Particulate matter concentration mapping from MODIS satellite data: a Vietnamese case study

You may also be interested in:

Vegetation fires, absorbing aerosols and smoke plume characteristics in diverse biomass burning regions of Asia
Krishna Prasad Vadrevu, Kristofer Lasko, Louis Giglio et al.

Using satellite data to develop environmental indicators
Alex de Sherbinin, Marc A Levy, Erica Zell et al.

Impact of the June 2013 Riau province Sumatera smoke haze event on regional air pollution
Sheila Dewi Ayu Kusumaningtyas and Edvin Aldrian

Mixing weight determination for retrieving optical properties of polluted dust with MODIS and AERONET data
Kuo-En Chang, Ta-Chih Hsiao, N Christina Hsu et al.

Synoptic and meteorological drivers of extreme ozone concentrations over Europe
N Otero, J Sillmann, J L Schnell et al.

Source attribution of particulate matter pollution over North China with the adjoint method
Lin Zhang, Licheng Liu, Yuanhong Zhao et al.

The UK particulate matter air pollution episode of March–April 2014: more than Saharan dust
M Vieno, M R Heal, M M Twigg et al.

Space-based measurements of air quality during the World Expo 2010 in Shanghai
N Hao, P Valks, D Loyola et al.

Environmental Research Letters

LETTER

Particulate matter concentration mapping from MODIS satellite data: a Vietnamese case study

Thanh T N Nguyen, Hung Q Bui, Ha V Pham, Hung V Luu, Chuc D Man, Hai N Pham, Ha T Le and Thuy T Nguyen

University of Engineering and Technology (UET), Vietnam National University Ha Noi (VNU). Building E3, 144 Xuan Thuy Str., Cau Giay Dist., Ha Noi, Vietnam

E-mail: thanhntn@vnu.edu.vn

Abstract

Particulate Matter (PM) pollution is one of the most important air quality concerns in Vietnam. In this study, we integrate ground-based measurements, meteorological and satellite data to map temporal PM concentrations at a 10×10 km grid for the entire of Vietnam. We specifically used MODIS Aqua and Terra data and developed statistically-significant regression models to map and extend the ground-based PM concentrations. We validated our models over diverse geographic provinces i.e., North East, Red River Delta, North Central Coast and South Central Coast in Vietnam. Validation suggested good results for satellite-derived $PM_{2.5}$ data compared to ground-based $PM_{2.5}$ ($n = 285, r^2 = 0.411, RMSE = 20.299 \, \mu g \, m^{-3}$ and $RE = 39.789\%$). Further, validation of satellite-derived $PM_{2.5}$ on two independent datasets for North East and South Central Coast suggested similar results ($n = 40, r^2 = 0.455, RMSE = 21.512 \, \mu g \, m^{-3}, RE = 45.236\%$ and $n = 45, r^2 = 0.444, RMSE = 8.551 \, \mu g \, m^{-3}, RE = 46.446\%$ respectively). Also, our satellite-derived $PM_{2.5}$ maps were able to replicate seasonal and spatial trends of ground-based measurements in four different regions. Our results highlight the potential use of MODIS datasets for PM estimation at a regional scale in Vietnam. However, model limitation in capturing maximal or minimal $PM_{2.5}$ peaks needs further investigations on ground data, atmospheric conditions and physical aspects.

1. Introduction

Asia is fast developing due to industrialization and urbanization. Together with the development, air pollution concerns on human health are also increasing (MacNee and Donaldson 2003, Atkinson *et al* 2012, Krzyzanowski *et al* 2014). $PM_{2.5-10}$ consists mainly of crustal particles mechanically generated from agriculture, mining, construction, road traffic, and related sources, as well as particles of biological origin. Particulate matter with diameter less than 2.5 μm ($PM_{2.5}$) consists mainly of combustion particles from motor vehicles and the burning of coal, fuel oil, and wood, but also contains some crustal particles from finely pulverized road dust and soils (Laden *et al* 2000). In several cities of Asia including Vietnam, the pollution levels exceed the World Health Organization (2006) Guideline values. In particular, exposure to $PM_{2.5}$ may lead to human respiratory diseases and even mortality (Pope *et al* 2009, Fann *et al* 2012, Kloog *et al* 2013).

Specific to Vietnam, a recent report from the Environmental Performance Index (EPI) suggests that the quality of the environment in Vietnam has steadily dropped compared to other nations (EPI 2014). As per the EPI report, in the general environmental index, Vietnam is ranked 136 out of 178. Further, air quality in Vietnam is lagging with a rank of 170 and it is forecasted to worsen in the near future (EPI 2014). Of the different pollutants in Vietnam, particulate matter pollution is serious because of rapid growth of industrial activities, traffic operations and forest/agricultural fires (Le *et al* 2014). Also, a recent National Environmental Report on air environment (Vietnam Environment Adminstration 2013) for Vietnam suggests that dust pollution based on total suspended particle (TSP), PM_{10}, $PM_{2.5}$ and $PM_{1.0}$ surpassed national standards in several cities with impacts on human health. Thus, continuous mapping and monitoring of PM in different regions of Vietnam gain significance.

PM$_{2.5}$ measurements through ground based stations although measure the concentrations with high level of accuracy and frequency, they are of limited geographic coverage in Vietnam. Also, point measurements from ground based monitoring stations are not necessarily representative of regional concentrations and regional variability is difficult to assess from point measurements alone (van Donkelaar *et al* 2010). Recent studies suggest the potential use of satellite remote sensing technologies for mapping and monitoring of pollutants and relating them to ground sources (Badarinath *et al* 2009, Monks and Bierelle 2011, Kharol *et al* 2012, Vadrevu *et al* 2014). Specific to PM, the use of satellite instruments to estimate surface PM concentration is considered as an effective way to extend point based measurements to wide spatial scales (Duncan *et al* 2014). Satellite-derived PM maps are gradually becoming basic layers for air quality, human health and disaster management (Anderson *et al* 2012, van Donkelaar *et al* 2011).

Satellite aerosol products, which represent spatial distribution of particles in the vertical direction of the atmosphere, provide a practical solution for PM estimation. Several researchers used aerosol optical depth (AOD) or aerosol optical thickness (AOT) derived from satellite data for PM estimation (Chu *et al* 2003, Wang and Christopher 2003, Engel-Cox *et al* 2004, Badarinath *et al* 2007, Gupta and Christopher 2008, Lee *et al* 2011, Pelletier *et al* 2007, Schaap *et al* 2009, Hirtl *et al* 2014, Zha *et al* 2010, Ma *et al* 2014). Methods for estimating PM vary from linear regression (LR) or multiple linear regression (MLR) to the non-linear regression methods such as artificial neural network (ANN), support vector regression (SVR) and self organizing map (SOM) (Gupta and Christopher 2009a, 2009b, Yahi *et al* 2011, Hirtl *et al* 2014). Recently, modelling systems such as GEOS-Chem or CMAQ are also used for relating AOT to PM (Liu *et al* 2007a, Liu *et al* 2007b, Liu *et al* 2007c).

In Vietnam, most of the earlier studies on PM characterization focused on ground measurements (Hien *et al* 2002). Within the framework of the Asian regional air pollution research network (AIRPET), PM$_{2.5}$ and PM$_{10}$ were studied by Kim Oanh *et al* (2006). Recently the effect of regional meteorology on mass and composition of PM was investigated in Ha Noi during December 2006–February 2007 (Hai and Kim Oanh 2013) and in a mining town in Quang Ninh province (Northern Vietnam) in both dry and wet seasons from 2009 to 2010 (Hang and Kim Oanh 2014). Compared to these studies, relatively few studies focused on using satellite remote sensing data for estimating pollution. For example, air pollution maps at high spatial resolution were estimated directly from SPOT 2, Landsat 5, Landsat 8 data for Quang Ninh and Ha Noi (Luong *et al* 2010, Nguyen and Tran 2014). Earlier, our team carried out initial investigations for PM estimation using of satellite aerosol products in Ha

Noi and obtained promising results (Nguyen *et al* 2014, Le *et al* 2014).

