Atmospheric Pollution Research 14 (2023) 101875

Contents lists available at ScienceDirect

Atmospheric Pollution Research

journal homepage: www.elsevier.com/locate/apr

Estimation of surface $PM_{2.5}$ concentrations from atmospheric gas species retrieved from TROPOMI using deep learning: Impacts of fire on air pollution over Thailand

Rackhun Son [a,b,f,g], Dimitris Stratoulias [b,c,**], Hyun Cheol Kim [d,e], Jin-Ho Yoon [f,*]

[a] Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Jena, Germany
[b] Geospatial Information Department, Asian Disaster Preparedness Center (ADPC), Bangkok, Thailand
[c] SERVIR-Mekong, Bangkok, Thailand
[d] Air Resources Laboratory, National Oceanic and Atmospheric Administration, College Park, MD, USA
[e] Cooperative Institute for Satellite Earth System Studies, University of Maryland, College Park, MD, USA
[f] School of Earth Science and Environmental Engineering, Gwangju Institute of Science and Technology, Gwangju, South Korea
[g] Department of Environmental Atmospheric Sciences, Pukyong National University, South Korea

ARTICLE INFO

Keywords:
$PM_{2.5}$
TROPOMI
Deep learning
TabNet

ABSTRACT

Surface $PM_{2.5}$ concentration is routinely observed at limited number of surface monitoring stations. To overcome its limited spatial coverage, space-borne monitoring system has been established. However, it also faces various challenges such as cloud contamination and limited vertical resolution. In this study, we propose a deep learning-based surface $PM_{2.5}$ estimation method using the attentive interpretable tabular learning neural network (TabNet) with atmospheric gas species retrieved from the tropospheric monitoring instrument (TROPOMI). Unlike previous applications that primarily used decision tree-based algorithms, TabNet provides interpretable decision-making steps to identify dominant factors. By incorporating five TROPOMI products (i.e., NO_2, SO_2, O_3, CO, HCHO), we have tested the system's capability to produce surface $PM_{2.5}$ concentration without aerosol optical property, which was used more traditionally. The proposed model successfully captures spatiotemporal variations over Thailand in the period of 2018–2020, and it outperforms other leading machine learning models, particularly at high concentrations. The interpretable decision-making steps highlight that carbon monoxide is the most influential chemical component, which relates to the seasonal burning in southeast Asia. It is found that air quality impacts from fire are stronger in the northern part of Thailand and fires in neighboring countries should not be neglected. The proposed method successfully estimates surface $PM_{2.5}$ concentration without aerosol optical property, implying its potential to advance monitoring air quality over remote regions.

1. Introduction

Ambient air pollution is a critical threat to public health, causing more than three million premature fatalities worldwide in 2012 (Organization, 2016) as well as various environmental issues (Gurjar et al., 2010). Among air pollutants, fine particulate matter ($PM_{2.5}$), namely, ambient airborne particulates sized under 2.5 μm, is well known to damage human health seriously. Due to its microscopic size, $PM_{2.5}$ can affect the respiratory and cardiovascular systems, causing or worsening major illnesses such as asthma, lung cancer and heart disease (Weichenthal et al., 2013). Rising public concern about air quality urges not only reductions in air pollutants, but also improvements to air quality monitoring at the ground level to assess the health and socioeconomic impacts.

Annual mean $PM_{2.5}$ concentrations in Thailand reached 21.4 $\mu g/m^3$ in 2020, making it the 34th most polluted country in the world.[1] An

Peer review under responsibility of Turkish National Committee for Air Pollution Research and Control.
* Corresponding author.
** Corresponding author.
E-mail addresses: dimitrios.stratoulias@adpc.net (D. Stratoulias), yjinho@gist.ac.kr (J.-H. Yoon).
[1] "World Air Quality Report, 2020." Retrieved (https://www.greenpeace.org/static/planet4-romania-stateless/2021/03/d8050eab-2020-world_air_quality_report.pdf).

https://doi.org/10.1016/j.apr.2023.101875
Received 19 February 2023; Received in revised form 3 August 2023; Accepted 3 August 2023
Available online 5 August 2023

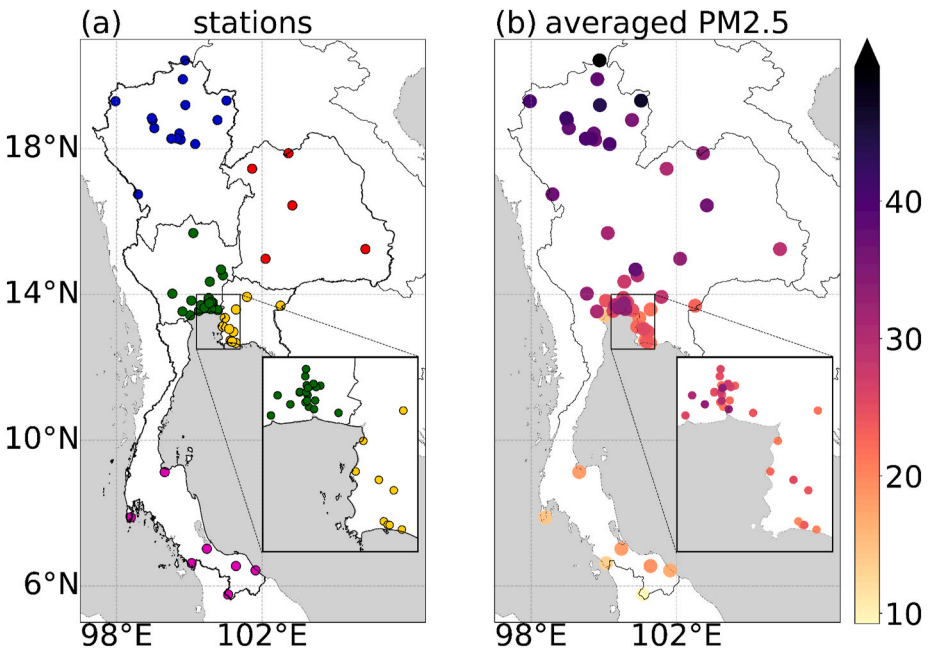

Fig. 1. Spatial distribution of (a) PM2.5 observation sites and (b) averaged PM2.5 concentrations (μg/m3) from 2018 to 2020. The stations are divided into 5 regions in each direction: north (blue), northeast (red), central (green), east (yellow) and south (magenta). The bounding box zooms up the most densely distributed area, including the capital Bangkok (100.2–101.4°E, 12.5–14.0°N).

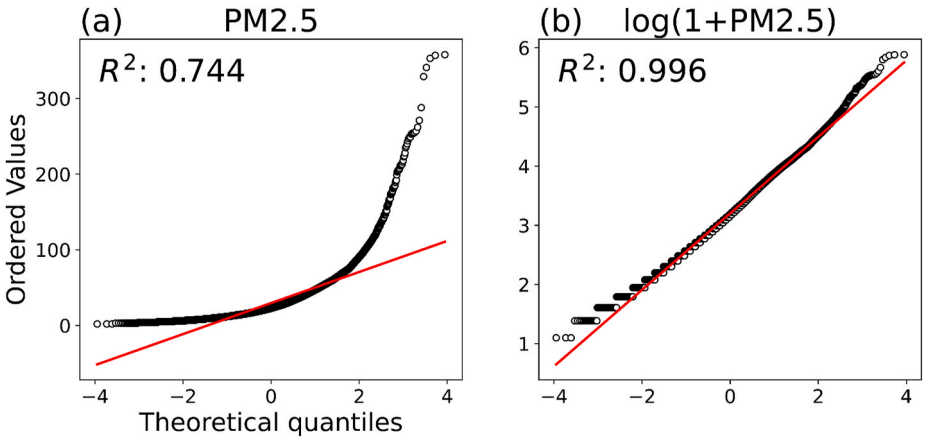

Fig. 2. The quantile-quantile (Q–Q) plot for (a) PM2.5 and (b) the log transformed PM2.5. The red line is 1:1 line between theoretical quantiles and ordered values.

estimated 40,000 deaths annually in Thailand are attributable to ambient air pollution (Pinichka et al., 2017), resulting in 0.74–1.33 million USD worth of economic costs (Vassanadumrongdee and Matsuoka, 2005). Although air pollution exposure in Thailand temporarily improved during the recent COVID-19 pandemic (Rodríguez-Urrego and Rodríguez-Urrego, 2020; Stratoulias and Nuthammachot, 2020), it remains high due to widespread smoke emissions from agricultural burnings and forest fires (Punsompong et al., 2021). $PM_{2.5}$ emissions from burning crop residue and forest fires are estimated to be 141,000 and 5000 tons per year, respectively, mostly concentrated in the central and northern regions of Thailand (Junpen et al., 2013; Kanabkaew and Kim Oanh, 2011). However, the insufficient number of in-situ $PM_{2.5}$ measurements, especially for the provinces in the north and northeast of the country (Fig. 1a), limits monitoring air quality and establishing a national plan for its management. Given that considerable financial and time resources are required to increase the number of air quality monitoring stations, satellite remote sensing-based $PM_{2.5}$ estimation is an alternative way to increase the limited spatial coverage.

