Multiple regression method to determine aerosol optical depth in atmospheric column in Penang, Malaysia

F Tan, H S Lim, K Abdullah, T L Yoon, M Zubir Matjafri and B Holben

1 School of Physics, Universiti Sains Malaysia
2 NASA Goddard Space Flight Center, USA

E-mail: fuyitan@yahoo.com

Abstract. Aerosol optical depth (AOD) from AERONET data has a very fine resolution but air pollution index (API), visibility and relative humidity from the ground truth measurements are coarse. To obtain the local AOD in the atmosphere, the relationship between these three parameters was determined using multiple regression analysis. The data of southwest monsoon period (August to September, 2012) taken in Penang, Malaysia, was used to establish a quantitative relationship in which the AOD is modeled as a function of API, relative humidity, and visibility. The highest correlated model was used to predict AOD values during southwest monsoon period. When aerosol is not uniformly distributed in the atmosphere then the predicted AOD can be highly deviated from the measured values. Therefore these deviated data can be removed by comparing between the predicted AOD values and the actual AERONET data which help to investigate whether the non uniform source of the aerosol is from the ground surface or from higher altitude level. This model can accurately predict AOD if only the aerosol is uniformly distributed in the atmosphere. However, further study is needed to determine this model is suitable to use for AOD predicting not only in Penang, but also other state in Malaysia or even global.

1. Introduction

Atmospheric aerosol plays a significant role on the environment and climate changes, and atmospheric correction in the case of space-borne remote sensing such as satellite. Moderate-Resolution Imaging Spectro-radiometer (MODIS) sensor on board the Terra and Aqua satellites (by NASA) has been widely used to study aerosol properties such as aerosol optical depth [1-5]. AOD estimated from MODIS is required to validate for the higher accuracy as a value of AOD is retrieved at 10 km spatial resolution of MODIS and [6, 7] also supported that MODIS AOD need for improvement. Therefore, larger validation network of ground-based sun-photometers (e.g. AErosol RObotic NETwork (AERONET)) are needed to be assessed by comparing satellite based AOD. The details of the AERONET such as measurement concept, instrument precision, instrument calibration, data accuracy, data transmission, processing system, data processing and archival browser was listed by [8].

Atmospheric correction in Malaysia always is a challenging job as Malaysia is located in equator region where cloud may be found most of the time, therefore to obtain a free cloud satellite image is a tough task. By this reason, atmospheric correction is very important for all the atmospheric study in Malaysia. AOD is a useful parameter for atmospheric correction of remote sensing images as AOD value can ensure the retrieval of accurate surface reflectance [9-11]. Currently, Malaysia only have two sites deployed with AERONET, one in Kuching, Sarawak and another one in our study site (USM Penang) to provide AOD values for satellite AOD validation.

3 To whom any correspondence should be addressed.
In this paper, we present a technique to determine the relationship between AOD measured from AERONET station in Penang and other weather station parameters such as API visibility, relative humidity, pressure and temperature. Several algorithms were tested for predicting AOD, the algorithm that produced the highest correlation was selected as the model to retrieve AOD. In finally, we can predict for AOD as long as we have API and weather station data are available.

2. Study area
Penang island is located in the northwestern region of Peninsular Malaysia and lies within latitudes 5º 12’ N to 5º 30’ N and longitudes 100º 09’ E to 100º 26’ E (shown in Fig. 1). API station and sun photometer equipment are located in the Universiti Sains Malaysia, Penang, Malaysia (marked as five-edge star) and oval shape in the figure is a point to represent the weather station location. Fig. 1 shows the topography in Penang region with different elevation and surrounded by water.

3. Data
The data used in the study were acquired from following sources:

3.1 AERONET
Columnar aerosol properties as AOD provided from the AERONET by NASA in 7 wavelengths retrieved from different filters in sun photometer but the values of wavelength 500 nm was selected for this study. This is because 500 nm lies within the visible wavelength, red (660 nm), green (550 nm) and blue (480 nm) and a strong correlation is found between AOD at 500 nm and the aerosol radiative forcing (ARF) [12]. All the AERONET data are available online in near real time [13], and level 2.0 was used in this study because this level data had produced by cloud-screened procedure [14] and data assured by Smirnov. The AERONET site is called as USM_Penang which was under part of the project instrument of 7-SEAS and Southeast Asia Composition, Cloud, Climate Coupling Regional Study (SEAC^4RS).

