2.7 A PROTOTYPE VERIFICATION SYSTEM FOR EXAMINING NDFD FORECASTS

Valery J. Dagostaro*, Wilson A. Shaffer,
Michael J. Schenk, Jerry L. Gorline
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

and

Arthur A. Taylor
RS Information Systems
McLean, Virginia

1. INTRODUCTION

Over the past few years, National Weather Service (NWS) forecast offices have significantly changed the way they produce and display forecast products. By the summer of 2003, all forecast offices in the conterminous United States (CONUS) produced forecasts for various weather elements on a high-resolution grid for the current hour to 7 days in the future. These forecast grids are centrally collected and merged into a national database called the National Digital Forecast Database (NDFD) (Glahn and Ruth 2003). The contents of the NDFD are available to NWS customers and partners on an experimental basis, enabling them to develop new and creative uses for NWS products.

To provide NDFD forecasters and users with feedback regarding the skill and accuracy of these new gridded forecasts, the Meteorological Development Laboratory (MDL) developed a prototype verification system consisting of a combination of traditional point verification plus grid comparisons. The point verification component is essentially a traditional verification system in which forecasts at specific points are matched with verifying observations of sensible weather elements. Where available, a standard of comparison such as climate, persistence, or model output can be used in a comparative verification. Currently, our primary standard of comparison is the Model Output Statistics (MOS) guidance (Dallavalle, et. al. 2004; Erickson 1999) based on the Global Forecast System (GFS) (NCEP 2003). The grid verification component is a simple gridpoint-by-gridpoint comparison of one or more forecast grids with a derived observational analysis on a grid. Similar to the point forecasts, the NDFD grids can be compared to some standard such as direct model output, as long as the grids coincide in resolution and orientation. The point and grid verification components are treated separately in the MDL prototype.

To prove the feasibility of the MDL prototype and to quickly provide some basic information regarding the skill and accuracy of NDFD forecasts, we limited our verification to two of the possible 24 hourly NDFD forecast releases, namely the 0000 and 1200 UTC releases. Note that the NDFD is actually updated continuously as forecast grids are received, but for practical purposes, the NDFD grids available to users are currently processed on an hourly basis (Boyer and Ruth 2003). Our intent is to verify all weather elements in the NDFD; however, at the time of this writing, only a few elements where the observations were readily available are verified. We currently provide some basic statistics for temperature, dewpoint, maximum/minimum (max/min) temperature, Probability of Precipitation (PoP), and wind speed. Forecasts from 3 hours to 7 days after the 0000 and 1200 UTC NDFD release times are verified. Given that the prototype's point and grid verification components are separate, there may be instances when point verification results are available for a given weather element, forecast projection, and score that are not available in the grid component and vice versa. Usually this occurs when a suitable observation source is not yet available.

Details regarding the point and grid verification components of MDL's NDFD verification prototype are described below. We discuss the current status as well as future enhancements.

2. POINT VERIFICATION COMPONENT

MDL's considerable experience in the point verification arena served as the starting point for the prototype's point verification component. We begin by obtaining point forecasts at 1279 CONUS locations for which GFS-based MOS is available. We use a modified nearest-neighbor technique to obtain forecasts at desired locations from the four surrounding gridpoint values. For locations over relatively flat terrain and not close to water, we choose the value at the nearest gridpoint. If the nearest gridpoint is over water or its elevation differs from the desired location by 500 ft or more, the second closest gridpoint is chosen. If that point is also unacceptable, the third or fourth surrounding gridpoint, respectively, is used. If all of the gridpoints exceed the 500 ft threshold, the forecast is assumed to be missing. The 500 ft threshold was chosen

**Corresponding author address*: Valery J. Dagostaro, W/OST25, Room 10442, SSMC2, NOAA, 1325 East-West Highway, Silver Spring, Maryland, 20910-3283; e-mail: Valery.Dagostaro@noaa.gov.

to match the value used in the NDFD forecast coordination process among neighboring NWS forecast offices; however, the value is customizable for each weather element.

