Retrieval of High-Resolution Grid Type Visibility Data in South Korea Using Inverse Distance Weighting and Kriging

Taeho Kang1) · Myoung-Seok Suh 2)

Abstract: Fog can cause large-scale human and economic damages, including traffic systems and agriculture. So, Korea Meteorological Administration is operating about 290 visibility meters to improve the observation level of fog. However, it is still insufficient to detect very localized fog. In this study, high-resolution grid-type visibility data were retrieved from irregularly distributed visibility data across the country. To this end, three objective analysis techniques (Inverse Distance Weighting (IDW), Ordinary Kriging (OK) and Universal Kriging (UK)) were used. To find the best method and parameters, sensitivity test was performed for the effective radius, power parameter and variogram model that affect the level of objective analysis. Also, the effect of data distribution characteristics (level of normality) on the performance level of objective analysis was evaluated. IDW showed a relatively high level of objective analysis in terms of bias, RMSE and correlation, and the performance is inversely proportional to the effective radius and power parameter. However, the two Krigings showed relatively low level of objective analysis, in particular, greatly weakened the variability of the variables, although the level of output was different depending on the variogram model used. As the level of objective analysis is greatly influenced by the distribution characteristics of data, power, and models used, care should be taken when selecting objective analysis techniques and parameters.

Key Words: Visibility, IDW, Ordinary Kriging, Universal Kriging, Sensitivity test

1. Introduction

A meteorological phenomenon in which horizontal visibility is less than 1 km due to fine water droplets or ice particles in the atmosphere floating near the surface is defined as fog (Eyre et al., 1984; NOAA, 2005; Gultepe et al., 2007a, 2007b; Koracin et al., 2014). As the operation of various public transportation systems such as automobiles, aircraft, and ships increases and becomes more common, sudden reduction of visibility caused by fog can cause large-scale human and economic damage. Low visibility due to fog was the
main cause of the 106 collisions accident at Yeongjong Bridge in 2015 and the 14 collisions accident at the West Coast Expressway in 2020. In addition, fog causes damage to crops, prevents the spread of pollutants, and affects radiation budget (Jhun et al., 1998; Tardif and Rasmussen, 2007; Lee and Suh, 2011; Egli et al., 2018; Lee and Suh, 2018). Therefore, the necessity of accurate fog detection is increasing in order to reduce various damages that can be caused by fog.

Early studies of fog in South Korea were mainly conducted using eye-observation data (Jhun et al., 1998; Cho et al., 2000; Heo and Ha, 2004). Although the available period of the eye-observation data is long, but the number of observation points and observation frequency are very limited, and the subjectivity of the observer is inevitably included to the data (Lee and Suh, 2018; Kang and Suh, 2019). Therefore, there are limits to the study of fog with characteristics of local and spatiotemporal variability only with eye-observation data. To minimize these problems, recently, the fog observation network was expanded and the characteristics of the fog were analyzed using visibility meter or satellite data with a higher spatiotemporal resolution than the eye-observation data (Heo et al., 2008; Lee and Suh, 2018; Kang and Suh, 2019; Han et al., 2020).

The Korea Meteorological Administration (KMA) has introduced visibility meters that allow simultaneous observation of the current weather from 2009. Currently, eye-observation has been discontinued since 2009, about 290 visibility meters have been installed and operated nationwide by 2019. The visibility meter calculates the visible distance by measuring the meteorological optical range (MOR) from the amount absorbed or scattered by air while the light emitted from the transmitter having a color temperature of 2,700 K reaches the detector. Visibility meters are scientific observation tools based on optical properties, and because the observation period is very short (1 minute), it is advantageous for fog observation compared to eye-observation (Lee et al., 2019). However, since the observation is performed in a very narrow space of about tens of cm between the transmitter and the detector, the spatial representation of the observation data is low. Also, the spatial distribution of the stations where the visibility system is installed is irregular. In addition, where equipment installation is difficult, such as the sea, observation is restricted (Lee and Suh, 2018; Kang and Suh, 2019; Oh and Suh, 2020). Another fog observation system, GEO-KOMPSAT-2A (GK2A), can produce regular high-resolution (2 km) data every 10 minutes for all observation areas regardless of geographic location. However, the accuracy is relatively low compared to field observation, and observation is impossible when occluded by clouds (Han et al., 2020).

