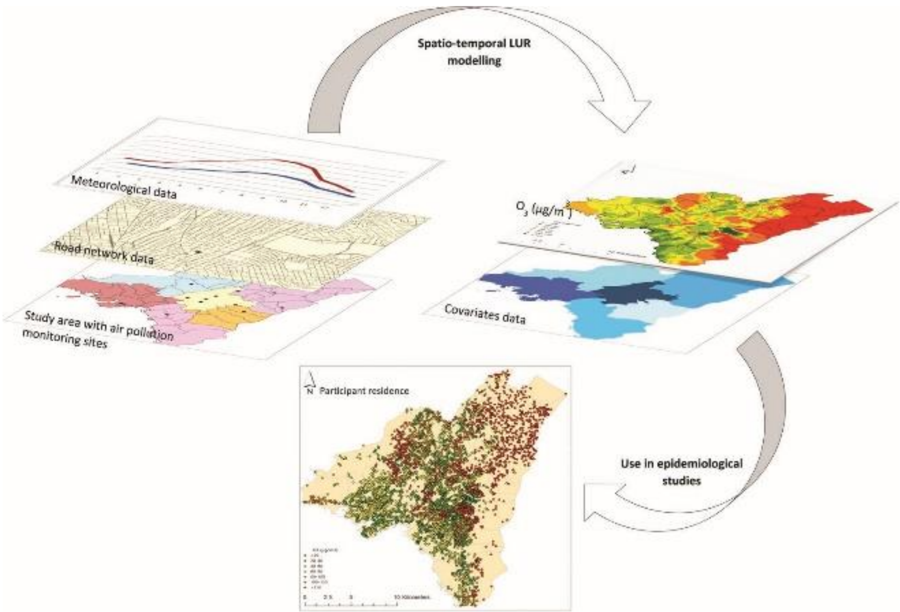
Global NEST Journal, Vol 22, No 1, pp 85-94


# Spatio-temporal land use regression modelling of ozone levels in Athens, Greece

**Dimakopoulou K.[1], Samoli E.[1] and Katsouyanni K.[1,2*]**
[1]Department of Hygiene, Epidemiology and Medical Statistics, School of Medicine, National and Kapodistrian University of Athens, 75 Mikras Asias, Athens, 11527, Greece
[2]School of Population Health and Environmental Sciences and MRC-PHE Centre for Environment and Health, King's College London, London, UK
Received: 14/10/2019, Accepted: 20/01/2020, Available online: 22/01/2020
*to whom all correspondence should be addressed: e-mail: kkatsouy@med.uoa.gr
https://doi.org/10.30955/gnj.003233

**Graphical abstract**




**Abstract**

Spatio-temporal methods have been developed for the estimation of concentrations of pollutants such as particulate matter and nitrogen dioxide for application in epidemiological studies. A limited number of city-specific spatio-temporal ozone ($O_3$) models have been proposed until today. Our aim was to develop a spatio-temporal land use regression (LUR) model that estimates daily concentrations of $O_3$, for the whole year, as well as the warm (April-September) and cold season (October-March), within the greater Athens area. We developed models using a semiparametric approach including linear and smooth functions of spatial and temporal covariates and a bivariate smooth thin plate function. The final set of explanatory variables was selected based on the adjusted-$R^2$. We tested the final model in temporal and spatial terms following a leave-one out monitor approach. The adjusted-$R^2$ in the leave-one-out cross validation was 0.73 for the annual model (warm: 0.65 and cold: 0.70). The spatial terms in our annual model explained 32.9% and the temporal 63.2% of the variability in $O_3$. The developed models showed good validity when comparing predicted and observed measurements for the 2015 data. Spatio-temporal LUR modeling provides a useful tool for estimating $O_3$ spatio-temporal variability with adequate accuracy for subsequent use in epidemiological studies.

**Keywords:** Spatio-temporal modeling, ozone, land use regression, exposure, epidemiology, air pollution.

## 1. Introduction

Epidemiological studies investigating the effects of air pollution exposure on health have increasingly been relying on modelling estimation techniques to provide an individualized exposure, primarily at participant home addresses and in large study samples (Beelen *et al.*, 2014; Dimakopoulou *et al.*, 2014; Kirrane *et al.*, 2015; Lee *et al.*, 2016; Ostro *et al.*, 2015; Raaschou-Nielsen *et al.*, 2013). The advantage of these exposure assessment methods is that they take into account the variability of air pollution concentrations at fine spatial scale, leading to a possible decrease in exposure measurement error and increase in statistical power.

Among the most common methods applied to estimate exposure concentrations are chemical transport models (CTM) (Jerret *et al.*, 2005), land use regression models (LUR) (Beelen *et al.*, 2013; Eeftens *et al.*, 2012; Gryparis *et al.*, 2007; Gryparis *et al.*, 2014) and models that use satellite-based aerosol optical depth (AOD) data (Dadvand *et al.*, 2014; Kloog *et al.*, 2014). All of the aforementioned approaches can be extended to account for both the temporal and spatial concentration variations. Therefore, they are able to predict both short- or long-term exposure estimates, a useful for different epidemiological study designs. Such spatio-temporal models have been developed primarily for particles (PM) and nitrogen dioxide ($NO_2$) and have been applied in epidemiological studies in the USA (Maynard *et al.*, 2007; Puett *et al.*, 2011) and in Europe (Dadvand *et al.*, 2013; Katsoulis *et al.*, 2013; Dimakopoulou *et al.*, 2017). The association between short-term $O_3$ exposure and health is considered as serious as the adverse health effects of PM exposure (Brunekreef *et al.*, 2012). Moreover, there is a recent debate about the effects of long-term exposure to $O_3$ on health (Schwartz, 2016). However, despite these issues only few studies have developed models for predicting ozone ($O_3$) concentrations.

$O_3$ is a main component of the photochemical air pollution cloud and a powerful oxidising agent (EEA, 2011). It is an extremely reactive gas created by the reaction of traffic-related pollutants, such as nitrogen oxides (NOx) and volatile organic compounds (VOCs) with sunlight (WHO, 2014). O3 is a secondary pollutant with spatio-temporal variations. $O_3$ concentrations are lower in urban areas compared with the suburbs and rural areas. This is because emissions of NO tend to scavenge $O_3$ and convert it into $NO_2$ and oxygen ($O_2$) (Stedman and Kent, 2008). $O_3$ variation also depends on meteorology (mainly on solar radiation but also on ambient temperature, relative humidity and wind speed) with $O_3$ concentrations displaying a summer maximum in urban areas (Monks, 2000).

