Environmental Research 228 (2023) 115836

Contents lists available at ScienceDirect

Environmental Research

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

Integrating large-scale stationary and local mobile measurements to estimate hyperlocal long-term air pollution using transfer learning methods

Zhendong Yuan[a,*], Jules Kerckhoffs[a], Youchen Shen[a], Kees de Hoogh[b,c], Gerard Hoek[a], Roel Vermeulen[a,d]

[a] Institute for Risk Assessment Sciences, Utrecht University, 3584 CK, Utrecht, Netherlands
[b] Swiss Tropical and Public Health Institute, Kreuzstrasse 2, 4123, Allschwil, Switzerland
[c] University of Basel, Petersplatz 1, Postfach, 4001, Basel, Switzerland
[d] Julius Centre for Health Sciences and Primary Care, University Medical Centre, University of Utrecht, 3584 CK, Utrecht, the Netherlands

ARTICLE INFO

Handling Editor: Jose L Domingo

Keywords:
Air pollution mapping
Mobile monitoring
Google street view
Transfer learning

ABSTRACT

Mobile air quality measurements are collected typically for several seconds per road segment and in specific timeslots (e.g., working hours). These short-term and on-road characteristics of mobile measurements become the ubiquitous shortcomings of applying land use regression (LUR) models to estimate long-term concentrations at residential addresses. This issue was previously found to be mitigated by transferring LUR models to the long-term residential domain using routine long-term measurements in the studied region as the transfer target (local scale). However, long-term measurements are generally sparse in individual cities. For this scenario, we propose an alternative by taking long-term measurements collected over a larger geographical area (global scale) as the transfer target and local mobile measurements as the source (Global2Local model). We empirically tested national, airshed countries (i.e., national plus neighboring countries) and Europe as the global scale in developing Global2Local models to map nitrogen dioxide (NO_2) concentrations in Amsterdam. The airshed countries scale provided the lowest absolute errors, and the Europe-wide scale had the highest R^2. Compared to a “global” LUR model (trained exclusively with European-wide long-term measurements), and a local mobile LUR model (using mobile data from Amsterdam only), the Global2Local model significantly reduced the absolute error of the local mobile LUR model (root-mean-square error, 6.9 vs 12.6 μg/m^3) and improved the percentage explained variances compared to the global model (R^2, 0.43 vs 0.28, assessed by independent long-term NO_2 measurements in Amsterdam, n = 90). The Global2Local method improves the generalizability of mobile measurements in mapping long-term residential concentrations with a fine spatial resolution, which is preferred in environmental epidemiological studies.

1. Introduction

Previous studies have shown that mobile monitoring campaigns, using vehicles equipped with high frequency monitors, are capable of capturing air pollution distribution patterns at a fine spatial granularity (Apte et al., 2017; Messier et al., 2018). However, applying land use regression models (LUR) developed on mobile measurements (referred to as the mobile LUR model) is not optimal to estimate outdoor long-term (e.g., annual) air pollution concentrations at residential addresses, due to the fundamental differences between these two domains (Messier et al., 2018; Chambliss et al., 2020). Mobile monitoring vehicles often measure air pollution on roads repeatedly for a few seconds and are often conducted during specific timeslots (e.g., working hours of weekdays). These differences in space (on-road vs residential), time (short-term vs long-term; working hours vs full day) and potentially instruments (different monitors) challenge the mobile data to optimally predict long-term residential air pollution concentrations (the knowledge gap). (Yuan et al., 2022).

In a previous paper, we demonstrated the possibilities of reducing the knowledge gap using transfer-learning LUR models by transferring the mobile knowledge into the local long-term domain approximated by long-term measurements (covering the entire period) at random subsets of 82 sites in the studied region (local scale) (Yuan et al., 2022). However, this method is not applicable in regions where no or only a few

[*] Corresponding author. Yalelaan 2, 3584 CM, Utrecht, Netherlands.
E-mail address: z.yuan@uu.nl (Z. Yuan).

https://doi.org/10.1016/j.envres.2023.115836
Received 19 January 2023; Received in revised form 31 March 2023; Accepted 1 April 2023
Available online 6 April 2023

local long-term measurements are available. This is a typical setting for many cities, as the number of stationary monitoring sites in a city is often small. Instead of local long-term measurements, our hypothesis, investigated in this paper, is using long-term measurements from a larger geographical area such as the entire country or even the whole continent as the transfer targe. We refer to this large area as the global scale, distinguishing it from the study area of interest (typically a city, local scale). However, it is uncertain whether using global measurements as the transfer target, will still improve the mobile LUR models in estimating local long-term air pollution concentrations as the concentration pattern captured in larger geographical areas may not accurately inform the pattern at local scale.

Another approach that has been broadly used to map air pollution with fine spatial resolution in cities are large-scale LUR models based upon long-term exposure measurements over a large geographic area, without specific local knowledge (Lu et al., 2020; de Hoogh et al., 2018; Shen et al., 2022). Targeted at fitting global measurements, global LUR models are optimized to capture more inter-urban variance and average the heterogeneity of local intra-urban variance (Lu et al., 2020; Meyer and Pebesma, 2022). In the ELAPSE project (Effects of Low-Level Air Pollution: A Study in Europe), the overall cross-validated R^2 of nitrogen dioxide (NO_2) at the European scale was about 0.49 achieved by the linear LUR model. The external validation R^2 varied however between 0.07 and 0.76 in different, more local, subregions (de Hoogh et al., 2018). In Lu et al., 2020, the world-wide daytime model of NO_2 achieved a cross-validated R^2 of 0.7 with a Random Forest (RF) LUR model. However, validated in the selected countries with external measurements, the local R^2 varied from 0.56 to 0.73 (Lu et al., 2020). Although the decrease in R^2 could also be due to the reduced variance at the local scale as compared to the global scale, it does indicate that by exclusively using global measurements, the global LUR model might be limited in estimating the local air pollution patterns.

In this paper, we explored the possibility of integrating global long-term measurements to supplement local mobile measurements for mapping hyperlocal long-term air pollution concentrations in a city. We chose the city of Amsterdam as our study region due to the availability of (i) local mobile monitoring NO_2 data; (ii) long-term measurements from the European monitoring network acting as the global data and (iii) sufficient number of local independent long-term measurements as external validation data (Palmes tubes, passive monitors, n = 90). We tested three larger geographic areas (i.e., national, airshed countries, and continental Europe) to identify the global scale in the Global2Local model that can estimate the most accurate NO_2 map for Amsterdam. The performance of Global2Local models were compared with LUR models trained exclusively using global long-term monitoring data or local mobile data. We then discussed how global measurements influence the local mobile LUR model.

2. Data and models

The training data for all NO_2 LUR models used in the analyses can be found in Table 1 and further explained in the next subsections. For the global model, we used the data and methodology from Shen et al., 2022, where a global LUR model was developed based on long-term measurements across Europe. For the local model, we used the Google mobile monitoring data in Amsterdam from Kerckhoffs et al., 2022. The transfer-learning models (Global2Local models) were developed using both local mobile measurements in Amsterdam and global long-term stationary measurements from three larger geographic areas (national, airshed countries (i.e., national plus neighboring countries), and continental EU). The selection of neighboring countries follows the concept of airshed which is loosely defined as part of atmosphere sharing similar emission and dispersion patterns (Anderson et al., 2013). Belgium, Luxembourg and Germany were selected as the neighboring countries. Paired with the NO_2 measurements, global, local (excluding Mobile_data_only) and Global2Local models shared the same categories and calculation scheme of predictor variables (details in section 2.1). All models are based upon measurements performed predominantly in 2019.

