EREM 79/4
Journal of Environmental Research, Engineering and Management
Vol. 79 / No. 4 / 2023
pp. 18–28
DOI 10.5755/j01.erem.79.4.33913

Estimates of PM2.5 Concentration Based on Aerosol Optical Thickness Data Using Ensemble Learning with Support Vector Machine and Decision Tree

Received 2023/04 | Accepted after revisions 2023/09

https://doi.org/10.5755/j01.erem.79.4.33913

Estimates of PM2.5 Concentration Based on Aerosol Optical Thickness Data Using Ensemble Learning with Support Vector Machine and Decision Tree

Satith Sangpradid[1*], Theeraya Uttha[1], Ilada Aroonsri[2]

[1] Department of Geoinformatics, Faculty of Informatics, Mahasarakham University, Thailand

[2] Department of Business Digital, Faculty of Management Sciences, Valaya Alongkorn Rajabhat University Under the Royal Patronage, Thailand

***Corresponding author:** satith.s@msu.ac.th

Air pollution, particularly fine particulate matter with a diameter of 2.5 micrometers or less (PM2.5), is a significant public health concern in many regions worldwide, including the northeastern region of Thailand. This study investigates the correlation between PM2.5 concentrations and meteorological spatial datasets such as surface relative humidity (SRH), surface wind speed (SPD), visibility (Vis), surface temperature (ST), and aerosol optical thickness (AOT) in the region. GIS techniques and the inverse distance weighting technique were used to create spatial maps of the meteorological datasets and ground station PM2.5 measurements. Pearson correlation analysis was performed to examine the relationship between PM2.5 and the meteorological datasets. Decision tree and support vector machine (SVM) algorithms were employed to estimate PM2.5 concentrations based on the spatial datasets. The results showed that Vis and ST have a moderate positive linear relationship with PM2.5, while AOT has a moderate negative linear relationship. SRH and SPD have weak relationships with PM2.5. The decision tree and SVM algorithms demonstrated a strong positive correlation between estimated and measured PM2.5 concentrations. The study shows that machine learning algorithms can be effective tools for estimating PM2.5 concentration based on AOT data, and feature selection can improve model performance. Ensemble learning could be employed to further improve model performance, particularly in regions with high spatial variability. Overall, the study provides a promising approach for estimating PM2.5 concentration using machine learning algorithms and AOT data.

Keywords: aerosol optical thickness, decision tree, machine learning, PM2.5, support vector machine (SVM).

Introduction

Air pollution has various health effects on human health and the environment. Air pollution is one of the leading environmental risk factors global for population health. The World Health Organization (WHO) estimates that about 7 million premature deaths, mainly from non-communicable diseases, are the result of the combined effects of ambient and household air pollution (WHO, 2018, 2020). According to the WHO, there are about 1.3 million premature deaths in the South East Asia region (WHO, 2018). Moreover, in Thailand, air pollution is found to be very high between December to May each year. Also, according to the WHO estimate, about 33 000 premature deaths in Thailand are caused by air pollution (WHO, 2018). In 2021, the measurement of air pollution of the annual average of PM2.5 concentration in Thailand is reported to be between 12.7 $\mu g/m^3$ to 31.7 $\mu g/m^3$ higher than the air quality guideline recommended by the WHO (Farrow et al., 2021). The northern region of Thailand has the highest level of PM2.5 which is higher than 24 $\mu g/m^3$, followed by the northeastern region where between 20 $\mu g/m^3$ and 24 $\mu g/m^3$ has been reported (Farrow et al., 2021). The northeastern region of Thailand has been experiencing high levels of PM2.5 pollution in recent years, particularly during the dry season from December to May. The main sources of PM2.5 in this region are open burning of agricultural waste, vehicle emissions, and industrial activities. According to the data from the Thailand Pollution Control Department, the average PM2.5 levels in the northeastern region during the dry season can exceed the safe threshold of 50 $\mu g/m^3$ set by the WHO. In some areas, the levels can reach hazardous levels of over 200 $\mu g/m^3$ (Thailand Pollution Control Department, 2021).

In recent years, geographic information system (GIS) has emerged as a powerful tool for estimating PM2.5 levels by seamlessly integrating various spatial data sources. These sources include information on pollution data, meteorological variable, population density and data derived from satellite images. The integration of remote sensing and GIS has gained significant traction among researchers for PM2.5 estimation. One prominent technique involves the use of remote sensing-based aerosol optical depth (AOD) estimation, allowing for estimations at various spatial and temporal scales (Handschub et al., 2022; Gupta et al., 2021). Various techniques employ a range of remote sensing algorithms available for AOD estimation from satellite images, such as the dark target (DT) algorithm, the deep blue (DB) algorithm, and the high-resolution aerosol retrieval algorithm with a priori land surface (HARLS) algorithm (He et al., 2021). The PM2.5 concentration can be estimated by using GIS and AOD as independent variables, which take into account factors such as relative humidity, air temperature, wind speed, land use types, and population density. These variables are known to be effective predictors of PM2.5 and can greatly enhance the accuracy of the model's estimations (Hua et al., 2013).

Traditionally, one common approach is to use statistical models that relate AOD to PM2.5 concentration. These models are typically developed using ground-based measurements of PM2.5 and AOD at a limited number of locations, and then applied to estimate PM2.5 concentration at other locations where only AOD data are available. Several statistical models have been developed for this purpose, including linear regression models (Kumar et al., 2007; Xie et al., 2015; Phuengsamran and Lalitaporn, 2021; Khamkhantee and Premanoch, 2021), and geographically weighted regression (Hua et al., 2013; You et al., 2016; Gou et al., 2017; Xu and Zhange, 2020). The linear regression approach is a well-established method that has been proven to be effective in providing accurate PM2.5 predictions using AOD data. This method is also relatively easy to implement in real-time applications. However, the lack of real-time measurements of planetary boundary layer (PBL) height and relative humidity (RH) profiles can limit the usefulness of nonlinear correlation models between AOD and surface PM2.5 concentration (Weber et al., 2010).

