scientific reports

OPEN

Enhancing the reliability of particulate matter sensing by multivariate Tobit model using weather and air quality data

Wan-Sik Won[1,2], Jinhong Noh[3], Rosy Oh[4], Woojoo Lee[5], Jong-Won Lee[6], Pei-Chen Su[1]✉ & Yong-Jin Yoon[1,3]✉

Low-cost particulate matter (PM) sensors have been widely used following recent sensor-technology advancements; however, inherent limitations of low-cost monitors (LCMs), which operate based on light scattering without an air-conditioning function, still restrict their applicability. We propose a regional calibration of LCMs using a multivariate Tobit model with historical weather and air quality data to improve the accuracy of ambient air monitoring, which is highly dependent on meteorological conditions, local climate, and regional PM properties. Weather observations and $PM_{2.5}$ (fine inhalable particles with diameters ≤ 2.5 μm) concentrations from two regions in Korea, Incheon and Jeju, and one in Singapore were used as training data to build a visibility-based calibration model. To validate the model, field measurements were conducted by an LCM in Jeju and Singapore, where R^2 and the error after applying the model in Jeju improved (from 0.85 to 0.88) and reduced by 44% (from 8.4 to 4.7 μg m^{-3}), respectively. The results demonstrated that regional calibration involving air temperature, relative humidity, and other local climate parameters can efficiently correct the bias of the sensor. Our findings suggest that the proposed post-processing using the Tobit model with regional weather and air quality data enhances the applicability of LCMs.

Rapid advances in computing systems and machine learning (ML) have increased the number of sensor technologies for real-time application by using abundant data and discovering hidden properties and patterns[1,2]. Data-driven approaches and scientific findings using big data analysis have shown great potential for applications in sensors and environmental industries[3–6]. Low-cost monitors (LCMs) are widely used in various types of urban areas to monitor air quality in real time. To improve the accuracy and applicability of LCMs, various environmental parameters have been proposed, showing promising improvements in sensing performance[7–9].

Particulate matter (PM) is a key parameter of air quality and its measurement is required for accurate air quality monitoring. In particular, the concentration of fine particulate matter ($PM_{2.5}$, fine inhalable particles with diameters ≤ 2.5 μm) is a reliable indicator of PM exposure, as it has a significant impact on mortality globally[10–12]. Regulatory authorities in charge of countries' environmental policies have generally used reference instruments that operate based on gravimetry or beta ray attenuation. These reference measurement systems have very high accuracy and precision, but their installation and maintenance costs are high; thus, LCMs, and especially LCMs with improved sensors and incorporated Internet of Things (IoT) technology, are attractive alternatives for high spatial density PM monitoring[13–15]. Studies have shown that LCMs are reliable and exhibit high accuracy during laboratory testing as well as when calibrated in the field[16–19]. Conversely, other studies have concluded that LCMs are sensitive to climate parameters such as air temperature and relative humidity (RH); however, long-term measurements combined with post-processing or ML can correct the bias of the sensors[7,20].

[1]School of Mechanical and Aerospace Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore. [2]Department of Aerospace Industrial and Systems Engineering, Hanseo University, Taean, Chungcheongnam-do 32158, Republic of Korea. [3]Department of Mechanical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. [4]Department of Mathematics, Korea Military Academy, Seoul 01805, Republic of Korea. [5]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul 08826, Republic of Korea. [6]Observer Foundation, Seoul 04050, Republic of Korea. ✉email: peichensu@ntu.edu.sg; yongjiny@kaist.ac.kr

LCMs are advantageous because they are easy to install and can perform real-time high-resolution PM monitoring[21,22]; however, their performance varies among sensors, and rigorous scientific verification of their reliability has so far been insufficient for regulatory use[20,23,24]. Compared to reference measurement systems that operate based on a filter-based gravimetric method or beta-attenuation monitoring (BAM), LCMs have inherently low accuracy, as they estimate mass concentration indirectly based on optical measurements of light scattering; thus, uncertainties arise, which are associated with meteorological factors, background PM concentration, aerosol chemical composition, and aerosol size distribution, which compromise the reliability of the sensors[25,26].

Another limitation is that LCMs typically do not perform conditioning of air temperature and RH of the sampled air. In the case of reference instruments, 24 h average air temperature and RH are conditioned during sampling between 20 and 23 °C and 30% and 40%, respectively[27–29]. LCMs typically do not control the sampling conditions (Fig. 1); hence, measurements are influenced by weather variations. The effects of air temperature and RH may be negligible in areas where PM concentrations are not significant, as well as during short-term field campaigns, depending on PM sources and background concentration[30,31]. However, weather may significantly impact the accuracy and repeatability of long-term monitoring in an area with highly variable PM concentrations during a high RH period due to $PM_{2.5}$ hygroscopicity[26,32,33]. For these reasons, LCMs perform differently in different areas; for example, the same sensor has exhibited different accuracies in the USA compared to India or China[16,34,35]. Hence, regional calibration incorporating local climate information is an effective solution[36].

As a part of the process of identifying regional differences that affect LCM performance, we present a post-processing method that uses airport visibility and $PM_{2.5}$ concentration as training data sets. Visibility is an indicator of air quality that may include information on the concentration and hygroscopic growth of aerosol particles[37–39]; therefore, although airport visibility and PM concentration measured with an LCM are different parameters, they are both based on light scattering. Similar to an LCM estimating PM concentration from light scattering intensity, airport visibility has been reported based on light-source properties and transmission factors that contain information on light scattering around airport runways[40]. Airports are ideal for studying visibility because it is reported every hour and even more frequently under adverse conditions for aviation safety[41,42]. After collecting all weather data, it is possible to build a model to train the relationship between weather parameters and $PM_{2.5}$, using a well-known basic light scattering principle[43], which enables visibility prediction. Conversely, $PM_{2.5}$ concentration can be estimated under various weather and visibility conditions by establishing empirical relationships regarding $PM_{2.5}$, RH, and visibility[44–48].

Here, we propose a regional calibration of LCMs that does not have an air-conditioning function using a multivariate Tobit model with airport weather and $PM_{2.5}$ concentration data. We collected data from two middle-latitude regions in Korea, Incheon and Jeju, and one equatorial region in Singapore, thereby assembling a training dataset of visibility, weather parameters, and $PM_{2.5}$ concentration. To calibrate the model to LCMs in different regions, we also conducted field measurements in Jeju and Singapore and compared the results before and after calibration, focusing on the differences in local climate that may affect LCM performance. Finally, we proposed better ways to use LCMs in different regions while overcoming their limitations during field measurements and could show that regional LCM calibration is feasible even without long-term field experiments.

