

1 **Western European land use regression incorporating satellite- and ground-**
2 **based measurements of NO₂ and PM₁₀**

3

4 Journal: Environmental Science & Technology

5 Danielle Vienneau^{1,2,3*}, Kees de Hoogh³, Matthew J. Bechle⁴, Rob Beelen⁵, Aaron van Donkelaar⁶,
6 Randall V. Martin^{6,7}, Dylan B. Millet⁴, Gerard Hoek⁵, Julian D. Marshall⁴

7 1. Swiss Tropical and Public Health Institute, Basel, Switzerland

8 2. University of Basel, Basel, Switzerland

9 3. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London,
10 United Kingdom

11 4. Department of Civil Engineering, University of Minnesota, Minneapolis, USA

12 5. Institute for Risk Assessment Sciences, Division Environmental Epidemiology, Utrecht University, Utrecht, the
13 Netherlands

14 6. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Canada

15 7. Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA

16

17 * Corresponding author: Danielle Vienneau, PhD

18 Tel: +41 (0)61 284 8398

19 Fax: +41 (0)61 284 8105

20 E-mail address: danielle.vienneau@unibas.ch

21 Mailing address: Department of Epidemiology and Public Health

22 Swiss Tropical and Public Health Institute

23 Socinstrasse 57

24 CH-4051, Basel, Switzerland

25

26 Word count: 4200 words excluding references and statement about available supporting information,
27 plus 2700 word-equivalents for tables and figures = 6900 total. (Basis: Table 1 is small, Tables 2 and
28 3 are large tables, Figures 1 and 2 are large figures. Word-equivalents: 300 per small table/figure, 600
29 per large table/figure.)

30

31 **Abstract**

32 Land use regression (LUR) models typically investigate within-urban variability in air pollution.
33 Recent improvements in data quality and availability, including satellite-derived pollutant
34 measurements, support fine-scale LUR modelling for larger areas. Here, we describe NO₂ and PM₁₀
35 LUR models for Western Europe (years: 2005–2007) based on >1500 EuroAirnet monitoring sites
36 covering background, industrial, and traffic environments. Predictor variables include land use
37 characteristics, population density, and length of major and minor roads in zones from 0.1km to
38 10km, altitude, and distance to sea. We explore models with and without satellite-based NO₂ and
39 PM_{2.5} as predictor variables and we compare two available land cover datasets (global; European).
40 Model performance (adjusted R²) is 0.48–0.58 for NO₂ and 0.22–0.50 for PM₁₀. Inclusion of satellite
41 data improved model performance (adjusted R²) by, on average, 0.05 for NO₂ and 0.11 for PM₁₀.
42 Models were applied on a 100m grid across Western Europe; to support future research, these datasets
43 are publicly available.

44

45 **1. Introduction**

46 Land use regression (LUR) has rapidly become a standard approach for estimating spatial variability
47 in air pollution, for example during exposure assessment in epidemiological studies. Since the
48 inception of LUR,¹ many studies have explored how well LUR can estimate within-city spatial
49 variability in pollutant concentrations.^{2,3} Recent attention has focused on comparing LUR to other
50 methods such as interpolation and dispersion modelling;^{4,5} applying LUR to specific constituents
51 (e.g., soot) and elements of PM_{2.5}^{6,7} and specific organic compounds (e.g., PAHs);^{3,8} and, evaluating
52 the transferability of models to other spatial and temporal contexts.⁹⁻¹⁴

53 LUR models are often derived from measurements made specifically to build the LUR. An alternative
54 approach is to employ data from existing monitors; this approach is well suited to modelling broad
55 geographic extents. Examples include individual European countries,^{11,15} continental USA,^{16,17}
56 Canada,¹⁸ and Western Europe.¹⁹

57 Here we develop NO₂ and PM₁₀ LUR models for Western Europe. Only one Europe-wide LUR has
58 previously been published.¹⁹ We improve on that investigation by offering two orders of magnitude
59 improvement in spatial resolution (1km² [prior¹⁹] versus 0.01km² [here]), and by including satellite-
60 derived estimates of ground-level air pollution. Investigations with large populations and geographic
61 extents, including epidemiological studies of air pollution and traffic-related air pollution,
62 environmental injustice studies, and health risk assessment, would benefit from continental-scale
63 models with a finer spatial resolution.

64 We investigate whether satellite-derived pollution measurements improve fine-scale concentration
65 estimates in European-wide LURs. Our approach incorporates GIS-derived land use, topographic
66 data, and satellite-derived estimates of ground-level concentrations for NO₂ and PM_{2.5}. We benefit
67 from the large number of regulatory monitoring stations (EuroAirnet) operating in Western Europe,
68 facilitating independent evaluation with reserved sites.

69 **2. Methods**

70 We develop land use regression (LUR) models for Western Europe (17 contiguous countries; Figure
71 1). Our dependent variables are ambient concentrations of NO₂ and PM₁₀, obtained from regulatory
72 monitoring. Our independent variables include several GIS-derived measures of land use and
73 topography (100m grids) and satellite-derived estimates of surface concentrations of NO₂ and PM_{2.5}
74 (not PM₁₀; despite the availability of satellite-derived PM_{2.5} estimates, there is an insufficient number
75 of ground-based monitoring sites to support modelling PM_{2.5}). We next describe the input data and
76 then our modelling approach.

77 **2.1. Data**

78 2.1.1. Ground-based monitoring data

79 We use annual mean NO₂ and PM₁₀ concentrations (years 2005–2007) from EuroAirnet, the
80 regulatory air pollution monitoring network in Europe. EuroAirnet comprises sites from national
81 networks²⁰ and is publicly reported in AirBase (version 5).²¹ NO₂ is monitored by chemiluminescence.
82 PM₁₀ is monitored by various methods including Tapered Element Oscillating Microbalance (TEOM),
83 Beta Attenuation, and Gravimetric methods.²² The network includes “background”, “industrial”, and
84 “traffic” sites; all site types are included here. Urban background sites are representative of the
85 exposure of the general urban population while rural background are sited away from major sources
86 of air pollution.²³ Annual measurements are excluded if a site captured <75% of the total hours (NO₂)
87 or days (PM₁₀). Table 1 presents summary statistics for retained monitoring sites. For each year,
88 monitoring data are randomly stratified (by country and site type) into five groups, each with 20% of
89 sites. Subset 1 (20%) is used for model evaluation; the remaining four subsets (80%) are combined
90 and used for model building. As a sensitivity analysis, we apply a five-fold cross-validation procedure
91 in which the 20% evaluation subset is rotated, thereby creating four additional models. *We a priori*
92 designate the first subset to model evaluation, reverting to the next subset only if spatial
93 autocorrelation is detected. We further evaluate models developed using 100% of the monitoring sites,
94 and undertake a sensitivity analysis including country to investigate potential differences in the
95 national networks comprising AirBase.

