92nd AMS Annual Meeting
21th Conf. on Probability and Statistics
22-26 January 2012, New Orleans, LA

# 8.4 OBJECTIVE ANALYSIS OF MOS FORECASTS AND OBSERVATIONS IN SPARSE DATA REGIONS

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

## 1. INTRODUCTION

Objective analysis is the process of interpolating from unevenly spaced data points to a grid of regularly spaced points, which could be in reference to a particular map projection over the earth's surface. Provided the data points are relatively dense over all portions of the grid and there are not complicating factors such as terrain elevation and land/water boundaries that can affect the analysis, the process is relatively straightforward. A number of techniques are available, such as the so-called Cressman (1959) and Barnes (1964), and other more sophisticated methods called data assimilation, including that used for the Real-Time Mesoscale Analysis (RTMA; De Pondeca et al. 2011).

The Cressman technique, actually originating with Bergthorssen and Doos (1955), has been expanded in major ways by the Meteorological Development Laboratory (MDL) into what is called the BCDG (Bergthorssen, Cressman, Doos, Glahn), method and is used for analyzing observations (Im et al. 2010, 2011; Glahn and Im 2011), MOS (Glahn et al. 2009) and Localized Aviation MOS Program (LAMP; Ghirardelli and Glahn 2011) forecasts over the conterminous United States (CONUS), Alaska, and Hawaii. This paper explains briefly the basic analysis technique, presents enhancements, and concentrates on the peculiarities of the analysis of MOS wind forecasts over Alaska and the surrounding ocean areas. While the emphasis here is on MOS forecasts, the technique can apply equally well to observations. The area of analysis is that used for the National Digital Forecast Database (NDFD; Glahn and Ruth 2003); the grid is on a polar stereographic map with a grid spacing of approximately 3.0 km at 60°N, and is oriented with 150°W.

The analysis examples are for 30-h MOS forecasts from the November 8, 2011, 0000 UTC run of the National Centers for Environmental Predication's (NCEP) Global Forecast System (GFS). An intense low existed over the Bering Sea and created tremendous winds and hazards along the western coast of Alaska. The winds were light over most of the interior and only about 20 kts in the Gulf of Alaska. The very strong winds over the Bering Sea created analysis challenges. This paper describes how these challenges were met for this case.

## 2. BASIC METHOD

The analysis starts with a "first guess" (FG)--a value at every grid point. This can be from some previous analysis or forecast, climatology, or just a constant value. Each data point is used to correct the grid points around it within a "radius of influence" R. The radius of influence is fairly large, and the result is a rather smooth field for which the data points are not fit very closely; the result could be considered a local area mean. The corrections due to the data points are indicated by the formula:

$$C = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{R^2 - d_i^2}{R^2 + d_i^2}\left(O_{x,y} - A_{x,y}\right)_i}{\frac{1}{n}\sum_{i=1}^{n}\frac{R^2 - d_i^2}{R^2 + d_i^2}}, \qquad d_i < R$$

where

C = the total correction to a grid point,
n = the number of point values affecting the grid point,
R = the radius of influence,
d = the distance between the grid point and the data point,
O = the value of the data point, and
A = the first guess value from the grid interpolated to the data point.

The correction C to a grid point is the average of the distance-weighted corrections due to all

---

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

data points within the radius of influence, divided by the average sum of the weights. (A similar correction could be made without the denominator, but the convergence to a fit is much slower because the denominator is $\leq 1$ and the corrections would be less.)

After this first pass over the data, another pass is made, where the process is the same except that the first guess A is now replaced with the analysis resulting from the previous pass. Generally four to six passes are made, the number of passes depending on the exactness of fit to the data desired; the more passes made with appropriate Rs (which vary, and are smaller for each subsequent pass), the closer the fit becomes. For wind analysis over Alaska, we use six passes.

## 3. VARIABLES ANALYZED

Wind is a vector and presents additional challenges to those for, say, temperature. Four MOS variables are analyzed–wind speed, U- and V-wind components, and "total wind," all at 10 m above the earth's surface. The total wind is defined for the NDFD and is the wind speed with the gusts added for those points where there are gusts. The two wind components are analyzed separately, and each of the four variables are treated as a scalar, like temperature. However, there are some interdependencies, as discussed later.

## 4. QUALITY CONTROL (QC) OF DATA

Any operational objective analysis method needs a robust method of quality controlling the data values. This is more important for observations, where errors can creep in from many sources. A preprocessor can be used to find gross errors by range checks and temporal changes that are not reasonable. However, spatial consistency is important, and BCDG has a rather elaborate method of ferreting out bad data. This is fully explained in Im et al. (2010, 2011). Basically, it consists of making sure the data value is not too different from the value interpolated from the previous pass of the analysis (or of the first guess for the 1st pass). If the difference exceeds a threshold which varies by pass, a buddy check is performed with the closest two stations so that stations that agree with each other are not thrown out, even if they do not at that stage of the analysis agree with the analysis. This process can be quite effective for a field that has high spatial continuity, such as sea level pressure, somewhat less so for temperature and dewpoint, and is even less effective for wind, which can be quite variable from station to station.

