TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 24, No. 6, December 2026, pp. 2103~2113
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v24i6.27751

Spatial estimation of PM2.5 concentrations using KNN based on data from an ESP32 mobile monitoring station

Retno Tri Wahyuni, Jesika Vebiola Manurung, Heri Subagiyo, M. Budi Satria Yonda
Study Program of Electronic System Engineering Technology, Department of Industrial Technology, Politeknik Caltex Riau, Riau, Indonesia

Article Info

Article history:

Received Dec 23, 2025
Revised Sep 25, 2026
Accepted Oct 11, 2026

Keywords:

K-nearest neighbors
Mobile station
Particulate matter 2.5
Spatial estimation
Web-based visualization

ABSTRACT

This study presents an integrated framework for spatial estimation of particulate matter 2.5 (PM2.5) concentrations using k-nearest neighbors (KNN) regression and data collected from a low-cost mobile air-quality monitoring system. The framework was implemented in Pekanbaru city, Indonesia, an urban area with limited air-quality monitoring coverage. The system integrates real-time data acquisition using portable sensors, server-based KNN spatial estimation, and interactive visualization through the AeroNav web platform. The KNN model applies the Haversine distance metric with a 1 km search radius, while the number of neighbors is dynamically determined based on the availability of nearby mobile observations. Field testing was conducted at four reference centroids during a single-day measurement campaign. Valid KNN estimates were obtained at three of the four reference centroids, with absolute errors ranging from 0.86 to 2.53 μg/m³ and a mean absolute error (MAE) of 1.65 μg/m³ compared with reference measurements. The estimated PM2.5 concentrations showed spatial variation among locations with different dominant land-use characteristics; however, the limited spatial and temporal coverage does not allow these differences to be attributed to specific land-use characteristics. The results demonstrate the feasibility of integrating low-cost mobile sensing, neighborhood-based spatial estimation, and web-based visualization into an environmental monitoring framework for data-scarce urban environments.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Retno Tri Wahyuni
Study Program of Electronic System Engineering Technology, Department of Industrial Technology
Politeknik Caltex Riau
Umbansari St. No. 1, Umbansari, Rumbai, Pekanbaru, Riau 28265, Indonesia
Email: retnotri@pcr.ac.id

1. INTRODUCTION

Air quality remains a critical environmental and public health concern. Among various air pollutants, fine particulate matter 2.5 (PM2.5), particles with an aerodynamic diameter ≤ 2.5 μm, has received particular attention due to its severe health impacts. Long-term exposure to PM2.5 significantly increases the risk of respiratory and cardiovascular diseases, making it one of the most hazardous pollutants in urban environments [1], [2]. Therefore, reliable monitoring of PM2.5 concentrations is essential for developing effective mitigation strategies and supporting evidence-based policymaking.

Over the past two decades, air-quality monitoring systems have evolved considerably. Early networks were primarily static, relying on permanently installed sensors at fixed locations. For instance, the wireless sensor network air pollution monitoring system (WAMPS) utilized thousands of stationary nodes to

measure and visualize air-quality data using the air quality index (AQI) [3]. Although such systems provide continuous measurements with high temporal resolution, their implementation and maintenance costs can limit spatial coverage [4], particularly in resource-constrained monitoring environments. To overcome these limitations, mobile monitoring systems have emerged as a flexible and cost-effective alternative, enabling spatially detailed air-quality observations. For example, a bicycle-mounted mobile station equipped with PM2.5 sensors and global positioning system (GPS) modules enabled high-resolution mapping of air quality along urban routes [5]. Similarly, community-based initiatives such as the citizen-science program in Charlotte, United States, engaged volunteers carrying low-cost AirBeam sensors to collect real-time PM2.5 data, later corrected using machine-learning (ML) algorithms to enhance accuracy [6].

The development of low-cost microcontrollers such as the ESP32, combined with internet of things (IoT) technologies, has further expanded the potential of portable air-quality monitoring systems [7]-[9]. The ESP32 enables integration of PM2.5 sensors, GPS modules, and wireless connectivity to collect and transmit data in real time from multiple locations, thereby improving spatial coverage at minimal cost. Several studies have demonstrated that ESP32-based systems can yield reliable measurements when properly calibrated [10]. Similar implementations using low-cost instrumentation such as Arduino wireless fidelity (Wi-Fi) (ESP8266) have been employed to monitor PM2.5 and meteorological parameters wirelessly [11]. These hardware innovations are often complemented by open-source Android-based applications, built using frameworks such as react native and connected to backend databases [12]. These applications not only enable real-time visualization of PM2.5 and meteorological data via accessible mobile interfaces, but also play a crucial role in data transmission to the database. Data from the mobile monitoring stations, including PM2.5, temperature, humidity, and other environmental parameters, are transmitted in real time to a centralized backend database, where they are processed, stored, and made available for further analysis and visualization.

Despite these advances, mobile sensing systems generate large and heterogeneous datasets that require intelligent processing to extract meaningful insights. The dynamic movement of mobile stations results in dense yet spatially irregular measurements. Without appropriate analysis, such data may become fragmented or redundant, diminishing its interpretive value. Therefore, analytical methods capable of structuring these heterogeneous observations are essential to transform raw sensor readings into coherent spatial information.

