U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
OFFICE OF SCIENCE AND TECHNOLOGY INTEGRATION
METEOROLOGICAL DEVELOPMENT LABORATORY

MDL OFFICE NOTE 18-1
(REVISED)

THE ANALYSIS OF CEILING HEIGHT AND VISIBILITY OBSERVATIONS AND PROBABILITY FORECASTS FOR THE LAMP/HRRR MELD

Bob Glahn and Adam D. Schnapp

May 2019 (Revised)

THE ANALYSIS OF CEILING HEIGHT AND VISIBILITY OBSERVATIONS AND PROBABILITY FORECASTS FOR THE LAMP/HRRR MELD

Bob Glahn and Adam D. Schnapp

1. INTRODUCTION

There are many places in the numerical weather prediction (NWP)-postprocessing marriage that an "objective analysis" of quasi-randomly spaced points to a regularly spaced grid is necessary, although in the NWP world the term data assimilation is now used and is a somewhat broader term. In LAMP (Localized Aviation MOS Program) (Ghirardelli and Glahn 2011), we use the BCDG method to analyze both observations and forecasts. The BCDG software has many tuning parameters that are specified according to the circumstances. This office note documents some of the procedures and parameter settings used in the analysis of ceiling height and visibility observations[1] and also of the LAMP probability forecasts of ceiling height and visibility.

2. THE BCDG ANALYSIS SYSTEM

The BCDG ancestry reaches back to the 1960's, but the specific code being discussed here was adapted from earlier versions in 2004, with improvements still being made. The technique is that of successive correction put forth by Bergerthorssen and Doos (1955) and made operational by Cressman (1959) for use at the National Meteorological Center, the forerunner of the National Centers for Environmental Prediction (NCEP). The method has become known as "the Cressman analysis." The current code, although still based on successive correction, has evolved so drastically that we gave it a new name, BCDG for the names of the primary contributors. Uses of it have been documented in Glahn et al. (2009), Im et al. (2010), Glahn and Im (2011; 2015), and Im and Glahn (2012).

A. Successive Correction

Successive correction means that a grid is given a value at each gridpoint, called the "first guess," and is corrected by each data value being analyzed in the vicinity of that datum on successive passes over the data. That is, each datum corrects the grid, and then the process is repeated. Usually the number of passes over the data is four to six. Each correction for each pass is made in the following manner:

1) Interpolate into the grid (the analysis at that point in the process) to the data point.
2) Find the difference between the interpolated value and the datum.
3) Apply that difference to all gridpoints within a (circular) radius of influence, usually with a weight depending on the distance between the datum and the gridpoint.

Fig. 1 shows three types of correction to make to the gridpoints. Type 1 weights each datum's correction to a gridpoint equally, Type 2 weights each datum's correction by a weight W_i that

[1] Ceiling height is not actually observed but is computed from related observations. It is treated here as "observed."

depends on the distance between the datum and the gridpoint, and Type 3 is the same as Type 2 except the sum of the weights is in the denominator. The weighting coefficient is shown in Fig. 1. Primarily, Type 3 is used to bring about convergence of the gird to the data more quickly. This brings up the question of what to use for a first guess. It would seem that if one started with a first guess that had some similarity to the data, convergence would be better and faster. Unfortunately, that is not generally the case. For instance, consider a grid of ceiling heights as a first guess. A datum with a value of 9,000 ft will indicate a decrease of 3,000 ft at that point from an initial interpolated value of 12,000 ft. That 3,000 ft is applied within the radius of influence. This is reasonable if the values within the radius of influence started at the same value of 12,000. But suppose the first guess was 1,000 ft at the gridpoint being corrected. This might indicate a ceiling at that gridpoint far below ground, depending on W_i.

$$C_1 = \frac{1}{n}\sum_{i=1}^{n} D_i, \text{ type1;}$$

$$C_2 = \frac{1}{n}\sum_{i=1}^{n} W_i D_i, \text{ type2; and}$$

$$C_3 = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i D_i}{\frac{1}{n}\sum_{i=1}^{n} W_i}, \text{ type3,}$$

where

$$W_i = \frac{R^2 - d_i^2}{R^2 + d_i^2},$$

Figure 1. The three types of corrections possible in BCDG. For each station i, D_i is the correction, W_i is the weight, R is the radius of influence for the station, and d_i is the distance from the station to the gridpoint being corrected. For variable radii, R varies by station.

