Article

Development of Land Use Regression Models for elemental, organic carbon, PAH and hopanes/steranes in 10 ESCAPE/TRANSPHORM European study areas

Aleksandra Jedynska, Gerard Hoek, Meng Wang, Marloes Eeftens, Josef Cyrys, Menno Keuken, Christophe Ampe, Rob Beelen, Giulia Cesaroni, Francesco Forastiere, Marta Cirach, Kees de Hoogh, Audrey De Nazelle, Wenche Nystad, Christophe Declercq, Kirsten Thorup Eriksen, Konstantina Dimakopoulou, Timo Lanki, Kees Meliefste, Mark J Nieuwenhuijsen, Tarja Yli-Tuomi, Ole Raaschou-Nielsen, Bert Brunekreef, and Ingeborg Kooter

Environ. Sci. Technol., **Just Accepted Manuscript** • DOI: 10.1021/es502568z • Publication Date (Web): 15 Oct 2014

	Lanki, Timo; National Institute for Health and Welfare, Department of Environmental Health Meliefste, Kees; Utrecht University, Institute for Risk Assessment Sciences (IRAS) Nieuwenhuijsen, Mark; IMIM (Hospital del Mar Research Institute), ; Center for Research in Environmental Epidemiology (CREAL), ; CIBER Epidemiología y Salud Pública , Yli-Tuomi, Tarja; National Institute for Health and Welfare, Department of Environmental Health Raaschou-Nielsen, Ole; Danish Cancer Society, Institute of cancer epidemiology Brunekreef, Bert; Utrecht University, Institute for Risk Assessment Sciences (IRAS); University Medical Center Utrecht, Julius Center for Health Sciences and Primary Care Kooter, Ingeborg; TNO, Netherlands Organization for Applied Scientific Research,

Development of Land Use Regression Models for elemental, organic carbon, PAH and hopanes/steranes in 10 ESCAPE/TRANSPHORM European study areas

Aleksandra Jedynska [†*], Gerard Hoek [‡], Meng Wang [‡,♦], Marloes Eeftens [‡,•,▪], Josef Cyrys [§,ǁ], Menno Keuken [†], Christophe Ampe [δ], Rob Beelen [‡], Giulia Cesaroni [%], Francesco Forastiere [%], Marta Cirach [@,&,○], Kees de Hoogh [Σ,•], Audrey De Nazelle [@,∞], Wenche Nystad [β], Christophe Declercq [α], Kirsten T. Eriksen [π], Konstantina Dimakopoulou [χ], Timo Lanki [ϕ], Kees Meliefste [‡], Mark J. Nieuwenhuijsen [@,&,○], Tarja Yli-Tuomi [ϕ], Ole Raaschou-Nielsen [π], Bert Brunekreef [‡,ψ], Ingeborg M. Kooter [†]

[†] TNO, The Netherlands Applied Research Organization, Utrecht, The Netherlands

[‡] Institute for Risk Assessment Sciences, Utrecht University, The Netherlands[§] Helmholtz Centre for Environmental Research, Munich, Germany

[ǁ] University of Augsburg, Environmental Science Center, Augsburg, Germany

[δ] AIRPARIF, Paris, France

[%] Epidemiology Department, Lazio Regional Health Service, Rome, Italy

[@] Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain

[&] IMIM (Hospital del Mar Research Institute), Barcelona, Spain

[○] CIBER Epidemiolog ía y Salud Pública (CIBERESP), Spain

[Σ] MRC-HPA Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, London, United Kingdom

[β] Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway

[α] French Institute for Public Health Surveillance (InVS), Saint-Maurice Cedex, France

[π] Danish Cancer Society Research Center, Copenhagen, Denmark

[χ] Department of Hygiene, Epidemiology & Medical Statistics, Medical School, National and Kapodistrian University of Athens, Greece

[ϕ] Department of Environmental Health, National Institute for Health and Welfare (THL), Kuopio, Finland

[ψ] Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

[•] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

[▪] University of Basel, Basel, Switzerland

[∞] Centre for Environmental Policy, Imperial College London, UK

[♦] Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, The United States

Abstract

Land use regression (LUR) models have been used to model concentrations of mainly traffic related air pollutants (nitrogen oxides (NO_x), particulate matter (PM) mass or absorbance). Few LUR models are published of PM composition, whereas the interest in health effects related to particle composition is increasing.

The aim of our study was to evaluate LUR models of polycyclic aromatic hydrocarbons (PAH), hopanes/steranes and elemental and organic carbon (EC/OC) content of $PM_{2.5}$.

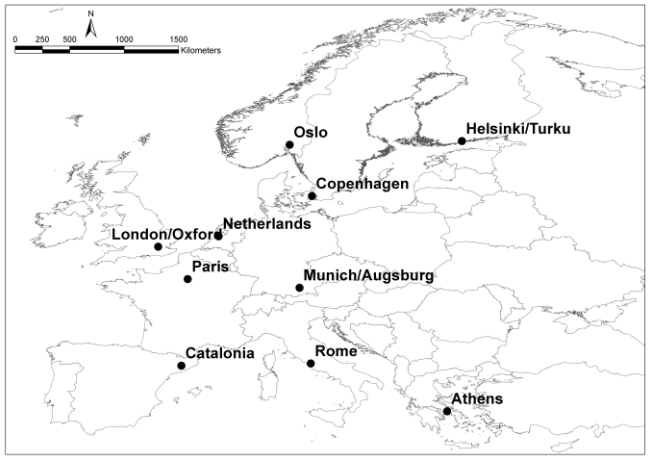
In 10 European study areas PAH, hopanes/steranes and EC/OC concentrations were measured at 16–40 sites per study area. LUR models for each study area were developed based on annual average concentrations and predictor variables including traffic, population, industry, natural land obtained from geographic information systems.

The highest median model explained variance (R^2) was found for EC – 84%. The median R^2 was 51% for OC, 67% for benzo[a]pyrene and 38% for sum of hopanes/steranes, with large variability between study areas. Traffic predictors were included in most models. Population and natural land were included frequently as additional predictors. The moderate to high explained variance of LUR models and the overall moderate correlation with $PM_{2.5}$ model predictions support the application of especially the OC and PAH models in epidemiological studies.

Keywords:
EC, OC, PAH, Land Use Regression model, PM2.5, Spatial variation

*

Abstract Art



INTRODUCTION

Many studies have documented adverse health effects of exposure to ambient air pollution [1-3]. Particulate matter (PM) with diameters smaller than 10 or 2.5 µm (PM_{10}, $PM_{2.5}$, respectively) is the most used parameter for assessment of air quality in epidemiological studies. However, PM is a chemically complex mixture and it has been suggested that observed adverse health effects depend on PM chemical composition [4,5]. Epidemiological studies have started to assess chemical composition of particles, but few studies have assessed the relationship between specific organic components and long-term adverse health effects. In the recently published studies from the US National Particle Components Toxicity (NPACT) Initiative on health effects of particle composition, the focus was on elemental composition supplemented with elemental and organic carbon measurements [6,7]. Limited availability of measurements with a sufficient spatial resolution and models of organic components of particles have contributed to the small number of studies that have assessed health effects of organic PM components.

One approach, which has been used in air quality studies, is to quantify elemental (EC) and organic carbon (OC) as more generic indicators of air quality. EC is a highly polymerized dark fraction which is resistant to oxidation at temperatures below 400°C [8]. EC is used as an indicator for traffic diesel emissions and correlates highly with black smoke (BS), black carbon (BC) and PM absorbance [9-12]. EC has been linked in epidemiological studies with short and long-term health effects [13]. OC in contrast is a mixture of organic compounds such as aliphatic and aromatic hydrocarbons and acids. OC has been included in epidemiological studies less often than elemental carbon. There is evidence of OC associations with short-term adverse health effects [14,15]. Recently long-term exposure to OC was associated with heart disease and pulmonary mortality [16] and cardiovascular mortality [6].

Polycyclic aromatic hydrocarbons (PAH) are persistent organic components formed during incomplete combustion processes. Sources include: domestic, transportation, industrial, and agricultural processes. Domestic emissions arise from burning wood, coal, oil and garbage for heat and waste disposal. Road and maritime transportation can be significant sources of PAH depending on engine type and age, and fuel used. Industrial PAH emissions originate mainly from metal and steel production, the petrochemical industry and power generation. PAH from agriculture derive mainly from open biomass burning [17]. PAH has been associated with various adverse health effects [18,19] but has not been evaluated in the major cohort studies on long-term air pollution exposure and mortality / lung cancer. The EU limits the annual average concentration of benzo[a]pyrene (B[a]P) as a proxy for all PAH in ambient air to 1 ng/m^3 [20]

Hopanes and steranes are components present in crude oil. They have been used as markers for motor vehicle exhaust particles in the atmosphere. Hopanes and steranes are also present in lubricating oil used by both gasoline- and diesel-powered motor vehicles, and are also found in diesel fuel [21]. Hopanes and steranes have been associated with lung toxicity [22], but have not been used as exposure metrics in epidemiological studies.

