P3.6

A CHARACTERIZATION OF NWP CEILING AND VISIBILITY FORECASTS FOR THE TERMINAL AIRSPACE†

Paul E. Bieringer*
Michael Donovan
Frank Robasky
David A. Clark
Jonathan Hurst

Massachusetts Institute of Technology Lincoln Laboratory
244 Wood Street
Lexington, MA 02420

1. INTRODUCTION

The Federal Aviation Administration (FAA) is sponsoring a Terminal Ceiling and Visibility (C&V) initiative to provide automated C&V guidance to the air traffic managers for both tactical (0-2 hour) and strategic (3-12 hour) decision making. To meet these requirements, particularly in the strategic time frame, it will most likely be necessary for the C&V system to incorporate guidance from an explicit numerical weather prediction (NWP) model. If NWP forecasts are found to be suitable for this application, they will be used as the backbone of the terminal C&V forecast system. More details on the terminal area C&V forecast product development for the FAA can be found in Allan et al. (2004). Before these NWP forecast products can be used, it is necessary to first characterize their accuracy relative to operational air traffic control (ATC) requirements. This makes it possible to exploit observed strengths, avoid weaknesses, and facilitate a better utilization of NWP forecast products.

This study provides an assessment tailored specifically to address the terminal C&V application. Consequently, the results represent forecast performance for relatively small geographic locations that for practical purposes can be considered point forecasts. It is our intention to answer four questions with this preliminary analysis:

1. How accurate are the NWP forecasts relative to the observational truth and a human generated forecast?

2. For the terminals of interest to this study (i.e. New York City Airports), are there any advantages to utilizing a non-hydrostatic mesoscale model run at horizontal resolutions of 3 km or less?

3. Do the NWP models exhibit forecast skill for non-traditional forecast metrics such as trends in C&V parameters and timings of threshold crossings associated with the onset and clearing of low ceiling and visibility conditions?

4. Are there obvious situations/conditions during which the NWP forecasts have more/less skill?

In addition to a report on the NWP terminal ceiling and visibility forecast accuracy, we provide preliminary recommendations on the direction we feel this line of research should pursue, and where we see opportunities to utilize NWP forecasts in an automated terminal C&V decision guidance system.

An ancillary goal of this study is to assemble the analysis software infrastructure required to quantitatively evaluate numerical forecast accuracy. We envision using these tools to develop and test modifications to the translation algorithms and techniques that will be necessary to integrate the NWP forecasts into the C&V guidance system. They will be instrumental in reducing the time required to make engineering turns during the upcoming development and implementation stages of this research.

---

†This work was sponsored by the Federal Aviation Administration under Air Force Contract FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

**Corresponding author address*: Paul E. Bieringer, MIT Lincoln Laboratory, 244 Wood Street, Lexington, MA 02420-9185; e-mail: paulb@ll.mit.edu

2. BACKGROUND

Changing ceiling and visibility conditions can have a significant impact on terminal ATC operations. The C&V conditions can influence how the airport operations are configured, and often will influence airport acceptance rates due to the necessity for increased arrival aircraft spacing. Air traffic managers currently rely on the terminal aerodrome forecast (TAF) when making operational decisions impacted by low ceiling and visibility conditions. This forecast is manually generated by a meteorologist at the National Weather Service forecast office (NWSFO) that is responsible for the aviation forecasts for that airport. The TAF provides a 24-hour forecast with atmospheric conditions specified for each terminal area at the top of the hour. It is issued every six hours starting at 00 UTC and updated when conditions warrant. The TAF provides forecasts of cloud ceiling coverage conditions in intervals of 100 feet up to 25,000 feet, and visibility in fractions of a statute mile up to 6 statute miles. Ceilings are derived from the lowest broken or overcast layer or vertical visibility forecast. Ceilings above 25,000 feet and visibilities above 6 miles are considered unlimited. Since the TAF is the product that is currently used operationally for C&V decision making it is also evaluated in this analysis and serves as a pseudo metric along with persistence (i.e. a continuation of current conditions) against which the NWP forecasts are evaluated.

Ceiling and visibility forecast output from two NWP forecast models, the Rapid Update Cycle (RUC) and the Pennsylvania State University (PSU)/National Center for Atmospheric Research (NCAR) fifth generation mesoscale model (MM5) are evaluated in this study. The RUC data were generated at the National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory (ESRL) Global Systems Division (GSD) and were taken from the 20 km horizontal resolution operational back-up version of the RUC model run by the National Center for Environmental Prediction (NCEP). Details regarding the operational configuration of the RUC model can be found in Benjamin et al. (2004a) and Benjamin et al. (2004b).

The MM5 forecasts were generated at Massachusetts Institute of Technology Lincoln Laboratory (MIT LL) and were initialized with the RUC analyses. RUC analyses were used because the forecast grids were not available for the offline simulations. Consequently, this analysis is not suitable to examine the overall forecast performance between the RUC and various MM5 simulations. MM5 was run in a nested configuration that provided 27, 9, and 3 km horizontal resolution forecasts centered over New York City (NYC). The model simulations were cold-started from RUC analysis and forecast grids and used no additional observations. Ceiling and visibility computations from the MM5 data were derived from translation algorithms based on the RUC translation algorithms. Additional details on the configuration of the 3 MM5 domains can be found in Appendix A. Details regarding the MM5 forecast model can be found in Grell et al.1994.

It is likely that the overall accuracy of the numerical forecasts could be improved using additional observations or alternative physics settings; however, since the forecast accuracy results are comparable to those from the operational RUC, we believe they are suitable for the purpose of this analysis. Our intention is not to show that one NWP model is better than another. The purpose of including the MM5 forecasts is to develop a preliminary sense of the benefits provided by higher horizontal grid resolutions in a non-hydrostatic weather forecast model.

NWP models do not forecast cloud ceiling height and horizontal visibilities directly; consequently translation algorithms are required to compute ceiling and visibility from the NWP prognostic variables. For this study, computed ceiling and visibility values were directly available from the GRIB formatted RUC data provided by ESRL GSD. The MM5 ceiling and visibility values were computed using the RUC translation algorithms (Brown 2004). The cloud base height ($H_{cld-base}$) is defined as the lowest level at which the combined cloud and ice mixing ratio exceeds $10^{-6} g/g$. The visibility translation algorithm is a modified version of the Stoelinga-Warner algorithm that includes a modification to use relative humidity to provide clear-air visibilities. The horizontal visibilities associated with precipitation hydrometeors follow that used by Stoelinga and Warner (1999), with the addition of an extinction coefficient for graupel developed specifically for the RUC. The clear-air visibilities are computed using the maximum of the relative humidity in the 2 levels of the model closest to the surface (rh_{1-2}). Visibility is defined as:

$$V_{rh} = 60.0 * \exp(-2.5 * q_{rh})$$

Where q_{rh} = the min (80.0, (rh_{1-2} /100.0 – 0.15)). This algorithm gives a visibility of 5.4 km at 95 % relative humidity (Brown, 2004).

