

U.S. Department of Commerce  
National Oceanic and Atmospheric Administration  
National Weather Service  
National Centers for Environmental Prediction  
5830 University Research Court  
College Park, MD 20740-3818

Office Note: 502

<https://doi.org/10.25923/esw8-5n84>

## Application of a Nonlinear Transformation Function to the Variational Analysis of Visibility and Ceiling Height

Runhua Yang<sup>1,2</sup>, R. James Purser<sup>\*1,2</sup>, Jacob R. Carley<sup>2</sup>, Manuel Pondeca<sup>1,2</sup>,

Yanqiu Zhu<sup>1,2</sup>, and Steven Levine<sup>3</sup>

<sup>1</sup>I.M. Systems Group, Inc., Rockville, Maryland; <sup>2</sup>NOAA/NWS/NCEP Environmental Modeling Center,  
College Park, Maryland; <sup>3</sup>Systems Research Group, College Park, Maryland, *present affiliation:*  
NOAA/NWS/MDL, Silver Spring, Maryland.

April 16, 2020

\*Corresponding author email: jim.purser@noaa.gov

## Abstract

In order to meet the growing requirements from general aviation services and helicopter emergency rescue, the Real-Time Mesoscale Analysis (RTMA) system provides analyses of surface visibility and cloud ceiling height since 2013 and 2016 respectively. Its analysis component uses NCEP's Grid-point Statistical Interpolation (GSI) configured to run in the two-dimensional surface mode. The fundamental assumption justifying the existence of an optimal analysis solution is that the innovation follows a Gaussian distribution. However, visibility and ceiling height fields are characterized by high degrees of discontinuity in both spatial and temporal dimensions and their conditional error statistics vary with states. Directly using these fields in a variational analysis system often undermines the relatively accurate observations in circumstances dominated by severe weather systems, where the first guess is likely to depart far from the observations. Many efforts have been made to transform non-Gaussian variables to have a Gaussian distribution, or close to it. Among the methods, a logarithmic transformation is often used. The logarithmic transformation in visibility and ceiling height improves the convergence in minimizing the cost function and uses more observation data, particularly in the areas occupied by stormy weather systems. However, this method often generates spuriously large analysis increments over the areas of clear weather.

In this study, a nonlinear transformation function (NLTF) is applied to visibility and ceiling height analysis in the RTMA. In this method, a function is derived to map the visibility and ceiling height into a space tending to a more uniform variance. Modulated by a varying parameter in the range of [0-1], the function family includes linear and logarithmic functions at the respective ends of the parameter range. A subjective approach, based on evaluating histograms of the variables, is used to determine the optimal value of the parameter. The statistical errors were estimated based on the statistics of innovation and then adjusted based on the statistics from single observation test and months-long analysis run. An experimental RTMA with the NLTF was conducted for eight months and compared to the control run that uses the previous analysis algorithm. The assessment and evaluation are carried out using a complete set of measures appropriate to categorized visibility and ceiling height according to FAA flight category definitions. The results showed that NLTF improves the visibility and ceiling height analysis consistently. Therefore, NLTF was implemented into the operational RTMA system on 5th December 2018.

## 1. Introduction

Information about frequently-updated surface visibility and cloud ceiling height (hereafter, simply 'ceiling height') at high spatial resolution is important to general aviation services and helicopter emergency rescue. The Real-Time Mesoscale Analysis (RTMA) has been developed at NCEP/NOAA since 2006 to provide gridded surface weather analysis for the atmospheric scientific studies and for public aviation services. The early RTMA product includes gridded analyses of surface pressure, temperature and moisture at 2 meters above the ground level, and wind at 10 meters (Pondeca et al. 2011). In its recent development, the RTMA includes more surface weather variables and areas, currently covering five regions, these being: the contiguous United States (CONUS); Alaska; Hawaii; Puerto Rico; and Guam. To meet the quickly growing requirements from aviation and the emergency rescue services, visibility and ceiling height analyses were added to RTMA in 2013 and 2016 respectively (Pondeca, et al. 2015). Furthermore, RTMA provides 15-minute analyses in the CONUS since the fall of 2018 (Yang et al. 2017). Figure 1 depicts the CONUS, which is the focus of this paper.

The advances in the following four areas make the RTMA successful: (1) the High-Resolution Rapid Refresh (HRRR) with a sophisticated cloud prediction scheme and hydrometeors analysis (Benjamin et al. 2016) from which visibility and ceiling height are derived; (2) the grid-space statistical interpolation system (GSI) developed at NCEP/NOAA (Wu et al. 2002); (3) the existence of a reliable and dense network of conventional observations, plus the recent sub-hourly special reports (SPECIs), which reflect flight-affecting weather systems in near real time; and (4) high performance computing.

The analysis component of the RTMA is the GSI configured to run in two-dimensional surface mode. The fundamental assumption justifying the existing optimal analysis solution is that the innovation collectively obeys a Gaussian distribution. However, both visibility and ceiling height fields are highly discontinuous in both spatial and temporal dimensions; the error statistics depend on states, which ranges from several meters to kilometers for visibility and less than a hundred feet to a thousand feet for ceiling height. Directly using these fields in a variational analysis system often leads to the rejection or undermining of relatively accurate observations in circumstances dominated by severe weather systems, where the first guess likely departs far from the observations. Many efforts have been made to transform non-Gaussian variables to a Gaussian distribution or close to it. Among these methods, a logarithmic transformation is often used. The logarithmic transformation in visibility and ceiling height improves the convergence in minimizing the cost function and the usage of the observations, particularly in the areas occupied by stormy weather systems. However, this method often generates spuriously large analysis increments over the areas with clear weather.

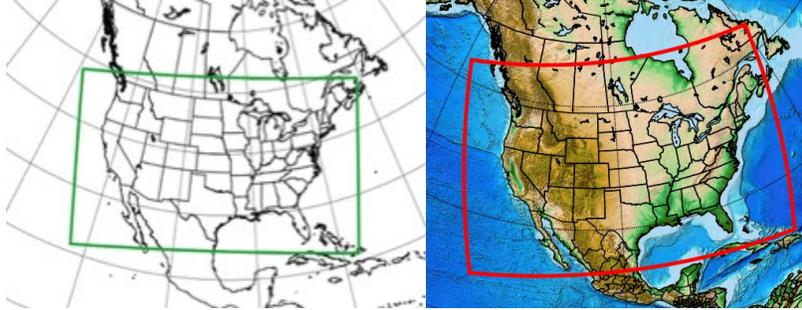
To solve this problem, a nonlinear variable transformation function (NLTF) is derived to transform a measured variable with non-uniform error to one with uniform error. In this paper we document the implementation of the NLTF to visibility and ceiling height analysis in the RTMA and demonstrate the improvements. In section 2, three components of the RTMA are described briefly. In section 3 the non-Gaussian statistical features in visibility and ceiling height are

depicted. Since the FAA flight category definitions are the guidance for aviation, the statistics are expressed in terms of the equivalent category. The previous analysis algorithm for visibility and ceiling height is also reviewed. Section 4 presents the equations of NLTF and the function graphics. It also documents a subjective approach to determine the parameter of the NLTF and a method to estimate error statistics for the transformed visibility and ceiling height. Section 5 is devoted to the assessment and evaluation of the experimental run with NLTF with respect to the control run. The evaluation is based on a complete set of measures, which are appropriate to categorized visibility and ceiling height. In the last section, we summarize the merits of the method and illustrate the remaining questions, and aspects to be improved.

## **2. Real-Time Mesoscale System**

The RTMA consists of the following three components: 1) A downscaling and first guess process leveraging short-term forecasts from the best-available convection-allowing model output, namely the High Resolution Rapid Refresh (HRRR) and Rapid Refresh (RAP) (Benjamin et al. 2016). The first guess for visibility and ceiling analysis is derived from the cloud and hydrometeor analysis of the aforementioned models. Figure 1 shows the areas covered by the HRRR and RTMA CONUS domains. The HRRR is the primary model providing the first guess, whereas the RAP is used to fill in the edges of the RTMA domain not covered by the HRRR; 2) The GSI system configured to run in two-dimensional variational mode. The observations come from a variety of platforms including surface observing systems (METARS, mesonets, buoys, and C-MAN), geostationary satellite cloud products, satellite cloud drift winds and scatterometer winds, and altimeter-derived significant wave heights. For visibility and ceiling height, METAR is the main observation source, which normally reports hourly. However, SPECIs are extremely important, because they are generated in time to report the advent of flight-affecting weather systems; 3) A postprocessing component to convert the first guess and analysis into GRIB2 format, as well as to estimate the analysis uncertainty based on diagnostics that use the Lanczos vectors provided by the GSI solution iterations (Pondeca, et al. 2011).

In the previous version of the RTMA, the visibility and ceiling height are converted into log-distance space while the corresponding observations remain in their natural (linear-distance) form. The implementation requires a linear approximation of the logarithm function to the observation quantities when combining the penalties calculated in logarithmic space with those calculated in the state space. This algorithm introduces nonlinearity into each minimization iteration, though in general the approximation is reasonable. In situations dominated by severe weather systems, where the first guess likely departs far from the observation, if the innovation is calculated in the physical space, i.e., in its natural form, then the innovation is likely to be too large to pass the prescribed quality criterion, so the observation is rejected. In section 5, we will show some examples.



**Figure 1: (Left) HRRR Domain and (right) RTMA-CONUS domain.**

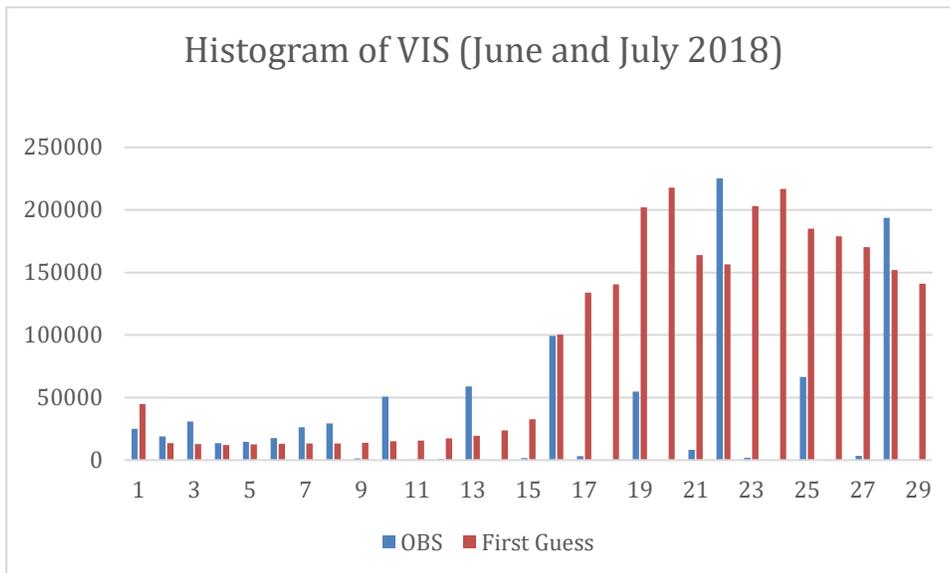
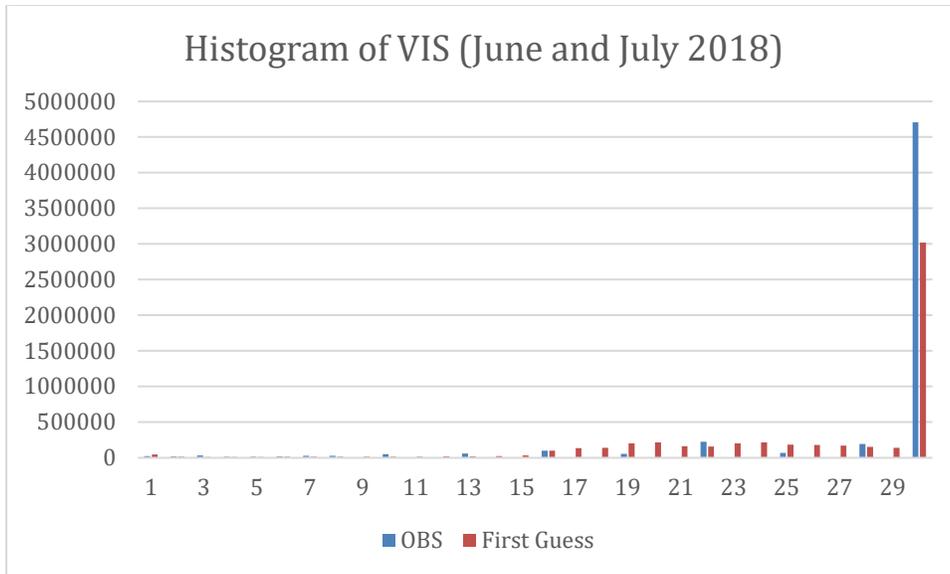
### 3. Statistical features in the observations and the model forecast

In this section, we first demonstrate the statistical features of high discontinuity and non-Gaussian distribution of the observed and forecast visibility and ceiling height in terms of histograms. We then present the distribution of the categorized fields according to the definition of the Federal Aviation Administration (FAA), as listed in Table 1. It is important to point out that the classifications are practically useful and meaningful.

