NGGPS
Verification and Validation Team Overview

NGGPS Program Meeting

Ivanka Stajner, Glenn White
August 5, 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 and DTC)
- 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 before NGGPS

- **NCEP**
  - Global verification focuses on large-scale flow pattern over the globe and long-term statistics of model performance,
  - Increasing emphasis on mesoscale verification: 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.

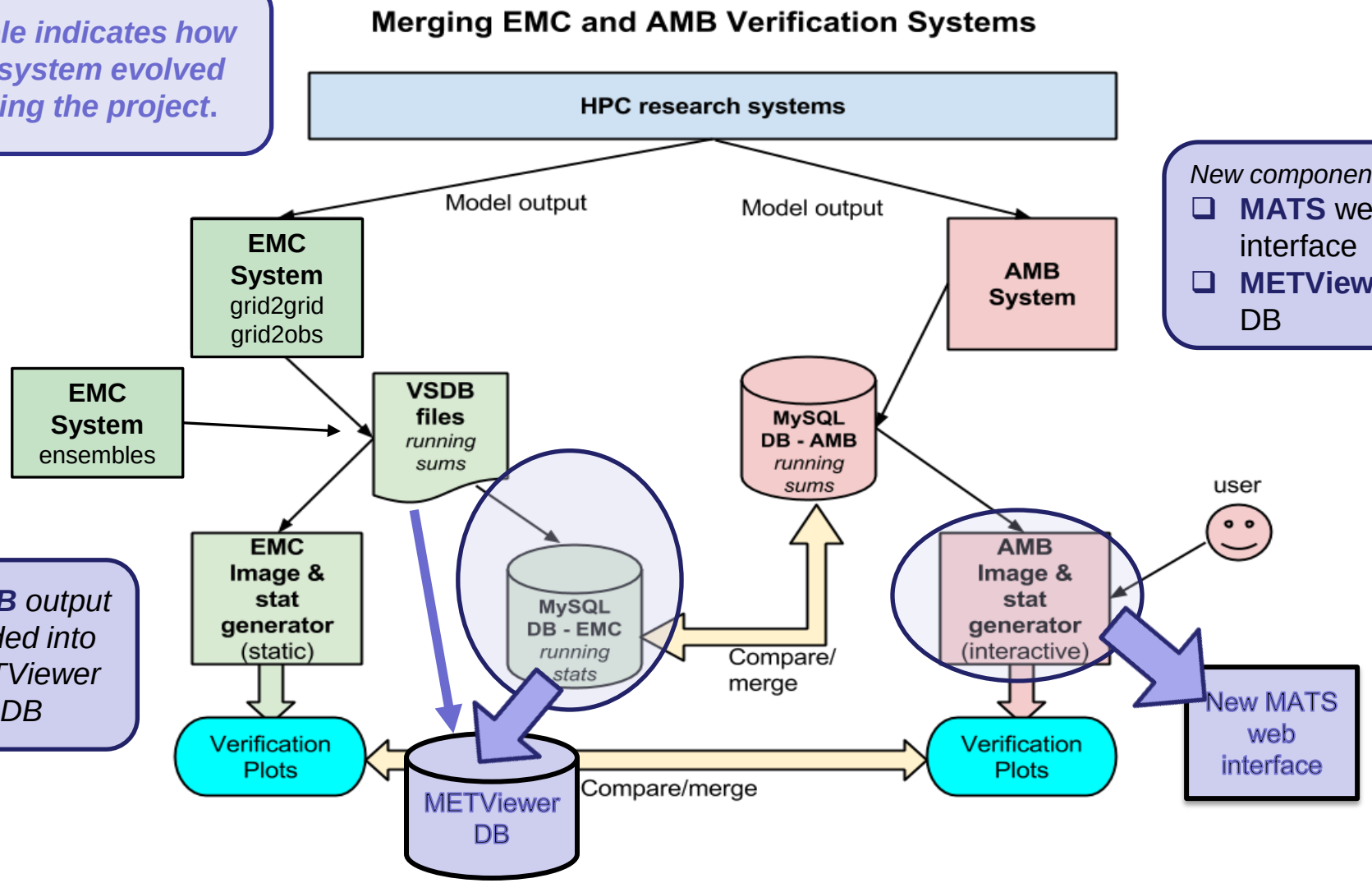
Data Assimilation Monitoring http://www.emc.ncep.noaa.gov/gmb/gdas/

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/GFS/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.

