Retrieval of Aerosol Optical Thickness from Satellite Images through Machine Learning

Jun Xiang CHEN[1], Soo Chin LIEW[1], and Santo SALINAS[1]
[1]Centre for Remote Imaging, Sensing and Processing (CRISP), National University of Singapore,10 Lower Kent Ridge Road, Blk S17, Level 2, Singapore 119076, Email: crscjx@nus.edu.sg, scliew@nus.edu.sg, crsscsv@nus.edu.sg

KEY WORDS: Aerosol, Aerosol Optical Thickness, Simulated Satellite Signals, Machine Learning and Neural Network

ABSTRACT: Aerosol measurements are routinely obtained through ground observation networks of sun-photometers such as the AERONET. Such implementation gives accurate and periodic observations of aerosol properties at each ground-based station. Satellite observations can provide wide areal coverage of aerosol optical thickness (AOT). However, the retrieval of AOT from satellite data usually involves execution of complex radiative transfer codes and is complicated by the influence of different ground covers. In this study, the retreival of AOT at 550 nm (AOT550) from satellite images using machine learning is presented. Simulated satellite signals were generated using the Py6S, a python interface to the 6S (Second Simulation of a Satellite Signal in the Solar Spectrum) Radiative Transfer Model (RTM). The simulated data were based on the different permutations of varying geometrical conditions, atmospheric models, spectral conditions and ground reflectances. A feedforward neural network (NN) with backpropagation algorithm was implemented. The dataset was used for training, validation and testing of the NN model. The trained NN model was then applied to retrieve the value of AOT550 from the satellite images. The plot of actual against retrieved AOT550 had achieved a coefficient of determination (R^2) of 0.716. The machine learning method has shown a promising capability for the spatial and temporal monitoring of aerosols distribution, especially during episodes of biomass burning.

1. INTRODUCTION

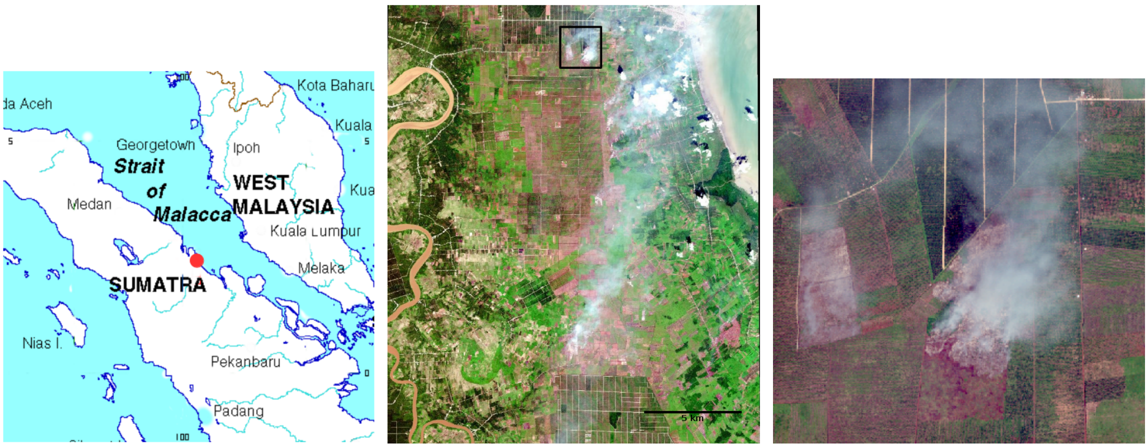
Remote sensing from satellite and ground observation networks should be used conjointly to monitor aerosols. (Kaufman et al., 1997) Aerosol observations obtained routinely through ground observation networks of sun-photometers, such as the AEROsol robotic NETwork (AERONET), give accurate and periodic measurements of the physical and optical characteristics of the in-situ vertical column of aerosols. Conversly, satellite observations provide wide areal coverage and could provide spatial information on the distribution of the aerosols. For instance, the Moderate Resolution Spectroradiometer (MODIS) payload on the Terra and Aqua satellites has an extensive spectral range that could be used to derive the Aerosol Optical Thickness (AOT) and aerosols size properties. (Remer et al., 2005) AOT measures the attenuation of the solar radation due to the atmospheric effects. (Radosavljevic et al., 2007) Therefore, monitoring of aerosols should include both satellite and ground-based measurements.

