Integrating augmented in-situ measurements and a spatiotemporal machine learning model to back extrapolate historical particulate matter pollution over the United Kingdom: 1980-2019

Riyang Liu[1,2,3], Zongwei Ma[1], Antonio Gasparrini[4], Arturo de la Cruz[4], Jun Bi[1], Kai Chen[2,3*]*

[1] State Key Laboratory of Pollution Control and Resource Reuse, School of the Environment, Nanjing University, Nanjing 210023, China

[2] Department of Environmental Health Sciences, Yale School of Public Health, New Haven, Connecticut 06520, United States

[3] Yale Center on Climate Change and Health, Yale School of Public Health, New Haven, Connecticut 06520, United States

[4] Environment & Health Modelling (EHM) Lab, Department of Public Health Environments and Society, London School of Hygiene & Tropical Medicine, London WC1H 9SH, U.K.

*** Corresponding Authors**

Abstract

Historical $PM_{2.5}$ data are essential for assessing the health effects of air pollution exposure across the life course or early life. However, a lack of high-quality data sources, such as satellite-based aerosol optical depth before 2000 has resulted in a gap in spatiotemporally resolved $PM_{2.5}$ data for historical periods. Taking the United Kingdom as an example, we leveraged the light gradient boosting model to capture the spatiotemporal association between $PM_{2.5}$ concentrations and multi-source geospatial predictors. Augmented $PM_{2.5}$ from PM_{10} measurements expanded the spatiotemporal representativeness of the ground measurements. Observations before and after 2009 were used to train and test the models respectively. Our model showed fair prediction accuracy from 2010 to 2019 [the ranges of coefficients of determination (R^2) for the grid-based cross-validation are 0.71-0.85] and commendable back extrapolation performance from 1998 to 2009 (the ranges of R^2 for the independent external testing are 0.32-0.65) at the daily level. The pollution episodes in the 1980s and pollution levels in the 1990s were also reproduced by our model. The 4-decade $PM_{2.5}$ estimates demonstrated that most regions in England witnessed significant downward trends in $PM_{2.5}$ pollution. The methods developed in this study are generalizable to other data-rich regions for historical air pollution exposure assessment.

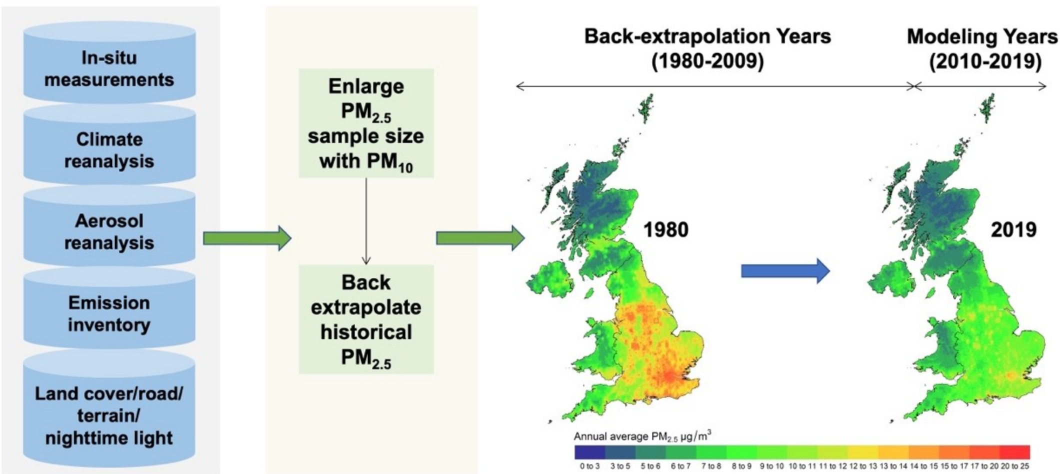
Keywords

$PM_{2.5}$, LightGBM, back extrapolation, U.K., SHAP, spatiotemporal patterns, exposure analysis

Synopsis

This study provides a generalizable case for estimating long-term spatiotemporal-resolved $PM_{2.5}$ estimates for life-course or early life exposure to air pollution in the U.K.

Abstract Graphic



1. Introduction

Extensive scientific evidence across disciplines has demonstrated that both short- and long-term exposure to fine particles with an aerodynamic diameter smaller than 2.5 μm ($PM_{2.5}$) is associated with a broad range of adverse health effects, including cardiovascular, respiratory and neurological effects, with varying severity at different stages of life[1-3]. To prevent the morbidity and mortality of these diseases, more detailed evidence is needed about the heterogeneity of the associations across sites and periods[4]. Long-term historical $PM_{2.5}$ data are essential to support such spatial and temporal exposure analyses. However, $PM_{2.5}$ in-situ measurements were scarce before the late 2000s even in developed countries like the United Kingdom [5-7]. Besides, partly due to lack of high quality model input like satellite-based aerosol optical depth (AOD)[8, 9], many long-term global[10-12], Europe-wide[8] or nationwide [6, 13, 14] $PM_{2.5}$ models only went back to around 2000, making it hard to assess early life or life-course exposure.

Although recent studies have attempted to extend the time span of $PM_{2.5}$ models to several decades, there are some important limitations. First, studies based on the atmospheric chemistry transport model (ACTM), which simulated air pollutant concentrations over several decades with surrogate meteorological input data[5, 15] were designed to evaluate policy effects rather than to reproduce actual historical pollution levels. Second, studies based on statistical models, that used long-term ground visibility observations as input to back extrapolate $PM_{2.5}$ concentrations [16, 17], were limited by the spatial coverage and uncertainty of the visibility data. Specifically, visibility data are limited by their relative inaccuracy in high values and inconsistency as they shifted from human observers to automated sensors [17, 18]. Third, the time span of the training data set in some previous statistical exposure studies was less than 3 years[17, 19], which could hardly capture the interannual difference in air pollution levels. Lastly, many studies estimated $PM_{2.5}$ concentrations at coarse spatiotemporal resolutions (e.g., 0.25°×0.25°[20] and annual mean[16], which could not produce spatiotemporal resolved exposure metrics based on different exposure durations.

Therefore, it is challenging to back extrapolate long-term spatiotemporally resolved $PM_{2.5}$ concentrations without high-quality satellite based AOD products and simulations from ACTMs. The U.K. has more than 20 years of regulatory monitoring in $PM_{2.5}$ and high-quality multi-source geospatial data sets that could reflect the historical variations of $PM_{2.5}$ pollution, making it a good example to investigate the method of back extrapolation in data-rich regions. In this study, we aim to utilize an advanced machine learning algorithm, the light gradient boosting model (LightGBM)[21], to capture reliable long-term spatiotemporal associations between daily $PM_{2.5}$ concentrations and multi-source geographical predictors in the U.K. The model is validated with cross validation (CV), external testing, and comparison to previous studies. We then derive a series of high-resolution (1×1 km) data sets for daily prediction of $PM_{2.5}$ from 1980 to 2019 and discuss the spatiotemporal patterns of $PM_{2.5}$ pollution.

2. Materials and Methods

2.1 Data Preparation

2.1.1 Study Area and Period

Our study includes the four countries of the U.K., namely, England, Wales, Scotland, and Northern Ireland, as well as the self-governing Isle of Man. A fishnet containing 245052 1 km grid cells was created to cover the whole study area (Figure S1) based on the Ordnance Survey National Grid.

The boundary data used in this study were from the U.K. government and are licensed under the Open Government License, version 3.0. We estimated the $PM_{2.5}$ concentrations from January 1, 1980 to December 31, 2019 as a result of data availability, which was described in detail below.

2.1.2 In-Situ Monitored Data

Measurements of hourly $PM_{2.5}$ and PM_{10} concentrations were obtained from seven monitoring network sources in the U.K. Automatic Urban and Rural Network (AURN), Air Quality England network, Air Quality Wales network, Air Quality Scotland network, Northern Ireland network, King's College London (KCL) network and locally managed AQ networks in England (hereafter referred to as "local networks"). We used R package openair[22] to download $PM_{2.5}$ observations from 1998 to 2019 and PM_{10} observations from 2010 to 2019. PM_{10} observations before 2009 were not included in the back extrapolation of historical $PM_{2.5}$ data due to the poor results of a preliminary analysis that attempted to augment the historical $PM_{2.5}$ measurements with PM_{10} observations from 1992 to 2009. We define the former five network sources as national networks and the latter two network sources as regional networks, depending upon whether they are part of the national monitoring strategy of the U.K. All of the observations from the national networks have been ratified[23] before download and used for model development, validation and testing. Observations from regional networks were not combined with those from the national networks because they may not be fully comparable. We used the observations from the regional networks for the external model testing to demonstrate the performance of our model on the best available data sets despite the regional networks' limited spatial coverage.

