Atmospheric Research 243 (2020) 104997

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Development of a cloud-screening algorithm for direct and diffuse AODs from the Skyradiometer Network

Yongjoo Choi[a], Young Sung Ghim[a,*], Sang-Woo Kim[b], Huidong Yeo[b]

[a] Department of Environmental Science, Hankuk University of Foreign Studies, Yongin 17035, Republic of Korea
[b] School of Earth and Environmental Sciences, Seoul National University, Seoul 08826, Republic of Korea

ARTICLE INFO

Keywords:
SKYNET
AERONET
Aerosol optical properties
Screening efficiency
Cloud-detection statistics

ABSTRACT

We developed a cloud-screening algorithm for direct and diffuse aerosol optical depths (AODs) from the Skyradiometer Network (CSDD). Variabilities of direct AODs were checked not only to screen cloud-contaminated (cloudy) data effectively but also to make direct AODs available for aerosol study, as in the Aerosol Robotic Network (AERONET). Skyradiometer data were used from the Seoul National University (SNU) site for three years from 2012 to 2014. CSDD tested the spectral and temporal variabilities according to the Ångström exponent at the first stage, and the temporal smoothness in the form of the coefficient of variation at the second stage. The algorithm for CSDD was constructed to minimize the differences with optical properties of cloudy data (and clear-sky data as well) based on the cloud cover from the synoptic station. A number of cloudy data that was not screened by the previous algorithms was removed, and the absolute value of the bias total could be substantially lowered. The performances of algorithms for cloud detection were also examined using lidar observations at the study site in terms of accuracy, probability of detection (POD), and false detection rate. The statistics of cloud detection for CSDD were generally comparable to those of AERONET for direct data, and the POD for diffuse data was improved to the level of direct data.

1. Introduction

During the past two decades, ground-based aerosol remote-sensing networks, including Aerosol Robotic Network (AERONET) and Skyradiometer Network (SKYNET), have been deployed throughout the world (Holben et al., 1998, 2001; Nakajima et al., 2003, 2007; Campanelli et al., 2012; Zhang et al., 2012; Li et al., 2014; Hamill et al., 2016; Yoon et al., 2016; Arola et al., 2017). The number of sites has expanded up to ~600 for AERONET (Giles et al., 2019) and ~ 50 for SKYNET (http://skynet-isdc.org/obs_sites.php) because of the ease of acquiring information on aerosols from the ground. The strength of such networks largely relies on imposing a standardization of measurements and data processing, which allows multi-year and large-scale studies. However, these networks have a crucial restriction, in that aerosol optical properties can be measured only under clear-sky conditions in the daytime (Christopher and Gupta, 2010; Choi and Ghim, 2017). Cloud screening (screening of cloud-contaminated data) is particularly important, because misperception of optically thin clouds as aerosols, even cirrus clouds, can seriously obscure aerosol optical properties.

The basic idea of cloud screening starts from the assumption that radiation changes due to clouds are rapid, whereas those due to aerosols are gradual. Smirnov et al. (2000) set up the cloud-screening algorithm for AERONET aerosol optical depths (AODs), consisting of a stability test for triplet values over a 1-min period, followed by statistical and smoothness tests for data in a day. They indicated that this algorithm eliminated ~20% to 50% of data as being cloud-contaminated. Kaufman et al. (2005, 2006) suggested the spectral-variability algorithm (SVA) for AERONET AODs for an Ångström exponent (AE) greater than 0.3 to keep highly variable aerosols such as smoke plumes, a large proportion of which would be rejected as clouds. It is known that AERONET version 2 data have been processed by the algorithm developed by Smirnov et al. (2000) after initial checking of the triplet values of raw signal to remove data with large temporal variations (Eck et al., 2014). Recently, AERONET released the version 3 database, in which fully automated and improved cloud screening and quality control checks were implemented. By using an elaborate cloud-screening algorithm compared to that for the version 2 based on Smirnov et al. (2000), Giles et al. (2019) reported that about 60% of data were removed from the complete Sun photometer database, which is similar to the coverage of clouds globally (about 68%).

For SKYNET, Khatri and Takamura (2009) developed an algorithm

* Corresponding author.
E-mail address: ysghim@hufs.ac.kr (Y.S. Ghim).

https://doi.org/10.1016/j.atmosres.2020.104997
Received 3 July 2019; Received in revised form 26 February 2020
Available online 11 April 2020
0169-8095/

for cloud screening of skyradiometer data (CSSR), using global irradiance measured by a pyranometer. Song et al. (2014) proposed an improved cloud-screening method (ICSM) consisting of a temporal-variability test using direct and diffuse AODs and a coarse-mode test to eliminate data that were likely contaminated by thin cirrus-type clouds. SKYNET recommended CSSR for cloud screening (http://skynet-isdc.org/methodology.php).

Whereas direct Sun AODs from AERONET have been utilized in various studies (Gobbi et al., 2007; Levy et al., 2010; Li et al., 2014; Zhang et al., 2016), little attention has been paid to direct AODs from SKYNET. Song et al. (2014) checked the variability of direct AODs but for separating cloud-contaminated diffuse AODs. We noted direct AODs from SKYNET because their time intervals (~1 min) are shorter than those of diffuse AODs from SKYNET (3–10 min), direct Sun AERONET AODs (3–15 min), and AODs derived from AERONET inversion (~1 h). It is apparent that short time-interval data have an advantage of providing detailed information on time variations. However, when dealing with data from remote-sensing networks, where individual data are frequently missing (primarily because valid data are available only under clear-sky conditions), it is more important to obtain more hourly data by using highly resolved data in short time intervals. More hourly data can provide more valid daily and monthly data.

In this study, we developed a cloud-screening algorithm for both direct and diffuse AODs from SKYNET. Variabilities of direct AODs were checked not only to screen cloud-contaminated (cloudy) data in short time intervals but also to make full use of direct AODs as in AERONET. The performance of cloud screening was assessed by comparing the optical properties of cloudy data with those based on the cloud cover during the development of the algorithm, and by comparing cloud detections with lidar observations after the development of the algorithm. The study site was the Seoul National University (SNU; 37.46°N, 126.95°E, 150 m asl; Fig. 1), where lidar and a pyranometer were installed. The study period was three years from 2012 to 2014.

2. Methods

2.1. Instruments

The skyradiometer, POM-01 (Prede Co. Ltd) at SNU measures direct and diffuse radiation with a 0.5° field of view at seven wavelengths of 315, 400, 500, 675, 870, 940, and 1020 nm. Diffuse radiation, also called sky radiation or (diffuse) skylight, was measured on the almucantar plane (with a fixed zenith angle and varied azimuth angle up to 180° on one sides) every 3 min. Direct radiation was measured at intervals of 1 min or less when diffuse radiation was not measured. Except for radiation at 315- and 940-nm wavelengths, absorbed by ozone and water vapor, respectively, radiation at five wavelengths is used for aerosol study. Diffuse AODs were retrieved from almucantar measurements using skyrad.pack version 5 (Hashimoto et al., 2012) that could suppress the concentration of coarse-mode particles with a radius of over 10 μm (considered cirrus-contaminated) in comparison with version 4.2 (Nakajima et al., 1996). Direct AODs were obtained from direct radiation using the monochromatic direct solar flux density equation (Nakajima et al., 1996; Hashimoto et al., 2012), subtracting optical depths caused by ozone and Rayleigh scattering.