Two different methods have been developed and widely used to estimate surface $PM_{2.5}$ concentration: Chemical Transport Models (CTMs) and statistical regression models. Based on physicochemical processes and atmospheric conditions, chemical transport models can approximate the quantity of air pollutants with continuous spatiotemporal coverage (Liu et al., 2004; Van Donkelaar et al., 2010). However, uncertainties in emission inventories and limited representation of chemical reactions in the ambient atmosphere remain major concerns (Shin et al., 2020). Among statistical approaches, multiple linear regression has been the most commonly applied in the early stages (Chu et al., 2016). Also, geographically weighted regression, an extension of multiple linear regression, has been proposed to assign distance-based weights to reflect spatial variability and local effects to provide regional estimations (Brunsdon et al., 1998; Jiang et al., 2017; You et al., 2016). Mixed-effect models adopt fixed and random effect terms to separate statistical relationship and variability by time and region (Kloog et al., 2012; Xie et al., 2015). In addition, the generalized additive model has been proposed to consider the nonlinear characteristics between input and target variables (Sorek-Hamer et al., 2013; Zou et al., 2017).

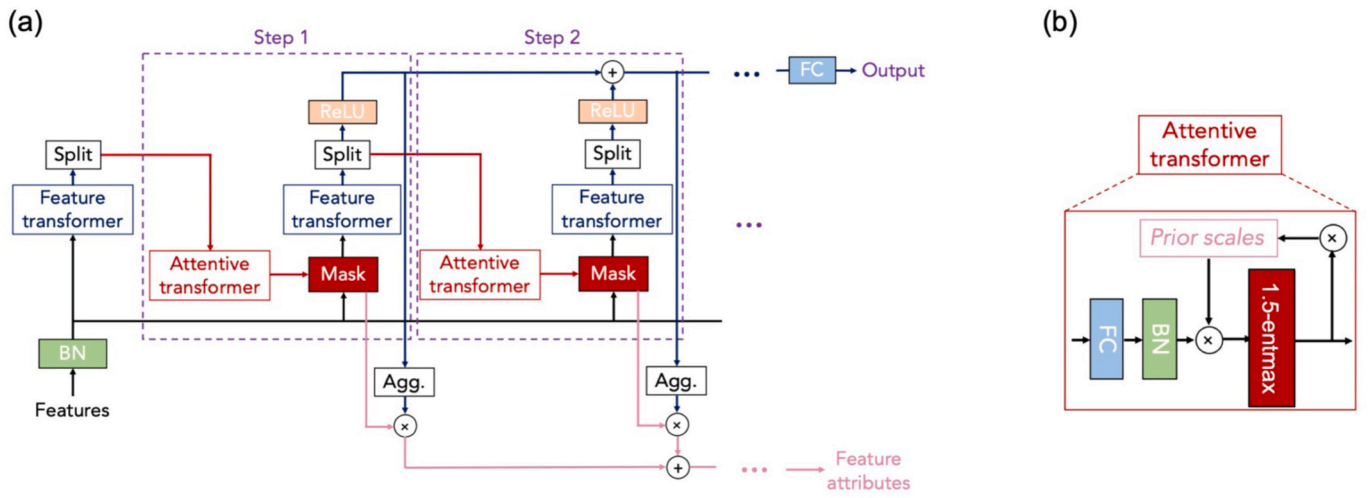

Fig. 3. Model architecture diagram (Fig. 4 from Arık and Pfister, 2021). (a) TabNet encoder, which employed as main model structure in this study. (b) A block of attentive transformer, generating a feature selection mask at each decision step.

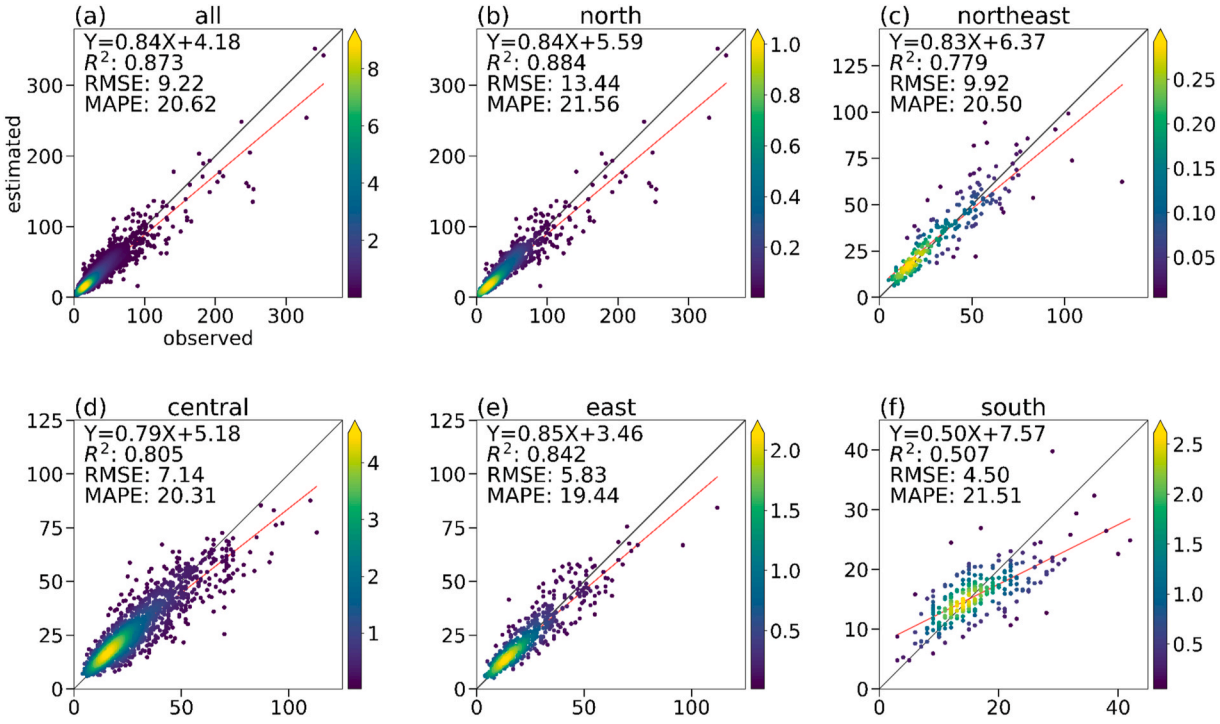

Fig. 4. Density scatterplots of validating PM2.5 estimation (μg/m3) results for (a) all regions in Thailand and the five divided regions in Fig. 1: (b) north, (c) northeast, (d) central, (e) east and (f) south. The black solid diagonal line is the best-fit line from linear regression. The red line is estimated regression line shown in the topleft of each panel.

Table 1
Comparison of model performance on the testing dataset.

	R^2	RMSE	MAPE	slope
SVM	0.821	10.96	25.37	0.76
Random Forest	0.829	10.70	22.74	0.74
CatBoost	0.857	9.78	18.89	0.78
XGBoost	0.867	9.44	19.10	0.79
LightGBM	0.868	9.42	**18.87**	0.80
TabNet	**0.873**	**9.22**	20.62	**0.84**

Machine learning (ML) algorithms have recently introduced as innovative developments in the bottom-up approaches to upscale data-driven in-situ models to spatially explicit gridded estimates. Random forest (RF), one of the most frequently applied algorithms, has further improved estimation accuracy and has higher interpretability at both national and regional scales (Chen et al., 2018; Hu et al., 2017; Wei et al., 2019). Elastic-net application has successfully expanded the spatiotemporal dimension with a large number of predictors (Xue et al., 2019). Support vector machine (SVM) can enhance spatial resolution at a 100 m scale by being merged into multiple modeling stages (de Hoogh et al., 2018). Other ML models such as Bayesian maximum entropy (Jiang and Christakos, 2018), gradient boosting machine models (Chen et al., 2019; Wang et al., 2021) and RF combined with ordinary kriging

