



Evaluation of Crowd-Sourced PM_{2.5} Measurements from Low-Cost Sensors for Air Quality Mapping in Stuttgart City

14

Joseph Gitahi and Michael Hahn

Abstract

Exposure to particulate matter (PM) pollution poses a major risk to the environment and human health. Monitoring PM pollution is thus crucial to understand particle distribution and mitigation. There has been rapid development of low-cost PM sensors and advancement in the field of Internet of Things (IoT) that has led to the deployment of the sensors by technology-aware people in cities. In this study, we evaluate the stability and accuracy of PM measurements from low-cost sensors crowd-sourced from a citizen science project in Stuttgart. Long-term measurements from the sensors show a strong correlation with measurements from reference stations with most of the selected sensors achieving Pearson correlation coefficients of $r > 0.7$. We investigate the stability of the sensors for reproducibility of measurements using five sensors installed at different height levels and horizontal distances. They exhibit minor variations with low correlation of variation (CV) values of between 10 and 14%. A CV of $\leq 10\%$ is recommended for low-cost sensors. In a dense network, the sensors enable extraction pollution patterns and trends. We analyse PM measurements from 2 years using space-time pattern analysis and generate two clusters of sensors that have similar trends. The clustering shows the relationship between traffic and pollution with most sensors near major roads being in the same cluster.

Keywords

Particulate matter · Low-cost sensors · Crowd-sourced air quality data · Citizen science

J. Gitahi (✉) · M. Hahn
Hochschule für Technik Stuttgart, Stuttgart, Germany
e-mail: joseph.gitahi@hft-stuttgart.de

14.1 Introduction

Ambient air pollution poses a crucial environmental risk to human health globally. This applies especially to urban centres which are characterized by high population densities, heavy vehicle traffic and high concentration of industries. Fine dust particles also referred to as particulate matter (PM) are a major component of air pollution whose sources include dust; combustion particles from power plants, vehicles and industries; and reactions of chemicals such as SO_2 and NO_x . They are categorized as PM_{10} and $\text{PM}_{2.5}$ for particles with a diameter of less than or equal to $10\ \mu\text{m}$ and $2.5\ \mu\text{m}$, respectively. $\text{PM}_{2.5}$ passes through the respiratory system with ease due to the smaller size, thus presenting a higher risk to human health. To reduce health impacts, the World Health Organization (WHO) has recommended $\text{PM}_{2.5}$ concentration thresholds of 10 and $25\ \mu\text{g}/\text{m}^3$ for annual and daily averages, respectively (WHO, 2016).

Two major sources of data are used for air quality monitoring, satellite remote sensing products and ground-based sensors. Columnar aerosol optical depth (AOD), a by-product of atmospheric correction of optical satellite images, is retrieved based on the inversion of radiative transfer (RT) equations which model the scattering and absorption of solar radiation by aerosols, gas and water molecules in the atmosphere. Readily available satellite AOD include the Moderate Resolution Imaging Spectroradiometer (MODIS) product MOD04 providing a high temporal resolution AOD for daily-based monitoring at 3 km (MOD04_3K) and 10 km (MOD04_L2) spatial resolution suited for global and regional scales. Under European Space Agency's (ESA) Copernicus programme, Sentinel-3 provides AOD at 300 m spatial resolution with a revisit time of 1–2 days.

Besides land monitoring satellites, satellite missions dedicated to air quality monitoring like ESA's Sentinel-5P measure gaseous and aerosol pollutants. Sentinel-5P continuously measures gaseous pollutants and aerosol index at a spatial resolution of $7\ \text{km} \times 3.5\ \text{km}$ with daily global coverage. While satellite remote sensing products have the inherent advantage of extensive spatial coverage, their spatial-temporal resolutions are not capable of mapping spatial-temporal air quality variations in detail. Ground-based air quality sensors such as reference monitoring stations, operated by environmental agencies and institutions, are highly accurate and reliable. However, due to their high installation costs, only a few reference stations are in use. Low-cost sensors present an opportunity to create dense air quality monitoring networks.

Air pollution in urban environments has large spatial and temporal variations which require a dense network of sensors for adequate monitoring. High-quality and accurate air quality monitoring stations are costly to install in large numbers. Citizen science initiatives have embarked on installing low-cost sensors for civic engagement in monitoring and controlling air pollution. Some of the initiatives in Europe include CITI-SENSE (www.citi-sense.eu), hackAIR (www.hackair.eu) and OK Lab Stuttgart (www.luftdaten.info). These sensors create a dense network which can supplement air quality data from the few reference stations. The sensors provide relative and indicative air quality measurements but at lower accuracies than required for regulatory purposes. They are prone to erroneous

measurements due to sensor faults, wrong handling by users and interference from meteorological parameters such as temperature and humidity. The measurements also have substantial data gaps hindering continuous air quality monitoring. Evaluation of the sensors is thus necessary before using them for mapping spatial and temporal variations of air pollution.

European countries are required to comply with EU air quality monitoring directives—Directive (AQD) 2008/50/EC on Air Quality (EU, 2008). The AQD outlines the criteria and reference measurement methods by member countries using fixed monitoring stations for legislative purposes. However, the directive also allows for supplementary indicative measurements from low-cost sensor platforms provided they meet the defined data quality objective (DQO). The DQO, a measure of the acceptable uncertainty of measurements, allows uncertainties of up to 50% for PM₁₀ and PM_{2.5} measurements.

