Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No.121
DOI: 10.11159/icsta25.121

Leveraging Spatial Interpolation for $PM_{2.5}$ Estimation: Constructing a Reference Framework for Calibration Models

Natthanidnan Sricharoen[1*], Patrinee Traisathit[2], Sukon Prasitwattanaseree[2], Pimwarat Srikummoon[2], Jeerasak Longmali[3], Titaporn Supasri[3]
[1]PhD Degree Program in Applied Statistics, Department of Statistics, Faculty of Science, Chiang Mai University, under the CMU Presidential Scholarship, Chiang Mai, Thailand
natthanidnan_s@cmu.ac.th
[2]Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, Thailand
patrinee.t@cmu.ac.th; sukon.pra@cmu.ac.th; pimwarat.n@cmu.ac.th
[3]National Astronomical Research Institute of Thailand, Chiang Mai, Thailand
jeerasak@narit.or.th; titaporn@narit.or.th

Abstract - Low-cost air quality sensors (LCSs) have become increasingly vital in augmenting traditional monitoring networks, particularly in regions constrained by limited reference-grade stations. Although their affordability and capacity for high-resolution measurements enable better insights into pollution patterns, calibrating pollution-monitoring data from LCS to ensure reliability remains a critical challenge. To addresses this, a two-phase framework was developed and applied to LCS-captured $PM_{2.5}$ data from Chiang Mai, Thailand, a region prone to elevated particulate matter from seasonal biomass burning. In the first phase, Inverse Distance Weighting generated localized $PM_{2.5}$ reference values in under-monitored areas, mitigating bias caused by distant reference stations. The IDW-based errors, with a root mean squared errors (RMSE) of 37.15 µg/m³, were marginally lower than direct reference-based RMSE (37.40 µg/m³). This minimal improvement underscores the limitations imposed by an insufficient number of reference monitors. In the second phase, four calibration models: Multiple Linear Regression, a Generalized Additive Model, Random Forest, and a Long Short-Term Memory (LSTM) neural network were evaluated to align the LCS readings with the interpolated reference. Meteorological and temporal factors were incorporated into each model to account for both seasonal trends and sensor-specific bias. Results showed that all calibrated outputs significantly better than the raw LCS data. Among them, the LSTM model consistently outperformed the other approaches by achieving coefficients of determination (R^2) as high as 0.96 and RMSE as low as 10 µg/m³. Overall, these enhancements were evident across multiple seasons and locations, highlight the pivotal role of robust machine learning approaches and integrated spatial techniques to enhance $PM_{2.5}$ monitoring, ultimately contributing to more effective pollution management and public health protection. In future research, we intend to explore alternative calibration strategies that do not rely on heavily on interpolation, particularly in resource-limited contexts and potentially integrating additional data sources such as satellite aerosol readings.

Keywords**:** spatial interpolation, calibration, low-cost sensor, air quality, bias correction

1. Introduction

Low-cost air quality sensors (LCSs) offer a promising approach to supplement reference-grade air quality monitoring. Recently, the adoption of LCSs has increased dramatically, driven by the need for monitoring pollution data with higher temporal and spatial resolution amid rising concerns about air quality [1-3]. Their affordability and ease of deployment enable the creation of denser pollution-monitoring networks that can provide more detailed insights into pollutant distribution and human exposure [4-6]. Moreover, local ambient pollution monitoring with LCSs could potentially increase the general awareness of the local citizenry towards pollution issues in their respective neighborhoods [7-8]. While LCSs can expand air-monitoring coverage and accessibility, calibrating and validating their readings against reference-grade instruments remains crucial to ensure accurate and reliable measurements [1-6]. This combination of LCS and conventional monitoring approaches could provide more comprehensive air quality assessments and informed decision-making than by using the latter exclusively.

