

Conference Proceedings Paper

Probabilistic Forecast of Visibility for Gimpo, Incheon and Jeju International Airports using Weighted Model Averaging

Keunhee Han¹, Hee-Wook Choi², Yeon-Hee Kim², and Chansoo Kim^{1,*}

¹ Department of Applied Mathematics, Kongju National University, 56 Gongjudaehak-ro, Gongju-si, 32588, Korea; kehan@kongju.ac.kr

² Applied Meteorology Research Division, National Institute of Meteorological Sciences, Seogwipo, 63631, Jeju, Korea; wook2845@korea.kr; yoenheekim@korea.kr

* Correspondence: chanskim@kongju.ac.kr

Abstract: In this study, weighted model averaging (WMA) was applied to calibrate ensemble forecasts generated from the limited area ENsemble prediction System (LENS). WMA is an easy to implement post-processing technique that assigns greater weight on an ensemble member forecast that exhibits better performance; it was used to provide probabilistic visibility forecasting in the form of a predictive probability density function for ensembles. The predictive probability density function is a mixture of discrete point mass and two-sided truncated normal distribution components. Observations were obtained from Gimpo, Incheon, and Jeju International Airports and 13 ensemble member forecasts derived from the LENS, for the period December 2018 to June 2019. Prior to applying to WMA, reliability analysis was conducted using rank histograms and reliability diagrams to identify the statistical consistency of ensembles with corresponding observations. The WMA method was then applied to each raw ensemble model and a weighted predictive probability density function was proposed. Performances were evaluated by mean absolute error, continuous ranked probability score, Brier score, and probability integral transform. The results showed that the proposed method provides improved performance compared to the raw ensembles, indicating that the raw ensembles are well calibrated by predicted probability density function.

Keywords: probabilistic visibility forecasting; reliability analysis; tow-sided truncated normal distribution; weighted model averaging

1. Introduction

The aviation sector requires new high-quality forecast information for all types of weather conditions. In particular, it is difficult to provide high-quality information on weather variables (e. g., wind direction, wind speed, cloud altitude, visibility, etc.) directly related to aviation safety with existing numerical weather prediction (NWP) systems. Visibility is one of the important aspects of aviation weather. A lack of visibility is hazardous for airplane landing operations and can be detrimental for its management, as it can lead to delays and cancellations. Efforts have been made to produce more reliable and accurate prediction information based on the ensemble NWP system. However, the visibility forecasts generated from the NWP can be biased and dispersive, due to the limitations of the models. Therefore, deterministic and probabilistic forecasts can be applied to provide improved results by calibrating these uncertainties.

Several methods for forecasting visibility and post-processing forecasts in short-time forecasting have been developed. Vislocky and Fritsch [1] compared the performance of observation-based, MOS-based, and persistence climatology models, for short-term deterministic ceiling and visibility

forecasts. Leyton and Fritsch [2,3] extended the observation-based approach using high-density and, later, high-frequency networks of surface observations, to produce probabilistic forecasts. Pasini et al [4], Bremnes and Michaelides [5], and Marzban et al. [6] applied neural networks for probabilistic visibility forecasting. Zhou et al. [7] described the use of a short-range ensemble forecast system to generate probabilistic visibility forecasts. Roquelaure and Bergot [8,9] and Roquelaure et al. [10] were the first to use Bayesian model averaging (BMA) in visibility forecasting. Their studies modeled binary low-visibility outcomes using a local ensemble prediction system at Charles de Gaulle Airport in Paris. However, none of the existing methods provides a general framework for generating a full predictive probability density function (PDF) for visibility. Chmielecki and Raftery [11] applied BMA for visibility, to probabilistic visibility forecasting, using a fully predictive PDF that is a mixture of discrete point mass and beta distribution components. BMA, introduced by Raftery et al. [12], helps in producing calibrated predictive PDFs for any weather parameter of interest. This method has been successfully used to generate probabilistic forecasts for temperature, sea level pressure, quantitative precipitation [13,14], wind speed [15,16], and wind direction [17].

We analyzed the probabilistic forecast of visibility using a statistical post-processing method based on the reliability analysis for ensemble forecasts of visibility. Various statistical post-processing techniques, for example, BMA [12] or ensemble model output statistics (EMOS [18]), are widely used to reduce systematic errors and uncertainties caused by initial conditions and parametrization in NWP systems [19, 16, 11, 20]. In addition, these methods can improve the accuracy by increasing the correlation between observations and ensemble forecasts, or by removing the existing biases. Among statistical post-processing methods, the most preferred method is to provide probabilistic forecasts that have a full probability distribution of an interest variable. The basic concept of probabilistic prediction is to derive probabilistic forecasts by applying a PDF to future weather variables for events, and to provide information on the magnitude of the occurrence and possibility of an interesting event (or threshold), through its probability generated from the predictive PDF. That is, the objective of probabilistic prediction is to reduce the variation of predictive distribution, hence leading to conformity between probability density functions of observations and those of corresponding ensemble forecasts [21, 22].

In this study, we propose weighted model averaging (WMA) as a way of generating probabilistic forecasts for visibility. The basic concepts of the WMA model are almost the same as those of the BMA model, but provides an easier means to estimate parameters compared to the BMA. A detailed description is provided in Section 3.

The remainder of this paper is organized as follows. In Section 2, we briefly describe the visibility forecasts generated by the LENS. In Section 3, we describe WMA (WMA), and its application to visibility forecasting. The results of the method for daily 1 h forecasts of visibility are presented in Section 4. Finally, conclusions are presented in Section 5.

2. Data

Visibility data obtained from Gimpo, Incheon, and Jeju International Airports in South Korea, and 13 ensemble member forecasts derived from the LENS, a numerical forecast model operated by the Korea Meteorological Administration, were used in this study. Datasets of hourly visibility, precipitation, and relative humidity were obtained between January 1 2017 and June 30, 2019, and the ensemble forecast was conducted at 0000 UTC and 1200 UTC for projection times of 1 h, 4 h and 24 h. Since the model resolution changed on November 29, 2018, only the datasets from December 2018 to June 2019 were analyzed. The information on the observation stations is presented in Table 1.

