Atmospheric Environment 223 (2020) 117238

Contents lists available at ScienceDirect

Atmospheric Environment

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

# Land use regression models revealing spatiotemporal co-variation in $NO_2$, NO, and $O_3$ in the Netherlands

Meng Lu [a,*], Ivan Soenario [a,1], Marco Helbich [b], Oliver Schmitz [a], Gerard Hoek [c], Michiel van der Molen [d], Derek Karssenberg [a]

[a] Department of Physical Geography, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[b] Department of Human Geography and Planning, Faculty of Geosciences, Utrecht University, Utrecht, the Netherlands
[c] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, the Netherlands
[d] Meteorology & Air Quality, Wageningen University, Wageningen, the Netherlands

## HIGHLIGHTS

- High resolution hourly LUR models at different temporal aggregations.
- Spatiotemporal variations and interactions are modelled for NO, NO2, and O3.
- Temporal variations of variable importance for air pollutant contributions.
- Hourly and seasonally varying explained variations of each predictor.
- Validation against an independent data set for $NO_2$.

## ABSTRACT

Land use regression (LUR) modeling has been applied to study the spatiotemporal patterns of air pollution, which when combined with human space-time activity, is important in understanding the health effects of air pollution. However, most of these studies focus either on the temporal or the spatial domain and do not consider the variability in both space and time. A temporally aggregated model does not reflect the temporal variability caused by traffic and atmospheric conditions and leads to inaccurate estimation of personal exposure. Besides, most studies focus on a single air pollutant (e.g., $O_3$, $NO_2$, or NO). These pollutants have a strong interaction due to photochemical processes. For studying relations between spatial and temporal patterns in these pollutants it is preferable to use a uniform data source and modelling approach which makes comparison of pollution surfaces between pollutants more reliable as they are produced with the same methodology. We developed temporal land use regression models of $O_3$, $NO_2$ and NO to study the co-variability of these pollutants and the relations with typical weather conditions over the year. We use hourly concentrations from the measurement network of the Dutch National Institute for Public Health and the Environment and aggregate them by hour, for weekday/weekend and month, and fit a regression model for each hour of the day. 70 candidate predictors that are known to have a strong relationship with combustion-related emissions are evaluated in the LUR modelling process. For all pollutants, the optimal LUR was identified with 4 predictors and the temporal variability was determined by the explained variance of each temporal model. Our temporal models for $O_3$, $NO_2$, and NO strongly reflect the photochemical processes in space and time. $O_3$ shows a high background value throughout the day and only dips in the (close) vicinity of roads. The diminishing rate is affected by traffic intensity. The $NO_2$ LUR is validated against $NO_2$ measurements from the Traffic-Related Air pollution and Children's respiratory HEalth and Allergies (TRACHEA) study, resulting in an $R^2$ of 0.61.

## 1. Introduction

Air pollution such as Nitrogen Oxides ($NO_2$ and $NO_x$)(Beelen et al., 2008) is associated with adverse effects on human health (Eeftens et al., 2012; Hoek et al., 2013). To study this relation, it is often required to quantify long-term personal exposures for a large number of people (Fischer et al., 2015; Cohen et al., 2009). This can be done by combining personal location information over time with spatiotemporal air pollution concentrations. To assess people's air pollution exposures along their daily path, spatially and temporally fine-grained air pollution data for a large spatial coverage (e.g., a country) are needed. Studies assessing air pollution exposure considering space-time paths of

* Corresponding author.
*E-mail address:* m.lu@uu.nl (M. Lu).
[1] Equal contribution

https://doi.org/10.1016/j.atmosenv.2019.117238
Received 18 July 2019; Received in revised form 13 December 2019; Accepted 18 December 2019
Available online 23 December 2019

individuals and spatiotemporal variation in air pollution have indicated considerable differences in calculated exposures compared to exposure assessments at the residential location (Park and Kwan, 2017; Gurram et al., 2015; Baxter et al., 2013).

The physical and statistical approaches have been developed to model spatiotemporal air pollution. The physical models attempt to describe atmospheric concentrations in a controlled volume as a function of local emission, horizontal and vertical mixing and chemistry. Coupled atmosphere – air quality models (EPA, 2019; de Hoogh et al., 2014) typically have a spatial resolution larger than 100 m and a temporal resolution of minutes. This approach has clear physical meanings. The level of accuracy mainly depends on how well the processes are described and the quality of the identification of the associated parameters. The main limitation of this approach is that the most up-to-date traffic exhaust list is needed and a higher than 100 m resolution product requires more sophisticated process descriptions, resulting in models becoming intractable and slow, the latter being a problem when large areas are considered. The statistical approach relates observed concentrations to environmental predictors of air pollution including road infrastructure, traffic loads, population, climate, topography, industry, land cover. Geostatistical and Land Use Regression (LUR, Hoek et al., 2008; Briggs et al., 2000) modelling, are the most commonly used statistical approaches. Detailed air quality maps could be obtained if the variables associated with the emitting source and clearance are available at high resolution. They are low in computation cost compared to the physical approaches and allow applications over larger areas. A disadvantage of statistical models is their lack of physical explanations as the physical process of emission dispersion is not simulated (von Klot, 2011). The major assumption of LUR models is that traffic-related air pollution decreases with distance to the source, usually roads. Air quality monitoring stations are preferably distributed at different distances from the roads.

As air pollution varies considerably over time (e.g., between seasons), an annual estimation of air pollution, as has been done in most of the LUR studies (Eeftens et al., 2012), is unable to provide sufficient information to assess exposure when the space-time activities are considered. A relatively small number of studies developed models considering temporal variability of air pollutant concentration. Boniardi et al. (2019) and Cordioli et al. (2017) studying seasonal LUR models of Black carbon, $PM_x$, and common gaseous pollutants found considerable differences between warm and cold seasons. Rahman et al. (2017) fit a LUR model to the residuals of a periodic model of the day of the year and the day of the week to predict daily average $NO_x$ concentrations. Dons et al. (2013) created a LUR model representing the diurnal cycle (hourly resolution) for Black Carbon. They developed different LUR models for each hour of the day, which has important implications in exposure studies.

The aim of this study is to develop a LUR-based approach to describe the spatiotemporal patterns of $O_3$, $NO_2$, and NO and their interactions in the Netherlands. We focus on what we refer to as air pollution "climate" – averaged air pollution over multiple years. The following key questions will be answered: 1) What is the diurnal cycle in the LUR model parameters and modelled air pollution climate at different levels of temporal aggregation? 2) How do vehicle emission, weather, and photochemical processes jointly affect air pollution? We used air pollution observed at 44 stations of the Dutch Air Quality Monitoring Network to identify LUR models describing air pollution climate for each pollutant. These stations were distributed over urban, rural, industrial areas, and positioned at various distances away from roads and thus capture spatial variation in air pollution across multiple environments and spatial scales.

## 2. Methods

### 2.1. Measurement data

We used hourly NO, $NO_2$, and $O_3$ measurement data from the Dutch Air Quality Monitoring Network, maintained by the National Institute for Public Health and the Environment (Rijksinstituut voor Volksgezondheid en Milieu, Landelijk Meetnet Luchtkwaliteit. Acronym: RIVM-LML) (National Institute for Public Health and the Environment, 2017). There are four types of measurement locations: urban, rural, traffic and industry. The geographical distribution of the different types of the stations is shown in supplement Fig. 1. NO and $NO_2$ were measured with the chemiluminescence method, using the Teledyne API M200E monitor (expanded uncertainty 8.3%). $O_3$ was measured with the UV absorption technique, using the Thermo Electron Instruments 49 W device (expanded uncertainty 7.0%).

We averaged for each hour, month, and weekend/weekday over RIVM-LML data from a period of five years, from 01 July 2006 to 01 July 2011. We chose five years because a longer period would result in a considerably smaller number of observational locations, as the number of measurements from the sensor network is not consistent over time. In addition, using this period of time provides an almost exact overlap with the time period of independent measurements from the TRACHEA study of Eeftens et al. (2011), which are used to validate our results.

The RIVM-LML started measuring air pollution variables in the 1980s. Around 50 measurement locations were active during the selected five year period. Stations with less than 20% data were left out of the analysis. Our analysis was executed using data from 44 stations. A map of the stations that are used in our study is given in Figure (supplement Fig. 2). The distribution of missing observations in these stations over time is shown in Figure (supplement Fig. 3). The monthly values of the $NO_2$, NO, and $O_3$ are shown in Figure (supplement Fig. 4).

### 2.2. Potential predictors

The potential predictors that are used in the LUR modelling process are based on environmental attributes such as traffic, land use, and population density (Table 1). These correspond to those defined by the ESCAPE project (Eeftens et al., 2012; Beelen et al., 2013). Data on traffic infrastructure was derived from the TeleAtlas MultiNet dataset for the year 2008. Traffic intensities for passenger and heavy traffic were provided by the Netherlands Environmental Assessment Agency (PBL, 2018). All of the predictors except population were derived from vector data (Schmitz et al., 2019). Land use data was derived from the CORINE dataset (Copernicus, 2018). Population data was from the INTARESE project dataset (INTARESE, 2018), which consists of 100 m resolution grids and these grids were resampled to 5 m resolution (Schmitz et al., 2019).

All source datasets were projected using a local projection of the Netherlands, the "Amersfoort/RD New" projection. The datasets were converted to raster datasets covering the entire land surface of the Netherlands at 5 m resolution. To account for large-scale effects and known spatial dispersion patterns of traffic-related pollutants (see, e.g., Hoek et al. (2008)), most of the variables were aggregated over buffers with radius of 25 m, 50 m, 100 m, 300 m, 500 m, 1000 m, and 5000 m for each air pollution monitoring location. We selected these distances as previous LUR modelling studies (in particular (Eeftens et al., 2012; Beelen et al., 2013)) of combustion related air pollutants for the Netherlands have shown these distances to be relevant in modelling. In addition, a distance to road predictor was created by calculating the distance from each cell-centre coordinate to the closest road segment. In total, 70 candidate predictor variables were calculated. Spatial data processing was done with PCRaster Python (Karssenberg et al., 2010) and the GDAL/OGR library (GDAL Development Team, 2018). Schmitz et al. (2019) gives technical details of the calculation of the predictor variables.