Most of the low-cost PM sensors in the market detect the number and size of dust particles in the air based on the light-scattering principle. For these sensors, accumulation of dust particles in the measuring chamber and extreme weather conditions, especially high humidity, are some of the factors affecting data quality (Castell et al., 2017; Badura et al., 2018; Bulot et al., 2020). The sensors are evaluated on several aspects: stability and accuracy of measurements, and their precision. The operational stability is crucial to determine sensors' performance over long-term measurement campaigns. They are assessed for stability and accuracy by comparing with measurements of co-located reference stations, while precision is determined by testing the reproducibility of data from different units of the same sensor model. The precision of sensors is evaluated using the coefficient of variation (CV) which is a ratio of the standard deviation and mean of measurements. A CV of zero shows a perfect agreement, and a CV of $\leq 10\%$ is acceptable for PM monitoring using low-cost sensors (Sousan et al., 2016; Bulot et al., 2020).

Different models of commercially available low-cost PM sensors have been subjected to tests in several studies to ascertain their accuracy and precision. The SDS011 sensor by Nova Fitness is a popular choice due to its low cost (< 20 €), low energy requirement and relatively stable performance. Badura et al. (2018) compared multiple units of four low-cost sensor models with a TEOM 1400a reference station for 6 months. Multiple units of SDS011 sensors were assessed for reproducibility where they scored a CV of 7% indicating good precision. The sensors also exhibited good agreement with the reference station with R^2 values of between 0.79 and 0.86. In another study, Liu et al. (2019) evaluated three SDS011 sensors co-located with a reference station over 4 months in Oslo. PM_{2.5} measurements from the sensors were highly correlated with the reference station having correlation values r of > 0.97 . On accuracy assessment, the sensors achieved good linearity with the reference station attaining R^2 values of between 0.55 and 0.71, and low RMSE values of < 6 $\mu\text{g}/\text{m}^3$.

In this study, we evaluate the suitability of PM_{2.5} measurements from a low-cost sensor network for spatial-temporal mapping of air quality in Stuttgart city. The sensors are evaluated on three aspects. Firstly, we perform an inter-sensor comparison by placing the sensors with different vertical and horizontal distances in the same location to determine

the influence of sensor's placement on performance. Secondly, we assess the stability and correlation of selected SDS011 sensors with the nearest reference station using a long-term dataset spanning over 1 year. Lastly, the dense network of sensors is analysed to identify PM distribution and patterns in a spatial-temporal context.

14.2 Methodology

14.2.1 Study Area

The city of Stuttgart suffers from high pollution; PM levels have in the past exceeded the thresholds set by WHO which attributed to high traffic and industrial activities. Geographically, the city centre and main industrial areas are in a valley which affects air pollution transport and dispersion (Fig. 14.1).

14.2.2 Datasets

In the study, we use two PM_{2.5} datasets. The first dataset is from five traffic and background air quality monitoring stations operated by the state institute for environment, Landesanstalt für Umwelt Baden-Württemberg (LUBW). The stations are distributed within and outside the city boundary and are used as reference stations in this study. From these reference stations, we obtain three PM measurements: PM₁₀ gravimetry, PM_{2.5} gravimetry and PM₁₀ photometry. The photometric measurements are available in real time for public information, while the more accurate gravimetric measurements are available after 10 days. Hourly averages of PM_{2.5} g gravimetric measurements are retrieved from LUBW API and stored in a spatial database. We use a dataset of measurements from June 2019 to June 2020.

The second dataset is from Luftdaten network of low-cost sensors by OK Lab Stuttgart (www.luftdaten.info) with approximately 200–350 operational sensors in the city at any given time. The primary sensor used in this network is the Nova PM sensor SDS011 which uses light scattering to measure the number and diameter of dust particles passing through the detector. OK Lab Stuttgart provides users with a list of components required to build the sensor as well as firmware and the configuration needed to set up the sensor and to upload recorded data to a central portal. The components include a micro-controller unit, SDS011 module, an optional temperature and humidity module and a pipe casing. The cost of the setup ranges from 25 € to 30 €. The sensors upload PM measurements every 2.5 min to the Luftdaten portal that is accessible via an API. We use scheduled scripts to retrieve and store measurements every 15 min. This dataset is available from August 2018 to August 2020 for sensors inside and near the city boundary. Table 14.1 shows the sensor specifications (Nova Fitness, 2015).

