Environment International 184 (2024) 108449

Contents lists available at ScienceDirect

Environment International

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

Full length article

Constructing transferable and interpretable machine learning models for black carbon concentrations

Pak Lun Fung [a,b,*], Marjan Savadkoohi [c,d,*], Martha Arbayani Zaidan [a,b,e], Jarkko V. Niemi [f], Hilkka Timonen [g], Marco Pandolfi [c], Andrés Alastuey [c], Xavier Querol [c], Tareq Hussein [a,h], Tuukka Petäjä [a]

[a] Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Helsinki FI-00560, Finland
[b] Helsinki Institute of Sustainability Science, Faculty of Science, University of Helsinki, Helsinki FI-00560, Finland
[c] Institute of Environmental Assessment and Water Research (IDAEA-CSIC), Barcelona, Spain
[d] Department of Mining, Industrial and ICT Engineering (EMIT), Manresa School of Engineering (EPSEM), Universitat Politècnica de Catalunya (UPC), Manresa 08242, Spain
[e] Department of Computer Science, Faculty of Science, University of Helsinki, Helsinki FI-00560, Finland
[f] Helsinki Region Environmental Services Authority (HSY), Helsinki FI-00066, Finland
[g] Atmospheric Composition Research, Finnish Meteorological Institute, Helsinki FI-00560, Finland
[h] Environmental and Atmospheric Research Laboratory (EARL), Department of Physics, School of Science, Amman 11942, Jordan

ARTICLE INFO

Keywords:
BC estimation
Virtual sensors
Relative importance
Neural network
SHAP
Traffic emission

ABSTRACT

Black carbon (BC) has received increasing attention from researchers due to its adverse health effects. However, in-situ BC measurements are often not included as a regulated variable in air quality monitoring networks. Machine learning (ML) models have been studied extensively to serve as virtual sensors to complement the reference instruments. This study evaluates and compares three white-box (WB) and four black-box (BB) ML models to estimate BC concentrations, with the focus to show their transferability and interpretability. We train the models with the long-term air pollutant and weather measurements in Barcelona urban background site, and test them in other European urban and traffic sites. Despite the difference in geographical locations and measurement sites, BC correlates the strongest with particle number concentration of accumulation mode (PN_{acc}, r = 0.73–0.85) and nitrogen dioxide (NO_2, r = 0.68–0.85) and the weakest with meteorological parameters. Due to its similarity of correlation behaviour, the ML models trained in Barcelona performs prominently at the traffic site in Helsinki (R^2 = 0.80–0.86; mean absolute error MAE = 3.90–4.73 %) and at the urban background site in Dresden (R^2 = 0.79–0.84; MAE = 4.23–4.82 %). WB models appear to explain less variability of BC than BB models, long short-term memory (LSTM) model of which outperforms the rest of the models. In terms of interpretability, we adopt several methods for individual model to quantify and normalize the relative importance of each input feature. The overall static relative importance commonly used for WB models demonstrate varying results from the dynamic values utilized to show local contribution used for BB models. PN_{acc} and NO_2 on average have the strongest absolute static contribution; however, they simultaneously impact the estimation positively and negatively at different sites. This comprehensive analysis demonstrates that the possibility of these interpretable air pollutant ML models to be transfered across space and time.

1. Introduction

Black carbon (BC) consists mostly of agglomerated sub-micron particulate matter (PM) generated primarily from incomplete combustion of fossil fuels, biomass, and other organic materials (Bond et al., 2013). In urban areas, BC often originates from residential burning

* Corresponding authors at: Institute for Atmospheric and Earth System Research / Physics, Faculty of Science, University of Helsinki, Helsinki FI-00560, Finland (P.L. Fung); Institute of Environmental Assessment and Water Researchh (IDAEA-CSIC), Barcelona, Spain (M. Savadkoohi).
E-mail addresses: pak.fung@helsinki.fi (P.L. Fung), marjan.savadkoohi@idaea.csic.es (M. Savadkoohi), martha.zaidan@helsinki.fi (M.A. Zaidan), jarkko.niemi@hsy.fi (J.V. Niemi), hilkka.timonen@fmi.fi (H. Timonen), marco.pandolfi@idaea.csic.es (M. Pandolfi), andres.alastuey@idaea.csic.es (A. Alastuey), xavier.querol@idaea.csic.es (X. Querol), tareq.hussein@helsinki.fi (T. Hussein), tuukka.petaja@helsinki.fi (T. Petäjä).

https://doi.org/10.1016/j.envint.2024.108449
Received 8 November 2023; Received in revised form 12 January 2024; Accepted 17 January 2024
Available online 22 January 2024

(particle size of ~ 300 nm) and vehicular combustion (~100–150 nm, Saarikoski et al., 2021). BC has been suggested to have strong link with adverse health effects, including respiratory and cardiovascular diseases, which have led to increasing attention from researchers and policymakers for formulating effective air quality (AQ) management strategies and understanding climate change dynamics (Briggs and Long, 2016). World Health Organization (WHO, 2021) has recommended regional authorities to make systematic measurements and subsequent reporting of BC. Besides, including BC as an additional component in comprehensive AQ index would allow analysis of the interconnected impacts of other gaseous and aerosol pollutants to AQ and the improved visualization of the overall AQ status to the general public. Following the trends of such comprehensive AQ index could be used by the authorities to implement policies that improve AQ and follow-ups with the observed impacts to the AQ index (e.g. Fung et al., 2022a).

To assess the concentrations of BC, various measurement techniques have been employed, including e.g. filter absorption photometers (FAPs) such as Aethalometers (AEs) and Multi-Angle Absorption Photometers (MAAPs) (Petzold and Schönlinner, 2004; Moosmüller et al., 2009; Petzold et al., 2013). These instruments measure the light absorption properties of particulate matter collected on filter substrate, allowing for the estimation of BC mass concentrations. However, the accuracy of these measurements depends on the specific method used. As a result, efforts have been made to improve these measurements and enhance their reliability through intercomparison of absorption photometers (Hitzenberger et al., 2006; Müller et al., 2011; Cuesta-Mosquera et al., 2021).

However, uncertainties persist due to the lack of agreement on methods and terminology (Petzold et al., 2013), as well as technical issues related to instrument software and correction algorithms (e.g. Weingartner et al., 2003; Collaud Coen et al., 2010; Luoma et al., 2021b). Applicability of the data is hindered by missing data e.g., due to instrument malfunction (Zaidan et al., 2019). Numerous studies have investigated the variability and trends of BC concentrations at local and global scales (e.g., Ahmed et al., 2014; Grange et al., 2020; Jafar and Harrison, 2021; Sun et al., 2021; Savadkoohi et al., 2023). Yet, due to the lack of measurement sites, existing research outcomes may not fully capture the whole story, particularly the spatial distribution and evolving trends over time, which are influenced by regional and global mitigation policies, changes in emission patterns, and meteorological factors (Collaud Coen et al., 2020).

To address these limitations, machine learning (ML) methods have emerged as virtual sensors for estimating BC concentrations (Zaidan et al., 2020). Various studies have explored the potential of ML methods in predicting air quality indicators and identifying emission sources, with a specific focus on BC emissions (e.g., Abu Awad et al., 2017; Fung et al., 2021b; Zhu et al., 2021; Rovira et al., 2022; Rubio-Loyola and Paul-Fils, 2022; Makkhan et al., 2023; Liu et al., 2023; Luo et al., 2023). ML algorithms can handle large datasets, identify complex patterns, and create predictive models with high accuracy, offering a cost-effective and efficient approach for real-time monitoring. The utilization of ML techniques in AQ research has improved due to the increasing availability of high-resolution monitoring data, advancements in computational resources, and the demand for accurate emission source characterization (e.g., Wang et al., 2020; Patil et al., 2022; Qiu et al., 2022; Méndez et al., 2023). Despite the complex interactions between emission sources, meteorological factors and aerosol properties, regression-based ML methods have been employed to estimate BC levels accurately and reliably incorporated with these components as input variables (e.g., Luo et al., 2018; Fung et al., 2021b; May and Li, 2022; Zhang et al., 2022).

Despite the accuracy and reliability in the abovementioned ML studies, ML models have been criticized, as common drawbacks for data-driven models, for the lack of accountability and generalizability. Data-driven models can be classified as white-box (WB) and black-box (BB) models where the classification of the two types of models is a continuum depending on their computational complexity and accountability (Zaidan et al., 2022). Generally speaking, WB models are transparent ML processes and often exist as a set of mathematical equations where the contribution of each input variable to the output is known (Rudin, 2019). One example is multiple linear regression (MLR, e.g., Zaidan et al., 2019; Liu et al., 2023). BB models, on the other hand, refer to systems which are viewed as deep learning processes through their inputs and outputs, without any knowledge of its internal workings or underlying principles (Rudin, 2019). The higher complexity of the model architecture usually results in a higher accuracy in model performance. These include Random Forest (RF, e.g., Qiu et al., 2022; Yu et al., 2023), Support Vector Machine (SVM, e.g., May and Li, 2022; Rovira et al., 2022), and Neural Networks (NNs, e.g., Bekkar et al., 2021; Duan et al., 2023). However, the generalizability of these data-driven models is highly subjected to the quality and representativeness of the training data. Previous studies (e.g., Ameer et al., 2019; Fung et al., 2021b) have argued that AQ models by this approach are site specific which fail to extend to a wider spatial context. Therefore, we strive to seek for transferable and interpretable ML models with high accuracy and, ideally consuming few computational resources. This would target at the core of the issue for BC models to work as virtual sensors to complement reference instruments in practice.

The aim of this study is to show the transferability and interpretability of selected data-driven models trained with the long-term BC measurements collected in an urban site in Barcelona and tested with four other external sites. We describe the measurement sites and instrumentation in Section 2.1 and 2.2 respectively. We further describe three WB and four BB models selected in this study and their corresponding relative importance metrics in Section 2.3, followed by trend analysis and evaluation metrics in Section 2.4 and 2.5, respectively. As part of the results, we first investigate the general BC seasonal, weekly and diurnal characteristics, and identify the key input parameters for the derivation of BC proxy in Barcelona in Section 3.1. In Section 3.2, we illustrate the performance of the various models in terms of accuracy and computational resources in different seasons and optimization combinations in Barcelona. The study further demonstrates the transferability of the models to other sites in Section 3.3. Furthermore, the study quantifies the relative importance of input variables using both WB and BB models. The effectiveness of the accountable proxies in estimating BC concentrations in other urban areas or regions will provide insights into the transferability and interpretability of the developed models.

2. Material and methods

A simplified workflow for the work is outlined in Fig. 1. This section first describes the observations used regarding the measurement sites and instrumentation in Section 2.1. The procedures of data preprocessing and the description of ML models used are elaborated in Section 2.2 and 2.3, respectively.

2.1. Observations

2.1.1. Measurement sites

The study focuses on data collected from five European monitoring sites covering different periods between 2009 and 2022, which were selected to ensure better spatial coverage and represent different climate zones and emission sources (Fig. S1). These sites comprise two urban background (UB) and three traffic (TR) sites located in Barcelona (BCN-UB), Helsinki (HEL-UB, HEL-TR), and Dresden (DDW-UB, DDN-TR), from three European countries (Spain, Finland, and Germany) where BCN-UB is the primary focus and serves as training data in ML modeling processes due to its longer term and more complete measurements. The other four sites are so-called external sites for testing the models. All these sites were selected from different geographic regions (South-

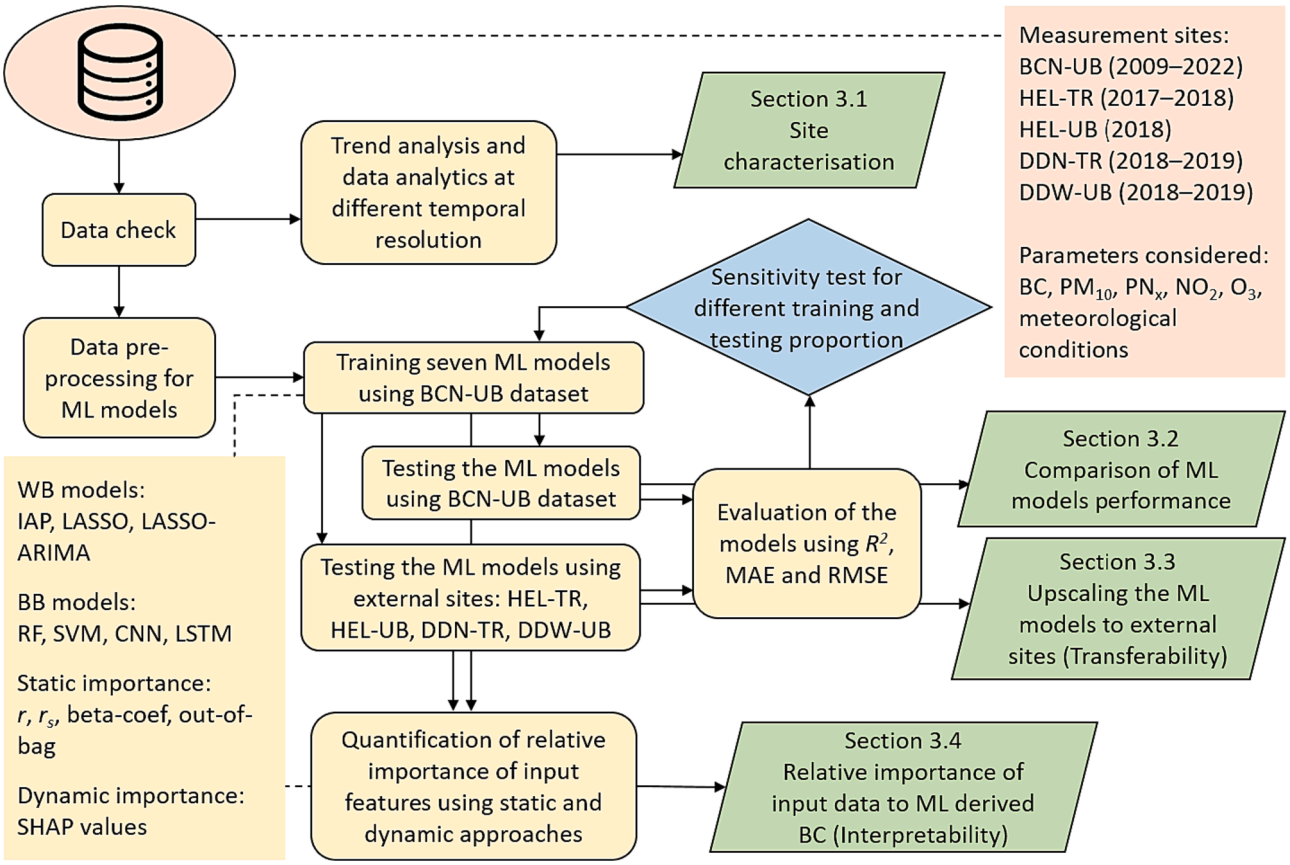

Fig. 1. Flow chart of the study.

western, Northern and Central Europe, respectively) and are characterized by aerosols with different physical and chemical properties to ensure better spatial coverage. Data collected include BC mass concentrations, particle number size distribution, meteorological data, gaseous pollutants, and PM concentrations.

