90th AMS Annual Meeting
20th Conf on Probability and Statistics
17–21 January 2010, Atlanta, Georgia

221 ESTIMATING THE ERROR OF THE BCDG ANALYSIS OF SURFACE DATA

Bob Glahn[1] and Jung-Sun Im
Meteorological Development Laboratory
Office of Science and Technology
National Weather Service, NOAA
Silver Spring, Maryland

1. INTRODUCTION

Analyses of surface-based meteorological observations have many uses. Such analyses are part of the assessment of the synoptic situation necessary for weather forecasting; this is as true today as it was 60 years ago. The methods of analyses have, however, changed since then. In the mid 1950's, data were being plotted by hand on maps and the analyses performed by humans. Since then, various methods of automated analyses have been developed in concert with the development of the digital computer.

One of the first techniques to appear was the least squares fitting of a polynomial over a fairly large area to the data (Panofsky 1949). Although this was refined by Gilchrist and Cressman (1954) to fit the data over a small area and was actually used in early numerical weather prediction experiments, it never became widely used. George Cressman, the first director of the National Meteorological Center (now the National Centers for Environmental Prediction, NCEP) and then director of its forerunner the Joint Numerical Weather Prediction Unit, recognized the potential for a technique developed by Bergthorsson and Doos (1955), and put a version of that into operation for analyzing upper air heights (Cressman 1959). This successive correction technique consisted of making multiple passes over the data, correcting each grid point on each pass by the data in the immediate vicinity. A very similar technique, basically differing only in the distance weighting factor, was proposed by Barnes (1964) and has been used extensively. Other very sophisticated methods of analysis, now called data assimilation, have been developed and are in operation at National Centers worldwide for providing initial conditions for numerical models. These latter methods employ relationships among free atmospheric variables that are not usually effective for use with surface observations.

[1]*Corresponding author address*: Harry R. Glahn, Meteorological Development Laboratory, National Weather Service, NOAA, 1325 East West Highway, Silver Spring, MD 20910. email: harry.glahn@noaa.gov.

As part of the Meteorological Development Laboratory's support to the aviation community, and the Next Generation Air Traffic Control System (NextGen) in particular, we have further developed the successive correction method, and have called it the BCDG method, the initials of the names of the primary developers. This method has been described by Glahn et al. (2009) for the analysis of MOS forecasts and by Im et al. (2010) for the analysis of surface data. While a considerable amount of effort has been placed on analysis methods, a more modest effort has been on estimating the analysis error associated with a particular method. Characteristically, errors associated with analysis schemes have been estimated with idealized data (e.g., a combination of sinusoidal waves) and/or upper air data where the patterns are relatively smooth. Barnes (1964) method is characteristically used with one or two passes. He suggested (op. cit.) "...direct application of the scheme to obtain maximum detail in regions wherein the data densities vary considerably is not recommended." Achtemeier (1989) suggested Barnes scheme be extended to three passes. Difficulties in making a good analysis and a good estimate of its error are evident by the exchanges between Smith and Leslie (1984) and Glahn (1987), Goodin et al. (1979, 1981) and Glahn (1981), and Fritsch (1971) and Glahn and McDonell (1971).

If one is going to estimate "analysis error," that error needs to be defined. One could be very interested in the location of fronts or other discontinuities, and not be overly concerned about "bland" areas. If such were the case, then a method would need to be defined that concentrated on that aspect. If one is concerned about making derivative calculations, then the method proposed by Achtemeier (1989) would be an option. Our use of the term "analysis error" is defined as a measure of the inability to recover the data values on which the analysis is based from the gridded analysis by linear interpolation anywhere within the extent of the grid. The measure used is absolute error (AE).

For an analysis algorithm such as BCDG, one can think of different ways of making such an estimate of the error. The simplest one is to interpolate into the completed analysis grid and compute the error at each data point. The average of these errors would be an overall measure of error. This has two undesirable attributes. First, the interpolated value itself has potential error. If the gridpoint representation of the data were perfect, one would still not, in general, recover the value at the data point exactly by interpolation. That is, the analysis process is not reversible. However, interpolation from a regularly-spaced grid to a random point is more exact than interpolating from randomly spaced points to a regular grid, especially when the data density is uneven and/or sparse; this error of interpolation is unavoidable, although one could adjust the interpolation method based on the relative densities of the grid points and the data points. The second, and major, difficulty is that any analysis process worth its salt can fit the data points rather closely, but still be poor where the data are sparse.