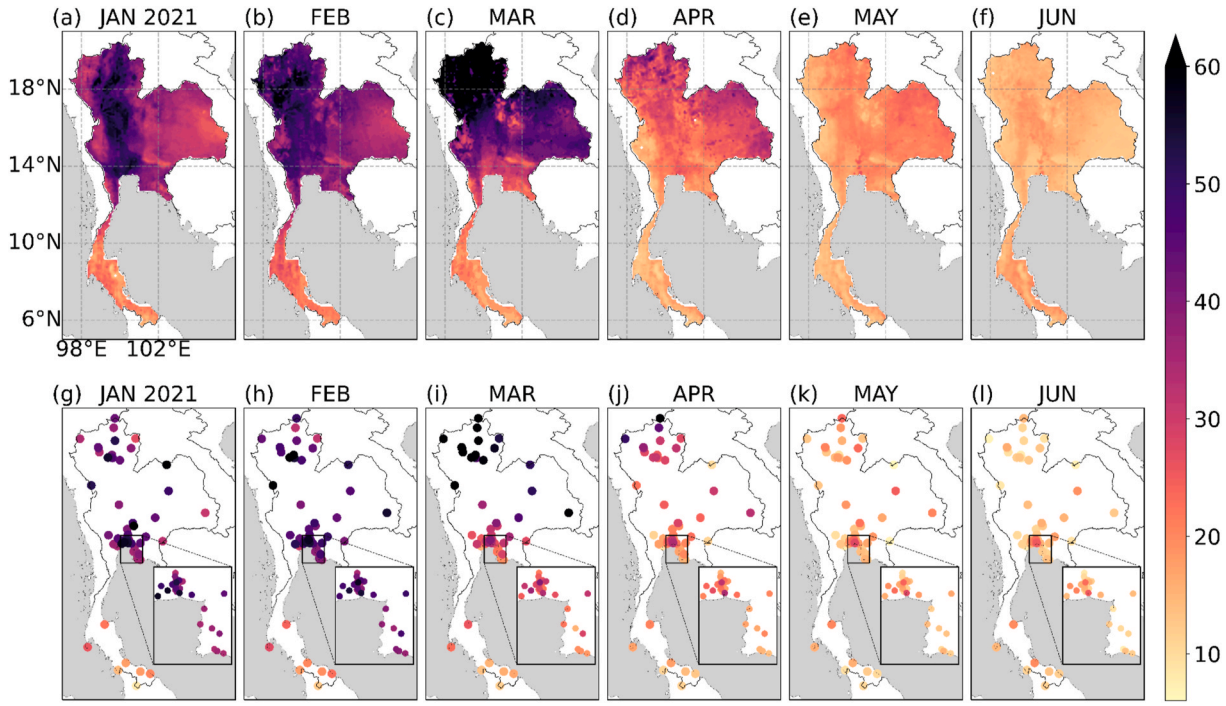

Fig. 5. (a–f) Monthly averaged spatial mapping of PM2.5 estimation and (g–l) observation for the year 2021.

(Han et al., 2022) have also been employed to incorporate satellite-derived products into ground-level observations.

As computing technology and resources have advanced, neural network-based approach has introduced deeper and wider layers, defined as deep learning (DL), and has begun to outperform classical ML models based on decision tree algorithms in various regression tasks (Devlin et al., 2018; He et al., 2016). DL based methods have also recently been attempted in remote sensing due to their high accuracy using large amounts of data (Ghahremanloo et al., 2021; Zhang et al., 2020; Zhu et al., 2020). However, compared with decision trees, the usability of this cutting-edge approach is yet to be explored in-depth for $PM_{2.5}$ satellite-based estimation.

Thus, this study aimed to develop a DL-based model to estimate daily ground-level $PM_{2.5}$ concentrations based in-situ observations in Thailand and satellite-derived atmospheric gas products. Regarding the DL network architecture, we implemented the Attentive Interpretable Tabular Learning neural network (TabNet) (Arık and Pfister, 2021), which is tailored for use with tabular datasets. We evaluated the model's performance through five different regions in Thailand and compared it with other popular machine learning algorithms such as SVM, RF, XGBoost (Chen and Guestrin, 2016), LightGBM (Ke et al., 2017) and CatBoost (Prokhorenkova et al., 2018). Furthermore, to shed light on the critical characteristics of $PM_{2.5}$ concentration in Thailand, we conduct an analysis of global/local feature selection and reasoning processes, as well as examined the impacts of fires on $PM_{2.5}$ concentration.

2. Study area and data

2.1. Study area

Thailand is located at the center of the Indochinese peninsula, has the 10th largest economy in Asia and hosts a population of almost 70 million people (Long and Ascent, 2020). The country is divided into 76 administrative provinces as primary local government units and the capital Bangkok. In this study, considering the meteorological division system in Thailand (https://www.tmd.go.th/en), we divided the country into five regions to analyze regional characteristics: north, northeast, central, east and south (Fig. 1a). Despite the low air quality in Thailand, ground monitoring stations are sparse and mostly concentrated in the central region, which contains 29 of the 67 stations used in this study (green points in Fig. 1a). For the remaining regions, 15, 5, 11 and 7 stations are distributed in the north (blue), northeast (red), east (yellow) and south (magenta), respectively.

2.2. Ground-level $PM_{2.5}$ observation

The Pollution Control Department in the Air Quality and Noise Management Bureau provides national air quality monitoring records for approximately 84 stations (as of 2021). Considering the consistency of the data availability during the experimental period from January 2018 to June 2021, we selected the daily measurements of $PM_{2.5}$ concentration from 67 stations (Fig. 1a) as the target dataset for the model training. The observed $PM_{2.5}$ concentration pattern has an exponential quantile-quantile distribution (Fig. 2a). This asymmetry can hamper model training by blurring the variance in the pollution levels over different input conditions. To transform the data to be closely fitted by a normal distribution, we thus took logs of the $PM_{2.5}$ values after adding one (Fig. 2b) and the results showed significantly higher R-squared coefficients (R^2) from 0.744 to 0.996.

2.3. TROPOMI

The Sentinel-5P mission is a precursor satellite measuring atmospheric chemical concentrations at high spatial and radiometric resolutions. The TROPOspheric Monitoring Instrument (TROPOMI) onboard Sentinel-5P is designed to record the reflectance of wavelengths using multispectral sensors. We utilized five TROPOMI products (Borsdorff et al., 2018; De Smedt et al., 2018; Garane et al., 2019; Theys et al., 2017; Van Geffen et al., 2019): the tropospheric NO_2 column (NO_2), SO_2 vertical column density at the ground level (SO_2), total atmospheric column of O_3 (O_3), vertically integrated column of CO (CO) and tropospheric formaldehyde column (HCHO); this was based on 354 of the 388 wavelength pairs. TROPOMI Level 2 products are accessible from the Copernicus Open Access Hub website (https://s5phub.copernicus.eu), and we retrieved a daily Level 3 pre-processed dataset from the Google Earth Engine using the quality assurance values of 0.75 for NO_2 and 0.5

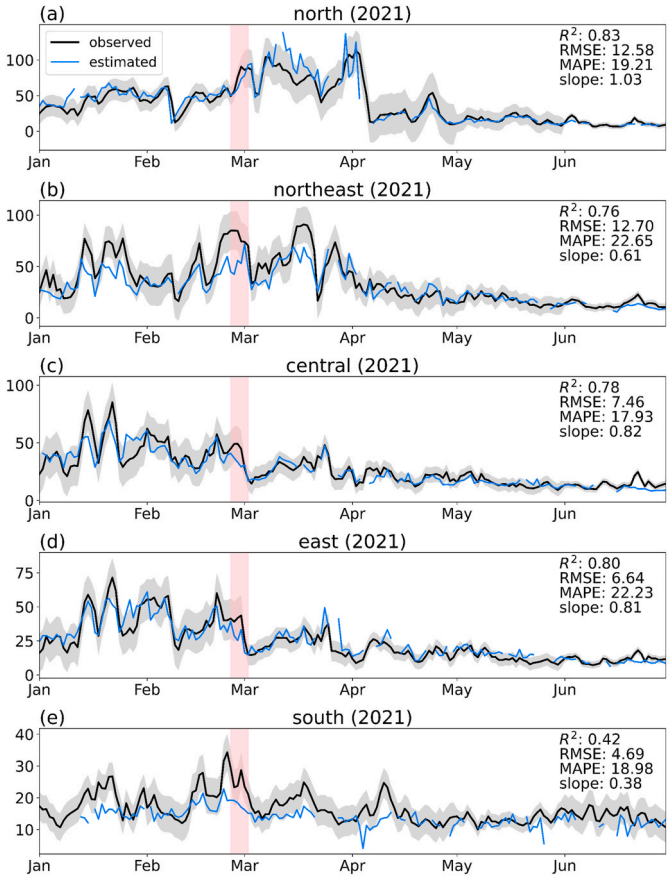

Fig. 6. Daily variations of observation (black) and estimated PM2.5 concentrations (blue) during the year 2021 in five parts of Thailand (Fig. 1a): (a) north, (b) northeast, (c) central, (d) east and (e) south. The scoring results are summarized in the topright of each panel. Gray shadings indicate one-sigma confidence interval based on monthly statistics. Red columns are marked for the period when sudden increase of fire radiative power (FRP) is detected for all the sub-regions from February 25th to March 2nd, 2021 (Figure S5-9).

