ON-FIELD PERFORMANCE TEST AND CALIBRATION OF TWO COMMERCIALLY AVAILABLE LOW-COST SENSORS DEVICES FOR CO_2 MONITORING

H. CHOJER, P.T.B.S. BRANCO, F.G. MARTINS & S.I.V. SOUSA
LEPABE – Laboratory for Process Engineering, Environment, Biotechnology and Energy, Faculty of Engineering, University of Porto, Portugal

ABSTRACT

The use of low-cost devices for air quality monitoring is rapidly growing, and the reason behind the growth might (at least partially) be the real-time monitoring at a lower fixed and operating cost, ease of use and portability. Nevertheless, the poor data reliability of low-cost sensors (LCS) remains a considerable challenge, especially when deployed in real-world conditions. This study aimed to evaluate and improve the performance of two commercially available indoor air quality monitoring LCS devices: AirVisual Pro and uRAD Monitor A3 (uRAD), which were used to monitor CO_2 via non-dispersive infrared technology. The analysis took place from June to July 2019 in several classrooms of an urban school in Porto city. Machine learning techniques such as multivariate linear, support vector, gradient boosting and XGBoost regression models were used to perform an on-field calibration for improving the data accuracy of the devices. The results showed that although both the devices showed a strong linear correlation ($r > 0.9$) with the reference device, they might indicate deviated CO_2 concentrations if used in their advertised plug and play format. Specifically, uRAD showed a steady offset compared to the reference values, while AirVisual Pro showed lower deviations than uRAD. The on-field calibration models improved the reliability and showed low root mean square error values (around 30 mg/m^3) and a high coefficient of determination (0.99).

Keywords: carbon dioxide, low-cost sensors, machine learning.

1 INTRODUCTION

The air quality sensor market is growing rapidly, with many companies today working on developing low-cost sensors (LCS), devices and networks [1]. In this rapidly growing market, LCS technology can be observed to be in a unique position of being a relatively new technology and fuelling a paradigm shift in air quality monitoring due to the inherent advantages that it delivers [2].

Their low-cost, portability, ease of use and real-time monitoring capability allow for novel opportunities in air pollution monitoring, which are not possible with traditional monitoring methods [3]. Citywide sensor networks for ambient environments and real-time monitoring in indoor environments such as homes, schools, hotels are possible with cheap and ubiquitous LCS. However, the design compromises leading to the cheap cost of LCS remain a largely unresolved issue. These sensors suffer from cross-sensitivity, have a short lifetime [4], exhibit drift in calibration over time and are sensitive to changes in ambient conditions [5, 6], and present inter-sensor variability [7]. These issues lead to weak data reliability of LCS, which is one of the biggest challenges facing LCS technology.

While there have been many sensor performance and calibration studies related to LCS used for outdoor air quality monitoring [5, 8–11], the studies evaluating the performance of commercially available LCS devices for indoor air quality monitoring are not so ubiquitous. Thus, the present study aimed to assess the performance of two commercially available LCS devices in a real-world indoor environment setting, measuring CO_2 concentrations.

ISSN: 2398-2640 (paper format), ISSN: 2398-2659 (online), http://www.witpress.com/journals
DOI: 10.2495/EI-V5-N1-15-22

Moreover, calibration models using supervised machine learning (ML) using LCS devices and reference grade instruments were performed.

2 METHODOLOGY

2.1 Deployed devices

The monitoring campaign was carried out in a nursery and a primary school in Porto city, Portugal. The monitoring devices were deployed in six different school classrooms from 3 June 2019 to 8 July 2019. Table 1 shows the commercially available LCS devices AirVisual Pro [12] and uRAD Monitor Model A3 (uRAD) [13] that were deployed with the research-grade device used as the reference instrument.

Apart from CO_2 measurements, all of the devices also measure temperature and humidity, data that was also used. Both the LCS devices used non-dispersive infrared based CO_2 sensors: uRAD contains a Winsen MH-Z19B sensor while the sensor for AirVisual was not mentioned in their technical specifications. Demanega et al. [14] used Air Visual Pro among several other LCS devices to perform a comparative assessment of LCS devices stated that it contains a SenseAir S8 or LP8 sensor. The researchers of this study did no prior lab calibration, and the sensor's performance was analysed with the factory settings (used in their advertised plug and play format).

2.2 Device performance

The acquired data showed several null data points and blank/missing values for uRAD. Blank/missing values were removed commonly for all datasets, and 10-minute means were used. Scatter plots were created to visualise the device performance compared to the reference values. Pearson correlation was evaluated to observe the existence of a linear relationship between the LCS devices and the reference instrument.