Since each fog observation network has distinct advantages and disadvantages that are contrasting with each other, it is necessary to combine the two data. However, in order to combine two data with different characteristics, it is necessary to transform the ground observation data into high-resolution grid data like satellite data. Researches are actively being conducted to produce grid-type climate data using objective analysis techniques to produce initial input data for models of various climate data such as temperature and precipitation (Hong et al., 2007; Park, 2009; Ly et al., 2011; Park and Kim, 2013; da Silva et al., 2019; Cho et al., 2020). However, studies on the application and assessment of objective analysis for visibility data have not been conducted until now due to problems such as lack and low reliability of visibility data. When considering the localized characteristics of fog generation, there is a limit to understanding the detailed distribution characteristics of fog occurrences using only observation data. Therefore, it is urgently necessary to generate high-resolution grid-type fog data when considering safe operation in areas where visibility is important, such as aviation, traffic, and maritime transportation, and quantitative verification of the KMA’s fog warning.
In this study, high-resolution grid-type visibility data was calculated by applying the objective analysis techniques of Inverse Distance Weighting (IDW), Ordinary Kriging (OK), and Universal Kriging (UK) to the visibility data. Since most objective analysis techniques are sensitive to major parameters or used models and are affected by the distribution characteristics of data, this study evaluated the level of objective analysis techniques through sensitivity analysis. Section 2 introduces the data and objective analysis techniques used in this study. In Section 3, the results of qualitative and quantitative assessments for the three objective analyses according to the main parameters (or used models) and data characteristics are presented. And performance level based on the sensitivity tests according to the key parameters and the results derived through this study are presented in Section 4 and Section 5, comprehensively.

2. Data and Method

1) Data

In this study, for the calculation of high-resolution grid-type visibility data, the visibility data \( m \) observed at 1-minute interval in about 290 visibility meters provided by the KMA were used. Fig. 1 shows the spatial distribution of the visibility meter stations used in this study. Among them, island regions such as Jeju Island, Ulleungdo, and Baengnyeongdo are far from inland, so, except for these stations, we used the visibility data from 243 points installed in the inland and adjacent islands of South Korea.

In the process of objective analysis of visibility data, in order to integrate with the GK2A fog output later, this study used the map projection and spatial resolution (2 km) identical to the GK2A fog output.

Fig. 1. Spatial distribution of visibility meters used in this study.
2) Method

Fig. 2 shows the process of calculating regular grid-type visibility data using irregularly distributed visibility data. By using the Oh and Suh (2020)’s quality control method, abnormal data included in the visibility meter data, such as outside the observation range or too strong spike or dip data, were removed. And the 1-minute frequency data was recalculated as 10-minute frequency data.

Most objective analysis techniques are affected by the distribution characteristics of data (normality, isotropy, etc.). Therefore, in order to evaluate the level of objective analysis technique according to the distribution characteristics of the data used, the characteristics of the data were classified in step 1. First, through Shapiro-Wilk’s normality test, it was confirmed whether the data follow a normal distribution. Next, statistical representative values such as mean, standard deviation (SD), skewness, and kurtosis were calculated to quantitatively identify the distribution characteristics of the data. After that, a suitable model was determined through variogram modeling to fit the spatial characteristics of the data. Variogram is a statistical measure that indicates the degree of dissimilarity between two data with distance, and is generally used by fitting a variogram model that expresses the spatial variation of variables. In this study, Exponential (Exp), Gaussian (Gau), and Matern correlation functions (Mat) models were adopted with reference to previous studies (Park et al., 2010; da Silva et al., 2019).

In the second step, grid-type data were calculated by three objective analysis techniques. Objective analysis techniques vary widely and the appropriate method is different depending on the data characteristics and resolution, so it is necessary to choose through experimentation (da Silva et al., 2019). Since there was no case of objective analysis of visibility data, we referred to techniques used for objective analysis of precipitation with various spatial variability as in the visibility data (fog). Previous researches mainly used IDW, Cokriging, and PRISM (David et al., 2009; Park, 2009; Ly et al., 2011; Kim et al., 2013). Cokriging and PRISM require another variable that has a linear relationship with the variable to be estimated. In the

![Fig. 2. Flow chart for the interpolation and assessments of visibility meter data.](image)
objective analysis of precipitation and temperature, altitude information is mainly used together, but in the case of visibility, it is difficult to apply the two techniques because there are few variables having a linear relationship. Therefore, in this study, IDW and univariate kriging techniques were selected and used for objective analysis of visibility data (Bostan et al., 2012; da Silva et al., 2019).

