Environmental Research 256 (2024) 119233

Contents lists available at ScienceDirect

Environmental Research

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

Hourly land-use regression modeling for NO_2 and $PM_{2.5}$ in the Netherlands.

Aisha Ndiaye[a,*], Youchen Shen[a], Kalliopi Kyriakou[a], Derek Karssenberg[b], Oliver Schmitz[b], Benjamin Flückiger[c,d], Kees de Hoogh[c,d], Gerard Hoek[a]

[a] Division of Environmental Epidemiology, Institute for Risk Assessment Sciences, Utrecht University, Pobox 80125, 3508, TC Utrecht, the Netherlands
[b] Department of Physical Geography, Faculty of Geosciences, Utrecht University, Princetonlaan 8a, 3584 CB Utrecht, the Netherlands
[c] Swiss Tropical and Public Health Institute, Kreuzstrasse 2 CH-4123 Allschwil, Switzerland
[d] University of Basel, Petersplatz 1, Postfach, 4001, Basel, Switzerland

ARTICLE INFO

Keywords:
Air pollution hourly models
Land-use regression
Random forest
Temporal variation
Temporal adjustment

ABSTRACT

Annual average land-use regression (LUR) models have been widely used to assess spatial patterns of air pollution exposures. However, they fail to capture diurnal variability in air pollution and consequently might result in biased dynamic exposure assessments. In this study we aimed to model average hourly concentrations for two major pollutants, NO_2 and $PM_{2.5}$, for the Netherlands using the LUR algorithm. We modelled the spatial variation of average hourly concentrations for the years 2016–2019 combined, for two seasons, and for two weekday types. Two modelling approaches were used, supervised linear regression (SLR) and random forest (RF). The potential predictors included population, road, land use, satellite retrievals, and chemical transport model pollution estimates variables with different buffer sizes. We also temporally adjusted hourly concentrations from a 2019 annual model using the hourly monitoring data, to compare its performance with the hourly modelling approach. The results showed that hourly NO_2 models performed overall well (5-fold cross validation R^2 = 0.50–0.78), while the $PM_{2.5}$ performed moderately (5-fold cross validation R^2 = 0.24–0.62). Both for NO_2 and $PM_{2.5}$ the warm season models performed worse than the cold season ones, and the weekends' worse than weekdays'. The performance of the RF and SLR models was similar for both pollutants. For both SLR and RF, variables with larger buffer sizes representing variation in background concentrations, were selected more often in the weekend models compared to the weekdays, and in the warm season compared to the cold one. Temporal adjustment of annual average models performed overall worse than both modelling approaches (NO_2 hourly R^2 = 0.35–0.70; $PM_{2.5}$ hourly R^2 = 0.01–0.15). The difference in model performance and selection of variables across hours, seasons, and weekday types documents the benefit to develop independent hourly models when matching it to hourly time activity data.

Funding

This project was founded by HEI Research Agreement Number: 4972-RFA19-1/20-6).

1. Introduction

Air pollution, including nitrogen dioxide (NO_2) and particulate matter smaller than 2.5 μm ($PM_{2.5}$), is one of the largest environmental risks to health (World Health Organization: WHO, 2022). Outdoor air pollution is estimated to have caused 4.2 million premature deaths worldwide in 2019 (World Health Organization: WHO, 2022). Research has shown relationships between exposure to polluted air and a large number of health outcomes (Landrigan et al., 2018). To determine human exposure, personal spatial information is usually combined with spatiotemporal data on air pollution concentrations (Molter et al., 2010). Most epidemiological studies have estimated air pollution concentrations annually (Shen et al., 2022; De Hoogh et al., 2016) or seasonally (Boniardi et al., 2019; De Hoogh et al., 2018), and have assigned concentrations to the residential address as exposure estimates. To do so, different statistical and physical models have been developed, land-use regression (LUR) models being the most used statistical ones. LUR models relate measured concentrations to environmental predictors of air pollution allowing the development of high spatial resolution maps of pollutants concentrations (Lu et al., 2020a). Furthermore, they have shown to perform well, to be time efficient and applicable for

* Corresponding author. Pobox 80125, 3508 TC, Utrecht, the Netherlands.
E-mail address: a.ndiaye@uu.nl (A. Ndiaye).

https://doi.org/10.1016/j.envres.2024.119233
Received 27 February 2024; Received in revised form 24 May 2024; Accepted 25 May 2024
Available online 25 May 2024

long-term exposure assessment (Molter et al., 2010; Shen et al., 2022).

However, mapping air pollution at residential addresses fails to capture time-activity patterns of the population (Lu et al., 2020a; Mölter et al., 2010). People spend time at other locations, including work, school, leisure time locations and commute. Moreover, annual and seasonal models do not capture diurnal variability in air pollution, as traffic and other sources and atmospheric conditions change during the day and pollutants concentrations have shown to vary in accordance to rush hours, night-time and daytime (Boniardi et al., 2019; Dons et al., 2013; Lu et al., 2020b). Therefore, linking time activity data to annual average spatial concentration patterns is not optimal. Instead, linking the spatiotemporal variability of air pollution to human time and space activity patterns may improve our understanding of the health effects of air pollution. For this reason, there is a need to develop hourly air pollution models to capture diurnal variation of air pollution. Few studies have already done so (Lu et al., 2020b; Boniardi et al., 2019; Dons et al., 2013; Masiol et al., 2018; Weissert et al., 2020). However, developing hourly models can be highly time-consuming and hourly average concentration data is not always available, for example with filter-based $PM_{2.5}$ measurements.

Lu et al. (2020b) fitted independent LUR models for hourly NO_2, NO and O_3 concentrations in the Netherlands. They modelled hourly averages, for each month and weekday type over a five-year period from 2006 to 2011. Their models were built on a set of four preselected predictor variables identified in the development of annual average air pollution models. Their hourly models allowed for temporally varying predictors' coefficients (Lu et al., 2020b). Dons et al. (2013) developed black carbon (BC) hourly LUR models for Belgium using different strategies: by means of dummy variables, with dynamic dependent variables and with dynamic and static independent variables and using hourly concentrations to recalibrate the annual model. They found that independent hourly models had higher cross-validation R^2, and lower root mean squared error (RMSE) than models developed with preselected predictors (Dons et al., 2013). Boniardi et al. (2019) developed BC morning rush hour models for a neighbourhood in Milan using only air pollution data from 7:00h to 9:00h on working days. Both studies found differences in the predictors selected during rush hours and other hours of the day (Boniardi et al., 2019; Dons et al., 2013).

Other studies have used mobile or low-cost sensor fixed monitoring data to estimate hourly concentrations for small geographical areas, such as neighborhoods or counties. (Hankey et al., 2019; Masiol et al., 2018; Yuan et al., 2024; Weissert et al., 2020). The study by Masiol et al. (2018) developed typical hourly LUR models for weekdays and weekends, for the Monroe County (New York, USA) using the deletion/substitution/addition (D/S/A) algorithm. They collected PM concentrations with low-cost monitors at 23 sites (LCMs) for 12 months. The study by Yuan et al. (2024) developed NO_2 LUR models for the city of Amsterdam, using supervised linear regression (SLR), random forest (RF) and Geographical and Temporal Weighted Regression (GTWR). They modelled the typical hourly averages based on a 10-month mobile monitoring campaign. The studies by Van Den Bossche et al. (2020) and Patton (2014) developed a spatiotemporal LUR model, using dynamic hourly covariates and hourly averaged concentrations as independent/dummy variables. Both studies found that both temporal and spatial variables contributed significantly to the model performance. The study by Hankey et al. (2019) used mobile monitoring data to develop daytime average models, independent hourly models and a spatiotemporal model for UFP and BC in a rural Appalachian community. The spatiotemporal model performed worse than the daytime average model for both pollutants (Hankey et al., 2019).

Finally, another approach to include temporal variability in LUR models is the temporal adjustment of an annual average model (Brauer et al., 2008; De Nazelle et al., 2013; Wu et al., 2011). By applying a temporal trend to spatially resolved air pollution predictions, annual LUR models are rescaled by a fixed ratio or difference. Temporal adjustment approaches have been considered time efficient methods to include temporal or diurnal variation in air pollution exposure (Brauer et al., 2008; De Nazelle et al., 2013; Stafoggia et al., 2022; Wu et al., 2011). De Nazelle et al. (2013) adjusted an annual dispersion model to obtain hourly exposures for 36 subjects in Barcelona. Brauer et al. (2008) and Wu et al. (2011) adjusted an annual LUR model to obtain monthly exposure for pregnant women in Vancouver, Canada, and in Orange Counties, California, respectively. However, by employing such techniques, it is assumed that every location has the same temporal trend and that changes in pollution over time are spatially uniform.

