Variations in optical properties of aerosols on monsoon seasonal change and estimation of aerosol optical depth using ground-based meteorological and air quality data

F. Tan¹, H. S. Lim¹, K. Abdullah¹, T. L. Yoon¹, and B. Holben²

¹School of Physics, Universiti Sains Malaysia, 11800 Penang, Malaysia
²NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

Correspondence to: F. Tan (fuyitan@yahoo.com)

Abstract

In this study, the optical properties of aerosols in Penang, Malaysia were analyzed for four monsoonal seasons (northeast monsoon, pre-monsoon, southwest monsoon, and post-monsoon) based on data from the AERosol RObotic NETwork (AERONET) from February 2012 to November 2013. The aerosol distribution patterns in Penang for each monsoonal period were quantitatively identified according to the scattering plots of the aerosol optical depth (AOD) against the Angstrom exponent. A modified algorithm based on the prototype model of Tan et al. (2014a) was proposed to predict the AOD data. Ground-based measurements (i.e., visibility and air pollutant index) were used in the model as predictor data to retrieve the missing AOD data from AERONET because of frequent cloud formation in the equatorial region. The model coefficients were determined through multiple regression analysis using selected data set from in situ data. The predicted AOD of the model was generated based on the coefficients and compared against the measured data through standard statistical tests. The predicted AOD in the proposed model yielded a coefficient of determination $R^2$ of 0.68. The corresponding percent mean relative error was less than 0.33% compared with the real data. The results revealed that the proposed model efficiently predicted the AOD data. Prediction of our model was compared against selected LIDAR data to yield good correspondence. The predicted AOD can beneficially monitor short- and long-term AOD and provide supplementary information in atmospheric corrections.

1 Introduction

Air quality issues in Asia can be attributed to unavoidable climate change impacts and the negative impact of human anthropogenic activities arising from rapid population growth, industrialization and urbanization (IPCC, 2007, 2013). Aerosol optical depth (AOD) derived from remote sensing has potential for assessing air quality under the right circumstances since the spatial and temporal variations in AOD are large due to production sources, transport and removal processes that are all modified by local and synoptic meteorological conditions. Many small-scale studies on the optical properties of aerosols have been conducted using sun and sky scanning radiometers of AERosol RObotic NETwork (AERONET) (Holben et al., 1998). However, these methods are limited spatially relative to satellite imagery and therefore
are complementary for comprehensive studies on atmospheric aerosols. Continuous measurement of AOD data is difficult because the atmosphere is frequently cloudy. To better monitor and understand the aerosol variation, sufficient measurements are necessary (Hansen et al., 1997; Tripathi et al., 2005; Kaskaoutis et al., 2007; Kaskaoutis and Kabzaidis, 2008; Russell et al., 2010).

Southeast Asia (SEA) stands out globally in this regard as it hosts one of the most complex meteorological and environmental conditions, making remote sensing difficult both for AERONET and satellites (Reid et al., 2013). Cloud cleared data leave gaps in our remote sensing data record and conversely residual cloud contamination of remotely sensed data cause challenging tasks to scientists studying aerosols (Campbell et al., 2013). Moreover, anthropogenic biomass burning activities have increased dramatically in recent decades for land preparation and forest clearance (Field et al., 2009). These fire activities result in trans-boundary and long-range transport of aerosols that often affect air quality in both source and surrounding regions (Hyer and Chew, 2010; Reid et al., 2013; Salinas et al., 2013; Lin et al., 2014b). These aerosols will combine with locally generated aerosols. Therefore, it is important to develop a regional/local model to estimate and monitor the AOD.

Development of an empirical model to produce reliable AOD estimates for temporal air-quality monitoring at local scales is novel and necessary for SEA with potential global applications (Chen et al., 2013; Fan et al., 2013). Several researchers have used models as alternative tools to predict AOD values by using various ground-based meteorology measurements (Wang et al., 2009; Qin et al., 2010; Liu et al., 2014a). However, this approach is a new-to-Penang. Previous studies indicate that AOD is proportional to air quality parameters such as particulate matter (PM) with diameters less than 10 or 2.5 μm (PM_{10} or PM_{2.5}) (Wang and Christopher, 2003; Cordero et al., 2012; Mielonen et al., 2012; Mogi et al., 2012; Müller et al., 2012) but inversely proportional to visibility (Vis) (Horvath, 1995; Li and Lu, 1997; Peppler et al., 2000; Bäumer et al., 2008; Singh and Dey, 2012), assuming most of the aerosol is at the surface. However, there are studies stating that AOD is not always highly correlated to surface or horizontal measurements especially when an elevated layer of AOD from transported dust or biomass burning (Mahowald et al., 2007; Baladeanu et al., 2012; Chen et al., 2013; Toth et al., 2014).

In this paper, we developed an AOD prediction model based on three types of measured data: namely (i) RIH, (ii) Vis and (iii) air pollution index (API). It is important because the stated parameters have been measured routinely at many ground-based stations. The AOD prediction model based on these routine measurements is necessary to establish a long-term database for (i) climatological studies, (ii) providing continuous atmospheric columnar AOD data, and (iii) monitoring aerosol variation. Meanwhile, it is important to understand the source and dominant type of aerosol in this study. There is an absence of understanding these factors on a local scale.

The AOD measurements were obtained through the AERONET site located in Universiti Sains Malaysia (USM) with geo-coordinates 5.36° N and 100.30° E. All AERONET data were level 2 quality assured (Smirnov et al., 2000). The Vis and API data were taken from the meteorological stations at the Penang International Airport and USM. All data were taken between 2012 and 2013. The aerosol characteristics in Penang were comprehensively
analyzed based on changes in seasonal monsoons. A near real-time AOD model was established based on multiple regression analysis of Vis and API. The accuracy and efficiency of the model were evaluated to assess air quality in Penang.

2 Methodology and statistical model

The present work was based on previous studies of Tan et al. (2014a, b). They predicted AOD using multiple regression analysis based on meteorological and air quality data. The AOD prediction model has been validated and successfully proven for the southwest monsoon period (June-September, 2012) in Penang Island. However, the following issues require reconciliation: (i) under- and overprediction of AOD were not validated because of the lack of available LIDAR data to monitor the variations in the vertical profile of the aerosol distribution; (ii) the algorithm was insufficiently robust because only a four month dataset was considered; and (iii) seasonal changes other than southwest monsoon were not included in their study. The present study uses a two-year dataset (2012, 2013) at Penang to efficiently validate the algorithms proposed by Tan et al. (2014a, b).

