Environmental Pollution 361 (2024) 124903

Contents lists available at ScienceDirect

Environmental Pollution

journal homepage: www.elsevier.com/locate/envpol

Detecting and quantifying $PM_{2.5}$ and NO_2 contributions from train and road traffic in the vicinity of a major railway terminal in Dublin, Ireland ☆

Shanmuga Priyan [a,*], Yuxuan Guo [a,b], Aonghus McNabola [a,b], Brian Broderick [a,b,c], Brian Caulfield [a,b], Margaret O'Mahony [a,b], John Gallagher [a,b,c]

[a] Department of Civil, Structural, and Environmental Engineering, Trinity College Dublin, University of Dublin, College Green, Dublin, 2, Ireland

[b] Centre for Transport Research and Innovation for People, Trinity College Dublin, University of Dublin, College Green, Dublin, 2, Ireland

[c] TrinityHaus, Trinity College Dublin, University of Dublin, College Green, Dublin, 2, Ireland

ARTICLE INFO

Keywords:
Railway station
Air quality sensors
Machine learning
Bivariate polar plot
Transport hubs

ABSTRACT

Air pollution from transport hubs is a recognised health concern for local urban inhabitants. Within the domain of transport hubs, significant attention has been given to larger airport and port settings, however concerns have been raised about emissions from urban railway hubs, especially those with diesel trains. This paper presents an approach that adopts low-cost monitoring (LCM) for fixed site monitoring (FSM) to quantify and disaggregate $PM_{2.5}$ and NO_2 contributions from railway station and road traffic on air quality in the vicinity of railway station in Dublin, Ireland. The NO_2 sensor showed larger discrepancies than the $PM_{2.5}$ sensor when compared to the reference monitor. Machine learning models (XGBoost and Random Forest (RF) regression) were applied to calibrate the LCM devices, with the XGBoost model (NO_2, $R^2 = 0.8$ and RSME = 9.1 μg/m^3 & $PM_{2.5}$, $R^2 = 0.92$ and RSME = 2.2 μg/m^3) deemed more appropriate than the RF model. Local wind conditions, pressure, $PM_{2.5}$ concentrations, and road traffic significantly impacted NO_2 model results, while raw $PM_{2.5}$ sensor readings greatly influenced the $PM_{2.5}$ model output. This highlights that the NO_2 sensor requires more input data for accurate calibration, unlike the $PM_{2.5}$ sensor. The monitoring results from the one-month monitoring campaign from May 25, 2023 to June 25, 2023 presented elevated NO_2 and $PM_{2.5}$ concentrations measured at the railway station, which translated to exceedances of the annual WHO limits ($PM_{2.5}$ = 5 μg/m^3, NO_2 = 10 μg/m^3) by 1.6–1.8 and 3.2–5.2 times respectively at the study site. A subsequent data filtering technique based on wind orientation, revealed that the railway station was the main $PM_{2.5}$ source and road traffic was the main NO_2 source when winds come from the railway station. This study highlights the value of LCM devices alongside robust machine learning techniques to capture air quality in urban settings.

1. Introduction

Air pollution has been associated with a wide number of serious health impacts, which include cardiovascular disease, lung disease, and chronic obstructive pulmonary disease, as well as premature death (Manisalidis et al., 2020; Phairuang et al., 2024; Saucy et al., 2021). Despite strong efforts to lower emissions through diverse methods, air pollution from the transportation sector continues to be a major source of elevated concentrations in the urban environment today (Cheng et al., 2011; Hakkim et al., 2021; Phairuang et al., 2021).

Rail is considered as a greener mode of transport when compared with air and road as it has a lower rate of emissions per passenger and thus a lesser relative impact on climate change and the environment (Font et al., 2020). The entire transportation industry is responsible for around 22% of worldwide emissions (Railway Technology, 2016), with railway transport only accounting for 0.8% of the sector. However, railway services, specifically powered by diesel emit significant amount of air pollutants along their routes and while stationed. In the European union (EU-27), diesel trains contribution to the mobile sources of $PM_{2.5}$, black carbon, and nitrogen oxides emissions were estimated to be 2.8%, 2.5% and 2%, respectively (Borken-kleefeld & Ntziachristos, 2012). 60% of the European rail network is already electrified with 80% of the train traffic running on these lines (European Commission, 2023). However, in Ireland, only 2% of railway lines are equipped with

☆ This paper has been recommended for acceptance by Calde

* Corresponding author.
E-mail address: rs10@tcd.ie (S. Priyan).

https://doi.org/10.1016/j.envpol.2024.124903
Received 2 July 2024; Received in revised form 30 August 2024; Accepted 5 September 2024
Available online 6 September 2024

electrification ranking it lowest in Europe (European Commission, 2023).

The impacts of major railway terminals on local ambient air quality equally requires further research. Many studies have reported air quality in subway systems (e.g. Grana et al., 2017; Kim et al., 2008; Martins et al., 2016; Passi et al., 2021; Vilcassim et al., 2014). Their focus has been on assessing the PM_{10}, $PM_{2.5}$ and ultrafine particles (UFP) concentrations and their sources in the subway network, where there is a sole presence of non-exhaust emission in the absence of diesel-powered trains. Only three studies have been performed in ground level railway station environments, all of which were conducted in the UK: in Birmingham New Street (Hickman et al., 2018); in London Paddington (Chong et al., 2015); and in Edinburgh Waverley and London King's Cross stations (Font et al., 2020). Hickman et al. (2018) and Chong et al. (2015) focused mainly on assessing the air quality within and outside the station environment, whilst Font et al. (2020) also applied machine learning to evaluate the impact of individual diesel-powered trains on measured concentrations. Comparatively no source apportionment work has been undertaken in the vicinity of the railway station to determine the railway station emissions impact on local air quality. Furthermore, no work has been undertaken in railway hubs of this nature with a high predominance of diesel trains as those currently existing in Ireland.

The concentration of air pollutants in the vicinity of the railway terminals is partially influenced by the emissions from the trains; these includes NO_2 and particle matter (PM) from the diesel-powered train exhaust at idling and notch 1 condition (accelerating and deaccelerating in station) (Ozigis et al., 2021), and particles generated by wheel and brake wear of the train (Font et al., 2020; Hickman et al., 2018). The emissions from diesel trains are lower during idling than during the notching operation. However, still the idling emissions are significant. The PM_{10}, NO_x, CO, and Hydrocarbon (HC) emissions from diesel locomotives during idling were estimated to be 43.4 mg/m^3, 147 ppm, 94 ppm, 4.3 ppm, respectively whilst during notch 1 condition were found to be 60 mg/m^3, 243 ppm, 53 ppm, 4 ppm, respectively (Kim et al., 2020). In addition, train emissions, eateries, and the resuspension of dust from commuters and train movement are also sources of emissions.

