npj | climate and atmospheric science

Article

Published in partnership with CECCR at King Abdulaziz University

https://doi.org/10.1038/s41612-024-00859-z

A novel spatiotemporal prediction approach to fill air pollution data gaps using mobile sensors, machine learning and citizen science techniques

Arunik Baruah[1,2], Dimitrios Bousiotis[3], Seny Damayanti[3], Alessandro Bigi[1], Grazia Ghermandi[1], O. Ghaffarpasand[3], Roy M. Harrison[3,4] & Francis D. Pope[3]

Particulate Matter (PM) air pollution poses significant threats to public health. We introduce a novel machine learning methodology to predict $PM_{2.5}$ levels at 30 m long segments along the roads and at a temporal scale of 10 seconds. A hybrid dataset was curated from an intensive PM campaign in Selly Oak, Birmingham, UK, utilizing citizen scientists and low-cost instruments strategically placed in static and mobile settings. Spatially resolved proxy variables, meteorological parameters, and PM properties were integrated, enabling a fine-grained analysis of $PM_{2.5}$. Calibration involved three approaches: Standard Random Forest Regression, Sensor Transferability and Road Transferability Evaluations. This methodology significantly increased spatial resolution beyond what is possible with regulatory monitoring, thereby improving exposure assessments. The findings underscore the importance of machine learning approaches and citizen science in advancing our understanding of PM pollution, with a small number of participants significantly enhancing local air quality assessment for thousands of residents.

Particulate Matter (PM) air pollution has a considerable negative influence on the human health, especially with respect to the cardiovascular and pulmonary systems. According to the European Environmental Agency (EEA, 2023)[1], 97% of the urban population in Europe is exposed to fine particulate matter ($PM_{2.5}$, i.e. PM with a diameter of 2.5 μm or smaller) concentrations above the World Health Organization's (WHO) 2021 recommendations of 5 μg m^{-3} for annual average. Within the EU, air pollution is estimated to lead to 238,000 premature deaths in 2021 and is the largest environmental health risk in Europe[1]. $PM_{2.5}$ is a critical air pollutant with primary $PM_{2.5}$ originating mainly from combustion processes and secondary $PM_{2.5}$ from the reaction of organic or inorganic gas compounds, finally contributing eventually up to more than 50% of $PM_{2.5}$ mass depending on the season and the location[2]. Also PM_{10} (PM with a diameter of 10 μm or smaller) is a critical air pollutant, with coarse particles, i.e. between 2.5 μm and 10 μm, resulting mainly from mechanical processes[3].

$PM_{2.5}$ has become the leading environmental contributor to the global burden of disease, representing a substantial departure from its position as the fifth major contributor among environmental risk factors in 1990[4]. Studies have shown that spending a substantial amount of time in areas even with low ambient $PM_{2.5}$ levels can have adverse effects on human health[5,6]. The health impact of air pollution is critical in urban areas, where most of the world's population resides, therefore rapid reduction strategies are required. For these strategies to be successful, they need to be targeted and hence an accurate description of the spatial-temporal variability of PM is required[7]. Urban areas exhibit high heterogeneity in PM concentrations due to the diversity of the emission sources, the variability in land use patterns, and of the interaction between the meteorological factors and the urban canopy, which influence air pollutants' dispersion[8]. This spatial and temporal variability poses challenges for exposure assessment and air quality management[9].

Regulatory monitoring networks, such as the UK's Automatic Urban and Rural Network (AURN), serve as the main UK infrastructure for ensuring compliance with ambient air pollution standards. Nevertheless, the acquisition and maintenance costs of regulatory-grade instruments are high, and the sparsely distributed station network fails to capture the small-scale spatial variations observed in pollutant concentrations in urban areas, as

[1]Dept. of Engineering 'Enzo Ferrari', University of Modena and Reggio Emilia, Modena, Italy. [2]University School for Advanced Studies IUSS Pavia, Palazzo Del Broletto, Pavia, Italy. [3]School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK. [4]Department of Environmental Sciences, Faculty of Meteorology, Environment and Arid Land Agriculture, King Abdulaziz University, Jeddah, Saudi Arabia. e-mail: f.pope@bham.ac.uk

indicated by numerous studies[10,11]. These localized variations contribute to differences in human pollutant exposures, ultimately influencing associated health impacts[12].

To detect and quantify the fine-scale spatial fluctuations in pollutant concentrations, there is a growing interest for utilizing low-cost sensor (LCS) networks. This interest is attributed to the improved capabilities of sensor technologies and the development of innovative methods for sensor calibration[13–15]. However, challenges remain in optimizing sensor placement strategies[16,17], in data quality assurance due to e.g. LCS drift or sensitivity to meteorological variables[18,19] and in interpreting LCS data in the context of regulatory air quality standards[20,21]. To accurately estimate population exposure, monitoring at a high spatial and temporal resolution should be pursued. Mobile low-cost sensors provide a cost-effective solution for monitoring air quality in areas with limited existing infrastructure, owing to their compact size and portability. Examples include $PM_{2.5}$ measurements performed by citizen-operated mobile sensors mounted on bikes[22], deployed on routine fleets of vehicles such as trash trucks[23], tram-based mobile sensor network in Zurich[24], taxi motorcycles in Kampala[25], etc. However, sampling every location continuously throughout a given geographic area is an unattainable goal.

A diverse array of models are utilized in the prediction of PM levels. Some are based on atmospheric processes and emissions, e.g. Chemical Transport Models (CTMs) or Lagrangian particle dispersion models. These models play a crucial role in simulating and understanding the complex dynamics of air pollutants, incorporating factors like atmospheric chemistry, emission sources, and dispersion patterns. For instance, Sokhi et al. [26] evaluated four regional chemistry transport models, with a horizontal resolution of ~20 km which systematically underestimated PM_{10} and $PM_{2.5}$ concentrations in Europe by 10–60%, varying with models and seasons, when compared with the European Monitoring and Evaluation Programme (EMEP) measurements. Zhang et al. [27] employed a simplified Lagrangian particle dispersion model (LPDM) with Bayesian-RAT (multiplicative ratio correction optimization) to enhance regional PM concentration predictions, demonstrating superior accuracy compared to other models (WRF-Chem and CAMX), showcasing the LPDM's advantage in forecasting PM and potentially other pollutants. However, both CTMs and LPDMs may encounter challenges in accurately predicting $PM_{2.5}$ concentrations due small size of the dataset, low predictive performance for small areas, high computational cost and achieving sufficient spatial and temporal resolution[28]. Other prediction approaches include the use of statistical approaches based on meteorological variables and emission proxies[29].

Data-driven models, in contrast to physically-driven models, have garnered significant attention due to their ease of implementation[30]. Machine learning (ML) models have been shown to be highly effective for PM prediction, showcasing robust performance with non-linear variables and flexible modelling[31]. Supervised learning involves the integration of tree-based algorithms (random forest, extreme gradient boosting, light gradient boosting, etc.) and vector-based algorithms (k-nearest neighbour, support vector regression, etc.), capable of learning label data through classifiers or regressors[32]. Nevertheless, classifier methods proved to be less suitable for PM prediction compared to regressor methods, and, in general, vector-based algorithms exhibited lower predictive power than their tree-based counterparts[33]. Hence, tree-based machine learning algorithms, known for their low computational costs and high prediction accuracy, are extensively employed in PM prediction research[34,35].

The existing literature reveals a notable research gap concerning the limitations of current air pollution prediction models, particularly in the context of fine-scale spatial and temporal variations. To address these gaps that demand innovative and cost-effective solutions for enhanced spatial resolution, especially in densely populated urban settings, our study proposes a novel methodology that leverages on ML tools, particularly tree-based models, to predict $PM_{2.5}$ levels with unprecedented precision at both spatial and temporal scales.

While our research maintains a broad scope, we conducted a comprehensive testing phase within a measurement campaign from Selly Oak, Birmingham, United Kingdom where we deployed a combination of static and mobile Optical Particle Counters. Our primary objective is to craft predictive models using tree-based ML algorithms that excel in estimating PM levels. To achieve this, we are harnessing the potential of a hybrid dataset, curated to integrate information from both static, mobile low-cost sensors and diverse ancillary datasets. Our focus extends beyond scenarios with active mobile sensor deployment, aiming to create models that can reliably forecast PM concentrations even when the mobile sensors are not in operation.

