17.2

VERIFICATION OF NDFD GRIDDED FORECASTS OVER COMPLEX TERRAIN

David T. Myrick and John D. Horel

NOAA Cooperative Institute for Regional Prediction
Department of Meteorology, University of Utah, Salt Lake City, Utah

1. INTRODUCTION

The National Weather Service (NWS) is undergoing a major shift in the way it creates and distributes its forecasts. Rather than having forecasters type text products, they now use graphical editors to create (currently experimental) high-resolution gridded forecasts of weather elements. Besides their utility for improved protection of life and property, such gridded forecasts have the potential to be of great benefit for many applications in mountainous terrain, including improved fire weather forecasts and prediction of the dispersion of air pollutants in urban basins. This Interactive Forecast Preparation System (IFPS; Glahn and Ruth 2003) is continuing to evolve, including the development of appropriate procedures to verify these forecasts on local and national scales (Dagastaro et al. 2004).

Developing an effective gridded verification scheme is critical to identifying the capabilities and deficiencies of this new forecast process, especially in areas of complex terrain. This paper investigates techniques to verify forecasts in the western United States using analyses created at the Cooperative Institute for Regional Prediction (CIRP) at the University of Utah.

2. IFPS GRIDDED FORECASTS

Forecast grids of various fields at resolutions of 1.25, 2.5, or 5 km are produced at each NWS Warning and Forecast Office (WFO) and cover their respective County Warning Area (CWA). These local grids are combined to form one National Digital Forecast Database (NDFD; Glahn and Ruth 2003) at 5-km resolution. Primary NDFD elements currently produced include maximum and minimum temperature, probability of precipitation, and weather. Other elements include but are not limited to temperature, dewpoint, and sky cover. Elements are available at up to hourly temporal intervals (with the exception of maximum and minimum temperature) with lead times up to 7 days.

Only gridded temperature, dewpoint and wind forecasts issued at 0000 UTC are evaluated here. The forecasts available from NDFD for a particular grid box are intended to be representative of the conditions throughout that area (a 5 x 5 km^2 region). Forecast skill can be assessed by either interpolating the forecasts to locations where observations are available or comparing directly the gridded forecasts to a gridded analysis of the current state based upon the available observations. We have used both approaches as part of this study and have found the sensitivity to “point” vs. “gridded” verification to be small. For brevity, we will present results for gridded verification only.

3. VERIFYING ANALYSES

Surface data from weather observing stations across the United States have been linked together into a common database as part of MesoWest (Horel et al. 2002). The Automated Surface Observing System network maintained by the NWS, Federal Aviation Administration, and the Department of Defense is supplemented by networks supported by over 120 government agencies and commercial firms. Our validation of the IFPS forecast grids relies upon 5 km objective analyses on the NDFD grid over the western United States that are derived from the Advanced Regional Prediction System Data Assimilation System (ADAS). ADAS employs the Bratseth method of successive corrections, an inexpensive analysis procedure that can be run in near-real time over a large horizontal domain at high horizontal resolution (Lazarus et al. 2002, Myrick et al. 2004). The background field used by ADAS is the 20 km version of the Rapid Update Cycle (RUC; Benjamin et al. 2004), which is downscaled to the 5km NDFD terrain grid. The ADAS analysis typically incorporates over 2,000 surface weather observations each hour from Mesowest to adjust the RUC background field. We have found that the RUC generally provides a good background from which to derive the high resolution analyses.

For stations in the continental United States that report hourly aviation observations, Benjamin et al. (2004) found RMS differences between the RUC surface analyses and observations to be on the order of 1.5^{o}C for temperature and 1.5 m s^{-1} for wind speed. As illustrated in Table 1, larger biases and RMS differences are evident between the RUC analyses and surface observations over

---

Corresponding author address: David T. Myrick, Dept. of Meteorology, Univ. of Utah, 135 S 1460 E Rm. 819, Salt Lake City, UT 84112; e-mail: dmyrick@met.utah.edu.

the mountainous terrain of the West. The RUC tends to be too warm in the early morning hours, especially over the valleys of the Intermountain West. In addition, the RUC wind speed tends to be higher than observed during the morning and afternoon. These large differences in wind speed between the RUC and observations depend in part upon the uncertainty and representativeness of the surface observations. For example, the surface wind field in mountainous regions during the night often becomes decoupled from the prevailing synoptic-scale flow, upon which the RUC focusses.

