Atmos. Chem. Phys. Discuss., 14, 19747–19789, 2014
www.atmos-chem-phys-discuss.net/14/19747/2014/
doi:10.5194/acpd-14-19747-2014

Atmospheric Chemistry and Physics
Discussions

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. Tan[1], H. S. Lim[1], K. Abdullah[1], T. L. Yoon[1], and B. Holben[2]

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

Received: 13 June 2014 – Accepted: 14 July 2014 – Published: 31 July 2014

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

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. Validation tests were performed on the model against selected LIDAR data and yielded good correspondence. The predicted AOD can beneficially monitor short- and long-term AOD and provide supplementary information in atmospheric corrections.

1 Introduction

The direct and indirect radiative influences of aerosols have been significant sources of uncertainty in climate change based on the report by the Intergovernmental Panel for Climate Change (IPCC, 2007, 2013). The consequences of aerosol–radiation and aerosol–cloud interactions cannot be fully elucidated because of their uncertainties. These interactions are increasingly complex and compounded by high degrees of variation~~s~~ in atmospheric aerosols because of meteorological and climatic factors (Reid et al., 2012). The trans-boundary and long-range transport of aerosols interact with their local counterparts ~~(e.g., cloud droplets)~~, enhance the microphysical properties of aerosols, and affect their radiative properties and precipitation processes (Ichoku et al., 2004; Lin et al., 2013; Rosenfeld, 2007; Andreae and Rosenfeld, 2008). The global effects of aerosols on the Earth's climate are ~~hardly~~ quantifiable because of the lack of extensive and reliable measurements in most world regions (Tripathi et al., 2005; Russell et al., 2010; Hansen et al., 1997; Kaskaoutis and Kambezidis, 2008; Kaskaoutis et al., 2007).

only coarsely

references are 1) chronological 2) alphabetical

characterized

Aerosol optical depth (AOD) ~~conveniently analyzes air quality/pollution, radiation budget and radiation forcing, climate change, atmospheric corrections in remote sensing from space, and aerosol characteristics~~. The spatial and temporal variations in AOD are large because of production sources, transport and removal processes, and prevalent meteorological conditions. Given the large uncertainty in aerosol characterization, local analyses essentially verify ~~the~~ satellite measurements ~~imageries~~ because the extraction of aerosol optical properties from remote sensing data exhibits limited accuracy despite its capability to provide global-scale coverage (Levy et al., 2005; Tripathi et al., 2005; Yoram et al., 2002; Gupta et al., 2013; Zhong et al., 2007). Local studies on the optical properties of aerosols have been conducted using sun photometers and sky radiometers (Salinas et al., 2009; Holben et al., 1998; Remer et al., 2008). However, these methods are limited spatially ~~to space coverage~~ in contrast to satellite imagery. Therefore, ground- and space-based measurements complementarily perform reliable and comprehensive studies on atmospheric aerosols.

The accuracy of satellite-derived daily AOD is often assessed by comparing satellite-based AOD with the AErosol RObotic NETwork (AERONET), a network of ground-based sun photometers. AERONET is widely used to monitor, investigate, and characterize the optical properties of aerosols (Holben et al., 1998). This network provides a database to ~~atmospherically~~ correct and validate satellite-based aerosol retrievals. However, cloud-contaminated data should be removed from the AERONET database

Huang et al. 2011
Chen et al. 2011

(Smirnov et al., 2000); the process is termed as cloud screening. Hence, only a limited dataset of level 2 AOD (data have been cloud screened and quality assured) ~~can be obtained~~ are available. Meanwhile, AODs obtained from satellites, such as those from MODIS (Retalis et al., 2010), are limited because these satellites are ~~orbiting~~ in sun-synchronous orbit. Continuous retrieval of AOD data is difficult. Thus, several models have been proposed to efficiently predict and retrieve AOD. REFERENCES

Previous studies have used single parameters from ground measurements to estimate the atmospheric columnar AOD, such as in situ horizontal visibility (Vis) or particulate matter (PM) with diameters less than 10 or 2.5 µm (PM_{10} or $PM_{2.5}$). REFS? The high concentrations of atmospheric aerosols increase the AOD to effectively scatter light and reduce Vis. PM_{10} and $PM_{2.5}$ are used to physically quantify the concentration of PM at ground level. High-quantity PM records imply high aerosol concentrations at the ground surface. AOD is proportional to air quality (Müller et al., 2012; Cordero et al., 2012; Mogo et al., 2012; Mielonen et al., 2012; Wang and Christopher, 2003) but inversely proportional to Vis (Horvath, 1995; Bäumer et al., 2008; Li and Lu, 1997; Peppler et al., 2000; Singh and Dey, 2012). Vis and air quality interact with columnar AOD; hence, these parameters should be considered into the algorithm to predict AOD through multiple regression analysis. The complementary combination increases the relative accuracy of prediction.

