Data Analysis of Low-cost Air Quality Sensors in Helsinki

Zaidan, Martha

Nordic Society for Aerosol Research (NOSA)
2020-06

Zaidan , M , Hossein Motlagh , N , Fung , P L , Timonen , H , Tarkoma , S , Hussein , T , Petäjä , T & Kulmala , M 2020 , Data Analysis of Low-cost Air Quality Sensors in Helsinki . Nordic Society for Aerosol Research (NOSA) , Denmark .

http://hdl.handle.net/10138/333822

acceptedVersion

DATA ANALYSIS OF LOW-COST AIR QUALITY SENSORS IN HELSINKI

M.A. ZAIDAN[1], N.H. MOTLAGH[2], P.L. FUNG[1], H, TIMONEN[3], S. TARKOMA[2], T. HUSSEIN[1], T. PETÄJÄ[1] and M. KULMALA[1]

[1] Institute for Atmospheric and Earth System Research (INAR), Helsinki University, Finland
[2] Department of Computer Science, University of Helsinki, Finland
[3] Finnish Meteorological Institute, Helsinki, Finland

Keywords: air quality, low-cost sensors, reference station, data analysis.
Aerosol Theme: Aerosol Measurement Techniques (Comparison of different measurement methods).

Air pollution is a major issue in urban areas. High population density is exposed with excess anthropogenic emissions impacting environment and health effects. Each year approximately 4.2 million deaths are attributed to exposure to ambient air pollution[1]. To perform air quality monitoring in urban areas, there is a need to measure the pollution with high resolution. Indeed, highly accurate and reliable air quality monitoring stations have been established to continuously monitor the air pollution. However, these monitoring stations are expensive and complex in operation and maintenance. Therefore, it is not feasible to deploy these stations massively in urban areas (Lagerspetz et al., 2019). Alternatively, dense deployment of low-cost air quality sensors in urban areas allows detecting the pollution hot spots in real-time. Nevertheless, these low-cost sensors suffer from sensing accuracy (Motlagh et al., 2020). In this paper, we present data analysis of two identical low-cost sensors provided by Clarity Movement Co[2]. The low-cost sensors measure temperature (Temp), relative humidity (RH), particulate matter ($PM_{2.5}$) and Carbon dioxide (CO_2). These low-cost sensors were installed next to the reference station in Kumpula Campus, University of Helsinki, Finland, called Station for Measuring Ecosystem-Atmosphere Relations (SMEAR III)[3]; and the data were collected from 13 Mar - 31 Dec 2018.

Figure 1 shows the relationship between the measured variables by low-cost sensors (Klc) and the measurements performed by SMEAR III (K). The different colors indicate the Pearson correlation coefficient levels (r). The correlation levels between the same measured variables of low-cost sensors provide almost perfect result, except for CO_2 with r value approximately equal to 0.85. The results demonstrate that the low-cost sensors provide consistent performance in terms of accuracy. The Figure also shows slight deviations (i.e. imperfect correlations) between the Temp (K) and Temp (Klc). Similar results are also obtained for measured RH between Klc and K by exhibiting slight offset. The results for Temp and RH show the need for calibrations of low-cost sensors by applying a simple correction factor. This is required to meet the SMEAR III data quality.

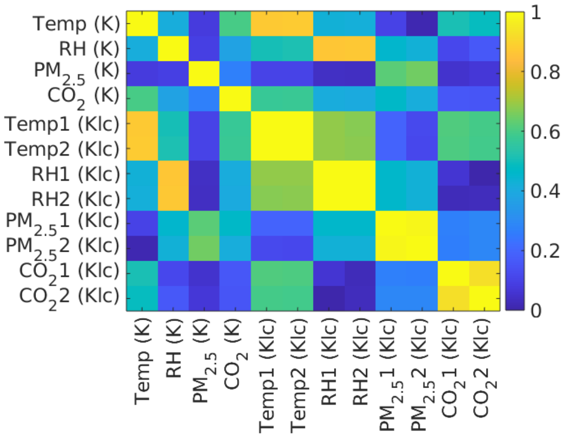

Figure 1: *C*orrelation analysis of the measurements of low-cost sensors (Klc) and SMEAR III (K).

The correlation for $PM_{2.5}$ between K and Klc presents lower r values almost equal to 0.65. However, this result indicates that measurements of Klc still follow the measurements pattern of (K), but with a bigger offset than Temp (Klc) and RH (Klc). Therefore, they need to be calibrated with not just a correction factor but with a mathematical model. The variable CO_2 presents almost no correlation between K and Klc with r almost equal to zero. This indicates that the measurements of Klc does not follow the pattern of K. Therefore, the sensor calibration is not a suitable approach for improving the accuracy of CO_2 measurements. Instead, mathematical models established using other measured variables, known as a proxy is requires to accurately estimate the CO_2 values (Zaidan et al., 2019). The future direction of this study is to improve the accuracy of low-cost sensors measurements by developing and deploying mathematical models for calibration and proxy.

Lagerspetz, E., et al. (2019). Megasense: Feasibility of low-cost sensors for pollution hot-spot detection, *2019 IEEE 17th International Conference on Industrial Informatics (INDIN)*, vol. 1, July 2019, pp. 1083-1090.

Motlagh, N.H. et al. (2020). Towards Massive Scale Air Quality Monitoring, *IEEE Communications Magazine*, February, 2020.

Zaidan, M.A. et al. (2019). Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies, *Applied Sciences*, 9, 4475.

[1] https://www.who.int/gho/phe/outdoor_air_pollution/burden_text/en/
[2] https://clarity.io/
[3] https://www.atm.helsinki.fi/SMEAR/index.php/smear-iii