LASER PULSE

Long-term Assistance and SErvices for Research (LASER)

Partners for University-Led Solutions Engine (PULSE)

TECHNICAL REPORT

SUPPLEMENT TO AGREEMENT NO. AID-7200AA18CA00009

AOR Name: Kevin Roberts

June, 2023

LASER PULSE Delivering Practical, Research-Driven Solutions to Global Development Challenges

VIETNAM NATIONAL UNIVERSITY

UNIVERSITY OF ENGINEERING AND TECHNOLOGY

TECHNICAL REPORT

Development of $PM_{2.5}$ concentration maps using multi-source datasets

PROJECT

IMPROVING AIR POLLUTION MONITORING AND MANAGEMENT OF VIETNAM WITH SATELLITE $PM_{2.5}$ OBSERVATION

No: F9002550402127

Host organization: University of Engineering and Technology, VNU

Project manager: Assoc. Prof. Thanh Thi Nhat Nguyen

Authors*:*

Assoc. Prof. Thanh Thi Nhat Nguyen, Falcuty of Information Technology, University of Engineering and Technology, VNU.

MSc. Vinh Tuan Tran, Institute of Information Technology, Hanoi Pedagogical University 2

MSc. Truong Xuan Ngo, Falcuty of Information Technology, University of Engineering and Technology, VNU.

Consultants:

Assoc. Prof. Thuy Bich Ly, School of Environmental Science and Technology, HaNoi University of science and technology.

Dr. Son Si Tong, University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology.

Dr. Trong Đuc Le, University of Engineering and Technology, VNU.

Ha Noi – 2022

LASER PULSE Delivering Practical, Research-Driven Solutions to Global Development Challenges

ABOUT LASER PULSE

LASER (Long-term Assistance and SErvices for Research) PULSE (Partners for University-Led Solutions Engine) is a $70M program funded through USAID's Innovation, Technology, and Research Hub, that delivers research-driven solutions to field-sourced development challenges in USAID partner countries.

A consortium led by Purdue University, with core partners Catholic Relief Services, Indiana University, Makerere University, and the University of Notre Dame, implements the LASER PULSE program through a growing network of 3,000+ researchers and development practitioners in 82 countries.

LASER PULSE collaborates with USAID missions, bureaus, and independent offices, and other local stakeholders to identify research needs for critical development challenges, and funds and strengthens the capacity of researcher-practitioner teams to co-design solutions that translate into policy and practice.

This report is made possible by the support of the United States Agency for International Development (USAID) via the Long-Term Assistance and SErvices for Research Partners for University-Led Solutions Engine (LASER PULSE) program. The contents herein are those of the authors and do not necessarily reflect the views of USAID, LASER PULSE or the United States Government.

TABLE OF CONTENTS

ABSTRACT ....

LIST OF ACRONYMS ....

LIST OF FIGURES ....

LIST OF TABLE ....

1. INTRODUCTION 1

2. DATA 3

2.1. Study area 3

2.2. $PM_{2.5}$ monitoring data at stations 4

2.3. Aerosol Optical Depth data 5

2.4. Meteorological data 6

2.5. Land surface data 8

3. METHODOLOGY 10

3.1.1. Data preprocessing 11

3.1.2. Enhance AOD data quality 13

3.1.3. Data integration 14

3.1.4. Model development 15

3.1.5. AOD data missing issue 17

3.1.6. Model and map validation 17

4. EXPERIMENT AND RESULT 20

4.1. MEM evaluations 20

4.2. $PM_{2.5}$ map evaluations 25

4.2.1. Daily map 25

4.2.2. Monthly map 29

4.2.3. Annual average map 36

CONCLUSION 41

REFERENCES 43

EXECUTIVE SUMMARY

General context

Air pollution poses a significant threat to the environment and human health in many countries. According to the World Health Organization (WHO), there are 7 million premature deaths attributed to both outdoor and indoor air pollution globally each year [1]. Recent environmental reports have shown that air pollution levels in Vietnam are among the highest in the world [2][3], with many provinces and cities exceeding national standards for annual average $PM_{2.5}$ concentrations [4].

To control the air pollution, it is important to monitor air pollution in order to assess the current situation, provide appropriate solutions, and evaluate the effectiveness of management policies on emissions and the environment. In Vietnam, $PM_{2.5}$ dust monitoring is currently carried out by various methods and technologies by different parties, including environmental protection agencies, embassies, research units, and the private sector. Although automatic and continuous monitoring by "standard" stations provides accurate results, the high installation and operating costs make it difficult to cover large areas with high density. In recent years, remote sensing (satellite) technology has been applied to research and provide information on $PM_{2.5}$ dust concentration, providing additional monitoring methods that show the spatial distribution of $PM_{2.5}$ dust over a large scale, especially in places where ground monitoring stations cannot be installed. $PM_{2.5}$ concentration data calculated from satellite images have a complete distribution in space and over a long period of time, but the frequency of data monitoring is low (mainly providing data on a daily basis) and the quality of data is affected by indirect and remote observations (atmospheric column observations), and cloudy weather conditions.

Objectives

This report introduces a method for creating and assessing a national-scale dataset of daily average $PM_{2.5}$ concentration maps from 2019 to 2021. This method has been inherited from the previous work of the research group [5] and updated according to the recently published study [6]. The $PM_{2.5}$ maps is utilized to generate deliverables for users under the project "Improving air pollution monitoring and management of Vietnam with satellite $PM_{2.5}$ observation" in the LASER PULSE program. The deliverables of the project include: (i) a $PM_{2.5}$ dataset from 2019-2021, (ii) a report on the status of $PM_{2.5}$ and health impact assessment in Vietnam in 2021, (iii) a video for educational and promotional purposes, and (iv) a WebGIS system that displays near-real-time air quality information.

Methodology

The daily map of $PM_{2.5}$ dust concentration in Vietnam was constructed using a Mixed Effect Model (MEM) that incorporates various data sources from the period of 2012 to 2021. The input data comprised of ground station monitoring data of $PM_{2.5}$ dust concentration, satellite image products (AOD), maps of meteorological parameters such as humidity and Planetary boundary layer height (PBLH), as well as land use maps indicating traffic density, vegetation index, and topography. The steps involved in constructing the $PM_{2.5}$ dust concentration map include: preprocessing the station and map data, improving the quality of the map data, integrating the map data with the station data, building and validating the Mixed Effect Model (MEM) to estimate daily $PM_{2.5}$ maps, synthesizing monthly/yearly maps of $PM_{2.5}$, evaluating the quality of the map data against station data, and comparing the results with global $PM_{2.5}$ mapping products.

Results

The daily $PM_{2.5}$ maps were estimated using two MEM models (one main and one auxiliary model) developed based on 10-year datasets (2012-2021). Preprocessing and quality enhancement techniques were applied to each data type to ensure data quality before modeling. The main model, which correlates with data from ground stations, had a Pearson correlation coefficient (r) of 0.83, R-squared (R^2) of 0.68, Root Mean Square Error (RMSE) of 14.14 µg/m³, and Mean Relative Error (MRE) of 40.99% out of a sample size of 10,614. The auxiliary model had a Pearson correlation coefficient (r) of 0.81, R-squared (R^2) of 0.65, RMSE of 14.8 µg/m³, and MRE of 48.58% on a larger sample size of 34,208. Although the evaluation parameters for the model obtained with ground station data are relatively good, the quality of the model cannot be fully and accurately reflected in terms of spatial limitations, completeness, and accuracy of datasets at ground stations.

The model estimated the daily $PM_{2.5}$ map (3x3 km) for Vietnam for the period from 2019 to 2021. The assessment of the daily $PM_{2.5}$ map with ground measurements during this period showed a Pearson r of 0.8, R^2 of 0.64, RMSE of 15.44 µg/m³, and MRE of 51.75%. The monthly and annual average maps for the period 2019-2021 were of better quality than the global $PM_{2.5}$ product when compared to monitoring stations in each year.

Recommendations

The $PM_{2.5}$ concentration maps generated from the model has provided valuable insights into the spatial distribution and seasonal variation of $PM_{2.5}$ concentrations in Vietnam for the period of 2019-2021. The data has a spatial resolution of 3x3 km, covering the entire area of Vietnam. This means that every 3x3 km region in Vietnam has an estimated $PM_{2.5}$ concentration value, which is convenient for analyzing $PM_{2.5}$ variations spatially (aggregating and comparing by commune, district, province, economic region, and country). The data is detailed on a daily basis and continuous over the period from 2019 to 2021, allowing for convenient analysis of $PM_{2.5}$ variations/trends over time (daily, weekly, monthly, quarterly, seasonal, yearly).

The spatial and temporal analysis of $PM_{2.5}$ concentration information can be applied to support decision-making and used in other related studies. The $PM_{2.5}$ datasets were used to assess the impact of air pollution on health [7][8] and environment [9] in Vietnam. As for policy development purposes, the $PM_{2.5}$ datasets of Hanoi and Ho Chi Minh City, Vinh Yen, Da Nang, Soc Trang, Ca Mau, Quang Tri, and Hai Phong City were used to assist with the policy briefs or recommendations for local government agencies. For instances, clean air action plans for Ho Chi Minh city and Vinh Yen city; policy brief and report on child-centered risks and access to community services for Soc Trang, Ca Mau, Quang Tri; vehicle emission inventory development for Da Nang; air quality management plan for 2022-2025 for Hai Phong. In this project, the map was used to develop and publish the report on the "Status of $PM_{2.5}$ and Health Impacts in Vietnam in 2021" [10], providing air pollution information to the aforementioned users. The map was also used to create educational and communication videos about air pollution. Additionally, the mapping method was applied in the WebGIS system to display near-real-time air quality and provide information to the community [11]. Prior to this study, there was a lack of results on automatic continuous monitoring and modeling of $PM_{2.5}$ concentrations in Vietnam, making this study a significant contribution to the field.

LIST OF ACRONYMS

Abbreviation	Definition (English)	Definition (Vietnamese)
AOD (AOT)	Aerosol Optical Depth/Thickness	Độ sâu quang học của sol khí
MEM	Mixed Effect Model	Mô hình ảnh hưởng hỗn hợp
NCEM	Northern Center for Environmental Monitoring	Trung tâm quan trắc môi trường miền Bắc
USE	US Embassy	Đại sứ quán Mỹ
WHO	World Health Organization	Tổ chức Y tế Thế giới
GDAL	Geospatial Data Abstraction Library	Thư viện Trích xuất Dữ liệu Không gian Địa lý
QCVN	Vietnam standard	Quy chuẩn Việt Nam
RMSE	Root Mean Square Error	Sai số toàn phương trung bình
MRE	Mean Relative Error	Sai số tương đối trung bình
EPI	Environmental Performance Index	Chỉ số hiệu quả môi trường
VNU-UET	University of Engineering and Technology – Vietnam National University	Trường Đại học Công Nghệ - Đại học Quốc gia Hà Nội
BTNMT	Ministry of Natural Resources and Environment	Bộ Tài nguyên và Môi Trường
CV	Cross Validation	Kiểm chứng chéo
PBLH	Planetary Boundary Layer Height	Độ cao lớp biên hành tinh
GWR	Geographical Weighted Regression	Hồi quy trọng số địa lý
GSO	General Statistics Office	Tổng cục Thống Kê
GBOD	Global Burden of Disease	Gánh nặng bệnh tật toàn cầu
DEM	Digital Elevation Model	Bản đồ mô hình độ cao số
WRF	Weather Research and Forecast model	Mô hình nghiên cứu và dự báo thời tiết
MODIS	Moderate Resolution Imaging Spectroradiometer	Máy quang phổ hình ảnh có độ phân giải vừa phải
VIIRS	Visible Infrared Imaging Radiometer Suite	Bộ đo phóng xạ hình ảnh hồng ngoại khả kiến
NDVI	Normalized difference vegetation index	Chỉ số thực vật

LIST OF FIGURES

Figure 2.1. Vietnam study area and air quality monitoring stations 4
Figure 2.2. Data at a monitoring station 5
Figure 2.3. Daily AOD Map over Vietnam 6
Figure 2.4. Meteorological maps over Vietnam 8
Figure 2.5. Land surface maps over Vietnam 9
Figure 3.1. The process of building $PM_{2.5}$ concentration maps [6] 10
Figure 3.2. Average resampling example 12
Figure 4.1. Example map generated from the primary model (with AOD) and auxiliary model (no AOD) 21
Figure 4.2. Scatter plot between map and station values 26
Figure 4.3. Variation of average $PM_{2.5}$ concentration in Vietnam by month during 2019, 2020, 2021 34
Figure 4.4. Distribution of $PM_{2.5}$ concentration on average monthly in 3 years 2019-2021 36
Figure 4.5. Evaluation of monthly map estimates at several stations for the period 2019-2021. 36
Figure 4.6. Comparision between the annual average $PM_{2.5}$ map and measurements at ground monitoring stations in the period of 2019-2021 38
Figure 4.7. Statistics of annual average $PM_{2.5}$ in Vietnam during the period of 2019-2021 39
Figure 4.8. Annual average maps of $PM_{2.5}$ dust concentration in 2019 - 2021 in Vietnam 40

LIST OF TABLES

Table 3.1. List of parameters and trend with $PM_{2.5}$ 15
Table 4.1. MEM model coefficients and statistical tests 20
Table 4.2. Model evaluations and cross-validation 22
Table 4.3. Related studies 23
Table 4.4. Comparison results between daily average $PM_{2.5}$ map and each monitoring station during 2019-2021 27
Table 4.5. Overall assessment of the monthly average map with observations at monitoring stations in the period of 2019 - 2021 30
Table 4.6. Comparision results of the monthly average $PM_{2.5}$ map with monitoring at each station during the period of 2019-2021 30
Table 4.7. Compare $PM_{2.5}$ maps with monitoring stations and compare with global maps 37

1. INTRODUCTION

Air pollution is a major global issue that poses a significant threat to public health and the environment. The World Health Organization's estimates of 7 million premature deaths globally each year due to air pollution [1]. According to the EPI 2022, Vietnam is ranked 130th out of 180 countries in terms of exposure to air pollution [2]. Similarly, the IQAir/Air Visual report indicates that the country ranks 36th in the world in terms of the population weighted average concentration of $PM_{2.5}$ dust [3]. The report on the state of the national environment for the period 2016-2020 shows an increasing trend in $PM_{2.5}$ concentrations from 2017 to 2019, with a slight decrease observed in 2020 [12]. The report on $PM_{2.5}$ dust status in Vietnam for the period 2019-2020 using multi-source data reveals that the Northern provinces and cities of Vietnam are particularly affected, with a significant number of them exceeding the national standards for $PM_{2.5}$ concentrations [4].

The Ministry of Natural Resources and Environment (MONRE) and provincial Departments of Natural Resources and Environment carry out automatic and continuous monitoring of $PM_{2.5}$ dust and other pollutants for state management purposes. $PM_{2.5}$ dust data is also provided by standard monitoring stations, sensor networks of embassies, scientific research institutions, and private entities. In recent years, modeling techniques have been widely used to supplement data from monitoring stations using satellite images (remote sensing) and auxiliary data to provide the spatial distribution of $PM_{2.5}$ dust concentrations, particularly in areas where stations are not installed [13]. $PM_{2.5}$ concentration data derived from satellite images provides complete spatial distribution over a long period of time, but the frequency of data monitoring is low, mainly providing data on a daily basis. The quality of the data is average because of indirect and remote observation.

To estimate $PM_{2.5}$ concentrations using satellite data, a statistical model is built to represent the relationship between satellite Aerosol Optical Depth (AOD) and ground $PM_{2.5}$. AOD is a measure of the attenuation of solar radiation caused by the scattering and absorption of aerosols in the atmospheric column, and it has a close relationship with ground $PM_{2.5}$ dust concentration [14]. AOD measures the atmospheric column, while $PM_{2.5}$ measures at ground level. To increase the interpretability of the model, factors affecting $PM_{2.5}$, such as meteorology (temperature, humidity, pressure, precipitation, wind), terrain, and land use (topography, population density, traffic density, tree cover, water cover, urban cover) are also used. Meteorological parameters affect the accumulation, dispersion, or spread of air pollution, while the parameters of terrain and land use characterize the sources that emit pollutants.

In this report, we present the method of creating a map of daily average $PM_{2.5}$ concentrations in Vietnam between 2019 and 2021 using multi-source data. This method has been inherited from the previous work of the research group [5] and updated according to the recently published study [6]. The resulting $PM_{2.5}$ map is used to generate deliverables for the "Improving air pollution monitoring and management of Vietnam with satellite $PM_{2.5}$ observation" project in the LASER PULSE program. These deliverables include: (i) a $PM_{2.5}$ dataset spanning the period of 2019-2021, (ii) a report on the status of $PM_{2.5}$ and its health impacts in Vietnam for 2021, (iii) an educational and promotional video, and (iv) a WebGIS system that provides near-real-time air quality information.

2. DATA

2.1. Study area

The geography of Vietnam places it in the East of the Indochina Peninsula, located near the center of Southeast Asia, and stretches from (8°27'N, 102°8'E) to (23°23'N, 109°27'E), with a land area of 331,236 km^2. Vietnam shares a 1,400 km border with China to the north, a 2,100 km border with Laos to the northwest, and a 1,100 km border with Cambodia to the southwest. To the east and south, Vietnam has a coastline that spans 3,260 km in total (as shown in Figure 2.1).

Depending on the topography that changes from North to South, each region has different climate characteristics. The North experiences a climate consisting of four seasons: spring, summer, autumn, and winter, with winter and summer being the two main seasons, while spring and autumn are considered transitional periods. Winter lasts from October to March and is characterized by cold and dry weather due to the influence of winter winds blowing from the northeast along the coast of China through the Gulf of Tonkin. Summer, from May to August, is dominated by hot and dry southwest winds as well as humid southeast winds [15]. In the Southern region, the weather is divided into a dry season with the primary northeast wind direction and a rainy season with the southwest wind direction.