Next, the NDFD point forecasts are matched with the latest GFS-based MOS available to the forecasters at the time the NDFD forecasts were prepared. Specifically, NDFD forecasts released at approximately 0000 UTC are compared to GFS-based MOS (i.e., the “MAV” guidance) from the previous 1200 UTC model forecast cycle for NDFD forecast projections from 3 to 60 hours. Beyond 60 hours, forecasters have only the GFS-based “MEX” guidance available to them, with the MEX MOS produced only from the 0000 UTC model run. So, for projections beyond 60 hours, the 0000 UTC NDFD forecasts are compared to MEX MOS based on the previous 0000 UTC GFS model run. Conversely, the NDFD forecasts released around 1200 UTC are compared to MAV and MEX MOS based on the previous 0000 UTC model forecast cycle. Note that only the forecast projections for which we had a matched sample of NDFD and MOS forecasts are verified. Temperature, dewpoint, and wind speed forecasts are verified directly against METAR observations. For max/min temperature and 12-h PoP, verifying observations are computed from hourly or 6-hourly information contained in METAR reports.

At this time, one or two basic scores are computed monthly for each weather element. For PoP, we compute the Brier Score (Brier 1950); for temperature, dewpoint, and wind speed, we compute mean algebraic error (bias) and mean absolute error (MAE). Scores are computed for a one-month period for each of the 1279 points, the four CONUS NWS regions, and the CONUS. Here, a month is defined as the calendar month in which a forecast was produced. In other words, forecasts produced on September 30 and verifying within the first 7 days of October are counted in the September statistics.

To fulfill users’ varying needs, monthly scores are displayed in various forms such as error maps, bar and line charts, and ASCII text. Fig. 1 shows a sample error

