FY 2010 Final Report, 1/13/2011 Edward Tollerud, Project lead

HMT/DTC Collaborative Project to Demonstrate Verification Utilities for the HMT-West and to Provide Objective Assessment of QPF for Operational Models

1. Introduction and Executive Summary

The Hydrometeorology Testbed (HMT) has the goal of accelerating the infusion of new technologies, models, and scientific results from the research community into daily forecasting operations of the National Weather Service (NWS) and it's River Forecast Centers (RFCs). The complementary mission of the DTC is to evaluate and verify the operational model predictions. Thus, the DTC has been selected by the HMT Management Council to explore the potential for the DTC and HMT collaboration in four areas of common interest: forecast verification (principally QPF), logical evaluation for ensemble forecasting, model physics, and data impact studies. To proceed on this objective, a real-time online verification utility based on MET (for Model Evaluation Tools) was established, which produced and displayed QPF verification results during the HMT Winter exercise between December 2009 and March 2010. Subsequent retrospective studies based on these results were undertaken and reported at the Test-bed meeting, the HMT annual meeting, AMS and AGU meetings, and during telecons and internal seminars and colloquiums. A list of some of these presentations is provided in the Appendix. In accordance with DTC and HMT missions, the principal goal has been to evaluate, compare, and thereby enhance precipitation forecasts provided by a research ensemble model and by operational forecast models (initially the GFS) used by EMC and by RFC and NWS forecasters. Funding for these tasks proceeded from the United States Weather Research Project (USWRP).

Principal accomplishments during 2010 include:

1) Installation of a complete demonstration verification workflow and website for HMT-West winter exercise QPF evaluation (http://verif.rap.ucar.edu/eval/hmt/2010);

2) Analysis and description of results from this verification process at tutorials and teleconferences;

3) Contributions to workshops and conferences focused on ensemble verification, on probabilistic forecasting, and on QPF;

4) Preliminary verification of operational GFS forecasts of precipitation and comparison with ensemble QPF.

5) Development of techniques to use spatial verification procedures from the MET/MODE verification package to assess historical forecasts of AR location, intensity, and size.

In the following section we describe several of these developmental accomplishment areas in greater detail. In Section 3 we provide preliminary analytic results for a principal project objective, that of intra-model QPF verification and comparison.

2. Development and Research Areas

a. Verification workflow development and MET/MODE

A key component of the DTC-HMT collaboration is the verification of the HMT ensemble produced by NOAA-GSD. The DTC leveraged the existing workflow manager framework for managing and running numerous automated tasks to define two separate workflows to perform this verification. In addition, the DTC wrote a handful of scripts to monitor the status of the HMT model output and verification results as well as archive the verification output.

The first of these workflows is used to ingest and preprocess several external models used for comparison with the HMT ensemble. These external models include the Global Forecasting System (GFS), North American Model (NAM), Short-Range Ensemble Forecast (SREF), North American Hires Window (NMM-B , and the Hi-Resolution Rapid Refresh (HRRR). The output for each of these models is interpolated to the HMT Ensemble 9-km parent domain and 3-km nest, and the accumulated precipitation is summed up into 6-hourly and 24-hourly accumulations. This workflow is run four times per day to process 00Z, 06Z, 12Z, and 18Z model initializations. The external model output as well as the HMT ensemble output are then verified using the second workflow.

The HMT verification workflow is run two times per day, once to verify the four 6-hour accumulations of precipitation for that day and once to verify the 24-hour accumulations. The models verified include all of the external models listed above, each of the 8 HMT ensemble members, and 2 HMT control runs. The HMT ensemble mean and probabilistic fields are also computed and verified as part of this workflow. The 6-hour and 24-hour accumulations are verified against Stage IV observations while the 24-hour accumulations are also verified against a set of gauges. The verification is performed over the 9-km parent domain as well as the 3-km inner nest. Verification scores are computed over several masking regions, including the land area (LAND), the California-

Nevada River Forecast Center (CNRFC), the area of the inner nest (NEST), and a handful of USGS Hydrologic Units (HUC4 regions).

A separate archival script is run twice daily looking for completed output from the HMT verification workflow. The completed output it finds is automatically transferred from the NOAA Jet machines, where the two workflows are run, over to a workstation residing in the DTC. After being transferred, the verification output is archived to the NCAR mass storage system and also automatically ingested into the METViewer database and display system. At this point, the verification output is available for plotting via the METViewer web interface.

Two other monitoring scripts are run through the crontab batch processor on the NOAA Jet machines. The first is run four times per day to monitor the output of the HMT ensemble runs. When expected model output is missing, an email is sent out to the interested parties listing the missing data files. The second status script is run twice per day to monitor the existence of the Stage IV and 24h accumulated gauge observation files to be used for the verification. Again, when any of the expected observation files are not present, an email to that effect is sent to the interested parties. These types of monitoring scripts provide a convenient way for monitoring the health of the HMT ensemble runs and verification which, in a real-time system, rely very heavily on the availability of the input data.

b. METViewer

The DTC developed a system for storing and visualizing MET verification statistics by integrating database and plotting software. The primary focus of this development was to provide a configurable tool for use across several different DTC testing and evaluation projects including HMT. METViewer was used to generate plots of verification statistics for a number of cases over the time period of the 2009-2010 HMT winter exercise. The cases represented different models, masking regions, accumulation intervals, interpolation types and thresholds. A web interface for accessing METViewer was made available over the world wide web, so that users outside the DTC could access the verification statistics database.

METViewer ingests MET output into a standard SQL database, which facilitates searching and sorting of the verification statistics. Using a user-supplied plot XML specification as input, METViewer gathers and aggregates statistics from the database and generates one or more plots. The three primary types of statistics handled are continuous, categorical and object based. Additional calculations, such as confidence intervals, bootstrap re-sampling and observation frequencies, are performed during the aggregation and plotting process. The primary focus, to date, has been to produce monthly and seasonal aggregations of statistics plotted over lead time, valid date and precipitation threshold, along with vertical profile plots, for a number of atmospheric variables, including: temperature, moisture and winds.