3.2 API
API is mainly use to determine the air pollution level in Malaysia in term of sulphur dioxide (SO$_2$), nitrogen dioxide (NO$_2$), carbon monoxide (CO), ozone (O$_3$), and suspended particulate matters of less than ten microns in size (PM 10) computed by US-EPA technique base on these five pollutants [15]. API data can be obtained directly from http://www.doe.gov.my/apims/index.php. The API data collected by the USM Penang station was used in this study.

3.3 Weather station in Penang International Airport
Weather Underground having a vast repository of freely available weather data recorded formed by thousands of individual weather stations and all the data will be uploaded to the website (http://www.wunderground.com) near-real-time up-to-date weather conditions [16, 17]. Several parameters provided by the station which located at the Penang International Airport (coordinate
5.30’N and 100.26’E) were used in this study such as pressure (hPa), temperature (°C), relative humidity (%) and visibility (km).

4. Methodology
The dataset from the above mentioned sources were acquired during the southwest monsoon period (August to September, 2012). These 3 different sources acquired data at different time intervals, therefore the dataset collected within 30 minutes between them were selected for the study. All the dataset were converted to Julian day then the dataset collected within 30 minutes between them by MATLAB programming were selected (78 data points). During preliminary study stage, pressure and temperature were found to be poorly correlated with AOD. Therefore, only API, relative humidity (RH), and visibility (Vis) were considered for multi-regression analysis to predict AOD. Several equations (listed in Table 1) were tested to figure out their relationship. The root mean square (RMS) and coefficients for each model were calculated. The equation that produced the highest correlation was then assigned as a new model to predict AOD. Moreover, the coefficient values will be listed out for the best model in this study.

Table 1. Equations tested in multi-regression analysis to create a new model to predict AOD.

| Model | Equation |
|-------|----------|
| 1     | AOD=a_0+a_1(RH)+a_2(RH)^2+a_3(RH)^3 |
| 2     | AOD=a_0+a_1(Vis)+a_2(Vis)^2+a_3(Vis)^3 |
| 3     | AOD=a_0+a_1(API)+a_2(API)^2+a_3(API)^3 |
| 4     | 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 |
| 5     | AOD=a_0+a_1(RH)+a_2(RH)^2+a_3(Vis)+a_4(API)+a_5(API)^2+a_6(API)^3 |
| 6     | AOD=a_0+a_1(Vis)+a_2(Vis)^2+a_3(Vis)^3+a_4(API)+a_5(API)^2+a_6(API)^3 |
| 7     | AOD=a_0+a_1(RH)+a_2(Vis)+a_3(RH*Vis)+a_4(Vis)^2+a_5(RH*Vis)+a_6(RH*Vis^2) |
| 8     | AOD=a_0+a_1(RH)+a_2(API)+a_3(RH*API)+a_4(Vis)+a_5(API)^2+a_6(RH^2*API)+a_7(RH*API^2) |
| 9     | AOD=a_0+a_1(Vis)+a_2(API)+a_3(Vis*API)+a_4(Vis)^2+a_5(API)^2+a_6(Vis^2*API)+a_7(Vis*API^2) |
| 10    | AOD=a_0+a_1(RH)+a_2(API)+a_3(Vis)+a_4(RH*Vis)+a_5(RH*API)+a_6(Vis^2)+a_7(Vis^3)+a_8(API)+a_9(API)^2+a_10(API)^3 |
| 11    | AOD=a_0+a_1(API)+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 |

4. Results and discussions
Eleven models were tested by the regression analysis, 1st to 3rd models from Table 1 was calculated by 3rd order polynomial regression analysis with one single parameter and 4th to 9th with two parameters and 10th to 11th with three parameters. Obviously, the multi-parameters have produced better results than the single parameter. During single parameter regression analysis, API showed the highest correlation with AOD and followed by Vis then only RH. During single parameter regression analysis, API showed the highest correlation with AOD as indicated in Table 1. The used of API in the 3rd order polynomial regression analysis with single parameter and 4th order polynomial regression analysis with two parameters showed higher correlation with AOD as indicated in Table 1. Therefore, the following model which included API parameter showed higher correlation with AOD as indicated in 5th, 6th and 7th models from Table 1. Even though the low correlation for models containing RH parameter was found, when all these three parameters were combined in the regression analysis the best model may be produced (10th and 11th model) compared to other models which only have one or two parameters. The 11th model was chosen as the best model created from the multi-regression analysis because it not only simpler compared to the 10th model but also have R^2 as high as 0.9595 and RMS as low as 0.1021 during predicting AOD value.
Table 2. Results of the model to predict AOD value including RMS, $R$, $R^2$.