The variables are analyzed in order: U component, V component, wind speed, and total wind. In order to get better consistency among the four variables analyzed, any data point tossed out because of the QC process in one analysis will not be used for any subsequent analysis. Note this is not complete interdependence, but only hierarchical.

## 5. FIRST GUESS

In some applications, a constant first guess for BCDG is actually preferable to other possibilities. However, the wide expanses of ocean with few or no MOS forecasts makes it advantageous to use the 10-m direct model output as a first guess over water, the projection matching the MOS forecasts being analyzed. This is especially necessary over the Arctic Ocean where there are no MOS forecasts at all. The first guess fields for the U- and V-wind components are taken from the GFS. These components are combined and used as the first guess for both speed and total wind. This first guess over land has negligible effect on the final analysis and a constant could just as well be used. However, the large scale of the gridded GFS data available, compared to the scale of the analysis (3 km), does not provide well defined land/ocean boundaries.

## 6. VARIABLE RADIUS

Most uses of the Cressman or Barnes methods use a radius of influence R that varies only by pass; that is, each station has the same R. This is impractical for areas that have highly variable densities of stations ("stations" and "data points" are used here interchangeably). Figure 1 shows the distribution of stations in Alaska for which we have MOS wind forecasts. The highly variable distribution is apparent. Vast stretches of ocean are almost devoid of data.

A preprocessor to BCDG determines the R to use for each station for each pass over the data. It is done in such a manner that each grid point will be affected by a sufficient number of stations to give a reasonable analysis. In addition, a user can specify an R for a specific station–an override feature. Figure 2 shows a few 1st and 6th pass radii.

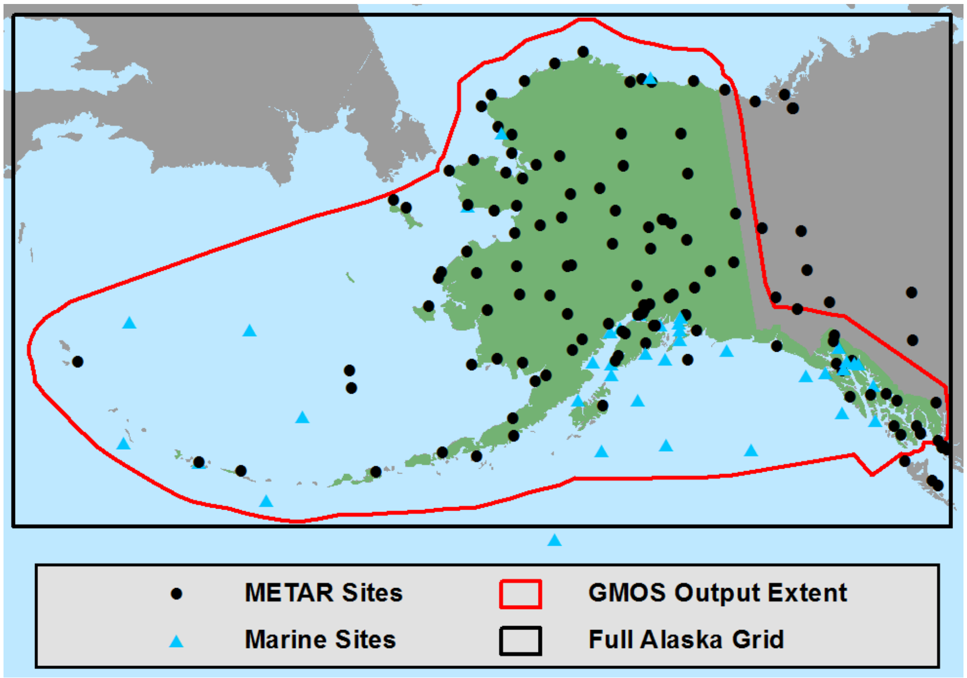

FIG. 1. The stations over the Alaska area where we have MOS wind forecasts. The analysis is done over the rectangle, but the analysis is made available only within the red-bounded area.