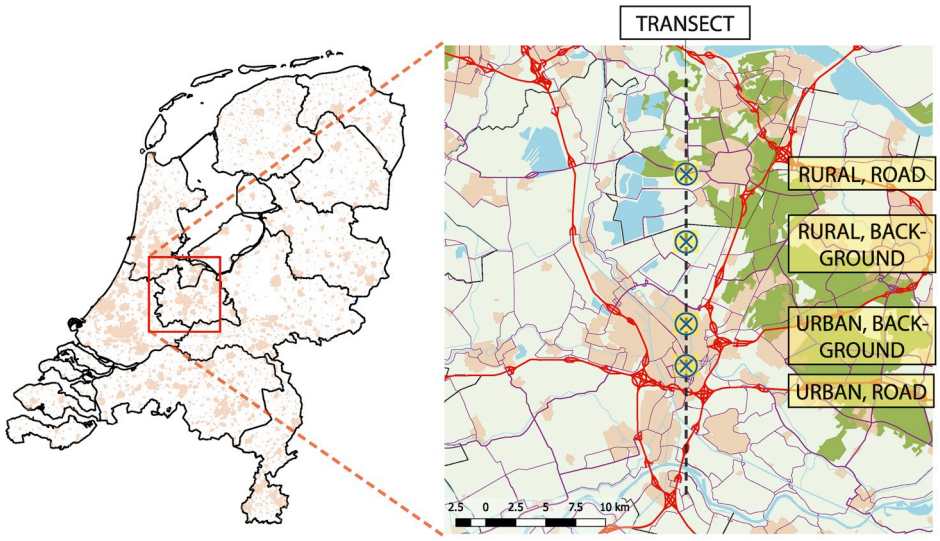


**Fig. 1.** Left: Study area, the Netherlands, the square indicates the area around the city of Utrecht used to illustrate the results. Right: transect crossing the city of Utrecht with four example locations. The Rose patches indicate urban area; dark green, forest; light green, arable land; blue, open water. The red lines indicate highways and purple lines other roads. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

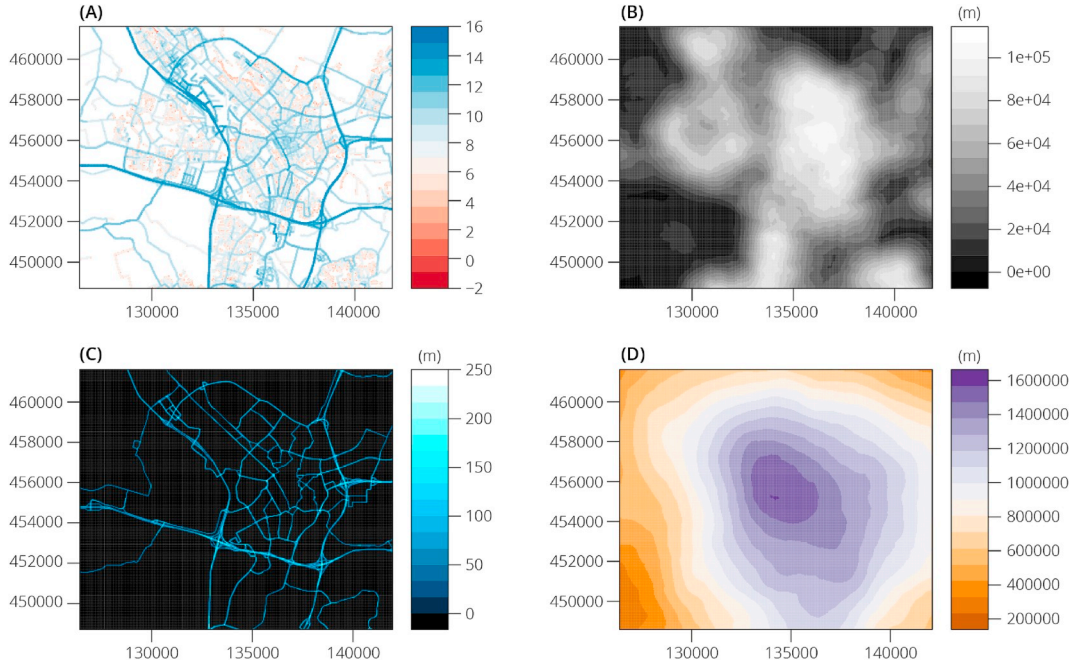


**Fig. 2.** The four predictors that are selected by the variable selection process for the $NO_2$ LUR model, displayed for the Utrecht area. A: Heavy traffic load within 50 m buffer. B: Major road length within 1000 m buffer. C: Major road length within 25 m buffer. D: Major road length within 5000 m buffer.

## 3. Regression model

### 3.1. Framework

The relationship between pollutant concentration and a set of predictors can be described statistically with a multiple linear regression:

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + c \quad (1)$$

In Equation (1), $y$ is the pollutant concentration ($\mu g/m^3$) and $x_i$, $i = 1, 2, .., n$, are the $n$ predictors. For each temporal model, the same set of predictors are used:

$$y_t = \beta_{1,t} x_1 + \beta_{2,t} x_2 + \ldots + \beta_{n,t} x_n + c \quad (2)$$

In Equation (2), $y_t$ is the five year average air pollution at a particular time $t$, which is the time of the day (h), in a particular month, at a weekday or weekend. Thus, 24 (hours) $\times$ 12 (months) $\times$ 2 (weekdays/weekends) = 576 regression equations are built.

Our air pollutant climates are built in two steps. First, we select predictor variables from the set of candidate predictors. Secondly, each temporal model is fitted using the selected predictors.

### 3.2. Predictor selection

To select candidate predictors, we fit the coefficients to the average measured air pollution per station over the given five-year time period (Equation (1)). As we have more potential predictors than observations, we used a shrinkage based method (James et al., 2013), elastic net (Zou

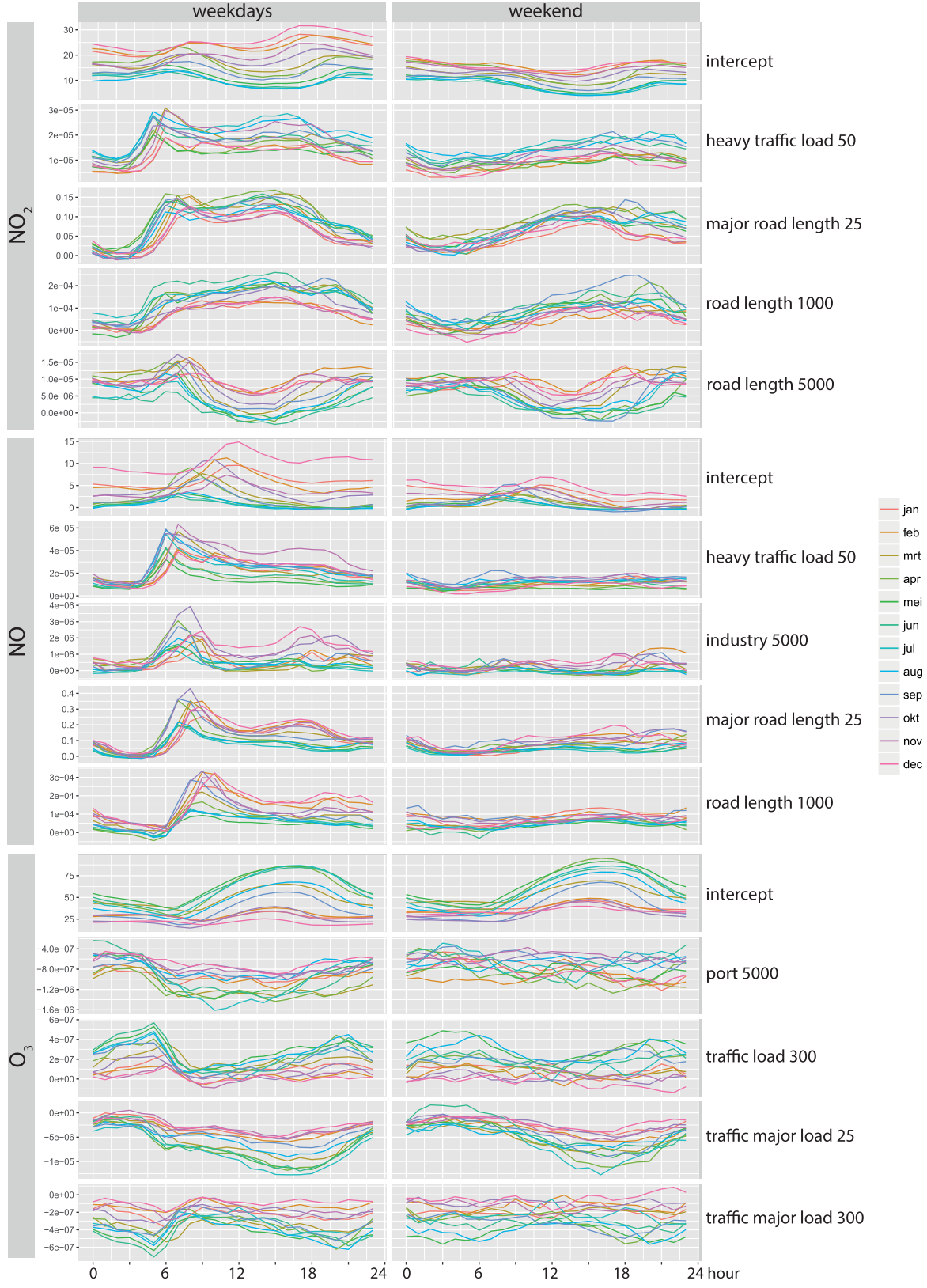


**Fig. 3.** Coefficients for the LUR models of $NO_2$ (top), NO (centre) and $O_3$ (bottom). Left, weekdays; right, weekend days. Each line represents a month and connects coefficient values calculated for each hour. The x-axis represents time of the day (hours).

