Abstract

Title of Document: Using Mahalanobis Distance to classify aerosol in Southeast Asia based on AERONET-retrieved optical properties

Hans Jarett Ong, 5 BS PS

Directed By: Dr. Nofel D. Lagrosas
Clint Dominic G. Bennett

Aerosol types over Southeast Asia (SEA) are determined from Aerosol Robotic Network (AERONET) derived aerosol optical properties for 25 sites using Mahalanobis method. Angstom exponent (*AE*), single scattering albedo (*SSA*), and real refractive index (*n*) are used in a three-dimensional specified clustering method that classified aerosol into 7 classes, namely: biomass burning white smoke (BB-W), polluted dust (PD), urban industrial developing economy (UI-D), urban industrial (UI), biomass burning dark smoke (BB-D), mineral dust (MD), and marine aerosols. The results show that most of the 25 sites are dominated by PD and UI-D. Specifically, sites from Indonesia, Singapore, and a part of Malaysia are dominated by reflective aerosols like UI and UI-D; sites from Thailand, Philippines, Malaysia, and southern Vietnam are dominated by more absorbing aerosols like PD and UI-D; sites from northern Vietnam and Taiwan are dominated by coarse aerosol like PD and UI-D.

Using Mahalanobis Distance to classify aerosol in Southeast Asia based on AERONET-retrieved optical properties

By

Hans Jarett J. Ong

Thesis submitted to the Faculty of the Physics Department of the
Ateneo de Manila University, in partial fulfillment
of the requirements for the degree of
Bachelor of Science in Physics
2019

Advisory Committee:

Thesis Advisers: Nofel D. Lagrosas, Ph.D., Clint Dominic G. Bennett

Panelists: Joel T. Maquiling, Ph.D., Patricio P. Dailisan

This is to certify that the undergraduate thesis titled. **“Using Mahalanobis Distance to classify aerosol in Southeast Asia based on AERONET-retrieved optical properties”** and submitted by Hans Jarett J. Ong to fulfill part of the requirements for the degree BS Physics was successfully defended and approved on the 5th of April 2019.

Joel T. Maquiling, Ph.D.
Panel member

Patricio P. Dailisan
Panel member

Clint Dominic G. Bennett
Thesis adviser

Benjamin O. Chan, Ph.D.
Chair
Physics department

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Significance: the need to classify aerosols in Southeast Asia

The study of aerosol particles is important in the study of earth and climate systems because they impact the climate by influencing the radiative and thermodynamic properties of the atmosphere [1]. This is especially true in the warm tropical environment of Southeast Asia (SEA henceforth) where there are still a lot of gaps in knowledge and large uncertainties regarding the relationship between aerosol radiative properties and the atmosphere's thermodynamic properties. Moreover, the SEA region is also known to be very vulnerable to climate impacts. [2] Thus, a better understanding of the climate system, which includes the study of aerosol, is necessary for better disaster risk reduction. To better understand the SEA climate system, the 7-Southeast Asian Studies (7SEAS) Mission was created to "facilitate interdisciplinary research into the integrated SEA aerosol environment via grass roots style collaboration" [3]. The 7-SEAS program uses ground based, remotely sensed, and modeled data sets to study the aerosol-environment interaction in the region of Java through the Malay Peninsula and SEA to Taiwan.

One important yet difficult problem in the study of aerosol is determining aerosol types which is usually done using chemical sampling and analysis, but such chemical methods are usually expensive and time-consuming. Furthermore, there are also many places where chemical sampling data are limited or simply unavailable. In such cases, remote sensing provides an advantage over chemical sampling. This is because remote sensing systems are automated which enables them to continuously gather large amounts of data while requiring low maintenance. One such remote sensing system widely used in 7-SEAS is NASA's Aerosol Robotic Network (AERONET) (see Section 2.2 AERONET). The AERONET sun photometers can measure aerosol optical properties such as absorption, scattering, optical depth and aerosol size distributions. [4]

1.2 Objectives

The objectives of this research are the following:

1. to determine and classify aerosol types over SEA from AERONET measurements using Mahalanobis distance; and
2. to identify the spatial and temporal variation of these aerosol types.

1.3 Scope

Only SEA sites with at least one year's worth of data shall be included in this study for the seasonal variations to be apparent. At the time of this study, 25 AERONET sites from Indonesia, Malaysia, Singapore, Thailand, Vietnam, Philippines and Taiwan (see Figure 1) meet this requirement. The dataset includes data from the years 1998 to 2017 as shown in Table 1.

Table 1. List of AERONET sites included in this study. The sites chosen here all have at least one year's worth of level 2.0 data.

Site Name	Country	Start Date	End Date	No. of Inversion Points
Bac_Giang	Vietnam	Mar-2003	Dec-2009	884
Bac_Lieu	Vietnam	Mar-2003	Sep-2015	175
Bandung	Indonesia	May-2009	Feb-2017	323
Chiang_Mai_Met_Sta	Thailand	Sep-2006	Mar-2017	2552
Chiayi	Taiwan	Sep-2013	Apr-2017	713
Dongsha_Island	Taiwan	Apr-1998	Apr-2016	151
EPA-NCU	Taiwan	Jul-2006	Jul-2016	668
Jambi	Indonesia	Jul-2012	Aug-2015	18
Kuching	Malaysia	Aug-2011	Oct-2015	38
Lulin	Taiwan	Aug-2006	Aug-2016	29
Manila_Observatory	Philippines	Jan-2009	Feb-2016	173
Mukdahan	Thailand	Nov-2003	May-2010	1125
NCU_Taiwan	Taiwan	Apr-1998	Jul-2013	323
ND_Marbel_Univ	Philippines	Dec-2009	Jan-2016	29
NGHIA_DO	Vietnam	Dec-2010	Dec-2016	471
NhaTrang	Vietnam	Nov-2011	Dec-2014	201
Omkoi	Thailand	Feb-2003	Apr-2017	445
Palangkaraya	Indonesia	Jul-2012	Feb-2016	42
Pontianak	Indonesia	Jul-2012	Feb-2016	50
Silpakorn_Univ	Thailand	Aug-2006	Mar-2017	2982
Singapore	Singapore	Nov-2006	Dec-2016	230
Songkhla_Met_Sta	Malaysia	Jan-2007	Apr-2016	160
Taipei_CWB	Taiwan	Oct-2002	May-2016	647
Ubon_Ratchathani	Thailand	Oct-2009	Jul-2016	836
USM_Penang	Malaysia	Nov-2011	Aug-2016	257

Figure 1. This map shows the locations of the 25 AERONET sites used in this study. (This figure was made using Google Maps)

1.4. Limitations

There are 2 main limitations encountered in this study. The first one is due to the AERONET sun photometer not being able to gather data when it's raining which results in generally less data during the rainy months. This is demonstrated in the case of Manila (Figure 2) where there is a lower number of retrievals during months when rainfall is high or when the cloud cover is thick. The second limitation is inherent in the inversion algorithm (see Section 2.3 AERONET Retrieved optical properties). The inversion algorithm only works for retrievals with $AOT_{440} \geq 0.4$ [5] (AOT stands for Aerosol Optical Thickness, see Section 2.3 AERONET Retrieved optical properties) which reduces the number of available data especially on sites where there isn't that much aerosol loading.

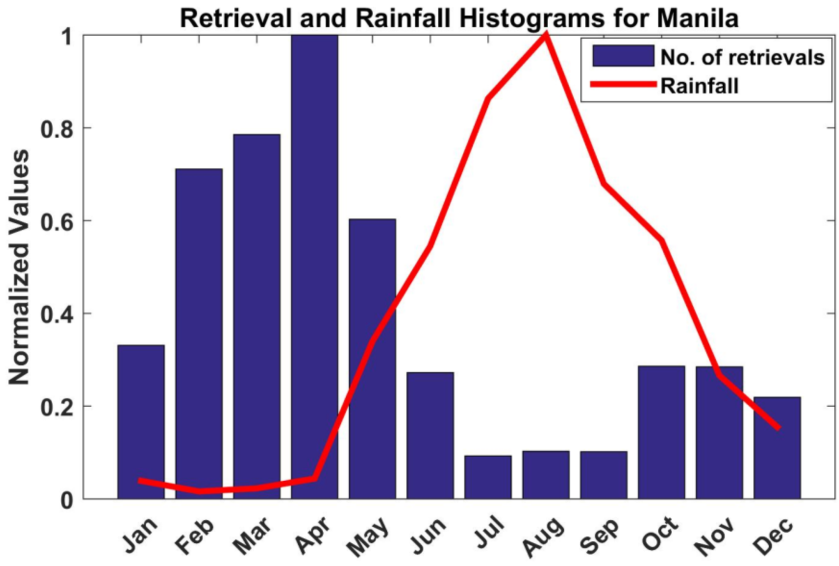


Figure 2. This is a graph comparing the monthly number of retrievals from AERONET's Manila Observatory site and the monthly average rainfall in Manila. The rainfall is measured in millimeters (mm), but both variables are normalized for better comparison. Rainfall data comes from *http://www.wunderground.com/*

This paper has 5 chapters. Chapter 2 is a review of the literature written about the topic which includes definition of aerosol types, description of the AERONET project, instrumentation of the AERONET sun photometer, AERONET data products, and aerosol classification used in other works. Chapter 3 is all about the methodology used in this paper; this chapter discusses the selection of the reference clusters, the definition and advantages of Mahalanobis distance, and the method used to clean/ filter the reference clusters. Chapter 4 is where is results are presented and discussed. To make the results more organized, the 25 sites are divided according to the following 4 regions: 1. Thailand and Malaysia; 2. Vietnam; 3. Taiwan; 4. Singapore, Indonesia, and Philippines. After discussing the results for each site, chapter 4 ends with a summary of the results for all 25 sites. Finally, chapter 5 gives the conclusions and recommendations of the paper.

Chapter 2

Review of Related Literature

2.1 Aerosols: Definition and Types

Aerosols, in general, are defined to be collections of solid or liquid particles suspended in the atmosphere excluding hydrometeors such as cloud droplets, ice crystals, raindrops, snowflakes, and graupel. [6] Individual aerosol particles are small so they are often studied in large concentrations. Common sources of aerosols are marine aerosols, desert dust, volcanic aerosols, biogenic aerosols, biomass burning aerosols, and aerosols from fossil fuel combustion. Table 2 summarizes the properties of these aerosol types.

Table 2. Summary of the different types of aerosol with their size ranges and common sources. [6]

Aerosol Type	**Typical Size**	**Common sources**
Marine	100nm – 10μm	Marine aerosols come from fine sea salt particles ejected into the atmosphere by the wind.
Desert/Mineral Dust	100nm – 10μm	Desert or Mineral Dust comes from soil particles ejected into the atmosphere by the wind.
Volcanic	1μm – 1mm	Volcanic aerosols come from volcanic eruptions where pulverized rocks and minerals are ejected into the atmosphere.
Biogenic	< 1μm or > 100μm	Biogenic aerosols come from plant and insect debris, pollen, spores, bacteria, and viruses.
Biomass Burning	< 1μm	Biomass burning aerosols come from the incomplete combustion of organic matter.
Urban Industrial	< 1μm	Urban industrial aerosols come from fossil fuel combustions common in cities and industrial areas.

2.2 AERONET

NASA's AERONET is a global network of sun photometers that provide continuous measurement of aerosol optical properties; Figure 3 shows the global distribution of AERONET sun photometers [https://aeronet.gsfc.nasa.gov/]. There are three data quality levels: levels 1.0, 1.5 and 2.0. The lowest data quality level, level 1.0 data, are raw, unprocessed data directly from the sun photometers' measurements without any form of processing. The next data quality level, level 1.5 data, is cloud-screened using a cloud-screening algorithm described elsewhere [7]. Finally, level 2.0 data cloud-screened and quality assured. Quality assurance is done by instrument tests and manual inspection using a visual data representation tool [5]. This makes level 2.0 data have the best data quality.

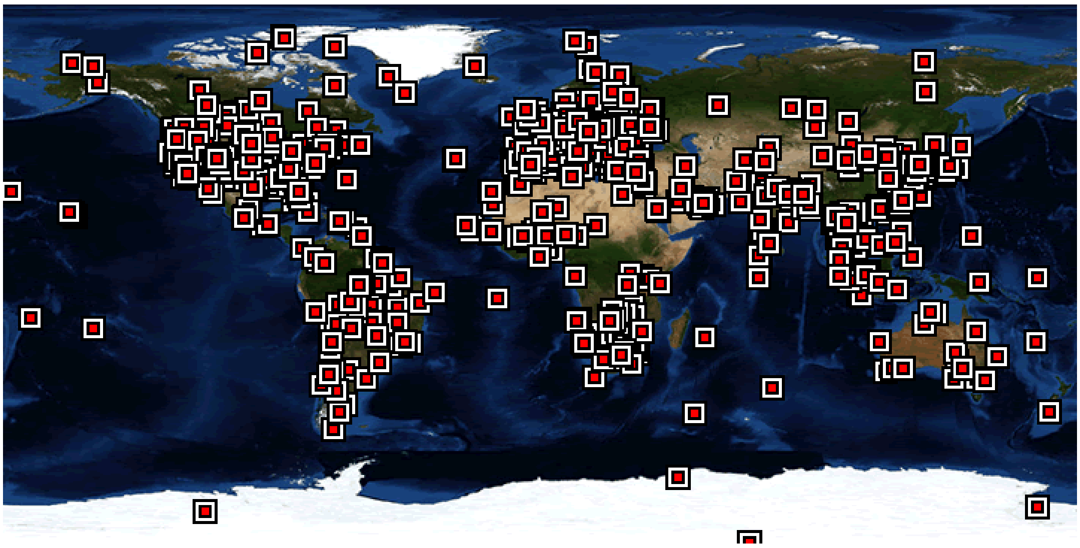


Figure 3. Global Distribution of AERONET sun photometers [https://aeronet.gsfc.nasa.gov/]

2.2 AERONET Instrumentation

The sun photometers used in AERONET are the CIMEL Electronique CE-318 sun-sky radiometer (Figure 4). The sun photometer has a 1.2° field of view, a 33-cm collimator to filter out stray light, a microprocessor which enables it to track the sun's location, and 2 detectors for measuring direct sun and sky radiance. The details of the specifications and operation of the sun photometer are described elsewhere [4]. The sun photometer is a passive remote sensing system, i.e., it relies on an outside light source (in this case, the sun) to measure aerosol optical properties.

The sun photometer has two modes of measurement: direct sun and sky radiance. Direct sun measurements are done by having the instrument pointed directly towards the sun. Direct sun measurements are performed every 15 minutes at wavelengths 340, 380, 440, 500, 675, 870, 940, and 1020 nm; this takes approximately 10 seconds per wavelength and this is done 3 times in 30-second intervals. On the other hand, sky radiance measurements are done with two sequences known as the almucantar and principal plane measurements. Almucantar measurements are taken at the elevation angle of the Sun and at varying azimuth angles (Figure 5a) while principal plane measurements are taken at the azimuth angle of the sun and at varying elevation angles (Figure 5b). Sky radiance measurements from various scattering angles can be used to deduce particle size distribution and phase functions.

Figure 4. The AERONET sun photometer installed in Manila Observatory.

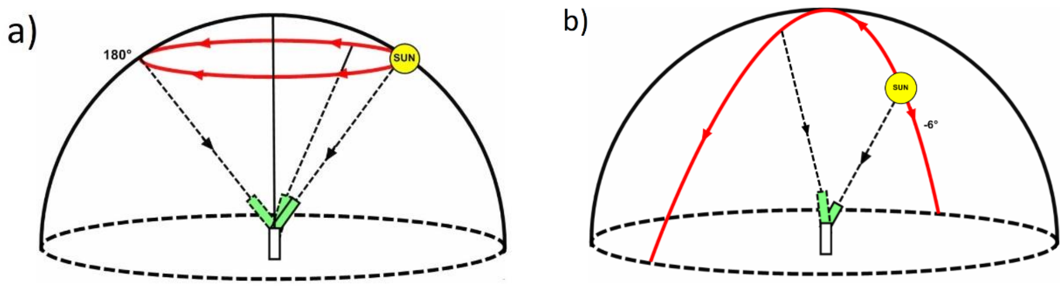


Figure 5. a) shows a schematic diagram of the sky radiance measurement along the Almucantar while b) shows the measurement along the Principal Plane [8].

2.3 AERONET Retrieved optical properties

Aerosol Optical Thickness (*AOT*) is a measure of extinction of light (from the sun in this case) due to aerosol. AERONET sun photometers calculate the *AOT* from the spectral extinction of the direct beam radiation according to the Beer-Lambert-Bouguer Law:

$$V_\lambda = V_{0\lambda} d^2 \exp(-\tau_\lambda m) \cdot t_y \tag{1}$$

where V is the digital voltage, V_0 is the extraterrestrial voltage, m is the optical air mass (which is approximately the secant of the zenith angle), τ is the total optical depth, λ is the wavelength, d is the ratio of the average to the actual Earth-Sun distance, and t_y is the transmission of the absorbing gasses. The AOT (or τ_a) is the τ minus the absorption by atmospheric gasses, water vapor, and the effects of Rayleigh scattering: [4]

$$\tau_a = \tau - \tau_{H_2O} - \tau_{Rayleigh} - \tau_{O_3} - \tau_{NO_2} - \tau_{CO_2} - \tau_{\mathrm{CH_4}} \tag{2}$$

The Angstrom Exponent (*AE*) can be derived from the *AOT* per wavelength. *AE* is defined to be the slope of the *AOT* with respect to the wavelength in a logarithmic scale:

$$\alpha = -\frac{d\ ln\tau_a}{d\ ln\lambda} \tag{3}$$

where α is the AE, τ_a is the AOT, and λ is the wavelength. [9] The AE is a particle size indicator where AE<1 suggests the dominance of coarse aerosols while AE≥2 suggests the dominance of fine aerosols. [10] The derivative of AE with the wavelength, α', is also a good indicator for particle size where $\alpha' > 0$ suggests the dominance of fine aerosols and $\alpha' < 0$ suggests the dominance of coarse aerosols. The α' is obtained from the second order polynomial fit of AOT vs wavelength in log-log space [10].

More parameters which describe particle size and absorption can be derived from direct sun and sky radiance measurements using the inversion algorithm and the spectral deconvolution algorithm (SDA). The inversion algorithm was developed using almucantar and principal plane measurements as inputs in a radiative transfer model [11]. It was further developed in succeeding works [12], [13], [14], [15], [16]. The inversion algorithm assumes that aerosol particles are partitioned into spherical and non-spherical components; the percentage of spherical particles is denoted by the asymmetry parameter ($g(\lambda)$). Aside from the sphericity, the algorithm also retrieves volume concentration (C_V), volume radius (r_V), and effective radius (r_{eff}) with their corresponding standard deviations (σ). Furthermore, the volume particle size distribution ($dV(r)/d\ lnr$) is retrieved for 22 logarithmically equidistant points (r_i) in the range $0.05\mu m \leq r \leq 15\mu m$. The single scattering albedo ($\omega(\lambda)$) retrieval, which assumes that a sunbeam is only reflected off a single particle, is the ratio of the scattering efficiency to the extinction efficiency. The real $(n(\lambda))$, imaginary $(k(\lambda))$ refractive indices and the single scattering albedo describe the scattering and absorbing properties of aerosol. It should be noted that retrievals of complex refractive index $(n + ik)$ require $AOT_{440} \geq 0.4$. Table 3 the data products derived from the inversion algorithm.

Table 3. Partial list of AERONET inversion products.

Optical Parameter	Symbol	Units/ Range of Values
Asymmetry Parameter	$g(\lambda)$	$0 \le g(\lambda) \le 1$
Effective Radius	r_{eff}	μm
Imaginary Refractive Index	$k(\lambda)$	$0.0005 \le k(\lambda) \le 0.5$
Real Refractive Index	$n(\lambda)$	$1.33 \le n(\lambda) \le 1.6$
Single Scattering Albedo	$\omega(\lambda)$	$0 \le \omega(\lambda) \le 1$
Standard Deviation	$\sigma; \sigma_f; \sigma_c$	μm
Volume Concentration	$C_V; C_{Vf}; C_{Vc}$	$\mu m^3/\mu m^2$
Volume Mean Radius	$r_V; r_{Vf}; r_{Vc}$	μm
Volume Size Distribution	$d\,V(r)/d\,lnr$	$\mu m^3/\mu m^2$

The spectral deconvolution algorithm (SDA) was developed by O'Neil et al. (2001 and 2003) [17], [18]. The SDA assumes the particle size distribution to be bimodal. This assumption enables it to separate the AOT (τ), AE (α), AE derivative (α') into fine and course modes. The Fine Mode Fraction (FMF) can be derived from the fine mode AOT (τ_f):

$$FMF = \frac{\tau_f}{\tau} \tag{4}$$

2.4 Existing aerosol classification methods

Specified clustering is a classification technique that uses predefined reference clusters as classes and subsequently assigns points to these clusters. The number of dimensions a cluster has would depend on the number of parameters of the data points. Data points are usually assigned to the reference clusters it is closest to in terms of some distance metric. The usual distance metrics used are the Euclidean distance and the Mahalanobis distance [19], [20], [21].

Previous work [22] used reference clusters from published work ([21]) and a 2-dimensional specified clustering method using AE (870-440nm) and SSA (440nm) as parameters and the scaled Euclidian (D) as metric:

$$D_i = \sqrt{(x - x_i)^2 + k(y - y_i)^2} \tag{5}$$

where x and y are the centers of the AE and SSA reference clusters respectively, and k is the scaling factor obtained by taking the ratio of the range of values of AE and SSA [22].

Specified clustering was used with 8 parameters: AE (491-863nm), SSA (491nm), difference in SSA at 863 and 491nm ($dSSA_{863,491}$), n(670nm), k(670nm), Absorption AE (AAE 491-863nm), % spheres, and volume FMF [21]. AERONET reference clusters are used to classify Polarization and Directionality of the Earth's Reflectances 3 (POLDER 3) products. The metric used was Mahalanobis Distance (D_M).

A recent work used Mahalanobis distance to classify AERONET products but with 5 parameters (AE, AAE, SSA, n, and k) [23]. The classification is used to all AERONET sites globally. The results showed promising results, however it leaves a lot to be improved. For instance, the classification method detected Marine aerosol were there shouldn't be marine aerosols like in Mexico City and in Manila Observatory. The problem of overestimating the mixed aerosol type is also encountered.

Chapter 3

Methodology

In this work, the methodology follows 3 main steps. Firstly, the reference clusters are created by selecting AERONET sites with a dominant aerosol type for certain months and by choosing the parameters from the AERONET derived optical properties. Secondly, the reference clusters are filtered for outliers and cross-validated to check for consistency. Finally, level 2 data from the 25 SEA AERONET sites are classified using Mahalanobis distance.

3.1 Establishing the reference clusters

The reference clusters used in this study are patterned after previous works [14], [24], [25], [21], [22]. There are 7 reference clusters, each of which would refer to an aerosol class (Table 4).

Table 4. Aerosol classes used as reference clusters in this study and the AERONET site they were taken from.