96 <<Table 1. Summary statistics for mean annual concentrations (µg/m³) at all monitoring sites with
97 ≥75% annual data capture >>

98 2.1.2. Satellite-derived estimates of ground-level concentrations

99 We employ satellite-derived estimates of ground-level NO₂¹⁷ and PM_{2.5}.²⁴ Tropospheric NO₂ columns
100 are from the OMI (Ozone Monitoring Instrument) instrument onboard the Aura satellite.²⁵ Aerosol
101 optical depth (AOD) retrieved from the MODIS (Moderate Resolution Imaging Spectroradiometer)²⁶
102 and MISR (Multiangle Imaging Spectroradiometer)²⁷ instruments onboard the Terra satellite is used to
103 estimate PM_{2.5}. As described elsewhere,^{17, 24, 28} satellite column-integrated retrievals were related to
104 surface concentrations at 0.1°×0.1° resolution (~10km grid) using scaling factors interpolated from the
105 GEOS-Chem chemical transport model (www.geos-chem.org) that account for the local vertical
106 distribution and scattering properties of each pollutant. Annual satellite-derived estimates for NO₂
107 were made for years 2005, 2006 and 2007. Satellite-derived humidity-corrected PM_{2.5} estimates for
108 2001–2006 were aggregated to improve accuracy by enabling sufficient data capture; estimates for
109 grid cells with <50 daily AOD measurements over the 6 years were removed.²⁴ In Europe, PM_{2.5}
110 represents a large fraction (40–80%) of PM₁₀ mass in ambient air,^{29, 30} motivating the use of satellite-
111 derived PM_{2.5} as an independent variable in a PM₁₀ LUR.

112 2.1.3. Predictor variables

113 Predictor variables are integrated into a 100m raster GIS database using ArcGIS10, employing the
114 European reference grid (ETRS Lambert Azimuthal Equal Area 52 10). Satellite-derived pollution
115 measurements and global land cover data are first resampled using nearest neighbour assignment;
116 altitude is resampled using bilinear interpolation (used for continuous data). Variables, described
117 below, are computed either as point estimates or zones. Zones of increasing radius (hereafter referred
118 to as “buffers”) from 0.1km to 10km are computed using the Focalsum command with the circle
119 option. Table 2 summarises the predictor variables.

120 <<Table 2. GIS predictor variables>>

121 Two land cover datasets are available: the 100m European Corine Land Cover³¹ and coarser global
122 datasets including 500m tree canopy³² and 1km impervious surfaces.³³ On the basis of the 44 land
123 classes available in Corine, we define six main groups, represented by individual classes (Hdr, Ldr,
124 Ind, Port; see Table 2) or aggregations of classes (Urbgr, Nat). We define two additional classes based
125 on further aggregation of the urban classes (Res, Tbu). For both datasets (European; global), the
126 percent area within in each buffer is computed for each land cover category. Population counts per
127 grid cell are based on the European Environment Agency 1km² population density grid.^{34,35}

128 We use the 1:10,000 EuroStreets digital road network (version 3.1, based on TeleAtlas MultiNet TM
129 for year-2008) to derive road density variables. EuroStreets includes 9 road classes, which we
130 aggregate into major roads (motorways, main roads and other major roads) and minor roads
131 (secondary and four types of local roads). Non-motorised tracks and paths are excluded. We intersect
132 the road data with a 100m base polygon, then calculate total length per grid cell and for each buffer.
133 Consistent traffic-volume data are not available for Europe.

134 We use altitude data from the SRTM Digital Elevation Database version 4.1.³⁶ The resolution of the
135 SRTM data is 3 arc second (approx. 90m), with vertical error <16m. SRTM is available for most of
136 the study area, up to 60°N latitude. For northern Scandinavia we use 1km resolution Topo30 data.
137 Distance to sea, a measure of continentality, differentiates coastal from inland areas which are not, for
138 example, influenced by coastal recirculation patterns and particulates from sea spray. We compute
139 this variable as the distance between centroids of a 1km grid and the open ocean 25km offshore as
140 defined by Corine land cover. Distance (in m) is then assigned to the 100m grid using inverse distance
141 weighed (1/d) interpolation. Interpolated distance was validated against direct calculation of distance
142 to sea, using NEAR, at the monitoring sites (r=99). Following Beelen et al.,¹⁹ we apply a nonlinear
143 transformation to altitude and distance to sea (see Table 2). We also include X and Y coordinates for
144 the cell centroids to reflect broad scale trends in background air pollution concentrations.^{9,11}

145 2.2. Modelling Approach

146 LUR model development follows the ESCAPE supervised stepwise selection to derive the multiple
147 linear regression equation.^{37, 38} Monitoring data (dependent variable), which are log-normally
148 distributed, are log-transformed prior to modelling. We exclude potential predictor variables with
149 $\geq 90\%$ null values. Univariate regressions of the natural logarithm (LN) of annual mean
150 concentrations and all available potential predictors variables are first developed, and the predictor
151 with the highest adjusted R^2 retained. In subsequent steps, the remaining predictor variables are
152 evaluated in turn; the variable offering the highest increase in adj- R^2 is retained if (1) the coefficient
153 conforms to the pre-specified direction of effect (see Table 2), (2) each additional predictor variable
154 increases the adj- R^2 by at least 0.01, and (3) the direction of effect for predictors already included in
155 the model does not change. *Post hoc*, variables with p-value > 0.10 or variance inflation factor (VIF)
156 > 5 are removed.¹⁷ When required, *post hoc* “ring” (i.e., annulus) variables are calculated by
157 differencing the component buffers, and the model is rerun to derive the final coefficients.^{11, 38} We
158 apply standard diagnostic tests for ordinary least squares regression, including checks on the
159 normality of residuals, heteroscedasticity, spatial autocorrelation of residuals using Moran’s I, and
160 influential observations using Cook’s D.

161 For models testing the inclusion of satellite-based measurements, that predictor variable is forced into
162 the model as the first variable, and the model is built according to the procedure above. Partial R^2
163 values are recomputed and reported after the final model is derived. Models are evaluated against the
164 independent subset of 20% sites reserved for this purpose; R^2 , root mean squared error (RMSE), error,
165 and bias¹⁷ are reported here.

