Western Region Technical Attachment
NO. 05-02
May 4, 2005

Notes on The National Digital Forecast Database Verification

Mark Mollner
Scientific Services Division, WRH, Salt Lake City, UT

I. Background

For a little over a year, National Weather Service Headquarter's Meteorological Development Laboratory (MDL) has produced numerical and graphical verification numbers on the 5 KM National Digital Forecast Database (NDFD) (Dagostaro et al, 2004). During this time, a number of questions have risen concerning the pertinence of these numbers and charts in accurately representing the performance of National Weather Service (NWS) Weather Forecast Office (WFO) forecasts – not only in relation to the ambient weather, but also in relation to Model Output Statistics (MOS). What follows is an examination of this verification process and a suggestion or two on how these verification numbers might be used to improve WFO forecast performance.

NDFD verification performs two types of verification: point and gridded verification. Point verification is performed by verifying MOS forecasts at approximately 1300 MOS sites in the continental United States (CONUS) using the METAR observations at the respective sites. The NDFD point verification is done by verifying the forecast at one of the four adjacent grid points from the MOS/METAR site which must be within 500 feet of elevation of the MOS/METAR site. Text verification numbers and plotted maps of monthly Mean Absolute Errors (MAE) and biases are produced for MOS and the NDFD forecast. In addition, comparison graphs of the NDFD forecast and MOS performance is presented in 12 hour intervals out to 168 hours. Elements verified are temperature, dew point temperature, wind speed, max/min temperature, and 12h probability of precipitation (POP). The 00Z and 12Z WFO forecast releases are verified each day.

Gridded verification is performed at all 5 KM NDFD grid points in the CONUS for temperature, dew point temperature, and wind speed. This is a grid-to-gird comparison of WFO 5 KM gridded forecasts to a 20 KM RUC analysis interpolated to the 5 KM NDFD grid. Forecasts at 12 hour intervals out to 168 hours are verified from the 00z and 12Z WFO forecast release times only.

The above point and gridded verification explanations provide an overview of the two methods, and for the sake of brevity, do not cover all the idiosyncracies of the two methods.

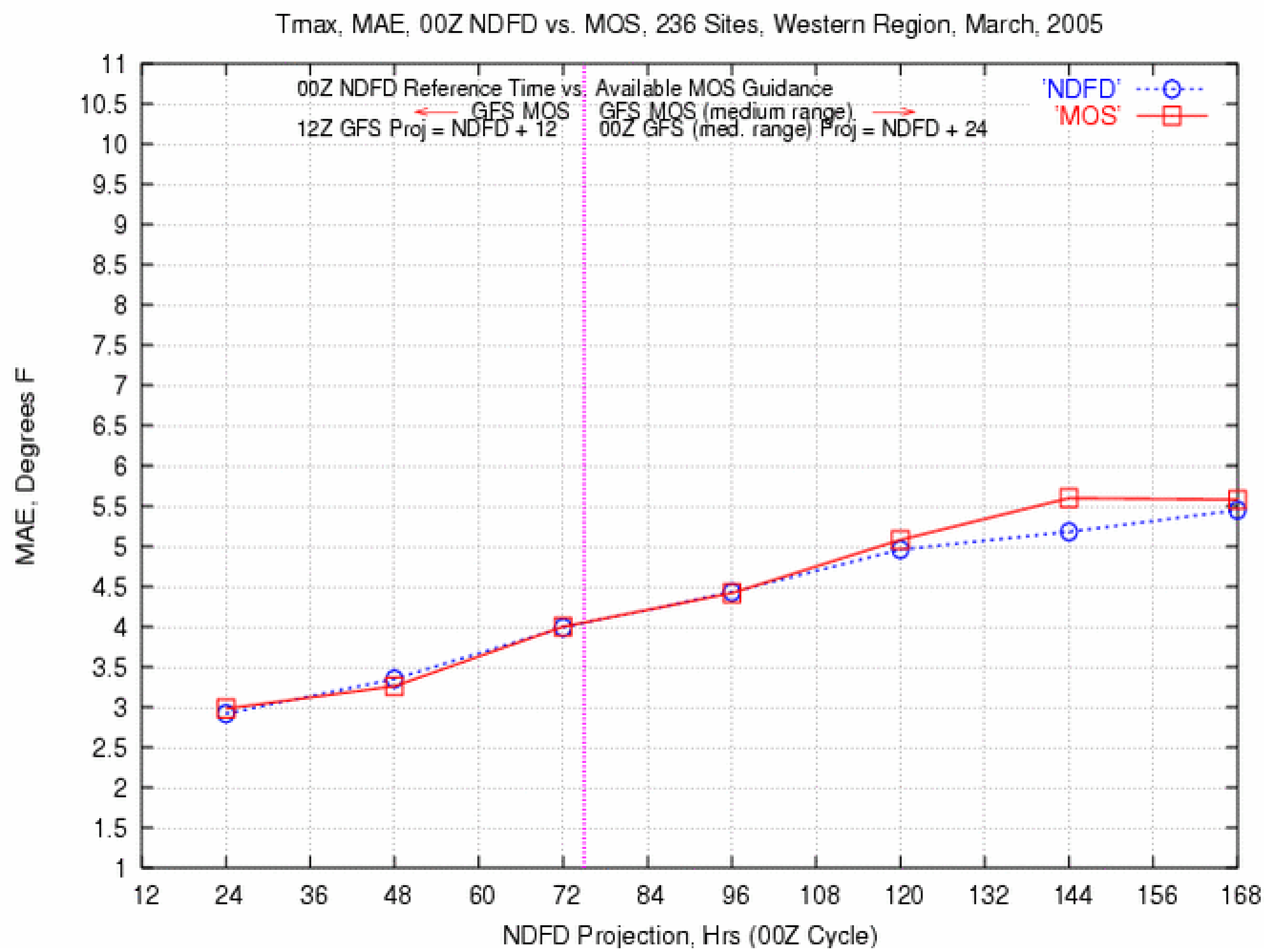
Although as advertised the above two methods do provide quick basic information regarding the skill and accuracy of NDFD forecasts, there are a number of shortcomings which all should be aware of when presenting the results to our forecasters and customers. This discussion will concentrate on temperature, dew point temperature, POP12, and wind speed forecasts.