Results

LCM dependency on RH. Figure 1 illustrates air conditioning in the case of BAM and LCM under high RH. Since the $PM_{2.5}$ concentration, due to ambient moisture and hygroscopicity, may be uncertain, the BAM maintains dry conditions by evaporating water with a heater at the inlet[29,49]. $PM_{2.5}$ concentration is precisely determined by beta-ray attenuation immediately after moisture has been removed from the sampled air. Conversely, the LCM detects light scattering by hygroscopic particles under high RH, which causes bias unless the LCM has undergone repeated calibration correctly.
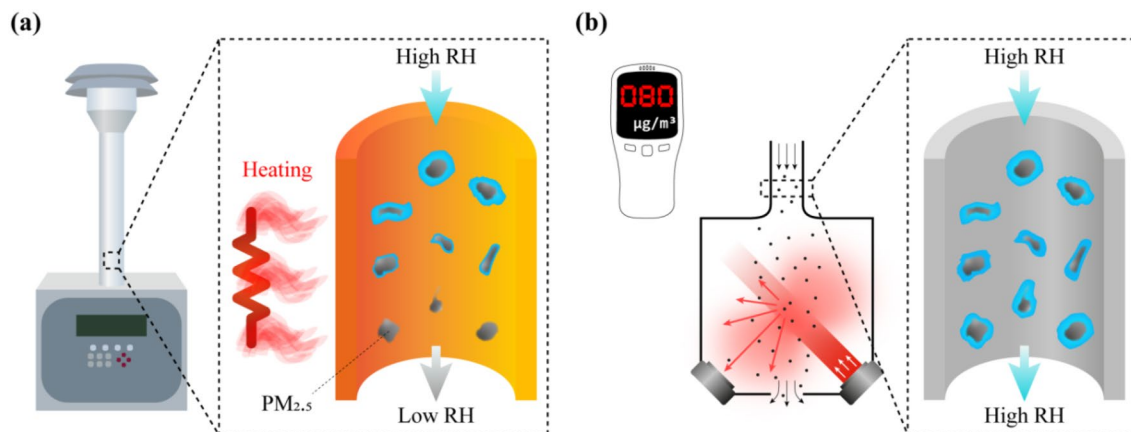
Figure 1. Schematic illustration of difference in $PM_{2.5}$ measurement between (**a**) a BAM and (**b**) a LCM, specifically under highRH. The inlet tube in the BAM is heated to control the sampling conditions to RH 35%. Conversely, the LCM does not have air-conditioning function, resulting in increased light scattering through hygroscopic effect under high RH.

To validate the model, LCM field testing was conducted over seven months, from March 25 to October 26, 2019, in Jeju, and over 22 months, from December 1, 2020, to September 30, 2022, in Singapore (Figs. S4–S6 and Table S5). Figure 2 shows the $PM_{2.5}$, concentrations by LCM, during high and low RH in the cases of Jeju and Singapore without post-processing calibration, criteria of high and low RH set as 70% and 40%, respectively. RH 70% and 40% are indicators of high and low RH, respectively. These criteria are not an absolute value but, rather, an arbitrarily set relative value. The vertical axis indicates the response of LCM at different RH conditions compared to the control conditions (RH 30−40%) measured by the BAM, which is in a relatively low RH range. R^2 indicates the coefficient of determination between the responses from LCM and BAM. The slope indicates the unit increase in the response from LCM to the unit increase in $PM_{2.5}$ concentration from BAM.

During the LCM field testing, the average humidity in Jeju and Singapore was 65% and 69%, respectively. Field measurements in Jeju comprised of 2218 hourly observations over seven months. 9983 hourly observations were made in Singapore over a period of 22 months. Among the 2218 observations in Jeju, the number of data with RH above 70% and below 40% was 924 (red plots) and 233 (blue plots), respectively. Among 9983 observations in Singapore, the number of high and low RH data were 4556 (red) and 466 (blue), respectively.

While comparing high and low RH, the slope in the high RH case was greater than that in the low RH, both in Jeju and Singapore (1.33 and 1.10, respectively, for high RH). In terms of regional differences, measurements had a negative zero drift tendency in Jeju, whereas they had little zero drift tendency in Singapore. These results indicate that the sensor located in a more humid climate tends to yield a higher mass concentration, showing a bias between the reference station and the low-cost sensor under high RH. Therefore, the two different instruments in Fig. 2 represent the characteristics of the BAM and LCM in relation to Fig. 1, implying that the LCM without the inlet heater results in an increased bias under high RH conditions compared to the control RH (< 40%) of BAM.

Airport and LCM in relation to light scattering.   LCMs operate based on light scattering; accordingly, airport visibility is evaluated based on a light scattering sensor and optical observations of aerosols around the airport, hourly or at more frequent intervals[40]. RH and $PM_{2.5}$ concentrations were the most correlated parameters with visibility (Fig. S3). The reported airport visibility and air quality data can elucidate the differences between the high and low RH cases during the LCM operation.

Figure 3 shows an illustrated flowchart of the calibration by the multivariate Tobit model of the airport weather and $PM_{2.5}$ concentration data to correct the LCM measurements under high RH, where the LCM sample scatters light from the transmitter, and the detector estimates the mass concentration from changing a signal such as voltage in response to the scattered light. Because hygroscopic particles scatter more light under humid conditions, light scattering and electrical signal reductions should be calculated if the RH is adjusted between 30 and 40% (the low-RH case was set to 35% in the selected model). Multi-annual airport weather observations were used for training and estimating the effect of RH on light scattering, and the result was applied to the calibration process.

A visibility prediction model was built using weather observations and $PM_{2.5}$ concentration data from Incheon and Jeju in Korea and Singapore. The details of each model are presented in Tables S2 and S3, and the subsequent results are provided in Table S4. Won et al.[50] showed that visibility prediction that considers $PM_{2.5}$ concentration, meteorological parameters, and their relationships improved compared to other existing models. From the relationships between $PM_{2.5}$ hygroscopicity, the visibility, and the extinction coefficient, airport observations and LCM measurements can be related to the hygroscopicity of $PM_{2.5}$, which depends on RH. Reduced visibility under high RH indicates increased light scattering and absorption, which means an increased signal from aerosol particles under high RH, as illustrated in Fig. 3.
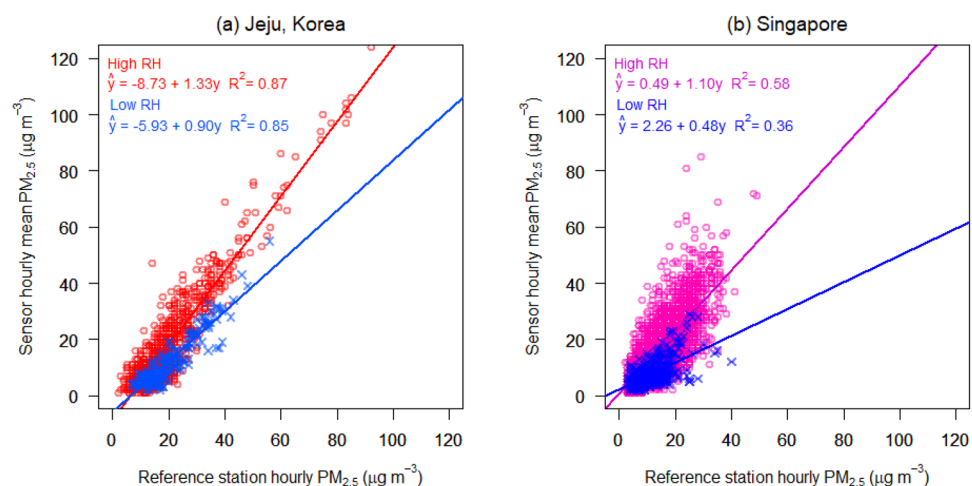