Barcelona urban background site (BCN-UB, 41°23′24.01″ N 02°06′58.06″E, 64 m a.s.l.) is located in a coastal city within the metropolitan area with maritime Mediterranean climate (Köppen-Geiger classification: Csa). Although being an urban site, this site is located 200 m from one of the city's busiest roads with a daily traffic count of >60 k vehicles. The measurements are strongly influenced by traffic emissions, as indicated by the daily patterns of particle number concentrations (PN) and BC (Rivas et al., 2020; Yus-Díez et al., 2022).

Helsinki, a coastal city in the south of Finland, with humid continental climate (Köppen-Geiger classification: Dfb), has two measurement sites: the Mäkelänkatu street canyon traffic site (HEL-TR, 60°11′N, 24°57′E, 25 m a.s.l.), located 3 km from the city centre and 0.5 m from the nearest street lane edge, characterized by high traffic volumes (~28 k vehicles per workday) and limited dispersion due to buildings surrounding the street; and the station SMEAR III in Kumpula (HEL-UB, 60°12′N, 24°57′E, 26 m a.s.l.) an urban background site situated in a heterogeneous environment with significant anthropogenic impacts with forest, buildings, parking lots, and a main road nearby, located 4 km northeast from the city centre (Järvi et al., 2009; Luoma et al., 2021a).

The two sites in Germany are located in the city of Dresden, state of Saxony. Dresden-Nord (DDN-TR, 51°03′54″ N, 13°44′29″ E, 116 m a.s.l.) is a roadside traffic site located at 7 m away from roadways with a daily traffic volume of 11 k vehicles in 2022. Another site Dresden-Winckelmannstraße (DDW-UB, 51°02′10″ N, 13°43′50″ E, 120 m a.s.l.) is an urban background site located 1.7 km away from the city centre where traffic, wood combustion and long-range transport account for a significant portion of atmospheric pollutant sources (Birmili et al., 2016; Sun et al., 2019). It has an oceanic climate (Köppen-Geiger classification: Cfb).

2.1.2. *Instrumentation*

The instrumentation covers Multi Angle Absorption Photometer (MAAP), scanning mobility particle sizer (SMPS), differential mobility particle sizer (DMPS), condensation particle counter (CPC), twin differential mobility particle sizer (TDMPS), optical particle counters, tapered element oscillating microbalance (TEOM), and other conventional instruments of gaseous pollutants in the urban environments.

BC mass concentrations (in μg m^{-3}) were measured by a Multi-Angle Absorption Photometer (MAAP, Thermo Scientific model 5012) at all the five sites. The determination of the particle light absorption coefficient was performed using an operating wavelength of 670 nm, as indicated by the manufacturer. However, it should be noted that the actual wavelength employed by the instrument is 637 nm (Müller et al., 2011). The optically measured absorption data was then converted to equivalent black carbon (eBC) mass concentrations employing the default mass absorption cross-section (MAC) setting of 6.6 m^2/g at nominal wavelength of 670 nm. Furthermore, the instrument considers filter-loading-related artifacts that may impact the calculation of the absorption coefficient (Petzold and Schönlinner, 2004). The BC concentration measured with this technique is generally accepted to be named eBC; however, we unify the term to be BC in the rest of the paper for simplicity.

Particle number size distribution containing various size bins was measured with different established instruments. At BCN-UB, they were measured by using a Scanning Mobility Particle Spectrometer (SMPS) TSI3080 with a CPC TSI3772. At HEL-UB, a TDMPS Hauke-type DMA 10.9 cm and a CPC TSI 3025 were used. At HEL-TR, the measurements were conducted by a DMPS Vienna-type DMA and a CPC Airmodus A20.

At DDW-UB, a TROPOS-MPSS with Vienna-type DMA 28 cm coupled with a CPC TSI3772, and at DDN-TR, a TROPOS-TSMPS, Vienna DMAs 11 and 28 cm + CPC TSI model 3025 and 3010 were used. In order to make the data comparable, particle number concentration (in cm^{-3}) is aggregated by their size bins: nucleation mode (PN_{nuc}, for particle diameter $10 < D_p < 25$ nm), Aitken mode (PN_{Ait}, $25 < D_p < 100$ nm), accumulation mode (PN_{acc}, $D_p > 100$ nm) and summation of all sizes (PN). As very small sized particles (e.g. $D_p \sim 5$–15 nm) have a significant impact on PN, and in particular PN_{nuc}, a lower cut-off size of 10 nm is chosen for better comparison.

Mass concentrations of PM_X (PM_1, $PM_{2.5}$, PM_{10}, in μg m^{-3}) were measured using Optical particle counters at BCN-UB using a GRIMM 180 monitor. At HEL-TR and HEL-UB, concentrations of PM_{10} and $PM_{2.5}$ were measured by using a Tapered Element Oscillating Microbalance (TEOM 1405, Thermo Scientific TM). At DDN-UB and DDN-TR, PM_{10} concentrations were measured by using a TEOM 1405.

Gaseous pollutant concentrations (in μg m^{-3}) were measured by conventional instrumentation: nitrogen dioxide (NO_2) was determined by the chemiluminescence and O_3 by UV absorption (and/or IR-absorption) photometers. Meteorological parameters of temperature (T), relative humidity (RH), wind speed (WS) and pressure (P) were recorded with standard weather instruments.

2.2. Data pre-processing

The raw data stored in our database first went through pre-processing procedures. These include outlier removal and gap filling for values below detection limits. The dataset at this stage was examined and investigated for site characterization as described in Section 3.1. Data gaps of short missing period less than three hours were filled by simple linear interpolation. The data were then normalized and standardized such that each variable would follow a normal distribution with mean of 0 and standard deviation of 1 as a setup for ML process. We pre-selected the input variables by two steps: (1) we considered only parameters available at all the studied sites and (2) we removed the parameter with worse correlation with BC in case of collinearity. We finally used PM_{10}, PN, PN_{nuc}, PN_{Ait}, PN_{acc}, O_3, NO_2, T, RH, WS and P as input variables. $PM_{2.5}$ and NO_x were once included, but the former was not measured at all sites and the latter had a strong collinearity with NO_2. To increase the reliability of our comparison, we partitioned the first 70 % of the time series as training set and the last 30 % as testing set. A few combinations of training and testing proportion (1: 70 to 30, 2: 75 to 25, 3: 80 to 20, 4: 85 to 25 and 5: 90 to 10) were applied as a sensitivity analysis. As precipitation is known to intensify the wet deposition of air pollutants in the atmosphere, we tested the model by filtering the data with non-negligible precipitation (Blanco-Alegre et al., 2019). Similarly, we also tested by limiting downwind and upwind situation for traffic sites in this study (Hilker et al., 2019). Flag vectors were created for workdays and weekends (also includes holidays) for comparison. Seasons were also classified into winter (December, January, and February), spring (March, April, and May), summer (June, July, and August) and autumn (September, October, and November). Data analysis was conducted using MATLAB R2021a.

2.3. Machine learning methods

We selected seven methods (three for WB and four for BB) in this study due to their proven performance addressed by several researchers (e.g., Cabaneros et al., 2019; Fung et al., 2021b; Yu et al., 2023). The brief description of model architecture and their optimization criteria are outlined for WB and BB models in Section 2.3.1 and 2.3.2, respectively. The methods to quantify the relative importance of each feature for individual model are elaborated in Section 2.3.3 (summarized in Table 1).

2.3.1. White-box (WB) models

2.3.1.1. Input-adaptive proxy (IAP). IAP was initially introduced by Fung et al. (2020), and subsequently applied for the estimation of various air pollutant parameters (e.g., Fung et al., 2022b). This technique effectively estimated continuous BC concentration, achieving a coefficient of determination (R^2) exceeding 0.8. The approach involves selecting highly correlated input features beforehand, generating sub-models with a maximum of three input features each, using ordinary least-squares (OLS) linear regression. The method incorporates additional regularization through a 'bisquare' weight function, which relies on residuals, leverages from OLS fittings, and incorporates estimates of error term standard deviations, with a tuning factor of 4.685 as a robust alternative for datasets with numerous outliers as commonly encountered in field measurements. The regression is executed, and each sub-model's performance is assessed. Sub-models are ranked based on their performance using the employed evaluation metrics, prioritizing higher performance. The model looks for the best available input features to impute missing data based on the ranks of the sub-models. It is important to note that IAP is designed to handle missing data within its modeling process, differing from other models where missing data imputation is typically done prior to modeling.

Table 1
Summary of the machine learning methods and the corresponding relative importance approaches.

Model type	Model name	Principles	Missing data filling	Auto-regressive	Relative importance method	Static	Dynamic
WB	IAP	Ordinary least squares (OLS) based with sub-model ranking and automatic input selection	x		Pearson (*r*) and Spearman's rank correlation coefficient (pho)	x	
	LASSO	OLS with L1-norm penalty using λ as a hyperparameter			Beta-coefficients of the corresponding input variables	x	
	LASSO-ARIMA	LASSO, with residuals modeled by autoregressive integrated and moving average parts		x			
BB	RF	Aggregated decision trees using tree bagging techniques			Out-of-bag variable importance	x	
	SVM	Radial basis kernel with a box constraint to panelize observations diverging from predefined criterion using Lagrange multiplier			SHapley Additive exPlanations (SHAP), naturally dynamic showing local contribution, behaving as static relative importance after aggregation	(x)	x
	CNN	A single convolutional layer of [1,24] with kernel size of 3, followed by a max pooling layer of [1,2], optimized by five training cycles using Bayesian regularization as training loss minimization				(x)	x
	LSTM	Neural networks of 60 hidden layers with input, output and forget gates to take into account of short memory using logistic sigmoid activation function, optimized by adaptive momentum with an initial learning rate of 0.005		x			

2.3.1.2. Least absolute shrinkage and selection operator (LASSO). Initially introduced by Tibshirani (1996), LASSO gained widespread adoption within the realm of air pollutant prediction (e.g., Van Roode et al., 2019; Sethi and Mittal, 2021; Fung et al., 2023), in particular for BC estimation (e.g., Järvi et al., 2023). It constitutes a multiple linear regression technique that incorporates regularization to prevent overfitting. The regularization mechanism enforces penalties on different model parameters to curtail model flexibility and eliminate unnecessary predictor variables, which could be particularly beneficial when dealing with a large number of potential predictors. LASSO employs an L1-norm penalty by employing a geometric sequence of λ, where λ functions as a hyperparameter regulating the penalty strength. Employing a five-fold cross-validation, the optimal λ value is determined by identifying the minimum mean squared error. The selection of an appropriate λ holds utmost significance for LASSO's performance, as it governs the extent of shrinkage and variable selection. When properly balanced, it can enhance both prediction accuracy and interpretability. Nonetheless, an excessive regularization could omit crucial variables from the model and excessively shrink coefficients.

2.3.1.3. AutoRegressive integrated moving average (ARIMA). To account for the autoregressive characteristics of time series data, the residuals from the LASSO model are further addressed through an ARIMA model. The choice of ARIMA aims to incorporate temporal dependencies, in particular trends and seasonality. Researchers have utilized ARIMA to model and forecast temporal variations in BC concentrations, influenced by factors such as emission sources, meteorological conditions, and anthropogenic activities (e.g., Patil et al., 2022; Duan et al., 2023; Kaur et al., 2023; Makkhan et al., 2023). The autogressive (AR), integrated (I) and moving average (MA) parts have been optimized with the number of time lags, degree of differencing and the order of the moving average model, respectively. Specifically, an ARIMA(0,0,0) model seasonally integrated with seasonal AR(12) and MA(12), exhibited the smallest absolute errors in predicting residuals compared to the baseline LASSO model. This ARIMA-based approach serves as a comparison alongside one of the BB models, which similarly considers autoregressive properties.

