10.3

A Performance Assessment of the National Ceiling and Visibility Analysis Product†

Andrew F. Loughe,[1,3*] Brian P. Pettegrew,[1,3] Judy K. Henderson,[3] Joan E. Hart,[1,3]
Sean Madine,[2,3] Jennifer Mahoney[3]

[1]Cooperative Institute for Research in Environmental Sciences, Boulder, CO
[2]Cooperative Institute for Research in the Atmosphere, Fort Collins, CO
[3]NOAA/Earth System Research Laboratory, Boulder, CO

1. INTRODUCTION

In the context of pre-flight planning, this evaluation is carried out to measure the ability of the National Ceiling and Visibility Analysis product (NCVA) to properly characterize categorical flight conditions and discrete ceiling and visibility across the Continental United States (CONUS). In performing this assessment, we evaluate individual attributes of the NCVA, including categorical flight conditions, confidence fields, ceiling, visibility, and issuance frequency. The skill assessment is performed using a cross-validation technique, in which portions of the observational dataset are selectively removed to arrive at a measure of interpolation error *in between* individual METAR sites (Neter 1996).

The main purposes of this evaluation are to:

1) Assess the quality of the NCVA product with respect to that of a baseline analysis.
2) Evaluate the effect of a satellite-based cloud mask on NCVA performance.
3) Measure the potential value of NCVA's frequent update cycle to the flight-planning process.
4) Determine if NCVA's performance compares well to the operational Weather Depiction Analysis, a product specifically referenced in the NCVA concept of use.

†This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

*Corresponding author: Andrew.Loughe@noaa.gov
Andrew Loughe ▪ 325 Broadway ▪ Boulder, CO 80305

2. DATA

The NCVA product (Herzegh et al. 2009), utilizes over 1,800 North American METAR observations (NWS 2000) to create an estimate of ceiling and visibility values every 5 minutes over the CONUS on a 5-km National Digital Forecast Database (NDFD) domain (Glahn et al. 2003). A categorical flight condition (**Figure 1**) is computed based upon ceiling and visibility values reported at METAR observation points, and at interpolated grid points located in between METAR reporting sites (**Table 1**). The NCVA product is updated every 5-minutes to ensure that significant weather changes are being captured.

Table 1: Determination of flight category from individual values of ceiling and visibility (FAA 2008).

Flight Category	Ceiling Height (Feet AGL)	Visibility (SM)
Visual Flight Rules (VFR)	ceil > 3000	vis > 5
Marginal Visual Flight Rules (MVFR)	1000 ≤ ceil ≤ 3000	3 ≤ vis ≤ 5
Instrument Flight Rules (IFR)	500 ≤ ceil < 1000	1 ≤ vis < 3
Low Instrument Flight Rules (LIFR)	ceil < 500	vis < 1

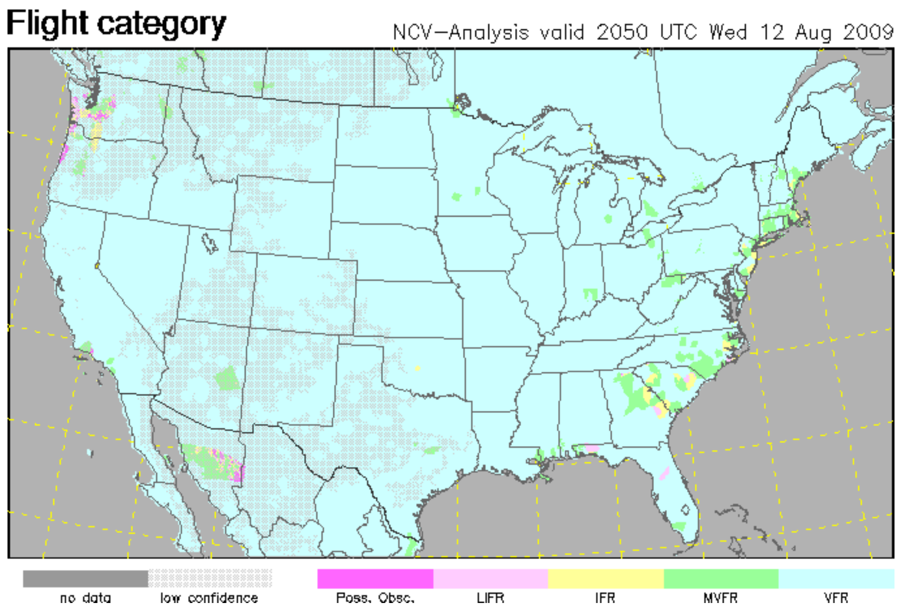

Figure 1: Sample NCV Analysis product from 2050 UTC on 12 August 2009 from the Experimental Aviation Digital Data Service (*http://weather.aero*). Displayed is the Flight Category with hatched areas representing areas of low confidence. The field is color coded based on severity.

