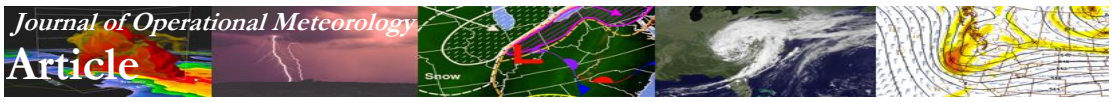
Glahn, B., and J.-S. Im, 2013: Error estimation of objective analysis of surface observations. *J. Operational Meteor.*, **1** (11), 114–127, doi: http://dx.doi.org/10.15191/nwajom.2013.0111.



Error Estimation of Objective Analysis of Surface Observations

BOB GLAHN and JUNG-SUN IM
Meteorological Development Laboratory, National Weather Service, Silver Spring, Maryland

(Manuscript received 26 March 2013; in final form 8 July 2013)

ABSTRACT

As part of the Localized Aviation Model Output Statistics (MOS) Program, the Meteorological Development Laboratory is analyzing surface data reports on an hourly basis. The Bergthórsson-Cressman-Döös-Glahn analysis program that is being used for gridding the MOS forecasts has been tailored to analyze surface observations. These analyses are available in the National Digital Guidance Database (NDGD) over the conterminous United States. This database is on the same grid as, and is interoperable with, the National Digital Forecast Database. It is desired to know the errors involved in these analyses. Whereas the actual errors are unknowable for several reasons, they can be estimated. On any given analysis, one would expect the error at a specific location to be a function of some knowable parameters, such as distances between the reporting locations and the grid points, the terrain roughness, the density of reporting locations, and the variability of the data values—all in the immediate vicinity.

We have made analyses of surface temperature and dewpoint over the conterminous United States every fifth hour for one year. On each analysis, 20 land stations and one water station were randomly withheld from the available data. For each withheld datum, the analysis value at that site was estimated by bilinear interpolation. The differences between these interpolated values and the actual observations were related to knowable parameters through least squares regression—one relationship for land and another for water—for each variable. These regression equations were then applied, respectively, to each land and water grid point for a specific hour. This produced an estimate of the error of the analysis for each grid point for that specific time. These grids of errors are, along with the analyses, available in the NDGD. This paper describes the process and shows the results.

1. Introduction

Analyses of surface-based meteorological observations have many applications. Such analyses are part of the assessment of the synoptic situation necessary for weather forecasting; this is as true today as it was 60 yr ago. The methods of analyses have, however, changed since then. In the mid 1950s, data were plotted by hand on maps and the analyses were performed by humans. Since then, various methods of automated analyses have been developed in concert with the development of the digital computer. Summarizing a mesoscale conference, Horel and Colman (2005) stated, "Thus, the NWS [National Weather Service] has an immediate and critical need to produce real-time and retrospective analyses at both a high spatial and temporal resolution in order to create the NDFD [National Digital Forecast Database] forecasts as well as to verify their accuracy." In addition, they also stated that additional research will be needed on how to quantify and express to the end user the uncertainty implicit in any analysis approach.

Objective (i.e., computer produced) analyses of meteorological data were being considered even before 1950. One of the first objective techniques to appear was the least-squares fitting of a polynomial to the data over a fairly large area (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, who was the director of the Joint Numerical Weather Prediction Unit of the Weather Bureau, the forerunner of the National Meteorological Center (now

Corresponding author address: Bob Glahn, Meteorological Development Laboratory, National Weather Service, 1325 East-West Highway, Silver Spring, MD 20910
E-mail: harry.glahn@noaa.gov

the National Centers for Environmental Prediction, NCEP), recognized the potential for a technique developed by Bergthórsson and Döös (1955), and put a version of that into operation for analyzing upper-air geopotential heights (Cressman 1959). This successive-correction technique consisted of making multiple passes over the data and correcting each grid point on each pass by the data in the immediate vicinity—immediate vicinity being defined by a radius of influence that was constant over the analysis area, but was decreased for successive passes. A very similar technique, basically differing only in the distance-weighting factor, was proposed by Barnes (1964) and has been used extensively. Barnes (1964) proposed his method be used with one or two passes. He suggested "…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.

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 (e.g., Kalnay 2003). These latter methods employ relationships among free atmospheric variables that are not as effective for use with variables observed at the earth's surface, although they have been adapted for surface variables and are used for the real-time mesoscale analysis (RTMA; Pondeca and Manikin 2009; Manikin and De Pondeca 2011) being run at NCEP. Numerical weather prediction centers in other countries have analysis systems tailored for their needs; for instance, Glowacki et al. (2012) presented a stage in the development of a system for Australia and reference other systems.

The method as used by Cressman has been used extensively in the Localized Aviation Model Output Statistics (MOS) Program (LAMP) and has been called the BCD method (see Glahn et al. 1985). As part of the Meteorological Development Laboratory's support to the aviation community, and the Next Generation Air Traffic Control System Program in particular (see Ghirardelli and Glahn 2010, 2011), we have further developed BCD, which now is called BCDG, after the initials of the last names of the primary developers (Bergthórsson, Cressman, Döös, and Glahn). This method has been described by Glahn et al. (2009) for the analysis of MOS forecasts and has been adapted for analysis of surface observations. Although the basic BCDG method is described in Glahn et al. (2009), there were several changes and improvements necessary for analysis of observations. These are described in Im et al. (2010, 2011), Glahn and Im (2011), and Im and Glahn (2012).

Although a considerable amount of effort has been placed on analysis methods, a more modest effort has been put 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. Difficulties in making a good analysis and a good estimate of its error are evident by the exchanges between Fritsch (1971) and Glahn and McDonell (1971), Goodin et al. (1979, 1981) and Glahn (1981), and Smith and Leslie (1984) and Glahn (1987). The RTMA process (De Pondeca et al. 2011) includes an estimate of the analysis error—the method being specific to the analysis process. Myrick and Horel (2008) have studied the sensitivity of surface analyses to a particular type of observation.

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 nondescript areas. If such were the case, then a method that concentrated on that aspect would need to be defined. If one is concerned about making derivative calculations, then the the analysis method proposed by Achtemeier (1989) would be an option, with errors defined appropriately. Our use of the term "analysis error" is defined as a measure of the inability to recover the data 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). Even though the definition is in terms of values at data points (that is where the error can be measured), it represents the difference between the true value (the value that would have been observed) and the analysis at *any* point. Note that this does not address the possible errors in the observations except to the extent they manifest themselves in the analysis.