Penang is an island located in the northwestern region of Peninsular Malaysia and lies within latitudes 5°12' to 5°30' N and longitudes 100°09' E to 100°26' E (Fig. 4). The weather is warm and humid year-round. However, two main monsoon seasons exist: northeast and southwest monsoons. Considering previous analyses on aerosol or air quality (Awang et al., 2000; Krishna Moorthy et al., 2007; Suresh Babu et al., 2007; Kumar and Devara, 2012; Chew et al., 2013; Xian et al., 2013), the monsoon period in this study was classified as follows: (i) northeast monsoon (December–March), (ii) transition period of northeast to southwest monsoon or pre-monsoon (April–May), (iii) southwest monsoon (June–September), and (iv) transition period of southwest to northeast monsoon or post-monsoon (October–November).

The AOD and Angstrom exponent were analyzed to identify the aerosol characteristics in Penang during each period. Meanwhile, the precipitable water (PW) was used to indicate the amount of the total water content in the atmosphere. The seasonal variations in AOD, Angstrom exponent, and precipitable water (PW) based on the frequency distribution patterns were identified. The aerosol types were seasonally discriminated from the scatter plot of AOD against the Angstrom exponent. Threshold values in the scatter plot for aerosol classification have been previously reported by Smirnov (2002b, 2003), Pace et al. (2006), Kaskaotis (2007), Toledano et al. (2007), Salinas et al. (2009), and Jalali et al. (2012). The data selection criteria proposed by Tan et al. (2014a) were used in this study. The seasonal back-trajectory frequency plots from the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT 4) model were used to identify the frequency occurrence of original sources of aerosol and transported pathways. Subsequently, the aforementioned datasets were used to examine the relation of the developed algorithm.

AOD, API, and Vis data were selected according to the procedure of Tan et al. (2014a) to generate predicted AOD data. AOD is computed from the solar extinction measured at 340, 380, 440, 500, 675, 1020, and 1640 nm using the automatic tracking sun and sky scanning radiometers (Holben et al., 1998). The AOD data can be obtained from AERONET (http://aeronet.gsfc.nasa.gov). AERONET data has three different levels. Level 1.0 is cloud-unscreened data, and Level 1.5 is cloud-screened data. Only level 2.0 was employed in this study because this data level is cloud screened and data assured (Smirnov et al. 2000). The Vis data were retrieved online from Weather Underground (http://www.wunderground.com)
or from NOAA satellite (http://www7.ncdc.noaa.gov/CDO/cdo). Hourly data free from
rainfall, thunderstorms, or fog during the calculations were utilized to predict the AOD data.
Air quality in Malaysia is reported in terms of API, which can be obtained from the
Department of Environment in Malaysia (http://api.ms.doe.gov.my/apims/). API is calculated
from carbon monoxide, ozone, nitrogen dioxide, sulfur dioxide and PM10. The Malaysian
Department of Environment provides a standardized procedure on how to calculate API
values (DOE, 1997).

A total of 790 data points from 2012 to 2013 were used. Initially, the datasets were separated
into (4+1) sets as follows: (i) December–March, (ii) April–May, (iii) June–September, and (iv)
October–November. The fifth or “overall” set comprised the annual data. The number of data
points for December–March, April–May, June–September, and October–November were 257,
132, 235, and 166, respectively. The data for each seasonal monsoon were further divided into
two subsets. For example, consider that data with a particular seasonal monsoon period takes
a sequential form of D1, D2, D3, D4, D5, ..., Dn where n is the total number of points. Thus, the
subsets are in the form of (D1, D2, D3, ..., Dn) and (D2, D3, D4, ..., Dn). The first data subset was used
to calibrate (Eq. 1) for AOD at 500 nm, given below:

\[
AOD = a_0 + a_1(RH) + a_2(RH)^2 + a_3(RH)^3 + a_4(Vis) + a_5(Vis)^2 + a_6(Vis)^3 + a_7/API + a_8(API)^2 + a_9(API)^3
\]  

where RH is the relative humidity (Tan et al., 2014a).

The root mean square error (RMSE), coefficient of determination (R^2), and percent mean
relative error (%MRE) between the measured and predicted AOD for each seasonal model
were calculated at 95% confidence level. The %MRE parameter was used to quantify the
systematic differences between the concentration levels. This parameter is given as
follows: %MRE = [(mean predicted AOD - mean measured AOD) / mean measured
AOD] x 100. The ability of the proposed model to produce reliable AOD estimates for
temporal air quality monitoring can be quantitatively justified or falsified based on the value
of the resultant %MRE.

Aerosols can be hydrophilic or hydrophobic, and these properties can give rise to non-trivial
contribution to AOD hydrophilic (Tang, 1996; Song et al., 2007; de Meij et al., 2012; Singh and
Dey, 2012; Ramachandran and Srivastava, 2013; Wang et al., 2013; van Beelen et al., 2014).
However, to discriminate between hydrophilic and hydrophobic aerosols requires addition
resources beyond the reach of the present study. Most fine mode aerosols such as a sulfates
that likely dominate urban industrial aerosol composition are hygroscopic and that one would
expect RH to exert a significant influence on the measured AOD. Given that Penang is
dominated by urban industrial aerosols, one would expect RH to be an important variable in
the model. However, our pre-analysis showed that RH does not contribute significantly to
AOD prediction in the proposed model. We suggest that the RH, which is very high year
round in Penang, exerts a much less influence on AOD than we would see in drier climates.
If RH was considered as a predictor, its related factors (e.g., aerosol stratification (dust or
smoke aloft), convection, and hysteresis in particles) should be taken into account. The
contribution of RH to the aerosol properties was integrated in the aerosol model (Srivastava
et al., 2012) because the net effect of RH on aerosol and related factors were difficult to
quantify. The RH contribution can be disregarded in the present model, yielding Eq. (2), given
as follows:
AOD = \( a_0 + a_1(\text{Vis}) + a_2(\text{Vis})^2 + a_3(\text{Vis})^3 + a_4(\text{API}) + a_5(\text{API})^2 + a_6(\text{API})^3 \) \hspace{1cm} (2)

The similar statistical measurements such as RMSE, \( R^2 \), \%MRE were calculated for Eq. (2) in each monsoon season. The second data subset was then used for cross-validation.

Lee et al. (2012) excluded days when the deviation between the measured and predicted values was greater than RMSE, or when the estimated AOD slope was negative because of measurement errors and cloud-contaminated AOD. Given the previous findings, the potential outliers in our model were removed using the approach of Lee et al. (2012). Then, the aforementioned procedures were repeated to calibrate and validate the AOD prediction model using new dataset (the potential outliers were removed). The predicted AOD was again compared with the measured counterpart from AERONET to determine the accuracy of the generated model.