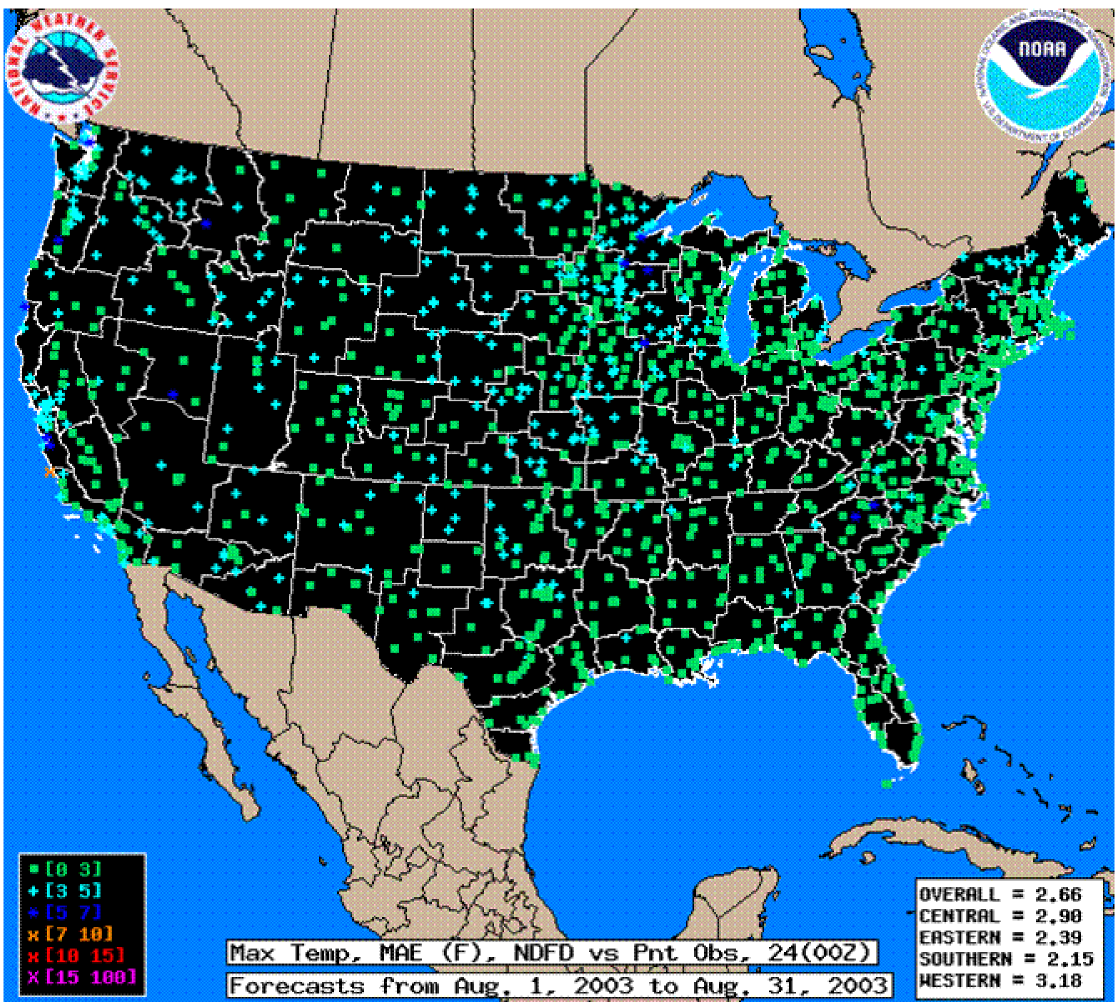

Figure 1. Sample of a CONUS NDFD point verification error map.

map for the CONUS in which errors are denoted by color-coded symbols plotted at the station locations. To view the actual values at specific locations, users may display either a similar error map zoomed in to the desired NWS Region (Fig. 2) or a file written in downloadable ASCII text (Fig. 3). Alternatively, we've also provided the scores in "shapefile" format for use in other programs and database applications. This format gives the user more flexibility to combine scores into user-defined regions and modify the display as desired.

Besides the map-based displays and downloadable data files, we also produce a series of graphs showing error as a function of forecast projection to give users a better overall picture of error trends over all forecast projections. These charts show the comparative results for NDFD and MOS for the CONUS (Fig. 4) and for each of the four CONUS NWS Regions.

Finally, for a limited number of forecast projections, weather elements, and stations, we also compared the NDFD and GFS-based MOS to the official NWS forecasts collected via the AWIPS Verification Program (AVP) (Morris and Kluepfel 1999). For max/min temperature and PoP only, we obtained the official NWS forecasts for the first four public forecast periods for as many AVP locations as possible (actually somewhat less than the 208 CONUS AVP sites because some transmit data for just aviation weather elements). A sample bar chart is shown in Fig. 5. Note that at the time of this writing, we had not yet decided whether to continue monthly production of these bar charts.

3. GRIDDED VERIFICATION COMPONENT

The second component of MDL's NDFD verification prototype is a basic grid-to-grid comparison. At each gridpoint on the 5-km NDFD grid, the forecast is compared to an analysis value at that gridpoint. In keeping with our desire to prove the feasibility of our system as well as provide a small set of statistics to users in a reasonable time, we used whichever analysis was readily available