In this study, we use satellite remote sensing for estimating PM for entire Vietnam. We mapped temporal PM$_{2.5}$ concentrations by integrating MODIS satellite data and ground based PM data from December 2010 to September 2014. We also validated our results for four regions.
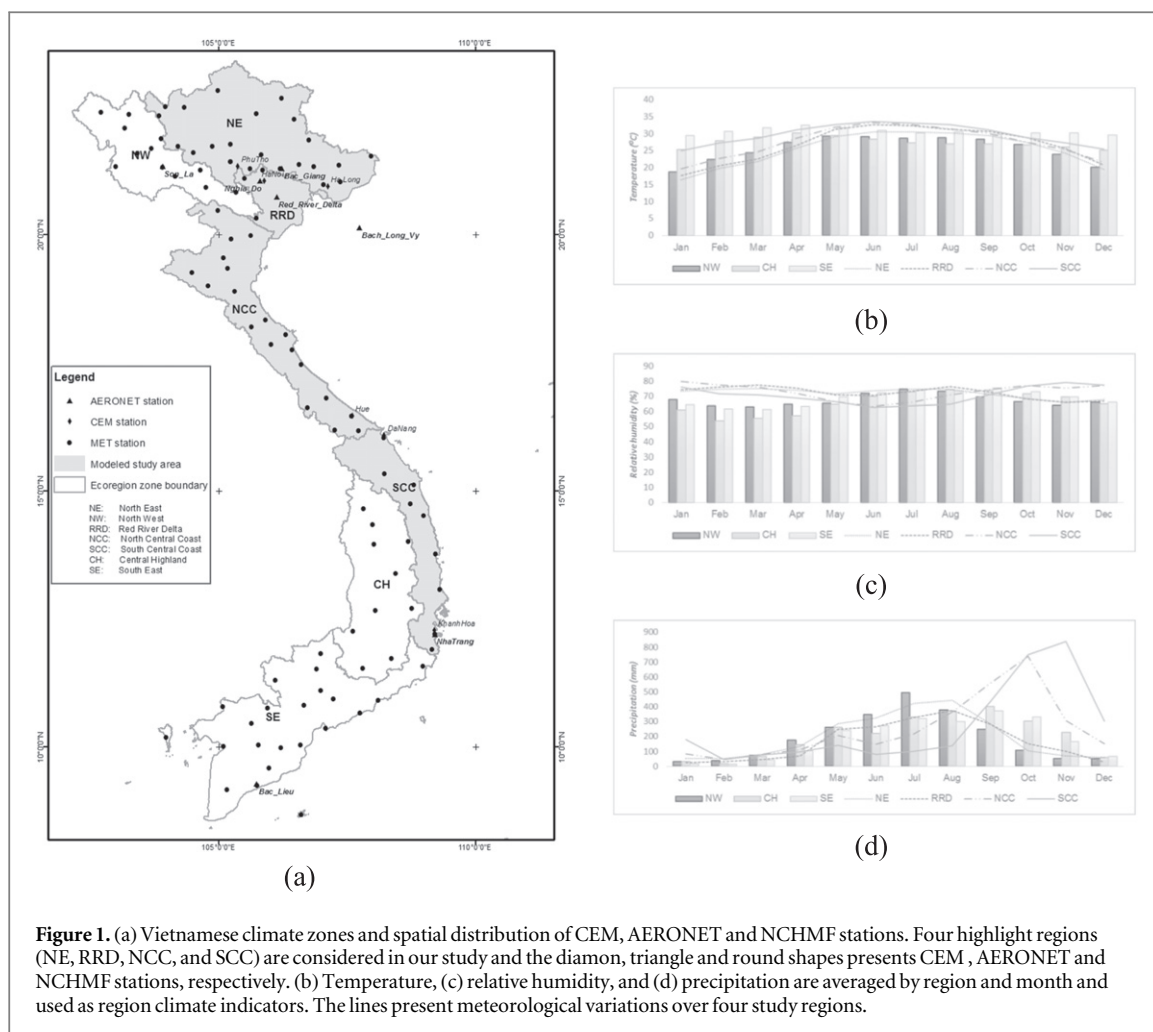
2. Study area

Vietnam is the easternmost country on the Indochina Peninsula in Southeast Asia. The total area is nearly 332 210 km^2 and extends from (8°27′ N, 102°8′ E) to (23° 23′ N, 109°27′ E) with the population of 90.5 million as of 2014. Vietnam is divided into seven different climatic zones which are based on summer and winter over northern regions (i.e.: North West—NW, North East—NE, Red River Delta—RRD, North Central Coast—NCC) and rainy and dry seasons to the South (South Central Coast—SCC, Central Highlands—CH and South East—SE). In this study, we applied our model and validated in four different regions i.e., NE, RRD, NCC and SCC (figure 1(a)).

The figures 1(b)–(d) present variation of meteorological parameters for different regions. Over four studied regions, NE and RRD have cold and dry winter from December–February. During summer from June–August, high temperature and rain are observed. Moving to the South, SCC has dry season from February–August and rainy period from September–December with peak rainy month in November. The average annual temperatures have not changed largely (from 25 °C–33 °C in average). However, SCC together with NCC is one of hottest regions in Vietnam during summer time from June–August. NCC is a special case with climate characteristics of both northern and southern regions. NCC has cold weather as NE and RRD during December and February but dry and rainy seasons which shifts one month before in comparison with SCC seasons.

3. Datasets and methodology

3.1. Datasets

MODIS instruments provide near-daily measurements of global coverage. The prefix MOD and MYD are reserved for data from MODIS onboard Terra (AM overpass) and Aqua (PM overpass), respectively. MODIS aerosol products at 10 km (i.e., Optical_Depth_Land_And_Ocean at 0.550 μm of MOD04 in Collection 5.1 and AOD_550_Dark_Target_Deep_Blue_Combin-ed of MYD04 in Collection 6) and MODIS meteorological products at 5 km (Skin Temperature of MOD07 and MYD07 in Collection 6) from 2009–2014 were investigated to estimate PM$_{2.5}$ maps for Vietnam. Currently, there are seven AERONET ground stations in Vietnam which include Bac_Giang,

Figure 1. (a) Vietnamese climate zones and spatial distribution of CEM, AERONET and NCHMF stations. Four highlight regions (NE, RRD, NCC, and SCC) are considered in our study and the diamon, triangle and round shapes presents CEM , AERONET and NCHMF stations, respectively. (b) Temperature, (c) relative humidity, and (d) precipitation are averaged by region and month and used as region climate indicators. The lines present meteorological variations over four study regions.

Bach_Long_Vy, Bac_Lieu, Nghia_Do, Nha_Trang, Red_River_Delta and Son_La (figure 1(a)). AERONET AOT from 2009–2014 at seven stations were compared to MOD04 and MYD04 AOT for accuracy.

The Center for Environmental Monitoring (CEM) at Vietnam Environment Administration (VEA) provides updated PM concentrations together with hourly meteorological parameters (i.e.: temperature, pressure, radiation, wind speed and relative humidity) at stations. The PM data at six CEM stations (Phu Tho, Ha Noi, Hue, Da Nang, Ha Long and Khanh Hoa) were considered from December 2010 to September 2014 (figure 1(a)) but period differed for stations. Data at Phu Tho, Ha Noi, Hue and Da Nang stations were used for modeling while data at Ha Long and Khanh Hoa stations in 2014 were utilized for independent validation process. Further, meteorological data from the National Center for Hydro-Meteorological Forecasting (NCHMF) covering 98 ground stations for temperature, relative humidity and precipitation from 2005–2013 has been used. At each meteorological station, temperature and relative humidity are measured at 13:00 but precipitation at 13:00 and 24:00 every day (Vietnam time zone).

3.2. Methodology

The methodology for $PM_{2.5}$ estimation over Vietnam from MODIS satellite images is carried out by integrating ground based $PM_{2.5}$ and MODIS data and then using Multiple Linear Regression and universal Kriging techniques. The detail processing steps are illustrated in figure 2.

3.2.1. Data pre-processing and integration

Since satellite and ground datasets have different temporal and spatial characteristics, they need to be integrated for modeling and testing process of PM estimation. Satellite data are first resampled to a grid of 10 km over Vietnam using a bilinear function and integrated using time and location constrains following Ichoku *et al* (2002). We considered only cloud-free aerosol data pixels which had distances to a ground station within a radius of 25 km but the directed temperature pixel over a ground station because of temperature coarse spatial resolution and lightly variant. Meanwhile, ground measurements are averaged within a temporal window of 60 min coinciding with the satellite overpasses. The optimal thresholds were selected by experiments.