To attempt to overcome the second difficulty, one could withhold a few data points when doing the analysis, then compute the AE only at those points. Then, the analysis would not be affected by the withheld points, and the AE should be a measure of the overall error at points on the grid between grid points where there were no data values. While the analysis is deprived of those withheld data, this is acceptable provided the number of withheld points is a very small fraction of the total points. Withholding data for error estimation was used as early as 1962 by Thomasell (1962). By replication with the same data, withholding different sets of points, one can estimate the mean absolute error (MAE) for a particular set of data. By performing analyses on many sets of data, with or without replication, one can estimate the MAE over that sample. But note that this is an overall error, and says nothing about the distribution of errors over the grid and its underlaying terrain.

Forecasters who ask about analysis errors are usually concerned about their specific area of interest, which may be rugged or not, be near water bodies or not, or be in sparse data regions or not. In this paper, we describe a method we developed to give an error estimate at specific grid points for each analysis and demonstrate it for surface temperature and dewpoint.

2. REGIONAL ERROR DEPENDENCIES

What is it that might cause analysis errors? Some prime candidates are discussed below.

2.1 *Data Density*

Obviously, if there are many data points relative to the spacing of the grid, the analysis will be better than if the data points are more sparse.

2.2 *Data Variability*

When data values are very nearly the same over some small region, say within a few gridlengths at most, then a grid point value should represent them very well, they should be highly recoverable, and the analysis error would be low. On the other hand, when there is high variability, one would not expect any particular value to be recoverable to a high degree of accuracy.

2.3 *Roughness of Terrain*

The values of most surface variables are influenced by the height of the terrain. For some variables, the affect of terrain can be anticipated. For instance, the observed temperature usually, but not always, decreases with altitude of the observing point within a local area. However, while the terrain may be a factor, its affect is not easy to anticipate for some variables. For example, cloud height above ground may well decrease with elevation in some restricted area, but cloud height of zero at an observing site may abruptly change to no clouds (clear or unlimited cloud height) above a certain elevation within the same local area.

2.4 *Land Use*

Land use may also affect analysis errors. For instance, as one goes from a grassy location to a sandy or rocky one, the value of the variable may well change.

2.5 *Land/Water Differences*

In addition to there being no terrain roughness over significant water bodies, the different surface itself may cause a difference in the analysis error.

2.6 *Synoptic Situation*

While the atmospheric stability and wind flow characteristics can undoubtedly affect analysis error, most of these affects should be captured in the data variability.

3. ERROR ESTIMATION PROCEDURE

As explained above, a mean error over a whole analysis area for a particular date and time can be obtained by withholding a few stations and replicating the analysis with different withheld stations. And, an average error over many dates and times can be obtained in the same way, perhaps without replication. But how does one get an estimate of analysis error for a particular date and time that varies realistically regionally, and do it efficiently in real time?

The approach we have taken is to develop two regression equations for each weather element (here, 2-m temperature and dewpoint), one for land locations and one for water, in which the predictors (independent variables) are functions of known data or constants, and the predictand (dependent variable) is the analysis error. To develop the regression equation, we have to compute the predictors and the predictand over a sample of sufficient size to yield stable equations.

3.1 *Computation of Predictand*

The predictand, the (absolute) error estimates at particular locations, is computed by making analyses over a large data set, randomly withholding a few stations from each analysis, and finding the absolute difference between the withheld station's value and the value interpolated from the analysis. This gives an AE at each withheld data point for each analysis. Specifically, we withheld 20 land stations per analysis and 1 water station, the latter from either ocean or Great Lakes buoys, or observing points judged to be more representative of a water location than land (see Im et al. 2010). The total number of land stations per analysis was on the order of 10,000, so the percentage withheld was about 0.2 percent. The total number of water points was on the order of 300 for temperature and 200 for dewpoint, so the percentage of withheld points was about 0.3 and 0.5 percent, respectively. Our sample consisted of all hours of the day for all days within a 5.5 month period June 3 to November 17, 2009.

3.2 *Computation of Predictors*

Nineteen potential predictors were computed for each withheld station, which incorporate data density, data variability, and terrain roughness. They are as follows:

Data Density

1. The distance to the closest station within 55 gridlengths.
2. The distance to the 2nd closest station within 55 gridlengths.

Data Variability

3. The data variability within a radius of 55 gridlengths of the station. Data variability is defined as the mean absolute difference between the data values and their mean. The withheld value itself is not included in the calculation.
4. Same as 3, except within 45 gridlengths.
5. Same as 3, except within 35 gridlengths.
6. Same as 3, except within 27 gridlengths.
7. Same as 3, except the vertical change with elevation (VCE) between the withheld and the other stations is applied (see Glahn et al. 2009 for a description of VCE).
8. Same as 7, except within 45 gridlengths.
9. Same as 7, except the distance between stations is weighted quadratically by the same weighting function used in the analysis (see Glahn et al. 2009).
10. Same as 9, except within 45 gridlengths.
11. Same as 9, except within 35 gridlengths.
12. Same as 9, except within 27 gridlengths.