According to the General Statistics Office, Vietnam's population is projected to reach 99.2 million people in 2022. The population is unevenly distributed across the territory, with concentrations in the plains, coastal areas, and sparsely populated urban, mountainous, and midland areas. Vietnam's GDP per capita is estimated to reach approximately 4,162.94 USD in 2022, ranking 117th in the world. In terms of economic structure, the agriculture, forestry, and fishery sector accounts for 11.88%; industry and construction account for 38.26%; the service sector accounts for 41.33%; and product tax account for 8.53% [16]. According to the draft National Land Use Planning Report for the period 2021-2030, with a vision to 2050, industrial parks (IPs) nationwide are divided into six regions: Northern Midlands and Mountains (30 IPs covering an area of 7,000 ha), Red River Delta (94 IPs covering an area of 26,000 ha), North Central and Central Coast (68 IPs covering an area of 22,000 ha), Central Highlands (10 IPs covering an area of 2,000 ha), South East (119 IPs covering an area of 44,000 ha), and the Mekong Delta (60 IPs covering an area of 13,000 ha).

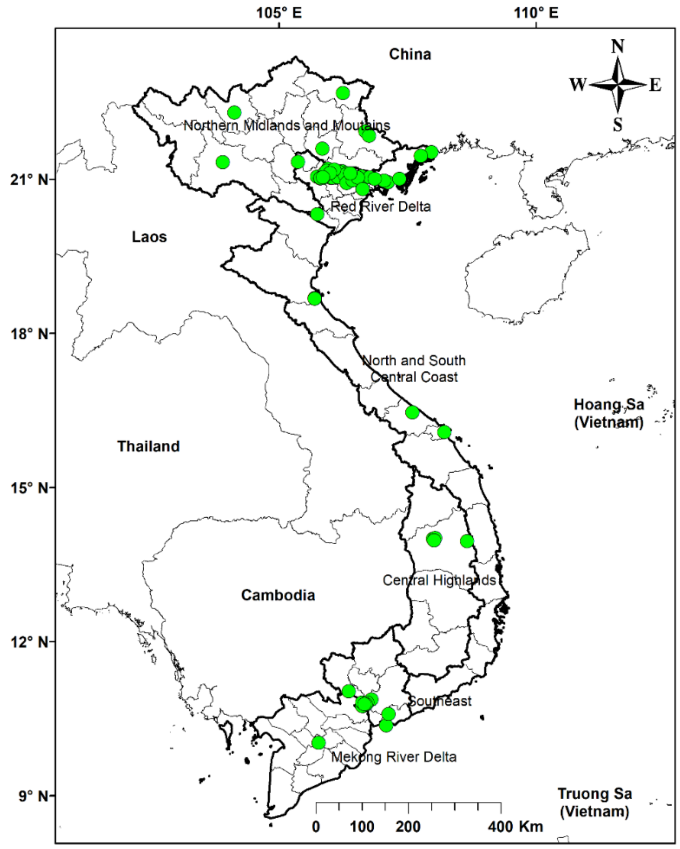


Figure 2.1. Vietnam study area and air quality monitoring stations

2.2. $PM_{2.5}$ monitoring data at stations

Hourly average measured $PM_{2.5}$ concentration data was collected from fixed and continuous monitoring stations nationwide between 2012 and 2021. Of the 85 stations, the Northern Center for Environmental Monitoring (NCEM) operated the majority, which are located in various provinces and cities across the country, including Hanoi, Bac Ninh, Phu Tho, Quang Ninh, and Ho Chi Minh City. In addition, two stations are operated by the US Embassy (US Embassy – USE) in Hanoi and Ho Chi Minh City, while six stations in Ho Chi Minh City are managed by the research group of Vietnam National University Ho Chi Minh, and one station is managed by the research group of University of Natural Sciences - Vietnam National University, Ho Chi Minh City. Station data was collected hourly in the form of tables (excel files), including fields (columns) such as time, air pollution parameters ($PM_{2.5}$, PM10, NO, NO2, CO, etc.), and meteorological parameters (temperature, pressure, wind, rain). Information about station name and location (longitude, latitude) was also collected (see Figure 2.2). $PM_{2.5}$ concentration data from these stations were used to build a model for estimating $PM_{2.5}$ dust concentration in Vietnam and verifying the resulting maps.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Datetime	NHIệT độ	NOX	SO2	O3	PM-10	PM-2-5	NO	NO2	CO	ÁP SUấT K	PM-1	RADIATIO	HướNG GI	TốC độ Gl	LượNG M	Độ ẩm	PM4	PM-TOTAL
00:00 01/0	26.761	5.9811	56.3913	49.2899	0.1887	0.1732	2.0683	3.9129	408.0035	1006.68	0.162	0	195.8511	2.6803	0	84.9608	0.188	0.1913
01:00 01/0	26.5952	6.1797	45.4446	46.6689	0.1523	0.1362	2.1849	3.9947	252.5	1005.802	0.1258	0	197.4704	2.3358	0	85.8094	0.1507	0.1551
02:00 01/0	26.3473	6.0624	34.6588	43.4097	0.2134	0.2058	2.0882	3.9742	433.4375	1006.747	0.1961	0	195.9981	3.1438	0	87.1536	0.2138	0.2143
03:00 01/0	26.2665	6.1526	34.7569	39.0899	0.2284	0.1966	2.1967	3.9559	328.2986	1005.393	0.1825	0	174.8088	1.9564	0	88.0308	0.2093	0.24
04:00 01/0	26.1344	6.3139	35.6098	34.3266	0.2011	0.1864	2.1537	4.1602	318.8715	1003.249	0.1761	0	197.2468	2.21	0	89.1409	0.1938	0.2057
05:00 01/0	26.117	6.3945	38.9008	33.1019	0.2271	0.2006	2.1293	4.2652	352.4132	1001.953	0.1876	0	183.046	1.2313	0	89.7554	0.2141	0.2349
06:00 01/0	26.101	5.9714	42.4452	28.6647	0.2498	0.2348	2.101	3.8704	849.3923	1005.065	0.217	14.6694	137.3312	0.6228	0	90.1429	0.2494	0.2518
07:00 01/0	26.8107	6.0383	43.0283	41.6766	0.2514	0.2257	1.9905	4.0478	783.8021	1006.044	0.2098	104.5354	197.3468	2.1419	0	88.9934	0.2444	0.2574
08:00 01/0	27.9565	6.318	36.8479	54.3286	0.2378	0.1758	2.1427	4.1753	822.5694	1005.676	0.1646	161.7259	197.3538	3.5608	0	84.4058	0.1764	0.3371
09:00 01/0	28.9875	6.2216	32.5508	58.6524	0.198	0.1509	2.064	4.1576	351.4236	1005.895	0.1321	273.9372	194.0961	4.2431	0	79.8766	0.1738	0.2361
10:00 01/0	30.169	6.1693	30.4395	58.5947	0.2377	0.187	2.0331	4.1362	312.3264	1005.785	0.1678	418.625	193.9129	4.2563	0	74.4608	0.2157	0.2517
11:00 01/0	31.5004	6.1747	32.3816	62.92	0.2501	0.2105	2.3774	3.7973	240.0347	1006.225	0.1901	765.5866	190.8387	5.8991	0	67.71	0.2314	0.2622
12:00 01/0	32.7115	6.2196	33.7869	73.4787	0.191	0.1814	2.0195	4.2001	213.5938	1006.152	0.1705	726.3216	192.7478	4.5471	0	62.1396	0.1914	0.1921
13:00 01/0	32.8969	6.2885	34.6966	114.1865	0.2229	0.192	2.0601	4.2284	197.1354	1003.109	0.171	534.5276	190.9288	5.275	0.0067	62.1783	0.2199	0.2283
14:00 01/0	33.7904	6.114	41.3484	133.2973	0.169	0.1609	2.002	4.112	193.7847	1001.719	0.1491	493.1329	197.6208	4.1818	0.0034	60.0158	0.1692	0.17
15:00 01/0	34.4941	5.8864	47.8204	125.5217	0.1491	0.1265	1.9013	3.985	174.0278	1001.854	0.115	642.9401	200.658	3.8295	0	57.0291	0.1422	0.1545
16:00 01/0	34.3529	5.9741	40.1397	111.0526	0.1885	0.1428	1.9905	3.9836	227.7951	1002.018	0.126	393.7217	196.8611	4.3686	0	56.3808	0.1627	0.2042
17:00 01/0	33.4337	6.3483	28.233	81.7212	0.2572	0.2059	2.096	4.2524	439.6007	998.6052	0.19	294.1842	192.1547	5.857	0	64.6549	0.2231	0.2761
18:00 01/0	31.7795	6.1161	26.9515	69.1734	0.3293	0.2119	2.2207	3.8955	726.6667	1001.242	0.1884	133.5201	192.6943	7.4848	0.0101	70.3487	0.2277	0.4244
19:00 01/0	29.3026	6.5031	28.061	55.9484	0.2469	0.2104	2.0524	4.4506	651.2847	1002.186	0.1839	19.5777	193.3257	5.3279	3.6093	82.6359	0.2444	0.2532
20:00 01/0	28.617	6.1887	28.3017	51.9633	0.269	0.2414	2.235	3.9537	555.3993	1002.436	0.2205	1.1302	126.7962	3.1783	0.398	83.6081	0.2625	0.275
21:00 01/0	25.4839	6.1934	32.5504	51.0918	0.2531	0.1022	2.1159	4.0775	368.4896	1002.041	0.0781	0.4259	195.6063	3.6015	1.8164	86.5465	0.1217	0.3849
22:00 01/0	25.4465	6.0159	34.6612	45.1441	0.0619	0.058	2.0742	3.9417	330.434	1002.375	0.0539	0.0001	142.2258	1.5482	3.1224	87.8827	0.0615	0.063
23:00 01/0	25.5475	6.0697	41.2539	45.1602	0.1736	0.0929	2.1557	3.914	279.2708	1002.703	0.0774	0	127.0918	2.4582	1.5592	88.3465	0.0981	0.2226
00:00 02/0	26.5589	5.9278	34.3335	52.1445	0.1451	0.1115	2.0965	3.8313	216.3889	1002.122	0.0993	0.0109	167.0271	3.5988	0.2986	88.9201	0.1265	0.1562
01:00 02/0	27.363	6.161	30.6795	47.3094	0.1764	0.1377	2.008	4.1531	189.4965	1001.911	0.1158	0.0853	198.0034	2.6456	0	87.8564	0.1613	0.1866
02:00 02/0	27.3753	6.3913	29.5422	30.3952	0.157	0.1443	2.0491	4.3422	207.691	1002.101	0.1308	0.1436	111.8793	1.1484	0	88.7087	0.1569	0.1588
03:00 02/0	27.0179	6.2516	31.1931	20.8576	0.1754	0.1529	2.3262	3.9254	219.4444	1002.509	0.1367	0	83.9706	0.7307	0	90.039	0.1721	0.1796
04:00 02/0	27.3623	6.1919	31.3956	21.6726	0.2035	0.1711	2.1898	4.0022	214.8785	1002.185	0.1501	0	204.1325	1.9341	0	90.9694	0.1979	0.2099
05:00 02/0	27.7638	6.1504	33.5959	32.4319	0.2455	0.2188	2.0723	4.078	285.2604	1002.357	0.2015	0	195.2795	3.5877	0	90.8289	0.2375	0.2518
06:00 02/0	28.0859	6.0517	38.4523	35.1805	0.2088	0.1797	2.1431	3.9086	450.7986	1001.793	0.1537	20.8238	193.9678	3.8034	0	89.7204	0.2073	0.2135
07:00 02/0	28.2777	6.0806	40.5496	36.9218	0.2132	0.1834	2.1181	3.9625	464.1493	1002.268	0.164	68.2244	194.087	3.6507	0.0034	89.298	0.1983	0.2226
08:00 02/0	28.8247	5.9514	34.2759	31.3824	0.2191	0.1992	2.2151	3.7363	560.3299	1002.054	0.1778	179.5901	192.5753	4.2043	0	87.3953	0.2169	0.2223
09:00 02/0	29.8361	6.2315	31.5927	35.7494	0.2224	0.1988	2.1974	4.0341	482.8472	1002.002	0.1827	296.8486	191.6427	4.0248	0	81.907	0.2092	0.2303

Figure 2.2. Data at a monitoring station

2.3. Aerosol Optical Depth data

Aerosols are small, solid or liquid particles suspended in the air that can cause optical and chemical changes in the atmosphere. They serve as a source of condensation nuclei that form clouds and rain. Aerosol Optical Thickness (AOT) or Aerosol Optical Depth (AOD) is a measure of the attenuation of solar radiation due to the absorption and scattering of aerosol particles at a point observed relative to the upper atmospheric limit. An AOD value of 0 represents clean air, while higher values indicate greater pollution. AOD is a parameter used to measure the transparency of the atmospheric column from satellite data, and is closely related to the concentration of $PM_{2.5}$ dust at the ground, which is used as input for the concentration map estimation model. The AOD is calculated using data collected by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors on the Aqua and Terra satellites, and the Visible Infrared Imaging Radiometer Suite (VIIRS) sensors on the Suomi National Polar-Orbiting Partnership (Suomi NPP) satellite. The Terra and Aqua satellites were launched in 1999 and 2002, respectively, and have polar orbits. Terra passes the equator in the morning while Aqua crosses it in the afternoon. The scan band width of the MODIS sensor is about 2,330 km, and the orbital period is 16 days [17]. The respective

AOD products of Aqua MODIS and Terra MODIS (Collection 6.1, level 2) are MOD04_3K and MYD04_3K, with a spatial resolution of 3x3 km and a temporal resolution of daily collection during the period 2012-2021 [18].

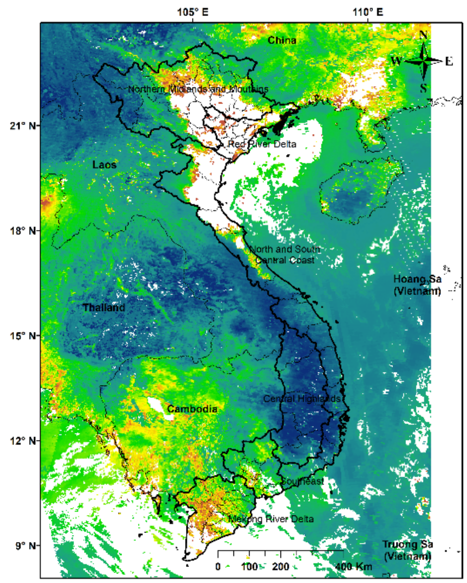


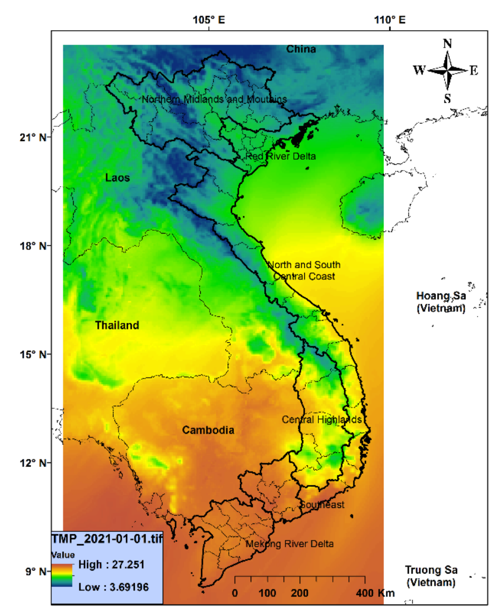
Figure 2.3. Daily AOD Map over Vietnam

VIIRS is a new generation sensor designed and improved after MODIS and mounted on the Suomi NPP satellite, which was launched on October 28, 2011. The satellite orbits at an altitude of 829 km, and the sweep width at the orthogonal position (swath width) reaches 3,040 km. The VIIRS sensor performs measurements of reflected and emitted radiation from the atmospheric system with 22 spectral bands, ranging from 412 nm to 12,050 nm. From VIIRS data, products of cloud, aerosol, surface temperature, fire, etc., are calculated and widely published globally [19]. The algorithm for calculating AOD for VIIRS data uses the same algorithm as for MODIS, but with adjustments and improvements. The VIIRS Environmental Data Record (EDR) is an official level 2 product provided by the US National Oceanic and Atmospheric Administration (NOAA). The product offers a 550 nm AOD with 6 km (2012-2020) and 750m (2020-2021) spatial resolutions, daily temporal resolution, with global coverage excluding areas with high cloud and unsuitable conditions. VIIRS AOD data was also collected for the period 2012-2021.

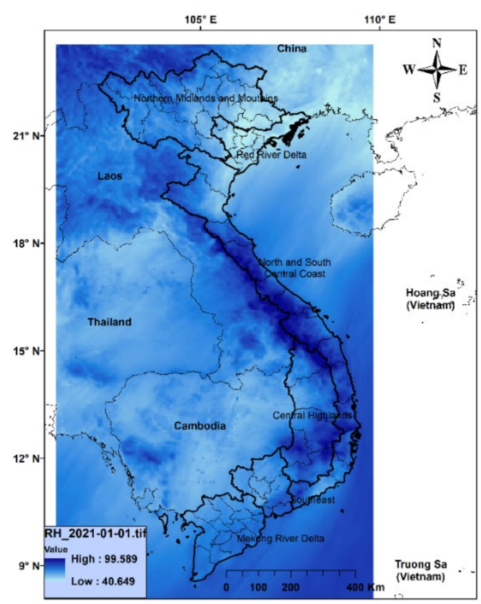
2.4. Meteorological data

Meteorological data, including temperature, humidity, pressure, wind speed, wind direction, and planetary boundary layer height (PBLH), were used as additional auxiliary parameters in the analysis. Meteorological factors have an impact on $PM_{2.5}$ dust concentration over time. For example, high humidity can increase the adhesion of dust in the air, making dust particles heavier and causing them to settle to the surface of the soil, which helps to reduce dust in the air. Strong winds can also move dust from one area to another, causing pollution levels to decrease in one area and increase in another. The lower the height of the planetary boundary layer, the higher the concentration of dust particles on the ground, which can cause more pollution, and vice versa.