Studies have implemented different source apportionment techniques to quantify the contribution of airport and port sources to local air quality (Carslaw et al., 2006; de Foy et al., 2021; Hagler et al., 2021; Han et al., 2022; Ledoux et al., 2018). For instance, Ledoux et al. (2018) analysed three months of air quality data using bivariate polar plots to quantify gaseous and PM emissions from ships in the neighbourhood of port of Calais. PM_{10}, SO_2, NO_x emissions from ships were quantified to be 2, 51 and 50%, respectively. A study conducted over 2 years at Santa Cruz de Tenerife filtered the particle number concentration data using inland sea breeze periods, and SO_2 as the indicative species of ship emissions to apportion particle sources in the vicinity of the port. The study found 65–70% of the particles detected are related to ship emissions (González et al., 2011). Han et al. (2022) used a generalized additive model to examine the effect of airport related emissions on hourly NO_x levels in Xinzheng International Airport, China. The study found 13% of the NO_x levels measured in the vicinity of the airport are attributed to the sources of airport activities. These studies have adopted high-end air quality instruments to discriminate the sources. However, the use of low-cost air quality sensors, diffusion tubes, and machine learning models can also aid in analysing polluting sources in a cost-effective manner, making such studies more accessible and scalable for wider implementation. An attempt has been made in this study to detect and quantify the contributions of train and road traffic with the help of low-cost sensors, diffusion tubes, and a valuable machine learning model in the vicinity of a railway terminal.

In 2019, a continuous monitoring station in Ireland's national air quality network operated by the Environmental Protection Agency (EPA) located adjacent to one of Dublin's major railway terminals, Heuston station, recorded exceedances of NO_2 (UTRAP, 2021). The railway station is located at an intersection where heavy rail, public transport services such as buses and taxis, and private cars intersect. The area also contains residential development within the vicinity of the station in several directions. This study aims to use low-cost monitoring (LCM) devices alongside calibration techniques to assess the impact of this large railway hub on localised air quality. It focuses on detecting, quantifying, and disaggregating the contributions of road traffic and railway station emissions to $PM_{2.5}$ and NO_2 levels measured in its vicinity. Machine learning models were employed to calibrate the low-cost devices, and data filtration methods were used to capture the contributions from railway station and road traffic. This combined measurement and modelling investigation can help key stakeholders determine the proportionality of impact on air pollution exceedances attributed to public and private sector transport modes in a cost-effective manner.

2. Methodology

2.1. Site description

Heuston railway station is a terminal station with tracks aligned in an East-West direction. The station is partially enclosed, with four platforms housed under a glazed roof (with longitudinal vent on its roof, approximately 30 cm height), while two platforms are not enclosed. The longitudinal vent in the roof is the major outlet through which pollutants from the train exhaust exit the railway station. The primary external opening for the enclosed platforms is at the western end where trains exit and enter the station. There are two additional openings, a pedestrian and vehicles access point located south and north of the station, respectively. Averaged over May–June 2023, Heuston had approximately 134 trains/day in the station between Monday-Saturday, reducing to approximately 73 trains/day on Sundays. All of the trains from Heuston station scheduled during the monitoring campaign were fuelled by diesel. St John's Road West is one of the major roads in Dublin connecting the city the with the arterial M50 motorway via the N4 corridor. The average traffic on this road in June ranged between 26,000–27,000 vehicles/day (Dublin City Council, 2023).

Adjacent to the railway station is a continuous EPA fixed monitoring site located along St John's Road station, and which is situated within 3 km from Dublin city centre. This site is hereafter referred to as St. John's FSM. This location observed NO_2 exceedances in 2019 and therefore serves as a point of interest in determining the contribution of pollution from the railway station at the monitoring site. This station is a statutory monitoring point for reporting air quality as part of the national network in line with EU directives. It records hourly PM_{10}, $PM_{2.5}$, and NO_2 concentrations and is located opposite to a major road (St John's Road West) and Heuston railway station (Fig. 1).

2.2. Measurement campaign

2.2.1. p.m.$_{2.5}$ & NO_2 measurement using low-cost monitoring device

The measurement of $PM_{2.5}$ and NO_2 were carried out using a low-cost monitoring (LCM) device (ELOS outdoor station, Elichens). The device uses light-scattering and an electrochemical sensor to measure $PM_{2.5}$ and NO_2, respectively. Beside $PM_{2.5}$ and NO_2, device also measures additional parameters such as temperature, humidity, pressure, and noise. Specification of the sensors in the LCM device are given in Supplementary Table S1. Each parameter was measured at different time resolutions. For evaluation and calibration purposes, all measurements were averaged over 1-h periods. Two LCM devices (LCM_A and LCM_B) were used in this study. Before and after deploying them at the study site, both devices were co-located for 28-days at a city centre and traffic site to evaluate the precision of the devices. Measured parameters from both the devices showed good correlation and relatively minimal error. The performance statistics of the devices are provided in Supplementary Table S2. Linear adjustments were applied for all the parameters measured with LCM_B, with reference to LCM_A to further reduce the error bias between them.