Figure 2. Different tendencies of the LCM caused by different RH in two different regions: Samples with high (> 70%) and low (< 40%) RH in (**a**) Jeju, Korea, and (**b**) Singapore. The LCM shows a positive bias of $PM_{2.5}$ concentration in the case of high RH in both locations.
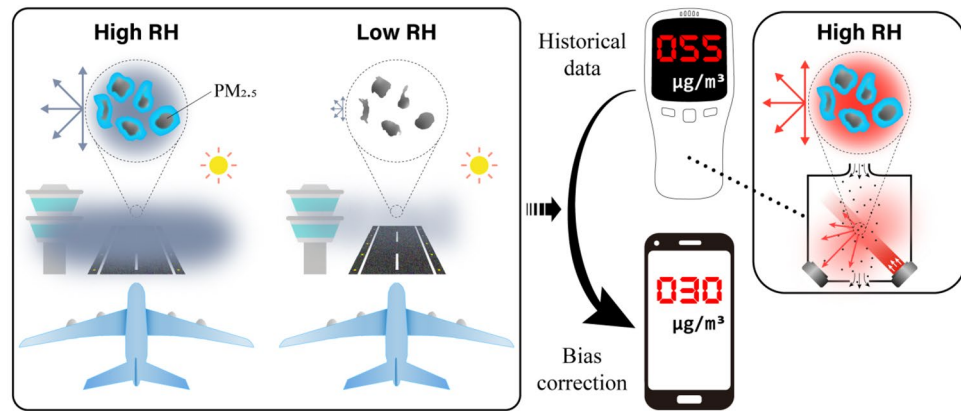
Figure 3. Illustrated flowchart of the calibration of PM$_{2.5}$ concentration measurements by multivariate Tobit model using long-term visibility, weather observations, and air quality data from airports. The biased signals recorded by the LCMs can be corrected if the PM$_{2.5}$ samples are conditioned to low RH (~30 to 40%); this is produced by the model using abundant historical data recorded at airports. Relationships between PM$_{2.5}$ concentration, RH, and visibility are presented in Fig. S3.

PM effect on light scattering depending on regional climate. Figure 4 shows the influence of air temperature (TMP), RH, and PM$_{2.5}$ concentration on visibility in three different regions, which was determined by a multivariate Tobit model using airport and air quality data (the effects of all parameters on visibility estimated by the model are summarized in Table S4). In the Korean regions, RH has the most pronounced effect on visibility (−2.31 and −2.26 km in Incheon and Jeju, respectively), followed by PM$_{2.5}$ concentration (−1.00 and −0.96 km in Incheon and Jeju, respectively), indicating that high RH and/or PM$_{2.5}$ concentration is associated with decreased visibility, while TMP has a positive effect on visibility. Meanwhile, the influence of each parameter on visibility was smaller in Singapore (−0.13 km for RH) than in the Korean regions (Fig. 4a). The difference between Singapore and Korea stems from the different local climates situated at the equator and middle latitudes, respectively. In Incheon and Jeju, the TMP variation was 53 °C (from −16 to 37 °C) and 42 °C (from −6 to 36 °C), while the RH variation was 90% (from 8 to 98%) and 88% (from 12 to 100%), respectively; while in Singapore, the TMP variation was 12 °C (from 22 to 34 °C) and the RH variation was 57% (from 43 to 100%) during the studied period (see Table S1). TMP and RH variations are not pronounced in the equatorial region; hence, their influence on visibility may be relatively small.

Figure 4b shows an enlarged view of the effect of weather parameters on visibility; the effects of RH, WS, and PM$_{2.5}$ on visibility are −0.13 km, −0.06 km, and −0.08 km in Singapore, respectively. Notably, unlike in Incheon and Jeju, the TMP effect on visibility is more pronounced in Singapore due to the local humid climate characteristics. The average RH during the study period was 81% in Singapore, and 62% and 67% in Incheon and Jeju, respectively (Table S1). At middle latitudes, RH variations were more pronounced than TMP variations. Conversely, at the equator, RH variation is quite small because the dew-point temperature is also relatively high,
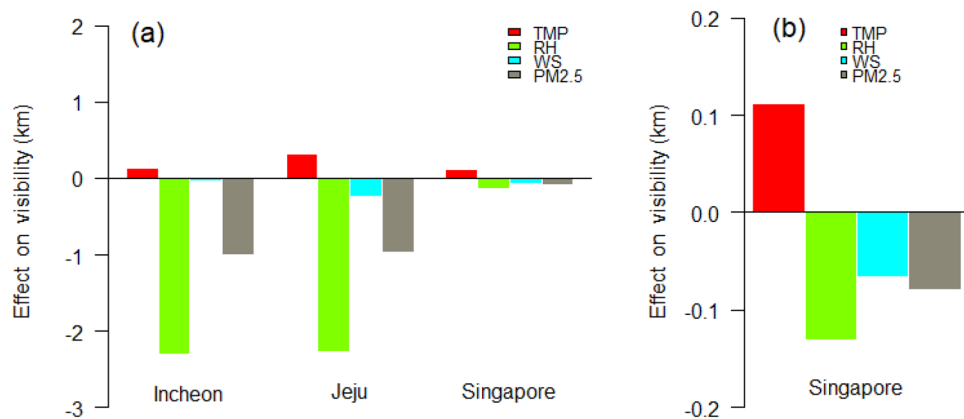

Figure 4. Regional visibility differences due to the effect of different weather parameters, i.e., air temperature (TMP), relative humidity (RH), and wind speed (WS), and PM$_{2.5}$ concentration: (**a**) The effects in Incheon and Jeju are similar to each other, while those in Singapore are less pronounced due to the different regional climate. (**b**) The figure on the right shows an enlarged view in Singapore. The RH effect is negatively most pronounced, while the TMP effect is much more positively pronounced compared to the others in Singapore than in Incheon and Jeju.

even when the TMP is high; thus, in Singapore, the influence of TMP on visibility is as pronounced as that of RH. The results from the model for the three regions reflect the similarities between neighboring mid-latitude regions and the climate characteristics at the equator.

Calibration result depending on regional models. LCM field testing combined with the calibration method was conducted in Jeju and Singapore (details are provided in "Methods" section and Fig. S5). Figure 5 shows the field measurement results for $PM_{2.5}$, after applying the regional calibration by the model. Figure 5a presents the raw hourly $PM_{2.5}$ concentration against the hourly measurements of the reference station, *Yeon-dong*, which is located near Jeju International Airport, and the other three panels present the calibrated $PM_{2.5}$ concentration by the Jeju, Incheon, and Singapore models, respectively. Similar to the data in Fig. 2, raw LCM data exhibits mostly positive bias compared to those of the reference station since the $PM_{2.5}$ concentration increases according to the 1.21 slope of linear regression. Conversely, low raw $PM_{2.5}$ concentrations exhibit a negative bias compared to those of the reference station, indicating that the sensor has both zero and sensitivity drift in Jeju.

The three post-processed $PM_{2.5}$ concentrations appear improved with linear regression slopes of 0.97, 0.98, and 1.11 respectively (Fig. 5b–d). Root Mean Square Error (RMSE) is 4.7, 4.8, and 6.1 after applying the Jeju, Incheon, and Singapore models; while the Jeju model exhibits, lowest error (Fig. 5b). The Incheon model result (Fig. 5c) is almost the same as that of the Jeju model. The Singapore model result (Fig. 5d) also shows improved accuracy; nevertheless, it exhibits a tendency toward higher values compared with the Jeju and Incheon model results.

