



Citizen-operated mobile low-cost sensors for urban PM_{2.5} monitoring: field calibration, uncertainty estimation, and application

Amirhossein Hassani^{a,*}, Núria Castell^{a,*}, Ågot K. Watne^b, Philipp Schneider^a

^a NILU - Norwegian Institute for Air Research, P.O. Box 100, Kjeller, 2027, Norway

^b IVL Swedish Environmental Research Institute, Valhallavägen 81, Stockholm 11428, Sweden

ARTICLE INFO

Keywords:

Air quality monitoring
Particulate matters (PM_{2.5})
Mobile low-cost sensors
Uncertainty analysis
Machine Learning
Snifferbike
Citizen science

ABSTRACT

Research communities, engagement campaigns, and administrative agents are increasingly valuing low-cost air-quality monitoring technologies, despite data quality concerns. Mobile low-cost sensors have already been used for delivering a spatial representation of pollutant concentrations, though less attention is given to their uncertainty quantification. Here, we perform static/on-bike inter-comparison tests to assess the performance of the Snifferbike sensor kit in measuring outdoor PM_{2.5} (Particulate Matter < 2.5 μm). We build a network of citizen-operated Snifferbike sensors in Kristiansand, Norway, and calibrate the measurements using Machine Learning techniques to estimate the concentrations of PM_{2.5} along the city roads. We also propose a method to estimate the minimum number of PM_{2.5} measurements required per road segment to assure data representativeness. The collocation of three Snifferbike kits (Sensirion SPS30) at the monitoring station showed a RMSD of 7.55 μg m⁻³. We approximate that one km h⁻¹ increase in the speed of the bikes will add 0.03 - 0.04 μg m⁻³ to the Standard Deviation of the Snifferbike PM_{2.5} measurements. We estimate that at least 27 measurements per road segment are required (50 m here) if the data are sufficiently dispersed over time. We recommend calibrating the mobile sensors when they coincide with reference monitoring stations.

1. Introduction

Over half of the world's population lives in urban areas, and urbanization is still increasing (Ritchie et al., 2018). Cities are the engines of social and economic growth, but they are responsible for more than 70% of air pollutant emissions (Lederbogen et al., 2011). Densely populated areas impose numerous environmental and policy-making challenges, such as air pollution, waste management, and resource usage (Chowdhuri, Pal, & Arabameri, 2022; Pal et al., 2021; Satterthwaite, 1993). According to the "European Environment Agency", ecological and sociological footprints of poor air quality are still environmental threats to European cities despite the recent achievements in reducing emission levels and pollutants concentration (Ortiz et al., 2020). It is estimated that 90% of city dwellers are exposed to pollutant concentrations above the thresholds accepted as healthy (World Health Organization, 2016); further, annually, 8 - 13 billion Euros is estimated to be required to address the air pollution-caused health effects. Among these pollutants is PM (Particulate Matter), which significantly impacts the human respiratory and cardiovascular systems. In 2018, it is estimated that 417,000 premature deaths were connected to long-term

exposure to PM_{2.5} (Particulate Matter < 2.5 μm) across all of Europe (Ortiz et al., 2020).

In Nordic countries, ~75% of the total population dwells in urban regions, and municipalities aim to provide clean air for all residents (Geels, 2015; Im et al., 2019). Reaching this aim demands assessing the urban air quality and emission sources and investments in devising sustainable plan actions to reduce the emissions. Many small and medium towns in Nordic countries, however, have limited access to resources and adequate funding to operate the traditional reference monitoring stations. Even in the bigger municipalities where such resources are available, reference stations are geographically sparse, and many urban environments within the municipality are under-represented in terms of air quality monitoring. The recent advances in Low-Cost Sensor systems/kits (LCSs) and mobile technologies for monitoring air quality have offered opportunities for real-time monitoring of air pollutants such as NO_x, CO, O₃, and mainly PM at higher densities than typically possible with traditional reference equipment (Aleixandre & Gerboles, 2012; Castell et al., 2017). In addition, the low cost of the LCS platforms allows for investigation of the individual air quality experience (Cheriyian et al., 2020; Sm et al., 2019)

* Corresponding authors.

E-mail addresses: ahas@nilu.no (A. Hassani), ncb@nilu.no (N. Castell).

<https://doi.org/10.1016/j.scs.2023.104607>

Received 9 January 2023; Received in revised form 20 March 2023; Accepted 21 April 2023

Available online 23 April 2023

2210-6707/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

and citizen engagement (Citizen Science) in air quality monitoring (Ekman et al., 2021; Mahajan & Kumar, 2020; Mahajan et al., 2020).

Compared to LCSs (throughout this paper, by LCS we mean LCS system/kit as they usually integrate a collection of sensors) used for measuring gaseous pollutants, PM LCS have been proven to show higher reliability and ease of operation in field applications (Lewis et al., 2018; Vogt et al., 2021). The use of PM LCS platforms in research projects, decision-making, and Citizen Science initiatives for PM monitoring has been rapidly growing in recent years (Giordano et al., 2021). Despite the substantial improvements in calibration and accuracy of the PM LCSs, they still have several limitations, and they are not adequate for regulatory purposes due to data quality and precision (Morawska et al., 2018). The application of PM LCS is mainly limited to cases where relative changes in particulate levels are desired over a relatively short period — days to a few months (Morawska et al., 2018) or when indicative measurements are enough. In some cases, data from LCS have been successfully assimilated with modeling data to improve the model predictions' accuracy, e.g. Schneider et al. (2017).

PM LCSs are prone to meteorology-driven inaccuracy, particularly Relative Humidity (RH), drifts in calibration, and limited lifetime (Alfano et al., 2020; Vogt et al., 2021); on the other hand, little information is usually provided by the sensor manufacturers about the actual real-world performance of the PM LCSs. The difficulty in the evaluation of the performance of PM LCS and all-in-one calibration procedures (Vogt et al., 2021; Wang et al., 2023) has led to numerous studies dedicated to calibrating and analysis of the reproducibility, scalability, and quantification of the performance of PM LCSs, under different operating conditions (Alfano et al., 2020; Kang et al., 2021; Liu et al., 2017; Venkatraman Jagatha et al., 2021; Vogt et al., 2021); however, the majority of this research is focused on static LCSs, and until recently, in-field evaluation of portable PM LCS in a non-stationary environment has received little attention.

Comparing the data from nine air monitoring stations with standard PM_{2.5} instruments with the data of 100 mobile SDS019-TRF PM LCSs mounted on taxis in Jinan, the capital city of Shandong Province, China (Qin et al., 2020) showed a Mean Absolute Error between 12.71 and 32.84 $\mu\text{g m}^{-3}$ (Correlation Coefficient = 0.70 - 0.88). A two-phase data calibration method was proposed by Lin et al. (2018) consisting of a linear and a nonlinear component (multiple least squares and Random Forest) to calibrate the Mosaic, a mobile air quality monitoring system mounted on urban buses that measures PM_{2.5}. They demonstrated a 6.1% - 133% improvement in precision using their method compared to traditional approaches for PM_{2.5} calibration. However, their sensors were calibrated against mid-cost Dyls DC1700 air quality sensors, mounted on the buses, not the stationary official monitoring stations. Some mobile networks do not offer such mid-cost sensors. Wang et al. (2023) conducted extensive stationary co-location experiments in two American cities, New York, and Boston, to evaluate the performance of various parametric and Machine Learning (ML) calibration algorithms. To assess their transferability, the best-performing calibration models were applied to a network of mobile sensors in Boston. They observed models trained in a stationary environment did not transfer well to mobile sensors.

Depending on the platform used for mobile measurements, such as cars, public transport vehicles, bicycles, and humans, measurements may be biased in space and time (Samad et al., 2020). The data are primarily measured during rush hours and sensors pass more frequently on specific routes. Additionally, contingent on the speed of movement, the measurement of portable LCSs may be subject to rapid changes in micro-climatic parameters, leading to lower accuracy and difficulty in differentiating signals from the noise. The difference between raw PM_{2.5} measurements and reference signals (Dyls DC1700) increased as speed increased from 3 km h⁻¹ to 9 km h⁻¹ in experiments conducted with three SDS011 sensors mounted on a rotary table at 5 cm, 10 cm, and 15 cm distances from the center (Lin et al., 2018). Despite these limitations, the mobility of LCSs for air quality monitoring purposes adds to the

spatial coverage and makes them suitable alternatives for personal exposure to air pollution studies (DeSouza et al., 2020; Lee et al., 2020; Lim et al., 2019; Lu et al., 2021; Mihăiță et al., 2019; Miskell et al., 2018; Santana et al., 2021; Van den Bossche et al., 2016; Wesseling et al., 2021). To take full advantage of mobile air-quality monitoring technology, it is necessary to address the sampling bias issue and analyze the reliability of the mobile LCSs in response to mobility speed or interference with meteorological parameters.

Here, we use the case of Kristiansand municipality (Norway) and 10 PM LCSs mounted on bikes, combined with official data from two reference monitoring stations to advance the current knowledge on the in-field calibration and application of mobile LCSs. Citizens operate the sensors to monitor urban air quality. The purpose of this paper is to evaluate the performance of portable LCSs for monitoring PM under a variety of different operating conditions and to make recommendations about how citizen science can efficiently utilize these tools. We investigate how reliable the Snifferbike PM LCSs are for measuring PM_{2.5}, particularly at higher cycling speeds, how we can improve their accuracy through smart calibration techniques, and how we can address the spatio-temporal sampling biases in the measurements. To answer these questions, we (1) analyze the results of a series of co-location and fixed/on-bike inter-comparison tests to quantify the reliability and accuracy of the Snifferbike portable sensors in measuring PM_{2.5}, specifically in response to sensor speed; (2) calibrate the mobile sensor measurements using ML techniques to enhance their accuracy and evaluate the transferability of the models, and (3) develop a method to find the optimal number of measurements per road segment to assure the validity of the spatial representativeness of the measurements. Finally, we discuss the technological and societal limitations involved in mapping the variability of PM_{2.5} along Kristiansand's roads using the measurements from citizen-operated mobile LCSs. This paper presents a comprehensive evaluation of the limitations of LCS for air quality monitoring and provides insights into how LCSs can be used to improve the spatial coverage of air quality data and enable citizen engagement.

2. Methods

Fig. 1 illustrates the work sequence for this paper. The flowchart outlines a process for evaluating the accuracy of ambient PM_{2.5} measured by Snifferbike sensors in Kristiansand, Norway. The process involves using reference monitoring station data, as well as Snifferbike sensor kits to collect on-bike data. The data is then evaluated by official PM_{2.5} data and sensor-to-sensor inter-comparisons, and a ML model is used for remote calibration. Lastly, the data accuracy is assessed and the PM_{2.5} concentrations along the roads are mapped.

2.1. Study location

We carried out this study in Kristiansand, a seaside city and municipality in southern Norway. With a population of ~112,000 inhabitants as of January 2020, Kristiansand is the fifth largest city in Norway. Kristiansand is a pilot city in the NordicPATH (Nordic participatory, healthy, and people-centered cities: <https://nordicpath.nilu.no/>, accessed on 16 Nov 2022) research project, focusing on co-monitoring and citizen-engaged urban planning for shaping healthier cities in Nordic countries.