Toth et al 2014

Three types of measurement data were used in this study, namely (i) AOD, (ii) Vis and (iii) air pollution index (API). The AOD measurements were obtained through ~~the~~ an AERONET site located in Universiti Sains Malaysia (USM). 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. The accuracy and efficiency of the model were validated and evaluated to assess ~~the~~ atmospheric pollution in Penang.

State clearly what you are doing and why

lat/lon coordinates

of what?

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. These studies have successfully proven and validated the algorithm during the southwest monsoon period. However, the following issues ~~should be addressed~~: (i) under- and overprediction of AOD were not validated because of the lack of available LIDAR data to ~~obtain~~ monitor the variations in the vertical profile of the aerosol distribution, (ii) the algorithm was insufficiently robust because only ~~the 4~~ a four month dataset ~~were~~ was considered; and (iii) seasonal changes in the southwest monsoon was only included. The present study uses a two-year dataset (2012, 2013) ~~in~~ at Penang to efficiently validate the algorithms proposed by Tan et al. (2014a, b).

where? when?

require reconciliation

ensure what you mean?

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), ~~which is~~ near the equator. ~~Seasons such as winter, spring, summer, and autumn are undefined; instead,~~ The weather is warm and humid year-round. However, two main monsoon seasons exist ~~in Penang~~, namely, northeast and southwest monsoons. Considering ~~the~~ previous analyses on aerosol or air quality (Suresh Babu et al., 2007; Krishna Moorthy et al., 2007; Kumar and Devara, 2012; Xian et al., 2013; Awang et al., 2000), the monsoon period ~~was~~ is 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 optical properties of aerosols were analyzed to identify the aerosol characteristics in Penang ~~in each monsoon~~ 'during each period'. 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 (2002, 2003, 2011), Pace

et al. (2006), Kaskaotis (2007), Toledano et al. (2007), Salinas et al. (2009), and Jalal et al. (2012). The data selection criteria proposed by Tan et al. (2014a) were used in this study. The seven-day seasonal plot of the back-trajectory frequency from the Hybrid Single-Particle Lagrangian Integrated Trajectory (HYSPLIT_4) model was used to identify the original sources of aerosol and transported pathways ~~because this model can suitably simulate air-mass movement~~. Subsequently, the obtained aerosol characteristics were used to examine the algorithm accuracy among the datasets.

AERONET, API, and Vis data were selected according to the procedure of Tan et al. (2014a) to generate ~~the~~ predicted AOD data. The in situ data were retrieved online from Weather Underground (http://www.wunderground.com) or from NOAA satellite (http://www7.ncdc.noaa.gov/CDO/cdo). ~~The data from Weather Underground were in METAR and AAXX formats, whereas those from NOAA were slightly different. Nevertheless, the information contents of both databases were essentially similar. Only the data in METAR format were used to standardize the calculation procedure.~~ 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, ~~API data~~ which can be obtained from the Department of Environment in Malaysia (http://apims.doe.gov.my/apims/). API is calculated from carbon monoxide, ozone, nitrogen dioxide, sulfur dioxide and PM_{10}. The Malaysian Department of Environment provides a standardized procedure on how to calculate API values (DOE, 1997).

AERONET data were recorded at the Coordinated Universal Time (UTC), whereas in situ and API data were recorded at local time (UTC + 8 h). All data were required as inputs in the proposed algorithm to predict the AOD data. To standardize the implementation of the proposed algorithm, the data of AERONET, in situ measurements, and API were converted to Julian days according to the UTC and compared with one another because they originated from different sources. Hence, the overlapped data within a time interval of ±30 min were retained; otherwise, these data were discarded.

is this really necessary?

