International Journal of GEOMATE, March, 2026 Vol.30, Issue 139, pp.116-125
ISSN: 2186-2982 (P), 2186-2990 (O), Japan, DOI: https://doi.org/10.21660/2026.139.g15298
Geotechnique, Construction Materials and Environment

MACHINE LEARNING BASED BIAS CORRECTION OF CMAQ USING EMISSION SOURCE DECOUPLED FEATURES IN THAILAND

*Supitcha Sukprasert[1], Pongpisit Thanasutives[2], Shin Araki[3], Katsushige Uranishi[4], Tomohito Matsuo[5] and Hikari Shimadera[6]

[1,3,5,6]Graduate School of Engineering, The University of Osaka, Japan; [2]RIKEN Center for Advanced Intelligence Project (AIP), Japan; [4]Department of Life and Environment Engineering, The University of Kitakyushu, Japan

*Corresponding Author, Received: 29 May 2025, Revised: 26 Dec. 2025, Accepted: 28 Dec. 2025

ABSTRACT: Machine learning (ML) models are commonly used to correct biases in chemical transport model simulations of $PM_{2.5}$ from multiple emission sources. However, $PM_{2.5}$ simulations are typically treated as a single predictor in ML models, limiting insight into how individual sources influence their predictions. In this study, we decomposed $PM_{2.5}$ concentrations, simulated by the Community Multiscale Air Quality (CMAQ) model, into individual contributions from biomass burning (BB), anthropogenic (AT), and other sources. These three source contributions were used as predictors in a Light Gradient Boosting Machine (LightGBM) and interpreted via Shapley Additive exPlanation (SHAP) values to diagnose their influence on $PM_{2.5}$ predictions over Thailand. The proposed model improved $PM_{2.5}$ prediction compared to the original CMAQ model. SHAP analysis suggested that the BB contribution was the most important predictor, followed by AT and other contributions, with mean absolute SHAP values of 8.43, 3.69, and 2.37 μg/m^3, respectively. On average, the BB contribution increased the predicted values by 4.68 ± 13.45 μg/m^3 in the dry season and decreased them by 7.44 ± 1.84 μg/m^3 in the wet season, relative to the model's expected output (around 24.56 μg/m³). SHAP interaction analysis suggested that CMAQ overestimation of $PM_{2.5}$ during high pollution episodes may stem from inaccurate BB and AT emissions. Findings highlight the need to prioritize refinement of the BB emission inventory (e.g., by tuning emission factors) to reduce $PM_{2.5}$ overestimation.

Keywords: Bias correction, Chemical transport model, Machine learning, $PM_{2.5}$, SHAP

1. INTRODUCTION

Fine particulate matter ($PM_{2.5}$) has remained a major air quality concern in Thailand for a decade. In 2024, the maximum daily average reached 218.6 μg/m³, far exceeding the national standard of 37.5 μg/m³ [1]. Epidemiological studies show that prolonged $PM_{2.5}$ exposure is linked to increased risks of respiratory diseases [2] and mortality risk from cardiovascular diseases [3]. Health Impact Assessment (HIA) studies build upon these findings to quantify the population health burden caused by $PM_{2.5}$ exposure. However, these studies often rely on data from sparse air quality monitoring (AQM) networks, resulting in low-resolution data (e.g., province-level averages) that limit detailed analysis [4]. Accurate and high-resolution $PM_{2.5}$ data are therefore critical not only for supporting HIA but also for spatially targeted air quality management and urban planning.

Chemical transport models (CTMs), such as the Community Multiscale Air Quality (CMAQ) model, are widely used to estimate atmospheric pollutants, including $PM_{2.5}$, and can help address the spatial limitations of sparse AQM networks. However, they are prone to biases due to model inputs (e.g., emission data and meteorology) and model parameterizations (e.g., chemical mechanisms and deposition). For example, the Fire INventory from NCAR (FINN) version 1.5 has been reported to cause underestimation of CTM-simulated $PM_{2.5}$, due to limitations in fire detection [5] and outdated emission factors [6], underscoring the need for post-processing bias correction.

Several methods have been applied to reduce biases in CTM outputs, including data assimilation techniques (e.g., Kalman filter [7]) and statistical models (e.g., regression-based models [8-9]). In recent decades, machine learning (ML) models have gained popularity. ML models adjust CTM outputs using observed data, similar to simple regression models, but they can also incorporate additional variables (e.g., meteorology and land-use data) and learn complex nonlinear relationships among them, often achieving higher predictive accuracy [10]. Common ML models used are tree-based models (TBMs) [10-11] and deep learning (DL) [12]. TBMs, such as Light Gradient Boosting Machine (LightGBM), are particularly favored for their ease of use and strong predictive capabilities [13].

Emission inventories are one of the major sources of CTM biases. However, existing studies usually treat CTM-simulated $PM_{2.5}$ concentrations from multiple emission source inputs as a single predictor [10], leaving the role of $PM_{2.5}$ source contributions in ML models unexplored. Using CTM-based source contributions as predictors allows ML models to be interpreted in terms of emission sources, providing an

opportunity to diagnose source-specific biases and potentially inform improvements to emission inventories.

In Thailand, the major sources of $PM_{2.5}$ include biomass burning (BB) and anthropogenic (AT) emissions [6]. Accordingly, this study aims to (1) decompose three source contributions from CMAQ-simulated $PM_{2.5}$ concentrations, namely AT, BB, and other sources, and assess their impact on LightGBM model performance; and (2) interpret their influence on $PM_{2.5}$ estimates via SHapley Additive exPlanations (SHAP) [14]. Other variables (e.g., meteorological and land-use data) are excluded so that SHAP attributions purely reflect the influence of CMAQ-based predictors, which constitutes the key novelty of this work.

The paper is organized as follows. Section 2 describes the research significance. Section 3 presents the methodology, including $PM_{2.5}$ data screening, CMAQ model configurations, the proposed approach for emission source decoupling, and the LightGBM development. Section 4 presents LightGBM model performance, SHAP analysis results, high-resolution $PM_{2.5}$ maps, and discussion. Finally, we conclude the study in Section 5.

2. RESEARCH SIGNIFICANCE

This study advances ML-based bias correction by incorporating CTM-based $PM_{2.5}$ source contributions as model predictors and interpreting their effects using SHAP. The analysis reveals the relative importance and directional influence of predictors on $PM_{2.5}$ estimates, as well as interaction effects between pairs of predictors, allowing diagnosis of source-specific biases. Interpretation based on source contributions improves the transparency of ML-based bias correction, an aspect often overlooked in previous studies, and offers guidance for refining emission inventories in CTM simulations.

