The influence of South East Asia Forest Fires on ambient Particulate Matter concentrations in Singapore: An Ecological Study using Random Forest and Vector Autoregressive Models

Jayanthi Rajarethinam (✉ janth2@yahoo.com)
National Environment Agency https://orcid.org/0000-0001-8789-960X

Joel Aik
National Environment Agency

Jing Tian
National University of Singapore

Research

Keywords: Air quality, forest fires, random forest model, vector autoregressive model

Posted Date: July 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-41720/v1

Version of Record: A version of this preprint was published at International Journal of Environmental Research and Public Health on December 14th, 2020. See the published version at https://doi.org/10.3390/ijerph17249345.

Abstract

Background In recent decades, haze, due to biomass burning, has become a recurring problem in *Southeast Asia* (SEA). Haze degrades air quality, thus, causing detrimental effects on human health. Exposure to atmospheric *particulate matter* (PM) remains an important public health concern.

Methods In this paper, we examined the long-term seasonality of $PM_{2.5}$ and PM_{10} in Singapore. To study the association between forest fires in SEA and air quality in Singapore, we built two machine learning models, including the *random forest* (RF) model and the *vector autoregressive* (VAR) model, using a benchmark air quality dataset containing daily $PM_{2.5}$ and PM_{10} from 2009 to 2018. Furthermore, we incorporated weather parameters as independent variables, to understand their effects on air quality.

Results We observed two annual peaks, one in the middle of the year and one at the end of the year for both $PM_{2.5}$ and PM_{10}. Singapore was more affected by fires from Kalimantan compared to fires from other SEA countries. In our experimental results, VAR models performed better than RF with *Mean Absolute Percentage Error* (MAPE) values being 0.8% and 6.1% lower for $PM_{2.5}$ and $PM_{10,}$ respectively.

Conclusions Our study findings suggest that air quality in Singapore can be reasonably anticipated with predictive models that incorporate information on forest fires and weather variations. The public communication of anticipated air quality at the national level benefit who are at higher risk of experiencing poorer health due to poorer air quality.

Background

Biomass burning is the burning of living and dead vegetation, and it can occur naturally or due to human activities. Biomass burning is a significant source of gases and particulates to the regional and global atmosphere [1–3]. Therefore, it is a substantial source of air pollution that affects local air quality as well as the air quality of distant places. Haze, generated by biomass burning, can have detrimental impacts on human health [4–8], climate, biodiversity, tourism and agricultural production [9] as well as degrade visibility [10].

Since 1960s, large fires have occurred in Sumatra; however, the first fire event in Kalimantan happened in the 1980s [10]. In recent decades, biomass burning has become a recurring phenomenon in mainland *Southeast Asia* (SEA) and the islands of Sumatra and Borneo [12–14]. The majority of biomass burning in Southeast Asia occur due to human initiated activities such as land clearing for oil palm plantations, other causes of deforestation, poor peatland management, and burning of agriculture waste [15, 16]. Haze can be felt even in downwind locations such as Singapore [17, 18].

Several studies have shown that meteorological conditions have significant influence on the formation of haze [19–24]. In 2012, Reid *et al.* [25], investigated relationships between fire hotspot appearance and various weather phenomena as well as climate variabilities in different timescales and found that the es over different parts of the Maritime Continent. Haze was also shown to be worse in El Niňo years [26]. Previous study conducted in Singapore demonstrated that haze varies across year, season, and location and is influenced by local and regional weather, climate, and regional burning [10]. A study on the 1997 Indonesia forest fires reported aerosols being transmitted from Kalimantan to other countries in SEA, including Singapore [27]. Differences in haze concentrations and variation in the relative contributions from the various source regions are seen between monitoring stations across Singapore, on a seasonal as well as on an inter-annual timescale [10]. Even across small scales, such as in Singapore, variation in local meteorology can impact concentrations of particulate matter significantly, and it emphasises the importance of the scale of modelling both spatially and temporally. The *Met Office* (MO) and the *Meteorological Service Singapore* (MSS) have previously established a haze forecast system to predict haze in Singapore [28]. Extreme haze conditions observed in the Maritime Continent and in mainland Southeast Asia in 2013 and 2014 was accurately reproduced by the modelling system that was developed. However, to the best of our knowledge, there is no long term study on the seasonality of air quality in Singapore, and predictive modelling that provides daily air quality predictions. A daily prediction of air quality will be useful for nationwide planning for community activities.

Other researchers have used several machine learning techniques to predict air quality. A novel spatiotemporal deep learning based air quality prediction method, was proposed by researchers in Beijing, and the study showed that the proposed method outperformed models using the artificial neural network, regression moving average, and support vector regression techniques [29]. Another study explored three methods: (i) laboratory univariate linear regression, (ii) empirical multiple linear regression, and (iii) machine-learning-based calibration models using *random forests* (RF), and concluded that combining RF models with carefully controlled state-of-the-art multipollutant sensor packages improves the performances of prediction models of air quality sensors [29]. Another study focussing on forecasting urban air pollution show that using different features in multivariate modelling with M5P algorithm yields the best forecasting performances [30].

In this present study, we examined the association between forest fires in SEA and air quality in Singapore using different statistical models. Daily air quality forecasts will help the community to be better prepared for outdoor activities, and is especially useful for vulnerable individuals.

Methods

Study setting

We conducted our study in Singapore (1° 17’N 103° 50’E), a city state with a land area of 724.2 square kilometer, and a population density of 7,804 people per square kilometer, one of the highest population densities in the world [31]. Singapore experiences a tropical climate with abundant rainfall, high and uniform temperatures and high humidity all year round [32].

Climate data

Daily mean temperature (in degrees celsius), minimum temperature (in degrees celsius), maximum temperature (in degrees celsius), relative humidity (in percentage), mean wind speed (meters per speed), minimum wind speed (in meters per speed), maximum wind speed (meters per speed), wind direction (0 to 360 degrees) and total rainfall (in millimeter) from 2009 to 2018 recorded in Changi weather station in Singapore is obtained from MSS.

Air quality data

Biomass burning contributes mainly to these two pollutants; *particulate matter 2.5* ($PM_{2.5}$) which are particles in the air that are 2.5 micrometers or less in diameter, and *particulate matter 10* (PM_{10}), which are particles in the air that are 10 micrometers or less in diameter. These two pollutants are chosen for this study. The 24-hour average of $PM_{2.5}$ and PM_{10} for Singapore is recorded daily from 2009 to 2018 (Fig. 1). The units for both $PM_{2.5}$ and PM_{10} are microgram per cubic meter.

Forest fire data

Daily forest fire hotspot counts in Malaysia (Peninsular Malaysia, Sabah and Sarawak) and Indonesia (Sumatra and Kalimantan) are obtained from Association of Southeast Asian Nations Specialised Meterological Centre for 2009 to 2018 [33] (Fig. 2). The hotspots depicted are derived from the NOAA satellite and they represent locations with possible fires. Some hotspots may go undetected due to cloudy conditions or incomplete satellite pass. Hotspot counts from year 2016 onwards are based on the NOAA-19 satellite, and for the period from year 2006–2015 is based on the NOAA-18 satellite. The fire detection algorithm is described in detailed in the website [33]. The illustration below shows how the hotspots are counted [33].

