were initialized every 36h over a 15-km North American domain.

To highlight the differences in forecast performance with model progression, objective model verification statistics were produced for surface and upper air temperature, dew point temperature and wind speed for the full CONUS domain and 14 sub-regions across the U.S. Examples of the results (in this case, 2 m temperature bias) are shown in the figures below. A consistent cold bias is seen for most lead times during the warm season for all versions (Fig. 3.1.3.1-1). While there was a significant degradation in performance during the overnight hours with versions 3.4.1 and newer, a significant improvement is noted for the most recent version (v3.6) during the daytime hours. Examining the distribution of 2 m temperature bias spatially by observation site (Fig. 3.1.3.1-2), it is clear that for the 30-hour forecast lead time (valid at 06 UTC), v3.6 is noticeably colder over the eastern CONUS. However, for the 42-hour forecast lead time (valid at 18 UTC), v3.4 is significantly colder across much of the CONUS. For the full suite of verification results, please visit: www.dtcenter.org/eval/meso_mod/version_tne.

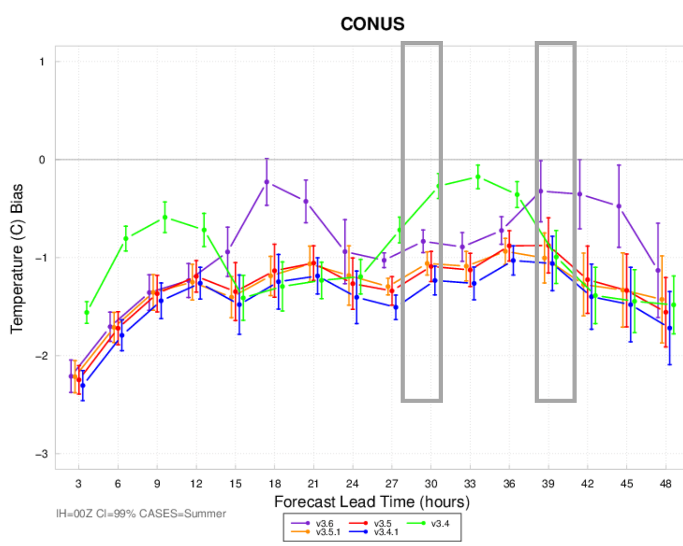


Figure 3.1.3.1-1. Time series of 2 m temperature (C) bias across the full CONUS domain over the warm season for WRF versions 3.4 (green), 3.4.1 (blue), 3.5 (red), 3.5.1 (orange), and v3.6 (purple). Median values of the distribution are plotted with the vertical bars representing the 99% confidence intervals. The gray boxes around forecast hours 30 and 42 correspond to the times shown in Fig. 3.1.3.1-2.

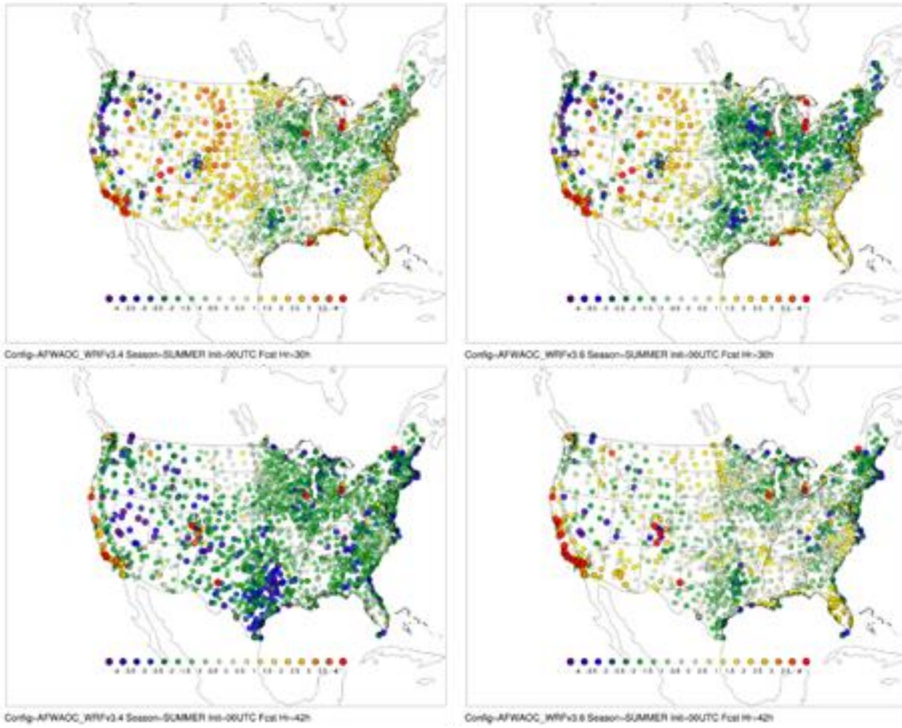


Figure 3.1.3.1-2. Average 2 m temperature (C) bias by observation station over the warm season for WRF version 3.4 (left) and 3.6 (right) at forecast hours 30 (top) and 42 (bottom).

3.1.3.2 WRF Inter-comparison T&E

For AOP 2014, the Air Force requested an assessment of a new combination of parameterizations for their operational configuration, including the Asymmetric Convective Model with non-local upward mixing and local downward mixing (ACM2) planetary boundary layer (PBL) scheme, the Pleim-Xiu land surface model (LSM) two-layer scheme with vegetation and sub-grid tiling, and the Pleim-Xiu surface layer scheme. A rigorous test and evaluation was conducted in a functionally similar operational environment; each configuration (Table 3.1.3.2-1) was initialized with a 6-hour "warm-start" spin up, including the GSI component. Forecasts were initialized every 36 hours from 1 August 2013 through 31 August 2014, for a total of 264 cases. Due to a large gap in input data from mid-June through mid-July, an additional summer month (August 2013) was included to provide a more complete analysis of the summer season. The core components of the end-to-end system included current versions of the WRF Pre-processing System (WPS), GSI, WRF, UPP, and MET software packages.

For the PBL and surface schemes inter-comparison, the largest impacts on forecast performance between the two configurations were seen at the surface and the lowest vertical levels, which was not unexpected given the focus of the test was PBL and surface schemes. A subset of the statistically significant (SS) and practically significant (PS) verification results are provided in Table 3.1.3.2-2. A large number of PS pair-wise differences were observed for the 2-m temperature and 2-m dew point temperature variables, while none of the pair-wise differences for 10-m wind speed were PS. For 2-m temperature, ACM2PX was the overall better performer, with exception of a few times valid around 15 – 21 UTC, where AFWAOC was superior. In the annual and summer aggregations, ACM2PX has higher median biases for 2-m dew point temperature than AFWAOC and has a wet bias at a majority of forecast lead times (e.g., Fig. 3.1.3.2-1). Due to the strong diurnal variability for both configurations, the favored configuration is dependent on forecast lead time and temporal aggregation. When considering the GO Index, AFWAOC is the better performer for summer and fall aggregations; whereas, ACM2PX has superior performance for the winter aggregation (Fig. 3.1.3.2-2). No SS differences are noted in performance between the two configurations in the annual and spring aggregations. A detailed report for the PBL and surface scheme inter-comparison is included in the supplemental materials. The full set of verification results are also available on the DTC Testing and Evaluation website (http://www.dtcenter.org/eval/meso_mod/afwa_test/). This website provides details on the inter-comparison results for the two WRF configurations tested during this period of performance (v3.6.1) along with inter-comparison testing and evaluation results from previous years.

Table 3.1.3.2-1. Physics suites used for AFWA Operational Configuration and ACM2PX replacement configuration testing and evaluation.

Physics Parameterization	AFWA Operational Configuration (AFWAOC)	ACM2PX Replacement Configuration (ACM2PX)
Microphysics	WRF Single-Moment 5 (opt. 4)	WRF Single-Moment 5 (opt. 4)
Surface Layer	Monin-Obukhov Similarity Theory (opt. 91)	Pleim-Xiu (opt. 7)
PBL	Yonsei State University (opt. 1)	Asymmetric Convective Model 2 (opt. 7)
Convection	Kain-Fritsch (opt. 1)	Kain-Fritsch (opt. 1)
Land-Surface Model	Noah (opt. 2)	Pleim-Xiu (opt. 7)
Radiation	New Rapid Radiative Transfer Model (LW/SW) (opt. 4)	New Rapid Radiative Transfer Model (LW/SW) (opt. 4)

Table 3.1.3.2-2-2. SS (light shading) and PS (dark shading) pair-wise differences for the AFWAOC and ACM2PX configurations run with WRFv3.5.1 (where the highlighted version is favored) for 2-m temperature and dew point temperature bias and 10-m wind speed bias by season and forecast lead time for the 00 UTC initializations.

