Overview of model development and validation at EMC

2018 DTC COMMUNITY UNIFIED FORECAST SYSTEM TEST PLAN AND METRICS WORKSHOP

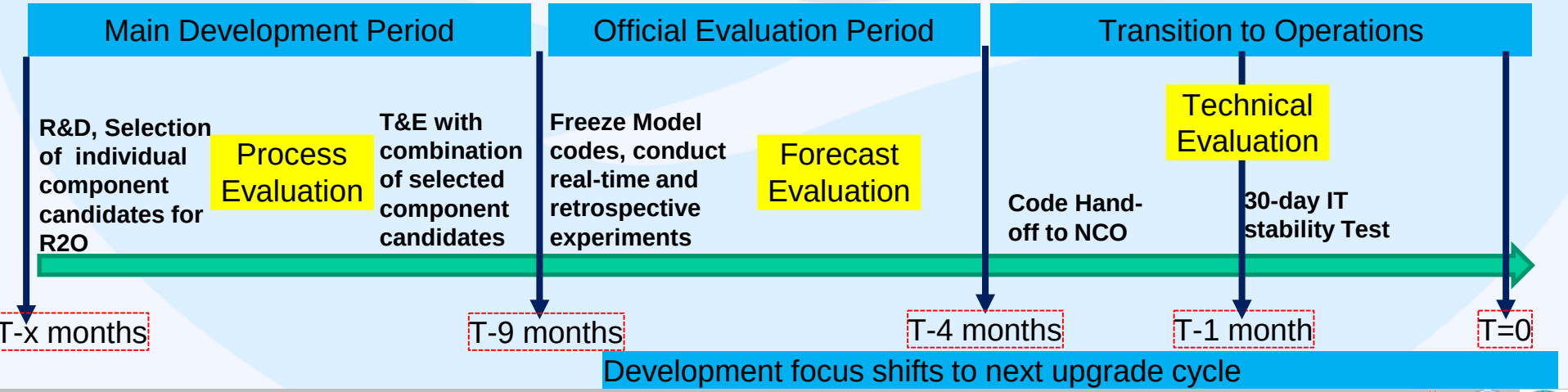
Vijay Tallapragada
Chief, Modeling and Data Assimilation Branch
NOAA/NWS/NCEP/EMC
July 30, 2018

Outline

- Model Development Process at EMC
- Verification, Validation and Evaluation
- Evidence Based Decision Making
- Requirements for Operational Model Implementations
- Priorities from R2O Perspective

Model Development and Implementation Process at EMC

- Model implementation process has finite life cycle that's repetitive
- Development efforts are continuous, R2O efforts play a big role
- Research Priorities are determined by past performance and user requirements
- Scientifically and technically mature developments pave their way into operations
- Model development focuses on Process Evaluation and potential for forecast improvements
- Pre-Implementation Testing and Evaluation focuses on establishing forecast improvements
- Verification and Validation happens throughout the development and implementation cycle



Verification Capabilities: Global Models

Core verification capabilities at EMC:

- Fit2Obs, Grid2Obs, and Grid2Grid verification using Verification System DataBase (VSDB) for current operational Global Models (GFS, GEFS, CFS, NGAC, WAM etc.)

1. Long term stats for historical performance of multiple models
2. Ensemble verification stats for multiple global ensembles
3. Assimilation monitoring and stats for data assimilation evaluation
4. Scorecard to summarize various conventional metrics (anomaly correlation, bias and RMSE)
5. Tropical Cyclone track, intensity and genesis verification
6. Extratropical storm track verification
7. QPF/Precip verification (including Equitable Threat Score; Bias Score; and object oriented QPF verification using MODE)
8. Murphy Relative Mean Square Error (MSE) Skill Scores
9. Ratio of Standard Deviation between Forecast and Analysis
10. RMSE from Mean Difference and Pattern Variation
11. Pattern correlation

2-D spatial plots and vertical cross-sections to augment objective verification

Verification Capabilities: Mesoscale Models

Core verification capabilities at EMC for North American Region (Forecast Verification System; FVS):

In addition to global model verification capabilities, operational verification of high-resolution models at EMC include:

1. Daily precipitation verification stats and skill scores for various operational models (NAM (12 km parent and nests), RAP, HRRR, GFS, International Models (CMC (regional and global), UK Met, ECMWF, JMA, DWD, Meteo France), HiResWindow runs, SREF mean, NAM/GFS parallels, DGEX, WPC QPF forecast)
2. Interactive Grid2Obs and Grid2Grid verification
3. Verification of cyclogenesis and extratropical cyclone tracks
4. Verification of Upper air, surface and severe weather elements

