https://doi.org/10.1016/j.envpol.2022.120119

Increased contribution to $PM_{2.5}$ from traffic-influenced road dust in Shanghai over recent years and predictable future

Meng Wang[1], Yusen Duan[2], Zhuozhi Zhang[1], Juntao Huo[2], Yu Huang[3], Qingyan Fu[2], Tao Wang[1], Junji Cao[3,4], Shun-cheng Lee[1, *]

[1]Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

[2]Shanghai Environmental Monitoring Center, Shanghai, China

[3]State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China

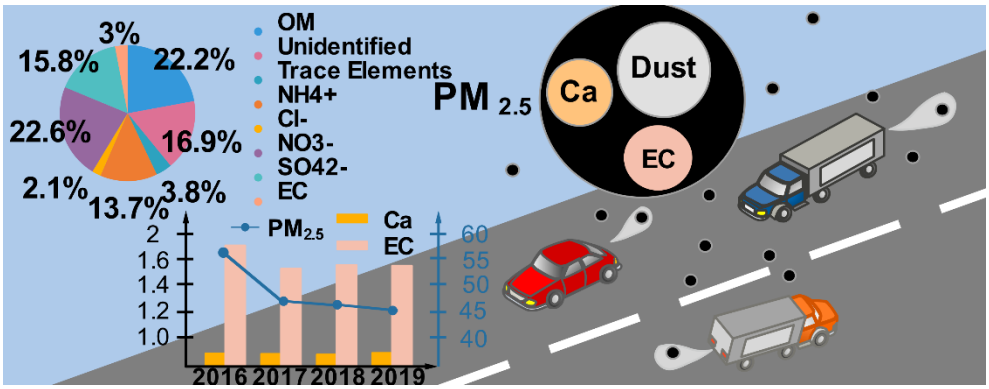
[4]Key Laboratory of Middle Atmosphere and Global Environment Observation, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

Correspondence to: Shun-cheng Lee (shun-cheng.lee@polyu.edu.hk).

Highlights:

1. The fractional contribution of road dust to $PM_{2.5}$ was found to be increasing year by year.
2. As a non-exhaust emission, road dust was found to be highly modulated by relative humidity.
3. Lower relative humidity is associated with higher road dust emissions.
4. Non-tailpipe emissions can be reduced by technology improvement and regulations.

Graphical Abstract



Abstract

Traffic contributes to fine particulate matter ($PM_{2.5}$) in the atmosphere through engine exhaust emissions and road dust generation. However, the evolution of traffic related $PM_{2.5}$ emission over recent years remains unclear, especially when various efforts to reduce emission e.g., aftertreatment technologies and high emission standards from China IV to China V, have been implemented. In this study, hourly elemental carbon (EC), a marker of primary engine exhaust emissions, and trace element of calcium (Ca), a marker of road dust, were measured at a nearby highway sampling site in Shanghai from 2016 to 2019. A random forest-based machine learning algorithm was applied to decouple the influences of meteorological variables on the measured EC and Ca, revealing the deweathered trend in exhaust emissions and road dust. After meteorological normalization, we showed that non-exhaust emissions, i.e., road dust from traffic, increased their fractional contribution to $PM_{2.5}$ over recent years. In particular, road dust was found to be more important, as revealed by the deweathered trend of Ca fraction in $PM_{2.5}$, increasing at 6.1% $year^{-1}$, more than twice that of EC (2.9% $year^{-1}$). This study suggests that while various efforts have been successful in reducing vehicular exhaust emissions, road dust will not abate at a similar rate. The results of this study provide insights into the trend of traffic-related emissions over recent years based on high temporal resolution monitoring data, with important implications for policymaking.

***Keywords*:** Random Forest; Traffic; Air pollution; Non-exhaust emission

1 Introduction

Vehicular emissions contribute to ambient fine particulate matter ($PM_{2.5}$) through engine exhaust and non-exhaust emissions e.g., road dust (An et al., 2019; Fuzzi et al., 2015; Gentner et al., 2017; Mukherjee et al., 2020). Exposure to emissions from heavily trafficked roadways can lead to adverse health effects, including dementia, Parkinson's disease, lung cancer, cardiovascular and respiratory diseases (Annavarapu and Kathi, 2016; Chen et al., 2017; Halonen et al., 2016). In particular, people living close (<500 m) to major roadways are directly impacted by vehicular emissions, exposing them to elevated $PM_{2.5}$ concentrations compared to the urban background level (Karner et al., 2010; Lin et al., 2020). Moreover, traffic-related $PM_{2.5}$ emissions containing elemental carbon (EC) and fugitive road dust (Chen et al., 2019) are likely more toxic than those e.g., sulfate and nitrate in the urban background $PM_{2.5}$ (Rappazzo et al., 2015; Sbihi et al., 2013).

With respect to primary $PM_{2.5}$ emissions from engine exhausts, gasoline vehicles generally pollute less than diesel vehicles, especially with the recent implementation of high emission standards (e.g., China IV and V) (Huang et al., 2022). Presently, over 90% of the vehicles in China are gasoline-powered, owing to the phasing out of vehicles with old emission standards (i.e., China I–III) (Wang et al., 2019). Moreover, exhaust gas recirculation (to reduce NO_x) and particulate filters have shown success in reducing vehicular emission (Ayodhya and Narayanappa, 2018; Cédric et al., 2016; Grange et al., 2017). Nevertheless, traffic emissions are still one of the most important sources of ambient $PM_{2.5}$ in urban China (An et al., 2019; Zhu et al., 2018). In real-world conditions, traffic mix and volume, vehicles ages, and vehicle speed can also impact the contribution of traffic to $PM_{2.5}$. On road or mobile measurements can capture the emission characteristics of various vehicle types (Huang et al., 2022). However, to monitor the overall trend of vehicular emissions, long-term multiple-year roadside measurements are necessary (Dabek-Zlotorzynska et al., 2019; Lin et al., 2020). Meteorological conditions, such as wind speed and direction, can also impact the observed $PM_{2.5}$, complicating the measurement process, and should be considered to reveal the deweathered trend in the evolution of vehicular emissions (Mukherjee et al., 2020; Wang et al., 2022).

Road dust caused by traffic-generated turbulence has attracted considerable research interest in recent years (Chen et al., 2019; Kong et al., 2011; Niu et al., 2019; Reff et al., 2009; Shakya et al., 2017). However, it has often been neglected or substantially underestimated by the current models (Chen et al., 2019). For example, in urban Lanzhou, China, Chen et al. (2019) constructed an emission inventory for road dust caused by traffic with an emission rate estimated to be approximately 1141±71 kg d^{-1}, accounting for 24.6% of total $PM_{2.5}$ emission. Such large road dust emissions were estimated to cause 234 premature deaths annually in urban Lanzhou (Chen et al., 2019). Through roadside filter measurements in the Kathmandu Valley, Nepal, Shakya et al. (2017) showed that road dust contained abundant elements, including silica, calcium, aluminum, and iron, the sum of which accounted for 10-19% of the $PM_{2.5}$ mass. Niu et al. (2019) compared the elemental compositions of road dust in 21 cities in China and revealed its spatial distribution across the country. However, most of these studies were based on filter measurements, influenced by artificial artifacts occurring during filter sampling and poor temporal resolution (e.g., 24 h mean), which have failed to provide sufficient information e.g., the diurnal variation. Moreover, lack of long-term roadside measurements of road dust limits our understanding of its recent trend and our ability to predict future emissions. In particular, as the number of vehicles is increasing (Jin and He, 2019; Wang et al., 2019), the corresponding road dust is expected to increase however, this remain poorly investigated in real-world conditions because of the lack of long-term roadside data.