Monitors with less than 18-h records were excluded when aggregating to daily average PM concentrations. The observations from different national networks in the same coordinates were in good agreement; we thus chose observations from AURN, the largest automatic monitoring network, for further analysis.

Measurements of $PM_{2.5}$ started in 1998 and had not been widespread until 2010[6]. In national networks, there were 196 co-located stations measuring both PM_{10} and $PM_{2.5}$, 25 $PM_{2.5}$-only stations, 174 PM_{10}-only stations from 2010 to 2019, and 72 $PM_{2.5}$ stations from 1998 to 2009 (see Figure S2). Regional networks have fewer and unevenly distributed stations, with 60 $PM_{2.5}$ stations from 2010 to 2019 and 14 $PM_{2.5}$ stations from 2001 to 2009 (Figure S3). All observations were assigned with a grid-cell ID. Mean values were calculated if a grid cell had more than one monitor. Grids with less than 7-day records per month and 9 months per year were excluded.

2.1.3 Auxiliary Predictors

Auxiliary predictors used in this study include meteorological factors, aerosol reanalysis, emission inventory, land cover data, road network, terrain data, anthropogenic activities (see Table S1 and Text S1 for details about data sources and preparations), and spatiotemporal weights. We utilized spatiotemporal weights to incorporate spatiotemporal heterogeneity and hidden predictors, such as the transboundary transport of pollutants from continental Europe which contributes significantly to $PM_{2.5}$ pollution in the U.K.[24] , as a previous study did[19]. The spatial weights were represented by the geographic distances to the four corners and the center of a rectangle around our study area using the Euclidean distance (see details in Figure S4). The temporal information was represented by the order of a day in a week and the time intervals to the middle of each season like a previous study did[25] (see details in Text S2 and Table S2).

2.2 Model Development and Validation

A two-stage model was developed to capture the long-term spatiotemporal association between $PM_{2.5}$ concentrations and multi-source predictors, as is shown in the upper panel in Figure 1. Each stage is described in detail below. In brief, stage 1 used co-located PM_{10} measurements to construct a model to augment $PM_{2.5}$ observations. Stage 2 used the LightGBM algorithm with the fusion of the original $PM_{2.5}$ observations and augmented $PM_{2.5}$ values to back extrapolate historical $PM_{2.5}$ concentrations. We chose LightGBM, which has been used in several previous studies[26-28], as the workhorse in our study for its strength in faster computation speed, lower memory consumption, and capability of handling big data when compared with other advanced algorithms like extreme gradient boosting[21] (see more details in Text S3). The model development was conducted with R package mlr3[29] and lightgbm[30].

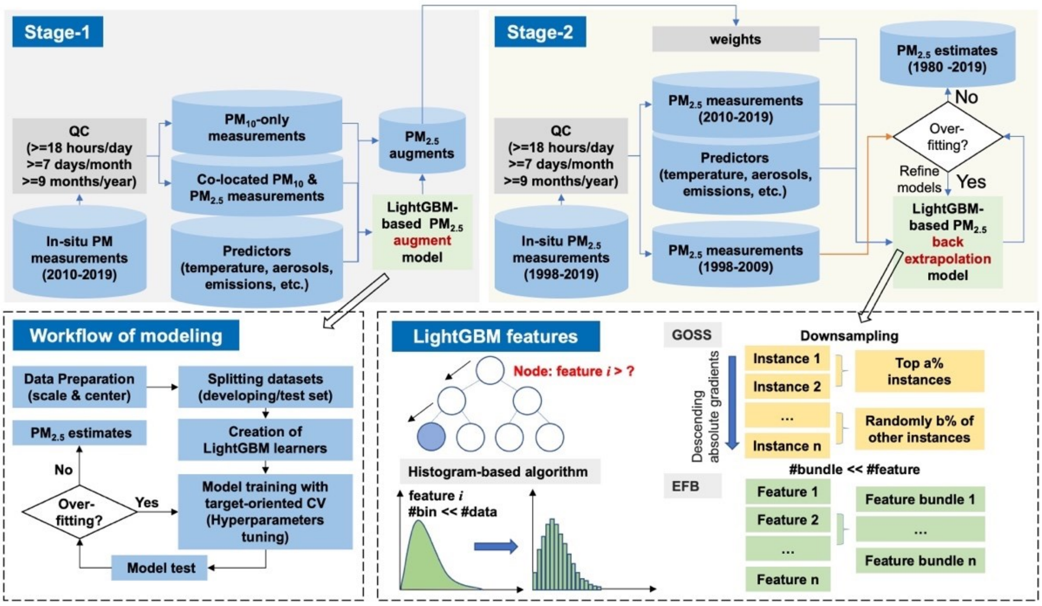


Figure 1. Schematics of the model developed in this study (upper panel), the workflow of modeling (left bottom panel), and optimization features of the LightGBM algorithm (right bottom panel). QC, quality control; LightGBM, the light Gradient Boosting model; CV: cross validation; GOSS: gradient-based one-side sampling; EFB, exclusive feature bundling; an instance means a data sample; a feature means a predictor variable; #bin, the number of bins; #data, the number of data samples; #bundle, the number of feature bundles; #feature, the number of features.

2.2.1 Stage 1: Augmenting $PM_{2.5}$ Observations Using Co-located PM_{10} Measurements

PM_{10} measurements are more widely distributed than $PM_{2.5}$ in the U.K.[6], as is shown in Figure S2. Stage 1 aims to improve the spatiotemporal distribution of data samples in the stage 2 model with PM_{10} observations. In this case, the spatiotemporal representativeness of the data samples will be enhanced, which could reduce the bias.

The workflow of modeling is shown in the left bottom panel of Figure 1. Correlation analysis was performed between the pollutant concentrations and the predictor variables and between each pair of predictor variables, respectively. The predictor variables with a lower correlation coefficient within paired predictors whose correlation coefficients were greater than 0.70 were excluded to mitigate the multicollinearity problem that could lead to overfitting[31, 32]. All of the predictors were scaled and centered before being fed into the models. All of the co-located PM_{10} and $PM_{2.5}$ data sets were used as the development set (see more details in Text S4).

There were 10 hyperparameters to tune in the LightGBM-based $PM_{2.5}$ augment model. Because the target of stage 1 is to estimate $PM_{2.5}$ concentrations in locations where only PM_{10} measurements were available, which is about spatial extrapolation, a target-oriented CV strategy, 10-fold grid-based CV (it was referred to as "spatial CV" in previous studies[12, 33]) was used to determine the optimal vector of hyperparameters. Data samples were divided into 10 groups randomly based on their grid IDs; i, e., samples from the same grid cell would not be split. In each iteration, nine groups of data were used as training data, while the other data were held out for validation. The training and validation process was repeated 10 times until the data of each group had been validated. Root mean square error (RMSE) was used as the loss function. We randomly compared 100 vectors of hyperparameters in this study, and the values of hyperparameters were shown in Table S3. Statistical indicators including the coefficient of determination (R^2), RMSE and mean absolute error (MAE) were calculated to demonstrate the model performance.

2.2.2 Stage 2: Back Extrapolating Historical $PM_{2.5}$

$PM_{2.5}$ augments derived from stage 1 could not simply be treated as ground observations for their uncertainty. Therefore, weights were needed to treat the original $PM_{2.5}$ measurements and augment differently to enhance the spatiotemporal representativeness of data samples without hurting the data quality. We used the RMSE with sample weights as the loss function during the tuning process, as shown in Equation 1. For data samples from original $PM_{2.5}$ measurements, we set the weight to 1, and for augments, we chose the weight from 0, 0.1, 0.3, 0.5, and 0.7 based on the model performance.

$$weighted\ RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} w_i\ (t_i - r_i)^2} \qquad \text{(Equation 1)}$$

where n is the number of samples, t_i and r_i represent the ground measurements (truth) and the prediction (response) of the model of a data sample i, respectively, and w_i represents the weight of a data sample i.