Persistence and NWP forecasts are compared with ceiling and visibility measurements taken by the Automatic Surface Observation System (ASOS) at Newark Intl Airport (EWR), John F. Kennedy Intl. Airport (JFK), LaGuardia Intl Airport (LGA), and Macarthur Islip Intl Airport (ISP). The ceiling and visibility observations taken nearest to the top of the hour but no further than +/- 15 minutes from the top of the hour were used. Ceiling observations taken by an automated ceilometer are reported in 252 50-foot interval bins and the system reports cloud layers up to an altitude of 12,000 feet. Visibility measurements are made with a Belfort model 6220 forward scatter visibility meter and report visibility in statute miles in the following categories: <1/4, 1/4, 1/2, 3/4, 1, 11/4, 11/2, 13/4, 2, 21/2, 3, 4, 5, 6, 7, 8, 9, 10+ (US Dept. of Commerce, 1992).

3. ANALYSIS AND RESULTS

Several criteria were used to identify the cases that were evaluated. An event was considered if the observed ceiling and/or visibility conditions deteriorated to a point where ATC operations at any of the four terminals would be significantly impacted. ATC impact was considered to be significant if C&V conditions were marginal visual flight rules (MVFR) or poorer for more than six hours. The FAA definition of a MVFR condition is when the cloud ceiling is between 1,000 and 3,000 feet and/or the surface visibility is between 3 and 5 statute miles. Precipitation, frozen or liquid, was not a deciding factor during case selection; however, in each of the 13 cases selected, precipitation was observed for several hours and contributed to the diminished ceiling and visibility conditions. In order to capture the entire C&V event, forecasts and observations were collected 12 hours prior to the deterioration of conditions to MVFR through 12 hours after the improvement of conditions to better than MVFR. Care is taken to not overpopulate the data set with unlimited ceiling and visibility conditions that might skew the statistical results.

Since the focus of this study is to evaluate forecast performance within the context of how operational decisions are currently made, this study categorized both the forecasts and observations. ASOS observations, TAF, and NWP forecasts of ceiling and visibility were broken into seven categories equivalent to those used operationally in the global forecast system (GFS) model output statistics (MOS). The GFS MOS category definitions for ceiling and visibility are illustrated in Table 1.

TAF and NWP Forecasts from 30 days (13 events) in the 03-04 winter C&V season where low ceiling and visibility conditions were present at the 4 airports listed above were selected for the analysis. Data were collected during a period extending from December 5, 2003 through May 31, 2004. Forecasts from the RUC model range from 1 – 24 hours at 1-hour intervals for hours 1 – 6 and at 3-hour intervals for hours 6 – 24. Forecasts from the MM5 range from 1 – 18 hours at 1-hour intervals. Since the TAF forecasts are only issued on a 6-hour cycle (00, 06, 12, and 18 Coordinated Universal Time (UTC)), NWP forecasts issued at these times were used so that comparable statistics could be compiled for all forecasts. Forecast performance was evaluated in four ways:

1. Categorical forecast performance was evaluated on a forecast-by-forecast basis. This allowed for a manual inspection of the data to prevent gross errors from being injected into a combined analysis. It also provided a means to characterize patterns in the forecast performance results that could be exploited or avoided.

2. Forecast accuracy distributions were computed for each C&V forecast at each hour. This provides an evaluation of the forecast in terms of bias and standard deviation of the forecast accuracy as a function of forecast length.

3. The categorical forecast accuracy, categorical mean error, and standard deviations of categorical error (from the hourly error distributions discussed above) were plotted as a function of time. These plots summarize the analysis results and illustrate the variations in performance of each forecast relative to other forecast time horizons.

4. Statistics representing the spatial variability characteristics of the high-resolution (3 km MM5) simulations were evaluated. This analysis was conducted to ascertain the relative value of using spatial variability to improve forecast accuracy over a point forecast and as a proxy for forecast error.

TABLE 1
Range of values defining each ceiling and visibility category level from the Global Forecast System (GFS) Model Output Statistics (MOS). The category is selected based on the lowest ceiling or visibility condition at the time

Category	Ceiling (feet)	Visibility (statute miles)
1	< 200	< ¼
2	200 < 500	¼ < ½
3	500 < 1,000	½ < 1
4	1,000 < 3,100	1 < 3
5	3,100 < 6,600	3 < 6
6	6,600 ≤ 12,000	6
7	> 12,000 or unlimited	> 6

Time series plots of the categorical C&V forecasts and observations were generated for each of the four NYC stations at each regularly scheduled TAF forecast cycle. Overall, comparisons between the observations and forecast data show that both the model and manually generated forecasts have some skill in capturing trends in conditions but often show latency in forecasting the onset of lowered (lifting) ceilings and reduced (increased) visibilities. This is illustrated in Figure 1, which shows an example of a winter storm approaching JFK on March 16, 2004. All forecasts in this case were 2 – 3 hours too slow in predicting the rapid decline of cloud ceiling and surface visibility. Timings of forecast condition transition times similar to the one shown in Figure 1 were quite variable and observed to be both early and late. In situations where the model ceiling forecasts are relatively steady state they are often within 1 category of the observational reports. Overall, visibility forecasts are typically too pessimistic and several hours late to forecast improving conditions. These tendencies were common throughout most of the events analyzed. Given the non-linear range of values between the ceiling and visibility categories (Table 1), it's common for model forecasts to be off by one or more levels during operationally significant conditions and vary less often when conditions are less restrictive. While the use of categorical forecasts does not explicitly characterize the overall forecast accuracies, it does provide an effective measure of forecast accuracy relative to operational C&V decision points.

The timing of the onset of changing conditions appears to be one significant source of error in the NWP forecasts. During these periods the forecasts of ceiling heights are often off by several categories. When conditions become more steady state, the NWP ceiling height forecasts appear to be more accurate. This same signal appears to be present in the visibility data as well; however, it is more difficult to discern due to the larger forecast errors. These results suggest at least two things:

1. NWP forecasts appear to be more robust when conditions are more steady state.

2. A reduction in the spatial/temporal biases may yield a reduction of forecast errors during this period during which conditions are transitioning.

It is often difficult to draw quantitative conclusions from a large set of separate time series plots. One method of consolidating results is to use frequency distribution histograms illustrating how often differences (i.e. forecast errors) in category levels between forecast and observations occur. This technique was used in this analysis to compile forecast performance statistics as a function of location and forecast length. Due to the limited number of data points per station, it was necessary to combine the data from all of the stations into a single analysis.