The NDFD point forecast verification scheme clearly gives the advantage to MOS and a few disadvantages to the forecast staff. MOS equations are tailored to each of the approximately 1300 MOS sites and use years of data to produce its regression equations, and to some extent, account for the effects of local topography. On the other hand, the WFO forecast comes from a grid point near the MOS point, which could be as much as 7 miles away and up to 500 feet in elevation change. In areas of complex terrain, a displacement of a few to several miles may produce differences in upslope/downslope wind regimes, ground cover, and aspect compared to the MOS site. MOS has statistical familiarity with its forecast site while the WFO forecaster has varying familiarity with the NDFD grid point, and in some instances, may not be issuing a point forecast ala MOS, but is issuing a grid box average forecast. The verifying observation for both MOS and neighboring NDFD grid point is the METAR observation at the MOS site. Hence a comparison between MOS and the WFO forecast staff under this current verification scheme seems less than judicious.

The NDFD 5 KM grid-to-grid verification uses the 20 KM RUC analysis interpolated horizontally to the NDFD 5 KM grid for its verifying surface observations. So concerns here range from the coarseness of the RUC analysis to properly define the terrain especially in complex areas, to accurately deriving surface fields, to its ability to spread data from data intense areas to data sparse areas. Although efforts have been made to use a background model lapse rate to better match surface observations with the RUC background when elevations are different, some model error still exists - the RMSE for temperature between the RUC surface analysis and hourly ASOS observations across the U.S. is around 2.7 degrees F (Benjamin et al, 2004). This error is highlighted more in the western U.S. when the RUC terrain field is examined and it shows the elevation at Salt Lake City is 1445 feet lower than reality, 1142 feet lower at Medford, and 835 feet lower than reality at Boise (NWS TPB, 2002). Myrick and Horel (2005) have observed in the winter that the RUC spreads valley cold air too far into the surrounding higher terrain. Other issues here involve WFO smart tools that use a diurnal curve formula to forecast hourly temperatures; the total number of surface observations used to define the RUC analysis; and only half of the WFO forecast is verified at 65% of Western Region WFOs, i.e. fifteen Western Region WFOs produce gridded forecasts at 2.5 KM resolution whereas the NDFD verification verifies a 5 KM grid.

