Atmospheric Environment 335 (2024) 120719

Contents lists available at ScienceDirect

Atmospheric Environment

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

Europe-wide high-spatial resolution air pollution models are improved by including traffic flow estimates on all roads

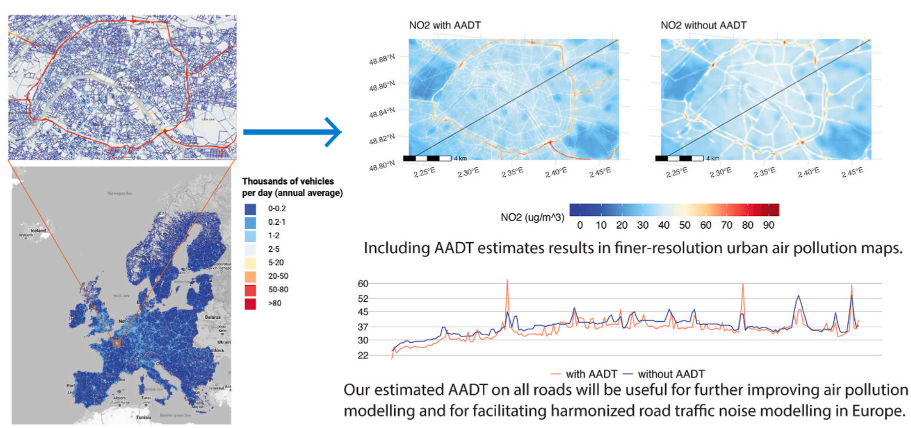
Youchen Shen [a,*], Kees de Hoogh [b,c], Oliver Schmitz [d], John Gulliver [e,f], Danielle Vienneau [b,c], Roel Vermeulen [a], Gerard Hoek [a,1], Derek Karssenberg [d,1]

[a] *Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands*
[b] *Swiss Tropical and Public Health Institute, Allschwil, Switzerland*
[c] *University of Basel, Basel, Switzerland*
[d] *Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands*
[e] *Centre for Environmental Health and Sustainability, School of Geography, Geology and the Environment, University of Leicester, Leicester, LE1 7HA, UK*
[f] *Population Health Research Institute, St George's, University of London, UK*

HIGHLIGHTS

- We used random forests models to estimate traffic counts on all roads across Europe.
- 5-fold cross-validation accuracy was good for most road types.
- Correlation between European and national traffic flow models was satisfactory.
- Our AADTs improved previous Europe-wide air pollution models.
- Incorporation of AADT resulted in finer-resolution urban air pollution maps.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Road traffic intensity
Machine learning
Geographic information system (GIS)
Land-use regression
Air pollution

ABSTRACT

Road traffic is an important source of noise and air pollution. Modelling of air pollution and noise therefore requires detailed information on annual average daily traffic (AADT) flows on all roads. Europe-wide estimates on traffic intensity are however not publicly available. This has hampered previous Europe-wide air pollution and noise modelling, used extensively in Europe-wide epidemiological studies of morbidity and mortality. We aim to estimate Europe-wide AADT and quantify potential improvements of previous Europe-wide air pollution models.

We built separate random forests (RF) models for different road types in OpenStreetMap (highway, primary, secondary and tertiary, and residential roads). We collected observations on annual average daily traffic (AADT)

* Corresponding author.
E-mail address: y.shen@uu.nl (Y. Shen).
[1] Shared last co-authors.

https://doi.org/10.1016/j.atmosenv.2024.120719
Received 18 April 2024; Received in revised form 10 July 2024; Accepted 24 July 2024
Available online 25 July 2024

from six European countries. We evaluated our AADT models using 5-fold cross-validation (CV) and by comparison of our Europe-wide traffic flow estimates with national traffic model estimates for Switzerland and the Netherlands. We evaluated whether adding our estimated AADT as predictors for Europe-wide air pollution models trained by more than 2000 routine monitoring sites improved the performance of the models based upon major road length in different buffer sizes.

The 5-fold cross-validation result showed our estimates overall captured variations in AADT between road types ($R^2 = 0.82$). Our result showed variability in AADT within and between road types, documenting the benefit of our model framework at a continental scale. Our AADT estimates modestly improved model performance of previous Europe-wide air pollution models for NO_2, PM_{10}, $PM_{2.5}$, and O_3, especially for NO_2 (3% improvement of geographically-weighted regression model). Improvement of model performance was larger in urban areas (5% and 8% increases in R^2 for NO_2 and O_3). Importantly, more detailed intra-city near-road variations were captured for traffic-related air pollution. The resulting AADT estimates of all roads across Europe will be useful for further improving air pollution modelling and facilitating harmonized road traffic noise modelling in Europe.

Software

The work was done in Ubuntu (18.04.6 LTS) and with R version 4.2.1.

1. Introduction

Road traffic intensity (AADT) is a key driver for levels of traffic-related air pollution and road traffic noise. Some epidemiological studies on the health effects of traffic-related pollution used AADT data as input for estimating air pollution and road traffic noise (Eeftens et al., 2012; Morley and Gulliver, 2016). However, most Europe-wide air pollution models do not include AADT but instead use road length of specific road types (de Hoogh et al., 2016; de Hoogh et al., 2018). For estimating air pollution exposure with land-use regression (LUR), where emission- and dispersion-related variables are used to explain the variations in measured air pollution concentrations, previous studies have shown that AADT data were important for improving the accuracy of exposure estimates (Beelen et al., 2013; Eeftens et al., 2012; Wang et al., 2014). One of the studies, with 23 mostly large urban areas in Europe, showed that including local AADT data improved the model R^2 on average by 10% for estimating nitrogen dioxide and nitrogen oxides compared to using road type information to represent road traffic (Beelen et al., 2013). For road traffic noise, more accurate AADT data has been shown to improve the noise estimates (Morley and Gulliver, 2016).

Although AADT data can be publicly available for some selected areas (e.g. local, regional, or national), or broadly commercially available, it is not publicly available nor affordable for researchers over larger areas (e.g. continentally or globally) on all roads. In previous air pollution modelling studies at larger scales, area-specific AADT data was collected in selected cities by local co-workers (Beelen et al., 2013; Eeftens et al., 2012; Wang et al., 2014). Studies instead used road length by type as a proxy for AADT (Chen et al., 2020; de Hoogh et al., 2018; Shen et al., 2022). Only a small number of studies estimated AADT as an input for assessing air quality and road traffic noise at a city- or nation-scale (Alvarado-Molina et al., 2023; Morley and Gulliver, 2016).

For epidemiological studies on health effects of noise and/or air pollution, exposure estimates need to consider road-related effects of AADT on all roads. This presents an additional challenge, as AADT is typically only observed on major roads. On minor roads, deterministic traffic models can estimate AADT to derive a complete coverage of all roads based on route choices (Ermagun and Levinson, 2019; Sheffi and Powell, 1981). On the other hand, when the AADT observations are available across major and minor roads, statistical models can also estimate AADT on all roads. With statistical regression models, AADT has been estimated by capturing variations in observed AADT with predictor variables that explain traffic flow patterns (Das and Tsapakis, 2020; Fu et al., 2017; Pun et al., 2019; Selby and Kockelman, 2013; Sfyridis and Agnolucci, 2020; Yi et al., 2021; Zhao and Park, 2004). Variables included in these models were road network characteristics (e.g., speed limit, number of lanes, road connectivity, and transportation pattern), distance to neighboring roads, and socio-demographic factors (e.g., population, number of vehicles in use, residential area density, workplace area density, and urbanization degree). With these attributes as spatial predictor variables, statistical regression methods were used to estimate AADT at city up to national scale in previous studies. Pun et al. (2019) found random forests (RF) outperformed artificial neural network (ANN), support vector regression (SVR), and linear regression based on root-mean-squared error (RMSE) when estimating AADT in Hong Kong. Das and Tsapakis (2020) reported similar outperformance of RF compared with SVR and generalized linear regression in Vermont in the USA. Sfyridis et al. (2020) also found better performance of RF than linear regression in England and Wales. In the UK, the study by Morley and Gulliver (2016) found a coefficient of determination (R^2) of 0.84 validated by external data in Norwich using an UK-wide statistical method that included road importance determined by a routing algorithm. These studies have shown the capability of statistical methods for predicting AADT on all roads at local to national scales.

However, estimating AADT across Europe comes with the additional challenge that Europe-wide AADT patterns are highly variable and complex. Long-distance, cross-country traveling patterns are mixed with medium and short-distance local commuting patterns which vary by country.

To fill these gaps, we developed a Europe-wide modelling framework using RF to capture variations in AADT and to estimate AADT on all roads across Europe. Our main goal is to capture the variations in the AADT across different road types using RF. Our second goal involves comparison between our Europe-wide AADT estimates and national AADT estimates available in limited countries. Our third goal is to evaluate whether our Europe-wide AADT estimates help improve previous Europe-wide air pollution models which used (major) road length summed in buffers to characterize traffic.

2. Methods

To estimate Europe-wide AADT on all roads, we developed RF AADT models using observed AADT counts. We collected observations of AADT counts from six European countries to train the RF AADT models. The models were trained separately for different road types, because one RF AADT model trained by all observations was dominated by the observed high-AADT counts on motorways in initial modelling and performed poorly on other roads. Thus, the RF AADT models were separated based on four road types: highway (motorway and trunk roads), primary roads, local roads (secondary and tertiary roads), and residential roads, defined in OpenStreetMap (OSM). We also compared the overall performance of our RF models with a fixed-value approach, where the average observed values were assigned for every OSM road type.

For the road-type specific model, we included several road-, population- and urbanization-related predictor variables with various circular buffer sizes (ranging from 50 m to 200 km). With the data and the models, we estimated AADT on 5 m × 5 m grids across Europe, where the road line segments from OSM overlapped with the 5 m × 5 m grid cells. We used the grid approach because it balances computational resources and provides an acceptable approximation of the road information. The resulting AADT map represented the annual average daily traffic (AADT) counts from all lanes and directions on each road.

2.1. Road traffic intensity (AADT) observations

The observed AADT counts were collected from both automatic continuous traffic counters and periodic manual counting. Both approaches captured the number of all types of vehicles passing in both directions on a specific (multi-lane) road segment (i.e., bidirectional AADT counts). The automatic traffic counters collect continuous traffic counts all year round, whereas manual counting collects periodic traffic counts during a day only available in the UK. The periodic manual counts were converted to AADT by the British Department for Transport using expansion factors derived from the observations of the automatic counters separated into road types (Department for Transport, 2019). The observations of the AADT counts were collected from national and local databases in Austria, Switzerland, Germany, France, Italy, and the United Kingdom (UK) where AADT observations were available publicly (Table 1). Data from the UK comprised 70% of the collected traffic count data. Counts on all types of roads were available in all countries (with 3651, 2861, 2356, and 1830 observations on highway, primary, local, and residential roads respectively), although for residential roads almost exclusively from the UK. We also removed observations with AADT higher than 5000 vehicles per day on residential roads to avoid the average estimates from being three times higher than the average observations on residential roads, based on our preliminary analysis described in S3.2 in the supplementary material.

Table 1
Description of traffic intensity observations for each country.

Country	Link to raw data	Observed year	Collecting approach	Number of observing locations
Austria	City of Vienna (City of Vienna, 2020)	2010	Automatic traffic counters	176
Switzerland	Federal Roads Office (Bundesamt für Strassen) (ASTRA, 2022)	2013	Automatic traffic counters	343
Germany	Federal Highway Research Institute (Bundesanstalt für straβenwesen) (Bast, 2022)	2013	Automatic traffic counters	1393
France	Department of Roads of Haute-Garonne, Nouvelle-Aquitaine, and Loiret (Department of Roads of Haute-Garonne, 2022; Department of Roads of; Department of Roads of Nouvelle-Aquitaine, 2020; Department of Roads of Loiret, 2022)	2013	Automatic traffic counters	485
Italy	Ministero delle infrastructure e dei trasporti (Ministry of Infrastructure and Transport, 2022)	2015	Automatic traffic counters	757
UK	Department for Transport (2022)	2013	Automatic traffic counters and manual counting[a]	8084

[a] In the UK, the manual counts were done between March and October by trained enumerators over 12 h. More detailed information can be found in their methodology document (Department for Transport, 2019).

We documented the number of observations by road types and countries in Table 3 and the observing locations in Fig. A1. The observed AADT counts represent the number of all vehicles passing a counting location from both directions per day. We did not attempt to model AADT counts separately for different vehicle types (e.g. heavy-duty vehicles), because the data were limited to fewer countries than the AADT counts of all vehicles.