In this study, markers for engine exhaust (EC), and road dust (calcium; Ca), were measured, along with the major components of $PM_{2.5}$, at an hourly resolution at a roadside near a highway in Shanghai from 2016 to 2019. A random forest-based machine learning algorithm (Grange and Carslaw, 2019; Grange et al., 2018; Grange et al., 2017) was applied to train the model to rebuild the measured EC, Ca, $PM_{2.5}$, and NO_x using meteorological and temporal variables as the model input. The SHapley Additive explanation (SHAP) algorithm (Lundberg et al., 2020; Oukawa et al., 2022) was used to understand the physical and chemical processes that govern the measured EC, Ca, and $PM_{2.5}$. Finally, a trend analysis was performed after meteorological normalization to reveal the deweathered changes over the four years.

2 Method

2.1 Sampling site

Hourly samples of EC and Ca in $PM_{2.5}$ were collected for four years at the Dianshan Lake (DSL) supersite (31.09° N,120.98° E, approximately 15 m above ground), located in Qingpu District in western Shanghai (Fig. S1). The DSL sampling site was approximately 7 km east of Dianshan Lake and 50 km west of downtown Shanghai, at the intersection of Jiangsu, Shanghai, and Zhejiang Provinces (Jia et al., 2020). The two highways (G318 and G50) were located approximately 1 km west of the sampling site. Windrose analysis showed that the sampling site could be affected by the two nearby highways (Figure S2). Owing to maintenance, data from July to September 2019 (6% of the four-year data) were not available.

2.2 Instruments

Hourly EC and organic carbon (OC) were measured on-line by a Sunset Carbon Analyzer (Model RT-4, Sunset Lab, USA) with a detection limit 0.04 and 0.2 μg m^{-3}, respectively. OC was converted to organic matter (OM) using a factor of 1.8 (Via et al., 2021; Wu et al., 2018). Ca with a size cut-off $PM_{2.5}$ was determined by a continuous multi-metals monitor (Model Xact 625, Cooper Environmental Services, Beaverton, OR, USA) along with other elements (Si, K, V, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Ag, Cd, Ba, Pb, and Hg). Because Ca showed a well-defined diurnal pattern with typical rush hour peaks (see Section 3), we focused on its trend analysis, which was used as a marker for road dust caused by vehicles passing through nearby highways.

The total $PM_{2.5}$ mass concentrations were measured using a tapered-element oscillating microbalance monitor (1405 TEOM, Thermo Fisher Scientific, Massachusetts, SUA), while an online Monitor for AeRosols and Gases (MARGA, Model ADI2080, Metrohm Applikon B.V., The Netherlands) was deployed to measure the water-soluble inorganic ions, including sulfate, nitrate, chloride, and ammonium in $PM_{2.5}$. A Visala automatic weather station (WXT520, Vaisala Ltd., Finland) was deployed to monitor air temperature (air_temp), wind direction (wd), wind speed (ws), relative humidity (RH), pressure, and rainfall. The gaseous pollutants nitrogen monoxide (NO) and nitrogen dioxide (NO_2) were measured with Thermo Scientific gas analyzers (Thermo 42i, Thermo Fisher Scientific, Massachusetts, USA).

2.3 Data analysis

2.3.1 Random Forest modeling set-up and validation

A decision-tree-based random forest model was developed to understand the trends of the observed EC, Ca, $PM_{2.5}$, and NO_x individually over the four years (2016–2019). EC was used as a marker of traffic exhaust emissions as traffic was its main contributor in Shanghai (Jia et al., 2021), whereas Ca was used as a marker of non-exhaust emissions (i.e., road dust) (Chang et al., 2018). The diurnal patterns of EC (Fig. S3), and Ca (Fig. S4) show elevated concentrations during rush hour, which is consistent with its

traffic-induced emission pattern. However, Ca may be occasionally associated with dust storms in spring, leading to the observed spikes. By comparing the time series trends of the original Ca and Ca after removing the days with high concentrations (top 5% of the data), we concluded that occasional dust storms did not impact the Ca trend.

Four random forests were developed, when the meteorological (ws, wd, air_temp, RH, rainfall, and pressure) and, time (date_unix, day of the year (day_julian), weekday, hour of the day, and day of the lunar year) variables for each random forest were used as model inputs. These variables can affect the observed $PM_{2.5}$ concentration, its components, and emission strengths from various sources. For example, time variables can act as proxies for emission strengths as they vary over time e.g., diurnal cycles and seasonal variations. The inclusion of the day of the lunar year Wang et al. (2020); Dai et al. (2021) to consider the effects of the Chinese New Year holiday. Eighty percent of the dataset was randomly selected as the training dataset, whereas the remaining 20% used to validate the models using the latest "rmweather" R package (Grange et al., 2018). To develop a tree, the number of independent/explanatory variables was set to three. The number of trees in the forest was set to 300, following Grange et al. (2018). The statistics of the model performance are summarized in Table S1 with root mean square error and mean absolute err in the range of 0.10–27.7 μg m^{-3} and 0.056–16.0 μg m^{-3}, respectively. The coefficient of determination (R^2) was 0.71–0.86, comparable to the reported values of similar studies (Qin et al., 2022; Zhou et al., 2022).

2.3.2 Meteorological normalization

For each target pollutant, meteorological normalization was performed by repeatedly resampling the explanatory variables (only meteorological variables included) and predicting the values at a specific time on a rolling basis over four years using deweathering techniques (Vu et al., 2019). The input meteorological variables (ws, wd, air_temp, RH, rainfall, and pressure) were randomly resampled from the original dataset (Vu et al., 2019). The resampled meteorological variables were subsequently fed to the random forest model to predict the dependent variables (EC, Ca, $PM_{2.5}$, and NO_x). This process was repeated 1000 times for each target variable. The corresponding 1000 predictions were arithmetically averaged to obtain meteorologically normalized or deweathered pollutants (Dai et al., 2021; Shi et al., 2021), i.e., the deweathered pollutants were not affected by a specific meteorological condition at a particular time.

2.3.3. Feature importance analysis

SHAP can be used to explain the output of any machine learning model (Lundberg et al., 2020; Oukawa et al., 2022). The SHAP (https://github.com/slundberg/shap) method was applied to evaluate the importance of the meteorological variables in predicting each target pollutant. SHAP uses an additive feature attribution method to produce an interpretable model (Lundberg et al., 2020), which quantified the contribution of the input meteorological variables to a single prediction at a specific time, producing a SHAP value in the same unit as the target pollutant. In this study, only the morning rush hour data (6:00–10:00) of the study period were used as model input to minimize the impact of daily variation on the pollutants.

2.3.4 Trend analysis using Theil-Sen algorithm

The long-term trends of EC, Ca, $PM_{2.5}$, and NO_x were assessed using the Theil-Sen regression algorithm (Carslaw and Ropkins, 2012), which is commonly used for long-term trend analysis as it can account for autocorrelation (Grange et al., 2018; Vu et al., 2019). It was used at the 95% confidence level to indicate

a significant trend. Slopes of the target pollutant (i.e., EC, Ca, $PM_{2.5}$ and NO_x) were calculated using the median value of all the possible slopes, based on bootstrap resampling, taken as a representative value for long-term trend analysis (Grange et al., 2018; Vu et al., 2019). The Theil-Sen functions provided in the "openair" package in R (version 4.1.2) (Carslaw and Ropkins, 2012) were applied in this study.

