3-D AIR POLLUTION ESTIMATION USING A HYBRID SPATIAL MODEL: A CASE STUDY OF ZHUNAN-MIAOLI AREA, TAIWAN

C.W. Hsu[1], Y. R. Chern[1], J. J. Su[1], C. Wijaya[2], Y. C. Chen[3], S. C. Lung[4], T. C. Hsiao[5], T.A. Teo[6], I.L. Shih[7], C. D. Wu [8*]

[1] Dept. of Geomatics, National Cheng Kung University, 701 Tainan, Taiwan - (p68111509, p66114022, 10901017)@gs.ncku.edu.tw
[2] Agricultural Engineering Research Center, 320 Taoyuan, Taiwan - rbhung@aerc.org.tw
[3] National Institute of Environmental Health Sciences, National Health Research Institutes, 350 Miaoli, Taiwan - yucheng@nhri.edu.tw
[4] Research Center for Environmental Changes, Academia Sinica, 115 Taipei, Taiwan – Department of Atmospheric Sciences, National Taiwan University, 106 Taipei, Taiwan - sclung@as.ntu.edu.tw
[5] Graduate Institute of Environmental Engineering, National Taiwan University, 106 Taipei, Taiwan – Research Center for Environmental Changes, Academia Sinica, 115 Taipei, Taiwan – tchsiao@ntu.edu.tw
[6] Dept. of Civil Engineering, National Yang Ming Chiao Tung University, 300 Hsinchu, Taiwan – tateo@nycu.edu.tw
[7] National Center for High-Performance Computing, 300 Hsinchu, Taiwan – 1703064@narlabs.org.tw
[8] Dept. of Geomatics, National Cheng Kung University, 701 Tainan, Taiwan – National Institute of Environmental Health Sciences, National Health Research Institutes, 350 Miaoli, Taiwan – Innovation and Development Center of Sustainable Agriculture, National Chung Hsing University, 402 Taichung, Taiwan – chidawu@mail.ncku.edu.tw

KEY WORDS: Air pollution, Fine Particulate Matter ($PM_{2.5}$), Hybrid spatial model, 3-D distribution, Unmanned Aerial Vehicle (UAV).

ABSTRACT:

The rapid global urbanization has transformed cityscapes, giving rise to iconic skyscrapers that define modern cities. However, alongside this urban evolution, a pressing concern arises – the air quality within these towering urban environments. Fine particulate matter, known as $PM_{2.5}$, poses a grave threat to human health and the environment. These tiny particles, measuring 2.5 micrometers or less, can penetrate deep into the human respiratory system, posing severe health risks. Due to the limitations of traditional land-use regression models in estimating the variation of air pollution with altitude, this study employs a novel hybrid spatial model to assess the three-dimensional distribution of $PM_{2.5}$ in the atmosphere. We employ a comprehensive methodology, integrating diverse datasets and advanced modelling techniques, to uncover significant findings. Our analysis reveals the non-uniform nature of $PM_{2.5}$ distribution, both horizontally and vertically. Variable selection identifies key factors influencing $PM_{2.5}$ levels, including Broadleaf Forest, Carbon Monoxide (CO), and Height. Our ensemble model demonstrates robust performance, with Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) exhibiting superior predictive capabilities. This study provides valuable insights into the complex interplay of environmental factors affecting $PM_{2.5}$ concentrations in high-rise urban environments, emphasizing the need for targeted air quality management strategies considering both horizontal and vertical variations.

1. INTRODUCTION

The rapid global urbanization has dramatically changed the appearance of cities worldwide. One striking symbol of this transformation is the proliferation of tall skyscrapers that have become iconic features of modern cities. However, amidst this urban evolution, a significant concern emerges – the air quality within these towering urban environments.

Fine particulate matter, often referred to as $PM_{2.5}$, stands out as one of the most serious threats to both human health and the natural environment. These tiny particles, measuring 2.5 micrometers or smaller in diameter, can penetrate deep into the human respiratory system, posing severe and potentially life-threatening health risks to exposed individuals.