Terrain Roughness

13. Roughness calculated on the grid centered on the grid point closest to the station within a radius of 8 gridlengths. Roughness is defined as the mean absolute difference between the terrain heights at the grid points and their mean.
14. Same as 13, except within 4 gridlengths.
15. Same as 13, except within 2 gridlengths.
16. Same as 13, except within 1 gridlength.

Data Density and Roughness

17. Absolute difference in elevation between the withheld station and its closest neighbor.

19 Product of 17 and 1.

Data Variability, Roughness, and Data Density

18. Absolute difference between the withheld station value and the value estimated from the closest station after applying the VCE calculated at the closest station and the elevation difference between the two.

For predictors 1-6, 9-12, and 18 dealing with gridlengths, the gridlengths quoted are for land; for water they are double those quoted.

Note there is nothing dealing with land use. We have not applied land use in any of the analysis procedures. It is likely any variation caused by land use is very localized, and is of a smaller scale than the analysis gridlength. Water/land differences are dealt with by having separate relationships for water and land.

4. RESULTS

Five and a half months of data have been processed, and a regression equation obtained by screening for land and for water. The screening process consists of choosing predictors in order according to their additional reduction of variance (RV) of the predictand (Lubin and Summerfield 1951). The development sample size for land was 69,100 and 3,455 for water. The values should be reasonably independent, at least spatially, and furnish stable equations, especially for a small number of predictors.

It became apparent that the best predictor by far is No. 18, which is the difference between the withheld station value and an estimate of it provided by its closest neighbor. This estimate includes the VCE procedure used in the analysis, which defaults to zero over water. No. 18 was selected first for all four equations and proved the bulk of the total RV.

4.1 *Land*

The means, standard deviations, and correlations with the predictand are given in Table 1 for temperature and dewpoint over land; Table 2 is the same, except for over water. Of these, with a 0.001 cutoff for additional RV, three predictors were chosen in order 18, 16, and 4 for temperature and 18, 5, and 15 for dewpoint. For temperature, the total RV was 0.505–about half the total variance--and the standard error was 2.36°F. For dewpoint, the total RV was a little less, 0.457, and the standard error was 2.50°F.

Table 1. The variable means and standard deviations in °F, and correlations with the predictand for temperature and dewpoint over land.

Variable No. (See above)	Temperature Mean	Temperature Std. Dev.	Temperature Correlation	Dewpoint Mean	Dewpoint Std. Dev.	Dewpoint Correlation
1	4.20	3.36	0.055	4.54	3.62	0.049
2	6.30	3.89	0.051	6.76	4.08	0.051
3	4.57	2.21	0.182	4.38	2.05	0.235
4	4.30	2.16	0.192	4.06	1.95	0.245
5	3.97	2.11	0.202	3.71	1.84	0.256
6	3.67	2.09	0.202	3.40	1.77	0.258
7	3.35	1.75	0.170	3.51	1.73	0.215
8	3.04	1.54	0.187	3.22	1.60	0.227
9	1.18	0.64	0.174	1.24	0.65	0.215
10	1.07	0.57	0.187	1.14	0.60	0.218
11	0.97	0.53	0.192	1.04	0.55	0.218
12	0.90	0.53	0.185	0.97	0.55	0.209
13	104.04	113.31	0.184	104.63	114.03	0.167
14	77.62	93.99	0.189	77.09	93.48	0.168
15	54.36	71.81	0.192	53.20	70.20	0.160
16	39.41	56.43	0.191	38.41	55.21	0.149
17	129.88	231.03	0.180	126.20	225.08	0.136
18	2.97	4.13	0.703	3.36	4.24	0.664
19	660.11	1477.49	0.146	747.41	1789.39	0.117
Predictand	2.50	3.35		2.77	3.39	

Table 2. Similar to Table 1 for over water. Predictors involving the lapse rate, Nos. 7 and 8, 13 through 17, and 19, do not exist over water because the elevation does not vary.