Figure 2. Methodology for $PM_{2.5}$ estimation from MODIS satellite images.
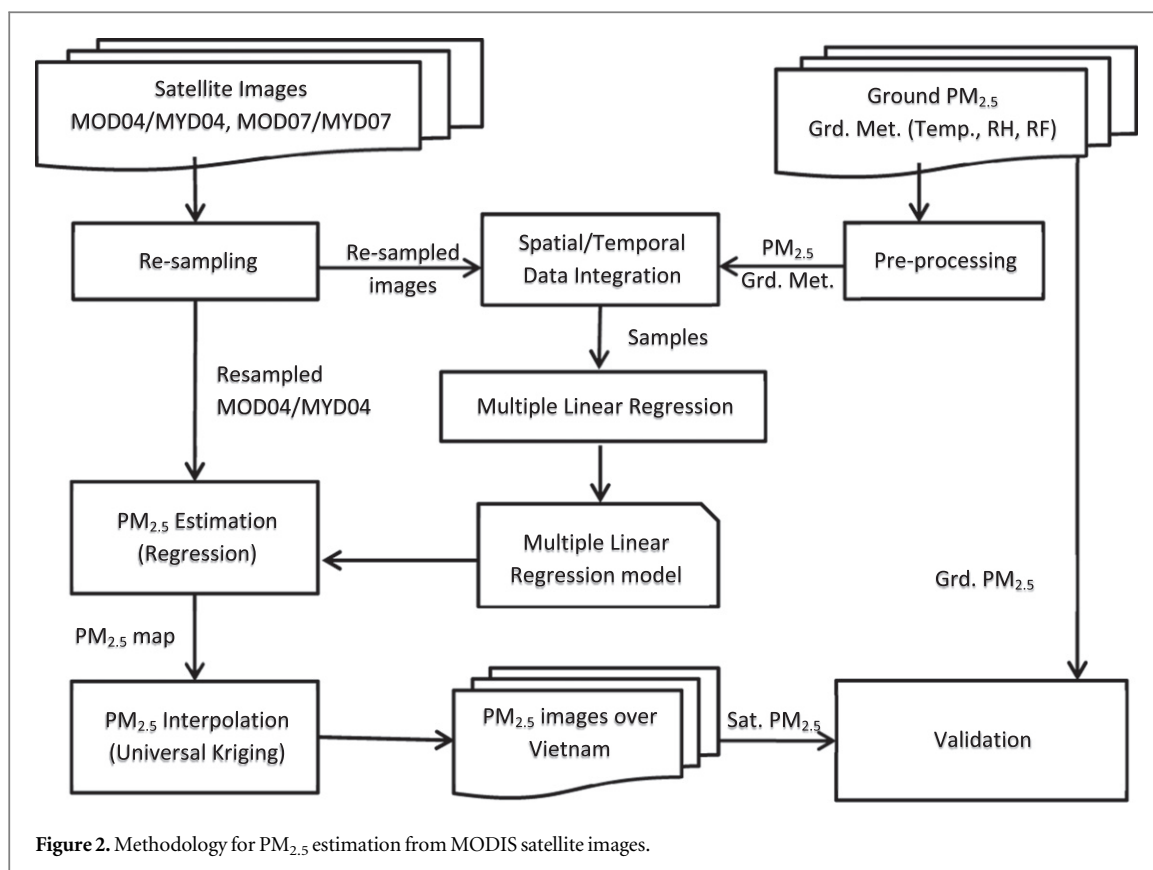
3.2.2. Feature selection

The selection of temporal parameters (i.e. aerosol and meteorological factors from satellite images) for regression model is based on correlation assessment of each factor to PM at ground level. Besides, monthly temperature, relative humidity and precipitation by region using all 98 NCHMF stations are considered for the regression model to characterize climate regions in Vietnam. The use of relative humidity as seasonal indicator is also found in Liu *et al* (2007b) and Liu *et al* (2007c).

3.2.3. Multiple linear regression

Linear regression or multiple linear regression (MLR) are considered as a common and valid methodology to estimate PM from aerosol and other meteorological factors (Chu *et al* 2003, Wang and Christopher 2003, Engel-Cox *et al* 2004, Gupta and Christopher 2009b, Schaap *et al* 2009, Zha *et al* 2010, Lee *et al* 2011). In our approach, different MLR models have been applied to estimate $PM_{2.5}$ concentrations from satellite-derived AOT, satellite-derived temperature and monthly and regional indicators (i.e. temperature, relative humidity and precipitation) normalized in range of $(-1, 1)$. The least square fitting technique determines MLR coefficients based on the least square errors of a model over a training dataset after that, the Bayesian model average (BMA) technique calculates Bayesian information criterion (BIC) and post probability for each MLR model and selects the best one that minimizes BIC and maximizes posterior probability. Finally, the

Cook's distance (Cook's D) is applied on the selected models to identify and remove the samples that have significant impacts on the estimated coefficients.

3.2.4. Universal kriging

The universal Kriging is applied on regression PM maps to interpolate values for the entire region that also included areas frequently impacted by clouds. The spatial correlation of $PM_{2.5}$ values is modeled by fitting a spherical variogram to the experimental semi-variance accounting for all valid geo-locations in a $PM_{2.5}$ image. Several models such as Gaussian, spherical, and exponential were tested to fit data. Results suggested the spherical model as the best with the lowest mean square error between the variogram model and experimental data.

After the same, we performed universal Kriging on each MOD and MYD $PM_{2.5}$ image to obtain an interpolated map together with its associated error map.

3.2.5. Evaluation

To evaluate the estimated PM concentration from our approach, we used four statistical indicators i.e., Pearson's correlation coefficient (R), coefficient of determination (r^2), root mean square error (RMSE) (i.e.: absolute error) and relative error (RE) (i.e.: percentage error). In addition, mean fractional bias (MFB) and mean fractional error (MFE) have been used to assess model performance (Boylan and Russell 2006).
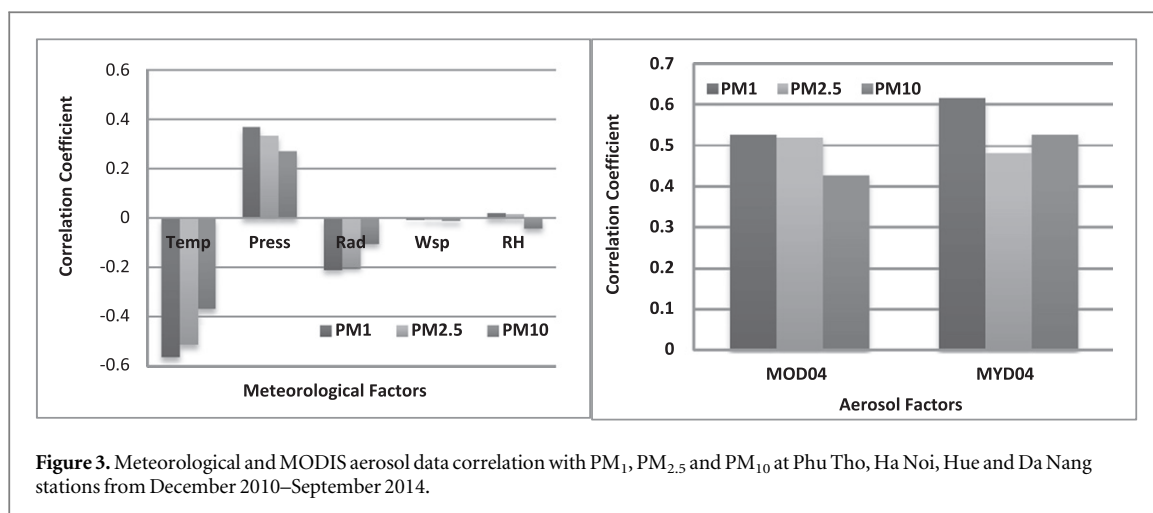
Figure 3. Meteorological and MODIS aerosol data correlation with PM$_1$, PM$_{2.5}$ and PM$_{10}$ at Phu Tho, Ha Noi, Hue and Da Nang stations from December 2010–September 2014.

$$MFB = \frac{2}{N}\sum_{i=1}^{N}\frac{\left(M_i - O_i\right)}{\left(M_i + O_i\right)} \times 100$$

PM: $\leqslant \pm 30\%$ (goal):
$\leqslant \pm 60\%$ (criteria)

$$MFE = \frac{2}{N}\sum_{i=1}^{N}\frac{|M_i - O_i|}{\left(M_i + O_i\right)} \times 100$$

PM: $\leqslant \pm 50\%$ (goal):
$\leqslant \pm 75\%$ (criteria)

Where N is number of samples. M_i and O_i. are modeled and observed PM mass concentration, respectively. In the above equation, *goal* is the level of accuracy close to the best estimation and *criteria* is the level of accuracy acceptable for standard modeling purposes.

4. Experiments and results

4.1. Meteorological and satellite data correlations with PM

The impact of meteorological variations, i.e., temperature (Temp), pressure (Pres), radiation (Rad), wind speed (Wsp) and relative humidity (RH) and satellite-derived aerosol (MOD04 and MYD04) on PM of 1, 2.5 and 10 μm diameters has been evaluated. The results clearly identify the influence of temperature and aerosol parameters on PM concentrations (see figure 3). Results in figure 3(a) suggested stronger influence of temperature on PM than pressure, radiation, wind speed and relative humidity and the relationship gradually decreasing from PM$_1$, PM$_{2.5}$ to PM$_{10}$. Correlation coefficient (R) of temperature with PM are -0.561, -0.509 and -0.366 for PM$_1$, PM$_{2.5}$ and PM$_{10}$, respectively. Hien *et al* (2002) has been confirmed that air temperature is an important determinant for his PM$_{2.5}$ estimation model in summer, meanwhile inverse correlations of temperature and PM$_{2.5}$ were explained as a trend of observing atmospheric dispersion under warm air than cold air

masses. The effect on PM at different sizes is similar to the temperature for MOD04 but MYD04 dataset (figure 2(b)). R of 0.527, 0.522 and 0.429 were obtained for MOD04 and 0.617, 0.482 and 0.592 for MYD04 datasets and PM$_1$, PM$_{2.5}$ and PM$_{10}$ datasets.