166 3. Results

167 3.1. Measured concentrations from ground-based monitoring

168 Variability in annual mean NO_2 and PM_{10} concentrations measured at the Airbase monitoring sites is
169 relatively consistent across the three years (Table 1). For both pollutants, the number of sites available
170 for modelling ($\geq 75\%$ annual data capture) increases each year, owing to network growth,
171 improvements in data capture, or both. The number of sites measuring continuously over the 3-year
172 period is lower than the number of sites for any individual year (23% [17%] less for NO_2 [PM_{10}],
173 relative to 2005). Given the longer temporal period of the $\text{PM}_{2.5}$ satellite data, we also include LUR
174 models based on the 3-year average concentrations. For both pollutants, the largest share of
175 monitoring sites, with ~ 100 – 400 each, are in Austria, Italy, Spain, Germany and France (see
176 Supplementary Table S1). Most countries have either a consistent number or experienced an increase
177 in number of sites by year. Great Britain is an exception, with a 60% (30%) reduction in NO_2 (PM_{10})
178 site number in year-2007 relative to 2006. Spain also exhibits a dip in monitor numbers for both

179 pollutants in 2006. Expansion in the network is greatest for Italy, with a 65% (86%) increase in NO₂
180 (PM₁₀) sites from 2005 to 2007. For both pollutants, Pearson's correlation between the ground- and
181 satellite-based measurements ranges from 0.33–0.37. The agreement between observed PM₁₀ and
182 satellite-derived PM_{2.5} is likely decreased by differences in sampling period, spatial representation and
183 aerosol size, but is sufficient to suggest applicability as a LUR predictor. Correlation is higher with
184 background sites, which are expected to be more representative of the larger area covered by each
185 satellite grid cell. Scatterplots are in Supplementary Figure S1 and S2.

186 **3.2. Model comparison**

187 Table 3 compares the models on the basis of coefficient of determination (R^2), mean error, and bias.
188 For both pollutants, models with satellite data outperformed the respective model without satellite
189 data, achieving higher model building and evaluation R^2 and lower error and bias. Increases in adj- R^2
190 attributable to including satellite estimates are 0.02–0.06 for NO₂, 0.07–0.13 for PM₁₀. Selection of
191 land cover dataset (Corine vs. global) yielded modest (at most 0.04) impacts to adj- R^2 .

192 The addition of satellite data did not substantially alter the structure of the NO₂ models (Table 3): road
193 and land cover variables remain largely unchanged; other variables (altitude, population density, and
194 distance to sea) only enter the models when satellite data is not included. By comparison, the PM₁₀
195 model structure is less stable both across and within years; a consistent pattern in variables entering
196 models with and without satellite data is not apparent.

197 <<Table 3. Comparison of all models>>

198 Model results are mapped in Figure 1 and 2 (models with satellite-derived pollution estimates) and
199 Figure S3 and S4 (models without satellite-derived pollution estimates). For both pollutants, the
200 models generally resolve expected patterns in air pollution, with higher concentrations in urban areas
201 and near roadways. There are detectable differences, however, in the specific spatial patterns for cities
202 (see map insets and profiles), because of differences in the overall structure of the models. At the
203 European scale, the maps show that known hotspots with frequently elevated regional background
204 levels (e.g., the Ruhr area, Po valley, and western Netherlands) are better captured in models that
205 include satellite-derived pollution estimates. Table S2 presents model evaluation by region. A striking
206 example from that table is for the Italy + Greece region (PM₁₀, n=309 monitors), R^2 is 0.07 without
207 satellite data, 0.45 with satellite data.

208 The sensitivity analysis of 80:20 subsets for annual models reveals that models are robust to changes
209 in the evaluation subset; differences in adj- R^2 are slight (<0.02 for NO₂ and <0.04 for PM₁₀; see Table
210 S3). Table S3 also shows the evaluation subset used to derive the models presented in Table 3. All
211 models, for both pollutants, show no spatial autocorrelation in the residuals. Models based on 100%

212 sites were similar in structure and performance (Tables S4 and S5). Including country indicators
213 generally improved models, although not all indicators were statistically significant. Furthermore, to
214 avoid the introduction of step changes in concentrations at country borders, we do not use country in
215 our final models. Improvement was marked, up to ~20%, for some of the PM₁₀ models. This
216 improvement in part is likely attributed to differences in PM monitoring equipment, but also reflects
217 differences in calibration²² and of site selection of the various countries.

218 << Figure 1. Map and profile plots of NO₂ concentration in 2005 using satellite data; scatterplot of
219 modelled vs. measured NO₂ at evaluation sites >>

220 << Figure 2. Map and profile plots of PM₁₀ concentration in 2007 using satellite data; scatterplot of
221 modelled vs. measured PM₁₀ at evaluation sites >>

222 **3.3. Final models**

223 **NO₂**

224 The best-performing NO₂ models by year are in Table S6. The variables in each NO₂ model are
225 consistent across years: in addition to satellite-derived surface NO₂, all models include the length of
226 minor roads in an intermediate buffer (1500 or 1800m) and in the outer ring to 10km, major road
227 length in a 100m buffer, and total built up land from Corine in a 300m buffer. The models also all
228 contain Corine semi-natural land with a negative coefficient in a 500 or 600m buffer. Minor roads in
229 the intermediate buffer contribute 59–65% to the model predictive power (partial R²: 0.3–0.4),
230 followed by satellite-based NO₂ at 17–23% (partial R²: 0.1). Those findings underscore the utility of
231 satellite-based NO₂ concentrations for NO₂ LUR.

232 Overall, the final NO₂ models explain 55–60% of the variation in log-transformed NO₂ at the more
233 than 400 reserved evaluation sites distributed across Europe (Table S8; Figure S5). Expressed and
234 mapped as concentrations (µg/m³), the explained variation is 50–56%. Error and bias are relatively
235 similar across years, with highest error + bias in year-2007: error (-1.3 – -1.8 µg/m³); absolute error
236 (8.1–8.5 µg/m³); mean bias (11–18%); and, absolute bias (34–41%). Minor road length and satellite
237 estimates of NO₂ are consistently the two most important predictors.

238 **PM₁₀**

239 The best-performing model for PM₁₀ by year is shown in Table S7. The variables in the final PM₁₀
240 models varied by year, with the global land cover models performing better than Corine in 2005 and
241 2007. All models contain satellite-based PM_{2.5}, the Y coordinate indicating the general decreasing
242 trend in concentrations from south to north, and major roads in the immediate buffer. As with NO₂,
243 for PM₁₀ the satellite measurement is consistently the first or second variable to enter the model.

244 Distance to sea enters all but the 2007 model, which instead has the altitude variable. The 2005, 2006
245 and 2005–2007 models include land cover classes representing both built up areas and remote areas.
246 The structure of the 2007 model is rather different, and includes minor roads in both a local (200m)
247 and intermediate (200–2500m) buffer, and percent tree canopy as the only land cover variable.
248 Satellite-based PM_{2.5} and the Y coordinate each contribute ~30–35% to the model predictive power in
249 each year (see Table S7; partial R²: 0.1–0.2).

