7.2

# An Automated National-Scale Ceiling and Visibility Forecast System: Development Progress

Jennifer L. Black, Richard E. Bateman, Paul H. Herzegh,
Gerry Wiener, James Cowie, Cathy J. Kessinger
National Center for Atmospheric Research, Boulder, CO, USA

## 1. Introduction

The Federal Aviation Administration (FAA) Aviation Weather Research Program (AWRP) National Ceiling and Visibility (NCV) Research Team has developed aviation products that depict 1 to 12-h forecasts of Ceiling, Visibility, and Flight Category (a combined field of the two former fields). These products are designed to aid human forecasters with an automated, initial forecast of ceiling and visibility (C&V) over the continental United States (CONUS). The NCV forecast products will be evaluated by the Aviation Weather Technology Transfer (AWTT) board for approval of experimental D3 status in August 2008. This paper presents an overview of the forecast products' architecture and tests done on elements of the system in preparation for the upcoming AWTT decision.

| Flight Category | Ceiling (ft) | Visibility (mi) |
|---|---|---|
| Visual (VFR) | > 3000 | > 5 |
| Modified Visual (MVFR) | ≤ 3000 | ≤ 5 |
| Instrument (IFR) | ≤ 1000 | ≤ 3 |
| Low Instrument (LIFR) | ≤ 500 | ≤ 1 |

***Table 1*** *Definition according to flight rules associated with C&V conditions of the NCV C&V forecast product Flight Category.*

## 2. Overview of the NCV Forecast System

The NCV forecast system outputs an hourly, site-based Ceiling (Fig. 1), Visibility (Fig. 2), and Flight Category (Fig. 3) forecast out to 10 h, and out to 12 h every 6 h. The C&V forecast fields are output as standard numeric values with units in feet and miles. Flight Category is defined by four categories using a combination of the ceiling and visibility fields and is defined in Table 1.

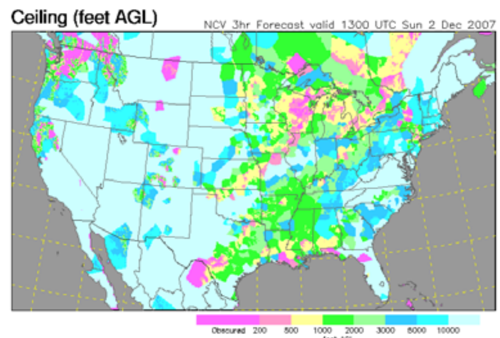



*Figure 1: Example of the NCV System Ceiling forecast field.*

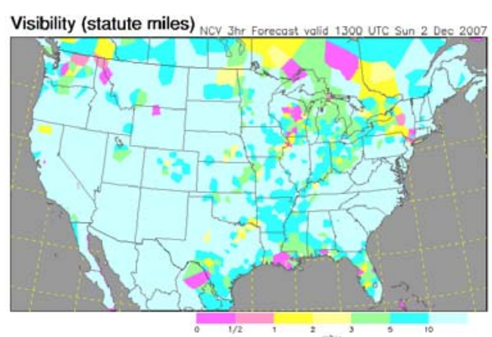



*Figure 2: Example of the NCV System Visibility forecast field.*

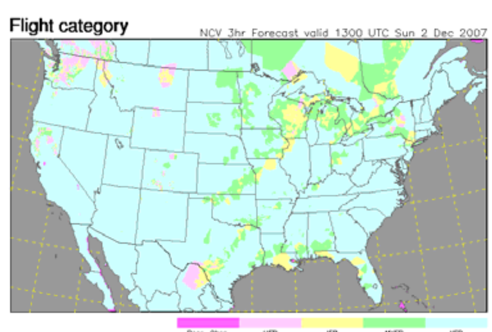



*Figure 3: Example of the NCV System Flight Category forecast field*.