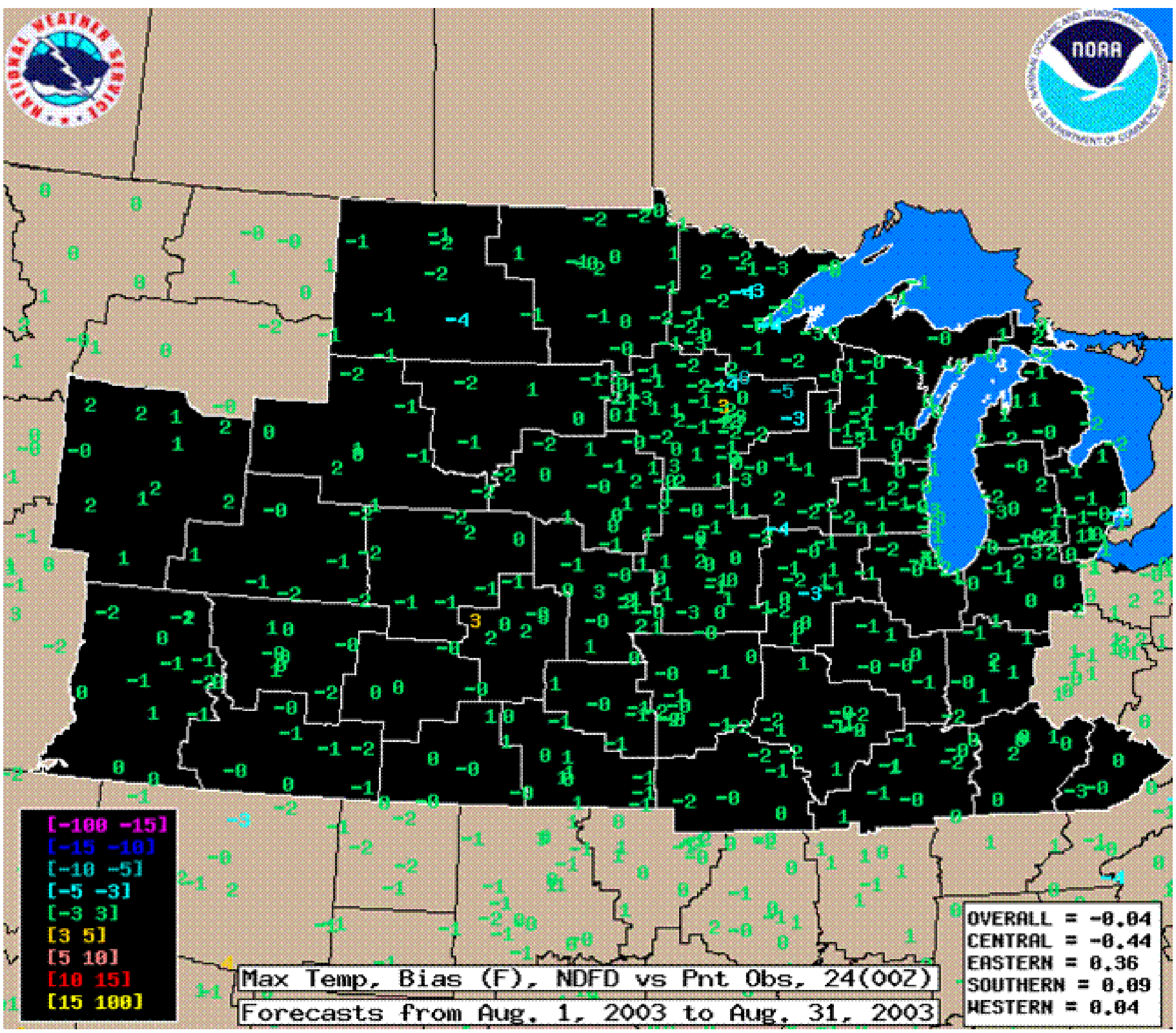

Figure 2. Sample regional NDFD point verification error map.

```
1200 UTC NDFD REFERENCE TIME 012-h PROJECTION MAX MAE

STATION        NUM_CASES       NDF_MAE      MOS_MAE
OVERALL      32283.00000       2.46000      2.73000
CENTRAL      11267.00000       2.60000      3.37000
EASTERN       7432.00000       2.28000      2.16000
SOUTHERN      7421.00000       2.04000      2.34000
WESTERN       6163.00000       2.92000      2.71000
K2WX            28.00000       3.36000      6.25000
K4BL              .00000    9999.00000   9999.00000
K8D3            25.00000       2.08000      3.24000
K9V9            28.00000       3.00000      5.32000
KAAO            27.00000       2.93000      3.56000
KABR            28.00000       2.79000      2.96000
KACB            26.00000       2.46000      1.77000
KADG            27.00000       2.22000      2.22000
KADU            28.00000       2.36000      3.54000
KAEL            27.00000       2.44000      3.93000
.
.
.
```
Figure 3. Sample of a partial ASCII text file containing point verification scores.