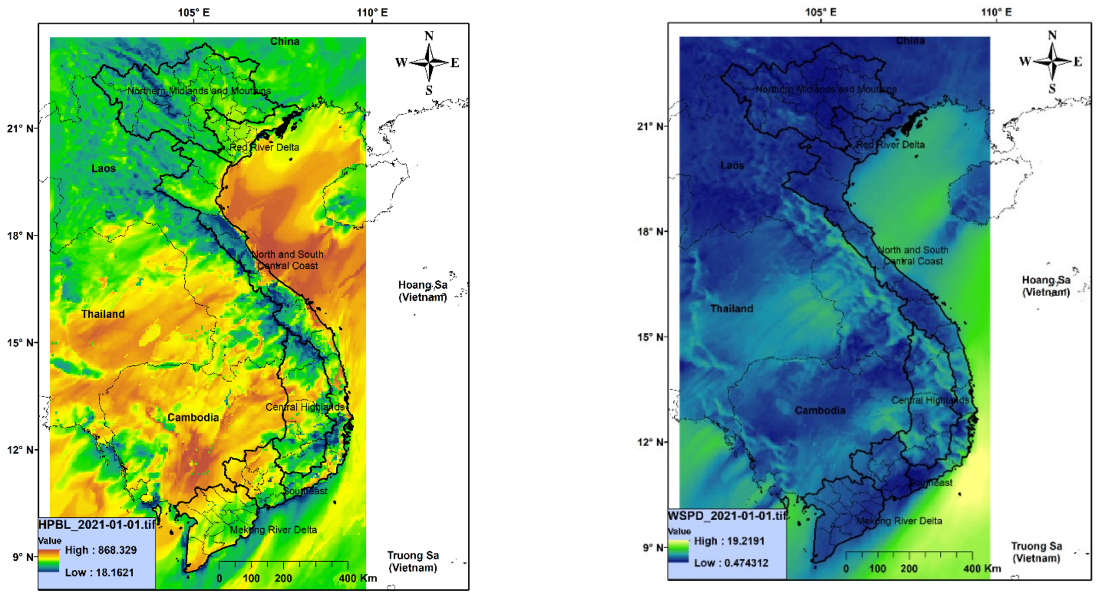
The meteorological maps are generated by the Weather Research and Forecasting (WRF) model [20], using the fifth generation ECMWF reanalysis (ERA-5) datasets as input. The ERA-5 datasets were collected between 2012 and 2021, with a spatial resolution of 0.25°×0.25° and a temporal resolution of hourly mean. The ERA-5 meteorological data is used as the initial and boundary conditions for the WRF model simulation. The WRF model is configured with two nested domains, with spatial resolutions of 15 km and 5 km, respectively. The output data of the WRF model is maps of meteorological variables, with a frequency of 4 images per day at 0h, 6h, 12h, and 18h (GMT+0), and a spatial resolution of 5 km (see Figure 2.4).



Temperature



Humidity



Planetary boundary layer height | Windspeed

Figure 2.4. Meteorological maps over Vietnam

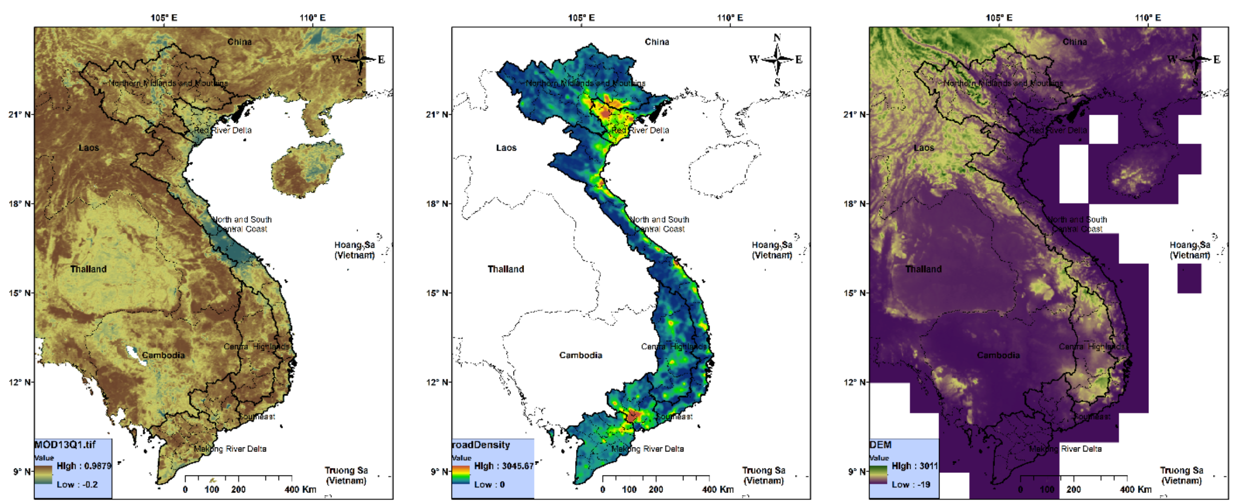
2.5. Land surface data

The land use information is utilized because it partly represents the sources of ground emissions, which is a crucial factor that affects the spatial distribution of $PM_{2.5}$ concentrations. For instance, areas with a lot of forests and trees tend to have cleaner air than areas with few trees. High traffic density areas with many vehicles tend to emit a lot of smoke and dust, thereby causing air pollution. Densely populated areas with many people's living and working activities can also result in increased air pollution. The land use data used in this study include the Normalized Difference Vegetation Index (NDVI), topographic map, and road map (as shown in Figure 2.5).

The NDVI map product used in this study is collected from Terra MODIS satellite images with the product name MOD13Q1, Collection 6, level 3. The spatial resolution of the product is 250 meters, and the time resolution is 16 days [21]. The product provides consistent observations of the vegetation status in the study area over the period of 2012 to 2021.

The road map used in this study was collected from the latest OpenStreetMap (OSM) source in 2022, in vector format containing the shapes of roads. OSM is a free and community-driven mapping service that provides global map objects, including data types such as node (point position on earth), way (polyline representing road objects, buildings, etc.), relation (relationship between objects), and tag (information about the object). Although the map is mainly edited by volunteers, the completeness and quality of the map are not uniform across different geographical locations. Despite this, OSM is still widely used in many applications in the geosciences, Earth observation, and environmental sciences [22].

The Digital Elevation Model (DEM) represents the topographic elevation of the Earth's surface in the form of a raster grid. The DEM ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) GLOBAL data third edition, provided by the Ministry of Economy, Trade, and Industry (METI) of Japan and the United States National Aeronautics and Space Administration (NASA) from 2019, was used in this study to describe the topography of Vietnam [23]. This data has a spatial resolution of 30m.



NDVI RoadDensity DEM

Figure 2.5. Land surface maps over Vietnam

3. METHODOLOGY

The method for constructing daily maps of $PM_{2.5}$ concentration is presented in Figure 3.1. This method has been inherited from the previous work of the research group [5] and updated according to the recently published study [6]. The input data includes $PM_{2.5}$ data from monitoring stations, AOD data from satellites, meteorological maps from the WRF model, vegetation index maps, road map, and topographic map. These data were preprocessed, and the maps and station data were integrated to create the model training dataset. A statistical model was then built for daily $PM_{2.5}$ mapping. The resulting daily $PM_{2.5}$ maps were aggregated into monthly and yearly averages and verified through comparison with station observations and with global $PM_{2.5}$ mapping products.

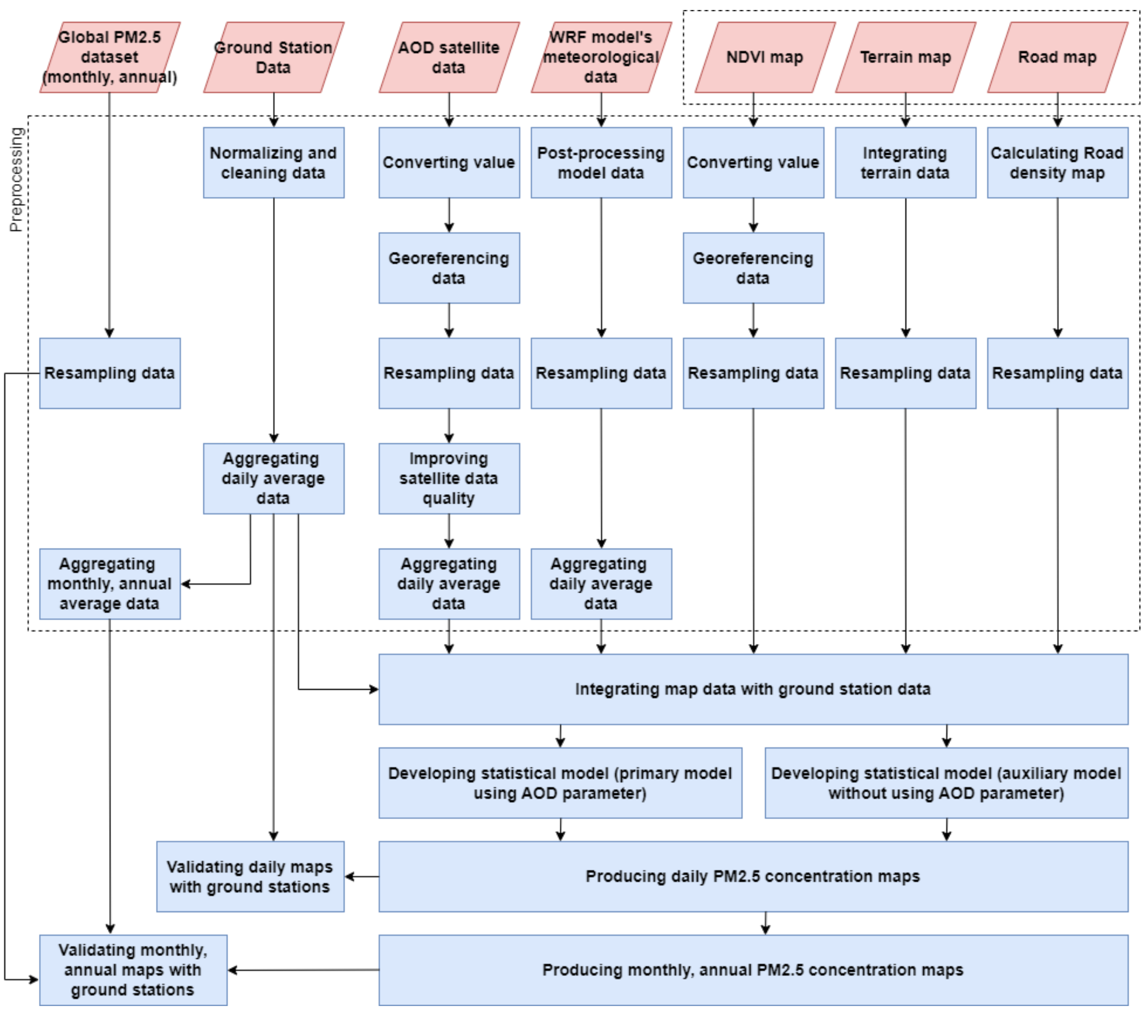


Figure 3.1. The process of building $PM_{2.5}$ concentration maps [6]

3.1.1. Data preprocessing

$PM_{2.5}$ concentration data from monitoring stations was normalized in terms of time format and uniform structure. The data was then cleaned by removing outliers, which was carried out in the following steps: (1) eliminating observations using a threshold with $PM_{2.5}$ values exceeding 500 µg/m³ or less than 3 µg/m³, (2) using statistical methods to detect outliers that are too high or too low compared to the measured data in the neighboring period (±15 days), (3) using statistical methods to detect outliers that are too high or too low compared to the measured data in the neighboring period (±15 hours) and in neighboring stations (within a 5km radius), (4) identifying outliers where the $PM_{2.5}$ value has not changed for an extended period and (5) identifying outliers where the measured $PM_{2.5}$ value is higher than the PM10 value [24]. These outliers were then manually checked for accuracy. After filtering the PM data, hourly data was aggregated into daily, monthly, and annual averages in preparation for data integration and modeling

Multi-source remote sensing data, including AOD products from Terra MODIS, Aqua MODIS, and VIIRS Suomi NPP, and NDVI products from Terra MODIS, shared the common feature of being a spatial grid of pixels. However, these data sources differed in format, temporal resolution, and spatial resolution. Therefore, preprocessing was required to standardize the remote sensing data into a consistent format and spatial grid. The preprocessing of satellite images involved extracting the dataset, converting pixel values, georeferencing, and resampling. This process was applied to both the AOD satellite image products and the NDVI products from Terra MODIS.

Dataset extraction and pixel value transformation involved extracting related data layers and recalculating pixel values based on metadata information such as offset and scale factor. Georeferencing is the process of mapping the internal coordinate system of a map to the geographic coordinate system on Earth. This involved using known geographic coordinates of ground control points (GCPs) to correct the geometric distortion of the map and determine the geographic coordinates of all other pixels.

Because satellite images have different coverage and resolutions, a common grid with uniform coverage and resolution was defined over the study area. Specifically, the grid covered the entire territory of Vietnam on the WGS84 geographic reference system, with each cell measuring 3 x 3 km. The multi-source spatial data was then resampled and projected onto this standard grid to create uniform map layers with the same spatial resolution and coverage over the Vietnam study area.

Nearest resampling was used for images with a spatial resolution greater than 3 km (such as VIIRS AOD and meteorological maps), while average resampling was used for images with a resolution less than 3 km (such as MODIS NDVI, road density map, and DEM). Resampling with the averaging method produces a smoother image than the nearest neighbor method and is typically used when converting input images of small spatial resolution to large spatial resolution. A pixel value of the output image is calculated as the average of the pixels in the corresponding window of the input image (Figure 3.2).

The GDAL tool was used to perform these processes. Finally, the AOD maps of each product category were aggregated into average daily maps.

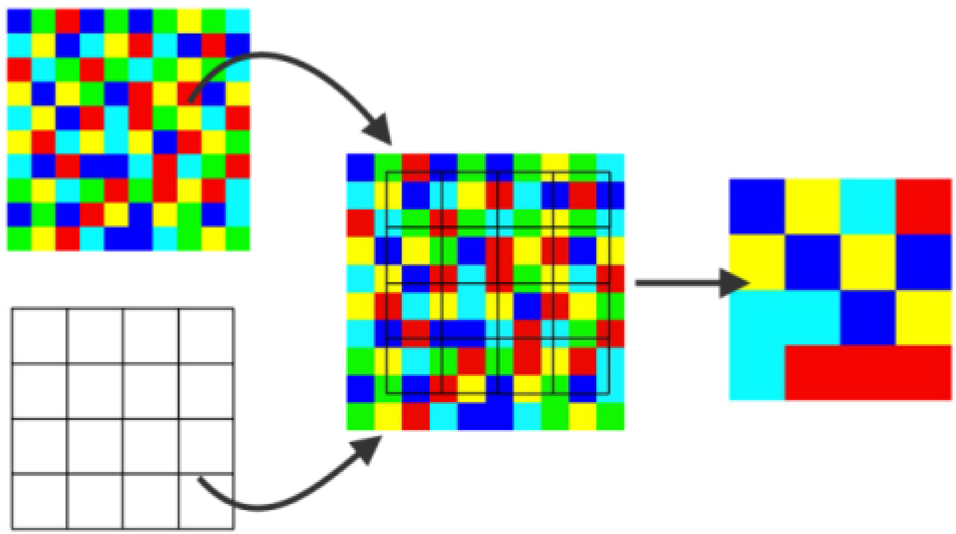


Figure 3.2. Average resampling example

The WRF model calculated meteorological data in *netcdf format. The Unified Post Processing (UPP) toolkit [25] was used to process the WRF model output data. UPP was developed at the National Center for Environmental Prediction (NCEP) and provides the ability to compute various fields and interpolate at different pressure levels with the results of running WRF models. We used the UPP tool to calculate (extract, interpolate) maps for temperature at 2m, humidity at 2m, pressure at 2m, planetary boundary layer height (PBLH), and wind vector components (U and V) at 10m. Then, these data were resampled to recalculate the values on a common coordinate grid consistent with other satellite image products in the study area. The meteorological maps were also aggregated into daily mean maps.

Vegetation index data (NDVI) is also a satellite image product (MODIS – MOD13Q1), so the preprocessing steps were similar to those described in AOD map preprocessing. The first step was to select the dataset to be extracted and exported, which was done by selecting the specific dataset for NDVI ("250m_16_days_NDVI"). Next, the dataset was converted and georeferenced. Finally, the dataset was resampled to match the common coordinate grid used for other satellite image products in the study area.

Road map data was provided in vector format, consisting of lines and road features. To use this data as input to the model, the Line Density tool from ArcMap 10.4.1 [26] was applied to convert the data to raster format. This tool calculated the ratio of the magnitude of the line features (within a radius around each cell) to the area of the whole circle. The radius was set to approximately 15 km. The output raster image was then resampled using the nearest neighbor method to ensure the road density map had the same coordinate grid as other maps.