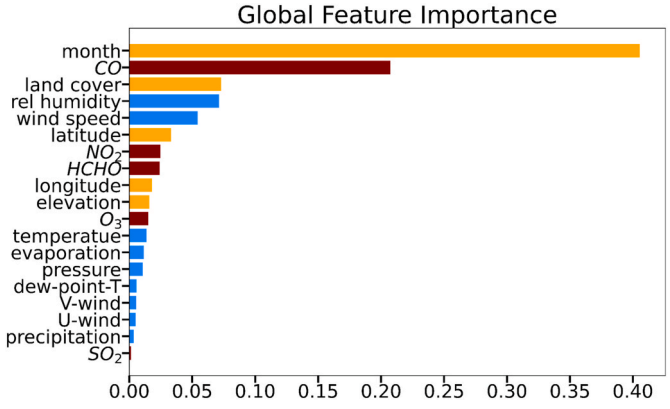

Fig. 7. Global feature importance assessment. The features are categorized in three group by colors: temporal and geospatial (orange), chemical (dark-red) and meteorological (blue) features.

for the other components except for O_3 and SO_2. Data points falling below the specified quality thresholds are considered as missing values and we exclude them from the model training and evaluation procedures.

The Sentinel-5P images were co-located with the ground station data and the values of the pixel encompassing the point location of the ground station were extracted to train the model. When the spatial mapping of $PM_{2.5}$ was inferred, the datasets were resampled to a 10 km grid to incorporate other auxiliary datasets. Subsequently, the variables, except for O_3, were transformed into a logarithmic scale similar to $PM_{2.5}$. Considering that the ranges of each variable varied, specified constants were multiplied and added before the log transform (Figs. S1a–d).

2.4. Meteorological dataset

ERA5-Land (Muñoz Sabater, 2019) provides a dataset for land components from ERA5, the fifth-generation climate reanalysis dataset provided by the Copernicus Climate Change Service at the European Centre for Medium-Range Weather Forecasts. Following previous studies (Chen et al., 2018; Wei et al., 2019), we adopted seven meteorological components from the reanalysis dataset: temperature and dew-point temperature at a 2 m height, total evaporation, surface pressure, precipitation and wind components at a 10 m height. We also approximated relative humidity and wind speed using Eqs. (1) and (2):

$$relhumidity = 100 \times \frac{e^{\frac{17.625 \times T_d}{243.04 + T_d}}}{e^{\frac{17.625 \times T}{243.04 + T}}} \quad (1)$$

$$windspeed = \sqrt{U^2 + V^2} \quad (2)$$

where T is temperature, T_d is dew-point temperature, U is the horizontal wind component (U-wind) and V is the meridional wind component (V-wind). For precipitation and wind speed, the scaled log transform was applied as mentioned above (Figs. S1e and f). Furthermore, we considered geographical factors such as elevation from ETOPO1 (Amante and Eakins, 2009) with a 1 arc-minute resolution to integrate the land topography and bathymetry and land cover classifications from GlobCover (Arino, 2010). These were categorized into 22 types based on observations from the ENVISAT satellite mission for 2009 with a spatial resolution of approximately 300 m.

3. Methodology

3.1. TabNet

TabNet is a novel neural network architecture specifically developed to handle tabular datasets effectively (Arık and Pfister, 2021). Built upon an encoder/decoder architecture, the model proficiently converts high-dimensional features into a meaningful representation using trainable embedding layers, obviating the necessity for preliminary preprocessing procedures. For instance, the layers can map categorical features into a numerical format as well as handle raw numerical features without normalizing global features. One salient strategy of the TabNet is to employ the sequential attentive transformer architecture to select the importance features in decision steps (Fig. 3a). In each step, learnable masks search for a subset of the relevant features by quantifying the contribution of the decision.

3.2. Interpretability

The feature attribution mask $\mathbf{M} \in \mathbb{R}^{B \times D}$ provides instance-wise interpretable insights for reasoning; B is the batch size and D is the dimension of the feature. At the ith decision step, the processed features from the preceding step $\mathbf{a}[i\text{-}1] \in \mathbb{R}^{B \times N_a}$ are given to an attentive transformer block (Fig. 3b), which includes a fully connected layer and batch normalization as a trainable nonlinear processing h_i. The mask is obtained through a sparse regulation function, which we set using α-entmax (Peters et al., 2019). The α-entmax is an extension of the softmax function, featuring an additional scalar parameter α that enables a continuous range of sparsity levels in the resulting probability distribution. When α is set to 1, the entmax function is equivalent to the

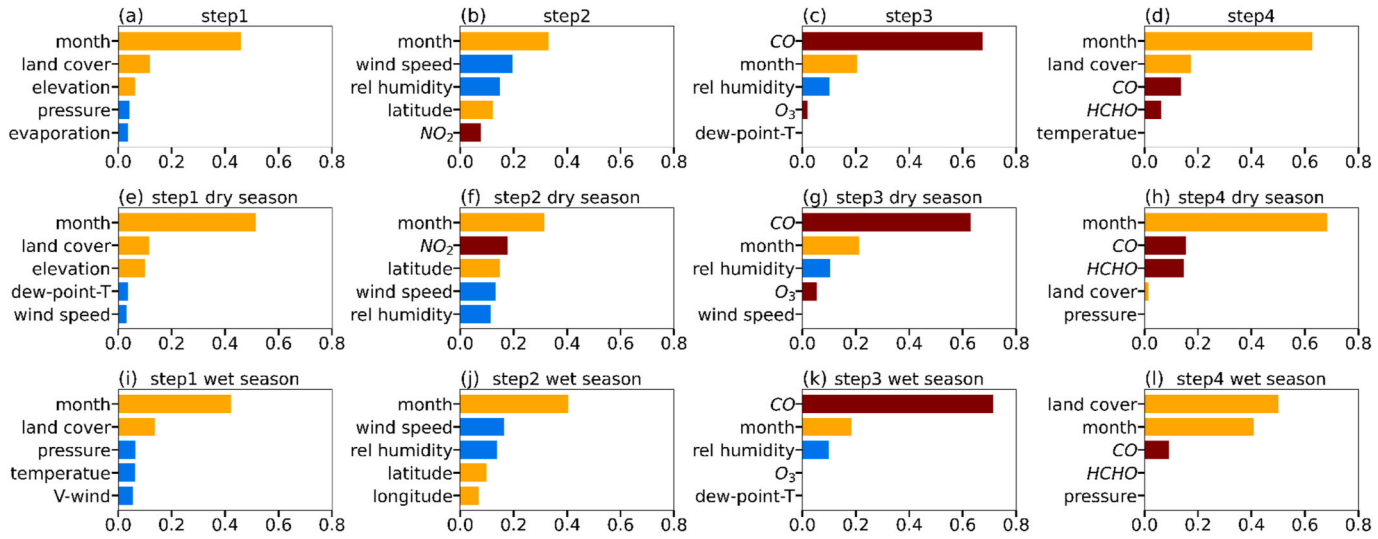

Fig. 8. Local feature importance assessment based on feature selection masks at each decision step. Top five important features are visualized with contribution ratios for (ad) all the period, (e–h) for dry season (January–March) and (i–l) for wet season (JulySeptember). The features are categorized in three group by colors: temporal and geospatial (orange), chemical (dark-red) and meteorological (blue) features.

softmax function. However, as the value of α increases, the entmax promotes sparsity in the output distribution, resembling the behavior of the sparsemax function (Martins and Astudillo, 2016) when α is set to 2. In this study, we adopt the α value of 1.5, which was validated for its efficacy in attention based sequence-to-sequence models (Peters et al., 2019). The formulation of the feature attribution mask is described in Eq. (3):

$$\mathbf{M}[i] = entmax(\mathbf{P}[i-1] \odot h_i(\mathbf{a}[i-1])) \tag{3}$$

$\mathbf{P}[i] \in \mathbb{R}^{B\times D}$ is the prior scale term to regulate the flexibility of feature selection in the multiple steps, as defined in Eq. (4):

$$\mathbf{P}[i] = \prod_{j=1}^{i} (\gamma - \mathbf{M}[j]) \tag{4}$$

where γ is a parameter to control the level of constraint on feature selection in the mask. For instance, when γ is set to 1, each feature is limited to being used at only one decision step, while increasing γ allows for more flexibility to be selected at multiple decision steps. $\mathbf{P}[0]$ is initialized as all ones, $\mathbf{1}^{B\times D}$, indicating that none of the features are used at the beginning. As a feature is considered thoroughly, its scale term is reduced to focus on the other features in the next steps. The weights of the trained mask represent the relative importance of each step in all instances. For example, if $\mathbf{M}_{b,j}[i] = 0$, then the jth feature should have no decision contribution in the ith step for the bth sample. Finally, the aggregated weights from the masks allow us to understand the importance of each feature in terms of its global behavior.