Prior to analysis, we checked the missing values of ensemble forecasts and their corresponding observations for the entire period. If either was absent, both were removed from the datasets. Because the property of visibility may vary depending on the season, therefore, to consider this seasonal characteristic, the datasets were converted into seasonal data (winter and spring) for three stations.

Table 1. Ensemble prediction system and stations used in the study.

Ensemble prediction system	Limited area ENsemble prediction System (LENS) with 13 ensemble members		
Data period	2018.12.01–2019.06.30		
UTC	00 UTC		
Projection time	4 h to 24 h		
Station	Station	Latitude	Longitude
	Gimpo Int. Airport (110)	37.5	126.4
	Incheon Int. Airport (113)	37.4	126.7
	Jeju Int. Airport (182)	33.5	126.5
Predictant	Visibility (km)		
Predictors	Visibility, relative humidity, and precipitation forecasts generated from LENS		

We first examined the empirical distributions of the observed visibility data and ensemble forecasts. Plots of the empirical distributions for Station 110 are presented in Figure 1. The first histogram presents the frequencies of all observed visibilities, and the second histogram shows frequencies only for observations below 10 km. The majority of all visibilities are recorded at exactly 10 km (Figure 1(1)), which indicates that observation is censored at 10 km, even if the visibility measure is more than 10 km. In addition, this plot shows a mixture of discrete probabilities of point masses at visibilities of 10 km; therefore, a continuous probability distribution is appropriate. In Figure 1(2), the observations are highly discretized. A histogram of the ensemble forecasts is presented in the Figure 1(3). The scales of the observation and ensemble forecasts vary considerably. Ensemble forecasts generate visibility values much greater than 10 km and are not constrained at 10 km, in contrast to values obtained from observations. In addition, ensemble forecasts are continuous and are skewed to the right. For the Stations 113 (Incheon) and 182 (Jeju), the empirical distributions of the observed visibility and ensemble forecast showed similar patterns as obtained at Station 110.

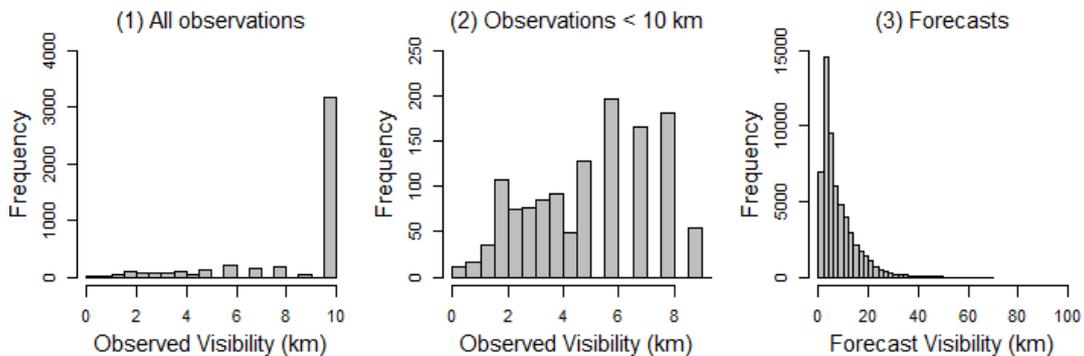


Figure 1. Histograms of visibility observations and forecasts for the Station 110: (1) all observations, (2) observation < 10 km, and (3) ensemble forecasts.

3. Materials and Methods

To set up the probabilistic prediction model for visibility, the characteristics of data from observed visibility and from the corresponding ensemble forecasts should be considered. The observed visibility is censored at 10 km, while the corresponding ensemble forecasts generated from the LENS have positive real numbers. In other words, even if visibility was observed for more than 10 km, the observed visibility was recorded as 10 km; in contrast, the corresponding values from the ensemble forecasts were used as they were generated. Therefore, we need a probabilistic model that reconciles the results obtained from these different datasets.

In this study, we consider WMA, which models the predictive PDF of a weather quantity of interest y , as a mixture of the conditional PDFs. Let f_k be the k th ensemble member forecast and $h_k(y|f_k)$ be a mixture of the conditional PDF given a specific forecast f_k . The WMA predictive PDF is given by

$$p(y|f_1, f_2, \dots, f_K) = \sum_{k=1}^K w_k h_k(y|f_k) \quad (1)$$

where w_k is the weight of the k th ensemble member forecast, and refers to its relative skill over the training period. The weights are constrained to be non-negative and sum to 1.

To determine the mixture of the conditional PDF $h_k(y|f_k)$, we consider a two-component model. The first part consists of a point mass at 10 km and corresponds to the probability that the recorded visibility is 10 km, which is conditional on the k th forecast in the ensemble. The second component of the model assigns a member-specific PDF to visibility, given that it is less than 10 km. We use a two-sided truncated normal distribution defined in (0, 10).

First, we apply a logistic regression model to estimate the probability that the observed visibility is 10 km, given the forecast of the k th ensemble member f_k , as follows

$$\text{logit } P(y = 10|f_k) = \log \frac{P(y=10|f_k)}{P(y<10|f_k)} = a_{0k} + a_{1k}f_k \quad (2)$$

where a_{0k} and a_{1k} are regression coefficients, and these parameters are estimated from a logistic regression model using the member forecasts in the training period as predictors, and a vector of binary indicator of $y=10$ as the response variable.