The NCV forecast system uses four input components/modules to generate the forecast: two numerical models -- Rapid Update Cycle (RUC, Benjamin, 2004) and Local AWIPS MOS Program (LAMP, Glahn, 2004); an internal observation-based data mining rule set (DM, Herzegh, 2005); and Meteorological Terminal Air Report (METAR) observations as persistence (PERSIS). Each component/module has a specified METAR site list for which it produces forecasts. The integration strategy assesses the performance of each component and selects the forecast of the module that is performing best as the final forecast. This requires a lookback period over which the components can be evaluated against METAR observations (that represent the "truth"). The components are evaluated independently for each site, forecast hour, and C&V fields. A contingency table is created for each module for each site in the specified site list, forecast hour and C&V fields. In essence for one site the system can choose the RUC forecast for the 3-h ceiling forecast and the LAMP for the 3-h visibility, etc. Lastly, the system uses a nearest-neighbor interpolation similar to the NCV Analysis product (Skiena, 1997); to interpolate the ~2100 site forecasts to the National Weather Service (NWS) National Digital Forecast Database (NDFD) 5-km grid.

### 2.1 System Components/Modules

**RUC**

Hourly files of the RUC 13-km grid are retrieved from the National Center for Environmental Prediction (NCEP) server, containing C&V forecasts for all of the 2110 sites in the system's CONUS site list. Due to an inherent latency of getting the data into the system, the RUC 3-h forecast replaces the 2-h forecast, etc., out to the 10-h forecast. The reason for this use of the data is because it is necessary to get the newest data into to system as soon as possible.

**LAMP**

For the time period over which the tests presented in this paper were done, the LAMP was only initialized every 6 h (3, 9, 15, and 21z). The LAMP C&V forecasts were used by the system when available. The files are also retrieved from NCEP for C&V forecasts for a smaller site list, 1462 sites; the remaining sites in the CONUS site list have only two components at this time. There is a similar latency in retrieval of the LAMP data hence the same effect as the RUC with respect to the forecast hour; the LAMP 3-h forecast is used for the system 2-h forecast.

**Data Mining**

Herzegh (2005) explains the DM techniques used in the NCV forecast system. There is additional information about the DM techniques used at http://www.rulequest.com. At this time, DM rule sets are available for fifty-two of the CONUS sites. The remaining sites use the other components that are available to them at this time, i.e., RUC, PERSIS, and LAMP. There is no latency with the DM data.

**Persistence**

Persistence uses the METAR observation at the initiation time of the forecast system and extends those conditions out to each forecast hour. This is available for all forecast hours, but is only used from hours one through ten.

### 2.2 Integration Schemes

Two integration schemes are evaluated and results discussed here: the Weighted Majority Vote (WMV; Blum, 1996) and Agile Selection (Agile). Both integration strategies run independently for each forecast field, site and forecast hour.

The forecast system uses a modified version of the WMV algorithm. In the version of WMV the entire list of forecast system components have a weighted vote on whether VFR or IFR conditions exist. To determine the current weights of the components' votes there is a specified lookback time period with METAR data to verify the votes. Components weights are penalized for wrong votes and rewarded for correct votes. There is a lower limit to how much a component can be penalized for wrong votes, and the component weights are normalized to the component with the highest weight. When the system is ready to issue a forecast, the weights determined in the lookback period are used and a vote is made by all components as to whether it will be VFR or IFR conditions. The category with the majority of votes wins, and the component that voted in the majority with the highest weight is chosen by the system. If there is a tie between two component weights in the majority vote a precedence table is used to decide the component forecast to be used. The precedence table is site and forecast-hour specific.

The Agile selection algorithm was developed by the NCV group specifically for the C&V forecast product. Agile selection uses a specific skill score technique, with METAR observations as truth, to find the highest scoring component for the given lookback time period. The highest scoring component's forecast is chosen by the system. If there is a tie for the highest skill score the same precedence table used in the WMV technique is used to decide the component forecast to be used.