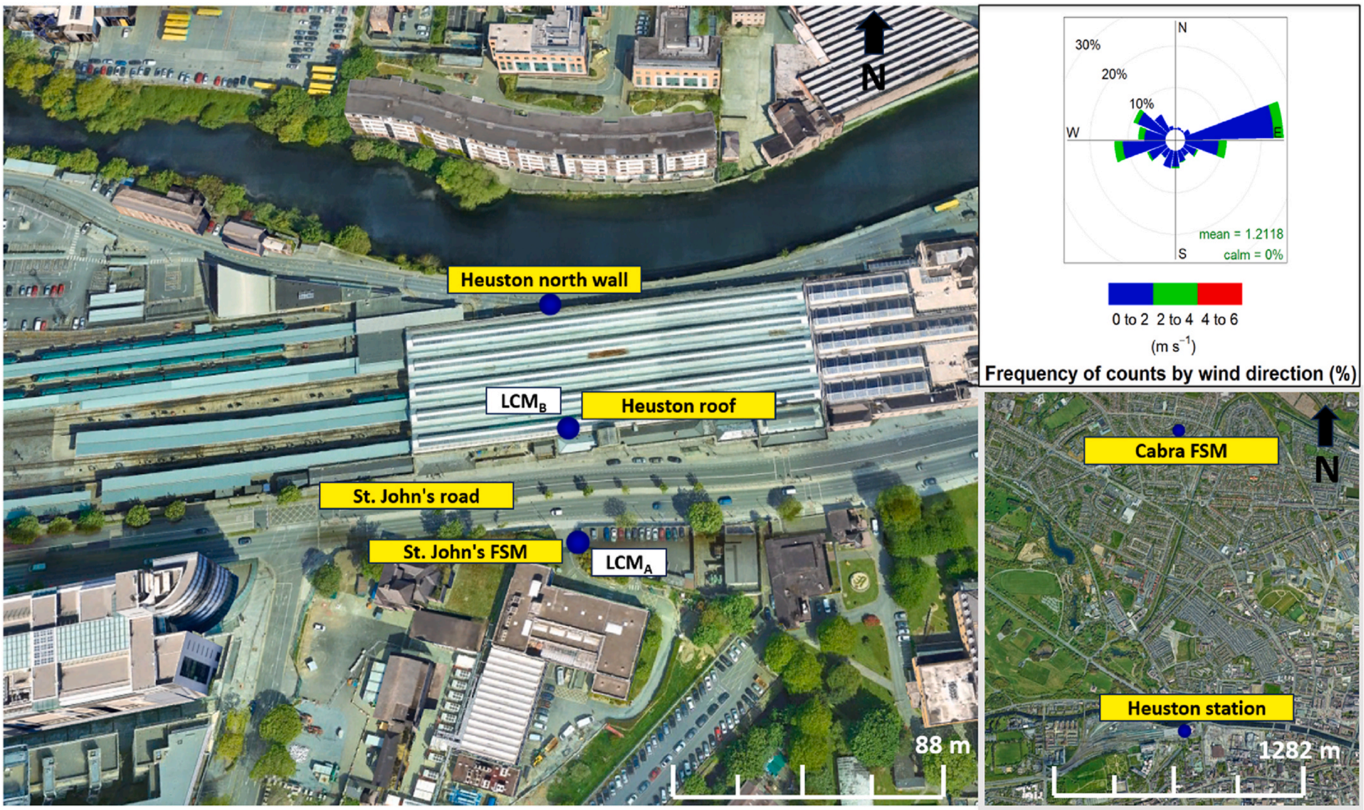

Fig. 1. Layout of monitoring site and insert of wind rose (25 May – June 26, 2023) at Heuston railway station in Dublin, Ireland.

During the monitoring campaign LCM$_A$ was co-located with St. Johns FSM, while LCM$_B$ was deployed on the roof of Heuston station (south end of the station and 6m above the ground level) and measurements were collected simultaneously. $PM_{2.5}$ and NO_2 concentrations were monitored continuously for a month from 25 May – June 26, 2023. The measurements collected by LCM$_A$ were used to develop the calibration model (with reference to St. Johns FSM measurements), while LCM$_B$ measurements were used in the analysis after calibration. The calibration model developed for LCM$_A$ was used for calibrating LCM$_B$ as the measurement performance of both the devices were very similar. A weather station (HOBO U30) was installed at Heuston roof to collect the hourly wind speed and direction for the aforementioned period. The winds were predominantly from the NEE direction (26%) during this period. Additionally, another nearby national EPA fixed station (Cabra community college) located approximately 2.1 km upwind of railway station was selected to act as a representation of the background air quality for $PM_{2.5}$. This site is hereafter referred to as Cabra FSM. Cabra FSM is a residential site surrounded by parks and educational establishments, with no major air pollution sources in its neighbourhood during the study period.

2.2.2. NO_2 measurement using diffusion tube

A Palmes-type diffusion tube was co-located with the LCM$_B$ device on the Heuston roof to measure average NO_2 concentrations over the monitoring period of one month. The accuracy of the NO_2 diffusion tubes (10–12% error) is higher than the calibrated NO_2 sensor values (Mueller et al., 2017). To further reduce the bias associated with the calibrated sensors, the low-cost sensor based NO_2 measurements were normalised using its average value (over the study period) after applying the calibration model and adjusted using average NO_2 value measured using diffusion tube. Additionally, a diffusion tube was installed in the northern wall of Heuston station to capture the upwind station background NO_2 concentration (Fig. 1). A set of diffusion tubes were also co-located alongside the St. Johns FSM. The results of these diffusion tubes were compared with St. Johns FSM NO_2 gas analyser and a correction factor was determined using the DEFRA bias adjustment tool (DEFRA, 2011). The correction factor (0.89) was then multiplied with diffusion tube NO_2 value measured at the Heuston station roof and northern wall.

2.3. Development of calibration model

The study evaluated two distinct machine learning algorithms (XGBoost and random forest), which are commonly found in the literature for calibrating low-cost sensors (Adong et al., 2022; Liu et al., 2021). XGBoost and random forest (RF) are ensemble models but employ boosting and bagging techniques, respectively (Pan et al., 2023). A k-fold cross validation approach was adopted to evaluate the two calibration models. In this approach, the dataset was split into k folds, and the model was trained and tested k times, using a different fold serving as the validation set during each iteration. This approach avoided performance overprediction, restricting leakage of data, and limiting prediction of past observations (Wang et al., 2023). The Python packages scikit-learn (sklearn) and XGBoost were used to perform RF and XGBoost regression, respectively. RandomizedSearchCV function was employed to perform k-fold cross validation and obtain the optimal hyperparameters for the models.

In the dataset, sensor readings of $PM_{2.5}$ and NO_2, meteorological variables such as wind speed (m/s), wind direction, humidity (%), temperature (Celsius), and pressure (millibar), temporal variables such as hours of the day, day of the month, and month of the year, enigineered variables such as rate of change of temperature per hour, rate of change of humidity per hour, and additional factors such as noise (Db), train schedule, and road traffic were all considered as the explanatory variables. The detailed information of the variables in the model are provided in the Supplementary Table S3. The reference monitor reading of $PM_{2.5}$ and NO_2 (St. Johns FSM values) were treated as the response variables. The models were trained with a test size of 20%, a random state of 42, and 5-fold cross-validation. A total of 750 data points were used to develop the models. Two calibration model were developed for $PM_{2.5}$: one based on a dataset with pollutant values less than or equal to median value; and one on a dataset with pollutant values above the median value to avoid overprediction of the response variables at lower concentration. However, the NO_2 calibration model was developed considering the whole dataset, as the NO_2 sensor exhibited a lower standard deviation (StDev) value. For each fold, trained models were evaluated against the test dataset using performance metrics such as root mean squared error (RMSE) and coefficient of determination (R^2). The result from each evaluation was then averaged to derive the final evaluation metric for the model, and a similar procedure was applied to determine the model's hyperparameters.

The Shapley Additive exPlanations (SHAP) method was then used to sort the variables based on their global impact and to compute the impact of each variables on the model output of each sample test, highlighting the complex relation between the response and explanatory variables (Wang et al., 2023).

2.3.1. *Source apportionment*

A data filtration technique helped disaggregate the contribution of road traffic and railway station emission sources. Data filtration is a quantitative technique that can be used to detect the sources contribution based on specific periods of interest such as selecting hours of the day when traffic is high, filtering data corresponding to upwind and downwind directions of monitoring sites, or based on specific activity such as filtering data corresponding to specific traffic movements (e.g., aircraft landing/take-off, ship hostelling and manoeuvring etc) (Carslaw et al., 2006; Merico et al., 2016) and transport characteristics (e.g., engine type, fuel type etc) (Schürmann et al., 2007). In this study, the pollutant concentrations were filtered based on corresponding upwind and downwind sectors of the Heuston railway station and St. Johns Road.

