Environmental Research 228 (2023) 115835

Contents lists available at ScienceDirect

Environmental Research

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

Machine learning model to predict vehicle electrification impacts on urban air quality and related human health effects

V. Calatayud [a,*], J.J. Diéguez [a], E. Agathokleous [b], P. Sicard [c]

[a] Fundación CEAM, Parque Tecnológico, C/Charles R. Darwin, 14, Paterna, Spain
[b] Institute of Ecology, Key Laboratory of Agrometeorology of Jiangsu Province, School of Applied Meteorology, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[c] ARGANS, 260 Route Du Pin Montard, Biot, France

ARTICLE INFO

Keywords:
Electric car
Air pollution
Cities
XGBoost
SHAP
Health impact assessment

ABSTRACT

Air pollution is a prevailing environmental problem in cities worldwide. The future vehicle electrification (VE), which in Europe will be importantly fostered by the ban of thermal engines from 2035, is expected to have an important effect on urban air quality. Machine learning models represent an optimal tool for predicting changes in air pollutants concentrations in the context of future VE. For the city of Valencia (Spain), a XGBoost (eXtreme Gradient Boosting package) model was used in combination with SHAP (SHapley Additive exPlanations) analysis, both to investigate the importance of different factors explaining air pollution concentrations and predicting the effect of different levels of VE. The model was trained with 5 years of data including the COVID-19 lockdown period in 2020, in which mobility was strongly reduced resulting in unprecedent changes in air pollution concentrations. The interannual meteorological variability of 10 years was also considered in the analyses. For a 70% VE, the model predicted: 1) improvements in nitrogen dioxide pollution (−34% to −55% change in annual mean concentrations, for the different air quality stations), 2) a very limited effect on particulate matter concentrations (−1 to −4% change in annual means of $PM_{2.5}$ and PM_{10}), 3) heterogeneous responses in ground-level ozone concentrations (−2% to +12% change in the annual means of the daily maximum 8-h average concentrations). Even at a high VE increase of 70%, the 2021 World Health Organization Air Quality Guidelines will be exceeded for all pollutants in some stations. VE has a potentially important impact in terms of reducing NO_2-associated premature mortality, but complementary strategies for reducing traffic and controlling all different air pollution sources should also be implemented to protect human health.

1. Introduction

Air pollution is a global environmental issue, especially for the urban population (Sicard et al., 2023). In 2016, 54% of the world's population lived in urban settlements, and these areas will host 68% of the population by 2050 (United Nations, 2019). Nitrogen dioxide (NO_2), particulate matter (particles with aerodynamic diameter <10 μm, PM_{10}, and <2.5 μm, $PM_{2.5}$) and ground-level ozone (O_3) produce harmful effects to human health and are associated with respiratory and cardiovascular diseases, cancer, and premature mortality (Anenberg, 2018; Liu et al., 2019; Sicard et al., 2021; Malashock et al., 2022; Keswani et al., 2022). Because of exposure to outdoor air pollution, 4.2 million premature deaths occur worldwide (World Health Organisation, 2021a) and, in the 28 countries of the European Union (EU-28) in 2018, premature deaths attributed to $PM_{2.5}$ exposure, NO_2 and O_3 were 379,000, 54,000 and 19, 400, respectively (European Environmental Agency, 2020). Focusing on cities, in 2018, the percentages European urban population exposed to exceedances of the World Health Organization 2005 Air Quality Guidelines values were 74% for $PM_{2.5}$, 48% for PM_{10}, 99% for O_3, and 4% for NO_2 (European Environmental Agency, 2020). Furthermore, people exposed to moderately above-average levels of PM in the two years before the COVID-19 pandemic were 51% more likely to suffer severe COVID-19, while for NO_2 the risk increased by 26% (Kogevinas et al., 2021).

To protect human health from air pollutants, ambient air quality standards and emission control policies have been developed worldwide. In Europe, reference limit values were established for $PM_{2.5}$, PM_{10}, and NO_2, and target values for O_3 (Air Quality Directive, 2008/50/EC,

* Corresponding author.
E-mail address: vicent@ceam.es (V. Calatayud).

https://doi.org/10.1016/j.envres.2023.115835
Received 6 March 2023; Received in revised form 31 March 2023; Accepted 1 April 2023
Available online 4 April 2023

hereafter AQD) (Table 1). Based on epidemiological studies on the impacts of air pollutants on human heath, the WHO established Air Quality Guidelines (hereafter WHO AQG) in 2005, recently revised in 2021 (World Health Organization, 2021b) (Table 1). These WHO AQG are more stringent than limits of the AQD and represent a fundamental reference for future revisions of the AQD. Compliance of the AQD limits and target values is based on measurements by reference measuring methods in networks of air quality stations. Exceedance of the AQD limits and target values in any monitoring station within established air quality zones, implies non-compliance with the legislation in the entire zone (AQD, Borge et al., 2022). In this case, air quality improvement plans have to be implemented.

In 2021, the global electric car sales surpassed 6.6 million, tripling the sales from two years earlier, with 16 million electric cars being on the road worldwide (Venter, 2022). In the framework of the EU strategy to achieve climate neutrality by 2050 and energy independency, vehicle electrification (VE) will be strongly promoted, and the sale of new petrol and diesel cars will be banned from 2035 (2021/0197(COD)). In Europe, electric car sales accounted for 17% of the total sales in 2021 (about half of them plug-in hybrids), with Norway and Sweden leading the market (72% and 45% of the sales, respectively); 25% of the sales are in Germany, about 15% in France, 9% in Italy and 7% in Spain (Paoli and Gül, 2022). Increasing VE is expected to have a significant impact on urban air quality, and therefore the prediction of its effects is a topic of great interest to support the development of strategies for the reduction of air pollution and compliance with air quality standards in urban environments.

Atmospheric chemical transport models (CTM) have been commonly applied to investigate the effect of changes of emissions and meteorological conditions on air pollutant concentrations (e.g., Lamarque et al., 2013; Young et al., 2013; De Marco et al., 2022). However, they require a lot of computational power and are subject to large uncertainties related to the missing or poorly parameterized physical and chemical processes, and also to the emission inventories, which, frequently, are inaccurate, incomplete or not updated (Liu et al., 2018). Source-receptor relationships have also been used to estimate the effect of changes in precursor emissions on pollutant concentrations. They mimic the behavior of a full CTM through a simplified statistical approximation and are much more flexible and faster than CTMs (Pisoni et al., 2019). However, depending on the purpose, their spatial resolution (typically $\geq$7 km^2 as in CTMs) may be insufficient for a fine city scale study, and prediction of non-linear relationships between precursors and concentrations as required for O_3 is problematic (Pisoni et al., 2019). In recent years, machine learning (ML) models have been revealed to be a powerful tool for data analysis and prediction (Jordan and Mitchell, 2015; Zhou, 2021), owing their flexibility for making use of the available data and computational efficiency (Jordan and Mitchell, 2015). In the field of urban air quality, ML models have been recently applied e.g., to investigate the importance of different predictive variables determining air pollutant concentrations (Ma et al., 2020), to predict changes in air quality under future scenarios (Yang et al., 2021) or to cancelling out the effect of meteorology on pollution concentrations during the COVID-19 lockdown and subsequent post-relaxation period (Querol et al., 2021). Data from the recent COVID-19 lockdown period, with unprecedented traffic reductions in many cities of the world (Aloi et al., 2020, and references therein) and associated changes in air pollution, are optimal for training ML models to investigate the impact of traffic reductions on air pollution (e.g., Lovric et al., 2021). In a recent study from Los Angeles, data including the lockdown period were also used to train a Random Forest model to predict changes in air quality associated with different scenarios of VE (Yang et al., 2021). This approach, that to our knowledge has only been applied in that study, is of great interest in the context of compliance of air quality targets by the cities. The XGBoost (eXtreme Gradient Boosting) system for tree boosting, which provides state-of-the-art results on many ML challenges (Chen and Guestrin, 2016), in combination with SHAP (SHapley Additive exPlanations) analysis, which helps interpreting predictions of complex models (Lundberg and Lee, 2017), are potential optimal tools for such a type of study.

In the present paper, we applied XGBoost model using the city of Valencia (Spain) as a study case. The objectives were: 1) to apply and assess the performance of XGBoost model for predicting air pollution levels; 2) to use this ML model to assess the importance of the different factors explaining air pollutant levels by using the SHAP analysis; 3) to predict the expected changes in air quality in city in a future scenario with increasing VE, considering the interannual meteorological variability; 4) to estimate the increase in VE needed to fulfill the air quality standards, at the level of air quality stations; 5) to predict the impact of increasing VE on human health at city level; and 6) to propose measures to effectively reduce air pollutants levels considering the predicted changes.

We hypothesized that air pollutant concentrations in the city of Valencia can be estimated at level of single air quality station by a ML-based model and that the effect of reduced emissions due to VE on air pollution can be predicted for a ten-year range of different meteorological conditions.

Table 1
Air Quality Limits (NO_2 and PM) and Target Values (O_3), according to the EU Air Quality Directive (Directive, 2008/50/EC) and WHO Air Quality Guidelines for the protection of human health (2005, updated in 2021). MDA8 is the daily maximum 8-h average of O_3 concentrations. The number of days per year (d y^{-1}) over the thresholds that should not be exceeded to fulfil the EU Air Quality Directive and WHO AQG are given in parenthesis, as well as the percentiles representing these numbers of days per year, i.e., P99.0 (3 days per year), P93.2 (25 days per year), and P90.4 (35 days per year).