3.3. Training details

The weather in Thailand has distinct seasonality; the rainy season, which usually lasts from June to October, can significantly affect the $PM_{2.5}$ concentration in the atmosphere (Fig. S2). Moreover, the mapping of averaged $PM_{2.5}$ displays a higher concentration in the northern area, above 40, than elsewhere (Fig. 1b). Considering these spatiotemporal characteristics, we added the observed month and geographical coordinates (longitude and latitude) of the station as input features. In total, 19 input variables were used in this study: NO_2, SO_2, O_3, CO, HCHO, temperature, dew-point temperature, relative humidity, U-wind, V-wind, wind speed, precipitation, pressure, evaporation, elevation, land cover type, month, longitude and latitude. Regarding the categorical variables, specifically the month and land cover type, we assigned embedding dimensions of 6 and 17, respectively.

Following convention, we randomly split the data from 2018 to 2020 into training and testing datasets using an 80:20 ratio; the number of samples were 14,069 and 3518, respectively. We also evaluated the functionality of upscaled mapping using a 10 km resolution grid format of the input dataset for 2021. To ensure robust training, a 5-fold cross-validation was set, and the final $PM_{2.5}$ estimation was calculated by averaging the results from the five trained models. The model was implemented using the *pytorch_tabnet* package (https://github.com/dreamquark-ai/tabnet) and trained with the Adam algorithm with weight decay using a 0.01 learning rate and a batch size of 64. Following the guidelines for hyperparameters (Arık and Pfister, 2021), we optimized the depth and width of TabNet based on grid search method (Table S1): N_d=N_a = 24, N_{steps} = 4, γ = 1.3 and λ_{sparse} = 0.001.

4. Results

4.1. Evaluation of general model performance

Fig. 4 presents the accuracy of the validation results for the estimated $PM_{2.5}$ concentration in Thailand as a whole and its five subdivided regions. For the entire study domain (Fig. 4a), three evaluation metrics show 0.873 of R^2, 9.22 of root mean square error (RMSE) and 20.62 of the mean absolute percentage error (MAPE). When these results are compared with other state-of-the-art ML algorithms, R^2 and RMSE of the proposed method show the best scores (Table 1). In terms of the linear relationship between the observations and estimated $PM_{2.5}$ concentrations, all the models show slope coefficient values below 1. These results indicate a tendency for the ML models to underestimate the PM2.5 concentration, consistent with previous studies (Ma et al., 2016; Wei et al., 2019). TabNet shows the highest slope coefficient value (0.84), suggesting its ability to partially mitigate this bias.

A noticeable improvement is the enhanced accuracy in predicting high levels of $PM_{2.5}$. Fig. S3 presents a comparison of prediction performances across various thresholds of high $PM_{2.5}$ concentrations. TabNet consistently outperforms other ML models, achieving the highest R^2 scores and the lowest RMSE across all thresholds (Figs. S3a and b). TabNet also demonstrates the lowest MAPE for concentrations exceeding 75 $\mu g/m^3$ (Fig. S3c). The performance gap between TabNet and other ML models becomes more pronounced as the degree of $PM_{2.5}$ pollution increases. Moreover, TabNet exhibits superior capability in addressing underestimation of high $PM_{2.5}$ concentrations, as evidenced

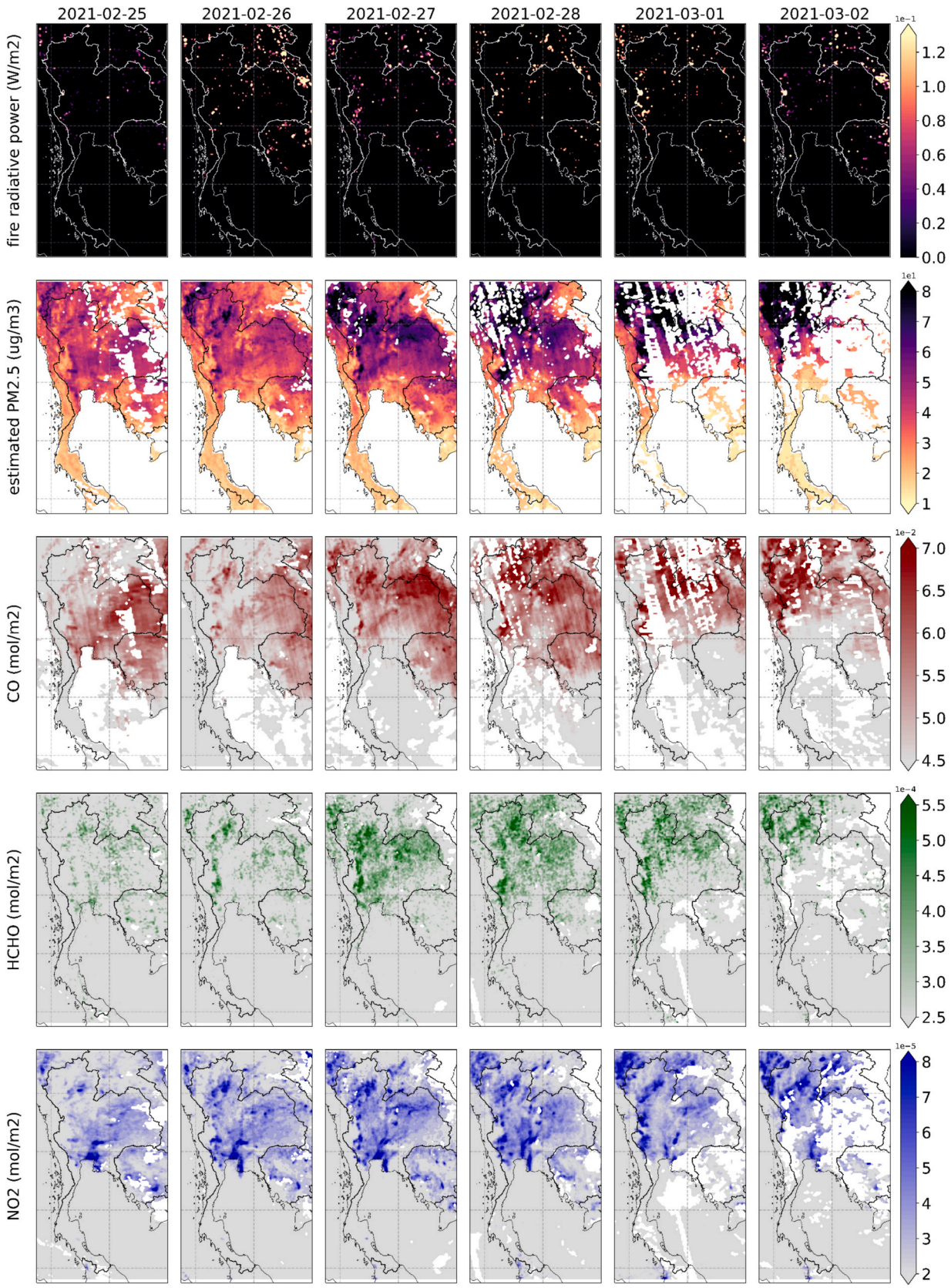

Fig. 9. The mapping of FRP and PM2.5/chemical concentrations from February 25th to March 2nd, 2021. White color is for missing values.

by its higher values of the slope coefficient (Fig. S3d).