	00 UTC Inits	f03	f06	f09	f12	f15	f18	f21	f24	f27	f30	f33	f36	f39	f42	f45	f48
2-m Temperature	Annual	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	--	--	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	--	--	ACM2PX *
	Summer	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	AFWAOC *	AFWAOC *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *
	Fall	--	--	--	ACM2PX *	AFWAOC *	AFWAOC *	AFWAOC *	ACM2PX *	--	--	ACM2PX *	ACM2PX *	AFWAOC *	AFWAOC *	AFWAOC *	ACM2PX *
	Winter	--	--	--	--	AFWAOC *	--	ACM2PX *	ACM2PX *	--	--	--	--	--	--	ACM2PX *	ACM2PX *
	Spring	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	--	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *
2-m Dew Point Temp	Annual	AFWAOC *	AFWAOC *	ACM2PX *	ACM2PX *	AFWAOC *	AFWAOC *	--	--	AFWAOC *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	--	--
	Summer	AFWAOC *	AFWAOC *	ACM2PX *	ACM2PX *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *
	Fall	AFWAOC *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	AFWAOC *	--	AFWAOC *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	ACM2PX *	--	--
	Winter	--	--	--	--	--	--	ACM2PX *	ACM2PX *	AFWAOC *	AFWAOC *	AFWAOC *	AFWAOC *	--	--	ACM2PX *	--
	Spring	AFWAOC *	AFWAOC *	ACM2PX *	ACM2PX *	AFWAOC *	--	ACM2PX *	ACM2PX *	--	--	--	--	--	--	--	--
10-m Wind Speed	Annual	--	AFWAOC	AFWAOC	--	AFWAOC	AFWAOC	AFWAOC	--	--	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC
	Summer	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC
	Fall	AFWAOC	AFWAOC	AFWAOC	--	AFWAOC	AFWAOC	AFWAOC	--	--	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	AFWAOC	--
	Winter	ACM2PX	ACM2PX	ACM2PX	ACM2PX	ACM2PX	AFWAOC	--	ACM2PX	ACM2PX	ACM2PX	ACM2PX	ACM2PX	ACM2PX	AFWAOC	--	ACM2PX
	Spring	ACM2PX	--	--	--	AFWAOC	AFWAOC	AFWAOC	AFWAOC	ACM2PX	--	--	--	AFWAOC	AFWAOC	AFWAOC	AFWAOC

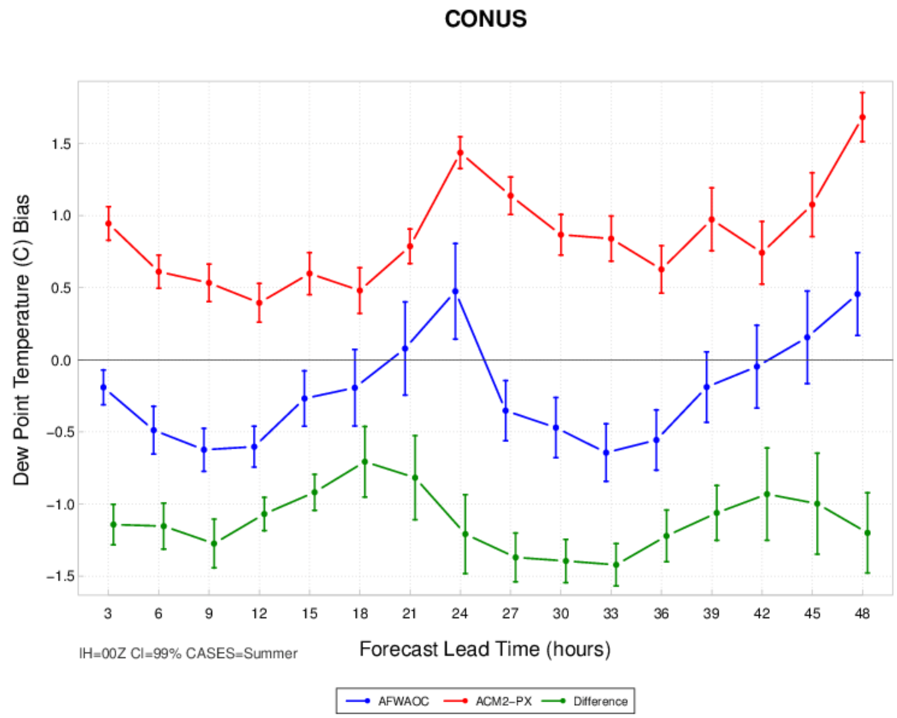


Figure 3.1.3.2-1. Time series plot of 2-m dew point temperature (°C) median bias for the CONUS domain aggregated across the summer season for the 00 UTC initializations. AFWAOC is in blue, ACM2PX in red, and the differences (AFWAOC-ACM2PX) in green. The vertical bars attached to the median represent the 99% CIs.

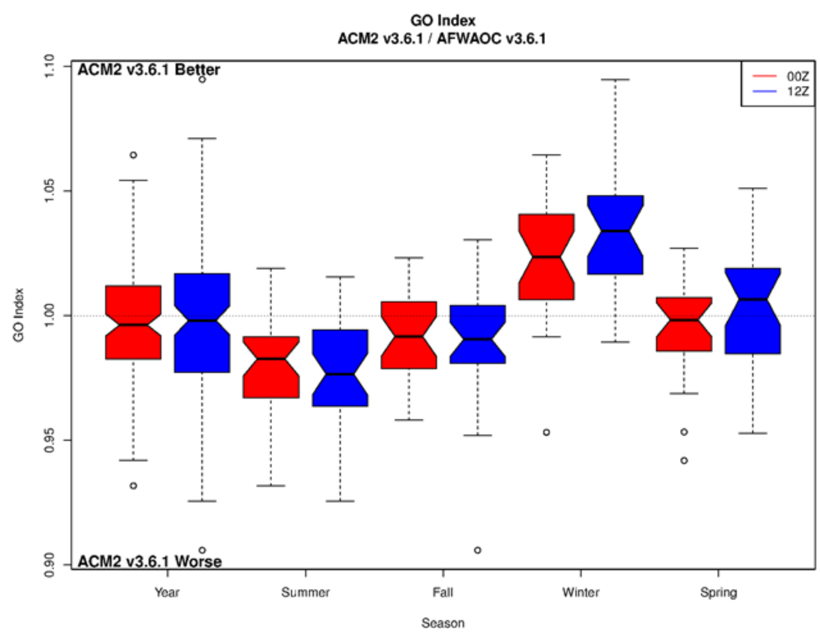


Figure 3.1.3.2-2. Boxplot of GO Index values aggregated across the entire year of cases and for all seasons, stratified by initialization time where 00 UTC is in red and 12 UTC is in blue. The median value is the thick black line located at the vertex of the notches, the notches around the median is 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

Work undertaken by the Hurricane team led to three papers that have been accepted for publication:

Bernardet, L., V. Tallapragada, S. Bao, S. Trahan, Y. Kwon, Q. Liu, M. Tong, M. Biswas; T. Brown, D Stark, L. Carson, R. Yablonsky, E. Uhlhorn, S. Gopalakrishnan, X. Zhang, T. Marchok; Y-H Kuo, and Robert Gall, 2015: Community Support and Transition of Research to Operations for the Hurricane Weather Research and Forecast (HWRF) Model, *Bull. Amer. Meteor. Soc.*, **96**, 953-960, doi: 10.1175/BAMS-D-13-00093.1 (http://journals.ametsoc.org/doi/pdf/10.1175/BAMS-D-13-00093.1).

Biswas, M. K., L. Bernardet, and J. Dudhia, 2014: Sensitivity of hurricane forecasts to cumulus parameterizations in the HWRF model, *Geophys. Res. Lett.*, **41**, 9113–9119, doi:10.1002/2014GL062071.

Yablonsky, R. M., I. Ginis, B. Thomas, V. Tallapragada, D, Sheinin, and L. Bernardet, 2014. Description and analysis of the ocean component of NOAA's operational HWRF model, *J. Atmos. Oceanic Technol.,* **32**, 144–163. (http://dx.doi.org/10.1175/JTECH-D-14-00063.1)

3.2.1 HWRF Quantitative Precipitation Forecasts

A comprehensive evaluation of HWRF Quantitative Precipitation Forecasts (QPF) was conducted for the 2014 HWRF model. The model output used for this evaluation, which was provided by EMC, consisted of pre-implementation runs for 22 storms from 2011 to 2013 Hurricane seasons and a subset of operational forecasts from the 2014 season. Three basic approaches were applied to gain insight into the performance of HWRF QPF: 1) a large-scale assessment looking at HWRF QPF accumulated over 24 h for the parent domain, 2) 24-h accumulations for a circular region with a 600-km diameter centered on the observed storm location (with and without accounting for storm location differences), and 3) run-total storm-scale QPF for the innermost domain with 3-km grid spacing. The Climate Prediction Center's CMORPH analyses and NCEP's Stage IV analyses (available only over the CONUS region) were used as sources of quantitative precipitation estimates (QPE) for these assessments.

For the large-scale assessment, the model QPF and the CMORPH and Stage IV analyses were re-gridded to a 0.25° "Mega Domain" (25S-60N, 150W-10E) defined by the range of HWRF parent domain locations for the sample. Operational Global Forecast System (GFS) forecasts were used as a baseline against which to benchmark the HWRF's parent domain precipitation. The full Mega Domain aggregations are based on CMORPH only. In addition to verifying for the full Mega Domain, statistics were also computed for sub-domains located over the Atlantic (AL) and eastern North Pacific (EP) basins and the CONUS. Statistics for the ocean basins are based on CMORPH QPE, whereas those for the CONUS are based on Stage IV QPE. These comparisons revealed that for most lead times and thresholds HWRF predicts precipitation events more frequently than observed, whereas GFS generally predicts precipitation events less frequently than observed (not shown). Over the CONUS, Equitable Threat Scores (ETS) for different thresholds indicate that GFS outperforms HWRF at all lead times, except for the higher thresholds (2 and 3") at 96 and 120 h (not shown).

The location of the forecasted precipitation associated with a tropical cyclone (TC) will be strongly dependent on the track forecast. Hence, assessments of storm-centric precipitation forecasts based on grid-to-grid comparisons will be influenced by both track errors and errors in the representation of the storm structure. To quantify the impact of differences in storm location on QPF skill for the storm-centric assessment, the entire QPF field was horizontally shifted to match the observed storm location (Marchok et al. 2007). Verification statistics for this approach used the CMORPH QPE regridded to the 0.25° Mega Domain for the entire evolution of the storm. Figure 3.2.1-1 shows the ETS for GFS and HWRF without and with the adjustments for storm location differences. In general, GFS shows higher forecast skill than HWRF. The increase in ETS shown for both models after the shifting to account for track errors reflects the contribution from storm location differences to the QPF error. It is important to keep in mind that these location adjustments do not account for QPF differences stemming from the storm interacting with a different local environment (e.g., terrain impacts, surface fluxes, etc.).

For the run-total storm-scale QPF (a.k.a swath data) assessment, the CMORPH and Stage IV data were re-gridded to 0.05° to match the grid-spacing of HWRF's innermost domain. Bands 50 km wide were drawn out to 400 km around the forecasted storm track for the HWRF data and around the Best Track for the observed datasets. Figure 3.2.1-2 shows these bands for a single Hurricane Sandy forecast. The QPE distributions are based on CMORPH when the forecasted track was over water, and Stage IV when it was over land. The boxplots in Fig. 3.2.1-2 show the QPF and QPE distributions, separately over land and water, for the different bands including the full swath (0-400 km). As shown in this example, the bulk of observed and predicted rainfall occurs within 0-100 km, which is consistent with previous studies. When combined for the whole storm, in general, HWRF over-estimates precipitation over land and water, but the over-prediction is larger over water (not shown).