2.1. Global model

The global model was implemented as a RF-based LUR model trained with European AirBase long-term measurements and land use covariates (referred to as EU_LUR) based on 25m*25m cells, following the same methodology and data as in Shen et al., 2022. Random forest is a bagging-based assemble learning algorithm which has been broadly used in air pollution modelling (Lu et al., 2020; de Hoogh et al., 2018; Shen et al., 2022). The European Environment Agency (EEA) AirBase routine monitoring data of NO_2 measurements across Europe in 2019

Table 1
Summary of the models compared and NO_2 monitoring data.

Model category	Training input	Model name	Algorithm
	Conventional models developed on a single scale (local or global)		
Global model[a]	European Airbase long-term data (n = 3,243 sites)	EU_LUR	Random Forest
Local model[a]	Amsterdam mobile data (n = 142,950, averaged to 25m * 25m cells)	Mobile_data_only	N.A. (average of measurements)
		AMS_LUR	Random Forest
	Transfer learning models		
Global2Local[a]	Amsterdam mobile monitoring data & European Airbase long-term data (n = 3,243 sites)	EU2AMS	TrAdaBoost
	Amsterdam mobile data & Airbase long-term data within the Netherlands and its neighboring counties-Belgium, Luxemburg and Germany (n = 787 sites)	NLBELUXDE2AMS	TrAdaBoost
	Amsterdam mobile data & Airbase long-term data within the Netherlands (n = 68 sites)	NL2AMS	TrAdaBoost

[a] The local scale refers to Amsterdam. The global scale here is an abstract concept referring to any geographic area that is spatially larger than the local scale.

Table 2
Overview of data locations.

Locations	Number of AirBase measurements			Number of external validations	Number of mobile measurements
	Europe	*NLBELUXDE*	*Netherlands*	*Palmes*	*Mobile*
Major road	2,650	589	10	26	116,816
Urban background	593	198	58	64	26,134

were averaged per site over the year (continuous monitoring in the full year, 24 h per day) (Air Quality e-Report, 2021). The monitoring locations at background and traffic locations are all considered relevant for assessing exposure at residential addresses on minor and major streets respectively. EEA AirBase data are measured by the national regular routine monitors from EEA-32 member countries. The NO_2 measurements are harmonized following the standard procedures developed over decades of regulatory monitoring (Guerreiro et al., 2014; Guerreiro, 2013). 3,243 sites, where at least 75% of the entire year was measured, were included in the modeling (in total of 3,253 sites downloaded from the EEA data port and 0.33% were dropped). The covariates of the global model were calculated based on 25m*25m cells centering at the geographic location of each AirBase monitoring site in different buffer sizes. They include the following six groups of predictors: 1) Land Use features, such as the areal ratios of CORINE land use classes (e.g., residential, urban green) (CORINE Land Cover, 2021), Copernicus impervious surfaces (Imperviousness, 2021), SRTM altitude information (SRTM 90m Digital Elevation Database, 2021); 2) road features from OpenStreetMap (OpenStreetMap Wiki, 2023) such as the length of major roads and residential roads; 3) population density from Eurostat (GEOSTAT - GISCO - Eurostat, 2021); 4) meteorological features from ECMWF ERA5 (Copernicus Climate Change Service, 2019), such as land temperature, wind speed and precipitation; 5) satellite-derived and deterministic modelling features, including the Ozone Monitoring Instrument (OMI) from Tropospheric (Boersma et al., 2011; Tropospheric Emission Monitoring Internet Service, 2022), Chemical Transport Model (CTM) from Danish Eulerian Hemispheric Model (DEHM_v31102016) (Brandt et al., 2012) and MACC-II ENSEMBLE Model (Marécal et al., 2015); 6) climate zone features from European bio-geographical regions 2018 (v1.0)[24]. The details of the covariates used are explained in Appendix Table S1 and Shen et al., 2022 (Shen et al., 2022).

2.2. Local model

We examined two local models, namely a "model" based on the mobile measurements only (Mobile_data_only) and a conventional mobile LUR model (referred to as AMS_LUR). The mobile monitoring data were obtained from the Air View campaign, in which two Google Street View cars continuously measured NO_2 at a frequency of 1 Hz in Amsterdam from May 25, 2019, to March 15, 2020 (stopped due to the COVID19 lock down policy). Most measurements were collected on weekdays between 08:00 and 22:00 (more details in Kerckhoffs et al., 2022).

Our previous mobile modelling work snapped the mobile measurements into 50m road segments as the basic spatial unit and made predictions based on them (Kerckhoffs et al., 2022, Yuan et al., 2022). Differently, in this work, the spatial unit of the local model is aligned with that of the global model. We divided Amsterdam into 25m*25m grid cells (n = 574,299; the same size and extent as the global grid). The mobile measurement points were aggregated to the spatially overlapping cells. A total of 142,950 cells were measured by Google Street View cars (25% of all 25m*25m cells in Amsterdam). The mean of measured mobile NO_2 was used as the mobile measurement of the corresponding cell, forming the Mobile_data_only model. These mobile measurements were then used to train a RF-based LUR model (AMS_LUR) following the implementation of the RF_LUR model in Yuan et al., 2022. But distinguished from Yuan et al., 2022, the AMS_LUR model did not contain the traffic intensity and other local predictors as they were not available at the global scale. Instead, the predictor variables used in the AMS_LUR model shared the same categories and calculation schema as the global variables in the aforementioned global model. In addition, to test the sensitivity of model performance with and without traffic intensity predictors, we trained a local mobile LUR model following the implementation of AMS_LUR, but with the addition of local traffic variable from Yuan et al., 2022.

2.3. Global2Local model

The Global2Local model was trained by directly merging the global long-term measurements from stationary routine monitors (AirBase) and local mobile measurements collected by the Google Street View cars in Amsterdam as the training input. The Amsterdam mobile data are treated as the transfer source and the AirBase data as the transfer target data. A weighted loss function was optimized by the instance-based transfer learning algorithm - TrAdaBoost (Kouw and Loog, 2021; Dai et al., 2007). By setting larger weights to the transfer target instances (i.e., global long-term AirBase measurements), TrAdaBoost can blend learning towards the desired long-term knowledge making the predictions more similar to the long-term measurements. The global scale can be any region geographically larger than the local scale. Selecting measurements from a larger geographic region brings more transfer target data but at a risk of more data deviating from the local data. Therefore, an empirical test is needed to identify which scale can balance the trade-offs to map the most accurate local air pollution (by the highest performance). We tested following three Global2Local models with subsequently increasing the geographical area: 1) NL2AMS, using the Dutch subset of AirBase measurements as the transfer target data (68 stationary long-term sites, locations provided in the map in Fig. S3.); 2) NLBELUXDE2AMS, considering neighboring countries within the same airshed as the global scale (includes AirBase measurements from the Netherlands, Belgium, Luxemburg and Germany, 787 sites); 3) EU2AMS, setting continental Europe as the global scale and using the full set of AirBase measurements across Europe (3,243 sites). All three Global2Local models shared the same categories of predictor variables as the global and local model.

TrAdaBoost inherits the key idea of the classic AdaBoost algorithm (Freund and Schapire, 1997) which is a boosting-based ensemble learning algorithm where multiple base models ("weak learners") are trained sequentially in an adaptive way: each base learner in the sequence is fitted by giving more weights to instances in the training dataset that caused higher errors by the previous learner in the sequence. Intuitively, each base learner focuses on fitting the most difficult instances in each boosting iteration. At the end, these weak learners are combined to form a "strong" model that is accurate at predicting all the cases learned from the training instances.