Equation (2) was applied to retrieve the AOD for specific days when no AOD values were available. The features of predicted AOD were compared against those of the measured counterpart. The under- and overpredicted AOD were examined by RAYMETRICS LIDAR system. However, examination can only be performed when LIDAR data were available. When LIDAR data were available for examination, only the data that can clearly elucidate the under- and over-predicted AOD were selected. The LIDAR signals were pre-analyzed based on the published works of Tan et al. (2013, 2014c). The backscatter coefficients of the aerosol from LIDAR data were determined using the method of Fernald (1984). Using the obtained aerosol backscatter coefficient and an assumed LIDAR ratio, aerosol extinction coefficients can be calculated. Integrating over these aerosol extinction coefficients, AOD values were estimated. The estimated AOD values so obtained were then compared against those predicted by our developed AOD prediction model, Eq. (2).

3 Results and discussion

3.1 Climatology of Penang, Malaysia

The climatological results derived from AERONET (http://aeronet.gsfc.nasa.gov/new_web/V2/climo_new/USM_Penang_500.html), based on the work of Holben et al., (2001), for USM Penang is tabulated in Table 1. The monthly AOD (referred to as AOD 500, second column) shows that the two lowest AOD values are 0.18 and 0.19 during the inter-monsoon period (October–November and May). During the southwest monsoon period (June–September), the smoke emitted by the local area and large-scale open burning activities in Sumatra, Indonesia was transported to Malaysia and yielded the highest AOD at approximately 0.31–0.73. However, the AOD was 0.21–0.24 during the northeast monsoon period (December–February). Small aerosol particles primarily contributed to the air pollution in Penang as the average Angstrom exponents (referred to as Angstrom exponent) were higher than 1.1 in humid atmospheres, because the precipitable water values (referred to as PW) were greater than 4.1 (Okulov et al., 2002).

3.2 Seasonal variations of AOD, Angstrom exponent, and PW based on frequency distribution patterns

AERONET parameters were plotted (Fig. 1) to reveal the relative frequency distributions at
Penang for each seasonal monsoon. Frequency histograms of AOD_500 and Angstrom_{440-870} (Fig. 1a–b, respectively) indicate changes in the optical properties of aerosols, whereas Fig. 1c shows the amount of water content in atmosphere column for each season. These histograms here helped distinguish aerosol types (Pace et al., 2006; Salinas et al., 2009; Smirnov et al., 2002a, 2011). Our results show that the distributed AOD mainly ranges from 0.2 to 0.4, contributing to approximately 71% of the total occurrence (Fig. 1a). Fig. 1b shows that the Angstrom exponent is typically between 1.3 and 1.7, translating to ~ 72% of the total. About 67% of the total occurrence of PW ranged from 4.5 cm to 5.0 cm (Fig. 1c).

The maximum AOD frequency was centered near 0.2 for all seasons. The clearest season was between October and November (Fig. 1a). Penang was most polluted from June to September, most likely due to the active open burning activities in Sumatra. The AOD peak was approximately 1.4, with three peaks distributed from AOD_500 = 0.1 to AOD_500 = 1.4 (Fig. 1a). The multiple peaks imply the presence of various aerosol populations, because AOD histograms follow log-normal distribution patterns (Salinas et al., 2009). By contrast, a single peak was observed for the clearest season (October–November).

The frequency distributions as function of Angstrom exponent display a trend (Fig. 1b), in which approximately 95% of the total occurrence fell within the range of 1 Å to 2 Å. This result implies that the effect of coarse particles (e.g., dust) on the study site was minimal. This statement is supported by Campbell et al. (2013), who showed that dust particles are uncommon in Southeast Asia. However, sometimes dust particles concentration may increase above boundary layer in southeast Asia. Two noticeable peaks were observed for the Angstrom exponent during the northeast monsoon period (blue curve, Fig. 1b). These aerosols originated from the northern part of Southeast Asia, particularly Indochina, transported by the monsoon wind and mixed with locally emitted aerosols. Lin et al. (2013) analyzed the aerosols in the northern region of Southeast Asia. They found that biomass burning aerosols from Indochina were transported in high- and low-level pathways to the west, and then later shift to the southwest by northeast monsoons. Hence, these aerosols were transported in the southwest. The biomass burning aerosols were continuously transported to our study site as the wind circulation flows toward the southwest direction, according to the monthly mean streamline charts of Lin et al. (2013) from 1979 to 2010. During and before southwest monsoon, the Angstrom exponents in Penang ranged between 1.4 and 1.8, indicating the likely presence of biomass burning aerosols (Holben et al., 2001; Gerasopoulos et al., 2003; Toledano et al., 2007). They are likely to originate from local and neighboring countries. Indonesia is known to be very active in open burning during this season. Furthermore, southwest monsoon wind is likely to have transported these biomass burning aerosols to Penang.

Although the southwest monsoon period is the driest season in Malaysia, PW frequency was approximately 20% lower than that of the northeast monsoon period for PW < 4.0 (Fig. 1c). Marked variations in the PW frequency were observed during the northeast monsoon period. Almost no frequency data were obtained for PW < 3.5, except the northeast monsoon period with about 14% less than this value. The most humid period took place in April–May, with PW ranging from 5.0 to 5.5 (approximately 74% of the total occurrence).
3.3 Seasonal discrimination of aerosol types based on the relationship between AOD and Angstrom exponent

Aerosol clusters have been developed using relative simple scatter plots of AOD and Angstrom exponent. Related studies have been analyzed using AERONET data. These datasets have been applied at different locations, such as the Persian Gulf (Smirnov et al., 2002a); several oceanic regions (Smirnov et al., 2002b); Brazil, Italy, Nauru, and Saudi Arabia (Kaskaoutis et al., 2007); Spain (Toledano et al., 2007); Singapore (Salinas et al., 2009); and the Multi-filter Rotating Shadowband Radiometer in Central Mediterranean (Pace et al., 2006). The scatter plot of AOD$_{500}$ or AOD$_{440}$ against Angstrom$_{440-870}$ was used to identify the aerosol type. The wavelength range of Angstrom$_{440-870}$ was used because of its nearness to the typical size range of aerosol based on spectral AOD (Eck et al., 1999). The relation between AOD values at 500 nm and Angstrom 440–870 is usually used for aerosol classification in scatter plot diagram. Many studies used AOD values at 500 nm (Cachorro et al., 2001; Smirnov et al., 2002b, 2003; Pace et al., 2006; Kaskaoutis et al., 2007; Salinas et al., 2009) to study aerosol turbidity conditions. Optically, 500 nm is an effective visible wavelength suitable for aerosol study (Stone, 2002). In this study, AOD$_{440}$–Angstrom$_{440-870}$ and AOD$_{500}$–Angstrom$_{440-870}$ plots were used.

