Science of the Total Environment 922 (2024) 171251

Contents lists available at ScienceDirect

Science of the Total Environment

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

LUR modeling of long-term average hourly concentrations of NO_2 using hyperlocal mobile monitoring data

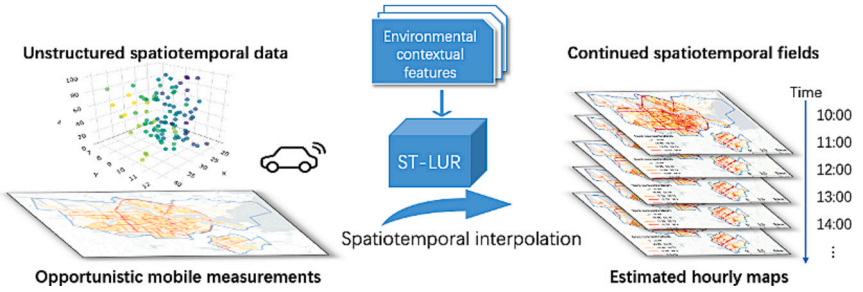
Zhendong Yuan [a,*], Youchen Shen [a], Gerard Hoek [a], Roel Vermeulen [a,b], Jules Kerckhoffs [a]

[a] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands
[b] Julius Centre for Health Sciences and Primary Care, University Medical Centre, University of Utrecht, the Netherlands

HIGHLIGHTS

- Mobile monitoring can measure fine spatiotemporal air pollution concentrations.
- Rare studies focus on producing temporally varying maps using mobile measurements.
- Mobile data follow the pattern of routine monitors, but with significant deviations.
- GTWR can efficiently map long-term average hourly air pollution concentrations.
- Hourly maps facilitate dynamic exposure assessment based on human activities.

GRAPHICAL ABSTRACT



ARTICLE INFO

Editor: Hai Guo

Keywords:
Hourly mapping
Mobile monitoring
Hyperlocal variations
NO_2
LUR
Geostatistics

ABSTRACT

Mobile monitoring campaigns have effectively captured spatial hyperlocal variations in long-term average concentrations of regulated and unregulated air pollutants. However, their application in estimating spatiotemporally varying maps has rarely been investigated. Tackling this gap, we investigated whether mobile measurements can assess long-term average nitrogen dioxide (NO_2) concentrations for each hour of the day. Using mobile NO_2 data monitored for 10 months in Amsterdam, we examined the performance of two spatiotemporal land use regression (LUR) methods, Spatiotemporal-Kriging and GTWR (Geographical and Temporal Weighted Regression), alongside two classical spatial LUR models developed separately for each hour. We found that mobile measurements follow the general pattern of fixed-site measurements, but with considerable deviations (indicating collection uncertainty). Leveraging heterogeneous spatiotemporal autocorrelations, GTWR smoothed these deviations and achieved an overall performance of an R^2 of 0.49 and a Mean Absolute Error of 6.33 μg/m^3, validated by long-term fixed-site measurements (out-of-sample). The other models tested were more affected by the collection uncertainty. We highlighted that the spatiotemporal variations captured in mobile measurements can be used to reconstruct long-term average hourly air pollution maps. These maps facilitate dynamic exposure assessments considering spatiotemporal human activity patterns.

* Corresponding author.
E-mail address: z.yuan@uu.nl (Z. Yuan).

https://doi.org/10.1016/j.scitotenv.2024.171251
Received 8 December 2023; Received in revised form 22 February 2024; Accepted 22 February 2024
Available online 27 February 2024

1. Introduction

With the recent development of location-based services, there is a growing trend toward incorporating time-activity patterns in environmental exposure estimation by considering human mobility (Wei et al., 2023; Lan et al., 2022; Kim and Kwan, 2018). So far, epidemiological studies primarily assign exposure estimates based on the home address using annual average concentration maps, neglecting human mobility patterns and the spatiotemporal variations of air pollution (Wei et al., 2023; Lan et al., 2022; Tang et al., 2018). Given the well-documented daytime variations in outdoor air pollution, there is a need to develop models with a better temporal resolution than the conventional annual average.

Spatiotemporal air pollution modeling typically relies on fixed-site measurements, which consist of complete time series with fine temporal resolution (van Zoest et al., 2020; Wai et al., 2022; Di et al., 2016; Di et al., 2019). However, they are often spatially sparse (commonly 0–5 sites per city). In contrast, mobile measurements have large spatial coverage and efficiently capture hyperlocal spatial variations (Apte et al., 2017; Kerckhoffs et al., 2022a). Equipped with high-frequency sensors (e.g., 1 Hz), the mobile collection platform measures concentrations directly on or next to streets. This produces dense spatial measurements but is often sparse in the temporal dimension at individual locations, posing challenges for interpolating continuous spatiotemporal fields (Mueller et al., 2016; Hofman et al., 2022; Hankey et al., 2019; Van den Bossche et al., 2020; Qin et al., 2022).

Generating long-term average air pollution maps with daytime variations can be achieved by developing spatial land-use regression (LUR) models separately for each hour. However, this strategy ignores the potential interconnections across hours (Mueller et al., 2016). A variant is to mix the instances from each hour to train a large model by taking the hour as a dummy variable (Hankey et al., 2019). Another option is to apply geostatistical methods that simultaneously model spatial and temporal autocorrelations, such as geographic temporal weighted regression (GTWR) and spatiotemporal kriging (ST-kriging) (Amato et al., 2020). These methods are broadly used to interpolate spatiotemporal environmental surfaces, including air pollution (van Zoest et al., 2020), precipitation (Hatvani et al., 2021) and soil pollutants (Zhao et al., 2023). Although primarily applied to consecutive fixed-site measurements, these methods can also effectively model unstructured spatiotemporal mobile monitoring data, with irregular observation points in space and time.

This paper evaluates the feasibility of utilizing mobile measurements in geostatistical-adapted spatiotemporal LUR models for mapping long-term average air pollution concentrations for daytime hours. Note that our focus is not on assessing air pollution concentrations for specific hours on specific dates. Two spatiotemporal LUR models (ST-kriging and GTWR) were compared with two spatial LUR models fitted separately for each hour: Random Forest (RF_hour) and Stepwise Linear Regression LUR models (SLR_hour). Considering the diverse spatiotemporal emission patterns in different seasons and road types, separating the situations may make spatiotemporal autocorrelations more pronounced and easier to model. The spatiotemporal LUR models were implemented in two ways: 1) using mobile measurements averaged over the entire collection period, and 2) dividing mobile measurements into four subsets based on road type and season, then training four submodels and averaging their predictions. A total of eight models were developed and the model performance was evaluated using independent routine monitoring data at nine external sites in Amsterdam.