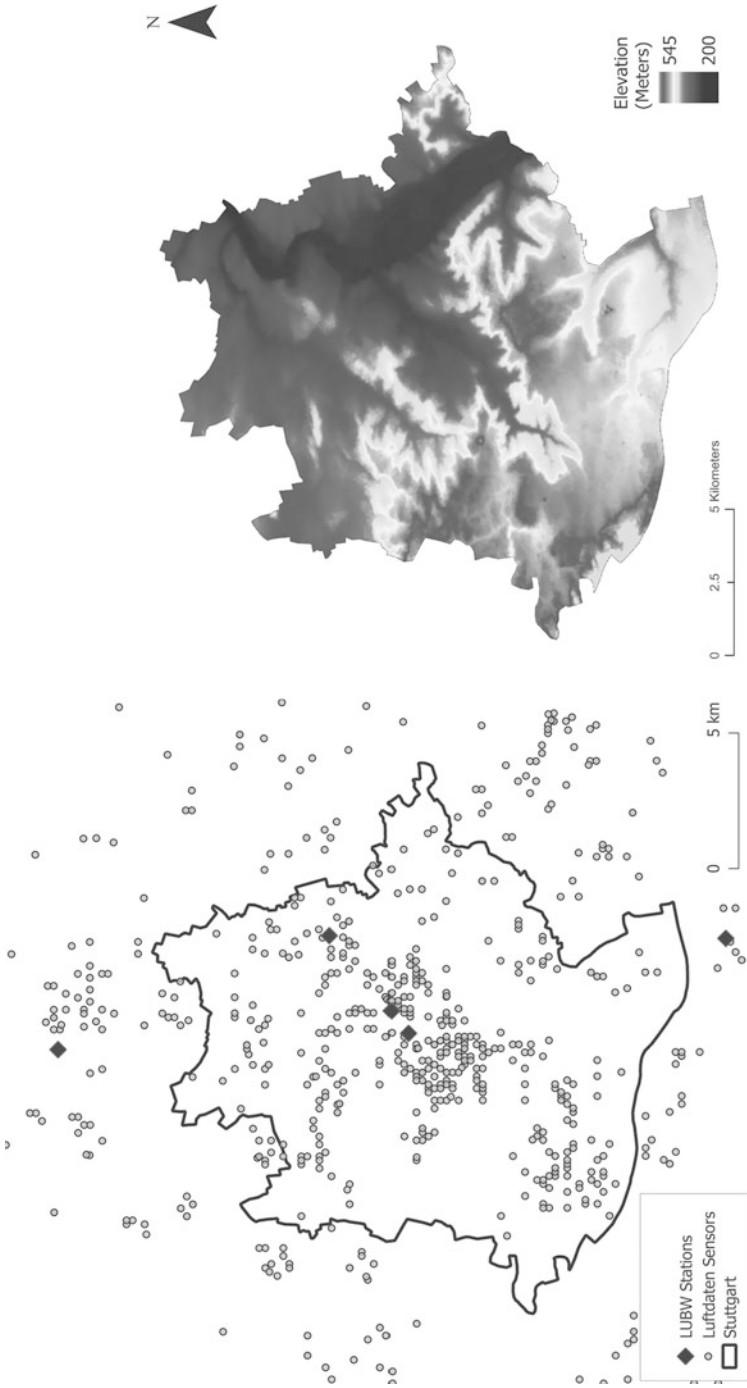
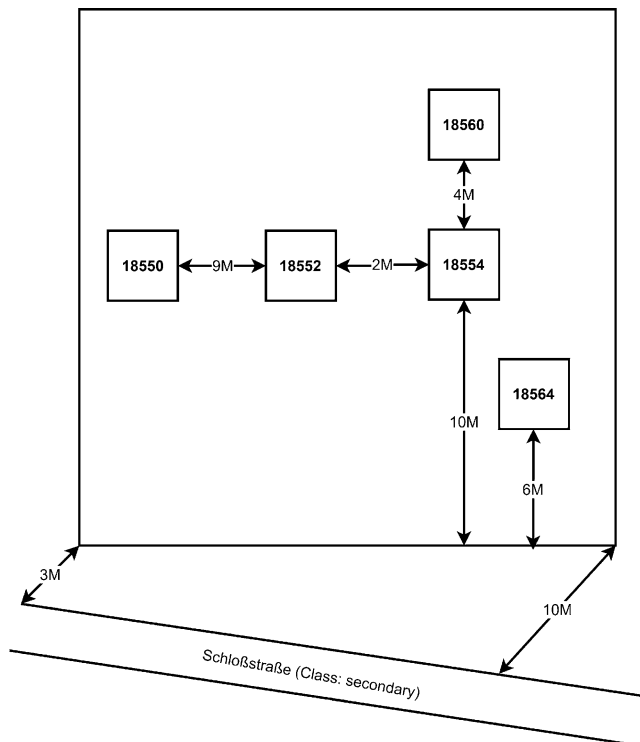


Fig. 14.1 A map (left) showing the location of low-cost sensors and reference monitoring stations by the state environmental authority. The map on the right shows an elevation map of the greater Stuttgart city region. The city centre and most industries are situated in the valley region

Table 14.1 Nova SDS011 PM sensor specifications

Item	Specification
Measuring parameters	PM _{2.5} , PM ₁₀
Measuring range	0.0–999.9 µg/m ³
Minimum resolution of dust particles	0.3 µm
Response time	< 10 seconds
Relative error	Maximum of ±15% and ± 10 µg/m ³ at 25 °C and 50% RH
Temperature range	–20–50 °C
Humidity range	0–70% relative humidity (RH)
Service life	8000 h

**Fig. 14.2** Installation of low-cost sensors at different points on the wall of building in University of Applied Sciences, Stuttgart. The building is adjacent to a secondary-class road

In Stuttgart University of Applied Sciences, we installed five SDS011 sensors for further investigations on their stability when placed at different heights and horizontal distances. They were installed on the facet of a building which is approximately 3–10 m adjacent to a secondary-class road. The placement of the sensors is shown in Fig. 14.2.

Since most of the sensors do not have a weather module, we use weather data from the OpenWeatherMap service. The data includes temperature, relative humidity, atmospheric pressure, wind direction and speed from 23 locations in the study area. An alternative weather dataset from Deutscher Wetterdienst (DWD) is available but has only one measurement location in the study area. This data is retrieved from the API at 15-min interval and is available from June 2019 to June 2020.

14.2.3 Data Preparation

In the first step, PM observations from the low-cost sensors are aggregated to hourly averages followed by removing measurements that lie outside the measuring range. The hourly aggregates are calculated to match the reference stations' sampling rate. We then create a new dataset by combining hourly PM measurements from the sensors and the reference stations, and the weather data. This fused dataset is created by spatially joining the low-cost PM measurements to the nearest weather and LUBW stations. Since the LUBW stations are few and sparsely distributed, a field containing the spatial distance in metres is calculated to allow analysis of low-cost sensors that are only within a specific distance from the high-quality stations. This combined dataset ranges from June 2019 to June 2020.