Results

The following section presents the outcomes of our study, which focused on predicting $PM_{2.5}$ levels using the three distinct ML calibration approaches. We evaluated the ML approaches beginning with the baseline model performance and then examining transferability across sensors and locations. The agreement between predicted and observed data points was assessed in terms of frequency distribution density plots for each calibration approach, to highlight areas of convergence or divergence between predicted and actual $PM_{2.5}$ levels. The values of the performance metrics are given in Table 1.

Calibration approach 1: standard random forest regression

The Standard Random Forest Regression, utilizing an 80-20 train-test split, demonstrated strong predictive performance at 10 s temporal resolution. The R^2 score was 0.85 and MAE of 1.60 μg m^{-3}. The corresponding RMSE of 2.40 μg m^{-3} further complements these results. As seen in Fig. 1a, it showcases a peak in density around the central values, indicating a strong alignment between predicted and actual $PM_{2.5}$ levels. Upon removing outliers ($PM_{2.5}$ > 40 μg/m^3), the slope moves slightly closer to 1, suggesting a marginally stronger correlation at lower concentration levels. Notably, the intercept is close to the stated sensitivity threshold of the OPC-N3 sensors (<1 μg/m^3), indicating that the model's predictions align with the lower detection limit of the instruments. This calibration approach serves as the reference point, representing the best achievable performance within the entire setup. It serves as our benchmark and provides a valuable standard against which we measure the effectiveness of other calibration approaches.

Calibration approach 2: sensor transferability evaluation

In the Sensor Transferability Evaluation, our focus shifted to the model's ability to generalize across different sensors. Specifically, the Random Forest model was initially calibrated on a specific mobile OPC-N3 unit and subsequently evaluated on an independent mobile OPC-N3 unit. The results present detailed insights into the model's performance across a range of sensor configurations. When employing a single-sensor test-and-train configuration, the model demonstrated good generalization capabilities. The R^2 score of 0.65 signifies a moderate yet substantial level of explained variance, while the MAE of 2.76 μg m^{-3} and an overestimated mean bias error of just 0.43 μg m^{-3} reflects the model's accuracy in predicting $PM_{2.5}$ levels across different sensors. Additionally, the KDP (Fig. 2a) indicates a notable pattern, with predicted $PM_{2.5}$ levels exhibiting a density peak at 0.14, while the actual $PM_{2.5}$ levels have a density peak at 0.7 suggesting systematic tendency of the model to slightly underestimate $PM_{2.5}$ levels in this scenario. However, it is crucial to note that not all sensor calibration/validation pairs

Table 1 | Performance metrics under various calibration configurations

Calibration approach	R^2	MAE (μg m^{-3})	MBE (μg m^{-3})	RMSE (μg m^{-3})	ρ
Standard random forest	0.85	1.60	−0.01	2.40	0.92
Sensor transferability evaluation	0.65	2.76	0.43	4.11	0.83
Road transferability evaluation	0.71	2.46	−1.14	3.22	0.81

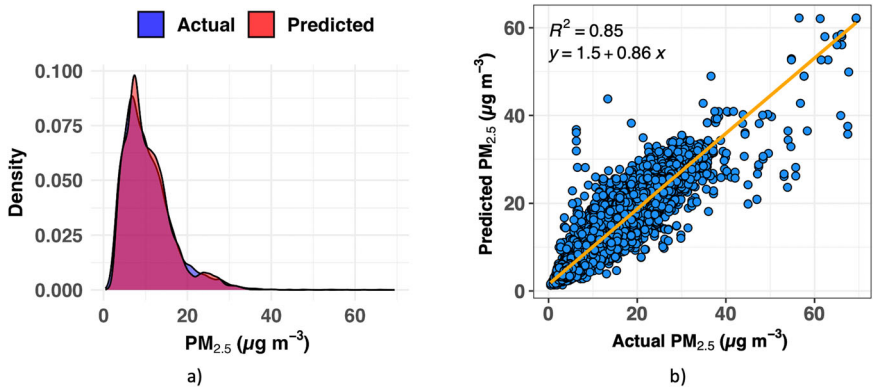

Fig. 1 | Results from the calibration approach 1: Standard Random Forest Regression. a Kernel density plot (KDP) with *x*-axis representing $PM_{2.5}$ levels and the *y*-axis representing the density of data points at each level and **b** linear regression between actual and predicted $PM_{2.5}$ concentrations for the Standard Random Forest Regression.

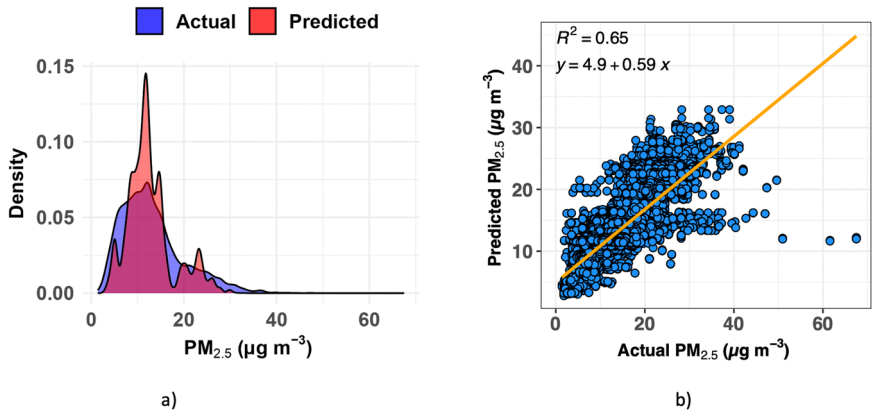

Fig. 2 | Results from the calibration approach 2: Sensor Transferability Evaluation. a Kernel density plot (KDP) with *x*-axis representing $PM_{2.5}$ levels and the *y*-axis represents the density of data points at each level and **b** linear regression between actual and predicted $PM_{2.5}$ concentrations for the best Sensor Transferability Evaluation.

perform equally under this approach. The variability in sensor pair performance becomes apparent, revealing the challenge for model's adaptability to specific OPC units. If the performance reported above is the best one among all pairs, swapping this same calibration/validation pair results in a significant decrease in model performance (MBE = −2.64 μg m^{-3}, MAE = 4.45 μg m^{-3} and R^2 = 0.25). When combining data from multiple sensors, a different perspective emerges (MBE 2.47 μg m^{-3}, MAE = 4.05 μg m^{-3}, and R^2 = 0.21). The combined sensor configuration yields suboptimal results compared to the single-sensor approach. These findings underscore the importance of considering sensor-specific behaviours and configurations. The model's adaptability across different sensors is evident, but the complexities of individual sensor performance, necessitate careful consideration in the interpretation of results. Factors such as differing time-of-mapping among sensors, variations in traffic patterns or $PM_{2.5}$ concentrations at specific locations, contribute to the observed variability in model performance. It should be noted that the results presented throughout the paper are based on a 10 s temporal resolution. However, it is interesting that after averaging the dataset into 5–15 min data intervals, the R^2 score showed improvement up to 0.4–0.5. This suggests a potential temporal sensitivity, emphasizing the consideration of the trade-off between temporal resolution and model performance, given the corresponding reduction in data points.

Calibration approach 3: road transferability evaluation

For the Road Transferability Evaluation, the model was calibrated on one road and evaluated on a different one. The results indicated a satisfactory ability to generalize across different road stretches, with an R^2 value of 0.71 and an MAE of 2.46 μg m^{-3} (Fig. 3 trained on Bristol Road and tested on Raddlebarn Road). Notably, the model's consistent performance across different roads within the area suggests a capacity to adapt to varying environmental conditions and road characteristics. This observed generalization may be attributed to the similarity in atmospheric and pollution background conditions across the roads, showcasing the model's adaptability to diverse but locally consistent factors. For instance, for scenarios with the road ranking explained in the "Hybrid Dataset Preparation" part of the later Methodology section, we calibrated the model on a secondary road (Raddlebarn Road) and tested it on a tertiary (Bournbrook Road), the results yielded comparable success with R^2 = 0.66, MBE = 0.21 μg m^{-3} and MAE = 2.20 μg m^{-3} or training and testing on tertiary and primary, etc. This versatility is particularly valuable in urban environments where road types can change rapidly, and atmospheric conditions may exhibit subtle variations.