The terrain data, after being downloaded, was already in GeoTiff format and georeferenced. However, the image range and spatial resolution (30m) were not consistent with other maps. We used the GDAL tool to reconstruct the map on a 3km grid in Vietnam, applying the average resampling method. The topographic data was then be normalized to the range of 0-1, and other parameters such as latitude and distance to the sea were also be normalized to the same range to ensure that no parameter took too much weight and to avoid negative values (<0) being affected by the model coefficient. The feature was computed using the following formula:

$$\sqrt[2]{\frac{Dist_{sea}}{DEM}} * Lat \quad (3.1)$$

Where DEM is the terrain elevation value, Dist_sea is the distance to the sea, and Lat is the geographical latitude. These are geographically related parameters that are combined into a single feature. Dist_sea and Lat have a positive relationship with air pollution, while DEM has a negative relationship with air pollution. Therefore, Dist_sea and Lat are placed above the denominator, and DEM is placed below the numerator. Dist_sea and DEM are added to the square root to reduce the spatial fluctuations of these two variables (which have large fluctuations) as the stations are not evenly distributed in Vietnam.

3.1.2. Enhance AOD data quality

Due to cloudy weather conditions in Vietnam, there was days when optical satellites cannot observe ground and atmospheric information, leading to a lack of AOD data. To enhance AOD data coverage and quality, we combined AOD products from different sources (MODIS Terra, MODIS Aqua, VIIRS Suomi NPP) on a day-to-day basis. The method used was the regression method on AOD calculated from MODIS Terra that had been previously studied in Vietnam [27]. First, the AOD map from MODIS Aqua and VIIRS Suomi NPP was co-located with the day-by-day MODIS Terra map, extracting the spatially overlapping pixels on the two products. Then, linear regression functions that calculated MODIS-Terra AOD from MODIS-Aqua AOD and VIIRS AOD were built based on the extracted data pairs. If a day had too few pairs of duplicate pixels, the window was enlarged to get the data of neighboring days. The linear regression function was calculated for each data type, on a day-to-day basis, and then applied to the MODIS-Aqua and VIIRS maps to generate the MODIS-Terra regression maps (Aqua regression to Terra, and VIIRS regression to Terra). This regression aimed to improve the quality of MODIS-Aqua AOD and VIIRS AOD data [27]. The regression maps and MODIS-Terra AOD maps were then combined to enhance the coverage of the AOD satellite data. At locations with MODIS-Terra AOD values, they remained the same, while at locations where this value did not exist (due to the influence of clouds or errors in the calculation process), they were replaced with the regression value from MODIS-Aqua AOD, or VIIRS AOD, or an average of both if available. The integrated map of AOD from multi-source products was used in the $PM_{2.5}$ concentration map estimation model in Vietnam

3.1.3. Data integration

After preprocessing the data, the map data layers were integrated with the station data to build the model training dataset. The model represented the relationship between the map parameters and the observed $PM_{2.5}$ concentration at ground level. The process of integrating map data and monitoring stations was subject to spatial and time constraints:

Spatial constraints: Data on map layers had to be extracted at the location of the monitoring station.

Time constraint: Data from the map layers and $PM_{2.5}$ monitoring at the station were respectively extracted over the same time period (daily average, monthly average, yearly average).

Monthly average monitoring station data was aggregated from daily average data if there was data for at least 50% of the days in that month. Annual average station data was synthesized from monthly average observations provided that there was monthly average data for at least 6 months.

3.1.4. Model development

The input data for the model included $PM_{2.5}$ monitoring data from the monitoring station, AOD satellite image maps, meteorological maps from the WRF model, vegetation index maps, road maps, and terrain data collected between 2012-2021. These data underwent preprocessing, after which the map and station data were integrated and normalized to create the model training dataset.

The dataset was first normalized, and then the supervised forward stepwise method was applied to select the optimal parameters for the computational model [28], [29]. Initially, univariate regression was performed between $PM_{2.5}$ stations and each input parameter to calculate the correlation coefficient R^2 for each case. The direction (+/-) of the coefficient showing positive/reverse correlation with $PM_{2.5}$ was then defined for each input parameter. Next, the input parameters were sorted in descending order of R^2, and the parameter with the highest R^2 and matching sign of the fitting coefficient was selected into the optimal parameter set of the model. These steps were repeated to add the parameters in sorted order, ensuring that the signs of the coefficients in the model matched and the adjust-R^2 index of the whole model improved by at least 1% (0.01). The process ended when no more parameters could be added to the model. After selecting the optimal set of parameters using this method, the research team checked and re-calibrated these parameters to suit the study area. However, due to the uneven distribution of monitoring stations across the country, the parameter selected from the statistical method based on station data may not have been representative in the entire study area, resulting in erroneous calculations of estimated $PM_{2.5}$ concentrations in areas without stations. Table 3.1 presents the list of parameters, sign of the coefficient.

Table 3.1. List of parameters and trend with $PM_{2.5}$

Parameters	**Sign (trend với $PM_{2.5}$)**
AOD	+
Temperature	-
Humidity	-

Planetary Boundary Layer Height	-
Pressure	+
Wind speed	-
NDVI	-
Road density	+
Terrain	+

A mixed-effects model (MEM) was used to estimate the $PM_{2.5}$ dust map. The model estimated daily $PM_{2.5}$ values at locations on the map based on AOD parameters, meteorological parameters, and land use with a random intercept. The formula for the MEM model is as follows:

$$PM_{2.5_{i,j}} = \sum_{k=1}^{N} \alpha_k \, X_{k,i,j} + (\alpha + \beta) \tag{3.2}$$

where $PM_{2.5\ i,j}$ is the estimated $PM_{2.5}$ at day i and spatial location j. $X_{k,i,j}$ is the kth parameter at day i and position j, N is the number of parameters used in the model. The coefficients α_k, α are fixed components (fixed effects), including the slope and intercept of input parameters, respectively. The parameter β is the random component of the random effect that varies from day to day.

Two statistical indicators were applied to test the model's parameters, including the p-value index to test the statistical significance of each parameter and the VIF (Variance Inflation Factor) index to test for multicollinearity among the parameters.

The p-value is a statistical index that measures the level of support or disapproval provided by the data sample for the null hypothesis H0. In this case, the null hypothesis H0 assumed that the coefficient of a model parameter (independent variable) is zero, and our goal was to reject this hypothesis. If the p-value was less than or equal to 0.05, then the null hypothesis H0 was then rejected. Conversely, if the p-value was greater than 0.05, there was no basis to reject the null hypothesis H0. Parameters in the model with p-values greater than 0.05 (no statistical significance) were removed from the model.

In case of VIF, the parameter with the highest VIF value (>3) in the model was removed (high VIF indicates that the independent variable is highly collinear with other variables in the model). After removing the parameter from the model, the VIF index was recalculated and filtered until all parameters in the model had a VIF < 3.

3.1.5. AOD data missing issue

Vietnam is an area characterized by cloudy weather conditions, especially in the North, where observations from optical satellites are limited. An assessment of monthly MODIS AOD and VIIRS AOD map coverage from 2012 to 2016 revealed that the monthly MODIS AOD coverage rate was quite low, ranging from 5% to 20%, while the VIIRS AOD coverage rate was higher, ranging from 25% to 75% [27]. Modeling using only the AOD parameter would result in some regions having no $PM_{2.5}$ estimates for many days. Therefore, to improve data coverage, we built a second auxiliary model that did not use the AOD parameter, but instead used the remaining parameters of the main model to estimate $PM_{2.5}$ in areas where AOD data was not available. The estimated $PM_{2.5}$ maps from these two models were averaged on a daily basis to create an integrated map with high data coverage. The integrated map from the two models was then verified and re-evaluated with observations from monitoring stations, compared with related research, and used for spatial and temporal analysis of air pollution.

3.1.6. Model and map validation

The model's quality was assessed using statistical indicators such as correlation (R^2, Pearson r) and error (RMSE, MRE) between the $PM_{2.5}$ estimation generated by the model and the actual observations from the monitoring stations. Additionally, the model was verified through two cross-validation techniques: 10-Fold Cross Validation and Leave One Station Out Cross Validation (LOSO CV). The 10-Fold CV technique involved randomly shuffling the dataset and dividing it into ten equal groups, using one group for model evaluation while the other nine groups were used for training. This process was repeated ten times, and the quality of the model was averaged from the ten evaluations. The trained and validated model was then used to generate daily $PM_{2.5}$ concentration maps with a spatial resolution of 3x3km. LOSO CV technique divided data into groups for each monitoring station, where one station's data was used for evaluation while the remaining stations' data was used for training. The quality of the model was averaged from N evaluations, where N was the number of monitoring stations.

The quality of daily $PM_{2.5}$ maps produced by the model was assessed by comparing them with the measurements from monitoring stations. The evaluation process involved extracting the $PM_{2.5}$ value at the location of the monitoring station from the daily $PM_{2.5}$ average map, and comparing it with the measured $PM_{2.5}$ value at the corresponding station. This process was repeated for all daily $PM_{2.5}$ maps that had monitoring station data, resulting in two $PM_{2.5}$ data series between the map and the station. Statistical indicators such as correlation (R^2, Pearson r) and error (RMSE, MRE) were used to evaluate these two data series collectively over the period 2019-2021, or evaluated individually for each year. The daily map data was then aggregated into monthly and annual average maps, which were also evaluated against the corresponding monthly/yearly average data at the stations. It should be noted that this comparison was limited due to the difference in scale, as the pixels on the map represented an area of 3x3km, while the monitoring stations represented a single point in the area.

In this report, the $PM_{2.5}$ maps generated were compared to a global mapping product developed by Van Donkelaar et al [30]. This product provided monthly and yearly average $PM_{2.5}$ datasets from 1998-2020, which were developed using the same method used to create the $PM_{2.5}$ exposure map for the Global Burden of Disease - GBD in 2010, 2013, and 2015. The global $PM_{2.5}$ maps were estimated by combining AODs from multiple satellites with the GEOs-Chem Chemical Transport Model and then corrected with station observations by using the Geographically Weighted Regression (GWR) method. To compare the two, the daily $PM_{2.5}$ maps generated in this study were combined into monthly/yearly mean maps. The average monthly/yearly map products and the global map were then compared and evaluated with the average monthly/yearly monitoring at stations in Vietnam for each year in the period from 2019-2021. Statistical parameters used to compare and evaluate models and maps include the Pearson correlation coefficient r, correlation coefficient R^2, Root Mean Square Error (RMSE), and Mean Relative Error (MRE):

$$\text{Pearson r} = \frac{\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x})}{\sqrt{\sum_{t=1}^{n}(y_t - \bar{y})^2}\sqrt{\sum_{t=1}^{n}(x_t - \bar{x})^2}} \quad (3.3)$$

$$\text{R2} = \frac{\left(\sum_{t=1}^{n}(y_t - \bar{y})(x_t - \bar{x})\right)^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2 \sum_{t=1}^{n}(x_t - \bar{x})^2} \quad (3.4)$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - x_t)^2} \quad (3.5)$$

$$\text{MRE} = \frac{1}{N}\sum_{t=1}^{N}\frac{|y_t - x_t|}{y_t}.100\% \quad (3.6)$$

Where x_t, y_t are the estimated values from the model (or extracted on the map) and the observed values at the station, respectively. $\bar{x}, \bar{y}$ is the average value of two data series with N elements. Correlation coefficient R^2, Pearson r represents the linear correlation relationship between the estimate from the model/map and the measurement at the station. The RMSE characterizes the absolute difference between the model/map estimate and the station measurement while the MRE error represents the relative difference between the model/map estimate and the station measurement.

4. EXPERIMENT AND RESULT

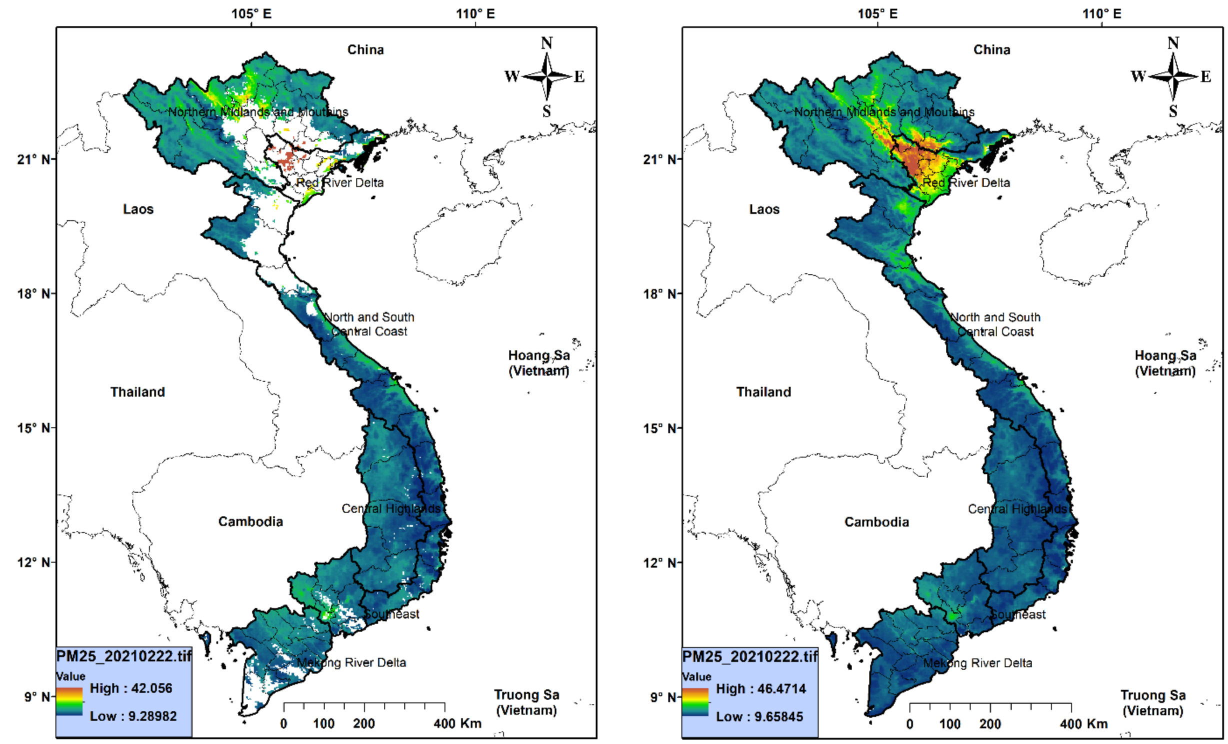
4.1. MEM evaluations

The parameters used in the MEM model after parameter selection include AOD, humidity, NDVI, PBLH, road density, and topography. Detailed information on the model parameters is presented in Table 4.1. The signs of the coefficients reflect the correlation relationship between the parameter and the $PM_{2.5}$ concentration. Additionally, all model parameters are statistically significant (p-value < 0.05) and there is no multicollinearity issue (VIF < 3).

Table 4.1. MEM model coefficients and statistical tests

Parameter	Coefficient	p-value	VIF
AOD	0.31	0.001	1.14
Humidity	-0.46	<2e-16	1.29
NDVI	-0.23	3.37e-07	1.53
Road Density	0.3	<2e-16	1.67
PBLH	-0.41	5.42e-10	1.2
Terrain	1.03	<2e-16	1.29

Based on the dataset of parameters selected during 2012-2021, the MEM model was developed to estimate the daily $PM_{2.5}$ map. Two models were built: the first model using the AOD parameter, and the second model that doesn't use the AOD parameter (as shown in Figure 4.1). The main model (using AOD) showed good quality, with a Pearson correlation coefficient r of 0.83, R^2 coefficient of 0.68, RMSE of 14.14 µg/m³, and MRE of 40.99% with 10,614 samples. On the other hand, the auxiliary model (without using AOD) demonstrated lower quality with a Pearson r of 0.81, R^2 of 0.65, RMSE of 14.8 µg/m³, and MRE of 48.58% with a sample number of 34208 (as summarized in Table 4.2)



AOD No AOD

Figure 4.1. Example map generated from the primary model (with AOD) and auxiliary model (no AOD)

The two MEM models mentioned above were cross-validated using 10-Fold Cross Validation and Leave-One-Station-Out Cross Validation. For the main model, 10-Fold CV resulted in a Pearson correlation coefficient of 0.7, R^2 coefficient of 0.49, RMSE error of 18.14 µg/m³, and MRE error of 52.71%. The auxiliary model had cross-validation results of Pearson coefficient r 0.75, R^2 coefficient 0.56, RMSE 16.64 µg/m³, and MRE 54.2%. The results of Leave-One-Station-Out cross-validation are presented in Table 4.2. It can be seen that the auxiliary model (without using AOD) had better results than the main model (using AOD). This may be due to the larger number of samples in the auxiliary model, which enabled the model to better learn the relationship between features and $PM_{2.5}$. Although the auxiliary model had better test results and higher map coverage, the main model using the AOD parameter was still important because AOD is closely related to terrestrial $PM_{2.5}$, and the spatial distribution of AOD contributes to explaining the subsurface $PM_{2.5}$ distribution. These evaluations were conducted on stations, which are unevenly distributed, with concentrations in certain areas.

