Thai Environmental Engineering Journal Vol. 35 No. 3 (2021) : 31-40
www.eeat.or.th

The Estimation of $PM_{2.5}$ Pollution Using Statistical Analysis and MERRA-2 Aerosol Reanalysis for Health Risk Assessment in Northern Thailand

Adchara Chinsorn and Seksan Papong*
Technology and Informatics Institute for Sustainability (TIIS), National Science and Technology Development Agency (NSTDA), Pathum Thani, Thailand
*E-mail : seksanp@mtec.or.th

Abstract

The landscape of northern Thailand consists of mountains, jungles, and valleys. Open burning, agricultural burning, and bushfires are the major sources of $PM_{2.5}$ (particles less than 2.5 micrometers in diameter) in the dry season that affect health via non-accidental mortality and morbidity. According to a report by the Geo-Informatics and Space Technology Development Agency (GISTDA), the MODIS satellite detected a fire hotspot of 9,859 points over the nine provinces of northern Thailand between January and May 2019. However, an estimation of $PM_{2.5}$ concentration over northern Thailand was limited due to the paucity of data. In this study, the method was developed to estimate the $PM_{2.5}$ concentration by applying a linear regression (MLR) of the $PM_{2.5}$ monthly data from the Pollution Control Department (PCD), MERRA-2 aerosol reanalysis, and meteorological factors such as temperature, relative humidity, wind speed, rainfall, and air pressure. In addition, the health risk was studied through relative risk (RR) using the risk function in SPSS to calculate the concentration-response coefficients (β values) between $PM_{2.5}$ concentration and non-accidental mortality and morbidity; namely chronic obstructive pulmonary disease (COPD), stroke, and ischemic heart disease (IHD). Finally, the concentration of $PM_{2.5}$ in 2019, over the nine provinces of northern Thailand, was 30.68 $\mu g/m^3$ while the coefficient of determination (R^2) was 0.90 and a root mean squared error (RMSE) was ±4.45. For the health risk, the results are shown that a 10 $\mu g/m^3$ $PM_{2.5}$ increase in northern Thailand was associated with an increase in the RR of mortality from COPD, stroke, and IHD about 20.9%, 24.3%, and 24.1%, respectively. In addition, increases in $PM_{2.5}$ concentration were also associated with the RR of morbidity on COPD, stroke, and IHD by 15.3%, 5.8%, and 11.5% per 10 $\mu g/m^3$, respectively. For the health burden, the results are shown that $PM_{2.5}$ contributed to mortality from COPD, stroke, and IHD accounting for 687, 1,818, and 1,095 cases, respectively. Moreover, that $PM_{2.5}$ caused 9,529, 1,080, and 3,916 cases of morbidity in COPD, stroke, and IHD, respectively. Thus, a decrease of $PM_{2.5}$ concentration in northern Thailand by 10 $\mu g/m^3$ could avoid 3,600 mortality and 14,525 morbidity cases.

Keywords : Health Risk; Mortality; Morbidity; $PM_{2.5}$; MERRA-2

Introduction

The problem of air pollution is a growing concern for the general public in northern Thailand because the landscape of northern Thailand consists of mountains, jungles, and valleys are subject to open burning, agricultural burning, and bushfires that are major sources of $PM_{2.5}$ (particles less than 2.5 micrometers in diameter) in the dry season. According to a report by the Department of Health and the Department of Disease Control [1], fine particulate matter can cause health effects through several systems, namely the respiratory system (coughing and acute lower respiratory infections), and the cardiovascular system (ischemic heart disease and cardiac arrhythmia). In addition, $PM_{2.5}$ can increase the risks of a stroke, and the risk rate will increase by with an increase in $PM_{2.5}$ concentration. Therefore, the announcement of the National Environment Committee [2] defined the 24-hour standard for PM2.5 at 50 $\mu g/m^3$, and the annual standard at 25 $\mu g/m^3$.

Although the related agencies tried to estimate $PM_{2.5}$ concentration, the estimation over northern Thailand was limited by the paucity of data. According to a report by the Pollutant Control Department [3], 16 air quality monitoring stations were installed that cover an area around sixteen sub-districts in the nine provinces of northern Thailand. Thus, the estimation of $PM_{2.5}$ concentration by several models was used for prediction in other areas. In a recent year, various models used the application of MODIS AOD to estimate $PM_{2.5}$ concentrations. However, the results have shown that an overall coefficient (R^2) that includes a root mean square error (RMSE) still provides high errors [4]. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2) can reconstruct major aerosol species of $PM_{2.5}$ and it may be mentioned that this expression is widely used for $PM_{2.5}$ estimates over Asia and Europe [5].

