Quality Assessment Report:

NATIONAL CEILING AND VISIBILITY ANALYSIS PRODUCT

Jamie T. Braid[1], Anne Holmes[1], Matthew J. Pocernich[1], Lacey D. Holland[1], Paul A. Kucera[1], Barbara G. Brown[1], Jennifer Luppens Mahoney[2]

[1]National Center for Atmospheric Research, Boulder, Colorado

[2]NOAA Research / Earth System Research Laboratory, Boulder, Colorado

Quality Assessment Product Development Team

FAA Aviation Weather Research Program

Minor changes:

15 March 2007

Quality Assessment Report:

NATIONAL CEILING AND VISIBILITY ANALYSIS PRODUCT

Jamie T. Braid, Anne Holmes, Matthew J. Pocernich, Lacey D. Holland, Paul A. Kucera, Barbara G. Brown, Jennifer Luppens Mahoney

15 March 2007

Quality Assessment Product Development Team

SUMMARY

The results of an evaluation of the National Ceiling and Visibility (NCV) analysis product (NCVA) are examined in this report. The NCVA combines surface observations with satellite data to produce analyses of current ceiling and visibility conditions. Routine observations from aviation surface reporting locations were utilized as the observation data set for verification. The NCVA was evaluated over the two periods, Summer 2006 and Fall 2006. Performance of human-generated forecasts of ceiling and visibility (AIRMETS) and 2–hr forecasts of ceiling and visibility from the Rapid Update Cycle (RUC) numerical weather prediction model are also presented in appendix to ensure the consistency of the NCVA performance with the performance of existing products. The results are summarized using standard verification statistics and presented through a variety of plots.

The main focus of this evaluation is on the performance of the NCVA at locations between the standard reporting sites. The results indicate that the NCVA has positive skill in diagnosing ceiling and visibility conditions for both evaluation periods. The NCVA only misdiagnosed reduced visibility 15% and lower ceilings 26% of the time. Comparisons of the NCVA ceiling analysis with METAR observations indicate good agreement up to 3,000 ft and a positive bias for higher ceilings. The NCVA visibility analysis has a slight high bias for visibilities less than 3 mi, which trends to a negative bias for higher visibility conditions. Both the NCVA and the AIRMET forecasts have positive skill, and have nearly equivalent skill for ceiling and visibility. The RUC 2 forecast has positive skill, with performance values that are similar to those for the NCVA.

Table of Contents

List of Figures

List of Tables

1 INTRODUCTION

The National Ceiling and Visibility Product Development Team has developed a product to analyze ceiling and visibility conditions on a grid across the continental U.S. (CONUS). This report summarizes the evaluation of the diagnostic capabilities of the National Ceiling and Visibility (NCV) analysis field (NCVA) as part of the process of determining whether the product should be granted D4 (operational) status through the Federal Aviation Administration (FAA) and National Weather Service (NWS) Aviation Weather Technology Transfer (AWTT) process. This study is a continuation of the evaluation of the NCVA performed and documented by Fowler et al (2005).

The NCVA was evaluated using the ceiling and visibility flight rule categories defined by the Federal Aviation Administration (FAA). The evaluation was conducted using data collected during summer 2006 and fall/early winter 2006, and included an evaluation of the NCVA across the CONUS and for four separate geographical regions (See Figure 1). The primary focus of the evaluation is on the ability of the NCVA to a diagnose flight category. In addition, the performance characteristics of the ceiling and visibility components of the NCVA grids are considered individually. Since the NCVA is an analysis of ceiling and visibility and not a forecast, the scores at the observation points should be close to perfect. Thus, this evaluation focuses on performance at locations between observation sites (where the NCVA values have been interpolated) using a cross-validation method that is explained in detail in Section 3. The CONUS verification statistics are compared to the statistics for the AIRMETs and for the 2-hr forecasts provided by the Rapid Update Cycle (RUC) operational model. Verification statistics are also presented for operational forecasts (i.e., Airmens' Meteorological Advisories, or AIRMETs) and for two-hr forecasts from the Rapid Update Cycle (RUC) operational model; these results, which are presented in an appendix, are included only to show whether the NCVA performance is consistent with the performance of other available products. This comparison is not intended as an evaluation of either the AIRMET sot the RUC 2-hr forecasts.

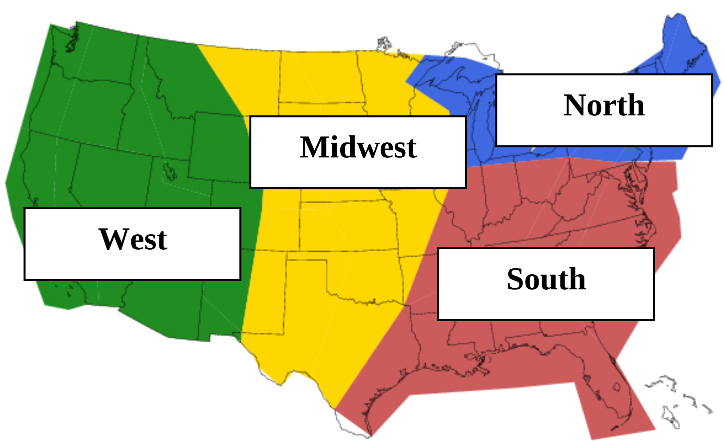

Figure 1: Map showing the four sub-regions.

Section 2 of the report describes the data used in the evaluation. The methodology is presented in Section 3, with the results presented in Section 4. Section 5 contains a summary of the conclusions. The evaluation of ceiling and visibility forecasts from the AIRMETs and 2-hr RUC forecasts are presented in the appendices.

2 DATA

For this study, the NCVA hourly analysis and surface ceiling and visibility observations (METARs) over the CONUS are examined for two separate periods, 29 June through 28 July 2006 (Summer-2006) and 1 September through 20 December 2006 (Fall-2006). In addition, RUC operational 2-hr forecasts and the AIRMETs are examined over the same period and their performance is reported in the appendix. Although these products are very different from the NCVA, it is important to consider the performance of existing capabilities when evaluating a new capability. All data used in the evaluation are described in more detail in the following subsections.

2.1 NCVA

The NCVA combines satellite and surface observational data to produce an analysis of ceiling and visibility conditions on a grid across the CONUS. Using a two-dimensional grid corresponding to the 5-km horizontal grid structure of the National Weather Service's National Digital Forecast Database (NDFD), the NCVA algorithm produces a grid of ceiling and visibility values by interpolating the satellite and surface observations. In particular, the NCVA uses ceiling and visibility observations to determine ceiling and visibility values at the observation sites, and then applies an interpolation scheme to estimate ceiling and visibility values between sites. The product is updated each 5 minutes as new satellite information becomes available. For this report, only the analyses produced on the hour are considered. An example of an NCVA grid is shown in Figure 2.

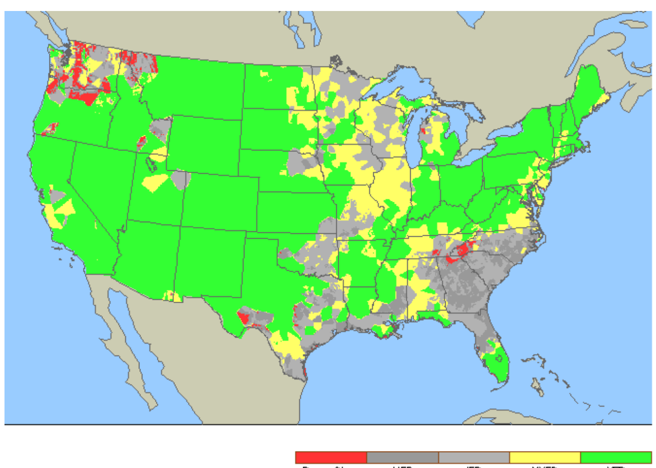

Figure 2: Example of an NCVA grid.

Once the ceiling and visibility values have been produced, they are converted to fit into the flight rules listed in Table 1. The "worst" of either the ceiling or visibility conditions are used to determine the flight rule. As an example, if the ceiling is 1,000 ft and the visibility is 5 mi, IFR flight rules would be applied because the ceiling value is the "worst" of the two values (1,000 ft equals IFR).

Table 1: Flight Rules and Associated Ceiling and Visibility Limits.

Flight Rules (FR)	Ceiling (ft)	Visibility (mi)
Visual (VFR)	> 3000	> 5
Modified Visual (MVFR)	< or = 3000	< or = 5
Instrument (IFR)	< or = 1000	< or = 3
Low Instrument (LIFR)	< or = 500	< or = 1

2.2 METARS

Surface-based observations of ceiling and visibility obtained from METARs (Aviation Routine Surface Weather Reports) are used to evaluate the NCVA. METAR observations are taken at a minimum of once per hour with the possibility of more than one in an hour if special weather conditions exist. Ceiling conditions are measured in hundreds of feet near the surface, changing to thousands of feet for higher altitudes. Visibility conditions are measured in fractions of a mile for lower levels of visibility, increasing to full miles as visibility increases. The map in Figure 3 shows the locations of the approximately 1,700 METAR stations across the CONUS.