The lidar at SNU is a two-wavelength Mie scattering lidar equipped with depolarization measurement capability (Kim et al., 2008; Kim et al., 2014). In addition to an analog detection system, the lidar employs an Nd:YAG laser as a light source that generates a fundamental output at 1064 nm and a second harmonic at 532 nm. It has been operated since March 2006 as a part of the National Institute for Environmental Studies (NIES) lidar network (http://www-lidar.nies.go.jp). The lidar produces a vertical profile up to 18 km with a resolution of 30 m for 5 min at 0, 15, 30, and 45 min every hour. The presence of clouds is detected in lidar observation by sharp changes of range-corrected signal returns at cloud bases and tops (Kim et al., 2008; Kim et al., 2014). Cloud boundaries are identified by the maximum slope of six sequenced points (altitude) of the returned lidar backscattering

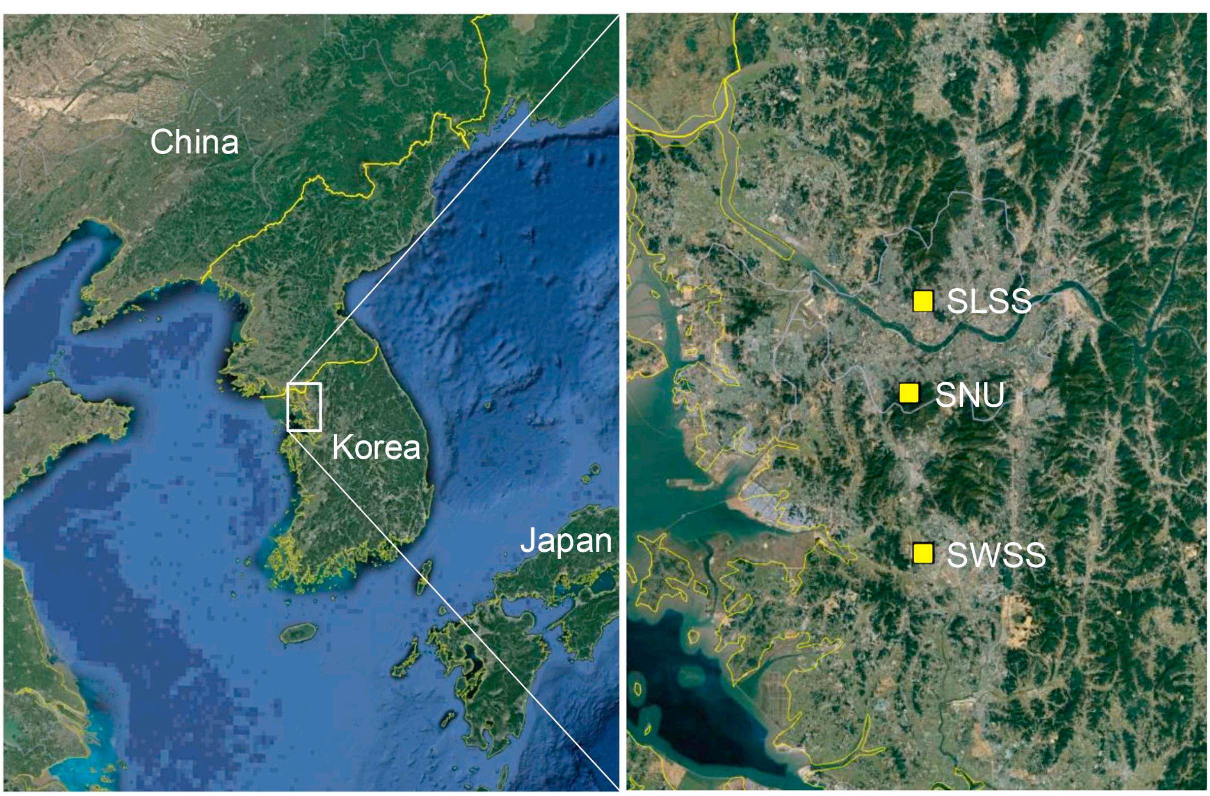


Fig. 1. Location of the Seoul National University (SNU) study site. The Seoul synoptic station (SLSS) and the Suwon synoptic station (SWSS) are about 12 km north of and 21 km south of the study site, respectively.

intensity at 532 nm. For optically thin cirrus clouds, a particle depolarization ratio at 532 nm (> 0.2) is additionally used to distinguish them from aerosols, because cirrus clouds are composed mostly of nonspherical ice crystals (Wang and Sassen, 2001).

The global solar irradiance was also measured approximately from 300 to 2800 nm using a collocated CM21 pyranometer (Kipp & Zonen) for the global irradiance test included in CSSR.

2.2. Determination of clear-sky and cloudy conditions

We first considered lidar to distinguish between clear-sky and cloudy conditions because it was collocated with skyradiometer. However, it sees the sky vertically within a narrow solid angle cone compared to skyradiometer measurements covering the hemispherical sky. Therefore, AODs only for the solar zenith angle less than 30° were classified into clear-sky and cloudy (Giles et al., 2019), depending on whether lidar detected clouds or not, over a 10-min period centered on the lidar observation time at 15-min intervals. However, due to limitation in the solar zenith angle, useful lidar data were biased to specific months (April–September) and hours (11:00–14:00 local time). Thus, we used these lidar observations to evaluate the performance of cloud detection by the cloud-screening algorithm developed in this study.

During the development of the cloud-screening algorithm, we used cloud cover observed at the synoptic station operated by the Korea Meteorological Administration to distinguish between clear-sky and cloudy conditions. Cloud cover, ranging from 0 to 10, is reported every hour through human observation. Diffuse and direct AODs over a 10-min period, spanning five minutes before and after each hour, were classified as either clear-sky for a cloud cover 0–2 or cloudy otherwise (Song et al., 2014). Because many studies have attempted to obtain information on physical and chemical characteristics of aerosols from optical properties (Eck et al., 2005; Gobbi et al., 2007; Kim et al., 2007; Russell et al., 2010; Zhang et al., 2012; Collaud Coen et al., 2013; Li et al., 2014; Yoon et al., 2016), we first compared AOD and AE for clear-sky and cloudy conditions determined by the cloud cover from the Seoul synoptic station (37.57°N, 126.97°E, 85.8 m asl; Fig. 1) with those from cloud-screening algorithms to evaluate the performance of the algorithm. However, because the Seoul synoptic station is about 12 km north of the study site, we performed the same comparison using the cloud cover from the Suwon synoptic station (37.27°N, 126.99°E, 34.1 m asl), which is about 21 km south of the study site, to confirm the validity of using the cloud cover at different locations.