There are two reference monitoring stations in Kristiansand: (1) *Bjørndalssletta*, a traffic station with proximity to the E18 highway, owned by the Norwegian Public Roads Administration, and (2) *Stener Heyerdahl*, an urban background station owned by Kristiansand municipality (Fig. 2b). These two reference stations are equipped with CEN (European Committee for Standardization)-approved PM analyzers, in line with the criteria of European Standards. The measured data are continuously monitored by NILU (Norwegian Institute for Air Research, <https://www.nilu.com/>, accessed on 16 Nov 2022) for quality assurance purposes. PM_{2.5} and PM₁₀ are routinely measured at the station (time

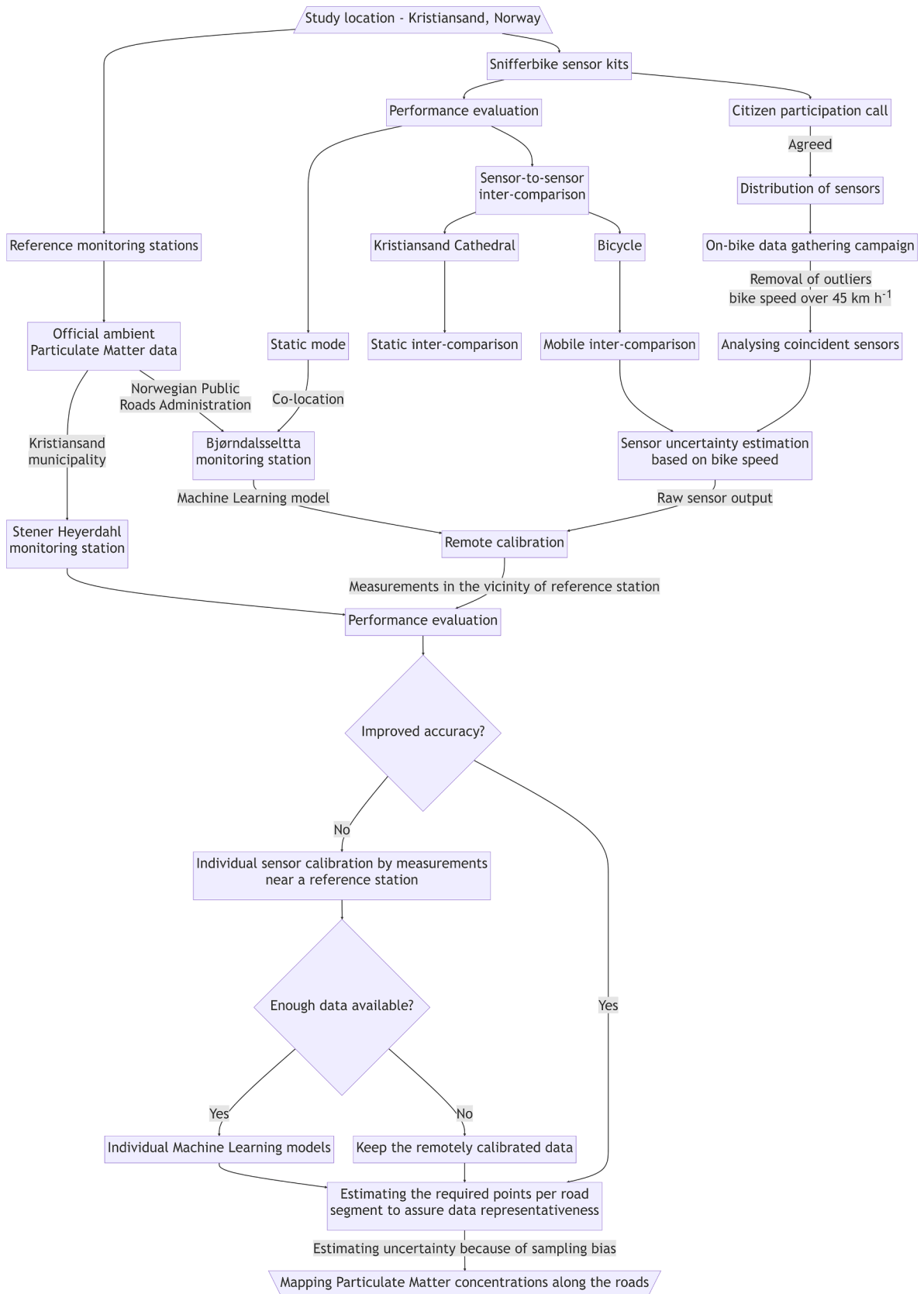


Fig. 1. The workflow proposed for mapping ambient PM_{2.5} using Snifferbike sensors along the roads in Kristiansand, Norway.

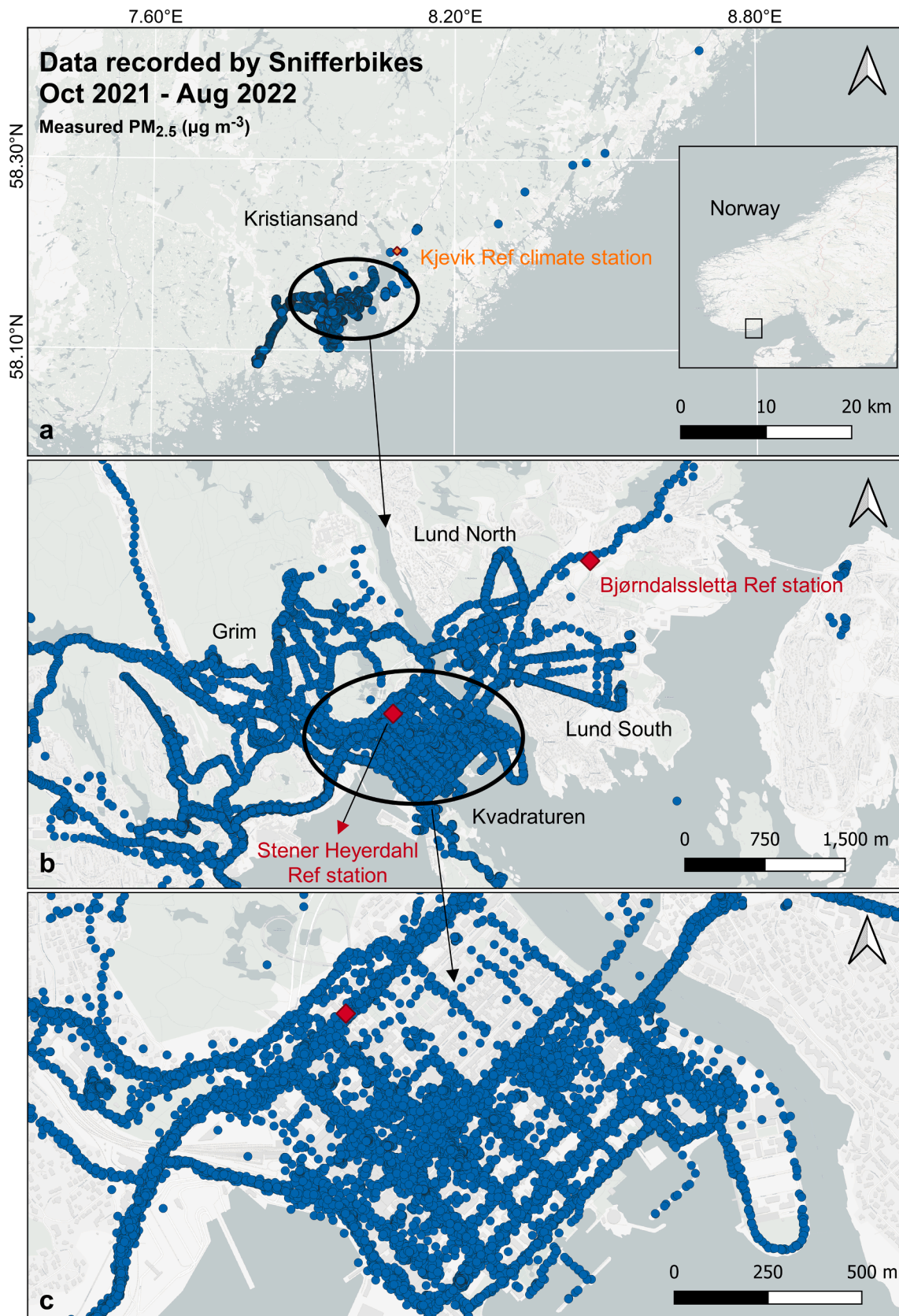


Fig. 2. Spatial distribution of the $PM_{2.5}$ concentrations measured by the 10 Snifferbike sensors in Kristiansand, Norway. a, b, and c show the city and data at different scales. The geographical location of the reference Kjevik (airport) climate station and reference monitoring stations is represented in a and b, respectively.

zone: UTC + 1) using GRIMM EDM 180 (with a reproducibility > 97% of the total measuring range).

2.2. Snifferbike sensor kits

In November 2020, we purchased 11 LTE-M-enabled Snifferbike sensor kits developed by SODAQ (<https://sodaq.com/>, accessed on 16 Nov 2022) and Civity (<https://civity.nl/en/>, accessed on 16 Nov 2022) with an approximate price of €270,00 per unit. SODAQ AIR is available in the market as an updated version of these sensors. Snifferbike sensor kits integrate a u-Blox EVA-M8 concurrent positioning engine, a Bosch BME680 temperature and humidity sensor, and a Sensirion SPS30 (<https://sensirion.com/products/product-categories/particulate-matter/>, accessed on 16 Nov 2022) dust sensor for measuring PM (PM₁, PM_{2.5}, and PM₁₀).

According to the manufacturer user guide, the most accurate measurements are the mass concentration of particles between 0.3 and 2.5 µm in size. The manufacturer-provided error for the mass concentrations in the range of 0 - 100 µg m⁻³ is ±10 µg m⁻³ and for the range of 100 - 1000 µg m⁻³, it is reported to be ±10%. Previous studies also have shown that Sensirion SPS30 is not efficient in measuring the coarse fraction of the PM (Kang et al., 2021; Kuula et al., 2020; Vogt et al., 2021). Accordingly, we focused here only on PM_{2.5} measurements.

Snifferbike sensors measure the air quality every 10s and send the raw data (pollutant concentration, temperature, RH, position, and Unix timestamp) to the City Innovation Data Platform (CIP) developed by the IoT Civity company for calibration and data quality purposes. The data then is pushed to the NILU (<https://nordicpathlive.nilu.no/>, accessed on 16 Nov 2022) sensor data platform operated by NILU under open data sharing standards to make the data accessible to third-party users and for research purposes. Each sensor received a unique ID number on the NILU sensor data platform, ranging between 155 and 165 (11 units).

2.3. Initial co-location and inter-comparison tests

To evaluate the performance of the sensors in static mode, we co-located three sensor units at the Bjørndalsseltta monitoring station from October 2020 to January 2021 (units 155, 156, 158). Due to technical issues in data transfer, the data from sensors in some periods, especially in November 2020 (Sensors 155 and 156), are missing

(Supplementary Fig. 4).

If a calibration scheme from one sensor is going to be applied to the rest of the sensors, it will be necessary to ensure that the sensors' accuracy is in the same range. To evaluate the consistency between the sensors, we performed a series of static and on-bike mobile sensor-to-sensor inter-comparison tests. First, all 11 sensor kits were statically installed and inter-compared at the Kristiansand Cathedral (Lon/Lat = 7.9946°E/58.1466°N) from 26 March 2021 to 28 April 2021 (Fig. 3a). Further, to examine the uncertainty analysis of the sensors' measurements against speed and inter-comparability of the sensors at mobile conditions, we mounted all 11 Snifferbike sensors on the same bike. We rode the bike on two episodes: (1) 31 May 2021, from 16:30 until 18:20 UTC, and (2) 10 June 2021, from 16:20 until 17:20 UTC (Fig. 3b).

2.4. Sensor network

Following the initial co-location and on-bike inter-comparison, an open call for citizen participation in air quality monitoring was held in Kristiansand; 63 out of 80 participants agreed to the statement "I want to install a sensor to monitor air quality where I live and/or travel." Eventually, ten mobile sensors (Snifferbike) were distributed among the most interested participants in October 2021. To reduce the impact of interference with cycling dust and other debris kicked up while cycling, we mounted the sensors on the bikes in such a way that their inlets faced forward on handlebars, away from the wheels and other moving parts.

The raw data measured by the mobile sensors from 01 September 2021 until 06 August 2022 retrieved from the NILU sensor platform is represented in Fig. 2 and Supplementary Fig. 1. A total of 69,914 measurements were recorded from the sensors during the mentioned period. We removed the measurements at high speeds, assuming 45 km h⁻¹ as the threshold, for two reasons: (1) it is rare that bike drivers reach higher speeds in urban environments, and (2) we assumed that the high speed of the bikes negatively affects the accuracy of the Snifferbike sensor measurements. We calculated the speed between every two consecutive measurements based on the time difference and cartesian distance; if the calculated speed was more than 45 km h⁻¹, both measurements were removed. Following this pre-processing step, 69,071 measurements remained in our analysis (1.2% data loss). A similar approach is adopted in other studies, for example, Wesseling et al. (2021) removed values over 45 km h⁻¹.