In Thailand, filter-based inertial or gravimetric techniques along with a well-constructed size-selection inlet comprising automatic beta ray attenuation or a tapered element oscillating microbalance are employed as a conventional reference method for $PM_{2.5}$ measurements. These methods are evaluated and indirectly operated by the Pollution Control Department

(PCD) under the Ministry of Natural Resources and Environment [9]. However, the high costs associated with installing and managing $PM_{2.5}$ monitoring sensors coupled with their limited detection range pose significant challenges to calibrating LCS data; these are exacerbated in areas where reference-grade monitors have been sparsely deployed or located far from sensor sites. Moreover, the lack of nearby conventional monitoring stations can lead to spatial mismatches, representativeness errors, and biased calibration results [2]. In Chiang Mai province, Thailand, $PM_{2.5}$ poses a significant environmental and public health challenge, particularly during the dry season spanning March to April; seasonal biomass burning in the surrounding northern mountains contributes substantially to elevated levels of $PM_{2.5}$, resulting in deteriorated air quality, reduced visibility, and adverse health impacts [10]. Although Chiang Mai province encompasses a vast geographical area (approximately 20,000 km^2), only a few reference-grade pollution-monitoring stations have been deployed therein, with only two such stations located within Chiang Mai city. This sparse distribution is insufficient to capture the extensive spatial variability of air quality across such a large region, especially in rural and suburban areas, and also complicates the calibration of data from the far more numerous LCSs. To address this, spatial interpolation methods have become essential. Of these, Inverse Distance Weighting (IDW) is a widely used deterministic technique that can used to estimate unknown $PM_{2.5}$ concentrations in unmonitored locations to construct more localized reference $PM_{2.5}$ values for improving air quality monitoring throughout the expansive Chiang Mai region.

In our study, we adopted a two-phase approach. In the first phase, we applied IDW to estimate $PM_{2.5}$ concentrations and create a localized reference framework tailored to the conditions at the LCS sites. By generating interpolated $PM_{2.5}$ values that reflect local variations more accurately, this approach mitigates the bias associated with distant reference monitors and provides a robust baseline for calibration. The second phase involved calibrating the LCS data using this newly constructed reference framework. To identify the most effective approach for enhancing LCS accuracy, we investigated employing several advanced calibration methods, including Multiple Linear Regression (MLR), a Generalized Additive Model (GAM), Random Forest (RF), and a Long Short-Term Memory (LSTM) neural network. The proposed methodology not only refines the pollution-monitoring calibration process for under-monitored regions but also strengthens the overall utility of LCS networks, ultimately contributing to more accurate air quality assessment and informed public health policies.

2. Materials and Methods

2.1. $PM_{2.5}$ Observation and Reference Data

Hourly $PM_{2.5}$ concentration data collected from January 1 to December 31, 2023, at two reference-grade PCD monitoring stations, 35T/CM and 36T/CM, in Chiang Mai city were supplemented with data from four LCSs stations. The LCS devices, which employ light-scattering techniques for real-time $PM_{2.5}$ measurements, were deployed as part of the Air Quality Awareness Raising under the American-Thai Collaboration (AQAAT) project. All four LCS stations are situated at varying distances from the reference monitors: LCS04 in Chom Thong is approximately 60 km away, LCS10 in Chiang Dao is around 90 km away, LCS13 in Mae Rim is roughly 11 km away, and LCS15 is approximately 4 km from the nearest reference station (Figure 1). Calibration of the sensor data involved relative humidity (RH) and temperature as covariates, in recognition of the significant influence of these two environmental factors on particulate matter measurements. Further details regarding the LCSs employed in this study are available at https://aqaat.narit.or.th/aqaat/index.php.

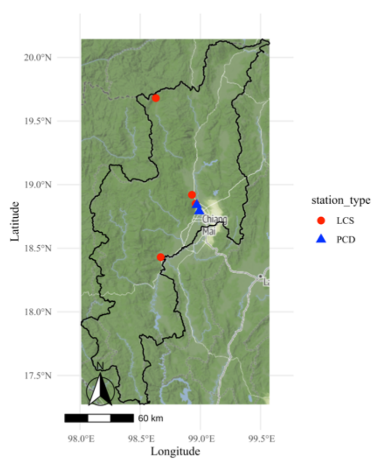

Fig. 1: Location of LCS and reference (PCD) $PM_{2.5}$ stations in the study area of Chiang Mai, Thailand.