II. Temperature

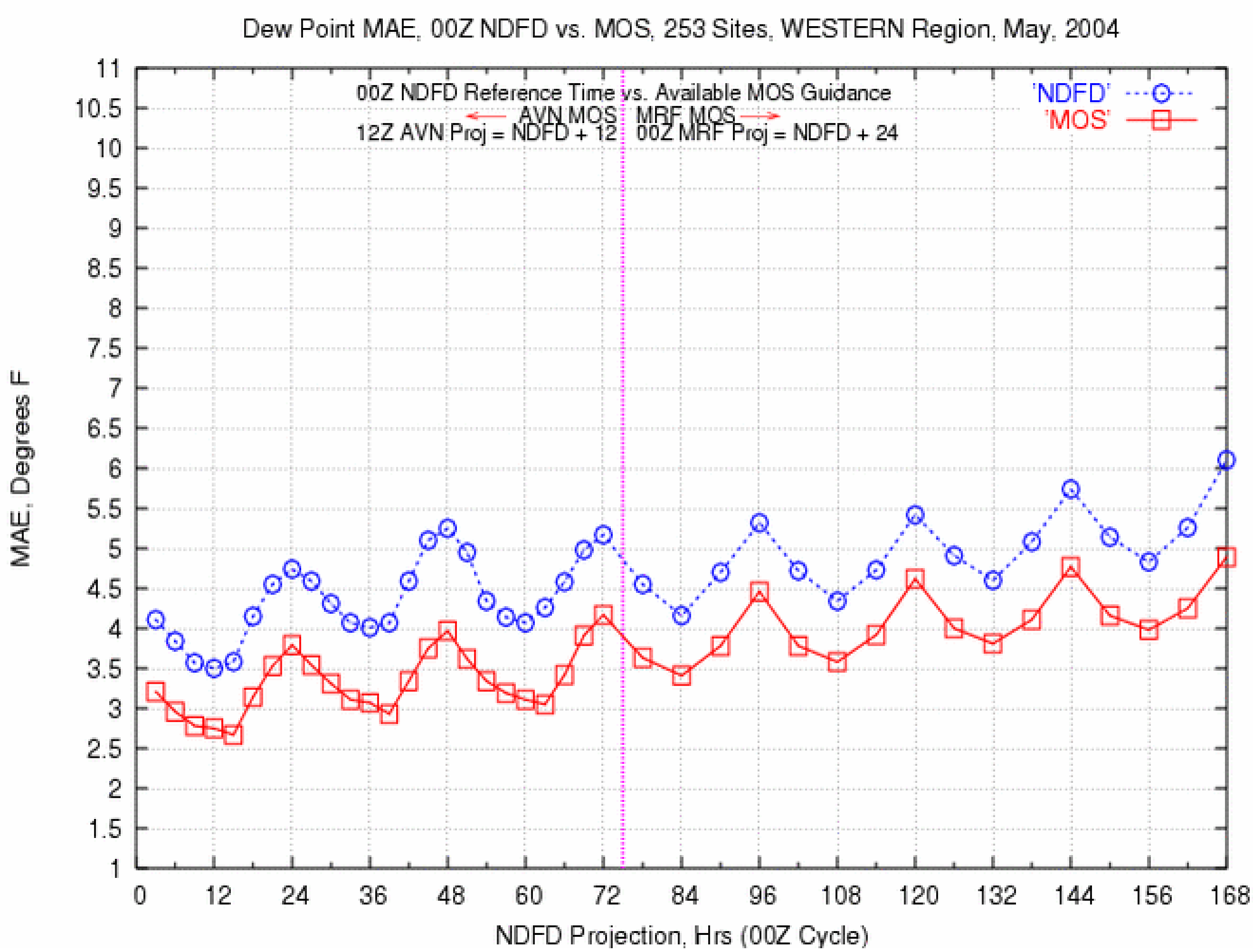
Examining the NDFD point verification for surface temperature and max/min temperature over roughly the last year, shows MOS to be doing better overall, but not significantly. MOS and NDFD MAE mostly vary between 2.5 degrees F and 5 degrees F with the larger errors in the later time projections and in the winter months. Of note, however, is that the MOS MAE at the Western Region MOS sites is just about always less than one degree F better than the NDFD forecast, and makes one wonder how much this is due to MOS being a point forecast and the NDFD forecast being an adjacent grid point forecast or average grid point forecast. There are clearly instances when Western Region forecasters out perform MOS as exhibited by the following Max Temp chart from March 2005.



Gridded temperature verification offers another set of challenges as noted in section one above. More comments on gridded temperature verification may be found in section IV.