**Table 1: FAA flight category definition for ceiling height (C) and Visibility (V).**

	<b>CEILING (ft)</b>	<b>VISIBILITY (mi)</b>
Low Instrument Flight Rules ( <b>LIFR</b> )	< 500	< 1
Instrument Flight Rules ( <b>IFR</b> )	500 to <1000	1 to <3
Marginal Visual Flight Rules ( <b>MVFR</b> )	1000 to 3000	3 to 5
Visual Flight Rules ( <b>VFR</b> )	> 3000	> 5

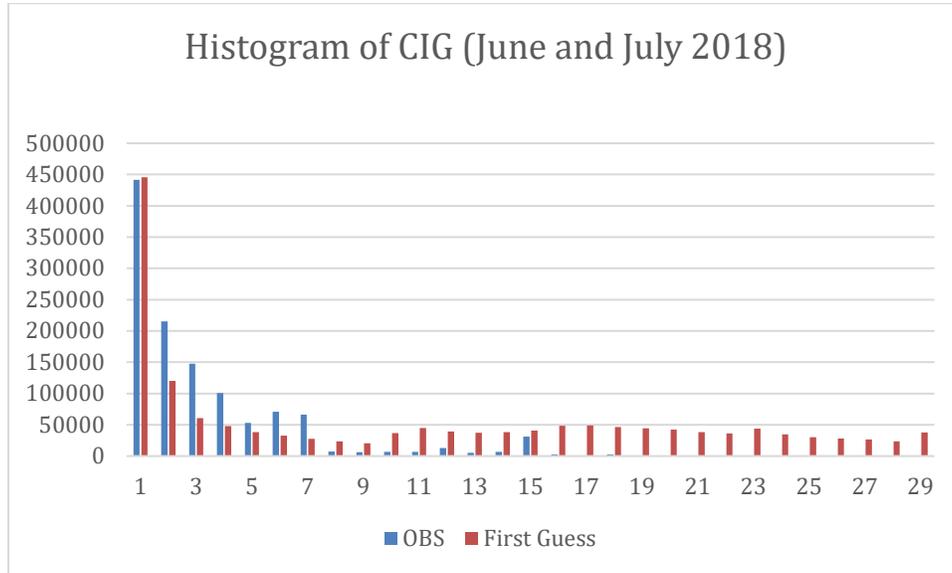
The top panel in Figure 2. shows the histogram computed from the visibility observation (in blue) and the first guess (in red) in June and July 2018. The horizontal axis spans the range up to the RTMA analysis capping value of 16 km, and is uniformly divided into 30 bins numbered 1-30. Notice that there is only one outlier, or mode, at the 30th bin corresponding to the capping value. This outlier is so dominating that values in other bins are essentially invisible in a direct graphical comparison. In order to see the histogram of the first 29 bins, they have to be replotted in the bottom panel of Figure 2. Two features are revealed from replotting: the restricted visibility events occurred rarely, and the event counts increase gradually between bins 16 and 29, corresponding to the range of about 8000 - 15,500 meters for both observation and the first-guess. Note that the values of the first guess spread much more uniformly than those of the observations.



**Figure 2: Top panel: Histograms of visibility (VIS) in all 30 bins spanning the 16 km total range. Bottom panel: the histogram in the first 29 bins, computed with data covering June and July 2018. The blue bars are for observation and the red bars for first-guess.**

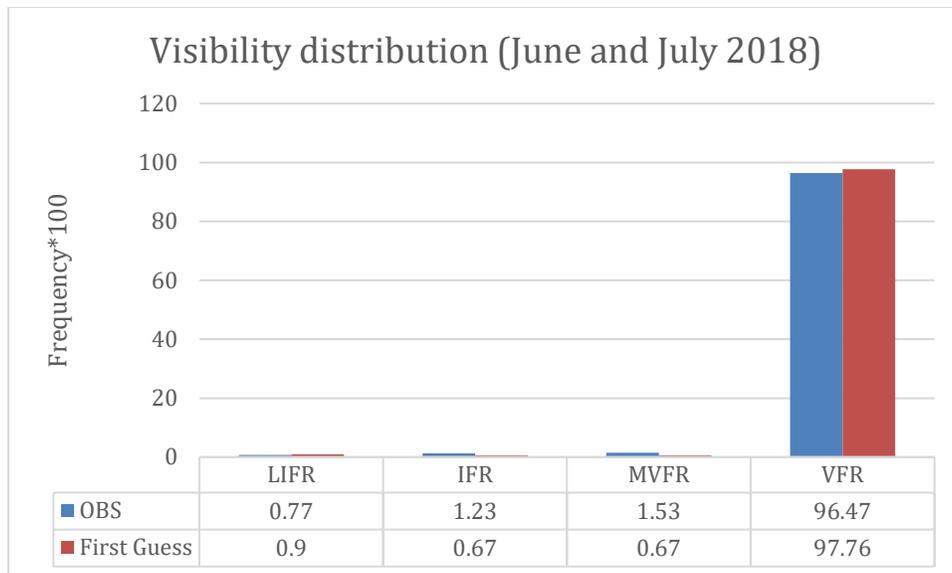
Figure 3 shows the histogram of the ceiling height in the range of the first bin up to the 29th bin. The capping value is 16000 meters as in visibility. Similar to visibility, the data count in the 30th bin is out of proportion relative to the rest of the bins to be plotted. There are three clear modes in the first three bins corresponding to the values less than about 1738 feet, less than about 3479 feet, and less than 5249 feet respectively. The bin range is too large to capture low ceiling height events, which include LIFR, IFR and a portion of the MVFR (referring to the categories values in Table 1). Note that the histogram values of the observed ceiling height, or event occurrences,

are so small within the range of bin 16 to bin 29 that they are barely visible at this vertical scale. However, the histograms of the first guess spread quite evenly across the range, indicating the artifact of the algorithm in HRRR analysis used to derive the ceiling height.

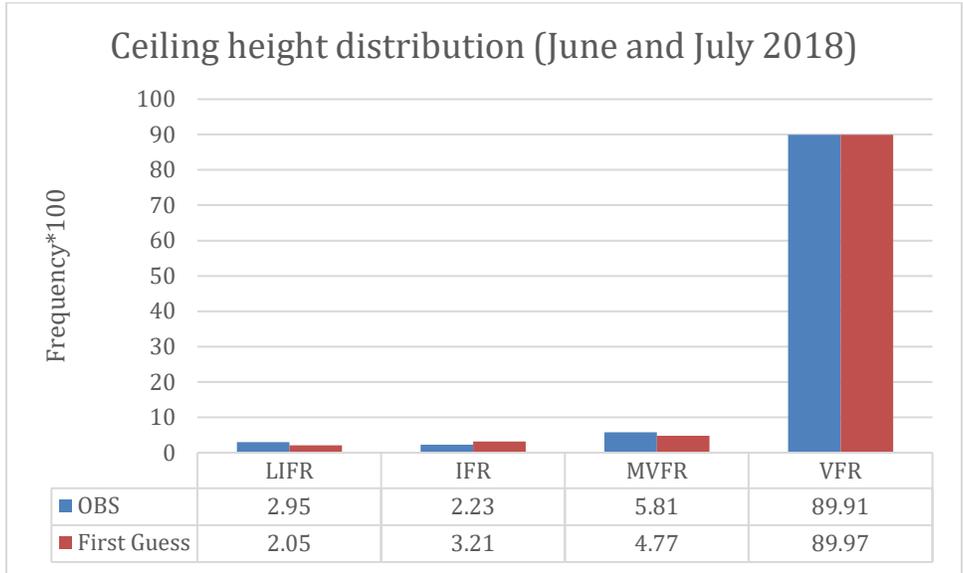


**Figure 3: The histograms of ceiling height (CIG) in the first 29 bins computed with data covering June-July 2018. The blue bars are for observation and red bars for first-guess.**

Figures 4 and 5 show the distribution of categorized visibility and ceiling height for observation (blue bars) and the first-guess (red bars) respectively. The data covers the period between June to July 2018. As expected, weather with poor visibility and ceiling height occurs rarely, while weather with clear visibility and ceiling height occurs often.

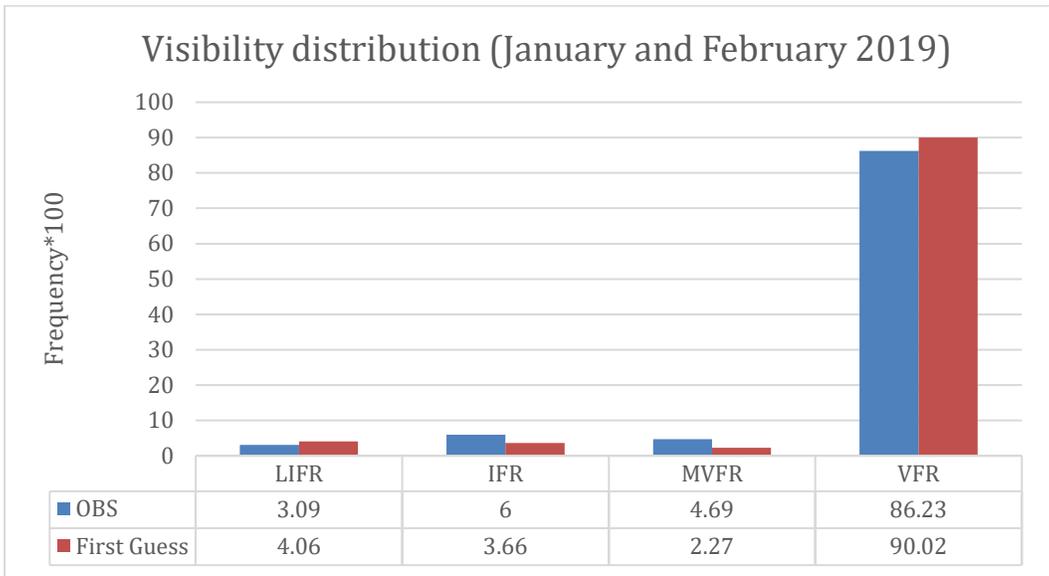


**Figure 4: Distribution of categorized visibility. Data (percent) cover June and July 2018.**

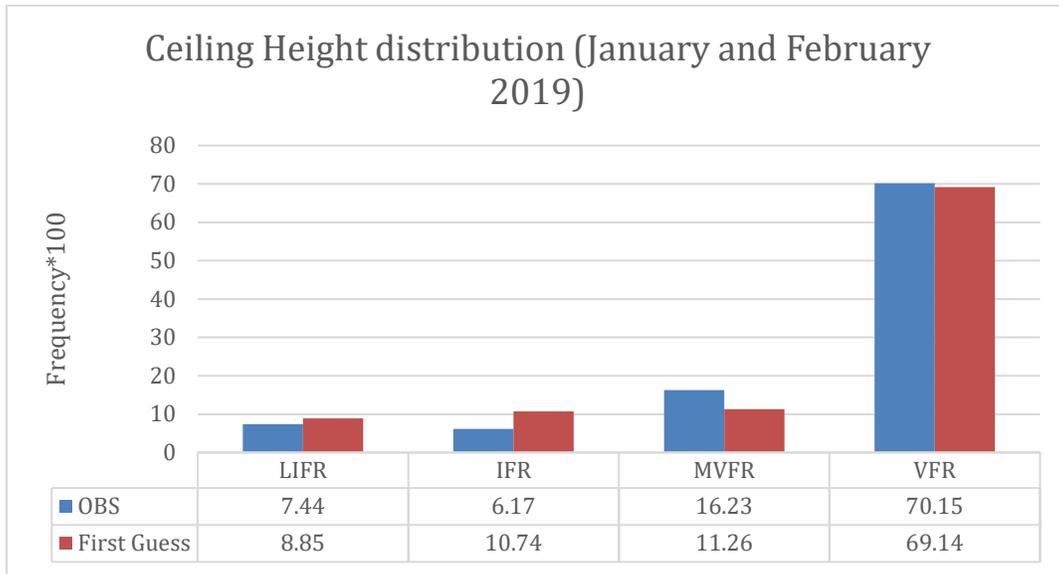


**Figure 5: Distribution of categorized ceiling height. Data (percent) cover June and July 2018.**

The statistics vary from warm months to cool months. To illustrate it, the same computation is conducted with the data of January and February 2019. The results are depicted in Figures 6 to 7. As expected, the values in LIFR and IFR are clearly increased, and the values in VFR remain dominant.