2. Methodology:

2.1. Mobile data and independent validation data

Mobile data were collected by two Google Street View cars from March 2019 to May 2020, exclusively on weekdays between 9:00 and 20:00 (no weekends), covering all accessible roads in Amsterdam. A total of 5.7 million 1-s GPS points with NO_2 measurements (at 1 Hz, using CAPS monitor from Aerodyne Research Inc., Massachusetts, USA) were collected and aggregated into 50 m road segments in hourly intervals. Extreme values were removed (by 2.5th and 97.5th percentile), and all valid measurements were adjusted to reduce the daily variations in background conditions using a reference monitoring station located in a suburban area of Amsterdam, following methods described previously (Kerckhoffs et al., 2022b; Yuan et al., 2022). Briefly, we used the difference between the overall mean concentration and the mean of specific time windows at the reference station to adjust all mobile measurements corresponding to those time windows. The reference station provided measurements for the entire time period, covering all days of the week including daytime and nighttime, allowing the adjustment to effectively reflect long-term concentrations. More details of data preprocessing are provided in Appendix Text S1. We limited our study scope to frequently monitored hours to minimize modeling uncertainties and ensure a balanced representation of road types (Appendix Fig. S1). Specifically, we target estimating long-term averaged air pollution concentration maps for each hour from 10:00 to 20:00 and on weekdays only.

For validation, we used data from nine fixed-site routine monitors, run by the Dutch National Institute for Public Health and the Environment (RIVM) and the Public Health Service of Amsterdam (GGD), that were within 20 m of a road (i.e., street locations). These routine monitors measured continuous NO_2. For each routine monitoring station, we averaged the fixed-site measurements on an hourly basis throughout the 10-month mobile monitoring campaign and compared the values to the mobile measurements and model predictions on a nearby single road segment for external validation. These hourly averages represent the long-term averaged NO_2 concentrations for each hour. Note that the reference site used for the temporal correction is not among these nine routine monitors.

To better inspect the spatiotemporal pattern of NO_2, we divided the long-term fixed-site and mobile measurements into four categories by different road types (major vs residential roads) and seasons (warm and cold, divided by 31st Oct. 2019). The Dutch national traffic databases (NWB) classified major roads as more than 10,000 cars per 24 h (Home:: Nationaal Wegenbestand, n.d.). Five fixed-site monitors are located near major roads and four near residential roads.

2.2. LUR model implementations

We mathematically formulate the estimation of hourly maps as follows: given training instances $(X_{s,t}, Y_{s,t})$, where $Y_{s,t}$ are the mobile air pollution measurements and $X_{s,t}$ are a set of environmental contextual covariates at location s and hour t, we try to find a proper function $f(X_{s,t})$ that relates $X_{s,t}$ to $Y_{s,t}$.

To make a fair comparison, all predictor variables ($X_{s,t}$) were identical for the models tested. The predictor variables are aligned with those used in our previous papers (Kerckhoffs et al., 2022a; Kerckhoffs et al., 2022b; Yuan et al., 2022). They are spatially varying but temporally static, as dynamic variables are not available in Amsterdam in such fine resolution or do not differ within the city. The predictor features used in this work were: 1) land use extracted from the Copernicus CORINE dataset (CORINE, n.d.), which is a harmonized pan-European land use dataset; 2) traffic information such as traffic counts and road types derived from NWB (Home:: Nationaal Wegenbestand, n.d.); and 3) population density downloaded from Central Bureau of Statistics Netherlands (CBS) (Netherlands, S. CBS, n.d.). The specific variables including evaluated buffer sizes are summarized in Appendix Table S1.

2.2.1. Spatial LUR models for each hour

Two classic and widely used spatial LUR methods – Stepwise Linear

Table 1
Summary of the model inputs and algorithms.

Model category	*Model names*	*Algorithms*	*Training inputs*
Spatial models fitted separately for each hour[a]	SLR_hour	Linear regression	All mobile measurements
	RF_hour	Random forest	All mobile measurements
Spatiotemporal models[a]	OK_all	Ordinary kriging	All mobile measurements
	OK_four		Mobile measurements split into four categories[b]
	RK_all	Regression kriging	All mobile measurements
	RK_four		Mobile measurements split into four categories[b]
	GTWR_all	GTWR	All mobile measurements
	GTWR_four		Mobile measurements split into four categories[b]

[a] All models were validated using data from nine independent fixed-site routine monitors.
[b] Mobile measurements are split into four subsets (categories) by road types (major vs residential roads) and seasons (warm vs cold).

Regression (SLR_hour) and Random-Forest-based land use regression (RF_hour) were trained separately for each hour as the baseline models. SLR assumes a linear relationship between predictor features and air pollution measurements, while RF can capture nonlinearity by fitting a tree structure. Neither method uses correlations across hours and space. The implementation of SLR_hour and RF_hour is based on our previous work (Kerckhoffs et al., 2022a; Kerckhoffs et al., 2022b; Yuan et al., 2022) using *R* Software. More implementation details are provided in Appendix Text S2.

2.2.2. ST-kriging

Kriging methods are designed to estimate the value at unobserved locations by a weighted average of the observed neighborhood based on a manually defined autocorrelation structure. ST-kriging extends the 2D autocorrelation structure into 3D by adding time as the third dimension. A spatiotemporal variogram function was used to describe the spatiotemporal autocorrelation structure (Gräler et al., 2016). Details of the calculation are provided in Appendix Text S2.

Two variants of ST-kriging methods were explored for each hour: ordinary-kriging (OK) and regression-kriging (RK). OK assumes the mean and variance of the values are constant across the spatial field. RK is broadly used in air pollution modeling (van Zoest et al., 2020; Apte et al., 2017; Zhan et al., 2018). It integrates a regression model (the trend component, $\mu_{s,t}$ in Eq. (1)) and an OK model for the regression residuals (the residual component, $\eta_{s,t}$). In this work, we fitted the trend component with the hourly SLR model mentioned in Section 2.2.1. Then, the resulting residuals are modeled by OK to capture the remaining spatiotemporal autocorrelations.

$$\widehat{Y}_{s,t} = \mu_{s,t} + \eta_{s,t} \quad (1)$$

Different spatiotemporal variogram functions were tested for kriging methods, such as *Separable*, *sum-metric* and *product-Sum* (Gräler et al., 2016). The best fit was selected and implemented using the R package *gstat* (Gräler et al., 2016; Pebesma, 2004).