It was impossible to collect the observations from the same year across different European countries. Therefore, we assumed that the AADT counts did not differ greatly at the locations where observations were collected between the years 2010 and 2015. Our assumption was supported by a high correlation and close-to-one slope between observations collected in 2013 and 2019 in Germany and the UK (as shown in scatterplots of Fig. A2 with Spearman's correlations of 0.99, and slopes of 1.06).

2.2. Predictor variables for AADT models

We included several predictor variables related to road, urbanization, and topology shown in Table 2 and Table A1 because population density and urbanization are shown to be related to mobility patterns and the number of on-road vehicles on the same road type in the road network (Frick and Grimm, 2014; Holz-Rau et al., 2014; Vienneau et al., 2009). Data sources of variables are shown in Table 2, and detailed information can be found in Table A1.

For the urbanization variables, we collected data on population, numbers of passenger cars and commercial vehicles, residential areas, impervious surfaces, and area types (rural and urban) shown in Table 2. We calculated the sum of population within various buffer sizes ranging from 1 km to 200 km at a 1 km spatial resolution to capture mobility patterns ranging from local daily transport patterns to medium-distance and long-distance journeys across cities or countries in Europe. The distance of 200 km represents the average distance of one-way long-distance journeys across cities in Europe (Frick and Grimm, 2014). The small buffer sizes represent the local daily traffic mobility patterns (e.g., 20-km equal to 30-min travel time). To obtain the numbers of passenger cars and commercial vehicles in the buffers, we combined the population density with the information on the density of passenger cars and commercial vehicles by country (ACEA, 2021). A passenger car is defined as a "*vehicle designed and constructed for the carriage of passengers and comprising no more than eight seats in addition to the driver seat, and having a maximum mass not exceeding 3.5 tons*." A commercial car is defined as a "*motor vehicle with at least four wheels used for the carriage of goods*."

We also included road-related variables such as road length, speed limit, and number of lanes from OSM as shown in Table 2. We calculated the sum of the road length of the OSM road segments. For the other variables, the OSM dataset also has information on the number of lanes and speed limit on each road segment. We replaced the missing values with nation-wide information, as described in Fig. 1 and Supplementary text (S3.1). In short, we created 20 m and 10 m buffers around the road segments of highway and non-highway roads to merge information from bidirectional roads. On the bidirectional roads, we averaged the traffic speed limit and number of lanes over the 5-m grids for locations where there were road segments. The same set of 5-m grids was then used to estimate AADT.

For the topology-related variables, we collected information on altitude from the Shuttle Radar Topography Mission (SRTM). The altitude data was then used to calculate the terrain slope data, which represents the local gradient of the 4-connected neighboring grids.

Because some variables had a different spatial resolution, we

Table 2
Overview of predictor variables used in Random Forests. Table A1 gives a detailed description.

Variable type	Predictor variables	Data source	Variable code
Urbanization variables	The sum of population within various buffer sizes (1–200 km)	GEOSTAT (EUROSTAT, 2015)	pop2011_[*buffer in meter*][a]
	The sum of the number of passenger car[b] within various buffer sizes (1–200 km)	ACEA (ACEA, 2021)	pcar2011_[*buffer in meter*][a]
	The sum of the number of commercial vehicles[c] within various buffer sizes (1–200 km)	ACEA (ACEA, 2021)	cveh2011_[*buffer in meter*][a]
	The sum of the number of 100 m grid cells of residential areas within various buffer sizes (0.1–10 km)	CORINE (CLC, 2018)	res_2012_[*buffer in meter*][a]
	The sum of the percentage of sealed impervious surfaces within various buffer sizes (0.1–10 km)	Copernicus (Copernicus, 2021)	imd_2012_[*buffer in meter*][a]
	Area type (Rural or urban)	See Fig. 1 part b.	area_type
Road-related variables	The sum of road length within various buffer sizes (50–500 m) separated by road types	OSM (OpenStreetMap contributors, 2017)	highway_[buffer in meter][a]; primary_[buffer in meter][a]; loca_[buffer in meter][a]; residential_[buffer in meter][a]
	The speed limit on a road segment separated by road types	OSM (see Fig. 1) (OpenStreetMap contributors, 2017)	maxv_20_highway[d]; maxv_10_primary; maxv_10_local[e]; maxv_10_residential
	The number of lanes on a road segment separated by road types	OSM (see Fig. 1) (OpenStreetMap contributors, 2017)	lane_20_highway[d]; lane_10_primary; lane_10_local[e]; lane_10_residential
Topology	Altitude	SRTM DTM (Jarvis et al., 2008)	Elevation
	Slope	SRTM DTM (Jarvis et al., 2008)	Slope

[a] [buffer in meter] is set differently for different variables (buffers of pop2011, pcar2011, and cveh2011: 1, 5, 8, 10, 12, 15, 20, 25, 30, 45, 50, 75, 100, 150, 200 km; buffers of res_2012 and mid_2012: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1, 1.2, 1.5, 1.8, 2, 2.5, 3.5, 4, 5, 6, 7, 8, 10 km; buffers of the sum of road length: 50, 100, 200, 300, 400, 500 m).
[b] Passenger car is defined as "vehicles designed and constructed for the carriage of passengers and comprising no more than eight seats in addition to the driver seat and having a maximum mass not exceeding 3.5 tons."
[c] Commercial car is defined as "motor vehicles with at least four wheels used for the carriage of goods."
[d] Highway consists of OSM road types of motorway, trunk, motorway link, and trunk link.
[e] Local road consists of OSM road types of secondary, secondary link, tertiary, and tertiary link.

resampled all raster maps of the predictor variables to the 5-m grids using the nearest neighbour approach. The 5-m resampled raster maps were used to extract the values of the predictor variables at the observed AADT locations.

2.3. AADT models: random forests (RF)

To capture potential non-linear relationships between predictor variables and AADT counts and to allow interaction between predictor variables, we used RF regression models (Breiman, 2001). RF is a tree-based ensemble algorithm. The algorithm samples a subset of predictor variables randomly with replacement in each split of an individual regression tree. This sampling approach allows highly correlated variables to be included in the RF. Each regression tree is built to minimize a loss function which was mean squared error (MSE) in this study. The sum of the reduction in the impurity at each split is calculated for each predictor variable to obtain the variable importance. The impurity for each node (i.e., the split point) is determined by the within-node sample variance in a regression RF model (Ishwaran, 2015).

We developed separate RF AADT models for each road type. Although developing one RF AADT model for all road types would benefit from the shared information of observations across road types, the model was dominated by the observed high AADT counts on motorways, and thus the significant difference in AADT patterns on other roads cannot be represented. The difference resulted in the poorer predictive power of one RF AADT model than of separate RF AADT models, based on our preliminary test (in Supplementary text S3.2). Moreover, previous studies have shown different mobility patterns across road types (Pulugurtha and Kusam, 2012). Therefore, we decided to develop separate road-type specific RF AADT models with the same set of predictor variables but trained by the observations from each road type separately (1) highway: motorway and trunk roads, 2) primary roads, 3) local: secondary and tertiary roads, and 4) residential roads).

We implemented RF in R using the library *ranger* (Wright and Ziegler, 2017) *version 0.14.1*. The variable importance was represented by the sum of the decrease in the impurity from each split. We used the variable importance to interpret how informative and important a variable was for explaining the variations in AADT. We set the number of trees (*ntree*) as 500 and the number of variables split at each node (*mtry*) as 12, which is the default setting: the square root of the number of variables.

With the trained road-type specific RF AADT models and the resampled 5-m predictor variables, we estimated our final Europe-wide AADT on 5 m × 5 m grids where the OSM road line segments overlapped with the 5 m × 5 m grids.

2.4. Evaluate model performance: five-fold cross-validation (5-fold CV)

To evaluate the performance of the AADT models, we performed five-fold CV, stratified by country and area type (rural and urban areas). The stratification ensured that the number of observations from each country and area type remained similar in different folds.

In the 5-fold CV analysis, we used four performance metrics: adjusted coefficient of determination (R^2), mean-square-error-based R^2 (MSE-R^2), root mean square error (RMSE), and relative RMSE (rRMSE) defined as Equations 1, 2 & 3, respectively. The MSE-R^2 reflects the prediction bias relative to the variance of the observations along the 1:1 line. In contrast, the RMSE reflects the absolute average difference between estimates and observations, and the rRMSE normalizes RMSE by the average of the observations.

$$MSE - R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2} \quad \text{Eq. 1}$$

Table 3
Descriptive statistics of annual average daily traffic (AADT) observations in vehicles per day separated by road types (column) and countries (row). Mean: average of observations; sd: standard deviation of observations; N: number of observations. Road types are defined in OpenStreetMap (OSM): highway (motorway and trunk roads), primary, local (secondary and tertiary roads), and residential roads.

	highway			primary			local			Residential		
	mean	sd	N	mean	sd	N	mean	sd	N	mean	sd	N
Austria	28,601	12,090	5	24,414	929	78	13,281	7207	90	3050	720	3
Switzerland	44,307	28,124	218	8193	5429	90	8865	6767	36	4196	130	2
Germany	47,839	33,407	769	8414	6266	582	7759	4719	37	*NA*	*NA*	0
France	38,338	28,757	85	8453	8061	244	4648	3473	155	*NA*	*NA*	0
Italy	21,800	22,384	220	7913	7041	424	2931	3116	112	*NA*	*NA*	0
UK	49,112	36,606	2354	16,839	11,060	1443	5679	4839	1926	1383	910	1825
Overall	46,632	35,196	3651	13,022	10,361	2861	5852	5118	2356	1392	923	1153

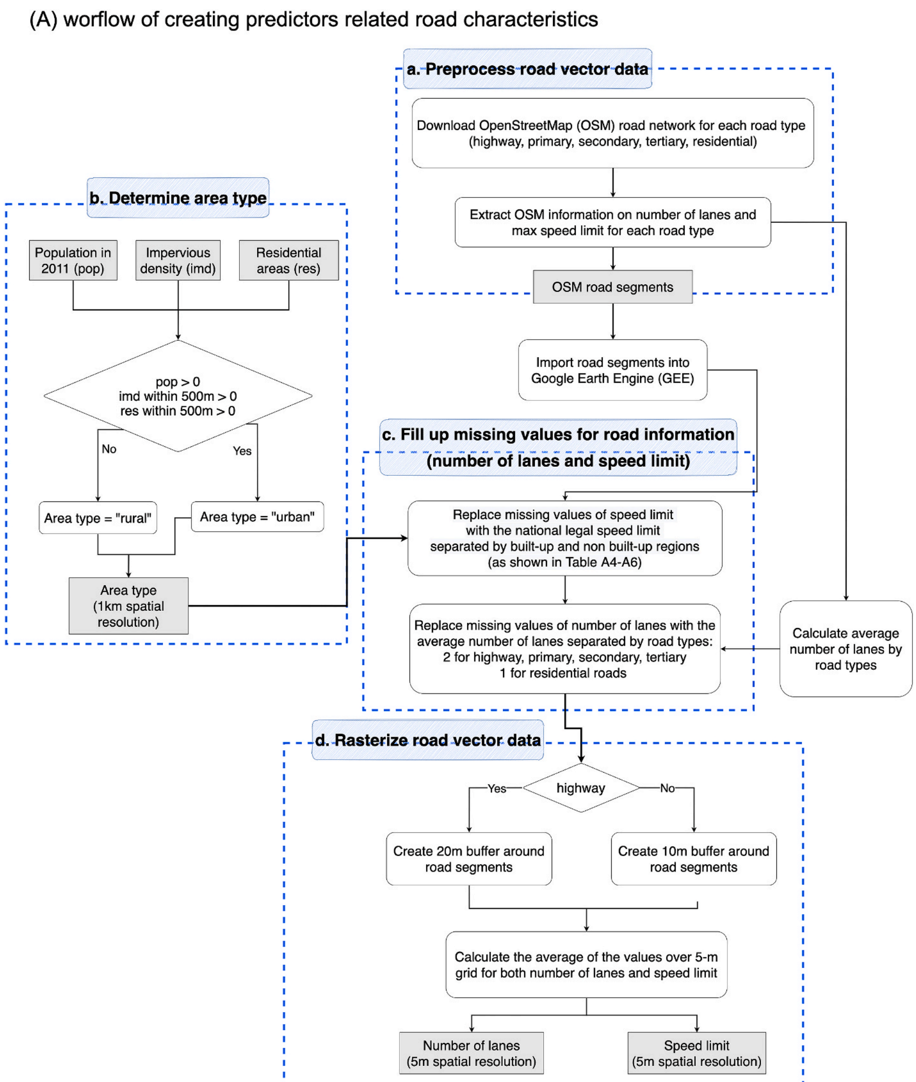


Fig. 1. Workflow of creating road-related variables. Rectangles with rounded corners represent processes, rectangles with sharp corners represent input or output, and diamonds represent decisions. Arrows connect orders and relationships of components.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad \text{Eq. 2}$$

$$rRMSE = \frac{RMSE}{\bar{y}} \qquad \text{Eq. 3}$$

where y_i is the observed AADT counts at location i, $\hat{y}_i$ is the estimated AADT counts at location i, and $\bar{y}$ is the average observations from all N observing locations.