The empirical data is a necessary factor for sustainable long-term solutions for $PM_{2.5}$. The related agencies can the use these data for announcements intended for the general public about understanding $PM_{2.5}$ and preparing themselves to cope with it. In this study, the method was developed using multiple linear regression or MLR [6]. The $PM_{2.5}$ monthly data from MERRA-2 aerosol reanalysis and meteorological factors such as temperature, relative humidity, wind speed, rainfall, and air pressure are the major parameters for the estimation of $PM_{2.5}$ concentration. In addition, the monthly $PM_{2.5}$ data from The Pollutant Control Department was used as a training sampling. Finally, the health risk was studied through relative risk (RR) using risk function in SPSS to calculate concentration-response coefficients (β values) between $PM_{2.5}$ concentration and non-accidental mortality and morbidity, namely chronic obstructive pulmonary disease (COPD), stroke, and ischemic heart disease (IHD) were included in the health burden analysis for northern Thailand.

Material and Methods

The specific steps to estimate the $PM_{2.5}$ concentration and to quantify mortality, and morbidity directly attributable to $PM_{2.5}$ in northern Thailand are illustrated in Figure 1. It begins with monthly $PM_{2.5}$ data and meteorological data, followed by an interpretation of the $PM_{2.5}$ values, and ends with the health risk analysis.