Class Name	Abbreviation	No. of points	Site	Period
Mineral Dust	MD	3239	Solar Village, Saudi Arabia	Mar-Jul (1999-2015)
Polluted Dust	PD	3471	Beijing, China	Whole Year (2000-2013)
Biomass Burning, Dark Smoke	BB-D	1620	Mongu, Nigeria	Aug-Nov (1995-2009)
Biomass Burning, White Smoke	BB-W	677	Alta Floresta, Brazil	Aug-Oct (1995-2013)
Urban/Industrial (Developed Economy)	UI	969	GSFC, Maryland, USA	Jun-Sept (1993-2013)
Urban/Industrial (Developing Economy)	UI-D	1075	Chen Kung Univ., Tainan, Taiwan	Whole Year (2002-2014)
Marine Aerosol	Marine	1513	Lanai, Hawaii, USA	Whole Year (1996-2004)

After choosing the sites which would be used as reference clusters, the next step is to select the parameters which would be used for classification. In selecting these parameters, one should note that aerosol types are determined primarily based on size and absorption. This means that quantitative parameters such as AOT shall be excluded because they give information on the amount of aerosol in the atmosphere rather than the type of aerosol. The percent sphericity parameter shall also be excluded because it produces huge error values for $AE > 1$ [26]. Thus, we are left with 5 candidate parameters: AE, SSA, AAE, n, and k.

Unlike the other reference clusters, the marine aerosol cluster had to be completed using a model because aerosols of this class rarely meet the inversion requirement $AOT_{440} \geq 0.4$ [5]. This is done by first getting the relative humidity (*RH*) values from *AE* using a model found in a previous research [27]. It should be noted that the model only worked for $AOT_{500} \leq 0.2$ and $0.1 \leq AE_{440,870} \leq 1$, so the data had to filtered accordingly. After getting *RH* from *AE*, *RH* values are inputted to a model called Optical Properties of Aerosol and Clouds (OPAC) which outputs *SSA*, *n*, and *k* [28], [29]. *AAE* is not defined for the marine aerosol class because marine aerosols are assumed to be non-absorbing, and this provide a problem [21]. Thus, we shall be excluding *AAE* as well.

When doing clustering, one should also exclude the parameters which are linearly correlated. Otherwise, there would be a bias in the classification towards the correlated parameters. One good measure for correlation is the Pearson R correlation coefficient which is a coefficient *r* which ranges from -1 to 1. Values of *r* close to 1 and -1 shows high direct or inverse correlations respectively [30]. Applying this to the 4 remaining parameters showed high inverse correlation between *SSA* and the imaginary refractive index *k* (see Table 5). This high inverse correlation between *SSA* and k can also be seen visually in their scatter plot shown in Figure 6. Thus, we must remove either *k* or *SSA*. We choose to remove the imaginary refractive index *k* because it has larger uncertainty values (30 – 50 % [11], [14]) compared to *SSA* (±0.03 [14]). Thus, we are left with 3 parameters: *AE*, *SSA*, and the real refractive index *n*.

Table 5. This table shows the Pearson R correlation coefficient between the *SSA* and *k* of each of the reference clusters. It shows that *SSA* and *k* have high inverse correlation.

Reference Cluster	Pearson R coefficient
BB-W	-0.91
PD	-0.86
UI-D	-0.97
UI	-0.96
BB-D	-0.93
MD	-0.86
Marine	-1.00

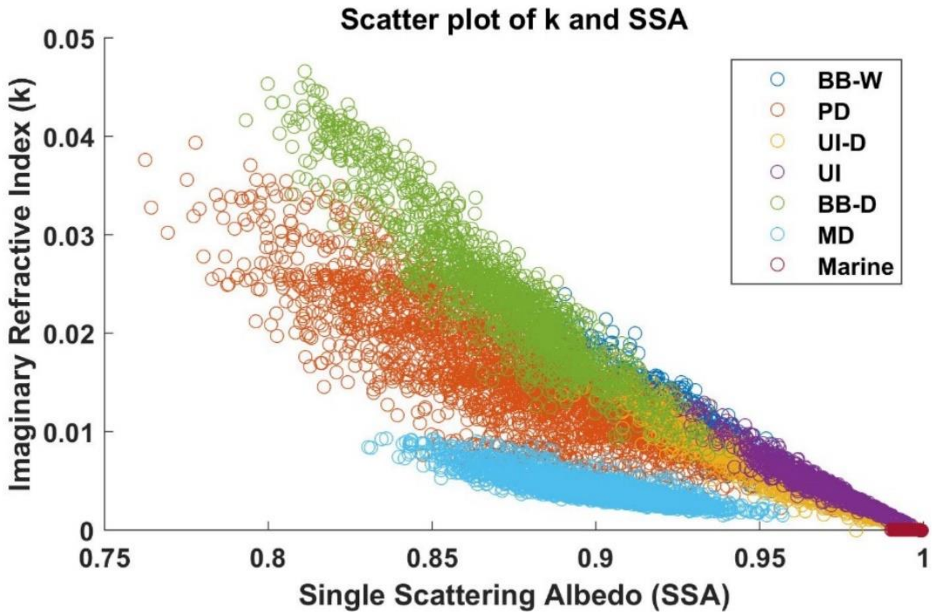


Figure 6. The scatter plot of *k* vs. *SSA* shows that these parameters have high inverse correlation.

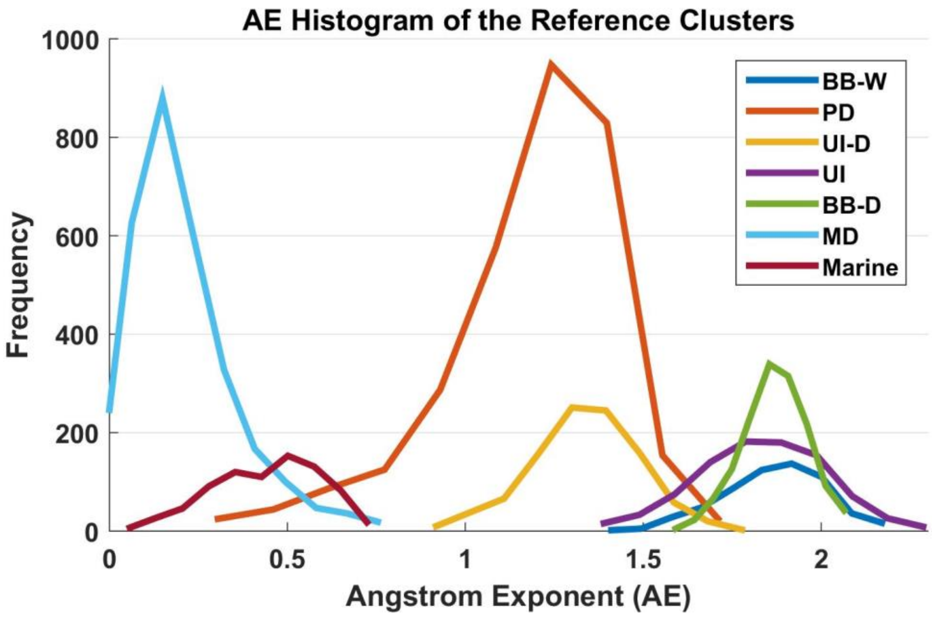


Figure 7. This shows the Angstrom exponent histogram of the reference clusters. From here we can infer the size composition of each aerosol class.

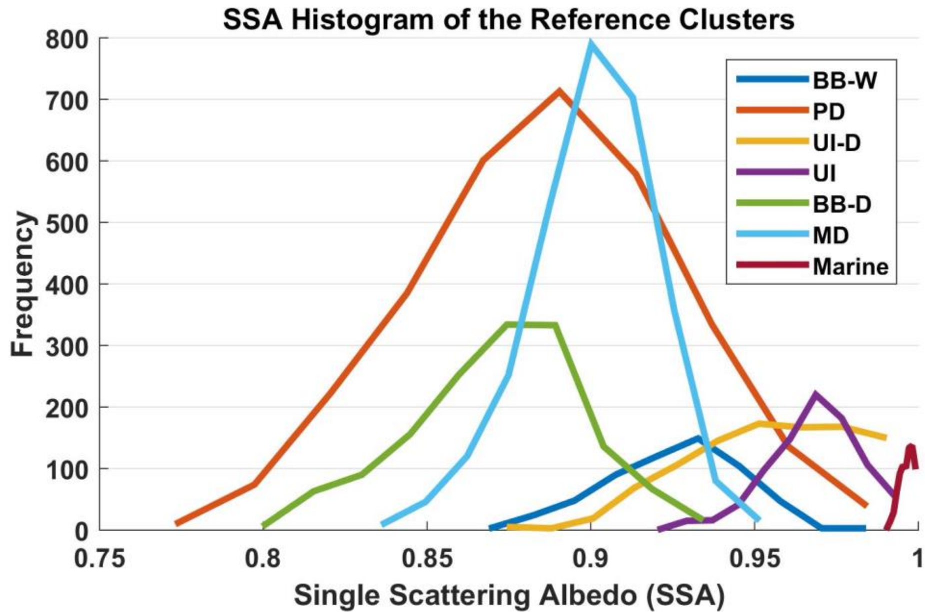


Figure 8. This shows the *SSA* histogram of the reference clusters. From this we can infer the absorptivity/reflectance of each aerosol class.

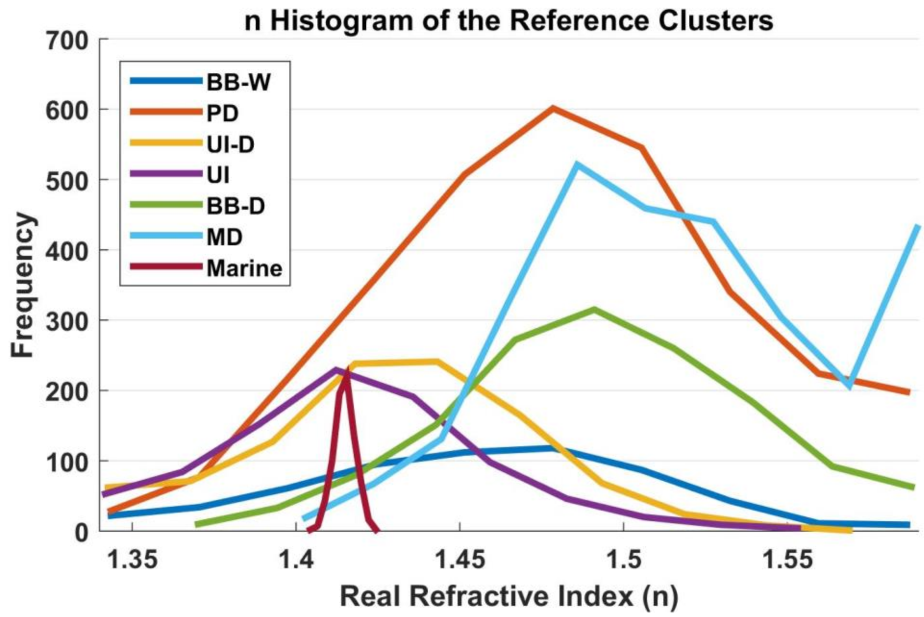


Figure 9. This shows the refractive index (n) histogram of the reference clusters. From here we may infer the water content of each aerosol class.

As mentioned in Section 2.3 AERONET Retrieved optical properties the Angstrom exponent is a measure of particle size where a smaller AE indicates the dominance of course aerosols while a larger AE indicates the dominance of fine aerosols. Figure 7 shows the AE histogram of the reference clusters; Here, it can be seen that mineral dust (MD) and marine aerosols (Marine) are composed of course particles, polluted dust (PD) and urban industrial for developing economies (UI-D) are composed of a mixture of fine and course particles, while biomass burning (BB-D and BB-W) and urban industrial (UI) are composed of fine aerosols.

The single scattering albedo (SSA) determines the reflectivity or absorptivity of the aerosol type. A low SSA indicates the dominance of absorbing aerosols while a high SSA indicated the dominance of reflective or scattering aerosols. Figure 8 shows the SSA histogram of the reference clusters. Here, it can be seen that BB-W, UI, UI-D, and the Marine aerosol classes are composed of highly reflective aerosols while the PD, BB-D, and MD classes are more absorbing compared to the former classes. We can also infer that the PD class is a mixture of absorbing and reflective aerosols.

The real refractive index (n) which has a range of 1.33<n<1.60 gives information about the water content of the aerosol [31]. A value of n close to 1.33 indicates that the aerosol class has high water content while higher values of n indicates different material. Figure 9 shows the histogram for the reference clusters. The histograms show UI, UI-D, and marine classes have high water content, BB-D and MD have lower content, and the PD and BB-W classes contain aerosols with a wide range of water content. Table 6 gives the mean of the reference clusters in AE-SSA-n space.

Table 6. Mean of the reference clusters.

Aerosol Class	Mean AE	Mean SSA	Mean n
BB-W	1.87	0.925	1.46
PD	1.19	0.886	1.48
UI-D	1.34	0.955	1.43
UI	1.83	0.968	1.42
BB-D	1.87	0.872	1.49
MD	0.21	0.901	1.52
Marine	0.44	0.996	1.41

3.2 Mahalanobis Distance

The Mahalanobis distance is a measure of the distance of a measurement from the mean of a cluster and is dependent on the covariance of the cluster – the formula for getting the Mahalanobis distance $D_M(\vec{x})$ is shown in (6 :

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \tag{6}$$

where $\vec{x}$ represents the position of the data point to be classified, $\vec{\mu}$ is the mean of the reference cluster and S is the covariance matrix of the cluster whose elements are given by

$$S_{ij} = \frac{1}{\mathrm{n}}\sum(x_i - \mu_i)(x_j - \mu_j) \tag{7}$$

where n is the total number of points in the cluster [19].

The Mahalanobis distance is particularly useful when working with clusters of high dimensionalities. This is because it considers the covariance of the clusters which results in two advantages: first, one can use any number of parameters even if they're not of the same units because the Mahalanobis distance is scale-invariant. Second, it considers the spread, obliqueness, and the orientation of the clusters. Intuitively, the Mahalanobis distance measures the number of standard deviations from the mean of the cluster. This contrasts with the absolute distance measured by the Euclidean distance (Figure 10). When using Euclidean distance, point A and point B would be equidistant to the cluster. On the other hand, when using Mahalanobis distance, point B would be closer to the cluster compared to A.

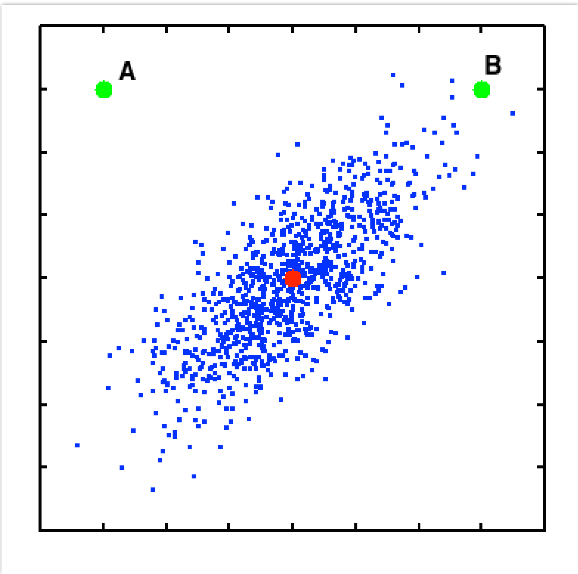


Figure 10. In this figure, point B is closer to the cluster in terms of the Mahalanobis distance. This demonstrates how the Mahalanobis distance takes into account the shape of the cluster.

3.3 Filtering the reference clusters

After setting up the reference clusters, it must be filtered for outliers. Filtering can be done by excluding points of a certain Mahalanobis distance (D_M) away from their clusters. Of course, there is a trade-off between the number of points to be included in the cluster and the accuracy we can get – i.e. smaller clusters would have less overlap which would result in better accuracies. To better understand this trade-off, a plot of the cumulative distribution function and the accuracies of the clusters with respect to D_M are shown in Figure 11. The cumulative distribution function can be calculated by simply getting the number of points of Mahalanobis distance 0 to D_M of their respective clusters (this is then normalized to get the percentage). The accuracy is computed using a k-fold cross-validation with k=10 which is done in the following steps: First, the data is partitioned randomly into 10 "folds". Then, cross-validation will be done 10 times where in each round one of the folds will be used as the testing set while the rest will be used as the training set. Finally, for each round, the accuracy is taken by getting the ratio of those correctly classified to the total number of points – the final accuracy is the average accuracy for the 10 rounds [32]. Since the partitioning is done randomly, the results are expected to vary a little for every trial. The plot of the accuracy in Figure 11 is the average of 5 trials. From Figure 11, we see that the critical point (where the two plots intersect) is at around $D_M = 2$. At this point, we get an accuracy of ~64% and we retain ~59% of the data points. Figure 12 shows the resulting clusters after filtering.

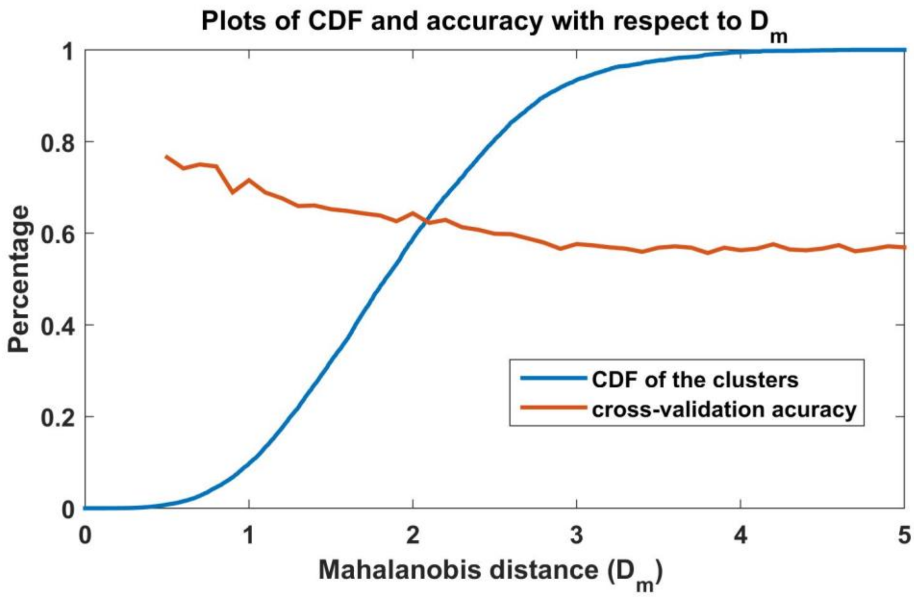


Figure 11. The plot of the normalized Cumulative Distribution Function and the accuracy from the 10-fold cross validation vs. D_M. Here, we see the trade-off: a smaller D_M would result to smaller number of points but will have better accuracy, and vice versa.

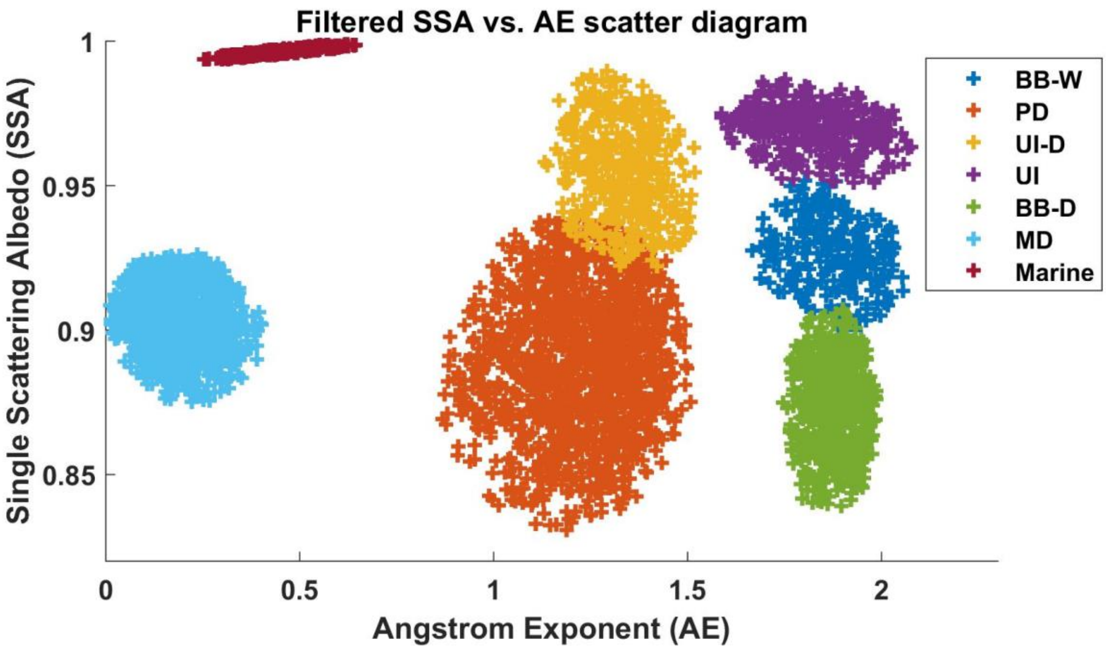


Figure 12. Scatter plot of the Single Scattering Albedo (*SSA*) vs. the Angstrom Exponent (*AE*) after filtering. Here we see that the clusters have good separation.

The optical properties of each of the 7 classes may be inferred qualitatively from Figure 12. Note that the AE describes particle size while SSA describes absorption/scattering. Figure 12 shows that the marine class comprises of coarse, reflective particles; the MD class comprises of coarse, slightly reflective particles; the PD class comprises of a mixture of fine and coarse particles with slight reflectivity; the UI-D and UI classes both comprise of reflective particles but the UI class has finer particles; the BB-W and BB-D classes both comprise of fine particles but the BB-D class is more absorbing than the BB-W class.

Chapter 4

Results and Discussion

In this chapter, the results of the classification will be presented and discussed. For the sake of organization, the results shall be arranged according to the following regions: Thailand and Malaysia; Vietnam; Taiwan; Singapore, Indonesia, and Philippines.

4.1 Aerosol types over Thailand and Malaysia

Descriptions of the different AERONET site locations in Thailand and Malaysia are presented in Table 7. Most of northern Thailand (which includes the sites ChiangMaiMetSta, Mukdahan, and Silpakorn) has a tropical wet and dry climate. This kind of climate has distinct wet and dry seasons but generally has less precipitation compared to a tropical monsoon climate. The wet season is brought about by the southwest monsoon which brings precipitation from the Indian Ocean from June to November. In contrast, the northeast monsoon brings dry winds from China and the north Pacific during from December to May. On the other hand, southern Thailand (SongkhlaMetSta) and Malaysia (Kuching and USMPenang) have tropical rain forest climates. This type of climate is characterized by the lack of a dry season and high precipitation all year round. [22]

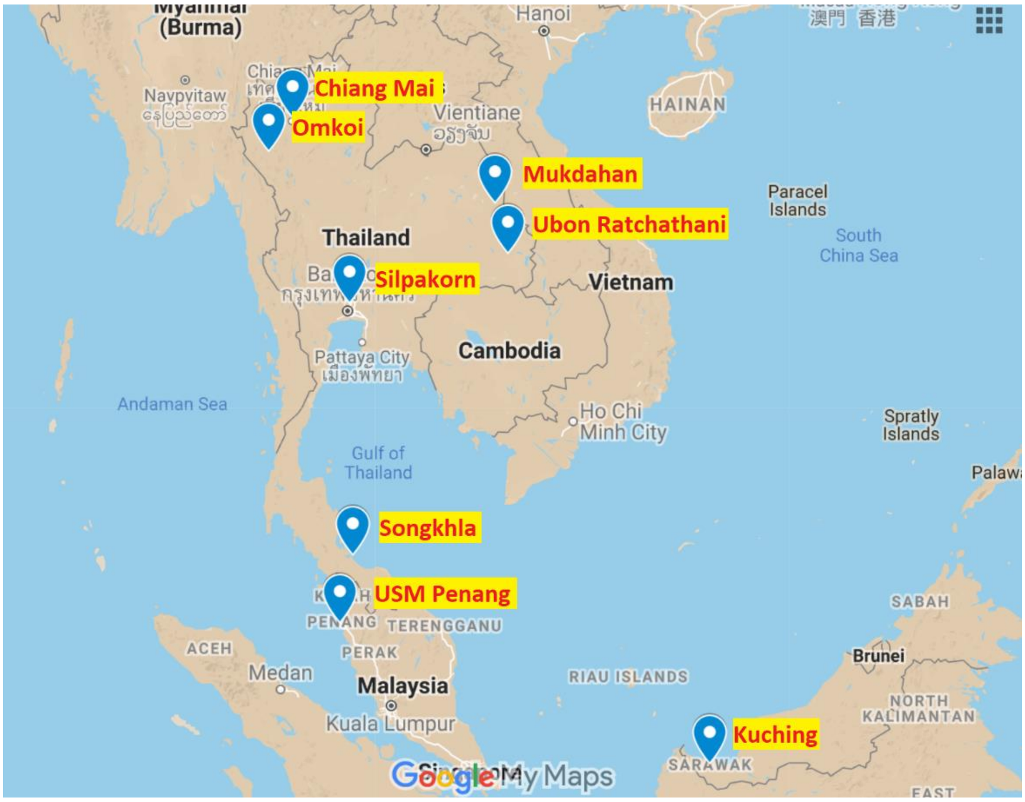


Figure 13. AERONET sites in Thailand and Malaysia (Map created using Google Maps).