2.2.3. GTWR

GTWR is a spatiotemporal modeling technique. It fits local ordinary least squares (OLS) at each space-time location using the weighted spatiotemporal neighboring observations ($Y_{s,t}$) and a set of environmental contextual features ($X_{k,s,t}$), following Eq. (2). The coefficients of fitted OLS can differ at each location based on the spatiotemporal dependency presented in the proximity of data points, allowing for localized and context-specific predictions (Huang et al., 2010). A detailed explanation is provided in Appendix Text S2. Compared to spatial SLR and ST-kriging methods, the advantage of GTWR is that it can capture both spatial and temporal heterogeneities (allows the spatiotemporal nonstationary) as its coefficients at each space-time location vary.

$$Y_{s,t} = \beta_{0,s,t} + \sum_k \beta_{k,s,t} X_{k,s,t} + \varepsilon_{s,t} \quad (2)$$

GTWR was implemented following Shen et al., 2022 (Shen et al., 2022) using the package *GWmodel (*Gollini et al., *2015)* in R. $Y_{s,t}$ denotes the measured hourly mobile air pollution on road segments. The predictor features were paired with mobile measurements and put in the model as $X_{k,s,t}$. A widely recognized problem is that GTWR may introduce local collinearity (Hagenauer and Helbich, 2022). To limit the influence of collinearity, we applied SLR to preselect the most informative covariates (pooled instances from all hours). Then, NO_2 measurements at the prior and posterior time steps were used to estimate the selected coefficients.

2.2.4. Four categories divided by seasons and road types

When measurements are aggregated for an extended period (i.e., 10 months), it smooths out heterogeneous temporal variations in the data (e.g., seasonal differences), resulting in a decreased strength of the spatiotemporal semi-variance. Also, the spatiotemporal patterns of traffic volumes (emission sources) on major roads and residential roads are typically different. We therefore divided the data into four subsets, categorized by season and road type. We labeled the spatiotemporal kriging and GTWR methods as all_observations and four_categories (Table 1). Models with the suffix *"_all"* were trained using the hourly mobile data averaged over 10 months. The *"_four"* means models were trained by mobile measurements divided into four categories (four submodels), and the four model predictions were then averaged as the final long-term predictions.

2.3. Performance evaluation

We used measurements from nine fixed regulatory air quality stations in Amsterdam for validation. Their measurements were averaged for each hour of the day throughout the entire studied period. Model performance was then separately assessed spatiotemporally, temporally, and spatially. First, spatiotemporal validation was performed by calculating the overall accuracy from all sites and all hours tested (n = 90, 9 sites * 10 h). Second, temporal variations were assessed by inspecting model performance for each hour (n = 10 h). Furthermore, a direct comparison of the similarity between model predictions and fixed-site measurements was conducted at each site (n = 9 sites). Third, spatial validation was performed by plotting the predictions on maps and comparing the distribution patterns. Model coefficients in each hour were inspected.

The squared Pearson correlation (R^2), mean absolute error (MAE) and root mean square error (RMSE) were used to assess model performance. R^2 reflects the percentage of the 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 compared to MAE.

3. Results and discussion

We first illustrate the spatiotemporal variations captured by mobile measurements and the associated collection uncertainty compared to the long-term fixed-site measurements (Section 3.1). Second, we demonstrate the robustness of GTWR to collection uncertainty,

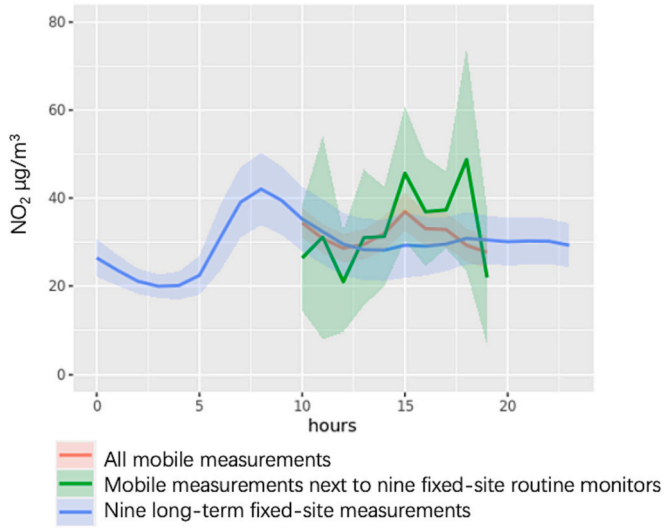


Fig. 1. Daytime variations of mobile and long-term fixed-site measurements of NO_2 averaged at nine routine monitoring locations, compared to the mean of all mobile measurements within Amsterdam. The shaded area reflects the 95 % confidence interval.

showcasing its ability to accurately reconstruct spatiotemporal maps of air pollution by spatiotemporal (Section 3.2), temporal (Section 3.3), and spatial validations (Section 3.4). Third, we discuss the strengths and limitations of applying mobile measurements to the spatiotemporal modeling of NO_2 (Section 3.5). Finally, implications of the produced spatiotemporal maps for the dynamic exposure assessment are discussed (Section 3.6).

3.1. Spatiotemporal variations for daytime hours

For each hour, our 10-month mobile campaign measured 29.4 %–55.9 % of all road segments between 10:00 and 20:00 in Amsterdam. For the measured road segments, on average, 1–3 drive passes (defined as distinct revisiting days) and 14–37 GPS points per road segment were monitored per hour (Appendix Table S2). The measured road segments were evenly distributed at major and residential locations, except for 9:00–10:00, 15:00–16:00 and 20:00–21:00. These were related to driving logistics such as driving in and out of the city using the main arteries. Considering the road types and the number of measurements, we decided to restrict the modeling period to the hours between 10:00 and 20:00. This restriction selected a balanced and representative dataset for developing spatiotemporal LUR models.

We observed significant daytime variations in NO_2 levels from mobile and long-term fixed-site measurements, as illustrated in Fig. 1. The mobile measurements close to the nine routine monitoring locations exhibited considerable deviations compared to the long-term fixed-site measurements, especially for two prominent peaks in the mobile data at 15:00 and 18:00. The deviations of mobile measurements are exaggerated as the comparison involves only nine road segments. These locations were occasionally under-sampled due to our collection strategy, where collection cars randomly traversed the city, leading to an uneven distribution of revisits per road segment.