When the scores of evaluation metrics are compared by region, the highest value of R^2 (0.884) is observed in the north (Fig. 4b). These results are consistent with the mapping of R^2 for each station showing higher than 0.8 of R^2 in all the stations in the north, including the Chiang Mai and Lampang provinces (Fig. S4a). On the other hand, the scale of biases is larger than other regions with 13.44 of RMSE, due to its wider range of the $PM_{2.5}$ concentration exceeding 300 μg/m^3 as a maximum (Fig. 4b and S4b). Given that the $PM_{2.5}$ concentrations in the north are generally higher (Fig. 1b) and extreme cases are more frequent due to agricultural burnings and forest fires (Punsompong et al., 2021), the large errors are typically caused by the underestimation mentioned previously, particularly for high concentration cases. When the regional differences in scale are diminished by considering the ration of the scale between the errors and actual values, some stations in Bangkok and neighbor cities show higher scores of MAPE (Fig. S4c). But the south region shows the lowest accuracy with 21.51 of MAPE and 0.507 of R^2 (Fig. 4f). The distinctively low slope coefficient in the south represents that its poor performance is mainly caused by underestimation (Fig. S4d). In the following section, we provide further discuss on the regional characteristics that are specific to the southern region.

4.2. Application on high-coverage mapping

One of the main purposes of employing remote sensing data is to enlarge the spatial coverage of $PM_{2.5}$ monitoring. Fig. 5 illustrates the monthly averaged results of the $PM_{2.5}$ estimation for 2021. The mapping results (Fig. 5a–f) generally agree with the observations (Fig. 5g–l) with respect to seasonal variation by region. In January, the central region of Thailand shows high levels of $PM_{2.5}$ concentrations. In the north, the concentrations significantly increase from January and peak at over 60 μg/m^3 in March. The variation in peak timing of air pollutants across region can be attributed to the seasonal differences in agricultural activities, such as harvesting and residue burning, which are carried out at varying times in each specific region (Kanabkaew and Kim Oanh, 2011). As the rainy season approaches, a decrease in pollutant concentrations is observed across all regions.

To evaluate the temporal variation of the $PM_{2.5}$, we compare the daily variations in the observed and estimated $PM_{2.5}$ over the five subdivided regions of Thailand (Fig. 6). The northern area shows the highest performance scores (0.83, 12.58 and 19.81 for R^2, RMSE and MAPE, respectively). The value of slope coefficient is almost 1 representing a significant improvement in the underestimation for extreme levels of $PM_{2.5}$, mostly within the one-sigma confidence interval for peak days during March and April. The other regions, except for the south, are also good fits with observation showing higher than 0.76 of R^2. In the southern region, despite the scale of error (4.69 of RMSE) is relatively smaller, the overall underestimation indicated by the slope value of 0.38 significantly undermines the performance across other skill scores.

The south region's extensive and encompassing coastlines contribute to its distinctive climate characteristics that differentiate it from other parts of Thailand. Firstly, the climate in the south is classified as maritime, in contrast to the prevailing continental climate in the other areas (Torsri et al., 2013). Secondly, the region is influenced by two regional monsoons: the southwest monsoon from May to October and the northeast monsoon from November to February (Manisan, 1995). Additionally, it has been known that the long-range transport of PM from Indonesian forest fires significantly influences PM levels during the dry season in southern Thailand (Mahasakpan et al., 2023; Phairuang et al., 2020). However, our dataset is limited to only 7 out of the total 67 stations in the southern region. This limitation appears to hinder progress in capturing the divers regional characteristics and comprehending the underlying mechanisms involved in estimating PM concentrations in southern Thailand.

5. Discussion

5.1. Model interpretation

Interpretability enables us to comprehend the behavior of the model at each learning step and identify important processes, thereby facilitating a more practical application. However, there isn't a perfect method to interpret ML and DL approaches, which is well recognized as a potential limitation. A major advantage of the TabNet is its attentive transformer structure, which provides post-hoc explanations by assessing the contribution of each feature from both global and local perspectives. First, the global importance of each feature is illustrated in Fig. 7. The observed month displays the highest ratio of contribution with approximately 40% of importance, which is expected according to the seasonality of $PM_{2.5}$ in Thailand (see Fig. S2). Among the geographical features, the land cover type, coordinates, and elevation sequentially demonstrate their respective importance. In terms of chemical components, NO_2 and SO_2, which are commonly known as precursors in the secondary formation of $PM_{2.5}$ (Baker and Scheff, 2007; Tucker, 2000), rank relatively low among all the features; SO_2 shows almost zero contribution to the estimation. Instead, CO accounts for about 20% of the contribution. We also compare R^2 scores between the chemical predictors, and CO shows the highest values for all regions (Table S3). Considering that CO is a by-product of carbon-containing fuel combustion, these results agree with the scenario that vehicular emissions and fires have a greater impact on the variation in air quality in Thailand than industrial emissions (ChooChuay et al., 2020).

Fig. 8 illustrates the top five important features on each decision step as the aspect of local feature importance. Consistent with the global perspective, the observed month, CO and land cover type are ranked as the most determining factors in all the steps, regardless of season. Interestingly, the second step displays different composition of importance, especially for meteorological features, by season. The importance of wind speed and relative humidity are relatively lower for dry season ranking fourth and fifth (Fig. 8f), while they are selected as the second and third most important features in wet season (Fig. 8j). Some other meteorological factors, such as pressure, evaporation and dew-point temperature, are also displayed in other steps, in spite of their low contribution (less than 5%). Considering that windy and humid weather can reduce pollution levels, the trained model locally employs weather information to identify the ideal conditions for lower $PM_{2.5}$ concentrations.

5.2. Impacts of fire on PM2.5 concentration in Thailand

To investigate the impact of fire on the air quality in Thailand, we analyze spatial distribution of fire radiative power (FRP) from the Global Fire Assimilation System (GFAS) in the Copernicus Atmosphere Monitoring Service (CAMS) and chemical components (Fig. 9) for a period when all the sub-regions show the rise of $PM_{2.5}$ concentration (from February 25th to March 2nd, 2021, red columns in Fig. 6). During the specific period, the central region of Thailand and the border areas in the northern and eastern parts, which are in close proximity to neighboring countries, such as Myanmar, Laos and Cambodia, exhibited high levels of FRP. Concurrently, there was an increase in the concentrations of $PM_{2.5}$ and major chemical components near fire hotspots, particularly during the detection of high FRP in the central-western region of Thailand on March 1st. However, in contrast to the satellite images, the in-situ measurements indicate an overall decreasing trend in $PM_{2.5}$ concentration within the central area, despite occasional sudden increases observed during the period (Fig. 6c). This inconsistency could be attributed to the spatially unbalanced distribution of in-situ measurements, which are primarily focus on the coastal region near Bangkok, the capital city of Thailand (Fig. 1). Unfortunately, no monitoring station exists in the central-western area where the high levels of FRP were observed, thereby imposing limitations on accurate $PM_{2.5}$ estimation.

The presence of statistically significant correlation between FRP and $PM_{2.5}$ suggests that fires play a critical role in explaining air quality variations throughout Thailand (Figs. S5–9). The increase in $PM_{2.5}$ concentration, accompanied by notably elevated levels of FRP, were particularly prominent in the northern and eastern regions, showing correlations exceeding 0.4. Nevertheless, it is important to acknowledge with care that the occurrence of fires alone cannot be regarded as the sole factor responsible for high $PM_{2.5}$ concentrations. For instance, there was a period of high $PM_{2.5}$ concentration observed from March 10th to April 2nd, despite low levels of FRP (Fig. S5).

The dispersion of fire emission across the neighbor countries can be another contributing factor leading to a significant increase in $PM_{2.5}$ concentration. For instance, when many hotspots were detected in the territory of Myanmar and Laos on February 26th and Marth 1st, the chemical and $PM_{2.5}$ concentrations distinctly increased in the north and northeast parts of Thailand the next day. Although U and V-wind components and wind speed show relatively smaller importance (Fig. 6), wind system across Southeast Asia also plays a crucial role on the pattern of emission propagation. The northeasterly wind is generally dominant in Thailand under the influence of the trade winds. However, two different monsoon systems in South Asia, such as the northeast and the southwest monsoon, cause seasonal changes in the wind direction and speed (Inthacha, 2011). Also, their variabilities are known to be modulated by the El Niño Southern Oscillation (Kirtphaiboon et al., 2014).