The road traffic emission contribution for NO_2 and $PM_{2.5}$ were estimated by considering the hourly concentrations in the specific wind sector of 110–250° when the road traffic emissions was considered more likely to have an effect on NO_2 and $PM_{2.5}$ concentrations (measured at Heuston roof) and by removing the background concentration by subtracting the pollutants concentration measured at St. Johns FSM (Table 1). Similarly, railway station emissions contribution for NO_2 and $PM_{2.5}$ were estimated by considering the wind sector ((0–70°) & (290–360°)) where railway station emissions have significant effect on NO_2 and $PM_{2.5}$ levels (measured at Heuston roof), after subtracting the background concentration (NO_2 - Northern wall of Heuston roof & $PM_{2.5}$ - Cabra FSM). However, the railway station emission contribution for NO_2 was not estimated explicitly as the upwind background NO_2 concentrations were reflective of the averaged value (measured using diffusion tube) over the monitoring period. At the St. Johns FSM, the railway station emissions contribution was estimated, after subtracting the previous estimated traffic contribution from the average $PM_{2.5}$ and NO_2 measured at St. Johns FSM. For estimating the railway station emissions, hours of the day when trains were scheduled to run (6:00–23:00) were only considered in the filtration analysis. Supplementary Fig. S1 shows the filtrating technique adopted at the site.

Table 1
Filtering technique employed at the sites.

Site	Sources to be estimated	Sources wind sector	Background site	Sources contribution
Heuston station roof	Road traffic	110–250°	St. Johns FSM	Road traffic = average pollutant concentration at Heuston roof - average pollutant concentration at background site
	Railway station emission	0–70° & 290–360°	NO_2 - Northern wall of Heuston roof & $PM_{2.5}$ - Cabra FSM)	Railway station emission = average pollutant concentration at Heuston roof - average pollutants concentration at background site
St John EPA station	Railway station emissions	0–70° & 290–360°	–	Railway station emissions = average pollutant concentration at St. Johns FSM - Road traffic

2.3.2. *Bivariate polar plot*

To interpret the dependency of pollutants on wind speed, the concentration of NO_2 and $PM_{2.5}$ measured at Heuston station roof and St. Johns FSM were associated with wind speed and direction using bivariate polar plots. Bivariate polar plots are a commonly employed method to discriminate between sources of pollutants emitted with significant levels of buoyancy (e.g., industrial stack emissions, ships and aircraft plumes) (Carslaw et al., 2006). This method can also be employed to understand the complex relationships between wind speed and pollutants concentration. For instance, flow in an urban street-canyon where wind-blown particle re-suspension is important and where concentration of particle increases with wind speed (Carslaw et al., 2006).

3. Results & discussion

3.1. *Performance analysis of stationary co-location of LCM_A with St. Johns FSM*

The raw data was measured using LCM_A in reference to the St. Johns FSM. The average $PM_{2.5}$ and NO_2 concentrations captured by LCM_A during the sampling period were found to underestimate the average concentrations measured by the reference monitor by 33% and 59%, respectively. In addition, LCM_A presented a significantly lower StDev for NO_2 of 2.8 μg/m^3 compared to the St. Johns FSM with a StDev of 16.3 μg/m^3. This suggests that the NO_2 sensor was less effective at capturing both high and very low NO_2 concentrations. Also, the LCM_A showed a high RMSE of 19.6 μg/m^3 and no discernible linear correlation (R^2 = 0.009) for NO_2 in relation to reference monitor. These discrepancies commonly occur in NO_2 measurements using electrochemical sensors because NO_2 is primarily a chemically reactive pollutant, and the accuracy of the electrochemical sensors for NO_2 measurements is influenced by various factors such as interference, pollutant sources, cross-sensitivity, and changing environmental conditions (Nowack et al., 2021). Conversely, StDev the value for the $PM_{2.5}$ sensor (6.1 μg/m^3) was similar to the St. Johns FSM (5.5 μg/m^3), indicating similar performance in terms of capturing extreme concentration. The error and linear correlation between the sensor and the reference monitor for $PM_{2.5}$ was found to be 3.7 μg/m^3 and 0.76, which indicated that the $PM_{2.5}$ sensor performance was in good agreement with that of the St. Johns FSM monitor.

3.2. *Performances of the calibration algorithm*

The statistical performance (R^2 and RMSE) of the calibration models developed for LCM_A for $PM_{2.5}$ and NO_2 is summarised in Supplementary Fig. S2. After applying the calibration model, sensor measurements demonstrated a good agreement with the St. Johns FSM. Primarily both algorithms provided good calibration results for both NO_2 and $PM_{2.5}$, while the performance of the XGBoost model was marginally better than the RF model. Also, the performance of the calibration models for $PM_{2.5}$ was better than NO_2. The R^2 and RMSE for $PM_{2.5}$ calibration model ranged between 0.9-0.92 and 2.2–2.3 μg/m^3, respectively, which indicated good estimation of PM concentration using the models. The PM calibration results tended to be better than gaseous pollutants, owing to a good linear relationship between the raw data from the LCM_A and the St. Johns FSM values (Nowack et al., 2021; Wang et al., 2023).

The R^2 and RMSE for the NO_2 calibration model ranged between 0.82-0.83 and 9.1–9.3 μg/m^3, respectively, which indicated a reasonable level of reconstruction accuracy for a gaseous pollutant like NO_2. Based on the performance of the model's, XGBoost was selected to calibrate both $PM_{2.5}$ and NO_2. The time series of the modelled, and measured sensor and reference NO_2 and $PM_{2.5}$ concentration are shown in Supplementary Fig. S3.

3.3. Impact of variables on model output

We investigated the impact of different variables on best performing XGBoost model for $PM_{2.5}$ and NO_2. The importance of each model output variable was represented using a SHAP summary plot, which consists of magnitude contribution of all points of training dataset of each variable, expressed by their SHAP value (Wang et al., 2023). The explanatory variable with positive SHAP values has positive prediction on the model and vice-versa, and higher absolute SHAP value of variables denotes greater effect on the output. Variables are ranked high to low based on the importance, with absolute SHAP values of all datapoints from high to low. Fig. 2 presents the violin SHAP summary plot for NO_2. In Fig. 2, wind speed had the highest variable importance on the NO_2 model with SHAP values increases at lower wind speed. A strong relation of positive SHAP values with lower wind speed clearly indicated that NO_2 was strongly influenced by local sources. Other than wind speed, $PM_{2.5}$, pressure, train and road traffic, NO_2 raw sensor readings, and days of the week were positively correlated with the reference NO_2 values. The wind sectors corresponding to negative sine functions (100–274°) had a significant impact on the positive SHAP values, indicating potential dominance of NO_2 sources in those directions. The effect of relative humidity, hours of the day, noise and temperature were not significant on the NO_2 model.

