Article

# Development of Europe-Wide Models for Particle Elemental Composition Using Supervised Linear Regression and Random Forest

Jie Chen,* Kees de Hoogh, John Gulliver, Barbara Hoffmann, Ole Hertel, Matthias Ketzel, Gudrun Weinmayr, Mariska Bauwelinck, Aaron van Donkelaar, Ulla A. Hvidtfeldt, Richard Atkinson, Nicole A. H. Janssen, Randall V. Martin, Evangelia Samoli, Zorana J. Andersen, Bente M. Oftedal, Massimo Stafoggia, Tom Bellander, Maciej Strak, Kathrin Wolf, Danielle Vienneau, Bert Brunekreef, and Gerard Hoek

Cite This: *Environ. Sci. Technol.* 2020, 54, 15698−15709

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

**ABSTRACT:** We developed Europe-wide models of long-term exposure to eight elements (copper, iron, potassium, nickel, sulfur, silicon, vanadium, and zinc) in particulate matter with diameter <2.5 $\mu$m ($PM_{2.5}$) using standardized measurements for one-year periods between October 2008 and April 2011 in 19 study areas across Europe, with supervised linear regression (SLR) and random forest (RF) algorithms. Potential predictor variables were obtained from satellites, chemical transport models, land-use, traffic, and industrial point source databases to represent different sources. Overall model performance across Europe was moderate to good for all elements with hold-out-validation $R$-squared ranging from 0.41 to 0.90. RF consistently outperformed SLR. Models explained within-area variation much less than the overall variation, with similar performance for RF and SLR. Maps proved a useful additional model evaluation tool. Models differed substantially between elements regarding major predictor variables, broadly reflecting known sources. Agreement between the two algorithm predictions was generally high at the overall European level and varied substantially at the national level. Applying the two models in epidemiological studies could lead to different associations with health. If both between- and within-area exposure variability are exploited, RF may be preferred. If only within-area variability is used, both methods should be interpreted equally.
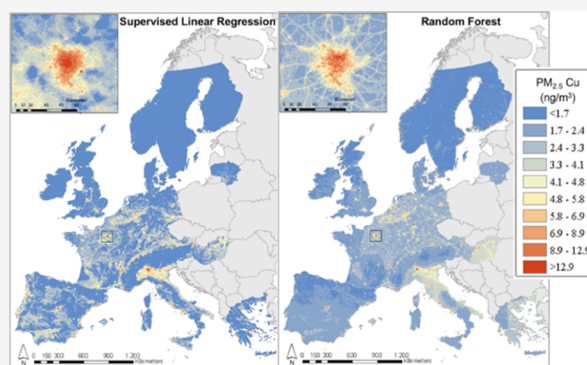
## 1. INTRODUCTION

Exposure to particulate matter (PM) is associated with adverse health outcomes.[1,2] PM is a complex mixture of components that differ spatially and temporally. Identifying which components are main contributors to adverse health effects is important for targeted policymaking. Multiple studies have attempted to associate health effects with PM components including metals, organic compounds, inorganic carbonaceous material, and inorganic secondary aerosols.[3−5] Findings, however, are inconsistent. Epidemiological studies have been limited because of the scarcity of air quality monitors that routinely measure PM composition. In Europe, a PM monitoring campaign was conducted in 20 ESCAPE (European Study of Cohorts for Air Pollution Effects) study areas following a common sampling protocol.[6] Most study areas consisted of a metropolitan area with some small towns around the main city. The PM samples were analyzed for elemental composition.[7] Based on the measurements from 20 sites in each study area, area-specific land use regression (LUR)

models were developed to assess long-term exposure to elemental composition.[8] The models were applied to cohorts within the study areas to assess health effects related to particle composition.[9]

The geographical extent of the study-area-specific ESCAPE models is limited and predictions from these models cannot reliably be used for other cohorts, such as large multicenter studies. The models were furthermore developed on 20 sites per area. Methodological studies suggested that more stable models can be developed based on larger number of sites in the model training dataset.[10,11] In addition, the ESCAPE

study-area specific models had rather good performance for traffic-related elements such as copper (Cu) and iron (Fe), but had poor performance for elements such as sulfur (S), nickel (Ni), and vanadium (V) for which spatial variation was limited within areas and key predictors were missing.[8] The lack of large-scale European models for particle elemental composition hampers large-area epidemiological studies. Our previous studies showed the possibility to develop European LUR models with good performance using a combined dataset from the ESCAPE study areas for PM with diameter <2.5 $\mu$m (PM$_{2.5}$), black carbon (BC) and nitrogen dioxide (NO$_2$).[12,13]

The supervised linear regression (SLR) algorithm is often used in air pollution modeling,[14,15] and was used to develop ESCAPE models of elemental composition[8] and European models of PM$_{2.5}$, NO$_2$, and BC.[12] The SLR algorithm shows good predictive ability and interpretability but has strong statistical assumptions such as linearity. SLR models can, however, take into account nonlinear relationships by offering a priori transformed predictor variables (e.g., inverse distance to a source), include only predictor variables following plausible direction of effect (e.g., a positive traffic slope) and add interaction terms. A number of more flexible algorithms including machine-learning algorithms have increasingly been applied in air pollution exposure assessment.[16,17] Random forest (RF) has been widely used in recent years.[18,19] RF is a classification tree analysis. It can model potentially complex relationships including nonlinearity and interactions within data but gives little information regarding the prediction process.[20] A key feature of RF is the "bagging" procedure adopted in both observation and variable selection: this allows even marginally important predictors to contribute, even in the presence of high multicollinearity. A previous study found RF outperformed linear regression in modeling spatial variation of particle elemental composition.[21]

LUR models for multiple particle components are more useful for epidemiological studies if they include more specific predictors. Developments in satellite and chemical transport modeling and availability of industrial point source data have made it possible to develop more specific models.

The aim of this study was to assess the performance of Europe-wide models for particle elemental composition, developed using SLR and RF algorithms. The models have been developed in the "Effects of Low-Level Air Pollution: A Study in Europe" (ELAPSE), a Europe-wide project investigating long-term health effects of low-level air pollution.

## 2. MATERIALS AND METHODS

**2.1. Air Pollution Data.** The PM$_{2.5}$ elemental composition concentration data originated from the ESCAPE monitoring campaigns conducted in 19 study areas across Europe (Figure S1). PM sampling and analysis methods have been described previously.[6,7] Briefly, measurements were made at 20 sites in each study area (40 in the large Catalunya and Netherlands/Belgium areas) for three 2-week periods in a 1-year period between October 2008 and April 2011. Monitoring sites were selected to represent pollution levels at regional background, urban background, and street locations using a common sampling protocol. PM$_{2.5}$ samples were collected on Teflon filters using Harvard Impactors and analyzed for elemental composition using energy-dispersive X-ray fluorescence. Annual average concentrations were calculated based on three 14-day average measurements spread over the seasons (warm, cold, and intermediate) with temporal adjustment from

a reference background site in each study area. Our measurement campaign was restricted temporally, as previous sampling campaigns used to develop LUR models.[22] While this design does not formally estimate absolute annual average concentrations as in regulatory monitoring, it has been shown to be useful to assess spatial contrast of long-term average concentrations because of specific design elements.[6,22,23] We performed temporal adjustment, using a continuous reference site located at a regional or urban background location (not directly influenced by local sources), where measurements were made for the full 12-month period. Three 14-day average samples were taken in different seasons at all locations, which are less sensitive to the very short-term variations caused by daily variation in weather. Five sites and the reference site were measured simultaneously representing all different site types (regional background, urban background, and street).[6,23]

Eight elements were a priori selected within ESCAPE to represent major pollution sources: Cu, Fe, and Zn representing nontailpipe traffic emissions, S representing long-range transport, Ni and V representing mixed oil burning/industry, silicon (Si) representing crustal material, and potassium (K) representing biomass burning.[7,8]

**2.2. Potential Predictor Variables.** *2.2.1. Traffic, Population, Altitude, and Land Use Variables.* We used the same road density, population, and elevation variables as in our previous exposure modeling paper for PM$_{2.5}$, NO$_2$, ozone (O$_3$), and BC across Europe.[12] In short, road data were extracted from the 1:10,000 EuroStreets digital road network (version 3.1 based on TeleAtlas MultiNet TM, year 2008), classified into "all" and "major" roads, and road density calculated in a 100 × 100 m grid. Population density in 1 × 1 km grid for 2011 were obtained from Eurostat.[24] Elevation was obtained from the SRTM Digital Elevation Database[25] version 4.1 with a resolution of 3 arc second (approximately 90 m) with vertical error of <16 m. *X* and/or *Y* coordinates were offered to represent the east−west/north−south gradient.

Land use variables were newly extracted from an updated European CORINE Land Cover surface in 100 × 100 m grid.[26] The initial 44 land cover classes were grouped to six main classes: residential, industry, ports, urban green space, total built up land, and natural land.

*2.2.2. Additional Component/Source-Specific Variables.* Special attention was taken in obtaining specific potential predictor variables representing different sources of the eight selected elements. We hypothesized that this would allow us to develop better and more specific models, such that the independent associations of the different elements with health could be studied better. For each component, only plausible variables were offered for model development. The restrictions of offering specific potential predictors are specified in Table S1.