2.2. Spatial Interpolation Methods for $PM_{2.5}$ Concentration Estimation

IDW is a representative conventional interpolation method for $PM_{2.5}$ estimation and a widely used deterministic technique for estimating unknown $PM_{2.5}$ values by assigning weights inversely proportional to the distance from a known data point [11]. This approach operates on the premise that locations closer to known measurements exert a greater influence on the estimated values than those farther away [12]. Various researchers have employed IDW due to its speed, ease of use, and ability to provide reasonable approximations of $PM_{2.5}$ distributions across different spatial scales [4, 12]. However, conventional IDW has inherent limitations, particularly when estimating local $PM_{2.5}$ concentrations; the assumption of constant distance decay may not adequately capture complex, short-term, or highly localized variations in $PM_{2.5}$ levels influenced by heterogeneous sources or environmental conditions [12]. Such assumptions can lead to inaccuracies when applying IDW in areas with intricate spatial patterns or where conditions change rapidly over small distances. Despite these challenges, IDW remains attractive for many applications because of its straightforward implementation and capacity to generate quick, interpretable results [11]. The concept of IDW can be expressed as follows:

$$Z = \sum_{i=1}^{n} W_i Z_i, \tag{1}$$

where Z is the estimated value used for the target interpolation location, n is the number of reference (measured) values near the target location, Z_i represents the measured or known value of the ith reference (measured) value, and W_i is the weight of the ith value defined as

$$W_i = \frac{\left(\frac{1}{d_i}\right)^{\alpha}}{\sum_{i=1}^{n}\left(\frac{1}{d_i}\right)^{\alpha}}, \tag{2}$$

where d_i is the distance between the target location and the ith point and α is a constant power used to adjust the diminishing strength in the relationship with increasing distance [11].

2.3. Calibration Method Selection

Before calibration, data preprocessing was undertaken to eliminate outliers in the LCS $PM_{2.5}$ concentration data. Extremely high $PM_{2.5}$ values can complicate calibration model development, so measurements exceeding three times the scaled median absolute deviation (MAD) were excluded. Scaled MAD is computed as follows:

$$MAD = k \times median\left(abs\left(M - median\left(M\right)\right)\right), \tag{3}$$

where M denotes the observed $PM_{2.5}$ concentration, *median* identifies the central tendency, and k is a scaling factor denoted as

$$k = \frac{1}{\sqrt{2} \times erfcinv\left(\frac{3}{2}\right)} \approx 1.4826, \tag{4}$$

where $erfcinv$ represents the inverse complementary error function.

Four regression algorithms were applied to calibrate the LCS data: MLR [6], GAM [13], RF [14], and LSTM neural network [15]. The temperature (°C), relative humidity (%), and temporal features (month and time of day) were incorporated as predictors. The time of day was stratified into morning (6:00 AM to 12:00 PM), noon (12:00 PM to 6:00 PM), evening (6:00 PM to 12:00 AM), and night (12:00 AM to 6:00 AM). The estimated $PM_{2.5}$ concentrations estimated via IDW at each LCS site were used as the dependent variable. Model performance was evaluated using root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). Data preprocessing, data analysis; MLR, GAM, RF and LSTM calculation; and model validation were conducted by leveraging packages “gam,” “ggplot2,” “mgcv,” “tidyr”, “plotly”, “dplyr”, “caret”, “randomForest”, “keras” and “car” in the R (Version 4.3.3) statistical software package.

3. Results and Discussion

3.1. Seasonal Variation in $PM_{2.5}$ Concentration

Hourly averaged $PM_{2.5}$ concentration, RH, and temperature data from January 1, 2023, to December 31, 2023, were obtained from LCS stations in four different districts and two reference stations in Chiang Mai. The data were stratified into three seasons: summer (January to April), rainy (May to August), and winter (September to December). Notably, the LCS measurements exhibited the highest mean $PM_{2.5}$ levels during the summer (96.8 ± 80.9 µg/m³), with a wide range (5.5–747.4 µg/m³) whereas the reference stations reported somewhat lower yet still elevated mean values (68.6 ± 47.9 µg/m³). These elevated concentrations are primarily attributable to the dry weather conditions and pervasive biomass burning in northern Thailand during this period. In particular, intense biomass burning during April in Doi Luang Chiang Dao provided the highest $PM_{2.5}$ measurements at the Chiang Dao LCS site. In contrast, both the LCS and reference stations record substantially lower mean $PM_{2.5}$ levels in the rainy and winter seasons, largely due to the mitigating effects of precipitation and cooler temperatures on particulate matter accumulation (Table 1).