The aim of this study was to assess the performance of independent air pollution LUR models of hourly averaged concentrations for the Netherlands for two major air pollutants, NO_2 and $PM_{2.5}$. We did not model specific hours of specific days individually, but rather we modelled the spatial variation of the hourly concentrations averages across three years combined (2016–2019), two seasons (cold, warm) and two weekday types (weekdays, weekends). We used two modeling methods, supervised linear regression (SLR) and the machine-learning algorithm random forest (RF). Regression models were fitted separately for each hourly average, for each season and weekday type. Thus, differently from previous studies, 96 models were built, with possibly different predictors and regression coefficients. Since temporal adjustment needs low data and computational input, we attempted to use this approach at a larger scale in this study. We compared the performance of the temporal adjustment approach to obtain hourly surfaces with the specific hourly SLR and RF models. As annual air pollution maps have already been produced for the Netherlands, it was sufficient to apply a diurnal trend to the annual predictions to obtain typical hourly concentrations, following the procedure operated in the ESCAPE project and by Gulliver et al. (2013).

The results of the study will be used in the MOBI-AIR project which aims at linking high spatial resolution maps with hourly time-activity patterns (MOBI-AIR, 2023). MOBI-AIR will investigate whether matching the spatiotemporal variability of air pollution with human time-activity patterns leads to less biased results in health studies compared to the traditional exposure assessment model based on the residential address only.

2. Methods

2.1. Design

Two different approaches were employed to obtain hourly concentrations for the Netherlands. Firstly, in our main approach we developed hourly models based on monitoring data from the Netherlands, and two neighbouring countries, Belgium, and Germany, to augment the number of monitoring locations. We calculated the average hourly concentration, for two seasons (cold and warm), for two weekday types (weekends and weekdays), over a three-year period, from 2016 until 2019. We modelled the hourly average concentrations for two pollutants, NO_2 and $PM_{2.5}$ with two approaches, supervised linear regression (SLR) and random forest (RF). The detailed technical implementations of SLR and RF can be found in Appendix A, section S1.1.1. To build the models 197 predictors were used, including population density, roads, land use, satellite retrievals, and chemical transport model estimates variables with varying buffer sizes (Table S1).

Secondly, we temporally adjusted the annual European model developed by Shen et al. (2022) to obtain hourly surfaces, for each season and weekday type. The temporal adjustment of annual average models is a widely used approach to temporally scale LUR models. In this study we compared the performance of our hourly LUR models with the temporal adjustment of the annual average European map.

2.2. Monitoring data

Hourly concentrations were downloaded from the European Environmental Agency (EEA) website for two pollutants: NO_2 and $PM_{2.5}$

(European Environmental Agency, 2020). Data was only used if more than 75% of the hourly observations were valid as defined by the EEA. We first looked at NO_2 and $PM_{2.5}$ patterns in the Netherlands for the year 2019, and aggregated hourly concentration by month, weekend, and weekdays. Previous studies showed that air pollution concentration and its variability differed between weekdays and weekends and between different parts of the year, especially between summer and winter months (Boniardi et al., 2019; De Hoogh et al., 2018; Lu et al., 2020b). By looking at the shape and the values of the monthly median concentrations and following the example of previous studies conducted in the Netherlands and elsewhere (Boniardi et al., 2019; De Hoogh et al., 2018), it was decided to aggregate the observations into two seasons: a warm season, from April until September, and a cold one from October until March.

For both NO_2 and $PM_{2.5}$ the mean concentration for each hour showed a large scatter across stations and hours. Moreover, the standard error (SE) of the mean varied a lot across months and stations, reaching high values, especially in the weekends, probably related to the small number of observations available to calculate time-specific averages. Thus, to reduce the SE and get more stable means, we aggregated hourly observations for a period of three years, from October 1, 2016 until September 30, 2019. This time frame was chosen to cover three full cold and warm seasons. In total, 26 stations were included for $PM_{2.5}$ and 69 for NO_2, with three types of locations: background, traffic and industrial.

Previous studies show that an increased number of observations provides more diverse monitoring data and allows the LUR models to better capture the relationships between predictors and pollutants, leading to improved performance and generalizability (Basagaña et al., 2012). Thus, to increase the number of observations to be used in training the LUR models in this study, we added stations from neighbouring countries, namely Belgium and Germany. The final dataset contained averaged hourly concentrations for the three countries over a three-year period, aggregated by weekend/weekdays and warm/cold seasons. The total number of stations for NO_2 and $PM_{2.5}$ was 544 and 227, respectively (Figures S1, S2).

2.3. Predictor variables

The predictors used in the study were the variables offered by Shen et al. (2022) for Europe-wide annual average air pollution models. They included several roads, land-use, population, satellite retrievals of NO_2 and $PM_{2.5}$, and chemical transport model estimates. The few time-varying predictors were averaged seasonally to align with the observations. These were temperature, precipitation, air pressure, wind speed, the satellite-derived seasonal tropospheric column density of NO_2, and the concentrations, aggregated from monthly estimates from the chemical transport model, Danish Eulerian Hemispheric Model (DEHM). Different circular buffer sizes (ranging from 50 m to 10,000 m) were applied to the predictors to capture the dispersion of the pollutants. In total 197 predictors were used to develop the models (Table S1). None of the predictors varied on an hourly basis.

2.4. Model development

In this study we developed independent LUR models for each hourly average. Other studies have implemented different approaches to develop typical hourly LUR models. One approach is to include hourly dummy variables (Dons et al., 2013). However, dummy LUR models assume the same spatial pattern throughout the day and preserve the same model structure (Dons et al., 2013). Single-hour LUR models instead, are more flexible, allowing for different slopes and predictors selected across hourly models. This comes at the expense of not using information from consecutive hours. Consequently, to test whether unrealistic changes occurred from hour to hour, we evaluated spatial prediction maps of consecutive hours.

To train the LUR models two algorithms were used: supervised linear regression (SLR) and random forest (RF), following the procedure used by Shen et al. (2022). Regression models were fitted separately for each hour of the day, for each season and for weekends and weekdays. Thus, for each pollutant, 96 models were built, with possibly different predictors and regression coefficients.

2.4.1. Supervised linear regression (SLR)

SLR is a widely used and standardized approach to develop LUR models (Chen et al., 2021; De Hoogh et al., 2018; Eeftens et al., 2012). In this study, we followed the ELAPSE protocol (De Hoogh et al., 2018) similarly to Shen et al. (2022) to train the SLR models. For each model, firstly the predictor variable with the highest R^2 value, and thus the one explaining the most variation in the concentrations, is added. In the following steps, additional predictor variables are considered for inclusion in the model if they improve the adjusted R^2, have the predetermined direction of effect, and if their coefficient value is statistically significant ($p < 0.1$). For each hour, a different SLR was built, with possibly different predictor variables.

2.4.2. Random forest (RF)

RF is a tree-based machine learning method that employs an ensemble approach (Breiman, 2001). Previous studies (Chen et al., 2019a, 2019b, 2020; Kerckhoffs et al., 2019, 2021; Lu et al., 2020a; Shen et al., 2022; Weissert et al., 2020) have demonstrated that RF performs similarly to linear regression in air pollution modeling. The advantage of RF is that it can handle non-linear associations, interactions between variables and highly correlated variables by randomly selecting a subset of variables in each split node of a tree. All variables in Table S1 were used to develop the models and variables' importance was calculated. The ranger package version 0.12.1 in R was used to train RF. Detailed information on how the RF was trained can be found in Shen et al. (2022).

2.5. Temporal adjustment

To calculate hourly estimates, we used the correcting factors obtained from the annual European model developed by Shen et al. (2022) and the hourly observations. First, we calculated the overall annual average of monitored concentrations for the years 2016–2019 combined for all stations in the Netherlands, grouped by seasons, weekends and weekdays. Then, we computed the hourly average of monitored concentrations for all stations, grouped by season, weekends and weekdays. Finally, we calculated the average difference in concentration between hourly means and annual averages, according to season and weekdays. This operation gave us the absolute difference correcting factor. Similarly, to obtain the ratio correcting factor, we divided the hourly means over the annual averages. We then added the resulting absolute difference factor to all predictions from the 2019 annual European model (Shen et al., 2022), and multiplied for the ratio factor. With this procedure we obtained hourly ratio and absolute difference temporally adjusted air pollution concentrations. In order to evaluate the performance of the temporal adjustment procedure, temporally adjusted concentrations at the monitoring sites were compared with the averaged measured concentrations at specific hours.

2.6. Validation

The R^2 and the root mean squared error (RMSE) of the 5-fold cross-validation (CV) were used to evaluate the performance of the models. CV included new model development. Observations were grouped based on station location and then randomly subdivided into 5 groups of equal size. Four groups were used to train the model, and one to test it. The 5-fold CV R^2 of the SLR and RF was calculated as an MSE-based R^2. To evaluate the performance of the temporal adjustment approach we calculated the MSE-based R^2, the correlation-based R^2, and the correlation between temporally adjusted estimates and the observations.

Moreover, to compare model predictions, we estimated air pollution concentrations using the hourly models developed by SLR and RF at twenty thousand locations randomly selected from the residential addresses of the BAG dataset. The BAG dataset contains all addresses from the Dutch address database.