In response, ML approaches have been increasingly applied to model and interpret mobile-sensing data. A range of algorithms has been explored for air-quality estimation. Song and Han [13] developed Deep-MAPS, a deep neural network framework integrating mobile and fixed sensors in Beijing to produce PM2.5 maps at 1×1 km resolution. Munir *et al.* [14] applied multiple linear regression and gradient boosting to analyze PM2.5 data from vehicular mobile stations and estimate inhaled exposure levels in urban environments. Hassani *et al.* [15] utilized a random forest model to calibrate data from citizen-operated mobile sensors in Norway, achieving substantial accuracy improvements. Li *et al.* [16] combined mobile and fixed monitoring networks using gradient boosting decision trees (GBDT) and convolutional neural networks (CNN), respectively, to produce high-resolution PM2.5 maps (≤ 500 m) while reducing monitoring costs.

Although these advanced models can provide strong predictive performance, their implementation may require relatively large datasets, multiple predictor variables, and greater computational resources. In such settings, simpler non-parametric methods may provide practical alternatives with lower modeling complexity. The k-nearest neighbors (KNN) algorithm offers a straightforward approach for estimating PM2.5 concentrations by exploiting the spatial proximity of nearby observations. Rather than relying on complex feature learning, KNN performs local regression based on surrounding observations, providing a relatively simple approach for spatially irregular datasets. Previous studies have explored neighborhood-based approaches for spatial air-quality estimation. For instance, Guo *et al.* [17] combined KNN with inverse distance weighting (KIDW) to improve PM2.5 interpolation in areas with sparse monitoring stations. Other studies have also demonstrated the potential of neighborhood information and spatiotemporal modeling for improving PM2.5 estimation and prediction [18], [19]. However, the integration of low-cost mobile sensing, neighborhood-based KNN spatial estimation, and interactive web-based visualization remains relatively underexplored in urban environments with limited monitoring infrastructure.

In this study, KNN is applied as a spatial regression method to estimate PM2.5 concentrations at predetermined grid centroids using nearby observations collected by the mobile monitoring station. Rather than generating a continuous spatial surface, each centroid serves as a target location for localized estimation based on the spatial proximity of available observations. This approach is integrated with ESP32-based mobile sensing and interactive web-based visualization through the AeroNav platform. The main contribution lies in the system-level integration of data acquisition, wireless transmission, KNN-based spatial estimation, and interactive visualization into a unified monitoring framework for data-scarce urban environments, rather than in the development of a new spatial estimation algorithm.

2. METHOD

This study employed a systematic approach to develop and evaluate an ESP32-based mobile air-quality monitoring system and to estimate PM2.5 concentrations using KNN regression. The methodological framework comprises mobile data acquisition, data transmission and storage in a centralized database, data preprocessing, KNN-based spatial estimation, and web-based visualization. As illustrated in Figure 1, the workflow begins with data collection using a mobile monitoring station that continuously records PM and environmental parameters. The collected data are transmitted to a centralized database and subsequently processed on the server for KNN-based spatial estimation at predefined target locations. The resulting PM2.5 estimates are then presented through the AeroNav web interface, enabling interactive monitoring and spatial visualization of air-quality conditions.

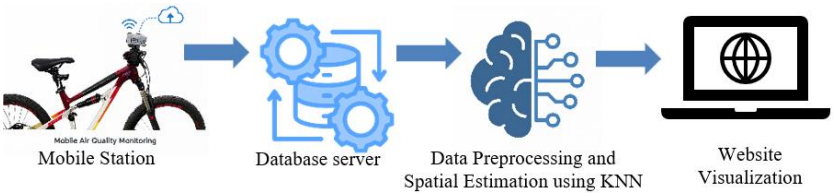


Figure 1. Overall methodological framework of the proposed system

2.1. Data acquisition process

The mobile monitoring station was designed as a compact and portable unit integrating multiple sensors and communication modules for real-time air-quality measurement. The system hardware comprised an ESP32 microcontroller as the central processing unit (CPU), a ZH03B dust sensor for measuring PM2.5 concentrations, a BME280 sensor for temperature, humidity, and atmospheric pressure, and a Ublox GPS Neo-6M module for geolocation tracking. Additionally, a BH1750 light sensor and an organic light-emitting diode (OLED) display were included to monitor ambient light intensity and display sensor readings locally. A rechargeable battery powered the system, and the ESP32's built-in Wi-Fi module handled data transmission to the server.

The ZH03B sensor used in this study is a low-cost optical particle sensor whose measurements may be affected by aerosol composition, relative humidity, and sensor drift. Prior to field deployment, the sensor was evaluated through a co-location comparison with an IQAir particle monitor to quantify its measurement uncertainty. Twelve paired PM2.5 measurements were collected under the same measurement conditions at 5-min intervals. Sensor performance was evaluated using the root mean square error (RMSE) between the mobile station and IQAir measurements.

Figure 2 shows the modular hardware configuration of the mobile monitoring station. Sensor data were first processed locally by the ESP32 before being transmitted to the web server and simultaneously stored in both onboard memory and the my structured query language (MySQL) cloud database to minimize the risk of data loss. For field deployment, the mobile station was securely mounted on a bicycle handlebar or carried on a motorcycle, depending on the measurement conditions. The mobile platform traveled along predefined urban routes at a controlled speed of 10–20 km h^{-1}, enabling repeated spatial sampling during movement.