Recently, machine learning algorithms have been applied in environmental modeling to estimate and predict various environmental parameters. For instance, machine learning methods have been widely used in satellite imagery analysis for land use classification, monitoring, change detection, and prediction (Aroonsri and Sangpradid, 2021). Machine learning has been applied to classify land use patterns by analyzing satellite imagery and other geospatial data. Artificial neural network (ANN) machines have been used to classify land cover and predict changes in land use over time (Aroonsri and Sangpradid, 2021). In addition, machine learning algorithms have been used to predict air quality levels by analyzing data from air quality monitoring stations, meteorological data, and satellite imagery. Machine learning algorithms are increasingly being used to estimate PM2.5 concentration from AOD or aerosol optical thickness (AOT) data. Some commonly used machine learning algorithms for estimated PM2.5 concentration such as decision tree (Vignesh et al., 2023; Letchumanan and Palanichamy, 2022), random forest (Vignesh et al., 2023; Grange, 2018), multi-layer perceptron (Gupta and Christopher, 2009), and support vector machine (Vignesh et al., 2023; Letchumanan and Palanichamy, 2022; Masood and Ahmad, 2020) can be utilized for estimating PM2.5 concentration. These algorithms are designed to analyze data and make predictions based on patterns they find within that data. By training the algorithms with relevant datasets, they can learn to predict PM2.5 concentrations with increasing accuracy (Vignesh et al., 2023).

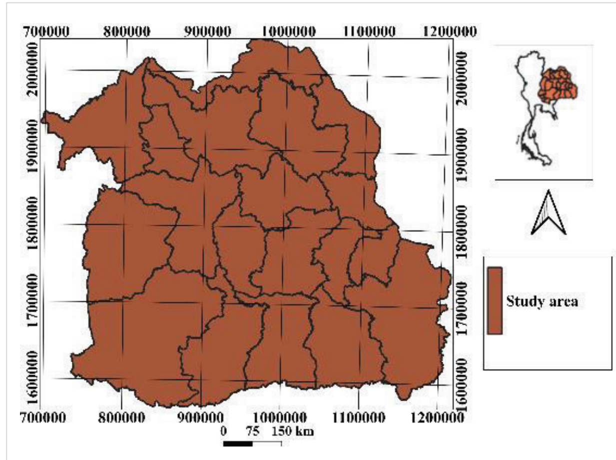
This paper proposes an ensemble learning approach that combines the strengths of support vector machines (SVM) and decision trees to estimate PM2.5 concentration in the northeastern region of Thailand using AOT data, along with the data from quality monitoring stations, meteorological data, and satellite imagery. PM2.5 is a major air pollutant that has adverse effects on human health, and accurately estimating its concentration is crucial for air quality management. Our approach uses a diverse set of data sources to train multiple models, which are then combined to improve the accuracy and robustness of the final estimation.

Data and Methods

Study area

The northeastern region of Thailand lies between 14.25° to 18.58° N Latitude and 100.97° to 105.55° E Longitude, as shown in *Fig. 1*. The study area is a large region that covers about one-third of Thailand's total land area. It is located in the northeastern part of the country and shares borders with Laos and Cambodia. The physical geography of the region is diverse, with a mix of highlands, plateaus, and river basins. The study area is known for its agricultural activities, and the burning of crop residues is a common practice. This has led to an increase in air pollution, particularly PM2.5, which can have serious health implications.

Fig. 1. *Location of the northeastern region of Thailand Data collection*



Data collection

MOD04 is a remote sensing data product that is part of the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument onboard NASA's Terra and Aqua satellites. The MOD04 product provides information on the amount of aerosols in the atmosphere, which can impact air quality and climate. Aerosol optical thickness (AOT) is a measure of the amount of light scattered and absorbed by aerosol particles in the atmosphere. AOT is typically reported as a dimensionless number, with higher values indicating more aerosols in the atmosphere. The MOD04 product provides AOT data on a global scale, with a spatial resolution of 10 kilometers.

The meteorological dataset used to estimate PM2.5 concentrations includes several variables: surface relative humidity (SRH), surface wind speed (SPD), horizontal visibility (Vis), surface air temperature (ST), and rainfall. The corresponding ground-based PM2.5 measurements were provided by the Pollution Control Department in Thailand. The spatial datasets for each variable were produced in the form of raster maps, a process facilitated through the application of geographic information system (GIS) techniques (Sangpradid, 2023). These variables and measurements are essential in proposing an ensemble learning approach for estimating PM2.5 concentrations accurately, by incorporating these variables and measurements into the approach.

Methods

Decision tree is a type of a predictive model that starts at the root of a tree to determine a target class for a given record. To create the tree structure, the algorithm uses the mean squared error (MSE) to split nodes into two or more sub-nodes. The tree is composed of three types of nodes: the root node, which is the initial starting point that can be split into further branches; intermediate nodes, which are created by recursively splitting the root node; and terminal nodes, also known as leaf nodes, which represent the final predictions or outcomes of the model (Dereswamy et al., 2020). Decision tree is a popular machine learning algorithm used to estimate PM2.5 concentration from aerosol optical depth (AOD) data. Decision trees are non-parametric models that recursively split the data based on the values of the predictor variables until a stopping criterion is met. The output is a tree-like model that can be used to predict PM2.5 concentration based on the values of the AOD data. One of the advantages of decision trees is that they can handle both categorical and continuous variables, which makes them well suited for modeling PM2.5 concentration from AOD data, often including a combination of continuous and categorical variables (Vignesh et al., 2023; Letchumanan and Palanichamy, 2022). Additionally, decision trees are relatively easy to interpret and can help identify the most important variables that are associated with PM2.5 concentration.

The support vector machine (SVM) algorithm is widely recognized as one of the most effective classification algorithms available. It operates by identifying the largest margin, or distance, between the points in a given dataset (known as support vectors) and the separation hyperplane. By maximizing this margin, SVM is able to accurately classify new data points based on their proximity to the hyperplane (Zhang et al., 2021; Debnath and Takahashi, 2010). SVM is a powerful supervised learning algorithm that has been successfully applied in several environmental studies. The use of SVM in PM2.5 estimation from AOD data is gaining popularity due to its ability to handle high-dimensional data and its good performance in predicting PM2.5 concentrations (Vignesh et al., 2023; Letchumanan and Palanichamy, 2022; Masood and Ahmad, 2020). The review provides insights into the effectiveness of SVM in PM2.5 estimation from AOD data and its potential for improving air quality monitoring and management.