Studies have shown that the MODIS Aerosol Algorithm can be used to derive AOT over land. The image is first screened to mask out water, cloud, snow and ice pixels. This is followed by the estimation of the surface and top-of-atmosphere (TOA) reflectances. The estimated AOT is then retrieved using a continental model Lookup Table (LUT). With the derived AOT, the path radiances are calculated. The aerosol model is then determined using the path radiances. Lastly, the finalized AOT is retrieved from the aerosol's model LUT. (Remer et al., 2005) Another study showed the use of data-driven method for the improvement of MODIS aerosols algorithm. (Han et al., 2006) It demonstrated the use of Neural Network (NN) to retrieve the AERONET data based on attributes from the MODIS data. It showed that the accuracy of the AOT retrieval from NN method (correlation coefficient, r = 0.662) was slightly better than the MODIS algorithm (r = 0.640). It had also implementated the Decision Tree method to analyze the MODIS aerosol retrieval algorithm. (Han et al., 2006) Futhermore, one of the studies proposed the use of a NN ensemble with adaptive cost function on data-driven method. (Radosavljevic et al., 2007) It had shown a retrieval accuracy of $r^2 = 0.84$. Besides, the retrieval of turbidity through TOA reflectance from satellite images had been explored. For example, such method had been performed on the Saharan Dust Outbreak. (Carlson, 1979) The retrieval was performed using the visible band of the NOAA 3 Very High Resolution Radiometer (VHRR). Screening the satellite image to identify pixels without contamination by clouds was first performed. The AOT was then retrieved using the LUT method. As stated in the study, retrieval using such a method might be possible only for a generally low AOT environment. As the ambient AOT increases, the challenge to discriminate cloud from AOT is evident.

In this paper, a novel approach using simulated satellite signal data along with machine learning (ML) method to retrieve AOT is demonstrated. Simulated data were first generated, followed by the use of ML. With the optimized parameters, the trained NN model was then used to convert the TOA reflectance from satellite image to the value of AOT at 550 nm (AOT550). Noteably, this retrieval is in pixel-level. Such pixel-level approach provides spatial information on the distribution of the aerosols of satellite images. Previous studies on data-driven approach were done using aggregated data of the satellite and ground-based station measurements which led to the retrieval over areas in the absence of ground-based stations being infeasible. This is not the case for this method as simulated satellite signals are used instead. Another disadvantage of using collocated data is the limited number of it due to the need for the satellite pass to coincide with the ground-based station measurement. The amount of data for higher AOT is also inadequate since the occurence of it is infrequent in most areas. The use of skewed data dominated by lower AOT values might result in some form of bias in retrieval. This problem could be overcome through this method which uses of simulated satellite data. Moreover, this method does not have pre-requisite for retrievals. MODIS Aerosol Algorithm requires at least 12 dark target pixels before the retrieval is deemed reliable. If this is not fulfiled, the selection criteria could be adjusted but the retrieval would classified as 'poor quality'. If the number of these pixels are not fulfiled with the relaxed selection criteria, no retrievals will be made for the entire image.

2. METHOD

The simulated satellite signal was first generated. The tabulated data were then fed into the ML. With the parameters generated by the ML, it was then used to retrieve the value of AOT550 from the satellite images. The satellite image was preprocessed from digital number to TOA reflectance (band 1-8) before the retrieval. The WV2 image used was over Sumatra, Indonesia, which is located south of the Malay Peninsula and west of Java. Figure 1a shows the location of the satellite image represented by a red dot. The satelite image is located at 2°21'49"N and 100°17'12"E (centre-coordinate) and captured on 13th May 2013 at 04:08:01 UTC. The full satellite image is shown in figure 1b. The AOT550 retrieval is performed over the cropped area of the full image and it captured a scene of biomass burning (Figure 1c) .



(a) (b) (c)

Figure 1: (a) Location of the satellite image (red dot), (b) Full satellite image and (c) cropped image.

Generation of Simulated Satellite Signal Data

Satellite signals were simulated using Py6S. Py6S is a Python interface to the 6S (Second Simulation of a Satellite Signal in the Solar Spectrum) Radiative Transfer Model (RTM). (Wilson, 2013) Solar radiance reflected from the surface is modified by atmospheric effects, absorption by the gases and scattering by aerosols and the molecules. (Vermote et al., 1997) The radiative transfer equation for a target with reflectance ρ_s, assumed to be a Lambertian surface and at sea level illuminated by solar radiation and viewed from satellite level is as follows:

$$\rho_{TOA}(\theta_s,\theta_v,\phi_s-\phi_v) = T_g(\theta_s,\theta_v)\left[\rho_{R+A} + T^{\uparrow}(\theta_s)T^{\uparrow}(\theta_v)\frac{\rho_s}{1-S\rho_s}\right] \tag{1}$$

where:

ρ_s is the reflectance of the target at sea level,

θ_v is the viewing zenith angle of the satellite,

ϕ_v is the viewing azimuth angle of the satellite,

θ_s is the zenith angle of the Sun,

ϕ_s is the azimuth angle of the Sun,

ρ_{R+A} is the intrinsic reflectance of the layer containing both molecule and aerosol,

$T_{(\theta_{s/v})}$ is the overall transmission of the atmosphere between the Sun/sensor and surface,

S is the spherical albedo of the atmosphere and

T_g is the gaseous transmission. (Vermote et al., 1997)

11742 simulated satellite signals data were generated. These data were based on a tropical atmospheric profile and generated by randomizing the different types of aerosol models. These aerosol models include Continental, Maritime, Urban and Biomass Burning. The values of AOT550 generated were based on the lognormal distribution of the AERONET data of Singapore from 2006 to 2016. The ground reflectance was obtained from the satellite image. Spectral profiles of vegetation and soil from the satellite image were collected. Pixel mixing of the vegetation and soil profiles to generate variations in ground reflectances were then performed. The tabulated dataset contained AOT550, solar zenith angle, solar azimuth angle, sensor zenith angle, sensor azimuth angle and TOA reflectance (band 1 - 8).

Feedforward Neural Network with Backpropagation

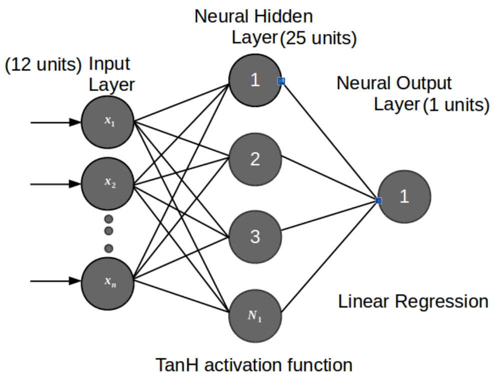


Figure 2: This illustrates the structure of the NN used in the retrieval of AOT550. It has an input layer, a hidden layer and an output layer with 12, 25 and 1 node respectively. A tanH function has been used in the hidden layer whereas a linear regression has been used to ouptut the AOT550.

NN was chosen as the ML method to retrieve AOT550. The NN consisted an input layer, a hidden layer and an output layer of 12, 25 and 1 unit respectively. TanH activation function was applied to the hidden layer and linear function for the output layer as shown in the figure 2. Least squares method with regularization terms was used as the cost function. The backpropagation algorithm was implemented and the Conjugate Gradient Method was used to optimize the parameters in the NN. The dataset was imported into the NN. The order of the data was first randomized. The data were then normalized by subtracting the mean and dividing by the standard deviation for each input feature. The data were then categorized into training, validation and test sets of 60%, 20% and 20% of the total number of data respectively. The training set was fed into the NN and trained with a range of regularization parameters. The trained model for each value of regularizer was the used to retrieve the AOT550 based on the features (angles and the TOA reflectances) from the validation set. The cost was computed between the actual and retrieved AOT550 from the validation set. The value of regularizer that resulted in the lowest cost was then selected. The NN then underwent another round of training with the chosen value of regularizer. With the trained model, it was used to retrieve the AOT550 based on the features from the test set. The correlation between the actual and retrieved AOT550 was then calculated. With the trained model, it was used to retrieve the AOT550 for the satellite image.

3. RESULTS AND DISCUSSION

Simulated Satellite Signal

Figure 3 illustrates some of the TOA reflectance spectra generated using the Py6S. Comparisions of the spectra under different conditions were plotted on the same graph. Spectra of 8 different conditions were plotted. They were high and low AOT550 at high and low Surface Brightness (SB) and Normalized Difference Vegetation Index (NDVI). SB was calculated by averaging band 2, 3 and 5 of the ground reflectances. Characteristics observed include an overall much gentler slope for spectra with lower NDVI as compared to the higher value of it, while other parameters being kept approximately constant. (Refer to figure 3c and 3d) The effect of the value of AOT550 on the spectra was also evident as showed in figure 3. A higher AOT550 in both high and low SB general caused an upward shift in the lower range of the spectra as observe in figure 3a and 3b. In addition, one could observed the spectra difference in characteristics among soil, vegetation and high SB target pixels.

