Team Co-Leads: Ivanka Stajner (NWS/OSTI) and Glenn White (NWS/EMC)

Verification and Validation of the NGGPS Modeling System

1.1 Objective

Verification and validation activities are critical for success of NGGPS as they provide tools and metrics for evaluation of model performance, thus providing the basis for evidence-based approach to decision making, modeling system development, and model improvement.

Objectives of the Verification and Validation team are to:

- Develop a comprehensive and flexible verification package for evaluation of progress in the development and operational readiness of NGGPS and of future NGGPS operational performance
- Enable stakeholder validation of NGGPS performance

The verification package is envisioned to provide quantitative measures to support evidence–based approach for decision making and NGGPS development. The primary users of the verification system and its products will be:

- NGGPS developers and users of NGGPS products
- NGGPS program office, NCEP operational centers, NOAA laboratories, and NOAA managers
- Research community, private sector, universities

Standard metrics, sensible weather metrics, tropical cyclone verification, and ensemble verification will be included as part of the verification package. Currently proposed verification metrics are included in Appendix A.

The validation is envisioned to gather input from interested stakeholders about their assessment if NGGPS is meeting their needs. These may be stakeholder specific needs such as the accuracy of predicting a variable for certain thresholds or the ability to predict specific phenomena.

1.2 State of the Science

There are three separate verification packages in current use that have been developed by NCEP, HIWPP and the DTC. A description of each is below, with a consolidation of current Verification Capabilities, still in draft form, listed in Appendix C. Also see http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/ and http://www.emc.ncep.noaa.gov/gmb/wx24fy/vsdb/gfs2016 for standard metrics used in NCEP-EMC Global Model Deterministic Forecast Verifications.

- **NCEP**
  - Global verification focuses on large-scale flow pattern over the globe and long-term statistics of model performance
  - Increased emphasis on mesoscale verification focusing on synoptic events and sensible weather elements
- **HIWPP**
  - Unifies metrics currently produced by NCEP/EMC and ESRL's Global System Division (GSD)/Earth Modeling Branch (EMB) for global models
- **DTC Model Evaluation Tools (MET)**
  - Community evaluation tools based on NCEP's grid2obs, grid2grid and FVS verification packages through a platform-independent and extensible software package.

1.3 State of Verification and Validation at NCEP

The verification of global, regional and ensemble forecasts at NCEP/EMC is carried out independently by different branches and groups. The Global Modeling Branch (GMB) focuses mostly on the verification of large-scale flow pattern over the globe and on long-term statistics of model performance. The Mesoscale Modeling Branch (MMB) focuses more on the verification of synoptic events and weather sensible elements. In recent years, EMC has made an effort to unify the verifications among different modeling systems. Since 2007, the global branch has been using the VSDB (Verification Statistics Data Base) approach for verifying global model forecasts, while continuing to run some of its legacy verification systems for consistency of historical records.

The Mesoscale Modeling Branch plots outputs of these files using either the Forecast Verification system (FVS), which takes the records and then calculates a wide variety of statistics from the partial sums, and then uses a Gempak-based code to plot them, or a web-based system created by Andy Loughe, formerly of the Development Testbed Center (DTC), that uses a Perl-based system to read a MySQL database containing data from the VSDB information.

Verification at NCEP/EMC, using VSDB output, has been created using one of two codes: grid2obs, which interpolates model data to observation location, and grid2grid, which compares model data against gridded analyses on the same grid as the model. EMC maintains a long history of these VSDB files for many models, and against several variables for many uses at EMC. EMC, with the help of the Verification Group at the DTC, is moving towards using the DTC-developed METViewer package to display the content of the VSDB files.

The grid2obs verification system is used to verify standard upper-air and surface variables, plus other variables such as cloud amount, cloud base height and ceiling, horizontal visibility, and PBL height. This verification system is used for the North American Mesoscale (NAM) forecasting system and its nests, Rapid Refresh (RAP), High Resolution Rapid Refresh (HRRR), the High-Resolution Window (HRW), and all the members of the Short-Range Ensemble Forecast (SREF) and its mean. In addition, grid2obs verification system is used to verify output from the Community Multi-scale Air Quality (CMAQ) model to verify ozone and 2.5-micron particulate matter predictions.

The grid2grid verification system is used to verify variables like cloud amount against both the Air Force World Wide Merged Cloud Analysis (WWMCA) and the CLAVR (cloud from AVHRR (Advanced Very High Resolution Radiometer)) cloud datasets, and also echo-top and reflectivity against analyses of these items. It is also used to verify ensemble systems SREF, Global Ensemble Forecast System (GEFS), Canadian ensemble, North America Ensemble Forecasting System (NAEFS), European Center for Medium range Weather Forecasting (ECMWF) and Navy ensemble, and their bias correction forecasts; and also used to verify NAM, RAP, HRRR, and the HRW against the Un-Restricted Mesoscale Analysis (URMA) dataset. Finally, the grid2grid system is used for verification of smoke and dust predictions from the Hybrid Single Particle Lagrangian Integrated Trajectory model.