Proposed method

An ensemble method in machine learning combines multiple individual models to produce a more accurate and robust prediction than any single model. It involves constructing a set of classifiers using a given algorithm, and each new example is classified by combining the predictions of every classifier in the ensemble. The predictions can be combined using methods such as averaging (for regression tasks) or majority voting (for classification tasks) (Madjarov et al., 2010). By combining the predictions of multiple models, ensemble methods can help reduce the impact of individual model weaknesses and increase the overall accuracy and robustness of the prediction. This paper proposes using ensemble learning with SVM and decision trees to estimate PM2.5 concentration based on AOT data, combined with a meteorological dataset using geographic information system. The proposed approach aims to improve the accuracy and robustness of PM2.5 concentration estimates by combining the strengths of both SVM and decision trees. The geographic information system is used to incorporate spatial information into the model, while the meteorological dataset provides additional features to enhance the prediction accuracy.

Evaluation

The proposed methods' models were evaluated for their performance using the mean absolute error (MAE), root mean square error (RMSE), and determination coefficient (R2) (Sangpradid et al., 2022). The calculation formulas are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (y_{is} - y_{it})^2}{N}} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^{N} |y_{is} - y_{it}|}{N} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_{is} - y_{it})^2}{\sum_{i=1}^{N} (y_{is} - \bar{y}_{is})^2} \quad (3)$$

where y_{is} is the estimate of PM2.5 concentration of the *i* value; y_{it} is the observed PM2.5 measurement of the *i* station; y_{is} is the average estimate of PM2.5 concentration for all the samples; and *N* is the size of the samples.

Results and Discussion

Results of correlation between the PM2.5 measurement and meteorological spatial datasets

To create a spatial map, GIS techniques were used in combination with meteorological spatial datasets such as SRH, SPD, Vis, and ST. The inverse distance weighting (IDW) technique was used for the spatial interpolation of meteorological datasets. The results of the SRH and SPD maps are shown in *Fig. 2(a)* and *(b)*, respectively, while the Vis and ST maps are shown in *Fig. 3(a)* and *(b)*. Additionally, the AOT datasets were examined to determine the surface PM2.5 concentration in the northeastern region of Thailand. The AOT data were processed to adjust the geometric data and coordinate system to match the meteorological spatial datasets. *Fig. 4(a)* shows the results of the AOT map. Ground station data on PM2.5 concentrations were also used to create a spatial map, which is presented in *Fig. 4(b)*.

The spatial distribution of SRH exposure in the central area, as shown in *Fig. 2(a)*, indicates that certain areas have higher SRH values than others, and this may be related to PM2.5 pollution. The visual interpretation of this distribution suggests that there may be a correlation