Land use regression models are used to describe spatial variation of the annual concentration of a pollutant, often as a tool for exposure assessment in epidemiological studies [23]. Most modeled pollutants are $PM_{2.5}$, PM_{10} and compounds strongly affected by traffic (NO_x, PM absorbance,

EC)[24,25]. Few LUR models have been developed for pollutants for which traffic is not a major source. Recently, LUR models were developed for the elemental composition of $PM_{2.5}$ and PM_{10} in 20 European study areas [26]. Three North American studies presented LUR models for wood smoke [27-29]. Within the framework of NPACT, LUR models for EC and OC were developed. LUR models for PAH concentrations were developed in an American study [30].

Within the framework of two European projects: ESCAPE (European Study of Cohorts for Air Pollution Effects) and TRANSPHORM (Transport related Air Pollution and Health impacts - Integrated Methodologies for Assessing Particulate Matter) concentrations of NO_x, NO_2, $PM_{2.5}$, PM_{10}, $PM_{2.5}$ absorbance and elemental composition of $PM_{2.5}$ and PM_{10}. were measured in 20 European study areas. The spatial variation and LUR models for these pollutants have been published [24-26,31]. The ESCAPE LUR models have been used to assess exposure for epidemiological studies [32]. Additionally specific organic components (EC/OC, PAH and hopanes/steranes) were measured in 10 of the areas. The spatial variation of concentrations within and between the ten European study areas was published [33].

The aim of this paper is the development and evaluation of LUR models for EC, OC, PAH and hopanes/steranes. A second aim was to assess the correlation of the predictions of the new models with the already published model predictions for $PM_{2.5}$ and NO_x.

METHODS

Sampling campaign and analyses. The ESCAPE sampling campaign including the organic component characterization was described previously[31,33,34]. Briefly, the monitoring campaign was conducted in 10 study areas across Europe between October 2008 and April 2011 (Table 1, online supplement figure S1) and common standardized procedures were used across the study areas. Maps of the study areas can be found in the supporting information of Eeftens et al. [31]. At each sampling site, three two-weekly samples were collected over a period of one year. Samples were taken during three different seasons: winter, summer and intermediate season – either spring or autumn. Monitoring was performed using the Harvard impactor, which collects $PM_{2.5}$ and PM_{10} on separate Teflon filters using an air flow of approximately 10 L/min. For extended $PM_{2.5}$ characterization two additional samples were collected: one on a Teflon coated glass fiber filter (T60A20, Pallflex) for analysis of organic components (PAH, hopanes/steranes) [33] and one on a quartz filter (QMA,Whatman) for EC/OC, oxidative potential and levoglucosan quantification. In most study areas, PM concentrations were measured at 20 monitoring sites. In Catalonia 40 sites were measured and in The Netherlands organic components were measured at 16 of the 40 ESCAPE sites because of lack of additional sampling equipment. In each study area, NO_x only measurements were conducted at 20 locations (40 in Catalonia and the Netherlands). Sampling sites were divided into three groups according to traffic intensity: street sites (S), urban background (UB) and regional background (RB), using common criteria. One reference site, located in an urban or rural background location depending on the study area, was established to measure continuously for 2-week periods during a full year to adjust for temporal variation. This site was used in the calculation of the temporally adjusted annual average concentrations.

Annual averages were calculated after adjusting for temporal variation measured at the continuous reference site. For each of the three sampling periods, the ratio of the concentration measured in that period from the overall annual mean at the reference site was used as adjustment, following previously published procedures [33].

Table 1. Description of study areas

Country	Study area	Sampling period	Sites	Site types		
				RB	UB	S
Norway	Oslo	05.02.2009 – 29.01.2010	19	2	9	8
Finland	Helsinki/Turku	27.01.2010 – 26.01.2011	20	2	10	8
Denmark	Copenhagen	19.11.2009 – 17.11.2010	20	3	6	11
United Kingdom	London/Oxford	26.01.2010 – 18.01.2011	20	1	12	7
The Netherlands	Rotterdam, Amsterdam, Groningen, Amersfoort	17.02.2009 – 19.02.2010	16	4	4	8
Germany	Munich/Augsburg	01.03.2009 – 05.11.2009	20	5	6	9
France	Paris	04.01.2010 – 04.01.2011	20	4	9	7
Italy	Rome	27.01.2010 – 26.01.2011	20	2	8	10
Spain	Catalonia (Barcelona, Girona, Sabadell)	14.01.2009 – 14.01.2010	40	4	13	23
Greece	Athens	21.04.2010 – 27.04.2011	20	1	12	7

Analytical methods. Analytical methods were previously described [33]. Briefly, quartz filters were used for EC/OC analyses, which were completed via a thermal-optical analyzer (Sunset Laboratory, Inc., Oregon, USA). The EUSAAR2 protocol was used for the temperature settings [35]. T60A20 filters were extracted via an accelerated solvent extraction method (ASE) with toluene. Extracts were fractioned into three fractions via a silica column. This separated hopanes/steranes from PAH and PAH derivatives. 16 EPA PAH and 13 hopanes/steranes were analyzed via gas chromatography in combination with mass spectrometric detection (GS/MS) in electron impact mode (GC/MS EI, Agilent 6890/5973N). From the 16 EPA PAH measured, eight particle-related PAH were quantified, as our sampling system did not quantitatively capture semi-volatile PAH well. ∑PAH was determined as the sum of eight particle-related PAH: benzo[a]anthracene, chrysene, benzo[b]fluoranthene, benzo[k]fluoranthene, benzo[a]pyrene, indeno[1,2,3-c,d]pyrene, dibenzo[a,h]anthracene and benzo[g,h,i]perylene. For budgetary reasons, hopanes/steranes were only measured in four study areas.

GIS predictor data. The description of available geographical information system (GIS) predictor variables has been presented in detail [24,25]. The predictor variables chosen for LUR development describe potential emission sources like traffic, industry or population density. The values of predictor variables were determined for each sampling site using GIS. Geographic data were obtained from two sources: central and local. Central data sets included: information on roads with a road

classification (EuroStreets version 3.1), land use (CORINE land cover 2000), altitude (SRTM 90m Digital Elevation Data), and population (enhanced EEA population density data using CORINE land cover 2000). If available, local GIS data were collected on road network with linked traffic intensities, land use, population density and altitude. Each variable was calculated in several circular buffers from the sampling site (e.g. 25, 50, 100). Traffic variable buffers started at 25m because significant variation has been found at small distances from roads. Land use variable buffers started at 100 meter as these variables may represent variability in urban background concentrations. Variables with at least 5 non-zero values were used in model development. A detailed description of predictor variables including the buffer sizes and the a priori specified direction of effect on the pollutant concentration are presented in online supplement table S1.

LUR model development. LUR models for all study areas were developed centrally at IRAS. We followed the method used in the ESCAPE study Briefly, adjusted annual average concentrations of pollutants and predictor variables were used for LUR development. A supervised stepwise method was used to obtain the linear regression model with the highest adjusted explained variance (R^2). First, all predictor variables were offered separately and the variable with the highest R^2 having a slope in the a priori specified direction was selected. Next variables were added to the model one by one based on the highest adjusted R^2 if it improved model's adjusted R^2 with at least 1% and had the same effect direction as decided a priori.

The following diagnostics were used for the final model check: significance of individual variables (p-value <0.1), collinearity test (variance inflation factor (VIF) lower than 3) and influential observation test (Cook's D value lower than 1) and Morans'I – spatial autocorrelation of the model residuals. If an influential observation determined by a too high Cook's D value was caused by extreme measured concentrations, we additionally developed LUR models without the observation. An outlier was defined as a concentration higher than the 75th percentile plus four times the interquartile range: $c_o > P75 + 4*(P75-P25)$, where P75 – 75th percentile, P25 – 25th percentile. The performance of the final model was evaluated with leave-one-out cross validation (LOOCV). Each site was sequentially left out from the model while the included variables were left unchanged.

Predicted pollutant concentration were compared with measured concentration at the excluded site. This procedure was repeated for all sites included in the model development.

We developed LUR models for EC, OC, ∑PAH, B[a]P, chrysene and ∑hopanes/steranes. As individual PAHs correlated with each other very highly in all study areas we developed LUR models for only two individual PAHs: B[a]P and chrysene in addition to ΣPAH.

B[a]P is used in air quality guidelines as a surrogate for all PAH[20]. Chrysene was chosen because of its relatively high concentrations in comparison with other PAH, its higher correlation with traffic markers (Table S2) [33].

To further assess the added value of the organic components to characterize the air pollution mixture, the correlation of LUR model predictions with predictions of the already published LUR models for fine particles was calculated. We selected $PM_{2.5}$ and $PM_{2.5}$ absorbance. The correlation was calculated at the (20 – 40) NO_2 measurement sites in each area which were not used for model development but did have GIS predictor variables.

RESULTS

Measured concentrations of pollutants. Substantial variability was found in measured concentrations of EC, OC, ΣPAH, B[a]P, chrysene and ∑hopanes/steranes within and between study areas (Figure 1, Table S2). The highest concentrations of EC and OC were found in southern Europe and the lowest in northern Europe. PAH concentrations exhibited a different trend in Europe, with similar concentrations in southern and northern Europe. For all components higher concentrations were measured at street locations in comparison to urban background and regional background locations [33].