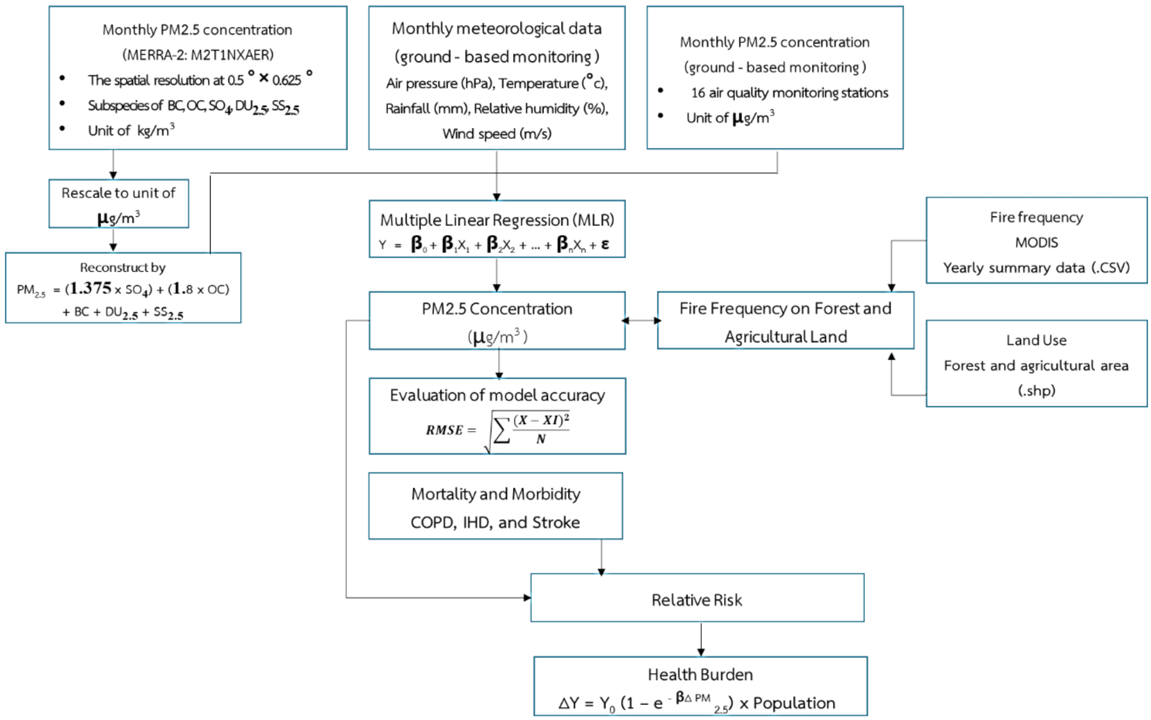


Figure 1 Workflow of the study

1. **$PM_{2.5}$ data and Meteorological data**

As seen in Figure 2 (A), fixed-site air monitoring stations are clustered in Lampang province due to the presence of a lignite-fired thermal power plant [7]. However, one or two air monitoring stations were also installed in each of the other provinces. Figure 2 (B) shows black carbon data, which is subspecies of $PM_{2.5}$ from MERRA-2, and the detect concentration coverage areas, which can solve the paucity of data from ground-based monitoring. Aerosol species are assumed to be external mixtures that do not interact with each other. Both dust and sea salt emissions depend on surface wind speed, while sulfate and carbonaceous species have emissions principally from fossil fuel combustion, biomass burning, and biofuel consumption, with additional biogenic sources of organic carbon. The MERRA-2 data which provides the spatial resolution at 0.5 ° x 0.66 ° grid is available from https://disc.gsfc.nasa.gov. The major aerosol species considered in MERRA-2 are SO_4, BC, $Dust_{2.5}$, $SS_{2.5}$, and OC. It is possible to reconstruct the $PM_{2.5}$ concentration using these subspecies [8]. However, all subspecies from MERRA-2 are detected in units of kg/m^3. Therefore, we have to rescale all subspecies to units of $\mu g/m^3$ which multiply by $1.0x10^9$. According to the study of characteristic species and bulk $PM_{2.5}$ mass using the interagency monitoring of protected visual environments (IMPROVE) (Hand et al., 2011), the following equation has been used to reconstruct the $PM_{2.5}$ concentration.

$$PM_{2.5} = (1.375 \times SO_4) + (1.8 \times OC) + BC + DU_{2.5} + SS_{2.5} \qquad (1)$$

Where, SO_4 is sulfate, OC is organic carbon, BC is black carbon, $DU_{2.5}$ is dust including aluminum (Al), silicon (Si), calcium (Ca), iron (Fe), and titanium (Ti), and $SS_{2.5}$ is sea salt particulate matter with a diameter of less than 2.5 which calculated with chloride ion data. The $PM_{2.5}$ concentration is reconstructed using the above equation and compared with the $PM_{2.5}$ data given by the ground-based monitoring of the PCD over 9 provinces, in 2019. It may be mentioned that this expression is widely used for $PM_{2.5}$ estimates over Asia and Europe.

Monthly mean meteorological data in 2019 were obtained from the Thai Meteorological Department (TMD). The meteorological parameters are air pressure, temperature, rainfall, relative humidity, and wind speed. The meteorological stations are clustered around northern Thailand which are shown in Figure 2 (A).

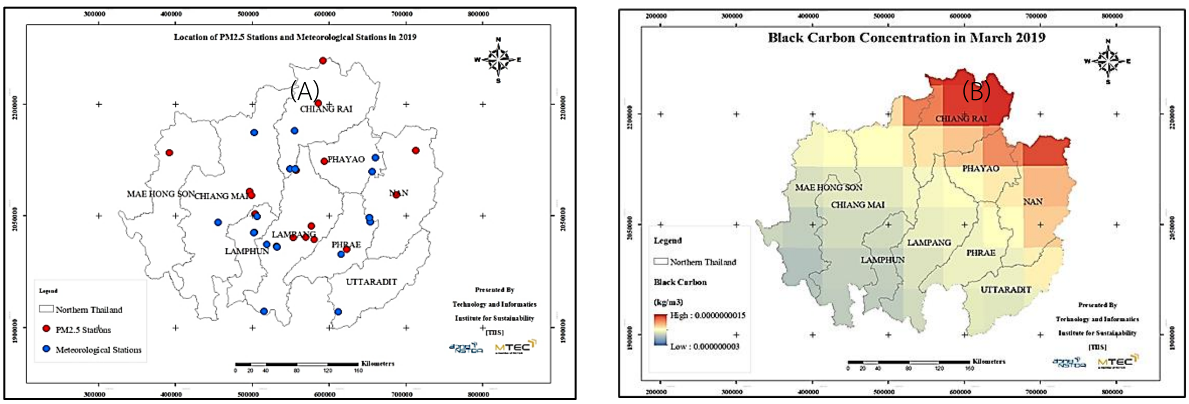


Figure 2 (A) Air quality monitoring stations and meteorological Stations in 2019

(B) PM2.5 data from MERRA-2

2. Multiple linear regression (MLR)

Multiple linear regression (MLR) model was to estimate the $PM_{2.5}$ concentration. First, the linear regression was used to determine the correlation between the $PM_{2.5}$ data from the PCD and the $PM_{2.5}$ data from MERRA-2. Second, we used meteorological parameters and $PM_{2.5}$ data from MERRA-2 for the estimation, using the multiple linear regression (MLR) model (95% confidence interval). For this study, the following equation has been used to estimate the $PM_{2.5}$ concentration [9].