3 Results

3.1 Overview of $PM_{2.5}$ from 2016 to 2019: impact of traffic

Figure 1 shows the daily average time series of $PM_{2.5}$, EC, Ca, and NO_x concentrations at the roadside sampling site of the Dianshan Lake in Shanghai over the four years from 2016 to 2019. The time series of daily $PM_{2.5}$ varied from 6.6 to 218.0 μg m^{-3}, with a mean of 47.0±33.3 (standard deviation; SD) μg m^{-3} (Fig. 1). The annual $PM_{2.5}$ was highest in 2016 (Table 1), with an annual average of 52.3±35.3 μg m^{-3} (median of 42.0 μg m^{-3}; Table 1). It decreased to 44.8±31.3 μg m^{-3} in 2017 and to 43.6±33.8 μg m^{-3} in 2018 but slightly increased in 2019 (47.1±31.5 μg m^{-3}), which is still lower than in 2016. Median values followed the same trend as the mean values over the study period (Table 1). Despite the decrease in the annual $PM_{2.5}$ concentration, it was still substantially (1.5–5 times) higher than the air quality standards or guidelines, when compared to China's national ambient air quality standard (NAAQS-II) of 35 μg m^{-3} and the WHO guideline of 10 μg m^{-3}.

Among the four seasons, winter was the most polluted, with an average $PM_{2.5}$ concentration of 60.8 ± 41.4 μg m^{-3}, followed by spring (50.7±30.2 μg m^{-3}), autumn (38.1±27.3 μg m^{-3}), and summer (35.4 ± 23.2 μg m^{-3}). The high $PM_{2.5}$ concentration in winter and spring is likely associated with low temperatures and a relatively stable atmosphere, favoring the build-up of pollutants. The average temperatures in winter and spring are 6.3 ºC and 16.5 ºC, respectively, lower than in autumn (19.2 ºC) and summer (27.5 ºC). The diurnal pattern of $PM_{2.5}$ showed an increase during the morning rush hour (6:00 – 10:00; Fig. 2), suggesting the direct impact of traffic emission. This is confirmed by the diurnal pattern of EC (discussed in Section 3.2) and NO_x, showing a typical increase during the morning rush hour (Fig. 2). The diurnal pattern of $PM_{2.5}$ over the four years were similar (Fig. S5), suggesting no significant changes in the source types but with different source strengths.

Of the major $PM_{2.5}$ components (Fig. 3), included nitrate, organic matter, nitrate, sulfate, ammonium, trace elements, EC, and chloride, accounting for 22.6%, 22.1%, 15.8%, 13,7%, 3.8%, 3.0%, and 2.1% of the measured $PM_{2.5}$, respectively. The unidentified fraction accounted for 16.9% of the $PM_{2.5}$ mass. The large unidentified fraction suggests that road dust (e.g., minerals) was likely an important fraction of the component. The measured Ca also showed a diurnal pattern with a morning rush hour spike (Fig. 2) similar to EC, suggesting the resuspension of road dust caused by passing vehicles. Notably, uncertainties of the measured species (10–15%) might also add up, leading to a high fraction of unidentified $PM_{2.5}$ components. Nevertheless, sum of the measured trace elements (Ca, Si, K, V, Cr, Mn, Fe, Ni, Cu, Zn, As, Se, Ag, Cd, Ba, Pb, and Hg) had a higher fraction of $PM_{2.5}$ than EC (3.8% vs. 3.0%), confirming the large fraction of road dust to $PM_{2.5}$.

3.2 Traffic-related EC and Ca

EC varied from <0.1 to 7.34 μg m^{-3}, with an average of 1.43 μg m^{-3} over the four years (Fig. 1). The diurnal pattern of EC showed rush hour peaks with concentration (>1.50 μg m^{-3}) above the average. Moreover, it also showed a smaller peak at around 21:00, the late evening rush hour. A similar pattern was observed for NO_x, confirming that the major source of EC was traffic emissions. Such trends were observed throughout the study period (Fig. S3), but with varying magnitudes. Moreover, no differences

were observed between the weekdays and weekends (Fig. S6), suggesting that the nearby highway was busy throughout the sampling period. On an average, EC accounted for 3.0% of the $PM_{2.5}$ mass, which may appear small. However, direct emissions from traffic also contain OC, which is usually over twice higher than the EC concentration for gasoline vehicles (DeWitt et al., 2015; Huang et al., 2022). Therefore, the total primary aerosol emissions from vehicles may account for >10% of the $PM_{2.5}$ mass. Moreover, NO_x emitted from traffic can also act as precursor gases for nitrate, which is also a major component of $PM_{2.5}$. In this study, EC was used as a marker for the primary aerosol emission from vehicles.

Ca concentration varied in the range of 0.01–1.64 μg m^{-3}, with an average of 0.16 μg m^{-3} over the four years (Fig. 1). Although the diurnal cycles of Ca and EC were similar (Fig. 2), their time series were poorly correlated with an R^2 of 0.12 (Fig. 3). This is reasonable considering that EC is a by-product of incomplete combustion emitted from engine exhaust, while Ca occurs is due to the re-suspension of road dust that can be caused by any type of vehicle, including electric vehicles (no EC emission) and/or gasoline vehicles with high emission standards, i.e., low EC emissions from China V vehicles (Huang et al., 2022). Even at low EC levels (<1.0 μg m^{-3}), the distribution of Ca for datapoints of >0.1 μg m^{-3} was highly scattered (Fig. S6) with a 75% percentile of >0.2 μg m^{-3}, suggesting clean vehicles with low EC emission can also cause high Ca, i.e., road dust in $PM_{2.5}$.

Ca concentration varied slightly during the study period (2016–2019), with annual values in the range of 0.15–0.17 μg m^{-3} (Table 1). However, it increased during rush hours (Fig. S4) in the study period. As Ca is sourced from the re-suspension of road dust caused by passing vehicles and assuming its amount released by each vehicle is similar, the observed increasing Ca trend suggests that the actual number of vehicles on the road also increased. However, such a trend needs to be confirmed after meteorological normalization because changes in RH, ws, and wd can also alter the Ca concentrations (discussed in Section 3.4).

3.3 Random Forest modelling and the importance of meteorological variables

For each of the random forests developed for $PM_{2.5}$, EC, Ca, and NO_x, the predicted values were well correlated with the observed values (R^2 >0.95) for the training dataset (Figure 4), while for the testing dataset, the predicted value was reasonably well correlated with observed value (R^2 >0.70), suggesting the model explained roughly 70–85% of the variation. Good correlation between the predicted and observed values demonstrated the explanatory power of the developed random forest model built, providing more reliability for subsequent meteorological normalization.

To quantify the contribution of meteorological variables to the random forest-based prediction, SHAP modelling was performed (see Methods section, Figure 5). A negative or positive SHAP value (μg m^{-3}) indicates that the meteorological variable contributes more, whereas an SHAP value closer to zero indicates less contribution to the prediction. For the random forest of $PM_{2.5}$, ws was the most important meteorological variable, followed by air_temp, wd, pressure, RH, and rainfall (Fig. 5a). In particular, when ws was low (i.e., <1.1 m s^{-1}; 25th percentile; Table 2), a SHAP value of >10 μg m^{-3} was observed. In contrast, higher ws (>2.8 m s^{-1}) negatively contributed to $PM_{2.5}$, with SHAP values of < -10 μg m^{-3}. Air_temp was the second most important variable, also demonstrating a negative relationship with $PM_{2.5}$ (Fig. 5a). Wd (in degrees) was the third-most important variable. The wind direction dependence is also shown as polar plots, with a wd of 200–300 being associated with higher concentrations (Fig. S8). Rainfall was the least important among the six variables, contributing negatively to $PM_{2.5}$, as well as other pollutants, consistent with its we-deposition effects on air pollutants.