The $PM_{2.5}$ model SHAP values presented in Supplementary Fig. S4 clearly show that the relationship between raw sensor readings and model output is linear. Besides sensor reading, pressure was the second most important variable that was positively correlated with reference $PM_{2.5}$ concentrations. The positive relation with humidity indicates the presence of aerosol hygroscopicity in the atmosphere. The association between pressure and model outcome was recognised as non-linear, consistent with previous literature (Wang et al., 2023). High air pressure was conducive in the accumulation of $PM_{2.5}$ to form polluted weather, thus the concentration of $PM_{2.5}$ increased. However, during periods with high PM concentrations, an inverse association was observed. It was also observed that air pressure was a good predictor rather than a good explainer for sensor readings. All other variables did not have significant impact on the $PM_{2.5}$ model.

It's important to note that the results from low cost air quality sensors are not transferable. While these sensors are affordable, they require a significant amount of additional input data to produce meaningful results. Additionally, the calibration model developed for one site cannot be generalized to other sites due to varying input characteristics. Furthermore, access to a reference sensor is required for calibration. This means one cannot use these sensors effectively without having access to a high-end instrument.

3.4. Concentrations profiles at different study sites

Table 2 presents the summary of the NO_2 and $PM_{2.5}$ concentration

Table 2
Summary of NO_2 and $PM_{2.5}$ concentrations measured at the study site.

Site	Average concentration		No. Of days exceeding 24-hr WHO limit	
	$PM_{2.5}$ (μg/m^3)	NO_2 (μg/m^3)	$PM_{2.5}$	NO_2a[a]
Heuston roof	8.9	51.8	2	30
EPA station	7.8	38.1	4	25
Upwind sites of Heuston station	5	31.8	–	–

[a] **Hourly and daily averaged NO_2 values (obtained using calibrated sensor) are adjusted using diffusion tubes data.**

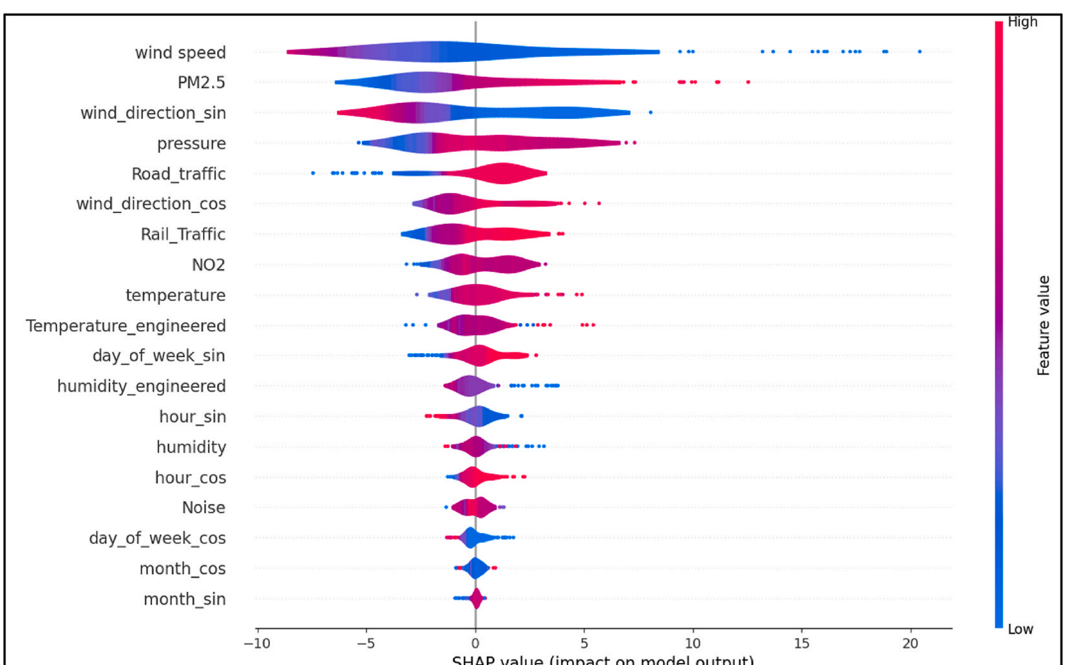

Fig. 2. SHAP summary plot for NO_2 concentrations.

measured at the study sites. The concentration of $PM_{2.5}$ and NO_2 over the study period at St. Johns FSM were found to be 7.8 and 38 μg/m^3, respectively. While the concentration of $PM_{2.5}$ and NO_2 at Heuston roof were found to be 8.9 and 51.8 μg/m^3, respectively. The concentration of $PM_{2.5}$ and NO_2 at the upwind sites of railway station were found to be 5 and 31.8 μg/m^3, respectively. Measured $PM_{2.5}$ concentrations averaged over the study period were exceeding the annual WHO limits ($PM_{2.5}$ = 5 μg/m^3, NO_2 = 10 μg/m^3) by 1.6–1.8 and 3.2–5.2 times respectively at the study site. However, $PM_{2.5}$ and NO_2 concentrations were well below the EU limits of 10 and 40 μg/m^3, respectively (except NO_2 level at Heuston roof). The diurnal variation of pollutants (Supplementary Fig. S5) at the study sites showed that elevated $PM_{2.5}$ concentration was observed during the early hours of the day (01:00–05:00) and during morning (07:00–9:00). The average $PM_{2.5}$ concentrations during these hours ranged between 5.2 and 9.9 μg/m^3, respectively.

The prevalence of low wind speed and low temperature at the site potentially led to PM accumulating and stagnating during these hours. The heterogeneous hydrolysis of NO_2 to nitrate is typically dominant during early hours of the day, and this may also add to elevated PM concentrations (Shanmuga et al., 2021). The diurnal variation of NO_2 showed a morning and evening peaks reflecting the schedule of traffic activities. The average NO_2 concentrations during these peak hours ranged between 30.3 and 50.7 μg/m^3, respectively. The weekday average concentrations ($PM_{2.5}$ = 7.8–9.3 μg/m^3 & NO_2 = 38.1–52.6 μg/m^3) were slightly higher than the weekend average concentrations ($PM_{2.5}$ = 6.8–8.1 μg/m^3 & NO_2 = 35.6–49.3 μg/m^3) at the study sites.

3.5. *Disaggregation of emission sources at St. Johns FSM*

3.5.1. *Contribution from road traffic*

Estimates of the road traffic contribution to NO_2 and $PM_{2.5}$ were made by considering the wind sector from which road traffic was likely to considerably effect on concentrations. Removing background concentrations aided in disaggregation of local sources. The road traffic contribution was estimated by computing the mean concentration of measurements in the considered wind sector range (110–250°) and the measurement frequency was estimated as a proportion of total number of hourly measurements in that sector (Table 1). This estimated value represented an upper limit of traffic contributions because there were no major sources of NO_2 and $PM_{2.5}$ between the background site (St. Johns FSM) and the analysed site (Heuston roof). An upper limit of traffic emission contributions to $PM_{2.5}$ and NO_2 from St. John's Road was estimated to be 1.7 and 18.7 μg/m^3, respectively. The proportion of measurements (both $PM_{2.5}$ & NO_2) in this wind sector was estimated to be 46%. The mean $PM_{2.5}/NO_2$ ratio was estimated to be 0.09 (on a mass basis), which was consistent with the $PM_{2.5}/NO_2$ ratio (0.05–0.09) measured from street level road traffic emissions in European cities of Gothenburg and Umea, interpreting the source of road traffic (Ferm & Sjöberg, 2015).