The workflow of the stage 2 model was similar to that of the stage 1 model. The differences lay in the predictors selected, splitting data sets, CV strategy, and assessment of the model performance.

A total of 10 years of data (from 2010 to 2019) from national networks were used to train and validate the models. Another target-oriented CV strategy, 10-fold by-year CV, which has been used in our previous study[34], was used to determine the optimal vector of hyperparameters for a reliable historical estimator. In this case, data samples were divided into 10 groups randomly based on the calendar year. We randomly compared 100 vectors of hyperparameters in this study; the values of hyperparameters were also shown in Table S3. Observations from 1998 to 2009 from both national networks and regional networks were used to test the spatiotemporal generalization capability of the models in years when only few regulatory measurements were available. Because some observations from the national networks from 1998 to 2009 were collected at stations that were also included in the development set from 2010 to 2019, the model performance could be overoptimistic if the observations from 1998 to 2009 were used directly as the testing set. Therefore, we use a spatiotemporal testing strategy by extending the grid-based CV method. Specifically, all of the observations from national networks were randomly divided into 10 groups based on their grid IDs. In each iteration, nine groups of data samples from 2010 to 2019 were used as training data, while data samples from 1998 to 2009 in the other group were kept for testing. This process mimics the prediction of historical $PM_{2.5}$ levels at locations not covered by monitors. Only the original $PM_{2.5}$ observations were used to calculate the CV results for comparison among models with different weights. We also applied another stricter spatiotemporal testing strategy, called 100 km grid-based CV. All of the observations from national networks were assigned to 100 km grids before being randomly divided into 10 groups based on 100 km grid IDs. This process mimics the prediction of $PM_{2.5}$ levels in the past at locations that are more than 100 km away from monitors. There are 28 agglomerations (large urban areas) and 16 non-agglomeration zones in the study region, which were divided for the purpose of assessing air quality compliance[23, 35]. R^2 values between daily $PM_{2.5}$ estimates and observations in each zone were calculated to show the difference in the model performance in urban and non-urban areas. Simulations from the European Monitoring and Evaluation Programme for U.K. model (EMEP4UK), a Eulerian model developed over the British Isles[13, 14], were used as a benchmark to explore how well our predictions could capture the temporal variability of in-situ measurements. For years before and around 1998, the statistics of $PM_{2.5}$ measurements were extracted from previous studies to test the reliability of the model. All of the observations from the development set from 2010 to 2019 were used to train the final estimator.

2.3 Interpretation of Models

Complex machine learning models are often considered "black box" models[36, 37]. To mitigate the effects of this lack of transparency on model credibility[37, 38], we applied two interpretation tools, feature importance and Shapley additive explanation (SHAP) [39, 40], to our models to explain how the models make predictions. Specifically, feature importance values were estimated using the intrinsic LightGBM gain method, which represent the total reduction in training loss gained when using a feature to split the data[21] and reflect the impact of a predictor on model performance. SHAP, which has been incorporated into LightGBM[41], can distribute individualized contribution of each predictor to the difference that each prediction deviates from the base value[39], as shown in Equation 2. SHAP has been used in previous studies[42, 43] to help explain the major driving factors of certain pollution levels.

$$f(x) = \emptyset_0(f) + \sum_{j=1}^{M} \emptyset_j(f, x) \qquad \text{(Equation 2)}$$

where $f(x)$ is the model output of a data sample x, $\emptyset_0(f)$ is the base value for the model output, M is the total number of predictors, $\emptyset_j(f, x)$ is the contribution of predictor j for a data sample x.

2.4 Spatiotemporal Patterns and Population Exposure Analysis

We hindcast the historical $PM_{2.5}$ concentration at a resolution of 1 km with the final estimator and derived the decadal, annual, and seasonal metrics of $PM_{2.5}$ pollution in the study period. Spatial patterns of pollution were identified based on the prediction maps. We also analyzed the trends in $PM_{2.5}$ pollution during the whole period based on the monthly average to avoid the relatively high uncertainty of daily estimates. $PM_{2.5}$ anomalies were derived by subtracting the long-term averages in the same month of the 4 decades from the monthly means in every grid cell and then calculating the linear trends for each grid cell and subregions with the least-squares approaches as a previous study did[44]. $PM_{2.5}$ estimates were matched with gridded population data to calculate the number of people exposed to specific levels of $PM_{2.5}$ pollution by year in the U.K. The groups of $PM_{2.5}$ concentrations were divided based on recommendations from the World Health Organization[45].

3. Results

3.1 Results of Augmenting $PM_{2.5}$ Observations Using Co-located PM_{10} Measurements

As shown in Figure S5, the overall value of R^2 for the grid-based CV was 0.91 at the day level, and the corresponding RMSE was 2.41 μg/m^3. During the period of the stage 1 model (2010-2019), the values of R^2 ranged from 0.88-0.93, with corresponding RMSE ranging from 1.88 to 3.05 μg/m^3, and MAE ranging from 1.18 to 2.20 μg/m^3 (see details in Table S4). PM_{10} was the most important predictor in the stage 1 model, playing a dominant role in both model predictions and model performance (see Figure S6 for details).

The stage 1 model was used to increase the sample size in the stage 2 model. After stage 1, the number of data samples increased by 118% (from 272216 to 592707), and the number of grid cells with data samples increased by 85% (from 226 to 417). The augmentation of $PM_{2.5}$ has significantly increased sample sizes outside of England, with 79, 72, and 61% of the data samples from Northern Ireland, Scotland, and Wales, respectively, coming from the stage 1 model.

3.2 Results of Back Extrapolating Historical $PM_{2.5}$

Stage 2 models were developed based on different weights to select the final weight for $PM_{2.5}$ augments. According to the testing results shown in Figure S7, the model with a weight of 0.3 showed the most robust performance. The difference in model performance between the model with a weight of 0.3 and the model with a weight of 0 revealed the improvement that the stage 1 model brought to our study.

According to the density scatterplots of the 10-fold by-year CV results (the upper panels in Figure 2), the values of R^2 were 0.72, 0.82, and 0.81 at the daily, monthly, and annual levels, respectively, and the corresponding RMSE values were 4.34, 2.13, and 1.42 μg/m^3. Table S5 showed that the ranges of R^2 and RMSE for the CV results are 0.63-0.78 and 3.73-5.36 μg/m^3 respectively at the daily level from 2010 to 2019.

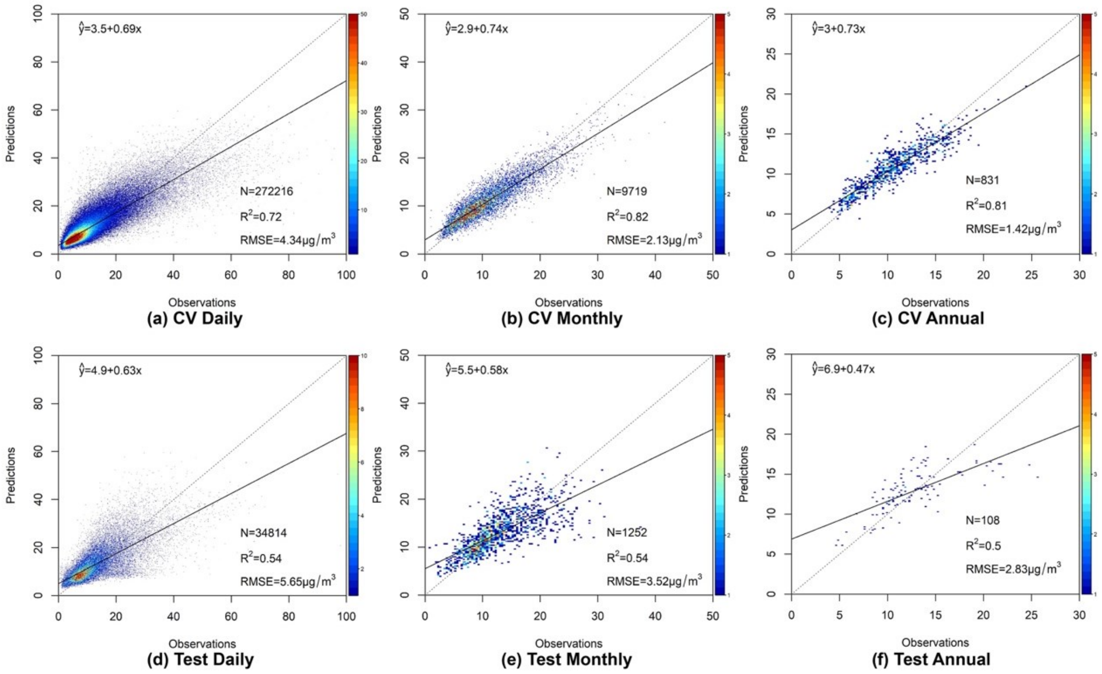


Figure 2. Density scatterplots of the by-year CV results for the stage 2 model at (a) daily, (b) monthly, and (c) annual levels from 2010 to 2019 and the testing results at (d) daily, (e) monthly and (f) annual levels from 1998 to 2009