$PM_{2.5}$ measurements using LCM in Singapore and the calibrated results from the three regional models are shown in Fig. S8. Similar to Fig. 5, the plots from Jeju and Incheon are similar. The Singapore model result is different from the Jeju and Incheon model results, with a positive bias still remaining. It can be seen from the empirical relation stemming from the middle latitudes[46], the extinction coefficient at 80–90% RH and average $PM_{2.5}$ concentration of 22 $\mu g\ m^{-3}$ is about 10–18 M m^{-1} (see "Methods" section, Eq. 3), while a study on hygroscopicity and visibility reported a value of 5.7–7.0 M m^{-1} in Singapore[51]. This difference may be reflected in the model, resulting in a more pronounced change after the calibration in Jeju than in Singapore.
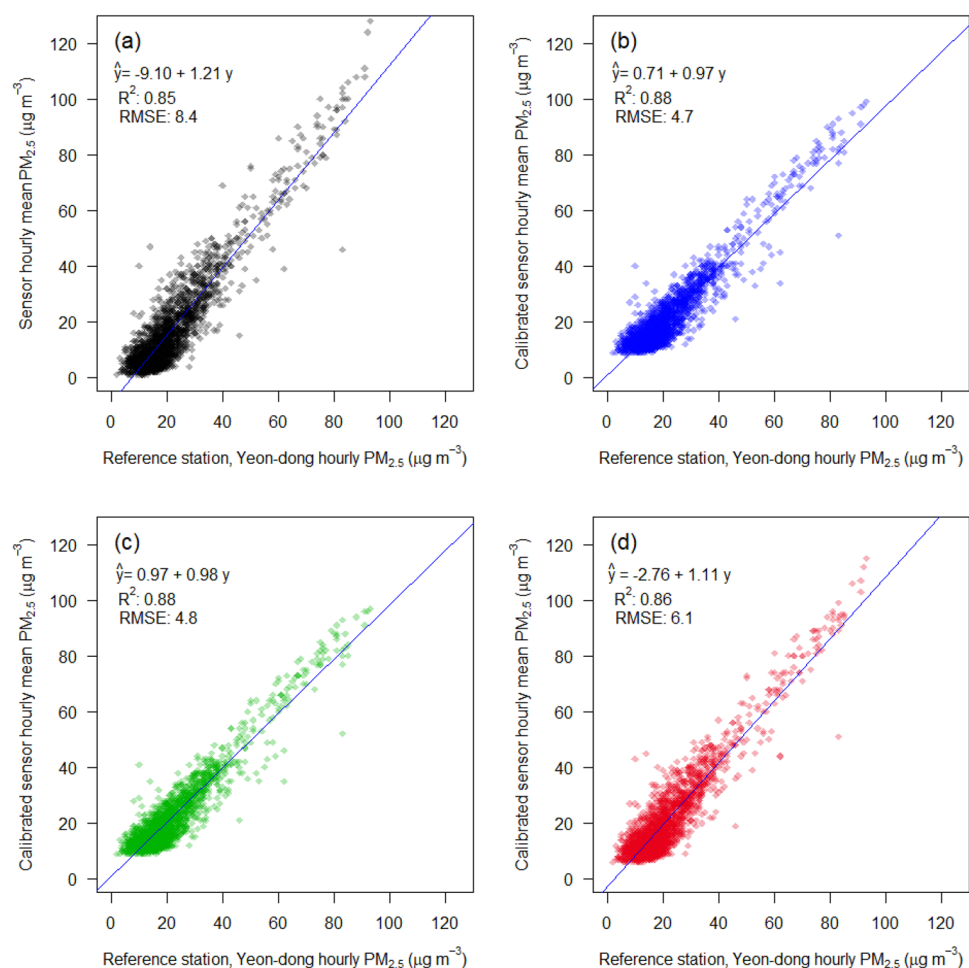
Figure 5. Scatterplots of hourly $PM_{2.5}$ mass concentration of the LCM against the hourly measurements of the reference station in Jeju, Korea: (**a**) Raw $PM_{2.5}$ data of the LCM; post-processed data according to the calibration method of the (**b**) Jeju, (**c**) Incheon, and (**d**) Singapore model.

All field measurement results over seven months in Jeju and 22 months in Singapore are summarized in Table 1. The average $PM_{2.5}$ concentration over the measurement period in Singapore (13.0 μg m^{-3}) is considerably less than that in Jeju (21.2 μg m^{-3}). The mean $PM_{2.5}$ concentration exhibits negative bias in Jeju and Singapore with 16.6 μg m^{-3} and 12.3 μg m^{-3}, respectively. Both field experiments in Jeju and Singapore show that the mean $PM_{2.5}$ concentration after calibration by the model (20.8–21.6 μg m^{-3} and 13.1–13.3 μg m^{-3}, respectively) is close to that of the reference station (21.2 μg m^{-3} and 13.0 μg m^{-3} respectively). The R^2 between the LCM and reference-station measurements is smaller in Singapore (0.51) than in Jeju (0.85), which may be due to the relatively low average concentration and relatively long distance between the reference station and the LCM in Singapore (Fig. S5). Regarding the results after calibration by the Jeju and Incheon models, RMSEs are nearly similar (4.7–4.8 μg m^{-3} in Jeju and 4.4 μg m^{-3} in Singapore) while the normalized error is smaller in Jeju (17%) than in Singapore (26%). The error of Jeju model is the lowest for field testing of Jeju and Singapore (4.7 μg m^{-3} and 17%; 4.4 μg m^{-3} and 26% respectively). However, the Jeju model has the lowest linear regression slope of 0.53 for field testing of Singapore, which means that the greater the ambient $PM_{2.5}$ concentration, the greater the RMSE by the model. For example, in the case of high $PM_{2.5}$ concentrations (in excess of 30 μg m^{-3}), RMSE of the Singapore model is the lowest (9.7 μg m^{-3}) while that of the Jeju model (12.3 μg m^{-3}) is still as high as raw data (12.5 μg m^{-3}).

Plots of the calibrated $PM_{2.5}$ mass concentrations for different RH levels, are shown in Fig. 6. The RH and TMP-adjusted results for high RH exhibit a reduced sensor bias compared to the raw data shown in Fig. 2; the linear regression slopes of 0.91–1.02 in Jeju show that the calibration method can efficiently correct the bias of the sensor, regardless of the RH levels.

Implications of modeling for regional calibration. The present study proposes a regional LCM calibration method that estimates $PM_{2.5}$, which is typically sensitive to weather, using a multivariate Tobit model with airport weather and air quality data. Instead of air conditioning to 20–23 °C TMP and 30–40% RH, this method predicts $PM_{2.5}$ concentration assuming that the LCM regulates TMP and RH at 21.5 °C and 35%, respectively. The validity of this airport weather-based calibration method was verified by constructing models for three regions, namely Incheon and Jeju in Korea at middle latitudes and Singapore at the equator, and complementary field measurements were conducted for several months in Jeju and Singapore.

The main argument against LCMs is that their accuracy is still not sufficiently high and is sensitive to the background concentration and regional environment, as also reported in the present study[16,26,32,34,35]. Although several studies have elaborated on the $PM_{2.5}$ hygroscopic properties and their influence on LCM performance[26,32,33], other studies have reported reduced TMP and RH effects on LCM performance during field calibration[30,52], which may be due to relatively short measurement periods, low average concentrations, and less pronounced TMP and RH variations in these regions. This study presents a novel approach for understanding regional differences by performing regional calibration using visibility-prediction models and LCM field testing in two different regions.