Similarly, ws was an important variable for EC and NO_x, with an overall negative impact (Fig. 5). In

contrast, RH was the most important parameter for the random forest grown of Ca (Fig. 5c), showing an inverse correlation (Fig. 5c), with low RH (<64%) associated with higher SHAP values of >0.1 μg m^{-3}. This reflects that the air was dry during Ca production (i.e., low RH) as Ca is primarily associated with the resuspension of road dust induced by passing vehicles, and humid air decreases its production rate. As demonstrated by the polar plots, ws and wd were also important variables, showing patterns similar to those of other pollutants, (Fig. S8).

3.4 Trend analysis of traffic exhaust and road dust emission

To decouple the effects of meteorological variables on the pollutants, meteorological normalization or deweathering analysis was performed (see the Methods section). Table 1 presents the yearly mean, one SD, and median $PM_{2.5}$, EC, Ca, and NO_x concentrations before and after meteorological normalization. For the mean value, the difference between the observed and weathered ((deweathered-observed)/observed) pollutants range from -1.8 to 7.7 % (Table 1), whereas for the median value, a larger difference (20–51.4 %) was observed. Large differences in the median values were likely caused by the uncertainties associated with machine learning training that failed to capture the extreme events, leading to significant deviations from the average values.

Figure 6 shows the trend analysis of monthly averaged $PM_{2.5}$ over the study period, as well as the fraction of EC in $PM_{2.5}$ and the fraction of Ca in $PM_{2.5}$ after meteorological normalization using the Theil-Sen algorithm; Figure S9 shows the trend before normalization. Specifically, for the deweathered $PM_{2.5}$, a trend of -6.4% $year^{-1}$ (or -3.65 μg m^{-3} $year^{-1}$) was estimated (Fig. 6), while it was -6.7% $year^{-1}$ (or -3.67 μg m^{-3} $year^{-1}$) for the observed $PM_{2.5}$ (Fig. S10). Therefore, trend of the deweathered $PM_{2.5}$ was only slightly more positive than the observed trend. The decreasing rate 6.4% $year^{-1}$ of $PM_{2.5,}$ was a result of all contributing factors, such as, industry, power plants, and residential sector (Zheng et al., 2018), as along with traffic (this study).

While the deweathered EC also showed a decreasing trend with a rate of -2.9% $year^{-1}$ (or -0.04 μg m^{-3} $year^{-1}$; Fig. S10), it was more than two times lower than the decreasing rate (-6.4 % $year^{-1}$) for $PM_{2.5}$. This suggests that the absolute primary particulate emissions from traffic have decreased over recent years, but the decreasing rate has been slower than that of other sources (e.g., power stations and industry (Zheng et al., 2018)) which contributed to a greater decrease in $PM_{2.5}$. The reduction in primary emissions from vehicles was confirmed by the decreasing trend (-3.1% $year^{-1}$) observed for the deweathered NO_x (Fig. 6d). The slightly lower decreasing rate for EC than NO_x (-2.9% $year^{-1}$ vs. -3.1% $year^{-1}$) suggests that vehicular particulate emissions were controlled in a less effective way than NO_x in recent years. Moreover, it is important to note that the decreasing trend in emissions from vehicles was not due to the decreasing number of vehicles on the road. This is corroborated by the fact that the normalized Ca, only showed a marginal decrease of -1.2% $year^{-1}$ (Fig. S10). According to the database from the Ministry of Transport, China, the number of privately owned motor vehicles in Shanghai increased from 2.4 million in 2016 to 3.2 million in 2019 (MinistryOfTransport). The larger decreasing trend observed for EC, compared to Ca, were likely due to the green technology and/or higher emission standards applied to the exhaust particulate emissions.

The fraction of EC in $PM_{2.5}$ showed an increasing rate of 2.5% $year^{-1}$ (Fig. 6) owing to the decreasing rate (-2.9% $year^{-1}$) of absolute EC concentration being roughly half of the decreasing rate of -6.4% $year^{-1}$ for $PM_{2.5}$. Thus, studying the direct particulate emission from traffic is important despite the decrease of traffic emission itself (but not as fast as other $PM_{2.5}$ sources). As shown in Figure 6, the fraction of EC in $PM_{2.5}$ had an increasing rate of 1.21% $year^{-1}$. The increasing rate of Ca in $PM_{2.5}$ at 6.1% $year^{-1}$, which is more than twice that of BC is more alarming. Therefore, road dust is projected to become even more

important in the near future as the number of vehicles on the road increases.

4 Discussion

Based on the random forest algorithm, we showed that the meteorologically normalized $PM_{2.5}$ decreased at a rate of -6.4% $year^{-1}$ from 2016 to 2019. This value was smaller than the values (-7.8% $year^{-1}$) reported in Beijing over 2013–2017 (Vu et al., 2019), although with different starting concentrations. Nevertheless, the results suggest that emission reductions in response to the Clean Air Act were less significant in Shanghai than in Beijing (Vu et al., 2019; Zhang et al., 2019; Zheng et al., 2018). As an important source of $PM_{2.5}$, vehicular exhaust emissions, i.e., EC, also decreased over the four years (2016–2019). The reduction in vehicular emissions was also confirmed by the trend analysis of the NO_x emissions. This is consistent with the implementation of aftertreatment technologies and the progressive tightening of emission standards (e.g., from China IV to China V) (Huang et al., 2022). A reduction in NO_x has also been observed in Europe (Grange et al., 2017). However, few studies have reported traffic-related EC trends based on high temporal resolution roadside monitoring data. To the best of our knowledge, this is the first study to report a long-term trend in EC near the highways.

We also showed that, despite various efforts to decrease vehicular exhaust emissions, road dust is becoming increasing important, with an increased contribution to $PM_{2.5}$ over the study periods and the predictable future under the current increase in the number of on-road vehicles (MinistryOfTransport). However, unlike EC or NO_x emissions, which can be controlled with aftertreatment technologies, road dust can be caused by passing vehicles, even for seemingly low-emission electric vehicles. According to a recent market report (Jin and He, 2019), the sales of electric vehicles has increased in recent years, with sales numbers reaching over 500,000 in China in 2017. In this study, we showed that, although switching to electric vehicles can reduce direct EC or NO_x emissions, road dust emissions do not seem to be abating at the same pace, which is likely to become increasingly important in the future. Using the machine learning algorithm, we showed that road dust, as suggested by the marker element of Ca, is inversely correlated with RH, with high Ca correlated with low RH levels. Therefore, as global warming may lead to a drier environment in certain locations (Byrne and O'Gorman, 2018), more road dust can be generated by future vehicles, which is a significant health concern for people living close to roadways. However, because road dust is sensitive to RH, policies on reducing road dust, such as street sweeping and washing, may help abate its emissions (Chang et al., 2018), in addition to technological advancements in road surface and tire designs.

5 Conclusion

In this study, hourly EC and Ca in $PM_{2.5}$ from highway sampling site in Shanghai were collected from 2016 to 2019, and were analyzed using a random forest-based machine learning algorithm. Four random forests were developed with satisfactory performances. Based on machine learning, the predicted values agreed well with the observed values, with R^2 >0.70. The effects of temperature and RH, as well as ws and wd on the observed values were revealed through the SHAP algorithm. After meteorological normalization, we showed that $PM_{2.5}$, decreased by -6.4% $year^{-1}$ over four years. As an important source of $PM_{2.5}$, vehicular exhaust emissions, i.e., EC, also decreased but at a rate slower than that of $PM_{2.5}$. Thus, direct particulate emissions from traffic were suggested to become increasingly important, despite the decrease in traffic emissions. Moreover, road dust, as opposed to engine exhaust, which cannot be directly controlled by aftertreatment technology, was suggested to be become increasingly important. The fraction of Ca in $PM_{2.5}$ showed an increasing rate of 6.1% $year^{-1}$, roughly twice that of BC. The results from this study suggest that road dust will not abate even with cleaner on-road traffic owing to

advances in aftertreatment technology and stricter emission standards. Such a trend may be exacerbated by global warming, leading to a drier atmosphere (i.e., low RH) in certain areas, inducing higher emissions of road dust from on vehicles.