| Model       | Parameter Involved | RMS       | $R^2$       |
|-------------|--------------------|-----------|-------------|
| 1           | RH                 | 1.066019  | 0.088167    |
| 2           | Vis                | 0.290544  | 0.672690    |
| 3           | API                | 0.205822  | 0.835744    |
| 4           | RH+Vis             | 0.284783  | 0.685545    |
| 5           | RH+API             | 0.211670  | 0.826488    |
| 6           | Vis+API            | 0.122372  | 0.941938    |
| 7           | RH+Vis             | 0.275984  | 0.704671    |
| 8           | RH+API             | 0.194236  | 0.853716    |
| 9           | Vis+API            | 0.143130  | 0.920568    |
| 10          | API+RH+Vis         | 0.121431  | 0.942827    |
| 11          | API+RH+Vis         | 0.102161  | 0.959535    |

The predicted AOD versus measured AOD was plotted using 11th model (Fig. 2). In this figure, most of the predicted data are very highly correlated to the measured data. The 11th Model created from these three ground measurements indicated that the aerosol was uniformly distributed in the atmospheric column during these period because the predicted AOD values were mostly found with standard deviation less than 0.1 and only few outliers were found at about 0.6 and 0.9 (x-axis in Fig. 2) with standard deviation higher than 0.1. Therefore, we assume this model was created under consideration of aerosol was uniformly distributed in the atmosphere column. Aerosol may uniformly distributed in the atmosphere was supported by several studies [18, 19].

To validate the reliability of this 11th model, the formula and coefficients generated was used with another data set (118 data points) for June to September of 2012 as shown in Fig. 3. The predicted AOD was obtained with $R^2 = 0.8201$ as there were 7 outliers (refer to red box in Fig. 3). The decrease of $R^2$ value from 0.9595 to 0.8201 when the model applied in other dataset was believed due to environment factors which may affected the AOD result in the atmosphere, therefore measured and predicted AOD was plotted as a function of time within the southwest monsoon period. From Fig. 4, predicted and measured AOD show a similar trend most of the time, but sometimes there were over- or under-predicted owing to non-uniformly loaded aerosol in the atmosphere at different altitude level. Therefore the 7 outliers in between 160 and 173 in Julian days (18 to 21 June of 2012) were removed from the original data to reach $R^2$ as high as 0.9337 (refer to Fig. 5). The result shows that this model
was successfully predict AOD values in Penang, Malaysia. However, uniformly distributed aerosol is a basic requirement to obtain highly reliable predicted AOD value.

5. Conclusion
This study showed that the ground measurement may able to predict the AOD in the atmosphere using the proposed empirical relationship. Therefore, this model can provide AOD value for the locations where the equipments for measuring AOD such as AERONET are not available because this model was successfully validated the predicted AOD in Penang with very high $R^2$. However, to gain the concrete support to above statement more studies are needed in future as the AERONET data still new in Penang, Malaysia since the AERONET was deployed on November 2011.

Acknowledgement
The authors gratefully acknowledge the financial support from RU grants 1001/PFIZIK/811228 and Universiti Sains Malaysia – Short term grant 304/PFIZIK/6310057 used to carry out this project.

References
[1] S. Liang, B. Zhong, and H. Fang, "Improved estimation of aerosol optical depth from MODIS imagery over land surfaces," vol. 104, ed, 2006, pp. 416-425.
[2] A. D. d. A. Castanho, R. Prinn, V. Martins, M. Herold, C. Ichoku, and L. T. Molina, "Analysis of Visible/SWIR surface reflectance ratios for aerosol retrievals from satellite in Mexico City urban area," Atmos. Chem. Phys., vol. 7, pp. 5467-5477, 2007.
[3] A. Retalis, D. G. Hadjimitsis, S. Michaelides, F. Tymvios, N. Chrysoulakis, C. R. I. Clayton, and K. Themistocleous, "Comparison of aerosol optical thickness with in situ visibility data over Cyprus," Natural Hazards and Earth System Science, vol. 10, pp. 421-428, 2010.
[4] Y. S. Bennouna, V. E. Cachorro, B. Torres, C. Toledano, A. Berjón, A. M. de Frutos, and I. Alonso Fernández Coppel, "Atmospheric turbidity determined by the annual cycle of the aerosol optical depth over north-center Spain from ground (AERONET) and satellite (MODIS)," Atmospheric Environment, vol. 67, pp. 352-364, 2013.
[5] H. S. Kim, Y. S. Chung, and S. G. Lee, "Analysis of spatial and seasonal distributions of MODIS aerosol optical properties and ground-based measurements of mass concentrations in the Yellow Sea region in 2009," Environmental Monitoring and Assessment, vol. 185, pp. 369-382, 2013.