**Figure 6: Distribution of categorized visibility. Data (percent) cover January and February 2019.**



**Figure 7: Distribution of categorized ceiling height. Data (percent) cover January and February 2019.**

Tables 2 and 3 are the multi-level contingency tables of visibility and ceiling height computed with the categorized analysis and the corresponding observations. Data covers the period of June and July 2018. In statistics, a contingency table displays the conditional distribution of a variable given the associated variable's distribution in a matrix format. The table header, 'OBSERVATION' denotes observation, and 'FORECAST' denotes HRRR's forecast. The value in each *i*th/*j*th cell is the total count of the forecast categorized as *i*th group, given the corresponding observation is categorized as *j*th group. We present the value in each cell as a percentage, i.e., a count in each cell is divided by the total observation count in the corresponding category. For example, a value of 29.07 in the intersection of the first row and first column of Table 2 denotes that the forecast catches 29.07% of observed LIFR events. The values in the diagonal cells are called probability of detection, or Hit rate, denoting the correct forecast ratio, which is the count of both observation and forecast falling into the same category, divided by the categorized observed events. Whereas the values in the off-diagonal cells represent the wrong forecast percentage relative to the total count of the categorized observed events, or marginal sum of each column. Obviously, the forecast system has difficulty catching the poor visibility events, as indicated by a value of 48.05 % (the intersection of the first column and the 4th row). These tables also present two features: 1) The forecast model has a high accuracy to forecast the clear weather: 84% for visibility and 97.55% for ceiling height; 2) forecasting visibility is more difficult than forecasting ceiling height, as shown by the much smaller Hit rate values of Table 2 than those of Table 3.

**Table 2: Contingency table for visibility: Percentage values (rounded to two decimal places) are computed with the data of June and July 2018. The forecast is produced by the HRRR.**

		OBSERVATION			
FORECAST		LIFR	IFR	MVFR	VFR
	LIFR	29.07	12.10	7.63	0.58
	IFR	11.49	9.48	8.26	0.95
	MVFR	11.39	13.11	12.21	1.64
	VFR	48.05	65.31	71.90	96.84

**Table 3: Contingency table for ceiling height: Percentage values are computed with the data of June and July 2018. The forecast is produced by the HRRR.**

		OBSERVATION			
FORECAST		LIFR	IFR	MVFR	VFR
	LIFR	57.82	20.60	4.84	0.30
	IFR	17.39	44.98	23.69	0.48
	MVFR	6.76	16.70	36.97	1.67
	VFR	18.03	17.73	34.50	97.55

To demonstrate the variation of those statistics between warm months and cool months, the same computation is conducted with the data of January and February 2019. Tables 4 and 5 show the resulting statistics. It is profoundly significant that the Hit rate are increased while the false alarm forecast ratio is decreased, particularly in the categories LIFR and IFR. Though it is common for the skill of the forecast and of the analysis to be clearly better in winter, the tabulated improvements seem too big to be explained by these ordinary seasonal effects; rather they seem to be a reflection of other factors. We speculate that the improvement in HRRR forecasts starting from late June 2018 contributed to this improvement to some degree.

**Table 4: Contingency table for visibility: Percentage values are computed with the data of January and February 2019. The forecast is produced by the HRRR.**

		OBSERVATION			
FORECAST		LIFR	IFR	MVFR	VFR
	LIFR	50.56	22.02	9.51	0.85
	IFR	14.29	21.15	13.43	1.53
	MVFR	5.90	9.06	9.03	1.30
	VFR	29.25	47.78	68.03	96.33

**Table 5: Contingency table for ceiling height: Percentage values are computed with data of January and February 2019.**

		OBSERVATION			
FORECAST		LIFR	IFR	MVFR	VFR
	LIFR	63.92	23.47	11.85	1.03
	IFR	27.17	54.63	27.89	1.16
	MVFR	6.26	16.47	40.15	4.65
	VFR	2.65	5.43	20.10	93.16

#### **4. Nonlinear Transformation Function and parameters estimation**

An inherent property of the visibility and ceiling variables which makes them problematic to analyze is their typically very large dynamical range. For variables of this type it is often advantageous to apply a nonlinear transformation to each variable before performing the analysis. Primarily, the motivation for doing this is to make the characteristic error variances in the background values more uniform over the range of values. Additionally, it is sometimes possible to make the distributions of these errors appear more Gaussian. While the logarithmic

transformation is one such widely used function, we have found that, for visibility and ceiling, it seems to overcompensate for the nonuniformity of the distributions of the original variables. Consequently, we seek a generalization that allows us to choose a nonlinear transformation that is, in some definite sense, intermediate between the linear and logarithmic functions.

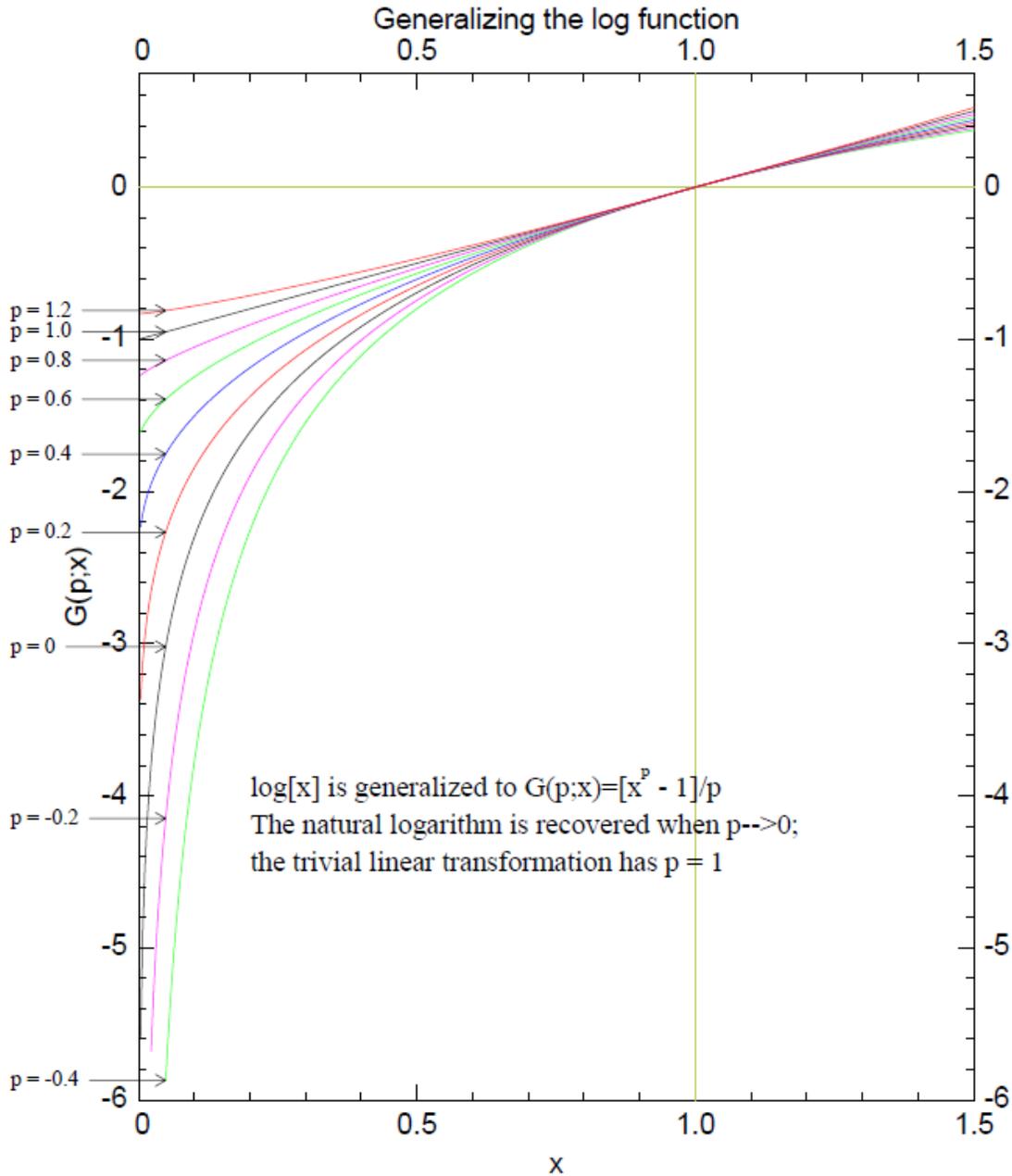
#### 4.1 Nonlinear Transformation Function

The derivation of the general nonlinear transformation function is described in detail in Appendix A. The transformation function transforms a measured variable with non-uniform error into a new variable tending to render the error of that function uniform.

The formula is:

$$G(p; x) = [x^p - 1]/p$$

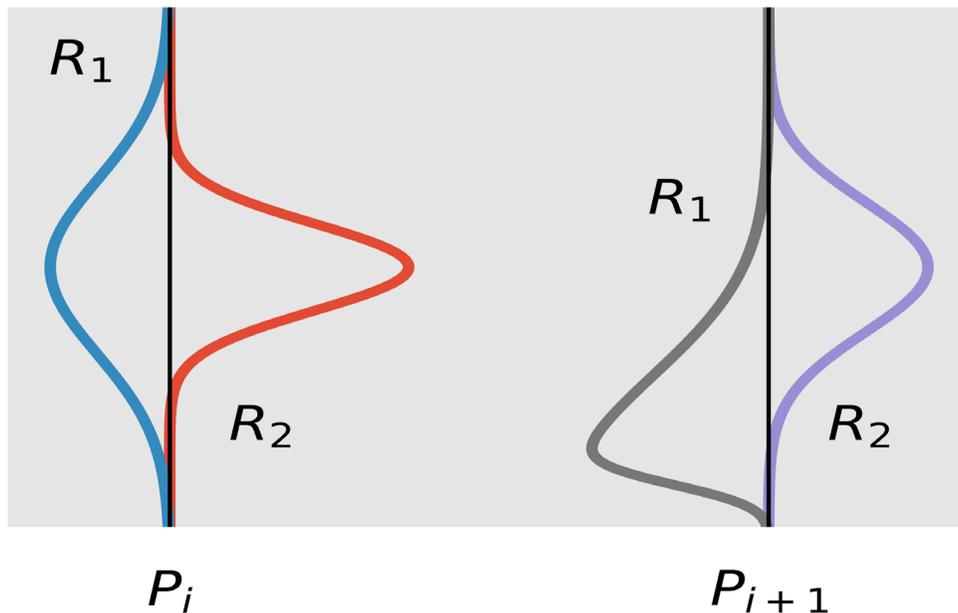
Here,  $x$  is the variable to be transformed and  $p$  is a real constant. The transformation converts  $x$ , which is not a Gaussian variable into the space of  $G(p; x)$ . The parameter  $p$  modulates the function: when  $p \rightarrow 0$ , it is the natural logarithm, whereas when  $p = 1$ , it is a linear function. Figure 8 shows the function family with several  $p$  values:



**Figure 8: Family of general nonlinear transformation function with different  $p$ .**

Our process starts to apply  $G(p;x)$  to both observations and the first guess, and computes their differences, i.e., the innovations, given a  $p$  in the range [0-1] with an interval of 0.1. Knowing that the weather systems associated with aviation-affecting weather are totally different from those associated with clear weather, we classify the innovation into two regimes,  $R_1$  with small values, and  $R_2$  with big values. The threshold is based on the upper value of the Marginal Visual Flight Rules plus a fluctuation range, 10,000 meters for visibility, and 1143 meters for ceiling height.

The essential question is whether there is an optimal  $p$ . We use a subjective detection by visually examining the variations of histogram shape: if the shapes of histograms of  $R_1$  and  $R_2$  change clearly with two adjacent values, a so-called optimal  $p$  exists, which lies within the range of these values. A schematic demonstration is given in Figure 9: given a value of  $p_1$ , the left panel shows a fat shape of the histogram for  $R_1$  and a narrow shape for  $R_2$ ; in the right panel, the fat shape for  $R_1$  becomes narrower, and the shape of  $R_2$  becomes broad when  $p_1$  is replaced by  $p_2$  in the sequence. We are wondering whether this discriminating method is analogous to the way to find a maximum or minimum value of a curve function, in which the derivative sign (considered as a metric), is reversed from one regime to another regime of the function.



**Figure 9: A schematic figure on determination of whether an optimal  $p$  exists. For explanation, see text.**

With real data, the histogram shapes are not as distinctive as in Figure 9. Figures 10-11 are the histograms computed from the visibility innovation data of June and July 2018. We have examined the histograms generated with a varying  $p$  in the range  $[0,1]$  with a 0.1 interval.

