P81

WRF Verification for Specific Needs: The DTC Connection with NOAA Testbeds

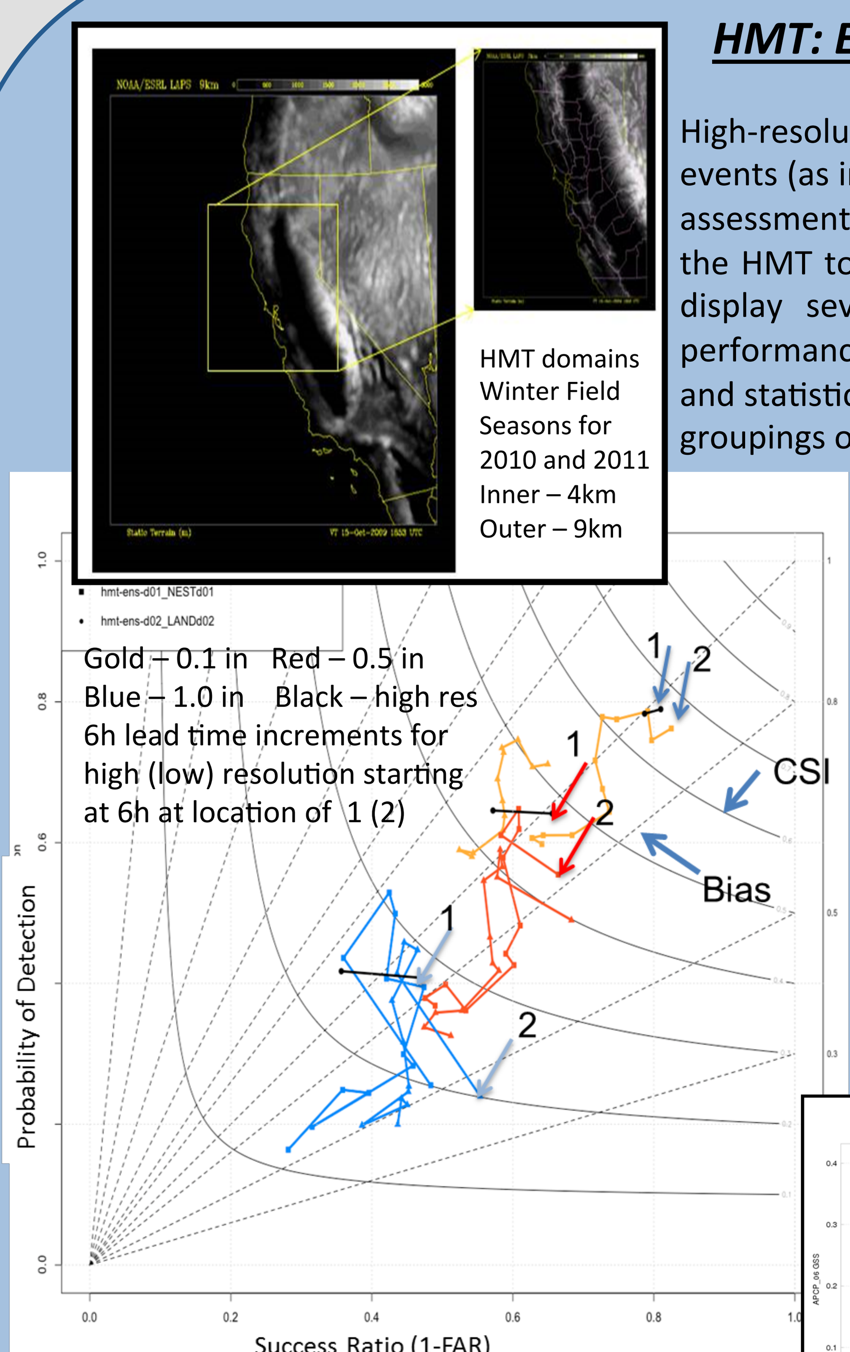
Tressa Fowler[1,2], Edward Tollerud[2,3,4], Tara Jensen[1,2], Eric Gilleland[1,2], Barbara Brown[1,2], Wally Clark[4,5], and Ligia Bernardet[2,6]