The values of R^2 for the testing result were 0.54, 0.54, and 0.50 at the daily, monthly, and annual levels, respectively, the corresponding RMSE values were 5.65, 3.52, and 2.83 μg/m^3 (the bottom panels in Figure 2). Table S6 showed that the ranges of R^2 and RMSE for the spatiotemporal testing at the daily level are 0.32-0.65 and 5.05-7.73 μg/m^3, respectively, from 1998 to 2009. The model performance shows a subtle decline back in time, which demonstrates that our historical predictions are reliable and robust. The model evaluation using the 100 km grid-based CV strategy in Table S7 showed comparable performance to that using the 1 km grid-based spatiotemporal CV, reflecting the robustness of our model.

The R^2 values between daily average $PM_{2.5}$ estimates and observations in 44 zones and agglomerations for the development set and testing set were shown in Figure S8. Densely populated urban agglomerations had better performance in both data sets than rural areas, with R^2 values for the development set larger than 0.70. North Wales showed the worst performance over the study period.

The time series plot of estimated and observed monthly $PM_{2.5}$ concentrations from 1998 to 2009 (Figure S9) demonstrated that our model could capture the long-term trends in $PM_{2.5}$ pollution in different subregions, with correlation coefficients larger than 0.7. However, an obvious overestimation occurred in the spring of 2003 in England. We selected four sites with more than 1000 observations before 2010 to compare the predictions from our model and the simulations from EMEP4UK, namely, London Bloomsbury (urban background), London Marylebone Road (urban traffic), Rochester Stoke (South East, rural background) and Harwell (South East, rural background). The time series plots in Figure S10 indicate that our model performed better in background sites than in the traffic site. The predictions in this study were better correlated with measurements than the simulations. Overestimation also occurred in 2003 in the time series of the simulations, which would be discussed in section 4.2.

Although observations from the regional networks may not be fully comparable to those from the national networks, our models had comparable performance on the data set from regional networks according to Figure S11 and Table S8. The ranges of R^2 and RMSE for the testing of KCL networks at the daily level are 0.42-0.77 and 5.98-13.22 μg/m^3 respectively from 2001 to 2009. For the local networks, the ranges of R^2 and RMSE for the testing at the daily level are 0.31-0.66 and 3.48-6.52 μg/m^3 respectively from 2002 to 2009. The correlations between the regional average of monthly mean $PM_{2.5}$ estimates and measurements were larger than 0.77 in subregions and periods, as shown in Figure S12, which were also comparable to those in Figure S10.

Table S9 shows the statistics of observed $PM_{2.5}$ concentrations extracted from previous studies and predictions produced in our study. The measurements were collected in Leeds (West Yorkshire, England), Birmingham (West Midlands, England), London, Rochester Stoke, Harwell, and Edinburgh (southeastern Scotland). The comparison shows that the model well reproduced the concentration levels in Birmingham, London, Rochester Stoke, Harwell, and Edinburgh in the 1990s. The model tended to be better at predicting period averages than at predicting peaks. Although the model did not perform well in predicting the absolute pollution levels in Leeds in the 1980s, it showed the same peak periods and peak dates of $PM_{2.5}$ pollution episodes in Figure S13 compared to the in-situ measurements[46].

3.3 Interpretation of Back Extrapolating Historical $PM_{2.5}$

Aerosol reanalysis data, boundary layer height, wind speed, temperature, and spatiotemporal terms were the most important predictors in the stage 2 model in terms of both model performance and prediction attribution (see Figure 3 for details). Both black carbon and sulfate, the two most important predictors, made robust contributions to $PM_{2.5}$ concentrations from 1998 to 2009, as shown in Figure S14. The SHAP dependence plots of wind in Figure S15 showed the spatial heterogeneity in the contributions of wind to $PM_{2.5}$ concentrations, reflecting the different effects of clean air and polluted air; e.g., a westerly wind often reduces $PM_{2.5}$ concentrations, with a greater magnitude in the west, reflecting the cleansing effects of air from the west (Wales or the Atlantic). Conversely, an easterly wind often increases $PM_{2.5}$ concentrations, also with a greater magnitude in the west, reflecting the transport of air pollutants from the east (England or continental Europe). The interpretations based on feature importance and SHAP values showed that our model is consistent with domain knowledge[47].

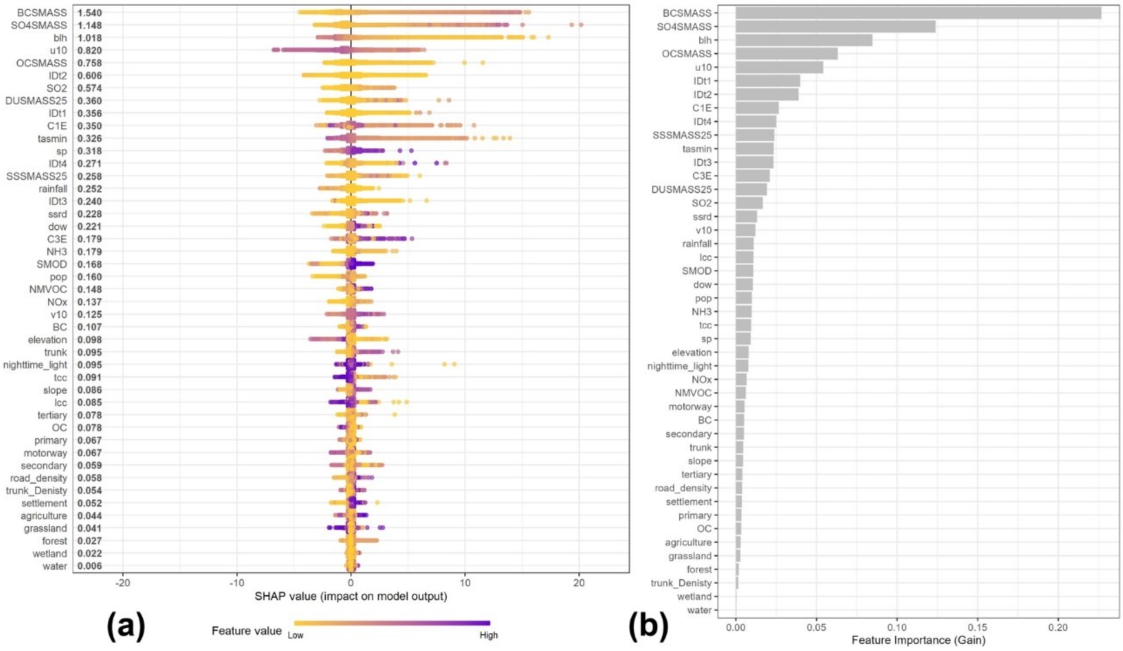


Figure 3. The interpretation of the stage 2 model with (a) the SHAP summary plot for $PM_{2.5}$ predictions in the development set which excludes augmented $PM_{2.5}$ and (b) feature importance of the predictors in relative percentage. The numbers next to the vertical axis in panel a represent the mean absolute SHAP value by predictor variable. In panel a, each dot in each row represents a data sample, where the x position of each dot is the effect of a predictor variable on the prediction of a model (i.e., the predicted $PM_{2.5}$ concentration of that data sample), and the color of the dot represents the value of that predictor variable. Dots that do not fit on the row are stacked to show density.