We notice the histograms with big  $p$  values are similar to each other, and that the significant shape changes seem to occur when  $p$  is less than 0.3. To 'zoom in' within this range, three histograms are presented with  $p \leq 0.2$  below. Figure 10 shows the histograms for visibility, in  $R_1$ , with 6 testing values of  $p$  labelled at the top of each panel. Figure 11 is similar to Figure 10 but for the region  $R_2$ . By visual comparison between Figure 10 and Figure 11, the major change in the shape seems to occur between the values of 0.3 to 0.2.

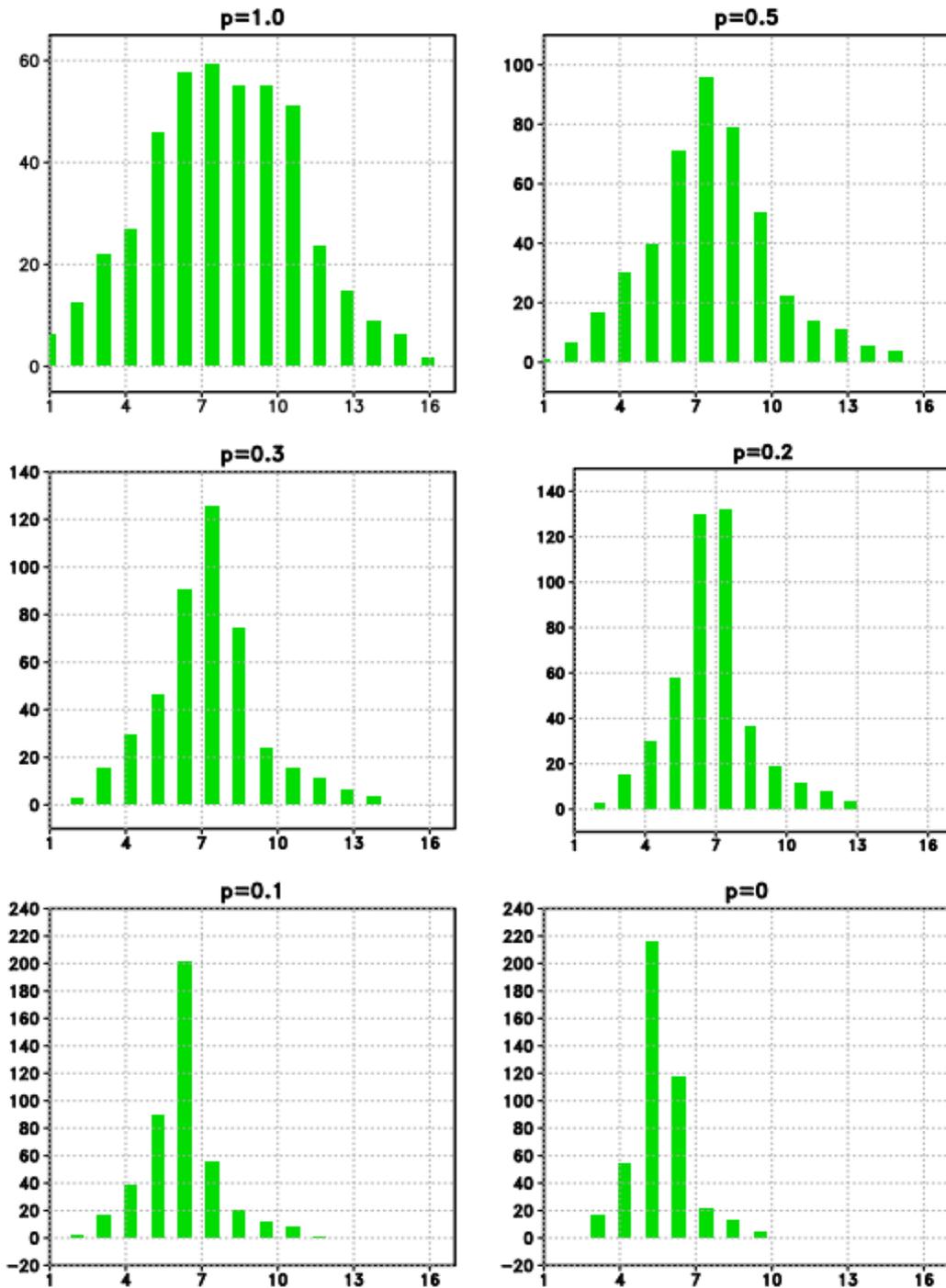


Figure 10: The histograms derived with the visibility in  $R_1$  with 6 testing values of  $p$  labelled at the top of each panel. The  $X$ -axis is for bin numbers, and the  $Y$ -axis for the count of occurring events within a bin, scaled by 1000. The criterion discriminating the two regimes is 10,000 meters.

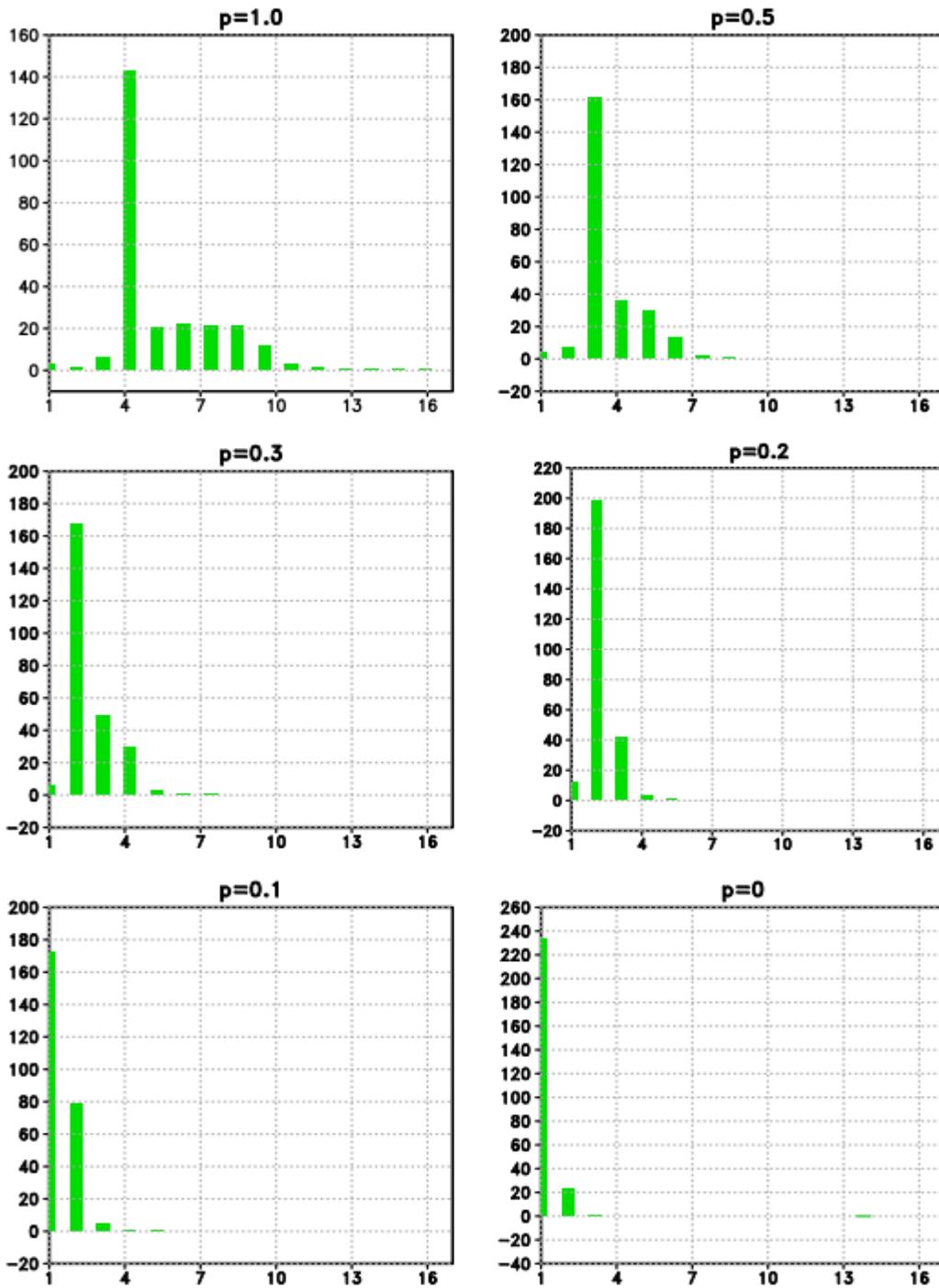


Figure 11: The histograms derived with the visibility in  $R_2$  with 6 testing values of  $p$  labelled at the top of each panel. The X-axis is for bin numbers, and the Y-axis for the count of occurring events within a bin, scaled by 20,000. The criterion discriminating the two regimes is 1000 meters.

Figures 12-13 are the histograms for ceiling height. The shapes of histograms in  $R_1$  are remarkably different from that of visibility by possessing two distinctive modes with  $p \leq 0.5$ . The mode positions gradually shift from the high end to the low end in  $x$ -axis. The major change in the shape seems to occur between the values of 0.3 to 0.2 for visibility, and 0.2 to 0.1 for cloud ceiling.

The parameter used in the final experiment is  $p=0.2$  for visibility and  $p=0.1$  for cloud ceiling, which were estimated based on the limited available data generated by the early study. It is important to point out that, as the experiment runs continued, we collected months-long data, such as the data of June and July 2018, as well as a couple of small data sets. We have computed the same statistics with those data sets, and found that the determined  $p$  fluctuated slightly with data samples, and noticed that the major change always likely occurs along with a small  $p$  value.

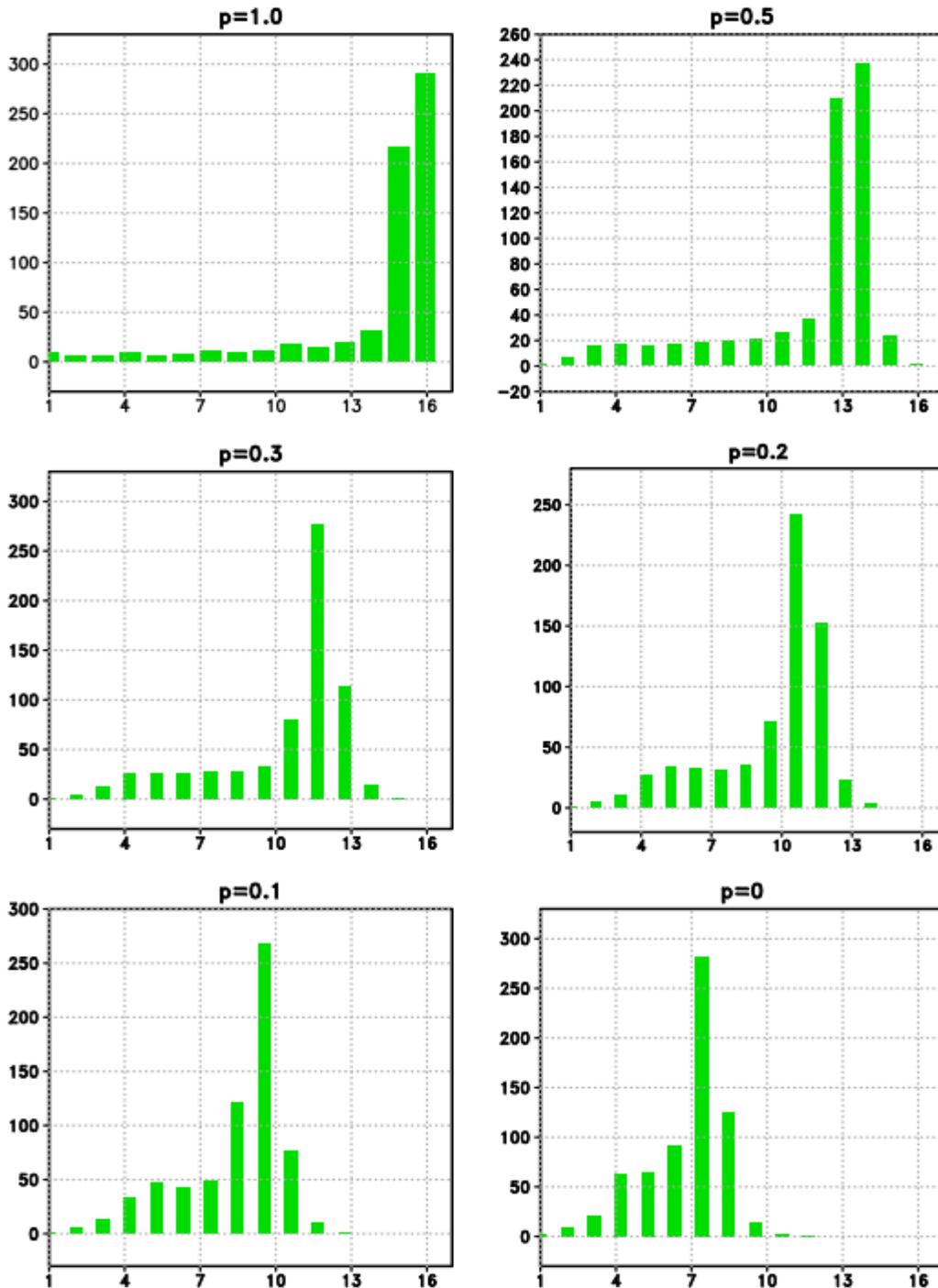


Figure 12: The histograms of ceiling height in  $R_1$  with 6 testing values of  $p$  labelled at the top of each panel. The X-axis is for bin numbers, and the Y-axis for the count of occurring events within a bin, scaled by 1000. The criterion discriminating the two regimes is 3000 feet.