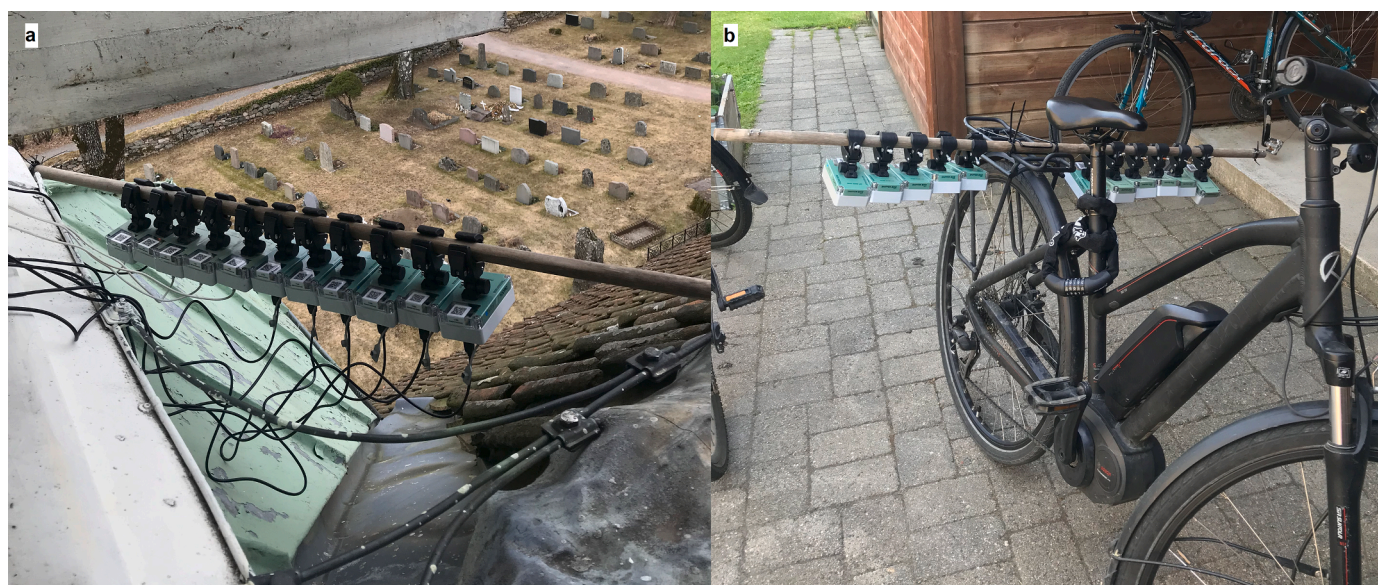


Fig. 3. 11 Snifferbike PM sensors (Sensirion SPS30) inter-comparison tests. a, Kristiansand Tower Church (Kristiansand Cathedral) – from 2021-03-26 to 2021-04-28. b, mounted on a bike and riding on two episodes: (1) 2021-05-31, from 16:30- until 18:20 UTC, and (2) 2021-06-10, from 16:20 until 17:20 UTC. Photo by Solvor B. Stølevik, Kristiansand municipality.

Theoretically, Snifferbike sensor kits should not deliver any information when immobile. Nearly 29% of the sensor records were calculated to have a speed of less than 2 km h^{-1} . These low speeds can be attributed to when the bikes are nearly immobile, or the cyclist has little movement, for example, behind a traffic light or during parking. Additionally, sensors do not stop immediately, and it takes some time with speed = 0 before they cease measuring. We keep those measurements as they were a considerable portion of our records.

Supplementary Fig. 2 shows the daily, weekly, and seasonal distribution and sampling bias in the recorded measurements, all in the local winter time (UTC + 1). There are two periods with no data, late December 2021, and Mid-April 2022; nearly 98% of the samples are recorded between October 2021 and June 2022. Close to 91% of the data were measured on the working days, and 97% were recorded between 6:00 and 20:00 local time.

2.5. Sensor data calibration

The preliminary analysis of the co-location and on-bike testing results showed a bias in the measurements relative to official measurements (discussed in detail in the results section). The temperature and RH at factory calibration environments differ from the actual meteorological field conditions, leading to bias/deviation of the commercially available optical PM sensors. To reduce such bias/deviation in measurements, we calibrated the data against the reference monitoring station data. Conventionally, some sensors are initially required to be co-located at the official monitoring stations and calibrated against reference-grade instrumentation. In-field measurements of the sensors can then be calibrated using the calibration equations. However, this calibration scheme may not be adequate for the long-term utilization of the sensor (networks) as sensors output may drift. Typically, the initial co-location of the sensors may not capture all environmental conditions that sensors can face in a particular environment (Alfano et al., 2020). Physical limitations, such as a lack of space or power supply near the reference stations, can hinder sensor co-location. Co-locating all sensors is also resource-intensive and time-consuming. Accordingly, we kept one Snifferbike unit (ID 156) constantly co-located at the Bjørndalssletta monitoring station from October 2021 and adopted remote calibration schemes to calibrate the sensors' output. We calibrated unit ID 156 using ML regression algorithms and applied the developed ML model to the rest of the sensors, operated by citizens.

We trained tree-based ML models using the hourly-averaged data from sensor 156 co-located at Bjørndalssletta from 2021-09-01 to 2022-08-06. We later applied the trained models to the measurements of the ten other sensors. Tree-based ML models are very efficient in capturing non-linearities and reducing the effect of outliers (Breiman, 2001). This efficiency, however, comes at the cost of higher computational load and less interpretability.

In addition to the signal from the LCS 156, we used hourly averaged air temperature ($^{\circ}\text{C}$), wind velocity (m s^{-1}), dew point temperature ($^{\circ}\text{C}$), and RH index ($^{\circ}\text{C}$), defined as air temperature minus dew point temperature as predictor data. Meteorological data were recorded at the Kjevik airport climate station (elevation: $\approx 17 \text{ m}$), located 17 km from the Kristiansand city center, northeast (Fig. 2a). Supplementary Fig. 3 provides information on the ambient wind speeds and directions. Data were retrieved from the Integrated Surface Dataset (Global) of the National Centers for Environmental Information (<https://www.ncei.noaa.gov/access/search/data-search/global-hourly>, retrieved in August 2022) in FM-15 Surface Meteorological Airways Format. The final input table prepared for model training had 6,837 rows. Air temperature and RH data also were available from the LCSs; however, we used the official and quality-controlled meteorological station data to keep the predictors consistent and avoid biases (especially in RH) in measurements from the LCSs.

We used the Python package "XGBoost" for the implementation of gradient-boosted trees ML approach: <https://xgboost.readthedocs.io>

</en/latest/index.html>, accessed on 16 Nov 2022 (Chen et al., 2016) to establish data-driven relation (model) between the co-located sensor measurements, auxiliary predictors, and the reference station $\text{PM}_{2.5}$ measurements as target value. "colsample-bytree" (searched randomly in the range between 0.001 and 1), "learning-rate" (0.001 - 1), "max-depth": (1 - 10), "no-estimators" (1 - 500), and "subsample" (0 - 1) were the optimized hyper-parameters using Sci-Kit Learn "Randomized Search CV" function (Pedregosa et al., 2011). Hyper-parameters were randomly sampled 15 times (number of iterations) using a 5-fold cross-validation splitting strategy to find the optimal set of hyper-parameters. We repartitioned the input data before each new iteration (with equal weights). We trained the models using the optimized set of hyper-parameters. 10-fold cross-validation evaluation metrics, including MAE (Mean Absolute Error), RMSE (Root Mean Square Error), R^2 (explained variance; $R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$); y_i is the true value, \hat{y}_i is the prediction, and \bar{y}_i is the mean true value), and maximum error (Prediction - Observation) were used to evaluate the performance of the trained models.

The trained models using unit 156 were deployed to calibrate the measurements from the rest of the Snifferbike sensors. To further assess the performance of the adopted calibration scheme, we compared the calibrated Snifferbike measurements against the official $\text{PM}_{2.5}$ measurements of the Stener Heyerdahl station (Fig. 2b). To do so, we found the calibrated measurements in a circle with a radius of 200 m around the Stener Heyerdahl. We calculated the RMSE and R^2 against the reference measurements for hourly averaged data. The comparisons showed that only some sensor units have sufficient measurements to extract meaningful inferences at such a radius (units 155, 157, 158, 160, and 164). This comparison also showed that the adopted remote calibration scheme does not necessarily improve the reliability of the measurements (this is discussed/quantified in detail in the results section). On this account, we applied the above-described calibration scheme only to the sensor units which did not have enough data records around the Stener Heyerdahl station, and we trained individual ML XGBoost models for the rest of the sensor units.

These models were prepared using the sensor measurements recorded inside a circle of 200 m radius around the Stener Heyerdahl station and the corresponding official $\text{PM}_{2.5}$ and predictors' values. We assumed that a 200 m radius around the monitoring station is short enough to assure that measurements are within the spatial representativeness of the monitoring station. This assumption is in line with previous studies, e.g., Shi et al. (2018) found that the $\text{PM}_{2.5}$ representative area of each station varies between 0.25 km^2 ($r \approx 282 \text{ m}$) and 16.25 km^2 ($r \approx 2,274 \text{ m}$), however, is less than 3 km^2 ($r \approx 977 \text{ m}$) for more than half of the stations, using high-resolution observations of 169 urban stations in North China (Nov 2015 to Feb 2016). A hyper-parameter tuning and evaluation approach, like what we explained for unit 156 was used for calibrating these individual sensors; only the hyperparameters were randomly sampled ten times to reduce the run time.

2.6. Minimum measurements per road segment calculation

We used the calibrated Snifferbike measurements to map the spatial distribution of $\text{PM}_{2.5}$ along the roads in Kristiansand. As mentioned, the measurements are geographically and temporally biased; simply averaging the measurements can be misleading. Previous studies have assumed a hypothetical lower limit for the required measurements per road segment to ensure the averaging reliability. For example, eight by Van den Bossche et al. (2016), although that study is focused on Black Carbon, or 25 and 50 by Wesseling et al. (2021).

However, we adopted a quantitative approach for defining the minimum required measurements per road segment. To do so, first, we divided the road map in Kristiansand into arbitrary segment lengths (50

m in this case). We assigned the Snifferbike measurements to the closest segment as the GPS coordinates received from the sensors are noisy; they do not locate precisely on the road segments. Following this, for each minimum measurement threshold (ranging between 1 and 150), we calculated the mean of the Standard Deviation (SD) of the measured $PM_{2.5}$ values in all road segments. For example, for the minimum measurement threshold 10, we calculated the SD of the measured $PM_{2.5}$ in all road segments with ≥ 10 $PM_{2.5}$ measurements (i.e., data count) from the sensors. We repeated this procedure for a minimum measurement threshold of 11 and continued until 150. Mean SDs (y) were then plotted against the minimum measurement thresholds (x , 1 - 150). Mean SD initially increases by an increase of minimum threshold and later begins to reduce. The minimum threshold was chosen where the mean of SDs starts to decline; in other words, the minimum threshold was the x value at which the mean SD reaches its maximum.

3. Results and discussion

3.1. Initial co-location at the reference monitoring station

The times series of the $PM_{2.5}$ measured by the sensor co-located at the Bjørndalssletta monitoring station (sensors 155, 156, and 158) and the reference monitoring station measurements are shown in Supplementary Fig. 4. Fig. 4 also visualizes the scatter plots of sensors factory-calibrated $PM_{2.5}$ signal (and PM_{10}) against the official data from the

station. During the co-location period, the average air temperature, RH, and wind velocity were 3.72°C , 87.3% , and 1.85 m s^{-1} , respectively. The average $PM_{2.5}$ measured by GRIMM EDM 180 was $10.65\text{ }\mu\text{g m}^{-3}$ during the co-location period. The mean Pearson Correlation Coefficient (r) between the sensors' $PM_{2.5}$ output and the reference measurements was 0.75. The mean factory-calibrated RMSD (Root Mean Square Deviation) of the three sensors from the reference measurements was $7.55\text{ }\mu\text{g m}^{-3}$. This value was reduced to $2.41\text{ }\mu\text{g m}^{-3}$ following a Robust Linear (using built-in MATLAB "fitlm" function) calibration of the sensor output. The robust estimator we used is less sensitive to outliers than traditional estimators like ordinary least squares (OLS). We used a method called Iteratively Reweighted Least Squares (Robust Linear Regression) with a "bisquare" (biweight) weight function and a default setting of 4.685 as the tuning constant.

Overall, three sensors underestimated the $PM_{2.5}$ during the co-location analysis. The matrix plot shown in Supplementary Fig. 5 also represents the factory-calibrated sensor-by-sensor comparison. The correlation between the sensors varies between 0.86 and 0.96. Supplementary Figs. 6 and 7 show similar results for the continuation of the co-location of Snifferbike unit 156 at Bjørndalssletta monitoring station, starting from October 2021 to August 2022. The sensor underestimates $PM_{2.5}$ values relative to the reference instrument (RMSD = $6.73\text{ }\mu\text{g m}^{-3}$). A PM pollution episode, which took place between 20 and 27 March 2022 (with PM_{10} daily mean concentrations of $50 - 100\text{ }\mu\text{g m}^{-3}$), is also distinctly captured by the sensors (Tsyro et al., 2022).