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 was perfect, one would still not, in general,

recover the value at the data point exactly by interpolation.[1] 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 relative to the grid spacing is uneven and/or sparse. This error of interpolation is unavoidable, although there are different methods of interpolation that could be used.[2] The second, and major, difficulty is that any good analysis process can fit the data points rather closely, but still be poor where the data are sparse; a calculation of error only at the data points may not well represent the error over the entire grid. In addition, this gives errors only at the data points, not at grid points.

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 would be a measure of the overall error at points on the grid between grid points where there were no data values. Although 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—and 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. Tyndall and Horel (2013, p. 256) used the adjoint of a two-dimensional variational surface analysis to identify high-impact observations and reference several cross-validation experiments.

[1] Suppose an analytic function were defined that could be evaluated at any point in the analysis area. Knowing the gridpoint values does not equate to knowing the values between grid points. Interpolation to a point from gridpoint values will give an estimate that will not, in general, be the same as the analytic function evaluated at that point.

[2] Biquadratic interpolation would work well in smoothly varying fields of data. Smooth fields could result from the data themselves not exhibiting much variablity, or the analysis process highly smoothing the data. For more variable data fields, bi-linear interpolation is probably best.

Forecasters who ask about analysis errors usually are concerned about their specific area of interest, which may or may not be in rugged terrain, near water bodies, or in data-sparse regions. This paper describes a process for generating an error map for a particular analysis and gives results for surface temperature and dewpoint over the conterminous United States (CONUS) on a 2.5-km grid. The grid is in relation to a Lambert conformal projection and is the one used in the NDFD (Glahn and Ruth 2003). Whereas the method is designed for BCDG, it could be used equally well for other analysis methods.

2. The error estimation method

As stated previously, our measure of error is the difference between a data value and the value obtained from the gridpoint analyses at that point by linear interpolation. For a particular meteorological variable, we make analyses over many observation times, randomly withholding a very small percentage of the observations. A particular analysis over the CONUS and its immediate water surroundings will contain over 10 000 observations. We withhold 20 land reports and one water report, and compute the AE for each location and analysis. These values become a predictand dataset for regression analysis.

For each of the withheld data points, we compute a set of predictors that might be related to the analysis errors. Then, two regression equations are computed by forward selection of predictors: one for land locations and one for water locations. The developed equations apply to observation points, but with certain assumptions they can be applied to each appropriate (land or water) grid point for a specific case—thereby giving a grid of estimated errors for that observation time.

3. Description of predictors

There are several factors that may cause errors at specific locations, which are discussed below. In all, 19 predictors were calculated; they are described and numbered in the following subsections, and summarized in Table 1.

a. 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—all other things

Table 1. Definitions for nineteen potential predictors. Those marked with an asterisk were the only ones used for water. Predictors 9–12 are used for water, but the *VCE* is zero. One grid length equals 2.5 km.

Predictor category	Predictor No.	Definition	Within-grid lengths
Data density	1*	Distance to the closest station	110
Data density	2*	Distance to the 2nd closest station	110
Data variability	3*	Data variability	110
Data variability	4*	Data variability	90
Data variability	5*	Data variability	70
Data variability	6*	Data variability	54
Data variability	7	Data variability with application of the *VCE*	110
Data variability	8	Same as previous	90
Data variability	9*	Data variability with application of the *VCE* and use of the distance between stations weighted quadratically by the same weighting function used in the analysis	110
Data variability	10*	Same as previous	90
Data variability	11*	Same as previous	70
Data variability	12*	Same as previous	54
Terrain roughness	13	Roughness	8
Terrain roughness	14	Roughness	4
Terrain roughness	15	Roughness	2
Terrain roughness	16	Roughness	1
Data density and roughness	17	Absolute difference in elevation between the withheld station and its nearest neighbor	110
Data density and roughness	18	Product of 17 and 1	110
Data variability, roughness, and data density	19*	Absolute difference in value between the station value and its estimate from the closest station	110

being equal. We computed two predictors related to just data density, with respect to the grid spacing. They are:

1. The number of grid distances from the withheld station to the closest station within 110 grid lengths.
2. The number of grid distances from the withheld station to the 2nd closest station within 110 grid lengths.

The choice of 110 grid lengths was based on the minimum data density so that the first predictor could always be computed. If the second could not be computed (i.e., there were not two stations within 110 grid lengths of the withheld station), that case was omitted; this rarely happened.

b. Data variability

When data values are very nearly the same over some small region, say within a few grid lengths at most, then (i) a gridpoint value should represent them very well, (ii) they should be highly recoverable, and (iii) 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. Several potential predictors related to data density were computed as indicated below.

3. The data variability within a circle with a radius of 110 grid lengths of the station. Data variability is defined as the mean absolute difference between the data value and the mean of the values within the circle. The withheld value itself is not included in the calculation of the mean.
4. Same as 3, except within 90 grid lengths.
5. Same as 3, except within 70 grid lengths.
6. Same as 3, except within 54 grid lengths.
7. Same as 3, except the vertical change with elevation (*VCE*) of the variable being analyzed between the withheld and the other stations is applied [the *VCE* is explained in Glahn et al. (2009), but because of its importance here, it is described in appendix A].
8. Same as 7, except within 90 grid lengths.
9. Same as 7, except the distance between stations is weighted quadratically by the same weighting function used in the analysis (Glahn et al. 2009).
10. Same as 9, except within 90 grid lengths.
11. Same as 9, except within 70 grid lengths.
12. Same as 9, except within 54 grid lengths.

c. Roughness of terrain and other factors

The values of most surface variables are influenced by the height of the terrain. Several potential predictors were calculated; some of them also are related to factors discussed above.

13. Roughness calculated on the grid centered on the grid point closest to the station within a radius of 8 grid lengths. Roughness is defined as the mean absolute difference between the terrain height at the grid point and the mean of the surrounding heights.
14. Same as 13, except within 4 grid lengths.
15. Same as 13, except within 2 grid lengths.
16. Same as 13, except within 1 grid length.
17. Absolute difference in elevation between the withheld station and its closest neighbor (combines data density and roughness).
18. Product of 17 and 1 (combines data density and roughness).
19. Absolute difference between the withheld station value and the value estimated from the closest station after applying the *VCE* of the analyzed variable calculated at the closest station and with the elevation difference between the two (combines data density, roughness, and data variability).

d. Other considerations

Land use also may affect analysis errors. For instance, as one goes from a grassy location to a sandy or rocky one, the value of the variable may change. We have not used predictors dealing with land use. It is likely any variation caused by land use is very localized, and is of a smaller scale than the analysis grid length.