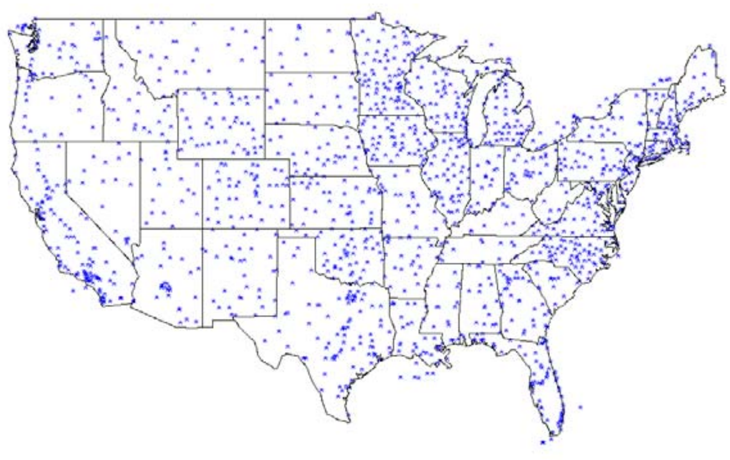

Figure 3: Distribution of METAR stations over the CONUS.

METAR values can be observed by human observers or by automated instruments. Although the current trend is to convert all METAR stations to the ASOS (Automated Surface Observing Systems) format (Bradley and Imbembo 1985; US DoC 1992), some stations in the network still use human observations of ceiling and visibility. This mix of observation methods leads to the possibility of inconsistencies in the ceiling and visibility values. The possibility of this inconsistency must be noted and taken into consideration when interpreting the results.

2.3 AIRMETS

AIRMETs are operational aviation advisories that are issued for a 0-to-6-hr period when IFR (ceilings below 1,000 ft and/or visibility below 3 mi) or worse flight rules are expected (NWS 1991). AIRMETs are issued every 6 hr and must cover an area of at least 3,000 sq mi. However, if conditions change during the time period, an AIRMET can be amended or completely cancelled. Figure 4 shows an example map of IFR and Mountain Obscurity (MTOS) AIRMETs from the ADDS website (http://adds.aviationweather.noaa.gov/). AIRMETs are included in this evaluation (in the appendix) *only for the purpose of judging whether the performance of the NCVA is consistent with existing capabilities; the goal is not to provide an evaluation of the AIRMETs themselves.* AIRMETs are only included in this evaluation because they are an operational product that could be used to make similar decisions as will be made with NCVA.

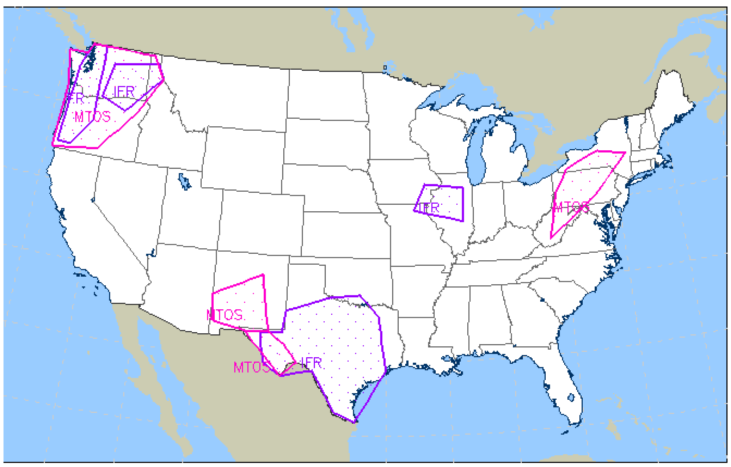

Figure 4: Example of AIRMETs for ceiling and visibility.

2.4 RUC operational 2-hr Ceiling and Visibility Forecast

The RUC operational 2-hr ceiling forecast is a METAR-augmented version of the RUC cloud analyses product. The RUC cloud analyses product combines data from the 1-hr RUC explicit 3-D hydrometeor (cloud water, rain, ice, snow graupel) forecast, and GOES/NESDIS cloud top product (pressure and temperature) to define the location of cloud tops. To produce cloud base information, the RUC cloud analysis product is then augmented with ceiling data from METAR stations to provide a better 3-D picture. A more complete description of the RUC operational 2 hr ceiling forecast is provided in the paper by Benjamin (2004).

The RUC operational 2-hr visibility forecast is a combination of Stoelinga-Warner (1999) algorithm using MM5/RUC hydrometeor mixing ratios and an RH-based factor that gives 5-mi visibility at a 95% RH and 50-mi visibility at a 15% RH. The RUC visibility algorithm gives the minimum of the modified Stoelinga-Warner algorithm value and the RH-based factor algorithm (NOAA 2006). The RUC forecasts, which could be used to make similar decisions as the NCVA, are included in this evaluation only for the purpose of judging whether the performance of the NCVA is consistent with existing capabilities; the goal is not to provide an evaluation of the RUC. Verification results for the RUC forecasts are presented in the Appendix.

3 METHODS

3.1 Matching the gridded NCVA to METAR sites

For this evaluation, the NCVA grids are matched to the METAR observations. For each METAR location used in verification, the minimum ceiling and visibility measurements from the four surrounding NCVA grid points are selected. In particular, for each variable (ceiling and visibility), the minimum value from one of the four surrounding grid points on the NCVA grid is matched to the ceiling/visibility measurement from the METAR station observation to create a verification pair. The ceiling and visibility measurements may come from different grid points. Since the NCVA is on the NDFD 5-km grid, the maximum distance between a METAR site and its matching grid location for the NCVA is less than 8 km. A time window of ±15 minutes around the top of the hour is used to match METARs to the NCVA; for locations where more than one METAR is reported within the time window, the most extreme value (i.e., the one representing the worst conditions) is used.

3.2 Cross-validation

Because the NCVA uses METAR observations to determine ceiling and visibility values at the METAR sites, verification using the same METAR observations as were used to create the analysis would produce verification statistics that are nearly perfect. Differences would result only from possible timing errors of the METARs, the data collection process (i.e., the list of stations used to create the NCVA may not exactly match the list of stations used in the verification process), or differences introduced by the satellite data. Thus, the goal of this study is to evaluate the performance at analysis locations between METAR locations. In order to do so, it was necessary to randomly leave some METAR observations out of the NCVA analysis procedure, so that these observations would be available for use in evaluating the performance of the NCVA. A cross-validation approach (Neter et al. 1996) makes this evaluation possible, and was applied in the evaluation of the NCVA gridded analyses.

The cross-validation technique involves randomly removing a small percentage (20% in this study) of the METAR stations from the process used to produce the NCVA; these held-out stations are then used in the verification process. The set of held-out METAR stations is called the “testing set”. The remaining stations, which are called the “training set,” are used to produce each NCVA. To mitigate the chance of getting a bad draw of METAR stations (e.g., all testing set stations located in the northeastern United States), which could in turn affect the statistical results, this procedure was repeated 10 different times with a NCVA grid produced each time. With hourly NCVA grids, this process resulted in 8,400 NCVA grids for the Summer-2006 period and 26,640 NCVA grids for the Fall-2006 period that were available for verification. The verification statistics are based on the ten testing sets of METAR reports, accumulated across all of the NCVA analyses included in the verification sample.

3.3 Verification statistics

Overall verification statistics were calculated based on binary event/non-event categories. The four flight categories listed in Section 2.1 were condensed into two categories by combining the bottom and top two categories, yielding the categories "IFR or worse" and "MVFR or better." The verification statistics computed include the probability of detection (POD), the probability of detection for non-events (PODn), Bias, and the False Alarm Ratio (FAR). In addition, three skill scores are included: (a) the Heidke Skill Score (HSS), (b) the Gilbert Skill Score (GSS), and (c) the True Skill Statistic (TSS). The average percent of the CONUS covered by an NCVA of IFR or worse (Percent Area) is also considered. Finally, the POD per unit area, known as Area Efficiency, is also included. Each statistic is calculated using the formulas listed in Table 3, based on a standard 2x2 contingency table as shown in Table 2.

The actual ceiling and visibility values are examined separately as well. In particular, the bias in the NCVA ceiling and visibility values is assessed. Boxplots, histograms, and a contour plot (essentially a 3-dimensional scatter plot) are used to examine errors in and agreement between NCVA and METAR values. Quantile-quantile (Q-Q) plots are used to compare the distributions of NCVA versus METAR values. Linear models are overlaid on the Q-Q plot to quantify the differences in distributions.

Table 2: Standard 2x2 contingency table for verification statistics. Entries in the table represent counts of each forecast/observation pair.

	METAR Flight Category	
NCVA Flight Category	IFR or worse	MVFR or better
IFR or worse	YY	YN
MVFR or better	NY	NN

Table 3: Verification statistics and associated formulas based on counts from Table 2.

