Verification Dataset Choices and their Impact on WRF QPF Forecasts for the 2009-2010 HMT Winter Exercise

Edward I. Tollerud[1], Tara Jensen[2], John Halley Gotway[2], Paul Oldenburg[2], Huiling Yuan[3], and Isidora Jankov[4]

[1]*NOAA Earth System Research Laboratory (ESRL), Global Systems Division, Boulder, Colorado*
[2]*Research Application Laboratory, National Center for Atmospheric Research, Boulder, Colorado*
[3]*Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado at Boulder, and NOAA Earth System Research Laboratory (ESRL), Global Systems Division, Boulder, Colorado*
[4]*Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, Colorado USA, and Earth System Research Laboratory (ESRL), Global Systems Division, Boulder, Colorado*

1. Introduction

During the 2009-2010 Hydrometeorological Testbed winter exercise (HMT-West), a real-time website was established to provide up-to-date and retroactive verification statistics for the 9 ensemble members of a high resolution (9 km) WRF modeling system situated over most of California and Nevada. This system allowed multiple scoring options including standard scores (equitable threat; false alarm; RMSE; bias; etc.) for runs at constant initialization time and constant valid times, as well as object based techniques that keyed on quantitative precipitation forecasts. In addition, summary score statistics were routinely displayed for the previous 30 day period to gain a sense of past model performance. One of the innovative features of the system was the opportunity to select from a choice of verification datasets (e.g., Stage IV grids at 6h accumulation periods, and Stage IV and gages at 24h periods) and regions (individual watersheds and the California Nevada River Forecast Center domain). In this paper we present results from this website that reveal some impacts presented by the choice of data. Since baseline GFS model simulations (at approximately 40 km resolution) were also verified, it is possible to compare verification results that proceed purely from resolution differences.

2. The 2009-2010 HMT Winter Exercise

Domains were selected for the winter exercise that included a large domain covering most of California and Nevada and extending several hundred km westward into the Pacific Ocean. Eight ensemble member forecasts were produced in the large domain using both ARW and NNM cores of the WRF model initiated with several randomly-selected GFS ensemble members for boundary conditions. Forecasts were output every three hours up to 5 day lead times. The spatial resolution of this domain was approximately 9 km. An ensemble mean was produced from these members, and a coarser-resolution GFS forecast was included in the verification for base-lining. In addition, forecasts within a smaller nested domain were produced, and another domain with high temporal resolution (1 hr) was produced for shorter duration forecasts. Verification results presented here are for the full domain.

Results shown here are from stormy periods in January 2010. During the week of 17-21 January, in particular, several storms moved onto the northern and central California coast resulting in heavy precipitation in most of the coastal mountains and the Sierra Nevada Mountains.

3. Diurnal Cycling of Verification Scores

The most remarkable feature on Fig. 1 is the clear diurnal pattern to the ETS scores. As the 6h fractional coverage values (shaded bars) suggest, the best scores closely reflect (but slightly lead) the maxima in area precipitation frequency. The likely explanation for this correlation is that forecast verification scores like the ETS are relatively more easily attained under conditions of substantial areal coverage of precipitation, especially for lower thresholds. It is of interest to investigate the nature of this somewhat unexpected diurnal sequence, which persists for several days during the period. Fig. 2 reveals that along the coast north of San Francisco, there is also a very strong diurnal cycle to the mid- and low-level winds, with strongest westerlies (and presumably strongest upslope flow) centered around 0000 UTC and very strong southerlies at 1200 UTC. This pattern is also evident at many other sites in California, particularly in the western half of the state and along the coast. Further analyses are necessary to determine if it is simply the result of chance waves moving on shore or if a true diurnally-driven circulation is in evidence. The conclusion to that question has strong implications for the development of relevant verification strategies.

Another result of Fig. 1 is the generally good performance of the ensemble mean during the full 4+ day period of the forecasts, a performance that is also reflected by the scores for the GFS. It cannot be ruled out that the GFS scores are simply an effect of coarser spatial resolution, a possibility also suggested by the relatively poor performance of the GFS for area-related scores (false alarm rates, for instance, and areal frequency bias).