In addition to there being no terrain roughness over water bodies, the different surface itself may cause a difference in the analysis error. As noted above, different relationships are developed for land and for water, so specific predictors are not necessary.

While the atmospheric stability and wind-flow characteristics undoubtedly can affect analysis error, most of these effects should be captured in the data variability and roughness of terrain. It also was desired to keep the analysis error estimation confined to the surface data themselves, so no other predictors were calculated.

4. Computation of the predictand

The predictand, which is the absolute-error estimate at particular locations, is computed by making analyses over a large dataset, 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. [The method of random selection is explained in appendix B.] This gives an AE at each withheld data point for each analysis. Specifically, we withheld 20 land stations per analysis and one water station, the latter from either the ocean or Great Lakes' buoys, or from observing points judged to be more representative of a water location than land.[3] The total number of land stations per analysis was on the order of 10 000, so the percentage withheld was about 0.2%. 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%, respectively. Our sample consisted of every fifth hour for all days within a 1-yr period from 2100 UTC 3 June 2009 to 2300 UTC 31 May 2010.[4]

5. Computation of the predictors

The 19 potential predictors discussed above were computed for each withheld station. The predictors incorporate data density, terrain roughness, and data variability, and are summarized in Table 1. Predictors dealing with elevation differences are not computed for over water. The grid lengths quoted are for land; the grid lengths for water are double those for land.

6. Results

One year of data was processed, and a regression equation was obtained for land and for water by screening the 19 predictors for land and 11 for water. The screening process consists of choosing predictors in order according to their additional reduction of variance (RV) of the predictand (Summerfield and Lubin 1951; Murphy and Katz 1985, chapter 8). The development sample size for land was 30 100 for temperature and 30 060 for dewpoint; the sample size for water was 1503 for temperature and 1508 for dewpoint. These values should be reasonably independent and furnish stable equations, especially for a small number of predictors.

It became apparent that the best predictor by far is 19, which is the difference between the withheld station value and an estimate of it provided by its

[3] Some coastal stations, while not over water, may be more representative of water than land. See Im and Glahn (2012) for a discussion of this issue.

[4] Using every fifth hour of the hourly sequence reduces redundancy, guarantees an even distribution of hours, and provides an adequate sample for stable regresson equations.

Table 2. The predictor variable means and standard deviations in °F (divide by 1.8 to get °C), and correlations with the predictand, for temperature and dewpoint over land.

Predictor No. (see Table 1)	Temperature			Dewpoint		
	Mean	Std. Dev.	Correlation	Mean	Std. Dev.	Correlation
1	8.49	6.64	0.049	9.11	7.22	0.055
2	12.70	7.82	0.054	13.66	8.30	0.057
3	4.64	2.10	0.158	4.53	2.06	0.232
4	4.33	2.04	0.170	4.18	1.95	0.242
5	3.97	1.98	0.182	3.81	1.86	0.255
6	3.64	1.96	0.180	3.47	1.78	0.253
7	3.42	1.67	0.156	3.60	1.74	0.218
8	3.10	1.49	0.173	3.29	1.60	0.234
9	1.19	0.61	0.160	1.27	0.65	0.216
10	1.08	0.55	0.172	1.16	0.59	0.222
11	0.98	0.52	0.180	1.06	0.55	0.223
12	0.91	0.53	0.179	0.99	0.56	0.210
13	78.31	95.46	0.146	76.82	93.54	0.155
14	54.23	72.60	0.148	52.54	70.09	0.149
15	36.31	52.86	0.150	34.94	51.35	0.138
16	26.06	41.30	0.142	24.96	40.81	0.123
17	128.48	227.48	0.149	125.75	221.06	0.140
18	1310.01	2817.16	0.118	1483.99	3385.04	0.121
19	2.97	4.36	0.726	3.36	4.23	0.655
Predictand	2.48	3.60	1.000	2.77	3.40	1.000

closest neighbor. This estimate includes the *VCE* procedure used over land in the analysis. Predictor 19 was selected first for all four equations (land/water, temperature/dewpoint) and provided the bulk of the total RV.

a. Land equations

The means, standard deviations, and correlations with the predictand are given in Table 2 for temperature and dewpoint over land. Of these, with a 0.1% cutoff for additional RV, three predictors were chosen in order of 19, 14, and 5 for temperature and 19, 5, and 14 for dewpoint.[5] For temperature, the total RV was 53.4% (about half the total variance) and the standard error was 2.46°F (1.37°C). For dewpoint, the total RV was less, 44.6%, and the standard error was 2.53°F (1.41°C).

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 0.29°F (0.16°C); this is the lower limit for the temperature error estimate, and includes the interpolation error. If each predictor had its mean value, the error estimate would be 2.48°F (1.38°C). If in addition, each predictor differed from its mean by one standard deviation *in a positive direction*, the error would be another 2.91°F (1.62°C), for a total of 5.39°F (2.99°C).

For dewpoint (see Table 4), the minimum error estimate is 0.20°F (0.11°C), and the estimate if each predictor had its mean value is 2.77°F (1.54°C). In addition, if each predictor differed from its mean by one standard deviation *in a positive direction*, the error would be another 2.65°F (1.47°C), for a total of 5.42°F (3.01°C)—very nearly the same as for temperature. Both temperature and dewpoint equations have a terrain roughness term (14), a data variability term (5), and a term that embodies data density, roughness, and

[5] Screening actually selected three more predictors for temperature, but the coefficient was negative for the fourth one. The predictors were devised so that each one should logically contribute positively to the error estimate; a negative coefficient could easily give inconsistent results (e.g., a negative absolute error). Negative coefficients can be caused by near multicolinearity among predictors, which occurs along with extremely low additional reductions of variance as additional variables are added to the equation.

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. Total reduction of variance (RV) for the 3-predictor equation is 53.4%. Units are °F (divide by 1.8 to get °C).

Predictor No. (see Table 1)	Coefficient (constant)	Mean	Contribution from mean and constant	Std. Dev.	Contribution from one std. dev.
Constant	0.2913		0.291		
19	0.5891	2.974	1.752	4.364	2.571
14	0.0026	54.234	0.141	72.595	0.189
5	0.0755	3.966	0.299	1.980	0.149
Sum			**2.483**		**2.909**

Table 4. Same as Table 3 except for the 3-predictor dewpoint equation. Total RV = 44.6%.