Even if the first guess is nearly the correct solution, and that is not generally the case, little is to be gained and much to be lost in starting with a spatially variable first guess. Starting from a constant first guess, each pass over the data will fit the data more closely and with more detail.

B. Treatment of Land/Water Boundaries

Many times the data values will vary markedly across a land/water boundary. BCDG makes essentially two analyses at the same time, one over water and one over land in which the boundaries can be treated as impervious or leaky. That is, data over water can be used to correct only water gridpoints and data over land can be used to correct only land gridpoints. Or the effect of water (land) data can be allowed to leak to the nearby land (water) gridpoints. Smoothing can also be treated in a similar fashion.

C. Adjustment of Corrections Based on Elevation

In many cases, the values on the grid should depend on the elevation of those points. This is best visualized for 2-m surface temperature, which usually decreases with altitude. Note, however, this is not in the free air where measured lapse rates might be preferred, but 2-m surface temperatures. BCDG computes a lapse rate to use based on the data. That is, data surrounding a datum point is surveyed, and based on the data values and their elevations, an average change with elevation is calculated; this change is used in the adjustment procedure (see Glahn et al. 2009 for more detail). In order to save time in running BCDG, a preprocessor U174 is used which calculates for each datum point a set of datum points in its vicinity to use in calculating the lapse rate in BCDG.

3. THE LAMP/HRRR MELD PROCESS FOR CEILING HEIGHT AND VISIBILITY

The current LAMP/HRRR meld (Glahn et al. 2017) that is in operation today is being improved by the addition of LAMP forecasts at stations and the extension of the hourly projections from 25 to 36 h. In addition, the area of extent is being greatly increased for input to the National Blend of Models (NBM). The meld is a 3-tier MOS process:

1) MOS equations based on the NCEP GFS model are developed for stations, and probability forecasts from them are made at **stations**.
2) LAMP (MOS) equations are developed for stations, one of the inputs being the MOS probability forecasts, and probability forecasts from them are made at **stations**.
3 LAMP/HRRR meld (MOS) equations are developed at stations, two of the inputs being the LAMP probability forecasts and current observations, and forecasts are made from them at **gridpoints**.

For the final meld step, in order to evaluate the meld probability equations at gridpoints, the inputs have to be on the grid. HRRR fields for input are already on a grid, but the current observations have to be gridded as well as the LAMP probability forecasts. The analysis details for ceiling and visibility observations are similar enough that they can be discussed together (Section 5). Also, the analysis details for ceiling and visibility probabilities are similar enough that they can be discussed together (Section 6).

4. PREPROCESSORS FOR THE ANALYSES

A. Determining Variable Radii

As mentioned previously, each datum affects gridpoints within its radius of influence. The radii of influence, one for each pass over the data, typically decrease with pass. The large first pass radius must assure that all gridpoints are affected by at least one datum, preferably more. The smaller last pass radius allows local detail to be captured; a datum may affect only a few gridpoints in its immediate vicinity. The original Cressman scheme and the process used for many years was to assign each datum the same radius for a particular pass. This is workable if the data density is rather uniform over the grid. Unfortunately, that is not the case in many real-world situations.

A preprocessor (U178C) derives a set of first pass radii, one for each datum (station[2]), that strives to guarantee that each gridpoint will be affected by at least five data points,[3] but also keeps radii small where data density is high. It also specifies up to five additional radii for subsequent passes, each of which is just a set fraction of the first. An input parameter limits the radius of search in U178C to find points, and it may be that not all gridpoints have a datum to affect them. Diagnostics are provided, one of which is a grid that can be viewed with gmos_plot that shows in colors the gridpoints that can be affected by 0, 1, 2, 3, and 4 or more data values.

[2] Most data are at stations, and "station" and "datum" or "datum value" are used interchangeably in this paper.
[3] The "5" is an input to U178C. A larger number causes larger radii. Each situation may be unique, and an iterative process may be needed in which radii are calculated and tried in the analysis to determine whether or not an adjustment is needed.