### 3. Data Set

The tests discussed in this paper were run for the 2110 CONUS sites over the six month period from 1 October 2006 to 31 March 2007, using available RUC, LAMP, METAR, and DM data. This data set was broken down into smaller site lists and time periods as required by the analyses.

### 4. Development Tests and Results

### 4.1 WMV vs. Agile

One of the initial decisions in generating the current version of the NCV forecast product was to decide between the two before-mentioned component integration strategies. This comparison was carried out by evaluating all statistics at various sites. Many of the sites evaluated showed very similar statistics for both integration schemes. As an example of the statistics figure 4 displays CSI of the Visibility for both integration schemes per forecast hour at the KLGA site. Because the Agile integration is more straightforward and its performance overall is somewhat better than WMV, it was chosen for the experimental version of the forecast system. Time constraints prohibited an extensive evaluation and configuration of the WMV to improve its statistical performance. However, the WMV is planned to be fully evaluated before the operational D4 AWTT decision.

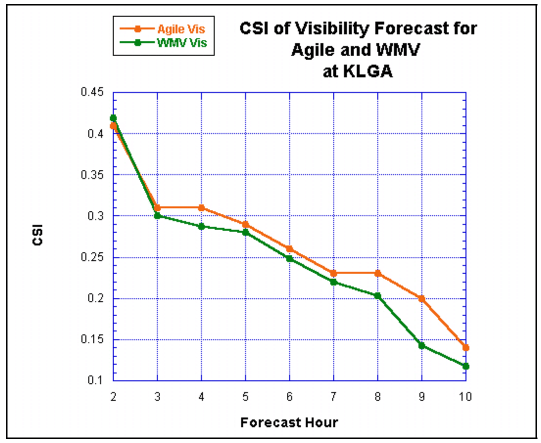



*Figure 4: Statistics from the 6-month test for the Agile (orange) and WMV (green) integration methods at KLGA.*

### 4.2 Agile Skill Score

The system started out using the Peirce Skill Statistic (PSS, Peirce 1884, Flueck 1987) as the skill score to determine the system's module choice for the Agile selection lookback time period. Through an over all statistical analysis of multiple sites it was determined that this skill score was not guiding the system to perform optimally. The system should perform at least as well as the best performing component/module. As seen in Fig. 5, using PSS as the optimizing skill score, the system is not producing that result. Plots of the False Alarm Rate (FAR), Probability of positive Detection (PODY), and Bias produced similar results -- as did the system's Visibility forecasts. Similar tests were conducted for the NCV forecast system optimized on Critical Success Index (CSI) and Heikde skill scores, and neither provided the performance that was expected.

A bi-modal skill score option was devised that used the probability of negative detection (PODN) to score the modules in the lookback time period when there were no IFR events (hits), and CSI when IFR events occurred. The system's performance improved from the 6-h to the 10-hr time periods, even though the CSI in the early hours is about the same as when the system was optimized on Pierce (Fig. 6). Plots of other sites also showed improved system performance with the Bi-Modal skill score. This technique is now used in the current version of the forecast system.

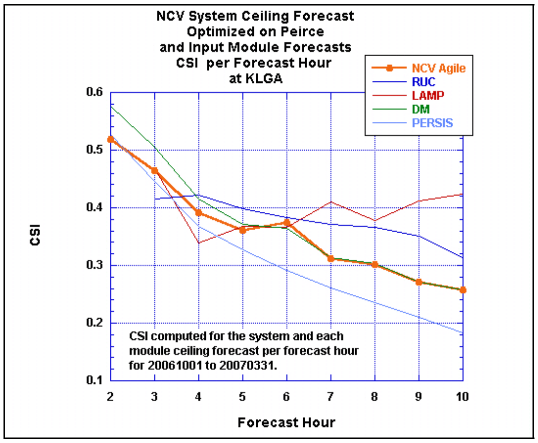



*Figure 5: Ceiling forecast CSI at KLGA over the 6-month time period from the NCV forecast system, optimized on the Peirce skill score, in orange and the input modules.*

