6A.2 GFS-BASED MOS OPAQUE SKY COVER AND CEILING HEIGHT FOR THE CONTIGUOUS UNITED STATES, ALASKA, HAWAII, AND PUERTO RICO

Wei Yan*
Meteorological Development Laboratory
National Weather Service, NOAA
Silver Spring, MD

Shengrong Zhao[1]
State Meteorological Administration
The People's Republic of China

1. INTRODUCTION

Sky cover and ceiling height are important weather elements in public weather forecasting, and they are also of special interest to the aviation industry. The Meteorological Development Laboratory of the National Weather Service (NWS) has developed a statistical guidance package of opaque sky cover and ceiling height by applying the Model Output Statistics (MOS) technique (Glahn and Lowry 1972) to model output from the National Centers for Environmental Prediction's (NCEP) Global Forecast System (GFS) model (Kanamitsu 1989). Sky cover is defined by the NWS as the cloud amount in eighths covering the celestial dome observed from the ground surface, and ceiling height is the height of the lowest non-surface based cloud layer when the sky cover is reported as broken or overcast. It can also be the vertical visibility into an indefinite ceiling.

The development of opaque sky cover and ceiling height MOS guidance was not done simultaneously. A hybrid approach was adopted for the opaque sky cover development to produce more skillful guidance for over 1700 stations across the U.S. Where developmental data were adequate, forecast equations were developed for single stations; regional, otherwise. Ceiling height development used only the regionalized approach. The MOS equations were developed to calculate the probabilities of five sky cover categories and eight ceiling height categories, described in Section 2. Categorical forecasts were generated from the probabilities by using probability thresholds. These thresholds were obtained from the developmental sample by using a selection scheme designed to maximize the threat score while keeping bias within a narrow range around unity.

The new MOS opaque sky cover guidance provides 3-h forecasts from 6 to 192 hours after the initial model runs at 0000 and 1200 UTC, and from 6 to 84 hours at 0600 and 1800 UTC. Ceiling height guidance is available at 3-h intervals from 6 to 84 hours in advance from all four model cycles. Both opaque sky cover and ceiling height guidance are available for the contiguous United States (CONUS), Alaska, Hawaii, and Puerto Rico.

Test results on independent data show significant improvement in forecast skill by the new MOS opaque sky cover guidance and moderate improvement by the new ceiling height guidance over the current operational total sky cover and ceiling height guidance, respectively. We attribute the progress to several possible factors: the application of the hybrid approach in the development of opaque sky cover forecast equations, the use of new monthly relative frequency data of total sky cover and ceiling height as forecast predictors in the development of ceiling height guidance, the non-simultaneous development of the sky cover and ceiling height, the availability of a longer and more stable sample of model data, and the improvement of the GFS model itself.

2. DEFINITION AND DEVELOPMENT

2.1 Predictands

Opaque sky cover is the surface-based observation of total cloud amount complemented by satellite estimates of high cloud amount (above 12,000 feet) and the Effective Cloud Amount (ECA) from Satellite Cloud Product (SCP) data (Kluepfel et al. 1994). There are five categories of sky cover, observed and reported in eighths: clear (0), few (>0 to 2/8), scattered (3/8 to 4/8), broken (5/8 to 7/8), and overcast (8/8).

**Corresponding author address*: Wei Yan, 1325 East-West Highway, W/OST22, Rm. 11364, Silver Spring, MD 20910 e-mail: wei.yan@noaa.gov

[1] Visiting scientist to the Meteorological Development Laboratory

To compute the opaque sky cover category at a station, the sky cover condition below 12,000 feet (low clouds) is first assessed with METAR (aviation routine weather report) data from Automated Surface Observing System (ASOS). Sky cover information above 12,000 feet is not available from most ASOS stations. High cloud information above 12,000 feet is obtained from SCP estimates and is modified by the ECA (in percentage) in a cascading downward scheme. If the high clouds indicate a category of overcast while the ECA is less than 66%, then the high clouds are assigned to the lower category of broken. If the high clouds are in the category broken while the ECA is less than 33%, then the high clouds are assigned the next lower category of scattered. When the high clouds are in the category of scattered and accompanied by an ECA of less than or equal to 16% while greater than or equal to 6%, the high cloud category is further lowered to few. If the high cloud category is few or scattered, and the ECA is less than 6%, then the high cloud category is further reduced to clear. Since the reduction of the high cloud amount by the cloud opacity is allowed to cascade downward by category, the highest category of overcast with a very low ECA could be lowered to a very low category, like clear, when the high clouds are very thin. This new high cloud amount complimented by SCP data is then compared to a cloud observation above 12,000 feet, if one exits (high cloud observations are reported by human observers at some METAR sites). The larger is taken as the high cloud amount. The final opaque sky cover takes the greater value of the low clouds from METAR reports and the high clouds which are computed in the scheme described above.