3. Results

3.1. Air pollution diurnal patterns

Figs. S3 and S4 display the NO_2 and $PM_{2.5}$ hourly distribution of the concentrations for all stations in the Netherlands, grouped by season and day of the week. The mean concentrations in the weekends and in the warm season were consistently smaller than the ones from the weekdays and the cold season. The different spatial patterns between hours and weekday type, supported the necessity to assess weekends and weekdays concentrations separately. To compare the concentrations diurnal patterns between the Netherlands, Belgium and Germany, we plotted the hourly averaged concentrations and the medians separately (Fig. 1, S5–S8) for each country for all stations, grouped by season, weekend and weekdays. All countries showed two peaks in the cold season, around 8:00h and 18:00h, that reached similar concentration levels. The warm season also presented two peaks, with the second one being less pronounced. Moreover, they all showed lower concentrations in the weekends, both in the cold and warm season. For $PM_{2.5}$, there were two peaks in the cold season, the first around 9:00h and the second around 21:00h, for all three countries (Fig. 1; Fig. S6, S8). Even though the shape of the median concentration in the cold season was similar, Belgium had slightly higher values. In the warm season, all countries showed one peak in the morning, around 6:00h-7:00h, with lower concentrations compared to the cold season. There was no significant difference between weekends and weekdays for $PM_{2.5}$, however for consistency with NO_2 we kept the distinction between weekday type.

The diurnal patterns of the pollutants concentrations at the measurement sites for the three countries combined are shown in Figs. S9 and S10. For NO_2 both the cold and warm season weekdays diurnal variation had two peaks. In the weekends the peaks were less pronounced compared to the weekdays, but the variation followed a similar pattern to the weekdays. The concentrations in the cold season were consistently higher than the ones in the warm season, and similarly weekdays concentrations were higher than weekends. For $PM_{2.5}$ there was no clear difference between the weekdays and weekends diurnal concentrations. However, there were significant differences between seasons. In the cold season we could still distinguish two peaks, while in the warm season there was only one major peak.

3.2. Hourly models performance

The R^2 for the linear models derived from 5-fold cross validation (CV) are summarized in Fig. 2 and Table 1. The corresponding RMSE is shown in Fig. S11. The hourly models showed pronounced diurnal differences in R^2 values especially in the warm season (Fig. 2). The NO_2 models performed overall well (5-fold CV R^2 range per hour: 0.50–0.78), while the $PM_{2.5}$ models performed moderately (5-fold CV R^2 range per hour: 0.24–0.62). Both for NO_2 and $PM_{2.5}$ the warm season models performed worse than the cold ones, and the weekends' worse than weekdays'. The NO_2 models that performed the worst were the ones between 11:00h and 16:00h in the cold season, and between 11:00h and 18:00h for the warm season. The highest R^2 were in the nighttime hours models.

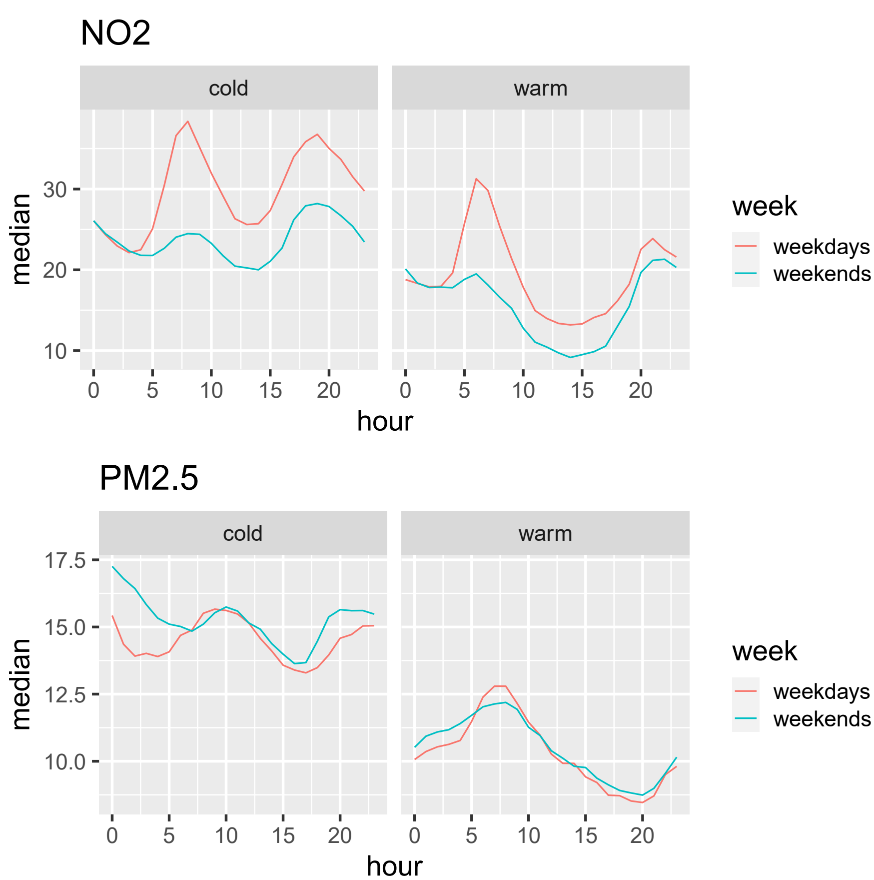

Fig. 1. Medians of the hourly average NO_2 and $PM_{2.5}$ concentrations observed in μg/m^3 from 2016 to 2019 for all stations in the Netherlands (NO_2 = 69 stations; $PM_{2.5}$ = 26 stations). Concentrations are averaged by season (cold = October till March, warm = April till September) and day of the week. (COLOR).

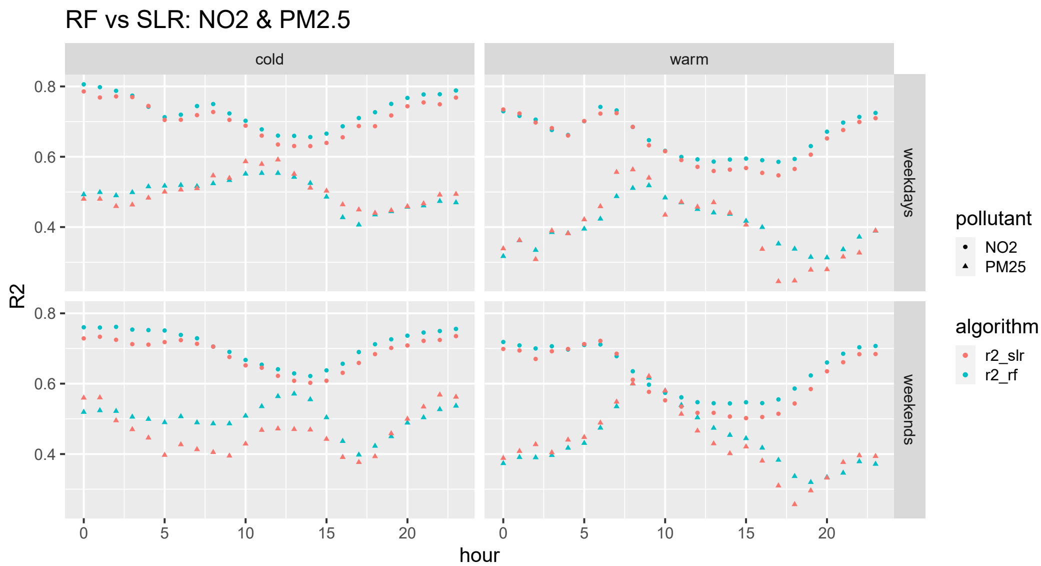

Fig. 2. R^2 of the 5-fold cross-validation hourly models developed through supervised linear regression (r2_slr) and random forest (r2_rf) for NO_2 and $PM_{2.5}$, for cold and warm season, weekdays, and weekends. The NO_2 models were built from monitoring data for the Netherlands, Belgium, and Germany, from 2016 to 2019 (N = 544 sites). The $PM_{2.5}$ models were built from monitoring data for the Netherlands, Belgium, and Germany, from 2016 to 2019 (N = 227 sites). The scales of NO_2 and $PM_{2.5}$ are different to illustrate hourly differences between methods, seasons and weekday types per pollutant. Note: the R^2 scale does not go from 0 to 1 to display fine differences in hourly R^2. (COLOR).

Table 1
5-fold cross validation R^2 of the NO_2 and $PM_{2.5}$ models developed through supervised linear regression (SLR) and random forest (RF). The mean_SLR and mean_RF columns show the mean R^2 of the hourly NO_2 and $PM_{2.5}$ models averaged by season (warm and cold), weekdays (No) and weekend (Yes). The columns SD_SLR and SD_RF show the standard deviation of the hourly R^2 values. The column Improvement shows the percentage increase in R^2 in RF compared to SLR.

