



NGGPS Verification and Validation Team Plan Overview

HIWPP/NGGPS Meeting

Ivanka Stajner, Glenn White
February 10, 2016



NGGPS Verification and Validation Team Objectives



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



Team Members

- Ivanka Stajner (NWS/STI), co-lead
- Glenn White (NWS/EMC), co-lead
- Geoff DiMego (NWS/EMC)
- Tara Jensen (NCAR and DTC)
- Bonny Strong (OAR/ESRL/GSD)
- Geoffrey Manikin (NWS/EMC)
- Stephen Weygandt (OAR/ESRL/GSD)
- Fanglin Yang (NWS/NCEP)



Purpose of NGGPS Verification Package



- Provide quantitative measures to support an evidence–based approach towards decision making and NGGPS development
- Primary users of the verification system and its products will be:
 - NGGPS developers and users of NGGPS products
 - NGGPS program office, EMC, NCEP operational centers, NOAA laboratories, and NOAA managers
 - Research community, private sector, and universities



NGGPS Validation



- Is NGGPS meeting stakeholder needs?
 - Accuracy of predicting variables for certain thresholds
 - Ability to predict specific phenomena
- Stakeholder assessment and feedback



State of Verification and Validation Systems



- **NCEP**

- Global verification focuses on large-scale flow pattern over the globe and long-term statistics of model performance, ***increasing emphasis on mesoscale verification***
- Mesoscale verification focuses on synoptic events and sensible weather elements

- **HIWPP**

- Unifies metrics currently produced by NCEP/EMC and GSD/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



NCEP Current Status



Main Verification Web Page http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/, including 1) verification statistics of AC, RMSE, Bias etc for major international NWP models and GFS implementation parallels in the past 31 days, 2) real-time weather forecast maps of GFS and GFS implementation parallels, 3) links to other verifications.

Grid-to-Obs Verification http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/g2o/ and <http://www.emc.ncep.noaa.gov/gmb/ssaha/>

Including 1) verifications of surface 2-m T, RH, Td, 10-m winds, SLP and total clouds against ground observations over the CONUS and its sub-regions and, 2) verifications of atmospheric T, Q, RH and Winds against rawinsonde and aircraft observations over the globe and its sub-regions.

Precipitation Verification

http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/www/rain2/rain.html

Including precipitation forecast maps verified against CCPA over the CONUS and CPC gauge observations over the globe, and precipitation Equitable Threat Scores for major international models

<http://www.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/>, including operational model scores (US and Int'l), links to parallel model scores and daily side-by-side precip field comparisons.



NCEP Current Status cont.



Objected-Oriented (MODE) Verification

http://www.emc.ncep.noaa.gov/gc_wmb/tdorian/

Including MODE verifications of precipitation over CONUS and jet streams over the globe.

Historical Performance http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/longterm/

Including annual review of GFS forecast skills and historical performances of major international NWP models.

Ensemble Forecast Verification

http://www.emc.ncep.noaa.gov/gmb/STATS_vsdb/ensm/, and

<http://www.emc.ncep.noaa.gov/GEFS/verif.php>

including GEFS, NAEFS and other international global ensemble forecasts.

Data Assimilation Monitoring <http://www.emc.ncep.noaa.gov/gmb/gdas/>

GFS Experimental Parallels Verification:

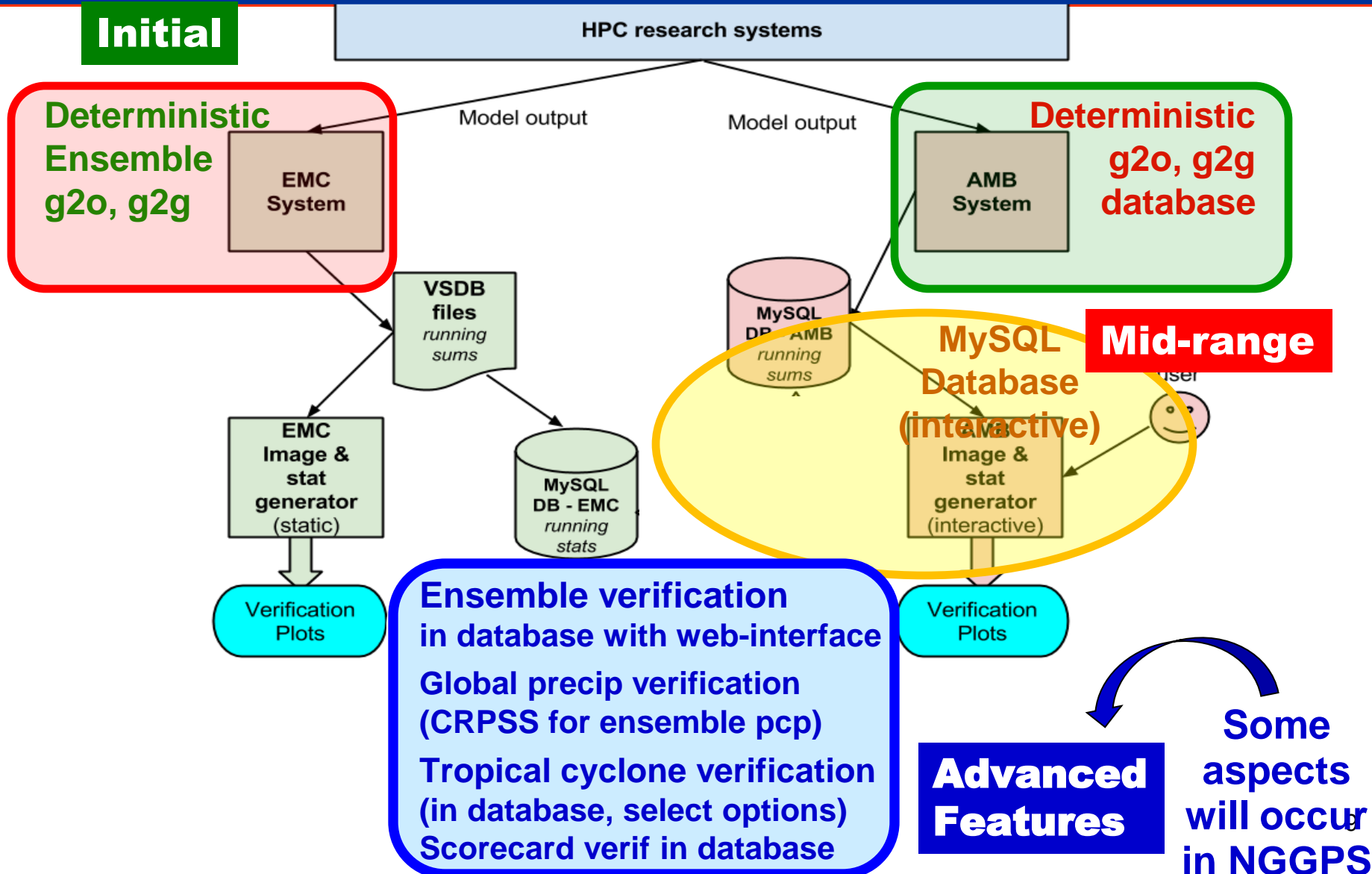
<http://www.emc.ncep.noaa.gov/gmb/wd20rt/vsdb/> and

<http://www.emc.ncep.noaa.gov/gmb/wx24fy/vsdb/>

Others : <http://www.emc.ncep.noaa.gov/GEFS/perf.php> contains a list of all verifications related to GFS and GEFS. <http://www.emc.ncep.noaa.gov/gmb/STATS/MAPS.html> presents daily weather forecast maps.

Merging HIWPP Verification Systems

Merging EMC and AMB Verification Systems



MET Package

- MET is community code supported by DTC that is free to download (registration required)
 - 2950+ registered users
 - 126 countries, 29% from USA
 - Universities, Government, Private Companies, Non-Profits
- Download MET release and compile locally.
 - Register and download: www.dtcenter.org/met/users
 - C++ with calls to some Fortran libraries
 - Linux with GNU, Portland Group (PGI), or Intel compilers
- Support
 - Online tutorial and in-person usually tutorials given yearly
 - met_help@ucar.edu help desk
 - 176 support tickets in past year

Originally developed to replicate the NOAA/NCEP verification package



Areas of Potential Improvement



- Treatment of uncertainty in the analyses and observational data set that forecasts are verified against (*a priority area in the current FFO*)
- Scorecard and weighted performance indices (global scale and sensible weather)
- Ensemble and probability verification metrics
- Unification of packages capturing strong elements from different systems
- Diagnostic tools (scales at which errors occur, energy spectra), physics-oriented metrics (radiation, fluxes, cloud verification)
- Measure of forecast consistency
- Metrics for extreme weather events
- Object-oriented metrics
- Component performance
- System performance (including fluxes and interface variables among components)
- Validation - enable early and comprehensive user involvement



Verification and Validation Strategy



- Develop a flexible and comprehensive verification package through unification of capabilities from NCEP, ESRL and DTC verification packages (*leveraging HIWPP efforts*)
- 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.
- Coordinate NGGPS validation with UMAC evaluation of NCEP production suite and Model Evaluation Group activities.