Air pollutant	Averaging period and metric	AQD limit and target values	WHO AQG 2005	WHO AQG 2021
		(μg m^{-3})	(μg m^{-3})	(μg m^{-3})
NO_2	Annual mean	**40**	**40**	**10**
O_3	24 h MDA8	**120** (25 d y^{-1}, P93.2)	**100** (3 d y^{-1}, P99.0)	**100** (3 d y^{-1}, P99.0)
$PM_{2.5}$	Annual mean	**25**	**10**	**5**
$PM_{2.5}$	24 h mean	–	**25** (3 d y^{-1}, P99.0)	**15** (3 d y^{-1}, P99.0)
PM_{10}	Annual mean	**40**	**20**	**15**
PM_{10}	24 h mean	**50** (35 d y^{-1}, P90.4)	**50** (35 d y^{-1}, P90.4)	**50** (3 d y^{-1}, P99.0)

2. Materials and methods

2.1. Datasets used

Two datasets were used, one for training and testing the performance of the model and the other for prediction. The first dataset included 5 years (2017–2021) of hourly air quality and meteorological data (Table S1). Data were obtained from seven monitoring stations of the Valencian Air Quality Network (http://www.agroambient.gva.es) (Table 2, Fig. 1). The main source of air pollutants is the traffic, while other sources include the harbor and a limited number of industries around the city; the airport is downwind of the city, being less relevant in the context of this study. The air pollutants considered were NO_2, O_3, $PM_{2.5}$ and PM_{10}. For O_3, running 8-h average concentrations (i.e., actual hour + previous 7 h, from now on 8-h O_3) were used instead of hourly values, as they are the basis for O_3 MDA8 calculation (the daily maximum 8-h average of O_3 concentrations). Analyses of the changes in air pollution during the lockdown and subsequent relaxation period in Valencia are available elsewhere (Sicard et al., 2020a; Querol et al., 2021). The days with Saharan dust intrusion, European intrusion or biomass burning were also included (downloaded from http://www.agroambient.gva.es). Saharan dust transported to the Mediterranean region are relatively frequent in Spain and may cause exceedances of the

Table 2
Air quality monitoring stations of the city of Valencia. Latitude and longitude in decimal degrees.

Station	Typology 1	Typology 2	Latitude	Longitude	Pollutants
Vivers	Urban	Background	39.4781	−0.3683	NO_2, O_3
Molí	Suburban	Traffic	39.4811	−0.4083	NO_2, PM, O_3
Politècnic	Suburban	Background	39.4797	−0.3375	NO_2, PM, O_3
Bulevard Sud	Urban	Traffic	39.4503	−0.3964	NO_2, O_3
Centre	Urban	Traffic	39.4707	−0.3765	NO_2, PM
Pista de Silla	Urban	Traffic	39.4581	−0.3767	NO_2, PM, O_3
Avd. de França	Urban	Traffic	39.4575	−0.3428	NO_2, PM, O_3

daily PM_{10} standards when added to the local and regional PM (Querol et al., 2019). Long-range PM transport from continental Europe is relevant for Northern Spain (Escudero et al., 2007), but of scarce importance for Valencia (6 days for the period 2012–2021). Biomass burning impacts to this city are mainly associated to agricultural practices, being particularly remarkable the period of the rice straw burning from October to December (Viana et al., 2008). The meteorological variables from Avd. de França station were used for the whole city: air temperature, relative humidity, wind speed, maximum wind speed, wind direction, total precipitation, and solar radiation, on an hourly basis. The boundary layer height was obtained for each hour from ERA5 reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5). This reanalysis provides hourly data with a spatial resolution of 30 km.

Other variables included in the dataset were the Average Hourly Traffic (AHT) values of streets within radii of 1000 m (AHT_1000) and 500 m (AHT_500) around the stations (Table S1). Hourly data from a total of 415 electromagnetic spires, used to measure traffic intensities in the streets, were used in the study (data provided by the Valencia City Hall upon request). As a reference, the total vehicle park of the city is 473,260 (Recull Estadístic València, 2021). Furthermore, two temporal indicators were included, the day of week and the hour of the day in local time. Also, the distance of each air quality stations to the harbor, the main pollutants source in Valencia after traffic, was included in the analyses (following Yang et al., 2021). In 2020, the harbor of Valencia welcomed 5539 ships and 416,228 passengers and acted as a hub for the transportation of 74, 584, 893 tons of goods (Recull Estadístic València, 2021). Finally, the stations themselves were included as predictor variables for the model. As XGBoost does not accept categorical variables, the column with the station names was transformed to seven "dummy" columns (Kaplan, 2020), one per station, with 0 and 1 values.

The second dataset was built for the XGBoost model to predict the concentrations of the four pollutants under scenarios of progressive increase of VE (through equivalent reductions in AHTs as described in 2.2). These predictions were carried out for the meteorological conditions of 10 different years (2012–2021), as air pollutant concentrations are affected not only by traffic but also by meteorological conditions (e.g., Acosta-Ramírez and Higham, 2022; Nguyen et al., 2022). In this way, the interannual variability in pollution levels due to meteorological

Fig. 1. Location of the air quality monitoring stations in Valencia. Suburban stations indicated by a paler box color than urban stations.

conditions was taken into account in the predictions. Data from intrusion and biomass burning variables of these 10 years were also included in this second dataset.

2.2. XGBoost model development and predictions

For the development of XGBoost models, all calculations were performed with Package "xgboost" for R (Chen et al., 2019). XGBoost is a supervised ML system for tree boosting, which is scalable, non-linear and capture deep interactions, being less prone to outliers (Chen and Guestrin, 2016; Nielsen, 2016). The model was trained with training dataset (75% of the hourly 2017–2021 data, randomly selected), and its performance for each air pollutant was evaluated by comparing observed and predicted data in an independent test dataset (the remaining 25%) by means of the root mean squared error (RMSE) and the coefficient of determination (R^2) (Fig. 2). Hyperparameters were tuned using a 5-fold cross-validation scheme and "train" function of "caret" package (Kuhn, 2021). The hyperparameter settings which gave the best estimated generalization were subsequently selected, further controlling for the absence of overfitting (Nielsen, 2016). A Tweedie regression with log-link was used to avoid model outcomes with unrealistic negative values for pollutant concentrations.

Model predictions of the changes in pollutant concentrations with increasing VE were estimated indirectly by changing AHT in the model input (similar to Yang et al., 2021), considering traffic flow of 2021 as the baseline. Traffic load in 2021 represents the new traffic conditions in the city after the lockdown year 2020, about 10% lower than those of pre-pandemic years (Fig. S1). For this study we assumed a simple scenario in which the percentages of reductions in AHT occurred homogeneously in the whole city and that vehicle travels inside the city were

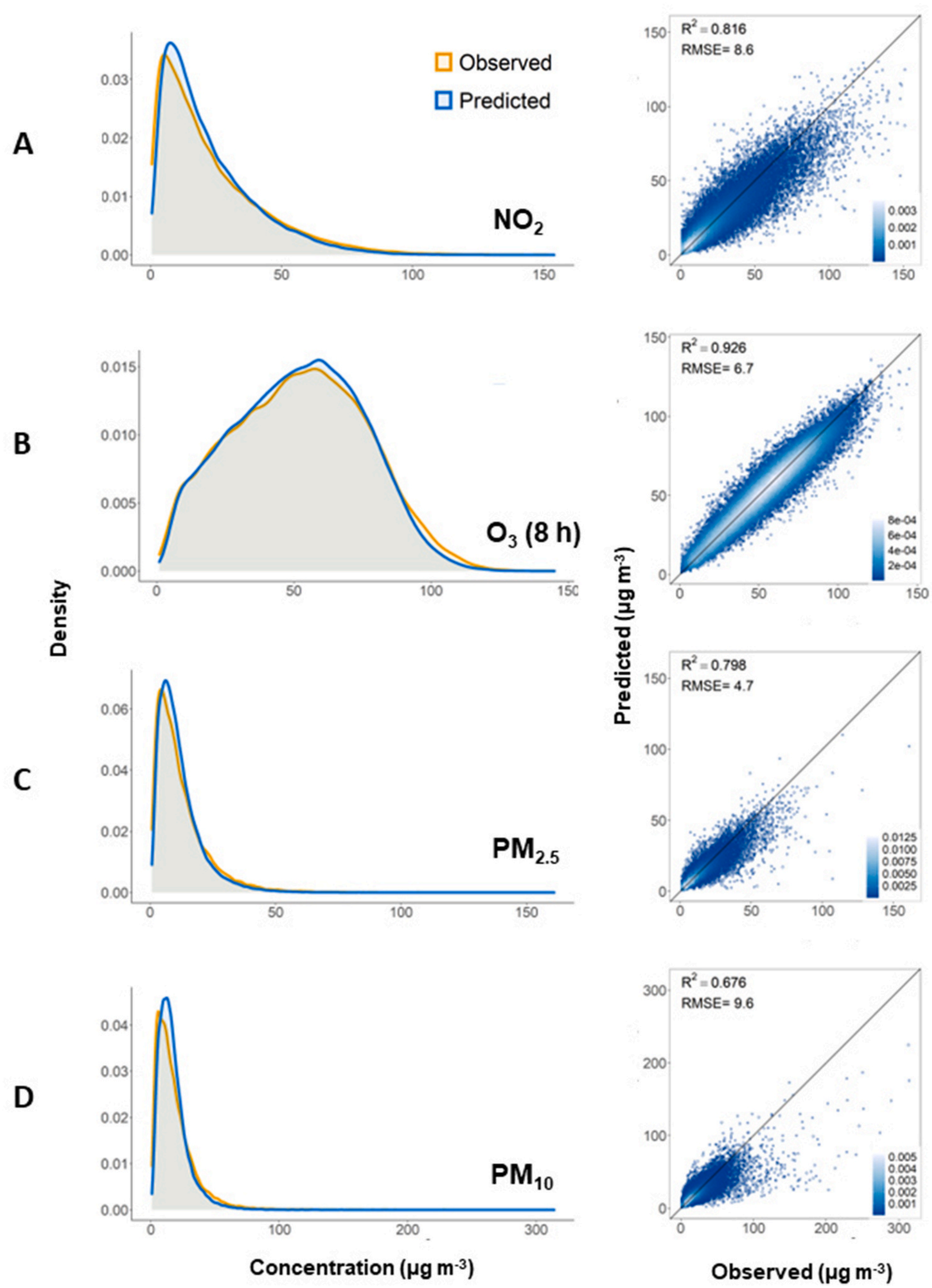

Fig. 2. Kernel density estimates (left) and regressions comparing observed and model-predicted values (right) for the four pollutants: A) NO_2, B) O_3, C) $PM_{2.5}$ and D) PM_{10}. Points represent hourly values (NO_2, $PM_{2.5}$ and PM_{10}) or running 8-h averages (O_3).

