ISSN 3027-8325 (Online)
RMUTTO Reserch Journal Vol.18 No.2 (July - December 2025)
DOI: 10.63271/rmuttorj.v18i2.265288

# Estimate Particulate Matter PM2.5 Concentration Impact of Wildfires Using Machine Learning in Chiang Mai Province, Thailand

**Thiwakorn Sena* and Phattraporn Soytong**

Department of Geoinformatics, Faculty of Humanities and Social Sciences, Burapha University, Thailand

**E-mail: thiwakornsena56@gmail.com Tel. (+66)934899370*

(Received 13/11/2024, Revised 24/12/2025, Accepted 26/12/2025)

## Abstract

Wildfires globally impact ecosystems, with open burning and forest fires being primary causes of air pollution. Thailand's PM2.5 levels rise annually from December to April, particularly in the central and northern regions like Chiang Mai, affecting health and the economy. However, air quality monitoring is often limited, and insufficient monitoring stations make current measurements less reliable and incongruent with the actual environmental conditions. This research directly towards assesses PM2.5 from wildfires using integrated Remote Sensing data. The Random Forest model outperformed XGBoost and CNN, with $R^2$ values of 0.74–0.91, RMSE 10.40–30.53 µg/m³, and MAPE 18.56–36.48%. On average across all stations, RF achieved $R^2$ of 0.89, RMSE 11.61 µg/m³, and MAPE 34.22%. Estimates of feature importance impacting features included AOD-MAIAC (40%), TOTEXTTAU (22%), DUSMASS25 (12%), and CO (11%), confirming wildfires and emissions drive pollution. In addition, RF spatial dynamics maps show concentration peaks in the north and south from January to April, aligned with wildfire activity, and fire heat maps confirmed PM2.5 spikes during intense wildfire periods, impacting air quality. However, due to the limited number of monitoring stations and their uneven distribution across the area, the accuracy of the data varies, leading to discrepancies at certain stations, highlighting the need for further research to improve the understanding of PM2.5 pollution. Future studies should refine RF training and incorporate additional factors to enhance model robustness for more accurate estimations.

**Keywords**: Wildfire, Particulate Matter PM2.5, Remote sensing, Machine learning

## 1. Introduction

Wildfire is a dominant global disturbance (Bowman et al., 2009) emitting dust, smoke, and PM2.5 which exerts detrimental impacts on the respiratory and cardiovascular systems (Weichenthal et al., 2013). Thailand ranks as the 34th most polluted nation globally (Narita et al., 2019; Phairuang et al., 2019; ChooChuay et al., 2020) with elevated PM2.5 concentrations linked to industrial emissions, biomass burning, and stagnant air masses during the dry season (Pollution Control Department, 2023). Numerous studies conducted in various regions of Thailand, notably in the North and Bangkok (BKK), have linked this particulate pollution to industrial and vehicular emissions, biomass burning activities, and meteorological conditions characterized by stagnant air masses induced by cold surges. High concentrations, attributed to agricultural burnings and forest fires, are linked to an estimated 40,000 deaths annually in Thailand (Pinichka et al., 2017). The emission of PM2.5 from burning crop residue and forest fires is estimated at 141,000 and 5000 tons per year, respectively, with concentrations predominantly concentrated in central and northern Thailand (Kanabkaew & Kim Oanh, 2011; Chudnovsky et al., 2013; Junpen et al., 2013). In Chiang Mai Province, PM2.5 has reached critical levels as forest

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

burning becomes more intense due to climatic modifications and the prolongation of summer heat. According to 2023 statistics, wildfire occurrences in Thailand resulted in over 4.7 million rai of burned areas or approximately 7,520 square kilometers of burned areas (Department of National Parks, Wildlife and Plant Conservation, 2023; Geo-Informatics and Space Technology Development Agency, 2023).

In this research, we integrated ground-surveyed wildfire area data from the Chiang Mai Forestry Department in 2023 as a vital parameter to estimate PM2.5 concentration levels (Geng et al., 2018). Utilizing advanced machine learning methodologies, this research endeavors to analyze diverse factors contributing to the variability of PM2.5 and Remote Sensing (RS) data, which is considered a serious challenge addressing the serious challenge of applying satellite data for accurate and reliable PM2.5 estimation. Utilizing satellite-derived AOD in Chiang Mai's mountainous terrain presents technical challenges. Terrain shadowing in steep valleys interferes with reflectance, while cloud contamination causes data gaps. Additionally, high surface reflectance from charred soil and ash distorts AOD retrievals and biases PM2.5 estimation. Understanding limitations is essential for evaluating machine learning robustness in topographically complex regions, including factors such as meteorological patterns, temperature fluctuations, seasonal variations, and other parameters influencing PM2.5 concentrations will be examined. Random Forest, XGBoost, and CNN models were used to improve accuracy in the relation of the severity of wildfire to PM2.5 concentration and prediction with the algorithm. Burn severity directly correlates with emissions; higher severity reflects higher temperatures and fuel consumption, increasing fine particulate matter release. Since different intensities (smoldering versus flaming) alter aerosol profiles, integrating severity data allows models to better simulate the non-linear relationship between wildfire energy release and atmospheric PM2.5 loading. Furthermore, this study develops algorithms integrating ground sensor PM2.5 data with satellite data (AOD, AAI, $NO_2$, $O_3$, CO) to enhance evaluation accuracy and refine understanding of wildfire-related PM2.5 and atmospheric conditions, consistent with previous research studying the distribution of PM.5 in the same area by RF model (Kawichai et al., 2025). And since the lack of monitoring stations and the high cost of network expansion pose substantial challenges, satellite remote sensing emerges as a viable alternative to augment spatial coverage (Son et al., 2023) This approach enables proactive measures and policies for PM2.5 aerosols by applying remote sensing to integrate diverse tools like ArcGIS 10.8, the Google Earth Engine (GEE) cloud platform, Python Jupyter Notebook, and machine learning.

### 1.1 Study area

Chiang Mai (17.25°-20.16° N, 98.08°-99.66° E), covering 20,107.057 $km^2$ (Figure 1) has an average temperature of 25.4°C and 1,100-1,200 mm/year rainfall. Its mountainous terrain, low winter moisture, and accumulated vegetation drive annual wildfires, influencing PM2.5 dynamics. During the dry season, nighttime temperature inversions and valley-basin circulation restrict the vertical and horizontal smoke dispersion. This causes PM2.5 accumulation within intermontane basins and localized spikes often missed by ground stations. These mechanisms pose critical model errors, as standard variables may not account for such microclimatic stabilities.

ISSN 3027-8325 (Online)
RMUTTO Reserch Journal Vol.18 No.2 (July - December 2025)
DOI: 10.63271/rmuttorj.v18i2.265288

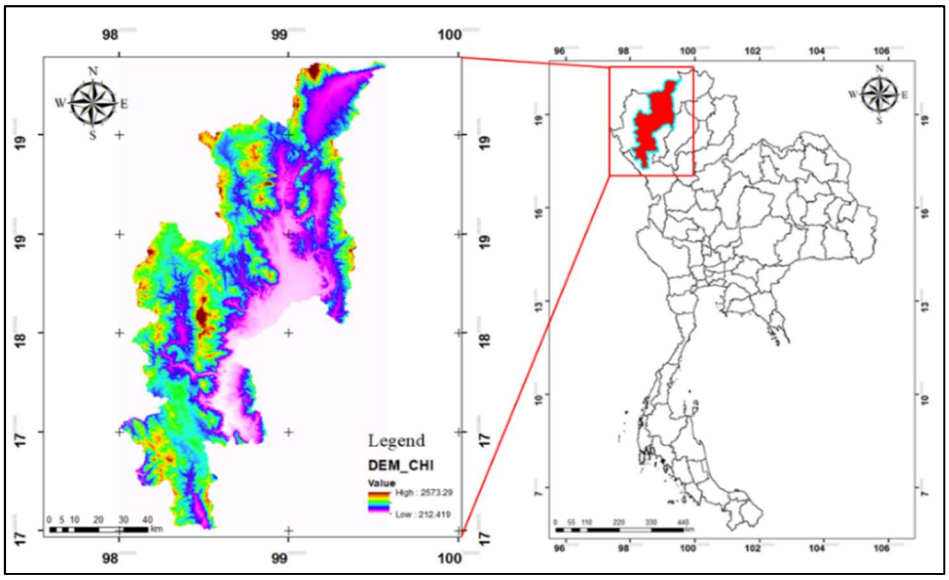



Figure 1 Study area the northeast of Thailand

### 1.2 Objectives

- To evaluate model performance in estimating PM2.5 at on-site and regional scales across Chiang Mai in 2023.
- To compare and determine optimal models for PM2.5 estimation and identify primary factors influencing pollution in wildfire-impacted regions.
- To generate spatial distribution maps of PM2.5 in areas lacking ground stations using remote sensing and machine learning.

### 1.3 Relation research

Researchers worldwide use machine learning and diverse datasets to analyze PM2.5 and PM10 during burning seasons. Using Sentinel-5P, GEOS, and ground data (Wang et al., 2021), an ensemble model estimated daily PM2.5 concentrations in China with high correlation ($R^2$ = 0.922) between PM2.5 and AOD (Gu et al., 2022). In Dalian, BPNN improved accuracy ($R^2$ = 0.752). by integrating meteorological factors. In Guanzhong, RF-XGBoost linked winter pollution to coal use ($R^2$ = 0.93) (Lin et al., 2022). Studies in India used TROPOMI and MODIS to analyze crop burning (Maurya et al., 2022), while research in Thailand benchmarked PM levels against standards, emphasizing multi-source data for air quality assessments (Dontree et al., 2018).