Variable No. (See above)	Temperature Mean	Temperature Std. Dev.	Temperature Correlation	Dewpoint Mean	Dewpoint Std. Dev.	Dewpoint Correlation
1	12.96	8.59	0.045	16.36	13.96	0.068
2	18.21	10.60	0.025	26.98	20.51	0.067
3	3.45	1.73	0.164	3.11	1.75	0.134
4	3.22	1.72	0.172	2.91	1.73	0.167
5	2.93	1.97	0.142	2.66	1.63	0.180
6	2.63	1.82	0.166	2.39	1.59	0.175
9	1.63	0.94	0.126	1.45	0.95	0.161
10	1.47	0.92	0.135	1.13	0.93	0.157
11	1.30	0.96	0.128	1.17	0.89	0.188
12	1.16	0.95	0.135	1.07	0.91	0.196
18	2.79	4.65	0.504	3.33	3.50	0.647
Predictand	2.16	2.69		3.01	2.99	

Both coefficients and the mean and range of the variable itself have to be considered in assessing the influence of a predictor on the error. From Table 3, we see that if all three predictors had a value of zero, not likely, but not impossible, the estimated temperature error would be only 0.35°F; this is the lower limit for the temperature error estimate. If each predictor had its mean value, the error estimate would be 2.50°F. If in addition, each predictor differed from its mean by one standard deviation in a positive direction, the error would be another 2.71°F, for a total of 5.21°F.

For dewpoint (see Table 4), the minimum error estimate is 0.25°F, and the estimate if each predictor had its mean value is 2.77°F, not too dissimilar from temperature. Both temperature and dewpoint equations have a terrain term (No. 15 and 16), a data variability term (Nos. 4 and 5), and No. 18, which embodies data density, roughness, and data variability.

4.2 ***Water***

Over water, there was only one predictor kept for temperature, No. 18, and two for dewpoint, Nos. 18 and 2. Screening actually selected three more for temperature, but the coefficient was negative for the second one. The predictors were devised so that each one should contribute positively to the error estimate; a negative coefficient could easily give inconsistent results, even a negative absolute error.

Tables 5 and 6 show that the minimum temperature and dewpoint estimates are somewhat larger over water than over land, being 1.34 and 0.90°F, respectively. Data over water are much more sparse than over land, but the variability is less. The effect of No. 18 is less for temperature than for dewpoint, likely because the data density for dewpoint is less than for temperature.

5. IMPLEMENTATION

The equations were developed for stations–points where we had data. To implement, we could compute the estimated error for a particular time at each station where there is data. For instance, for the temperature/land equation, we could compute the absolute difference between the station's value and the value estimated by the closest station, taking into account the VCE (predictor No. 18), compute the roughness (predictor No. 16), and compute the data variability (predictor No. 4). These values can be used with the equation constant and coefficients to compute the error. However, this does not give values on a grid–what we really want. We could analyze these values with the BCDG analysis method, but that would give questionable error values in the same areas where we had a questionable analysis. This doesn't seem to be an acceptable solution.

Alternatively, we can, with some reasonable assumptions, apply the equation at grid points. We do it in the following manner. Predictor No. 18 is calculated by finding the absolute value of the difference of the analysis value at a grid point and the value for that grid point estimated by the closest station, taking into account the VCE at the station. The roughness (predictor No. 16) can be calculated

Table 3. The constant and coefficients, means, and standard deviations for the three predictor variables in the temperature equation over land, together with the predictor contributions to the total estimate. Units are °F.

Variable No. (See above)	Coefficient (Constant)	Mean	Contribution from Mean and Constant	Std. Dev (sd.)	Contribution from 1 sd.
Constant	0.3472		0.347		
18	0.5564	2.972	1.653	4.126	2.296
16	0.0046	39.411	0.181	56.432	0.260
4	0.0733	4.298	0.315	2.163	0.158
Total			2.497		2.714

Table 4. Same as Table 3 except for dewpoint equation.

Variable No. (See above)	Coefficient (Constant)	Mean	Contribution from Mean and Constant	Std. Dev (sd.)	Contribution from 1 sd.
Constant	0.2498		0.250		
18	0.5112	3.363	1.719	4.236	2.165
5	0.1855	3.711	0.688	1.844	0.342
15	0.0021	53.204	0.112	70.199	0.147
Total			2.769		2.655

Table 5. The constant and the coefficient, mean, and standard deviation for the one variable in the temperature equation over water, together with the predictor contributions to the total estimate. Units are °F.

Variable No. (See above)	Coefficient (Constant)	Mean	Contribution from Mean and Constant	Std. Dev (sd.)	Contribution from 1 sd.
Constant	1.3443		1.344		
18	0.2918	2.794	0.815	4.652	1.357
Total			2.159		1.357

Table 6. Same as Table 5, except for the 2-predictor dewpoint equation.