Season-specific averages for temperature (approximately 26–28 °C in the summer) and RH (around 56–57% in the summer versus higher RH later in the year) further demonstrate the climatic influence on air pollution levels. Figure 2 reveals a pronounced $PM_{2.5}$ peak from late February through April across all monitoring sites, followed by a sharp decline during the rainy season and relatively moderate concentrations throughout the winter. These trends align with studies in other regions of Southeast Asia, which similarly report peak $PM_{2.5}$ concentrations during dry and biomass-burning periods, followed by substantial reductions once seasonal rainfall commences [10, 16]. However, in contrast to some urban settings where vehicular emissions predominate year-round $PM_{2.5}$ levels, the Chiang Mai data underscore the significant role of seasonal biomass burning in driving high pollution episodes [10]. Our observations extend and reinforce the importance of both meteorological conditions and local land-use practices (e.g., biomass burning) in shaping the ambient air quality profile for northern Thailand.

Table 1: Seasonal $PM_{2.5}$ levels measured at the LCS and reference stations.

Station/ Variable	Summer		Rainy		Winter	
	Mean±SD	Range	Mean±SD	Range	Mean±SD	Range
LCS						
$PM_{2.5}$ (µg/m³)	96.8 ± 80.9	5.5 – 747.4	14.5 ± 14.9	0.1 – 355.8	19.1 ± 14.4	0.1 – 282.6
T (°C)	26.2 ± 6.0	9.8 – 44.6	28.8 ± 3.7	21.7 – 43.3	26.4 ± 3.9	13.9 – 40.7
RH (%)	56.5 ± 16.0	15.2 – 98.0	72.2 ± 14.0	31.0 – 98.9	75.7 ± 13.1	38.4 – 99.0
Reference						
$PM_{2.5}$ (µg/m³)	68.6 ± 47.9	6.0 – 378.0	16.9 ± 10.5	1.4 – 88.0	16.0 ± 8.1	1.4 – 63.3

Abbreviation: SD, Standard deviation; T, temperature; RH, relative humidity. Seasons are defined as Summer (January to April), Rainy (May to August), and Winter (September to December).

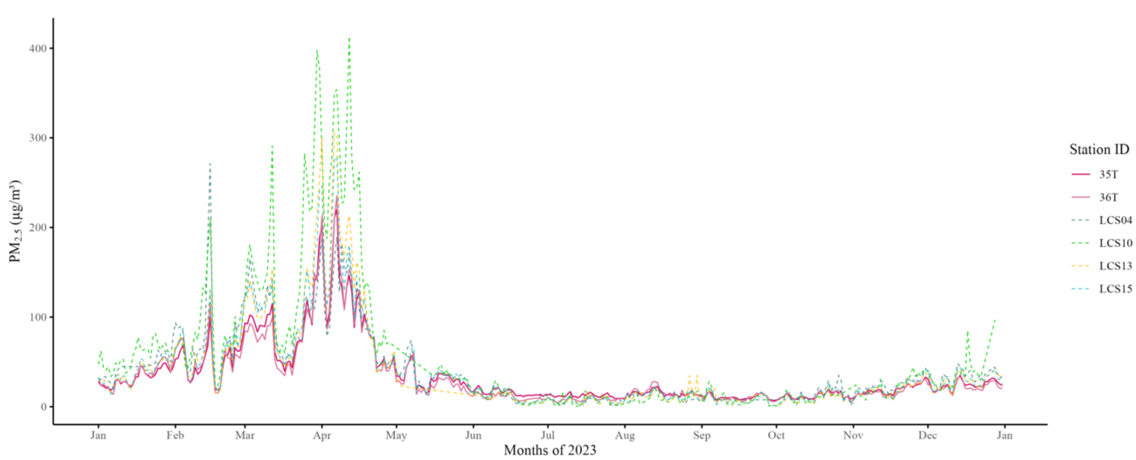

Fig. 2: 24-hourly averaged $PM_{2.5}$ concentrations measured at the reference (35T and 36T) and LCS (LCS04, LCS10, LCS13, and LCS15) stations from January 1, 2023, to December 31, 2023, in Chiang Mai, Thailand.

3.2. IDW for Estimating Reference $PM_{2.5}$ Values and Constructing the Reference Dataset

We employed IDW to estimate $PM_{2.5}$ concentrations at unmonitored locations, thereby creating a more localized reference framework for subsequent analyses. Error metrics (RMSE, MAE, and MAPE) derived from the IDW-interpolated $PM_{2.5}$ values were compared with those obtained directly from reference station (Table 2). Overall, the IDW-derived data were nearly identical to the reference-based values. For example, the all-stations average RMSE for IDW (37.15 µg/m³) was only marginally lower than that for the reference-based (37.40 µg/m³), indicating that IDW can effectively approximate $PM_{2.5}$ in areas lacking real-time reference data.