2-D spatial plots and vertical cross-sections to augment objective verification

Verification Capabilities: Ensemble Systems

Core verification capabilities at EMC for Global Ensemble Systems:

1. Climatology is derived from NCEP/NCAR Reanalysis
2. Talagrand Distribution for probabilistic evaluation
3. Ranked Probability Score (RPS) and Ranked Probability Skill Score (RPSS)
4. Continuous Rank Probability Score (CRPS)
5. Brier Score (BS) and Brier Skill Score (BSS)
6. Reliability diagrams and Relative Operating Characteristics (ROC)
7. Economic Value of Forecasts
8. Anomaly Forecast and Extreme Forecast Index
9. Performance diagram (success ratio vs. POD)

- Verification of probabilistic forecasts is based on a combination of deterministic verification (ensemble mean) and probabilistic verification of ensemble members (spread, bias, error)
- Bias correction (systematic error correction) is a required step for long-term forecast evaluation

Verification Capabilities: Coupled System Components

Core verification capabilities at EMC for Marine Systems (Deep Ocean; Waves; Sea Ice; Coastal Waters):

Weather Scales:

1. SST, SSS, SSH (surface fields)
2. Longitude-Depth Cross-Sections for temperature, salinity and velocity
3. MLD; OHC; Barrier Layer Depth (BLD); and TCHP
4. Significant Wave Height and Peak Period for Waves
5. Sea Ice concentration; thickness; drift; snow thickness on ice; Sea Ice/Snow temperatures
6. Coastal Ocean: RMSE; Central Frequency; Outlier Frequency; Taylor Diagrams

S2S Scales (Coupled Systems):

1. Ocean-Wave coupling: MLD; Stokes drift
2. Air-Ocean-Ice coupling: Heat flux modulations
3. Air-Ice coupling: Ice temperature and 2m Temperature correlations
4. Wave-Ice coupling: Scattering and damping
5. Large-scale indices including MJO; ENSO; IOD; NAO; PNA; QBO; SSW; Blocking;
6. Teleconnections; Weather Regime transitions; Rossby/Kelvin Waves; Monsoons

Verification Metrics for High Impact Weather Forecasts

- High Impact Weather Forecasts form a core component of model improvement validation
- Sensible weather elements are often non-prognostic quantities (e.g., CAPE)

- Need to capture typical errors in timing / location / intensity of events
- Reliability is an important factor for making better decisions
- Guides model developers to focus on areas for improvement

- Complicated by rarity of the events; sampling error; measurement error; non-reports; observation uncertainty; and representativeness error
- Spatial verification; Object based verification (e.g., MODE); Neighborhood verification
- Forecast Value metrics

Model Validation Process

Multiple levels of validation occurs during model development:

- **Sanity checks:**

1. Is the code working accurately?
2. Pass the regression tests
3. Bit-wise reproducibility
4. Butterfly tests

- **Performance validation:**

1. Is the code performing accurately and reliably?
2. Cross-platform compliance
3. Exception handling

- **Forecast validation (Hierarchical testing):**

1. Single column models
2. Low-resolution experiments
3. Forecast only experiments
4. Testing of combination of multiple components

Model Evaluation Process

Multiple levels of validation occurs during model development (before transition to operations):