carried out in full electric mode. However, while electric vehicles do not emit NOx, this is not the case of PM, as about 90% of the PM_{10} and 85% of the $PM_{2.5}$ emissions of the vehicles are non-exhaust ones, i.e., from tire, brake, road wears and re-suspended dust (Timmers and Achten, 2016). Therefore, in terms of NO_2, PM_{10} and $PM_{2.5}$ emissions, a *x*% increase in VE was considered to be equivalent to AHT percent reductions of *x*%, *x*% *0.1 and *x*% *0.15, respectively. The effect of increasing VE from 0 to 70% in steps of 5% was then estimated by the model through equivalent changes in the AHT. A 70% value was chosen as the maximum because the ML model can only learn from situations included in the training dataset, and days with traffic below this value were very scarce in the dataset (Fig. S1). To also account for interannual meteorological variability in air pollutant concentration predictions, for each of the mentioned steps, the model was run 10 times, using the meteorological data of the 10 different years (2012–2021). In these predictions, we refer to "current situation" as the outputs of the model for a 0% increase in electrification (Fig. 4). At present, the number of electric and hybrid vehicles in the Valencian Community is still low (<2%).

2.3. SHAP and smoothed density graphs

SHAP analysis is based on Shaply values from game theory and provides the importance of each feature (a ML concept related to explanatory variable), considering their marginal contribution to the model outcome (Lundberg and Lee, 2017). Representation of the SHAP contribution dependency summary plots provide comprehensive visual information for comparing the SHAP contribution of different features to the model prediction, and also to identify, for this prediction, if the features have a positive or negative contribution and whether it differs for larger or smaller values of the feature (Chen et al., 2022) (Fig. 3).

Smoothed Kernel density estimates were also graphically represented to compare the distribution of measured and modeled concentrations of air pollutants. This type of graph is a smoothed alternative to the histogram for continuous data (Fig. 2).

2.4. Changes in pollution-related premature mortality due to vehicle electrification

Calculation of the pollution-related premature deaths for different levels of VE was carried out using the impact assessment module of the AirQ + software tool (World Health Organization, 2020). Heath endpoints, incidence, relative risk, and cut-off values used are provided in Table S2. This approach is widely used for health risk assessment worldwide (e.g., Amoatey et al., 2020; Xu et al., 2022; Cakaj et al., 2022). The analysis was also repeated for different levels of traffic reduction, because, for PM, it was considered interesting to compare these results with those of VE. Calculations were carried out at city level.

3. Results

3.1. Model performance

Overall, the XGBoost model was able to predict 8-h O_3 average concentrations very accurately, with an $R^2 = 0.926$ and RMSE = 6.7 μg m^{-3} (Fig. 2B). Prediction for NO_2 hourly concentrations was also good, with $R^2 = 0.816$ and RMSE = 8.6 μg m^{-3} (Fig. 2A). For $PM_{2.5}$, the accuracy of the model was slightly lower ($R^2 = 0.798$ and RMSE = 4.7 μg m^{-3}), especially underestimating higher values, some of them over 100 μg m^{-3} (Fig. 2C). The lowest model performance was that obtained for PM_{10}, with $R^2 = 0.676$ and RMSE = 9.6 μg m^{-3} (Fig. 2D).

Kernel density estimates were also represented graphically for comparing the distribution of pollutant concentrations in observed (measured) and predicted (modeled) values of the test database. The results showed an overall good fitting of the distributions of the predicted values against the reference observed ones (Fig. 2). The shape of the density estimates distribution varied among the different pollutants. NO_2 and PM distributions were right skewed, with an asymmetric pattern and a long tail on the right side (scarce events of high pollutant values). In contrast, that of the daily maximum 8-h concentration was more symmetrical.

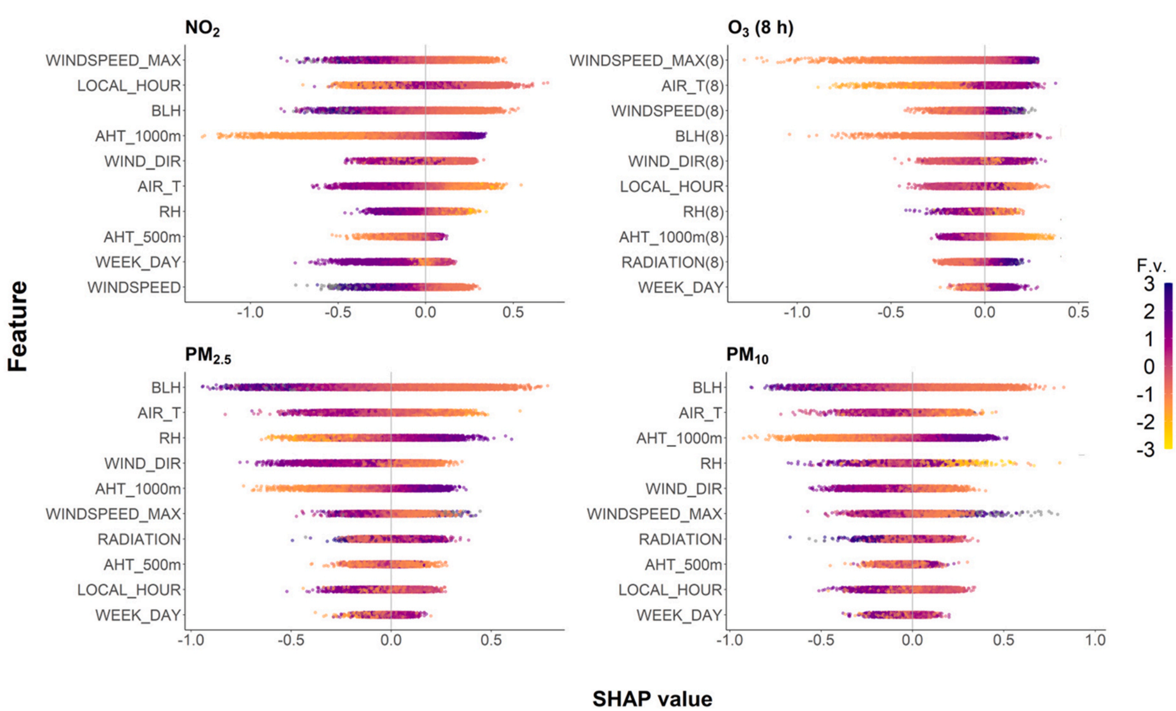

Fig. 3. SHAP analysis of the features used in the XGBoost model. F. v. = Feature value (low to high). Feature descriptions are provided in Table S1.

3.2. Feature importance

The impact of each feature in the model output was examined by means of SHAP graphs (Fig. 3 and, more detailed, Fig. S2). Traffic flows (AHTs within 1000 m and 500 m radii) were among the most important features explaining local air pollutant concentrations. An increase in traffic increased NO_2 and PM but decreased O_3. Lower boundary layer heights and lower temperatures, especially <500 m, as it is typical of wintertime conditions, were associated with higher NO_2 and PM concentrations, but the contrary was observed for O_3. PM concentrations also increased for temperatures over 30 °C, in the summertime. An increase in solar radiation was especially reflected in rising O_3 concentrations. Wind speed, especially the maximum hourly values, were also among the most important features, with lower NO_2 concentrations when wind speed was higher (over 10 m s^{-1}), and reductions for PM until 5 m s^{-1} but with an increase again for higher wind speeds. Wind directions from east to southwest, typical from daylight hours including the rush hours, consistently showed higher NO_2 and PM concentrations. Higher humidity values were related to lower NO_2 concentrations but to higher $PM_{2.5}$ concentrations. On the contrary, lower humidity values contributed to increase PM_{10}. Daily and weekly temporal patterns were also observed, with higher concentrations during the morning and somewhat less in the weekends for NO_2 and PM. For O_3, decreases during the night were reflected in a decrease of the 8-h values until midday, and an increase in the 8-h values during the weekend was also observed.

The model also included as variables the 7 air quality stations, the distance to the harbor and the days with Saharan dust intrusion, European intrusion or biomass burning. For the 5 target years, the number of days with Saharan intrusions ranged between 79 and 111, the days of biomass burning were 51 in 2017, but between 11 and 18 in 2018–2021, and European intrusions were rare, only 6 in 2020. Overall, 32% of the PM_{10} values over 100 μg m^{-3} were associated with intrusion or biomass burning events, while this percentage was lower for $PM_{2.5}$ >100 μg m^{-3}, only 9%. Neither of these variables were ranked among the most important ones to explain the observed hourly changes in pollutant concentrations.

3.3. Effect of vehicle electrification

The XGBoost model estimated the effect of VE up to a 70%. The results are represented in Fig. 4, including the reference AQD and WHO AQG thresholds (Table 1). Note that graphs in percentiles provide information on the total number of days per year with exceedances of AQD and WHO AQG thresholds (Fig. 4, Table 1).