Statistical analyses

The outcome variables for this study are $PM_{2.5}$ and PM_{10}. The independent variables are i) mean temperature, ii) minimum temperature, iii) maximum temperature, iv) relative humidity v) mean wind speed, vi) minimum wind speed, vii) maximum wind speed, viii) wind direction, ix) total rainfall, x) counts of hotspots in Kalimantan, xi) counts of hotspots in Sumatra, xii) counts of hotspots in Sabah and Sarawak and xiii) counts of hotspots in Peninsular Malaysia. Each independent variable has 31 variations, with lags from 0 days to 31 days (Additional File 1). Correlations tests are carried out using the “corrr” package in the R statistical language [34] to determine the association between the outcome variables and each of the independent variables. We evaluated the trend and seasonality of the daily values of $PM_{2.5}$ and PM_{10} in separate time series models using the “ts” and “decompose” package implemented in the R statistical language [34]. The *Kwiatkowski–Phillips–Schmidt–Shin* (KPSS) was used to test if the time series was stationary. KPSS test for both $PM_{2.5}$ and PM_{10} showed they were both stationary over time ($p\text{-value} < 0.05$). Therefore, the subsequent models used for prediction in this study are appropriate.

Model parameters and evaluation

Several models such as backward stepwise multivariate regression model, Holtwinters Time Series model, Seasonal Autoregressive Integrated Moving Average model, RF and VAR models were explored for the analyses. We chose RF and VAR model for the following reasons. RF model was chosen due to the ease of interpreting results; predictors that affect the outcomes most can be easily interpreted based on the importance calculation. Comparing the different time series models, VAR stands out as we can incorporate multiple independent variables into the model, which was relevant for our dataset. Hence, separate statistical models using RF and VAR techniques were built for both $PM_{2.5}$ and PM_{10}. The independent variables that were incorporated into the models can be found in Additional File 1. All dataset (2009–2018) were randomly split into training (70%) dataset and testing (30%) dataset to evaluate the accuracy of the models. The accuracy of the models was tested by calculating the *mean absolute percentage error* (MAPE) for each model using the following equation, where n is the total number of fitted points:

$$\frac{1}{n}\left(\sum \frac{Actualvalue - Predictedvalue}{Actualvalue}\right)*100$$

All data and statistical analyses were performed using R software version 3.6.1 [34]. Statistical significance was assessed at the 5% level. All results, where indicated, are computed for 95% *confidence intervals* (CI).

RF model

RF is an ensemble machine learning method that uses an ensemble of decision trees [35]. In RF, several bootstrap samples are drawn from the training set data, and an unpruned decision treeis fitted to each bootstrap sample. At each node of the decision tree, variable selection is carried out on a small random subset of the predictor variables. The best split on these predictors is used to split the node.

To find the best split for the model, we plotted the Out of Bag Error estimates and the error calculated on the test set [36]. We chose the split that gives the lowest error. We also calculated the percentage *mean squared error* (MSE) for each independent variable to determine the importance of each variable. MSE is calculated by the following equation:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Actualvalue - predictedvalue)^2$$

Percentage MSE is computed by calculating the percent increase in MSE of the RF model when the data for each variable were randomly permuted. For each tree, the MSE on test is recorded. Then the same is done after permuting each predictor variable. The difference between the MSE on test and the MSE of the new model, from permuting each predictor variable are then averaged over all trees, and normalized by the standard deviation of the differences. If the standard deviation of the differences is equal to 0 for a variable, the division is not done. Higher the difference is, more important the variable. We categorised the

he predicted response is obtained by averaging the predictions of all trees. RF analyses are performed using the “Random Forest” package implemented in the R statistical language [34].

VAR model

The VAR model extends the idea of univariate autoregression to multi time series regressions, where the lagged values of all series appear as regressors. The model can be seen as a linear prediction model that predicts the current value of a variable based on its own past value on the previous point in time and the past values of the other variables [37]. For example, the VAR model of two variables X_t and Y_t ($k = 2$) with the lag order p is defined as

$$Y_t = \beta_{10} + \beta_{11}Y_{t-1} +\ldots. + \beta_{1p}Y_{t-p} + \gamma_{11}X_{t-1}+\ldots. + \gamma_{1p}X_{t-p} + \mu_{1t},$$

$$X_t = \beta_{20} + \beta_{21}Y_{t-1} +\ldots. + \beta_{2p}Y_{t-p} + \gamma_{21}X_{t-1}+\ldots. + \gamma_{2p}X_{t-p} + \mu_{2t}.$$

The βs and γs can be estimated using ordinary least squared on each equation [38]. Analyses is carried out under the assumption of normality of the data. The function “VARselect” is first used to select the maximum lag which has the lowest *Akaike information criterion* (AIC). The AIC is an estimator of out-of-sample prediction error and it estimates the quality of each model, relative to each of the other models. VAR analyses are conducted using the “vars” package implemented in the R statistical language [34].

Results

Association between $PM_{2.5}$ and PM_{10} with climate and hotspots variables

The independent variables had weak correlation with $PM_{2.5}$ and PM_{10}, however, we noticed that for both $PM_{2.5}$ and PM_{10}, counts of hotspots in Kalimantan with lags between 1 to 18 days had an average correlation coefficient of 0.2, and p-value < 0.05. The correlations coefficients and corresponding p-values between the outcome variables ($PM_{2.5}$ and PM_{10}) and the climate and hotspot variables are listed in Additional File 2.

Time-series analyses of daily 24-hour average of $PM_{2.5}$ and PM_{10}

There are seasonal fluctuations in both $PM_{2.5}$ and PM_{10} over the study period. We observed two annual peaks, one in the middle of the year and one at the end of the year for both $PM_{2.5}$ and PM_{10}. There was no discerning trend, but we noticed two episodes of very poor air quality in mid-2013 and mid-2015, and they appeared to be outliers.

RF model

The RF models are built using 500 trees, and the number of variables splits that gives the lowest error for model $PM_{2.5}$ and model PM_{10} are 193 and 89, respectively. Among the independent variables, relative humidity with lags of 0, 1 and 2 days are top-ranked for $PM_{2.5}$ and PM_{10}. In addition, counts of hotspots in Kalimantan with lags of 8 and 11 days are top-ranked for $PM_{2.5}$, whilst, counts of hotpots in Kalimantan with lags of 1, 8 and 9 days are top-ranked for PM_{10}. The MSEs calculated for the rest of the variables are listed in Additional File 3. Figure 4 shows graphical comparison of the predicted and actual values for $PM_{2.5}$ and PM_{10}.

VAR model

To get the lowest AIC, the VAR model for $PM_{2.5}$ and PM_{10} was built using maximum lags of 8 and 9 respectively. The variables used in the models $PM_{2.5}$ and PM_{10} are listed in Additional File 4. Tables 1 and 2 summarizes the coefficients of the variables that were significant ($p < 0.05$) for $PM_{2.5}$ and PM_{10}, respectively.