Adding to the complexity is the fact that $PM_{2.5}$ concentrations do not follow a uniform distribution pattern. Unlike some other pollutants, $PM_{2.5}$ levels vary not only horizontally across geographical areas but also vertically within different layers of the atmosphere. This three-dimensional aspect of $PM_{2.5}$ distribution adds complexity to the assessment of air quality.
Traditionally, environmental scientists and policymakers have relied on two-dimensional models, specifically Land-Use Regression (LUR) models, to analyze and address air pollution (Wu et al., 2017; Wong et al., 2021; Chen et al., 2022). While these models have been valuable in many ways, they do have limitations. These limitations become particularly evident in areas with dense clusters of tall buildings, where traditional models struggle to capture the intricate relationships among factors influencing air quality. To define the variation of air pollution in vertical scale, some scholars have lately recognized the vertical variability of air pollution, leading to the development of various derivative methodologies for estimation (Xu et al., 2022; Eeftens et al., 2018; Cichowicz et al., 2021; Liu et al., 2020). However, most of these methodologies rely on interpolation or direct visualization of air pollution changes in three-dimensional space through ratio relationships, without assessing the relationship between land use and vertical heterogeneity of air pollution.

In cities like Zhunan-Miaoli in Taiwan, characterized by their towering skyscrapers, the task of estimating air quality becomes even more crucial. Understanding the intricate web of 3-D air quality variations in such urban settings becomes a top priority. This is precisely where our research comes into play.

* Corresponding author

Our study aims to leverage a hybrid spatial model to unravel the three-dimensional complexity of $PM_{2.5}$ variations in Zhunan, Taiwan by conducting a three-dimensional land use regression. By delving into the vertical dimension of air quality, our goal is to emphasize the critical importance of conducting comprehensive assessments that go beyond traditional boundaries.

2. METHODOLOGY

2.1 Study Area

Our study concentrated on the Zhunan Town and Toufen City areas, both situated in proximity to the Hsinchu Science Park Zhunan Campus, renowned for its technological and research significance. Adjacent to this, the Toufen Industrial Zone, with its potential for environmental pollution, also fell within our study scope.

Figure 1 shows a visual representation of our study region and the sampling locations. In this illustrative diagram, various land use categories are represented by distinct colors: green denotes forested regions, yellow signifies industrial and commercial land utilization, gray corresponds to transportation infrastructure, and blue represents bodies of water.

Figure 1. Research area and sampling points

2.2 Data Source

This study employed the AS-LUNG-P system, developed by Taiwan's Academia Sinica, to measure real-time $PM_{2.5}$ concentrations. We mounted this equipment atop a hexacopter drone, enabling us to capture 3-D spatial air samples of $PM_{2.5}$.

As for the sampling route, we divided the sampling process into three different heights: 30 meters, 60 meters, and 100 meters, at each sampling site. Additionally, for each height, the drones completed two circles with radius of 100 meters and 200 meters. In terms of sampling site selection, we chose a total of 12 locations. These sites were categorized into four main types of land use: forest, commercial, traffic, and water.

Sampling was conducted during two different periods: the first round of sampling took place from January 12th, 23 to January 17th, 23 during the winter season, while the second round of sampling occurred from July 13th, 23 to July 20th, 23 in the summer. In this study, the primary dataset for analysis and modelling was derived from the first round of sampling. The data obtained from the second round of sampling was utilized for external validation purposes.

Various datasets were collected for model development as independent variables, including EPA Air Quality Monitoring Database, meteorological observation database, the Land Use Survey from National Land Surveying and Mapping Center, Ministry of the Interior (NLSC), road network digital map, multisource remote sensing image data and three-dimensional building from NLSC.

Figure 2. hexacopter and AS-LUNG-P system

2.3 Method

Our research began with the creation of a comprehensive database called the Concentration and Exposure Database. This database brought together the datasets we mentioned earlier, merging them into a single repository of important information. During this phase, we used ArcGIS Pro version 3.1 to perform complex calculations. Specifically, we calculated the proportion of each variable within the sampling point's area. This detailed process was followed by building the database itself using Python scripting, ensuring efficiency and accuracy in handling data.