Satellite-model (SAT) estimates of 2010 annual average sulfate (SO$_4^{2-}$), organic matter (OM), BC, and mineral dust (SOIL) in PM$_{2.5}$ were extracted from a gridded surface (0.01° × 0.01°, ~1.11 km) over Europe. These estimates are an application of simulated relative composition to the total PM$_{2.5}$ estimates produced by the methods described elsewhere,[27] and do not incorporate compositional ground-based measurements over Europe. In brief, PM$_{2.5}$ mass estimates were produced by relating a combined aerosol optical depth (AOD) retrieval involving multiple satellite products and simulation to near-surface PM$_{2.5}$ concentrations using the spatiotemporally varying geophysical relationship simulated by the GEOS-

Chem chemical transport model (CTM). Ground-based observations of total $PM_{2.5}$ were then incorporated into these initial values using geographically weighted regression, and the resulting total mass estimates partitioned into chemical composition using their relative contributions according to the GEOS-Chem CTM simulations.

CTM estimates of BC AOD, Sulphate AOD, total column $SO_2$, and sea-salt AOD were obtained from the European Centre for Medium-Range Weather Forecasts.[28] Daily estimates in 2010 were extracted from a gridded surface ($0.125° \times 0.125°$, ~13.9 km) produced by the MACC-II ENSEMBLE model,[29] and then aggregated to derive the annual average.

In addition to the unspecific industry land-use category, information on major industrial point sources was obtained including facility location, pollutant, and emission amount from the European Pollutant Release and Transfer Register.[30] The industrial facility points were intersected with a 100 m base polygon, and then the number of facility sites and emissions were summed within each $100 \times 100$ m cell. Density of general industries and industries emitting specific aerosols (metal, Cu, Ni, $PM_{10}$, $SO_x$, and Zn) were calculated accordingly. Sum of emissions were calculated for $PM_{10}$, Cu, Ni, $SO_x$, and Zn.

All predictor variables were integrated into a 100 m gridded GIS database covering Europe. For the road density, land use and industrial information, a moving window procedure was used to calculate the sum of values for selected buffers (focal statistics using sum within a circle). The influence of industrial point sources was calculated by inverse distance weighting ($1/d$). The processing of variable surfaces was done in ArcMap 10.6.

**2.3. Model Development.** The number of monitoring sites available for particle composition ranged from 400 to 414 because of failed PM composition measurements.[8] We used both SLR and RF algorithms to develop models.

The SLR approach has been described in detail before.[12] Briefly, a univariate linear regression model was applied for each potential predictor to find the predictor that explained the maximum variance in the measurements. At each subsequent step, the significant predictor variable ($P < 0.1$) that generated the highest increase in the model adjusted coefficient of determination (adjusted $R^2$) was added. Predictors only entered the model if they adhered to the plausible direction of the effect (Table S1). This process was repeated until the model adjusted $R^2$ could not be increased anymore. Predictor variables with variance inflation factor larger than 3 were removed from the model to avoid multicollinearity.

RF is an ensemble machine learning technique based on decision trees.[31] It builds independent trees in parallel, each based on a random sample drawn from the full set of measurements. At each node, a random subset of potential predictors is split. The final predictions are derived by averaging predictions from all decision trees. RF does not perform variable selection. It produces variable importance, calculated as percentage increase in mean squared errors after a random permutation of the values of a variable. We used the R package "randomForest" to develop the RF models.

One-step and two-step modeling processes were used to offer geographical coordinates ($X$ and $Y$) to the models. Following our previous exposure modeling procedure,[12] for SLR we used a 2-step approach, in which we first developed a SLR model without offering $X$ or $Y$, then added $X$ and $Y$ only if

they increased the model adjusted $R^2$. The rationale for the 2-step procedure is that we preferred spatial variation to be explained first by specific predictor variables and the residual variation to be further explained by the $X$, $Y$ coordinates added in the second step. In RF, we applied one-step modeling as our primary approach: $X$ and $Y$ were offered together with the other predictor variables. This allowed us to take advantage of the possibilities of RF algorithm to model the potential interactions between coordinates and other predictors. For comparison, we also developed one-step models for SLR and 2-step models for RF: we first developed a RF model without offering $X$ or $Y$, then developed a second RF model with $X$, $Y$ coordinates only, explaining variations in the residuals of the step1 RF model. The predictions of these two RF models were later added together. We further performed a sensitivity analysis offering a few nonlinear transformations of the $X$- and $Y$-coordinates to the SLR model, including $X^2$, $Y^2$, $\sqrt{X}$, $\sqrt{Y}$, and $XY$, to allow more flexible functions of the coordinates than the linear function. We were not able to perform kriging because of the clustered nature of the monitoring data.[12]

**2.4. Model Evaluation and Comparison.** For each model, we calculated model $r^2$ (squared Pearson correlation) and root-mean-square error (RMSE) by comparing main model predictions to the measurements.

We performed five-fold hold-out validation (HOV). The full set of measurements were randomly divided into five groups (20% each), stratified by site type (street, rural, and urban background) and region (north, west, central, and south). For each element-model combination, five additional HOV models were built, each based on 80% of the monitoring sites, with the remaining 20% for validation. HOV regression-based $r^2$ and RMSE were computed by comparing the stacked predictions at the five HOV test sets to the corresponding measurements. We also calculated mean square error-based $R^2$ (MSE-$R^2$), defined as

$$\text{MSE} - R^2 = 1 - \frac{\text{MSE}}{\left( \frac{1}{n} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right)}$$

where $\bar{y}$ is the average of the measurements. MSE-$R^2$ can be seen as a rescaling of MSE. It measures fit about the 1:1 line rather than fit about the best fit line in regression-based $r^2$. The HOV $r^2$ and RMSE are relevant for multicity studies that exploit both within and between city variability of air pollution contrasts.

To test how the European models predict within-area variability, we calculated within-area $r^2$ and RMSE by comparing the stacked HOV predictions and measurements within each individual study area. Because the monitors are spatially clustered over Europe and nearby locations might have auto-correlations in their measurements, we additionally performed leave-one-area-out cross-validation (LOAOCV).[32,33] We developed Europe-wide models by excluding all observations from one study area at a time and applied the models to the sites that were left out. Therefore, 19 additional models were developed for each pollutant-algorithm combination. Within-area $r^2$ and RMSE were computed by comparing the predictions and measurements in the area that was excluded from model development. We focus interpretation on the average of the within-area $r^2$s and RMSEs because the performance statistics of the individual study areas may be affected strongly by random error because they were based on only 20 sites in each study area.

### Table 1. Performance of PM$_{2.5}$ Composition Models over Europe[a]

| | | component | Cu | Fe | K | Ni | S | Si | V | Zn |
|---|---|---|---|---|---|---|---|---|---|---|
| | inclusion of X, Y coordinates | no. of sites | 414 | 413 | 414 | 402 | 404 | 400 | 402 | 413 |
| **Model Building** | | | | | | | | | | |
| SLR | one-step | model $r^2$ | 0.56 | 0.55 | 0.61 | 0.62 | 0.79 | 0.52 | 0.70 | 0.48 |
| | | model RMSE[b] | 3.3 | 65.5 | 64.6 | 0.9 | 146.5 | 59.7 | 1.7 | 11.8 |
| | two-step, step1 | model $r^2$ | 0.52 | 0.53 | 0.52 | 0.56 | 0.80 | 0.48 | 0.66 | 0.47 |
| | | model RMSE | 3.4 | 67.4 | 71.1 | 1.0 | 142.2 | 61.9 | 1.8 | 11.9 |
| | two-step, step2 | model $r^2$ | 0.56 | 0.53 | 0.60 | 0.60 | 0.82 | 0.50 | 0.69 | 0.48 |
| | | model RMSE | 3.3 | 67.4 | 65.0 | 0.9 | 135.4 | 61.1 | 1.7 | 11.8 |
| RF[c] | one-step | model $r^2$ | 0.95 | 0.95 | 0.97 | 0.95 | 0.98 | 0.95 | 0.97 | 0.95 |
| | | model RMSE | 1.1 | 20.8 | 16.8 | 0.3 | 40.2 | 19.9 | 0.5 | 3.5 |
| | two-step, step1 | model $r^2$ | 0.95 | 0.95 | 0.97 | 0.95 | 0.98 | 0.94 | 0.97 | 0.96 |
| | | model RMSE | 1.1 | 20.9 | 17.4 | 0.3 | 41.8 | 20.3 | 0.5 | 3.4 |
| | two-step, step2 | model $r^2$ | 0.98 | 0.98 | 0.99 | 0.98 | 0.99 | 0.98 | 0.99 | 0.99 |
| | | model RMSE | 0.6 | 12.4 | 9.5 | 0.2 | 27.0 | 12.2 | 0.3 | 1.8 |
| **HOV** | | | | | | | | | | |
| SLR | one-step | HOV $r^2$ | 0.47 | 0.48 | 0.58 | 0.57 | 0.76 | 0.50 | 0.63 | 0.41 |
| | | HOV RMSE | 3.6 | 70.5 | 66.4 | 1.0 | 156.4 | 60.8 | 1.8 | 12.5 |
| | two-step, step1 | HOV $r^2$ | 0.44 | 0.46 | 0.50 | 0.51 | 0.76 | 0.46 | 0.60 | 0.42 |
| | | HOV RMSE | 3.7 | 71.7 | 72.6 | 1.0 | 154.9 | 63.4 | 1.9 | 12.4 |
| | two-step, step2 | HOV $r^2$ | 0.48 | 0.48 | 0.59 | 0.56 | 0.79 | 0.46 | 0.63 | 0.41 |
| | | HOV RMSE | 3.6 | 70.5 | 66.1 | 1.0 | 147.0 | 62.9 | 1.8 | 12.5 |
| RF | one-step | HOV $r^2$ | 0.60 | 0.60 | 0.82 | 0.74 | 0.91 | 0.62 | 0.85 | 0.68 |
| | | HOV RMSE | 3.2 | 61.7 | 44.1 | 0.7 | 97.0 | 52.9 | 1.2 | 9.3 |
| | two-step, step1 | HOV $r^2$ | 0.59 | 0.59 | 0.79 | 0.74 | 0.90 | 0.60 | 0.84 | 0.68 |
| | | HOV RMSE | 3.2 | 62.4 | 47.4 | 0.7 | 102.1 | 54.2 | 1.2 | 9.2 |
| | two-step, step2 | HOV $r^2$ | 0.59 | 0.61 | 0.80 | 0.76 | 0.90 | 0.62 | 0.86 | 0.71 |
| | | HOV RMSE | 3.2 | 61.3 | 45.8 | 0.7 | 99.5 | 53.1 | 1.1 | 8.7 |