The Weather Depiction Analysis is an hourly product generated by the National Centers for Environmental Prediction (NCEP) for the purpose of providing useful flight category information. Hourly analyses are augmented by forecasters every 3 hours to adjust flight-category contours and to add fronts.

3. METHODOLOGY

Cross validation is used to create independence between METARs that are used to verify the NCVA product and those that are used for the NCVA algorithm itself. By design, and with respect to operationally available data, NCVA skill is perfect at individual METAR reporting sites– the analysis does not deviate from values reported *at* those sites. The cross-validation technique utilizes METAR data that are selectively withheld, and provides a measure of how well the analysis performs *in between* METAR reporting sites. In utilizing this technique, we assume that there is minimal effect on the overall quality of the NCVA from withholding 20% of the input data, and we also make an effort to ensure that the entire geographic domain is effectively sampled.

For this study, we employ a random-repeat cross validation with 10 iterations, utilizing a pool of 1700 surface stations over the CONUS and Canada that report at least 90% of the time. From these frequently-reporting stations, we create 10 separate lists of randomly selected withholding sites, comprising 340 stations per iteration. For each valid time in the study, 10 separate analyses are computed using the remaining 1360 stations (**Figure 2**). Once selective data holes have been created within the analysis grids, we gather analysis output from the 10 iterations and compare results from the analysis with reported values from those METARs that were held out. We then assess the error in estimating ceiling, visibility, and flight category at each of the withheld METAR locations. This study is carried out over June-August 2008 (summer) and December 2008-February 2009 (winter).

4. QUALITY OF THE NCVA PRODUCT

In order to measure quality of the NCVA product, and in particular its ability to properly identify IFR conditions, a level of skill (positive or negative) must be explicitly defined. In this evaluation, a baseline measure is computed by

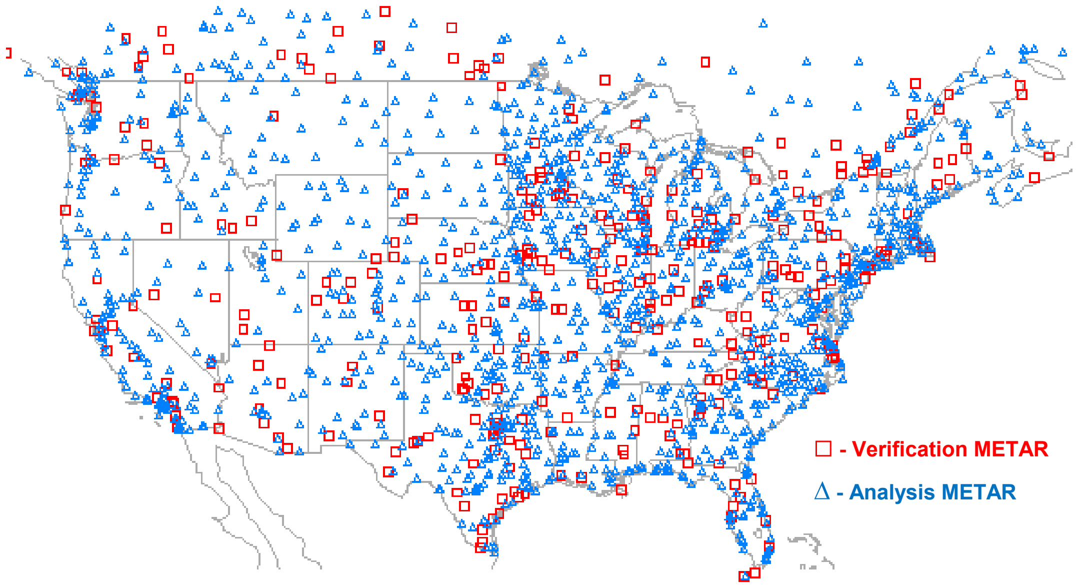

Figure 2: Distribution of METAR sites throughout the CONUS. These METARs have been determined to report at least 90% of the time. Blue triangles represent those METARs that were included in the set #10 cross-validation runs (~1360 stations), while red squares are those METARs that were withheld from that particular run (~340 stations).

creating a surrogate analysis that is also compared with values from withheld METARs. The approach utilizes a comparison of the withheld METAR to its nearest neighbor in arriving at a baseline measure of skill. We refer to this baseline analysis as the nearest-neighbor analysis (NN-A). The actual performance of the NCVA against withheld METARs (using a 10km radius, best flight category matching technique) is then compared with the performance of NN-A against withheld METARs.