250 Based on model R², the year-2007 model explains ~47% of the variation in measured concentrations
251 (~50% of the variation in log-transformed NO₂) at the sites reserved for model evaluation; models for
252 earlier years explain 38–44% of variation in measured concentrations (Table S8; Figure S6). Error and
253 bias are relatively similar across years, with lower error and bias in later years: error (-0.2 – -1.2
254 µg/m³); absolute error (4.4– 6.0 µg/m³); mean bias (3–5%); absolute bias (17–22%).

255 **4. Discussion**

256 LUR models given here explain 46–56% (36–48%) of the variation in annual mean NO₂ (PM₁₀)
257 concentration at independent sites. For both pollutants, satellite data are consistently the first or
258 second variable into the model, and those data improve LUR model performance. Based on model R²,
259 satellite data contribute more to the PM₁₀ models than the NO₂ models, despite the difference in
260 particle sizes (using PM_{2.5} satellite data to model PM₁₀ measurements). This finding is likely because
261 the satellite data provide estimates averaged over a ~10km grid, and thus reflects regional background
262 rather than local variations in concentrations. Compared to NO₂, ambient concentrations of PM₁₀ are
263 much more affected by long-range transport; that transport is detected by the PM_{2.5} satellite data.

264 The overall performance of the NO₂ model is better than for PM₁₀, perhaps owing to other more local
265 predictor variables, consistent with observations in the ESCAPE study.^{37, 38} Furthermore, in the EU
266 methodological consistency of monitoring is greater for NO₂ (chemiluminescence) than PM₁₀
267 (multiple methods). Recent spatiotemporal LURs for the USA reported an R² of 0.78 for NO₂¹⁷ and
268 0.63 for PM_{2.5}.¹⁶ As indicated by our models with country indicators (Tables S4 and S5) and the
269 evaluation by region (Table S2), however, there are differences between countries which cannot be
270 explained by the variables in our final PM₁₀ models. This perhaps points to the need for regional
271 models, especially for PM₁₀.

272 We expect that meteorological conditions also play a role in PM₁₀ model performance. In Europe, for
273 example, 2006 was a year with several air pollution episodes including that associated with the July
274 heat wave. Here, unlike in our previous work,^{19, 39} we did not specifically include coarse-scale
275 meteorological variables. We took this *a priori* decision because the effects of meteorology are
276 generally captured by the satellite-derived air pollution data, yet at a higher spatial resolution than for
277 meteorological data. While daily meteorological variability is incorporated into the satellite-derived

278 PM_{2.5} estimates, year-to-year variability, however, is not captured by the long-term mean (2001–
279 2006) we use in the PM₁₀ LUR models. If year-to-year model variation is in fact mainly driven by
280 meteorological factors, model performance may benefit from including meteorological variables in
281 the LUR models or, like NO₂, using annual satellite data.

282 In general, the models described here exhibit comparable performance as previous LUR models at the
283 European scale: Beelen et al.,¹⁹ report validation R²s of 0.61 (0.45) for NO₂ (PM₁₀) using a hybrid
284 Kriging-LUR approach. Our NO₂ models may explain less of the variation in measured
285 concentrations relative to the work of Beelen et al. in part because we model all site types, including
286 traffic, rather than only background sites. We found that evaluation R²s for independent monitoring
287 sites is very similar to the model R², consistent with methodological work showing that model R² can
288 exceed independent evaluation R²s for small datasets, but less so for large datasets such as the ones
289 we use here.^{40, 41}

290 An important next step for this research would be to model PM_{2.5}, a pollutant which is subject to
291 recent EU guideline limits²³ and, based on the Global Burden of Disease estimates, is responsible for
292 3.2 million deaths and 76 million years of lost healthy life worldwide.⁴² Although site numbers for
293 PM_{2.5} are slowly increasing, for this time period and study area, too few sites are available to derive
294 reliable LUR models (146 and 195, respectively, in year-2005 and 2007 with sufficient annual data
295 capture). A large fraction of the spatial variation of PM₁₀ is related to variation of PM_{2.5}. The
296 ESCAPE study reported an average R² between spatial variation of PM₁₀ and PM_{2.5} of 0.74 (range
297 0.44–0.95).³⁰

298 Modelling over large areas at fine spatial resolutions is an attractive solution for a variety of
299 applications with large study populations, including health risk assessment. Given that LUR models
300 generally cannot be directly transferred to other spatial domains,^{10, 11, 14} our approach addresses a
301 particular need for reliable and consistent models at the continental level. From our models we
302 estimate the mean population-weighted exposure in 2007 was 27 (25) µg/m³ for NO₂ (PM₁₀).
303 Furthermore, we estimate that 9% (NO₂) and 1% (PM₁₀) of the European population reside in areas
304 exceeding the annual guideline limit of 40 µg/m³ (current annual guidelines are the same for NO₂ and
305 PM₁₀).²³ Some caution is needed in interpretation of these results given differences in model
306 performance by region (Table S2). These regional differences in model performance may in part be
307 attributable to known deficiencies in the monitoring network (uneven distribution and clustering of
308 sites in EuroAirmet, which is an assembly of sites from existing country networks; use of different
309 PM₁₀ monitoring methods and correction factors by country) or discrepancies in the definition of land
310 cover or road classes across Europe.⁴³

311 There are several challenges in producing suitable models for air pollution exposure assessment
312 across large areas. We aim here for models at a spatial scale fine enough to estimate within-city and

313 near-roadway contrasts in pollution while also accounting for long-range transport and other large
314 scale variability in pollution. Most studies evaluating exposures over large areas use a vector-LUR
315 approach whereby estimates are then made at census centroids, a coarse mesh, or home addresses;
316 should a map or estimates at additional locations be required, interpolation is then used to produce a
317 continuous surface.^{17, 18, 44-46} A strength of our models is that we take a raster-based LUR approach,
318 which enables direct prediction at the 100m grid (Figure 1 and 2). We thus eliminate the need for
319 interpolation which can over-smooth estimates. In this study, a 100m resolution is justified given the
320 quality and resolution of the source information as well as the dense network of monitoring sites
321 distributed in different exposure environments across Europe. Although not always reflected in the R²
322 as a performance measure, this attribute (large number of monitors, located in diverse environments)
323 is an important advance over the previous models for Europe.^{19, 39}

324
325 As was previously demonstrated in Canada¹⁸ and the USA,^{16, 17, 47} we show here that combining LUR
326 models with worldwide, satellite-based pollution measurements can offer improved continental-scale
327 exposure models for Europe. To support future research, model results are publicly available.