- **Sanity checks:**

1. Is the code working accurately?
2. Pass the regression tests
3. Bit-wise reproducibility
4. Butterfly tests

- **Performance validation:**

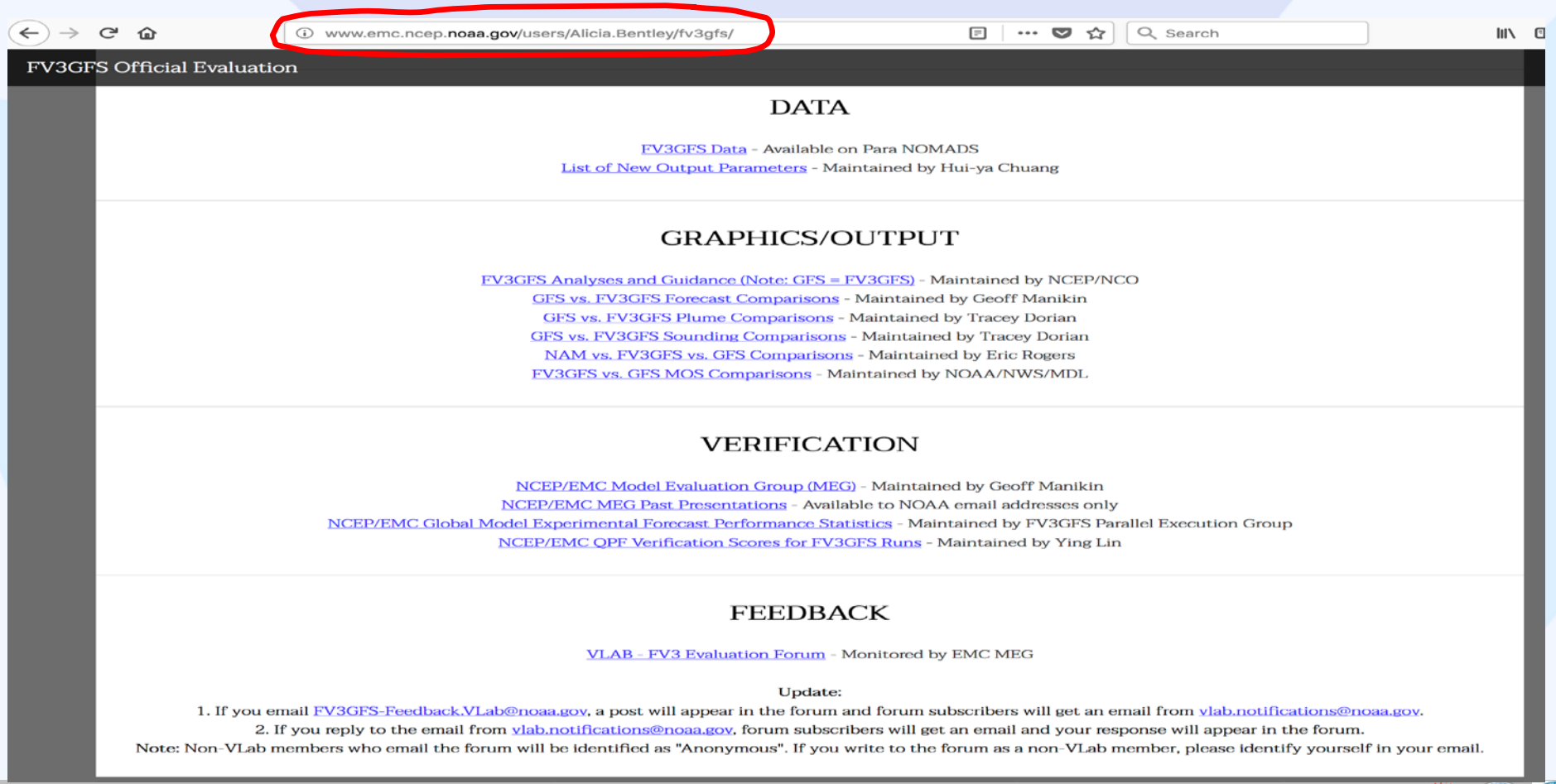
1. Is the code performing accurately and reliably?
2. Cross-platform compliance
3. Exception handling

- **Forecast validation (Hierarchical testing):**

1. Single column models
2. Low-resolution experiments
3. Forecast only experiments
4. Testing of individual and combination of multiple components

- **Pre-implementation T&E:**
- **Focused on forecast skill evaluation**
- **Based on large-scale retrospective and real-time experiments**
- **Includes case studies; subjective evaluation; and objective verification**

FV3GFS Official Evaluation Website



Evidence Based Decision Making

Multiple decisions are made before finalizing a model upgrade package:

- **Decisions are driven by:**

1. Resource availability (will it fit into operations?)
2. Programmatic priorities/sponsored research
3. User requirements/Mission goals
4. Maturity of scientific research
5. Technical compatibility
6. Scope vs. Schedule
7. Accuracy vs. reliability
8. Forecast evaluation at various stages of testing (individual/combined)
9. Downstream dependencies
10. Feedback from stakeholders

- **Hand-offs:**

1. Often researchers hand off codes to EMC for transition to operations
2. Most of the time is spent in re-engineering (operational standards; environmental equivalence)
3. Limited time and resources (human and compute) to validate all scientific aspects
4. T&E need to be repeated if not done within the operational framework