Following this methodology, we have determined (see **Table 2**) that the NCVA could add significant value to the flight planning process compared with the NN-A:

- By providing a higher probability of detection (PODy, detection rate), thus more effectively detecting IFR events and reducing risk over the CONUS. The NCVA was also more effective at reducing false alarms on IFR events, resulting in more efficient use of the airspace.

Table 2: Performance in detecting IFR events by the NCVA and by a Nearest-neighbor Analysis (NN-A).

	NCVA vs NN-A Detection Rate	**NCVA vs NN-A FARatio**
All Seasons	0.71 vs 0.60	0.25 vs 0.39
Summer	0.60 vs 0.47	0.31 vs 0.51
Winter	0.77 vs 0.66	0.22 vs 0.34

We also determined that the quality of the NCVA and the NN-A differ during the wintertime and summertime:

- NCVA has a significantly higher PODy of IFR events in the wintertime than in the summertime, and a lower (better) FARatio (False Alarm Ratio) in the wintertime than in the summertime.

A regional analysis of NCVA performance shows generally better performance in the Midwest, East, and along the Southwest Coast than in the Intermountain West where performance is generally poor (**Figure 3**). While there are issues associated with data quality and density over these large regions, we have also determined that skill is higher whenever the frequency of long-lived, large scale IFR events is greatest, as is the case along the west coast in the summertime (**Figure 4**).

When considering the root mean square error (RMSE, Wilks 2006) of NCVA versus NN-A for the ceiling analyses, the NCVA was found to have a RMSE that was 161 ft less than the NN-A error for all seasons, and a corresponding visibility analysis error that was 0.16 statute miles less than that measured for the NN-A.

5. EFFECT OF THE SATELLITE CLOUD MASK

We also sought to determine effects of a satellite cloud mask on skill and confidence of the NCVA. Results are presented for a set of cloud mask/no cloud mask analysis runs that were carried out for a summer month and a winter month (July 2008 and January 2009). The night/day results presented in this section include the 0600, 0900 UTC analyses as part of nighttime, and the 1500, 1800 UTC analyses as part of daytime. The 1200 UTC analyses are entirely omitted from this comparison. The cloud mask can alter ceiling or visibility values at a

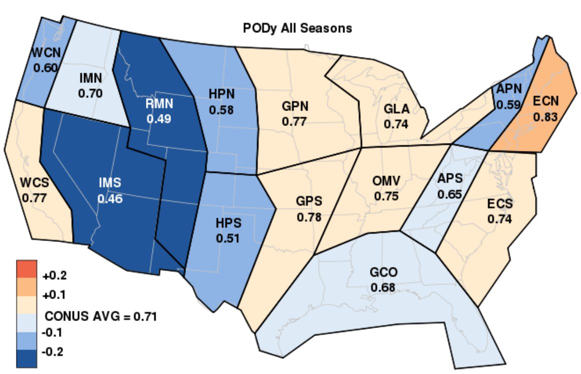

Figure 3: Regional performance (PODy) for all seasons. Colors represent deviation from the CONUS average. Note the relatively high PODy values in the Midwest, East, and Southwest Coast. The Intermountain West reflects a much lower PODy.

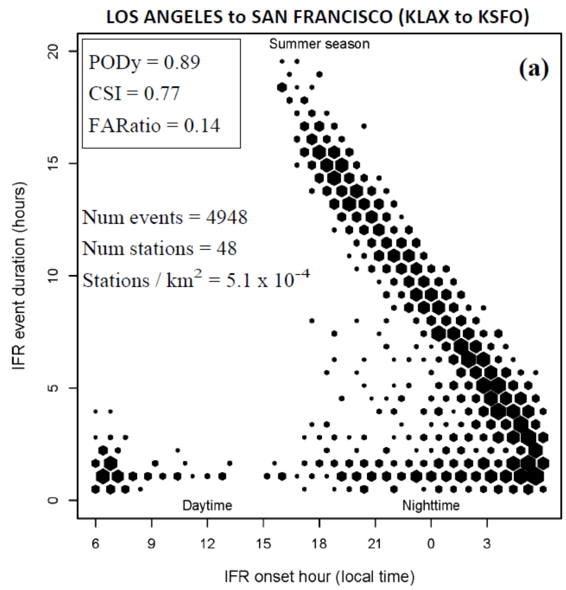

Figure 4: Frequency analysis of IFR event duration vs. onset hour, with NCVA performance measures for the summer from KLAX to KSFO. Note that long-duration IFR events are initiated with great frequency at night.