4.2. Satellite-derived aerosol and temperature validation

The MOD04 and MYD AOT at 0.550 μm are compared to AERONET aerosol data at 0.550 μm calculated using log-linear interpolation from two AOT values of two closest channels 0.500 and 0.675 μm (Nguyen *et al* 2014). The correlation coefficient between MOD04 and MYD04 to AERONET AOT are 0.867 and 0.887, respectively. Figure 4 presents their scatter plots.

We compared MODIS temperature (MOD07 and MYD07 skin temperature) and CEM temperature from December 2010 to September 2014 over four stations (Phu Tho, Ha Noi, Hue, and Da Nang). Satellite data showed strong correlation with CEM temperature (0.878 and 0.799 for MOD07 and MYD07s temperature, respectively). The MOD07 temperature and PM correlation (table 1) was even higher than station temperature and PM data (figure 3).

4.3. Predictor variables and model selection

The BMA technique suggests five and four regression models for MOD and MYD dataset, respectively (see table 2). Each regression model with different variables (i.e.: MODIS derived aerosol—AOT$_t$, MODIS derived temperature—Temp$_t$, regional monthly temperature/relative humidity/Precipitation—Temp$_{mr}$, RH$_{mr}$, Prec$_{mr}$) is evaluated using r^2 and posterior probability of model correction. The best regression models are MOD #1 and MYD #1 which use satellite-derived aerosol (AOT$_t$) and monthly and regional temperature (Temp$_{mr}$) as predictor variables and then, gain high r^2 and highest posterior probability. The low r^2 for MYD models can be explained by outliers in the MYD dataset. Therefore, Cook's distance using $4/(n\text{-}p\text{-}1)$
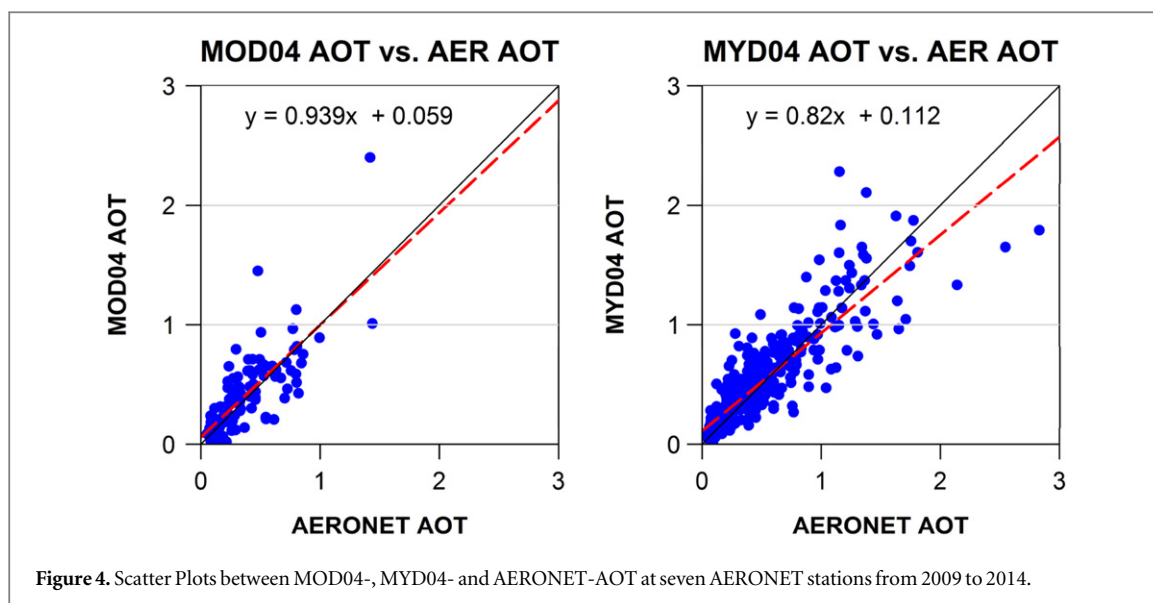
Figure 4. Scatter Plots between MOD04-, MYD04- and AERONET-AOT at seven AERONET stations from 2009 to 2014.

Table 1. Correlation Coefficient between MOD07 and MYD07 temperature with CEM temperature.

	# Samples	CEM Temp	CEM PM_1	CEM $PM_{2.5}$	CEM PM_{10}
MOD07 Temp	722	0.878	−0.617	−0.605	−0.495
MYD07 Temp	1142	0.799	−0.529	−0.477	−0.408

Table 2. Regression model selection for MOD and MYD dataset using BMA techniques. Five and four regression models with corresponding coefficients, r^2 values and post probabilities are suggested for MOD and MYD, respectively. MOD#1 and MYD#1 are the best models which use satellite-derived AOT and monthly and regional temperature as variables of $PM_{2.5}$ concentration and maximize model's posterior probability.

	MOD					MYD			
	#1	#2	#3	#4	#5	#1	#2	#3	#4
Intercept	27.063	27.063	27.063	27.063	27.063	20.591	20.591	20.591	20.591
AOT_t	27.839	28.228	28.374	27.593	28.418	29.449	29.838	29.216	29.455
$Temp_t$	—	—	—	—	2.371	—	—	—	−2.559
$Temp_{mr}$	−36.645	−45.753	−39.611	−37.484	−38.434	−18.266	−23.267	−17.992	−17.263
RH_{mr}	—	−7.933	−3.639	—	—	—	−5.966	—	—
$Prec_{mr}$	—	12.761	—	4.052	—	—	—	−2.898	—
r^2	0.529	0.545	0.534	0.531	0.530	0.197	0.206	0.197	0.197
Post Prob	0.510	0.274	0.130	0.052	0.034	0.607	0.322	0.038	0.034

threshold is applied on MOD and MYD datasets for MOD#1 and MYD#1 regression models to remove outliers. The final MOD and MYD regression models are as follows:

$$PM2.5_{t-MOD} = 21.444 \times AOT_{t-MOD} - 26.984 \times Temp_{mr} + 25.287$$

$$PM2.5_{t-MYD} = 27.401 \times AOT_{t-MYD} - 18.909 \times Temp_{mr} + 18.993$$

in which $PM2.5_{t-MOD}$. and $PM2.5_{t-MYD}$ are $PM_{2.5}$ estimated from MOD and MYD models. AOT_{t-MOD} and AOT_{t-MYD} are MOD04 and MYD04 AOT. $Temp_{mr}$ is mean monthly temperature for each region.

Table 3 presents regression results of the MOD and MYD regression models. Results suggested r^2 and RE of 0.602 and 33.348% for MOD data and 0.577 and

53.353% for MYD data. Figure 5 shows scatter plots of fitted $PM_{2.5}$ mass coentrations with ground based PM data from CEM stations. The above results are comparable to the other results found iliterare. For example, Gupta and Christopher 2009a, 2009b achieved r^2 of 0.462 and 0.547 for hour $PM_{2.5}$ estimation from MODIS AOT and meterological factors using multiple linear regression and neural network techniques over southeastern US. Liu *et al* (2007b) predicted daily $PM_{2.5}$ from MODIS AOT, high RH season in Eastern and Western US with adjusted r^2 of 0.42 and 0.21, respectively. Without meteorological data, Lee *et al* (2011) calculated daily $PM_{2.5}$ directly from MODIS AOT with r^2 varying from 0.12–0.88 using linear regression and 0.82–1.00 using mixed effect models in New England region.

Table 3. MOD and MYD regression model's results on filtered dataset using $4/(n\text{-}p\text{-}1)$ threshold in which n and p are number of samples and degree of freedom, respectively.

	MOD				MYD		
#Sample	r^2	RMSE ($\mu g\ m^{-3}$)	RE (%)	#Sample	r^2	RMSE ($\mu g\ m^{-3}$)	RE (%)
274	0.602	8.527	33.348	385	0.577	8.777	53.353
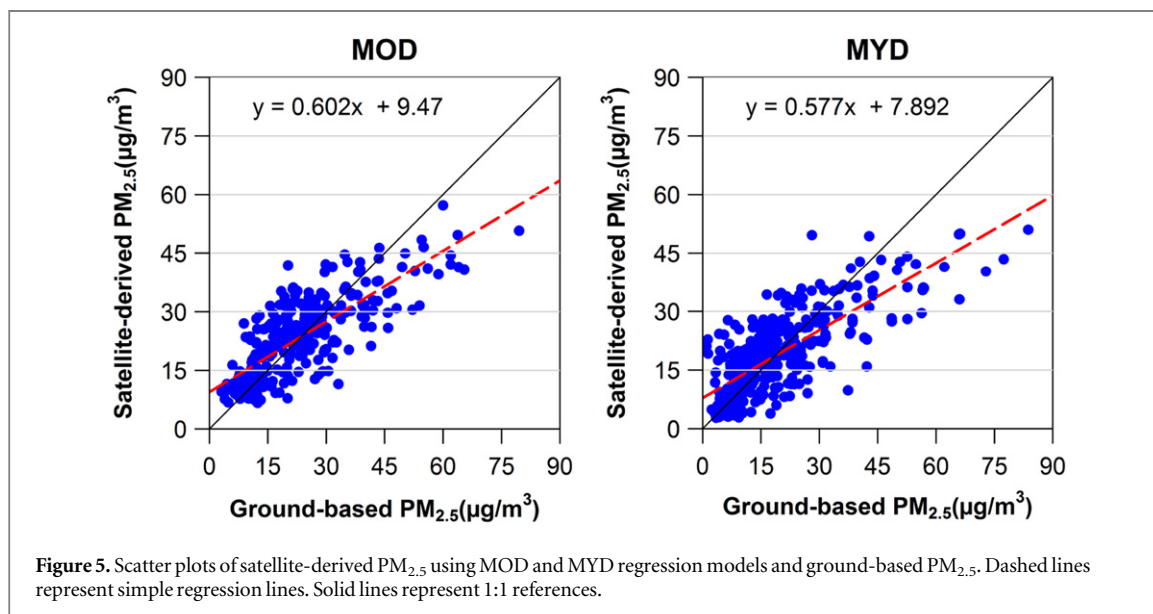
Figure 5. Scatter plots of satellite-derived $PM_{2.5}$ using MOD and MYD regression models and ground-based $PM_{2.5}$. Dashed lines represent simple regression lines. Solid lines represent 1:1 references.