Statistic	**Formula**
POD	YY / (YY + NY)
PODn	NN / (NN + YN)
Bias	(YY + YN) / (YY + NY)
FAR	YN / (YN + YY)
HSS	(YY + NN – C1) / (N – C1) { where C1 = [(YY + YN) (YY + NY) + (NY + NN) (YN + NN)] / (YY + YN + NY + NN) }
GSS	(YY– C2) / (YY - C2 + YN +NY) [where C2 = (YY + YN) (YY + NY) / (YY + YN + NY + NN)]
TSS	POD + PODNo -1
Percent Area	Average Event Area * 100 / Total CONUS Area
Area Efficiency	100 * POD / (Percent Area)

4 RESULTS

4.1 Flight Category results from the cross-validation analyses

Table 4 shows the verification statistics obtained for the NCVA for the Summer-2006 time period using the cross validation technique. Table 5 shows (for the same time period) the scores that would be obtained if all the METARs available were used in both the production of the NCVA and the verification scoring process[1]. The values in Table 5 also represent the best verification scores that can be obtained by the NCVA using the data available. METAR data often have inconsistent sampling and are not uniform in spatial coverage. These METAR limitations along with additional temporal and spatial limitations from the satellite data lead to verification scores that can only be "practically" perfect. This idea of "practically" perfect forecasts is described in more detail in Brooks et al. (1998). Using this idea of "practically" perfect forecasts, the scores obtained using the cross-validation technique will not be compared to a perfect value for each verification statistic (i.e., 1.00 for POD) but instead will be compared to the values obtained from the "practically perfect" adjustment of NCVA method described above.

Over the entire domain (labeled CONUS in the tables) during the Summer-2006 time period, the NCVA has adjusted POD and PODn values of 0.523 (0.399/0.763) and 0.982 (0.972/0.990), respectfully. The adjusted FAR of 0.329 is a little high for this time period. The high FAR for the Summer-2006 time period could be due to the variability in the types of weather systems that occur. The NCVA has positive skill as indicated by the HSS, GSS and TSS values. The Bias of 0.0.782 for the cross-validated sample indicates that the NCVA identifies IFR-or-worse conditions around 22% fewer times than they occur, but even in the "perfect" case the Bias is less than 1 (0.910). The average NCVA IFR-or-worse covers about 2% of the area.

In the smaller regions, the values of the verification measures are similar to what was found for the CONUS, with the North region recording a larger adjusted POD value (0.483/0.806 = 0.599) value than the other regions. The PODn values are quite similar across all of the sub-regions. The NCVA is skillful in all of the sub-regions, and all but the west region have comparable bias values.

Table 4: NCVA results for the Summer-2006 period using cross-validation technique.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.399	0.972	0.490	0.782	0.413	0.260	0.371	2.303	17
North	0.483	0.952	0.410	0.818	0.471	0.308	0.435	6.163	29
South	0.353	0.961	0.548	0.781	0.349	0.212	0.315	3.455	8
Midwest	0.363	0.986	0.545	0.797	0.387	0.240	0.349	1.259	10
West	0.345	0.986	0.471	0.652	0.397	0.248	0.332	1.093	32

Table 5: NCVA scores for the Summer-2006 period using all the observations in the analysis.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.763	0.990	0.161	0.910	0.786	0.647	0.753	2.333	33
North	0.806	0.982	0.137	0.934	0.810	0.681	0.788	6.246	55
South	0.727	0.986	0.186	0.893	0.750	0.599	0.713	3.513	13
Midwest	0.697	0.995	0.183	0.853	0.745	0.594	0.692	1.265	20
West	0.821	0.994	0.140	0.955	0.832	0.713	0.814	1.105	74

[1] Note that these statistics are not perfect, due primarily to the time window used to collect METARs used in the evaluation.

During the Fall-2006 season, the adjusted POD and PODn values for the CONUS are 0.685 and 0.975 respectively (Tables 6 and 7). The Bias for the cross-validated set is 0.876, compared to 0.962 when all stations are included. The NCVA shows positive skill and covers an area roughly 6% of the CONUS.

The Midwest and North have the largest adjusted POD values (0.783 and 0.708) while the West has the smallest (0.546) adjusted POD value. All regions have similar PODn values and all regions show positive skill. The percent areas range from a high of just over 10% in the North to a low of around 5% in the West. The reduced forecast skill over the Western region may be partially influenced by the weather regimes that exist in that region.

Table 6: NCVA results for the Fall-2006 period using the cross-validation technique.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.574	0.959	0.344	0.876	0.564	0.393	0.534	5.997	10
North	0.602	0.942	0.325	0.893	0.569	0.398	0.545	10.006	11
South	0.536	0.957	0.351	0.827	0.532	0.363	0.493	6.168	6
Midwest	0.661	0.971	0.283	0.934	0.651	0.483	0.632	6.265	9
West	0.460	0.959	0.463	0.856	0.447	0.288	0.419	4.971	9

Table 7: NCVA scores for the Fall-2006 period using all the observations in the analysis.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.838	0.984	0.128	0.962	0.836	0.718	0.822	6.018	14
North	0.850	0.975	0.128	0.975	0.834	0.727	0.826	10.192	13
South	0.821	0.982	0.128	0.941	0.824	0.715	0.803	6.240	8
Midwest	0.844	0.987	0.131	0.972	0.842	0.701	0.831	6.315	13
West	0.843	0.987	0.127	0.965	0.843	0.729	0.830	4.849	17

4.2 Categorical evaluation of ceiling and visibility

The flight rules associated with ceiling and visibility are dictated by the more restrictive of two conditions: vertical ceiling level and horizontal visibility level. Hence, in actuality only one of the two values needs to be correct for a correct decision to be made. In this section, we examine the joint probability of one or more of the parameters being correct. Initially, NCVA values and METAR observations are binned into two categories based on whether or not flight rules apply. The contingency tables in Table 8 and Table 9 show the performance of the NCVA with respect to each parameter. From 2x2 contingency tables like these, many verification statistics can be computed, as shown in the previous section.

For visibility (Table 8), about two-thirds of the time the observed visibility was greater than 5 mi. When the observed visibility was less than 5 mi, there was a 56% (19/34 x 100) chance that the associated NCVA visibility value was correct. Restricted visibility was identified by NCVA 33% of the time. From a safety perspective, 15% of the time low visibility was observed but not identified by NCVA (highlighted in red in Table 8). Table 9 contains similar information for the ceiling values produced by NCVA. For the ceiling analyses, low ceilings were identified by the NCVA 51% of the time but were observed 68% of the time. Twenty-six p percent of the time low ceilings were observed, but were not identified by NCVA (highlighted in red in Table 9), which is slightly larger than the percentage associated with the visibility analyses.

Table 8: Summary of the two-category binning of the NCVA visibility values. The entries in the table represent the percentage of occurrence each category.

		NCVA visibility		
		0-5-mi	5 mi - unlimited	Total
Observed visibility	0-5 mi	19	15	34
	5 mi - unlimited	14	53	67
	Total	33	67	100

Table 9: Summary of two-category NCVA ceiling values. The entries in the table represent the percentage of occurrence each category.

		NCVA ceiling		
		0-3 K ft	Greater than 3 Kft	Total
Observed ceiling	0-3 Kft	42	26	68
	Greater than 3Kft	9	22	31
	Total	51	48	100

Since warnings are issued using both fields, we need a way to combine the information from the two parameters. The NCVA ceiling and visibility values were turned into ordinal values by assigning the lowest visibility or ceiling bin a value of 1 and the higher category a value of 2. Using this numbering convention, we created an index of values representing both the NCVA value and the observed value. For example, a value of 11 for visibility means that the observed visibility was less than 1 mi and the NCVA visibility value was less than 1 mi. Coding the observations and forecasts in this manner allows us to examine the effects of the ceiling and visibility parameters together. It is important to remember that small values for the NCVA values are more restrictive and represent conditions that are potentially more dangerous.

Contingency tables can readily summarize performance related to the combined NCVA categories. To illustrate how these tables can be useful, consider the following examples. In Table 10, yellow indicates regions where only the NCVA ceiling value was correct, blue indicates regions where only the visibility value was correct and green indicates where both parameters were correct. Table 11, Table 12 and Table 13 show regions where observed visibility and observed ceiling or either observed visibility or ceiling were restrictive. This leads to the conditions defined in Table 14. "Danger" indicates situations where a restrictive ceiling or visibility value was not indicated but was observed. In the cell labeled "Danger 1", both NCVA parameters were wrong. In "Danger 2" the unrestricted visibility was correctly identified by NCVA, but the more critical ceiling condition was missed. The opposite condition is represented in "Danger 3." "FP" indicates situations in which one or both of the values was incorrectly restrictive. "W" indicates situations in which an incorrect NCVA value for an unrestricted condition resulted in the correct warning.