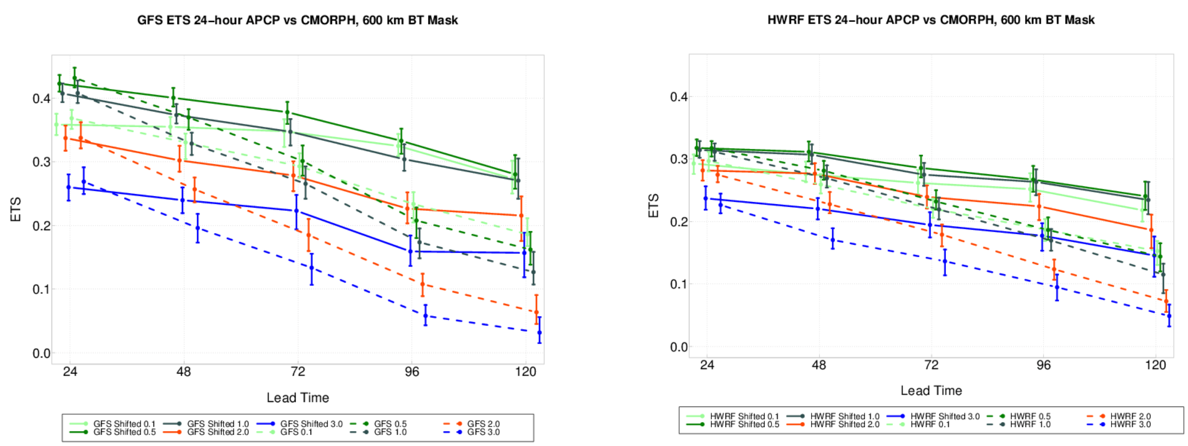


Figure 3.2.1-1. Equitable threat scores with respect to lead time for GFS (left) and HWRF (right) before shifting (dashed lines) and after shifting (solid lines) of the grids. Verification was computed using CMORPH data over a 600 km mask around the observed storm center. The thresholds are 0.1" (light green), 0.5" (dark green), 1.0" (army green), 2.0" (red) and, 3.0" (blue).

3.2.2 Rapid Intensification Forecasts

Rapid intensity change is a major challenge for TC prediction. Rapid intensification (RI) is defined as an intensity increase of 30 kt or more over-water in 24 h (Kaplan and Demaria 2003). These events are rare and difficult to predict. For this study, evaluations were done for the AL and EP basins using HWRF retrospective runs from the 2014 Stream 1.5 exercise (sample includes storms from 2011-2013 Hurricane seasons), as well as real-time HWRF forecasts during the 2014 Hurricane season. Given the higher frequency of RI events in the western North Pacific (WP) basin, HWRF's ability to capture RI events was also evaluated over the WP region for the 2013 and 2014 seasons. The Probability of Detection (POD) for the AL and EP basins combined was 7.3%, whereas the POD for the WP basin was 28.7%. The False Alarm Rates (FAR) were 66.4% for the AL and EP basins combined and 43% for the WP basin. Thus, HWRF appears to have higher skill for RI events over the WP basin than that for the AL and EP basins. Note that the sample size for the AL/EP aggregation is significantly larger than that for the WP aggregation. Furthermore, the AL/EP basin sample includes storms from four Hurricane seasons, and the WP basin sample only includes storms from two seasons. Hence, the AL/EP sample likely captures a larger degree of inter-seasonal variability. Given forecast skill can vary from year-to-year due to inter-seasonal variability in TCs, caution should be exercised when comparing the HWRF skill for these basins based on these samples.

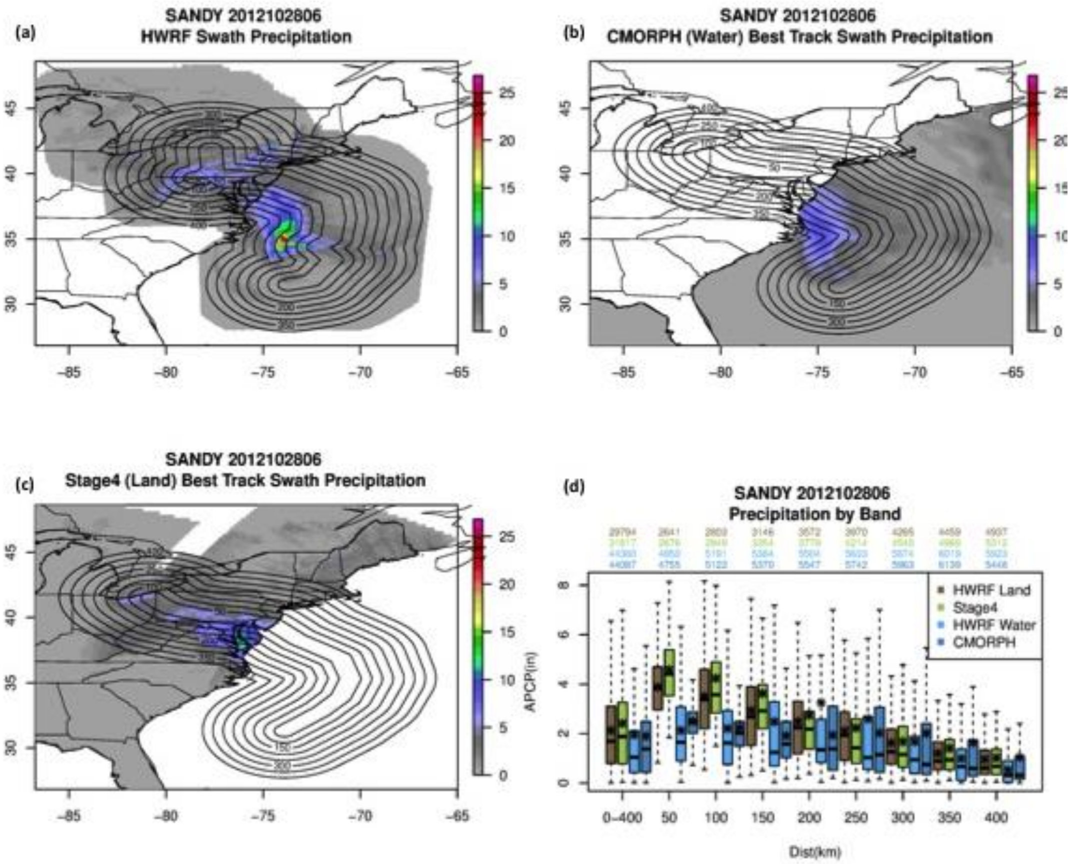


Figure 3.2.1-2. Precipitation (mm) swath data accumulated over 5 days for the Hurricane Sandy forecast initialized at 06 UTC on 28 October 2012 for (a) HWRF, (b) CMORPH over water, (c) Stage IV over land. The bands shown are 0-50, 50-100, etc. out to 350-400 km around the forecasted storm track for (a) and Best Track for (b) and (c). The boxplot distributions of the precipitation are shown in (d). The mean is shown as asterisk (*). Precipitation (inches) for HWRF over land (brown), Stage IV over land (green), HWRF over water (light blue) and CMORPH over water (dark blue) are shown for different bands shown on the x-axis. The numbers of points taken into consideration for the four boxes are shown on the top of each band.

One drawback to categorical statistics is they do not provide information about the degree to which the forecast missed or matched the defined threshold. Considering the difference between the 24-h intensity change predicted by HWRF and that analyzed in the Best Track for the four categories in the contingency table provides useful information for this type of assessment. The boxplots in Fig. 3.2.2-1 show the distributions of these differences by category and lead time. Distributions for which the median is not distinguishable from zero (i.e., waist or notch of box includes zero) indicate the predicted and analyzed 24-h intensity change are similar. Distributions with a median greater than zero (i.e., entire notch is above zero) indicate HWRF over-predicted the intensity change, whereas distributions with a median less than zero (i.e., entire notch is below zero) indicate HWRF under-predicted the intensity change. For the AL and EP basins combined, the medians of the hits (green) are below the zero line indicating that when HWRF detects RI, it under-predicts its magnitude. HWRF over-predicts intensity change by 15-20 kt for False Alarms (red) and under-predicts by 20-25 kt when it misses (orange) RI events. On the other hand, when validated over the WP basin, the median of the hits is not statistically distinguishable from zero. The misses and the false alarms for the WP basin are similar to that for the AL and EP basins. An evaluation was also conducted using 2015 pre-implementation tests provided by EMC. This evaluation, which was based on a homogeneous comparison, indicated the 2015 HWRF performs better at capturing intensity change than the 2014 HWRF. The data set for the Stream 1.5 exercise also provided the opportunity to compare the HWRF performance to other HFIP models. This comparison indicated HWRF has the highest skill for predicting intensity change (not shown).

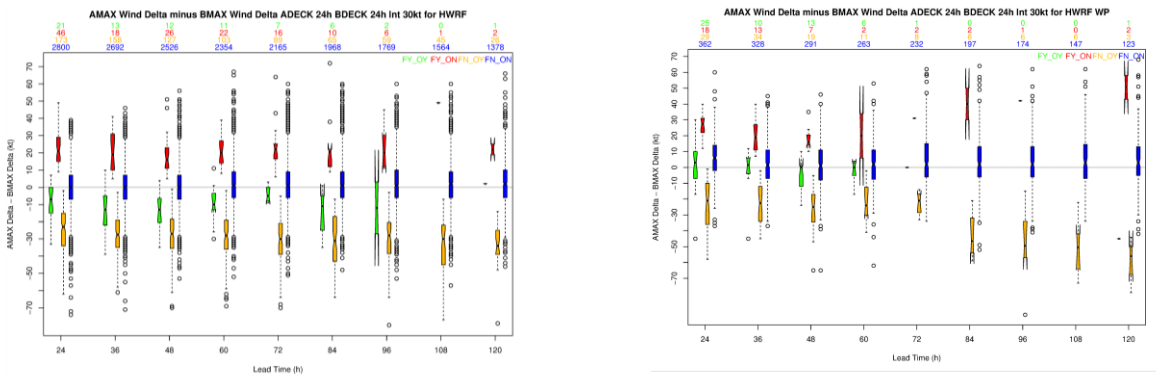


Figure 3.2.2-1 Boxplots showing the 24-h intensity-change differences between the HWRF model and the Best Track estimate for the AL and EP basins combined (left) and the WP basin (right). The four boxplots for each lead-time are hits (green), false alarms (red), misses (orange) and correct negatives (blue). The number of cases is shown on the top for each category.