## 2. Methodology

### 2.1 Processing workflow

This study developed PM2.5 predictive models using data from Google Earth Engine (GEE) and ArcGIS 10.8. Independent layers were reprocessed to WGS 84 and linked to a 1x1 km grid. To ensure consistency and reduce satellite noise for robust feature learning, a 1-km radius around PM2.5 monitoring stations was used with bilinear resampling and normal distribution. This suitability for RF and XGBoost models, helps reduce noise from satellite measurement errors, enabling robust feature learning of the relationship between AOD/other factors and PM2.5. Although it may slightly underestimate peak values or cause smoothing in high spatial

DOI: 10.63271/rmuttorj.v18i2.265288

gradients, it effectively matches different spatial and temporal resolutions matches satellite overpasses at 10:30 AM and 1:30 PM. Values were extracted via ArcGIS (multi-values to points), and vector data was converted to raster in Python for training and mapping. The RF model assessed feature importance (Figure 4) using a 70/30 training-testing split. Data sources included MODIS (MCD19A2), MERRA-2 (TOTEXTTAU, DUSMASS25), Sentinel-5P, and auxiliary datasets (fire counts, LST, DEM) from GEE and ArcGIS. Over 39% of aerosol data was missing after calculation. To address this, backfilling was performed using MERRA-2 aerosol products (total aerosol extinction AOT 550 nm) and data from previous studies. Additionally, MERRA-2 data was used to analyze total surface mass (PM2.5) emissions from surface-level combustion. These were matched with ground PM2.5 data to form independent and dependent variables. A total of 1,180 records were used to train RF, XGBoost, and CNN models for PM2.5 estimation. The resulting data was combined with ground observations to create 1x1 km resolution raster layers in ArcGIS 10.8 to produce training and testing sets for regional-scale predictions (excluding six specific stations). The model generated 1 km resolution spatial distribution maps for Chiang Mai in 2023, accounting for monthly and seasonal variations from January to June (Figure 2).

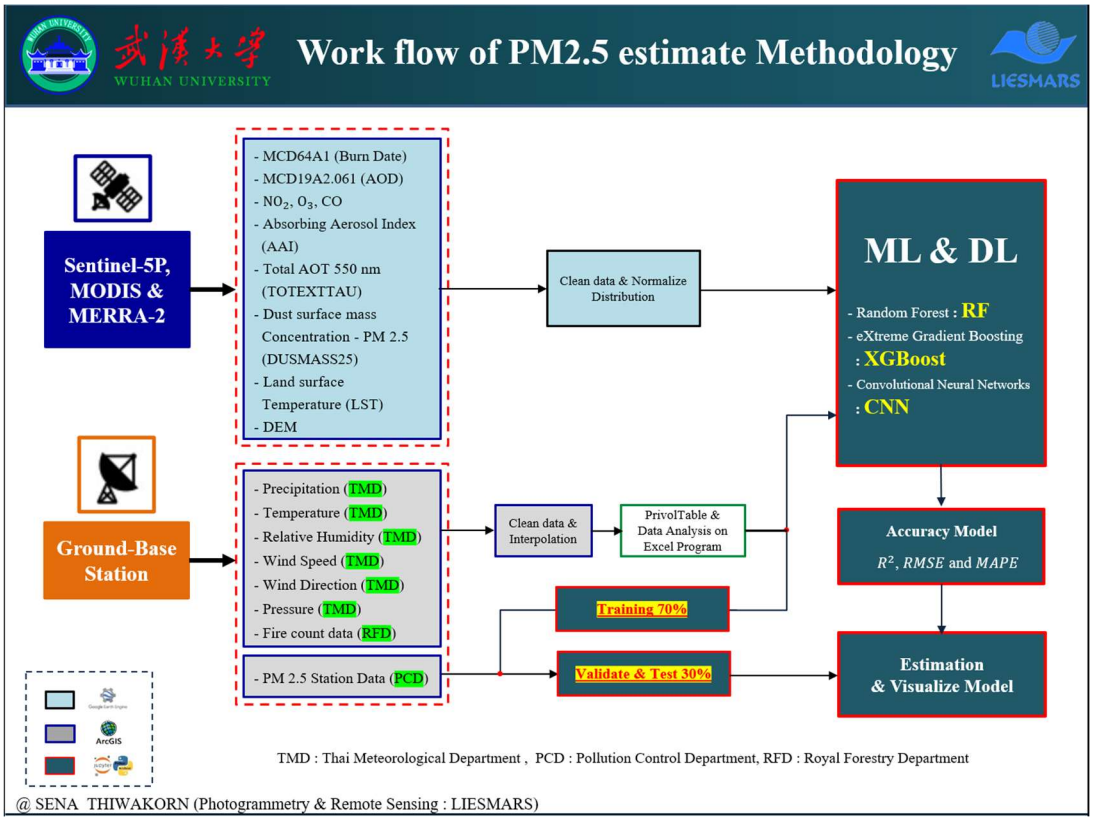



**Figure 2 Flowchart of Methodology**

Integrating ground truth, satellite imagery, and auxiliary factors through machine learning enables efficient PM2.5 spatial estimation at 1x1 km resolution, covering areas without monitoring stations. Although accuracy varies with wildfire intensity, this process ensures assessments are more consistent with actual environmental conditions than relying solely on ground-based monitoring stations.

### 2.2 PM2.5 Ground station measurements

Hourly and daily concentrations of air pollution (PM2.5 (μg/m³)) with 6 monitoring ground-based stations, show in Table 1, the Pollution Control Department in Thailand, track the concentration of pollutants on an hourly and daily basis. To demonstrate how the severity of the wildfire affects the concentration of particulate matter (PM2.5). In this study using the Nearest-Hour Matching method, the hourly average PM2.5 concentration value between 10:30 AM and 13:30 PM daily was selected to align with the satellite data during the daytime overpass period and measurement, which helps achieve more accurate matching of station readings with satellite data from the first 6 months of 2023 to illustrate the period of high concentration and wide spread following the fires, encompassing both pre-fire and post-fire periods. The accuracy of the satellite data was assessed by the use of ground measurement stations.

### 2.3 Remote sensing data

#### 2.3.1 Sentinel-5P TROPOMI

The Sentinel-5P TROPOMI monitors atmospheric pollutants from wildfires at 0.01° (1.11 km) resolution. Nitrogen dioxide ($NO_2$) and ozone ($O_3$) are used to estimate PM2.5 concentrations, as $NO_2$ forms nitrate particles and $O_3$ oxidizing VOCs. These parameters were computed using Google Earth Engine (GEE) over fire-affected areas.

#### 2.3.2 Aerosol Optical Depth product of MCD19A2

This study used MODIS AOD data (MAIAC, MCD19A2) at 0.55 μm and 1 km resolution. The AOD values above 4 were excluded due to interference (Di et al., 2016), leaving approximately 39% of the dataset missing. Inverse Distance Weighting (IDW) interpolation was used to fill gaps, prioritizing nearby data points, to analyze PM2.5 spread. Processing was done in GEE for Jan 1 to Jun 30, 2023.

#### 2.3.3 Total AOT 550 nm and Dust surface mass PM2.5 from MERRA-2 Reanalysis

NASA's MERRA-2 provides aerosol data at 0.5 × 0.625° resolution (Gelaro et al., 2017), integrating MODIS, AVHRR, MISR, and AERONET (Gelaro et al., 2017) via a 3D assimilation algorithm with a 6-hour GSI cycle (Wu et al., 2002; Kleist et al., 2009) provides aerosol data from 1980 to present (Rienecker et al., 2011; Molod et al., 2015). In this study, MERRA-2 data retrieved via GEE replaced missing AOD-MAIAC data to ensure model accuracy (Sayeed et al., 2022).

#### 2.3.4 The product of MCD64A1 resolutions 500 m

MCD64A1 MODIS product in GEE integrates surface reflectance and active fire datasets for daily fire detection at 500 m. Using the Burn Sensitive Vegetation Index (VI) from shortwave infrared and reflectance bands (Giglio et al., 2018), it improves small burned area detection compared (Rodrigues et al., 2019) to previous collections.

### 2.4 Meteorological conditions

Meteorological variables, including TEMP, RH, WS, WD, PRES, and PRE, affect PM2.5 characteristics and dispersion. Daily and hourly data from the Thai Meteorological Department (Thai Meteorological Department, 2023) were matched with satellite overpass times using station coordinates. Since rainfall and cross-ventilation facilitate PM2.5 sedimentation and dispersion (Jiao et al., 2016). This study paired meteorological data with PM2.5 images by location and date. To address the coarse resolution of

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

meteorological data, interpolation was used to match the 1 × 1 km spatial resolution of satellite data for machine learning model training.

### 2.5 Fire count ground data

The Burn Area Ground Base dataset from the Royal Forestry Department of Chiang Mai documents 2023 fire occurrences. Mapped using ArcGIS 10.8 with ~1,600 data points, each entry includes fire dates and damage extent. Fires are classified as "fire days" or "non-fire days" within a 1 km radius of PM2.5 stations, stored in shapefile format as an independent variable for the analysis model.