Considering that visibility is a simple indicator of air quality, Molnár et al.[39] showed that visibility-based PM hygroscopic growth is in good agreement with filter-based mass growth rate and can be applied to low-cost PM monitoring. Datta et al.[19] showed that a calibration equation can be applied to another LCM in the network of the region because TMP, RH, and $PM_{2.5}$ concentration trends in the region are similar, which is in line with the findings of this study, which showed that the Jeju model is most effective in field testing in Jeju. Zusman et al.[36] showed that regional calibration may increase LCM reliability because meteorological conditions and PM sources

| | Reference Station | Raw data | After calibration by each model | | |
			Jeju	Incheon	Singapore
Jeju					
Mean (μg m^{-3})	21.2	16.6	21.2	21.6	20.8
R^2		0.85	0.88	0.88	0.86
Slope		1.21	0.97	0.98	1.11
RMSE (μg m^{-3})		8.4	**4.7**	4.8	6.1
Normalized error (%)		33	**17**	**17**	22
Singapore					
Mean (μg m^{-3})	13.0	12.3	13.1	13.5	13.3
R^2		0.51	0.51	0.53	0.50
Slope		0.91	0.53	0.56	0.84
RMSE (μg m^{-3})		5.8 (12.5*)	**4.4** (12.3*)	**4.4** (11.7*)	5.5 (**9.7***)
Normalized error (%)		33	**26**	**26**	31

Table 1. Mean $PM_{2.5}$ concentration, coefficient of determination (R^2), slope, RMSE, and normalized error between $PM_{2.5}$ concentration measured by the LCM and measurements of the reference station over seven months (March to October 2019) in Jeju and 22 months (December 2020 to September 2022) in Singapore: each raw LCM value is compared with the respective calibrated values produced by three calibration methods developed by multivariate Tobit model using historical weather and air-quality data from Jeju, Incheon, and Singapore, respectively. Bold font indicates lowest errors. *In case of high $PM_{2.5}$ concentrations (in excess of 30 μg m^{-3} from the reference station in Singapore).
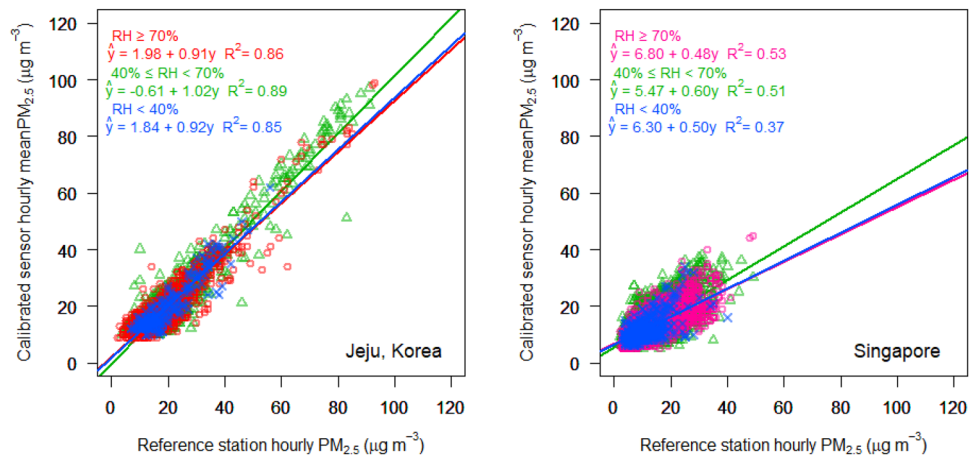
Figure 6. RH and TMP-adjusted $PM_{2.5}$ concentrations by the calibration model, namely Jeju model, against those of the reference station for the different RH ranges in Jeju and Singapore. Data under different RH levels (low, moderate, and high RH) are shown in blue, green, and red, respectively.

may differ from one region to another. Onal et al.[3] showed that the IoT Big Data framework and machine learning can identify regional climate differences from complex datasets. The regional calibration method using the Tobit model with multi-annual visibility, meteorological factors, and air quality data in this study can be supported by these studies because it reflects various local factors affecting LCM performance.

Routinely correcting the bias of LCMs using filter-based mass concentrations as a reference is an easy way to increase accuracy, but this is not always applicable, as in the case of variable PM concentrations/compositions and meteorological conditions. Such a correction is also not practically feasible because of the difficult acquisition of long-term observations in all regions.

As an attempt to focus on LCM air conditioning, similar to reference instruments, and regional calibration using a multivariate Tobit model, the significance of this study is as follows. First, we showed that there is abundant data for post-processing and generating a calibration model. The multivariate model generally requires as much data as possible; and many airports and air quality authorities possess decades of weather and $PM_{2.5}$ concentration data. We developed a regional calibration formula by comparing several years of airport visibility reported every hour with PM sensor measurements, showing how post-processing techniques can be applied to airport meteorological data and air quality measurements around the world. Second, we demonstrated that the calibration method can reflect local characteristics without requiring long-term field testing. The proposed method can reproduce an effect similar to that of creating empirical relations through long-term field experiments by building a visibility-prediction model using accumulated historical data. Third, the accuracy of PM monitoring, which is region-dependent, was quantified using a specific type of LCM instrument. By comparing two regions with different climate characteristics, located in the middle latitudes and equator respectively, we revealed the LCM requirements for calibration according to the local climate parameters for more reliable ambient air monitoring. LCMs still have limitations in terms of accuracy and reliability, but their potential is invaluable, when combined with advances in postprocessing methodologies[7,8,20].

East Asia has been experiencing increased $PM_{2.5}$ concentrations due to climate change and related stagnant atmospheric conditions[53]. Southeast Asia, in particular, currently experiences severe haze every few years and faces the major challenge of mitigating any damage from such climate crises[54,55]. Future investigations should focus on these high-impact pollution events, and the use of LCMs for this purpose should allow more communities to have easy access to air quality information. The proposed regional calibration of the postprocessing method can enhance the applicability of outdoor air quality monitoring using LCMs.

Methods

Data. To establish relationships between meteorological parameters and $PM_{2.5}$ concentration, airport weather observations and $PM_{2.5}$ concentration data were collected from two regions in Korea, Incheon and Jeju, and one in Singapore (Figs. S1 and S2)[56,57]. The data cover four years, from January 2015 to December 2018 at Incheon and Jeju airports, and two years and six months, from April 2020 to September 2022 at Changi Airport in Singapore. Airport observations consist of hourly wind direction and wind speed (WS), visibility, *'present weather'* (WX)[40], air temperature (TMP), and dew-point temperature (DPT)[58]. Relative humidity (RH) was calculated using the TMP–DPT relation equation[59]. The selected air-quality monitoring stations are situated five kilometers northeast of Incheon airport, three kilometers south of Jeju airport, and five kilometers west of Changi airport in Singapore respectively; in all cases, the $PM_{2.5}$ concentration was regularly monitored by reference instruments. The ranges of the meteorological parameters and $PM_{2.5}$ concentration are summarized in Table S1. The correlation coefficients between the variables and their scatter plots are shown in Fig. S3.