Table 1

Coefficients for variables associated with $PM_{2.5}$ that are significant ($p < 0.05$) using VAR model

Variables	Estimate (CI)
Mean temp with 2 days lag	-2.77 (-1.58 to -3.94)
PM25 with 1 day lag	0.76 (0.56 to 0.95)
Mean wind speed with 2 days lag	0.56 (0.10 to 1.01)
PM25 with 5 days lag	0.12 (-0.10 to 0.33)
Relative humidity with 1 day lag	-0.36 (-0.75 to 0.03)
Mean wind speed with 1 day lag	-0.44 (-0.87 to -0.01)
Mean temp with 1 day lag	-2.81 (-3.91 to -1.72)
Count of hotspots in Kalimantan with 3 days lag	0.01 (-0.08 to 0.11)
Max temp with 2 days lag	-0.63 (-1.3 to 0.04)
Count of hotspots in Kalimantan with 8 days lag	0.01 (-0.08 to 0.09)
Rainfall with 1 day lag	-0.0008 (-0.03 to 0.02)
PM25 with 6 days lag	-0.06 (-0.28 to 0.16)
Min temp with 1 day lag	0.69 (-0.05 to 1.44)
Mean wind speed with 5 days lag	0.24 (-0.21 to 0.7)
Mean wind speed with 4 days lag	-0.24 (-0.69 to 0.22)
Count of hotspots in Sabah/Sarawak with 8 days lag	-0.04 (-0.21 to 0.14)
Count of hotspots in Kalimantan with 6 days lag	0.01 (-0.08 to 0.10)
Count of hotspots in Kalimantan with 1 day lag	0.01 (-0.08 to 0.09)
Max wind speed with 2 days lag	-0.05 (-0.28 to 0.17)
Count of hotspots in Sabah/Sarawak with 6 days lag	-0.04 (-0.22 to 0.15)

Table 2
Coefficients for variables associated with PM_{10} that are significant ($p < 0.05$) using VAR model

Variables	Estimate (CI)
PM10 with 1 day lag	0.75 (0.59 to 0.91)
Mean temp with 1 day lag	-3.53 (-2.49 to -4.56)
PM10 with 5 days lag	0.08 (-0.08 to 0.24)
Relative humidity with 1 day lag	-0.52 (-0.93 to -0.10)
Mean wind speed with 2 days lag	0.68 (0.19 to 1.16)
Mean temp with 2 days lag	-3.72 (-2.58 to -4.87)
Relative humidity with 2 days lag	0.31 (-0.09 to 0.72)
Mean wind speed with 1 day lag	-0.35 (-0.79 to 0.09)
Counts of hotspots in Kalimantan with 8 days lag	0.01 (-0.07 to 0.09)
Counts of hotspots in Sabah/Sarawak with 8 days lag	-0.05 (-0.24 to 0.13)
Min temp with 4 days lag	0.61 (-0.01 to 1.23)
Mean wind speed with 4 days lag	-0.33 (-0.78 to 0.13)
Rainfall with 1 day lag	-0.001 (-0.03 to 0.03)
Min temp with 1 day lag	0.84 (0.06 to 1.62)
Min temp with 2 days lag	-0.85 (-1.65 to -0.05)
Mean wind speed with 5 days lag	0.23 (-0.19 to 0.65)
Max temp with 2 days lag	-0.57 (-1.23 to 0.10)
Mean wind speed with 3 days lag	-0.23 (-0.68 to 0.22)
Counts of hotspots in Sumatra with 3 days lag	0.001 (-0.08 to 0.09)
Counts of hotspots in Sabah/Sarawak with 6 days lag	-0.04 (-0.21 to 0.14)
Rainfall with 7 days lag	0.0007 (-0.03 to 0.03)
Min temp with 9 days lag	-0.44 (-1.09 to 0.21)
Max wind speed with 2 days lag	-0.06 (-0.29 to 0.18)
Counts of hotspots in Kalimantan with 1 day lag	0.006 (-0.07 to 0.09)

Comparison of models

Table 3 shows the MAPE values for each of the four models. From Table 3, we can see that VAR models have lower MAPE performance compared to that of the RF models for both $PM_{2.5}$ and PM_{10} experiments.

Table 3
Mean Absolute Percentage Error of the Random Forest and Vector Autoregressive models for both $PM_{2.5}$ and PM_{10}

	MAPE (%)	
Outcome variable	**Random forest**	**VAR**
$PM_{2.5}$	26.8	26.0
PM_{10}	21.3	15.2

Discussion

In this study, we sought to examine the association between forest fires and air quality in Singapore. We found a positive association between ambient air particulate concentrations in Singapore and counts of instances of forest fires. This association was observed with a 1 to 8 days' lag depending on the location of the forest fires. Our study findings were consistent with other studies. Significant build-up of aerosol and black carbon concentrations was observed in the Tibetan plateau due to the occurrence of fires and transport of pollution from the nearby regions of Southeast Asia and the northern part of the Indian Peninsula [39]. Similarly, forest fires in Serbia resulted in air pollution through Mongolia, eastern China, down to the Korean peninsula [40]. This finding is not unexpected. Past research has shown that forest fire emissions were the largest contributors to the air pollution problem in regions tens of kilometers away from the fire source [41]. Our RF model picked up counts of hotspots in Kalimantan up to 11 days' lag as significant variable that affects $PM_{2.5}$ and PM_{10} concentration in the air. A similar study on the 1997 Indonesia forest fires corroborates with our results that Singapore was more affected by fires from Kalimantan compared to fires from other countries, due to the shifting of the monsoons [42]. Although Malaysia and Sumatra are closer to Singapore in terms of distance than Kalimantan [43], the models show that climatic factors are important in influencing the impact of forest fires in the air quality.

The seasonality shows that the peaks in poor air quality in Singapore occurs twice, once in the middle of the year, and one at the end of the year. This finding corresponds with other studies that show that high values of $PM_{2.5}$ and PM_{10} are reported in the middle of the year, which corresponds to the burning season [39]. Similarly, it is also reported that the burning season in SEA peaks from July to October [44]. High amounts of $PM_{2.5}$ and PM_{10} not only aggravate health issues, they also degrade visibility. Hence, these results can be used to guide tourism as well as large scale community programmes.

Based on our RF models importance plot, relative humidity is another significant variable that affects $PM_{2.5}$ and PM_{10} concentration in the air. Other studies have also concluded that relative humidity is a key factor in influencing the distribution of air quality [45, 46].

In contrast, the VAR models picked up mean temperature lagging $PM_{2.5}$ and PM_{10} by one and two days having significant negative effect on the concentration of $PM_{2.5}$ and PM_{10} in the air. The effect of mean temperature on air quality has however, been inconsistent, with several studies showing conflicting findings. Some studies have observed a negative correlation between mean temperature and concentrations of $PM_{2.5}$ and PM_{10} [47, 48]. However, there are other studies that have shown that there is a combined effect of climatic factors on the concentration of $PM_{2.5}$ and PM_{10}. For example, a study in Nagasaki, Japan concluded that temperature is positively correlated with $PM_{2.5}$ and PM_{10} during monsoons and negatively correlated during other seasons [49]. Another study in Dhaka also showed variable response of relative humidity with air pollutants according to seasonal variation [50]. Hence, machine learning methods are relevant for the predictions of air quality, due to the mixed effects of climatic factors.