3.2.3 Advancing the Connections between Radiation and Clouds in HWRF

During AOP 2014, the DTC extended its work toward connecting the RRTMG radiation scheme with the SAS convection parameterization. To account for the missing connection between convective clouds and the RRTMG radiation scheme, a cloud fraction scheme based on Sundqvist et al (1989) was implemented in WRF combined with a first-guess "scale-aware" relative humidity threshold that requires higher humidity values to make sub-grid clouds as the resolution increases. This scheme includes an algorithm to specify the liquid and ice water content of the partial clouds based on the vertical profile of temperature and humidity at each grid point, which provides the information needed by the RRTMG scheme. The water vapor and explicit cloud variables were not modified. The only changes were the incorporation of the partial cloudiness scheme into the shortwave and longwave radiation treatment that subsequently causes the explicit microphysics scheme to produce its own clouds. The existing WRF RRTMG scheme has a cloud fraction scheme attributed to Xu and Randall (1996), however, in practice, the scheme produces a binary 0 or 100% cloud fraction based on the absence or presence of explicit cloud condensate by the microphysics scheme.

An extensive T&E activity comparing an HWRF control configuration using the GFDL radiation scheme and an experimental HWRF configuration using the RRTMG radiation scheme and the new cloud fraction parameterization found the tracks errors were similar for the two configurations, whereas the experimental configuration was able to reduce the intensity bias associated with the control configuration in the EP basin. An evaluation of the HWRF large-scale forecasts on the parent domain against GFS analyses, performed using MET, indicated small improvements when the RRTMG and partial cloudiness parameterizations were employed. The innovative display of results highlighting the geographical distribution of bias and RMSE, along with summary plots displaying the distribution of errors with vertical level and forecast lead time (see Fig. 3.2.3-2), was well received, and led to requests by EMC for evaluations of their HWRF 2015 pre-implementation results. Collaborators from CIRA-CSU also participated in this evaluation by comparing the control and experimental HWRF forecasts against GOES water vapor and infrared channels brightness temperatures. Results indicated the experimental configuration vastly improved the IR synthetic satellite brightness temperatures generated by HWRF, while the water vapor results were somewhat degraded (not shown). Based on the outcome of this test and additional tests conducted by EMC, the RRTMG combined with the new cloud fraction scheme were accepted for operational implementation in the 2015 HWRF.

Details of the cloud fraction scheme implementation and the results of the tests are available at http://www.dtcenter.org/eval/hwrf_hdrf_hdgf/.

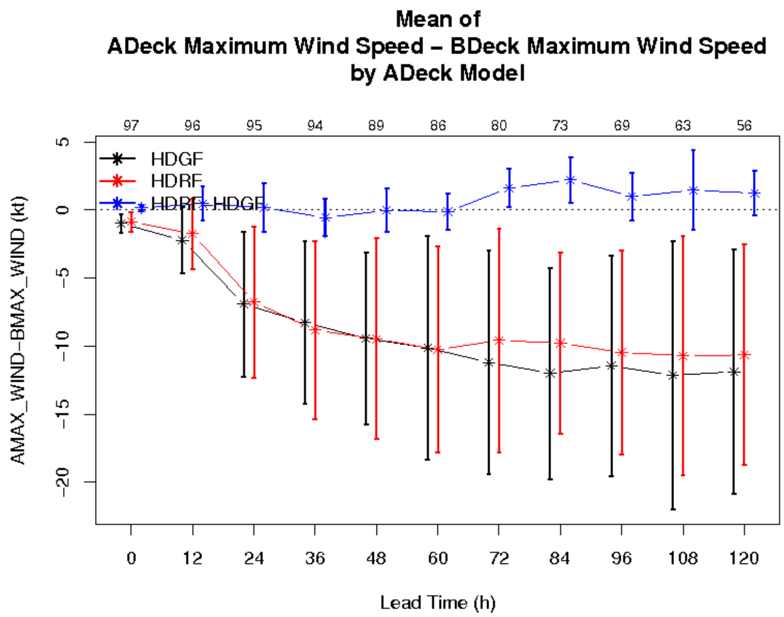


Figure 3.2.3-1. Intensity (maximum 10-m wind) bias in the EP basin for the HWRF control configuration using the GFDL radiation parameterization (black) and the HWRF experimental configuration using the RRTMG radiation and cloud fraction parameterizations (red) as a function of forecast lead time (h). Pairwise differences are shown in blue, and the sample size is listed above the plot.

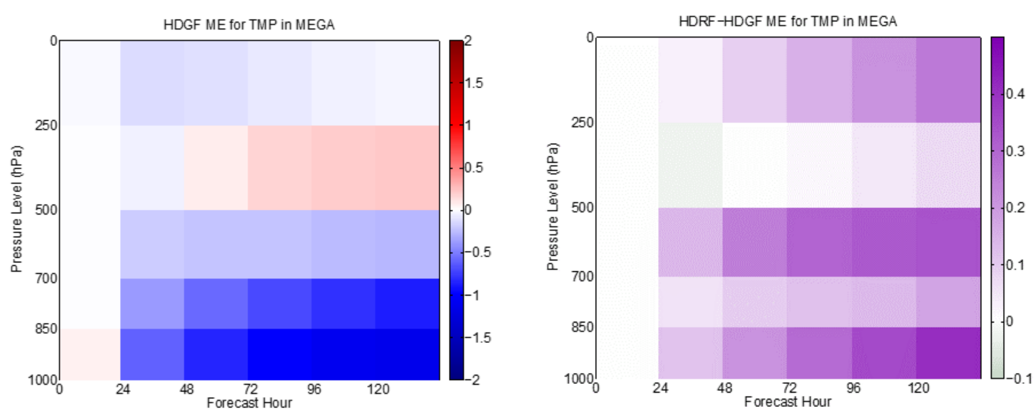


Figure 3.2.3-2. Verification of HWRF temperature forecasts over the mega-domain as a function of forecast lead time (x-axis) and vertical level (y-axis). Left plots displays bias (K) for the control configuration using the GFDL radiation parameterization, and right plot displays the difference in bias (K) between the experimental configuration using the RRTMG radiation and cloud fraction parameterizations and the control. Results indicate that the control is cold (blue) for most levels and lead times, and that the experimental configuration alleviates this bias.

3.3 Data Assimilation

During AOP 2014, the DA team submitted the following paper to the *Bulletin of the American Meteorology Society*:

The Community Gridpoint Statistical Interpolation (GSI) Data Assimilation System: Bridging the Gap between the Research and Operational Communities, by Hui Shao, John Derber, Xiang-Yu Huang, Ming Hu, Donald Stark, Michael Lueken, Kathryn Newman, Chunhua Zhou, Louisa Nance, Ying-Hwa Kuo, and Barbara Brown.

One paper associated with past DA activities was accepted by *Weather and Forecasting*:

Newman, K. M., C. S. Schwartz, Z. Liu, H. Shao, and X.-Y. Huang, 2015: Evaluating Forecast Impact of Assimilating Microwave Humidity Sensor (MHS) Radiances with a Regional Ensemble Kalman Filter Data Assimilation System

3.3.1 GSI-Hybrid System for Hurricane WRF

In consultation with EMC, the DTC AOP 2014 T&E activities related to the application of the GSI-hybrid system for the HWRF application focused on diagnostics for the 2014 HWRF spin-down issue and investigating the current two-way GSI-hybrid capabilities. As a first step, the DTC completed baseline tests for the HWRF 2014 pre-implementation cases to "reproduce" the HWRF pre-implementation results. The DTC then performed further diagnostics and, subsequently, selected Hurricane Irene (2011) for more detailed study. This detailed study indicated the spin-down behavior stems from an imbalance between the vortex initialization and the DA steps. The inner core DA and its interaction with the vortex initialization are the key components for this spin-down issue. Figure 3.3.1-1 shows the surface pressure change rate for each time step. The DA step is the largest contributor to the sharp changes in surface pressure at the initial stage of the forecasts. This outcome indicates the analysis from the DA step shocked the forecast model due to lack of appropriate balancing. Based on these results, the DTC decided to further investigate potential measures to improve the inner core DA and its balancing.

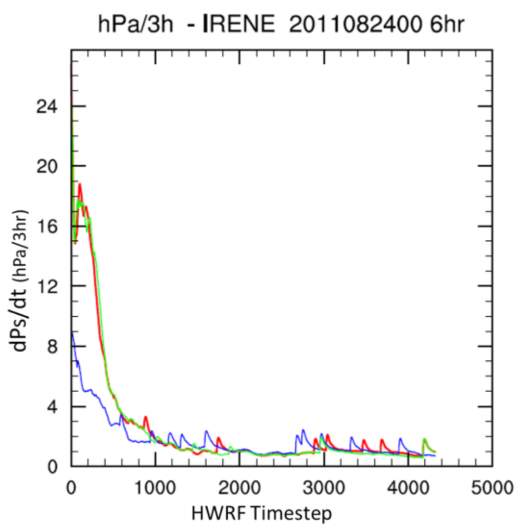


Figure 3.3.1-1. Change in surface pressure (dPs/dt) with time steps. Red curves indicate the results from the runs with HWRF 2014 default configuration (including both vortex initialization and data assimilation). Blue and green curves indicate the results from the denial experiments with only vortex initialization (blue) and only data assimilation (green) in the system configuration.

The DTC studied the balance algorithm as part of the HWRF vortex initialization procedure and the potential to apply this information to the GSI DA as a post-processing step. The DTC also studied and tested the impacts of applying a tangent-linear normal-model constraint (TLNMC) to the analysis for a TC forecast. Studies showed that it is possible to apply TLNMC to the TC case and the constraint may help reduce the noise level produced by the inner core DA (Figure 3.3.1-2) and improve the intensity forecasts (Figure 3.3.1-3).