For NO_2, the current situation (0% Change, Fig. 4) in Valencia meets the requirements of the EU Directive and WHO 2005, which are coincident, but the World Health Organization, 2020 is far exceeded (Fig. 4A). The model showed a close to linear decline in NO_2 concentrations with increasing VE for most of the stations. With a 70% VE, NO_2 annual means would be reduced by 34–55% for the different stations, but only the 2 suburban stations (Molí and Politècnic) would show annual means < 10 μg m^{-3} and only in part of the target years (Fig. 4A).

For the O_3 MDA8, the AQD is currently fulfilled but not the WHO 2025 and 2021 AQG (Fig. 4B and C). With increasing VE, the 3 urban traffic stations equipped with O_3 analyzers showed an increase (Avd. França) in the O_3 MDA8 values (P93.2 and P99.0), or an initial increase but a stabilization (P. Silla) or decrease (Blv. Sud) when VE>40% (Fig. 4B and C). The urban background station Vivers, set in a large public urban garden, showed a slight reduction due to VE, while the suburban background station Politècnic, located in a garden in the periphery of the city, exhibited a marked decline in O_3 MDA8 values (Fig. 4B and C). Overall, the model predicted that VE is expected to produce small and heterogeneous changes in O_3 concentrations in the city, in the range of −2%–12% in O_3 MDA8 annual means for the different stations at a 70% VE. Even with this high level of VE, the WHO 2025 and 2021 AQG would not be met.

Particulate matter was measured at 5 stations. All stations currently exhibit values below the AQD limits for $PM_{2.5}$ (Fig. 4D) and PM_{10} (Fig. 4F). If the WHO AQG is considered instead, for $PM_{2.5}$, all stations exceed the annual and 24 h WHO AQG 2021 (Fig. 4D and E) and only part of the stations fulfill the WHO AQG 2005 (Fig. 4E). For PM_{10}, WHO AQG 2005 AQG are fulfilled but the WHO AQG 2021 are exceeded in several stations (Fig. 4F and H). VE is expected to reduce PM concentrations very little, and not enough to be in compliance with the WHO AQG 2021 even at a 70% share. Predicted reductions for the different stations were in the ranges of 2%–4% and 1%–4% for $PM_{2.5}$ and PM_{10} annual means, respectively.

3.4. Impact vehicle electrification on premature mortality

Estimates of changes in premature mortality due to VE at city level are provided in Fig. 5. As a reference, mortality by natural causes in adults in Valencia was 7315 persons (7271 in older than 30 years) in 2018. As the model used does not allow a multipollutant effect estimation, the results are given for each pollutant separately. For the current situation (0% Change, Fig. 5), premature mortality was the highest for NO_2 (320 [95% CI 150 to 493] persons), followed by that of $PM_{2.5}$ (220 [109 to 333] persons), while for O_3 and PM_{10} it was much lower (31 [15 to 46] and 12 [5 to 20] persons, respectively). VE was predicted to produce an important reduction in NO_2-associated premature mortality. As a reference, a shorter-term objective of a 20% VE would result in 65 (−20%) [31 to 101] less premature deaths, while a 50% increase in VE would reduce associated premature mortality to more than half (reduction of 179 (−56%) [84 to 274] persons). For the other pollutants, however, VE would have a much smaller effect on premature mortality: a 50% VE would increase the number of cases with 2 (+6%) [1 to 3] more for O_3 and would reduce them with 8 (−62%) [3 to 12] for PM_{10} and 9 (−4%) [5 to 14] less cases for $PM_{2.5}$.

Comparison of the results of increasing VE with the effects of equivalent percent reductions in traffic are also provided in Fig. 5. It was assumed that for NO_2 and O_3, the effect would be equivalent, but for PM traffic reduction would be much more beneficial than VE in terms of human health. A 10% traffic reduction would eliminate all PM_{10}-associated mortality in the city (32 cases). For $PM_{2.5}$, 26 (−12%) [13 to 39] and 66 (−30%) [33 to 99] less persons would die prematurely with traffic reductions of 20% and 50%, respectively.

4. Discussion

4.1. Model performance

The study revealed how a ML model followed by a SHAP analysis can be used for investigating the importance and effect of different features (traffic, meteorological variables, air quality stations, and distance to sources of air pollutants) in determining the hourly air pollutant concentrations in a city. This approach overcomes the serious limitations of linear models in establishing relationships between predictive and response variables in temporal series of autocorrelated data.

The study clearly shows that XGBoost model can accurately predict air pollutant concentrations in the city of Valencia, with $R^2 > 0.80$ for NO_2 and $R^2 > 0.90$ for 8-h O_3 average concentrations. The observed lower accuracy for $PM_{2.5}$ ($R^2 = 0.80$) and PM_{10} ($R^2 = 0.68$), especially for the highest values may be related to several issues. On one hand, events with high PM were scarce and more difficult to be well captured during the training of the model. On the other hand, it may be indicative that there were sources of PM not considered in the model (e.g., works in buildings or streets near the stations), responsible for part of the higher PM episodes. Finally, as mentioned above, intrusion and biomass burning days were included in the model, but these data were not available as hourly values. The accuracy of XGBoost is critical when the objective is to investigate changes in air pollution in relation to

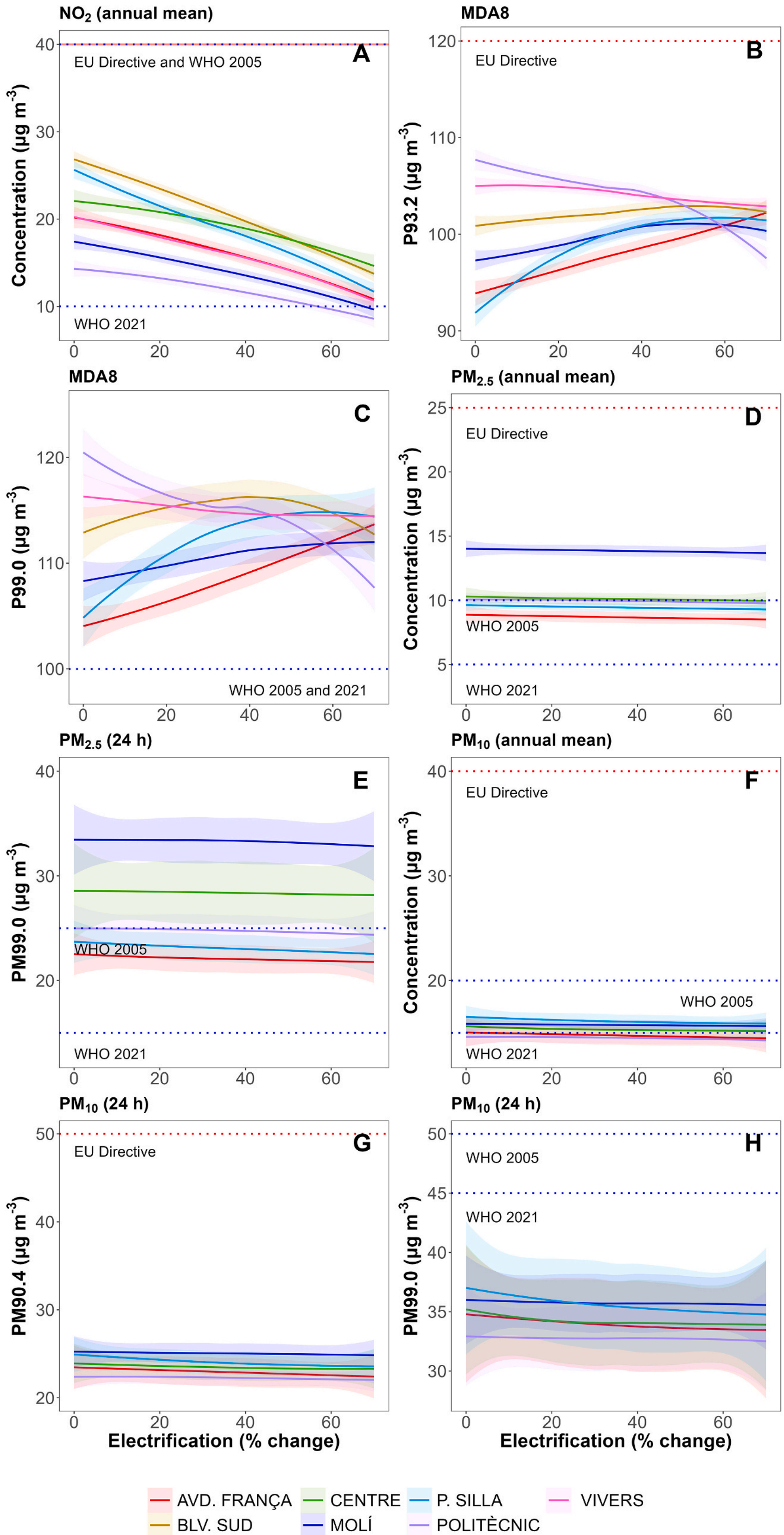

Fig. 4. XGBoost estimations of the effect of electrification in air pollutant concentrations in the air quality stations of the city of Valencia in relation to EU Air Quality Directive limit (NO_2, PM) and target (O_3) values (A, B, D, F and G) and WHO 2005 and 2021 Air Quality Guidelines (A, C, D, E, F and H). The effect of the interannual variability in meteorological conditions between 2012 and 2021 is taken into account by the confidence intervals. See Table 1 and the text for further information.