Fig. 1 illustrates one particularly useful HMT example of METViewer capabilities. From an interface (such as the opening screen shown on Fig. 2), uncertainty information indicated by the boxplots for frequency bias can be illustrated for several rainfall thresholds. To a large extent, the plotting options available (scoring metrics; forecast periods; masking regions, verification domains; verification datastream choices; etc.) are much like the original website options which incorporated pre-prepared images.

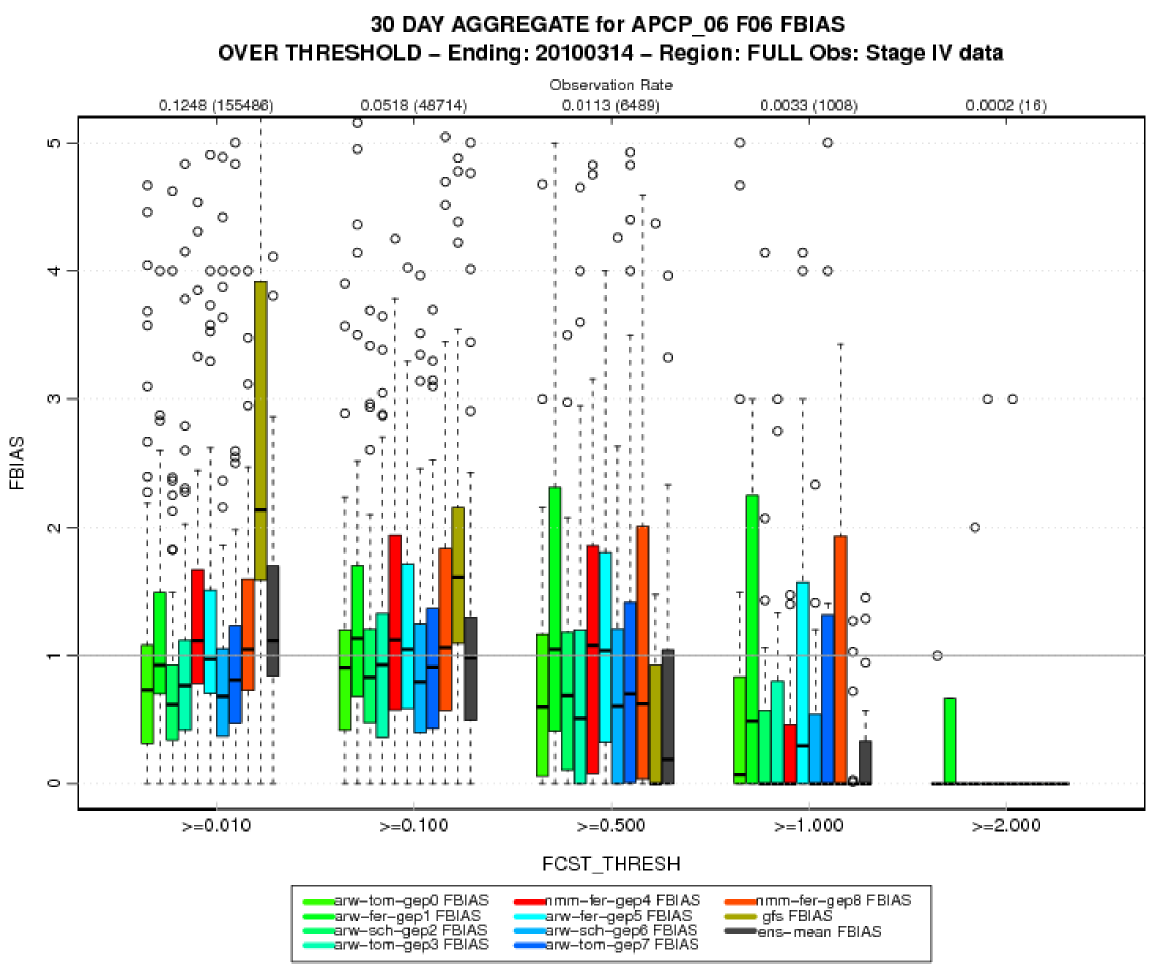

Fig. 1. METViewer example of frequency bias boxplots.

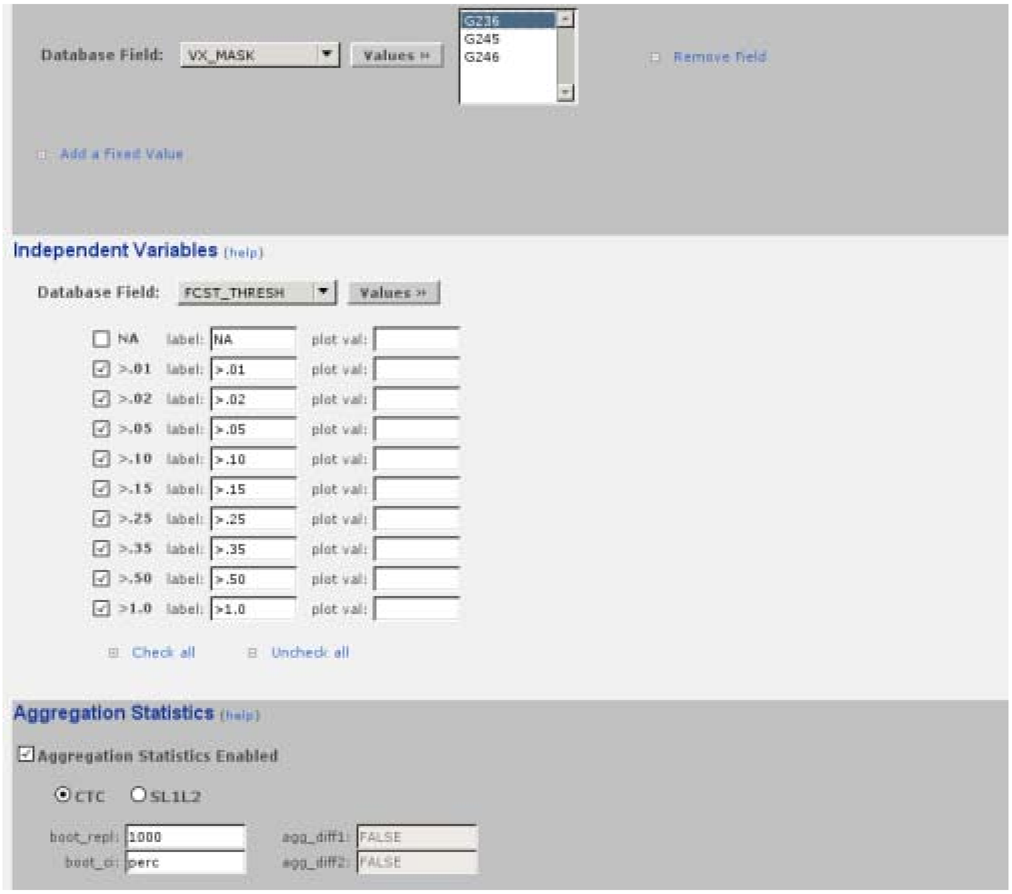

Fig. 2. Example of the METViewer interface screen.