Variable No. (See above)	Coefficient (Constant)	Mean	Contribution from Mean and Constant	Std. Dev (sd.)	Contribution from 1 sd
Constant	0.9004		0.900		
18	0.5522	3.325	1.836	3.501	1.933
2	0.0101	26.980	0.272	20.515	0.207
Total			3.008		2.140

at each grid point. Also, the data density at the grid point can be calculated in the same manner it was calculated at stations in the development.

Is implementation at grid points substantially different from implementation at stations? The data density calculation should not suffer. The number of stations within the specified radius (here, a substantial 55 gridlengths) will vary whether the calculation is at stations or at grid points. The roughness calculation is done at grid points in development (see Table 1), so there is no difference there. The major difference is for Predictor No. 18; in development the value at the station was known (observed), but in implementation the value is the analysis value.

The fact that predictor No. 18 is calculated at grid points in implementation and at stations in development may cause a low bias in the estimates for grid points. Given that the density of grid points, in our application, is greater than the density of stations, the distance between a station and its closest neighbor will be, in general, greater than the distance between a grid point and its closest station. This may tend to underestimate the value of predictor No. 18, and the error at grid points, compared to errors calculated at stations.

Temperature and dewpoint analyses are shown in Figs. 1 and 3 (see also Im et al. 2010) and the corresponding error maps in Figs. 2 and 4. The error maps show many errors in the eastern and central part of the U.S. are < 1.0°F, as might be expected from Table 3, and only isolated spots where the errors are > 4.0°F. In areas where the terrain is flat (terrain not shown), the larger errors are undoubtedly due to greater data variability. Examples of correspondence between the analysis "detail" and error can be found. For instance, for temperature in extreme eastern Texas, temperatures of 70 to 80°F poke into an area more generally ~ 90°F. This shows up as spots and a crescent on the error map. For dewpoint, a spot in southwestern Indiana shows up.

The largest temperature errors, for these analyses, are along the western seacoast and the nearby mountains. There is a sharp temperature contrast near the coast. Data are fairly dense, but the roughness is pronounced and the data variability is high. On the other hand, the largest dewpoint errors are associated with terrain in the West and the error along the coast is not particularly high, indicating the less variable dewpoint there. The high dewpoint errors in high terrain, as compared to temperature, is likely due to the more consistent change of temperature with elevation than dewpoint.

6. DISCUSSION AND CONCLUSIONS

A method to estimate the errors associated with the BCDG analysis of temperature and dewpoint has been developed and demonstrated. It should be recognized, any estimate of analysis error is just that–an estimate. The truth cannot be known (the values of the element being dealt with at each grid point) unless some data set is fabricated at both grid points (ground truth) and quasi-random points (data to analyze) with an analytic function. This fabrication route has been taken in analysis studies (e.g., Smith and Leslie 1984 and Goodin et al. 1979), but it is difficult to devise an analytic function that simulates the real world with elevation differences, data with unknown errors, and data densities that are variable and reasonable.

This method, which we call BCDGE (BCDG Error), furnishes an estimate of error which is physically reasonable, is specific to the data set being analyzed, and is relatively easy to implement. To emphasize a previous point, the "error" used in the development included the interpolation error–the estimate of the station value from the regularly spaced grid. This itself, can be a considerable cause for error, especially in the West. It is also recognized that the development can be carried out only for the elevations where there are stations. For higher elevations, the estimated errors are essentially extrapolations from stations at lower elevations with similar terrain roughness and data density. This is also true of the analysis; the true values at high elevations are not known.

The error maps look reasonable in terms of pattern, and also in terms of absolute value, although there is no way to know how close the estimates really are. It is believed these error maps will help pinpoint where the problems are with the associated analyses.

7. REFERENCES

Achtemeier, G. L., 1989: Modification of a successive corrections objective analysis for improved derivative calculations. *Mon. Wea. Rev*., **117**, 78-86.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis*. J. Appl. Meteor*., **3,** Amer. Meteor. Soc., 396-409.