Modeling. The TMP, RH, WS, WX, and PM$_{2.5}$ concentration are used as explanatory variables to predict visibility, which is determined as follows:

$$Z_{x_i} = \frac{x_i - \overline{x_i}}{s_{x_i}} \tag{1}$$

where Z_{x_i} is the standardized explanatory variable of x_i, x_i represents the explanatory variable (i.e., meteorological parameters or PM concentration) for observation i, $\overline{x_i}$ is the mean value of the explanatory variable, and s_{x_i} is the standard deviation of x_i. Since visibility observations in airports have an upper limit of 9999 m, it has right-censored data characteristics; therefore, the Tobit model is utilized as an post-processing tool with vector generalized linear model (VGLM) function in the programming language R[60,61]. The Tobit model is expressed as follows:

$$y_i = \begin{cases} Z_{x_i}\beta + \varepsilon_i & (y_i < u) \\ u & (y_i \geq u) \end{cases} \tag{2}$$

where y_i is the airport visibility for observation i, Z_{x_i} represents the predictor variables for observation i, β is a vector of regression parameters, $\varepsilon_i \sim N(0, \sigma^2)$ is a random error for observation i, and u is the censoring value, that is 9999 m. The selected parameters that constitute the model and subsequent results are summarized in Tables S2, S3, and S4. The RH and PM$_{2.5}$ concentration have the greatest influence on the extinction coefficient; TMP and WS also has a significant effect on visibility[62,63]. The relationships between PM$_{2.5}$ concentration, and weather parameters have also been incorporated into the model[50].

Calibration equations. The predicted visibility from the above-mentioned model corresponds to the extinction coefficient in the equation below, which is derived from field measurements in various PM concentration ranges[46]:

$$\sigma_{ext_i}(RH_i) = 3.97 \times PM_{2.5_i} \times \left(1 + 8.8 \times \left(\frac{RH_i}{100}\right)^{9.7}\right) + 0.62 \times PM_{2.5_i} + 25 \tag{3}$$

where σ_{ext_i} is the extinction coefficient for observation i; and the visibility corresponding to σ_{ext_i} is expressed as follows[43]:

$$VIS_i = \frac{3.912}{\sigma_{ext_i}} \tag{4}$$

where VIS_i is visibility (km) for observation i. Subsequently, we can determine visibility from Eqs. (3) and (4) using RH and PM$_{2.5}$ concentration. For example, VIS$_2$ in Fig. 7 represents visibility at a PM$_{2.5}$ concentration of m$_2$ under low RH. If the LCM yields m$_2$ µg m^{-3} under high RH, we find the expected visibility variation (i.e., $\overline{DC'}$ in Fig. 7) by assuming that TMP and RH change to 21.5 °C and 35%, respectively, to calibrate the misread concentration. The expected visibility variation due to changes in TMP and RH was calculated using the model as follows:

$$\Delta VIS_{TMP_i} = \left(Z_{21.5 °C} - Z_{TMP_i}\right) \times \left(\beta_{TMP} + \beta_{TMP:PM_{2.5}} \times Z_{PM_{2.5_i}}\right) \tag{5}$$
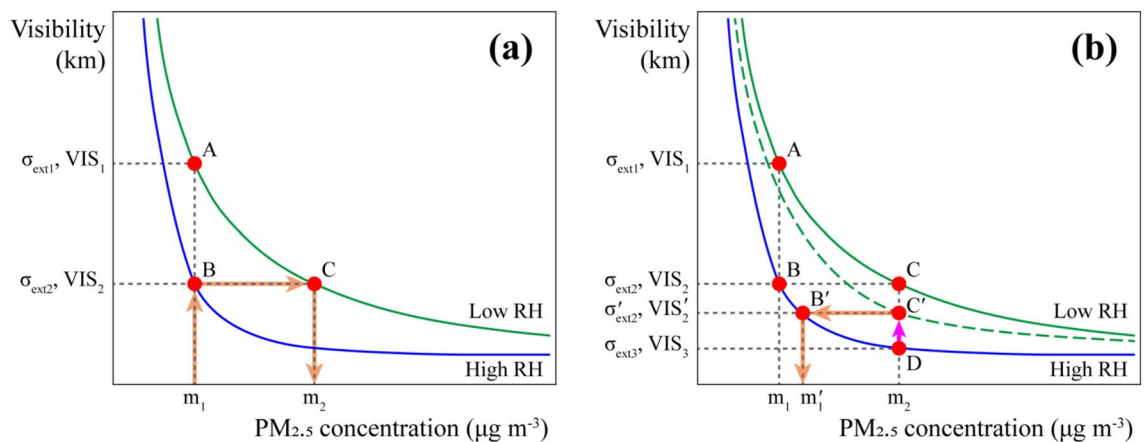
Figure 7. Schematic diagram of the proposed calibration method using visibility prediction by relative humidity (RH) air conditioning: (**a**) The graph shows the process of misreading m$_1$ as m$_2$ under high RH. (**b**) The arrows indicate how to correct the misread PM$_{2.5}$ concentration on the equivalent curve of visibility, PM$_{2.5}$ concentration, and RH, by predicting the change in visibility while assuming sample air conditioning in a LCM.

$$\Delta VIS_{RH_i} = \left(Z_{35\%} - Z_{RH_i}\right) \times \left(\beta_{RH} + \beta_{RH:PM_{2.5}} \times Z_{PM_{2.5_i}}\right) \tag{6}$$

$$VIS'_i = VIS_i + \Delta VIS_{TMP_i} + \Delta VIS_{RH_i} \tag{7}$$

where $VIS_i\prime$ is the visibility predicted by the change in TMP and RH to 21.5 °C and 35%, respectively. Subsequently, a new extinction coefficient, $\sigma_{ext\,i}\prime$, was obtained using Eq. (4). Finally, the calibrated concentration, $PM_{2.5_i}\prime$, was calculated using Eq. (3). It is the calibrated concentration m'_1 (Fig. 7b) by air-conditioning TMP and RH to 21.5 °C and 35%, respectively. According to these equations, increased visibility changes due to air conditioning lead to more pronounced changes in the extinction coefficient, eventually resulting in high $PM_{2.5}$ concentration in need of calibration. The empirical formula consists of RH and $PM_{2.5}$ concentration, but the model enhances prediction performance by considering the effects of TMP and RH as well.

Figure 7 shows the process of correcting a misread $PM_{2.5}$ concentration using visibility prediction under high RH. The two curves depict empirical relationships between RH, $PM_{2.5}$, and visibility under high and low RH. The orange arrows in Fig. 7a indicates the sequence by which the LCM misreads the actual $PM_{2.5}$ concentration, m_1, as m_2. Low $PM_{2.5}$ concentration of m_1 corresponds to high visibility of VIS_1 under low RH (point A). Under high RH, the m_1 concentration has a higher extinction coefficient, σ_{ext_2}, resulting in lower visibility of VIS_2, which results in the LCM calibrated according to low RH by the manufacturer indicating a higher $PM_{2.5}$ concentration of m_2 (point C). Conversely, when the LCM indicates a $PM_{2.5}$ concentration of m_2 under high RH conditions, it should be calibrated to m_1; hence, an LCM bias by RH can be corrected using the empirical relations if it is already known in the monitoring area; however, it is unknown in most cases.