Table 4. Results of universal Kriging cross validation are considered separately by model and year. Overall assessments are calculated on the total dataset.

	MOD				MYD				Total			
Year	#	r^2	RMSE ($\mu g\ m^{-3}$)	RE (%)	#	r^2	RMSE ($\mu g\ m^{-3}$)	RE (%)	#	r^2	RMSE ($\mu g\ m^{-3}$)	RE (%)
2010	4	0.918	1.371	2.402	1	0.929	0.946	2.436	5	0.924	1.158	2.419
2011	11	0.948	1.382	3.434	5	0.901	1.266	3.585	16	0.925	1.324	3.510
2012	10	0.937	1.231	3.118	7	0.950	1.312	4.031	17	0.943	1.272	3.575
2013	31	0.940	1.328	3.485	14	0.925	1.500	4.638	45	0.933	1.414	4.061
2014	29	0.953	1.337	3.696	16	0.945	1.748	6.202	45	0.949	1.542	4.949
Total	85	0.939	1.330	3.227	43	0.930	1.354	4.178	128	0.935	1.342	3.703

4.4. Interpolation model

We carried out a 3-fold cross-validation using all valid pixels for each image. Interpolated $PM_{2.5}$ values are compared with regression $PM_{2.5}$ values to evaluate Kriging performance (table 4). For interpolation model, r^2 and RE is 0.935 and 3.703% on average of total 128 datasets, which suggests robustness of our approach.

5. Validation

5.1. Overall assessment

Validation is carried out using 85 MOD and 43 MYD $PM_{2.5}$ images from December 2010 to September 2014. First, we projected each MODIS image on the Vietnamese grid of 10 km to create a $PM_{2.5}$ map using the regression model. Because percentage of available data ($\geqslant$30%) in each image is limited, the data has been interpolated from valid cells. We extracted $PM_{2.5}$ over four CEM automatic ground stations (i.e.: Phu Tho, Ha Noi, Hue and Da Nang) which are representative of four regions NE, RRD, NCC and SCC, respectively and compared them with ground $PM_{2.5}$. Table 5 shows different results of MOD and MYD model performance in which MOD model is slightly dominant. Satellite derived $PM_{2.5}$ have moderate correlation and error to ground-based $PM_{2.5}$ ($r^2 = 0.427$ and RE = 39.957%) for MOD dataset but lower correlation and error ($r^2 = 0.337$ and RE = 39.459%) for MYD dataset. Both model performances meet the goals proposed by (Boylan and Russell 2006) for MFB ($\leqslant \pm$ 30%) and MFE ($\leqslant \pm$ 50%). Figure 6 shows satellite-derived $PM_{2.5}$ versus ground measured $PM_{2.5}$ at Phu Tho, Ha Noi, Hue and Da Nang stations. Satellite-derived $PM_{2.5}$ was able to replicate average of ground $PM_{2.5}$ spatial

Table 5. Overall validation of satellite-derived $PM_{2.5}$ maps over Phu Tho, Ha Noi, Hue and Da Nang in NE, RRD, NCC and SCC regions, respectively. Results are separated by MOD and MYD datasets and accumulated in total.

	#Images	#Samples	r^2	RMSE (μg m^{-3})	RE (%)	MFB (%)	MFE (%)
MOD	85	189	0.427	21.709	39.957	0.491	34.954
MYD	43	96	0.337	17.188	39.458	3.639	34.799
Total	128	285	0.411	20.299	39.789	1.552	34.902
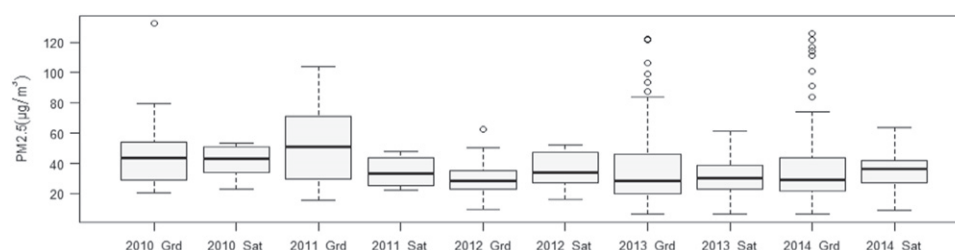
Figure 6. Boxplots shows the variation of satellite derived $PM_{2.5}$ (_Sat) and ground based $PM_{2.5}$ (_Grd) from 2010–2014 at Phu Tho, Ha Noi, Hue and Da Nang stations. The means values are shown as the thick solid line in the plots.

Table 6. Satellite-derived $PM_{2.5}$ is validated with ground-based $PM_{2.5}$ by station.

Station	#Samples	r^2	RMSE (μg m^{-3})	RE (%)	MFB (%)	MFE (%)
Phu Tho	68	0.412	26.146	39.693	-7.047	36.274
Ha Noi	52	0.158	26.448	36.195	-14.485	38.269
Hue	60	0.300	19.786	32.747	-9.323	33.038
Da Nang	105	0.281	10.276	45.656	21.276	33.410

Table 7. Independent validation of satellite-derived $PM_{2.5}$ and ground-based $PM_{2.5}$ at Ha Long and Khanh Hoa stations.

Station	#Samples	r^2	RMSE (μg m^{-3})	RE (%)	MFB (%)	MFE (%)
Ha Long	40	0.455	21.512	45.236	-2.975	38.695
Khanh Hoa	45	0.444	8.551	46.446	25.988	33.960

patterns, however limited in capturing maximal or minimal peaks. We also analyzed results for different locations (table 6). Satellite-derived $PM_{2.5}$ at Phu Tho station has the best correlation ($r^2 = 0.412$ and RE = 39.693%) in comparison with ground-based $PM_{2.5}$. At Ha Noi station, predicted $PM_{2.5}$ has low quality with $r^2 = 0.158$ and RE = 36.195% mainly due to large variations in the Ha Noi dataset. Large error was observed at Da Nang station (RE = 45.656% and $r^2 = 0.281$) while moderate results were obtained for Hue station (RE = 32.747% and $r^2 = 0.300$). MFBs show that $PM_{2.5}$ is underestimated in Phu Tho, Ha Noi, and Hue datasets but overestimated in the Da Nang dataset. However, four station's MFBs and MFEs still meet the goals. We attribute the errors to geolocation. For example, all CEM stations are located closely to rd sides whereas satellites signals represent an average of 10 km leading to high errors on satellite-derived $PM_{2.5}$.

5.2. Independent validation
We used Ha Long and Khanh Hoa stations for independent validation. They are located in the NE and SCC regions respectively. Ha Long is in Quang Ninh which is a mountainous and coastal province with four distinct seasons and dominated by tourism and coal mining (90 percent of coal output of Vietnam). Khanh Hoa is mostly mountainous and coastal with two distinct seasons (i.e. rainy season from September to December and dry season on other months). It has strong industry and services and a small agriculture sector.