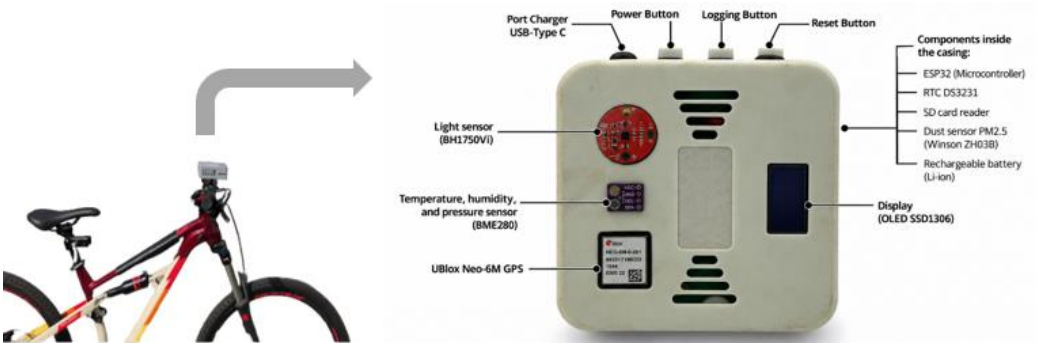


Figure 2. Design of mobile monitoring station and installation

Data collection was conducted within the area of Pekanbaru city, which was spatially divided into observation zones in the form of 5×5 km grids to ensure uniform spatial coverage and facilitate spatial analysis. This division was intended to systematically represent the spatial distribution of PM2.5 concentrations across the study area. Each grid contained a centroid designated as the target location for spatial estimation.

In fact, various strategies have been employed in determining spatial coverage by previous researchers. Several earlier studies used different grid resolutions, ranging from 500 meters [20] to 10 km [21], where finer grids allow for more detailed spatial analysis but require greater data density and computational resources. In this study, a 5×5 km grid resolution was selected as a practical compromise between spatial detail, field-data availability, and computational requirements.

Based on this division, a total of 27 observation grids were generated, as illustrated in Figure 3. From these 27 grids, four reference points, namely ID 5, 10, 11, and 15 were selected as field sampling locations. The selection of these points was solely intended for the implementation of the estimation model and the testing of the KNN-based mobile monitoring system, rather than representing specific environmental variations. Measurements collected around these points produced field observation data used in the modeling process and the evaluation of PM2.5 concentration estimates.

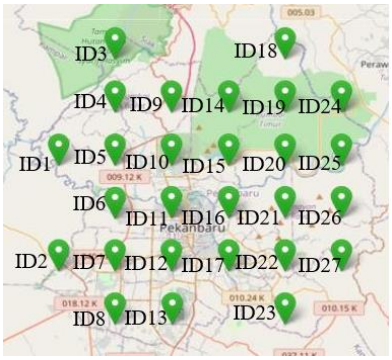


Figure 3. Observation grid

2.2. Data preprocessing

Before the modeling process was conducted, the system applied a data preprocessing mechanism to ensure the consistency and reliability of data received from the mobile monitoring stations. This process was executed on the web server after the data were transmitted from the devices via wireless connection. Each incoming record was checked for invalid formats, missing spatial or temporal attributes, and invalid sensor readings. Records failing these checks were excluded before spatial estimation.

The main database was structured to record each measurement in an organized manner, including the timestamp and geographic coordinates (latitude and longitude). Other measurement parameters stored in the database included PM2.5 and PM10 concentration, air temperature, relative humidity, atmospheric pressure, and light intensity (lux). The structure of the database table is shown in Table 1.

Table 1. The structure of the database

Column name	Data type	Description
autoIncrement	INT	Primary key: auto-increment identifier for each record
Datetime	TIMESTAMP(6)	Timestamp of when the measurement was taken
Latitude	DECIMAL(10,6)	Latitude coordinate of the measurement location
Longitude	DECIMAL(10,6)	Longitude coordinate of the measurement location
PM2.5	FLOAT	PM2.5 concentration in micrograms per cubic meter (µg/m³)
PM10	FLOAT	PM10 concentration in micrograms per cubic meter (µg/m³)
Temperature	FLOAT	Air temperature in celsius (°C)
Pressure	FLOAT	Atmospheric pressure in hectopascals (hPa)
Humidity	FLOAT	Relative humidity in percentage (%)
Lux	FLOAT	Light intensity in lux
Sign	VARCHAR(50)	Device status or signal indicator
Region	VARCHAR(100)	Administrative region or district
Location	VARCHAR(100)	Specific location name

2.3. KNN algorithm for spatial estimation of PM2.5

To estimate PM2.5 concentrations at unsampled locations, this study applied distance-weighted KNN regression. The number of neighbors (K) was determined dynamically based on observations within a specified search radius. The Haversine distance was used to calculate the geographic distance between observation and prediction locations. For two coordinates (ϕ_1, λ_1) and (ϕ_2, λ_2), the distance d is calculated using (1).

$$d = 2R\arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1)\cos(\varphi_2)\sin^2\left(\frac{\Delta\lambda}{2}\right)}\right) \quad (1)$$

where $R = 6371$ km is the Earth's radius, and φ and λ are the latitude and longitude in radians.

In the developed system, the search radius can be adjusted through the web interface. For the analysis presented in this study, a radius of 1 km was used to prioritize observations located close to each target centroid. Consequently, the value of K was not fixed but corresponded to the number of valid observations available within 1 km of each prediction location. The 1 km radius provides a localized neighborhood relative to the 5×5 km grid cells while limiting the contribution of more distant observations. Because the spatial distribution of mobile measurements was irregular, the number of available neighbors varied among target centroids according to local data availability. If no observations were available within the specified radius, no KNN estimate was generated for that target location.