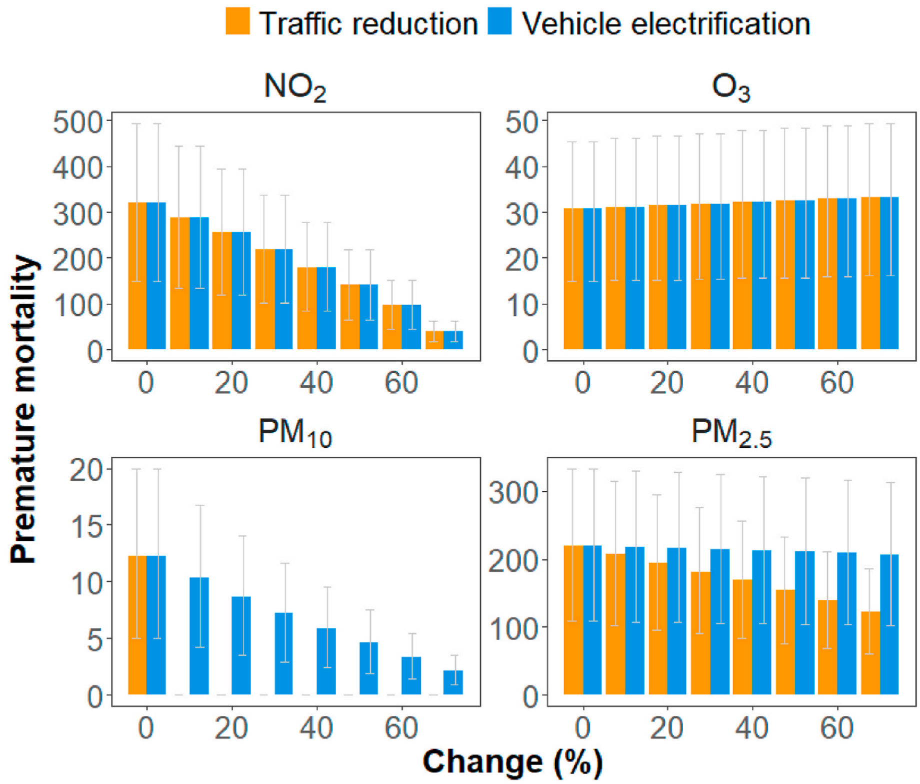

Fig. 5. Comparison of the changes in estimated premature mortality (Means ± 95% Coefficient Intervals) in the city of Valencia for scenarios of increasing traffic reduction and vehicle electrification.

compliance of air quality standards. It should be noted that the model could be further trained when more data are available in the future or when the pollution situations change, improving predictability.

4.2. Importance of predictor variables

As expected, the traffic load (AHT) was among the most important features explaining daily, weekly, and seasonal changes in air pollutant concentrations in Valencia. Air pollutant concentrations were better predicted by considering both local traffic flows (within a 500 m radius), with those within a larger area (1000 m radius). Both scales are complementary and necessary given that, depending on the pollutant considered and station type (background or traffic), the representative area of an air quality station varies (Kracht et al., 2017). Our study show that lower traffic resulted in lower NO_2 and PM concentrations, but also in a tendency for the O_3 to increase. These changes are overall consistent with the patterns observed during the weekends in many cities of the world and also during the COVID-19 lockdown period, when traffic was strongly reduced (e.g., Sicard et al., 2020a,b; Grange et al., 2021; Zhang et al., 2022).

The boundary layer height and temperature were also identified among the most important variables for the models of the different air pollutants. With lower boundary layer heights, as it is typical of anticyclonic and cold conditions of winter, NO_2 and PM were confined in a smaller air volume, and thus their concentrations increased. For O_3, despite the dilution effect caused by a thicker boundary layer in spring and summer, the more active photochemical formation of secondary pollutants in combination with the coastal to inland recirculation of the O_3-rich polluted air masses (Millán et al., 1997, 2000; Massagué et al., 2019), resulted in a seasonal relative increase of O_3 concentrations. Higher temperatures in spring and summer also know to increase plant emission of Biogenic Volatile Organic Compounds (BVOCs), which are O_3 precursors, fostering the production of this pollutant (Monks et al., 2009; Feng et al., 2019). An increase in PM concentrations for temperatures over 30 °C, in the summertime, could also be partly explained by secondary photochemical processes (e.g., Amato et al., 2016).

Wind speed and direction were also identified among the most important features. The SHAP analyses showed how high gusts of wind (hourly maximum wind speed) were particularly effective in advecting city pollution but at the same time may increase PM resuspension (as observed above 10 m s^{-1}). The dominant southeast winds, typical of daily hours when traffic is more intense, were in general associated with higher air pollutant concentrations.

Relative humidity was another important feature. NO_2 decrease with increasing relative humidity, especially over 60%, which can be explained because water react with NO_2 to form HONO and HNO_3 (Goodman et al., 1999; Valuntaitė et al., 2012). A decrease in O_3 was also observed with higher RH: high RH may be associated to cloudy days, less appropriate for O_3 formation, but it is also known that surface wetness increases O_3 deposition (Lamaud et al., 2002). The most important effect of relative humidity occurred with $PM_{2.5}$, which increased at higher relative humidity values. In cities like Beijing, it was observed that the daily variation in $PM_{2.5}$ tracked the patterns of relative humidity (Cheng et al., 2015), and humidity is known to facilitate multiphase reactions for aerosol formation and growth (Le et al., 2020). Interestingly, although precipitation remove pollutants by wet deposition, it was not among the most important variables for the model, owing that the number of days with precipitations in the city was low (46–72 days per year for the period 2017–2021). For the larger particles PM_{10}, the observed increase under dry conditions with low humidity values was most likely related to favorable conditions for dust generation and PM resuspension.

The rationale for including a temporal feature such as the hour of the day is that pollutant concentrations of a given hour not only depends on traffic emissions of that hour but also on the built up of pollutants from previous hours and on the chemical and photochemical activity, that varies along the day. It results in repeated patterns of pollutant accumulation during the day that can be included in the model through the hour of the day, increasing model predictability. This feature was especially important for NO_2, but also for O_3. The day of the week was also included but turned out to be a less relevant feature.

It must be noted that variables such as the air quality stations, the distance to the harbor and the days with Saharan dust intrusion, European intrusion or biomass burning were also included in the model. However, according to the SHAP analysis, these variables were not ranked among the 10 most important ones to predict hourly pollutant concentrations. Nevertheless, Saharan dust intrusions, European intrusions and days with biomass burning from agriculture events are relevant for compliance of the AQD limits as about a third of the highest PM_{10} values and 9% for $PM_{2.5}$ were associated with them, contributing to part of the observed exceedances of the air quality limit values. A proper quantification of the contribution of the harbor to the urban pollution of Valencia would require of a tailored study and model techniques which are beyond the scope of the present study.

4.3. Effect of vehicle electrification on air pollution and human health

Yang et al. (2021), applied the Random Forest prediction method to estimate the effect of electrification on air quality in Los Angeles Basin. Compared to that study, a different type of ML model was used by us, and pollutant concentrations were predicted for each air quality station (as requested by the AQD), under a 10-year range of meteorological conditions. Furthermore, the effects of VE on human health were also evaluated. Air quality in Valencia is already in compliance with the AQD, but not with the most recent WHO AQG, which anticipate future changes in the AQD. The present study provides insights into the feasibility of using VE as a tool for meeting the WHO AQG requirements in Valencia.

The results of the model suggested that there is a high potential for NO_2 concentration reduction (to about half in some stations) by increasing VE up to a 70%. This is consistent with the characteristics of the city, in which traffic represents the bulk of the NO_2 emissions and NOx-emitting industries are scarce in the vicinity. In Valencia, reductions of up to 71.6% in NO_2 concentrations were observed during the lockdown period (Sicard et al., 2020a), when traffic flow was reduced up to ~80%. Even though, a 70% VE will not be enough to reduce NO_2 concentrations to 10 μg m^{-3} (annual mean, WHO AQG 2021). This suggests that, even in cities like Valencia in which the AQD limits are already fulfilled, human health risks due to NO_2 will persist for a long time. However, any effort for reducing vehicle NO_2 emissions would have an important effect on human heath, e.g., an affordable 20% VE would result in a 20% reduction of premature deaths.

For PM, increases of up to 70% VE would not be enough to be in compliance with the WHO AQG as reductions in PM emissions by cars and associated premature mortality were limited. This was expected as electric vehicles are 24% heavier than their conventional counterparts increasing non-exhaust emissions, which results in PM emissions comparable or only slightly lower than those of conventional vehicles (Timmers and Achten, 2016; Liu et al., 2021). Furthermore, in an apportionment study conducted in several cities from southern Europe including Barcelona, only 30–35% of the urban annual PM_{10} and $PM_{2.5}$ averages were directly attributable to road traffic, with a considerable fraction of the PM being secondary (50–60% of PM_{10} and 65–70% of $PM_{2.5}$) (Amato et al., 2016). As the climatic conditions of the Mediterranean favor photochemistry and therefore formation of secondary PM, and given that sources of anthropogenic PM are diverse, including traffic exhaust emissions, vehicle wear, industry, construction, demolition and works, and biomass burning, a tight control of all sources of emissions will be needed to achieve the WHO AQG. It also has to be mentioned that Saharan intrusions represent an important source of non-anthropogenic PM affecting southern Europe that cannot be avoided, and that together with the transport of mineral dust load, locally emitted or co-transported anthropogenic pollutants are also accumulated (Querol et al., 2019). Given that PM is the most important cause of air pollution-related mortality in European cities (Khomenko et al., 2021) and traffic is a key emission source within the cities, actions to reduce PM necessarily will imply a reduction of urban traffic rather than VE, but also a control of the other anthropogenic PM emissions.