Inspecting the city-wide distribution pattern of air pollution, we averaged mobile measurements across all locations in Amsterdam. We found that the mean value of the full set of mobile measurements aligns with the general patterns observed in long-term fixed-site measurements (Fig. 1), as the mean trendline fluctuates within the 95 % confidence interval bounds of the mean of long-term measurements at the nine routine monitoring sites. Remarkably, the second peak at 18:00 vanished. However, the 15:00 peak remained, although less pronounced. This persistence could be attributed to the unbalanced sampling, as drivers often started or ended their shifts around 15:00, leading to increased drives on the major ring road, which is often more polluted (illustrated in the spatial map of 15:00 in Appendix Fig. S1 and the high ratio between major and residential roads in Appendix Table S2). Furthermore, separating mobile measurements by season and road type, the 15:00 peak was exclusively observed at major road locations (Appendix Fig. S2).

3.2. Spatiotemporal validation

The model accuracy of the top-performing kriging and GTWR models compared to hour-specific spatial LUR models is presented in Table 2. The performance of the other models listed in Table 1 is shown in the Appendix Table S3. GTWR_four models outperformed the other LUR models with a moderate overall accuracy of $R^2 = 0.49$ and an MAE of 6.33 µg/m^3 in the spatiotemporal validation against long-term fixed-site measurements (Table 2). This performance is validated by independent long-term monitoring measurements. In the literature, a difference is often observed between the accuracy derived from cross-validation using mobile measurements (representative of the short-term on-road training domain) and the accuracy estimated through external fixed-site measurements (reflecting the long-term roadside application domain) (Yuan et al., 2022; Kerckhoffs et al., 2019). However, when assessing the environmental exposure for residents, long-term roadside air pollution concentrations are often more informative.

The performance of GTWR can be attributed to three key advantages. First, GTWR considers information from the spatiotemporal neighbors. Potential connections between prior and subsequent hours provide extra information. Therefore, GTWR_four modestly outperformed the spatial models (i.e., SLR_hour and RF_hour), which rely solely on spatial neighbors for each hour (Table 2).

Second, GTWR allows coefficients to vary across different spacetime

Table 2
Model performance of spatial and spatiotemporal LUR models calculated using long-term fixed-site measurements (spatiotemporal validation, $n = 90$, 9 sites * 10 h[a]).

Mean(min-max)[b]	Spatial models		Spatiotemporal models	
	SLR_hour	RF_hour	ST_OK_four[c]	GTWR_four[c]
R^2	0.47 (0.3–0.77)	0.33 (0.18–0.67)	0.35 (0.27–0.85)	0.49 (0.32–0.82)
MAE (µg/m^3)	6.48 (4.22–7.69)	10.33 (6.18–14.45)	9.46(7.02–14.59)	6.33 (4.70–8.11)
nMAE[c]	0.21 (0.14–0.25)	0.34 (0.20–0.48)	0.31 (0.23–0.48)	0.20 (0.15–0.26)
RMSE(µg/m^3)	8.04 (5.20–10.32)	14.74 (7.56–18.24)	12.14 (8.31–18.39)	7.88 (5.95–9.43)
nRMSE[d]	0.26 (0.17–0.33)	0.49 (0.25–0.60)	0.40 (0.27–0.60)	0.25 (0.19–0.30)

[a] Long-term average measurements (10 months) for daytime hours, as opposed to a specific day and hour.
[b] Min-max refers to the range of model performance during the tested daytime hours.
[c] ST_OK_four and GTWR_four refer to ordinary kriging and GTWR trained separately for the four categories as they achieved better performance than using all measurements.
[d] nMAE and nRMSE are the MAE and RMSE normalized by the mean of nine fixed-site validation measurements.

locations which can capture more heterogeneities of spatiotemporal relationships between covariates and the response. This adaptability is a primary factor contributing to the superior performance of GTWR_four, especially in comparison to kriging methods (e.g., ST_OK_four). Kriging methods assume that the spatiotemporal relationship (joint probability distribution) remains stationary across space and time. However, natural environmental factors often are complex and may violate this assumption (Hagenauer and Helbich, 2022). When divided into four categories based on road type and season, GTWR's flexible structure becomes even more pronounced, enabling it to capture more local heterogeneities within each category.

Third, the core of GTWR is the ensemble of multiple linear regressions which yields a higher generalization ability to mitigate the knowledge gap between the training and predicting domains (Yuan et al., 2022; Yuan et al., 2023). Contrarily, RF_hour is more affected by this knowledge gap, compared with linear regressions (Yuan et al., 2022). Random forest is an excellent machine learning regressor that can be efficiently parameterized according to the training loss. However, in our case, the training loss was approximated by mobile data that differs from the long-term validation data (short-term on-road vs long-term near-road) (Kerckhoffs et al., 2019; Kerckhoffs et al., 2017).

The features used to implement the GTWR_four model are listed in Appendix Text S3, with most of them being traffic-related. The fitted coefficients of GTWR_four vary strongly in space and time (Figs. S9-S12). For ST_OK_four, the sampled and fitted semi-variograms are shown in Appendix Fig. S3, with the spatiotemporal variogram function *Separable* as the best fit. This function assumes that the spatial and temporal variabilities can be modeled independently and combined to form the spatiotemporal variogram.

The implemented SLR_hour model revealed that different covariates were selected for each hour, as indicated by the heat map of selected features shown in Appendix Fig. S8. Aligned with previous mobile modeling studies (Kerckhoffs et al., 2019; Messier et al., 2018), traffic-related features were consistently identified as the most influential features across hours. Additionally, different land-use features such as the area of industry, green space, port, and airport were frequently selected across different hours. The top-10 feature importance of the best-fitted RF_hour per hour is presented in the Appendix Fig. S8. Like SLR_hour, traffic features were also constantly selected as the top-10 important features.

Notably, the spatiotemporal models do not consistently outperform the simple spatial models, as indicated in the spatiotemporal validation. ST_OK_four performed worse than the SLR trained separately each hour (i.e., SLR_hour). In contrast, GTWR_four slightly outperformed SLR_hour. This can be explained with the following two perspectives. First, the spatiotemporal model in this study was limited by the static covariates meaning they do not vary by hours, which failed to capture the intricate spatiotemporal variations comprehensively. This is recognized as a common issue of spatiotemporal modeling in the literature (Hankey et al., 2019; Van den Bossche et al., 2020). Second, SLR has been demonstrated as a robust and generalizable method in earlier mobile monitoring studies. With limited repeats and spatial coverage, SLR consistently produces accurate long-term average pollution maps (Apte et al., 2017; Kerckhoffs et al., 2022a). In our study, dividing mobile data into hourly intervals (10 parts), the number of measurements per hour remained sufficiently large to derive robust SLR models. Nevertheless, in the more common case of mobile campaigns with fewer collection days, spatiotemporal models may outperform SLR_hour more significantly.