Table 7 shows the high correlation for both station datasets ($r^2 = 0.455$ and 0.444 for Ha Long and Khanh Hoa respectively). RMSEs are 21.512 μg m^{-3} and 8.551 μg m^{-3} for Ha Long and Khanh Hoa datasets but corresponding REs are nearly the same (45.236% and 46.446%). Our models could estimate $PM_{2.5}$ quite well in Khanh Hoa (slope = 0.609) but underestimated $PM_{2.5}$ in the Ha Long (slope = 0.271) which can be explained by appearance of big $PM_{2.5}$ values ($\geqslant 90$ μg m^{-3}) in January at Ha Long station, although MFBs and MFEs still meet the goals. Figure 7 shows satellite-derived $PM_{2.5}$ values versus observed $PM_{2.5}$ at Ha Long and Khanh Hoa stations. Time-series patterns are matched well in the Khanh Hoa dataset.
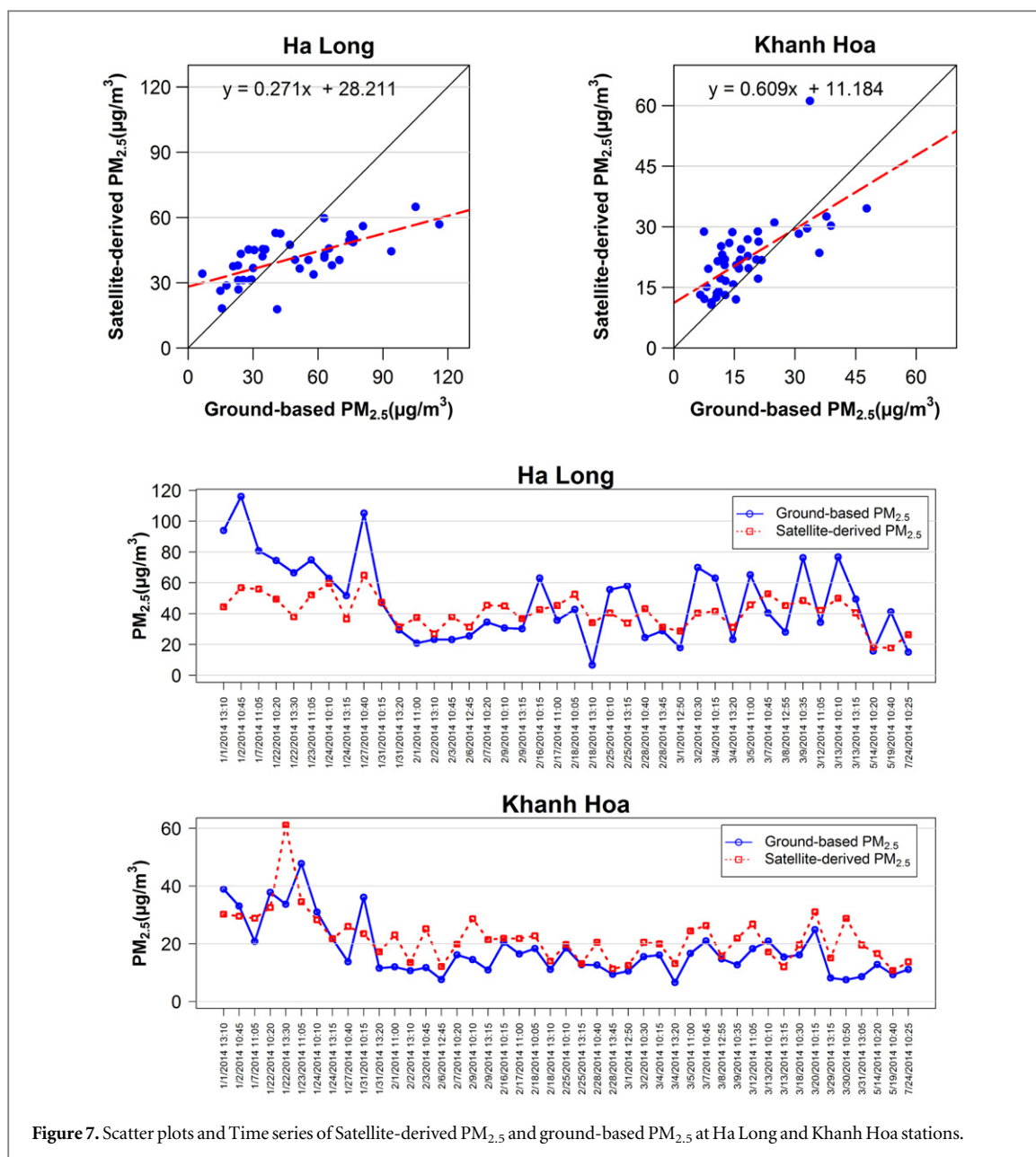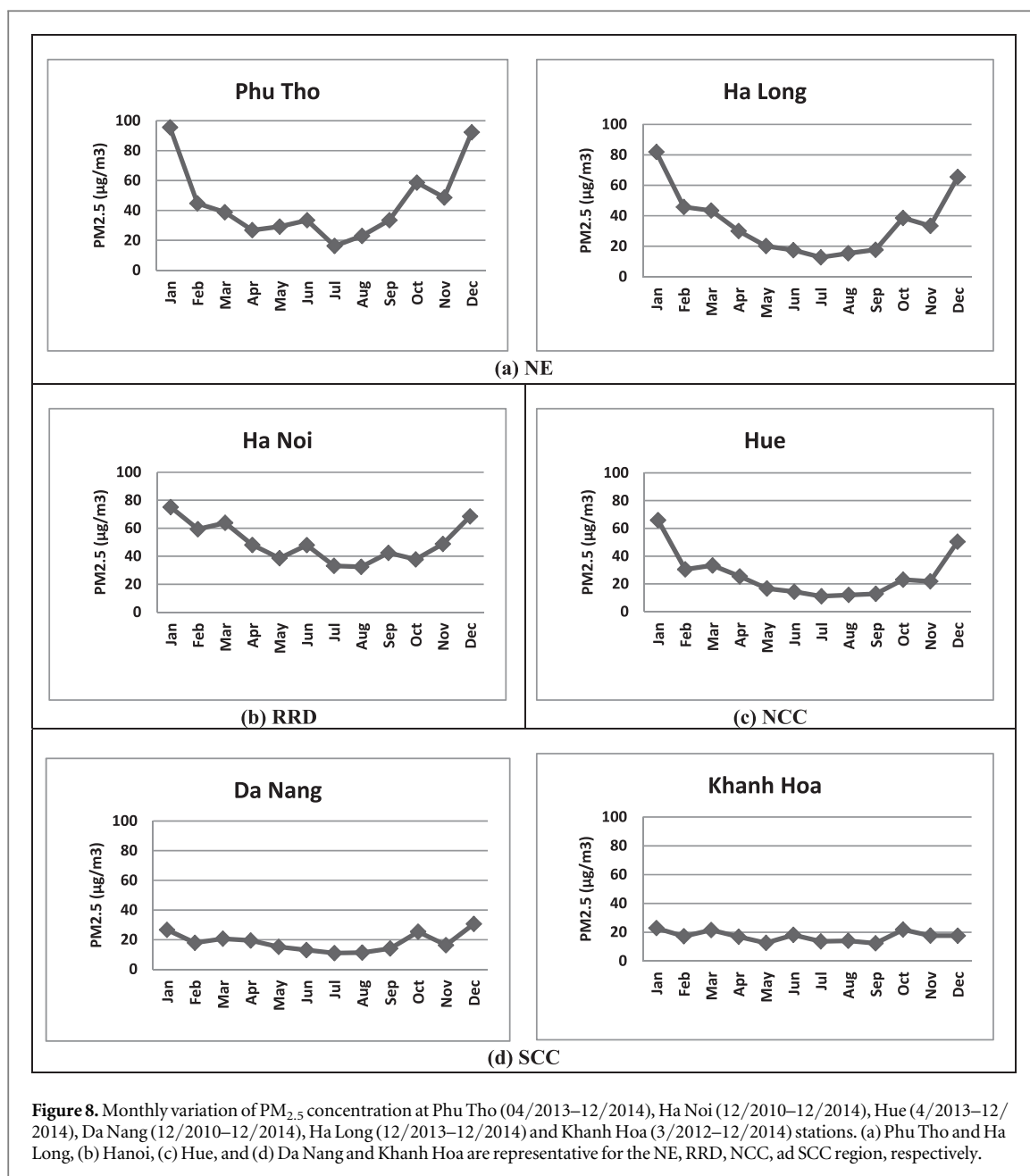
Figure 7. Scatter plots and Time series of Satellite-derived PM$_{2.5}$ and ground-based PM$_{2.5}$ at Ha Long and Khanh Hoa stations.

5.3. Seasonal and spatial trends

Figure 8 presents monthly PM$_{2.5}$ concentrations in NE, RRD, NCC and SCC representative of Phu Tho—Ha Long, Ha Noi, Hue, Da Nang—Khanh Hoa stations. Figures 9 and 10 are corresponding monthly PM$_{2.5}$ maps averaged from several individual maps (table 8). Regarding ground measurements, two stations in the same region has same data trends (i.e., Phu Tho, Ha Long in NE (figure 8(a)) and Da Nang, Khanh Hoa in NCC (figure 8(d)) which show consistence of PM$_{2.5}$ concentrations over region.