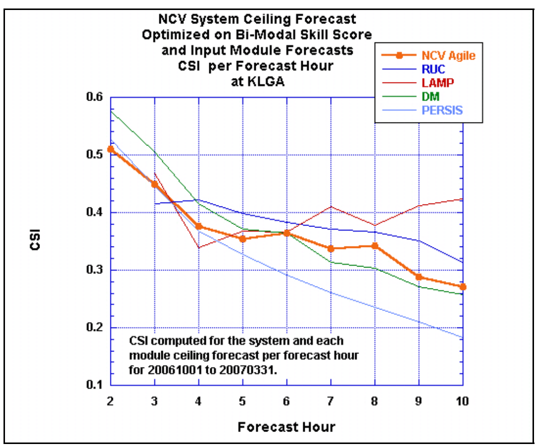



*Figure 6: Ceiling forecast CSI at KLGA over the 6-month time period from the NCV forecast system, optimized on the Bi-Modal skill score, in orange and the input modules.*

### 4.3 Statistical Analysis of the System Components

Many analyses were performed with the data from each of the system components. The initial reason for these tests was to explore how each of the NCV forecast system input component's overall statistics performed over different site lists.

Since LAMP was not available as often as the other components, a statistical analysis was conducted for all components only when LAMP was available (Fig. 7). LAMP is the best performer from 4 to 10 h, with DM and PERSIS coming a close second from 4 to 6 h. Looking at the KLGA site, LAMP does not become the best performer until 7 h but continues to be the top through 10 h. DM also shows good performance in early forecast hours. Both DM and LAMP add value to the system as a whole. Visibility plots of the same skill scores (not shown), showed similar results. The Visibility plots still show that LAMP and DM are top performers and add value to the system.

A smaller site list was used to compare DM with the other components. Through these tests we found that DM and LAMP both improved the statistical performance of the forecast system when they were available. Two precedence tables were composed from this data set. One table was created for the DM sites and another for the rest of the CONUS sites.

### 4.4 Lookback Time Periods

A statistical analysis of four lookback periods, 3, 6, 12, and 24 h, were conducted to find the time period that produced optimum performance of the forecast product. Statistics from different site lists were evaluated to ascertain the effect of lookback time period on the forecast system performance.

The difference in the skill scores between the different lookback time periods is very small (see Fig. 8). Overall, the smaller lookback time periods, 3 and 6-h, perform slightly better. The system is now using a 6-h lookback time period to take advantage of the better performance, but to allow enough data for the Bi-Modal skill score to work properly.

## 5. System Performance

Since all of the input components are only available at the 52 DM sites, a statistical analysis of CSI, FAR, PODY, and Bias was done for only these sites for the 6-month time period. The LAMP was still only available four times a day for the time period the system was run, so it is not giving the system the total added benefit of its performance for this evaluation; in the future when it is available hourly, the LAMP is expected to prove its worth based on previous performance.

As seen in Figure 9, the forecast system performed very close to the top performers for both C&V fields, excluding the LAMP (red lines) due to its different run-time schedule.. For the forecast system, both the C&V field's CSI were within 0.04 of the top two component performers per forecast hour, This result shows the ability of the agile selection method to capture the best performance of its input components. The performance of the forecast system will reach its maximum potential when it has well performing components available to use.