The results over Thailand and Malaysia show that the northern sites – Chiang Mai, Omkoi, Silpakorn, Mukdahan, and UbonRatchathani (refer to Figure 13) – show similar trends in aerosol type. They exhibit data scarcity in the months of June to December which is due to these months being the rainy months in this region. Also, these northern sites are generally dominated by polluted dust followed by biomass burning for the months of January to May – this corresponds with the fact that biomass burning in the Indochina region peaks during these months according to MODIS satellite fire hotspot count. The biomass burning in this region is mostly anthropogenic and is usually due to agricultural activities such as agricultural waste disposal and land clearing [3]. On the other hand, the southern sites – Songkhla and USM Penang – have data all year round, although the number of data points are fewer from May to December; most of these aerosols are polluted dust and there is very few biomass burning compared to the northern sites. Finally, the southernmost site, Kuching, is located far from the other sites which is why it has a different trend.

Table 7. Description of the instrument location for each AERONET site in Thailand and Malaysia.

Site	Country	Description (from http://aeronet.gsfc.nasa.gov)
ChiangMaiMetSta	Thailand	The instrument is on the roof of a meteorological building near Chiang Mai Airport.
Mukdahan		The instrument is on top of a building ~10km from the town of Mukdahan and the Mekong River. The surrounding land is a mixture of forest and agriculture mostly dominated by rice and sugar cane fields.
Omkoi		The instrument is in a field in Omkoi which is a rural area in the Chiang Mai province and is mostly covered by vegetation.
SilpakornUniv		The instrument is on the roof of the Department of Physics building at Silpakorn University located in the urban city of Bangkok.
SongkhlaMetSta		The instrument is on top of a meteorological building in Songhkla City- a fishing town in Southern Thailand near Malaysian border.
UbonRatchathani		The instrument is on the roof of a meteorological building near the Ubon Ratchathani Airport.
Kuching	Malaysia	The instrument is mounted on the observation tower of the Kuching Meteorological Station located at the Kuching International Airport.
USMPenang		The instrument is in the Universiti Sains Malaysia (USM) main campus in Penang which is a few kilometers from the industrial town of Bayan Lepas.

Chiang Mai Meteorological Station

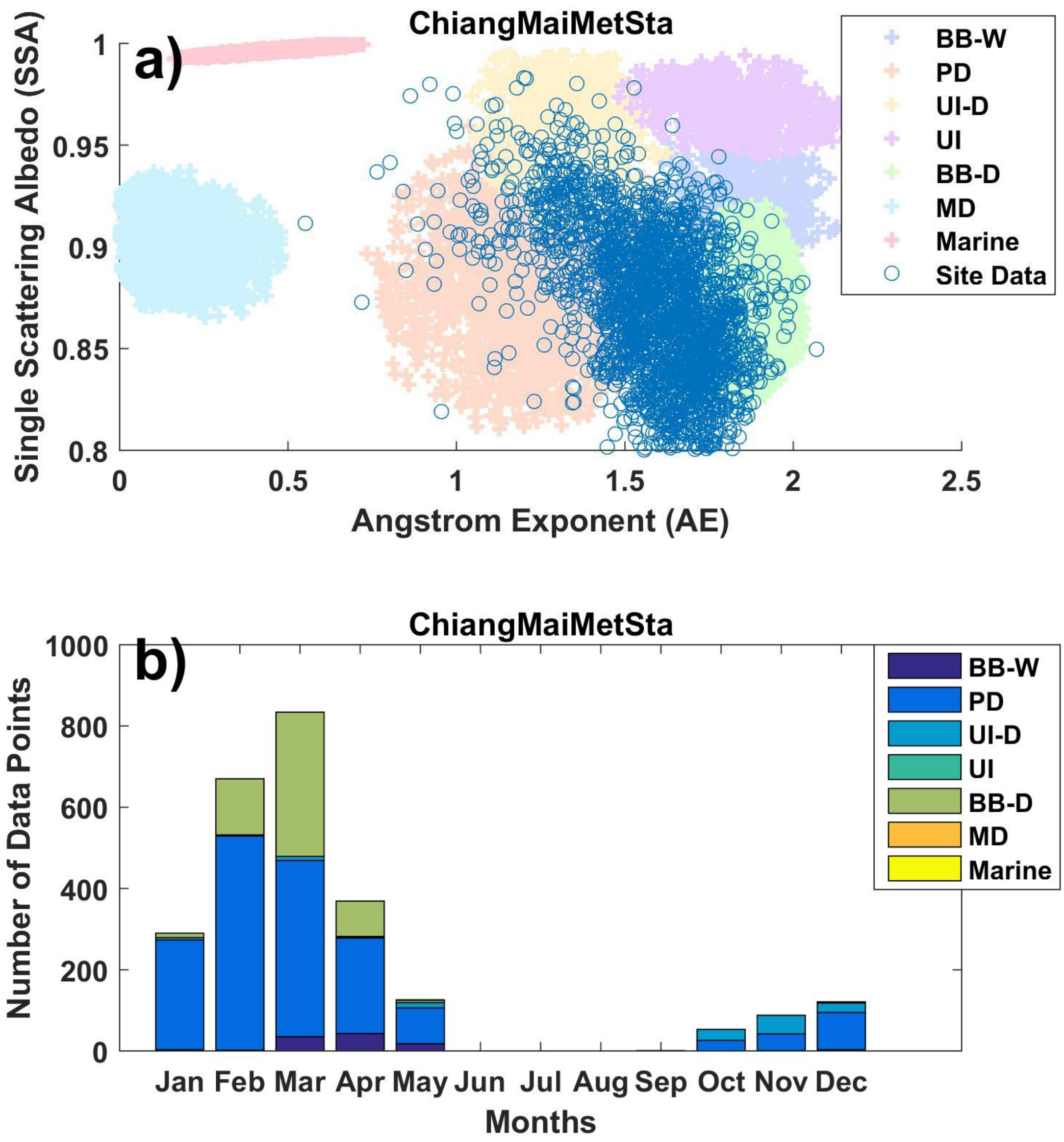


Figure 14. a) Scatter plot for the data points from Chiang Mai Meteorological Station. b) Monthly distribution of aerosol types in Chiang Mai Meteorological Station.

Figure 14 (a) shows that Chiang Mai is comprised mostly of fine aerosol. In term of absorption, the aerosols in this site are more absorbing than scattering. The monthly distribution (Figure 14 (b)) shows the dominance of polluted dust – this may be because the instrument is located near an airport. Furthermore, there are also traces of biomass burning from

January to May. The scarcity of data from June to September corresponds with the rainy months in Chiang Mai.

<u>Mukdahan</u>

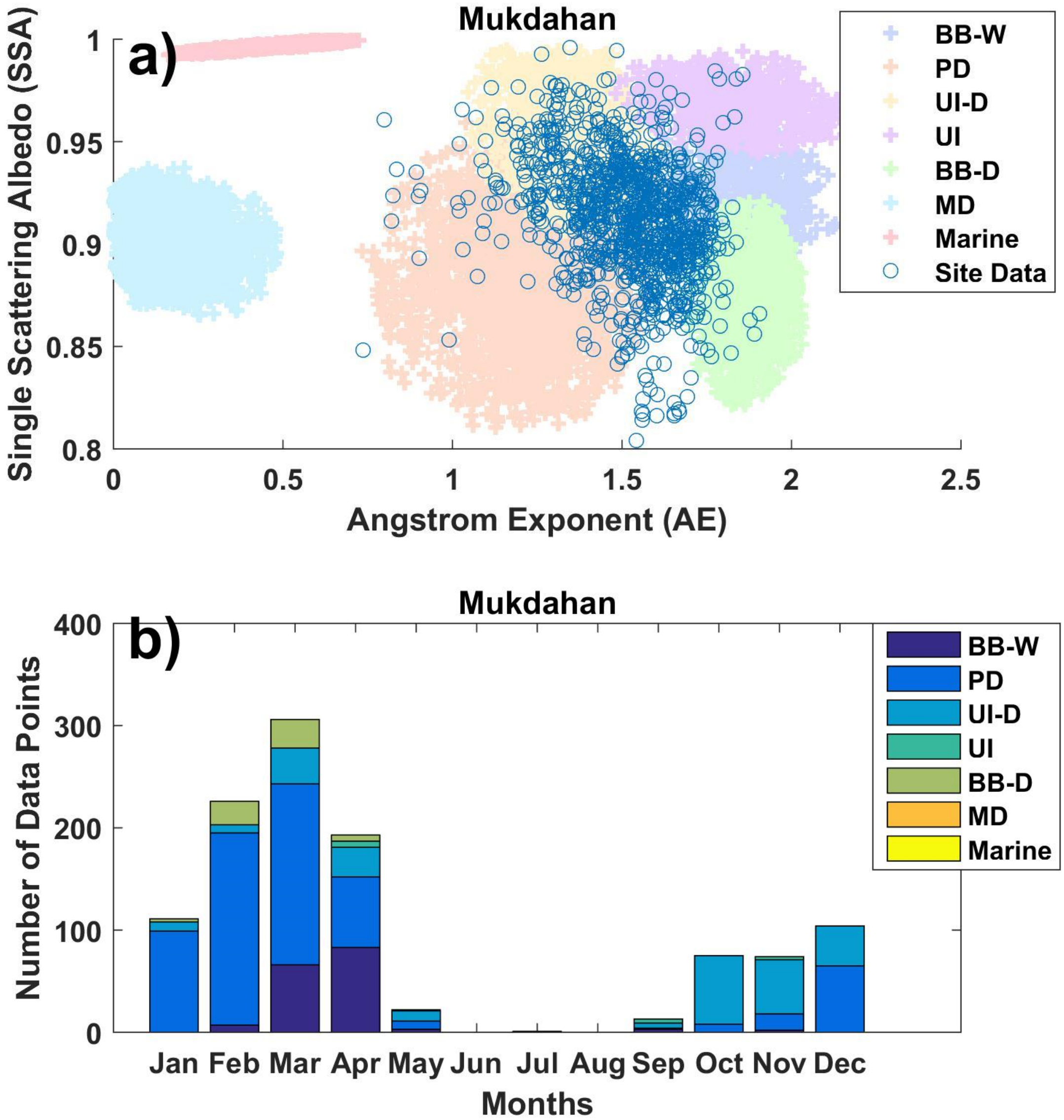


Figure 15. a) Scatter plot for the data points from Mukdahan. b) Monthly distribution of aerosol types in Mukdahan.

Figure 15 (a) shows that Mukdahan is comprised of a mixture of fine and coarse aerosols. In term of absorption, the aerosols in this site are a mix of absorbing and scattering aerosols. The monthly distribution (Figure 15 (b)) shows the dominance of polluted dust from January to April. The biomass burning detected from January to April may be attributed to agricultural activities during these months. The scarcity of data is due to the rainy months restricting data acquisition.

<u>Omkoi</u>

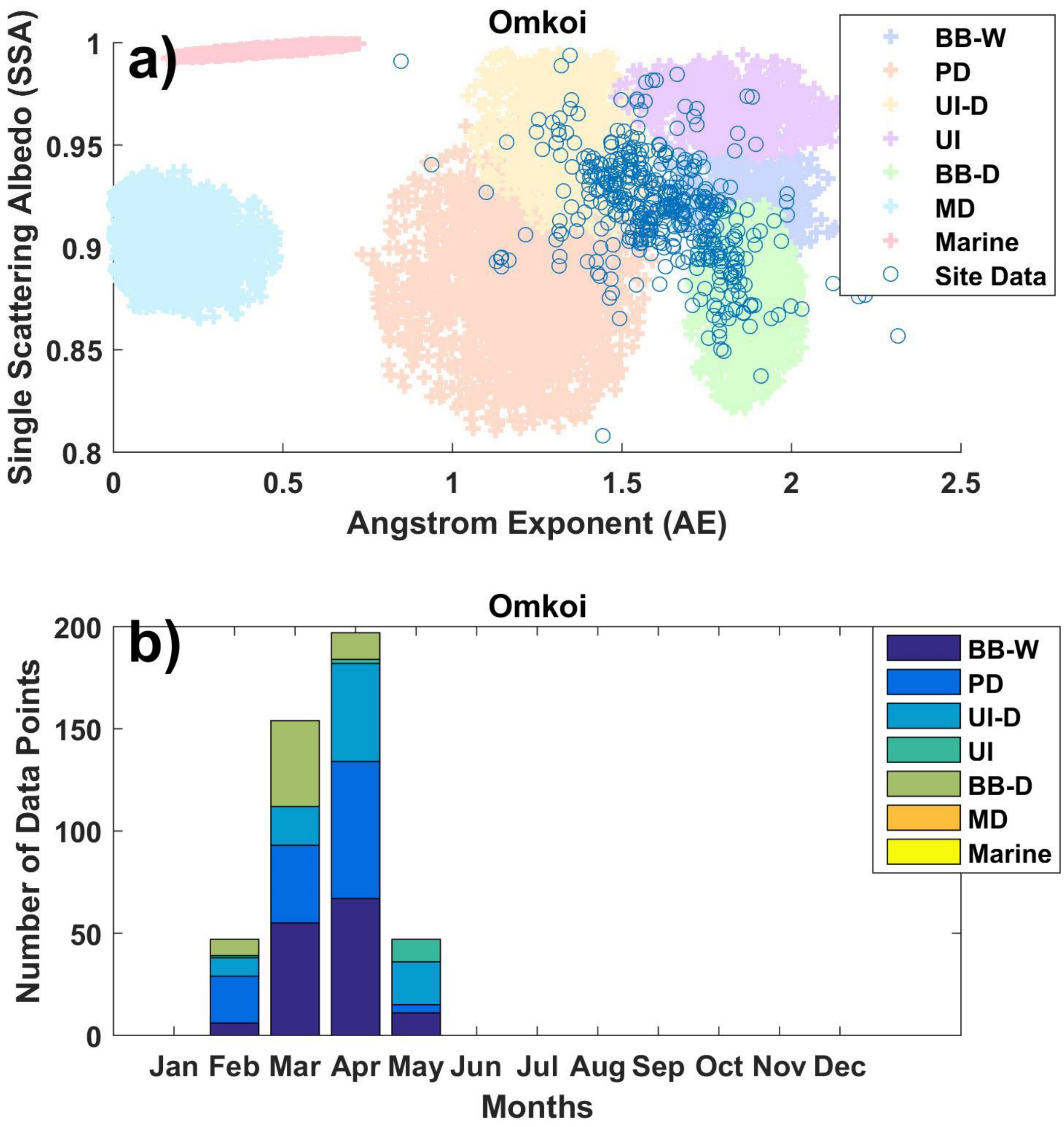


Figure 16. a) Scatter plot for the data points from Omkoi. b) Monthly distribution of aerosol types in Omkoi.

Figure 16 (a) shows that Omkoi is comprised of a mixture of fine and coarse aerosols leaning more towards fine. The aerosols in this site mostly scattering aerosols. The presence of biomass burning in the monthly distribution (Figure 16 (b)) may be attributed to burning vegetation or agricultural activities. The presence of polluted dust and urban industrial (developing) is a bit odd since the instrument is far from major sources of these aerosols.

<u>Silpakorn University</u>

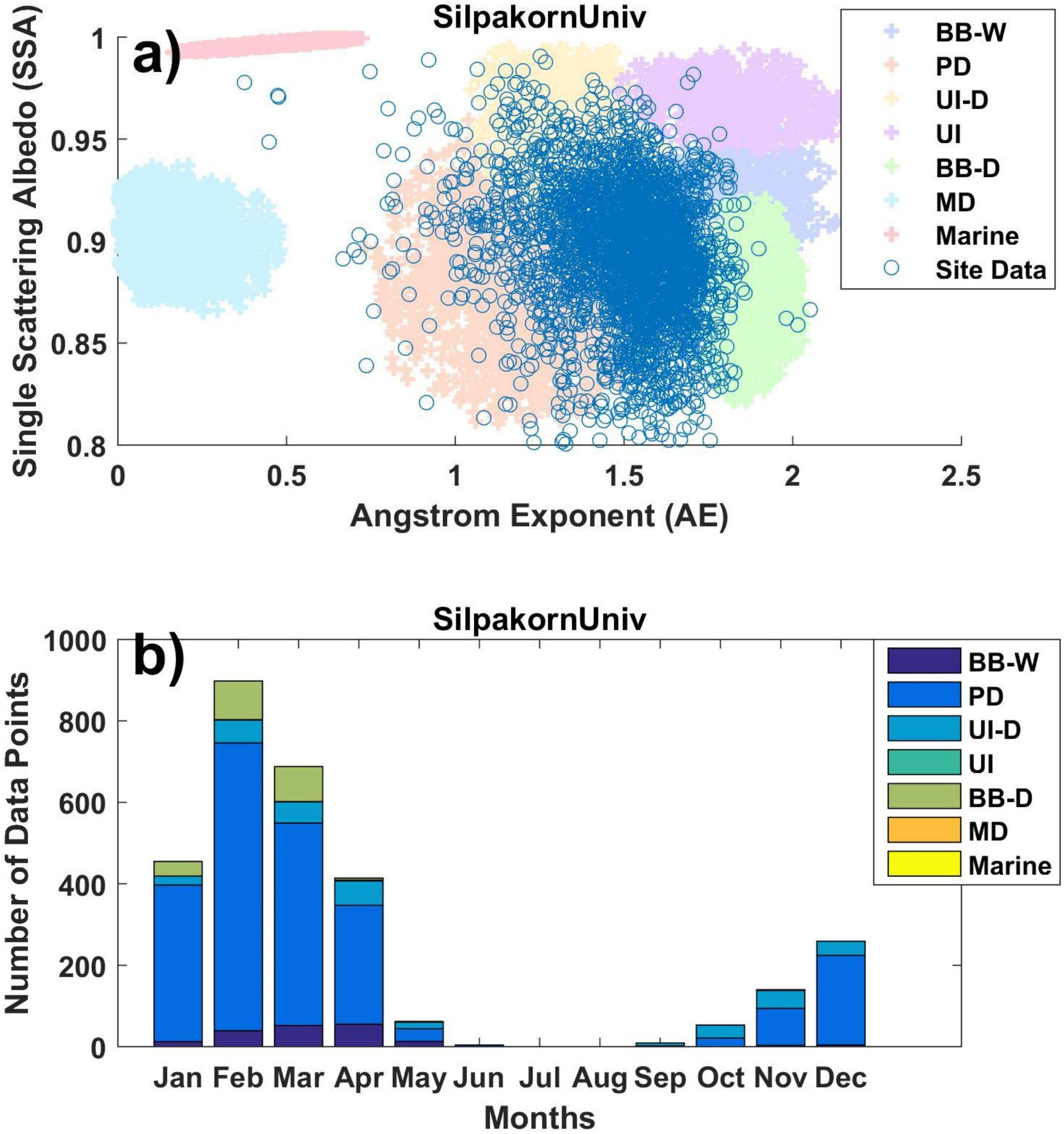


Figure 17. a) Scatter plot for the data points from Silpakorn University. b) Monthly distribution of aerosol types in Silpakorn University.

Figure 17 (a) shows that Silpakorn University is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are a mixture of scattering and absorbing aerosols. The monthly distribution (Figure 17 (b)) shows the dominance of polluted dust for the months with available data. There are also traces of biomass burning in the months of January to May – these, again, may be attributed to agricultural activities.