c. Demonstration website development

Part of the goals for the first full winter season of the HMT-DTC objective evaluation was the establishment of a demonstration website to disseminate results of the evaluation. The website achieved full functionality in January 2010. Figure 3 shows the entry-page to the verification plots. Tabs along the top of the page provided access to traditional statistics calculated using MET and provided via pre-generated plots. These graphics could be used to investigate how the nine HMT WRF ensemble members and the ensemble mean performed versus the NCEP Global Forecast System deterministic 0.5 deg model for a given threshold and :

1) Initialization time (plotted against lead time to provide quick look at run performance)
2) Valid time (plotted against lead time to provide quick look at event performance)
3) Over a 30-day window (plotted against valid time)

4) Median value of that 30-day window (plotted against lead time)
5) Median value of that 30-day window (plotted against threshold)

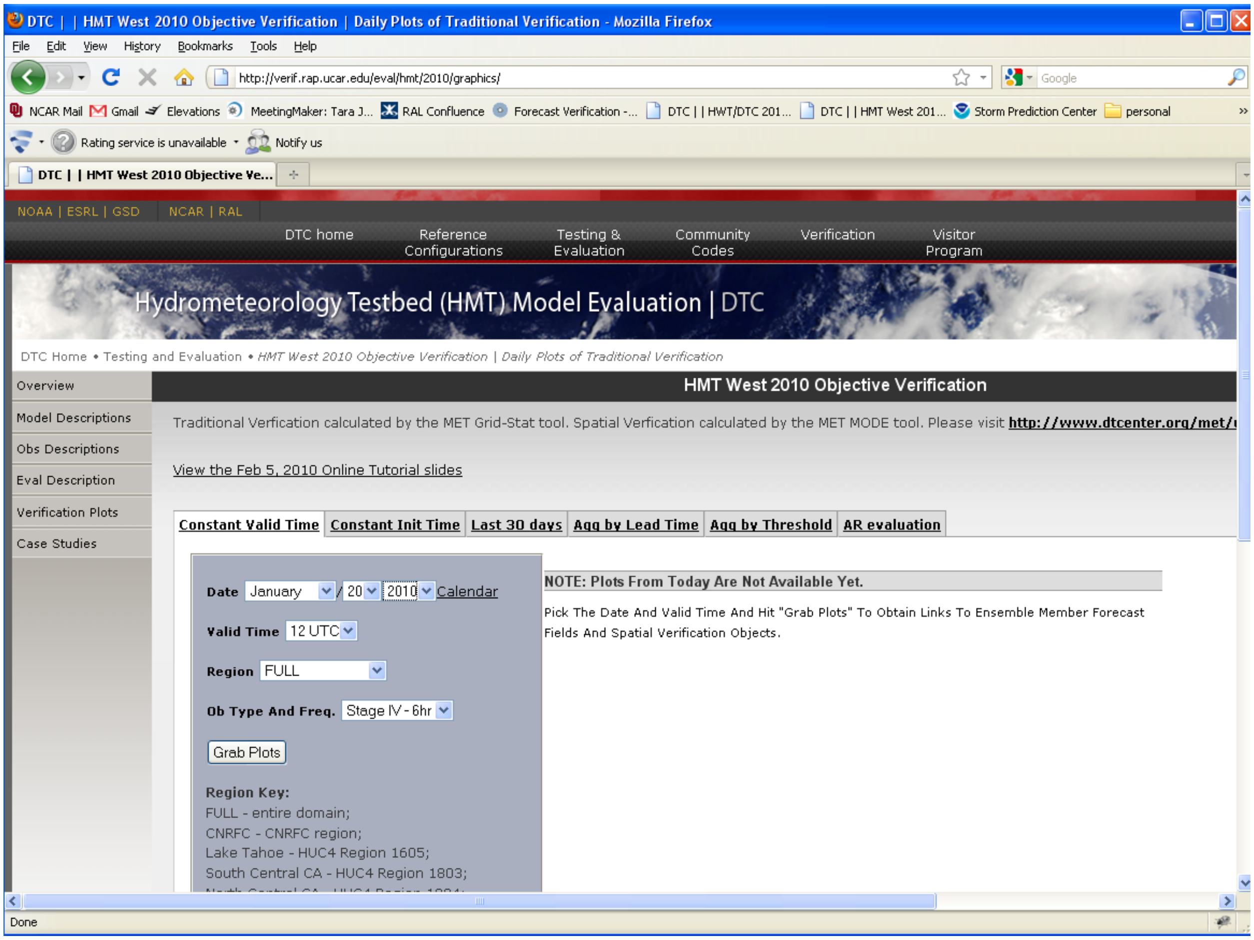

Fig. 3. Screenshot of entry page to DTC verification plots page for HMT-West 2010.

Additionally, MET's object-oriented evaluation tool, called MODE (Method for Object-based Diagnostic Evaluation), was used to identify objects for two thresholds (0.25 and 1.0 inch over 6 and 24 hours). Postage-stamp images of the forecast and observed fields, along with the objects were displayed using an animated web-based tools. Figure 4 shows a snap-shot of that display. Forecaster feedback suggested that these animations were very useful for gaining an understanding of how each member was performing. Additionally, these animations helped identify a problem that was occurring in the post-processing of HMT-WRF ensemble products being disseminated to the NWS Weather Forecast Offices (WFOs) via the Advanced Linux Prototype System (ALPS) boxes.