Associate content

Supporting Information
Supplementary figures (Fig. S1-S10).

Credit authorship contribution statement

Meng Wang: designed the study, conducted data analysis, prepared the manuscript with contributions from all co-authors.
Yusen Duan: Investigation, Methodology, and Review.
Zhuozhi Zhang: Formal analysis, Writing, Review and Editing.
Juntao Huo: Investigation.
Yu Huang: Writing, Review and Editing.
Qingyan Fu: Review and Editing.
Tao Wang: Review and Editing.
Junji Cao: Review and Editing
Shun-cheng Lee: Writing - review and editing, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no conflicting interests.

Acknowledgements

This work was supported by the Environment and Conservation Fund-Environmental Research, Technology Demonstration and Conference Projects (ECF 63/2019), the RGC Theme-based Research Scheme (T24-504/17-N), the RGC Theme-based Research Scheme (T31-603/21-N).

References:

An Z, Huang R-J, Zhang R, Tie X, Li G, Cao J, et al. Severe haze in northern China: A synergy of anthropogenic emissions and atmospheric processes. Proceedings of the National Academy of Sciences 2019; 116: 8657.
Annavarapu RN, Kathi S. Cognitive disorders in children associated with urban vehicular emissions. Environmental Pollution 2016; 208: 74-78.
Ayodhya AS, Narayanappa KG. An overview of after-treatment systems for diesel engines. Environ Sci Pollut Res Int 2018; 25: 35034-35047.
Byrne MP, O'Gorman PA. Trends in continental temperature and humidity directly linked to ocean warming. Proceedings of the National Academy of Sciences 2018; 115: 4863-4868.
Carslaw DC, Ropkins K. openair — An R package for air quality data analysis. Environmental Modelling & Software 2012; 27-28: 52-61.

Cédric L, Goriaux M, Tassel P, Perret P, André M, Liu Y. Impact of Aftertreatment Device and Driving Conditions on Black Carbon, Ultrafine Particle and NOx Emissions for Euro 5 Diesel and Gasoline Vehicles. Transportation Research Procedia 2016; 14: 3079-3088.
Chang Y, Huang K, Xie M, Deng C, Zou Z, Liu S, et al. First long-term and near real-time measurement of trace elements in China's urban atmosphere: temporal variability, source apportionment and precipitation effect. Atmos. Chem. Phys. 2018; 18: 11793-11812.
Chen H, Kwong JC, Copes R, Tu K, Villeneuve PJ, van Donkelaar A, et al. Living near major roads and the incidence of dementia, Parkinson's disease, and multiple sclerosis: a population-based cohort study. The Lancet 2017; 389: 718-726.
Chen S, Zhang X, Lin J, Huang J, Zhao D, Yuan T, et al. Fugitive Road Dust PM(2.5) Emissions and Their Potential Health Impacts. Environ Sci Technol 2019; 53: 8455-8465.
Dabek-Zlotorzynska E, Celo V, Ding L, Herod D, Jeong C-H, Evans G, et al. Characteristics and sources of PM2.5 and reactive gases near roadways in two metropolitan areas in Canada. Atmospheric Environment 2019; 218: 116980.
Dai Q, Hou L, Liu B, Zhang Y, Song C, Shi Z, et al. Spring Festival and COVID-19 Lockdown: Disentangling PM Sources in Major Chinese Cities. Geophysical Research Letters 2021; 48: e2021GL093403.
DeWitt HL, Hellebust S, Temime-Roussel B, Ravier S, Polo L, Jacob V, et al. Near-highway aerosol and gas-phase measurements in a high-diesel environment. Atmospheric Chemistry and Physics 2015; 15: 4373-4387.
Fuzzi S, Baltensperger U, Carslaw K, Decesari S, Denier Van Der Gon H, Facchini M, et al. Particulate matter, air quality and climate: lessons learned and future needs. Atmospheric Chemistry and Physics 2015; 15: 8217-8299.
Gentner DR, Jathar SH, Gordon TD, Bahreini R, Day DA, El Haddad I, et al. Review of Urban Secondary Organic Aerosol Formation from Gasoline and Diesel Motor Vehicle Emissions. Environmental Science & Technology 2017; 51: 1074-1093.
Grange SK, Carslaw DC. Using meteorological normalisation to detect interventions in air quality time series. Science of The Total Environment 2019; 653: 578-588.
Grange SK, Carslaw DC, Lewis AC, Boleti E, Hueglin C. Random forest meteorological normalisation models for Swiss PM10 trend analysis. Atmos. Chem. Phys. 2018; 18: 6223-6239.
Grange SK, Lewis AC, Moller SJ, Carslaw DC. Lower vehicular primary emissions of NO2 in Europe than assumed in policy projections. Nature Geoscience 2017; 10: 914-918.
Halonen JI, Blangiardo M, Toledano MB, Fecht D, Gulliver J, Anderson HR, et al. Long-term exposure to traffic pollution and hospital admissions in London. Environmental Pollution 2016; 208: 48-57.
Huang H, Zhang J, Hu H, Kong S, Qi S, Liu X. On-road emissions of fine particles and associated chemical components from motor vehicles in Wuhan, China. Environmental Research 2022; 210: 112900.
Jia H, Huo J, Fu Q, Duan Y, Lin Y, Jin X, et al. Insights into chemical composition, abatement mechanisms and regional transport of atmospheric pollutants in the Yangtze River Delta region, China during the COVID-19 outbreak control period. Environmental Pollution 2020; 267: 115612.
Jia H, Pan J, Huo J, Fu Q, Duan Y, Lin Y, et al. Atmospheric black carbon in urban and traffic areas in Shanghai: Temporal variations, source characteristics, and population exposure. Environmental Pollution 2021; 289: 117868.
Jin L, He H. Comparison of the electric car market in China and the United States. https://theicct.org/sites/default/files/publications/ICCT_US-China_EV-mkt-%20comp_20190523.pdf (last access: April 01 2022), 2019.
Karner AA, Eisinger DS, Niemeier DA. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. Environmental Science & Technology 2010; 44: 5334-5344.
Kong S, Ji Y, Lu B, Chen L, Han B, Li Z, et al. Characterization of PM10 source profiles for fugitive dust in Fushun-a city famous for coal. Atmospheric Environment 2011; 45: 5351-5365.
Lin C, Ceburnis D, Xu W, Heffernan E, Hellebust S, Gallagher J, et al. The impact of traffic on air quality in Ireland: insights from the simultaneous kerbside and suburban monitoring of submicron aerosols. Atmos. Chem. Phys. 2020; 20: 10513-10529.
Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence 2020; 2: 56-67.
MinistryOfTransport. https://www.ceicdata.com/en/china/no-of-motor-vehicle-private-owned/cn-no-of-motor-vehicle-private-owned-shanghai (Last Access: 5 August 2022).
Mukherjee A, McCarthy MC, Brown SG, Huang S, Landsberg K, Eisinger DS. Influence of roadway