Anticipations from UFS Test Plan and Metrics Workshop

- **We don't want to get Right Answer for Wrong Reason**
- **Decision Support System for Model Development and R2O**

- **Model Diagnostics:**

1. Ensure robust diagnostics that connect various validation metrics with process level evaluation
2. Easy to use tools in support of model development

- **Metrics for Evaluation:**

1. Common set of metrics applicable at each stage of model development
2. Separation of concerns: Diagnostic/Model Development Metrics vs. Forecast Evaluation Metrics

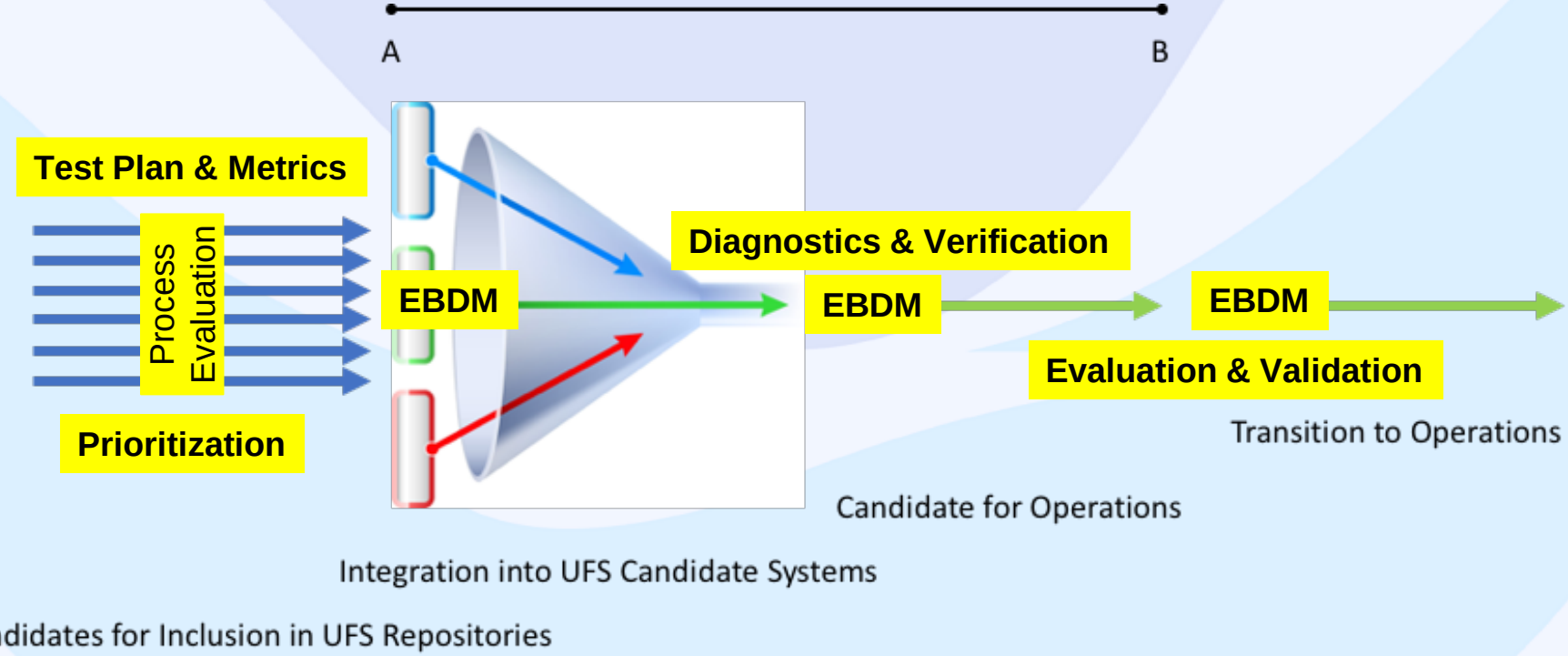
- **Governance for making Informed Decisions:**

1. Appropriate governance for guiding the model development, T&E and verification activities
2. Prioritization of scientific improvements: Physics, DA, Products etc.

- **Development of NWP Index equivalent Metrics:**

1. **Scorecards are not sufficient - need for consolidation of various critical metrics of relevance to aid decision making process at different stages of development**

Anticipation from UFS Test Plan and Metrics Workshop



Develop standard metrics and procedures for test plan, verification and validation to accelerate R2O

Thanks for your attention!

Questions?