2.3 Calibration strategy

To improve data accuracy of the devices, the on-field calibration was performed using ML techniques. CO_2 concentrations from the LCS and reference grade instruments, as well as T and RH were used. The first step was mapping a correlation matrix to evaluate acceptable Pearson correlations, r, between the independent variables; if $|r| > 0.5$, then one of them was dropped to avoid multi-collinearity. The threshold value was set to be at 0.5 or higher as it

Table 1: The devices deployed in the field (inside the classrooms).

Device	Type	Concentration range	Monitoring interval
AirVisual Pro	Low-cost	400–10,000 ppm	10 s
uRADMonitor Model A3	Low-cost	400–5000 ppm	1 min
Reference gases: Hazscanner	Research-grade	0–10,000 ppm	1 min

represents moderate to good correlations beyond that point [15]. As a rule, the LCS CO_2 variable was never dropped from the analysis as the improvement in data reliability was intended to be built upon it using other variables. The data were resampled randomly into training (80%) and validation (20%) datasets with a fixed seed before regression modelling. The resampling seed was fixed to validate the models in a uniform way for the comparative analysis between the four models: multivariate linear regression (MLR), support vector regression (SVR), gradient boosting regression (GBR) and extreme gradient boosting (XGB).

Figure 1 shows the supervised learning process. The model training was done using the reference values as the output (or dependent variable) while the LCS CO_2 values and the temperature and relative humidity (RH) were considered as the input (or independent/explanatory variables). The temperature and RH data used were those acquired from the reference device to increase reliability of the developed models. The statistical significance check of the variables was performed for MLR models. The level of statistical significance was set at a p-value of 0.05.

Hyperparameters were optimised for SVR, XGB and GBR models via an exhaustive grid search performed with threefold cross validation. The different hyperparameters that were selected to be optimised are shown in Table 2. The hyperparameters exhibiting high influence on the model were selected to be optimised.

After the models were trained with the optimised hyperparameters, the validation of models (on 20% of the unseen dataset) was evaluated. Statistical indexes used to evaluate the performance of the models were coefficient of determination R^2, root mean squared error (RMSE) and mean bias error (MBE).

The entire data analysis was done using Python 3.7 with Jupyter Notebook interface [16]. *Scikit-learn* library was used to train all the regression models and to optimise the hyperparameters [17]. The statistical significance tests were done using *statsmodels* module in Python [18]. The visualisation was done using two libraries: *matplotlib* and *seaborn* [19, 20].

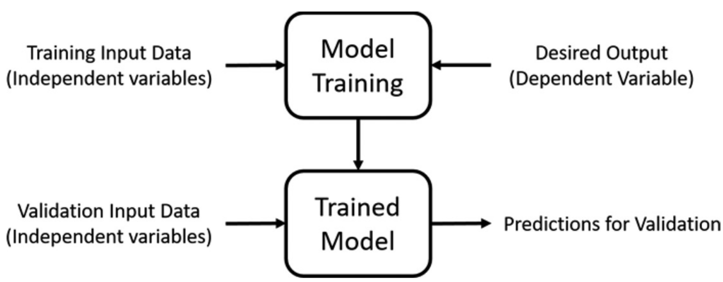

Figure 1: Supervised training and validation process.

Table 2: Hyperparameters optimised for the models.

Model	Hyperparameters
SVR	Regularisation parameter C
GBR	Number of boosting stages, learning rate, maximum depth
XGB	Number of boosting stages, learning rate, maximum depth, subsample, gamma

3 RESULTS AND DISCUSSION

Results for various classrooms were similar; hence, results from one representative classroom are presented and discussed. The datasets, after merging and taking 10 min mean, yielded 712 data points.

3.1 Device performance

Figure 2 shows the performance of low-cost devices compared to the research-grade reference instrument in the form of scatter plots. The Pearson correlation shows the linear correlation in the figure. Both the devices show a strong correlation to the reference. AirVisual Pro shows a few erratic peaks in measurement. The results obtained for AirVisual Pro resemble the recently published results of Demanega et al. [14], who got a strong linear relationship with Pearson correlation coefficient of 0.975.

Although uRAD shows an almost perfect linear correlation, it has a consistent offset compared to the reference. The implication arising from this observation is that such devices might show an increase in the concentration levels leading to wrong mitigation/prevention measures. LCS devices also tend to understate pollutant concentration levels (especially particulate matter) in some cases, as observed by the authors and stated elsewhere.