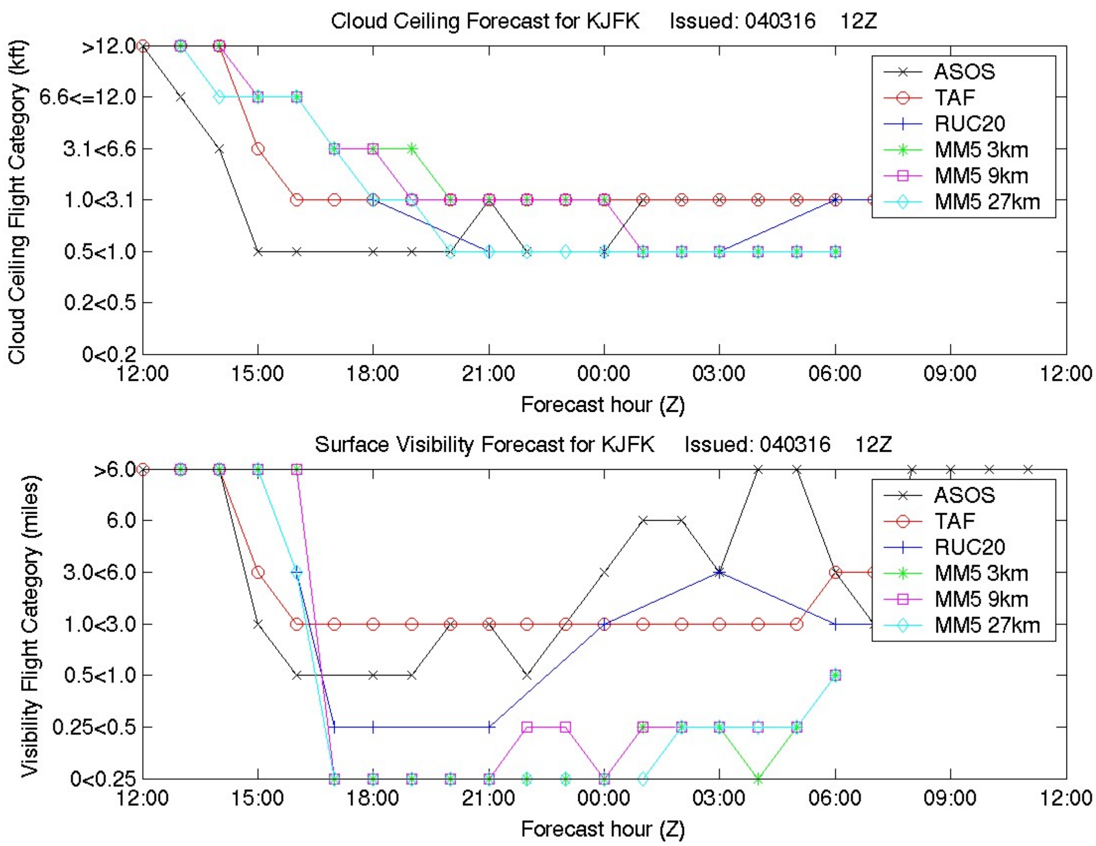

Figure 1. Time series plot of the categorical observations, TAF, and NWP model forecasts of cloud ceiling (top) and visibility (bottom) beginning at the 12:00 Z TAF forecast cycle for a late season winter snowstorm approaching JFK on March 16, 2003.

Although each location has its own local weather idiosyncrasies, combining the stations is still reasonable given the grid resolutions of the forecast models and the fact that they are all relatively close to the coast. With the addition of more cases from the 04-05 C&V season it should be possible to examine the forecast performances by location.

Examples of the relative frequency distribution of category level differences between ceiling and visibility forecasts and observations for the 6-hour forecasts are shown in Figures 2 and 3, respectively. The results from all of the frequency distributions as a function of forecast length are summarized in Figures 4 and 5. As expected, the distributions of TAF ceiling forecasts show the best symmetry and are typically the most accurate. The only exception to this is for forecasts of 3 hours or less where the persistence forecasts were the most accurate (Figure 4). TAF forecasts of visibility have similar accuracy distributions but are significantly more accurate than comparable NWP model forecasts. The visibility persistence forecast provides a comparable or better forecast than the TAF for forecasts out to six hours (Figure 5). The results also indicate that a bimodal distribution is present in the NWP categorical visibility forecast errors in the NWP forecasts (Figure 3).

The specific source of this bimodal distribution is unknown at this time, and unfortunately a more in-depth investigation was beyond the scope of this preliminary study. We suspect that the erroneous mode in the visibility forecast may be associated with the component of the visibility associated with model precipitation. In this study the MM5 forecasts were cold started from the RUC analyses and typically have less precipitation in the 1-2 hour forecasts than in longer forecast times. The absence of a significant bimodal distribution in the 1 and 2 hour MM5 forecasts supports the hypothesis that precipitation is not being properly translated into visibility estimates.

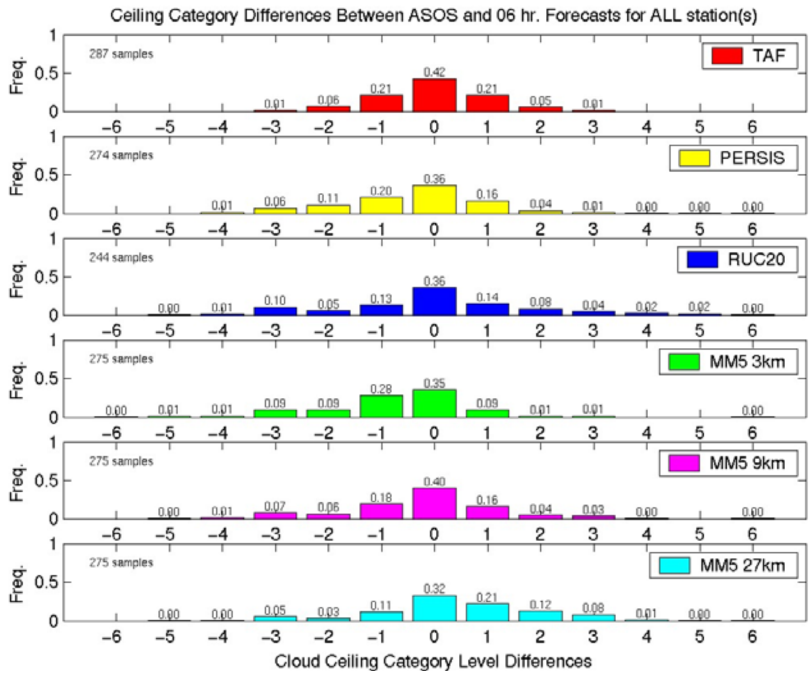

Figure 2. Relative frequency distribution of all category differences between cloud ceiling observations and forecast data for 6 hour forecasts for all NYC stations. Negative (positive) category differences represent ceilings that were observed to be lower (higher) than what was forecasted.