Cloud observation data were used, whenever available, to compute the opaque sky cover amount as the predictand data sample valid at 6 through 192 hours, at an interval of 3 hours, after model cycle starting time, 0000 and 1200 UTC, and at 6 through 84 hours after 0600 and 1800 UTC. (There were no SCP data available for any of the Alaska stations when the development was completed, thus sky cover guidance issued for Alaska is the same as that of the current operational MOS guidance, which is total sky cover by definition, for this region.)

The ceiling height in the new MOS development is defined by eight categories: less than 200 feet; 200 to 400 feet; 500 to 900 feet; 1,000 to 1,900 feet; 2,000 to 3,000 feet; 3,100 to 6,500 feet; 6,600 to 12,000 feet; and above 12,000 feet. Ceiling height observation data were processed to form the predictand data sample at 3-hourly intervals.

2.2 Predictors

The same predictors were offered in the forward selection screening process in developing the opaque sky cover and ceiling height forecast equations to promote consistency. These included GFS model forecasts, geoclimatic variables, and surface weather observations. The model based predictors, available at 3-h intervals for the developmental sample, consisted mainly of relative humidity, wind speed, u- and v-wind components, vertical velocity, relative vorticity, 3-h precipitation amount, column total precipitable water, the product of relative humidity and vertical velocity to depict the vertical transport of moisture, height of saturation level reaching 85% and 90%, and percentage of model levels reaching 85% relative humidity. These predictors were valid at certain predefined mandatory levels, e.g., 1000 mb, or for a layer average. For better opaque sky cover and ceiling height forecast of the synoptic scale, some of the small-scale variability found in model forecasts was reduced by using a spatial 25-point smoothing for all model based predictors.

The geoclimatic predictors used include monthly relative frequency of total sky cover of clear, few-scattered-broken combined, and overcast, monthly relative frequency of ceiling height below 1,000 feet, and the first and second harmonics of the day of the year. These predictors are intended to explain seasonal variations in cloud amount not accounted for by other predictors. Monthly relative frequencies used in the opaque sky cover development were based on only three years of data (April 1997 through March 2000). A new set of monthly relative frequencies for the same variables was compiled over a much longer period (January 1997 through December 2007) covering more stations and was used in the ceiling height development.

Surface weather observations were offered as potential predictors for the 6-h, 9-h, 12-h, …, and 24-h forecast equations as a means of incorporating persistence influence into the regression process. They included the observed ceiling height, non-SCP complemented total sky cover, wind speed and direction, all reported at 3 hours after

0000, 0600, 1200, and 1800 UTC. We used observations valid at 0300, 0900, 1500, and 2100 UTC because they are usually available when the GFS MOS program is run operationally.

While many of the potential predictors were continuous variables, others were either point-binary or grid-binary variables (Jensenius 1992). The point-binary technique applies a binary cutoff to the predictor valid at a specific station. The resulting value of the predictor is either 0 or 1. The observed ceiling height and total sky cover are point-binary predictors. The grid-binary technique first applies the binary cutoff at gridpoints, and then the gridded field of 1's and 0's is smoothed and interpolated to stations. The resulting predictor values range from 0 to 1. This technique provides a smoother transition, both spatially and temporally, between the extremes of the predictor than does the point-binary approach.

2.3 Seasons and Developmental Data Sample

The development of sky cover was completed in 2007 and the ceiling height in 2009, resulting in two different sample lengths. The sample used in opaque sky cover forecast equation development consisted of data from October 2000 through March 2006. All these data were stratified into two, 6-month seasons: cool (October-March) and warm (April-September). As a result, for the opaque sky cover development, the warm season had a 5-year sample while the cool season a 6-year sample for forecast projections ranging from 6-h to 84-h for the 0000 and 1200 UTC cycles, and from 6-h to 78-h for the 0600 and 1800 UTC cycles. For longer forecast projections, sample sizes varied with model cycles and seasons. Four years worth of data were available for the 87-h through 192-h projections of the 0000 and 1200 UTC model cycles for both the warm and cool seasons. For the 81-h and 84-h projections of the 0600 and 1800 UTC model cycles, data were available for 4 years for the warm season and 5 years for the cool season. For the ceiling height MOS development, 7 years of model data were available for both the warm and cool seasons (April 2002 through March 2009).