3.2 Calibration

Figure 3 shows the Pearson correlation matrix plot of the variables ascertained from all the devices deployed during the research campaign in the form of a heat plot. Threshold correlation of 0.5 or higher was assessed and RH was removed from further analyses as it showed

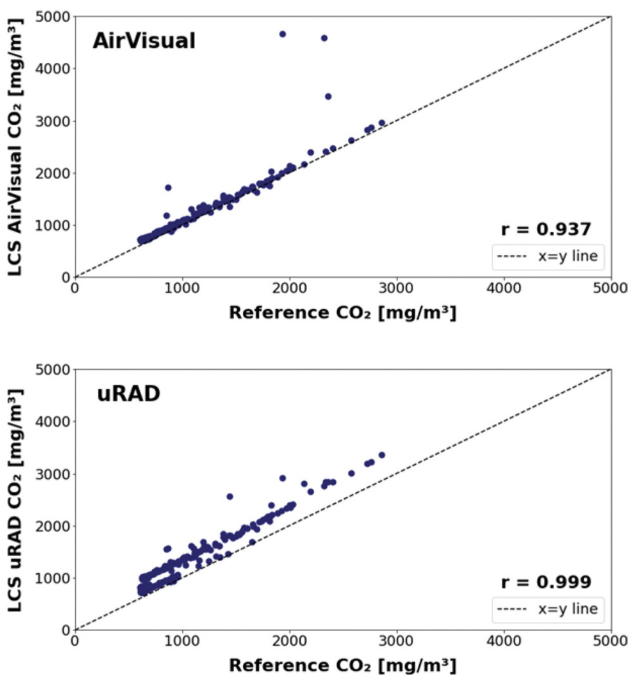

Figure 2: Scatter plot of the low-cost devices and the reference device.

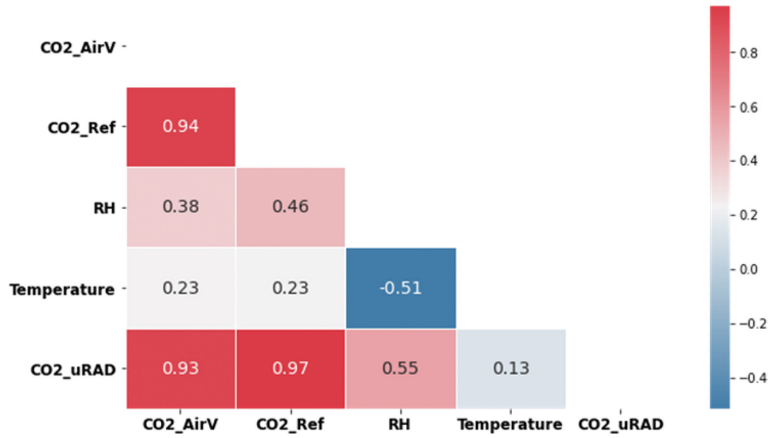

Figure 3: Pearson correlation matrix plot of all the variables involved. CO2_AirV: CO_2 AirVisual Pro; CO2_uRAD: CO_2 uRAD Monitor A3.

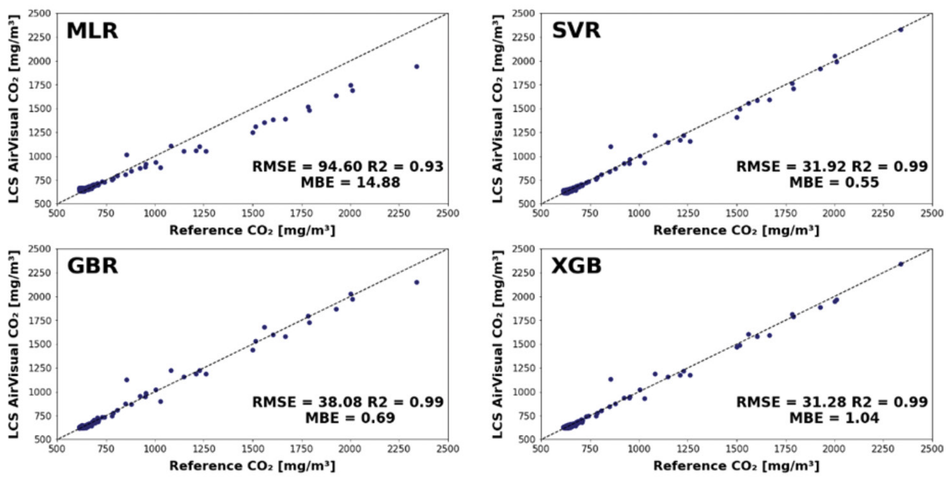

Figure 4: Model validation scatter plot after hyperparameter optimisation for all four models for AirVisual Pro. MLR: multiple linear regression; SVR: support vector regression; GBR: gradient boosting regression; XGB: extreme gradient boosting.

high correlation to the other input variables. The models, after data resampling, were trained with LCS CO_2 and temperature as the input variables for both the devices. The high correlation between the LCS and reference CO_2 values is evident from the correlation matrix.