Table 15 presents results from the binning of NCVA values into two categories. Assuming that the center cells indicate correct values, the table shows that 44.5% of the time both the ceiling and visibility values provided by NCVA were correct. Dangerous conditions (highlighted in red) existed 26.5% of the time and false positive events (highlighted in green) occurred 12.5% of the time. The marginal values can be used to relate these values with those in Table 8 and Table 9. These results only represent a simple two category evaluation, but could be expanded to examine multiple categories of conditions if desired.

Table 10: Categorical table of visibility and ceiling values. Blue indicates correct ceiling values; yellow indicates correct visibility values and green indicates both parameters are correct. "12" indicates a low observed value and high NCVA value; "11" indicates both the observation and the NCVA values were low; "22" indicates both were high; and "21" indicates the observation was high and the NCVA value was low.

		Visibility				
		12	11	22	21	Totals
Ceiling	12					
	11					
	22					
	21					
	Total					

Table 11: Red indicates region where observed visibility is restrictive.

		Visibility				
		12	11	22	21	Totals
Ceiling	12					
	11					
	22					
	21					
	Total					

Table 12: Red indicates region where observed ceilings are restrictive.

		Visibility				
		12	11	22	21	Totals
Ceiling	12					
	11					
	22					
	21					
	Total					

Table 13: Red indicates region where the observed ceiling or visibility is restrictive.

		Visibility				
		12	11	22	21	Totals
Ceiling	12					
	11					
	22					
	21					
	Total					

Table 14: Dangerous conditions may exist because neither restrictive ceiling nor visibility conditions are indicated by the NCVA. False positive (FP) indicates that one or both of the parameters indicates restrictive conditions. Dangerous conditions may also be identified by incorrect analyses for the other parameter (W).

		Visibility				
		12	11	22	21	Totals
Ceiling	12	Danger 1		Danger 2	W	
	11					
	22	Danger 3			FP	
	21	W		FP	FP	
	Total					

Table 15: Value from a two category grouping of ceiling and visibility. Red values indicate potentially dangerous conditions. Values are expressed as a percent.

		Visibility				
		12	11	22	21	Totals
Ceiling	12	**4.4**	3.7	**16.7**	1.6	26.4
	11	4.3	9.9	23.2	5.0	42.4
	22	**5.4**	4.1	7.3	**5.4**	22.2
	21	0.4	1.5	**5.5**	**1.6**	9.0
	Total	14.5	19.1	52.8	13.6	100.0

4.3 Ceiling results from the cross-validation analyses

For this comparison, NCVA ceiling values and METAR ceiling observations for the cross-validated (testing) data are compared. Only cases that have measurable NCVA or METAR ceilings are examined.[2] This adjusted sample size is still too large to deal with in a meaningful way, so a random sample of 10,000 cases was used for the analysis and plots.

In addition, all ceilings above 20,000 ft were set to 20,000 ft. This was done to prevent large differences in values from overwhelming the results and to accommodate human observations of ceiling height at some METAR sites: instrument-recorded ceiling values are often capped, which is not the case for human observers. A level of consistency can be reached by censoring the data in this way.

Box plots[3] showing the distributions of differences between the NCVA and METAR ceiling values for the Summer-2006 time period are shown in Figure 5. These plots show that for ceiling less than 3,000 ft, the NCVA is biased low, while in the range between 3,000 to 10,000 ft the NCVA ceiling values exhibit a high bias. In the 10,000- to 20,000-ft range the NCVA ceiling values also have a high bias overall, although there is a greater frequency in this range of NCVA ceiling values being too small.

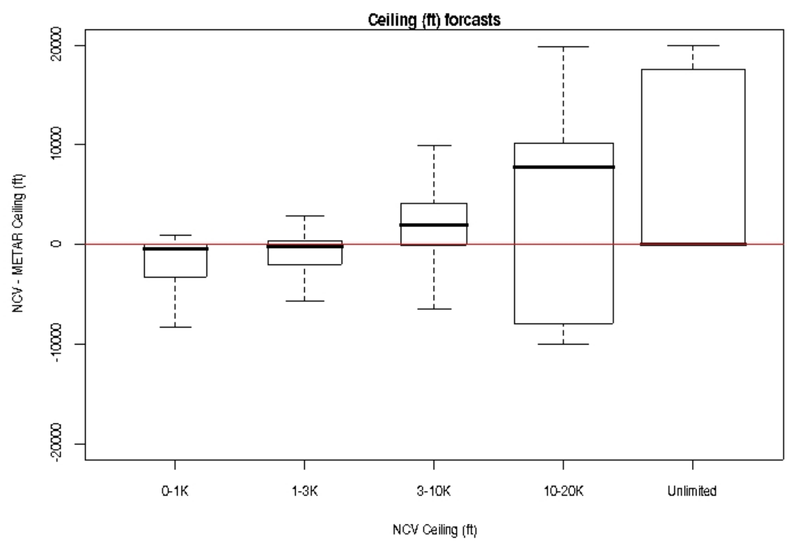

Figure 5: Box plots showing the distributions of differences between the NCVA ceiling values and the METAR ceiling observations as a function of NCVA ceiling during the Summer-2006 time period.

---

[2] The majority of observed and analyzed ceiling heights are "unlimited". These correctly identified non-event cases are the least interesting and are difficult to analyze since "unlimited" is not a numeric value; thus, they were excluded from this comparison.

[3] Box plots show the distribution of values. The line at the center of each box is the median, while the top and bottom of the box represent the 75th and 25th percentiles, respectively. Thus, the box shows the range of the center half of the data. The whiskers extend to the maximum and minimum values that are not outliers, each showing the range of the top and bottom quarters of the data. The dots above or below the whiskers are outliers. The width of each box is scaled to the number of cases represented by that category. Thus, narrower boxes represent fewer cases than wider boxes.

The histogram in Figure 6 shows distribution of errors in the ceiling field (NCVA ceiling minus METAR ceiling) during the Summer-2006 time period. The red line in the center provides guidance regarding under- or over-forecasting. Errors to the left of zero are an indication of under-forecasting and errors on the right indicate over-forecasting. Most of the errors are small, within a couple of thousand feet in absolute value. However, the results in Figure 6 also indicate that the NCVA has a bias toward higher ceiling values, as indicated by the positive skewness of the histogram. The large spike in errors around -8,000 ft is a remapping artifact for ceiling values beyond the maximum operational limits of the ceilometers. The NCV algorithm maps these values to 12,000 ft and the QA analysis maps these values to 20,000 ft, giving the difference of -8,000 ft.

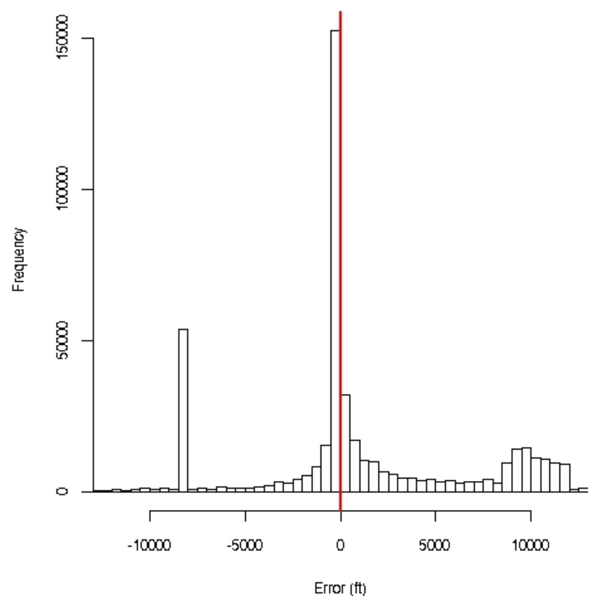

Figure 6: Histogram showing the distribution of errors in ceiling height (NCVA-METAR) during the Summer-2006 time period.

A contour plot of the density of a random sample of 100,000 ceiling observations below 5,000 ft is given in Figure 7. Regions of this plot with a great number of points are shaded in warmer colors. Cooler colors indicate areas with fewer points. This plot is an alternative to a scatter plot. In a scatter plot, the areas with warm colors would be an indecipherable mass of points; the blue areas would have some points and the purple areas would be nearly empty. Ideally, warm colors should fall along the one-to-one line (in red) with cooler colors filling the remaining areas of the plot, which would indicate a good correspondence between the NCVA ceiling values and the observed ceilings provided by the METARs. As shown in Figure 7, this is indeed the case as the warmer colors do lie along the one-to-one line. Away from the one-to-one line, a greater number of lighter colors are located above (to the left) of the one-to-one line. This result is another indication of the NCVA's tendency identifying ceiling values that are somewhat too large. In addition, many NCVA ceiling values indicate ceilings of 12,000 and 20,000 feet (not shown), which is reflected in the large bias seen for the 10,000-20,000-ft and Unlimited categories in Figure 5. This characteristic also is a result of truncation of the NCVA ceiling values.