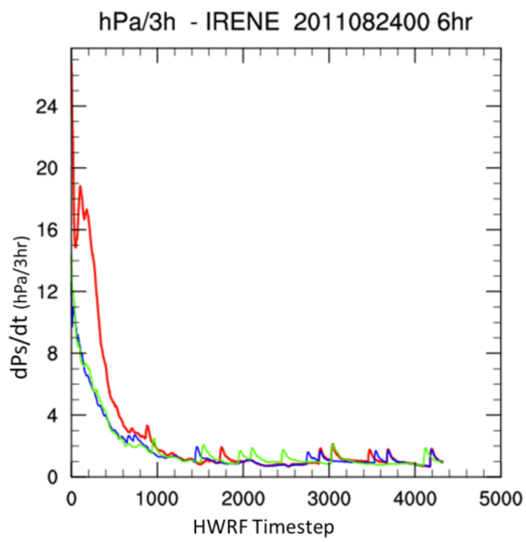


Figure 3.3.1-2. Same as Fig. 3.3.1-1, except the blue and green curves indicate the results with TLNMC options 1 (blue) and 2 (green) applied.

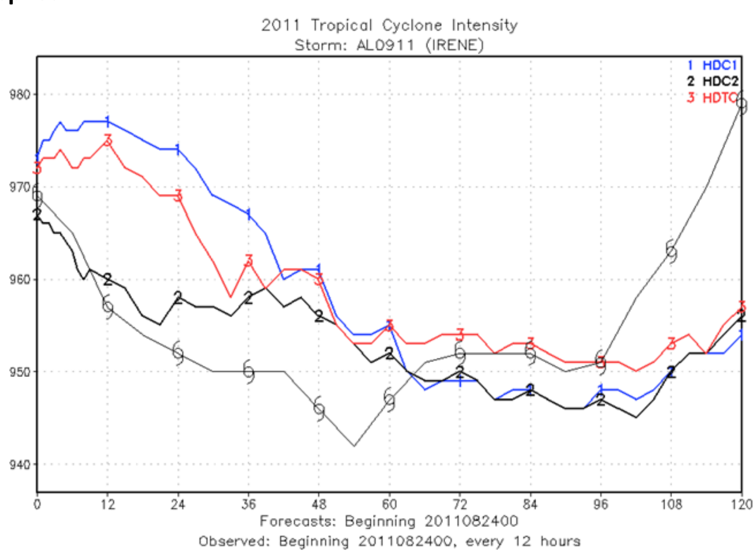


Figure 3.3.1-3. Analyses and forecasts of the TC intensity generated by the runs with default configuration (red), TLNMC option 1 (blue), and TLNMC option 2 (black).

During AOP 2014, the DTC also investigated ways to improve how the flow-dependent background error is prescribed. A quick study performed by the DTC indicated using the HWRF regional ensemble (instead of GFS ensemble) in the one-way hybrid DA to define the flow-dependent portion of the background error may improve the TC intensity forecasts. Cycling update of the regional ensemble may further improve the forecasts. Therefore, the DTC continued to build a complete two-way hybrid system for

HWRF. Initial tests of the system produced reasonable results. The DTC is currently running experiments with both one-way and two-way hybrid techniques and working on the evaluation of their impacts on the inner core DA. This work is expected to be complete by end of June 2014, at which time the full report will be posted on the DTC website.

3.3.2 Regional Ensemble for Data Assimilation

Most of NCEP's current NWP suites use the GSI DA in hybrid mode, combining background error covariance from static sources (computed using forecast errors averaged over long periods) with flow-dependent sources (computed using the perturbations of the ensemble forecast for specific locations and times). The NCEP regional NWP suites, such as the Rapid Refresh (RAP) and HWRF, ingest a global low-resolution ensemble, the GFS EnKF ensemble. We hypothesize that a high-resolution ensemble could improve the analysis, and therefore the forecast, by providing additional information to the DA system. This activity was aimed at developing an initial capability toward a high-resolution ensemble for DA. To save computational resources, this high-resolution ensemble was produced only over the domain of the regional forecast model that will use it for DA.

For demonstration purposes, the DTC selected the RAP suite (which utilizes the ARW model) for development and preliminary testing of the regional ensemble for DA. Perturbations from the GFS EnKF ensemble 6-h forecasts were used to create diversity in initial conditions for the ARW ensemble in the RAP domain. Forecasts from this ensemble was used in the RAP DA. A substantial part of the DTC work was devoted to developing a utility that reads the spectral coefficients for the GFS ensemble forecasts, converts the meteorological fields to the ARW grid, subtracts the GFS ensemble members from the mean to create perturbations, and creates the ARW initial conditions.

A preliminary evaluation was performed on a case study to compare the DA increments of RAP ingesting the GFS ensemble (control) versus the RAP ensemble. Figure 3.3.2-1, which shows the analysis increments at the 15th model level from both runs, demonstrates that the GSI-hybrid using the RAP ensemble generated increments provides a reasonable distribution. While the analysis increments from the runs ingesting GFS and RAP ensembles have similar patterns, the run ingesting GFS ensemble has larger analysis increments and more flow-dependent details. Therefore, the RAP ensemble had less impact than the GFS ensemble in the RAP analysis. This result warrants additional investigation, and future work could include an in-depth examination of the differences between perturbations generated with the regional RAP ensemble versus the GFS ensemble.

3.3.3 GSI Sensitivity Tests

Observation sensitivity tests were performed for the Air Force to assist with data and system configurations that optimize the use of new and proposed sources of data for GSI. The Air Force requested sensitivity tests for at least two data types from the following list:

- Global Change Observation Mission-W1 (GCOM-W1) Advanced Microwave Scanning Radiometer 2 (AMSR2)
- NPOESS Preparatory Project (NPP) Cross-Track Infrared Sounder (CrIS)
- NOAA-16/18/19 Solar Backscatter UV/2 (SBUV/2)
- METOP-A Global Ozone Monitoring Experiment-2 (GOME-2)

GCOM-W1 AMSR2 was eliminated from the list after the DTC confirmed with Joint Center for Satellite Data Assimilation (JCSDA) and National Centers for Environmental Prediction (NCEP) Environmental Modeling Center (EMC) colleagues that this data type was not yet available for ingest into GSI and would not be ready within the period of performance. The DTC performed sensitivity studies for the remaining three data types in the list. To enable the radiance and ozone DA, the DTC also performed tests to evaluate the latest ARW release with the model top increased from 10 hPa to 2 hPa.

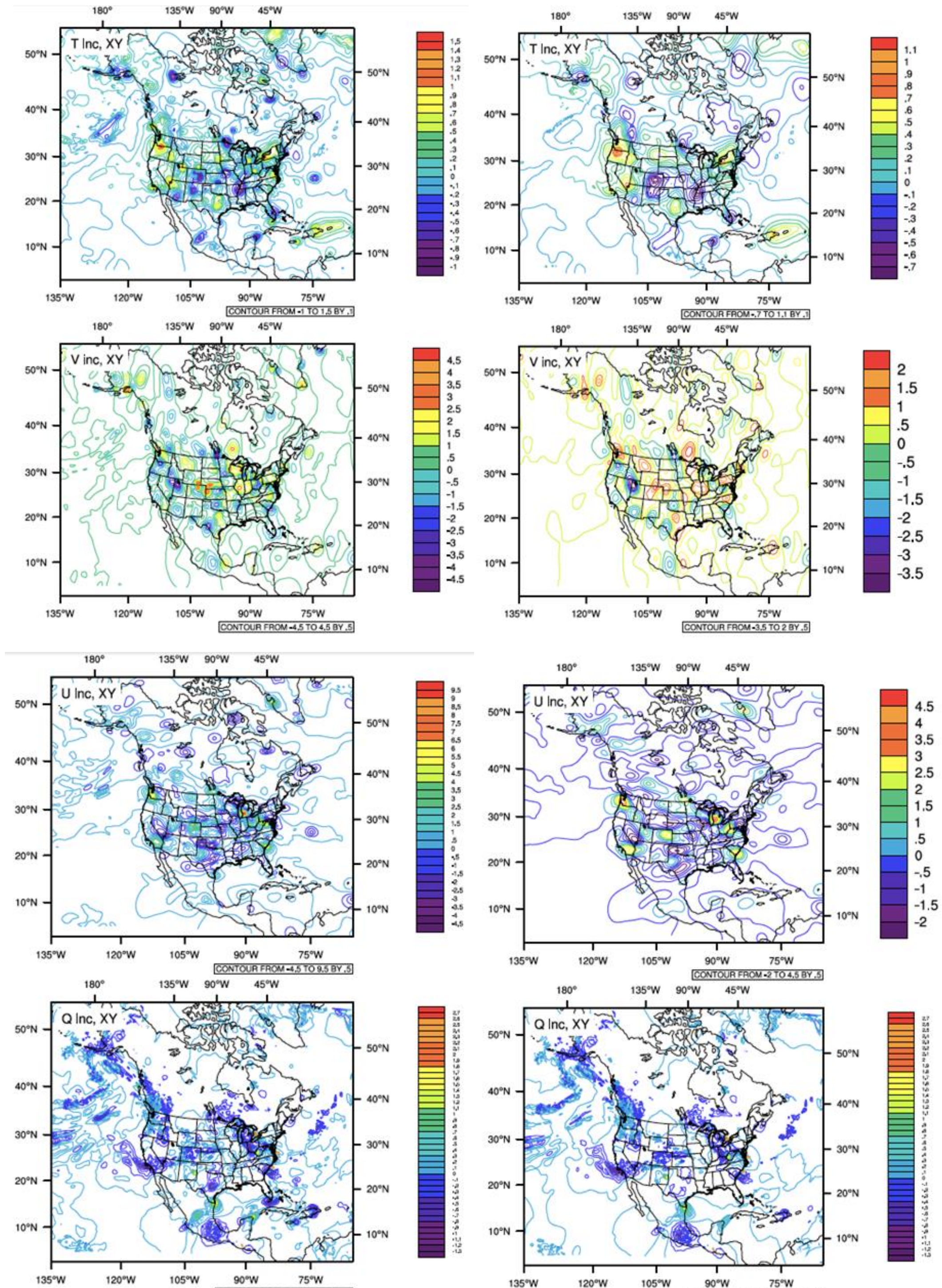


Fig 3.2.3-1: GSI-hybrid analysis increments at the 15th model level valid at 12Z on 25 March 2105 with the GFS (left) and RAP (right) ensembles. The rows depict, from top to bottom, temperature (K), meridional wind (ms^{-1}), zonal wind (ms^{-1}), and mixing ratio (gkg^{-1}). Note the color scales differ among the plots.