5.3. Potential further improvements

Traditionally, aerosol optical depth (AOD) has been played as an essential factor to estimate the surface level of $PM_{2.5}$ concentrations. However, the presence of cloud and snow along with the limited vertical resolution causes unfeasibility for archive reliable AOD limiting its spatial coverage (Hsu et al., 2013; Levy et al., 2007). Our approach based on atmospheric gas composition offers a viable alternative to address the spatial limits. To validate the $PM_{2.5}$ estimation skill without aerosol information, we conducted a comparison of skill scores by incorporating blue band (at 0.47 μm) and green band (at 0.55 μm) AOD from the Moderate Resolution Imaging Spectroradiometer (MODIS) Terra and Aqua combined level 2 product (MCD19A2 version 6.1). Fig. S10 depicts the comparable level of prediction performance, even slightly inferior to our model in some regions. However, it is important to note that due to higher frequency of missing values in the AOD dataset, the size of training dataset is significantly reduced, and the evaluation dataset cannot be identical. As an alternative, we also examined the inclusion of ultraviolet aerosol index, which the TROPOMI provides instead of AOD (Fig. S11). The results show almost equivalent performance supporting that surface $PM_{2.5}$ can be accurately estimated only with atmospheric trace gases. When comparing to recent studies using AOD as a predictor, our model shows better prediction skills, although evaluation period and regression models are different (Amnuaylojaroen, 2022; Peng-In et al., 2022). Nevertheless, the spatial constraints still remain in this study due to excluding data below the quality threshold. For instance, we only could collect less than 5 days of the predictors over the majority of the southern part of Thailand during January 2021 (Fig. S12a). As the rainy season approaches, the frequency of missing data tends to rise, resulting in nearly all regions experiencing a lack of data for more than 10 days in June 2021 (Fig. S12f). Thus, our future work will focus on the development of a model to handle the low-quality data aiming to achieve more reliable full coverage for $PM_{2.5}$ estimation.

While there have been prior attempts to apply DL-based modeling to estimate $PM_{2.5}$, its performance is lower than that of other algorithms (Chen et al., 2022; Pu and Yoo, 2021; Wong et al., 2021). Importantly, these results could be linked to their simple model structures, which mostly consist of a series of fully connected hidden layers with nonlinear activation functions. In the current study, we showcase how by fusing TROPOMI data with other geospatial sources and incorporating an advanced DL algorithm to provide an accurate representation of $PM_{2.5}$ concentration; consequently, an air pollution indicator can be developed. Previous studies have reported the potential of DL algorithms such as CNN and LSTM to improve estimation performance (Chen et al., 2021; Lu et al., 2021), and our results also support this by adopting a state-of-the-art DL algorithm. Numerous advanced DL methods have recently been developed and have achieved remarkable progress in diverse fields (Devlin et al., 2018; He et al., 2016); however, applying DL to estimate $PM_{2.5}$ concentration has not yet been widely explored. Thus, monitoring air quality by implementing DL approaches has considerable room for improvement.

6. Conclusion

Estimating ground-level $PM_{2.5}$ concentration is crucial for air quality monitoring and management. In this study, we introduced a data driven $PM_{2.5}$ estimation model based on a novel DL algorithm, called TabNet with atmospheric gases from the TROPOMI on the Sentinel-5P. Our model also considers meteorological variabilities and land properties as predictors, and targets 67 station records across Thailand.

The main findings of this study are as follows: 1) our model demonstrates statistically robust performance, achieving competitive skill scores (e.g. 0.873 of R^2) when compared to other state-of-the-art ML algorithms; 2) the spatially upscaled mappings exhibit good agreement with the observed data in terms of spatio-temporal variations; 3) the decision process in TabNet accounts for importance of each predictor, eliminating the need for additional post-processing steps typically required by other ML algorithms.

The evaluation of the predictor importance indicate that monthly variation is the leading predictor in $PM_{2.5}$ estimation. Geospatial characteristics, such as land cover type and latitude, also show a notable contribution from a global perspective. Among the chemical components from TROPOMI, we found that CO has the highest degree of importance, indicating that emissions from biomass burning and transport from neighboring countries exert a substantial influence on air quality in Thailand. The enhanced estimation capability and findings through its application are expected to not only boost other air quality studies, but also contribute to air quality management by providing advanced monitoring and evaluation techniques.

Author statement

Rackhun Son: Conceptualization, Methodology, Data, Visualization, Writing-Reviewing & Editing, Dimitris Stratoulias: Conceptualization, Data, Writing-Reviewing & Editing, Supervision, Hyun Cheol Kim: Writing-Reviewing & Editing, Jin-Ho Yoon: Writing-Reviewing & Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was conducted as a part of the Int'l Meteorological Expert Training Program supported by the Korea Meteorological Administration. JYoon was funded by Korea Environment Industry & Technology Institute (KEITI) through "Climate Change R&D Project for New Climate Regime.", funded by Korea Ministry of Environment (MOE) (RE202201655), and the GIST Research Institute (GRI) in 2023. H.C.K. was also supported by NOAA grant NA19NES4320002 (CISESS).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.apr.2023.101875.

References

Amante, C., Eakins, B.W., 2009. ETOPO1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis.

Amnuaylojaroen, T., 2022. Prediction of PM2. 5 in an urban area of northern Thailand using multivariate linear regression model. Adv. Meteorol. 2022, 3190484. https://doi.org/10.1155/2022/3190484.

Arino, O., 2010. GlobCover, 2009.

Arık, S.O., Pfister, T., 2021. Tabnet: Attentive Interpretable Tabular Learning. AAAI, pp. 6679–6687.

Baker, K., Scheff, P., 2007. Photochemical model performance for PM2. 5 sulfate, nitrate, ammonium, and precursor species SO2, HNO3, and NH3 at background monitor locations in the central and eastern United States. Atmos. Environ. 41, 6185–6195.

Borsdorff, T., Aan de Brugh, J., Hu, H., Aben, I., Hasekamp, O., Landgraf, J., 2018. Measuring carbon monoxide with TROPOMI: first results and a comparison with ECMWF-IFS analysis data. Geophys. Res. Lett. 45, 2826–2832.

Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. J. Roy. Stat. Soc.: Series D (The Statistician) 47, 431–443.

Chen, B., Song, Z., Pan, F., Huang, Y., 2022. Obtaining vertical distribution of PM2. 5 from CALIOP data and machine learning algorithms. Sci. Total Environ. 805, 150338.

Chen, B., You, S., Ye, Y., Fu, Y., Ye, Z., Deng, J., Wang, K., Hong, Y., 2021. An interpretable self-adaptive deep neural network for estimating daily spatially-continuous PM2. 5 concentrations across China. Sci. Total Environ. 768, 144724.

Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L.D., Abramson, M.J., Guo, Y., 2018. Spatiotemporal patterns of PM10 concentrations over China during 2005–2016: a satellite-based estimation using the random forests approach. Environ. Pollut. 242, 605–613.

Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, pp. 785–794.

Chen, Z.-Y., Zhang, T.-H., Zhang, R., Zhu, Z.-M., Yang, J., Chen, P.-Y., Ou, C.-Q., Guo, Y., 2019. Extreme gradient boosting model to estimate PM2. 5 concentrations with missing-filled satellite data in China. Atmos. Environ. 202, 180–189.

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. Atmos. Pollut. Res. 11, 1657–1667.

Chu, Y., Liu, Y., Li, X., Liu, Z., Lu, H., Lu, Y., Mao, Z., Chen, X., Li, N., Ren, M., 2016. A review on predicting ground PM2. 5 concentration using satellite aerosol optical depth. Atmosphere 7, 129.

de Hoogh, K., Héritier, H., Stafoggia, M., Künzli, N., Kloog, I., 2018. Modelling daily PM2. 5 concentrations at high spatio-temporal resolution across Switzerland. Environ. Pollut. 233, 1147–1154.

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., Van Roozendael, M., Richter, A., Hilboll, A., Peters, E., 2018. Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. Atmos. Meas. Tech. 11, 2395–2426.

Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding arXiv preprint arXiv: 1810.04805.

Garane, K., Koukouli, M.-E., Verhoelst, T., Lerot, C., Heue, K.-P., Fioletov, V., Balis, D., Bais, A., Bazureau, A., Dehn, A., 2019. TROPOMI/S5P total ozone column data: global ground-based validation and consistency with other satellite missions. Atmos. Meas. Tech. 12, 5263–5287.

Ghahremanloo, M., Lops, Y., Choi, Y., Yeganeh, B., 2021. Deep learning estimation of daily ground-level NO2 concentrations from remote sensing data. J. Geophys. Res. Atmos. 126, e2021JD034925.

Gurjar, B.R., Molina, L.T., Ojha, C.S.P., 2010. Air Pollution: Health and Environmental Impacts. CRC press.

Han, S., Kundhikanjana, W., Towashiraporn, P., Stratoulias, D., 2022. Interpolation-based fusion of sentinel-5P, SRTM, and regulatory-grade ground stations data for producing spatially continuous maps of PM2. 5 concentrations nationwide over Thailand. Atmosphere 13, 161.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Hsu, N.C., Jeong, M., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., Tsay, S., 2013. Enhanced Deep Blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118, 9296–9315.

Hu, X., Belle, J.H., Meng, X., Wildani, A., Waller, L.A., Strickland, M.J., Liu, Y., 2017. Estimating PM2. 5 concentrations in the conterminous United States using the random forest approach. Environ. Sci. Technol. 51, 6936–6944.