Variable importance (VIMP) of the input features

Figure 4 shows the relative VIMP from the standard RF regression model for $PM_{2.5}$ predictions. It highlights the high importance assigned to the 'Mobile Sensors' variable, which is a categorical variable representing sensor ID. The model attributes considerable weight to the sensor ID, indicating that specific sensors consistently capture important patterns or localized pollution sources. However, it is important to acknowledge the potential introduction of sampling biases due to the choice of routes taken by individuals carrying the sensors. Routes through busy intersections or areas with construction activities could result in higher variability in the measured $PM_{2.5}$ concentrations, thus influencing the perceived importance of certain sensors. Figure 4 also underscores the significance of static sensors such as 'Static Sensor 2', 'Static Sensor 1', and so forth. Unlike mobile sensors, static sensors offer continuous and consistent measurements at specific locations. The importance of these static sensors may be attributed to their strategic placement in areas that serve as key indicators of the overall air quality. 'Static Sensor 2', for instance, was in a place with high vehicular traffic, while 'Static Sensor 1' was positioned near an urban background location. Despite its proximity to the primary road, the Static sensor 4 exhibited lower

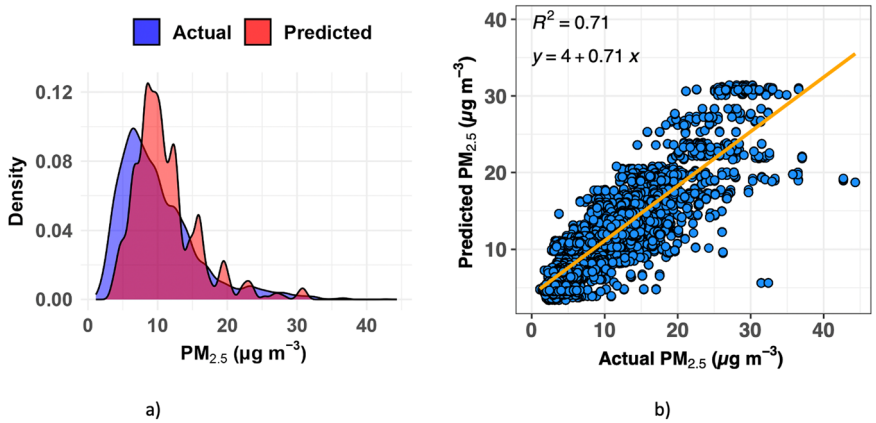

Fig. 3 | Results from the calibration approach 3: Road Transferability Evaluation. a Kernel density plot (KDP) with *x*-axis representing $PM_{2.5}$ levels and the *y*-axis represents the density of data points at each level and **b** linear regression between actual and predicted $PM_{2.5}$ concentrations for the best Road Transferability Evaluation.

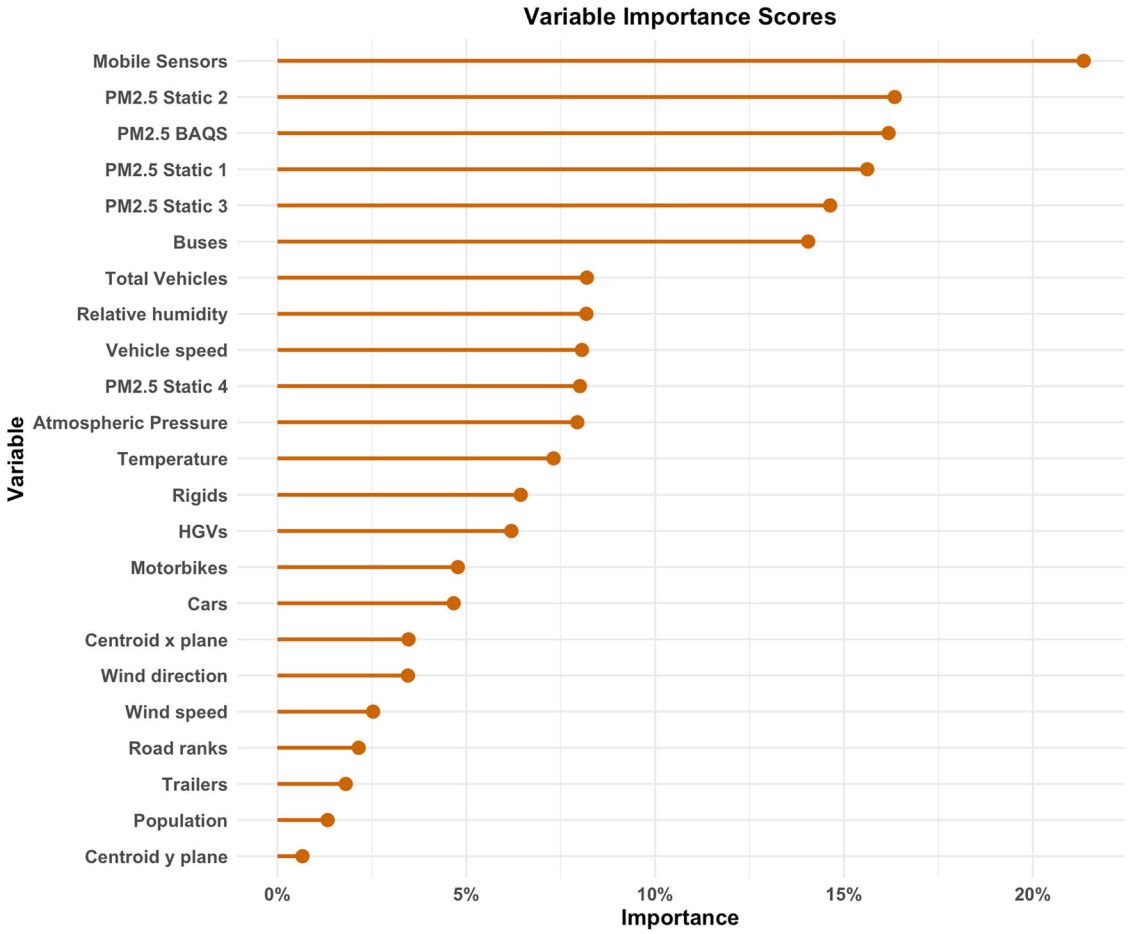

Fig. 4 | Relative importance of different input parameters on the Random Forest regression model. The points on the plot provide the relative importance (*x*-axis) of each of the named input variables (*y*-axis) within the Standard Random Forest regression model.

importance in the model, potentially due to its position in a densely populated area with local emission sources like nearby vehicular traffic and a pub, creating micro-environmental conditions that poorly correlate with expected $PM_{2.5}$ levels based on the regression model. The observed high concentrations of $PM_{2.5}$ measurements from Static Sensor 4 suggest that the local sources of emissions in that specific area introduce complexities that are not fully captured by the model's inputs. Consequently, the model may have difficulty accurately associating the $PM_{2.5}$ measurements from Static 4 with the predictor variables, leading to its lower importance in the overall model. The model likely identifies these static sensors as indicators of the average pollution conditions in the area, and the mobile units as effective indicators of localized pollution sources and patterns. Moreover, the consistent nature of data collection from static sensors provides a stable reference for understanding baseline pollution levels in specific regions. $PM_{2.5}$ measurements from BAQS also emerged as a crucial predictor variable across all calibration models. Situated within the university campus at an urban background site less affected by emission peaks at roadside, BAQS is anticipated to predominantly reflect the regional background.