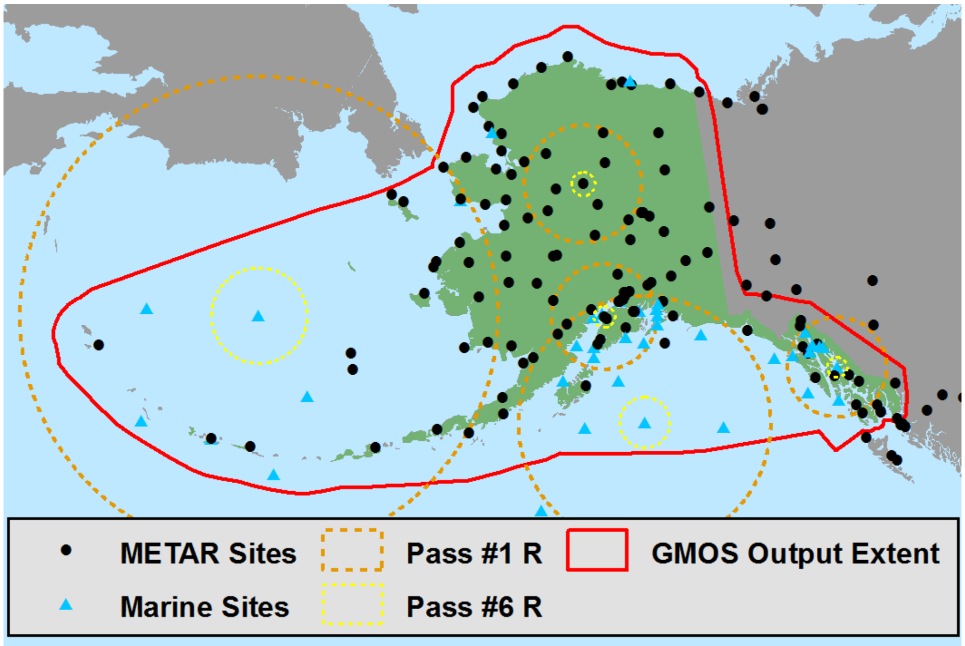

FIG. 2. The extents of the 1st (orange) and 6th pass (yellow) radii of influence are shown for a few stations. The ones over the interior were calculated by the preprocessor. The others were determined empirically based on how far the influence of the station should extend. The sizes of the circles are highly dependent on the local data density.

The ones in interior Alaska were calculated in the preprocessor; the ones over the ocean were user-specified.

The user-specified radii are determined subjectively to make sure all grid points represent the data points as well as possible, and are sometimes chosen to limit the effect of a station. For instance, no station in the Bering Sea influences the Arctic Ocean. The station-specific R for each pass is used for all four analyses–U, V, speed, and total wind.

## 7. LAND/WATER BOUNDARIES

Usually, because of the large differences in land and water temperatures, the air temperatures over water and land differ considerably in many instances. The same is true, but to a somewhat lesser extent, for wind. To account for these possible differences, two analyses are made, one over the ocean and one over land. Forecasts over land have minimal impact over the ocean, and ocean forecasts have minimal impact over land.

## 8. VARIATION WITH ELEVATION

Wind generally varies in both speed and direction with elevation. When there are mountain ridges or peaks, the wind at higher elevations is not well represented by the winds at low elevation where the data points tend to be. In areas of dense data, the data themselves can be used to calculate an expected change with elevation; this Vertical Change with Elevation (VCE) is explained in Glahn et al. (2009).

However, when the data are sparse as they are over Alaska, this process is not workable, so the winds from a numerical model, in this case the GFS, are used. The preprocessor that calculates the station-specific R (see Section 6), also calculates for each land station the highest land elevation grid point that will be affected by the station with the 1st pass R. The difference between the MOS forecast at the station and the model wind at that elevation divided by the elevation difference gives a VCE that can be used in the same way as a VCE calculated from only surface data in dense data regions. This VCE is calculated for wind speed and for both components separately. The GFS does not provide gusts, so the same VCE is used for total wind as is used for wind speed.

The VCE is used in the analysis as explained in Glahn et al. (2009). Essentially, the process is the same as explained above in Section 2 except that the change to a grid point due to a specific station is modified by the VCE and the difference in elevation between the grid point and the station. Also, the full VCE is not used for every pass, but is weighted for the six passes 1.0, 0.75, 0.50, 0, 0, and 0, respectively. In addition, because wind speed is expected to increase with elevation, negative VCEs are limited to radii of 5 grid lengths and have weights for the six passes of 0.2, 0.1, 0, 0, 0, and 0, respectively. This is just a climatological consideration; it does not keep low wind speeds observed at higher elevations from being analyzed, but in general they will not be less than winds at lower elevations. Negative VCEs can also cause negative analyzed wind speeds, and while all negative speeds are set to zero, this would be a bit unusual especially at high latitudes. VCEs for the wind components do not have these limits, because the individual components may well decrease as well as increase in elevation.

Note that by having a different VCE for U- and V-wind, a change in direction with elevation is also possible and likely, depending of course on the upper air forecast pattern.

## 9. BOGUS VALUES

The use of bogus values goes back to the very beginnings of objective analysis. A bogus value is a “manufactured” value of the variable being analyzed placed at some critical location. The manufacturing process depends on the situation. Two methods are used for wind in the Alaskan region.

### *9.1 Averages from Other Points*

This process is used in a couple of locations over the Alaska land mass and also in the Gulf of Alaska. The relatively flat area around Fort Yukon is represented by only one data point, and to get that point’s influence adequately represented, it is duplicated at four other locations. This is shown in Fig. 3. The vast expanse north of the Brooks Range is almost devoid of forecasts. Several bogus points are inserted in this area to try to capture the difference between that area, the coastal regions to the north and west, and the mountains to the south. These bogus points, weighted values from other real data points, are also shown in Fig. 3.