HIWPP Verification

Purple indicates how the system evolved during the project.



New components:

- **MATS** web interface
- **METViewer** DB

***VSDB** output loaded into METViewer DB*

MET Package

- MET is community code supported by DTC that is free to download (registration required)
  - 3100+ registered users from 126 countries, 29% from USA
- Download MET release and compile locally
  - Register and download: **www.dtcenter.org/met/users**
- Support
  - Online tutorial and
  - Next in-person tutorial to be given at NCWCP in early 2017
  - **met_help@ucar.edu** help desk
- Approved to be transferred to run at EMC

Originally developed to replicate the NOAA/NCEP verification package

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.

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

Priorities established with EMC in January 2016

- Transition MET to run at EMC and extend it to reproduce all current EMC global metrics.
- Identify and begin including metrics for ocean, ice, land, aerosols and system coupling
- Identify and begin including process oriented metrics
- Document requirements
- Develop a high performance database for large amounts of model and observational data that is suitable for NCEP's operational environment

In-depth User Meetings at NCEP

- In order to get full understanding of current usage and future needs
- Held May 2-6, 2016
- V&V Team members Tara Jensen and Bonny Strong and DTC/MET developer John Halley Gotway
- 50 participants from NCEP
- 18 information-gathering sessions held
- METViewer training for ~20 people

Information-gathering Sessions

- Global Deterministic
- Global Ensemble
- Mesoscale - Model Developers
- Mesoscale - Verification Developers
- Mesoscale - Ensembles
- Model Evaluation Group
- Fit-to-Obs Developers
- Data Assimilation
- UGCS Developers
- Tropical and Extra-Tropical Cyclones
- Sea Ice
- Aerosols
- Oceans and Waves
- Process-oriented Metrics
- NCO – METViewer installation
- NCO – IDP and general
- CPC
- WPC

**Information gathering with Land Surface, Hydrology and Space Weather teams are being planned*

Verification Focus

- **Global Deterministic and Ensemble**
  - –Initial focus on needs for global model development and evaluation
- **Regional Deterministic and Ensemble and MEG**
  - –Also ensuring these needs can be met as requirements are defined
- **Coupled components**
  - –Early reviews completed, but will be further elaborated over the next year
  - –In coordination with
    - •Unified Global Coupled Analysis and Forecast System (UGCS)
    - •Other NGGPS teams
  - –Including:
    - Aerosols
    - Sea Ice
    - Ocean and Waves
    - Land Surface
    - Space
    - Hydrology

Verification Focus - cont.

- **Process-oriented metrics**
  - –Initial review begun
  - –Will require further elucidation and coordination with other groups
- **Database**
  - –Year 1 efforts to address only essential requirements for support to EMC
- Coordination and communication with all NGGPS teams is **essential**
- Requirements document draft completed and sent to EMC POCs and NGGPS team leads for review

Verification status update

- End-to-end system will be called **MET+** includes
  - MET
  - METViewer
  - Python wrappers
- MET installed on development side of WCOSS and IDP, and on Theia and Jet
- METViewer installed for EMC internal use only
- Work has begun on python-based system to replicate Fanglin's auto-generated global model images
- MET/METViewer in use currently by GMTB
- Reviewing NCO requirements for database and other operational constraints

NGGPS Validation

- Coordinate with model evaluation group (MEG) activities
  - –MEG synoptic model evaluations presented weekly mainly by EMC
  - –Also involved in 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

NGGPS Validation progress

- MEG progress
  - –New procedures followed for GFS2016 evaluation and implementation
  - –Longer period for evaluation, case studies proposed by field
  - –Field found evaluation greatly improved
  - –Increased participation by EMC in
    - •HWT experiment in Norman, Oklahoma May 2016
    - •Flash flood experiment WPC summer 2016
  - –Visits by EMC to forecast offices being planned