This map facilitates making adjustments as needed. The grid for mapping is in file fort.99 and the stations (locations) on which the calculated radii are based are in file fort.98. These are used in gmos_plot (use the color bar for ceiling). The plotted values are the "station types," (0 = ocean, 3 = lake, 9 = land).

U178C provides for an "override" feature which gives the user the ability to specify the first pass radius for specific stations. This can be used in unusual cases to insure all gridpoints have a correction and to target specific gridpoints to be affected by bogus stations (points where data are manufactured to assist in the analysis).

B. Determining the Datum Pairs for Calculating Lapse Rates

Preprocessor U174 determines for each land station a set of other land stations each of which would be good to pair with the base station to calculate the lapse rate at that station point.[4] U174 attempts to get 60 data points to pair with the datum point. The search is made, and the stations in the output lists are such that the stations with the smaller horizontal and larger vertical distances from the base station are first in the list, the exact parameters of search being specified in U174. Some stations in flat terrain areas will have few or even no pairs because of similar elevations.

5. ANALYSIS OF CEILING AND VISIBILITY OBSERVATIONS

U155 is run to make an analysis and it calls U405A; each have writeups and .CN control files. In explaining the specific settings, subroutine names and the variable names in the writeups and .CN files are used (capitalized). They can be identified in the U405A writeup. Water and land are analyzed separately, with no bleeding between them (land does not affect water and vice versa; WTLTW = 0 and WT WTL = 0).

A. First Guess Grids

GUESS is set to 120 (hds of ft) for ceiling and 10 (mi) for visibility. However, NBLEND is set to 3, so the land gridpoints are set to GUESS and the water gridpoints are set to the first guess values specified in the 4th entry of ITABLE, which is 008000003 for ceiling and 008100003 for visibility. These IDs are for the RAP (rapid refresh) model 1-h forecast from the previous (to the analysis time) cycle. The HRRR would have been used, but it didn't cover the whole area.

B. Treatment of Data

The ceilings are accessed in hundreds of ft and the visibilities in miles. Ceilings are unique because they can be "unlimited," which means there are no clouds of sufficient coverage to constitute a ceiling. These are coded as 888, and for the analysis are set to 130. Also, for the analysis, all values > 120 are set in CIGFRQ to 130. They cannot be left at a very high number (compared to legitimate values) or they will overpower the smaller values. It was determined

---

[4] I am using lapse rate here to mean the change with elevation, either positive or negative, of the variable being analyzed.

that 130 worked well. A lower number, like 120, gave too many gridpoints with values under 120.

Reported visibilities are generally ≤ 10 mi. ASOS (Automated Surface Observation System) reports ≤ 10 mi, but manual observations and reports from outside the U.S. can be > 10 mi. For analysis purposes, all reported values > 10 are set in VISFRQ to 10.

C. Bogus Points

There are very few observations over the water in the area analyzed except in the northern Gulf of Mexico. We wanted to preserve these values. However, because there was no way to integrate these few points well, we decided to use the HRRR model forecasts unchanged over all water areas. That is, we just put the HRRR values at gridpoints; no analysis or smoothing was done. This included values in the Great Lakes, Lake Pontchartrain, Lake Okeechobee, Lake Winnipeg, and Hudson Bay.

D. Treatment of Lapse Rates

Lapse rates are calculated and used in the analyses. This is more problematic for ceiling and visibility than for temperature. Many times as altitude increases, the clouds are lower (above ground), the ceiling decreases and the visibility decreases (clouds may be at ground level). However, there can be low clouds with low tops, and the ceiling can be unlimited at higher elevations. In the same manner, there can be low-lying fog and low visibilities at low elevations, and be good visibility at higher elevations.

In the adjustment for elevation, which for a particular station the computed lapse can be either positive or negative, U405A can give preference to the sign of the lapse. For instance with temperature, the lapse is expected to be negative with elevation; if the computed lapse is positive, it is used but in a restricted sense. For both ceiling and visibility, preference is given to negative lapse rates; positive lapse rates are considered "unusual," and are used only in a restricted sense (IBKPN = 1). Allowable maximum and minimum lapse rates are specified in LAPSE. If the calculated lapse rate exceeds the maximum or is lower than the minimum, it is set to that allowable value.

In calculating lapses for ceiling, values of 130 are not used. Values of 120 (which represent observed values of 12,000 ft) are used.