The prediction of change with increasing VE for O_3 represents an interesting case. In comparison with the current levels, the model predicted an increasing trend for most stations, more remarkable in traffic stations. This effect of VE on O_3 is obviously indirectly, driven by the associated changes in NOx concentrations in a VOC-limited environment, as it is the case of urban areas (Sicard, 2021). The O_3 concentrations in cities are usually lower than in rural areas, due to O_3 titration by freshly emitted NO by traffic (e.g, Monks et al., 2009; Sicard, 2021), and also to higher aerosol concentrations, which decrease O_3 production efficiency by decreasing UV radiation flux by scattering or absorption (Wang et al., 2014). Consistently, a rise in O_3 is observed in some cities during the weekends when both NOx and PM emissions by traffic decline (e.g., Seguel et al., 2012; Wang et al., 2014; Sicard et al., 2020b) and increasing trends in O_3 levels have been reported worldwide in parallel with NOx reductions (Sicard, 2021). Remarkable O_3 increases were also observed worldwide during the COVID-19 lockdown (e.g., Sicard et al., 2020a; Siciliano et al., 2020; Grange et al., 2021; Lovric et al., 2021; Zhang et al., 2022). For the two suburban stations, however, the model predicted different response patterns to increasing VE. In a downwind station (Molí), fully under the city plume influence during the central hours of the day, O_3 was predicted to increase, but in another station receiving less NO emissions (Politècnic, data now shown), O_3 decreased. This heterogeneity in the responses regarding O_3 represents a challenge for predictions. The chemistry of O_3 formation and destruction is complex and non-linear, furthermore involving on a myriad of VOCs which contribute both to ozonolysis and O_3 formation (Monks et al., 2009). Depending on characteristics of the city and stations, including local NO emissions, meteorological conditions, and NOx-to-VOCs ratios and their seasonal and daily variations (Sillman and He, 2002; Yang et al., 2019; Zhang et al., 2022; Akimoto and Tanimoto, 2022), increasing VE may result in a rise or decrease of O_3 concentrations. Indeed, some of stations studied in Valencia even showed increases until a given point of VE, and then a decrease (Fig. 4). The results from Valencia suggest that even with 70% VE the WHO AQG will not be fulfilled, and that, contrary to the other air pollutants, O_3 will remain stable or even increase in parts of the city, while improvements can occur in some suburban areas. In terms of O_3-related premature mortality, future VE will not importantly change the current situation.

For Valencia, NO_2 was concluded to be currently the most important pollutant in terms of premature mortality in the city, followed by $PM_{2.5}$ and far from O_3 and PM_{10}. However, uncertainty of the effects of NO_2 and PM_{10} is higher than that of $PM_{2.5}$ and O_3 (Gamarra et al., 2020), and the estimations of premature mortality are strongly dependent on the cut-off value and relative risks values chosen (Table S2). Also note that, as concentrations of the pollutants are frequently correlated, a simple addition of cases is not possible as it may lead to double counting, e. g, up to 30% in the case of $PM_{2.5}$ and NO_2 (World Health Organisation, 2013).

4.4. Applicability and measures to reduce air pollutants

This study shows how ML models can be applied to investigate the impacts of VE and traffic changes in air quality in the cities basically using meteorological, air quality and traffic data, which are frequently available, and with a low computational cost. Furthermore, in comparison with other models like CTMs or source-receptor relationship models, there is no need for an updated emission inventory which is not available for many cities. This increases the potential applicability of ML models in many cities in support of policy and decision making, also considering the expected health benefits, as shown in the present study. Admittedly, the output of both types of models is different, as our approach evaluate the effects on the existing air quality stations, while the CTMs and other models can provide city-wide spatial estimations of pollutant concentrations. Ideally, both approaches should be combined.

Measures to reduce urban air pollution include a range of solutions that should be implemented in parallel, with VE being only one of them. An unpublished pilot study in Valencia using GPS tracks in 50 different types of vehicles showed that about 65% of total km during the measuring period belonged to freight vehicles, taxis, and similar services, against only about 25% to cars from resident people. Therefore, electrification of this kind of vehicles should be a priority. Development of policies to reduce car weight to limit non-exhaust PM emissions has also been recommended (Timmers and Achten, 2016; Beddows and Harrison, 2021).

However, while VE will provide clear benefits to cities' air quality, to achieve climate mitigation goals and making urban areas more livable, it is unavoidable to reduce cars, as they have important associated problems of congestion, noise, traffic injuries and space-intense use (Gössling, 2020). In parallel, green solutions should be promoted, as plants provide a range of ecosystem services including the reduction of air pollution (Nowak, 2006; Chaparro and Terradas, 2009). However, plants also emit BVOCs that contribute to O_3 and secondary aerosol formation (Calfapietra et al., 2013), with BVOCs-related Ozone Formation Potential overpassing that of the anthropogenic VOCs in cities like Los Angeles (Gu et al., 2021). Therefore, selecting the most appropriate vegetation, maximizing air pollutant removal, at the same time avoiding BVOC emission, is especially relevant for limiting the formation of secondary pollutants in cities with Mediterranean climate, where photochemical activity is particularly intense (Sicard et al., 2018, 2022; Gu et al., 2021).

5. Conclusions

A case study of interpretation of the factors determining air pollution levels in the city of Valencia has been presented. The XGBoost coupled with SHAP analysis provides an optimal approach for identifying the importance and effect of the different factors. The use of this approach is very suitable for investigating the effect of future VE in the cities under different meteorological conditions, being very efficient in terms of computational requirements and interpretability of the results. Data from the COVID-19 lockdown provided invaluable information for training ML models and make future predictions on mobility changes in cities. For the case of Valencia, the model predicted that VE can importantly reduce the NO_2 pollution levels but will have a very limited effect on PM and O_3 concentrations. Although the stringent WHO 2021 AQG will not be fulfilled for NO_2, O_3, and $PM_{2.5}$, even with 70% VE, any further steps for reducing these pollutants by VE and other strategies are highly relevant in terms of protecting human health.

Credit author statement

V.C, P.S. and E.A.: Concept of the project. V.C. and J.D.: Data curation and collation. V.C., P.S. and J.D.: Data analysis. V.C. wrote the original draft. All authors participated in writing, reviewing and editing the manuscript.

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

Data will be made available on request.

Acknowledgements

V.C. thanks projects IMAGINA (PROMETEU 2019; Generalitat Valenciana) and URBFLUX (PID 2021-125941OB-I00), the latter funded by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe". This work was carried out with the contribution of the LIFE financial instrument of the European Union in the framework of the AIRFRESH project "*Air pollution removal by urban forests for a better human well-being*" (LIFE19 ENV/FR/000086). P.S acknowledges the Research Group 8.04.00 "*Air Pollution and Climate Change*" and Working Party 8.04.05 "*Ground-level ozone*" under the International Union of Forest Research Organizations. Data regarding PM intrusions and PM from biomass burning have been kindly provided by the Ministry for the Ecological Transition and the Demographic Challenge, in the framework of "Encargo del Ministerio para la Transición Ecológica a la Agencia Estatal Consejo Superior de Investigaciones Científicas para la detección de episodios naturales de aportes transfronterizos de partículas y otras fuentes de contaminación de material particulado, y de formación de ozono troposférico".

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2023.115835.

References

Acosta-Ramírez, C., Higham, J.E., 2022. Effects of Meteorology and Human-mobility on UK's Air Quality during COVID-19. Meteorological Applications, vol. 29.
Akimoto, H., Tanimoto, H., 2022. Rethinking of the adverse effects of NOx-control on the reduction of methane and tropospheric ozone – challenges toward a denitrified society. Atmos. Environ. 277, 119033.
Aloi, A., Alonso, B., Benavente, J., Cordera, R., Echániz, E., González, F., Ladisa, C., Lezama-Romanelli, R., López-Parra, Á., Mazzei, V., Perrucci, L., Prieto-Quintana, D., Rodríguez, A., Sañudo, R., 2020. Effects of the COVID-19 lockdown on urban mobility: empirical evidence from the city of santander (Spain). Sustainability 12.
Amato, F., Alastuey, A., Karanasiou, A., Lucarelli, F., Nava, S., Calzolai, G., Severi, M., Becagli, S., Gianelle, V.L., Colombi, C., Alves, C., Custódio, D., Nunes, T., Cerqueira, M., Pio, C., Eleftheriadis, K., Diapouli, E., Reche, C., Minguillón, M.C., Manousakas, M.-I., Maggos, T., Vratolis, S., Harrison, R.M., Querol, X., 2016. AIRUSE-LIFE+: a harmonized PM speciation and source apportionment in five southern European cities. Atmos. Chem. Phys. 16, 3289–3309.
Amoatey, P., Sicard, P., De Marco, A., Omidi Khaniabadi, Y., 2020. Long-term exposure to ambient $PM_{2.5}$ and impacts on health in Rome, Italy. Clin. Epidemiol. Global Health 8, 531–535.
Anenberg, S.C., 2018. Estimates of the global burden of ambient $PM_{2.5}$, ozone, and NO_2 on asthma incidence and emergency room visits. Environ. Health Perspect. 126, 107004)107001-107014.
Beddows, D.C., Harrison, R.M., 2021. PM_{10} and $PM_{2.5}$ emission factors for non-exhaust particles from road vehicles: dependence upon vehicle mass and implications for battery electric vehicles. Atmos. Environ. 244, 117886.
Borge, R., Jung, D., Lejarraga, I., de la Paz, D., Cordero, J.M., 2022. Assessment of the Madrid region air quality zoning based on mesoscale modelling and k-means clustering. Atmos. Environ. 287.
Cakaj, A., Lisiak-Zielinska, M., Khaniabadi, Y.O., Sicard, P., 2022. Premature deaths related to urban air pollution in Poland. Atmos. Environ. 301, 119723.
Calfapietra, C., Fares, S., Manes, F., Morani, A., Sgrigna, G., Loreto, F., 2013. Role of Biogenic Volatile Organic Compounds (BVOC) emitted by urban trees on ozone concentration in cities: a review. Environ. Pollut. 183, 71–80.
Chaparro, L., Terradas, J., 2009. Barcelona Ecosystem Analysis. CREAF, p. 96.
Chen, T., Guestrin, C., 2016. Xgboost: a scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794.
Chen, T., He, T., Benesty, M., Khotilovich, V., 2019. Package 'xgboost'. R Version 90.
Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., Li, M., Xie, J., Lin, M., Gend, Y., Li, Y., Yuan, J., 2022. Xgboost: Extreme Gradient Boosting. R package version 1. 5. 2. 1. https://CRAN.R-project.org/package=xgboost.
Cheng, Y., He, K.-b., Du, Z.-y., Zheng, M., Duan, F.-k., Ma, Y.-l., 2015. Humidity plays an important role in the $PM_{2.5}$ pollution in Beijing. Environ. Pollut. 197, 68–75.
De Marco, A., Garcia-Gomez, H., Collalti, A., Khaniabadi, Y.O., Feng, Z., Proietti, C., Sicard, P., Vitale, M., Anav, A., Paoletti, E., 2022. Ozone modelling and mapping for risk assessment: an overview of different approaches for human and ecosystems health. Environ. Res. 211, 113048.
Escudero, M., Querol, X., Ávila, A., Cuevas, E., 2007. Origin of the exceedances of the European daily PM limit value in regional background areas of Spain. Atmos. Environ. 41 (4), 730–744.