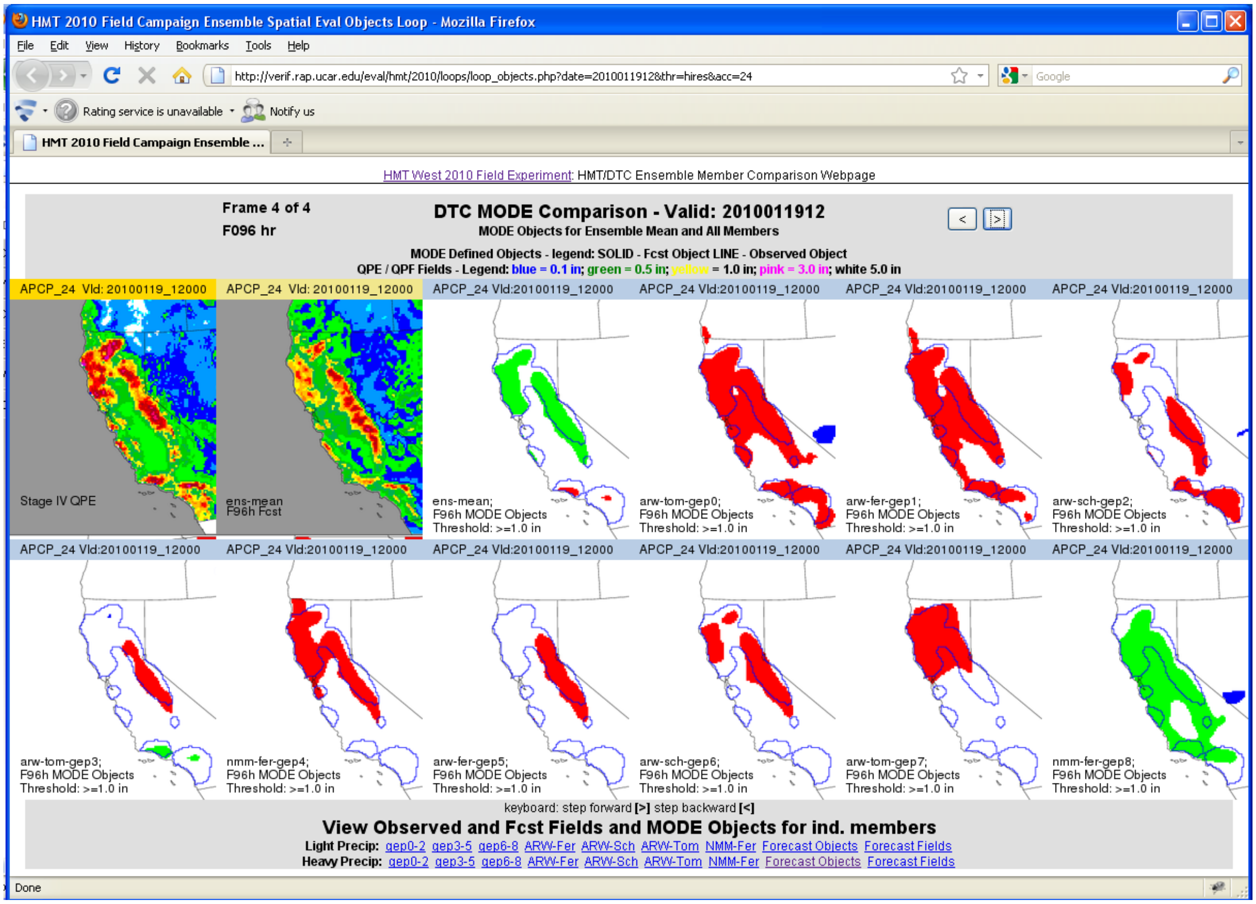

Fig. 4. Screenshot of MODE evaluation page with animation for 96 hour forecast valid at 12 UTC on 19 January 2010. First two postage stamps are the observed 24 hour precipitation and the ensemble mean 24 hour precipitation forecast. Other postage stamps are the forecast objects (solid) and observed objects (blue lines) for ensemble mean, arw-tom-gep0, arw-fer-gep1, arw-sch-gep2 (top row, left to right) and arw-tom-gep3, nmm-fer-gep4, arw-fer-gep5, arw-sch-gep6, arw-tom-gep7, and nmm-fer-gep8 (bottom row, left to right).

The HMT-West 2010 DTC Evaluation website has remained active to the present. Placeholders for Atmospheric River evaluations and Case-Studies were provided on the website. The HMT team is currently in the process of identifying what content should be included on this portion of the website.

A HMT-West DTC 2011 Evaluation website has been developed following the template of the HMT-West 2010 site. To date, all functionality available during 2010 has been implemented for the 2011 season. An expansion of capability to include plots of probabilistic metrics such as Brier Score, Receiver Operator Characteristic (ROC) curves, Area under the ROC Curve, and Spread-Skill diagrams is expected during February 2011.

Also, plots of attributes identified by MODE are also planned for inclusion in the 2011 website.

d. Atmospheric River Analyses

The meteorological precursors for U. S. West Coast AR driven extreme precipitation events are generally detectable well out to sea via satellite observation, allowing initialization of state of the art NWP models that provide warnings several days in advance that a land-falling AR event is imminent. However, traditional verification metrics (skill scores, mean and rmse differences, etc.) indicate that, taken together, the uncertainties in current AR event forecasts of occurrence, magnitude, and location of extreme precipitation along the West Coast are still too large for efficient storm effect mitigation. In detail, this uncertainty leads to significant economic and societal costs due to unnecessary mitigation efforts when events are over estimated. While full time forecasters in this region can and do develop an intuitive feel for the biases and uncertainties in these forecasts, there is a need to supplement this hard gained and subjective intuition with objective metrics that speak directly and quantitatively to these errors. Such metrics, provided in near real time, should prove invaluable to those developing, improving, and troubleshooting models. To this end, one task of the HMT-DTC Collaboration Project has been to explore the use of the DTC MET/MODE object attributes as building blocks for developing such metrics. An initial effort was carried out over the 2009-2010 cool season using real time analyses and forecasts from the GFS model. Fields from the GFS analysis were defined as the observation fields and 24h, 48h, 72h, and 96h forecast fields were compared with this analysis to see how much of the uncertainty in the forecasts is due solely to the model and/or its initialization.