[a]SLR = supervised linear regression; RF = random forest; $r^2$ = squared Pearson correlation; RMSE = root-mean-square error; HOV = fivefold hold-out validation. [b]Unit of RMSE: ng/m$^3$. [c]Performance of RF on training set cannot be interpreted.

For each main model, predictor variables selected in SLR models and the 15 most important variables in RF models were compared.

Each main model was mapped at a 100 × 100 m resolution across the whole study area, allowing for visual comparison between maps. Additionally, we compared predictions from models at 41,936 random locations across Europe used previously.[12] Comparisons of model predictions were made for the entire study area and at the national scale reporting the Pearson correlation coefficient ($r$) and RMSE. Truncations were performed to deal with unrealistic predictions of the SLR approach: predictions at the high end were truncated to the maximum final two-step modeled value, calculated by fitting the model with the maximum predictor values at monitoring sites for positive slopes (or the minimum predictor values for negative slopes); the negative predictions were set to zeros.

## 3. RESULTS AND DISCUSSION

### 3.1. Distribution of PM$_{2.5}$ Component Measurements.
Boxplots of the annual mean concentration for PM$_{2.5}$ components in the full dataset and in individual study areas are shown in Figure S2. For the majority of pollutants, pollution concentrations varied substantially within and between study areas. A positive north−south gradient was observed with higher pollution levels in southern study areas. A more detailed interpretation of the measured concentrations can be found elsewhere.[7]

### 3.2. Model Performance.
Performance of models across Europe is shown in Table 1. Models for most components had moderate to good performance based upon HOV. Model performance was almost the same evaluating by regression-based $r^2$ or MSE-based $R^2$ (Table S2), consistent with the observation that the fitted regression slopes between observed versus predicted values are close to the 1−1 line (Figure S3). Models with the highest HOV $r^2$s were developed for PM$_{2.5}$ S, having large between-area concentration variability for which large-scale predictor variables from CTM were available to explain the contrast. Sulfate (represented by S) is a secondary pollutant formed by the oxidation of sulfur dioxide for which the ratio of between- and within-area variability is larger than for the other elements.[7] RF models consistently outperformed SLR models for all elements. This agrees with a previous study, which found more accurate exposure assessed for elemental components by RF than SLR, based on 24 monitoring sites.[21] The better performance of RF is different from two previous comparisons,[16,17] where similar performance of spatial models was observed for SLR and RF. One study compared Europe-wide models for PM$_{2.5}$ and NO$_2$ developed using similar predictor variables as in the current study,[16] the second study compared LUR models for ultrafine particles based upon mobile monitoring.[17] One possible explanation for the difference in findings is that there might be more complex relationships between predictors and elemental composition than with the mass of PM$_{2.5}$, NO$_2$, and UFP. RF can capture unknown nonlinear relationships and interactions not predefined in SLR, without introducing overfitting of the data. Another important difference is that in the current study, the data were clustered within Europe, whereas in the earlier study on PM$_{2.5}$ and NO$_2$, models were developed based upon routine monitoring, with a more even distribution of sites

**Table 2. Performance of PM$_{2.5}$ Composition Models to Assess within-Area Variation: Average within-Area $r^{2a}$**

| avg. WA $r^2$ | inclusion of $X$, $Y$ coordinates | evaluation method | Cu | Fe | K | Ni | S | Si | V | Zn |
|---|---|---|---|---|---|---|---|---|---|---|
| SLR | one-step | five-fold HOV | 0.34 | 0.35 | 0.09 | 0.18 | 0.14 | 0.18 | 0.21 | 0.20 |
| | | LOAOCV | 0.37 | 0.38 | 0.09 | 0.15 | 0.22 | 0.21 | 0.23 | 0.18 |
| | two-step, step1 | five-fold HOV | 0.34 | 0.34 | 0.08 | 0.17 | 0.14 | 0.20 | 0.18 | 0.21 |
| | | LOAOCV | 0.35 | 0.35 | 0.09 | 0.15 | 0.22 | 0.20 | 0.20 | 0.18 |
| | two-step, step2 | five-fold HOV | 0.35 | 0.36 | 0.07 | 0.17 | 0.14 | 0.20 | 0.19 | 0.19 |
| | | LOAOCV | 0.36 | 0.36 | 0.09 | 0.15 | 0.22 | 0.20 | 0.21 | 0.18 |
| RF | one-step | five-fold HOV | 0.31 | 0.31 | 0.05 | 0.21 | 0.21 | 0.19 | 0.27 | 0.24 |
| | | LOAOCV | 0.35 | 0.35 | 0.12 | 0.18 | 0.21 | 0.17 | 0.27 | 0.18 |
| | two-step, step1 | five-fold HOV | 0.31 | 0.30 | 0.06 | 0.21 | 0.22 | 0.17 | 0.27 | 0.24 |
| | | LOAOCV | 0.34 | 0.34 | 0.07 | 0.16 | 0.21 | 0.16 | 0.23 | 0.19 |
| | two-step, step2 | five-fold HOV | 0.29 | 0.29 | 0.07 | 0.21 | 0.23 | 0.17 | 0.29 | 0.25 |
| | | LOAOCV | 0.34 | 0.34 | 0.07 | 0.16 | 0.21 | 0.16 | 0.23 | 0.20 |

$^a$SLR = supervised linear regression; RF = random forest; $r^2$ = squared Pearson correlation; avg. WA $r^2$ is the average of 19 study area-specific $r^2$s (area-specific $r^2$s evaluated by five-fold HOV are shown in Figure S4); HOV = hold-out validation; LOAOCV = leave-one-area-out cross-validation.
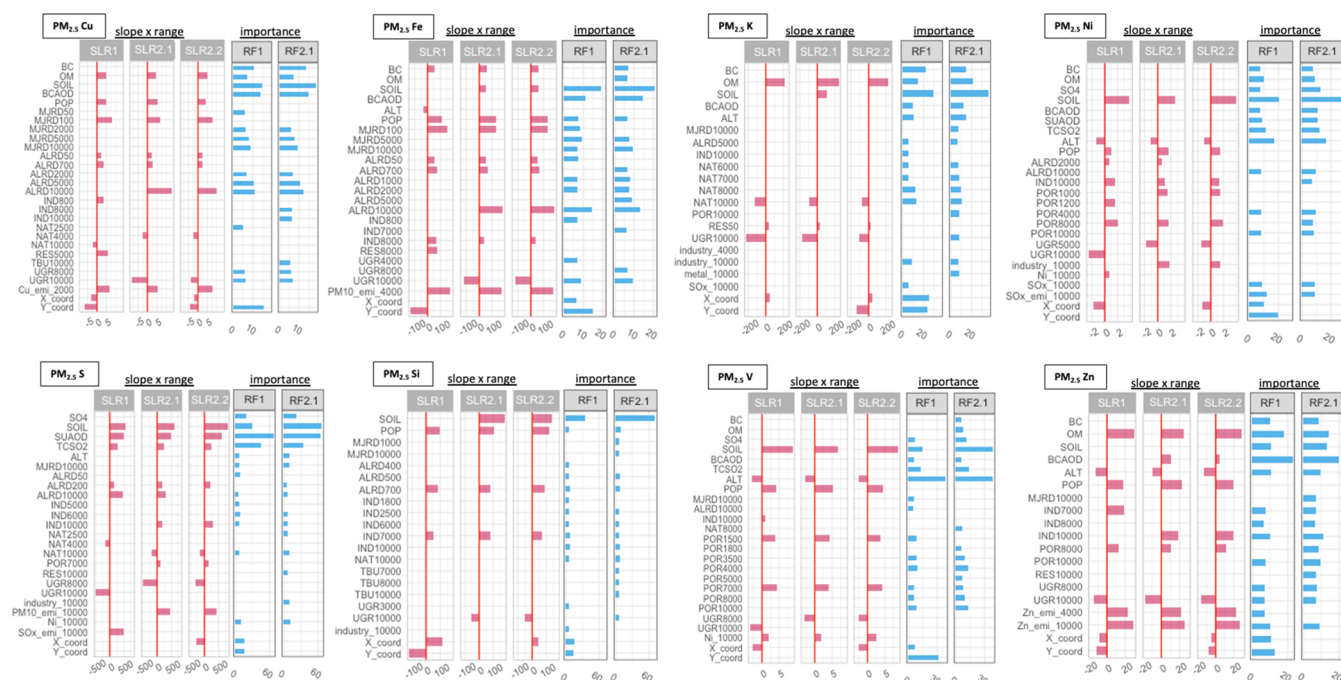


**Figure 1.** Regression slopes (shown in red) of predictors selected in SLR and relative variable importance (shown in blue) of the 15 most important predictors in RF.

across Europe. We hypothesize that the RF model accounted for spatial trends across Europe better than the linear model.