328 **Acknowledgements**

329 The research leading to these results has received funding from the European Community's Seventh
330 Framework Program (FP7/2007-2011) under Grant Agreement No. 211250, and is based upon work
331 supported by the National Science Foundation under Grant No. 0853467. We thank Dr. Eric Novotny
332 for assistance in obtaining and processing the global land use data. We acknowledge the free use of
333 tropospheric NO₂ column data from the OMI sensor via www.temis.nl.

334 **Supporting Information Available**

335 Table S1: Number of monitoring sites by year. Table S2: Evaluation statistics by regions for best
336 models (based on concentration ($\mu\text{g}/\text{m}^3$)). Table S3: Sensitivity analysis - rotating 20% evaluation
337 subsets. Table S4: Sensitivity analysis - NO₂ models based on all monitoring sites. Table S5:
338 Sensitivity analysis - PM₁₀ models based on all monitoring sites. Table S6: Final NO₂ models by year.
339 Table S7: Final PM₁₀ models by year. Table S8: Summary of model building and evaluation statistics.
340 Figure S1: Measured ground-based NO₂ vs. satellite-derived NO₂, at all monitoring sites. Figure S2:
341 Measured ground-based PM₁₀ vs. mean 2001-2006 satellite-derived PM_{2.5}, at all monitoring sites.
342 Figure S3: Map and profile plots of NO₂ concentration in 2005 without satellite data; scatterplot of
343 modelled vs. measured NO₂ at evaluation sites. Figure S4: Map and profile plots of PM₁₀
344 concentration in 2007 without satellite data; scatterplot of modelled vs. measured PM₁₀ at evaluation
345 sites. Figure S5: Modelled vs. measured NO₂ concentration ($\mu\text{g}/\text{m}^3$) at evaluation sites for final
346 models. Figure S6: Modelled vs. measured PM₁₀ concentration ($\mu\text{g}/\text{m}^3$) at evaluation sites for final
347 models. This information is available free of charge via the Internet at <http://pubs.acs.org/>.

348 **References**

- 349 1. Briggs, D. J.; Collins, S.; Elliott, P.; Fischer, P.; Kingham, S.; Lebret, E.; Pyl, K.; van
350 Reeuwijk, H.; Smallbone, K.; van der Veen, A., Mapping urban air pollution using GIS: a
351 regression-based approach. *Int. J. GIS.* **1997**, *11* (7), 699-718.
- 352 2. Jerrett, M.; Arain, A.; Kanaroglou, P.; Beckerman, B.; Potoglou, D.; Sahsuvaroglu, T.;
353 Morrison, J.; Giovis, C., A review and evaluation of intraurban air pollution exposure models.
354 *J. Exposure Sci. Environ. Epidemiol.* **2005**, *15*, (2), 185-204.
- 355 3. Hoek, G.; Beelen, R.; de Hoogh, K.; Vienneau, D.; Gulliver, J.; Fischer, P.; Briggs, D., A
356 review of land-use regression models to assess spatial variation of outdoor air pollution.
357 *Atmos. Environ.* **2008**, *42*, (33), 7561-7578.
- 358 4. Marshall, J. D.; Nethery, E.; Brauer, M., Within-urban variability in ambient air pollution:
359 Comparison of estimation methods. *Atmos. Environ.* **2008**, *42*, (6), 1359-1369.
- 360 5. Gulliver, J.; de Hoogh, K.; Fecht, D.; Vienneau, D.; Briggs, D., Comparative assessment of
361 GIS-based methods and metrics for estimating long-term exposures to air pollution. *Atmos.*
362 *Environ.* **2011**, *45*, (39), 7072-7080.
- 363 6. de Hoogh, K.; Wang, M.; Adam, M.; Badaloni, C.; Beelen, R.; Birk, M.; Cesaroni, G.;
364 Cirach, M.; Declercq, C.; Dèdelè, A.; Dons, E.; de Nazelle, A.; Eeftens, M.; Eriksen, K. T.;
365 Eriksson, C.; Fischer, P.; Gražulevičienė, R.; Gryparis, A.; Hoffmann, B.; Jerrett, M.;
366 Katsouyanni, K.; Iakovides, M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Mosler, G.;
367 Nádor, G.; Nieuwenhuijsen, M. J.; Pershagen, G.; Peters, A.; Phuleria, H.; Probst-Hensch, N.;
368 Raaschou-Nielsen, O.; Quass, U.; Ranzi, A.; Stephanou, E. G.; Sugiri, D.; Schwarze, P.; Tsai,
369 M.-Y.; Yli-Tuomi, T.; Varró, M. J.; Vienneau, D.; Weinmayr, G.; Brunekreef, B.; Hoek, G.,
370 Development of land use regression models for particle composition in 20 study areas in
371 Europe. *Environ. Sci. Technol.* **2013**, *47*, (11), 5778-5786.
- 372 7. Saraswat, A.; Apte, J.; Kandlikar, M.; Brauer, M.; Henderson, S.; Marshall, J., Spatiotemporal
373 land use regression models of fine, ultrafine and black carbon particulate matter in New
374 Delhi, India. *Environ. Sci. Technol.* **2013**, DOI: 10.1021/es401489h.
- 375 8. Noth, E. M.; Hammond, S. K.; Biging, G. S.; Tager, I. B., A spatial-temporal regression
376 model to predict daily outdoor residential PAH concentrations in an epidemiologic study in
377 Fresno, CA. *Atmos. Environ.* **2011**, *45*, (14), 2394-2403.
- 378 9. Gulliver, J.; de Hoogh, K.; Hansell, A.; Vienneau, D., Development and back-extrapolation of
379 NO₂ land use regression models for historic exposure assessment in Great Britain. *Environ.*
380 *Sci. Technol.* **2013**, DOI: 10.1021/es4008849.
- 381 10. Allen, R. W.; Amram, O.; Wheeler, A. J.; Brauer, M., The transferability of NO and NO₂
382 land use regression models between cities and pollutants. *Atmos. Environ.* **2011**, *45*, (2), 369-
383 378.
- 384 11. Vienneau, D.; de Hoogh, K.; Beelen, R.; Fischer, P.; Hoek, G.; Briggs, D., Comparison of
385 land-use regression models between Great Britain and the Netherlands. *Atmos. Environ.* **2010**,
386 *44*, (5), 688-696.
- 387 12. Poplawski, K.; Gould, T.; Setton, E.; Allen, R.; Su, J.; Larson, T.; Henderson, S.; Brauer, M.;
388 Hystad, P.; Lightowlers, C.; Keller, P.; Cohen, M.; Silva, C.; Buzzelli, M., Intercity
389 transferability of land use regression models for estimating ambient concentrations of
390 nitrogen dioxide. *J. Exposure Sci. Environ. Epidemiol.* **2009**, *19*, (1), 107-117.
- 391 13. Mercer, L. D.; Szpiro, A. A.; Sheppard, L.; Lindström, J.; Adar, S. D.; Allen, R. W.; Avol, E.
392 L.; Oron, A. P.; Larson, T.; Liu, L. J. S.; Kaufman, J. D., Comparing universal kriging and
393 land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO_x) for the
394 Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos. Environ.* **2011**,
395 *45*, (26), 4412-4420.
- 396 14. Liu, L. J. S.; Tsai, M.-Y.; Keidel, D.; Gemperli, A.; Ineichen, A.; Hazenkamp-von Arx, M.;
397 Bayer-Oglesby, L.; Rochat, T.; Künzli, N.; Ackermann-Liebrich, U.; Straehl, P.; Schwartz, J.;
398 Schindler, C., Long-term exposure models for traffic related NO₂ across geographically
399 diverse areas over separate years. *Atmos. Environ.* **2012**, *46*, 460-471.