[1]National Center for Atmospheric Research (NCAR) [2]Developmental Testbed Center (DTC)
[3]Cooperative Institute for Research in the Atmosphere (CIRA) [4]Earth System Research Laboratory/NOAA (ESRL)
[5]Science and Technology Corporation (STC) [6]Cooperative Institute for Research in Environmental Sciences (CIRES)

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.

HMT: Ensemble QPF for Severe Rainfall

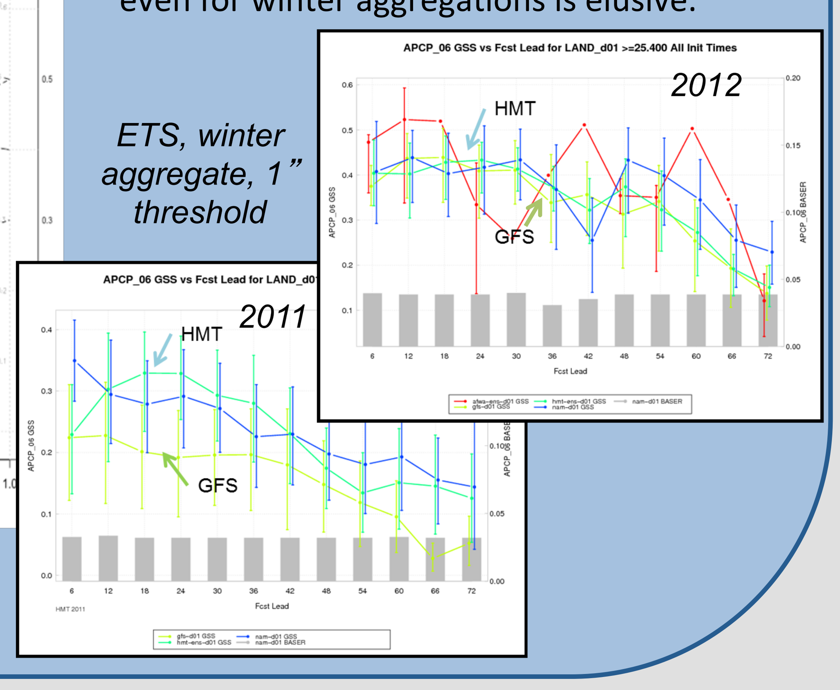


Performance diagram for ensemble mean QPF with attributes shown.

High-resolution forecast verification for severe precipitation events (as in the CA HMT domain to the left) presents several assessment challenges that the DTC has been working with the HMT to address. These include techniques to effectively display several critical verification scores together on a performance diagram as shown below left, and aggregation and statistical significance methods to appropriately compare groupings of different model precipitation forecasts (below).