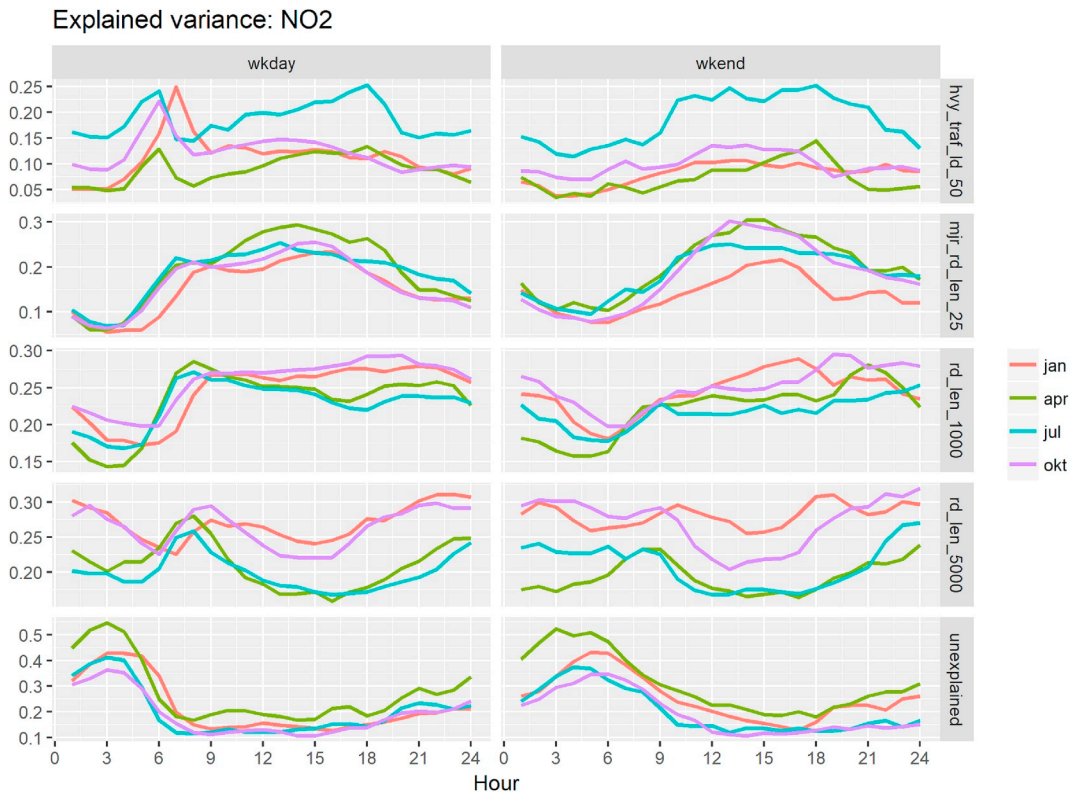


**Fig. 4.** Time series of explained variance in each of the predictors, separating between weekdays (wkday) and weekend (wkend), and the unexplained variables for $NO_2$, the time series are aggregated to jan (Jan.–Mar.), apr (Apr.–June), jul (July–Sep.), and okt (Okt. - Dec.). hvy_traf_ld_50: heavy traffic loads in 50 m buffer, mjr_rd_len_25: major road lengths in 25 m buffer, rd_len_1000: road lengths in 1000 m buffer, rd_len_5000 road lengths in 5000 m buffer.

**Table 1**
Candidate predictor variables. The value of each predictor is calculated within 7 different buffer sizes.

| Candidate predictor | Name | Unit | Source data |
|---|---|---|---|
| Total length of major roads | Major Road Length | m | Eeftens et al. (2012) |
| Total length of all roads | Road Length | m | Eeftens et al. (2012) |
| Area of residential land | Residential | $M^2$ | Copernicus (2018) |
| Surface area of semi-natural and forested areas | Natural | $m^2$ | Eeftens et al. (2012) |
| Area of industry | Industry | $m^2$ | Copernicus (2018) |
| Port area | Port | $m^2$ | Copernicus (2018) |
| Total heavy-duty traffic load of roads | Heavy Traffic Load | veh. $day^{-1}$ m | PBL (2018) |
| Total traffic load of roads | Traffic Load | veh. $day^{-1}$ m | PBL (2018) |
| Total traffic load of major roads | Traffic Major Load | veh. $day^{-1}$ m | PBL (2018) |
| Number of inhabitants | Pop | n | INTARESE (2018) |

and Hastie, 2005), to reduce the number of predictors. Elastic net can be seen as a combination of the lasso and ridge methods; it, therefore, maintains the features of both methods and fits the model more flexibly. We set the elastic net mixing parameter to 0.5, indicating selecting a penalty between lasso and ridge. The response type is set to the Gaussian, and the elastic net sequence of the model is fit by coordinate descent (Friedman et al., 2010a), as is implemented in (Friedman et al., 2010b). The response is standardized (i.e. have unit variance) before computing the sequence.

Applying the elastic net alone leads to 20, 17 and 11 predictors for $NO_2$, NO and $O_3$, respectively. This large number of predictors may still conceal the relative importance of predictors, cause overfitting, and result in low model interpretability. It is possible to set the elastic net to select fewer variables at a sacrifice of prediction accuracy, but the shrinkage is only based on the prediction accuracy and not all possible models from a specific set of predictor candidates are evaluated. Therefore, we used the Best Subset Selection (BSS, James et al., 2013) to further reduce the number of predictors by evaluating all possible models with elastic net selected variables. The criterion used to select variables is based on the $R^2$, BIC (Bayesian Information Criterion), and Mallows's $C_p$.

For implementation, we used the elastic net from the R glmnet package (Friedman et al., 2010b), and the BSS from the R leaps package (Lumley and Lumley, 2013).

### 3.3. Fitting temporal models

After the candidate predictors are selected, regression models are fitted to the observations separately for each hour of the day, for each month, and for weekdays or weekend days. We use the predictors identified using the 5-year average air pollution in the temporal models. With this approach, it comes with the advantage that the time series of regression coefficients can be directly interpreted as temporal variation in mechanisms of air pollution explained by the selected predictors. For instance, one would expect higher coefficient values for traffic-related predictors during rush hours.

The station measurements were temporally aggregated into hourly averages representing weekdays and weekends of each month of the 5 years. Before fitting the coefficients, the measurement data (43800 measurements per location) are averaged within each of the 576 groups created by crossing the temporal variables. We used the adjusted $R^2$ as the indicator of the goodness-of-fit of the final temporal LUR models.

To study the contribution of each of the predictors in explaining the

total variation in the dataset, we calculated the relative variable importance of the predictors following the lmg method (Lideman et al., 1980), which decomposes the R square using the ANOVA test. The lmg method is implemented in the R (R Core Team, 2017) package relaimpo (Grömping et al., 2006). This was done for each regression (576 models).

### 3.4. Validation

The RMSE (Root Mean Squared Error) of the Leave-one-out Cross-Validation (LOOCV) and the R-squared are used to evaluate the hourly models of $NO_2$, NO, and $O_3$ ($\mu g/m^3$).

In addition, we used an external validation dataset, the TRACHEA measurements (Traffic-related Air pollution and Children's respiratory HEalth and Allergies) to further validate the $NO_2$ models. In TRACHEA, average concentrations of $NO_2$ were measured for four week-long periods in four different seasons of 2007 (Eeftens et al., 2011). We predicted $NO_2$ over the study period using our temporal LUR models at each location of the TRACHEA stations and averaged the predictions to enable comparison with TRACHEA. For this comparison, the TRACHEA measurements were averaged over the four measurement periods described in Eeftens et al. (2011).

## 4. Results

### 4.1. Predictor selection

Four predictors were selected using the BSS for all the pollutants. The optimal number of variables of the BSS variable selection criterion is shown in Table 2. For NO, all the BSS variable selection criteria indicate 4 variables is optimal. For $NO_2$, there is a small increase in explained variation (Adjusted $R^2$) using 9 variables comparing to using 4 variables; however, the BIC and Mallow's Cp increased considerably from using 3 variables to 9 variables. For $O_3$, the BIC and Mallow's Cp are the lowest with 3 parameters and increased slightly with an additional variable. Also the adjusted $R^2$ shows considerable improvement with 4 parameters. The value of the criteria corresponding to the optimal model of the different number of predictors is shown in supplement Fig. 5.

As shown in Table 3, the LUR models of $NO_2$ and NO both include road length within the 1000 m buffer, major road length within the 25 m buffer and heavy traffic load within the 50 m buffer. $NO_2$ uses entirely road and traffic-related predictors. NO is also related to the industry. $O_3$ is related to a different set of traffic-related predictors, mainly the traffic load, but the port as well.

We use the area around the city of Utrecht (population ca. 345,000) to illustrate the results (Fig. 1). The predictors in the $NO_2$ model for the Utrecht area are given in Fig. 2. The $NO_2$ model includes predictors with small and larger buffer sizes. A variable with a large buffer size represents a smooth surface and thus larger scale variation (e.g., road length within 5000 m buffer), whereas variables with a small buffer size more closely represent the road patterns.

### 4.2. Temporal models

Fig. 3 shows the coefficients of the LUR models for NO, $NO_2$, and $O_3$ fitted by using the set of predictor variables selected in the previous step (Table 3). In our LUR models, air pollution is expressed as a linear combination of these predictors (Equation (2)). The coefficients, p-values, and the adjusted R-Squared for each model are provided at https://github.com/pcraster/gghdc-spatio-temporal-lur-nl/. For all pollutants and most predictors, relatively smooth diurnal patterns in the predictors can be observed, and the pattern appears consistently over months. This may suggest relatively small random errors in the identification of the values for the coefficients. Exceptions are, for instance, the coefficients related to the variables port within the 5000 m buffer and traffic major load within the 300 m buffer, which contain inconsistent variation between hours which seems random, that is, due to the uncertainty in the fitting of the coefficients. The coefficients for NO show different diurnal variation but with a comparable smooth trend over the day.

Although the values of the coefficients give an indication of the importance of each of the predictors, care should be taken by comparing them as the units of predictors are different and absolute values may not always be indicative for the variation explained by the predictor. Instead, the relative importance values (Figs. 4–6) can be interpreted as the proportion of the total variation in the data set explained by each predictor. A high importance value implies that the spatial pattern of that predictor variable is a large component of the total spatial variation in air pollution. For $NO_2$, the order of magnitude of the importance values of each of the predictors is the same (their values are around 0.25), which indicates that the patterns represented by each of the predictors (Fig. 2) are all clearly present in the spatial pattern of the $NO_2$ air pollution. Differences in the importance values occur between seasons, but they do not show an obvious trend. An obvious diurnal pattern occurs, however, in the importance values of the $NO_2$ regressions. During weekdays, a sharp increase in the importance value of short-range buffers occurs between 4 a.m. and 7 a.m., with a similar increase of the long-range buffers occurring slightly later. This could indicate that in the morning and close to roads (short-range buffers), air pollution rises first and disperses gradually, resulting in high importance values for the long-range buffers as well, but slightly later in the morning. During weekends, the rapid increase of the importance values early in the morning is replaced by a gradual increase over the day. Both in weekdays and weekend, the importance values of the short-range buffers tend to decrease in the evening while those of the long-range buffers increase, indicating that peaks close to roads disappear and large scale patterns become prevalent. While during the day the unexplained variance ($1 - R^2$) is relatively low (below 0.2), it increases considerably during the night, indicating that the occurring spatial patterns are not represented by the predictor variables used. The magnitude and temporal patterns of the regression coefficients and importance values of NO are comparable to those of $NO_2$ with the main difference that an Industry predictor is included.