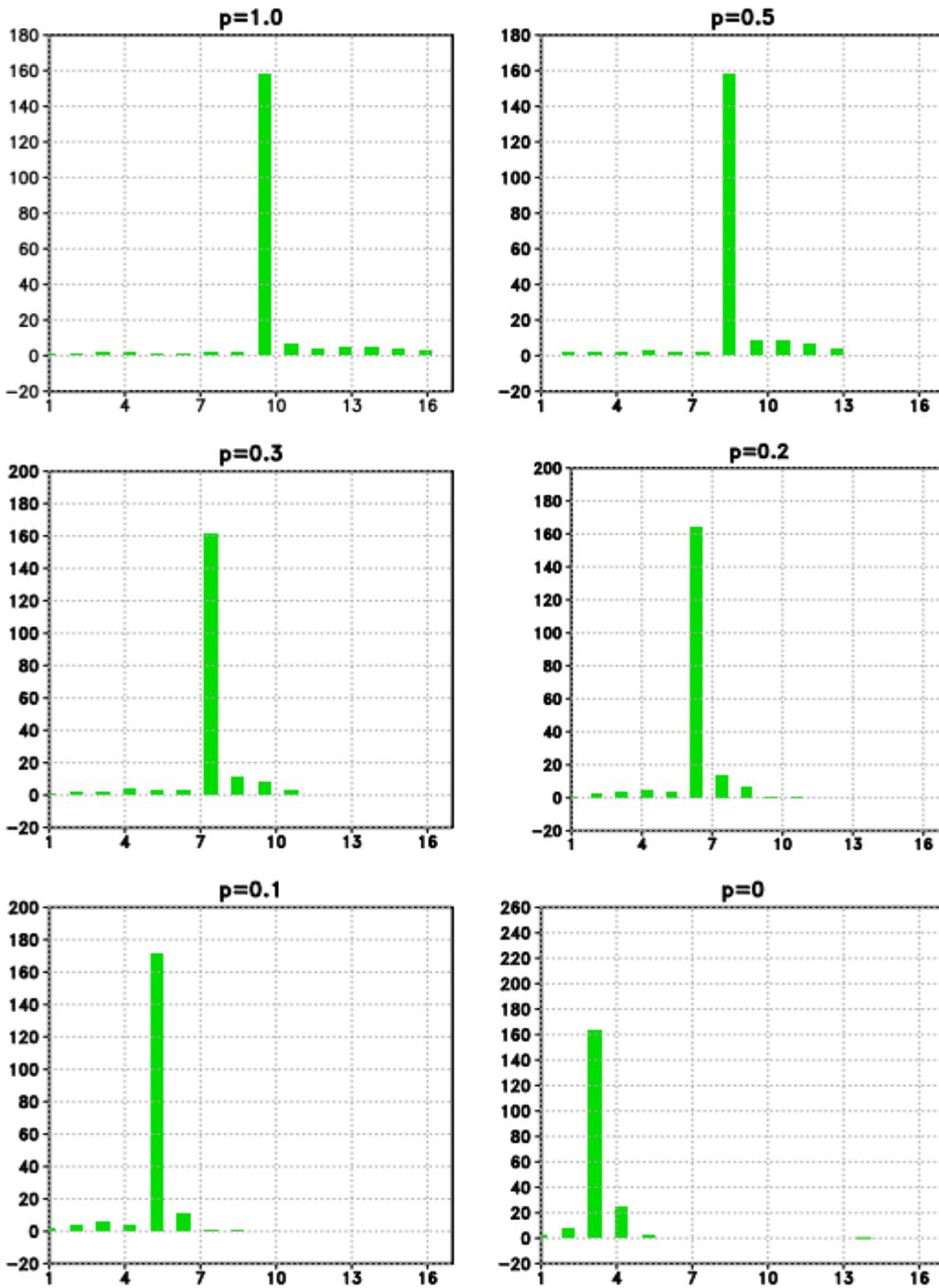


Figure 13: The histograms of ceiling height in  $R_2$  with 6 testing values of  $p$  labelled at the top of each panel. The X-axis is for bin numbers, and the Y-axis for the count of occurring events within a bin, scaled by 20,000. The criterion discriminating the two regimes is 1143 meters.

## 4.2 Estimated error statistics

The estimated statistical errors in the observation and the first guess were obtained by analyzing the data sets of the ‘innovation’, or observation minus forecast (OMF), and the difference between the observation and the analysis (OMA), generated by the RTMA analysis procedures for several time periods. Assuming the statistical error of each observation is independent of that of the forecast first guess, the variance of OMF is partitioned into the variance of each part. Though the estimation is subjective, it still provides a reasonable range to the statistical error. To refine the estimate, single observation tests were used to check and adjust these statistics, and the real-time experimental runs spanning several months were also used to adjust the statistics based on the overall analysis that fits to the observations.

## 5. Examining and evaluating the impact of NLTF

We examine the results of RTMA-RU runs performed since March 2018. The data used in the following computation covers the period of October and November 2018. Because the experimental runs were performed in the computer platforms for developers, there were missing dates in the experimental run when the computers were not accessible. Those missing dates are excluded from the control runs to form the matched data set.

The assessments consist of the comparisons between the experiment and the control runs, throughout evaluating the overall statistics, visual inspection of the 2D-fields, and the categorical statistics. We want to emphasize the use of categorical statistics, because the general statistics for continuous variables are not suitable measures for visibility and ceiling height.

The evaluation of the hourly RTMA experimental results is not presented here, because the experiment also uses a new data selection scheme in addition to using NLTF. Therefore it is hard to purely discern the impact of the NLTF on the improvements of hourly RTMA performance. However, we include those results in the Appendix B2 for a complete document for this upgrade of the RTMA system.

### 5.1 15-minute averaged critical success index and false alarm ratio

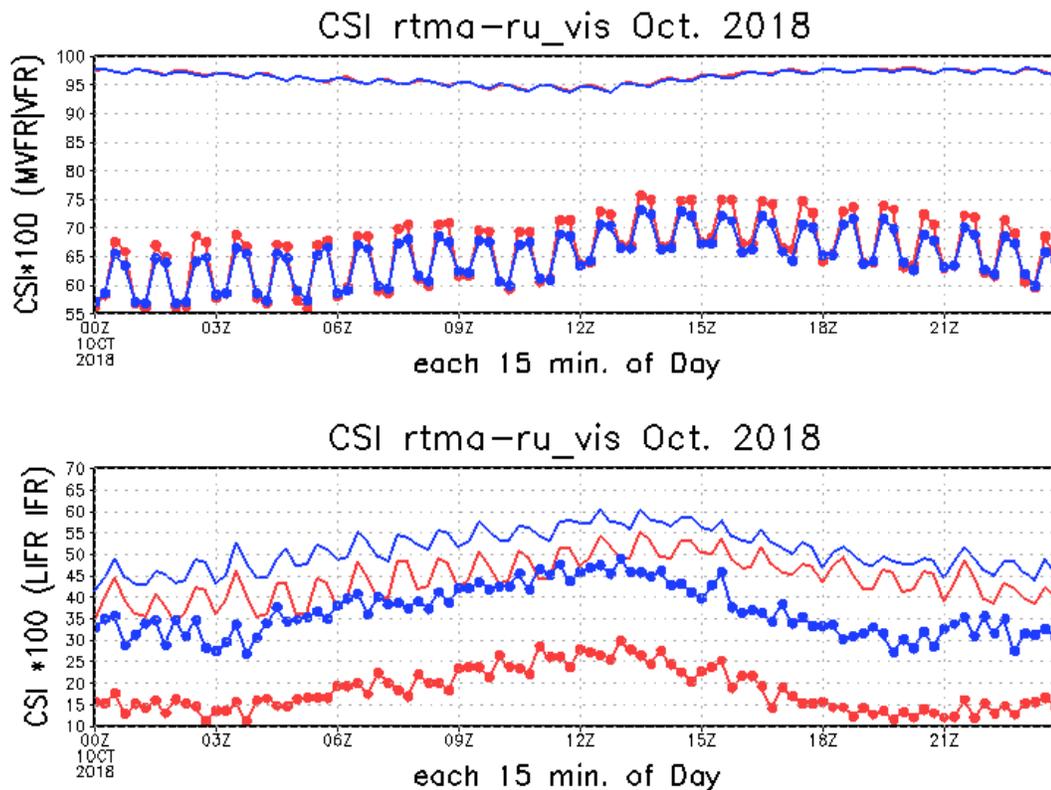
Critical success index (CSI) and false alarm ratio (FAR) are used to measure the categorized analysis fields. The CSI and FAR are computed according to the classification shown in Table 6.

**Table 6: A schematic table lists the components in a two-level contingency table.**

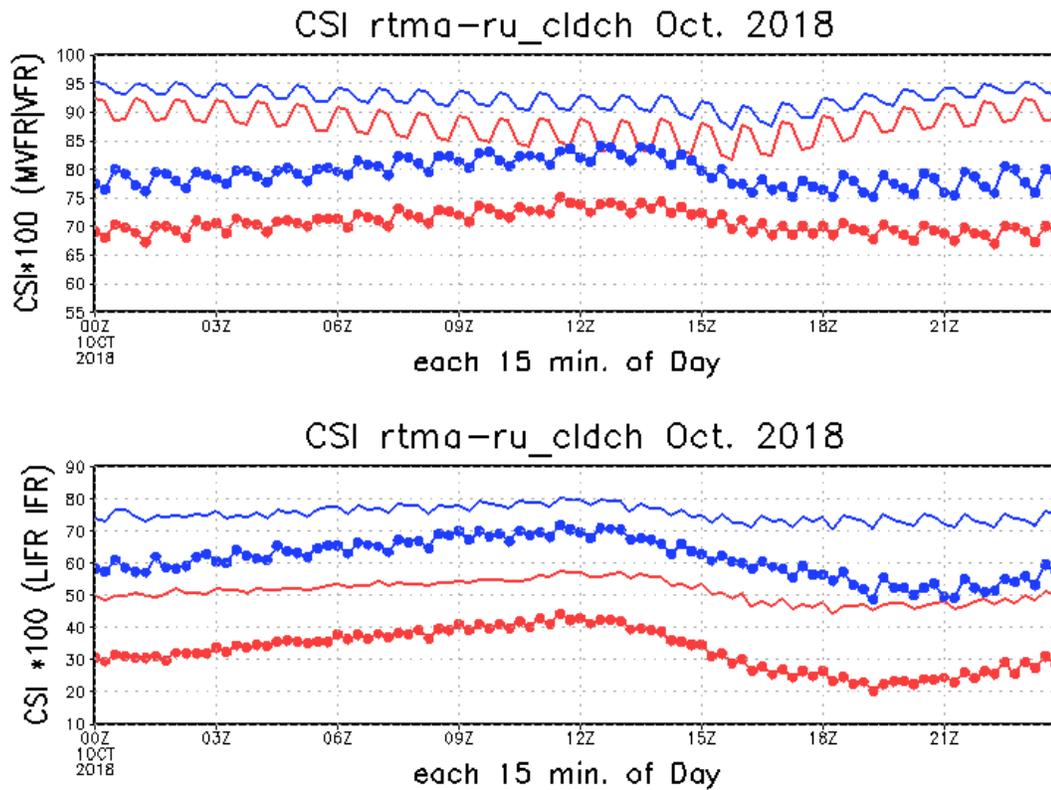
		OBSERVED		
		YES	NO	TOTAL
FORECAST	YES	a	b	a + b
	NO	c	d	c + d
	TOTAL	a + c	b + d	n

CSI =  $a/(a + b + c)$ , represents the analysis accuracy, whereas FAR =  $b/(a + b)$ , represents the count of wrong classification versus the total counts of classification. Note the denominator is the counts of forecasting YES.