E. The Correction Passes

Six corrective passes are made over the data. For ceiling, the first four are made with the radii specified by U178C. The last two are made with the default radii in U405A, namely 9 and 5 gridlengths. The switch to the default radii is made to allow the values to be fit more closely; the radii calculated by U178C for the last passes are larger because a lower limit of 40 was specified. For visibility, the first three passes are made with the U178C radii, and the last 3 with the default values of 14, 9, and 5.

Type 3 corrections are made except in certain cases when there is only one datum to affect a gridpoint, Type 2 is used.

Water and land are analyzed separately (ILS=1), and there is no leakage from one to another (WTWTL = 0, WTLTW = 0). It was found that using corrections to water gridpoints from (numerous) land observations created too much havoc with the model first guess we are using.

F. Post Correction Smoothing

The spot remover (SPOTRM) is run after passes 5 and 6 over land (only) for visibility but only after pass 6 for ceiling. It smooths out bogus points. It could be run over the Gulf, where there are real and bogus points, except the facility does not exist for running if over only part of the ocean, and there are no data over the Atlantic and Pacific

G. Making the Grid Agree with the Data Points

For some purposes it is desired to retrieve the original observations from the grid. To facilitate this, each observed value is inserted into the grid at the closest gridpoint to the station. This is controlled by ISETP = 3. Only in cases where two stations are closest to the same gridpoint is an exact station value not retrievable for one of the pair. In case of conflicts, preference is given to METAR stations defined to have identifiers of KXXXbbbb (X is any character, b is blank) in the CONUS. Bogus points are not inserted.

H. Output Grids

The output grids written to unit KFILIO are in hundreds of ft for ceiling and miles for visibility. For ceiling, values > 120 and ≤ 121 are set to 120; values > 121 are set to 888 by POST88. For visibility, values > 9.1 are set to 10 in POSTPM.[5]

There is no reflection of values reported > 12,000 ft or > 10 mi in the grids.

6. ANALYSES OF LAMP CEILING AND VISIBILITY PROBABILITY FORECASTS

A. First Guess Grids

HRRR MOS equations were developed at stations for each of the categories. These equations predict the probability of the categories based on the HRRR model forecasts of the last three runs (a time-lagged ensemble). The HRRR is not available for the outer extremities of the grid, so the

[5] Note the difference in the way high values are treated for ceiling and visibility. For ceiling analysis, the cap is set at 130 for reports of unlimited (888), and then gridpoint values > 121 are set to 888 (unlimited) and values > 120 and ≤ 121 are set to 120. For visibility, the maximum of 10 is analyzed as such, then at the end gridpoint values > 9.1 are set to 10 to get flat areas of 10. Both methods seem to work, with no clear preference. There is a difference for ceiling because of the unlimited reports that need to be set at some value > 120. There is no counterpart for visibility, even though there are a few reports of > 10.

RAP was used as a substitute. At the HRRR-RAP boundaries, strong discontinuities exist, so the RAP was blended into the HRRR using the logit function as a weighting mechanism. This blended field was used as input to the equations to produce a first guess over the whole area. Settings of IGUESS = 3 and NBLEND = 0 were used. In retrospect, it might have been better to use NBLEND = 1 to use a constant over land.