Land use regression modelling. The final models of EC, OC, ∑PAH, B[a]P, chrysene and ∑hopanes/steranes are presented in Tables 2 to 4 and Supplementary information tables S4, S5 and S6. Figures S2 to S7 show predicted versus measured concentrations.

EC models. In all 10 study areas LUR models could be developed (Table 2). Explained variance was high in all areas, ranging from 73% for Catalonia to 95% in Oslo and the Netherlands (median 87%). The median LOOCV R^2 was 80%. Differences between R^2 and LOOCV R^2 was lower than 11% in all study areas indicating that models were stable. In all models traffic related variables were used, mostly describing local traffic (nearest road or buffer ≤100m). In 8 of the 10 models, traffic intensity was included, sometimes in combination with road length variables. The Rome model contained only traffic variables, but all other models included non-traffic predictors such as population or address density often in large buffers (1000 or 5000m). In Helsinki, Rome and Paris single street sites were very influential because of very high EC concentrations which were identified as outliers (Figure 1). Models are presented without the outliers.

OC models. In all study areas LUR models for OC could be developed (Table 3). In general explained variance of OC LUR models was lower than for EC models. The explained variance varied from 27% in Rome to 81% in London/Oxford (median 59%). The difference between model R^2 and LOOCV R^2 was larger than for EC indicating less stable models. In 8 models traffic variables were used, of which 2 models contained only traffic predictors (Rome, London). Most models contained non-traffic predictors particularly natural land and population / address density variables, often in smaller buffers than for EC. In Copenhagen and Helsinki/Turku, models with a single non-traffic predictor were identified.

B[a]P models. In two locations -Copenhagen and Helsinki/Turku- LUR models could not be developed due to influential sites which were not identified as outliers (Table 4). The median model R^2 was 67%, with a range from 31% (Munich/Augsburg) to 87% (London/Oxford). In all 8 developed models traffic variables were used. Three models included only traffic predictors (Munich, Rome and Athens). In five models, population or natural land use were additionally included.

ΣPAH models. In general, ΣPAH model performance and structure was very similar to B[a]P models (Supplemental Information Table S5). Because of the same sites as in the case of B[a]P models it was not possible to develop models in Helsinki/Turku and Copenhagen. For the eight ΣPAH models a moderate median R^2 of 65% was found.

Chrysene models. Chrysene models were slightly better than B[a]P models (Table S5). Fewer areas had low cross-validation R^2 compared to B[a]P. Models could be developed for 9 study areas. Only for Copenhagen it was not possible to build a LUR model due to an influential site. The median explained variance of the chrysene models was 71%. In Helsinki/Turku and Athens models with only one variable were developed with the low R^2. In eight models, traffic variables were included. Only in Paris three non-traffic variables were used. The better performance compared to B[a]P is likely due to the larger impact of traffic on chrysene.

Σhopanes/steranes models. Σhopanes/steranes models were developed in all four study areas with measurements available (Supplemental Information Table S6). The models had the lowest explained variance of all modeled components (median R^2 = 43%). In Oslo and Catalonia, models with low R^2 (<30%) were developed including one non-traffic variable. In The Netherlands, the model with the highest R^2 was developed – 78%. The model included one traffic variable, population and they coordinate indicating lower concentrations in the north. The Munich/Augsburg model included two traffic variables and a coordinate indicating higher concentrations in the east.

Additional analyses

Most models explained more variation for street than for urban background locations (figures S2 – S7). The average model R^2 was 19% higher at street sites. Averaged over the five pollutants measured in 10 areas, the models with the lowest R^2 were found in Helsinki and Copenhagen and the models with the highest R^2 in London and Paris (Table S7).

Comparison of LUR PM component model predictions and LUR predictions from $PM_{2.5}$, $PM_{2.5}$ absorbance models.

Correlations between predictions of the newly developed LUR PM component models with already published model predictions for $PM_{2.5}$, $PM_{2.5}$ and absorbance at monitoring sites not used in model development are shown in Table5. Correlations varied with study area and component, e.g. in London most correlations were high and in Munich correlations with $PM_{2.5}$ model predictions were very low. Moderate to high correlations were found between the EC

and PAH models and $PM_{2.5}$ absorbance. OC correlated less with $PM_{2.5}$ absorbance, but showed a moderately high correlation with $PM_{2.5}$.

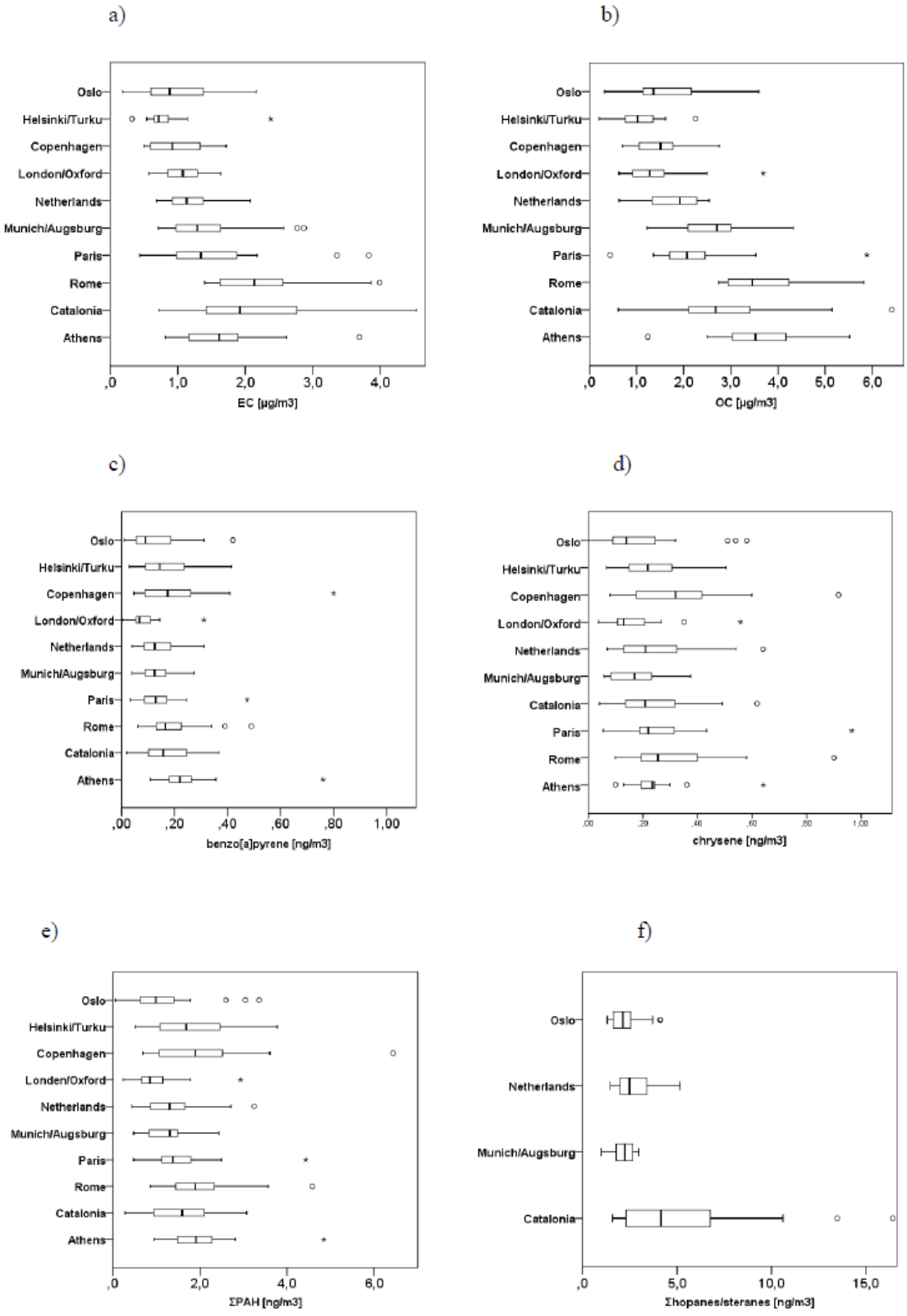


Figure 1. Distribution of the annual average measured concentrations within study areas a) EC, b) OC, c) B[a]P, d) chrysene, e) ΣPAH, f) Σhopanes /steranes [34]

Table 2. Description of land use regression models for EC ($\mu g/m^3$)