3. Optical properties of aerosols from selected algorithms

3.1. Optical properties for clear-sky and cloudy conditions based on cloud cover

Fig. 2 shows the frequency distributions of data points in the domain of AE vs. AOD from diffuse and direct radiation. AOD is the value at 500 nm, and AE was calculated from AODs at 870 and 500 nm. Clear-sky and cloudy conditions were determined based on the cloud cover from the Seoul synoptic station. Among 9,858 and 31,994 data from diffuse and direct radiation, 23% (2,289) and 48% (15,326) of them, respectively, were classified as cloudy. The frequency of data from diffuse radiation (diffuse data) for clear-sky condition peaks around 0.2–0.3 for AOD and 1.2–1.6 for AE (Fig. 2a); the peak area of direct data is broader but similar to that of diffuse data (Fig. 2c). Note that much more direct data are distributed to high AOD and low AE than diffuse data, as seen in the lower right portion of Fig. 2c. This is in part because direct data in shorter time intervals can capture a wider range of aerosols with higher AOD and lower AE. However, considering that diffuse AODs are inversion products, the difference in the frequency distribution between diffuse and direct data in Fig. 2a and c could result because some diffuse data with high AODs and low AEs were discarded as abnormal during the retrieval process.

The tendency is clearer for cloudy data. Whereas the upper limit of AE for direct data decreases with increasing AOD, the lower limit of AE approaches zero (Fig. 2d). In contrast, diffuse data are scarce at low AEs, because AEs near zero, which were cloud-contaminated and thus classified as cloudy, had already been discarded as abnormal during the retrieval process (Fig. 2b). Several data points at high AEs particularly for direct data are also worthy of note (Fig. 2d). Since cloudy data will have lower AEs, these data points are unusual even in small numbers. It is likely that determination of clear-sky and cloudy conditions based on the cloud cover would misclassify cloudy data.

3.2. Comparison of biases in optical properties from selected algorithms

In this section, we compared the frequency distributions of data from each cloud-screening algorithm (A) with those based on the cloud cover (B) to examine the performances of existing algorithms. The biases in frequencies of data from each algorithm were calculated by subtracting (B) from (A). Fig. 3 shows the biases in frequencies of cloudy data only, because if the biases are positive for cloudy data, they are negative for clear-sky data, and vice versa. The biases would be zero if the frequencies of cloudy data selected by a cloud-screening algorithm coincide with those based on the cloud cover at given values of AE and AOD. The algorithm developed by Smirnov et al. (2000) is designated as the temporal-variability algorithm (TVA), correspondingly to the SVA developed by Kaufman et al. (2006).

Both TVA and SVA check the triplet variability, which is defined as the difference between maximum and minimum AODs, to screen the cloud-contaminated triplet values over 1 min (Smirnov et al., 2000; Kaufman et al., 2006). In this study, we used both direct and diffuse AODs, which were obtained from direct and diffuse radiation measured every 1 to 2 min. Because one or two data were common for 1 min, we checked the variabilities over a 2-min period. Nevertheless, the number of data was still one or two during a 2-min period in half the cases, since the skyradiometer does not measure radiation when it cannot track the sun because of clouds or when the rain sensor detects rain. If two data were available for 2 min, the difference between the two values was checked, but if only one datum was available, it was discarded as being cloud-contaminated.

For TVA, we carried out two checks and three criteria tests in the order described in Smirnov et al. (2000), including the triplet stability criterion test using the aforementioned triplet variabilities during a 2-min period for five wavelengths (400, 500, 675, 870, and 1020 nm). The triplet variabilities for SVA were calculated using AODs at 440, 675 and 870 nm, where AOD at 440 nm was obtained by linear interpolation of AODs at 400 and 500 nm. SVA also tested the variabilities of three AODs using the maximum difference between the current AOD and the next or previous one at ~15 min intervals. However, we did not perform this test for SVA because it could not screen additional cloudy data. We added an asterisk (*) to the acronyms to indicate that we modified TVA and SVA for skyradiometer data.

Biases tend to be large where the frequencies based on the cloud cover are large in Fig. 2 for both diffuse and direct data (Fig. 3a and b) because absolute, instead of relative, values of biases are presented. Despite a few positive biases, negative biases prevail for both diffuse and direct data, showing that the algorithms selected in this study did not screen much cloudy data distinguished based on the cloud cover. This is more true for SVA* than for TVA*. The discrepancy between SVA* and TVA* is amplified in the distribution for direct data, as negative biases are intensified for SVA* while TVA* exhibits some balance between positive and negative biases (Fig. 3b). In CSSR, negative biases are mostly found for AE > 1.0, probably because data for AE > 1.0 were classified as cloud-free without spectral-variability test (Khatri and Takamura, 2009).

Note that Figs. 2 and 3 show the distributions of frequencies and biases, respectively, in the domain of AE and AOD between 0 and 3. Negative and positive biases were separately summed over the entire