During winter in December and January, PM$_{2.5}$ concentration is high in NE and RRD. It can be explained by influence of the high pressure concuring with NE monsoon and cold weather, which leads to poor atmospheric dispersion and therefore, enhances a high buildup of air pollutants (Kim Oanh *et al* 2006, Hoang *et al* 2014). Air pollution is more serious in NE

than RRD during this season, especially in industrial provinces including Phu Tho, Thai Nguyen, Bac Giang and Quang Ninh. Local air pollution sources in the NE region are coal mining and thermal power in Quang Ninh, steel factories (account for 2857% of Vietnam as of 2009) and cement production in Thai Nguyen and Phu Tho, agriculture activities and long-range transport pollution from China by NE monsoon during winter (Vietnam Environment Administration 2013, 2014). Meanwhile, air pollution sources in RRD are more complicated since they are often from steel production (account for 4372% of Vietnam as of 2009), cement production in Ha Nam, crafts villages in Ninh Binh, Bac Ninh, Hung Yen, Nam Dinh and Hanoi, agriculture, transportation, construction and domestic activities (Vietnam Environment Administration 2013, 2014). Otherwise, during the summer from May to August, PM$_{2.5}$

Figure 8. Monthly variation of $PM_{2.5}$ concentration at Phu Tho (04/2013–12/2014), Ha Noi (12/2010–12/2014), Hue (4/2013–12/2014), Da Nang (12/2010–12/2014), Ha Long (12/2013–12/2014) and Khanh Hoa (3/2012–12/2014) stations. (a) Phu Tho and Ha Long, (b) Hanoi, (c) Hue, and (d) Da Nang and Khanh Hoa are representative for the NE, RRD, NCC, ad SCC region, respectively.

concentration is less in NE than in RRD. However, in both regions, level of $PM_{2.5}$ is lower in this period than in winter due to summer rains. Considering $PM_{2.5}$ satellite maps in NE and RRD (figure 8), similar trends in winter and summer can be observed.

SCC with rainy and dry seasons has low $PM_{2.5}$ concentration level and no seasonal trend which is confirmed by Da Nang, Khanh Hoa station measurements (figure 8(d)), $PM_{2.5}$ maps (figure 10) and the previous study (Vietnam Environment Administration 2013).

NCC is a special case with two seasons as SCC (i.e., dry season from March to August and rainy season from August to the next January) but cold and still effected by very strong NE monsoon from the South during rainy season. As presented in figures 8(c) and 10, the seasonal pattern in NCC is

more similar to NE and RRD than SCC. The significantly cold months are from December to January. Due to low temperature and precipitation, NCC climate be similar to NE and RRD's ones during these months.

Over four regions, $PM_{2.5}$ concentrations decrease from the South to the North. In addition, stronger $PM_{2.5}$ is observed in March especially in NE and RRD and we attribute the same to vegetation fires, especially agricultural residue burning (Le *et al* 2014). More data is needed to substantiate the causative factors of particulate pollution. In conclusion, satellite-derived PM2.5 maps were able to replicate seasonal and spatial trends of ground-based measurements in four different regions. Results highlight the potential use of MODIS datasets for PM estimation at a regional scale in Vietnam.
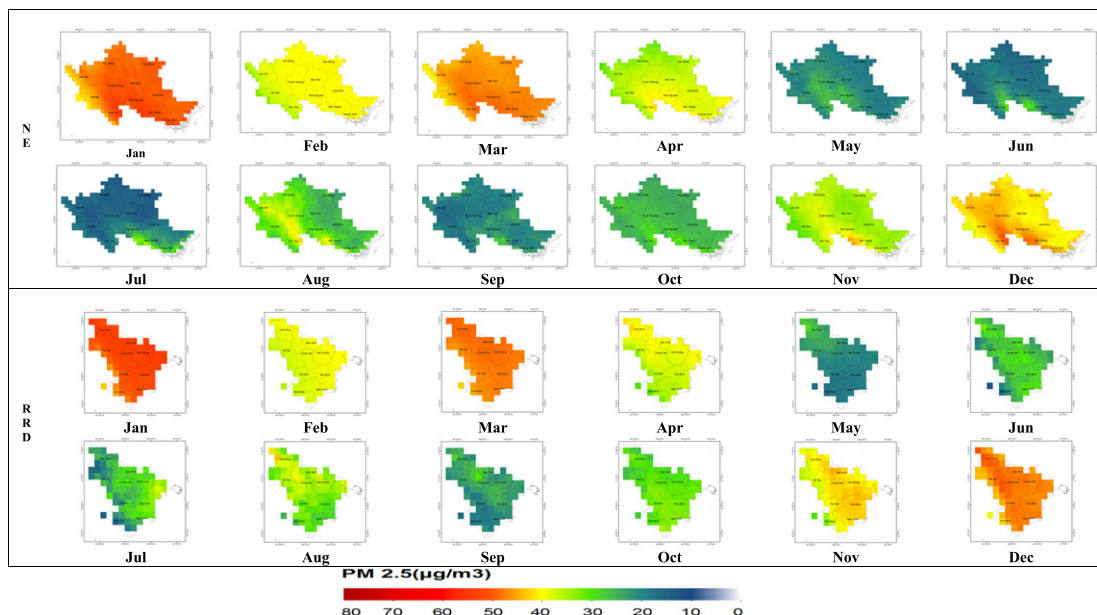
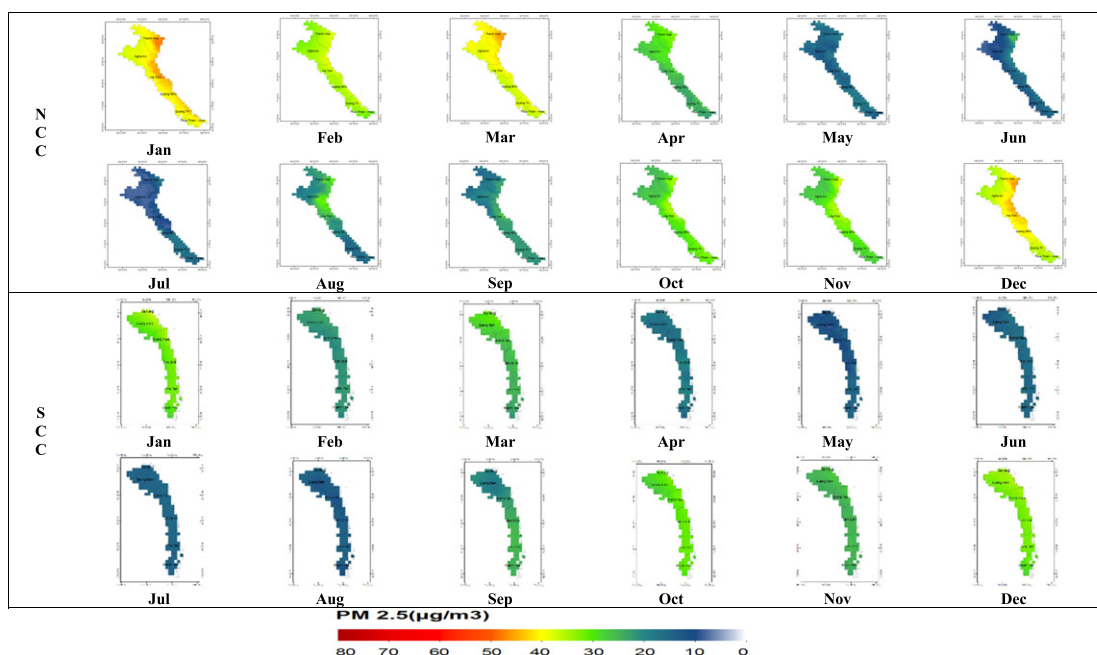**Figure 9.** Monthly variation of PM$_{2.5}$ concentration in the NE and RRD regions from 2010–2014.

Figure 10. Monthly variation of PM$_{2.5}$ concentration in the NCC and SCC regions from 2010–2014.

Table 8. Number of PM$_{2.5}$ maps by month in the validation dataset.

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
14	29	34	5	2	5	1	2	1	12	5	18

6. Conclusion

In this study, we developed an approach for temporal particulate matter concentration mapping over Vietnam using MODIS satellite images. PM$_{2.5}$ maps from December 2010 to September 2014 at a 10 km spatial resolution were developed through integrating ground based PM and MODIS data using multiple linear regression and universal Kriging techniques. Data representing four climatology regions in Vietnam i.e., North East, Red River Delta, North Central Coast and South Central Coast are used for modeling and

validation. Validation on satellite derived PM$_{2.5}$ datasets showed significant results in comparison with ground-based measurements. In addition, validation on two independent datasets (i.e.: Ha Long in North East and Khanh Hoa in South Central Coast region) showed similar results suggesting spatial stability and robustness of our methodology. Seasonal trends of PM$_{2.5}$ concentration over NE, RRD, NCC and SCC regions showed consistency between ground stations and satellite-derived PM$_{2.5}$ maps. Over four regions, PM$_{2.5}$ concentrations decreased from the South to the North and we attribute the patterns to regional climatology. Further, limitation in capturing maximal and minimal PM$_{2.5}$ peaks is being addressed. More concerning on ground data, atmospheric conditions and physical aspects for regression model should be done to improve model performance.