### 2.6 Auxiliary Data

Auxiliary data, including 30 m Copernicus DEM (GLO-30) and Landsat 8 Land Surface Temperature (LST) was calculated via Planck's law and NDVI-derived emissivity (EM). Both datasets, accessed via GEE, were resampled to 1 km to align with AOD and PM2.5 data were used to improve model performance

**Table 1 Variables used in the model development and validation**

| Variables | Short name | Unit | Temporal resolution | Spatial resolution | Source |
|---|---|---|---|---|---|
| MCD19A2 | AOD-MAIAC | | Daily | 1 km | MODIS (Terra & Aqua) |
| Air Pollutions | AAI | $\mu g/m^3$ | 24 h average | 0.01 arc degree (1.11km) | Sentinel-5P TROPOMI |
| | $NO_2$ | $\mu g/m^3$ | 24 h average | 0.01 arc degree (1.11km) | |
| | $O_3$ | $\mu g/m^3$ | 24 h average | 0.01 arc degree (1.11km) | |
| | CO | $\mu g/m^3$ | 24 h average | 0.01 arc degree (1.11km) | |
| Dust surface mass concentration PM < 2.5 μm | DUSMASS25 | $\mu g/m^3$ | 24 h average | 0.5 x 0.625° | MERRA-2 Reanalysis |
| Total Aerosol Extinction AOT 550 nm | TOTEXTTAU | $\mu g/m^3$ | 24 h average | 0.5 x 0.625° | |
| MCD64A1 | BA | M | Daily | 500 m | MODIS (Terra & Aqua) |
| Fire count data | | $km^2$ | Daily | | Royal Forestry Department |
| Topographic features | DEM | m | Annual | 30 m | Copernicus GLO-30 |
| Land Surface Temperature | LST | °C | Seasonal | 30 m | Landsat-8 |
| Estimated feature | PM 2.5 | $\mu g/m^3$ | Hourly & Daily | | Pollution Control Department |
| Meteorological Conditions | TEMP | °C | Hourly & Daily | | Thai Meteorological Department (TMD) |
| | RH | % | Hourly & Daily | | |
| | WS | m/s | Hourly & Daily | | |
| | WD | m/s | Hourly & Daily | | |
| | PRES | hPa | Hourly & Daily | | |
| | PRE | mm | Hourly & Daily | | |

### 2.7 The correlation coefficients and feature importance

Correlation analysis examined PM2.5 relationships with various factors (Month, AOD, $NO_2$, $O_3$, CO, DUSMASS25, TOTEXTTAU, Meteorology, LST, DEM, and fire counts). PM2.5 correlates negatively with PRE, PRS, RH, TEM, and WD, but positively with WS. Satellite data shows positive correlations with most pollutants (except $O_3$), while wildfires significantly increase PM2.5, greenhouse gases, and temperatures (Figure 3). Consequently, experiments demonstrate high PM2.5 estimation accuracy, contingent upon these factors. To ensure robustness, Pearson correlation (r) and p-values were evaluated through multiple iterations. High feature importance assigned to AOD and CO in the Random Forest model validates that wildfires and biomass burning are primary drivers of Chiang Mai's PM2.5 crisis, involving complex non-linear relationships. Although AOD is the most influential factor, integrating hotspots and CO is essential to identify sources and mitigate statistical bias. Consequently, RF was selected for its robustness against overfitting and superior ability to capture non-linear patterns compared to XGBoost and CNN.

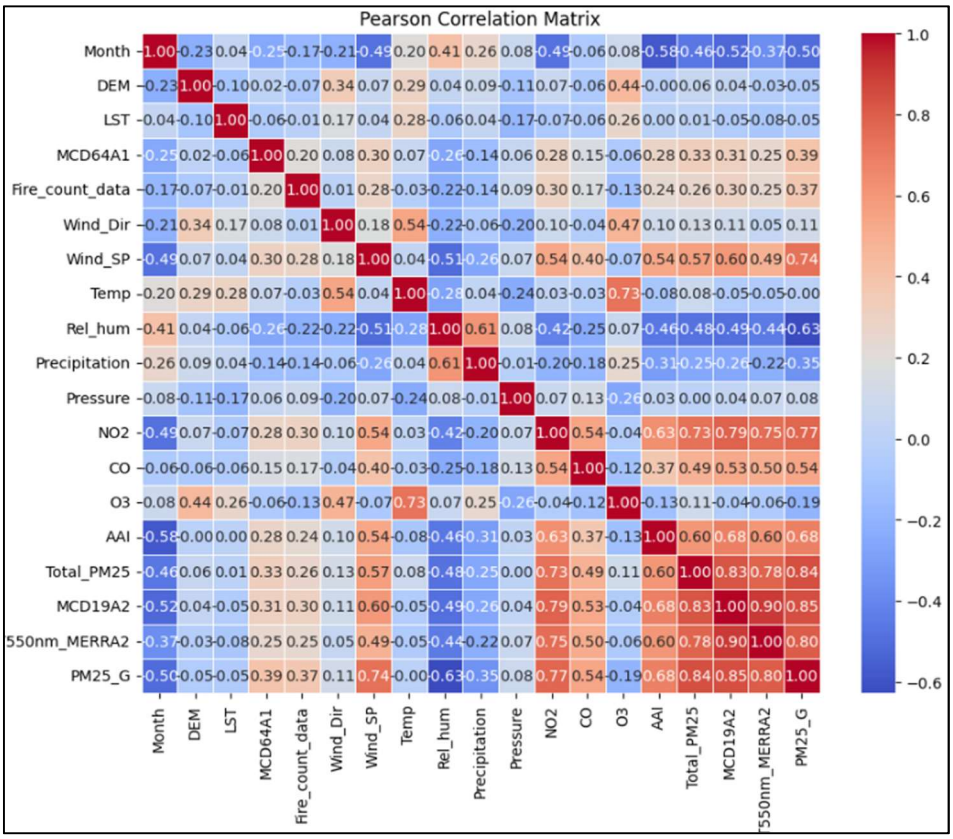



**Figure 3 Correlation coefficient matrix between PM2.5 observations and various independent variables**

DOI: 10.63271/rmuttorj.v18i2.265288

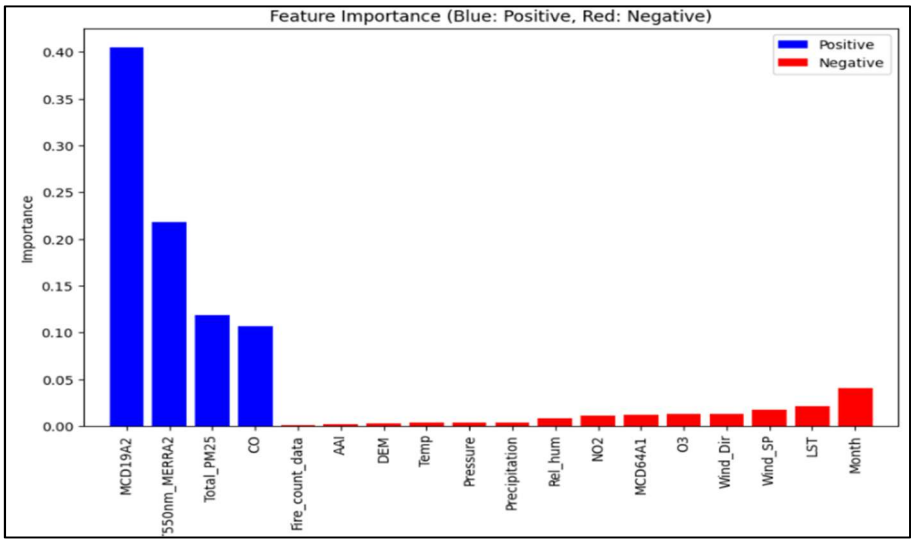



Figure 4 Feature importance

## 2.8 Machine learning Models and Architecture

### 2.8.1 Random Forest Regression

To provide accurate, fast estimations, this study employs Random Forest Regression (RFR) (Li et al., 2023). Superior to conventional methods, RFR predicts continuous functions, requires less training, resists overfitting, and models nonlinear relationships (da Silva Chagas et al., 2016; Chang et al., 2023). The algorithm combines numerous regression trees with hierarchical constraints for final predictions (Rodriguez-Galiano et al., 2014; You et al., 2016). Eq. (1) presents the general RFR formula, where $\widehat{f_{RF}^{K}}(X)$ denotes the RFR predictors, $X$ denotes the input vector, $K$ denotes the number of decision trees, and $\{T(X)\}_1^K$ denotes the constructed decision trees (Rodriguez-Galiano et al., 2014).

$$\widehat{f_{RF}^{K}}(X) = \frac{1}{K}\sum_{k=1}^{K} T(X) \quad (1)$$

In performing RFR, 70% of the data are used as the training sample, while 30% are used for validation, utilizing Python within a Jupyter notebook.

### 2.8.2 eXtreme Gradient Boosting

XGBoost originates from Gradient Boosting Machine (GBM), an ensemble-learning algorithm that adjusts weights each iteration adding weight to misclassified errors and subtracting from correctly classified ones (Bisri & Wahono, 2015). GBM employs second-order gradient statistics to minimize the regularized objectives in Eq. (2).

$$\mathcal{L}(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

where $\Omega(f) = \gamma T + 1/2\,\lambda \|w\|\text{^}2$

The function $\mathcal{L}$ represents a differentiable convex loss function quantifying the disparity between prediction $yi$ and the target $yi$, while $\Omega$ penalizes the model's complexity (Chen & Guestrin, 2016). To find optimal parameters, a hyperparameters grid was created. Table 3 presents a comprehensive summary of the ten pivotal parameters of XGBoost, encompassing their respective ranges and tuning values.