Predictor No. (see Table 1)	Coefficient (constant)	Mean	Contribution from mean and constant	Std. Dev.	Contribution from one std. dev.
Constant	0.2009		0.201		
19	0.5048	3.363	1.698	4.229	2.135
5	0.2033	3.814	0.775	1.859	0.378
14	0.0019	52.539	0.100	70.090	0.133
Sum			**2.774**		**2.646**

data variability (19). The only difference is that the order of selection for predictors 5 and 14 was reversed.

The correlations in Table 2 indicate the error is more linearly related to data variability for dewpoint than for temperature (predictors 3–12). Also, it does not matter much over what area the variability is calculated. Temperature and dewpoint are about equally related to roughness (predictors 13–16), mattering little over what area they are calculated. Predictor 19, being by far the best, indicates the importance of the *VCE* in the analysis process. It also may indicate that if a better *VCE* process could be found, the error would decrease and therefore the analysis would be better—an obvious conclusion from logic alone.

b. Water equations

Table 5 is similar to Table 2, and Tables 6 and 7 are similar to Tables 3 and 4, respectively, except they are for over water. There were only two predictors retained for temperature over water (19 and 3), and three for dewpoint (19, 2, and 5).

Tables 6 and 7 show that the minimum temperature and dewpoint error estimates are somewhat larger over water than over land, being 0.65°F (0.36°C) and 0.75°F (0.42°C), respectively. The contribution to error from the constant and means of the temperature and dewpoint are 2.29°F (1.27°C) and 3.58°F (1.99°C), respectively. If each predictor is different from the mean by one standard deviation *in a positive direction*, the total error is 4.17°F (2.32°C) and 6.60°F (3.67°C), for temperature and dewpoint, respectively. Data over water are much more sparse than over land, but the spatial variability is less.[6]

7. Implementation

The equations were developed for stations–points where we had data. To implement the equations, we could compute the estimated error for a particular time at each station where there are data. For instance, for the temperature/land equation we could (i) compute the absolute difference between the station's value and the value estimated by the closest station, taking into account the *VCE* (predictor 19), (ii) compute the roughness (predictor 14), and (iii) compute the data variability (predictor 5). These values can be used with the equation constant and coefficients to compute the error. However, this does not give values on a grid, which is 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 does not seem to be an acceptable solution.

Alternatively, we can, with some reasonable assumptions, apply the appropriate equations at grid points. We do it in the following manner. Predictor 19 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

[6] As with temperature over land, more predictors were selected, but the next one selected had a negative coefficient.

Table 5. Similar to Table 2 but for over water. Predictors involving the *VCE* and terrain roughness (7, 8, and 13–18) do not exist over water because the elevation does not vary.

Predictor No. (see Table 1)	Temperature			Dewpoint		
	Mean	Std. Dev.	Correlation	Mean	Std. Dev.	Correlation
1	27.00	16.99	0.042	32.38	27.78	0.050
2	38.61	21.44	0.003	54.43	39.86	0.065
3	3.73	1.77	0.218	3.73	2.15	0.143
4	3.44	1.71	0.217	3.54	2.14	0.184
5	3.14	1.71	0.203	3.30	2.39	0.216
6	2.76	1.77	0.191	2.88	2.22	0.170
9	1.73	0.92	0.177	1.76	1.14	0.148
10	1.55	0.90	0.178	1.59	1.08	0.160
11	1.35	0.87	0.176	1.45	1.26	0.184
12	1.19	0.89	0.172	1.30	1.27	0.187
19	2.98	4.96	0.518	4.00	4.59	0.701
Predictand	2.29	3.13	1.000	3.58	3.79	1.000

Table 6. The constant and coefficients, means, and standard deviations for the two predictor variables in the temperature equation over water, together with the predictor contributions to the total estimate. Total RV for the 2-predictor equation is 27.9%. Units are °F (divide by 1.8 to get °C).

Predictor No. (see Table 1)	Coefficient (constant)	Mean	Contribution from mean and constant	Std. Dev.	Contribution from one std. dev.
Constant	0.6466		0.647		
19	0.3121	2.983	0.931	4.956	1.547
3	0.1898	3.729	0.708	1.773	0.337
Sum			**2.286**		**1.884**

Table 7. Same as Table 6, except for the 3-predictor dewpoint equation. Total RV = 49.7%.

Predictor No. (see Table 1)	Coefficient (constant)	Mean	Contribution from mean and constant	Std. Dev.	Contribution from one std. dev.
Constant	0.7492		0.749		
19	0.5685	4.003	2.276	4.587	2.608
2	0.0059	54.429	0.321	39.862	0.235
5	0.0718	3.296	0.237	2.392	0.172
Sum			**3.583**		**3.015**

station, taking into account the *VCE* at the station. The roughness (predictor 14) can be calculated 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 process.

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 will vary whether the calculation is at stations or at grid points. In development of the equations, whether the density was computed at stations or at the closest grid point for a station on a 2.5-km grid, it would not vary much; thus, the station itself is not entering into the calculation. The roughness calculation is done at grid points in the development stage, so there is no difference there. The major difference is for predictor 19; in the development stage the value at the station was known (observed), but during implementation the value is the analysis value.

The fact that predictor 19 is calculated at grid points during implementation and at stations in development may cause a low bias in the estimates for grid points. Because 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 19, and thus the errors at grid points compared to errors calculated at stations.

Temperature and dewpoint analyses made with BCDG are shown in Figs. 1 and 3 and the corresponding error maps in Figs. 2 and 4 for 0000 UTC 23 November 2010. The number of temperature (dewpoint) observations used was 11 896 (10 295). The surface map for the same date taken from NCEP's Hydrometeorological Prediction Center archives is shown in Fig. 5. A dominant feature is a low-pressure center near the western Great Lakes, with a strong cold front trailing down through Illinois, Missouri, Oklahoma, and into Texas. Another low-pressure system is moving into the Seattle, Washington, area. A weak low pressure exists in western Oklahoma and Kansas along with an associated strong moisture gradient.

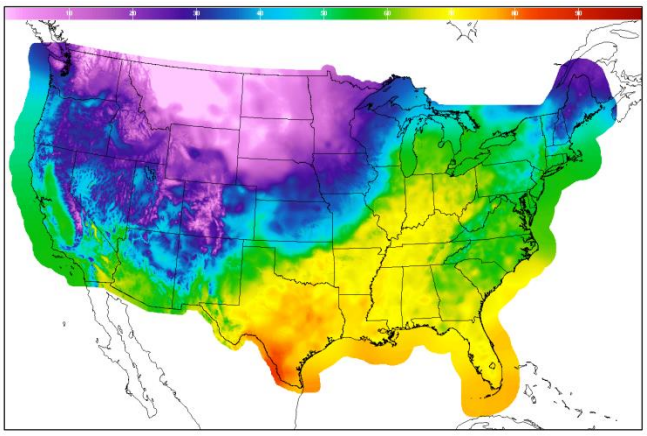


Figure 1. BCDG analysis of temperature (°F) valid at 0000 UTC 23 November 2010. *Click image for an external version; this applies to all figures hereafter*.

