

Fast and operational gap filling in satellite-derived aerosol optical depths using statistical techniques

Kyunghwa Lee^{1b},^{a,*} Mijeong Kim,^a Myungje Choi^{1b},^{b,c} Jhoon Kim,^d
Yunsoo Choi,^e Jaehoon Jeong,^a Kyung-Jung Moon,^a and Sojin Lee^f

^aNational Institute of Environmental Research (NIER), Environmental Satellite Center,
Climate and Air Quality Research Department, Incheon, Republic of Korea

^bUniversity of Maryland, Baltimore County, Baltimore, Maryland, United States

^cNASA Goddard Space Flight Center, Greenbelt, Maryland, United States

^dYonsei University, Department of Atmospheric Sciences, Seoul, Republic of Korea

^eUniversity of Houston, Department of Earth and Atmospheric Sciences, Houston, Texas,
United States

^fThe Seoul Institute, Department of Safety and Environment Research, Seoul, Republic of Korea

Abstract. Satellite observations, used worldwide in the atmospheric sciences, are extremely useful for providing aerosol information within a wide spatial range. However, the coverage of aerosol data by satellite observations is sometimes of inferior quality because of the effects of surface reflectivity and clouds. To fill the gaps in aerosol optical depths (AODs) retrieved from geostationary ocean color imager observations, this study applies operational statistical techniques, including radial basis functions (RBFs) with four different weightings (i.e., linear, multi-quadratic, thin-plate, and inverse), Poisson, and ordinary Kriging. Based on computation time and accuracy of the individual gap-filling techniques, Poisson and the linear RBF are selected as the two best methods and then averaged with weights using one-dimensional weighted average (1D-WAVG) and two-dimensional weighted average (2D-WAVG) root mean square errors. All methods produce reliable results, yielding a correlation coefficient between 0.74 and 0.87 over the entire research domain. Out of the individual techniques, the Poisson, with an initial estimation from a zonal mean of AODs, is the most accurate with the lowest computational costs, even for a large number of missing pixels and most regions, excluding East China (EC). The Poisson's high bias over EC is compensated in 1D- and 2D-WAVGs by taking more accurate estimations of the linear RBF than those of the Poisson over the region. If we consider 1D- and 2D-WAVGs in our analysis, the highest correlation is obtained from the 2D-WAVG over all regions. Because of its reliability and fast computation time, applying the 2D-WAVG can be a good solution to provide spatial-temporal continuous aerosol information. In addition to air pollution studies, such as real-time air quality predictions, estimation of ground-level particulate matter concentrations, and other applications, the fast and operational gap-filling technique can also be expanded to remote sensing data obtained from satellite observations to provide helpful and useful information for the public. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JRS.16.044507](https://doi.org/10.1117/1.JRS.16.044507)]

Keywords: geostationary ocean color imager; satellite observations; gap-filling; statistics; aerosol optical depth; aerosols.

Paper 220321G received Jun. 2, 2022; accepted for publication Oct. 5, 2022; published online Oct. 25, 2022.

1 Introduction

Aerosols, which are comprised of solid or liquid particles suspended in air, influence visibility, public health, and the climate.¹ As the impact of aerosols is considerable in areas worldwide, researchers have devoted numerous efforts to quantifying aerosols using ground-based and airborne measurements and satellites. The satellite observations, state-of-the-art techniques used to

*Address all correspondence to Kyunghwa Lee, lk1515@gmail.com; nierkhlee@korea.kr

monitor the Earth's surface and its atmosphere, provide wide spatial coverage and applicability to several research fields. Specifically, the satellite observations over oceans are useful because of the lack of ground-based measurements over such regions.

Satellite observations are also capable of obtaining concentrations of aerosols in terms of the aerosol optical depth (AOD), but the AOD retrievals from the satellite observations are limited with regard to bright pixels, such as clouds, ice/snow, turbid water, or Sun glint. These bright pixels lead to a large portion of missing pixels in AOD data, owing to the low sensitivity of aerosols over these areas.

To overcome these limitations, a number of researchers have conducted studies to provide aerosol information for missing AOD pixels. Chen et al.² applied a simple statistical technique, the inverse distance weighting method, to reduce the rate of missing values from 87.91% to 13.83% in daily AODs retrieved from moderate resolution imaging spectroradiometer (MODIS) observations. This method, however, cannot provide complete coverage, which is typically achieved by assimilating satellite data with chemical transport model (CTM) simulations.³⁻⁶ Recent studies have applied artificial intelligence (AI) techniques to address the limitations of AOD data by training satellites with the CTM data. They have imputed AODs derived from MODIS observations by combining two techniques: random forest (a popular nonparametric machine learning algorithm) and lattice Kriging (a multiresolution Gaussian process model).⁷ To fill gaps in geostationary ocean color imager (GOCI) AODs using CTM simulations, Lops et al.⁸ proposed a novel deep learning technique, the partial convolutional neural network, which showed acceptable performance. However, including CTM simulations in the training process, as in the aforementioned studies, can increase computational costs, especially when they use the model to achieve high spatial resolution. In Ref. 9, the researchers filled gaps in MODIS AODs over the Sichuan Basin of southwest China by applying the random forest technique and then estimated PM_{2.5} (atmospheric particulate matter with an aerodynamic diameter of <2.5 μm).

Although the aforementioned techniques have yielded acceptable results, operational methods based on statistics are still efficient as they can quickly provide complete satellite AODs. Fast processing times for filling the gaps in AOD data can be crucial to city officials who are able to alert their citizens, regarding high concentrations of long-range transported aerosols in advance. The fast and operational gap-filling techniques can also lead to reliable air quality predictions by providing more accurate initial and boundary conditions for CTM simulations rather than using the default constant values built into the models. In addition, they can be used in a variety of other applications, such as estimations of ground-based PM concentrations from satellite AODs,¹⁰ the fusion of satellite data,¹¹ and studies of long-range transported aerosols.^{12,13}

To develop the fast gap-filling method, this study applied six operational techniques based on statistics. These are radial basis functions (RBFs) with four weighting functions, Poisson, and ordinary Kriging, to impute gaps in GOCI AODs. These methods are considered useful because of their overall performance and computational costs, as well as their ability to bypass the expensive computational costs by conducting CTM simulations, especially at high spatial resolutions. In addition, AODs (level 2 products) were used in this study because the operational methods can fill gaps in data, having physical homogeneity more effectively than using level 1 products of satellites that contain inhomogeneous spectroscopic information. We then evaluated the performance of the gap-filling results by comparing them with reference data of this study. From this analysis, the two best methods were averaged with weights based on their errors and compared with the individual gap-filling techniques. We expect that the best method assessed in this study will provide aerosol information in near-real time by acquiring continuous gap-filled AODs in time and space and be expanded to various research fields using remote sensing data from satellites.

Section 2 of this paper describes details in the GOCI data, the gap-filling techniques, and the statistical evaluation metrics. Section 3 discusses the results of the complete gap-filled AODs and their performance, and Sec. 4 summarizes and concludes this study.

2 Materials and Methods

To fill the gaps in GOCI AODs and to evaluate various gap-filling techniques, this study used the following six steps: (1) monthly GOCI AODs were averaged from January to December 2019 to

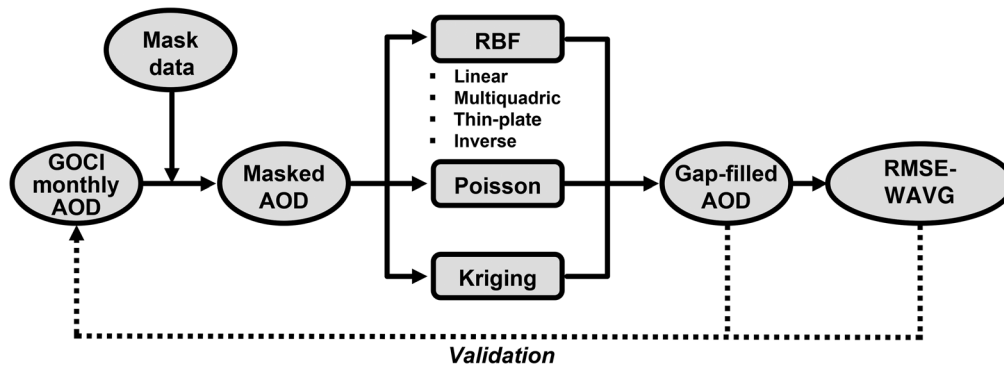


Fig. 1 Schematic illustration of the process for filling gaps in AOD data. The dashed line marks the validation process used in this study.

produce reference datasets that included fewer missing pixels than individual hourly AODs; (2) mask data were produced based on the randomly chosen and daily averaged unretrieved GOCI aerosol pixels, during the research period; (3) using the mask product, we masked out the monthly averaged AODs; (4) applying six gap-filling techniques, we filled the gaps in the masked AODs; (5) the gap-filled AODs were evaluated by comparing them to the monthly averaged AODs over masked pixels; and (6) the two best methods found by conducting the fifth step were averaged with weights based on their errors to compensate for the two methods and then compared with the six individual gap-filling techniques.