2.5. Comparison with national AADT estimates on road segments and within circular buffers

To support future applications, we investigated the differences between our Europe-wide estimates of AADT counts and existing predictions from national models. We compared our final Europe-wide AADT estimates with national model estimates on road segments (hereafter called the 'road-segment' approach), and we also compared the sum of AADT values within different buffer sizes (hereafter called the 'buffer' approach). These two approaches are related to the applications for estimating road traffic noise (Khan et al., 2021; Morley et al., 2015; Morley and Gulliver, 2016) and for estimating air pollution concentrations using empirical methods (Bechle et al., 2015; Cattani et al., 2017; de Hoogh et al., 2016; Gonzales et al., 2012; Kerckhoffs et al., 2019; Lu et al., 2020; Wolf et al., 2017).

From the national models, we obtained national AADT estimates for the Netherlands and Switzerland. The Swiss national model estimates were obtained from the National Passenger Traffic Model (Nationales Personenverkehrsmodell, NPVM) which is a population-based traffic model built for the year 2017 (das Geoportal des Bundes, 2022). In the Swiss model, the Tomtom road network was used in a simplified version where all road functionalities remained. Then the simplified road network was used to estimate AADT by modelling choices of routes and transportation choices (ARE, 2023). The Dutch data was obtained from the National Road Database (Nationaal WegenBestand, NWB) (NDW, 2021) for the year 2008. In the Dutch population-based traffic model, almost all roads were included in the database. There are intrinsic differences between the datasets. The national traffic models estimated traffic intensity for each road segment of the national road networks, whereas our Europe-wide models estimated traffic intensity on each 5-m grid cell containing a road segment. In addition, the road network used in the national models was different from the OSM road network.

For the 'road-segment' approach, to minimize the intrinsic differences, we discretize the road segments with national AADT estimates using the same approach as described in Fig. 1. This produced AADT values as the bidirectional sum of traffic counts comparable to our Europe-wide AADT estimates at 5-m grids. Then we extracted the 5 m-gridded AADT estimates of both the Europe-wide and national data on a randomly selected subset of OSM road segments. To evaluate the variabilities of 5 m-gridded AADT estimates on each road segment, we further calculated the coefficients of variation ($\frac{standard\ deviation\ (\sigma)}{mean\ (\mu)}$) of the 5-m gridded AADT estimates overlapped with each of the centerlines.

For the 'buffer' approach, for all pixels on the land area of Europe within various circular buffer sizes (50 m, 500 m, 1 km, 2 km, 10 km), we summed up the 5-m gridded AADT values within the buffers. Then we extracted the sum of AADT values of the Europe-wide and national data at randomly selected residential addresses of the European population (Shen et al., 2022).

2.6. Improve previous Europe-wide air pollution models using the modelled AADT estimates

In previously published air pollution models (Shen et al., 2022), we used several predictor variables in the regression models including road-related variables such as the sum of road length in specific buffer sizes but without traffic intensities. Here, we included the newly developed Europe-wide AADT estimates as additional potential predictor variables to investigate whether the air pollution models can be further enhanced. The models were developed for four pollutants (NO_2: nitrogen dioxide, O_3: ozone, PM_{10}: particulate matter <10 μm, $PM_{2.5}$: particulate matter <2.5 μm) in the year 2019.

The Europe-wide air pollution modelling framework, as reported previously, was not changed (Shen et al., 2022). Two sets of potential predictor variables were offered: one excluding AADT variables (i.e., the original setting in the previous study), and the other including them. The original set of predictor variables included several land-use variables, population data, road data that served as a traffic indicator, meteorological and satellite-retrieved data, and chemical transport model estimates. The included AADT variables were our on-road AADT estimates and the sum of the on-road AADT estimates within various circular buffer sizes (50 m, 100 m, 200 m, 300 m, 400 m, 500 m, 700 m, 1 km, 2 km, 5 km). A detailed description of predictor variables can be found in Table A7 and Shen et al. (2022).

For air pollution modelling, our study used two linear regression models—Supervised Linear Regression (SLR) and Geographically Weighted Regression (GWR)—alongside a nonlinear machine learning method: Random Forest (RF).

Supervised linear regression (SLR)

Potential variables were selected step-wise; in each step, the variable is selected if it contributes most to explaining the variance in the response variable (i.e., air pollution concentrations) and only if its direction of effect is realistic (e.g., positive for road variables and negative for urban green). Variables with high p-values (>0.1) and high Variance Inflation Factors (VIF >3) were excluded to ensure model interpretability and to mitigate multicollinearity.

Geographically weighted regression (GWR)

The variables selected by SLR were used, and spatially varying regression coefficients were estimated using an exponentially weighted function decaying by distance. This method addresses spatial heterogeneity in the data.

Random forest (RF)

All potential variables were included. RF, a tree-based ensemble method, handles multicollinearity by randomly selecting a subset of variables at each node split. The hyperparameters—mtry (number of variables split at each node) and ntree (number of trees)—were optimized using out-of-bag (OOB) data, which is unused data since each tree is built using bootstrapped training data and thus not all training data is used in each regression tree.

Details on these models are available in Shen et al. (2022).

Some of the predictors used in AADT modelling were also offered as predictors for air pollution, raising issues of co-linearity. In the SLR and thus GWR procedures (SLR used to select the potential predictor variables for GWR), we checked for variance inflation (VIF) and excluded variables with high VIF (>3). This approach effectively avoids collinearity in the models. The RF trains each tree by randomly selecting subsets of predictor variables, thereby reducing the impact of multicollinearity.

To quantify the improvement of model accuracy, we conducted 5-fold CV, stratifying observations by their station type (background, industrial and traffic) and climate zone (Alpine, Atlantic, Continental, Northern and Southern), following the method described previously (Shen et al., 2022). The evaluation metrics included MSE-R^2 (Eq. (1)) and RMSE (Eq. (2)). We also inspected and compared the spatial maps of the air pollution estimates from models with and without the AADT estimates.

3. Results

We collected observations of AADT from Austria, Switzerland, Germany, France, Italy, and the UK. Most observations were limited to higher classifications of roads such as highway, primary, and local roads (Table 3). In the UK, observations were distributed quite evenly across different road types. In contrast, other countries had limited observations unevenly distributed on highway (35.5%), primary (49.6%), local (18.3%), and residential (0.27%) roads compared to the UK. The variations of the observed AADT were distinct across different road types and countries (shown in Table 3 and Fig. A3), supporting the need for assessing AADT beyond road type.

3.1. Model structure

In our RF AADT models using all observations separated by road types, the most important (top 10) variables varied with road types, as shown in Fig. 2. For the highway model, the information on highway (e.g., the number of lanes, speed limit and the sum of road length within 50 m–500 m) explained most variation in the AADT. Population in large buffers (>10 km) added further prediction. For the primary road model, the sum of primary road length within 50 m–100 m was the most informative variable, but population-related variables within 4 km–8 km were also important in explaining the variation in AADT. For the local road model, the population-related variables 1.8 km–5 km were the most informative. For the residential road model, the information on residential roads and local roads, population-related variables at a local scale (1 km) as well as at a large scale (200 km), and elevation were important in explaining the variation in AADT on the residential roads.

3.2. Model performance

The road-type specific RF AADT models demonstrated strong performance, achieving an overall MSE-R^2 of 0.81, as shown in Table 4. This performance significantly outperformed the fixed-value approach, which had an MSE-R^2 of 0.44. For highway and primary roads, the 5-fold CV model performances were satisfactory, with MSE-R^2 values ranging from 0.60 to 0.69. The performance on local roads was moderate, with a 5-fold CV MSE-R^2 of 0.44. However, the residential roads model struggled to capture AADT variability, as indicated by the close-to-zero MSE-R^2 and R^2 values. The rRMSE values followed similar trends, with higher rRMSE values for residential roads and lower values for other road types, as detailed in Table 4. Excluding the five observations from Austria and Switzerland in the residential model yielded similar model performance, e.g. with MSE-R2 of −0.01 and R2 of 0.01.

The scatterplot (in Fig. 3 and Fig. A4) shows that the agreement was high between the observed and hold-out predicted values (with close-to-1 regression slope) on highway, primary, and local roads. On residential roads, however, the agreement was low and at most observed points (66% of the residential observations) the RF model overestimated the AADT values.

Boxplots in Fig. 4 illustrate that our RF AADT models predict substantial variability of AADT within road type. The averages and variability of AADT counts were similar between observed and estimated values for highway, primary, and local roads. However, for residential roads, the estimated AADT values were, on average, higher than the observed values in the UK at 1825 observation locations, and lower than the observed values on five residential roads in Austria and Switzerland. Consistent with the poor predictive model accuracy, the residential estimates exhibited a narrower distribution compared to the observed values. Overall, for most road types, the distributions of the estimated values closely matched those of the observed values.

Table 4
Model performance of traffic flow models from 5-fold cross-validation separated by road type (highway: motorway and trunk roads, primary, local: secondary and tertiary, and residential roads) and on all road types ("overall"). N: number of observation points.

	RMSE	rRMSE	MSE-R^2	R^2	N
highway	20,072	0.43	0.67	0.69	3651
primary	6532	0.50	0.60	0.60	2861
local	3845	0.66	0.44	0.44	2356
residential	907	0.92	−0.01	0.02	1830
Overall	12,340	0.59	0.81	0.82	10,698

Moreover, we examined variations in the top 15 most informative variables from the residential model. We found high variations in the values of these informative variables within the corresponding models (Fig. A5).

3.3. AADT estimates from RF

To illustrate the intra-and inter-city spatial variations in our 5-m gridded AADT estimates, we show spatial maps of our AADT estimates in Fig. 5 and the extracted values on a subset of road segments in boxplots (Fig. A6) in some selected European cities (Amsterdam, Athens, Barcelona, Basel, Lodz, Munich, and Rome). These cities were selected for visualization because of the potential future use of our AADT estimates for air pollution modelling using mobile data collected in these cities. Motorways in Barcelona and Athens, for example, had higher AADT estimates than other cities shown in the figures, whereas Basel and Lodz had lower AADT estimates in general. Within the cities, the contrasts in AADT were also reflected across road types.

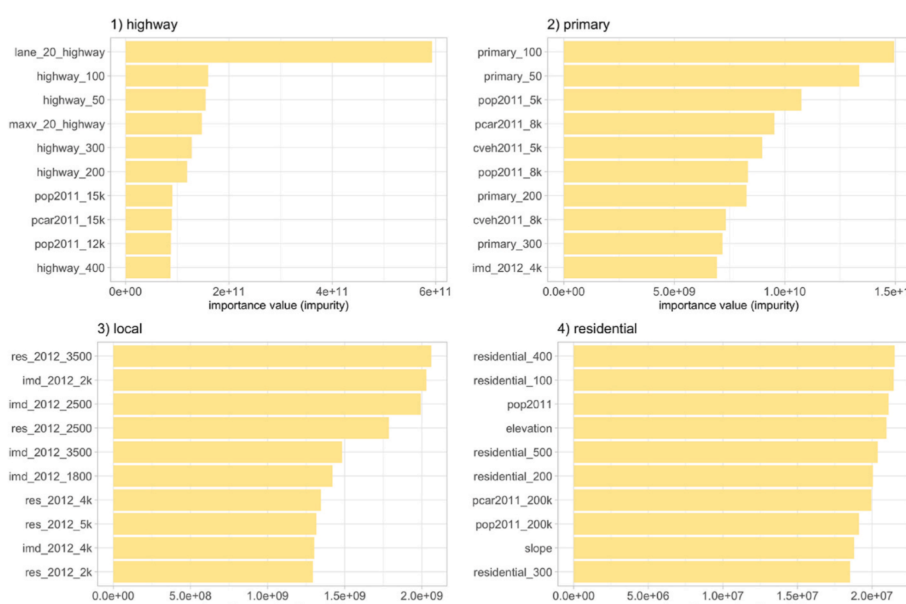


Fig. 2. Top 10 variables with the highest variable importance measured by impurity in four European road-type specific random forest models of road traffic flow: 1) highway (motorway and trunk roads), 2) primary, 3) local (secondary and tertiary roads), and 4) residential roads. An overview of the variables' description can be found in Table 2, and a detailed explanation can be found in Table A1.