The model top increase generally showed positive impacts on the data assimilation and forecasts. Figure 3.3.3-1 shows the root-mean-square errors (RMSE) of temperature forecasts before and after the model top was increased. Statistically significant (SS) improvements associated with simply increasing the model top were found at both upper and lower levels.

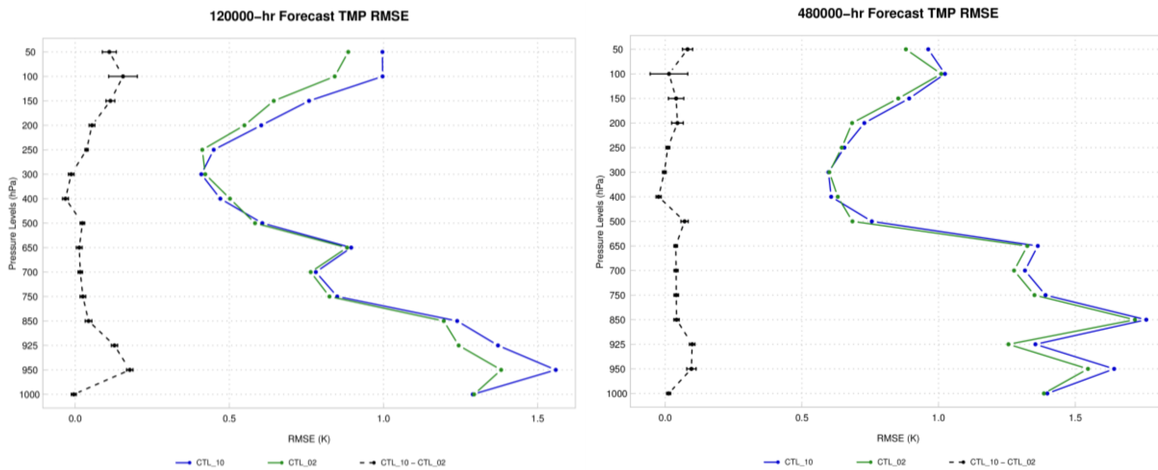


Figure 3.3.3-1: Vertical profiles of RMSE for 12 (left) and 48 (right) hour temperature forecasts with 2 hPa model top (green) and 10 hPa model top (blue). The pairwise differences are indicated by the black dashed line, where the difference is SS if the confidence intervals do not encompass zero. Positive differences correspond to the 2 hPa model top producing lower RMSE. SS is determined at the 99% level.

When focusing on the utility of ozone data in GSI coupled with ARW, there are signs of improvement, particularly in the upper levels of the earliest lead times for temperature and wind. It should be noted that ozone is not a prognostic variable in ARW. Therefore, to assimilate ozone data, NCEP's Global Forecast System (GFS) ozone was added as the background. In addition, the RRTMG radiation scheme was used instead of RRTM/Dudhia schemes for both the control and ozone DA experiments. Impacts of ozone data assimilation on forecasts are indirect through the impacts of the radiation computation inside the radiative transfer model used by the GSI. Compared with GOME, assimilation of SBUV shows more promise with a slightly stronger signal for improvement and larger SS differences over the control configuration. Figure 3.3.3-2 shows times series of the RMSE for the temperature analyses and forecasts at 50 hPa and 500 hPa. Clear improvement is present due to the SBUV assimilation in this figure.

The DTC discovered neutral forecast impacts from CrIS data assimilation over the current operational suite, which includes both AIRS and IASI. The data overlap of CrIS and AIRS data may lessen the impact of CrIS data. Also, the DTC discovered the channel selection for the CrIS configuration may need some careful consideration to improve the utility of these data (through the FSO tool as follows).

While the DTC continued to verify results through traditional scores (bias, RMSE, etc), the DTC also investigated the potential to utilize the Forecast Sensitivity to Observation (FSO) tool developed by the National Center for Atmospheric Research (NCAR) Mesoscale and Microscale Meteorology (MMM) Laboratory. The adjoint model required by FSO is not available for the latest community GSI release v3.3, which was used for all of the above-mentioned sensitivity tests. Therefore, the DTC used the matched version GSI v3.2 for this FSO study. While the mismatch between the GSI versions used by the tests discussed above and FSO will alter the FSO results, most of the updates between these two versions are not related to the ozone and CrIS data assimilation. Hence, the FSO results can still be used as a reference. The DTC discovered the FSO tool shows relatively consistent results to those in the impact studies. The advantage of using the FSO is the ability to examine the impact of data in much more detail. However, the lack of a timely update for the adjoint code could be an issue for future use of this tool.

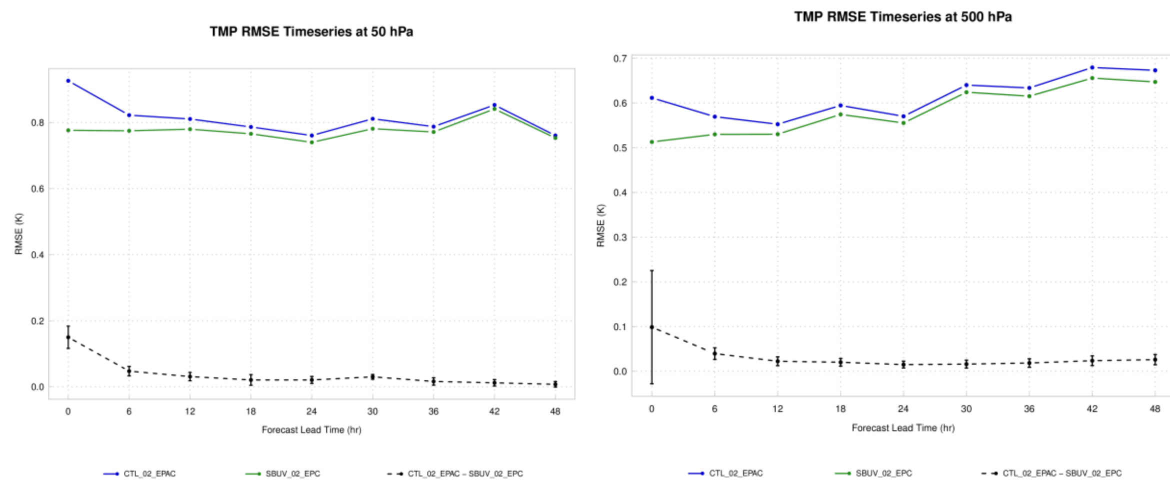


Figure 3.3.3-2: Time series of the RMSE for 50 hPa temperature (left) and 500 hPa temperature (right) generated from the control runs (blue) and SBUV assimilation runs (green). The pairwise differences are indicated by the black dashed line, where the difference is SS if the confidence intervals do not encompass zero. Positive differences correspond to the 2 hPa model top producing lower RMSE. SS is determined at the 99% level.

The FSO capability was applied to the sensitivity studies for both SBUV ozone and CrIS radiance data assimilation. Figure 3.3.3-3 shows forecast sensitivities to observations (per observation point) for each of the CrIS channels. Clearly, certain channels impose negative impacts on the forecasts. Therefore, the DTC recommends further study be performed to examine the current channel selection for CrIS data assimilation.

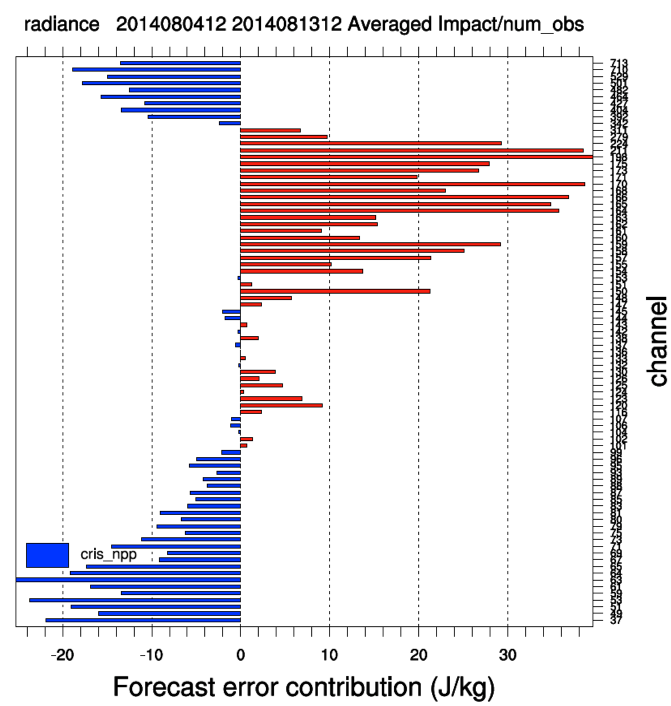


Figure 3.3.3-3: Observation impact from different CrIS channels on the 12-hour forecast initialized at both 00 UTC and 12 UTC for the period of 04-13 August 2014. Negative values (blue) indicate positive data impacts and positive values (red) indicate negative data impacts.

3.3.4 Air Force GSI Support

Two mitigation requests from the Air Force were received and solved during this performance period. The issues were associated with the reported Sea Level Pressure (SLP) anomaly and the operational CrIS data usage.

The DTC received the original request from the Air Force regarding problems with the SLP field in July 2013 when implementing GSI on the T4 southwest (SW) Asia domain. Significant progress on this issue was made during the last performance period, but that work pointed to the need for further investigation beyond the data assimilation step because SLP is neither an analysis variable nor a forecast variable (not directly updated by the data assimilation system, GSI or the forecast model, ARW). As requested by the Air Force, the DTC continued to investigate this issue through examination and testing of the associated surface pressure post-processing procedure and diagnostic field formulation inside ARW, GSI, and the former production DA system, WRFDA. The DTC discovered the mismatch between the GSI analysis variables and

1) the ARW prognostic and diagnostic variables
2) the SLP computation formulation in the post-processing procedure

contributed to this SLP problem. Prior to performing forecasts, ARW computes full fields for its prognostic and diagnostic variables based on the input file at analysis time. This input file can come from either the WRF Preprocessing System (WPS) or a data assimilation system (GSI or WRFDA). Table 3.3.4-1 shows a summary of the control and prognostic variables for the DA and forecast systems, respectively, as well as the computed or diagnostic variables for each system. As shown in the table, ARW expects an update of geopotential height (ϕ), dry air mass in column (μ), and potential temperature (ϑ) from the input file. Geopotential height is not among the analysis variables used in GSI or WRFDA. However, WRFDA updates geopotential height through an extra subroutine posterior to the analysis update. Therefore, WRFDA does not produce the same issues that GSI does due to lack of the update of geopotential height from GSI. To seek a solution for this issue, the DTC referred to the NOAA RAP system, which is similar to the Air Force operational system in that the GSI is coupled with the ARW dynamical core. This investigation revealed that the RAP system has similar issues unless an extra rebalance step is performed posterior to running the GSI analysis, which computes full field diagnostic variables based on the data assimilation analysis. The DTC followed a similar procedure and tested the rebalance code for the Air Force case.