<u>Songkhla Meteorological Station</u>

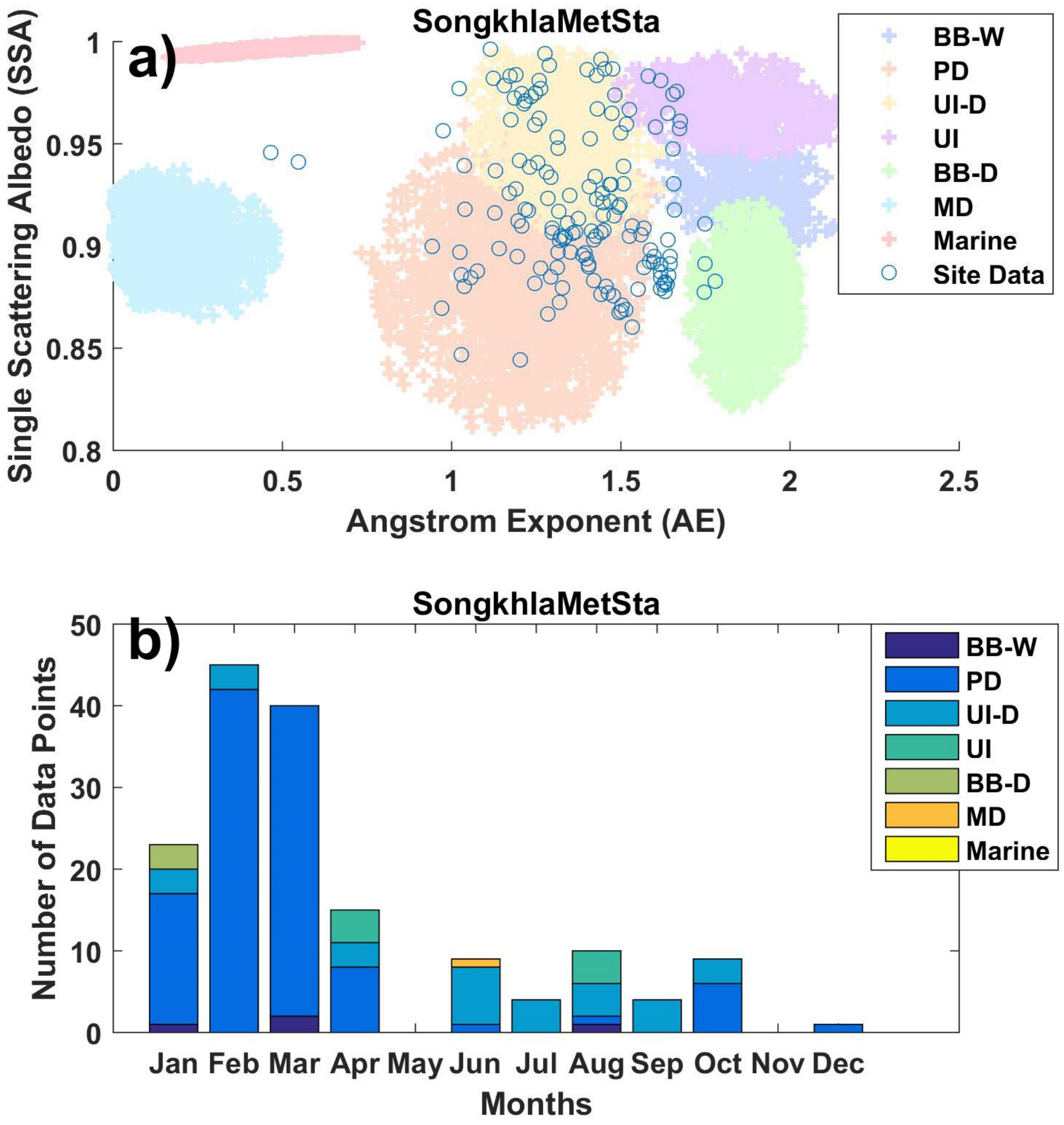
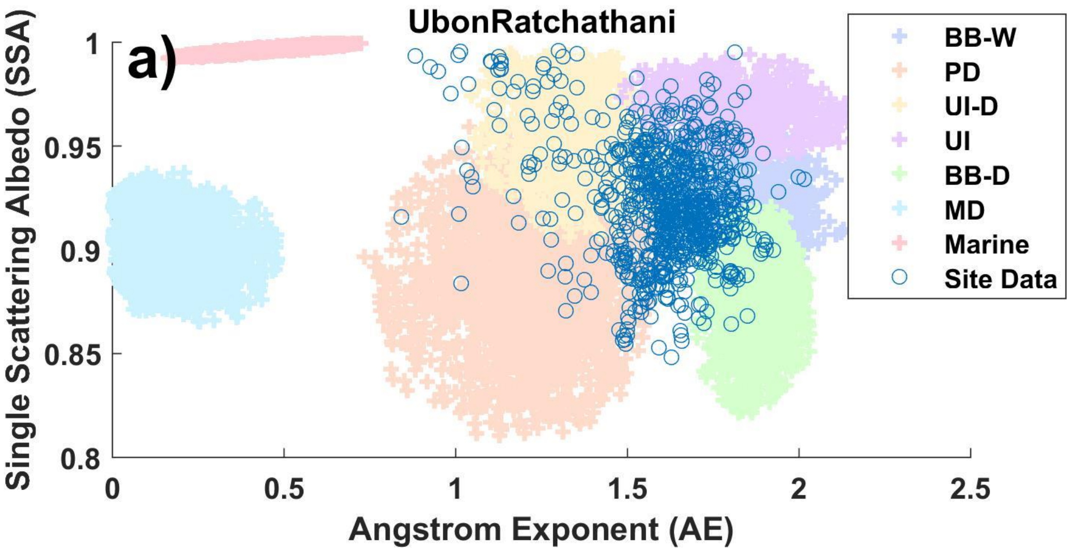


Figure 18.a) Scatter plot for the data points from Songkhla Meteorological Station. b) Monthly distribution of aerosol types in Songkhla Meteorological Station.

Figure 18 (a) shows that Songkhla Meteorological Station is comprised of a mixture of coarse aerosols. The aerosols in this site are a mixture of slightly scattering and highly scattering aerosols. The monthly distribution (Figure 18 (b)) shows the dominance of polluted dust from January to April and the dominance of urban industrial (developing economy) from June to September.

<u>Ubon Ratchathani</u>



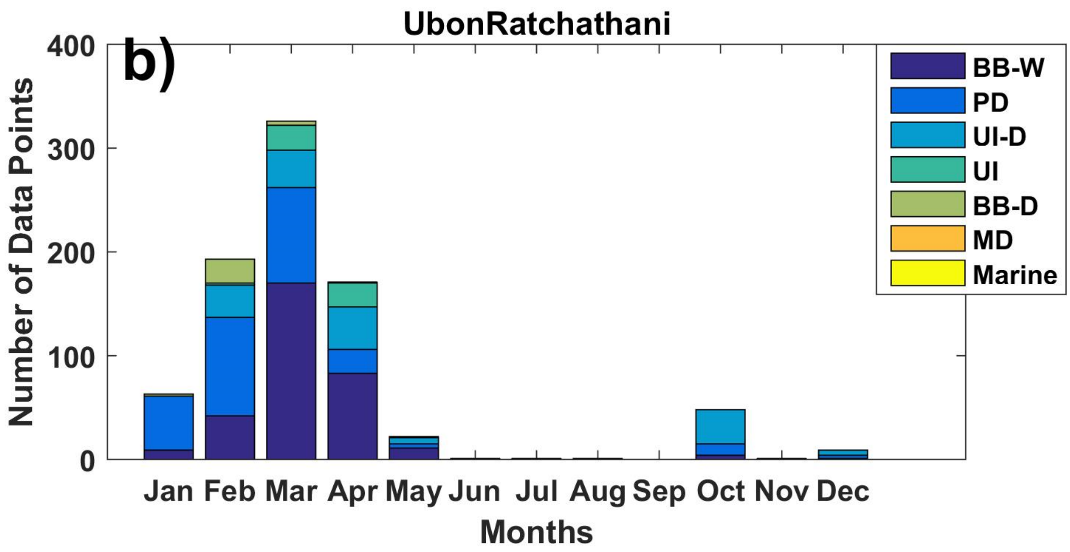


Figure 19. a) Scatter plot for the data points from Ubon Ratchathani. b) Monthly distribution of aerosol types in Ubon Ratchathani.

Figure 19 (a) shows that Ubon Ratchathani is comprised mostly of fine aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 19 (b)) shows that January to May is dominated by biomass burning white smoke which may be attributed to agricultural activities near the site or from wind transport. There are also polluted dust and urban industrial (developing economy) present – these can be attributed to the site being located near an airport.

<u>Kuching</u>

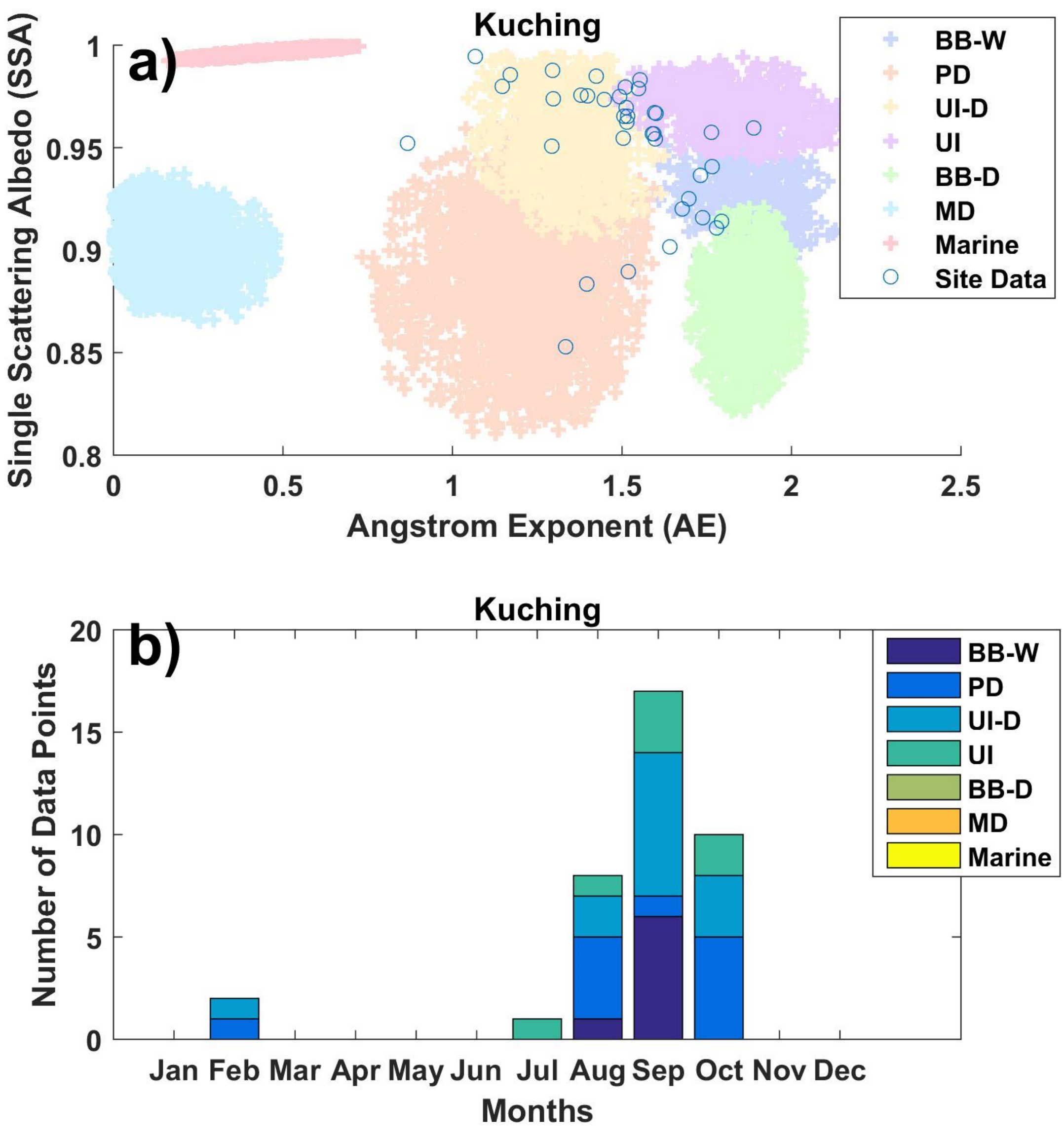


Figure 20. a) Scatter plot for the data points from Kuching. b) Monthly distribution of aerosol types in Kuching.

Figure 20 (a) shows that Kuching, despite having very few data points, is comprised of a mixture of fine and coarse aerosols that are mostly reflective. The monthly distribution (Figure 20 (b)) shows that data is only available in the months of February and July to October. The presence of urban industrial and polluted dust during these months may be attributed to the emissions from the airport near the site.

<u>USM Penang</u>

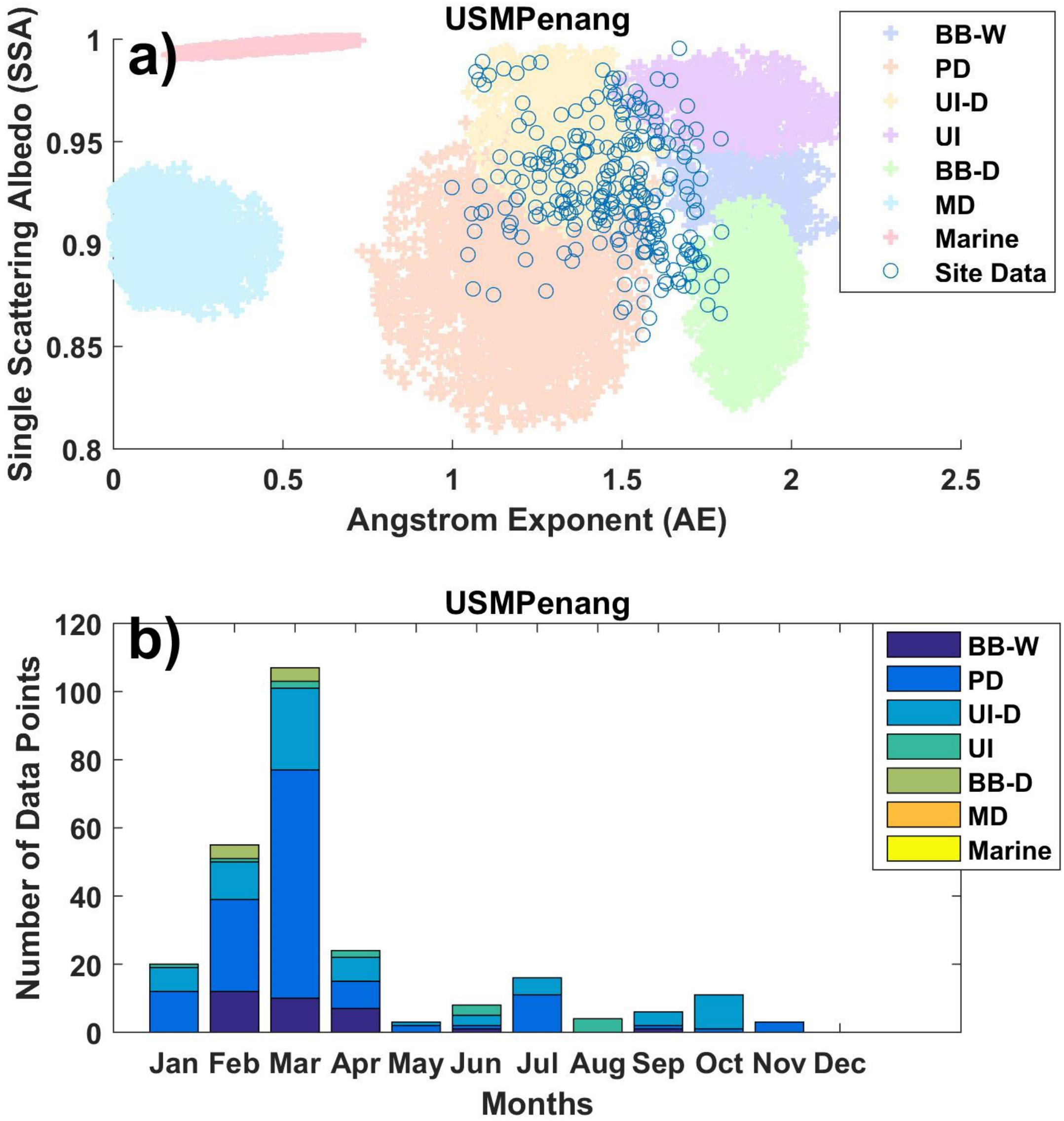


Figure 21. a) Scatter plot for the data points from USM Penang. b) Monthly distribution of aerosol types in USM Penang.

Figure 21 (a) shows that USM Penang is comprised of a mixture of fine and coarse aerosols leaning more towards fine. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 21 (b)) shows that January to April is dominated by polluted dust but we also see some biomass burning during these months. Also, urban industrial is consistently present all throughout the year – these may have been transported from the industrial town of Bayan Lepas to Penang.

4.2 Aerosol types over Vietnam

Descriptions of the different AERONET site locations in Vietnam are presented in Table 8. Vietnam's climate is divided in to tropical and temperate zones. This kind of climate is strongly influenced by the monsoons. Vietnam can experience dry seasons as well as periods of high rainfall and high humidity. Southern Vietnam has two distinct seasons: the cold season from November to April; and the warm season from May to October. On the other hand, northern parts of Vietnam have 4 seasons with warm summers and cold winters. [33]

Figure 22. AERONET sites in Vietnam (Map created using Google Maps).

Table 8. Description of the instrument location for each AERONET site in Vietnam.

Site	Country	Description (from http://aeronet.gsfc.nasa.gov)
BacGiang	Vietnam	The instrument is on the roof of the Bac Giang Institute of Geophysics in Bac Giang which is a province that lies in the Red River Delta.
BacLieu		The instrument is on the roof of the Bac Lieu Geophysics Observatory. Bac Lieu is a coastal province located in the Mekong Delta region.
NGHIADO		The instrument is on a roof of a building at the Vietnam Academy of Science and Technology in the Hanoi capital.
NhaTrang		The instrument is at the roof of the R&D for Sea Resources Area in Hon Chong, Nha Trang city. The site is located at about 100 meters from the sea.

Bac Giang

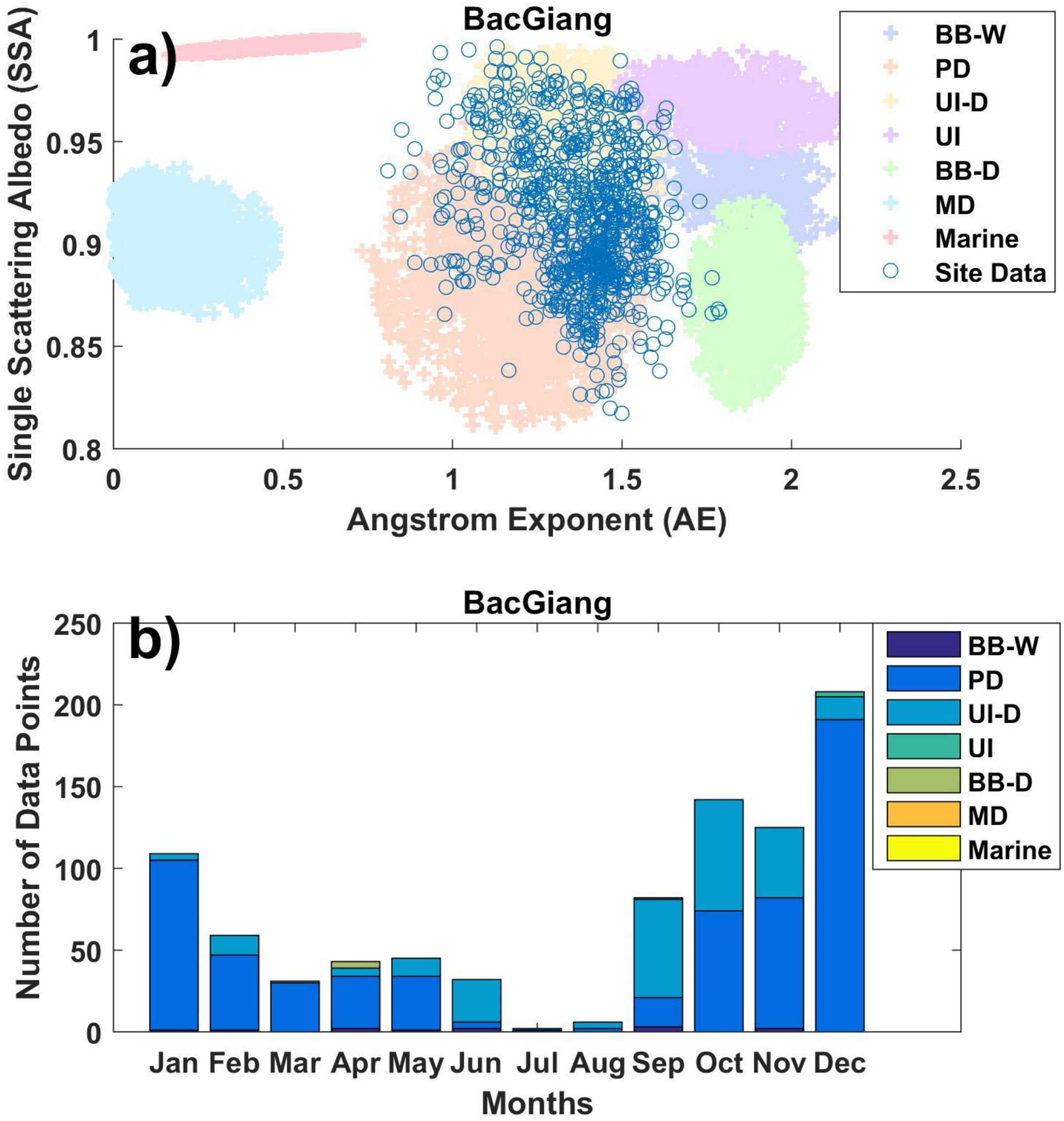


Figure 23. a) Scatter plot for the data points from Bac Giang. b) Monthly distribution of aerosol types in Bac Giang.

Figure 23 (a) shows that Bac Giang is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are a mixture of absorbing and scattering aerosols. The monthly distribution (Figure 23 (b)) shows that this site is dominated by urban industrial (developing economy) and polluted dust all year round.

Bac Lieu

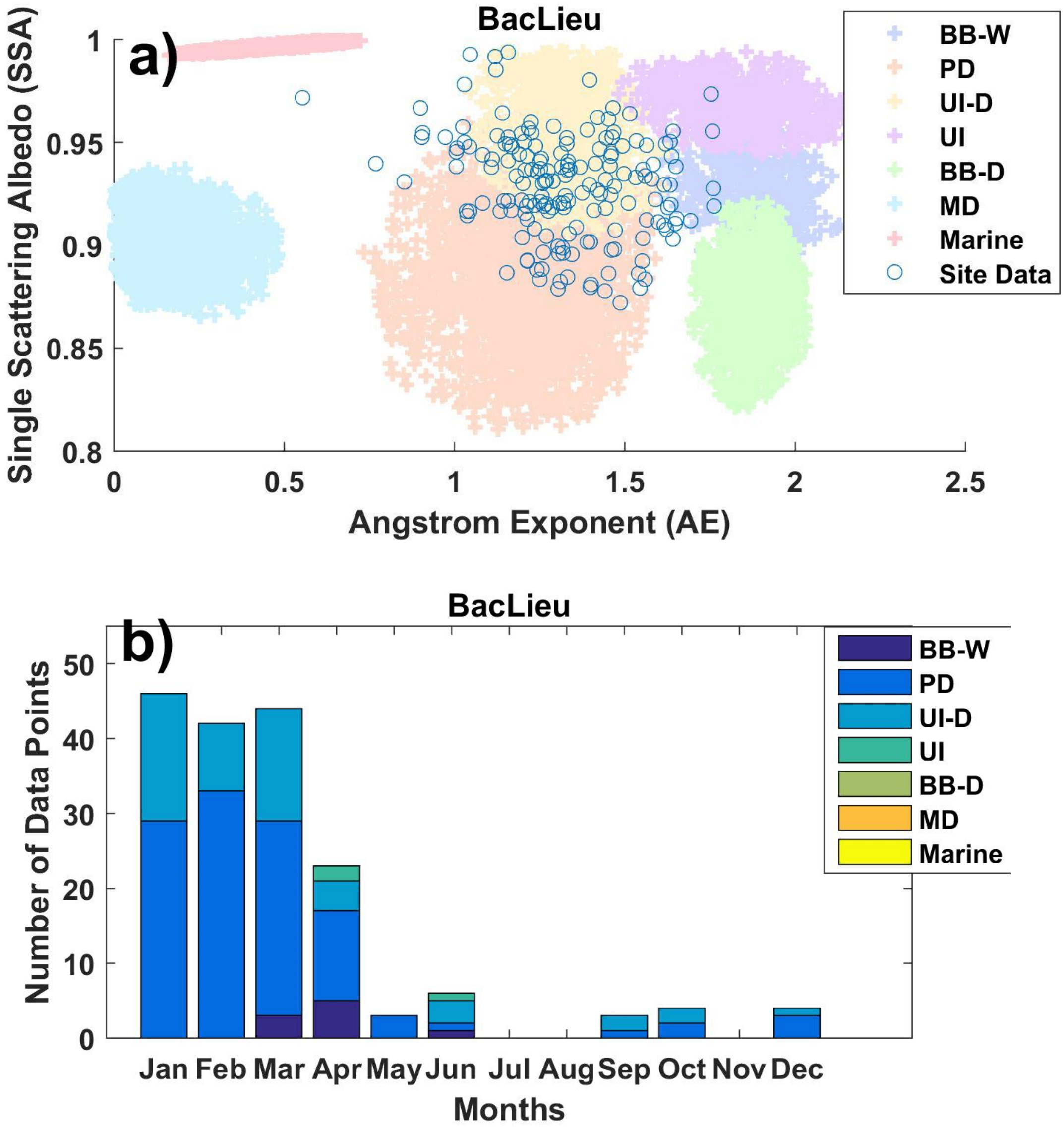


Figure 24. a) Scatter plot for the data points from Bac Lieu. b) Monthly distribution of aerosol types in Bac Lieu.