IDW is a basic objective analysis technique with the simplest calculation method, and calculates the weight in inversely proportional to the distance between the estimated point and the observation point. Kriging calculates the covariance or variogram between the neighboring points along with the distance weight, and then statistically derives the correlation between the neighboring points and reflects this in the estimation (Park, 2009; Park, 2010). There are various techniques for univariate kriging depending on how the local mean value at the location to be estimated. In this study, we selected ordinary kriging (OK) and universal kriging (UK), which assume that the regional average is constant, and the regional average is constant but changes slightly with a tendency, respectively.

Since the level of objective analysis is directly affected by the parameters involved, a sensitivity experiment is essential. In this study, the calculated levels according to the effective radius, the minimum number of points in the radius, the power parameter, and the threshold value of the variogram model were compared. Table 1 summarizes the sensitivity experiments for various parameters performed in this study.

In step 3, the grid data calculated with the selected parameters and models were evaluated through two assessment methods. Since IDW and Kriging have the characteristics of reproducing the given observation data as it is, it was assessed that the value was actually maintained through a one-to-one comparison between the observation and the estimation obtained on the nearest grid through an stationarity property. Correlation coefficient (Corr.), bias, and RMSE between observed and estimated values were calculated. Next, the Jackknife method was performed to compare the estimated value with the observed value under assumption that missing occurred at an arbitrary point. After performing non-recovery extraction into 5 groups of 20 points, Corr., bias, and RMSE between observations and estimates were calculated for each group. The equations for each statistical estimate are as shown in Eqs. (1) to (3), where $z(u)$ is the estimated value, $z(u_{\alpha})$ is the observed value of the $\alpha$ point, and $n$ is the total number of points in the effective radius.

$$
\text{Corr.} = \frac{\text{cov}(z(u), z(u_{\alpha}))}{\sigma_z(u) \sigma_z(u_{\alpha})}
$$

(1)

$$
\text{Bias} = \frac{\sum(z(u) - z(u_{\alpha}))}{n}
$$

(2)

$$
\text{RMSE} = \sqrt{\frac{\sum(z(u) - z(u_{\alpha}))^2}{n}}
$$

(3)

3. Results

1) Case selection

To evaluate the level of objective analysis according to the spatial distribution characteristics of visibility, two cases of fog and non-fog were selected, respectively.

---

Table 1. Parameters for sensitivity study for each interpolation technique

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Definition</th>
<th>Thresholds</th>
<th>Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effective Radius [km]</td>
<td>Maximum distance from estimation point when performing interpolation</td>
<td>30, 35, 40</td>
<td>IDW, OK, UK</td>
</tr>
<tr>
<td># of Minimum STN</td>
<td>Minimum number of observation points within the effective radius</td>
<td>3, 4, 5</td>
<td>IDW, OK, UK</td>
</tr>
<tr>
<td>Power Parameter</td>
<td>Positive real number used in the exponential term of IDW weights</td>
<td>1.0, 1.5, 2.0</td>
<td>IDW</td>
</tr>
<tr>
<td>Variogram Model</td>
<td>Variogram model to be applied in Krigings</td>
<td>Exp, Gau, Mat</td>
<td>OK, UK</td>
</tr>
</tbody>
</table>

---
Fig. 3 shows the spatial distribution of visibility for the selected two fog (05:00 KST Sept. 29, 2019; 05:00 KST Aug. 30, 2019) and non-fog cases (15:00 KST Jul. 4, 2019; 15:30 KST Mar. 1, 2019). In the cases of fog, it can be seen that red colors (visibility < 1 km) are scattered around the central and southern regions of South Korea. In the cases of non-fog, it can be seen that blue and green colors (visibility > 5 km) appear widely in most areas.

In order to perform objective analysis using Kriging, the optimal variogram model must be selected first. Fig. 4 shows a three types of fitted variogram model for the selected four cases. In the case of Sept. 29, the exponential model (Exp) has a small nugget effect and a tendency to follow the trend of empirical variograms well. On the other hand, in the case of Aug. 30, the nugget effect is the smallest in the Matern correlation function (Mat), but the Gaussian model (Gau) follows the tendency of the empirical variogram to increase rapidly at close range. Therefore, it was determined that...
the ‘Exp’ was suitable for the Sept. 29 case and the ‘Gau’ for the Aug. 30 case. In the non-fog cases, it can be seen that there is no suitable variogram model as the visibility is almost constant regardless of distance.

2) Qualitative assessment

In order to select the optimal objective analysis technique, a sensitivity test on the major parameters (models) was conducted for the four selected cases (Table 1). Fig. 5 shows spatial distribution of visibility calculated by the IDW, OK, and UK based on the major impacting parameters (effective radius, power, model) for the fog case of Aug. 30. At this time, the minimum number of stations in the effective radius was set to three in order to minimize the grid points that can not be calculated.