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 that were sourced from alternatively selected data (in temporal sequence) for cross-validation. For example, consider that data with a particular seasonal monsoon period takes a sequential form (D1, D2, D3, D4, D5, ...). Thus, the subsets are in the form of (D1, D3, D5, ...) and (D2, D4, D6,...). The first data subset was used to determine the correlation between the parameters and AOD at 500 nm (Eq. 1), which was the original model of Tan et al. (2014a), and given as follows:

$$\begin{aligned}\mathrm{AOD} = {} & a_0 + a_1(\mathrm{RH}) + a_2(\mathrm{RH})^2 + a_3(\mathrm{RH})^3 + a_4(\mathrm{Vis}) + a_5(\mathrm{Vis})^2 + a_6(\mathrm{Vis})^3 \\ & + a_7(\mathrm{API}) + a_8(\mathrm{API})^2 + a_9(\mathrm{API})^3 \end{aligned} \tag{1}$$

¶ - new paragraph

where RH is the relative humidity. The second data subset was used to predict AOD in each seasonal monsoon and validate the accuracy of the prediction based on the parameters (e.g., a_0 and a_1) obtained from the correlation procedure. The algorithm of Tan et al. (2014a) was tested to determine the correlation at 95 % confidence level for each seasonal monsoon. 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 period were calculated. 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] × 100. The ability of the proposed model to produce reliable AOD estimates for temporal air monitoring can be quantitatively justified or falsified based on the quality of the resultant %MRE.

Aerosols ~~could~~ can be hydrophilic or hydrophobic, and these properties ~~could~~ can give rise to non-trivial contribution to AOD retrieval (Ramachandran and Srivastava, 2013; Singh and Dey, 2012; de Meij et al., 2012; Tang, 1996; Song et al., 2007; van Beelen et al., 2014; Wang et al., 2013). However, to discriminate whether the aerosols are hydrophilic

or hydrophobic requires addition resources beyond the reach of the present study. On the other hand, our pre-analysis showed that RH does not contribute significantly to AOD prediction in the proposed model. 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 hardly quantifiable. The RH contribution can be disregarded in the present model, yielding Eq. (2). ~~The~~ Results were obtained from ~~the~~ correlation analysis based on Eq. (2) given as follows:

$$\mathrm{AOD} = a_0 + a_1(\mathrm{Vis}) + a_2(\mathrm{Vis})^2 + a_3(\mathrm{Vis})^3 + a_4(\mathrm{API}) + a_5(\mathrm{API})^2 + a_6(\mathrm{API})^3 \tag{2}$$

Lee et al. (2012) excluded ~~the~~ 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 outliers in our model were removed using the approach of (Lee et al., 2012). The predicted AOD was 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 backscatter coefficients of the aerosol were determined using the method of Fernald (1984). The LIDAR signals were pre-analyzed based on the published works of Tan et al. (2013, 2014c).

3 Results and discussion

3.1 Climatology of Penang, Malaysia

~~Given the~~ From local climatology results ~~from the aerosol robotic network~~ derived by AERONET (http://aeronet.gsfc.nasa.gov/new_web/V2/climo_new/USM_Penang_500.html), the monthly AOD (referred to as AOD_500) in USM Penang showed that the lowest AOD ranged from 0.18–0.19 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 ~~by the monsoon wind~~ 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 ~~because~~ the average Angstrom exponents (referred to as Angstrom$_{440-870}$) were higher than 1.1 in humid atmospheres, and the precipitable water values (referred to as PW) were greater than 4.1.

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

~~The aerosol properties~~ AERONET parameters were plotted (Fig. 1) to reveal the relative frequency distributions ~~of the atmospheric aerosols in~~ at Penang for each seasonal monsoon. ~~The~~ Frequency histograms of AOD_500, Angstrom$_{440-870}$, and PW (Fig. 1a–c, respectively) indicate~~d~~ changes in the optical properties of aerosols with season~~al variations~~; these histograms helped ~~identify the~~ aerosol types (Pace et al., 2006; Salinas et al., 2009; Smirnov et al., 2011, 2002a). Our results show~~ed~~ that the distributed AOD mainly range~~d~~s from 0.2 to 0.4, ~~and~~ contribut~~ed~~ing to approximately 71 % of the total occurrence (Fig. 1a). Fig. 1b shows that the Angstrom exponent is typically between 1.3 and 1.7, ~~which~~ translat~~es~~ing to ∼ 72 % of the total occurrence. About 67 % of the total occurrence of PW ranged from 4.5 cm to 5.0 cm (Fig. 1c).