2.3.2. Black-box (BB) models

2.3.2.1. Random forest (RF) – Bagging ensemble method. RF model is formulated by amalgamating outcomes from individual decision trees across various subsets (e.g., Yu et al., 2023). It shows good prediction performances with high-dimensional data inputs that has been used for evaluating trends in air quality under changing meteorological conditions (Qiu et al., 2022). The aggregation of multiple trees reduces overfitting and improves the model's generalization ability, making it robust for handling complex and noisy data, understanding of the relationship between BC levels and different variables. Employing a bagging technique, distinct random subsets are drawn with replacement from the original dataset. Each of these samples is subjected to the same learning method, culminating in the weighted combination of the resultant models. This ensemble method, rooted in bootstrap aggregation, serves to alleviate bias, reduce error variance, and enhance generalization, a principle outlined by Van Roode et al. (2019). The determination of split decisions relies on the same curvature test across the subsets. Additionally, Breiman's random forest algorithm is implemented to ascertain the number of variables to be selected randomly for each decision split, following the principles delineated in Breiman (1996).

2.3.2.2. Support vector machine (SVM). SVM constitutes a statistical learning framework formulated by Vapnik (1997), widely harnessed within the domain of air quality prediction (e.g., Fung et al., 2021b; May and Li, 2022; Rovira et al., 2022). By training the SVM on historical data of BC concentrations and relevant predictors, it can be used to forecast future BC levels. SVM can also be utilized for classification tasks to categorize AQ conditions based on BC concentrations and their threshold values. Operating on the principle of regression, SVM seeks a kernel function that optimizes the margin of tolerance for the regression fit. These pivotal vectors that define the kernel are termed support vectors. The underlying objectives of the SVM model are twofold: firstly, to identify a function that deviates from the training data's output variables by no more than a specified value for each training point; and secondly, to minimize flatness determined by a box constraint, a positive numerical parameter governing the penalty applied to observations diverging from the predefined criterion. This is achieved through the utilization of two Lagrange multipliers associated with support vectors and a radial basis kernel function, mirroring the approach adopted in Fung et al., (2021b). Recent research on predicting AQ index has identified that the performance of the SVM model is significantly influenced by three main factors: the penalty factor, the regularization parameter, and the choice of kernel function (Leong et al., 2019).

2.3.2.3. Convolutional neural network (CNN). Neural network models have been applied in the prediction of AQ (e.g., Cabaneros et al., 2019; Van Roode et al., 2019; Zaidan et al., 2019; Bekkar et al., 2021; Fung et al., 2021a; Duan et al., 2023). Among them, one dimensional convolutional neural network (1D-CNN) has been effectively applied on time series data mining (e.g. Zhu and Xie, 2023). A typical CNN as a regularized type of feed-forward neural network has three layers: convolutional layer, activation layer, and pooling layer. According to the insights shared in the review paper by Cabaneros et al. (2019), a shallow neural network with a solitary hidden layer containing an ample number of neurons can effectively model any finite input–output mapping issue involving non-linear relationships. To maintain simplicity, a single convolutional layer of [1,24] with kernel size of three was implemented. Maximum pooling with a sliding window of [1,2] was used as suggested by Mao and Lee (2019). Finally, the output was sent to a fully connected layer before prediction. We used ReLU as the activation function, which dictates the output value for each neuron, which subsequently becomes the input for neurons in the succeeding connected hidden layer. The weights were initialized randomly, and these weights were updated via gradient descent optimization, which might potentially lead to the vanishing gradient problem. To mitigate this issue, five training cycles were conducted, each comprising multiple iterations. The objective is to minimize the training loss using mean squared error function while incorporating Bayesian regularization with a default prior setting using a normal-inverse-gamma conjugate distribution within the Levenberg-Marquardt algorithm.

2.3.2.4. Long short-term memory (LSTM). LSTM was initially introduced by Hochreiter and Schmidhuber (1997), marking the start of its extensive exploration in AQ estimation (e.g., Cabaneros et al., 2019; Bekkar et al., 2021). On top of the architecture of neural networks, LSTM units effectively address the challenges of vanishing gradients and long-term dependencies by enabling the unhindered flow of gradients. This makes it potentially applicable to time series data with autoregressive properties like BC concentrations, which may exhibit complex temporal patterns and trends (Duan et al., 2023). A common architecture incorporates a cell (responsible for memory) and three regulators that govern information flow within the LSTM unit: an input gate, an output gate, and a forget gate. The cell maintains dependency relationships among input sequence elements. The input gate modulates the influx of new values into the cell, while the forget gate determines information to discard from the cell state. The output gate controls the cell value's contribution to computing the output activation of the LSTM unit block at a given timestamp. In this study, we used a sequence layer with length of 11 (same size as the number of input parameters) and LSTM layers of 60 hidden layers. We also used logistic sigmoid function as the activation function for the three LSTM gates. Several connections, including

recurrent ones, link into and out of the LSTM gates. The weights of these connections, learned during training, governed the gate behaviour. The output of the final step within the current LSTM block wass passed through two additional layers: a fully connected layer and a regression layer, culminating in the predicted output of the current block. For optimization, adaptive momentum ('adam', Freeman et al., 2018) with initial learning rate of 0.005 and gradient threshold of 1 was employed in this paper. Mean squared error was used as a loss function. No additional regularization was used.

2.3.3. Determination of relative importance of different ML models

To demonstrate the interpretability of the different WB and BB models, we introduced a set of methods to determine the relative importance of input variables depending on their different model structures. Whenever possible, static relative importance approach that only depends on training data should be used as this is most straightforward way to explain the overall contribution of a feature to the output. For the OLS-based IAP, we simply used Pearson correlation coefficient (r) as an indicator of the importance of the input variables as the selection basis for the input adaptive function is determined by the value of r (Fung et al., 2021b). Spearman's rank correlation coefficient (r_s), which is less sensitive to non-linear datasets, is additionally included as a baseline value. For LASSO(-ARIMA), since the datasets were normalized and standardized, the individual coefficient of variables could reflect their relative importance in the LASSO models. Similar method has been applied in explaining BC (e.g. Järvi et al., 2023). Furthermore, since RF model that used a bagging technique, an out-of-bag variable importance value that estimates by permutation measure how influential the predictor variables in the model are at predicting the response was used. The influence of a predictor increases with the value of this measure. If a predictor is influential in prediction, then permuting its values should affect the model error. If a predictor is not influential, then permuting its values should have little to no effect on the model error (Loh, 2002). For easier comparison, the relative importance of input variables calculated by individual methods were normalized to a range of 0 and 1 where input variable having 1 has a strongest contribution and vice versa.

There are no similar simple ways to quantify the relative importance of parameters of BB models like SVM and CNN. In this case, we calculated SHapley Additive exPlanations (SHAP), a Shapley-value-based explanation method based on the coalitional game theory introduced by Lundberg and Lee (2017), as a unifying framework to interpret and compare different types of data-driven BB models. The key idea of using SHAP in AQ models is to calculate the Shapley values for each feature of the sample to be interpreted, where each SHAP value represents the impact that the feature to which it is associated, generates in the prediction (e.g., Wang et al., 2020; Gu et al., 2021). SHAP values which illustrate the predictor variable's contribution to each data point. The dynamic contribution to each point is dependent on each training and testing combination sets.

2.4. Trend analysis

To explore the long-term trends within air monitoring data and assess their significance, we employed the Mann-Kendall test and Sen's slope estimator, both of which are nonparametric statistical methods capable of handling missing data points. These methods find widespread application in the analysis of environmental data (e.g., Collaud Coen et al., 2020; Savadkoohi et al., 2023). The Mann-Kendall test evaluates whether a consistent long-term trend in a given variable holds statistical significance and whether it demonstrates a monotonic increase or decrease for $p < 0.05$. Meanwhile, Sen's slope estimator gauges the magnitude of the trend.

To account for the challenges stemming from cyclic data patterns, like seasonal variations, weekend effects, and diurnal cycles linked to factors such as boundary layer dynamics or traffic rates, we employed a seasonal version of the Mann-Kendall test and Sen's slope estimator. This adaptation overcomes autocorrelation concerns inherent to cyclic data. Our analysis focused on monthly median values. To be included in the trend analysis, valid data spanning a minimum of 14 days within each month were required; otherwise, the month was excluded. The trend analysis was specifically conducted for Barcelona's measured BC values and their relative importance.

2.5. Evaluation attributes

In order to evaluate and compare the accuracy of the models, coefficient of determination (R^2), together with mean absolute percentage error (MAE) and root mean square percentage error (RMSE), are used as diagnostic evaluation attributes. R^2 (ranged from 0 to 1) is a measure of how close the data lie to the fitted regression line. It, however, does not consider the biases in the estimation. Therefore, we further validate the models with MAE and RMSE. With both metrics expressed in percentage (ranged from 0 to 100 %), they easily show how much errors the models generate in comparison with the original data. The difference of them is that MAE measures the arithmetic mean of the absolute differences between the members of each pair while RMSE calculates the square root of the average squared difference between the estimate and the observation pairs. RMSE is more sensitive to larger errors than MAE. In addition to accuracy, the performance of a model could be described in respect of its simplicity/complexity as measured by the computational time of training the model.

3. Results and discussion

3.1. Site characterization at the urban site in Barcelona and other testing sites

Fig. 2 presents the time series of various pollutants, including BC, NO_2, PM_{10} and PN at the BCN-UB site. These data are visualized with daily mean, monthly mean, and yearly mean variations in pollutant concentrations over the period from 2013 to 2022. The figure also incorporates trend lines of Sen's slope estimator, which show the overall magnitude of changes in pollutant levels during this timeframe, tested with Mann-Kendall test for their statistical significance. Notably, the BC concentration at the BCN-UB site exhibits a statistically significant trend of –0.09 μg m^{-3} per year over the 2013–2022 interval. These findings are consistent with recent research on BC trends, which disclosed a decrease of approximately 4.7 % per year in BC concentrations from 2010 to 2020 at this site (Savadkoohi et al., 2023). This trend also aligns with earlier reports of a substantial BC reduction of around 18 % from 2014 to 2018 (Via et al., 2021). These observations underscore the role of traffic emission mitigation policies in driving the observed decrease in BC mass concentrations.

Additionally, PN and NO_2 concentrations display statistically significant annual decrease of 598.63 cm^{-3} and 1.52 μg m^{-3}, respectively. The latter trend agreeing with previous research over the period 2003–2014 reported significant declining trends in Barcelona of up to 30 % of NO_2 (Casquero-Vera et al., 2019). While O_3 and PM_{10} concentrations also showed decreasing trends, these trends are not statistically significant ($h = 0$). Regarding O_3, previous studies have revealed an increase of urban O_3 concentrations within the Barcelona metropolitan area ranging from 0.4 % per year to 3.2 % per year. The spatio-temporal distribution of O_3 hotspots in Spain has not been exhibited clear trends in previous studies, emphasizing the need for targeted local and regional mitigation measures to address chronic and episodic O_3 exposure (Massagué et al., 2023).

In order to compare the training site BCN-UB with other external sites, we illustrated the distribution of data points for BC (Fig. S2). They all appear to follow a log-normal distribution. The highest average of hourly BC concentrations over the whole period was detected in BCN-UB (1.31 μg m^{-3}), followed by the two traffic sites HEL-TR (1.04 μg m^{-3})

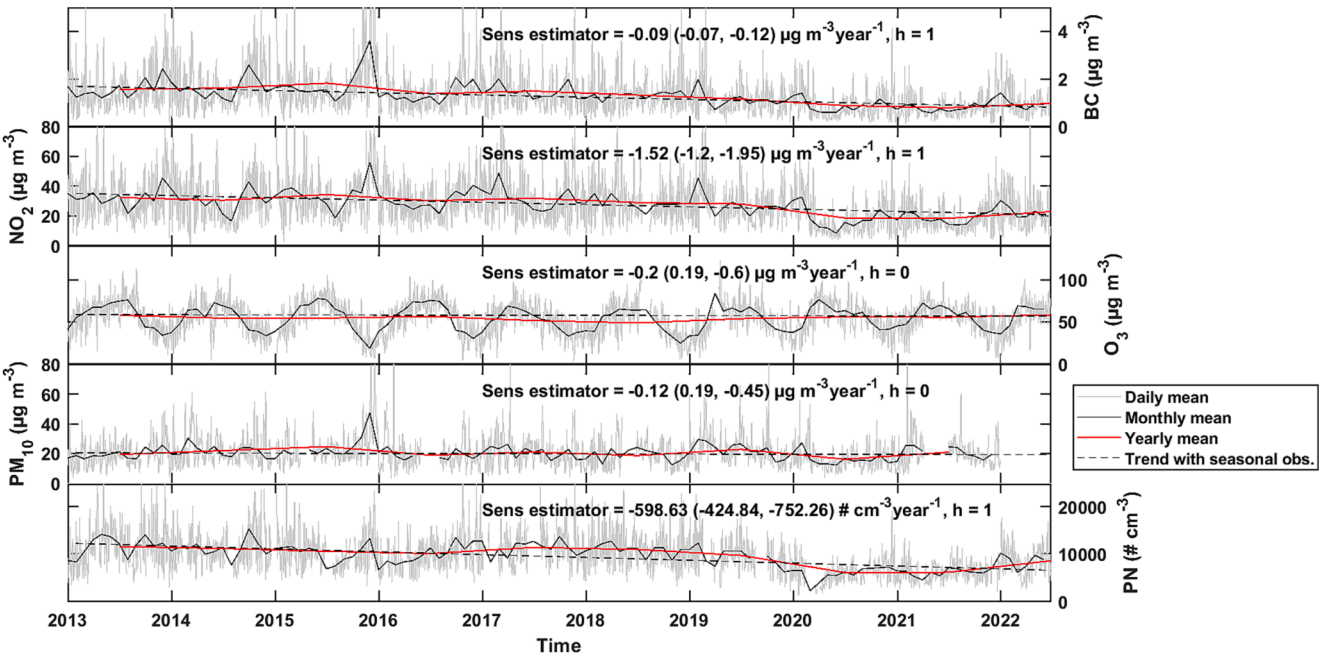

Fig. 2. Timeseries (daily mean, monthly mean, yearly mean and trend) of air pollutants BC, NO_2, O_3, PM_{10} and PN in BCN-UB. Similar timeseries plots for testing sites could be found in supplementary materials.

and DDN-TR (0.97 μg m^{-3}). DDW-UB comes fourth (0.74 μg m^{-3}) and HEL-UB appears to record the lowest average BC (0.48 μg m^{-3}). Besides, time series plots (daily mean and monthly mean) of air pollutants BC, NO_2, O_3, PM_{10} and PN in these testing sites site are presented in Fig. S3 in the year between 2017 and 2019. It is worth noting that these datasets lasted for less than three years; therefore, it is not statistically significant to make direct comparison with the dataset retrieved in BCN-UB. The potential for analysing long-term trends in pollutant concentrations at these sites is also hindered for the same reason.