The specified ratio of the observation to background field RMS error controls the degree to which the ADAS analyses fit the observations (Lazarus et al. 2002). Overfitting and spurious analyses in data void regions can result if the observational error is assumed to be too small. As shown in Table 1, we have constrained the ADAS analyses to be within 1-2^{o}C and 1-1.5 m s^{-1} for the observations available in the West.

4. WINTER 2003-2004 VERIFICATION

NDFD experimental forecast grids of temperature, dewpoint and wind speed for the 2003-2004 winter season (18 November 2003 - 29 February 2004) are verified every twelve hours at 0000 and 1200 UTC for forecast lead times from 0.5 - 7 days. As an illustration of the analysis and forecast grids, the ADAS analysis at 0000 UTC 14 January 2004 and the 48 h NDFD temperature forecast valid at the same time are shown in Fig. 1. A prolonged period of upper level ridging over the western United States and extensive snow cover in mountain valleys led to the development of persistent cold pools in many basins in the Intermountain West. For example, the ADAS temperature analysis (Fig. 1b) depicts inverted temperature profiles with height with low temperature in the valleys and higher temperature along nearby slopes in northern Utah, southwestern Wyoming, northeastern Nevada, and southern Idaho. Temperatures at high elevations (e.g. Wasatch and Uinta Mountains in northern Utah) are as much as 10°C higher than nearby valley locations.

	RUC 0000 UTC	RUC 1200 UTC	ADAS 0000 UTC	ADAS 1200 UTC
T BE (°C)	0.1	1.5	0.0	-0.2
T MAE (°C)	2.0	2.9	1.0	1.3
T RMS (°C)	2.7	3.9	1.6	2.1
WS BE (ms^{-1})	1.4	1.9	-0.1	-0.1
WS MAE (ms^{-1})	2.3	2.6	0.9	0.9
WS RMS (ms^{-1})	3.1	3.5	1.5	1.5

Table 1. Temperature (T) and wind speed (WS) bias error (BE), mean absolute error (MAE) and root mean squared error (RMS) of RUC and ADAS analyses at 0000 and 1200 UTC verified against MesoWest observations during the 2003-2004 winter season (18 November 2003 - 29 February 2004).

The 48h NDFD temperature forecast grid valid 0000 UTC 14 January 2004 is shown in Fig. 1a. For the most part, forecasters at WFOs in the Great Basin predicted the valley cold pools to persist but they underforecast their strength. Sharp discontinuities in the temperature forecast along some boundaries between adjacent CWA’s demonstrates the challenge to produce a seamless forecast in regions of complex terrain. For example, the Elko, NV WFO forecasters predicted higher temperature in the valleys compared to the temperature predicted by forecasters at the neighboring Salt Lake City, UT WFO to the east and Boise, ID WFO to the north.

The seasonally-averaged 48h forecast bias highlights CWA boundary coordination problems in some locales (Fig. 2a). For example, the Elko, NV and Medford, OR WFO’s tended to overforecast 48h temperatures during the winter season, while the neighboring Boise, ID and Salt Lake City, UT WFOs tended to have smaller biases (Fig. 2a).

The 48h root mean squared difference (RMS) between the forecast and analysis grids across the western United States are shown in Fig. 2b. The smallest RMS errors (< 2°C) are located over the desert southwest near the California/Arizona border and offshore. The largest RMS errors (4-7°C) are located over the higher terrain of the Sierra Nevada Mountains and Great Basin.

In order to focus upon the skill of the forecasts on the synoptic and meso- scales, departures from the seasonal mean forecast and analysis grids at each grid point are determined. Hence, the seasonal forecast bias shown in Fig. 2a is removed using this procedure. Figure 1d shows the departure from the seasonal mean of the 0000 UTC 14 January 2004 ADAS analysis. Valleys of the northern Great Basin are significantly colder than the seasonal average while nearby higher elevations are warmer than the winter season mean. Cold pools are evident in many other regions of the West, including the Central Valley of California and valleys in Wyoming, Washington and northern Oregon. As a general rule, the NWS forecasters in the West did not anticipate the intensity of the cold pools 48h earlier (Fig. 1c) and at longer lead times expected the cold pools to weaken substantially (not shown).