DOI: 10.63271/rmuttorj.v18i2.265288

**Table 2 Hyperparameter tuning of RF algorithms**

| Parameter | Definition | Tuning |
|---|---|---|
| Booststrap | method for sampling data points (with or without replacement) | True |
| Max_depth | max number of levels in each decision tree | 10 |
| Max_features | max number of features considered for splitting a node | auto |
| Min_samples_leaf | min number of data points allowed in a leaf node | 1 |
| Min_samples_split | min number of data points placed in a node before the node is split | 2 |
| Min_weight_fraction_leaf | min weighted fraction of the sum total of weights (of all the input samples) required to be at a leaf node. | 0 |
| n_estimators | number of trees in the forest | 550 |
| random_state | seed of the random number generator used to initialize the random state. | 70 |

**Table 3 Hyperparameter tuning of XGBoost algorithms**

| Parameter | Definition | Tuning |
|---|---|---|
| nrounds | Number of trees to be used in the model | 300 |
| eta ($\eta$) | Learning rate | 0.1 |
| gamma ($\gamma$) | Minimum spilt loss reduction | 3 |
| Max_depth | Maximum depth of trees | 9 |
| Minimum_Child_weight | Minimum observation for a child | 1 |
| subsample | Ratio of random samples to be considered for training | 1 |
| Column_sample_tree | Ratio of features (columns) to be used to train each tree | 0.8 |
| Column_sample_level | Ratio of features to be used to train each split of the tree | 0.8 |
| Lambda ($\lambda$) | Regularization term on the weights for the $L^2$ norm Eq.(2) | 2 |
| Alpha ($\alpha$) | Regularization term on the weights for the $L^1$ norm Eq.(2) | 60 |

#### 2.8.3 Convolutional Neural Networks

This research employs a 2D convolutional neural network (Conv2D) with a 3x3 image patch to learn complex relationships. By transforming 1D vectors into a grid-like structure, Conv2D layers extract features and capture spatial relationships more effectively than linear processing. This technique enables the Deep Learning model to access in-depth correlations between meteorological factors and AOD for more accurate PM2.5 estimates. Its architecture comprises two primary components: (1) The convolution tool for feature extraction, which discerns and segregates image features, and the connected layer for prediction, utilizing the convolutional output to make predictions based on the identified features.

This model employs a hybrid technique: Conv1D and Max_pooling1D extract features, followed by reshaping into Conv2D and Max_pooling2D to detect grid-like relationships. Finally, data is flattened for the Dense Layer to output a PM2.5 value. Data were divided into 70% training and 30% testing sets, using ground-based data for validation.

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

**Table 4 Parameter of CNN algorithms by Keras using Sequential API**

| Layer (type) | Output Shape | Param |
|---|---|---|
| Conv1D | 0, 11, 64 | 256 |
| Max_pooling1D | 0, 5, 64 | 0 |
| Conv2D | 0, 3, 64 | 12352 |
| Max_pooling2D | 0, 1, 64 | 0 |
| Flatten | 0, 64 | 0 |
| Dense_1 | 0, 128 | 8320 |
| Dense_2 | 0, 1 | 129 |

### 2.9 Model Assessment Accuracy

The datasets were split into a 70% training dataset and a 30% validation dataset. GridSearchCV in scikit-learn was employed to identify optimal parameter values through a 5-fold cross-validation method an international standard for checking overfitting. This method partitions the data into 5 subsets, using 4 for training and 1 for testing in each iteration, repeated 5 times so each subset is used exactly once for validation. This thorough testing ensures the model's robustness across all split data variations. Optimizing parameters using GridSearchCV significantly enhances performance and flexibility, with evaluation metrics such as $R^2$, RMSE, and MAPE employed to assess the robustness of all models (Eqs. 3–5). The results are average values from testing with data the model has 'never seen before.' If overfitting truly occurred, these statistical values would be significantly poor when predicting data in the validation set.

$$R^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y}_i)^2} \quad (3)$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|^2 \quad (5)$$

where $\boldsymbol{y_i}$ represents the PM2.5 observations, $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ stands for the predicted value, $\bar{\boldsymbol{y}}$ denotes the average of the PM2.5 observations, and $\boldsymbol{n}$ indicates the overall sample size.

## 3. Results and Discussion

### 3.1 Statistics of Variables

Total data with 1,180 matched samples for 6 ground monitoring stations over a 6-month study period in 2023 was plotted in a histogram with basic statistics (Figure 5). The daily PM2.5 concentration curve is similar to AOD-MAIAC and TOTEXTTAU (Wei et al., 2019). while DUSMASS25 obtained from the MERRA2 reanalysis model indicates that aerosol data used in this study has a good relationship with PM2.5 estimation and contributes greatly to model accuracy. The measured mean PM2.5 concentration was 55.60 µg/m³, ranging from 2.21 µg/m³ in the rainy seasonals to 378.45 µg/m³ in the winter and dry seasonals (Figure 5a) affected

by annual wildfires and smog from neighboring Burma. The annual average AOD-MAIAC value was 0.49 (range: 0.08-1.77), while the average TOTEXTTAU was 0.38 (range: 0.04-1.58), showing values very similar to AOD-MAIAC. And in terms of DUSMASS25 had an average of 64 μg/m³ (range: 6–277 μg/m³), which is very similar to the ground station PM2.5 values. Moreover, the curve of carbon monoxide (CO) (Figure 5g) is very consistent with the Fire Count Data from the Royal Forestry Department (RFD) and Burn Date analysis from the MCD64A1 from MODIS product (Figure 5o-p). This confirms that wildfires play a very important role in the release of air pollutants and PM2.5 into the atmosphere, consistent with both ground station and remote sensing data.

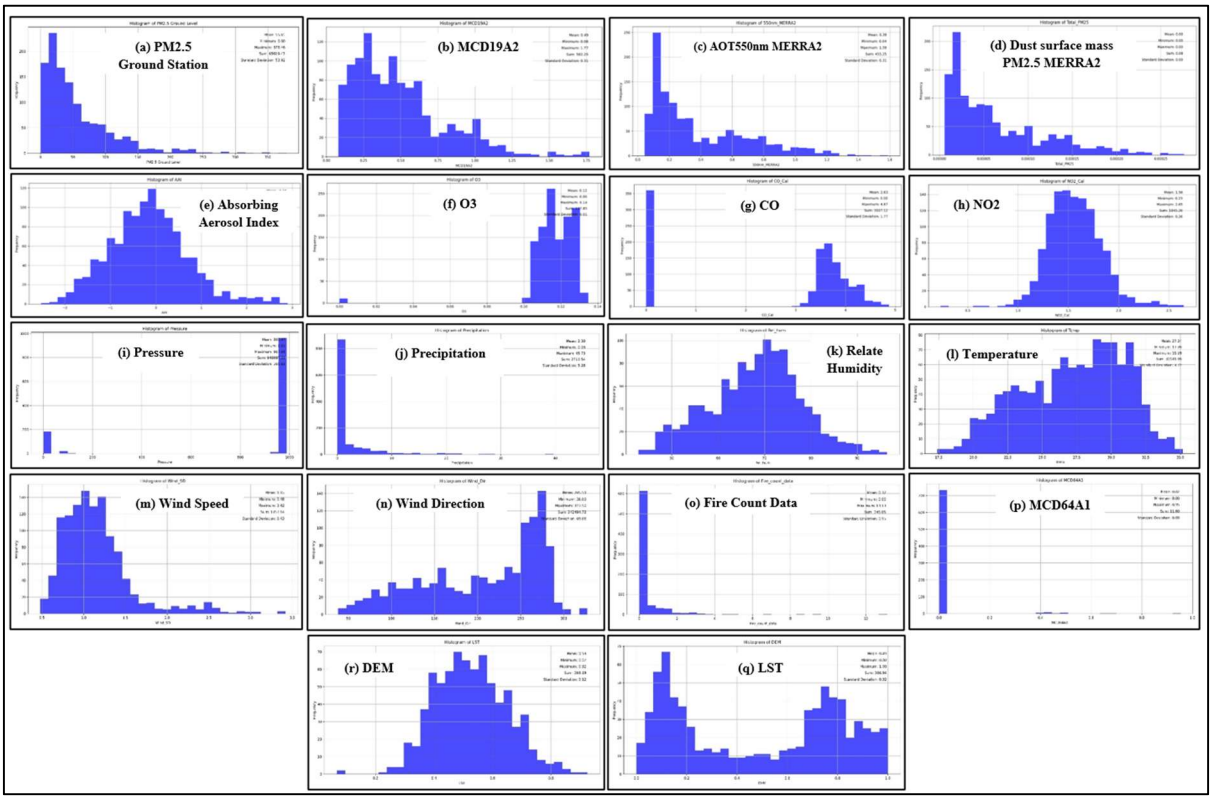



**Figure 5 Shows histograms and descriptive statistics (minimun, maximum, sum, mean, and standard deviation) for PM2.5 concentration ground station and independent variables**

## 3.2 Performance of the Model

### 3.2.1 On-site scale

The accuracy of PM2.5 concentration estimates in Chiang Mai Province is assessed using combined data from all 6 stations across three model evaluations. The Random Forest model demonstrates excellent performance (Figure 6), achieving an R-squared (RMSE), The eXtreme Gradient Boosting (XGBoost) model shows inferior performance, achieving an R-squared ($R^2$) of 0.81 an RMSE of 12.51, and a MAPE of 37.38. Finally, the Convolutional Neural Network (CNN) model showed the poorest performance, with an R-squared ($R^2$) of 0.67 RMSE of 14.70, and MAPE of 37.78, respectively (Table 5), reflecting that the CNN architecture in this study may not yet be as capable of handling the complexity of independent factors as the RF and XGBoost models, as shown in Table 2. Cutting-edge machine learning algorithms underscore the efficacy of the RF model, particularly evident in its superior $R^2$ and RMSE scores, indicative of a stronger linear relationship between observed

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

and estimated PM2.5 concentrations. The PM2.5 scatter plot shows the relationship and accuracy between the training and testing generated by the machine learning RF model for each and every station in 2023.

The assessment through scatter plots in Figure 6 demonstrates that the estimated values align accurately with the observed data along the trend line. Conversely, if the model were overfitting, the data points would exhibit anomalies (artifacts) or unstructured clustering. This consistency indicates that the model possesses strong generalization capabilities rather than merely memorizing a specific dataset, leading to overfitting.