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n + \varepsilon \qquad (2)$$

Where, Y is the prediction of $PM_{2.5}$ ($\mu g/m^3$) which calculate by a linear combination of a coefficient, β_0 is y-intercepts for $PM_{2.5}$ prediction, β_1-β_n are regression coefficients for the predictor variables, X_1-X_n are value of meteorological parameters namely the temperature (°c), wind speed (m/s), air pressure (hPa), relative humidity (%), and rainfall (mm.), and ε is the model's error term (also known as the residual).

3. Evaluation of model accuracy

The various statistical evaluators, namely a root mean square error or RMSE and the coefficient of determination (R^2), were applied to the monthly mean $PM_{2.5}$ concentration for evaluating the MERRA-2 data and meteorological data. For this study, the following equation was used to evaluate the model accuracy.

$$\mathrm{RMSE} = \sqrt{\sum \frac{(X - Xi)^2}{N}} \qquad (3)$$

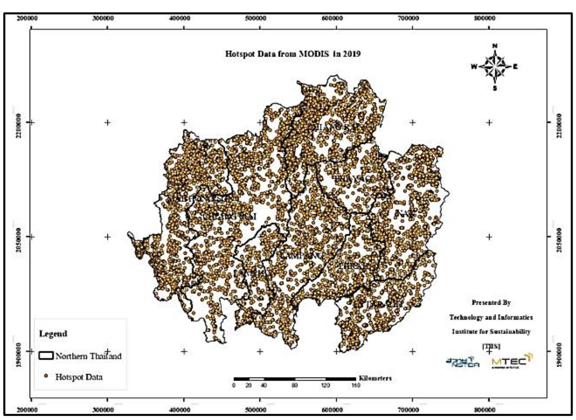
The RMSE is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line the data points are. In addition, RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Finally, RMSE is commonly used in climatology, forecasting, and regression analysis to verify experimental results [10]. For the Eq. (3), X is the forecasts (expected values or unknown results), X_i is the observed values (known results), and N is amount of data.

4. Fire Frequency

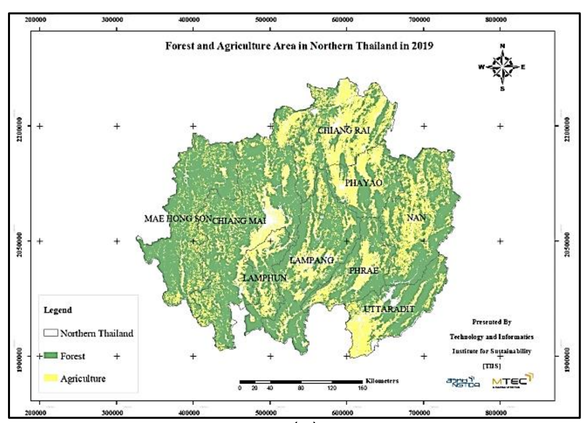
Fire frequency is one of the main components of the fire regime, together with the pattern and the intensity of open burning, agricultural burning, and bushfires [11]. For Thailand, fire hotspots are the major sources of $PM_{2.5}$ (particles less than 2.5 micrometers in diameter) in northern Thailand. The hotspot data are available from https://firms.modaps.eosdis.nasa.gov/download. For this study, yearly summary data of the MODIS instrument was used for the detection of fire frequency. In addition, land use data were collected from the Land Development Department (LDD). First, we classify hotspot data by month in each province. Second, the land use of forest, and agriculture was overlaid by hotspot data for study the relationship between open burning, agricultural burning, and bushfires and the subsequent $PM_{2.5}$ concentration in northern Thailand. Finally, the hotspot data (A) and land use data (B) are shown in Figure 3.