Aerosols were classified into five types, including dust, maritime, continental/urban/industrial, biomass burning, and mixed aerosols (Ichoku et al., 2004). Mixed aerosols in practice represent an indistinguishable type that cannot be categorized into any of the previous types. To effectively identify the aerosol distribution types in our study sites, the results were compared using different threshold criteria (Table 2). The results are presented in Fig. 2.

The thresholds proposed by Pace et al. (2006) and Kaskaoutis et al. (2007) failed to determine the maritime aerosol (MA) and dust aerosol (DA) for each season. Instead, they showed that mixed-type aerosols (MIXA) were dominant at Penang (50–72%). Urban and industrial (UIA) and biomass burning (BMA) aerosols were grouped into a single class (28–50 % of the total occurrence). Meanwhile, the threshold suggested by Smirnov et al. (2002b, 2003) failed to identify DA, UIA, and BMA, but efficiently identified MA. As a result, a large amount of MIXA was obtained (> 80 % of the total occurrence). These results reveal the extent of uncertainty, the indistinguishable aerosol types in the study sites were large.

Salinas et al. (2009) suggested that the determination of DA and BMA did not correspond entirely to the range of threshold used in our study, in which the amount of MIXA (approximately 43 % of the total occurrence) was large. Jalal et al. (2012) efficiently identified aerosol types using an alternative threshold criterion. Using their threshold, we yielded a low amount of MIXA; approximately 21 %. However, the determination of DA was unsatisfactory. The threshold criteria of Toledano et al. (2007) provided the least MIXA (< 5 %; Fig. 2). All thresholds consistently increased from June to September (Fig. 2c) and coincided with the occurrence of haze. UIA was constantly and highly distributed over Penang. Overall, the thresholds provided by Toledano et al. (2007) were selected for our study.

Based on the criteria suggested by Toledano et al. (2007), UIA class was determined as the highest frequency of occurrence in overall study period (Fig. 3). This could be as a result of Penang being an urban area. The next highest was the MA class, because of its geolocation (i.e., surrounded by the sea). BMA is also one of the major pollutants in Penang which was
produced by active burning in local and neighboring countries. These results were in
accordance with the records from our Department of Meteorological, DOE (2010). The study
site was minimally affected by coarse particles and DA, which were less than 5% in each
seasonal monsoon. These results are supported by Campbell et al. (2013) who suggest UIA,
MA, and BMA is likely the most common in Southeast Asia and maritime continent.

BMA, UIA, and MA obtained in our study during the southwest monsoon were about 45, 24,
and 19%, respectively. During the northeast monsoon period, UIA (approximately 38%) was
the major aerosol in Penang, followed by MA (30%), BMA (20%), dust (4%), and
unidentified substances (8%). However, MIXA reached 17% from April to May, which was
the highest among the seasonal monsoons. MA and UIA were 38%; the MA level was
significant from October to November (51%), followed by UIA (40%) and BMA (<1%).
The aerosol distribution in Penang was highly seasonal dependent.

3.4 Seasonal flow patterns of air parcel from the HYSPLIT_4 model for
identification of aerosol origins

From seven-day seasonal plots of the back-trajectory frequency sourced from the HYSPLIT_4
model, flow patterns reached in the Penang site were obtained (Fig. 4) for each monsoon season
averaged between the ground surface up to an altitude of 5000 m. Residence time analysis
was performed to generate the frequency plot and determine the time percentage of a specific
air parcel in a horizontal grid cell across the domain.

During the northeast monsoon period, air parcels flow southwestward from the northern part
of Southeast Asia (Fig. 4a), including Indochina, transported through the South China Sea to
reach Penang. The aerosols during the northeast monsoon period were also locally produced,
whereas those observed during the southwest monsoon period were from the Andaman Sea,
Malacca Strait, Sumatra (site of open active burning), and other more local areas.

Fig. 1b indicates the differences in the patterns (bimodal distribution pattern) of the seasonal
relative frequency of occurrence for Angstrom440-870 during the northeast monsoon compared
to other monsoon period. These differences are likely attributable to the mixing of various
aerosol sources from the northern (e.g., Indochina, Philippines, Taiwan, and eastern China)
and southern (e.g., Malaysia and Indonesia) parts of Southeast Asia (refer Fig. 4a). The
biomass burning aerosol is likely different for northern and southern SEA because of different
types of burning process. As a result, bimodal pattern was only observed for the northeast
monsoon period from the frequency distribution pattern of Angstrom440-870 (Fig. 1b).

Figure 1b reveals that the distribution patterns of Angstrom exponent between the post-
monsoon and northeast monsoon are similar. Figure 4a and d also indicate the similarities of
the air flow patterns for these monsoon seasons. Hence, a clear correspondence was observed
between Fig. 1b with Fig. 4a and d. The similarity in the patterns of Angstrom exponents for
the post-monsoon and northeast monsoon may be attributed to the mixture of aerosols from
the northern and southern parts of Southeast Asia. Given the classification results (Fig. 3), the
occurrence frequency of MA was higher during the post-monsoon and northeast monsoon
compared to the southwest and pre-monsoon period. The large amount of MA is originating
from the South China Sea and Andaman Sea.

For the pre-monsoon period, aerosols observed at Penang originated from the Malacca Strait,
Andaman Sea, the northern and some eastern areas of Sumatra, and the western part of
peninsular Malaysia, especially the local regions marked in yellow (Fig. 4b). During this
season, the air flow patterns were similar to those during the southwest monsoon (Fig. 4c).
However, a small percentage of aerosols were transported from the northern part of Southeast
Asia to Penang. A clear correlation is observed between Fig. 1b and Fig. 4b and c during pre-
monsoon and southwest monsoon.

The dominant aerosol types were UIA and MA (Fig. 3). The yellow portions in Fig. 4c
indicate that Penang, the second largest city in Malaysia and one of the most industrially-
concentrated cities, therefore UIA is a major aerosol type in this area. MA contribution to the
overall aerosol distribution is likely significantly influenced by proximity of the surrounding
sea.

3.5 Examination of predicted AOD values

The optical properties of aerosol for each monsoonal season are obtained by analyzing the
relative frequency occurrence of AOD_500 and Angstrom_{440-870}. The relative frequency plot
of PW value also shown for each monsoonal season has different precipitable water amounts.
We hypothesize that the proposed AOD prediction model should exhibit different accuracies
seasonally because the sensitivity for AOD prediction depends on the distribution patterns of
the measured AOD; these values were used as inputs to derive the correlation parameters of
the model. The sensitivity of AOD prediction is affected when the major occurrence
frequency is clustered around small AOD values. The insensitivity of the aerosol models to
clear atmospheric conditions was also previously observed (Zhong et al., 2007).