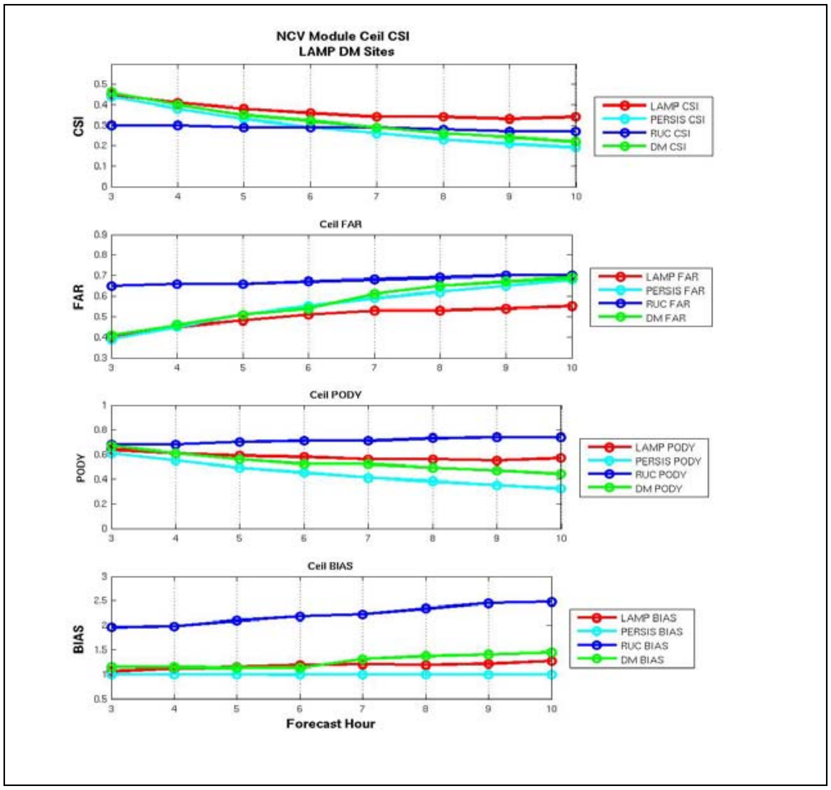
### 6. Summary and Future Work

This paper described the process by which the automated NCV forecast technique was developed. The current version of the forecast system has the following features:

- Four components: RUC, persistence, data mining, and LAMP
- All components available hourly with the exception of LAMP (only available xx times daily, soon to change to hourly)
- Hourly forecasts to 10 h
- Forecasts out to a 12-h when the LAMP is available. With planned upgrades to the LAMP issuance schedule, the forecast system will eventually produce a forecast out to 12h every hour.
- Agile selection
- 6-h lookback
- Optimized on a Bi-Modal CSI/PODN skill score.
- Two-dimensional, gridded output over the CONUS of ceiling, visibility and flight category (Fig. 6)

In the next year we plan these analyses and upgrades:

- Expand the number of data mining sites and evaluate data mining methodologies more appropriate to the infrequency of IFR events
- Perform an extensive evaluation of the WMV strategy and its inner workings; and use case studies to determine how well it handles prediction of IFR events compared to the agile selection strategy
- Examine other skill score strategies to score the different components
- Divide the CONUS into regions for statistical analysis to ascertain regional characteristics.
- Improve the IFR-VFR discrimination skill by creating an IFR-only data set to be used to evaluate the performance of the forecast system and its individual components
- Compare the forecast product against TAFs
- Include other components that have demonstrated positive skill in forecasting C&V