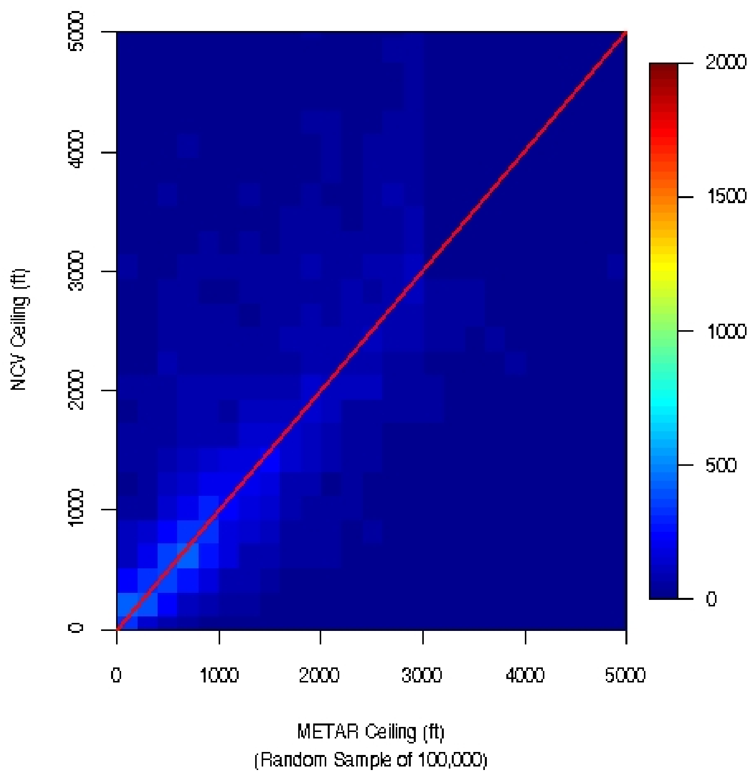

Figure 7: Contour plot showing density of NCVA and METAR ceiling pairs during the Summer-2006 time period.

The Quantile-Quantile (Q-Q) plot in Figure 8 compares the distributions of NCVA and METAR ceiling fields on a log-log scale. This type of plot shows the relationships between the overall distributional characteristics of each variable (e.g., the range, variance) rather than characteristics of their individual differences. The vertical stacks of values on the lower left of the plot, near the origin, are due to the discreteness of the METAR ceiling values, which is especially noticeable near the surface on the log scale. Multiple NCVA ceiling points match each discrete METAR ceiling value at those levels.

If the distributions of these two fields were identical, all points would fall along the one-to-one line. Instead the points are shifted above the 1-to-1 line in an almost linear fashion. This result indicates that the distributions of the NCVA and METAR ceiling values are about the same except that the NCVA ceiling values are shifted higher. The points located at the ends of the plot have a somewhat non-linear shape. The reason for this non-linear shape at the lower end of the scale may be attributed to discreteness of the METAR measurements rather than any real difference in the distributions of ceiling values. On the upper end of the plot, the departure from linearity implies that the differences between NCVA and METAR ceiling values are even larger than would be expected based on the shift of the rest of the distribution (i.e., the NCVA is more biased in the higher ranges).

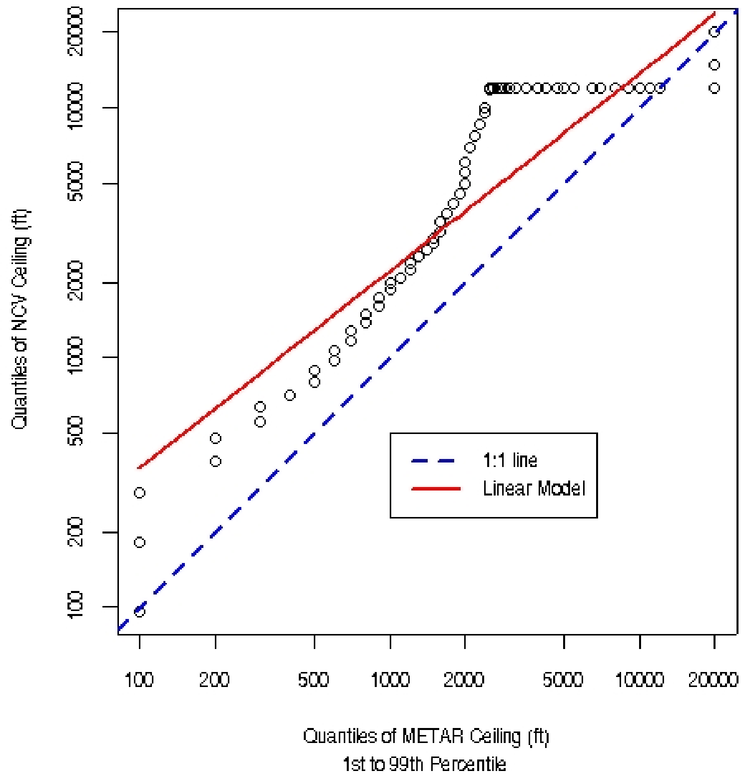

Figure 8: Q-Q plot of ceiling vs. METAR ceiling for Summer-2006 time period.

The box plots in Figure 9 show distributions of the difference between the NCVA and METAR ceiling values for the Fall-2006 time period. This plot shows that in the surface to 3,000-ft range there is essentially no difference between the ceiling values produced by the NCVA and the ceiling measurements provided by the METARs. In the greater than 3,000-ft range, the NCVA ceiling values exhibit a large positive bias.

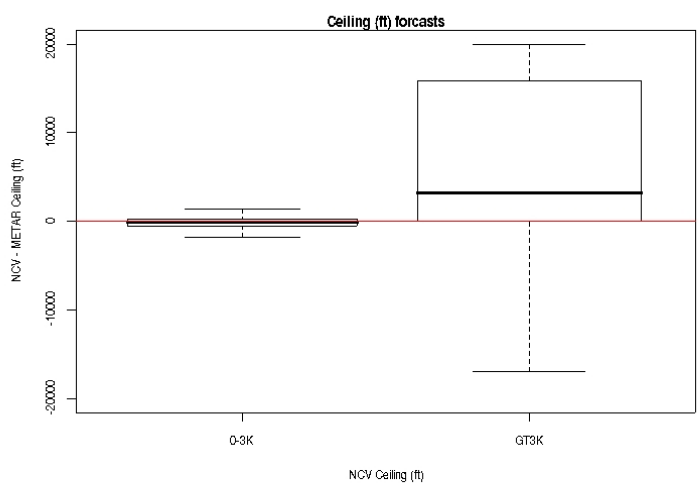

Figure 9: Box plots of the differences between the NCVA ceiling analyses and observations of ceiling provided by the METARs during the Fall-2006 time period.

The histogram in Figure 10 shows the errors in the ceiling field (NCVA ceiling minus METAR ceiling) for the Fall-2006 time period. The red line in the center provides guidance regarding under- or over-forecasting, with errors related to under forecasting on the left side of the red line and areas of over forecasting on the right side of the line. Most of the errors are absolute values of the errors are small, less than a couple of thousand feet. However, the results in Figure 10 also indicate that the NCVA has a bias toward higher ceilings for this period, as was the case for the summer period. This aspect of NCVA performance is indicated by the positive skewness of the histogram. The large spike in errors around -8,000 ft is due to the same artifact discussed in the Summer-2006 results (e.g. Figure 6).

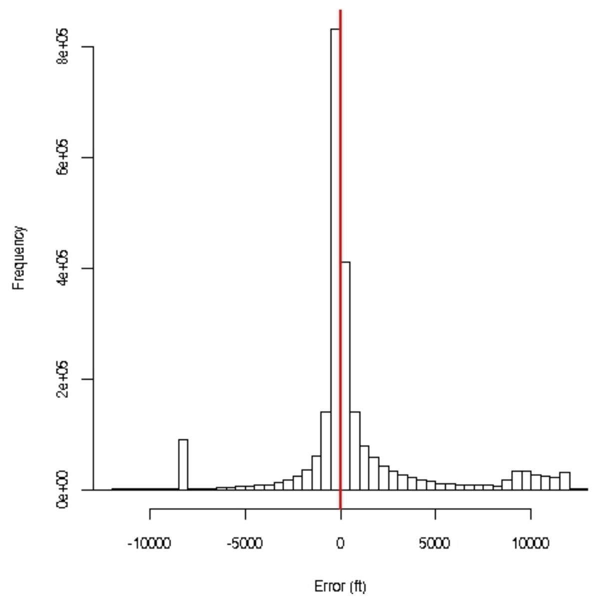

Figure 10: Histogram showing error in ceiling height (NCVA – METAR) during the Fall-2006 time period.

A contour plot of the density of a random sample of 100,000 ceiling observations below 5,000 ft is presented in Figure 11. The plot indicates that the NCVA ceiling values and observations are in good agreement up to a height of 3,000 ft. There is a slight positive scatter in the plot that is indicative of a slight high bias in the NCVA ceiling values.