[6] R. C. Levy, L. A. Remer, J. V. Martins, Y. J. Kaufman, A. Plana-Fattori, J. Redemann, and B. Wenny, "Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS," Journal of the Atmospheric Sciences, vol. 62, pp. 974-992, 2005.

[7] S. N. Tripathi, S. Dey, A. Chandel, S. Srivastava, R. P. Singh, and B. N. Holben, "Comparison of MODIS and AERONET derived aerosol optical depth over the Ganga Basin, India," Annales Geophysicae, vol. 23, pp. 1093-1101, 2005.

[8] B. N. Holben, T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, "AERONET - A federated instrument network and data archive for aerosol characterization," Remote Sensing of Environment, vol. 66, pp. 1-16, 1998.

[9] G. Fedosejevs, N. T. O'Neill, A. Royer, P. M. Teillet, A. I. Bokoye, and B. McArthur, "Aerosol optical depth for atmospheric correction of AVHRR composite data," Canadian Journal of Remote Sensing, vol. 26, pp. 273-284, 2000.

[10] T. Hilker, A. I. Lyapustin, C. J. Tucker, P. J. Sellers, F. G. Hall, and Y. Wang, "Remote sensing of tropical ecosystems: Atmospheric correction and cloud masking matter," Remote Sensing of Environment, vol. 127, pp. 370-384, 2012.

[11] S. C. McCarthy, R. W. Gould, J. Richman, C. Kearney, and A. Lawson, "Impact of aerosol model selection on water-leaving radiance retrievals from satellite ocean color imagery," Remote Sensing, vol. 4, pp. 3638-3665, 2012.

[12] A. Prasad, S. Singh, S. Chauhan, M. Srivastava, R. Singh, and R. Singh, "Aerosol radiative forcing over the Indo-Gangetic plains during major dust storms," Atmospheric Environment, vol. 41, pp. 6289-6301, 2007.

[13] NASA. (2012). AERONET. Available: http://aeronet.gsfc.nasa.gov/new_web/index.html

[14] A. Smirnov, B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker, "Cloud-Screening and Quality Control Algorithms for the AERONET Database," Remote Sensing of Environment, vol. 73, pp. 337-349, 2000.

[15] DOE, "A Guide to Air Pollutant Index in Malaysia (API)," D. O. Environment, Ed., 3rd ed. Kuala Lumpur, Malaysia, 1997, pp. 1-20.

[16] M. Williams, D. Cornford, L. Bastin, R. Jones, and S. Parker, "Automatic processing, quality assurance and serving of real-time weather data," Computers and Geosciences, vol. 37, pp. 353-362, 29 Dec 2012 2011.

[17] T. Bedrina, A. Parodi, A. Quarati, and A. Clematis, "ICT approaches to integrating institutional and non-institutional data services for better understanding of hydro-meteorological phenomena," Natural Hazards and Earth System Science, vol. 12, pp. 1961-1968, 2012.

[18] J. S. Reid, S. J. Piketh, A. L. Walker, R. P. Burger, K. E. Ross, D. L. Westphal, R. T. Bruintjes, B. N. Holben, C. Hsu, T. L. Jensen, R. A. Kahn, A. P. Kuciauskas, A. A. Mandoos, A. A. Mangoosh, S. D. Miller, J. N. Porter, E. A. Reid, and S. C. Tsay, "An overview of UAE2 flight operations: Observations of summertime atmospheric thermodynamic and aerosol profiles of the southern Arabian Gulf," Journal of Geophysical Research D: Atmospheres, vol. 113, 2008.

[19] R. D. Cadle and G. W. Grams, "Stratospheric Aerosol Particles and Their Optical Properties," Rev Geophys Space Phys, vol. 13, pp. 475-501, 1975.