The model performance for each monsoonal season was tested (Table 3). The pre-monsoon
and southwest monsoon periods exhibited $R^2$ of 0.65 (RMSE = 0.114) and 0.77 (RMSE = 0.172).
However, for the transition period between post-monsoon to northeast monsoon, $R^2 < 0.45$
and RMSE ranged from 0.06 to 0.11. The increased amount of atmospheric aerosol enhanced
the predicted AOD and vice versa. This result was in agreement with the aforementioned
hypothesis. Overall, the 22 month data were satisfactory with $R^2 = 0.72$ and RMSE = 0.133.
The low value of %MRE (< 1) indicates that the model yielded accurate results for all seasons.
Given the criteria that a low %MRE corresponded to a good prediction, the “overall” dataset
yielded the least biased prediction.

High correlation was observed between the measured and predicted AOD for pre-monsoon
and southwest monsoon, in which similar air flow patterns occurred (Fig. 4b and c). Figure 1b
displays the relative frequencies of occurrence of Angstrom_{440-870}. The frequency spectra for the
pre-monsoon and southwest monsoon also indicated the same patterns for AOD (Fig. 4b and c). The spectrum of Angstrom frequency exhibited narrow peaks at 1.6 and 1.7 Å for pre-
monsoon and southwest monsoon, respectively.

The accuracy of the prediction of the AOD model in post-monsoon and northeast monsoon
was moderate when the aerosols in Penang were locally mixed with those from transported
sources, because of the wind flow pattern during these two seasons (Fig. 4a and d).
Correlation between Fig. 1b and Fig. 4a and d represent these monsoonal periods. The
spectrum of the Angstrom frequency exhibited a broad region from 1.3 Å to 1.7 Å for post-
monsoon and northeast monsoon.
By comparing the types of dominant aerosol in each monsoon, we observed that the results as
obtained in Table 3 are consistent with the information from Fig. 3. Table 3 shows higher
coefficients of determination of the proposed AOD prediction model, which can be associated
with higher amount of BMA but lower LLA and MA during pre-monsoon and southwest
monsoon period. Such observation implies that the aerosol types are possibly related to the
AOD prediction model. This similar observation result was also noticed by Chen et al. (2013).
However, the relationship between the predicted AOD and aerosol type as observed in our
model is qualitative and preliminary. Further study is needed. In addition, as mentioned in Lee
et al. (2012) and Gupta et al. (2013), the relationship between AOD and air quality at ground
level depends also on environmental factors such as extent of atmospheric mixing, chemical
composition, relative humidity, aerosol size distribution, etc.

3.6 Validation of the predicted AOD

Optimized coefficients, $a_i$ (Eq. 2), were obtained from the first subset in the overall dataset.
To validate the model accuracy, $a_i$ was used to predict AOD from the second subset (Fig. 5).
The predicted AOD exhibited high correlation to the measured AOD ($R^2 = 0.68$). In addition,
the temporal characteristics of the predictions between 2012 and 2013 were similar to those of
the measured AOD.

To examine bias, the approach proposed by Lee et al. (2012) was performed to remove the
outliers when the deviation of the predicted AOD was larger than the overall RMSE (0.133).
Approximately 21% of the total data were removed using this method. After filtering out 21% of
the potential outliers, the left-over data were used to calibrate Eq. (2). $R^2$ of this fitting
significantly increased to 0.92, with RMSE = 0.059 and % MRE = 1.17$\times$10^{-4}. After filtering
the outliers, $R^2$ and RMSE were enhanced, but % MRE remained at 10^{-4} level. Thus,

Subsequently, these new coefficients obtained were used to predict AOD data (Subset 2),
which were then compared against the measured counterpart for validation. The prediction
failed to improve in terms of $R^2$ between the predicted and measured AOD (compare the red
and black lines in Fig. 5). The %MRE increased from 0.33 to 5.99.$\times$10^{-4} As a result, the removed
data might not be the genuine outliers. In fact, the errors were attributed to the non-uniformly-
loaded atmospheric aerosols at different altitudes. We believe that the non-uniform
atmospheric mixing caused the high deviations in our predicted results, according to previous
studies (Qiu and Yang, 2000). Considering that the proposed model was established based on
ground-based sources, the aerosols are assumed to be well-mixed in the atmosphere to obey
congruency with the columnar measurement of the sun photometer. The predicted AOD were
subjected to some uncertainties, however, that were quantified in terms of RMSE because the
atmosphere is not always well-mixed.

Figure 5 indicates that most of the predicted AOD values were lower than the measured
counterparts. Tan et al. (2014c) analyzed the underprediction in these values. They used
a LIDAR system to determine the vertical profile of aerosols in Penang and found that the
aerosol concentration decreased with height up to the planetary boundary layer (PBL). This
layer was less than 2 km during the study period. The large amount of transported aerosols
above boundary layer yielded residual layers (Toth et al., 2014). Significant underestimation
of AOD occurred for thick residual layers. Only a few points were significantly
underpredicted because of the aerosol residual layer beyond PBL. Studies in Cyprus (Retalis
et al., 2010) suggested that the extent of atmospheric mixing was relatively homogeneous on scales of a few meters to tens of kilometers. Hence, the predicted results were representative of the large samples. The predicted AOD was underestimated because all measured data were taken from the ground. However, overprediction would be significant if local burning were to occur near the measurement station.

To properly validate the prediction, these data should coincide in time with those measured from API, Vis, and AOD level 2. In our case, the LIDAR data coincided only once at 12 July 2013 (Fig. 6). Figure 6a shows the vertical profile of the aerosol backscatter coefficient as a function of time (morning to evening). The brown vertical line represented the instance when both the measured and predicted AOD could be compared with the LIDAR data. Figure 6b illustrates the normalized range corrected signal (RCS) at different altitudes from 10.00 a.m. and 11.00 a.m. local time. RCS was normalized through calibration based on the theoretical molecular backscatter (USSA976 standard atmospheric model) to calibrate the performance of the LIDAR system.

Figure 6c displays the profiles of the aerosol backscatter coefficient obtained at 10:00 and 11:00 a.m. local time. Aerosols had accumulated near the ground at 10:00 a.m., which was consistent with a slightly increased value in the predicted AOD of about 0.039. By contrast, most aerosols at 11:00 a.m. were at a higher level. This result corresponds with the lower value in the predicted AOD of approximately 0.044. Therefore, the predicted AOD values were acceptable because they exhibited small deviations against the measured AOD. This result was thus valid as long as the aerosols did not considerably differ at altitude levels beneath the planetary boundary layer. The LIDAR data should be therefore considered as an independent validation method for ground-based prediction models. Aerosols are not always well-mixed in the atmosphere over Penang. Several environmental factors can cause ambiguity in the predictions (Gupta et al., 2013; Lee et al., 2012). Propagating particles within the free troposphere is a factor (Toth et al., 2014). If a significant number of elevated aerosol plumes (equivalent to aerosol residual layer) occurred over the region, then a large deviation from the predicted will be produced. Therefore, it can be inferred that a small group of highly underpredicted results (Fig. 5) may be attributed to a significant layer of high-level transported aerosol.