However, the small difference could be partially attributed to the limited number of reference stations available to construct the IDW surface. When reference data are sparse, interpolation may not fully capture local variability, potentially constraining IDW's accuracy. This limitation aligns with other studies suggesting that while IDW performs robustly under data-limited scenarios, the addition of more reference monitors or the use of advanced interpolation techniques can further improve accuracy of $PM_{2.5}$ estimation [12, 17]. Despite these constraints, our findings demonstrate that IDW remains a practical method for expanding spatial coverage where standard monitoring infrastructure is inadequate. Our results are consistent with research indicating that IDW can maintain satisfactory performance even in settings characterized by sparse reference-grade coverage, while contrasting with studies reporting greater gains from alternative methods when denser monitoring networks are available [17].

Table 2: Comparison of $PM_{2.5}$ estimation using IDW interpolation versus reference measurements across the LCS stations.

Station	IDW Interpolation			Reference Data		
	RMSE (µg/m³)	**MAE (µg/m³)**	**MAPE (%)**	**RMSE (µg/m³)**	**MAE (µg/m³)**	**MAPE (%)**
Chiang Dao	62.21	27.37	73.4	62.55	27.41	75.1
Chom Thong	27.51	13.97	44.4	27.94	14.16	47.9
Mae Rim	22.92	11.04	37.2	23.13	11.25	37.9
Mueang Chiang Mai	19.19	9.91	33.2	19.24	9.96	33.4
All stations	**37.15**	**15.62**	**47.3**	**37.40**	**15.65**	**48.4**

3.3. Seasonal Variation-Based Model Calibration Performance Evaluation

To ensure a robust and reliable performance assessment, the calibration performances of the four models were investigated using 10-fold cross-validation. MLR, GAM, RF, and LSTM were employed to capture the influence of seasonal patterns on LCS measurements. Table 3 provides the results for performance metrics: R^2, RMSE, MAE, and MAPE before and after calibration. Prior to calibration, the LCS data exhibited only moderate alignment with the reference-grade measurements, with R^2 values ranging from 0.42 to 0.50 across all seasons and relatively high RMSE and MAE scores. However, after calibration, all four models showed notable improvement. LSTM consistently achieved the highest R^2 values (ranging from 0.75 to 0.86) and the lowest RMSE, thus illustrating the advantage of incorporating meteorological factors (temperature, RH) and temporal variables (month and time of day) into the calibration process. For example, for the summer season, LSTM attained an R^2 value of 0.86, an RMSE value of 18.48 µg/m³, and a MAPE decrease of 30%, thereby indicating a significant enhancement over using the uncalibrated LCS $PM_{2.5}$ data. Similar gains were observed for the rainy and winter periods, albeit against lower overall $PM_{2.5}$ concentrations.

These findings underscore the efficacy of advanced machine learning methods in adjusting for seasonal dynamics and sensor-specific biases inherent in LCS data. By including meteorological and temporal covariates, models like LSTM can capture variations that simpler regressions might overlook [6]. Previous studies in regions with marked seasonal trends in $PM_{2.5}$ (e.g., areas subject to biomass burning or monsoonal cycles) have reported comparable success when integrating meteorological information into calibration frameworks [3, 10]. In contrast, research conducted in settings with more uniform climate conditions or lower baseline pollution has sometimes found smaller improvements, suggesting that local environmental factors significantly influence calibration outcomes [8].

The high R^2 values across multiple seasons bolstered by the 10-fold cross-validation design that mitigates overfitting highlight the consistency of our calibration approach and help validate the generalizability of our results. Nevertheless, one limitation of our study is the relatively sparse network of reference-grade monitors in Chiang Mai, which may constrain the spatial accuracy of any calibration effort. Future work could involve expanding the reference network or exploring alternative data-fusion approaches to capture finer-scale variations in air quality. Despite these constraints, the present results reveal that a careful combination of machine learning calibration methods, meteorological inputs, and validation procedures can substantially enhance the reliability of LCS-derived $PM_{2.5}$ measurements.