For target locations with available neighboring observations, PM2.5 concentration was estimated using a distance-weighted average as expressed in (2).

$$\hat{y}(x) = \frac{\sum_{i=1}^{K} w_i y_i}{\sum_{i=1}^{K} w_i}, w_i = \frac{1}{d_i^p} \quad (2)$$

where $\hat{y}(x)$ is the estimated PM2.5 concentration at prediction location x, y_i is the observed PM2.5 concentration at the i-th neighboring point, d_i is the Haversine distance between prediction location x and its i-th neighbor, and K is the number of valid observations within the specified search radius. The distance-weighting exponent p was set to 2 in this study.

The weighting function $w_i = 1/d_i^p$ is adapted from the inverse distance weighting (IDW) method, which has been widely applied in spatial modeling of environmental variables such as temperature, rainfall, and PM [22]. This weighting function allows the influence of each neighboring point to decrease smoothly with distance, resulting in stable estimates that are free from extreme fluctuations. The distance-weighting exponent p was set to 2. This selection was informed by previous IDW-based interpolation research showing that $p = 2$ can provide suitable performance for relatively small datasets [23]. However, the optimal exponent may vary depending on the characteristics and spatial distribution of the input data. Model performance was evaluated by comparing the KNN-estimated PM2.5 concentrations with direct field measurements at reference points for which valid KNN estimates were available. For each valid reference point, the absolute error (AE) was calculated using (3).

$$E_a = | P_{pred} - P_{obs} | \quad (3)$$

where $\hat{y}_i$ represents the estimated PM2.5 concentration and y_i represents the observed PM2.5 concentration at the i-th valid reference point. The absolute error quantifies the magnitude of the difference between the estimated and observed concentrations at each reference location. In addition to the site-specific absolute errors, the mean absolute error (MAE) was calculated to summarize the overall estimation error across all valid reference comparisons, as expressed in (4).

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

where y_i represents the observed value, $\hat{y}_i$ represents the estimated value, and n denotes the number of valid reference points.

2.4. Web-based visualization

The web-based visualization component was developed as the final part of the monitoring system to display both measurement data and prediction results interactively. The application, named AeroNav, functions to present real-time air-quality data as well as spatial regression results of PM2.5 concentrations generated by the KNN algorithm. This web-based approach enables continuous environmental monitoring,

visualization, and user interaction through an online interface, which is essential for improving the accessibility of air-quality information [24]. The visualization system architecture consists of three main layers: (i) the database, (ii) the server-side processing, and (iii) the user interface (UI). Measurement data and regression outputs from the KNN algorithm are stored in a MySQL database. These data are then retrieved by hypertext preprocessor (PHP) scripts on the server side and transmitted to the client interface, which is built using hypertext markup language (HTML), cascading style sheets (CSS), and JavaScript.

The AeroNav interface comprises three main modules: (i) the dashboard, which displays system status and monitoring activities; (ii) the data monitoring module, which presents real-time variations of PM2.5, temperature, and humidity in graphical form; and (iii) the prediction map, which visualizes spatial regression results on an interactive Google Maps platform. The map module provides several parameter-control options, allowing users to adjust the neighbor search radius, which determines the spatial neighborhood and consequently the number of observations (K) included in the regression, and to select data sources either from the database or through comma-separated values (CSV) file upload. This integration of predictive modeling with a web-based mapping platform improves the visualization and interpretation of spatial air-quality data, enabling users to interactively explore pollutant distributions and trends across urban areas. Such interactive web-mapping systems have been shown to enhance data accessibility and support informed decision-making for researchers, policymakers, and the public [25].

3. RESULTS AND DISCUSSION

3.1. Data acquisition performance

Prior to field deployment, the PM2.5 sensor performance was evaluated through a co-location comparison with an IQAir particle monitor. Across 12 paired measurements, the mobile station recorded a mean PM2.5 concentration of 31.65 μg/m³, compared with 31.75 μg/m³ measured by the IQAir monitor, with an RMSE of 0.31 μg/m³. This relatively small RMSE indicates limited measurement deviation within the evaluated concentration range. However, because the comparison covered a limited concentration range and was not used to derive a formal calibration equation, the result should be interpreted as a quantitative evaluation of sensor measurement uncertainty rather than evidence of reference-equivalent performance.

Field measurements were subsequently conducted using the ESP32-based mobile monitoring station at four reference locations, namely ID5, ID10, ID11, and ID15, located near the centroids of selected 5×5 km grid cells in Pekanbaru city. The geographical coordinates and dominant surrounding land-use characteristics of the reference locations are presented in Table 2.

Table 2. Geographical information and dominant land-use characteristics of the reference locations

Reference ID	Location name	District	Latitude	Longitude	Dominant land-use characteristics
ID5	Palas	Rumbai Barat	0.578309	101.390139	Residential and plantation
ID10	Lembah Damai	Rumbai	0.578337	101.434986	Residential and secondary swamp forest
ID11	Padang Terubuk	Senapelan	0.533189	101.435013	Residential and peatland
ID15	Lembah Sari	Rumbai Timur	0.578366	101.479836	Plantation and residential

For each reference location, mobile measurements were collected along designated routes within a maximum radius of 5 km from the corresponding centroid, with a sampling interval of one minute. During field deployment, the mobile station continuously recorded georeferenced PM2.5 measurements together with environmental parameters, including temperature, humidity, pressure, and light intensity. The acquired geographic coordinates and PM2.5 measurements were subsequently used for spatial estimation. The Haversine distance was calculated between each observation and the corresponding target centroid to determine the spatial proximity of the observations used in the KNN regression.