As an adapted version of TrAdaBoost, Two_stage_TrAdaBoost was implemented in the Global2Local model (Pardoe and Stone, 2010). It has been shown to outperform other transfer learning algorithms tested in our earlier paper of modelling long-term NO_2 concentrations using mobile measurements (Yuan et al., 2022). It works by directly merging the local mobile monitoring instances (as source instances – (X_s, Y_s)) with the global long-term instances (as target instances – (X_t, Y_t)) to form a single training dataset and assign the same initial weights to each instance. Then, these weights update following the design of AdaBoost and take place in two stages. In the first stage, during each boosting step, TrAdaBoost decreases the relative weights of source instances that are different from the target instance during each boosting iteration. In the second stage, the weights of all instances are frozen while TrAdaBoost increases the weights of target instances that are different from source instances (Pardoe and Stone, 2010). This approach emphasizes the target instances and keeps also characteristics of the source instances.

2.4. External validation

We fine-tuned the global, local and Global2Local models by the full-gridded hyperparameter searching and selected the best combination. The performance of all models was validated by external long-term measurements collected by Palmes tubes in Amsterdam. Palmes tubes are passive samplers used in the routine monitoring network that measure NO_2 on street lanterns and building facades in Amsterdam (spatial locations in Fig. 4A), which differ from the on-road mobile measurement. They are deployed and calibrated following the standard procedure by Dutch national institute for public health and the environment (RIVM) (Dijkema et al., 2011). 90 Palmes monitoring sites covering the complete period from May 2019 to March 2020 (repeated four-weeks sampling covering the full year, 24 h per day) were assigned to the overlapping 25m*25m cells (the same grid as used to calculate the covariates at the local and global scales). The sites include both major traffic ($n = 26$) and urban background locations ($n = 64$), covering Amsterdam and its surroundings. Together with the corresponding predictors, Palmes measurements represent the targeted local long-term residential knowledge.

The squared Pearson correlation (R^2) mean absolute error (MAE), and root mean square error (RMSE) were used to assess model performance. Accuracy metrics reflect accuracy in different aspects. R^2 reflects the percentage of total variation in measurements that can be explained. MAE and RMSE reflect the level of absolute errors. In RMSE, the errors are squared before they are averaged giving a relatively high weight to extreme errors as compared to MAE. The choice of which accuracy metric to use largely depends on the applications. The primary objective of an exposure model for application in an epidemiological study is to rank exposures of subjects into low, medium and high categories. Hence, the R^2 is a very important metric. The RMSE and MAE are also important metrics, since we also want to assign health effects to specific concentration values. For other applications, for example, comparison with legal limits, assigning the correct absolute value may be most important.

In addition to the prediction accuracy, the training accuracy of the global and local models (Table 4) reflect the accuracy of the model in the training domain approximated by the local mobile or global stationary instances. Since accuracy metrics can only provide an overview of performance, to comprehensively understand the prediction uncertainty, the density and spatial distributions of model predictions were inspected.

To further quantify the spatial pattern of the model accuracy, model performance on major traffic locations and in urban background areas were separated. Major roads consisted of a series of OpenStreetMap tags, including motorway (a restricted access major divided highway, freeway, autobahn etc.), trunk (the most important roads in a country's traffic system that not necessarily be a divided highway) and primary (the second most important roads and often link larger towns) (OpenStreetMap Wiki, 2023). A major traffic location was defined as a cell that is within 50m of a major road (major_roads_50 > 0, $n = 26$). The remaining cells were considered as urban background locations ($n = 64$).

3. Results

3.1. Differences between global and local measurements

Ranges of global, local and validation NO_2 measurements are summarized in Table 3. Aggregated to 25m*25m grid cells, mobile measurements consisted of on average 36 s on 6 different days per cell (see the histogram in Appendix Fig. S1. A, B).

The density distributions of the global Airbase long-term and local Google mobile data differed from the local Palmes long-term measurements (Fig. 1). The variability in the local mobile and global long-term measurements in larger geographic areas was wider than the external validation data, indicative of the knowledge gap problem.

In addition, in this work, we set the mean of the GPS measurements within a cell as the mobile measurement. But the mean based mobile

Table 3
An overview of averaged NO_2 measurements contributing to global, local and validation datasets.

Data	Scale	Time frame	Number of measured cells (25m * 25m)	1st quantile ($\mu g/m^3$)	Median ($\mu g/m^3$)	3rd quantile ($\mu g/m^3$)
Global data EU Airbase data	Europe	2019, 24 h per day	3,243	12.0	18.9	27.7
Global data NLBELUXDE Airbase data	The Netherlands and its neighboring countries - Belgium, Luxemburg and Germany	2019, 24 h per day	787	16.1	24.8	34.8
Global data NL Airbase data	The Netherlands	2019, 24 h per day	68	15.4	21.7	26.7
Local mobile data	Amsterdam	March 2019 to May 2020, diurnal hours	142,950	16.2	23.7	34.3
Local external long-term validation data	Amsterdam	March 2019 to May 2020, 24 h per day	90	21.2	25.9	31.3

The local scale refers to Amsterdam. The global data EU, NLBELUXDE and NL stand for the stationary AirBase data in different regions. They are defined as the global scale in different Global2Local models.

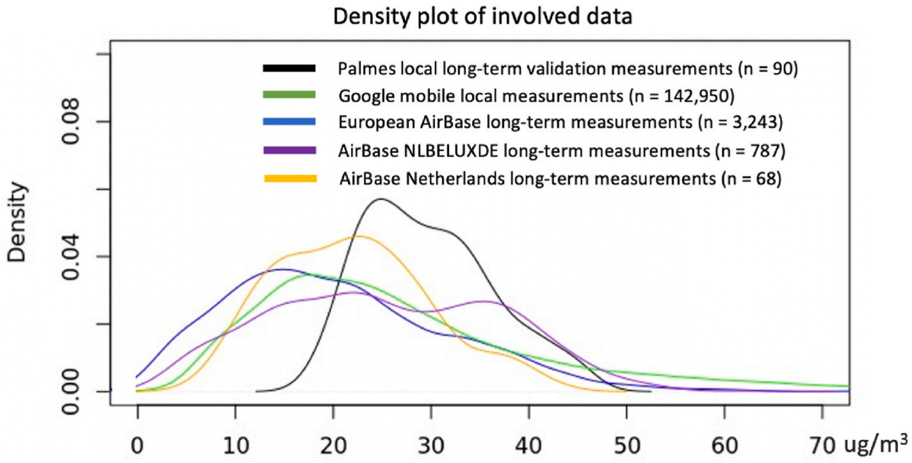


Fig. 1. Density plot of NO_2 from global long-term (European, NLBELUXDE and Dutch AirBase data), local mobile (AirView mobile data in Amsterdam) and local external validation data (Palmes data in Amsterdam).

measurements can be affected by "high-polluting vehicles" during individual mobile monitoring drives and the median value is more outlier-resistant. We calculated the median of the mobile measurements and compared it with the mean. We found typically small differences between mean and median and a very high correlation (Pearson correlation = 0.95, density and scatter plot in Appendix Fig. S2). There are two reasons for this. First, extreme outliers in mobile data have been removed in our preprocessing (NO_2 measurements larger than 500 μg/m^3). Second, the median is less representative, as each cell consisted of only a small number of observations: the median number of GPS points is 11 and the number of drive-pass per cell is 5 (Appendix Fig. S1). We preferred the mean, as air quality guidelines are all represented as annual means, to include peak concentrations in the exposure metrics associated with health effects.