Comparing RF and VAR models, the VAR models performs slightly better with MAPE values being 0.8% and 6.1% lower for $PM_{2.5}$ and PM_{10}, respectively. Hence, the VAR model can be reliably used for future predictions of the concentration of $PM_{2.5}$ and PM_{10} in urban atmosphere in Singapore. To improve the communication of predictions to the community, we can categorise the predicted values according to the Table 4 [51]. Singapore uses this category to show the levels of pollutants in the air. It will be useful to release a daily prediction of $PM_{2.5}$ and PM_{10} for community preparedness.

Table 4
Breakdown used to define the index for $PM_{2.5}$ and PM_{10}

Index category	**24-hr $PM_{2.5}$ (μg/m^3)**	**24-hr PM_{10} (μg/m^3)**
Good	0–12	0–50
Moderate	13–55	51–150
Unhealthy	56–150	151–350
Very unhealthy	151–250	351–420
Hazardous	251–350	421–500
	351–500	501–600

There are several study limitations. Other than climatic factors, there are other factors that can affect the air quality in Singapore. The models did not account for other anthropogenic sources of PM such as those from the industry and shipping. Data on these factors should be collected and included into the models, to see if they can improve the predictions. In addition, currently, the dataset for independent variables were collected from Changi Meteorological Station, which is the eastern meteorological station in Singapore. Daily news reports on pollutants have shown that different parts of Singapore can be

intensities [52]. It will be useful to provide predictions for the five areas in Singapore (north, south, east, west and central). In order to achieve that, we need to collect climate data in different meteorological stations around the island which is spatially representative, and also obtain the measurements from the hotspots to the stations as one of the variable. The models can be further developed for better spatial resolution. Lastly, analyses were done using average values for a daily prediction. It might be more useful to the community to predict the air quality on an hourly basis. Hence, moving forward we could collect hourly data and run the models.

Conclusions

Our study findings suggest that air quality in Singapore can be reasonably anticipated with predictive models that incorporate information on forest fires and weather variations. The public communication of anticipated air quality at the national level benefit who are at higher risk of experiencing poorer health due to poorer air quality.

Abbreviations

Southeast Asia: SEA, The Met Office: MO, Meteorological Service Singapore: MSS, Random forests: RF, Particulate matter 2.5: $PM_{2.5}$, Particulate matter 10: PM_{10}, Kwiatkowski–Phillips–Schmidt–Shin: KPSS, Mean Absolute Percentage Error: MAPE, Confidence Intervals: CI, Mean Squared Error: MSE, Akaike information criterion: AIC

Declarations

Ethics Approval And Consent To Participate

Not applicable

Consent For Publication

Not applicable

Availability Of Data And Materials

Data on forest fire hotspots in South East Asia can be obtained from http://asmc.asean.org/asmc-haze-hotspot-daily#Hotspot. Data on air quality and climate are owned by a third party. They are available upon reasonable request from the Meteorological Services Singapore of the National Environment Agency (email: Contact_NEA@nea.gov.sg)

Competing interests

The authors declare that they have no competing interests.

Funding

The study was funded by NEA, Singapore. The funding sources of this study had no role in the study design, data collection, data analysis, data interpretation, writing of the report, or in the decision to submit

the paper for publication.

Author Contribution

Conceptualization: Jayanthi Rajarethinam, Joel Aik, Jing Tian; **Data curation**: Jayanthi Rajarethinam; **Formal analysis**: Jayanthi Rajarethinam; **Methodology**: Jayanthi Rajarethinam, Joel Aik, Jing Tian; **Project Administration**: Joel Aik; **Resources**: Jayanthi Rajarethinam; **Software**: Jayanthi Rajarethinam; **Supervision**: Joel Aik, Jing Tian; **Writing-Original Draft**: Jayanthi Rajarethinam; **Writing-Review and Editing**: Jayanthi Rajarethinam, Joel Aik, Jing Tian

Acknowledgements

The authors thank MSS for providing the climate data for the study. We thank the informatics colleagues, Annabel Seah, Janet Ong and Stacy Soh for their comments during the brainstorming meetings as well as subsequent meetings to improve the study. We also thank the examiners for the suggestions provided during the Professional Conversion Programme for Data Analysts (Faculty of ISS-NUS) presentations, which was useful to guide the study. We are also grateful to Dr Ng Lee-Ching from EHI, NEA for her support in conducting the study.

References

1. Crutzen PJ, Heidt LE, Krasnec JP, Pollock WH, Seiler W. Biomass burning as a source of atmospheric gases CO, H2, N2O, NO, CH3Cl and COS. Nature. 1979 Nov 15;282(5736):253–6.
2. Seiler W, Crutzen PJ. Estimates of gross and net fluxes of carbon between the biosphere and the atmosphere from biomass burning. Climatic change. 1980 Sep 1;2(3):207 – 47.
3. Crutzen PJ, Andreae MO. Biomass burning in the tropics: Impact on atmospheric chemistry and biogeochemical cycles. Science. 1990 Dec 21;250(4988):1669-78.
4. Crippa P, Castruccio S, Archer-Nicholls S, Lebron GB, Kuwata M, Thota A, Sumin S, Butt E, Wiedinmyer C, Spracklen DV. Population exposure to hazardous air quality due to the 2015 fires in Equatorial Asia. Scientific reports. 2016 Nov;16(1):1–9. 6(.
5. Sigsgaard T, Forsberg B, Annesi-Maesano I, Blomberg A, Bølling A, Boman C, Bønløkke J, Brauer M, Bruce N, Héroux ME, Hirvonen MR. Health impacts of anthropogenic biomass burning in the developed world. European Respiratory Journal. 2015 Dec 1;46(6):1577-88.
6. Youssouf H, Liousse C, Roblou L, Assamoi EM, Salonen RO, Maesano C, Banerjee S, Annesi-Maesano I. Non-accidental health impacts of wildfire smoke. Int J Environ Res Public Health. 2014 Nov;11(11):11772–804.
7. Reddington CL, Butt EW, Ridley DA, Artaxo P, Morgan WT, Coe H, Spracklen DV. Air quality and human health improvements from reductions in deforestation-related fire in Brazil. Nat Geosci. 2015 Oct;8(10):768–71.
8. Aik J, Chua R, Jamali N, Chee E. The burden of acute conjunctivitis attributable to ambient particulate matter pollution in Singapore and its exacerbation during South-East Asian haze ment. 2020 Jun;11:140129.