Figure 14 shows the Critical success index (CSI) of visibility analysis aggregated by each 15 minutes for October 2018. The red lines are for the control runs and the blue for the experimental runs. In the bottom panel, the dotted lines are for visibility less than 1 mile, and the solid lines are for visibility less than 3 miles. In the top panel, dotted lines are for visibility less than 5 miles and solid lines for larger than 5 miles. Figure 15 is similar to Figure 14 but for cloud ceiling. Data are classified using the upper values of four categories. It is obvious that the CSIs of the experiments are consistently higher than those of the control, except in the VFR category.



**Figure 14: Critical success index of visibility analysis aggregated over each 15 minutes in October 2018. The red lines are for the control runs and the blue for the experimental runs. In the bottom panel, the dotted lines are for visibility less than 1 mile, and the solid lines are for visibility less than 3 miles. In the top panel, dotted lines and solid lines are for visibility less than 5 miles and larger than 5 miles respectively.**



**Figure 15: Critical success index of ceiling height analysis aggregated over each 15 minutes in October 2018. The red lines are for the control runs and the blue for the experimental runs. In the bottom panel, the dotted lines are for ceiling height less than 500 feet; the solid lines are for ceiling height less than 1000 feet. In the top panel, dotted lines and solid lines are for ceiling height less than 3000 feet and larger than 3000 feet respectively.**

As a counterpart of the CSI, the false alarm ratio (FAR) of visibility and ceiling height of the experimental runs, computed with the same data set, is consistently smaller than that of the control run, as shown in Figures 16 and 17.

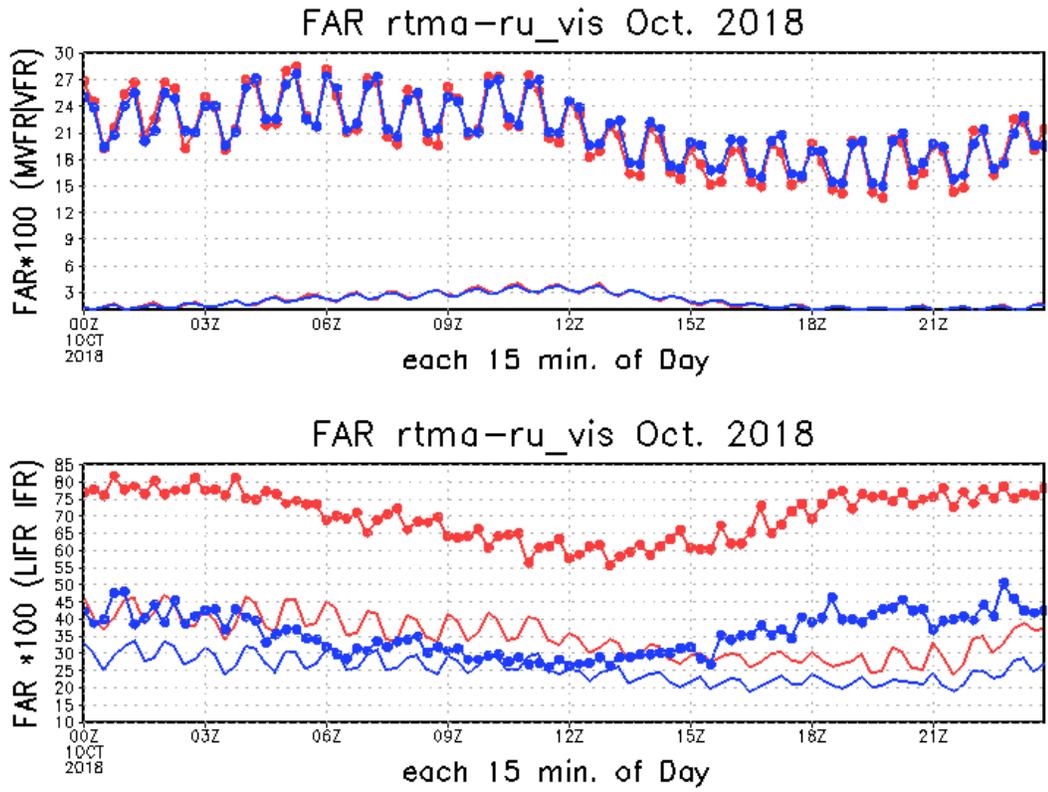
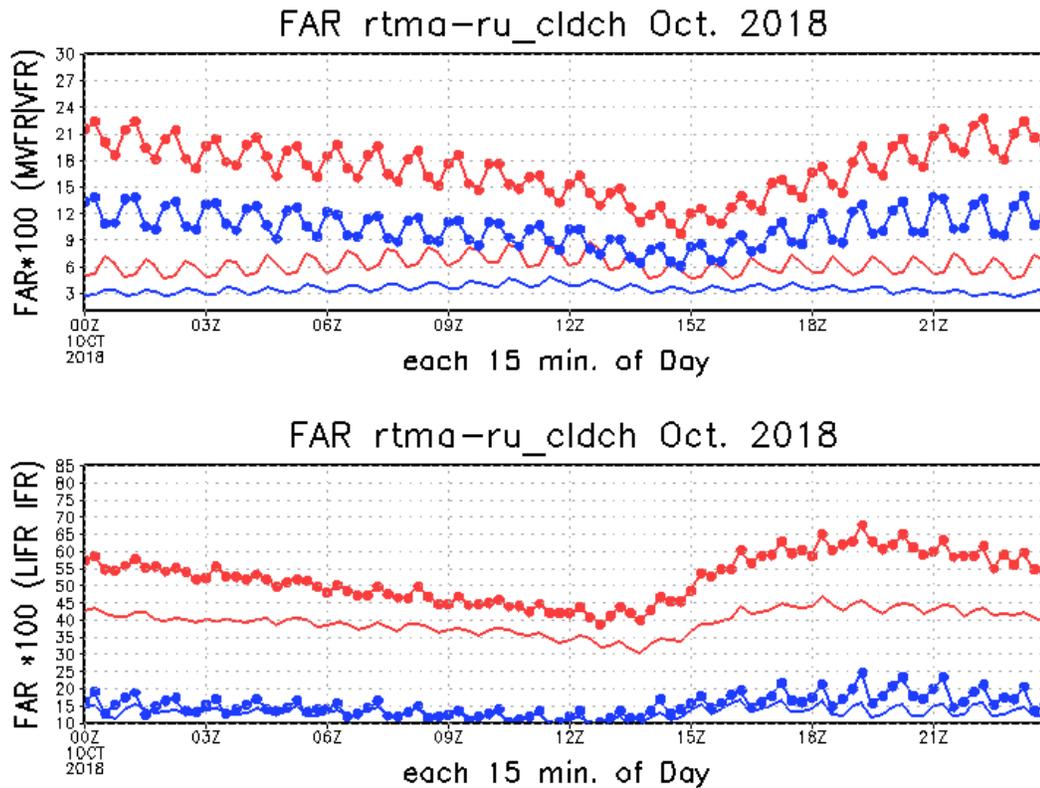


Figure 16: Similar to Figure 14 for visibility, but the statistic is the false alarm ratio.

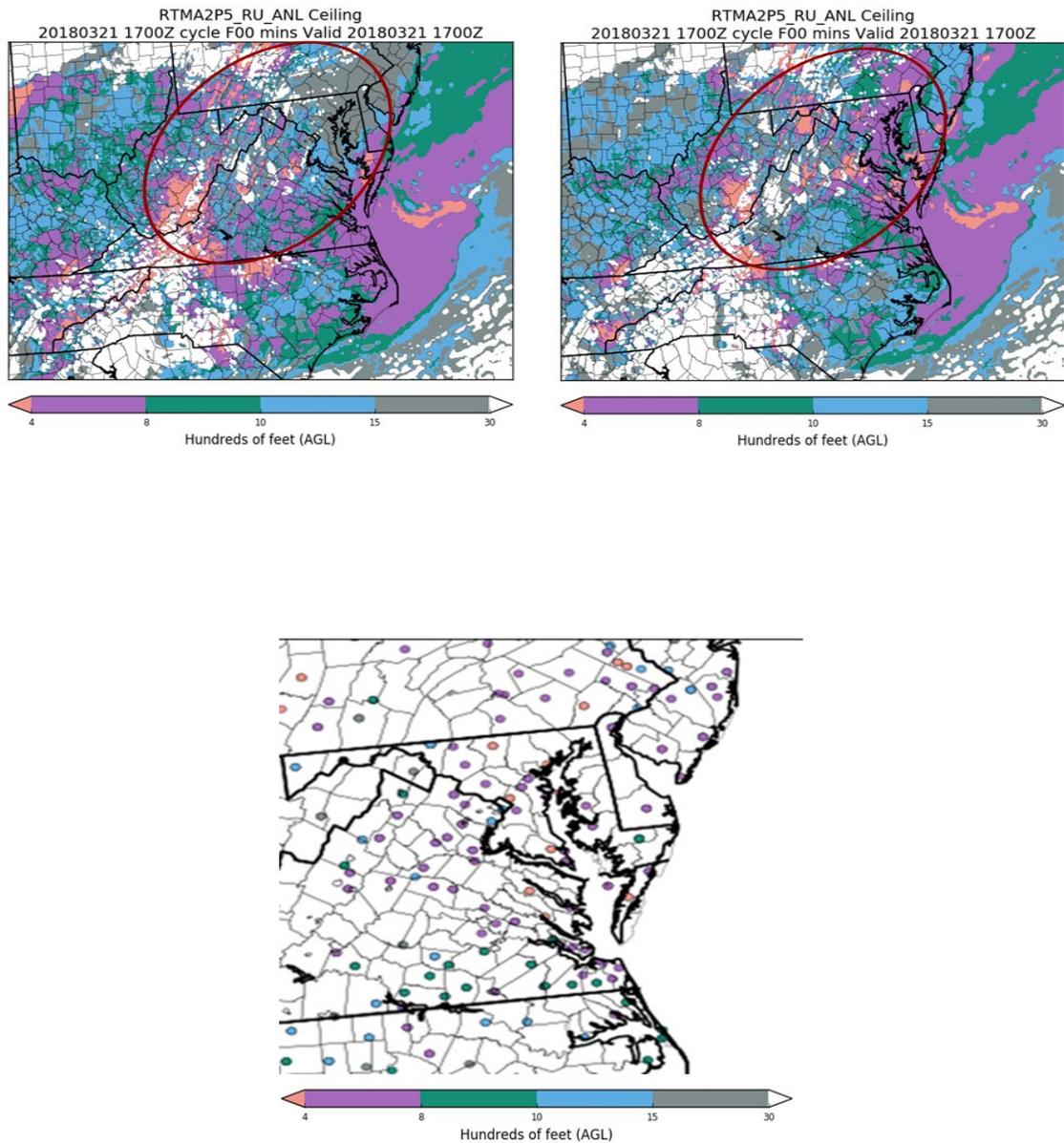


**Figure 17: Similar to Figure 15 for ceiling height, but the statistic is the false alarm ratio.**

The same computations are computed with the data of November 2018. The improvements remain in the experimental runs: higher CSI and lower FAR relative to the control runs (not shown).

## 5.2 Case study

Figure 18 shows a comparison of ceiling height analysis between the control (left panel) and the experiment (right panel) for a case, 1700Z 21 March 2018. The bottom panel shows the corresponding observations, and the dot color denotes the value as in the color scales. Many stations reported the ceiling heights with 400--800 feet, especially in the Mid-Atlantic Region. But the analyzed values of the control (left panel) mainly represent the high values in the first guess, and did not reflect some observations. With the new algorithm the analyzed ceiling height values (right panel) are much closer to the observations.



**Figure 18: Comparison in ceiling height analysis between the control (top left panel) and the experiment (top right panel) for the case 1700Z 21 March, 2018. Bottom panel shows the observed ceiling height, and values are denoted by the color as in the scale, in feet.**

Table 7 shows the comparison between the observed and the analyzed ceiling height at six selected airports in the Central Eastern area for the same case as the above. *Old Analysis* denotes control, and the *New Analysis* the experiment. The differences between the observed

and forecast background are quite large, the old scheme did not reflect much of the observation, whereas the experimental run reflects the observed poor ceiling height situations effectively in the NLTF space.