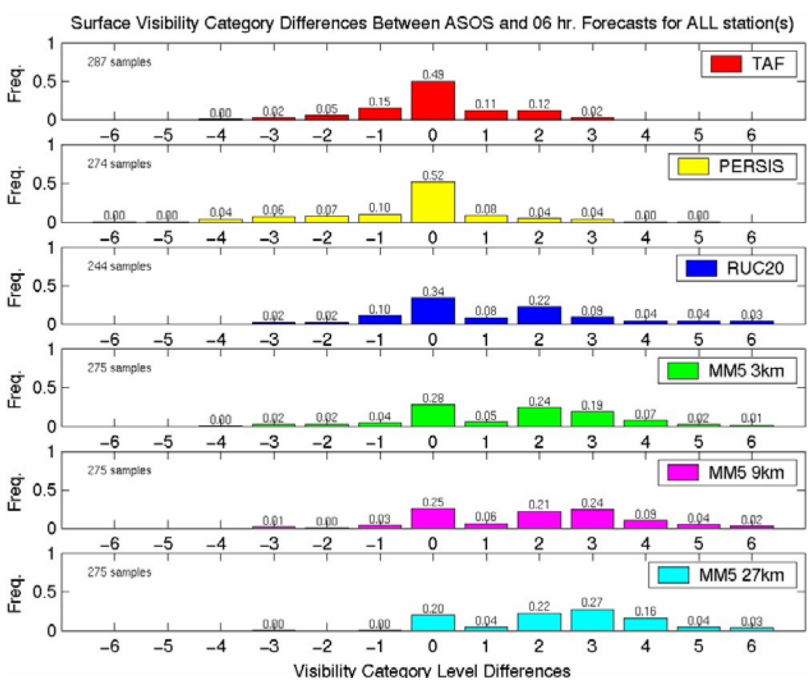

Figure 3. Relative frequency distribution of all category differences between visibility observations and forecast data for 6-hour forecasts for all NYC stations. Negative (positive) category differences represent visibilities that were observed to be lower (higher) than what was forecasted.

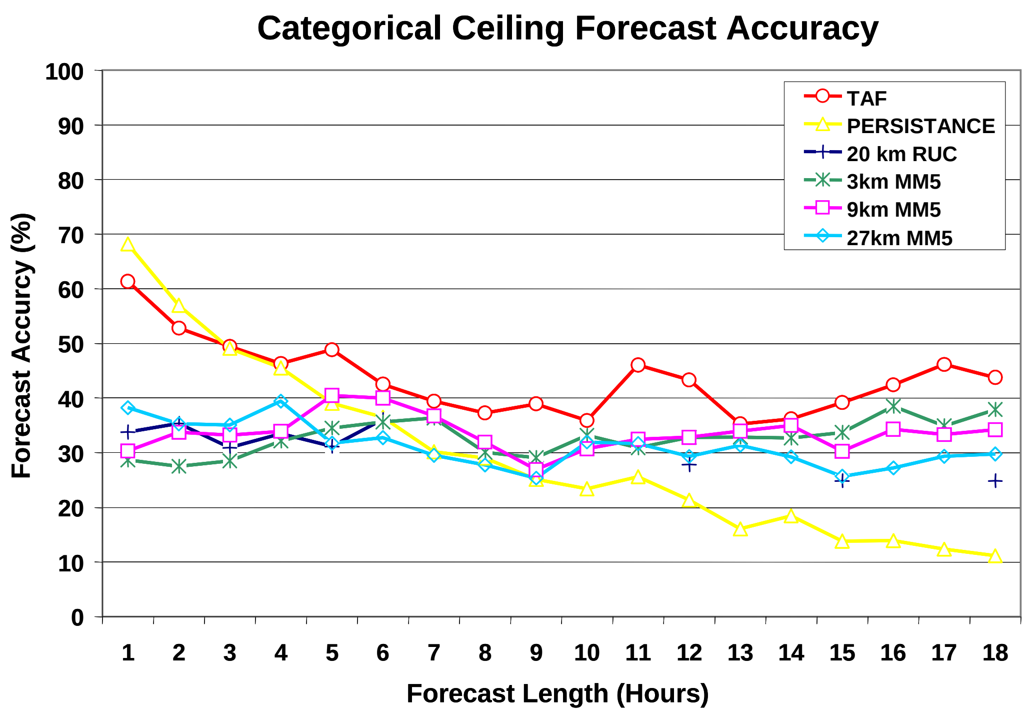

Figure 4. A plot of percentage of accurate categorical ceiling forecasts versus forecast length.

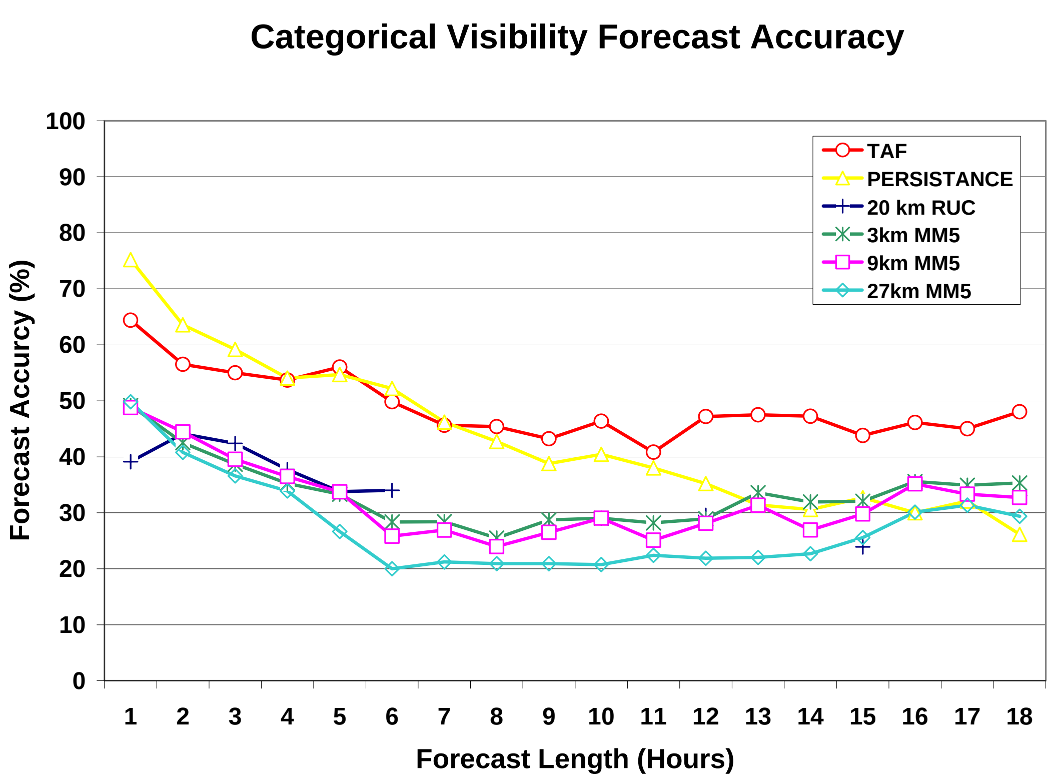

Figure 5. A plot of percentage of accurate categorical visibility forecasts versus forecast length.