Study area	LUR model	n	R^2 (%)	LOOCV R^2 (%)	RMSE ($\mu g/m^3$)	Measured concentration ($\mu g/m^3$)
Oslo	0.915 + 3.444 x 10^{-12} x TRAFLOAD_1000 + 8.185 x DISTINVMAJOR1 - 0.0342 x SQRALT	19	95	92	0.141	0.99 [0.18 - 2.16]
Helsinki/Turku	0.738 - 2.383 x 10^{-7} x URBNATURAL_1000 + 0.000293 x ROADLENGTH_100 + 0.000595 x ROADLENGTH_50	19	79	68	0.087	0.81 [0.32 - 2.38]
Copenhagen	0.526 + 0.0002764 x MAJORROADLENGTH_300 + 3.622 x 10^{-8} x TRAFMAJORLOAD_100 - 1.805 x 10^{-7} x URBNATURAL_1000	20	87	78	0.146	0.98 [0.49 - 1.72]
London/Oxford	0.692 + 0.000237 x INTMAJORINVDIST + 0.000143 x HEAVYTRAFMAJOR + 43.793 x DISTINVMAJORC2 + 1.111 x 10^{-8} x HDRES_5000	19	85	77	0.131	1.32 [0.57 - 5.84]
Netherlands	0.583 + 0.00116 x ROADLENGTH_50 + 3.114 x DISTINVMAJOR1 + 9.992 x 10^{-7} x POP_5000	16	95	90	0,103	1.21 [0.68 - 2.07]
Munich/Augsburg	0.821 + 0.00572 x MAJORROADLENGTH_50 + 0.0000139 x TRAFNEAR + 7.699 x 10^{-7} x POP_5000	20	86	81	0.265	1.45 [0.71 - 2.87]
Paris	0.689 + 3.968 x 10^{-7} x TRAFLOAD_50 + 9.750 x 10^{-6} x POP_1000 + 3.422 x 10^{-8} x INDUSTRY_5000	19	95	91	0.215	1.82 [0.44 - 8.08]
Rome	1.582 + 3.313 x 10^{-7} x TRAFMAJORLOAD_50 + 0.000330 x MAJORROADLENGTH_300	20	75	65	0.430	2.26 [1.40 - 3.99]
Catalonia	1.579 + 0.000521 x INTMAJORINVDIST1- 2.343 x 10^{-8} x NATURAL_5000 + 2.536 x 10^{-7} x HDRES_1000	40	73	66	0.604	2.15 [0.72 - 5.50]
Athens	0.125 + 9.366 x 10^{-7} x TRAFLOAD_25 + 2.820 x 10^{-7} x PORT_5000 + 0.0108 x DISTINVNEARC2 + 0.00001837 x TRAFMAJOR + 1.683 x 10^{-8} x HDRES_5000	20	93	91	0.207	1.62 [0.81 - 3.69]
Median			87	80		

Description of variables used in the models: DISTINVMAJOR1 Inverse distance to the nearest major road (local road network), major road – road with intensity > 5,000 mvh/24h, DISTINVMAJOR2 Inverse distance to the nearest major road squared (local road network), DISTINVMAJORC1 Inverse distance to the nearest major road (central road network), DISTINVMAJORC2 Inverse distance to the nearest major road squared (central road network), INTMAJORINVDIST1 Product of traffic intensity on nearest major road and inverse of distance to the nearest major road, DISTINVNEARC2 Distance to the nearest major road squared (central road network), INTINVDIST Product of traffic intensity on nearest road and inverse of distance to the nearest road, HEAVYINTINVDIST2 Product of Heavy-duty traffic intensity on nearest road and inverse of squared distance to the nearest road, heavy duty traffic – trucks. TRAFLOAD Total traffic load of all roads in a buffer (sum of (traffic intensity * length of all segments), TRAFNEAR Traffic intensity on nearest road, TRAFMAJOR Traffic intensity on nearest major road, ROADLENGHT Road length of all roads in a buffer, MAJORROADLENGTH Road length of major roads in a buffer, TRAFMAJORLOAD Total traffic load of major roads in a buffer (sum of (traffic intensity * length of all segments), HEAVYTRAFMAJOR Heavy-duty traffic intensity on nearest major road, URBNATURAL Urban green area, NATURAL Semi-natural and forested areas, UGNL Combined urban green and natural land, PORT Port, HDRES High density residential land, LDRES Low density residential land, POP Number of inhabitants, INDUSTRY Industry area, SQRALT Squared altitude, BUILDINGS_L Area of buildings, YCOORD, XCOORD, XPLUSY Coordinate variables

Table 3. Description of land use regression models for OC ($\mu g/m^3$)

Study area	LUR model	n	R^2 (%)	LOOCV R^2 (%)	RMSE ($\mu g/m^3$)	Measured concentration ($\mu g/m^3$)
Oslo	0.732 - 0.00000573 x NATURAL_300 + 0.00683 x ROADLENGTH_50 + 2.429 x INTMAJORINVDIST	19	59	34	0.621	1.65 [0.32 - 3.60]
Helsinki/Turku	1.393 - 0.00000580 x URBNATURAL_300L	20	42	32	0.362	1.08 [0.21 - 2.25]
Copenhagen	0.410 + 0.00000488 x BUILDINGS_L_300	20	35	25	0.441	1.45 [0.70 -2.75]
London/Oxford	0.886 + 109.599 x DISTINVMAJORC2 + 0.000181 x HEAVYTRAFMAJOR	20	81	74	0.326	1.39 [0.63 - 3.69]
Netherlands	0.480 + 4.798 x 10^{-7} x LDRES_1000 + 0.00397 x ROADLENGTH_25	16	80	71	0.279	1.78 [0.63 - 2.54]
Munich/Augsburg	-112.391 - 0.406 x SQRALT + 0.000224 X INTINVDIST + 0.0000126 x XPLUSY	20	59	39	0.506	2.66 [1.22 - 4.33]
Paris	0.894 + 0.000314 x MAJORROADLENGTH_500 + 0.0000278 x LDRES_100	20	79	69	0.528	2.23 [0.44 - 5.89]
Rome	3.396 + 3.3128 x 10^{-7} x TRAFLOAD_50	20	27	10	0.778	3.70 [2.73 - 5.82]
Catalonia	2.602 + 0.00360 x INTMAJORINVDIST2 - 8.237 x 10^{-7} x NATURAL_1000	40	44	30	0.895	2.76 [0.62 - 6.41]
Athens	0.812 + 0.00177 x MAJORROADLENGTH_100 + 0.000896 x ROADLENGTH_100 + 5.199 x 10^{-9} x TRAFLOAD_500	20	77	68	0.472	3.53 [1.24 - 5.52]
Median			59	37		

Table 4. Description of land use regression models for B[a]P (ng/m^3)

Study area	LUR model	n	R^2 (%)	LOOCV R^2 (%)	RMSE (ng/m^3)	Measured concentration (ng/m^3)
Oslo	0.158 + 9.402 x 10^{-13} x TRAFLOAD_1000 - 0.00848 x SQRALT	19	70	57	0.0713	0.14 [0.01 - 0.42]
Helsinki/Turku*	-	20				0.78 [0.03 - 0.42]
Copenhagen*	-	20				0.21 [0.05 - 0.80]
London/Oxford	0.0666 + 0.0000350 x INTMAJORINVDIST - 2.165 x 10^{-9} x URBGREEN_5000 + 0.0000131 x MAJORROADLENGTH_500	20	87	78	0.0251	0.09 [0.00 - 0.31]
Netherlands	0.143 + 0.000695 x MAJORROADLENGTH_50 - 4.151 x 10^{-9} x UGNL_5000	16	64	39	0.0489	0.14 [0.04 - 0.31]
Munich/Augsburg**	0.108 + 0.000255 x MAJORROADLENGTH_100	20	31	14	0.0501	0.13 [0.04 - 0.28]
Paris**	0.0108 + 0.0000250 x MAJORROADLENGTH_500 + 0.00000144 x LDRES_100 - 7.793 x 10^{-9} x URBGREEN_5000 + 3.208 x 10^{-9} x LDRES_5000	20	87	66	0.0400	0.14 [0.03 - 0.47]

Rome	0.111 + 2.169 x DISTINVMAJOR1 + 0.000174 x MAJORROADLENGTH_100	20	77	69	0.0548	0.19 [0.06 - 0.49]
Catalonia	0.0752 + 0.0000311 x INTMAJORINVDIST1 + 1.115 x 10^{-7} x HDRES_500	40	39	29	0.0732	0.17 [0.02 - 0.37]
Athens	0.176 + 1.363 x 10^{-8} x TRAFMAJORLOAD_25 + 7.517 x 10^{-11} x TRAFMAJORLOAD_500	19	38	18	0.0574	0.25 [0.11 - 0.76]
Median			67	48		

* no model possible due to an influential site
** too high Cook's D value

338 Table. 5 Correlations (R^2) between predicted values of PM component and LUR $PM_{2.5}$ and $PM_{2.5}$absorbance at
339 the NO_x only sites not used for model development in each study area
340