A separate grid2grid application is used to verify precipitation for various forecast ranges and hours of accumulation of precipitation against the Climatology Calibrated Precipitation Analysis (CCPA). In addition, object-based verification of precipitation is done by calculating the Fractions Skill Score.

The GMB runs an instance of the grid2obs to create VSDB files needed for regions of the globe. Other statistics used by the GMB, such as anomaly correlation coefficients, are displayed on a web-page that shows pre-generated graphics.

The goals of the EMC global model verification group are to:

- Monitor GFS historical performance and compare GFS with other international NWP models (e.g., http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/)
- Evaluate the performance of GFS parallel experiments and provide management decision guidance for GFS upgrade and,

- Provide model developers quick and easy access to model evaluation tools and verification results.

A prototype Global NWP Model Verification Package was developed at NCEP EMC in the past few years to accomplish these goals. This package is now used by all GFS developers at NCEP. It is also used by other model developers at ESRL, NESDIS, KIAPS, CMA, and Indian Meteorological Department etc. For instance, it was used for verifying T1534 GFS parallel experiments (http://www.emc.ncep.noaa.gov/gmb/wx24fy/vsdb/gfs2015/) before the model was implemented into operation in January 2015. It uses the VSDB-based tools grid2grid and grid2obs developed at the EMC MMB for computing verification statistics of forecasts verified against analyses and observations, respectively. In addition, it also contains verification of hurricane track and intensity, precipitation ETS and Bias scores verified against CPC gauge observations, 2-D forecast and verification maps, analysis increments, ENKF ensemble spread, and a Fit-to-Obs tool that is independently developed by Jack Woollen and Suranjana Saha and uses GFS/GDAS output with model native grids as input. This Global NWP package uses GrADS as its plotting tool to make pre-generated graphics, and uploads all graphics along with html and java templates to web servers for displaying and viewing of results. It also generates a "scorecard" that summarizes the improvements and degradations between two forecast experiments based on selected verification metrics. In the past few months EMC has also started to use the Method for Object-based Diagnostic Evaluation (MODE) within the MET package developed by DTC to carry out object-oriented verification of global model forecasts of precipitation and jet streams (http://www.emc.ncep.noaa.gov/gc_wmb/tdorian/).

While the current NCEP-EMC Global NWP Model Verification Package is quite useful and includes many skill metrics, there is a growing list of additional desired capabilities, especially as the resolution of the global model increases and the demand by user community grows for more detailed evaluation prior new model implementation.

It is expected that new gaps in the verification package will be identified and added to the list of desired capabilities for the NGGPS verification plan. Further, it is recognized that the longer-term plan will focus on synergies with other verification software development efforts of EMC, ESRL, and DTC.

1.4 State of HIWPP Verification and Validation

The verification deliverable from the High Impact Weather Prediction Project (HIWPP) is a framework that unifies metrics currently produced by NCEP/EMC for global models and metrics produced by ESRL's Global Systems Division (GSD)/Earth Modeling Branch (EMB) for global models. This framework, publically available at [http://hiwpp.noaa.gov/verify/,](http://hiwpp.noaa.gov/verify/) is in use to verify output from the hydrostatic global models participating in HIWPP. The EMC portion displays static images produced in VSDB format. The EMB portion is an interactive interface that allows the user to dynamically select the plot to be displayed, using a backend MySQL database which stores running sums (partial sums) from model output along with observations. To date, the VSDB images are available for HIWPP models. Parts of the EMB system are implemented; additional products are also being included. The full suite of EMB products is listed in Appendix D.

In the process of implementing the web application to interactively view EMB metrics, a serious security issue was encountered around Java Applets, which are used in the application. The application was redeveloped without Java Applets, and a new modular and easily extensible web framework was developed for use in HIWPP, based on html5 technology. This has been named the Model Assessment Tool Suite (MATS).

Tropical Cyclone verification is supported through software developed by Mike Fiorino at ESRL/GSD and displayed on the Hurricane Forecast Improvement Project (HFIP) and Weather and Climate Data (WXMAP) websites. It is anticipated that these web pages will be integrated into the HIWPP site.

Beyond the unified framework of existing metrics, the HIWPP verification task is also working to address three areas for advancement of verification metrics: 1) ensemble verification, 2) a multi-parameter scorecard, and 3) enhanced sensible weather verification, in particular relating to global precipitation verification. Ensemble verification is being addressed using NCEP's grid2grid ensemble verification tool with the capabilities developed by Yuejian Zhu at NCEP and by the MET team within DTC. Specifications for a scorecard are in development. To support global precipitation verification, the HIWPP team has developed processing to ingest and apply quality control to global SYNOP reports.