Just like $NO_2$ and NO, the spatial variation in $O_3$ is represented by short and long-range road buffers, but the temporal pattern is different. The road related buffers show a sinusoidal shape diurnal pattern in importance values with the highest values just after the noon time. The port 5000 m predictor is more constant over the day. Between seasons, the importance values are similar, but in spring and summer, importance values of the 25 m buffer for the traffic major road are considerably higher compared to the autumn and winter values. At night, the largest part of the variation remains unexplained by the predictors, but note that total spatial variation during the night is considerably lower as well (supplement page S3–S14).

### 4.3. Predicted air pollutant concentration

Considerable variation in air pollution concentration occurs for all pollutants considered (Figs. 7–9), with spatial and temporal variation components that are comparable in magnitude. The $NO_2$ concentrations show a pronounced spatial variation, with concentrations that are approximately two times higher in urban areas compared to rural areas, and an increase in concentration values towards roads up to about twice

**Table 2**
Optimal number of predictors of different selection criterion in Best Subset Selection.

| | NO | $NO_2$ | $O_3$ |
|---|---|---|---|
| Adj $R^2$ | 4 | 9 | 4 |
| BIC | 4 | 3 | 2 |
| Mallows Cp | 4 | 4 | 2 |

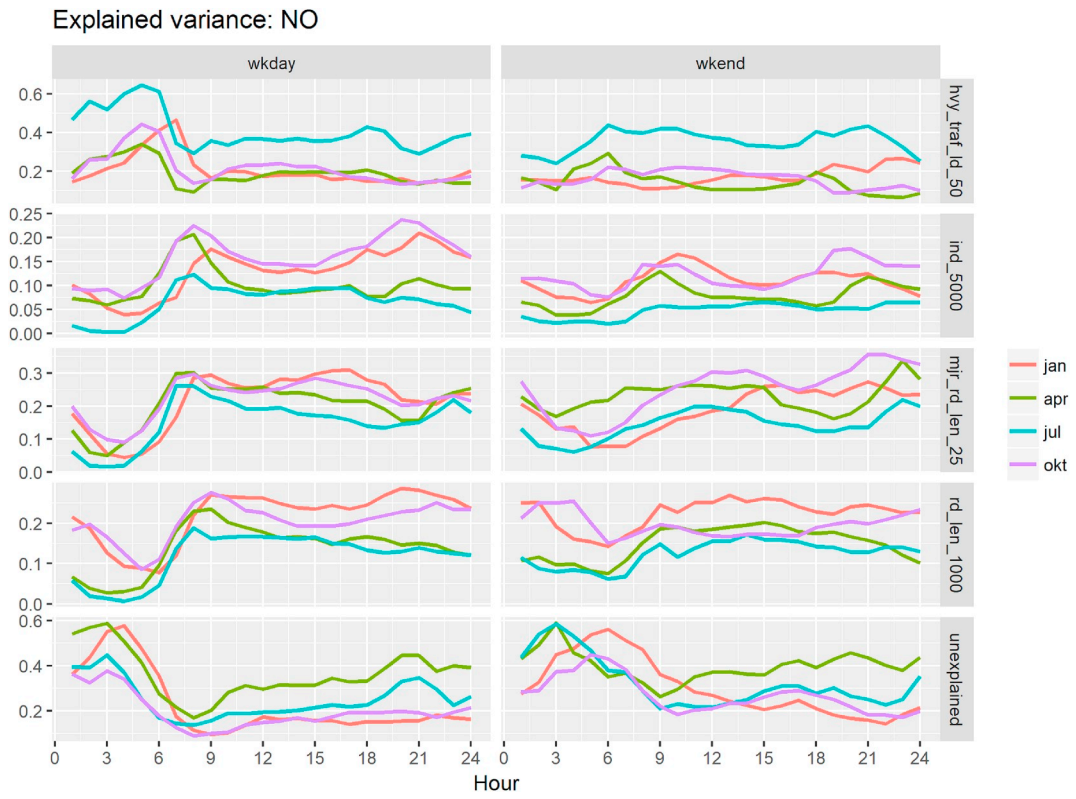


**Fig. 5.** Time series of explained variance in each of the predictors, separating between weekdays (wkday) and weekend (wkend), and the unexplained variables for NO, the time series are aggregated to jan (Jan.–Mar.), apr (Apr.–June), jul (July–Sep.), and okt (Okt. - Dec.). hvy_traf_ld_50: heavy traffic loads in 50 m buffer, ind_5000: industrial area in 5000 m buffer, mjr_rd_len_25: major road lengths in 25 m buffer, rd_len_1000: road lengths in 1000 m buffer.

**Table 3**
Selected LUR model predictors and coefficients of linear regression models fitting the 5-year averages of $NO_2$, NO, and $O_3$ respectively to the selected predictors.

| Predictor | buffer (m) | $NO_2$ | NO | $O_3$ |
|---|---|---|---|---|
| Road Length | 1000 | $1.09 \times 10^{-4}$ | $1.79 \times 10^{-4}$ | |
| Road Length | 5000 | $6.50 \times 10^{-6}$ | | |
| Major Road Length | 25 | $7.66 \times 10^{-2}$ | $6.96 \times 10^{-2}$ | |
| Traffic Load | 300 | | | $1.44 \times 10^{-7}$ |
| Heavy Traffic Load | 50 | $1.47 \times 10^{-5}$ | $1.59 \times 10^{-5}$ | |
| Traffic Major Load | 25 | | | $7.23 \times 10^{-6}$ |
| Traffic Major Load | 300 | | | $4.04 \times 10^{-8}$ |
| Industry | 5000 | | $2.31 \times 10^{-7}$ | |
| Port | 5000 | | | $1.21 \times 10^{-6}$ |

the concentration away from the roads. $NO_2$ concentrations are generally lowest during the night, with during weekdays a relatively steep rise in concentration at 7 a.m. resulting in a peak at around 8 a.m. A second, lower peak occurs at 6 p.m. Fig. 11 shows that the temporal variability of our estimations is in general higher close to roads. On weekend days, concentrations are lower and gradually increase over the day. There is also a seasonal variation component in $NO_2$ with concentrations that in winter time up to two times those during summer (Figs. 8 and 9). NO follows a pattern similar to that of $NO_2$, however with a more pronounced diurnal and seasonal pattern. Both in space and in time, $O_3$ concentrations tend to show a pattern(Fig. 10) that is the opposite of $NO_2$ or NO, with a marked reduction (up to three times lower) in concentration values close to roads (Figs. 7 and 10). The difference in $O_3$ between weekend and weekdays as well as the variation over the day is relatively small, although a distinct diurnal pattern in $O_3$ is apparent in summer, where afternoon values rise everywhere except in city centres near roads. A strong seasonal trend occurs (Fig. 9)), with an increase in spring up to May, when concentrations are approximately three times of those during the winter.

Fig. 12 shows the relationships between predicted $NO_2$, NO, the ratio between NO and $NO_2$ ($NO/NO_2$), and $O_3$ and their distributions in the N–S transect across Utrecht in January (Fig. 7). The $NO/NO_2$ should be theoretically directly related to the volume $O_3$, but from the scatterplot and the correlation the relationship with $O_3$ is similar in NO, $NO_2$ and $NO/NO_2$, with the NO the most correlated with the $O_3$. $O_3$ and $NO_x$ are negatively correlated, and the NO and $NO_2$ are highly positively correlated.

### 4.4. Validation

The RMSE (Root Mean Squared Error) of the leave-one-out cross validation (LOOCV) is used to evaluate the hourly models of $NO_2$, NO, and $O_3$ ($\mu g/m^3$). The RMSE that represent different seasons and hours, namely the January, April, June, October at respectively 05:00 a.m., 09:00 a.m., 14:00 p.m., 18:00 p.m., for $NO_2$, NO, and $O_3$ ($\mu g/m^3$) are shown in supplement Table 1.

The scatter plot between our prediction and the TRACHEA measurement is shown in (Fig. 13). Our model underestimates at high $NO_2$ concentration and overestimates at low $NO_2$ concentration. A linear regression fit between our model predictions and the TRACHEA measurements results in an $R^2$ of 0.61.