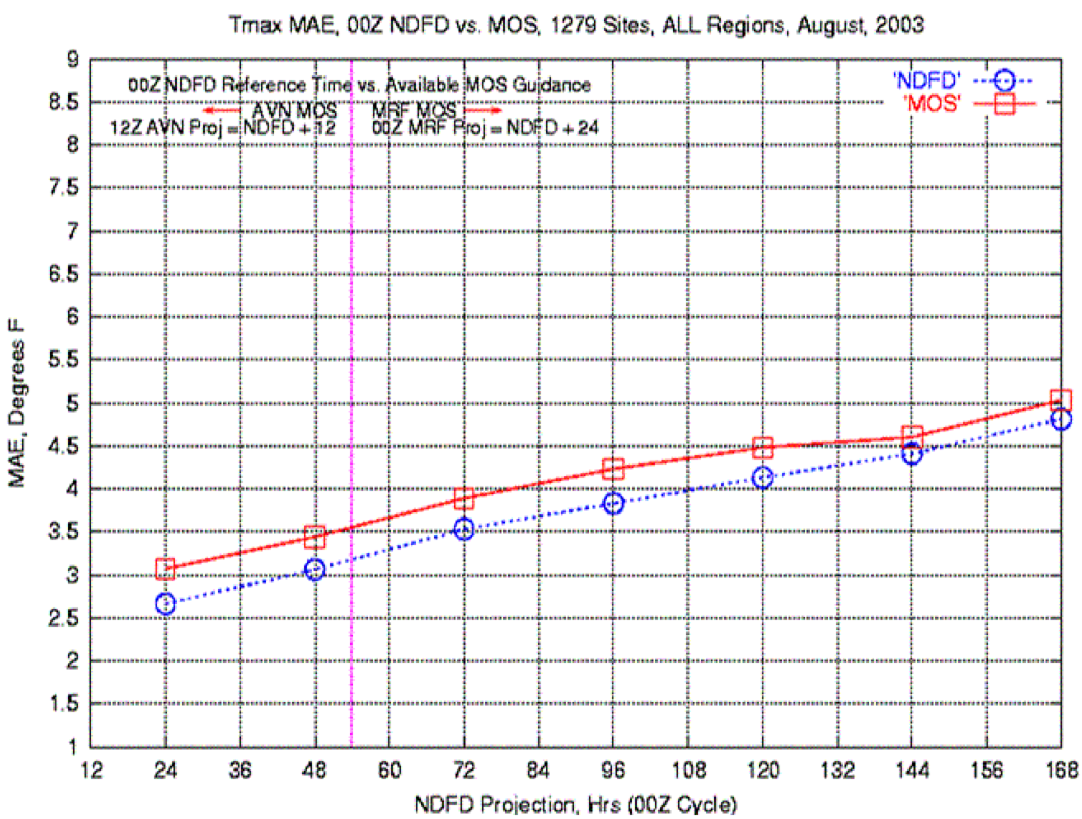

Figure 4. Sample line chart showing NDFD and GFS MOS as a function of forecast projection.