To predict visibility when the observed visibility is less than 10 km, we consider a two-sided truncated normal distribution. The observed visibility y has a normal distribution, with mean μ and variance σ^2 defined for $0 < y < 10$, and the PDF for $0 < y < 10$ is given by

$$p(y|\mu, \sigma) = \frac{\phi\left(\frac{y-\mu}{\sigma}\right)}{\sigma\left(\Phi\left(\frac{10-\mu}{\sigma}\right) - \Phi\left(-\frac{\mu}{\sigma}\right)\right)}, \quad 0 < y < 10 \quad (3)$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution and $\Phi(\cdot)$ is its cumulative distribution function.

Combining the two components of the model, we build a final conditional PDF for visibility, given the k th ensemble member forecast as follows

$$h_k(y|f_k) = P(y < 10)g_k(y|f_k)I(y < 10) + P(y = 10|f_k)I(y = 10) \quad (4)$$

where $g_k(y|f_k)$ is a member-specified truncated normal distribution. The final WMA model for the predictive probability density function of visibility y is given by

$$p(y|f_1, \dots, f_K) = \sum_{k=1}^K w_k [P(y < 10|f_k)g_k(y|f_k)I(y < 10) + P(y = 10|f_k)I(y = 10)] \quad (5)$$

where

$$g_k(y|f_k) = \frac{\phi\left(\frac{y-\mu_k}{\sigma_k}\right)}{\sigma_k\left(\Phi\left(\frac{10-\mu_k}{\sigma_k}\right) - \Phi\left(-\frac{\mu_k}{\sigma_k}\right)\right)}, \quad 0 < y < 10 \quad (6)$$

with the mean $\mu_k = b_{0k} + b_{1k}f_k$, and standard deviation σ_k of the truncated normal distribution.

For the given observation of visibility less than 10 km, the parameters b_{0k} and b_{1k} , and the standard deviation σ_k are estimated using the method of maximum likelihood.

The parameters w_1, \dots, w_K are estimated as follows: After estimating the parameter $\mu_k = b_{0k} + b_{1k}f_k$, standard deviation σ_k , and probability that the observed visibility is 10 km given the forecast of the k th ensemble member f_k over the training period, the median value is derived from the truncated normal distribution based on Eq. (6) to estimate the visibility for less than 10 km. The corresponding estimates of observed visibility \mathbf{y} during the training period are obtained as

$$\hat{\mathbf{y}} = \begin{bmatrix} \hat{\theta}_{11} & \cdots & \hat{\theta}_{1K} \\ \vdots & \ddots & \vdots \\ \hat{\theta}_{n1} & \cdots & \hat{\theta}_{nK} \end{bmatrix}$$

where $\hat{\delta}_{i,k} = \begin{cases} 10, & P(y_i = 10|f_{i,k}) \geq 0.5, \\ \text{median value,} & \text{otherwise} \end{cases}$, $i = 1, \dots, n$, $k = 1, \dots, K$, and K is the total number of ensemble members and n is the total number of observations. We used mean absolute error (MAE) and non-negative least squares to determine the weights, w_1, \dots, w_K . The weight based on MAE is used to assign the largest weight to the ensemble member forecast with the smallest prediction error in the training period. In contrast, the weights based on non-negative least squares are determined by minimizing the weighted combinations

$$\hat{\mathbf{w}} = (\hat{w}_1, \dots, \hat{w}_K) = \arg \min \sum_{i=1}^n (y_i - \sum_{k=1}^K w_k \hat{\delta}_{ik})^2, w_k \geq 0, \sum_{k=1}^K w_k = 1. \quad (7)$$

To select one of the two estimated weights, each prediction error is calculated for the training period, and then the weights $\mathbf{w} = (w_1, \dots, w_K)$ that provide the smallest prediction error are finally selected.

4. Results

We compared the performance of the models on the ensemble forecasts from December 2018 to June 2019. The point forecast from the WMA is obtained by evaluating the median of the predictive probability density function. Similarly, we take the median of ensemble forecasts to be the point forecast associated with the raw ensembles. In the calculation of prediction performances, all ensemble forecasts greater than 10 km were set to 10 km to facilitate the comparisons between WMA and ensemble forecasts.

A rank histogram (RH) (Hamil 2001; Wilks 2011) was used to assess the reliability of visibility ensemble forecasts and their corresponding observations at the three stations. The RH is a very useful visual tool for evaluating the reliability of ensemble forecasts and for identifying errors related to their mean and spread.

The RH for 13 ensemble forecasts and the corresponding observation visibilities for three stations, that is, (Gimpo (110), Incheon (113), and Jeju (183)) are presented in Figure 2. In general, the RH shows different trends and dispersions for each station. For Station 110 (Gimpo), an RH with high counts near the right extreme and low frequency counts near the left extreme presents a systematic error in the data of the ensembles; ensemble forecasts have a strong negative bias, which indicates an under-estimation. For Station 113 (Incheon), the RH shows nearly similar frequency counts at both extremes but has a slightly higher frequency on the right. This implies that the ensemble forecasts may be under-dispersive and have a weak negative bias. However, RH for Station 182 (Jeju) tends to have an almost uniform distribution, although it has a slightly higher frequency on the extreme left of the histogram.

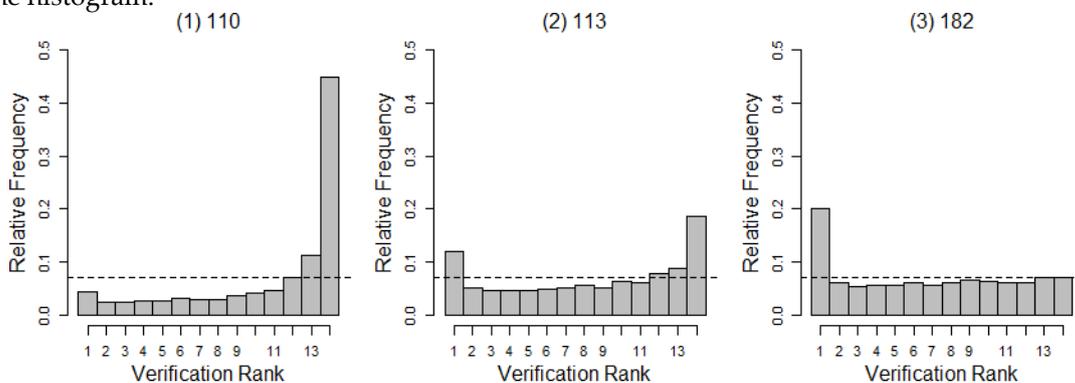


Figure 2. Rank histograms for Stations 110, 113 and 182.