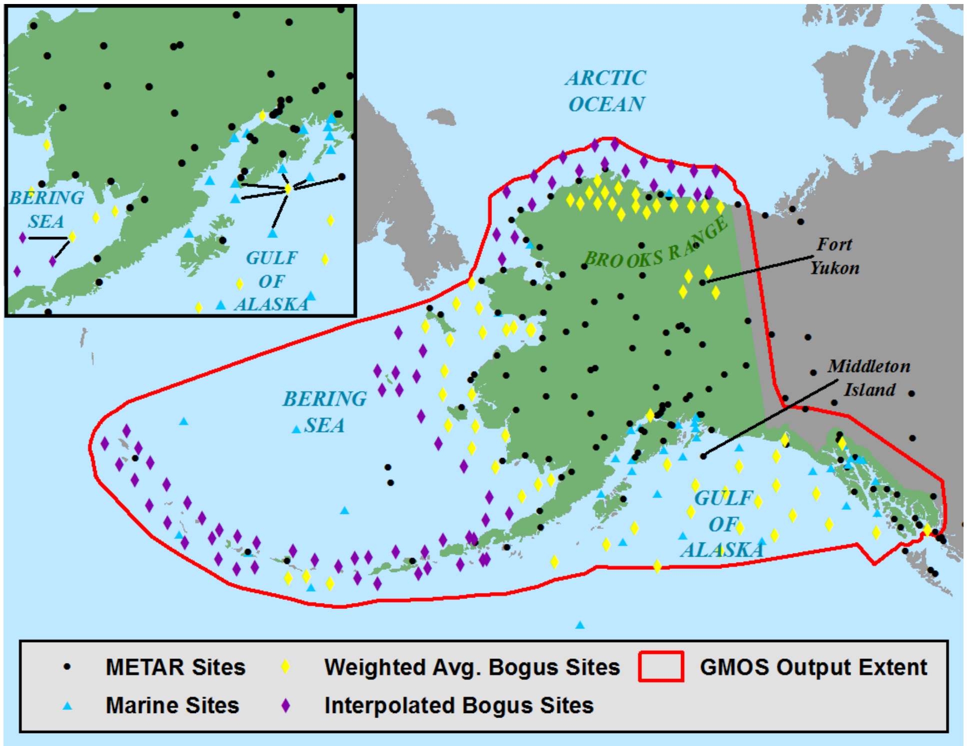

FIG. 3. The total set of points used in the analysis, including MOS at METAR and marine sites and the bogus values created either by interpolation from the model first guess or as a weighted average from MOS and/or other bogus sites. The inset shows how two bogus points were ceated from neighboring points.

The Gulf of Alaska has several buoys, but none in the central Gulf, so to adequately fair these reports into the first guess, several bogus points are located in the Gulf that are weighted averages of other points. Middleton Island is treated as a water point, because it is more representative of water than land. These points are also shown in Fig. 3.

One of the difficulties in dealing with real data is that one or more data points may be missing for a particular analysis. This is a regular occurrence for observations, but can also occur when MOS produces a forecast for a station for one time of day, but not another. So, the averaging of stations to provide bogus values must take that possibility into consideration. The bogussing algorithm calculates a value as the sum of the product of the data values and corresponding weights, divided by the sum of the weights. If a value is missing, it is "normalized" to the number of points available. For instance, if the bogus value is desired to be equal to a real value, the weight given may be 0.98, but two other real values are averaged with weights of 0.01. When all are present, the total weight is 1.0, and the bogus value will be nearly equal to the one desired real value. If the one value wanted is not present, the bogus value will be the average of the other two points–a less desirable circumstance, put probably better than having a missing value.

### *9.2 Values from the GFS*

There are no MOS forecasts over the Arctic Ocean, and very few over the Bering Sea, so GFS forecasts need to be used. However, the large scale model data available do not delineate the land/water boundaries well, and rather show a broad band of transition that extends over both the land and water. To get better western coastline definition, bogus values from the GFS are found by interpolation out away from the coast, and then bogus points closer to the coast are calculated by averaging points in this outer row unaffected in the

GFS by the land. These points are also shown in Fig. 3.

The GFS provides bogus values for wind speed and the components, but not for gusts. In order to get gusts into the grid over water, two things are done:

(1) Every water *grid point* that has a speed of > 20 kt is modified by the formula:

$$G = S[1. + Ae^{B*S}]$$

where

G = the gust speed,
S = sustained speed,
A = 0.6, and
B = -0.011

This general formula was recommended by Tattelman (1975) and the coefficients were used in the computer worded forecasts (Glahn, 1978). The coefficients undoubtedly vary with many factors, including, but not limited to, the definitions of G and S (e.g., elevation and averaging times). This formula is probably as exact as many other uncertainties in the surface wind forecasting and analysis process. The enhancement is 48 % for a 20 kt wind and is 25 % for an 80 kt wind. Especially at the lower speeds, the formula agrees well with other studies (e.g., Cook and Gruenbacher 2008).

(2) The water *bogus points* are “enhanced” by the average difference between the gust and sustained wind forecasts at the land stations in the vicinity.

The “stations in the vicinity” are provided by a preprocessor to avoid having to do searches in real time. This process is aimed at assuring the total wind over water in the vicinity of land stations is in agreement with the land stations. Note that this enhancement process is not necessary for bogus values that are formed from MOS forecasts, because whenever the MOS total wind is greater than the wind speed, this builds into the bogus points also.

## 10. INSURING CONSISTENCY BETWEEN LAND AND WATER WIND SPEEDS

For general synoptic conditions, one would expect the wind speed over water to be at least as great as over nearby land. Using a FG from a model over water without MOS forecasts does not guarantee this. For instance in the case shown, the MOS wind speeds at the two sites on St. Lawrence Island and one on Nunivak Island are stronger than the FG and the resulting analysis over surrounding water. This would be an unacceptable analysis pattern.