3. METHODOLOGY

3.1 Study area and measured $PM_{2.5}$ data

The study area, Thailand, was partitioned into five regions (Central, East, North, Northeast, and South) as defined by the Thai Meteorological Department (Fig. 1). Measured $PM_{2.5}$ data were obtained from the Pollution Control Department and Bangkok Metropolitan Administration for 2019 to 2021. Hourly $PM_{2.5}$ data were aggregated to daily averages only for days with at least 75% data availability, to preserve daily representativeness in the presence of diurnal fluctuations, which can arise from traffic-related emissions, particularly in Bangkok [15], as well as other factors. This threshold is consistent with the U.S. Environmental Protection Agency's criteria for 24-h $PM_{2.5}$ averages [16]. Hourly $PM_{2.5}$ over 450 $\mu g/m^3$, accounting for around 0.00047% of the dataset, were considered anomalies via histogram-based distributional analysis and removed before averaging. Finally, AQM data from a station in a given year were excluded if the annual coverage of daily $PM_{2.5}$ samples fell below 70%, to maintain annual representativeness. This excluded data from 38, 23, and 7 stations in 2019, 2020, and 2021, respectively, with most exclusions occurring in the data-abundant Central region. Thus, it is unlikely to introduce additional spatial bias. The final screened data included N=103,639 samples from 125 stations.

3.2 WRF-CMAQ simulation

CMAQ v5.3.3 [17] was employed to simulate $PM_{2.5}$ concentrations. Meteorological fields used as CMAQ input were simulated by the Weather Research and Forecasting (WRF) model v.4.3 [18]. The WRF-CMAQ was set up with two nested domains that cover Asia (D1) with a 45-km grid and Thailand (D2) with a 15-km grid resolution, the latter of which serves as the study area. Original 15-km CMAQ simulations were resampled using a bilinear interpolation to achieve the desired 1-km resolution. The LightGBM model was tasked with performing bias correction after resampling. The WRF model configurations followed the methodology in [19], and details of CMAQ settings are provided in Table 1.

The CMAQ model was run in three cases: C1) all emissions included, C2) AT emission excluded in D2, and C3) BB emission excluded. Table 2 presents how emission contributions were decoupled and their labels. The other contribution represents $PM_{2.5}$ contributions from emission sources not perturbed in the simulations, including AT emissions outside D2, biogenic, and volcanic emissions. Numerical noise associated with differencing CMAQ simulations was assessed using the mean absolute value of negative source contributions divided by simulated $PM_{2.5}$ concentration and was found to be small.

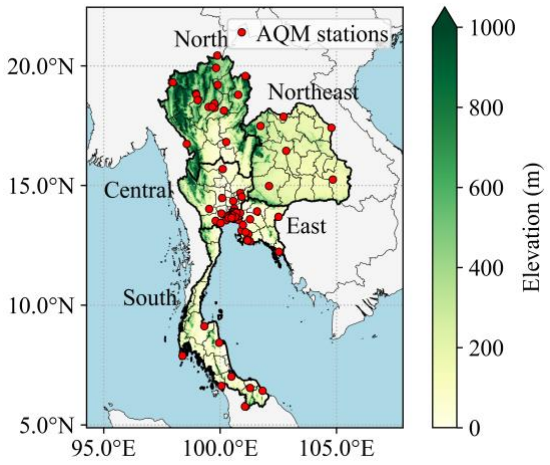

Fig. 1 The study area

Table 1 CMAQ model configurations

	Configuration
Vertical resolution	30 vertical layers (top pressure of 100hPa, first layer thickness 30 m)
D1 boundary condition	WACCM[a] [20]
Advection	PPM[b]/WRF-based scheme
Diffusion	Multiscale/ACM2[c] [21-22]
Gas-phase chemistry	SAPRC07tc[d] [23]
Aerosol process	AERO6[e]
Dry deposition	M3DRY
Wet deposition	RADM[f]
Emission inventories	AT: REAS[g] v3.2.1 [24], HTAP[h] v2.2 [25], Biogenic: MEGAN[i]v2.04 [26], BB: FINN v2.5 [27], Volcano: Carn et al. [28]

Note: [a] Whole Atmosphere Community Climate Model; [b] Piecewise Parabolic Method; [c] Asymmetrical Convective Model Version 2; [d] State Air Pollution Research Center Version 07tc; [e] The 6th generation CMAQ aerosol module; [f] Regional Acid Deposition Model; [g] Regional Emission inventory in Asia; [h] Hemispheric Transport of Air Pollution; [i] Model of Emissions of Gases and Aerosols from Nature

Table 2 Summary of CMAQ simulation cases and source decoupling calculations

		Label
CMAQ running case		
C1	All emissions included	CMAQ_PM25
C2	AT emissions excluded in D2	CMAQ_PM25_No_AT
C3	BB emissions excluded	CMAQ_PM25_No_BB
Decoupling individual emission contribution		
C1–C2	AT contribution	CMAQ_PM25_AT
C1–C3	BB contribution	CMAQ_PM25_BB
C3–(C1– C2)	Other contribution	CMAQ_PM25_Othr

CMAQ_PM25 was used as the baseline predictor, whereas CMAQ_PM25_AT, CMAQ_PM25_BB, and CMAQ_PM25_Othr were used collectively as multiple source-decoupled predictors (hereafter referred to as AT-BB predictors). Fig. 2 illustrates daily averages of measured $PM_{2.5}$ concentrations, along with baseline and AT-BB predictors across all AQM stations.

3.3 LightGBM model development

The data were spatially split by AQM stations into training (80%) and test (20%) sets. Stratified random sampling was applied to maintain proportional representation of stations from each region in both sets. This splitting process was repeated five times with different random seeds to obtain reliable results. Two LightGBM models (baseline and AT-BB models) were developed for each seed, with names reflecting CMAQ predictor types.

For each seed, hyperparameters (HPs) were tuned independently on the training set using the baseline predictor and Bayesian optimization implemented in the Python library Scikit-Optimize [29]. The search space for key HPs was identical across all seeds (max_depth = 3–6, learning_rate = 0.01–0.3, n_estimators = 100–500, and num_leaves = 8–64). The tuning process employed 5-fold spatial cross-validation (CV), in which the data were split based on stations. Optimal HPs were applied to validate the two models using spatial CV and overall CV (sample-based splitting). Performance metrics included the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Model performance was evaluated over the full period, covering all data samples, and across wet (mid-May to mid-October) and dry (mid-October to mid-May) seasons.