3.4 Spatial Patterns of $PM_{2.5}$ Pollution in the U.K.

The spatial distribution of decadal average $PM_{2.5}$ estimates in the U.K. from 1980 to 2019 (Figure 4) revealed strong spatial and temporal variation in $PM_{2.5}$ pollution. $PM_{2.5}$ concentrations were higher in England than in other subregions over the 4 decades, with areas with relatively high $PM_{2.5}$ pollution (annual average of > 10 μg/m^3 [45, 48]) concentrated in urban agglomerations in England, such as Greater London, Birmingham, Manchester, etc. The relatively higher concentrations in southeastern background areas shown in Figure 4 and Figure S16 were partly due to the transboundary transport of pollutants from continental Europe, as previous studies revealed[23, 35, 49]. The spatial distribution of annual mean $PM_{2.5}$ anomalies (using the averages in each grid over the entire period as the baseline) in Figure S17 clearly showed that $PM_{2.5}$ concentrations in the U.K. had decreased significantly over the whole study period despite significant fluctuations in some particular years, such as 1996, 2003 and 2011. The winter and spring months had the largest areas of pollution, while the summer months had cleaner ambient air, as shown in Figure S18. The spatiotemporal patterns of back-extrapolated $PM_{2.5}$ were very similar to in-situ measurements, as shown in Figure S19.

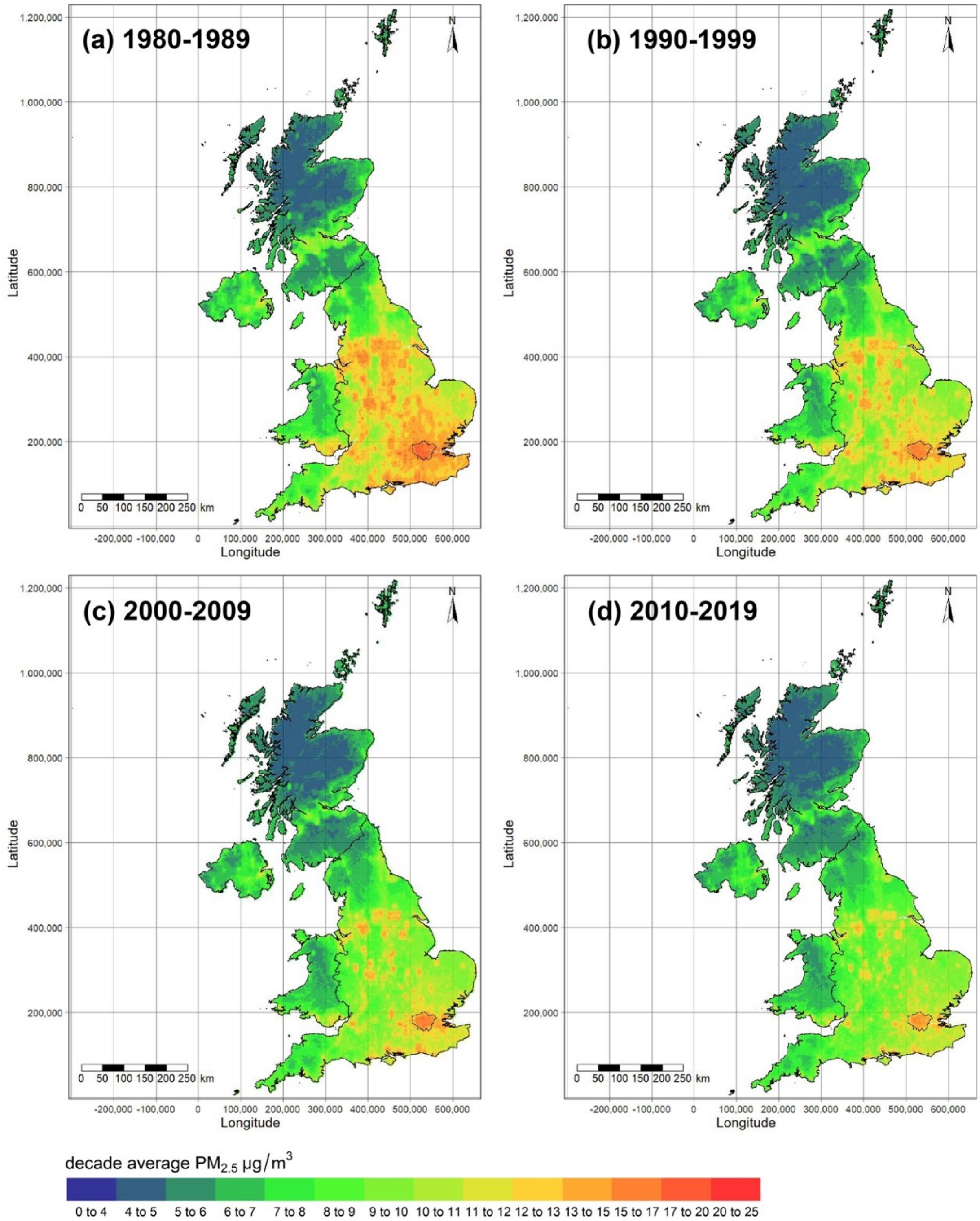


Figure 4. Spatial distribution of decadal average $PM_{2.5}$ estimations in the U.K. from 1980 to 2019

3.5 Trends of $PM_{2.5}$ Pollution in the U.K.

The gridded monthly mean $PM_{2.5}$ anomaly trends in Figure 5 present that most areas in the U.K. showed significantly downward trends in $PM_{2.5}$ pollution over the study period. England showed the most rapid decrease among all of the subregions, with the fastest rate of decline of more than 0.15 µg/m^3 per year. Areas with upward trends were scarce and only distributed in low-concentration areas. Some of the least polluted areas, such as the Highland and Outer Hebrides in Scotland, had increased $PM_{2.5}$ concentrations with no significant trends.

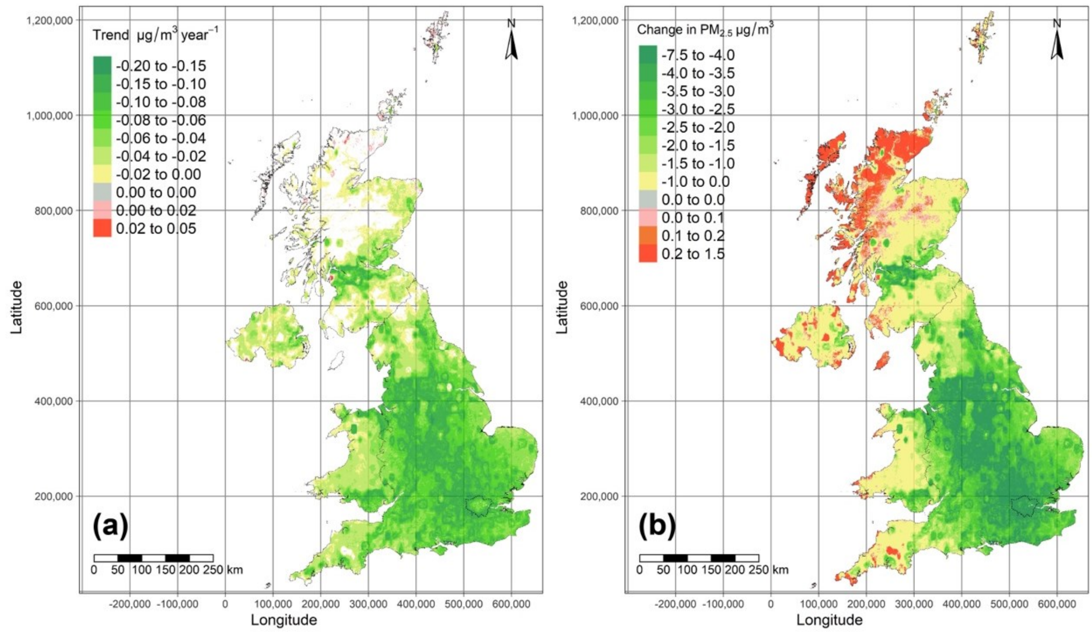


Figure 5. Spatial distribution of the (a) monthly mean $PM_{2.5}$ anomaly trends and (b) changes in annual $PM_{2.5}$ concentrations from 1980 to 2019. The white areas in panel a indicate the significance level $p \geq 0.05$.