14.2.4 Low-Cost Sensors' Evaluation

Repeatability of PM measurements is crucial when using low-cost sensors for air quality monitoring. The coefficient of variation (CV) is calculated for hourly average PM_{2.5} measurements to assess sensors' precision for the sensors installed in the university building as shown in Eq. (14.1). Temporary CV is calculated for corresponding hourly average measurements and a final CV determined as an average of all temporary CVs. Two sets of sensors were compared: sensors placed at the same height but varying horizontal distances and sensors placed at different heights on the building.

$$CV_t = \frac{\sigma_t}{\mu_t} \cdot 100 \quad (14.1)$$

where CV_t is the coefficient of variation at time t and σ_t and μ_t are the standard deviation and mean at time t , respectively.

The sensors' performance is further assessed by comparing with the LUBW reference stations by calculating the Pearson correlation coefficient (r) and the root mean square error (RMSE). In this assessment, PM_{2.5} measurements from low-cost sensors that are within 1 km radius of the reference stations and within the operating range of 0–70% RH are selected for analysis. To further examine the quality of the low-cost sensors'

measurements, we select one low-cost sensor for each reference station with the highest correlation and perform multilinear regression. In the linear fitting, reference station $PM_{2.5}$ is the dependent variable, and low-cost sensors $PM_{2.5}$, temperature and humidity are the independent variables as shown in Eq. (14.2). The relationships are evaluated using coefficients of determination (R^2) and RMSE.

$$y = \beta_0 + \beta_x + \beta_{RH} + \beta_T \quad (14.2)$$

Multilinear regression fitting where y is the reference station $PM_{2.5}$, x is the low-cost sensor PM , RH is the relative humidity and T is the temperature.

A dense network of sensors provides a chance to extract underlying air pollution spatial patterns using long-term measurements. We use ArcGIS Pro Space-Time Pattern Mining toolbox to analyse $PM_{2.5}$ distribution and patterns in space and time. First, we create space-time bins by aggregating $PM_{2.5}$ measurements into daily averages per sensor location. Data gaps due to sensor malfunction and transmission issues are filled by interpolating values based on the temporal trend of $PM_{2.5}$ values for each sensor. The space-time cubes are then used to analyse PM concentrations using the time series clustering technique. In this technique, similar sensors are grouped based on either similar $PM_{2.5}$ values, increase and decrease of values at the same time or having similar repeating patterns. We extract cluster patterns based on $PM_{2.5}$ values using the long-term dataset from August 2018 to August 2020.

14.3 Results and Discussion

Figure 14.3 represents the results of $PM_{2.5}$ measurements from five sensors placed at varying horizontal and vertical distances. Three sensors placed at the same height of 10 m and short horizontal distances of 2, 9 and 11 m from each other had stable measurements throughout the 1-month testing period. There were no significant variations and the sensors showed good precision with a mean CV of 10%. For the vertical assessment on sensors placed at different heights of 6, 10 and 14 m, the measurements followed a similar trend but with minor variations and a mean CV of 14%. Sensor 18,560 at 14 m generally has slightly lower values compared to the other two, but sensor 18,554 at 10 m has slightly higher values than the sensor at 6 m.

A trend analysis of $PM_{2.5}$ measurements by the sensors over 1 year shows a good correlation with the reference stations. Out of the 52 sensors selected, 23 had a correlation coefficient r values of >0.7 , and only 13 had r values of <0.5 . In both urban and suburban settings, most of the sensors were able to detect peaks recorded by the reference stations with minor variances. The higher variations are observed in cold months starting from October to March as seen in Fig. 14.4. In the 1 km radius, distance from the reference station has little influence on the correlation, but the location of sensors has a greater impact. In the city centre map shown in Fig. 14.5, sensors in similar settings as the