Figure 4 shows the scatter plots of all four models implemented after hyperparameter optimisation for AirVisual Pro. All models seemingly performed well. For SVR, linear kernel was used for the model training, and the regularisation parameter C was optimised. For GBR and XGB, number of boosting iterations and learning rate had the biggest impact on the models. A very high R^2 score can be observed for SVR, GBR and XGB. These three models also showed low RMSE and MBE values.

Table 3: Performance indexes of all models for the low-cost devices compared to the reference.

Dataset	Model	AirVisual			uRAD		
		R^2	RMSE	MBE	R^2	RMSE	MBE
Training	MLR	0.857	123.807	−1.758	0.952	71.910	−5.085
	SVR-Linear	0.774	155.374	−10.780	0.951	72.693	−9.349
	GBR	0.999	7.769	−7.752	0.995	20.990	−6.593
	XGB	0.988	34.675	−0.552	0.999	10.167	−0.027
Validation	MLR	0.932	94.604	14.875	0.954	77.753	−0.838
	SVR-Linear	0.992	31.920	0.552	0.953	79.077	−10.010
	GBR	0.989	38.075	0.692	0.985	44.829	2.337
	XGB	0.992	31.277	1.039	0.989	36.733	1.855

After the implementation of models, the erratic peaks exhibited by the device have been rectified as well. Hence, the application of models using supervised ML techniques have improved the data reliability of AirVisual Pro. The models also improved the data reliability of uRAD. Table 3 shows the results of the training and validation datasets for both devices. For uRAD, the validation results show that XGB improved the data accuracy the most, with a very high R^2 score of 0.989 and very low error values.

For both the devices, XGB really stood out as the best performing model among the four models implemented in this study. It showed excellent training scores, which implies that it could encompass enough information from the training phase to make a model complex enough to capture accurate relationships between the target variable and the input variables while also showing good results on unseen data implying that it was not overfitting.

In comparison with other studies that performed on-field calibration of CO_2 monitoring LCS, the validation results obtained by this study show a much higher R^2 score. Spinelle, et al. [21] performed field calibration of commercially available LCS with linear, MLR and artificial neural network (ANN) models and their highest validation set R^2 score was 0.732 with ANN versus reference measurements. It should be noted that their study included CO_2 sensor and field calibration (among several others), but was focused towards outdoor monitoring.

4 CONCLUSIONS

There are three major takeaways from the present study. The first is regarding the performance of LCS devices for CO_2 monitoring. Although not very high in their plug and play format, the data accuracy can still be concluded as useful for indicative purposes. Secondly, the on-field calibration yielded very good results and could improve the data accuracy and subsequently the data reliability of the two LCS devices. Supervised ML techniques implemented for the proposed calibration strategy developed models that closely resembled the reference values. Thirdly, the comparative analysis showed that the XGB model was very consistent and outperformed all the other models to improve the data accuracy of CO_2 monitoring for both low-cost devices.

Future work on this topic may involve calibrating other sensors of these devices, incorporating more models for a bigger comparative analysis, and more extended monitoring periods to observe how the models perform on long-term measurements once deployed. It will also test if the models can cope up with the calibration drift associated with LCS.

ACKNOWLEDGEMENTS

The authors express immense gratitude to the school for their cooperation. This work was financially supported by: Base Funding – UIDB/00511/2020 of the Laboratory for Process Engineering, Environment, Biotechnology and Energy – LEPABE – funded by national funds through the FCT/MCTES (PIDDAC); Project PTDC/EAM-AMB/32391/2017, funded by FEDER funds through COMPETE2020 – Programa Operacional Competitividade e Internacionalização (POCI) and by national funds (PIDDAC) through FCT/MCTES; and Project 2SMART - engineered Smart materials for Smart citizens, with reference NORTE-01-0145-FEDER-000054, supported by Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF). H. Chojer thanks the Portuguese Foundation for Science and Technology (FCT) for the individual research grant SFRH/BD/05092/2021. Sofia I.V. Sousa thanks the Portuguese Foundation for Science and Technology (FCT) for the financial support of her work contract through the Scientific Employment Stimulus - Individual Call - CEEC-IND/02477/2017.

REFERENCES

[1] InsightPartners, Air Quality Sensor Market Forecast to 2027 – COVID-19 Impact and Global Analysis by Type, Location, and End User, 2020.