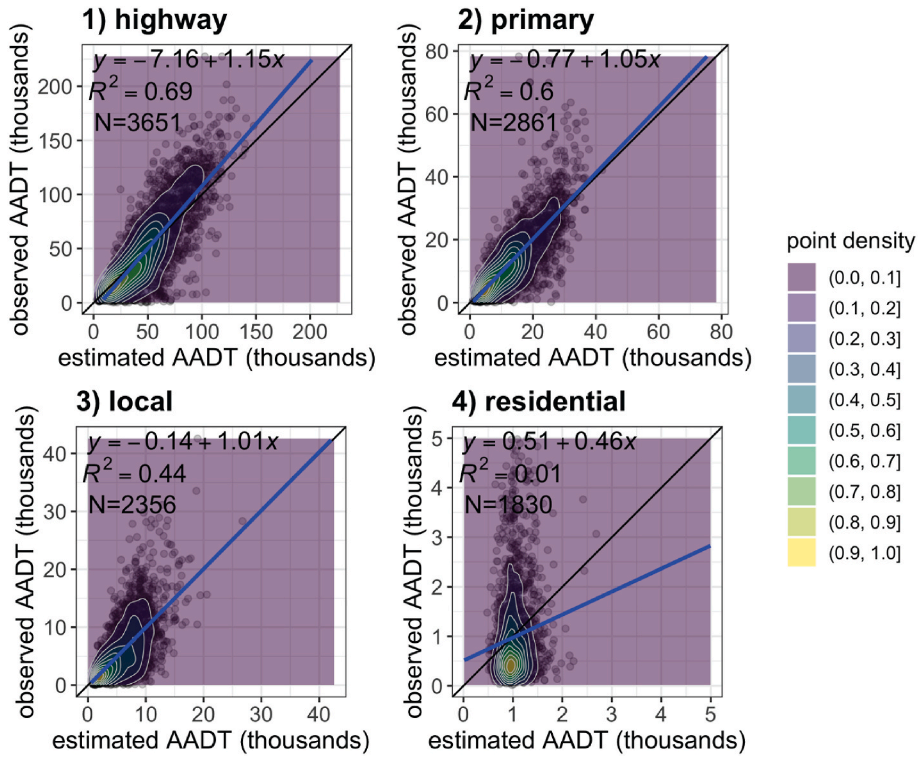


Fig. 3. Scatterplots between observed and estimated annual average daily traffic (AADT) counts in thousands (unit: thousand vehicles/day) on 1) highway, 2) primary, 3) local (secondary, tertiary), and 4) residential roads. The best-fit linear line is in blue with fitted linear regression and coefficient of determination (R^2) indicated in the upper left corner, and the 1:1 line is in black. The number of points (N) is also shown in upper left corner of each panel.

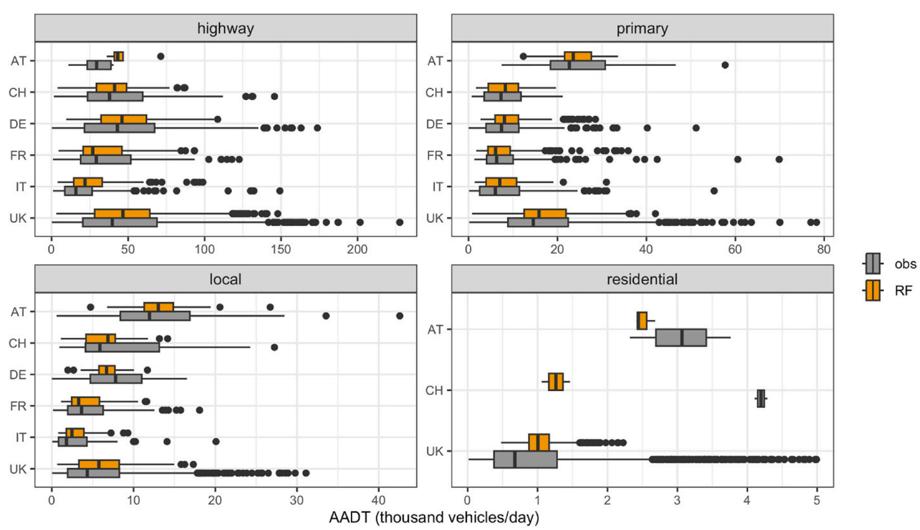


Fig. 4. Boxplots of observed (obs) and estimated (RF) AADT values in thousands (unit: thousand vehicles/day) across different road types (highway: motorway and trunk roads; primary roads; local: secondary and tertiary roads; residential roads as defined in OpenStreetMap) and countries (AT: Austria, CH: Switzerland, DE: Germany, FR: France, IT: Italy, UK: United Kingdom).

3.4. Model comparison: EU AADT model estimates vs national traffic model estimates

3.4.1. Comparison of centerline road segments

Overall, in both Switzerland and the Netherlands, the agreement between national model and Europe-wide AADT model estimates was high (correlation = 0.81–0.84, shown in Fig. A7), higher than a fixed-value approach of using the average by road types (correlation = 0.58–0.59). The agreement was different between road types (Fig. 6).

For the Netherlands, on highway, primary, and local roads, the overall correlation between the Dutch national model and European model estimates was moderate to high (correlation = 0.60–0.73) with proportional bias, based on the scatterplot and Bland Altman plots (Fig. 6 & Fig. A8). The proportional bias was only limited to some road segments (15%–20%) represented by the points falling out of the 95% limits of agreement lines (i.e., the dotted black lines in Fig. A8). On residential roads, however, the agreement between the two datasets was low (correlation = 0.16) with the European traffic model giving overall higher traffic flow estimates than the Dutch national model (on 85% of the selected residential roads) (Fig. A8).

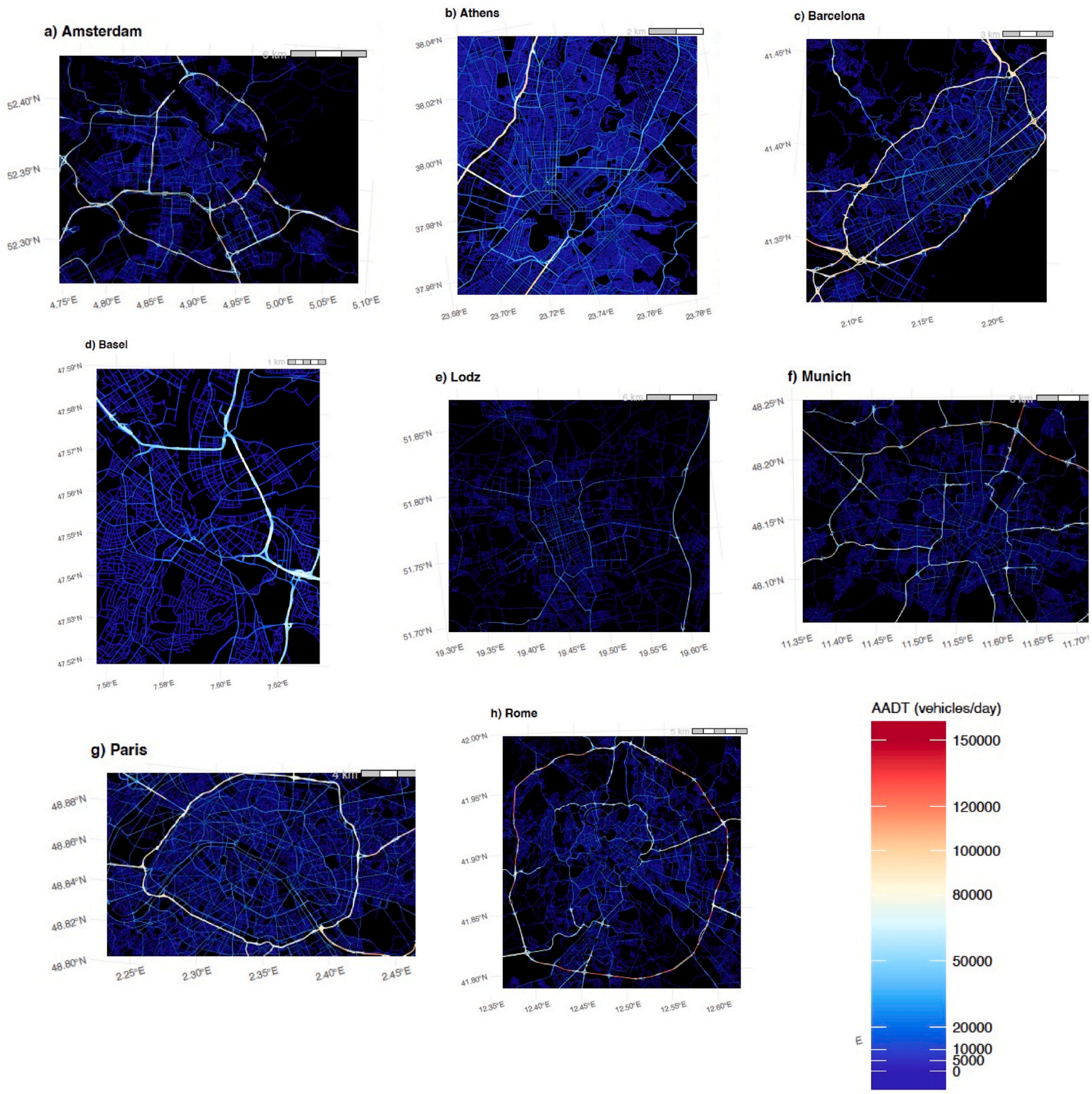


Fig. 5. Spatial maps of annual average daily traffic (AADT) counts in thousands (unit: thousand vehicles/day) estimated by road-type specific RF AADT models in a) Amsterdam, b) Athens, c) Barcelona, d) Basel, e) Lodz, f) Munich, g) Paris, and h) Rome. Coordinates in meter are shown in the x-axis and y-axis in projection EPSG: 3035.

For Switzerland, on highway, primary, and local roads, the overall agreement of the traffic flow estimates from the national Swiss model and European model was moderate to high, as indicated by the correlation (0.60–0.77) (Fig. 6). On these three types of roads, the estimates from the two had no obvious systematic bias, based on the Bland Altman plot (Fig. A8). On the residential roads, there was little agreement between the two models, (correlation = 0.03). On most of the selected residential roads (80% of the selected residential roads), the European model gave higher traffic flow estimates than the Swiss national model (Fig. A9).

In both the Netherlands and Switzerland, the variabilities of the 5-m gridded Europe-wide AADT estimates were overall low (with coefficients of variation lower than 0.15 shown in Fig. A10). Only very few road segments near the coastlines in the Netherlands had high coefficients of variation, caused by the missing pixel values near the coastlines.

3.4.2. *Comparison of sum within circular buffers at random points*

For the Netherlands, the correlation was high (0.85–0.99) between Europe-wide traffic flow estimates and national model estimates at 1795 random points, although overall the sum of national Dutch traffic flow estimates was three times higher than the sum of Europe-wide estimates with different circular buffer sizes (Fig. A11). The difference was propagated from the higher Dutch traffic flow estimates on some road segments observed in section 3.4.1. Such systematic differences between the two datasets were also observed in the Bland Altman plot (Fig. A12), which shows proportional bias between the two.

For Switzerland, the correlation was similarly high (0.91–0.98) between the Europe-wide and national model estimates at 999 random points. For all buffer sizes, locations with a high AADT sum consistently had higher values from the Swiss national model than from the European model, but at locations with a low AADT sum this was turned around (AADT of Swiss model < AADT of European model). Proportional bias was also found in the Bland Altman plot (Fig. A12) for points with a higher sum of traffic intensity within all buffers.