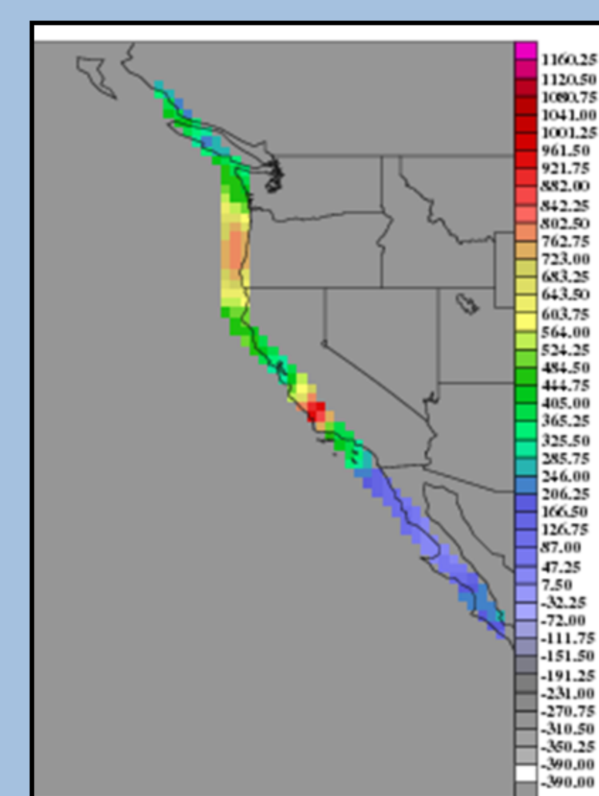
In the former, an advantage for finer resolution forecasts (1 compared to 2) for different precipitation thresholds and lead times is suggested. In the latter, equitable threat scores for heavy precipitation events from 2011 and 2012 demonstrate remarkable improvement between the years, particularly for the GFS. However, statistical significance even for winter aggregations is elusive.

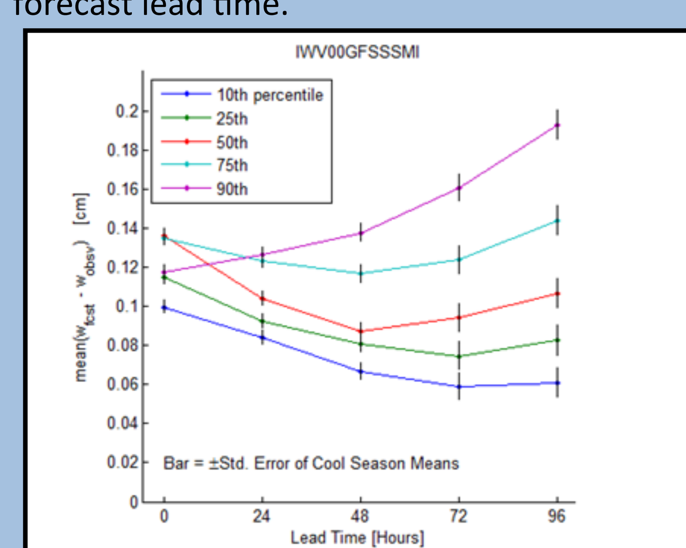
ETS, winter aggregate, 1" threshold



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 below for objects produced from fields of integrated water vapor (IWV). The figure suggests a GFS forecast positive bias that increases with rain rates while showing more complicated behavior with forecast lead time.