2.4 Development Methods and Regions

Prior MOS development demonstrated that single-station based forecast equations performed better than those based on regionalized development, if a long enough data sample was available. In light of this, an effort was made to maximize the benefits the single-station approach could bring us and a mixed single-station and regionalized development method was adopted in the opaque sky cover MOS development.

Among the MOS stations, about 60% were qualified as full time observation sites with enough data to develop single-station forecast equations. The remaining 40% of the stations did not have enough observations to qualify as full time. We combined them into geographical regions to compose a large data sample for each region so more stable forecast equations could be developed. To that end, we also incorporated the data of the full time stations located in the same geographical regions into the developmental sample to help the regionalized equation development.

The boundaries of the geographical regions defined in this development were the same as in the AVN-based MOS total cloud amount and ceiling height development (Weiss 2001). There were 22 warm season and 21 cool season CONUS regions, 7 Alaska regions and 1 combined Hawaii and Puerto Rico region for both seasons (Fig.1.). The same geographical regions as described above are used in the regionalized development for ceiling height.

2.5 Equation Development

Opaque sky cover forecast equations were developed every 3 hours from 6 to 192 hours after 0000 and 1200 UTC, and ending at 84-h projection for 0600 and 1800 UTC, for 1751 stations across the U.S. For the ceiling height MOS development, equation development stopped at 84-h projection for all four model cycles. There were 2280 stations across the U.S. and southern Canada involved in the ceiling height MOS development.

For the projections from 6 through 24 hours, surface weather observations, along with model variables and geoclimatic data, were used to develop a so called “primary” set of the forecast equations to capitalize on persistence. A “secondary” set of equations was also developed for the same projections, in which only the model variables and geoclimatic data were used, in order for it to be used in an operational environment when real-time observations were not available as predictors.

During equation development, we allowed the regression process to continue until a maximum of 20 predictors were chosen or until the selection of any of the remaining predictors contributed an additional half a percentage point to the reduction of variance with respect to the predictand. An equation was generated when at least 200 cases were present in the developmental sample. For most of the stations and forecast regions, surface observations and some of the model variables were among the top predictors selected for short forecast projections; for later forecast hours, geoclimatic variables were gradually picked more often as important predictors. In developing ceiling height equations, it was observed that the new monthly relative frequency of ceiling height was picked as the second most important predictor in almost every forecast equation, which was not the case in the opaque sky cover development.

2.6 Threshold Development

The cumulative probabilistic forecast values of the first four categories of the opaque sky cover from the equations, applied to the dependent data sample, were used in generating single-station threshold values for each of the categories. We found that overall station-based threshold values as opposed to regionalized thresholds provided better skill in determining opaque sky cover categories. The threshold values were generated by maximizing the threat score while keeping the bias within a very narrow range around unity. The threshold values were generated for every projection and every station for which MOS forecasts were issued.

The discrete probabilistic forecast values of the eight categories produced by the ceiling height equations, applied to the dependent data sample of a region, were used to generate threshold values for each of the categories for a station in the region. It followed the same procedure as described above.

3. VERIFICATION OF CATEGORICAL OPAQUE SKY COVER AND CEILING HEIGHT FORECASTS

For opaque sky cover, verification was conducted based on one year of independent data (April 2006 - March 2007) for 300 CONUS stations and 30 Alaskan sites selected for their quality and consistency of reporting. Comparison of Heidke Skill Score (HSS) of the forecasts was made between the newly developed system and that of the current operational MOS categorical total sky cover. The use of HSS eliminates the influence of forecasts made correctly by chance and higher HSSs indicate greater forecast skill. Our verification results showed that the new MOS opaque sky cover categorical forecasts were more skillful than the operational ones (Figs. 2, 3). For certain forecast projections, the improvement in HSS was very significant. We attribute this improvement mainly to the use of the hybrid approach in forecast equation development. The application of the single-station approach greatly enhanced the forecasting capabilities of the probability equations which were well attuned to the characteristics of the locales.

In Alaska, as noted earlier, a total sky cover system was developed and verified (Figs. 4, 5). The verification results were very promising as it demonstrated significant improvement in HSSs of the new system over the operational one for both the warm and the cool seasons.

Data from April 2008 to March 2009 were used for the ceiling height forecast verification. Forecasts were verified for the same 300 CONUS stations, 30 Alaskan sites, and 5 Hawaii and Puerto Rico sites. The HSSs plots (Figs. 6, 7) showed consistent but moderate improvement in forecast skill by the newly developed ceiling height MOS system over the current operational system. In general there is not much difference shown in HSS between the warm and cool seasons.