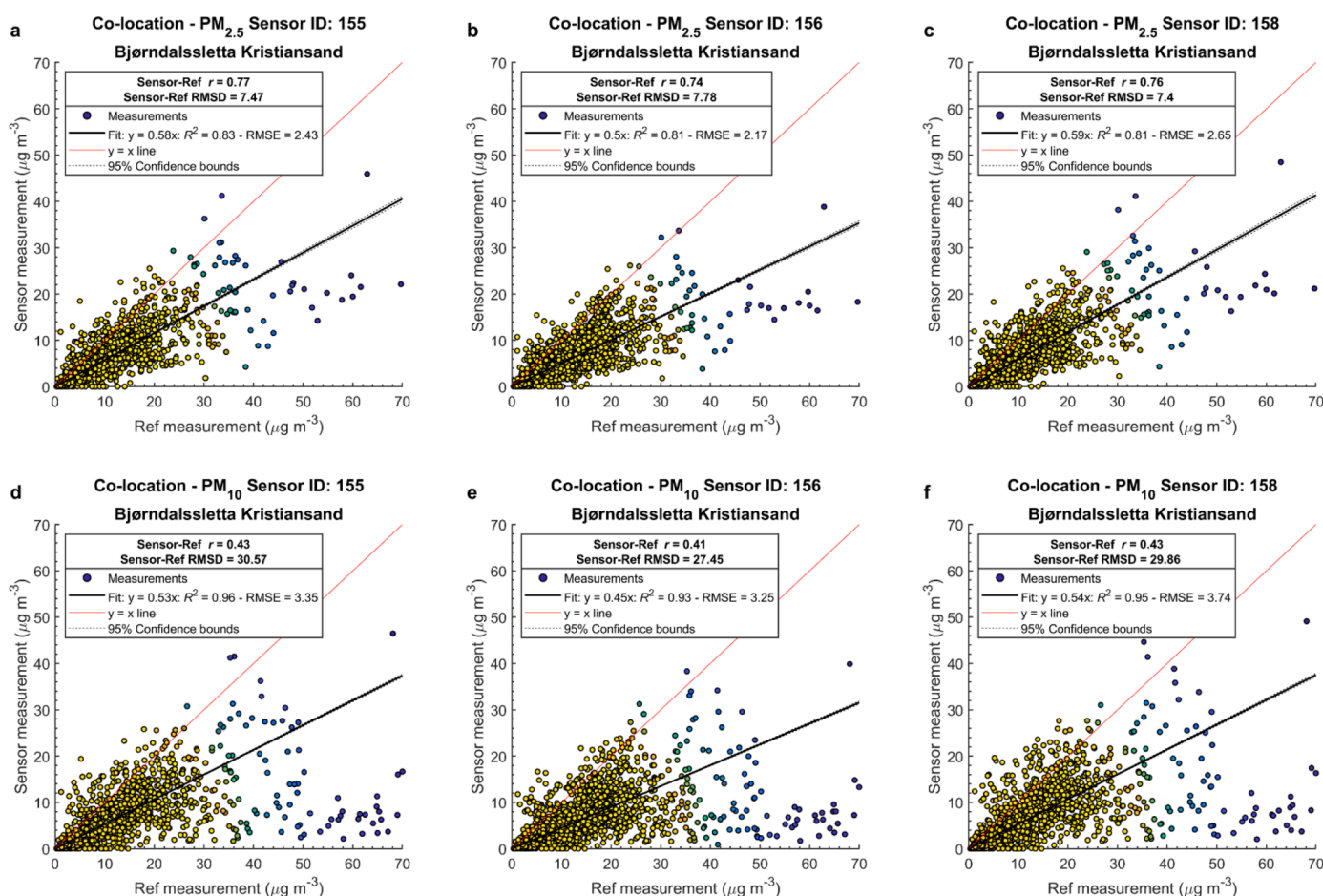


Fig. 4. Comparison of reference-equivalent (Bjørndalssletta station) $PM_{2.5}$ and PM_{10} data against the measurements of the three co-located Snifferbike sensors. The dots' color shows the scatter density. The blue color represents lower density while yellow is a high-density region. We have used Iteratively Reweighted Least Squares fitting (Robust Linear regression) using the "bisquare" ("biweight") weight function with the default tuning constant of 4.685 to reduce the outlier effect. For further details, readers are referred to MATLAB "fitlm" documentation (<https://uk.mathworks.com/help/stats/fitlm.html>, retrieved in December 2022). r : Pearson Correlation Coefficient. R^2 : Coefficient of determination. RMSD: Root Mean Square Deviation. Sensor-Ref r and RMSD values shown in the legends are calculated based on factory-calibrated sensor outputs (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

In a review of the outdoor environmental setting's influence on the correlation of PM_{2.5} LCS with the reference instrument, Kang et al. (2021) found a median r^2 (including r_s that were converted to r^2) equal to 0.72 (25th Percentile = 0.53, 75th Percentile = 0.85) according to 80 studies. Parameters influencing the performance of PM LCSs sensors, such as environmental conditions, co-location procedure, and ambient PM concentrations are site-dependent. Thus, comparing the Sensirion SPS30 accuracy observed here with similar studies may be only an overall indication of the sensors' performance. Hong et al. (2021) reported an average factory-calibrated RMSD = 5.64 $\mu\text{g m}^{-3}$ for five Sensirion SPS30 PM_{2.5} sensor units co-located at Penzhen reference station (reference instrument: BAM-1020 FEM), Taiwan compared to official hourly measurements from 2019-11-13 to 2020-12-30. Roberts et al. (2022) collected Sensirion SPS30 PM_{2.5} data from 29 July 2019 to 12 December 2019, in a reference site in Columbia, South Carolina, and observed a RMSD = 3.029, and a bias = -1.61 $\mu\text{g m}^{-3}$ for one-hour averaged measurements; however, outliers were detected/removed in their analysis. Overall, the factory-calibrated PM_{2.5} measurements of the Sensirion SPS30 used in our analysis in static mode show a good correlation with the reference instrument and bias compared to the

literature. Yet, the Robust Linear calibration of the sensors in static mode substantially improves the accuracy of the measurements.

Kristiansand is a seaside city, and RH is over 60% most of the year. Due to the sensitivity of the optical PM LCSs sensors to RH (Feenstra et al., 2019), we particularly analyzed the role of RH on the factory-calibrated measurements of Snifferbike sensor kits with integrated Sensirion SPS30. It is widely accepted in the literature that PM LCS bias increases with a higher RH due to the negative impact of water vapor on light scattering properties of aerosol particles (Brattich et al., 2020; Jayaratne et al., 2018). The validity of Sensirion SPS30 PM_{2.5} measurements at various environmental conditions is evaluated in other studies. It has been shown that they have a superior performance in handling high humidity levels compared to similar sensors. For example, relative to AirSenseEUR and Airly PM LCSs, Vogt et al. (2021) concluded that Ensense PM_{2.5} LCSs (which integrates the Sensirion SPS30) show the lowest bias (less than 5 $\mu\text{g m}^{-3}$) and dependence on changes in RH, through a two-month co-location of the sensors at a reference monitoring station in Oslo. Hong et al. (2021) likewise observed high fluctuations for Mean Normalized biases (MNBs) or deviation from reference instrument data (BAM-1020) for the Plantower PMS5003 and

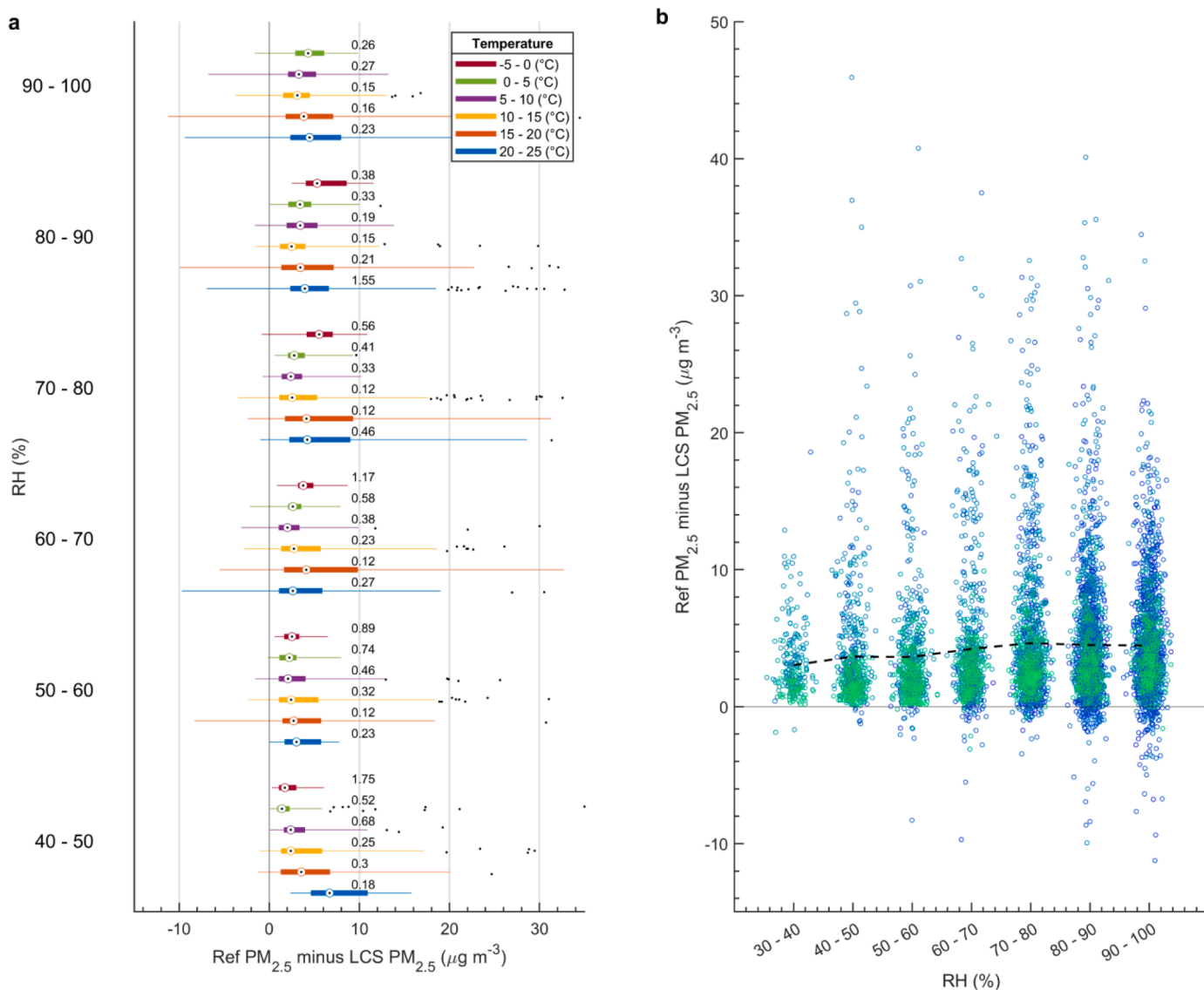


Fig. 5. Role of Relative Humidity (RH) on the absolute Snifferbike PM_{2.5} sensor bias from official measurements at Bjørndalsselta station. a, the magnitude of bias for each RH bin depends on the air temperature. b, swarm chart of PM_{2.5} bias against RH bins. The distribution of the bias data is visualized against each RH bin based on the kernel density estimate of bias. The black dotted line represents the average bias for each discrete RH bin.

Honeywell HPM115S0 at increasing RH values while those of Sensirion SPS30 $PM_{2.5}$ showed slight increase when the RH was higher than 80%.

Our results align with those studies as we observe here that the mean bias of the three sensors (during all co-location periods) against different RH bins is less than $10 \mu\text{g m}^{-3}$ with respective mean and median of 4.17 and $2.94 \mu\text{g m}^{-3}$, respectively (Fig. 5). Additionally, our results suggest that air temperature plays a significant role in this bias from the reference measurements. Typically, the bias (reference measurement minus LCS measurement) at the lower temperature bands ($-5 - 0^\circ\text{C}$) and higher air temperature bands ($20 - 25^\circ\text{C}$) is more than the medium temperature bands ($10 - 15^\circ\text{C}$). Summary statistics are presented in Supplementary Table 1. Our results are relevant to only three Sensirion SPS30 in Kristiansand (Norway); this provokes further investigation on other sensor brands and other environmental/meteorological conditions.