Fig. 3 represents the diurnal cycles of BC concentrations across the five urban traffic sites in Europe in different seasons during both workdays and weekends. Notably, BC exhibits diurnal cycles characterized by peak concentrations during traffic rush hours on workdays, particularly evident at the BCN-UB site. This site, located along one of the city's busiest roads, is marked by the direct emission of particles from vehicle exhausts (Rivas et al., 2020). These diurnal cycles are most pronounced in the autumn and winter seasons. Similar but less distinctive diurnal patterns were also observed at the Northern and Central European TR sites with slightly lower BC concentrations (Savadkoohi et al., 2023). Conversely, at HEL-UB and DDW-UB, variations in the BC diurnal cycles are less pronounced on weekdays. However, evening peaks appear to be more distinctive, primarily attributed

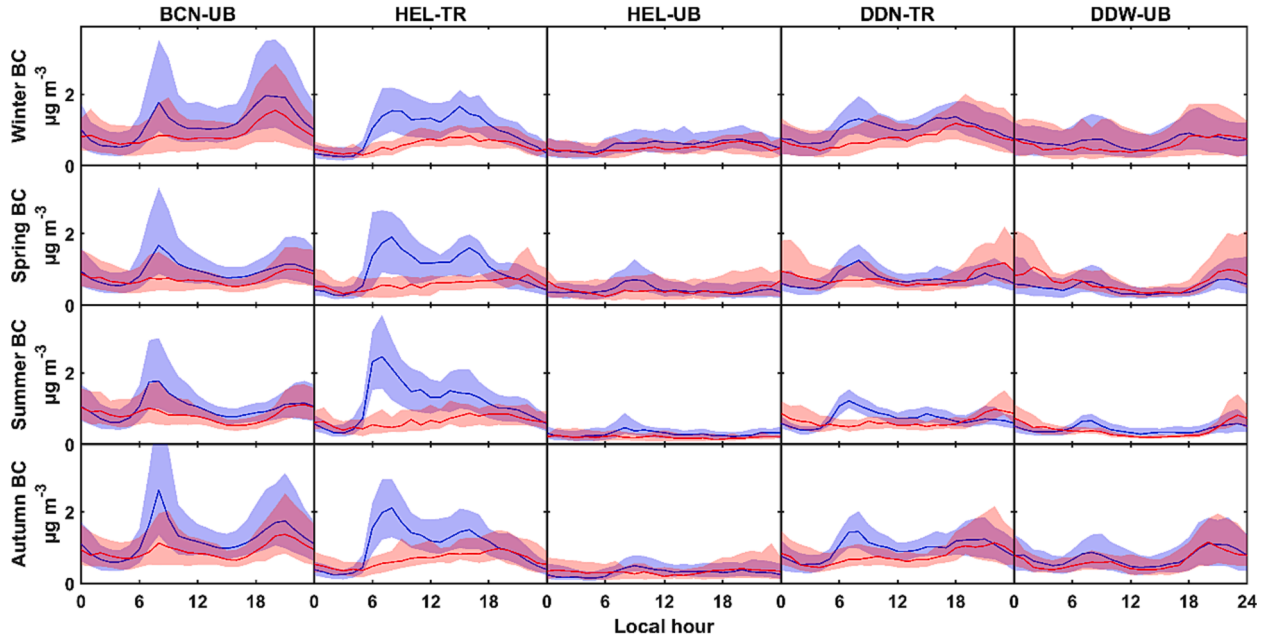

Fig. 3. Diurnal cycles of BC concentrations at five locations (columns) and in four seasons (rows) during workdays (shaded in blue) and weekends (shaded in red). Similar graphs for NO_2, O_3, PM_{10} and PN could be found in supplementary materials. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the elevated residential burning (Fung et al., 2020). Similar diurnal cycles of NO_2, O_3, PM_{10} and PN are also, respectively, presented in Fig. S4–S7.

Despite the difference in geographical locations and measurement sites, the BC concentrations measured at the individual site show high correlation with other aerosol and gaseous compounds measured at the same site (Fig. 4). The strongest correlation on average is, in descending order, PN_{acc} ($r = 0.73$–0.85) and NO_2 ($r = 0.68$–0.85), followed by PM_{10} ($r = 0.54$–0.76) and PN ($r = 0.48$–0.72), with the weakest correlation being O_3 ($|r| = 0.36$–0.58), PN_{Ait} ($r = 0.39$–0.66) and PN_{nuc} ($r = 0.11$–0.61). Among them, the correlations with PN_{nuc} have a strongest variability within the five sites, scattering from HEL-TR with r of 0.61 to DDW-UB of r being 0.11. PM_{10} in the two sites in Dresden also have a considerably stronger correlation with BC compared to the sites in other location. While in Helsinki the road dust induced PM_{10} is very high in the springtime (Fig. S6) due to studded winter tyres and winter sanding of the streets, the PM_{10} concentrations in Barcelona could be attributed to Saharan dust and re-suspension of urban dust. Dresden, on the other hand, is less impacted by dust but more by emission from traffic and wood combustion, as indicated by the stronger correlation with BC. These show that the particle size distribution constituting the BC concentrations is different to a certain extent depending on their geographic locations and types of sites. Furthermore, weather conditions correlate less well with BC. While wind speeds demonstrate a substantial negative correlation of r between –0.52 and –0.29, temperature, RH and P correlate much less with BC either positively or negatively ($|r| = 0.04$–0.27). That being said, the overall strong correlation of BC with other parameters suggests the potentiality for the estimation proxies using ML methods.

3.2. Comparison of different ML methods to derive the BC concentration

Table S1 presents the best combination of proportion of training and testing set in different measurement sites and different seasons using different ML methods. None of the selected combinations dominates in all groups. Combination 4 appears to outperform in CNN while combination 1 and 5 are often found to generate results of higher accuracy in some specific measurement sites. Overall, the mode of the best combination is 2 across all measurement sites, seasons and ML methods. Therefore, for easier comparison, combination 2 of training and testing proportion will be adopted in the rest of the paper. Furthermore, based on our prior knowledge that BC concentrations would be strongly influenced under certain weather conditions, such as precipitation which enhances the deposition and scavenging process of BC (Blanco-Alegre et al., 2019) and downwind situation for street canyon type of TR sites which restricts the dispersion of air pollutants (Hilker et al., 2019), we tried to optimize the estimation by excluding data with precipitation for all measurement sites and including data only with downwind conditions for TR sites. However, these procedures did not show distinct improvement to the models; therefore, we only show the results without these considerations.

Graphically, the BC concentrations by all the models tested in Barcelona mostly follow the 1:1 line as shown with a dark red color indicating the highest density of data points (Fig. 5). In the meantime, scarce amounts of points scatter from the central line as outliers. All the seven models illustrate similar patterns. Statistically, Tables 2, S2 and S3 show that the four BB models perform better on average ($R^2 = 0.78$–0.83, MAE = 4.23–4.61 %, RMSE = 5.00–5.75 %) compared to the three WB models ($R^2 = 0.75$–0.76, MAE = 4.93–5.01 %, RMSE = 6.02–6.15 %). Among them, LSTM outperforms the rest in all the four seasons plausibly because this model considers the time-dependency properties of the dataset. In general, estimations in winter and summer ($R^2 > 0.8$) surpass those in autumn and the worst is in spring ($R^2 = 0.67$–0.70).

Fig. 6 compares and evaluates the seven ML methods in terms of their accuracy indicated by R^2 (x-axis) and complexity indicated by computation time (y-axis). Each dot with solid color representing the overall performance of individual machine learning method is the average of the dots with the corresponding oblique color that represent individual runs for each season of that method. The figure demonstrates that the more complex the model is, the better the estimates the model calculates. The cluster of the simpler models (average computation time < 50 s), IAP, LASSO and CNN, shows a lower accuracy ($R^2 < 0.775$) compared to the other cluster of more complex models (average computation time > 50 s), RF, SVM and LSTM. Among all the models used, LASSO-ARIMA behaves as an outlier such that it has a relatively long computation time, yet the accuracy performance is not as good as some of the less complex models. This is because, on top of LASSO, ARIMA was built based on the model residuals and its optimization of the three components of ARIMA consumed an extensive period of computing time. In addition, all except LSTM have a consistent performance within all the individual runs. Although all runs for LSTM have similar computation time, their accuracy in terms of R^2 range from 0.76 to 0.81 which are more scattered compared to the other methods.

3.3. Upscaling the BC proxies to different environments

It is obvious that the models trained with BCN-UB data work well for the BCN-UB testing data as these models have learnt the site-specific data. The transferability to the other sites has been demonstrated to be relatively uncertain with different site classification and geographical locations (e.g., Ameer et al., 2019; Fung et al., 2021b). Surprisingly, the models in our study work well also in other testing sites. In particular locations, the performance is even better than the one at BCN-UB (HEL-TR: $R^2 = 0.80$–0.86; MAE = 3.90–4.73 %; RMSE = 4.92–5.89 % and DDW-UB: $R^2 = 0.79$–0.84; MAE = 4.23–4.82 %; RMSE = 5.26–5.92 %). Fig. 7 show the scatter plots of calculated BC trained with BCN-UB data against measured BC at the two respective testing sites. The former shows the models work typically well at the testing site HEL-TR where most data points lie along the 1:1 line. The other one representing DDW-UB show a cut-off of measured BC concentration at the detection limit of the instrument 0.1 μg m^{-3}, which is coarser than the one used at BCN-UB. However, this did not hinder the model performance in terms of

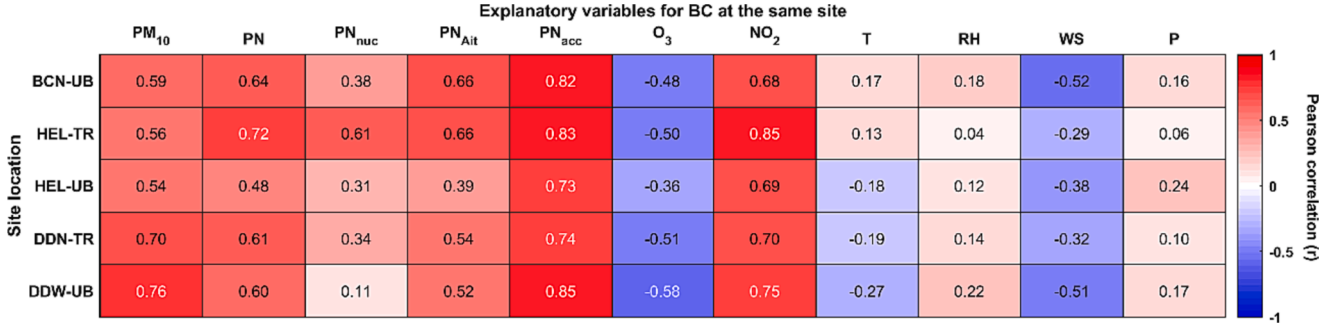

Fig. 4. Pearson correlation (r) of BC with other pollutant and meteorological parameters measured at the same site.

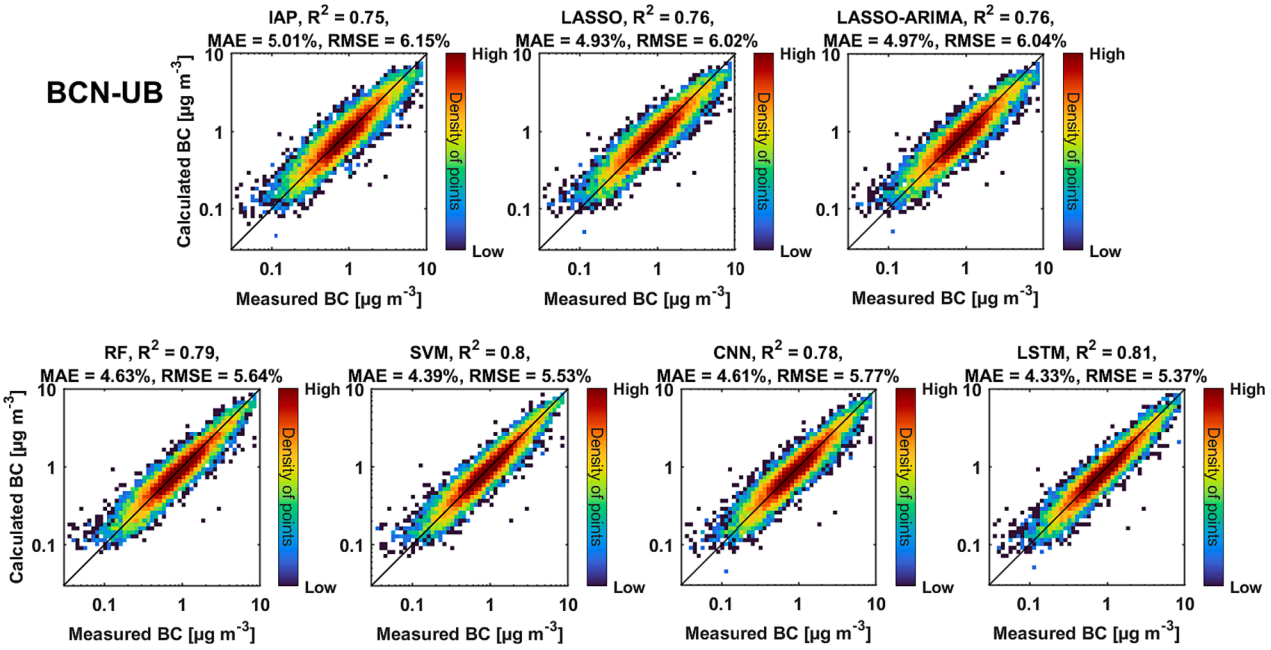

Fig. 5. Scatter plots of the testing data at BCN-UB calculated BC against measured BC for different methods used colored by the point density. R^2, MAE and RMSE are presented on the subplot titles. Similar scatter plots of the training data at BCN-UB can be found in Fig. S8.