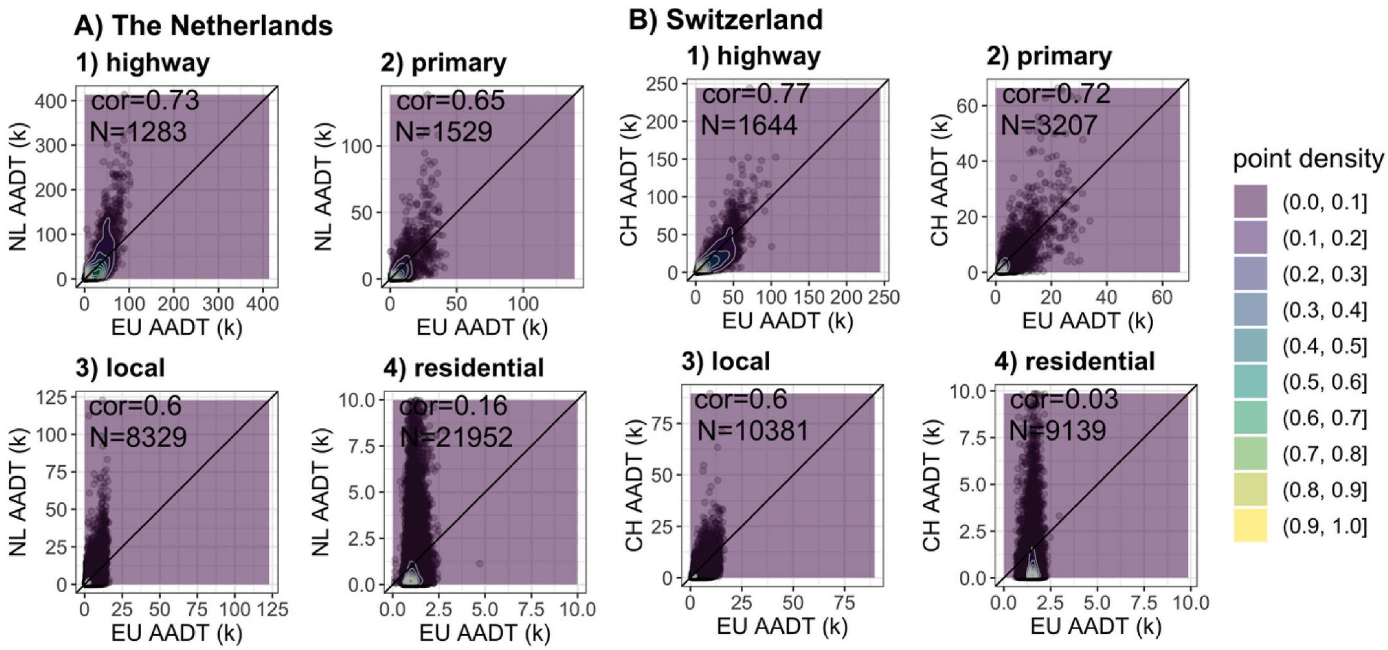


Fig. 6. Scatterplots between national traffic flow model estimates and Europe-wide random forest model estimates on annual average daily traffic (AADT) counts in thousands (unit: thousand vehicles/day) on a randomly selected subset of road segments separated by road types (highway: motorway and trunk roads, primary, local: secondary and tertiary roads, residential) in A) The Netherlands and B) Switzerland. Values were extracted on randomly selected road segments (with AADT counts' unit as thousand vehicles per day). Black lines are 1:1 line, and blue lines are regression lines. Correlation (cor) and number of points (N) are presented in the upper left corner of each scatterplot.

3.5. Improvement of previous air pollution modelling by adding the AADT estimates

We offered the Europe-wide AADT estimates as potential predictor variables in our previous air pollution models for the year 2019. In each pollutant model (NO_2, O_3, PM_{10}, and $PM_{2.5}$), AADT-related variables were selected and used as predictor variables. On-road AADT estimates were selected by SLR for NO_2 and PM_{10}. For O_3 and $PM_{2.5}$, the sum of AADT estimates within 50 m were selected. The inclusion of traffic flow variables replaced some road variables (such as sum of length of major roads within 100 m for NO_2 and $PM_{2.5}$), as shown in Fig. 7. With traffic flow variables provided, for NO_2, we observed that more land-use variables with both small and large buffer sizes were selected, while the outdated but high-resolution (10 × 10km) NO2 MACC chemical transport estimates were replaced by up-to-date but coarser-resolution (13 × 24km) satellite retrieved OMI (Ozone Monitoring Instrument) data. Similarly, for PM_{10}, with traffic flow variables included, more land-use related variables were further selected while satellite retrieved data from Aerosol Optical Depth (AOD) MODIS Blue band (0.47 μm) were replaced by MODIS green band (0.55 μm).

Shown in Fig. A13, the top 10 informative variables in the RF air pollution models remained similar with the inclusion of traffic flow data for all pollutants, except for NO_2. For NO_2, the informative variables in the RF models were dominated by traffic flow data (Fig. A13). Interestingly, the models with AADT offered still included a buffer with major road length in addition to the AADT at the nearest road.

When the AADT variables were included, the 5-fold CV results showed improvement in model performance evaluated by MSE-R^2 and RMSE. The MSE-R^2 increased by 0–4% (3% for NO_2, 4% for O_3, 3% for PM_{10}, and 0% for $PM_{2.5}$), and the RMSE decreased by 0.02–0.42 μg/m^3 (Fig. A14). The improvement was larger for more traffic-related pollutant NO_2 than for more regionally-varying pollutant $PM_{2.5}$ for which no gain in performance was found. The increases in MSE-R^2 were larger (NO_2: 5% increase, O_3: 8%, PM_{10}: 4%, $PM_{2.5}$: 2%) in urban areas with population density larger than 5000 km^{-2} (Fig. 8) than in rural areas (where increases ranged from 0 to 3%, as shown Fig. A15).

The spatial maps of NO_2, a road traffic-related pollutant, were broadly similar for models with and without AADTs included, as shown in Fig. 9. However, AADT maps for most cities showed more granularity than maps with road length (i.e., without AADTs). Consistently, distinct differences between the two models were found in near-road concentrations. More intra-city variations were captured when our AADT estimates were used in the model (Fig. 9). The contrasts in the estimated concentration (presented by 95th minus 5th percentile) in most cities increased when AADT was included (with increases of 0.18–1.55 μg/m^3), shown in Fig. A16.

4. Discussion

Our traffic flow models showed overall high performance based on 5-fold CV ($R^2 = 0.82$). The resulting traffic flow estimates for all European roads further enhanced our previous air pollution models by capturing more detailed near-road variations and spatial contrasts, and modest improvements in model performance for NO_2, PM_{10} and O_3 especially in urban areas (up to 8% increase in R^2).

4.1. Europe-wide traffic models

4.1.1. Model structure

The RF models selected different important predictor variables consistent with the type of road. Population with large buffers (>5 km) were informative in all models. This information helped distinguish the differences between rural and urban areas with distinct traffic patterns (Morley and Gulliver, 2016; Sfyridis and Agnolucci, 2020).

4.1.2. Model performance and variability in AADT estimates

Based on the 5-fold CV result, our RF models performed well overall, and we observed good to moderate performance on highway, primary, and local roads. On residential roads, the performance was no better than the overall average AADT on residential roads. Similarly, the comparison between national AADT model estimates and our Europe-wide AADT estimates yielded an acceptable agreement on highway,

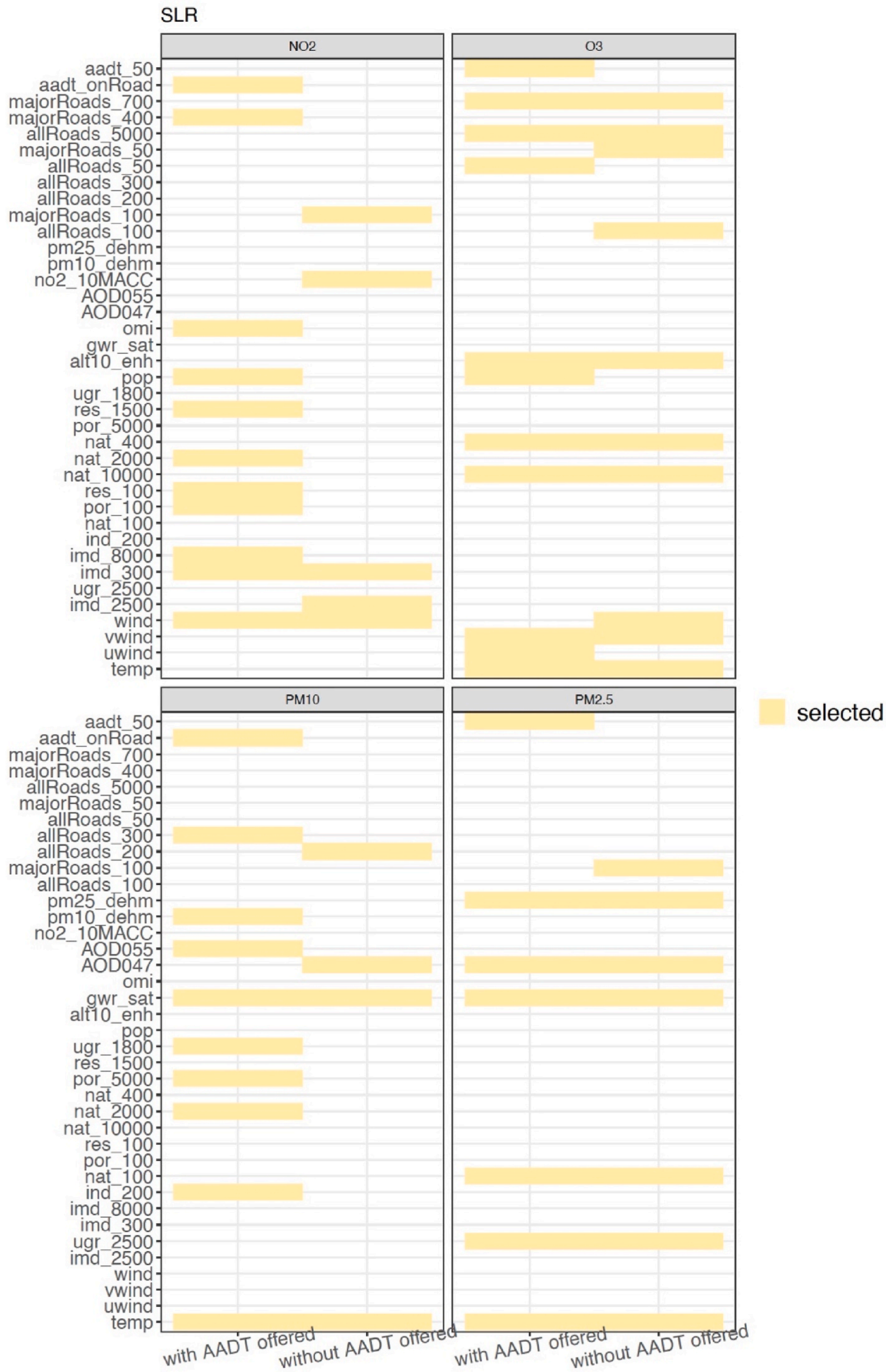


Fig. 7. Variables selected in air pollution SLR models with and without traffic flow variables offered (i.e., aadt_onRoad and aadt_[buffer in m]). A detailed explanation of predictor variables can be found in Table A7.

primary, and local roads but a low agreement on residential roads. The poor performance of our residential model may be partly due to higher measurement errors from manual counting on residential roads compared to the other road types. Another explanation is that we did not include routing information or road connectivity as predictor variables, which are influential in distinguishing AADT variations on low-traffic local and residential roads (Alvarado-Molina et al., 2023; Morley and Gulliver, 2016). The informative variables in the residentials were mainly population, topology, and sum of residential road lengths. Although these variables showed great variations (Fig. A5), such variations can only reflect on the overall average AADT on residential roads rather than the local variations in AADT.

Overall, the model predicted substantial variability of AADT within road-type sites, documenting the benefit of estimating traffic intensity beyond using road type to represent motorized traffic in air pollution and noise modelling. While we developed models per road type, all roads will be combined for the future application. Therefore, the low model accuracy for residential roads is not a major problem, as road traffic intensity is low on residential roads relative to other road types. Assigning an average low road traffic intensity to all residential roads is a reasonable approximation for future application of air pollution and road traffic noise modelling.