Table 4.2. Model evaluations and cross-validation

		N	r	R^2	RMSE (µg/m³)	MRE (%)
Model validation	Primary	10,614	0.83	0.68	14.14	40.99
	Auxiliary	34,208	0.81	0.65	14.8	48.58
10-fold cross validation	Primary	1,061	0.7	0.49	18.14	52.71
	Auxiliary	3,420	0.75	0.56	16.64	54.2
Leave One Station Out Cross Validation	Primary	-	0.67	0.49	14.96	57.9
	Auxiliary	-	0.67	0.5	14.11	58.14

In the next section, the model's performance was compared with the results of related studies both domestically and internationally (Table 4.3). In Vietnam, the number of related studies is limited. Thanh et al's research on building $PM_{2.5}$ maps in Vietnam during the period from 2010-2014 yielded model efficiency using MODIS Terra AOD with R^2 = 0.6, RMSE = 8.5 µg/m³, MRE = 33.3% and MODIS Aqua AOD with R^2 = 0.58, RMSE = 8.8 µg/m³, MRE = 53.4% [5]. The MEM model's performance is competitive with other studies worldwide. In the US, the MEM model showed good results in cross-validation tests (10 fold CV) with R^2 of 0.82, 0.89, 0.66 and RMSE of 1.38 µg/m³, 2.33 µg/m³, 5.69 µg/m³ for datasets from 2000-2008 (161 stations) [31], 2003-2011 (161 stations) [32], and 2006-2012 (123 stations) [33], respectively. However, the MEM model showed lower accuracy in studies conducted in China with cross-validation tests yielding R^2 values from 0.77 to 0.8, NME (normalized mean error) values from 15.2% to 22.4%, and RMSE values from 12.47 to 23.07 µg/m³ [34] in 2013 with a dataset on 3 regions BTH (Beijing-Tianjin-Hebei), YRD (Yangtze River Delta), and PRD (Pearl River Delta) and corresponding numbers of stations 101, 85, 89. A study in 2014 using the GWR model yielded cross-validation results of R^2 = 0.64, RMSE = 32.98 µg/m³ on 835 stations [35] and R^2 = 0.85, RMSE = 24.86 µg/m³, MRE = 19.7% on 943 stations [36]. Another study in the Beijing-Tianjin-Hebei region in China combined MEM and GWR based on AOD product, meteorological, and land use parameters to estimate $PM_{2.5}$ from 2013 to 2020. The model had good cross-validation results with R^2 values ranging from 0.85–0.94, RMSE values ranging from 13.14–29.90 µg/m³, and MRE values ranging from 21.42%–32.71% [37]. The higher model error in the studies conducted in China may be due to the higher monitoring data used in the model (1-795 µg/m³) compared to the observed data in the US study (<40 µg/m³, rarely >60 µg/m³), resulting in higher errors. Another possible cause is that the observed values were local and not well represented in a raw pixel grid, causing the model to tend to underestimate when compared with the observations at the station [35]. In India, the MEM model has also been used to estimate a 10x10 km daily $PM_{2.5}$ map [38]. The model's accuracy improved when using more meteorological and land-use parameters, with R^2 increasing from 0.66 to 0.75 and RMSE decreasing from 29.11 µg/m³ to 20.9 µg/m³

Compared to the aforementioned studies, the constructed models in this study had an R^2 value between 0.65-0.68, an RMSE value between 14.14-14.8 µg/m³, and an MRE value between 40.99-48.58%. The 10-fold cross-validation R^2 value ranged between 0.49-0.56, the RMSE ranged between 16.64-18.14 µg/m³, and the MRE ranged between 52.71%-54.2%. The model error is similar to studies conducted in China, but higher than those in the US. This is because the observed $PM_{2.5}$ levels at stations in Vietnam were relatively high. Moreover, the error of the model also depended on various factors, such as the study area, year of study, quality and quantity of station data, and others.

Table 4.3. Related studies

Authors	**Study Area**	**Study period**	**Number of stations**	**Results**
Kloog et al, 2012[31]	U.S	2000-2008	161	MEM: -CV: R^2 = 0.82, RMSE = 1.38 µg/m³
Ma et al, 2014[39]	China	2014	835	GWR model: R^2 = 0.71, RMSE = 29.58 µg/m³ CV: R^2 = 0.64, RMSE = 32.98 µg/m³
Kloog et al, 2014[32]	U.S	2003–2011	161	MEM CV: R^2 = 0.88, RMSE = 2.33 µg/m³

Zheng et al, 2016[40]	China (BTH, YRD, and PRD)	2013	101, 85, 89	MEM: CV: R^2: 0.77 ~ 0.8, Normalized Mean Error (NME): 15.2%~22.4% RMSE: 12.47 ~ 23.07 µg/m³
Youet al, 2016[36]	China	2014	943	GWR model: - R^2 = 0.76, RMSE = 22.26 µg/m³, MRE = 22.5% CV: - R^2 = 0.79, RMSE = 20.85 µg/m³, RE = 20.8%
Lee et al, 2016[33]	U.S	2006-2012	123	MEM - R^2 = 0.75 - CV:R^2 = 0.66, RMSE = 5.69 µg/m³
Unnithan et al (2020)[38]	India	2017	33	MEM R^2=0.75, RMSE = 20.09 µg/m³
Yang et al, 2022[37]	China (Beijing–Tianjin–Hebei)	2013-2020	80	MEM + GWR R^2: 0.89-0.97, RMSE: 6.85 - 24.6 µg/m^3 CV: R^2: 0.85–0.95, RMSE: 7.87–29.90 µg/m^3
Peng et al, 2022[41]	Thailand	2020	23	Multiple linear regression R^2 = 0.48, RMSE = 9.16 µg/m³
Nguyễn et al, 2015[5]	Vietnam	2014	6	Multiple linear regression - MODIS Terra AOD: R^2 = 0.602, RMSE = 8.527 µg/m³, MRE = 33.348% - MODIS Aqua AOD: R^2 = 0.577, RMSE = 8.777 µg/m³, MRE = 53.353%

Current model	**Vietnam**	**2012-2021**	**65**	**MEM** - R^2 **= 0.65 – 0.68, RMSE = 14.14 – 14.8 µg/m³, MRE = 40.99% - 48.58%** **CV** - R^2 **= 0.49 – 0.56, RMSE = 16.64 - 18.14 µg/m³, MRE= 52.71% - 54.2%**

4.2. $PM_{2.5}$ map evaluations

4.2.1. Daily map

During the period of 2019-2021, the daily $PM_{2.5}$ integrated maps were evaluated by comparing observations from monitoring stations. The number of stations used for evaluation was reduced due to data preprocessing, with stations having abnormal data or stations with insufficient data for more than 50% of the aggregate time discarded. The overall comparison results for all stations during this period are shown in Figure 4.2. The integrated map showed a high correlation with the monitoring stations, with the Pearson correlation coefficient (r) reaching 0.8 and the coefficient of determination (R^2) reaching 0.64. The RMSE error was 15.44 µg/m³ and the MRE error was 51.75%

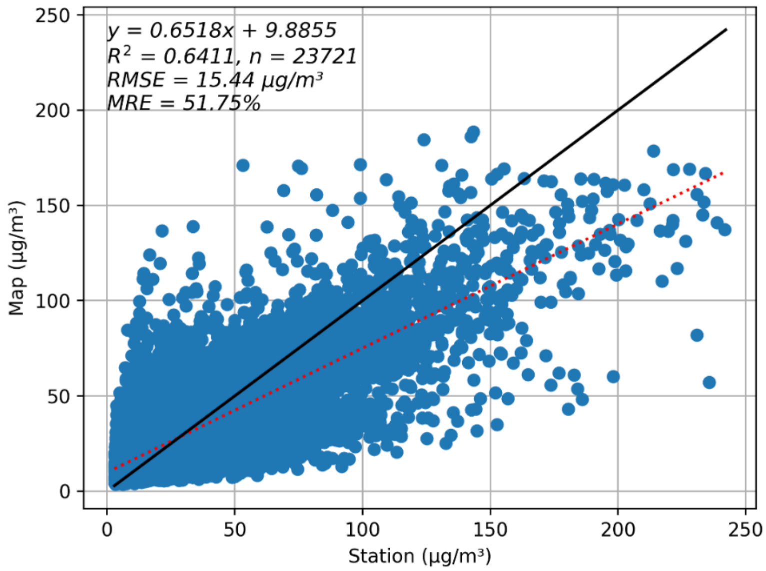


Figure 4.2. Scatter plot between map and station values

However, the evaluation results varied across the different monitoring stations (as shown in Table 4.4). The standard monitoring stations are concentrated in the provinces/cities of Hanoi, Bac Ninh, Quang Ninh, and Hai Duong in the North, and in Ho Chi Minh City in the South. The R^2 index assessed at the stations ranged from 0.07 to 0.95, the Pearson r index ranged from 0.27-0.97, the RMSE error ranged from 3.4 to 43.1 µg/m³, and the MRE error ranges from 17.6% to 268%. The integrated map had a high correlation rating at stations in Hanoi and Bac Ninh (most $R^2 > 0.7$), and a low correlation at stations located in Can Tho, Gia Lai, and Son La ($R^2 < 0.1$). The $PM_{2.5}$ value had a low relative error of MRE at some stations in Hanoi, Bac Ninh, and Ho Chi Minh (MRE < 30%), while a very high error (MRE > 100%) appeared at some stations in Quang Ninh, Hai Duong, and Nghe An.

Out of a total of 65 assessed stations, most of the new stations were installed, operated, and provided data from 2020. The old stations (9 stations) were installed before (2012-2017) and have relatively good and stable results. These include stations in Hanoi: Nguyen Van Cu (2010-present), Hanoi Environmental Protection Agency (2017-present), and US Embassy (2015-present); stations in Phu Tho (2013-present), Quang Ninh (2013-present), Hue (2013-present), Da Nang (2010-present), and stations in Ho Chi Minh City: US Embassy (2016-present) and University of Natural Sciences (2013-2019).

Monitoring stations with low evaluation results may be due to the lack of accurate estimation models at those locations, but it could also be due to the quality of data at some stations not being stable or guaranteed. Therefore, there is a need to improve the quality of the model as well as have a better station data QA-QC process in the future.

Table 4.4. Comparison results between daily average $PM_{2.5}$ map and each monitoring station during 2019-2021.

ID	City/ Province	Station	N	R^2	Pearson r	RMSE	MRE	Mean (station)	Mean (map)
11	Bac Ninh	Bac Ninh: Chau Khe - Tu Son	418	0.85	0.92	10.6	25.8	36.2	38.4
16	Bac Ninh	Bac Ninh: Phu Lang - Que Vo	382	0.83	0.91	16.5	35.6	38	32.8
42	Bac Ninh	Bac Ninh: gan KCN Que Vo	439	0.75	0.86	16.3	63.1	37	37.6
43	Bac Ninh	Bac Ninh: gan KCN Tien Son	182	0.91	0.95	21.5	72.6	35.4	37.4
44	Bac Ninh	Bac Ninh: UBND xa Dai Dong - Tien Du	568	0.67	0.82	20.8	74.2	36.9	36.9
69	Bac Ninh	Bac Ninh: Binh Dinh - Luong Tai	241	0.93	0.97	8.4	24.2	39.8	41.8
70	Bac Ninh	Bac Ninh: UBND xa Xuan Lam - Thuan Thanh	519	0.79	0.89	18.4	62.7	33.7	33.4
81	Bac Ninh	Bac Ninh: Khu lien co Thuan Thanh	414	0.93	0.96	16.1	20.8	41.3	33.8
82	Bac Ninh	Bac Ninh: UBND P.Dong Nguyen - Tu Son	399	0.95	0.97	20	20.3	52.5	40.4
83	Bac Ninh	Bac Ninh: gan KCN Yen Phong	618	0.79	0.89	12.6	33.6	34.6	39
84	Bac Ninh	Bac Ninh: UBND huyen Yen Phong	569	0.75	0.87	15.1	36.1	40.5	43.6
85	Bac Ninh	Bac Ninh: Phu Xa - Van Mon	394	0.18	0.43	31.7	75.3	45.7	37.9
86	Bac Ninh	Bac Ninh: Phong Coc-Long Duc-Que Vo	583	0.85	0.92	10.1	24.6	34.3	33.8
87	Bac Ninh	Bac Ninh: Tieu hoc Dai Bai - Gia Binh	586	0.86	0.93	9.3	27	33	32.3
88	Bac Ninh	Bac Ninh: UBND huyen Que Vo	671	0.87	0.93	12.4	19.4	39.4	35.1
94	Bac Ninh	Bac Ninh: TT Quan trac P. Suoi Hoa	540	0.92	0.96	8.3	26.4	28.8	29.8
95	Bac Ninh	Bac Ninh: UBND xa Cao Duc - Gia Binh	436	0.93	0.96	10.4	21.9	36.6	34.3
110	Can Tho	Can Tho: Ninh Kieu	126	0.07	0.27	4.6	47.4	8.6	10
22	Cao Bang	Cao Bang: TTQMT - P. Song Hien	451	0.66	0.81	43.1	62.3	54.6	19.4
78	Da Nang	Da Nang: 41 duong Le Duan	460	0.18	0.42	10.8	51.3	19.7	20.8
73	Gia Lai	Gia Lai: UBND TX. An Khe	420	0.11	0.34	8.4	60.2	12.3	12.6
92	Gia Lai	Gia Lai: TTQT TN&MT - TP Pleiku	621	0.08	0.28	11.5	69.7	13.3	13.3

101	Gia Lai	Gia Lai: BQL KCN Tra Da	371	0.13	0.36	11.5	55.7	14.2	12.4
108	Gia Lai	Gia Lai: Phu Dong - TP. Pleiku	92	0.18	0.43	3.4	37.9	8.2	8.7
3	Ha Noi	Ha Noi: Dai su quan Phap	23	0.79	0.89	10.8	28.9	36.4	44.7
71	Ha Noi	Ha Noi: UBND P. Minh Khai, Bac Tu Liem	522	0.85	0.92	12	20	46.8	41.5
72	Ha Noi	Ha Noi: Chi cuc BVMT	983	0.85	0.92	10.4	18.2	44.4	41.8
77	Ha Noi	Ha Noi: Lang Thuong - Dong Da	128	0.75	0.86	7.7	52.9	17.4	23.4
107	Ha Noi	Ha Noi: Hong Ha - Hoan Kiem	123	0.77	0.88	6.5	36.7	19.9	24.3
130	Ha Noi	Ha Noi: 556 Nguyen Van Cu	958	0.8	0.89	15.3	64.8	28.6	39.5
200	Ha Noi	Ha Noi: DSQ My	799	0.92	0.96	10.6	17.6	44.8	39
41	Hai Duong	Hai Duong: UBND TP Chi Linh	142	0.75	0.86	12.5	37.1	30.8	31.3
96	Hai Duong	Hai Duong: so TNMT - Tan Binh	297	0.35	0.59	32.3	268	10.1	35.3
97	Hai Duong	Hai Duong: UBND xa Co Dung - Kim Thanh	297	0.28	0.53	22.9	154.1	19.5	31.6
99	Hai Duong	Hai Duong: Tram Y te Duy Tan - Kinh Mon	301	0.69	0.83	16.8	133	11.4	24.7
103	Hai Phong	Hai Phong: Phu Lien, Nui Kha Lam, Kien An	115	0.26	0.51	6.9	75.2	10.2	14.5
201	Ho Chi Minh	Ho Chi Minh: DSQ My	1055	0.31	0.55	10.6	32.2	23.6	20.8
202	Ho Chi Minh	Ho Chi Minh: DH KHTN	85	0.55	0.74	7.2	27.8	20.3	23.7
211	Ho Chi Minh	Ho Chi Minh: DHQG TP. HCM - Thu Duc	173	0.1	0.32	10	37.4	22.1	16.7
212	Ho Chi Minh	Ho Chi Minh: gan phong GDDT Q Binh Tan	249	0.3	0.54	7.8	30.2	18.8	16.6
213	Ho Chi Minh	Ho Chi Minh: Tram phat song MobiFone - KCN Tan Binh	226	0.24	0.49	12	34.9	23.6	17.6
214	Ho Chi Minh	Ho Chi Minh: THCS Cu Chinh Lan - Q. Binh Thanh	260	0.17	0.42	11.9	33.8	24.9	18.1
215	Ho Chi Minh	Ho Chi Minh: Bao Thanh Nien - Nguyen Dinh Chieu - Q3	105	0.21	0.46	5.8	22.7	17	16.4
216	Ho Chi Minh	Ho Chi Minh: MobiFone Thanh Thai, Q10	213	0.3	0.55	7.9	29.7	18	18.6
125	Hue	Thua Thien Hue: 83 duong Hung Vuong	454	0.37	0.61	9.5	67.1	14.2	19.6
115	Lang Son	Lang Son: phuong Chi Lang	249	0.8	0.89	29.7	59.1	40.5	15.8
116	Lang Son	Lang Son: xa Hong Phong - Cao Loc	318	0.76	0.87	5.1	30.5	13.7	15.5
123	Lao Cai	Lao Cai: Tram gan KCN Tang Loong 2	145	0.72	0.85	15.4	32.5	26	17
100	Nghe An	Nghe An: TTQT TN&MT Nghe An	320	0.47	0.69	9.5	55.4	17.7	17.6
105	Nghe An	Nghe An: Truong Thi - TP.Vinh	102	0.12	0.35	7.3	100.1	9.1	13.5

104	Ninh Binh	Ninh Binh: Cuc Phuong, Nho Quan	113	0.53	0.73	8	91.1	11.6	17.7
50	Phu Tho	Phu Tho: duong Hung Vuong - Viet Tri	708	0.7	0.84	15.6	80.6	23.3	34.4
17	Quang Ninh	Quang Ninh: Cam Thinh - Cam Pha	92	0.28	0.53	12.7	74.8	17	23.7
58	Quang Ninh	Quang Ninh: UBND P. Mao Khe - Dong Trieu	308	0.48	0.69	22.7	180.1	13.6	31.5
59	Quang Ninh	Quang Ninh: UBND TP Uong Bi	475	0.19	0.44	15.5	86.9	13.5	22
63	Quang Ninh	Quang Ninh: Thanh uy Mong Cai	377	0.4	0.63	11.4	74.5	13.8	20.3
65	Quang Ninh	Quang Ninh: Gan KCN Cai Lan	409	0.46	0.68	8.5	37.8	16.9	18
66	Quang Ninh	Quang Ninh: UBND huyen Hai Ha	229	0.4	0.64	11	109.1	10.2	18.6
67	Quang Ninh	Quang Ninh: K11 - Minh Thanh	384	0.34	0.58	9.3	60.5	12.8	16.8
79	Quang Ninh	Quang Ninh: P. Hong Ha - Ha Long	621	0.52	0.72	9.4	39.5	20	21.4
102	Son La	Son La: P. To Hieu	97	0.08	0.29	5.3	40.9	8.5	8.7
75	Tay Ninh	Tay Ninh: TX. Trang Bang	41	0.33	0.57	4.7	27.2	12.8	11.2
93	Thai Nguyen	Thai nguyen: duong Hung Vuong	647	0.79	0.89	20.9	32.9	41.9	28.5
45	Vung Tau	Vung Tau: Nga tu Gieng nuoc	34	0.45	0.67	6.3	52.2	14.9	15.5
46	Vung Tau	Vung Tau: Tieu hoc Toc Tien - TX.Phu My	44	0.28	0.53	11.7	43.5	19.7	12

4.2.2. Monthly map

The daily average $PM_{2.5}$ maps were aggregated into a monthly average map and compared with ground-based measurement data for the period of 2019-2021. The monthly research maps were compared with the global average monthly product map of Van Donkelaar et al. [42] using the same set of ground measurement data to evaluate the quality of the two products. The results of the comparison are presented in Table 4.5. The monthly research maps had higher quality than the global production monthly maps for each year. In 2019-2020, our maps had an R^2 ranging from 0.62-0.81, RMSE error from 7.68-10.67 µg/m³, and MRE from 26.14-34.55%, while the global map had an R^2 ranging from 0.44-0.69, RMSE from 9.3-13.11 µg/m³, and MRE from 29.78-48.06%. In 2021, the global map had not been published yet, and our map had an R^2 rating of 0.6, RMSE of 13.43 µg/m³, and MRE of 43.14%.