Table 2. Reliability index for each station.

Station	110	113	182
Reliability index	0.837	0.375	0.261

The reliability index (Delle Monache et al., 2006) was used to quantify the deviation of the rank distribution from uniformity. The reliability index (RI) is defined as,

$$\sum_{k=1}^K \left| p_k - \frac{1}{K} \right|,$$

where K denotes the number of classes in the rank histogram, and p_k denotes the observed relative frequency in class k . If the ensemble forecasts and observations were obtained from the same distribution, the RI should be zero. The RI for each station in Figure 2 is listed in Table 2, and their values show that they are far from uniformity.

To evaluate the reliability of the forecast probability of the ensemble forecasts, a reliability diagram was constructed. The reliability diagram is a highly useful tool that shows how often a forecast probability actually occurs. For accurate reliability, the forecast probability and frequency of occurrence should be equal, and the plotted points should align on the diagonal.

The forecast probability of the ensemble forecasts can be obtained using a particular threshold. Since the observed visibility is curtailed at 10 km in this study, the occurrence of observed visibility is defined as

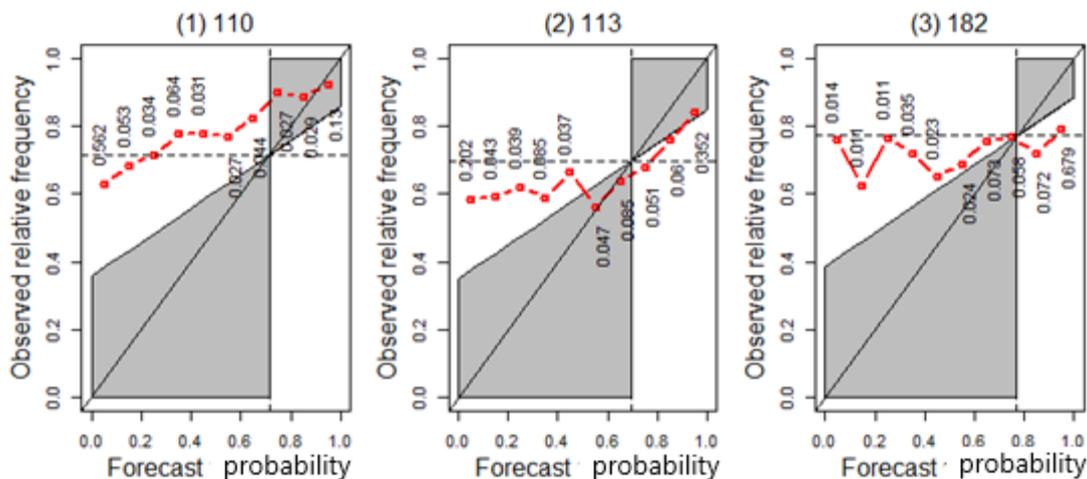
$$o_i = \begin{cases} 1, & \text{if } y_i \geq 10 \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

In addition, the forecast probability of ensemble forecasts is computed using 13 ensemble members generated from LENS

$$p(y = 10) = \frac{n_x}{n_t}, \quad (9)$$

where n_t and n_x denote the number of ensemble members and the number of ensemble members that are greater than or equal to 10 km, respectively.

The occurrence of the observed visibility and their forecast probability defined in Eqs. (8) and (9) are calculated using a visibility ensemble forecast and its corresponding observation for each station. A reliability diagram is shown in Figure 3. In the figure, for Station 110 (Gimpo), the reliability curve is located above the diagonal line, which implies under-forecasting. This result is the same as that obtained for its RH. For Stations 113 (Incheon) and 182 (Jeju), the reliability diagram shows over-confident forecasting, which indicates that the ensemble forecast is severely under-dispersive with respect to the forecasts of $p(y = 10)$. Therefore, the reliability diagrams of all stations show poor performance.