$PM_{2.5}$ concentrations in England had been significantly declining all over the study period, with a faster rate of decline up to 0.12 μg/m^3 per year in the first 2-decade period (1980-1999) than in the second 2-dacade period (2000-2019). Scotland, Wales, and Northern Ireland had a much slower rate of decline and only witnessed significant downward trends from 1980 to 1999, as shown in Figure S20 and Table S10.

3.6 Population Exposure

Figure S21 shows the number of people exposed to specific levels of $PM_{2.5}$ pollution by year. The annual average of $PM_{2.5}$ concentrations was seldom larger than 20 μg/m^3 in the U.K., as shown in Figure S21a. The proportion of people who were exposed to $PM_{2.5}$ greater than 20 μg/m^3 was usually less than 0.05%, except for 0.07% in 1982 and 0.12% in 2003. Most people lived in areas where the annual average ranged from 10 to 15 μg/m^3 over the study periods. The changes in the proportion of people living in areas with $PM_{2.5}$ concentrations above 10 μg/m^3 were ranged from 67.00 % in 2019 to 92.39% in 2003. The threat to the population from long-term $PM_{2.5}$ exposure decreased during the study period. Figure S21b showed a more fluctuated time series over years, which indicated that short-term $PM_{2.5}$ pollution episodes still posed a severe threat to population health in the U.K..

4. Discussion

4.1 Strengths and Innovations

Our study exhibits several strengths and innovations. First, we incorporated in-situ $PM_{2.5}$ measurements from seven monitoring networks and estimated $PM_{2.5}$ concentrations at PM_{10} monitoring sites to enhance the spatiotemporal representativeness of the training data samples as much as possible. To balance the data quantity and quality, we selected a weight for augmented $PM_{2.5}$ samples based on trials and errors. To better capture the historical trends, the time span of the training data was set at 10 years, longer than that used in previous studies[17, 19, 20]. Second, we collected recently available multi-source geospatial data sets to represent drivers or spatial proxies for $PM_{2.5}$ pollution to compensate for the role of satellite based AOD. An advanced tree-based ensemble algorithm, LightGBM, combined with target-oriented CV strategies, was used to efficiently capture the nonlinear and high-order relationship between these predictors and $PM_{2.5}$ concentrations. Third, we adopted a comprehensive testing strategy, comprising independent external testing and comparison with statistics from previous studies, to evaluate the back extrapolation capability of the model during the years with few regulatory monitors (1998-2009) and the years when $PM_{2.5}$ measurements were extremely scarce (before 2000). Fourth, we used interpretation methods such as feature importance and SHAP to peer into the LightGBM model, which showed that our model is in good agreement with domain science. Lastly, we obtained historical daily continuous $PM_{2.5}$ pollution levels at a resolution of 1 km over 4 decades in the U.K., which is one of the first to the best of our knowledge.

4.2 Comparison to Previous Studies

Schneider et al. reconstructed daily $PM_{2.5}$ concentrations at horizontal resolution of 1 × 1 km across Britain from 2008 to 2018 using year-specific satellite-based machine learning models, which performed well, with overall CV R^2 for the models ranging from 0.704 to 0.821 and RMSE ranging from 3.275 to 4.547 μg/m^3 [6]. Our model showed comparable performance in the modeling years when using the grid-based CV strategy (the ranges of R^2 and RMSE for the CV results are 0.71-0.85 and 3.04-4.73 μg/m^3, respectively, at the daily level from 2010 to 2019, see details in Table S11), indicating that the vector of hyperparameters tuned by the by-year CV strategy could also capture the spatial variations of $PM_{2.5}$ pollution in the modeling years.

The spatiotemporal patterns of $PM_{2.5}$ pollution derived from the predictions in this study were also consistent with findings from previous studies. The pollution hotspots were clustered in urban areas in England, which was also found in previous studies[6, 50]. The downward trends of $PM_{2.5}$ were greater before the 2000s than those in the early years of the 21st century, which was also summarized in another study focusing on NO_2 pollution. The reason was attributed to increasing NO_x emissions from road traffic[5].

The overestimation in the spring of 2003 in England could be partly attributed to relatively high concentrations of $PM_{2.5}$ composition from aerosol reanalysis data (Figure S22), which were among the most important predictor variables in terms of prediction attribution, as shown in Figure S23. The peaks of $PM_{2.5}$ also occurred in the ACTM simulations, as shown in Figure S10. We are not sure whether the overestimation of our predictions and the simulations was biased because ground observations were scarce. The year 2003 was recorded as a high pollution year for PM_{10} [7], and nitrate and SO_2 emissions were also high in 2003[51], therefore, the reasons for the discrepancy need further careful investigation.

4.3 Limitations

This study has some limitations. First, the way to determine the values of the weights was based on trials and errors instead of theoretical analysis of the characteristics of the data samples. Since the training samples are high-dimensional, new approaches are needed to determine which part of the augments contributes more to the model performance. Second, evidence of the reliability of the model prior to 2000 was relatively sparse, consisting of statistics or sporadic samples. We did not use in-situ measurements of PM_{10}, black smoke, visibility data, and gas pollutants like SO_2 before 2000 to estimate the historical trends of $PM_{2.5}$ in this study because of their inconsistency in monitoring techniques[18] and locations. As a next step, we could try to figure out more patterns of $PM_{2.5}$ pollution from these observations.

4.4 Implications

The methods developed in this study which fuse long-term in-situ measurements and various geospatial factors could be applied to other regions with abundant long-term data, such as the United States and Western Europe. More in-situ observations, such as meteorological factors and black smoke could be further incorporated to assist in capturing the historical trends.

The predictions derived in this study could benefit health effect studies in the U.K. in several ways. First, spatiotemporally resolved $PM_{2.5}$ estimates could be aggregated to various exposure metrics (e.g., seasonal mean, and the 99th percentile of the annual distribution of the 24-h average) depending upon different study objectives. Second, our robust historical estimates over 4 decades could be combined with long-term cohorts in the U.K. to assess the life-course or early exposure of participants to air pollution. Third, the model performed better in densely populated urban agglomerations whereas ACTMs often have the highest uncertainty level in urban areas[49], making predictions from our study a good input for epidemiological studies focusing on urban populations.

Data Availability

Data and code to replicate all results in the main text and supplementary materials are available upon request.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.3c05424

- Additional model details about model inputs, model tuning and model assessments, including interpretations of the model predictions, assessment of augmented $PM_{2.5}$ and historical $PM_{2.5}$ estimates, model performance, spatiotemporal patterns of $PM_{2.5}$ pollution, trends in aerosol reanalysis and exposure analysis of $PM_{2.5}$ pollution (PDF)

Notes

The authors declare no competing financial interest.

Acknowledgements

The work of Riyang Liu, Zongwei Ma, and Jun Bi was supported by the National Natural Science Foundation of China (Grants. 71761147002 and 72234003). The data acquisition was supported by the DataHub Support Team at the Pacific Northwest National Laboratory (PNNL) and Centre for Environmental Data Analysis (CEDA) Support in the U.K.

References

This article references 51 other publications.

(1) U.S. EPA. *Supplement to the 2019 Integrated Science Assessment for Particulate Matter (Final*

Report, 2022); EPA/635/R-22/028; U.S. Environmental Protection Agency, Washington, D.C., 2022. https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=354490.

(2) Braithwaite, I.; Zhang, S.; Kirkbride, J. B.; Osborn, D. P. J.; Hayes, J. F. Air Pollution (Particulate Matter) Exposure and Associations with Depression, Anxiety, Bipolar, Psychosis and Suicide Risk: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives* **2019**, *127* (12), 126002. DOI: https://doi.org/10.1289/EHP4595.