3.2. Overall performance

The variable importance of EU_LUR, AMS_LUR and NLBELUXDE2AMS were recorded in Appendix Fig. S4. The variables related to major roads were important for all models. The R^2 of EU_LUR in the training accuracy was the same as the cross-validated performance of the rf00-19 model (a RF LUR model in 2019) reported in Shen et al., 2022.

In terms of predicting long-term air pollution concentrations in Amsterdam, the AMS_LUR model had a better accuracy compared to Mobile_data_only (Table 4). In the additional sensitivity test, with the traffic predictor variables, the performance of AMS_LUR increased (R^2 0.56; MAE 6.7 μg/m^3; RMSE 8.5 μg/m^3). Comparing global and local models, with the same algorithm and categories of covariates (Appendix Table S1), AMS_LUR had a better R^2 but EU_LUR achieved lower MAE and RMSE of prediction accuracy. The transfer-learning-based Global2Local models had the best combination of R^2, RMSE, MAE compared to the local and global models.

Comparing the three Global2Local models, NLBELUXDE2AMS achieved the lowest absolute error but its R^2 was marginally less than EU2AMS. The RMSE and MAE of NLBELUXDE2AMS were lower than that of AMS_LUR and EU_LUR. The R^2 of NLBELUXDE2AMS was 0.15 higher than the EU_LUR model.

The density and scatter plots of the predictions at validation sites demonstrate that the NLBELUXDE2AMS model fits the long-term validation measurements better than EU_LUR and AMS_LUR (Fig. 2). Predictions of the AMS_LUR model covered a wider range than the external validation data, while EU_LUR predictions were concentrated between 30 and 40 μg/m^3: a narrower range compared to Palmes ground-truth measurements.

3.3. Spatial distribution of predictions and variable importance

Trained exclusively with on-road mobile measurements in Amsterdam, all top-10 important variables of the AMS_LUR model are related to road length in various buffer sizes (from 50m to 1000m) and a large total built-up area buffer (6000m, Appendix Fig. S4). This resulted in the highest NO_2 predictions of AMS_LUR mainly distributing along roads (Fig. 3A). NO_2 predictions from EU_LUR are highest in the city center and major roads, with values gradually decreasing to the suburbs (Fig. 3B). EU_LUR relies heavily on impervious surface areas in buffers between 1800m and 6000m, Chemical Transport Model (the MACC-II CTM model) and the small-scaled road length (i.e., major_roads_50 and allRoads_100). Within Amsterdam the MACC-II CTM (10*10 km scale) does not result in sizable differences in concentrations. The impervious surface variables in large buffer sizes explained gradual changes over space. The spatial distribution of NLBELUXDE2AMS predictions shows the combined properties of AMS_LUR and EU_LUR. The highest concentrations were predicted at major-traffic locations. The city center is more polluted as compared to the suburbs (Fig. 3C). The variable importance plot of NLBELUXDE2AMS suggests that both small-scaled traffic and large-scaled environmental context features jointly dominate the predictions, such as port, nature, industrial, residential

Table 4
Performance of the model tested.

Model category	Model name	Training accuracy[c] (estimated by 10-folds cross validations)			Prediction accuracy (evaluated by local external long-term measurements; GGD Palmes; n = 90)		
		R^2	MAE (μg/m^3)	RMSE (μg/m^3)	R^2	MAE (μg/m^3)	RMSE (μg/m^3)
Local model[a]	Mobile_data_only	–	–	–	0.29[b]	11.3[b]	15.1[b]
	AMS_LUR	0.55	7.3	12.3	0.44	8.8	12.6
Global model[a]	EU_LUR	0.67	4.4	6.1	0.28	6.4	7.6
Global2Local[a]	EU2AMS	–	–	–	0.45	5.7	7.3
	NLBELUXDE2AMS	–	–	–	0.43	5.5	6.9
	NL2AMS	–	–	–	0.41	6.1	7.5

[a] The local model refers to models trained using local (Amsterdam) mobile information only. The Global model stands for models trained with global (Europe) stationary data. The Global2Local models were trained with both global and local information. Three different larger geographic areas were explored as the global scale such as Europe (EU2AMS), airshed countries (includes the Netherlands, Belgium, Luxemburg and Germany, NLBELUXDE2AMS) and the Netherlands (NL2AMS).

[b] The accuracy of Mobile_data_only was calculated based on the 79 cells where mobile measurements and Palmes data were available concurrently.

[c] The training accuracy for Global2Local models is not informative, as it is based on the weighted dataset that merges mobile and AirBase measurements. It is not informative to compare with models based solely on local mobile or global long-term measurements.

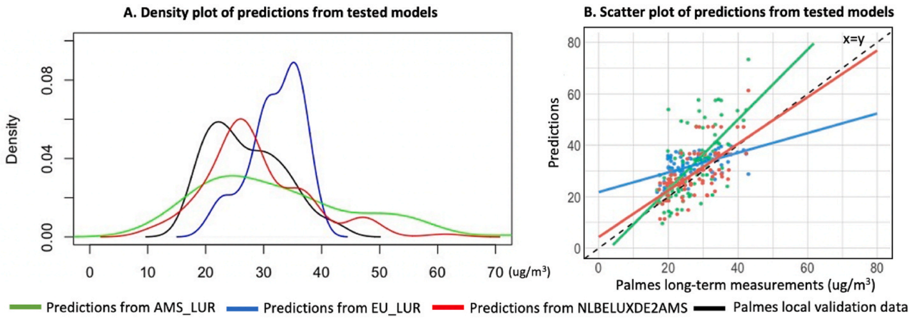
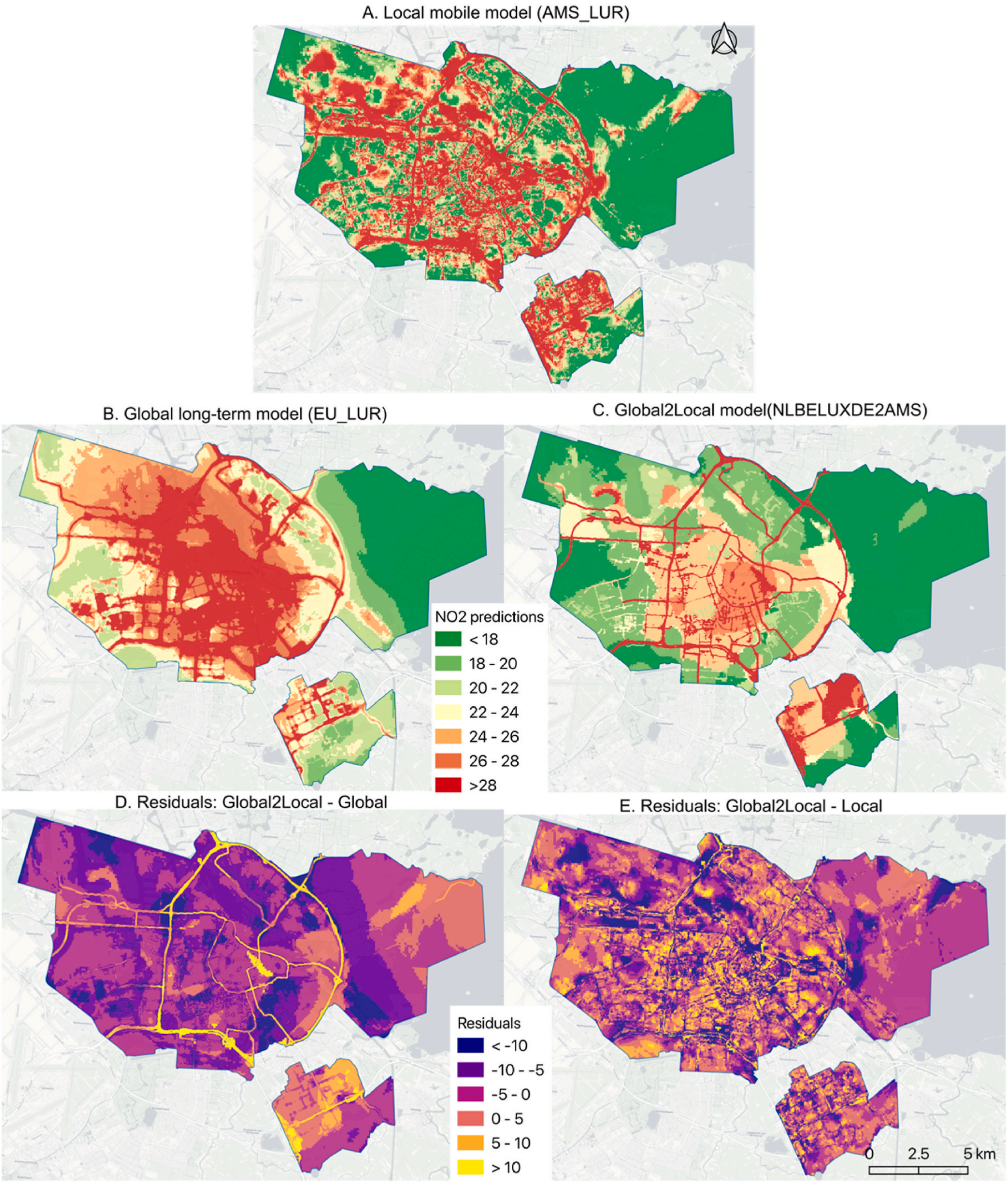