emissions on near-road PM2.5: Monitoring data analysis and implications. Transportation Research Part D: Transport and Environment 2020; 86: 102442.
Niu XY, Ho KF, Hu TF, Sun J, Duan J, Huang Y, et al. Characterization of chemical components and cytotoxicity effects of indoor and outdoor fine particulate matter (PM2.5) in Xi'an, China. ENVIRONMENTAL SCIENCE AND POLLUTION RESEARCH 2019; 26: 31913-31923.
Oukawa GY, Krecl P, Targino AC. Fine-scale modeling of the urban heat island: A comparison of multiple linear regression and random forest approaches. Science of The Total Environment 2022; 815: 152836.
Qin Y, Ye J, Ohno P, Liu P, Wang J, Fu P, et al. Assessing the Nonlinear Effect of Atmospheric Variables on Primary and Oxygenated Organic Aerosol Concentration Using Machine Learning. ACS Earth and Space Chemistry 2022.
Rappazzo KM, Daniels JL, Messer LC, Poole C, Lobdell DT. Exposure to Elemental Carbon, Organic Carbon, Nitrate, and Sulfate Fractions of Fine Particulate Matter and Risk of Preterm Birth in New Jersey, Ohio, and Pennsylvania (2000-2005). Environmental health perspectives 2015; 123: 1059-1065.
Reff A, Bhave PV, Simon H, Pace TG, Pouliot GA, Mobley JD, et al. Emissions Inventory of PM2.5 Trace Elements across the United States. Environmental Science & Technology 2009; 43: 5790-5796.
Sbihi H, Brook JR, Allen RW, Curran JH, Dell S, Mandhane P, et al. A new exposure metric for traffic-related air pollution? An analysis of determinants of hopanes in settled indoor house dust. Environmental Health 2013; 12: 48.
Shakya KM, Rupakheti M, Shahi A, Maskey R, Pradhan B, Panday A, et al. Near-road sampling of PM2. 5, BC, and fine-particle chemical components in Kathmandu Valley, Nepal. Atmos. Chem. Phys. 2017; 17: 6503-6516.
Shi Z, Song C, Liu B, Lu G, Xu J, Vu TV, et al. Abrupt but smaller than expected changes in surface air quality attributable to COVID-19 lockdowns. Science Advances 2021; 7: eabd6696.
Via M, Minguillón MC, Reche C, Querol X, Alastuey A. Increase in secondary organic aerosol in an urban environment. Atmos. Chem. Phys. 2021; 21: 8323-8339.
Vu TV, Shi Z, Cheng J, Zhang Q, He K, Wang S, et al. Assessing the impact of clean air action on air quality trends in Beijing using a machine learning technique. Atmos. Chem. Phys. 2019; 19: 11303-11314.
Wang J, Wu Q, Liu J, Yang H, Yin M, Chen S, et al. Vehicle emission and atmospheric pollution in China: problems, progress, and prospects. PeerJ 2019; 7: e6932-e6932.
Wang M, Zhang Z, Yuan Q, Li X, Han S, Lam Y, et al. Slower than expected reduction in annual PM2.5 in Xi'an revealed by machine learning-based meteorological normalization. Science of The Total Environment 2022; 841: 156740.
Wang Y, Wen Y, Wang Y, Zhang S, Zhang KM, Zheng H, et al. Four-Month Changes in Air Quality during and after the COVID-19 Lockdown in Six Megacities in China. Environmental Science & Technology Letters 2020.
Wu X, Vu TV, Shi Z, Harrison RM, Liu D, Cen K. Characterization and source apportionment of carbonaceous PM2.5 particles in China - A review. Atmospheric Environment 2018; 189: 187-212.
Zhang Q, Zheng Y, Tong D, Shao M, Wang S, Zhang Y, et al. Drivers of improved PM2.5 air quality in China from 2013 to 2017. Proceedings of the National Academy of Sciences 2019; 116: 24463-24469.
Zheng B, Tong D, Li M, Liu F, Hong C, Geng G, et al. Trends in China's anthropogenic emissions since 2010 as the consequence of clean air actions. Atmos. Chem. Phys. 2018; 18: 14095-14111.
Zhou W, Lei L, Du A, Zhang Z, Li Y, Yang Y, et al. Unexpected increases of severe haze pollution during the post COVID-19 period: effects of emissions, meteorology, and secondary production. Journal of Geophysical Research: Atmospheres 2022; n/a: e2021JD035710.
Zhu Y, Huang L, Li J, Ying Q, Zhang H, Liu X, et al. Sources of particulate matter in China: Insights from source apportionment studies published in 1987–2017. Environment International 2018; 115: 343-357.

Table 1. Annual mean, one standard deviation (SD), and median values for $PM_{2.5}$, EC, Ca, and NO_x concentration (in μg m^{-3}) before and after meteorological normalization (deweathered).

	$PM_{2.5}$		EC		Ca		NOx	
	observed	deweathered	observed	deweathered	observed	deweathered	observed	deweathered
Mean								
2016	52.3	53.6	1.57	1.63	0.163	0.174	58.4	62.9
2017	44.8	46.1	1.32	1.37	0.161	0.161	60.1	59.0
2018	43.6	44.2	1.41	1.42	0.151	0.158	50.9	54.5
2019	47.2	48.6	1.41	1.44	0.172	0.179	60.8	62.7
SD								
2016	35.3	26.4	1.26	0.995	0.194	0.133	61.4	47.0
2017	31.3	16.9	1.08	0.809	0.195	0.0894	56	36.1
2018	33.8	16.9	1.04	0.821	0.182	0.0784	48.9	32.0
2019	31.5	15.0	0.951	0.747	0.197	0.071	57.1	32.2
Median								
2016	42.0	54.6	1.20	1.59	0.105	0.159	40.0	55.1
2017	37.0	44.6	1.05	1.28	0.104	0.142	42.0	52.0
2018	34.0	44.5	1.11	1.35	0.0998	0.147	37.0	48.8
2019	38.0	47.6	1.14	1.38	0.113	0.168	45.0	56.5

Table 2. Statistics including mean, SD, median, 25th percentile (Q1), and 75th percentile (Q3) for the meteorological variables over the four years.

	mean	SD	median	Q1	Q3
ws (m s^{-1})	2.04	1.29	1.9	1.1	2.8
RH (%)	75.8	18.7	79	64	92
air_temp (°C)	17	9.19	17.6	9.5	24.3
pressure (hPa)	1017	9.15	1017	1009	1024
rainfall (mm)	0.14	0.09	0	0	0