4. Gages vs. Stage IV Analyses

What impacts can the choice of verification datasets have in a real-life setting? One indication is given by Fig. 3, which demonstrates significant PODY differences that originate solely from the choice of 24h gages vs. that of 6h analysis from the Stage IV product as verification data. Two factors may be relevant to this difference: rainfall during 6h accumulation periods cannot reach given thresholds as easily, reducing sampling and negatively affecting ETS scores; and gages are predominantly located in California as opposed to Nevada whereas Stage IV analyses extend across the full domain (excluding Pacific Ocean grid points of course), resulting in verification in poorly-observed geographic regions.

5. Conclusions and Further Research

The extensive verification results obtained during the winter experiment in California represent a rich source for studies like those briefly introduced here. In addition to dataset options, the real-time and retrospective scores also offer opportunities for comparing verification within different regions and over various meteorological scenarios.

Acknowledgments

We thank Paul Neiman, Seth Gutman, Allan White, Jian-Wen Bao, and Dan Gottas for supplying some of the graphics presented in this paper. The USWRP provided funding for the website development and other research shown here.

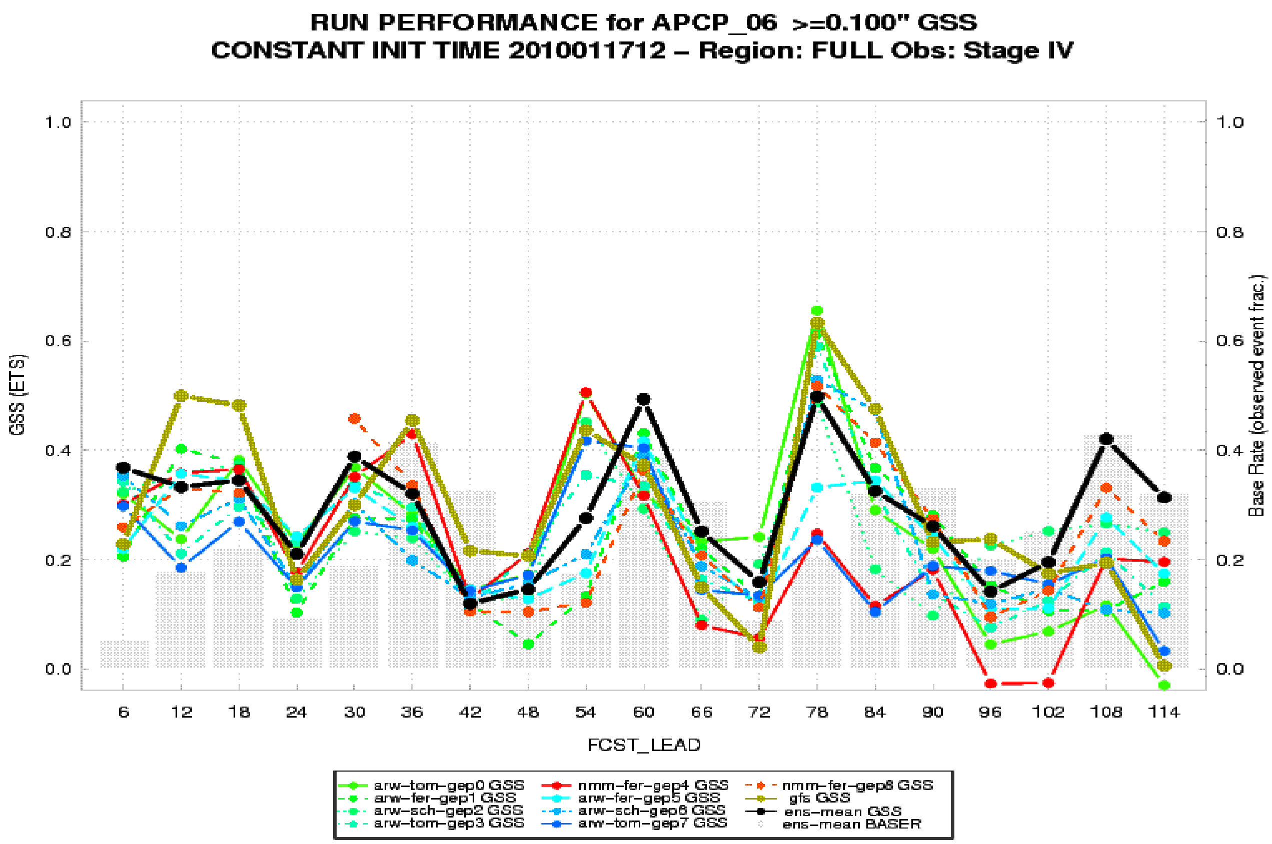

Fig. 1. Equitable Threat Scores (ETS) for ensemble model runs initiated at 1200 UTC 17 January   in the HMT-West domain. Individual ensemble ARW and NNM core members are as shown in legend; black curve is for the ensemble mean; and brown curve is for the deterministic GFS forecast. Lead times are in hours. Verification was performed using Stage IV 6h analyzed precipitation. Shaded bars indicate areal frequency of observed precipitation for each 6h period.