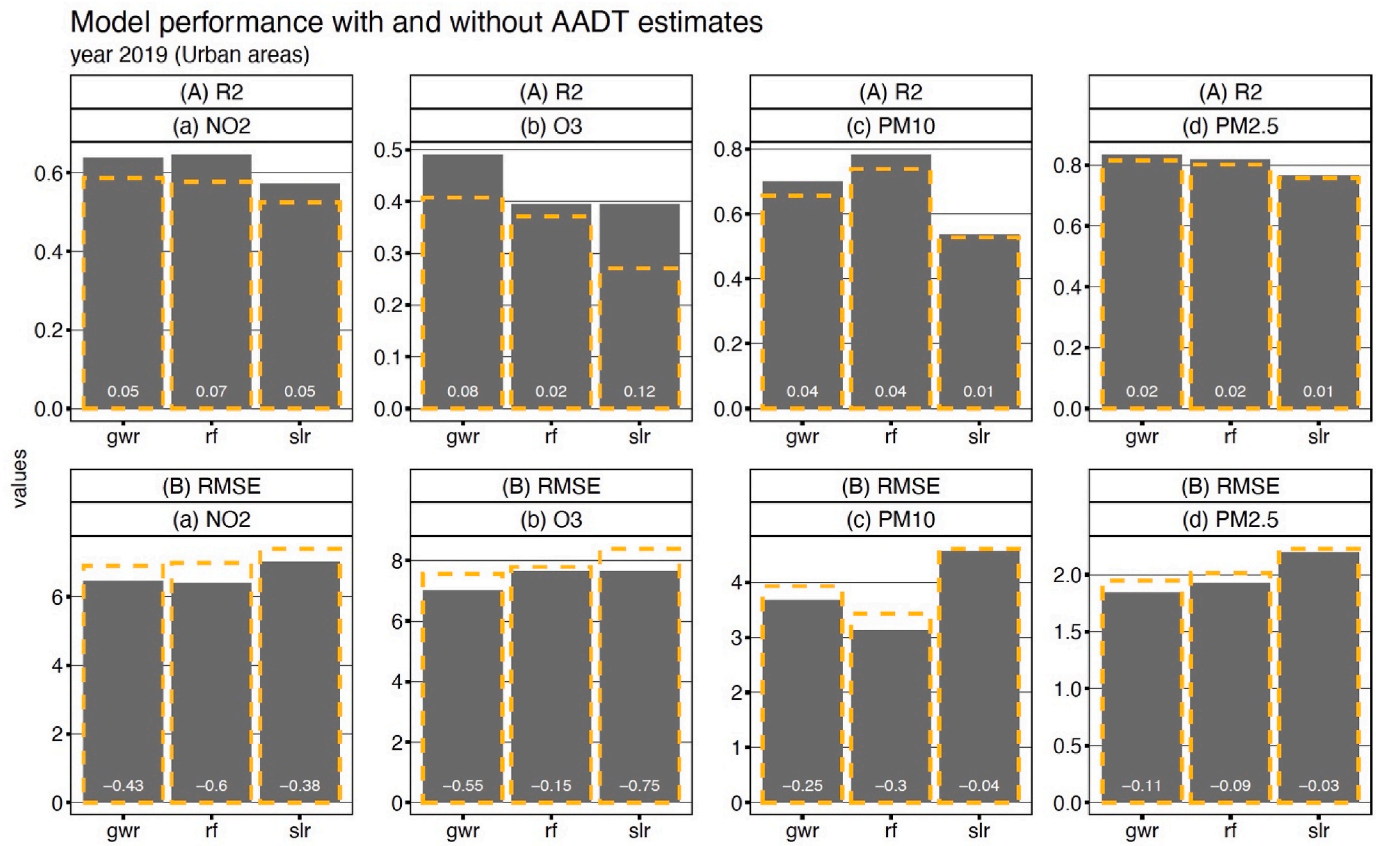


Fig. 8. Five-fold cross-validated model performance of air pollution models with and without AADT estimates, evaluated by two metrics: (A) mean-square-error-based R2 (MSE-R2) and (B) root-mean-square-error (RMSE) at stations located in urban areas where population density is larger than 5000 km^{-2}. Models were built for four pollutants: (A) NO2, O3, PM10 and PM2.5 for year 2019. Bars with dashed lines present models with AADT estimates and with solid grey fills present models without AADT estimates. Regression approaches include gwr (geographically weighted regression), rf (random forests), and slr (supervised linear regression). Text in white at the bottom of each bar shows the difference of metric values between models with and without AADT estimates.

4.1.3. Comparison between European and national AADT estimates

When we compared our Europe-wide AADT estimates with national AADT estimates on a subset of road segments in Switzerland and the Netherlands, the overall correlation was high, better than the fixed-value approach. We found moderate agreement on highway, primary and local roads, but the agreement was low on residential roads. The Europe-wide traffic flow estimates exhibited proportional bias on a small subset of the selected residential roads compared with national traffic flow estimates. We note that national AADT estimates are also based on modelling, but from deterministic traffic flow models (ARE, 2023; NDW, 2023). These models have considerable uncertainty for individual road segments, similar to our models (ARE, 2024; NDW, 2023). This indicates that the agreement between our and national AADT estimates is reasonable.

When the Europe-wide AADT estimates were used as a raster of the sum within circular buffer, the Europe-wide traffic flow estimates showed good agreement with national traffic flow estimates within all buffer sizes.

The comparison result indicates that similar prediction results for air pollution modelling are likely when using the sum of Europe-wide AADT estimates with buffers compared to the sum of national traffic flow estimates. But if Europe-wide traffic flow estimates are used on road line segments for air pollution or road traffic noise modelling, a limited number of road segments may produce spatially heterogenous predictions compared to national estimates. Europe-wide estimates also capture less variability in residential road traffic emissions (both air pollution and noise) than national estimates. Regardless, due to the high correlation between the datasets, overall results are expected to remain similar for all road types in air pollution modelling, although local differences may arise when using traffic flow predictor variables from European and local national traffic models in the application of estimating air pollution concentrations or road traffic noise levels.

4.2. Improvement of air pollution modelling with modelled AADT estimates

When including estimated traffic flow data in the LUR air pollution models, we found modest increases in MSE-R^2 for more traffic-related pollutants (NO_2, O_3) and only minor increases for $PM_{2.5}$, which is not strongly influenced by local traffic. The modest increase in MSE-R^2 suggests that widely available traffic flow proxies, such as road length of (major) roads, already represent the important impact of local traffic emissions reasonably well. However, including estimated traffic flow provided more granularity in near-road air pollution concentrations compared to using proxies like road length.

Traffic flow data has been shown to be crucial for air pollution modelling at a city- and multi-cities scales in Europe, North America and Australia (Beelen et al., 2013; Boniardi et al., 2019; De Hoogh et al., 2019; Dijkema et al., 2011; Eeftens et al., 2012; He and Huang, 2018; Jones et al., 2020; Lu et al., 2020; Naughton et al., 2018; Shi et al., 2020; Van den Bossche et al., 2018; Wang et al., 2014; Weissert et al., 2020; Wen et al., 2023; Yang et al., 2020). Using only road type and road length as proxies for traffic flow can miss variations in traffic-related emissions.

While traffic flow estimates were available for some countries, there were no publicly available continental-wide traffic flow estimates for air

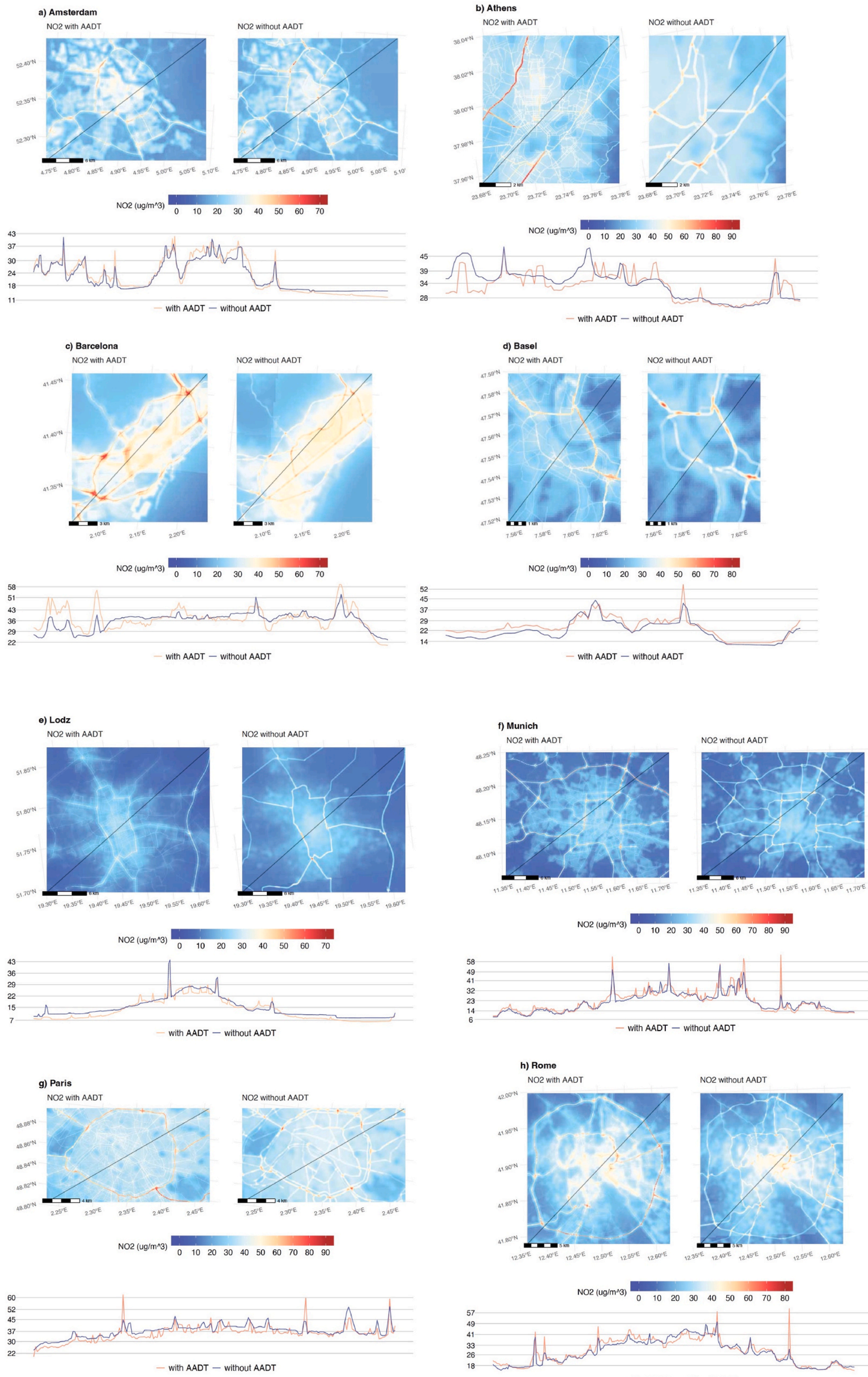


Fig. 9. Spatial maps of estimated NO_2 concentrations (unit: $\mu g/m^3$) without and with AADT estimates included as predictor variables in the geographically weighted regression model for a) Amsterdam, b) Athens, c) Barcelona, d) Basel, e) Lodz, f) Munich, g) Paris, and h) Rome. Estimated concentrations along a diagonal transect are shown in the line chart below spatial maps for each city.

pollution modelling. Here our traffic flow estimates were available on every road segment of OSM (i.e., motorway, trunk, primary, secondary, tertiary, and residential roads defined in OSM). The completeness of the OSM road network data is high in most European countries (>95%) (Barrington-Leigh and Millard-Ball, 2017), and the OSM data quality is comparable to national road data (Graser et al., 2014; Neis et al., 2011). In both our studies and other national studies, the improvement of traffic flow data was found similarly. In some national or city-wide studies, traffic flow estimates were estimated by regression models. For example, a recent study in Australia showed that estimating traffic flow on minor roads improves the modelling of traffic-related pollutants like black carbon and ultrafine particles (Alvarado-Molina et al., 2023). Another study in Los Angeles in U.S. also demonstrated significant improvement in estimated concentrations for NO_2, O_3, and $PM_{2.5}$ (increases of 21%, 3% and 7% in CV-R^2 respectively) with traffic flow data considered in LUR models (Wen et al., 2023). In other national studies, they used national traffic intensity data estimated by traffic demand models to improve air pollution modelling, such as studies from the ESCAPE project in Europe for air pollution models for NO_2 and PM (Beelen et al., 2013; Eeftens et al., 2012; Wang et al., 2014) in multiple European cities. They reported significantly improved performance of air pollution models, particularly for NO_2 where a 10% increase in R^2 was observed when including local traffic intensity data (Beelen et al., 2013). The minor R^2 increase in $PM_{2.5}$ models can be attributed to its nature as a regional pollutant with complex aerosol formation (Wang et al., 2014) and the substantial contribution from satellite-retrieved data and chemical transport model estimates (de Hoogh et al., 2016). For O_3 in our study, the relatively larger impact of traffic flow data on air pollution model accuracy reflects scavenging of O_3 by nitric oxide (NO), which is emitted from combustion of fossil fuels.