The Q-Q plot in Figure 12 compares the distributions of NCVA and METAR ceiling fields on a log-log scale for the Fall-2006 time period. As with the Summer-2006 time period, the Q-Q plot for the Fall-2006 time period has a linear shape that is shifted to the left of the one-to-one line. The values located at the ends of the plots are somewhat non-linear in shape; this characteristic may be attributed to discreteness of the METAR measurements rather than any real difference in the distributions of ceiling values. On the upper end of the plot the departure from linearity implies that the differences between NCVA and METAR ceiling values are even larger than would be expected based on the shift of the rest of the distribution (i.e., the NCVA is even more biased in the higher ranges).

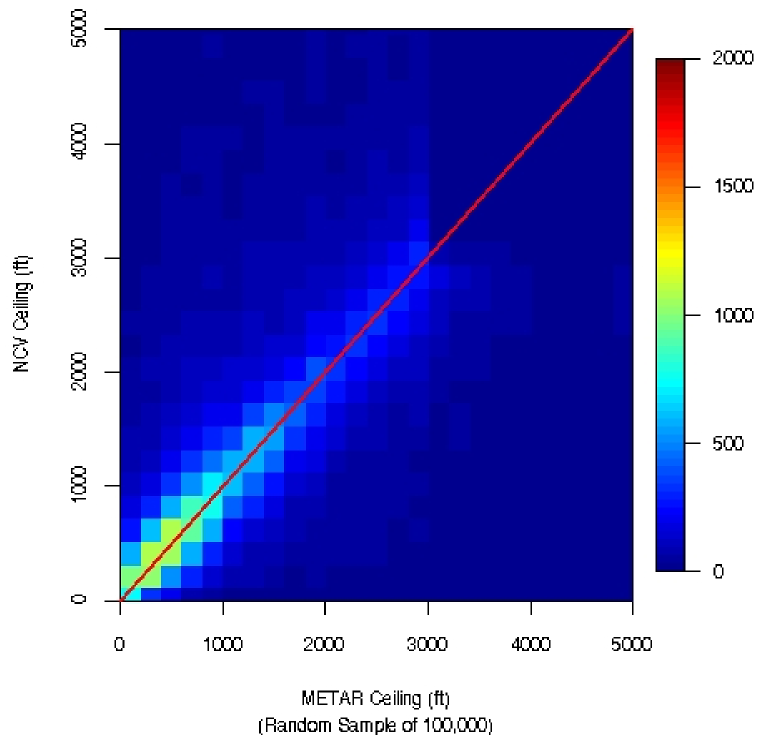

Figure 11: Contour plot showing density of METAR and NCVA ceiling pairs during the Fall-2006 time period.

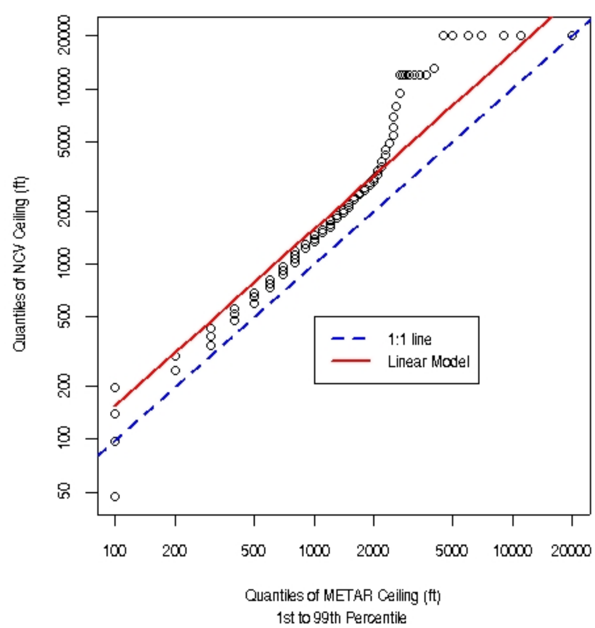

Figure 12: Q-Q plot of NCVA ceiling vs. METAR ceiling for Fall-2006.

4.4 Visibility results from the cross-validation analyses

This section presents a comparison of METAR and NCVA visibility values, again for locations representing the interpolation points between METAR stations. As was done for the ceiling analyses, cases where both the METAR and NCVA reported “unlimited” visibility were excluded from this analysis. Visibility measures are censored at 10 mi, as visibility greater than 10 mi is essentially considered unlimited. Visibility data are binned in quarter-mile increments up through 2 mi. Visibilities greater than 2 mi have been binned every mile with the exception that a bin for 2.5 mi has been included.

The box plots in Figure 13 show distributions of the differences in visibility values between the NCVA visibility values and the METAR visibility observations for the summer period, as a function of observed visibility. These plots show that for observed visibility values up to about 5 mi the NCVA visibility product has a bias toward visibility values that are too large. For observed visibilities greater than 5 mi, the NCVA visibility values are biased toward visibility values that are too small.

The histogram in Figure 14 shows the distribution of errors in the visibility field (NCV-METAR) for the Summer-2006 time period. The red line in the center provides guidance regarding under- or over-forecasting, with errors related to under-forecasting by the NCVA on the left side of the red line, and areas of over-forecasting on the right side of the line. This plot is very uniform with a high percentage of the error values between ±3 mi. Overall, the NCVA visibility values do not have an apparent bias.

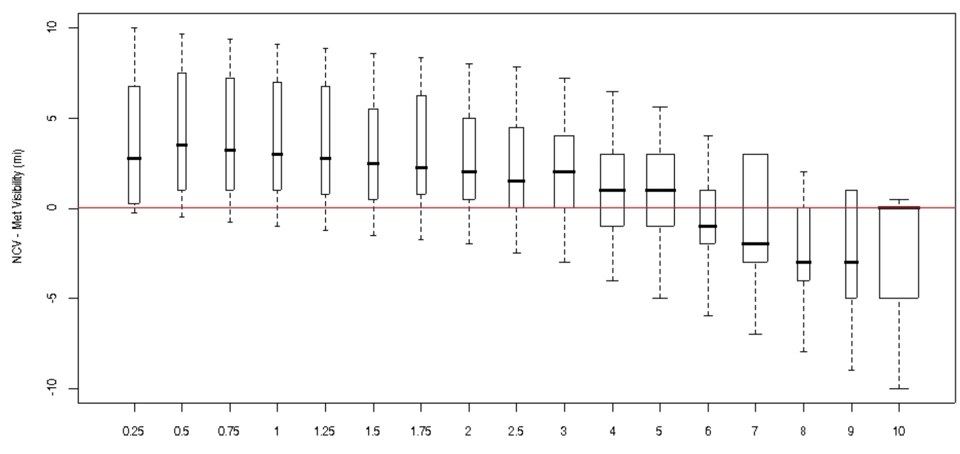

Figure 13: Box plots showing distributions of the differences between the NCVA visibility values and the METAR visibility observations for the summer time period.

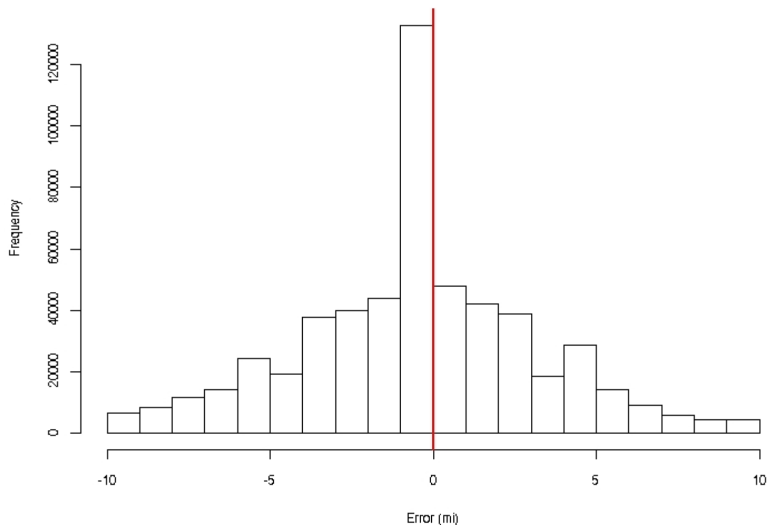

Figure 14: Histogram of the differences between the NCVA visibility values and the METARs for the Summer-2006 time period.

The Q-Q plot in Figure 15 compares the distributions of NCVA and METAR visibility values on a log-log scale for the Summer-2006 time period. The visibility values are very closely to the 1-to-1 line, with a slight bias toward higher values for the NCVA. The linear model line is shifted slightly to the left at lower values of visibility and runs right along the 1-to-1 line as visibility values increase. The results show that the distribution of NCVA visibility values is in good agreement with the distribution of the observations.