	Component	EC	OC	ΣPAH	B[a]P	Chrysene	Σhopanes/steranes
Oslo	PM2.5	0.60**	0.36**	0.54**	0.55**	0.47**	0.34**
	PM2.5abs	0.31*	0.06	0.30*	0.30*	0.14	0.28*
Helsinki/Turku	PM2.5	0.18	0.07	NM	NM	0.27*	NA
	PM2.5abs	0.24*	0.25*	NM	NM	0.47**	NA
Copenhagen	PM2.5	0.49**	0.67**	NM	NM	NM	NA
	PM2.5abs	0.83**	0.26*	NM	NM	NM	NA
London/Oxford	PM2.5	0.85**	0.60**	0.66**	0.84**	0.64**	NA
	PM25abs	0.73**	0.86**	0.84**	0.68**	0.81**	NA
Netherlands	PM2.5	0.18*	0.02	0.21**	0.35**	0.26**	0.11
	PM2.5abs	0.42**	0.06	0.11*	0.23**	0.15*	0.18*
Munich/Augsburg	PM2.5	0.00	0.07	0.00	0.00	0.03	0.39**
	PM2.5abs	0.54**	0.00	0.59**	0.59**	0.86**	0.25*
Paris	PM2.5	0.59**	0.69**	0.61**	0.60**	0.13	NA
	PM2.5abs	0.92**	0.43**	0.36**	0.36**	0.09	NA
Rome	PM2.5	0.49**	0.79**	0.91**	0.85**	0.10	NA
	PM2.5abs	0.68**	0.85**	0.71**	0.70**	0.23*	NA
Catalonia	PM2.5	0.71**	0.47**	0.50**	0.50**	0.50**	0.19**
	PM2.5abs	0.91**	0.56**	0.72**	0.72**	0.72**	0.30**
Athens	PM2.5	0.56**	0.47**	0.46**	0.27*	0.17	NA
	PM25abs	0.67**	0.38**	0.36**	0.56**	0.51**	NA
Median	**PM2.5**	**0.53**	**0.47**	**0.52**	**0.53**	**0.22**	**0.27**
	PM2.5abs	**0.68**	**0.38**	**0.48**	**0.58**	**0.37**	**0.27**

341
342 NA - Σhopanes/steranes not measured
343 NM – no model possible for organic component
344 ** p value< 0.01
345 * p value < 0.05
346

DISCUSSION

We developed LUR models for PAH, EC, OC and Σhopanes/steranes in ten study areas across Europe. LUR models were developed with high (EC), moderate (OC, ∑PAH) and low (Σhopanes/steranes) explained variance. For EC explained variance of the models was high in all study areas. For the other components large differences were found in levels of explained variance. Correlations between predictions of the newly developed LUR PM component models with model predictions for $PM_{2.5}$, $PM_{2.5}$ absorbance varied between components and areas. EC and PAH model predictions correlated moderately to highly with especially $PM_{2.5}$ absorbance. OC model predictions correlated highest with $PM_{2.5}$ model predictions.

EC models. The explained variability for EC was the same (R^2=87%) as the R^2 reported previously for $PM_{2.5}$ absorbance for the same study areas [25] and higher than for the NO_x and $PM_{2.5}$ models (R^2=80% and 74%, respectively). EC model predictions correlated high with the $PM_{2.5}$ absorbance model predictions (median R^2 0.68 and 0.72, respectively). Predictor variables were similar though not identical as in the absorbance models. The high correlations between absorbance and EC model predictions is consistent with those of the measurements [33]. Traffic-related pollutants can be modeled effectively with land use regression models, because availability of traffic intensity and-or road category and length data representing source strength is relatively good (e.g. compared to wood burning); traffic emissions are emitted at low height and traffic emission factors are relatively homogeneous (e.g. compared to industrial emissions). In our study, we further overrepresented traffic sites (35-50% of the all sites) because of prior evidence that motorized traffic was an important source of intra-urban spatial variation. Previous studies have also generally reported high explained variances for models for EC or surrogates of EC [23]. In some areas more moderate explained variances

were reported, e.g. in Vancouver, attributed to the impact of difficult to characterize wood burning emissions [23].

OC and PAH models. The lower explained variance of the OC and PAH (ΣPAH, B[a]P, chrysene) models compared to EC is probably due to the contribution of less well characterized sources in addition to traffic to OC and PAH concentrations. In the NPACT study, cross-validation R^2 values for the spatial part of intra-urban spatiotemporal models were higher for EC than for OC (0.81 versus 0.56 for the six cities combined). One PAH LUR model was previously reported for daily PAH concentrations in Fresno, California[30]. The spatiotemporal model included length of highway in a 500m buffer, proximity to roads and neighborhood use of gas for heating as spatial variables, in addition to temporal weather data. The authors reported an index of agreement (which can range from 0 to 100%) of 67% suggesting overall good fit of the PAH models.

In most of our models traffic variables were included but OC and PAH have multiple sources including wood burning and industrial emissions [33]. OC further includes both primary and secondary organic components [37]. Land use regression models cannot represent atmospheric formation processes well, hence the secondary component of OC was not accounted for. We did not have detailed information on sources such as wood burning and industrial emissions available. Non-traffic sources were represented by more general variable such as: population density or industrial land use. The available industrial land use data does not include type of industry, e.g. steel and metal industry which is one of the PAH sources. The limited number of monitoring sites in the direct neighborhood of industry further contributed to the limited inclusion of industry in our models. For most study areas, industry was present within 5000 m buffers for the majority of sites, but in buffers of 1000m and smaller no industry was present for a large fraction of sites. Consistently, in the three models with industry, the 5000 m buffer was included. Identification of large industrial point sources and better databases incorporating type of industry and stack height of emissions may be needed.

We previously observed lower explained variance of LUR models for elements with significant non-traffic sources (e.g. potassium, silica, sulfur) compared to elements for which traffic is the dominant source (copper, iron) [26]. The moderate explained variance was attributed as well to lack of variables describing specific sources of the elements but also to low intra-urban contrasts in elemental concentrations [26]. The latter explanation did not apply to our OC and PAH observations. The three previous LUR studies on modeling of wood smoke also recognized the problem of obtaining good data on the spatial distribution of wood burning emissions [27-29]. In the Seattle and Vancouver studies, neighborhood data from property databases was used [27,28]. In the Montreal study, chimney density was used as a proxy for wood burning [29]. The three studies reported only moderate levels of explained variance – 57% in Seattle [28,38], 58% in Vancouver [27,28] and 40% for the global model in Montreal [29].

Representation of traffic predictors

The most common representations are road length and traffic intensity (load) in buffers and distance to a (major) road. We also used the product of traffic intensity and inverse distance, because this combined variable better represents the processes of emission and dispersion than separate traffic intensity and distance variables [36]. The combined variables were included in only few models, possibly due to insufficient quality of the variables derived from GIS e.g. distance [36].

Several LUR models included both traffic intensity and road length variables, though typically not at the same scale. Because of concern about the completeness and quality of traffic intensity data, we offered both traffic load and road length in model development. Traffic load is more specific. The inclusion of traffic intensity in many models suggests quality was sufficient in most study areas. Road length represents traffic emissions but also emissions related to population, which may explain that mixed models performed reasonably well for urban background sites as well.

422

423 **Differences between study areas**

424 We found substantially larger differences in model R^2 between study areas for OC and PAH
425 compared to EC. Model performance and structure may differ because of variability in measured
426 concentrations, size and complexity of the study area, completeness and quality of predictor
427 variables, quality of geo-coding and availability of predictor variables representing the major source
428 of a component.[24,26]. On average, the models with the lowest R^2 were found in Helsinki and
429 Copenhagen. The lack or poor performance of models for PAH and OC may be due to a smaller
430 influence of motorized traffic emissions relative to other sources. No difference in measured PAH
431 concentrations between street and urban background locations was found in these two cities only [33].
432 The OC models for these two cities were the only models without traffic predictors. As discussed
433 before, sources other than traffic were less well represented by variables used during LUR models
434 development. The high R^2 in London and Paris may be due to the larger diversity in the study areas
435 incorporating two very large cities and smaller towns. The moderate model R^2 for Catalonia could be
436 due to the limited availability of traffic intensity data for the cities outside Barcelona [24].

437

438 **Performance at street and background sites**

439 The generally better performance of models for street compared to urban background locations is
440 likely due to the higher contrast of concentrations at street locations and better representation of the
441 source. The overall model is affected heavily by the street locations. In studies with more sampling
442 sites, investigators have built models separately for the background and local scale [23]. Relatively poor
443 models for particularly the urban background sites were often related to models with a single
444 predictor; models with predictors with a sizable (>50%) number of zero values e.g. natural land in
445 small buffers; models with predictors largely restricted to major roads such as traffic load in 50 meter
446 buffers. We included predictors with at least five non-zero observations, but could have been more
447 restrictive.

448

449 **Hopanes/steranes models.** Although traffic is the main source for hopanes/steranes, we found
450 relatively low level of explained variance (median R^2 = 38%). The low model R2 might be caused by
451 low within study area contrast in hopanes/steranes concentrations, as indicated by the relatively low
452 street to urban background ratio of 1.31 at the monitoring sites [33]. In Catalonia where the
453 concentration contrast was highest, traffic related variables were not included in the
454 ∑hopanes/steranes model, possibly because the traffic variables we used do not describe motor oil
455 emission. Taken together, hopanes /steranes do not contribute much to additional characterization of
456 human exposure of particulate matter air pollution.