Figure 24 (a) shows that Bac Lieu is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 24 (b)) shows that most of the data points are present from January to April which is dominated by polluted dust and urban industrial (developing economy) during these months.

Nghia Do

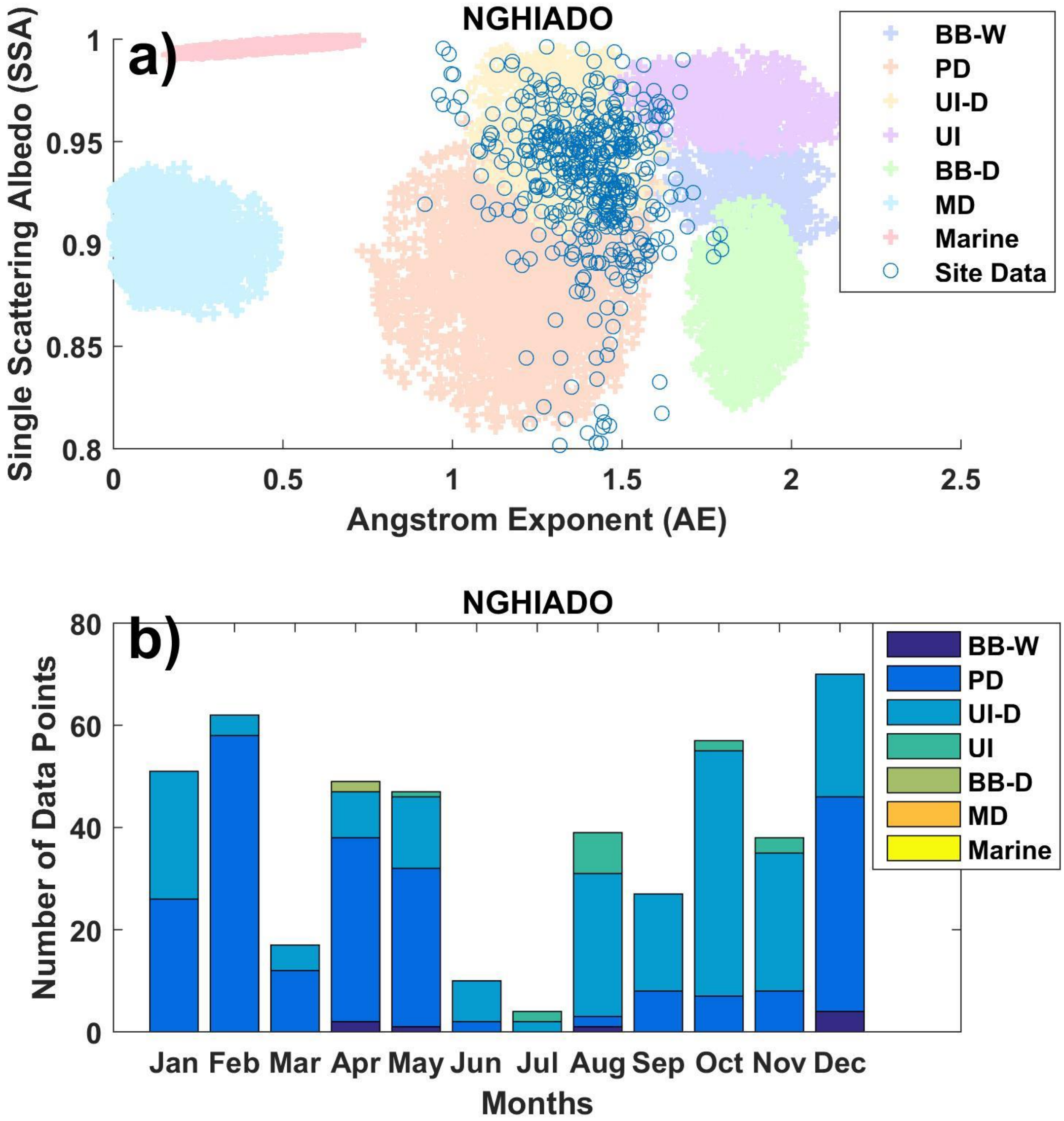


Figure 25. a) Scatter plot for the data points from Nghia Do. b) Monthly distribution of aerosol types in Nghia Do.

Figure 25 (a) shows that Nghia Do is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 25 (b)) shows that the whole year is dominated by polluted dust and urban industrial (developing economy) with polluted dust more dominant from December to May and the urban industrial (developing economy) more dominant from June to November.

<u>Nha Trang</u>

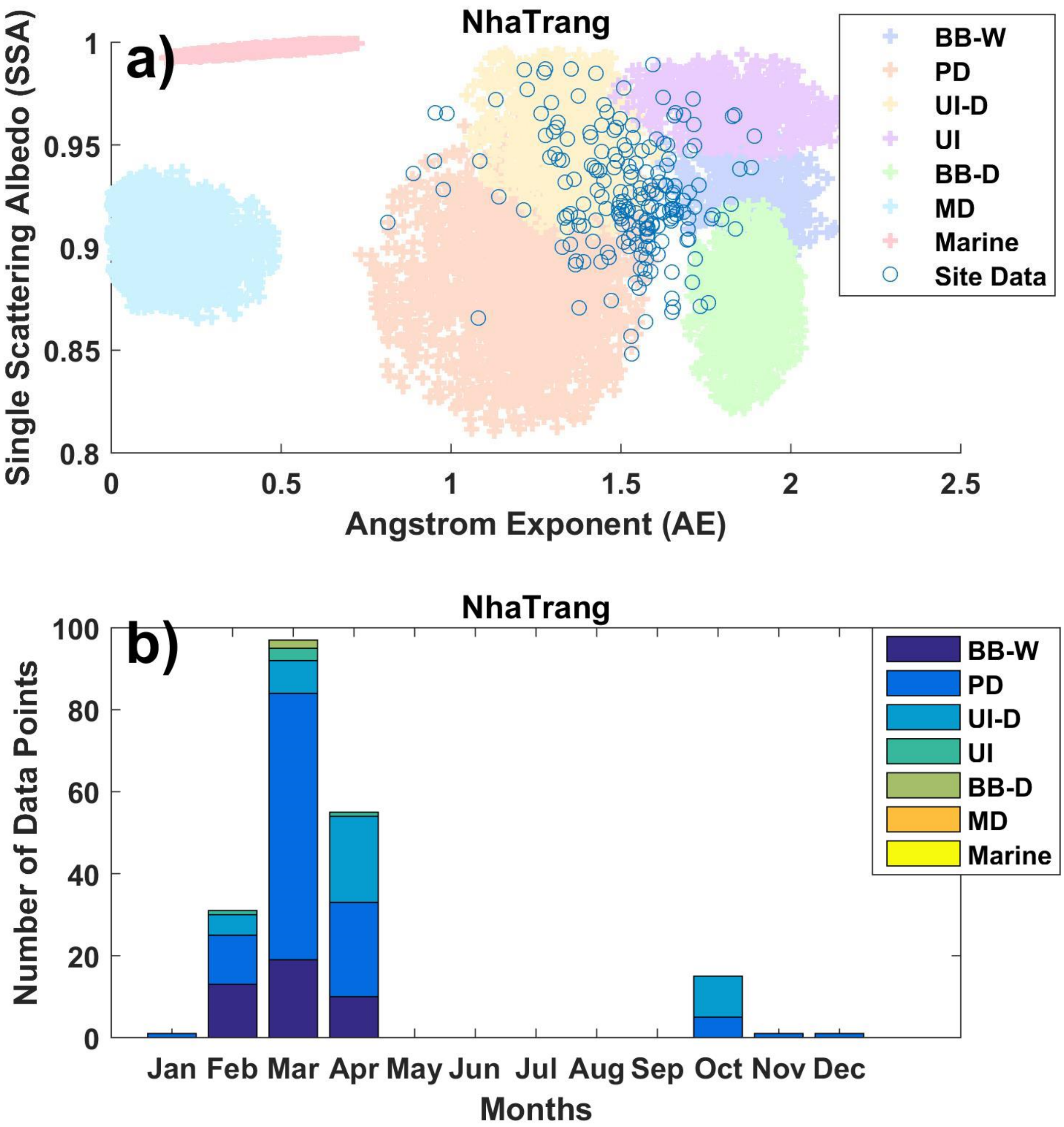


Figure 26. a) Scatter plot for the data points from Nha Trang. b) Monthly distribution of aerosol types in Nha Trang.

Figure 26 (a) shows that Nha Trang is comprised of a mixture of fine and coarse aerosols leaning more towards fine. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 26 (b)) shows that this site is dominated by polluted dust and urban industrial (developing economy) followed by biomass burning white smoke from February to April.

4.3 Aerosol types over Taiwan

Descriptions of the different AERONET site locations in Taiwan are presented in Table 9. It is worth noting and Lulin, located in a mountainous region, is representative of a clean atmosphere. [22] Also, Dongsha Island is referred to as a marine site. [14] Taiwan experiences a generally warm climate all year round. The northern part of Taiwan has a sub-tropical climate while the southern part has a tropical climate. This means that Taiwan experiences warm winters and hot and wet summers. The wet season is usually from July to September. [34]

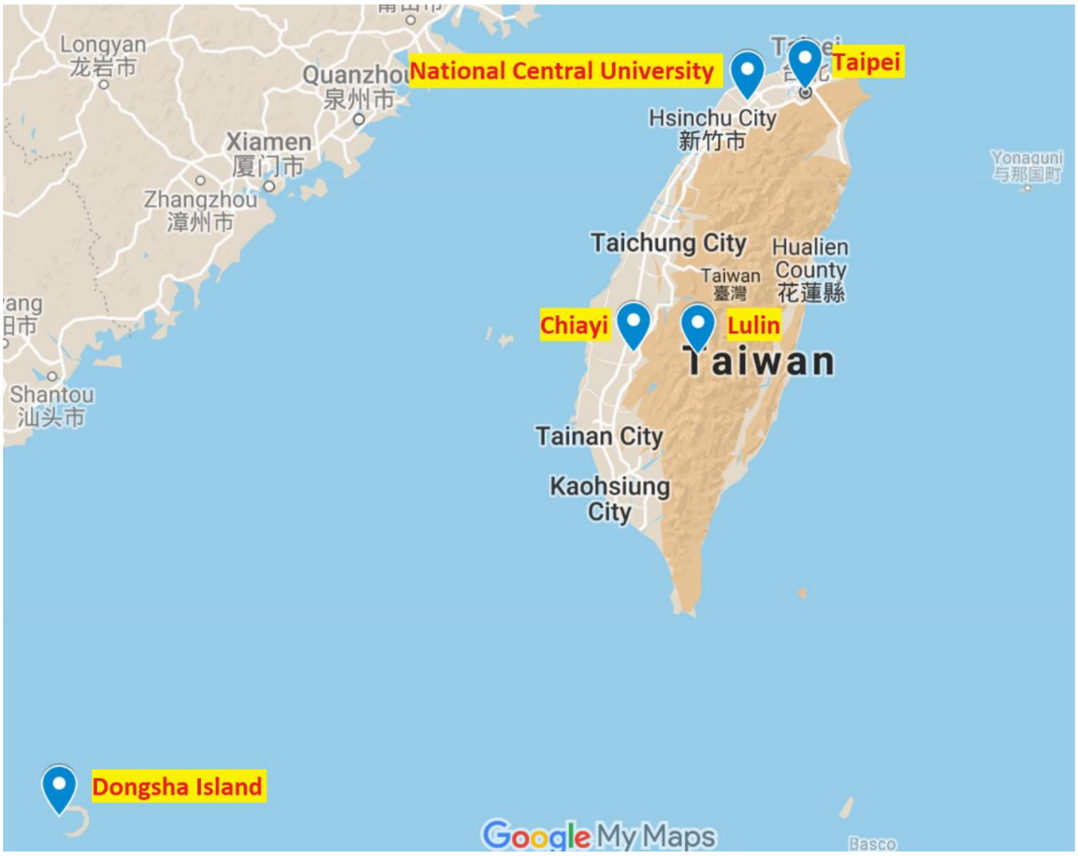


Figure 27. AERONET sites in Taiwan. (Map created using Google Maps)

Table 9. Description of the instrument location for each AERONET site in Taiwan.

Site	Country	Description (from http://aeronet.gsfc.nasa.gov)
Chiayi	Taiwan	The site is surrounded by mountains to the east and buildings to the west. Potential sources of aerosol here are agricultural and industrial activities.
DongshaIsland		Dongsha is a small island in the South China Sea. The only possible source of anthropogenic aerosol here is an airport. This site is considered as a marine background site.
EPANCU		The instrument is at the National Central University in the city of Chungli. Chungli city has industrial areas as well as green reserves.
Lulin		The instrument is at the Lulin Atmospheric Background Station (LABS) in Nantou County. Around 83% of Nantou County is covered by hills and mountains.
NCUTaiwan		The instrument is on the same location as the EPANCU site.
TaipeiCWB		The instrument is on the roof of the Central Weather Bureau (CWB) headquarters located at the capital city of Taipei.

Chiayi

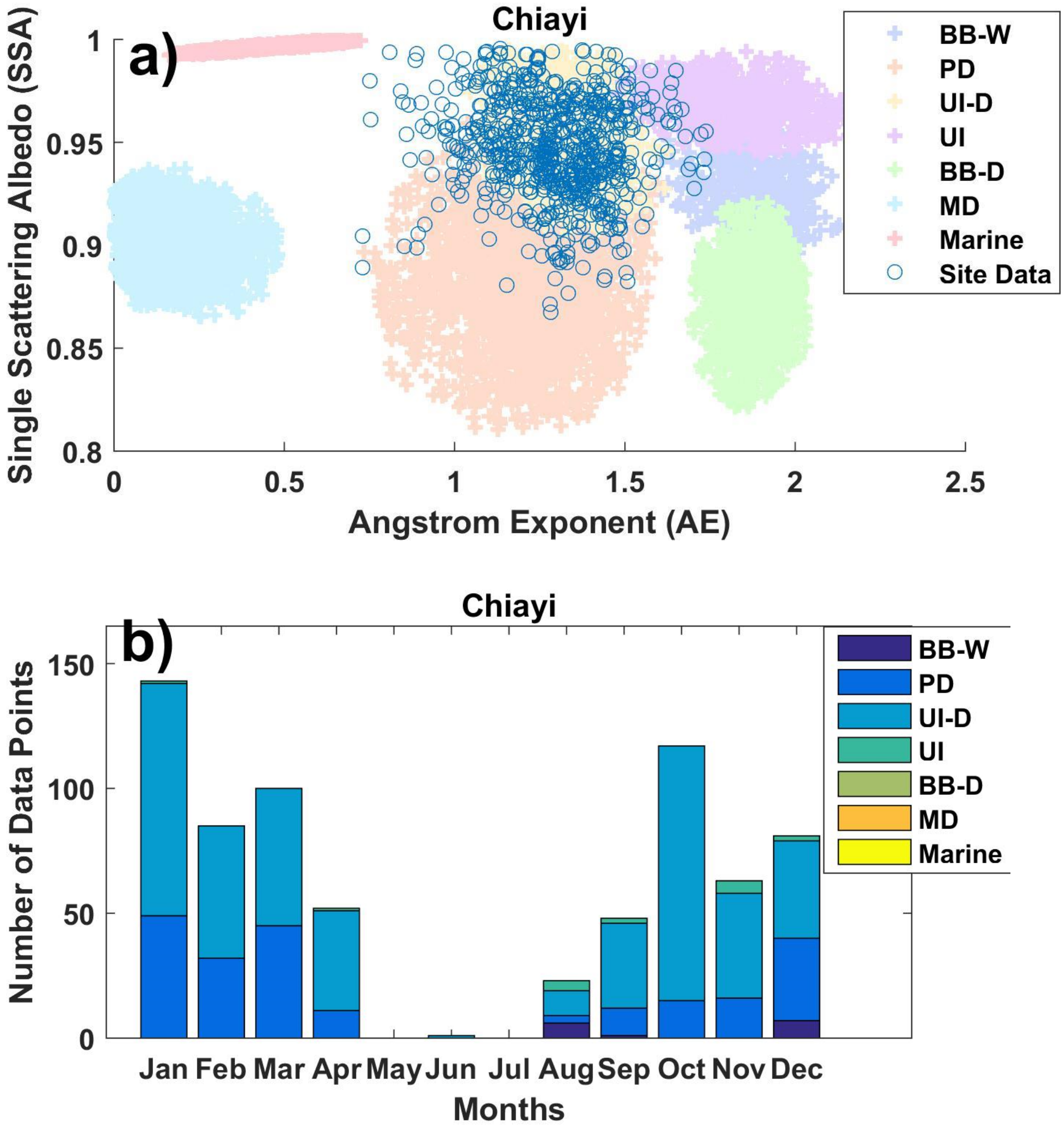


Figure 28. a) Scatter plot for the data points from Chiayi. b) Monthly distribution of aerosol types in Chiayi.

Figure 28 (a) shows that Chiayi is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 28 (b)) shows that the whole year (except for May to July which lacks data points) is dominated by urban industrial (developing economy) followed by polluted dust.

Dongsha Island

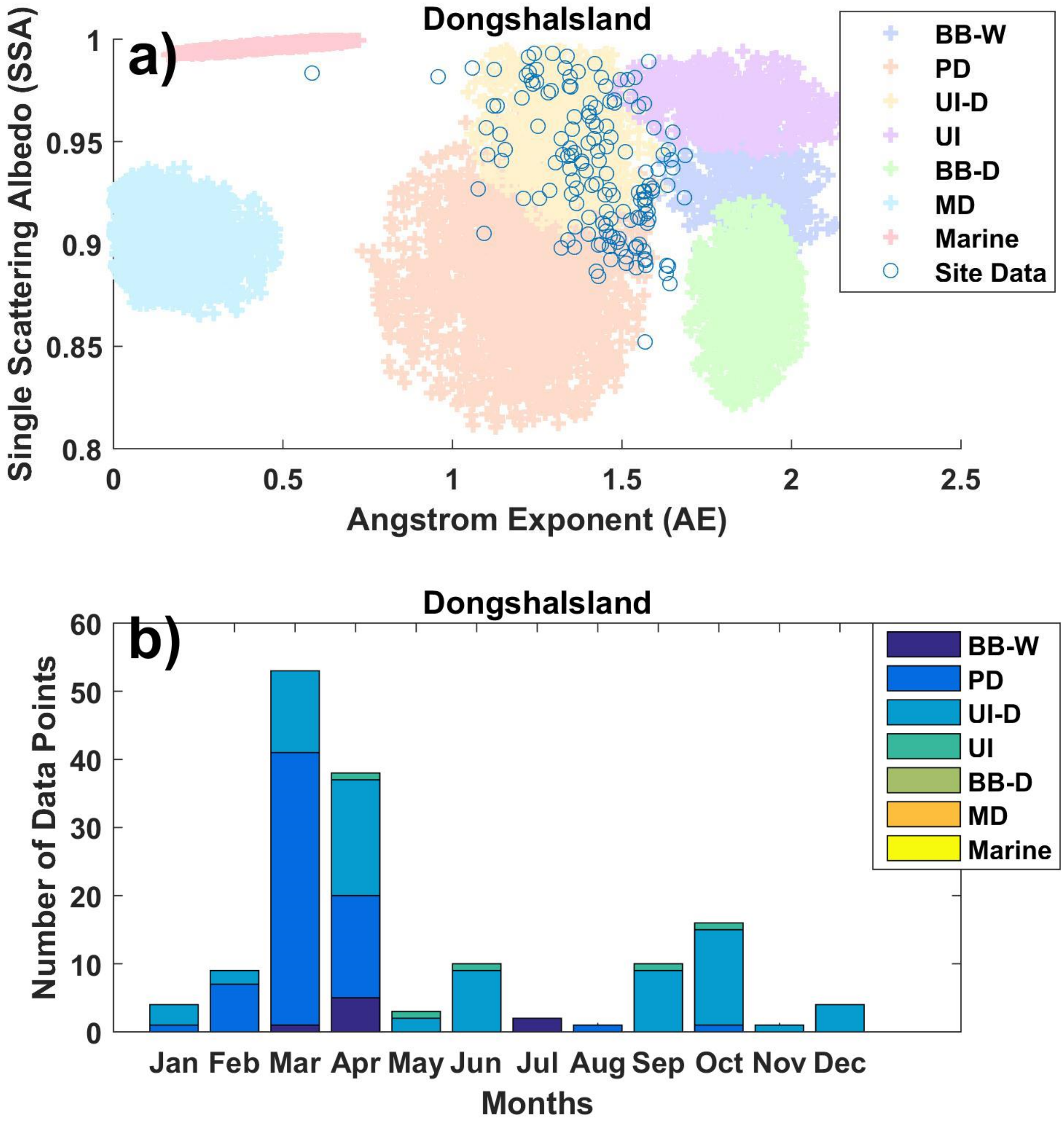


Figure 29. a) Scatter plot for the data points from Dongsha Island. b) Monthly distribution of aerosol types in Dongsha Island.

Figure 29 (a) shows that Dongsha Island is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 29 (b)) shows that urban industrial (developing economy) is present all year round while significant amounts of polluted dust is present from February to April – these may be due to the airport in the airport in the island. The results show that there is no marine aerosol detected even though Dongsha Island is expected to be a marine aerosol site.

<u>EPA NCU</u>

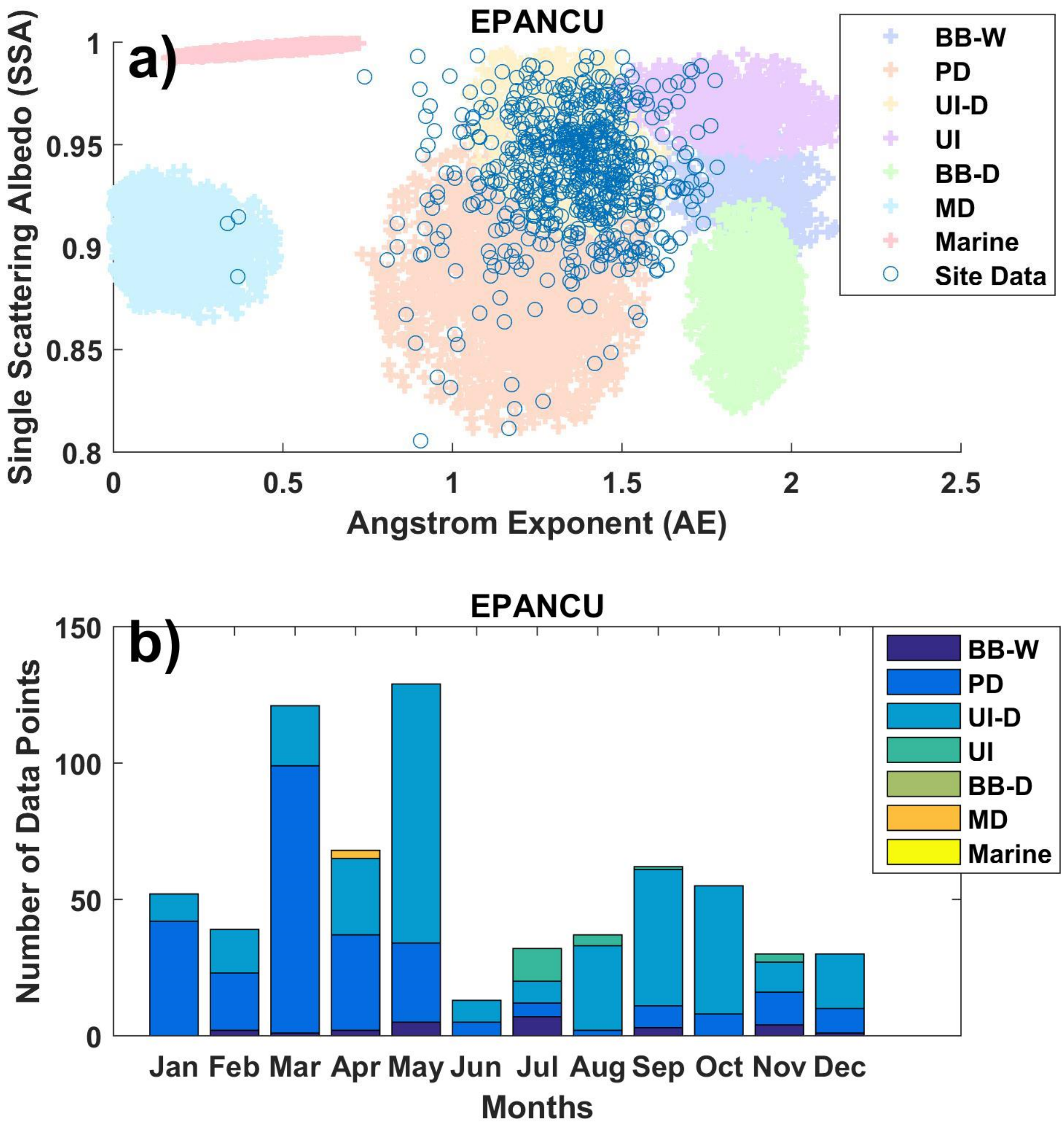


Figure 30. a) Scatter plot for the data points from EPA NCU. b) Monthly distribution of aerosol types in EPA NCU.