Previous studies have shown that short-term exposure to $O_3$ has adverse effects on pulmonary function, respiratory symptoms, while it is associated with increased medication usage, morbidity and mortality (WHO, 2008,2013; EPA, 2009; Karakatsani *et al.*, 2017; Samoli *et al.*, 2017). On the other hand, the evidence on mortality effects of long-term exposure to $O_3$ is inconclusive (Jerrett *et al.*, 2009; Brunekreef *et al.*, 2012; Atkinson *et al.*, 2016), while recent studies suggest an association between long-term exposure and reduced lung capacity and increased asthma incidence (WHO, 2013).

Only few models have been developed for $O_3$ exposure assessment in epidemiological studies. Adam-Poupart *et al.* (2014) developed three different spatio-temporal models (a LUR mixed effects model, a Bayesian maximum entropy (BME) and a kriging method model) to predict summer ground-level $O_3$ in Quebec, Canada. They compared the models with leave-one-out cross validation (LOOCV) and found that the combination of LUR and BME methods reduced the estimation errors. A recent study in the USA (Wang *et al.*, 2016) improved the accuracy of the spatio-temporal estimates of $O_3$ in the Los Angeles Basin, by including a smoothed spatial CTM output in their spatio-temporal LUR model. To our knowledge, only spatial LUR models have been developed in Europe in order to assess $O_3$ exposure for long-term epidemiological studies in Sweden (Malmqvist *et al.*, 2014) and the Netherlands (Kerckhoffs *et al.*, 2015).

In the present paper we develop and validate a spatio-temporal LUR model for $O_3$ concentrations, using data from 2001 to 2014, which estimates daily concentrations of $O_3$, for the whole year, warm (April 1st to 30th September) and cold season (October 1st and 31st March), within the greater Athens area, Greece. Our goal is to develop a useful tool that can predict daily maximum 8-hour average $O_3$ concentrations in urban settings and that can be used in various epidemiological study designs.

## 2. Materials and methods

### *2.1. $O_3$ monitoring data*

We obtained hourly ground-level $O_3$ observations for 2001 through 2014 from the fixed air pollution monitoring network, run by the Ministry of Environment and Energy (http://www.ypeka.gr/) in the greater Athens area.

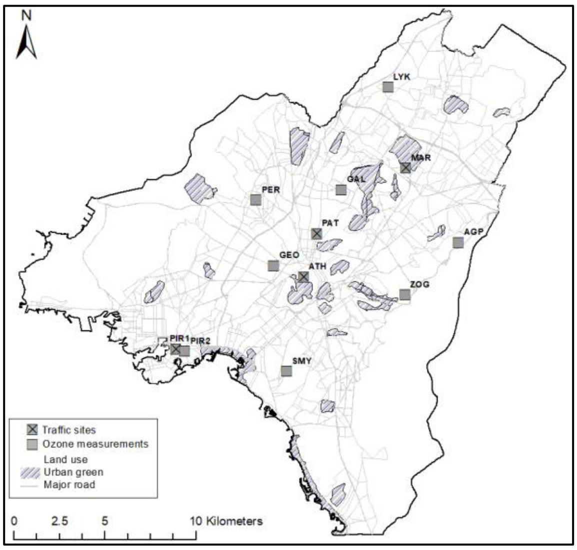



**Figure 1**. Map of the geographical location of the 12 fixed monitoring sites operated by the Ministry of Environment and Energy, at the greater Athens area, Greece and the land use data included in the developed model

We calculated the daily maximum 8-hour average ozone concentration to represent daily exposure and included all available data. All sites used the same monitoring technology and complied with the E.U. Directives (Directive, 2008/50/EC). UV absorption was used to measure $O_3$. In Athens, the number of $O_3$ monitoring stations ranged from 9 to 12 stations during 2001–2014, resulting in 48,137 total observations (station-days). Figure 1 shows the study area and the geographical location of the monitoring sites.

*2.2. Road density data*

We used traffic counts from field measurements carried out by the Ministry of Infrastructure, Transport and Networks (MITN) (http://www.yme.gr/), traffic data from projects conducted by the municipality authorities and our own on-site measurements. The integration method of all the available traffic data for the study area and the validation procedures applied have been described in detail elsewhere (Gryparis *et al.*, 2014).

*2.3. Meteorological data*

We obtained meteorological data from the fixed meteorological site located at the center of Athens (National Observatory of Athens, Thissio) for the years 2001–2014. We extracted mean daily temperature (°C), relative humidity (%), wind speed (m/s), barometric pressure (mb), total solar radiation ($Wh/m^2/day$), average solar radiation between hours 4 am and 7 pm ($Wh/m^2/day$), average solar radiation between hours 10 am and 3pm ($Wh/m^2/day$), cloud coverage (oktas) and wind direction (north, south, east, west).

*2.4. Other predictor variables for spatial variability*

In total, 87 potential predictor variables of $O_3$ concentrations spatial variability were extracted by conducting GIS analyses. These variables are traffic-related, characterize land use and population density in different buffers around the fixed air pollution monitoring sites. The geographical coordinates of each monitoring site were obtained through the Ministry of Environment and Energy. The buffer zones used for traffic-related variables were: 25, 50, 100, 300, 500 and 1000 m and were selected to take account of known dispersion patterns (Beelen *et al.*, 2013). Total traffic load was calculated as length of the road segment multiplied by traffic intensity and divided by road segment for all roads within each buffer zone. We also calculated the length of the road segments in different buffer zones around the fixed monitoring sites, traffic intensity on the nearest road to the fixed monitoring site and inverse distance to the nearest road to the fixed monitoring site. Moreover, we extracted the same traffic variables by including only major roads (road-width category ≥ 5). Land use data was available from CORINE (COordination and INformation on the Environmental programme, initiated by the European Commission) for year 2000. Definitions of the ESCAPE study were used (Beelen *et al.*, 2013) to characterize land use (high density residential areas, low density residential areas, industry, port, urban green and semi natural plus forested areas). In addition, we used building and population density data for 2001 (data obtained from the Hellenic Statistical Authority – EL. STAT.). The buffer zones used for land use variables were 100, 300, 500, 1000 and 5000 m. Areas of different land cover use, building and population density were calculated in $m^2$, within each buffer zone. Finally, the altitude of the monitoring sites was obtained from the Ministry of Environment and Energy.