Fig. 2. Density and scatter plots of the NO_2 predictions made from the model tested against the local Palmes validation data at Palmes locations (n = 90).


A. Local mobile model (AMS_LUR)

B. Global long-term model (EU_LUR)

C. Global2Local model(NLBELUXDE2AMS)

D. Residuals: Global2Local - Global

E. Residuals: Global2Local - Local

Fig. 3. A-C. Spatial distributions of NO_2 predictions (μg/m^3) from the models tested. D,E. the residuals of NLBELUXDE2AMS against EU_LUR and AMS_LUR. The EU_LUR model was trained with European AirBase long-term measurements. The AMS_LUR model was trained with mobile AirView measurements in Amsterdam. The NLBELUXDE2AMS model was trained by combining the long-term and mobile measurements from the global (NLBELUXDE) and local scales.

and urban green areas.

AMS_LUR, EU_LUR and NLBELUXDE2AMS overestimated NO_2 at both major traffic and urban background locations. The averaged residuals were all positive (see Fig. 4C). AMS_LUR predictions resulted in significantly higher residuals at major traffic than urban background locations. EU_LUR predictions resulted in similar residuals at both locations. The residuals from NLBELUXDE2AMS were intermediate between EU_LUR and AMS_LUR at major traffic locations and achieved the lowest residuals at urban background locations.

4. Discussion

The knowledge gap between the training domain and the application domain potentially limits the generalizability of mobile LUR models to predict long-term air pollution concentrations at residential addresses. The presented work illustrates that the long-term concentration information extracted from a larger geographical area can narrow this knowledge gap. The selection of the global scale in the Global2Local method requires empirical tunning to find a balance between the similarity to the local scale and the number of long-term measurements at the global scale. The scale of airshed countries was identified as the most

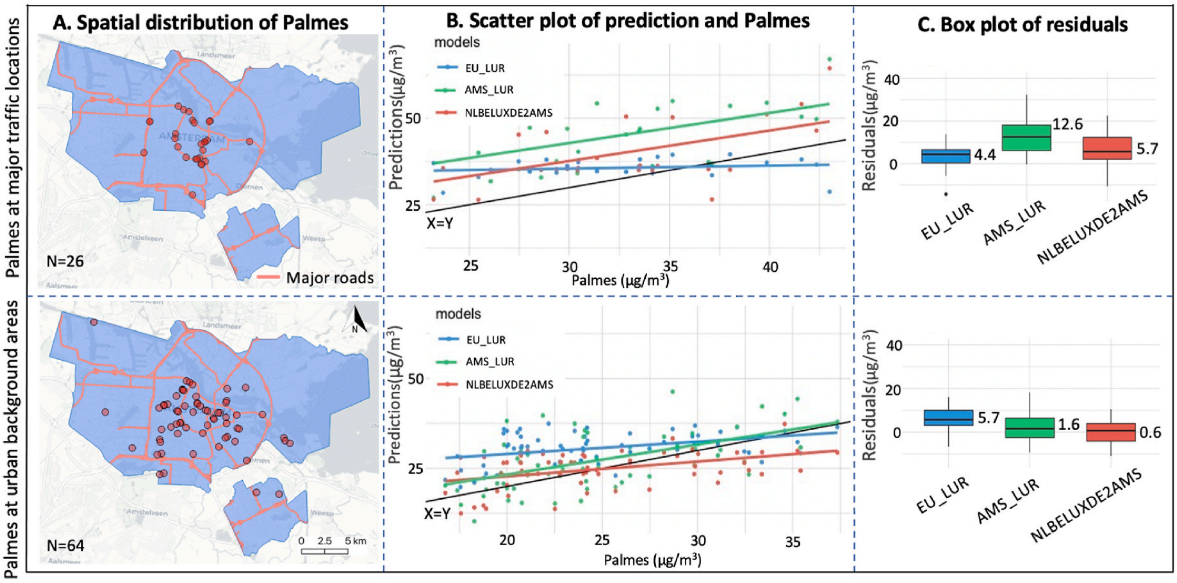


Fig. 4. A. The spatial distribution of Palmes validation sites at major traffic locations and in urban background areas. B. Scatter plot between Palmes data and the corresponding predictions (NO_2) from the tested models at Palmes locations. Black line is the 1:1 line. C. Boxplot of the residuals between model predictions and Palmes data at Palmes locations (calculated by Predictions - Palmes). Median values are marked.

appropriate global scale for the Global2Local model to estimate long-term concentrations in Amsterdam.

4.1. Global measurements can narrow the knowledge gap

In this work, the mobile data differed from the desired local long-term measurements in space (on-road vs residential), time (short-term vs long-term, working hours vs full day) and instruments (different sensors). Following our previous paper, these differences between training and prediction domains are identified as the knowledge gap (Yuan et al., 2022). It brings issues of mismatching between the training and prediction accuracy (especially validated by external long-term measurements), which have been broadly observed in many mobile monitoring studies (Chambliss et al., 2020; Kerckhoffs et al., 2019). Aligned with them, we found the knowledge gap in this work also limited the generalizability of the mobile LUR model (i.e., AMS_LUR) on estimating local long-term residential concentrations. Back to our previous paper, this knowledge gap has been demonstrated to be mitigated by transferring the mobile knowledge learned from mobile measurements into the long-term domain approximated by local long-term measurements (Yuan et al., 2022).