Summaries of the bias and RMS statistics averaged over the entire West as a function of forecast lead time

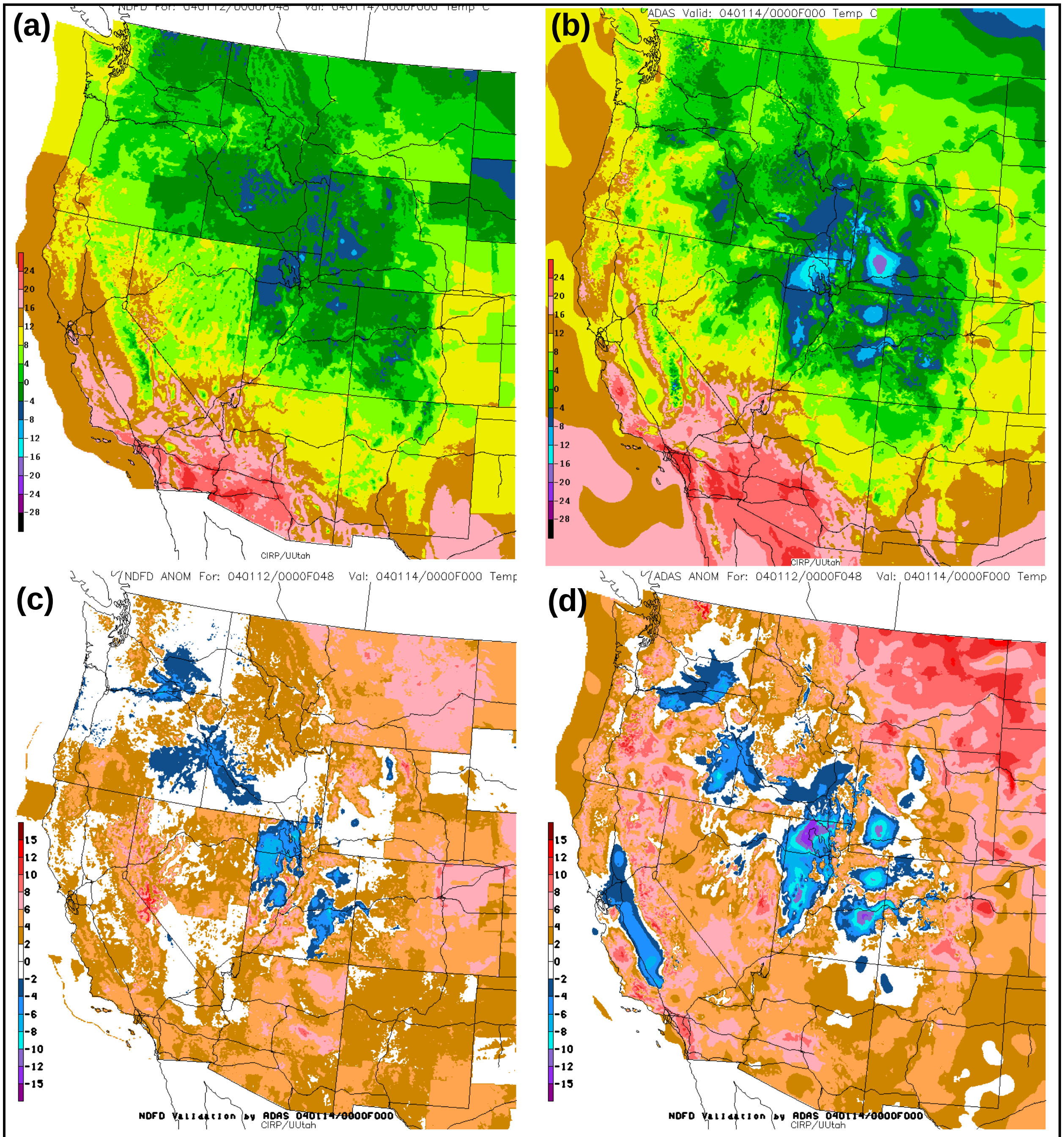

Figure 1. (a) Experimental 48h NDFD forecast of temperature (°C) valid at 0000 UTC 14 January 2004, (b) ADAS analysis of temperature (°C) at 0000 UTC 14 January 2004, (c) Same as (a) except the seasonal mean of the forecasts at 48h lead time has been removed, (d) Same as (b) except the seasonal analysis mean has been removed.