## 5. Discussion

We developed hourly LUR models at different levels of temporal aggregation for $NO_2$, NO and $O_3$. Our LUR models allow us to identify how these air pollutants vary spatiotemporally and how the relations with the predictors change over time, which can possibly be explained

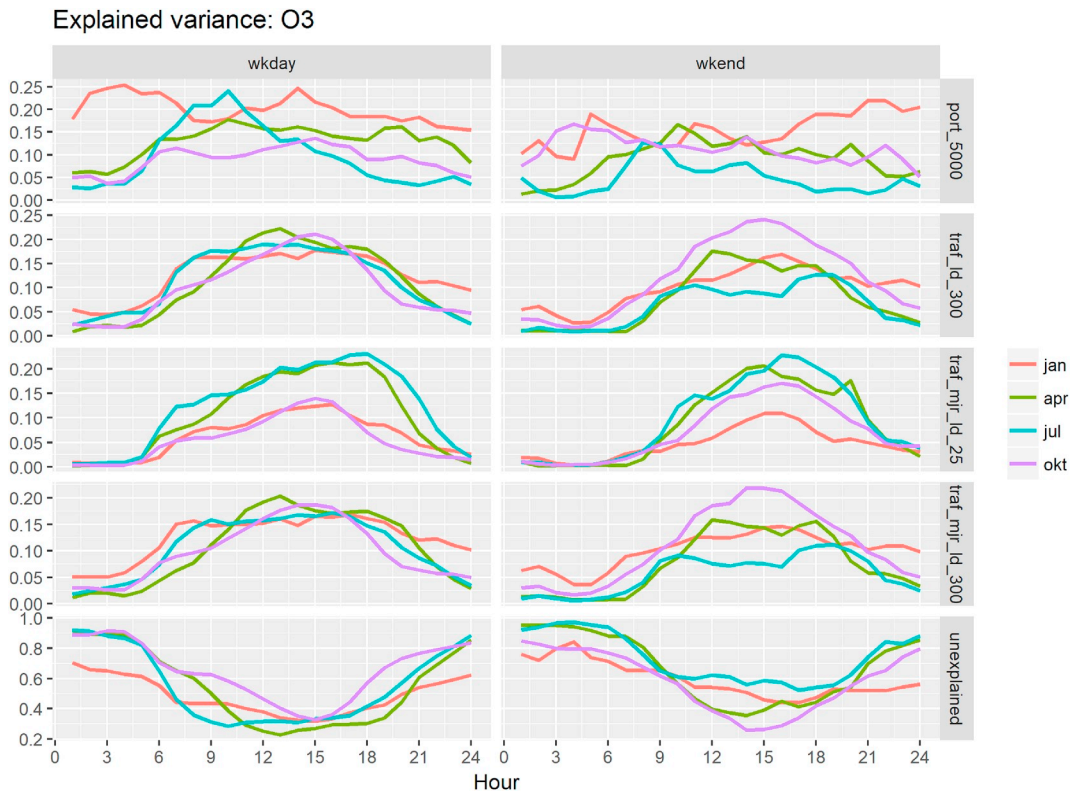


**Fig. 6.** Time series of explained variance in each of the predictors, separating between weekdays (wkday) and weekend (wkend), and the unexplained variables for $O_3$, the time series are aggregated to jan (Jan.–Mar.), apr (Apr.–June), jul (July–Sep.), and okt (Okt. - Dec.). port_5000: port area in 5000 m buffer, traf_ld_300: traffic loads in 300 m buffer, traf_mjr_ld_25: traffic loads on major roads in 25 m buffer, traf_mjr_ld_300: traffic loads on major roads in 300 m buffer.

by mechanisms that steer spatiotemporal variations in air pollution. We found that both spatial variation and diurnal variation in air pollution concentration varies between seasons, and between weekdays and weekends. $NO_2$ and NO show a comparable spatiotemporal pattern, while $O_3$, in general, follows a pattern that is the opposite of these two pollutants. In this section, we discuss the spatial and temporal patterns that are modelled, the strengths and limitations of our modelling process, and directions for future research.

### 5.1. *Spatial and temporal patterns*

The spatial and temporal patterns observed here can be explained by patterns in combustion emissions, climate, and photochemical reactions. Spatiotemporal patterns in combustion emission seem to cause a large proportion of the variation in air pollution in our study area. Diurnal variation in air pollution is clearly related to the traffic intensity, in particular for $NO_2$ and NO. During weekdays, air pollution concentrations peak during hours that coincide with rush hours in the morning and late afternoon. These peaks in air pollution concentration are less pronounced or absent during weekends, which is explained by the more uniform distribution of traffic over the day during weekdays. Spatial patterns in air pollution are strongly related to the road network as well, which has been shown in many other LUR studies. From our descriptive approach, it is hard to disentangle the effect of atmospheric processes from other mechanisms causing spatio-temporal variation in air pollution. It is obvious however that photochemical reactions are important determinants of air pollution levels, in particular for $O_3$. The chemical coupling of the pollutants considered here generally causes $O_3$ levels to decrease when $NO/NO_2$ ratios increase. Also, at higher air temperatures $O_3$ concentrations tend to be high as well. The former mechanism causes $O_3$ concentrations to decrease towards roads, while the latter mechanism results in a seasonal variation in $O_3$ with highest concentrations during spring and summer. Similar patterns have been described and observed by Han et al. (2011); Hagenbjörk et al. (2017).

The spatial pattern in air pollution changes over time, as shown by the considerable temporal variation in the values of the regression coefficients and importance values. For instance, at night, spatial patterns in air pollution tend to be more smooth, while during the day, patterns show smaller scale variation more closely related to the road pattern. This implies that it is important to incorporate changes in the spatial pattern of air pollution over time in statistical models of air pollution. Approaches that rely on a constant spatial pattern (e.g., Cordioli et al. (2017)) may oversimplify the spatial and temporal structure of the variation in air pollutants.

### 5.2. *Strengths and limitations of our method*

The main strengths of our modelling approach are the inclusion of temporal variability in LUR modelling, the evaluation of the changes in the predictor variable contributions over time, and the relative simplicity of the modelling approach which eases interpretation and, due to the short run time of our models, allows ad-hoc calculation of human exposures integrated over the activity domain of individual persons. The relative simplicity of the modelling is however also a limitation as our approach does not integrate spatial and temporal correlations, neither does it incorporate photochemical reactions between pollutants. Another limitation is the lack of air pollution stations very close to roads (25 m or closer) which made it challenging to model air pollution in areas close to roads. Finally, the model performance at nighttime was relatively low compared to the daytime. Below we discuss these strengths and limitations in more detail and provide suggestions for future research.

#### 5.2.1. *Temporal variability*

Our study has important implications for long-term exposure assessment. Thus far, long-term exposure has mainly been assessed using

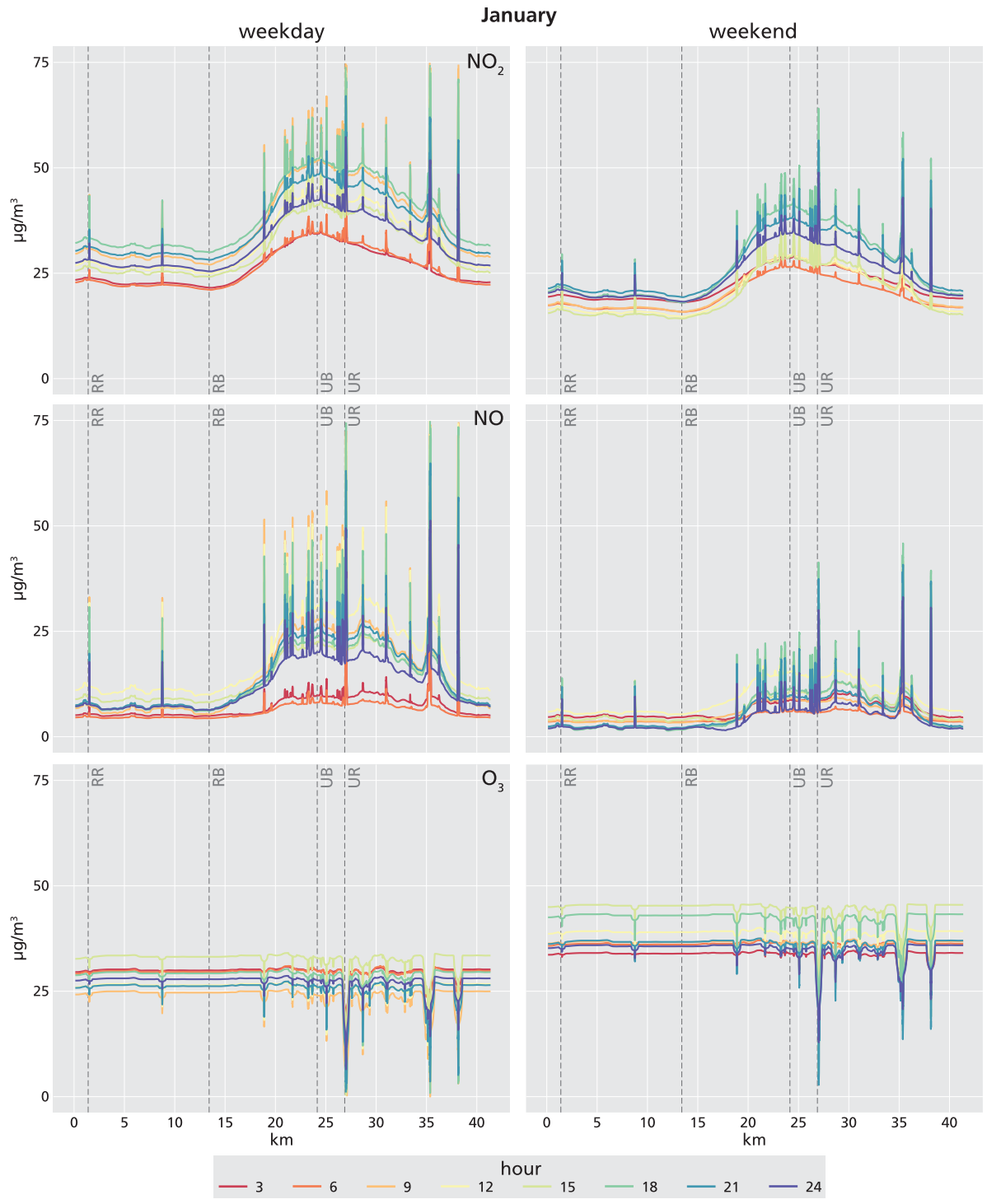


**Fig. 7.** Modelled $NO_2$ concentrations for each hour of the day, on a North-South transect across the province of Utrecht (Fig. 1), intersecting the country side (0–17 km along x-axis) and the city (17–42 km along x-axis), for January. The spatial pattern for January is comparable to that of other months. For a clearer visualization, every third hour is printed, and the pollutant concentration of the Y-axis is clipped at 75 ($\mu g/m^3$, the original maximum value is 130 $\mu g/m^3$). The dashed lines indicate the locations (also given in Fig. 1) for which time series are plotted in Fig. 8, RR: rural near road, RB: rural background, UR: urban near road, UB: urban background.

one temporally aggregated air pollution weighted by population density. Our study shows that temporal variability in air pollution can be considerable. The spatiotemporal predictions of air pollution presented here enable more sophisticated exposure assessment taking human space-time activity into account, as shown for example in Lu et al. (2019).