Both CV approaches were repeated 50 times with different randomizations. The resulting 50 sets of scores for the two models were statistically compared using a paired t-test at a confidence level (CI) of 95%. Since the statistical test involves five seeds, five p-values were generated. However, partial overlap among training samples across seeds led to dependent tests. The harmonic mean method defined in Eq. (1) was applied to appropriately combine five p-values into a single combined p-value (p_c) [30]. In Eq. (1), w_i represents the weight assigned to the p-value (p) from seed i, where $\sum_{i=1}^{n} w_i = 1$. In this study, all weights were set equally to $1/n$, where n denotes the total number of seeds.

$$p_c = \frac{\sum_{i=1}^{n} w_i}{\sum_{i=1}^{n} \frac{w_i}{p_i}} \quad (1)$$

Reported validation scores were averaged across 50 repeated CVs and five seeds. For each seed, the two models were retrained on the entire training set and then used to predict the hold-out test set. Reported test scores are means across five seeds.

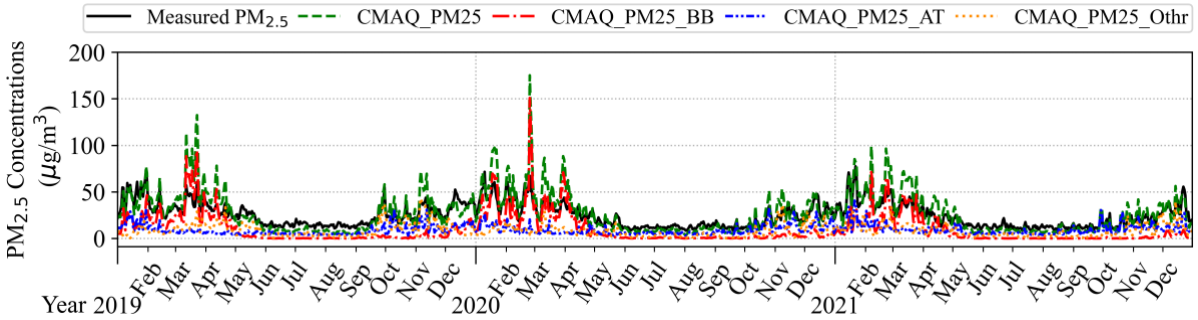

Fig. 2 Daily average of measured and CMAQ-simulated $PM_{2.5}$ across all AQM stations from 2019 to 2021

3.4 SHAP analysis

SHAP was applied to quantify feature contributions to the AT-BB model's estimates $f(x)$. Grounded in cooperative game theory, SHAP estimates Shapley values as defined in Eq. (2).

$$\phi_f^j = \sum_{S \subseteq M \setminus \{j\}} w_S [f_{S \cup \{j\}}(x_{s \cup \{j\}}) - f_S(x_S)] \quad (2)$$

ϕ_f^j denotes the Shapley value for feature j, M is the set of all features, and S is a subset of features. $f_{S \cup \{j\}}(x_{S \cup \{j\}})$ and $f_S(x_S)$ represent the model estimates with and without the feature j, respectively. The weighting term $w_S = \frac{|S|!(|M|-|S|-1)!}{|M|!}$ ensures fair average marginal contributions among features.

SHAP belongs to the class of additive feature attribution methods, which construct an explanation model $g(\tilde{x})$ that approximates $f(x)$ for any i-th instance [14]. The explanation model represents a linear combination of feature contributions as shown in Eq. (3).

$$g(\tilde{x}) = \phi^0 + \sum_{j=1}^{|M|} \phi^j \tilde{x}^j \quad (3)$$

In Eq. (3), $\tilde{x}^j$ is a simplified feature where $\tilde{x}^j \in \{0,1\}^{|M|}$, ϕ^0 is the base value (the expected model estimate), ϕ^j denotes a SHAP value for feature j, where $\phi^j \in \mathbb{R}$. SHAP values for CMAQ_PM25_BB, CMAQ_PM25_AT, and CMAQ_PM25_Othr are denoted as ϕ^{BB}, ϕ^{AT}, and ϕ^{Othr}, respectively.

Various algorithms have been developed to estimate Shapley values, such as DeepSHAP, and TreeSHAP, each tailored to different model architectures. TreeSHAP, which provides computationally efficient and exact computation of SHAP values for TBMs [31], was used in this study. The tree-path dependent approach was selected as the feature perturbation method. For each seed, the AT-BB model was retrained on the entire training set, and TreeSHAP was then applied to compute SHAP values on the corresponding test set. The resulting five sets of SHAP values were averaged across overlapping samples to produce final SHAP values for interpretation.

For a global interpretation, SHAP values for a feature j were aggregated across all test instances ($i = 1, \ldots, N_{test}$) to obtain its mean contribution ($\bar{\phi}^j$), as defined in Eq. (4). This aggregation provides a global direction of influence. Computing means of absolute SHAP values ($\bar{\phi}_+^j$) across all test instances, as defined in Eq. (5), yields a measure of global feature importance. The Standard Deviation (SD) of $\bar{\phi}^j$ and $\bar{\phi}_+^j$ is reported as ±SD and is also shown as error bars in the bar plot (Fig. 5).

$$\bar{\phi}^j = \mathbb{E}_i[\phi_i^j] \quad (4)$$

$$\bar{\phi}_+^j = \mathbb{E}_i[|\phi_i^j|] \quad (5)$$

4. RESULTS AND DISCUSSION

4.1 The LightGBM model performance

4.1.1 Validation performance

As shown in Table 3, the AT-BB model significantly outperformed the baseline model for both overall and spatial CV ($p_c < 0.05$). This improvement arises from the source decoupling, which reveals source-specific patterns and helps the model better identify and correct CMAQ biases. For example, CMAQ_PM25_BB shows a consistent seasonal pattern over three years (Fig. 2), with higher $PM_{2.5}$ levels in the dry season and lower levels in the wet season due to reduced agricultural burning and more frequent rainfall. This seasonal pattern enhances the effectiveness of the source decoupling, particularly in the dry season when $PM_{2.5}$ concentrations and variability are higher.

Interestingly, when using only CMAQ-simulated $PM_{2.5}$ as a predictor, the spatial and overall CV scores were nearly identical. This contrasts with usual trends, where spatial CV scores are usually lower due to location generalization challenges. Removing other features, such as meteorological and land-use data, simplifies the model and can reduce the risk of overfitting caused by spatial patterns learned from these features.