[2] E. G. Snyder *et al.*, The changing paradigm of air pollution monitoring. *Environmental Science & Technology,* 47(20), pp. 11369–11377, 2013. Art no. Journal Article Review, doi: https://doi.org/10.1021/es4022602.

[3] N., Castell, M., Viana, M. i. C., Minguillón, C., Guerreiro and X., Querol, Real-world application of new sensor technologies for air quality monitoring. in *ETC/ACM Technical Paper*, vol. 16, 2013.

[4] R. M., White, I., Paprotny, F., Doering, W. E., Cascio, P. A., Solomon and L. A. G., Sensors, Sensors and Apps for Community-Based Atmospheric Monitoring. *In: Air and Waste Management Association's Magazine for Environmental Managers,* pp. 36–40, 2012.

[5] P. J. D. Peterson *et al.*, Practical Use of Metal Oxide Semiconductor Gas Sensors for Measuring Nitrogen Dioxide and Ozone in Urban Environments, (in eng). *Sensors (Basel, Switzerland),* 17(7), 2017. Art no. Journal Article The Zephyr air quality sensor is currently being developed by Earthsense systems Ltd., a company in which Roland Leigh has direct commercial involvement., doi: https://doi.org/10.3390/s17071653.

[6] L. Morawska *et al.*, Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International,* 116, pp. 286–299, 2018/07/01/ 2018. doi: https://doi.org/10.1016/j.envint.2018.04.018.

[7] L., Zhang, F.-C., Tian, X.-W., Peng and X., Yin, A rapid discreteness correction scheme for reproducibility enhancement among a batch of MOS gas sensors. *Sensors and Actuators A: Physical,* 205, pp. 170–176, 2014/01/01/ 2014. doi: https://doi.org/10.1016/j.sna.2013.11.015.

[8] J. M., Cordero, R., Borge and A., Narros, Using statistical methods to carry out in field calibrations of low cost air quality sensors. *Sensors and Actuators B: Chemical,* 267, pp. 245–254, 2018/08/15/ 2018. doi: https://doi.org/10.1016/j.snb.2018.04.021.

[9] L., Spinelle, M., Gerboles, M. G., Villani, M., Aleixandre and F., Bonavitacola, Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical,* 215, pp. 249–257, 2015/08/01/ 2015. doi: https://doi.org/10.1016/j.snb.2015.03.031.

[10] H. Cui *et al.*, A new calibration system for low-cost Sensor Network in air pollution monitoring. *Atmospheric Pollution Research,* 12(5), p. 101049, 2021/05/01/ 2021. doi: https://doi.org/10.1016/j.apr.2021.03.012.

[11] L., Liang, Calibrating low-cost sensors for ambient air monitoring: techniques, trends, and challenges. *Environmental Research,* vol. 197, p. 111163, 2021/06/01/ 2021, doi: https://doi.org/10.1016/j.envres.2021.111163.

[12] IQAir., AirVisual Pro. https://www.iqair.com/air-quality-monitors/airvisual-pro (accessed).

[13] Radu., uRAD Monitor Model A3., https://www.uradmonitor.com/uradmonitor-model-a3/ (accessed).

[14] I., Demanega, I., Mujan, B. C., Singer, A. S., Anđelković, F., Babich and D., Licina, Performance assessment of low-cost environmental monitors and single sensors under variable indoor air quality and thermal conditions. *Building and Environment,* 187, p. 107415, 2021/01/01/ 2021. doi: https://doi.org/10.1016/j.buildenv.2020.107415.

[15] M. M. Mukaka, “Statistics corner: A guide to appropriate use of correlation coefficient in medical research,” (in English), *Malawi Med J,* vol. 24, no. 3, pp. 69–71, 2012. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23638278

[16] Jupyter., https://www.python.org/ (accessed 24/03/2021.

[17] F. Pedregosa *et al.*, Scikit-learn: machine learning in Python. *Journal of Machine Learning Research,* 12(85), pp. 2825–2830, 2011.

[18] S., Seabold and J., Perktold, Statsmodels: econometric and statistical modeling with python. 2010.

[19] M., Waskom, Seaborn: statistical data visualization, *Journal of Open Source Software.* 6(60), p. 3021, 2021. doi: 10.21105/joss.03021 M4 – Citavi.

[20] J. D., Hunter, Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering,* 9(3), pp. 90–95, 2007. doi: 10.1109/MCSE.2007.55.

[21] L., Spinelle, M., Gerboles, M. G., Villani, M., Aleixandre and F., Bonavitacola, Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2, *Sensors and Actuators B: Chemical.* 238, pp. 706–715, 2017/01/01/ 2017. doi: https://doi.org/10.1016/j.snb.2016.07.036.