Since a primary ingredient to forecasting AR events is an accurate representation of integrated water vapor IWV, this field was chosen for the initial effort. The analysis was restricted to a region of the North East Pacific beginning about 1000 km offshore to the west and effectively bounded by the West Coast on the east. Objects within this region were built by bounding adjacent pixels in latitude and longitude that represented more than 20 cm IWV, a threshold closely related to that needed to produce an AR extreme precipitation event. The centroid results (Fig. 5), where the centroid is the mean geographic center of the object (which we hypothesize serves as a good metric for estimating the location accuracy for tracking an event), indicates that the uncertainty of the location of IWV objects increases with forecast lead time, that 50 % of the forecast versus analysis centroids differ by more than several 10s of km, and that there seems to be a GFS southerly bias of 10s of km for lead times greater than 24h. The uncertainty of the peak value of IWV within the object, as represented by the 90th percentile of IWV found within the object (Fig. 6), increases with lead time, but does not seem to vary with intensity, except that it appears to be anonymously large between 40 and 55 cm. The uncertainty in the sum of pixel IWV values over the object (not shown) also increased with lead time, but did not increase with amount. In the analysis under way for the 2010-2011 cool season the GFS MODE analysis will be expanded to cover the flux field (IVT) derived from the IWV and wind fields. It is expected that the uncertainties found will increase as the errors inherent in the two fields will be additive.

.

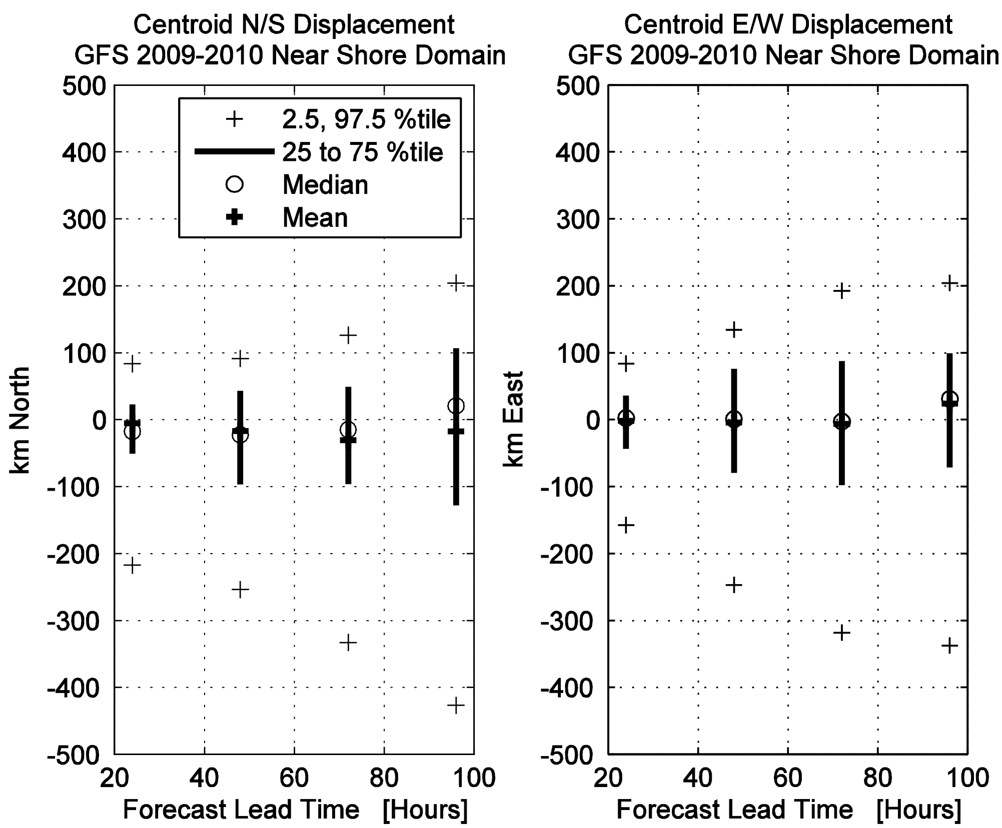

Fig. 5. Illustration of the use of the MET/MODE centroid attribute as a metric closely related to the precision of forecast AR event location. This uncertainty conflates errors in time and shape with errors in location. As expected, the difference in centroid location between the analysis and the forecast objects increases with lead time. There appears to be a southerly bias indicated by the median bars in the left panel. Note that 50 % of the differences are larger in magnitude than the tips of the vertical black bars.