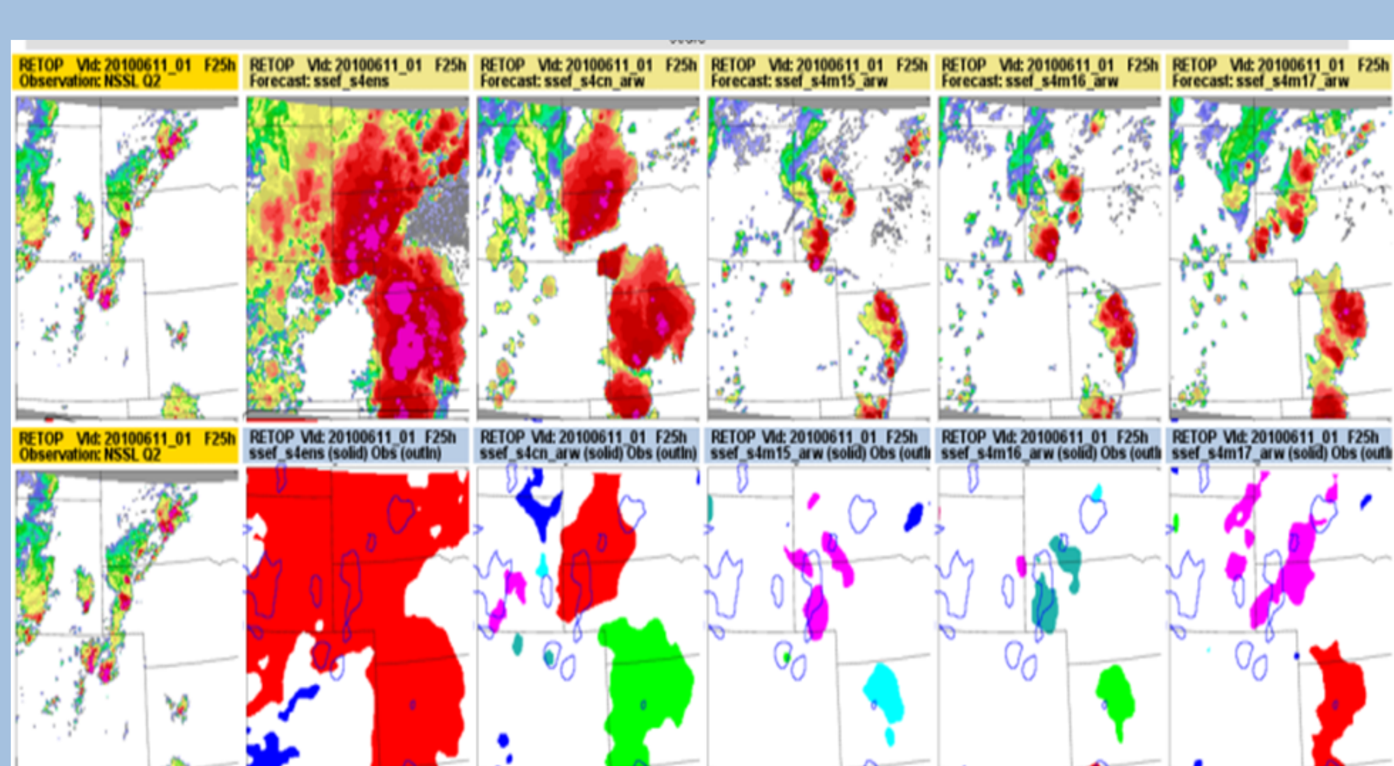


Average domain IWV bias as a function of percentile value category (colors) and lead time for GFS forecasts verified against SSMI satellite obs.

Another innovative use of MODE involves the definition of objects based on computed fields of IWV transport (IVT; above). Since the timing of the arrival of ARs is critical, MODE IVT objects have been identified in a narrow banded domain along the coast. Diagnosis 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.

HWT: MODE-based Verification for Radar Echo Forecasts

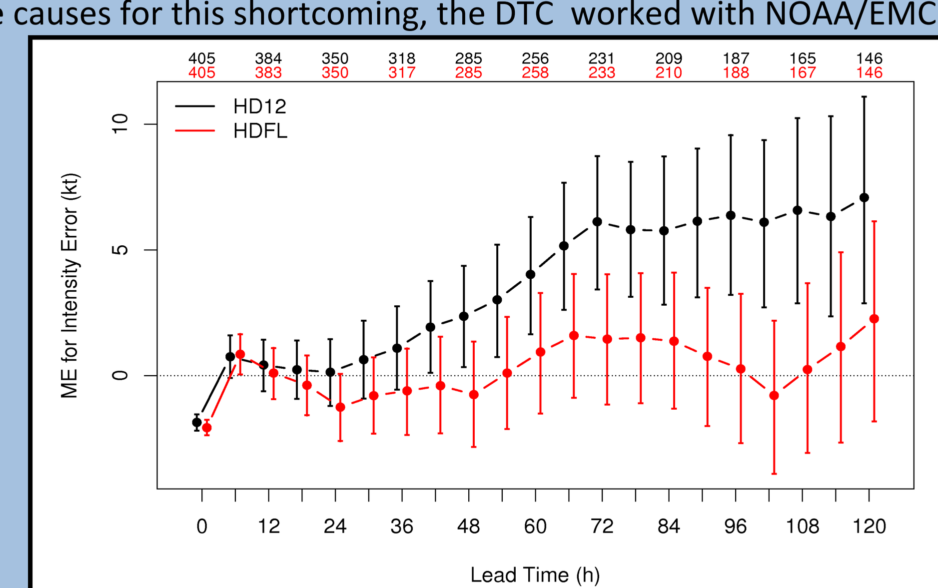
Standard verification procedures for severe storm attributes like reflectivity are limited in usefulness by the inherent matching penalties associated with high-resolution forecasts and observation fields. For HWT verification during several Spring Exercises, MODE spatial methods have been employed in novel ways to alleviate these penalties. The application of these methods are illustrated below, notably as applied to radar echo top height probability objects.