1.5 State of DTC MET Verification and Validation

The Model Evaluation Tools (MET) software package was developed to address the general need for model evaluation and to provide the scientific community with a comprehensive set of forecast evaluation tools for diagnostic evaluation of NWP and climate prediction systems. MET and the accompanying database and display system, METViewer, were developed with direction from the Developmental Testbed Center (DTC) with funding from National Oceanic and Atmospheric Administration (NOAA), the Air Force (AF) and the National Center for Atmospheric Research (NCAR), and is supported to the community by the Developmental Testbed Center (DTC). MET and METViewer provide capabilities for model evaluations to the broad NWP community, including university researchers, model developers and scientists at government agencies, as well as members of the commercial sector. A more detailed description along with access to the MET software package can be obtained from the DTC website: http://www.dtcenter.org/met/users/. MET has over thousands of registered users, both nationally and internationally, and is also at the foundation of many of the DTC testing and evaluation (T&E) systems.

MET was originally developed based on the NCEP Mesoscale Modeling Branch (MMB) grid2obs, grid2grid and FVS verification packages with the goal of supporting these capabilities to the community through a platform independent and extensible software package. The MET package has been designed to be modular and adaptable. For example, individual modules can be applied without running the entire set of tools. New tools can easily be added to the MET package due to this modular design. MET computes over 50 traditional continuous, categorical, and probability statistics such as bias, root-mean squared error (RMSE), and mean absolute error (MAE), Probability of Detection (POD), Probability of False Detection (POFD), False Alarm Ratio (FAR), and Critical Success Index (CSI), Brier Score (BS), rank histograms and reliability diagrams. MET also includes the Ensemble-Stat tool which is used to formulate simple ensemble products and calculate various ensemble performance attributes (e.g., rank histograms). Additionally, MET includes tools for object-based verification using Method for Object-based Evaluation (MODE), error scale decomposition using Wavelet-Stat, neighborhood methods using Grid-Stat, and evaluation of tropical cyclone intensity and track forecasts through the MET-TC package. MET-TC was adapted from the National Hurricane Center (NHC) verification package and added to MET with the same goal of supporting it out to the community.

For statistics aggregation, MET has several analysis tools (Stat-Analysis, MODE-Analysis, TC-Stat, and Series Analysis) to read output from MET, stratify and aggregate the results. Additionally, the METViewer database and display system is a useful supplement to the MET system. METViewer has two core interfaces: (1) a batch engine for systematic generation of user defined plots; and (2) a web-based user interface to allow researchers to further stratify and analyze the data. Both components will be leveraged to provide static plots and in-depth analysis and stratification of evaluation statistics. METViewer was recently updated to read VSDB files from grid2obs, grid2grid and many of the fields from grid2grid_e (for ensembles). The user interface for METViewer provides a great deal of flexibility for the user to stratify the data and produce publication quality plots. METViewer is now being used by some staff within EMC and GSD as well as for T&E activities in the DTC.

The DTC Verification Team within the DTC has been working closely with DTC teams as well as the operational community to enhance MET to better support both internal T&E activities and testing performed at EMC and within testbeds and centers (e.g. EMC, WPC, Hydrometeorology Testbed, Hazardous Weather Testbed). These interactions have led to the improved capability to streamline verification and provide support to components of the modeling systems beyond the core model. For example, MET can now automatically regrid one or both grids passed to it prior to computing the matched pairs. This eliminates the need to use regridding tools, such as copygb in the NCEP Unified Post-Processing package (UPP) and to store additional files. It has enhanced support for reading GSI diagnostic files and uses the increments to compute matched pairs. Support for verification of aerosol components has been made available and the use of climatologies to compute scores such as anomaly correlation and skill-scores had been expanded recently. Finally, MET developers try to incorporate cutting-edge non-traditional tools to the package whenever one is identified as mature enough to be in a community release.

2.0 Verification and Validation Strategy

Verification and validation strategy for NGGPS will rely on the three-prong approach listed here:

1) Develop a flexible and comprehensive verification package through unification of capabilities from NCEP, ESRL and DTC verification packages.
2) Gather input from other NGGPS teams to develop a package that meets their needs and has a stable portion that would be routinely run by NCEP to produce standard statistics over time and allow backward compatibility for historical statistics.
3) Coordinate NGGPS validation with UMAC's (University Corporation for Atmospheric Research Community Advisory Committee for NCEP (UCACN) Model Advisory Committee (UMAC)) evaluation of NCEP's production suite, Model Evaluation Group activities, and collection of stakeholder input to define the approach to NGGPS validation.