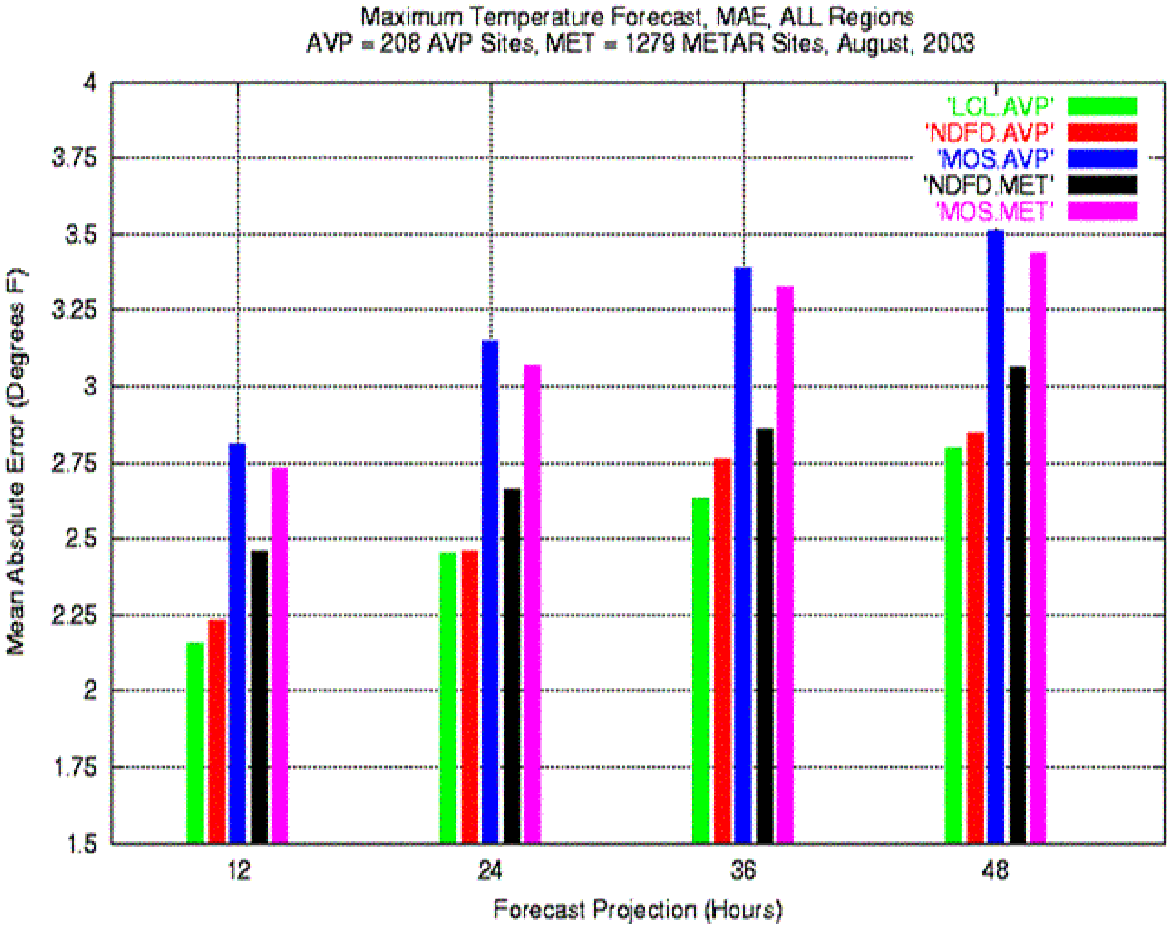

Figure 5. Sample bar chart showing comparative verification results for official NWS forecasts ("LCL AVP"), NDFD ("NDFD AVP"), and GFS MOS ("MOS AVP") for approximately 208 AVP sites along with NDFD ("NDFD MET") and GFS MOS ("MOS MET") statistics based on a larger sample of approximately 1279 METAR sites.

for the weather elements we currently verify. At this time, we use fields from the 20-km RUC analysis interpolated to the 5-km NDFD grid (and, for wind, turned to earth orientation) as the verifying observation grid. Note that the lack of certain observational fields such as max/min temperature and 12-h precipitation amount limits the number of weather elements we can currently verify. Work is underway to identify and obtain other observational data sets for those weather elements missing from the RUC analysis. Similar to the point verification component, our prototype can accommodate a standard of comparison such as direct model output though we don't currently include it in the monthly verifications.

Each month we compute MAE and bias for three weather elements: temperature, dewpoint, and wind speed. Since the gridded verification is resource-intensive, we chose to verify only selected forecast projections at 12-h intervals from the 0000 and 1200 UTC NDFD release times. For temperature, we chose projections that most closely match the time of day when the max or min would normally occur over the largest portion of the CONUS. For example, for the NDFD release time of 0000 UTC, we verified the temperature grids at 12-, 24-, 36-, ... 168-hr. For continuity, we verify dewpoint and wind speed forecasts for the same projections. As in the point verification system, we verify all forecasts produced in that calendar month, including those whose verifying observations occur in the following month. Monthly average errors are computed at each gridpoint, and are displayed on a map background using a custom-designed GIS display program. Fig. 6 shows a sample gridded verification error map with color-coded error ranges.