location where recent satellite data strongly indicates that the skies are clear, while a distant yet influential METAR report suggests a contrary value. It is important to note that the cloud mask has only a clearing effect on ceiling, and cannot alter values *at* METAR reporting sites. Overall, we found that the cloud mask does not harm NCVA's ability to identify IFR conditions.

In the aggregate for all flight category confidence levels and for the situation referencing IFR/non-IFR flight category classifications, it can be shown from the viewpoint of a statistical contingency table (**Figure 5**), that the migration of NCVA analysis points from hits to misses and from false alarms to correct negatives are the only changes possible with application of the "clearing" cloud mask. This would result in an ***unfavorable*** decrease in hit rate (PODy), and a ***favorable*** decrease in false alarm rate (FAR). Raw numbers of analysis points in addition to percentages are shown in **Table 3** to highlight migration of analysis points ***to*** the "miss" cell of the contingency table versus those that migrate ***from*** the "false alarm" cell.

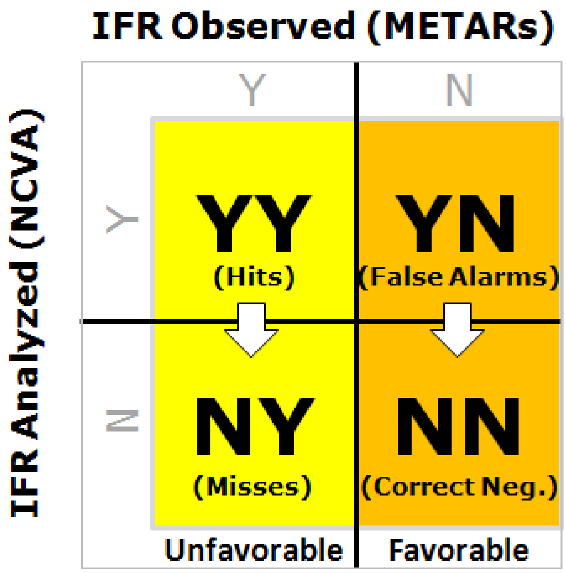

Figure 5: The satellite-based cloud mask has the potential for altering analysis grid point flight category classifications from IFR to non-IFR, but *not* from non-IFR to IFR. Thus in the aggregate for all flight category confidence values, application of the cloud mask can move NCVA values from the top row to the bottom row of the contingency table, which leads to an unfavorable increase in misses, but a favorable decrease in the number of false alarms.

As seen in **Table 3,** from application of the cloud mask, we measure a small decrease in hits (increase in misses), along with a somewhat appreciable decrease in false alarms (increase in correct negatives), thus yielding a slight increase in analysis risk (decrease in the hit rate), but a more significant increase in overall efficiency (measurable decrease in false alarms) over the CONUS.

Note, however, even with this slight increase in risk, that the overall NCVA risk measurement is still significantly lower than that of the baseline (better PODy for the NCVA represents lower risk). Utilization of the cloud mask for both the summer and winter months yields an increase of only 137 (0.52%) misses migrating from the hit cell of the contingency table, and a much larger decrease in the number of false alarms (1356, a 5.87% reduction).

We have also noted a large re-assignment of NCVA analysis flight category confidence levels from low confidence to normal confidence with application of the cloud mask (see **Table 3**). For the summer and winter months combined, 9.26% (74,504) of the low flight category confidence analysis points were re-assigned a normal flight category confidence value with utilization of the cloud mask.

The percent increase in misses is slightly greater at night than during the day, with a 0.72% (81) increase in misses at night compared to a 0.33% (28) increase in misses during the day. The percent decrease in false alarms is nearly the same during the day and night, with a 5.52% (524) decrease at night and a 6.52% (503) decrease during the day.