The next important step was to understand the significance of the variables we had collected. To do this, we used a concept called SHapley Additive exPlanations (SHAP), which helps us understand how much each variable contributes to our final predictions. These SHAP values were then ranked to identify the most influential variables in our model.

Once we determined the importance of the variables, we proceeded with model development. We used various machine learning algorithms like Extreme Gradient Boosting Regression (XGBR), Gradient Boosting Regression (GBR), Light Gradient Boosting Machine Regression (LGBMR), CatBoost Regression (CBR), and Random Forest Regression (RFR) to build models that could effectively connect variables to $PM_{2.5}$ levels. We applied these algorithms step by step, selecting the best ones based on the previously calculated SHAP values. Finally, we chose the top-performing algorithms to work together in an ensemble learning approach to enhance our predictive capabilities.

The validation process was rigorous and included both internal and external validation methods. We used a well-known technique called ten-fold cross-validation to assess the model's performance. This method involves dividing the data into ten parts, training the model on nine of them, and testing it on the remaining part. We repeated this process ten times to get a comprehensive evaluation of the model's accuracy and reliability. We then conducted stratified validation by dividing our database into different groups based on factors like weekdays, height, or median concentration levels. These approaches helped us understand how well the model would perform with new data and under various conditions, making it more applicable and reliable.

To create a visual representation of $PM_{2.5}$ distribution, we used ArcGIS Pro's 3-D mapping tool. We created 3-D points for the study area and converted raster data into point attributes using the "XY Table to Point" function. These points were then used in the ensemble learning process to calculate predicted values for each point. To visualize the results in 3D, we used the Empirical Bayesian Kriging 3D functionality. For rendering graphics into a cubic format, we employed the "GA Layer 3D To NetCDF" feature to make the results visually informative.

The flowchart is presented in Figure 3, which outlines the step-by-step process of our study, from building the database to developing and validating the model, ultimately leading to a better understanding of $PM_{2.5}$ dynamics in the study area.

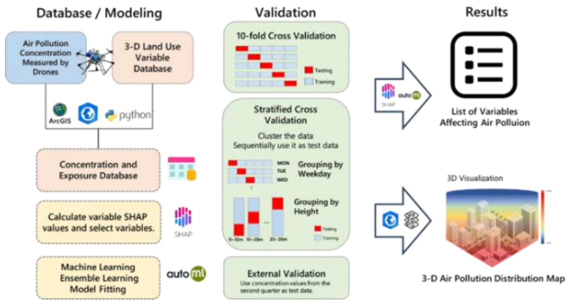

Figure 3. Flowchart

3. RESULT

3.1 Descriptive Statistics

Table 1 illustrates the descriptive statistics of our measurements, provides valuable insights into the distribution of $PM_{2.5}$ concentrations across our sampling points. Notably, it is evident that sampling points located in proximity to bodies of water consistently exhibit significantly lower $PM_{2.5}$ concentrations. Conversely, sampling points situated within areas designated for traffic or commercial use tend to demonstrate notably higher concentrations of $PM_{2.5}$. This preliminary statistical result underscores the potential impact of environment
al factors associated with land use on $PM_{2.5}$ levels within our study area, emphasizing the intricate nature of the environmental dynamics contributing to variations in air quality.

Date	Site	Heights	Buffer	$PM_{2.5}$ (μg/m³)
				Mean ± std
2023/1/13	**風璽**	Vertical only	Vertical only	31.72 ± 3.04
2023/1/13	**竹南海濱濕地**	120、80、60	200、100	33.27 ± 12.8
2023/1/13	**竹南海濱自然公園**	120、90、60	200、100	30.7 ± 14.24
2023/1/16	**竹南綜合運動公園**	100、60、30	trapezoid	28.28 ± 5.57
2023/1/16	**蘆竹社區運動中心**	100、80、60	200、100	37.31 ± 6.26
2023/1/17	**東興橋人工濕地**	100、60、30	200、100	21.41 ± 4.39
2023/1/17	**東興河濱公園**	100、60、40	200、100	19.72 ± 4.34
2023/2/8	**國家衛生研究院**	100、80、60	200、100	23.39 ± 6.67
2023/2/8	**頂埔國小**	120、100、80	200、100	30 ± 13.58
2023/2/9	**新南國小**	100、60、30	200、100	40.69 ± 34.65
2023/2/9	**頭份市活動中心**	Vertical only	Vertical only	31.61 ± 3.9