for temperature, dewpoint and wind speed are shown in Fig. 3. NDFD forecasts issued at 0000 UTC from November 18, 2003 to February 29, 2004 are compared to the 20km RUC analyses downscaled to the NDFD 5 km grid, ADAS analyses, and regions within the ADAS domain where the observations are more likely to make significant adjustments to the background (ADAS_C). The latter restriction is intended to focus upon the roughly 2/3 of the grid where at least 1 or more observations are available to adjust the RUC background. Areas of the grid that are eliminated are scattered throughout the domain but the largest regions omitted are offshore, to the east of the Sierra Nevada Mountains (e.g., Death Valley) and over northeastern Arizona.

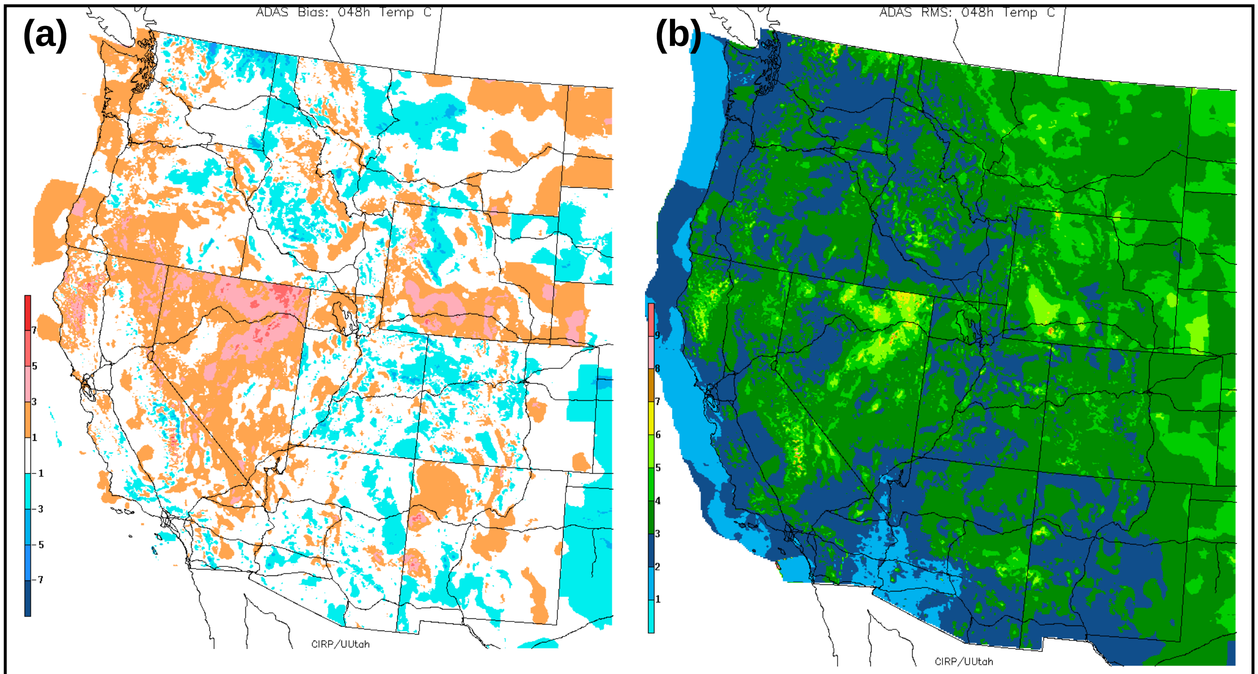

Figure 2. 48h NDFD temperature forecast (a) bias (°C; NDFD-ADAS) and (b) root mean squared difference (°C) for the period 18 November 2003 - 29 February 2004.