have distinguish

The maximum peak of AOD frequency was centered ~~at~~ near 0.2 for all seasons. The clearest season was between October and November ~~(Fig. 1a(i) and (ii))~~. Penang was most polluted from June to September, ~~because of~~ most likely due to the active open burning activities in Sumatra. The AOD peak was approximately 1.4, with ~~about~~ three peaks distributed from AOD_500 = 0.1 to AOD_500 = 1.4 ~~(Fig. 1a(i) and (ii))~~. The multiple peaks ~~implied~~ imply likely 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 for the Angstrom exponent display~~ed~~ noticeable seasonal trends (Fig. 1b), ~~and ranged from~~ values range from 1 to 2 with approximately 95 % of the total occurrence. This result impl~~ied~~ies that the effects of coarse particles (e.g., dust) on the study site was minimal, ~~which was attributed to the absence of desert areas or the lengthy distance from these areas. The absence of desert areas impeded the transport of dust to the study site.~~ However, 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 winds, 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 ~~by~~ during west and northeast monsoons. Hence, these aerosols were transported in the southwest ~~direction~~. 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) from Indonesia.

Although the southwest monsoon period ~~was~~ is the driest season in Malaysia, ~~the recorded~~ PW frequency was approximately 21 % lower than that of the northeast monsoon period for PW < 4.0 (Fig. 1c). Marked variations in the PW frequency were

Campbell et al. 2013 describe occurrence of elevated dusts over region

These parameters are only telling you what is likely to be occurring must be careful!

* See Omar et al. 2005 (JGR) He uses more info than AOD and AE, incl. size distribution. You need to either add references to support this claim, or you need to be more circumspect w. respect to your tone and application of these parameters

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

clusters developed using relatively simple

Aerosol~~s~~ have been ~~widely classified by the~~ scatter plots of AOD and Angstrom exponent~~. AOD provides information of aerosol loading in the atmosphere column through the extinction of radiation rate for a specific wavelength. The Angstrom exponent determines the aerosol size in coarse and fine modes from the slope with wavelengths that depend on AOD in logarithmic coordinates. Therefore, the AOD–Angstrom exponent scatter plots indicate the amount and dimension of the observed aerosols. The corresponding distribution pattern was grouped into a few clusters to determine the aerosol species~~. 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); Brazil, Italy, Nauru, and Saudi Arabia (Kaskaoutis et al., 2007); Spain (Toledano et al., 2007); Singapore (Salinas et al., 2009); several oceanic regions (Smirnov et al., 2011); Kuching (Jalal et al., 2012); 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. The AOD values at 500 nm are normally used to indicate the turbidity conditions (Cachorro et al., 2001; Smirnov et al., 2002b, 2003; Kaskaoutis et al., 2007; Pace et al., 2006; Salinas et al., 2009). Optically, 500 nm is an effective visible wavelength suitable for aerosol study (Stone, 2002). In this study,

you've said all of this?

AOD_440–Angstrom$_{440–870}$ and AOD_500–Angstrom$_{440–870}$ plots were used ~~to classify the aerosols.~~

~~The~~ Aerosols were classified into five types, ~~namely~~ included in dust, maritime, continental/urban/industrial, biomass burning, and mixed aerosols (Ichoku et al., 2004); mixed aerosols in practice represents an indistinguishable ~~aerosol~~ 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 of aerosol classification using different threshold criteria~~ are presented in Fig. 2.

The thresholds proposed by Kaskaoutis et al. (2007) and Pace et al. (2006) failed to determine the maritime aerosol (MA) and dust aerosol (DA) for each season. Instead, they showed that mixed-type aerosols (MIXA) were dominant ~~in~~ set 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. (2002, 2003, 2011) 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. ~~Thus, other options should be considered.~~

order of references

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 ~~the~~ 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 ~~properly~~ best suited for our study ~~site to determine the aerosol types~~.