Table 1. Descriptive statistics of measurement

3.2 Vertical Distribution of Pollutants

In our investigation of the vertical distribution of pollutants across sites categorized by four distinct land uses, our results indicate that $PM_{2.5}$ concentration does not consistently decrease with higher altitudes. Surprisingly, in specific areas, we have observed sudden elevational increases in $PM_{2.5}$ concentration levels, as illustrated in Figure 4.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-301-2024 | © Author(s) 2024. CC BY 4.0 License.

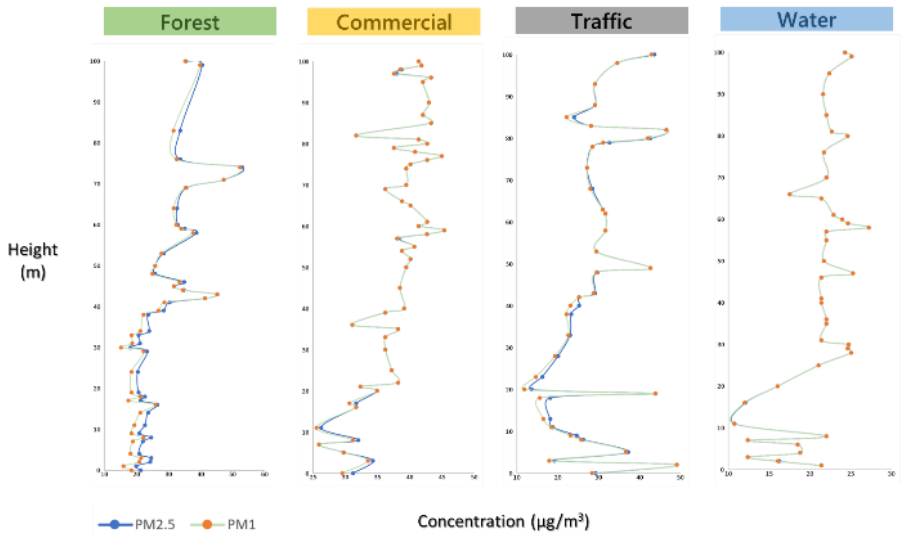

Figure 4. Vertical distribution of pollutants in each type of land-use

3.3 Variable Selection

As for the variable selection, we present the SHAP values of Gradient Boost, where color signifies high and low values, with red indicating high and blue representing low values. Along the horizontal axis, we depict the relationship between the variables and $PM_{2.5}$ concentrations.

Our analysis reveals that the top three crucial variables influencing $PM_{2.5}$ levels are Broadleaf Forest, CO (Carbon Monoxide), and Height. Moreover, we have included variables such as manufacturing industry building volume and closest distance from buildings in our assessment, shedding light on the substantial impact of urban structures on air pollutants.
Visual representation of these findings was shown in Figure 5.

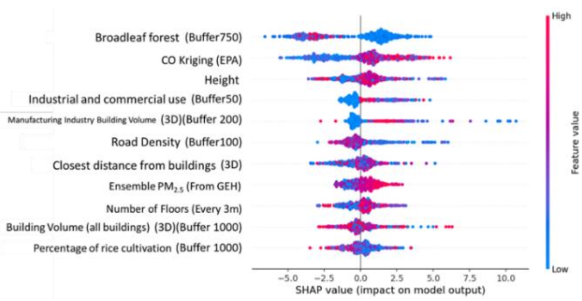

Figure 5. Variable selection of GBR

3.4 Model Performance

Regarding Model Performance, our evaluation indicates that all algorithms exhibited commendable performance, boasting high R^2 and adjusted R^2 values. This suggests their proficiency in elucidating data variability and emphasizing their suitability for our analysis.

In our pursuit of preventing overfitting, we diligently compared training and testing R^2 values. Notably, Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) emerged as the algorithms with the lowest overfitting rates, bolstering their credibility in modelling $PM_{2.5}$ concentrations.