For most components, HOV $r^2$s were similar for the one-step model and the final two-step model and higher than for the first step of the two-step model, documenting that spatial trends account for the residual variance not explained by the available predictors. Offering a priori transformed $X$- and $Y$-coordinates did not further improve performance for SLR models. The differences between model $r^2$ and HOV $r^2$ in SLR models were small, suggesting the models do not overfit. The perfect performance on training set (model $r^2$) for RF models is "by design" and basically meaningless. This is because the RF algorithm generally does not prune the individual trees, relying instead on the ensemble of trees to control overfitting.[20]

While the models performed well to explain overall variability across Europe, models performed less well in explaining variation within individual study areas (Table 2). Results are similar by performing fivefold HOV and LOAOCV,

suggesting the model performance is stable regardless of CV methods. The better overall performance is explained by a combination of larger variability of concentrations between areas than within areas and the better availability of predictor variables for describing between- compared to within-area variability. Specifically, the addition of large-scale satellite and CTMs has contributed to assess the study-area background. The average within-area $r^2$s were moderate for Cu and Fe and relatively poor for other components. Cu and Fe represent mechanically generated traffic-related particles and thus their particle size distribution within PM$_{2.5}$ is skewed toward coarse particles.[7,34] Therefore, Cu and Fe do not travel far and may show large within-area variation. The better within-area performance for Cu and Fe is thus possibly because of the combination of higher within-area variation of the concentrations in most areas and the availability of data on traffic networks within individual areas. Within-area $r^2$s were poor for components that have limited within-area variation such as S. S

represents secondary inorganic aerosols (sulfates) produced by atmospheric chemistry of precursor gases (sulfur oxides) originating from combustion of sulfur-containing fossil fuels (e.g., in power plants).[35] Much of transported sulfate are in the submicron range and travel far, resulting in fairly uniform spatial variation in the scale of cities. Ni and V are often emitted from coal, oil, or residual oil burning in buildings and ships. The emission height of buildings and ships are relatively low so that within-city variation can be observed. Despite the consistently better performance of RF models than SLR models in overall HOV, the average within-area $r^2$s were similar across models for each element. This further supports our hypothesis that the RF model accounted for spatial trends across Europe better than the SLR model. Within-area $r^2$s varied substantially across study areas and were low in areas with small contrasts in measured concentrations shown by low RMSE (Figure S4).

In summary, the generally moderate within-area performance of the developed models is likely related to a combination of limited availability of predictor variables, for example, targeting especially nonexhaust traffic emissions, the clustered nature of the monitoring data and the lack of exposure contrasts within specific areas. Especially predictor variables at the local scale are insufficient.

### 3.3. Model Structure.

Predictor variables selected in SLR models and the 15 most important variables in RF models are shown in Figure 1. For each element, some consistency was found between SLR models and RF models in terms of the variable categories that were included. To some extent, however, different buffer sizes were included. Variables within each algorithm were very similar. X, Y coordinates usually contributed to the models when offered and were considered relatively important in the one-step RF.

The major predictors in the models differed substantially between the eight elements, broadly reflecting the different sources.

In Cu and Fe models, traffic-related predictor variables dominated the other source categories in SLR models while they were also considered relatively important in RF models. This is consistent with previous LUR models of Cu and Fe where a large proportion of the variability in the measured concentrations was explained by traffic-related variables.[8,36,37] Some of the industrial point sources were picked up in the SLR models for Cu and Fe, possibly reflecting emission released by metallurgic industries.[35] A previous study suggested that industrial sources were major predictors for Cu and Fe models in PM with diameter <1 $\mu$m $(PM_1)$.[38]

In Zn models, predictors representing industrial Zn emission and combustion sources contributed a large proportion to the overall $r^2$. This is consistent with LUR models in other studies.[36−38] In ESCAPE modeling, specific industrial predictors were not available.[8] The large contribution of industrial point sources to the Zn models is consistent with results of source apportionment analyses in MESA (Multi-Ethnic Study of Atherosclerosis) showing that Zn-rich features were indicative of incinerators at nearby fixed locations.[34] Previous studies have used Zn as a tracer for metallurgic industries and nonmetallurgic industries for frit production.[35]

In Ni and V models, ports were important predictors, as a proxy for shipping emissions. Density of Ni-emitting industries and more general industrial density predictors were included in the SLR model for Ni and V, consistent with the identification that Ni and V shared the same mixed industrial/fuel-oil

combustion source.[39] Large-scale SAT dust showed a large contribution in the Ni and V models, which possibly accounts for the observed north−south trend in the absence of a specific large-scale Ni and V CTM or satellite predictors. We offered SAT dust to all elements as windblown dust can be a source for all components.

In S models, variation in the measured concentrations was predominantly explained by large-scale satellite and CTM estimates and predictors in large buffers. Sulfate from the CTM and SAT dust were virtually equally important in the models. SAT sulfate did not enter the model possibly because sulfate from the CTM was in the model and they are highly correlated. SAT dust likely accounts for the observed north−south trend in concentration. In area-specific ESCAPE models, less well performing models were developed for S mainly because of the small within-study area variability.[8] Predictors representing industrial point sources also contributed to the S models, indicative of the transformation of emissions from combustion.[34]

In K models, SAT estimates for OM explained a large proportion of the variation, indicative of the main source of biomass burning for fine particle K.[35] Small-scale variables contributed little to K models, resulting in limited ability in explaining within-area variability. In our current models, we are still missing fine spatial scale biomass burning source terms because of the lack of reliable predictor variables.

Si models were dominated by SAT dust estimates and the population density, reflecting its crustal dust source.[35] Road length and industry areas from CORINE land cover also contributed to the models. These variables contributed a large fraction also in models for Si in $PM_1$.[38] In a previous study in New York, Si was strongly associated with an indicator for areas of industrial structures. This indicator includes a wide range of industrial, manufacturing, and commercial activities, thus it is difficult to identify the main source.[36]

Values between two algorithms are not quantitatively comparable. Regression slopes in SLR were multiplied by the range of each predictor to allow comparison across predictors. Relative variable importance in RF was calculated as percentage increase in mean squared errors after a random permutation of the values of a variable. $SO_4$ = satellite sulfate, OM = satellite organic matter, SOIL = satellite dust; BC = satellite black carbon; BCAOD = CTM black carbon, SUAOD = CTM sulphate, $TCSO_2$ = CTM $SO_2$, POP = population, ALT = altitude, MJRD = major roads, ALRD = all roads, TBU = total build up, NAT = natural land, IND = industry, POR = ports, UGR = urban green, RES = residential, Cu_emi = Cu emission amount, $PM_{10}$_emi = $PM_{10}$ emission amount, $SO_x$_emi = $SO_x$ emission amount, Zn_emi = Zn emission amount, industry = number of total industrial sites, Ni = number of industrial sites emitting Ni, X_coord = east−west gradient, and Y_coord = north−south gradient. Number in subscript depicts the buffer size SLR1 = one-step SLR; SLR2.1 = two-step SLR, step one; SLR2.2 = two-step SLR, step two; RF1 = one-step RF; RF2.1 = two-step RF, step one.

### 3.4. Maps and Prediction at Random Locations.

The truncation frequency for prediction at random locations is shown in Table S3. A large number of negative SLR predictions were truncated to zero for some elements—for example, 41.3% of the 41,936 random locations across Europe for Cu in the final two-step SLR model predictions. Most of the negative values were located in the low population density areas of Northern Europe, covered mostly by natural land.
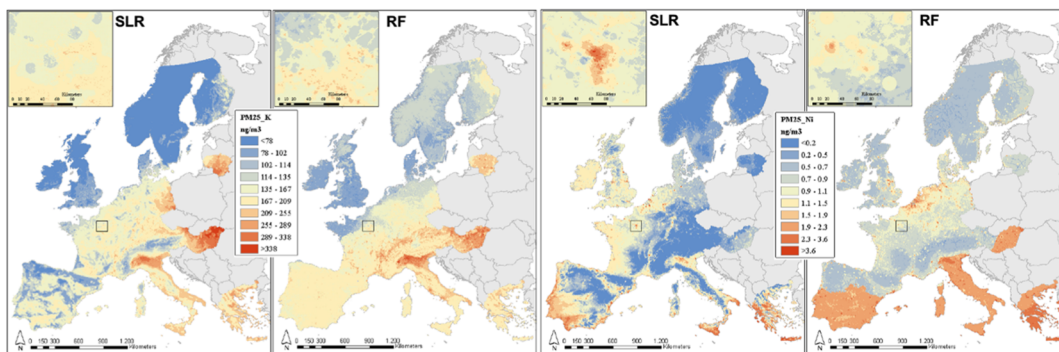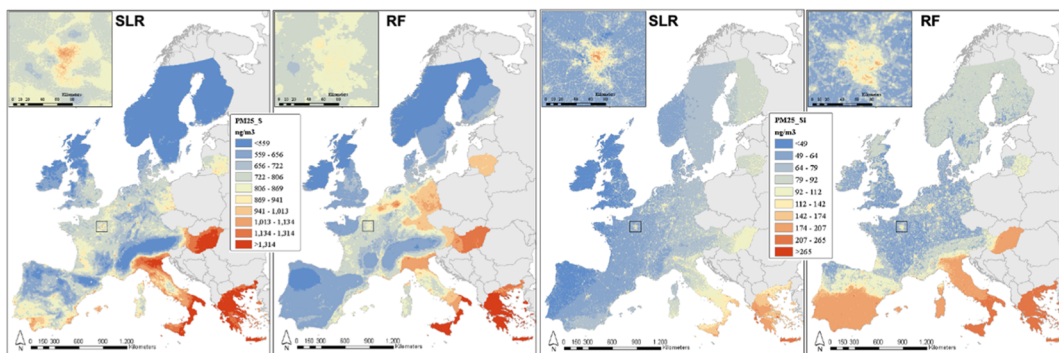
(1) PM$_{2.5}$ Cu