The model (Standard RF regression) highlights the pivotal role of traffic-related variables in predicting $PM_{2.5}$ concentrations, with buses

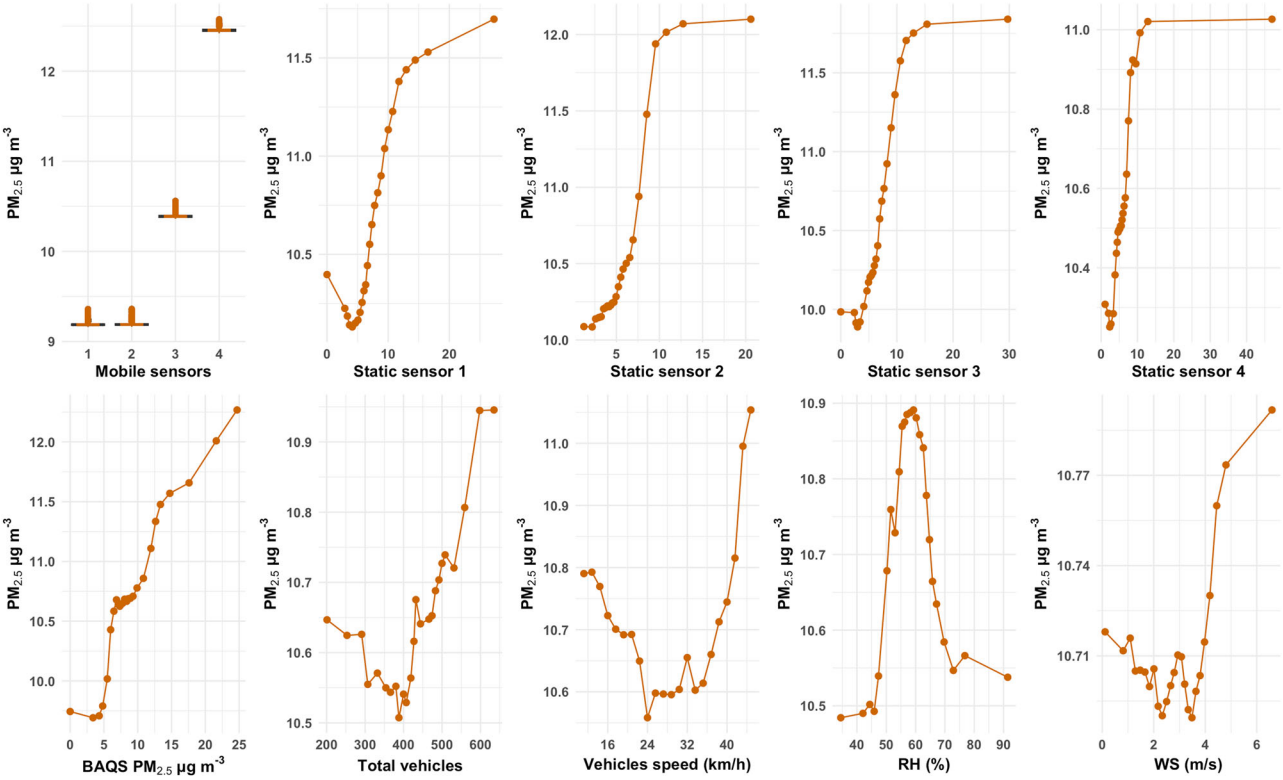

Fig. 5 | Partial dependence of $PM_{2.5}$ with the studied input features. The seperate panels provide the Partial Dependence Plots (PDPs) of $PM_{2.5}$ with the different studied input features.

emerging as the most influential factor. This underscores their substantial contribution to urban air pollution, given their diesel-fuelled engine emissions and consistent routes (see Fig. S6). Notably, the presence of seven bus stops along the primary road further amplifies the impact of bus-related traffic on the local air quality. 'Vehicle Speed' has significant importance, emphasizing the influence of traffic flow dynamics on air quality. Higher speeds contribute to increased pollutant levels. Additionally, the variable 'Total Vehicles' which is inclusive of all the vehicles in the area is an important predictor, suggesting areas with a higher traffic volume, exhibit larger $PM_{2.5}$ concentrations. Meteorological variables also contributed to the prediction of $PM_{2.5}$ concentrations confirming the role of meteorology on atmospheric dynamics and on the variability of particulate matter. While relative humidity can influence the physical properties and measurement of particulate matter through aerosol hygroscopicity, this effect was accounted for during the calibration of the $PM_{2.5}$ data used in the model. Atmospheric pressure shapes general atmospheric dispersion patterns, impacting the vertical mixing of pollutants, and is reflected in the VIMP rankings. Temperature and relative humidity, acting as proxies for time of day and boundary layer height, influence chemical reactions, particle volatility, and pollution dispersion dynamics. Meanwhile, wind direction and speed are critical for directly determining pollutant transport pathways, revealing potential sources of $PM_{2.5}$. The population density exhibited a relatively small contribution as a predictor variable, despite its typical association with anthropogenic emission sources impacting particulate matter levels. This low importance of population could potentially be due to the model capturing the impacts of population-related emissions through other more direct predictor variables, such as mobile/static sensor data, traffic counts, etc.

Figure 5 displays the Partial Dependence Plots (PDPs) for the Standard RF regression of $PM_{2.5}$ with important variables from the input dataset. PDPs illustrate the relationship between a specific feature and the model's predictions for the target variable[36,37]. PDPs have been used in air quality studies to explore relationships with meteorological variables[38] and other non-linear factors[39–42]. These plots depict how changes in each feature influence the average predicted $PM_{2.5}$ while holding other variables constant, providing valuable insights into the impact of individual factors on the model's outcomes. Among the PDPs generated, particular attention is drawn to the plot for the 'Mobile Sensors' variable which reveals significant fluctuations in $PM_{2.5}$ predictions, indicating varying impacts of different sensor IDs on the model's predictions. This suggests that certain sensors consistently capture crucial patterns or localized pollution sources, as evidenced in the discernible fluctuations in the plot. However, it's important to note that OPC 4 consistently records high PM levels, while OPC 1 and OPC 2 exhibit similar patterns, and OPC 3 falls somewhere in between. This observation raises questions about the transferability of sensor data between OPC 1 or OPC 2 to OPC 4, suggesting potential limitations in predictive capabilities across different sampling conditions and the model's adaptability to specific OPC units. Nonetheless, the unexpectedly high importance attributed to sensor ID underscores the significance of the location of individual sensors in predicting $PM_{2.5}$ levels. A consistent pattern was seen among the static sensors, where changes in the static sensor data led to uniform shifts in the model's predicted $PM_{2.5}$ levels. This suggests that the static sensors exhibit similar behaviour in capturing $PM_{2.5}$ levels, despite potential differences in their exact locations or environmental conditions. Static Sensor 4 shows a limited local effect at higher concentrations, suggesting it may not be capturing site-specific variations effectively in the model. Similarly, Static Sensor 1 located at Tiverton Academy associates well with the model-predicted $PM_{2.5}$ at lower concentrations but not at higher levels, indicating that the model may not fully capture the broader area's $PM_{2.5}$ levels effectively. This can also suggest a limited spatial representativeness, which could be due to the sensor's specific location or local environmental factors that differ significantly from the general conditions of the study area. In contrast, the BAQS data is consistently captured by the model and is associated till the highest $PM_{2.5}$ concentrations. This suggests that the BAQS location is highly representative of the area's overall air quality, effectively capturing the spatial distribution of $PM_{2.5}$.The strong

correlation between BAQS data and model predictions validates its value as a reference point for area-wide $PM_{2.5}$ levels. From the PDP of Total vehicles, we observe the largest contribution to $PM_{2.5}$ at highest traffic volumes, lowest $PM_{2.5}$ correlated to low-moderate traffic volumes (300–400 vehicles $hour^{-1}$) and an increase in $PM_{2.5}$ with minimum traffic rates. This is consistent with the relation between vehicular emissions and vehicular speed: during traffic peaks, $PM_{2.5}$ is highest because of traffic queues and jams; lower $PM_{2.5}$ is associated to a moderate traffic flow, with vehicles moving at optimal cruise speed in terms of consumptions and emissions. Finally, the increase in $PM_{2.5}$ associated with low traffic volume is probably due to the higher speed of vehicles in this traffic condition, which is also associated with higher emission rates. This is consistent with the PDP for vehicular speed, where highest predicted $PM_{2.5}$ is associated with vehicle speed >40 km h^{-1} and lowest $PM_{2.5}$ occurs for a speed in the range 25–30 km h^{-1}. The pattern seen in the PDP of wind speed with the model-predicted $PM_{2.5}$ is typical, consistent with a study conducted in London[43], the detected high concentrations during both low and high wind speeds are indicative of stagnation and resuspension phenomena respectively. Conversely, the noted decrease in concentrations amidst moderate wind speeds suggests effective dilution processes. In examining the various meteorological variables, the differences in predicted concentrations appear relatively small. Despite numerical discrepancies as modest as ±0.4 μg m^{-3} in certain variables, these variations become meaningful within the broader context of our analysis. It is important to recognize the sensitivity of our model to scale, where seemingly small numerical changes may hold practical implications based on the specific range of our data. These fine-scale variations can significantly impact air quality compliance assessments and exposure calculations for public health monitoring.