Table 2
R^2 of models trained in BCN-UB and tested in various sites in different seasons using different ML models.

		IAP	LASSO	LASSO-ARIMA	RF	SVM	CNN	LSTM
BCN-UB	Winter	0.83	0.85	0.85	0.87	0.88	0.87	0.88
	Spring	0.67	0.68	0.68	0.69	0.67	0.67	0.70
	Summer	0.80	0.82	0.82	0.84	0.85	0.84	0.87
	Autumn	0.75	0.77	0.77	0.80	0.79	0.80	0.82
	All	0.75	0.76	0.76	0.79	0.80	0.78	0.83
HEL-TR	Winter	0.91	0.90	0.90	0.89	0.88	0.86	0.88
	Spring	0.88	0.88	0.88	0.86	0.85	0.83	0.85
	Summer	0.82	0.83	0.83	0.84	0.83	0.78	0.83
	Autumn	0.82	0.83	0.83	0.83	0.81	0.76	0.82
	All	0.86	0.85	0.85	0.85	0.83	0.80	0.84
HEL-UB	Winter	0.77	0.77	0.77	0.74	0.74	0.76	0.72
	Spring	0.86	0.87	0.87	0.84	0.82	0.86	0.81
	Summer	0.58	0.63	0.63	0.62	0.59	0.61	0.56
	Autumn	0.60	0.62	0.62	0.6	0.57	0.57	0.57
	All	0.66	0.68	0.68	0.66	0.65	0.65	0.61
DDN-TR	Winter	0.64	0.67	0.67	0.69	0.70	0.70	0.67
	Spring	0.71	0.73	0.72	0.74	0.73	0.75	0.72
	Summer	0.60	0.62	0.61	0.66	0.68	0.66	0.63
	Autumn	0.25	0.30	0.30	0.36	0.39	0.35	0.35
	All	0.59	0.62	0.62	0.65	0.67	0.66	0.63
DDW-UB	Winter	0.85	0.85	0.85	0.84	0.84	0.82	0.85
	Spring	0.85	0.86	0.85	0.84	0.83	0.81	0.86
	Summer	0.79	0.80	0.79	0.79	0.77	0.74	0.78
	Autumn	0.67	0.69	0.68	0.73	0.72	0.71	0.76
	All	0.82	0.82	0.82	0.82	0.81	0.79	0.84

accuracy. The models manage to catch the BC diurnal patterns both during workdays and weekends at the two locations. Results by LASSO and LSTM representing WB and BB models can be seen in Fig. 8. The other two testing sites with less prominent results are HEL-UB (R^2 = 0.61–0.68; MAE = 7.06–7.99 %; RMSE = 8.74–9.77 %) and DDN-TR (R^2 = 0.59–0.67; MAE = 4.59–5.13 %; RMSE = 5.63–6.30 %) where the models trained in BCN-UB could still explain more than 60 % of the data at these two sites. This very good transferability across space and time could be attributed to the similar strong correlation of BC with other pollutant parameters measured at the same site. This similarity in correlation behaviour provides a strong basis for the machine learning models to work as they learn the hidden trends and patterns from the training dataset. This strong yet hidden patterns of BC in turn trivialize the external factors such site location and measurement time. In respect

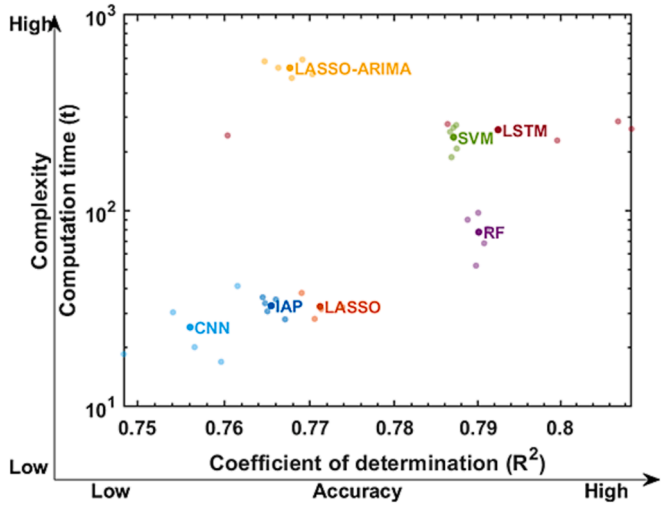

Fig. 6. Comparison of different methods used in terms of accuracy (x-axis, R^2) and complexity (y-axis, computation time in logarithm scale).

with which individual model work best for these external testing sites, the results show inconsistency. The two WB models IAP and LASSO works best at HEL-TR and HEL-UB, respectively, while the two BB models SVM and LSTM has the highest R^2 at DDN-TR and DDW-UB, respectively. Unlike testing at BCN-UB described in Section 3.2, estimations in spring show the highest R^2 when upscaling to the other testing sites while the trained models explain only 25–39 % of the testing data at DDN-TR in the autumn.

3.4. Relative importance of input data to ML derived BC

In addition to the model transferability, we improved the interpretability of models, both WB and BB models, by calculating the relative importance of the input variable using various methods for individual model. Fig. 9 illustrates the normalized static relative importance of each input variables calculated for individual model based on the training data. They are presented in descending order where PN_{acc} is the most important input variable in all models used evaluated by their respective metrics. This parameter was also found to have played a consistently major role in estimating BC for the past decade, as demonstrated in the trend analysis of its yearly relative importance. Although vehicular emission reduction technologies have been advancing in recent years (e.g. Brewer, 2019; Xu et al., 2021), no statistically significant trends of relative importance were found. NO_2 ranks the second on average (~0.6) where LASSO, SVM and IAP consider NO_2 as the second important variable. Although the correlation is high between BC and NO_2, CNN does not consider NO_2 to explain the variability of BC at all. This is due to the high collinearity of PN_{acc} and NO_2, both of which come from similar anthropogenic source that constitutes a considerate proportion of BC. By first examining these two input variables, the variability of BC is already well explained by PN_{acc} in CNN, and NO_2 fails to supply new information to contribute to the estimation. The other ML models, on the other hand, might have taken the approach to retrieve the patterns partially from both input variables; therefore, both normalized relative importance values are high. Moreover, the other aerosol variables PM_{10} and PN, although having a relative strong correlation with BC, contribute very little to the estimation, as the information they could provide overlap with PN_{acc} and NO_2. Another interesting point from Fig. 9 is that O_3 has similar relative contribution (~0.2) in all the models. This indicates that O_3 is able to supplement moderately to the BC variability. Although the contribution being moderately low, it provides unique piece of information to the estimation regardless of model architecture. Ground-level O_3 is a secondary pollutant formed through chemical reactions (Massagué et al., 2023), which is different from the emission source of other aerosol and gaseous compounds used in the study. However, similar to BC measurements, particle size distribution is not always included as part of the regulated pollutants within cities' air monitoring network. The construction of the BC estimation models could be more useful yet less accurate in practice if only regulated pollutants, such as O_3, NO_2 and $PM_{2.5}$.

As expected, meteorological parameters contribute a relatively trivial part of the estimation for most of the models. Although WS correlates negatively at a moderate degree, the relative importance analysis indicates its negligible contribution, let alone the other meteorological parameters of even lower correlation. However, unlike the other models, RF acknowledges meteorological conditions as important parameters (>0.6). This is plausibly due to its unique model architecture that RF consists of many decisions trees that use bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree (Yu et al., 2023).

As for dynamic relative importance, which indicates the local contribution of each input variable to each data point, Fig. 10 presents the average of SHAP values of two BB models, SVM and CNN, with error bars for all testing sets. Although LSTM has an overall higher accuracy (see Fig. 6), the performance over different cross validation runs is inconsistent. This inconsistency might impose significant uncertainties on the SHAP analysis. Therefore, we chose to present the SHAP values of SVM and CNN over LSTM. The results of static relative importance are comparable with the dynamic relative importance at BCN-UB with aerosol parameters like PN_{acc} the strongest and meteorological parameters like P the lowest. However, when it comes to other external testing sites, contrasting results are found at individual sites. For example, the best predictor deduced from the static relative importance, PN_{acc}, demonstrates to have a highly positive SHAP values at the two TR sites using both SVM and CNN while it has an overall negative SHAP value at the two UB sites. These values represent the local contribution of a feature to the output of a model. This serves as one of the limitations of SHAP values, which is the fact that factor contributions could not be represented as a single digit, meaning that the outcome of factors in a model is usually impacted by the other factors so that all the factors in a model will not have a constant impact on the output of a model (Lundberg and Lee, 2017). The representation of SHAP values would be a better choice if the purpose is to understand the feature importance at a particular timestamp or for a certain period which could be contradictory from the static importance. With that said, the introduction of SHAP values in explaining the local impacts a feature exert on the output of a model provides an alternative and new insights into the interpretability of a model. Furthermore, with the elevated interpretability of BB models, aerosol scientists might be able to find out the unknown patterns or rules that are learned by the more complex yet accurate algorithms. BB models would become more generalizable through transfer learning based on the hidden patterns. Therefore, model interpretability could benefit in updating and upscaling the model to other environments. The results from the model transferability could in turn further validate the model interpretability. This synergy might require several stages of trial and error, but the co-benefits it brings would be significant in explaining the characteristics of air pollutants like BC.

4. Conclusion

Receiving increasing attention from health experts and policymakers, black carbon (BC) as an air pollutant has gained recognition of its health impact, and thus its importance to be recommended as one of the regulated parameters within air quality (AQ) monitoring network. Machine learning (ML) models, although being criticized for its lack of generalizability and accountability, have been suggested to supplement BC reference measurements as virtual sensors in the absence of data due to financial constraints or instrument failure. In this study, we aim to show the transferability and interpretability of the selected data-driven

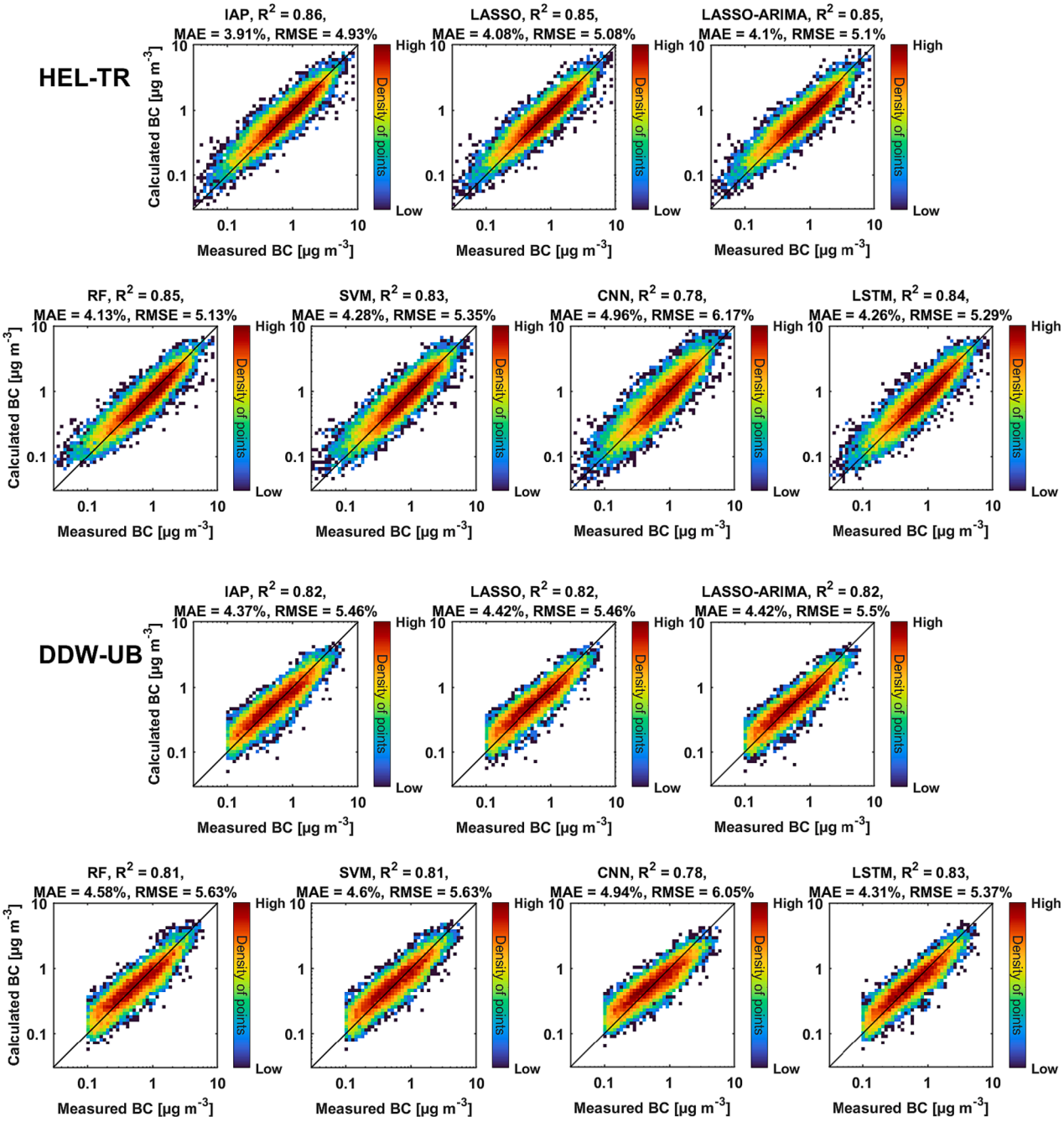

Fig. 7. Scatter plots of the testing data at HEL-TR (upper panel) and DDW-UB (lower panel) calculated BC against measured BC for different methods used colored by the point density. R^2, MAE and RMSE are presented on the subplot titles.

models using long-term BC measurements collected in an urban site in Barcelona (BCN-UB). We investigated the general BC characteristics and tested the feasibility of BC proxies by calculating the correlation of BC with other parameters measured at the same site. Trained using the data at BCN-UB, we tested the machine learning models of different architectures at four external sites in Northern and Central Europe. We evaluated which ML model works best and which parameters contribute most to the estimation.

