

NCEP Metrics for Evaluating Climate Models and Forecasts

May 15, 2015

This document provides guidelines for evaluating the improvements in climate model performance to NCEP operational Climate Forecast System (CFS) and for evaluating the forecast tools (either statistical or dynamical) compared to Climate Prediction Center (CPC) operational forecasts or some other common baselines (such as, persistence). The forecast metrics include deterministic, categorical, and probabilistic measures. The model performance metrics are intended to include measures of model performance in presenting mean climate, variability (i.e., ENSO, MJO), and key physical processes (e.g., convection, fluxes). Researchers or tool developers can select appropriate metrics and variables from this document accordingly.

1. Climate model performance metrics

The metrics are designed for the following four sets of diagnostics during the model development process. The proposed diagnostics includes two aspects (a) characteristics of climate variability in free simulations (both atmosphere alone and coupled), and (b) initialized weather and seasonal predictions. The strategy behind model evaluation would be to develop a hierarchical evaluation procedure with appropriate decision points to better utilize resources.

- **AMIP (~30 year run)** to determine basic characteristics of mean climate and its variability on various time scales
 - Mean biases (u, v, T, q, soil moisture, snow,...)
 - Characteristics of atmospheric modes of variability (PNA, NAO, MJO, ...)
 - ENSO tele-connection
- **CMIP (~50 year run)**
 - Mean biases (u, v, T, q, SST, soil moisture, snow, sea ice,...)
 - Characteristics of atmospheric modes of variability (PNA, NAO, MJO, ...)
 - ENSO variability
 - ENSO tele-connection
- **Short to medium range weather forecast performance evaluations**

The performance of model upgrades in the atmosphere component of CFS (e.g., GFS) short to medium range weather forecasts will be assessed using the Global NWP Model Verification Package developed at NCEP EMC. The evaluation requires two years of retrospective medium-range forecasts. The main metrics include

 - Anomaly correlations, biases, RMSE (u, v, T, P, SLP, q, cloud)
 - Hurricane track and intensity errors
- **Initialized climate prediction evaluations**

The performance of model upgrades in the CFS monthly and seasonal forecast will be assessed based on hindcast of selected period.

 - Skill of SSTs
 - ENSO skill
 - Surface temperature and precipitation skill

2. Climate forecast evaluation metrics

The forecast metrics and protocol are for evaluations after the model development is settled. The metrics include different aspects of verifications including error/accuracy, skill scores, conditional statistics, reliability, and biases for three types of forecasts -- deterministic, categorical, and probabilistic. The metrics are consistent with those used in the CPC real-time verification system and have been used at many places elsewhere, for example, the WMO Standardized Verification System for Long-Range Forecasts (<http://www.bom.gov.au/wmo/lrfvs/>).

- Deterministic/Continuous:
 - Anomaly Correlation (AC)
 - Root Mean Square Error (RMSE)
 - Mean Absolute Error
 - Amplitude
 - Biases

- Categorical:
 - Contingency Table
 - Heidke Skill Score

- Probabilistic:
 - Brier Skill Score (BSS)
 - Rank Probability Skill Score (RPSS)
 - Reliability