5. Health Burden Assessment

Individual mortality, and morbidity records include data on the location of death, and primary causes of death in 2019, were obtained from the Thailand Ministry of Public Health [12] for the entire northern Thailand area. There are shown that mortality by COPD, Stroke, and IHD was 1,716, 3,403, and 2,082 respectively. In addition, morbidity by COPD, Stroke, and IHD was 43,439, 32,964, and 31,115 respectively. The mortality and morbidity records are shown in Table 1.



(A)



(B)

Figure 3 (A) Hotspot data in 2019
(B) Land use data in 2019

Table 1 The mortality, and morbidity records by Stroke, COPD, and IHD in northern Thailand

Provinces	Mortality			Morbidity			Total
	COPD	Stroke	IHD	COPD	Stroke	IHD	
Chiangrai	295	543	293	7,663	6,888	5,195	20,877
Chiang Mai	395	816	481	12,478	8,106	7,142	29,418
Nan	269	262	211	4,128	2,152	1,658	8,680
Phayao	145	246	209	3,228	2,430	2,296	8,554
Phrae	143	350	172	3,074	3,026	2,728	9,493
Mae Hong Son	48	86	43	1,738	680	334	2,929
Lampang	253	577	322	6,013	5,402	7,251	19,818
Lamphun	106	226	179	2,782	1,993	1,500	6,786
Auttaradit	62	297	172	2,335	2,287	3,011	8,164
9 Provinces	1,716	3,403	2,082	43,439	32,964	31,115	

The Relative risk (RR) function in SPSS is used for comparisons in this case. The sampling was separated by two groups, mortality and morbidity of the $PM_{2.5}$ exposed group, and the mortality and morbidity of a group with no $PM_{2.5}$ exposure. The slope of the natural log of RR versus $PM_{2.5}$ is called β, and it is frequently used across different studies to compare the strength of the relative risk for a similar change in $PM_{2.5}$ exposure ($\Delta PM_{2.5}$). β can also be calculated from the ln (RR)/($\Delta PM_{2.5}$). In addition, Δ $PM_{2.5}$ of 10 µg/m^3 is often used. For this study, the following equation has been used to assess the health burden [13].

$$\Delta Y = Y_0 \left(1 - e^{-\beta \Delta PM2.5}\right) \times \text{Population} \qquad (4)$$

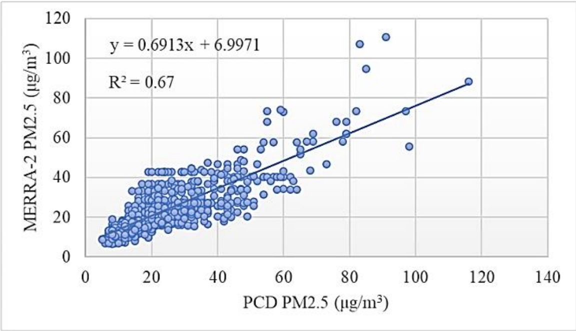
Where, ΔY is the change in incidence rate, Y_0 is the baseline incidence rate of the health effects, β is the C-R coefficient, pop is the exposed population, and Δ $PM_{2.5}$ is the change in $PM_{2.5}$ concentration to some target or health standard value.

Results

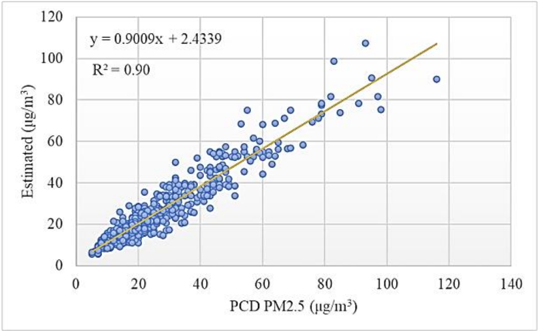
1. Monthly Average $PM_{2.5}$ Concentration

The evaluation model of $PM_{2.5}$ concentration was developed using the modern-era retrospective analysis for research and applications, version 2 (MERRA-2), and meteorological factors for the higher accuracy of the model. The coefficient of determination (R^2) between $PM_{2.5}$ concentration from ground measured data and MERRA-2, which was reconstructed from the major aerosol species SO_4, BC, $Dust_{2.5}$, $SS_{2.5}$, and OC, was 0.67. However, the application of meteorological factors namely temperature, air pressure, relative humidity, wind speed, and rainfall to the $PM_{2.5}$ data from MERRA-2 result in an increased coefficient of determination (R^2) of 0.90. In addition, a root mean square error (RMSE) between the $PM_{2.5}$ concentration by model and the ground measured data was ±4.45 µg/m^3. Scatterplots between the multiple linear regression (MLR) model fitting (cross-validation) and the ground-measured of $PM_{2.5}$ are shown in Figure 4.