(2) PM$_{2.5}$ Fe
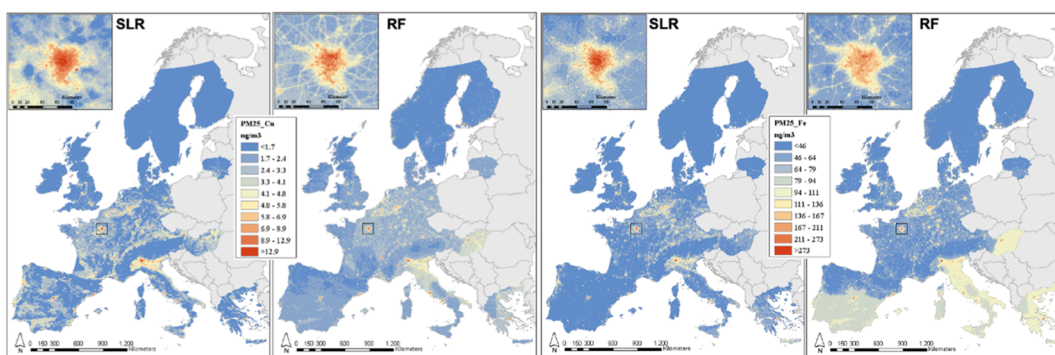
(3) PM$_{2.5}$ K

(4) PM$_{2.5}$ Ni

(5) PM$_{2.5}$ S

(6) PM$_{2.5}$ Si

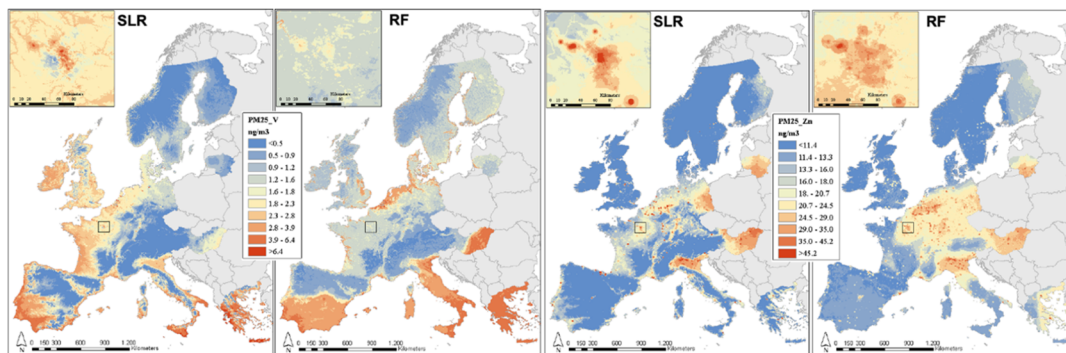(7) PM$_{2.5}$ V

(8) PM$_{2.5}$ Zn



**Figure 2.** Maps of PM$_{2.5}$ components developed by our main SLR (two-step, step2) and RF (two-step, step1) models.

**Table 3. Correlations between Predictions by Our Main SLR (Two-Step, step2) and RF (Two-Step, step1) Models at 41,936 Random Locations[a]**

| region | PM$_{2.5}$ Cu r | RMSE[b] | PM$_{2.5}$ Fe r | RMSE | PM$_{2.5}$ K r | RMSE | PM$_{2.5}$ Ni r | RMSE | PM$_{2.5}$ S r | RMSE | PM$_{2.5}$ Si r | RMSE | PM$_{2.5}$ V r | RMSE | PM$_{2.5}$ Zn r | RMSE | N |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| all European countries | 0.72 | 1.1 | 0.66 | 20.5 | 0.75 | 51.3 | 0.56 | 0.5 | 0.88 | 162.6 | 0.56 | 22.8 | 0.64 | 1.0 | 0.73 | 7.1 | 41,936 |
| ELAPSE Countries | | | | | | | | | | | | | | | | | |
| combined | 0.79 | 1.0 | 0.77 | 19.1 | 0.70 | 56.1 | 0.53 | 0.5 | 0.89 | 142.0 | 0.75 | 12.5 | 0.66 | 0.8 | 0.73 | 7.0 | 27,411 |
| Austria | 0.77 | 1.1 | 0.80 | 16.2 | 0.80 | 35.0 | 0.28 | 0.1 | 0.95 | 69.5 | 0.08 | 9.5 | 0.82 | 0.2 | 0.89 | 4.4 | 1051 |
| Belgium | 0.82 | 0.9 | 0.89 | 11.9 | 0.05 | 21.0 | 0.83 | 0.4 | 0.79 | 47.4 | 0.71 | 8.2 | 0.76 | 0.8 | 0.70 | 11.4 | 355 |
| Switzerland | 0.80 | 1.1 | 0.89 | 13.1 | 0.76 | 24.1 | 0.36 | 0.1 | 0.48 | 95.0 | 0.38 | 9.2 | 0.49 | 0.2 | 0.65 | 5.8 | 500 |
| Germany | 0.74 | 1.0 | 0.80 | 14.9 | 0.43 | 30.6 | 0.66 | 0.3 | 0.67 | 60.1 | 0.59 | 8.2 | 0.80 | 0.4 | 0.58 | 6.9 | 4233 |
| Denmark | 0.77 | 0.3 | 0.77 | 9.3 | 0.68 | 15.0 | 0.21 | 0.2 | 0.73 | 36.0 | 0.45 | 7.2 | 0.24 | 0.3 | 0.49 | 2.0 | 522 |
| France | 0.62 | 1.1 | 0.75 | 14.2 | 0.31 | 26.9 | 0.39 | 0.4 | 0.72 | 69.0 | 0.48 | 9.0 | 0.77 | 0.6 | 0.59 | 6.8 | 6476 |
| Italy | 0.72 | 1.3 | 0.53 | 23.2 | 0.69 | 33.6 | 0.51 | 0.6 | 0.83 | 183.0 | 0.70 | 19.5 | 0.63 | 1.3 | 0.67 | 9.0 | 3550 |
| Netherlands | 0.74 | 0.9 | 0.83 | 14.2 | 0.55 | 17.2 | 0.64 | 0.4 | 0.60 | 45.0 | 0.68 | 8.9 | 0.48 | 0.7 | 0.66 | 14.2 | 451 |
| Norway | 0.37 | 0.0 | 0.61 | 5.6 | -0.68[c] | 8.6 | 0.43 | 0.2 | 0.25 | 104.9 | 0.03 | 4.3 | 0.80 | 0.3 | 0.46 | 4.0 | 2649 |
| Sweden | 0.56 | 0.2 | 0.77 | 5.8 | -0.68[c] | 22.6 | 0.68 | 0.1 | 0.88 | 86.7 | 0.31 | 4.9 | 0.73 | 0.4 | 0.38 | 3.6 | 4786 |
| United Kingdom | 0.85 | 0.6 | 0.87 | 12.0 | -0.74[c] | 25.9 | 0.59 | 0.3 | 0.86 | 72.1 | 0.12 | 9.6 | 0.57 | 0.5 | 0.68 | 4.1 | 2838 |
| Non-ELAPSE Countries | | | | | | | | | | | | | | | | | |
| Greece | 0.55 | 1.0 | 0.57 | 16.8 | 0.38 | 28.7 | 0.44 | 0.8 | 0.18 | 123.5 | 0.34 | 25.6 | 0.46 | 1.9 | 0.54 | 7.0 | 1541 |
| Finland | 0.25 | 0.3 | 0.58 | 7.4 | 0.49 | 28.7 | 0.48 | 0.1 | 0.85 | 67.3 | 0.28 | 4.9 | 0.33 | 0.3 | 0.55 | 4.2 | 3208 |
| Hungary | 0.68 | 0.9 | 0.61 | 11.7 | 0.74 | 29.0 | 0.34 | 0.2 | 0.60 | 87.4 | 0.57 | 8.2 | 0.67 | 0.3 | 0.79 | 5.0 | 1123 |
| Ireland | 0.52 | 0.3 | 0.65 | 5.8 | -0.70[c] | 16.9 | 0.46 | 0.2 | 0.45 | 48.5 | -0.08 | 4.8 | 0.41 | 0.4 | 0.26 | 1.8 | 844 |
| Lithuania | 0.63 | 0.8 | 0.68 | 7.0 | 0.65 | 25.3 | 0.26 | 0.1 | 0.16 | 52.5 | 0.59 | 4.7 | 0.29 | 0.2 | 0.89 | 1.8 | 783 |
| Luxembourg | 0.76 | 0.8 | 0.83 | 8.4 | -0.08 | 16.2 | 0.82 | 0.1 | 0.82 | 26.5 | 0.77 | 4.9 | 0.80 | 0.2 | 0.83 | 2.1 | 33 |
| Portugal | 0.41 | 1.6 | 0.01 | 21.0 | -0.03 | 23.5 | 0.71 | 0.4 | 0.72 | 72.9 | 0.22 | 7.3 | 0.80 | 0.7 | 0.60 | 6.6 | 1021 |
| Spain | 0.59 | 1.2 | 0.34 | 17.2 | 0.06 | 32.7 | 0.41 | 0.6 | 0.59 | 102.0 | 0.49 | 10.7 | 0.59 | 1.1 | 0.56 | 8.0 | 5972 |

[a] r = Pearson correlation coefficient; RMSE = root-mean-square error. [b] Unit of RMSE: ng/m³. [c] We do not have clear explanations of these high negative correlations. These values possibly reflect the poor performance of both models at low concentrations. Scatter plots documented the poor agreement between predictions by the two models with lots of scatters.