The thresholds of AOD–Angstrom$_{440–870}$ scatter ~~plots~~ by Toledano et al. (2007) ~~used to classify the aerosol types~~ revealed that higher amount of pollutants in UIA class were identified, ~~and directly affected the air quality in Malaysia~~ (Fig. 3). The MA observed in Penang was high because of its geolocation (i.e., surrounded by the sea). The study site was minimally affected by coarse particles and DA, which were less than 5 % in each seasonal monsoon. BMA ~~was~~ is one of the major pollutants in Penang because of ~~the~~ active burning ~~activities~~. ~~Furthermore, haze occurred during the southwest monsoon because of the trans-boundary aerosols from Indonesia~~. These results were in accordance with the records from DOE (2010). 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.

when?

?

haze is STILL aerosol!

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

~~Given the~~ From seven-day seasonal plots of the back-trajectory frequency ~~by~~ solved from the HYSPLIT_4 model, ~~the~~ flow patterns ~~of the air parcel in~~ reaching the Penang site were obtained (Fig. 4) for each monsoon season ~~in terms of percentage~~ 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, ~~the~~ air parcels flow~~ed~~ southwestward from the northern part of southeast Asia (Fig. 4a), ~~which illustrated that the aerosol sources to Penang were from the former (open burning season, Lin et al., 2013)~~, including Indochina, ~~and~~ transported through the South China Sea to reach Penang. The aerosols

observed

during the northeast monsoon period were ~~also~~ locally produced, whereas those ~~obtained~~ during the southwest monsoon period were directly more from Andaman Sea, Malacca Strait, Sumatra (site of open active burning), and other local areas. more

~~The patterns in seasonal relative frequency of air parcel movement were significantly different (Fig. 4a)~~. Comparison with Fig. 1b indicated the differences in the patterns of the seasonal relative frequency of occurrence for Angstrom$_{440–870}$ during the northeast monsoon. These differences ~~were~~ are likely able attribut~~ed~~ 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. As a result, ~~the bimodel~~ pattern was only obtained during the northeast monsoon period (December–March) ~~because the local aerosol sources were mixed with several sources from Indochina that contained different sizes~~ compared with those from the southern counterpart.

??

bimodal (of what?)

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 post-monsoon and northeast monsoon ~~was attributed~~ may be attributable to the mixture of aerosols from northern and southern parts of Southeast Asia. Given the classification results (Fig. 3), MA was the major aerosol during the post-monsoon and northeast monsoon, ~~The large amount of MA~~ originat~~ed~~ing from the South China Sea and Andaman Sea.

you don't know anything for certain from your data

For the pre-monsoon period, the 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 aerosol was transported from the northern part of Southeast Asia to Penang. A clear ~~correspondence~~ correlation ~~was~~ is observed between Fig. 1b with 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. 4a–e indicate that Penang, the second largest city in Malaysia and one of the most industrially concentrated cities, was a major aerosol trap because of local and industrial emissions. MA contribution to the overall aerosol distribution ~~could be significant because the study site was surrounded by the~~ sea.

is likely significantly influenced by proximity of the surrounding sea

? source ?

3.5 Examination of predicted AOD values

averaged? imply

~~Various Monthly AOD and Angstrom exponents from climatological data implied that each period exhibit different aerosol distributions in Penang~~. Seasonal analysis of the relative frequency occurrence of AOD_500, Angstrom$_{440–870}$, and PW clearly distinguished the dominant optical properties of aerosol for each monsoonal season. We hypothesized that the proposed model should exhibit different accuracies each season because the sensitivity for AOD prediction depended 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 was affected when the major occurrence frequency was clustered around small AOD values. The insensitivity of the aerosol models in clearing atmospheric conditions was also previously observed (Zhong et al., 2007). Conversely, the model appropriately predicted the AOD ~~data when~~ the corresponding input data were clustered around large values.

proposed model of what?

The model performance for each monsoonal season was tested (Table 3). The pre-monsoon and southwest 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~~ previous hypothesis. ~~Meanwhile, the~~ "overall" 22 month data was satisfactory with $R^2 = 0.72$ and RMSE = 0.133. The low value of %MRE (< 1) indicated 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.