Table 3.3.4-1 Summary of differences in WRFDA, GSI, Rebalance, and ARW formulas

	WRFDA	GSI	RAP Rebalance	ARW
Control/prognostic variables	$\Delta T\ \Delta P_s\ \Delta q$	$\Delta T\ \Delta P_s\ \Delta q\ \Delta\mu$	$T\ \mu\ q$	$\phi\ \mu\ \theta$
Computed/diagnostic variables	$\Delta\theta\ \Delta P\ \Delta\phi\ \Delta\mu$	$\Delta\theta$ (from ΔT)	$P\ \alpha\ \phi$	$\alpha\ P$

For the posting-processing procedure, the DTC discovered the reported SLP spurious anomaly was not present unless the following aspects were considered in the post-processing procedure (using either the Air Force WRF Post-Processing (WPP) code or the Unified Post-Processing (UPP) system):

1) The input file for WPP/UPP must be the ARW output file at the analysis time (i.e. 0-hour forecast file generated by ARW, not wrfinput directly generated from GSI analysis);

2) Perturbation pressure at the lowest model level needs to be used for surface pressure in the SLP computation, not dry air mass or the surface pressure directly from GSI.

The DTC modified the UPP code to meet the above requirements and evaluated the impact of the rebalance step on the SLP computation. Figure 3.3.4-1 shows the mean SLP (MSLP) with and without rebalance. The results show a smoother MSLP field with fewer erroneously high MSLP values. Though the MSLP field shows improvement, the rebalance step actually pushes the resulting analysis closer to the background. This results in a degradation of the forecast when compared with the observations as shown in Fig. 3.3.4-2. Further study is recommended to examine how to appropriately perform such a rebalance (e.gl, applied to increment fields or full fields) and to which fields it should be applied. Furthermore, surface observation internal QC (FY2013 work) should be implemented to prevent bad surface observations from being ingested into GSI.

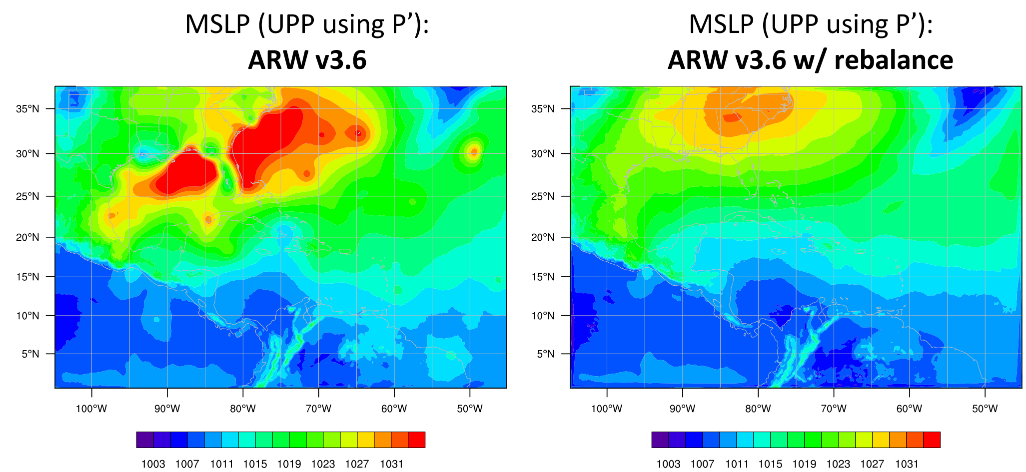


Figure 3.3.4-1: Mean SLP field calculated using 1st level pressure perturbation (P') for the surface pressure without (left) and with (right) rebalance applied.

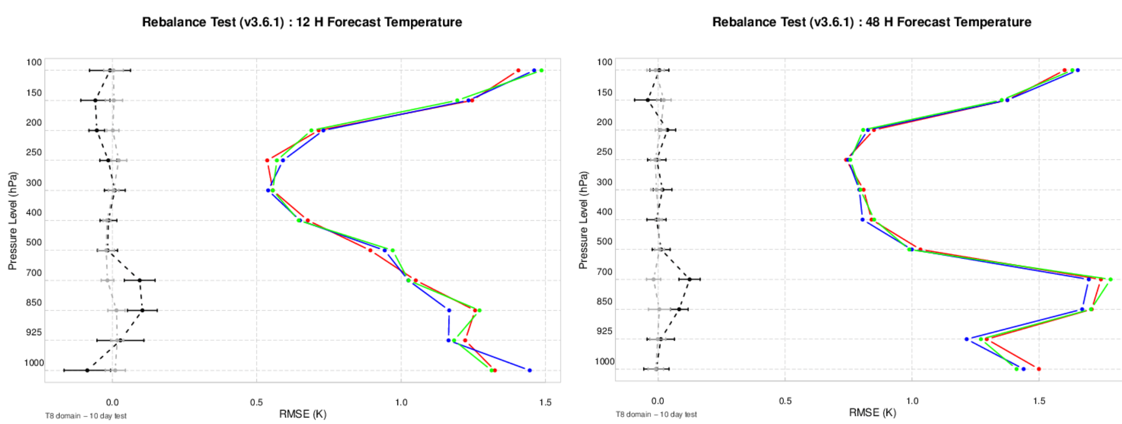


Figure 3.3.4-2: Rebalance test forecasts for temperature at 12-hour (left) and 48-hour (right) forecasts. The GSI with rebalance (red) is compared to the GSI without rebalance (blue) with the pairwise difference of these two runs in black. The no-data assimilation run (green) and GSI with rebalance pairwise difference is in grey. Differences are not SS when the confidence interval (CI) encompasses zero.

The Air Force submitted a mitigation request pertaining to a significant decrease in the CrIS data that is assimilated when using GSI v3.2 (with Forecast Sensitivity to Observations; FSO) rather than GSI v3.1 (Air Force operational configuration at time of request). DTC requested a case and was able to reproduce the decrease in number of observations assimilated, shown in Table 3.3.4-2.

Table 3.3.4-2: DTC comparison to Air Force case for CrIS mitigation request

			# Read	# Kept	# Assimilated
GSI v3.1	**T4 DTC repeat**				
	npp	cris	1695750	468825	12797
	T4 Air Force				
	npp	cris	551019	156009	15954
GSI v3.2	**FSO DTC repeat**				
	npp	cris	1231713	318402	1074
	FSO Air Force				
	npp	cris	1231713	318402	1074

The problem was identified to be related to the use of the GSI namelist option 'dval'. In GSI v3.1, dval was set to one, whereas in GSI v3.2, dval was set to zero for CrIS (other types dval=1 for v3.2). When dval is set to zero, most of the CrIS data is removed by the radiance thinning process. This option allows for relative weighting of different satellite radiance instruments in a thinning box. Discussions with EMC led to the suggestion of setting dval to zero for all radiance data types so no specific types are unequally weighted during the thinning process (therefore increasing the CrIS usage). EMC plans to remove the 'dval' option in the near future; removing this option will eliminate future confusion.

3.4 Ensembles

3.4.1 Pre-NARRE T&E

An immediate need to work on North American Rapid Refresh Ensemble (NARRE) is the result of an expansion of computing resources for operations at NCEP through support of the Sandy Supplemental Program. NARRE became a part of a plan on how to best utilize these resources to address important forecasting questions. One of the items on the EMC roadmap was an option to create an extension of the existing Short Range Ensemble Forecasting (SREF) system by adding the rapid refresh component. The idea is to have SREF continuing to run on 6-hourly cycles out to 84 h and NARRE will be a subset of 6 to 8 SREF members updated hourly and running out to 18-24 h. Having these members as a subset of SREF means that model uncertainty, at least at the beginning, would be addressed by the use of two dynamic cores ARW (RAP) and NMMB (NAM) and variations in physics. According to the EMC roadmap, this new system, consisting of 6 members, is expected to be implemented in operations during 2017 for improved aviation and probabilistic forecasts for other short-range applications. The work involves a very close collaboration between GSD, EMC and DTC staff.

Initial work included development of preliminary configurations that were tested in retrospective experiment mode. More detailed information about the experiment design and results can be found in the report available on the DTC website (http://www.dtcenter.org/eval/ensembles/EN6_2014_report.pdf). An example of physics variations that were examined as a part of this configuration testing is presented in Table 3.4.1-1. A sample of results and ensemble spread/error ratio is illustrated in Fig 3.4.1-1. Initialization will ultimately be hourly ("rapid"), but for the initial testing, less frequent updates were employed. The forecast length used in the testing was 24 h. The final configuration will depend on computing resources dedicated to this task and discussion with EMC colleagues on pre-NARRE design. NARRE will be configured with 13-km horizontal grid spacing and 60 vertical levels. For the initial testing, the DTC used the RAP operational domain, which is somewhat smaller than the domain on which the operational system will run. DTC staff participated in the Weather Prediction Center (WPC) Hydrometeorological Testbed

Winter Weather Experiment representing work done under the preliminary NARRE configuration. Following the experiment, DTC staff continued to work with WPC colleagues on exchanging experiment data and communicating some objective analysis results.

Table 3.4.1-1. List of members tested. The green color indicates NAM members and blue indicates RAP members selected based on the results of the experiment.