3.2. Sensor uncertainty estimation

The time series of the $PM_{2.5}$ measured by all 11 sensors inter-compared at the Kristiansand Cathedral shows a very good sensor-to-sensor correlation between the Snifferbike sensor kits ranging from $r = 0.96$ to $r = 0.98$ (Supplementary Figs. 8 and 9). A similar analysis for Sensirion SPS30 conducted by Vogt et al. (2021) also revealed an excellent inter-sensor correlation ($r > 0.99$) for $PM_{2.5}$ measured at the Kirkeveien air quality monitoring station in Oslo, Norway (from 28 August to 19 October 2020). The results of the on-bike sensor inter-comparison tests are presented in Supplementary Fig. 10. Air

temperature/RH measured at the Kjevik airport station was $20.5^\circ\text{C} / 59.21\%$ and $18.5^\circ\text{C} / 75.07\%$ during the 2021-05-31 and 2021-06-10 on-bike tests, respectively. Measurements of $PM_{2.5}$ at different speeds show high variability during the on-bike tests with an average SD of $0.81 \mu\text{g m}^{-3}$ (Fig. 6). We estimate an increase of 1 km h^{-1} will increase the SD of the 11 sensors $PM_{2.5}$ measurements by $0.03 \mu\text{g m}^{-3}$.

Similar to the on-bike inter-comparison test, we attempted to estimate the variability in the coincident Snifferbike (ridden by citizens) sensors' measurements as a function of the bike speed. Bike-based measurements are assumed to be coincident if they are located within a $50 \times 50 \text{ m}$ grid at the same minute. The SD of the factory-calibrated $PM_{2.5}$, measured by bikes passing the same grid cell of $50 \times 50 \text{ m}$ at the same time (i.e., at the same minute), shows an increase of $0.04 \mu\text{g m}^{-3}$ with one 1 km h^{-1} increase in the average speed of the bikes (Fig. 7a). $PM_{2.5}$ SD was calculated only when at least five coincident measurements were available in the same grid-cell-minute. This rate is $0.01 \mu\text{g h km}^{-1} \text{ m}^{-3}$, higher than the estimated uncertainty from the on-bike inter-comparison tests. This rate is $0.08 \mu\text{g m}^{-3}$ per km h^{-1} for calibrated coincident bike measurements (Fig. 7b). At a speed of 25 km h^{-1} (SD = 1, and $2 \mu\text{g m}^{-3}$), this represents normalized standard deviations of 19% and 28.5% (normalized to the interquartile range of measured and calibrated data), respectively. The variability in sensor measurements also shows a statistically meaningful relation with $PM_{2.5}$ concentration (Fig. 7c and d). The SD of coincident data increases by $0.04 \mu\text{g m}^{-3}$ and $0.02 \mu\text{g m}^{-3}$ by $1 \mu\text{g m}^{-3}$ increase in the average $PM_{2.5}$ measured by the coincident Snifferbike sensors before and after

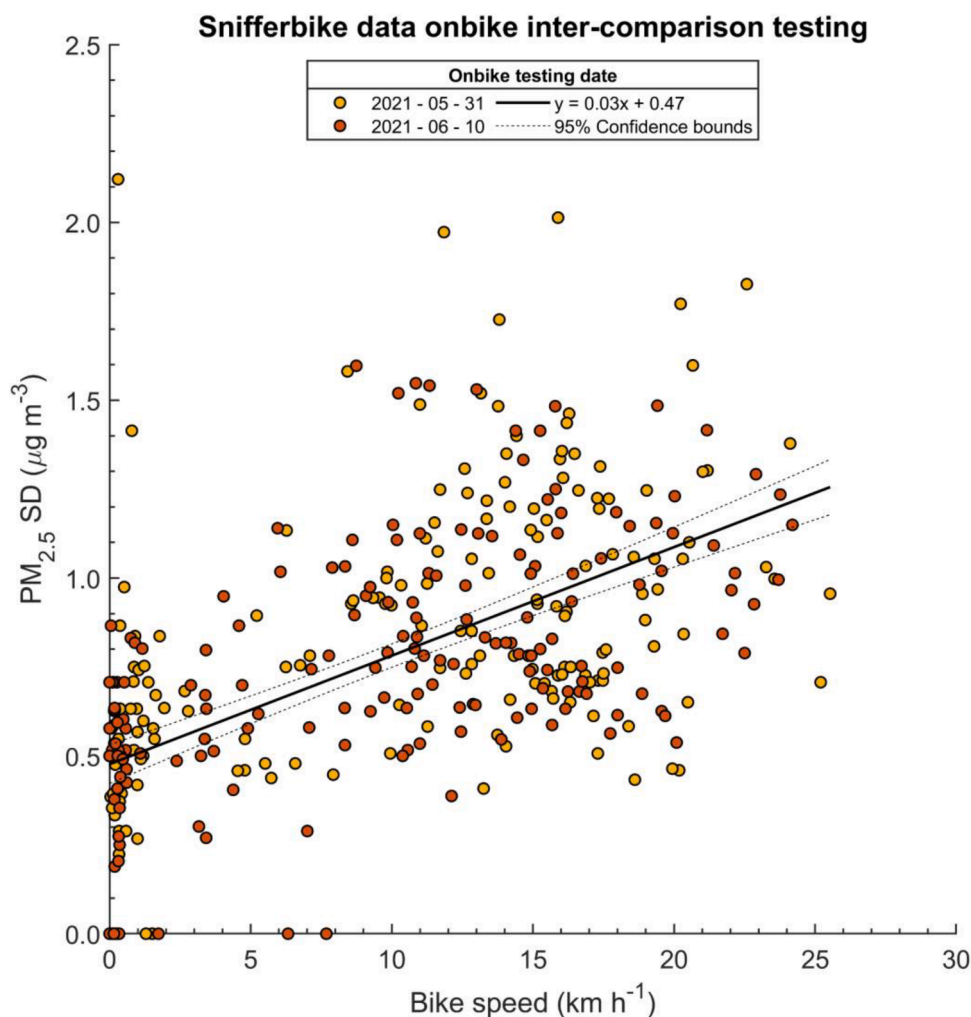


Fig. 6. Standard Deviation (SD) of the factory-calibrated measurements of all 11 Snifferbikes against the bike speed during the 2021-05-31 and 2021-06-10 on-bike tests.

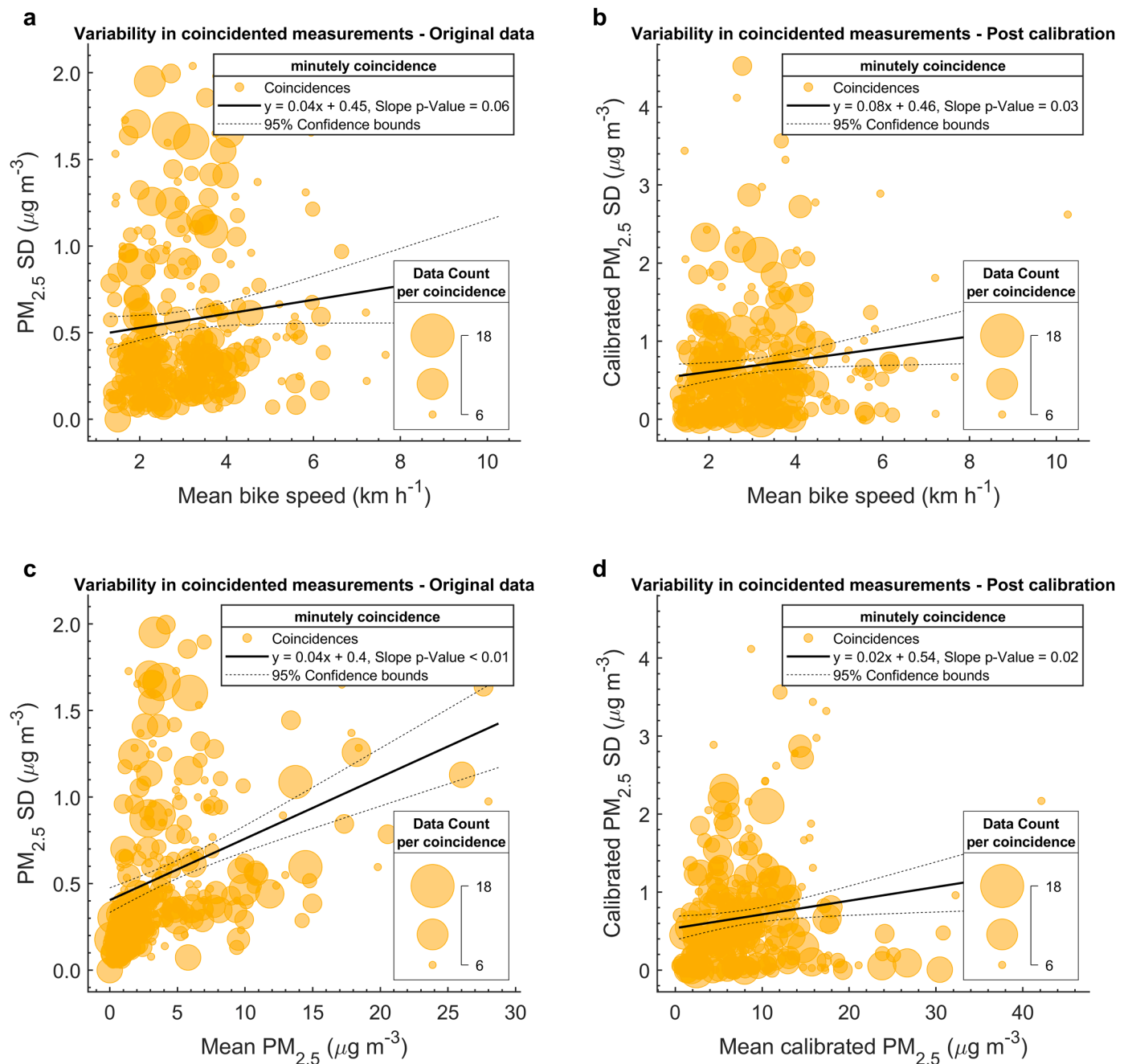


Fig. 7. Standard Deviation (SD) of coincident Snifferbike $PM_{2.5}$ measurements. Bike-based measurements are assumed to be coincident if they are located within a 50×50 m grid at the same minute. a and b, SD of pre- and post-calibration of $PM_{2.5}$ signal against the mean bike speed of coincident bikes. c and d, SD of pre- and post-calibration of the $PM_{2.5}$ signal against the average raw and calibrated $PM_{2.5}$ signal from the coincident cycles. The size of bubbles represents the number of coincident bike measurements located within a grid at the same minute used for calculating SD and mean speed/ $PM_{2.5}$. $PM_{2.5}$ SD was calculated only when at least five coincident measurements were available in the same grid-cell-minute. P-values are rounded to the nearest two decimal digits.

calibration, respectively. According to the factory documentation, this increase in sensor uncertainty can be due to the higher error of Snifferbikes at higher $PM_{2.5}$ concentrations.

Overall, the results here confirm that the uncertainty of the $PM_{2.5}$ measurements of Sensirion SPS30 sensors increases at higher speeds; this increase in variability can be assumed negligible at usual bike speeds (5 - 15 $km\ h^{-1}$), contingent on the low ambient $PM_{2.5}$ concentrations (less than $30\ \mu g\ m^{-3}$). In a relevant study, Wesseling et al. (2021) analyzed $5000\ h\ cell^{-1}$ coincident $PM_{2.5}$ measurements in grids of 50×50 m in Utrecht, Netherlands. They assumed uncertainty on the order of $6\ \mu g\ m^{-3}$ in the average concentration of the $PM_{2.5}$ measured by Snifferbikes in mobile mode. This assumption was based on the

similarity of the scatter of the hourly coincident sensor data mean (x) against SD (y) to the scatter of co-located Snifferbike sensor data (x) against reference instrument (y). According to the range of observations, they also concluded that the difference among sensors in mobile mode is limited. However, the increase rates calculated above are general and may occasionally hide the high levels of variation in coincident bike measurements within each grid. The SD of both the calibrated and factory-calibrated response of the coincident sensors may be up to $2\ \mu g\ m^{-3}$, even at low speeds of $2\ km\ h^{-1}$.