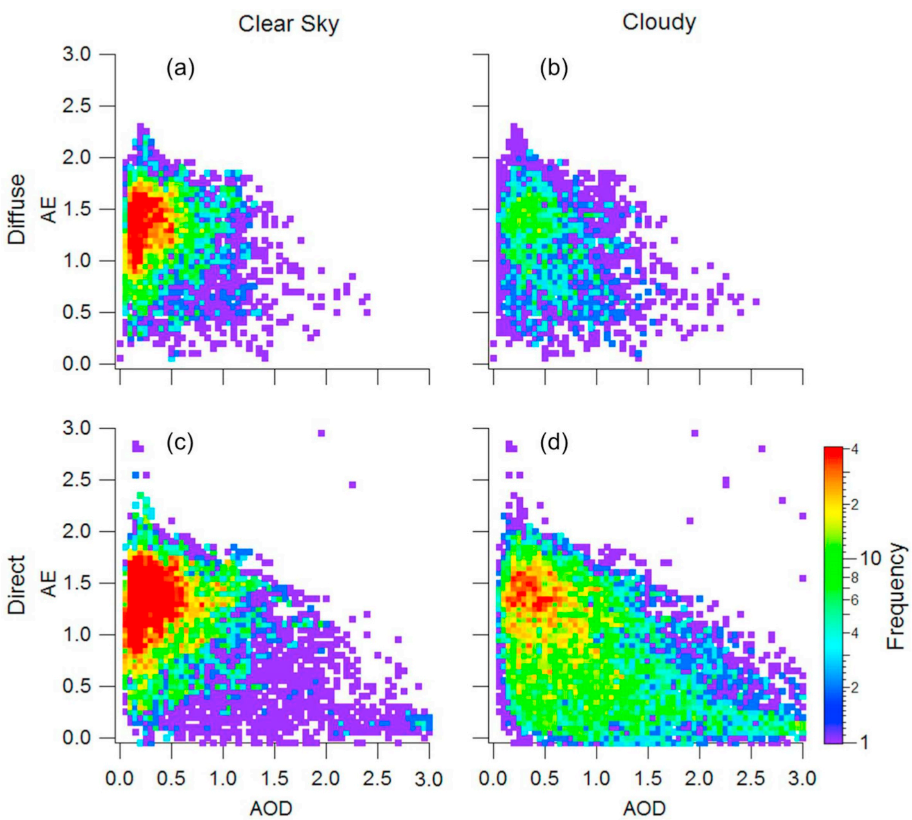


Fig. 2. Comparison of frequency distributions of data points from diffuse and direct radiation under clear-sky and cloudy conditions distinguished based on the cloud cover from the Seoul synoptic station. The numbers of data points under clear-sky and cloudy conditions are 7,569 and 2,289, respectively, for diffuse radiation and 16,668 and 15,326, respectively, for direct radiation. Note that the color scale is logarithmic to show the variations in the frequencies.

ranges of AE and AOD (without limiting the values) and are provided as a percentage in Table 1. The sums of the biases for the diffuse data set were divided by the total number of diffuse data (before cloud screening) to calculate the percentage; the same was done for the direct data set. The bias total is the simple sum of positive and negative bias sums, whereas the deviation is given by the sum of absolute values of positive and negative bias sums.

As seen in Fig. 3, all the algorithms demonstrate the negative bias total. Absolute value of the bias total is lowest for both diffuse and direct data from TVA*, and so is the deviation. For diffuse data, it is interesting to note that the deviation of CSSR, which employs the spectral-variability test as a part of the algorithm, is similar to that of SVA*, and the deviation of ICSM, which encompasses the concept of the temporal-variability test, is similar to that of TVA*. However, the deviations of TVA* and SVA* originally developed for AERONET are lower than those of ICSM and CSSR developed for SKYNET, respectively, although the difference is small. In contrast, the absolute value of the bias total of SVA* is larger than that of CSSR.

4. Development of a new algorithm

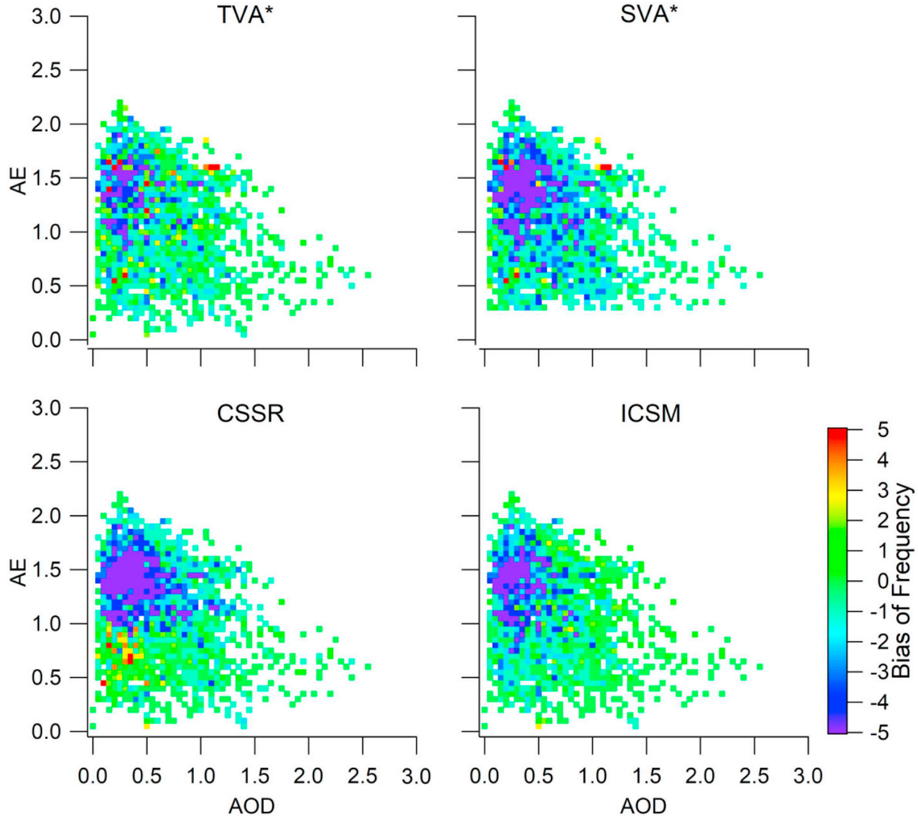
4.1. Structure of a new algorithm

Fig. 4 shows the flow diagram of a cloud-screening algorithm for SKYNET, developed in this study. It was designated as a cloud-screening algorithm for diffuse and direct AODs (from SKYNET) (CSDD), which is common to TVA and SVA for AERONET, but not to CSSR and ICSM for SKYNET. CSDD starts with a spectral-variability test for AE $>$ 0.3 or a temporal-variability test otherwise. We noted spectral- and temporal-variability tests for AERONET mainly because CSSR and ICSM cannot deal with direct data. However, TVA* and SVA* in Table 1 reveal high negative biases, indicating that they did not screen much cloudy data based on the cloud cover. We presumed that the 2-min windows employed in TVA* and SVA* could not produce sufficient variability to distinguish between clear-sky and cloudy data. Therefore, we adopted a 15-min window for the spectral-variability test; that is, the difference between maximum and minimum AODs over a 15-min period was checked with the 15-min criterion given by Kaufman et al. (2006). This test differs from the original 15-min test performed in Kaufman et al. (2006), which used the difference between the current AOD and the next or previous one at ~15 min intervals. As mentioned earlier, we tried that for SVA* but found it to be ineffective.

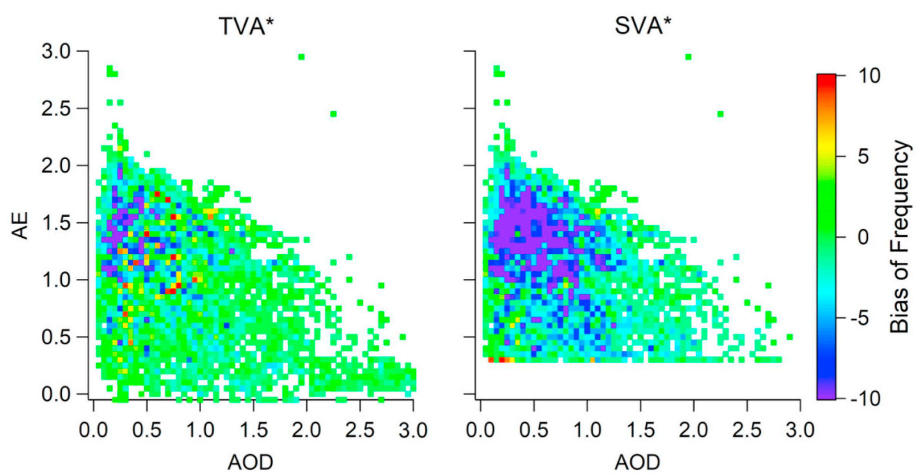
Compared to using the 15-min window for the spectral-variability test, we used the 2-min window for the temporal-variability test, as in TVA*, considering relatively small negative bias sum of TVA*. Table 2 presents the performance of screening cloudy data in each step of CSDD. A high percentage of cloudy data removed by the spectral-variability test indicates that the adoption of the 15-min window was suitable. We presumed that a small number of cloudy data were removed by the temporal-variability test because the number of data for AE $\leq$ 0.3 was small, that is, 73 (0.7%) for diffuse data and 4,456 (13.9%) for direct data. Because of a larger portion of direct data for AE $\leq$ 0.3, the percentage of cloudy data removed by the temporal-variability test is higher for direct data than for diffuse data.