3.7 Applications of the proposed model in the absence of measured AOD data

Our proposed model generates AOD data when those from AERONET are unavailable. We described the procedure to predict AOD data. Only the API data for 7.00 a.m., 11.00 a.m., and 5.00 p.m. (local time) were available (http://apims.doe.gov.my) before 24 June 2013. The API data were provided hourly beyond this date. In this study, approximately 5% of the data were discarded due to fog, rain, or thunderstorms, and only 4493 data points were retained. Figure 7 shows the predicted results from 2012 to 2013, which overlapped with the measured AOD data to simplify the comparison. The average AOD was 0.31 based on 4493 predicted data for the entire study period, which was near that of AERONET (about 0.29).

As an illustration, we selectively examine into three separate data windows (28 September, 17 October, and 30–31 October 2013; Fig. 8a–c) to analyze variations in the predicted and measured AOD values. The predicted AOD and CIMEL sun photometer data are shown as
blue and red dotted lines, respectively. AOD variations were continuously generated by the
proposed model based on the hourly data from ground-based measurements. The unrecorded
information by the sun photometer could be reproduced by the proposed method (Fig. 8). The
model coefficients were trained under cloud-free conditions. Hence, the hourly AOD data
could be generated anytime to compensate for the absence of measured AOD data during
cloudy periods.

The proposed model was independently verified using four selective sets of LIDAR data. We
generated these data and compared them against the temporal plots of the aerosol
backscattering coefficient signal (Fig. 9). The rectangles in Fig. 9a correspond to the
window periods for the LIDAR signal (Fig. 9b). The variability in the retrieved AOD for the
given window periods (Fig. 9a) correspond well to the intensity variations in the aerosol
backscattering coefficient signal (Fig. 9b). The LIDAR signals reveal the fidelity of our
predicted AOD because the low (high) intensities of aerosol backscattering coefficient signal
corresponded to low (high) AOD. The high intensities at 1–1.5 km altitudes (low cloud
distributions) are represented by green ovals. Although clouds were present within the
selected time windows, the retrieved AOD remained invariant.

To strengthen our AOD prediction model, the variability in the retrieved AOD for the given
window periods (Fig. 9a), were compared to AOD retrieved from the LIDAR signal. Our
LIDAR uses a laser pulse of wavelength 355 nm, whereas the AERONET data are taken at a
different wavelength. A conversion is performed to obtain AOD data from AERONET at 355
nm as described in the following:

\[ \kappa = -\left( \frac{\ln \tau_{355}}{\ln \lambda_1} \right) \]

(3)

Therefore, AOD at wavelength 355 nm can be calculated as

\[ \tau_{355} = \tau_{355} \times \left( \frac{\lambda_2}{\lambda_1} \right)^{-\kappa} \]

(4)

After the conversions, we repeat the procedure in Section 2 to obtain a new set of coefficients
at 355 nm for the AOD predicting model.

Next, AOD value is obtained from the LIDAR signal. A LIDAR ratio (L) is a constant,
defined as the ratio of aerosol extinction coefficient (\(a_a\)) and backscatter coefficient (\(b_s\)) (see
Eq. (5)). The value of L depends on the particle size distribution, shape and composition of the
aerosols in the atmosphere. R in Eq. (5) is the range or altitude. \(a_a\) and \(b_s\) are obtained once \(b_s\) and
L are known. The value of L has to be assumed for an elastic LIDAR system (He et al., 2006;
Lopes et al., 2012). Normally, L values can range from 20–40 sr for clean and polluted marine
aerosol particles or dust, urban aerosols (40-60 sr), and biomass burning aerosols (60-80 sr) as
suggested by Chew et al. (2013). In our case, we set L = 70 sr, because this window period is
commonly affected by the biomass burning aerosol (refer to the relative frequency of
dominant of aerosol types in the southwest monsoon, in Fig. 3). Additionally, other studies conducted by Tesche et al. (2011) and Lopes et al. (2012) also suggested L = 79 sr for biomass burning aerosols. AOD values \( \tau_a \) can be obtained using Eq. (6), where \( R_{\text{max}} \) is the maximum height of aerosol distribution, and \( R_h \) is height where the overlap function, \( O(R) = 1 \). Inaccurate assumption of L can lead to large errors in the retrieval of \( \alpha_a \) and \( \tau_a \) (He et al., 2006) especially under inhomogeneous atmospheric conditions. Therefore, 10% uncertainty of L and typical values of 7% uncertainty for \( \alpha_a, \beta_a \) are set to estimate potentially erroneous values of \( \alpha_a \) at any given \( R \) in an atmospheric profile. Finally, all uncertainties in the profile are summed to obtain the uncertainty of the estimated columnar AOD.

\[
L(R) = \frac{\alpha_a(R)}{\beta_a(R)} \quad (5)
\]

\[
\tau_a = \int_{R_h}^{R_{\text{max}}} \alpha_a(R) \, dr \quad (6)
\]

If the LIDAR signal is affected by cloud, the AOD data calculated from the LIDAR signal will be removed. Then the predicted AOD from our model and that calculated from LIDAR signal was compared. The result of comparison between the predicted AOD (by our model) and that derived from LIDAR is shown in Fig. 10a and b. Fig. 10a show the correlation between these two sets of data is high, as \( R^2 \) obtained is 0.86 with RMSE = 0.20. Fig. 10b also indicated that the predicted AOD values from our model are within the error bars of estimated AOD from the LIDAR signal. However, the AOD prediction model is less sensitive during clear atmospheric conditions on 13 Aug (as shown in Fig. 10b). The comparison indicated that the results agreed with the aforementioned hypothesis made in Section 3.5. Via this independent check, the robustness of the AOD prediction model has been further clarified.