Table 4.5. Overall assessment of the monthly average map with observations at monitoring stations in the period of 2019 - 2021

	Study's maps					Global maps				
Year	**N**	R^2	**Pearson r**	**RMSE (µg/m³)**	**MRE (%)**	**N**	R^2	**Pearson r**	**RMSE (µg/m³)**	**MRE (%)**
2019	65	0.81	0.9	7.68	26.14	65	0.69	0.83	9.30	29.78
2020	321	0.62	0.79	10.67	34.55	321	0.44	0.66	13.11	48.06
2021	447	0.6	0.77	13.43	43.14	-	-	-	-	-

The monthly average $PM_{2.5}$ concentration maps during the period of 2019-2021 were evaluated using data from 65 monitoring stations nationwide. Stations that did not have at least 50% of the data for the month were removed before aggregating on a monthly average. The overall evaluation index on all stations, including a Pearson correlation coefficient (r) of 0.79, an R^2 coefficient of 0.62, an RMSE error of 12.05 µg/m³, and an MRE error of 38.5%, showed that the results of this assessment varied from station to station (Table 4.6).

The monthly average $PM_{2.5}$ map had a low RMSE error (<5 µg/m³) at stations in Tay Ninh, Gia Lai, Ho Chi Minh, Son La, Vung Tau, and Can Tho, but showed high RMSE errors (> 30 µg/m³) at some stations in Bac Ninh, Hai Duong, Cao Bang, and Lang Son. The map had a low relative error of MRE (<10%) in Hanoi, Bac Ninh, and Ho Chi Minh and a relatively high error (MRE > 100%) appeared at some stations in Quang Ninh and Hai Duong.

During this period, some stations had very little average monthly data to compare with the map, such as Hanoi: French Embassy (1 month), stations in Vung Tau (1 month), Tay Ninh (1 month), Son La: To Hieu Ward (3 months), Nghe An (3 months), Ho Chi Minh: University of Science (3 months), and others. This indicates that during the period of 2019-2021, some stations operated unstably, and the data was not provided continuously, which may have affected the map evaluation results.

Table 4.6. Comparison results of the monthly average $PM_{2.5}$ map with monitoring at each station during the period of 2019-2021

ID	City/ Province	Station	N	RMSE	MRE	Mean (station)	Mean (map)

22	Cao Bang	Cao Bang: TTQMT - P. Song Hien	16	39.1	65.4	53.5	18.4
123	Lao Cai	Lao Cai: Tram gan KCN Tang Loong 2	5	12.1	30.7	28.9	19
116	Lang Son	Lang Son: xa Hong Phong - Cao Loc	10	2.6	11.6	14	15.8
115	Lang Son	Lang Son: phuong Chi Lang	9	27.2	60.9	40.2	15.9
93	Thai Nguyen	Thai nguyen: duong Hung Vuong	22	15.9	31.2	41.3	28.3
63	Quang Ninh	Quang Ninh: Thanh uy Mong Cai	16	9.4	53.3	13.7	20.7
66	Quang Ninh	Quang Ninh: UBND huyen Hai Ha	9	8.8	91.5	9.8	17.7
50	Phu Tho	Phu Tho: duong Hung Vuong - Viet Tri	27	12.7	73.4	22.8	34
102	Son La	Son La: P. To Hieu	3	1.7	18.6	7.6	9
84	Bac Ninh	Bac Ninh: UBND huyen Yen Phong	19	12.4	30.6	40.3	43.8
94	Bac Ninh	Bac Ninh: TT Quan trac P. Suoi Hoa	19	5.9	12.4	28.8	28.6
85	Bac Ninh	Bac Ninh: Phu Xa - Van Mon	13	21.3	38.3	44.9	38.3
16	Bac Ninh	Bac Ninh: Phu Lang - Que Vo	14	9.3	21.6	39.4	37.5
42	Bac Ninh	Bac Ninh: gan KCN Que Vo	16	11.3	40.6	37.6	37.4
88	Bac Ninh	Bac Ninh: UBND huyen Que Vo	23	8.1	12.4	38.4	34.3
83	Bac Ninh	Bac Ninh: gan KCN Yen Phong	21	8	16.7	34.5	36.6
82	Bac Ninh	Bac Ninh: UBND P.Dong Nguyen - Tu Son	14	16.6	20.7	53.4	40.9
86	Bac Ninh	Bac Ninh: Phong Coc-Long Duc-Que Vo	21	4.2	12.4	34.3	33.8
43	Bac Ninh	Bac Ninh: gan KCN Tien Son	6	20	63	40.3	40.7
41	Hai Duong	Hai Duong: UBND TP Chi Linh	5	4.6	26.3	29.3	30.3
11	Bac Ninh	Bac Ninh: Chau Khe - Tu Son	14	7.4	20.2	36.3	39.2
44	Bac Ninh	Bac Ninh: UBND xa Dai Dong - Tien Du	18	15.4	63.2	34.9	36.3
95	Bac Ninh	Bac Ninh: UBND xa Cao Duc - Gia Binh	15	6.1	12.4	36	34
58	Quang Ninh	Quang Ninh: UBND P. Mao Khe - Dong Trieu	11	19.7	147	13	30.1
71	Ha Noi	Ha Noi: UBND P. Minh Khai, Bac Tu Liem	18	9.1	16.5	46.9	40.9
130	Ha Noi	Ha Noi: 556 Nguyen Van Cu	34	13.3	55.2	28.4	39.3
107	Ha Noi	Ha Noi: Hong Ha - Hoan Kiem	4	4.9	23.2	19.9	24.5
87	Bac Ninh	Bac Ninh: Tieu hoc Dai Bai - Gia Binh	20	5.4	14.4	31.9	31.4
59	Quang Ninh	Quang Ninh: UBND TP Uong Bi	17	13.2	71	13.6	21.7
69	Bac Ninh	Bac Ninh: Binh Dinh - Luong Tai	8	4.4	11.5	40.1	43.8
99	Hai Duong	Hai Duong: Tram Y te Duy Tan - Kinh Mon	11	15.2	110.9	11.2	23.7
81	Bac Ninh	Bac Ninh: Khu lien co Thuan Thanh	14	13.9	23.8	42	31.1

70	Bac Ninh	Bac Ninh: UBND xa Xuan Lam - Thuan Thanh	19	16.8	53.5	35.3	33.5
200	Ha Noi	Ha Noi: DSQ My	27	6.8	13.1	44.7	39.2
3	Ha Noi	Ha Noi: Dai su quan Phap	1	3.8	10.7	36	39.8
77	Ha Noi	Ha Noi: Lang Thuong - Dong Da	4	6.5	36.5	17.4	23.8
72	Ha Noi	Ha Noi: Chi cuc BVMT	32	5.9	12.6	44.7	41.9
17	Quang Ninh	Quang Ninh: Cam Thinh - Cam Pha	4	7.4	40.5	16.7	23.7
67	Quang Ninh	Quang Ninh: K11 - Minh Thanh	13	6.4	37.1	13.2	16.8
97	Hai Duong	Hai Duong: UBND xa Co Dung - Kim Thanh	11	18.9	129.8	18.7	30.1
65	Quang Ninh	Quang Ninh: Gan KCN Cai Lan	15	5.2	21.1	16.8	18
79	Quang Ninh	Quang Ninh: P. Hong Ha - Ha Long	23	6.5	30	20.3	21.5
96	Hai Duong	Hai Duong: so TNMT - Tan Binh	12	26	207.6	10.3	31.8
103	Hai Phong	Hai Phong: Phu Lien, Nui Kha Lam, Kien An	4	4.8	43.6	10.2	14.5
104	Ninh Binh	Ninh Binh: Cuc Phuong, Nho Quan	4	6.7	60.6	11.5	18.1
100	Nghe An	Nghe An: TTQT TN&MT Nghe An	11	5.4	34.9	17.2	17.7
105	Nghe An	Nghe An: Truong Thi - TP.Vinh	3	5.3	69.4	7.9	13.1
125	Hue	Thua Thien Hue: 83 duong Hung Vuong	16	6.2	49.6	13.9	19.4
78	Da Nang	Da Nang: 41 duong Le Duan	16	6	24.9	19.7	20.7
101	Gia Lai	Gia Lai: BQL KCN Tra Da	13	7.1	33.2	14.2	12.1
92	Gia Lai	Gia Lai: TTQT TN&MT - TP Pleiku	22	7.5	43.5	13.2	13.3
108	Gia Lai	Gia Lai: Phu Dong - TP. Pleiku	4	1.2	15.2	8.2	9.2
73	Gia Lai	Gia Lai: UBND TX. An Khe	15	4.2	24.6	12.3	12.6
75	Tay Ninh	Tay Ninh: TX. Trang Bang	1	0.3	3.7	9.4	9
211	Ho Chi Minh	Ho Chi Minh: DHQG TP. HCM - Thu Duc	6	6.4	27.4	22.5	17.7
213	Ho Chi Minh	Ho Chi Minh: Tram phat song MobiFone - KCN Tan Binh	8	6.5	25	23.9	18.3
214	Ho Chi Minh	Ho Chi Minh: THCS Cu Chinh Lan - Q. Binh Thanh	10	7.7	27.5	24.3	17.7
201	Ho Chi Minh	Ho Chi Minh: DSQ My	36	4.9	17	23.7	20.8
216	Ho Chi Minh	Ho Chi Minh: MobiFone Thanh Thai, Q10	8	3.1	10.4	17.6	18.2
215	Ho Chi Minh	Ho Chi Minh: Bao Thanh Nien - Nguyen Dinh Chieu - Q3	5	1.4	6.5	16.6	16.4
202	Ho Chi Minh	Ho Chi Minh: DH KHTN	3	4.4	22.4	20	24.3
212	Ho Chi Minh	Ho Chi Minh: gan phong GDDT Q Binh Tan	9	2.8	13.6	18.4	16.1

46	Vung Tau	Vung Tau: Tieu hoc Toc Tien - TX.Phu My	1	6.8	34.5	19.6	12.9
45	Vung Tau	Vung Tau: Nga tu Gieng nuoc	1	1.9	14.2	13.4	15.3
110	Can Tho	Can Tho: Ninh Kieu	4	2	23.3	8.6	10.2

The map showing the average monthly concentration of $PM_{2.5}$ dust during the period from 2019-2021 in Vietnam can be seen in Figures 4.3 and 4.4. $PM_{2.5}$ concentrations were found to be high during the winter period from November to March, with January being the month with the highest concentration. This phenomenon can be attributed to the climatic conditions, in addition to the influence of the northeast monsoon, which brings $PM_{2.5}$ pollution from the North to this area [43]. The concentration of $PM_{2.5}$ usually decreases during the summer months from May to September due to the rainy season with high temperatures and strong winds. Rain helps to reduce the concentration of $PM_{2.5}$ in the air, while the summer weather makes pollutants easy to diffuse. The monthly fluctuations in $PM_{2.5}$ concentrations were found to be similar across the years (Figure 4.2). During the last few months of 2019 (August – December), Vietnam's $PM_{2.5}$ concentration tended to be significantly higher than in 2020 and 2021. Additionally, $PM_{2.5}$ concentration in January 2021 was unusually high, as observed in monitoring stations located in Hanoi, Bac Ninh, and Ho Chi Minh (Figure 4.5)

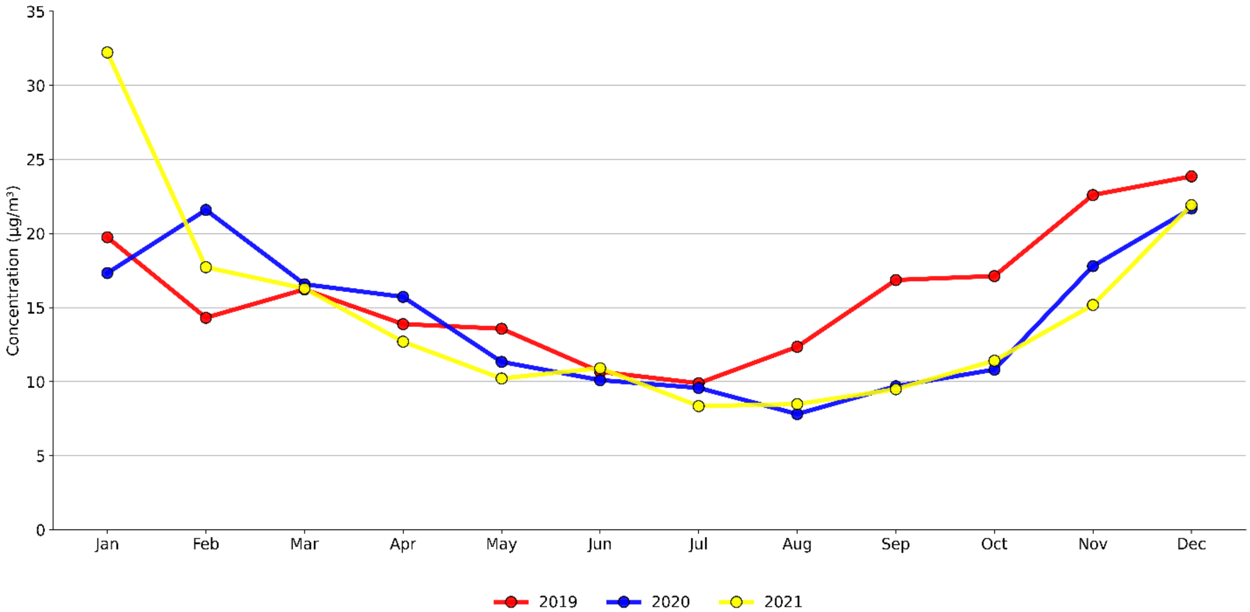
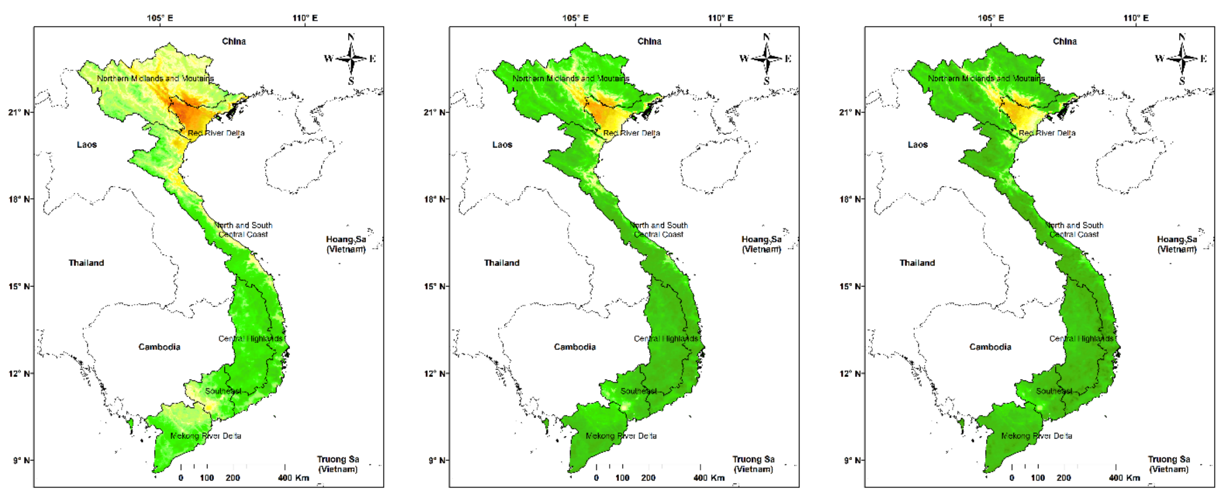
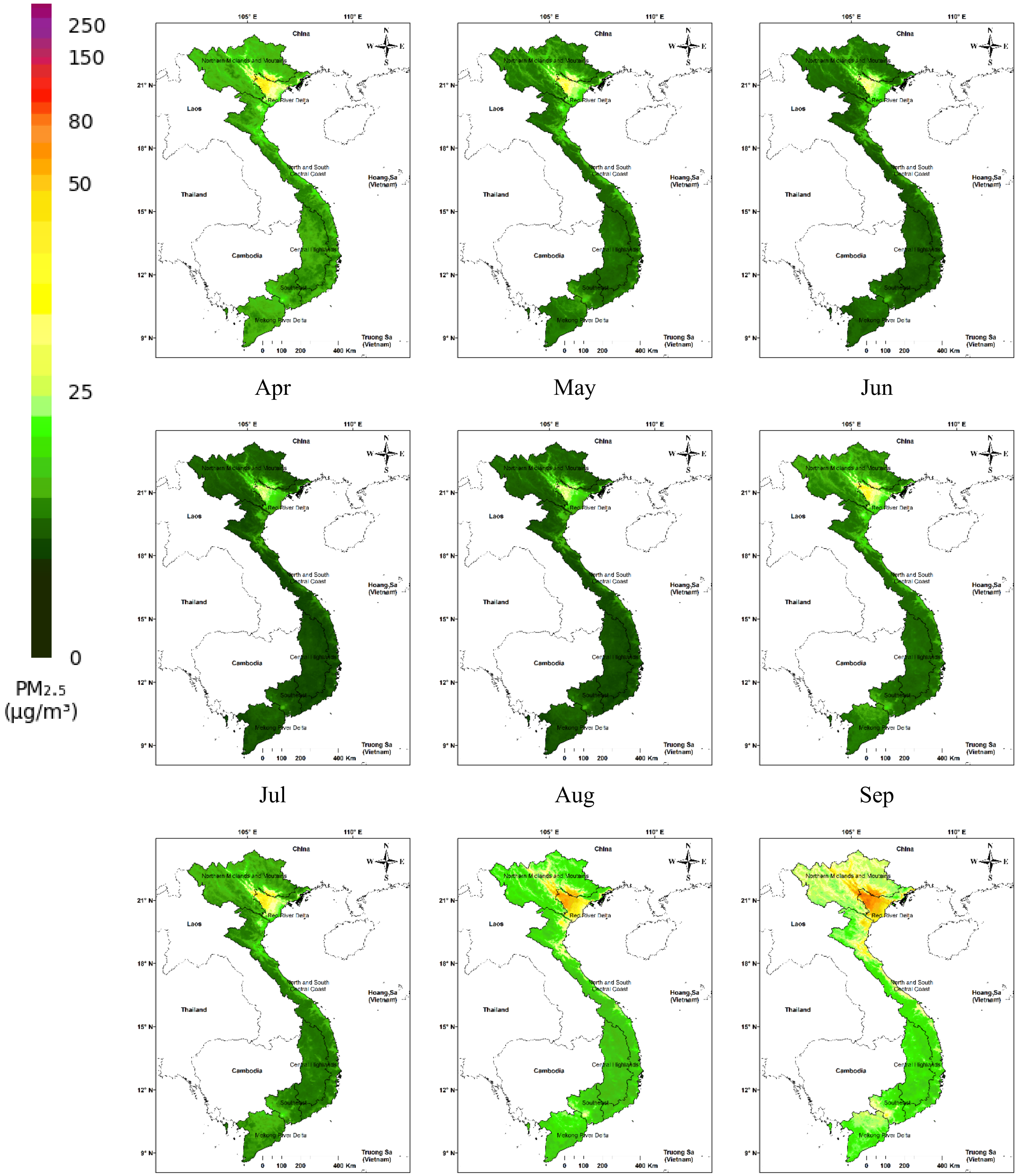