The spatial distribution of the collected measurements around the four reference locations is presented in Figure 4. In the spatial maps, the blue points represent mobile PM2.5 measurement locations, while the green and red markers indicate the reference centroid and reference measurement location, respectively. The map panels are displayed at different spatial extents to clearly show the distribution of measurements at each location; therefore, the apparent distances between points in the maps should not be directly compared across locations. The accompanying scatter plots provide the corresponding quantitative distances between each mobile measurement and its target centroid and show the measured PM2.5 concentrations as a function of these distances. These plots provide the distance information used to characterize the spatial distribution of observations for the subsequent KNN estimation.

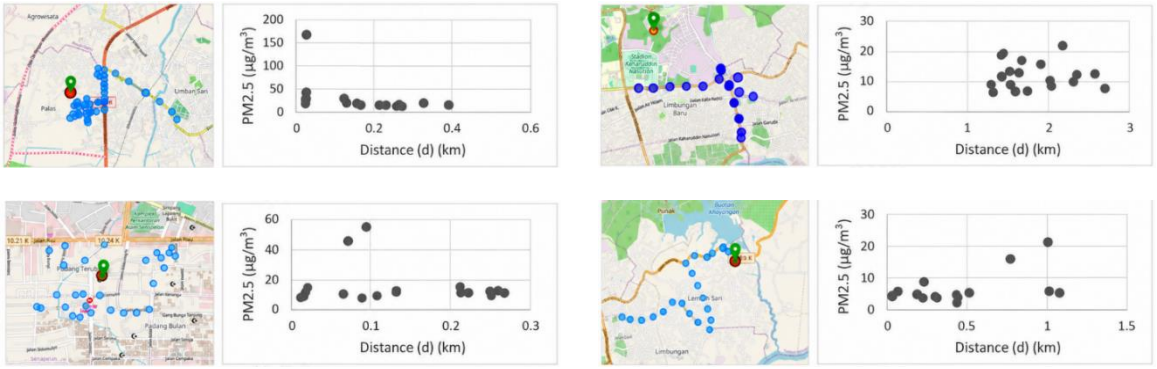


Figure 4. Spatial distribution of mobile PM2.5 measurements and their distances from the reference centroids

3.2. KNN model result and validation

The KNN regression used in this study determines the number of neighbors (K) dynamically based on the observations available within a specified search radius. A search radius of 1 km was applied, such that all valid measurement points within 1 km of each centroid were included in the estimation. Consequently, the value of K varied among centroids depending on the local availability of measurement data. The 1 km radius was selected to prioritize observations located close to each centroid while limiting the contribution of more distant measurements.

Data sampling with a larger radius up to a maximum of 5 km has the potential to provide a more comprehensive representation of conditions within each grid, considering that the grid size in this study was defined as 5×5 km. Based on the measurement data obtained and visualized in Figure 4, the number of observation points located within a 1 km radius from the centroid was found to be 19 points for ID5, 0 points for ID10, 22 points for ID11, and 18 points for ID15. These data were subsequently used in estimating the PM2.5 concentration at each centroid location using the KNN algorithm. For comparison purposes, direct measurements were also performed at each centroid using an IQAir particle monitor. The observed PM2.5 concentrations obtained at the designated reference locations during the field measurement period were used as reference values for evaluating the KNN estimates, as presented in Table 3.

Table 3. Comparison between observed and estimated PM2.5 concentrations

Reference ID	PM2.5 observed (µg/m³)	PM2.5 estimated (µg/m³)	$\|E_a\|$
5	30.06	28.51	1.55
10	18.48	-	-
11	13.32	15.85	2.53
15	4.02	4.88	0.86

The comparison between observed and estimated PM2.5 concentrations in Table 3 showed absolute errors ranging from 0.86 to 2.53 µg/m³ at the three reference points for which KNN estimates were successfully generated (ID5, ID11, and ID15). However, the validation was limited to four reference locations and a single-day measurement campaign. Although the selected locations represent different surrounding land-use characteristics, including residential, plantation, peatland, and secondary swamp forest areas, this limited spatial and temporal coverage is insufficient to establish the generalizability of the proposed framework across broader urban environments. Therefore, the reported errors should be interpreted as pilot-scale performance under the specific conditions evaluated. The absence of an estimate at ID10 further demonstrates the dependence of the KNN regression on local data availability within the specified search radius. Future studies should extend the validation through repeated measurements across multiple days and a larger number of reference locations representing diverse land-use conditions to evaluate the spatial and temporal robustness of the proposed approach.

To provide a summarized measure of overall estimation deviation, the MAE was calculated across the three valid reference comparisons. The resulting MAE was 1.65 µg/m³, representing the average magnitude of deviation between observed and estimated PM2.5 concentrations under the evaluated pilot-scale conditions. Because only three valid reference comparisons were available, the error values were evaluated descriptively rather than through inferential statistical analysis. The limited sample size was insufficient to establish a reliable error distribution or confidence interval; therefore, the reported absolute

errors and MAE should be interpreted as descriptive indicators of estimation performance under the specific conditions evaluated.

Although the main validation applied a fixed 1 km radius, the estimation behavior remains sensitive to neighborhood definition because the radius directly controls neighbor availability and spatial locality. The unequal distribution of neighboring observations across centroids demonstrates that estimation stability depends strongly on local data density. In particular, insufficient neighbors may prevent estimation, while excessive spatial coverage may reduce locality through smoothing effects across heterogeneous land-use zones.