In Fig. 5, the spatial distribution of visibility interpolated by IDW differs according to the effective radius and power parameter. The locality of the interpolated values tends to be enhanced as the radius of influence decreases and the power parameter increases. This suggests that in order to reproduce a phenomenon with strong locality such as fog, the radius of influence should be small but the power parameter should be large. The spatial distribution of visibility interpolated by two krigings (OK, UK) also shows different spatial distributions according to the effective radius and variogram model. Here the impact of the used model is more pronounced than the radius of influence. There is no significant difference between the two krigings, but the Gaussian model results show a large difference from the other two models. In
addition, the spatial distribution of the visibility interpolated by two krigings shows less spatial variability than that of the interpolated visibility by IDW.

Fig. 6 shows the non-fog cases as shown in Fig. 5. The impact of interpolation method and major parameters on the interpolated visibility are similar to that of fog cases. However, in this case, there is a large difference at the shorelines between the interpolation methods (OK and UK) rather than the used models. This seems to be because, unlike OK, it reflects the spatial change trend of visibility in the UK.

As a result of the objective analysis of the two cases, it seems that IDW interpolates relatively better local phenomena than OK and UK. In particular, OK and UK are not only affected by the use model and data characteristics, but also show problems that weaken the spatial variability of data. This is because kriging has a characteristic that minimizes the variance of the data under the condition that the interpolation value is not biased. Although it depends on the case, it seems that the UK interpolates the spatial variability of visibility a little more realistically than OK.
3) Quantitative assessment

Table 2 shows the results of stationarity test for the selected four cases by power or variogram model and interpolation method (IDW, OK, and UK) values for five sets of 20 randomly extracted observation points from the 243 observation points. IDW has a correlation coefficient (Corr.) greater than 0.98 and an RMSE less than 1.315 km, regardless of cases and power parameters. On the other hand, the correlation coefficients, bias and RMSE of OK and UK are relatively lower and greater than that of IDW, respectively. In addition, the level of their output varies greatly from case to case. Although there are differences
depending on the cases, but regardless of the used model, the UK has a slightly higher level of interpolation in terms of Corr. and RMSE than OK. As explained in the qualitative evaluation, the interpolation level of IDW is getting better as the power parameter increases (Mean Corr.: 0.98, 1.00, 1.00; Mean RMSE: 1.15, 0.39, 0.14 km). However, the levels of interpolation of OK and UK differ depending on the cases, in particular, the used variogram model, as shown in cases of Sept. 29 (Exp) and Aug. 30 (Gau).

Table 3 shows the results as shown in table 2 except for the Jackknife method. As with the exactitude property test, the level of interpolation appears in the order of IDW, UK, and OK, regardless of the cases and major parameters. In IDW, when the power is 2, the level of objective analysis and consistency are the highest. However, unlike in exactitude property test, the level of objective analysis of IDW varies greatly.