Acknowledgments

References

Anderson H R *et al* 2012 Satellite-based estimates of ambient air pollution and global variations in childhood asthma prevalence *Environ. Health Perspect.* **120** 1333–9

Atkinson R W, Cohen A, Mehta S and Anderson H R 2012 Systematic review and meta-analysis of epidemiological time-series studies on outdoor air pollution and health in Asia *Air Quality, Atmosphere & Health* **5** 383–91

Badarinath K V S *et al* 2007 Multiyear ground-based and satellite observations of aerosol properties over a tropical urban area in India *Atmospheric Science Letters* **8** 7–13

Badarinath K V S *et al* 2009 Analysis of aerosol and carbon monoxide characteristics over Arabian Sea during crop residue burning period in the Indo-Gangetic plains using multi-satellite remote sensing datasets *Atmospheric and Solar-Terrestrial Physics* **71** 1267–76

Boylan J W and Russell A G 2006 PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models *Atmos. Environ.* **40** 13

Chu D A *et al* 2003 Global monitoring of air pollution over land from the Earth observing system-terra moderate resolution imaging spectroradiometer (MODIS) *J. Geophys. Res.* **108** 4661

Duncan B N *et al* 2014 Satellite data of atmospheric pollution for US air quality applications: examples of applications, summary of data end-user resources, answers to FAQs, and common mistakes to avoid *Atmos. Environ.* **94** 647–62

Engel-Cox J A, Holloman C H, Coutant B W and Hoff R M 2004 Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality *Atmos. Environ.* **38** 2495–509

EPI 2014 (http://epi.yale.edu/epi)

Fann N, Amy D L, Susan C A, Karen W, David R and Bryan J H 2012 Estimating the national public health burden associated with exposure to ambient PM$_{2.5}$ and ozone *Risk Anal.* **32** 81–95

Gupta P and Christopher S A 2008 Seven year particulate matter air quality assessment from surface and satellite measurements *Atmos. Chem. Phys.* **8** 3311–24

Gupta P and Christopher S A 2009a Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: a neural network approach *J. Geophys. Res.* **114** D20205

Gupta P and Christopher S A 2009b Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: Multiple regression approach *J. Geophys. Res.* **114** D14205

Hai C D and Kim Oanh N T 2013 Effects of local, regional meteorology and emission sources on mass and compositions of particulate matter in Ha Noi *Atmos. Environ.* **78** 105–12

Hang N T and Kim Oanh N T 2014 Chemical characterization and sources apportionment offine particulate pollution in a mining town of Vietnam *Atmos. Res.* **145–6** 214–25

Hien P D, Bac V T, Tham H C, Nhan D D and Vinh L D 2002 Influence of meteorological conditions on PM$_{2.5}$ and PM$_{2.5–10}$ concentrations during the monsoon season in Ha Noi, Vietnam *Atmos. Environ.* **3473–84** 3473–84

Hirtl M, Mantovani S, Krüger B C, Triebnig G and Flandorfer C 2014 Improvement of air quality forecasts wth satellite and ground based particulate matter observations *Atmos. Environ.* **84** 20–7

Hoang X C, Nghiem T D, Nguyen H P and Kim Oanh N T 2014 Air quality management in Hanoi City, Vietnam Improving air quality in Asian developing countries *Compilation Res. Findings* p 113–31

Ichoku C, Chu D A, Mattoo S and Kaufiman Y J 2002 A spatio-temporal approach for global validation and analysis of MODIS aerosol products *Geophys. Res. Lett.* **29** 1616

Kharol S K *et al* 2012 Black carbon aerosol variations over Patiala city, Punjab, India—a study during agriculture crop residue burning period using ground measurements and satellite data *Atmospheric and Solar-Terrestrial Physics* **84** 45–51

Kim Oanh N T *et al* 2006 Particulate air pollution in six Asian cities: spatial and temporal distributions, and associated sources *Atmos. Environ.* **40** 3367–80

Kloog I, Ridgway B, Koutrakis P, Coull B A and Schwartz J D 2013 Long-and short-term exposure to PM$_{2.5}$ and mortality: using novel exposure models *Epidemiology* **24** 555–61

Krzyzanowski M, Apte J S, Bonjour S P, Brauer M, Cohen A J and Pruss-Ustun A M 2014 Air pollution in the mega-cities *Current Environ. Health Rep.* **1** 185–91

Laden F *et al* 2000 Association of fine particulate matter from different sources with daily mortality in six US cities *Environ. Health Persp.* **108** 941

Le T H, Nguyen T N T, Laskob K, Ilavajhalac S, Vadrevub P K and Justice C 2014 Vegetation fires and air pollution in Vietnam *Environ. Pollut.* **195** 267–75

Lee H J, Liu Y, Coull B A, Schwartz J and Koutrakis P 2011 A novel calibration approach of MODIS AOD data to predict PM$_{2.5}$ concentrations *Atmos. Chem. Phys.* **11** 7991–8002

Liu Y, Franklin M, Kahn R and Koutrakis P 2007a Using aerosol optical thickness to predict ground-level PM$_{2.5}$ concentrations in the St. Louis area: a comparison between MISR and MODIS *Remote Sens. Environ.* **107** 33–44

Liu Y, Koutrakis P, Kahn R, Turquety S and Yantosca R M 2007b Estimating fine particulate matter component concentrations and size distributions using satellite-retrieved fractional aerosol optical depth: 2. A case study *Air & Waste Manage. Assoc.* **57** 1360–9

Liu Y, Koutrakis P, Kahn R, Turquety S and Yantosca R M 2007c Estimating fine particulate matter component concentrations and size distributions using satelliteretrieved fractional

aerosol optical depth: 1. Method development *Air & Waste Manage. Assoc.* **57** 1351–9

Luong C K, Ho T V T, Tran N T and Nguyen L D 2010 Detecting air polution in Vietnam by optical satellite images *The 31st Asian Conf. on Remote Sensing*

Ma Z, Hu X, Huang L, Bi J and Liu Y 2014 Estimating ground-level $PM_{2.5}$ in China using satellite remote sensing *Environ. Sci. Technol.* **48** 7436–44

MacNee W and Donaldson K 2003 Mechanism of lung injury caused by PM_{10} and ultrafine particles with special reference to COPD *Environ. Sci. Technol.* **21** 47S–51S

Monks P S and Beirle S 2011 Applications of satellite observations of tropospheric composition *The Remote Sensing of Tropospheric Composition from Space, Physics of Earth and Space Environments* ed J P Burrows *et al* (Berlin, Heidelberg: Springer- Verlag)

Nguyen N H and Tran V A 2014 Estimation of PM10 from AOT of satellite LandSat 8 image over Ha Noi city *Int. Symp. on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences* pp 2014

Nguyen T N T, Ta V C, Le T H and Mantovani S 2014 Particulate matter concentration estimation from satellite aerosol and meteorological parameters, data-driven approaches *Knowledge and Systems Engineering* **224** 351–62

Pelletier B, Santer R and Vidot J 2007 Retrieving of particulate matter from optical measurements: a semiparametric approach *J. Geophys. Res.* **112** D06208

Pope C A I, Ezzati M and Dockery D W 2009 Fine-particulate air pollution and life expectancy in the United States *The New England J. of Medicine* **360** 376–86

Schaap M, Apituley A, Timmermans R M A, Koelemeijer R B A and Leeuw G D 2009 Exploring the relation between aerosol optical depth and $PM_{2.5}$ at Cabauw, the Netherlands *Atmos. Chem. Phys.* **9** 909–25

Vadrevu K P *et al* 2014 Analysis of Southeast Asian pollution episode during June 2013 using satellite remote sensing datasets *Environ. Pollut.* **195** 245–56

van Donkelaar A, Martin R V, Brauer M, Kahn R, Levy R, Verduzco C and Villeneuve P J 2010 Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application *Environ. Health Persp.* **118** 847–55

van Donkelaar A, Martin R V, Levy R, Da Silva A M, Krzyzanowski M, Chubarov N E, Semutnikovag E and Cohen A J 2011 Satellite-based estimates of ground-level fine particulate matter during extreme events: a case study of the Moscow fires in 2010 *Atmos. Environ.* **45** 6225–32

Vietnam Environment Adminstration 2013 National Environmental Report on Air Environment *NARENCA*

Vietnam Environmental Adminstration 2014 National Environmental Report on Rural Environment *NARENCA*

Wang J and Christopher S A 2003 Intercomparison between satellite-derived aerosol optical thickness and $PM_{2.5}$ mass: implication for air quality studies *Geophys. Res. Lett.* **30** 2095

World Health Organization 2006 Working together for health—the World Health Report (http://who.int/whr/2006/whr06_en.pdf)

Yahi H, Santer R, Weill A, Crepon M and Thiria S 2011 Exploratory study for estimating atmospheric low level particle pollution based on vertical integrated optical measurements *Atmos. Environ.* **45** 3891–902

Zha Y, Gao J, Jiang J, Lu H and Huang J 2010 Monitoring of urban air pollution from MODIS aerosol data: effect of meteorological parameters *Tellus* B **62** 109–16