Discussion

This research addresses a critical gap in air pollution monitoring by introducing a novel spatiotemporal prediction approach to fill missing data in low-cost particulate matter sensors, specifically focusing on $PM_{2.5}$ concentrations. Our primary objective was to predict $PM_{2.5}$ concentrations at fine spatial and temporal scales, even in areas lacking direct observations. One key aspect of our contribution lies in the ability of our methodology to predict missing $PM_{2.5}$ measurements even when mobile sensors were not actively collecting data or, in other words, no one was walking around with a sensor in the area. This feature is crucial for extending the applicability of the model to scenarios where continuous data collection may not be feasible or practical. The robust performance of the RF model, demonstrated through calibration and evaluation processes, emphasizes its reliability in estimating $PM_{2.5}$ concentrations beyond active sensor deployment.

The study has several limitations, including the relatively short duration of the campaign (one month) and the relatively small study area in Selly Oak, Birmingham. The short duration and limited geographical scope may limit the capture of the full range of seasonal variations and diverse atmospheric conditions that can impact $PM_{2.5}$ levels. To enhance the generalizability of the model, future studies should incorporate data collected over more extended periods and from diverse locations, spanning different seasons and weather conditions. Additionally, the citizen science approach employed in data collection, while fostering community engagement, introduces certain challenges. The variability in data collection frequency and routes taken by individuals carrying mobile sensors may introduce biases. Moreover, participant behaviours such as smoking can also introduce biases that affect the measurements. Future research should explore strategies to standardize data collection procedures and address potential biases associated with the citizen science aspect.

The hybrid dataset integration encompassed meteorological parameters, aerosol properties, and spatially resolved proxy variables, contributing to the model's accuracy. Our approach demonstrated the ability to predict $PM_{2.5}$ levels but also showcased the adaptability of the model across different sensors and road types. The importance of the study lies in its potential to significantly enhance spatial resolution beyond regulatory monitoring infrastructure, providing refined air quality predictions and improving exposure assessments. The findings underscore the performance of machine learning-based approaches, particularly tree-based algorithms, allowing to advance our understanding of $PM_{2.5}$ dynamics and of their implications for public health.

Our methodology also distinguishes itself through its remarkable computational efficiency and speed. The utilization of low-cost Optical Particle Counters in both static and mobile configurations offers a cost-effective solution for air quality monitoring and significantly reduces the computational burden. In contrast to more complex models, our approach excels in delivering high prediction accuracy while maintaining relatively lower computational costs. Notably, the second and the third ML calibration approaches showcase that relatively few mobile monitors can effectively characterize air pollution levels across an entire area at different times of day when measurements are made. This finding has profound implications for citizen science initiatives, as it suggests that the contributions of a small number of participants can significantly enhance our understanding of local air quality patterns. The efficiency of the RF algorithm ensures swift model training and prediction, making it an ideal choice for real-time air quality assessments. This computational efficiency enhances the speed of our predictions and makes the methodology more accessible for widespread implementation. The developed approach achieves a favourable balance between prediction accuracy and computational performance, rendering it a practical and readily deployable solution for air quality monitoring across diverse urban landscapes. The demonstrated success of the machine learning approach encourages further exploration and application in urban air quality assessments, ultimately contributing to improved public health outcomes.

Methodology

Study area

The area of study is an approximately 1 × 1 km^2 block in Selly Oak, situated approximately 3 km south-west of city centre of Birmingham, UK, which is a major city with population of 1.14 million. Selly Oak is highly influenced by its close proximity to the University of Birmingham, which is just to the north. The community is deeply influenced by the academic institution, hosting the main Edgbaston campus, and serving as a prominent residential area for students (hosting around 38,000 students and 8000 staff). The northern and southern boundaries of the block are delineated by two busy roads, namely Bristol Road and Raddlebarn Road, respectively (Fig. 6). The area has a railway station to the west of the block that connects it to the city centre and other parts of the West Midlands.

Low-cost optical particle counters

During the measurement campaign from April 15th to June 20th, 2023, Alphasense optical particle counters (OPC-N3, Alphasense Ltd), which cost approximately 350 GBP per unit, were employed in both stationary and mobile configurations. The OPC-N3 employs a Class 1 laser (wavelength ~658 nm) to detect, size, and count particles within the range of 0.35–40 μm, distributed across 24 size bins. An embedded algorithm translates this size distribution into estimated mass concentrations within the PM_1, $PM_{2.5}$, and PM_{10} size fractions. Employing an internal fan, the OPC-N3 draws the air sample through the detection region, with a total flow rate of 5.5 L min^{-1}. For this instrument, the manufacturer declares a sensitivity of <1 μg m^{-3} and a measurement range spanning from 0 to 2000 μg m^{-3}.

The OPC-N3 was configured with the default settings for particle refractive index and particle density, with values set at 1.5 and 1.65 g cm^{-3}, respectively. Measurements were stored at 10-second intervals and included the following parameters: date, number size distribution, flow rate, relative humidity (RH), temperature and estimated concentrations for PM_1, $PM_{2.5}$, and PM_{10}. In static configurations, the sensor drew power from a car battery housed within a robust all-weather box (Fig. 7a), providing reliable operation for a span of 18–20 days. All the four static sensors were strategically positioned to capture comprehensive data across the study area. The need for battery replacement, conducted as part of routine maintenance was needed to sustain the continuous functionality of the sensors in

Fig. 6 | Maps of the study area in Selly Oak, Birmingham. Left hand map plot provides the location of the city of Birmingham within the UK. The upper plot provides the location of the Selly Oak area within the city of Birmingham. The bottom (zoomed in) plot provides the locations of 4 static LCS (blue pinpoints), the area where the mobile measurements were conducted (outlined in red), and the location of the Birmingham Air Quality Supersite (BAQS) inside the University of Birmingham campus. Maps ©Google Maps.

Fig. 7 | Photographs of the $PM_{2.5}$ monitoring equipment. a Static sensor setup including car battery for power, **b** Mobile sensor setup placed within backpack.

environments where access to the mains was not feasible. Additionally, 4 mobile sensors were deployed at street level, using a user-friendly backpack setup that was both lightweight (<1 kg) and easy to handle (Fig. 7b). These sensors ran on a mobile power bank (5000 mAh), requiring a recharge every two days to keep them working smoothly without any interruptions. Using a citizen science approach, local businesses and schools actively contributed to data collection from static sensors throughout the 2-month period, while university students similarly engaged in the collection of data from mobile sensors over a 1-month period (15th May–20th June 2023). Within the enclosed box, a Bosch BME-280 sensor was present for measuring the relative humidity and temperature. Additionally, a GSM module was included in the microcontroller connected to the OPC-N3 and served as real-time clock and allowed the transmissions of the measurements to a dedicated cloud server.