9. Jones DS. ASEAN and transboundary haze pollution in Southeast Asia. Asia Europe Journal. 2006 Sep 1;4(3):431–46.
10. Hansen AB, Witham CS, Chong WM, Kendall E, Chew BN, Gan C, Hort MC, Lee SY. Haze in Singapore–source attribution of biomass burning PM10 from Southeast Asia. Atmospheric Chemistry & Physics. 2019 Apr 15;19(8).
11. Gellert PK. A brief history and analysis of Indonesia's forest fire crisis. Indonesia. 1998 Apr 1(65):63–85.
12. Langner A, Miettinen J, Siegert F. Land cover change 2002–2005 in Borneo and the role of fire derived from MODIS imagery. Glob Change Biol. 2007 Nov;13(11):2329–40.
13. Carlson KM, Curran LM, Ratnasari D, Pittman AM, Soares-Filho BS, Asner GP, Trigg SN, Gaveau DA, Lawrence D, Rodrigues HO. Committed carbon emissions, deforestation, and community land conversion from oil palm plantation expansion in West Kalimantan, Indonesia. Proceedings of the National Academy of Sciences. 2012 May 8;109(19):7559-64.
14. Van der Werf GR, Randerson JT, Giglio L, Collatz GJ, Mu M, Kasibhatla PS, Morton DC, DeFries RS, Jin Y, van Leeuwen TT. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). Atmospheric Chemistry and Physics. 2010 Dec 1;10(23):11707-35.
15. Marlier ME, DeFries RS, Kim PS, Koplitz SN, Jacob DJ, Mickley LJ, Myers SS. Fire emissions and regional air quality impacts from fires in oil palm, timber, and logging concessions in Indonesia. Environmental Research Letters. 2015 Aug 12;10(8):085005.
16. Lee HH, Bar Or RZ, Wang C. Biomass burning aerosols and the low-visibility events in Southeast Asia. 2017.
17. Koe LC, Arellano AF Jr, McGregor JL. Investigating the haze transport from 1997 biomass burning in Southeast Asia: its impact upon Singapore. Atmospheric Environment. 2001 May 1;35(15):2723-34.
18. Heil A, Langmann B, Aldrian E. Indonesian peat and vegetation fire emissions: Study on factors influencing large-scale smoke haze pollution using a regional atmospheric chemistry model. Mitigation and adaptation strategies for global change. 2007 Jan 1;12(1):113 – 33.
19. Flocas H, Kelessis A, Helmis C, Petrakakis M, Zoumakis M, Pappas K. Synoptic and local scale atmospheric circulation associated with air pollution episodes in an urban Mediterranean area. Theoret Appl Climatol. 2009 Mar;95(3–4)(1):265–77.
20. Wang L, Zhang N, Liu Z, Sun Y, Ji D, Wang Y. The influence of climate factors, meteorological conditions, and boundary-layer structure on severe haze pollution in the Beijing-Tianjin-Hebei region during January 2013. Advances in Meteorology. 2014 Jan 1;2014.
21. Zhao XJ, Zhao PS, Xu J, Meng W, Pu WW, Dong F, He D, Shi QF. Analysis of a winter regional haze event and its formation mechanism in the North China Plain. Atmospheric Chemistry & Physics Discussions. 2013 Jan 1;13(1).
22. Song LC, Rong G, Ying LI, Guo-Fu W. Analysis of China's haze days in the winter half-year and the
2. Advances in Climate Change Research. 2014 Jan 1;5(1):1–

6.

23. Renhe Z, Li Q, Zhang R. Meteorological conditions for the persistent severe fog and haze event over eastern China in January 2013. Science China Earth Sciences. 2014 Jan 1;57(1):26–35.
24. Fu GQ, Xu WY, Yang RF, Li JB, Zhao CS. The distribution and trends of fog and haze in the North China Plain over the past 30 years. Atmos. Chem. Phys. 2014 Nov 13;14(21):11949–58.
25. Reid JS, Xian P, Hyer EJ, Flatau MK, Ramirez EM, Turk FJ, Sampson CR, Zhang C, Fukada EM, Maloney ED. Multi-scale meteorological conceptual analysis of observed active fire hotspot activity and smoke optical depth in the Maritime Continent. Atmospheric Chemistry and Physics. 2012 Feb 15;12(4):2117.
26. Marlier ME, DeFries RS, Voulgarakis A, Kinney PL, Randerson JT, Shindell DT, Chen Y, Faluvegi G. El Niño and health risks from landscape fire emissions in southeast Asia. Nature climate change. 2013 Feb;3(2):131–6.
27. Roswintiarti O, Raman S. Three-dimensional simulations of the mean air transport during the 1997 forest fires in Kalimantan, Indonesia using a mesoscale numerical model. InAir Quality 2003 (pp. 429–438). Birkhäuser, Basel.
28. Hertwig D, Burgin L, Gan C, Hort M, Jones A, Shaw F, Witham C, Zhang K. Development and demonstration of a Lagrangian dispersion modeling system for real-time prediction of smoke haze pollution from biomass burning in Southeast Asia. Journal of Geophysical Research: Atmospheres. 2015 Dec;27(24):12605–30. 120(.
29. Li X. Ling Peng, Yuan Hu, Jing Shao, and Tianhe Chi. "Deep learning architecture for air quality predictions. Environ Sci Pollut Res. 2016;23(22):22408–17.
30. Shaban KB, Kadri A, Rezk E. Urban air pollution monitoring system with forecasting models. IEEE Sensors Journal. 2016 Jan 4;16(8):2598–606.
31. Population and Population Structure. Department of Statistics Singapore. 27. Sep 2018. 2018. https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data. Accessed at 13 Jul 2020.
32. Climate of Singapore. Meteorological Service Singapore. 2019. http://www.weather.gov.sg/climate-climate-of-singapore/. Accessed at 13 Jul 2020.
33. Transboundary Haze. Hotspot information. 2019. http://asmc.asean.org/asmc-haze-hotspot-daily#Hotspot. Accessed at 13 Jul 2020.
34. Team RC. R: A language and environment for statistical computing. 2012. Vienna, Austria: R Foundation for Statistical Computing. 2012 Oct;10.
35. Breiman L. Random forests. Machine learning. 2001 Oct 1;45(1):5–32.
36. Breiman L. Out-of-bag estimation. 1996.
37. Wild B, Eichler M, Friederich HC, Hartmann M, Zipfel S, Herzog W. A graphical vector autoregressive modelling approach to the analysis of electronic diary data. BMC medical research methodology. 2010 Dec 1;10(1):28.