457

458 **Strengths and limitations.** The advantage of our study was the standardization of every stage of
459 the project. Annual averages of the pollutants used for LUR models development were obtained
460 according to standardized protocols across all study areas. LUR models were developed centrally
461 according to a standardized protocol. This allowed us to obtain comparable LUR models in ten
462 European study areas for various pollutants
463 Care must be taken if using some of the poorer LUR models in health studies, for example by
464 incorporating the cross validation R^2 in a sensitivity analysis. Also use of models with high R^2, but
465 giving predicted values which correlate highly with other pollutants (e.g. EC vs $PM_{2.5}$ or NO_x) must
466 be evaluated carefully in order to distinguish health effects caused by the pollutant of interest.
467 A limitation of our study was the lack of specific GIS variables for especially industry or wood
468 smoke. More specific GIS data for these sources are difficult to obtain. We evaluated source-specific
469 emission data in some areas, but these did not explain variation over the available GIS variables,
470 probably because of a too-large spatial scale.
471 A further limitation of our study was the small number of sites available in the study areas for the
472 LUR model development. It has been reported that small number of site selected for LUR models
473 development can cause overestimation of results of models validation used in our study (LOOCV)

474 [39,40]. Since we had only one measurement per season we could not develop reliable models for each
475 season.
476 **Application in epidemiology.** Though EC model performance was good in all areas, predictions
477 of the EC models correlated highly with the predictions from the $PM_{2.5}$ absorbance models which
478 have already been applied in ESCAPE epidemiological studies. In most study areas, the squared
479 correlation was above 0.5 (~R=0.7), used as a rule of thumb to determine whether the independent
480 effect of two pollutants can be assessed [41]. Though application of model predicted EC in
481 epidemiological studies likely will result in similar associations with health as the model predicted
482 PM2.5 absorbance, there may be some benefit of applying EC models to compare with previous US
483 studies that have used EC as a metric [6]. Furthermore, as EC to absorbance concentration ratios varied
484 across study areas (Supplementary information Table S8), effect estimates could be more
485 homogenous between studies, if EC is associated more closely with health than $PM_{2.5}$ absorbance.
486 Model performance was moderate for OC and the correlations of OC model predictions with the
487 prediction of mostly traffic-related $PM_{2.5}$ absorbance were moderate overall. Correlations of OC with
488 $PM_{2.5}$ were higher, but still moderate and highly variable across study areas. In six areas the squared
489 correlation was below 0.5. Depending on the locations of cohort studies, OC may be a useful
490 additional indicator. In the NPACT study, some associations between OC and cardiovascular
491 morbidity were found [6].
492
493 Performance of PAH models was also moderate and their predictions correlated on average
494 moderately with $PM_{2.5}$ absorbance and $PM_{2.5}$ predictions with very high variation of the correlations
495 across study areas. Application in epidemiological studies will depend on the locations of cohorts.
496 PAH have been implicated as causal agents for lung cancer and non-malignant respiratory disease
497 and not merely as another indicator for traffic tailpipe emissions [30].
498
499

▪ ASSOCIATED CONTENT

Supporting information

Additional information containing: (1)Map with 10 European study areas, (2) description of predictor variables, (3) Spearman correlation between B[a]P , chrysene and traffic and wood smoke markers, (4) mean and contrast of annual averages of organics components for 10 European study areas, (5) description of land use regression model for ∑PAH, (6) description of land use regression model for chrysene (7) description of land use regression model for Σhopanes/steranes, (8) average of the model R2 and LOOCV R2 for EC, OC, B[a]P, ΣPAH and chrysene per study area, (9) linear regression between EC and PM2.5absorbance for 10 study areas, (10) graph of measured vs modelled concentration of presented pollutants. This material is available free of charge via the Internet at http://pubs.acs.org

▪ AUTHOR INFORMATION

Corresponding Author

* Phone: +31 (0)65 109 62 97

E-mail: aleksandra.jedynska@tno.nl

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

We would like to thank everybody who participated in the air sampling, measurements, data management and the project supervision. This work was supported by the Netherlands Ministry of Infrastructure and Environment and the 7th European Framework projects TRANSPHORM (ENV.2009.1.2.2.1) and ESCAPE under grant agreement number: 211250.

526

527 ## References

528 1. Brunekreef, B.; Holgate, S. T. Air pollution and health. *The Lancet* **2002***, 360*, 1233-1242.

529 2. Pope, C. A.,3rd; Dockery, D. W. Health effects of fine particulate air pollution: lines that connect.
530 *J. Air Waste Manag. Assoc.* **2006***, 56*, 709-742.

531 3. WHO Regional Office for Europe, Review of evidence on health aspects of air pollution -
532 REVIHAAP Project. **2013**.

533 4. Stanek, L. W.; Sacks, J. D.; Dutton, S. J.; Dubois, J. B. Attributing health effects to apportioned
534 components and sources of particulate matter: An evaluation of collective results. *Atmos.*
535 *Environ.* **2011***, 45*, 5655-5663.

536 5. Kelly, F. J.; Fussell, J. C. Size, source and chemical composition as determinants of toxicity
537 attributable to ambient particulate matter. *Atmos. Environ.* **2012***, 60*, 504-526.

538 6. Vedal, S.; Campen, M. J.; McDonald, J. D.; Larson, T. V.; Sampson, P. D.; Sheppard, L.;
539 Simpson, C. D.; Szpiro, A. A. National Particle Component Toxicity (NPACT) initiative report
540 on cardiovascular effects. *Res. Rep. Health Eff. Inst.* **2013***, (178)*, 5-8.

541 7. Lippmann, M.; Chen, L. C.; Gordon, T.; Ito, K.; Thurston, G. D. National Particle Component
542 Toxicity (NPACT) Initiative: integrated epidemiologic and toxicologic studies of the health
543 effects of particulate matter components. *Res. Rep. Health Eff. Inst.* **2013***, (177)*, 5-13.

544 8. Penner, J. E.; Novakov, T. Carbonaceous particles in the atmosphere: A historical perspective to
545 the Fifth International Conference on Carbonaceous Particles in the Atmosphere. *Journal of*
546 *Geophysical Research: Atmospheres (1984–2012)* **1996***, 101*, 19373-19378.

547 9. Cyrys, J.; Heinrich, J.; Hoek, G.; Meliefste, K.; Lewne, M.; Gehring, U.; Bellander, T.; Fischer, P.;
548 van Vliet, P.; Brauer, M.; Wichmann, H. E.; Brunekreef, B. Comparison between different
549 traffic-related particle indicators: elemental carbon (EC), PM2.5 mass, and absorbance. *J. Expo.*
550 *Anal. Environ. Epidemiol.* **2003***, 13*, 134-143.

551 10. Heal, M. R.; Quincey, P. The relationship between black carbon concentration and black smoke:
552 A more general approach. *Atmos. Environ.* **2012***, 54*, 538-544.

553 11. Schaap, M.; Denier van der Gon, H. A. C. On the variability of Black Smoke and carbonaceous
554 aerosols in the Netherlands. *Atmos. Environ.* **2007***, 41*, 5908-5920.

555 12. Keuken, M.; Jonkers, S.; Zandveld, P.; Voogt, M. Elemental carbon as an indicator for evaluating
556 the impact of traffic measures on air quality and health. *Atmos. Environ.* **2012**, *61,* 1-8.

557 13. Janssen, N. A.; Hoek, G.; Simic-Lawson, M.; Fischer, P.; van Bree, L.; ten Brink, H.; Keuken,
558 M.; Atkinson, R. W.; Anderson, H. R.; Brunekreef, B.; Cassee, F. R. Black carbon as an
559 additional indicator of the adverse health effects of airborne particles compared with PM10 and
560 PM2.5. *Environ. Health Perspect.* **2011***, 119*, 1691-1699.

14. Kim, S. Y.; Peel, J. L.; Hannigan, M. P.; Dutton, S. J.; Sheppard, L.; Clark, M. L.; Vedal, S. The temporal lag structure of short-term associations of fine particulate matter chemical constituents and cardiovascular and respiratory hospitalizations. *Environ. Health Perspect.* **2012**, *120*, 1094-1099.

15. Son, J. Y.; Lee, J. T.; Kim, K. H.; Jung, K.; Bell, M. L. Characterization of fine particulate matter and associations between particulate chemical constituents and mortality in Seoul, Korea. *Environ. Health Perspect.* **2012**, *120*, 872-878.

16. Ostro, B.; Reynolds, P.; Goldberg, D.; Hertz, A.; Burnett, R. T.; Shin, H.; Hughes, E.; Garcia, C.; Henderson, K. D.; Bernstein, L. Assessing Long-Term Exposure in the California Teachers Study. *Environ. Health Perspect.* **2011**, *119*, A242.

17. Ravindra, K.; Sokhi, R.; Van Grieken, R. Atmospheric polycyclic aromatic hydrocarbons: Source attribution, emission factors and regulation. *Atmos. Environ.* **2008**, *42*, 2895-2921.

18. Nielsen, T.; Jørgensen, H. E.; Larsen, J. C.; Poulsen, M. City air pollution of polycyclic aromatic hydrocarbons and other mutagens: occurrence, sources and health effects. *Sci. Total Environ.* **1996**, *189–190*, 41-49.

