Mitigating the Impact of Humidity on Low-Cost PM Sensors

Martina Casari[1], Laura Po[1]

[1]UNIMORE, Dipartimento di Ingegneria "Enzo Ferrari", via P. Vivarelli, Modena, 41125, Italy

Abstract

This preliminary study, conducted in Italy, aims to investigate the potential of growth functions and multi-layer perceptron neural networks (MLP NN) in reducing the impact of humidity on low-cost particulate matter (PM) sensors, with a focus on the sustainability of low-cost sensors compared to reference stations. All over the world, low-cost sensors are increasingly being used for air quality monitoring due to their cost-effectiveness and portability. However, low-cost sensors are susceptible to high humidity, which can lead to inaccurate measurements due to their hygroscopic property. This issue is particularly relevant in Italy, where many cities such as Rome, Milan, Naples, and Turin experience high mean relative humidity levels (>70%) for most months of the year. To improve data quality and gain useful data for quantitative analysis, techniques must be developed to reduce the impact of humidity on the final data. The sensors used in this study were placed in proximity to a reference station, solely for validation purposes in the case of corrective functions and involved in the training phase in the case of MLP NN.

Keywords

Air quality, PM, Relative humidity, Low-cost sensor, Machine learning

1. Introduction

The issue of air pollution is becoming increasingly prevalent due to various factors such as urbanization, industrial activities, and transportation [1]. It has been widely acknowledged that air pollution can negatively impact public health and contribute to global warming, acid rain, and environmental degradation [2]. As a result, there is a pressing need for sustainable solutions to combat air pollution. One such solution is the use of low-cost sensors (LCS) for air quality monitoring.

These sensors have the potential to make air quality monitoring more accessible and widespread, particularly in developing countries or areas with limited resources [3]. The affordability and portability of these sensors have opened up new possibilities for citizen science projects and community-led monitoring efforts. Additionally, low-cost sensors can provide real-time data, enabling prompt actions to be taken to address air pollution hotspots. However, the drawback of LCSs is their restricted technology, which makes quantitative evaluation difficult due to the fact that these sensors are very sensitive to environmental factors, compared to reference stations [4], [5].

When the relative humidity (RH) exceeds a certain threshold, the water in the air can be detected by the sensor, leading to inaccurate measurements of particulate matter mass concentration due to the properties of particles and the presence of condensed water droplets. RH starts to significantly impact the concentration detected when it exceeds 80-85%, although the impact may start lower for some sensors [6]. This limitation is particularly relevant for low-cost PM sensors, which utilize laser scattering technology to measure particle concentration.

The purpose of this paper is to explore the feasibility of using growth functions and multi-layer perceptron neural networks to mitigate the effects of humidity on low-cost particulate matter sensors. The study is preliminary and is specifically focused on evaluating the sustainability of low-cost sensors in comparison to reference stations. Experiments are carried out in Italian cities.

The paper is divided into five sections, each covering a different aspect of the study. In Section 2, laser scattering sensors are introduced and their operation is explained. Section 3 presents the use of growth functions as a method for mitigating the effects of humidity on low-cost sensors. The section describes how these functions can be used to correct the presence of water in the air. In Section 4, the use of multi-layer perceptron neural networks as an alternative method for correcting humidity effects is discussed. Section 5 presents the concept of a cooperative technique, which is a method proposed to combine data from the application of the corrective function as input to the MLP NN. Finally, in Section 6, evaluation metrics are presented. This section outlines the various metrics that were used to evaluate the performance of the different correction methods, and how they can be used to assess the accuracy of low-cost

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.

martina.casari@unimore.it (M. Casari); laura.po@unimore.it (L. Po)

0000-0003-0406-3036 (M. Casari); 0000-0002-3345-176X (L. Po)

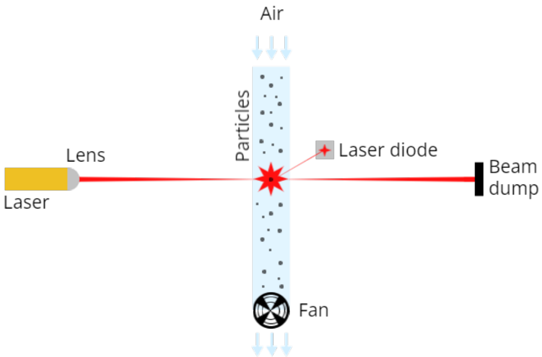

Figure 1: The figure illustrates the operation of the Sensirion SPS30 low-cost air quality sensor, which uses laser scattering to detect particulate matter. The sensor operates by pulling in ambient air through a blower fan, which then flows through a laser diode channel. Particles in the air cause the laser light to scatter and hit the light sensor. The non-scattered laser light gets absorbed into a special surface to avoid any additional scattering. Finally, the air is exhausted from the unit.

sensors.