38. Zivot E, Wang J. Vector autoregressive models for multivariate time series. Modeling financial time series with S-PLUS®. 2006:385–429.
39. Engling G, Zhang YN, Chan CY, Sang XF, Lin M, Ho KF, Li YS, Lin CY, Lee JJ. Characterization and sources of aerosol particles over the southeastern Tibetan Plateau during the Southeast Asia biomass-burning season. Tellus B: Chemical and Physical Meteorology. 2011 Jan 1;63(1):117 – 28.
40. Lee KH, Kim JE, Kim YJ, Kim J, von Hoyningen-Huene W. Impact of the smoke aerosol from Russian forest fires on the atmospheric environment over Korea during May 2003. Atmospheric Environment. 2005 Jan 1;39(1):85–99.
41. Lazaridis M, Latos M, Aleksandropoulou V, Hov Ø, Papayannis A, Tørseth K. Contribution of forest fire emissions to atmospheric pollution in Greece. Air quality, atmosphere & health. 2008 Nov 1;1(3):143–58.
42. Koe LC, Arellano AF Jr, McGregor JL. Investigating the haze transport from 1997 biomass burning in Southeast Asia: its impact upon Singapore. Atmospheric Environment. 2001 May 1;35(15):2723-34.
43. Distance between cities and places. URL: https://www.distancefromto.net/ Accessed at 13 July 2020.
44. BBC World News
Indonesia haze: Why do forests keep burning? URL
Accessed at 13 July 2020
BBC World News. Indonesia haze: Why do forests keep burning? URL: https://www.bbc.com/news/world-asia-34265922 Accessed at 13 July 2020.
45. Zhao CX, Wang YQ, Wang YJ, Zhang HL, Zhao BQ. Temporal and spatial distribution of PM2. 5 and PM10 pollution status and the correlation of particulate matters and meteorological factors during winter and spring in Beijing. Huan jing ke xue = Huanjing kexue. 2014 Feb 1;35(2):418 – 27.
46. Lou C, Liu H, Li Y, Peng Y, Wang J, Dai L. Relationships of relative humidity with PM 2.5 and PM 10 in the Yangtze River Delta, China. Environmental monitoring and assessment. 2017 Nov 1;189(11):582.
47. Hernandez G, Berry TA, Wallis S, Poyner D. Temperature and humidity effects on particulate matter concentrations in a sub-tropical climate during winter. 2017.
48. Akyüz M, Çabuk H. Meteorological variations of PM2. 5/PM10 concentrations and particle-associated polycyclic aromatic hydrocarbons in the atmospheric environment of Zonguldak, Turkey. Journal of Hazardous Materials. 2009 Oct 15;170(1):13–21.
49. Wang J, Ogawa S. Effects of meteorological conditions on PM2. 5 concentrations in Nagasaki, Japan. Int J Environ Res Public Health. 2015 Aug;12(8):9089–101.
50. Kayes I, Shahriar SA, Hasan K, Akhter M, Kabir MM, Salam MA. The relationships between meteorological parameters and air pollutants in an urban environment. Global Journal of Environmental Science and Management. 2019 Jul 1;5(3):265 – 78.
51. Computation of the Pollutants Standard Index (PSI). 2014. https://www.haze.gov.sg/docs/default-source/faq/computation-of-the-pollutant-standards-index-(psi).pdf Accessed at 13 July 2020.

52. National Environment Agency. Resources. Pollutants concentrations. 2019. https://www.haze.gov.sg/resources/pollutant-concentrations Accessed at 13 July 2020.