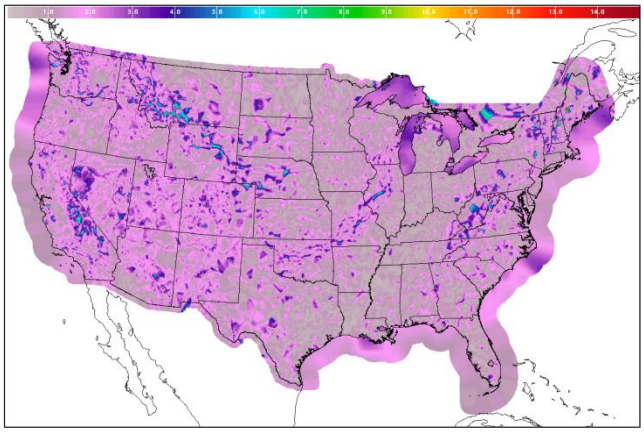


Figure 2. Error estimation (°F) of the BCDG temperature analysis valid at 0000 UTC 23 November 2010.

The dominant large-scale feature is the cold front in the central part of the country [note that the color scale is not the same on the temperature and dewpoint analyses, one being shifted 20°F (11.1°C) to the other]. The error maps show that while the areas well ahead and behind the front have small estimated errors—generally less than 2°F (1.1°C) for both temperature and dewpoint—the frontal area stands out with values at individual grid points being in question by as much as 8°F (4.4°C) or so for dewpoint. The front is in an area with minimal terrain differences, so the larger estimated errors for dewpoint (relative to temperature) are due to the variability of the observations, and to a much lesser extent the dewpoint values being slightly less dense.

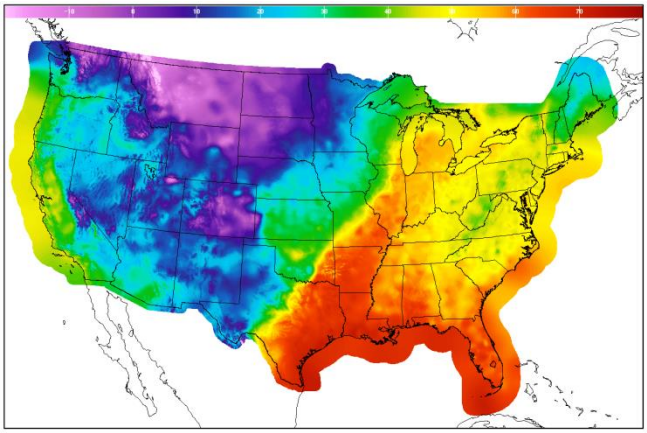


Figure 3. Same as Fig. 1 except for dewpoint.

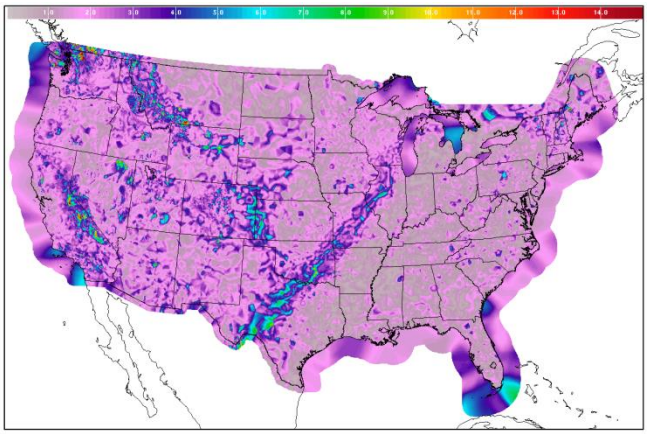


Figure 4. Same as Fig. 2 except for dewpoint.

The irregular streak of relatively higher estimated temperature errors running from northwestern Wyoming to eastern Wyoming is along the frontal boundary shown in Fig. 5. A more irregular pattern is also shown in the dewpoint estimated errors.

The California Central Valley is noticeable especially on the temperature map, and the Snake

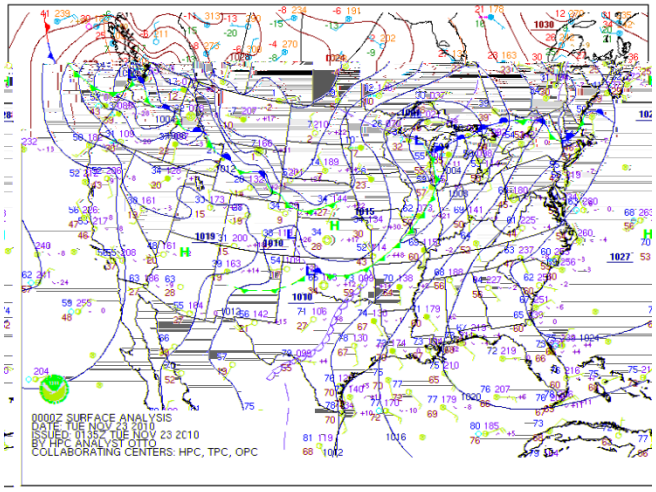


Figure 5. Surface weather map valid at 0000 UTC 23 November 2010 (available online at www.hpc.ncep.noaa.gov/html/sfc_archive.shtml#CONUS).

River Valley stands out, especially on the dewpoint analysis; being rather broad features with adequate observations, the estimated errors are small. The Sierra Nevada Mountains near the border of California and Nevada indicate cold and dry conditions, and because of a scarcity of observations as well as large variations in elevation, the errors are variable and larger than in surrounding areas. Small-scale features can be noted in the analyses, such as Death Valley in southern California (warm and dry) and the Grand Canyon in northwestern Arizona (warm in the temperature analysis).

The largest temperature errors for these analyses are in mountainous regions, as expected. Dewpoint errors are generally larger than temperature errors because of somewhat less dense and more variable dewpoint values, especially in mountainous areas.

For a closer look, we will concentrate on (i) an area in west-central Illinois where the estimated errors are larger than for other areas in the vicinity along the front, (ii) another area along the front in Oklahoma, where the estimated dewpoint errors are higher than surrounding areas, and (iii) an area in southern Ohio where the dewpoint analysis shows an area with lower values relative to the immediate surrounding area.