REF

were

~~The~~ aerosol types and distribution patterns could be elucidated from the ~~model results~~. ~~These~~ parameters strongly depended on the ~~changes in~~ wind flow from one seasonal monsoon ~~to the next~~. Aerosols ~~were transported~~ by monsoonal flows which were combined into the ~~atmosphere from~~ different sources. Haze could become noticeable (i.e., AOD ~~value~~ is high) when higher amount ~~of aerosols~~ was injected into the atmosphere from different sources, especially during large-scale open burning activities. High correlations were 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 the occurrence of Angstrom$_{440–870}$. The frequency spectra for pre-monsoon and southwest monsoon also indicated the same patterns for AOD (Fig. 4b and c). The spectrum of the Angstrom frequency exhibited narrow peaks at 1.6 and 1.7 Å for pre-monsoon and southwest monsoon, respectively.

when

The accuracy of the prediction of the AOD model was moderate the aerosols in Penang were locally mixed with those from foreign sources because of the winds during post-monsoon and northeast monsoon ~~characterized by similar air flow patterns~~ (Fig. 4a and d). Correlations between Fig. 1b with Fig. 4a and d ~~were observed for~~ 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. ~~The broadened region implied that the particle size was largely distributed~~. The relationship between AOD to the air quality at ground surface depended on environmental factors, such as RH, aerosol size distribution, and chemical composition. These factors were disregarded in the AOD model, yielding deviations in the predicted values (Gupta et al., 2013; Lee et al., 2012).

represent

3.6 Validation of the predicted AOD

~~The~~ 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 the AOD from the second subset ~~of the "overall" dataset~~ (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.

~~However, the predicted AOD values were over- or underpredicted.~~ To examine ~~these~~ biases, 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 ~~calculated~~ "overall" RMSE (0.133). Approximately 21 % of the total data were removed using this method. Without the outliers, the measured AOD data (subset 1) were against a_i (Eq. 2). R^2 of this fitting significantly increased to 0.92 with RMSE = 0.059 and % RME = 1.17×10^{-4}. After filtering the outliers, R^2 and RMSE were enhanced, but % RME remained at 10^{-4} level.

Not a sentence?

the

The coefficients without the outliers were used to predict AOD data, which were ~~then~~ compared against the measured counterpart (subset 2) for validation. The prediction failed to improve in terms of R^2 between the predicted and measured AOD (Fig. 5). However, the %MRE increased from 0.33 (with outliers) to 5.99 (without outliers) based on the comparison between the predicted and measured AOD.

Getting confused. Isn't this what you say here?

~~We removed the outliers based on the suggestion of Lee et al. (2012) to improve our AOD prediction. However, this model also failed to improve based on the previous analysis.~~ The removed data might not be the genuine outliers. The data exhibited large RMSE that should otherwise be removed (Lee et al., 2012); but in fact was 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 should be well-mixed in the atmosphere to obey congruency with the vertical 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. ~~In other words, the predicted values of AOD were within an error of ± RMSE.~~

however,

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 yielded residual layers ~~because of convection effects~~. Significant underestimation of AOD occurred for thick residual layers. ~~Figure 5 shows good AOD prediction.~~ 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 occurred near the measurement station.

you don't know this!

Toth et al. 2014 !!

~~LIDAR data could be used to independently validate the predicted data. However, the available LIDAR data was limited.~~ 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. RCS was normalized through calibration based on the theoretical molecular backscatter (USSA976 standard atmospheric model) to calibrate the performance of the LIDAR system.

local time

Figure 6c displays the profiles of the aerosol backscatter coefficient (beta) obtained at 10:00 and 11:00 a.m. ~~The~~ Aerosols had accumulated near the ground ~~surface~~ at 10:00 a.m., which ~~was consistent~~ with a slightly increased value in the predicted AOD of about 0.039. By contrast, ~~the accumulated~~ most aerosols at 11.00 a.m. were at a higher level

profiled

~~than those at the ground level~~. This result corresponds ~~was in accordance~~ with the lower value in the predicted AOD of approximately 0.044, ~~which was consistent with the expected result~~. Therefore, the predicted AOD values were acceptable because they exhibited small deviations against the measured AOD; thus this result was valid as long as the aerosols did not considerably differ at altitude levels beneath the planetary boundary layer. The LIDAR data should therefore be considered as an independent validation method for ground-based prediction models. ~~Comparing the consistency between the predicted results against LIDAR data could falsify or verify the correctness of the prediction model with high confidence~~. In reality, aerosols are not frequently well mixed in the atmosphere; several environmental factors can cause ambiguity in the predictions (Gupta et al., 2013; Lee et al., 2012). The small group of highly underpredicted results (Fig. 5) was attributed to the significant heterogeneity of aerosols in the atmosphere (e.g., aerosol residual layers) and the large amount of high-level transported aerosol (Tan et al., 2014b, c).