Fig. 2. *Spatial mapped of (a) SRH (b) SPD*

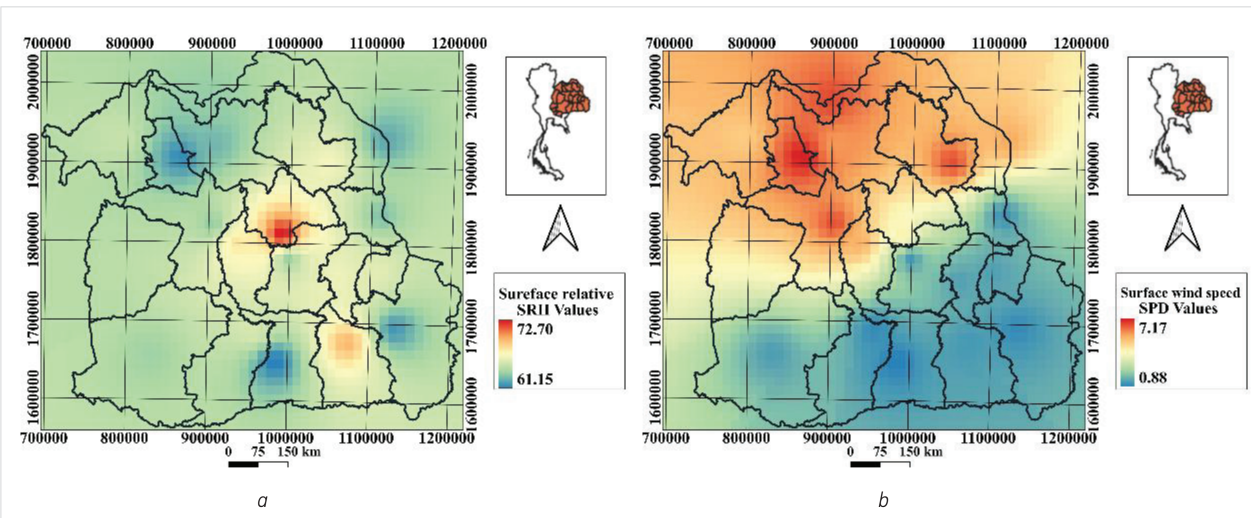


Fig. 3. *Spatial mapped of (a) Vis (b) ST*

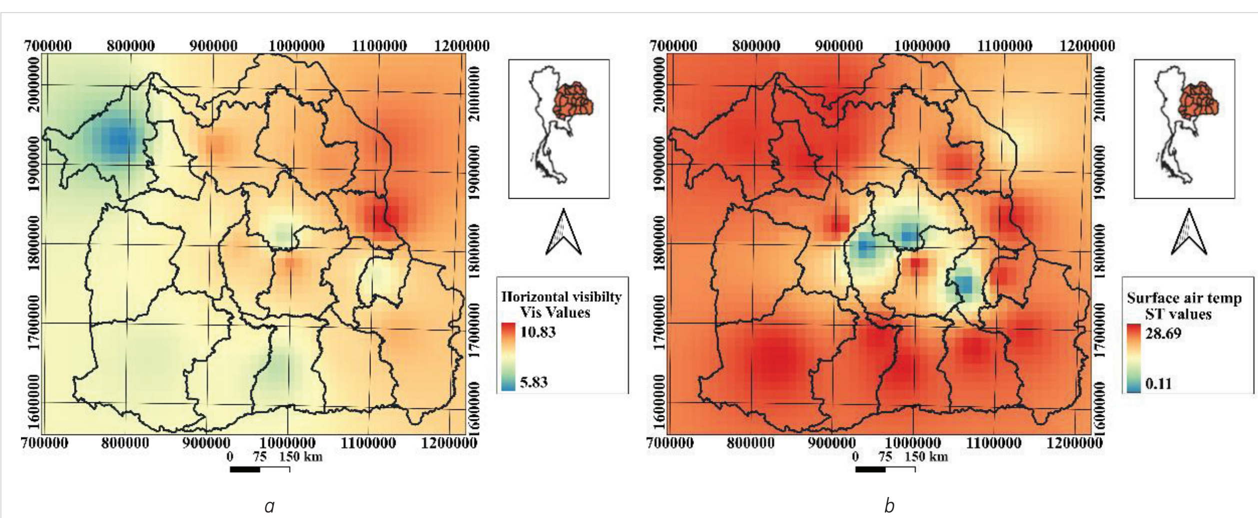
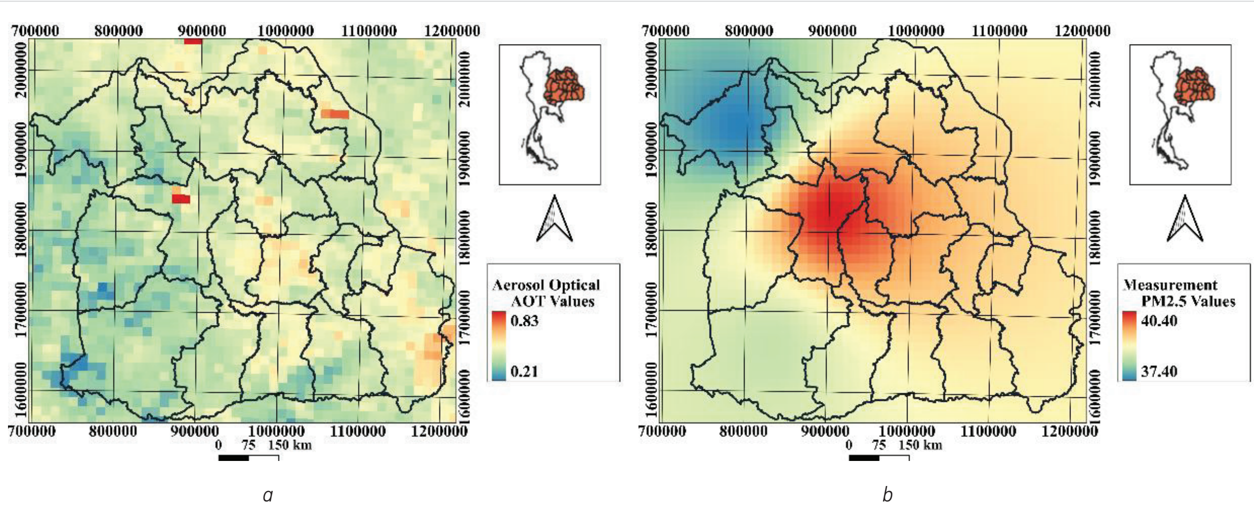


Fig. 4. *Spatial map of (a) AOT (b) PM2.5 measurement*



between PM2.5 levels and SRH exposure in the study area. The spatial distribution of SPD exposure in the north area was found to have higher values than in the south area, as shown in *Fig. 2(b)* and *4(b)*. The relationship between SPD and PM2.5 in the northwest area suggests that higher surface wind speeds may help reduce PM2.5 concentrations in this region. The spatial distribution of Vis exposure in the northeast and east area was found to have higher values than in the south and west area, as shown in *Fig. 3(a)* and *4(b)*. The relationship between Vis and PM2.5 in the central area suggests that the higher levels of PM2.5 can reduce visibility by scattering and absorbing light in the atmosphere. The spatial distribution of ST exposure in the central area was found to have lower values compared with other areas, as shown in *Fig. 3(b)* and *4(b)*. The relationship between ST and PM2.5 in the central area suggests that the higher levels of PM2.5 can reduce ST values.

The correlation between ground station PM2.5 measurements and satellite data such as SRH, SPD, Vis, ST, and AOT was examined. *Fig. 5(a)* and *(b)* display the results of the correlation between PM2.5 and SRH and SPD, respectively. Meanwhile, *Fig. 6(a)* and *(b)* depict the correlation between PM2.5 and Vis and ST, respectively. Additionally, *Fig. 7* shows the correlation between PM2.5 and AOT. In each figure, the x-axis represents the ground station PM2.5 measurements.

Fig. 5(a) illustrates a gradual linear pattern along the x-axis in the correlation between PM2.5 and SRH, with a Pearson correlation value of −0.053. Similarly, *Fig. 5(b)*