19. Delfino, R. J.; Staimer, N.; Tjoa, T.; Arhami, M.; Polidori, A.; Gillen, D. L.; George, S. C.; Shafer, M. M.; Schauer, J. J.; Sioutas, C. Associations of primary and secondary organic aerosols with airway and systemic inflammation in an elderly panel cohort. *Epidemiology* **2010**, *21*, 892-902.

20. EUD (European Union Directive) Directive 2004/107/EC of the European Parliament and of the Council of 15 December 2004 relating to arsenic, cadmium, mercury, nickel and polycyclic aromatic hydrocarbons in ambient air. *Official Journal L* **2005**, *23*, 0003-0016.

21. Cass, G. R. Organic molecular tracers for particulate air pollution sources. *TrAC Trends in Analytical Chemistry* **1998**, *17*, 356-366.

22. McDonald, J. D.; Eide, I.; Seagrave, J.; Zielinska, B.; Whitney, K.; Lawson, D. R.; Mauderly, J. L. Relationship between composition and toxicity of motor vehicle emission samples. *Environ. Health Perspect.* **2004**, *112*, 1527-1538.

23. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmos. Environ.* **2008**, *42*, 7561-7578.

24. Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.; Künzli, N.; Schikowski, T.; Marcon, A.; et al. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmos. Environ.* **2013**, *72*, 10-23.

25. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dedele, A.; Dons, E.; de Nazelle, A.; et al. 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.* **2012**, *46*, 11195-11205.

599 26. de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.; Cirach,
600 M.; Declercq, C.; Dedele, A.; et al. Development of Land Use Regression Models for Particle
601 Composition in Twenty Study Areas in Europe. *Environ. Sci. Technol.* **2013**,47, 5778 - 5786.

602 27. Larson, T.; Su, J.; Baribeau, A.; Buzzelli, M.; Setton, E.; Brauer, M. A spatial model of urban
603 winter woodsmoke concentrations. *Environ. Sci. Technol.* **2007***, 41*, 2429-2436.

604 28. Su, J. G.; Buzzelli, M.; Brauer, M.; Gould, T.; Larson, T. V. Modeling spatial variability of
605 airborne levoglucosan in Seattle, Washington. *Atmos. Environ.* **2008***, 42*, 5519-5525.

606 29. Smargiassi, A.; Brand, A.; Fournier, M.; Tessier, F.; Goudreau, S.; Rousseau, J.; Benjamin, M. A
607 spatiotemporal land-use regression model of winter fine particulate levels in residential
608 neighbourhoods. *Journal of Exposure Science and Environmental Epidemiology* **2012***, 22*, 331-
609 338.

610 30. Noth, E. M.; Hammond, S. K.; Biging, G. S.; Tager, I. B. A spatial-temporal regression model to
611 predict daily outdoor residential PAH concentrations in an epidemiologic study in Fresno, CA.
612 *Atmos. Environ.* **2011***, 45*, 2394-2403.

613 31. Eeftens, M.; Tsai, M.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.; Cirach,
614 M.; Cyrys, J.; de Hoogh, K.; et al. Spatial variation of PM2.5, PM10, PM2.5 absorbance and
615 PMcoarse concentrations between and within 20 European study areas and the relationship with
616 NO2 – Results of the ESCAPE project. *Atmos. Environ.* **2012***, 62*, 303-317.

617 32. Beelen, R.; Raaschou-Nielsen, O.; Stafoggia, M.; Andersen, Z. J.; Weinmayr, G.; Hoffmann, B.;
618 Wolf, K.; Samoli, E.; Fischer, P.; Nieuwenhuijsen, M.; et al. Effects of long-term exposure to air
619 pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre
620 ESCAPE project. *The Lancet* **2014***, 383*, 785-795.

621 33. Jedynska, A.; Hoek, G.; Eeftens, M.; Cyrys, J.; Keuken, M.; Ampe, C.; Beelen, R.; Cesaroni, G.;
622 Forastiere, F.; Cirach, M.; et al. Spatial variations of PAH, hopanes/steranes and EC/OC
623 concentrations within and between European study areas. *Atmos. Environ.* **2014***, 87*, 239-248.

624 34. Cyrys, J.; Eeftens, M.; Heinrich, J.; Ampe, C.; Armengaud, A.; Beelen, R.; Bellander, T.;
625 Beregszaszi, T.; Birk, M.; Cesaroni, G.; et al. Variation of NO2 and NOx concentrations
626 between and within 36 European study areas: Results from the ESCAPE study. *Atmos. Environ.*
627 **2012***, 62*, 374-390.

628 35. Cavalli, F.; Viana, M.; Yttri, K. E.; Genberg, J.; Putaud, J. Toward a standardised thermal-optical
629 protocol for measuring atmospheric organic and elemental carbon: the EUSAAR protocol.
630 *Atmos. Meas. Tech. Discuss.* **2009***, 2*, 2321-2345.

631 36. Hoek, G.; Beelen, R.; Kos, G.; Dijkema, M.; Zee, Saskia C van der; Fischer, P. H.; Brunekreef,
632 B. Land use regression model for ultrafine particles in Amsterdam. *Environ. Sci. Technol.* **2010***,*
633 *45*, 622-628.

634 37. Saarikoski, S.; Timonen, H.; Saarnio, K.; Aurela, M.; Järvi, L.; Keronen, P.; Kerminen, V.;
635 Hillamo, R. Sources of organic carbon in fine particulate matter in northern European urban air.
636 *Atmos.Chem.Phys* **2008***, 8*, 6281-6295.

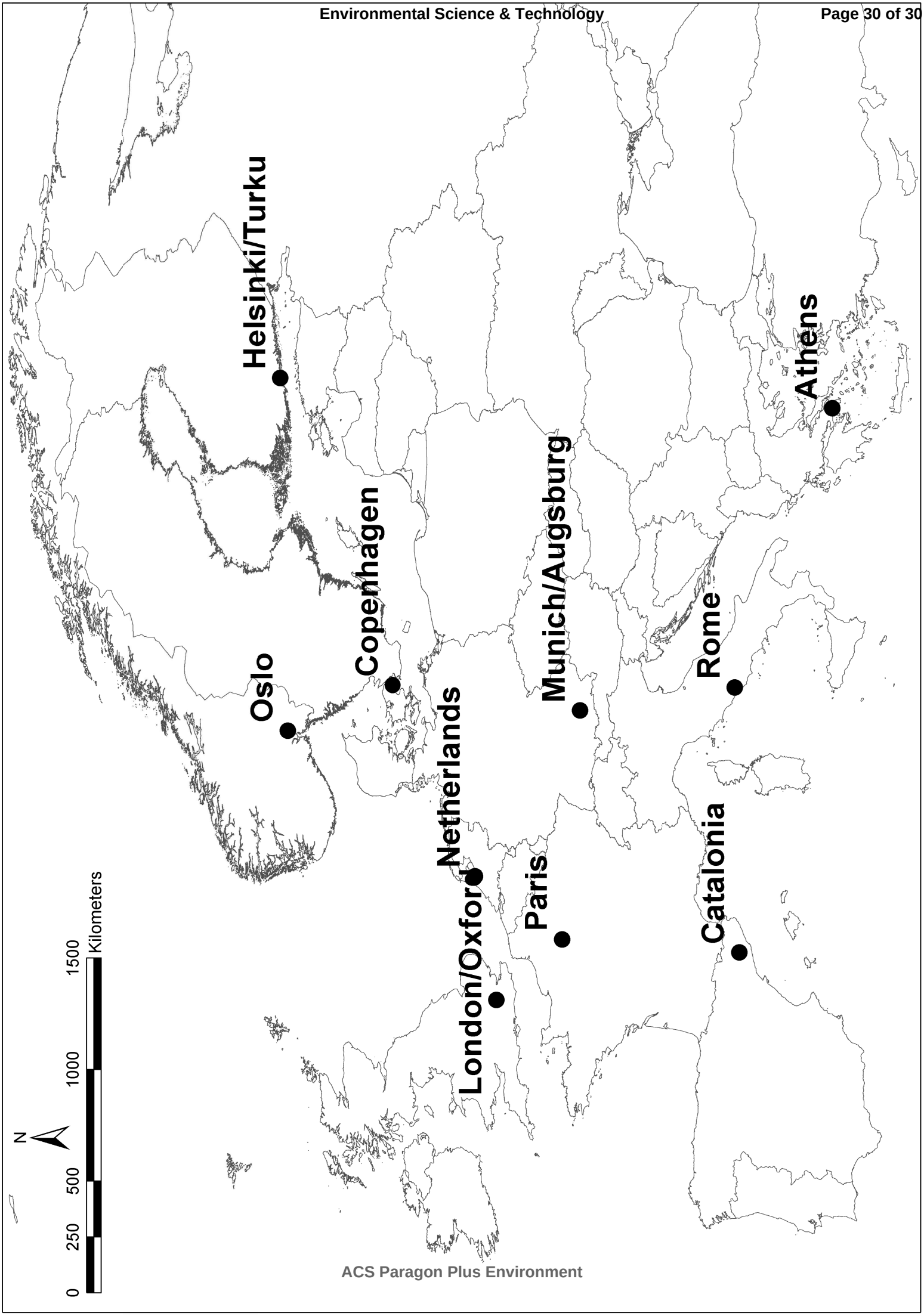
637 38. Larson, T.; Henderson, S. B.; Brauer, M. Mobile monitoring of particle light absorption
638 coefficient in an urban area as a basis for land use regression. *Environ. Sci. Technol.* **2009***, 43*,
639 4672-4678.