III. Dew Point Temperature, POP12, and Wind Speed

The verification scheme for surface dew point temperatures is similar to that for surface temperatures. Hence, overall MOS does better in the point verification, but not significantly better at most sites. The Western Region composite error score overall is 1 to 1.5 degrees F worse than MOS which again may be predominately due to the superior MOS dataset. The graphs of the MOS and NDFD errors are almost always parallel and show a diurnal nature in most instances, i.e. dew point temperature errors are almost always higher in the afternoons. The parallelism of the MOS and NDFD errors may indicate that Western Region forecasters on the whole rarely deviate from the MOS dew point forecasts. The following chart is a monthly example of the above thoughts.



For POP12 verification, only point verification is performed by the NDFD verification system at the approximately 1300 MOS sites. As with temperature, MOS and the NDFD forecasts are verified at the MOS/METAR site, but the NDFD forecast is from an adjacent grid point. Verifying observations are computed from hourly or 6-hourly information contained in the METAR reports. Brier scores are comparable for the two forecasts, and overall MOS performs better in the first 72 hours while Western Region WFOs on many occasions better MOS in the 84 to 132 hour time frame.

Wind speed forecasts verified at MOS points and using the RUC analysis for grid verification show Western Region WFOs doing an excellent job at all time projections out to 168 hours. For the gridded forecasts, MAEs over approximately 90-95% of the Western Region shows errors of 6 knots or less while the majority of the windspeed biases are 3 knots or less. Point wind speed forecasts are only verified if the NDFD forecast, MOS forecast, or observed wind speeds are 8 knots or more. MOS forecasts are only verified out to 72 hours. Three trends are observed from the point forecasts. Western Region forecasts have larger MAEs in the morning, mostly no more than 1 to 1.5 knots more than the afternoon hours, while both morning and afternoon biases are positive (generally 2-4 knots in the morning and 1-2 knots in the afternoon). On average MOS forecasts are a little better by 1-3 knots than the Western Region NDFD forecasts, and again, the NDFD forecast is from an adjacent grid point versus the MOS statistical point forecast.

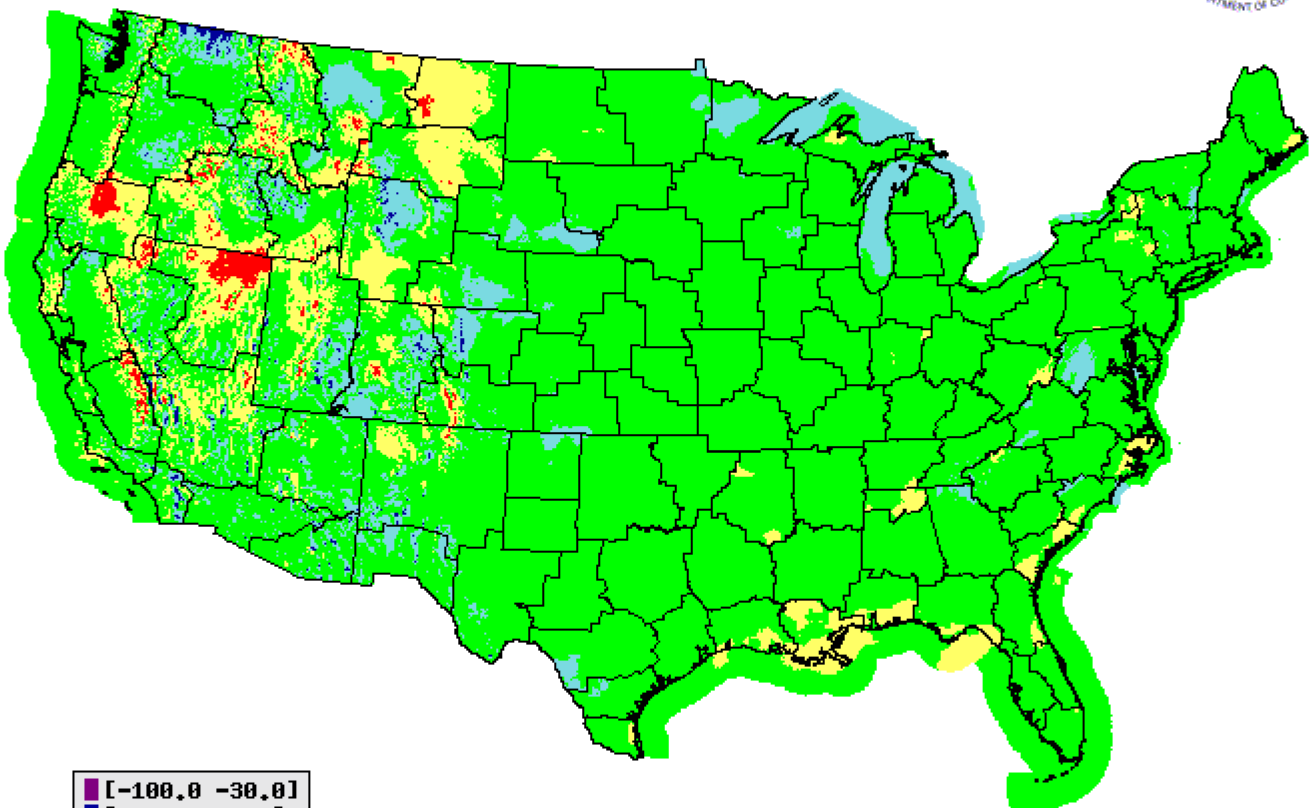
IV. Some Uses and Thoughts for the NDFD Verification