Future of Verification and Validation System



- Unify the approach based on MET and METviewer
- Examples of some skill metrics/capabilities to consider/add:
 - What other verification fields are desired?
 - What other types of error measures?
 - Preset vs. on-the-fly skill score assessments (or both)?
 - Database and web interface aspects?
 - Precipitation and reflectivity verification (also novel fields like solar irradiance, etc.)?
 - Ensemble, tropical cycle and scorecard verification?
- Recent Federal Funding Opportunity released included request for proposals to address the treatment of observation uncertainty.



NGGPS Validation



- Coordinate with model evaluation group (MEG) activities
 - STI-funded project
 - Scope:
 - Form three groups for global model evaluation, development of convection-permitting ensemble, communications and dissemination
 - Develop improved model evaluation and implementation processes
 - Provide recommendations for determining membership of planned NCEP convection-permitting ensemble and assist with initial evaluation and testing
 - Establish trial visitors' program between EMC and rest of NWS
 - Improve access to operational and experimental model output
 - Unify verification with improved public access
 - Strategy:
 - Involve field in real-time and retrospective evaluation of science upgrades
 - Identify case studies and provide data for extended evaluation period beyond last 30-day parallel
 - NCO 30-day parallel is only for IT evaluation



MEG: Role of users in model evaluation



- EMC has learned the hard way we need forecasters' perspective to spot problems and assess their significance and to evaluate significance of changes
- Forecasters' concerns lead to new verification metrics
- Forecasters have own metrics
- EMC forecast systems subject to forecaster review before implementation
- Plan:
 - Forecasters work with developers to identify and prioritize problems and to develop improvements
 - Forecasters review tests of fixes to problems and tests of proposed improvements
 - Forecasters review real time and retrospective tests of experimental forecast systems
 - Real time tests on AWIPS
 - Retrospectives maps generated, data available



MEG: GFSX Independent Field Evaluation



Region/Center	Remarks
Western Region	<i>Evaluation continuing. Preliminary assessment neutral.</i>
Central Region	<i>Neutral to positive.</i>
Southern Region	<i>Evaluation continuing. Not sure yet.</i>
Eastern Region	<i>Evaluation continuing. Not sure yet.</i>
Pacific Region	<i>Not participating in the evaluation due to staffing issues</i>
Alaska Region	<i>Evaluation continuing. Preliminary assessment neutral</i>
WPC	<i>Evaluation continuing. Not sure yet.</i>
NHC	<i>Evaluation continuing. Not sure yet.</i>
SPC	<i>Evaluation continuing. Not sure yet.</i>
AWC	<i>Preliminary Assessment: IMPLEMENT</i>
CPC	<i>Evaluation continuing. Not sure yet. Upper stratos. T forecasts worse</i>
OPC	<i>Evaluation continuing. Not sure yet.</i>
SWPC	<i>Interested in participating</i>
National Water Center	<i>Evaluating GFS hourly output, proceed with implementation</i>
MDL	<i>No harm done... proceed with implementation</i>



FY16 Tasks/Milestones



- **Begin unification of the verification approach starting from MET and MET viewer (0.75 FTE)**
 - 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 MET viewer system at NCO (0.5 FTE)**
 - Set up initial system and evaluate outputs
 - Set-up initial capability for scorecarding and visualization of statistics



FY16 Tasks/Milestones (cont.)



- **Identify database schema that is consistent and suitable for use at NCO (0.75 FTE)**
- **Develop a procedure for inclusion of new verification metrics (0.75 FTE)**
- **User support, training, documentation for MET and MET viewer (0.75 FTE)**
- **Coordination and planning (0.75 FTE)**
 - Establish a Focus Group of stakeholders
 - Gather stakeholder needs



Identifying NGGPS metrics



Compilation began with existing EMC metrics and will identify specific gaps

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.

Top of the table in the appendix of the implementation plan is shown above



Long Term FY17-FY19 Tasks/Milestones



- **Comprehensive verification system for operational and developer use (3 FTE/year)**
 - 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
 - System performance
- **Database development and optimization (1.5 FTE/year)**
- **User Support (2 FTE/year)**



Long Term FY17-FY19 Tasks/Milestones (cont.)



- **Validation and implementation decision support (1.5 FTE/year)**
- **Data repository for verification data sets and quality control (1.5 FTE/year)**
- **Inclusion of additional verification datasets (3 FTE/year)**
- **Visualization and user interface improvements (2 FTE/year)**
- **Treatment of uncertainty in the analyses and observational data set that forecasts are verified against (topic for FFO)**



NGGPS Verification and Validation

Three Main Points



- One Major Accomplishment:
 - NGGPS Verification team and EMC management agreed on MET as the unified verification tool

UMAC recommends that "NCEP unify its verification systems, and migrate toward a community verification system, based on infrastructure such as MET and METviewer, with comprehensive and regionally specific statistics."