640 39. Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de Nazelle,
641 A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N. Effect of the number of measurement sites on land
642 use regression models in estimating local air pollution. *Atmos. Environ.* **2012***, 54*, 634-642.

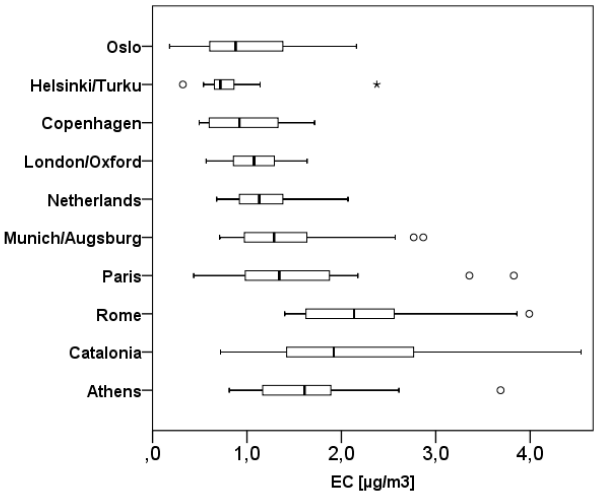
643 40. Wang, M.; Beelen, R.; Basagana, X.; Becker, T.; Cesaroni, G.; de Hoogh, K.; Dedele, A.;
644 Declercq, C.; Dimakopoulou, K.; Eeftens, M.; et al. Evaluation of land use regression models for
645 NO2 and particulate matter in 20 European study areas: the ESCAPE project. *Environ. Sci.*
646 *Technol.* **2013**.

647 41. Wang, M.; Beelen, R.; Stafoggia, M.; Raaschou-Nielsen, O.; Andersen, Z. J.; Hoffmann, B.;
648 Fischer, P.; Houthuijs, D.; Nieuwenhuijsen, M.; Weinmayr, G.; et al. Long-term exposure to
649 elemental constituents of particulate matter and cardiovascular mortality in 19 European cohorts:
650 Results from the ESCAPE and TRANSPHORM projects. *Environ. Int.* **2014***, 66*, 97-106.

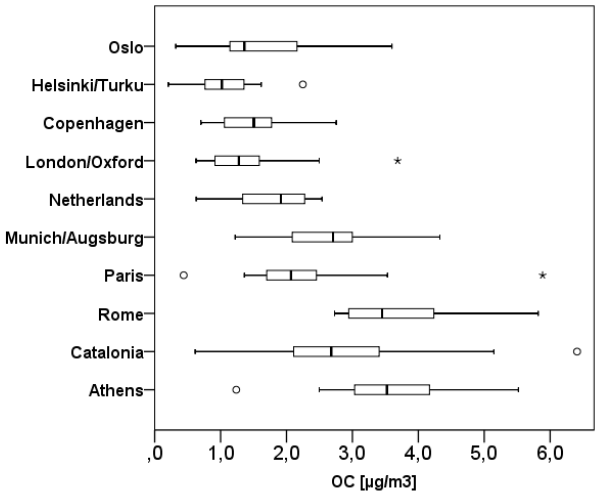
651


Helsinki/Turku
Athens
Copenhagen
Munich/Augsburg
Oslo
Rome
Netherlands
Paris
London/Oxford
Catalonia
N
0
250
500
1000
1500
Kilometers

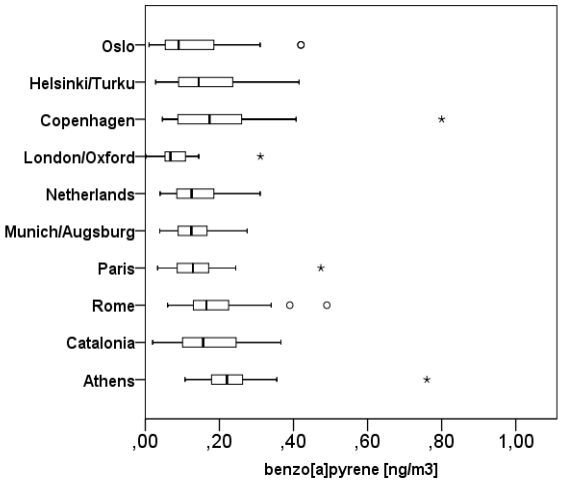
a)


Oslo
Helsinki/Turku
Copenhagen
London/Oxford
Netherlands
Munich/Augsburg
Paris
Rome
Catalonia
Athens
,0
1,0
2,0
3,0
4,0
EC [µg/m3]

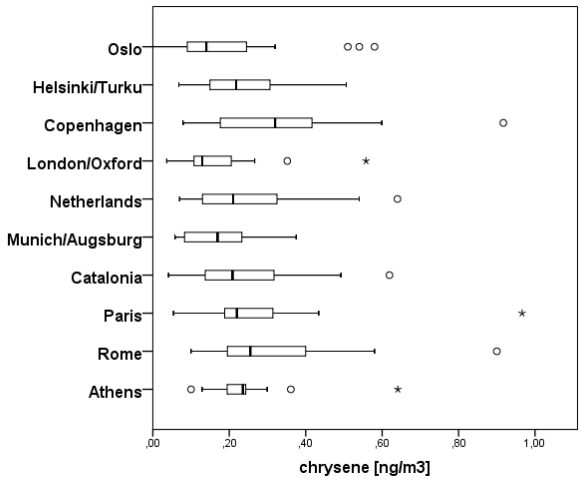
b)


Oslo
Helsinki/Turku
Copenhagen
London/Oxford
Netherlands
Munich/Augsburg
Paris
Rome
Catalonia
Athens
,0
1,0
2,0
3,0
4,0
5,0
6,0
OC [µg/m3]

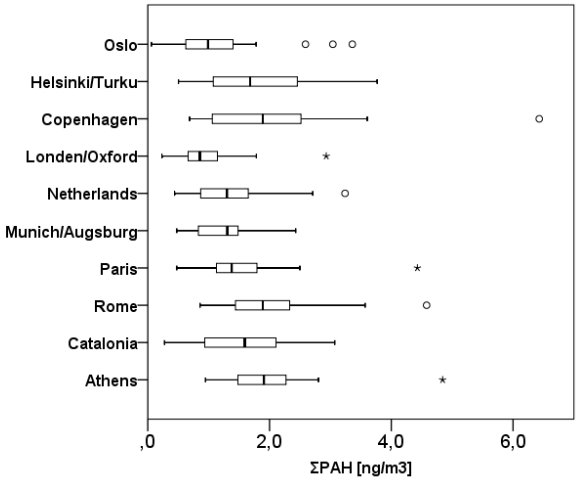
c)


Oslo
Helsinki/Turku
Copenhagen
London/Oxford
Netherlands
Munich/Augsburg
Paris
Rome
Catalonia
Athens
,00
,20
,40
,60
,80
1,00
benzo[a]pyrene [ng/m3]

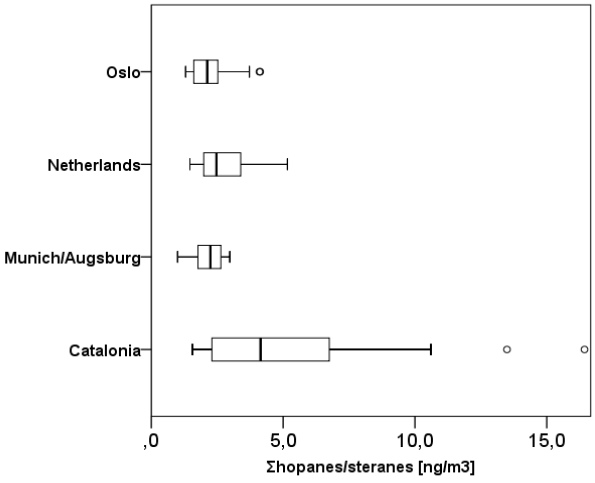
d)


Oslo
Helsinki/Turku
Copenhagen
London/Oxford
Netherlands
Munich/Augsburg
Catalonia
Paris
Rome
Athens
,00
,20
,40
,60
,80
1,00
chrysene [ng/m3]

e)


Oslo
Helsinki/Turku
Copenhagen
Londen/Oxford
Netherlands
Munich/Augsburg
Paris
Rome
Catalonia
Athens
,0
2,0
4,0
6,0
ΣPAH [ng/m3]

f)


Oslo
Netherlands
Munich/Augsburg
Catalonia
,0
5,0
10,0
15,0
Σhopanes/steranes [ng/m3]