The net re-assignment of flight category confidence from low confidence to normal confidence with application of the cloud mask is reversed at night, with 0.054% (180) of the

Table 3: Change in misses and false alarms for the statistical contingency table as drawn in **Figure 5**, along with the net decrease in low flight category confidence points with re-assignment to normal flight category confidence through utilization of the cloud mask.

Time Period of the Study	**Increase in Misses with Application of the Cloud Mask *(Unfavorable)***	**Decrease in False Alarms with Application of the Cloud Mask *(Favorable)***	**Net Decrease in *Low* Flight Category Confidence Values (re-assignment to *Normal* Flight Category Confidence)**
Summer and Winter	137 (0.52%)	1,356 (5.87%)	74,504 (9.26%)
Daytime (excludes 1200 UTC)	28 (0.33%)	503 (6.52%)	72,659 (24.3%)
Nighttime (excludes 1200 UTC)	81 (0.72%)	524 (5.52%)	-180 (-0.054%)

nighttime low confidence analysis points having been re-assigned ***from*** the normal confidence classification at night, and 24.3% (72,659) of the daytime low confidence analysis points having been re-assigned ***to*** the normal confidence classification during the day. At night, the cloud mask is rather active in converting analysis values from IFR classification as noted above, but the cloud mask is much less active at night in altering confidence values of the analysis field (**Table 3**).

6. FREQUENCY OF THE ANALYSIS UPDATES

To assess the impact of frequent NCVA updates, we compared analyses that were issued in successive 5-minute intervals from the top of one hour through to the top of the next hour, in an effort to measure the correlation of a fixed reference analysis against those issued every 5 minutes thereafter (00-05, 00-10, ..., 00-65 minutes).

In performing this analysis for a summertime month in 2008 and a wintertime month in 2009, we found that:

- The linear correlation between an NCVA issuance and its successor one hour later is ~0.8, representing a significant change in flight conditions over the CONUS.
- The five-minute updates that occur between hourly issuances appear to effectively capture the incremental changes over the CONUS, as indicated by linear changes in correlation between the initial and intermediate issuances.

An example of correlation decay over the computational NDFD grid for the NCVA is illustrated in **Figure 6**. This difference field illustrates the change in flight categories from 0700 to 0800 UTC on 15 January 2009. Note, for this particular example, that the correlation between the two grids is measured to be 0.85.

When comparing adjacent-in-time analyses, there is a steady reduction in correlation throughout the time period of an hour, down to the more abrupt change in correlation at 60 minutes, which is the beginning of the next analysis period (**Figure 7a-c**). The decrease in correlation measured between 55 minutes past

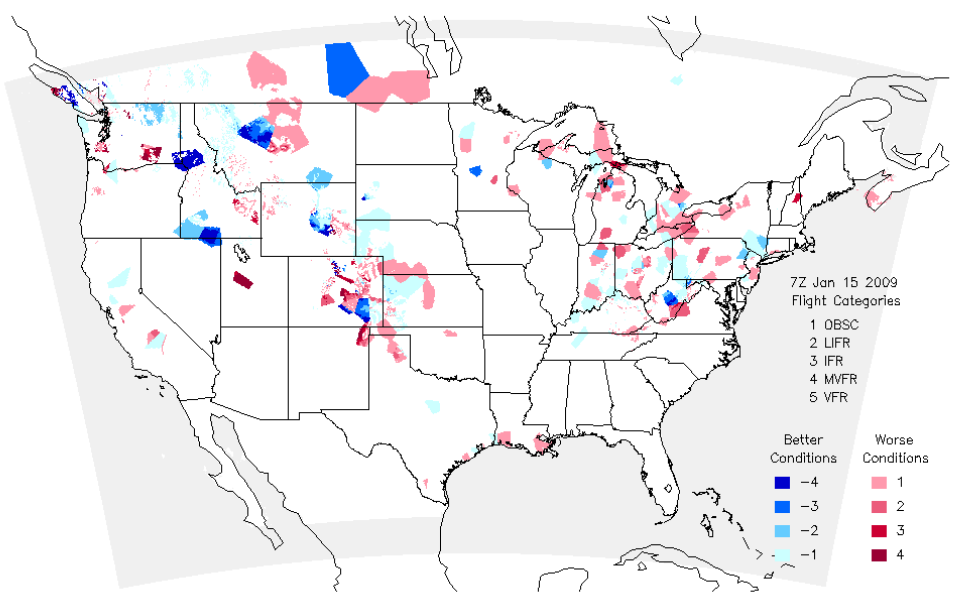

Figure 6: Difference field illustrating change in flight categories from 0700 to 0800 UTC 15 Jan 2009. Shades of red reveal worsening conditions while shades of blue reveal improving conditions. The correlation between the 0700 and 0800 UTC NCVA grids was measured to be 0.85.