#### 5.2.2. *Variable influence over time*

Our statistical setup using a single set of predictor variables for all time steps allowed us to analyse the temporal variation in variable influence. The variable influence is calculated to study the impact of a predictor on the concentration of a pollutant instead of the model coefficients because the model coefficients are fit to predictors of different

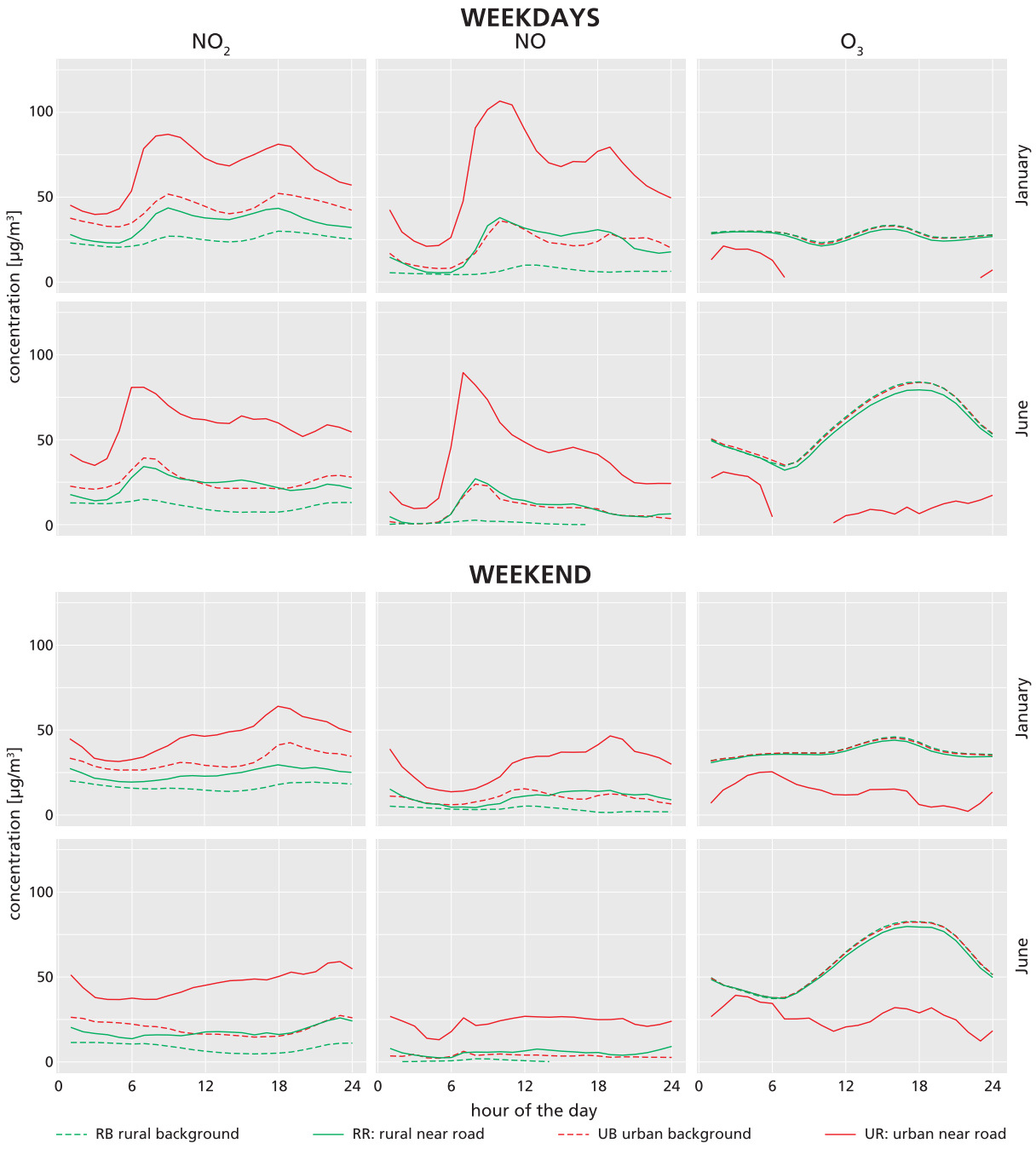


**Fig. 8.** Time series for the three air pollutants at four locations along a N–S transect as illustrated in 7, on weekdays, in January and June, and for weekdays and weekend. January and June are used as they represent the extremes in the yearly cycle of the air pollution climate.

scales and lack physical meaning. Future studies may need to use a different set of predictors for day and night and include atmospheric variables (Bertazzon et al., 2015) and temporally varying predictor variables (Son et al., 2018). Station measurements from nearby countries (e.g., Germany) and mobile sensor measurements may be used to better identify a more flexible model. With sufficient data, buffer sizes could be selected (Grant et al., 2015) using a data-driven approach and could reflect the variation of vehicle emission over time.

### 5.2.3. *Nighttime model performance*

Our models show less satisfactory performance during nighttime, with lower values of explained variation compared to daytime, although cross-validation results did not show a considerable increase in errors during nighttime. A similar problem with LUR modelling of night time concentrations was encountered in the study by van Donkelaar et al. (2015). A possible cause may be that nighttime variation in space is less pronounced compared to day time spatial variation and thus harder to explain in a statistical model. It might also be that the predictors used in our study do not represent spatial patterns in air pollution occurring during the night. Emissions, being lower at night time, become a less important determinant of air pollution patterns during the night and other spatial patterns, related to dispersion, may become more important. Future studies may need to evaluate other predictor variables, including possibly those that relate to spatial patterns in photochemical reactions (e.g., temperature) or atmospheric variables (e.g., roughness, distance to the coast).

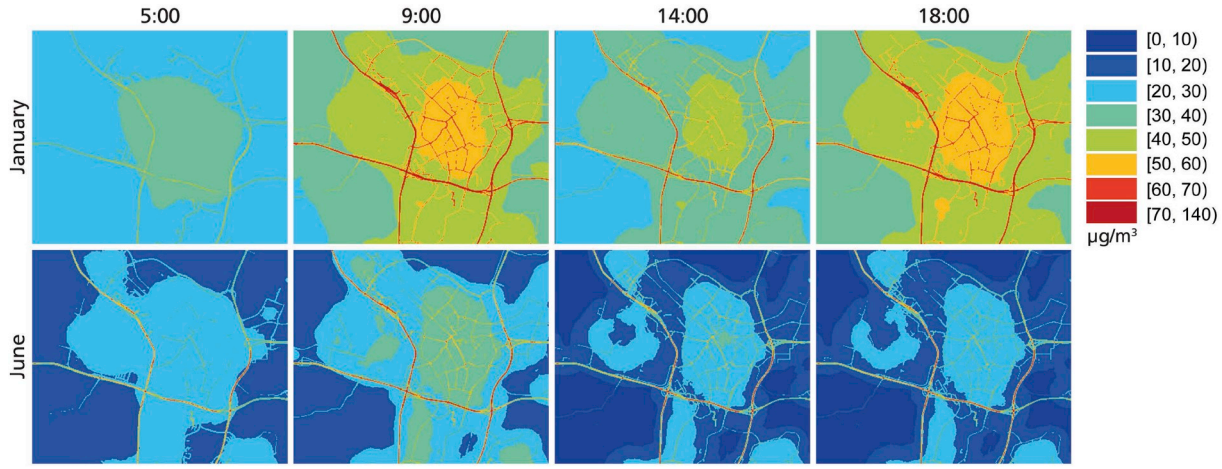


**Fig. 9.** Maps of predicted $NO_2$ concentrations on weekdays for Utrecht city and surrounding area, during rush hour and outside rush hour. The distances in map are 15.4 km horizontally and 12.9 km vertically.

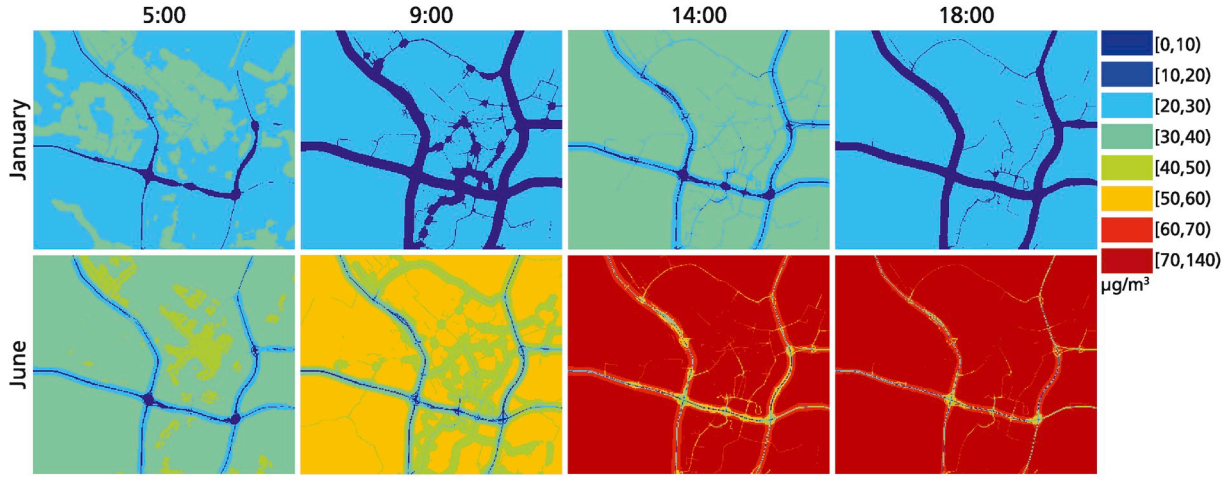


**Fig. 10.** Maps of predicted $O_3$ concentrations on weekdays for Utrecht city and surrounding area, during rush hour and outside rush hour. The distances in map are 15.4 km horizontally and 12.9 km vertically.

*5.2.4. Interactions between pollutants*

Our LUR models fit a separate linear regression model for each air pollutant and each time step. Future studies aiming at improved prediction may need to consider including more interactions in the models. One possibly promising approach would be to define a single statistical model for the three air pollutants considered. This most likely requires incorporation of mechanistic rules describing the chemical reactions between the different pollutants. The modelling would become more process-oriented, which is advantageous for process understanding, but it is not guaranteed it would lead to predictions with a higher accuracy. The identification of the parameters of such a model may need inverse modelling techniques, which makes the fitting of the parameters a considerably larger challenge.