Figure 4.3. Variation of average $PM_{2.5}$ concentration in Vietnam by month during 2019, 2020, 2021

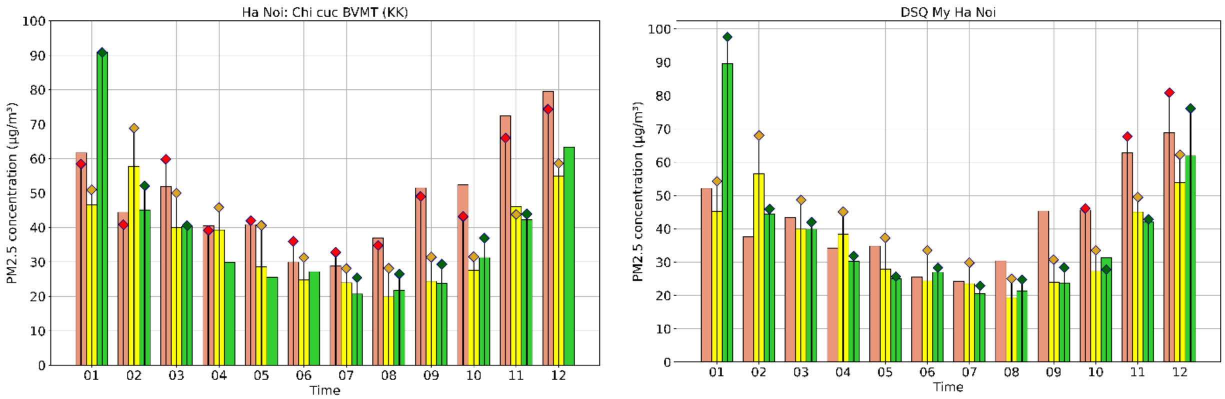


Jan Feb Mar

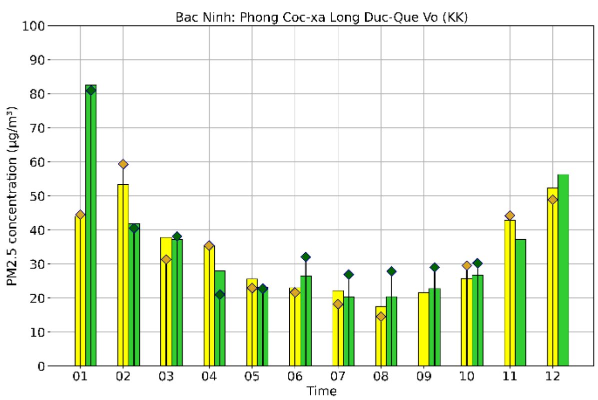

250
150
80
50
25
0
PM2.5 (µg/m³)
China
Laos
Thailand
Cambodia
Northern Midlands and Moutains
Red River Delta
North and South Central Coast
Central Highlands
Southeast
Mekong River Delta
Hoang Sa (Vietnam)
Truong Sa (Vietnam)
105° E
110° E
21° N
18° N
15° N
12° N
9° N
0 100 200 400 Km
Apr
May
Jun
Jul
Aug
Sep

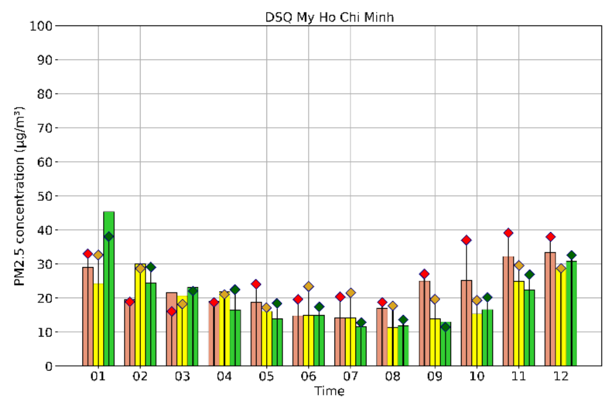
Oct Nov Dec

Figure 4.4. Distribution of $PM_{2.5}$ concentration on average monthly in 3 years 2019-2021



Hanoi





Bac Ninh Ho Chi Minh

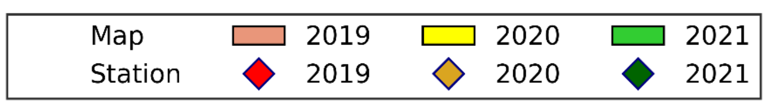


Figure 4.5. Evaluation of monthly map estimates at several stations for the period 2019-2021.

4.2.3. Annual average map

Similar to the monthly average map assessment, the global product [42] and our annual mean map [36] was compared with the corresponding ground observations in Vietnam for the period 2019-2021 (Table 4.7). Overall, the study's annual maps performed better than the global maps. In the year 2019-2020, our maps had R^2 ranging from 0.5 - 0.9, RMSE error from 4.32 - 8.79 µg/m³, MRE from 10.98 - 30.76 %. While the global map for

2019-2020 had an R^2 quality ranging from 0.49 – 0.75, RMSE from 4.72 – 9.42 µg/m³, MRE from 14.21 – 37.19%. In 2021, the global map has not been published, our map has an R^2 of 0.4, RMSE of 9.78 µg/m³, MRE of 29.15%

Table 4.7. Compare $PM_{2.5}$ maps with monitoring stations and compare with global maps

	Study's maps					**Global maps**				
Year	**N**	R^2	**Pearson r**	**RMSE**	**MRE**	**N**	R^2	**Pearson r**	**RMSE**	**MRE**
2019	5	0.9	0.95	4.32	10.98	5	0.75	0.87	4.72	14.21
2020	29	0.5	0.71	8.79	30.76	29	0.49	0.70	9.42	37.19
2021	37	0.4	0.63	9.78	29.15	-	-	-	-	-

The annual average $PM_{2.5}$ dust concentration maps for 2019, 2020, and 2021 were compared with the annual average $PM_{2.5}$ measurements at 47 monitoring stations throughout Vietnam, and the results are presented in Figure 4.6. The overall evaluation results include a Pearson r coefficient of 0.67, R^2 coefficient of 0.46, RMSE of 9.1 µg/m³, and MRE of 28.5%. In 2019, there were a limited number of ground monitoring stations (Hanoi, Quang Ninh, Da Nang, Ho Chi Minh), and the average annual map had good quality when compared to these stations, with an R^2 of 0.9, Pearson r of 0.95, RMSE of 4.32 µg/m³, and MRE of 10.98%. In 2020, the number of monitoring stations increased significantly, and the average annual map assessment with 29 stations had a Pearson r coefficient of 0.71, R^2 coefficient of 0.5, RMSE of 8.79 µg/m³, and MRE of 30.76%. The annual mean map tended to estimate lower than those observed at stations in Hanoi and Bac Ninh, but higher than those in Quang Ninh and Hai Duong. Large deviations between the estimated map and the stations appeared in some stations in Bac Ninh and Hai Duong. In 2021, the average annual map assessment with 37 stations had a Pearson r coefficient of 0.63, R^2 coefficient of 0.4, RMSE of 9.78 µg/m³, and MRE of 29.15%. Similar to 2020, the average map in 2021 tended to estimate lower than those observed at stations in Hanoi and Bac Ninh, but higher than those in Quang Ninh. Some stations in Cao Bang, Lang Son, and Thai Nguyen have large deviations between the estimated map and the monitoring station, and the reason for their high $PM_{2.5}$ observations needs to be further investigated. Some stations in Quang Ninh, Hanoi, Bac Ninh, and the central and southern provinces had low mean annual $PM_{2.5}$ observations and low variance between the map and station estimates.

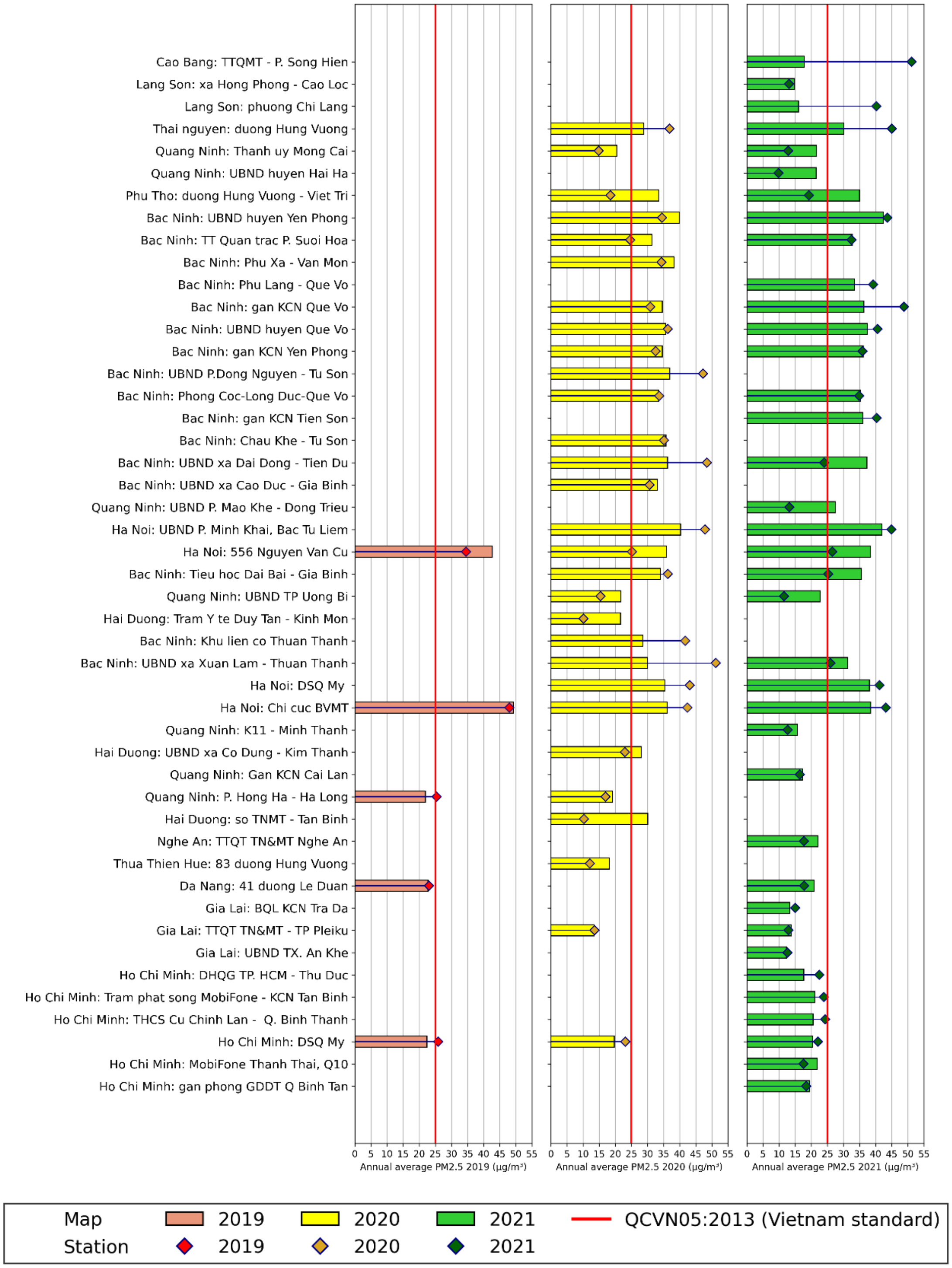


Figure 4.6. Comparison between the annual average $PM_{2.5}$ map and measurements at ground monitoring stations in the period of 2019-2021

On a national scale during the period of 2019-2021, areas with high $PM_{2.5}$ concentrations were mainly concentrated in the Red River Delta. In this region, in addition to the influence of emission sources, air quality was also partly influenced by meteorological conditions, especially during cold winters, resulting in high concentrations at certain times [44]. The central and southern regions had lower values, with the eastern coastal area in the central region having a higher concentration of $PM_{2.5}$ than the western hilly area. TP. Ho Chi Minh City in the southern region had a higher concentration of $PM_{2.5}$ than the entire region (Figure 4.8). In 2019, the average annual $PM_{2.5}$ concentration in Vietnam ranged from 11.44 μg/m³ to 50.77 μg/m³, with an average of 15.9 μg/m³. In 2020, $PM_{2.5}$ concentrations ranged from 10.26 μg/m³ to 44.07 μg/m³, averaging 14.17 μg/m³. In 2021, $PM_{2.5}$ concentrations ranged from 10.2 μg/m³ to 46.3 μg/m³, averaging 14.57 μg/m³. The average $PM_{2.5}$ concentration in Vietnam in 2020 and 2021 tended to decrease compared to 2019, and in 2021, it tended to increase slightly compared to 2020 (Figure 4.7)