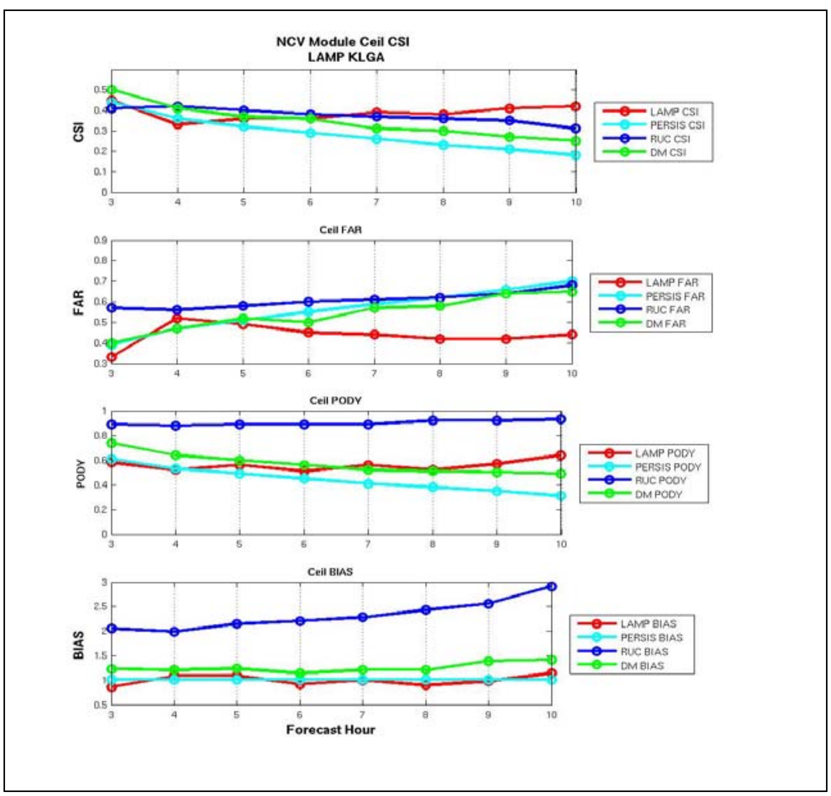
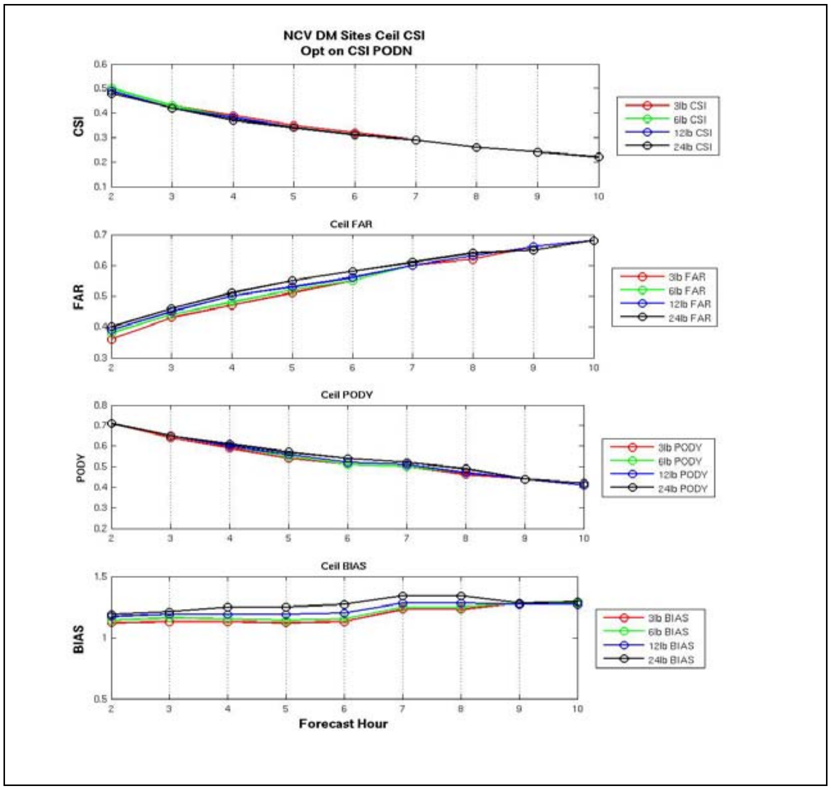




*Figure 7: Ceiling forecast CSI, FAR, PODY, and BIAS for all of the DM sites for 3, 9, 15, and 21z initialization times for all of the components per forecast hour from 3hr to 10hr. Top: all sites; bottom: KLGA.*