*5.2.5. Spatial and temporal correlations*

Another direction of improvement of spatiotemporal LUR models would be the inclusion of terms describing the spatiotemporal correlations in air pollution, in addition to the components describing the relations with predictor variables. To investigate the relevance of incorporating spatiotemporal correlations, we computed space-time variograms (supplement Fig. 7) of two months in different seasons to quantify spatiotemporal correlations. We found that the spatial continuity is not captured by the ground sensor measurements. Future studies incorporating spatial correlations to improve predictions may need to assume a variogram describing spatial correlations in air pollution, possibly taken from existing studies in the Netherlands, or include more air pollution ground sensors (mobile or static) for better spatial sampling. The space-time variogram (supplement Fig. 7) indicates temporal correlations in long-term hourly aggregations, in particular for weekdays. Thus, including, for instance, a lag one correlation may increase model prediction accuracy in future studies. The temporal correlations were not used in our study as we focus on identifying and evaluating the differences between the spatial prediction of each temporal aggregation/time-step (e.g., differences between each hour of the day, differences between weekday vs. weekend). Integrating the temporal correlations between observations for the spatial prediction makes the coefficients of land use predictors between models at different time-steps less comparable as each model becomes dependent on models describing air pollution at a preceding time-step. For example, if an AR (1) (first order auto-regressive) model is used, model prediction at 5:00 a.m. becomes dependent on model prediction at 4:00 a.m., and similarly the prediction at 6:00 a.m. will depend on the 5:00 a.m. prediction. Thus, comparing contributions of land use predictors between time steps becomes more difficult as they will become related.

*5.2.6. Variable selection*

Even though we select variables statistically, the predictor selection process cannot be seen as completely data-driven. The Lasso is

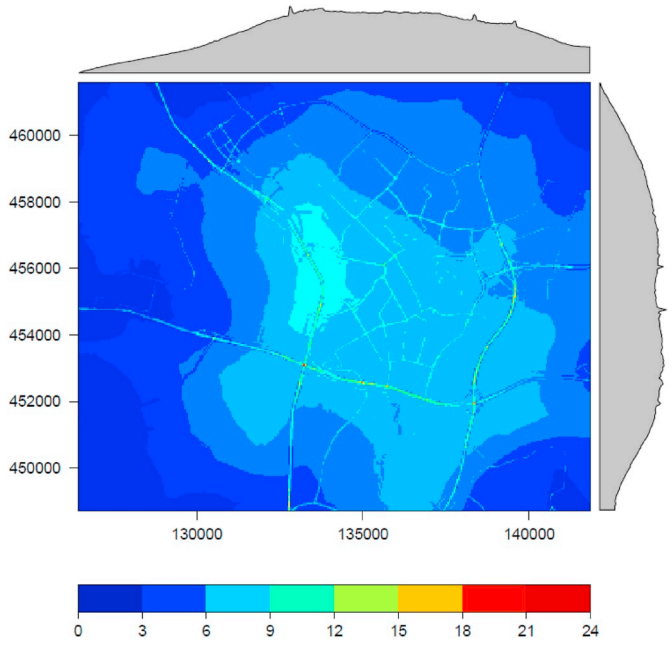


**Fig. 11.** Standard deviation of temporal predictions of $NO_2$ that represent different seasons and hours, namely January, April, June, October at respectively 05:00 a.m., 09:00 a.m., 14:00 p.m., 18:00 p.m. The row and column averages are shown at the top and right sides, respectively. The same standard deviation map for the whole Netherlands is shown in the supplement Fig. 8.

stochastic, meaning it could select other (most likely highly correlated) variables. In addition, the Lasso selected variables are further reduced not by increasing the penalty term to reduce model variance but by applying the BBS to the second stage. This has the benefit of considering bias, variance, information criterion, and $R^2$ in making the decision.

*5.2.7. Ground monitoring stations*

It is important to note that the variable selection process is partly a function of the number of samples and the sampling scheme, and the exclusion of the small buffer variables may be due to the relatively low number of ground monitors. In our study, the ground monitors are located at least 25 m away from the roads, which makes it impossible to identify coefficients for buffers with a radius smaller than 25 m. This is a limitation for modelling the near-road variability of the highly traffic-related and localised air pollutants $NO_2$ and NO. Future studies aiming at more accurate spatiotemporal prediction of traffic related pollution

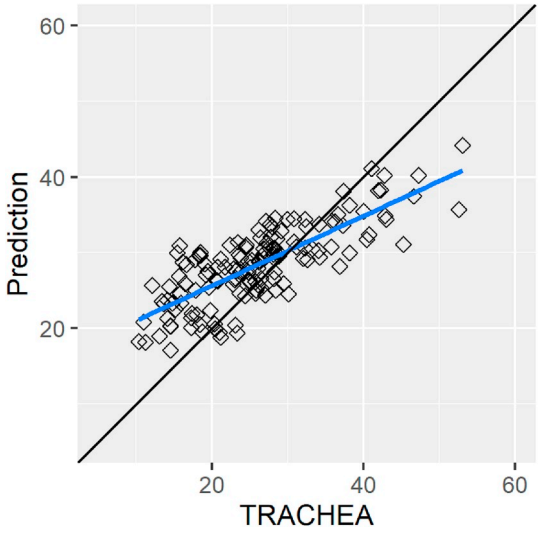


**Fig. 13.** Comparison of the average of our temporal model predicted $NO_2$ values and TRACHEA measurements. Blue line: fitted linear regression line, $R^2$: 0.61. Black line: 1 to 1 line. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

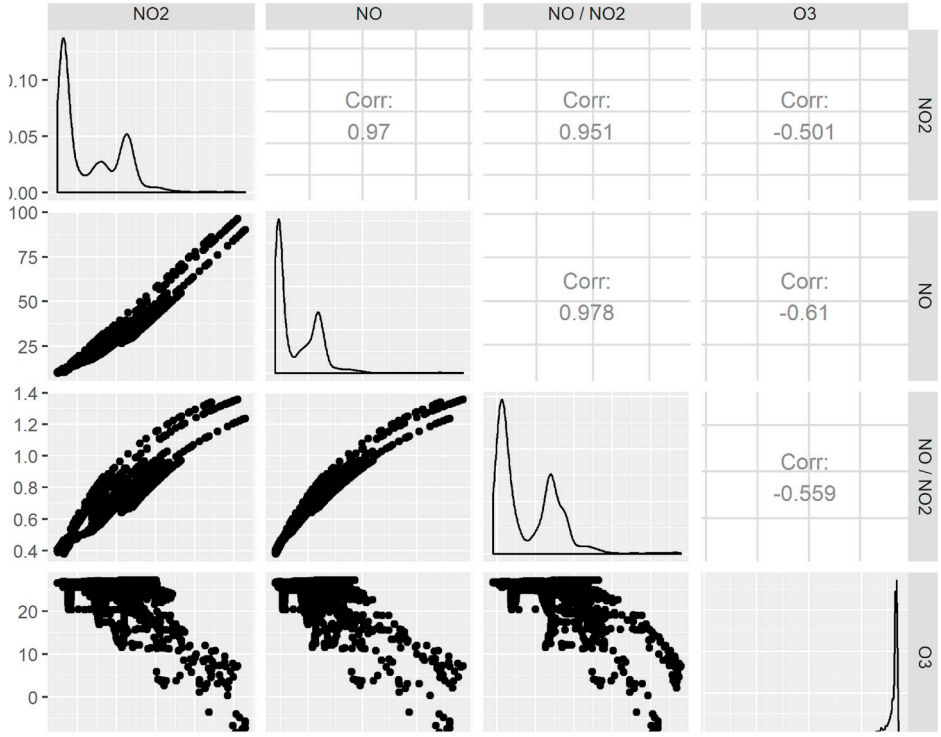


**Fig. 12.** Co-variability of predicted $NO_2$, NO, $NO/NO_2$ (ratio between NO and $NO_2$), and $O_3$. The diagonal plots shows the histogram. All values are from the N–S transect (Fig. 1) across Utrecht, in January, for weekdays. The x-axis is shown on top of the first row of the figure.

need to involve air pollution measurements close to roads.

## 6. Conclusion

We developed high resolution hourly LUR models at different temporal aggregations to study the spatiotemporal variations and interactions of NO, $NO_2$, and $O_3$. The amount of spatiotemporal variation in air pollution implies they need to be considered in personal exposure calculation. The temporal models that are developed with the same predictors selected using an annual model explain less variation of the pollutants at nighttime. The amount of variation each predictor could explain varies hourly and differ between seasons. These assessments of model fitting are important for understanding the contribution of air pollutant sources and for applying these models to exposure assessment and epidemiology studies. We found NO and $NO_2$ having similar patterns and the peaks close to urban roads during rush hours varies in different seasons. The $O_3$ shows a contrasting pattern to NO and $NO_2$. The spatiotemporal patterns can be explained by traffic, weather, and photochemical processes, e.g., the midweek peaks in $NO_2$ strongly overlap with rush hour patterns and the differences between summer and winter times.

## 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.

## Acknowledgements

This research is funded by the Global Geo Health Data Centre (Utrecht University) and the Startimpulsprogramma Meten en Detecteren van Gezond Gedrag (Dutch Science Foundation). We thank Ton Marcus for his advises and contributions on improving the figures. We are grateful to the contributions from the reviewers.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2019.117238.

## References

Baxter, L.K., Dionisio, K.L., Burke, J., Sarnat, S.E., Sarnat, J.A., Hodas, N., Rich, D.Q., Turpin, B.J., Jones, R.R., Mannshardt, E., et al., 2013. Exposure prediction approaches used in air pollution epidemiology studies: key findings and future recommendations. J. Expo. Sci. Environ. Epidemiol. 23 (6), 654.

Beelen, R., Hoek, G., van Den Brandt, P.A., Goldbohm, R.A., Fischer, P., Schouten, L.J., Jerrett, M., Hughes, E., Armstrong, B., Brunekreef, B., 2008. Long-term effects of traffic-related air pollution on mortality in a Dutch cohort (NLCS-AIR study). Environ. Health Perspect. 116 (2), 196.