The following have been identified as potential areas for improvement needed to develop comprehensive verification tools and validation approaches (additional details on some technical aspects are provided in Appendix A):

1) Treatment of uncertainty in the analyses and observational data set that forecasts are verified against
2) Scorecard and weighted performance indices (global scale and sensible weather)
3) Ensemble and probability verification metrics
4) Unification of packages capturing strong elements from all systems
5) Diagnostic tools (scales at which errors occur, energy spectra), physics-oriented metrics (radiation, fluxes, cloud verification)
6) Measure of forecast consistency
7) Metrics for extreme weather events (extend hurricane track to day 7)
8) Object-oriented metrics
9) Component performance
10) System performance (including fluxes and interface variables among components)
11) Validation - enable early and comprehensive user involvement

A stronger linkage is envisioned between verification and data assimilation systems for NGGPS. The operational data assimilation system will be upgraded in the next few years with observation database (ODB) software that makes access to the observations and of the model forecast fit to the observations much more straightforward. In addition to verifying forecasts against analyses, direct forecast comparisons against all observation types, in-situ and remote, will be available to model developers. The NGGPS verification team thus expects to work with the data assimilation team to develop the ODB-based diagnostics that can be used for a variety of purposes. In particular, the diagnostics software may be useful for model development. It will be possible to more accurately determine where biases are occurring in the forecast, and by associating this with, say, what parameterizations were active in that region at that time, determine the aspect of the forecast model that should be improved to ameliorate this bias.

Forecast verification can also guide the model-development process. In order to enable that application, additional data may need to be provided and saved from a model run. For example, commonly the forecast models do not have the ability to archive instantaneous model tendencies associated with the dry dynamics and each parameterization. The capability to save such data when requested can be very helpful for model development. With such tendencies, it is possible to determine what parameterizations were active in a region with substantial bias, and what the sign and magnitude of the tendencies were, so as to determine potential sources of the forecast bias. Model design and storage needs are listed in more detail in Appendix A.

3.0 Milestones and Outcomes for Near-Term Objectives (FY2016)

- **Begin unification of the verification approach starting from MET and METViewer**
  - Define initial metrics to be used
  - Identify location of data sources (forecast and obs)
  - Identify location to run system
  - Identify additional metrics for inclusion in comprehensive verification system
- **Set up the initial MET and METViewer system (hereafter, MET+) at EMC/NCO**
  - Set up initial system and evaluate outputs
  - Set-up initial capability for scorecarding and visualization of statistics
- **Identify database schema that is consistent and suitable for use at NCO**
- **Document requirements and develop a procedure for inclusion of new verification metrics**
  - Identify and begin including metrics for ocean, ice, land, aerosols and system coupling
  - Identify and begin including process oriented metrics
- User support, training, documentation for MET+
- Coordination and planning
  - Establish a Focus Group of stakeholders
  - Gather stakeholder needs

4.0 Milestones and Outcomes for Long-Term Objectives (FY 2017-FY 2019)

- **Comprehensive verification system for operational and developer use**
  - Additional metrics (e.g. ensemble, cyclone, scorecard, high-resolution, object oriented, forecast consistency, process-oriented, global index, sensible weather index)
  - Evaluation tools (e.g. 2D maps, timeseries)
  - Component performance (e.g. ocean, ice, land, aerosol)
  - System performance (system coupling metrics)
- **Database development and optimization**
- **User support**
- **Validation and implementation decision support**
- Data repository for verification data sets and quality control
- Inclusion of additional verification datasets
- Visualization and user interface improvements
- Treatment of uncertainty in the analyses and observational data set that forecasts are verified against (topic for external FFO)

Appendix A

Goals and Considerations for a Verification and Validation System

A. General requirements for verification software

1. All verification system components should be accessible to all identified users, with technical support. The users should have the ability to modify or extend the core capabilities quickly and easily, therefore a common core language such as C++, Fortran, or python, would be beneficial.
2. The code should be portable to various computational platforms and must be acceptable to NCO.
3. Protocols will be in place so that any errors can be fixed quickly, without waiting for a system release; documentation will be readily available.
4. Regression testing will be performed after any code changes using a product similar to, or the set of, standardized test cases that MET has developed.
5. As community code, the community will need to be defined and the code managed in a shared repository accessible to non-NOAA collaborating institutions (e.g. NCAR, universities, NRL).
   The verification system must be able to evolve and have the ability to add new verification metrics as they are determined.

B. Generating metrics

1. Metrics are needed for atmospheric, ocean, sea ice, wave, land surface, aerosol and chemistry components, and for system coupling.
2. As metrics are developed and defined, there will be a need to support both retrospective and real-time models, with ability to be integrated into Rocoto or other (e.g. ecFlow) workflow management system (WMS).
3. The metrics generated will provide distributions of observations, forecasts, and errors, not just averages, and from there several statistics should be available, such as mean, median, quartiles, confidence intervals etc.
4. Conditional verification, for example, what is the temperature error for sites with precipitation observations > 1 in in 24h, should also be available.