Fig. 5. *The correlation between PM2.5 measurement with the (a) SRH and (b) SPD*

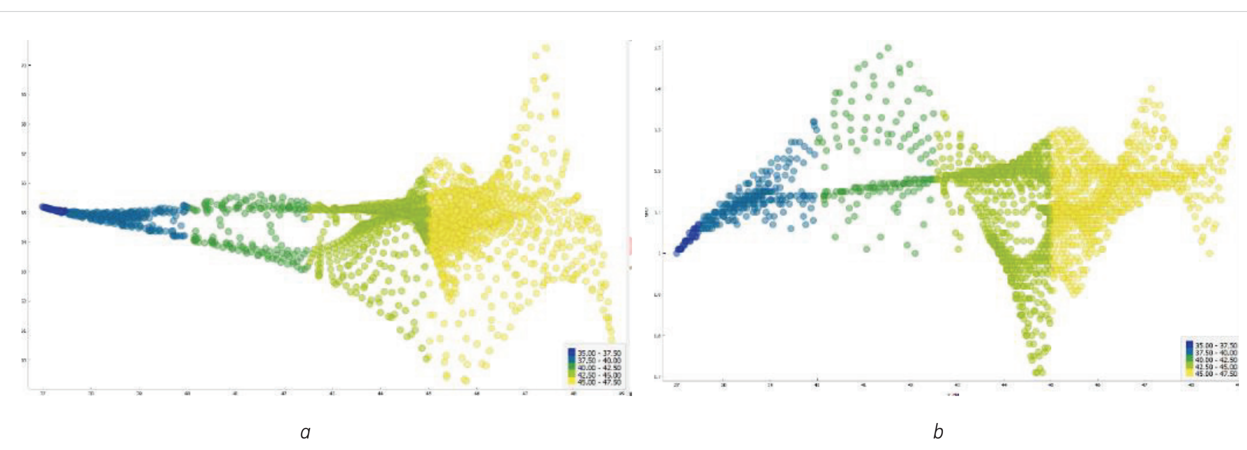
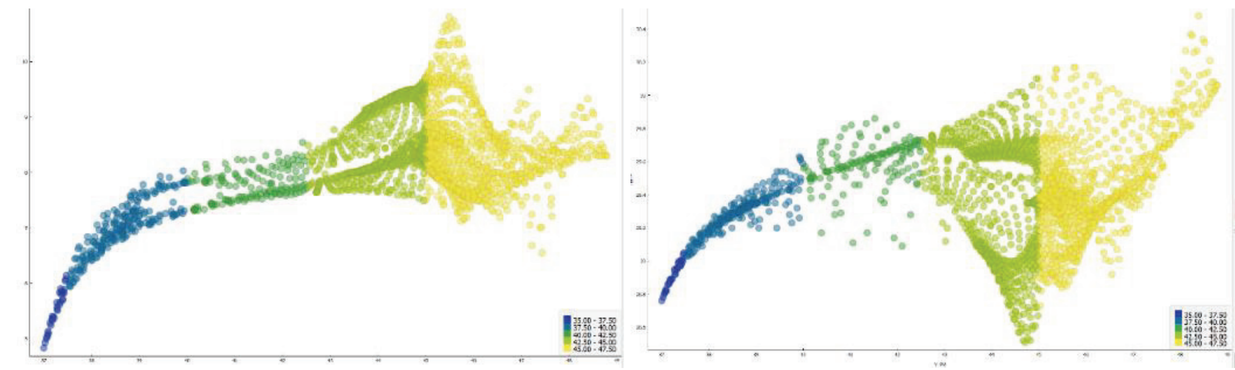
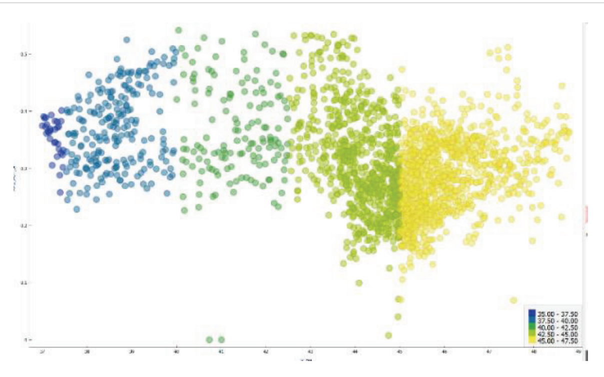


Fig. 6. *The correlation between PM2.5 measurement with the (a) Vis and (b) ST*



a *b*

Fig. 7. *The correlation between PM2.5 measurement with the AOT*



shows a gradual linear pattern along the x-axis in the correlation between PM2.5 and SPD, with a Pearson correlation value of 0.03. In *Fig. 6(a)*, the relationship between PM2.5 and Vis, as well as ST, shows a positive trend. The Pearson correlation values for Vis and ST are 0.56 and 0.21, respectively. Finally, *Fig. 7* displays the distribution of the relationship between PM2.5 and AOT. The Pearson correlation value is –0.33.

The results of the Pearson correlation analysis between SRH and SPD with PM2.5 revealed a weak relationship, indicating that the values of SRH or SPD do not have a significant effect on PM2.5 levels. However, the Pearson correlation values of Vis and ST showed a moderate positive linear relationship with PM2.5, indicating that they have a significant effect on PM2.5 levels. In contrast, the Pearson correlation values of AOT showed a moderate negative linear relationship with PM2.5, suggesting that they also have a significant effect on PM2.5 levels, but in the opposite direction.

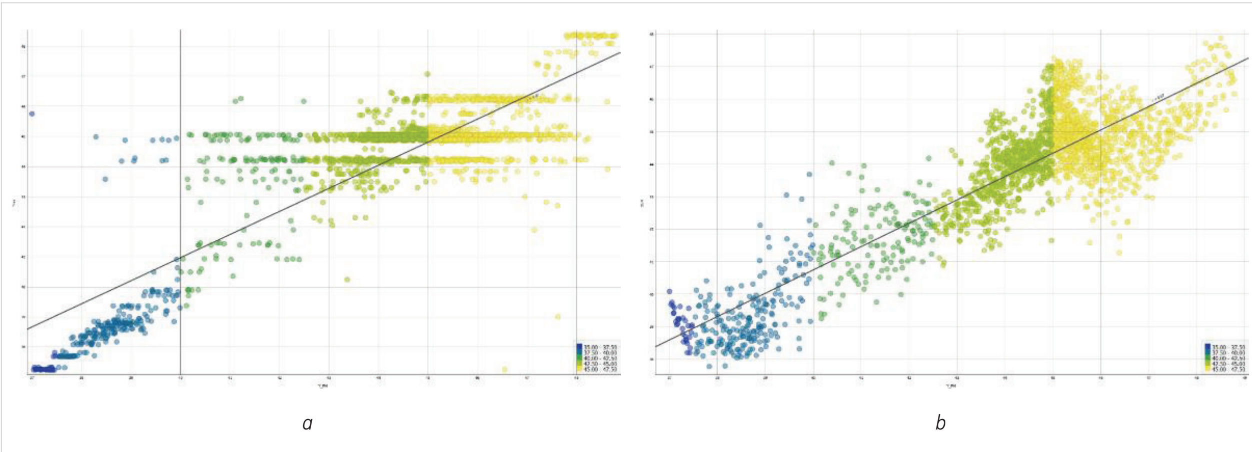
The results of estimates of PM2.5 concentration using ensemble learning with support vector machine and decision trees

The decision tree learning algorithm was utilized to estimate PM2.5 concentration based on spatial datasets of SRH, SPD, Vis, ST, and AOT. The algorithm was trained on a training dataset and used to construct a decision tree by recursively splitting the data into nodes based on selected features, with the goal of maximizing information gain at each step. The performance of the decision tree was evaluated using 10-fold cross-validation, where the data were repeatedly split into 10 training and testing sets. The experimental results demonstrated a strong positive correlation between the estimated PM2.5 concentration obtained from the decision tree learning algorithm and the PM2.5 measurements, with a Pearson correlation value of 0.86, as depicted in *Fig. 8(a)*.