Therefore, the 1 km radius was retained in this pilot-scale study to prioritize spatial locality while allowing the number of neighboring observations to vary according to local data availability. The estimated PM2.5 concentrations showed spatial variation across the four reference locations with different dominant land-use characteristics. At reference ID5 (Palas, Rumbai Barat), characterized by residential and plantation land use, the estimated PM2.5 concentration was 28.51 μg/m³, the highest among the estimated reference locations. At reference ID11 (Padang Terubuk, Senapelan), characterized by residential and peatland areas, the estimated PM2.5 concentration was 15.85 μg/m³. At reference ID10 (Lembah Damai, Rumbai), characterized by residential and secondary swamp forest, the observed PM2.5 concentration was 18.48 μg/m³; however, no KNN estimate was generated because no neighboring observations were available within the 1 km search radius. In contrast, reference ID15 (Lembah Sari, Rumbai Timur), characterized by plantation and residential land use, showed the lowest estimated PM2.5 concentration of 4.88 μg/m³. These results indicate spatial variation in PM2.5 concentrations among the reference locations; however, given the limited number of reference locations and the single-day measurement campaign, the observed differences cannot be directly attributed to specific land-use characteristics or emission sources. The observed spatial differences may reflect variations in local emission activities and surrounding environmental conditions. Residential areas may be influenced by local traffic, domestic activities, and other nearby combustion sources, whereas plantation and vegetated areas may experience different emission and dispersion conditions. However, because emission-source data and meteorological variables were not included in the present analysis, these potential mechanisms cannot be quantitatively distinguished from the effects of land-use characteristics.

The spatial variation observed in this pilot study is broadly comparable with the spatial patterns reported in [26], which applied a random forest-based spatial modeling framework using temporally adjusted portable sensor data, land-use parameters, and aerosol optical depth (AOD) across 5×5 km grids in Pekanbaru. Both approaches reveal that residential and peatland areas consistently exhibit higher PM2.5 concentrations, while vegetation-dominated regions such as plantations and secondary forests maintain lower particulate levels.

The current pilot-scale evaluation focused on the performance of the implemented KNN approach within the proposed mobile monitoring framework. A comparative evaluation against conventional spatial interpolation methods, such as IDW and Kriging, was not included in the present study. Therefore, the reported results should not be interpreted as demonstrating the superiority of KNN over alternative interpolation methods. Future studies should include comparative evaluations of multiple spatial estimation approaches using the same measurement dataset.

3.3. Web-based visualization

The AeroNav web application was developed to visualize both real-time measurement data and PM2.5 concentrations estimated using the KNN regression model. The Prediction Map module integrates the spatial estimation results with an interactive Google Maps interface, allowing users to perform on-demand estimation at unsampled grid centroids and adjust the search radius. The selected radius determines the spatial neighborhood and, consequently, the number of observations (K) included in the estimation. The results are displayed using ISPU-based colored markers and interactive pop-up windows containing the estimated PM2.5 concentration, air-quality category, number of neighboring observations, and distance to the nearest observation.

Figure 5 illustrates an example of the AeroNav prediction map at Lembah Sari. In the example shown, the system estimated a PM2.5 concentration of 13.84 μg/m³, corresponding to the "Good" air-quality category. Using a 1 km search radius, 39 neighboring observations were identified, with the nearest observation located 0.28 km from the target location. The interactive pop-up therefore provides not only the estimated concentration but also information about the neighboring observations used in the estimation. This visualization connects the mobile sensing and KNN estimation components of the proposed framework by enabling users to access and interpret localized PM2.5 estimation results through a web-based map.

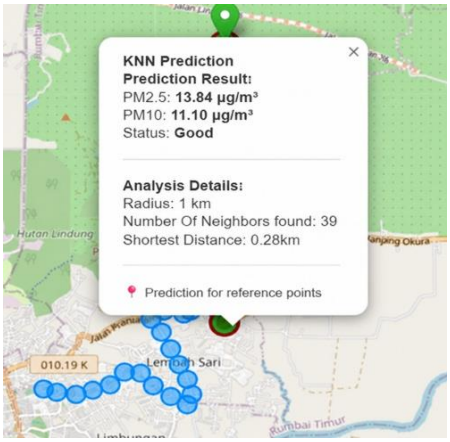


Figure 5. Example of an interactive pop-up window on the AeroNav prediction map

4. CONCLUSION

This study developed and demonstrated an integrated ESP32-based mobile air-quality monitoring system combined with KNN regression for spatial estimation of PM2.5 concentrations in Pekanbaru city, Indonesia. The proposed system integrates mobile data acquisition, server-side spatial estimation, and web-based visualization within a unified monitoring framework. Field implementation demonstrated the operational capability of the mobile station to collect and transmit particulate and environmental data during mobile measurements. Using a 1 km search radius and distance-weighted averaging, the KNN regression generated valid PM2.5 estimates at three of the four reference locations. The absolute errors at these locations ranged from 0.86 to 2.53 µg/m³, with a MAE of 1.65 µg/m³. No estimate was generated at one reference location because no neighboring observations were available within the specified search radius, highlighting the dependence of the method on local data availability. The estimated PM2.5 concentrations also showed spatial variation among locations with different dominant land-use characteristics; however, these differences should not be interpreted as evidence of a causal relationship between land use and PM2.5 concentrations.