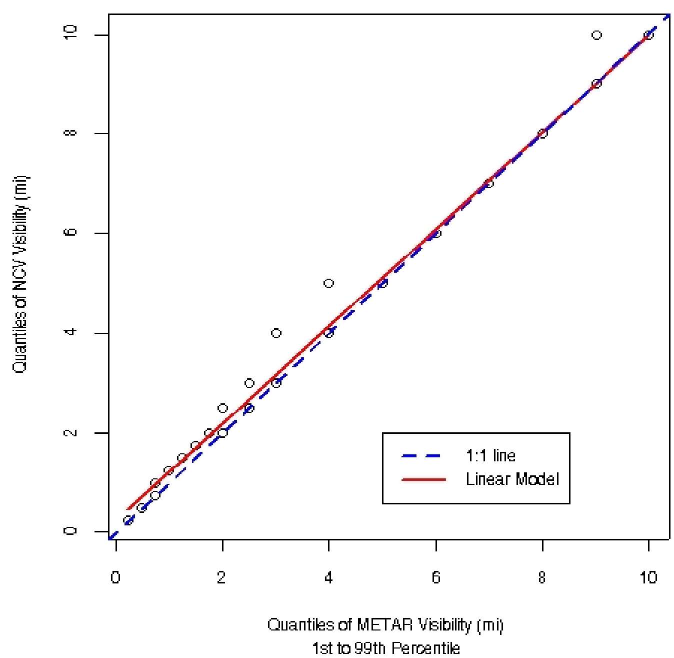

Figure 15: Q-Q plot of NCVA visibility vs. METAR visibility for the summer time period.

The box plots in Figure 16 show distributions of the differences between the NCVA visibility values and the METAR visibility observations for the Fall-2006 time period. For visibilities less than or equal to 5 mi, the NCVA has a bias toward visibility values that are somewhat too large when compared to METARs. Visibility values greater than 6 mi show a bias toward visibility values that are too small, with an exception at 7 mi where there is relatively good agreement between the NCVA and METAR visibility values.

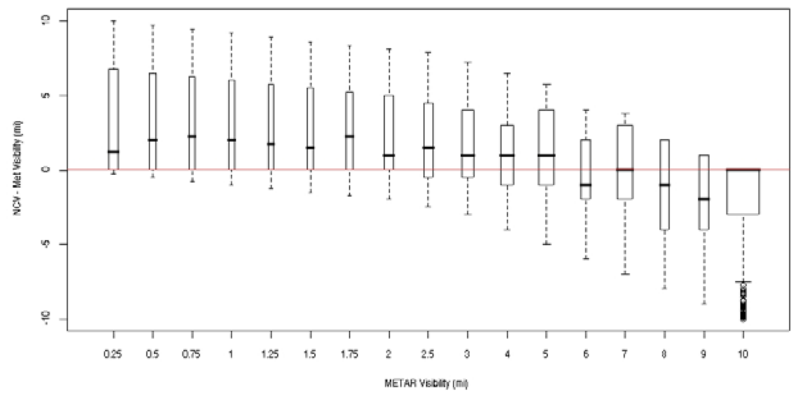

Figure 16: Box plots showing distributions of the differences between the NCVA visibility values and the METAR visibility observations for the Fall-2006 time period.

The histogram in Figure 17 shows the errors in the visibility field (NCV-METAR) for the Fall-2006 period. As with the results for summer, most of the absolute differences are quite small, indicating there is relatively little error in the NCVA visibility grid. The plot also gives little indication of bias in the NCVA visibility values.

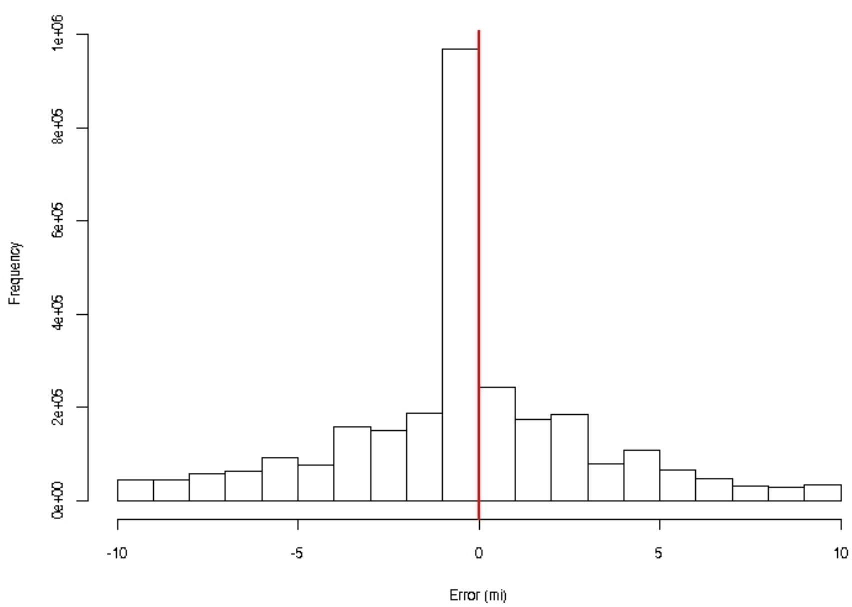

Figure 17: Histogram NCV-METAR visibility differences for Fall-2006.

The Q-Q plot in Figure 18 compares the distributions of NCVA and METAR ceiling values on a log-log scale for the Fall-2006 time period. As was the case for the Summer-2006 visibility values, the Fall-2006 results are very close to the 1-to-1 line with a slight bias of higher visibility values for the NCVA near the lower end of the scale.

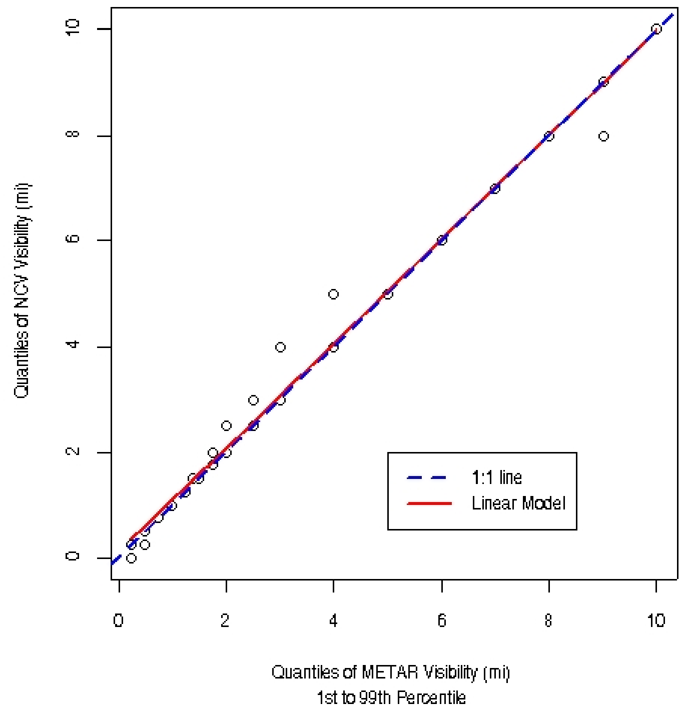

Figure 18: Q-Q plot of NCVA visibility vs. METAR visibility for the Fall-2006 time period.

5 CONCLUSIONS

The NCVA was evaluated for two periods, Summer-2006 and Fall-2006. A cross-validation approach was used to evaluate the performance of the NCVA at interpolated locations between METAR stations, which was the main focus of this evaluation.

The results of the evaluation indicate that the NCVA has positive skill in diagnosing categories of ceiling and visibility conditions. While conditions between METAR locations are not perfectly identified, the algorithm is quite efficient. The NCVA misdiagnosed reduced visibility 15% and lower ceilings 26% of the time. IFR (or worse) conditions at the interpolated locations between METARs were missed about half the time.

Comparisons of NCVA ceiling values with METAR ceiling observations indicates good agreement up to 3,000 ft and a positive bias for higher ceilings. The NCVA visibility values are slightly biased in the high direction for visibilities less than 3 mi, and in the low direction for larger visibility values.

The performances of human-generated forecasts of IFR conditions (i.e., AIRMETs) as well as RUC 2-hr forecasts of ceiling and visibility are also considered in the appendix. These forecasts were considered only to provide some indication of a meaningful level of performance for the NCVA. The results of these evaluations indicate that all three approaches are skillful, with somewhat different performance attributes. Both the AIRMETs and RUC forecasts tend to be positively biased (i.e., IFR conditions are identified more often that they occur) which is the opposite of the NCVA; this result is as expected for these products. Variations in the performance of the NCVA, AIRMETs, and RUC forecasts can be attributed to differences in the basic underlying forms of these three products.

In summary, the NCVA shows positive skill in identifying IFR conditions and ceiling and visibility values. The performance of the NCVA is consistent with the performance of other products that might be used to infer current ceiling and visibility conditions.

6 ACKNOWLEDGEMENTS

This research is in response to requirements and funding by the Federal Aviation Administration (FAA). The views expressed are those of the authors and do not necessarily represent the official policy of the FAA.