In this study, the SVM learning approach was employed to classify the attributes of spatial datasets that include SRH, SPD, Vis, ST, and AOT, by separating them with a hyperplane. A radial basis function (RBF) kernel was utilized to map the input data into a higher-dimensional feature space. The experimental results exhibited a strong positive correlation between the estimated PM2.5 concentration obtained from the SVM learning algorithm and the PM2.5 measurements, with a Pearson correlation value of 0.87, as shown in *Fig. 8(b)*.

An ensemble learning approach with SVM and the decision tree was performed on the spatial datasets, including SRH, SPD, Vis, ST, and AOT, to estimate the PM2.5 concentration. The ensemble learning model exhibited the closest correlation to both SVM and decision tree

Fig. 8. *Scatter plot result of (a) the decision tree and (b) SVM with the PM2.5 measurements*



models with a Pearson correlation value of 0.87. The scatter plots of ensemble learning exhibited similarities to the decision tree results, as shown in *Fig. 9*.

The experimental results of spatial distribution map of PM2.5 concentration revealed by the decision tree (DT), SVM and ensemble learning are shown in *Fig. 10(a)*, *(b)*, and *Fig. 11(a)*, respectively. The comparison of the Pearson correlation of the PM2.5 measurement and the estimated models is shown in *Table 1*. The results of a spatial distribution map of the estimated model show that the PM2.5 concentration was found to be a higher level of PM2.5 in the central area compared with other areas. Additionally, it is noteworthy that a moderate level of PM2.5 concentration coincided with regions that aligned with the meteorological dataset of ST values. This observation

Fig. 9. *Scatter plot results of the ensemble learning algorithm and the PM2.5 measurements*

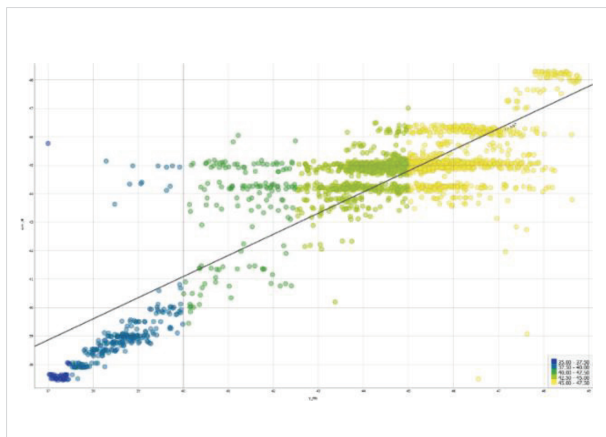
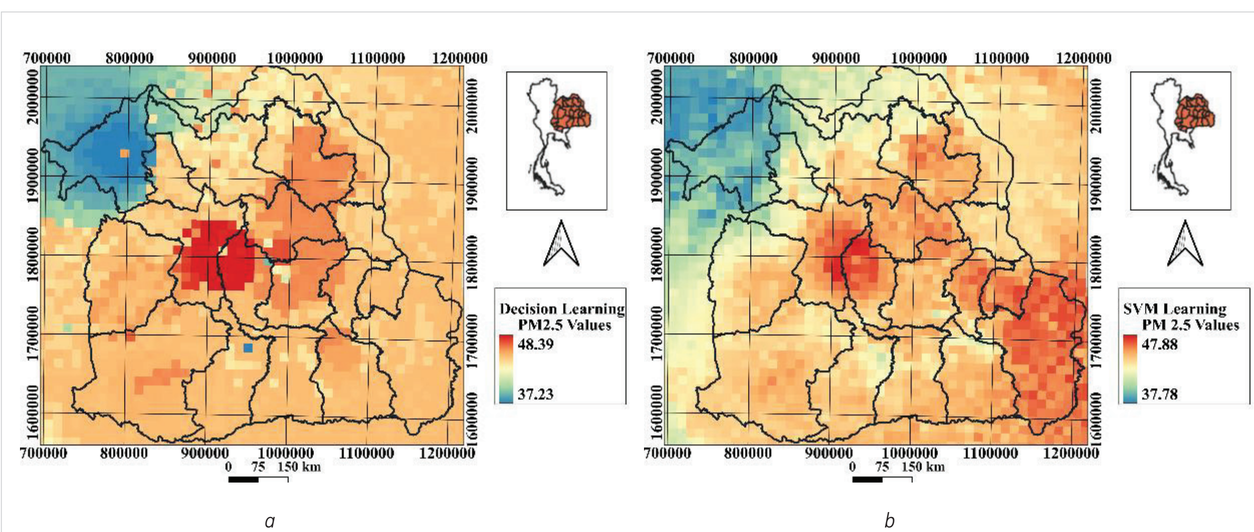


Fig. 10. *Spatial distribution of the estimated PM2.5 concentrations by the (a) DT and (b) SVM*



suggests a potential correlation between PM2.5 levels and ST values in those specific area, implying that the variation in surface temperature may influence PM2.5 concentration. *Fig. 10(a)*, *(b)*, and *Fig. 11(a)* show that the area of the moderate level of PM2.5 concentration is greater than the PM2.5 measurement area. According to the correlation of the meteorological datasets and the PM2.5 measurement, a Pearson correlation of the ST values of estimated models was higher than PM2.5 measurement in *Fig. 11(a)*.