- One priority focus effort for FY16:
 - Enhancing MET to include all current EMC verification statistics
- One most important issue or coordination need:
 - Identification of metrics important for ocean, sea ice, wave, land surface, aerosol and chemistry components and for system coupling for inclusion in the verification package



Summary



- Develop a flexible and comprehensive verification package for the evaluation of progress in the development and operational readiness of NGGPS and of future NGGPS operational performance through:
 - unification of verification approaches at NCEP, ESRL and DTC using MET (*consistent with UMAC recommendation*)
 - leveraging HIWPP verification effort
 - use of a more flexible database approach
 - addition of standard, sensible weather, ensemble, extreme weather, object-oriented verification, diagnostic and process-oriented tools (prioritized lists requested from NGGPS implementation team leads)
 - inclusion of NGGPS component and system verification
- Validation will be coordinated with Model Evaluation Group activities
 - extended involvement of stakeholders early in the process of testing of proposed model upgrades



Backup slides



Examples of GFS verification gaps



- Forecast consistency from cycle to cycle
- Monitoring of extreme cold temperatures near the surface
- Hurricane track and intensity verification to day 7 (currently to day 5) and significance
- Sensible weather over the globe (currently for CONUS and Alaska)
- Quantitative Precipitation Forecast over the globe
- Cloud verification against satellite products



Identifying NGGPS Variables



- Initial variables from EMC, identify additional potential variables
 - Height
 - Temperature
 - Pressure
 - Winds
 - Precipitation
 - Temperature
 - Winds
 - Tropical cyclone track and intensity verification



NCEP Verification System



- Verifications of global, regional and ensemble forecasts at NCEP/EMC are carried out independently by different branches and groups
- The current GFS verification system evaluates several metrics
- NCEP's Global NWP Model Verification package includes:
 - Computation of model forecast statistics for global NWP model simulations
 - Comparison of statistics among different model simulations
 - The data is saved in Verification Statistics Data Base (VSDB) format
- Examples of statistics include:
 - Anomaly Correlation (AC)
 - Root Mean Square Error (RMSE) for Geo-Potential Height (HGT), Temperature (T) and Vector Wind (Wind)
- NCEP's Mesoscale Modeling Branch (MMB) plots VSDB output using Forecast Verification System (FVS) and is moving towards using the METViewer package to display VSDB files



HIWPP Verification System



- The current system verifying model output from the hydrostatic global models participating in HIWPP is publically available at <http://hiwpp.noaa.gov/verify/>
- 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
- Products continue to be added
- Security issues led to modified deployment of Model Assessment Tool Suites (MATS) based on html5 technology
- Addressing three areas to advance metrics:
 - 1) ensemble verification
 - 2) a multi-parameter scorecard
 - 3) enhanced sensible weather verification, in particular relating to global precipitation verification



DTC MET Verification System



- 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
- Originally developed based on the NCEP grid2obs, grid2grid and FVS verification packages with the goal of supporting these capabilities to the community through a platform independent and extensible software package
- Designed to be modular and adaptable
 - Individual modules can be applied
 - New tools can be added
- Computes over 50 traditional statistics including:
 - Bias, Root Mean Square 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), and Brier Score (BS)



HIWPP Verification Development

- 1. Initial System – EMC verification package (VSDB output)**
-- run within HIWPP (basic stats, reference) **Initial**
- 2. Mid-range System -- MySQL database system** **Mid-range**
-- Incorporates EMC and AMB verification
global: upper-air, AC (work toward surface / precip)
conversion package from VSDB → database in place
-- Basic verification system with interactive database)
- 3. Advanced System -- Fully merged system with additional capabilities (ensemble verification, global surface, global gridded and station-based precipitation)** **Advanced**



HIWPP Verification metrics / attributes



Variable	Levels	Area	Scores	Forecast Range (Hours)
Height	500 hPa	NH, SH	ACC, RMSE, Spread, CRPS	0 to 384
Height	1000 hPa	NH, SH	CRPS	0 to 384
Temperature	850 hPa	NH, SH	CRPS	0 to 384
Pressure	Surface	NH, Tropics	Track Error	0 to 120
Winds	850, 200 hPa	NH, Tropics	CRPS, RMSE	0 to 384
Precipitation		GLOBAL	ETS, CRPS, Bias	0 to 384
Temperature	2 meter	NH	RMSE, bias, CRPS	0 to 384
Winds	10 meter	NH	RMSE, bias, CRPS	0 to 384
Tropical cyclone track and intensity verification				