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 ~~from the web site~~ (http://apims.doe.gov.my) before 24 June 2013. The API data were provided hourly beyond this date. ~~Any in situ visibility data with a value of −9999 and those recorded as fog, rain, or thunderstorms were removed~~. 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). ~~The good agreement between the predicted and measured average AOD suggested that the model was sufficiently feasible to perform predictions.~~

As an illustration, we selectively ~~zoom into~~ examine three separate data windows (28 September, 17 October, and 30–31 October 2013; Fig. 8a–c) to analyze ~~the~~ variations in the

predicted and measured AOD values ~~at the scale of days~~. The predicted AOD and CIMEL sun photometer data are ~~were~~ shown as blue and red dotted lines, respectively. ~~The~~ availability of the measured data points are often limited because of the unavailability of ~~AERONET~~ data caused by the presence ~~of clouds and~~ robotic errors. The predicted graphs ~~exhibited temporal~~ variation trends that tally with those measured at ~~the same time~~ scale (days).

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. In addition, the proposed model can generate daytime and nighttime temporal data in contrast to AERONET. ~~Our model can be highly beneficial in monitoring the air concentration cycle because it generates continuous hourly data; hence, complementary information are provided.~~

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 corresponded to the window periods for the LIDAR signal (Fig. 9b). The ~~variation patterns~~ variability in the retrieved AOD for the given window periods (Fig. 9a) corresponded well to the intensity variations in the aerosol backscattering coefficient signal (Fig. 9b). The LIDAR signals revealed the fidelity ~~correctness~~ 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. Therefore, this result strengthened the robustness of the proposed model to perform reliable and accurate prediction and retrieval of AOD. ~~Our model could provide complementary retrieval of AOD data when AERONET data are unavailable because of the presence of clouds.~~

3.8 Comparison with other linear regression models

~~Following quantitative and qualitative validation of our model,~~ T the proposed model was compared against other AOD-predicting models in the literature. Table 4 shows the R^2 values of ~~some~~ 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). Retalis et al. (2010) suggested a simple linear regression analysis to predict AOD from the Vis data. Mahowald et al. (2007) suggested a similar linear regression model for the AOD prediction model, in which the Vis data were converted to surface extinction coefficients b_{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 ~~subjected to comparison~~ compared (Gao and Zha, 2010; Chen et al., 2013). In these models, linear regression analysis for AOD and PM_{10} 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 PM_{10} via the guidance on air pollutant index from DOE (1997). The obtained PM_{10} values were inputted into the linear regression formula to predict AOD. The linear regression yielded $R^2 \leq 0.6$, which was much lower than that of our model (≤ 0.72) based on the comparison of R^2 values for the "overall" dataset in Table 3 against those in Table 4. This result implied the dominance of the proposed model in terms of R^2.

4 Conclusions

Seasonal variations in the primary aerosol types and their characteristics in Penang were analyzed from February 2012 to November 2013. The aerosol types for a specific monsoonal period were determined by applying a threshold criteria on the scatter plots between AOD and $\text{Angstrom}_{440-870}$. The threshold criteria from Smirnov (2002, 2003, 2011), Pace et at. (2006), Kaskaotis (2007), Toledano et al. (2007), Salinas et al. (2009), and Jalal et al. (2012) determined the aerosol types. The testing results

I recommend redefining all acronyms in the Conclusion. You should write this portion as if the reader never read a word of the rest of the paper

indicated that the threshold criteria by Toledano et al. (2007) were the most reliable because of the minimal value of the predicted MIXA. For the entire study period, the 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. UIA and MA were the major pollutants in Penang throughout the year. DA negligibly contributed to the emissions in Penang ~~because deserts were nonexistent and the location was sufficiently far from known desert areas~~. ~~The small amount of DA particles was caused by vehicles and construction activities~~. The variations in aerosol types for different monsoon seasons yielded distinct optical properties.

The original prototype model of Tan et al. (2014a) feasibly predicted the AOD values based on the measured API, Vis, and RH data through multiple regression analysis. In this study, the algorithm of Tan et al. (2014a) was used and slightly modified by neglecting the RH contribution. Our results suggested that the removal of the RH contribution caused no changes in the predictability of the proposed model. The modified algorithm was quantitatively and qualitatively validated. The retrieved AOD data in the proposed model were in agreement with those measured.