The performance comparison of the other spatiotemporal models is provided in Appendix Table S3. The models fitted separately within the four categories and subsequently combined as the final predictions, yielded higher accuracy than fitting the averaged measurements over the entire period, particularly for ST_OK methods. Partitioning the data into the four categories based on road type and season, ST_OK_four was able to capture localized patterns of spatial and temporal heterogeneity, resulting in a more appropriate fit for the spatiotemporal semi-variance than ST_OK_all (Appendix Fig. S3). In contrast, the model accuracy of GTWR showed only slight differences between the *all* and *four* versions. This can be attributed to the varying coefficients of GTWR at individual space-time locations. The inherent flexibility relaxes the assumption of stationarity, thereby providing a higher generalization ability to fit local patterns. Further analysis of the performance of spatiotemporal models is provided in Appendix Text S4.

3.3. Temporal validation

Leveraging the dependencies among spatiotemporal neighbors, GTWR_four predictions showed resilience to collection uncertainty and exhibited smoother daytime variations than mobile measurements in Fig. 2. In contrast, the other LUR models generally inherited the uncertainty from mobile measurements. For example, at 15:00 and 18:00, the two peaks of mobile measurements exceed long-term fixed-site measurements by 16 and 18 $\mu g/m^3$. All LUR models were able to reduce this overestimation, especially GTWR_four. It smoothed out these two peaks with the smallest overestimation of only 3 $\mu g/m^3$ (about 10 % of the mean of long-term fixed-site measurements). This is further evident when comparing variations at individual sites (Appendix Fig. S5). Note that mobile measurements reflect on-road concentrations, and the remaining overestimation of GTWR_four could be partially attributed to the dispersion decay in NO_2 levels due to the distance from the road centerline to the routine monitoring locations (ranging from 5 to 20 m). In previous studies, we found that this part of overestimations is approximately 25 % (Kerckhoffs et al., 2022b).

The model performance for each hour was quantified in the temporal validation (Table 3). The accuracy varied across hours for all models. SLR_hour performed generally well during morning hours. GTWR_four exhibited a more consistent level of accuracy across daytime hours.

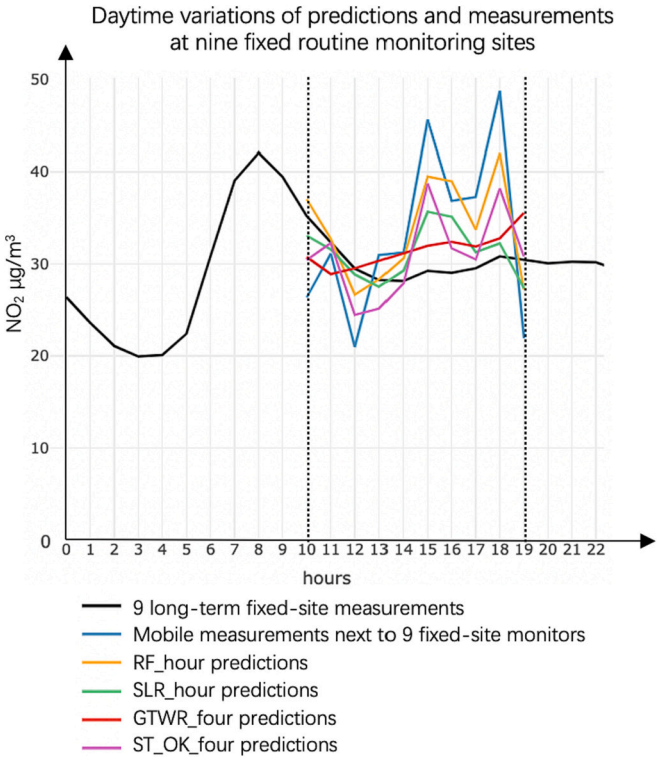


Fig. 2. Daytime variations of average NO_2 concentration predictions and measurements at nine routine monitoring sites.

Table 3
Model performance for each hour validated by long-term fixed-site measurements of NO_2 at nine routine monitoring sites (temporal validation, $n = 9$, nine sites for each hour of the day).

Hours[a]	*SLR_hour*			*RF_hour*			*GTWR_four*			*ST_OK_four*		
	R^2	*MAE*[b]	*RMSE*[b]	R^2	*MAE*[b]	*RMSE*[b]	R^2	*MAE*[b]	*RMSE*[b]	R^2	*MAE*[b]	*RMSE*[b]
Hour10	0.55	5.93	7.45	0.67	8.29	8.99	0.55	6.81	8.39	0.42	7.35	9.49
Hour11	0.77	4.22	5.20	0.50	12.86	17.32	0.39	7.51	9.21	0.45	12.97	17.14
Hour12	0.55	5.68	6.93	0.37	6.35	10.33	0.54	5.81	7.06	0.43	7.97	10.33
Hour13	0.47	6.78	7.74	0.65	6.18	7.56	0.69	4.70	6.06	0.36	10.02	11.18
Hour14	0.58	5.73	6.62	0.51	8.31	9.64	0.82	5.19	5.95	0.45	7.12	9.11
Hour15	0.56	7.69	10.32	0.54	14.45	18.24	0.60	5.86	7.42	0.85	11.03	12.45
Hour16	0.75	6.73	8.49	0.36	11.76	15.32	0.32	7.45	9.07	0.37	8.69	11.08
Hour17	0.30	7.95	8.57	0.22	8.15	10.20	0.45	5.32	7.22	0.29	7.02	8.31
Hour18	0.30	6.17	8.85	0.18	17.02	26.88	0.52	7.17	8.38	0.27	14.59	18.39
Hour19	0.42	5.49	7.12	0.32	9.91	11.83	0.74	8.11	9.43	0.27	7.87	9.44
Mean	0.53	6.24	7.73	0.43	10.33	13.63	0.56	6.39	7.82	0.42	9.46	11.69
STD[c]	0.16	1.10	1.41	0.17	3.60	5.90	0.16	1.16	1.27	0.17	2.64	3.43

[a] Hour refers to the start hour. For example, hour 10 means 10:00–11:00.
[b] In the unit of $\mu g/m^3$.
[c] STD: standard deviation.