Instead of using long-term measurements from the studied local region, in this paper, we found that using global long-term measurements from a larger geographic area in the transfer-learning approach also can narrow the knowledge gap. Instead of optimizing the loss function in the mobile domain (broadly used in training mobile LUR models (Messier et al., 2018; Kerckhoffs et al., 2019; van de Beek et al., 2020) as well as AMS_LUR in this work), the loss function optimized by the Global2Local model was weighted to represent the pattern between the land-use variables and global long-term measurements (the combination of mobile and long-term data as training inputs). This way pushed the LUR model to learn more from global long-term instances while simultaneously preserving local knowledge from local mobile measurements. Trained solely with local mobile measurements, AMS_LUR captured fine-scaled spatial variations, leading to a high R^2 (Table 4). However, compared to the targeted long-term residential air pollution, the difference in space, time and instruments led to high absolute errors. The capability of explaining the local variations is inherited from the local mobile measurements, yielding the Global2Local model a high R^2 (a similar level to AMS_LUR).

The characteristics of the global measurements (e.g., spatially - near-road monitoring and temporally - 24h per day) complement short-term on-road mobile measurements, which reduced the discrepancy in absolute values between the predictions and local long-term concentrations. This explains the significantly lower absolute errors of the Global2Local models compared to the AMS_LUR model (Table 2). Inspecting the density curves in Fig. 2, the Global2Local model shows better goodness-of-fit among concentrations between 20 and 30 μg/m^3, where the median of Palmes validation measurements is located (Table 2.). More importantly, NO_2 at most urban background locations also falls within this range (Fig. 4B).

The improved R^2 of Global2Local models was significant compared to the long-term Europe-wide model (EU_LUR). It can be explained by the heterogeneous local variability in different cities. The European model has been optimized to explain the combination of mainly inter-urban variability and partly local intra-urban variability in cities. In terms of explaining intra-urban concentration contrasts, the performance of an Europe-wide (global scale) model varied across smaller areas and was generally lower than the overall performance at the global scale (Lu et al., 2020; de Hoogh et al., 2018). Meyer and Pebesma, 2021, 2022 also pointed out that large-scaled models are challenging to capture and assess the detailed local variability, especially at the region with conditions that are very different from the training data. Given the local variance was well captured by fine-grained mobile measurements collected in the studied region, incorporating detailed and specifical local information improved the European model.

4.2. The choice of global scale

The core point of selecting the optimal global scale is to maximize the representability of global measurements to represent the true local long-term concentrations. A balance is needed between two factors, namely the similarity to local long-term concentrations and the number of global long-term measurements.

Increasing the global scale, more diverse instances are included in training, which complexes the instance space resulted in capturing more variances and a larger range of measurements. This approach increased the R^2 of Global2Local models (R^2 of EU2AMS > NLBELUXDE2AMS > NL2AMS). Meanwhile, the similarity is decreasing as more sites outside of the studied region are included. NO_2 was previously found to vary strongly in space (McAdam et al., 2011; Karner et al., 2010). Areas nearby tend to reflect a similar mix of emission sources (e.g., car fleets, fuel types etc) and a similar dispersion pattern such as similar meteorology and topography. Involving more sites that are not similar to the local measurements increased the absolute error (the MAE and RMSE of NLBELUXDE2AMS < EU2AMS).

However, decreasing the global scale may suffer from a lower number of long-term monitoring sites. Several studies explored the minimal number of stationary sites for developing LUR models to capture the distribution patterns of air pollution (Wang et al., 2012; Basagaña et al., 2012). Model robustness increased substantially until about 40–80 sites were used and more modestly with further increase in number of training sites (Basagaña et al., 2012). However, they were conducted at the same scale where the highest similarity applied. Higher similarity requires a smaller number of sites. In our previous transfer-learning work, the idea of Local2Local was implemented where 82 local long-term measurements within the study area were divided into a training (as the transfer target) and a validation set (50/50 random split). With 41 local long-term measurements as the transfer target, our previous work reported a good performance ($R^2 = 0.54$), although different spatial units and covariates were used (Yuan et al., 2022). When using long-term measurements from a larger geographic area as the transfer target, the decreased similarity demands more sites. For example, the absolute error of NL2AMS predictions was expected to be the lowest. But in our result, the absolute error was higher than that of the NLBELUXDE2AMS model. This is because the Dutch long-term measurements consist only out of 68 sites, limiting its representability (3,243 and 878 sites for EU2AMS and NLBELUXDE2AMS respectively). Empirical tuning of global scale thus is necessary to find the balance between the number of sites and the similarity when applying the Global2Local method to other regions.

For estimating air pollution concentrations in Amsterdam, NLBELUXDE2AMS met the balanced point and achieved the lowest RMSE and MAE (Table 4). Meanwhile, the R^2 remained at a similar level to AMS_LUR. We thus recommend using NLBELUXDE2AMS to map long-term air pollution in Amsterdam. Using sites from the Netherlands and its neighboring countries, the Global2Local model benefits from a large number of long-term measurements. In addition, the selected neighboring countries are part of the same airshed where similar emission sources and dispersion patterns are expected (Anderson et al., 2013), which further increases the similarity.

4.3. Model performance at major traffic and urban background locations

The mobile LUR model predicted higher concentrations than long-term validation measurements, especially for the major traffic locations (Fig. 4C.). The overestimation is mainly related to the on-road monitoring, while the validation sites are located at road-side or building façade (spatial differences) which has been previously documented by several studies (Kerckhoffs et al., 2022; McAdam et al., 2011; Karner et al., 2010; Richmond-Bryant et al., 2017). The used outlier-sensitive aggregation value (the mean based mobile measurements) may also contribute to the overprediction as the mean value is sensitive to a random high value. The median value was on average 2 μg/m^3 lower than the mean aggregated mobile measurements. However, previous studies did not find that the overpredictions of the mobile LUR model is significantly higher at major traffic than urban background locations (12.6 vs 1.6 μg/m^3 in Fig. 4C., overpredicted 38.8% and 6.5% to the mean of Palmes measurements respectively). The higher level of overprediction at traffic locations than urban background locations can be explained by the fact that at major traffic locations, the traffic volume, acting as the major emission source, decreased more dramatically from daytime to nighttime than at urban background locations. But the mobile campaign was performed only during the working hours.

The reduced residuals at major traffic locations can be attributed to the combined efforts of narrowing differences in space and time. Adding in the global long-term information, NLBELUXDE2AMS substantially reduced the residuals at major traffic locations compared to AMS_LUR (5.7 vs 12.6 μg/m^3 in Fig. 4C.). Meanwhile, low prediction errors at background sites were maintained (0.6 μg/m^3). This suggests that the proposed Global2Local model can inherit the detailed spatial information from local mobile measurements and simultaneously capture the pattern of temporal variations by leveraging temporally rigorous global measurements.

4.4. Strength and limitations

An advantage of our proposed Global2Local model is that it extends the application boundary of transfer-learning LUR models to regions with sparse local long-term measurements. This is a very common situation, as the number of long-term monitoring sites per city is generally small. Our previous paper indicates that with 41 sites at the local scale, transfer learning models can achieve a good performance on estimating Amsterdam concentrations (Yuan et al., 2022). While quantifying the required number of global sites is challenging due to the complex interactions with the forehead-mentioned similarity as well as the size of the studies region. Two basic rules of application are 1) sufficient global long-term measurements must be available; 2) global knowledge must contain a certain level of local knowledge. In Europe and North America, regulated pollutants such as NO_2 are general intensively monitored. However, for unregulated pollutants such as Ultra Fine Particles (UFP) and Black Carbon (BC), long-term measurements are not readily available at a national or continental scale. For these pollutants, specific cohort studies may provide some global data (Saha et al., 2021). Furthermore, within Europe, routine monitoring of UFP is increasing, suggesting more future possibilities of applying the transfer learning approach. For the second rule, the European AirBase measurements cover many different regions with different climatic, socioeconomic and demographic situations. This diverse information provides the Global2Local model the potential to be applied in other European regions with local mobile monitoring data.