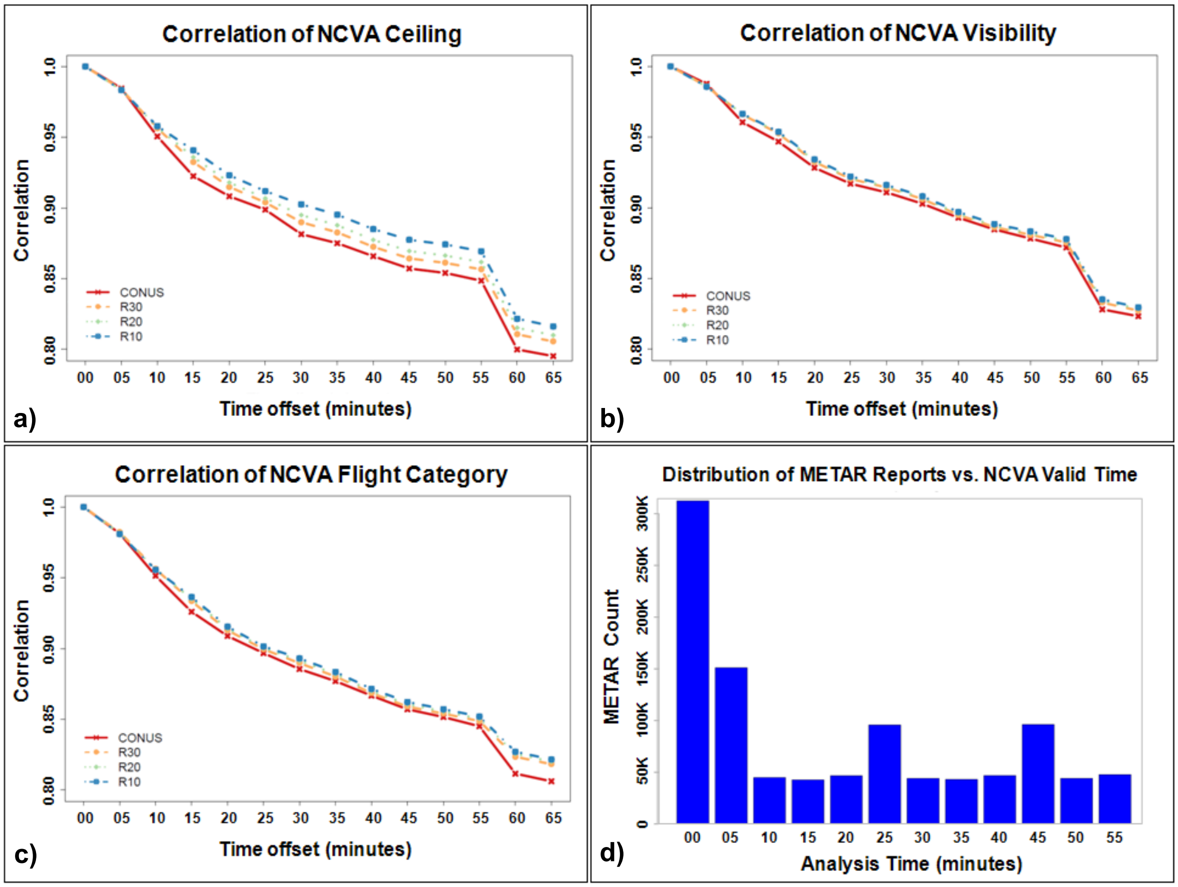

Figure 7: Decrease in correlation over the course of 65 minutes from reference analyses taken from the top of the hour during July 2008 and January 2009, and for various radius matching distances. Note that the inclusion of new METAR data every five minutes leads to a significant and steady decline in correlation over the course of one hour, with a larger drop in correlation noted near the time of the hourly infusion of a sizeable number of new METAR reports (60 minutes past the top of the hour).

the hour (0.85) and 60 minutes past the hour (0.80) represents a change in flight category conditions not captured by the intermediate 5-minute updates. The frequent updates provide a consistent and steady change with infusion of new METAR reports throughout the course of the hour.