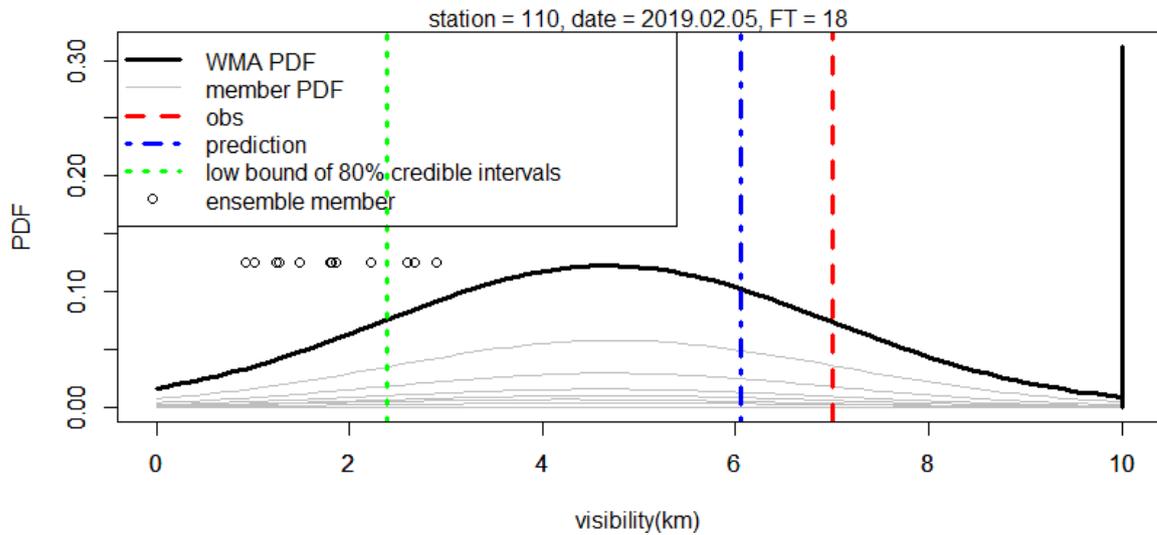


Figure 4. WMA predictive probability density function (PDF) for Station 110. The vertical line at 10 km represents the WMA probability of visibility of 10 km, the vertical line (red) is the verifying observation, the blue vertical line is the WMA forecast, and the green vertical line is the 10th percentile of the WMA predictive PDF. The thick curve is the WMA predictive PDF of visibility given that it is less than 10 km, and the thin curves represent the individual ensemble contributions toward the WMA. Dots represent the ensemble member forecasts.

Table 4 lists each ensemble forecast and the corresponding WMA output details, and Figure 4 depicts the WMA predictive PDF and contributions from the weighted ensemble PDFs. From Table 4 and Figure 4, it can be seen that the range of ensemble forecasts was approximately in the range of 0.9–3.0 km, and the verifying observation value was 7 km. Although the observation value lay outside the range of ensemble forecasts, it was within 80% of the central predictive interval provided by the WMA.

The comparison of the prediction performances of the ensemble median and WMA median forecast for each station in terms of MAE, CRPS, and BS according to the seasons are listed in Table 5. As shown in the table, although the prediction skills of the models differ slightly depending on the season, it can be seen that the prediction skills of the WMA forecast are better than those of the ensemble for all stations. Among them, it can be seen that the prediction error was significantly improved at Station 110, and the prediction error for Station 182 (Jeju) was less improved. These improvements can also be inferred from the results of the previous reliability analysis. Since RH for Station 110 had a strong trend (negative bias), this bias was significantly calibrated in the predictive probability model, whereas RH for Station 182 (Jeju) is generally uniform, indicating that it is relatively less calibrated for bias.

Table 5. Comparison of prediction skills of WMA and ensemble forecasts according to the seasons.

(a) 2018-19 December, January, and February (DJF)						
Station	MAE		CRPS		BS (y=10)	
	Ensemble	WMA	Ensemble	WMA	Ensemble	WMA
110	2.842	1.610	3.914	2.806	0.355	0.211
113	2.263	1.854	3.502	3.272	0.302	0.255
182	0.967	0.942	0.901	0.776	0.223	0.196

(b) 2019 March, April, and May (MAM)						
Station	MAE		CRPS		BS (y=10)	
	Ensemble	WMA	Ensemble	WMA	Ensemble	WMA

110	3.843	0.847	3.248	0.677	0.489	0.156
113	2.048	1.272	1.607	0.909	0.274	0.181
182	0.744	0.592	0.643	0.548	0.165	0.138

We then assessed whether adding additional predictors result in improved WMA forecasts. The available independent variables (predictors) in this study are relative humidity and quantitative precipitation. A quantitative precipitation is mostly 0, which is not useful for estimating visibility. Therefore, only relative humidity is significantly associated with observed visibility, and we limit the inclusion to this additional predictor.

We consider a relative humidity variable for inclusion in each component of the model. The model for binary outcome $y = 10$ is represented by

$$\text{logit } P(y = 10 | f_k, r_k) = a_{0k} + a_{1k}f_k + a_{2k}r_k \quad (10)$$

where f_k and r_k are the member-specified visibility and relative humidity forecasts, respectively. Given that $y < 10$, the mean and standard deviation of the associated member-specific two-sided truncated normal distribution is specified as

$$\mu_k = b_{0k} + b_{1k}f_k + b_{2k}r_k$$

and

$$\sigma_k = c_k$$

By inserting these two components into Equation (6), we consider the predictive PDF that takes into account the relative humidity variable.

Table 6 lists the model performance scores for the WMA for visibility (denoted by WMA (vis)), the WMA for visibility and relative humidity (denoted by WMA (vis, rh)), and the raw ensembles. From the table it can be seen that the WMA models perform better than the raw ensembles across all scores; in particular, the WMA(vis, rh) model shows a slight improvement over the WMA(vis) model. Moreover, at Station 110 (Gimpo), it can be seen that the prediction error improved significantly; the prediction error was less improved for Station 182 (Jeju). The results show that it is similar to the reliability analysis that was mentioned above. We can see that the prediction performance in spring is better than that in winter, and this implies that prediction performance can be affected by season, that is, different prediction performances may be delivered according to the season.

We then evaluated the performance of the complete predictive probability density function generated by the WMA model. Figures 5-10 show the verification RHs for raw ensembles, and the probability integral transform (PIT) histograms for the WMA forecasts for each station, for winter and spring during the test period. To generate the PIT histogram, each WMA cumulative distribution function was evaluated at the corresponding observation that was less than 10 km. For observation of 10 km, the resulting probability was sampled randomly from a uniform distribution in the interval between the quantity $1 - P(y = 10)$ and 1.

Table 6. Comparison of model performance in terms of MAE, CRPS and BS.

station		2018-19 DJF			2019 MAM			
		110	113	182	110	113	182	
MAE	Ensemble	2.842	2.263	0.967	3.843	2.049	0.744	
	WMA	(vis)	1.610	1.854	0.942	0.847	1.272	0.592
		(vis,rh)	1.267	1.300	0.936	0.715	1.213	0.593
CRPS	Ensemble	2.342	1.882	0.901	3.248	1.607	0.643	
	WMA	(vis)	1.191	1.434	0.776	0.677	0.909	0.548
		(vis,rh)	0.898	0.901	0.786	0.545	0.855	0.459
BS	Ensemble	0.355	0.302	0.223	0.490	0.274	0.165	
	WMA	(vis)	0.211	0.255	0.196	0.156	0.181	0.138
		(vis,rh)	0.160	0.144	0.196	0.119	0.166	0.114

The PIT histogram for the test period was analyzed to evaluate the performance of the complete predictive PDF of the WMA models. Through the PIT histogram, we can analyze the improvement of the probabilistic forecasts generated by WMA models compared to the raw ensembles. The RH for raw ensembles and PIT histograms over the test period for each station in winter are presented in Figures 5–7.