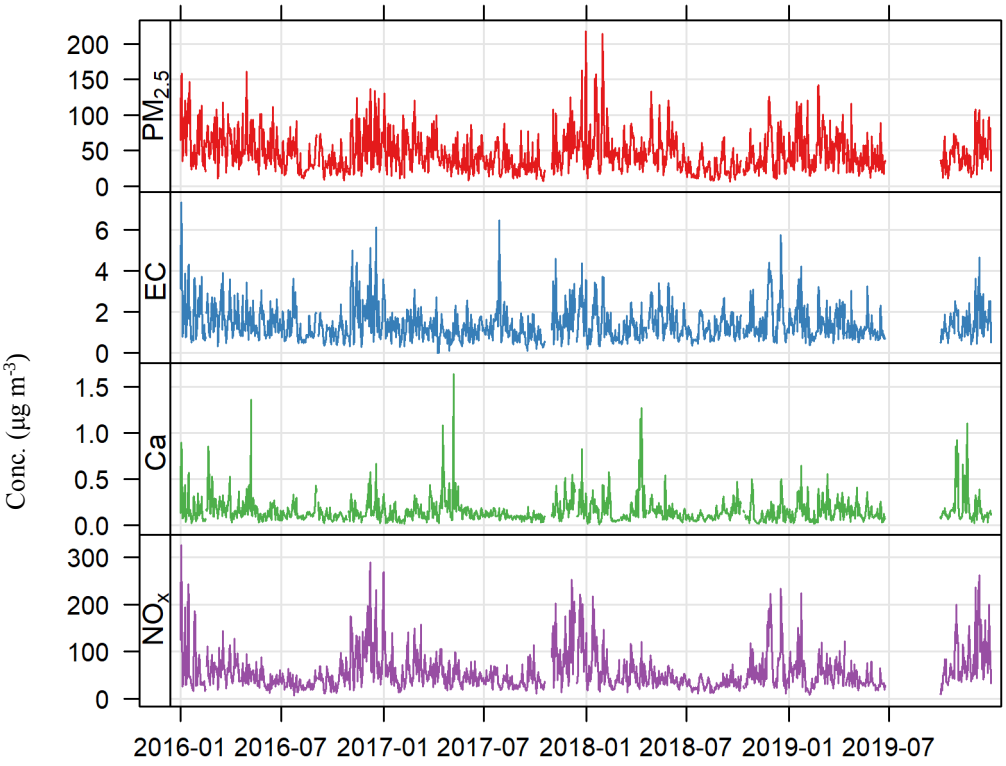


Figure 1. Time series of $PM_{2.5}$, elemental carbon (EC), Ca, and NO_x from 2016 to 2019. Data are daily averaged from the original hourly resolution.

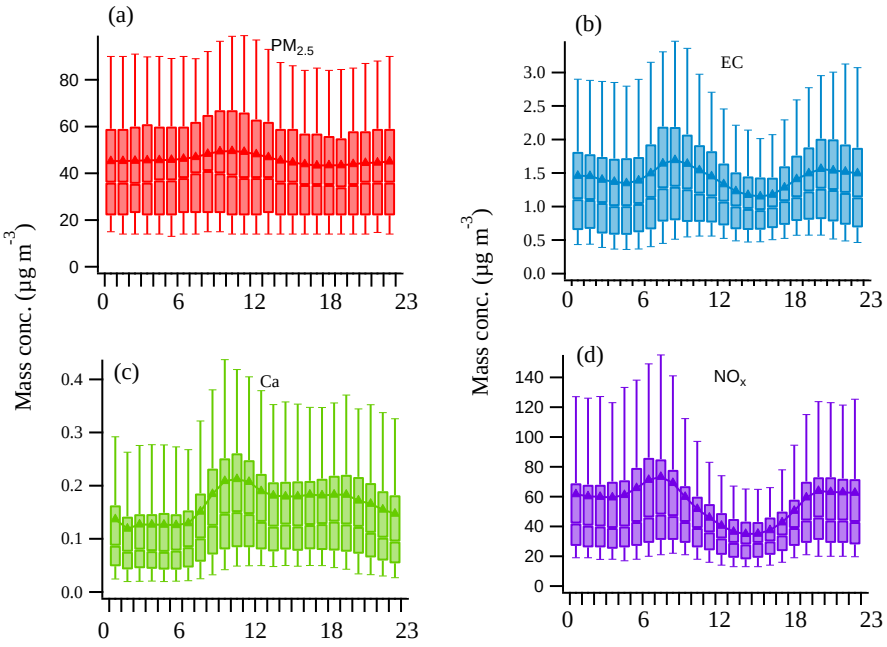


Figure 2. Diurnal cycle of (a) $PM_{2.5}$, (b) EC, (c) Ca, and (d) NO_x. Box plots show the 5th, 25th, median, 75^{th,} and 95th percentile. The triangle inside the box represents the mean value over the four years.

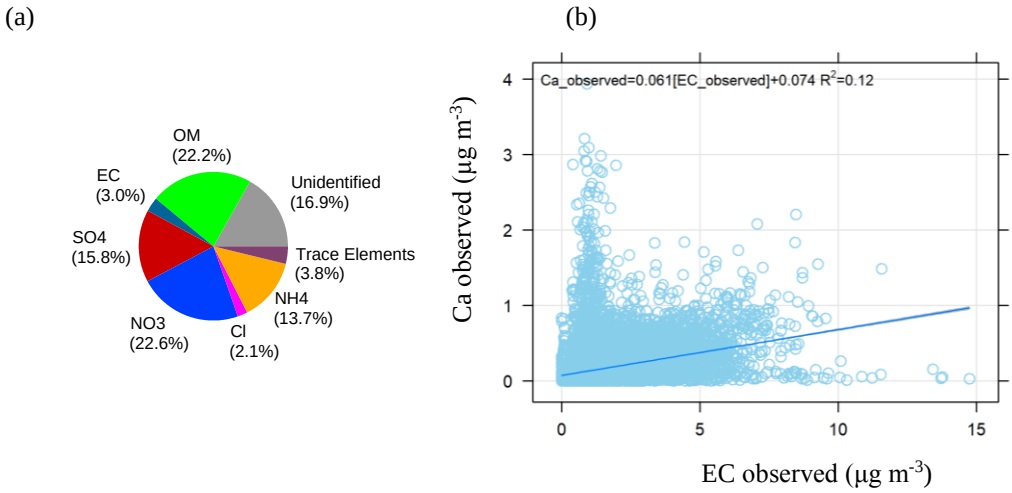


Figure 3. (a) Average chemical composition of $PM_{2.5}$ including organic matter (OM = 1.8 × OC), EC, sulfate (SO_4), nitrate (NO_3), chloride (Cl), ammonium (NH_4), and trace elements; the unidentified part is the gap between the measured and rebuilt PM2.5. (b) scatter plot between the observed Ca and EC (in µg m^{-3}) over the four years, showing poor coefficient of determination R^2 of 0.12.