*2.5. Development of models*

We developed semi-parametric spatio-temporal land use regression models to predict $O_3$ concentrations measured at fixed monitoring sites for the whole year, warm and cold season. The model development methods have been described in detail elsewhere (Gryparis *et al.*, 2014). In brief, we developed spatio-temporal semiparametric models, of the form:

$$\text{log.poll}_{ij} = W_{ij}^T \beta + \sum_{l=1}^{q} f_l(s_{l,ij}) + h(\text{geog}_{ij}) + \varepsilon_{ij}, \quad (1)$$

where $\text{log.poll}_{ij}$ is the log-transformed measurement of $O_3$ at location i on day j, $f_l(.)$ l = 1,2,…,q, is an unspecified smooth function reflecting the non-linear effect of covariate $s_{l,ij}$ on log-transformed pollutant's concentration $\text{log.poll}_{ij}$, $s_{l,ij}$ stands for the $l^{th}$ smoothed covariate, $\text{geog}_{ij}$ = ($\text{latitude}_i$, $\text{longitude}_j$), h is a bivariate smooth function of geographical coordinates (latitude and longtitude), and $W_{ij}$ is the vector of covariates that have a linear effect on $\text{log.poll}_{ij}$. In summary, our model consists of covariates that have either a linear effect or a smooth effect on the outcome and of a bivariate smooth function of geography accounting for the remaining residual correlation. We assume that the errors $\varepsilon_{ij}$ are independent normal variables, with mean 0 and constant variance $\sigma^2_\varepsilon$. Degrees of freedom were estimated via REML.

The final variables included in the annual model were different years (2002 to 2014; 2001 is the reference category), the day of the week (Monday to Saturday; Sunday is the reference category), wind direction (east, south, west; north is the reference category) and cloud coverage (oktas). Also, we used penalized splines to model temperature (3 degrees of freedom), relative humidity (3 degrees of freedom), wind speed (3 degrees of freedom), average solar radiation between 10 am and 3 pm (8 degrees of freedom), day count (10 degrees of freedom). A bivariate smooth function (thin plate spline) of geography was used to account for remaining residual spatial correlation. Traffic load on the nearest major road ($veh\ day^{-1}\ m$), length of the major road segments (m) in a buffer of 500m around the fixed $O_3$ monitoring sites and Urban Green land use class ($m^2$) in a buffer of 300 m around the fixed sites, were expressed as linear terms.

The final spatio-temporal LUR model for warm & cold season $O_3$ concentrations accounted for the same variables at the temporal scale as the annual model. Significant spatial covariates were for warm season model: traffic load on the nearest major road ($veh\ day^{-1}\ m$), urban green land cover in a buffer of 100 meters ($m^2$), inverse distance of the monitoring sites to the nearest major road (included as a non-linear term; $m^{-1}$), around each monitoring site and geographical location (longitude, latitude) of the monitoring sites; for cold season model: traffic load on the nearest major road ($veh\ day^{-1}\ m$), household density (N) in a buffer of 1000 meters around each monitoring site, nitrogen dioxide concentrations ($\mu g/m^3$) and geographical location (longitude, latitude) of the monitoring sites.

To check whether there was any remaining spatial or temporal residual autocorrelation, we used the partial autocorrelation function plots (Box *et al.*, 2008), for each monitoring site separately and calculated Moran's I (Moran, 1950) for all the days of the study period (2001–2014), respectively.

*2.6. Model validation*

We performed both temporal and spatial validation as described in Gryparis *et al.* (2014). Briefly, we calculated the overall and by-site bias by comparing the daily $O_3$ predictions with the measurements for the year 2015 (not used in the model development) at the same fixed site monitoring locations. We validated our developed models with leave-one-out cross validation (LOOCV). Furthermore, we estimated the percentage explained spatial and temporal variability of the annual model's, by excluding all temporal and spatial terms respectively and comparing adjusted-$R^2$ values.

All analysis was conducted using the R statistical software (version 2.10.1; R Development Core Team, 2009) and the R library "SemiPar" version 1.0-2.

## 3. Results and discussion

*3.1. Distribution of $O_3$ concentrations and covariates*

Data on $O_3$ concentrations were available from a maximum of 12 fixed stations during the period 2001 to 2014. Table 1 presents the summary statistics for the available 8-hour maximum $O_3$ data for each fixed site. Also, a description of site type, distance to nearest road segment and altitude of fixed sites is given in Table 1. Annual concentrations presented an increase in 2011 followed by a decline by about 7% in 2008–2014. Monitoring sites at traffic locations showed significantly (p-value < 0.001) lower mean $O_3$ concentrations compared to background locations (50.7 $\mu g/m^3$ vs 75.0 $\mu g/m^3$, respectively). The spatial variability of the average $O_3$ concentrations in the study area was larger compared to the temporal variability (per site range: -36.9% to 59.7% & -11.4% to 14.1% of the overall mean value, respectively).

Table 2 presents summary statistics for the meteorological covariates included in the final models. In addition to the geographical location of the monitoring sites, Figure 1 also shows the "Urban Green" land cover class available from CORINE and the major road network in the study area.

*3.2. Model selected, Spatial and temporal variability and model performance*

Table 3 summarizes the estimated coefficients for the final set of linear predictors included in the annual $O_3$ spatio-temporal LUR model. Traffic load on the nearest major road, length of the major road segments in a buffer of 500 m around the fixed $O_3$ monitoring sites, urban green land use class in a buffer of 300 m around the fixed $O_3$ monitoring sites, cloud coverage and day of the week (compared to Sunday = reference category) were negatively associated with $O_3$ levels. A significant annual incline in $O_3$ concentrations was shown, compared to year 2001 (reference category). The lowest $O_3$ concentrations were associated with west winds. The adjusted-$R^2$ of the developed annual model was 0.76, while for the warm and cold season it was 0.70 and 0.71, respectively. Model assumptions were not violated. Inspection of the PACF plots separately per site did not reveal any major temporal residual autocorrelation. There was no significant spatial autocorrelation in the residuals of the models, since only 10% of the days from the 14-year study period resulted in Moran's I p-value <0.05.

*3.3. Model validation*

The predicted values were compared with the actual $O_3$ measurements for the year 2015 (not used in the model building procedure), per fixed monitoring site. We found that the daily mean model predictions were on average higher compared to the daily mean observed values (Figure 2). The overall bias of the $O_3$ prediction (calculated as observed-predicted values for year 2015) was -2.5 $\mu g/m^3$, while the by site-type bias was -7.3 $\mu g/m^3$ for the traffic sites and was -2.6 $\mu g/m^3$ for the background sites. Therefore, the developed model seems to perform better for background monitors.

Figure 3 shows the error bar plot of the annual mean observed and predicted $O_3$ values, for 2015, per fixed monitoring site. A couple of sites (i.e. ATH, PIR1, SMY) displayed differences between measured and estimated values.