Pollutant	Season	Weekend	mean_SLR	SD_SLR	mean_RF	SD_RF	Improvement
NO_2	cold	No	0.71	0.05	0.73	0.05	2%
	cold	Yes	0.69	0.04	0.71	0.05	2%
	warm	No	0.64	0.07	0.66	0.06	2%
	warm	Yes	0.61	0.08	0.64	0.07	3%
$PM_{2.5}$	cold	No	0.50	0.04	0.50	0.04	0%
	cold	Yes	0.46	0.06	0.50	0.04	4%
	warm	No	0.39	0.09	0.40	0.06	1%
	warm	Yes	0.43	0.09	0.44	0.09	1%

For $PM_{2.5}$, the explained variance was lower from 15:00h to 20:00h in the cold season, and from 16:00h to 21:00h in the warm season, when it dropped below 0.3. In the cold season, the lowest R^2 were found around 17:00h, while in the warm season we observed two drops: one around 00:00h and one in the afternoon around 18:00h.

In the case of NO_2, the models that performed the worst in both seasons were the late morning and early afternoon models. In the same time frame NO_2 concentrations showed a higher variability across stations (indicated by high RMSE). High RMSE values thus corresponded to lower R^2 values, suggesting that the model had difficulty explaining the substantial fluctuations in NO_2 levels (Fig. S11). On the other hand, for $PM_{2.5}$, lower concentration variability corresponded to lower R^2 values, implying that the model performed worse in capturing the limited variations in $PM_{2.5}$ levels (Fig. S11).

3.2.1. *Comparison between SLR and RF 5-fold CV performance*

The performance of the RF and SLR models was similar and comparable at all hours, with the RF 5-fold CV R^2 being slightly higher for most hourly models. None of the differences in R^2 between SLR and RF were statistically significant (*t*-test $p > 0.05$). In Fig. 2 the hourly R^2 of SLR and RF are plotted showing similar diurnal trends. For NO_2, RF performed overall slightly better than SLR. For $PM_{2.5}$, RF improved models' performance only slightly in the cold-weekdays, and more consistently in the weekends. The most significant improvement in performance was found on the cold-weekends, with a 4% increase in RF R^2 compared to SLR (Table 1).

3.2.2. *SLR model structure*

The patterns of predictors selection for NO_2 exhibit variations for different hours, seasons, and weekday types (Figs. S12, S13, S14, Section S2.1.1). During cold and warm season, weekdays and weekends models, roads and chemical transport model predictors were consistently chosen. Night-time hours models included prominently land use variables with larger buffer sizes, such as urban green, natural areas, and industry. In contrast, daytime hours selected more population-related predictors like residential areas and population density. In the weekday models, road predictors with small buffers demonstrated a consistent influence throughout the day, while larger buffer road predictors were chosen more frequently in the weekends. In the warm season, more variables were selected during the night-time models than in the daytime ones. Compared to the cold season models, the selection of industry predictors occurred more uniformly throughout the day. Additionally, natural areas were chosen with larger buffer sizes during early morning hours, in contrast to the cold season where they appeared in both morning and evening with small buffers. For both seasons, background land use predictors were selected until later morning hours (around 9:00h) in the weekends, with respect to weekdays.

The selected predictors for $PM_{2.5}$ are summarized in Figures S15, S16 and S17. The weekdays and weekends models were more similar compared to the ones for NO_2. Chemical transport models and chemical retrieval variables were selected in both season and weekday types for most models. Road predictors with small buffers were selected in the daytime models, consistently in the weekdays and less in the weekends, both in the cold and warm season. Population predictors were selected until 9:00h and after 18:00h in the weekdays, while in the weekends they were present with different buffer sizes in almost every model.

3.2.3. RF model structure

Figs. S18 and S19 give the top-10 variables in reducing the mean squared error (MSE) the most for each RF hourly model. As all 197 variables were kept to build RF models, variable importance is shared between highly correlated variables in different buffer sizes. Thus, it must be interpreted carefully.

For NO_2, most variables selected in the SLR had the highest variable importance in RF models. However, while in SLR the chemical transport model predictor was selected for almost all models, in RF it was present only for few models, from around 00:00h till 3:00h. Moreover, few land use predictors that were not selected in SLR were important for RF. For example, the total built-up area (tbu) predictor, which was not selected in the SLR, appeared quite consistently in the RF models. In the cold and warm weekdays, tbu was present with relatively large buffers from 5:00h in the morning until evening rush hours, while in the cold and warm weekends from around 8:00h till 22:00h. Both in the warm and cold season, roads predictors were present with similar patterns. Small buffers appeared mostly in the daytime, earlier for the weekdays (around 5:00h and 6:00h) and later in the weekends (from 9:00h); larger buffers replaced them in the evening, around the evening rush hours for weekdays and around 20:00h-21:00h in the weekends. Finally, the warm season models also had residential areas predictors in the early afternoon, that were not present in the cold season.

Estimates from chemical transport models and estimates from the satellite retrievals were present for the $PM_{2.5}$ like in SLR models, however, less consistently. In the warm season models, background land use variables were mostly present in the night-time, especially in the weekends, while in the cold season they were present with different buffer sizes throughout the day. Climate variables were present in the early afternoon hours in the cold season, while in the warm season mostly during the early morning and evening.

3.2.4. Spatial patterns

Fig. 3, S20, S21, S22 show the maps of predicted NO_2 and $PM_{2.5}$ concentrations estimated with RF, at two spatial resolutions: the whole

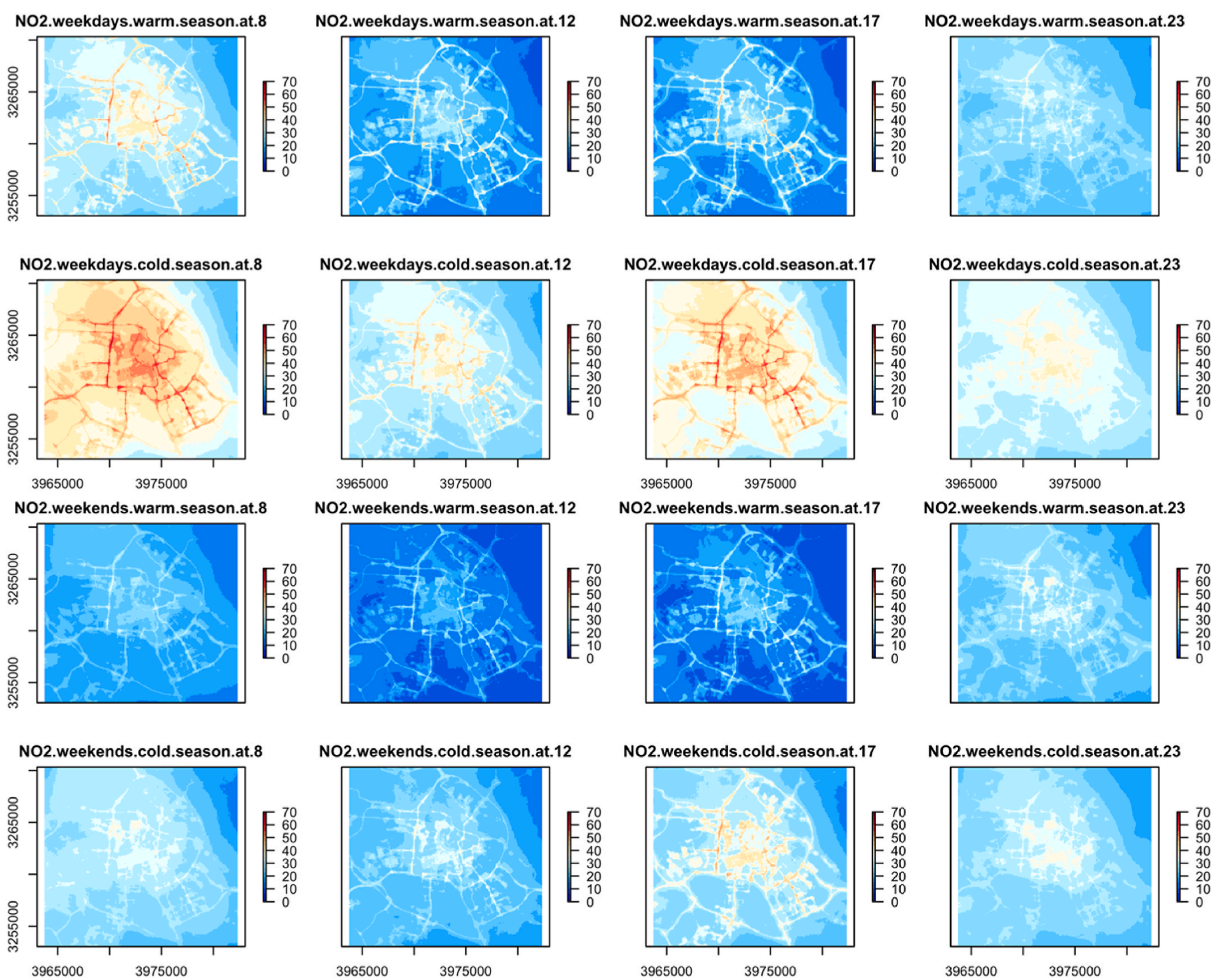

Fig. 3. Maps of predicted NO_2 concentrations in $\mu g/m^3$ for Amsterdam and surrounding area using RF, for different selected hours in the warm and cold season, weekdays and weekends. (COLOR).

of the Netherlands and the city of Amsterdam. Fig. 5 shows the spatial variation of NO_2 for the city of Amsterdam for selected hours across the day. From the maps we could see a clear spatial variation, with the urban areas being more polluted compared to rural areas. In both season and weekday types, an increased concentration towards the roads was visible, with the road network being visible at all hours. However, the differences in concentration were more pronounced during daytime hours. Moreover, during weekdays a peak in concentrations was visible at rush hours around the road network. From the maps it was also possible to see the seasonal and day of the week component, with concentrations in the warm season and weekends being lower and the difference in concentrations being less pronounced, possibly due to the more uniform distribution of traffic during the day (Fig. 3; Fig. S20). Finally, NO_2 concentrations were lower during night-time hours when spatial variation was also less pronounced.