C. Specific features to be addressed

1. Additional evaluation tools to be considered
   - 2D maps of verification vs. grid, overplotted with verification vs. observations when available
   - Timeseries
   - Diurnal cycle figures
2. Proposed verification scorecard
   - Expand to include verification against observations, precipitation
   - Consider display of subsets of the score card, e. g. for upper levels only
3. Proposed indices
   - Global index - based on standard global metrics
   - Sensible weather index - based on sensible weather metrics
   - Provide formulas with metrics and weights (Potentially start with Met office approach and modify as needed with stakeholder inputs?)
4. Proposed software solution should be evaluated for:
   - security
   - speed
   - flexibility
   - maintainability

D. Web interface

1. The web interface will need to conform to current security requirements
2. The interface should have a “quick-start” for scientists to be able to use without extensive training, as well as a more advanced interface with more capabilities.
3. It should include “on-the-fly stratification selection”
4. A “quick-look” summary scorecard (dashboard) with an agreed upon set of summary statistics (eg: UA, sfc, precip, etc.) for several models would be beneficial.
5. Advanced web-interaction could include options such as clicking on a plot or section of a page in order to trigger a pop up the “on-the-fly” page for that verification area, replacing the long list of verification choices on the EMB page.
6. Clear naming convention for models, with separation of real-time and retro runs would be established; ideally, with two entries for each NCEP operational model in a category - 1) the NCEP operational version and 2) the primary real-time parallel.
7. Integrating static images with dynamic content, such as thumbnails that user could click to “go deeper” and access dynamic content would provide an informative user interface.

E. Plotting

1. A method of generating verification statistics and plots interactively, as well as pre-generating a configurable set of plots would be needed.
2. Modelers/users should be able to make necessary changes to specific final plotting scripts to produce plots that meet publication requirements, and allow adjustments to labels.
3. Users should be able to save plots as gif/jpeg files.

F. Data and data storage

1. The amount of data generated, as well as the previous data to be archived, will need to be available and stored, so data management and storage needs will be considered.
2. Database technology that can handle large amounts of data should be identified, and it can evolve to meet additional needs and models as they become available, will be evaluated.
3. Consideration will be given to whether running/partial sums could be stored in a peripheral database in an agreed convention.
4. The ability to save historical model output, as well as gridded data (forecast and obs) will also be evaluated to determine if it is optimal and/or feasible.
5. Observations requirements, including access and maintenance of observational data, will be addressed, since gaining access to NOAA computers/data is often a challenge to users outside of NOAA, as well as determining whether a separate database of observations would be warranted and whether the formats and any obs conversions would be supported.
6. There would be a need to determine whether there would be verification against non-standard observations, such as satellite radiances and radar reflectivity, and what organization would be responsible to QC the data

G. Data and data storage for process-oriented metrics

Additional model output and storage will be needed for process-oriented metrics. It is recommended that as NGGPS is developed:

1. Software engineers should actively consult with model developers about the extra diagnostic information, such as tendencies, they may wish to have saved as a forecast model is developed. These data need not be saved routinely, but might be triggered with a namelist option.
2. Software developers should engineer the NGGPS system to be able to save these data if requested.

3. Parameterization developers develop a set of process verification best practices and use them. What diagnostics should be routinely generated and examined to illuminate deficiencies in a particular parameterization?
4. Disk storage should be increased sufficiently to permit the storage of extra diagnostic information as discussed above.

Appendix B

Table of Verification Metrics for Consideration

Variable (temperature, precipitation, geopotential height, …)	**Dataset used in verification** (radiosonde observations, NCEP analysis at ? degree resolution,...)	**Metric** (mean, RMS, AC, …)	**Regions** (global, NH, SH, North America,...)	**Levels** (surface, 500 hPa, ..)	**Forecast times** (every 6h for 0 to 10 days, …)
Grid-to-Grid Verification Based on VSDB Partial Sums					
Z, T, SLP, U, V, and vector wind	GFS analysis, 2.5-deg resolution	AC	Global, NH, SH, Tropics, and PNA	1000, 700, 500, 250 hPa (except for SLP)	every 6-h up to 10 days, and then every 12-h up to 16 days
Z, T, O3, U, V, and wind, Total cloud	GFS analysis, 2.5-deg resolution 1 deg AFWA and CLAVR satellite analysis data	Bias, RMSE, RMSE by Mean Difference, RMSE by Pattern Variation, Murphy's MSE Skill Score, Ratio of standard deviations between forecasts and analysis, pattern correlation	Global, NH, SH, Tropics, and PNA	1000, 850, 700, 500, 200, 100, 50, 20, 10 hPa	every 6-h up to 10 days, and then every 12-h up to 16 days
Precipitation	CCPA	Fractions skill score; contingency table (FHO)-based scores (ETS, bias, FAR, POD, EDI *etc.*), SL1L2 stats	ConUS (some OConUS FHO/SL1L2 using other, less reliable analysis)	Surface	FSS: daily and 6-hourly up to 84h. FHO/SL1L2: 3-hourly up to 84h; daily up to 8 days.
Grid-to-Grid Ensemble Verification Based on VSDB Partial Sums					
T, Z, U, V	2.5 and 1 degree GFS analysis NDAS, 2.5 and 1-degree climatology data	Reliability-diagram, RMSE/spread, ROC, histogram, BS,BSS, resolution, reliability, RPS, RPSS, CRPS, CRPSS, econo- value, etc	Global, NH, SH, CONUS, NA, Tropical, Europe, Asia, etc	Surface, 1000mb, 850mb, 700mb, 500mb	Every 6hr forecast up-to 14 days for global ensemble, every 3hr up-to 87 hr for SREF