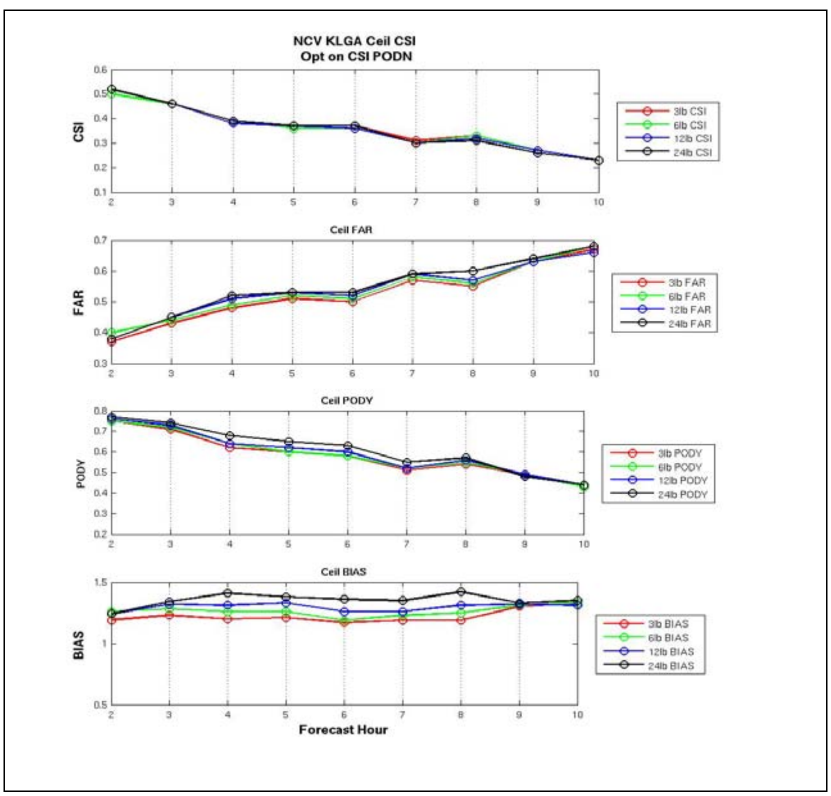




*Figure 8: CSI, FAR, PODY, and BIAS for each forecast hour over the 6-month time period of the NCV forecast system 3, 6, 12, and 24-h lookback time periods over the DM site lists and for KLGA.*