3.3. Calibration using Machine Learning models

Supplementary Tables 2 and 3 summarize the hyper-parameter tuning and cross-validation results of calibrating co-located sensor 156 (XGBoost ML model). 10-fold cross-validation R^2 and RMSE for the final trained ML model were 0.78 and $3.97 \mu\text{g m}^{-3}$, respectively. For $\text{PM}_{2.5}$ sensor calibration, Mahajan and Kumar (2020) found that Support Vector ML models outperformed other ML models including Linear regression, Random Forest, and Artificial Neural Networks with an average RMSE of $3.39 \mu\text{g m}^{-3}$ and an average R^2 of 0.87. The study used 10 Smart Citizen sensor kits (Plantower PMS 5003) co-located with a GRIMM EDM 107 reference-grade instrument. To estimate the concentration of pollutants based on real-time data, Hemamalini et al. (2022) presented a Smart Drone monitoring solution with bidirectional gated recurrent unit (Bi-GRU) network modeling. As a result of the model evaluation, RMSE values of 1.63, 1.35, and $1.47 \mu\text{g m}^{-3}$ were found for the dumpsites of solid waste, residential areas, and industrial areas, respectively.

We estimated the importance of predictors for the XGBoost model. The signal from Snifferbike was the most important predictor (relative variable importance = 100%), followed by RH index (10.75%), dew temperature (10.63%), wind velocity (8.46%), and temperature (7.45%). The significance of predictors is primarily case-specific and depends on the size of the input training set, the hyper-parameter tuning scheme, and the choice of predictors.

The final trained model using sensor 156 was applied to factory-calibrated output from the 10 Snifferbike sensors. A total of 1,672 measurements were located within a circle of 200 m radius around the Stener Heyerdahl reference station. We assumed that the number of measurements from each Snifferbike should be more than 50 to evaluate that sensor against the official measurements. This threshold limited our analysis to sensor units 155, 157, 158, 160, and 164. Since the reference station measurements are hourly, we averaged the Snifferbike sensor measurements at hourly resolution and compared them with reference data (Supplementary Figs. 1 and 12). Compared to the reference measurements from the Stener Heyerdahl, applying the model calibration model from the sensor unit 156 to Snifferbikes 160 and 164 (respective RMSD values of 4.33 and $5.77 \mu\text{g m}^{-3}$) reduced the factory-calibrated measurement accuracy (respective RMSD values of 4.71 and $7.84 \mu\text{g m}^{-3}$). As mentioned in the methods, we trained separate XGBoost ML models for the sensors with enough measurements around Stener Heyerdahl (using Stener Heyerdahl official $\text{PM}_{2.5}$ data) to address this issue; the other five sensors were kept calibrated using unit 156.

Hyper-parameter optimization and 10-fold cross-validation results for individually calibrated sensors 155, 157, 158, 160, and 164 are summarized in Supplementary Table 4 and Supplementary Table 5, respectively. Similar to the predictor importance order for unit 156, factory-calibrated signal output (average relative variable importance = 100%), RH index (48.2%), dew temperature (32.98%), wind velocity (32%), and temperature (27.7%) were the most important predictors. Training these five individual ML models improved the calibration against the official measurement; all five models' hourly-averaged 10-fold cross-validated predictions are illustrated in Supplementary Figs. 13 and 14. Unfortunately, the measurements in the vicinity of the Bjørndalselva monitoring station were insufficient to evaluate these five models independently.

Based on our results, the initial co-location and/or remote calibration of Snifferbike sensors does not necessarily improve the accuracy of the measurements in the mobile mode. A similar conclusion is drawn by Wang et al. (2023), who concluded that model transferability is limited for the models trained in the stationary settings to the mobile setting due to differences in urban environments and climates.

3.4. Spatial mapping of the $\text{PM}_{2.5}$

The mean distance of the points to the center of 50 m road segments

was 7.7 m. According to the proposed method for estimating the minimum required points for each road segment, we found that at least 27 measurements per road segment are needed to make a robust estimation of $\text{PM}_{2.5}$ concentration along the roads. This value is represented as a cut-off line in Fig. 8 which shows the mean SD of $\text{PM}_{2.5}$ measurements across in road segment against the minimum measurement count in each road segment. Our results show that this cut-off depends on the road segment length choice. The higher the road segment length, the higher the cut-off (Supplementary Figs. 15 and 16).

The spatial distribution of the calibrated $\text{PM}_{2.5}$ measurements and the corresponding standard error (SD divided by $\sqrt{\text{sample size}}$) for each segment is shown in Fig. 9. The road segments in this map are 50 m, and at least 27 measurements are averaged in each segment. Overall, the (calibrated) $\text{PM}_{2.5}$ measured by the Snifferbike sensors was $9.36 \mu\text{g m}^{-3}$ per road segment (October 2021 - Aug 2022) over the whole city. Additionally, Supplementary Fig. 17 shows the spatial distribution of the calibrated $\text{PM}_{2.5}$ measurements at different day periods (with at least 27 measurements per road segment). The average of $\text{PM}_{2.5}$ measurements per road segment during the rush hour in the morning (6:00 - 10:59) and afternoon (16:00 - 22:00) was $9.5 \mu\text{g m}^{-3}$ and $10.87 \mu\text{g m}^{-3}$, respectively. In general, $\text{PM}_{2.5}$ concentration was higher in the denser region of the city (Kvadraturen). Wesseling et al. (2021) and Sm et al. (2019) found a similar spatio-temporal pattern in $\text{PM}_{2.5}$ concentrations and attributed this to the higher traffic load. However, our evaluation showed no statistically meaningful relationship between the road type/traffic load and measured $\text{PM}_{2.5}$ concentrations (Supplementary Figs. 18 and 19, and Supplementary Table 6). To do so, we compared the hourly-averaged calibrated Snifferbike $\text{PM}_{2.5}$ measurements with the corresponding traffic count data summed per hour (data were retrieved from Statens Vegvesen: The Norwegian Public Roads Administration, September 2022; <https://www.vegvesen.no/trafikkdata>). Sensor data were at a maximum of 200 m away from the Kristiansand municipality Inductive Loop vehicle counters. Results obtained here may partly be due to spatio-temporal sampling bias and insufficient data available from our sensor network, while Wesseling et al. (2021) analyzed 500 sniffer bikes in their analysis.

$\text{PM}_{2.5}$ concentrations show two diurnal peaks (bimodal distribution), one between 8:00 and 12:00 and the other between 16:00 and 23:00. The morning peak is likely caused by a lower boundary layer and fumigation effect (American Meteorological Society, 2020; Kompalli et al., 2014; Tiwari et al., 2013). In the afternoon, wood burning for household heating accounts for most of the peak. It has been found that wood burning is one of the main sources of fine PM emissions in Norway during the winter months (Wolf et al., 2021).

3.5. Study limitations and future research recommendations

- We conducted the on-bike inter-comparison tests only for two 1.5 h periods. More extended on-bike inter-comparison tests during different meteorological and environmental conditions can better evaluate the uncertainty in sensor measurements.
- We here quantified the uncertainty of the sensors against the bike speed. By more extended on-bike inter-comparison tests at controlled bike speeds, it would be feasible to investigate the role of varying micro-climatic parameters (e.g., air temperature and RH) on the uncertainty of the sensor PM measurements.
- There may be some errors in the estimation of air pollutant concentrations at the site due to drag caused by the bicycle's motion.
- The ambient wind speed and direction when the bikes are operating can affect the accuracy of the measurements, even at low speeds. Hyper-local data on wind speed and direction are necessary for a robust uncertainty estimation which were not available by Snifferbike sensor kits. The cost of installing and maintaining hyper-local wind speed sensors can be high.
- One shortcoming of this analysis can be the small size of the sensor network and the high temporal and spatial sampling bias in the

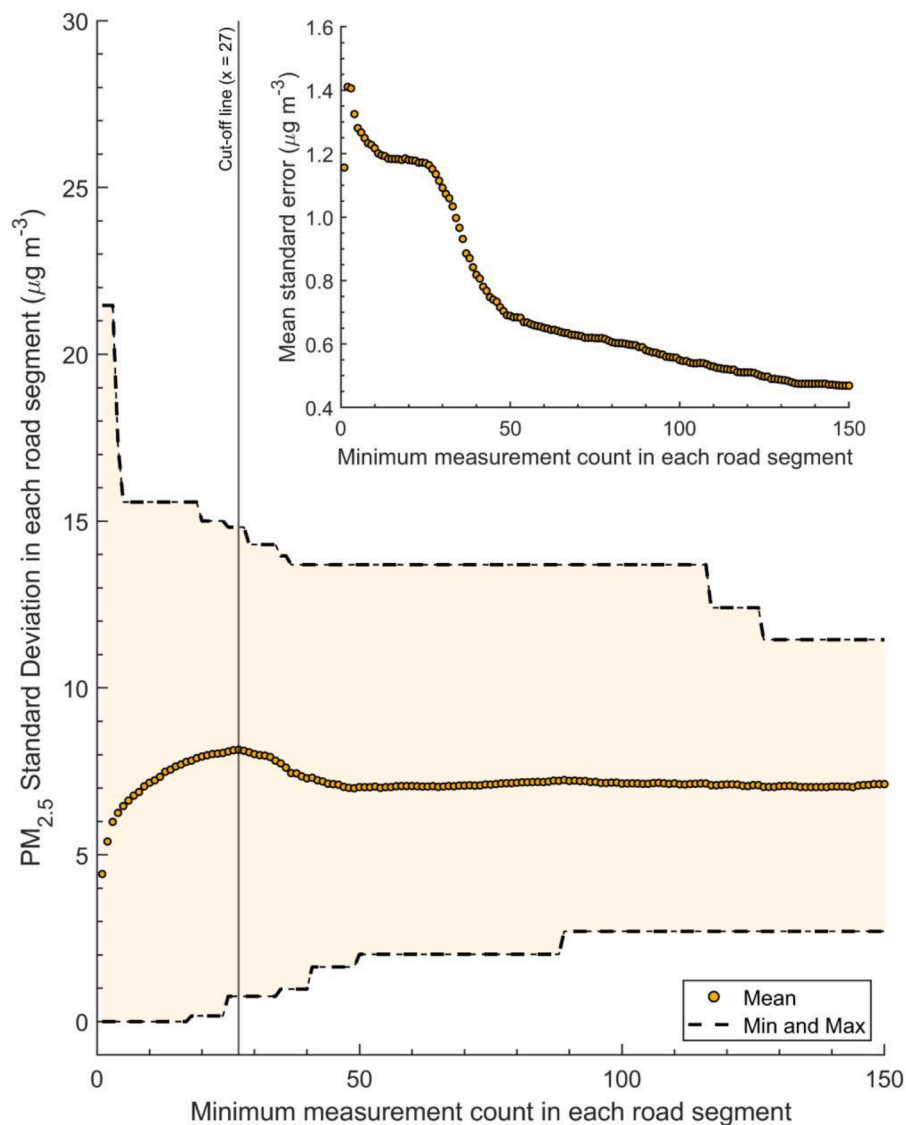


Fig. 8. Standard Deviation (SD) and standard error of Snifferbike $PM_{2.5}$ measurements in each road segment against the minimum measurement count in each road segment.

measurements. We had only ten sensors-participants in our sensor network, with five participants providing 81.8% of the data.

- Relevant to the previous point, due to a lack of enough data in the vicinity of the Stener Heyerdahl, we could not appropriately evaluate the performance of the remote calibration scheme for some of the Snifferbike sensors.
- The Snifferbike measurements are at 10s temporal resolution, while the official data from the reference instruments are at hourly resolution.
- Due to a lack of funding resources, the Snifferbike sensor tool kits were not inspected during the analysis period. A good practice will be that sensors are regularly inspected and co-located in reference monitoring stations to provide more robust calibration results.
- The reproducibility of co-dependency of the Sensirion SPS30 bias on RH and temperature needs further investigation under different environmental conditions. Additionally, analysis of such behavior from other PM LCSs is worthy of research.
- In the method that we proposed for the estimation of minimum required measurements per road segment, all measurements might relate to a short period. For example, assume that there are 50 measurements available for a road segment, and all of them are recorded during an hour. In that case, the provided mean of the

measurements cannot be representative of the real PM level in that road segment. One approach to address this issue is to calculate the SD of the measurements' date of sampling in each road segment; here, the SD for only three road segments was less than one week (Fig. 10). The mean of SD for measurements' date of sampling in each road segment was 63 days and 30 min.