For data passing the spectral- and temporal-variability tests, we checked the temporal smoothness of the data. Similar tests have been conducted using an index D (first derivatives difference) for data in a day (Smirnov et al., 2000), the standard deviation (σ) over a 10-min period (Song et al., 2014), and the relative rate of AOD change per minute (Giles et al., 2019). We used the coefficient of variation (σ divided by the mean) to check the smoothness of data, and accepted data when the coefficient of variation over a 10-min period was $\leq$0.10. The percentages of cloudy data removed by the smoothness test are 7.6% for diffuse data and 12.3% for direct data in Table 2.

Table 2 shows that 76.4% of diffuse data and 54.9% of direct data were accepted as clear-sky data. Summing diffuse and direct data, 40.0% was removed as cloudy data. The flow diagram in Fig. 4 was constructed in a way to reduce the bias total and deviation in Table 1. The lengths of the time windows for the tests and the coefficient of variation for the smoothness test were chosen in the same way.



(a) Diffuse radiation



(b) Direct radiation

Fig. 3. Distributions of biases in the frequency of cloudy data from selected algorithms compared to that based on the cloud cover from the Seoul synoptic station (Fig. 2b and d). TVA* and SVA* are modified TVA and SVA, respectively, for skyradimeter. The bias of frequency for SVA* was calculated for AE > 0.3, because SVA is applicable to those data (Kaufman et al., 2006). Note that AE was calculated for 870 and 500 nm instead of using 870 and 440 nm in Kaufman et al. (2006).

However, Table 1 was prepared by using the cloud cover from the Seoul synoptic station, which is located about 12 km north of the study site (Fig. 1). To examine the validity of using the cloud cover at a different location, Table S1 was prepared by using the cloud cover from the Suwon synoptic station, which is about 21 km south of the study site. Despite the distance between two synoptic stations, biases and deviations in Tables 1 and S1 are very similar; differences in the bias total range from −0.5% to 1.1%, while those in the deviation range from 0.4% to 1.6% (Table S2). This reveals that differences in the cloud cover between the study site and synoptic stations are small and that our approach to reduce the bias total and deviation in Table 1 is appropriate.

The distributions of biases in the frequency of cloudy data from CSDD compared to that based on the cloud cover from the Seoul synoptic station are shown in Fig. S1. The bias totals and deviations of cloudy data frequencies from CSDD are listed in Table 3. The statistics for AE > 0.3 are provided separately to compare them with those from SVA* in Table 1. In comparison with Table 1, the absolute value of the bias total is substantially reduced particularly for diffuse data. Only the deviation for direct data is increased compared to that from TVA*. The absolute value of the negative bias sum is mainly reduced, indicating that much cloudy data, which was not screened by the selected

Table 1
Bias total and deviation of cloudy data frequencies from selected algorithms compared to those based on the cloud cover from the Seoul synoptic station.

	Diffuse AODs[a]			Direct AODs[b]		
	TVA*	SVA*[,c]	CSSR	ICSM	TVA*	SVA*[,c]
Positive bias sum	2.8%	1.1%	2.0%	1.4%	3.4%	0.9%
Negative bias sum	−10.6%	−16.8%	−16.0%	−12.1%	−7.2%	−18.0%
Bias total	−7.8%	−15.7%	−14.0%	−10.7%	−3.8%	−17.2%
Deviation	13.4%	17.9%	18.0%	13.5%	10.7%	18.9%

[a] Percentage of the total number of diffuse data before cloud screening.
[b] Percentage of the total number of direct data before cloud screening.
[c] Calculated from the data for AE > 0.3.

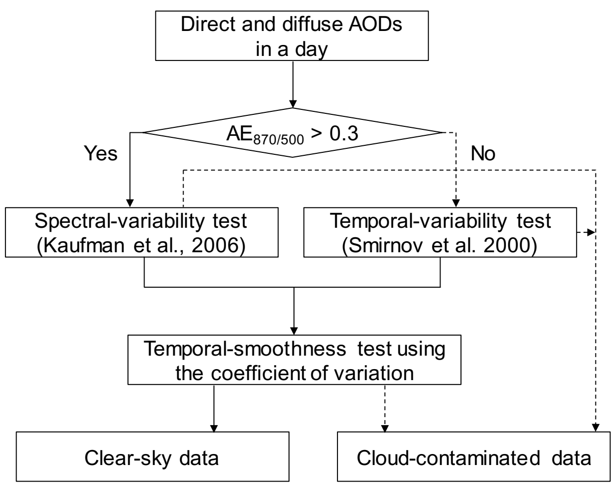


Fig. 4. Flow diagram of a cloud-screening algorithm for diffuse and direct AODs from SKYNET (CSDD).

algorithms in Table 1, was screened by CSDD. However, the positive bias sum is somewhat increased, because some clear-sky data was also screened by CSDD.

As for Table S1, Table S3 was prepared by using the cloud cover from the Suwon synoptic station. Although important parameters of CSDD were determined to reduce the bias total and deviation in Table 1 (based on the cloud cover from the Seoul synoptic station), the values in Table S3 are similar to those in Table 3 and differences between the two are small in Table S4.

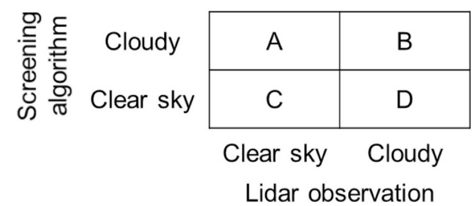
4.2. Cloud-detection statistics

Up to now, we have examined the performance of the algorithm in terms of biases in optical properties based on the cloud cover, because many studies have attempted to obtain information on aerosol characteristics by using a large amount of data from long-term

Table 3
Bias total and deviation of cloudy data frequencies from CSDD compared to those based on the cloud cover from the Seoul synoptic station.