Monthly average $PM_{2.5}$ concentrations by provinces are shown in Figure 5. The consideration of $PM_{2.5}$ concentration found that Chiangrai province contributed the highest $PM_{2.5}$ concentrations of 96 $\mu g/m^3$, and 90 $\mu g/m^3$ in March and April, respectively. In addition, a comparison of annual average $PM_{2.5}$ concentrations in nine provinces found that all provinces contributed $PM_{2.5}$ concentrations that were higher than the annual standard of Thailand at 25 $\mu g/m^3$. According to yearly summary data in section 2, the MODIS satellite detected 10,343 fire hotspots over nine provinces of northern Thailand in 2019, while 9,859 points occurred between January to May. Moreover, most fire hotspots occurred in Mach about 3,288 points, and 3,051 points in April respectively. Therefore, we cannot deny that open burning, agricultural burning, and bushfires are also major sources of $PM_{2.5}$ in the area. Finally, results of fire frequency on forest and agricultural land are presented in section 2.



(A) N = 623, RMSE = ±9.35 $\mu g/m^3$



(B) N = 623, RMSE = ±4.45 $\mu g/m^3$

Figure 4 (A) Cross-validation result of monthly average $PM_{2.5}$ by MERRA-2.
(B) Cross-validation result of monthly average $PM_{2.5}$ by the model

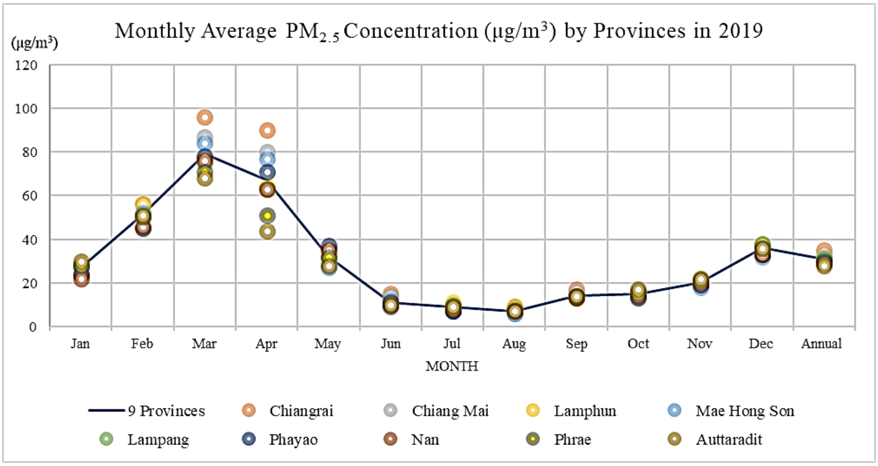


Figure 5 Monthly average $PM_{2.5}$ concentration by provinces

2. Fire Frequency on Forest and Agriculture

According to results in section 1, Chiangrai province contributed the highest $PM_{2.5}$ concentration along with the highest fire hotspots of 1,991 points in forest and agriculture. The second $PM_{2.5}$ concentration is Chiang Mai province where contributed fire hotspots about 1,593 points, and the third $PM_{2.5}$ concentration is Mae Hong Son where contributed fire hotspots about 1,555 points respectively.

Figure 6 shows that period of most fire hotspots in Chiangrai province occurred in March and April which relate to the highest $PM_{2.5}$ concentration. The noticeable results are the total number of hotspots that occurred twice as much in the forest as opposed to in agriculture. The Royal Thai Government mandated that farmers did not allow to burn agriculture residues between February to April in 2019 [14] and it is, therefore, possible to control agricultural burning in northern Thailand. However, brushfires are still the main contributor to $PM_{2.5}$ in northern Thailand that needs to be solved.