All of the ceiling height NWP forecasts tend to be optimistic in the short range, and forecast ceiling heights that are higher than observed. Conversely, at longer forecast time ranges, the models are typically more pessimistic and forecast ceiling heights that are lower than observed. Frequency distributions of visibility category differences indicate there are distinct positive biases at all forecast hours and a higher frequency of large errors in the model forecasts. When comparing the NWP models among themselves the MM5-27km model initially outperforms the others in the first few forecast hour intervals. For the longer forecast lengths the higher resolution MM5 simulations outperform the rest of the NWP forecasts.

Figure 6 illustrates the mean and standard deviation of the distributions of categorical ceiling height forecast errors versus forecast length. At most forecast times, the TAF forecast has lower categorical mean errors than any of the NWP forecasts; however, over most forecast lengths, the NWP forecasts had mean categorical errors of less than one category. Unfortunately the standard deviation in the NWP forecast accuracy is not as good, and was routinely greater than 1.5 categories. The TAF forecasts had a standard deviation that ranged from 1 to 1.5 categories and was typically 1 category lower than the NWP forecasts. This suggests that there is a fair amount of variability in the NWP forecasts and that they are not as robust as the TAF. Among the NWP forecasts, the RUC forecasts had the lowest mean errors but showed the most forecast variability. While the higher resolution MM5 (3 km) forecast had larger mean errors when compared to the lower resolution model forecasts in the short range, it also had the least variability at nearly every forecast hour. Additionally, the 3km-MM5 model consistently forecast higher ceilings than observed.

Figure 7 illustrates the mean and standard deviation of the distributions of the categorical visibility forecast errors versus forecast length. As in the comparable ceiling plots, the human generated TAF performed best. Mean categorical errors of 1-2 categories were typical for all forecast time horizons, and standard deviations ranged from 1.5 to 2.25 categories for all of the NWP model forecasts. Overall these results suggest that the NWP visibility forecasts were poor. Of the NWP models the RUC forecasts outperform the others with the lowest mean error and standard deviation. One interesting finding, however, was that while the variability among the MM5 forecasts is relatively consistent at the various resolutions, the mean error decreases as the model resolution is increased. It is important to also note that the MM5 is cold-started in this analysis. The rapid increase in mean error in the 1 to 2 hour MM5 forecasts results from the lack of precipitation during model spin-up. This rapid increase coincides with the increase in precipitation in the model and occurs because precipitation is used by the visibility translation algorithms. This decrease in accuracy corresponds to the increase in the erroneous mode in the categorical forecast accuracy frequency distributions discussed earlier.

In theory, higher resolution simulations should be better able to capture fluctuations in the C&V forecasts associated with mesoscale features and forcings due to localized variability in the land sea boundary along the coast. Furthermore, high-resolution simulations provide significantly more data points for a given area that can then be used in a translation algorithm that converts the model forecast to a terminal ceiling and visibility forecast. We hypothesize that information derived from the NWP forecast from multiple points around the terminal area may prove to be more accurate than a single point and provide an error metric that can be used to ascertain the accuracy of a point forecast. The belief is that this may be particularly beneficial during times where a transition in ceiling and visibility conditions is occurring. To test this hypothesis spatial statistics from the 3 km MM5 forecasts were evaluated over a 54 km x 54 km evaluation box centered over the 4 NYC area airport ASOS observation sites. This analysis covered both ceiling and visibility forecasts; however, due to the gross inaccuracies currently present in the visibility forecasts we believe the visibility results from this analysis are of limited use at this time.

As a first step we evaluated time series plots of mean error and standard deviation in the ceiling and visibility forecasts from the 3 km MM5 forecasts over the evaluation areas. Due to model spin-up in the MM5 forecasts, it was necessary to focus the analysis on MM5 forecasts greater than 2-3 hours. Throughout the entire set of time series plots examined, there was often an inverse relationship between the standard deviation in the ceiling forecast and the mean ceiling height forecast error. Figure 8 illustrates an example of this inverse relationship. The signature is characterized by increases in standard deviation that coincide with times where the forecast error deviates from 0. In Figure 8 this signature occurs between forecast hour 7 and 11 and again between forecast hours 13 and

15. Overall, it was encouraging to find that spatial variability in the forecast may be used as a proxy for forecast accuracy; however, further examination of these fields will be required in order to quantify this finding and use it in the operational terminal ceiling and visibility system.

In addition to the mean error and standard deviation of the NWP forecasts inside the evaluation box, the categorical mean, maximum, and minimum values of the NWP forecasts were evaluated. The maximum and minimum values were used to compute a dynamic range of the forecasts in the evaluation box that along with the categorical mean value, was contrasted against the categorical point forecast. An example of a time series plot illustrating a typical result of this analysis is shown in Figure 9. An initial review of these time series results suggest that the dynamic range may possibly be used to improve the output from the translation algorithms. Here dynamic range would be used as a surrogate of standard deviation inside the evaluation box. Based on results like those shown in Figure 8 we hypothesize that if there is low variability among neighboring values, the atmospheric conditions are in steady state and the mean NWP forecast value may be most representative. Conversely, the NWP point value forecast may be best to use when large spatial variability in the forecast fields are present that may tend to skew the mean value forecast.

The benefits of using dynamic range of the forecast to highlight its accuracy are illustrated in Figure 10. In this plot the categorical forecast accuracy of the point and mean forecast for cases where the dynamic range in the NWP forecast is less than or equal to 2 categories are contrasted with the accuracies of a persistence, TAF, and NWP point value forecast. In these situations the forecast accuracies are improved when compared to the overall average for the point forecasts (Figure 4) and relative to times when the dynamic range is greater than 2 (Figure 11). These plots indicate that the mean forecast value provides a forecast of comparable or better accuracy to the TAF and persistence forecasts at forecast hours beyond 5 hours. The mean forecast in these situations is in most cases more accurate than the point forecast by 10-15% for all forecast hours. These results along with the standard deviation results shown in Figure 9 suggest that spatial variability in the ceiling forecasts from high-resolution NWP models has the potential to be used to characterize NWP forecast accuracy.