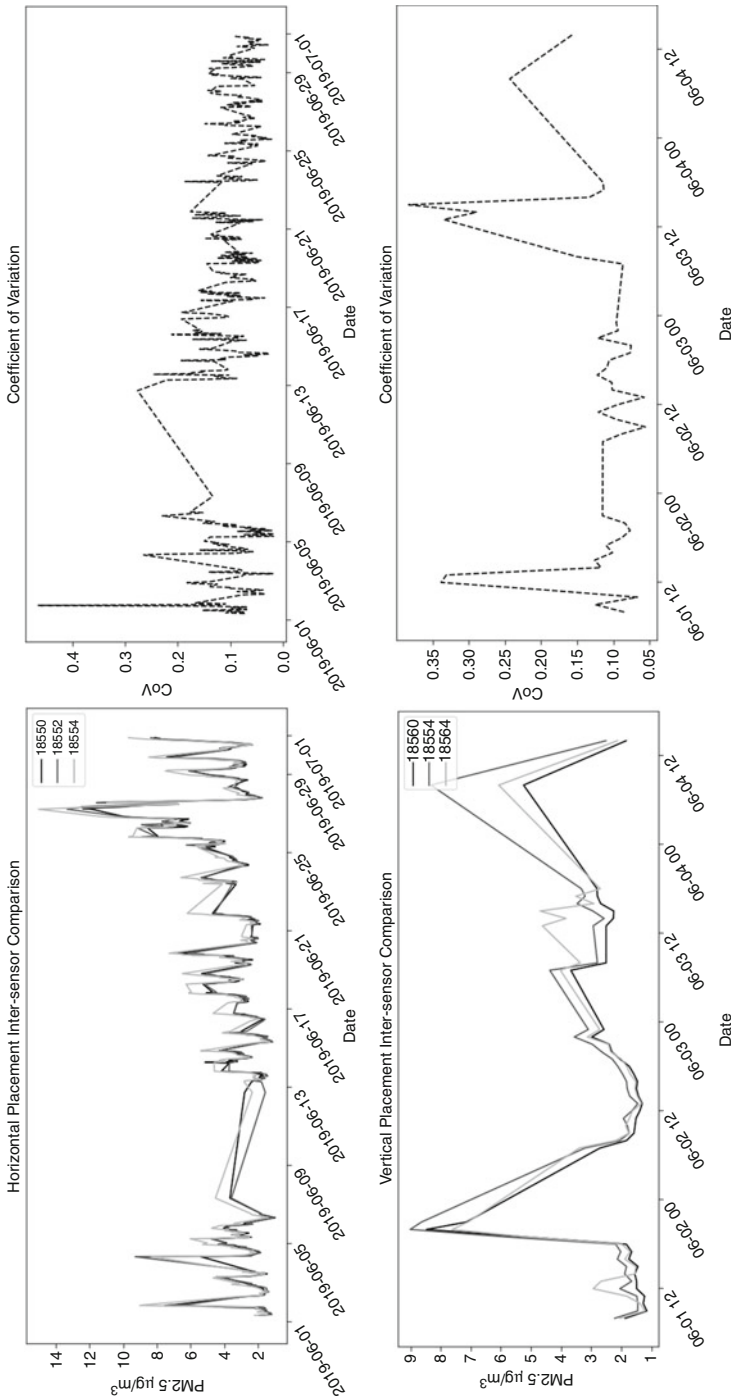


Fig. 14.3 PM_{2.5} measurements and coefficients of variations from an inter-sensor comparison exercise on sensors placed at varying horizontal and vertical distances

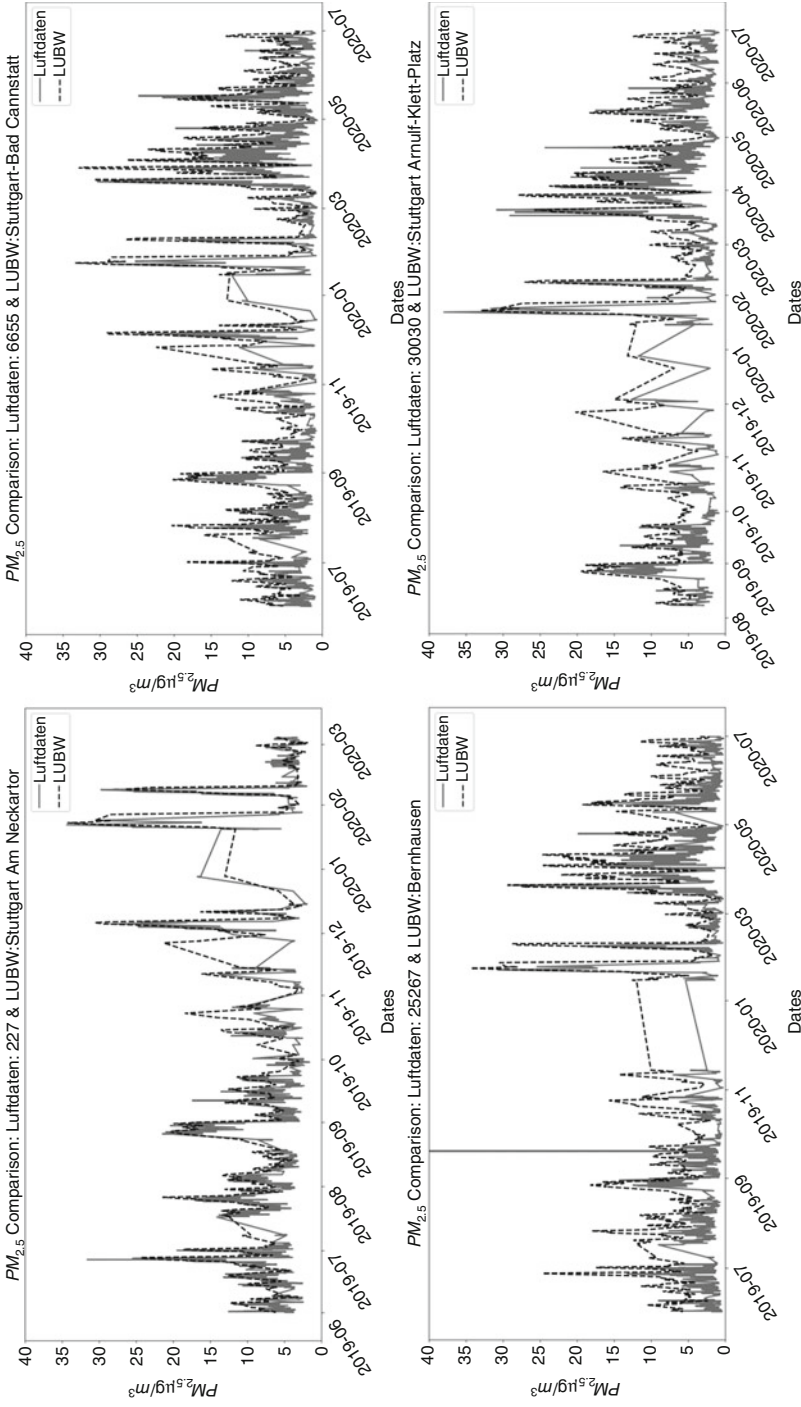


Fig. 14.4 A comparison of low-cost sensors' (solid line) and LUBW reference stations' (dashed line) $PM_{2.5}$ measurements. The four plots are from low-cost sensors that have the highest correlation with the respective nearest reference station measurements