2. Laser Scattering Sensors

Sensor scattering is a technique (shown in Figure 1) commonly utilized to measure the concentration of PM in the air. This technique involves the use of lasers to scatter light off particles in the air, which are then detected by a sensor. The amount of scattered light is proportional to the number of particles in the air, allowing for an estimation of the PM concentration. However, this technique is limited by the sensitivity and accuracy of the sensor, particularly in the case of low-cost sensors.

Low-cost sensors lacking a drying function or installed in humid environments, such as coastal areas, are more susceptible to humidity interference. Particles with hygroscopic properties absorb water from the air [7], resulting in larger particle sizes and an increase in light scattering within the sensor [8]. This leads to overestimated PM concentration levels (Figure 2) for high levels of RH.

It is important to note that pollution and humidity are not directly correlated and that water vapor is not harmful to human health. Therefore, to accurately measure pollution levels, it is essential to clean the data from this artifact. The EU air quality standards [9] and the WHO air quality guidelines [10] and other governmental organizations measure pollution impact based on the dry PM concentration.

Calibrating a low-cost sensor using data from a reference station and humidity levels is possible if a reference station is available. However, it may not work well for PM sensors because the problem is highly localized, with different types of pollutants having different hygroscopic properties based on surrounding environmental emissions [11], [12], [13]. Placing the sensor near a reference station, even if it is far from the final location of the LCS, is another possibility for calibration, but it may also result in poor data quality. Therefore, the proposed methods aim to clean data in a location-agnostic manner.

In addition to the previously proposed methods that utilize reference stations [14], two potential techniques were proposed for cleaning low-cost sensor data in a location-agnostic manner. The first technique involves using a corrective function to reduce the correlation between PM concentration and humidity [15]. By reducing the correlation, the concentration level detected can be reduced by a particular factor. The second technique involves training a multi-layer perceptron neural network using data from a reference station, taking into account not only the relative humidity but also other atmospheric variables, such as pollutants and meteorological factors. This could enable the NN to learn the relationship between humidity, pollutants, and the growth of PM concentration. Although the network is trained using data from a reference station, the learned function is generic and not location-specific. To make this technique work, the network needs to be exposed to multiple contexts to be used in locations far from the training site.

3. Growth functions

High relative humidity causes low-cost air particulate matter sensors to measure higher values than professional sensors, due to the condensation effect and hygroscopic properties of the particles. To correct this effect, a growth function that estimates the increase in PM values due to humidity can be applied [15]. In the context of air quality monitoring, a growth function is used to estimate the increase in particulate matter concentration due to the absorption of humidity by the particles. Specifically, the growth function is used to estimate the amount by which the concentration of wet particulate matter exceeds the concentration of dry particulate matter at a given relative humidity level.

The growth function, given the level of the RH, returns the coefficient to use in the correction. Dividing the PM concentration detected by the low-cost sensor, which we define as PM_{wet} is returned the PM concentration without the humidity contribution, defined as PM_{dry}. See equation 1.

$$PM_{dry} = PM_{wet}/gf(RH) \tag{1}$$

The use of growth functions for humidity correction can be implemented over the entire dataset or with a