This study represents a pilot-scale implementation conducted under limited spatial and temporal sampling conditions. Validation was based on four reference locations within a single-day measurement campaign, with valid KNN estimates available at only three locations. Consequently, the limited number of valid reference comparisons did not allow robust statistical characterization of the error distribution or confidence intervals, and the reported errors should be interpreted as descriptive indicators of performance under the evaluated conditions. In addition, the KNN approach was not directly compared with alternative spatial interpolation methods. Future studies should therefore include repeated multi-day measurements across a larger number of reference locations and diverse land-use conditions, extended sensor calibration and uncertainty evaluation, statistical characterization using a larger validation dataset, and comparative evaluation with conventional spatial interpolation methods such as IDW and Kriging.

FUNDING INFORMATION

This research was supported by the Internal Research Grant of Politeknik Caltex Riau 2025. The authors gratefully acknowledge the financial assistance, which enabled the completion of this study.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Retno Tri Wahyuni	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓
Jesika Vebiola Manurung		✓	✓	✓							✓			
Heri Subagiyo				✓				✓		✓				
M. Budi Satria Yonda	✓											✓		

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known financial, personal, or professional conflicts of interest that could have influenced the work reported in this paper.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] L. Zhang, J. P. Wilson, N. Zhao, W. Zhang, and Y. Wu, "The dynamics of cardiovascular and respiratory deaths attributed to long-term PM2.5 exposures in global megacities," *Science of the Total Environment*, vol. 842, Oct. 2022, doi: 10.1016/j.scitotenv.2022.156951.

[2] Y. Zhang *et al.*, "Estimating the mutually adjusted health effects of short- and long-term exposure to PM2.5 on respiratory mortality in a population-based study," *Atmospheric Pollution Research*, vol. 15, no. 5, May 2024, doi: 10.1016/j.apr.2024.102091.

[3] K. K. Khedo, R. Perseedoss, and A. Mungur, "A Wireless Sensor Network Air Pollution Monitoring System," *International Journal of Wireless & Mobile Networks*, vol. 2, no. 2, pp. 31–45, May 2010, doi: 10.5121/ijwmn.2010.2203.

[4] W. Y. Yi, K. M. Lo, T. Mak, K. S. Leung, Y. Leung, and M. L. Meng, "A survey of wireless sensor network based air pollution monitoring systems," *Sensors (Switzerland)*, vol. 15, no. 12, pp. 31392–31427, Dec. 2015, doi: 10.3390/s151229859.

[5] R. Hart, L. Liang, and P. Dong, "Monitoring, mapping, and modeling spatial–temporal patterns of PM2.5 for improved understanding of air pollution dynamics using portable sensing technologies," *Int. J. Environ. Res. Public Health*, vol. 17, no. 14, pp. 1–18, Jul. 2020, doi: 10.3390/ijerph17144914.

[6] M. D. Adams, F. Massey, K. Chastko, and C. Cupini, "Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction," *Atmospheric Environment*, vol. 230, 2020, doi: 10.1016/j.atmosenv.2020.117479.

[7] D. Hercog, T. Lerher, M. Truntič, and O. Težak, "Design and Implementation of ESP32-Based IoT Devices," *Sensors*, vol. 23, no. 15, Aug. 2023, doi: 10.3390/s23156739.

[8] S. Mahetaliya, D. Makwana, A. Pujara, and S. Hanumante, "IoT based Air Quality Index Monitoring using ESP32," *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 4, pp. 5186–5191, Apr. 2021.

[9] V. Hura and L. Monastyrskii, "IoT-based solution for detection of air quality using ESP32," *Artif. Intell.*, vol. 28, no. 3, pp. 86–93, Nov. 2023, doi: 10.15407/jai2023.03.086.

[10] C. Muñoz, J. Huircan, F. Jaramillo, and Á. Boso, "Calibration of Sensor Network for Outdoor Measurement of PM2.5 on High Wood-Heating Smoke in Temuco City," *Processes*, vol. 11, no. 8, Aug. 2023, doi: 10.3390/pr11082338.

[11] Q. Li and C. Hao, "Design of PM2.5 Detector Based on Internet of Things (IoT) Technology," *Proceedings of the 4th Workshop on Advanced Research and Technology in Industry (WARTIA 2018). Atlantis Press*, 2018. doi: 10.2991/wartia-18.2018.30.

[12] R. T. Wahyuni, B. Novriadi, and Z. Arifin, "Android Application Design for Monitoring Weather Parameter and PM 2.5," *2021 International Conference on Computer Science and Engineering (IC2SE)*, Padang, Indonesia, 2021, pp. 1-5, doi: 10.1109/IC2SE52832.2021.9792022.

[13] J. Song and K. Han, "Deep-MAPS: Machine Learning based Mobile Air Pollution Sensing," 2019, *arXiv*. doi: 10.48550/ARXIV.1904.12303.

[14] M. M. Munir, M. Adrian, C. Saputra, and P. Lestari, "Utilizing Low-cost Mobile Monitoring to Estimate the PM2.5 Inhaled Dose in Urban Environment," *Aerosol Air Qual. Res.*, vol. 22, no. 6, Jun. 2022, doi: 10.4209/aaqr.220079.

[15] A. Hassani, N. Castell, Å. K. Watne, and P. Schneider, "Citizen-operated mobile low-cost sensors for urban PM2.5 monitoring: field calibration, uncertainty estimation, and application," *Sustainable Cities and Society*, vol. 95, Aug. 2023, doi: 10.1016/j.scs.2023.104607.

[16] T. Li *et al.*, "Machine learning-guided integration of fixed and mobile sensors for high resolution urban PM2.5 mapping," *NPJ Clim. Atmos. Sci.*, vol. 8, no. 1, Mar. 2025, doi: 10.1038/s41612-025-00984-3.