Some predictor variables thematically overlapped in our traffic flow and air pollution models, potentially introducing collinearity issues in air pollution models with traffic flow estimates offered as potential variables. To address this, in our air pollution modelling, we used SLR to select the most contributing variables for GWR, excluding those with a high variance inflation factor (VIF>3). Consequently, this exclusion reduces the impact of multi-collinearity. The RF air pollution model is robust to collinearity because RF randomly selects subsets of predictor variables, minimizing the impact of multi-collinearity. Thus, our air pollution models mitigated collinearity issues, even with overlapping predictor variables between the traffic flow and air pollution models.

4.3. Strengths and limitations

We developed a modelling framework using the random forest algorithm to develop models stratified by road type. The results showed several strengths. Firstly, our Europe-wide RF models, separated by road type, effectively captured variations in AADT across different road types with a high overall 5-fold CV accuracy. Secondly, the road-type specific RF models outperformed the traditional fixed-value approach that assumes the same AADT per road type. Thirdly, the availability of harmonized AADT estimates can improve air pollution and road traffic noise modelling, especially when higher-quality commercial AADT data is not available nor affordable for academic research. However, across different road types, the accuracy of the RF models varied. For all road types except residential roads, the model performance was moderate to good. The residential road model failed to capture local variations in AADTs and only reflected on regional variation, with an accuracy as good as using average observed values. Furthermore, there are some limitations to consider for future applications.

Firstly, harmonized road networks and traffic information (e.g., number of lanes, speed limit) are lacking across Europe. Collecting and creating harmonized information would be cumbersome but not impossible. Some of this information is available in OSM, but data quality and availability, especially for speed limits and number of lanes, vary across Europe.

This variation is reflected in the proportion of missing values separated by country and road type (Tables A2 and A3). Despite these data quality issues, the completeness of the OSM road network is high in most European countries (>95%) based on a previous study (Barrington-Leigh and Millard-Ball, 2017). OSM provides the most well-organized and freely accessible database for all of Europe.

Secondly, the AADT observations were dominated by UK data, especially on local and residential roads. We assumed that the mobility patterns observed in the UK could be extrapolated to represent the mobility patterns across Europe (Vienneau et al., 2009). Prediction accuracy on residential roads could be further improved with more observations from more European countries.

Thirdly, we obtained road length within different circular buffer sizes in Euclidian distance. The Euclidian distance, however, cannot reflect the actual road distance. On some residential roads in the centres of European cities, vehicles are not allowed or limited (e.g., low emission zones in some cities such as London and Barcelona). This information was missing across Europe for our RF models. Therefore, it could result in overestimating traffic intensity and further overestimating in air pollution concentrations and road traffic noise for potential application in these limited low-emission zones.

Moreover, our RF models relied on OSM definitions to represent the importance or connectivity of road traffic. Due to local variations in road importance or connectivity, traffic patterns on the same OSM road type may differ within and between countries. Instead of grouping observations by the OSM-defined road types, clustering observations based on observed values could potentially better capture variations in AADT caused by road importance (Sfyridis and Agnolucci, 2020). However, this approach has difficulty estimating AADT on roads without observations that cannot be assigned to existing clusters.

In addition, for modelling road traffic noise and air pollution, diurnal variations in AADT are important. Although our modelling framework was built for estimating long-term AADT counts, it can be extended to estimate daily or hourly AADT by including corresponding observations and/or predictor variables.

Finally, there might be some biases in the comparison result due to information loss when we converted between raster and vector data of road segments. However, these issues can be mitigated by the small raster grid cells we used (5 m). Despite these limitations, our Europe-wide RF AADT models showed overall good accuracy based on CV and agree generally well with national AADT models.

5. Conclusion

The Europe-wide AADT models demonstrated satisfactory overall accuracy, despite variations in predictive accuracy across road types. Our findings highlight the variability of AADT both within and across road types, documenting the advantages of our approach over the use of fixed values. This study introduces a novel methodology for estimating traffic flow at a continental scale. Our AADT estimates showed overall strong agreement with national AADT model estimates. The 5-m gridded AADT estimates for all roads across Europe will be highly beneficial for future modelling applications, enhancing air pollution modelling and addressing critical gaps in harmonized road traffic noise modelling across Europe. We have demonstrated enhanced air pollution modelling performance, particularly in urban areas, with more refined spatial patterns.

CRediT authorship contribution statement

Youchen Shen: Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Kees de Hoogh:** Writing – review & editing, Supervision, Methodology, Investigation, Data curation, Conceptualization. **Oliver Schmitz:** Software, Investigation, Formal analysis, Data curation, Conceptualization. **John Gulliver:** Writing – review & editing, Methodology, Formal analysis, Data curation, Conceptualization. **Danielle Vienneau:** Writing – review & editing, Validation, Investigation, Data curation. **Roel Vermeulen:** Supervision, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Gerard Hoek:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization. **Derek Karssenberg:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization.

Declaration of competing interest

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

Data availability

All code and data to build models in R programming language are available on Github: https://github.com/co822ee/eu_roadTraffic

Acknowledgements

This work was supported by EXPANSE and EXPOSOME-NL projects. The EXPANSE project is funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 874627. The content of this article is not officially endorsed by the European Union. The EXPOSOME-NL project is funded through the Gravitation programme of the Dutch Ministry of Education, Culture, and Science and the Netherlands Organization for Scientific Research (NWO grant number 024.004.017). The authors declare no competing financial interest. We thank technical and information support from Benjamin Flueckiger from Swiss Tropical and Public Health Institute (Swiss TPH) and Marta Cirach from Barcelona Institute for Global Health (ISGlobal).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2024.120719.

Abbreviations

AADT	annual average daily traffic
CV	cross-validation
GWR	geographically weighted regression
MSE	mean square error
OSM	OpenStreetMap
RF	random forests
SLR	supervised linear regression
RMSE	root mean square error

References

ACEA, 2021. Report - Vehicles in Use, Europe 2021. ACEA - European Automobile Manufacturers' Association, European Automobile Manufacturers' Association.

Alvarado-Molina, M., Curto, A., Wheeler, A.J., Tham, R., Cerin, E., Nieuwenhuijsen, M., Vermeulen, R., Donaire-Gonzalez, D., 2023. Improving traffic-related air pollution estimates by modelling minor road traffic volumes. Environ. Pollut. 338, 122657 https://doi.org/10.1016/j.envpol.2023.122657.

ARE, 2024. Model validation and calibration [WWW Document]. URL. https://www.are.admin.ch/are/de/home/mobilitaet/grundlagen-und-daten/verkehrsmodellierung/npvm/validierung.html. (Accessed 3 April 2024).

ARE, 2023. Introduction NPVM [WWW Document]. URL. https://www.are.admin.ch/are/de/home/mobilitaet/grundlagen-und-daten/verkehrsmodellierung/npvm/einfuhrung.html. (Accessed 19 January 2024).

ASTRA, 2022. Bundesamt für Strassen (ASTRA) [WWW Document]. URL. https://www.astra.admin.ch/astra/de/home.html. (Accessed 10 October 2023).

Barrington-Leigh, C., Millard-Ball, A., 2017. The world's user-generated road map is more than 80% complete. PLoS One 12, e0180698. https://doi.org/10.1371/journal.pone.0180698.

Bast, 2022. BASt - automatische straßenverkehrszählung - automatische Zählstellen 2017 [WWW Document]. URL. https://www.bast.de/DE/Verkehrstechnik/Fachthemen/v2-verkehrszaehlung/Daten/2017_1/Jawe2017.html?nn=1819490. (Accessed 6 February 2022).

Bechle, M.J., Millet, D.B., Marshall, J.D., 2015. National spatiotemporal exposure surface for NO2: monthly scaling of a satellite-derived land-use regression, 2000-2010. Environ. Sci. Technol. 49, 12297–12305. https://doi.org/10.1021/acs.est.5b02882.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dedele, A., Gražulevičiene, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - the ESCAPE project. Atmos. Environ. 72, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037.

Boniardi, L., Dons, E., Campo, L., Van Poppel, M., Int Panis, L., Fustinoni, S., 2019. Annual, seasonal, and morning rush hour Land Use Regression models for black carbon in a school catchment area of Milan, Italy. Environ. Res. 176 https://doi.org/10.1016/j.envres.2019.06.001.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Cattani, G., Gaeta, A., Di Menno di Bucchianico, A., De Santis, A., Gaddi, R., Cusano, M., Ancona, C., Badaloni, C., Forastiere, F., Gariazzo, C., Sozzi, R., Inglessis, M., Silibello, C., Salvatori, E., Manes, F., Cesaroni, G., 2017. Development of land-use regression models for exposure assessment to ultrafine particles in Rome, Italy. Atmos. Environ. 156, 52–60. https://doi.org/10.1016/j.atmosenv.2017.02.028.

Chen, J., De Hoogh, K., Gulliver, J., Hoffmann, B., Hertel, O., Ketzel, M., Weinmayr, G., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Atkinson, R., Janssen, N.A.H., Martin, R.V., Samoli, E., Andersen, Z.J., Oftedal, B.M., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Brunekreef, B., Hoek, G., 2020. Development of europe-wide models for particle elemental composition using supervised linear regression and random forest. Environ. Sci. Technol. 54, 15698–15709. https://doi.org/10.1021/acs.est.0c06595.

City of Vienna, 2020. Road traffic census 2010 Vienna - data set - data.gv.at [WWW Document]. URL. https://www.data.gv.at/katalog/dataset/a0279fdf-230e-493d-bd50-3865ad29f825. (Accessed 6 February 2022).

CLC, 2018. CORINE land cover 2018 — copernicus land monitoring service [WWW Document] Copernicus L. Monit. URL. https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=metadata%0Ahttps://land.copernicus.eu/pan-european/corine-land-cover/clc2018%0Ahttps://land.copernicus.eu/pan-european/corine-land-cover/clc2018%0Ahttps://land.copernicus.eu/pan-europea. (Accessed 25 January 2021).

Copernicus, 2021. Imperviousness [WWW Document] Copernicus Progr. URL. https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness. (Accessed 9 March 2021).

das Geoportal des Bundes, 2022. data.geo.admin.ch [WWW Document]. URL. https://data.geo.admin.ch/. (Accessed 11 August 2023).

Das, S., Tsapakis, I., 2020. Interpretable machine learning approach in estimating traffic volume on low-volume roadways. Int. J. Transp. Sci. Technol. 9, 76–88. https://doi.org/10.1016/j.ijtst.2019.09.004.

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., 2018. Spatial PM2.5, NO2, O3 and BC models for Western Europe – evaluation of spatiotemporal stability. Environ. Int. 120, 81–92. https://doi.org/10.1016/j.envint.2018.07.036.

de Hoogh, K., Gulliver, J., Donkelaar, A. van, 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., 2016. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10. https://doi.org/10.1016/j.envres.2016.07.005.

De Hoogh, K., Saucy, A., Shtein, A., Schwartz, J., West, E.A., Strassmann, A., Puhan, M., Röösli, M., Stafoggia, M., Kloog, I., 2019. Predicting fine-scale daily NO2 for 2005-2016 incorporating OMI satellite data across Switzerland. Environ. Sci. Technol. 53, 10279–10287. https://doi.org/10.1021/acs.est.9b03107.

Department for Transport, 2022. GB Road Traffic Counts - Data.gov.uk [WWW Document]. URL. https://www.data.gov.uk/dataset/208c0e7b-353f-4e2d-8b7a-1a7118467acc/gb-road-traffic-counts. (Accessed 15 June 2022).

Department for Transport, 2019. Road Traffic Estimates Methodology Note.