When we applied the final two-step SLR models to a large Europe-wide pooled dataset of ESCAPE cohorts with 393,064 subjects (including a Swedish and Danish cohort), truncation frequencies were much smaller: 10.5% for $PM_{2.5}$ Cu, 0.5% for $PM_{2.5}$ Fe, 11.3% for $PM_{2.5}$ Ni, 14.2% for $PM_{2.5}$ V, and 2.7% for $PM_{2.5}$ Zn. Therefore, we do not expect this to be a big issue when applying the SLR models to participants in epidemiological studies. No truncation was needed for RF models.

Although we a priori considered one-step RF models as our main RF models, we observed large concentration jumps along horizontal or vertical lines in several maps (Figure S5). This counterintuitive pattern possibly reflects the role of the X, Y coordinates in RF modeling and relative importance attributed to these variables. Using X and Y in RF introduces strong boundary effects because, depending on the value where trees are split, large difference in predictions will be produced below and above that value. The concentration jumps were also observed in the final two-step RF model maps with X, Y coordinates. The RF models without offering X, Y coordinates produced clearly different maps while the HOV $r^2$s were marginally lower than for the RF models with coordinates (Table 1). We, therefore, prefer the first step in the two-step RF and the final two-step SLR (maps in Figure 2), and show maps deriving from the other procedures in the appendix (Figure S5). The maps showing strong boundary effects might require smoothing before application in epidemiological studies. Our results clearly indicate the value in evaluating plausibility of maps as an important last step in air pollution exposure assessment studies. Comparing models solely by HOV statistics is not sufficient. We did not observe sharp gradients in the SLR model maps.

There are clear agreements between maps produced by our main SLR and RF for some elements and differences for other elements (Figure 2). In both maps for Cu, high levels of pollution are shown in big cities, and transport networks can be clearly seen in the inset map of the area around Paris. Maps for $PM_{2.5}$ S are broadly similar with higher pollution levels in the south and east, while quite different patterns were observed for East Germany and Spain. Both maps for Zn show high concentrations close to industrial sites. The same industrial sites were picked up in the area around Paris shown in the inset. Comparing the predictions at a total of 41,936 random locations, agreement was high at all European countries-level and the ELAPSE countries combined-level for $PM_{2.5}$ Cu, $PM_{2.5}$ K, $PM_{2.5}$ S, $PM_{2.5}$ Zn ($r > 0.7$), and moderately high for other elements (Table 3). Correlation between all model predictions at the ELAPSE countries combined-level is presented in Figure S6. For most components, correlations were high for predictions derived from the same algorithm, and lower for predictions derived from different algorithms.

While correlations between predictions derived from SLR and RF were moderate to high at the European level, they are lower than the very high correlations ($r$ generally >0.9) reported previously for Europe-wide models of $PM_{2.5}$ and $NO_2$.[16] Agreement between predictions from the two algorithms at the national level varied substantially across countries (Table 3). There was no consistently good agreement between predictions for a specific country. Poor agreement between predictions were observed for area-component combinations that had small contrasts in measured concentrations shown by low RMSE (e.g., most components in Norway and Sweden).

Computation time for mapping differed substantially for RF and SLR—around 40 h for RF and less than 1 h for SLR to map pollution concentrations across Europe on a standard office computer.

**3.5. Strengths and Limitations.** With the development of Europe-wide models, we are able to assess long-term exposures to $PM_{2.5}$ components in a large European project, which consists of several nation-wide cohorts and smaller cohorts in which participants were recruited in specific study areas. The use of a single harmonized model allows a standardized exposure assessment in international multicenter studies.

Our Europe-wide models had the advantage of a large training dataset with large contrasts in measured concentrations by combining measurements from individual ESCAPE study areas. In contrast, the previous ESCAPE area-specific models could not be developed for some composition-area combinations because of missing data (e.g., in Lugano), small within-area variability (e.g., S) and poor precision of the measurements in areas with low concentrations (Ni and V).[8] The moderate to good performance of our models across Europe suggests that the models would perform well in multicenter studies that exploit both within and between area variability of air pollution contrasts.

Another strength of our study is that we made efforts in collecting specific large-scale predictors, from satellites and CTMs, representing different pollution sources such as soil, industrial sources, and biomass burning, which could not be applied in prior area-specific models. The availability of these predictors increased the specificity of our models, which is useful to study associated health effects of specific single components.

While inclusion of industrial point source data was an improvement over the simple land use categories available in CORINE land cover, a dispersion model for point sources would have been the method of choice. We did not have the possibility to use Europe-wide small-scale dispersion modeling, and we did not have information on chimney height and wind direction around chimneys. We therefore used inverse distance weighting to create variables from industrial point sources, which can lead to overestimation of pollution levels in areas very close to the industrial sites.[40] The misclassification is expected to be minimal given that it is unlikely that many people live very close to the large point source chimneys included in the European Pollutant Release and Transfer Register databases. The small truncation frequency (<0.1%) above the maximum values in a total of around 42,000 random locations across Europe suggested the overestimation might not have a large impact.

The prediction ability of our Europe-wide model at small-scale, however, is limited, especially in areas without main sources present. The lack of specificity of the small-scale land use predictors might have contributed to the poor predictive ability for some elements. The poor within-area predictive ability suggests our Europe-wide models should be applied with caution in small-scale individual study areas, with the possible exception of the Cu and Fe models. Our model is more suited for multicenter studies.

Moderate to high overall correlations between our Europe-wide SLR and previous area-specific ESCAPE model predictions at monitoring sites were observed except for K and Zn (Table S4). The within-area correlations between SLR and ESCAPE varied considerably and the average correlations were high for Cu and Fe. The results suggested applying the

newly developed Europe-wide models in epidemiological studies could lead to different findings from the ESCAPE study. The leave-one-out-cross-validation (LOOCV) $r^2$s of the area-specific ESCAPE models are not quantitatively comparable with the within-area fivefold cross-validation $r^2$s in this study as the LOOCV is based on a small number of sites and tends to overestimate predictive ability.[10,11]

Given the discrepancies in predictions derived from the two methods, applying the two sets of models in epidemiological studies could lead to different associations with health. SLR and RF model performances were similar for the within-area concentration variability, while RF model explained overall concentration variability (including between-area variability) better than SLR. In SLR, we did not add fixed or random intercepts for study area as such models could not be applied outside the specific study areas. In a previous study on PM$_{2.5}$ and NO$_2$,[33] we found that adding indicators for study area or the measured regional background in each study area, improved the overall explained variability. Therefore, when applied in epidemiological studies, it depends on the contrast exploited in the epidemiological study which method is the preferred method. If both between- and within-area variability are exploited, RF would be the method of choice based on the cross-validation statistics. If an epidemiological study only includes within-area exposure contrast, then both methods should be interpreted equally, without a prior preference for one of the methods. Given the moderate performance of both models, it would be important to observe robustness of the findings in epidemiological studies. If health effects are found with only one model, this should be interpreted cautiously.

Because of the lack of external validation data, we cannot draw strong conclusions about the preferred method. We note that RF models might be more difficult to interpret in terms of how predictor variables act in the models, although the "importance" statistics provide useful information on the relative importance of individual predictors. The classification nature of RF led to visible boundary effects in some exposure maps, which might require smoothing before application in epidemiological studies. On the other hand, SLR might fail to capture some complex nonlinear relationships and/or interactions between predictors and pollutants, or might induce overfitting if multiple nonlinear and interaction terms were added to the model. Despite the discrepancies in predictions, we believe our models are stable and the results are robust, as different cross-validation methods and several sensitivity analyses showed moderate to good performance, especially at the overall Europe-wide scale and similar results.

## ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.0c06595.

Overview of potential predictor variables; performance of PM2.5 composition models over Europe; truncation frequency for model predictions at random locations; correlation of predictions at monitoring sites; distribution of 416 ESCAPE monitoring sites; boxplots of annual mean concentrations for PM2.5 composition; scatter plots of the stacked predictions at five held-out sites versus measurements; within-area $r^2$s and RMSEs of PM2.5 composition models; maps of PM2.5 components; and Pearson correlation between model predictions at random locations across ELAPSE countries (PDF)

## AUTHOR INFORMATION

### Corresponding Author

**Jie Chen** − *Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TC Utrecht, The Netherlands;* ⓘ orcid.org/0000-0001-9644-8026; Email: j.chen1@uu.nl

### Authors

**Kees de Hoogh** − *Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland; University of Basel, 4001 Basel, Switzerland;* ⓘ orcid.org/0000-0001-5974-2007

**John Gulliver** − *Centre for Environmental Health and Sustainability, School of Geography, Geology and the Environment, University of Leicester, LE1 7RH Leicester, U.K.;* ⓘ orcid.org/0000-0003-3423-2013

**Barbara Hoffmann** − *Institute for Occupational, Social and Environmental Medicine, Centre for Health and Society, Medical Faculty, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany*

**Ole Hertel** − *Department of Environmental Science, Aarhus University, 4000 Roskilde, Denmark*

**Matthias Ketzel** − *Department of Environmental Science, Aarhus University, 4000 Roskilde, Denmark; Global Centre for Clean Air Research (GCARE), Department of Civil and Environmental Engineering, University of Surrey, GU2 7XH Guildford, U.K.*

**Gudrun Weinmayr** − *Institute of Epidemiology and Medical Biometry, Ulm University, 89081 Ulm, Germany*

**Mariska Bauwelinck** − *Interface Demography–Department of Sociology, Vrije Universiteit Brussel, 1050 Brussels, Belgium*

**Aaron van Donkelaar** − *Department of Physics and Atmospheric Science, Dalhousie University, B3H 4R2 Halifax, Nova Scotia, Canada; Department of Energy, Environmental & Chemical Engineering, Washington University in St. Louis, 63130 St. Louis, Missouri, United States*

**Ulla A. Hvidtfeldt** − *Danish Cancer Society Research Center, 2100 Copenhagen, Denmark*

**Richard Atkinson** − *St George's University of London, SW17 0RE London, U.K.*

**Nicole A. H. Janssen** − *National Institute for Public Health and the Environment (RIVM), 3720 BA Bilthoven, The Netherlands*

**Randall V. Martin** − *Danish Cancer Society Research Center, 2100 Copenhagen, Denmark; Department of Physics and Atmospheric Science, Dalhousie University, B3H 4R2 Halifax, Nova Scotia, Canada; Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, 02138 Cambridge, Massachusetts, United States*

**Evangelia Samoli** − *Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, 115 27 Athens, Greece*

**Zorana J. Andersen** − *University of Copenhagen, 1165 Copenhagen, Denmark*

**Bente M. Oftedal** − *Department of Environmental Health, Norwegian Institute of Public Health, N-0403 Oslo, Norway*

**Massimo Stafoggia** − *Department of Epidemiology, Lazio Region Health Service/ASL Roma 1, 00147 Rome, Italy; Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden*

**Tom Bellander** − *Institute of Environmental Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden*

**Maciej Strak** − *Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TC Utrecht, The Netherlands; Atomic and Molecular Physics Division, Harvard-Smithsonian Center for Astrophysics, Cambridge, 02138 Cambridge, Massachusetts, United States*

**Kathrin Wolf** − *Helmholtz Zentrum München, German Research Center for Environmental Health (GmbH), Institute of Epidemiology, D-85764 Neuherberg, Germany;* orcid.org/0000-0002-4343-201X

**Danielle Vienneau** − *Swiss Tropical and Public Health Institute, 4051 Basel, Switzerland; University of Basel, 4001 Basel, Switzerland*

**Bert Brunekreef** − *Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TC Utrecht, The Netherlands; Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3584 CX Utrecht, Netherlands*

**Gerard Hoek** − *Institute for Risk Assessment Sciences (IRAS), Utrecht University, 3508 TC Utrecht, The Netherlands*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.est.0c06595

**Notes**

The authors declare no competing financial interest.

■ **REFERENCES**

(1) Adar, S. D.; Filigrana, P. A.; Clements, N.; Peel, J. L. Ambient coarse particulate matter and human health: a systematic review and meta-analysis. *Curr. Environ. Health Rep.* **2014**, *1*, 258−274.

(2) Vodonos, A.; Awad, Y. A.; Schwartz, J. The concentration-response between long-term PM2.5 exposure and mortality; A meta-regression approach. *Environ. Res.* **2018**, *166*, 677−689.

(3) Adams, K.; Greenbaum, D. S.; Shaikh, R.; van Erp, A. M.; Russell, A. G. Particulate matter components, sources, and health: Systematic approaches to testing effects. *J. Air Waste Manage. Assoc.* **2015**, *65*, 544−558.

(4) Badaloni, C.; Cesaroni, G.; Cerza, F.; Davoli, M.; Brunekreef, B.; Forastiere, F. Effects of long-term exposure to particulate matter and metal components on mortality in the Rome longitudinal study. *Environ. Int.* **2017**, *109*, 146−154.

(5) Pennington, A. F.; Strickland, M. J.; Gass, K.; Klein, M.; Sarnat, S. E.; Tolbert, P. E.; Balachandran, S.; Chang, H. H.; Russell, A. G.; Mulholland, J. A.; Darrow, L. A. Source-Apportioned PM2.5 and Cardiorespiratory Emergency Department Visits: Accounting for Source Contribution Uncertainty. *Epidemiology* **2019**, *30*, 789−798.

(6) Eeftens, M.; Tsai, M.-Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dèdelè, A.; Eriksen, K.; Galassi, C.; Gražulevičienè, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki, T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening, M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Spatial variation of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20 European study areas and the relationship with NO2—Results of the ESCAPE project. *Atmos. Environ.* **2012**, *62*, 303−317.

(7) Tsai, M.-Y.; Hoek, G.; Eeftens, M.; de Hoogh, K.; Beelen, R.; Beregszászi, T.; Cesaroni, G.; Cirach, M.; Cyrys, J.; De Nazelle, A.; de Vocht, F.; Ducret-Stich, R.; Eriksen, K.; Galassi, C.; Gražuleviciene, R.; Gražulevicius, T.; Grivas, G.; Gryparis, A.; Heinrich, J.; Hoffmann, B.; Iakovides, M.; Keuken, M.; Krämer, U.; Künzli, N.; Lanki, T.; Madsen, C.; Meliefste, K.; Merritt, A.-S.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M. J.; Pershagen, G.; Phuleria, H.; Quass, U.; Ranzi, A.; Schaffner, E.; Sokhi, R.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Taimisto, P.; Tewis, M.; Udvardy, O.; Wang, M.; Brunekreef, B. Spatial variation of PM elemental composition between and within 20 European study areas–Results of the ESCAPE project. *Environ. Int.* **2015**, *84*, 181−192.

(8) de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dèdelè, A.; Dons, E.; de Nazelle, A.; Eeftens, M.; Eriksen, K.; Eriksson, C.; Fischer, P.; Gražulevičienė, R.; Gryparis, A.; Hoffmann, B.; Jerrett, M.; Katsouyanni, K.; Iakovides, M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Mosler, G.; Nádor, G.; Nieuwenhuijsen, M.; Pershagen, G.; Peters, A.; Phuleria, H.; Probst-Hensch, N.; Raaschou-Nielsen, O.; Quass, U.; Ranzi, A.; Stephanou, E.; Sugiri, D.; Schwarze, P.; Tsai, M.-Y.; Yli-Tuomi, T.; Varró, M. J.; Vienneau, D.; Weinmayr, G.; Brunekreef, B.; Hoek, G. Development of land use regression models for particle composition in twenty study areas in Europe. *Environ. Sci. Technol.* **2013**, *47*, 5778−5786.

(9) Beelen, R.; Hoek, G.; Raaschou-Nielsen, O.; Stafoggia, M.; Andersen, Z. J.; Weinmayr, G.; Hoffmann, B.; Wolf, K.; Samoli, E.; Fischer, P. H.; Nieuwenhuijsen, M. J.; Xun, W. W.; Katsouyanni, K.; Dimakopoulou, K.; Marcon, A.; Vartiainen, E.; Lanki, T.; Yli-Tuomi, T.; Oftedal, B.; Schwarze, P. E.; Nafstad, P.; De Faire, U.; Pedersen, N. L.; Östenson, C.-G.; Fratiglioni, L.; Penell, J.; Korek, M.; Pershagen, G.; Eriksen, K. T.; Overvad, K.; Sørensen, M.; Eeftens, M.; Peeters, P. H.; Meliefste, K.; Wang, M.; Bueno-de-Mesquita, H. B.; Sugiri, D.; Krämer, U.; Heinrich, J.; de Hoogh, K.; Key, T.; Peters, A.; Hampel, R.; Concin, H.; Nagel, G.; Jaensch, A.; Ineichen, A.; Tsai, M.-Y.; Schaffner, E.; Probst-Hensch, N. M.; Schindler, C.; Ragettli, M. S.; Vilier, A.; Clavel-Chapelon, F.; Declercq, C.; Ricceri, F.; Sacerdote, C.; Galassi, C.; Migliore, E.; Ranzi, A.; Cesaroni, G.; Badaloni, C.; Forastiere, F.; Katsoulis, M.; Trichopoulou, A.; Keuken, M.; Jedynska, A.; Kooter, I. M.; Kukkonen, J.; Sokhi, R. S.; Vineis, P.; Brunekreef, B. Natural-cause mortality and long-term exposure to particle components: an analysis of 19 European cohorts within the multi-center ESCAPE project. *Environ. Health Perspect.* **2015**, *123*, 525−533.