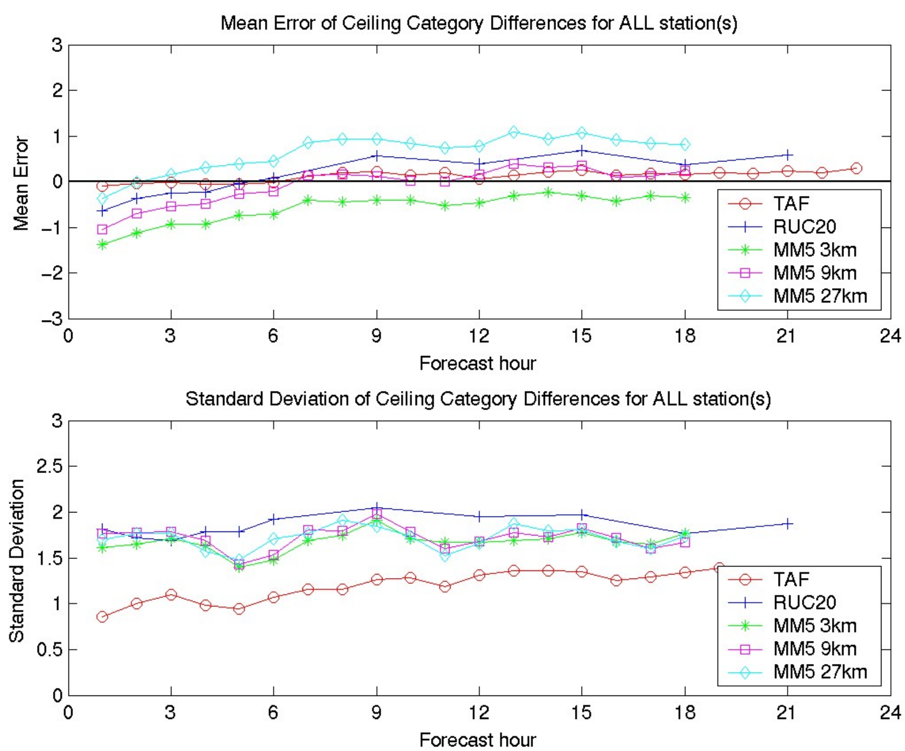

Figure 6. A plot of mean categorical forecast error (top) and standard deviation (bottom) in the ceiling forecasts. Categorical error is defined as observation category – forecast category.

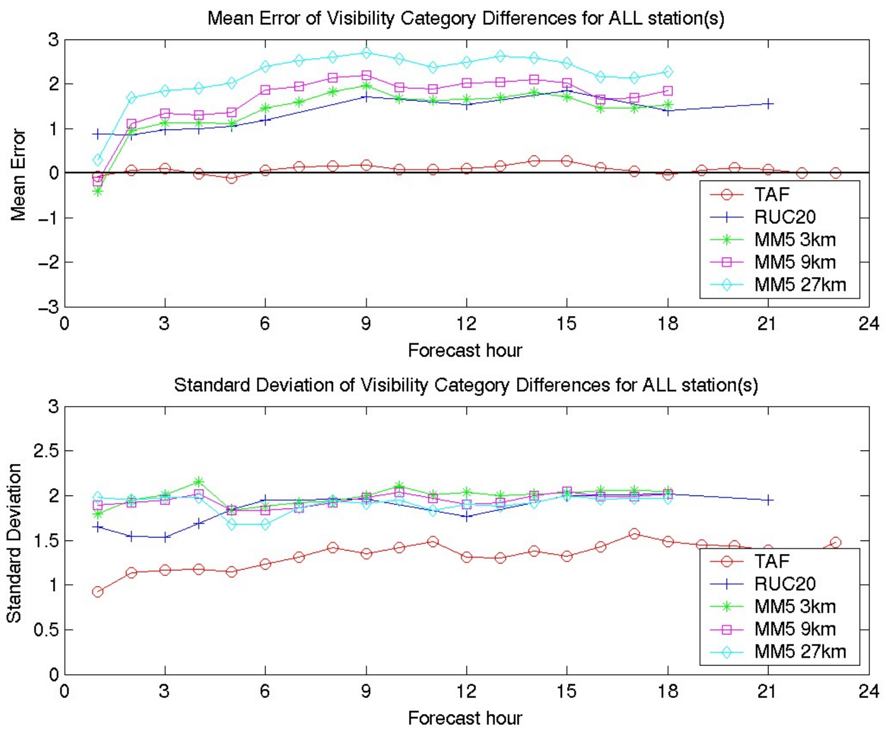

Figure 7. A plot of mean categorical forecast error (top) and standard deviation (bottom) in the visibility forecasts. Categorical error is defined as observation category – forecast category.

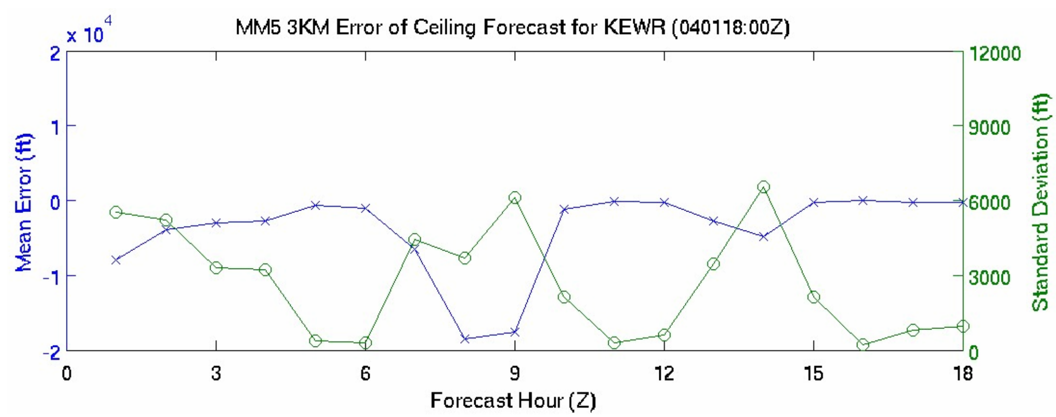

Figure 8. A plot of the mean ceiling height and standard deviation of the NWP forecasts inside a 54km x 54 km evaluation box. Increases in the standard deviation often correspond to deviations in the mean forecast errors.