Table 3: Comparison of model fitting with 10-fold cross-validation for hourly $PM_{2.5}$ concentrations calibrated using MLR, GAM, RF, and LSTM models

Season	Scenario	Model	R^2	RMSE (µg/m³)	MAE (µg/m³)	MAPE (%)	Equation
Summer	Before calibration	-	0.50	62.29	32.35	47.08	-
	After calibration	MLR	0.76	23.56	15.30	25.59	$PM_{2.5\_REF}$ =0.64($PM_{2.5\_LCS}$)+1.58(T)+0.31(RH)+ 1.33(Feb)+1.05(Mar)+10.97(Apr)+9.19(morning) +10.5(noon)-0.08(evening)-52.51
		GAM	0.77	22.83	14.62	24.81	-
		RF	0.80	21.25	12.73	19.96	-
		LSTM	0.86	18.48	11.24	17.31	-
Rainy	Before calibration	-	0.47	11.05	7.35	48.33	-
	After calibration	MLR	0.60	6.41	4.22	26.88	$PM_{2.5\_REF}$ = 0.40($PM_{2.5LCS}$)+0.11(T)+0.02(RH)- 6.47(Jun)-7.56(Jul)-4.98(Aug)+1.30(morning) +0.89(noon)+1.66(evening)+10.46
		GAM	0.65	6.09	4.00	26.04	-
		RF	0.72	5.36	3.68	24.33	-
		LSTM	0.77	4.84	3.39	22.50	-
Winter	Before calibration	-	0.42	13.77	7.50	49.68	-
	After calibration	MLR	0.66	4.54	3.38	24.38	$PM_{2.5\_REF}$ = 0.30($PM_{2.5\_LCS}$)-0.14(T)- 0.03(RH)+0.76(Oct)+3.82(Nov)+5.78(Dec) +0.57(morning)-0.43(noon)+2.98(evening)+13.25
		GAM	0.69	4.33	3.21	22.89	-
		RF	0.70	4.29	3.20	23.20	-
		LSTM	0.75	3.87	2.88	20.86	-

3.4. Post-Calibration LCS $PM_{2.5}$ Data

Figure 3 illustrates scatter plots comparing the post-calibration LCS $PM_{2.5}$ measurements across the four stations at Chiang Dao, Chom Thong, Mae Rim, and Mueang Chiang Mai with the reference-grade readings from the PCD. The post-calibration R^2 values ranged from 0.94 to 0.96, thereby indicating a strong correlation between the LCS and reference $PM_{2.5}$measurements. Moreover, the RMSE values ranging from 10.92–13.84 µg/m³ confirm the effectiveness of the calibration procedure in reducing deviation from the reference readings. These high R^2 results underscore the success of integrating spatial interpolation based on IDW with a robust calibration model to mitigate systematic bias and location-specific environmental variations.

The improvements in the error metrics (RMSE and R^2) suggest that the post-calibration data distributions are more tightly aligned with the reference observations, reflecting greater consistency when measuring $PM_{2.5}$ levels across diverse settings. In Mueang Chiang Mai, which is urbanized, the calibration yielded highly accurate results, likely due to a combination of denser emission sources and relatively closer proximity to the reference stations. However, in the suburban and rural locations (Chiang Dao and Chom Thong), the slightly wider distributions in the $PM_{2.5}$ readings before calibration were substantially reduced after calibration. The reduced RMSE values across all of the LCS sites highlight the adaptability of the calibration framework, even in areas with sparse monitoring coverage. These findings align with several studies in which the researchers documented significant improvements in LCS accuracy by applying spatially informed calibration methods, especially in heterogeneous regions with varying emission profiles [7]. However, some research conducted in less topographically diverse or more uniformly urbanized regions has revealed more modest benefits from calibration, thereby suggesting that local emission patterns and the spatial density of reference monitors can influence outcomes [7, 8]. In the context of the Chiang Mai region, incorporating IDW with an advanced calibration models appear especially valuable given the patchy distribution of reference-grade stations and the stark contrasts between urban and suburban pollution sources.