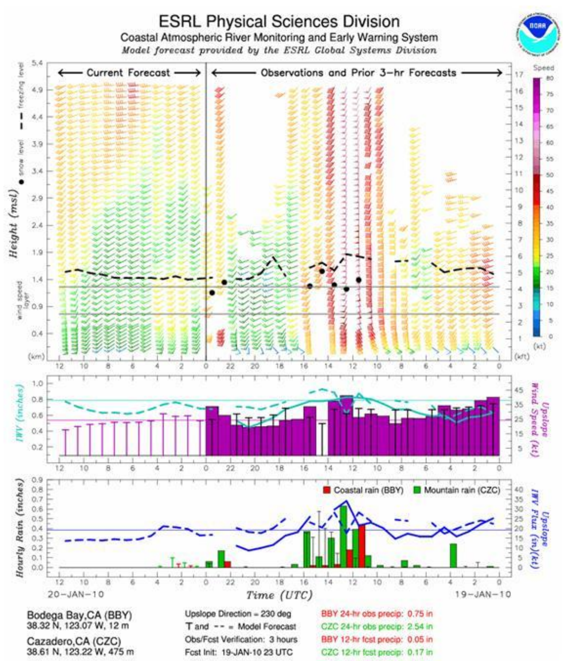

Fig. 2. Time series of precipitation, winds, and other quantities, as shown, for the period 0000 UTC 19 January to 1200 UTC 20 January 2010. Stations BBY and CZC are located on the Pacific coastline and close by in the coastal mountain range, respectively, about 50 km N. of San Francisco. Plots are generated by the Physical Sciences and the Global Systems Divisions of the Earth System Research Laboratory and displayed at http://www.esrl.noaa.gov/psd/data/obs/.

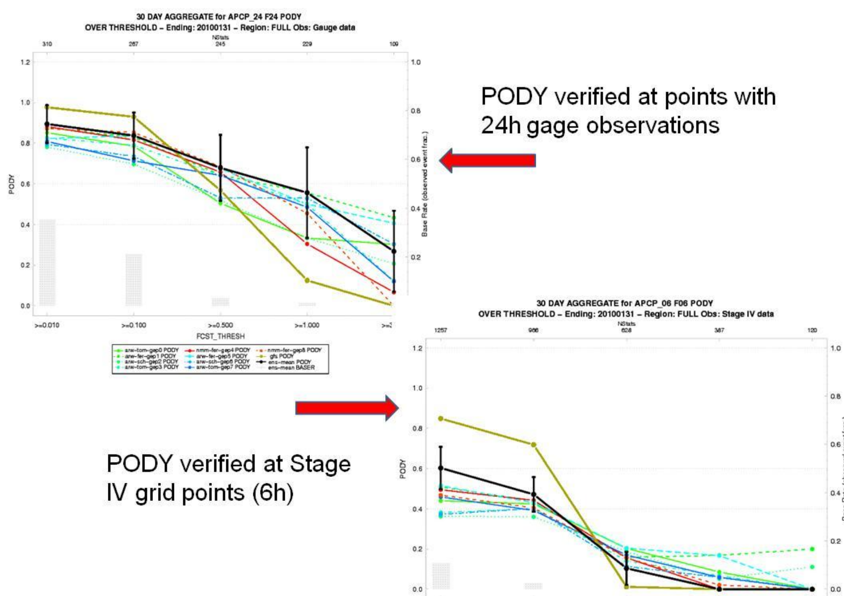

Fig. 3. Comparison of Probability of Detection-yes (PODY) scores for January 2010 in the full HMT-West domain as verified using 24h gage totals (top) and 6h Stage IV estimates (bottom). Designation of individual ensemble members coded by color and line type is as in Fig. 1.