Figure 1 shows a schematic illustration of the gap-filling processes. We gridded all of the data used in this study onto $0.1 \text{ deg} \times 0.1 \text{ deg}$ longitudinal and latitudinal grids before conducting interpolation using the six aforementioned techniques. The research domain analyzed in this study was 113°E to 148°E and 24°N to 48°N . The following sections address details in the GOCI data, gap-filling techniques, and evaluation metrics.

2.1 GOCI Data

The GOCI, onboard the Korean geostationary satellite and the communication, ocean, and meteorological satellite, was launched on June 26, 2010 as the first ocean color observation instrument placed into geostationary Earth orbit (GEO). GOCI has six visible (i.e., 412, 443, 490, 555, 660, and 680 nm) and two near-infrared (i.e., 745 and 865 nm) spectral channels and $0.5 \times 0.5 \text{ km}^2$ spatial resolution with a horizontal coverage of $2500 \times 2500 \text{ km}^2$ over East Asia. Hourly AODs at 550 nm were retrieved from GOCI observations via the Yonsei aerosol retrieval version 2 algorithm¹⁴ with a spatial resolution of $6 \times 6 \text{ km}^2$ from 00:30 to 07:30 coordinated universal time. The GOCI mission was officially completed in March 2021. In this study, we analyzed GOCI AODs from December 2015 to March 2021 to count the missing rate at each grid pixel regionally and seasonally.

2.2 AEROSOL ROBOTIC NETWORK DATA

The AEROSOL ROBOTIC NETWORK (AERONET) AODs were used for independent evaluation of the gap-filling techniques. The recent version 3 level 2.0 (cloud screened and quality assured) AOD data of AERONET, a federated global ground-based remote sensing network,¹⁵ were downloaded from Ref. 16. Because AERONET provides AODs at different wavelengths depending on the observation site, AODs at 550 nm were computed using the AODs at 500 nm and the Ångström exponent obtained with AODs at 440 and 675 nm in this study.

2.3 Gap-Filling Techniques

Figure 2 shows the statistical gap-filling methods applied to fill gaps in the AOD data derived from GOCI observations. To calculate values at the target missing pixels shown in Fig. 2,

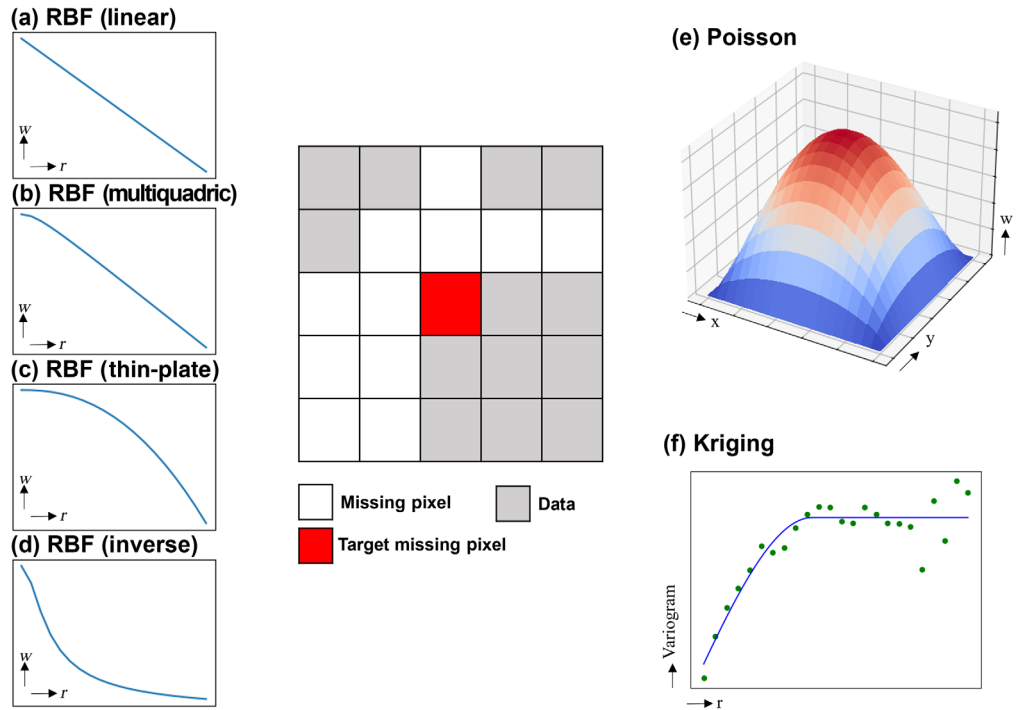


Fig. 2 Gap-filling techniques applied in this study. To fill the gaps in the target missing pixel, weights (w) based on the distance (r) between the target missing pixel and the data were determined based on the different techniques: (a)–(f) the linear, multiquadric, thin-plate, and inverse RBFs, Poisson, and Kriging, respectively.

weights, according to the distance between the target pixel and the AOD pixel, were determined using the equations presented in the following sections.

2.3.1 Radial basis functions

The RBFs of SciPy¹⁷ in Python¹⁸ were applied to fill gaps in the AODs. The RBF is a real-valued function in N -dimensional space whose data at x can be explained by the following equation

$$r = \|x - c\|, \tag{1}$$

where c is the center of the RBF. Among the various RBF functions, the following four functions based on distance (r) are commonly used [see Figs. 2(a)–2(d)], so we applied them to produce the missing pixel data in this study

$$-r, \tag{2}$$

$$-\text{sqrt}(1 + r^2), \tag{3}$$

$$-r^2 \times \log(r), \tag{4}$$

$$\frac{1}{\text{sqrt}(1 + r^2)}. \tag{5}$$

The RBFs, with each function defined in Eqs. (2)–(5), are referred to as the linear, multiquadric, thin-plate, and inverse RBFs, respectively.

2.3.2 Poisson

In addition to using the RBF methods, we filled the missing pixels with values derived using relaxation to solve Poisson’s equation [see Fig. 2(e)], which is an elliptic partial differential equation with an iterative relaxation scheme.¹⁹

We applied a “Poisson grid fill” function in the National Center for Atmospheric Research command language (NCL)²⁰ in this study and set an initial guess as a start off the zonal averages of AODs to achieve more accurate performance rather than using a default constant of 0. Retrieved AODs from GOCI were used as boundary conditions. For every interior grid point, the quantity is calculated using the following equation:

$$\varphi_{i,j}^* = \frac{1}{4}(\varphi_{i+1,j} + \varphi_{i-1,j} + \varphi_{i,j+1} + \varphi_{i,j-1} - \sigma(x_i, y_i)\delta x^2), \tag{6}$$

where $\varphi_{i,j}$ and $\varphi_{i,j}^*$ represent new and old initial guesses, respectively, and $\sigma(x_i, y_i)$ is the source term. Through iteration, this process is repeated until the difference between the two approximations is less than the tolerance that is used to end relaxation before reaching the maximum number of iterations. In our experiments, we set the maximum number of iterations used by relaxation, the tolerance, and the relaxation constants 2000, 0.001, and 0.6, respectively.

2.3.3 Kriging

Kriging has been widely used in geosciences to estimate values over a continuous spatial field from neighboring data points. As it has the lowest computational cost among various Kriging algorithms, ordinary Kriging, one of the most commonly used Kriging methods,^{21,22} has been applied in this study. As the number of data points to be interpolated is huge, ordinary Kriging is more effective in overcoming memory limitations than other Kriging methods. Indeed, ordinary Kriging is defined as the best linear unbiased estimator by means of which interpolated values are estimated by minimizing the variance of errors. Ordinary Kriging is conducted based on the stationarity assumption that the mean and variance of data are constant across a spatial field and its estimates are weighted by linear combinations of input data. Hereafter, ordinary Kriging is referred to as “Kriging” in this study.