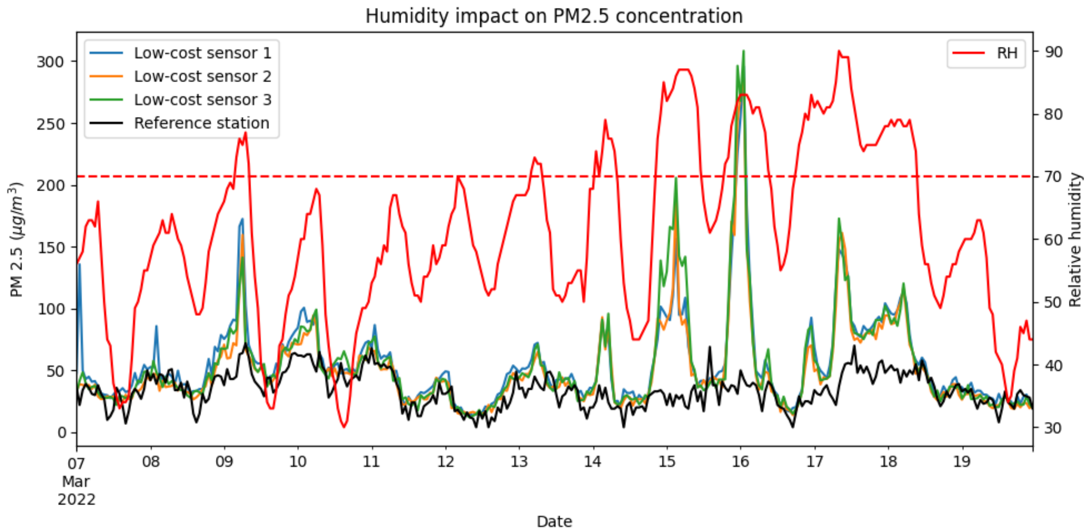

Figure 2: The figure illustrates the impact of relative humidity on the detection of PM concentration by low-cost sensors. The dotted line represents an RH threshold of 70%, which is used to distinguish between the RH levels where the effect is less pronounced, based on the LCS Sensirion SPS30.

threshold approach, depending on the sensor and environmental factors. Applying the growth function across the entire dataset may result in coefficients being applied to low RH levels, which can negatively impact already corrected data. Therefore, a threshold approach is preferred, whereby the growth function is only applied to RH values above the chosen threshold. The threshold level can be determined based on the specific sensor being used or by taking into consideration environmental knowledge. A commonly used threshold is around 70%.

Different corrective functions have been developed to address the problem of humidity affecting PM concentration measurements. Soneja et al. [16] research suggests that there is an overestimation bias that becomes significant at around 75% relative humidity, while an underestimation bias exists at very low RH levels (below 30%). To address this issue, they have proposed humidity adjustment equations that cover the entire RH range.

In this paper [15], the author proposes a comprehensive approach to correct nephelometric [1] PM for humidity-related bias. The paper starts with an overview of different sources that explain the principle behind the hygroscopic growth of particulates [17], [18].

The author proposes a new approach, named "combo", described in 2, to address humidity-related bias in PM measurements. This approach wants to combine some of the existing methods.

[1]Nephelometry is a method used to measure the concentration of suspended particles in a liquid or gas.

$$gf_{combo} = 1 + \alpha \cdot \frac{rh^2}{(1-rh)^{\beta}} \tag{2}$$

The growth functions can be customized to fit measured data by selecting appropriate values for the parameters α and β. A key feature of all growth functions is that they have a value of 1 when the relative humidity is 0 and a significantly larger value as the relative humidity approaches 100%.

3.1. Parameters optimization

One approach to choosing the growth function parameters is to use reference station values as a ground truth. Different α and β parameters are chosen within a certain range and the parameters that best improve the PM detected compared to the data detected by the reference station are selected, as in [8] and [19]. However, this approach links the parameters to the specific location of the reference station and is limited to the period studied. This approach is only useful if the low-cost sensor is co-located and fixed near the reference station and never moved. In this way, it is possible to update the parameters over time and use the reference station as ground truth, [20]. However, this approach involves the optimization of the growth function parameters concerning the reference station values, which may not be feasible in all situations.

To address this limitation, an alternative approach is to select growth function parameters that result in the lowest correlation between corrected PM concentrations and RH [15]. This method does not require optimization of the growth function parameters concerning reference station values and may be more practical in certain situations where a reference station is not available or the low-cost sensor is not co-located with a reference station.

In this approach, it is necessary to have data on the PM concentration as well as the relative humidity detected in the same location and at the same time. However, since the accuracy of the sensors used is lower than that of a reference station, it is always better to preprocess the initial data and remove anomalies using typical anomaly removal methods. It is also important to note that this approach works on the assumption that relative humidity and PM concentration are not strongly correlated, as shown by the lack of correlation observed in reference station data.

4. Neural network

One alternative that we propose as a solution is to use a multi-layer perceptron neural network to model and generalize the relationship between relative humidity and PM concentration growth. To achieve this, the neural network is trained on a dataset containing input-output pairs of RH, PM concentrations, and other relevant variables like meteorological and atmospheric variables. Once trained, the neural network can be used to correct PM concentrations. This method has the advantage of being able to capture complex, nonlinear relationships between RH and PM concentrations and has the potential to generalize well to different locations and periods. However, it requires a significant amount of high-quality training data and careful tuning of the network architecture.

To effectively generalize the relationship between RH and PM hygroscopic growth using a neural network, it is necessary to feed the network with additional variables beyond just RH and PM concentration data. Hygroscopic growth is a process that occurs as water vapor accumulates on the surface of aerosol particles with increasing relative humidity. The extent to which this process occurs depends on the chemical composition of the particles, which can vary widely in time and space [21].