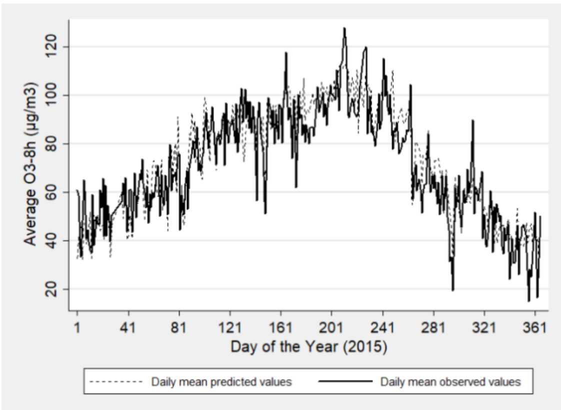



**Figure 2.** Plot of the daily average $O_3$-8h ($\mu g/m^3$) observed (solid line) and predicted (dash line) values for the year 2015

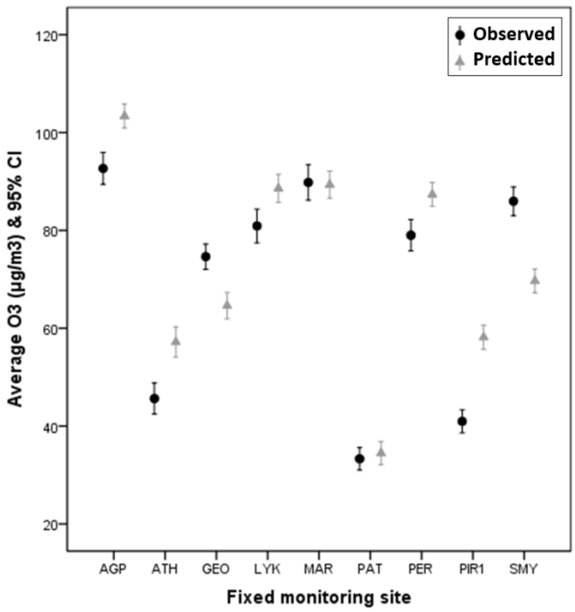



**Figure 3.** Error bar for the 2015 annual mean $O_3$-8h ($\mu g/m^3$) observed (black color) & predicted (grey color) values along with their 95% C.I., by monitoring site

**Table 1.** Description of the fixed monitoring sites operated by the Ministry of Environment and Energy in the greater Athens area and summary statistics for the available 8-hour maximum $O_3$ measurements by site during 2001–2014

| | | Distance (m) to the nearest | | | $O_3$ (μg/m$^3$) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Site ID** | **Site type** | **Street** | **Major street** | **Altitude (a.m.s.l.)*** | **Mean** | **SD** | **25% percentile** | **Median** | **75% percentile** | **Min–Max** |
| ATH | Urban Street Site | 12 | 12 | 100 | 48 | 26.0 | 27 | 46 | 66 | 2–171 |
| GEO | Suburban–Industrial | 30 | 70 | 40 | 63 | 30.7 | 38 | 63 | 87 | 1–158 |
| LYK | Suburban | 50 | 50 | 234 | 82 | 33.5 | 56 | 83 | 107 | 1–217 |
| MAR | Urban street site | 30 | 150 | 170 | 78 | 30.6 | 54 | 76 | 100 | 1–195 |
| SMY | Urban background | 10 | 100 | 50 | 82 | 34.2 | 56 | 83 | 110 | 2–224 |
| PAT | Urban Street Site | 3 | 3 | 105 | 28 | 20.6 | 11 | 22 | 39 | 1–105 |
| PIR1 | Urban Street Site | 15 | 15 | 20 | 48 | 26.0 | 27 | 45 | 66 | 2–142 |
| PER | Urban background | 8 | 80 | 80 | 77 | 31.5 | 51 | 78 | 102 | 1–192 |
| AGP | Suburban background | 200 | 200 | 290 | 94 | 31.3 | 69 | 93 | 117 | 9–218 |
| ZOG | Suburban background | 800 | 800 | 245 | 83 | 30.0 | 61 | 84 | 104 | 1–217 |
| PIR2 | Urban background | - | - | 25 | 57 | 29.8 | 35 | 54 | 74 | 3–160 |
| GAL | Suburban background | 7 | 75 | 154 | 75 | 35.7 | 48 | 74 | 99 | 1–211 |

**above mean sea level*

**Table 2**. Summary statistics for the meteorological data (daily values; 2001–2014) from the National Observatory site in the center of the greater Athens area

| Meteorological variables | Mean (SD) | Min-Max |
|---|---|---|
| Temperature (°C) | 19 (7.4) | -7–36 |
| Relative humidity (%) | 64 (14.7) | 24–100 |
| Wind speed (m/sec) | 3.2 (1.6) | 0.3–12.7 |
| Solar radiation between 4am to 7pm (Wh/m$^2$/day) | 289 (146.0) | 0–554 |
| Cloud coverage (oktas) | 4 (2.5) | 0–8 |
| Wind direction (daily prevailing) | **Frequency (%)[1]** | |
| North | 2,289 | 51.3 |
| East | 205 | 4.6 |
| South | 1,554 | 34.9 |
| West | 410 | 9.2 |

*[1]Number of days within the study period 2001–2014, with prevailing wind direction and corresponding percentage*

LOOCV of the $O_3$ annual model resulted in a cross-validation adjusted-$R^2$ of 0.73 (warm: 0.65 and cold: 0.70). The overall bias of the CV predictions (calculated as observed-CV predicted values) was -0.3 μg/m$^3$, while the by site-type bias was -0.8 μg/m$^3$ for the traffic sites and was -0.3 μg/m$^3$ for the background sites. Therefore, also when taking into account the CV predictions the developed model performs slightly better at background sites. The exclusion of all the spatial covariates of the annual $O_3$ model (traffic load on the nearest major road, length of the major road segments in a buffer of 500 m around the fixed $O_3$ monitoring sites, urban green land use class in a buffer of 300 m around the fixed sites and the bivariate smooth term of fixed sites coordinates) resulted in an adjusted-$R^2$ of 0.51. On the other hand, the exclusion of all temporal covariates resulted in an adjusted-$R^2$ of 0.28. Therefore, all spatial terms together accounted for approximately 32.9% in the $O_3$ of the explained variability, while all temporal term for 63.2%.

**Table 3.** Estimates of the linear predictors in the developed $O_3$ spatio-termporal model. Additionally, the model included penalized splines for daily average temperature, relative humidity, wind speed, solar radiation and a variable for day count and a bivariate term of monitoring site geography