Figure 30 (a) shows that EPA NCU is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 30 (b)) shows that January to April is dominated by polluted dust followed by urban industrial (developing economy). On the other hand, the opposite is true for the rest of the year (May to December) where urban industrial (developing economy) is more dominant compared to polluted dust.

Lulin

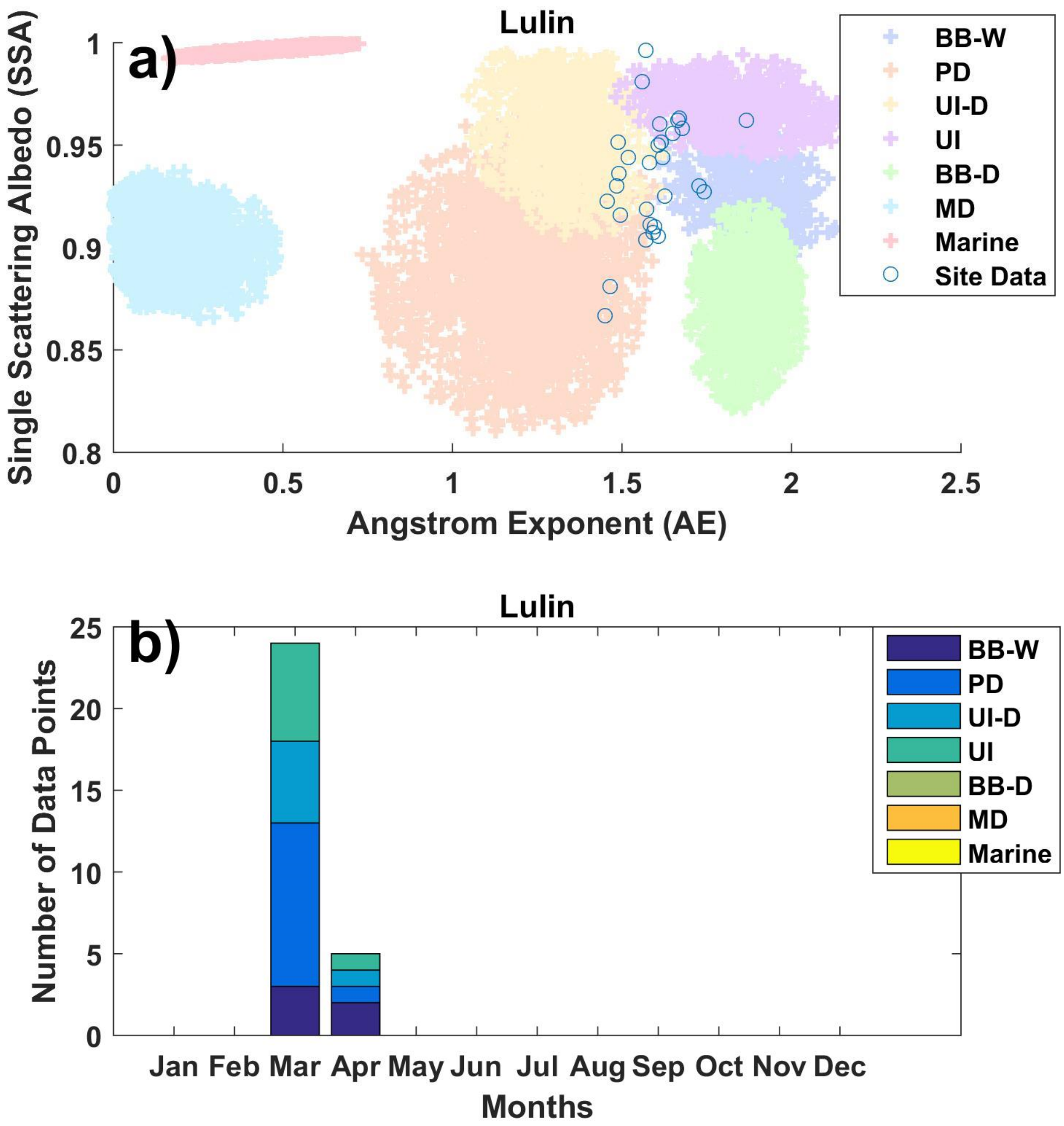


Figure 31. a) Scatter plot for the data points from Lulin. b) Monthly distribution of aerosol types in Lulin.

Figure 31 (a) shows that whatever little data was available in Lulin were mostly fine reflective aerosols. The monthly distribution (Figure 31 (b)) shows that only the months of March and April have available data which is mostly polluted dust and urban industrial with some traces of biomass burning white smoke.

NCU Taiwan

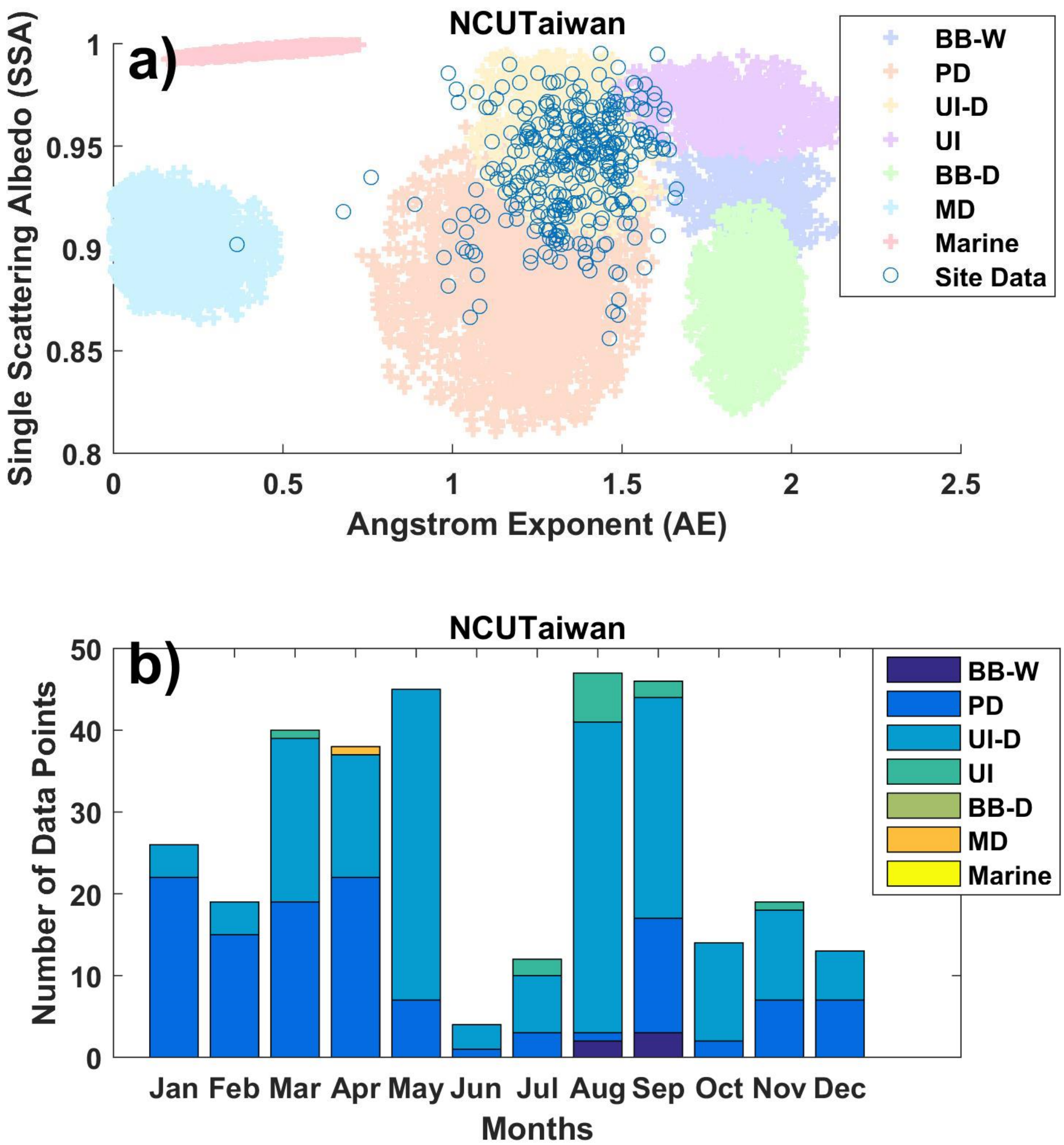


Figure 32. a) Scatter plot for the data points from NCU Taiwan. b) Monthly distribution of aerosol types in NCU Taiwan.

Figure 32 (a) shows that NCU Taiwan is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 32 (b)) shows that December to April is dominated by polluted dust followed by urban industrial (developing economy) while the opposite is the case for the months of May to November. This is the same as in EPA NCU which makes sense considering that both instruments are located close to each other.

<u>Taipei CWB</u>

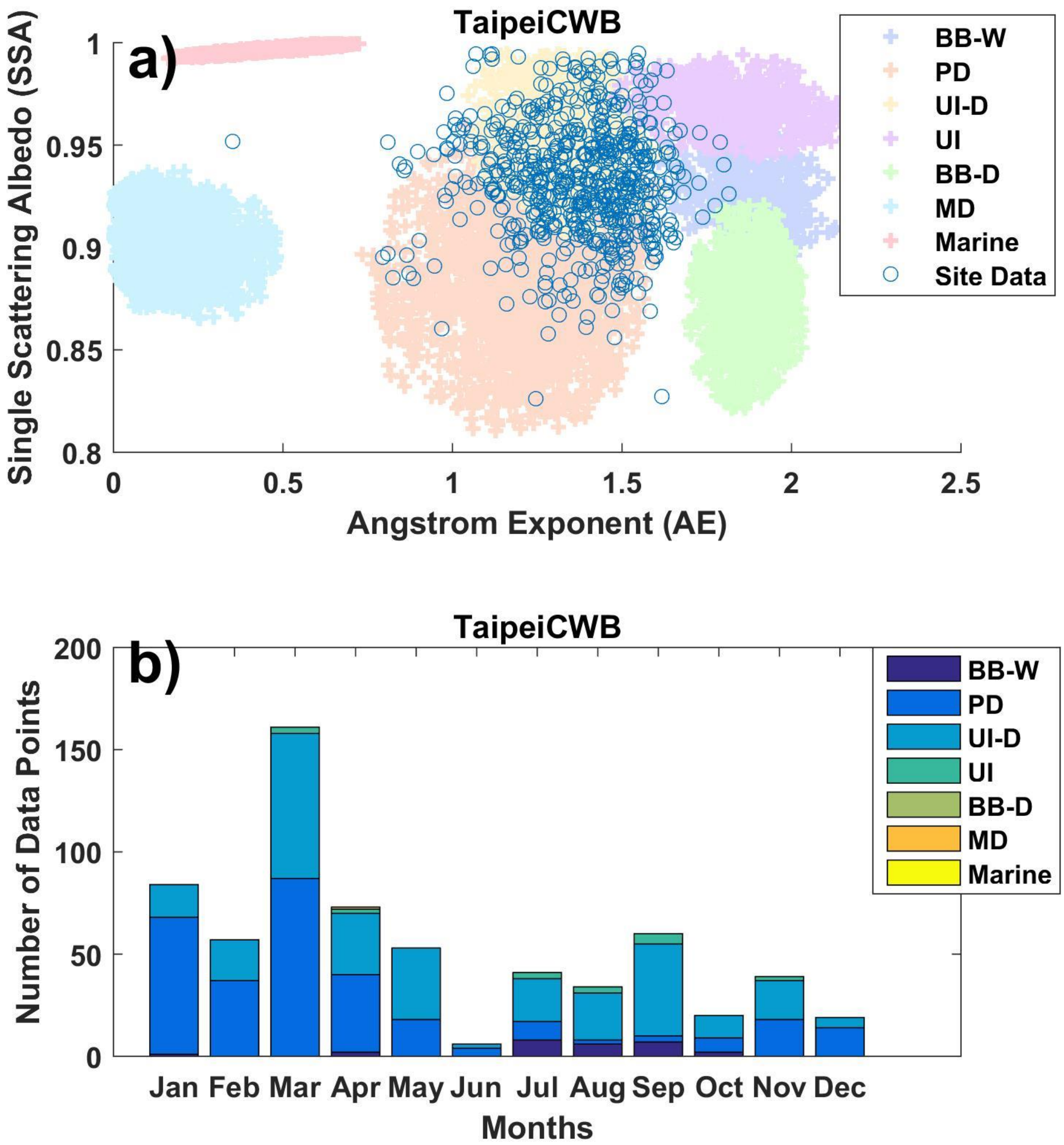


Figure 33. a) Scatter plot for the data points from Taipei CWB. b) Monthly distribution of aerosol types in Taipei CWB.

Figure 33 (a) shows that Taipei CWB is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 33 (b)) shows that the whole year is dominated by urban industrial and polluted dust which may be attributed to vehicular exhaust and industrial activities in Taipei.

4.4 Aerosol types over Singapore, Indonesia, and the Philippines

Descriptions of the different AERONET site locations in Singapore, Indonesia, and the Philippines are presented in Table 10. Almost all sites here except for Manila Observatory experience a tropical climate with high rainfall all year round. Manila Observatory has a tropical monsoon climate and experiences trade winds from the southwest and northeast monsoons. [22] Singapore, located near the equator and having a tropical climate, experiences rainfall and high humidity all year round with little variation between months. [35] On the other hand, Indonesia, which also has a tropical climate, experiences extreme variations in rainfall. Generally, the dry season in Indonesia is from June to September while the wet season is from December to March. [36] Finally, in the Philippines, Manila has what PAGASA (Philippine Atmospheric, Geophysical and Astronomical Services Administration) refers to as Type I climate while ND Marbel Univ has a Type IV climate. A Type I climate is characterized by two distinct seasons: a dry season from November to April and a wet season for the rest of the year. Meanwhile, a Type IV climate is characterized by even rainfall distribution throughout the year. [37]

Figure 34. AERONET sites in Singapore, Indonesia, and Philippines. (Map created using Google Maps)

Table 10. Description of the instrument location for each AERONET site in Singapore, Indonesia, and Philippines.

Site	Country	Description (from http://aeronet.gsfc.nasa.gov)
Bandung	Indonesia	The instrument is on top of the Institute of Technology building in Bandung. Bandung is the fourth most populous city in Indonesia.
Jambi		No information is given about this site.
Palangkaraya		The instrument is on the roof of the Tjilik Riwut Meteorology Station at the Tjiik Riwut Airport.
Pontianak		Pontianak is a major port city west of Borneo and Palangkaraya.
ManilaObservatory	Philippines	The instrument is on the roof deck of the Manila Observatory located inside the Ateneo de Manila campus in Quezon City. It is 15km northeast from the capital city Manila and 20km from Manila Bay.
NDMarbelUniv		The instrument is at the rooftop of OMER hall of Notre Dame of Marbel University (NDMU). The university is located almost at the center of Korondal city.
Singapore	Singapore	The instrument is mounted on the roof deck of Block S17 in the National University of Singapore (NUS).

<u>Bandung</u>

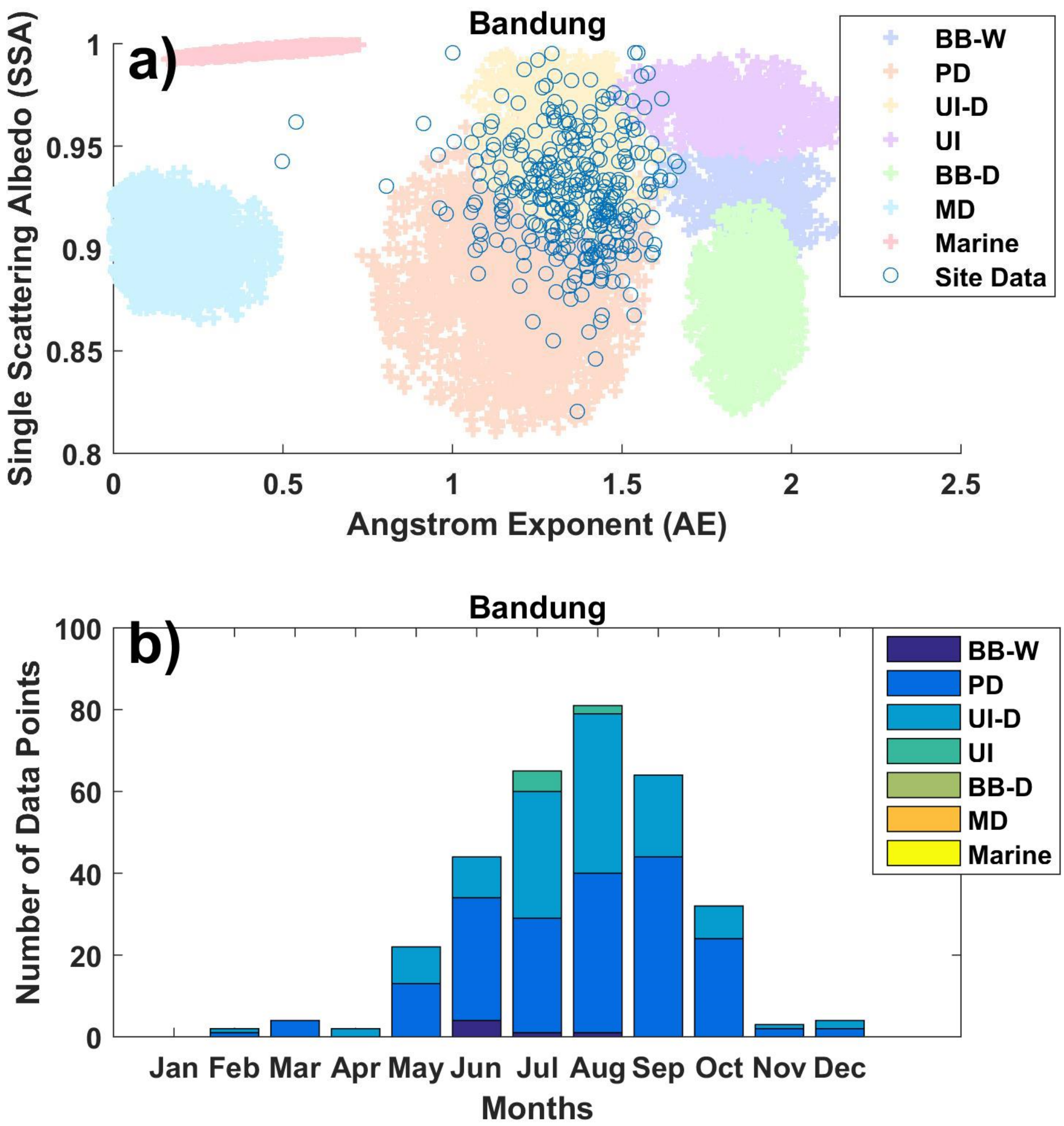


Figure 35. a) Scatter plot for the data points from Bandung. b) Monthly distribution of aerosol types in Bandung.

Figure 35 (a) shows that Bandung is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 35 (b)) shows that the whole year is dominated by urban industrial (developing economy) and polluted dust which may be attributed to vehicular exhaust and industrial activities in Bandung. The scarcity of data from November to April roughly corresponds to the wet season in Indonesia.

<u>Jambi</u>

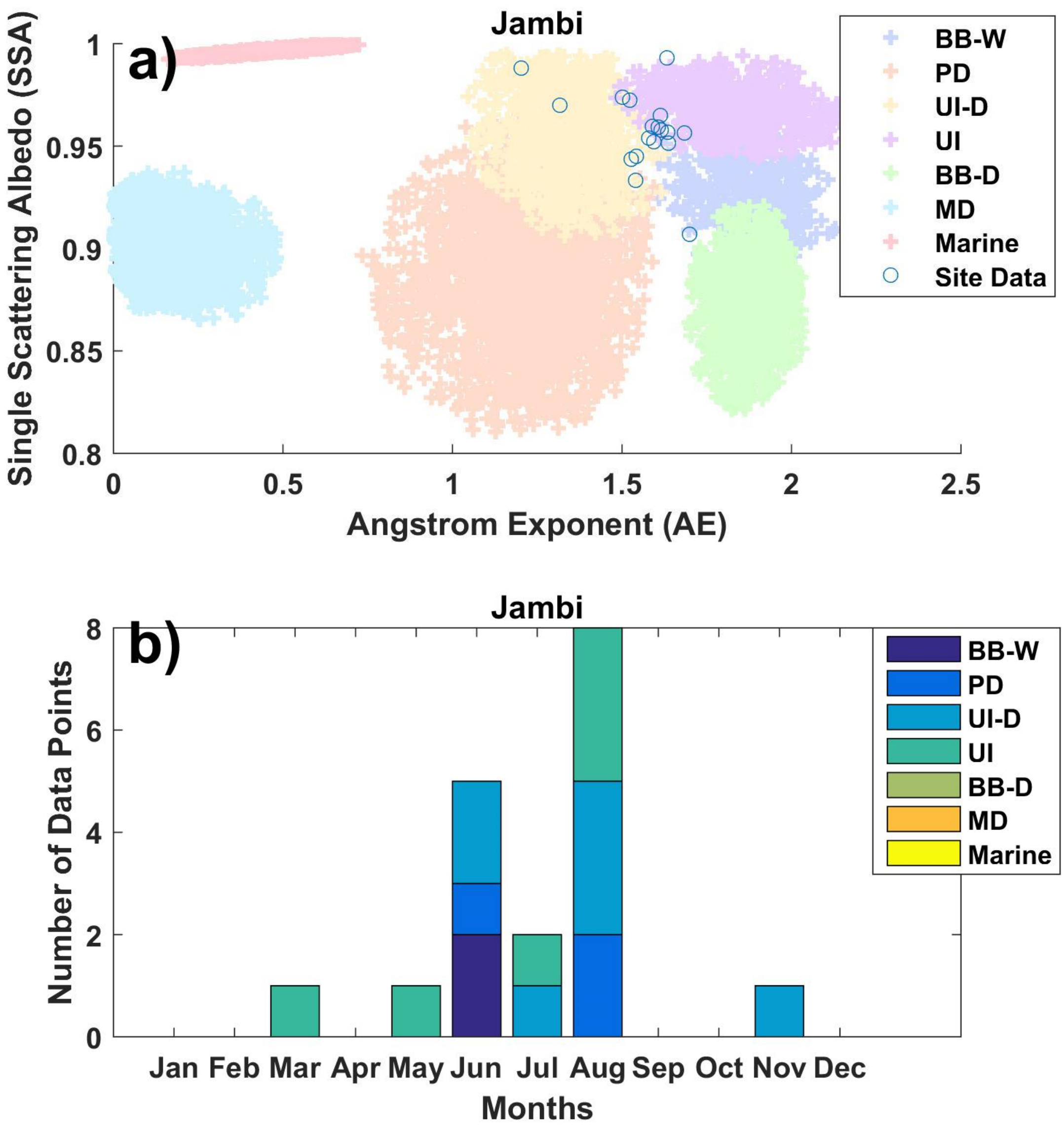


Figure 36. a) Scatter plot for the data points from Jambi. b) Monthly distribution of aerosol types in Jambi.