Figure 7b shows the calibration process. From the empirical relation, the difference in visibility between high and low RH at the $PM_{2.5}$ concentration of m_2 is $\overline{DC}$, and the difference in $PM_{2.5}$ concentration between high and low RH at VIS_2 is $\overline{BC}$. However, because the empirical relation varies depending on regional and environmental conditions, it is impractical to construct every empirical relation in all regions where LCMs operate. Here, we calculate the anticipated visibility difference ($\overline{DC\prime}$) between high and low RH at $PM_{2.5}$ concentration of m_2 using the model. Subsequently, the difference in $PM_{2.5}$ concentration before and after calibration is $\overline{B\prime C\prime}$, which is dependent on $\overline{DC\prime}$. For example, in the Jeju case on March 29, 2019, 06:00 KST (i.e., Korea standard time), $\overline{DC}$ from Fig. 7b was calculated as 4.0 km according to the empirical relation, and accordingly the $PM_{2.5}$ concentration was calibrated as 52 μg m^{-3} (m_1). Conversely, $\overline{DC\prime}$ is calculated as 3.0 km by the model, assuming that the sampled air in the LCM is conditioned at 21.5 °C TMP and 35% RH, thereby producing a final calibrated value of 57 μg m^{-3} (m'_1). (Table S6).

As described above, the proposed calibration method can be applied in a region where the empirical relation is unknown, but has sufficient historical weather and air quality data for predicting visibility.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References
1. Moldovan, D., Cioara, T., Anghel, I. & Salomie, I. in *13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP),* 147–154 (2017).
2. Praveen Kumar, D., Amgoth, T. & Annavarapu, C. S. R. Machine learning algorithms for wireless sensor networks: A survey. *Inf. Fusion* **49**, 1–25. https://doi.org/10.1016/j.inffus.2018.09.013 (2019).
3. Onal, A. C., Sezer, O. B., Ozbayoglu, M. & Dogdu, E. in *IEEE International Conference on Big Data (Big Data)* 2037–2046 (2017).
4. Worland, S. C., Farmer, W. H. & Kiang, J. E. Improving predictions of hydrological low-flow indices in ungaged basins using machine learning. *Environ. Model. Softw.* **101**, 169–182. https://doi.org/10.1016/j.envsoft.2017.12.021 (2018).
5. Goldstein, A. *et al.* Applying machine learning on sensor data for irrigation recommendations: Revealing the agronomist's tacit knowledge. *Precis. Agric.* **19**, 421–444. https://doi.org/10.1007/s11119-017-9527-4 (2018).
6. Sun, A. Y. & Scanlon, B. R. How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environ. Res. Lett.* **14**, 073001. https://doi.org/10.1088/1748-9326/ab1b7d (2019).
7. Lim, C. C. *et al.* Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environ. Int.* **131**, 105022. https://doi.org/10.1016/j.envint.2019.105022 (2019).
8. Zimmerman, N. *et al.* A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* **11**, 291–313. https://doi.org/10.5194/amt-11-291-2018 (2018).
9. Lee, M. *et al.* Forecasting air quality in Taiwan by using machine learning. *Sci. Rep.* **10**, 4153. https://doi.org/10.1038/s41598-020-61151-7 (2020).
10. Lepeule, J., Laden, F., Dockery, D. & Schwartz, J. Chronic exposure to fine particles and mortality: An extended follow-up of the Harvard Six Cities study from 1974 to 2009. *Environ. Health Perspect.* **120**, 965–970 (2012).
11. Cohen, A. J. *et al.* Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: An analysis of data from the Global Burden of Diseases Study 2015. *Lancet* **389**, 1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6 (2017).
12. Burnett, R. *et al.* Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci.* **115**, 9592. https://doi.org/10.1073/pnas.1803222115 (2018).
13. Snyder, E. G. *et al.* The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* **47**, 11369–11377. https://doi.org/10.1021/es4022602 (2013).
14. Hagler, G., Solomon, P. & Hunt, S. *New Technology for Low-cost, Real-time Air Monitoring* (EM: Air and Waste Management Association's Magazine for Environmental Managers; Air & Waste Management Association, 2014).
15. Kumar, P. *et al.* The rise of low-cost sensing for managing air pollution in cities. *Environ. Int.* **75**, 199–205. https://doi.org/10.1016/j.envint.2014.11.019 (2015).