Birmingham Air Quality Supersite (BAQS)

Birmingham Air Quality Supersite (BAQS) is one of three highly instrumented air quality stations in the United Kingdom. The regulatory station, characterised as an urban background is located at the grounds of the University of Birmingham (52.45°N, 1.93°W), about 3 km southwest from the city centre[44]. The reference instrument at BAQS was the PALAS Fidas 200 providing continuous and simultaneous measurements of PM_1, $PM_{2.5}$, PM_4, PM_{10}, TSP (PM_{tot}) and the particle number concentration with a time resolution of 1 minute.

Calibration

The sensors underwent calibration through comparison with research-grade instruments at BAQS during two collocation periods, conducted before and after the campaign, totalling 4 days of simultaneous measurements. During this process, outliers (~5% of measurements) were removed, primarily those occurring under extreme humidity conditions, and an exponential model was developed to calibrate each OPC-N3 by comparing the ratio of its $PM_{2.5}$ measurements to those of the Fidas instrument under varying humidity conditions. While low-cost optical particle counters typically overestimate $PM_{2.5}$ and PM_{10} concentrations in high humidity due to particle hygroscopicity effects[13,45,46], the relatively low humidity during this campaign minimized discrepancies between sensor and reference instrument readings. Nevertheless, the calibration significantly improved the precision of the low-cost OPC estimates, achieving Pearson correlation coefficient (r) for $PM_1 = 0.81$ to 0.84, $PM_{2.5} = 0.63$ to 0.75, $PM_{10} = 0.32$ to 0.57. The meteorological sensor measurements, though affected by their container placement which led to humidity underestimation and temperature overestimation, showed strong correlation with the Elms Road meteorological station readings at the University of Birmingham ($r > 0.95$), allowing for effective calibration through simple linear regression[47]. BAQS proved to be representative of the broader study area, as demonstrated by strong correlations between BAQS and the study sites for both relative humidity ($r > 0.95$) and particulate matter measurements (r up to 0.68).

Hybrid dataset and ML model

Building upon the comprehensive understanding of the study area and the deployment of a set of low-cost OPC-N3 detailed in Sections 'Study Area' and 'Low-cost Optical Particle Counters', our methodology proceeds to the construction of a hybrid dataset and the implementation of machine learning (ML) models.

Hybrid dataset preparation. To address the challenge of predicting $PM_{2.5}$ levels at fine spatial and temporal scales, we compiled a hybrid dataset incorporating additional information along with that gathered by both static and mobile low-cost sensors. The road network of the whole study area was ranked into primary, secondary, and tertiary segments based traffic volume characteristics using telematics data collected for the area[48,49]. Furthermore, continuous traffic data was captured through a traffic monitoring camera next to a supermarket from Bristol Road, a key thoroughfare in the study area. These data were provided by the Birmingham City Council and included vehicle count and vehicle speed differentiated by vehicle type (motorbike, passenger car, Trailer, Rigid, Heavy Good Vehicle (HGV), bus). This hybrid dataset also integrated demographic (census) data and meteorological variables from BAQS, including wind speed, wind direction, atmospheric pressure, RH, and atmospheric temperature. $PM_{2.5}$ from the Fidas at BAQS was also incorporated into the dataset, finally resulting in a comprehensive and multi-faceted analytical framework to investigate the area. All measurements from the static sensor network and BAQS were aggregated into the hybrid dataset, temporally aligned with the mobile sampling intervals. VSP (vehicle specific power)[48], representing the instantaneous total power demand per vehicle mass, was initially included in the hybrid dataset from the telematics data but later excluded due to its reduced importance as determined by the machine learning model (see Fig. S5).

For spatial analysis and integration, the mobile sensor dataset was converted into a spatial object and reprojected to the British National Grid. To allow for a detailed analysis of the spatial distribution of $PM_{2.5}$ along the road network, the reprojection was divided into 30 m-long segments, and for each segment, the position of the corresponding centroid was computed (Fig. 8). The centroids served as focal points for the assignment of all spatially resolved proxy variables (e.g. demographic data, average traffic counts) and $PM_{2.5}$ by pedestrian mapping. The spatial analysis in this study was conducted using the open-source software QGIS. Details of the input data used can be found in Table 2.

ML model development and calibration approaches. Random Forest (RF) is a powerful ensemble supervised learning algorithm, introduced by Breiman (2001)[50], leveraging on the classification and regression trees

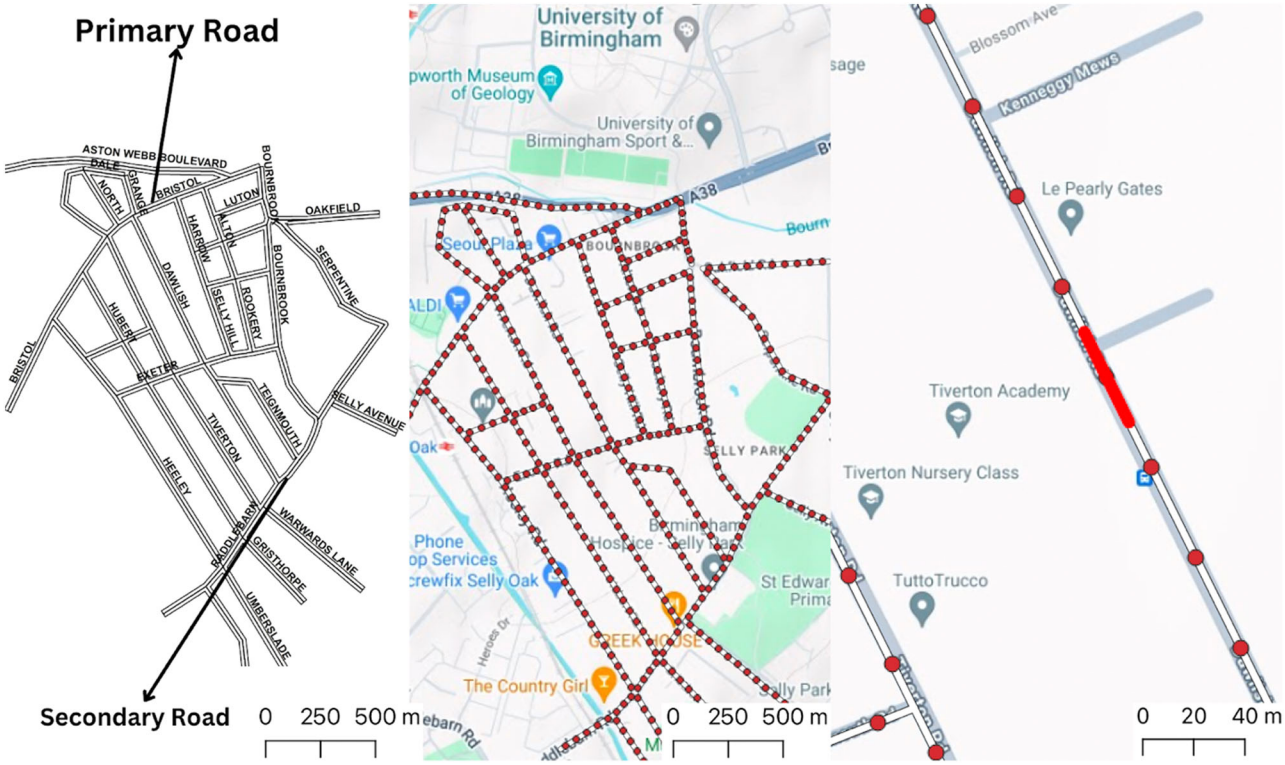

Fig. 8 | Maps of the Selly Oak area. The left hand panel provides the locations and names of the different studied roads. The middle panel provides the the locations of the calculated centroids which are placed at 30 m intervals. The right hand panel provides a zoomed in portion of the middel panel. Maps ©Google Maps.