	Diffuse AODs[a]		Direct AODs[b]	
	All	AE > 0.3	All	AE > 0.3
Positive bias sum	5.1%	5.1%	5.4%	6.1%
Negative bias sum	−4.8%	−4.4%	−8.3%	−3.3%
Bias total	0.4%	0.7%	−2.8%	2.8%
Deviation	9.9%	9.6%	13.7%	9.4%

[a] Percentage of the total number of diffuse data before cloud screening.
[b] Percentage of the total number of direct data before cloud screening.



$$\text{Accuracy} = \frac{B+C}{A+B+C+D}$$

$$\text{Probability of detection (POD)} = \frac{B}{B+D}$$

$$\text{False detection rate (FDR)} = \frac{A}{A+B}$$

Fig. 5. Definition of parameters used in the statistical analysis of cloud detections by algorithms compared to those by collocated lidar observations.

measurements of optical properties. However, when the study period is not long enough or the amount of data is not large, identifying the cloud contamination for each datum is important, because aerosol characteristics could be obscured by the inclusion of cloudy data or exclusion of clear-sky data individually. Therefore, we examined whether cloud detections by the algorithm agreed with those by collocated lidar observations in each case.

We distinguished four groups from "A" to "D" as shown in Fig. 5. "A" indicates that the algorithm judged clear-sky data for lidar observations to be cloud-contaminated, whereas "B" indicates that both the algorithm and lidar judged data to be cloud-contaminated. We defined three parameters—the accuracy, probability of detection (POD), and false detection rate (FDR)—, which concepts were taken from the verification statistics used to evaluate the air quality forecasts (United States Environmental Protection Agency, 2003).

The statistics of cloud detections by algorithms compared to those by lidar observations are summarized in Table 4. For diffuse data, SVA* and CSSR work better than TVA* and ICSM, showing higher accuracy and lower FDR. However, PODs of SVA* and CSSR are lower, as expected from higher absolute values of negative bias sums (meaning that much cloudy data was not screened) in Table 1. Similar results are

Table 2
Performance of screening cloudy data in each step of CSDD.

	Diffuse AODs		Direct AODs	
	Clear-sky[a]	Cloudy[b]	Clear-sky	Cloudy
Total number before CSDD	9,858 (100.0)		31,994 (100.0)	
Spectral-variability test	7,709 (78.2)	2,149 (92.4)	20,413 (63.8)	11,581 (80.3)
Temporal-variability test	7,708 (78.2)	1 (0.0)	19,353 (60.0)	1,060 (7.3)
Smoothness test	7,531 (76.4)	177 (7.6)	17,572 (54.9)	1,781 (12.3)
Total number after CSDD		2,327 (100.0)		14,422 (100.0)

[a] Number of clear-sky data after each step. The figure in the parentheses indicates a percentage of the total number of data before applying CSDD.
[b] Number of cloudy data screened by each step. The figure in the parentheses indicates a percentage of the total number of cloudy data screened by CSDD.

Table 4
Statistics of cloud detections[a,b] by new and existing algorithms compared to those by lidar observations for the solar zenith angle less than 30°.

(a) Existing algorithms

	Diffuse AODs			Direct AODs		
	TVA*	SVA*[,c]	CSSR	ICSM	TVA*	SVA*[,c]
Accuracy	72.2%	80.9%	85.3%	78.0%	59.0%	67.1%
POD	55.8%	27.0%	41.0%	67.3%	84.4%	55.5%
FDR	63.9%	53.4%	28.6%	54.4%	54.8%	57.7%

(b) CSDD and AERONET[d].

	Diffuse AODs		Direct AODs		
	CSDD		CSDD		
	All	AE > 0.3	All	AE > 0.3	AERONET
Accuracy	74.2%	74.9%	66.2%	66.9%	70.8%
POD	79.9%	84.0%	84.1%	90.5%	83.3%
FDR	58.5%	59.0%	49.2%	55.2%	55.4%

[a] See Fig. 5 for the definition of the statistical parameters.
[b] The number of lidar observations is 3,662 for diffuse AODs and 17,201 for direct AODs. The number of direct AODs from AERONET is 839.
[c] Calculated using the data for AE > 0.3.
[d] For one and a half years from Jan. 2012 to July 2013 in comparison with the 3-year study period from 2012 to 2014.

found for direct data in Table 4. SVA* with a higher absolute value of the negative bias sum in Table 1 shows higher accuracy and lower POD, but in this case, FDR of SVA* is slightly higher than that of TVA*.

A sun-and-sky scanning radiometer, CE 318 (Cimel Electronique) of AERONET measured direct and diffuse radiation at the study site (Seoul_SNU) for one and a half years from Jan. 2012 to July 2013, compared with the three-year study period of Jan. 2012 to Dec. 2014. We calculated the statistics of cloud detections for direct data (version 3 AOD) using level 1.0 and 1.5 data sets. Whereas only cloudy data were removed from level 1.0 through cloud screening to obtain level 1.5 in version 2, anomaly data due to instrument malfunction were also removed in version 3 as part of quality control after cloud screening (Giles et al., 2019). In other words, clear-sky data classified by the cloud screening algorithm can be discarded during the instrument quality control procedures, which means that a part of clear-sky data (Giles et al. (2019) estimated ~6% for worldwide AERONET sites during 1993–2018) fell into the cloudy data category shown in Fig. 5. Thus, if we distinguish the data removed only by cloud screening from level 1.5 data set, a part of the cloudy data, "A" and "B", should be reassigned to the clear-sky category, "C" and "D", respectively. This will reduce "A" and "B" and increase "C" and "D". As a result, accuracy and FDR will either increase or decrease depending on the rate of change from "A" to "C" and "B" to "D", while POD decreases. Therefore, the accuracy and FDR in Table 4, calculated from level 1.0 and 1.5 data sets without considering the removal of data due to instrument malfunction, could represent approximate values; however, POD should be regarded as an upper limit.

Table 4 shows that POD and FDR of AERONET are similar to those of TVA* for direct data with slightly lower POD and higher FDR. Note that, different from existing algorithms, AERONET exhibits similarly high POD with TVA* despite higher accuracy. However, considering that POD of AERONET is an upper limit, it likely decreases, and the difference in POD between AERONET and TVA* becomes larger than that shown in Table 4. Compared to SVA*, AERONET has slightly higher accuracy and low FDR, while POD is significantly higher although the difference in POD can reduce with decreasing POD of AERONET. We could not calculate the statistics of cloud detections for diffuse data, because only level 1.5 data were available for inversion products of AERONET.

As shown in Table 4, the statistics of CSDD are generally comparable to those of AERONET. The increase in POD is remarkable even for diffuse data although the accuracy is slightly lowered and FDR is increased in comparison with the other three algorithms except TVA*. It is worthy of note that CSSR, using additional global irradiance data, exhibits higher accuracy and much lower FDR. For direct data, CSDD works better than TVA* and SVA* with higher accuracy and lower FDR than TVA*, and higher POD and lower FDR than SVA* for AE > 0.3. Compared to AERONET, the accuracy is lowered, POD is slightly higher (while the difference can increase with decreasing POD of AERONET), but FDR is lowered.