NDFD temperature bias and RMS errors are largest for forecasts valid at 1200 UTC (Fig. 3a). Forecasters tend to predict temperatures that are colder than that observed as determined either by the objective analyses (Fig. 3) or when compared directly to the observations (not shown). This larger bias in the morning may result in part from several interrelated factors. First, there is a difference between what is being predicted and what is being verified. The forecasters predict the minimum and maximum temperature for the day and then a diurnal curve is fitted to those extremes to obtain the hourly temperature grids. The observations to which the ADAS analyses are constrained are not observations of minimum and maximum temperature; instead they are a mixture of instantaneous observations (i.e., aviation reports) and temperature averages over periods from 5-60 min. Second, the interpolation step from the daily temperature extremes may lead to an underprediction of the temperature valid at 1200 UTC, since the minimum temperature (often a brief temperature spike downward that lasts only a few minutes) usually occurs within an hour or two of 1200 UTC. Third, the maximum temperature usually occurs several hours earlier than 0000 UTC so that the fit to the daily temperature extremes results in a downward trend in temperature around 0000 UTC and, hence, a smaller bias at 0000 UTC.

Relative to ADAS, NDFD RMS errors increase from 3-4^{o}C during the first day to 4.5-5.5^{o}C after 6 days (Fig. 3a). When the verification is limited to the regions where there is higher confidence in the verifying analyses, the errors are similar to those computed from the entire grid (contrast the length of the bars labeled ADAS to ADAS_C).

The bias and RMS errors for the NDFD experimental dewpoint forecasts exhibit less diurnal variation compared to the temperature forecasts (cf. Figs. 3b and 3a). While the biases are relatively small, the RMS differences increase from 4-6^{o}C as the forecast duration increases from 1-7 days.

Relative to the ADAS analyses (Fig. 3c) or the observations (not shown), the NDFD experimental forecasts tend to overpredict the magnitude of wind speed. The smaller bias of the NDFD forecasts relative to the RUC is due to the fact that the RUC wind speeds also tend to be higher than those observed at most locations in the West (see Table 1). The RMS errors of wind speed forecasts begin to saturate after 4 days.

The anomaly pattern correlation is computed between each pair of spatial anomaly maps in order to assess forecast skill as a function of large-scale weather patterns and forecast duration. For example, the correlation between the 48 h NDFD temperature forecast in Fig. 1c and its verifying analysis in Fig. 1d is 0.58, which suggests that the two maps share many features in common. We will use a threshold correlation of 0.5 as an indicator of forecast skill; a higher cutoff of 0.6 is often used for spatially smooth fields such as 500 mb geopotential height.

