608

EVALUATION FOR SPECIFIC NEEDS: THE DTC CONNECTION WITH NOAA TESTBEDS

Edward Tollerud [2,3,4], Tressa Fowler [1,2], Tara Jensen [1,2], Wally Clark [4,5], Eric Gilleland [1,2], Ligia Bernardet [2,6], and Barbara Brown [1,2]

[1] *National Center for Atmospheric Research (NCAR), Boulder, CO*
[2] *Developmental Testbed Center (DTC), Boulder, CO*
[3] *Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO*
[4] *Earth System Research Laboratory (ESRL), Global Systems Division, Boulder*
[5] *Science and Technology Corporation (STC), Boulder, CO*
[6] *Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO*

1. Introduction

NOAA Testbeds and Programs have the responsibility to improve forecasts of extreme and high-impact events: heavy precipitation at the Hydrometeorological Testbed (HMT), severe storms at the Hazardous Weather Testbed (HWT), and hurricane intensity for the Hurricane Forecast Improvement Program (HFIP), for example. Verification of forecasts of these high impact (and often rare) weather phenomena presents a unique array of requirements. To meet these needs, the DTC has participated in development of software packages such as the Model Evaluation Tools (MET), Method for Object-based Diagnostic Evaluation (MODE), and the SpatialVx R-package. These utilities provide a variety of evaluation methods, covering the range of traditional to spatial techniques. MET and MODE in particular have been used extensively in various NOAA testbeds, often in collaborative projects with the DTC, and enhancements to these tools at the DTC have also evolved as they were adapted to meet project needs. We describe several of these collaborations and discuss their relevance and contribution to high-impact weather research at the NOAA testbeds.

2. HMT: Ensemble QPF for Severe Rainfall

High-resolution forecast verification for severe precipitation events (as in the California HMT domain in Fig. 1) presents several assessment challenges that the DTC has been working with the HMT to address. One such technique effectively displays several critical verification scores together on a performance diagram (Fig. 2). The figure succinctly illustrates one of the key motivations of the HMT forecast exercises by demonstrating that finer resolution does in fact lead to better winter forecasts of precipitation in the California domain.

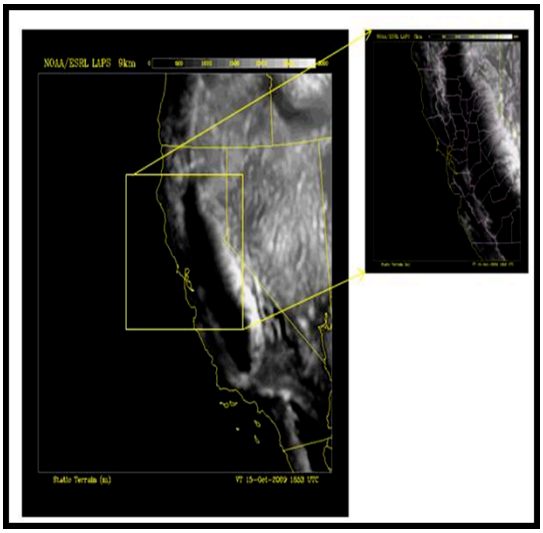

FIG. 1. HMT domains on the west coast of the United States during winter modeling exercises for 2010 and 2011. The inner and outer domains have spatial resolutions of 4 and 9 km, respectively.

The HMT has also made use of (and contributed to development of) MET- and MODE-based methods to aggregate and assess the statistical significance of groupings of ensemble model members with different attributes. On Figs. 3 and 4, for instance, there is strong evidence for a significant positive advantage at early lead times for LAPS hot-started members as compared to those with cold starts for the set of WRF-ARW ensemble members.

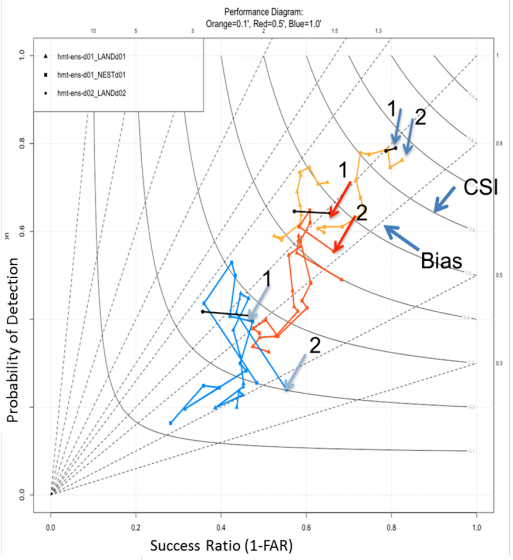

FIG. 2. Performance diagram for ensemble mean QPF with attributes as shown. Gold, red, and blue represent precipitation verification thresholds of 0.1, 0.5, and 1.0 in, respectively, for low-resolution forecasts. Black points and lines represent high-resolution forecasts. Line segment breaks are at 6h lead time increments initiating generally from upper right part of diagram at positions 1 for high-resolution inner domain and positions 2 for the inner domain using low-resolution forecasts.