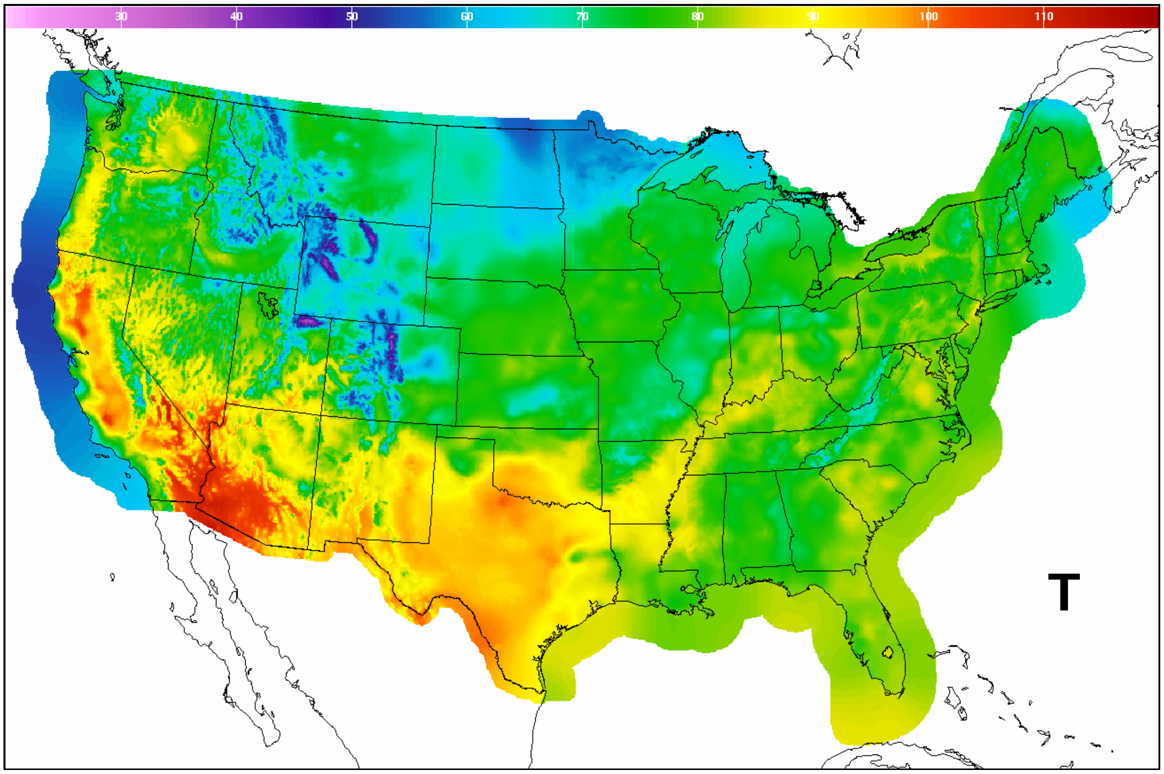

FIG. 1. BCDG analysis of temperature (°F) produced for 0000 UTC 18 August 2009.

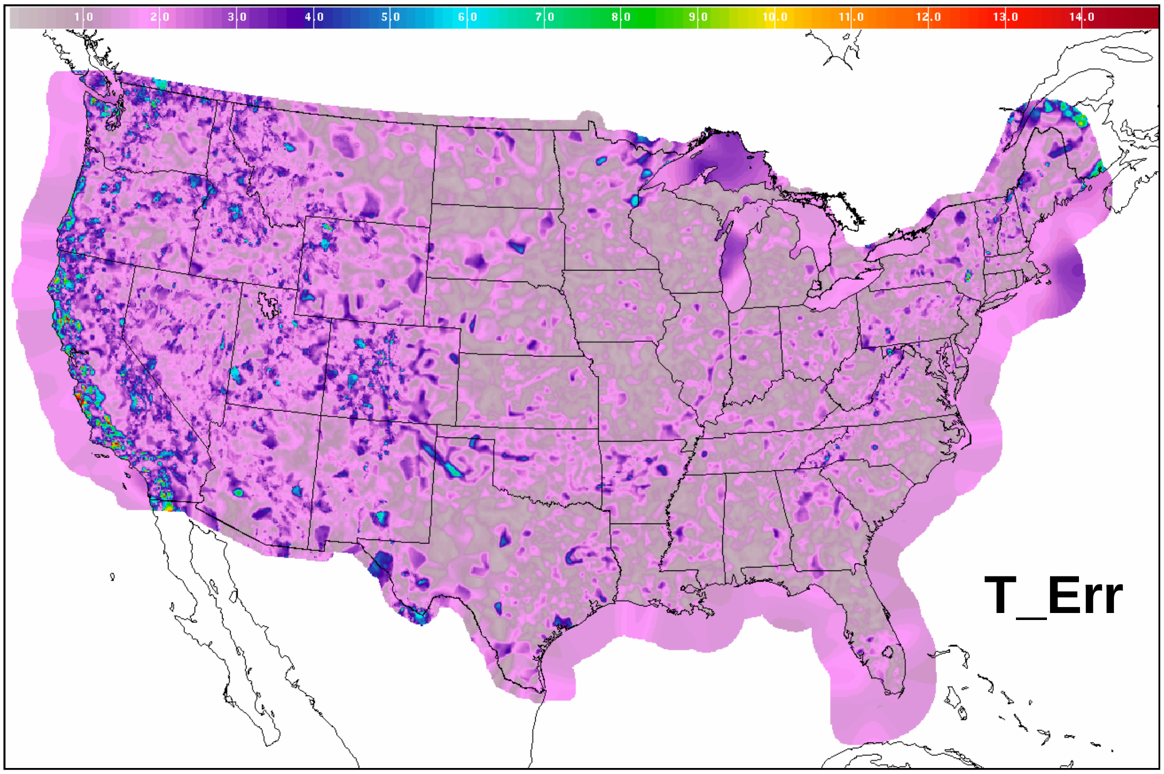

FIG. 2. Error estimation (°F) of the BCDG analysis for 0000 UTC 18 August 2009.