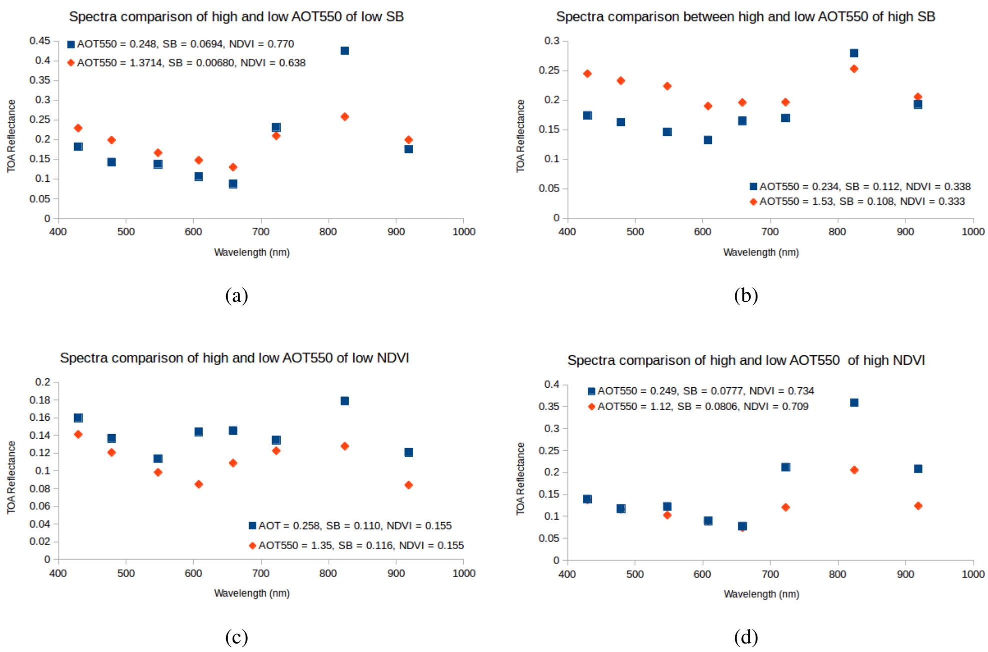


Figure 3: TOA reflectance spectra generated under different conditions by Py6S. It shows the spectra comparison between (a) high and low AOT550 of low SB, (b) high and low AOT550 of high SB, (c) high and low AOT550 of low NDVI, (c) high and low AOT550 of high NDVI and (d) high and low SB of low NDVI

AOT550 Retrieval

Plot of the actual against retrieved AOT550 is shown in figure 4. The coefficient of determination (R-squared) for the plot is 0.716 for AOT550. This shows good correlation between the actual and retrieved data.

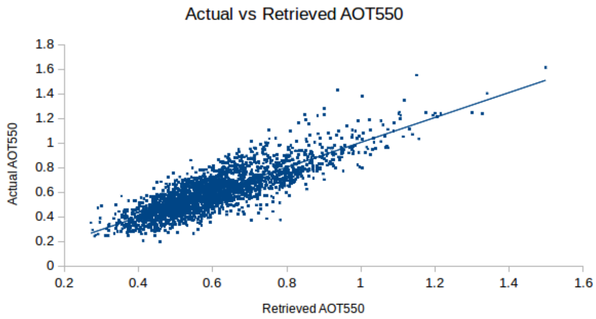
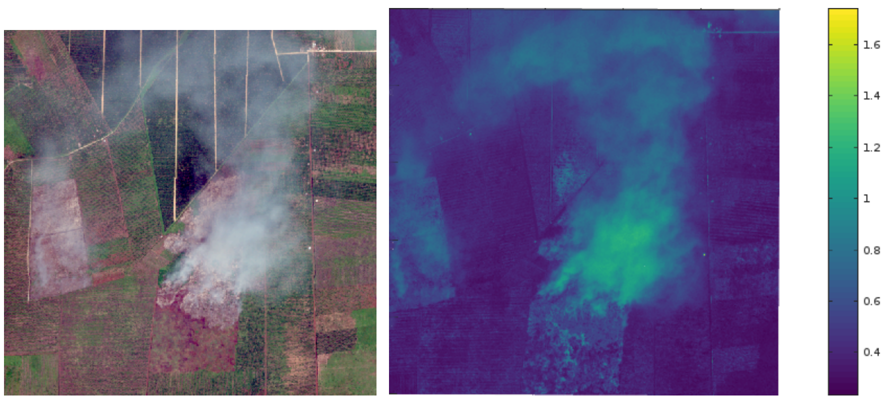


Figure 4: This illustrates the plot of the actual against retrieved AOT550. The coefficient of determination (R-squared) is of 0.716.