- 400 15. Gulliver, J.; Morris, C.; Lee, K.; Vienneau, D.; Briggs, D.; Hansell, A., Land use regression
401 modeling to estimate historic (1962-1991) concentrations of black smoke and sulfur dioxide
402 for Great Britain. *Environ. Sci. Technol.* **2011**, *45*, (8), 3526-3532.
- 403 16. Beckerman, B. S.; Jerrett, M.; Serre, M.; Martin, R. V.; Lee, S.-J.; van Donkelaar, A.; Ross,
404 Z.; Su, J.; Burnett, R. T., A hybrid approach to estimating national scale spatiotemporal
405 variability of PM_{2.5} in the contiguous United States. *Environ. Sci. Technol.* **2013**, *47*, (13),
406 7233-7241.
- 407 17. Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D., National satellite-based land-use
408 regression: NO₂ in the United States. *Environ. Sci. Technol.* **2011**, *45*, (10), 4407-4414.
- 409 18. Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van
410 Donkelaar, A.; Lamsal, L.; Martin, R.; Jerrett, M.; Demers, P., Creating national air pollution
411 models for population exposure assessment in Canada. *Environ. Health Perspect.* **2011**, *119*,
412 (8), 1123-1129.
- 413 19. Beelen, R.; Hoek, G.; Pebesma, E.; Vienneau, D.; de Hoogh, K.; Briggs, D. J., Mapping of
414 background air pollution at a fine spatial scale across the European Union. *Sci. Total Environ.*
415 **2009**, *407*, (6), 1852-1867.
- 416 20. Larssen, S.; Sluyter, R.; Helmis, C. *Criteria for EUROAIRNET. The EEA Air Quality*
417 *Monitoring and Information Network*; European Environment Agency: 1999; pp 2-56.
- 418 21. EEA AirBase - The European air quality database, version 5. [http://www.eea.europa.eu/data-](http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-4)
419 [and-maps/data/airbase-the-european-air-quality-database-4](http://www.eea.europa.eu/data-and-maps/data/airbase-the-european-air-quality-database-4).
- 420 22. ETC/ACC *Correction factors and PM₁₀ measurements in AirBase*; Draft technical paper
421 2004/x; European Topic Centre on Air and Climate Change, Bilthoven, Netherlands.
- 422 23. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on
423 ambient air quality and cleaner air for Europe. In Official Journal of the European Union: EU,
424 EC 2008; p 44.
- 425 24. van Donkelaar, A.; Martin, R. V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve,
426 P. J., Global estimates of ambient fine particulate matter concentrations from satellite-based
427 aerosol optical depth: Development and application. *Environ. Health Perspect.* **2010**, *118*, (6),
428 847-855.
- 429 25. Boersma, K. F.; Eskes, H. J.; Veeffkind, J. P.; Brinksma, E. J.; van der A, R. J.; Sneep, M.;
430 van den Oord, G. H. J.; Levelt, P. F.; Stammes, P.; Gleason, J. F.; Bucsela, E. J., Near-real
431 time retrieval of tropospheric NO₂ from OMI. *Atmos. Chem. Physics* **2007**, *7*, (8), 2103-2118.
- 432 26. Levy, R. C.; Remer, L. A.; Mattoo, S.; Vermote, E. F.; Kaufman, Y. J., Second-generation
433 operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate
434 Resolution Imaging Spectroradiometer spectral reflectance. *J. Geo Res.-Atmos.* **2007**, *112*,
435 (D13).
- 436 27. Diner, D. J.; Beckert, J. C.; Reilly, T. H.; Bruegge, C. J.; Conel, J. E.; Kahn, R. A.;
437 Martonchik, J. V.; Ackerman, T. P.; Davies, R. G.; erstl, S. A. W.; Gordon, H. R.; Muller, J.-
438 P.; Myneni, R. B.; Sellers, P. J.; Pinty, B.; Verstraete, M. M., Multi-angle Imaging
439 SpectroRadiometer (MISR) instrument description and experiment overview. *IEEE Trans.*
440 *Geosci. Remote Sens.* **1998**, *36*, (4), 1072-1087.
- 441 28. Bechle, M. J.; Millet, D. B.; Marshall, J. D., Remote sensing of exposure to NO₂: Satellite
442 versus ground-based measurement in a large urban area. *Atmos. Environ.* **2013**, *69*, 345-353.
- 443 29. EEA *Air quality in Europe - 2012 report*, No 4/2012; European Environment Agency:
444 Copenhagen, 2012; p 104.
- 445 30. Eeftens, M.; Tsai, M.-Y.; Ampe, C.; Anwander, B.; Beelen, R.; Bellander, T.; Cesaroni, G.;
446 Cirach, M.; Cyrus, J.; de Hoogh, K.; De Nazelle, A.; de Vocht, F.; Declercq, C.; Dèdelè, A.;
447 Eriksen, K.; Galassi, C.; Gražulevičienė, R.; Grivas, G.; Heinrich, J.; Hoffmann, B.;
448 Iakovides, M.; Ineichen, A.; Katsouyanni, K.; Korek, M.; Krämer, U.; Kuhlbusch, T.; Lanki,
449 T.; Madsen, C.; Meliefste, K.; Mölter, A.; Mosler, G.; Nieuwenhuijsen, M.; Oldenwening,
450 M.; Pennanen, A.; Probst-Hensch, N.; Quass, U.; Raaschou-Nielsen, O.; Ranzi, A.;
451 Stephanou, E.; Sugiri, D.; Udvardy, O.; Vaskövi, É.; Weinmayr, G.; Brunekreef, B.; Hoek,
452 G., Spatial variation of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations
453 between and within 20 European study areas and the relationship with NO₂ – Results of the
454 ESCAPE project. *Atmos. Environ.* **2012**, *62*, 303-317.