Table 2 | Input parameters used in the ML-model

Context feature	Parameter	Data source
Meteorology	Temperature Relative humidity Wind speed Wind direction Atmospheric pressure	Birmingham Air Quality Supersite (BAQS)
Road network	Road type	OpenStreetMap
Traffic data	Traffic count by vehicle type (Motorbikes, Cars, Trailers, Rigids, HGVs, Buses). Frequency distribution of vehicle speed	Birmingham City Council
Population	Population	Office for National Statistics
Air Quality	$PM_{2.5}$	BAQS, Static Sensors
Categorical	Mobile sensor ID	Mobile sensors

algorithm (CART) for prediction. Employing a bagging approach with replacement, RF draws samples from various subsets of variables. Each decision tree within the RF is generated using the bootstrap method, with node splitting occurring through random subsets of variables (set by the *mtry* hyperparameter). The forest is expanded using a specified number of trees (hyperparameter *ntree*) to mitigate bias error[50]. The final predicted result is derived by aggregating and averaging predictions from individual trees, selected at random (Fig. 9). Random Forest simulations were conducted to select and optimize the hyperparameters. The model configuration involved the utilization of 1000 trees, limiting each split to a maximum of 5 randomly selected features (*mtry* = 5), and establishing a minimum node size of 5. In this study, RF modelling was performed using the 'RandomForestSRC' package in R. The modelling process took place on a dedicated High-Performance Computing (HPC) ARIES platform of the University of Modena and Reggio Emilia.

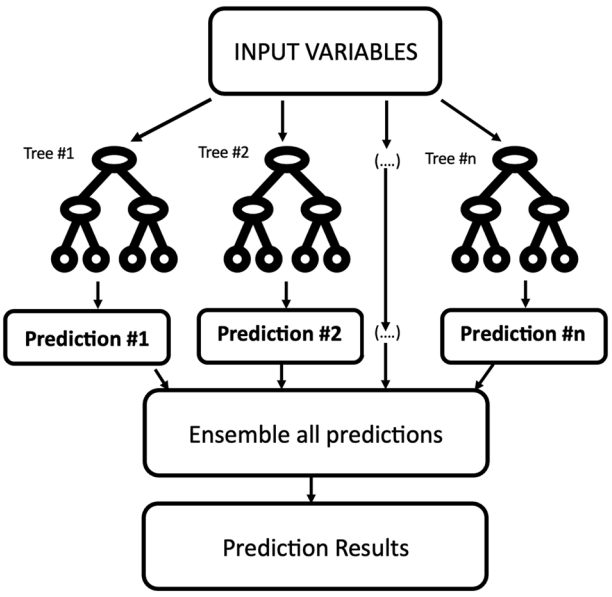

Fig. 9 | Schematic diagram of the random forest model employed in the study. The figure provides a simple flow diagram of the steps involved in the modelling process.

The calibration of predictive models for $PM_{2.5}$ concentrations involved the implementation of three specific approaches: 1) Standard Random Forest Regression (RF): This method employed the standard RF regression technique, utilizing an 80-20 train-test split. The model was trained on 80% of the hybrid dataset (95,072 data points) and tested on a 20% (23,769 data points) of the total (118,841 observations) to evaluate its predictive performance based on the provided input features. The total number of observations (118,841) includes only the relevant static

measurements tied to the mobile data i.e. only the static measurements matching the mobile timestamps were included in the dataset, rather than a continuous count of all possible static measurements. 2) Sensor Transferability Evaluation: This approach involved calibrating the RF model using the hybrid dataset but using only one of the four OPC-N3 involved in the pedestrian mapping. The performance of this approach was then assessed based on an independent OPC-N3 sensor. This evaluation aims to gauge the model's ability to generalize across different OPC-N3 sensors. 3) Road Transferability Evaluation: In this calibration approach, the model was calibrated using the hybrid dataset, but including only one road and its performance was evaluated on a different road. This evaluation explores the capability of the model to generalize across distinct roads in the area. The results of sensor calibration/validation pairs and selected road calibration/validation pairs evaluation are shown in the SI for clarity. The observed variability can be attributed to the diverse nature of citizen-collected data, with participants following different walking patterns, sampling times, and route choices. These variations in data collection inherently affect model predictions, emphasizing the importance of standardized protocols for future campaigns.

For each approach, the individual contribution of the variables to the model prediction was analysed and assessed using the variable importance (VIMP). This is achieved by quantifying the change in model error when a single variable is permuted, i.e., randomly shuffled.

This technique is widely adopted and has been extensively utilized across the machine-learning literature[50–54]

To validate these models, various performance metrics were computed. The coefficient of determination (R^2) was computed for the linear least-squares regression of predicted concentrations versus observed concentrations. Also, the Spearman Rank Coefficient (ρ) was computed, to assess the non-parametric linear correlation between model estimates and observations. Additionally, Mean Bias Error (MBE), the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE) were calculated to assess the average deviation between predicted and observed concentrations, to indicate the overall model bias. Detailed equations for each metric are provided in SI for clarity.

Data availability

The datasets generated and/or analysed during the current study are available in the the University of Birmingham Institutional Research Archive (UBIRA), https://doi.org/10.25500/edata.bham.00001144.

Code availability

All the relevant code can be obtained upon reasonable request from the corresponding author.

Received: 1 July 2024; Accepted: 26 November 2024;
Published online: 19 December 2024

References

1. European Environment Agency. *Europe's Air Quality Status 2021*. (Publications Office, LU, 2022).
2. Gelencsér, A. et al. Source apportionment of PM2.5 organic aerosol over Europe: primary/secondary, natural/anthropogenic, and fossil/biogenic origin. *J. Geophys. Res. Atmos.* **112**, 2006JD008094 (2007).
3. Pöschl, U. Atmospheric aerosols: composition, transformation, climate and health effects. *Angew. Chem. Int. Ed.* **44**, 7520–7540 (2005).
4. Murray, C. J. L. et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1223–1249 (2020).
5. Liu, C. et al. Ambient particulate air pollution and daily mortality in 652 cities. *N. Engl. J. Med.* **381**, 705–715 (2019).
6. Pappa, A. & Kioutsioukis, I. Forecasting particulate pollution in an urban area: from Copernicus to Sub-Km scale. *Atmosphere* **12**, 881 (2021).
7. Southerland, V. A. et al. Global urban temporal trends in fine particulate matter (PM2·5) and attributable health burdens: estimates from global datasets. *Lancet Planet. Health* **6**, e139–e146 (2022).
8. Minguillón, M. C., Querol, X., Baltensperger, U. & Prévôt, A. S. H. Fine and coarse PM composition and sources in rural and urban sites in Switzerland: local or regional pollution? *Sci. Total Environ.* **427–428**, 191–202 (2012).
9. Jerrett, M. et al. A review and evaluation of intraurban air pollution exposure models. *J. Expo. Sci. Environ. Epidemiol.* **15**, 185–204 (2005).
10. Li, H. Z., Dallmann, T. R., Gu, P. & Presto, A. A. Application of mobile sampling to investigate spatial variation in fine particle composition. *Atmos. Environ.* **142**, 71–82 (2016).
11. Baruah, A., Zivan, O., Bigi, A. & Ghermandi, G. Evaluation of low-cost gas sensors to quantify intra-urban variability of atmospheric pollutants. *Environ. Sci. Atmos.* **3**, 830–841 (2023).
12. Di, Q. et al. Air pollution and mortality in the medicare population. *N. Engl. J. Med.* **376**, 2513–2522 (2017).
13. Crilley, L. R. et al. Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmos. Meas. Tech.* **11**, 709–720 (2018).
14. Zimmerman, N. et al. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmos. Meas. Tech.* **11**, 291–313 (2018).
15. Bigi, A., Mueller, M., Grange, S. K., Ghermandi, G. & Hueglin, C. Performance of NO, NO_2 low cost sensors and three calibration approaches within a real world application. *Atmos. Meas. Tech.* **11**, 3717–3735 (2018).
16. Morawska, L. et al. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone? *Environ. Int.* **116**, 286–299 (2018).
17. Fattoruso, G. et al. Site suitability analysis for low cost sensor networks for urban spatially dense air pollution monitoring. *Atmosphere* **11**, 1215 (2020).
18. Zheng, T. et al. Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. *Atmos. Meas. Tech.* **11**, 4823–4846 (2018).
19. Badura, M., Batog, P., Drzeniecka-Osiadacz, A. & Modzel, P. Evaluation of low-cost sensors for ambient $PM_{2.5}$ monitoring. *J. Sens.* **2018**, 1–16 (2018).
20. Castell, N. et al. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ. Int.* **99**, 293–302 (2017).
21. Sayahi, T., Butterfield, A. & Kelly, K. E. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. *Environ. Pollut.* **245**, 932–940 (2019).
22. Hassani, A., Castell, N., Watne, Å. K. & Schneider, P. Citizen-operated mobile low-cost sensors for urban PM2.5 monitoring: field calibration, uncertainty estimation, and application. *Sustain. Cities Soc.* **95**, 104607 (2023).
23. deSouza, P. et al. Air quality monitoring using mobile low-cost sensors mounted on trash-trucks: Methods development and lessons learned. *Sustain. Cities Soc.* **60**, 102239 (2020).
24. Mueller, M. D., Hasenfratz, D., Saukh, O., Fierz, M. & Hueglin, C. Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network. *Atmos. Environ.* **126**, 171–181 (2016).
25. Singh, A. et al. Air quality assessment in three East African cities using calibrated low-cost sensors with a focus on road-based hotspots. *Environ. Res. Commun.* **3**, 075007 (2021).
26. Prank, M. et al. Evaluation of the performance of four chemical transport models in predicting the aerosol chemical composition in Europe in 2005. *Atmos. Chem. Phys.* **16**, 6041–6070 (2016).
27. Guo, L., Chen, B., Zhang, H. & Zhang, Y. A new approach combining a simplified FLEXPART model and a Bayesian-RAT method for