16. Gao, M., Cao, J. & Seto, E. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *Environ. Pollut.* **199**, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013 (2015).
17. Kelly, K. E. *et al.* Ambient and laboratory evaluation of a low-cost particulate matter sensor. *Environ. Pollut.* **221**, 491–500. https://doi.org/10.1016/j.envpol.2016.12.039 (2017).
18. Sayahi, T., Butterfield, A. & Kelly, K. E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **245**, 932–940. https://doi.org/10.1016/j.envpol.2018.11.065 (2019).
19. Datta, A. *et al.* Statistical field calibration of a low-cost PM25 monitoring network in Baltimore. *Atmos. Environ.* **242**, 117761. https://doi.org/10.1016/j.atmosenv.2020.117761 (2020).
20. Zamora, M. L., Rice, J. & Koehler, K. One year evaluation of three low-cost PM2.5 monitors. *Atmos. Environ.* **235**, 117615. https://doi.org/10.1016/j.atmosenv.2020.117615 (2020).
21. Hankey, S. & Marshall, J. D. Land use regression models of on-road particulate air pollution (Particle number, black carbon, PM2.5, particle size) using mobile monitoring. *Environ. Sci. Technol.* **49**, 9194–9202. https://doi.org/10.1021/acs.est.5b01209 (2015).
22. Bi, J., Wildani, A., Chang, H. H. & Liu, Y. Incorporating low-cost sensor measurements into high-resolution PM2.5 modeling at a large spatial scale. *Environ. Sci. Technol.* **54**, 2152–2162. https://doi.org/10.1021/acs.est.9b06046 (2020).
23. Rai, A. C. *et al.* End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci. Total Environ.* **607–608**, 691–705. https://doi.org/10.1016/j.scitotenv.2017.06.266 (2017).
24. Castell, N. *et al.* Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?. *Environ. Int.* **99**, 293–302. https://doi.org/10.1016/j.envint.2016.12.007 (2017).
25. Mukherjee, A., Stanton, L. G., Graham, A. R. & Roberts, P. T. Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama Valley of California. *Sensors* https://doi.org/10.3390/s17081805 (2017).
26. Jayaratne, R., Liu, X., Thai, P., Dunbabin, M. & Morawska, L. The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmos. Meas. Tech.* **11**, 4883–4890 (2018).
27. Chow, J. C. & Watson, J. G. Guideline on speciated particulate monitoring, in *Report Prepared for US Environmental Protection Agency, Research Triangle Park, NC, by Desert Research Institute, Reno, NV* (1998).
28. Carlton, A. G. & Teitz, A. Design of a cost-effective weighing facility for PM2.5 quality assurance. *J. Air Waste Manag. Assoc.* **52**, 506–510. https://doi.org/10.1080/10473289.2002.10470802 (2002).
29. U.S.EPA. Quality assurance guidance document 2.12. (2016).
30. Jiao, W. *et al.* Community air sensor network (CAIRSENSE) project: Evaluation of low-cost sensor performance in a suburban environment in the southeastern United States. *Atmos. Meas. Tech.* **9**, 5281–5292. https://doi.org/10.5194/amt-9-5281-2016 (2016).
31. Bulot, F. M. J. *et al.* Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Sci. Rep.* **9**, 7497. https://doi.org/10.1038/s41598-019-43716-3 (2019).
32. Crilley, L. R. *et al.* Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmos. Meas. Tech.* **11**, 709–720. https://doi.org/10.5194/amt-11-709-2018 (2018).
33. Zamora, M. L. *et al.* Field and laboratory evaluations of the low-cost plantower particulate matter sensor. *Environ. Sci. Technol.* **53**, 838–849. https://doi.org/10.1021/acs.est.8b05174 (2019).
34. Johnson, K. K., Bergin, M. H., Russell, A. G. & Hagler, G. S. Field test of several low-cost particulate matter sensors in high and low concentration urban environments. *Aerosol Air Qual. Res.* **18**, 565–578 (2018).
35. Zheng, T. *et al.* Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. *Atmos. Meas. Tech.* **11**, 4823–4846. https://doi.org/10.5194/amt-11-4823-2018 (2018).
36. Zusman, M. *et al.* Calibration of low-cost particulate matter sensors: Model development for a multi-city epidemiological study. *Environ. Int.* **134**, 105329. https://doi.org/10.1016/j.envint.2019.105329 (2020).
37. Hyslop, N. P. Impaired visibility: The air pollution people see. *Atmos. Environ.* **43**, 182–195. https://doi.org/10.1016/j.atmosenv.2008.09.067 (2009).
38. Singh, A., Bloss, W. J. & Pope, F. D. 60 years of UK visibility measurements: Impact of meteorology and atmospheric pollutants on visibility. *Atmos. Chem. Phys.* **17**, 2085–2101. https://doi.org/10.5194/acp-17-2085-2017 (2017).
39. Molnár, A., Imre, K., Ferenczi, Z., Kiss, G. & Gelencsér, A. Aerosol hygroscopicity: Hygroscopic growth proxy based on visibility for low-cost PM monitoring. *Atmos. Res.* **236**, 104815. https://doi.org/10.1016/j.atmosres.2019.104815 (2020).
40. WMO. *Guide to Meteorological Observing and Information Distribution Systems for Aviation Weather Services.* Vol. WMO-No. 731 (World Meteorological Organization, 2014).
41. WMO. *Aerodrome Reports and Forecasts: A Users' Handbook to the Codes.* Vol. WMO-No. 782 (World Meteorological Organization, 2014).
42. WMO. *Manual on Codes: International Codes.* 2011 edn, Vol. WMO-No. 306 (World Meteorological Organization, 2017).
43. Koschmieder, H. Theorie der horizontalen Sichtweite. *Beitr. Phys. Freien Atmos.* **12**, 33–53 (1924).
44. Pan, X. L. *et al.* Observational study of influence of aerosol hygroscopic growth on scattering coefficient over rural area near Beijing mega-city. *Atmos. Chem. Phys.* **9**, 7519–7530. https://doi.org/10.5194/acp-9-7519-2009 (2009).
45. Liu, X. G. *et al.* Formation and evolution mechanism of regional haze: A case study in the megacity Beijing, China. *Atmos. Chem. Phys.* **13**, 4501–4514. https://doi.org/10.5194/acp-13-4501-2013 (2013).
46. Liu, X. *et al.* Increase of aerosol scattering by hygroscopic growth: Observation, modeling, and implications on visibility. *Atmos. Res.* **132–133**, 91–101. https://doi.org/10.1016/j.atmosres.2013.04.007 (2013).
47. Xia, C. *et al.* Observational study of aerosol hygroscopic growth on scattering coefficient in Beijing: A case study in March of 2018. *Sci. Total Environ.* **685**, 239–247. https://doi.org/10.1016/j.scitotenv.2019.05.283 (2019).
48. Zhao, P., Ding, J., Du, X. & Su, J. High time-resolution measurement of light scattering hygroscopic growth factor in Beijing: A novel method for high relative humidity conditions. *Atmos. Environ.* **215**, 116912. https://doi.org/10.1016/j.atmosenv.2019.116912 (2019).
49. Gobeli, D., Schloesser, H. & Pottberg, T. in *The Air & Waste Management Association (A&WMA) Conference, Kansas City, MO.* (Citeseer).
50. Won, W.-S. *et al.* Impact of fine particulate matter on visibility at incheon international airport, South Korea. *Aerosol Air Qual. Res.* **20**, 1048–1061. https://doi.org/10.4209/aaqr.2019.03.0106 (2020).
51. Lee, S.-Y., Gan, C. & Chew, B. N. Visibility deterioration and hygroscopic growth of biomass burning aerosols over a tropical coastal city: A case study over Singapore's airport. *Atmos. Sci. Lett.* **17**, 624–629. https://doi.org/10.1002/asl.712 (2016).
52. Holstius, D. M., Pillarisetti, A., Smith, K. R. & Seto, E. Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmos. Meas. Tech.* **7**, 1121–1131. https://doi.org/10.5194/amt-7-1121-2014 (2014).
53. Kim, H. C. *et al.* Recent increase of surface particulate matter concentrations in the Seoul metropolitan area, Korea. *Sci. Rep.* **7**, 4710. https://doi.org/10.1038/s41598-017-05092-8 (2017).
54. Lee, J. S. H. *et al.* Toward clearer skies: Challenges in regulating transboundary haze in Southeast Asia. *Environ. Sci. Policy* **55**, 87–95. https://doi.org/10.1016/j.envsci.2015.09.008 (2016).
55. Tacconi, L. Preventing fires and haze in Southeast Asia. *Nat. Clim. Chang.* **6**, 640–643 (2016).
56. KEC. *Statistics Information: Air Quality Data Retrieve*, https://www.airkorea.or.kr/web/pastSearch (2020).
57. Data.gov.sg. *Retrieve the Latest PM2.5 Information*, https://data.gov.sg/dataset/pm2-5 (2020).
58. Washington, W. *The World Area Forecast System (WAFS) Internet File Service (WIFS) Users Guide* (2018).

59. Lawrence, M. G. The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bull. Am. Meteorol. Soc.* **86**, 225–234. https://doi.org/10.1175/BAMS-86-2-225 (2005).
60. Tobin, J. Estimation of relationships for limited dependent variables. *Econometrica* **26**, 24–36. https://doi.org/10.2307/1907382 (1958).
61. Yee, T. W. VGAM: Vector generalized linear and additive models; 2018. *Available from: R Package Version 1.0–6* (URL: http://CRAN.R-project.org/package=VGAM) (2018).
62. Tsai, Y. I. Atmospheric visibility trends in an urban area in Taiwan 1961–2003. *Atmos. Environ.* **39**, 5555–5567. https://doi.org/10.1016/j.atmosenv.2005.06.012 (2005).
63. Lin, M. *et al.* Regression analyses between recent air quality and visibility changes in megacities at four haze regions in China. *Aerosol Air Qual. Res.* **12**, 1049–1061. https://doi.org/10.4209/aaqr.2011.11.0220 (2012).

Acknowledgements

Author contributions

W.W., P.S., and Y.Y. designed the research, W.W. and R.O. built the model, R.O., W.L analyzed the data and provided comments on methods, W.W. and J.N. wrote the paper, J.L., P.S., and Y.Y. provided discussions. Everyone edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-023-40468-z.

Correspondence and requests for materials should be addressed to P.-C.S. or Y.-J.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.