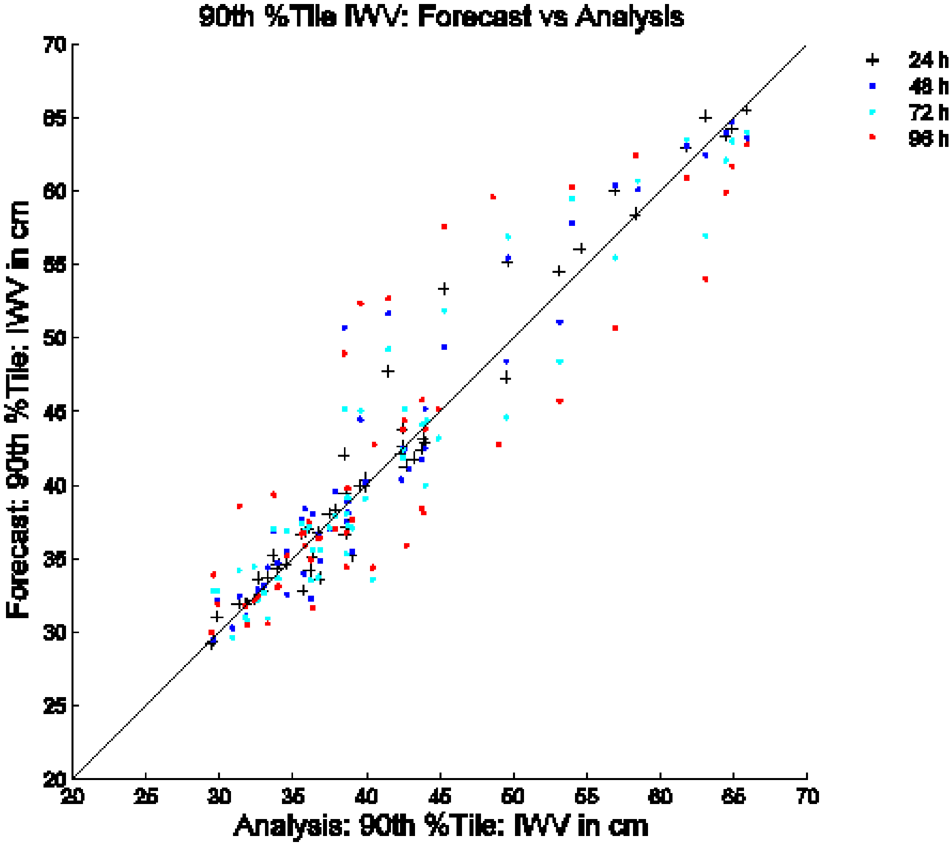

Fig. 6. The peak (i.e., 90th percentile) object values for the forecast objects are plotted against the peak observed (analysis) object values. The difference between the peak values appears to increase with forecast lead time. The reason for the abrupt increase in uncertainty over the band between 40 and 55 cm is not understood.

3. Winter Exercise Results

As previously emphasized, the principal DTC mission that this collaborative effort has thus far addressed is the verification of QPF for modeling systems that (1) are either operational at EMC or are potential operational research projects, or (2) are important and present or potential guidance products used by forecasters in the HMT-West region. During 2010, based on cases processed by the verification workflow and website display utilities, a start has been made on this important task by explicitly comparing the WRF HMT-West ensemble members and their mean with a important guidance model historically used by California forecasters, the GFS deterministic model. Fig. 7 is a useful summary image to summarize some preliminary conclusions; several more aspects of this comparison are included in the presentations and documents listed below. As the difference boxplots and the ensemble mean vs. GFS boxplot comparison in the figure suggest, the performance of the ensemble (as measured by the equitable threat score, for which values from worst to best run between zero and one) has distinct advantages at the extreme precipitation threshold (2.0 inches) presented in the figure. This finding has been borne out for other verification metrics and aggregations as well, and suggests that the improved spatial resolution of the WRF ensemble members and/or the improvement due to ensemble averaging have a telling positive impact on forecast quality. In 2011, several other EMC operational models which have potentially greater forecast value performance than the GFS will be added to the comparison.

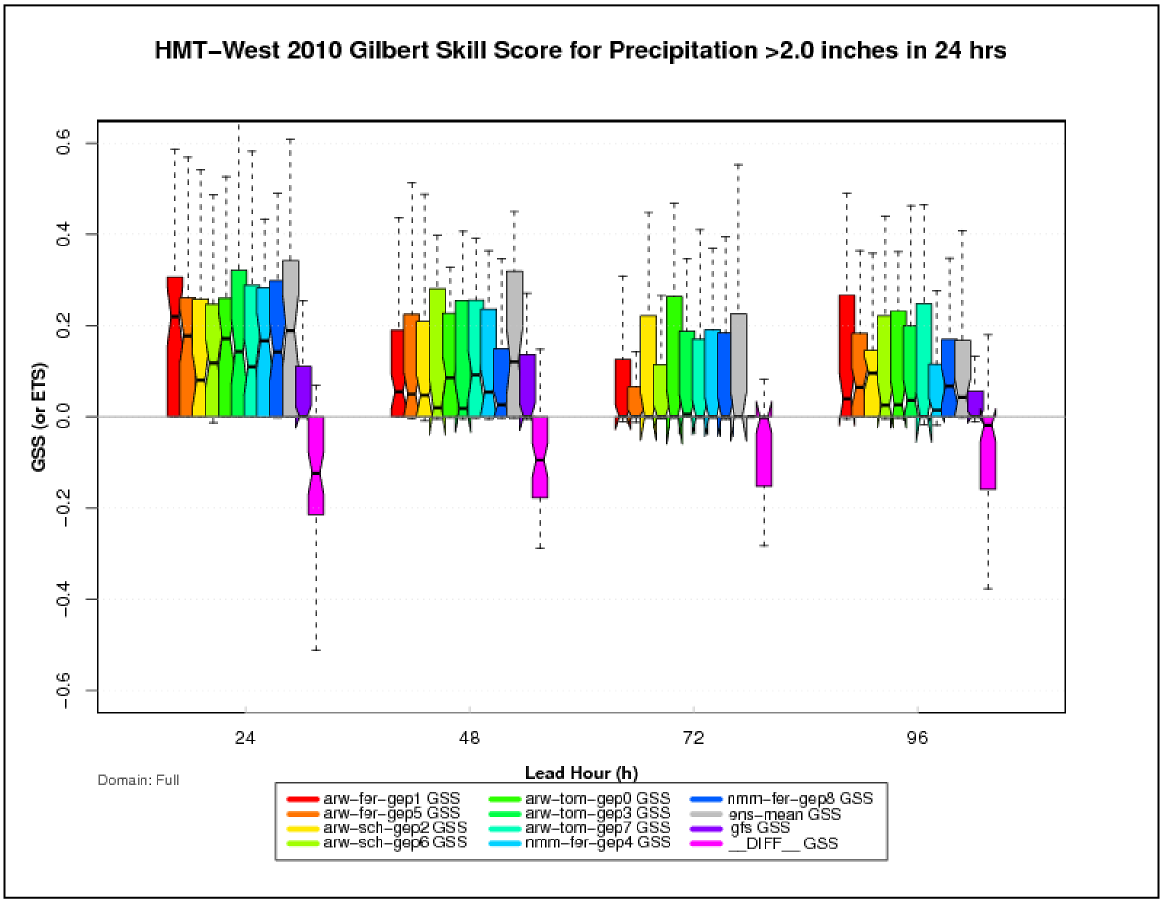

Fig. 7. Equitable threat score distributions (represented as boxplots of inter-quartile range, medians, extrema, and confidence interval estimates) at 1-4 day lead times. Left to right groupings indicated in the legend are for 9 HMT-West ensemble members, ensemble mean, GFS, and difference between the GFS and the ensemble mean. Statistics are aggregated over the 2009-10 exercise (December-March) for 24h precipitation totals greater than 2.0 inches, and verification data are gridded Stage IV precipitation analyses.