Figure 36 (a) shows that Jambi only has a few data points and these data points are mostly fine reflective aerosols. The monthly distribution (Figure 36 (b)) shows that this site is comprised of urban industrial and polluted dust with some biomass burning white smoke in June.

Palangkaraya

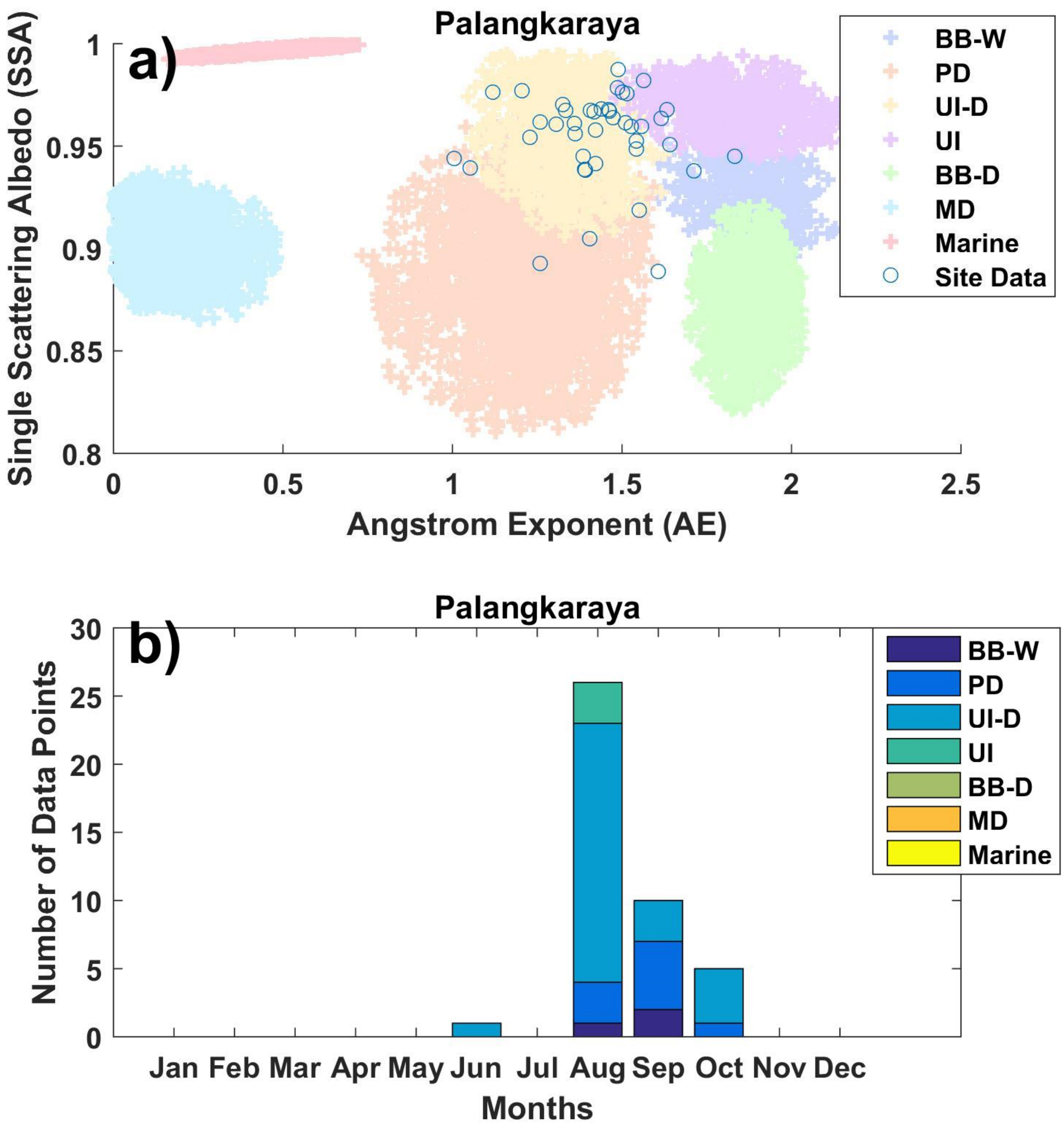


Figure 37. a) Scatter plot for the data points from Palangkaraya. b) Monthly distribution of aerosol types in Palangkaraya.

Figure 37 (a) shows that Palankaraya, like Jambi, only has a few data points and most of these are fine reflective aerosols. The monthly distribution (Figure 37 (b)) shows that the available data points are mostly urban industrial (developing economy) in the months of June and August to October. These may be attributed to the Tjiik Riwut Airport near the site.

Pontianak

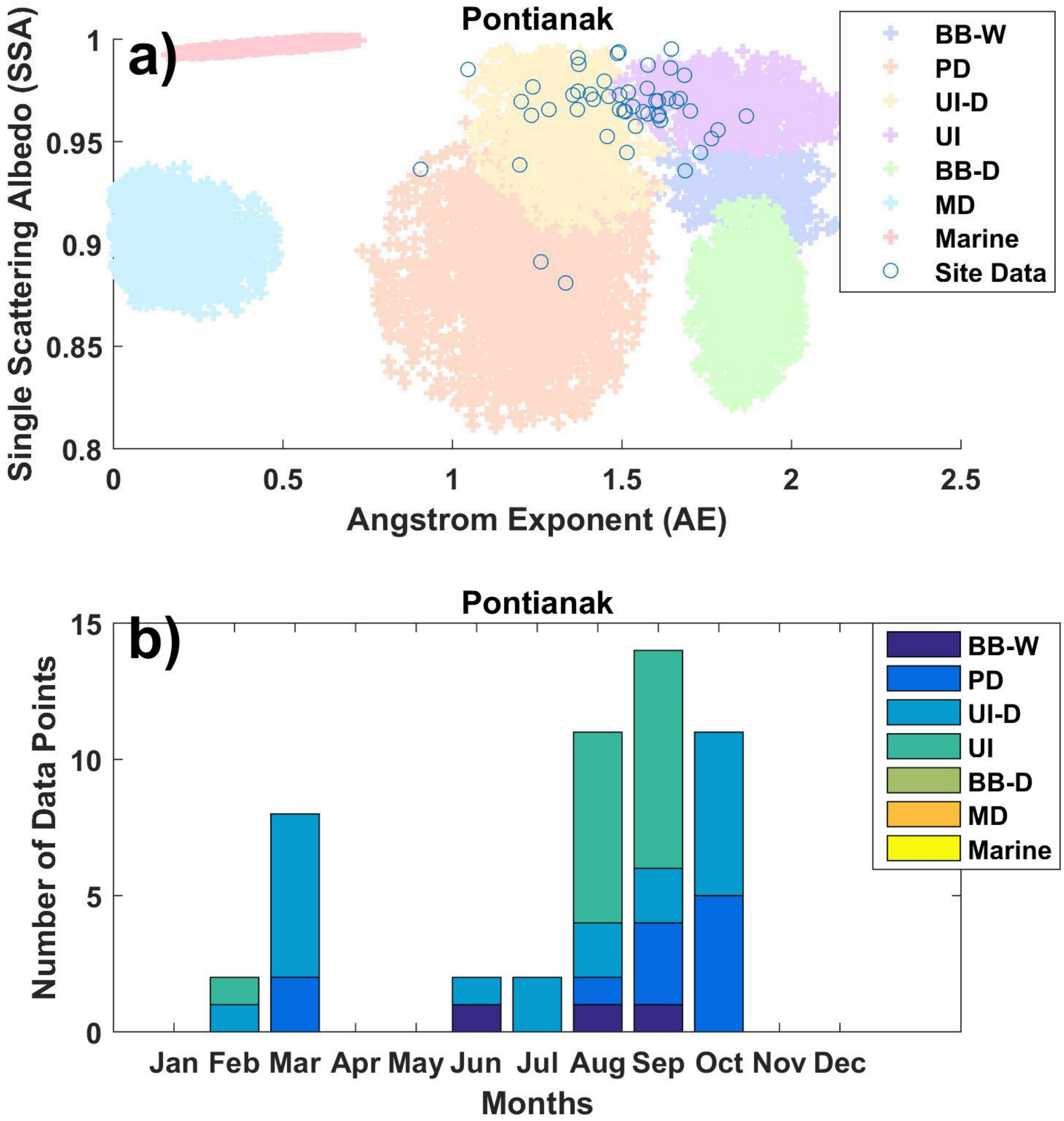


Figure 38. a) Scatter plot for the data points from Pontianak. b) Monthly distribution of aerosol types in Pontianak.

Figure 38 (a) shows that Pontianak is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are mostly scattering or reflective aerosols. The monthly distribution (Figure 38 (b)) shows that urban industrial (both kinds) are dominant in this site. These aerosols may be attributed to industrial activities near the site.

Manila Observatory

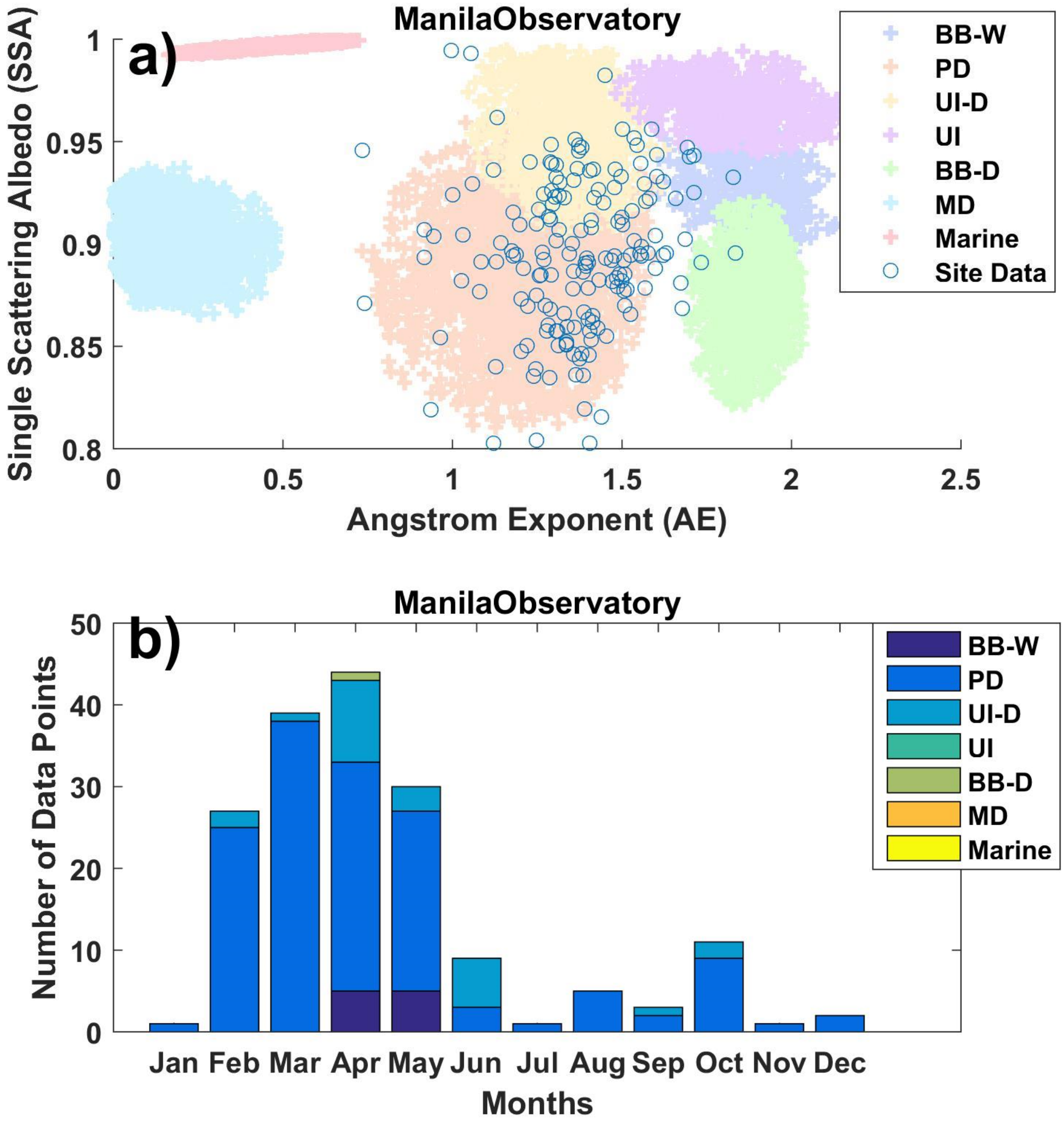


Figure 39. a) Scatter plot for the data points from Manila Observatory. b) Monthly distribution of aerosol types in Manila Observatory.

Figure 39 (a) shows that Manila Observatory is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are a mixture of scattering and absorbing aerosols. The monthly distribution (Figure 39 (b)) shows that the whole year is dominated by polluted dust and there are some traces of urban industrial (developing) in most of the months. The scarcity of data for the latter half of the year is due to the rainy season.

Notre Dame of Marbel University

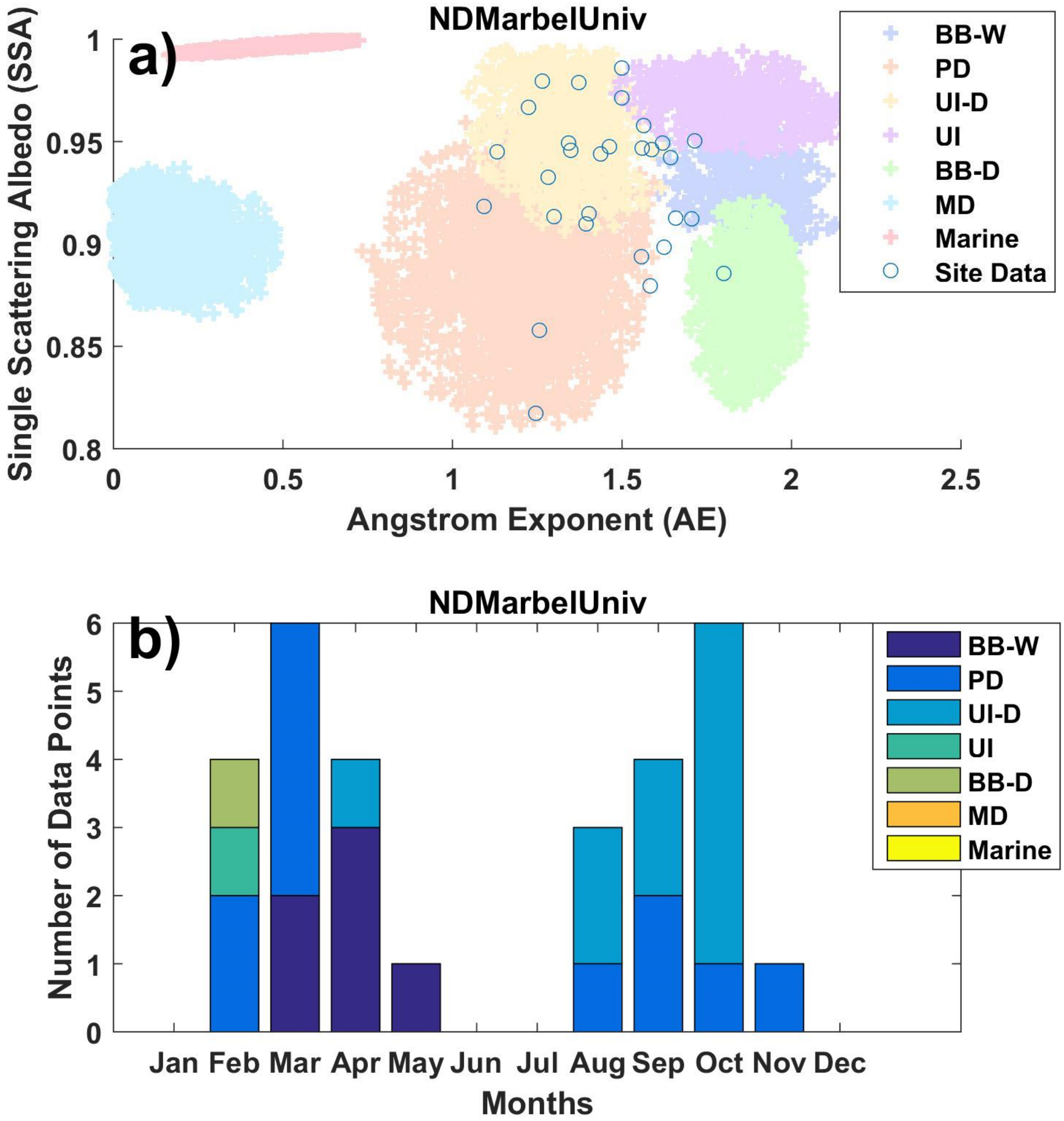


Figure 40. a) Scatter plot for the data points from Notre Dame of Marbel University. b) Monthly distribution of aerosol types in Notre Dame of Marbel University.

Figure 40 (a) shows that the Notre Dame of Marbel university site is comprised of a mixture of fine and coarse aerosols. The aerosols in this site are a mixture of scattering and absorbing aerosols. The monthly distribution (Figure 40 (b)) shows that March to May has mostly biomass burning white smoke and the rest of the year is dominated by polluted dust and urban industrial (developing economy).

Singapore

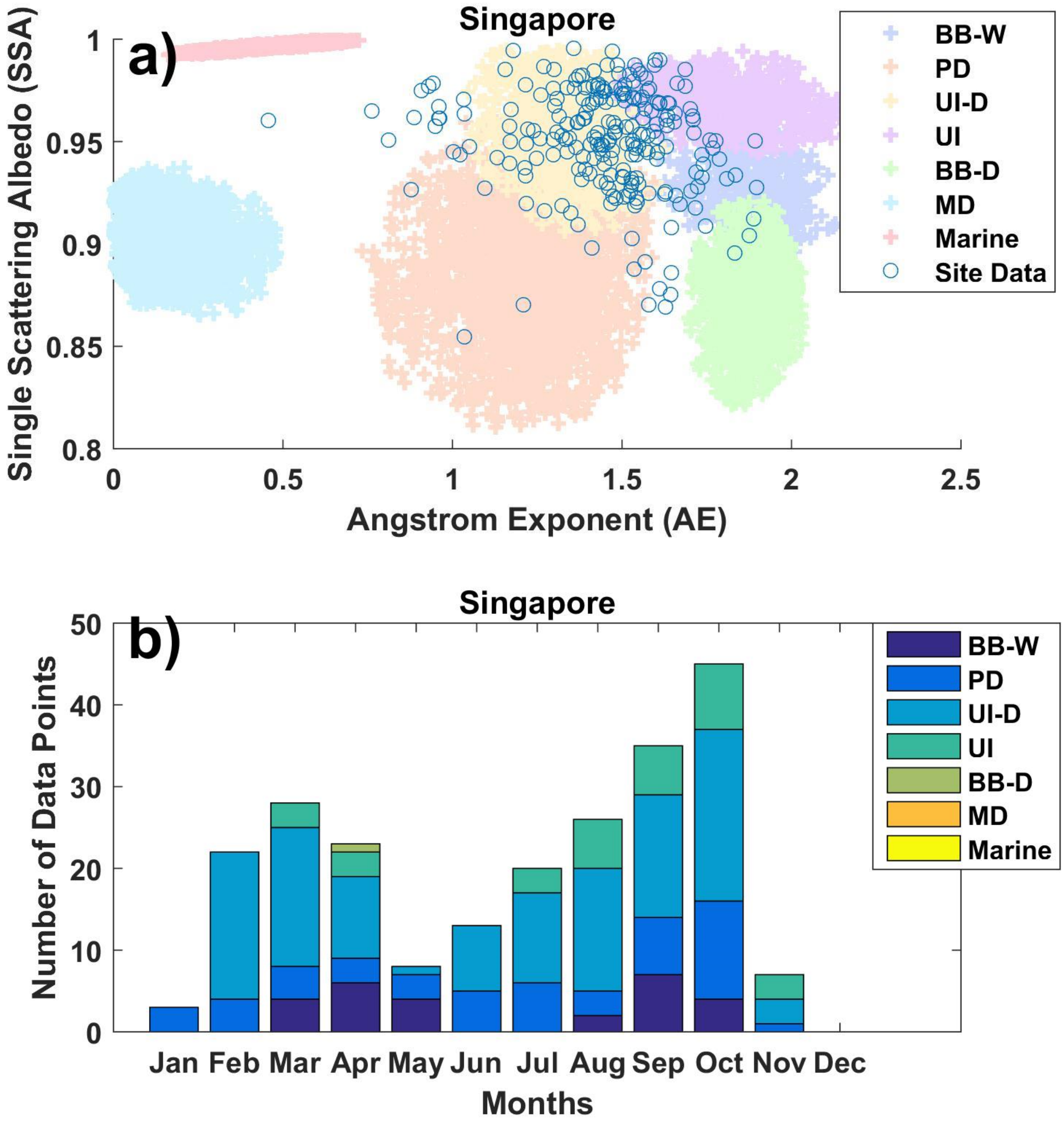


Figure 41. a) Scatter plot for the data points from Singapore. b) Monthly distribution of aerosol types in Singapore.

Figure 41 (a) shows that Singpore is comprised of a mixture of fine and coarse aerosols which are mostly reflective. The monthly distribution (Figure 41 (b)) shows that the whole year is dominated by urban industrial (both types) and polluted dust. There are also some traces of biomass burning white smoke from March to May and from August to October. These results may be because Singapore is a major urban port city.