Different from the previous gap-filling techniques, Kriging determines weight by considering the spatial correlation between sampled data points to interpolate the values in the spatial field with the following semivariogram equation:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} (z(u_\alpha) - z(u_\alpha + h))^2, \tag{7}$$

where $\gamma(h)$ is a semivariance, $z(u_\alpha)$ represents a value at a target sampled point, $z(u_\alpha + h)$ implies the value of the neighbor at distance h , and N is the number of data points included in distance h . Among the various Kriging toolkits in different scientific languages, PyKrige^{23,24} in Python was applied in this study.

2.4 Evaluation Metrics

Statistical evaluation metrics that are used to evaluate the accuracy of the gap-filling techniques are as follows:

$$\text{Correlation coefficient } (R) = \frac{1}{(n-1)} \sum_1^n \left(\left(\frac{R_{-f} - \overline{R_{-f}}}{\sigma_{R_{-f}}} \right) \left(\frac{C - \overline{C}}{\sigma_C} \right) \right), \tag{8}$$

$$\text{Root mean square error (RMSE)} = \sqrt{\frac{\sum_1^n (C - R_{-f})^2}{n}}, \tag{9}$$

$$\text{Mean bias (MB)} = \frac{1}{n} \sum_1^n (C - R_f), \quad (10)$$

where C denotes gap-filled AODs produced using different gap-filling techniques, and R_f represents the reference data. In addition, n and σ indicate the number of datasets and the standard deviation, respectively. The overbars of R and C refer to the arithmetic mean of the data.

2.5 Root Mean Square Error-Weighted Average

Among the six gap-filling techniques, the two best methods were selected based on the computational speed and evaluation metrics and then averaged using weights with RMSE values, as written in Eq. (11), where l indicates each gap-filling technique and k is the number of methods used in averages. RMSE weighted averages (WAVGs) were calculated based on 1D-RMSE, that is, constant values for all pixels, or 2D-RMSE, that is, different weights at each pixel. Henceforth, the WAVGs using 1D-RMSE and 2D-RMSE are referred to as “1D-WAVG” and “2D-WAVG,” respectively

$$\text{WAVG} = \frac{\sum_1^k \left(\frac{C_l^2}{\text{RMSE}_l^2} \right)}{\sum_1^k \frac{1}{\text{RMSE}_l^2}}, \quad (11)$$

where C denotes the gap-filled AODs produced using different gap-filling techniques (l).

3 Results and Discussion

We analyzed gaps in monthly averaged AODs derived from GOCI observations in terms of the season and the region and then filled them by applying the gap-filling techniques. Performances of the gap-filling techniques were then evaluated via statistical metrics. In addition, all the gap-filling techniques were applied to the daily AODs to estimate the missing values, and the results were compared with the AERONET daily AODs. The detailed results of the analyses are described in the following sections.

3.1 Analysis of AOD Coverage

Although satellite data are useful in numerous ways, a large number of missing pixels exist in satellite aerosol products, owing to the limitations in retrievals over bright surfaces, such as desert, ice, turbid water, and/or Sun glint, and under those blocked by the presence of clouds and fogs. Figure 3 shows the spatial distributions of an average coverage of GOCI AODs for (a) the entire year, (b) spring (March to May), (c) summer (June to August), (d) fall (September to November), and (e) winter (December to February) from December 2015 to March 2021. The highest coverage (>50%) was found over the Bohai Sea and northern part of the East Sea of Korea during the spring. Overall, the data coverage over the oceanic area was relatively low at lower latitudes due to the frequent occurrence of clouds and Sun-glint over the area.

Table 1 lists the pixel coverage of the GOCI AOD data over the entire research domain and four regions—D01 Seoul metropolitan area (SMA), D02 South Korea (SK), D03 Yellow Sea (YS), and D04 East China (EC). Almost 22.3% pixels of the data remain for the entire domain and period of the GOCI observations. Seasonally, the data coverage was high as follows: spring (27.9%) > summer (23.7%) > fall (23.2%) > winter (22.3%). During winter, loss of some GOCI AODs occurred due to the increase in bright surfaces of snow and ice over Manchuria and the higher solar zenith angle where aerosol retrieval is limited. Filling the gaps in AODs during the winter is imperative as the long-range transported high AODs were frequently observed over East Asia because of increased burning of fossil fuels and weather conditions. Regionally, the percentages of data coverage over the four regions were high in the following order: SK (28.6%) > YS (27.0%) > EC (25.9%) > SMA (23.2%). If we closely examine the regions, the highest pixel coverage of 39.5% was observed over SK during the spring and the lowest, 17.2%, over

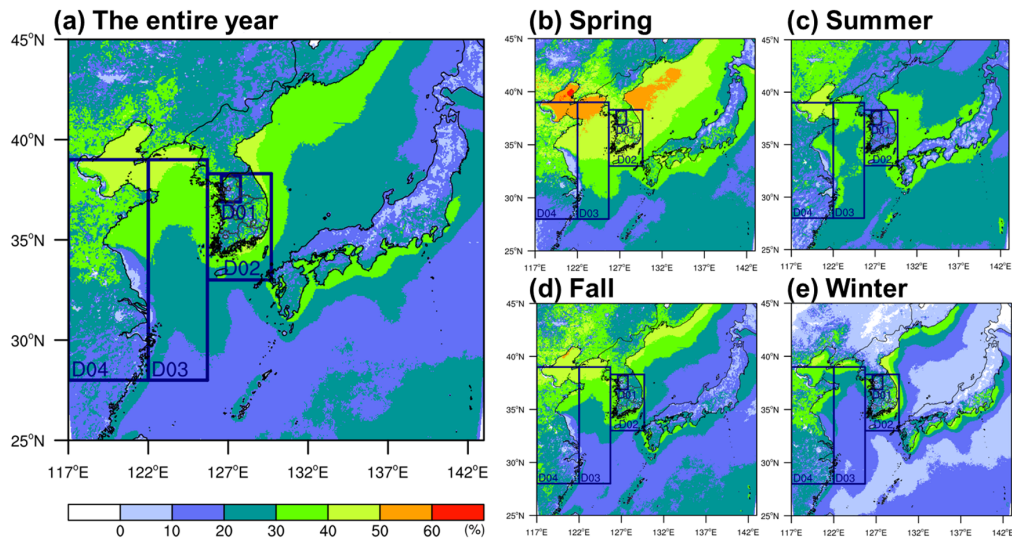


Fig. 3 Average coverage of GOCI AOD data for (a) the entire year, (b) the spring, (c) the summer, (d) the fall, and (e) the winter from December 2015 to March 2021. The darker blue boxes show the domains applied in this study. D01, D02, D03, and D04 indicate the SMA, SK, YS, and EC, respectively.

Table 1 Average pixel coverage of GOCI AODs from December 2015 to March 2021 in percentages (%). D01, D02, D03, and D04 indicate regions of the SMA, SK, YS, and EC, respectively.

	Spring	Summer	Fall	Winter	Entire year
D01 (SMA)	34.8	17.0	21.5	18.3	23.2
D02 (SK)	39.5	24.7	26.4	22.9	28.6
D03 (YS)	37.4	28.5	24.9	17.2	27.0
D04 (EC)	30.4	22.3	26.4	24.1	25.9
Entire domain	27.9	23.7	23.2	13.5	22.3

YS during the winter. Based on these results, filling the gaps in AODs over YS can be extremely useful in providing information for a deeper understanding of long-range transported aerosols from abroad to SK. Considering the frequent sampling of GEO instruments, this issue should be more significant for low Earth orbit instruments.

3.2 Gap-Filling Results

3.2.1 Computational time

Because the aim of this study is to investigate the fast and operational gap-filling technique, the computational time of the individual gap-filling processes is equally as important as their accuracy. Using identical computer resources as the Intel(R) Xeon(R) Platinum 8280 system per socket clocked at 2.70 GHz, the gap-filling techniques resulted in different computation times, as shown in Table 2. The fastest computation, <1 min, was obtained by the Poisson algorithm, whereas the slowest computation time ~1 h was acquired by the Kriging algorithm. A considerable amount of computation time is required via Kriging to calculate the spatial correlation between sampled data points and target pixels, which requires the use of a large amount of computer resources. It is important to note that Kriging’s computational time can be decreased if parallel computing architectures, such as the message passing interface, using the graphics processing unit are applied because this study compares gap-filling techniques, under the same

Table 2 An average of computational times taken to fill gaps in AODs.