Figures 6a and 6b show the temperature and dewpoint analyses, respectively, for an area in Illinois just east of the Illinois-Iowa-Missouri triple point. The area of higher estimated error is a narrow southwest–northeast oriented band along the northern portion of the front shown in Figs. 2 and 4. This is an area of relatively sparse data and where the gradient is tight as shown in the figures. There is a 10–20°F (5.6–11.1°C) difference in both temperature and dewpoint observations across the front, so it is reasonable that the exact analysis in this small, relatively data-sparse area would be more questionable than in surrounding areas.

Figures 7a and 7b show the dewpoint analysis and its estimated error over southern Oklahoma. Here, the observations are not sparse, but the difference of two values across the boundary is as much as 28°F (15.6°C), so there is a narrow zone of uncertainty.

Figures 8a and 8b show an area in south-central Ohio where the dewpoint analysis has a couple of

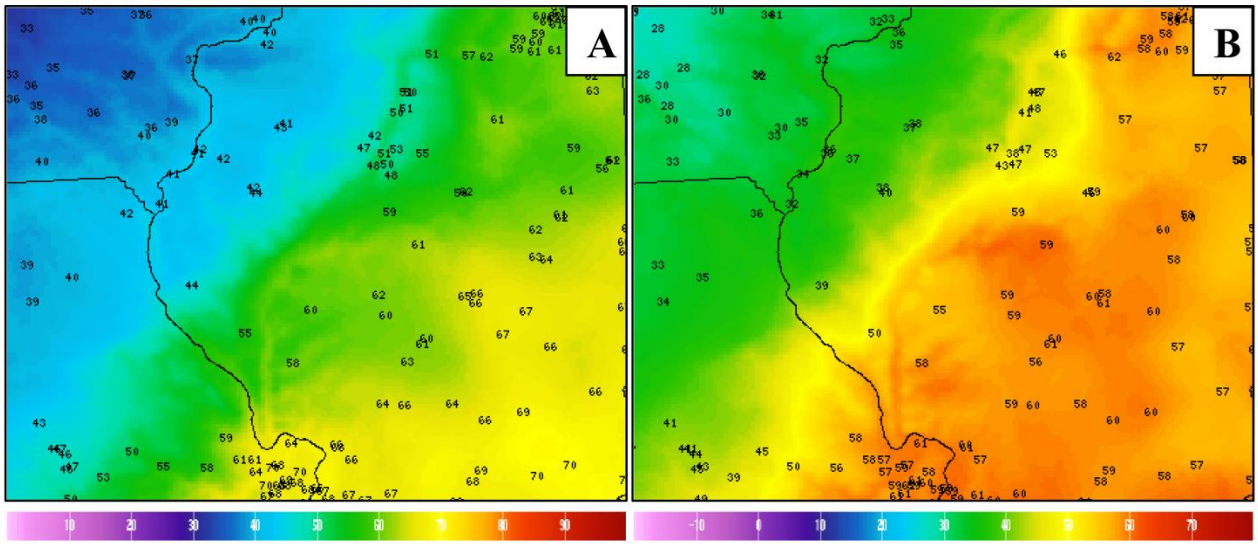


Figure 6. Temperature (A) and dewpoint (B) analyses in west-central IL valid at 0000 UTC 23 November 2010.

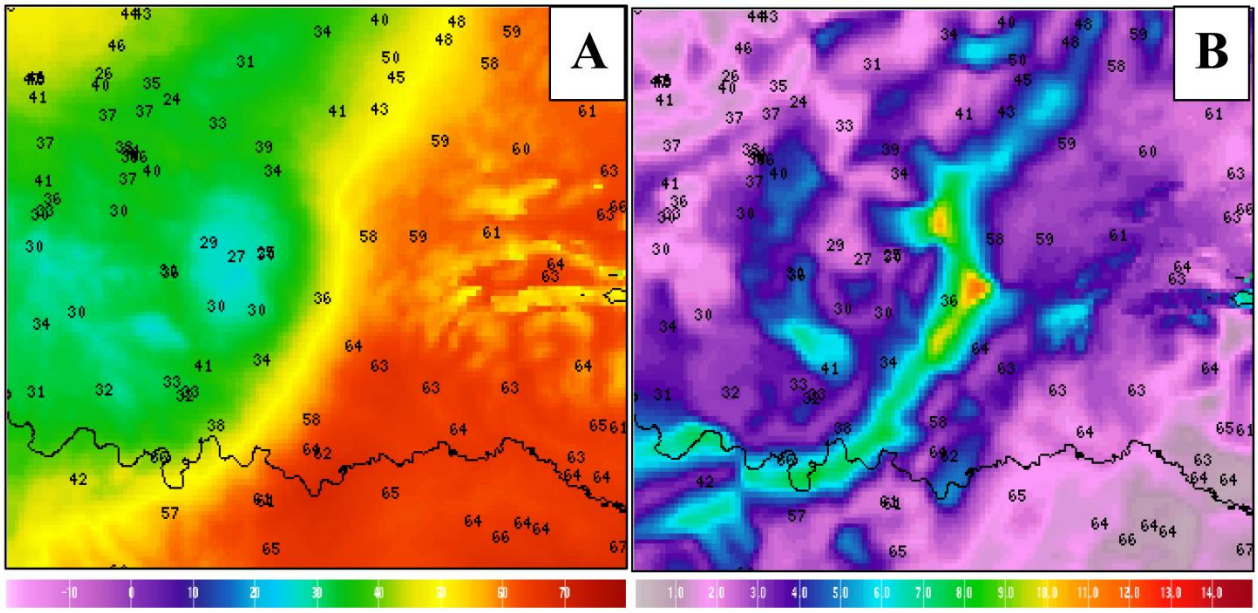


Figure 7. Dewpoint analysis (A) and its estimated error (B) in southern OK valid at 0000 UTC 23 November 2010.