Table 2 shows the performance of the three models based on these three metrics. The DT model has the lowest RMSE and MAE values, which means that it has the best predictive performance based on these metrics. However, the SVM_DT model has a slightly higher RMSE and MAE values compared with DT, but it has the same R2 value as DT. The SVM model has the highest RMSE and MAE values, indicating poorer performance compared with the other two models.

Fig. 11. *Spatial distribution of the estimated PM2.5 concentrations by (a) the ensemble learning and (b) the PM2.5 measurement*

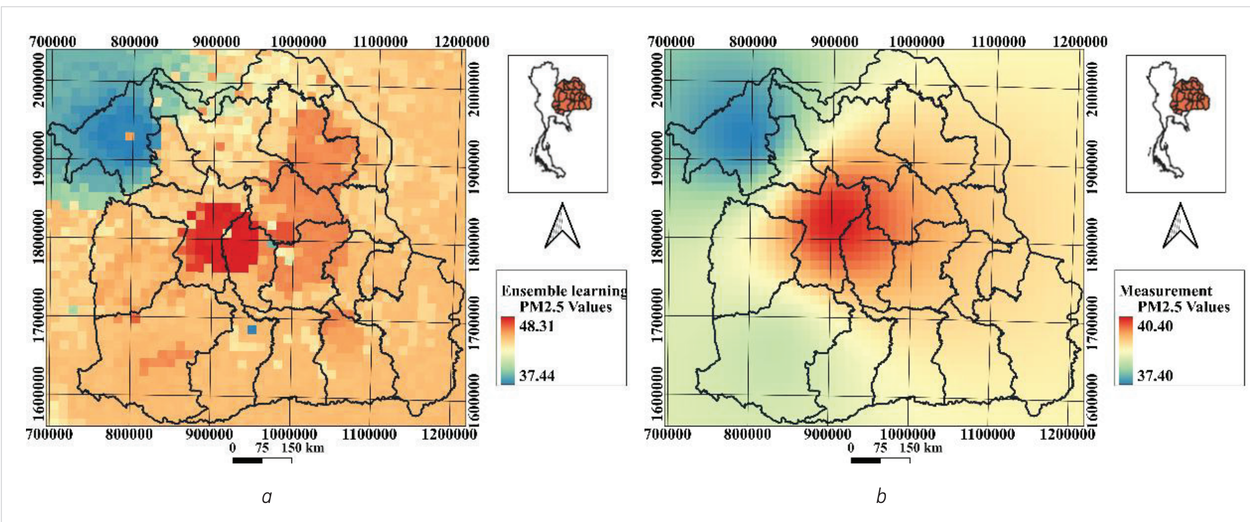


Table 1. *Comparison of the Pearson correlation of the PM2.5 measurement and the estimated models*

Variables	Measurement	DT	SVM	SVM_DT
	Pearson correlation values			
SRH	−0.053	−0.075	−0.054	−0.069
SPD	0.027	0.057	−0.028	0.057
Vis	0.565	0.596	0.603	0.590
ST	0.027	0.249	0.135	0.245
AOT	−0.33	−0.345	−0.445	−0.349

Table 2. *Comparison of the evaluation of the performance estimated models*

Models	RMSE	MAE	R^2
DT	1.295	0.865	0.750
SVM	1.336	1.048	0.734
SVM_DT	1.293	0.869	0.751

Conclusions

This study investigated the correlation between PM2.5 concentrations and meteorological spatial datasets such as SRH, SPD, Vis, ST, and AOT in the northeastern region of Thailand, using GIS techniques and the inverse distance weighting technique to create spatial maps of the meteorological datasets and ground station PM2.5 measurements. Pearson correlation analysis was performed to examine the relationship between PM2.5 and the meteorological datasets. The decision tree learning algorithm and SVM learning approach were employed to estimate PM2.5 concentration based on the spatial datasets.

The results showed that there was a moderate positive linear relationship between Vis and ST and PM2.5, while AOT showed a moderate negative linear relationship with PM2.5. The Pearson correlation values of SRH and SPD revealed a weak relationship with PM2.5. The decision tree and SVM learning algorithms demonstrated a strong positive correlation between the estimated PM2.5 concentration and the PM2.5 measurements.

Furthermore, this study demonstrated the effectiveness of machine learning algorithms, specifically decision tree and SVM, in estimating PM2.5 concentration based on AOT data, with feature selection significantly improving the performance of the models. Additionally, the analysis of the importance of AOT spectral bands for PM2.5 concentration estimation provided insights into the physical relationship between AOT and PM2.5.

To further improve the performance of the models, an ensemble machine learning approach could be employed. Ensemble learning combines the results of multiple models to generate more accurate predictions. In the context of PM2.5 estimation, ensemble learning could be used to combine the strengths of decision tree and SVM models, resulting in more accurate and robust PM2.5 predictions, particularly in regions with high spatial variability in AOT and PM2.5 concentrations.

Finally, this study provides a promising approach for estimating PM2.5 concentration using machine learning algorithms and AOT data. Further research is needed to explore the potential of ensemble machine learning for this application and to evaluate the performance of the models on independent datasets.{Gurauskiene, 2006, Eco-design methodology for electrical and electronic equipment industry}

Acknowledgements

This research project was financially supported by Mahasarakham University.

References

Aroonsri I., Sangpradid S. (2021) Artificial neural networks for classification of shrimp farm from satellite imagery. Geographia Technica 16(2): 149-159. https://doi.org/10.21163/GT_2021.162.12

Debnath R., Takahashi H. (2010) A new ensemble learning with support vector machines. International Conference on Computer and Information Application (ICCIA 2010): 33-35. https://doi.org/10.1109/ICCIA.2010.6141529

Doreswamy, Harishkumar K.S., Yogesh K.M., and Gad I. (2020) Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. Procedia Computer Science 171: 2057-2066. https://doi.org/10.1016/j.procs.2020.04.221

Farrow A., Anhäuser A., and Moun-Ob A. (2021) The burden of air pollution in Thailand 2021 conservation, Greenpeace Thailand. Available at: https://www.greenpeace.org/static/planet4-southeastasia-stateless/2022/06/67375e28-the-burden-of-air-pollution-in-thailand_2021.pdf (accessed 20 March 2022).