| Variable | Coefficient | p-value | 95% C. I. |
|---|---|---|---|
| Intercept | $-1.06 \times 10^4$ | 0.679 | $(-6.09 \times 10^4, 3.96 \times 10^4)$ |
| Urban green in a buffer of 300 m around | $-2.82 \times 10^{-4}$ | 0.009 | $(-4.94 \times 10^{-4}, -7.01 \times 10^{-5})$ |
| Road length of major roads in a buffer of 500 m around monitoring sites (m) | $-2.35 \times 10^{-3}$ | 0.015 | $(-4.24 \times 10^{-3}, -4.52 \times 10^{-4})$ |
| Traffic intensity on nearest major road | $-5.36 \times 10^{-3}$ | <0.001 | $(-6.06 \times 10^{-3}, -4.65 \times 10^{-3})$ |
| *Wind direction (daily prevailing)* | | | |
| North | Reference category | | |
| East | 0.55 | 0.193 | (-0.28, 1.38) |
| South | 1.63 | <0.001 | (1.20, 2.07) |
| West | -1.06 | 0.001 | (-1.68, -0.45) |
| Cloud coverage (oktas) | -0.30 | <0.001 | (-0.41, -0.19) |
| *Day of the week* | | | |
| Sunday | Reference category | | |
| Monday | -8.78 | <0.001 | (-9.40, -8.16) |
| Tuesday | -8.91 | <0.001 | (-9.53, -8.29) |
| Wednesday | -8.72 | <0.001 | (-9.34, -8.10) |
| Thursday | -8.08 | <0.001 | (-8.70, -7.46) |
| Friday | -8.78 | <0.001 | (-9.40, -8.16) |
| Saturday | -3.13 | <0.001 | (-3.75, -2.51) |
| *Year* | | | |
| 2001 | Reference category | | |
| 2002 | 6.18 | <0.001 | (5.04, 7.32) |
| 2003 | 11.05 | <0.001 | (9.06, 13.04) |
| 2004 | 14.04 | <0.001 | (11.48, 16.60) |
| 2005 | 20.37 | <0.001 | (17.06, 23.68) |
| 2006 | 26.42 | <0.001 | (22.23, 30.61) |
| 2007 | 33.18 | <0.001 | (28.11, 38.25) |
| 2008 | 44.69 | <0.001 | (38.72, 50.66) |
| 2009 | 50.31 | <0.001 | (43.40, 57.23) |
| 2010 | 46.34 | <0.001 | (39.94, 52.74) |
| 2011 | 44.18 | <0.001 | (39.06, 49.30) |
| 2012 | 37.44 | <0.001 | (32.64, 42.24) |
| 2013 | 29.45 | <0.001 | (25.59, 33.31) |
| 2014 | 23.75 | <0.001 | (20.29, 27.22) |

*3.4. Discussion*

A LUR model was developed for the greater Athens area, that explained the 76% of the spatio-temporal variability in annual $O_3$-8h concentrations and the 70% and 71% of the spatio-temporal variability in warm- and cold- season $O_3$-8h concentrations. The model is a useful tool that can be used in different epidemiological study designs. For example, in a time series or panel study it can predict daily $O_3$-8h values for the time period 2001 to 2014. For subsequent use in studies assessing long-term effects an average of $O_3$-8h exposure over the time period of interest can easily be calculated from daily predicted values. Moreover, the model can provide an $O_3$-8h estimate for any geographical point in the study area. Therefore, in case address history, work address or personal time activity patterns are known, it can be used to calculate a weighted average of exposure to $O_3$-8h concentrations. This may lead towards a decrease in exposure measurement error.

Previous studies in Europe have developed models for $O_3$ exposure assessment based on a LUR-model approach for capturing fine scale patterns of air pollution. Kerckhoffs *et al.* (2015) developed a national fine spatial scale LUR model for the Netherlands that explained 71% of the spatial variation in summer average $O_3$ concentrations. Malmqvist *et al.* (2014) developed LUR models for two cities in Sweden (Malmo and Umea), with model $R^2$ values of 0.40 and 0.67, respectively. Both studies used similar spatial predictor variables as in the present paper (traffic intensity, major road length and urban green space). A study in Quebec, Canada developed a spatio-temporal LUR model that explained 47% of the variability in summer ground-level $O_3$ (Adam-Poupart *et al.*, 2014). The temporal covariates included in their final model (temperature, precipitation, day of the year, year) were also similar to our temporal predictors. Subsequently, combining the LUR with the Bayesian maximum entropy model improved the model fit ($R^2$ = 0.65). Recent studies on $O_3$ exposure assessment combine CTM outputs with LUR models to improve model fit (Akita *et al.*, 2014; De Nazelle *et al.*, 2010; Wang *et al.*, 2016). The integration of a large-scale CTM model and LUR model, in a study by Wang *et al.* (2016) in the USA, resulted in an improvement in the predictability of estimates for $O_3$ (spatio-temporal LUR model: $R^2$ = 0.75 vs combined model: $R^2$ = 0.78). While, another study in Canada (Hystad *et al.*, 2012), which included dispersion estimated $O_3$ concentrations as an additional predictor variable in the LUR model, explained 56% of the spatio-temporal variability in $O_3$ concentrations. The $R^2$ values of our models are within the range of those previously reported in the literature. Furthermore, our model is locally generated and it is able to capture fine scale spatio-temporal patterns of $O_3$ concentrations in contrast to LUR models that have been developed over different and large study areas (i.e. different countries) (de Hoogh *et al.*, 2018).

The aim of the present study was to capture the spatio-temporal variation of $O_3$ concentrations in the greater Athens area, Greece. A map of long-term average of $O_3$ concentrations estimated from the annual model for 1,000 randomly selected geographical points within the study area, shows higher concentrations in the suburban areas compared to the urban areas (Figure 4).

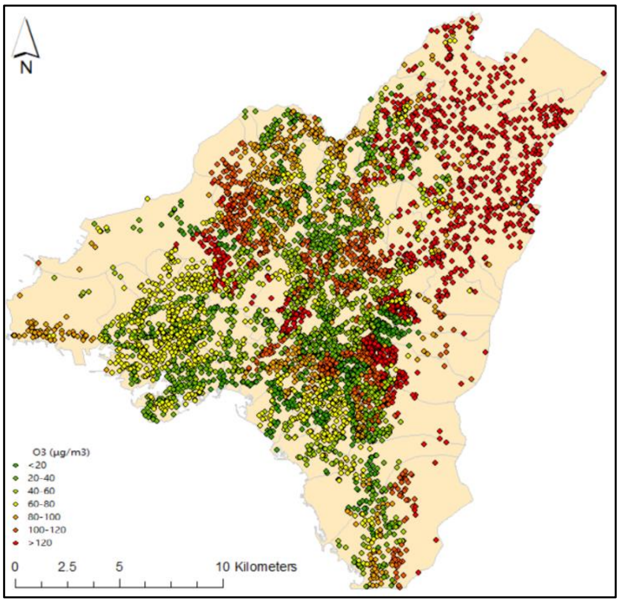



**Figure 4.** Long-term average of $O_3$ (daily 8-hour maximum) concentrations estimated by the annual spatio-temporal LUR model, at 1,000 randomly selected geographical points.