4. OPERATIONAL PRODUCTS

The operational MOS opaque sky cover and ceiling height forecasts are part of a more comprehensive GFS-based MOS forecast package (Dallavalle and Cosgrove 2004a, b) which is issued, as an alphanumeric message, four times a day, at 0000, 0600, 1200, and 1800 UTC. Guidance is provided for projections of 6 to 72 hours for most weather elements, including opaque sky cover and ceiling height, to the general public. Forecasts for longer projections are only available to the weather forecast community for being used internally.

Many of the NWS official forecasts are available in digital format. To support the NWS mission and the growing demand for digital weather information, MOS products are increasingly available on high resolution grids. Gridded opaque sky cover guidance is available operationally as part of the National Digital Guidance Database. Further detailed information about gridded MOS products is available in the literature (Glahn et al. 2009).

5. ACKNOWLEDGEMENT

I am grateful to have constructive discussions and in depth exchange of ideas with Mr. Mitchell Weiss during the course of preparation and development of the opaque sky cover MOS product.

6. REFERENCES

Dallavalle, J. P., and R. L. Cosgrove, 2004: GFS-based MOS guidance - The short-range alphanumeric messages from the 0000/1200 UTC forecast cycles. *MDL Technical Procedures Bulletin* No. 05-03, NOAA, U.S. Dept. of Commerce, 13 pp.

Dallavalle, J. P., and R. L. Cosgrove, 2004: GFS-based MOS guidance - The short-range alphanumeric messages from the 0600/1800 UTC forecast cycles. *MDL Technical Procedures Bulletin* No. 05-04, NOAA, U.S. Dept. of Commerce, 13 pp.

Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor*., 11, 1203-1211.

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

Jensenius, J. S., Jr., 1992: The use of grid-binary variables as predictors for statistical weather forecasting. Preprints *Twelfth Conference on Probability and Statistics in the Atmospheric Sciences*, Toronto, Canada, Amer. Meteor. Soc., 225-230.

Kanamitsu, M., 1989: Description of the NMC global data assimilation and forecast system. *Wea. Forecasting*, **4,** 335-342.

Kluepfel, C. K., A. J. Schreiner, and D. A. Unger, 1994: The satellite-derived cloud cover product (sounder). *NWS Technical Procedures Bulletin* No. 410, NOAA, U.S. Department of Commerce, 15 pp.

Weiss, M., 2001: AVN-based MOS ceiling height and total sky cover guidance for the contiguous United States, Alaska, Hawaii, and Puerto Rico. *NWS Technical Procedures Bulletin* No. 483, NOAA, U.S. Department of Commerce, 22 pp.