Table 3 LightGBM model performance evaluated on the validation set

Period	Model	Overall CV			Spatial CV		
		R^2	RMSE (μg/m^3)	MAE (μg/m^3)	R^2	RMSE (μg/m^3)	MAE (μg/m^3)
Full period	Baseline	0.67	10.82	6.99	0.66	10.94	7.04
	AT-BB	0.72*	9.92*	6.45*	0.71*	10.18*	6.55*
Wet season	Baseline	0.44	5.46	4.14	0.43	5.53	4.25
	AT-BB	0.47*	5.33*	4.03*	0.45*	5.40*	4.11*
Dry season	Baseline	0.58	13.43	9.06	0.57	13.58	9.07
	AT-BB	0.65*	12.21*	8.20*	0.64*	12.55*	8.32*

Note: A Star superscript (*) indicates that the AT-BB model performed significantly better than the baseline model at the 95% CI ($p_c < 0.05$)

Table 4 LightGBM model performance evaluated on the test set

Period	Model	R^2	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)
Full period	CMAQ	0.07	17.70	9.98
	Baseline	0.66	10.75	6.90
	AT-BB	0.71	9.94	6.44
Wet season	CMAQ	-0.08	7.40	5.77
	Baseline	0.38	5.44	4.17
	AT-BB	0.42	5.29	4.05
Dry season	CMAQ	-0.22	22.38	13.03
	Baseline	0.58	13.32	8.88
	AT-BB	0.64	12.25	8.13

4.1.2 Test performance

Test scores of both LightGBM models closely aligned with validation scores, indicating that the models did not overfit to the training set (Table 4). The CMAQ model, evaluated on the same test set, performed worse than both LightGBM models. The scatter plot of measured versus predicted $PM_{2.5}$ (Fig. 3) shows that both LightGBM models reduced severe overestimation of the CMAQ model at high $PM_{2.5}$ concentrations (> 100 μg/m³). Nevertheless, at extreme concentrations (> 200 μg/m³), predictions of both LightGBM models no longer increased with observations. This is because the models were intentionally constrained (i.e., low maximum depth) to reduce overfitting, given the small number of predictors, and trained on a right-skewed $PM_{2.5}$ distribution. Consequently, their ability to generalize to rare extreme events is limited. Future work, which incorporates more features and complex models, could improve performance at extreme $PM_{2.5}$ concentrations.

4.2 SHAP analysis

4.2.1 SHAP value trends

To facilitate visualization of the SHAP value trends, mean SHAP values were computed within feature j value bins ($\bar{\phi}_b^j$) rather than displayed as individual scatter points (Fig. 4). For features with relatively low value ranges (CMAQ_PM25_AT and CMAQ_PM25_Othr), the bin width was uniformly set to 20. CMAQ_PM25_BB exhibited a wider value range. Unequal bin widths were therefore applied: a bin size of 20 for values below 100 μg/m³ and 200 for values exceeding 100 μg/m³.

All decoupled predictors displayed similar trends in their $\bar{\phi}_b$ (Fig. 4). The higher predictor values were associated with higher $\bar{\phi}_b$. Notably, $\bar{\phi}_b^{BB}$ plateaued beyond 200 μg/m³, providing further evidence of the model's limited ability to generalize at extreme $PM_{2.5}$ levels.

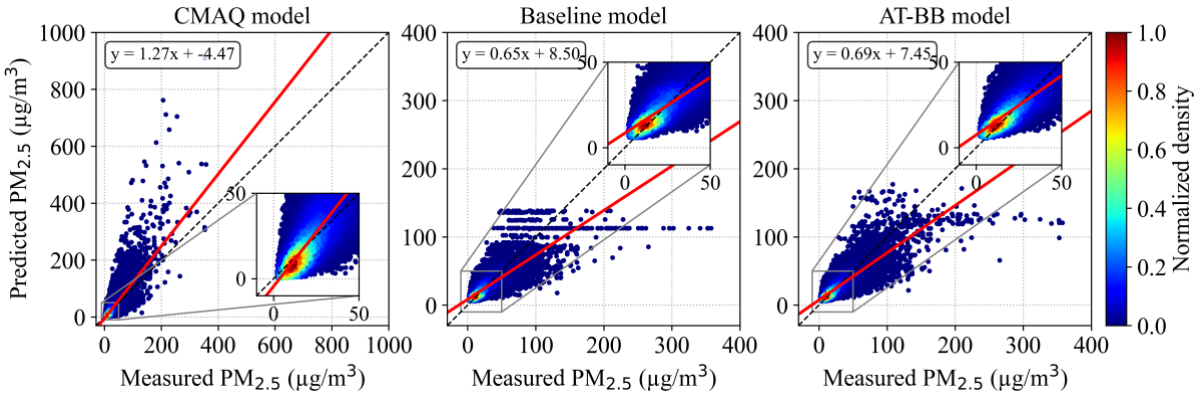

Fig. 3 Scatter plot between measured and predicted $PM_{2.5}$ of CMAQ, baseline, and AT-BB models on test set

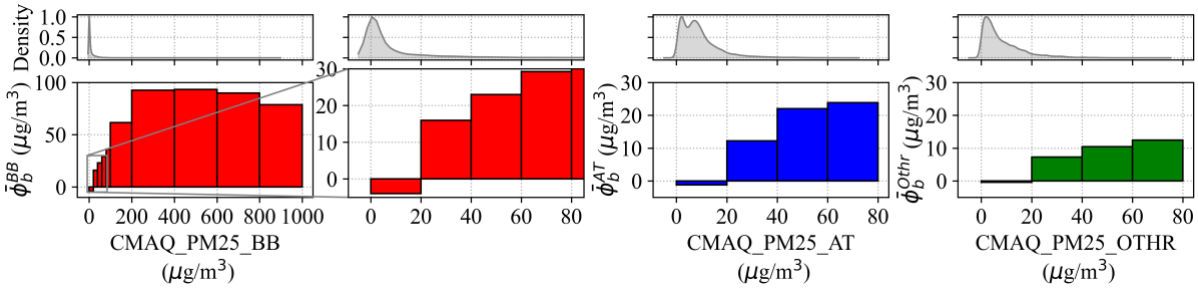

Fig. 4 Bar plot between mean SHAP values and corresponding CMAQ predictor values. The upper subplots show the normalized density distributions of the predictors obtained via kernel density estimation

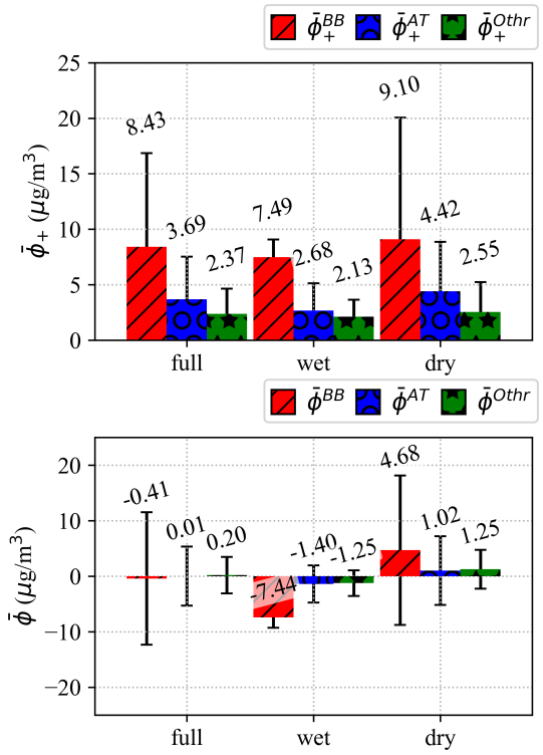

Fig. 5 $\bar{\phi}_+$ of CMAQ predictors (upper) and $\bar{\phi}$ of CMAQ predictors (lower)