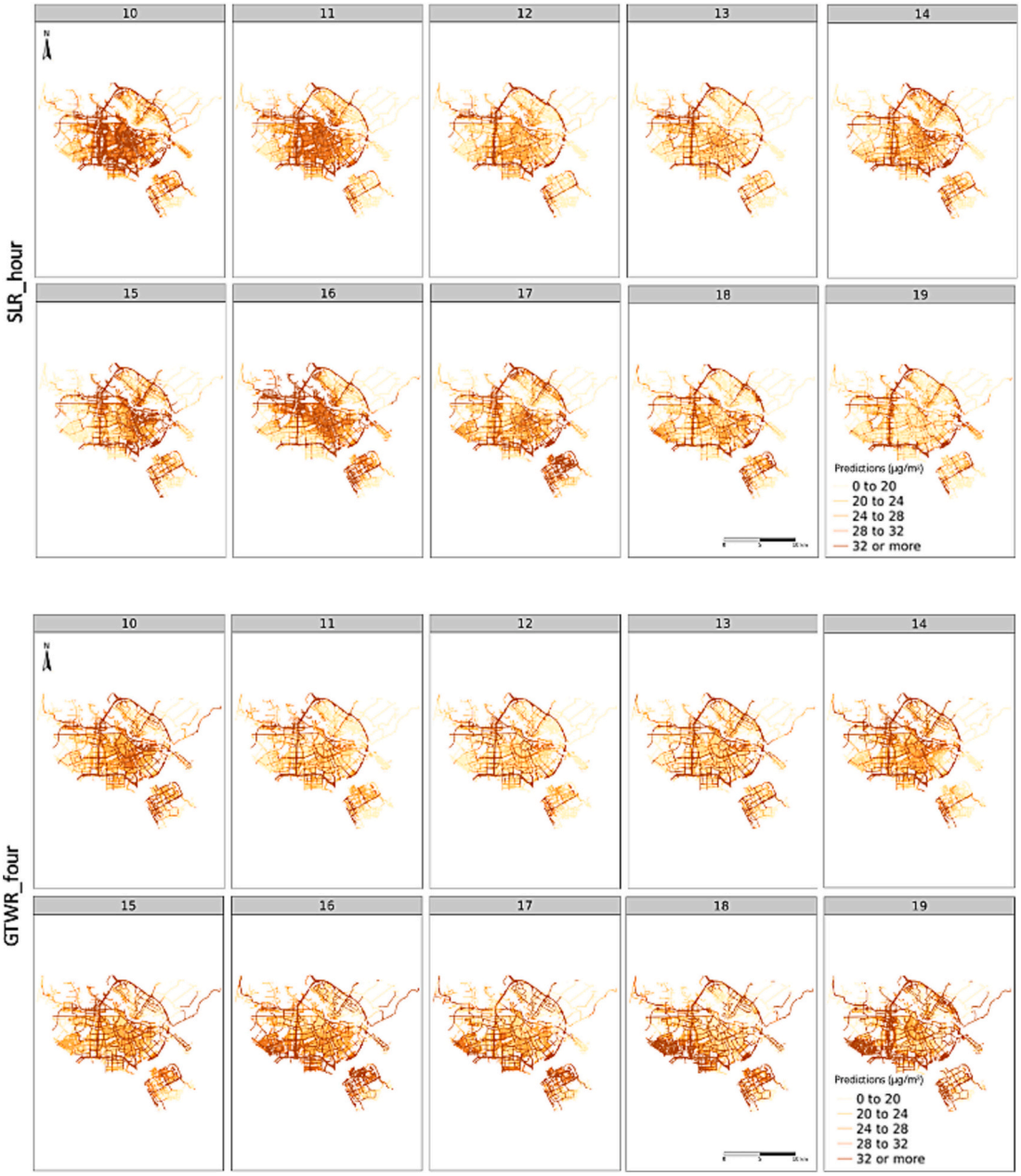


Fig. 3. Spatial maps of model predictions of hourly averaged NO_2 concentrations during daytime hours (SLR_hour and GTWR_four). The enlarged version and the maps of RF_hour and ST_OK_four are provided in Appendix Fig. S6.

3.4. Spatial validation

The spatial distribution of hourly NO_2 predictions is plotted in Fig. 3 and Appendix Fig. S6. Overall, all models indicate that the ring road surrounding Amsterdam was highly polluted throughout all daytime hours. Predictions from all models and mobile measurements show high NO_2 concentrations clustering in the center at 10:00 and gradually decreasing after that (from 11:00 to 13:00). Between 14:00 and 16:00, NO_2 concentrations increase again in the center. Differences between spatial and spatiotemporal LUR models occur mainly between 16:00 and 19:00. Specifically, RF_hour and SLR_hour tended to predict higher concentrations in the city center. In contrast, GTWR and ST-kriging models estimated higher levels in the southwest and southeast corners, where population density is higher (population density map in Appendix Fig. S7.). Notably, the predictions from GTWR_four correspond to the daily commuting patterns in Amsterdam, in which citizens travel to work in the center in the morning and return home in the afternoon.

3.5. Strengths & limitations

We demonstrated the feasibility of applying mobile measurements to map hyperlocal hourly NO_2 concentrations covering a large city (Amsterdam) during daytime hours. To date, pollution maps heavily relied on spatially sparse fixed-site observations. These methods often assume stationarity or simplified distributions for unobserved locations, resulting in spatiotemporally averaged predictions. Mobile measurements capture more heterogeneous spatial variation than fixed-site monitors by directly measuring air pollution on each road segment.

We illustrated that geostatistics-adapted spatiotemporal LUR models, especially GTWR, are suitable to fit the data structure of mobile measurements. Mobile measurements show a random distribution in space and time, lacking consecutive time series for each location (temporal sparsity). GTWR and ST-kriging are also easy to interpret and require less extensive training data than other machine learning models.

We acknowledge that we limited the scope to Amsterdam and only during daytime hours potentially limiting the generalizability of these results. However, the daytime hours covered in this study are the hours with most human activity outside the home. Next, using temporally static covariates capped the goodness-of-fit of spatiotemporal models. It has been documented in a previous study that incorporating more spatiotemporally resolved traffic and meteorologic information (e.g., wind direction and speed) can enhance the performance of LUR models predicting air pollution concentrations for a specific day and hour (Van den Bossche et al., 2020). Nevertheless, such data are often not available in fine spatiotemporal resolution. Furthermore, the number of our external validation data is limited, comprising only nine fixed long-term sites in Amsterdam. However, this number is typically considered substantial in routine monitoring for a city. Prior spatiotemporal modeling work was validated on a smaller number of fixed sites, typically 2–3 sites, as seen in studies conducted in Antwerp and Oakland (Hofman et al., 2022). Additionally, except for RF, the LUR models tested in this work were based on the linear relationship between environmental covariates and air pollution. Another variant of GWR such as geographically weighted artificial neural network (GWANN) can model the non-linear and non-stationary relationship (Hagenauer and Helbich, 2022; Wu et al., 2021). However, parameterizing GWANN requires a substantial amount of mobile measurements, and currently, no spatiotemporal version of this approach is available. Finally, although linear regression is more generalizable, it can only partially bridge specific knowledge gaps between short-term on-road mobile measurements and the target long-term residential concentrations.