4. DISSEMINATION OF SCORES

Monthly verification results produced by our prototype are currently accessible by NWS personnel via two websites. Due to the experimental status of the NDFD and verification prototype, at the time of this writing, the statistics were available only to internal NWS users. Access to the verification results by outside users is discussed in more detail in Section 5. Upon completion of a given month's processing, scores are posted on the MDL Evaluation Branch's limited-access NDFD verification website (MDL 2003). The same information is also posted

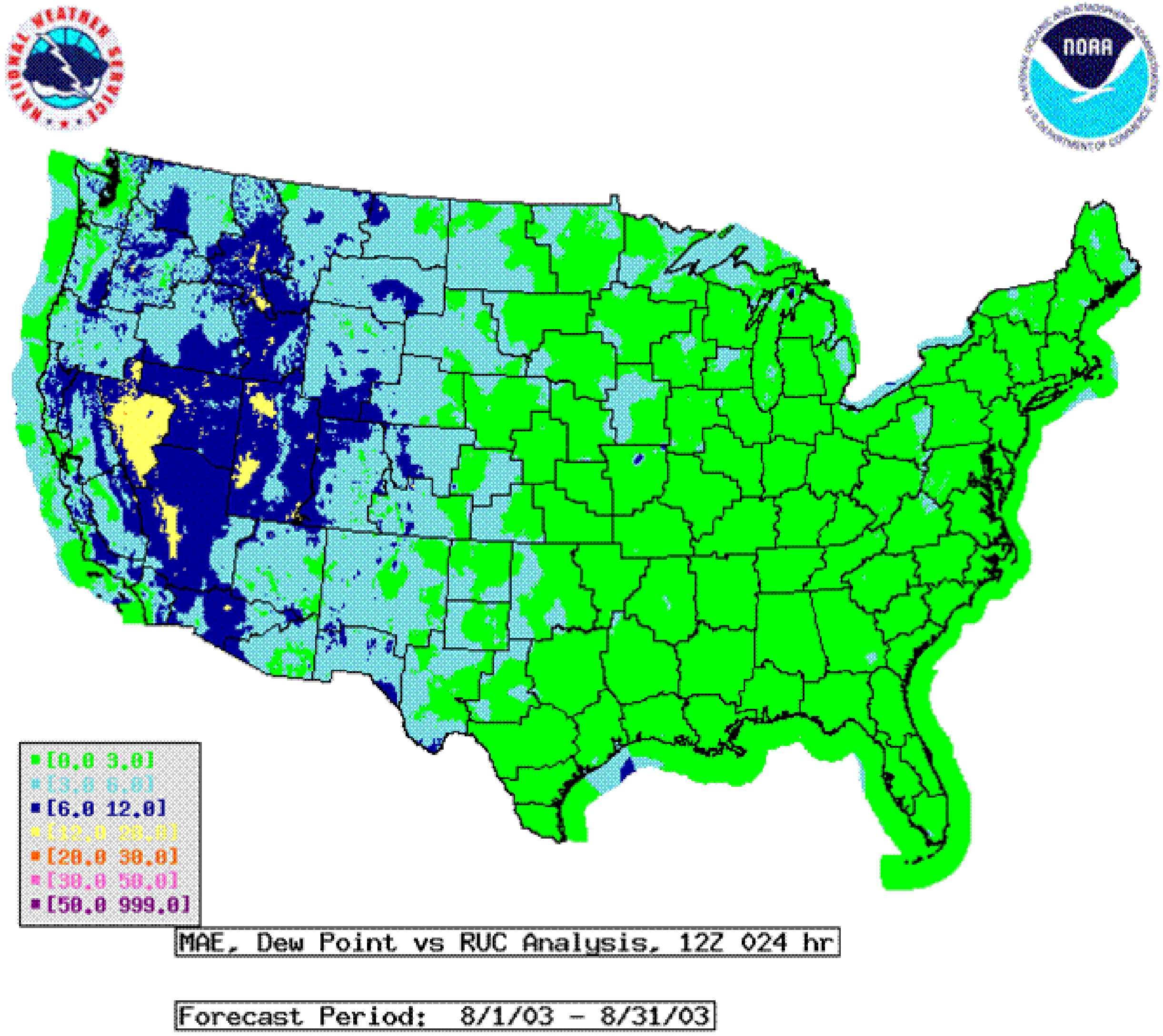

Figure 6. Sample gridded NDFD verification error map.

on the NWS's internal AWIPS NDFD website. Both websites are updated around the middle of the month to contain verification statistics for the previous month. These websites currently contain static images of error maps, plots, and tables, meaning that the user has no control over such things as the starting and ending dates of the data sample, how individual station or gridpoint scores are aggregated, and zoom capability.