Figure 5b illustrates a map of AOT550 retrieved from figure 5a. The map was derived from retrievals of every pixels in the satellite image. Aside from allowing retrieval of AOT550 for specific pixel or location, it showed capability in mapping of the smokes plume from the biomass burning. This clear mapping of the smoke plume provides knowledge on the distribution of the aerosols in the satellite image. Having said that, some of the ground features could be observed from the map. This is not desirable as these ground features have affected the retrieval of the AOT550 for those pixels. This could be due to training data being limited to only mixed pixels of soil and vegetation of varying ratios. Despite shortcomings, this method does provides a tool to map the distribution of the AOT550.

Furthermore, the extraction of the ground reflectance from the satellite image was challenging. The spectra were chosen in area where smoke appeared absent. The extraction of ground reflectances were based on a 'kink' in the spectrum to distinguish the already mixed pixel from single-component dominated ones. One of the precautions taken was to extract as much of the ground reflectance spectra as possible to average out the possible error caused by such subjective selection process. Futhermore, the AOT550 for the dataset was generated using a lognormal distribution assumption. This distribution was derived from the AERONET data for Singapore from 2006 to 2016. This might have put a ceiling on the magnitude of AOT550 retrieval. The highest recorded AOT550 in the dataset recorded was 1.87 and the number of data with AOT550 above 1.50 was limited. The use of a different distribution could be explored. Moreover, the magnitude of the AOT550 in this image could not be cross-validated in the absence of ground-based measurement.



(a) (b)

Figure 5: (a) The WorldView-2 satellite image over Sumatra, Indonesia and (b) mapping of the retrieved AOT550.

4. CONCLUSION

The use of simulated satellite signals along with NN method had shown promising capability for the retrieval of AOT550. A coefficient of determination (R-squared) of 0.716 between the actual and retrieved AOT550 has been obtained. Some of the key advantages of this method includes providing pixel-level retrieval and allowing for spatial mapping of the AOT550 distribution. The latter could be observed from the mapping of the smoke plume of the biomass burning from the image over Sumatra, Indonesia. This technique helps to overcome many challenges faced by previous studies. This includes the limited number of data available aggregated from satellite and ground-based station measurements for data-driven retrieval approach. It also allows for retrieval outside the vicinity of the ground-based stations which is one of the constraints of using collocated data. Having said that, measurements from ground-based station are needed to validate the retrieved AOT550. The influence on retrievals by heavy smoke plumes and ground features could also be observed from the map indicating room for improvement for this method, such as having a combined dataset containing more variations of ground covers. A more objective way of extracting ground reflectance from satellite images and other ML methods could also be explored.

REFERENCES

Carlson, T. N., 1979, Atmospheric turbidity in saharan dust outbreaks as determined by analyses of satellite brightness data. Monthly Weather Review, 107(3):pp. 322–335.

Han, B., Obradovic, Z., Li, Z. and Vucetic, S., 2006, Data mining support for the improvement of modis aerosol retrievals. In: Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on, IEEE, pp. 2453–2456.

Kaufman, Y., Tanré, D., Remer, L. A., Vermote, E., Chu, A. and Holben, B., 1997, Operational remote sensing of tropospheric aerosol over land from eos moderate resolution imaging spectroradiometer. Journal of Geophysical Research: Atmospheres, 102(D14):pp. 17051–17067.

Radosavljevic, V., Vucetic, S. and Obradovic, Z., 2007, Aerosol optical depth retrieval by neural networks ensemble with adaptive cost function. In: Proceedings of the 10th International Conference on Engineering Applications of Neural Networks, pp. 266–275.

Remer, L. A., Kaufman, Y., Tanré, D., Mattoo, S., Chu, D., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R., Kleidman, R. et al., 2005, The modis aerosol algorithm, products, and validation. Journal of the atmospheric sciences, 62(4):pp. 947–973.

Vermote, E. F., Tanré, D., Deuze, J. L., Herman, M. and Morcette, J.-J., 1997, Second simulation of the satellite signal in the solar spectrum, 6s: An overview. IEEE transactions on geoscience and remote sensing, 35(3):pp. 675–686.

Wilson, R. T., 2013, Py6s: A python interface to the 6s radiative transfer model. Computers & Geosciences, 51(2):p. 166.