Gou Y., Tang Q., Gong D.Y., and Zhang Z. (2017) Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. Remote Sensing Environment 198: 140-149. https://doi.org/10.1016/j.rse.2017.06.001

Grange S.K., Carslaw D.C., Lewis A.C., Boleti E., and Hueglin C. (2018) Random forest meteorological normalisation models for Swiss PM10 trend analysis. Atmospheric Chemistry and Physics 18: 6223-6239. https://doi.org/10.5194/acp-18-6223-2018

Gupta P., Christopher S.A. (2009) Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. a neural network approach. Journal of Geophysical Research: Atmospheres 114: 2025. https://doi.org/10.1029/2008JD011497

Gupta P., Zhan S., Mishra V., Aekakkararungroj A., Markert A., Paibong S., and Chishtie F. (2021) Machine learning algorithm for estimating surface PM2.5 in Thailand. Aerosol and Air Quality Research 21(11): 210105. https://doi.org/10.4209/aaqr.210105

He Y., Sun L., Sun Z., and Hu X. (2021) MODIS Aerosol optical thickness product algorithm verification and analysis.

Aerosol and Air Quality Research 21(11): 210019. https://doi.org/10.4209/aaqr.210019

Hua X., Waller L.A., Al-Hamdan M.Z., Crosson W.L., Ester M.G., Ester S.U., Quattrochi D.A., Sarnat J.A., and Liu Y. (2013) Estimating ground-level PM2.5 concentrations in the southeastern U.S. using geographically weighted regression. Environmental Research 121: 1-10. https://doi.org/10.1016/j.envres.2012.11.003

Khamkhantee N., Premanoch P. (2021) Forecasting PM2.5 concentrations using mathematical models and satellite imagery to find guidelines for setting standards under the environmental law, in an industrial area in Samut Prakan Province. Thai Environmental Engineering Journal 35(1): 53-61.

Kumar N., Chu A., and Foster A. (2007) An empirical relationship between PM2.5 and aerosol optical depth in Delhi Metropolitan. Atmospheric Environment 41(21): 4492-4503. https://doi.org/10.1016/j.atmosenv.2007.01.046

Letchumanan K.A.L., Palanichamy N. (2022) Machine learning regression models to predict particulate matter (PM2.5). Proceedings of the International Conference on Computer, Information Technology and Intelligent Computing (CITIC 2022) 10: 458-468. https://doi.org/10.2991/978-94-6463-094-7_36

Madjarov G., Gjorgjevikj D., and Delev T. (2010) Ensembles of Binary SVM Decision Trees. ICT Innovations 2010 Web Proceedings:1857-7288.

Masood A., Ahmad K. (2020) A model for particulate matter (PM2.5) prediction for Delhi based on machine learning approaches. Procedia Computer Science 167: 2101-2110. https://doi.org/10.1016/j.procs.2020.03.258

Phuengsamran P., Lalitaporn P. (2021) Estimating particulate matter concentrations in central Thailand using satellite data. Thai Environmental Engineering Journal 35(3): 1-11.

Sangpradid S. (2023) Application of a multi-layer perceptron neural network to simulate spatial-temporal land use and land cover change analysis based on cellular automata in Buriram province, Thailand. Environment Engineering and Management Journal 22(5): 917-930. https://doi.org/10.30638/eemj.2023.074

Sangpradid S., Prasertsri N., and Aroonsri I. (2022) Geographically weighted regression modelling for above-ground biomass assessment from satellite imagery, in Tad Sung waterfall park forest, Thailand. International Journal of Conservation Science 13(2): 695-704.

Thailand Pollution Control Department (2021) Announces air quality on March 29. Available at: https://www.pcd.go.th/pcd_news/12568 (accessed 30 March 2022).

Vignesh P.P., Jiang J.H., and Kishore P. (2023) Predicting PM2.5 concentrations across USA using machine learning. Available at: https://www.authorea.com/users/579292/articles/627112-predicting-pm2-5-concentrations-across-usa-using-machine-learning (accessed 20 March 2023). https://doi.org/10.22541/au.167631720.03707478/v1

Weber S.A., Engel-Cox J.A., Hoff R.M., Prados A.I., and Zhang H. (2010) An improved method for estimating surface fine particle concentrations using seasonally adjusted satellite aerosol optical depth. Journal of the Air and Waste Management Association 60: 574-585. https://doi.org/10.3155/1047-3289.60.5.574

WHO (World Health Organization) (2018) Burden of disease from ambient air pollution for 2016, version 2. April 2018. Available at: https://www.who.int/airpollution/data/AAP_BoD_results_May2018_final.pdf (accessed 17 March 2022).

WHO (World Health Organization) (2021) WHO global air quality guidelines particulate matter. Available at: https://apps.who.int/iris/handle/10665/345329 (accessed 20 March 2022).

Xie Y., Wang Y., Zhang K., Dong W., Lv B., and Bai Y. (2015) Daily estimation of ground-Level PM2.5 concentrations over Beijing using 3 km resolution MODIS AOD. Environment Science and Technology 49: 12280-12288. https://doi.org/10.1021/acs.est.5b01413

Xu X., Zhang C. (2020) Estimation of ground-level PM2.5 concentration using MODIS AOD and corrected regression model Over Beijing, China. PLoS ONE 15(10): 1-15. https://doi.org/10.1371/journal.pone.0240430

You W., Zhang Z., Zhang L., Li Y., Pan X., and Wang W. (2016) National-scale estimates of ground-level PM2.5 concentration in China using geographically weighted regression based on 3 km resolution MODIS AOD. Remote Sensing 8(3): 184. https://doi.org/10.3390/rs8030184

Zhang P., Ma W., Wen F., Liu L., Yang L., Song J., Wang N., and Li Q. (2021) Estimating PM2.5 concentration using the machine learning GA-SVM method to improve the land use regression model in Shaanxi, China. Ecotoxicology and Environmental Safety: 225. https://doi.org/10.1016/j.ecoenv.2021.112772