First, the RH for Station 110 (Gimpo Airport) shows that raw ensembles have a strong negative bias and weak under-dispersion. PIT histograms for the WMA forecasts have nearly uniform distributions, indicating that WMA models are well calibrated and showed substantial improvement over raw ensembles. In particular, in comparison to the WMA models, it can be seen that the WMA (vis, rh) model is more calibrated than WMA(vis). In the case of Station 113 (Incheon Airport), RH shows an under-dispersive and weak positive bias. However, the PITs for WMA forecasts indicate that these biases and dispersions are considerably reduced by the WMA predictive model. In addition, more uniform patterns of the PIT for WMA (vis, rh), indicate that the WMA (vis, rh) is calibrated fairly well. Jeju Airport (182) shows a slightly different pattern from the two stations. RH has a slightly higher frequency on the low extreme, but it shows an almost uniform pattern overall. Although there is not much change in the overall pattern for PIT, we can see that the extreme frequency is decreased by the WMA models. This indicates that the ensemble was less calibrated.

Figures 8–10 show the PIT histograms for the WMA forecasts over the test period of spring. In Figure 10, the RH for Station 110 shows a strong negative bias, whereas PITs for the WMA forecasts show nearly uniform distributions, implying that the WMA is well calibrated over the range of the predictive PDF. As mentioned above, it can be seen that the WMA(vis, rh) model is better calibrated compared to WMA(vis). Station 113 has patterns that are almost similar to the results analyzed during spring. In the case of Station 182, it can be seen that the PITs in spring are much more calibrated in comparison to the PITs in winter. This implies that the prediction error improved significantly in spring compared to winter.

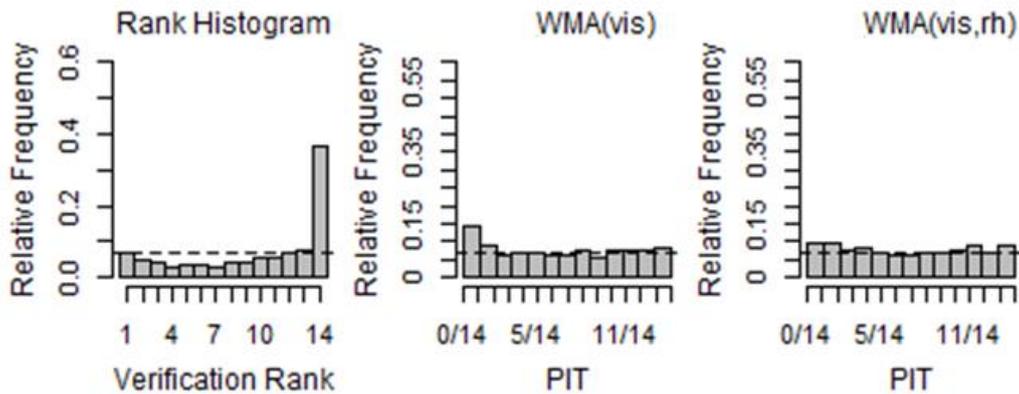


Figure 5. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 110 (2018-19 DJF)

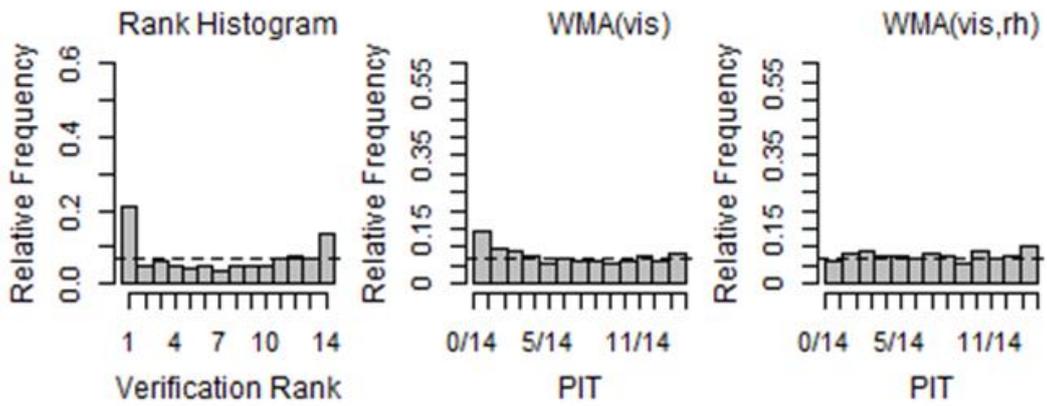


Figure 6. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 113 (2018–2019 DJF)

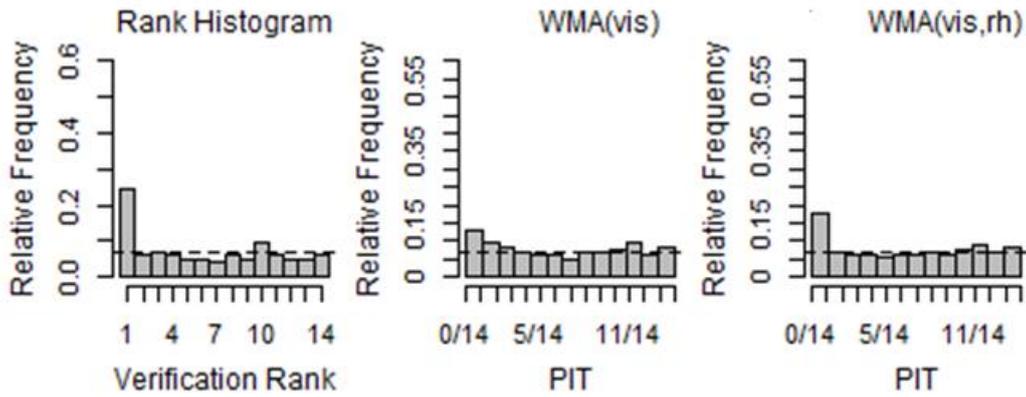


Figure 7. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 182 (2018–19 DJF)