Although the MDL NDFD verification provides important initial feedback, as noted, problems abound with the technique. Much work is needed to improve the system to more accurately depict WFO performance versus MOS and the ambient weather. Advancements here range from an improved verifying analysis, the so-called Analysis of Record, to switching to a more level playing field versus MOS. Concerning the latter, gridded MOS forecasts at 5 KM for the CONUS will be available on AWIPS in the fall of 2005. These should be verified along with, and compared to, the WFO NDFD 5 KM forecasts. Both systems will have to contend with the strengths and weaknesses of whichever verifying surface analysis is used. Progress is being made on a "proof of concept" Analysis of Record at 5 KM resolution which is currently scheduled for no later than AWIPS OB7 (fall of 2006). Both of these efforts would be a step in the right direction.

Still with the weaknesses of the current NDFD verification system, WFOs may extract useful information that can improve their forecasts – especially in the 4-7 day range. The monthly temperature MAEs of the point verification show instances when the MOS errors and NDFD errors at individual sites exhibit, or are close to, double digit errors in the 4-7 day range. Since MOS in this time frame moves toward climatology, it seems there are instances when an NDFD forecast could take advantage especially in regimes of anomalously high or low geopotential height fields. In addition, WFO's could put more time into examining these poor performing MOS sites to determine why they do worse than at other MOS sites in their CWFA. Perhaps better statistical relationships could be made between the better and poorer MOS performing locations.