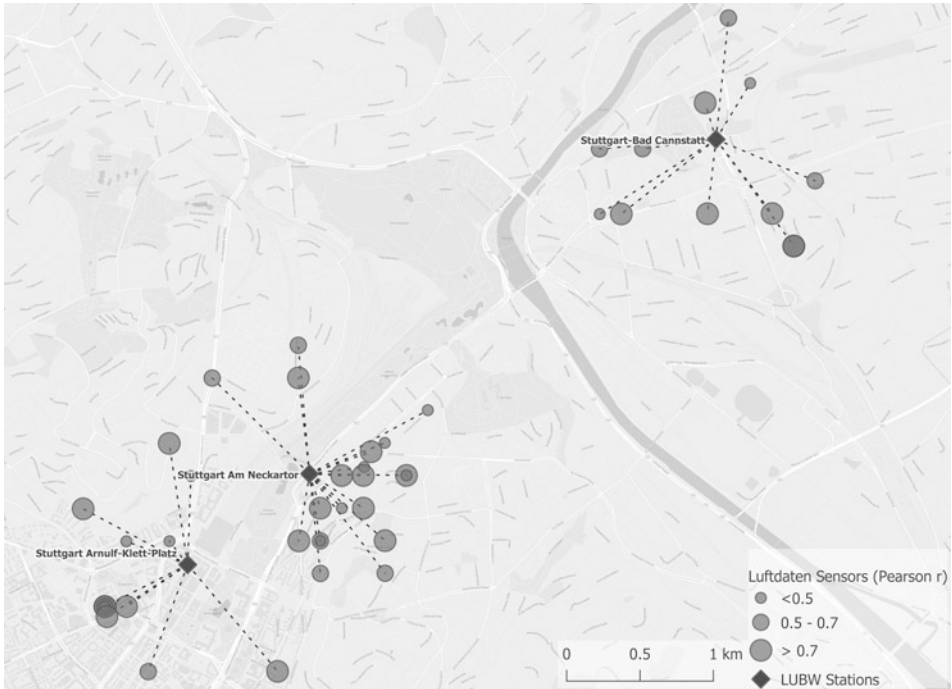


Fig. 14.5 A map of Stuttgart city centre showing Luftdaten low-cost sensors that are within 1 km radius of LUBW sensors. The symbol size represents the Pearson correlation coefficient (r) of PM_{2.5} measurements recorded between June 2019 and June 2020

reference stations are highly correlated regardless of the distance. In such an urban environment, the sensors are influenced by the distance to the road network and the type of roads in the vicinity (Table 14.2).

We select the highest correlated sensors for each reference station and fit the measurements using multilinear regression shown in Eq. (14.2). We examine the relationship between the sensors in Table 14.3 and reference stations for the period June 2019–June 2020, summer months June–September 2019 and winter months December 2019–March 2020. The sensors showed good linear correlation over the whole period (R^2 values 0.52–0.64) but lower correlations during winter (R^2 values 0.38–0.58) as seen in the comparison charts in Fig. 14.6. This is due to SDS011 sensors not having a heating mechanism to eliminate water droplets in the measuring chamber which negatively affects their performance. The best results are the warmer period with R^2 values ranging from 0.62 to 0.71. The scatterplots of the multilinear fittings are shown in Fig. 14.7.

A space-time cube created using ArcGIS pro with data aggregated to daily averages shows that for the period between August 2018 and August 2020, there were 758 unique sensors in the study area. However, all the sensors had data gaps, and 47% of the

Table 14.2 Long PM_{2.5} measurement comparison for low-cost sensors within a 1 km radius of each LUBW reference station

Reference station	Bernhausen	Ludwigsburg	Stuttgart Am Neckartor	Stuttgart Arnulf-Klett-Platz	Stuttgart-Bad Cannstatt	
<i>Code</i>	<i>DEBW042</i>	<i>DEBW024</i>	<i>DEBW118</i>	<i>DEBW099</i>	<i>DEBW013</i>	
<i>Type</i>	<i>Background</i>	<i>Background</i>	<i>Traffic</i>	<i>Traffic</i>	<i>Background</i>	
<i>Setting</i>	<i>Suburban</i>	<i>Suburban</i>	<i>Urban</i>	<i>Urban</i>	<i>Urban</i>	
Low-cost sensors within 1 km radius	3	2	20	15	12	
Pearson <i>r</i>	MIN	0.22	0.69	0.13	-0.15	0.18
	MEAN	0.45	0.72	0.57	0.52	0.64
	MAX	0.68	0.75	0.77	0.76	0.78
RMSE	MIN	6.04	6.09	4.14	5.50	4.55
	MEAN	39.25	7.06	16.01	6.77	6.76
	MAX	104.76	8.03	121.01	9.46	15.23

Table 14.3 Low-cost sensors with the highest correlation for each reference station