It was mentioned earlier that useful lidar data were biased to specific months and hours due to inherent limitation in the solar zenith angle. On the other hand, the cloud cover was available for the entire period of the skyradiometer measurement, and its variations between the study site and synoptic stations were revealed small in Tables S1 to S4. Thus, the statistics of cloud detection by the algorithms were compared to those by the cloud cover from the Seoul and Suwon synoptic stations in Tables S5 and S6, respectively. Unlike Table 4, in which the comparison results differed by parameter, CSDD mostly outperforms other algorithms for all three parameters, because CSDD was constructed to minimize the differences between the optical properties from the algorithm and those based on the cloud cover from the synoptic station. Compared to AERONET, CSDD still shows higher POD, but is accompanied by higher accuracy and lower FDR, resulting in a better performance than Table 4. As in Tables S2 and S4, Tables S5 and S6 using the cloud cover from the Seoul and Suwon synoptic stations, respectively, were similar.

5. Summary and conclusions

We developed a cloud-screening algorithm for both direct and diffuse AODs from SKYNET (CSDD) using skyradiometer data collected at the Seoul National University (SNU) site for the three years 2012–2014. CSDD starts with a spectral-variability test for AE > 0.3 or a temporal-variability test otherwise. The key element of CSDD was the spectral-variability test over a 15-min period for AE > 0.3. Summing diffuse and direct data, 82.0% of cloudy data was removed by this test. Overall, CSDD eliminated 40.0% of the original data as being cloud-contaminated.

We attempted to minimize the differences between the optical properties of cloudy data (and clear-sky data as well) based on the cloud cover and those from the algorithm. We found that CSDD could substantially reduce the absolute value of the bias total by removing much cloudy data that had not been removed by the existing algorithms. However, the positive bias sum was somewhat increased because some clear-sky data was screened as well. We also examined the performance of CSDD in terms of cloud detection statistics compared to lidar observations for the solar zenith angle less than 30°. The statistics of CSDD were comparable to those of AERONET estimated for direct data. As the removal efficiency of cloudy data increased, POD for diffuse data was improved to the level of direct data although the accuracy was lowered and FDR was increased compared to other existing algorithms.

Despite the performance of CSDD described above, it should be noted that CSDD was developed using data at a single site over three years. Meteorology and aerosol characteristics vary by site and by period, which can alter the CSDD performance. Further analyses using data from various sites and periods are warranted to use CSDD at worldwide SKYNET sites.

Acknowledgements

This study was supported by the Korea Meteorological

Administration Research and Development Program under the Grant KMIPA 2015-6010, the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1C1B6008004), and the Hankuk University of Foreign Studies Research Fund. We are grateful to B.-J. Sohn for maintaining the SKYNET site at the Seoul National University during the study period. We also thank anonymous reviewers for precise and valuable comments that greatly improved the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosres.2020.104997.

References

Arola, A., Eck, T.F., Kokkola, H., Pitkänen, M.R.A., Romakkaniemi, S., 2017. Assessment of cloud-related fine-mode AOD enhancements based on AERONET SDA product. Atmos. Chem. Phys. 17, 5991–6001. https://doi.org/10.5194/acp-17-5991-2017.

Campanelli, M., Estelles, V., Smyth, T., Tomasi, C., Martìnez-Lozano, M.P., Claxton, B., Muller, P., Pappalardo, G., Pietruczuk, A., Shanklin, J., Colwell, S., Wrench, C., Lupi, A., Mazzola, M., Lanconelli, C., Vitale, V., Congeduti, F., Dionisi, D., Cardillo, F., Cacciani, M., Casasanta, G., Nakajima, T., 2012. Monitoring of Eyjafjallajökull volcanic aerosol by the new European Skynet Radiometers (ESR) network. Atmos. Environ. 48, 33–45. https://doi.org/10.1016/j.atmosenv.2011.09.070.

Choi, Y., Ghim, Y.S., 2017. Assessment of the clear-sky bias issue using continuous PM10 data from two AERONET sites in Korea. J. Environ. Sci. 53, 151–160.

Christopher, S.A., Gupta, P., 2010. Satellite Remote Sensing of Particulate Matter Air Quality: the Cloud-Cover Problem. J. Air Waste Manage. Assoc. 60, 596–602.

Collaud Coen, M., Andrews, E., Asmi, A., Baltensperger, U., Bukowiecki, N., Day, D., Fiebig, M., Fjaeraa, A.M., Flentje, H., Hyvärinen, A., Jefferson, A., Jennings, S.G., Kouvarakis, G., Lihavainen, H., Lund Myhre, C., Malm, W.C., Mihapopoulos, N., Molenar, J.V., O'Dowd, C., Ogren, J.A., Schichtel, B.A., Sheridan, P., Virkkula, A., Weingartner, E., Weller, R., Laj, P., 2013. Aerosol decadal trends – part 1: In-situ optical measurements at GAW and IMPROVE stations. Atmos. Chem. Phys. 13, 869–894.

Eck, T.F., Holben, B.N., Dubovik, O., Smirnov, A., Goloub, P., Chen, H.B., Chatenet, B., Gomes, L., Zhang, X.-Y., Tsay, S.-C., Ji, Q., Giles, D., Slutsker, I., 2005. Columnar aerosol optical properties at AERONET sites in central eastern Asia and aerosol transport to the tropical mid-Pacific. J. Geophys. Res.-Atmos. 110 (D6), D06202. https://doi.org/10.1029/2004JD005274.

Eck, T.F., Holben, B.N., Reid, J.S., Arola, A., Ferrare, R.A., Hostetler, C.A., Crumeyrolle, S.N., Berkoff, T.A., Welton, E.J., Lolli, S., Lyapustin, A., Wang, Y., Schafer, J.S., Giles, D.M., Anderson, B.E., Thornhill, K.L., Minnis, P., Pickering, K.E., Loughner, C.P., Smirnov, A., Sinyuk, A., 2014. Observations of rapid aerosol optical depth enhancements in the vicinity of polluted cumulus clouds. Atmos. Chem. Phys. 14, 11633–11656. https://doi.org/10.5194/acp-14-11633-2014.

Giles, D.M., Sinyuk, A., Sorokin, M.G., Schafer, J.S., Smirnov, A., Slutsker, I., Eck, T.F., Holben, B.N., Lewis, J.R., Campbell, J.R., Welton, E.J., Korkin, S.V., Lyapustin, A.I., 2019. Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. Atmos. Meas. Tech. 12, 169–209. https://doi.org/10.5194/amt-12-169-2019.

Gobbi, G.P., Kaufman, Y.J., Koren, I., Eck, T.F., 2007. Classification of aerosol properties derived from AERONET direct sun data. Atmos. Chem. Phys. 7, 453–458. https://doi.org/10.5194/acp-7-453-2007.