- 455 31. ETC-LC Corine land cover (CLC2000), raster database (version 12/2009).
456 [http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-](http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-12-2009)
457 [version-12-2009](http://www.eea.europa.eu/data-and-maps/data/corine-land-cover-2000-clc2000-100-m-version-12-2009).
- 458 32. Hansen, M.; DeFries, R.; Townshend, J. R.; Carroll, M.; Dimiceli, C.; Sohlberg, R.,
459 *Vegetation Continuous Fields MOD44B, 2001 Percent Tree Cover, Collection 4, version*
460 *2001*: University of Maryland, College Park, Maryland 2006.
- 461 33. Elvidge, C.; Tuttle, B.; Sutton, P.; Baugh, K.; Howard, A.; Milesi, C.; Bhaduri, B.; Nemani,
462 R., Global Distribution and Density of Constructed Impervious Surfaces. *Sensors* **2007**, *7*, (9),
463 1962-1979.
- 464 34. Gallego, F. J., A population density grid of the European Union. *Popul. Environ.* **2010**, *31*,
465 (6), 460-473.
- 466 35. INTARESE EU age/sex stratified population: 100 metre grid. [http://www.integrated-](http://www.integrated-assessment.eu/resource_centre/eu_agesex_stratified_population_100_metre_grid)
467 [assessment.eu/resource_centre/eu_agesex_stratified_population_100_metre_grid](http://www.integrated-assessment.eu/resource_centre/eu_agesex_stratified_population_100_metre_grid).
- 468 36. CGIAR-CSI SRTM 90m Digital Elevation Data. <http://srtm.csi.cgiar.org/>.
- 469 37. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.;
470 Dèdelè, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.;
471 Galassi, C.; Gražulevičienė, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek,
472 M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Nádor, G.; Nieuwenhuijsen, M.;
473 Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.;
474 Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M.-Y.; Yli-Tuomi, T.; Varró,
475 M. J.; Vienneau, D.; Klot, S. v.; Wolf, K.; Brunekreef, B.; Hoek, G., Development of land use
476 regression models for PM2.5, PM2.5 absorbance, PM10 and PMcoarse in 20 European study
477 areas; Results of the ESCAPE Project. *Environ. Sci. Technol.* **2012**, *46*, (20), 11195-11205.
- 478 38. Beelen, R.; Hoek, G.; Vienneau, D.; Eeftens, M.; Dimakopoulou, K.; Pedeli, X.; Tsai, M.-Y.;
479 Künzli, N.; Schikowski, T.; Marcon, A.; Eriksen, K. T.; Raaschou-Nielsen, O.; Stephanou, E.;
480 Patelarou, E.; Lanki, T.; Yli-Tuomi, T.; Declercq, C.; Falq, G.; Stempfelet, M.; Birk, M.;
481 Cyrus, J.; von Klot, S.; Nádor, G.; Varró, M. J.; Dèdelè, A.; Gražulevičienė, R.; Mölter, A.;
482 Lindley, S.; Madsen, C.; Cesaroni, G.; Ranzi, A.; Badaloni, C.; Hoffmann, B.; Nonnemacher,
483 M.; Krämer, U.; Kuhlbusch, T.; Cirach, M.; de Nazelle, A.; Nieuwenhuijsen, M.; Bellander,
484 T.; Korek, M.; Olsson, D.; Strömngren, M.; Dons, E.; Jerrett, M.; Fischer, P.; Wang, M.;
485 Brunekreef, B.; de Hoogh, K., Development of NO2 and NOx land use regression models for
486 estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmos.*
487 *Environ.* **2013**, *72*, 10-23.
- 488 39. Vienneau, D.; de Hoogh, K.; Briggs, D., A GIS-based method for modelling air pollution
489 exposures across Europe. *Sci. Total Environ.* **2009**, *408*, (2), 255-266.
- 490 40. Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B., Systematic
491 evaluation of land use regression models for NO2. *Environ. Sci. Technol.* **2012**, *46*, (8), 4481-
492 4489.
- 493 41. Basagaña, X.; Rivera, M.; Aguilera, I.; Agis, D.; Bouso, L.; Elosua, R.; Foraster, M.; de
494 Nazelle, A.; Nieuwenhuijsen, M.; Vila, J.; Künzli, N., Effect of the number of measurement
495 sites on land use regression models in estimating local air pollution. *Atmos. Environ.* **2012**,
496 *54*, 634-642.
- 497 42. Lim, S. S.; Vos, T.; Flaxman, A. D.; Danaei, G.; Shibuya, K.; Adair-Rohani, H.; Amann, M.;
498 Anderson, H. R.; Andrews, K. G.; Aryee, M.; Atkinson, C.; Bacchus, L. J.; Bahalim, A. N.;
499 Balakrishnan, K.; Balmes, J.; Barker-Collo, S.; Baxter, A.; Bell, M. L.; Blore, J. D.; Blyth, F.;
500 Bonner, C.; Borges, G.; Bourne, R.; Boussinesq, M.; Brauer, M.; Brooks, P.; Bruce, N. G.;
501 Brunekreef, B.; Bryan-Hancock, C.; Bucello, C.; Buchbinder, R.; Bull, F.; Burnett, R. T.;
502 Byers, T. E.; Calabria, B.; Carapetis, J.; Carnahan, E.; Chafe, Z.; Charlson, F.; Chen, H.;
503 Chen, J. S.; Cheng, A. T.; Child, J. C.; Cohen, A.; Colson, K. E.; Cowie, B. C.; Darby, S.;
504 Darling, S.; Davis, A.; Degenhardt, L.; Dentener, F.; Des Jarlais, D. C.; Devries, K.; Dherani,
505 M.; Ding, E. L.; Dorsey, E. R.; Driscoll, T.; Edmond, K.; Ali, S. E.; Engell, R. E.; Erwin, P.
506 J.; Fahimi, S.; Falder, G.; Farzadfar, F.; Ferrari, A.; Finucane, M. M.; Flaxman, S.; Fowkes, F.
507 G.; Freedman, G.; Freeman, M. K.; Gakidou, E.; Ghosh, S.; Giovannucci, E.; Gmel, G.;
508 Graham, K.; Grainger, R.; Grant, B.; Gunnell, D.; Gutierrez, H. R.; Hall, W.; Hoek, H. W.;
509 Hogan, A.; Hosgood, H. D., 3rd; Hoy, D.; Hu, H.; Hubbell, B. J.; Hutchings, S. J.; Ibeanusi,