5. PLANS

MDL's NDFD verification prototype is still very much a work in progress. Some enhancements we're currently pursuing are: streamlining the monthly processing system to post scores more quickly; adding software to handle new weather elements (wind direction, sky cover, and precipitation amount) and scores; aggregating scores by forecast offices in addition to NWS Regions; and incorporating mesonet observational data sources. We expect all but the last goal to be completed relatively soon. As for the mesonet observational data, we've taken steps to acquire a sample high-resolution analysis of mesonet data over the Western U.S. (Lazarus, et. al. 2002) for possible use in the gridded verification.

We are developing a new, limited version of the current website that will be accessible to our non-NWS customers and the public. This website will contain images of monthly error maps for the CONUS and NWS Regions for all verified weather elements.

Finally, as we obtain user feedback regarding the usefulness of the information we provide, we'll consider redesigning websites to provide users maximum flexibility to extract scores for the exact combination of locations, forecast projections, dates, etc., they need. In addition, as resources permit, we could expand the number of NDFD release times verified. A logical first step would be to add the 0600 and 1800 UTC releases. At this time, it's unlikely NDFD forecasts for every release time would be verified until an automated, operational verification system is in place.

6. SUMMARY

MDL's NDFD verification prototype is currently capable of providing simple statistics for both point and gridpoint locations. Forecasts from a subset of the NDFD release times, forecast projections, and weather elements are verified on a monthly basis. For max/min temperature, temperature, dew point, and wind speed, we compute the MAE and bias for a comparative verification of NDFD point forecasts and GFS-based MOS at about 1279 locations. Similarly, we compute the NWS Brier Score for point PoP forecasts. For a smaller sample of AVP sites, weather elements, and projections, we include the official local forecasts in the comparative verification. Point verification results are displayed in the form of error maps, line and bar charts, and downloadable ASCII text. Scores are computed for individual stations, NWS Regions and the CONUS. Similar to the point verification component, only a subset of gridded forecasts are verified by using RUC analysis fields interpolated to the 5-km NDFD grid. MAE and bias are computed for temperature, dewpoint, and wind speed. MDL's prototype will continue to evolve as we gain expertise in the gridded verification arena.

7. REFERENCES

Boyer, T. R. and D. P. Ruth, 2003: National digital forecast database design and development. *Preprints 19th Int. Conf. on Interactive Information and Processing Systems for Meteorology, Oceanography, and Hydrology*, Long Beach, CA, Amer. Meteor. Soc., CD-ROM, 12.3.

Brier, G.W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1-3.

Dallavalle, J. P., M. C. Erickson, and J. C. Maloney, III, 2004: Model Output Statistics (MOS) guidance for short-range projections. *Preprints 20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., CD-ROM, 6.1.

Erickson, M. C., 1999: Updated MRF-based MOS guidance: Another step in the evolution of objective medium-range forecasts. *Preprints 17th Conf. on Weather Analysis and Forecasting*, Denver, CO, Amer. Meteor. Soc., 190-195.

Glahn, H. R. and D. P. Ruth, 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-201.

Lazarus, S. M., C. M. Ciliberti, J. D. Horel, and K. A. Brewster, 2002: Near-real-time applications of a mesoscale analysis system to complex terrain. *Wea. Forecasting*, **17**, 971-1000.

Meteorological Development Laboratory, cited 2003: NDFD verification scores. [Available online at http://slosh.nws.noaa.gov.]

Morris, K .R. and C. K. Kluepfel, 1999: WFO forecast verification in AWIPS: initial capabilities and future plans. *Preprints 15th Int. Conf. on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Dallas, TX, Amer. Meteor. Soc., 360-363.

NCEP Environmental Modeling Center, cited 2003: The GFS Atmospheric model. [Available online at http://wwwt.emc.ncep.noaa.gov/gmb/moorthi/gam.html.]