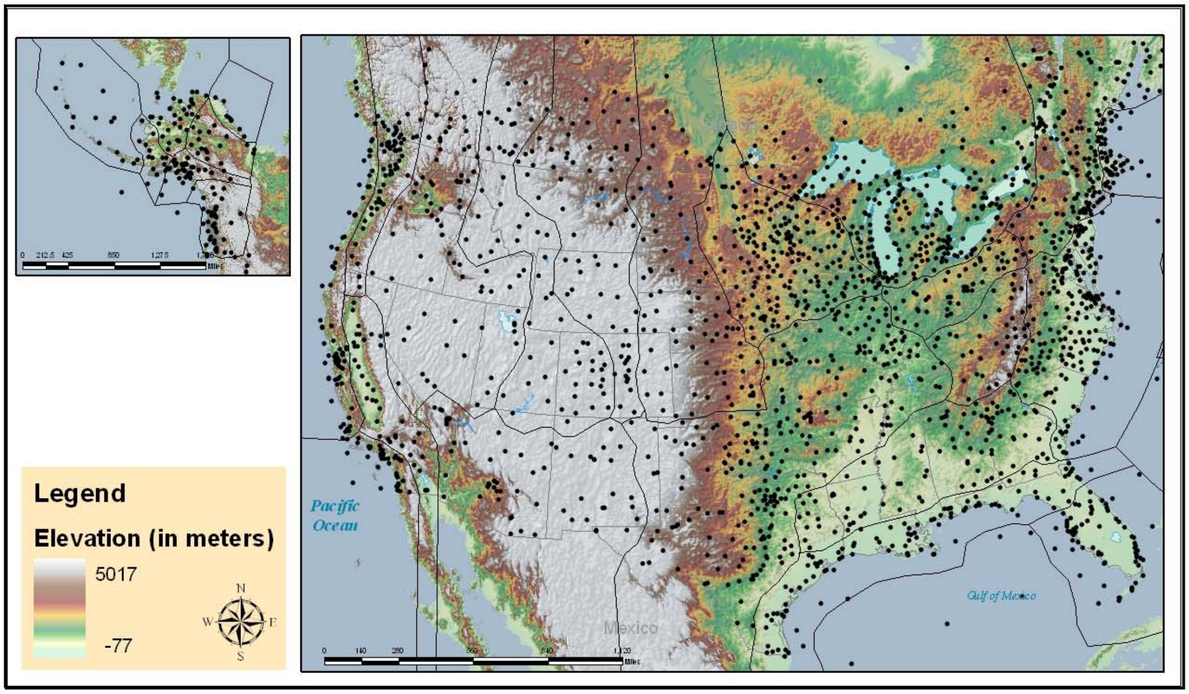

Fig.1a Regions in the CONUS and Alaska used for opaque sky cover and ceiling height development for the warm season (April - September)

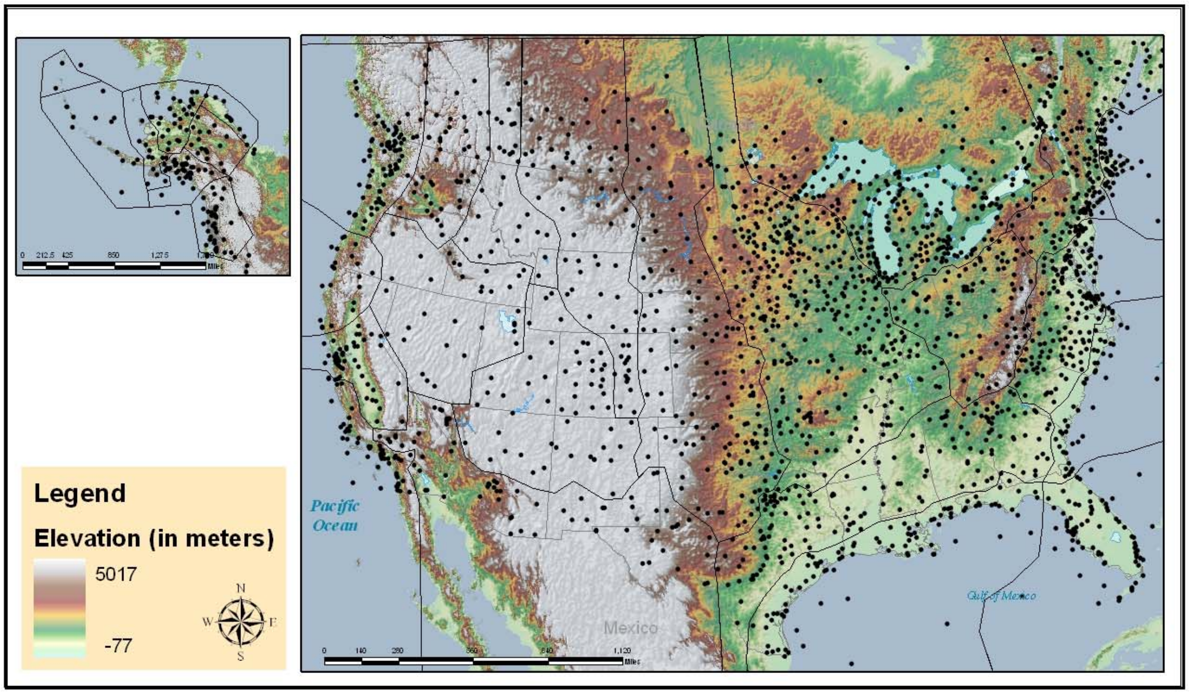

Fig.1b Same as Fig.1a except for cool season (October - March)