To further enhance the predictive power of our model, we employed an ensemble approach, which proved to be superior to individual algorithms across various metrics. Consequently, we have chosen the ensemble model for our subsequent analysis.
Detailed reference to our model performance metrics was shown in Table 2.

Algorithm	**Model**	**Validation**	**R^2**	**Adjusted R^2**	**RMSE**	**MSE**	**MAE**	**Overfitting Rate (Training–Testing) (Training–CV)**
Ensemble Model	**GBR**	Training	0.98	0.98	1.28	1.63	0.52	0.00
	+	Testing	0.77	0.77	4.58	20.95	2.08	0.21
	RFR	10-Fold CV	0.91	0.91	2.77	7.68	0.86	0.07

Table 2. Model performance of ensemble model

3.5 Visualization

In this study, 3-D visualization of air pollution concentrations within the research area was conducted using ArcGIS Pro. Four distinct land-use types were selected for visualization. Figure 6 shows the result of one site from the land-use of Forest, represented by Zhunan Seaside Nature Park.

Based on this visualization, it can be observed that in the near-surface atmospheric layer, air pollution concentrations are significantly higher in the eastern part, closer to the Miaoli County waste incineration plant, compared to the western part. Conversely, in the higher atmospheric layers, air pollution concentrations are higher than at other altitudes, indicating that emission sources impact air quality at different elevations. Notably, in the higher atmospheric layers, air pollution concentrations are higher in the west compared to the east, suggesting that the influence of air pollutants varies with altitude due to daily wind patterns, leading to dispersion in different directions.

These visualizations provide a clear depiction of the spatial distribution of air pollution concentrations across different land-use types within the study area.

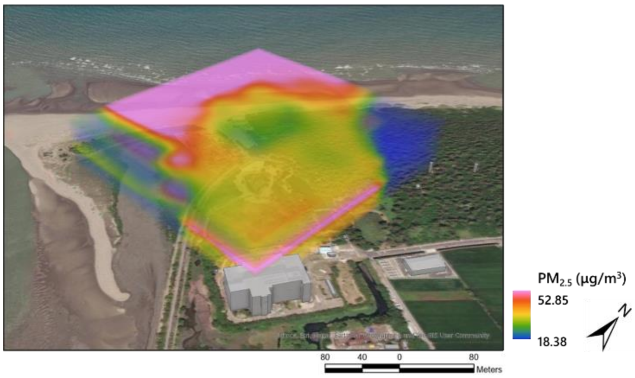

Figure 6. Visualization of air pollution concentrations in Zhunan Seaside Nature Park

4. CONCLUSION

In conclusion, our study delved into the intricate dynamics of $PM_{2.5}$ distribution in urban environments, particularly in areas characterized by a high density of towering structures. Through a comprehensive methodology that integrated diverse datasets and advanced modelling techniques, we uncovered several noteworthy findings.

Our analysis highlighted the non-uniform nature of $PM_{2.5}$ distribution, both horizontally and vertically. Notably, proximity to bodies of water appeared to significantly influence lower $PM_{2.5}$ concentrations, while areas designated for traffic or commercial use exhibited elevated concentrations. This underscores the pivotal role of land use in shaping air quality.

Moreover, our investigation of the vertical distribution of pollutants unveiled unexpected patterns, with certain locations experiencing abrupt increases in $PM_{2.5}$ concentrations at specific elevations. These findings emphasize the necessity of three-dimensional air quality assessments in urban environments characterized by towering structures.

Furthermore, our variable selection process identified the key factors influencing $PM_{2.5}$ levels, including Broadleaf forest, Carbon Monoxide (CO), and Height, along with the impact of urban infrastructure.

In terms of model performance, our ensemble approach demonstrated superior predictive power, with Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) exhibiting the lowest overfitting rates.

In summary, our study provides valuable insights into the complex interplay of environmental factors affecting $PM_{2.5}$ concentrations in high-rise urban environments, paving the way for more targeted and effective air quality management strategies.

5. DISCUSSION

Our research has uncovered compelling insights into the distribution of $PM_{2.5}$ in urban environments characterized by towering structures. The non-uniform distribution of $PM_{2.5}$, both horizontally and vertically, emphasizes the need for comprehensive assessment approaches.