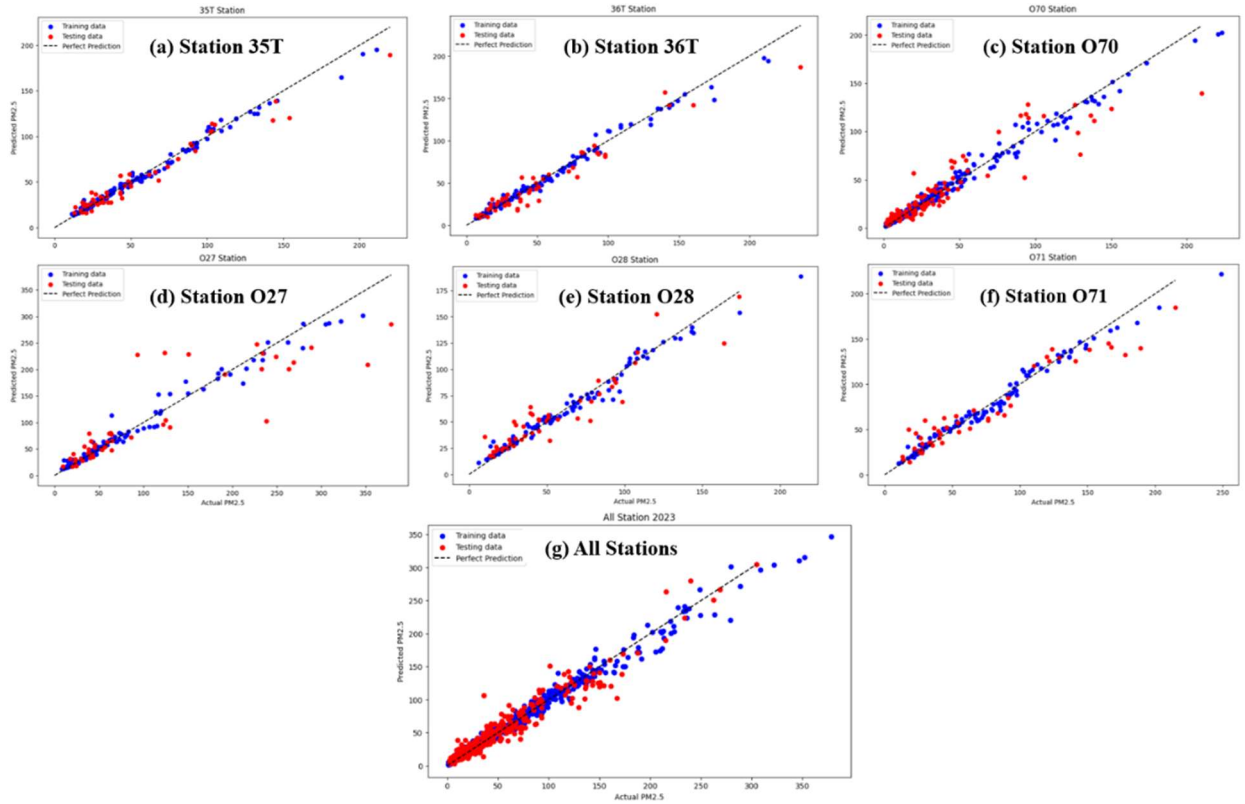



**Figure 6 Scatter plots of the observed and estimated PM2.5 concentrations of each station and all station 2023**

**Table 5 Comparison of model performance on the testing dataset at on-site scale.**

| Station | RF | | | XGBoost | | | CNN | | |
|---|---|---|---|---|---|---|---|---|---|
| | $R^2$ | RMSE | MAPE | $R^2$ | RMSE | MAPE | $R^2$ | RMSE | MAPE |
| 35T | 0.91 | 10.40 | 15.58 | 0.71 | 11.95 | 14.53 | 0.82 | 15.16 | 16.23 |
| 36T | 0.88 | 14.05 | 18.56 | 0.86 | 15.87 | 19.66 | 0.58 | 23.37 | 22.55 |
| O27 | 0.81 | 30.53 | 25.67 | 0.73 | 34.15 | 29.88 | 0.49 | 32.79 | 31.85 |
| O28 | 0.74 | 16.91 | 24.19 | 0.67 | 18.51 | 27.26 | 0.47 | 26.41 | 27.57 |
| O70 | 0.81 | 14.83 | 36.48 | 0.80 | 16.36 | 39.83 | 0.61 | 22.66 | 38.97 |
| O71 | 0.78 | 21.03 | 20.63 | 0.77 | 19.40 | 21.62 | 0.36 | 39.23 | 24.63 |
| All Station | 0.89 | 11.61 | 34.22 | 0.81 | 12.51 | 37.38 | 0.67 | 14.70 | 37.78 |

DOI: 10.63271/rmuttorj.v18i2.265288

This comprehensive analysis focuses on elevated PM2.5 concentrations (Figure 7). Consistent with previous regional studies (Kawichai et al., 2025), the Random Forest model outperforms other machine learning (ML) models, achieving the highest $R^2$ and low RMSE (Figure 7a, 7b). RF also demonstrates the lowest MAPE for concentrations exceeding 20 µg/m³, as shown in Figure 7c. As pollution escalates, the performance disparity between RF and other ML models becomes more conspicuous. Furthermore, RF shows a remarkable capability in addressing PM2.5 overestimation, affirming its superior predictive prowess.

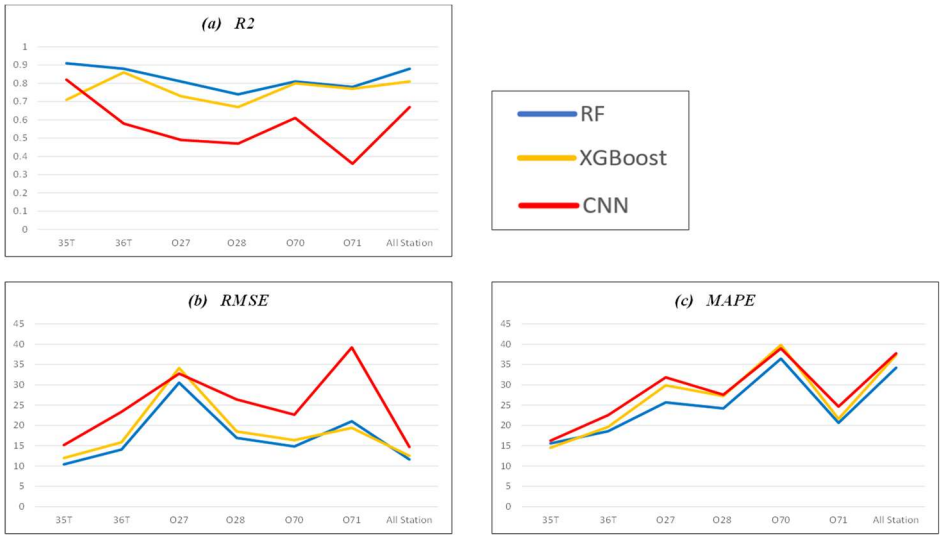



**Figure 7 Model comparison across various thresholds of high PM2.5 concentrations: (a) $R^2$, (b) RMS, and (c) MAPE**

Evaluation indicators across six monitoring stations (Table 2) show notable bias discrepancies, particularly in stations O27, O28, O70, and O71. These exhibit higher values than urban stations (35T, 36T), with RF model RMSE (30.53, 16.91, 14.83, 21.03 µg/m³) and MAPE (25.67, 24.19, 36.48, 20.63 µg/m³) increasing as concentrations exceed 295 µg/m³. From a physics perspective, thick smoke plumes during severe wildfires cause discrepancies between satellite columnar integrated values and ground station measurements. While ground station measurements are in situ at ground level concentration people breathe differs significantly from that visible to satellites, especially at maximum values, resulting in drastically different bias values (Son et al., 2023; Kawichai et al., 2025). These elevated levels, attributed to agricultural and wildfires (Punsompong et al., 2021), Higher error metrics at specific stations are likely due to complex topographical constraints impeding aerosol dispersion. Stations near active wildfires at high altitudes showed significant PM2.5 volatility. This phenomenon is exacerbated by anemological conditions transporting concentrated smoke plumes into localized basins, inducing discrepancies between in-situ ground measurements and satellite columnar values. Such extreme values lead to substantial errors from model underestimation or parameter inaccuracies, particularly at exceptionally high concentrations. However, this deviation stems largely from wildfire intensity, aligning with the study's second objective.

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

**3.2.2 Month and seasonal on regional scale**

As the RF model showed the highest potential on-site scale for PM2.5 estimation, it was used to create spatial distribution maps based on time of month and season on a regional scale for Chiang Mai in 2023. Data were divided into monthly and seasonal training/testing sets. Monthly and averaged test accuracies are shown in Table 6. The winter (January-February) and dry season (March-April) models performed well, with $R^2$ of 0.79 - 0.85, an RMSE of 13.51 - 19.31 $\mu g/m^3$, and a MAPE of 22.14 - 25.54 $\mu g/m^3$, respectively, outperforming the rainy season (May-June) models $R^2$ of 0.74 - 0.81, an RMSE of 12.43 - 21.02 $\mu g/m^3$, and a MAPE of 20.29 - 20.37 $\mu g/m^3$, respectively. This is due to air pollution and higher PM2.5 densities in winter and summer compared to the rainy season, when precipitation and high humidity in this tropical monsoon region decrease concentrations. The 6-month average test accuracy was an $R^2$ of 0.81, an RMSE of 14.45 $\mu g/m^3$, and a MAPE of 21.25 $\mu g/m^3$.

**Table 6 RF model performance of month and difference seasonals on regional scale**

| Month | RF | | |
|---|---|---|---|
| | $R^2$ | RMSE | MAPE |
| Jan | 0.85 | 18.25 | 23.71 |
| Feb | 0.81 | 16.03 | 22.14 |
| Mar | 0.79 | 13.51 | 22.76 |
| Apr | 0.80 | 19.31 | 25.54 |
| May | 0.81 | 12.43 | 20.37 |
| Jun | 0.74 | 21.02 | 20.29 |
| Average periods | 0.81 | 14.45 | 21.25 |

## 3.3 Spatial-temporal distribution pattern

**3.3.1 Estimation PM2.5 concentration at On-site scale**

The spatial distribution of PM2.5 concentrations was investigated by averaging data from January 01 to June 30, 2023, at six sites, excluding missing values (Figure 8a). High concentrations were observed at north (O27) (60 to over 90 $\mu g/m^3$) and south (O71) stations, while lower values occurred at central stations (35T, 36T, O28, O70), around 35 to 50 $\mu g/m^3$ or less, along with Figure 8b, which shows station-level PM2.5 concentrations estimated via the optimized RF model. This demonstrates significant consistency between site-level observations and model evaluations.