	MP	Sfclay	Sfcphy	PBL	CU	IC/LBs
rap ctl	Thompson	MYNN	RUC	MYNN	GF	GFS
rap1	Thompson	MO-MYJ	RUC	MYJ	BMJ	GEP01
rap2	Ferrier	MO-YSU	RUC	YSU	BMJ	GEP02
rap3	Ferrier	MO-MYJ	RUC	MYJ	BMJ	GEP03
rap4	Ferrier	MO-MYJ	NOAH	MYJ	BMJ	GEP01
rap5	Ferrier	MYNN	RUC	MYNN	GF	GEP02
rap6	Ferrier	MYNN	RUC	MYNN	BMJ	GEP03
rap7	Thompson	MO-YSU	RUC	YSU	BMJ	GEP04
rap8	Ferrier	MO-YSU	RUC	YSU	GF	GEP01
nmmb ctl	Ferrier	MYJ	NOAH	MYJ	BMJ	GFS
nmmb1	Ferrier	MYJ	NOAH	MYJ	BMJ	GEP01
nmmb2	Ferrier	MYJ	NOAH	MYJ	BMJ	GEP02
nmmb3	Ferrier	MYJ	NOAH	MYJ	BMJ	GEP03

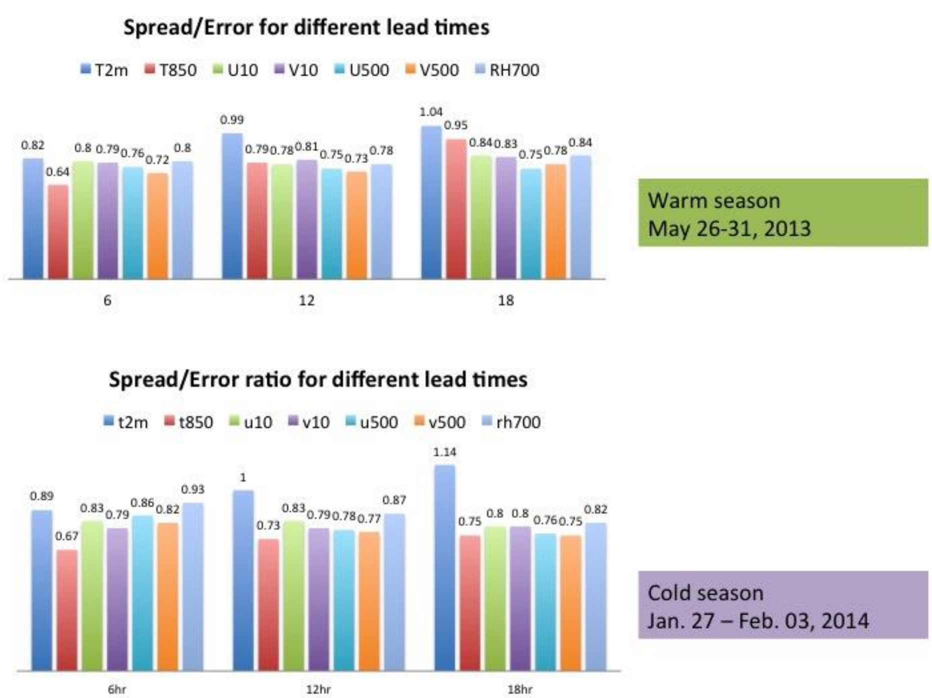


Figure 3.4.1-1. Spread/Error ratio for 2-m temperature, 850 mb temperature, 10-m wind, 500 mb wind and 700 mb relative humidity for different lead times and two periods of interest.

An additional part of this T&E activity was to assess resources needed to run NARRE on the NAM operational domain. The task was accomplished. Interestingly, a decision was made that the expanded domain will become the operational RAP domain in the next implementation.

3.4.2 Neural Network

Many physics schemes, especially microphysics, have become very advanced and often need significant computing resources when run at the frequent time steps required by the schemes. In addition, ensemble modeling has shown that the use of multiple physics parameterizations offers an improvement over the use of a single scheme, suggesting that no one scheme captures the physics necessary to represent certain phenomena. Given these two concerns, a neural network (NN) approach to physics may provide a comprehensive method to capturing the uncertainty within complex physics processes while at the same time reducing the computing resources necessary to run the scheme. The goal of this T&E activity is to generate and test a NN microphysics scheme by using Thompson microphysics output generated in the NMMB model for a series of cases representing all four seasons. Once a NN scheme has been created, the Thompson scheme will be replaced by the NN scheme and these cases will be rerun for verification purposes. This proof-of-concept research will show that NN may be applicable for other physics schemes in order to capture necessary uncertainty and to improve the speed with which operational models can be run. Progress on this project is delayed, and the DTC expects to have results by the end of summer 2015.

4 References

Kaplan, J. and M. DeMaria, 2003: Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. *Wea Forecasting*, **18**, 1093- 1108.

Marchok, T., R. Rogers, and R. Tuleya, 2007: Validation schemes for tropical cyclone quantitative precipitation forecasts: Evaluation of operational models for U.S. landfalling cases. *Wea. Forecasting*, **22**, 726–746.

Xu, K.-M., and D. A. Randall, 1996: A semiempirical cloudiness parameterization for use in climate models. J. Atmos. Sci., **53**, 3084-3102.

Sundqvist, H., E. Berge, and J. E. Kristjánsson, 1989: Condensation and Cloud Parameterization Studies with a Mesoscale Numerical Weather Prediction Model. *Mon. Wea. Rev.*, **117**, 1641–1657.

5 Acronyms and Abbreviations

ABSerr	Absolute Error
AF	Air Force
AFWA	Air Force Weather Agency
AL	Atlantic
AMSR2	Advanced Microwave Scanning Radiometer 2
AOML	Atlantic Oceanographic and Meteorological Laboratory
AOP	Annual Operating Plan
AR	Atmospheric River
ARW	Advanced Research WRF
BAMS	Bulletin of the American Meteorological Society
BS	Brier Score
BSS	Brier Skill Score
CIRES	Cooperative Institute for Research in Environmental Sciences
CMORPH	CPC MORPHing technique

CONUS	Contiguous United States
CrIS	Cross-Track Infrared Sounder
CRPS	Continuous Ranked Probability Score
CRPSS	Continuous Ranked Probability Skill Score
CWB	Central Weather Bureau
DA	Data Assimilation
DRC	Data Assimilation Review Committee
DTC	Developmental Testbed Center
EC	Executive Committee
ECMWF	European Center for Medium-Range Weather Forecasts
EMC	Environmental Modeling Center
EnKF	Ensemble Kalman Filter
EP	Eastern North Pacific
ESRL	Earth System Research Laboratory
ESTDEV	Error Standard Deviation
ETS	Equitable Threat Score
FAR	False Alarm Rate
FUR	Friendly User Release
FSOE	Functionally Similar Operational Environment
GCOM-W1	Global Change Observation Mission-W1
GFDL	Geophysical Fluid Dynamics Laboratory
GFS	Global Forecasting System
GMAO	Global Modeling and Assimilation Office
GOME-2	Global Ozone Monitoring Experiment-2
GSD	Global Systems Division
GSI	Gridpoint Statistical Interpolation
GSS	Gilbert Skill Score
H214	2014 HWRF Retrospective Test
HEDAS	HWRF Ensemble Data Assimilation System
HFIP	Hurricane Forecast Improvement Project
HMT	Hydrometeorology Testbed
HRD	Hurricane Research Division
HWRF	Hurricane WRF
LW	Longwave
MB	Management Board
MCC	Mesoscale Convective Complexes
MAE	Mean Absolute Error
ME	Mean Error
MET	Model Evaluation Tools
MET-TC	Model Evaluation Tools – Tropical Cyclone
MODE	Method for Object-based Diagnostic Evaluation
MMET	Mesoscale Model Evaluation Testbed
MMM	Mesoscale and Microscale Meteorology (Division at NCAR)
MPIPOM-TC	Message Passing Interface Princeton Ocean Model for Tropical Cyclones
NAM	North American Mesoscale
NAM-RR	NAM Rapid Refresh
NARRE	North American Rapid Refresh Ensemble
NASA	National Aeronautics and Space Administration

NCAR	National Center for Atmospheric Research
NCEP	National Centers for Environmental Prediction
NCO	NCEP Central Operations
NCODA	Navy's Coupled Data Assimilation
NCWCP	NOAA Center for Weather and Climate Predication
NEMS	NOAA Environmental Modeling System
NetCDF	Network Common Data Form
NGGPS	Next Generation Global Prediction System
NHC	National Hurricane Center
NITE	NWP Information Technology Environment
NMMB	Nonhydrostatic Multiscale Model on the B grid
NN	Neural Network
NOAA	National Oceanic and Atmospheric Administration
NPP	NPOESS Preparatory Project
NPS	NMMB Preprocessing System
NRL	Naval Research Laboratory
NSF	National Science Foundation
NSSL	National Severe Storms Laboratory
NWP	Numerical Weather Prediction
NWS	National Weather Service
OAR	Office of Oceanic and Atmospheric Research
PAC	Pattern Anomaly Correlation
POD	Probability of Detection
PBL	Planetary Boundary Layer
PSD	Physical Sciences Division
PSU	Pennsylvania State University
QPE	Quantitative Precipitation Estimate
QPF	Quantitative Precipitation Forecast
RAP	Rapid Refresh
R2O	Research to Operations
R&D	Research and Development
RAMADDA	Repository for Archiving, Managing and Accessing Diverse DAta
RC	Reference Configuration
RI	Rapid Intensification
RMSE	Root Mean Square Error
ROC	Receiver Operator Characteristic Curve
RRTM	Rapid Radiative Transfer Model
RRTMG	Rapid Radiative Transfer Model for Global Climate Models
RW	Rapid Weakening
SAB	Science Advisory Board
SAS	Simplified Arakawa-Schubert
SBUV/2	Solar Backscatter UV/2
SREF	Short-Range Ensemble Forecast
SS	Statistical significance
SW	Shortwave
T&E	Testing and Evaluation
TC	Tropical Cyclone
TLNMC	Tangent-Linear Normal-Model Constraint

TTFRI	Taiwan Typhoon and Flood Research Institute
UCLA	University of California – Los Angeles
UKMET	United Kingdom MET Office
UPP	Unified Post-Processor
URI	University of Rhode Island
USWRP	US Weather Research Program
VSDB	Verification Statistic DataBase
WP	Western North Pacific
WPC	Weather Prediction Center
WRF	Weather Research and Forecasting