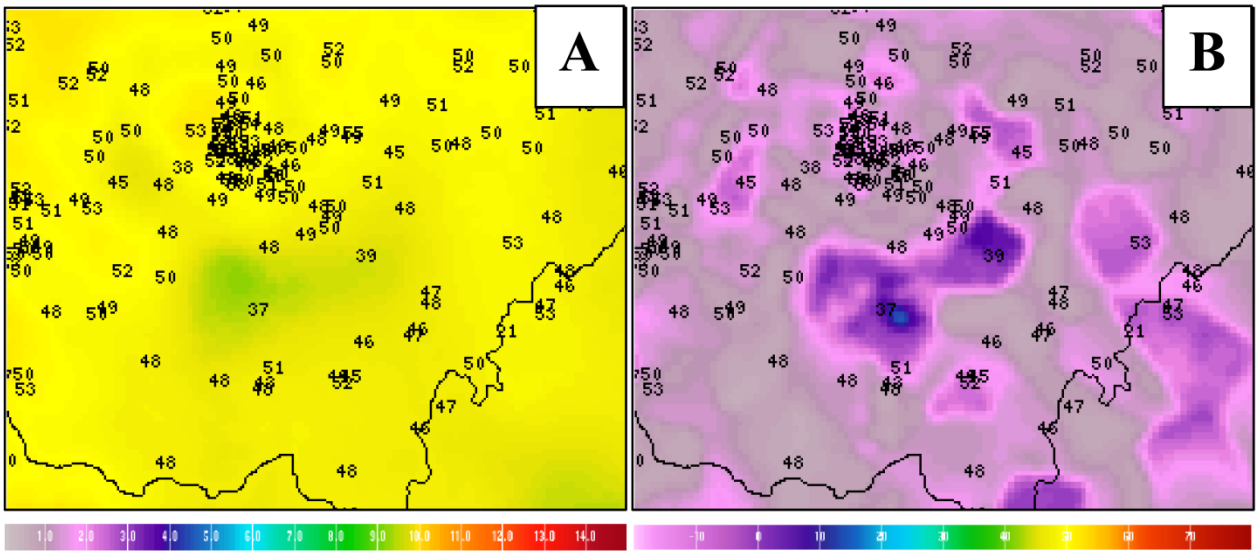


Figure 8. Dewpoint analysis (A) and its estimated error (B) in southern OH valid at 0000 UTC 23 November 2010.

spots with lower values than the surrounding area. Two stations with values of 37°F (2.8°C) and 39°F (3.9°C) are in the midst of values in the high 40s °F (~9°C) and low 50s °F (~11°C). They are not extreme enough to be discarded by the analysis process and show up in the analysis, but are not prominent. The error analysis suggests a closer look at these two values. Persistent spots in the error maps can indicate station reporting problems. In this case, the spot in the analysis is closely associated with twin spots in the error grid. On the other hand, the dominant spot in western Oklahoma with high dewpoints (Fig. 3) is not closely associated with a specific spot on the error map (Fig. 4) because the clustered observations there support each other. That is, a "spot" in an analysis is not necessarily associated with an area of higher uncertainty. Figure 8b also shows another less noticeable spot associated with the station observation of 53°F (11.7°C), which is a few degrees higher than surrounding values, and just to the east of the other two spots. While some meteorologists may consider such spots as blemishes on a national map and would prefer they be smoothed out, differences of this order can and do occur owing to mesoscale conditions such

as clouds, rain, etc. We have tried to be true to the observations so that forecasters at individual weather forecast offices can see the fine-scale detail that may be indicative of weather features that suggest careful scrutiny.

Hourly analyses of temperature and dewpoint and the associated errors are in the National Digital Guidance Database[7] as part of the LAMP suite of forecasts (see Ghirardelli and Glahn 2010, 2011) about 45 min after each hour. Such analyses can be considered to be 0-h forecasts and compared to the LAMP 1-h forecasts for continuity. Areas of small analysis error should indicate where the consistency from analysis to forecast should be relatively good. However, because a station value not only affects the analysis but also contributes substantially to a 1-h forecast, both the analysis and forecast may exhibit error in the area indicated by the error map.

8. Discussion and conclusions

A method to estimate the errors associated with the analysis of temperature and dewpoint has been developed and demonstrated with the BCDG analysis method. It should be recognized that 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 dataset 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., Goodin et al. 1979; Smith and Leslie 1984), but it is difficult to devise an analytic function that simulates (i) the real world with elevation differences, (ii) data with unknown errors, and (iii) data densities that are variable and reasonable. Also, this does not address the real world, day-to-day synoptic situations.

This method, which we call BCDGE (BCDG error), furnishes an estimate of error that is physically reasonable, is specific to the dataset being analyzed, and is relatively easy to implement. To emphasize a previous point, the error used in the development included the interpolation error (i.e., the estimate of the station value from the regularly spaced grid). This itself can be a considerable cause of error, especially in the western CONUS. 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 also is true of the analysis; the true values at high elevations are not known. While the error estimation method presented here was applied to the BCDG analysis method, it is general and could be applied to any analysis method.

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 at grid points. De Pondeca et al. (2011) presented an example error map from the RTMA in their Fig. 8. It shows similar characteristics to our error maps, inasmuch as the smaller errors are close to observation points, and the errors are greater between observation points. The magnitudes of the errors presented for the RTMA example and the BCDG example are about the same. The patterns in the BCDGE maps help pinpoint where the problems are with the associated analyses; the actual values are not as important in this respect as the spatial variability. For instance, the analysis of dewpoint does not by itself indicate a potential problem in Ohio (Fig. 8a), but the error-map pattern calls attention to it (Fig. 8b).

The detailed pattern or errors, which are not obvious at the scale shown in Figs. 2 and 4, are more "choppy" than desired. This is due to the discrete nature of some of the predictors. For instance, predictors 2 and 19, which include a contribution from the closest or second closest station, can switch abruptly from grid point to grid point because the closest station to the grid point switches abruptly. The result is a boundary about halfway between two stations. If different predictors that have a less discrete nature could be derived, the error estimates would not switch so abruptly. Because of this discrete nature, the analysis errors should be viewed as highlighting an *area* of possible error rather than focusing on individual gridpoint values.

This method for determining error also is applicable to wind speed, possibly with some adaptation. However, the high variability of some variables, like ceiling height and visibility, make application of this method—or actually any method—questionable for these variables.

Acknowledgements. The authors thank Scott Scallion for implementing the error-estimate products into operation. We also appreciate the careful technical editing of Matthew Bunkers.

[7] The National Digital Guidance Database is the guidance counterpart of the NDFD, and can be accessed by the same methods as the NDFD.

APPENDIX A

The VCE Method

The vertical change with elevation (*VCE*), sometimes called the lapse rate, plays a major role in how a datum affects a grid point. Consider station A with a specific temperature value *TA* at elevation *EA*, and another station B a short distance away with a temperature value *TB* at elevation *EB*. To apply a correction based on station A to a grid point near station B, one needs to consider that the correction should be based on the *VCE*, defined as:

———. (1)

The *VCE* is computed for each station for each analysis. The specific value for a station is based on several stations (B_i) that are close in horizontal distance and far apart in vertical distance, with the more stations the better. Thus,

Σ /Σ (2)

is the VCE summed over all designated close stations, B_i. This process is quite robust, computing only one statistic from several pieces of data. It is important to find a list of stations that are close in horizontal distance but far apart in vertical distance. A preprocessor finds such lists of stations by making several passes over the vertical and horizontal station locations, searching for the desired combinations.