	RBF				Poisson	Kriging	1D-WAVG	2D-WAVG
	Linear	Multiquadric	Thin-plate	Inverse				
Computational time	5 min 11 s	9 min 10 s	7 min 25 s	7 min 40 s	55 s	55 min 5 s	3 s	21 s

conditions, using a single core of the central processing unit. After conducting the six gap-filling methods, calculation of the 1D- and 2D-WAVGs only took 3 and 21 s, respectively, suggesting the addition of the averaging processes seems to be a feasible approach.

3.2.2 Overall statistical analysis

To evaluate the performance of the gap-filling techniques, we conducted a statistical analysis by comparing the gap-filled AODs to the reference data. Figure 4 shows the scatter plots of the

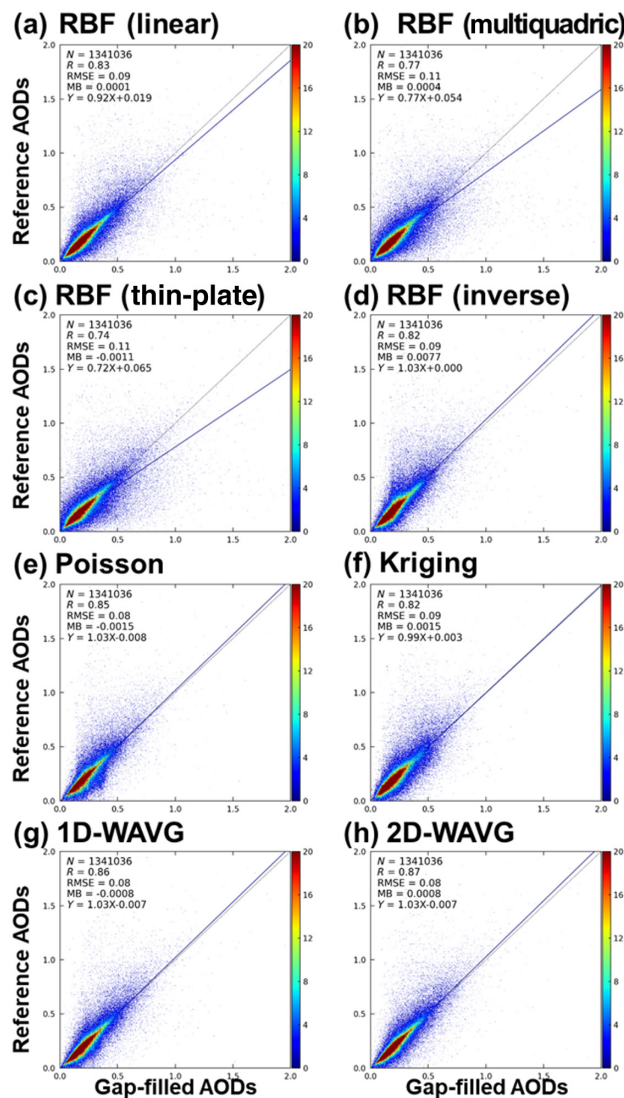


Fig. 4 Statistical analysis of (a)–(f) gap-filled AODs when applying the four RBFs, Poisson, and Kriging techniques, respectively. Weighted averages of the two best methods (Poisson and the linear RBF) using (g) 1D- and (h) 2D-RMSE are also shown. Values on the x-axis show the gap-filled AODs and values on the y-axis indicate the reference AODs. The density of the data is visualized by coloring the markers.

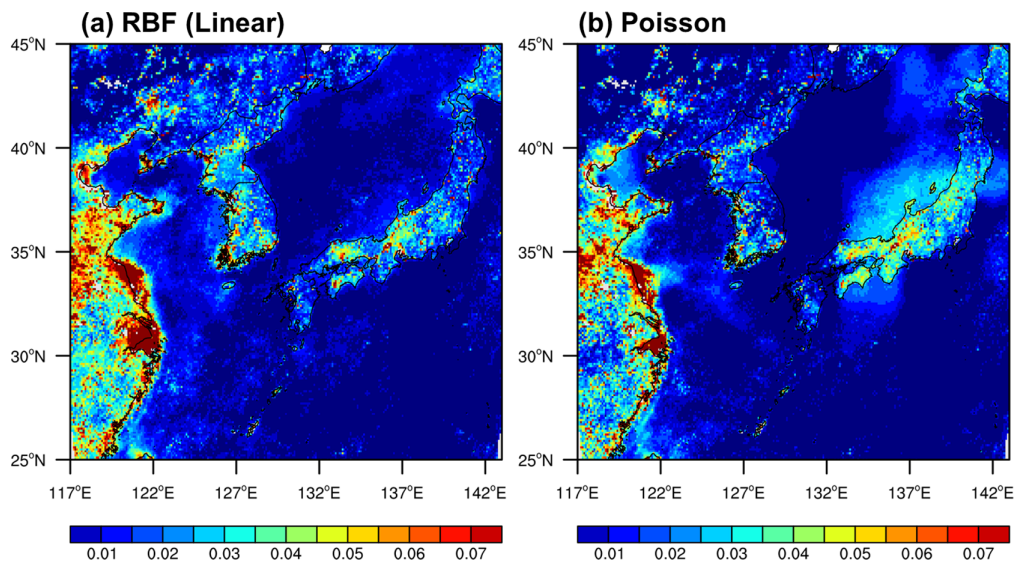


Fig. 5 Spatial distributions of RMSE: (a) the linear RBF and (b) Poisson.

gap-filled AODs compared with the original AODs with statistical indices, such as Pearson's correlation coefficient (R), the RMSE, and the MB. We identified relatively poorer performance when the multiquadric ($R = 0.77$ and $\text{RMSE} = 0.11$) and thin-plate ($R = 0.74$ and $\text{RMSE} = 0.01$) RBFs were used to fill the gaps in the AODs than when using the linear and inverse RBFs, Poisson, and Kriging, which showed similar acceptable performances. The highest accuracy was obtained by Poisson in terms of the R (0.85) and RMSE (0.08) values. In case of MB, the linear RBF showed the best results of 0.0001 (almost zero).

Based on those statistical results and computational costs (see Sec. 3.2.1), we selected Poisson and the linear RBF as the two best methods and then calculated RMSE WAVGs using them. For the 1D-WAVG, RMSE values of 0.08 and 0.09 were applied for Poisson and the linear RBF, respectively, as shown in Fig. 4. The WAVGs for the 2D-WAVG were calculated using RMSE distributions (see Fig. 5) of Poisson and the linear RBF obtained by comparing the gap-filled AODs to the reference data. As shown in Fig. 5, Poisson yielded less RMSE values over Korea and southeast China and larger RMSE values near and over Japan and the Bohai Sea, than those from the linear RBF. The different RMSE values of the two best gap-filling techniques over the regions indicate that the 2D-WAVG may provide more reliable results compared with the 1D-WAVG and individual gap-filling techniques by complementing the two methods.

Furthermore, our expectations turned out to be true with the highest R of 0.87 from the 2D-WAVG among the eight gap-filling results, including the 1D-WAVG, as shown in Fig. 4. As well as the 2D-WAVG, the 1D-WAVG yielded a better result in terms of R (0.86) and MB (-0.0008) than those from Poisson. If we look into the spatial distributions of R from all eight gap-filling results (see Fig. 6), we can see that the 2D-RMSE values worked well to provide reliable gap-filled AODs. Low R over southeast China and the oceans near Japan and Bohai Sea from Poisson were improved after averaging Poisson and the linear RBF with RMSE weights. Compared with the linear RBF, the correlations of the RMSE WAVGs increased over SK and EC. The high correlations obtained over EC from the 1D- and the 2D-WAVGs can play a significant role to capture occurrences of high concentrations of aerosols, which can travel long distances and affect neighboring countries.