Therefore, to improve the neural network's ability to accurately correct PM concentrations and generalize its predictions to new contexts, additional variables that capture information about the chemical composition of the particles should be included as inputs to the network. However, these variables are often not available at the sensor location and must be obtained from online resources like Copernicus [2].

[2]Copernicus is a European Union Earth observation and monitoring program that provides free and open access data (www.copernicus.eu)

In training the network, data from one or more sensors located near a reference station can be used, including PM concentration, meteorological and atmospheric variables, and the additional variables as inputs, with the reference station as the output. It's critical to ensure that the additional variables used are also available in other locations where low-cost sensors are placed. The elemental analysis is the most crucial additional variable, describing the concentration of each element present. To improve the network's ability to generalize and correct PM concentrations accurately, the network should learn the correlation between RH and PM growth in different pollutant contexts, as the hygroscopic properties are dependent on the specific pollutants present at the time of detection.

It is common in the literature to find studies that aim to calibrate low-cost PM sensors against reference stations. In addition to using an MLP NN, there may be other suggestions in the literature [22]. Some possible methods include regression analysis, decision tree models, and support vector machines. Each method has its advantages and limitations, and the choice of method depends on the specific application and data available.

We believe that MLP NN has the potential to correct PM concentration levels more effectively than other methods. To achieve this, a wide range of scenarios, including those that the sensors are likely to encounter, and different periods should be presented to the network during the training phase. This will improve the network's ability to generalize and correct PM concentration levels accurately, even for sensors located far from the reference station.

5. Cooperative techniques

One alternative and potential way to improve the accuracy of PM concentration measurement is to combine different methods. Preprocessing techniques can be applied to eliminate obvious anomalies in the raw data before applying the growth function. The resulting corrected data can then be used as input to train a neural network, which can further improve the accuracy of the PM concentration measurements. This approach has the potential to benefit from the strengths of each method, leading to more accurate and reliable results. However, the success of this approach depends on the quality and consistency of the data, as well as the effectiveness of the preprocessing.

6. Evaluation metrics

The wide range of possibilities for calibrating sensor devices makes it challenging to assess their performance and suitability for specific applications, especially since

Table 1
Recommended Performance Metrics and Target Values for PM2.5 Air Sensors

Performance Metric	Target Value
Standard Deviation (SD) or Coefficient of Variation (CV)	$SD \leq 5\mu g/m^3$ or $CV \leq 30\%$
Slope Intercept (b)	1.0 ± 0.35 $-5 \leq b \leq 5\mu g/m^3$
Coefficient of Determination (R^2)	$R^2 \geq 0.70$
Root Mean Square Error (RMSE) or Normalized Root Mean Square Error (NRMSE)	$RMSE \leq 7\mu g/m^3$ or $NRMSE \leq 30\%$

data from these devices are often made public and used for monitoring pollutant levels. To address this issue, the U.S. EPA has proposed guidelines [23], containing metrics of Table 1, for evaluating the data quality of low-cost sensors, which can be useful for non-regulatory purposes such as identifying local air quality trends and hotspots, promoting environmental awareness, and providing supplemental monitoring.

The goal of the report is to establish a consistent set of testing protocols, metrics, and target values for evaluating the performance of PM2.5 air sensors in outdoor, fixed-site environments specifically for non-regulatory supplemental and informational monitoring (NSIM). The metrics proposed can be used to assess the performance of these methods and guide future developments in this field.

7. Conclusion

In conclusion, it is expected that correction functions based on hygroscopic growth factor can be applied in locations where reference stations are not available. However, a neural network trained in various contexts may still be superior. Therefore, the winning approach may eventually come from a combination of these two techniques. Future work will focus on exploring the potential of these methods in mitigating the effects of humidity on low-cost PM sensors and improving the accuracy of their measurements.

References

[1] HEI, State of global air 2020, special report, Health Effects Institute (2020).

[2] I. Manisalidis, E. Stavropoulou, A. Stavropoulos, E. Bezirtzoglou, Environmental and health impacts of air pollution: a review, Frontiers in public health (2020) 14.

[3] S. Bishop, How low-cost air sensors empower economically developing nations to improve their air quality monitoring infrastructure, 2021. URL: https://www.clarity.io/blog.

[4] X. Liu, R. Jayaratne, P. Thai, T. Kuhn, I. Zing, B. Christensen, R. Lamont, M. Dunbabin, S. Zhu, J. Gao, et al., Low-cost sensors as an alternative for long-term air quality monitoring, Environmental research 185 (2020) 109438.