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 high robustness of multiple regression analysis in contrast to the simple regression approach, AOD data based on previous simple models were retrieved (Gao and Zha, 2010; Chen et al., 2013; Retalis et al., 2010; Mahowald et al., 2007). The R^2 values in our model were compared with those what about the impact of the vertical?

Impact of study? who will use this information? Future work and/or application?

previously proposed. The results indicated that the quality of AOD prediction of our model was more dominant than those of the simple models.

Our algorithm could 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 aptly real-time AOD data despite the availability of the measured data for limited time points. The predicted AOD data are beneficial to monitor 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

Andreae, M. O. and Rosenfeld, D.: Aerosol–cloud–precipitation interactions, Part 1. The nature and sources of cloud-active aerosols, Earth-Sci. Rev., 89, 13–41, doi:10.1016/j.earscirev.2008.03.001, 2008.

Awang, M. B., Jaafar, A. B., Abdullah, A. M., Ismail, M. B., Hassan, M. N., Abdullah, R., Johan, S., and Noor, H.: Air quality in Malaysia: impacts, management issues and future challenges, Respirology, 5, 183–196, 2000.

Bäumer, D., Vogel, B., Versick, S., Rinke, R., Möhler, O., and Schnaiter, M.: Relationship of visibility, aerosol optical thickness and aerosol size distribution in an ageing air mass over South-West Germany, Atmos. Environ., 42, 989–998, doi:10.1016/j.atmosenv.2007.10.017, 2008.

Cachorro, V. E., Vergaz, R., and de Frutos, A. M.: A quantitative comparison of α-Å turbidity parameter retrieved in different spectral ranges based on spectroradiometer solar radiation measurements, Atmos. Environ., 35, 5117–5124, doi:10.1016/S1352-2310(01)00321-1, 2001.

Chen, B. B., Sverdlik, L. G., Imashev, S. A., Solomon, P. A., Lantz, J., Schauer, J. J., Shafer, M. M., Artamonova, M. S., and Carmichael, G.: Empirical relationship between par-

ticulate matter and aerosol optical depth over Northern Tien-Shan, Central Asia, Air Quality, Atmosphere and Health, 6, 385–396, 2013.

Cordero, L., Wu, Y., Gross, B. M., and Moshary, F.: Use of passive and active ground and satellite remote sensing to monitor fine particulate pollutants on regional scales, in: Advanced Environmental, Chemical, and Biological Sensing Technologies IX, Baltimore, MD, 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 Mediterranean region, Atmos. Res., 114–115, 38–69, 2012.

DOE: A Guide to Air Pollutant Index in Malaysia (API), 3rd Edn., D. O. Environment, Ed., Kuala Lumpur, Malaysia, 1–20, 1997.

DOE: Pollution sources inventory, in: Malaysia Environment Quality Report 2010, ch. 5, Department of Environment, Petaling Jaya, Malaysia, 2010.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophys. Res.-Atmos., 104, 31333–31349, 1999.

Fernald, F. G.: Analysis of atmospheric lidar observations: some comments, Appl. Optics, 23, 652–653, 1984.

Gao, J. and Zha, Y.: Meteorological influence on predicting air pollution from MODIS-derived aerosol optical thickness: a case study in Nanjing, China, Remote Sensing, 2, 2136–2147, 2010.

Gerasopoulos, E., Andreae, M. O., Zerefos, C. S., Andreae, T. W., Balis, D., Formenti, P., Merlet, P., Amiridis, V., and Papastefanou, C.: Climatological aspects of aerosol optical properties in Northern Greece, Atmos. Chem. Phys., 3, 2025–2041, doi:10.5194/acp-3-2025-2003, 2003.

Gupta, P., Khan, M. N., da Silva, A., and Patadia, F.: MODIS aerosol optical depth observations over urban areas in Pakistan: quantity and quality of the data for air quality monitoring, Atmospheric Pollution Research, 4, 43–52, 2013.

Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, J. Geophys. Res.-Atmos., 102, 6831–6864, doi:10.1029/96jd03436, 1997.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – a federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, doi:10.1016/s0034-4257(98)00031-5, 1998.