The dataset in BCN-UB shows a statistically significant declining trend for BC, NO_2 and PN in the interval of 2013–2022, which are in alignment with previous studies. BC exhibits diurnal cycles characterized by peak concentrations during traffic rush hours on workdays, particularly evident at BCN-UB and two traffic (TR) sites. Regardless of the geographic locations and types of sites, BC has high correlation with other aerosol and gaseous compounds measured at the same site, with the strongest being accumulation mode (r = 0.73–0.85) and NO_2 (r = 0.68–0.75) and the weakest being the meteorological parameters. The strong correlation suggests the potentiality for the estimation proxies using different machine learning methods.

Four BB models perform better on average (R^2 = 0.78–0.83, MAE = 4.23–4.61 %, RMSE = 5.00–5.75 %) compared to the three WB models (R^2 = 0.75–0.76, MAE = 4.93–5.01 %, RMSE = 6.02–6.15 %). Among them, LSTM outperforms the rest in terms of accuracy, yet consumes most computational time, in all the four seasons plausibly because this model includes additional layers for the consideration of the time-dependency properties of the dataset. From the perspective of transferability, the model performs even better in some external locations (HEL-TR: R^2 = 0.80–0.86; MAE = 3.90–4.73 %; RMSE = 4.92–5.89 %

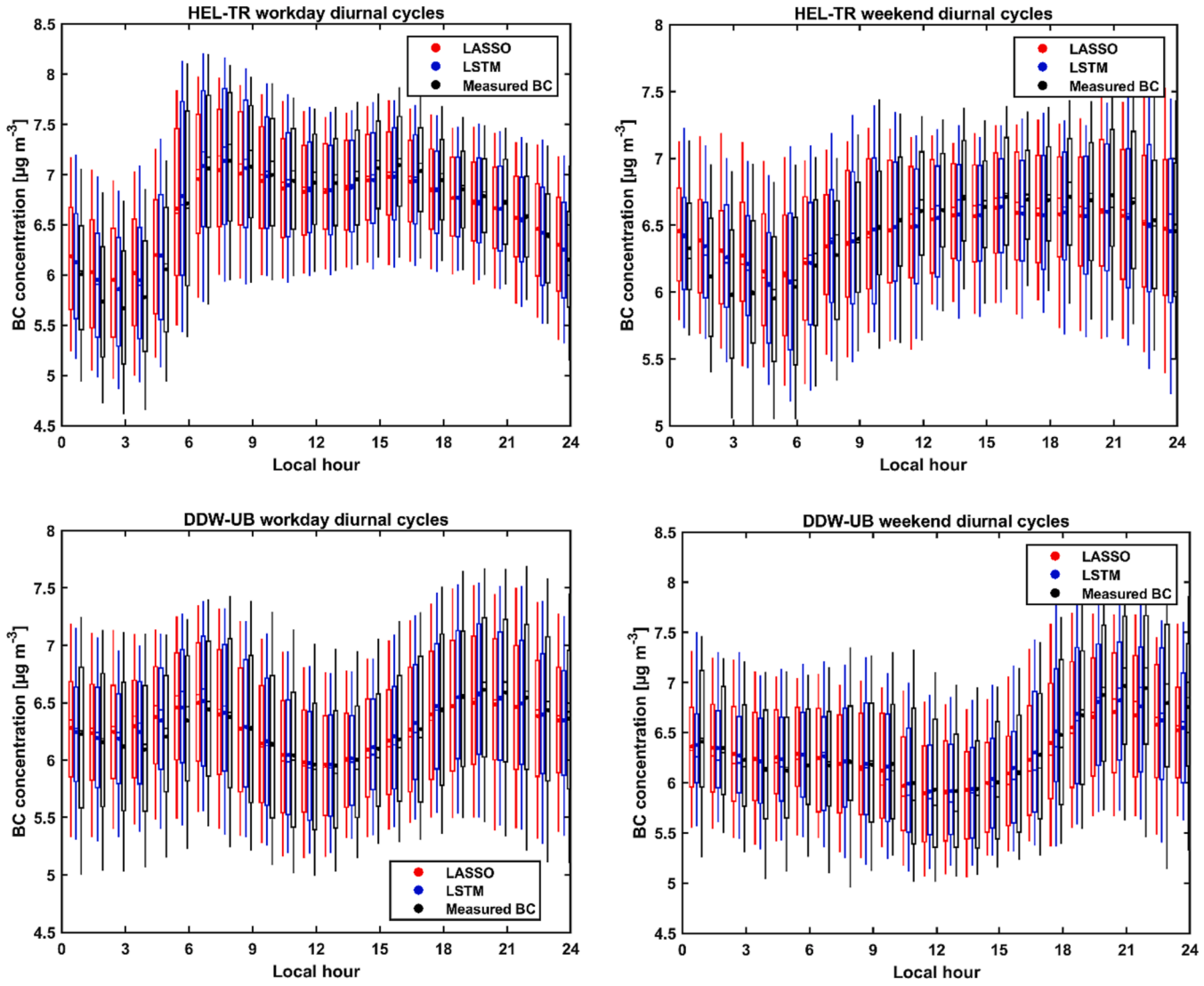

Fig. 8. Diurnal cycles of calculated BC by LASSO (red) and LSTM (blue) in comparison with the measured BC concentration (black) at two testing sites with prominent results in the form of box plot. The first row is the cycles for HEL-TR and the second row is DDW-UB while the first column illustrates workday condition and the second is weekend. The box plot has the component of lower whisker, lower box, median, upper box and upper whisker, which correspond to 10th, 25th, 50th, 75th and 90th percentiles of the BC distribution, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

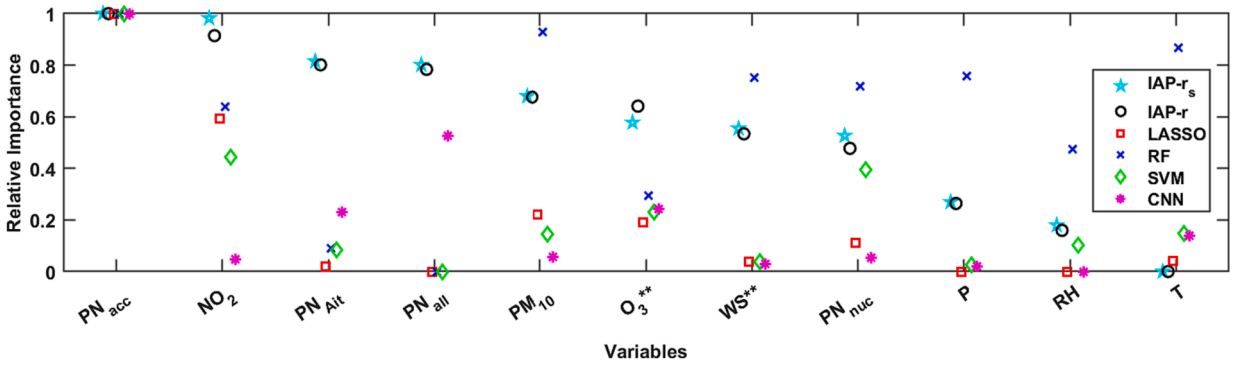

Fig. 9. Static relative importance of explanatory parameters using different machine learning methods.

and DDW-UB: $R^2 = 0.79–0.84$; MAE = 4.23–4.82 %; RMSE = 5.26–5.92 %) than the one in BCN-UB. This very good transferability could be attributed to the similar strong correlation of BC with other parameters measured at the same site.

In terms of interpretability, the static normalized relative importance that tells the overall contribution using respective metrics show PN_{acc}

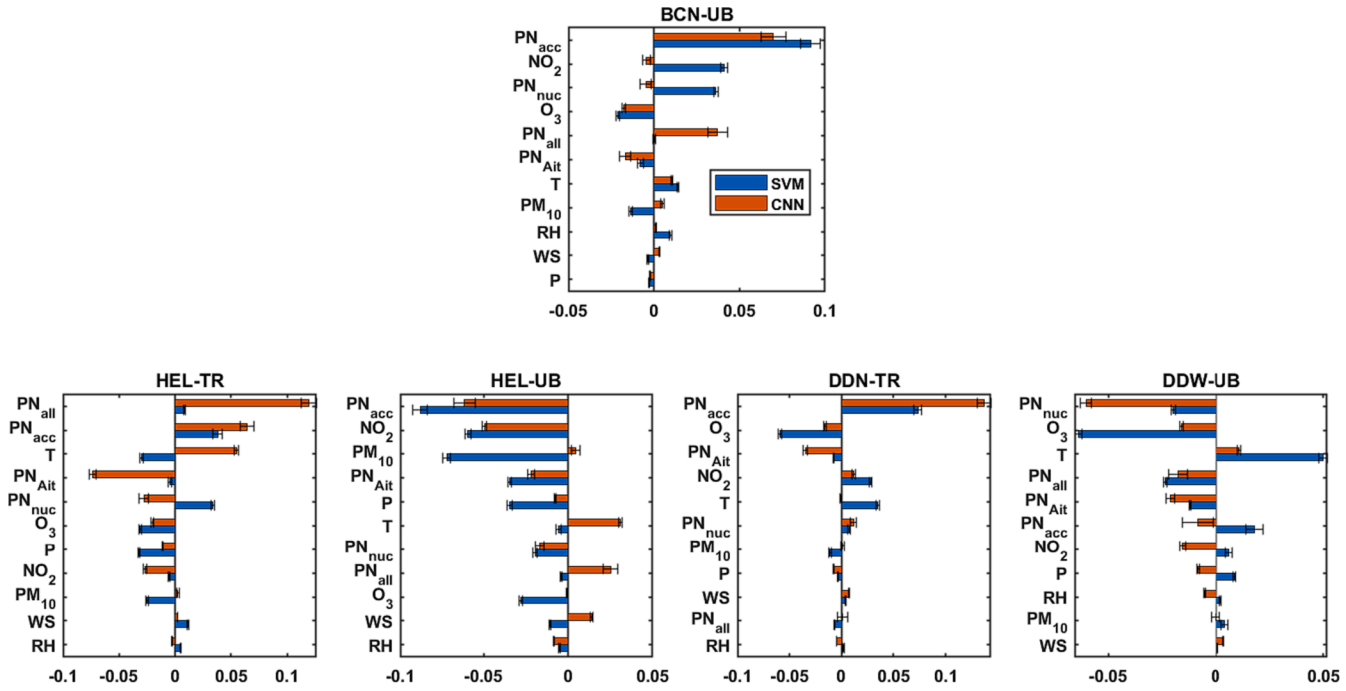

Fig. 10. Relative importance of explanatory parameters as SHAP values using SVM (blue bars) and CNN (orange bars) for the five testing sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and NO_2 are the two most important parameters for the estimation (>0.6). However, the dynamic relative importance SHAP values that represent the local contribution of a feature to the output of a model give varying results. They demonstrate to have positive impacts on BCN-UB and the two TR sites using both while it has an overall negative local impact at the two urban background (UB) sites. The introduction of SHAP values provides new insights into the overall and local interpretability of a BB model.

Although data-driven models have been long regarded to be site specific and lack of accountability, this comprehensive analysis shows that the BC model trained in Barcelona works well in terms of accuracy in other European sites with comprehensive information to explain the model. This transferable and interpretable proxy serves as an important supplement in case of missing data due to instrument failure. So far, the model transferability and interpretability were only tested at four external sites (urban background and traffic) in Europe. To enhance the generalization and representativeness of the model to the next level, it would be valuable to include sites with diverse emission profiles (e.g. detached housing areas with residential wood combustion, harbours and airports) in Europe and on other continents.

CRediT authorship contribution statement

Pak Lun Fung: Data curation, Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Marjan Savadkoohi:** Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. **Martha Arbayani Zaidan:** Validation, Writing – review & editing. **Jarkko V. Niemi:** Data curation, Validation, Writing – review & editing. **Hilkka Timonen:** Data curation, Validation, Writing – review & editing. **Marco Pandolfi:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition. **Andrés Alastuey:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition. **Xavier Querol:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition. **Tareq Hussein:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition. **Tuukka Petäjä:** Validation, Resources, Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

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

Data availability

The data that support the findings of this study are available from the corresponding authors upon request. Any requests for materials should be addressed to the corresponding authors.