Problem—How to get more, higher resolution forecast products and experimental forecasts to forecast offices
Problem being addressed through AWIPS II gradually

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: GFS16 Independent Field Evaluation

Region/Center	Recommendation	Remarks
Western Region	**Implement**	Neutral
Central Region	**Implement with reservations**	*Little improvement*
Southern Region	**Implement**	No striking differences
Eastern Region	**Implement**	Minor improvements
Pacific Region	**Implement**	Models performed well with Winston
Alaska Region	**Implement**	No specific problems
WPC	**Implement**	Similar, GFSX slightly better sometimes
NHC	**Neither endorse nor oppose**	Improved tropical forecasts, downstream tests for HWRF incomplete *(70% of the retrospectives completed as of today)*

MEG: GFS16 Independent Field Evaluation

Region/Center	Recommendation	Remarks
AWC	**Implement**	Better winds, temperatures
CPC	**Implement**	*Large errors upper stratosphere*
OPC	**Implement**	Extratropical storm tracks better
SWPC	**Implement**	*Need improvements in upper atmosphere*
MDL	**Implement**	Redeveloped MOS better
NWC	**Implement**	Hourly files should improve NWC fcsts
SPC	**Implement**	Improved in warm season
Weather It Is Ltd. (Prof. Barry Lynn)	**under situations where the observational network is more dense, there has been improvement in the initial state (and lateral boundary conditions) of the GFSX compared to GFS**	
AccuWeather	**Hourly output is of significant value for Weather Industry**	

FY16 Tasks/Milestones

- **Begin unification of the verification approach starting from MET and METViewer**
  - Now called MET+
  - 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+ system at NCO**
  - Set up initial system and evaluate outputs
  - Set-up initial capability for scorecarding and visualization of statistics

FY16 Tasks/Milestones

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

Progress in FY16

- Implementation plan finalized and posted on NGGPS website
- NCAR, ESRL and DTC team visited NCEP to collect requirements and submitted an NGGPS Verification User Needs and System Requirements document to EMC POCs and NGGPS team leads for review
- MET and METViewer enhanced with some of the missing EMC Global verification capabilities and development began on python scripts for the MET+ unified system
- NCEP verification system was used to evaluate forecast skill of candidate NGGPS dycores with GFS physics http://www.emc.ncep.noaa.gov/gmb/wx24fy/nggps/web
- 2015 FFO grant funding: *Developing Physics-oriented Diagnostic Tools for Model Evaluation and Improvement*, Zhuo Wang, Univ. of Illinois
- Verification is one of the topics in the recent 2016 FFO

Developing Physics-oriented Diagnostic Tools for Model Evaluation and Improvement

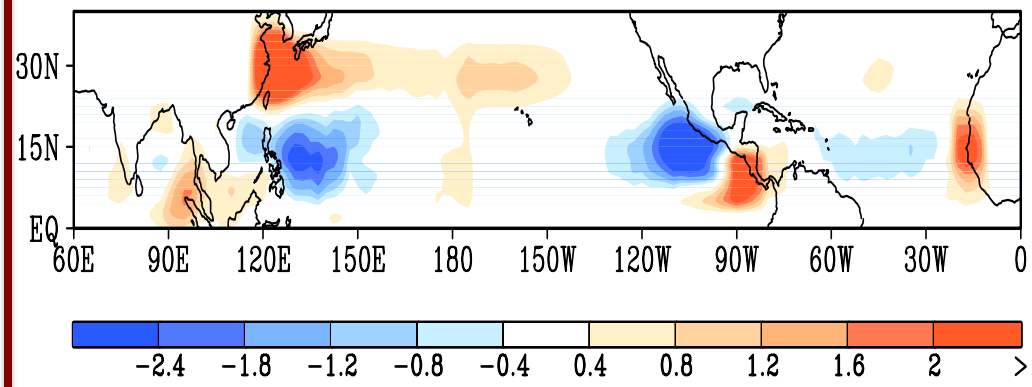
PI: Zhuo Wang, University of Illinois at Urbana-Champaign
Co-Pis: Stan Benjamin, Melinda Peng and Ming Zhao

- **Objective**: develop physics-oriented diagnostic tools to assist the development of the NOAA's Next Generation Global Prediction System (NGGPS).
- **Significance**: Physics-oriented evaluation not only provides information on how well a model performs, but also on why a model may fail in a certain aspect.
- **A special focus** of the project:
  - high impact weather: such as tropical cyclones, heat waves and drought
  - systems that play an important role in extended range forecasts: such as the MJO, midlatitude blocking, and teleconnections.
- The effort will help to improve operational forecasts on the synoptic time scale and extend skillful forecasts over a longer time range (up to 30 days).