4. Conclusions

Citizen science initiatives can foster collaboration and knowledge sharing between researchers, policymakers, and the general public. Using mobile LCSs, citizen science initiatives can provide valuable data on air pollution in urban areas. This data can be used by individuals and communities to save official monitoring costs, increase public awareness, devise better policy decisions, and eventually reduce air pollution. Calibration and estimation of measurement uncertainty, however, are crucial to ensuring the accuracy and reliability of data collected by these sensors.

We aimed at quantifying the uncertainty and evaluation of a specific LCS brand (Sensirion SPS30 integrated into Snifferbike sensor kit) in a mobile mode which is less considered in previous studies. We conducted co-location and sensor-to-sensor inter-comparison tests on Snifferbike

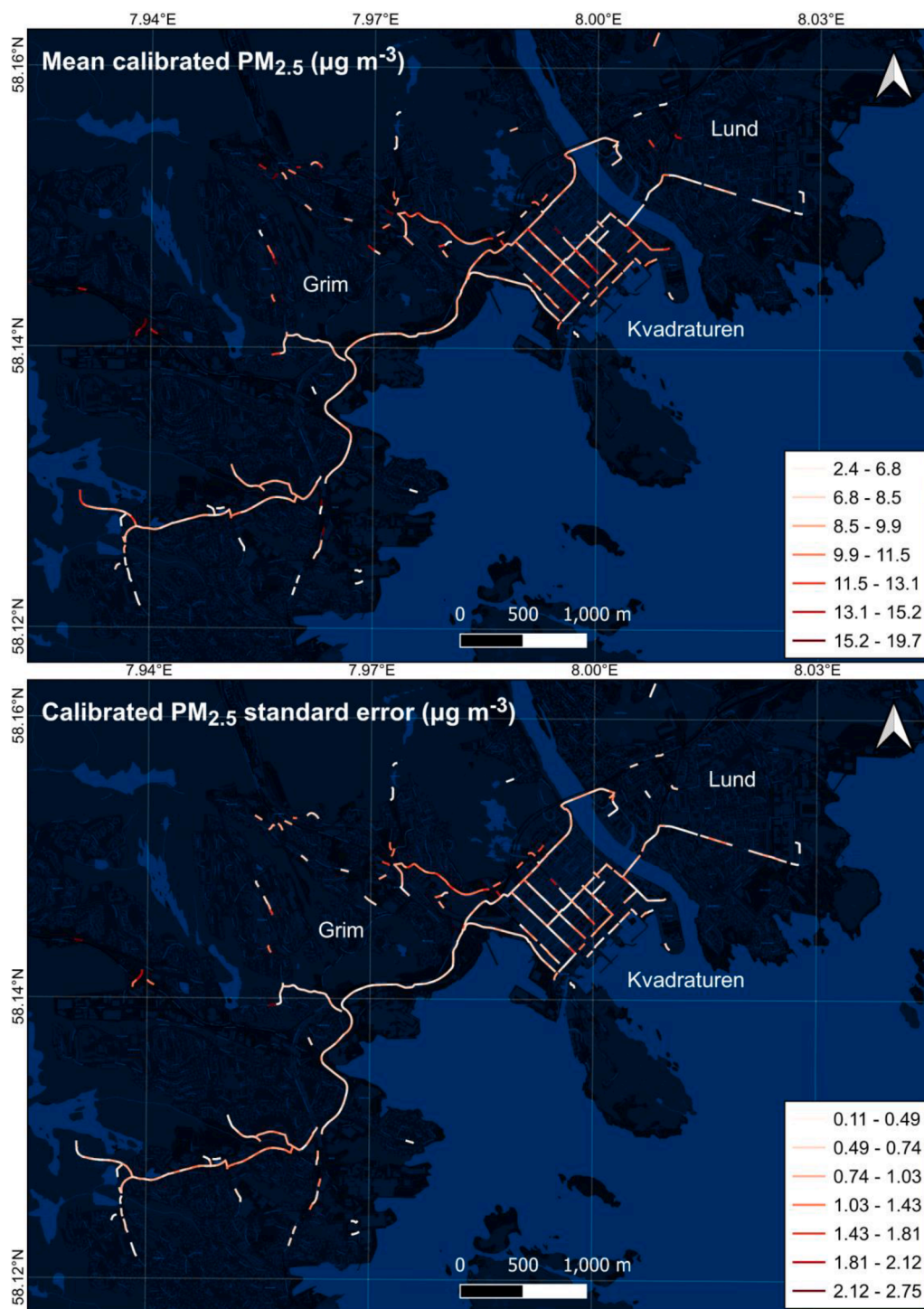


Fig. 9. Spatial distribution of calibrated PM_{2.5} measured by the 10 Snifferbike sensors mounted on bikes between Oct 2021 and Aug 2022, Kristiansand, Norway. Values are the means of at least 27 measurements across each road segment (50 m length).

PM_{2.5} sensors. We developed a network of sensors mounted on citizens' bikes to map the PM_{2.5} variability along the roads in Kristiansand Norway.

In static mode, the co-location of the three Snifferbike sensors at the reference monitoring station showed a RMSD of 7.55 µg m⁻³ and a *r* of 0.75, compared to reference measurements. In static mode and different RH bins, we observed a lower deviation from the reference measurements at moderate air temperatures (5 - 10°C) compared to low or high air temperature ranges (-5 - 0°C and 20 - 25°C), respectively. At the same time, it is generally accepted that deviation increases with RH increase.

Our sensor-to-sensor inter-comparison of the Snifferbike sensors in static mode revealed a high correlation (*r* ranging from 0.94 to 0.99) between the output of the sensors. In contrast, on-bike/coincident bike inter-comparison tests showed an increase in SD of factory-calibrated measured PM_{2.5} between 0.03 and 0.04 µg m⁻³ per one km h⁻¹ increase in the bike speed. The factory-calibrated SD of PM_{2.5} measurements also increased by 0.04 µg m⁻³ with a 1 µg m⁻³ increase in ambient PM_{2.5} levels. We conclude that uncertainty in Snifferbike PM_{2.5} measurements increases by the device speed; at low to moderately low ambient PM_{2.5} concentrations (0 - 30 µg m⁻³) and low to moderate

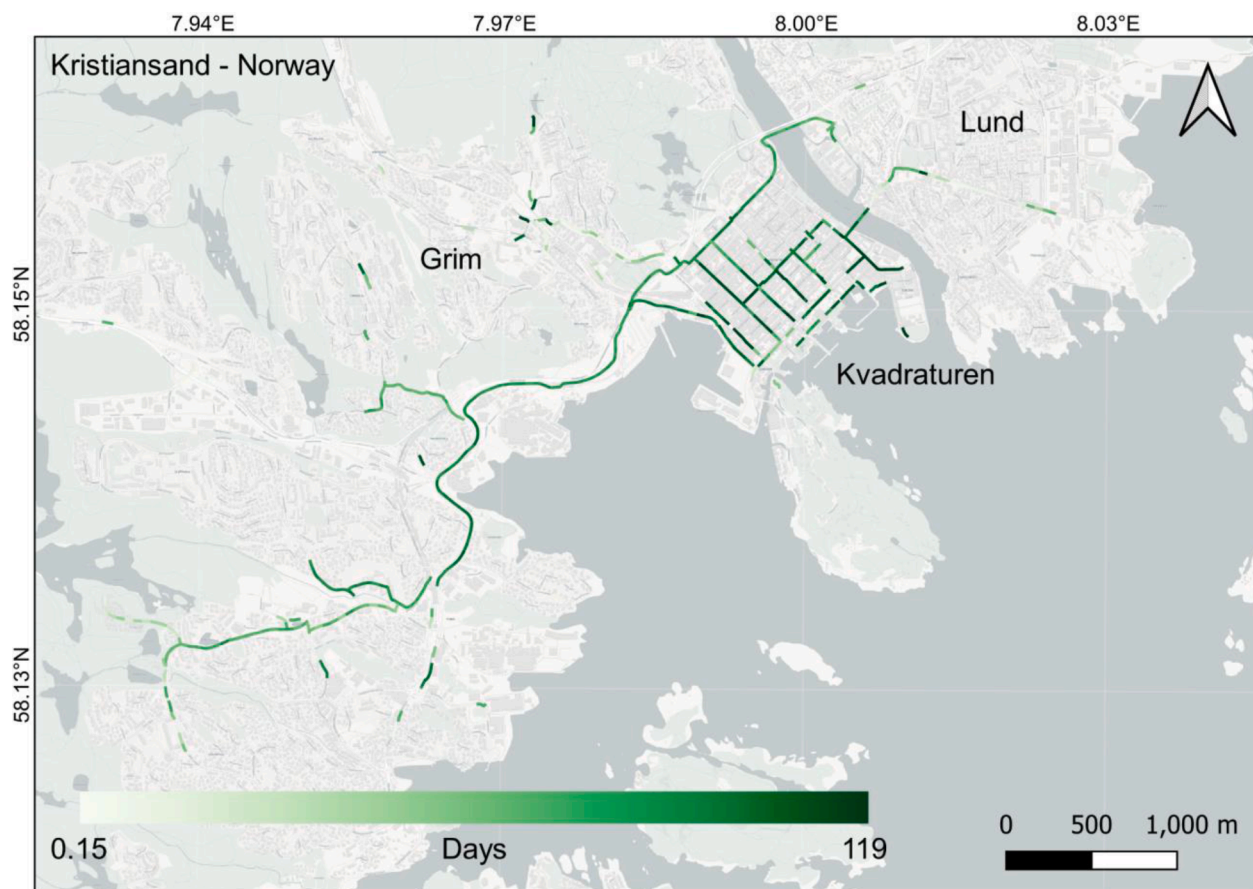


Fig. 10. Standard Deviation (SD) of the times at which $PM_{2.5}$ samples were measured by the ten citizen-operated Snifferbike sensors in Kristiansand, Norway. SDs are calculated for road segments of 50 m in length.

sensor speeds ($0 - 15 \text{ km h}^{-1}$), this uncertainty (or SD of $PM_{2.5}$ measurements) can be assumed to be negligible.

The Snifferbike network was successfully used in mapping the spatial distribution of $PM_{2.5}$ pollution along the roads in Kristiansand. Overall, the average (calibrated) $PM_{2.5}$ measured by the Snifferbike sensors per road segment was $9.36 \mu\text{g m}^{-3}$ in the analysis period (October 2021 - Aug 2022). We did not find any correlation between the high concentrations of $PM_{2.5}$ and traffic load in the city center of Kristiansand. The study proposes a method for estimating the minimum number of $PM_{2.5}$ measurements required per road segment to ensure data representativeness, which can be useful for designing air quality monitoring networks. Assuming the data are scattered adequately in the time, we approximated that at least 27 measurements per road segment (50 m here) are required to make robust estimations of $PM_{2.5}$ concentration along the roads. The calculated minimum number of points depends on the road segment length and needs to be higher at longer road segments.

By comparing the calibrated measurements against independent official measurements, we conclude that for the case of Snifferbike sensors, in particular mobile mode, in-field pre-calibration (of limited sensors) and/or remote calibration of the sensors do not necessarily improve the accuracy of the measurements. In case of data availability of both mobile sensors and official stations, we recommend evaluating/calibrating individual sensors when they pass or are located in the vicinity of the reference monitoring stations. We conclude that on-bike Snifferbike $PM_{2.5}$ measurements are useful in combination with static sensors for non-regulatory air quality monitoring purposes and personal exposure to PM evaluation, despite the spatio-temporal sampling bias in data as the limitation in our analysis (as volunteer citizens operated the sensors). The results of this study provide a practical, replicable, and scalable guide for designing comparable citizen-operated low-cost

mobile sensor networks in other urban regions.

Data availability

Sensor data required for replication of the results provided in this paper are available at the NILU's sensor data platform, accessible at <https://nordicpathlive.nilu.no/>. Bulk data download requires a data access token that can be obtained from the first/corresponding Authors. Reference monitoring station data is available at <https://luftkvalitet.nilu.no/en/historical>.

CRediT authorship contribution statement

Amirhossein Hassani: Conceptualization, Methodology, Funding acquisition, Formal analysis, Software, Investigation, Visualization, Validation, Data curation, Writing – original draft, Writing – review & editing. **Núria Castell:** Conceptualization, Methodology, Funding acquisition, Supervision, Investigation, Validation, Project administration, Resources, Writing – review & editing. **Ågot K. Watne:** Funding acquisition, Writing – review & editing. **Philipp Schneider:** Methodology, Investigation, Validation, Conceptualization, Visualization, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The link to data is provided.