[17] C. Guo, G. Liu, L. Lyu, and C.-H. Chen, "An Unsupervised PM2.5 Estimation Method With Different Spatio-Temporal Resolutions Based on KIDW-TCGRU," *IEEE Access*, vol. 8, pp. 190263–190276, 2020, doi: 10.1109/ACCESS.2020.3032420.

[18] L. Feng, Y. Li, Y. Wang, and Q. Du, "Estimating hourly and continuous ground-level PM2.5 concentrations using an ensemble learning algorithm: The ST-stacking model," *Atmospheric Environment*, vol. 223, Feb. 2020, doi: 10.1016/j.atmosenv.2019.117242.

[19] G. Chen, S. Chen, D. Li, and C. Chen, "A hybrid deep learning air pollution prediction approach based on neighborhood selection and spatio-temporal attention," *Sci. Rep.*, vol. 15, no. 1, Jan. 2025, doi: 10.1038/s41598-025-88086-1.

[20] J. Xiong, R. Yao, W. Wang, W. Yu, and B. Li, "A spatial-and-temporal-based method for rapid particle concentration estimations in an urban environment," *Journal of Cleaner Production*, vol. 256, pp. 1–7, May 2020, doi: 10.1016/j.jclepro.2020.120331.

[21] G. Geng *et al.*, "Tracking Air Pollution in China: Near Real-Time PM2.5 Retrievals from Multisource Data Fusion," *Environ. Sci. Technol.*, vol. 55, no. 17, pp. 12106–12115, Sep. 2021, doi: 10.1021/acs.est.1c01863.

[22] J. Li and A. D. Heap, "Spatial interpolation methods applied in the environmental sciences: A review," *Environmental Modelling and Software*, vol. 53, pp. 173–189, Mar. 2014, doi: 10.1016/j.envsoft.2013.12.008.

[23] U. Barudžija, J. Ivšinović, and T. Malvić, "Selection of the Value of the Power Distance Exponent for Mapping with the Inverse Distance Weighting Method—Application in Subsurface Porosity Mapping, Northern Croatia Neogene," *Geosciences (Switzerland)*, vol. 14, no. 6, Jun. 2024, doi: 10.3390/geosciences14060155.
[24] P. Chen, "Visualization of real-time monitoring datagraphic of urban environmental quality," *EURASIP J. Image Video Process.*, vol. 2019, no. 1, Feb. 2019, doi: 10.1186/s13640-019-0443-6.
[25] M. Ashouri, F. K. H. Phoa, C. H. Chen, and G. Shmueli, "An Interactive Clustering-Based Visualization Tool for Air Quality Data Analysis," *Aerosol Air Qual. Res.*, vol. 23, no. 12, Dec. 2023, doi: 10.4209/aaqr.230124.
[26] R. T. Wahyuni, D. Hanafi, M. Razali Tomari, and D. S. S. Sahid, "Spatial Estimation of PM2.5 Using Random Forest with Temporally Adjusted Portable Sensors, Land Use, and AOD in a Data-Scarce Urban Area," *International Journal of Intelligent Engineering and Systems*, vol. 18, no. 10, pp. 751–763, Nov. 2025, doi: 10.22266/ijies2025.1130.48.

BIOGRAPHIES OF AUTHORS



Retno Tri Wahyuni received a Bachelor's degree and a Master in Electrical Engineering from Sepuluh November Institute of Technology, Indonesia. She has also been a lecturer in the Electronics Systems Engineering Technology program at Politeknik Caltex Riau since 2006. As a lecturer, her activities include teaching, research, and community service. Her current research focuses on environmental topics, both in the design and development of embedded systems and in system modeling using various methods, including machine learning. To support her knowledge of environmental topics, she collaborates with relevant institutions such as the Meteorology and Geophysics Agency and also holds a competency certification in air pollution control. She can be contacted at email: retnotri@pcr.ac.id.



Jesika Vebiola Manurung is an undergraduate student in the Electronics Systems Engineering Technology program at Politeknik Caltex Riau, Indonesia, enrolled in 2021. As part of her final project, she is working on a lecturer-led research project related to PM2.5 estimation systems. Her work focuses on developing a structured database and a web-based platform for PM2.5 monitoring, which enables data storage, visualization, and accessibility to support environmental analysis and informed decision-making. Her academic interests include environmental monitoring systems and web-based information systems. She can be contacted at email: jesikavebiolamanurung@gmail.com.



Heri Subagiyo received a Bachelor's degree and a Master in Electrical Engineering from Bandung Institute of Technology, Indonesia in 1999 and 2005. He has also been a lecturer in the Electronics Systems Engineering Technology program at Politeknik Caltex Riau since 2001. As a lecturer, his activities include teaching, research, and community service. His research interests are in instrumentation, control, and intelligent systems for environmental, industrial, and educational application. He also holds a competency certification in air pollution control and LabVIEW associate developer. He can be contacted at email: heri@pcr.ac.id.



M. Budi Satria Yonda received a Bachelor's degree in Applied Electrical Engineering from Politeknik Caltex Riau, Indonesia. He has been working as an Educational Laboratory Assistant at Politeknik Caltex Riau since 2016, where he is involved in teaching, applied research, and product development. His work includes supporting laboratory instruction, developing prototype systems, and guiding student innovation projects. His current research interests focus on FPGA-based digital design, the internet of things (IoT), motor control and sensor integration in environmental monitoring systems. He can be contacted at email: budi@pcr.ac.id.