7 REFERENCES

Benjamin, S. J. and S. S. Weygandt, J. M. Brown, T. L. Smith, T. G. Smirnova, B. E. Schwartz, E. J. Szoke, and K. J. Brundage, 2004: Assimilation of METAR Cloud and Visibility Observations in the RUC. *11th Conference on Aviation, Range, and Aerospace*, Hyannis, MA. Amer. Meteor. Soc., Boston

Bradley, J.J. and Imbembo, S.M., 1985: "Automated Visibility Measurements for Airports," American Institute of Aeronautics and Astronautics: Preprints, Reno, NV.

Brooks, H. E. and M.P. Kay and J.A. Hart, 1998: Objective limits on forecasting skill of rare events. Preprints, *18th Conf. on Severe Local Storms*, Minneapolis, MN. Amer. Meteor. Soc., 552-555.

Fowler, T.L., A. Holmes, M.J. Pocernich, J. T. Braid, B.G. Brown, and J.L. Mahoney, 2005: Quality Assessment Report for the National Ceiling and Visibility Analysis Product. Report to the FAA Aviation Weather Research Program; available from B.G. Brown (bgb@ucar.edu), 30 pp.

Neter, J., M. H. Kunter, C. J. Nachtsheim, and W. Wasserman, 1996: Applied Linear Statistical Models. McGraw Hill / Irwin, Chicago, Ill.

NOAA, 2006: RUC diagnosed variables. NOAA/Earth System Research Laboratory (Available at Website http://ruc.fsl.noaa.gov/vartxt.html)

NWS, 1991: National Weather Service Operations Manual, D-22. National Weather Service. (Available at Website http://wsom.nws.noaa.gov).

Stoelinga, M. T., and T. T. Warner, 1999: Nonhydrostatic, mesobeta-scale model simulations of cloud ceiling and visibility for an East Coast winter precipitation event. *J. Appl. Met.*, **38**, 385-404.

Takacs, A., L. Holland, M. Chapman, B. Brown, T. Fowler, and A. Holmes, 2004: Graphical turbulence guidance 2 (GTG2): Results of the 2004 post analyses. Quality Assessment Report. Report to the FAA, October 2004.

U.S. Dept. of Commerce, NOAA, 1992: ASOS Users Guide, Government Printing Office.

APPENDIX: Verification statistics for IFR AIRMETs and RUC 2-hr forecasts

In any verification study, it is important to examine verification results for comparable forecasts, both to establish the credibility of the forecasts being evaluated and to ensure that the verification statistics are reasonable. This section includes verification results for IFR AIRMETs as well as 2-hr ceiling and visibility forecasts based on output of the RUC model. These forecast products are available to the same decision makers who might use the NCVA. Thus, although they are very different in form and character from the NCVA, it makes sense to examine their performance characteristics in this context.

Nevertheless, it is important to recall the fundamental differences between these forecasts and the NCVA, including the different requirements guiding their formulation. Thus, a direct comparison of the performance of the NCVA with the performance of the AIRMETs and the RUC forecasts is not meaningful. Specific aspects of these performance differences are outlined in the subsections below, along with the results of the evaluations.

A. AIRMETS

As noted above, AIRMETs are very different from the NCVA both in their form and their purpose. For example, AIRMETs are *forecasts* of ceiling and visibility conditions for a six-hr period (i.e., for 0-6 hr), while the NCVA is an analysis of current conditions. Although the AIRMETs can be amended during their 6-hr valid period, the NCVA is updated every 5 min. Moreover, AIRMETs must cover a minimum of 3,000 mi^2, whereas the NCVA has a granularity of 5 km. Thus, the AIRMETs would be expected to cover broader areas and to capture a larger percentage of the IFR observations than NCVA. In general, smoother forecasts often achieve better statistics than a less smooth forecast (e.g., Takacs et al. 2004). The AIRMETs are required to be a smoother product, while the NCVA may or may not produce smooth areas.

Table 16 and Table 17 present verification statistics for the AIRMETs and the NCVA, respectively, for the full CONUS domain.

Table 16: AIRMET scores for both the Summer-2006 and Fall-2006 time periods.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
Summer	0.586	0.907	0.570	1.361	0.425	0.270	0.492	9.572	6
Fall	0.772	0.845	0.508	1.570	0.502	0.335	0.617	15.700	5

Table 17: NCVA score for both the Summer-2006 and Winter-2006 time periods.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
Summer	0.399	0.972	0.490	0.782	0.413	0.260	0.371	2.303	17
Fall	0.574	0.959	0.344	0.876	0.564	0.393	0.534	5.997	10

As expected, the POD values for the AIRMETs are larger than those for the NCVA for both the Summer-2006 and Fall-2006 time periods; in contrast, the NCVA values of PODn are somewhat larger than the PODn values for the AIRMETs. In addition (as expected), the FAR and Bias values for the AIRMETs are larger than the corresponding values for the NCVA. The larger area efficiency values for the NCVA reflect the fact that the AIRMETs are 6-hr forecasts, in contrast to the NCVA which has a very frequent update rate and thus does not need to cover as much area. These differences between the products also lead to the differences in Bias values: the NCVA tends to underforecast IFR areas, while the AIRMETs tend to overforecast them.

Both the AIRMETs and NCVA have positive skill based on the three skill scores (HSS, GSS, and TSS). Thus, the AIRMETs and NCVA have similar *overall* performance, but the performance is characterized by different attributes (e.g., Bias, detection rate) due to the differences in the basic forms of these forecasts.

B. RUC 2-hr forecast

The comparison of the RUC 2-hr forecast and the NCVA also is not an “apple to apple” comparison. The RUC grid is 13 km, which is quite coarse compared to the 5-km NCVA grid. In addition, the NCVA is updated every 5 min, while the RUC forecast is based on a single output at the top of the hour.

Table 18 and Table 19 list the verification results for RUC 2-hr forecasts and the NCVA for the Summer-2006 time period. The RUC and NCVA POD and PODn values are very similar. In contrast, the FAR value for the NCVA for the CONUS is somewhat smaller than the FAR value for the RUC forecast.

The RUC forecasts have a Bias close to 1 for the summer period but significantly greater than 1 for the fall period. Thus, the RUC forecasts tended to over-forecast the frequency of IFR or worse conditions during the fall period. This tendency is also indicated by the large area values associated with the RUC forecasts. In contrast the NCVA tended to underforecast IFR or worse conditions in both seasons, with more underforecasting in the summer.

The areas covered by the RUC forecasts tend to be larger than those covered by the NCVA, as would be expected due to the differences in grid scales. This characteristic led to smaller Area Efficiency values for the RUC.

Both the RUC and NCVA products show positive skill, with the NCVA values for all three skill measures somewhat larger than the values for the RUC, as would be expected since the RUC is a forecast and the NCVA is a nowcast. The results for the sub-regions have the same general pattern as the results for the CONUS.

Table 18: RUC 2-hr forecast scores for the Summer-2006 time period.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.367	0.956	0.608	0.936	0.333	0.200	0.323	4.771	8
North	0.460	0.922	0.520	0.959	0.389	0.241	0.382	2.119	20
South	0.260	0.949	0.682	0.816	0.227	0.128	0.208	11.531	4
Midwest	0.423	0.972	0.635	1.160	0.367	0.225	0.395	4.913	5
West	0.312	0.972	0.640	0.866	0.304	0.179	0.284	2.760	11

Table 19: NCVA results for the Summer-2006 period using cross-validation technique.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.399	0.972	0.490	0.782	0.413	0.260	0.371	2.303	17
North	0.483	0.952	0.410	0.818	0.471	0.308	0.435	6.163	29
South	0.353	0.961	0.548	0.781	0.349	0.212	0.315	3.455	8
Midwest	0.363	0.986	0.545	0.797	0.387	0.240	0.349	1.259	10
West	0.345	0.986	0.471	0.652	0.397	0.248	0.332	1.093	32

Table 20: RUC 2-hr forecast scores for the Fall-2006 time period.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.702	0.889	0.514	1.445	0.497	0.331	0.591	12.630	6
North	0.748	0.853	0.481	1.442	0.512	0.345	0.602	10.674	7
South	0.634	0.902	0.501	1.272	0.481	0.317	0.536	18.644	4
Midwest	0.767	0.919	0.449	1.392	0.586	0.414	0.686	10.943	6
West	0.632	0.853	0.664	1.881	0.349	0.211	0.485	13.838	5

Table 21: NCVA results for the Fall-2006 period using the cross-validation technique.

	POD	PODn	FAR	Bias	HSS	GSS	TSS	Percent Area	Area Efficiency
CONUS	0.574	0.959	0.344	0.876	0.564	0.393	0.534	5.997	10
North	0.602	0.942	0.325	0.893	0.569	0.398	0.545	10.006	11
South	0.536	0.957	0.351	0.827	0.532	0.363	0.493	6.168	6
Midwest	0.661	0.971	0.283	0.934	0.651	0.483	0.632	6.265	9
West	0.460	0.959	0.463	0.856	0.447	0.288	0.419	4.971	9