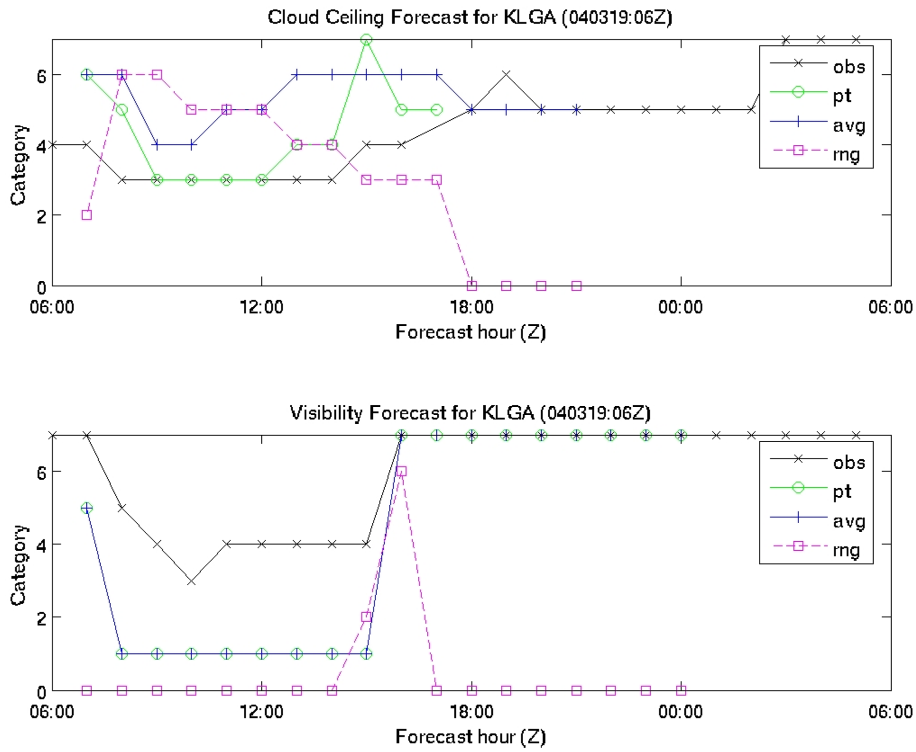

Figure 9. A plot of the categorical ceiling (top) and visibility (bottom) observations and the MM5 (3 km) categorical point, mean, and dynamic range forecasts for the 06Z forecast cycle at KLGA on March 19, 2004.

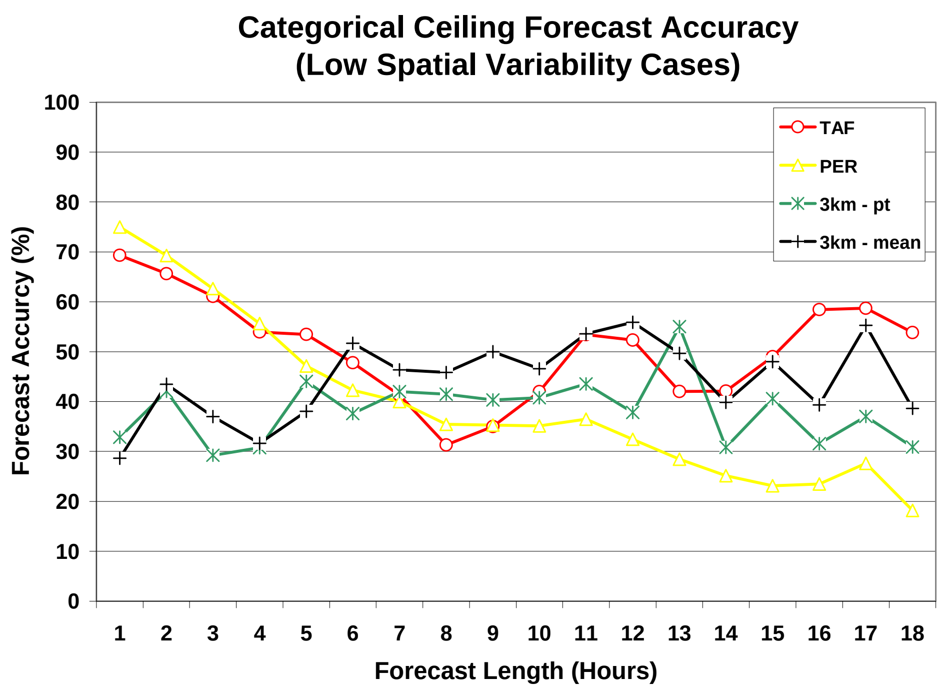

Figure 10. A plot of percentage of accurate categorical ceiling forecasts from the 3km MM5 simulations versus forecast length for cases where the MM5 forecast exhibited low spatial variability. Low spatial variability was defined as situations in which the range of categorical forecasts was less than equal to 2 categories. The MM5 mean value forecasts are typically 15-20% more accurate than during high spatial variability situations and beyond 5 hours are on par or more accurate than the TAF.

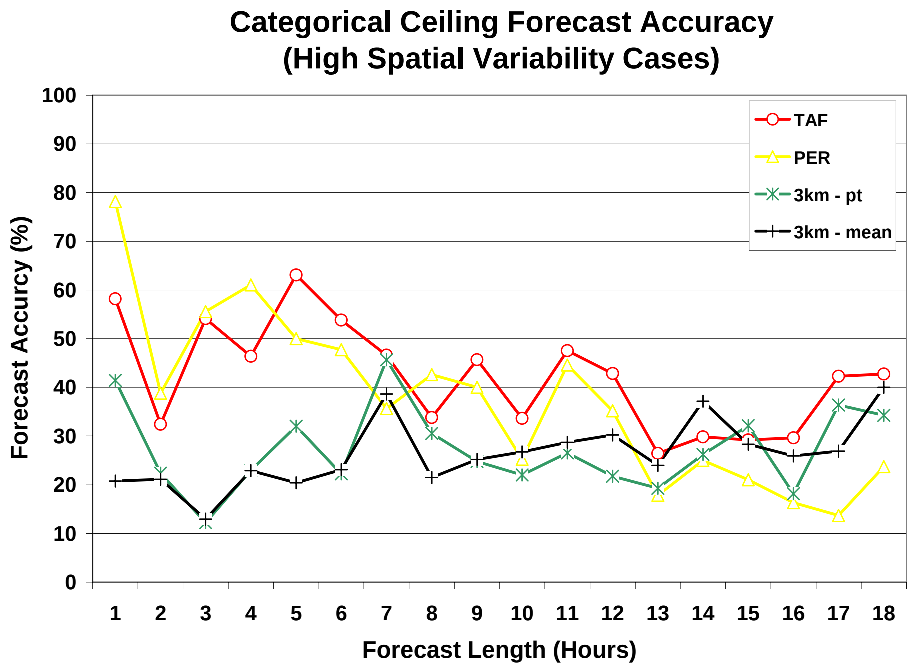

Figure 11. A plot of percentage of accurate categorical ceiling forecasts from the 3km MM5 simulations versus forecast length for cases where the MM5 forecast exhibited low spatial variability. High spatial variability was defined as situations in which the range of categorical forecasts was greater than 2 categories. The MM5 forecasts are typically 15-20% less accurate than during low spatial variability situations.