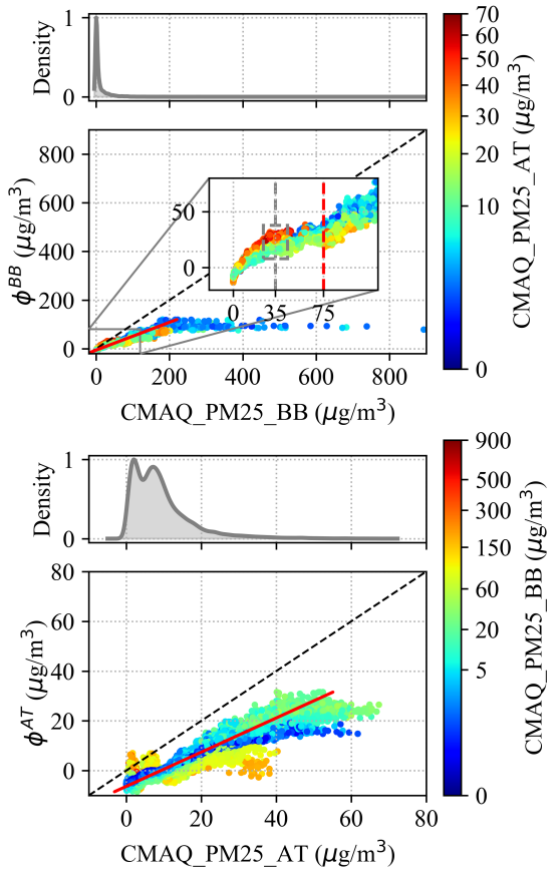

Fig. 6 SHAP dependence plot for CMAQ_PM25_BB (upper) and CMAQ_PM25_AT (lower)

4.2.2 Global interpretation of SHAP values

CMAQ_PM25_BB was the most important predictor based on its $\bar{\phi}_+$ for the full period, followed by CMAQ_PM25_AT and CMAQ_PM25_Othr as illustrated in the upper plot of Fig. 5. The wet and dry seasons also presented the same order. Note that the SD values of $\bar{\phi}_+^{BB}$, $\bar{\phi}_+^{AT}$ and $\bar{\phi}_+^{Othr}$ across five seeds were 0.27, 0.13, and 0.09 µg/m³, respectively, which are very small compared to the mean values of 8.43, 3.69, and 2.37 µg/m³, respectively, indicating that the predictor ranking is stable across seeds.

For the direction of influence, seasonal averaging was preferred because it grouped SHAP values into more homogeneous ranges and mitigated cancellations, as shown in the lower plot of Fig. 5. During the wet season, all predictors contributed negatively to $PM_{2.5}$ estimates on average, reducing them relative to ϕ^0 (~24.56 µg/m³). CMAQ_PM25_BB showed the strongest contribution, followed by CMAQ_PM25_AT and CMAQ_PM25_Othr. Conversely, in the dry season, they exhibited positive contributions on average, with CMAQ_PM25_BB again leading, followed by CMAQ_PM25_Othr and CMAQ_PM25_AT. Overall, CMAQ_PM25_BB emerged as the most influential predictor, consistent with its strong seasonal variability.

CMAQ_PM25_AT and CMAQ_PM25_Othr showed similar mean SHAP values, and they had lower contributions than CMAQ_PM25_BB in both seasons. This is due to their relatively lower concentrations and lack of a distinctive temporal pattern, unlike CMAQ_PM25_BB. Despite similar mean SHAP values, ϕ^{AT} yielded a greater SD compared to ϕ^{Othr} (Fig. 5). One possible reason is spatial variability of CMAQ_PM25_AT, which tends to be higher in densely populated areas such as Bangkok in Central Thailand. This spatial pattern makes CMAQ_PM25_AT more useful in predicting $PM_{2.5}$ in these areas.

4.2.3 Feature interaction effects

Feature interaction effects could be another reason for the higher SD of ϕ^{AT}. SHAP interaction values, derived from the Shapley interaction index for measuring how pairs of features jointly influence model estimates, were computed to quantify this effect. Further details on SHAP interaction values can be found in [31].

Among all feature pairs, the interaction between CMAQ_PM25_BB and CMAQ_PM25_AT was found to be the strongest. Fig. 6 visualizes this interaction. For example, in the upper plot, a vertical spread in ϕ^{BB} can be observed at a CMAQ_PM25_BB value of 35 µg/m³. Investigating the color distribution revealed that higher CMAQ_PM25_AT values were observed with increased ϕ^{BB}. This positive relationship generally held for CMAQ_PM25_BB values below roughly 75

μg/m³, suggesting a synergistic effect in which both predictors jointly increase $PM_{2.5}$ estimates. When CMAQ_PM25_BB values exceeded roughly 75 μg/m³, this relationship reversed: higher values of CMAQ_PM25_AT were associated with lower ϕ^{BB}. A similar pattern can be observed in the lower plot of Fig. 6.

CMAQ_PM25_BB was the most important predictor and exhibited relatively high variability in ϕ^{BB}. Because it also showed a strong interaction with CMAQ_PM25_AT, this interaction likely contributed to the relatively high variability observed in ϕ^{AT}.

4.2.4 SHAP values vs. CMAQ's biases

Although SHAP values do not provide direct quantification of biases in the CMAQ model, they offer insights into how the AT-BB model adjusts its outputs, which can be interpreted to form hypotheses about potential biases in the CMAQ-simulated $PM_{2.5}$ by emission source. The observed interaction between CMAQ_PM25_BB and CMAQ_PM25_AT suggests that the AT-BB model learns to use these predictors cooperatively to correct CMAQ biases. During severe pollution episodes, characterized by elevated values of both predictors, the model assigns relatively lower ϕ^{BB} and ϕ^{AT}. This pattern suggests that the model is correcting overestimation in the CMAQ simulations, which may arise from inaccurate BB and AT emissions. This finding is consistent with a previous WRF-Chem modeling study over Thailand, which reported that FINN v2.5 (BB emissions used in this study) tends to overestimate $PM_{2.5}$ concentrations, despite improvements over FINN v1.5 [6]. As a result, refining emission inventories, particularly for BB (the most important predictor), could enhance the accuracy of CMAQ simulations. Moreover, this interaction suggests that future ML frameworks adopting the source decoupling technique in areas with mixed emission sources should consider decoupling both BB and AT emissions to improve bias correction efficiency.