For $PM_{2.5}$ there was also a spatial variation in concentrations, but less defined compared to NO_2 (Figs. S21 and S22). Spatial patterns differed modestly between hours, seasons and day of the week, particularly in the magnitude of the contrast. In the country level maps we could see that concentrations are higher in more populated areas, in the south and long the coast (Fig. S22). By looking at the maps for Amsterdam, the spatial patterns were less defined compared to NO_2, probably due to the different characteristics of the pollutants (Fig. S21). $PM_{2.5}$ concentrations are not only influenced by emission-related sources but also meteorological factors. Moreover, $PM_{2.5}$ can be suspended in the air for long periods of time and can travel long distances, contrarily to NO_2, that reacts with other compounds rapidly after it is released and does not get transported as far from its source as $PM_{2.5}$ (Eeftens et al., 2012).

3.3. Comparison of predictions of SLR and RF models at external locations

The SLR and RF models were also evaluated by comparing

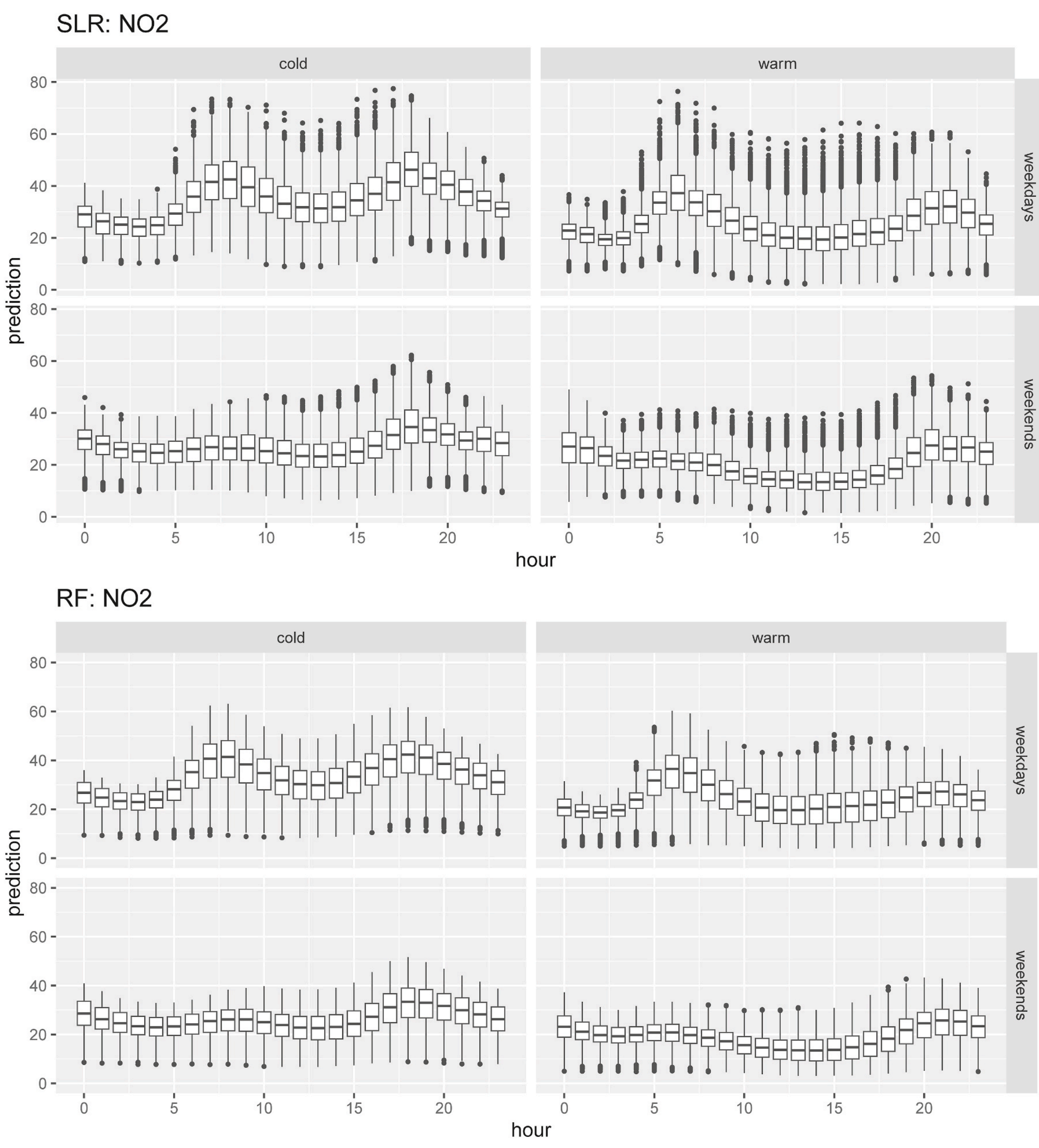

Fig. 4. NO_2 estimated concentrations in µg/m^3 at twenty thousand randomly selected residential addresses across the Netherlands, for the cold and warm season, weekdays and weekends, using SLR in the first graph and RF in the second.

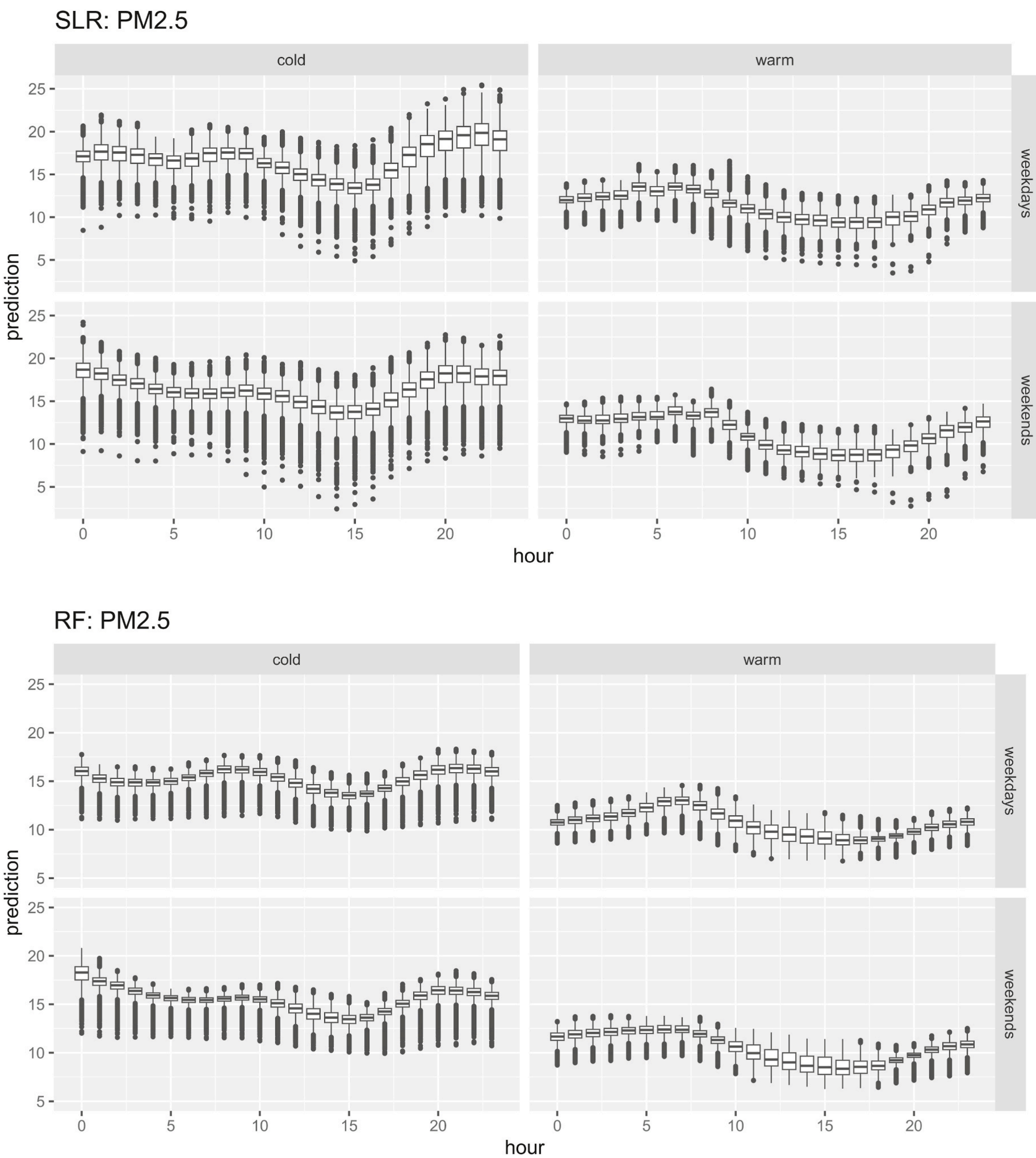

Fig. 5. $PM_{2.5}$ estimated concentrations in μg/m^3 at twenty thousand randomly selected residential addresses across the Netherlands, for the cold and warm season, weekdays (0) and weekends (1), using SLR in the first graph and RF in the second.