3. Mortality and Morbidity by $PM_{2.5}$

As seen in Table 2, relative risk (RR) was used to compare non-accidental deaths, and non-accidental illness by Stroke, COPD, and IHD. A 10 $\mu g/m^3$ $PM_{2.5}$ increase in northern Thailand was associated with an increase in the RR of mortality from COPD, stroke, and IHD of about 20.9%, 24.3%, and 24.1%, respectively. In addition, increases in $PM_{2.5}$ concentration were associated with the RR of morbidity on COPD, stroke, and IHD by 15.3%, 5.8%, and 11.5% per 10 $\mu g/m^3$, respectively. For the health burden, $PM_{2.5}$ contributed 687, 1,818, and 1,095 cases to mortality from COPD, stroke, and IHD, respectively. Moreover, that $PM_{2.5}$ was also caused morbidity in COPD, stroke, and IHD in about 9,529, 1,080, and 3,916 cases, respectively. Thus, a decrease of $PM_{2.5}$ concentration in northern Thailand by 10 $\mu g/m^3$ could avoid 3,600 mortality and 14,525 morbidity cases.

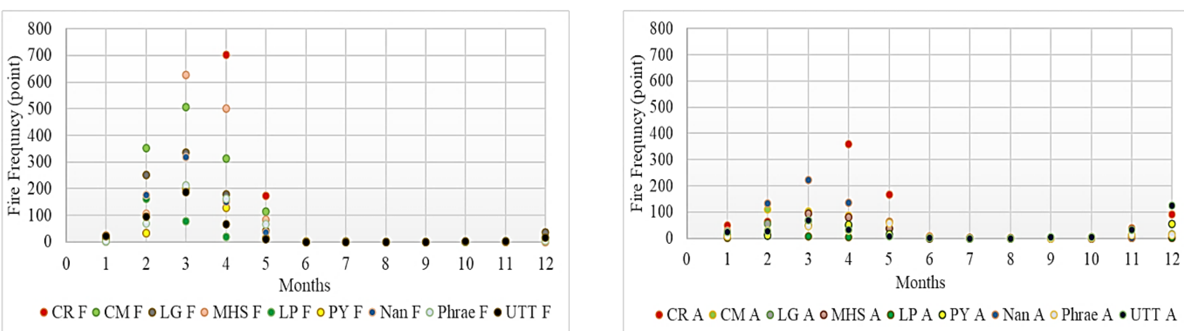


Figure 6 The fire frequency each province by months
Remark: F is forest, A is agriculture, CR. is Chiangrai, CM is Chiang Mai, LP is Lamphun, MHS is Mae Hong Son, LG is Lampang, PY is Phayao, and UTT is Auttaradit

Table 2 Relative risk, percentage increment, and concentration-response coefficients in COPD, Stroke, and IHD by 10 $\mu g/m^3$ $PM_{2.5}$ increase in northern Thailand

Health Effects	RR (significant 95%)			(% increment)			β Value (Concentration-response coefficients)		
	COPD	Stroke	IHD	COPD	Stroke	IHD	COPD	Stroke	IHD
Mortality	1.209	1.243	1.241	20.9	24.3	24.1	0.018979	0.021753	0.021592
Morbidity	1.153	1.058	1.115	15.3	5.8	11.5	0.014237	0.005638	0.010885

Discussion

There is much research that studies the estimation of $PM_{2.5}$ concentration in northern Thailand due to the fact that $PM_{2.5}$ pollution is the major cause of health problems for the general public for a long period of time. The MODIS AOD is widely used for $PM_{2.5}$ estimates over northern Thailand. However, the overall coefficient (R^2) still provides high levels of error. For example, the result of the coefficient of determination (R^2) between AOD from MODIS and hourly PM ($PM_{2.5}$ and PM_{10}) were 0.22 and 0.21 respectively in the study about the prediction of hourly particulate matter concentrations in Chiangmai, Thailand [15]. Thus, in this study, the methods were developed using the modern-era retrospective analysis for research and applications, version 2 (MERRA-2) to estimate $PM_{2.5}$ concentration. The major aerosol species measured in MERRA-2 as reconstructed in Eq. (1) appears to successfully estimate the $PM_{2.5}$ concentration. Moreover, the application of meteorological data in a multiple linear regression model accurately estimated $PM_{2.5}$ concentrations. The application of the MERRA-2 data along with meteorological data through MLR models can provide a reliable estimation of $PM_{2.5}$ concentrations. In addition, results of $PM_{2.5}$ concentration are used to assess relative the risk of COPD, Stroke, and IHD diseases by exposed group and unexposed group.