Figure 7 shows the spatial distributions of the monthly averaged AODs before and after conducting gap-filling for June 2019. Here, we filled in the missing pixels of the target dataset [see Fig. 7(a)] by conducting statistical computations via the RBFs [see Figs. 7(c)–7(f)], Poisson [see Fig. 7(g)], and Kriging [see Fig. 7(h)]. The results of 1D- and 2D-WAVGs are also shown in Figs. 7(i) and 7(j), respectively. Despite the large portion of missing pixels, all gap-filled AODs shown in Figs. 7(c)–7(j) were well matched with the original AODs [i.e., reference data, in Fig. 7(b)]. In particular, similar distributions of high AODs over EC were adequately captured by the gap-filling techniques. In the case of thin-plate RBF, it overestimated the AODs over EC

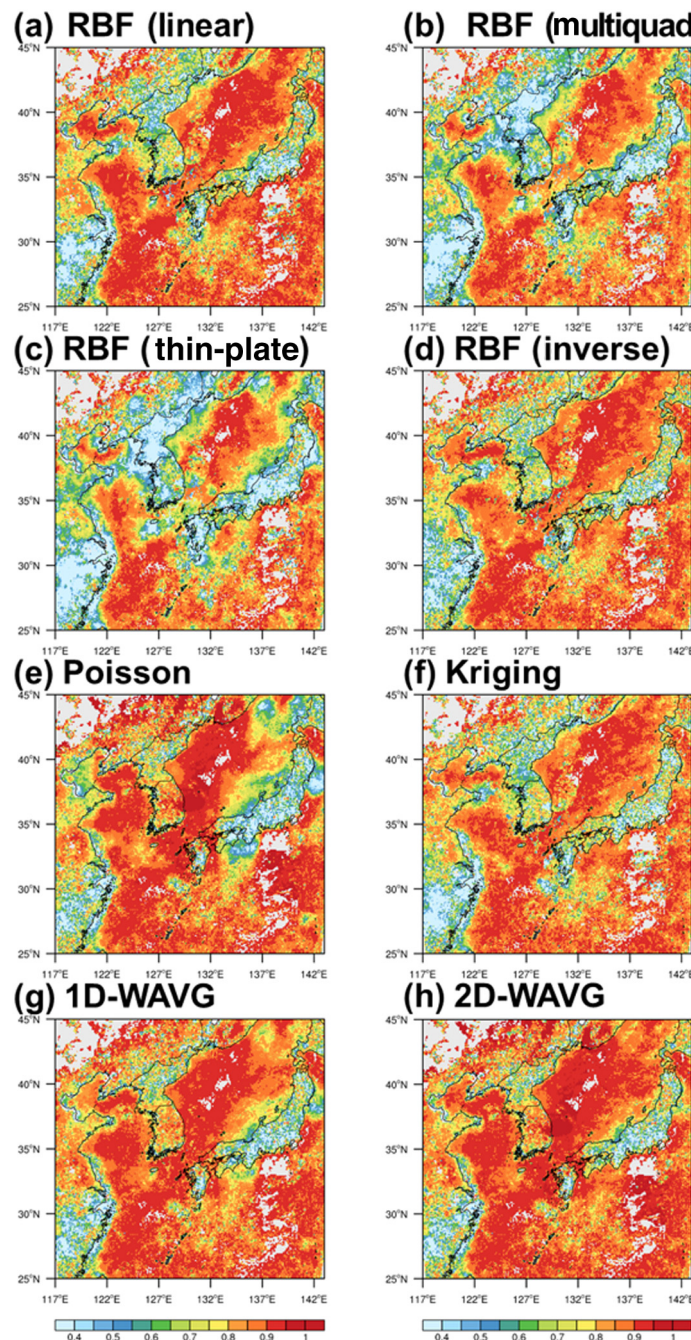


Fig. 6 Spatial distributions of correlation coefficient (R): (a)–(f) gap-filled AODs by applying the linear, multiquadric, thin-plate, and inverse RBFs, Poisson, and Kriging techniques, respectively, (g) 1D- and (h) 2D-RMSE WAVGs using the linear RBF and Poisson.

because it estimates gap-filling data using a weight function based on a plane, not on a specific point. The AOD distributions of Kriging were similar to the reference AODs but tended to be more smoothed out than those from other imputation results. This is because Kriging forces the original values to be data corrected by removing the measurement error while interpolation is carried out, because a nugget is considered a measurement error. The high gap-filled AODs can play an important role in providing reliable aerosol information to citizens and improving the accuracy of air quality prediction models by serving as their initial fields. All the methods appear to have limitations in capturing every detail of the features in the reference AODs. However, most of the AOD distribution features were well preserved for the cases in Fig. 7.

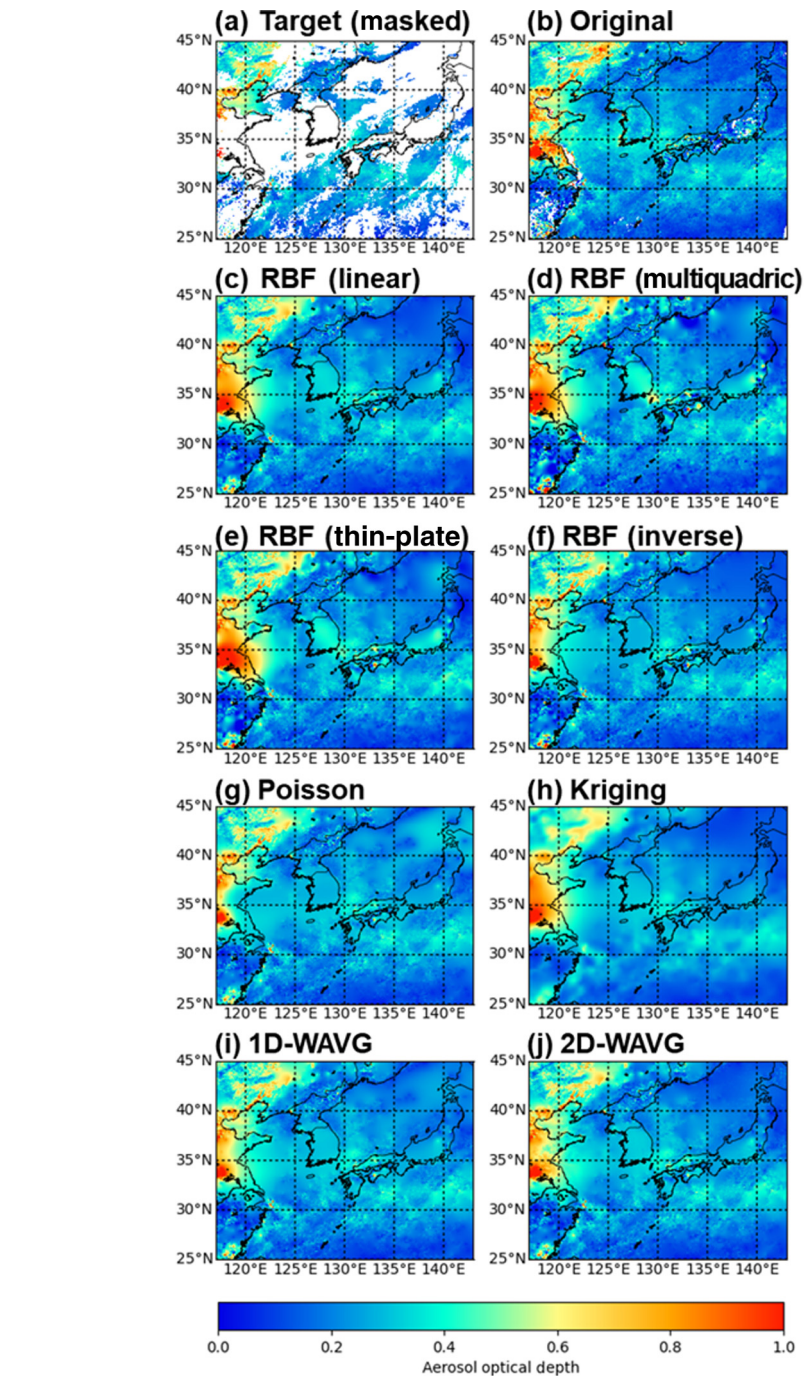


Fig. 7 Spatial distributions of (a) the masked target with clouds; (b) the original (reference); (c)–(h) the gap-filled monthly averaged AODs when applying the four RBFs, Poisson, and Kriging techniques; and (i) and (j) 1D- and 2D-RMSE WAVGs for June 2019.