**Table 7: Ceiling height (in feet) at six airports in the Central Eastern Area on 1700Z 21 March 2018 as observed by the ASOS (“Observed”), Forecast by the HRRR (“Background”) or analyzed by the RTMA (“Old Analysis” and “New Analysis”). “Old Analysis” denotes the control (operational RTMA) and “New Analysis” denotes the experimental RTMA with NLTF.**

Airport	Observed	Background	Old Analysis	New Analysis
KLNS	700	1890	1757	597
KNAK	800	1879	1622	487
KILG	500	1931	1791	609
KMQS	100	1911	1757	311
KPHL	1200	1900	1766	787
KBWI	700	1900	1650	633

### 5.3 Hit rate and false alarm ratio from multi-level contingency tables

We computed the contingency tables for both visibility and ceiling height categorized according to FAA categories. The values are computed based on the analysis products generated by the control and the experimental runs. Tables 8-9 lists the Hit rate in percent for visibility and ceiling height, respectively, for October 2018; similarly, Tables 10-11 are for the results computed data of November 2018. The values in brackets denote false alarm ratio. The values in the last row show the differences between the experiment and the control, where the bold red values indicate that the control run is better than the experimental run.

**Table 8: Hit rate and false alarm ratio ( $\times 100$ ) of visibility computed from the control and the experiment runs respectively. Data cover the period of October 2018. False alarm ratio is annotated in brackets.**

	LIFR	IFR	MVFR	VFR
Control	48.98 [32.50]	41.13 [54.97]	43.13 [64.93]	98.10 [1.84]
Experiment	49.02 [32.26]	41.15 [54.94]	43.10 [64.97]	98.10 [1.84]
Exp – Control	0.04 [-0.24]	0.02 [-0.03]	0.03 [ <b>0.04</b> ]	0.00 [0.00]

**Table 9: Hit rate and false alarm ratio ( $\times 100$ ) of ceiling height computed from the control and the experiment runs respectively. Data cover the period of October 2018. False alarm ratio is annotated in brackets.**

	LIFR	IFR	MVFR	VFR
Control	70.60 [14.67]	66.39 [40.73]	66.92 [31.07]	96.63 [4.20]
Experiment	70.69 [14.27]	66.52 [40.41]	67.16 [30.94]	96.66 [4.18]
Exp - Control	0.09 [-0.40]	0.13 [-0.32]	0.24 [-0.13]	0.03 [-0.02]

**Table 10: Hit rate and false alarm ratio ( $\times 100$ ) of visibility computed from the control and the experiment runs respectively. Data cover the period of November 2018. False alarm ratio is annotated in brackets.**

	LIFR	IFR	MVFR	VFR
Control	56.34 [30.81]	50.77 [46.19]	47.23 [61.06]	97.64 [2.06]
Experiment	56.39 [30.62]	50.79 [46.16]	47.18 [61.11]	97.64 [2.06]
Exp - Control	0.05 [-0.19]	0.02 [-0.03]	<b>-0.05 [0.05]</b>	0.0 [0.0]

**Table 11: Hit rate and false alarm ratio ( $\times 100$ ) of ceiling height computed from the control and the experiment runs respectively. Data cover the period of November 2018. False alarm ratio is annotated in brackets.**

	LIFR	IFR	MVFR	VFR
Control	72.66 [16.71]	65.37 [44.31]	65.52 [27.68]	95.38 [6.58]
Experiment	72.77 [16.18]	65.50 [43.95]	65.81 [27.57]	95.41 [6.56]
Exp - Control	0.11 [-0.53]	0.13 [-0.36]	0.29 [-0.11]	0.03 [-0.02]

In general, the differences are small, but the experimental runs perform consistently better than the control runs. This is particularly noticeable in ceiling height in LIFR and IFR for both October and November 2018. Since the traditional significance test method is not suitable for visibility and ceiling height, we do not do examine the significance. Though the results between the experiment and control are similar, the dataset these categorical values were evaluated with have several million samples. It is doubtful that these differences are due to noise in the dataset. In addition, even a 0.1% improvement in visibility hit rate would mean at least 73 more LIFR events caught, or 144 more for IFR events, given the total number of observed LIFR is about 73,000, or about 144,000 for IFR in October 2018. Lastly, the data sample in this study might not be diverse or

large enough to draw conclusions in these categories considering that the LIFR/IFR are rare events compared to the MVFR and VFR (referring to Figures 4-7).

It is important to mention the contradictory features between the statistics of Tables 8-11 and the Table 1 of a previous article (Yang et al. 2019). Tables 8-11 show very small differences between the control and the experimental runs, whereas the Table 1 of Yang et al. (2019) shows profound improvements in the experimental run. We speculate that the differences are due to the different data sample sizes used for Tables 8-11 and the Table 1 of Yang et al.; for the latter, though the total data samples were in the thousands, the time period only covered three days. We re-computed the Hit rate and the FAR with the same three-day data. The Hit rate values in the LIFR and IFR are the same as in that Table 1. The tiny changes in the MVFR and VFR are due to a 'bug' in the previous computation: the computation for the MVFR category did not count the visibility equal to 3 miles or equal to 5 miles. We also found that the FAR in that Table 1 was the false alarm ratio not expressed as a percentage. The false alarm ratio denotes the probability of false detection. The re-computed hit rate and false alarm ratio, after the error was corrected, are now listed in Table 12 below.

**Table 12: Hit rate and false alarm ratio ( $\times 100$ ) computed from observed and analyzed visibility generated by the control and the experiment runs. Data covers the period of 31 March 2018 to 03 April 2018. The false alarm ratio is annotated in brackets.**

	LIFR	IFR	MVFR	VFR
Control	48.85 [55.55]	50.28 [42.21]	49.80 [60.95]	98.40 [1.15]
Experiment	71.99 [23.44]	70.04 [29.92]	61.86 [51.99]	98.61[0.61]

## 6. Summary and discussion

In this office note, we document the implementation of a nonlinear transformation function to surface visibility and ceiling height analysis in RTMA. For assessments, we compared the categorical statistics, visual inspection of the 2D-fields, and multi-level contingency tables between the control and the experimental runs. The comparison shows the consistent improvement obtained by the experimental runs, particularly in the situations with bad weather events.

As explained in the introduction, ceiling height and visibility are characterized by high degrees of discontinuity in both spatial and temporal dimensions; therefore, their direct assimilation in a variational system does not yield realistic results. Our study suggests that the NLTF transformation mitigates the problem. We are aware of another application of NLTF: Chen et al. (2020) have applied the NLTF directly to radar reflectivity data assimilation within the GSI En3DVar framework, by transforming the hydrometeor mixing ratios into new control variables,

named CVpq. The performance of assimilating radar data using CVpq is compared with those using hydrometeor mixing ratios as control variables and those using logarithmic hydrometeor mixing ratios as control variables for five convective storm cases in spring of 2017. Their study shows that CVpq with  $p=0.4$  gives the best performance in terms of 0-3 hour forecasts of composite reflectivity and updraft helicity tracks.

We recognize some limitations in our study, which are:

- There is uncertainty in estimating the  $p$  values. We detect an “optimal”  $p$  using empirical and subjective methods, as described in Section 4. Ideally, we would like to find an objective way to determine the existence of an optimal  $p$ , such as performing a set of parallel runs and comparing the results respectively, as was done by Chen et al. (2020) and documented in that paper.
- In the early stage of this work, a small data sample was used to estimate the  $p$  value and the statistical errors for the observation and the first guess. Later those statistical errors were adjusted based on a months-long experimental run, but  $p$  was not adjusted.
- The statistical errors and error correlation length used in this study do not vary with seasons as they should (Glahn et al. 2017, Carley et al. 2018).
- We did not do much on the modification of the forecast error variance. The forecast error variance is still prescribed as a constant in NLTF space; i.e. the variation with topography is neglected for simplicity. Some recent studies prescribed the development of a heterogeneous forecast error covariance (Michel et al. 2011), specifically, the error statistics are formulated based on stratified state groups. We think these methods may be suitable for ceiling and visibility, particularly in the mountain and valley regions.
- Because the variances of the converted visibility and ceiling height are still varying clearly with the new states, we wonder whether there is a function that can convert visibility and ceiling height into near-Gaussian variables. Note the main measures are for categorical statistics. In Glahn and Im (2015), they explain why they build up a statistical forecast model for categorical visibility and ceiling height forecast, rather than using them as continuous variables, quoting here: “Many attempts to statistically deal with highly non-normal variables as continuous, especially where the rare values are the most important, have met with limited success, if not outright failure”.

## **Acknowledgements**

*Disclaimer: 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.*

*We are grateful to Michael Lueken for his assistance in merging the upgrade RTMA, including the NLTF, into the GSI. Thanks also to Jenny Colavito, Annette Gibbs, and Binbin Zhou for their insightful reviews and suggestions, and to Gang Zhao for his insightful comments on the original manuscript.*

## Appendix A. Nonlinear Transformation Formula

This appendix describes transformations of a measured variable that tend to render the error in that function uniform when the error in the original variable isn't. In the case where the error in the original variable is proportional to that variable, the needed transformation is the logarithm; for other cases where the error in the original variable is proportional to some power of the original variable, then the needed function can be thought of as a generalization of the logarithm,  $(x^p - 1)/p$ , say. We can also always scale the function to make the standard error in the function of the measurement unity.

Let the error  $\sigma(s)$  of a measured variable  $s$  be a nonlinear function of  $s$  itself. Then, if  $f(s)$  is some nonlinear function of  $s$ , the standard error,  $\phi$ , of  $f$  is, to a first approximation, given by:

$$\phi = \frac{df}{ds} \sigma(s). \quad (\text{A.1})$$

So, if we want the function to be such that  $\phi$  is constant over the range of  $s$ , then all we have to do is to integrate the differential equation,

$$\frac{df}{ds} = \frac{\phi}{\sigma(s)}. \quad (\text{A.2})$$

Moreover, we can choose a scaling of this function,  $f$ , to make the magnitude of  $\phi$  conveniently equal to one, whereupon the needed function is simply defined by the integral:

$$f(s) = \int^s \frac{1}{\sigma(s')} ds', \quad (\text{A.3})$$

For example, if  $\sigma(s) = as$ , for some constant,  $a$ , sufficiently smaller than one, the function,

$$f(s) = \frac{1}{a} \ln(s), \quad (\text{A.4})$$

will provide the variable transformation of the observation  $s$  that now has a standard error approximately unity for the whole range of  $s$ .

If instead, the error in  $s$  is proportional to some power of  $s$ , say,

$$\sigma(s) = ax^{1-p}, \quad (\text{A.5})$$

for some  $p$ , then the needed function,  $f(s)$ , will be:

$$f(s) = \frac{1}{a} G(p; s), \quad (\text{A.6})$$

where  $G(p; s)$  is the 'generalization of the logarithm',

$$G(p; s) = \frac{(x^p - 1)}{p}. \quad (\text{A.7})$$

For other, more general forms of the function  $\sigma(s)$ , we can still use the same integration trick, but we won't then be using the  $G$  function. Some of the functional forms we obtain by varying the exponent  $p$  are illustrated in Figure 8.

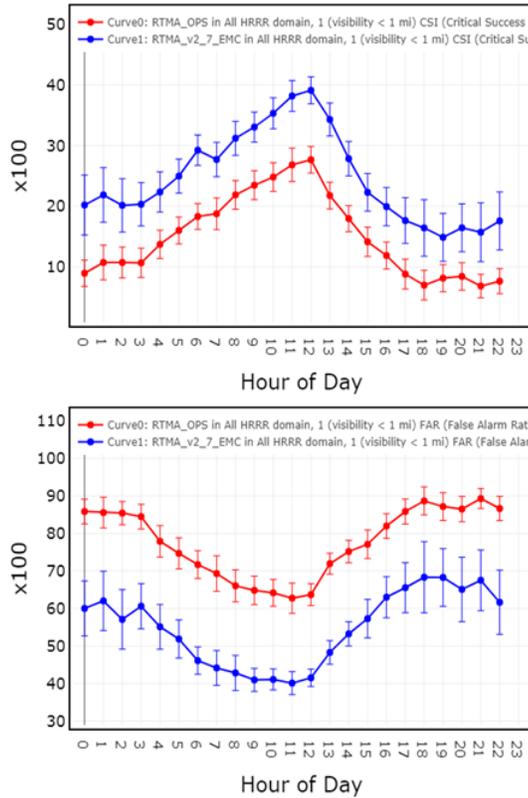
## **Appendix B. Comparison of hourly analysis between the control and the experiment**

As mentioned in Section 5, in addition to NLTF, a different data selection scheme is used in the hourly RTMA experimental run, which only chooses the observation closest to the analysis time. Therefore the comparison between the experimental and control runs is not an “oranges to oranges” comparison for detecting the impact of NLTF. However, it is necessary to demonstrate these results, because the NLTF is used in the latest operational RTMA system.