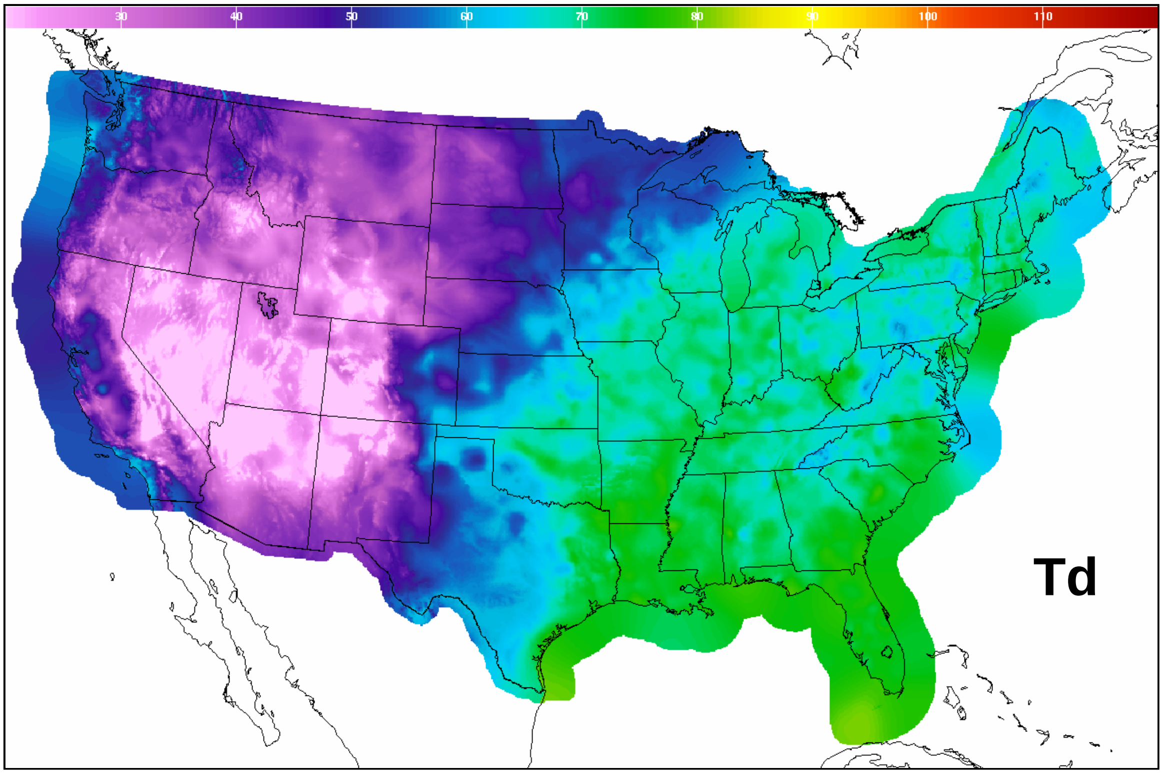

FIG. 3. Same as Fig. 1 except for dewpoint.