4.3 High-resolution $PM_{2.5}$ mapping

4.3.1 Prediction maps

For a given seed, two LightGBM models were retrained on the full training set to generate daily $PM_{2.5}$ estimates at 1×1 km² resolution for the entire study area. Daily estimates were first averaged across five seeds and then averaged over the full period and by seasons. Differences were calculated as AT-BB estimates minus baseline estimates (AT-BB – Baseline). Additionally, differences between each LightGBM model and the CMAQ model (LightGBM – CMAQ) were also computed.

The AT-BB model predicted higher $PM_{2.5}$ concentrations in heavily polluted regions, particularly in the Northern region, with this pattern amplified during the dry season (Fig. 7). Conversely, the AT-BB model predicted slightly lower concentrations across other regions. During the wet season, the difference between the AT-BB and baseline models was minimal. The AT-BB model predicted slightly lower concentrations across most regions, with the Southern region showing the largest decrease.

Both LightGBM models generally predicted lower $PM_{2.5}$ concentrations across the Central to Northern regions compared to the CMAQ model, with this effect being more pronounced during the dry season. In the Northern region, where $PM_{2.5}$ levels were particularly elevated during the dry season, mainly due to BB emissions, the CMAQ model tended to substantially overestimate $PM_{2.5}$ concentrations. The LightGBM models helped reduce this overestimation.

4.3.2 Coefficient of variation (CoV) maps

Since both models were run across five random seeds, with varying data splits and HPs, epistemic uncertainty in these factors was assessed using CoV. Fig. 8 presents spatial CoV maps for three periods: the full period, dry, and wet seasons. Although the AT-BB model exhibited higher CoV values than the baseline model, most remained below 5%. Relatively higher CoV values were observed in Thailand-Myanmar border areas during the wet season. The maximum CoV, however, did not exceed 10.92%, indicating good model stability. Combined with the improvements in R^2, RMSE, and MAE, these results suggest that the AT-BB model enhances predictive accuracy while maintaining stable performance.

The stability of model outputs across seeds demonstrates that $PM_{2.5}$ spatial patterns derived from CMAQ-decoupled predictors are robust to variations in training conditions. Future works can adopt CMAQ-decoupled predictors as robust inputs, and when supplemented with meteorological and land-use predictors, they can support the development of accurate and high-resolution $PM_{2.5}$ maps. To support reliable decision-making in applications, such as exposure assessment, urban planning, and air quality management, future works should further estimate intervals of $PM_{2.5}$ predictions, accounting for both epistemic and aleatoric (model and data) uncertainties.

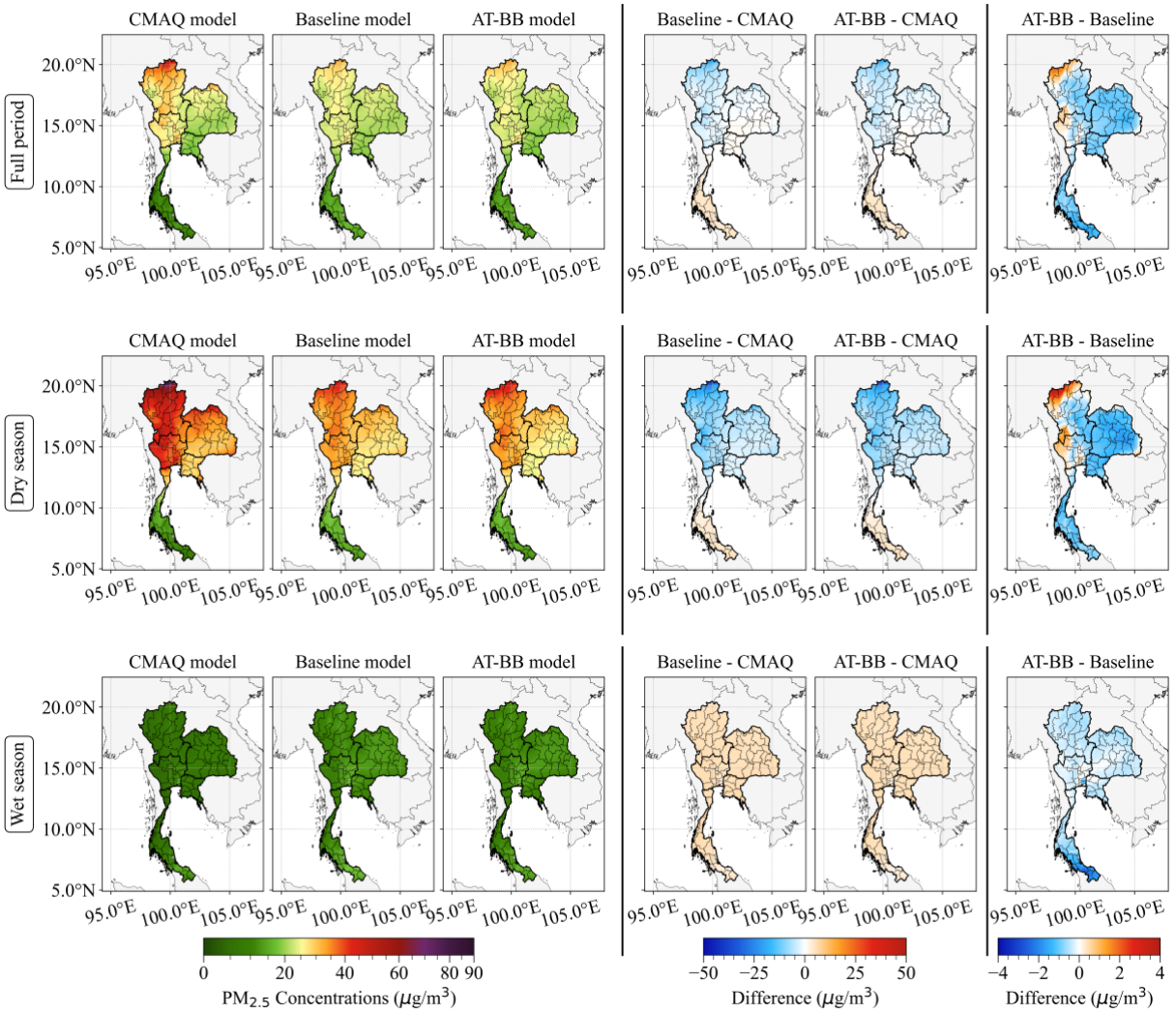

Fig. 7 Spatial distribution of $PM_{2.5}$ estimates from the CMAQ, baseline, AT-BB models, and their pairwise differences in full period, dry, and wet seasons