This is consistent with urban decrement, since emissions of primary pollutants such as NO scavenge $O_3$. Moreover, similar maps of long-term average of $O_3$ concentrations estimated from the warm- and cold-season model shows higher $O_3$ concentrations during the warm season and lower $O_3$ concentrations in the cold season (Figures 5 and 6)

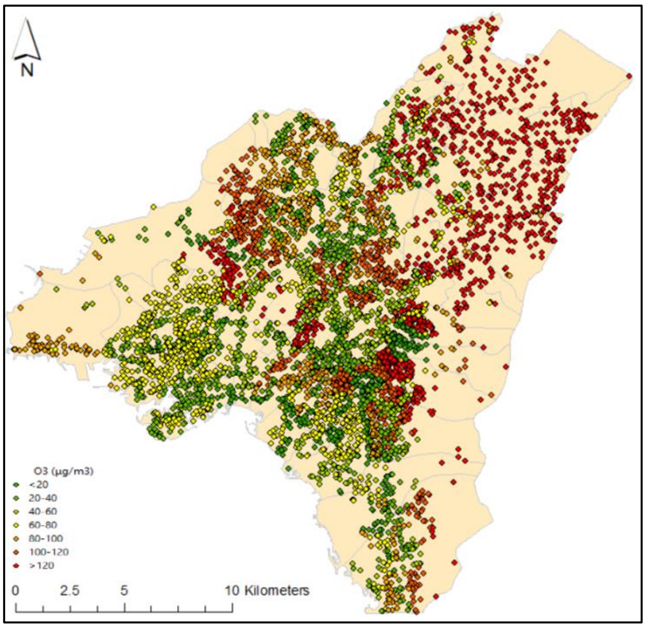



**Figure 5**. Long-term average of $O_3$ (daily 8-hour maximum) concentrations estimated by the warm season spatio-temporal LUR model, at 1,000 randomly selected geographical points

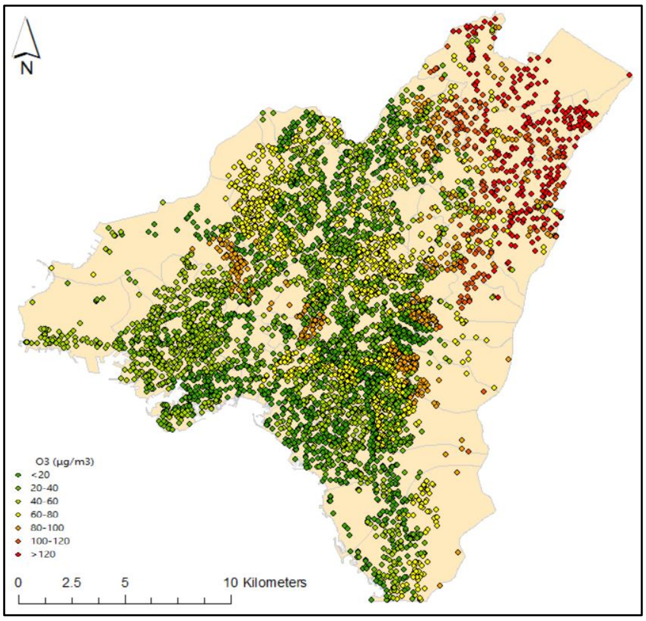



**Figure 6**. Long-term average of $O_3$ (daily 8-hour maximum) concentrations estimated by the cold season spatio-temporal LUR model, at 1,000 randomly selected geographical points

This is consistent with the known $O_3$ seasonal variation pattern. These findings suggest that the developed models account for both spatio-temporal $O_3$ concentrations variation and therefore may decrease exposure measurement error in epidemiological studies.

The declining trend observed in concentration levels of $O_3$ in the study area is linked to Greece's financial crisis from 2010, which resulted in a 100% rise in petrol prices, followed by an immediate traffic decrease of about 10% (MITN). However, despite this decreasing trend, climate change could lead to an increase in the levels of $O_3$ concentrations. Temperature and $O_3$ formation are strongly dependent. As temperatures rise and stay elevated for longer periods of time (EPA, 2009), it is probable that the number of days that are conducive to $O_3$ formation will increase. Therefore, variables that capture long-term trends should be considered as possible predictors during the model development procedure.

A limitation of our study is that we used LOOCV in order to validate our developed models, rather than applying other validation methods such as k-fold cross validation. This was due to the number of available air pollution monitoring sites located in the study area and may led to overestimation of the predictive ability of the models (Wang *et al.*, 2013). Also, a disadvantage was that daily $O_3$ concentrations estimated from a CTM were not available for the study area and time period. Hence, we could not apply a combined modeling approach as in other studies (Hystad *et al.*, 2012; Wang *et al.*, 2016) which suggest that exposure estimates at a location which is not close to a monitoring site may be improved by applying a hybrid model approach.

## 4. Conclusions

We developed and assessed the validity of an annual, warm- and cold- season spatio-temporal LUR model for $O_3$-8h, for the greater Athens area, Greece. Our developed models are capable of providing fine-scale daily $O_3$ concentration estimates, for a 14-year period and for an urban area that is the most populated in Greece (3,752,973 inhabitants; according to Census 2011, EL. STAT.). Subsequently, such predictions can be used in health effects analysis of air pollution. Moreover, a weighted average of the overall exposure to $O_3$-8h can be calculated for each participant in case address history or personal time activity diary information is available. Since our developed models predict daily values, they can be used for either short-term or long-term exposure health effects analysis, the latter by averaging the daily estimates over the time interval of interest. Therefore, spatio-temporal LUR modeling is a promising method to predict $O_3$ annual and warm- or cold- season spatio-temporal variability with adequate accuracy for use in epidemiological studies.

**Acknowledgements**

The work presented was funded by the State Scholarships Foundation (IKY) and the Greek government by the National Strategic Reference Framework 2014–2020. Contract Ref: 2016-050-0503-8976.

**References**

Adam-Poupart A., Brand A., Fournier M., Jerrett M. and Smargiassi A. (2014), Spatiotemporal modeling of ozone levels in Quebec (Canada): a comparison of kriging, land-use regression (LUR), and combined Bayesian maximum entropy–LUR approaches, *Environmental Health Perspectives*, 122, 970–976.

Akita Y., Baldasano J.M., Beelen R., Cirach M., de Hoogh K., Hoek G., Nieuwenhuijsen M., Serre M.L. and de Nazelle A. (2014), Large Scale Air Pollution Estimation Method Combining Land Use Regression and Chemical Transport Modeling in a Geostatistical Framework, *Environmental Science & Technology*, 48, 4452–4459.

Atkinson R.W., Butland B.K., Dimitroulopoulou C., Heal M.R., Stedman J.R., Carslaw N., Jarvis D., Heaviside C., Vardoulakis S., Walton H. and Anderson H.R. (2016), Long-term exposure to ambient ozone and mortality: a quantitative systematic review and meta-analysis of evidence from cohort studies, *BMJ Open*, 23, e009493.

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 MJ, Dëdelë A., Graþulevièienë 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. and de Hoogh K**.** **(**2013), Development of $NO_2$ and $NO_x$ land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project, *Atmospheric Environment*, **72**, 10–23.