Acknowledgments

This work was (partly) supported by NordForsk through funding to NordicPATH: Nordic Participatory, Healthy, and People-Centered Cities, Project no 95326. Special thanks to Solvor B. Stølevik from Kristiansand Kommune (municipality) for her work on the co-location of the LCSs, cycling with the 11 LCSs, and distributing the LCSs to participants. Thanks also to Håkan Fridén from IVL Swedish Environmental Research Institute for the valuable inputs during the analysis of the data. Also, our acknowledgment to the 10 citizen scientists that participated in the study, collecting air quality data with their bikes.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.scs.2023.104607.

References

- Alexandre, M., & Gerboles, M. (2012). Review of small commercial sensors for indicative monitoring of ambient gas. *Chemical Engineering Transactions*, 30, 169–174.
- Alfano, B., Barretta, L., Del Giudice, A., De Vito, S., Di Francia, G., Esposito, E., et al. (2020). A review of low-cost particulate matter sensors from the developers' perspectives. *Sensors*, 20(23), 6819.
- American Meteorological Society. (2020). Fumigation. *Glossary of Meteorology*. <http://glossary.ametsoc.org/wiki/fumigation>.
- Brattich, E., Bracci, A., Zappi, A., Morozzi, P., Di Sabatino, S., Porcù, F., et al. (2020). How to get the best from low-cost particulate matter sensors: Guidelines and practical recommendations. *Sensors*, 20(11), 3073.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Castell, N., Dauge, F.R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., et al. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?. 99, 293–302.
- Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the Paper presented at the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*.
- Cheriyian, D., & Choi, J.-h. (2020). Estimation of particulate matter exposure to construction workers using low-cost dust sensors. *Sustainable Cities and Society*, 59, Article 102197.
- Chowdhuri, I., Pal, S. C., Arabameri, A., et al. (2022). Have any effect of COVID-19 lockdown on environmental sustainability? A study from most polluted metropolitan area of India. *Stochastic Environmental Research and Risk Assessment*, 36, 283–295.
- DeSouza, P., Anjomshoa, A., Duarte, F., Kahn, R., Kumar, P., & Ratti, C. (2020). Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustainable Cities and Society*, 60, Article 102239.
- Ekman, K., & Weilenmann, A. (2021). Behind the scenes of planning for public participation: Planning for air-quality monitoring with low-cost sensors. *Journal of Environmental Planning and Management*, 64(5), 865–882.
- Feenstra, B., Papapostolou, V., Hasheminassab, S., Zhang, H., Der Boghossian, B., Cocker, D., & Polidori, A. (2019). Performance evaluation of twelve low-cost PM_{2.5} sensors at an ambient air monitoring site. *Atmospheric Environment*, 216, Article 116946.
- Geels, C. (2015). *Future Air Quality and Related Health Effects in a Nordic Perspective: The Possible Impacts of Future Changes in Climate, Anthropogenic Emissions, Demography and Building Structure*. Nordic Council of Ministers.
- Giordano, M. R., Malings, C., Pandis, S. N., Presto, A. A., McNeill, V. F., Westervelt, D. M., et al. (2021). 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, Article 105833.
- Hemamalini, R. R., Vinodhini, R., Shanthini, B., Partheeban, P., Charumathy, M., & Cornelius, K. (2022). Air quality monitoring and forecasting using smart drones and recurrent neural network for sustainable development in Chennai city. *Sustainable Cities and Society*, 85, Article 104077.
- Hong, G.-H., Le, T.-C., Tu, J.-W., Wang, C., Chang, S.-C., Yu, J.-Y., et al. (2021). Long-term evaluation and calibration of three types of low-cost PM_{2.5} sensors at different air quality monitoring stations. *Journal of Aerosol Science*, 157, Article 105829.
- Im, U., Christensen, J. H., Nielsen, O.-K., Sand, M., Makkonen, R., Geels, C., et al. (2019). Contributions of Nordic anthropogenic emissions on air pollution and premature mortality over the Nordic region and the Arctic. *Atmospheric Chemistry and Physics*, 19(20), 12975–12992.
- Jayarate, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). 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(8), 4883–4890.
- Kang, Y., Aye, L., Ngo, T. D., & Zhou, J. (2021). Performance evaluation of low-cost air quality sensors: A review. *Science of The Total Environment*, Article 151769.
- Kompalli, S. K., Moorthy, K. K., & Babu, S. S. (2014). Rapid response of atmospheric BC to anthropogenic sources: Observational evidence. *Atmospheric Science Letters*, 15(3), 166–171.
- Kuula, J., Mäkelä, T., Aurela, M., Teinilä, K., Varjonen, S., González, Ó., & Timonen, H. (2020). Laboratory evaluation of particle-size selectivity of optical low-cost particulate matter sensors. *Atmospheric Measurement Techniques*, 13(5), 2413–2423.
- Lederbogen, F., Kirsch, P., Haddad, L., Streit, F., Tost, H., Schuch, P., et al. (2011). City living and urban upbringing affect neural social stress processing in humans. *Nature*, 474(7352), 498–501.
- Lee, C. C., Tran, M.-V., Choo, C. W., Tan, C. P., & Chiew, Y. S. (2020). Evaluation of air quality in Sunway City, Selangor, Malaysia from a mobile monitoring campaign using air pollution micro-sensors. *Environmental Pollution*, 265, Article 115058.
- Lewis, A., Peltier, W.R., & von Schneidmesser, E. (2018). Low-cost sensors for the measurement of atmospheric composition: Overview of topic and future applications.
- Lim, C. C., Kim, H., Villacassim, M. J. R., Thurston, G. D., Gordon, T., Chen, L. C., et al. (2019). Mapping urban air quality using mobile sampling with low-cost sensors and machine learning in Seoul, South Korea. *Environment International*, 131, Article 105022.
- Lin, Y., Dong, W., & Chen, Y. (2018). Calibrating low-cost sensors by a two-phase learning approach for urban air quality measurement. In *2. Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* (pp. 1–18).
- Liu, D., Zhang, Q., Jiang, J., & Chen, D.-R. (2017). Performance calibration of low-cost and portable particulate matter (PM) sensors. *Journal of Aerosol Science*, 112, 1–10.
- Lu, M., Dai, R., de Boer, C., Schmitz, O., Kooter, I., Cristescu, S., & Karssenberg, D. (2021). External validation for statistical NO₂ modelling: A study case using a high-end mobile sensing instrument. *Atmospheric Pollution Research*, 12(11), Article 101205.
- Mahajan, S., & Kumar, P. (2020). Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustainable Cities and Society*, 57, Article 102076.
- Mahajan, S., Kumar, P., Pinto, J. A., Riccetti, A., Schaaf, K., Camprodon, G., et al. (2020). A citizen science approach for enhancing public understanding of air pollution. *Sustainable Cities and Society*, 52, Article 101800.
- Mihăiță, A. S., Dupont, L., Chery, O., Camargo, M., & Cai, C. (2019). Evaluating air quality by combining stationary, smart mobile pollution monitoring and data-driven modelling. *Journal of Cleaner Production*, 221, 398–418.
- Miskell, G., Salmond, J. A., & Williams, D. E. (2018). Use of a handheld low-cost sensor to explore the effect of urban design features on local-scale spatial and temporal air quality variability. *Science of the Total Environment*, 619, 480–490.
- Morawska, L., Thai, P. K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., et al. (2018). Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environment International*, 116, 286–299.
- Ortiz, A., Guerreiro, C., & Soares, J. (2020). Air Quality in Europe—2020 Report. *European environment agency (EEA) and European topic centre on air pollution, noise, transport and industrial pollution (ETC/ATNI)*, 9, 1–162.
- Pal, S. C., Chowdhuri, I., Saha, A., Chakraborty, R., Roy, P., Ghosh, M., & Shit, M. (2021). Improvement in ambient-air-quality reduced temperature during the COVID-19 lockdown period in India. *Environment, Development and Sustainability*, 23, 9581–9608.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, 12, 2825–2830.
- Qin, X., Hou, L., Gao, J., & Si, S. (2020). The evaluation and optimization of calibration methods for low-cost particulate matter sensors: Inter-comparison between fixed and mobile methods. *Science of The Total Environment*, 715, Article 136791.
- Ritchie, H., & Roser, M. (2018). *Urbanization*. Our world in data.
- Roberts, F. A., Van Valkinburgh, K., Green, A., Post, C. J., Mikhailova, E. A., Commodore, S., et al. (2022). Evaluation of a new low-cost particle sensor as an internet-of-things device for outdoor air quality monitoring. *Journal of the Air & Waste Management Association*, 72(11), 1219–1230.
- Samad, A., & Vogt, U. (2020). Investigation of urban air quality by performing mobile measurements using a bicycle (MOBAIR). *Urban Climate*, 33, Article 100650.
- Santana, P., Almeida, A., Mariano, P., Correia, C., Martins, V., & Almeida, S. M. (2021). Air quality mapping and visualisation: An affordable solution based on a vehicle-mounted sensor network. *Journal of Cleaner Production*, 315, Article 128194.
- Satterthwaite, D. (1993). The impact on health urban environments. *Environment and Urbanization*, 5(2), 87–111.
- Schneider, P., Castell, N., Vogt, M., Dauge, F. R., Lahoz, W. A., & Bartonova, A. (2017). Mapping urban air quality in near real-time using observations from low-cost sensors and model information. *Environment International*, 106, 234–247.
- Shi, X., Zhao, C., Jiang, J. H., Wang, C., Yang, X., & Yung, Y. L. (2018). Spatial representativeness of PM_{2.5} concentrations obtained using observations from network stations. *Journal of Geophysical Research: Atmospheres*, 123(6), 3145–3158.
- Sm, S. N., Yasa, P. R., Narayana, M. V., Khadirmaikar, S., & Rani, P. (2019). Mobile monitoring of air pollution using low cost sensors to visualize spatio-temporal variation of pollutants at urban hotspots. *Sustainable Cities and Society*, 44, 520–535.
- Tiwari, S., Srivastava, A. K., Bisht, D. S., Parmita, P., Srivastava, M. K., & Attri, S. D. (2013). Diurnal and seasonal variations of black carbon and PM_{2.5} over New Delhi, India: Influence of meteorology. *Atmospheric Research*, 125, 50–62.
- Tsyro, S., Schulz, M., Mortier, A., Valdebenito, A., & Benedictow, A. (2022). High PM₁₀ levels: Episode of 20-27 March 2022 CAMS2.71 episode analysis report N°02 in 2022 Reference: CAMS71_D3.2.1-2022-2-B.20-27MarchEpisode. *Norwegian Meteorological Institute, Copernicus Atmosphere Monitoring Service*, 1-23.

- Van den Bossche, J., Theunis, J., Elen, B., Peters, J., Botteldooren, D., & De Baets, B. (2016). Opportunistic mobile air pollution monitoring: A case study with city wardens in Antwerp. *Atmospheric Environment*, *141*, 408–421.
- Venkatraman Jagatha, J., Klausnitzer, A., Chacón-Mateos, M., Laquai, B., Nieuwkoop, E., van der Mark, P., et al. (2021). Calibration method for particulate matter low-cost sensors used in ambient air quality monitoring and research. *Sensors*, *21*(12), 3960.
- Vogt, M., Schneider, P., Castell, N., & Hamer, P. (2021). Assessment of low-cost particulate matter sensor systems against optical and gravimetric methods in a field co-location in Norway. *Atmosphere*, *12*(8), 961.
- Wang, A., Machida, Y., deSouza, P., Mora, S., Duhl, T., Hudda, N., et al. (2023). Leveraging machine learning algorithms to advance low-cost air sensor calibration in stationary and mobile settings. *Atmospheric Environment*, Article 119692.
- Wesseling, J., Hendricx, W., de Ruiter, H., van Ratingen, S., Drukker, D., Huitema, M., et al. (2021). Assessment of PM_{2.5} exposure during cycle trips in the Netherlands using low-cost sensors. *18*(11), 6007.
- Wolf, T., Pettersson, L. H., & Esau, I. (2021). Dispersion of particulate matter (PM_{2.5}) from wood combustion for residential heating: Optimization of mitigation actions based on large-eddy simulations. *Atmospheric Chemistry and Physics*, *21*(16), 12463–12477.
- World Health Organization. (2016). *Global report on urban health: Equitable healthier cities for sustainable development*. World Health Organization.