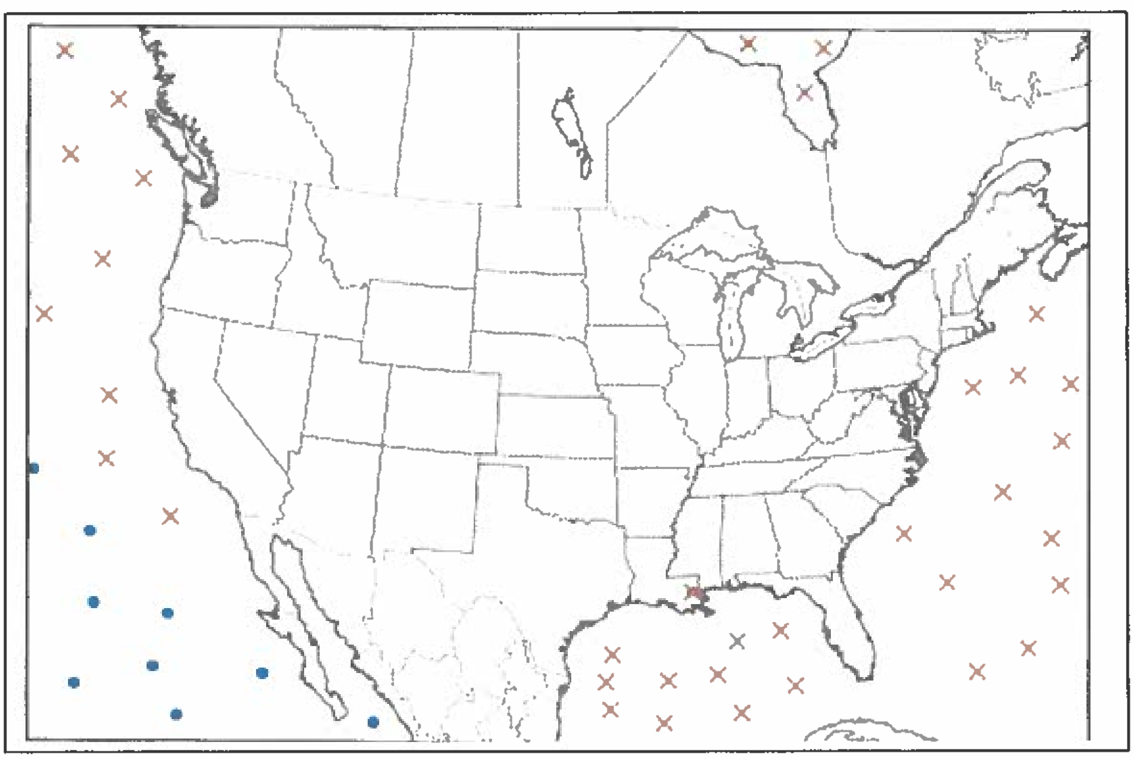
B. Treatment of Data

The probabilities are accessed in fractions, the way they are produced by the regression equations. Values exist at land stations and for some points in the water. These latter are made with LAMP backup equations developed at land stations in regions and with no observations input. These are at points generally near the Atlantic and Pacific coasts and in the northern Gulf of Mexico.

C. Bogus Points

We added bogus points in the Atlantic, Pacific, Gulf, Lake Pontchartrain, and Hudson Bay (see Fig. 2). Most points are just a weighted average of two or more nearby points. The nearby points are preferably over water, but not always is that possible. For instance, the three bogus points in Hudson Bay are from land stations, and being in northern Canada, they are from backup equations.

A bogus point is defined for Lake Pontchartrain from nearby land points so that this water point will have a land influence as well as a water influence from the Gulf. The Great Lakes are not given bogus points because they have some LAMP forecasts and are influenced from land



points because of the land to water leakage. Points in the Pacific Ocean in the lower left of the map are interpolated from the first guess.

Figure 2. Bogus points for analyses of LAMP probabilities. The blue dots are interpolated from probability forecasts made with MOS equations based on three cycles of RAP forecasts. The red X's are based on other points with LAMP probabilities.

D. Treatment of Lapse Rates

Lapse rates are calculated and used in the analyses. As with observations, the probability of ceiling or visibility could increase or decrease with height, but it seems more likely than not that the probability of a particular cloud base would increase with elevation. The computation of lapse rates is done as a combination of all levels analyzed (seven for ceiling and six for visibility) in order to be consistent among levels. There were almost twice as many positive as negative rates computed on the case I examined (for both ceiling and visibility), which means more often than not the probability of a level increased with height. Because negatives were more rare than positives, negatives are treated as "unusual" and IBKPN is set to -1. This makes the higher elevations tend to stand out with higher probabilities, but not always. Allowable maximum and minimum lapse rates are specified in LAPSE. If the calculated lapse rate exceeds the maximum or is lower than the minimum, it is set to that allowable value.

E. The Correction Passes

Six corrective passes are made over the data. For ceiling, the first four are made with the radii specified by U178C. The last two are made with the default radii in U405A, namely 9 and 5 gridlengths. The switch to the default radii was made to allow the values to be fit more closely; the radii calculated by U178C for the last passes are larger because a lower limit of 40 was specified. For visibility, the first three passes are made with the U178C radii, and the last 3 with the default values of 14, 7, and 5.