Hamill, P., Giordano, M., Ward, C., Giles, D., Holben, B., 2016. An AERONET-based aerosol classification using the Mahalanobis distance. Atmos. Environ. 140, 213–233. https://doi.org/10.1016/j.atmosenv.2016.06.002.

Hashimoto, M., Nakajima, T., Dubovik, O., Campanelli, M., Che, H., Khatri, P., Takamura, T., Pandithurai, G., 2012. Development of a new data-processing method for SKYNET sky radiometer observations. Atmos. Meas. Tech. 5, 2723–2737.

Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I., Smirnov, A., 1998. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. Remote Sens. Environ. 66, 1–16.

Holben, B.N., Tanré, D., Smirnov, A., Eck, T.F., Slutsker, I., Abuhassan, N., Newcomb, W.W., Schafer, J.S., Chatenet, B., Lavenu, F., Kaufman, Y.J., Castle, J.V., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N.T., Pietras, C., Pinker, R.T., Voss, K., Zibordi, G., 2001. An emerging ground-based aerosol climatology: aerosol optical depth from AERONET. J. Geophys. Res. 106, 12067–12097. https://doi.org/10.1029/2001jd900014.

Kaufman, Y.J., Remer, L.A., Tanre, D., Rong-Rong, L., Kleidman, R., Mattoo, S., Levy, R.C., Eck, T.F., Holben, B.N., Ichoku, C., Martins, J.V., Koren, I., 2005. A critical examination of the residual cloud contamination and diurnal sampling effects on MODIS estimates of aerosol over ocean. IEEE Trans. Geosci. Remote Sens. 43, 2886–2897.

Kaufman, Y.J., Gobbi, G.P., Koren, I., 2006. Aerosol climatology using a tunable spectral variability cloud screening of AERONET data. Geophys. Res. Lett. 33, L07817.

Khatri, P., Takamura, T., 2009. An Algorithm to Screen Cloud-Affected Data for Sky Radiometer Data Analysis. J. Meteorol. Soc. Jpn. 87, 189–204.

Kim, S.-W., Yoon, S.-C., Kim, J., Kim, S.-Y., 2007. Seasonal and monthly variations of columnar aerosol optical properties over East Asia determined from multi-year MODIS, LIDAR, and AERONET Sun/sky radiometer measurements. Atmos. Environ. 41 (8), 1634–1651.

Kim, S.-W., Berthier, S., Raut, J.C., Chazette, P., Dulac, F., Yoon, S.C., 2008. Validation of aerosol and cloud layer structures from the space-borne lidar CALIOP using a ground-based lidar in Seoul, Korea. Atmos. Chem. Phys. 8, 3705–3720.

Kim, Y., Kim, S.-W., Kim, M.-H., Yoon, S.-C., 2014. Geometric and optical properties of cirrus clouds inferred from three-year ground-based lidar and CALIOP measurements over Seoul, Korea. Atmos. Res. 139, 27–35.

Levy, R.C., Remer, L.A., Kleidman, R.G., Mattoo, S., Ichoku, C., Kahn, R., Eck, T.F., 2010. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. Atmos. Chem. Phys. 10, 10399–10420. https://doi.org/10.5194/acp-10-10399-2010.

Li, J., Carlson, B.E., Dubovik, O., Lacis, A.A., 2014. Recent trends in aerosol optical properties derived from AERONET measurements. Atmos. Chem. Phys. 14, 12271–12289. https://doi.org/10.5194/acp-14-12271-2014.

Nakajima, T., Tonna, G., Rao, R., Boi, P., Kaufman, Y., Holben, B., 1996. Use of sky brightness measurements from ground for remote sensing of particulate polydispersions. Appl. Opt. 35, 2672–2686.

Nakajima, T., Sekiguchi, M., Takemura, T., Uno, I., Higurashi, A., Kim, D., Sohn, B.J., Oh, S.-N., Nakajima, T.Y., Ohta, S., Okada, I., Takamura, T., Kawamoto, K., 2003. Significance of direct and indirect radiative forcings of aerosols in the East China Sea region. J. Geophys. Res. Atmos. 108. https://doi.org/10.1029/2002JD003261.

Nakajima, T., Yoon, S.-C., Ramanathan, V., Shi, G.-Y., Takemura, T., Higurashi, A., Takamura, T., Aoki, K., Sohn, B.-J., Kim, S.-W., Tsuruta, H., Sugimoto, N., Shimizu, A., Tanimoto, H., Sawa, Y., Lin, N.-H., Lee, C.-T., Goto, D., Schutgens, N., 2007. Overview of the Atmospheric Brown Cloud East Asian Regional Experiment 2005 and a study of the aerosol direct radiative forcing in east Asia. J. Geophys. Res. 112, D24S91. https://doi.org/10.1029/2007jd009009.

Russell, P.B., Bergstrom, R.W., Shinozuka, Y., Clarke, A.D., DeCarlo, P.F., Jimenez, J.L., Livingston, J.M., Redemann, J., Dubovik, O., Strawa, A., 2010. Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition. Atmos. Chem. Phys. 10, 1155–1169. https://doi.org/10.5194/acp-10-1155-2010.

Smirnov, A., Holben, B.N., Eck, T.F., Dubovik, O., Slutsker, I., 2000. Cloud-screening and quality control algorithms for the AERONET database. Remote Sens. Environ. 73, 337–349.

Song, H.-J., Sohn, B.-J., Chun, H.-W., Chun, Y., Lee, S.-S., 2014. Improved cloud screening method for the analysis of sky radiometer measurements and application to asian dust detection. J. Meteorol. Soc. Jpn. 92A, 167–183.

United States Environmental Protection Agency, 2003. Guidelines for developing an air quality (Ozone and PM2.5) forecasting program. EPA-456/R-03-002.

Wang, Z., Sassen, K., 2001. Cloud type and macrophysical property retrieval using multiple remote sensors. J. Appl. Meteorol. 40, 1665–1682.

Yoon, J., Pozzer, A., Chang, D.Y., Lelieveld, J., Kim, J., Kim, M., Lee, Y.G., Koo, J.H., Lee, J., Moon, K.J., 2016. Trend estimates of AERONET-observed and model-simulated AOTs between 1993 and 2013. Atmos. Environ. 125 (Part A), 33–47. https://doi.org/10.1016/j.atmosenv.2015.10.058.

Zhang, W., Gu, X., Xu, H., Yu, T., Zheng, F., 2016. Assessment of OMI near-UV aerosol optical depth over Central and East Asia. J. Geophys. Res. Atmos. 121, 382–398. https://doi.org/10.1002/2015JD024103.

Zhang, Y., Yu, H., Eck, T.F., Smirnov, A., Chin, M., Remer, L.A., Bian, H., Tan, Q., Levy, R., Holben, B.N., Piazzolla, S., 2012. Aerosol daytime variations over North and South America derived from multiyear AERONET measurements. J. Geophys. Res. Atmos. 117. https://doi.org/10.1029/2011JD017242.