Beelen R., Raaschou-Nielsen O., Stafoggia M., Andersen Z.J., Weinmayr G., Hoffmann B., Wolf K., Samoli E., Fischer P., Nieuwenhuijsen M., Vineis P., Xun W.W., Katsouyanni K., Dimakopoulou K., Oudin A., Forsberg B., Modig L., Havulinna

A.S., Lanki T., Turunen A., Oftedal B., Nystad W., Nafstad P., De Faire U., Pedersen N.L., Ostenson C.G., Fratiglioni L., Penell J., Korek M., Pershagen G., Eriksen K.T., Overvad K., Ellermann T., Eeftens M., Peeters P.H., Meliefste K., Wang M., Bueno-de-Mesquita B., Sugiri D., Krämer U., Heinrich J., de Hoogh K., Key T., Peters A., Hampel R., Concin H., Nagel G., Ineichen A., Schaffner E., Probst-Hensch N., Künzli N., Schindler C., Schikowski T., Adam M., Phuleria H., Vilier A., Clavel-Chapelon F., Declercq C., Grioni S., Krogh V., Tsai M.Y., Ricceri F., Sacerdote C., Galassi C., Migliore E., Ranzi A., Cesaroni G., Badaloni C., Forastiere F., Tamayo I., Amiano P., Dorronsoro M., Katsoulis M., Trichopoulou A., Brunekreef B. and Hoek G. (2014), Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project, *Lancet*, **383**, 785–795.

Brunekreef B., Annesi-Maesano I., Ayres J.G., Forastiere F., Forsberg B., Künzli N., Pekkanen J. and Sigsgaard T. (2012), Ten principles for clean air, *European Respiratory Society*, **39**, 525–528.

Dadvand P., Figueras F., Basagaña X., Beelen R., Martinez D., Cirach M., Schembari A., Hoek G., Brunekreef B. and Nieuwenhuijsen M.J**. (**2013), Ambient air pollution and preeclampsia: a spatiotemporal analysis, *Environmental Health Perspectives*, **121**, 1365–1371.

de Hoogh K., Chen J., Gulliver J., Hoffmann B., Hertel O., Ketzel M., Bauwelinck M., van Donkelaar A., Hvidtfeldt U.A., Katsouyanni K., Klompmaker J., Martin R.V., Samoli E., Schwartz P.E., Stafoggia M., Bellander T., Strak M., Wolf K., Vienneau D., Brunekreef B. and Hoek G. (2018), Spatial $PM_{2.5}$, $NO_2$, $O_3$ and BC models for Western Europe - Evaluation of spatiotemporal stability, *Environment International*, **120**, 81–92.

De Nazelle A., Arunachalam S. and Serre M.L**. (**2010), Bayesian Maximum Entropy Integration of Ozone Observations and Model Predictions: An Application for Attainment Demonstration in North Carolina, *Environmental Science & Technology*, **44**, 5707–5713.

Dimakopoulou K., Gryparis A. and Katsouyanni K. (2017), Using spatio-temporal land use regression models to address spatial variation in air pollution concentrations in time series studies, *Air Quality, Atmosphere & Health*, **10**, 1139–1149.

Dimakopoulou K., Samoli E., Beelen R., Stafoggia M., Andersen Z.J., Hoffmann B., Fischer P., Nieuwenhuijsen M.J., Vineis P., Xun W.W., Hoek G., Raaschou -Nielsen O., Oudin A., Forsberg B., Modig L., Jousilahti P., Lanki T., Turunen A., Oftedal B., Nafstad P., Schwarze P.E., Penell J., Fratiglioni L., Andersson N., Pedersen N., Korek M., De Faire U., Eriksen K.T., Tjønneland A., Becker T., Wang M., Bueno-de-Mesquita B., Tsai M.Y., Eeftens M., Peeters P.H., Meliefste K., Marcon A., Krämer U., Kuhlbusch T.A., Vossoughi M., Key T., de Hoogh K., Hampel R., Peters A., Heinrich J., Weinmayr G., Concin H., Nagel G., Ineichen A., Jacquemin B., Stempfelet M., Vilier A., Ricceri F., Sacerdote C., Pedeli X., Katsoulis M., Trichopoulou A., Brunekreef B., Katsouyanni K. and on behalf of the ESCAPE group. (2014), Air pollution and non-malignant respiratory mortality in 16 cohorts within the ESCAPE project, *American Journal of Respiratory and Critical Care Medicine*, **189**, 684–696.

EEA (European Environment Agency), Air quality in Europe - 2011 report, 2011 Luxembourg: Publications Office of the European Union.

Eeftens M., Beelen R., de Hoogh K., Bellander T., Cesaroni G., Cirach M., Declercq C., Dedele A., Dons E., de Nazelle A., Dimakopoulou K., Eriksen K., Falq G., Fischer P., Galassi C., Grazuleviciene R., Heinrich J., Hoffmann B., Jerrett M., Keidel D., Korek M., Lanki T., Lindley S., Madsen C., Molter A., Nador 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., Varro M.J., Vienneau D., Klot S., Wolf K., Brunekreef B. and 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, *Environmental Science & Technology*, **46**, 11195–11205.

EPA (Environmental Protection Agency) (2009), Assessment of the impacts of global change on regional U.S. air quality: A synthesis of climate change impacts on ground-level ozone(An interim report of the U.S. EPA Global Change Research Program). ePA/600/r-07/094f Washington, Dc: ePA. online at cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=203459 #Download, accessed on October 1, 2018.

Gryparis A., Coull B.A., Schwartz J. and Suh H. (2007), Latent variable semiparametric regression models for spatio-temporal modelling of mobile source pollution in the Greater Boston Area, *Applied Statistics*, **56**, 183–209.

Gryparis A., Dimakopoulou K., Pedeli X. and Katsouyanni K. (2014), Spatio-temporal semiparametric models for $NO_2$ and $PM_{10}$ concentration levels in Athens, Greece, *Science of the Total Environment*, **479–480**, 21–30.

Hystad P., Demers P.A., Johnson K.C., Brook J., van Donkelaar A., Lamsal L., Martin R. and Brauer M. (2012), Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study, *Environmental Health*, **11**, 22.

Jerrett M., Arain A., Kanaroglou P., Beckerman B., Potoglou D. and Sahsuvaroglu T. (2005), A review and evaluation of intraurban air pollution exposure models, *Journal of Exposure Analysis and Environmental Epidemiology*, **15**, 185–204.

Jerrett M., Burnett R.T., Pope CA 3rd, Ito K., Thurston G., Krewski D., Shi Y., Calle E. and Thun M. (2009), Long-term ozone exposure and mortality, *The New England Journal of Medicine*, **360**, 1085–1095.