**3.3.2 Estimation and mapping the spatial distribution of PM2.5 concentrations across different months and seasons on regional scale.**

The 1 km resolution RF model maps PM2.5 concentrations, showing temporal and spatial heterogeneity consistent with in-situ measurements. The RF model is optimized for mountainous terrains and basins like Northern Thailand, Myanmar, and Laos; the model performs highly. However, applying it to lowlands (e.g., Central Thailand) may require additional variables and retraining.

Spatial coverage in 2023 (Figure 9a–h) shows seasonal changes, with winter being the most polluted. In January-February, dry air and agricultural clearing raise PM2.5 above 120 $\mu g/m^3$ (Figure 9a-b), consistent with site O71 Hkod (Figure 10c). Concentrations peaked (>150 $\mu g/m^3$) in late February-March, particularly in the

DOI: 10.63271/rmuttorj.v18i2.265288

northern mountains (Figure 9c-d) due to severe 10-year record forest fires and agricultural activities. Transboundary smoke from Myanmar also significantly impacts levels, aligned with positive wind speed correlations. By late May, concentrations drop except in cities (Figure 9e–g), where traffic and energy consumption rather than wildfires are primary sources. The RF model's positive interaction with $CO_2$ and biomass combustion confirms separate urban and highland sources. Figure 9h shows the 6-month average, with the highest concentrations in the north and south due to wildfires, consistent with RFD 2023 fire counts. However, the limitation of only six unevenly distributed stations hinders capturing the underlying mechanisms for PM estimation.

To assess temporal variation, daily and monthly observed PM2.5 from 6 stations were compared against model estimates (Figure 10a-f). Results throughout the year (Figure 10g) show concentrations beginning in January and peaking in March, demonstrating the model's value in estimating PM2.5. The time series line chart confirms that remote sensing data driven models estimate PM2.5 concentrations with high efficiency and reliability.

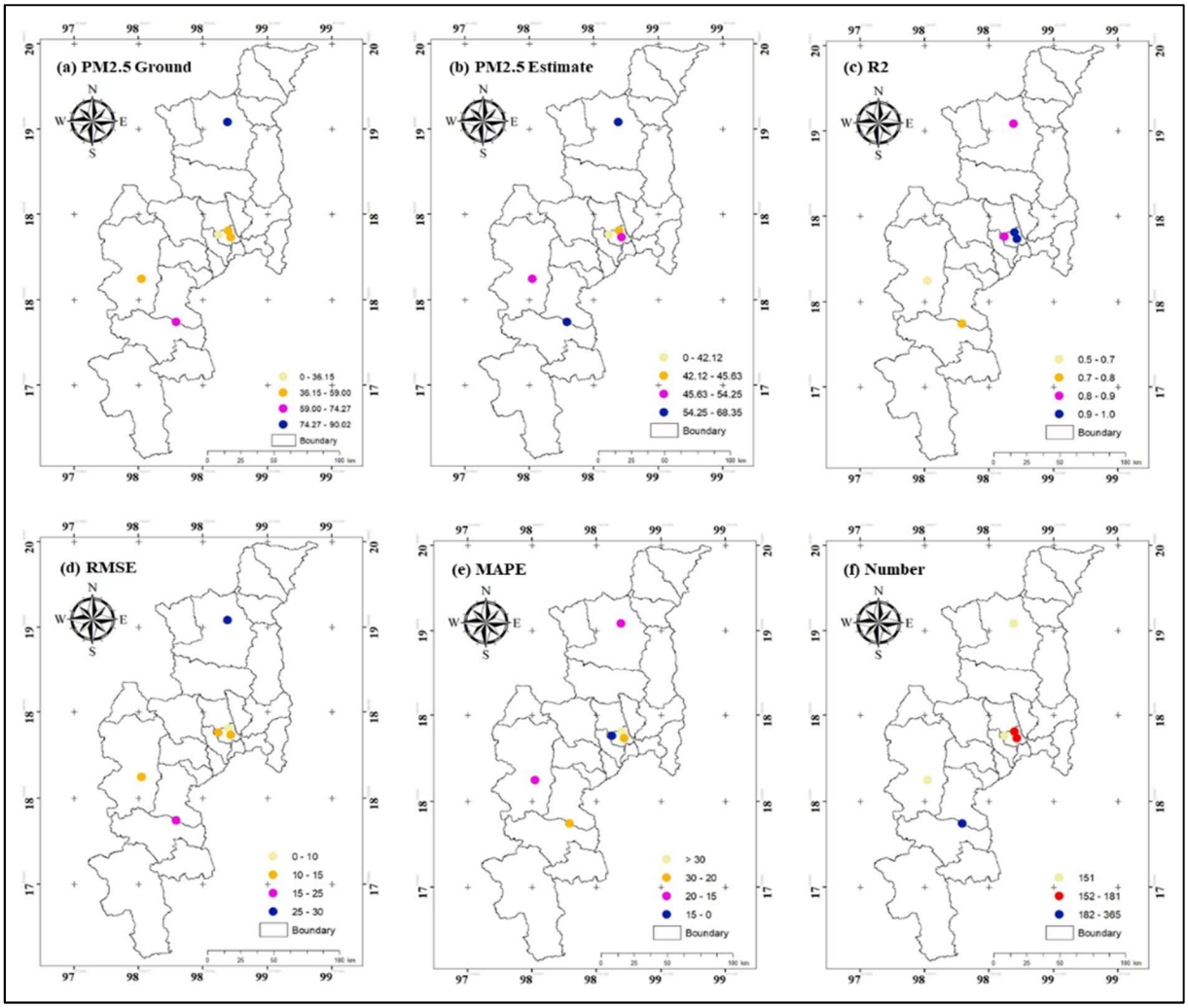



**Figure 8 Spatial distribution of (a) PM2.5 ground station and (b) Estimate PM2.5 concentrations (µg/m3) (c) is $R^2$ (d) RMSE (e) MAPE and (f) numbers.**

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

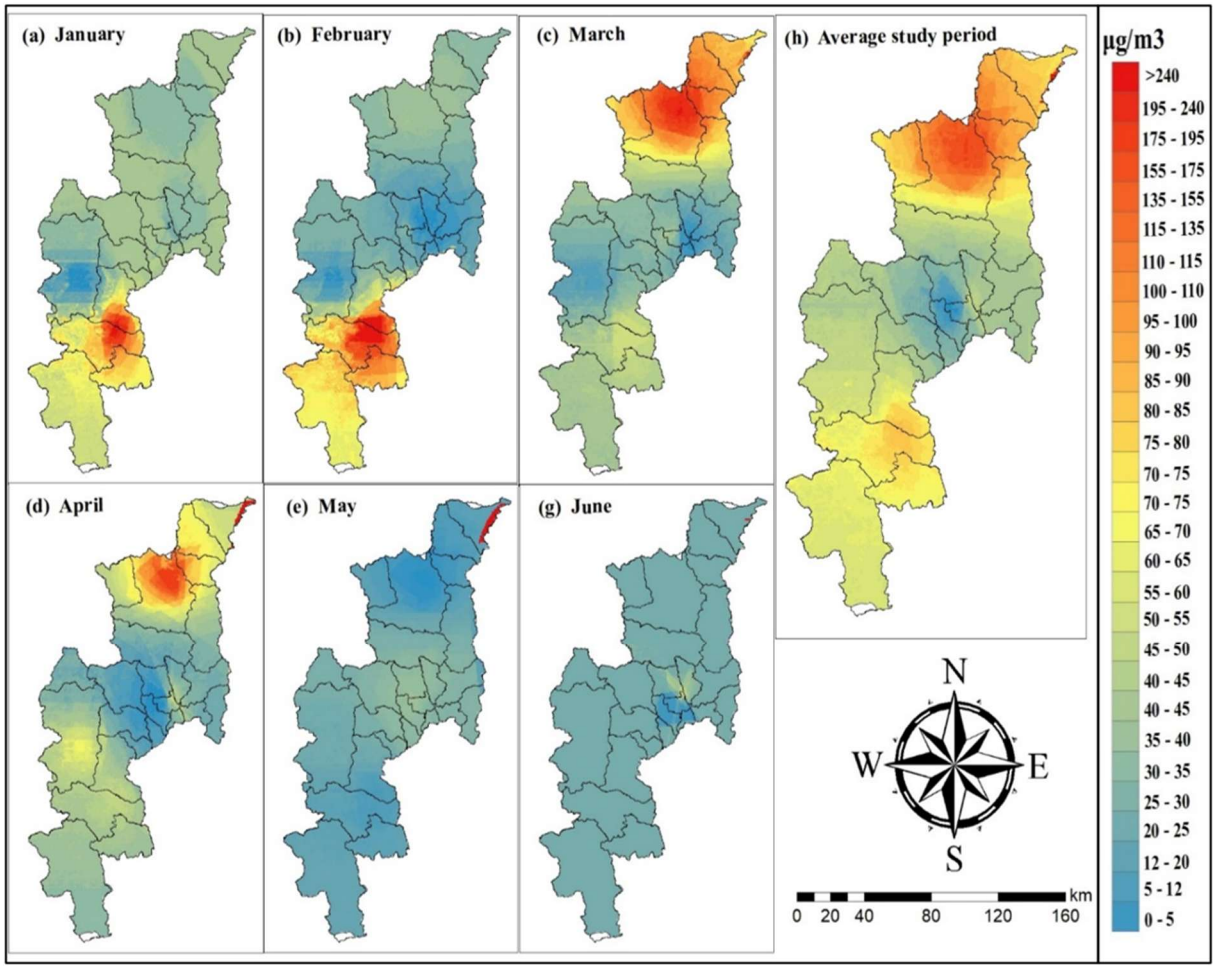



Figure 9 (a–h) Monthly averaged spatial-temporal distributions of the estimated PM2.5