As a remedy, each bogus point over water that is determined from the FG is paired with one or more nearby land stations. As a preprocessing step, each bogus value is changed, if necessary, to be the maximum of itself and its paired land stations. The first guess values at the grid points are not changed, but the bogus values will correct the FG in the areas near land, and because the radii of influence are fairly large over water, the influence will be felt quite a distance from land. Figure 4 shows the analysis without this correction. The final analysis, shown in Section 13 below, is with this correction, and the wind over the ocean is at least as great as the nearby land.

## 11. SMOOTHING

Most analysis schemes have some form of smoothing after all other analysis procedures have been completed, and also sometimes between passes. This is especially important when data densities are highly variable. The successive correction process can leave blemishes between data points, that may be in the nature of arcs at the extremities of a particular station’s R at one or more passes. Usually, smoothers are very localized, a grid point becoming dependent on only a very few grid points in the immediate vicinity (e.g., see Glahn et al. 2009).

Two smoothers are used, one for land and one for water. These are quite effective for removing blemishes left by the analysis process proper and allow the data values to be fit more closely.

### *11.1 Smoothing Over Land*

For each grid point over land, the closest station is found. Then every grid point within a radius of that distance is averaged with a weight of 1/D, where D is the distance to the closest station, provided the elevation difference between the grid point and the grid point being averaged is not greater than 75 m.

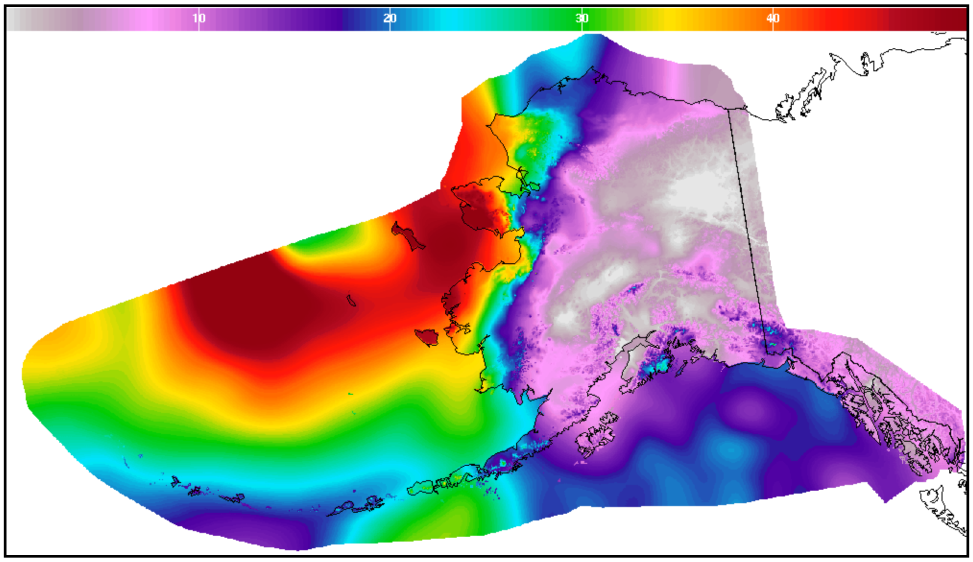

Fig. 4. The analysis before the correction showing St. Lawrence and Nunivk Islands (just off the west coast in the Bering Sea) having greater wind speeds from MOS than the surrounding water from the FG.

With special consideration very close to the station, the four grid points surrounding the station are left unchanged, while the ones far away from the station may be considerably smoothed. This provides a smooth pattern, except as interrupted by terrain features. Rather than smoothing out the terrain features put in by use of the VCE, this process can actually enhance them, even to the extent that some subsequent small, grid-length scale smoothing may be beneficial. That is, the terrain may actually be enhanced too much. For Alaska, we follow the variable distance smoothing by three passes of a terrain following smoother (Glahn et al. 2009); this makes the terrain features "softer." Because the true values of 10-m wind at high elevations are not known, whether this latter smoothing is desirable or not is largely a matter of judgment. An example of the effect of smoothing over land is shown in Fig. 5.