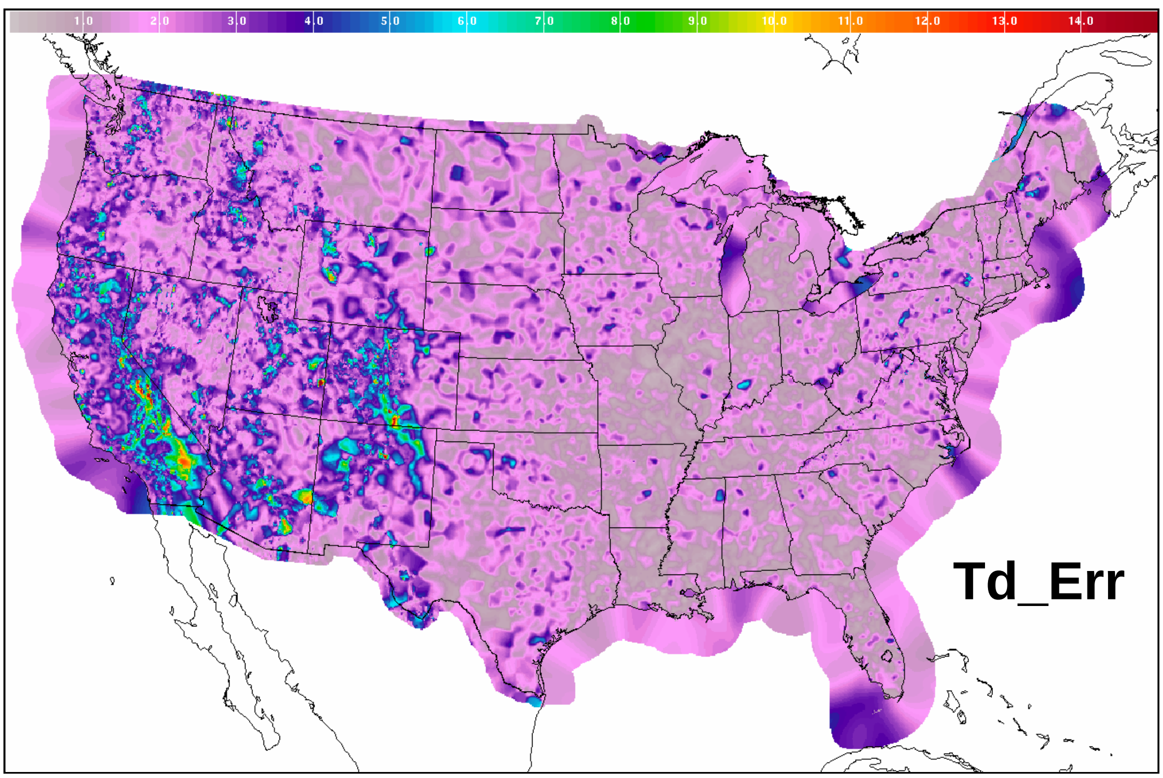

FIG. 4. Same as Fig. 2 except for dewpoint.

Bergthorssen, P., and B. R. Doos, 1955: Numerical weather map analysis. *Tellus*, **7**, 329-340.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.*, **87**, 367-374.

Fritsch, J. M., 1971: Objective analysis of a two-dimensional data field by the cubic spline techniques. *Mon. Wea. Rev*., **99**, 379-386.

Gilcrest, B., and G. P. Cressman, 1954: An experiment in objective analysis. *Tellus*, **6**, 309-318.

Glahn, B., K. Gilbert, R. Cosgrove, D. P. Ruth, and K. Sheets, 2009: The gridding of MOS. *Wea. Forecasting*., **24**, 520-529.

Glahn, H. R., 1981: Comments on "A comparison of interpolation methods for sparse data: Application to wind and concentration fields." *J. Appl. Meteor*, **20**, Amer. Meteor. Soc., 88-91.

———, 1987: Comments on "Error determination of a successive correction type objective analysis scheme." *J. Appl. Meteor*. **2**, 348-350.

———, and J. E. McDonell, 1971: Comments on "Objective analysis of a two-dimensional data field by the cubic spline techniques. *Mon. Wea. Rev*., **99**, 977-978.

Goodin, W. R., G. J. McRae, and J. H. Seinfeld, 1979: A comparison of interpolation methods for sparse data: Application to wind and concentration fields. *J. Appl. Meteor*, **18**, Amer. Meteor. Soc., 761-771.

———, 1981: Reply. *J. Appl. Meteor*, **20**, Amer. Meteor. Soc., 92-94.

Im, J.-S., B. Glahn, and J. Ghirardelli, 2010: Real-time objective analysis of surface data at the Meteorological Development Laboratory. Preprints, *20th Conf. on Probability and Statistics in the Atmospheric Sciences,* Atlanta, GA, Amer. Meteor. Soc., 219.

Lubin, A., and A. Summerfield, 1951: A square root method of selecting a minimum set of variables in multiple regression: I. The method. *Physchometrika*, **16**, 271-284.

Panofsky, H. A., 1949: Objective weather map analysis. *J. of Meteor*. **6**, Amer. Meteor. Soc., 386-392.

Smith, D. R., and F. W. Leslie, 1984: Error determination of a successive correction type objective analysis scheme. *J. Appl. Meteor*, **1**, 120-130.

Thomasell, A., Jr., 1962: The areal-mean-error method of analysis verification. The Travelers Research center, Inc., Contract FAA/BRD-363, *Tech. Pub*. No. **5**, Hartford, Connecticut, 18 pp.