The steady reduction in correlation between analysis times indicates that the 5-minute cycle is achieving its goal of providing useful updates throughout the course of an hour, with a much larger drop in correlation noted near the time of hourly infusion of a sizeable number of new METAR reports (60 minutes past the top of the hour). It also demonstrates that the analysis system is stable and not prone to dramatic changes in results between fixed time increments. The decrease in correlation is steeper during the first 30 minutes of the hour as a result of the larger number of new METAR reports that are utilized during that time period than in the latter half hour.

The discretized nature of flight category acts to dampen fluctuations in this field since particular thresholds of ceiling and visibility must be crossed to record a change in category. For ceiling, the further out in time from the reference analysis, out to 55 minutes, the greater the separation in correlation results by proximity to the analysis points. It is worthwhile noting that the reported sky cover is representative of a broader horizontal area than is the reported visibility value.

7. COMPARISON OF NCVA WITH THE OPERATIONAL WEATHER DEPICTION ANALYSIS

The primary issue with regard to comparison of the NCVA to the operational hourly Weather Depiction Analysis is to determine if the automated product performs at least as well as manually generated products, and if it is consistent with those products (FAA/ATO 2007). During the course of this analysis, it was found that:

- NCVA and the Weather Depiction Analysis have an overall correlation greater than 0.6, indicating that these disparate products may perform in a consistent manner.
- With no suitable way to directly measure the quality of the Weather Depiction Analysis, the study finds indications that NCVA performs at least as well as the Weather Depiction Analysis.

The correlation between the gridded NCVA product and the operational hourly Weather Depiction Analysis was calculated for a period of July 2008 and January 2009, using a grid-to-grid approach.

The skill of the NCVA, using CSI, is plotted against its correlation with the Weather Depiction Analysis. Results are color-coded by season (JAN=blue dots ; JULY=green dots) and distributions are calculated. The numbers printed in each quadrant represent the relative distribution of points. This distribution of results shows that more than 90% of compared values have a significant positive correlation (>0.5). Only 45% of those comparisons show lower skill (**Figure 8**).

A majority of the comparisons where NCVA showed better skill (CSI > 0.5), occur during the month of January (blue dots), a month shown to have a greater frequency of restricted flight category events. Further examination reveals that on average, the NCVA tends to analyze a greater spatial quantity of events MVFR or worse compared to the Weather Depiction Analysis (**Figure 9**).

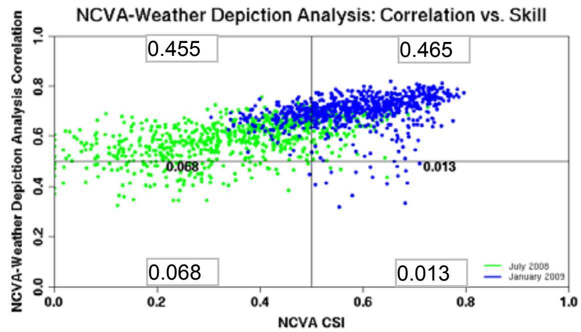

Figure 8: Scatter plot of NCVA-Weather Depiction Analysis correlation vs. NCVA skill, color coded by season (January 2009 in blue and July 2008 in green) with quadrant ratio in the upper center of each quadrant.

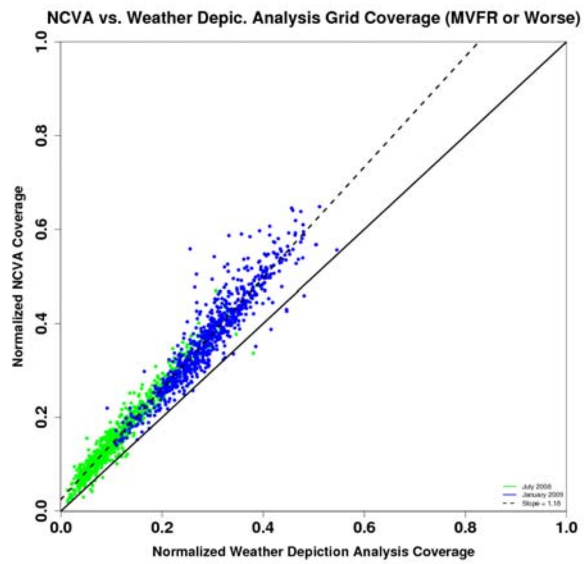

Figure 9: Scatter plot diagram of NCVA grid coverage vs. Weather Depiction Analysis grid coverage. Solid line represents a slope=1 line and the dashed line represents the linear regression of points (dashed-line slope=1.18).