Variable (temperature, precipitation, geopotential height, …)	**Dataset used in verification** (radiosonde observations, NCEP analysis at ? degree resolution,...)	**Metric** (mean, RMS, AC, …)	**Regions** (global, NH, SH, North America,...)	**Levels** (surface, 500 hPa, ..)	**Forecast times** (every 6h for 0 to 10 days, …)
Grid-to-Obs Verification Based on VSDB Partial Sums					
T-2m, RH-2m, Td-2m, Wind-10m, Total Cloud, and SLP	Surface observations saved in prebufr files	Bias, and RMSE	Alaska, CONUS and its sub-regions	surface	every 3-h up to 7 days.
T, Q, RH, Wind	ADPUPA, AIRCAR, AIRCFT etc saved in prepbufr files (resolution of forecast data is one degree)	Bias and RMSE	Global, NH, SH, Tropics, and CONUS	1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, and 50 hPa	every 6-h up to 7 days
Fit-to-Obs Verification Using GSI Prefits on Model Native Grid					
Z, T, Q, Wind, and Ps	ADPUPA, ADPSFC, AIRCAR, AIRCFT and SFCSHP etc saved in GSI prepfits files	Bias and RMSE	Global, NH, SH, Tropics, North America, Europe, and Asia	uses GFS output on model native vertical grid, and surface output at model Gaussian grid.	every 12-h up to 5 days
precipitation	CPC-Gauge observations (resolution of forecast data is about 13km)	ETS and BIAS Scores	CONUS	surface	24-h accumulations up to 7 days
precipitation	CCPA analysis (forecast and obs are both on 0.25 deg)	object-oriented MODE Verification	CONUS	surface	24-h accumulations up to 7 days

Variable (temperature, precipitation, geopotential height, …)	**Dataset used in verification** (radiosonde observations, NCEP analysis at ? degree resolution,...)	**Metric** (mean, RMS, AC, …)	**Regions** (global, NH, SH, North America,...)	**Levels** (surface, 500 hPa, ..)	**Forecast times** (every 6h for 0 to 10 days, …)
Jet Streams	model analyses at 1-degree resolution	object-oriented MODE Verification	NH	250 hPa	every 6-h up to 7 days
Hurricanes	NHC best tracks (resolution of forecast data is 0.-5 degree)	Track and intensity errors	Atlantic, Eastern Pacific, and Western Pacific basins	surface	every 12-h up to 5 days
Highs/Lows	model analyses at 1-degree resolution	object-oriented MODE Verification	NH	500mb HGT, MSLP	every 6-h up to 7 days
Forecast Consistency of a field	model analyses at 1-degree resolution	Wald Wolfowitz test and other measures TBD	NH?	Precip fields, Jet Stream strength, TC location and strength, Extra-TC location and strength	Every cycle

Appendix C

Capability Summary Table as of April 2016

Notes: for MET+ entries, V5.2 will be released June 2016, V6.0 will be released in October 2016, V6.1 and V6.2 will be released in FY2017-2019. "In separate pkg" means capability in NHC TC display developed under HFIP funds for MET, or in a package developed by Mike Fiorino for GSD.

Capability	**Description**	**NCEP**	**GSD/MATS**	**MET+**
Time Series	X-axis is either lead time or specific valid times	x	X	x
Vertical profiles	Y-axis is pressure	x	X	x
Diurnal Plots	X-axis is valid times	x	X	x
Plan View - Geographic Representation	2D grid of statistics	x	X	x
Scatter Plots of Scores	e.g. CSI vs. Bias	x	X	V5.2 or v6.0
Performance Diagrams		x	X	X
Reliability Diagram		x	X	X
Taylor Diagram				V5.2 or v6.0
ROC Diagram		x		x
Binned Spread Skill Diagram				x
Matching Cyclone Tracks		x		V6.1
Threshold Quilt plot	Threshold on X-axis and Forecast Hour on Y-axis	x		V6.0
Die-off Plot	AC - Decomposed into wavenumbers	x		V6.0
T-p map plot	AC – fcst lead vs. initialization	x		V6.0
Regridding within tool	As opposed to regridding with copygb outside tool			x
Aggregation over time at each grid point	e.g. Monthly mean charts	x		X
Scorecard		x		V5.2
Fields/Formats				
Upper Air		x	x	X
PrepBUFR		x		x
AMDAR			x	x
Surface Variables - Hourly		x	x	x
Surface Stations		x	x	x
Precip Type			x	x
Gauge QPE		x		x