Department of Roads of Haute-Garonne, 2022. Comptage routier sur la voirie départementale en Haute-Garonne — Haute-Garonne Open Data [WWW Document]. URL. https://data.haute-garonne.fr/explore/dataset/comptage-routier-sur-la-voirie-departementale/map/?location=6,44.73113,2.15332&basemap=jawg.streets. (Accessed 6 March 2022).

Department of Roads of Loiret, 2022. Trafic routier - Sections de comptage - Département du Loiret - 2020 - data.gouv.fr [WWW Document]. URL. https://www.data.gouv.fr/fr/datasets/trafic-routier-sections-de-comptage-departement-du-loiret-2020/. (Accessed 6 March 2022).

Department of Roads of Nouvelle-Aquitaine, 2020. Nouvelle-Aquitaine : Traffic routier des réseaux autoroutiers non concédé. Nat, et Départ. - Localisat. (ponctuel) - data.

gouv.fr [WWW Document]. URL. https://www.data.gouv.fr/fr/datasets/nouvelle-aquitaine-trafic-routier-des-reseaux-autoroutiers-non-concede-national-et-departemental-localisation-ponctuel/. (Accessed 6 March 2022).
Dijkema, M.B., Gehring, U., van Strien, R.T., van der Zee, S.C., Fischer, P., Hoek, G., Brunekreef, B., 2011. A comparison of different approaches to estimate small-scale spatial variation in outdoor NO2 concentrations. Environ. Health Perspect. 119, 670–675. https://doi.org/10.1289/EHP.0901818/ASSET/6D416C21-7986-4C99-B608-F953BE278AB2/ASSETS/GRAPHIC/EHP-119-670F2.JPG.
Eeftens, M., Beelen, R., De Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dedele, A., Dons, E., De Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičiene, 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. Von, Wolf, K., Brunekreef, B., Hoek, G., 2012. Development of land use regression models for PM2.5, PM 2.5 absorbance, PM10 and PMcoarse in 20 European study areas; Results of the ESCAPE project. Environ. Sci. Technol. 46, 11195–11205. https://doi.org/10.1021/es301948k.
Ermagun, A., Levinson, D.M., 2019. Development and application of the network weight matrix to predict traffic flow for congested and uncongested conditions. Environ. Plan. B Urban Anal. City Sci. 46, 1684–1705. https://doi.org/10.1177/2399808318763368.
EUROSTAT, 2015. Geostat 2011 grid dataset [WWW Document]. URL. http://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography.
Frick, R., Grimm, B., 2014. Long-distance Mobility - Current Trends and Future Perspectives.
Fu, M., Kelly, J.A., Clinch, J.P., 2017. Estimating annual average daily traffic and transport emissions for a national road network: a bottom-up methodology for both nationally-aggregated and spatially-disaggregated results. J. Transport Geogr. 58, 186–195. https://doi.org/10.1016/j.jtrangeo.2016.12.002.
Gonzales, M., Myers, O., Smith, L., Olvera, H.A., Mukerjee, S., Li, W.W., Pingitore, N., Amaya, M., Burchiel, S., Berwick, M., 2012. Evaluation of land use regression models for NO2 in El Paso, Texas, USA. Sci. Total Environ. 432, 135–142. https://doi.org/10.1016/j.scitotenv.2012.05.062.
Graser, A., Straub, M., Dragaschnig, M., 2014. Towards an open source analysis toolbox for street network comparison: indicators, tools and results of a comparison of OSM and the official Austrian reference graph. Trans. GIS 18, 510–526. https://doi.org/10.1111/tgis.12061.
He, Q., Huang, B., 2018. Satellite-based high-resolution PM2.5 estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model. Environ. Pollut. 236, 1027–1037. https://doi.org/10.1016/j.envpol.2018.01.053.
Holz-Rau, C., Scheiner, J., Sicks, K., 2014. Travel distances in daily travel and long-distance travel: what role is played by urban form? Environ. Plann. 46, 488–507. https://doi.org/10.1068/a4640.
Ishwaran, H., 2015. The effect of splitting on random forests. Mach. Learn. 99, 75–118. https://doi.org/10.1007/s10994-014-5451-2.
Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E., 2008. Hole-filled SRTM for the globe version 4, available from the CGIAR-CSI SRTM 90m database [WWW Document] CGIAR CSI Consort. Spat. Inf. URL https://srtm.csi.cgiar.org/.
Jones, R.R., Hoek, G., Fisher, J.A., Hasheminassab, S., Wang, D., Ward, M.H., Sioutas, C., Vermeulen, R., Silverman, D.T., 2020. Land use regression models for ultrafine particles, fine particles, and black carbon in Southern California. Sci. Total Environ. 699 https://doi.org/10.1016/j.scitotenv.2019.134234.
Kerckhoffs, J., Hoek, G., Portengen, L., Brunekreef, B., Vermeulen, R.C.H., 2019. Performance of prediction algorithms for modeling outdoor air pollution spatial surfaces. Environ. Sci. Technol. 53, 1413–1421. https://doi.org/10.1021/acs.est.8b06038.
Khan, J., Ketzel, M., Jensen, S.S., Gulliver, J., Thysell, E., Hertel, O., 2021. Comparison of road traffic noise prediction models: CNOSSOS-EU, Nord2000 and TRANEX. Environ. Pollut. 270, 138577 https://doi.org/10.1016/j.envpol.2020.116240.
Lu, M., Soenario, I., Helbich, M., Schmitz, O., Hoek, G., van der Molen, M., Karssenberg, D., 2020. Land use regression models revealing spatiotemporal co-variation in NO2, NO, and O3 in The Netherlands. Atmos. Environ. 223, 117238 https://doi.org/10.1016/j.atmosenv.2019.117238.
Ministry of Infrastructure and Transport, 2022. Average daily traffic - ANAS road and motorway network - dataset - open data - Ministry of infrastructure and transport [WWW Document]. URL. https://dati.mit.gov.it/catalog/dataset/traffico-giornaliero-medio-anas. (Accessed 6 September 2022).
Morley, D.W., De Hoogh, K., Fecht, D., Fabbri, F., Bell, M., Goodman, P.S., Elliott, P., Hodgson, S., Hansell, A., Gulliver, J., 2015. International scale implementation of the CNOSSOS-EU road traffic noise prediction model for epidemiological studies. Environ. Pollut. 206, 332–341. https://doi.org/10.1016/j.envpol.2015.07.031.
Morley, D.W., Gulliver, J., 2016. Methods to improve traffic flow and noise exposure estimation on minor roads. Environ. Pollut. 216, 746–754. https://doi.org/10.1016/j.envpol.2016.06.042.
Naughton, O., Donnelly, A., Nolan, P., Pilla, F., Misstear, B.D., Broderick, B., 2018. A land use regression model for explaining spatial variation in air pollution levels using a wind sector based approach. Sci. Total Environ. 630, 1324–1334. https://doi.org/10.1016/j.scitotenv.2018.02.317.
NDW, 2023. Verkeersintensiteiten voor verkeersveiligheidsanalyses | Verkeersveiligheid. Nat. Dataport. Wegverkeer [WWW Document]. URL. https://www.ndw.nu/onderwerpen/verkeersveiligheid/intensiteitsgegevens-voor-verkeersveiligheid. (Accessed 4 March 2024).
NDW, 2021. Nationaal dataportaal wegverkeer | nationaal dataportaal wegverkeer [WWW Document]. URL. https://www.ndw.nu/. (Accessed 11 August 2023).
Neis, P., Zielstra, D., Zipf, A., 2011. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. Future Internet 4, 1–21. https://doi.org/10.3390/fi4010001.
OpenStreetMap contributors, 2017. OpenStreetMap [WWW Document] OpenStreetMap. URL. https://www.openstreetmap.org.
Pulugurtha, S., Kusam, P., 2012. Modeling annual average daily traffic with integrated spatial data from multiple network buffer bandwidths. Transport. Res. Rec. 2291, 53–60. https://doi.org/10.3141/2291-07.
Pun, L., Zhao, P., Liu, X., 2019. A multiple regression approach for traffic flow estimation. IEEE Access 7, 35998–36009. https://doi.org/10.1109/ACCESS.2019.2904645.
Selby, B., Kockelman, K.M., 2013. Spatial prediction of traffic levels in unmeasured locations: applications of universal kriging and geographically weighted regression. J. Transport Geogr. 29, 24–32. https://doi.org/10.1016/j.jtrangeo.2012.12.009.
Sfyridis, A., Agnolucci, P., 2020. Annual average daily traffic estimation in England and Wales: an application of clustering and regression modelling. J. Transport Geogr. 83, 102658 https://doi.org/10.1016/j.jtrangeo.2020.102658.
Sheffi, Y., Powell, W., 1981. A comparison of stochastic and deterministic traffic assignment over congested networks. Transport. Res. Part B 15, 53–64. https://doi.org/10.1016/0191-2615(81)90046-1.
Shen, Y., de Hoogh, K., Schmitz, O., Clinton, N., Tuxen-Bettman, K., Brandt, J., Christensen, J.H., Frohn, L.M., Geels, C., Karssenberg, D., Vermeulen, R., Hoek, G., 2022. Europe-wide air pollution modeling from 2000 to 2019 using geographically weighted regression. Environ. Int. 168, 107485 https://doi.org/10.1016/j.envint.2022.107485.
Shi, T., Dirienzo, N., Requia, W.J., Hatzopoulou, M., Adams, M.D., 2020. Neighbourhood scale nitrogen dioxide land use regression modelling with regression kriging in an urban transportation corridor. Atmos. Environ. 223, 117218 https://doi.org/10.1016/J.ATMOSENV.2019.117218.
Van den Bossche, J., De Baets, B., Verwaeren, J., Botteldooren, D., Theunis, J., 2018. Development and evaluation of land use regression models for black carbon based on bicycle and pedestrian measurements in the urban environment. Environ. Model. Software 99, 58–69. https://doi.org/10.1016/j.envsoft.2017.09.019.
Vienneau, D., de Hoogh, K., Briggs, D., 2009. A GIS-based method for modelling air pollution exposures across Europe. Sci. Total Environ. 408, 255–266. https://doi.org/10.1016/j.scitotenv.2009.09.048.
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., 2014. Performance of multi-city land use regression models for nitrogen dioxide and fine particles. Environ. Health Perspect. 122, 843–849. https://doi.org/10.1289/ehp.1307271.
Weissert, L., Alberti, K., Miles, E., Miskell, G., Feenstra, B., Henshaw, G.S., Papapostolou, V., Patel, H., Polidori, A., Salmond, J.A., Williams, D.E., 2020. Low-cost sensor networks and land-use regression: interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California. Atmos. Environ. 223, 117287 https://doi.org/10.1016/J.ATMOSENV.2020.117287.
Wen, Y., Zhang, S., Wang, Y., Yang, J., He, L., Wu, Y., Hao, J., 2023. Dynamic traffic data in machine-learning air quality mapping improves environmental justice assessment. Environ. Sci. Technol. 58, 3128. https://doi.org/10.1021/acs.est.3c07545.
Wolf, K., Cyrys, J., Harciníková, T., Gu, J., Kusch, T., Hampel, R., Schneider, A., Peters, A., 2017. Land use regression modeling of ultrafine particles, ozone, nitrogen oxides and markers of particulate matter pollution in Augsburg, Germany. Sci. Total Environ. 579, 1531–1540. https://doi.org/10.1016/j.scitotenv.2016.11.160.
Wright, M.N., Ziegler, A., 2017. {ranger}: a fast implementation of random forests for high dimensional data in {C++} and {R. J. Stat. Software 77, 1–17. https://doi.org/10.18637/jss.v077.i01.
Yang, Z., Freni-Sterrantino, A., Fuller, G.W., Gulliver, J., 2020. Development and transferability of ultrafine particle land use regression models in London. Sci. Total Environ. 740, 140059 https://doi.org/10.1016/j.scitotenv.2020.140059.
Yi, Z., Liu, X.C., Markovic, N., Phillips, J., 2021. Inferencing hourly traffic volume using data-driven machine learning and graph theory. Comput. Environ. Urban Syst. 85, 101548 https://doi.org/10.1016/J.COMPENVURBSYS.2020.101548.
Zhao, F., Park, N., 2004. Using geographically weighted regression models to estimate annual average daily traffic. In: Transportation Research Record. National Research Council, pp. 99–107. https://doi.org/10.3141/1879-12.