(10) Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N. Effect of the number of measurement sites on land use regression models in estimating local air pollution. *Atmos. Environ.* **2012**, *54*, 634−642.

(11) Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B. Systematic evaluation of land use regression models for NO2. *Environ. Sci. Technol.* **2012**, *46*, 4481−4489.

(12) De Hoogh, K.; Chen, J.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Klompmaker, J.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Brunekreef, B.; Hoek, G. Spatial PM2.5, NO2, O3 and BC models for Western Europe—Evaluation of spatiotemporal stability. *Environ. Int.* **2018**, *120*, 81−92.

(13) de Hoogh, K.; Gulliver, J.; van Donkelaar, A.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.;

Eeftens, M.; Forsberg, B.; Galassi, C.; Heinrich, J.; Hoffmann, B.; Jacquemin, B.; Katsouyanni, K.; Korek, M.; Künzli, N.; Lindley, S. J.; Lepeule, J.; Meleux, F.; de Nazelle, A.; Nieuwenhuijsen, M.; Nystad, W.; Raaschou-Nielsen, O.; Peters, A.; Peuch, V.-H.; Rouil, L.; Udvardy, O.; Slama, R.; Stempfelet, M.; Stephanou, E. G.; Tsai, M. Y.; Yli-Tuomi, T.; Weinmayr, G.; Brunekreef, B.; Vienneau, D.; Hoek, G. Development of West-European PM 2.5 and NO 2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environ. Res.* **2016**, *151*, 1−10.

(14) Brauer, M.; Hoek, G.; van Vliet, P.; Meliefste, K.; Fischer, P.; Gehring, U.; Heinrich, J.; Cyrys, J.; Bellander, T.; Lewne, M.; Brunekreef, B. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. *Epidemiology* **2003**, *14*, 228−239.

(15) Henderson, S. B.; Beckerman, B.; Jerrett, M.; Brauer, M. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environ. Sci. Technol.* **2007**, *41*, 2422−2428.

(16) Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U. A.; Katsouyanni, K.; Janssen, N. A. H.; Martin, R. V.; Samoli, E.; Schwartz, P. E.; Stafoggia, M.; Bellander, T.; Strak, M.; Wolf, K.; Vienneau, D.; Vermeulen, R.; Brunekreef, B.; Hoek, G. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environ. Int.* **2019**, *130*, 104934.

(17) Kerckhoffs, J.; Hoek, G.; Portengen, L.; Brunekreef, B.; Vermeulen, R. C. H. Performance of Prediction Algorithms for Modeling Outdoor Air Pollution Spatial Surfaces. *Environ. Sci. Technol.* **2019**, *53*, 1413−1421.

(18) Meng, X.; Hand, J. L.; Schichtel, B. A.; Liu, Y. Space-time trends of PM2.5 constituents in the conterminous United States estimated by a machine learning approach, 2005-2015. *Environ. Int.* **2018**, *121*, 1137−1147.

(19) Hu, X.; Belle, J. H.; Meng, X.; Wildani, A.; Waller, L. A.; Strickland, M. J.; Liu, Y. Estimating PM2.5 Concentrations in the Conterminous United States Using the Random Forest Approach. *Environ. Sci. Technol.* **2017**, *51*, 6936−6944.

(20) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.

(21) Brokamp, C.; Jandarov, R.; Rao, M. B.; LeMasters, G.; Ryan, P. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmos. Environ.* **2017**, *151*, 1−11.

(22) Hoek, G.; Beelen, R.; De Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561−7578.

(23) Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dėdelė, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.; Galassi, C.; Gražulevičienė, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Nádor, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M.-Y.; Yli-Tuomi, T.; Varró, M. J.; Vienneau, D.; Klot, S. v.; Wolf, K.; Brunekreef, B.; Hoek, G. Development of Land Use Regression Models for PM2.5, PM2.5 Absorbance, PM10 and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012**, *46*, 11195−11205.

(24) *GEOSTAT* (version 2.0.1). 2011, https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat (last accessed July 19, 2019).

(25) CGIAR-CSI. *Srtm 90m Digital Elevation Data*, 2013.

(26) CLC. 2012, https://land.copernicus.eu/pan-european/corine-land-cover/clc-2012 (last accessed July 19, 2019).

(27) van Donkelaar, A.; Martin, R. V.; Li, C.; Burnett, R. T. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from

Satellites, Models, and Monitors. *Environ. Sci. Technol.* **2019**, *53*, 2595−2611.

(28) *ECMWF.* https://apps.ecmwf.int/datasets/data/macc-reanalysis/levtype=sfc/ (last accessed July 23, 2019).

(29) Inness, A.; Baier, F.; Benedetti, A.; Bouarar, I.; Chabrillat, S.; Clark, H.; Clerbaux, C.; Coheur, P.; Engelen, R. J.; Errera, Q.; Flemming, J.; George, M.; Granier, C.; Hadji-Lazaro, J.; Huijnen, V.; Hurtmans, D.; Jones, L.; Kaiser, J. W.; Kapsomenakis, J.; Lefever, K.; Leitão, J.; Razinger, M.; Richter, A.; Schultz, M. G.; Simmons, A. J.; Suttie, M.; Stein, O.; Thépaut, J.-N.; Thouret, V.; Vrekoussis, M.; Zerefos, C. The MACC reanalysis: an 8 yr data set of atmospheric composition. *Atmos. Chem. Phys.* **2013**, *13*, 4073−4109.

(30) E-PRTR. https://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-18 (version 16, last accessed July 19, 2019).

(31) Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18−22.

(32) Murray, N. L.; Holmes, H. A.; Liu, Y.; Chang, H. H. A Bayesian ensemble approach to combine PM2.5 estimates from statistical models using satellite imagery and numerical model simulation. *Environ. Res.* **2019**, *178*, 108601.

(33) Wang, M.; Beelen, R.; Bellander, T.; Birk, M.; Cesaroni, G.; Cirach, M.; Cyrys, J.; de Hoogh, K.; Declercq, C.; Dimakopoulou, K.; Eeftens, M.; Eriksen, K. T.; Forastiere, F.; Galassi, C.; Grivas, G.; Heinrich, J.; Hoffmann, B.; Ineichen, A.; Korek, M.; Lanki, T.; Lindley, S.; Modig, L.; Mölter, A.; Nafstad, P.; Nieuwenhuijsen, M. J.; Nystad, W.; Olsson, D.; Raaschou-Nielsen, O.; Ragettli, M.; Ranzi, A.; Stempfelet, M.; Sugiri, D.; Tsai, M.-Y.; Udvardy, O.; Varró, M. J.; Vienneau, D.; Weinmayr, G.; Wolf, K.; Yli-Tuomi, T.; Hoek, G.; Brunekreef, B. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. *Environ. Health Perspect.* **2014**, *122*, 843−849.

(34) Vedal, S.; Campen, M. J.; McDonald, J. D.; Kaufman, J. D.; Larson, T. V.; Sampson, P. D.; Sheppard, L.; Simpson, C. D.; Szpiro, A. A. National Particle Component Toxicity (NPACT) Initiative Report on Cardiovascular Effects. Research Report 178. Health Effects Institute: Boston, MA, 2013.

(35) Belis, C. A.; Karagulian, F.; Larsen, B. R.; Hopke, P. K. Critical review and meta-analysis of ambient particulate matter source apportionment using receptor models in Europe. *Atmos. Environ.* **2013**, *69*, 94−108.

(36) Ito, K.; Johnson, S.; Kheirbek, I.; Clougherty, J.; Pezeshki, G.; Ross, Z.; Eisl, H.; Matte, T. D. Intraurban Variation of Fine Particle Elemental Concentrations in New York City. *Environ. Sci. Technol.* **2016**, *50*, 7517−7526.

(37) Tripathy, S.; Tunno, B. J.; Michanowicz, D. R.; Kinnee, E.; Shmool, J. L. C.; Gillooly, S.; Clougherty, J. E. Hybrid land use regression modeling for estimating spatio-temporal exposures to PM2.5, BC, and metal components across a metropolitan area of complex terrain and industrial sources. *Sci. Total Environ.* **2019**, *673*, 54−63.

(38) Zhang, J. J. Y.; Sun, L.; Barrett, O.; Bertazzon, S.; Underwood, F. E.; Johnson, M. Development of land-use regression models for metals associated with airborne particulate matter in a North American city. *Atmos. Environ.* **2015**, *106*, 165−177.

(39) Viana, M.; Kuhlbusch, T. A. J.; Querol, X.; Alastuey, A.; Harrison, R. M.; Hopke, P. K.; Winiwarter, W.; Vallius, M.; Szidat, S.; Prévôt, A. S. H.; Hueglin, C.; Bloemen, H.; Waåhlin, P.; Vecchi, R.; Miranda, A. I.; Kasper-Giebl, A.; Maenhaut, W.; Hitzenberger, R. Source apportionment of particulate matter in Europe: A review of methods and results. *J. Aerosol Sci.* **2008**, *39*, 827−849.

(40) Michanowicz, D. R.; Shmool, J. L. C.; Tunno, B. J.; Tripathy, S.; Gillooly, S.; Kinnee, E.; Clougherty, J. E. A hybrid land use regression/AERMOD model for predicting intra-urban variation in PM2.5. *Atmos. Environ.* **2016**, *131*, 307−315.