Figures

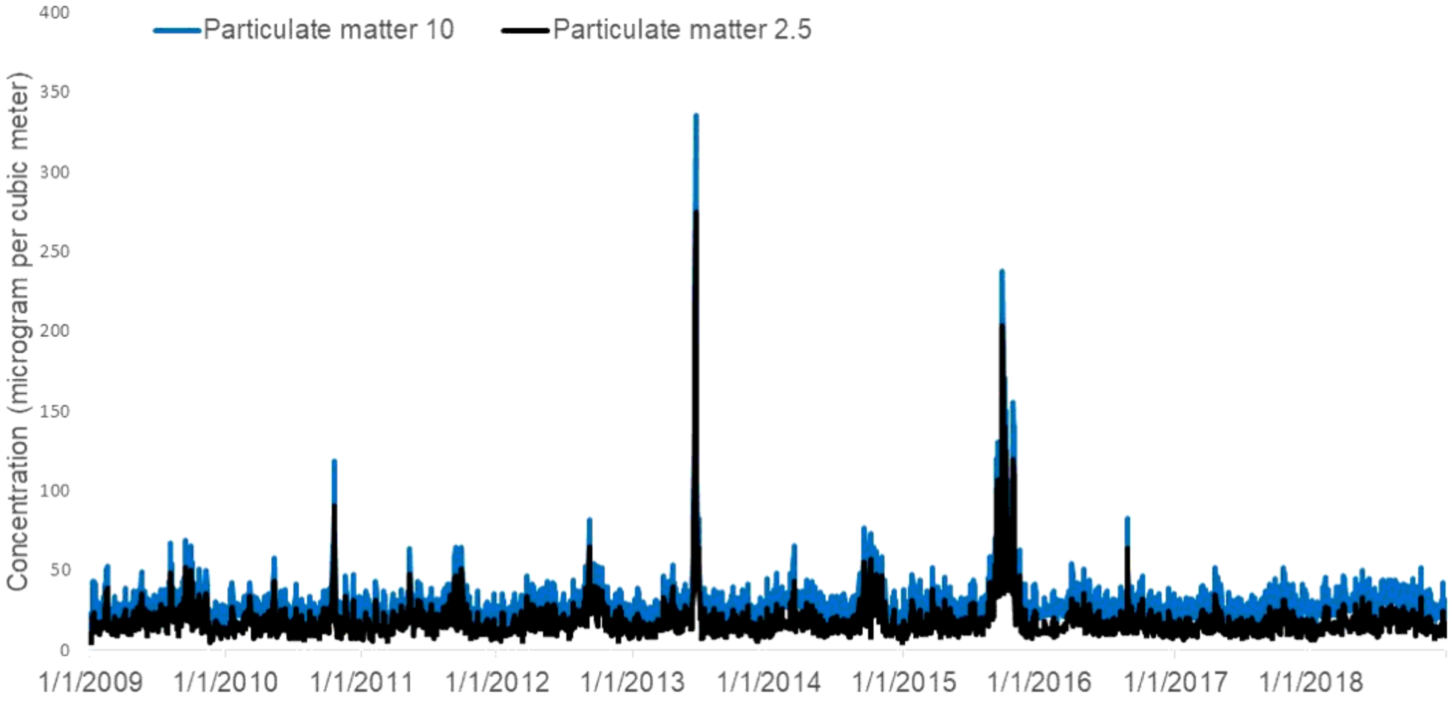

Figure 1

Daily distribution of PM2.5 and PM10 from 2009 to 2018

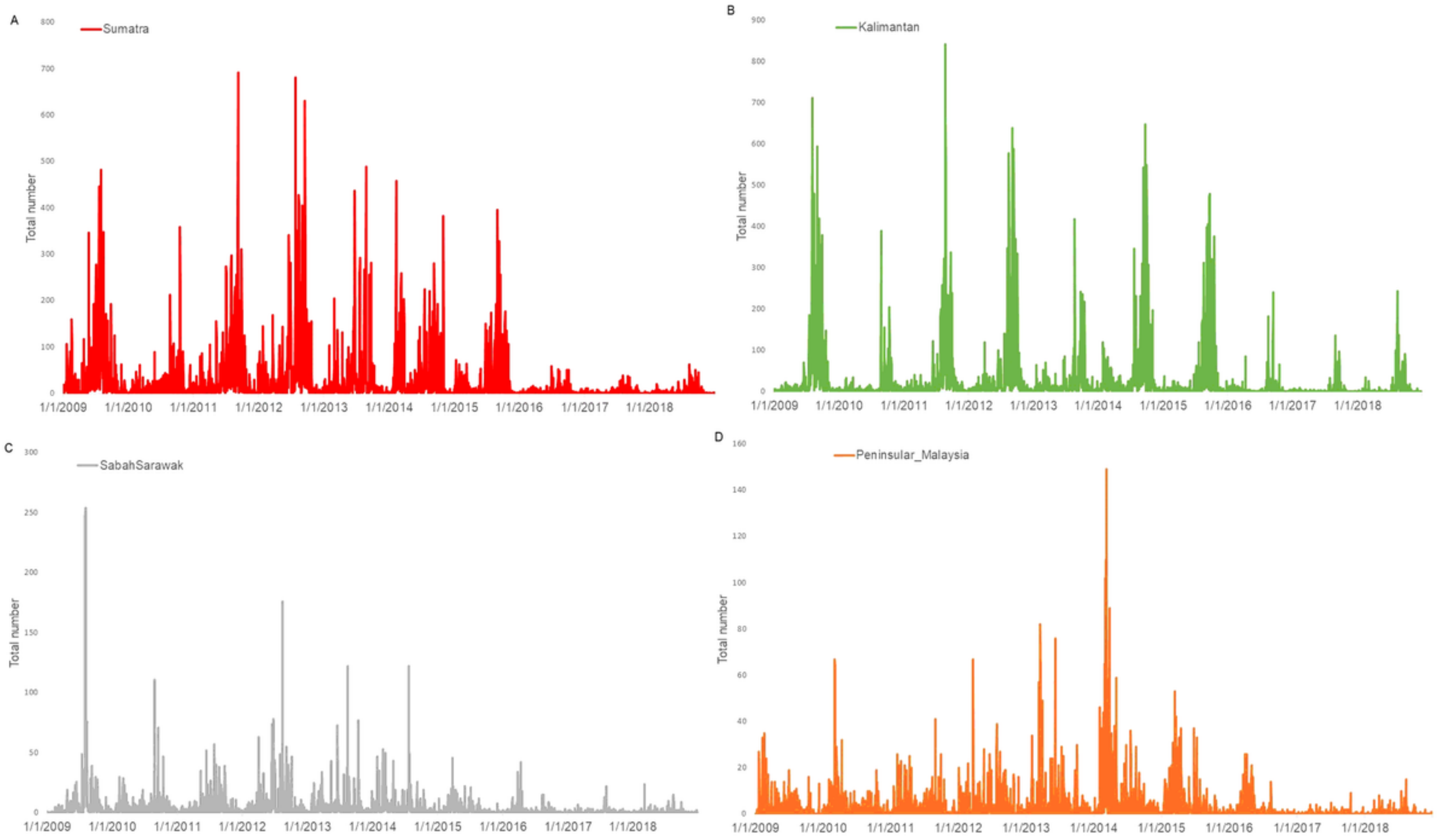

Figure 2

Daily distribution of forest fires hotspots counts (A) Sumatra (B) Kalimantan (C) Sabah/Sarawak (D) Peninsular Malaysia from 2009 to 2018

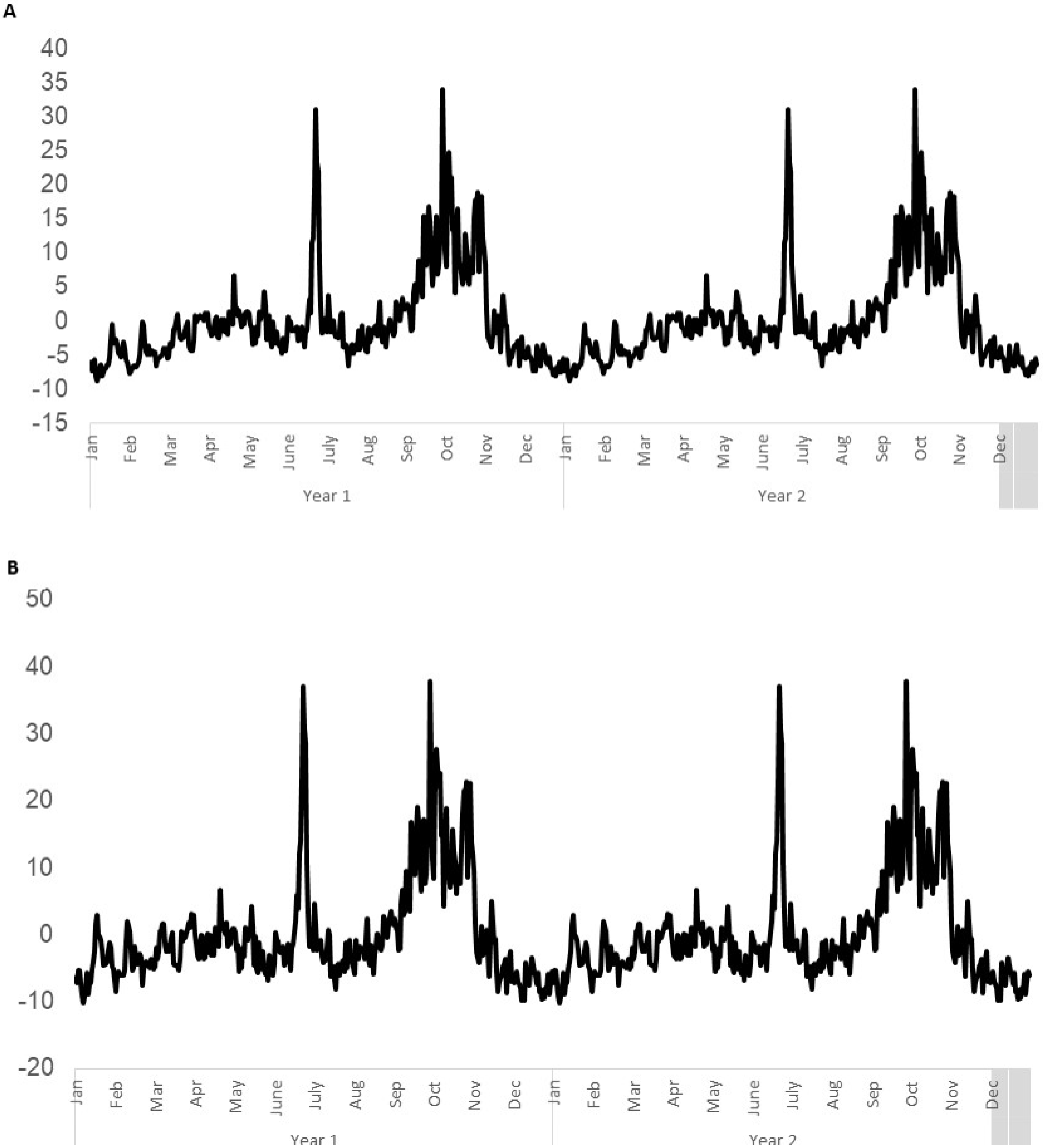

Figure 3

The seasonality of (A) PM2.5 and (B) PM10. The first two years are shown for easier visualization.