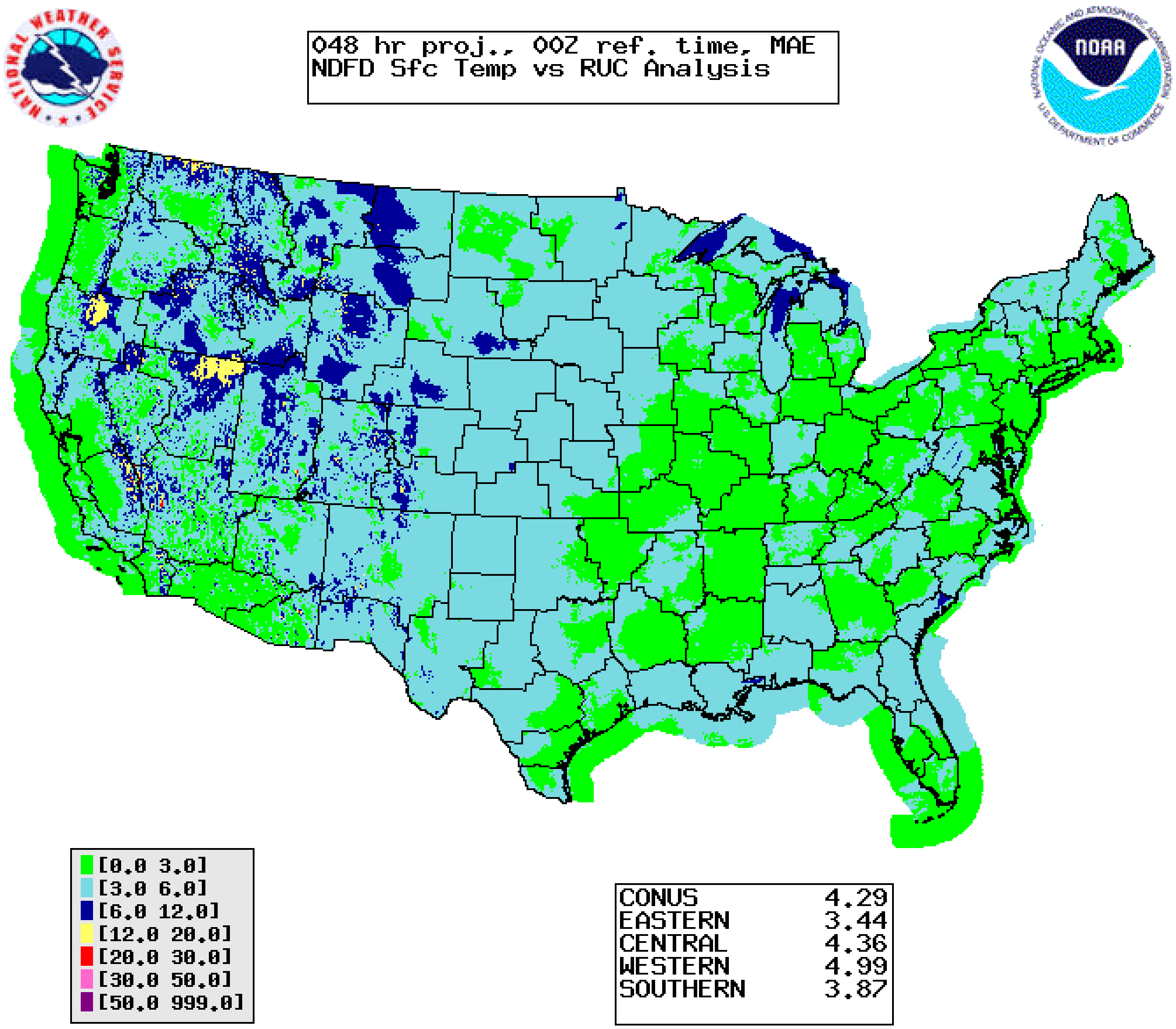
In the same vein, examining the NDFD temperature bias and MAE maps in the gridded verification may provide added insight into WFO performance. The first map below is an example of a NDFD winter temperature bias versus the RUC verifying analysis that's typical of the past two winters. The second map is the corresponding temperature MAE for the same time projection (48 hr) and month (February 2004).

048 hr proj., 00Z ref. time, Bias (fcst-obs)
NDFD Sfc Temp vs RUC Analysis



[-100.0 -30.0]	
[-30.0 -10.0]	
[-10.0 -3.0]	
[-3.0 3.0]	
[3.0 10.0]	
[10.0 30.0]	
[30.0 100.0]	

CONUS	.26
EASTERN	-.05
CENTRAL	-.63
WESTERN	1.30
SOUTHERN	.33



WFOs might examine their CWFAs for persistent, large bias and MAE areas and determine how much is due to the RUC analysis and how much is due to their forecast performance. Irregardless of the two weightings, modifying the forecast to account for a portion of the large bias and MAE would work to mitigate the large areas and improve the forecast for the user. This might also work to lessen the NDFD temperature error at some of the poorer performing MOS sites.

One agreement in the gridded forecast verification issue is that there is a long way to go to provide a robust, more accurate, and acceptable verification system. Users of the current NDFD verification system need to understand its limitations and properly measure and weigh its feedback to the forecast staff and the customer. In addition, all stakeholders should continue to push for a more level verification playing field, a better Analysis of Record, and more education to ensure there’s a real understanding of the issues involved in point and grid field verification.

V. References

Benjamin, S.G., and CoAuthors, 2004: An Hourly Assimilation-Forecast Cycle: The RUC. Mon. Wea. Rev, 132, 495-518.

Dagostaro, V.J., W.A. Shaffer, M.J. Schenk, J.L. Gorline, A.A. Taylor, 2004: A Prototype Verification System for Examining NDFD Forecasts. Preprints, 17th Conf. on Probability and Statistics in the Atmospheric Sciences, Seattle, WA, Amer. Meteor. Soc., CD-ROM 2.7.

Myrick, David T, and Horel, John D.: Verification over the Western United States of Surface Temperature Forecasts from the National Digital Forecast Database. Under Review for Weather and Forecasting, Spring 2005.

NWS Technical Procedures Bulletin No. 490: RUC20 – The 20-km Version of the Rapid Update Cycle, April 2002.