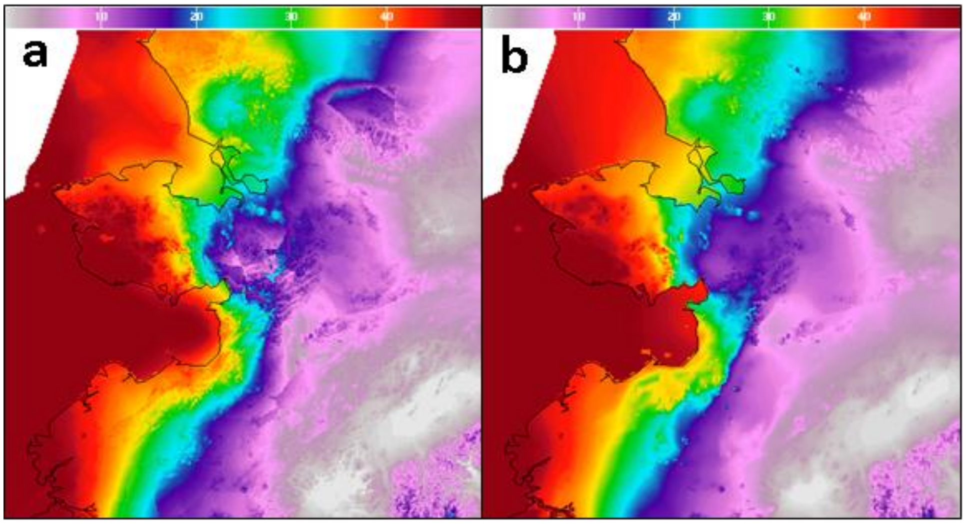

FIG. 5. The unsmoothed (a) and smoothed (b) analysis over the western part of Alaska. Some obvious blemishes have been removed without significant loss of detail.

### *11.2 Smoothing Over Water*

Smoothing over water is simpler and consists of smoothing over a circle of 30 grid length radius.

## 12. FINAL POSTPROCESSING

After the six corrective passes and smoothing, some consistency checks are made. If at any grid point the total wind is less than the wind speed, the total wind grid point is set to the speed grid point. Even though the data points already have this restriction, it is not yet guaranteed the analysis will be consistent at every grid point. In addition, any negative wind speed is set to zero. Also, the grid point closest to the station is given the station value; with the small radii of the last pass, along with sparse data, large changes are not generally made. This is done to make the gridded analysis consistent with the specific station values distributed to users by other means. Finally, the wind direction is calculated at grid points from the U and V analyses; this is the only purpose for analyzing the components.

## 13. EXAMPLES OF ANALYSES

Figure 6 shows the analysis of wind speed over the whole region. This can be compared to Fig. 4 without the consistency adjustment described in Section 10. Figure 7 shows a more detailed segment of the western coastal area with MOS forecasts and bogus values plotted.

Figure 8 shows the analysis of total wind over the whole region; Figure 9 shows a more detailed segment of the western coastal area with MOS forecasts and bogus values plotted.

Figure 10 shows wind speed and direction analysis over the whole region; Figure 11 shows a more detailed segment with MOS forecasts and bogus values plotted. Wind direction gridpoint values (the analysis) are indicated by selected grid points being plotted with a direction from the U and V analyses and the speed from the speed analysis. Actual forecasts are plotted as direction/speed.

U- and V-wind analyses are not shown; their sole purpose is to compute the direction.

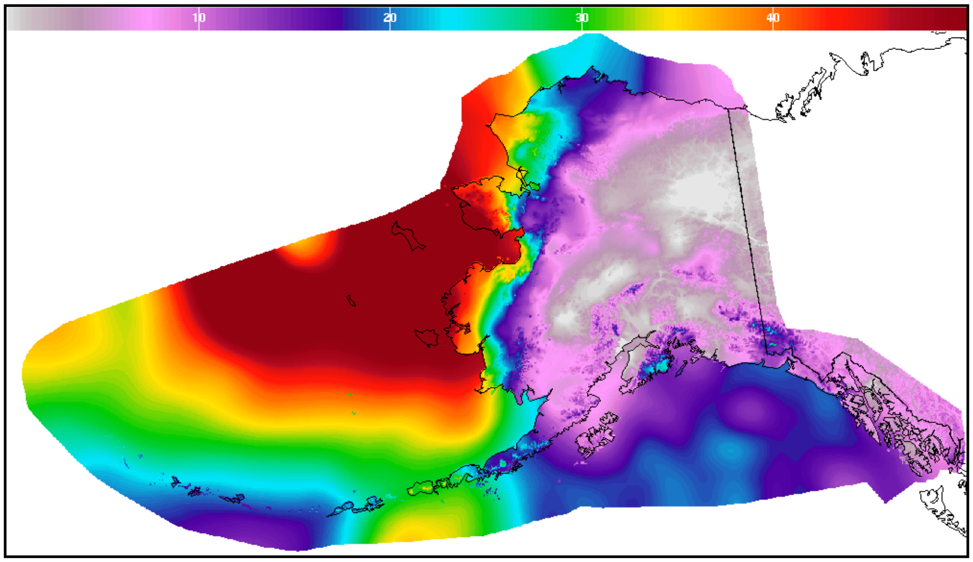

Fig. 6. The final analysis of wind speed over the Alaska area. Note the land does not have speeds greater than close-by water, and that a reasonable amount of detail is present.

## 14. SUMMARY

The capabilities of the BCDG analysis method have been demonstrated as they apply to wind over Alaska; other capabilities are used over the CONUS and the Pacific region covering Hawaii. Analysis over Alaska is very challenging, and great pains have been taken to provide a good analysis of MOS forecasts that is consistent with the station values disseminated in other ways.