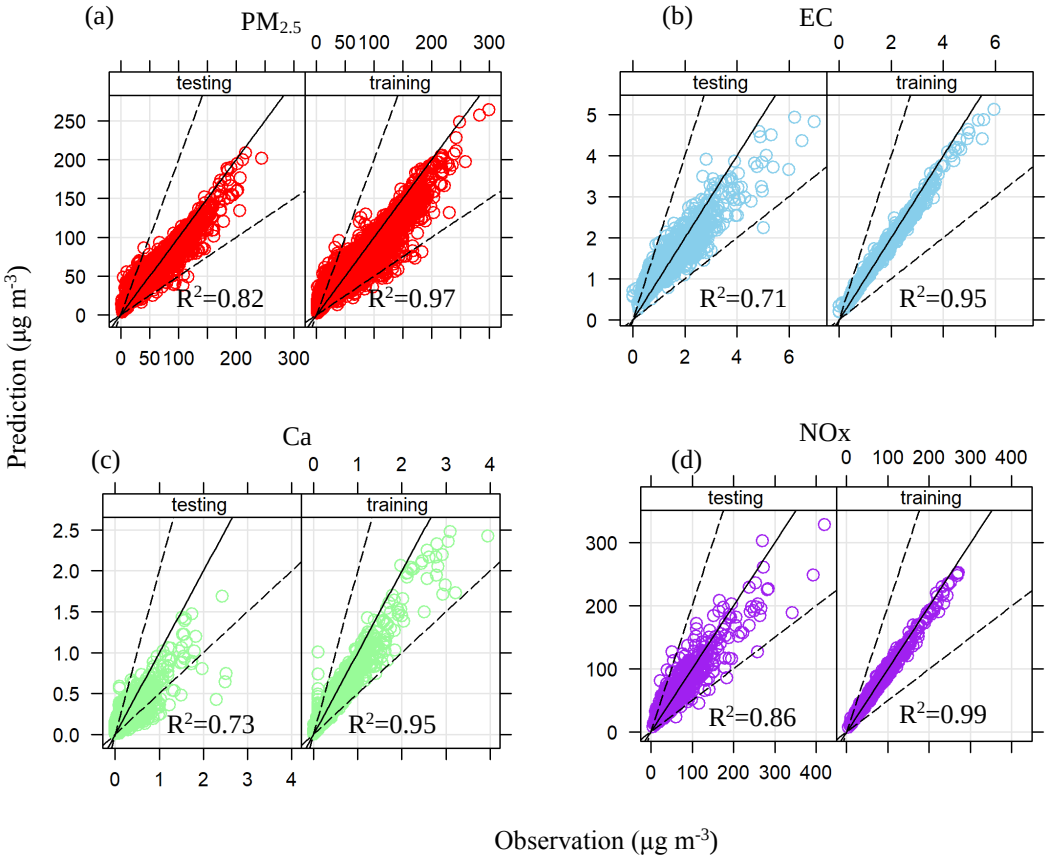


Figure 4. Correlation between the predicted and observed values (in µg m^{-3}) during the build of random forest for (a) $PM_{2.5}$, (b) EC, (c) Ca, and (d) NO_x.

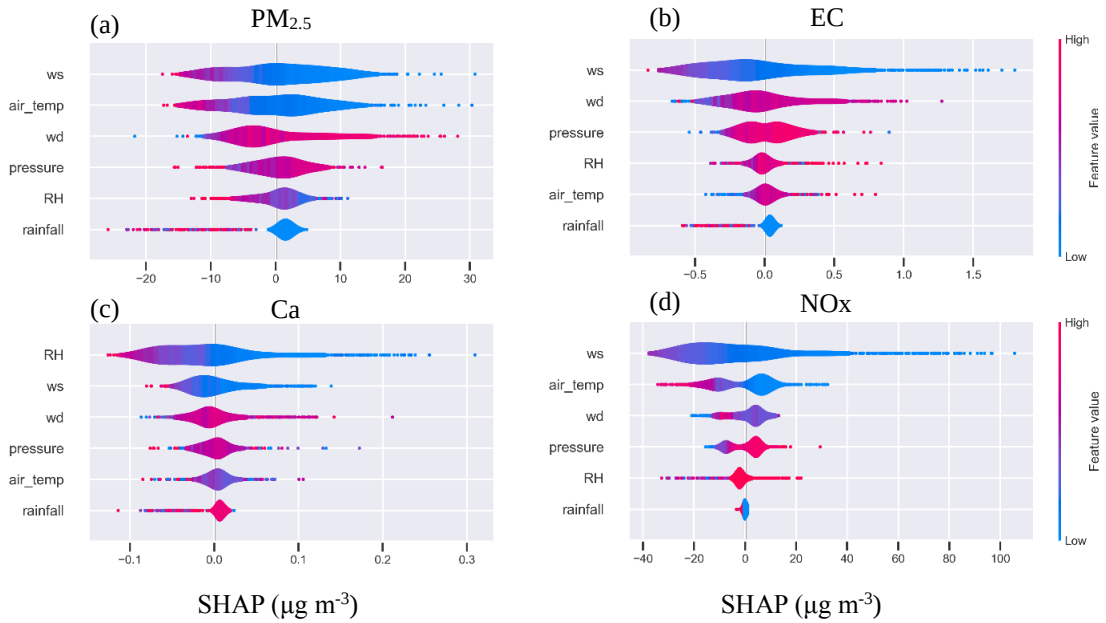


Figure 5. SHAP values for the analysis of the importance of meteorological variable for the random forest built for (a) $PM_{2.5}$, (b) EC, (c) Ca, and (d) NO_x.

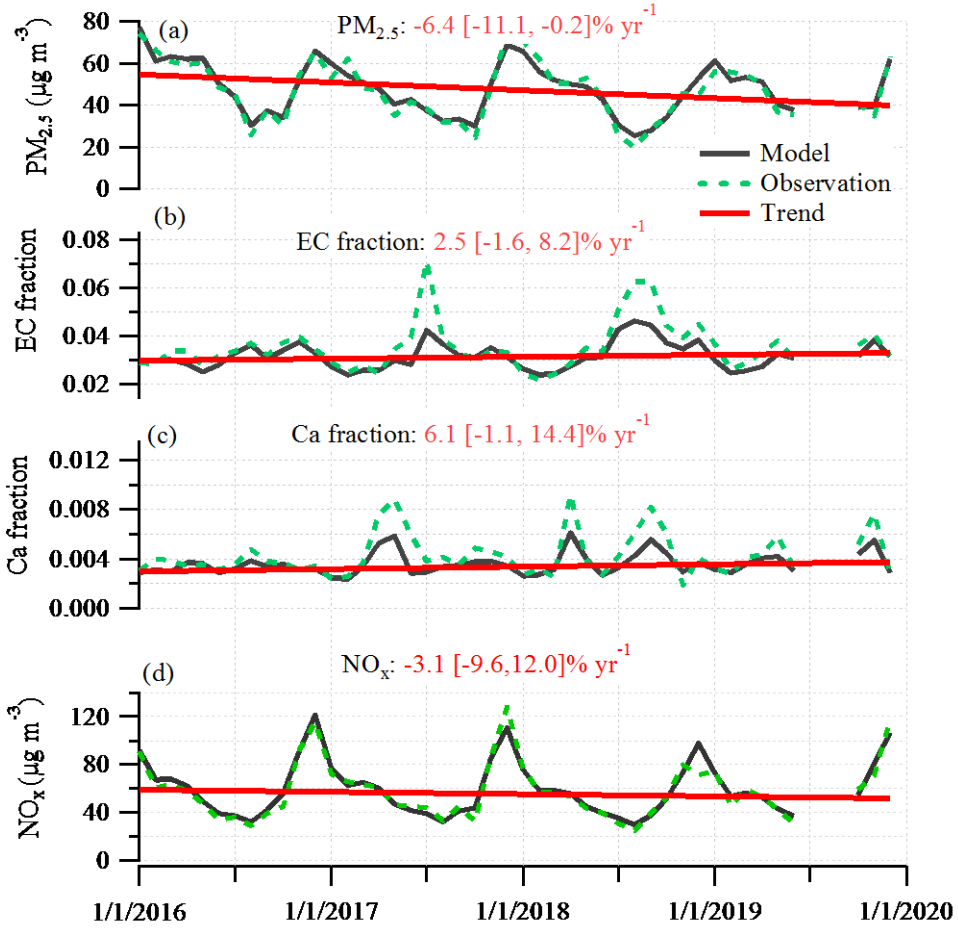


Figure 6. Monthly average (a) $PM_{2.5}$ (μg m^{-3}) (b) fraction of EC in $PM_{2.5}$; and (c) fraction of Ca in $PM_{2.5}$, and (d) NO_x before (i.e., observation) after meteorological normalization (i.e., model). The red line represents the trend analysis of $PM_{2.5}$ using the Theil-Sen estimator, with values above representing the modelled trend and the 95% confidence level.

Author statement

Meng Wang: designed the study, conducted data analysis, prepared the manuscript with contributions from all co-authors.
Yusen Duan: Investigation, Methodology, and Review.
Zhuozhi Zhang: Formal analysis, Writing, Review and Editing.
Juntao Huo: Investigation.
Yu Huang: Writing, Review and Editing.
Qingyan Fu: Review and Editing.
Tao Wang: Review and Editing.
Junji Cao: Review and Editing
Shun-cheng Lee: Writing - review and editing, Funding acquisition, Supervision.