Beelen, R., Hoek, G., Vienneau, D., Eeftens, M., Dimakopoulou, K., Pedeli, X., Tsai, M.-Y., Künzli, N., Schikowski, T., Marcon, A., Eriksen, K.T., Raaschou-Nielsen, O., Stephanou, E., Patelarou, E., Lanki, T., Yli-Tuomi, T., Declercq, C., Falq, G., Stempfelet, M., Birk, M., Cyrys, J., von Klot, S., Nádor, G., Varró, M.J., Dėdelė, A., Gražulevičienė, R., Mölter, A., Lindley, S., Madsen, C., Cesaroni, G., Ranzi, A., Badaloni, C., Hoffmann, B., Nonnemacher, M., Krämer, U., Kuhlbusch, T., Cirach, M., de Nazelle, A., Nieuwenhuijsen, M., Bellander, T., Korek, M., Olsson, D., Strömgren, M., Dons, E., Jerrett, M., Fischer, P., Wang, M., Brunekreef, B., de Hoogh, K., 2013. Development of $NO_2$ and NOx land use regression models for estimating air pollution exposure in 36 study areas in europe - the ESCAPE project. Atmos. Environ. 72, 10–23. https://doi.org/10.1016/j.atmosenv.2013.02.037. ISSN 1352-2310. http://www.sciencedirect.com/science/article/pii/S1352231013001386. URL.

Bertazzon, S., Johnson, M., Eccles, K., Kaplan, G.G., 2015. Accounting for spatial effects in land use regression for urban air pollution modeling. Spatial and spatio-temporal epidemiology 14, 9–21.

Boniardi, L., Dons, E., Campo, L., Poppel, M.V., 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. ISSN 0013-9351. http://www.sciencedirect.com/science/article/pii/S0013935119303093. URL.

Briggs, D.J., de Hoogh, C., Gulliver, J., Wills, J., Elliott, P., Kingham, S., Smallbone, K., 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. Sci. Total Environ. 253 (1–3), 151–167.

Cohen, M.A., Adar, S.D., Allen, R.W., Avol, E., Curl, C.L., Gould, T., Hardie, D., Ho, A., Kinney, P., Larson, T.V., et al., 2009. Approach to estimating participant pollutant exposures in the multi-ethnic study of atherosclerosis and air pollution (mesa air). Environ. Sci. Technol. 43 (13), 4687–4693.

Copernicus, 2018. CORINE land cover datasets. URL. https://land.copernicus.eu/pan-european/corine-land-cover. Accessed 20 March 2018.

Cordioli, M., Pironi, C., Munari, E.D., Marmiroli, N., Lauriola, P., Ranzi, A., 2017. Combining land use regression models and fixed site monitoring to reconstruct spatiotemporal variability of no2 concentrations over a wide geographical area. Sci. Total Environ. 574, 1075–1084. https://doi.org/10.1016/j.scitotenv.2016.09.089. ISSN 0048-9697. http://www.sciencedirect.com/science/article/pii/S0048969716320083. URL.

de Hoogh, K., Korek, M., Vienneau, D., Keuken, M., Kukkonen, J., Nieuwenhuijsen, M.J., Badaloni, C., Beelen, R., Bolignano, A., Cesaroni, G., Pradas, M.C., Cyrys, J., Douros, J., Eeftens, M., Forastiere, F., Forsberg, B., Fuks, K., Gehring, U., Gryparis, A., Gulliver, J., Hansell, A.L., Hoffmann, B., Johansson, C., Jonkers, S., Kangas, L., Katsouyanni, K., Künzli, N., Lanki, T., Memmesheimer, M., Moussiopoulos, N., Modig, L., Pershagen, G., Probst-Hensch, N., Schindler, C., Schikowski, T., Sugiri, D., Teixidó, O., Tsai, M.-Y., Yli-Tuomi, T., Brunekreef, B., Hoek, G., Bellander, T., 2014. Comparing land use regression and dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies. Environ. Int. 73, 382–392. https://doi.org/10.1016/j.envint.2014.08.011. ISSN 0160-4120. http://www.sciencedirect.com/science/article/pii/S0160412014002591. URL.

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.

Eeftens, M., Beelen, R., Fischer, P., Brunekreef, B., Meliefste, K., Hoek, G., 2011. Stability of measured and modelled spatial contrasts in $NO_2$ over time. Occup. Environ. Med. 68 (10), 765–770. https://doi.org/10.1136/oem.2010.061135.

Eeftens, M., Beelen, R., de Hoogh, K., Bellander, T., Cesaroni, G., Cirach, M., Declercq, C., Dėdelė, A., Dons, E., de Nazelle, A., Dimakopoulou, K., Eriksen, K., Falq, G., Fischer, P., Galassi, C., Gražulevičienė, R., Heinrich, J., Hoffmann, B., Jerrett, M., Keidel, D., Korek, M., Lanki, T., Lindley, S., Madsen, C., Mölter, A., Nádor, G., Nieuwenhuijsen, M., Nonnemacher, M., Pedeli, X., Raaschou-Nielsen, O., Patelarou, E., Quass, U., Ranzi, A., Schindler, C., Stempfelet, M., Stephanou, E., Sugiri, D., Tsai, M.-Y., Yli-Tuomi, T., Varró, M.J., Vienneau, D., von Klot, S., Wolf, K., Brunekreef, B., 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. URL, PMID: 22963366.

EPA, 2019. Air quality dispersion modeling - preferred and recommended models. https://www.epa.gov/scram/air-quality-dispersion-modeling-preferred-and-recommended-models. Accessed Feb. 18, 2019.

Fischer, P.H., Marra, M., Ameling, C.B., Hoek, G., Beelen, R., de Hoogh, K., Breugelmans, O., Kruize, H., Janssen, N.A., Houthuijs, D., 2015. Air pollution and mortality in seven million adults: the Dutch environmental longitudinal study (duels). Environ. Health Perspect. 123 (7), 697.

Friedman, J., Hastie, T., Tibshirani, R., 2010a. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1.

Friedman, J., Hastie, T., Tibshirani, R., 2010b. Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33 (1), 1–22. URL. http://www.jstatsoft.org/v33/i01/.

GDAL Development Team, 2018. GDAL - geospatial data abstraction library. URL. http://www.gdal.org/. Accessed 1 May 2018.

Grant, L.P., Gennings, C., Wheeler, D.C., 2015. Selecting spatial scale of covariates in regression models of environmental exposures. Canc. Inf. 14 (Suppl. 2), 81.

Grömping, U., et al., 2006. Relative importance for linear regression in r: the package relaimpo. J. Stat. Softw. 17 (1), 1–27.

Gurram, S., Stuart, A.L., Pinjari, A.R., 2015. Impacts of travel activity and urbanicity on exposures to ambient oxides of nitrogen and on exposure disparities. Air Quality, Atmosphere & Health 8 (1), 97–114.

Hagenbjörk, A., Malmqvist, E., Mattisson, K., Sommar, N.J., Modig, L., 2017. The spatial variation of o 3, no, no 2 and no x and the relation between them in two Swedish cities. Environ. Monit. Assess. 189 (4), 161.

Han, S., Bian, H., Feng, Y., Liu, A., Li, X., Zeng, F., Zhang, X., 2011. Analysis of the relationship between o3, no and no2 in tianjin, China. Aerosol Air Qual. Res 11 (2), 128–139.

Hoek, G., Beelen, R., de Hoogh, K., Vienneau, D., Gulliver, J., Fischer, P., Briggs, D., 2008. A Review of Land-Use Regression Models to Assess Spatial Variation of Outdoor Air Pollution. Atmospheric Environment, pp. 7561–7578. https://doi.org/10.1016/j.atmosenv.2008.05.057.

Hoek, G., Krishnan, R.M., Beelen, R., Peters, A., Ostro, B., Brunekreef, B., Kaufman, J.D., 2013. Long-term air pollution exposure and cardio-respiratory mortality: a review. Environ. Health 12 (1), 43.

INTARESE, 2018. INTARESE population data. URL. http://www.integrated-assessment.eu/eu/index.html. Accessed 3 May 2018.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. An Introduction to Statistical Learning, vol 112. Springer.

Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K., Bierkens, M.F.P., 2010. A software framework for construction of process-based stochastic spatio-temporal models and

data assimilation. Environ. Model. Softw 25 (4), 489–502. https://doi.org/10.1016/j.envsoft.2009.10.004.
Lideman, R., Merenda, P., Gold, R., 1980. Introduction to Bivariate and Multivariate Analysis Scott. Scott Foresman: Glenview, IL, USA.
Lu, M., Schmitz, O., Vaartjes, I., Karssenberg, D., 2019. Activity-based air pollution exposure assessment: differences between homemakers and cycling commuters. Health Place 60, 102233.
Lumley, T., Lumley, M.T., 2013. Package 'leaps'. *Regression Subset selection. Thomas Lumley Based on fortran Code by alan miller*. Available online: http://CRAN.R-project.org/package=leaps. accessed on 18 March 2018.
National Institute for Public Health and the Environment, 2017. Air quality monitoring network. https://www.luchtmeetnet.nl/. Accessed July 1, 2017.
Park, Y.M., Kwan, M.-P., 2017. Individual exposure estimates may be erroneous when spatiotemporal variability of air pollution and human mobility are ignored. Health Place 43, 85–94.
PBL, 2018. PBL Netherlands environmental assessment agency. URL. http://www.pbl.nl/en/. Accessed 3 May 2018.
R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL. https://www.R-project.org/.
Rahman, M.M., Yeganeh, B., Clifford, S., Knibbs, L.D., Morawska, L., 2017. Development of a land use regression model for daily NO2 and NOx concentrations in the brisbane metropolitan area, Australia. Environmental Modelling & Software 95, 168–179.
Schmitz, O., Beelen, R., Strak, M., Hoek, G., Soenario, I., Brunekreef, B., Vaartjes, I., Dijst, M.J., Grobbee, D.E., Karssenberg, D., 2019. High resolution annual average air pollution concentration maps for The Netherlands. Scientific Data 6. https://doi.org/10.1038/sdata.2019.35, 190035.
Son, Y., Osornio-Vargas, Á.R., O'Neill, M.S., Hystad, P., Texcalac-Sangrador, J.L., Ohman-Strickland, P., Meng, Q., Schwander, S., 2018. Land use regression models to assess air pollution exposure in Mexico city using finer spatial and temporal input parameters. Sci. Total Environ. 639, 40–48.
van Donkelaar, A., Martin, R.V., Brauer, M., Boys, B.L., Februari 2015. Use of satellite observations for long term exposure assessment of global concentrations of fine particulate matter. Environ. Health Perspect. 123 (2), 135–142.
von Klot, S., 2011. Equivalence of using nested buffers and concentric adjacent rings as predictors in land use regression models. Atmos. Environ. 45 (24), 4108–4110.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B 67 (2), 301–320.