3.8 Comparison with other linear regression models

The proposed model was compared against other AOD-predicting models in the literature. Table 4 shows the \( R^2 \) values of selected AOD-predicting models calculated using the first data subset by our model (Sect. 2). The \( R^2 \) values in Table 4 were compared with those of the overall dataset (Table 3). Retailis et al. (2010) suggest a simple linear regression analysis to predict AOD from the Vis data. Mahowald et al. (2007) suggest a similar linear regression model for the AOD prediction model, in which the Vis data were converted to surface extinction coefficients \( b_{\text{ext}} \) using the Koschmieder equation \( \text{Vis} = K b_{\text{ext}} \), where \( K (= 3.912) \) is the Koschmieder constant (Koschmieder, 1924). Two other AOD-predicting models were also compared (Gao and Zha, 2010; Chen et al., 2013). In these models, linear regression analysis for AOD and PM10 was carried out to predict the surface air quality. The approaches can also be used to retrieve AOD after appropriate conversion procedures. Initially, we converted the API data into PM10 via the guidance on air pollutant index from DOE (1997). The obtained PM10 values were inputted into the linear regression formula to predict AOD. The linear regression yielded \( R^2 \leq 0.6 \) with RMSE approximately 0.16 and above, which was much lower than that of our model (\( \leq 0.72 \) with RMSE = 0.13) based on the comparison of \( R^2 \) values. This result implied the dominance of the proposed model in terms of \( R^2 \) and RMSE.
4 Conclusions

Seasonal variation in the primary aerosol types and their characteristics in Penang were analyzed from February 2012 to November 2013. The aerosol types for a specific monsoon period were determined by applying threshold criteria on the scatter plots between aerosol optical depth (AOD) and Angstrom exponent. The threshold criteria from Smirnov et al. (2002b, 2003), Pace et al. (2006), Kaskaoutis et al. (2007), Toledano et al. (2007), Salinas et al. (2009), and Jalil et al. (2012) determined the aerosol types. The testing results indicated that the threshold criteria by Toledano et al. (2007) were the most reliable, because of the minimal occurrence value of the indistinguishable aerosols (referred as mixed-type aerosols, MIXA). For the entire study period, the biomass burning aerosols (BMA) abruptly increased during the southwest monsoon period because of active open burning activities in local areas and neighboring countries. During the northeast monsoon period, the optical properties (e.g., size distribution patterns) of the aerosols were unique. Two noticeable peaks were observed in the occurrence frequency of the Angstrom exponents compared with the single peaks for other monsoon seasons. These results were attributed to the mixing of aerosols from local sources with those from the northern part of Southeast Asia, caused by the northeast monsoon winds. Urban and industrial aerosols (UIA) and marine aerosol (MA) were the major aerosols in Penang throughout the year. Dust aerosols (DA) negligibly contributed to the emissions in Penang. The variation in aerosol types for different monsoon seasons yielded distinct optical properties.

Previous models used simple regression analysis between AOD and meteorological parameters to predict the corresponding AOD data. In this study, multiple regression analysis was used in the proposed model. Two predictors (API and Vis) were introduced to increase the statistical reliability. To verify the robustness of multiple regression analysis, in contrast to the simple regression approach, AOD data based on previous simple models were retrieved (Mahowald et al., 2007; Gao and Zha, 2010; Retalis et al., 2010; Chen et al., 2013). The R² and RMSE values in our model are ≤ 0.72 and 0.13. These figures are comparable with the results of other relevant work which obtained R² ≥ 0.60 and RMSE approximately 0.16 and above (see Table 4). The comparison indicates that the quality of our AOD prediction is statistically better than those simple models.

In addition, predicted AOD from our model were compared with the data derived from LIDAR system. The values of R² and RMSE (0.86 and 0.20) indicate very favorable between our model and LIDAR-derived data at wavelength 355 nm. This has added additional weight to the robustness of the developed AOD prediction model.

Our algorithm can properly predict the AOD data during non-retrieval days caused by the frequent occurrence of clouds in the equatorial region. The proposed model yielded reliable and apply real-time AOD data despite the availability of the measured data for limited time points. The predicted AOD data are beneficial to monitor aerosols in short- and long-term behavior and provide supplementary information in atmospheric correction.

Acknowledgements

The authors gratefully acknowledge the financial support provided by RU (grant no.
1001/PFIZIK/811228) and RUI-PRGS grants (grant no. 1001/PFIZIK/846083). The authors would like to thank the members of the NASA Goddard Space Flight Center for setup assembly, as well as the site members who maintained the AERONET in Penang. The authors also acknowledge A. Smirnov from NASA for fruitful discussions on certain issues.

References


Barlaucanu, R., Stefan, S., and Radulescu, R.: Correlation between the particulate matter (PM10) mass concentrations and aerosol optical depth in Bucharest, Romania, Romanian Reports in Physics, 64, 1085-1096, 2012.


de Meij, A., Pozzer, A., Pringle, K. J., Tost, H., and Lelieveld, J.: EMAC model evaluation and analysis of atmospheric aerosol properties and distribution with a focus on the


14 Holben, B. N., Tanré, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N.,


Kumar, S. and Devara, P. C. S.: A long-term study of aerosol modulation of atmospheric and surface solar heating over Pune, India, Tellus B, 64, 18420, doi:10.3402/tellusb.v64i0.18420, 2012.


Salinas, S. V., Chew, B. N., and Liew, S. C.: Retrievals of aerosol optical depth and


### Table 1. Average values of model-related parameters from the database collected from November 2011 to November 2013 in USM Penang (latitude, 05°21’ N; longitude, 100°18’ E; elevation, 51 m).