3.5.2. *Railway station contributions*

In a similar way, the measurements were filtered based on wind sectors (0–70° & 290–360°) where railway station emissions were likely to have significant effect (Heuston roof), while also considering the hours of the day when trains are scheduled to run. The railway station emissions contribution to $PM_{2.5}$ and NO_2 at Heuston station roof was estimated to be 6.3 and 36.7 μg/m^3, respectively after removing the upwind pollutant concentration. However, NO_2 emissions from the railway station was not explicitly estimated, as upwind NO_2 concentration reflected the averaged value over the entire monitoring period (irrespective of wind sector). The proportion of measurements (both $PM_{2.5}$ & NO_2) in this wind sector was 12%.

The estimated mean concentration of $PM_{2.5}$ at St. Johns FSM (9.3 μg/m^3) in the wind sector (0–70° & 290–360°) nearly aligns with the estimated mean concentration at the Heuston roof (11.8 μg/m^3) within the same sector. This implies that the $PM_{2.5}$ emissions from the railway station are effectually transported to the St. Johns FSM in this specific wind sector with slight dispersion. Therefore, the railway station emissions contribution to $PM_{2.5}$ estimated at Heuston roof (6.3 μg/m^3) is projected to be the same at St. Johns FSM. However, a similar trend was not observed for NO_2, making the results inconclusive for NO_2 dispersion.

The contribution of railway station emissions to NO_2 at the St. Johns FSM was estimated to be approximately 20.1 μg/m^3, after subtracting the previously determined road traffic contribution from the mean NO_2 St. Johns FSM concentration in this wind sector (Supplementary Table S4). It should be noted that the estimates include the upwind concentration of the railway station.

3.6. *Effect of wind speed on pollutants source*

The bivariate polar plots for pollutants at the study sites considering the wind sectors for which road and railway station emission are likely to have considerable effect on measured concentrations are shown in Fig. 3. At Heuston Station roof (considering wind directions in the sector 110–250°), the highest concentrations of NO_2 are observed with calm wind speed (0–0.5 m/s) and concentrations decrease with increase in wind speed (Fig. 3). The average concentration of NO_2 (42.1 μg/m^3) lies predominantly in the lower wind speed range (1–1.5 m/s) which could possibly implicate local source (i.e., road traffic in this case) as a dominant source. In general, low wind speed in bivariate polar plot corresponds to local sources of pollutants and high wind speed helps in identifying long distant transport of pollutants (Uria-Tellaetxe & Carslaw, 2014; Xing et al., 2021). A similar pattern was observed in the NO_2 concentrations measured at St. Johns FSM (considering concentrations in the sector 0–70° & 290–360°), however a slight spike in concentrations was observed at higher wind speed bins, likely owing to transport of pollutants from distant sources. The average concentration of NO_2 (38.8 μg/m^3) was distributed over both high (>2 m/s) and low wind speed ranges (1–1.5 m/s), reflecting dominant contributions from both local sources (i.e., railway station and on-road emissions in this case) and distant background sources. The polar plots clearly support that the local traffic is one of the dominant sources of NO_2 at both St. Johns FSM and Heuston Station roof.

Concerning $PM_{2.5}$, concentrations at St. Johns FSM increases with calm wind speed condition (0–1 m/s) and decreases with increase in wind speed (1.5–2.5 m/s) (Supplementary Fig. S6). The spike in concentration was observed at higher wind speed range (>2.5 m/s), where particles are transported to ground level through buoyancy effects (likely from the distant sources). The average concentration of $PM_{2.5}$ (9.1 μg/m^3) was noted in both low and high-speed ranges, indicating emission from local sources (on-road and railway station emissions)) and distant sources. While average concentration of $PM_{2.5}$ (11.1 μg/m^3) at Heuston roof dominantly lies in the higher wind speed (>1.5 m/s) range indicating the upwind background as a dominant source.

4. Conclusions

This study demonstrated the potential of using low-cost environmental monitors as part of a 4-week monitoring campaign to examine the impact of railway hub (Heuston railway station) contributions to local air pollution in Dublin city. The study used machine learning models (RF & XGBoost) to calibrate the $PM_{2.5}$ and NO_2 measurements acquired by the low-cost sensors. A data filtration technique (wind direction corresponding to sources) was employed to disaggregate road traffic and railway station emission contributions to $PM_{2.5}$ and NO_2 concentrations at the study site (St. Johns FSM). In addition, the results from the bivariate polar plots helped to understand the origin of pollution sources.

The performance of the XGBoost model was slightly better than the RF model for both $PM_{2.5}$ and NO_2. Local wind speeds presented the variable with the greatest impact on NO_2 model results, followed by

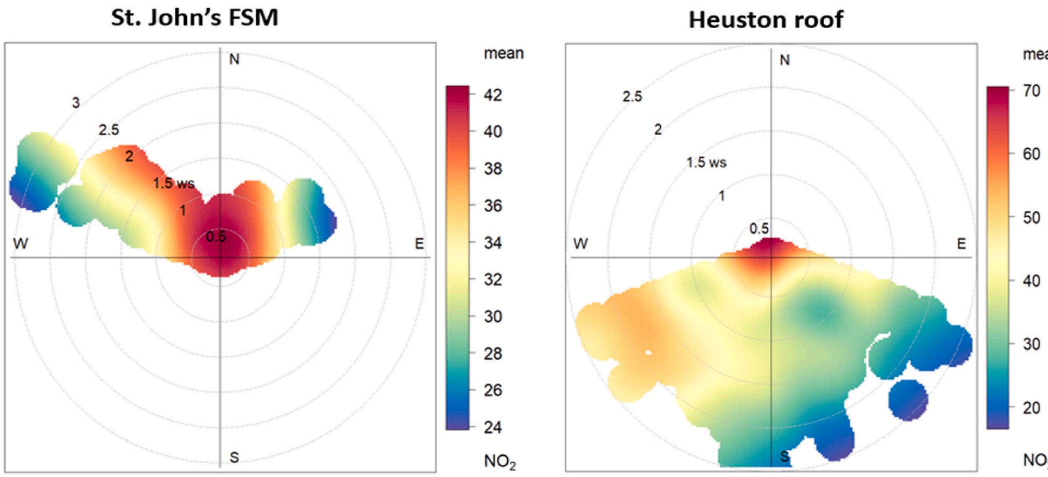

Fig. 3. Bivariate polar plots for NO_2 at the study site.

equivalent $PM_{2.5}$ concentrations, wind direction and pressure. The strong relation of lower wind speeds was observed for the NO_2 model which clearly indicated that NO_2 was strongly influenced by the local sources. Linear relationship between the raw sensor readings and the $PM_{2.5}$ model was identified, indicating that sensor readings have a significant effect on the model output. These results highlight that the NO_2 sensor requires more input data for accurate calibration compared to the $PM_{2.5}$ sensor.