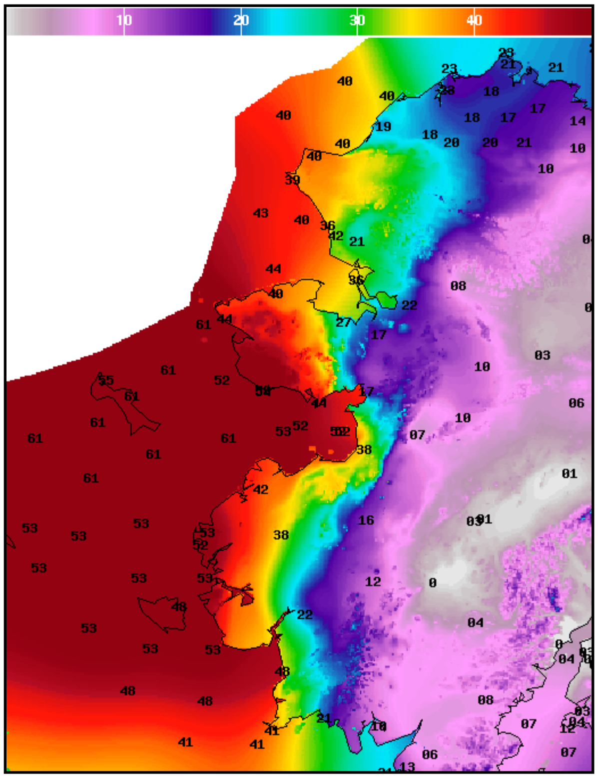

FIG. 7. A detailed segment of the final analysis of wind speed over the western coastal area of Alaska. Wind speeds at MOS and bogus points are plotted in knots.

However, limited experience has been gained either at MDL or in the field to indicate where improvement can be made. We will make improvement as needed as the Alaskan field forecasters give us feedback on the product. While the emphasis here is on MOS forecasts, the techniques can apply equally well to observations.

MOS wind analyses are available on the AWIPS Satellite Broadcast Network. This paper describes enhancements that will be implemented in the future, although final testing and refinements are not complete.

## 15. ACKNOWLEDGMENTS

The MOS wind forecasts are produced by MDL's Statistical Modeling Branch headed by Kathryn Gilbert. The analysis demonstrated in this paper will be implemented by that branch.

## 16. REFERENCES

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

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

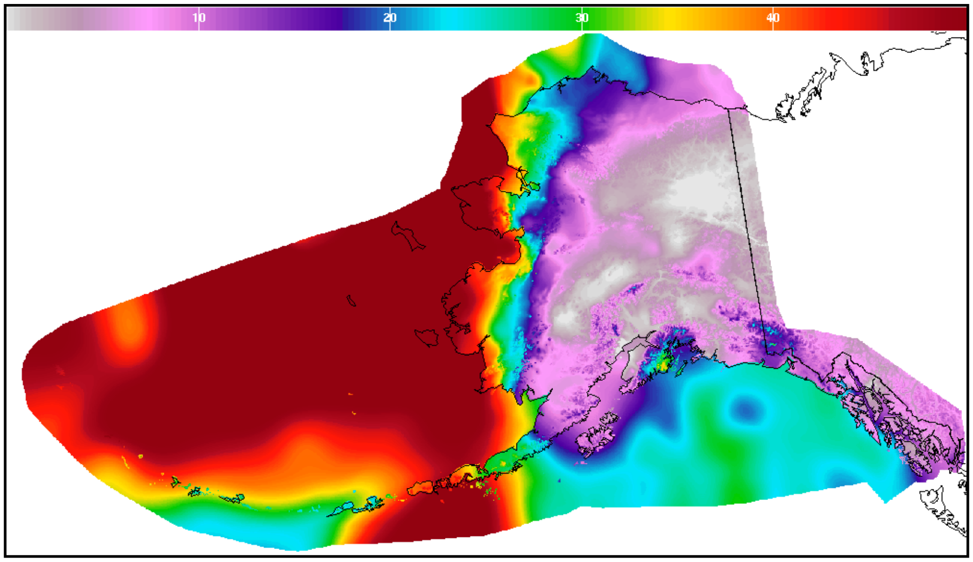

FIG. 8. The final analysis of total wind speed over the Alaska area.

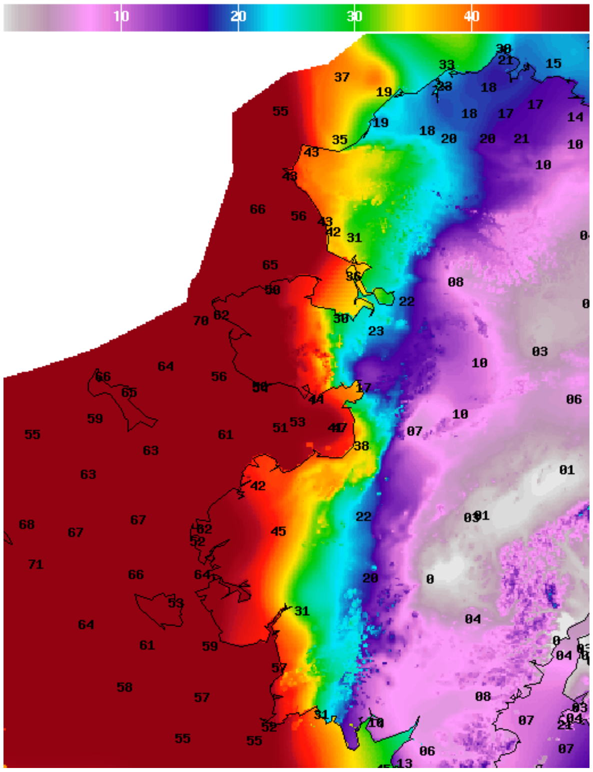

FIG. 9 A detailed segment of the final analysis of total wind speed over the western coastal area of Alaska. Total wind speeds at MOS and bogus points are plotted in knots.