Karakatsani A., Samoli E., Rodopoulou S., Dimakopoulou K., Papakosta D., Spyratos D., Grivas G., Tasi S., Angelis N., Thirios A., Tsiotsios A. and Katsouyanni K. (2017), Weekly Personal Ozone Exposure and Respiratory Health in a Panel of Greek Schoolchildren, *Environmental Health Perspectives*, **125**, 077016.

Katsoulis M., Dimakopoulou K., Pedeli X., Trichopoulos D., Gryparis A., Trichopoulou A. and Katsouyanni K. (2014), Long-term exposure to traffic-related air pollution and cardiovascular health in a Greek cohort study, *Science of the Total Environment*, 490, 934–940.

Kerckhoffs J., Wang M., Meliefste K., Malmqvist E., Fischer P., Janssen N.A., Beelen R. and Hoek G. (2015), A national fine spatial scale land-use regression model for ozone, *Environmental Research*, **140**, 440–448.

Kirrane E.F., Bowman C., Davis J.A., Hoppin J.A., Blair A., Chen H.L., Patel M.M., Sandler D.P., Tanner C.M., Vinikoor-Imler L., Ward M.H., Luben T.J. and Kamel F. (2015), Associations of Ozone and PM2.5 Concentrations With Parkinson's Disease Among Participants in the Agricultural Health Study, *Journal of Occupational and Environmental Medicine*, **57**, 509–517.

Kloog I., Nordio F., Zanobetti A., Coull B.A., Koutrakis P. and Schwartz J.D. (2014), Short term effects of particle exposure on hospital admissions in the Mid-Atlantic states: a population estimate, *PLoS One*, **9**, e88578.

Lee M., Koutrakis P., Coull B., Kloog I and Schwartz J. (2016), Acute effect of fine particulate matter on mortality in three southeastern states 2007–2011, *Journal of Exposure Science & Environmental Epidemiology*, **26**, 173–179.

Malmqvist E., Olsson D., Hagenbjörk-Gustafsson A., Forsberg B., Mattisson K., Stroh E., Strömgren M., Swietlicki E., Rylander L., Hoek G., Tinnerberg H. and Modig L. (2015), Assessing ozone exposure for epidemiological studies in Malmö and Umeå, Sweden, *Atmospheric Environment*, **94**, 241–248.

Maynard D., Coull B.A., Gryparis A. and Schwartz J. (2007), Mortality risk associated with short-term exposure to traffic particles and sulfates, *Environmental Health Perspectives*, **115**, 751–755.

Monks P.S**. (**2000), A review of the observations and origins of the spring ozone maximum, *Atmospheric Environment*, **34**, 3545–3561.

Ostro B., Hu J.L., Goldberg D., Reynolds P., Hertz A., Bernstein L. and Kleeman M.J**. (**2015), Associations of Mortality with Long-Term Exposures to Fine and Ultrafine Particles, Species and Sources: Results from the California Teachers Study Cohort, *Environmental Health Perspectives*, **123**, 549–556.

Puett R.C., Hart J.E., Schwartz J., Hu F.B., Liese AD and Laden F. (2011), Are particulate matter exposures associated with risk of type 2 diabetes? *Environmental Health Perspectives*, **119**, 384–389.

Raaschou-Nielsen O., Andersen Z.J., Beelen R., Samoli E., Stafoggia M., Weinmayr G., Hoff mann B., Fischer P., Nieuwenhuijsen M.J., Brunekreef B., Xun W.W., Katsouyanni K., Dimakopoulou K., Sommar J., Forsberg B., Modig L., Oudin A., Oftedal B., Schwarze P.E., Nafstad P., De Faire U., Pedersen N.L., Ostenson C.G., Fratiglioni L., Penell J., Korek M., Pershagen G., Eriksen K.T., Sorensen M., Tjonneland A., Ellermann T., Eeftens M., Peeters P.H., Meliefste K., Wang M., Bueno-de-Mesquita B., Key T.J., de Hoogh K., Concin H., Nagel G., Vilier A., Grioni S., Krogh V., Tsai M.Y., Ricceri F., Sacerdote C., Galassi C., Migliore E., Ranzi A., Cesaroni G., Badaloni C., Forastiere F., Tamayo I., Amiano P., Dorronsoro M., Trichopoulou A., Bamia C., Vineis P. and Hoek P. (2013), Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE), *The Lancet Oncology*, **14**, 813–822.

Samoli E., Dimakopoulou K., Evangelopoulos D., Rodopoulou S., Karakatsani A., Veneti L., Sionidou M., Tsolakoglou I., Krasanaki I., Grivas G., Papakosta D. and Katsouyanni K. (2017), Is daily exposure to ozone associated with respiratory morbidity and lung function in a representative sample of schoolchildren? Results from a panel study in Greece, *Journal of Exposure Science & Environmental Epidemiology*, 27, 346–351.

Stedman J.R. and Kent A.J**. (**2008), An analysis of the spatial patterns of health related surface ozone metrics across the UK in 1995, 2003 and 2005, *Atmospheric Environment*, **42**, 1702–1716.

Schwartz J. (2016), The Year of Ozone, *American Journal of Respiratory and Critical Care Medicine*, **193**, 1077–1079.

Wang M., Beelen R., Basagana X., Becker T., Cesaroni G., de Hoogh K., Dedele A., Declercq C., Dimakopoulou K., Eeftens M., Forastiere F., Galassi C., Gražulevičienė R., Hoffmann B., Heinrich J., Iakovides M., Künzli N., Korek M., Lindley S., Mölter A., Mosler G., Madsen C., Nieuwenhuijsen M., Phuleria H., Pedeli X., Raaschou-Nielsen O., Ranzi A., Stephanou E., Sugiri D., Stempfelet M., Tsai M.Y., Lanki T., Udvardy O., Varró MJ, Wolf K., Weinmayr G., Yli-Tuomi T., Hoek G. and Brunekreef B. *(***2013),** Evaluation of Land Use Regression Models for NO2 and Particulate Matter in 20 European Study Areas: The ESCAPE Project, *Environmental Science & Technology,* ***47*****,** 4357–4364.

Wang M., Sampson P.D., Hu J., Kleeman M., Keller J.P., Olives C., Szpiro A.A., Vedal S. and Kaufman J.D. (2016), Combining Land-Use Regression and Chemical Transport Modeling in a Spatiotemporal Geostatistical Model for Ozone and PM2.5, *Environmental Science & Technology*, 50, 5111–5118.

WHO (World Health Organization).(2008), Health Risks of Ozone from Long-Range Transboundary Air Pollution, Regional Office for Europe, Copenhagen, Denmark.

WHO (World Health Organization). (2013), Review of Evidence on Health Aspects of Air Pollution REVIHAAP Project: Final Technical Report, WHO Europe, Copenhagen, Denmark.