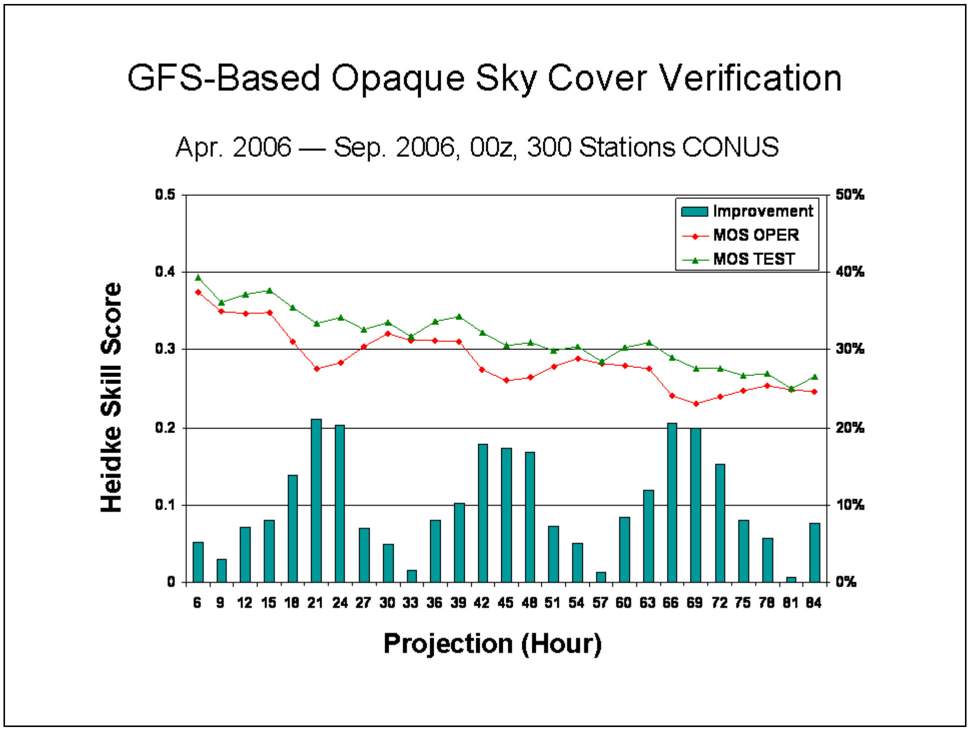

Fig. 2 Heidke skill score for opaque sky cover verification for warm season in the CONUS