3.2.3 Regional statistical analysis

To evaluate the regional dependency of the gap-filling techniques, the results were analyzed according to the four domains of D01 (SMA), D02 (SK), D03 (YS), and D04 (EC), as depicted in Fig. 3. Figure 8 and Table 3 show statistical indicators based on the defined regions. Among the six individual gap-filling techniques, Poisson showed the best correlation for all the regions analyzed in this study. In particular, it showed such a good performance over the SMA ($R = 0.78$ and $RMSE = 0.12$) compared with that of the other gap-filling techniques, which demonstrated poorer results over the region, yielding R 's < 0.6 and $RMSE$'s > 0.15 . A similar tendency

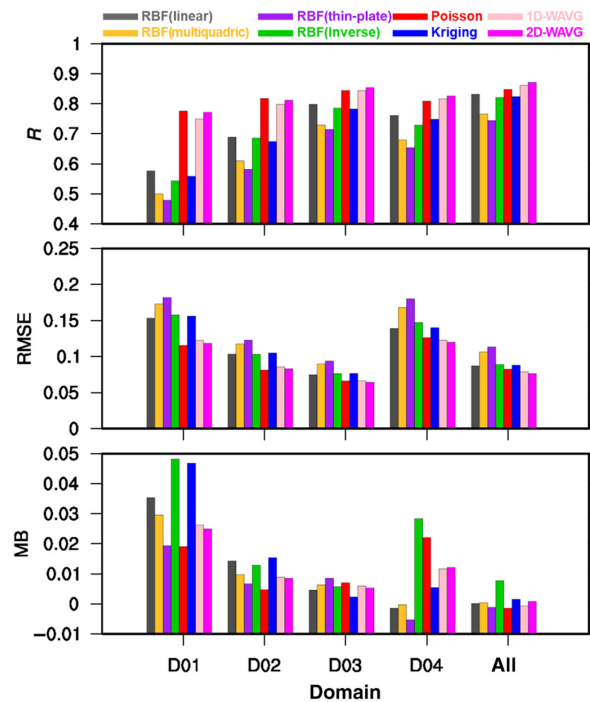


Fig. 8 Variations in the three performance metrics (i.e., the R , the RMSE, and the MB) over the four study regions—D01 (SMA), D02 (SK), D03 (YS), and D04 (EC)—and the entire domain.

Table 3 Statistical analysis of the performance evaluation of the gap-filling techniques via Pearson’s correlation coefficient (R), the RMSE, and the MB for the entire study period and the four regions applied in this study. D01, D02, D03, and D04 indicate the SMA, SK, YS, and EC, respectively.

		RBF				Poisson	Kriging	1D-WAVG	2D-WAVG
		RBF							
		Linear	Multiquadric	Thin-plate	Inverse				
D01 (SMA)	N					4384			
	R	0.58	0.50	0.48	0.54	0.78	0.56	0.75	0.77
	RMSE	0.15	0.17	0.18	0.16	0.12	0.16	0.12	0.12
	MB	0.035	0.030	0.020	0.048	0.019	0.047	0.026	0.025
D02 (SK)	N					71527			
	R	0.69	0.61	0.58	0.69	0.82	0.67	0.80	0.81
	RMSE	0.10	0.12	0.12	0.10	0.08	0.10	0.09	0.08
	MB	0.014	0.0097	0.0067	0.013	0.0047	0.015	0.0089	0.0085
D03 (YS)	N					135611			
	R	0.80	0.73	0.71	0.79	0.84	0.78	0.84	0.85
	RMSE	0.07	0.09	0.09	0.08	0.07	0.08	0.07	0.06
	MB	0.0046	0.0063	0.0085	0.0057	0.0070	0.0023	0.0059	0.0053
D04 (EC)	N					268640			
	R	0.76	0.68	0.65	0.73	0.81	0.75	0.82	0.83
	RMSE	0.14	0.17	0.18	0.15	0.13	0.14	0.12	0.12
	MB	-0.0015	-0.0003	-0.0053	0.028	0.022	0.0054	0.012	0.012

appeared over the SK. The results for EC, where high concentrations of aerosols were observed, were the most diverse, and over this area, Poisson showed the highest accuracy among the gap-filling techniques in terms of the R (0.81) and RMSE (0.13) values; however, Poisson also produced a relatively high MB of 0.022. Out of the four RBF methods, the linear RBF was found to be the best solution in terms of R and RMSE and yielded smaller MB of -0.0015 compared with Poisson.

Through weighted averaging of Poisson and the linear RBF (i.e., 1D- and 2D-WAVGs), R increased and RMSE became similar or smaller than those of Poisson over all the regions. Over EC, the MB of 1D- and 2D-WAVGs was 0.012, which is smaller than Poisson’s MB of 0.022. A large discrepancy of the statistical evaluation metrics between the eight gap-filling results was observed over SMA and SK, and good results were obtained from Poisson, 1D-WAVG, and 2D-WAVG across the regions. As a whole, the 2D-WAVG was the most accurate in terms of R and RMSE relative to the six individual methods and the 1D-WAVG. This indicates that taking into account the errors on each pixel of the average AODs significantly contributed to the reliable gap-filled AODs by compensating for the limitations of individual techniques in Poisson and the linear RBF.

3.2.4 Pixel statistical analysis

Here, we evaluate the results of the gap-filling techniques over SK and EC, according to the missing pixel rates of data, that is, the percentage of missing data in each pixel, as shown in Fig. 9. Apart from Poisson, 1D-WAVG, and 2D-WAVG, the accuracy of the gap-filled AODs tended to dramatically decrease as the rate of missing data increased. In contrast, Poisson, 1D-WAVG, and 2D-WAVG yielded the highest R , the lowest RMSE, and almost 0 MB for large missing pixel rates from 50% to 100%, respectively. Because aerosols affect cloud formation by serving as cloud condensation or ice nuclei,¹ many AOD pixels in satellite data are frequently contaminated by clouds,

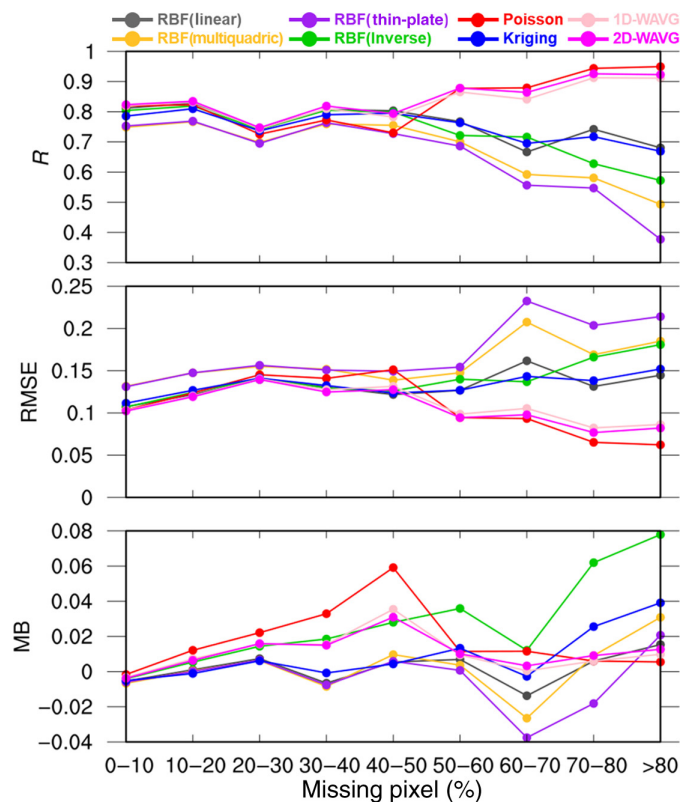


Fig. 9 Variations among the three performance metrics of the Pearson’s correlation coefficient (R), the RMSE, and the MB with respect to the percentage of missing pixels in the GOCI AOD data over D02 (SK) and D04 (EC).

leading to a large number of missing pixels in AOD data. The good performance of Poisson for the large missing pixel rates indicates that an initial guess using the zonal mean in solving Poisson's equation produced reliable gap-filled AODs, especially when we estimate AODs over either the SK or EC region where higher AODs are often observed than those over oceans. By considering the reliable AODs obtained from Poisson, 1D-WAVG and 2D-WAVG also exhibited good performance. However, additional studies over a longer period of time are needed to better understand the tendency with large missing pixels because this study did not cover all cases observed from GOCL.