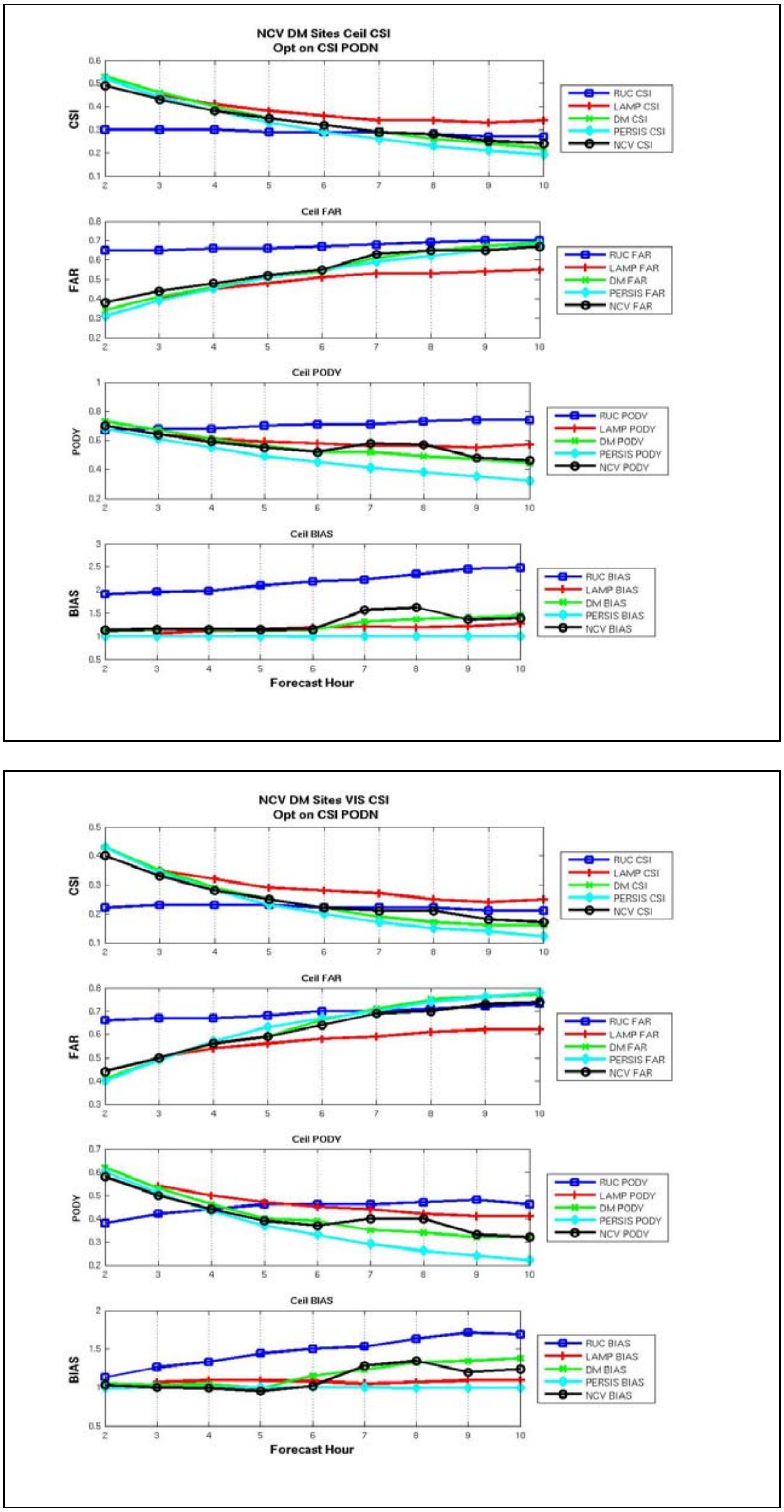



*Figure 9: CSI, FAR, PODY, and BIAS for each forecast hour over the 6-month time period of the experimental version of the NCV forecast system and the input components over the DM site lists. Top: Ceiling, Bottom: Visibility.*

**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 or position of the FAA.

**References**

Benjamin, S., T. G. Smirnova, K. Brundage, S. S. Weygandt, T. L. Smith, B. Schwartz, D. Devenyi, J. M. Brow, G. A. Grell, 2005: A 13-km RUC and beyond: recent developments and future plans, *22nd Conference on Severe Local Storms,* Seattle, WA.

Blum, A., 1996: On-Line Algorithms in Machine Learning, *Lecture Notes in Computper Science*, Vol 1442, pages 306-325.

Glahn, J. E. Ghirardelli, 2004: The new and improved Localized Avaition MOS Program (LAMP) analysis and prediction system, *17th Conf. on Probability and Statistics in the Atmoshperic Sciences,* Seattle, WA.

Herzegh, P., G. Wiener, R. Bateman, B. Weekley, 2005: *Automated forecast of ceiling and visibility for avaitaion: exploration of a data mining-based forecast method,* 21st Conf. on Interactive Information Processing System for Meteorology, San Diego, CA.

Flueck, J.A., 1987: A study of some measures of forecast verification. *Preprints, 10th Conf. Probability and Statistics in Atmospheric Sciences*, Edmonton, Alberta, Amer. Meteor. Soc., 69-73.

Peirce, C.S., 1884: The numerical measure of the success of predictions. *Science*, **4**, 453-454.

Skiena, Steven S.: 1997m, *The Algorithm Design Manual*, Telos.