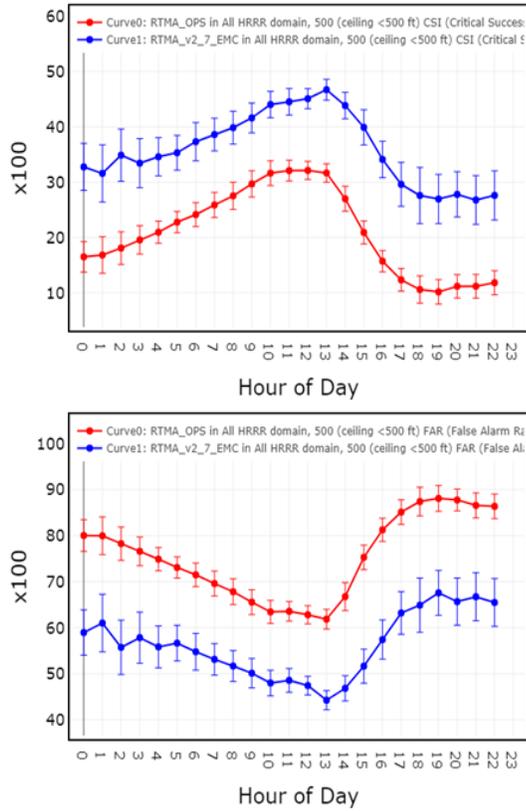
### **B.1 Hourly averaged critical success index and false alarm ratio**

Figures B1-B3 shows the Critical success index (CSI, top panels) and false alarm ratio (FAR, bottom panels) for visibility less than one mile, less than three miles, and less than or equal to five miles, respectively. The values are hourly averaged for the period of August 1 to October 12, 2018. These figures are generated by utilizing the Model Analysis Tool Suite (MATS, 2018) developed by GSD, NOAA. The red line is for the control run and the blue is for the experimental runs. CSI measures the fraction of correctly predicted events, a score of 1 denoting a perfect forecast model, whereas FAR is the percentage of the wrong forecast out of the total forecast records for a given event, a score of 0 is perfect. It is important to point out that the upper values of each category are used to classify the data into two groups, for instance, a group with visibility less than three miles or with visibility less than or equal to five miles. The performance information in an FAA-defined category can be inferred by comparing the figures of adjacent groups.

The top panels of Figures B1-B2 show that the experimental run improves the CSI, while significantly reducing the FAR in category LIFR for visibility and ceiling height respectively. The improvements significantly pass the 95% significance bars.

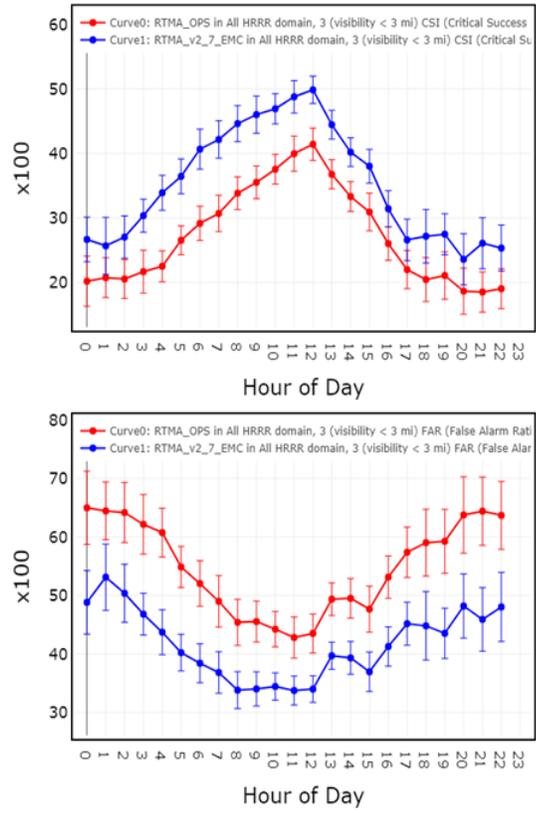


**Figure B1: Critical success index (top panel) and false alarm ratio (bottom panel) for visibility < 1 mile aggregated by time of day over a period covering August 1 to October 12, 2018. The red line is for the control run and the blue for the experimental run. The vertical bars give the range of 95% significance.**

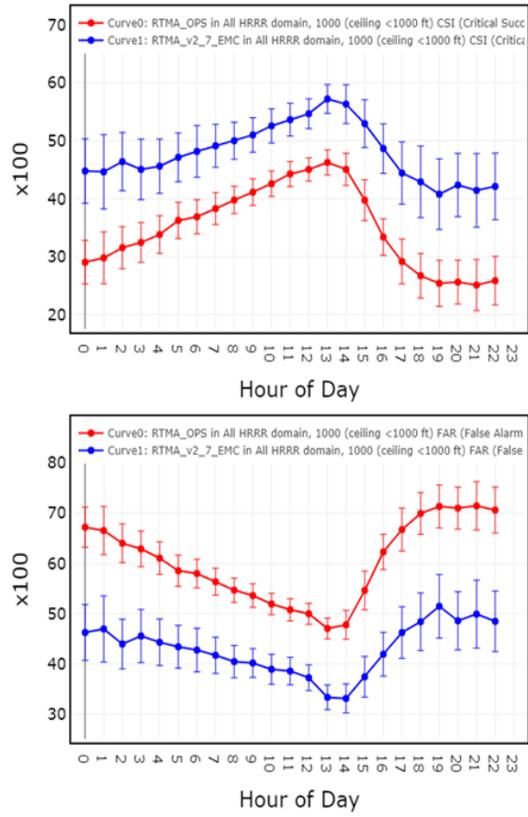


**Figure B2: Similar to Figure B1 but for ceiling height < 500 feet.**

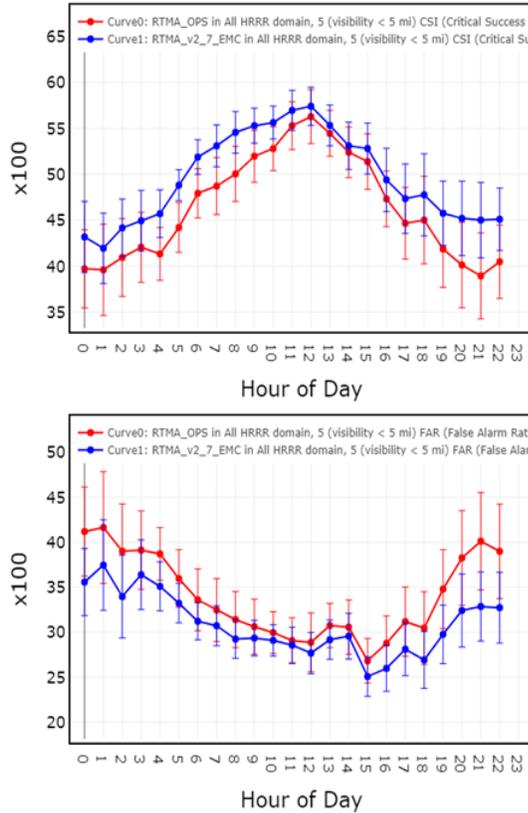
Figure B3 shows the CSI and FAR for visibility less than three miles. Though the experiment improves the CSI and reduces the FAR, the improvement in the CSI is not significant during some afternoon periods. Figure B4 shows the CSI and FAR for ceiling height less than 1000 feet, the improvements are clearly significant. When visibility and ceiling height values reach the upper value of MVFR, the experimental run does not show much improvement, specifically for visibility, as shown by Figure B5 where the significance bars overlap.



**Figure B3: Similar to Figure B1 but for visibility < 3 miles.**



**Figure B4: Critical success index (top panel) and False alarm ratio (bottom panel) for ceiling height < 1000 feet.**



**Figure B5: Critical success index (top panel) and false alarm ratio (bottom panel) for visibility < 5 miles.**

## B.2 Hit rate and false alarm ratio

Tables B1 and B2 list the Hit rate and false alarm ratio (annotated in brackets) between the control and the experiment for visibility and ceiling height respectively, computed with the data of June and July 2018. Data are matched between the control and the experiment, and only assimilated data are considered in the computation. Those two measures demonstrate the improvements, consistently with the aforementioned. Tables B3 and B4 list the same statistics but with data from October and November 2018. In short, the hourly RTMA experimental runs improve the visibility and ceiling height based on the evaluation with large data sets.

**Table B1: Hit rate and false alarm ratio (annotated in brackets) computed from the analysis visibility generated by hourly RTMA experimental runs and the control runs respectively, with data from June and July 2018.**

	LIFR	IFR	MVFR	VFR
Control	39.03 [35.77]	23.87[72.87]	28.87[74.72]	98.93[1.31]
Experiment	40.05[32.97]	24.69[72.77]	28.83[74.76]	98.96[1.25]

**Table B2: Similar to Table B1, but for ceiling height.**

	LIFR	IFR	MVFR	VFR
Control	70.28[24.58]	54.18[50.07]	51.47[39.11]	98.40[2.55]
Experiment	71.25[21.08]	55.40[47.88]	53.26[38.16]	98.50[2.45]

**Table B3: Hit rate and false alarm ratio (annotated in brackets) computed from the analysis visibility generated by hourly RTMA experimental runs and the control runs respectively, with data from October and November 2018.**

	LIFR	IFR	MVFR	VFR
Control	53.39[32.73]	46.06[50.07]	43.50[63.18]	98.15[1.82]
Experiment	54.34[30.45]	46.68[50.02]	43.54[63.16]	98.18[1.78]

**Table B4: Similar to Table B2 for ceiling height, but the analysis data is from October and November 2018.**

	LIFR	IFR	MVFR	VFR
Control	71.67[21.50]	66.09[46.05]	65.81[30.13]	95.78[4.27]
Experiment	72.97[17.04]	67.41[43.40]	68.11[29.25]	96.04[4.08]

## References

Benjamin, S. G., and coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694. <https://doi.org/10.1175/MWR-D-15-0242.1>.

Bocquet, M., C. Pires, and L. Wu, 2010: Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation. *Mon. Wea. Rev.*, **138**, 2997--3023.

Carley, J. and coauthors, 2018: Revisions and Extensions to C&V Processing Based on Evaluation of 15-minute RTMA-RU Products. <https://docs.google.com/document/d/1FOzne6HKLgfz0DBmxQmGhtAwCTAQgfWoeW4loy1rB6k/edit?usp=sharing>

Chen, L., C. Liu, M. Xue, R. Kong, and Y. Jung, 2020: Use of Power Transform Mixing Ratios as Hydrometeor Control Variables for Direct Assimilation of Radar Reflectivity in GSI-based En3DVar and Tests with Five Convective Storms Cases. *Mon. Wea. Rev.*, (to be submitted).

De Ponca, M. S. F. V., and coauthors, 2011: The Real-Time Mesoscale Analysis at NOAA's National Centers for Environmental Prediction: Current status and development. *Wea. and Forecasting*, **26**, 593--612.

Glahn, B. and J.-S. Im, 2015: Objective analysis of visibility and ceiling height observations and forecasts. MDL Office Note 15-2, NWS/Meteorological Development Laboratory. [Available online at: [https://www.weather.gov/media/mdl/MDL\\_OfficeNote15-2.pdf](https://www.weather.gov/media/mdl/MDL_OfficeNote15-2.pdf) ]

MATS, 2018: <https://www.esrl.noaa.gov/gsd/mats/>

Michel, Y., T. Auligné, and T. Montmerle, 2011: Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables. *Mon. Wea. Rev.*, **139**, 2994–3015.

Wu, W.-S., R. J. Purser, and D. F. Parrish, 2002: Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905--2916.

Yang, R., M. Ponca, S. Levine, G. DiMego, J. Whiting, J. Carley, G. Manikin, and A. Gibbs, 2017: Development of a rapid-update real-time mesoscale analysis of ceiling and visibility. *Working Group on Numerical Experimentation Blue Book*. [Available online at [http://wmc.meteoinfo.ru/bluebook/uploads/2017/docs/01\\_Yang\\_Runhua\\_real\\_time\\_mesoscale\\_analysis.pdf](http://wmc.meteoinfo.ru/bluebook/uploads/2017/docs/01_Yang_Runhua_real_time_mesoscale_analysis.pdf).]

Yang, R., R. J. Purser, J. R. Carley, M. Ponca, Y. Zhu, S. Levine, and W.-S. Wu, 2019: Applying a general nonlinear transformation to the analysis of surface visibility and cloud ceiling height. *Working Group on Numerical Experimentation Bluebook*. [Available online at [http://bluebook.meteoinfo.ru/uploads/2019/docs/01\\_Yang\\_Runhua\\_RTMA\\_sensitivity.pdf](http://bluebook.meteoinfo.ru/uploads/2019/docs/01_Yang_Runhua_RTMA_sensitivity.pdf).]