4. CONCLUSIONS AND FUTURE WORK

This study provided a brief statistical review of the forecast accuracies of the TAF, NWP, and persistence forecasts relative to the terminal C&V operational decision points. In addition to assessments of mean and standard deviations of the categorical forecast accuracy, this analysis also examined non-standard forecast accuracy associated with trends and spatial variability in the high-resolution simulations. An ancillary goal of this study was to assemble the analysis software infrastructure required to quantitatively evaluate numerical forecast accuracy.

Ceiling height forecasts were the more accurate of the two NWP C&V parameters evaluated. For forecasts greater than six hours, the analysis indicates that overall the NWP forecasts of cloud ceiling heights are accurate between 30 and 40 % of the time while the categorical visibility forecasts are accurate between 20 and 30 % of the time. These results reflect what is likely the worst-case performance for these models. This is particularly true for the MM5 forecasts that were cold started and utilized no additional observations.

Some of the results however are encouraging. The NWP forecasts were observed to capture the same overall trends that were present in the observational data. In particular, the NWP models showed more skill in forecasting cloud ceiling heights during situations of steady state low ceiling conditions. Without any tuning of the analyses or translation algorithms, the distributions of the categorical forecasts are Gaussian with mean values near zero. Furthermore the percentage of accurate categorical model forecasts approach the accuracy of the TAF for forecasts of 5-6 hours and greater. Unfortunately, the standard deviation of the error distributions for the NWP forecasts are much broader than the TAF forecasts. This is likely the major contributor to the lower percentage forecast accuracy statistics. As mentioned previously, large ceiling height forecast errors occur during the transition periods when conditions are degrading or improving from periods of extended steady state conditions. This behavior opens up the opportunity to reduce the impact of these errors on the terminal C&V forecast by reducing these errors in the forecasts. This could be accomplished through a systematic identification of the conditions that lead to these errors and then the subsequent avoidance of them in the implementation of the NWP forecasts.

We recommend exploiting this characteristic of the NWP forecast accuracy to the extent possible when implementing NWP ceiling forecasts in the terminal C&V forecast system.

Unfortunately, the statistics are not as encouraging for the visibility NWP forecast products. The initial results suggest that the translation algorithms used to characterize visibility from the NWP forecasts have a distinct bias in many situations to forecast lower visibilities than observed. The standard deviation of the categorical forecast errors is very high for the visibility forecasts suggesting that the forecasts and/or translation algorithms are not very robust. This is confirmed by an examination of the raw distributions, which often show a bimodal distribution (Figure 3). Again these results reflect the forecast performance of the NWP models and their unmodified translation algorithms, and should be considered worst case. In their current state, the visibility forecasts from the NWP models would be only marginally useful. Before implementation, the forecast accuracy will need to be significantly improved. We recommend a deeper examination of both the translation algorithms and the type of situations which lead to the bimodal distribution. If the erroneous mode can be identified, and then corrected or avoided, some subset of the visibility forecasts may be useful for implementation.

Although not conclusive, there is some evidence that increased horizontal resolution may yield improvements in the NWP C&V forecast products. Overall the MM5 forecasts produced results that were on par with the RUC forecast accuracy performance. The higher resolution MM5 forecasts yielded equivalent and in certain instances marginally better performance statistics than the more coarse resolution simulations (including the RUC). There are a number of factors that most likely contribute to these performance results. We suspect significant variability will be associated with model cloud microphysics settings used. These settings will strongly influence the production of the hydrometeors that then factor into the translation algorithms computation of cloud ceiling heights and surface visibility. An examination of the relationships between model microphysics and ceiling and visibility forecast was beyond the scope of this study. Furthermore, it is likely that variability in hydrometeor production for a given cloud microphysics scheme could be tied to horizontal and vertical grid resolutions. Numerous studies have shown that grid resolutions influence the model vertical velocity forecasts, which in turn will factor into the hydrometeor production. It may be necessary to tune the translation algorithms based on cloud microphysics and grid resolutions. This was also not evaluated in this study, but could be investigated as a means to improve the C&V forecast guidance derived from NWP.

The analysis of spatial variability in the high-resolution NWP forecasts appears to have potential use as a forecast error metric. A preliminary review of time series data depicting mean cloud ceiling forecast error and the standard deviation in the forecast over a 54km x 54km evaluation box centered over the ASOS observation site identified a common signature. Increases in forecast standard deviation typically coincide with deviations in mean forecast errors from values near 0 (Figure 5). Further evaluation will be necessary to quantify this relationship and identify thresholds that could be used as an error metric in an operational C&V system. We recommend refining the analysis of standard deviation and forecast error to better quantify its relationship and potential value. Further evidence of this relationship is found in NWP forecasts with low dynamic range in the forecast values. In these situations the mean ceiling forecast meets or exceeds the TAF, persistence, and NWP point forecast at nearly every forecast interval. NWP ceiling forecast are considerably less accurate during periods of high dynamic range in the forecast, and in these situations the point forecast appears to be slightly better. We recommend using spatial variability (either standard deviation or categorical dynamic range) as an error metric that would be passed along with the NWP C&V forecast to the decision support tool.

5. REFERENCES

Allan, S., R. DeLaura, B. Martin, D. A. Clark, and C. Gross, 2004: Advanced Terminal Weather Products Demonstration in New York. 11th Conference on Aviation, Range, and Aerospace Meteorology, Hyannis, MA. American Meteorological Society, Boston MA.

Benjamin, S.G., G. A. Grell, J. M. Brown, and T. G. Smirnova, 2004a: Mesoscale Weather Prediction with the RUC Hybrid Isentropic Terrain-Following Coordinate Model. *Mon Wea. Rev.*, 131. 473-494.

Benjamin, S. G., D. Devenyi, S. S. Weygandt, K. J. Brundage, J. M. Brown, G. A. Grell, D. Kim, B. E. Schwartz, T. G. Smirnova, T. L. Smith, and

G. S. Manikin, 2004b: An Hourly Assimilation-Forecast Cycle: The RUC. *Mon Wea. Rev.*, 131. 495-518.

Brown, J, 2005: Personal communication. NOAA Forecast System Laboratory, Boulder, CO.

Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A Description of the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). *NCAR Technical Note,* NCAR/TN-398+STR, 117 pp.

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. *Journal of Applied Meteorology.,* 38, 385-404.

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