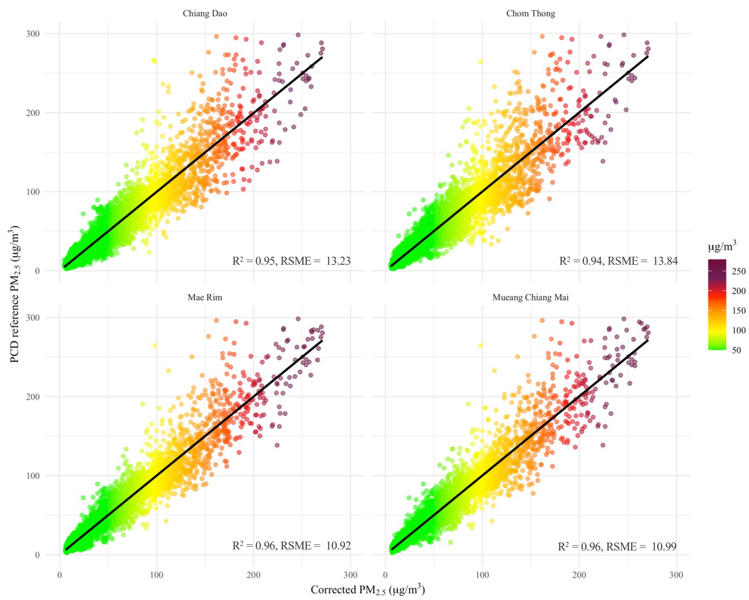

Fig. 3: Comparison between the reference and corrected LCS $PM_{2.5}$ data across all four stations

4. Conclusion

We successfully applied a two-phase approach to improve the accuracy of LCS $PM_{2.5}$ measurements in a region with limited reference station monitoring. In the first phase, IDW interpolation was adopted to construct a localized reference framework, thereby mitigating spatial mismatches caused by distant reference-grade stations. In the second phase, calibration models MLR, GAM, RF, and LSTM incorporating meteorological and temporal factors were applied to address sensor-specific bias and seasonal variability. Combining IDW-based interpolation with an advanced calibration model substantially enhanced the reliability of LCS $PM_{2.5}$ measurements, as evidenced by high R^2 values up to 0.96 and RMSE values as low as ~10 µg/m³ relative to the reference-grade data. These improvements were evident across diverse conditions, from urban areas like Mueang Chiang Mai to suburban and rural districts such as Chiang Dao and Chom Thong where localized burning and other factors initially produced greater discrepancies. Although the sparse distribution of reference monitors presents challenges, the proposed two-phase framework effectively reduces systematic errors and can adapt to variation in the local environment. In practical terms, this methodology enables more accurate air quality assessments and supports evidence-based interventions, even where reference stations are limited. In future research, we will consider alternative calibration strategies that are designed to function with more distantly located reference monitors that potentially bypass the need for the initial IDW step. Such an approach might involve combining machine-learning algorithms with supplementary data (e.g., satellite-based emission inventories) to develop calibration models that directly accommodate sparse monitoring networks. Overall, these findings highlight the potential of integrated spatial and calibration techniques to enhance $PM_{2.5}$ monitoring, ultimately contributing to more effective pollution management and public health protection.

Acknowledgements

We gratefully acknowledge the Pollution Control Department of Thailand for providing the essential data used in this study. We also extend our sincere thanks to the National Astronomical Research Institute of Thailand (NARIT) for their support in supplying and maintaining the low-cost sensor (LCS) instruments, as well as for providing the corresponding data. Additionally, we are deeply appreciative of Chiang Mai University for awarding the CMU Presidential Scholarship, which has supported my PhD studies. This research would not have been possible without the generous support and contributions from these institutions.

References

[1] M. E. Mathieu-Campbell, C. Guo, A. P. Grieshop and J. Richmond-Bryant, “Calibration of PurpleAir low-cost particulate matter sensors: model development for air quality under high relative humidity conditions,” *Atmos. Meas. Tech.,* vol. 17, pp. 6735–6749, 2024.

[2] R. Byrne, J. C. Wenger and S. Hellebust, “Spatial analysis of PM2.5 using a concentration similarity index applied to air quality sensor networks,” *Atmos. Meas. Tech.,* vol. 17, pp. 5129–5146, 2024.

[3] L. F. Weissert, G. S. Henshaw, A. L. Clements, R. M. Duvall and C. Croghan, “Seasonal effects in the application of the MOMA remote calibration tool to outdoor PM2.5 air sensors,” *EGUsphere [preprint],* 2024.

[4] H.J. Chu, M.Z. Ali and Y.C. He, “Spatial calibration and PM2.5 mapping of low-cost air quality sensors,” *Scientifc Reports,* vol. 10, pp. 22079, 2020.