Source apportionment results showed that railway station emissions as the main $PM_{2.5}$ source, and road traffic as the main NO_2 source, when winds are from station wind sector. Bivariate polar plot also revealed the dominance of local sources in this sector for $PM_{2.5}$ and NO_2. However, only 12% of the winds were from the direction of the station during the study period, which may not reflect the annual average conditions. The majority of measured wind conditions were upwind of the St. Johns FSM (46%), with upwind background NO_2 and $PM_{2.5}$ being the major contributors to the St. Johns FSM. Implementing control strategies such as low-emission zones and reducing traffic congestion would improve air quality in this control area. This study outlines the importance and value of LCM devices and complementary modelling techniques to improve our understanding of spatio-temporal conditions around the transport hub.

CRediT authorship contribution statement

Shanmuga Priyan: Conceptualization, Formal Analysis, Writing – original draft, Visualization, Methodology. **Yuxuan Guo:** Conceptualization, Writing – review & editing. **Aonghus McNabola:** Writing – review & editing. **Brian Broderick:** Writing – review & editing. **Brian Caulfield:** Writing – review & editing. **Margaret O'Mahony:** Writing – review & editing. **John Gallagher:** Conceptualization, Writing – review & editing, Resources, Supervision.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Shanmuga Priyan reports financial support was provided by Environmental Protection Agency. Yuxuan Guo reports financial support was provided by Environmental Protection Agency. If there are other authors, they 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

Data will be made available on request.

Acknowledgements

This STATION-AIR project is funded under the EPA Research Programme 2021–2030 and co-funded by the Department of Transport (2021-HE-1041). The EPA Research Programme is a Government of Ireland initiative funded by the Department of the Environment, Climate, and Communications.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envpol.2024.124903.

References

Adong, P., Bainomugisha, E., Okure, D., Sserunjogi, R., 2022. Applying machine learning for large scale field calibration of low-cost PM 2.5 and PM 10 air pollution sensors. Applied AI Letters 3 (3), 1–16. https://doi.org/10.1002/ail2.76.

Borken-kleefeld, J., Ntziachristos, L., 2012. The potential for further controls of emissions from mobile sources in Europe. URL. https://pure.iiasa.ac.at/id/eprint/10161/1/XO-12-014.pdf. (Accessed 30 August 2024).

Carslaw, D.C., Beevers, S.D., Ropkins, K., Bell, M.C., 2006. Detecting and quantifying aircraft and other on-airport contributions to ambient nitrogen oxides in the vicinity of a large international airport. Atmos. Environ. 40 (28), 5424–5434. https://doi.org/10.1016/j.atmosenv.2006.04.062.

Cheng, Y.H., Chang, H.P., Hsieh, C.J., 2011. Short-term exposure to PM10, PM2.5, ultrafine particles and CO2 for passengers at an intercity bus terminal. Atmos. Environ. 45 (12), 2034–2042. https://doi.org/10.1016/j.atmosenv.2011.01.061.

Chong, U., Swanson, J.J., Boies, A.M., 2015. Air quality evaluation of London Paddington railway station. Environ. Res. Lett. 10 (9) https://doi.org/10.1088/1748-9326/10/9/094012.

DEFRA, 2011. Local bias adjustment factors. London. URL. https://laqm.defra.gov.uk/air-quality/air-quality-assessment/local-bias/. (Accessed 30 August 2024).

de Foy, B., Heo, J., Kang, J.Y., Kim, H., Schauer, J.J., 2021. Source attribution of air pollution using a generalized additive model and particle trajectory clusters. Sci. Total Environ. 780, 146458 https://doi.org/10.1016/j.scitotenv.2021.146458.

Dublin City Council, 2023. Transport infrastructure Ireland. URL. https://data.cso.ie/table/TII04. (Accessed 30 August 2024).

European Commission, 2023. Electrification of rail infrastructure. URL. https://alternative-fuels-observatory.ec.europa.eu/transport-mode/rail. (Accessed 30 August 2024).

Ferm, M., Sjöberg, K., 2015. Concentrations and emission factors for PM2.5 and PM10 from road traffic in Sweden. Atmos. Environ. 119, 211–219. https://doi.org/10.1016/j.atmosenv.2015.08.037.

Font, A., Tremper, A.H., Lin, C., Priestman, M., Marsh, D., Woods, M., Heal, M.R., Green, D.C., 2020. Air quality in enclosed railway stations: quantifying the impact of diesel trains through deployment of multi-site measurement and random forest modelling. Environ. Pollut. 262, 114284 https://doi.org/10.1016/j.envpol.2020.114284.

González, Y., Rodríguez, S., Guerra García, J.C., Trujillo, J.L., García, R., 2011. Ultrafine particles pollution in urban coastal air due to ship emissions. Atmos. Environ. 45 (28), 4907–4914. https://doi.org/10.1016/j.atmosenv.2011.06.002.

Grana, M., Toschi, N., Vicentini, L., Pietroiusti, A., Magrini, A., 2017. Exposure to ultrafine particles in different transport modes in the city of Rome. Environ. Pollut. 228, 201–210. https://doi.org/10.1016/j.envpol.2017.05.032.

Hagler, G., Birkett, D., Henry, R.C., Peltier, R.E., 2021. Three years of high time-resolution air pollution monitoring in the complex multi-source harbor of New York and New Jersey. Aerosol Air Qual. Res. 21 (3), 1–15. https://doi.org/10.4209/aaqr.2020.02.0069.

Hakkim, H., Kumar, A., Annadate, S., Sinha, B., Sinha, V., 2021. RTEII: a new high-resolution (0.1° × 0.1°) road transport emission inventory for India of 74 speciated NMVOCs, CO, NOx, NH3, CH4, CO2, PM2.5 reveals massive overestimation of NOx and CO and missing nitromethane emissions by existing inventories. Atmos. Environ. X 11 (February). https://doi.org/10.1016/j.aeaoa.2021.100118.

Han, B., Wang, L., Deng, Z., Shi, Y., Yu, J., 2022. Source emission and attribution of a large airport in Central China. Sci. Total Environ. 829, 154519 https://doi.org/10.1016/j.scitotenv.2022.154519.

Hickman, A.L., Baker, C.J., Cai, X., Delgado-Saborit, J.M., Thornes, J.E., 2018. Evaluation of air quality at the Birmingham new street railway station. Proc. Inst. Mech. Eng. F J. Rail Rapid Transit 232 (6), 1864–1878. https://doi.org/10.1177/0954409717752180.

Kim, K.Y., Kim, Y.S., Roh, Y.M., Lee, C.M., Kim, C.N., 2008. Spatial distribution of particulate matter (PM10 and PM2.5) in Seoul Metropolitan Subway stations. J. Hazard Mater. 154 (1–3), 440–443. https://doi.org/10.1016/j.jhazmat.2007.10.042.