forecasting PM10 and PM2.5. *Environ. Sci. Pollut. Res.* **27**, 2165–2183 (2020).
28. Ma, J., Yu, Z., Qu, Y., Xu, J. & Cao, Y. Application of the XGBoost machine learning method in PM2.5 prediction: a case study of Shanghai. *Aerosol Air Qual. Res.* **20**, 128–138 (2020).
29. Taheri Shahraiyni, H. & Sodoudi, S. Statistical modeling approaches for PM10 prediction in urban areas; a review of 21st-century studies. *Atmosphere* **7**, 15 (2016).
30. Bai, L., Liu, Z. & Wang, J. Novel hybrid extreme learning machine and multi-objective optimization algorithm for air pollution prediction. *Appl. Math. Model.* **106**, 177–198 (2022).
31. Yang, G., Lee, H. & Lee, G. A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. *Atmosphere* **11**, 348 (2020).
32. Kim, B.-Y., Cha, J. W., Chang, K.-H. & Lee, C. Estimation of the visibility in Seoul, South Korea, based on particulate matter and weather data, using machine-learning algorithm. *Aerosol Air Qual. Res.* **22**, 220125 (2022).
33. Tella, A. & Balogun, A.-L. GIS-based air quality modelling: spatial prediction of PM10 for Selangor State, Malaysia using machine learning algorithms. *Environ. Sci. Pollut. Res.* **29**, 86109–86125 (2022).
34. Kim, B.-Y., Lim, Y.-K. & Cha, J. W. Short-term prediction of particulate matter (PM10 and PM2.5) in Seoul, South Korea using tree-based machine learning algorithms. *Atmos. Pollut. Res.* **13**, 101547 (2022).
35. Barthwal, A., Acharya, D. & Lohani, D. Prediction and analysis of particulate matter (PM2.5 and PM10) concentrations using machine learning techniques. *J. Ambient Intell. Humaniz. Comput.* **14**, 1323–1338 (2023).
36. Greenwell, B. M., Boehmke, B. C. & McCarthy, A. J. A simple and effective model-based variable importance measure. https://doi.org/10.48550/ARXIV.1805.04755 (2018).
37. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat*. **29**, 1189–1232 (2001).
38. Grange, S. K., Carslaw, D. C., Lewis, A. C., Boleti, E. & Hueglin, C. Random forest meteorological normalisation models for Swiss PM10 trend analysis. *Atmos. Chem. Phys.* **18**, 6223–6239 (2018).
39. Liu, M., Chen, H., Wei, D., Wu, Y. & Li, C. Nonlinear relationship between urban form and street-level PM2.5 and CO based on mobile measurements and gradient boosting decision tree models. *Build. Environ.* **205**, 108265 (2021).
40. Hassani, A. et al. Low-cost sensors and machine learning aid in identifying environmental factors affecting particulate matter emitted by household heating. *Atmos. Environ.* **314**, 120108 (2023).
41. Houdou, A. et al. Interpretable machine learning approaches for forecasting and predicting air pollution: a systematic review. *Aerosol Air Qual. Res.* **24**, 230151 (2024).
42. Shi, H., Yang, N., Yang, X. & Tang, H. Clarifying relationship between PM2.5 concentrations and spatiotemporal predictors using multi-way partial dependence plots. *Remote Sens.* **15**, 358 (2023).
43. Jones, A. M., Harrison, R. M. & Baker, J. The wind speed dependence of the concentrations of airborne particulate matter and NOx. *Atmos. Environ.* **44**, 1682–1690 (2010).
44. Alam, M. S. et al. Diurnal variability of polycyclic aromatic compound (PAC) concentrations: Relationship with meteorological conditions and inferred sources. *Atmos. Environ.* **122**, 427–438 (2015).
45. Khreis, H., Johnson, J., Jack, K., Dadashova, B. & Park, E. S. Evaluating the performance of low-cost air quality monitors in Dallas, Texas. *Int. J. Environ. Res. Public. Health* **19**, 1647 (2022).
46. Crilley, L. R. et al. Effect of aerosol composition on the performance of low-cost optical particle counter correction factors. *Atmos. Meas. Tech.* **13**, 1181–1193 (2020).
47. Bousiotis, D. et al. Pinpointing sources of pollution using citizen science and hyperlocal low-cost mobile source apportionment. *Environ. Int*. 109069 https://doi.org/10.1016/j.envint.2024.109069 (2024).
48. Ghaffarpasand, O. & Pope, F. D. Telematics data for geospatial and temporal mapping of urban mobility: Fuel consumption, and air pollutant and climate-forcing emissions of passenger cars. *Sci. Total Environ.* **894**, 164940 (2023).
49. Ghaffarpasand, O. & Pope, F. D. Telematics data for geospatial and temporal mapping of urban mobility: New insights into travel characteristics and vehicle specific power. *J. Transp. Geogr.* **115**, 103815 (2024).
50. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
51. Louppe, G., Wehenkel, L., Sutera, A. & Geurts, P. Understanding variable importances in forests of randomized trees. in *Advances in Neural Information Processing Systems* (eds. Burges, C. J., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. Q.) vol. 26 (Curran Associates, Inc., 2013).
52. Ishwaran, H. Variable importance in binary regression trees and forests. *Electron. J. Stat*. **1**, 519–537 (2007).
53. Genuer, R., Poggi, J.-M. & Tuleau-Malot, C. Variable selection using random forests. *Pattern Recognit. Lett.* **31**, 2225–2236 (2010).
54. Williamson, B. D., Gilbert, P. B., Carone, M. & Simon, N. Nonparametric variable importance assessment using machine learning techniques. *Biometrics* **77**, 9–22 (2021).

Acknowledgements

This paper and related research have been conducted during and with the support of the Italian national inter-university PhD course in Sustainable Development and Climate change (link: www.phd-sdc.it). The materials and study were funded by the European Commission as part of the RI-URBANS project (grant no. 1010362450).

Author contributions

A.B. and F.D.P. conceived the idea for this paper. A.B. led the processing of the data and wrote the first draft. F.D.P. and A.B. supervised the project. S.D. and D.B. helped in data collection. D.B. performed the sensor calibrations. A.B. helped with the refinement of the methodology. D.B., S.D., A.B., G.G, O.G., R.M.H and F.D.P. provided help and feedback on the analysis and edited the final draft.

Competing interests

All authors declare no financial or non-financial competing interests. Author Roy M. Harrison is Co-Editor-in-Chief of npj Climate and Atmospheric Science and was not involved in the journal's review of or decisions related to this manuscript.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41612-024-00859-z.

Correspondence and requests for materials should be addressed to Francis D. Pope.