Reference station	Bernhausen	Ludwigsburg	Stuttgart Am Neckartor	Stuttgart Arnulf-Klett-Platz	Stuttgart-Bad Cannstatt
<i>Type</i>	<i>Background</i>	<i>Background</i>	<i>Traffic</i>	<i>Traffic</i>	<i>Background</i>
<i>Setting</i>	<i>Suburban</i>	<i>Suburban</i>	<i>Urban</i>	<i>Urban</i>	<i>Urban</i>
<i>Luftdaten sensor ID</i>	25,267	34,589	227	30,030	6655
<i>Distance (metres)</i>	204	920	246	506	510
Pearson <i>r</i>	0.68	0.75	0.77	0.76	0.78
RMSE	6.95	6.09	4.14	5.50	6.06
No. of observations	2994	1245	1746	2507	3206

observations had to be estimated by interpolating values based on the temporal trend of each sensor. Out of the 758 sensors, only 452 sensors were transmitting data as of August 2020. This could indicate a high failure rate of the sensors or mishandling by users. A different number of clusters were evaluated to extract patterns in the dataset with two clusters giving the optimum results. From the time series clustering results in Fig. 14.8, sensors that are near major roads, Cluster 2, have similar trends with higher values than sensors in the background. Leveraging on the large number of sensors shows potential in mapping pollution trends in space and time. For example, in Fig. 14.9 the sensors in Cluster 2 were able to detect lower PM_{2.5} concentration levels during the lockdown period (March–August 2020) due to Covid-19 compared to the same period in 2019.

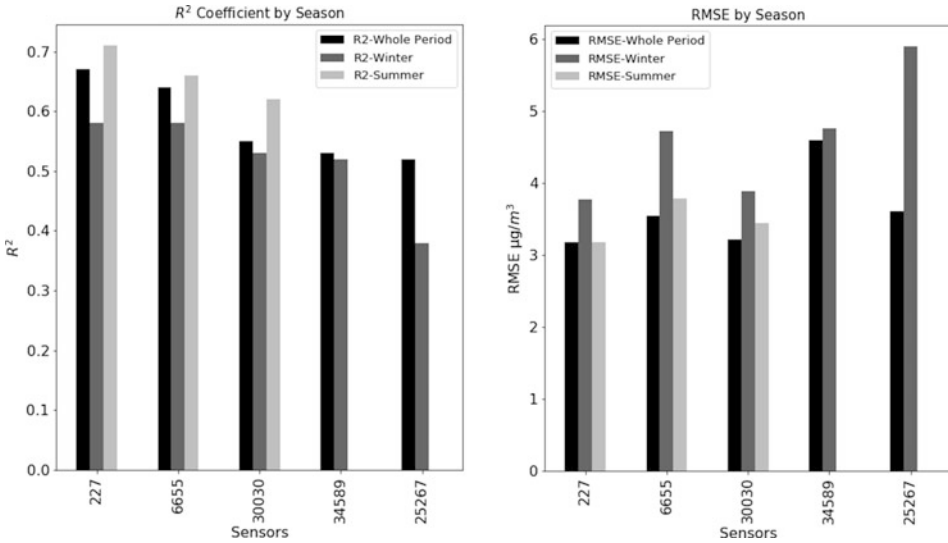


Fig. 14.6 Seasonal comparison of R^2 and RMSE statistics

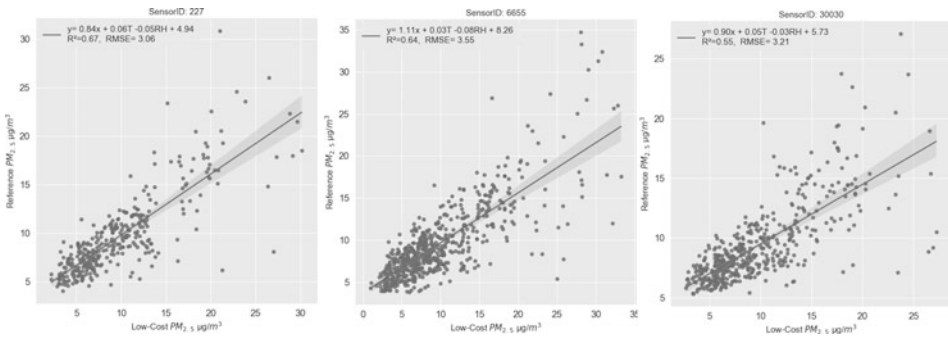


Fig. 14.7 Multilinear fitting results of hourly average reference and PM_{2.5} measurements after correcting for relative humidity and temperature effects

14.4 Conclusions

Low-cost sensors have the potential to improve spatial-temporal monitoring of PM pollution and supplement information from the costlier reference monitoring stations. The sensors exhibit stability and high correlation to reference measurements with varying degrees of accuracy. Whereas the sensors have lower accuracies and data gaps, using them in a dense network provides a wide coverage necessary for analysing pollution patterns and trends. One major challenge is outlier detection since it is hard to separate high pollution events from erroneous recordings due to the sensor’s fault. In a crowd-