In the analysis, the modifications to grid points from the stations within the radii of influence use not only the observation, but also the individual *VCEs*.

APPENDIX B

Method of Selection of Withheld Points

In order that the regression equations relating error to predictor variables represent the map equally and not be based on, for instance, data density, a latitude and longitude within the confines of the grid were each selected randomly. Their combination defined a point on the analysis grid. To be used, the point was required to be within the NDFD domain. The station closest to the point was selected to be withheld. Canadian stations were not used.

One disadvantage of this selection method is that, while regions of the country are represented approximately equally, a value in sparse data areas might be selected to contribute to the computation much more than one in a dense data area. It is not possible to weight each geographic region *and* each datum equally. It was thought that using a datum more often than others would not invalidate the results, unless the observation had unusual error characteristics.

Note that this selection process is in distinction to randomly selecting from the station list. If that were done, the areas of dense data would be weighted too heavily and create a bias toward areas of high-density observations.

The seed for the pseudo random-number generator can be the same in each checkout run to get consistent results. However, it must be different in actual withholding runs, or the same stations will be withheld on each run. Rather, the seed can be based on the system clock to get different withheld stations.

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**, 396–409.

Bergthórsson, P., and B. R. Döös, 1955: Numerical weather map analysis. *Tellus*, **7**, 329–340.

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

De Pondeca, M. S. F. V., and Coauthors, 2011: The real-time mesoscale analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. Forecasting*, **26**, 593–612.

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

Ghirardelli, J. E., and B. Glahn, 2010: The Meteorological Development Laboratory's aviation weather prediction system. *Wea. Forecasting*, **25**, 1027–1051.

____, and ____, 2011: Gridded Localized Aviation MOS Program (LAMP) guidance for aviation forecasting. Preprints, *15th Conf. on Aviation, Range, and Aerospace Meteorology*, Los Angles, CA, Amer. Meteor. Soc., 4.4. [Available online at ams.confex.com/ams/14Meso15ARAM/webprogram/Manuscript/Paper191169/AMS_glmp_ghirardelliandglahn_2011_finalsubmitted.pdf.]

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

Glahn, H. R., 1981: Comments on "A Comparison of Interpolation Methods for Sparse Data: Application to Wind and Concentration Fields." *J. Appl. Meteor.*, **20**, 88–91.

____, 1987: Comments on "Error Determination of a Successive Correction Type Objective Analysis Scheme." *J. Atmos. Oceanic Technol.*, **4**, 348–350.

____, and J. E. McDonell, 1971: Comments on "Objective Analysis of a Two-Dimensional Data Field by the Cubic Spline Technique." *Mon. Wea. Rev*., **99**, 977–978.

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

____, and J.-S. Im, 2011: Algorithms for effective objective analysis of surface weather variables. Preprints, *24th Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 10A.4. [Available online at ams.confex.com/ams/91Annual/webprogram/Manuscript/Paper179368/ams_seattle_2011.pdf.]

____, T. L. Chambers, W. S. Richardson, and H. P. Perrotti, 1985: Objective map analysis for the Local AFOS MOS Program. NOAA Tech. Memo., NWS TDL 75, Techniques Development Laboratory, Silver Spring, MD, 34 pp. [Available online at www.nws.noaa.gov/mdl/pubs/Documents/TechMemo/TechMemo75.pdf.]

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

Glowacki, T. J., Y. Xiao, and P. Steinle, 2012: Mesoscale surface analysis system for the Australian domain: Design issues, development status, and system validation. *Wea. Forecasting*, **27**, 141–157.

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**, 761–771.

____, ____, and ____, 1981: Reply. *J. Appl. Meteor.*, **20**, 92–94.

Horel, J., and B. Colman, 2005: Real-time and retrospective mesoscale objective analyses. *Bull. Amer. Meteor. Soc*., **86**, 1477–1480.

Im, J.-S., and B. Glahn, 2012: Objective analysis of hourly 2-m temperature and dewpoint observations at the Meteorological Development Laboratory. *Natl. Wea. Dig*., **36**, 103–114.

____, ____, and J. Ghirardelli, 2010: Real-time objective analysis of surface data at the Meteorological Development Laboratory. Preprints, *20th Conf. on Probability and Statistics*, Atlanta, GA, Amer. Meteor. Soc., 219. [Available online at ams.confex.com/ams/pdfpapers/159910.pdf.]

____, ____, and ____, 2011: Real-time hourly objective analysis of surface observations. NOAA Tech. Memo., NWS MDL 84, Meteorological Development Laboratory, Silver Spring, MD, 19 pp. [Available online at www.nws.noaa.gov/mdl/pubs/Documents/TechMemo/TechMemo84.pdf.]

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Manikin, G. S., and M. S. F. V. De Pondeca, 2011: Recent upgrades to and ongoing challenges for the real-time mesoscale analysis (RTMA). Preprints, *15th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans and Land Surface (IOAS-AOCS)*, Seattle, WA, Amer. Meteor. Soc., J7.3. [Available online at ams.confex.com/ams/91Annual/webprogram/Manuscript/Paper178899/RTMA_AMS_2011.pdf.]

Murphy, A. H., and R. W. Katz, 1985: *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. Westview Press, 545 pp.

Myrick, D. T., and J. D. Horel, 2008: Sensitivity of surface analyses over the western United States to RAWS observations. *Wea. Forecasting*, **23**, 145–158.

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

Pondeca, M., and G. S. Manikin, 2009: Recent improvements to the real-time mesoscale analysis. Preprints, *23rd Conf. on Weather Analysis and Forecasting/19th Conf. on Numerical Weather Prediction*, Omaha, NE, Amer. Meteor. Soc., 15A2. [Available online at ams.confex.com/ams/pdfpapers/1_99_.pdf.]

Smith, D. R., and F. W. Leslie, 1984: Error determination of a successive correction type objective analysis scheme. *J. Atmos. Oceanic Technol.*, **1**, 120–130.

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

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.

Tyndall, D. P., and J. D. Horel, 2013: Impacts of mesonet observations on meteorological surface analyses. *Wea. Forecasting*, **28**, 254–269.