The influence of land use on $PM_{2.5}$ levels underscores the significance of urban planning and environmental management. Proximity to bodies of water appeared to have a mitigating effect on $PM_{2.5}$ concentrations, highlighting the potential role of green spaces and natural features in urban air quality improvement.

The unexpected vertical variations in $PM_{2.5}$ concentrations challenge traditional air quality assessment methods and call for more sophisticated models that consider three-dimensional dynamics. This finding is particularly relevant for cities like Zhunan-Miaoli, Taiwan, with a dense cluster of high-rise structures.

Variable selection revealed that factors such as Broadleaf forest, Carbon Monoxide (CO), and Height play pivotal roles in shaping $PM_{2.5}$ levels. Additionally, the influence of urban infrastructure, including Manufacturing Industry Building Volume and Closest distance from buildings, further underscores the impact of human activities on air quality.

Our ensemble model exhibited robust performance, with Gradient Boosting Regression (GBR) and Random Forest Regression (RFR) demonstrating superior predictive capabilities and lower overfitting rates.

In conclusion, our study contributes valuable insights into the intricate dynamics of $PM_{2.5}$ distribution in high-rise urban environments, highlighting the need for targeted urban planning and air quality management strategies that consider both horizontal and vertical variations.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor C. D. Wu for their invaluable guidance and support throughout this research. I also extend my appreciation to my fellow colleagues in the laboratory for their technical collaboration and assistance. Additionally, I want to thank my co-authors, C. Wijaya, Y. C. Chen, S. C. Lung, T. C. Hsiao, T. A. Teo, and I. L. Shih, for their guidance and contributions to this work.

REFERENCES

Wu, C.D., Chen, Y.C., Pan, W.C., Zeng, Y.T., Chen, M.J., Guo, Y.L., Lung, S.C.C., 2017. Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM2.5 spatial-temporal variability. Environmental Pollution. doi.org/10.1016/j.envpol.2017.01.074.

Wong, P.Y., Su, H.J., Lee, H.Y., Chen, Y.C., Hsiao, Y.P., Huang, J.W., Teo, T.A., Wu, C.D., Spengler, J.D., 2021. Using land-use machine learning models to estimate daily NO2 concentration variations in Taiwan. Journal of Cleaner Production. doi.org/10.1016/j.jclepro.2021.128411.

Chen, W., Zhang, F., Luo, S., Lu, T., Zheng, J., He, L., 2022. Three-Dimensional Landscape Pattern Characteristics of Land Function Zones and Their Influence on PM2.5 Based on LUR Model in the Central Urban Area of Nanchang City, China. Int. J. Environ. Res. Public Health. doi.org/10.3390/ijerph191811696.

Xu, X., Qin, N., Zhao, W., Tian, Q., Si, Q., Wu, W., Iskander, N., Yang, Z., Zhang, Y., Duan, X., 2022. A three-dimensional LUR framework for PM2.5 exposure assessment based on mobile unmanned aerial vehicle monitoring. Environmental Pollution. doi.org/10.1016/j.envpol.2022.118997.

Eeftens, M., Odabasi, D., Flückiger, B., Davey, M., Ineichen, A., Feigenwinter, C., Tsai, M.Y., 2018. Modelling the vertical gradient of nitrogen dioxide in an urban area. Science of The Total Environment. doi.org/10.1016/j.scitotenv.2018.09.039.

Cichowicz, R., Dobrzański, M., 2021. 3D Spatial Analysis of Particulate Matter (PM10, PM2.5 and PM1.0) and Gaseous Pollutants (H2S, SO2 and VOC) in Urban Areas Surrounding a Large Heat and Power Plant. Energies. doi.org/10.3390/en14144070.

Liu, Y., Nie, J., Li, X., Ahmed, S.H., Lim, W.Y.B., Miao, C., 2020. Federated Learning in the Sky: Aerial-Ground Air Quality Sensing Framework with UAV Swarms. IEEE Internet of Things Journal. doi.org/10.48550/arXiv.2007.12004.

https://doi.org/10.5194/isprs-archives-XLVIII-4-W8-2023-301-2024