[5] D. H. Hagan, J. H. Kroll, Assessing the accuracy of low-cost optical particle sensors using a physics-based approach, Atmospheric Measurement Techniques (2020). doi:10.5194/amt-13-6343-2020.

[6] M. R. Giordano, C. Malings, S. N. Pandis, A. A. Presto, V. McNeill, D. M. Westervelt, M. Beekmann, R. Subramanian, From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors, Journal of Aerosol Science 158 (2021) 105833.

[7] A. K. Lee, T. Ling, C. K. Chan, Understanding hygroscopic growth and phase transformation of aerosols using single particle raman spectroscopy in an electrodynamic balance, Faraday Discussions 137 (2008) 245–263.

[8] R. Jayaratne, X. Liu, P. Thai, M. Dunbabin, L. Morawska, The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog, Atmospheric Measurement Techniques 11 (2018) 4883–4890.

[9] EEA, Eu air quality directives, 2021. URL: https://www.eea.europa.eu/themes/air/air-quality-concentrations/air-quality-standards.

[10] W. H. Organization, et al., Who global air quality guidelines: particulate matter (pm2. 5 and pm10), ozone, nitrogen dioxide, sulfur dioxide and carbon

monoxide: executive summary, World Health Organization (2021).

[11] T. Wu, Z. Li, J. Chen, Y. Wang, H. Wu, X. Jin, C. Liang, S. Li, W. Wang, M. Cribb, Hygroscopicity of different types of aerosol particles: Case studies using multi-instrument data in megacity beijing, china, Remote Sensing 12 (2020) 785.

[12] U. Lohmann, Aerosol-cloud interactions and their radiative forcing, Encyclopedia of atmospheric sciences, 1 (2015) 17–22.

[13] Q. Jin, J. Wei, W. K. Lau, B. Pu, C. Wang, Interactions of asian mineral dust with indian summer monsoon: Recent advances and challenges, Earth-Science Reviews 215 (2021) 103562.

[14] L. R. Crilley, A. Singh, L. J. Kramer, M. D. Shaw, M. S. Alam, J. S. Apte, W. J. Bloss, L. Hildebrandt Ruiz, P. Fu, W. Fu, et al., Effect of aerosol composition on the performance of low-cost optical particle counter correction factors, Atmospheric Measurement Techniques 13 (2020) 1181–1193.

[15] N. Streibl, Influence of humidity on the accuracy of low-cost particulate matter sensors, Techn. Ber. Technical report. DOI 10 (2017).

[16] S. Soneja, C. Chen, J. M. Tielsch, J. Katz, S. L. Zeger, W. Checkley, F. C. Curriero, P. N. Breysse, Humidity and gravimetric equivalency adjustments for nephelometer-based particulate matter measurements of emissions from solid biomass fuel use in cookstoves, International journal of environmental research and public health 11 (2014) 6400–6416.

[17] A. Skupin, Optische und mikrophysikalische Charakterisierung von urbanem Aerosol bei (hoher) Umgebungsfeuchte, Ph.D. thesis, Verlag nicht ermittelbar, 2013.

[18] G. Hänel, The properties of atmospheric aerosol particles as functions of the relative humidity at thermodynamic equilibrium with the surrounding moist air, in: Advances in geophysics, volume 19, Elsevier, 1976, pp. 73–188.

[19] A. Di Antonio, O. A. Popoola, B. Ouyang, J. Saffell, R. L. Jones, Developing a relative humidity correction for low-cost sensors measuring ambient particulate matter, Sensors 18 (2018) 2790.

[20] J. Shi, F. Chen, Y. Cai, S. Fan, J. Cai, R. Chen, H. Kan, Y. Lu, Z. Zhao, Validation of a light-scattering pm2. 5 sensor monitor based on the long-term gravimetric measurements in field tests, PloS one 12 (2017) e0185700.

[21] I. Tang, H. Munkelwitz, Aerosol phase transformation and growth in the atmosphere, Journal of Applied Meteorology and Climatology 33 (1994) 791–796.

[22] V. Kumar, M. Sahu, Evaluation of nine machine learning regression algorithms for calibration of low-cost pm2. 5 sensor, Journal of Aerosol Science 157 (2021) 105809.

[23] R. Duvall, A. Clements, G. Hagler, A. Kamal, V. Kilaru, L. Goodman, S. Frederick, K. Barkjohn, I. VonWald, D. Greene, et al., Performance testing protocols, metrics, and target values for fine particulate matter air sensors: Use in ambient, outdoor, fixed sites, non-regulatory supplemental and informational monitoring applications, US EPA Office of Research and Development (2021).