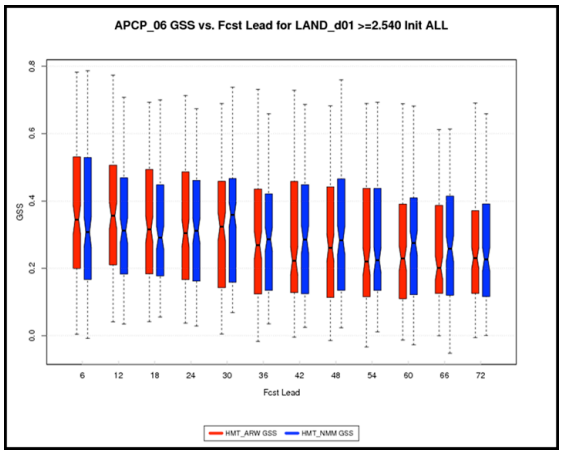

FIG. 3. Gilbert skill score (GSS) distributions at forecast lead times for ARW ensemble members with LAPS hot-start (red) and NMM ensemble members (blue).

3. HMT: Object-Based Verification for Atmospheric Rivers

Accurately predicting winter-season precipitation along the U.S. West Coast requires good forecasts for the relatively narrow streamers of moist air that impinge on the high terrain and produce dangerous levels of rain and snow. Special techniques are required to usefully verify the ability of numerical models to capture these 'atmospheric rivers' (ARs). The spatial verification methods in MODE have been applied in several ways. One is to identify biases in (for instance) GFS model analyses and forecasts. Results of one such test are shown in Fig. 5 for objects produced from fields of integrated water vapor (IWV). The figure suggests a GFS forecast positive bias that increases at higher rain rates but not necessarily at longer forecast lead times.

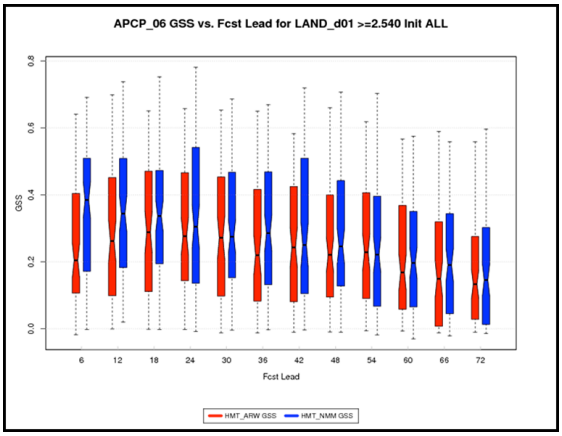

FIG. 4. As in Fig. 3 except for ARW members without LAPS hot-start.

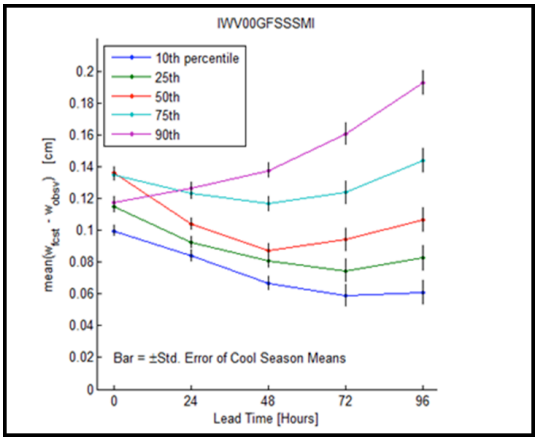

FIG. 5. Average domain IWV bias as a function of percentile value category (colors) and lead time for GFS forecasts verified against SSMI satellite observations.

Another innovative use of MODE involves the definition of objects based on computed fields of IWV transport (IVT; see Fig. 6). Since the timing of the arrival on the coast of ARs and other precipitation-producing systems is critical, MODE IVT objects have been identified in a narrow banded domain along the coast. Diagnoses of these objects have shown a tendency for forecast IVT maxima to be dislocated a marginally significant distance south along the coastline compared to satellite observations.

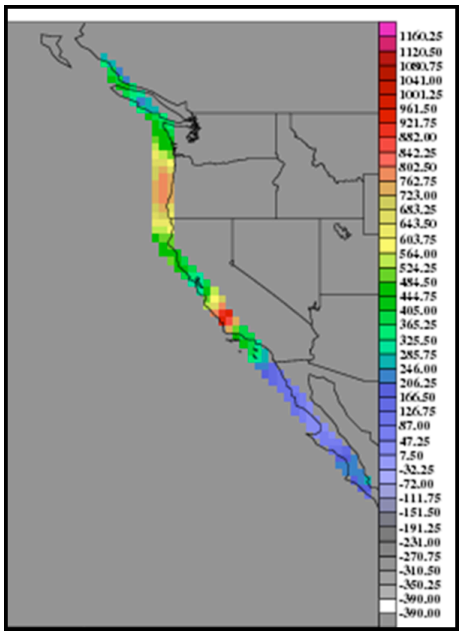

FIG. 6. Coastline domain and example of IVT contours for precipitation events during an HMT field exercise.