3.6. Implications for exposure assessment studies

An ongoing debate in environmental epidemiology pertains to the accuracy of home-address-based versus dynamic time-activity-based exposures, and the value of developing the latter. Our research contributes to this discourse by illustrating the pronounced hyperlocal variations in hourly averaged NO_2 levels across space and time. These variations highlight the necessity of incorporating spatiotemporal variations of air pollution into exposure assessments. Notably, substantial disparities in absolute values and spatiotemporal patterns were observed concerning different road types and seasons (as depicted in Appendix Fig. 2 and detailed in Appendix Text S5). These imparities emphasize the necessity of splitting models into four categories to capture local heterogeneities. Although the proposed GTWR model demonstrated moderate accuracy, the spatiotemporally varying maps produced are essential to enable dynamic exposure assessment as the hourly spatiotemporal variations of NO_2 are combined with citizens' time-allocation information.

4. Conclusions

Most mobile monitoring work in air pollution research has primarily focused on developing long-term average air pollution maps, emphasizing spatial variations (Kerckhoffs et al., 2022b; Yuan et al., 2023; Messier et al., 2018; Xu et al., 2022; Shi et al., 2016; Blanco et al., 2022). Only limited attention has been given to studying spatiotemporal variations (Hankey et al., 2019; Van den Bossche et al., 2020; Lane et al., 2015). Tackling this gap, we found that GTWR can capture the heterogeneity of spatiotemporal autocorrelations in mobile measurements. It achieved better overall performance compared to hour-specific spatial LUR models, validated by the external data (out-of-sample), in the context of mapping long-term average air pollution concentrations for daytime hours across the day rather than specific days and hours. Additionally, GTWR demonstrates resilience to the temporal collection uncertainty, highlighting its robustness. The interpolated hyperlocal hourly NO_2 maps hold the potential to enrich our understanding of dynamic exposure patterns and their implications in environmental health studies.

Funding

The project received funding from the Environmental Defense Fund, Google, EXPOSOME-NL (NWO; project number 024.004.017) and EXPANSE (EU-H2020 Grant number 874627). This work used the Dutch national e-infrastructure with the support of the SURF Cooperative using grant no. EINF-3851.

CRediT authorship contribution statement

Zhendong Yuan: Writing – original draft, Methodology, Formal analysis, Conceptualization. **Youchen Shen:** Writing – review & editing, Methodology, Data curation. **Gerard Hoek:** Writing – review & editing, Supervision. **Roel Vermeulen:** Writing – review & editing, Supervision, Funding acquisition. **Jules Kerckhoffs:** Writing – review & editing, Supervision, Data curation.

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.

Acknowledgments

The authors gratefully acknowledge the help and coordination (data,

discussions) of Karin Tuxen-Bettman and Natalie Smailou, Google Inc., USA. We thank Kees Meliefste for the car's design and instrumental setups, data processing by Fares Al Hasan and discussion of kriging methods with Professor Meng Lu from University of Bayreuth, Germany.

References

Amato, F., Guignard, F., Robert, S., Kanevski, M., 2020. A novel framework for spatio-temporal prediction of environmental data using deep learning. Sci. Rep. 10 (1), 22243 https://doi.org/10.1038/s41598-020-79148-7.

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.

Blanco, M.N., Gassett, A., Gould, T., Doubleday, A., Slager, D.L., Austin, E., Seto, E., Larson, T.V., Marshall, J.D., Sheppard, L., 2022. Characterization of annual average traffic-related air pollution concentrations in the greater Seattle area from a year-long Mobile monitoring campaign. Environ. Sci. Technol. 56 (16), 11460–11472. https://doi.org/10.1021/acs.est.2c01077.

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

Di, Q., Kloog, I., Koutrakis, P., Lyapustin, A., Wang, Y., Schwartz, J., 2016. Assessing PM2.5 exposures with high spatiotemporal resolution across the continental United States. Environ. Sci. Technol. 50 (9), 4712–4721. https://doi.org/10.1021/acs.est.5b061 21.

Di, Q., Amini, H., Shi, L., Kloog, I., Silvern, R., Kelly, J., Sabath, M.B., Choirat, C., Koutrakis, P., Lyapustin, A., Wang, Y., Mickley, L.J., Schwartz, J., 2019. An ensemble-based model of PM2.5 concentration across the contiguous United States with high spatiotemporal resolution. Environ. Int. 130, 104909 https://doi.org/10.1016/j.envint.2019.104909.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P., 2015. GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. J. Stat. Softw. 63, 1–50. https://doi.org/10.18637/jss.v063.i17.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using Gstat. R J. 8 (1), 204. https://doi.org/10.32614/RJ-2016-014.

Hagenauer, J., Helbich, M., 2022. A geographically weighted artificial neural network. Int. J. Geogr. Inf. Sci. 36 (2), 215–235. https://doi.org/10.1080/13658816.2021.1871618.

Hankey, S., Sforza, P., Pierson, M., 2019. Using mobile monitoring to develop hourly empirical models of particulate air pollution in a rural Appalachian community. Environ. Sci. Technol. 53 (8), 4305–4315. https://doi.org/10.1021/acs.est.8b05249.

Hatvani, I.G., Szatmári, G., Kern, Z., Erdélyi, D., Vreča, P., Kanduč, T., Czuppon, G., Lojen, S., Kohán, B., 2021. Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary. Environ. Int. 146, 106263 https://doi.org/10.1016/j.envint.2020.106263.

Hofman, J., Do, T.H., Qin, X., Bonet, E.R., Philips, W., Deligiannis, N., La Manna, V.P., 2022. Spatiotemporal air quality inference of low-cost sensor data: evidence from multiple sensor testbeds. Environ. Model Softw. 149, 105306 https://doi.org/10.1016/j.envsoft.2022.105306.

Home:: Nationaal Wegenbestand. (n.d.) https://www.nationaalwegenbestand.nl/ (accessed 2023-07-08).

Huang, B., Wu, B., Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. Int. J. Geogr. Inf. Sci. 24, 383–401. https://doi.org/10.1080/13658810802672469.

Kerckhoffs, J., Hoek, G., Vlaanderen, J., van Nunen, E., Messier, K., Brunekreef, B., Gulliver, J., Vermeulen, R., 2017. Robustness of intra urban land-use regression models for ultrafine particles and black carbon based on mobile monitoring. Environ. Res. 159, 500–508. https://doi.org/10.1016/j.envres.2017.08.040.