ISSN 3027-8325 (Online)
RMUTTO Reserch Journal Vol.18 No.2 (July - December 2025)
DOI: 10.63271/rmuttorj.v18i2.265288

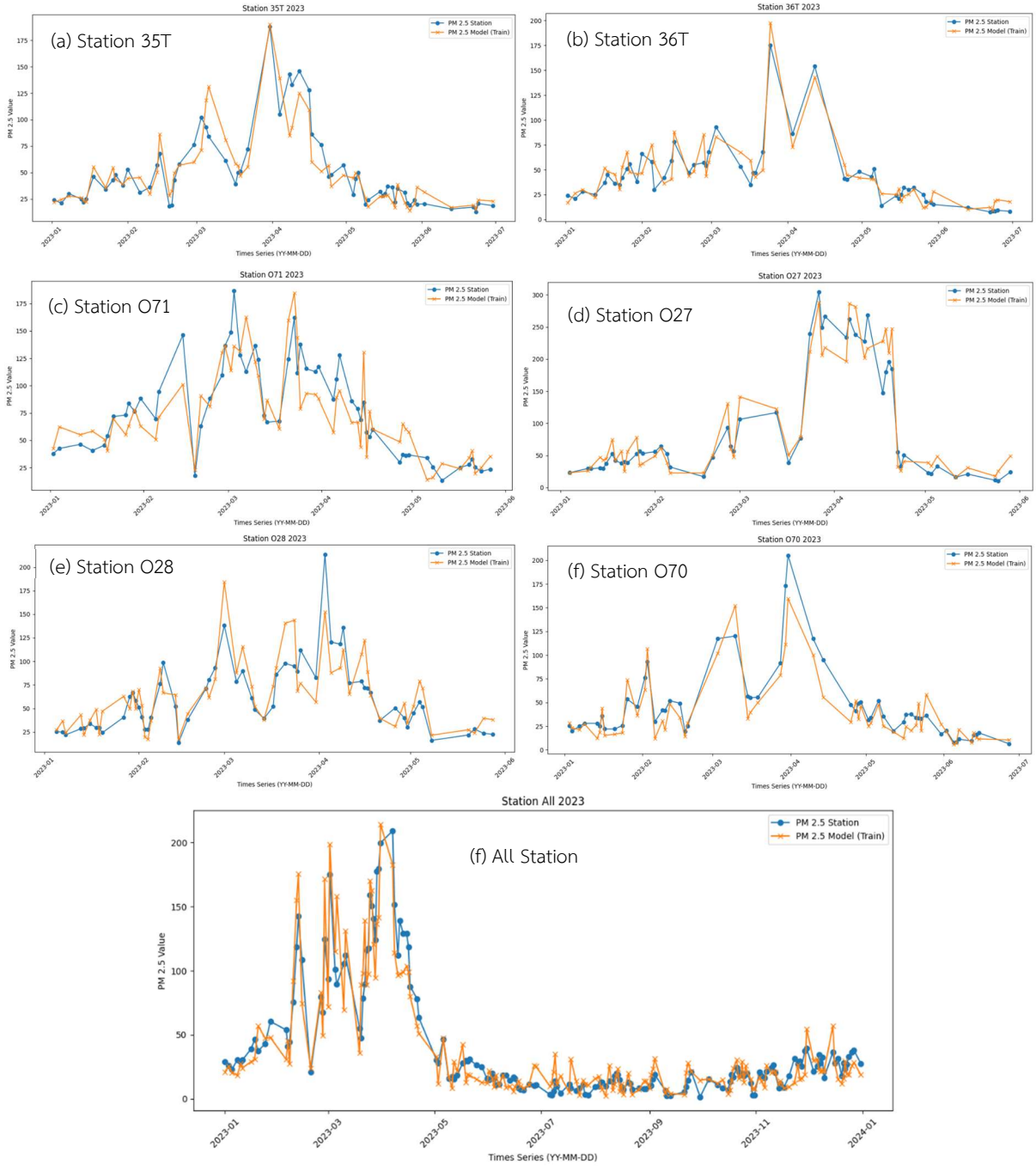



Figure 10 Shows plot times series line chart of PM2.5 concentration daily and month variations of observation (blue) compare with estimation PM2.5 value of the model (orange)

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

### 3.4 Impacts of wildfire on PM2.5 concentration in Chiang Mai Province, Thailand

To investigate 2023 fire impacts on Chiang Mai's air quality, this study reveals significant seasonal and spatial variations. High-resolution PM2.5 mapping from a RF model shows close correlation with in-situ measurements, highlighting heterogeneous pollution distribution of pollution across the region.




**Figure 11 Heat map of severe wildfires based on ground truth data for 1,600 fire incident during 1 Jan to 30 Jun, in 2023.**

Heat maps based on 1,600 ground-truth fire areas (Forestry Department) from 1 Jan to 30 Jun 2023 (Figure 11) show high levels spreading from the southern zone in January-February (Figure 11a-b) to the northern zone in March-April (Figure 11c-d). PM2.5 concentrations peak between February 20th and April 10th due to wildfire activity and agricultural burning during the dry season, exacerbated by low humidity and stable atmospheric conditions, which exacerbate the accumulation of particulate matter in the air. Correlation between the heat map burn area and PM2.5 confirms fires' critical role (Figure 9 and 11), though feature importance values indicate fires are not the sole factor evident from the statistical correlation analysis of the feature importance values derived from the model outcomes. Notably, the report in 2023 experienced the most severe wildfires in a decade. Additionally, transboundary pollution from Burma impacts levels, supported by positive wind speed correlations. While wind components show smaller importance, the northeast and southwest monsoons modulated by El Niño Southern Oscillation (Kirtphaiboon et al., 2014) influence regional emission propagation (Inthacha, 2011).

# 4. Conclusions

## 4.1 Model performance

Model Performance RF outperformed XGBoost and CNN with the highest $R^2$ (0.89), lowest RMSE (11.61 μg/m³), and MAPE (34.22%). Consequently, RF was used for spatial-temporal distribution maps PM2.5 concentrations of time of month and seasonal differences on regional scale (Jan-Jun), yielding monthly an $R^2$ of 0.74–0.85, RMSE 12.13–21.02 μg/m³, and MAPE 20.29–25.54%, respectively. Overall, the regional estimate for the entire study period achieved an $R^2$ of 0.81, RMSE of 14.45 μg/m³, and MAPE of 21.25%.

## 4.2 Spatial-temporal patterns

RF showed superior performance in areas with elevated PM2.5 from agricultural burning and forest fires. Spatial analysis revealed higher concentrations in north and south zones compared to central areas, with peaks during the dry season (January-April) driven by biomass burning, and in late May, concentrations drop except in cities where traffic and energy consumption rather than wildfires are primary sources. The RF model's positive interaction with $CO_2$ and biomass combustion confirms separate urban and highland sources.

## 4.3 Correlation & feature importance

Correlation analysis identified relationships between PM2.5 correlated negatively with precipitation, air pressure, and humidity, but positively with wind speed and satellite data. Feature importance analysis identified Aerosol Optical Depth (AOD), TOTEXTTAU, DUSMASS25, and Carbon Monoxide (CO) as the most significant variables for the RF model.

## 4.4 Technical limitations

Despite the high performance of the models, several technical limitations persist. First, there is a significant data gap in AOD-MAIAC retrieval during the wildfire season, primarily due to cloud masking and heavy smoke interference. Second, the limited number of ground-based monitoring stations in Chiang Mai restricts the model's calibration and validation in remote mountainous areas. Lastly, the temporal resolution of satellite overpasses may not fully capture the rapid diurnal fluctuations of PM2.5 concentrations during extreme burning events.

## 4.5 Overall and recommendations

The study's findings underscored the importance of accurately estimating PM2.5 concentrations for understanding air quality dynamics in the region. However, limitations such as uneven distribution of monitoring stations were noted, highlighting the need for further research to improve the understanding of PM2.5 pollution in Chiang Mai Province. The RF model can support zone-level monitoring based on spatial risk and early warning systems by integrating WRF/PBLH forecasts and real-time thermal anomalies. However, the limited number of stations makes accurate monitoring in mountainous areas difficult. Therefore, an in-situ low-cost sensor network (e.g., PurpleAir, DustBoy) is recommended to enhance data density, and refined model calibration and analyzing spatial autocorrelation (Moran's I) of residuals are recommended for future calibration.

ISSN 3027-8325 (Online)
DOI: 10.63271/rmuttorj.v18i2.265288

## 5. References

Bisri, A. M., & Wahono, R. S. (2015). Penerapan Adaboost untuk Penyelesaian Ketidakseimbangan Kelas pada Penentuan Kelulusan Mahasiswa dengan Metode Decision Tree. *Journal of intelligent systems, 1*, 27–32.

Bowman, D. M. J. S., Balch, J. K., Artaxo, P., Bond, W. J., Carlson, J. M., Cochrane, M. A., D'Antonio, C. M., DeFries, R. S., Doyle, J. C., Harrison, S. P., Johnston, F. H., Keeley, J. E., Krawchuk, M. A., Kull, C. A., Marston, J. B., Moritz, M. A., Prentice, I. C., Roos, C. I., Scott, A. C., Swetnam, T. W., van der Werf, G. R., & Pyne, S. J. (2009). Fire in the Earth System. *Science, 324*(5926), 481-484. https://doi.org/doi:10.1126/science.1163886

Chang, X., Xing, Y., Gong, W., Yang, C., Guo, Z., Wang, D., Wang, J., Yang, H., Xue, G., & Yang, S. (2023). Evaluating gross primary productivity over 9 ChinaFlux sites based on random forest regression models, remote sensing, and eddy covariance data. *Science of The Total Environment, 875*, Article 162601. https://doi.org/10.1016/j.scitotenv.2023.162601

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

ChooChuay, C., Pongpiachan, S., Tipmanee, D., Suttinun, O., Deelaman, W., Wang, Q., Xing, L., Li, G., Han, Y., & Palakun, J. (2020). Impacts of PM2. 5 sources on variations in particulate chemical compounds in ambient air of Bangkok, Thailand. *Atmospheric Pollution Research, 11*(9), 1657–1667.

Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., & Koutrakis, P. (2013). A critical assessment of high resolution aerosol optical depth (AOD) retrievals for fine particulate matter (PM) predictions. *Atmospheric Chemistry and Physics, 13*(21), 10907–10917.

da Silva Chagas, C., de Carvalho Junior, W., Bhering, S. B., & Calderano Filho, B. (2016). Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena, 139*, 232–240.

Department of National Parks, Wildlife and Plant Conservation. (2023). *The final report wildfire suppression statistics project 2023*. https://portal.dnp.go.th/Content/firednp?contentId=15705

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., & Schwartz, J. (2016). Assessing PM2. 5 exposures with high spatiotemporal resolution across the continental United States. *Environmental science & technology, 50*(9), 4712–4721.

Dontree, S., Prapamontol, T., Chantara, S., & Pardthaisong, L. (2018). *Smog in the North: Severity, effects, causes and solutions*. Health Systems Research Institute. [in Thai]

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., & Reichle, R. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate, 30*(14), 5419–5454.

Geng, G., Murray, N. L., Chang, H. H., & Liu, Y. (2018). The sensitivity of satellite-based PM2.5 estimates to its inputs: Implications to model development in data-poor regions. *Environment International, 121*, 550–560. https://doi.org/10.1016/j.envint.2018.09.051

Geo-Informatics and Space Technology Development Agency. (2023). *Summary of forest fires and smog situations with satellite images*. https://www.gistda.or.th/more_news.php?c_id=580&lang=TH

Giglio, L., Boschetti, L., Roy, D. P., Humber, M. L., & Justice, C. O. (2018). The Collection 6 MODIS burned area mapping algorithm and product. *Remote sensing of environment, 217*, 72–85.

Gu, J., Wang, Y., Ma, J., Lu, Y., Wang, S., & Li, X. (2022). An estimation method for PM2.5 based on aerosol optical depth obtained from remote sensing image processing and meteorological factors. *Remote Sensing, 14*(7), Article 1617. https://doi.org/10.3390/rs14071617

Inthacha, S. (2011). *The climatology of Thailand and future climate change projections using the regional climate model PRECIS* [Doctoral dissertation, University of East Anglia].

Jiao, L., Zhang, B., Xu, G., & Zhao, S. (2016). Spatio-temporal variability of correlation between aerosol optical depth and PM2.5 concentration. Journal of Arid Land Resources and Environment, 30, 34–39.

Junpen, A., Garivait, S., & Bonnet, S. (2013). Estimating emissions from forest fires in Thailand using MODIS active fire product and country specific data. *Asia-Pacific Journal of Atmospheric Sciences, 49*, 389–400.

Kanabkaew, T., & Kim Oanh, N. T. (2011). Development of spatial and temporal emission inventory for crop residue field burning. *Environmental Modeling & Assessment, 16*, 453–464.

Kawichai, S., Sripan, P., Rerkasem, A., Rerkasem, K., & Srisukkham, W. (2025). Long-term retrospective predicted concentration of PM2.5 in upper northern Thailand using machine learning models. *Toxics, 13*(3), Article 170. https://doi.org/10.3390/toxics13030170

Kirtphaiboon, S., Wongwises, P., Limsakul, A., Sooktawee, S., & Humphries, U. (2014). Rainfall variability over Thailand related to the El nino-southern oscillation (ENSO). *J. Sustain. Energy Environ, 5*, 37–42.

Kleist, D. T., Parrish, D. F., Derber, J. C., Treadon, R., Errico, R. M., & Yang, R. (2009). Improving incremental balance in the GSI 3DVAR analysis system. *Monthly Weather Review, 137*(3), 1046–1060.

Li, Y., Sha, Z., Tang, A., Goulding, K., & Liu, X. (2023). The application of machine learning to air pollution research: A bibliometric analysis. *Ecotoxicology and Environmental Safety, 257*, 114911. https://doi.org/10.1016/j.ecoenv.2023.114911

Lin, L., Liang, Y., Liu, L., Zhang, Y., Xie, D., Yin, F., & Ashraf, T. (2022). Estimating PM2.5 concentrations using the machine learning RF-XGBoost model in Guanzhong urban agglomeration, China. *Remote Sensing, 14*(20), Article 5239. https://doi.org/10.3390/rs14205239

Maurya, N. K., Pandey, P. C., Sarkar, S., Kumar, R., & Srivastava, P. K. (2022). Spatio-temporal monitoring of atmospheric pollutants using Earth observation Sentinel 5P TROPOMI data: Impact of stubble burning a case study. *ISPRS International Journal of Geo-Information, 11*(5), Article 301. https://doi.org/10.3390/ijgi11050301

Molod, A., Takacs, L., Suarez, M., & Bacmeister, J. (2015). Development of the GEOS-5 atmospheric general circulation model: Evolution from MERRA to MERRA2. *Geoscientific Model Development, 8*(5), 1339-1356.

Narita, D., Oanh, N. T. K., Sato, K., Huo, M., Permadi, D. A., Chi, N. N. H., Ratanajaratroj, T., & Pawarmart, I. (2019). Pollution Characteristics and Policy Actions on Fine Particulate Matter in a Growing Asian Economy: The Case of Bangkok Metropolitan Region. *Atmosphere, 10*(5), Article 227. https://doi.org/10.3390/atmos10050227

Phairuang, W., Suwattiga, P., Chetiyanukornkul, T., Hongtieab, S., Limpaseni, W., Ikemori, F., Hata, M., & Furuuchi, M. (2019). The influence of the open burning of agricultural biomass and forest fires in Thailand on the carbonaceous components in size-fractionated particles. *Environmental Pollution, 247*, 238–247.

Pinichka, C., Makka, N., Sukkumnoed, D., Chariyalertsak, S., Inchai, P., & Bundhamcharoen, K. (2017). Burden of disease attributed to ambient air pollution in Thailand: A GIS-based approach. *PLOS ONE, 12*(12), Article e0189909. https://doi.org/10.1371/journal.pone.0189909

Pollution Control Department. (2023). *Report air pollution and PM2.5 concentration of Thailand 2022-2023*. https://pcd.gdcatalog.go.th/dataset/air-quality-bangkok-metropolitan-region

Punsompong, P., Pani, S. K., Wang, S.-H., & Pham, T. T. B. (2021). Assessment of biomass-burning types and transport over Thailand and the associated health risks. *Atmospheric Environment, 247*, Article 118176. https://doi.org/10.1016/j.atmosenv.2020.118176

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., & Kim, G.-K. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate, 24*(14), 3624–3648.

Rodrigues, J. A., Libonati, R., Pereira, A. A., Nogueira, J. M., Santos, F. L., Peres, L. F., Santa Rosa, A., Schroeder, W., Pereira, J. M., & Giglio, L. (2019). How well do global burned area products represent fire patterns in the Brazilian Savannas biome? An accuracy assessment of the MCD64 collections. *International Journal of Applied Earth Observation and Geoinformation, 78*, 318–331.

Rodriguez-Galiano, V., Mendes, M. P., Garcia-Soldado, M. J., Chica-Olmo, M., & Ribeiro, L. (2014). Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Science of the Total Environment, 476*, 189–206.

Sayeed, A., Lin, P., Gupta, P., Tran, N., Buchard, V., & Chirstopher, S. (2022). *Hourly and daily PM2.5 estimations using MERRA-2: A machine learning approach*. Earth and Space Science Open Archive. https://doi.org/10.1002/essoar.10511086.1

Son, R., Stratoulias, D., Kim, H. C., & Yoon, J.-H. (2023, 2023/10/01/). Estimation of surface PM2.5 concentrations from atmospheric gas species retrieved from TROPOMI using deep learning: Impacts of fire on air pollution over Thailand. *Atmospheric Pollution Research, 14*(10), 101875. https://doi.org/https://doi.org/10.1016/j.apr.2023.101875

Thai Meteorological Department. (2023). *Daily weather and meteorological report for Thailand*. https://www.tmd.go.th

Wang, Y., Yuan, Q., Li, T., Tan, S., & Zhang, L. (2021). Full-coverage spatiotemporal mapping of ambient PM2.5 and PM10 over China from Sentinel-5P and assimilated datasets: Considering the precursors and chemical compositions. *Science of The Total Environment, 793*, Article 148535. https://doi.org/10.1016/j.scitotenv.2021.148535

Wei, J., Li, Z., Guo, J., Sun, L., Huang, W., Xue, W., Fan, T., & Cribb, M. (2019). Satellite-derived 1-km-resolution PM1 concentrations from 2014 to 2018 across China. *Environmental Science & Technology, 53*(22), 13265–13274.

Weichenthal, S. A., Godri Pollitt, K., & Villeneuve, P. J. (2013). PM 2.5, oxidant defence and cardiorespiratory health: A review. *Environmental Health, 12*, Article 49. https://doi.org/10.1186/1476-069X-12-49

Wu, W.-S., Purser, R. J., & Parrish, D. F. (2002). Three-dimensional variational analysis with spatially inhomogeneous covariances. *Monthly Weather Review, 130*(12), 2905–2916.

You, W., Zang, Z., Zhang, L., Li, Y., Pan, X., & 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), Article 184. https://doi.org/10.3390/rs8030184