As with observations, Type 3 corrections are made, except in certain cases when there is only one datum to affect a gridpoint, Type 2 is used.

Water and land are analyzed separately (ILS=1), but there is leakage from land data to water areas (WTWTL = 0, WTLTW = 1). This leakage gives credence to the near shore areas. The leakage also gives values to the smaller lakes with no data or bogus points.

F. Post Correction Smoothing

The spot remover is run over land after passes 5 and 6 for visibility but only after pass 6 for ceiling. It could be run over the water, but it is quite computer intensive. Instead, the ray smoother (ORVMTH) is run over the water.

G. Making the Grid Agree with the Data Points

For some purposes it is desired to retrieve the original probabilities from the grid. To facilitate this, each data value is inserted into the grid at the closest gridpoint to the station. Only in cases where two stations are closest to the same gridpoint can this not be done. This is controlled by ISETP = 3. In case of conflicts, preference is given to METAR stations defined to have identifiers of KXXXbbbb (X is any character, b is blank). Bogus points are not inserted.

H. Output Grids

For writing to KFILIO, POST sets probabilities < .01 to 0 and sets values > 1.0 to 1.0. The values are then scaled times 100 by POST, so that the output is in percent when written to the disposable file; units of percent are required for gmos_plot. However, for operations, the units are left in fractions when written to the archive file.

7. SUMMARY AND CONCLUSIONS

Only a few of the control parameters and analysis options in BCDG are described in this document. The ones described are what were paramount in this current tuning process. The options for land are very similar to what are currently used in LAMP operations. The process described herein was implemented in the spring of 2019.

It is noted that these analyses are made as part of the LAMP system of forecasts. As such, ceilings > 12,000 ft and visibilities > 10 mi are not represented in the grids. A casual observer might deduce gross errors in the analyses, but the decision to truncate was made by management a number of years ago to conform to the ASOS reporting capabilities.

A. Characteristics of the Analyses of Observations

Over land, the analyses fit the data rather closely, ceiling ranging from 0 to 120 hds ft, then 888 signifying clear or unlimited, and visibility ranging from 0 to 10 mi. There are few reports over water, except in the northern Gulf of Mexico. The analyses over water consist primarily of the RAP model, with observations in the northern Gulf of Mexico being analyzed.

B. Characteristics of the Analyses of LAMP Probability Forecasts

Over land, the LAMP station probability forecasts are represented quite well. The probability levels (the different thresholds) are analyzed together with the same lapse rates to foster consistency, and after the analyses, each level is made compatible with the layer below. Over water, a HRRR-RAP MOS is used as a first guess, and that is supplemented in the analysis by LAMP forecast points along the coasts and outward into the ocean by bogus points as combinations of the forecast points. The lower left portion of the grid (Pacific Ocean) is left intact as the HRRR-MOS RAP.

REFERENCES

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.

Ghirardelli, J. E., and B. Glahn, 2011: Gridded localized aviation MOS program (LAMP) guidance for aviation forecasting. *Preprints 15th Conference on Aviation, Range, and Aerospace Meteorology,* Los Angeles, CA, Amer. Meteor. Soc., **4.4**, 22pp.

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

_____, and J.-S. Im, 2011: Algorithms for effective objective analysis of surface weather variables. *Preprints 24th Conference on Weather and Forecasting/20th Conference on Numerical Weather Prediction*, Seattle, WA, **10A.4**, 11 pp.

_____ and _____, 2015: Objective analysis of visibility and ceiling height observations and forecasts. MDL Office Note 15-2, National Weather Service, NOAA, U.S. Department of Commerce, 17 pp.

_____, A. D. Schnapp, J. E. Ghirardelli, and J.-S. Im, 2017: A LAMP-HRRR meld for improved aviation guidance. *Wea. Forecasting*, **32**, 391-405.

Im, J.-S., and B. Glahn, 2012: Objective analysis of hourly 2-m temperature and dew point observations at the Meteorological Development Laboratory. *Nat. Wea. Dig.*, **36(2),** 103-114.

_____, _____, and J. E. Ghirardelli, 2010: Real-time objective analysis of surface data at the Meteorological Development Laboratory. *Preprints 20th Conference on Probability and Statistics in the Atmospheric Science*, Atlanta, GA, **219**, 11 pp.