- **Deliverables**: a suite of diagnostic tools with general applicability across models.
- **Transition**: Some of the developed tools have been transitioned to the DTC Global Model Test Bed

An Example: Link Tropical Cyclogenesis Biases to the Model Physics

Biases of Tropical Cyclogenesis in GEFS

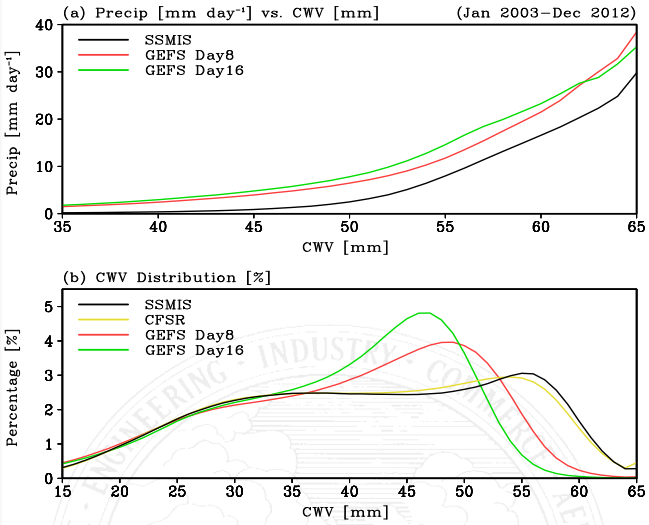


Large biases in tropical cyclogenesis exist on the regional scale.

TC genesis biases are related to biases in diff. large-scale circulations over diff. basins

- a weaker-than-observed monsoon trough over the West Pacific
- a southward displaced ITCZ over the East Pacific
- hyperactive Africa easterly waves over the East Atlantic

Evaluation Against Satellite Data



Evaluation reveals **a dry bias in column water vapor (CWV)** and indicates that **precipitation is initiated too early with respect to CWV accumulation** in the GEFS → **Deficiencies in the cumulus scheme**.

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

- Major Accomplishments
  - 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."

  - Transition of MET+ to NCEP has begun
  - Implementation plan completed
- Priority focus effort for FY16:
  - Enhancing MET to include all current EMC verification statistics
- 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 – input for *NGGPS Verification User Needs and System Requirements document*

Summary

- Implementation plan posted
- 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
  - inclusion of NGGPS component and system verification
  - NGGPS verification user needs and system requirements document drafted: input requested from EMC and implementation teams (e.g. addition of standard, sensible weather, ensemble, extreme weather, object-oriented verification, diagnostic and process-oriented tools)
- 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

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.

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

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)

Recent Additions to MET+ that support NGGPS needs

- MET
  - Computation of anomaly correlation and skill scores using climatologies
  - Added cosine latitude weighting
  - Ability to read non-NCEP grib tables (e.g. ECMWF, UKMET files)
  - Support for extended GRIB PDS for ensemble member metadata
  - Additional masking options (e.g. circle or square about a lat/lon point)
  - Allow Multiple thresholds in MODE config files to speed up processing
  - Improved handling of WWMCA cloud analysis files
  - Finer control of point-observation handling to allow more representative pre-processing of sporadic data (e.g. AeroNET data)

- METViewer
  - Ability to load VSDB files into database
  - Scripts to help EMC staff with precise database management
  - Performance ***(completed)*** and Taylor Diagrams ***(in testing)***
  - Derived curves now supports Difference, Ratio and Skill Score capability
  - Improved event equalization logic to produce a homogeneous dataset
  - User support

*Leveraging general development of MET+

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**
   -- Incorporates EMC and AMB verification **Mid-range**
   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				