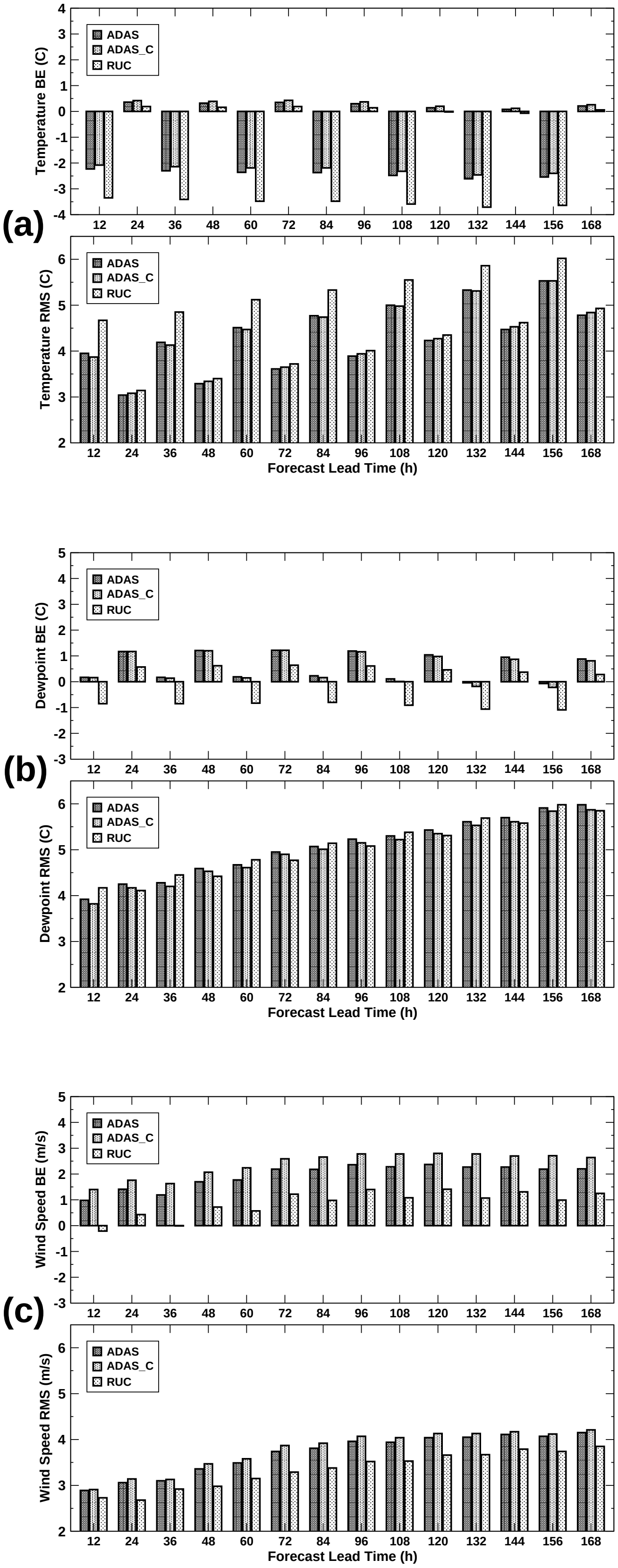

Figure 3. NDFD (a) temperature (°C), (b) dewpoint (°C) and (c) wind speed (ms^{-1}) bias error (BE) and root mean squared error (RMS) for the period 18 November 2003 - 29 February 2004 as a function of forecast lead time. Verification relative to the entire ADAS domain (ADAS), regions of the ADAS domain where the confidence in the analysis is higher (ADAS_C) and the 20km RUC surface analysis downscaled to the 5km grid. (RUC).

Figure 4 shows the anomaly pattern correlations between the 48h temperature forecast anomaly maps and verifying analyses. The spatial anomaly correlation between pairs of ADAS analyses separated by 48 h is also shown to provide a reference persistence forecast. For the most part, the 48h NDFD temperature forecasts exhibit skill, since the anomaly correlations are greater than 0.5. The day-to-day differences between the anomaly correlations are large compared to the differences arising from the various analysis approaches (i.e., compare the ADAS, ADAS_C and RUC values for any particular forecast). Hence, once the seasonal biases are removed (Fig. 2a), the day-to-day variations in the spatial patterns of the ADAS and RUC are similar to one another when considered over the West as a whole. During several active weather situations (e.g., cold front sweeping across the West from 21-23 November 2003 and an arctic outbreak to the east of the Rockies after 1 January 2004), the NDFD temperature forecasts at 48 h exhibited considerable skill, especially relative to the 48 h persistence forecasts. During the persistent upper-level ridging episode from 10-15 January 2004, the NDFD forecasts exhibited high skill but persistence forecasts had equal skill, and occasionally even greater skill. The particularly low NDFD skill on 25 November 2003 following immediately after a period of high skill resulted from the failure to capture several different synoptic and mesoscale details around the West, including rapid warming to the east of the Rockies.

The average over the entire winter season of the spatial anomaly correlations for temperature, dewpoint and wind speed are shown in Fig. 5 as a function of forecast duration. Little difference is seen between the spatial anomaly correlations derived from the entire ADAS analysis (ADAS) vs. the 2/3 of the domain for which the confidence in the analysis is higher (ADAS_C). The