Kim, M.K., Park, D., Kim, M., Heo, J., Park, S., Chong, H., 2020. A study on characteristic emission factors of exhaust gas from diesel locomotives. Int. J. Environ. Res. Publ. Health 17 (11), 1–12. https://doi.org/10.3390/ijerph17113788.

Ledoux, F., Roche, C., Cazier, F., Beaugard, C., Courcot, D., 2018. Influence of ship emissions on NOx, SO2, O3 and PM concentrations in a North-Sea harbor in France. J. Environ. Sci. (China) 71 (x), 56–66. https://doi.org/10.1016/j.jes.2018.03.030.

Liu, B., Tan, X., Jin, Y., Yu, W., Li, C., 2021. Application of RR-XGBoost combined model in data calibration of micro air quality detector. Sci. Rep. 11 (1), 1–14. https://doi.org/10.1038/s41598-021-95027-1.

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., Bezirtzoglou, E., 2020. Environmental and {health} {impacts} of {air} {pollution}: {A} {review. Front. Public Health 8 (February), 1–13. https://doi.org/10.3389/fpubh.2020.00014.

Martins, V., Moreno, T., Minguillón, M.C., Van Drooge, B.L., Reche, C., Amato, F., De Miguel, E., Capdevila, M., Centelles, S., Querol, X., 2016. Origin of inorganic and organic components of PM2.5 in subway stations of Barcelona, Spain. Environ. Pollut. 208, 125–136. https://doi.org/10.1016/j.envpol.2015.07.004.

Merico, E., Donateo, A., Gambaro, A., Cesari, D., Gregoris, E., Barbaro, E., Dinoi, A., Giovanelli, G., Masieri, S., Contini, D., 2016. Influence of in-port ships emissions to gaseous atmospheric pollutants and to particulate matter of different sizes in a Mediterranean harbour in Italy. Atmos. Environ. 139, 1–10. https://doi.org/10.1016/j.atmosenv.2016.05.024.

Mueller, M., Meyer, J., Hueglin, C., 2017. Design of an ozone and nitrogen dioxide sensor unit and its long-Term operation within a sensor network in the city of Zurich. Atmos. Meas. Tech. 10 (10), 3783–3799. https://doi.org/10.5194/amt-10-3783-2017.

Nowack, P., Konstantinovskiy, L., Gardiner, H., Cant, J., 2021. Machine learning calibration of low-cost NO2 and PM10 sensors: non-linear algorithms and their impact on site transferability. Atmos. Meas. Tech. 14 (8), 5637–5655. https://doi.org/10.5194/amt-14-5637-2021.

Ozigis, I.I., Oche, J.I.O., Lawal, N.M., 2021. Locomotive engines and the future of railway automotive power in Africa: a review. Nigerian Journal of Technology 40 (4), 660–673. https://doi.org/10.4314/njt.v40i4.13.

Pan, Q., Harrou, F., Sun, Y., 2023. A comparison of machine learning methods for ozone pollution prediction. Journal of Big Data 10 (1). https://doi.org/10.1186/s40537-023-00748-x.

Passi, A., Nagendra, S.M.S., Maiya, M.P., 2021. Characteristics of indoor air quality in underground metro stations : a critical review. Build. Environ. 198 (February), 107907 https://doi.org/10.1016/j.buildenv.2021.107907.

Phairuang, W., Chetiyanukornkul, T., Suriyawong, P., Amin, M., Hata, M., Furuuchi, M., Yamazaki, M., Gotoh, N., Furusho, H., Yurtsever, A., Watanabe, S., Sun, L., 2024. Characterizing chemical, environmental, and stimulated subcellular physical characteristics of size-fractionated PMs down to PM0.1. Environ. Sci. Technol. 58 (28), 12368–12378. https://doi.org/10.1021/acs.est.4c01604.

Phairuang, W., Suwattiga, P., Hongtieab, S., Inerb, M., Furuuchi, M., Hata, M., 2021. Characteristics, sources, and health risks of ambient nanoparticles (PM0.1) bound metal in Bangkok, Thailand. Atmos. Environ. X 12, 100141. https://doi.org/10.1016/j.aeaoa.2021.100141.

Railway Technology, 2016. The big stink: how much do trains really emit? URL. https://www.railway-technology.com/features/featurethe-big-stink-how-much-do-trains-really-emit-4807131/?cf-view. (Accessed 30 August 2024).

Saucy, A., de Hoogh, K., Vienneau, D., Tangermann, L., Schäffer, B., Wunderli, J.M., Probst-Hensch, N., Röösli, M., 2021. Mutual effects of fine particulate matter, nitrogen dioxide, and fireworks on cause-specific acute cardiovascular mortality: {A} case-crossover study in communities affected by aircraft noise. Environ. Pollut. 291 (April) https://doi.org/10.1016/j.envpol.2021.118066.

Schürmann, G., Schäfer, K., Jahn, C., Hoffmann, H., Bauerfeind, M., Fleuti, E., Rappenglück, B., 2007. The impact of NOx, CO and VOC emissions on the air quality of Zurich airport. Atmos. Environ. 41 (1), 103–118. https://doi.org/10.1016/j.atmosenv.2006.07.030.

Shanmuga, R., Elizbath, A., Jyothi, P., Mohan, S., Shiva, N., 2021. Vertical distribution of PM 10 and PM 2 . 5 emission sources and chemical composition during winter period in Delhi city. Air Qual. Atmos. Health. https://doi.org/10.1007/s11869-021-01092-w.

Uria-Tellaetxe, I., Carslaw, D.C., 2014. Conditional bivariate probability function for source identification. Environ. Model. Software 59, 1–9. https://doi.org/10.1016/j.envsoft.2014.05.002.

UTRAP, 2021. Urban transport related air pollution. URL. https://www.gov.ie/en/publication/3f634-urban-transport-related-air-pollution-utrap-working-group/. (Accessed 30 August 2024).

Vilcassim, M.J.R., Thurston, G.D., Peltier, R.E., Gordon, T., 2014. Black carbon and particulate matter (PM2.5) concentrations in New York city's subway stations. Environ. Sci. Technol. 48 (24), 14738–14745. https://doi.org/10.1021/es504295h.

Wang, A., Machida, Y., deSouza, P., Mora, S., Duhl, T., Hudda, N., Durant, J.L., Duarte, F., Ratti, C., 2023. Leveraging machine learning algorithms to advance low-cost air sensor calibration in stationary and mobile settings. Atmos. Environ. 301 (February) https://doi.org/10.1016/j.atmosenv.2023.119692.

Xing, C., Liu, C., Hu, Q., Fu, Q., Wang, S., Lin, H., Zhu, Y., Wang, S., Wang, W., Javed, Z., Ji, X., Liu, J., 2021. Vertical distributions of wintertime atmospheric nitrogenous compounds and the corresponding OH radicals production in Leshan, southwest China. J. Environ. Sci. (China) 105 (July), 44–55. https://doi.org/10.1016/j.jes.2020.11.019.