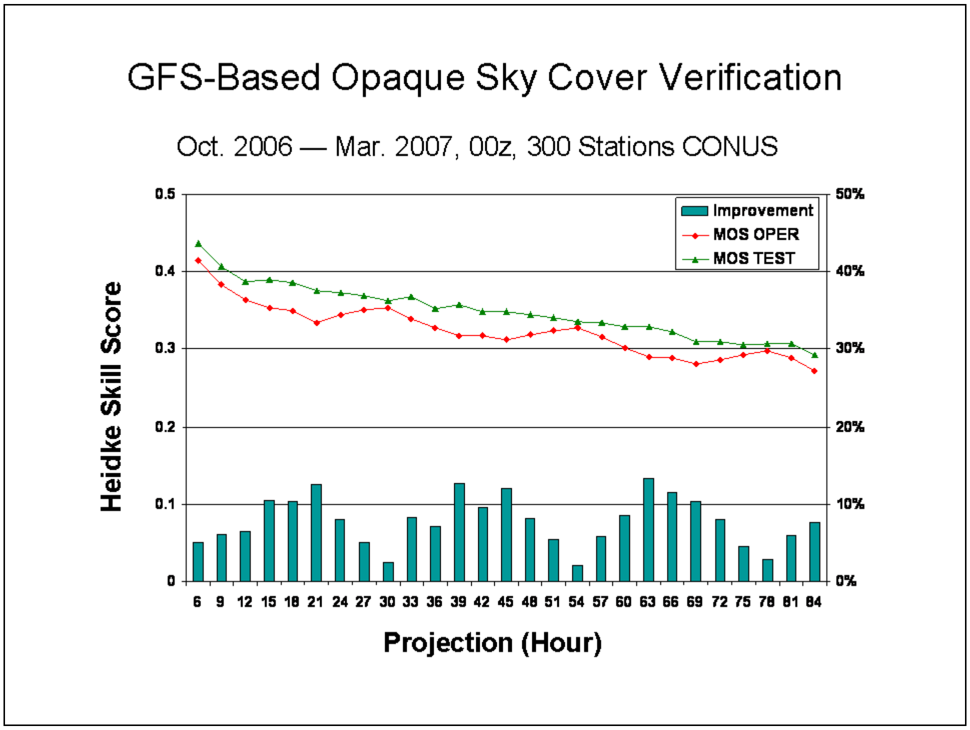

Fig. 3 Same as Fig. 2 except for cool season

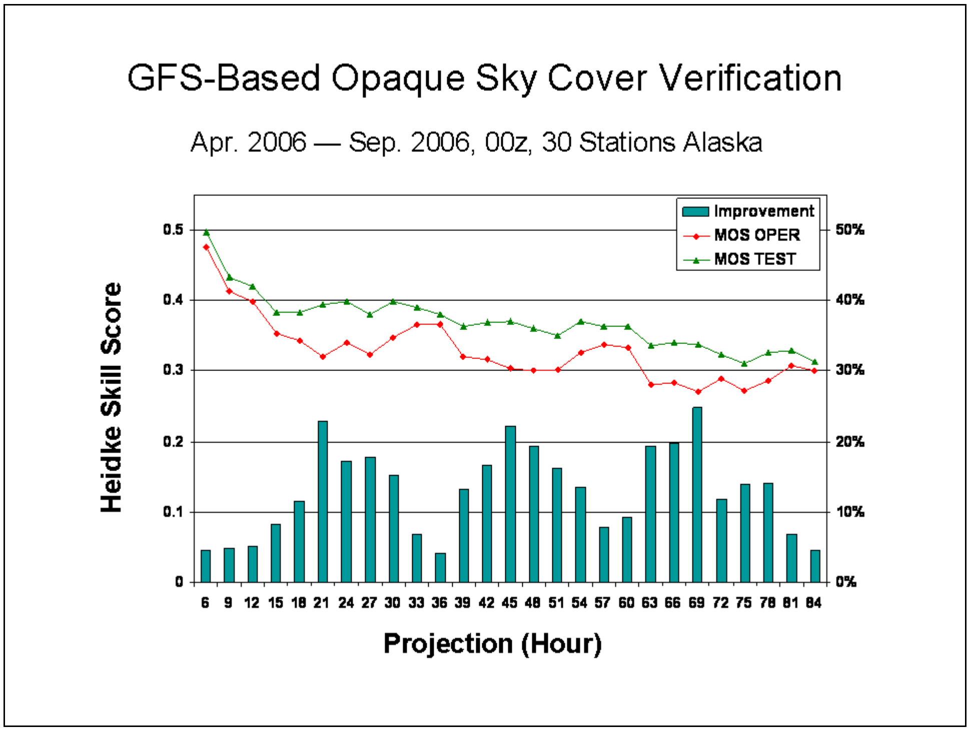

Fig. 4 Heidke skill score for opaque sky cover verification for warm season in Alaska.