(3) Al-Kindi, S. G.; Brook, R. D.; Biswal, S.; Rajagopalan, S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nature Reviews Cardiology* **2020**, *17* (10), 656-672. DOI: https://doi.org/10.1038/s41569-020-0371-2.

(4) Power, M. C.; Lamichhane, A. P.; Liao, D.; Xu, X.; Jack, C. R.; Gottesman, R. F.; Mosley, T.; Stewart, J. D.; Yanosky, J. D.; Whitsel, E. A. The Association of Long-Term Exposure to Particulate Matter Air Pollution with Brain MRI Findings: The ARIC Study. *Environmental Health Perspectives* **2018**, *126* (2), 027009. DOI: https://doi.org/10.1289/EHP2152.

(5) Carnell, E.; Vieno, M.; Vardoulakis, S.; Beck, R.; Heaviside, C.; Tomlinson, S.; Dragosits, U.; Heal, M. R.; Reis, S. Modelling public health improvements as a result of air pollution control policies in the UK over four decades—1970 to 2010. *Environmental Research Letters* **2019**, *14* (7), 074001. DOI: https://doi.org/10.1088/1748-9326/ab1542.

(6) Schneider, R.; Vicedo-Cabrera, A. M.; Sera, F.; Masselot, P.; Stafoggia, M.; de Hoogh, K.; Kloog, I.; Reis, S.; Vieno, M.; Gasparrini, A. A Satellite-Based Spatio-Temporal Machine Learning Model to Reconstruct Daily PM2.5 Concentrations across Great Britain. *Remote Sensing* **2020**, *12* (22), 3803. DOI: https://doi.org/10.3390/rs12223803.

(7) Harrison, R. M.; Pope, F. D.; Shi, Z. Trends in Local Air Quality 1970–2014. In *Still Only One Earth: Progress in the 40 Years Since the First UN Conference on the Environment*, The Royal Society of Chemistry, 2015; pp 58-106.

(8) Shen, Y.; de Hoogh, K.; Schmitz, O.; Clinton, N.; Tuxen-Bettman, K.; Brandt, J.; Christensen, J. H.; Frohn, L. M.; Geels, C.; Karssenberg, D.; et al. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. *Environment International* **2022**, *168*, 107485. DOI: https://doi.org/10.1016/j.envint.2022.107485.

(9) Ma, Z.; Dey, S.; Christopher, S.; Liu, R.; Bi, J.; Balyan, P.; Liu, Y. A review of statistical methods used for developing large-scale and long-term PM2.5 models from satellite data. *Remote Sensing of Environment* **2022**, *269*, 112827. DOI: https://doi.org/10.1016/j.rse.2021.112827.

(10) Hammer, M. S.; van Donkelaar, A.; Li, C.; Lyapustin, A.; Sayer, A. M.; Hsu, N. C.; Levy, R. C.; Garay, M. J.; Kalashnikova, O. V.; Kahn, R. A.; et al. Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998–2018). *Environmental Science & Technology* **2020**, *54* (13), 7879-7890. DOI: https://doi.org/10.1021/acs.est.0c01764.

(11) van Donkelaar, A.; Hammer, M. S.; Bindle, L.; Brauer, M.; Brook, J. R.; Garay, M. J.; Hsu, N. C.; Kalashnikova, O. V.; Kahn, R. A.; Lee, C.; et al. Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. *Environmental Science & Technology* **2021**, *55* (22), 15287-15300. DOI: https://doi.org/10.1021/acs.est.1c05309.

(12) Yu, W.; Ye, T.; Zhang, Y.; Xu, R.; Lei, Y.; Chen, Z.; Yang, Z.; Zhang, Y.; Song, J.; Yue, X.; et al. Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study. *The Lancet Planetary Health* **2023**, *7* (3), e209-e218. DOI: https://doi.org/10.1016/S2542-5196(23)00008-6.

(13) Vieno, M.; Dore, A. J.; Wind, P.; Marco, C. D.; Nemitz, E.; Phillips, G.; Tarrasón, L.; Sutton, M. A.

Application of the EMEP Unified Model to the UK with a Horizontal Resolution of 5 × 5 km2. In *Atmospheric Ammonia: Detecting emission changes and environmental impacts*, Sutton, M. A., Reis, S., Baker, S. M. H. Eds.; Springer Netherlands, 2009; pp 367-372.

(14) Lin, C.; Heal, M. R.; Vieno, M.; MacKenzie, I. A.; Armstrong, B. G.; Butland, B. K.; Milojevic, A.; Chalabi, Z.; Atkinson, R. W.; Stevenson, D. S.; et al. Spatiotemporal evaluation of EMEP4UK-WRF v4.3 atmospheric chemistry transport simulations of health-related metrics for NO2, O3, PM10, and PM2. 5 for 2001–2010. *Geoscientific Model Development* **2017**, *10* (4), 1767-1787. DOI: https://doi.org/10.5194/gmd-10-1767-2017.

(15) Russ, T. C.; Cherrie, M. P. C.; Dibben, C.; Tomlinson, S.; Reis, S.; Dragosits, U.; Vieno, M.; Beck, R.; Carnell, E.; Shortt, N. K.; et al. Life Course Air Pollution Exposure and Cognitive Decline: Modelled Historical Air Pollution Data and the Lothian Birth Cohort 1936. *Journal of Alzheimer's Disease* **2021**, *79*, 1063-1074. DOI: https://doi.org/10.3233/JAD-200910.

(16) Kim, S.-Y.; Olives, C.; Sheppard, L.; Sampson, P. D.; Larson, T. V.; Keller, J. P.; Kaufman, J. D. Historical Prediction Modeling Approach for Estimating Long-Term Concentrations of PM2.5 in Cohort Studies before the 1999 Implementation of Widespread Monitoring. *Environmental Health Perspectives* **2017**, *125* (1), 38-46. DOI: https://doi.org/10.1289/EHP131.

(17) Liu, M.; Bi, J.; Ma, Z. Visibility-Based PM2.5 Concentrations in China: 1957–1964 and 1973–2014. *Environmental Science & Technology* **2017**, *51* (22), 13161-13169. DOI: https://doi.org/10.1021/acs.est.7b03468 From NLM Medline.

(18) Singh, A.; Bloss, W. J.; Pope, F. D. 60 years of UK visibility measurements: impact of meteorology and atmospheric pollutants on visibility. *Atmospheric Chemistry and Physics* **2017**, *17* (3), 2085-2101. DOI: https://doi.org/10.5194/acp-17-2085-2017.

(19) Araki, S.; Shima, M.; Yamamoto, K. Estimating historical PM2.5 exposures for three decades (1987–2016) in Japan using measurements of associated air pollutants and land use regression. *Environmental Pollution* **2020**, *263*, 114476. DOI: https://doi.org/10.1016/j.envpol.2020.114476.

(20) Zhong, J.; Zhang, X.; Gui, K.; Liao, J.; Fei, Y.; Jiang, L.; Guo, L.; Liu, L.; Che, H.; Wang, Y.; et al. Reconstructing 6-hourly PM2.5 datasets from 1960 to 2020 in China. *Earth System Science Data* **2022**, *14* (7), 3197-3211. DOI: https://doi.org/10.5194/essd-14-3197-2022.

(21) Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. Eds.; Vol. 30; Curran Associates, Inc., 2017.

(22) Carslaw, D. C.; Ropkins, K. openair — An R package for air quality data analysis. *Environmental Modelling & Software* **2012**, *27-28*, 52-61. DOI: https://doi.org/10.1016/j.envsoft.2011.09.008.

(23) Brookes, D.; Bush, T.; Cooke, S.; Eaton, S.; Fraser, A.; Grice, S.; Griffin, A.; Kent, A.; Loader, A.; Martinez, C.; et al. *Air Pollution in the UK 2010*; London, 2011. https://uk-air.defra.gov.uk/library/annualreport/viewonline?year=2010_issue_3#report_pdf.

(24) Air Quality Expert Group. *Fine Particulate Matter (PM 2.5) in the United Kingdom*; Department for Environment, Food and Rural Affairs; Scottish Executive; Welsh Government; and Department of the Environment in Northern Ireland, London, 2012. https://uk-air.defra.gov.uk/assets/documents/reports/cat11/1212141150_AQEG_Fine_Particulate_Matter_in_the_UK.pdf (accessed 2022-03-28).