4.5 Summary

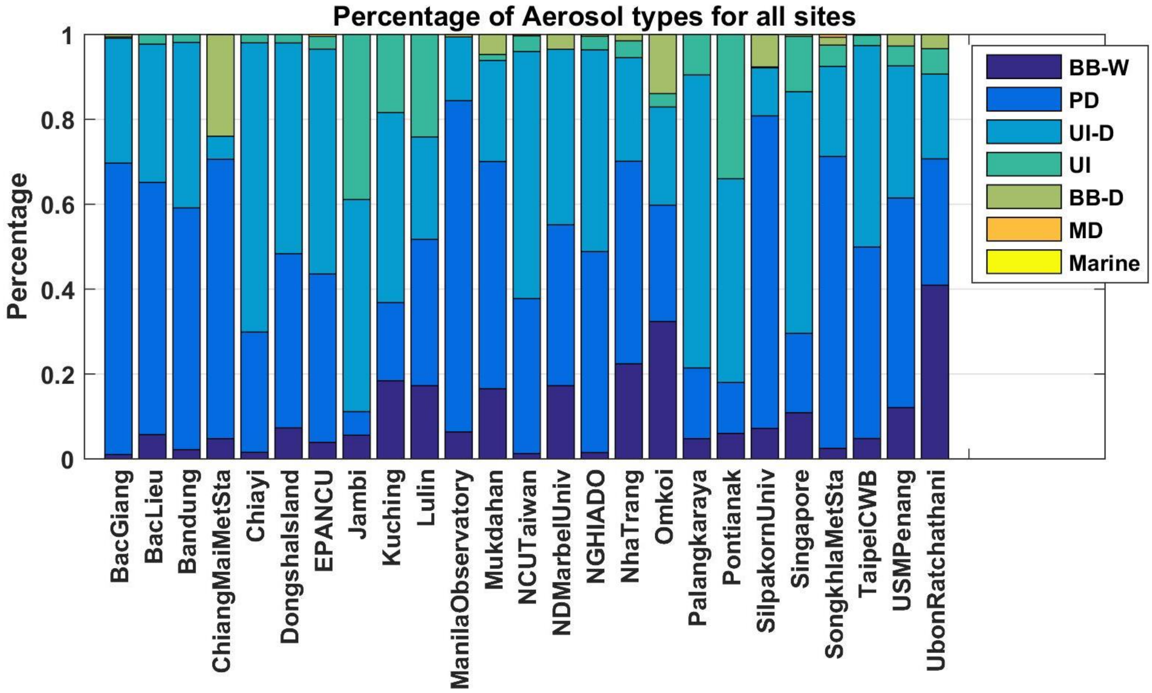


Figure 42. Summary of the results. Percentage of the aerosol types for all sites.

Figure 42 shows that the most dominant aerosol type in SEA is polluted dust (PD) followed by urban-industrial developing (UI-D). PD and UI-D aerosol types are usually due to lax environmental regulations where fossil fuel is not fully combusted. [38] Biomass burning white smoke (BB-W) can be attributed to harvesting activities and is dominant in Ubon Ratchathani (in the months of January to May), Omkoi (February to May), and Kuching (July to September). Biomass burning dark smoke (BB-D) comes from the burning of objects which contains huge amounts of carbon; BB-D is dominant in Chiang Mai Meteorological Station (January to April), Omkoi (February to April), and Nghia Do (January to April). Overall, the large amounts of polluted dust and urban industrial (developing economy) is consistent with other studies in SEA which link the densely populated regions of SEA with high emissions of soil dust (due to constructions and development) and incomplete combustions (from diesel and coal) which produce coarse absorbing aerosols. [38], [22], [39], [40]

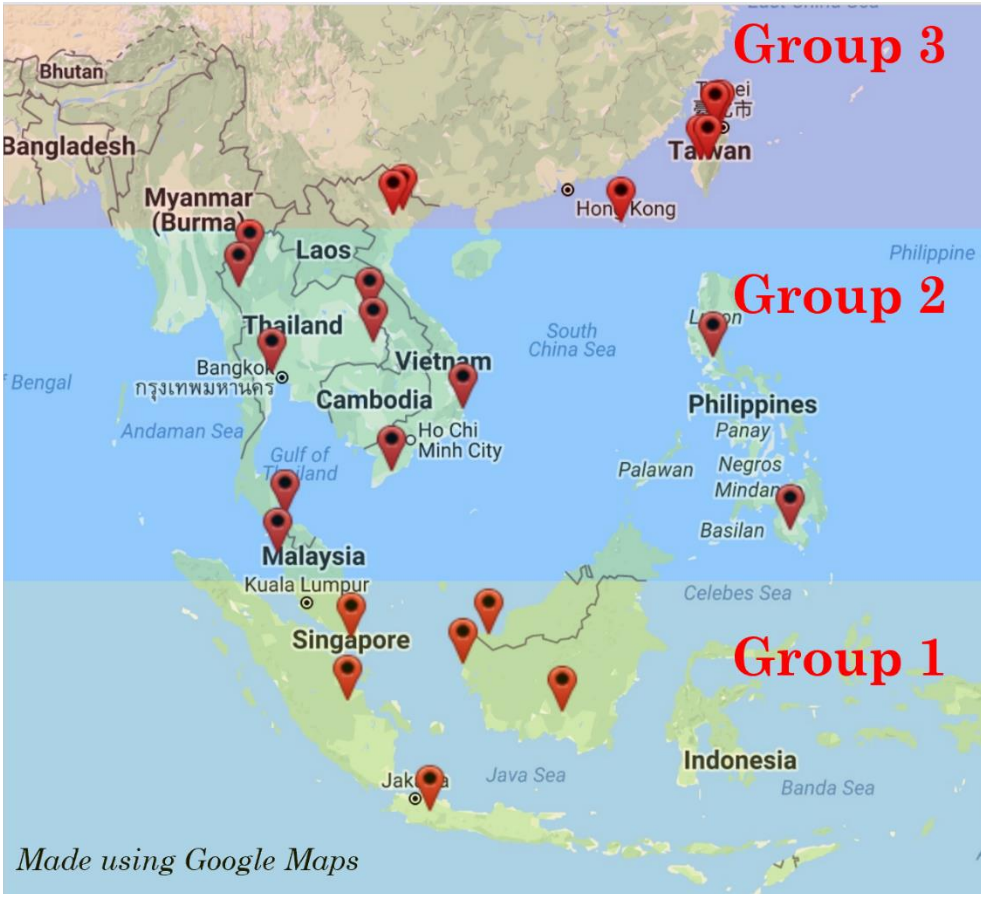


Figure 43. Map showing the grouping of the AERONET sites according to latitude. Those highlighted in yellow, blue, and red correspond to groups 1, 2, and 3 respectively.

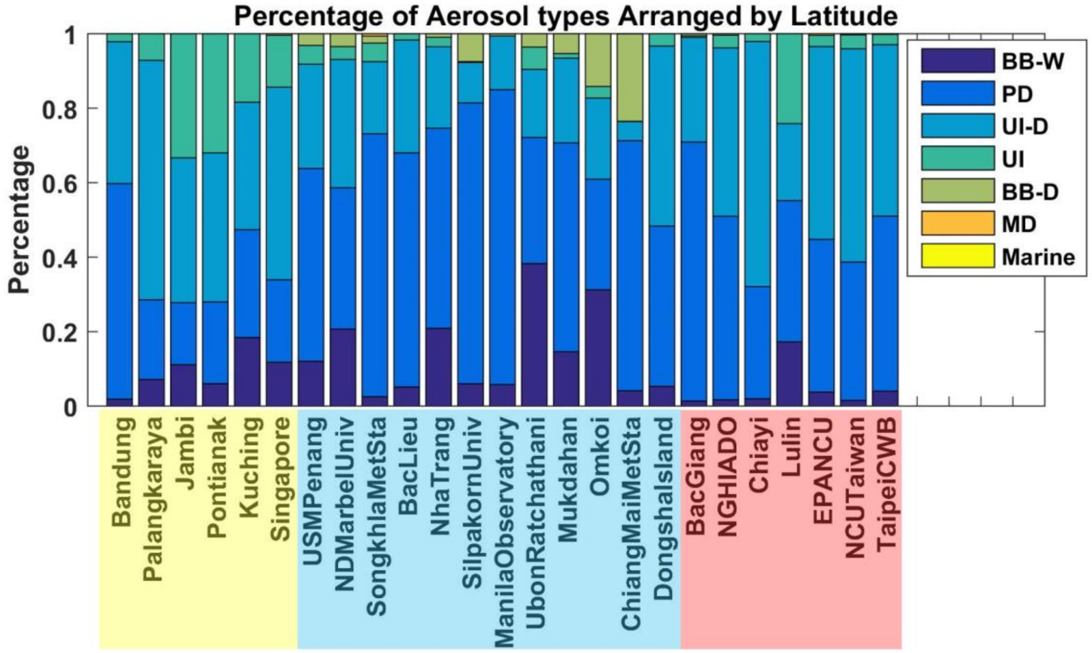


Figure 44. These are the same results as the summary (Figure 42) but arranged by latitude. The site names highlighted in yellow, blue, and red correspond to groups 1, 2, and 3 respectively (like Figure 43).

Arranging the sites by latitude also gives us insight on the geographical trends of aerosol types. In Figure 44, the sites are arranged by latitude and are grouped into three groups. The rationale behind this grouping is that sites belonging to the same group exhibit similar trends in aerosol type dominance. Group 1 consists of sites in the lower latitudes, included here are all sites from Indonesia, the southernmost site from Malaysia (Kuching), and the site from Singapore. Next, group 2 includes all the sites from Thailand, Malaysia (except Kuching), Philippines, and southern Vietnam (Bac Lieu). Lastly, group 3 includes all the sites from Taiwan and northern Vietnam.

Figure 44 shows that the sites belonging to group 1 are predominantly UI-D. Also, the UI levels are generally higher in group 1 than in other groups. Another interesting thing in this group is the presence of BB-W and the absence of BB-D which indicates that the smoke in this region is mostly more scattering. In short, the aerosols detected in the group 1 sites are mostly scattering aerosols. On the other hand, group 2 sites are mostly dominated by PD. Unlike the other groups, the percentage of UI and UI-D in group 2 is small. The unique thing about group 2 is the presence of both BB-W and BB-D. All these characteristics of group 2 show that most aerosols in this region are more absorbing. Finally, group 3 aerosols are similar to group 1 except that the percentages of UI and BB-W are smaller – this implies that the aerosols found in group 3 are coarser compared to group 1.

Chapter 5

Conclusions

Aerosol classification using AERONET-derived aerosol optical properties was demonstrated in this paper. This paper is successful in terms of accomplishing its objectives which has been to classify aerosol sites over 25 SEA AERONET sites and to determine the spatial and temporal trends of these aerosol types. Some of the aerosol optical properties used in this paper are the Angstrom exponent (*AE*), single scattering albedo (*SSA*), and real refractive index (*n*) which determine aerosol particle size, reflectivity, and water content respectively. Three-dimensional specified clustering using AE, SSA, and n was utilized in order to achieve classification. Furthermore, Mahalanobis distance was used as the distance metric because it provides advantages such as being scale invariant and being able to take the covariance (which corresponds to the obliqueness of the cluster) into account.

The result of the classification showed that the aerosol emissions from Indonesia, Singapore, and a part of Malaysia (Kuching) are dominated mostly by more reflective aerosols like urban industrial (UI and UI-D) and biomass burning white smoke (BB-W). Meanwhile, aerosol emissions from Thailand, Philippines, Malaysia, and southern Vietnam (Bac Lieu) are dominated by more absorbing aerosols such as polluted dust (PD) and biomass burning dark smoke (BB-D). Furthermore, aerosol emissions from northern Vietnam and Taiwan are dominated by coarse aerosols like polluted dust (PD) and urban industrial developing economy (UI-D).

Results also showed that biomass burning white smoke (BB-W) is most dominant in Ubon Ratchathani, Omkoi, and Kuching; biomass burning dark smoke (BB-D) is most dominant in Chiang Mai, Omkoi, and Nghia Do. But generally, the aerosol emissions in SEA is dominated by polluted dust (PD) and urban industrial developing economy (UI-D) – this is consistent with the findings of previous works. [38], [22], [39], [40]

What distinguishes this paper from other similar works is the use of Mahalanobis distance, cross-validation, and the number of dimensions used for clustering. The work of Chan [22] only used 2 dimensions (*AE* and *SSA*), and Euclidean distance. Arguably, using 3 dimensions is better than 2 because it takes more properties into account; also, as discussed in Section 3.2 Mahalanobis Distance, Mahalanobis distance is more advantageous than Euclidean distance for this kind of classification. Meanwhile, the work of Hamill [23] made use of Mahalanobis distance and 5 dimensions (*AE*, *SSA*, *AAE*, *n*, and *k*). Generally, having more dimensions is better, but the flaw in Hamill's work is that some of these parameters are correlated with each other (as discussed in Section 3.1 Establishing the reference clusters) – this produces bias towards the correlated parameters. Furthermore, none of the other works (Chan and Hamill) used cross-validation to check the validity of their reference clusters.

Future works on this topic should explore AERONET's version 3 data sets. The difference between versions 2 (used in this study) and 3 is the automated quality assurance and the lunar retrievals. The automated quality assurance might result in more consistent data retrievals, and the lunar retrievals would increase the number of data points. Additionally, one of the problems with Mahalanobis distance is that none of the data points were classified under the Marine aerosol class. This problem arises because the Mahalanobis distance gets the inverse of the covariance matrix. Since the marine aerosol cluster came from a model, the inverse covariance between its dimensions become large, so none of the points were classified into the marine cluster. To address this problem, future works should investigate other classification methods such as k-nearest neighbors.

References

[1] Solomon, S., Qin, D., Manning, M, Chen, Z., Marquis, M., Averyt, K.B., Tingor, M., Miller, H.L., "IPCC, 2007b. The physical science basis," *Climate Change 2007,* 2007.

[2] Parry, M.L., Canziani, O.F., Palutikof, J.P., van der Linden, P.J., Hanson, C.E., "Impacts, adaptation, and vulnerability In: Climate Change 2007: Impacts, Adaptation and Vulnerability Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, United Kingdom, 2007.

[3] Reid Jeffrey S., Edward J. Hyer, Randall S. Johnson, Brent N. Holben, Robert J. Yokelson, Jianglong Zhang, James R. Campbell, Sundar A. Christopher, Larry Di Girolamo, Louis Giglio, Robert E. Holz, Courtney Kearney, Jukka Miettinen, Elizabeth A. Reid, F., "Observing and understanding the Southeast Asian aerosol system by remote sensing: An initial review and analysis for the Seven Southeast Asian Studies (7SEAS) program," *Atmospheric Research,* vol. 122, pp. 403-468, 2013.

[4] Holben, B.N., T.F. Eck, I. Slutsker, D. Tanre, J.P. Buis, A. Setzer, E. Vermote, J.A Reagan, Y.J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, A. Smirnov, "AERONET - A federated instrument network

and data archive for aerosol characterization," *Remote Sensing of Environment,* vol. 66, pp. 1-16, 1998.

[5] Holben, B. N., T. F. Eck, I. Slutsker, A. Smirnov, A. Sinyuk, J. Schafer, D. Giles, and O. Dubovik, "AERONET's Version 2.0 quality assurance criteria," *Remote Sensing of the Atmosphere and Clouds,* 2006.

[6] O. Boucher, Atmospheric Aerosols, Springer Netherlands, 2015.

[7] Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker, "Cloud-screening and quality control algorithms for the AERONET database," *Remote Sensing of Environment,* no. 73, pp. 337-349, 2000.

[8] *Multiband Photometer CE318-N User's Manual,* 5th ed., Paris: Cimel Electronique, 2015, pp. 38-40.

[9] A. Ångström, "The parameters of atmospheric turbidity," *Tellus A,* no. 16, pp. 64-75, 1964.

[10] Eck, T.F., B.N. Holben, J.S. Reid, O. Dubovik, A. Smirnov, N.T. O'Neill, I. Slutsker, and S. Kinne, "Wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols," *Journal of Geophysical Research,* no. 104, pp. 31333-31350, 1999.

[11] Dubovik, O. and M. D. King, "A flexible inversion algorithm for retrieval of aerosol," *Journal of Geophysical Research,* vol. 20, no. 105, pp. 673-696, 2000.

[12] Dubovik, O., A. Smirnov, B.N. Holben, M.D. King, Y. J. Kaufman, T.F. Eck and I. Slutsker, "Accuracy assessment of aerosol optical properties retrieval from AERONET sun and sky radiance measurements," *Journal of Geophysical Research,* no. 105, pp. 9791-9806, 2000.

[13] Dubovik, O., B. N. Holben, T. Lapyonok, A. Sinyuk, M. I. Mishchenko, P. Yang and I. Slutsker, "Non-spherical aerosol retrieval method employing light scattering by spheroids," *Geophysical Research Letters,* no. 10, 2002a.

[14] Dubovik, O., B. N. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanré, and I. Slutsker, "Variability of absorption and optical properties of key aerosol types observed in worldwide locations," *Journal of Atmospheric Science,* no. 59, pp. 590-608, 2002b.

[15] Dubovik, O., A. Sinyuk, T. Lapyonok, B. N. Holben, M. Mishchenko, P. Yang, T. F. Eck, H. Volten, O. Munoz, B. Veihelmann, van der Zander, M Sorokin, and I. Slutsker, "Application of light scattering by spheroids for accounting for particle nonsphericity in remote sensing of desert dust," *Journal of Geophysical Research,* no. 111, 2006.

[16] Sinyuk, A., O. Dubovik, B. Holben, T.F. Eck, F-M Breon, J. Martonchik, R. Kahn, D. J. Diner, E. F. Vermote, J-C Roger, T. Lapyonok, and I. Slutsker, "Simultaneous retrieval of aerosol and surface properties from a combination of AERONET and satellite," *Remote Sensing of the Environment,* no. 107, 2007.

[17] O'Neill, N. T., Dubovik, O., Eck, T. F., "A modified Angstrom coefficient for the characterization of sub-micron aerosols," *Applied Optics,* vol. 40, no. 15, pp. 2368-2374, 2001.

[18] O'Neill, N. T.,T. F., Eck, A. Smirnov, B. N.Holben, S. Thulasiraman, "Spectral discrimination of coarse and fine mode optical depth," *Journal of Geophysical Research,* vol. 108, no. D17, pp. 4559-4573, 2003.

[19] Mahalanobis, P. C., "On the generalized distance in statistics," *Proceedings of the National Academy of Sciences, India,* no. 2, pp. 49-55, 1936.

[20] Burton, S. P., R. A. Ferrare, C. A. Hostetler, J. W. Hair, R. R. Rogers, M. D. Obland, C. F. Butler, A. L. Cook, D. B. Harper, and K. D. Froyd, "Aerosol classification using airborne High Spectral Resolution Lidar measurements—Methodology and examples," *Atmospheric Measurement Techniques,* no. 5, pp. 73-98, 2012.

[21] Russell, P. B., Meloë Kacenelenbogen, John M. Livingston, Otto P. Hasekamp, "A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry," *Journal of Geophysical Research,* no. 119, 2014.

[22] J. D. Chan, *Characterization of aerosol optical properties over southeast asia from AERONET sun photometer measurements,* Quezon City: Unpublished Thesis, 2015.

[23] Hamill P., Giordano M., Ward C., Giles D., Holben B., "An AERONET-based aerosol classification using the Mahalanobis distance," *Atmospheric Environment,* no. 140, pp. 213-233, 2016.

[24] Cattrall, C., J. A. Reagan, K. Thome, and O. Dubovik, "Variability of aerosol and spectral lidar and backscatter and extinction ratios of key aerosol types derived from selected Aerosol Robotic Network locations," *Journal of Geophysical Research,* no. 110, 2005.

[25] Giles, D.M., B. N. Holben, T. F. Eck, A. Sinyuk, A. Smirnov, I. Slutsker, R. R., "An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions," *Journal of Geophysical Research,* no. 117, 2012.

[26] Giles, D. M., Holben, B. N., Tripathi, S. N., Eck, T. F., Newcomb, W. W., Slutsker, I., Dickerson, R. R., Thompson, A. M., Mattoo, S., Wang, S. H., Singh, R. P., Sinyuk, A., Scafer, J. S., "Aerosol properties over the Indo-Gangetic plain: mesoscale perspective from the TIGERZ experiment," *Journal of Geophysical Research,* no. 116, 2011.

[27] Sayer, A. M., A. Smirnov, N. C. Hsu, and B. N. Holben, "A pure marine aerosol model, for use in remote sensing applications," *Journal of Geophysial Research,* no. 117, 2012.

[28] Hess, M., P. Koepke, and I. Schult, "Optical properties of aerosols and clouds: The software package OPAC," *Bulletin of the Americal Meteorological Society,* vol. 5, no. 79, p. 831–844, 1998.

[29] Shettle, E. P., and R. W. Fenn, "Models for the aerosols of the lower atmosphere and the effects of humidity variations on their optical properties," Air Force Geophysics Lab, Hanscom, Massachusetts, 1979.

[30] Zaiontz, C., "Real Statistics," [Online]. Available: http://www.real-statistics.com/correlation/basic-concepts-correlation/. [Accessed 13 11 2018].

[31] Schuster, G.L., Dubovik, O., Holben, B.N., Clothiaux, E.E., "Inferring black carbon content and specific absorption from AERONET aerosol retrievals," *Journal of Geophysical Research,* no. 110, 2005.

[32] McCormick, Chris, "K-Fold Cross-Validation, With MATLAB Code," 1 August 2013. [Online]. Available: http://mccormickml.com/2013/08/01/k-fold-cross-validation-with-matlab-code/. [Accessed 14 November 2018].

[33] "Weather Online," [Online]. Available: https://www.weatheronline.co.uk/reports/climate/Vietnam.htm. [Accessed 12 January 2019].

[34] "China Highlights: Taiwan Weather," [Online]. Available: https://www.chinahighlights.com/taiwan/weather.htm. [Accessed 13 January 2019].

[35] "Meteorological Service Singapore: Climate of Singapore," [Online]. Available: http://www.weather.gov.sg/climate-climate-of-singapore/. [Accessed 13 January 2019].

[36] "Weather Online: Indonesia," [Online]. Available: https://www.weatheronline.co.uk/reports/climate/Indonesia.htm. [Accessed 13 January 2019].

[37] "PAGASA: Climate of the Philippines," [Online]. Available: http://bagong.pagasa.dost.gov.ph/information/climate-philippines. [Accessed 13 January 2019].

[38] Cattrall, C., J. Reagan, K. Thome, and O. Dubovik, "Variability of aerosol and spectral lidar and backscatter and extinction ratios of key aerosol types derived from selected Aerosol Robotic Network locations," *Journal of Geophysical Research,* vol. 110, 2005.

[39] Maenhaut, W., D.J.A. De Ridder, M.T. Fernandez-Jimenez, M.A. Hooper, B., Hooper, M. Nurhayati, "Long-term observations of

regional aerosol composition at two sites in Indonesia," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms,* vol. 189, pp. 256-259, 2002.

[40] Santoso, M., P. Hopke, A. Hidayat, D.L. Dwiana, "Sources identification of the atmospheric aerosol at urban and suburban sites in Indonesia by positive matrix factorization," *Science of the Total Environment,* vol. 397, no. 1-3, pp. 229-237, 2008.

[41] Boucher, O., D. Randall, P. Artaxo, C. Bretherton, G. Feingold, P. Forster, V.-M. Kerminen, Y. Kondo, H. Liao, U. Lohmann, P. Rasch, S.K. Satheesh, S. Sherwood, B. Stevens and X.Y. Zhang, "Clouds and Aerosols. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change," Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[42] Oanh, N.T. Kim; N. Upadhyaya, Y.-H. Zhuang, Z.-P. Haob, D.V.S. Murthy ,P. Lestarid, J.T. Villarin, K. Chengchua, H.X. Co, N.T. Dung, E.S. Lindgren, "Particulate air pollution in six Asian cities: Spatial and temporal distributions, and associated sources," *Atmospheric Environment,* vol. 40, pp. 3367-3380, 2006.

[43] Lohmann, U., Feicher, J., "Global indirect aerosol effects: a review," *Atmospheric Chemistry and Physics ,* vol. 5, pp. 715-735, 2005.