Another important concern for QPF verification is the impact of verification dataset choice on inherent uncertainty of verification metrics. For the HMT winter exercises, several choices have been examined: verification against gridded Stage IV precipitation analyses at 6h and 24h accumulation periods, and point gage measurements at a 24h accumulation period. Results for one such comparison for January 2010 (which was strongly influenced by a particularly intense one-week period of severe and widespread rainfall) are described in the report included in the Appendix. For this case, at least, the scoring uncertainty apparently introduced by the choice of verification products is significant and persistent.

4. Reports, Publications, etc.

The following is a list of presentations and reports that have been produced or planned as part of the activities of this project during 2010. Some of these reports and other presentations that present more details of for this report can be directly accessed at the 'Documents' link on http://verif.rap.ucar.edu/eval/hmt/2010.

2011 AMS Weather and Forecasting Conference, Seattle, WA: Comparison of traditional and neighborhood verification statistics for the Hydrometeorology Testbed Forecast Demonstration Project using the Model Evaluation Tools (MET). John Halley Gotway, Tara Jensen, Ed Tollerud , Paul Oldenburg, Huiling Yuan, Isidora Jankov.

2010 WRF-Users Workshop, Boulder, CO: 2010 Hydrometeorology Testbed (HMT) Forecast Demonstration Project,:Verification Using the Model Evaluation Tools (MET). John Halley Gotway, Tara Jensen, Ed Tollerud, Paul Oldenburg, Huiling Yuan, Isidora Jankov.

2010 BACIMO Conference, Omaha, NE: Application of the Model Evaluation Tools Verification Package to the Hydrometeorology Testbed Forecast Demonstration Project. John Halley Gotway, Tara Jensen, Ed Tollerud, Paul Oldenburg, Huiling Yuan, Isidora Jankov.

2nd Testbed Workshop, May 4-5, 2010, NOAA, Boulder, CO: MODE object analyses of integrated water vapor and integrated vapor transport fields. Wallace Clark, Ed Tollerud, Huiling Yan, Gary Wick, Ellen Sukovich, Tara Jensen, John Halley-Gotway, Randy Bullock, and Paul Oldenburg

AMS 2011 Annual Meeting: Evaluation of GFS water vapor forecast errors during the 2009-2010 West Coast cool season using the MET/MODE object analyses package. Wallace L. Clark, H. Yuan, T. L. Jensen, G. Wick, E. I. Tollerud, R. G. Bullock, and E. Sukovich.

AMS NWP Conference, Seattle, WA. The Performance of High-resolution WRF Ensemble QPF Forecasts in California. Edward tollerud, Tara Jensen, Isidora Jankov, John Halley Gotway, and Paul Oldenburg.

2nd Testbed Workshop, Boulder CO: Evaluation of QPF during the HMT-West Winter Exercise: A DTC/HMT Collaboration with USWRP. Edward Tollerud, Tara Jensen, John Halley Gotway, Huiling Yuan, Wally Clark, Ellen Sukovich, Paul Oldenburg, Randy Bullock, Gary Wick.

WRF Users Conf., Boulder CO: Verification Dataset Choices and their Impact on WRF QPF Forecasts for the 2009-2010 HMT Winter Exercise. Edward Tollerud, Tara Jensen, John Halley Gotway, Paul Oldenburg, Huiling Yuan, Isidora Jankov.

3rd International Meeting on QPF and QPE, Nanjing, China: Scale-related Issues Involving Verification of QPF: Results from a DTC Assessment of WRF Ensemble Forecasts during the HMT-West Winter. Edward Tollerud, Tara Jensen, John Halley Gotway, Paul Oldenburg, Huiling Yuan, Isidora Jankov.

5. Project Budget and Personnel

Funding from USWRP to the DTC to perform these tasks in FY 2010 totaled $300K, with $100K to GSD personnel and $200K to RAL.

Personnel involved in the various aspect of this project are:

GSD: Ed Tollerud, Huiling Yuan, Stanislav Stoytchev

UCAR: Tara Jensen, John Halley Gotway, Paul Oldenburg, Randy Bullock,

Tressa Fowler, Barb Brown, Louisa Nance

PSD: Wally Clark, Ellen Sukovich, Gary Wick

6. Future Directions

As the third funding cycle for this project begins, a new research focus is in the process of development. QPF verification and validation remains a high priority, and during 2011 we will continue to build on the demonstration utilities and verification workflow infrastructure that have been the major focus of the project. In addition, a new research area has been identified that will begin to address the last two of the major project objectives, data impacts and physical parameterization assessment. The latter will initially concentrate on microphysics parameterizations. This work has been implicitly involved in the HMT effort for several years by virtue of the definitions of ensemble membership, which has included different microphysical packages. However, this new focus will focus more directly on this research area by involving the several sets of observatory and other observational assets of the HMT-West experiments as verification data streams. A project plan for testing and evaluation is in the process of development, which will include personnel with observational expertise from PSD and model parameterization developers from EMC and elsewhere. The HMT/DTC project involvement will first emphasize standard verification of the model versions that will be evaluated to baseline performance. Subsequently, the specific evaluation of forecasts of microphysical properties will require new options for verification that can match the representations of the new research observations, many of which are not griddd data sets or standard observation quantities. Some of capabilities are currently available in MET and some are still in development. It is hoped that a start on specification task for these new focal areas can occur in 2011, but will be a major emphasis in following years.

Appendix

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.