<table>
<thead>
<tr>
<th>Method</th>
<th>Date (KST)</th>
<th>Value</th>
<th>Corr.</th>
<th>Bias [km]</th>
<th>RMSE [km]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>P=1.0 M=Exp</td>
<td>P=1.5 M=Gau</td>
<td>P=2.0 M=Mat</td>
</tr>
<tr>
<td>IDW</td>
<td>09.29 (05:00)</td>
<td>Mean</td>
<td>0.973</td>
<td>0.990</td>
<td>0.991</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>08.30 (05:00)</td>
<td>Mean</td>
<td>0.968</td>
<td>0.983</td>
<td>0.984</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>07.04 (15:00)</td>
<td>Mean</td>
<td>0.951</td>
<td>0.977</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.005</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>03.01 (15:30)</td>
<td>Mean</td>
<td>0.942</td>
<td>0.962</td>
<td>0.963</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Mean</td>
<td>0.959</td>
<td>0.978</td>
<td>0.980</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.013</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>OK</td>
<td>09.29 (05:00)</td>
<td>Mean</td>
<td>0.988</td>
<td>0.951</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.001</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>08.30 (05:00)</td>
<td>Mean</td>
<td>0.924</td>
<td>0.973</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>07.04 (15:00)</td>
<td>Mean</td>
<td>0.695</td>
<td>0.516</td>
<td>0.513</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.006</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>03.01 (15:30)</td>
<td>Mean</td>
<td>0.598</td>
<td>0.548</td>
<td>0.692</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.024</td>
<td>0.025</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Mean</td>
<td>0.801</td>
<td>0.747</td>
<td>0.780</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.160</td>
<td>0.215</td>
<td>0.189</td>
</tr>
<tr>
<td>UK</td>
<td>09.29 (05:00)</td>
<td>Mean</td>
<td>0.988</td>
<td>0.938</td>
<td>0.944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.001</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>08.30 (05:00)</td>
<td>Mean</td>
<td>0.948</td>
<td>0.989</td>
<td>0.971</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>07.04 (15:00)</td>
<td>Mean</td>
<td>0.661</td>
<td>0.475</td>
<td>0.521</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.009</td>
<td>0.018</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>03.01 (15:30)</td>
<td>Mean</td>
<td>0.635</td>
<td>0.735</td>
<td>0.690</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.016</td>
<td>0.012</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>Mean</td>
<td>0.808</td>
<td>0.784</td>
<td>0.782</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SD</td>
<td>0.161</td>
<td>0.202</td>
<td>0.186</td>
</tr>
</tbody>
</table>
from case to case. This seems to be related to the different rate of change in visibility with distance when fog occurs or not. However, in the case of the two Krigings, the level and consistency of objective analysis are different depending on the model used and the evaluation elements, so it is difficult to say that a specific model is better than others. And in all interpolation methods and powers (Variogram model), the level of interpolation is slightly lowered, in terms of mean and their standard deviation, compared to that of exactitude property test.

4. Discussions
In general, the objective analysis technique assumes that the data follow a normal distribution (Ro and Yoo, 2016). In this study, the normality of data used for objective analysis was examined to analyze the cause of the low level of objective analysis of the two kriging techniques (OK, UK). Fig. 7 shows the histogram distribution of fog and non-fog cases used for objective analysis. Excluding Fig. 7(d) case, the distribution of the three cases data does not follow the normal distribution at all. In order to quantitatively evaluate the distribution characteristics of the used visibility data, the Shapiro-Wilk’s normality test and statistical elements including the mean and standard deviation were analyzed. In the 95% confidence interval, the P-values were all significantly smaller than 0.05 regardless of the presence of fog, suggesting that none of these data follow the

Fig. 7. Histogram of visibility distribution: (a) 05:00 KST on September 29, 2019, (b) 05:00 KST on August 30, 2019, (c) 15:00 KST on July 4, 2019 and (d) 15:30 KST on March 1.
normal distribution (Not shown). This distribution characteristic of the visibility data used for objective analysis seems to have contributed to the low level of objective analysis of the two kriging techniques.

5. Conclusions

In this study, as a preliminary step to create a fusion data prototype with GK2A fog output and grid-type visibility data, the level of three objective techniques (IDW, OK, and UK) was evaluated with four sets of visibility data with diverse spatial variabilities. In this study, visibility data from 243 visibility meters operated by the KMA were used. In addition, in order to objectively evaluate the output level of the objective analysis techniques, sensitivity evaluation was performed on the major parameters (power, effective radius, and used variogram model) that influence the level of objective analysis. The level of objective analysis techniques was evaluated by the methods of exactitude property test and Jackknife method.

Regardless of the presence or absence of fog and major parameters (power, effective radius, and used variogram model), the level and consistency of objective analysis of IDW were found to be superior to that of OK and UK in all evaluation factors. Objective analysis techniques have different sensitivity to the selected parameters, the IDW is primarily more sensitive to powers, while both Kriging are more sensitive to the used variogram model. In IDW, when the power is 2, the level of objective analysis and consistency are the highest. However, the level and consistency of the two Kriging objective analyses are different depending on the case, models used and the evaluation elements, so it is difficult to say that a specific model is better than others.

In general, the objective analysis technique, in particular, Kriging, assumes that the data follow a normal distribution. The normality test of visibility data using the Shapiro-Wilk’s normality test and histogram distribution showed that most of the visibility data do not follow the normality distribution. The fact that the visibility data do not follow the normality distribution seems to have lowered the level of objective analysis of the two Kriging techniques. This suggests that it is very important to select an appropriate objective analysis technique and parameters through various sensitivity tests including the normality test of used data.

Acknowledgements

This research was carried out with the support of the KMI2018-06510 and KMI2020-01414 project of the Korea Meteorological Industry and Technology Institute. In addition, we would like to thank Professor No-Wook Park of the Department of Spatial Information Engineering at Inha University for providing knowledge related to Krigings.

References