8. CONCLUSIONS

We have summarized results from a formal quality assessment of the National Ceiling and Visibility Analysis product (NCVA), a gridded analysis that evaluates reported ceiling and visibility information for the purpose of improving the flight planning process. On behalf of the Federal Aviation Administration's Aviation

Weather Research Program, and in support of an Aviation Weather Technology Transfer (AWTT) D4 (operational) decision point, this study was carried out to examine the following:

- The quality of the NCVA product with respect to that of a baseline analysis.
- The effect on NCVA performance from application of a satellite-based cloud mask.
- The potential value of NCVA's frequent update cycle to the flight planning process.
- NCVA's performance compared to the operational Weather Depiction Analysis, a product specifically referenced in the NCVA Concept of Use.

Quality of the NCVA product

Overall, the NCVA could add significant value to the flight-planning process compared with the baseline analysis (NN-A):

- By more effectively detecting IFR events and reducing risk over the CONUS (NCVA Probability of Detection of 0.71 vs. 0.60 for the NN-A).
- By more effectively reducing false alarms of IFR events, resulting in more efficient use of the airspace (NCVA False Alarm Ratio of 0.25 vs. 0.39 for the NN-A), with a lower False Alarm Ratio being more favorable.
- Continuous measures of the error in ceiling and visibility attributes concur with the observations stated above.

NCVA performance was also found to vary greatly by region, sub-region, and weather regime:

- Performance of the NCVA was found to be most favorable along the East Coast, the Southwest Coast, and in the Midwest. It performed much less favorably in the Intermountain West and Plains region. Consistent among the sub-regions is demonstrated superiority of the NCVA over the NN-A.
- Skill was found to be greater in sub-regions possessing high METAR station density and where long-lived, large-scale IFR events occur frequently. This is evident, for example, along the West Coast in the summertime.

Effect of the satellite-based cloud mask

This study indicates that when the cloud mask is applied there is**:**

- An overall decrease in false alarms (5.9%) that outweighs an increase in misses (0.5%), resulting in more efficient use of the airspace while only slightly increasing the risk; the NCVA risk is still significantly lower than that of the baseline.
- A measurable difference between the daytime and nighttime, as a large number of analysis grid points possessing *low* flight category confidence are actively re-assigned to *normal* flight category confidence during the daytime (24.3%), but only a negligible change during the nighttime.

Potential value of the frequent update cycle

Incremental changes in the updates of NCVA every five minutes appear to contain information useful to planners:

- The linear correlation between an NCVA issuance and its successor one hour later is ~0.8, representing a significant change in flight conditions over the CONUS.
- The five-minute updates that occur between hourly issuances appear to effectively capture the incremental changes over the CONUS, as indicated by linear changes in correlation between the initial and intermediate issuances.

Comparison to the weather depiction analysis

This study shows that the NCVA, in accordance with its Concept of Use, performs as well as and is consistent with the weather depiction analysis:

- NCVA and the weather depiction analysis have an overall correlation greater than 0.6, indicating that the two products may perform in a consistent manner.
- With no suitable way to directly measure the quality of the Weather Depiction Analysis, the study found indications that the NCVA performs at least as well as the Weather Depiction Analysis.

9. ACKNOWLEDGEMENTS

This research is in response to requirements and funding provided by the Federal Aviation Administration. The views expressed are those of the authors and do not necessarily represent the official policy and position of the U.S. Government.

10. REFERENCES

Federal Aviation Administration, 2008: Aeronautical Information Manual (section 7-1-7).http://www.faa.gov/air_traffic/publications/ATpubs/AIM.

Federal Aviation Administration Air Traffic Organization, 2007: National Ceiling and Visibility (NCV) CONUS Concept of Use, version 2.0.

Glahn, H.R. and D. P. Ruth, 2003: The New Digital Forecast Database of the National Weather Service, Bull. Amer. Meteor. Soc., 84, 195-201.

Herzegh, P., G. Wiener, R. Bateman, J. Cowie, J.Black, 2009: NCVA – The NCV CONUS Analysis Product, available from herzegh@ucar.edu.

National Weather Service, 2000: National Weather Service Operations Manual. National Weather Service, http://www.nws.noaa.gov/wsom/manual/archives/NC750006.pdf.

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

Wilks, Daniel S., 2006: Statistical Methods in the Atmospheric Sciences, Vol. 91, 2d ed. Academic Press, 627 pp.