predictions of hourly concentrations at 20.000 Dutch residential addresses. The diurnal variation in predicted concentrations is displayed in Fig. 4 for NO_2 and 5 for $PM_{2.5}$. The predictions from SLR and RF followed a similar trend and took on a similar range of values for NO_2 and $PM_{2.5}$. For $PM_{2.5}$, the concentrations were slightly lower for RF compared to SLR. The predictions from both algorithms followed a diurnal trend similar to the one of the observations for the Netherlands (Figs. S3 and S4). Compared to the observations, the predictions took on slightly higher values and showed a greater hourly variation. This is possibly because the number of stations used to build the models was 227 for $PM_{2.5}$ and 544 for NO_2, while the predictions were made for 20.000 locations, and thus possibly included more polluted or remote points.

The correlation between predictions of the SLR and the RF modeling approach are shown in Figs. S23 and S24. For NO_2, the correlation was high, in both the cold and warm season, ranging from 0.85 to 0.95. For $PM_{2.5}$, the correlation was not as consistently strong as for NO_2 ($r = 0.21$–0.8). In the cold season the correlation was the lowest around 18:00h (0.21–0.51), while in the warm season around 21:00h (0.26–0.44). Around these hours both the SLR and the RF $PM_{2.5}$ hourly models performed the worst (Fig. 2). The difficulties in capturing the limited variations in $PM_{2.5}$ levels during these hours showed by the lower RMSE and lower R^2, resulted in a less strong correlation between SLR and RF estimates.

3.4. Temporal adjustment performance

In Figs. S25 and S26 the correlation between the observations and the temporal adjusted estimates from the European annual average model is shown for both seasons and weekday types. For NO_2 the graphs showed a good correlation in the cold season ($r = 0.67$–0.84) and in the warm one (r = 0.60–0.81). For $PM_{2.5}$, the correlation was less evident than for NO_2, in particular in the weekend hours, where it took on negative values. The correlation in the weekday hours ranged from 0.1 to 0.37 in the cold season, and from 0.02 to 0.4 in the warm one. In the weekends instead, it ranged from −0.14 – 0.39 in the cold season, and from −0.33 – 0.25 in the warm one. This was in line with the results of the SLR and RF, with the $PM_{2.5}$ models performing worse than the NO_2 ones, and thus predicting the concentrations less accurately. Moreover, the small negative correlations between the $PM_{2.5}$ observations and temporal adjusted estimates, suggest that not only the temporal adjustment did not explain the variability in the data, but also made predictions in the opposite direction of the actual outcomes.

Fig. S27 shows the hourly MSE-based R^2 of the temporal adjustment, which were consistently lower that the ones from the SLR and RF especially in the night-time models. As the validation monitoring data were also used to develop the diurnal correction factors, the presented R^2 values are too optimistic. Thus, because of the already lower R^2, we conclude that independent LUR models outperformed temporal adjustment. The MSE-based R^2 of NO_2 (−0.87 – 0.70) were higher than the ones for $PM_{2.5}$ (−0.68 – 0.15), most of which were below 0. The R^2 of NO_2 was especially low in the nigh-time hours. For $PM_{2.5}$, the prevalence of negative R^2, combined with the negative correlations, suggested that the models predicted worse than the average concentration, and sometimes in the opposite direction. For NO_2 instead, the relatively low R^2 and the high correlations could mean that while the model did not properly predict concentrations values, it mimicked the trend across stations.

To gain more understanding of the temporal adjustment performance, we also calculated the correlation-based R^2. The correlation-based hourly R^2 ranged between 0.35 and 0.70 for NO_2 and between 0.01 and 0.15 for $PM_{2.5}$ (Fig. S28). The R^2 of NO_2 was especially low in the nigh-time hours, when also the MSE-based R^2 was the lowest. For $PM_{2.5}$, the R^2 was still very low, especially in the weekends. The negative correlations and the low R^2, might indicate that the temporal adjusted annual model could not reproduce the diurnal variation at the Dutch measurement sites for $PM_{2.5}$.

4. Discussion

We developed hourly models for NO_2 and $PM_{2.5}$ using monitoring data from the Netherlands, Germany, and Belgium, from hourly averages over a three-year period (2016–2019) stratified by season and weekday type. The results of the study showed that it was possible to develop hourly models that performed well for NO_2 and moderately well for $PM_{2.5}$. RF and SLR performed similarly. Overall, the 5-fold CV R^2 of the RF models was modestly higher for most hours. The performance of hourly models was better than that of the temporal adjustment of annual the average model.

4.1. Difference in model performance and across hours for NO_2 and $PM_{2.5}$

The differences in model structure and performance observed across different hours, highlighted the benefit of developing independent hourly models. Moreover, the variations in model performance between seasons and weekday types indicated the benefit of temporal models tailored to specific seasons and days of the week. One possible risk of developing single-hour models is that of obtaining unrealistic changes in spatial patterns from one hourly model to the following one. However, our results showed that predictors selected changed gradually during the day (Figs. S13, S14, S16, S17). Moreover, the spatial prediction maps, showed stable spatial patterns with concentrations gradually increasing and decreasing from 1 h to the next one (Fig. S29).

For NO_2, the models for most hours performed well. The models that performed the worst in both seasons were the late morning and early afternoon models. In the same timeframe NO_2 concentrations showed a higher variability across stations, suggesting that the model had difficulty explaining the substantial fluctuations in NO_2 levels. On the other hand, for $PM_{2.5}$, lower concentration variability corresponded to lower R^2 values, implying that the model performed poorly in capturing the limited variations in $PM_{2.5}$ levels. Especially in the cold season, R^2 drops happened around morning and evening rush hours. The study by Masiol et al. (2018) also found lower R^2 for PM during rush hours. Their suggestion is that the emission of fresh carbonaceous particles during busy hours can affect both the way light interacts with particles in the air and the accuracy of reported measurements of particulate matter mass due to potential biases caused by the timing of measurements (Masiol et al., 2018). Moreover, our $PM_{2.5}$ models performed worse in the warm compared to the cold season. This was, potentially due to the presence of mechanisms unrelated to local sources that contribute to $PM_{2.5}$ levels during summer, which are difficult to tackle with our statistical models (Masiol et al., 2018).

Even though we did not observe a difference in performance in our Europe-wide annual modelling paper (Shen et al., 2022), in this study NO_2 models performed better than the $PM_{2.5}$ ones. This could be due to the following reasons. Firstly, the number of observations used to build the models was larger for NO_2, possibly leading to more robust models. The influence of the number of monitoring stations may be more important in hourly models, because the process of calculating averages per hour, season and day of the weekend results in less robust averages. A second difference is the larger concentration variability between stations and the more pronounced diurnal variation for NO_2. Finally, as NO_2 is more affected by local combustion sources than $PM_{2.5}$, the available source-related predictors may be more explanatory. Contrarily, LUR models do not represent particle formation processes very well (Jones et al., 2020).

Hourly LUR models have demonstrated to out-perform the temporal adjustment method. While the approach has been effective in predicting historical concentrations (Molter et al., 2010; Gulliver et al., 2013), it appeared to be less effective when extrapolating hourly concentrations from an annual model. This was evident from the R^2 values, which indicated that the absolute concentrations values were more accurately predicted using specific hourly models. Thus, relying on hourly models proved to be a more reliable approach for accurately estimating concentrations at different time intervals. This agrees with the comparison of BC models in Belgium (Dons et al., 2013).

RF and SLR hourly models performed similarly, consistent with other studies (Chen et al., 2019a; Shen et al., 2022). Most variables selected in the SLR models had high variable importance in RF models, though often with different buffer sizes. Variables with different buffer sizes are highly correlated and therefore may result in very similar model predictions, as documented by the high correlation between SLR and RF predictions and the similar model performance. A few land use predictors were important for RF and were not selected in SLR. The small differences in variables entering the models is partly due to differences in how variables are selected in SLR and rated for importance in RF.