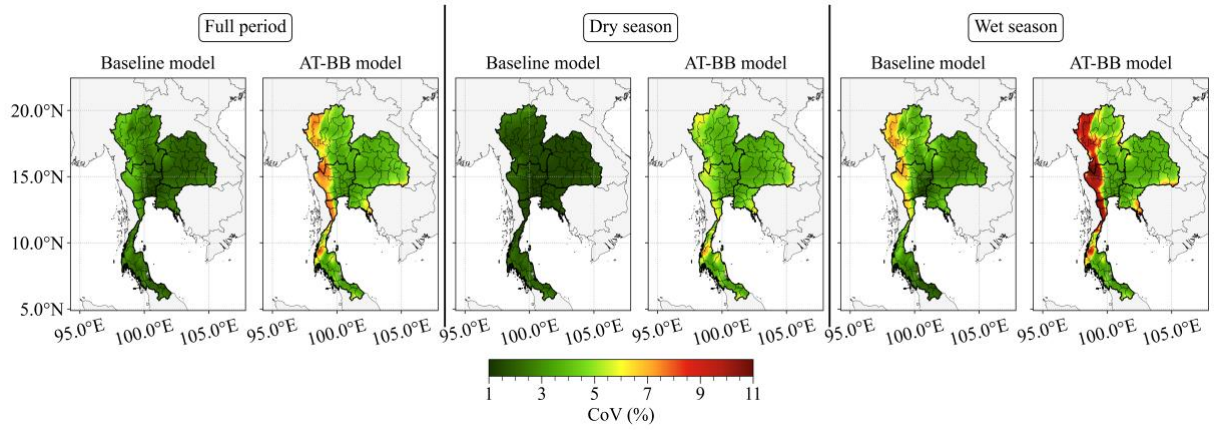

Fig. 8 Spatial distribution of %CoV of the baseline and AT-BB models across five random seeds in full period, dry, and wet season

5. CONCLUSION

This study developed an interpretable LightGBM model to correct CMAQ-simulated $PM_{2.5}$ in Thailand by using three source-decoupled predictors: CMAQ_PM25_BB, CMAQ_PM25_AT, and CMAQ_PM25_Othr. The model showed strong predictive performance and effectively reduced CMAQ biases, particularly in the dry season. SHAP analysis revealed the dominant influence of CMAQ_PM25_BB and suggested that CMAQ

overestimation during high pollution episodes is linked to both AT and BB emissions. Findings highlight the effectiveness of source decoupling in improving bias correction of CMAQ simulations and suggest that priority should be given to refining BB emission inventories (e.g., adjusting emission factors) to reduce $PM_{2.5}$ overestimations.

Because this analysis relied on a specific set of emission inventories, future work could evaluate alternative emission inventories, particularly for BB emissions given their dominant role as model predictors. Moreover, the use of bilinear interpolation to resample data to 1-km resolution may introduce spatial artifacts in the resampled fields, primarily due to the smoothing effect that can distort local concentration patterns. Therefore, sensitivity analyses using different interpolation methods and target resolutions are warranted. Finally, extensions to DL can be explored, with careful consideration of trade-offs between predictive performance, model complexity, and interpretability.

6. ACKNOWLEDGEMENTS

This work was supported by the Environment Research and Technology Development Fund of the Environmental Restoration and Conservation Agency of Japan [grant number JPMEERF20215005], and the JSPS KAKENHI, Japan [grant number JP22H0375].

7. REFERENCES

1. Pollution Control Department, Thailand State of Pollution Report 2024, Bangkok, AP CONNEX Co., Ltd., 2025, pp. 1-250. (In Thai)
2. Yan M., Ge H., Zhang L., Chen X., Yang X., Liu F., Shan A., Liang F., Li X., Ma Z., Dong G., Liu Y., Chen J., Wang T., Zhao B., Zeng Q., Lu X., Liu Y., and Tang N., Long-term $PM_{2.5}$ exposure in association with chronic respiratory diseases morbidity: A cohort study in Northern China, Ecotoxicology and Environmental Safety, Vol. 244, 2022, 114025. DOI: 10.1016/j.ecoenv.2022.114025
3. Du Y., Xu X., Chu M., Guo Y., and Wang J., Air particulate matter and cardiovascular disease: the epidemiological, biomedical and clinical evidence, Journal of Thoracic Disease, Vol. 8, Issue 1, 2016, pp. E8–E19. DOI: 10.3978/j.issn.2072-1439.2015.11.37
4. Mueller W., Vardoulakis S., Steinle S., Loh M., Johnston H. J., Precha N., Kliengchuay W., Sahanavin N., Nakhapakorn K., Sillaparassamee R., Tantrakarnapa K., and Cherrie J. W., A health impact assessment of long-term exposure to particulate air pollution in Thailand, Environmental Research Letters, Vol. 16, Issue 5, 2021, 055018. DOI: 10.1088/1748-9326/abe3ba
5. Chen L., Gao Y., Ma M., Wang L., Wang Q., Guan S., Yao X., and Gao H., Striking impacts of biomass burning on $PM_{2.5}$ concentrations in Northeast China through the emission inventory improvement, Environmental Pollution, Vol. 318, 2023, 120835. DOI: 10.1016/j.envpol.2022.120835
6. Thongsame W., Henze D. K., Kumar R., Barth M., and Pfister G., Evaluation of WRF-Chem $PM_{2.5}$ simulations in Thailand with different anthropogenic and biomass-burning emissions, Atmospheric Environment: X, Vol. 23, 2024, 100282. DOI: 10.1016/j.aeaoa.2024.100282
7. Tang X., Zhu J., Wang Z. F., and Gbaguidi A., Improvement of ozone forecast over Beijing based on ensemble Kalman filter with simultaneous adjustment of initial conditions and emissions, Atmospheric Chemistry and Physics, Vol. 11, Issue 24, 2011, pp. 12901–12916. DOI: 10.5194/acp-11-12901-2011
8. Van Donkelaar A., Martin R. V., Spurr R. J. D., and Burnett R. T., High-Resolution Satellite-Derived $PM_{2.5}$ from Optimal Estimation and Geographically Weighted Regression over North America, Environmental Science & Technology, Vol. 49, Issue 17, 2015, pp. 10482–10491. DOI: 10.1021/acs.est.5b02076
9. Guillas S., Tiao G. C., Wuebbles D. J., and Zubrow A., Statistical diagnostic and correction of a chemistry-transport model for the prediction of total column ozone, Atmospheric Chemistry and Physics, Vol. 6, Issue 2, 2006, pp. 525–537. DOI: 10.5194/acp-6-525-2006
10. Thongthammachart T., Araki S., Shimadera H., Eto S., Matsuo T., and Kondo A., An integrated model combining random forests and WRF/CMAQ model for high accuracy spatiotemporal $PM_{2.5}$ predictions in the Kansai region of Japan, Atmospheric Environment, Vol. 262, 2021, 118620. DOI: 10.1016/j.atmosenv.2021.118620
11. Bi J., Knowland K. E., Keller C. A., and Liu Y., Combining Machine Learning and Numerical Simulation for High-Resolution $PM_{2.5}$ Concentration Forecast, Environmental Science & Technology, Vol. 56, Issue 3, 2022, pp. 1544–1556. DOI: 10.1021/acs.est.1c05578
12. Singh D., Choi Y., Park J., Salman A. K., Sayeed A., and Song C. H., Deep-BCSI: A deep learning-based framework for bias correction and spatial imputation of $PM_{2.5}$ concentrations in South Korea, Atmospheric Research, Vol. 301, 2024, 107283. DOI: 10.1016/j.atmosres.2024.107283
13. Dai H., Huang G., Wang J., Zeng H., and Zhou F., Spatio-Temporal Characteristics of $PM_{2.5}$ Concentrations in China Based on Multiple Sources of Data and LUR-GBM during 2016–2021, International Journal of Environmental