Capability	Description	NCEP	GSD/MATS	MET+
Stage II and Stage IV		x		x
MRMS				x
CCPA		x		x
CPC-Gauge		x		x
Alaska QPE		x		x
CMORPH			x	x
SYNOP-based Precip			x	V6.1 or 6.2
Sub-24hr precip		x	x	x
24hr precip		x	x	x
Ceilings		x	x	x
Visibility		x	x	x
PBL Height		x		x
Skin Temp		x		x
Snow Depth		x		x
Soil Temp		x		x
Soil Moisture		x		x
Cloud and solar insolation			x	x
SURFRAD/ISIS		x	x	x
Composite Reflectivity		x	x	x
1km AGL reflectivity		x		x
Hybrid scan reflectivity				
Echo Top Height		x	x	x
Convective Probability			x	x
CAPE		x		x
Best CAPE		x		V6.0
Updraft Helicity				X
VIL			x	X
Wind Profiler		x	x	X
Wind Tower			x	X
GPS-Met			x	X
Total Cloud - AF	WWMCA	x		X
Total Cloud from CLAVR		x		V6.1
U, V		x	planned	X
Wind Speed		x	x	X
Wind Direction		x	x	X
Hurricane Track		x	In separate pkg	x
Hurricane Intensity		x	In separate pkg	x
Extratropical Cyclone Track		x		x
Ozone		x		x

Capability	Description	NCEP	GSD/MATS	MET+
Smoke		x		x
Dust		x		x
Trop Height		x		x
Trop Temp		x		x
Transport Wind		x		v6.1
Ventilation Rate		x		v6.1
80-m wind		x		v6.1
Haines Index		x		v6.1
0-6km Wind Shear		x		v6.1
Brooks-Craven Severe Wx Index		x		v6.1
GSI Diagnostic CONV obs				x
GSI Diagnostic AMSU-A obs				x
GSI Diagnostic AMSU-B obs				x
Continuous Statistics				
Partial Sums		x	x	x
Mean Forecast and Observed Values	x		x	x
Standard Deviation		x	x	x
Ratio of Standard Deviation		x		x
Pearson's (conventional) Correlation Coefficient		x		x
Spearman's Rank Correlation				x
Kendall's Tau				x
Number of ranks, tied forecasts/obs for Kendall's Tau				x
Mean Error	Bias	x	x	x
Standard Error		x	x	x
Quantiles of Error				x
Inner Quartile Range				x
Multiplicative Bias				x
Mean Absolute Error		x		x

Capability	Description	NCEP	GSD/MATS	MET+
Total Mean Square Error (MSE)		x		x
MSE by Mean Difference		x		x
MSE by Pattern Deviation		x		x
Murphy's Skill Score		x		x
Bias-corrected Mean Square Error				x
Root Mean Square Error		x	x	x
Mean Absolute Deviation				x
Anomaly Correlation using User Derived Mean Climo field				x
Anomaly Correlation using NCEP binned Climo		x		V5.2
Anomaly Correlation using NCEP mean and standard deviation Climos		x	x	V5.2
Cyclone Cross-Track		x	In separate pkg	x
Cyclone Along-Track		x	In separate pkg	x
Categorical Statistics				
Contingency Table Counts (CTC)/ Forecast Hits Obs (FHO)		x		x
Forecast rate		x		x
Hit rate		x		x
Observation rate		x		x
Base Rate				x
Forecast Mean			x	x
Accuracy				x
Frequency Bias		x	x	x
Probability of Detection - Yes			x	x
Probability of Detection - No				x
Probability of False Detection (aka False		x	x	x

Capability	Description	NCEP	GSD/MATS	MET+
Alarm Rate)				
False Alarm Ratio		x	x	x
Critical Success Index (CSI - aka Threat Score)		x	x	x
Gilbert Skill Score (GSS - aka ETS)		x	x	x
Bias-Adjusted GSS		x		x
Odds Ratio		x		x
Log-Odds Ratio		x		x
Odds-Ratio Skill Score		x		x
Hanssen-Kuipers Discriminant (or True Skill Score)		x	x	x
Heidke Skill Score		x	x	x
Extreme Dependency Score and Index		x		x
Symmetric Extreme Dependency Score and Index				x
Identification of RI/RW events	e.g.NHC definition: 30kts over 24hr			X
Identification of Ramp Events	Large change in a field over a given time			x
Probability and Ensemble Related Statistics				
Rank Histogram		x		x
Probability Integral Transform (PIT) and PIT Histogram				x
Binned Ensemble Spread-Skill				x
Total Brier Score		x		x
Brier Score Reliability		x		x
Brier Score Resolution		x		x
Brier Score Uncertainty		x		x
Brier Skill Score using User Defined Probability Climatology				x
Brier Skill Score using NCEP methodology		x		V5.2
BSS using another model as reference		x		V5.2
CRPS		x		x
CRP Skill Score using				x