Acknowledgements

This study is supported by the RI-URBANS project (Research Infrastructures Services Reinforcing Air Quality Monitoring Capacities in European Urban & Industrial Areas, European Union's Horizon 2020 research and innovation program, Green Deal, European Commission, contract 101036245). RI-URBANS is implementing the ACTRIS (http://actris.eu) strategy for the development of services for improving air quality in Europe. The authors would also like to thank the support from "Agencia Estatal de Investigación" from the Spanish Ministry of Science and Innovation under the project CAIAC (PID2019-108990RB-I00), AIRPHONEMA (PID2022-142160OB-I00), and the Generalitat de Catalunya (AGAUR, SGR-447), Technology Industries of Finland Centennial Foundation to Urban Air Quality 2.0 project, Research Council of Finland Flagship funding (project number: 337549, 337552), Research Council of Finland Research Fellowship funding (project number: 355330) and European Commission via on-CO2 Forcers And Their Climate, Weather, Air Quality And Health Impacts (FOCI, project number: 101056783). P.L. Fung would like to acknowledge Artificial Intelligence for Urban Low-Emission Autonomous Traffic (AIforlessAuto) funded under the Green and Digital transition call from the Research Council of Finland (project numbers: 347197, 347198) for the support. M. Savadkoohi would like to thank the Spanish Ministry of

Science and Innovation for her FPI grant (PRE-2020-095498). The authors also take the opportunity to thank Dr. Susanne Bastian from the Saxon State Office For Environment for contributing to data collection in the study. Open access funded by Helsinki University Library.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2024.108449.

References

Abu Awad, Y., Koutrakis, P., Coull, B.A., Schwartz, J., 2017. A spatio-temporal prediction model based on support vector machine regression: ambient Black Carbon in three New England States. Environ. Res. 159, 427–434. https://doi.org/10.1016/j.envres.2017.08.039.

Ahmed, T., Dutkiewicz, V.A., Khan, A.J., Husain, L., 2014. Long term trends in Black Carbon Concentrations in the Northeastern United States. Atmos. Res. 137, 49–57. https://doi.org/10.1016/j.atmosres.2013.10.003.

Ameer, S., Shah, M.A., Khan, A., Song, H., Maple, C., Islam, S.U., Asghar, M.N., 2019. Comparative analysis of machine learning techniques for predicting air quality in smart cities. IEEE Access 7, 128325–128338. https://doi.org/10.1109/ACCESS.2019.2925082.

Bekkar, A., Hssina, B., Douzi, S., Douzi, K., 2021. Air-pollution prediction in smart city, deep learning approach. J. Big Data 8, 161. https://doi.org/10.1186/s40537-021-00548-1.

Birmili, W., Weinhold, K., Rasch, F., Sonntag, A., Sun, J., Merkel, M., Wiedensohler, A., Bastian, S., Schladitz, A., Löschau, G., Cyrys, J., Pitz, M., Gu, J., Kusch, T., Flentje, H., Quass, U., Kaminski, H., Kuhlbusch, T.A.J., Meinhardt, F., Schwerin, A., Bath, O., Ries, L., Gerwig, H., Wirtz, K., Fiebig, M., 2016. Long-term observations of tropospheric particle number size distributions and equivalent black carbon mass concentrations in the German Ultrafine Aerosol Network (GUAN). Earth Syst. Sci. Data 355. https://doi.org/10.5194/essd-8-355-2016.

Blanco-Alegre, C., Calvo, A.I., Coz, E., Castro, A., Oduber, F., Prévôt, A.S.H., Močnik, G., Fraile, R., 2019. Quantification of source specific black carbon scavenging using an aethalometer and a disdrometer. Environ. Pollut. 246, 336–345. https://doi.org/10.1016/j.envpol.2018.11.102.

Bond, T.C., Doherty, S.J., Fahey, D.W., Forster, P.M., Berntsen, T., DeAngelo, B.J., Flanner, M.G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P.K., Sarofim, M.C., Schultz, M.G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S.K., Hopke, P.K., Jacobson, M.Z., Kaiser, J.W., Klimont, Z., Lohmann, U., Schwarz, J.P., Shindell, D., Storelvmo, T., Warren, S.G., Zender, C.S., 2013. Bounding the role of black carbon in the climate system: a scientific assessment. JGR Atmos. 118, 5380–5552. https://doi.org/10.1002/jgrd.50171.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1007/BF00058655.

Brewer, T.L., 2019. Black carbon emissions and regulatory policies in transportation. Energy Policy 129, 1047–1055. https://doi.org/10.1016/j.enpol.2019.02.073.

Briggs, N.L., Long, C.M., 2016. Critical review of black carbon and elemental carbon source apportionment in Europe and the United States. Atmos. Environ. 144, 409–427. https://doi.org/10.1016/j.atmosenv.2016.09.002.

Cabaneros, S.M., Calautit, J.K., Hughes, B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. Environ. Model. Softw. 119, 285–304. https://doi.org/10.1016/j.envsoft.2019.06.014.

Casquero-Vera, J.A., Lyamani, H., Titos, G., Borrás, E., Olmo, F.J., Alados-Arboledas, L., 2019. Impact of primary NO2 emissions at different urban sites exceeding the European NO2 standard limit. Sci. Total Environ. 646, 1117–1125. https://doi.org/10.1016/j.scitotenv.2018.07.360.

Collaud Coen, M., Weingartner, E., Apituley, A., Ceburnis, D., Fierz-Schmidhauser, R., Flentje, H., Henzing, J.S., Jennings, S.G., Moerman, M., Petzold, A., Schmid, O., Baltensperger, U., 2010. Minimizing light absorption measurement artifacts of the Aethalometer: evaluation of five correction algorithms. Atmos. Meas. Technol. 3, 457–474. https://doi.org/10.5194/amt-3-457-2010.

Collaud Coen, M., Andrews, E., Alastuey, A., Arsov, T.P., Backman, J., Brem, B.T., Bukowiecki, N., Couret, C., Eleftheriadis, K., Flentje, H., Fiebig, M., Gysel-Beer, M., Hand, J.L., Hoffer, A., Hooda, R., Hueglin, C., Joubert, W., Keywood, M., Kim, J.E., Kim, S.W., Labuschagne, C., Lin, N.H., Lin, Y., Lund Myhre, C., Luoma, K., Lyamani, H., Marinoni, A., Mayol-Bracero, O.L., Mihalopoulos, N., Pandolfi, M., Prats, N., Prenni, A.J., Putaud, J.P., Ries, L., Reisen, F., Sellegri, K., Sharma, S., Sheridan, P., Sherman, J.P., Sun, J., Titos, G., Torres, E., Tuch, T., Weller, R., Wiedensohler, A., Zieger, P., Laj, P., 2020. Multidecadal trend analysis of in situ aerosol radiative properties around the world. Atmos. Chem. Phys. 20, 8867–8908. https://doi.org/10.5194/acp-20-8867-2020.

Cuesta-Mosquera, A., Močnik, G., Drinovec, L., Müller, T., Pfeifer, S., Minguillón, M.C., Briel, B., Buckley, P., Dudoitis, V., Fernández-García, J., Fernández-Amado, M., Ferreira De Brito, J., Riffault, V., Flentje, H., Heffernan, E., Kalivitis, N., Kalogridis, A.C., Keernik, H., Marmureanu, L., Luoma, K., Marinoni, A., Pikridas, M., Schauer, G., Serfozo, N., Servomaa, H., Titos, G., Yus-Díez, J., Zioła, N., Wiedensohler, A., 2021. Intercomparison and characterization of 23 Aethalometers under laboratory and ambient air conditions: procedures and unit-to-unit variabilities. Atmos. Meas. Technol. 14, 3195–3216. https://doi.org/10.5194/amt-14-3195-2021.

Duan, J., Gong, Y., Luo, J., Zhao, Z., 2023. Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer. Sci. Rep. 13, 12127. https://doi.org/10.1038/s41598-023-36620-4.

Freeman, B.S., Taylor, G., Gharabaghi, B., Thé, J., 2018. Forecasting air quality time series using deep learning. J. Air Waste Manage. Assoc. 68, 866–886. https://doi.org/10.1080/10962247.2018.1459956.

Fung, P.L., Zaidan, M.A., Sillanpää, S., Kousa, A., Niemi, J.V., Timonen, H., Kuula, J., Saukko, E., Luoma, K., Petäjä, T., Tarkoma, S., Kulmala, M., Hussein, T., 2020. Input-adaptive proxy for black carbon as a virtual sensor. Sensors 20, 182. https://doi.org/10.3390/s20010182.

Fung, P.L., Zaidan, M.A., Surakhi, O., Tarkoma, S., Petäjä, T., Hussein, T., 2021a. Data imputation in in situ-measured particle size distributions by means of neural networks. Atmos. Meas. Technol. 14, 5535–5554. https://doi.org/10.5194/amt-14-5535-2021.

Fung, P.L., Zaidan, M.A., Timonen, H., Niemi, J.V., Kousa, A., Kuula, J., Luoma, K., Tarkoma, S., Petäjä, T., Kulmala, M., Hussein, T., 2021b. Evaluation of white-box versus black-box machine learning models in estimating ambient black carbon concentration. J. Aerosol Sci 152, 105694. https://doi.org/10.1016/j.jaerosci.2020.105694.

Fung, P.L., Al-Jaghbeer, O., Pirjola, L., Aaltonen, H., Järvi, L., 2023. Exploring the discrepancy between top-down and bottom-up approaches of fine spatio-temporal vehicular CO2 emission in an urban road network. Sci. Total Environ. 901, 165827 https://doi.org/10.1016/j.scitotenv.2023.165827.

Fung, P.L., Sillanpää, S., Niemi, J.V., Kousa, A., Timonen, H., Zaidan, M.A., Saukko, E., Kulmala, M., Petäjä, T., Hussein, T., 2022a. Improving the current air quality index with new particulate indicators using a robust statistical approach. Sci. Total Environ. 844, 157099 https://doi.org/10.1016/j.scitotenv.2022.157099.

Fung, P.L., Zaidan, M.A., Niemi, J.V., Saukko, E., Timonen, H., Kousa, A., Kuula, J., Rönkkö, T., Karppinen, A., Tarkoma, S., Kulmala, M., Petäjä, T., Hussein, T., 2022b. Input-adaptive linear mixed-effects model for estimating alveolar lung-deposited surface area (LDSA) using multipollutant datasets. Atmos. Chem. Phys. 22, 1861–1882. https://doi.org/10.5194/acp-22-1861-2022.

Grange, S.K., Lötscher, H., Fischer, A., Emmenegger, L., Hueglin, C., 2020. Evaluation of equivalent black carbon source apportionment using observations from Switzerland between 2008 and 2018. Atmos. Meas. Techol. 13, 1867–1885. https://doi.org/10.5194/amt-13-1867-2020.

Gu, J., Yang, B., Brauer, M., Zhang, K.M., 2021. Enhancing the evaluation and interpretability of data-driven air quality models. Atmos. Environ. 246, 118125 https://doi.org/10.1016/j.atmosenv.2020.118125.

Hilker, N., Wang, J.M., Jeong, C.H., Healy, R.M., Sofowote, U., Debosz, J., Su, Y., Noble, M., Munoz, A., Doerksen, G., White, L., Audette, C., Herod, D., Brook, J.R., Evans, G.J., 2019. Traffic-related air pollution near roadways: discerning local impacts from background. Atmos. Meas. Technol. 12, 5247–5261. https://doi.org/10.5194/amt-12-5247-2019.

Hitzenberger, R., Petzold, A., Bauer, H., Ctyroky, P., Pouresmaeil, P., Laskus, L., Puxbaum, H., 2006. Intercomparison of thermal and optical measurement methods for elemental carbon and black carbon at an urban location. Environ. Sci. Tech. 40, 6377–6383. https://doi.org/10.1021/es051228v.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Jafar, H.A., Harrison, R.M., 2021. Spatial and temporal trends in carbonaceous aerosols in the United Kingdom. Atmos. Pollut. Res. 12, 295–305. https://doi.org/10.1016/j.apr.2020.09.009.

Järvi, L., Hannuniemi, H., Hussein, T., Junninen, H., Aalto, P.P., Hillamo, R., Mäkelä, T., Keronen, P., Siivola, E., Vesala, T., 2009. The urban measurement station SMEAR III: Continuous monitoring of air pollution and surface–atmosphere interactions in Helsinki, Finland. Boreal Environ. Res. 14, 86–109.

Järvi, L., Kurppa, M., Kuuluvainen, H., Rönkkö, T., Karttunen, S., Balling, A., Timonen, H., Niemi, J.V., Pirjola, L., 2023. Determinants of spatial variability of air pollutant concentrations in a street canyon network measured using a mobile laboratory and a drone. Sci. Total Environ. 856, 158974 https://doi.org/10.1016/j.scitotenv.2022.158974.

Kaur, J., Singh, S., Parmar, K.S., Soni, K., 2023. Development of a mathematical model to forecast black carbon concentration using ARIMA and soft computing. Arab. J. Geosci. 16, 258. https://doi.org/10.1007/s12517-023-11321-4.

Leong, W.C., Kelani, R.O., Ahmad, Z., 2019. Prediction of air pollution index (API) using support vector machine (SVM). J. Environ. Chem. Eng. 8 (3), 103208 https://doi.org/10.1016/j.jece.2019.103208.

Liu, X., Concas, F., Motlagh, N.H., Zaidan, M.A., Fung, P.L., Varjonen, S., Niemi, J.V., Timonen, H., Hussein, T., Petäjä, T., 2023. Estimating Black Carbon Levels with Proxy Variables and Low-Cost Sensors. TechRxiv. https://doi.org/10.36227/techrxiv.24152931.v1.

Loh, W.-Y., 2002. Regression tress with unbiased variable selection and interaction detection. Statistica Sinica 361–386. https://www.jstor.org/stable/24306967.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017.