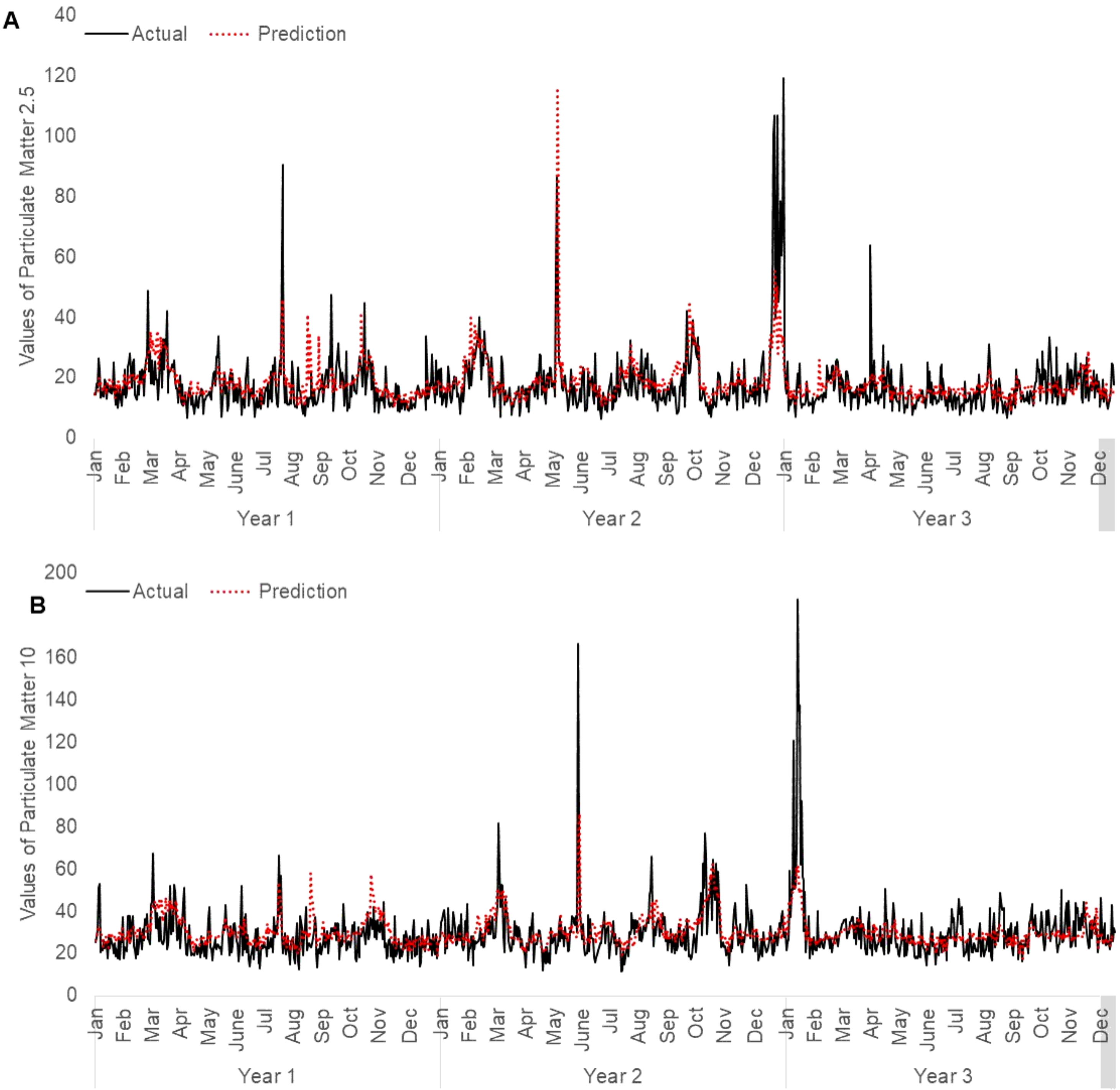

Figure 4

Comparison of actual and predicted air quality values using random forest model in Singapore: (A) PM2.5 and (B) PM10 Testing data (30%) is randomly selected from the dataset (2007-2018)

Loading [MathJax]/jax/output/CommonHTML/jax.js

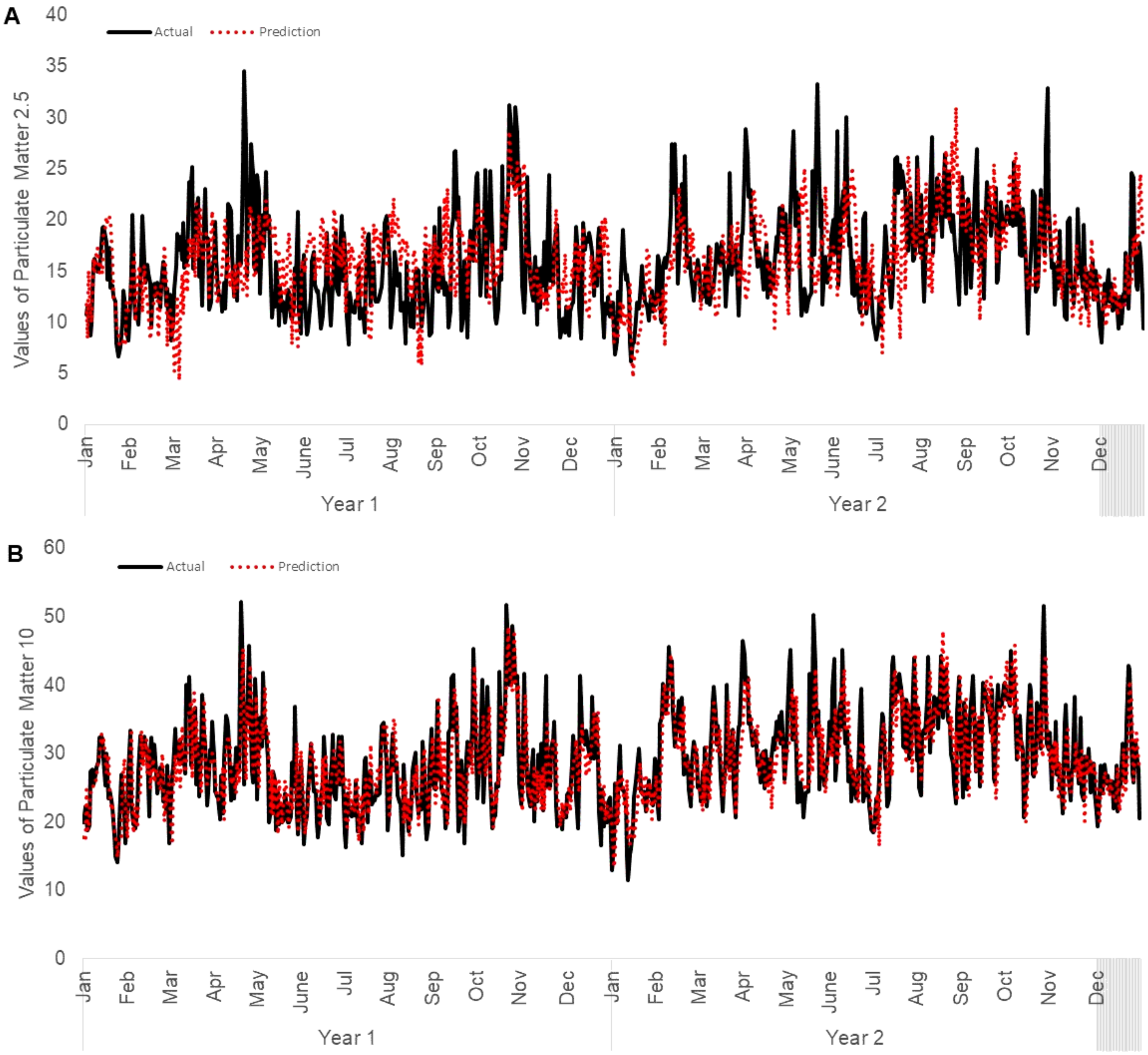

Figure 5

Comparison of actual and predicted air quality values using Vector Autoregressive model in Singapore: (A) PM2.5 and (B) PM10 Testing data is two years from 1st Jan 2017 to 31st Dec 2018.

Loading [MathJax]/jax/output/CommonHTML/jax.js