Capability	Description	NCEP	GSD/MATS	MET+
User Defined Probability Climatology				
CRP Skill Score using NCEP Probability Climatology		x		V5.2
CRPSS using another model as reference		x		x
RPS		x		V5.2
RP Skill Score using User Defined Probability Climatology				V5.2
RP Skill Score using NCEP Probability Climatology		x		V5.2
RPSS using another model as reference		x		V5.2
Economic Value		x		V5.2
Points for Receiver Operating Characteristic (ROC) Curve		x		x
Area Under ROC				x
Points for Reliability Diagram		x		x
Calibration				x
Refinement				x
Likelihood				x
Base Rate for Probability Forecast				x
Relative position		x		V5.2
Low 10% Brier Score		x		V5.2
High 10% Brier Score		x		V5.2
Probability Anomaly Correlation (PAC)		x		x
Neighborhood Statistics				
Fraction Brier Score				x
Fraction Skill Score		x		x
ETS with Max value match		x		
Asymptotic Fraction Skill Score				x
Forecast/Observed Event Frequency				x
Spatial Decomposition				

Capability	Description	NCEP	GSD/MATS	MET+
Decomposition of wavenumber using Fourier Transform		x		V6.0
Decomposition of wavenumber using Haar Wavelet				x
MSE for each scale				x
Intensity Skill Score				x
Forecast/Observed Energy Squared				x
Base Rate				x
Frequency Bias				x
Object Based Methods				
Method for Object-based Diagnostic Evaluation (MODE)				x
MODE-TD (time domain)				x
Contiguous Rain Area (CRA)				
Image Warping Techniques				
FIQAS object based methods				
Confidence Intervals (CIs)				
Normal Approximation		x	x	x
Bootstrapped		x		x
Removing Autocorrelation from CIs using Variance Inflation Factor				x
Removing Autocorrelation from CIs using Trend Analysis				
Student-t		x		
Regional Verification				
NCEP Regions		x	x	x
Storm-centric				x
Intersections, Unions, Symmetric Differences				x
Land Use			x	x

Capability	**Description**	**NCEP**	**GSD/MATS**	**MET+**
Alaska		x	x	x
NH		x	x	x
SH		x	x	x
Trop		x	x	x
PNA		x	x	x
TC basins		x	x	x

Appendix D

GSD/Earth Modeling Branch Verification products:

Cluster	Product
RAOB	Upper Air Time-series
	Model-RAOB residuals (text)
	Vertical profiles
AMDAR	AMDAR time-series
	AMDAR vertical profiles
AC	Anomaly correlation - 5-day, 6-day, 7-day
	Anomaly correlation time-series
Ceiling	Ceiling hourly - time-series
	Ceiling hourly - plan view
Surface	Surface statistics - hourly time series
	Surface statistics by land use
	Individual station and network stats (restricted) (Text and plan view)
Precip	Precip type - time-series and diurnal
	SYNOP-based Precipitation time-series
Visibility	Visibility hourly - time-series
	Visibility hourly - plan view
Cloud/Solar Insolation	Cloud and Solar Insolation - time-series
	Cloud and Solar Insolation - diurnal
	Cloud and Solar Insolation - plan-view GSIP, HRRR for specific times
	Cloud and Solar Insolation - plan-view from surface stns
	Cloud and Solar Insolation - SURFRAD/ISIS stations
Composite Reflectivity	Composite Reflectivity - hourly time-series
	Composite Reflectivity - valid time

Cluster	Product
	Composite Reflectivity - Lead time
	Composite Reflectivity - performance diagrams
Vertically Integrated Liquid	Vertically Integrated Liquid - hourly time-series
	Vertically Integrated Liquid - valid time
	Vertically Integrated Liquid - Lead time
	Vertically Integrated Liquid - performance diagrams
Echo Top Height	Echo Top Height - hourly time-series
	Echo Top Height - valid time
	Echo Top Height - Lead time
	Echo Top Height - performance diagram
Sub-24hr Precip	Sub-24hr precip - time-series
	Sub-24hr precip - threshold
24hr Precip	24HR precip - time series
	24HR precip - threshold
Convective Probability	Convective Probability - hourly time-series
	Convective Probability - CSI vs Bias
	Convective Probability - ROC
	Convective Probability - Reliability Diagram
Wind	Wind profiler - time-series (non-java)
	Wind profiler - vertical structure
	Wind tower - time-series (restricted, non-java)
GPS-Met	No longer in use