European Environmental Agency, 2020. Air quality in Europe - 2020 report. In: EEA Report No 09/2020. Copenhagen, p. 160.
Feng, Z., Yuan, X., Fares, S., Loreto, F., Li, P., Hoshika, Y., Paoletti, E., 2019. Isoprene is more affected by climate drivers than monoterpenes: a meta-analytic review on plant isoprenoid emissions. Plant Cell Environ. 42, 1939–1949.
Gamarra, A., Lechón, Y., Vivanco, M., Garrido, J., Martín, F., Sánchez, E., Theobald, M., Gil, V., Santiago, J., 2020. Benefit analysis of the 1st Spanish air pollution control programme on health impacts and associated externalities. Atmosphere 12.
Goodman, A.L., Underwood, G.M., Grassian, V.H., 1999. Heterogeneous reaction of NO2: characterization of gas-phase and adsorbed products from the reaction, $2NO_2(g) + H_2O(a) \rightarrow HONO(g) + HNO_3(a)$ on hydrated silica particles. J. Phys. Chem. 103, 7217–7223.
Gössling, S., 2020. Why cities need to take road space from cars - and how this could be done. J. Urban Des. 25, 443–448.
Grange, S.K., Lee, J.D., Drysdale, W.S., Lewis, A.C., Hueglin, C., Emmenegger, L., Carslaw, D.C., 2021. COVID-19 lockdowns highlight a risk of increasing ozone pollution in European urban areas. Atmos. Chem. Phys. 21, 4169–4185.
Gu, S., Guenther, A., Faiola, C., 2021. Effects of anthropogenic and biogenic volatile organic Compounds on Los Angeles air quality. Environ. Sci. Technol. 55, 12191–12201.
Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349, 255–260.
Kaplan, J., 2020. fastDummies: Fast Creation of Dummy (Binary) Columns and Rows from Categorical Variables. R package version 1.6.3. https://CRAN.R-project.org/package=fastDummies.
Keswani, A., Akselrod, H., Anenberg, S.C., 2022. Health and clinical impacts of air pollution and linkages with climate change. NEJM Evidence 1, EVIDra2200068.
Khomenko, S., Cirach, M., Pereira-Barboza, E., Mueller, N., Barrera-Gómez, J., Rojas-Rueda, D., de Hoogh, K., Hoek, G., Nieuwenhuijsen, M., 2021. Premature mortality due to air pollution in European cities: a health impact assessment. Lancet Planet. Health 5, e121–e134.
Kogevinas, M., Castano-Vinyals, G., Karachaliou, M., Espinosa, A., de Cid, R., Garcia-Aymerich, J., Carreras, A., Cortes, B., Pleguezuelos, V., Jimenez, A., Vidal, M., O'Callaghan-Gordo, C., Cirach, M., Santano, R., Barrios, D., Puyol, L., Rubio, R., Izquierdo, L., Nieuwenhuijsen, M., Dadvand, P., Aguilar, R., Moncunill, G., Dobano, C., Tonne, C., 2021. Ambient air pollution in relation to SARS-CoV-2 infection, antibody response, and COVID-19 disease: a cohort study in catalonia, Spain (COVICAT study). Environ. Health Perspect. 129, 117003.
Kracht, O., Santiago, J.L., Martin, F., Piersanti, A., Cremona, G., Righini, G., et al., 2017. Spatial Representativeness of Air Quality Monitoring Sites - Outcomes of the FAIRMODE/AQUILA Intercomparison Exercise", 978-92-79-77218-4, JRC108791. Publications Office of the European Union, Luxembourg, 2017.
Kuhn, M., 2021. Caret: Classification and Regression Training. R Package Version 6.0-90. https://CRAN.R-project.org/package=caret.
Lamarque, J.F., Shindell, D.T., Josse, B., Young, P.J., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W.J., Doherty, R., Dalsoren, S., Faluvegi, G., Folberth, G., Ghan, S.J., Horowitz, L.W., Lee, Y.H., MacKenzie, I.A., Nagashima, T., Naik, V., Plummer, D., Righi, M., Rumbold, S.T., Schulz, M., Skeie, R.B., Stevenson, D.S., Strode, S., Sudo, K., Szopa, S., Voulgarakis, A., Zeng, G., 2013. The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): overview and description of models, simulations and climate diagnostics. Geosci. Model Dev. (GMD) 6, 179–206.
Lamaud, E., Carrara, A., Brunet, Y., Lopez, A., Druilhet, A., 2002. Ozone fluxes above and within a pine forest canopy in dry and wet conditions. Atmos. Environ. 36, 77–88.
Le, T., Wang, Y., Liu, L., Yang, J., Yung, Y.L., Li, G., Seinfeld, J.H., 2020. Unexpected air pollution with marked emission reductions during the COVID-19 outbreak in China. Science 369, 702–706.
Liu, C., Chen, R., Sera, F., Vicedo-Cabrera, A.M., Guo, Y., Tong, S., Coelho, M., Saldiva, P. H.N., Lavigne, E., Matus, P., Valdes Ortega, N., Osorio Garcia, S., Pascal, M., Stafoggia, M., Scortichini, M., Hashizume, M., Honda, Y., Hurtado-Diaz, M., Cruz, J., Nunes, B., Teixeira, J.P., Kim, H., Tobias, A., Iniguez, C., Forsberg, B., Astrom, C., Ragettli, M.S., Guo, Y.L., Chen, B.Y., Bell, M.L., Wright, C.Y., Scovronick, N., Garland, R.M., Milojevic, A., Kysely, J., Urban, A., Orru, H., Indermitte, E., Jaakkola, J.J.K., Ryti, N.R.I., Katsouyanni, K., Analitis, A., Zanobetti, A., Schwartz, J., Chen, J., Wu, T., Cohen, A., Gasparrini, A., Kan, H., 2019. Ambient particulate air pollution and daily mortality in 652 cities. N. Engl. J. Med. 381, 705–715.
Liu, F., van der A, R.J., Eskes, H., Ding, J., Mijling, B., 2018. Evaluation of modeling NO2 concentrations driven by satellite-derived and bottom-up emission inventories using in situ measurements over China. Atmos. Chem. Phys. 18, 4171–4186.
Liu, Y., Chen, H., Gao, J., Li, Y., Dave, K., Chen, J., Federici, M., Perricone, G., 2021. Comparative analysis of non-exhaust airborne particles from electric and internal combustion engine vehicles. J. Hazard Mater. 420, 126626.
Lovric, M., Pavlovic, K., Vukovic, M., Grange, S.K., Haberl, M., Kern, R., 2021. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environ. Pollut. 274, 115900.
Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30.
Ma, J., Ding, Y., Cheng, J.C.P., Jiang, F., Tan, Y., Gan, V.J.L., Wan, Z., 2020. Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. J. Clean. Prod. 244, 118955.
Malashock, D.A., DeLang, M.N., Becker, J.S., Serre, M.L., West, J.J., Chang, K.-L., Cooper, O.R., Anenberg, S.C., 2022. Estimates of ozone concentrations and attributable mortality in urban, peri-urban and rural areas worldwide in 2019. Environ. Res. Lett. 17.
Massagué, J., Carnerero, C., Escudero, M., Baldasano, J.M., Alastuey, A., Querol, X., 2019. 2005–2017 ozone trends and potential benefits of local measures as deduced from air quality measurements in the north of the Barcelona metropolitan area. Atmos. Chem. Phys. 19, 7445–7465.
Millán, M.M., Mantilla, E., Salvador, R., Carratalá, A., Sanz, M.J., Alonso, L., Gangoiti, G., Navazo, M., 2000. Ozone cycles in the western mediterranean basin: interpretation of monitoring data in complex coastal terrain. J. Appl. Meteorol. 39, 487–508.
Millán, M.M., Salvador, R., Mantilla, E., Kallos, G., 1997. Photooxidant dynamics in the Mediterranean basin in summer: results from European research projects. J. Geophys. Res. Atmos. 102, 8811–8823.
Monks, P.S., Granier, C., Fuzzi, S., Stohl, A., Williams, M.L., Akimoto, H., Amann, M., Baklanov, A., Baltensperger, U., Bey, I., Blake, N., Blake, R.S., Carslaw, K., Cooper, O.R., Dentener, F., Fowler, D., Fragkou, E., Frost, G.J., Generoso, S., Ginoux, P., Grewe, V., Guenther, A., Hansson, H.C., Henne, S., Hjorth, J., Hofzumahaus, A., Huntrieser, H., Isaksen, I.S.A., Jenkin, M.E., Kaiser, J., Kanakidou, M., Klimont, Z., Kulmala, M., Laj, P., Lawrence, M.G., Lee, J.D., Liousse, C., Maione, M., McFiggans, G., Metzger, A., Mieville, A., Moussiopoulos, N., Orlando, J.J., O'Dowd, C.D., Palmer, P.I., Parrish, D.D., Petzold, A., Platt, U., Pöschl, U., Prévôt, A.S.H., Reeves, C.E., Reimann, S., Rudich, Y., Sellegri, K., Steinbrecher, R., Simpson, D., ten Brink, H., Theloke, J., van der Werf, G.R., Vautard, R., Vestreng, V., Vlachokostas, C., von Glasow, R., 2009. Atmospheric composition change – global and regional air quality. Atmos. Environ. 43, 5268–5350.
Nguyen, D.-H., Lin, C., Vu, C.-T., Cheruiyot, N.K., Nguyen, M.K., Le, T.H., Lukkhasorn, W., Vo, T.-D.-H., Bui, X.-T., 2022. Tropospheric ozone and NOx: a review of worldwide variation and meteorological influences. Environ. Tech. Innov. vol. 28.
Nielsen, D., 2016. Tree boosting with xgboost-why does xgboost win" every" machine learning competition? Master Thesis, NTNU.
Nowak, D.J., 2006. Institutionalizing urban forestry as a "biotechnology" to improve environmental quality. Urban For. Urban Green. 5, 93–100.
Paoli, L., Gül, T., 2022. Electric Cars Fend off Supply Challenges to More than Double Global Sales. International Energy Agency. https://www.iea.org/commentaries/electric-cars-fend-off-supply-challenges-to-more-than-double-global-sales.
Pisoni, E., Thunis, P., Clappier, A., et al., 2019. Application of the SHERPA source-receptor relationships, based on the EMEP MSC-W model, for the assessment of air quality policy scenarios. Atmos. Environ. 4, 100047.
Querol, X., Perez, N., Reche, C., Ealo, M., Ripoll, A., Tur, J., Pandolfi, M., Pey, J., Salvador, P., Moreno, T., Alastuey, A., 2019. African dust and air quality over Spain: is it only dust that matters? Sci. Total Environ. 686, 737–752.
Querol, X., Massague, J., Alastuey, A., Moreno, T., Gangoiti, G., Mantilla, E., Dueguez, J. J., Escudero, M., Monfort, E., Perez Garcia-Pando, C., Petetin, H., Jorba, O., Vazquez, V., de la Rosa, J., Campos, A., Munoz, M., Monge, S., Hervas, M., Javato, R., Cornide, M.J., 2021. Lessons from the COVID-19 air pollution decrease in Spain: now what? Sci. Total Environ. 779, 146380.
Recull Estadístic València, 2021. Ajuntament de València. https://www.valencia.es/estadistica/Recull/Recull2021_Castellano.pdf.
Seguel, R.J., Morales, S.R., Leiva, G.M., 2012. Ozone weekend effect in Santiago, Chile. Environ. Pollut. 162, 72–79.
Sicard, P., 2021. Ground-level ozone over time: an observation-based global overview. Curr. Opin. Environ. Sci. Health 19, 100226.
Sicard, P., Agathokleous, E., Araminiene, V., Carrari, E., Hoshika, Y., De Marco, A., Paoletti, E., 2018. Should we see urban trees as effective solutions to reduce increasing ozone levels in cities? Environ. Pollut. 243, 163–176.
Sicard, P., De Marco, A., Agathokleous, E., Feng, Z., Xu, X., Paoletti, E., Rodriguez, J.J.D., Calatayud, V., 2020a. Amplified ozone pollution in cities during the COVID-19 lockdown. Sci. Total Environ. 735, 139542.
Sicard, P., Paoletti, E., Agathokleous, E., Araminiene, V., Proietti, C., Coulibaly, F., De Marco, A., 2020b. Ozone weekend effect in cities: deep insights for urban air pollution control. Environ. Res. 191, 110193.
Sicard, P., Agathokleous, E., De Marco, A., Paoletti, E., Calatayud, V., 2021. Urban population exposure to air pollution in Europe over the last decades. Environ. Sci. Eur. 33, 28.
Sicard, P., Agathokleous, E., De Marco, A., Paoletti, E., 2022. Ozone-reducing urban plants: choose carefully. Science 377, 585, 585.
Sicard, P., Agathokleous, E., Anenberg, S.C., De Marco, A., Paoletti, E., Calatayud, V., 2023. Trends in urban air pollution over the last two decades: a global perspective. Sci. Total Environ. 858, 160064.
Siciliano, B., Dantas, G., da Silva, C.M., Arbilla, G., 2020. Increased ozone levels during the COVID-19 lockdown: analysis for the city of Rio de Janeiro, Brazil. Sci. Total Environ. 737, 139765.
Sillman, S., He, D., 2002. Some theoretical results concerning O_3-NOx-VOC chemistry and NOx-VOC indicators. J. Geophys. Res. Atmos. 107. ACH 26-21-ACH 26-15.
Timmers, V.R.J.H., Achten, P.A.J., 2016. Non-exhaust PM emissions from electric vehicles. Atmos. Environ. 134, 10–17.
United Nations, 2019. World Urbanization Prospects: the 2018 Revision. United Nations (Department of Economic and Social Affairs Population Division), New York. ST/ESA/SER.A/420).
Valuntaitė, V., Šerevičienė, V., Girgždienė, R., Paliulis, D., 2012. Relative humidity and temperature impact to ozone and nitrogen oxides removal rate in the experimental chamber. J. Environ. Eng. Landsc. Manag. 20, 35–41.
Venter, I., 2022. Electric Car Sales Hit 6.6-million in 2021, Tripling Their Market Share from 2019 – IEA. Engineering News, 11th March 2022.
Viana, M., López, J.M., Querol, X., Alastuey, A., García-Gacio, D., Blanco-Heras, G., López-Mahía, P., Piñeiro-Iglesias, M., Sanz, M.J., Sanz, F., Chi, X., Maenhaut, W.,