Luo, J., Li, Z., Qiu, J., Zhang, Y., Fan, C., Li, L., Wu, H., Zhou, P., Li, K., Zhang, Q., 2023. The Simulated Source Apportionment of Light Absorbing Aerosols: Effects of Microphysical Properties of Partially-Coated Black Carbon. JGR Atmospheres 128, e2022JD037291. https://doi.org/10.1029/2022JD037291.

Luo, J., Zhang, Y., Wang, F., Wang, J., Zhang, Q., 2018. Applying machine learning to estimate the optical properties of black carbon fractal aggregates. J. Quant. Spectrosc. Radiat. Transf. 215, 1–8. https://doi.org/10.1016/j.jqsrt.2018.05.002.

Luoma, K., Niemi, J.V., Aurela, M., Fung, P.L., Helin, A., Hussein, T., Kangas, L., Kousa, A., Rönkkö, T., Timonen, H., Virkkula, A., Petäjä, T., 2021a. Spatiotemporal

variation and trends in equivalent black carbon in the Helsinki metropolitan area in Finland. Atmos. Chem. Phys. 21, 1173–1189. https://doi.org/10.5194/acp-21-1173-2021.

Luoma, K., Virkkula, A., Aalto, P., Lehtipalo, K., Petäjä, T., Kulmala, M., 2021b. Effects of different correction algorithms on absorption coefficient – a comparison of three optical absorption photometers at a boreal forest site. Atmos. Meas. Technol. 14, 6419–6441. https://doi.org/10.5194/amt-14-6419-2021.

Makkhan, S.J.S., Singh, S., Parmar, K.S., Kaushal, S., Soni, K., 2023. Comparison of hybrid machine learning model for the analysis of black carbon in air around the major coal mines of India. Neural Comput. Appl. 35, 3449–3468. https://doi.org/10.1007/s00521-022-07909-8.

Mao, Y., Lee, S., 2019. Deep convolutional neural network for air quality prediction. J. Phys. Conf. Ser. 1302, 032046 https://doi.org/10.1088/1742-6596/1302/3/032046.

Massagué, J., Escudero, M., Alastuey, A., Mantilla, E., Monfort, E., Gangoiti, G., García-Pando, C.P., Querol, X., 2023. Spatiotemporal variations of tropospheric ozone in Spain (2008–2019). Environ. Int. 176, 107961 https://doi.org/10.1016/j.envint.2023.107961.

May, A.A., Li, H., 2022. Application of machine learning approaches in the analysis of mass absorption cross-section of black carbon aerosols: aerosol composition dependencies and sensitivity analyses. Aerosol Sci. Technol. 56, 998–1008. https://doi.org/10.1080/02786826.2022.2114312.

Méndez, M., Merayo, M.G., Núñez, M., 2023. Machine learning algorithms to forecast air quality: a survey. Artif. Intell. Rev. 56, 10031–10066. https://doi.org/10.1007/s10462-023-10424-4.

Moosmüller, H., Chakrabarty, R.K., Arnott, W.P., 2009. Aerosol light absorption and its measurement: a review. J. Quant. Spectrosc. Radiat. Transf. 110, 844–878. https://doi.org/10.1016/j.jqsrt.2009.02.035.

Müller, T., Henzing, J.S., de Leeuw, G., Wiedensohler, A., Alastuey, A., Angelov, H., Bizjak, M., Collaud Coen, M., Engström, J.E., Gruening, C., Hillamo, R., Hoffer, A., Imre, K., Ivanow, P., Jennings, G., Sun, J.Y., Kalivitis, N., Karlsson, H., Komppula, M., Laj, P., Li, S.M., Lunder, C., Marinoni, A., Martins dos Santos, S., Moerman, M., Nowak, A., Ogren, J.A., Petzold, A., Pichon, J.M., Rodriquez, S., Sharma, S., Sheridan, P.J., Teinilä, K., Tuch, T., Viana, M., Virkkula, A., Weingartner, E., Wilhelm, R., Wang, Y.Q., 2011. Characterization and intercomparison of aerosol absorption photometers: result of two intercomparison workshops, Atmospheric. Meas. Tech. 4, 245–268. https://doi.org/10.5194/amt-4-245-2011.

Patil, R., Bedekar, G., Tergundi, P., Goudar, R.H., 2022. An Efficient Implementation of ARIMA Technique for Air Quality Prediction. In: Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021. Springer Nature Singapore, Singapore, pp. 441–451.

Petzold, A., Schönlinner, M., 2004. Multi-angle absorption photometry—a new method for the measurement of aerosol light absorption and atmospheric black carbon. J. Aerosol. Sci. 35, 421–441. https://doi.org/10.1016/j.jaerosci.2003.09.005.

Petzold, A., Ogren, J.A., Fiebig, M., Laj, P., Li, S.M., Baltensperger, U., Holzer-Popp, T., Kinne, S., Pappalardo, G., Sugimoto, N., Wehrli, C., Wiedensohler, A., Zhang, X.Y., 2013. Recommendations for reporting "black carbon" measurements. Atmos. Chem. Phys. 13, 8365–8379. https://doi.org/10.5194/acp-13-8365-2013.

Qiu, M., Zigler, C., Selin, N.E., 2022. Statistical and machine learning methods for evaluating trends in air quality under changing meteorological conditions. Atmos. Chem. Phys. 22, 10551–10566. https://doi.org/10.5194/acp-22-10551-2022.

Rivas, I., Beddows, D.C., Amato, F., Green, D.C., Järvi, L., Hueglin, C., Reche, C., Timonen, H., Fuller, G.W., Niemi, J.V., Pérez, N., Aurela, M., Hopke, P.K., Alastuey, A., Kulmala, M., Harrison, R.M., Querol, X., Kelly, F.J., 2020. Source apportionment of particle number size distribution in urban background and traffic stations in four European cities. Environ. Int. 135, 105345 https://doi.org/10.1016/j.envint.2019.105345.

Rovira, J., Paredes-Ahumada, J.A., Barceló-Ordinas, J.M., García-Vidal, J., Reche, C., Sola, Y., Fung, P.L., Petäjä, T., Hussein, T., Viana, M., 2022. Non-linear models for black carbon exposure modelling using air pollution datasets. Environ. Res. 212, 113269 https://doi.org/10.1016/j.envres.2022.113269.

Rubio-Loyola, J., Paul-Fils, W.R.S., 2022. Applied Machine Learning in Industry 4.0: Case-Study Research in Predictive Models for Black Carbon Emissions. Sensors 22, 3947. https://doi.org/10.3390/s22103947.

Rudin, C., 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat. Mach. Intell. 1, 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Saarikoski, S., Niemi, J.V., Aurela, M., Pirjola, L., Kousa, A., Rönkkö, T., Timonen, H., 2021. Sources of black carbon at residential and traffic environments obtained by two source apportionment methods. Atmos. Chem. Phys. 21, 14851–14869. https://doi.org/10.5194/acp-21-14851-2021.

Savadkoohi, M., Pandolfi, M., Reche, C., Niemi, J.V., Mooibroek, D., Titos, G., Green, D. C., Tremper, A.H., Hueglin, C., Liakakou, E., Mihalopoulos, N., Stavroulas, I., Artiñano, B., Coz, E., Alados-Arboledas, L., Beddows, D., Riffault, V., De Brito, J.F., Bastian, S., Baudic, A., Colombi, C., Costabile, F., Chazeau, B., Marchand, N., Gómez-Amo, J.L., Estellés, V., Matos, V., van der Gaag, E., Gille, G., Luoma, K., Manninen, H.E., Norman, M., Silvergren, S., Petit, J.-E., Putaud, J.-P., Rattigan, O.V., Timonen, H., Tuch, T., Merkel, M., Weinhold, K., Vratolis, S., Vasilescu, J., Favez, O., Harrison, R.M., Laj, P., Wiedensohler, A., Hopke, P.K., Petäjä, T., Alastuey, A., Querol, X., 2023. The variability of mass concentrations and source apportionment analysis of equivalent black carbon across urban Europe. Environ. Int. 178, 108081 https://doi.org/10.1016/j.envint.2023.108081.

Sethi, J.K., Mittal, M., 2021. An efficient correlation based adaptive LASSO regression method for air quality index prediction. Earth Sci. Inf. 14, 1777–1786. https://doi.org/10.1007/s12145-021-00618-1.

Sun, J., Birmili, W., Hermann, M., Tuch, T., Weinhold, K., Spindler, G., Schladitz, A., Bastian, S., Löschau, G., Cyrys, J., Gu, J., Flentje, H., Briel, B., Asbach, C., Kaminski, H., Ries, L., Sohmer, R., Gerwig, H., Wirtz, K., Meinhardt, F., Schwerin, A., Bath, O., Ma, N., Wiedensohler, A., 2019. Variability of black carbon mass concentrations, sub-micrometer particle number concentrations and size distributions: results of the German Ultrafine Aerosol Network ranging from city street to High Alpine locations. Atmos. Environ. 202, 256–268. https://doi.org/10.1016/j.atmosenv.2018.12.029.

Sun, J., Hermann, M., Yuan, Y., Birmili, W., Collaud Coen, M., Weinhold, K., Madueño, L., Poulain, L., Tuch, T., Ries, L., Sohmer, R., Couret, C., Frank, G., Brem, B.T., Gysel-Beer, M., Ma, N., Wiedensohler, A., 2021. Long-term trends of black carbon and particle number concentration in the lower free troposphere in Central Europe. Environ. Sci. Eur. 33, 47. https://doi.org/10.1186/s12302-021-00488-w.

Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. J. r. Statist. Soc. Ser. B Statist. Methodol. 58, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Van Roode, S., Ruiz-Aguilar, J., González-Enrique, J., Turias, I., 2019. An artificial neural network ensemble approach to generate air pollution maps. Environ. Monit. Assess. 191, 727. https://doi.org/10.1007/s10661-019-7901-6.

Vapnik, V.N., 1997. The support vector method. In: International conference on artificial neural networks. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 261–271.

Via, M., Minguillón, M.C., Reche, C., Querol, X., Alastuey, A., 2021. Increase in secondary organic aerosol in an urban environment. Atmos. Chem. Phys. 21, 8323–8339. https://doi.org/10.5194/acp-21-8323-2021.

Wang, A., Xu, J., Tu, R., Saleh, M., Hatzopoulou, M., 2020. Potential of machine learning for prediction of traffic related air pollution. Transp. Res. Part D: Transp. Environ. 88, 102599 https://doi.org/10.1016/j.trd.2020.102599.

Weingartner, E., Saathoff, H., Schnaiter, M., Streit, N., Bitnar, B., Baltensperger, U., 2003. Absorption of light by soot particles: determination of the absorption coefficient by means of aethalometers. J. Aerosol Sci 34, 1445–1463. https://doi.org/10.1016/S0021-8502(03)00359-8.

WHO, 2021. WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization, 2021.

Xu, H., Ren, Y.A., Zhang, W., Meng, W., Yun, X., Yu, X., Li, J., Zhang, Y., Shen, G., Ma, J., 2021. Updated global black carbon emissions from 1960 to 2017: improvements, trends, and drivers. Environ. Sci. Technol. 55, 7869–7879. https://doi.org/10.1021/acs.est.1c03117.

Yu, W., Ye, T., Zhang, Y., Xu, R., Lei, Y., Chen, Z., Yang, Z., Zhang, Y., Song, J., Yue, X., Li, S., Guo, Y., 2023. Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study. Lancet Planetary Health 7, e209–e218. https://doi.org/10.1016/S2542-5196(23)00008-6.

Yus-Díez, J., Via, M., Alastuey, A., Karanasiou, A., Minguillón, M.C., Perez, N., Querol, X., Reche, C., Ivančič, M., Rigler, M., Pandolfi, M., 2022. Absorption enhancement of black carbon particles in a Mediterranean city and countryside: effect of particulate matter chemistry, ageing and trend analysis. Atmos. Chem. Phys. 22, 8439–8456. https://doi.org/10.5194/acp-22-8439-2022.

Zaidan, M.A., Wraith, D., Boor, B.E., Hussein, T., 2019. Bayesian proxy modelling for estimating black carbon concentrations using white-box and black-box models. Appl. Sci. 9, 4976. https://doi.org/10.3390/app9224976.

Zaidan, M.A., Motlagh, N.H., Fung, P.L., Lu, D., Timonen, H., Kuula, J., Niemi, J.V., Tarkoma, S., Petäjä, T., Kulmala, M., 2020. Intelligent calibration and virtual sensing for integrated low-cost air quality sensors. IEEE Sens. J. 20, 13638–13652. https://doi.org/10.1109/JSEN.2020.3010316.

Zaidan, M.A., Motlagh, N.H., Fung, P.L., Khalaf, A.S., Matsumi, Y., Ding, A., Tarkoma, S., Petäjä, T., Kulmala, M., Hussein, T., 2022. Intelligent air pollution sensors calibration for extreme events and drifts monitoring. IEEE Trans. Ind. Inf. 19, 1366–1379. https://doi.org/10.1109/TII.2022.3151782.

Zhang, Y., Wen, M., Sun, Y., Chen, H., Cai, Y., 2022. Black carbon emission prediction of diesel engine using stacked generalization. Atmos. 13, 1855. https://doi.org/10.3390/atmos13111855.

Zhu, J.-J., Chen, Y.-C., Shie, R.-H., Liu, Z.-S., Hsu, C.-Y., 2021. Predicting carbonaceous aerosols and identifying their source contribution with advanced approaches. Chemosphere 266, 128966. https://doi.org/10.1016/j.chemosphere.2020.128966.

Zhu, M., Xie, J., 2023. Investigation of nearby monitoring station for hourly PM2.5 forecasting using parallel multi-input 1D-CNN-biLSTM. Expert Syst. Appl. 211, 118707 https://doi.org/10.1016/j.eswa.2022.118707.