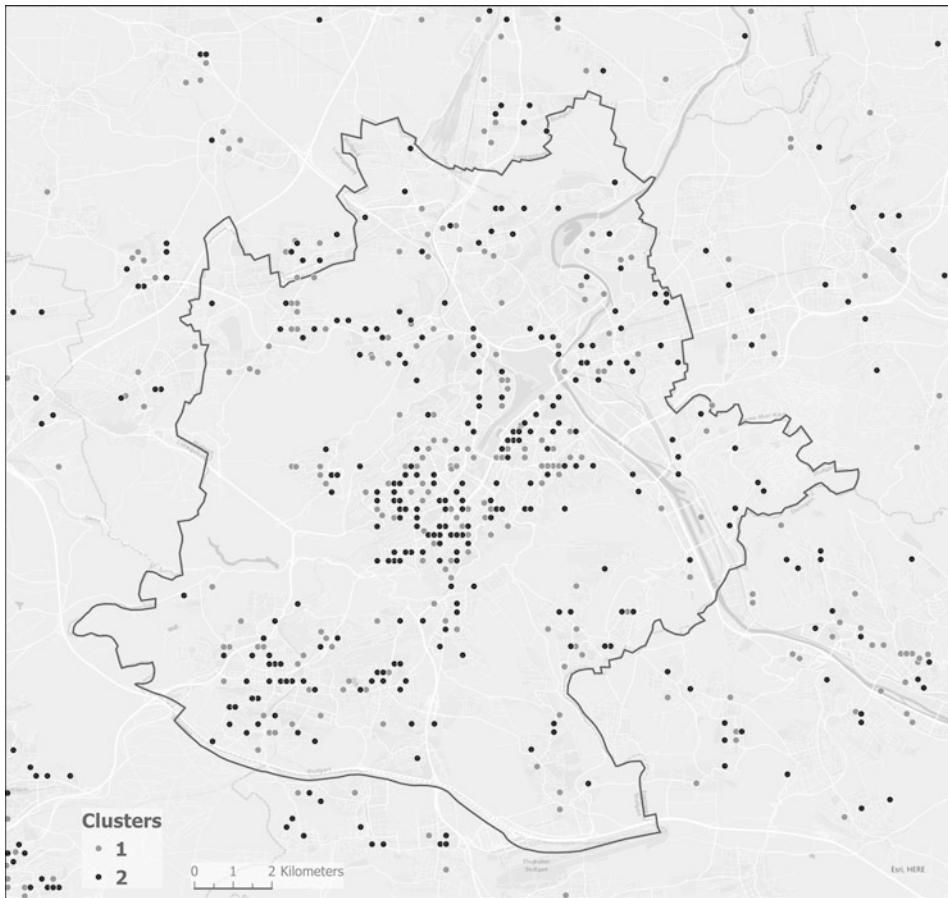


Fig. 14.8 A map showing the sensors clustered based on $PM_{2.5}$ values trends over 2 years from August 2018 to August 2020

sourced project like OK Lab Stuttgart, sensor installation and placement by the users is not standardized leading to measurements that are not representative of the location. For more accurate data collection, extensive sensor calibration, testing and robust outlier detection and removal techniques should be applied. Machine learning techniques could also be used to predict and fill in data gaps.

Acknowledgement The project “iCcity: intelligent city” is sponsored by the Federal Ministry of Education and Research (BMBF) under the promotion code 13FH9I011A and supervised by the project executing organization VDI Technologiezentrum GmbH for the BMBF.

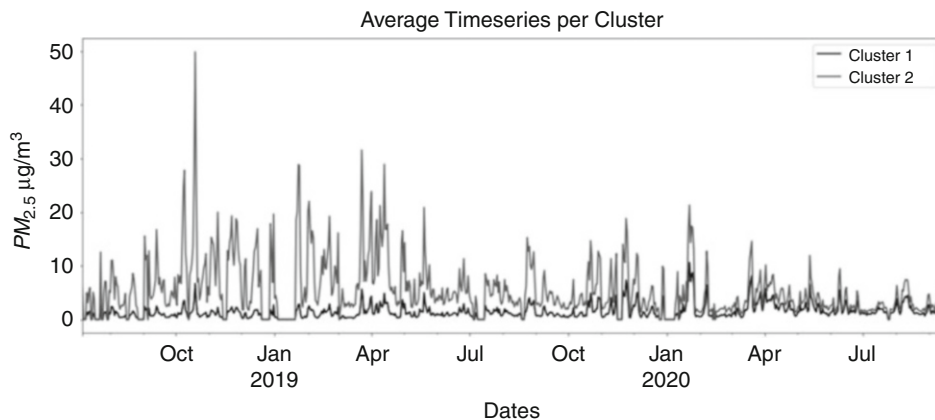


Fig. 14.9 Time series plot showing the average PM_{2.5} values for each cluster

References

- Badura, M., Batog, P., Drzeniecka-Osiadacz, A. and Modzel, P. (2018), “Evaluation of Low-Cost Sensors for Ambient PM 2.5 Monitoring”, *Journal of Sensors*, 2018, 1-16, *Journal of Sensors*, Vol. 2018, pp. 1–16.
- Bulot, F.M.J., Russell, H.S., Rezaei, M., Johnson, M.S., Ossont, S.J.J., Morris, A.K.R., Basford, P.J., Easton, N.H.C., Foster, G.L., Loxham, M. and Cox, S.J. (2020), “Laboratory Comparison of Low-Cost Particulate Matter Sensors to Measure Transient Events of Pollution”, *Sensors (Basel, Switzerland)*, Vol. 20 No. 8.
- Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D. and Bartonova, A. (2017), “Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?”, *Environment international*, Vol. 99, pp. 293–302.
- EU (2008), “Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe”, *Official Journal of the European Union*.
- Liu, H.-Y., Schneider, P., Haugen, R. and Vogt, M. (2019), “Performance Assessment of a Low-Cost PM2.5 Sensor for a near Four-Month Period in Oslo, Norway”, *Atmosphere*, 10(2), 41, *Atmosphere*, Vol. 10 No. 2, p. 41.
- Nova Fitness (2015), “Laser PM2.5 Sensor specification. Product Model: SDS011” (accessed 10 October 2020).
- Sousan, S., Koehler, K., Thomas, G., Park, J.H., Hillman, M., Halterman, A. and Peters, T.M. (2016), “Inter-comparison of Low-cost Sensors for Measuring the Mass Concentration of Occupational Aerosols”, *Aerosol science and technology the journal of the American Association for Aerosol Research*, Vol. 50 No. 5, pp. 462–473.
- WHO (2016), *Ambient air pollution: A global assessment of exposure and burden of disease*, Geneva, Switzerland.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