Inthacha, S., 2011. The Climatology of Thailand and Future Climate Change Projections Using the Regional Climate Model PRECIS. PhD Thesis. University of East Angila.

Jiang, M., Sun, W., Yang, G., Zhang, D., 2017. Modelling seasonal GWR of daily PM2. 5 with proper auxiliary variables for the Yangtze River Delta. Rem. Sens. 9, 346.

Jiang, Q., Christakos, G., 2018. Space-time mapping of ground-level PM2. 5 and NO2 concentrations in heavily polluted northern China during winter using the Bayesian maximum entropy technique with satellite data. Air Qual Atmos Health 11, 23–33.

Junpen, A., Garivait, S., Bonnet, S., 2013. Estimating emissions from forest fires in Thailand using MODIS active fire product and country specific data. Asia Pac J Atmos Sci 49, 389–400.

Kanabkaew, T., Kim Oanh, N.T., 2011. Development of spatial and temporal emission inventory for crop residue field burning. Environ. Model. Assess. 16, 453–464.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y., 2017. Lightgbm: a highly efficient gradient boosting decision tree. Adv. Neural Inf. Process. Syst. 30.

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.

Kloog, I., Nordio, F., Coull, B.A., Schwartz, J., 2012. Incorporating local land use regression and satellite aerosol optical depth in a hybrid model of spatiotemporal PM2. 5 exposures in the Mid-Atlantic states. Environ. Sci. Technol. 46, 11913–11921.

Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J. Geophys. Res. Atmos. 112.

Liu, Y., Park, R.J., Jacob, D.J., Li, Q., Kilaru, V., Sarnat, J.A., 2004. Mapping annual mean ground-level PM2. 5 concentrations using Multiangle Imaging Spectroradiometer aerosol optical thickness over the contiguous United States. J. Geophys. Res. Atmos. 109.

Long, A., Ascent, D., 2020. World Economic Outlook. International Monetary Fund.

Lu, X., Wang, J., Yan, Y., Zhou, L., Ma, W., 2021. Estimating hourly PM2. 5 concentrations using Himawari-8 AOD and a DBSCAN-modified deep learning model over the YRDUA, China. Atmos. Pollut. Res. 12, 183–192.

Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM2. 5 concentrations: China, 2004–2013. Environ. Health Perspect. 124, 184–192.

Mahasakpan, N., Chaisongkaew, P., Inerb, M., Nim, N., Phairuang, W., Tekasakul, S., Furuuchi, M., Hata, M., Kaosol, T., Tekasakul, P., 2023. Fine and ultrafine particle- and gas-polycyclic aromatic hydrocarbons affecting southern Thailand air quality during transboundary haze and potential health effects. J. Environ. Sci. 124, 253–267.

Manisan, V., 1995. Geography and Climatology in Every Season of Various Parts in Thailand. Technical document of Meteorological Department, Bangkok.

Martins, A., Astudillo, R., 2016. From softmax to sparsemax: a sparse model of attention and multi-label classification. In: International Conference on Machine Learning. PMLR, pp. 1614–1623.

Muñoz Sabater, J., 2019. ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) 10.

Organization, W.H., 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease.

Peng-In, B., Sanitluea, P., Monjatturat, P., Boonkerd, P., Phosri, A., 2022. Estimating ground-level PM2. 5 over Bangkok Metropolitan Region in Thailand using aerosol optical depth retrieved by MODIS. Air Qual Atmos Health 1–12.

Peters, B., Niculae, V., Martins, A.F.T., 2019. Sparse Sequence-To-Sequence Models arXiv preprint arXiv:1905.05702.

Phairuang, W., Inerb, M., Furuuchi, M., Hata, M., Tekasakul, S., Tekasakul, P., 2020. Size-fractionated carbonaceous aerosols down to PM0. 1 in southern Thailand: local and long-range transport effects. Environ. Pollut. 260, 114031.

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, e0189909.

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A.V., Gulin, A., 2018. CatBoost: unbiased boosting with categorical features. Adv. Neural Inf. Process. Syst. 31.

Pu, Q., Yoo, E.-H., 2021. Ground PM2. 5 prediction using imputed MAIAC AOD with uncertainty quantification. Environ. Pollut. 274, 116574.

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. Atmos. Environ. 247, 118176.

Rodríguez-Urrego, D., Rodríguez-Urrego, L., 2020. Air quality during the COVID-19: PM2. 5 analysis in the 50 most polluted capital cities in the world. Environ. Pollut. 266, 115042.

Shin, M., Kang, Y., Park, S., Im, J., Yoo, C., Quackenbush, L.J., 2020. Estimating ground-level particulate matter concentrations using satellite-based data: a review. GIsci Remote Sens 57, 174–189.

Sorek-Hamer, M., Strawa, A.W., Chatfield, R.B., Esswein, R., Cohen, A., Broday, D.M., 2013. Improved retrieval of PM2. 5 from satellite data products using non-linear methods. Environ. Pollut. 182, 417–423.

Stratoulias, D., Nuthammachot, N., 2020. Air quality development during the COVID-19 pandemic over a medium-sized urban area in Thailand. Sci. Total Environ. 746, 141320.

Theys, N., De Smedt, I., Yu, H., Danckaert, T., van Gent, J., Hörmann, C., Wagner, T., Hedelt, P., Bauer, H., Romahn, F., 2017. Sulfur dioxide retrievals from TROPOMI onboard Sentinel-5 Precursor: algorithm theoretical basis. Atmos. Meas. Tech. 10, 119–153.

Torsri, K., Octaviani, M., Manomaiphiboon, K., Towprayoon, S., 2013. Regional mean and variability characteristics of temperature and precipitation over Thailand in 1961–2000 by a regional climate model and their evaluation. Theor. Appl. Climatol. 113, 289–304.

Tucker, W.G., 2000. An overview of PM2. 5 sources and control strategies. Fuel Process. Technol. 65, 379–392.
Van Donkelaar, A., Martin, R.V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., Villeneuve, P.J., 2010. Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. Environ. Health Perspect. 118, 847–855.
Van Geffen, J., Eskes, H.J., Boersma, K.F., Maasakkers, J.D., Veefkind, J.P., 2019. TROPOMI ATBD of the Total and Tropospheric NO2 Data Products. DLR document.
Vassanadumrongdee, S., Matsuoka, S., 2005. Risk perceptions and value of a statistical life for air pollution and traffic accidents: evidence from Bangkok, Thailand. J. Risk Uncertain. 30, 261–287.
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. Sci. Total Environ. 793, 148535.
Wei, J., Huang, W., Li, Z., Xue, W., Peng, Y., Sun, L., Cribb, M., 2019. Estimating 1-km-resolution PM2. 5 concentrations across China using the space-time random forest approach. Remote Sens. Environ. 231, 111221.
Weichenthal, S.A., Godri Pollitt, K., Villeneuve, P.J., 2013. PM2. 5, oxidant defence and cardiorespiratory health: a review. Environ. Health 12, 1–8.
Wong, P.-Y., Lee, H.-Y., Chen, Y.-C., Zeng, Y.-T., Chern, Y.-R., Chen, N.-T., Lung, S.-C.C., Su, H.-J., Wu, C.-D., 2021. Using a land use regression model with machine learning to estimate ground level PM2. 5. Environ. Pollut. 277, 116846.
World Air Quality Report, 2020. https://www.greenpeace.org/static/planet4-romania-stateless/2021/03/d8050eab-2020-world_air_quality_report.pdf.
Xie, Y., Wang, Y., Zhang, K., Dong, W., Lv, B., Bai, Y., 2015. Daily estimation of ground-level PM2. 5 concentrations over Beijing using 3 km resolution MODIS AOD. Environ. Sci. Technol. 49, 12280–12288.
Xue, T., Zheng, Y., Tong, D., Zheng, B., Li, X., Zhu, T., Zhang, Q., 2019. Spatiotemporal continuous estimates of PM2. 5 concentrations in China, 2000–2016: a machine learning method with inputs from satellites, chemical transport model, and ground observations. Environ. Int. 123, 345–357.
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. Rem. Sens. 8, 184.
Zhang, X., Han, Liangxiu, Han, Lianghao, Zhu, L., 2020. How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery? Rem. Sens. 12, 417.
Zhu, L., Huang, L., Fan, L., Huang, J., Huang, F., Chen, J., Zhang, Z., Wang, Y., 2020. Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network. Sensors 20, 1576.
Zou, B., Chen, J., Zhai, L., Fang, X., Zheng, Z., 2017. Satellite based mapping of ground PM2. 5 concentration using generalized additive modeling. Rem. Sens. 9, 1.