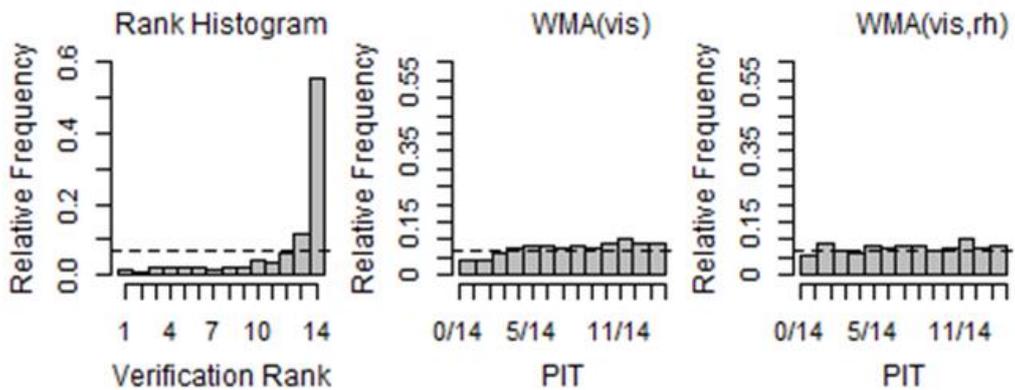


Figure 8. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 110 (2019 MAM)

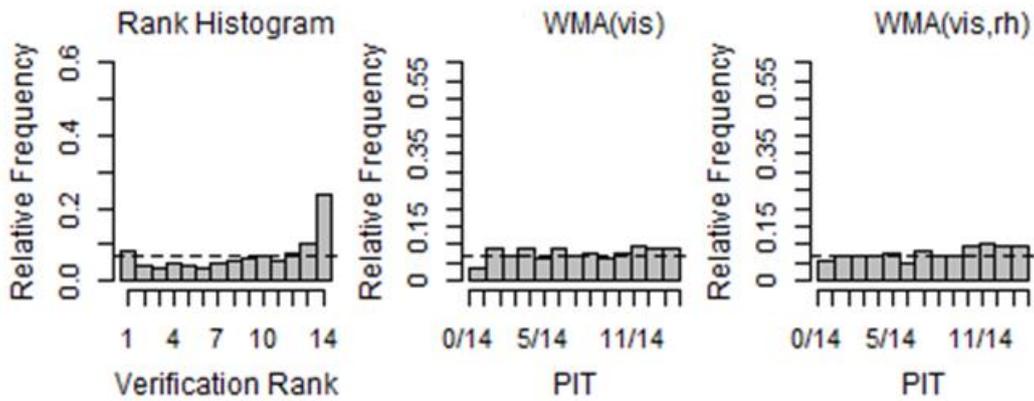


Figure 9. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 113 (2019 MAM)

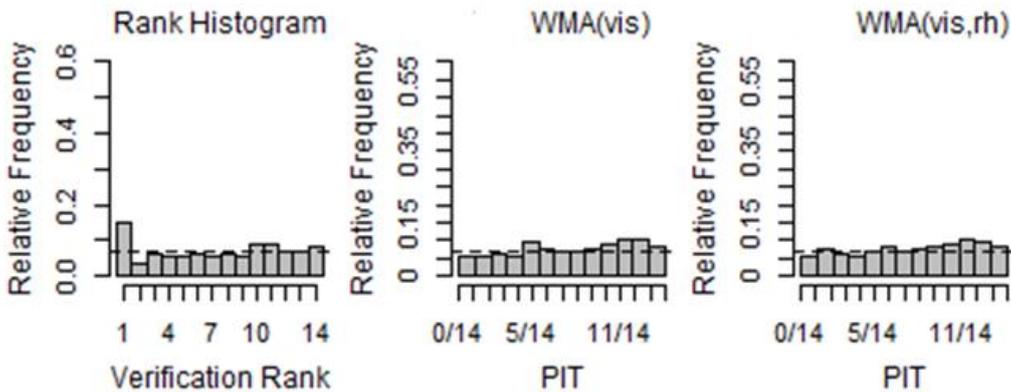


Figure 10. Verification rank histogram for raw ensemble forecasts and PIT histograms for WMA models for Station 182 (2019 MAM)

5. Conclusions

As visibility is a risk to operations and management and causes airplane delays and cancellations, it is one of the important variables in aviation weather. Therefore, significant effort has been made to provide more reliable and accurate visibility forecasts based on ensemble numerical prediction systems. In this respect, this study provides a statistical post-processing method for verifying the performance of an ensemble prediction system, and calibrating the biases and dispersions existing in the ensemble prediction system.

The characteristics of the data from ensemble member forecasts generated from the LENS and from observations were examined, and the bias and dispersion existing in the ensemble forecasts were analyzed to construct an appropriate statistical model. Based on these results, we suggested a simple WMA method that provides a fully predictive PDF, making it is easier to estimate the parameters of the model using 13 ensemble member forecasts and relative humidity generated from LENS. WMA is almost similar to BMA; however, WMA can estimate weights more easily.