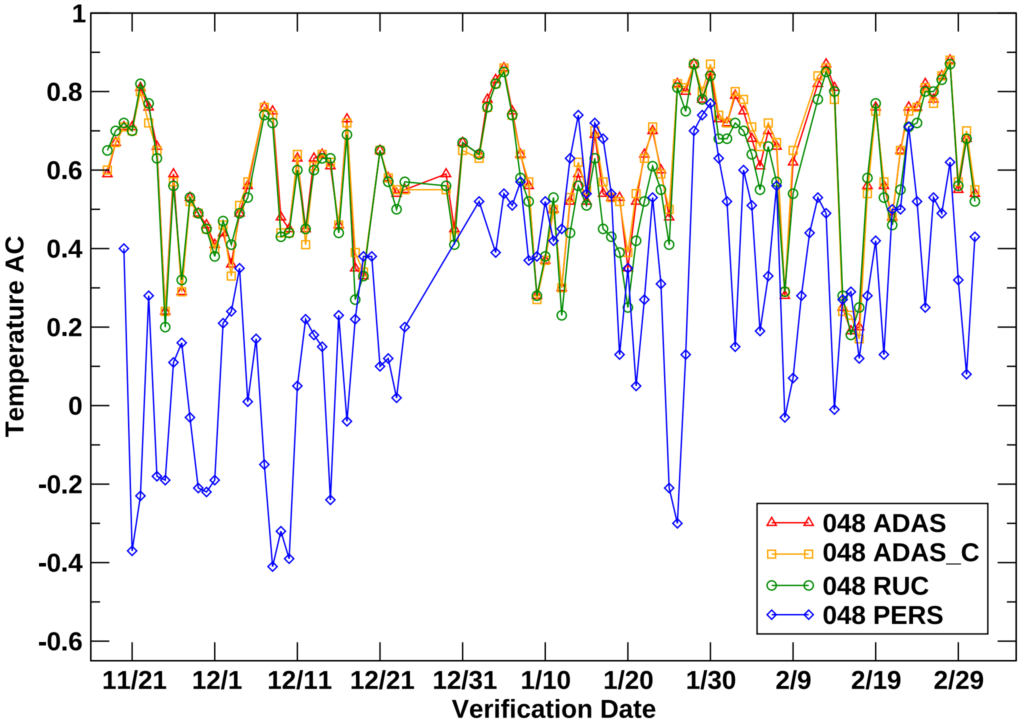

Figure 4. Anomaly correlations between each 48h NDFD temperature forecast at 0000 UTC during the period 18 November 2003 - 29 February 2004 and the analysis for the entire ADAS domain (ADAS), regions of the ADAS domain with higher confidence (ADAS_C), and the 20km RUC surface analysis (RUC). Anomaly correlations between the ADAS analysis 48 h earlier and the verifying ADAS analysis are also shown (PERS).

skill of the NDFD temperature (wind speed) forecasts would be estimated to be slightly lower (higher) if the RUC is used for the verification. Using the 0.5 value as a crude skill threshold, NDFD temperature forecasts exhibit skill out to 3 days (Fig. 5a), dewpoint temperature out to 1.5 days (Fig. 5b), and wind speed is on the verge of exhibiting skill during the first day only (Fig. 5c). The NDFD 72h temperature forecasts exhibit skill comparable to 24h persistence forecasts.

5. SUMMARY

Our results demonstrate the challenge inherent in creating high-resolution gridded forecasts, and creating them with spatial consistency and accuracy across the entire West. Revealing and resolving deficiencies through verification is a critical step to improve the IFPS system and NDFD products. Other preliminary verification studies underway indicate that NDFD forecasts provide added value and skill compared to the model guidance from which the forecasts are generated (personal communication, L. Cheng and L. Dunn, Salt Lake City, UT WFO). Verification needs to proceed within each CWA as well as on a national level. Local verification provides feedback to the forecasters that will result in improved forecasts; national verification will provide feedback to the user community on the utility of the NDFD products.

The verification of NDFD forecasts at the present time appears to be relatively insensitive to several aspects of the verification methodology. The conclusions obtained from our results are not significantly altered if the forecasts are evaluated at points or on a grid, if the forecasts are compared to ADAS analyses or downscaled RUC analyses, or if the verification is restricted to portions of the analysis domain where the analysis is thought to have higher quality. In other words, at the present time, the errors exhibited by the NDFD forecasts tend to be larger than the uncertainty in the verification data sets. Nonetheless, for the development of the IFPS forecasts as well as for verification of those forecasts, considerable research and development is required to improve the observational database, estimates of the errors of those observations, and analysis techniques in mountainous regions. Quantifying the distribution of observed precipitation on a grid in mountainous terrain is one of the biggest challenges.