Cook, K. R, and B. Gruenbacker, 2008: Assessment of methodologies to forecast wind gust speed. Unpublished manuscript, NWS Wichita Forecast Office, 6 pp.

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.

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

Glahn, B., 1978: Computer worded public weather forecasts. *NOAA Tech. Memo.* NWS TDL **67**, NOAA, U.S. Department of Commerce, 24 pp.

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

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

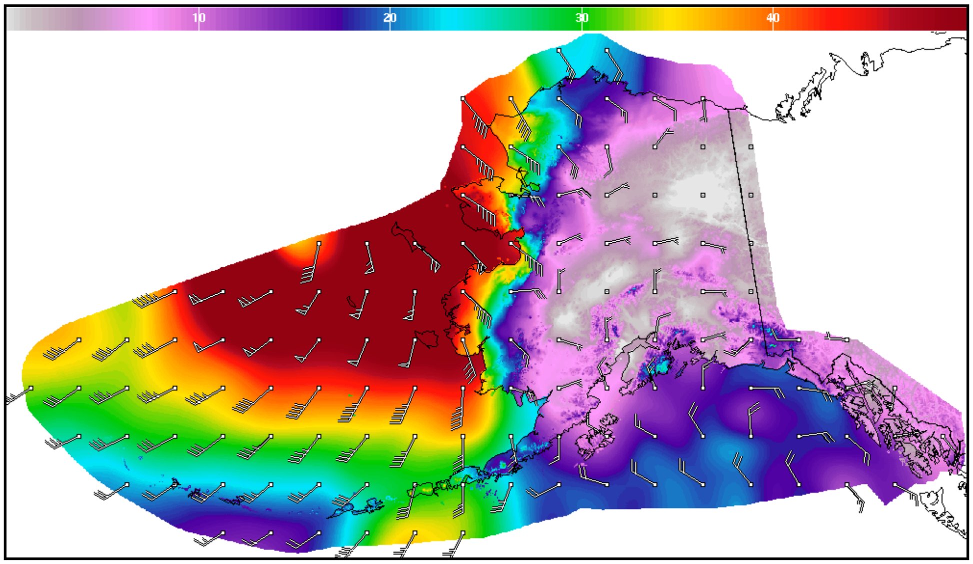

FIG. 10. The same as the wind speed analysis shown in Fig. 6, except wind barbs are plotted at selected grid points.

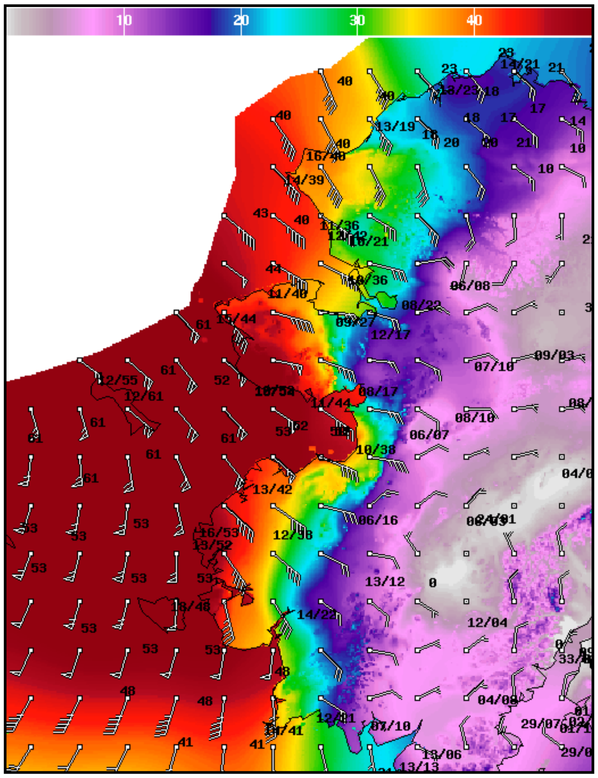

FIG. 11. The same as the wind speed analysis shown in Fig. 7, except the MOS wind directions are plotted along with the directions and speeds at selected grid points as wind barbs. Full flags on wind barbs are 10 kts. Actual MOS forecasts are plotted as direction (two digits)/speed.

Glahn, B., and J.-S. Im, 2011: Algorithms for effective objective analysis of surface weather variables. Preprints, *24$^{th}$ Conf. on Weather and Forecasting/20th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc. **10A.4**.

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

______, ______, ______, 2011: Real-time hourly objective analysis of surface observations. *NOAA Tech. Memo*. NWS MDL **84**, NOAA, U.S. Department of Commerce,19 pp.

Tattelman, P., 1975: Surface gustiness and wind speed range as a function of time interval and mean wind speed. *J. App. Meteor*., **14**, 1271-1276.