[5] J. Hua, Y. Zhang, B. de Foy, X. Mei, J. Shang, Y. Zhang, I.D. Sulaymon and D. Zhou, “Improved PM2.5 concentration estimates from low-cost sensors using calibration models categorized by relative humidity,” *Aerosol Science and Technology*, vol. 55, no. 5, pp. 600-613, 2021.

[6] C. McFarlane, G. Raheja, C. Malings, E. K. E. Appoh, A.F. Hughes and D.M. Westervelt, “Application of gaussian mixture regression for the correction of low cost PM2.5 monitoring data in Accra, Ghana,” *ACS Earth and Space Chemistry,* vol. 5, no. 9, pp. 2268-2279, 2021.

[7] S. De Vito, A. Del Giudice, G. D’Elia, E. Esposito, G. Fattoruso, S. Ferlito, F. Formisano, G. Loffredo, E. Massera, P. D’Auria and G. Di Francia, “Future low-cost urban air quality monitoring networks: insights from the EU’s AirHeritage Project,” *Atmosphere*, vol.15, no. 11, pp. 1351, 2024.

[8] R. K. Huda, P. Kumar, R. Gupta, A. K. Sharma, G. S. Toteja and B. V. Babu, “Air quality monitoring using low-cost sensors in urban areas of Jodhpur, Rajasthan,” *Int. J. Environ. Res. Public Health.*, vol. 21, no. 5, pp. 623, 2024.

[9] K. Jainontee, P. Pongkiatkul, Y. L. Wang, R. J. F. Weng, Y. T. Lu, T. S. Wang and W. K. Chen, “Strategy design of PM2.5 controlling for Northern Thailand,” *Aerosol and Air Quality Research,* vol. 23, no. 6, pp. 220432, 2023.

[10] R. Dejchanchaiwong, P. Tekasakul, A. Saejio, T. Limna, T. C. Le, C. J. Tsai, G. Y. Lin and J. Morris, “Seasonal field calibration of low-cost PM2.5 sensors in different locations with different sources in Thailand,” *Atmosphere*, vol. 14, no. 3, pp. 496, 2023.

[11] G.Y. Lu and D.W. Wong, “An adaptive inverse-distance weighting spatial interpolation technique,” *Comput. Geosci.,* vol. 34, pp. 1044–1055, 2008.

[12] K. Choi and K. Chong, “Modified inverse distance weighting interpolation for particulate matter estimation and mapping,” *Atmosphere,* vol. 13, no. 5, pp. 846, 2022.

[13] T. Hastie and R. Tibshirani, “Generalized additive models,” *Statistical Science*, vol. 1, no. 3, pp. 297-318, 1986.

[14] G. Raheja, J. Nimo, E. K. E. Appoh, B. Essien, M. Sunu, J. Nyante, M. Amegah, R. Quansah, R. E. Arku, S. L. Penn, M. R. Giordano, Z. Zheng, D. Jack, S. Chillrud, K. Amegah, R. Subramanian, R. Pinder, E. Appah-Sampong, E.N. Tetteh, M.A. Borketey, A.F. Hughes and D.M. Westervelt, “Low-cost sensor performance intercomparison, correction factor development, and 2+ years of ambient PM2.5 monitoring in Accra, Ghana,” *Environ. Sci. Technol*., vol. 57, no. 29, pp. 10708-10720, 2023.

[15] Y. Su, J. Li, L. Liu, X. Guo, L. Huang and M. Hu, “Application of CNN-LSTM Algorithm for PM2.5 Concentration Forecasting in the Beijing-Tianjin-Hebei Metropolitan Area,” *Atmosphere,* vol. 14, no. 9, pp. 1392, 2023.

[16] S.K. Pani, S. Chantara, C. Khamkaew, C.T. Lee and N.H. Lin, “Biomass burning in the northern peninsular Southeast Asia: Aerosol chemical profile and potential exposure,” *Atmospheric Research*, vol. 224, pp. 180-195, 2019.

[17] A. D. Piazza, F. L. Conti, F. Viola, E. Eccel and L. V. Noto, “Comparative analysis of spatial interpolation methods in the mediterranean area: application to temperature in Sicily,” *Water,* vol.7, no. 5, pp. 1866-1888, 2015.