Screenshot of a 18dBZ radar echo top height and spatial verification display. Plots are the 12hr forecast valid at 8 June 2010 12UTC. The top row: 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 objects.

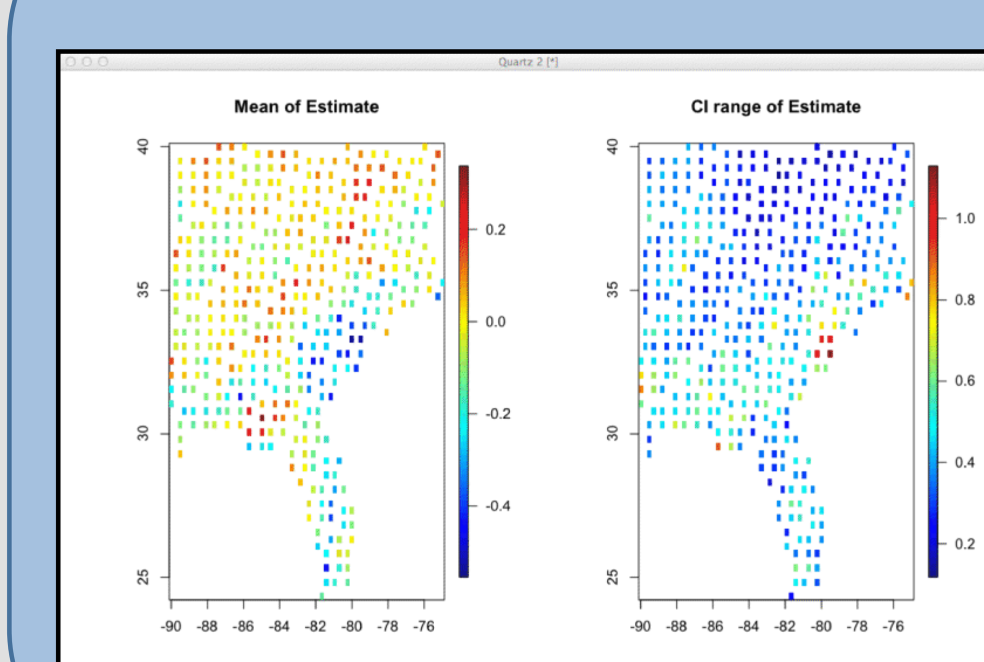
HFIP: Significance of Verification Scores for Hurricane Intensity Forecasts

In a comparison to 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 in the figure (generated using the National Hurricane Center verification system) suggest that the modified code led to an improved forecast bias, and on that basis the change was accepted in the 2013 operational HWRF. When verification results like these are used to confirm improvements in model performance, statistical assessment of the scores (the error bars on the figure) is critical.



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 on the mean. A routine to generate these metrics has been installed in MET-TC, an extension of MET designed for verification of tropical storm forecasts.

New 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. These figures illustrate spatial block bootstrap results at grid locations in the SE US 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.

Summary and Future Development

As illustrated by these examples, MET and MODE continue to evolve to meet the sometimes unconventional verification requirements of numerical modeling projects, including those in several other 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 program, and procedures to estimate and display statistical assessment of verification scores.