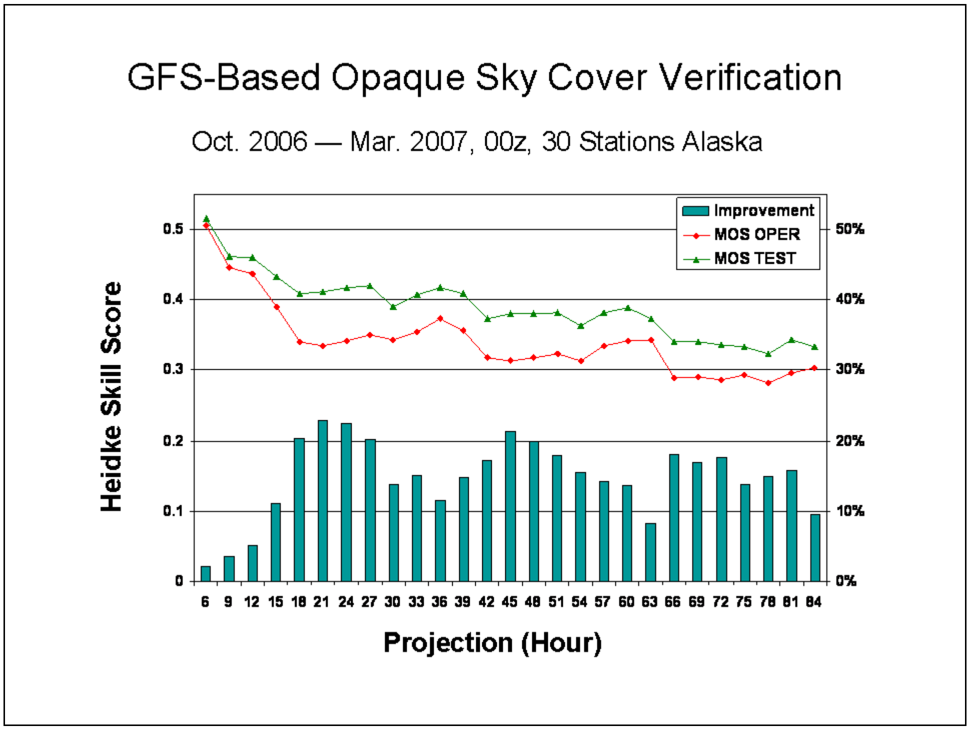

Fig. 5 Same as Fig. 4 except for cool season

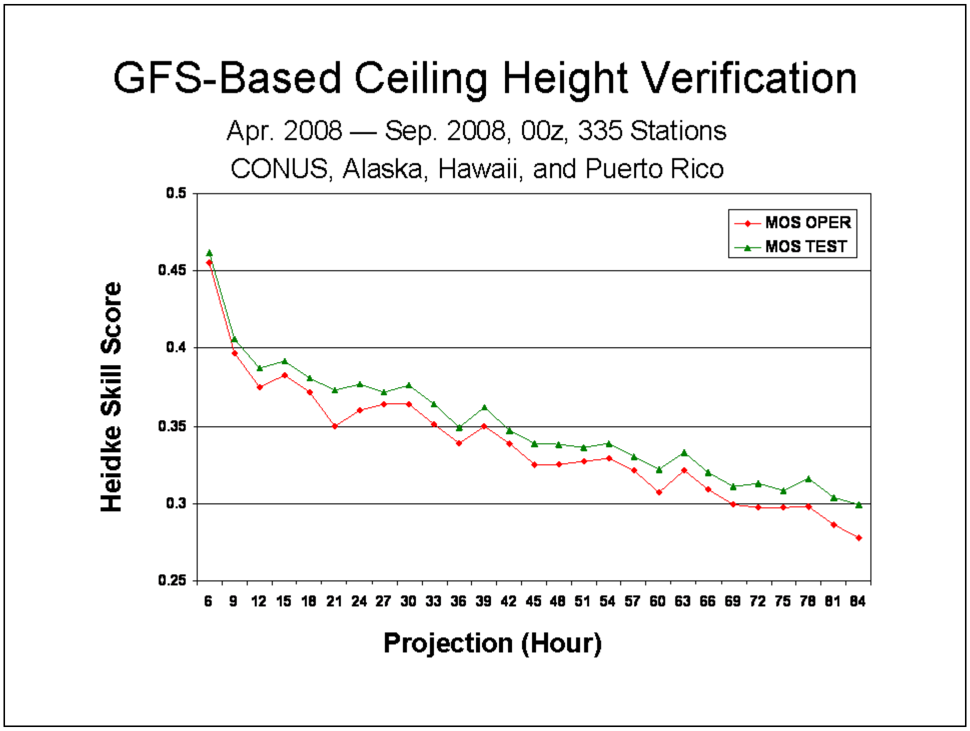

Fig. 6 Heidke skill score for ceiling height verification for warm season in the CONUS, Alaska, Hawaii, and Puerto Rico

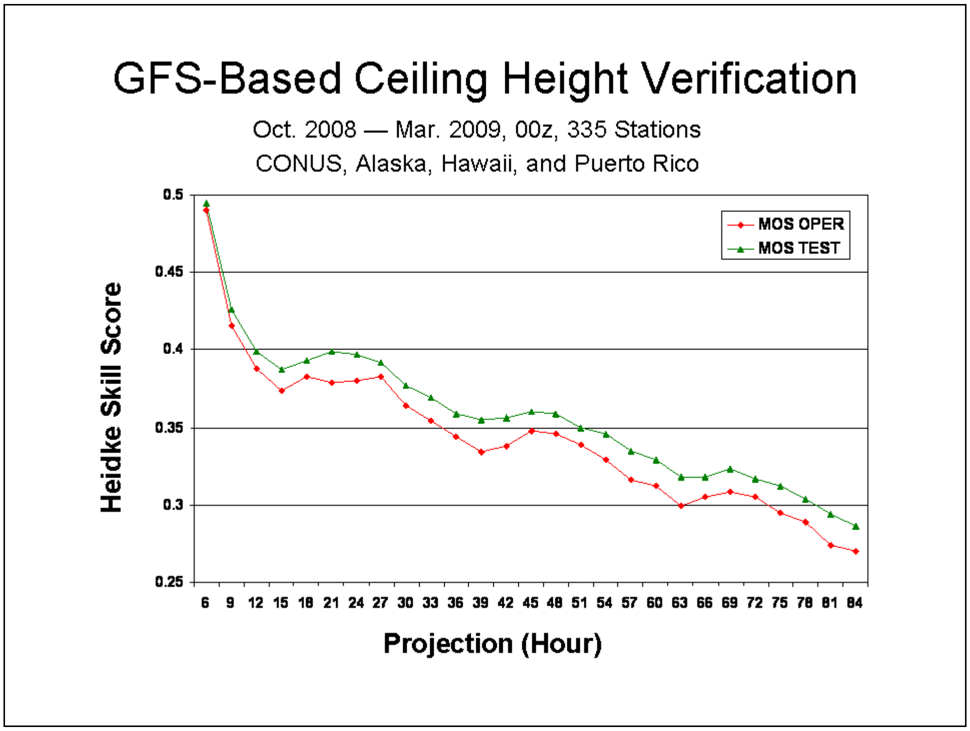

Fig. 7 Same as Fig. 6 except for cool season