For the rates of missing pixels <50%, the thin-plate RBF was the least accurate in terms of *R* and RMSE, whereas the Poisson yielded the highest MB even though Poisson's *R* and RMSE

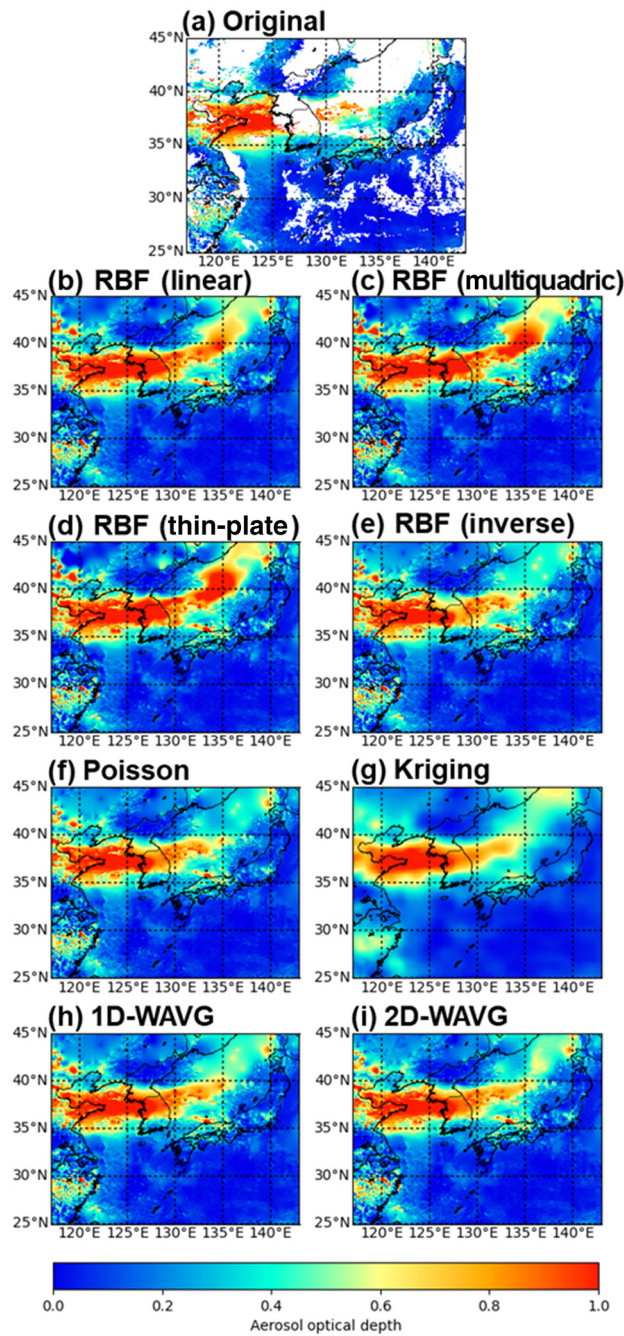


Fig. 10 Spatial distributions of (a) the original; (b)–(g) gap-filled daily averaged AODs following application of the four RBF, Poisson, and Kriging techniques; and (h) and (i) 1D- and 2D-RMSE WAVGs for April 18, 2019.

were comparable to other results. As a result of RMSE WAVG, MB decreased to a level similar to the inverse RBF for the missing rates <50% and became close to zero for the missing rates >50%. In addition, slightly better results were obtained with the 2D-WAVG than with the 1D-WAVG. Based on our analysis of the missing rates, the 2D-WAVG still appears to be the best choice in terms of accuracy.

3.3 Applications

The gap-filling techniques were applied to estimate the daily AODs over missing pixels in 2019. Figure 10 shows the spatial distributions of AODs on April 18, 2019. High AODs were estimated

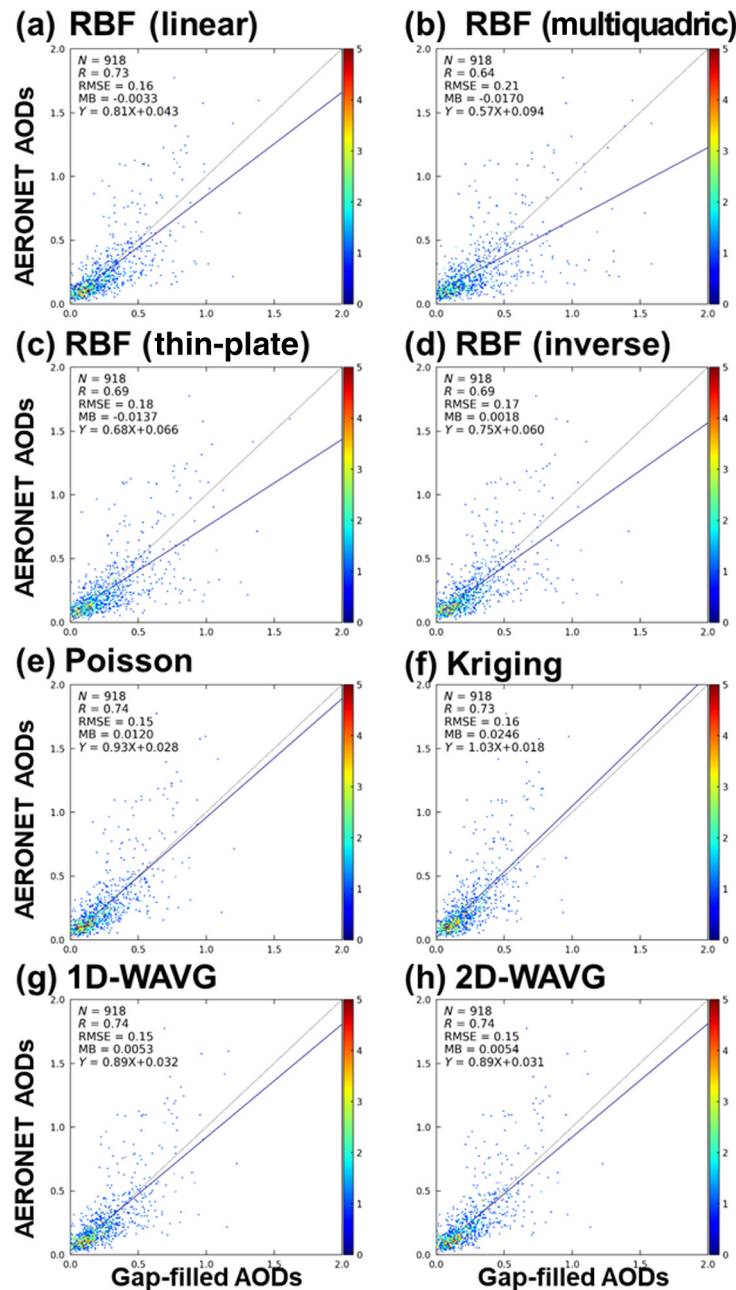


Fig. 11 Statistical analysis of (a)–(f) gap-filled daily AODs following application of the four RBF, Poisson, and Kriging techniques, respectively. The WAVGs of the two best methods (Poisson and the linear RBF) using (g) 1D- and (h) 2D-RMSEs are also shown. Values on the x-axis show the gap-filled daily AODs, and values on the y-axis show the AERONET daily AODs. The density of the data is visualized by the colored markers.

over SK and the East Sea from the gap-filling results. As shown in Fig. 7, the thin-plate RBF technique estimated the most intense AODs over a wide range. The accuracy of the filled daily AODs over missing pixels was evaluated by comparing them with the AERONET data for independent validation (see Fig. 11), in addition to the evaluation conducted on a monthly scale. The highest R of 0.74 was obtained when Poisson, 1D-WAVG, and 2D-WAVG were applied. In terms of MB, the linear RBF produced the lowest value of -0.003 , as also found in Fig. 4. Note that 1D- and 2D-WAVGs effectively compensated for the linear RBF and Poisson by yielding lower MB than those of Poisson and higher R than those of the linear RBF. Because AERONET screened out cloud effects in its data, our results indicate that there are masked AOD pixels in satellite data and gap-filling techniques performed acceptably on those pixels.

4 Conclusion

In this study, we used mask data to create gaps in monthly averaged GOCI AODs in 2019 and then filled the gaps using the following six operational statistical methods: linear, multiquadric, thin-plate, and inverse RBFs, Poisson, and ordinary Kriging. Among the six techniques, the two best methods (Poisson and the linear RBF) were selected in terms of computational time and accuracy and then averaged using weights based on 1D- and 2D-RMSE values. We refer to the 1D- and 2D-RMSE WAVGs as 1D-WAVG and 2D-WAVG, respectively. By comparing the gap-filled AODs to the reference data (i.e., monthly averaged GOCI AODs before masking of missing pixels), we evaluated the accuracy of the eight gap-filling results with statistical evaluation metrics, regarding regions and missing rates of pixels. The gap-filling techniques were also applied to fill gaps in the daily AODs, and the results were evaluated by comparing them with the AERONET AODs.

By analyzing the gap-filling results to estimate AODs for missing pixels using statistical evaluation metrics, we found that the Poisson, the linear RBF, and Kriging were the most accurate of the six techniques investigated. Among the three aforementioned methods, the Poisson yielded the best results for all the regions, including the YS with regard to the R and the RMSE. The reliable gap-filled AODs over the YS can be used for a better understanding in long-range transported aerosols from abroad to SK. However, over EC where high AODs can be frequently observed, bias of Poisson was larger than those of other techniques, whereas the linear RBF showed almost zero bias. The 1D- and 2D-WAVGs showed the highest correlation over the whole domain and the four regions (D01 to D04), in that the WAVG effectively reduced errors found from Poisson and the linear RBF.