<table>
<thead>
<tr>
<th>Month</th>
<th>AOD_500</th>
<th>sigma</th>
<th>Angstrom 440–870</th>
<th>sigma</th>
<th>PW</th>
<th>sigma</th>
<th>PW</th>
<th>N</th>
<th>Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>JAN</td>
<td>0.24</td>
<td>0.09</td>
<td>1.33</td>
<td>0.18</td>
<td>4.19</td>
<td>0.47</td>
<td>21</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>FEB</td>
<td>0.21</td>
<td>0.09</td>
<td>1.39</td>
<td>0.23</td>
<td>4.44</td>
<td>0.58</td>
<td>18</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MAR</td>
<td>0.36</td>
<td>0.16</td>
<td>1.41</td>
<td>0.19</td>
<td>4.15</td>
<td>0.58</td>
<td>31</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>APR</td>
<td>0.32</td>
<td>0.19</td>
<td>1.42</td>
<td>0.16</td>
<td>4.78</td>
<td>0.53</td>
<td>29</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MAY</td>
<td>0.19</td>
<td>0.07</td>
<td>1.10</td>
<td>0.33</td>
<td>4.48</td>
<td>0.43</td>
<td>11</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>JUN</td>
<td>0.48</td>
<td>0.35</td>
<td>1.30</td>
<td>0.33</td>
<td>4.56</td>
<td>0.37</td>
<td>14</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>JUL</td>
<td>0.31</td>
<td>0.18</td>
<td>1.39</td>
<td>0.21</td>
<td>4.50</td>
<td>0.49</td>
<td>14</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>AUG</td>
<td>0.73</td>
<td>0.39</td>
<td>1.50</td>
<td>0.19</td>
<td>4.58</td>
<td>0.25</td>
<td>13</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>SEP</td>
<td>0.35</td>
<td>0.23</td>
<td>1.40</td>
<td>0.17</td>
<td>4.78</td>
<td>0.45</td>
<td>14</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>OCT</td>
<td>0.19</td>
<td>0.08</td>
<td>1.31</td>
<td>0.19</td>
<td>4.48</td>
<td>0.32</td>
<td>16</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>NOV</td>
<td>0.18</td>
<td>0.07</td>
<td>1.31</td>
<td>0.20</td>
<td>4.72</td>
<td>0.41</td>
<td>24</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DEC</td>
<td>0.21</td>
<td>0.04</td>
<td>1.41</td>
<td>0.20</td>
<td>4.67</td>
<td>0.27</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>YEAR</td>
<td>0.31</td>
<td>0.16</td>
<td>1.36</td>
<td>0.10</td>
<td>4.53</td>
<td>0.20</td>
<td>213</td>
<td>22</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Threshold values of AOD and Angstrom$_{440}$ for aerosol classification. Abbreviations: MA = maritime, DA = dust, UIA = urban and industrial, BMA = biomass burning, MIXA = mixed-type aerosols. MIXA represents indistinguishable aerosol type that lies beyond the threshold ranges.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Angstrom$_{440}$</td>
<td>AOD 440</td>
<td>Angstrom$_{440}$</td>
<td>AOD 440</td>
<td>AOD 500</td>
</tr>
<tr>
<td>MA</td>
<td>0.5–1.7</td>
<td>≤0.3</td>
<td>0–2</td>
<td>≤0.2</td>
<td>0.5–1.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤0.15</td>
</tr>
<tr>
<td>DA</td>
<td>≤1.0</td>
<td>≥0.4</td>
<td>≤1.05</td>
<td>≤0.11 (only this value is for AOD$_{870}$)</td>
<td>≤1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≤0.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.2</td>
</tr>
<tr>
<td>UIA</td>
<td>≥1.0</td>
<td>0.2–0.4</td>
<td>≥1.05</td>
<td>0.2–0.4</td>
<td>≥1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥1.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.4</td>
</tr>
<tr>
<td>BMA</td>
<td>≥1.0</td>
<td>≥0.7</td>
<td>≥1.4</td>
<td>≥0.35</td>
<td>≥1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>≥0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Calculated results for the AOD prediction model [Eq. (2)] from 2012 and 2013 data.

<table>
<thead>
<tr>
<th>Seasonal monsoon months</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>% MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast monsoon</td>
<td>0.41</td>
<td>0.110</td>
<td>$8.34 \times 10^{-4}$</td>
</tr>
<tr>
<td>Pre-monsoon</td>
<td>0.64</td>
<td>0.114</td>
<td>$8.33 \times 10^{-4}$</td>
</tr>
<tr>
<td>Southwest monsoon</td>
<td>0.77</td>
<td>0.172</td>
<td>$-1.50 \times 10^{-3}$</td>
</tr>
<tr>
<td>Post-monsoon</td>
<td>0.42</td>
<td>0.061</td>
<td>$-7.50 \times 10^{-4}$</td>
</tr>
<tr>
<td>Overall</td>
<td>0.72</td>
<td>0.133</td>
<td>$-1.11 \times 10^{-4}$</td>
</tr>
</tbody>
</table>
Table 4. $R^2$ values of the AOD predicted by selected linear regression models from the literature.

<table>
<thead>
<tr>
<th>Model</th>
<th>Author(s)</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{AOD} = a_0 + a_1(\text{Vis})$</td>
<td>(Retalis et al., 2010)</td>
<td>0.56</td>
<td>0.166</td>
</tr>
<tr>
<td>$\text{AOD} = a_0 + a_1(\text{b}_{\text{ext}})$</td>
<td>(Mahowald et al., 2007)</td>
<td>0.58</td>
<td>0.162</td>
</tr>
<tr>
<td>$\text{AOD} = a_0 + a_1(\text{PM10})$</td>
<td>(Gao and Zha, 2010; Chen et al., 2013)</td>
<td>0.60</td>
<td>0.159</td>
</tr>
<tr>
<td>$\text{AOD} = a_0 + a_1(\text{Vis}) + a_2(\text{Vis})^2$</td>
<td>Current Study</td>
<td>0.72</td>
<td>0.133</td>
</tr>
<tr>
<td>$+ a_3(\text{Vis})^3 + a_4(\text{API}) + a_5(\text{API})^2 + a_6(\text{API})^3$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**Figure 1.** Seasonal relative frequencies of occurrences of (a) AOD\_500, (b) Angstrom\_{440-870}, and (c) PW in Penang for February 2012 to November 2013. Each curve was smoothed by using moving average technique.
Figure 2. Classification of aerosol types for a) December–March, b) April–May, c) June–September, and d) October–November based on AOD–Angstrom440–870 scatter plots by proposed thresholds.
Figure 3. Seasonal classification of aerosol types based on AOD–Angstrom$_{440-870}$ scatter plots by the threshold proposed by Toledano et al. (2007).
Figure 4. Seasonal back-trajectory frequency plot by the HYSPLIT_4 model for a) northeast monsoon, b) pre-monsoon, c) southwest monsoon, d) post-monsoon, and e) overall study period at Penang, which was marked as a five-edged star.
Figure 5. Predicted and measured AOD at 500 nm against Julian days in 2012 and 2013.
Figure 6. a) Profiles of the aerosol backscatter coefficients (km$^{-1}$sr$^{-1}$) recorded on 12 July 2013. No data were acquired from 12:00 PM to 2:00 PM. The brown lines represent the moment of acquisition of sun photometer; b) normalized range corrected signals at different altitudes; c) profiles of the aerosol backscatter coefficient (beta) obtained from 10 AM to 11 AM for the brown lines in a).
Figure 7. Predicted AOD_500 data plotted against the period from 2012 to 2013. Rectangles 1 and 2 correspond to the data recorded on 24–25 July and 13–14 August 2013, respectively. These data were used for comparison with those obtained from LIDAR (Fig. 9).
Figure 8. Hourly AOD recorded on a) 28 September, b) 17 October, and c) 30–31 October 2013 from AERONET (red dotted line) and predicted AOD_500 (blue dotted line). The predicted graphs reveal temporal variations that tally with those of the measured data points.
Figure 9. Hourly retrieved AOD recorded on a) 24–25 July and 13–14 August 2013 (rectangles, Fig. 7). b) Temporal plots of the aerosol backscattering coefficient signal from the LIDAR system (morning to evening) for the corresponding periods in the rectangles of a). Green ovals represent low cloud distributions.
Figure 10. a) A scatter plot for AOD_355 predicted from our model versus the AOD calculated from Raymetrics LIDAR system. b) Predicted AOD from our model and estimated AOD from LIDAR was plot versus UTC time and date. Error bars for estimated AOD from LIDAR are shown.