Conclusion

In this study, an innovative method was used to estimate the $PM_{2.5}$ concentrations. The monthly average $PM_{2.5}$ data from MERRA-2, which is constructed from SO_4, BC, $Dust_{2.5}$, $SS_{2.5}$, and OC was used for estimation of the $PM_{2.5}$ concentration coverage in all grid cells, in northern Thailand. In addition, the accuracy of results was improved by the addition of meteorological data such as air pressure, temperature, relative humidity, rainfall, and wind speed into the MLR model. Thus, the overall coefficient (R^2) and root mean square error (RMSE) were improved to 0.90, and ±4.45 $\mu g/m^3$ respectively. For the health burden study, we used methods which can construct concentration-response coefficients for the northern population. Further studies should focus on other factors, such as smoking, which also relate to diseases such as COPD, stroke, and IHD to improve the accuracy of the health burden results.

Acknowledgement

This study was financial supported by the National Science and Technology Infrastructure (NSTI), National Science and Technology Development Agency (NSTDA) (RD&E Fund. No.: P2050336). In addition, the monthly average $PM_{2.5}$ data, and meteorological data were supported by the Pollution Control Department (PCD), and the Thai Meteorological Department (TMD), respectively including the land use data was supported by the Land Development Department (LDD).

References

[1] The Ministry of Public Health (MOPH). 2015. Guideline for awareness of risk area from air pollution [Online]. Available at: http://www.oic.go.th/FILEWEB/CABINFOCENTER17/DRAWER002/GENERAL/DATA0000/00000200.PDF [Accessed on 28 July 2021].

[2] The announcement of the National Environment Committee No.36 B.E.2553. The Standard of $PM_{2.5}$ Concentration in Atmosphere. Thai Government Gazette volume 172 on 24 March B.E.2553.

[3] The Pollution Control Department (PCD). 2020. Air monitoring stations [Online]. Available at http://air4thai.pcd.go.th/webV2/download.php?grpIndex=1 [Accessed on 5 September 2020].

[4] Wang, Q. et al, 2019. Estimating PM2.5 Concentrations based on MODIS AOD and NAQPMS data over Beijing–Tianjin–Hebei 19(5): 1207.

[5] Hand, J.L. et al., 2011. Spatial and seasonal patterns and temporal variability of haze and its constituents in the United States: Report. Interagency Monitoring of Protected Visual Environments. in the United States: Implications for the sensitivity of PM2.5 to climate change.

[6] Tai, A.P.K. et al., 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. Atmospheric Environment 44(32): 3976-3984.

[7] Electricity Generating Authority of Thailand (EGAT). 2019. Maemoh [Online]. Available at: https://www.egat.co.th/index.php?option=com_content&view=article&id=2494&Itemid=117 [Accessed on 20 July 2021].

[8] Navinya, C.D., Vinoj, V., and Pandey, S.K. 2020. Evaluation of PM2.5 surface concentrations simulated by NASA's MERRA version 2 aerosol reanalysis over India and its relation to the air quality index. Aerosol and Air Quality Research 20(6): 1329-1339.

[9] Zhao, R. et al., 2018. Short period PM2.5 prediction based on multivariate linear regression model. PLOS ONE 10: 1317.

[10] Chai, T. and Draxler, R. R. 2014. Root mean square error (RMSE) or mean absolute error (MAE). Geoscientific Model Development Discussions 7: 1525-1534.

[11] Curt, T. 2018. Fire Frequency. Encyclopedia of Wildfires and Wildland-Urban Interface (WUI) Fires 10: 1007.

[12] The Ministry of Public Health (MOPH). 2020. Data of Non-Communicable Diseases. [Online]. Available at http://www.thaincd.com/2016/mission/documents.php?tid=32&gid=1-020 [Accessed on 27 July 2021].

[13] Nathaniel, R.F. et al., 2020. An assessment of annual mortality attributable to ambient PM2.5 in Bangkok, Thailand. International Journal of Environmental Research and Public Health 17(10): 7298.

[14] The Geo-Informatics and Space Technology Development Agency (GISTDA). 2019. Bushfire and smog situation over Thailand in 2019 [Online]. Available at http://fire.gistda.or.th/fire_ report/Fire_2562.pdf [Accessed on 15 January 2021].

[15] Kanabkaew, T. 2013. Prediction of hourly particulate matter concentrations in Chiangmai, Thailand using MODIS aerosol optical depth and ground-based meteorological data. Environment Asia 6(2): 65-70.