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.8b 06038.

Kerckhoffs, J., Khan, J., Hoek, G., Yuan, Z., Hertel, O., Ketzel, M., Jensen, S.S., Al Hasan, F., Meliefste, K., Vermeulen, R., 2022a. Hyperlocal variation of nitrogen dioxide, black carbon, and ultrafine particles measured with Google street view cars in Amsterdam and Copenhagen. Environ. Int. 170, 107575 https://doi.org/10.1016/j.envint. 2022.107575.

Kerckhoffs, J., Khan, J., Hoek, G., Yuan, Z., Ellermann, T., Hertel, O., Ketzel, M., Jensen, S.S., Meliefste, K., Vermeulen, R., 2022b. 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.1c058 06_.

Kim, J., Kwan, M.-P., 2018. Beyond commuting: ignoring individuals' activity-travel patterns may lead to inaccurate assessments of their exposure to traffic congestion. Int. J. Environ. Res. Public Health 16 (1), 89. https://doi.org/10.3390/ijerph16010089.

Lan, Y., Roberts, H., Kwan, M.-P., Helbich, M., 2022. Daily space-time activities, multiple environmental exposures, and anxiety symptoms: a cross-sectional mobile phone-based sensing study. Sci. Total Environ. 834, 155276 https://doi.org/10.1016/j.scitotenv.2022.155276.

Lane, K.J., Levy, J.I., Scammell, M.K., Patton, A.P., Durant, J.L., Mwamburi, M., Zamore, W., Brugge, D., 2015. Effect of time-activity adjustment on exposure assessment for traffic-related ultrafine particles. J. Expo. Sci. Environ. Epidemiol. 25 (5), 506–516. https://doi.org/10.1038/jes.2015.11.

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.

Mueller, M.D., Hasenfratz, D., Saukh, O., Fierz, M., Hueglin, C., 2016. Statistical modelling of particle number concentration in Zurich at high spatio-temporal resolution utilizing data from a mobile sensor network. Atmos. Environ. 126, 171–181. https://doi.org/10.1016/j.atmosenv.2015.11.033.

Netherlands, S. CBS. (n.d.) Statistics Netherlands. https://www.cbs.nl/en-gb (accessed 2023-07-09).

Pebesma, E.J., 2004. Multivariable geostatistics in S: the Gstat package. Comput. Geosci. 30 (7), 683–691. https://doi.org/10.1016/j.cageo.2004.03.012.

Qin, X., Do, T.H., Hofman, J., Bonet, E.R., La Manna, V.P., Deligiannis, N., Philips, W., 2022. Fine-grained urban air quality mapping from sparse mobile air pollution measurements and dense traffic density. Remote Sens. 14 (11), 2613. https://doi.org/10.3390/rs14112613.

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, Y., Lau, K.K.-L., Ng, E., 2016. Developing street-level PM2.5 and PM10 land use regression models in high-density Hong Kong with urban morphological factors. Environ. Sci. Technol. 50 (15), 8178–8187. https://doi.org/10.1021/acs.est.6b01807.

Tang, R., Tian, L., Thach, T.-Q., Tsui, T.H., Brauer, M., Lee, M., Allen, R., Yuchi, W., Lai, P.-C., Wong, P., Barratt, B., 2018. Integrating travel behavior with land use regression to estimate dynamic air pollution exposure in Hong Kong. Environ. Int. 113, 100–108. https://doi.org/10.1016/j.envint.2018.01.009.

Van den Bossche, J., De Baets, B., Botteldooren, D., Theunis, J., 2020. A spatio-temporal land use regression model to assess street-level exposure to black carbon. Environ. Model Softw. 133, 104837 https://doi.org/10.1016/j.envsoft.2020. 104837.

Wai, T.H., Apte, J.S., Harris, M.H., Kirchstetter, T.W., Portier, C.J., Preble, C.V., Roy, A., Szpiro, A.A., 2022. Insights from application of a hierarchical spatio-temporal model to an intensive urban black carbon monitoring dataset. Atmos. Environ. 277, 119069 https://doi.org/10.1016/j.atmosenv.2022.119069.

Wei, L., Kwan, M.-P., Vermeulen, R., Helbich, M., 2023. Measuring environmental exposures in people's activity space: the need to account for travel modes and exposure decay. J. Expo. Sci. Environ. Epidemiol. 1–9. https://doi.org/10.1038/s41370-023-00527-z.

Wu, S., Wang, Z., Du, Z., Huang, B., Zhang, F., Liu, R., 2021. Geographically and temporally neural network weighted regression for modeling spatiotemporal non-stationary relationships. Int. J. Geogr. Inf. Sci. 35 (3), 582–608. https://doi.org/10.1080/13658816.2020.1775836.

Xu, J., Zhang, M., Ganji, A., Mallinen, K., Wang, A., Lloyd, M., Venuta, A., Simon, L., Kang, J., Gong, J., Zamel, Y., Weichenthal, S., Hatzopoulou, M., 2022. Prediction of short-term ultrafine particle exposures using real-time street-level images paired with air quality measurements. Environ. Sci. Technol. 56 (18), 12886–12897. https://doi.org/10.1021/acs.est.2c03193.

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.2c 05036.

Yuan, Z., Kerckhoffs, J., Shen, Y., De Hoogh, K., Hoek, G., Vermeulen, R., 2023. Integrating large-scale stationary and local mobile measurements to estimate hyperlocal long-term air pollution using transfer learning methods. Environ. Res. 228, 115836 https://doi.org/10.1016/j.envres.2023.115836.

Zhan, Y., Luo, Y., Deng, X., Zhang, K., Zhang, M., Grieneisen, M.L., Di, B., 2018. Satellite-based estimates of daily NO2 exposure in China using hybrid random forest and spatiotemporal kriging model. Environ. Sci. Technol. 52 (7), 4180–4189. https://doi.org/10.1021/acs.est.7b05669.

Zhao, M., Wang, H., Sun, J., Tang, R., Cai, B., Song, X., Huang, X., Huang, J., Fan, Z., 2023. Spatio-temporal characteristics of soil Cd pollution and its influencing factors: a geographically and temporally weighted regression (GTWR) method. J. Hazard. Mater. 446, 130613 https://doi.org/10.1016/j.jhazmat.2022.130613.

van Zoest, V., Osei, F.B., Hoek, G., Stein, A., 2020. Spatio-temporal regression kriging for modelling urban NO $_2$ concentrations. Int. J. Geogr. Inf. Sci. 34 (5), 851–865. https://doi.org/10.1080/13658816.2019.1667501.