Unlike on-road mobile measurements, AirBase monitors spread over more diverse locations, including on-road and off-road residential locations (e.g., parks) and industrial sites (Air Quality e-Reporting, 2021). The diversity in AirBase locations increases the generalization and robustness of the Global2Local model to estimate air pollution at all locations covering the whole city. Note that although the AirBase measurements primarily include traffic locations, both background and major traffic monitoring locations provide useful information for residential exposures, for example, in minor and major streets respectively. The 24-h measured Global AirBase data also complements diurnal mobile data in terms of narrowing the temporal difference. Narrowing these gaps in space and time, the Global2Local model outperforms the other LUR models based solely on measurements from a single geographic scale.

Our Global2Local model resulted in an R^2 of 0.45 indicating a moderate performance in explaining variability, taking into account that the validation was performed at external validation points and not based on cross-validation in the same monitoring domain. In Amsterdam, the state-of-the-art external validated accuracy reported in several previous studies were $R^2 = 0.45–0.5$, nMAE = 0.10–0.20, nRMSE = 0.20–0.30 (Yuan et al., 2022; Kerckhoffs et al., 2022). The NLBELUXDE2AMS in this work achieved a similar level: $R^2 = 0.43$, MAE = 5.5, nMAE = 0.21, RMSE = 6.9 and nRMSE = 0.27. Moreover, quite a few models have been applied in epidemiological studies and reported robust associations with health effects that have similar R^2 values (de Hoogh et al., 2018; Chen et al., 2020). We acknowledge that applying a model that explains only a moderate proportion of the true variability could lead to some misclassification of air pollution exposure. When applied in epidemiological studies, the most common consequence is that health risks may be underestimated.

Two main factors limit the performance of Global2Local. First, global long-term measurements are still biased to represent the local long-term measurements (limited similarity). Measurements from other regions outside the studied region partially introduce additional noise (see the difference to Palmes validation data in Fig. 1). Additionally, the different instruments between global AirBase and local Palmes measurements also contributed to the dissimilarity, although the differences between the instruments are probably small because they are all carefully calibrated. Second, the predictors used in the Global2Local model must be consistent with the global model. The algorithm of TrAdaBoost precludes adding additional features. For example, the traffic density covariates involved in our previous paper were found as the most important variable to delineate the spatial distribution patterns of NO_2 concentrations (Kerckhoffs et al., 2022; Yuan et al., 2022) . But these traffic related covariates were not available at the global scale in this study. Therefore, they were not added in the Global2Local model. In our sensitivity test, we found that if including traffic variables, the R^2 of the local mobile LUR model increased from 0.44 to 0.56, consistent with an earlier study from Beelen et al., 2013, where the difference of LUR model performance for NO_2 was about 0.1 in R^2 comparing LUR models with and without traffic density variables). Further improvement of performance is expected by adding global traffic variables to Global2-Local models.

5. Conclusion

The presented work demonstrates that integrating global long-term measurements with local mobile data, the Global2Local method mitigates the ubiquitous shortcomings of applying mobile measurements to estimate long-term concentrations at residential addresses. Our proposed Global2Local model can inherit the advantages of data instances from both scales and outperform traditional LUR models trained with measurements exclusively at the global or local scale. Given the increasing demand for hyperlocal air pollution mapping, more mobile monitoring campaigns are planned globally. Our proposed Global2Local model can be widely used to transfer mobile measurements to optimize long-term estimation of residential concentrations with fine spatial resolution, which is preferred in environmental epidemiological studies.

Credit author statement

Z.Y.: Conceptualization, Methodology, Formal analysis, Writing – original draft; J.K.: Data curation, Supervision, Writing – review & editing; Y.S.: Resources, Writing – review & editing; K.H.: Writing – review & editing; G.H.: Supervision, Writing – review & editing; R.V.: Data curation, Supervision, Project administration, Writing – review & editing.

Funding

The project received funding from the Environmental Defense Fund (United States), Google (United States), EXPOSOME- NL (NWO; project number 024.004.017) and EXPANSE (EU-H2020 Grant number 874627). This work made use of the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-1923.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

The authors gratefully acknowledge the help and coordination (data, discussions) of Karin Tuxen-Bettman and Natalie Smailou, Google, United States. Help, support, and coordination (discussions) of Harry van Bergen, Paul Coops, and Imke van Moorselaar, the Municipality of Amsterdam. We thank Kees Meliefste for the design and instrumental setups of the car and processing of data by Fares Al Hasan.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2023.115836.

References

Air Quality e-Reporting (AQ e-Reporting) — European Environment Agency . https://www.eea.europa.eu/data-and-maps/data/aqereporting-9 (accessed 2021-11-23).

Anderson, H.R., Favarato, G., Atkinson, R.W., 2013. Long-term exposure to outdoor air pollution and the prevalence of asthma: meta-analysis of multi-community prevalence studies. Air Qual. Atmosphere Health 6 (1), 57–68. https://doi.org/10.1007/s11869-011-0145-4.

Apte, J.S., Messier, K.P., Gani, S., Brauer, M., Kirchstetter, T.W., Lunden, M.M., Marshall, J.D., Portier, C.J., Vermeulen, R.C.H., Hamburg, S.P., 2017. High-resolution air pollution mapping with Google street View cars: exploiting big data. Environ. Sci. Technol. 51 (12), 6999–7008. https://doi.org/10.1021/acs.est.7b00891.

Basagaña, X., Rivera, M., Aguilera, I., Agis, D., Bouso, L., Elosua, R., Foraster, M., de Nazelle, A., Nieuwenhuijsen, M., Vila, J., Künzli, N., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642. https://doi.org/10.1016/j.atmosenv.2012.01.064.

van de Beek, E., Kerckhoffs, J., Hoek, G., Sterk, G., Meliefste, K., Gehring, U., Vermeulen, R., 2020. Spatial and spatiotemporal variability of regional background ultrafine particle concentrations in The Netherlands. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.0c06806.

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., Dėdelė, A., Gražulevičienė, 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.

Boersma, K.F., Eskes, H.J., Dirksen, R.J., van der A, R.J., Veefkind, J.P., Stammes, P., Huijnen, V., Kleipool, Q.L., Sneep, M., Claas, J., Leitão, J., Richter, A., Zhou, Y., Brunner, D., 2011. An improved tropospheric NO_2 column retrieval algorithm for the Ozone monitoring instrument. Atmos. Meas. Tech. 4 (9), 1905–1928. https://doi.org/10.5194/amt-4-1905-2011.

Brandt, J., Silver, J.D., Frohn, L.M., Geels, C., Gross, A., Hansen, A.B., Hansen, K.M., Hedegaard, G.B., Skjøth, C.A., Villadsen, H., Zare, A., Christensen, J.H., 2012. An integrated model study for Europe and North America using the Danish eulerian hemispheric model with focus on intercontinental Transport of air pollution. Atmos. Environ. 53, 156–176. https://doi.org/10.1016/j.atmosenv.2012.01.011.