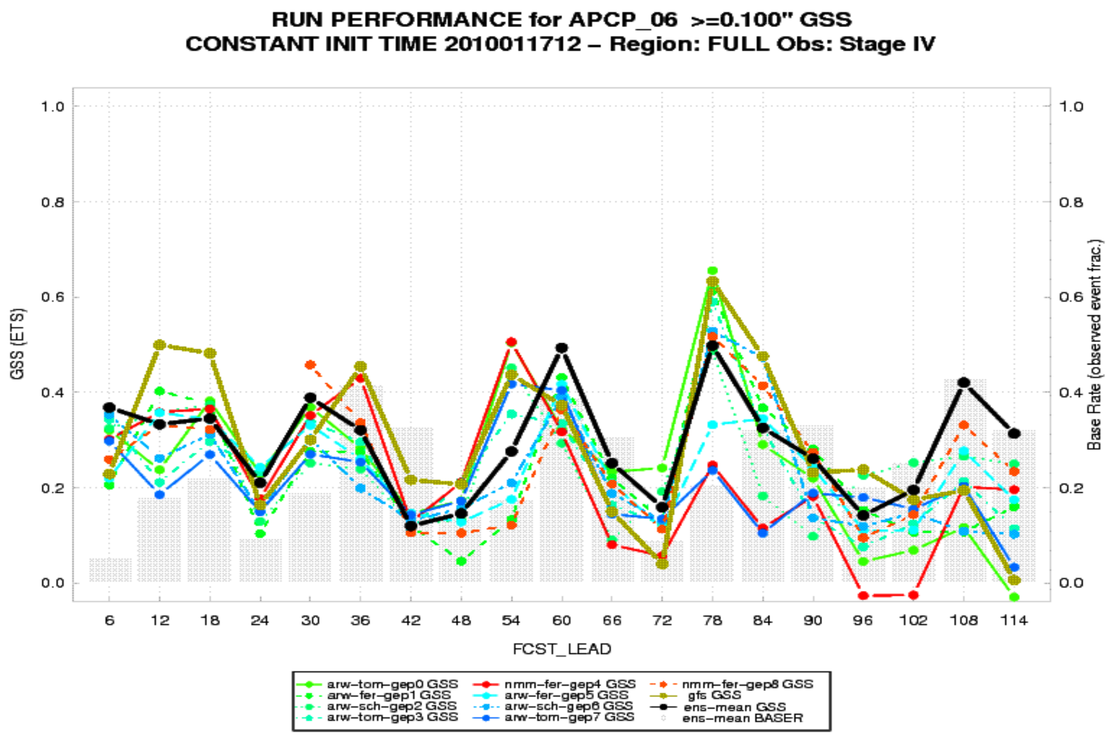

Fig. 1. Equitable Threat Scores (ETS) for ensemble model runs initiated at 1200 UTC 17 January in the HMT-West domain. Individual ensemble ARW and NNM core members are as shown in legend; black curve is for the ensemble mean; and brown curve is for the deterministic GFS forecast. Lead times are in hours.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

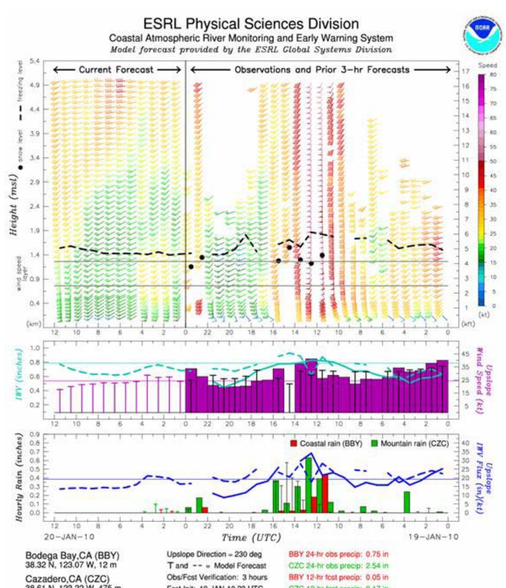

Fig. 2. Time series of precipitation, winds, and other quantities, as shown, for the period 0000 UTC 19 January to 1200 UTC 20 January 2010. Stations BBY and CZC are located on the Pacific coastline and close by in the coastal mountain range, respectively, about 50 km N. of San Francisco. Plots are generated 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/.

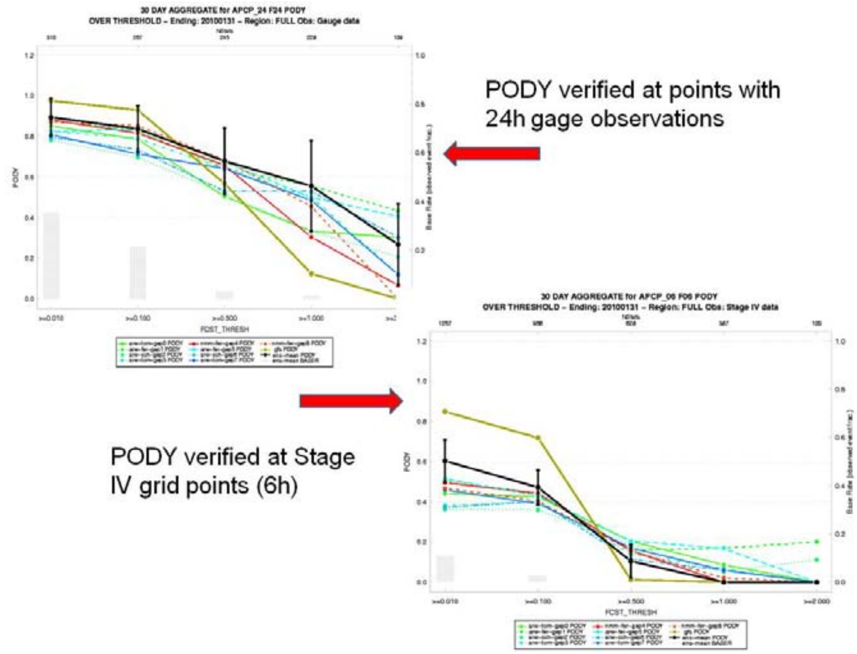

Fig. 3. Comparison of Probability of Detection-yes (PODY) scores for January 2010 in the full HMT-West domain as verified using 24h gage totals (top) and 6h Stage IV estimates (bottom). Designation of individual ensemble members coded by color and line type is as in Fig. 1.