(25) Wei, J.; Liu, S.; Li, Z.; Liu, C.; Qin, K.; Liu, X.; Pinker, R. T.; Dickerson, R. R.; Lin, J.; Boersma, K. F.; et al. Ground-Level NO2 Surveillance from Space Across China for High Resolution Using

Interpretable Spatiotemporally Weighted Artificial Intelligence. *Environmental Science & Technology* **2022**, *56* (14), 9988-9998. DOI: https://doi.org/10.1021/acs.est.2c03834.

(26) Wei, J.; Li, Z.; Pinker, R. T.; Wang, J.; Sun, L.; Xue, W.; Li, R.; Cribb, M. Himawari-8-derived diurnal variations in ground-level PM2.5 pollution across China using the fast space-time Light Gradient Boosting Machine (LightGBM). *Atmospheric Chemistry and Physics* **2021**, *21* (10), 7863-7880. DOI: https://doi.org/10.5194/acp-21-7863-2021.

(27) Zhong, J.; Zhang, X.; Gui, K.; Wang, Y.; Che, H.; Shen, X.; Zhang, L.; Zhang, Y.; Sun, J.; Zhang, W. Robust prediction of hourly PM2.5 from meteorological data using LightGBM. *National Science Review* **2021**, *8* (10). DOI: https://doi.org/10.1093/nsr/nwaa307.

(28) Zhang, S.; Mi, T.; Wu, Q.; Luo, Y.; Grieneisen, M. L.; Shi, G.; Yang, F.; Zhan, Y. A data-augmentation approach to deriving long-term surface SO2 across Northern China: Implications for interpretable machine learning. *Science of The Total Environment* **2022**, *827*, 154278. DOI: https://doi.org/10.1016/j.scitotenv.2022.154278.

(29) Lang, M.; Binder, M.; Richter, J.; Schratz, P.; Pfisterer, F.; Coors, S.; Au, Q.; Casalicchio, G.; Kotthoff, L.; Bischl, B. mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software* **2019**, *4* (44), 1903. DOI: https://doi.org/10.21105/joss.01903.

(30) *lightgbm: Light Gradient Boosting Machine*; 2022. https://CRAN.R-project.org/package=lightgbm (accessed 2022-03-28).

(31) Gu, Y.; Li, B.; Meng, Q. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing* **2022**, *468*, 123-136. DOI: https://doi.org/10.1016/j.neucom.2021.09.051.

(32) Betancourt, C.; Stomberg, T. T.; Edrich, A. K.; Patnala, A.; Schultz, M. G.; Roscher, R.; Kowalski, J.; Stadtler, S. Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties. *Geoscientific Model Development* **2022**, *15* (11), 4331-4354. DOI: https://doi.org/10.5194/gmd-15-4331-2022.

(33) Xiao, Q.; Chang, H. H.; Geng, G.; Liu, Y. An Ensemble Machine-Learning Model To Predict Historical PM2.5 Concentrations in China from Satellite Data. *Environmental Science & Technology* **2018**, *52* (22), 13260-13269. DOI: https://doi.org/10.1021/acs.est.8b02917.

(34) Liu, R.; Ma, Z.; Liu, Y.; Shao, Y.; Zhao, W.; Bi, J. Spatiotemporal distributions of surface ozone levels in China from 2005 to 2017: A machine learning approach. *Environment International* **2020**, *142*, 105823. DOI: https://doi.org/10.1016/j.envint.2020.105823.

(35) Buckland, T.; Bush, T.; Eaton, S.; Kilroy, E.; Kent, A.; Loader, A.; Morris, R.; Norris, J.; Stedman, J.; Vincent, K.; et al. *Air Pollution in the UK 2014*; London, 2015. https://uk-air.defra.gov.uk/library/annualreport/viewonline?year=2014_issue_1#report_pdf.

(36) Liu, X.; Lu, D.; Zhang, A.; Liu, Q.; Jiang, G. Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. *Environmental Science & Technology* **2022**, *56* (4), 2124-2133. DOI: https://doi.org/10.1021/acs.est.1c06157.

(37) Zhong, S.; Zhang, K.; Bagheri, M.; Burken, J. G.; Gu, A.; Li, B.; Ma, X.; Marrone, B. L.; Ren, Z. J.; Schrier, J.; et al. Machine Learning: New Ideas and Tools in Environmental Science and Engineering. *Environmental Science & Technology* **2021**, *55* (19), 12741-12754. DOI: https://doi.org/10.1021/acs.est.1c01339.

(38) Zhu, J.-J.; Yang, M.; Ren, Z. J. Machine Learning in Environmental Research: Common Pitfalls and Best Practices. *Environmental Science & Technology* **2023**. DOI: https://doi.org/10.1021/acs.est.3c00026.

(39) Lundberg, S. M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California,

USA; 2017.

(40) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* **2020**, *2* (1), 56-67. DOI: 10.1038/s42256-019-0138-9.

(41) Lundberg, S. M.; Erion, G. G.; Lee, S.-I. Consistent individualized feature attribution for tree ensembles. *arXiv preprint arXiv:1802.03888* **2018**. DOI: https://doi.org/10.48550/arXiv.1802.03888.

(42) Ren, X.; Mi, Z.; Cai, T.; Nolte, C. G.; Georgopoulos, P. G. Flexible Bayesian Ensemble Machine Learning Framework for Predicting Local Ozone Concentrations. *Environmental Science & Technology* **2022**, *56* (7), 3871-3883. DOI: https://doi.org/10.1021/acs.est.1c04076.

(43) Hou, L.; Dai, Q.; Song, C.; Liu, B.; Guo, F.; Dai, T.; Li, L.; Liu, B.; Bi, X.; Zhang, Y.; et al. Revealing Drivers of Haze Pollution by Explainable Machine Learning. *Environmental Science & Technology Letters* **2022**, *9* (2), 112-119. DOI: https://doi.org/10.1021/acs.estlett.1c00865.

(44) Ma, Z.; Hu, X.; Sayer, A. M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y. Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004-2013. *Environmental Health Perspectives* **2016**, *124* (2), 184-192. DOI: https://doi.org/10.1289/ehp.1409481.

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

(46) Clarke, A. G.; Willison, M. J.; Zeki, E. M. A comparison of urban and rural aerosol composition using dichotomous samplers. *Atmospheric Environment (1967)* **1984**, *18* (9), 1767-1775. DOI: https://doi.org/10.1016/0004-6981(84)90352-4.

(47) Graham, A. M.; Pringle, K. J.; Arnold, S. R.; Pope, R. J.; Vieno, M.; Butt, E. W.; Conibear, L.; Stirling, E. L.; McQuaid, J. B. Impact of weather types on UK ambient particulate matter concentrations. *Atmospheric Environment: X* **2020**, *5*, 100061. DOI: https://doi.org/10.1016/j.aeaoa.2019.100061.

(48) Department for Environment Food & Rural Affairs. *CLEAN AIR STRATEGY 2019*; Department for Environment Food & Rural Affairs, 2019. https://www.gov.uk/government/publications/clean-air-strategy-2019.

(49) Aleksankina, K.; Reis, S.; Vieno, M.; Heal, M. R. Advanced methods for uncertainty assessment and global sensitivity analysis of an Eulerian atmospheric chemistry transport model. *Atmospheric Chemistry and Physics* **2019**, *19* (5), 2881-2898. DOI: https://doi.org/10.5194/acp-19-2881-2019.

(50) Chemel, C.; Fisher, B. E. A.; Kong, X.; Francis, X. V.; Sokhi, R. S.; Good, N.; Collins, W. J.; Folberth, G. A. Application of chemical transport model CMAQ to policy decisions regarding PM2.5 in the UK. *Atmospheric Environment* **2014**, *82*, 410-417. DOI: https://doi.org/10.1016/j.atmosenv.2013.10.001.

(51) Harrison, R. M.; Stedman, J.; Derwent, D. New Directions: Why are PM10 concentrations in Europe not falling? *Atmospheric Environment* **2008**, *42* (3), 603-606. DOI: https://doi.org/10.1016/j.atmosenv.2007.11.023.