4. HWT: MODE-based Verification for Radar Echo Forecasts

Standard verification procedures for severe storm attributes like reflectivity are limited in usefulness by the inherent matching penalties associated with high-resolution forecasts and observation fields. For HWT verification during several Spring Exercises, MODE spatial methods have been employed in novel ways to alleviate these penalties. These methods are illustrated in Fig. 7, notably as applied to radar echo top height probability objects. During the spring exercise illustrated, several models with varying techniques for microphysical forecasts and with different input data assimilation options (including radar reflectivity) were compared in operational and retrospective research environments. Object verification methods were employed using the MODE package in order to provide a non-standard comparison metric that was less critically sensitive to verification idiosyncracies produced by point-to-point comparisons at high-resolutions.

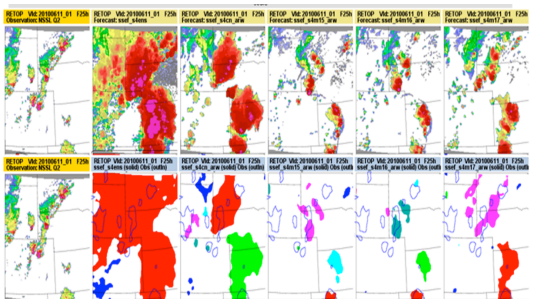

FIG. 7. Screenshot of an 18dBZ radar echo top height and spatial verification display. Plots are 12hr forecasts valid at 8 June 2010 12UTC. Top row, respectively: Q2 observed field; CAPS simple probability field; SREF simple probability field; CAPS probability neighborhood field; NAM deterministic QPF field; and CAPS probability matched QPF field. The bottom row shows forecast (solid) and observed (blue line) objects identified by MODE for 18dBZ echo top height > 25000 ft.. Different colors indicate forecast cluster of objects matched with underlying observation.

5. HFIP: Significance of Verification Scores for Hurricane Intensity

In a comparison with buoy data performed by NOAA/AOML, the National Weather Service Hurricane WRF model (HWRF) was shown to have insufficient surface cooling and a subsequent degradation of intensity forecasts. To determine the causes for this shortcoming, the DTC worked with NOAA/EMC and the University of Rhode Island to formulate a test to determine the effect of adjustments to the momentum flux in the HWRF ocean model. The results shown on Fig. 8 reveal that the modified code reduced the original forecast bias, and on the basis of these results the suggested changes were accepted in the 2013 HWRF model baseline. When verification results like these are used to confirm improvements in model performance, or to make important changes in model routines, statistical assessment of the scores themselves (the error bars on Fig. 8) is critical. In this particular case, the confidence intervals for the two forecasts overlap at most lead times, suggesting only marginal significance at this high level (95%).

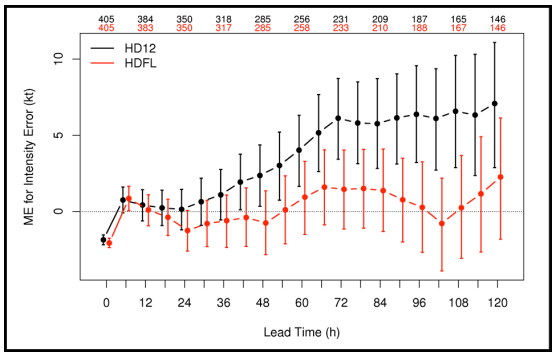

FIG. 8. Mean intensity error (kt) as a function of forecast lead time (h) for all Atlantic storms of 2012. The black curve is the control forecast and the red curve is the forecast with modified fluxes. Vertical bars denote the 95% confidence intervals around the mean.

6. Verification Displays in Community Code

The capability to easily and efficiently compute and display verification products has become increasingly important as numerical models produce ever more types of forecasts at higher spatial and temporal resolutions. Fig. 9 illustrates spatial block bootstrap results at grid locations in the Southeast United States along with the associated confidence interval estimates. Advanced techniques such as this have been implemented in community software available for use by NOAA testbeds and others.

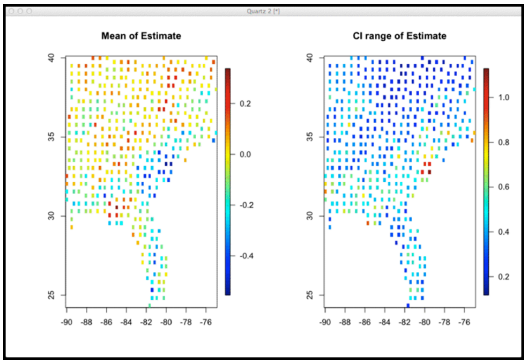

FIG. 9. Regional display of verification scores (left) and their bootstrap-based confidence intervals (right).

7. Summary and Future Development

As illustrated by these examples, MET and MODE continue to evolve to meet the sometimes unconventional verification requirements of NOAA testbeds. These needs are especially acute for severe and extreme phenomena. Of particular current interest are new display techniques in development for the METViewer utility in the MET package, and innovative procedures to estimate and display statistical assessments of verification scores.