Upon examining the accuracy of the gap-filling techniques with respect to the missing pixel rates of the GOCI AODs, we found that Poisson demonstrated extreme accuracy, particularly for large missing pixel rates ($>50\%$), whereas other techniques showed poorer accuracy as the number of missing pixels increased. In case of the missing pixel rates $<50\%$; however, MB of Poisson was larger than those from other methods. The 1D- and 2D-WAVGs effectively reduced the bias found from Poisson for large missing pixel rates ($<50\%$) and from the linear RBF for small missing pixel rates ($>50\%$).

According to the regional and pixel-based monthly analysis and daily results, overall the 2D-WAVG showed the best performance when considering the different errors by compensating for the two methods, namely Poisson and the linear RBF. Because Poisson and the linear RBF produced the fastest estimates with the same computing environment, and the 2D-RMSE WAVG taking only 21 s, conducting the 2D-WAVG can be a good solution and play a significant role in providing the public with continuous spatial–temporal aerosol information.

The fast gap-filling techniques of this study can be applied in remote sensing data, having missing pixels, e.g., forests, soils, land, etc. Particularly for air pollution studies, gap-filled AODs can be utilized to improve the accuracy of air quality predictions by serving as reliable initial and boundary conditions of the models, estimating ground-level PM concentrations, combining various satellite data, analyzing long-range transported aerosols, and so on.

As ongoing and future work, we try to expand the gap-filling techniques in the following two ways: (1) using AODs retrieved from two payloads onboard the GEO-KOMPSAT-2B satellite, which was recently launched on February 2020, i.e., GOCI-II^{25,26} and the Geostationary

Environment Monitoring Spectrometer (GEMS)^{27,28} and (2) developing a gap-filling technique by combining our operational methods with AI techniques.

Acknowledgments

This work was supported by a grant from the National Institute of Environment Research funded by the Ministry of Environment of the Republic of Korea (Grant No. NIER-2022-01-01-102). The authors would like to thank people who were involved in developing programming languages, such as NCL and Python, including their libraries. We also sincerely appreciate the hard work of all researchers who have been involved in satellite observations. The authors declare no conflict of interest.

Code, Data, and Materials Availability

GOCI data were provided by Korea Institute of Ocean Science and Technology (KIOST) with an algorithm developed by the Atmospheric Radiation Laboratory at Yonsei University. The AERONET AODs were provided by the National Aeronautics and Space Administration (NASA) at <https://aeronet.gsfc.nasa.gov>. A detailed description of SciPy and PyKriging in Python can be found at <https://scipy.org/> and <https://geostat-framework.readthedocs.io/projects/pykriging/en/stable>, respectively.

References

1. IPCC, *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, V. Masson-Delmotte et al., Cambridge University Press (2021).
2. Z.-Y. Chen et al., “Extreme gradient boosting model to estimate PM_{2.5} concentrations with missing-filled satellite data in China,” *Atmos. Environ.* **202**, 180–189 (2019).
3. K. Lee et al., “Development of Korean Air Quality Prediction System version 1 (KAQPS v1) with focuses on practical issues,” *Geosci. Model Dev.* **13**, 1055–1073 (2020).
4. K. Lee and C. E. Chung, “Observationally-constrained estimates of global fine-mode AOD,” *Atmos. Chem. Phys.* **13**(5), 2907–2921 (2013).
5. P. E. Saide et al., “Assimilation of next generation geostationary aerosol optical depth retrievals to improve air quality simulations,” *Geophys. Res. Lett.* **41**(24), 9188–9196 (2014).
6. Y. Tang et al., “A case study of aerosol data assimilation with the Community Multi-scale Air Quality Model over the contiguous United States using 3D-Var and optimal interpolation methods,” *Geosci. Model Dev.* **10**(12), 4743–4758 (2017).
7. B. Kianian, Y. Liu, and H. H. Chang, “Imputing satellite-derived aerosol optical depth using a multi-resolution spatial model and random forest for PM_{2.5} prediction,” *Remote Sens.* **13**(1), 126 (2021).
8. Y. Lops et al., “Application of a partial convolutional neural network for estimating geostationary aerosol optical depth data,” *Geophys. Res. Lett.* **48**(15), e2021GL093096 (2021).
9. R. Zhang et al., “A nonparametric approach to filling gaps in satellite-retrieved aerosol optical depth for estimating ambient PM_{2.5} levels,” *Environ. Pollut.* **243**(Pt B), 998–1007 (2018).
10. M. Ghahremanloo et al., “Estimating daily high-resolution PM_{2.5} concentrations over Texas: Machine Learning approach,” *Atmos. Environ.* **247**, 118209 (2021).
11. H. Lim et al., “Integration of GOCI and AHI Yonsei aerosol optical depth products during the 2016 KORUS-AQ and 2018 EMeRGe campaigns,” *Atmos. Meas. Tech.* **14**(6), 4575–4592 (2021).
12. S. Lee et al., “Analysis of long-range transboundary transport (LRTT) effect on Korean aerosol pollution during the KORUS-AQ campaign,” *Atmos. Environ.* **204**, 53–67 (2019).

13. G.-H. Choo et al., “Optical and chemical properties of long-range transported aerosols using satellite and ground-based observations over Seoul, South Korea,” *Atmos. Environ.* **246**, 118024 (2021).
14. M. Choi et al., “GOCI Yonsei aerosol retrieval version 2 products: an improved algorithm and error analysis with uncertainty estimation from 5-year validation over East Asia,” *Atmos. Meas. Tech.* **11**(1), 385–408 (2018).
15. B. N. Holben et al., “AERONET—a federated instrument network and data archive for aerosol characterization,” *Remote Sens. Environ.* **66**(1), 1–16 (1998).
16. <https://aeronet.gsfc.nasa.gov>
17. P. Virtanen et al., “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat. Methods* **17**(3), 261–272 (2020).
18. G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, California (2009).
19. R. S. Varga, *Matrix Iterative Analysis*, 2nd ed. 2000, 2nd printing 2009 ed., Springer, Berlin (2009).
20. NCL, *The NCAR command language (Version 6.6.2) [Software]*, UCAR/NCAR/CISL/TDD, Boulder, Colorado (2019).
21. H. Saito et al., “Geostatistical interpolation of object counts collected from multiple strip transects: Ordinary Kriging versus finite domain Kriging,” *Stoch. Environ. Res. Ris. Assess.* **19**(1), 71–85 (2005).
22. H. Wackernagel, “Ordinary Kriging,” in *Multivariate Geostatistics: An Introduction with Applications*, H. Wackernagel, Ed., pp. 74–81, Springer Berlin Heidelberg, Berlin, Heidelberg (1995).
23. B. Murphy, S. Müller, and R. Yurchak, *GeoStat-Framework/PyKrige: v1.6.1*, Zenodo (2021).
24. B. S. Murphy, “PyKrige: development of a Kriging toolkit for Python,” in *AGU Fall Meeting Abstr.* pp. H51K–0753 (2014).
25. Y.-H. Ahn et al., “Missions and user requirements of the 2nd geostationary ocean color imager (GOCI-II),” *Korean J. Remote Sens.* **26**(2), 277–285 (2010).
26. S.-S. Yong et al., “Current status and results of in-orbit function, radiometric calibration and INR of GOCI-II (geostationary ocean color imager 2) on Geo-KOMPSAT-2B,” *Korean J. Remote Sens.* **37**(5_2), 1235–1243 (2021).
27. W. J. Choi et al., “Introducing the geostationary environment monitoring spectrometer,” *J. Appl. Remote Sens.* **12**(4), 044005 (2018).
28. J. Kim et al., “New era of air quality monitoring from space: geostationary environment monitoring spectrometer (GEMS),” *Bull. Am. Meteorol. Soc.* **101**(1), E1–E22 (2020).

Kyunghwa Lee received her MS and PhD degrees in environmental sciences from Gwangju Institute of Science and Technology in 2013 and 2018, respectively. Currently, she is working as an environmental researcher at Environmental Satellite Center of the National Institute of Environmental Research, Republic of Korea. Her research interests include filling gaps in satellite data, analysis in optical and chemical properties of aerosols, and data assimilation using CTMs and observations, including satellites and ground-based *in situ* measurements.

Biographies of the other authors are not available.