4.2. Comparison with previous hourly modelling studies

Spatiotemporal patterns were in line with previous studies. The diurnal NO_2 variability across stations during all hours showed night-time lower concentrations corresponding to low traffic source intensity, and weekdays rush hours corresponding to air pollution peaks. For $PM_{2.5}$ there was no clear difference between the weekdays and weekends diurnal concentrations while there were significant differences between seasons.

Previous studies have developed more temporally refined LUR models than the annual average through different approaches. Studies have modelled average hourly concentrations to account for diurnal variation in source strength and weather, based on fixed site routine monitoring (Lu et al., 2020b), low-cost sensor fixed site networks (Masiol et al., 2018; Weissert et al., 2020) and mobile monitoring

(Hankey et al., 2019; Van Den Bossche et al., 2020; Yuan et al., 2024). One approach, which is the one implemented in this study, is to develop independent LUR model for each typical hour (Dons et al., 2013; Lu et al., 2020b) using monitoring data. The other studies that have used this approach found that night-time models performed the worst, differently from our results. In the study by Dons et al. (2013), 48 hourly models for BC were built, for weekend and weekdays separately based on monitoring data. The R^2 ranged from 0.07 to 0.8, with the night-time models performing the worst (Dons et al., 2013). Similarly, in the fixed-site routine monitoring study by Lu et al. (2020b) the NO_2 hourly models performance was lower during the night-time for NO_2 (hourly R^2 range: 0.39–0.89). Lu et al. (2020b) used preselected predictor variables to build the LUR models for all hours of the day. This choice comes with the advantage of assessing the relative importance of every variable in each model by comparing the coefficients. Both papers suggested that poor nigh-time models performance could be attributed to the lower pollutants concentration variation between locations during night-time, and to the inability of the preselected variables used to build the models to describe the night-time variations (Dons et al., 2013; Lu et al., 2020b). Our models, however, do perform well even in the evening hours, and indeed by looking at the concentrations variation in Figs. S3 and S4 the range of values taken by each boxplot is quite stable, especially in the weekends and especially for $PM_{2.5}$. Moreover, by allowing each model to choose independent and possibly different variables, the variation in air pollution is likely better represented for each hour. With regards to variable selection, the study by Lu et al. (2020b) found that during weekdays there was a sharp increase in the importance value of short-range buffers in the daytime, and long-range buffer in nigh-time models. Similar trends were also observed in our models, which could be explained by diurnal variation of local source strength.

In the mobile monitoring study by Yuan et al. (2024) the researchers found that NO_2 hourly models performed overall well and similarly using different approaches (SLR validation R^2 0.3–0.77; RF validation R^2 0.18–0.67; GTWR validation R^2 0.32–0.82). However, models could only be developed for the daytime period, between 9:00h and 20:00h. Similarly to Dons et al. (2013) and Lu et al. (2020b), the study by Yuan et al. (2024) found that SLR performed better during morning hours. Masiol et al. (2018) developed typical hourly models for PM based on 23 low-cost sensor sites, which performed well (average hourly R^2 0.7). Hankey et al. also developed single-hour models for UFP (10-fold CV R^2 0.11–0.69) and BC (0.12−0.57). The models that performed the best for BC were the morning and afternoon peak traffic models, when variation in concentrations could be better explained by the selected predictor variables (Hankey et al., 2019). The results from these studies revealed that different covariates were selected for each hourly model (Hankey et al., 2019; Masiol et al., 2018; Yuan et al., 2024). Similarly to our study, traffic related variables and average daytime traffic intensity were consistently identified across hours for NO_2. Even though the studies using mobile monitoring successfully developed typical hourly LUR models and may better capture spatial air pollution variability compared to fixed monitoring sites, their applicability is still limited. Indeed, mobile monitoring studies measure air pollution only during daytime hours and are usually applied in a relatively small study area (Hankey et al., 2019; Yuan et al., 2024).

Another approach is to add hourly dummy variables to an average LUR model (Dons et al., 2013). In the fixed-site monitoring based study in Flanders, the hourly dummy LUR model performed moderately well with a R^2 of 0.44, but with only two traffic predictors (Dons et al., 2013). Even though the dummy models performed moderately, the same spatial pattern was assumed across all hours and the hourly concentrations were not independent of each other, violating the assumption of independence in linear regression (Dons et al., 2013).

A third approach is that to include dynamic temporal covariates to account for temporal variability in air pollution. The study by Patton et al. (2014) developed a regression model (R^2 0.43) for measured UFP using temporal (1 h resolution) and spatial variables (20 m resolution). The study by Van Den Bossche et al. (2020) also developed a LUR model with mobile monitoring data (R^2 0.49), using hourly BC concentrations as an independent variable. These models however differ from our models in that they modelled the concentration at specific hours and thus needed to incorporate truly temporal variables.

4.3. Strengths and limitations

The main strength of this study is the inclusion of seasonal, diurnal and week variability in LUR models built with two algorithms, SLR and RF. In particular, our models took into consideration that sources of air pollution change on an hourly basis. Another strength of this study is the large number of locations used to train the SLR and RF models. Our independent hourly models did allow for different coefficients and different variables for each hour. Moreover, our models did not demand more data compared to the annual ones. Finally, the 5-fold CV R^2 is comparable with the one found in the other studies which developed hourly models, especially for NO_2. The good performance exhibited by our models suggests that the investment of time in their development is both feasible and valuable.

The main limitation of this study is that we did not use dynamic covariates to develop the LUR models. While some predictors had a seasonal or monthly resolution, no predictor had an hourly one. However, the study conducted in Belgium by Dons et al. (2013), which evaluated the performance of dynamic models for BC, found that adding dynamic covariates did not improve the performance of single-hour models with static covariates. In truly spatio-temporal models, where the aim is to model the concentration of a specific hour and date, temporal covariates on source strength (such as traffic intensity) and weather would be crucial. However, the aim of this study was to model the typical hourly average concentration. Moreover, most of the available predictors used, such as land use and population variables, were not available with hourly resolutions. Another limitation of the predictors used in the study is that did not include traffic volume and heating predictors because they were not available for the entire study area. With regards to the temporal adjustment methodology, the limitation of our application is that we used the most straightforward approach to temporally adjust the annual LUR model. Other approaches have been explored, that take into consideration spatial variation across stations. However, in this study we calculated only one correcting factor for each hour, assuming the same spatial pattern across stations.

In follow-up studies the same methodology could be used to predict hourly estimates of other pollutants. It would also be worthwhile to apply this approach in smaller study areas, where more detailed hourly information is available on predictors, and to compare our approach with other statistical approaches, including the dummy variable model, and the spatially and temporally refine temporal adjustment of the annual average.

Finally, our results showed that night-time hourly LUR models, especially for NO_2, performed better than daytime models. Future health studies using these hourly air pollution surfaces for exposure assessment in combination with time activity data should take this into account. At night, people are more likely to be at home in contrast to daytime hours when people are more mobile. Exposure miss-classification might therefore be higher during the day than during the night, affecting miss-classification of the overall exposure. However, the hourly R^2 values were still high for the daytime models, and it is equally relevant to model the air pollution variation of the hours that people tend to spend at home. We thus hypothesize that exposure miss-classification will be more driven by errors in mobility data (i.e. how well can we simulate or characterize time activity of a population) than by errors in air pollution models.

5. Conclusions

In this study, we showed that hourly LUR models perform overall well and that their performance is comparable with other studies. SLR and RF had similar 5-fold CV R^2, and both outperformed the temporal adjustment approach. The differences in models' performance and predictor variables selected across different hours, seasons, and weekday type has reinforced the importance of developing independent hourly models. Combining the NO_2 and $PM_{2.5}$ hourly models with hourly time-activity data will allow to develop dynamic exposure models. This will allow assessment of their added value in epidemiological studies compared to exposures assessed at the residential address only.

CRediT authorship contribution statement

Aisha Ndiaye: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Youchen Shen:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Kalliopi Kyriakou:** Writing – review & editing, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Derek Karssenberg:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization. **Oliver Schmitz:** Writing – review & editing, Visualization, Software, Resources, Methodology, Formal analysis, Data curation. **Benjamin Flückiger:** Writing – review & editing, Visualization, Software, Resources, Methodology, Formal analysis, Data curation. **Kees de Hoogh:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Gerard Hoek:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This project was founded by HEI Research Agreement Number: 972-RFA19-1/20-6). The authors declare no competing financial interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2024.119233.

References

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.

Boniardi, L., Dons, E., Campo, L., Van Poppel, M., Panis, L.I., Fustinoni, S., 2019. Annual, seasonal, and morning rush hour Land Use Regression models for black carbon in a school catchment area of Milan, Italy. Environ. Res. 176, 108520 https://doi.org/10.1016/j.envres.2019.06.001.

Brauer, M., Lencar, C., Tamburic, L., Koehoorn, M., Demers, P.A., Karr, C.J., 2008. A Cohort study of Traffic-Related Air pollution Impacts on birth outcomes. Environ. Health Perspect. 116 (5), 680–686. https://doi.org/10.1289/ehp.10952.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/a:1010933404324.