The resulting WMA predictive PDF was well calibrated with respect to the raw ensembles. The WMA model could resolve the problems arising from ensemble forecasts, including both systematic and random errors, and the discrepancy in scale between ensembles and observations. In addition,

we considered additional variables to increase the precision in WMA forecasts. We also wanted to use quantitative precipitation and relative humidity as these are the most informative variables available in our datasets; however most of the 3-h accumulative precipitation have a value of zero and therefore these could not be used. In this application, we found that the relative humidity forecast added some amount of information to the visibility forecasts compared to the use of only visibility ensembles. Because we did not compare the WMA and BMA results, we cannot predict the superiority of the prediction performances between these two methods; however, it is known that the WMA model has the advantage of being much easier to apply than the BMA model. In a further study, we will compare the performances between the two methods using various scoring rules. Finally, the method based on the research is expected to be useful for bias-correction of other aviation variables and for probabilistic forecast analysis.

Acknowledgments: This work was funded by the Korea Meteorological Administration Research and Development Program “Advanced Research on Aviation Meteorology and Urban Meteorology (KMA2018-00621)”.

References

1. Vislocky, R.L.; Fritsch, J.M. An automated, observations-based system for short-term prediction of ceiling and visibility. *Wea. Forecasting* **1997**, *12*, 116–122.
2. Leyton, S.M.; Fritsch, J.M. Short-term probabilistic forecasts of ceiling and visibility utilizing high-density surface weather observations. *Wea. Forecasting* **2003**, *18*, 891–902.
3. Leyton, S.M.; Fritsch, J.M. The impact of high-frequency surface weather observations on short-term probabilistic forecasts of ceiling and visibility. *J. Appl. Meteor.* **2004**, *43*, 145–156.
4. Pasini, A.; Pelino, V.; Potesta, S. A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *J. Geophys. Res.* **2001**, *106*, 14951–14959
5. Bremnes, J.B.; Michaelides, S.C. Probabilistic visibility forecasting using neural networks, *Pure Appl. Geophys.* **2007**, *164*, 1365–1382.
6. Marzban, C.; Leyton, S.M.; Colman B. Ceiling and visibility forecasts via neural networks, *Wea. Forecasting* **2007**, *22*, 466–479.
7. Zhou, B.; Du, J.; McQueen, J.; Dimego, G. Ensemble forecast of ceiling, visibility, and fog with NCEP Short-Range Ensemble Forecast system (SREF). In Preprints, Aviation, Range, and Aerospace Meteorology Special Symp. on Weather-Air Traffic Management Integration, Phoenix, AZ, United States of America, 2009; Meteor. Soc., 4.5.
8. Roquelaure, S.; Bergot, T. A local ensemble prediction system for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *J. Appl. Meteor. Climatol.* **2008**, *47*, 3072–3088.
9. Roquelaure, S.; Bergot, T. Contributions from a Local Ensemble Prediction System (LEPS) for improving low cloud forecasts at airports, *Wea. Forecasting* **2009**, *24*, 39–52.
10. Roquelaure, S.; Tardif, R.; Remy, S.; Bergot, T. Skill of a ceiling and visibility Local Ensemble Prediction System (LEPS) according to fog-type prediction at Paris-Charles de Gaulle Airport. *Wea. Forecasting* **2009**, *24*, 1511–1523.
11. Chmielecki, R.M.; Raftery, A.E. Probabilistic visibility forecasting using Bayesian model averaging. *Mon. Wea. Rev.* **2011**, *139*, 1626–1636.
12. Raftery, A.E.; Gneiting, T.; Balabdaoui, F.; Polakowski, M. Using Bayesian model averaging to calibrate forecast ensembles, *Mon. Wea. Rev.* **2005**, *133*, 1155–1174.
13. Slughter, J.M.; Raftery, A.E.; Gneiting, T.; Fraley, C. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.* **2007**, *135*, 3209–3220.
14. Han, K.; Choi, J.; Kim, C. Comparison of prediction performance using statistical postprocessing methods, *Asia-Pac. J. Atmos. Sci.* **2016**, *52*, 495–507.
15. Slughter, J. M., Gneiting, T; Raftery, A.E. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.* **2010**, *105*, 25–35
16. Thorarindottir, T.L.; Gneiting, T. Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *J. Roy. Stat. Soc.* **2010**, *173*, 371–388.

17. Bao, I.; Gneiting, T.; Gritmit, E.P.; Guttorp, P.; Raftery, A.E. Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Mon. Wea. Rev.*, **2010**, *138*, 1811–1821.
18. Gneiting, T.; Raftery, A.E.; Westveld, A.H.; Goldman, T. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.* **2005**, *133*, 1098–1118.
19. Glahan, H.R.; Lowry, D.A. The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.* **1972**, *11*, 1203–1211.
20. Han, K.; Choi, J.; Kim, C. Comparison of statistical post-processing methods for probabilistic wind speed forecasting. *Asia-Pac. J. Atmos. Sci.* **2018**, *54*, 91–101.
21. Murphy, A.H.; Winkler, R.L. A general framework for forecast verification. *Mon. Wea. Rev.* **1987**, *115*, 1330–1338.
22. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.* **2007**, *102*, 359–378.
23. Hamil, T.M. Interpretation of rank histogram for verifying ensemble forecasts. *Mon. Wea. Rev.* **2001**, *129*, 550–560.
24. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*, 3rd ed.; Elsevier Academic Press, 2011; p. 113.
25. Delle Monache, I., Hacker, J.P.; Zhou, Z.; Deng, X.; Stull, R.B. Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *J. Geophys. Res.* **2006**, *111*, D23407, doi:10.1029/2005/JD006971.
26. Brier, G.W. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.* **1950**, *78*, 1–3.
27. Murphy, A.H. A new vector partition of the probability score. *J. Appl. Meteorol.* **1973**, *12*, 595–600.
28. Gritmit, E.P., Gneiting, T.; Berrocal, V.I.; Johnson, N.A. The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification. *Quart. J. Roy. Meteor. Soc.*, **2006**, *132*, 3209–3220.



© 2020 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).