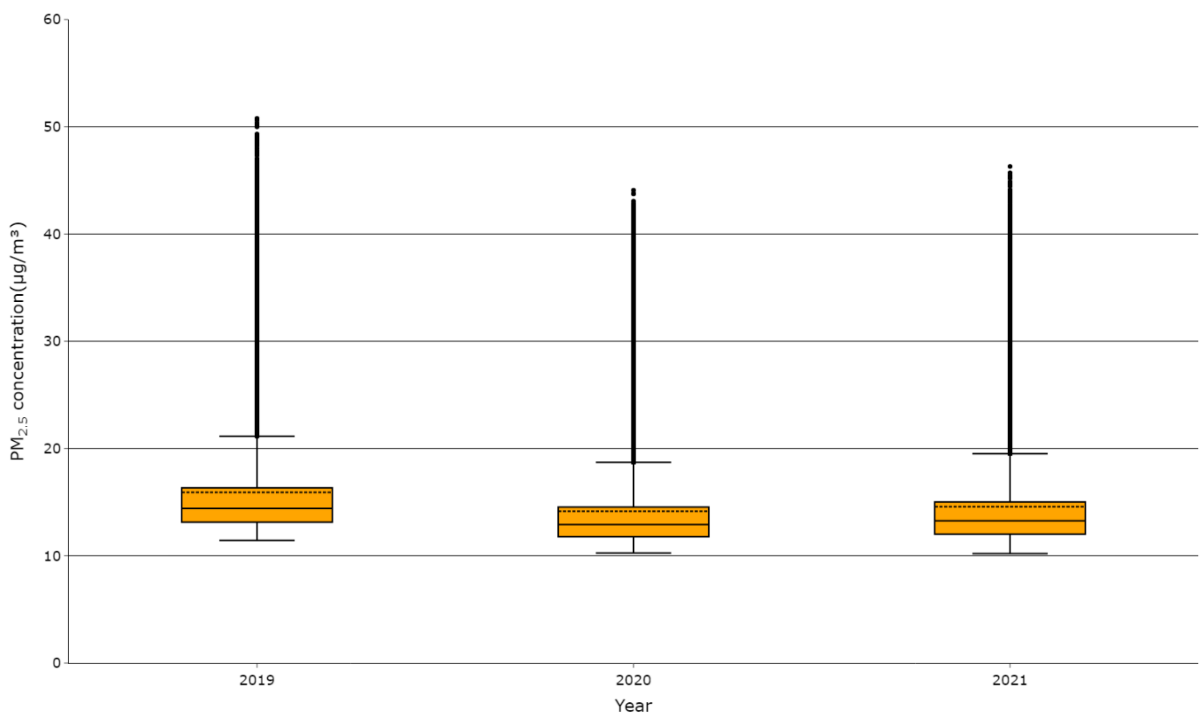
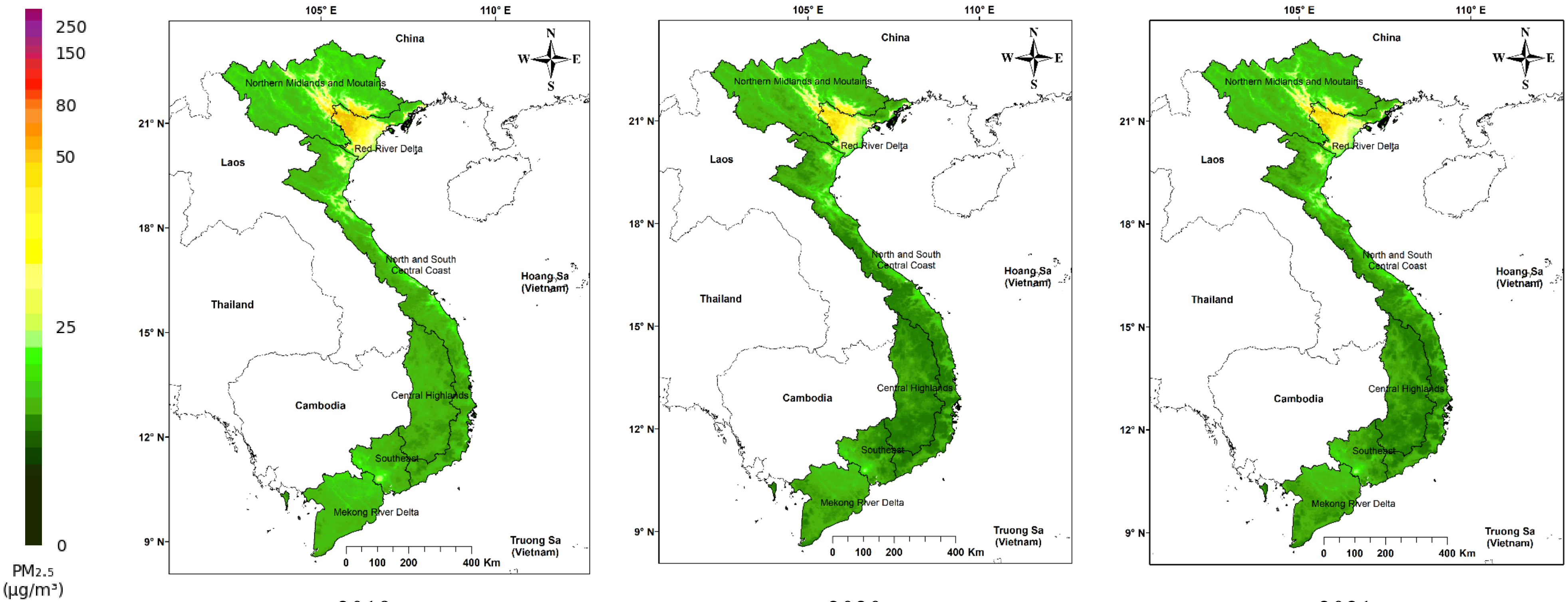


Figure 4.7. Statistics of annual average $PM_{2.5}$ in Vietnam during the period of 2019-2021



2019 2020 2021

Figure 4.8. Annual average maps of $PM_{2.5}$ dust concentration in 2019 - 2021 in Vietnam

CONCLUSION

In this report, daily $PM_{2.5}$ maps were estimated using two MEM models (one main model and one auxiliary model) developed on a 10-year dataset (2012-2021) including $PM_{2.5}$ measurements from monitoring stations, satellite AOD data, meteorological data, and land use maps. Pre-processing and quality enhancement techniques have been proposed and applied to ensure data quality before modeling. The main model had evaluation results with a Pearson correlation coefficient of 0.83, R-squared coefficient of 0.68, RMSE of 14.14 µg/m³, and MRE of 40.99% on a dataset of 10,614 samples. The ancillary model had evaluation results with a Pearson correlation coefficient of 0.81, R-squared coefficient of 0.65, RMSE of 14.8 µg/m³, and MRE of 48.58% on a larger dataset of 34,208 samples. The model represents an improvement over previous efforts by Thanh et al. in Vietnam [5], and is comparable to the results of similar studies in China, but inferior to those reported in studies in the United States. The error of the model depends on several factors, such as the study area, year of study, quality of station data, and number of stations.

The daily $PM_{2.5}$ map at 3x3 km spatial resolution for Vietnam was generated from the model for the period from 2019 to 2021. Results of daily $PM_{2.5}$ map assessment with ground measurements in this period showed Pearson r reaching 0.8, coefficient of determination R^2 reaching 0.64, RMSE reaching 15.44 µg/m³, MRE reaching 51.75%. The monthly aggregated and annual average maps from 2019-2021 were of better quality than the global $PM_{2.5}$ product when compared with the monitoring station in each year. The $PM_{2.5}$ concentration map has shown the spatial distribution and seasonal variation of $PM_{2.5}$ concentrations in Vietnam during the period of 2019-2021 which was not possible before due to the lack of important stations measurements and modeling results.

The map data is intended solely for the purpose of delivering the LASER PULSE project product to end users. The data has a spatial resolution of 3x3 km, covering the entire area of Vietnam. This means that every 3x3 km region in Vietnam has an estimated $PM_{2.5}$ concentration value, which is convenient for analyzing $PM_{2.5}$ variations spatially (aggregating and comparing by commune, district, province, economic region, and country). The data is detailed on a daily basis and continuous over the period from 2019 to 2021, allowing for convenient analysis of $PM_{2.5}$ variations/trends over time (daily, weekly, monthly, quarterly, seasonal, yearly). The spatial and temporal analysis of $PM_{2.5}$ concentration information can be applied to support decision-making and used in other related studies. The $PM_{2.5}$ datasets were used to assess the impact of air pollution on health [7][8] and environment [9] in Vietnam. As for policy development purposes, the $PM_{2.5}$

datasets of Hanoi and Ho Chi Minh City, Vinh Yen, Da Nang, Soc Trang, Ca Mau, Quang Tri, and Hai Phong City were used to assist with the policy briefs or recommendations for local government agencies. For instances, clean air action plans for Ho Chi Minh city and Vinh Yen city; policy brief and report on child-centered risks and access to community services for Soc Trang, Ca Mau, Quang Tri; vehicle emission inventory development for Da Nang; air quality management plan for 2022-2025 for Hai Phong. In this project, the $PM_{2.5}$ map was used to develop and publish the report on the "Status of $PM_{2.5}$ and Health Impacts in Vietnam in 2021" [10], providing air pollution information to the aforementioned users. The map was also used to create educational and communication videos about air pollution. Additionally, the mapping method was applied in the WebGIS system to display near-real-time air quality and provide information to the community [11].

REFERENCES

[1] WHO, "7 million premature deaths annually linked to air pollution," 2014. https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution (accessed Oct. 02, 2022).

[2] M. J. Wolf, J. W. Emerson, D. C. Esty, A. de Sherbinin, and Z. A. Wendling, "Environmental Performance Index 2022," 2022. [Online]. Available: https://epi.yale.edu/downloads/epi2022technicalappendix.pdf.

[3] IQAir/AirVisual, "World Air Quality Report 2021," 2022.

[4] VNU-UET, Live&Learn, and USAID, "State of PM2.5 in Vietnam during 2019-2020 based on multi-source data.," 2021. [Online]. Available: https://drive.google.com/file/d/1NkGCpXIWlcfP8qoG4h_YFuvARtCSop7-/view?usp=sharing.

[5] T. N. T. Nguyen *et al.*, "Particulate matter concentration mapping from MODIS satellite data: A Vietnamese case study," *Environ. Res. Lett.*, vol. 10, no. 9, Sep. 2015, doi: 10.1088/1748-9326/10/9/095016.

[6] T. X. Ngo, H. V. Pham, H. D. T. Phan, A. T. N. Nguyen, H. T. To, and T. T. N. Nguyen, "A daily and complete PM2.5 dataset derived from space observations for Vietnam from 2012 to 2020," *Sci. Total Environ.*, vol. 857, no. April 2022, p. 159537, 2023, doi: 10.1016/j.scitotenv.2022.159537.

[7] N. Nhung *et al.*, "Mortality burden due to exposure to outdoor fine particulate matter in Hanoi, Vietnam: Health Impact Assessment," *Int. J. Public Health*, 2022.

[8] N. T. T. Nhung *et al.*, "Mortality benefits of reduction fine particulate matter in Vietnam, 2019," *Front. Public Heal.*, vol. 10, 2022, doi: 10.3389/fpubh.2022.1056370.

[9] N. Tuyet, C. T. Tran, T. B. Ly, and T. T. N. Nguyen, "Developing a provincial environmental performance index for the environmental health of Vietnam," *Environ. Dev. Sustain.*, no. 0123456789, 2022, doi: 10.1007/s10668-021-02072-x.

[10] T. T. N. Nguyen *et al.*, "The Status of PM2.5 and Its Impacts on Public Health in Vietnam 2021," 2022. [Online]. Available: https://drive.google.com/open?id=1aIaZTQiOSsVDTeIKcc5apVbButHfivse&usp=drive_fs.

[11] T. N. N. Thanh, X. H. Tung, H. N. Tuan, V. P. Ha, X. N. Truong, and D. T. P. Hieu, "System to monitor daily average PM2.5 concentration in Vietnam," 2023. https://popgis.vnu.edu.vn/ (accessed Feb. 03, 2023).

[12] Vietnam Environment Administration, "Report on the State of the National

Environment for the period 2016-2020," 2021.

[13] A. J. Cohen *et al.*, "Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015," *Lancet*, vol. 389, no. 10082, pp. 1907–1918, 2017, doi: 10.1016/S0140-6736(17)30505-6.

[14] J. Wang and S. A. Christopher, "Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies," *Geophys. Res. Lett.*, 2003, doi: 10.1029/2003GL018174.

[15] T. Nhat *et al.*, "Current Status of PM 2 . 5 Pollution and its Mitigation in Vietnam," *Glob. Environ. Res.*, no. June, pp. 65–72, 2018.

[16] GSO Vietnam, "Tình hình kinh tế - xã hội 2022," 2022. https://www.gso.gov.vn/tin-tuc-thong-ke/2022/12/thong-cao-bao-chi-ve-tinh-hinh-kinh-te-xa-hoi-quy-iv-va-nam-2022/ (accessed Dec. 10, 2022).

[17] "MODIS," 2022. https://modis.gsfc.nasa.gov/about/index.php (accessed Jan. 26, 2022).

[18] "MOD04_3K," 2022. https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD04_3K (accessed Jan. 26, 2022).

[19] "NOAA NPP VIIRS," 2022. https://ncc.nesdis.noaa.gov/VIIRS/ (accessed Jan. 26, 2022).

[20] "Weather Research and Forecasting Model," 2022. https://www.mmm.ucar.edu/weather-research-and-forecasting-model (accessed Jan. 26, 2022).

[21] "MOD13Q1," 2022. https://lpdaac.usgs.gov/products/mod13q1v006/ (accessed Jan. 26, 2022).

[22] J. E. Vargas-Munoz, S. Srivastava, D. Tuia, and A. X. Falcao, "OpenStreetMap: Challenges and Opportunities in Machine Learning and Remote Sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 1, pp. 184–199, 2021, doi: 10.1109/MGRS.2020.2994107.

[23] ASTER Global V3, "LP DAAC - ASTGTM," 2019. https://lpdaac.usgs.gov/products/astgtmv003/ (accessed Oct. 02, 2022).

[24] H. Wu *et al.*, "Probabilistic Automatic Outlier Detection for Surface Air Quality Measurements from the China National Environmental Monitoring Network," *Adv. Atmos. Sci.*, vol. 35, no. 12, pp. 1522–1532, 2018, doi: 10.1007/s00376-018-8067-9.

[25] NCEP UPP, "NCEP Unified Post Processing System (UPP)," 2022.

https://dtcenter.org/community-code/unified-post-processor-upp (accessed Jan. 26, 2022).

[26] ArcMap, "ArcMap Resources for ArcGIS Desktop," 2022. https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources (accessed Apr. 26, 2022).

[27] P. Van Ha, N. X. Truong, D. Laffly, A. Jourdan, and N. T. Nhat Thanh, "Evaluation of maximum likelihood estimation and regression methods for fusion of multiple satellite aerosol optical depth data over Vietnam," 2019, doi: 10.1109/KSE.2019.8919417.

[28] R. Beelen *et al.*, "Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project," *Atmos. Environ.*, vol. 72, no. 2, pp. 10–23, 2013, doi: 10.1016/j.atmosenv.2013.02.037.

[29] C. T. Cowie *et al.*, "Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO2 for a birth cohort in Sydney, Australia," *Environ. Res.*, 2019, doi: 10.1016/j.envres.2019.03.068.

[30] A. van Donkelaar *et al.*, "Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors," *Environ. Sci. Technol.*, vol. 50, no. 7, pp. 3762–3772, Apr. 2016, doi: 10.1021/ACS.EST.5B05833.

[31] I. Kloog, F. Nordio, B. A. Coull, and J. Schwartz, "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.*, 2012, doi: 10.1021/es302673e.

[32] I. Kloog *et al.*, "A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data," *Atmos. Environ.*, 2014, doi: 10.1016/j.atmosenv.2014.07.014.

[33] H. J. Lee, R. B. Chatfield, and A. W. Strawa, "Enhancing the Applicability of Satellite Remote Sensing for PM2.5 Estimation Using MODIS Deep Blue AOD and Land Use Regression in California, United States," *Environ. Sci. Technol.*, vol. 50, no. 12, pp. 6546–6555, 2016, doi: 10.1021/acs.est.6b01438.

[34] Y. Zheng, Q. Zhang, Y. Liu, G. Geng, and K. He, "Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements," *Atmos. Environ.*, 2016, doi: 10.1016/j.atmosenv.2015.06.046.

[35] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating Ground-Level PM2.5 in

China Using Satellite Remote Sensing," *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7436–7444, Jul. 2014, doi: 10.1021/ES5009399.

[36] W. You, Z. Zang, L. Zhang, Y. Li, and W. Wang, "Estimating national-scale ground-level PM25 concentration in China using geographically weighted regression based on MODIS and MISR AOD," *Environ. Sci. Pollut. Res.*, vol. 23, no. 9, pp. 8327–8338, 2016, doi: 10.1007/s11356-015-6027-9.

[37] X. Yang, D. Xiao, H. Bai, J. Tang, and W. Wang, "Spatiotemporal Distributions of PM2.5 Concentrations in the Beijing–Tianjin–Hebei Region From 2013 to 2020," *Front. Environ. Sci.*, vol. 10, no. March, 2022, doi: 10.3389/fenvs.2022.842237.

[38] S. L. K. Unnithan and L. Gnanappazham, "Spatiotemporal mixed effects modeling for the estimation of PM2.5 from MODIS AOD over the Indian subcontinent," *GIScience Remote Sens.*, vol. 57, no. 2, pp. 159–173, Feb. 2020, doi: 10.1080/15481603.2020.1712101.

[39] Z. Ma, X. Hu, L. Huang, J. Bi, and Y. Liu, "Estimating ground-level PM2.5 in china using satellite remote sensing," *Environ. Sci. Technol.*, vol. 48, no. 13, pp. 7436–7444, 2014, doi: 10.1021/es5009399.

[40] Y. Zheng, Q. Zhang, Y. Liu, G. Geng, and K. He, "Estimating ground-level PM2.5 concentrations over three megalopolises in China using satellite-derived aerosol optical depth measurements," *Atmos. Environ.*, vol. 124, pp. 232–242, Jan. 2016, doi: 10.1016/J.ATMOSENV.2015.06.046.

[41] B. Peng-in, P. Sanitluea, P. Monjatturat, P. Boonkerd, and A. Phosri, "Estimating ground-level PM2.5 over Bangkok Metropolitan Region in Thailand using aerosol optical depth retrieved by MODIS," *Air Qual. Atmos. Heal.*, no. 0123456789, 2022, doi: 10.1007/s11869-022-01238-4.

[42] A. Van Donkelaar *et al.*, "Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty," *Environ. Sci. Technol.*, vol. 55, no. 22, pp. 15287–15300, Nov. 2021, doi: 10.1021/ACS.EST.1C05309.

[43] P. D. Hien, V. T. Bac, and N. T. H. Thinh, "PMF receptor modelling of fine and coarse PM10 in air masses governing monsoon conditions in Hanoi, northern Vietnam," *Atmos. Environ.*, vol. 38, no. 2, pp. 189–201, 2004, doi: 10.1016/j.atmosenv.2003.09.064.

[44] P. D. Hien, V. T. Bac, H. C. Tham, D. D. Nhan, and L. D. Vinh, "Influence of meteorological conditions on PM2.5 and PM2.5-10 concentrations during the monsoon season in Hanoi, Vietnam," *Atmos. Environ.*, vol. 36, no. 21, pp. 3473–3484, 2002, doi: 10.1016/S1352-2310(02)00295-9.