The goals for the NDFD products need to continue to be examined carefully. Are the forecasts intended to be representative of the entire 5x5 km^2 box? To what extent

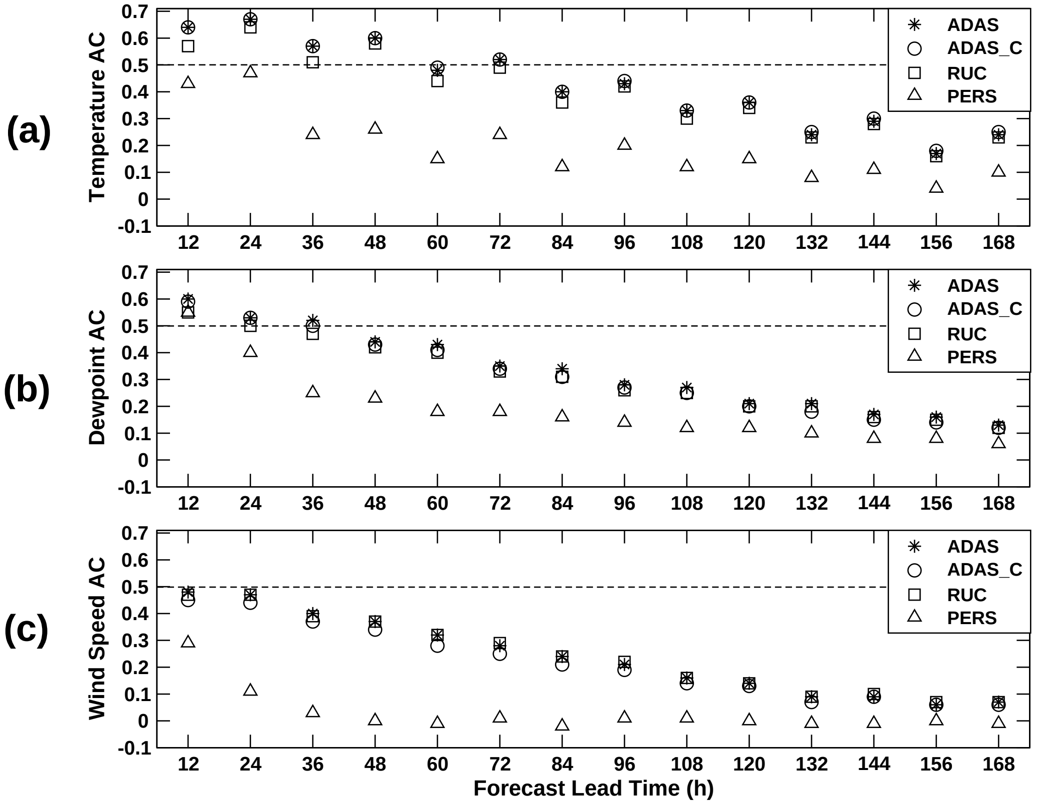

Figure 5. NDFD (a) temperature (b) dewpoint and (c) wind speed anomaly correlations for the period 18 November 2003 - 29 February 2004 as a function of forecast lead time. Verification of the NDFD forecasts relative to the entire ADAS domain (ADAS), regions of the ADAS domain with higher confidence (ADAS_C), and the RUC. Anomaly correlations between pairs of ADAS analyses as a function of lead time are also shown (PERS).

is that feasible when weather in mountainous regions can vary significantly within distances of 5 km? To what extent does a curve fit to the maximum and minimum temperature reflect the diurnal variation in temperature sampled every hour? What does the forecast for surface wind speed and direction represent in complex terrain? For example, in mountain valleys at night, the surface winds becomes largely decoupled from the larger-scale flow. Should the forecast focus upon the often highly localized winds or the synoptic-scale or mesoscale forcing?

6. 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, and 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.

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.

Horel, J. D., and Coauthors, 2002: MesoWest: Cooperative Mesonets in the Western United States. *Bull. Amer. Meteor. Soc.*, **83**, 211-226.

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.

Myrick, D. T., J. D. Horel, and S. M. Lazarus, 2004: A modification to the Bratseth method for use in complex terrain. Submitted to *Wea. Forecasting*. [Available online at http://www.met.utah.edu/jhorel/homepages/dmyrick/WAF_Myrick.pdf].