Chen, J., De Hoogh, K., Gulliver, J.S., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A., Martin, R.M., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Hoek, G., 2019a. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ. Int. 130, 104934 https://doi.org/10.1016/j.envint.2019.104934.

Chen, J., De Hoogh, K., Gulliver, J.S., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Janssen, N.A., Martin, R.M., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Hoek, G., 2019b. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. Environ. Int. 130, 104934 https://doi.org/10.1016/j.envint.2019.104934.

Chen, J., De Hoogh, K., Gulliver, J.S., Hoffmann, B., Hertel, O., Ketzel, M., Weinmayr, G., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Atkinson, R.C., Janssen, N.A., Martin, R.M., Samoli, E., Andersen, Z.J., Oftedal, B., Stafoggia, M., Bellander, T., Strak, M., Hoek, G., 2020. Development of Europe-wide models for particle elemental composition using supervised linear regression and random forest. Sci. Technol. 54 (24), 15698–15709. https://doi.org/10.1021/acs.est.0c06595.

Chen, J., Rodopoulou, S., De Hoogh, K., Strak, M., Andersen, Z.J., Atkinson, R.C., Bauwelinck, M., Bellander, T., Brandt, J., Cesaroni, G., Concin, H., Fecht, D., Forastiere, F., Gulliver, J.S., Hertel, O., Hoffmann, B., Hvidtfeldt, U.A., Janssen, N.A., Jöckel, K., Hoek, G., 2021. Long-term exposure to fine particle elemental components and natural and cause-specific mortality—a pooled analysis of eight European cohorts within the ELAPSE project. Environ. Health Perspect. 129 (4) https://doi.org/10.1289/ehp8368.

De Hoogh, K., Gulliver, J.S., Van Donkelaar, A., Martin, R.M., Marshall, J.D., Bechle, M. J., Cesaroni, G., Pradas, M.C., Dedele, A., Eeftens, M., Forsberg, B., Galassi, C., Heinrich, J., Hoffmann, B., Jacquemin, B., Katsouyanni, K., Korek, M., Künzli, N., Lindley, S., Hoek, G., 2016. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10. https://doi.org/10.1016/j.envres.2016.07.005.

De Hoogh, K., Chen, J., Gulliver, J.S., Hoffmann, B., Hertel, O., Ketzel, M., Bauwelinck, M., Van Donkelaar, A., Hvidtfeldt, U.A., Katsouyanni, K., Klompmaker, J.O., Martin, R.S., Samoli, E., Schwartz, P.E., Stafoggia, M., Bellander, T., Strak, M., Wolf, K., Vienneau, D., Hoek, G., 2018. Spatial PM2.5, NO2, O3 and BC models for western Europe – evaluation of spatiotemporal stability. Environ. Int. 120, 81–92. https://doi.org/10.1016/j.envint.2018.07.036.

De Nazelle, A., Seto, E., Donaire-González, D., Mendez, M.A., Matamala, J., Nieuwenhuijsen, M., Jerrett, M., 2013. Improving estimates of air pollution exposure through ubiquitous sensing technologies. Environ. Pollut. 176, 92–99. https://doi.org/10.1016/j.envpol.2012.12.032.

Dons, E., Van Poppel, M., Kochan, B., Wets, G., Panis, L.I., 2013. Modeling temporal and spatial variability of traffic-related air pollution: hourly land use regression models for black carbon. Atmos. Environ. 74, 237–246. https://doi.org/10.1016/j.atmosenv.2013.03.050.

Eeftens, M., Beelen, R., De Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dèdelè, A., Dons, E., De Nazelle, A., Dimakopoulou, K., Eriksen, K.T., Falq, G., Fischer, P., Galassi, C., Grazuleviciene, R., Heinrich, J., Hoffmann, B., Jerrett, M., Hoek, G., 2012. Development of land use regression models for $PM_{2.5}$, $PM_{2.5}$ absorbance, PM_{10} and PM_{coarse} in 20 European study areas; results of the ESCAPE project. Environ. Sci. Technol. 46 (20), 11195–11205. https://doi.org/10.1021/es301948k.

European Environmental Agency, 2020. AirBase - The European air quality database — European Environment Agency (EEA) europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-8.

Gulliver, J.S., De Hoogh, K., Hansell, A., Vienneau, D., 2013. Development and back-extrapolation of NO_2 land use regression models for historic exposure assessment in great Britain. Environ. Sci. Technol. 47 (14), 7804–7811. https://doi.org/10.1021/es4008849.

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.

Jones, R.R., Hoek, G., Fisher, J.A., Hasheminassab, S., Wang, D., Ward, M.H., Sioutas, C., Vermeulen, R., Silverman, D.T., 2020. Land use regression models for ultrafine particles, fine particles, and black carbon in Southern California. Sci. Total Environ. 699, 134234 https://doi.org/10.1016/j.scitotenv.2019.134234.

Kerckhoffs, J., Hoek, G., Portengen, L., Brunekreef, B., Vermeulen, R., 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., Hoek, G., Gehring, U., Vermeulen, R., 2021. Modelling nationwide spatial variation of ultrafine particles based on mobile monitoring. Environ. Int. 154, 106569 https://doi.org/10.1016/j.envint.2021.106569.

Landrigan, P.J., Fuller, R.A., Acosta, N.J.R., Adeyi, O., Arnold, R.M., Basu, N., Baldé, A. B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breysse, P.N., Chiles, T.C., Mahidol, C., Coll-Seck, A.M., Cropper, M.L., Fobil, J.N., Fuster, V., Greenstone, M., Haines, A., Zhong, M., 2018. The Lancet Commission on pollution and health. Lancet 391 (10119), 462–512. https://doi.org/10.1016/s0140-6736(17)32345-0.

Lu, M., Schmitz, O., De Hoogh, K., Kai, Q., Karssenberg, D., 2020a. 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.

Lu, M., Soenario, I., Helbich, M., Schmitz, O., Hoek, G., Van Der Molen, M.K., Karssenberg, D., 2020b. Land use regression models revealing spatiotemporal co-variation in NO2, NO, and O3 in The Netherlands. Atmos. Environ. 223, 117238 https://doi.org/10.1016/j.atmosenv.2019.117238.

Masiol, M., Zikova, N., Chalupa, D., Rich, D.Q., Ferro, A.R., Hopke, P.K., 2018. Hourly land-use regression models based on low-cost PM monitor data. Environ. Res. 167, 7–14. https://doi.org/10.1016/j.envres.2018.06.052.

MOBI-AIR, 2023. MOBI-AIR. Universiteit Utrecht. https://mobi-air.sites.uu.nl/.

Mölter, A., Lindley, S., De Vocht, F., Simpson, A., Agius, R., 2010. Modelling air pollution for epidemiologic research – Part II: predicting temporal variation through land use regression. Sci. Total Environ. 409 (1), 211–217. https://doi.org/10.1016/j.scitotenv.2010.10.005.

Patton, A.P., Collins, C., Naumova, E.N., Zamore, W., Brugge, D., Durant, J.L., 2014. An hourly regression model for ultrafine particles in a Near-Highway urban area. Environ. Sci. Technol. 48 (6), 3272–3280. https://doi.org/10.1021/es404838k.

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.

Stafoggia, M., Oftedal, B., Chen, J., Rodopoulou, S., Renzi, M., Atkinson, R.C., Bauwelinck, M., Klompmaker, J.O., Mehta, A., Vienneau, D., Andersen, Z.J., Bellander, T., Brandt, J., Cesaroni, G., De Hoogh, K., Fecht, D., Gulliver, J.S., Hertel, O., Hoffmann, B., Janssen, N.A., 2022. Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. Lancet Planet. Health 6 (1), e9–e18. https://doi.org/10.1016/s2542-5196(21)00277-1.

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. Software 133, 104837. https://doi.org/10.1016/j.envsoft.2020.104837.

Weissert, L., Alberti, K.A., Miles, E., Miskell, G., Feenstra, B., Henshaw, G.S., Papapostolou, V., Patel, H., Polidori, A., Salmond, J., Williams, D.R., 2020. Low-cost sensor networks and land-use regression: interpolating nitrogen dioxide concentration at high temporal and spatial resolution in Southern California. Atmos. Environ. 223, 117287 https://doi.org/10.1016/j.atmosenv.2020.117287.

World Health Organization: WHO, 2022. Ambient (outdoor) air pollution. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health?gclid=Cj0KCQjw7PCjBhDwARIsANo7CgknejjPplCOnAsVK23maLqWPZ0hFyywkDr0Dr0vLnyoeSLQMtnTGl4aAubwEALw_wcB.

Wu, J., Wilhelm, M., Chung, J.H., Ritz, B., 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. Environ. Res. 111 (5), 685–692. https://doi.org/10.1016/j.envres.2011.03.008.

Yuan, Z., Shen, Y., Hoek, G., Vermeulen, R., Kerckhoffs, J., 2024. LUR modeling of long-term average hourly concentrations of NO2 using hyperlocal mobile monitoring data. Sci. Total Environ., 171251 https://doi.org/10.1016/j.scitotenv.2024.171251.