Research and Public Health, Vol. 19, Issue 10, 2022, 6292. DOI: 10.3390/ijerph19106292
14. Lundberg S., and Lee S.-I., A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, Vol. 30, 2017, pp. 4765–4774.
15. Phanukarn P., BLACK CARBON IN PM2.5 AT ROADSIDE SITE IN BANGKOK, THAILAND, International Journal of GEOMATE, Vol. 19, Issue 72, 2020, DOI: 10.21660/2020.72.9245
16. U.S. Environmental Protection Agency, Appendix N to Part 50 - Interpretation of the National Ambient Air Quality Standards for $PM_{2.5}$, Title 40, 2025.
17. US EPA Office Of Research And Development, CMAQ, 2021, DOI: 10.5281/ZENODO.5213949
18. Skamarock W. C., Klemp J. B., Dudhia J., Gill D. O., Liu Z., Berner J., Wang W., Powers J. G., Duda M. G., Barker D. M., and Huang X.-Y., A Description of the Advanced Research WRF Model Version 4.3, 2019. DOI: 10.5065/1DFH-6P97
19. Chantaraprachoom N., Shimadera H., Uranishi K., Mui L. V., Matsuo T., and Kondo A., A Nation-by-Nation Assessment of the Contribution of Southeast Asian Open Biomass Burning to $PM_{2.5}$ in Thailand Using the Community Multiscale Air Quality-Integrated Source Apportionment Method Model, Atmosphere, Vol. 15, Issue 11, 2024, 1358. DOI: 10.3390/atmos15111358
20. Atmospheric Chemistry Observations & Modeling/National Center for Atmospheric Research/University Corporation for Atmospheric Research, Whole Atmosphere Community Climate Model (WACCM) Model Output, 2020. DOI: 10.5065/G643-Z138
21. Pleim J. E., A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part I: Model Description and Testing, Journal of Applied Meteorology and Climatology, Vol. 46, Issue 9, 2007, pp. 1383–1395. DOI: 10.1175/JAM2539.1
22. Pleim J. E., A Combined Local and Nonlocal Closure Model for the Atmospheric Boundary Layer. Part II: Application and Evaluation in a Mesoscale Meteorological Model, Journal of Applied Meteorology and Climatology, Vol. 46, Issue 9, 2007, pp. 1396–1409. DOI: 10.1175/JAM2534.1
23. Carter W. P. L., Development of the SAPRC-07 chemical mechanism, Atmospheric Environment, Vol. 44, Issue 40, 2010, pp. 5324–5335. DOI: 10.1016/j.atmosenv.2010.01.026
24. Kurokawa J., and Ohara T., Long-term historical trends in air pollutant emissions in Asia: Regional Emission inventory in ASia (REAS) version 3, Atmospheric Chemistry and Physics, Vol. 20, Issue 21, 2020, pp. 12761–12793. DOI: 10.5194/acp-20-12761-2020
25. Janssens-Maenhout G., Crippa M., Guizzardi D., Dentener F., Muntean M., Pouliot G., Keating T., Zhang Q., Kurokawa J., Wankmüller R., Denier Van Der Gon H., Kuenen J. J. P., Klimont Z., Frost G., Darras S., Koffi B., and Li M., HTAP_v2.2: a mosaic of regional and global emission grid maps for 2008 and 2010 to study hemispheric transport of air pollution, Atmospheric Chemistry and Physics, Vol. 15, Issue 19, 2015, pp. 11411–11432. DOI: 10.5194/acp-15-11411-2015
26. Guenther A., Karl T., Harley P., Wiedinmyer C., Palmer P. I., and Geron C., Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmospheric Chemistry and Physics, Vol. 6, Issue 11, 2006, pp. 3181–3210. DOI: 10.5194/acp-6-3181-2006
27. Wiedinmyer C., Kimura Y., McDonald-Buller E. C., Emmons L. K., Buchholz R. R., Tang W., Seto K., Joseph M. B., Barsanti K. C., Carlton A. G., and Yokelson R., The Fire Inventory from NCAR version 2.5: an updated global fire emissions model for climate and chemistry applications, Geoscientific Model Development, Vol. 16, Issue 13, 2023, pp. 3873–3891. DOI: 10.5194/gmd-16-3873-2023
28. Carn S. A., Fioletov V. E., McLinden C. A., Li C., and Krotkov N. A., A decade of global volcanic SO_2 emissions measured from space, Scientific Reports, Vol. 7, Issue 1, 2017, 44095. DOI: 10.1038/srep44095
29. Head T., Kumar M., Nahrstaedt H., Louppe G., and Shcherbatyi I., holgern/scikit-optimize, 2024. DOI: 10.5281/ZENODO.10804382
30. Wilson D. J., The harmonic mean p -value for combining dependent tests, Proceedings of the National Academy of Sciences, Vol. 116, Issue 4, 2019, pp. 1195–1200. DOI: 10.1073/pnas.1814092116
31. Lundberg S. M., Erion G., Chen H., DeGrave A., Prutkin J. M., Nair B., Katz R., Himmelfarb J., Bansal N., and Lee S.-I., From local explanations to global understanding with explainable AI for trees, Nature Machine Intelligence, Vol. 2, Issue 1, 2020, pp. 56–67. DOI: 10.1038/s42256-019-0138-9