2008. Tracers and impact of open burning of rice straw residues on PM in Eastern Spain. Atmos. Environ. 42, 1941–1957.
Wang, Y.H., Hu, B., Ji, D.S., Liu, Z.R., Tang, G.Q., Xin, J.Y., Zhang, H.X., Song, T., Wang, L.L., Gao, W.K., Wang, X.K., Wang, Y.S., 2014. Ozone weekend effects in the Beijing–Tianjin–Hebei metropolitan area, China. Atmos. Chem. Phys. 14, 2419–2429.
World Health Organisation, 2013. Health Risks of Air Pollution in Europe—HRAPIE Project, Recommendations for Concentration–Response Functions for Cost–Benefit Analysis of Particulate Matter, Ozone and Nitrogen Dioxide. WHO, Geneva, Switzerland, 2013.
World Health Organization, 2020. Health Impact Assessment of Air Pollution: Introductory Manual to AirQ+. World Health Organization. Regional Office for Europe. https://apps.who.int/iris/handle/10665/337681. Regional Office for Europe.
World Health Organization, 2021a. Ambient (Outdoor) Air Pollution Fact Sheet [22/10/2021, consulted 11/11/2022]. Available from: https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health.
World Health Organization, 2021b. WHO Global Air Quality Guidelines. Particulate Matter ($PM_{2.5}$ and PM_{10}), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. World Health Organization, Geneva.
Xu, C., Zhang, Z., Ling, G., Wang, G., Wang, M., 2022. Air pollutant spatiotemporal evolution characteristics and effects on human health in North China. Chemosphere 294, 133814.
Yang, Y., Liu, X., Zheng, J., Tan, Q., Feng, M., Qu, Y., An, J., Cheng, N., 2019. Characteristics of one-year observation of VOCs, NOx, and O_3 at an urban site in Wuhan, China. J. Environ. Sci. 79, 297–310.
Yang, J., Wen, Y., Wang, Y., Zhang, S., Pinto, J.P., Pennington, E.A., Wang, Z., Wu, Y., Sander, S.P., Jiang, J.H., Hao, J., Yung, Y.L., Seinfeld, J.H., 2021. From COVID-19 to future electrification: assessing traffic impacts on air quality by a machine-learning model. Proc. Natl. Acad. Sci. U. S. A. 118.
Young, P.J., Archibald, A.T., Bowman, K.W., Lamarque, J.F., Naik, V., Stevenson, D.S., Tilmes, S., Voulgarakis, A., Wild, O., Bergmann, D., Cameron-Smith, P., Cionni, I., Collins, W.J., Dalsøren, S.B., Doherty, R.M., Eyring, V., Faluvegi, G., Horowitz, L.W., Josse, B., Lee, Y.H., MacKenzie, I.A., Nagashima, T., Plummer, D.A., Righi, M., Rumbold, S.T., Skeie, R.B., Shindell, D.T., Strode, S.A., Sudo, K., Szopa, S., Zeng, G., 2013. Pre-industrial to end 21st century projections of tropospheric ozone from the atmospheric chemistry and climate model intercomparison project (ACCMIP). Atmos. Chem. Phys. 13, 2063–2090.
Zhang, K., Liu, Z., Zhang, X., Li, Q., Jensen, A., Tan, W., Huang, L., Wang, Y., de Gouw, J., Li, L., 2022. Insights into the significant increase in ozone during COVID-19 in a typical urban city of China. Atmos. Chem. Phys. 22, 4853–4866.
Zhou, Z.-H., 2021. Machine Learning. Springer Nature.