Chambliss, S.E., Preble, C.V., Caubel, J.J., Cados, T., Messier, K.P., Alvarez, R.A., LaFranchi, B., Lunden, M., Marshall, J.D., Szpiro, A.A., Kirchstetter, T.W., Apte, J.S., 2020. Comparison of mobile and fixed-site Black Carbon measurements for high-resolution urban pollution mapping. Environ. Sci. Technol. 54 (13), 7848–7857. https://doi.org/10.1021/acs.est.0c01409.

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 (24), 15698–15709. https://doi.org/10.1021/acs.est.0c06595.

Copernicus Climate Change Service, 2019. ERA5 Monthly Averaged Data on Pressure Levels from 1979 to Present. https://doi.org/10.24381/CDS.6860A573.

CORINE Land Cover — Copernicus Land Monitoring Service . https://land.copernicus.eu/pan-european/corine-land-cover (accessed 2021-07-28).

Dai, W., Yang, Q., Xue, G.-R., Yu, Y., 2007. Boosting for transfer learning. In: Proceedings of the 24th International Conference on Machine Learning - ICML '07. ACM Press: Corvalis, Oregon, pp. 193–200. https://doi.org/10.1145/1273496.1273521.

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 (5), 670–675. https://doi.org/10.1289/ehp.0901818.

Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 55 (1), 119–139. https://doi.org/10.1006/jcss.1997.1504.

Geostat - Gisco - Eurostat. accessed 2021-11-22. https://ec.europa.eu/eurostat/web/gisco/geodata/reference-data/population-distribution-demography/geostat.

Guerreiro, C., 2013. Air Quality in Europe : 2013 Report. Publications Office of the European Union.

Guerreiro, C.B.B., Foltescu, V., de Leeuw, F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384. https://doi.org/10.1016/j.atmosenv.2014.09.017.

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.

Imperviousness — Copernicus Land Monitoring Service . https://land.copernicus.eu/pan-european/high-resolution-layers/imperviousness (accessed 2021-11-22).

Karner, A.A., Eisinger, D.S., Niemeier, D.A., 2010. Near-roadway air quality: synthesizing the findings from real-world data. Environ. Sci. Technol. 44 (14), 5334–5344. https://doi.org/10.1021/es100008x.

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 (3), 1413–1421. https://doi.org/10.1021/acs.est.8b06038.

Kerckhoffs, J., Khan, J., Hoek, G., Yuan, Z., Ellermann, T., Hertel, O., Ketzel, M., Jensen, S.S., Meliefste, K., Vermeulen, R., 2022. Mixed-effects modeling framework for Amsterdam and copenhagen for outdoor NO $_2$ concentrations using measurements sampled with Google street View cars. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.1c05806 acs.est.1c05806.

Kouw, W.M., Loog, M., 2021. A review of domain adaptation without target labels. IEEE Trans. Pattern Anal. Mach. Intell. 43 (3), 766–785. https://doi.org/10.1109/TPAMI.2019.2945942.

Lu, M., Schmitz, O., de Hoogh, K., Kai, Q., Karssenberg, D., 2020. Evaluation of different methods and data sources to optimise modelling of NO2 at a global scale. Environ. Int. 142, 105856 https://doi.org/10.1016/j.envint.2020.105856.

Map features - OpenStreetMap Wiki . https://wiki.openstreetmap.org/wiki/Map_features#Highway. (accessed 2023-03-30).

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., Chéroux, F., Colette, A., Coman, A., Curier, R.L., Denier van der Gon, H.a.C., Drouin, A., Elbern, H., Emili, E., Engelen, R.J., Eskes, H.J., Foret, G., Friese, E., Gauss, M., Giannaros, C., Guth, J., Joly, M., Jaumouillé, E., Josse, B., Kadygrov, N., Kaiser, J.W., Krajsek, K., Kuenen, J., Kumar, U., Liora, N., Lopez, E., Malherbe, L., Martinez, I., Melas, D., Meleux, F., Menut, L., Moinat, P., Morales, T., Parmentier, J., Piacentini, A., Plu, M., Poupkou, A., Queguiner, S., Robertson, L., Rouïl, L., Schaap, M., Segers, A., Sofiev, M., Tarasson, L., Thomas, M., Timmermans, R., Valdebenito, Á., van Velthoven, P., van Versendaal, R., Vira, J., Ung, A., 2015. A regional air quality forecasting system over Europe: the MACC-II daily ensemble production. Geosci. Model Dev. (GMD) 8 (9), 2777–2813. https://doi.org/10.5194/gmd-8-2777-2015.

McAdam, K., Steer, P., Perrotta, K., 2011. Using continuous sampling to examine the distribution of traffic related air pollution in proximity to a major road. Atmos. Environ. 45 (12), 2080–2086. https://doi.org/10.1016/j.atmosenv.2011.01.050.

Messier, K.P., Chambliss, S.E., Gani, S., Alvarez, R., Brauer, M., Choi, J.J., Hamburg, S.P., Kerckhoffs, J., LaFranchi, B., Lunden, M.M., Marshall, J.D., Portier, C.J., Roy, A., Szpiro, A.A., Vermeulen, R.C.H., Apte, J.S., 2018. Mapping air pollution with Google street View cars: efficient approaches with mobile monitoring and land use regression. Environ. Sci. Technol. 52 (21), 12563–12572. https://doi.org/10.1021/acs.est.8b03395.

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.13650, 2041-210X.13650.

Meyer, H., Pebesma, E., 2022. Machine learning-based global maps of ecological variables and the challenge of assessing them. Nat. Commun. 13 (1), 2208. https://doi.org/10.1038/s41467-022-29838-9.

Pardoe, D., Stone, P., 2010. Boosting for regression transfer. In: Proceedings of the 27th International Conference on International Conference on Machine Learning. ICML'10; Omnipress, Madison, WI, USA, pp. 863–870.

Richmond-Bryant, J., Chris Owen, R., Graham, S., Snyder, M., McDow, S., Oakes, M., Kimbrough, S., 2017. Estimation of on-road NO2 concentrations, NO2/NOX ratios, and related roadway gradients from near-road monitoring data. Air Qual. Atmosphere Health 10 (5), 611–625. https://doi.org/10.1007/s11869-016-0455-7.

Saha, P.K., Hankey, S., Marshall, J.D., Robinson, A.L., Presto, A.A., 2021. High-spatial-resolution estimates of ultrafine particle concentrations across the continental United States. Environ. Sci. Technol. 55 (15), 10320–10331. https://doi.org/10.1021/acs.est.1c03237.

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.

SRTM 90m Digital Elevation Database v4.1. CGIAR-CSI . https://cgiarcsi.community/data/srtm-90m-digital-elevation-database-v4-1/(accessed 2021-11-22).

Tropospheric Emission Monitoring Internet Service . https://www.temis.nl/(accessed 2022-01-18).

Wang, M., Beelen, R., Eeftens, M., Meliefste, K., Hoek, G., Brunekreef, B., 2012. Systematic evaluation of land use regression models for NO_2. Environ. Sci. Technol. 46 (8), 4481–4489. https://doi.org/10.1021/es204183v.

Yuan, Z., Kerckhoffs, J., Hoek, G., Vermeulen, R., 2022. A knowledge transfer approach to map long-term concentrations of hyperlocal air pollution from short-term mobile measurements. Environ. Sci. Technol. https://doi.org/10.1021/acs.est.2c05036.