510 S. E.; Jacklyn, G. L.; Jasrasaria, R.; Jonas, J. B.; Kan, H.; Kanis, J. A.; Kassebaum, N.;
 511 Kawakami, N.; Khang, Y. H.; Khatibzadeh, S.; Khoo, J. P.; Kok, C.; Laden, F.; Lalloo, R.;
 512 Lan, Q.; Lathlean, T.; Leasher, J. L.; Leigh, J.; Li, Y.; Lin, J. K.; Lipshultz, S. E.; London, S.;
 513 Lozano, R.; Lu, Y.; Mak, J.; Malekzadeh, R.; Mallinger, L.; Marcenes, W.; March, L.; Marks,
 514 R.; Martin, R.; McGale, P.; McGrath, J.; Mehta, S.; Mensah, G. A.; Merriman, T. R.; Micha,
 515 R.; Michaud, C.; Mishra, V.; Hanafiah, K. M.; Mokdad, A. A.; Morawska, L.; Mozaffarian,
 516 D.; Murphy, T.; Naghavi, M.; Neal, B.; Nelson, P. K.; Nolla, J. M.; Norman, R.; Olives, C.;
 517 Omer, S. B.; Orchard, J.; Osborne, R.; Ostro, B.; Page, A.; Pandey, K. D.; Parry, C. D.;
 518 Passmore, E.; Patra, J.; Pearce, N.; Pelizzari, P. M.; Petzold, M.; Phillips, M. R.; Pope, D.;
 519 Pope, C. A., 3rd; Powles, J.; Rao, M.; Razavi, H.; Rehfuess, E. A.; Rehm, J. T.; Ritz, B.;
 520 Rivara, F. P.; Roberts, T.; Robinson, C.; Rodriguez-Portales, J. A.; Romieu, I.; Room, R.;
 521 Rosenfeld, L. C.; Roy, A.; Rushton, L.; Salomon, J. A.; Sampson, U.; Sanchez-Riera, L.;
 522 Sanman, E.; Sapkota, A.; Seedat, S.; Shi, P.; Shield, K.; Shivakoti, R.; Singh, G. M.; Sleet, D.
 523 A.; Smith, E.; Smith, K. R.; Stapelberg, N. J.; Steenland, K.; Stockl, H.; Stovner, L. J.; Straif,
 524 K.; Straney, L.; Thurston, G. D.; Tran, J. H.; Van Dingenen, R.; van Donkelaar, A.; Veerman,
 525 J. L.; Vijayakumar, L.; Weintraub, R.; Weissman, M. M.; White, R. A.; Whiteford, H.;
 526 Wiersma, S. T.; Wilkinson, J. D.; Williams, H. C.; Williams, W.; Wilson, N.; Woolf, A. D.;
 527 Yip, P.; Zielinski, J. M.; Lopez, A. D.; Murray, C. J.; Ezzati, M.; AlMazroa, M. A.; Memish,
 528 Z. A., A comparative risk assessment of burden of disease and injury attributable to 67 risk
 529 factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global
 530 Burden of Disease Study 2010. *Lancet* **2012**, *380*, (9859), 2224-60.
 531 43. Vienneau, D.; Briggs, D. J., Delimiting affinity zones as a basis for air pollution mapping in
 532 Europe. *Environ. Int.* **2013**, *51*, 106-115.
 533 44. Hystad, P.; Demers, P.; Johnson, K.; Brook, J.; van Donkelaar, A.; Lamsal, L.; Martin, R.;
 534 Brauer, M., Spatiotemporal air pollution exposure assessment for a Canadian population-
 535 based lung cancer case-control study. *Environ. Health* **2012**, *11*, (1), 22.
 536 45. Ross, Z.; Jerrett, M.; Ito, K.; Tempalski, B.; Thurston, G. D., A land use regression for
 537 predicting fine particulate matter concentrations in the New York City region. *Atmos.*
 538 *Environ.* **2007**, *41*, (11), 2255-2269.
 539 46. Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J.; Garshick, E.;
 540 Laden, F., Spatial modeling of PM10 and NO2 in the continental United States, 1985-2000.
 541 *Environ. Health Perspect.* **2009**, *117*, (11), 1690-6.
 542 47. Kloog, I.; Nordio, F.; Coull, B. A.; Schwartz, J., Incorporating local land use regression and
 543 satellite aerosol optical depth in a hybrid model of spatiotemporal PM2.5 exposures in the
 544 Mid-Atlantic States. *Environ. Sci. Technol.* **2012**, *46*, (21), 11913-11921.

Table 1. Summary statistics for mean annual concentrations ($\mu\text{g}/\text{m}^3$) at all monitoring sites with $\geq 75\%$ annual data capture

Year	N	Min	5%	95%	Max	Mean	SD	GM	GSD
NO₂									
2005	2010	0.8	7.1	60.8	112.3	29.3	16.5	24.5	1.9
2006	2099	0.9	7.9	61.8	121.3	29.8	16.8	25.1	1.9
2007	2236	0.3	7.5	58.7	106.5	28.8	15.9	24.3	1.9
2005-2007	1670	0.9	8.0	57.9	108.5	28.5	15.5	24.2	1.9
PM₁₀									
2005	1487	7.8	14.8	44.9	70.9	26.6	9.2	25.2	1.4
2006	1584	7.7	15.7	45.7	71.7	27.7	9.2	26.3	1.4
2007	1664	3.6	15.2	44.1	77.4	26.7	8.7	25.4	1.4
2005-2007	1151	7.7	16.1	43.5	61.7	26.7	8.3	25.5	1.4

GM = geometric mean; GSD = geometric standard deviation (unit less)

Table 2. GIS predictor variables

Dataset	Variable ^a	Code	Buffer ^b or point estimate
OMI derived NO ₂ (ppb): ~10km	Surface NO ₂ concentration	SNO2	Point
Terra derived PM _{2.5} (µg/m ³): ~10km	Surface PM _{2.5} concentration	SPM	Point
Corine land cover ^c (% area)	Continuous urban fabric - high density	Hdr	Buffer
	Discontinuous urban fabric - low density	Ldr	
	Industry	Ind	
	Ports	Port	
	Urban green	Urbgr	
	Total built up (Res + Ind + Port + transport infrastructure, airports, mines, dumps and construction sites)	Tbu	
	Semi-natural land	Nat	
Global land cover (% area)	Residential (Hdr + Ldr)	Res	Buffer
	Impervious surface	Isurf	
EuroStreets roads (length in m)	Tree canopy	Tree	Buffer
	Major roads	Majrd	
Modelled Population (N)	Minor roads	Minrd	Buffer
	Population	Pop	
Topography: 90m SRTM DTM	Altitude - transformed ^d	Talt	Point
Modelled distance to sea (m)	Distance to sea - transformed ^e	Tsea	Point
Coordinates (m)	XY coordinates for 100m cell centroids	Xcoord Ycoord	Point

- a. Pre-specified direction of effect is negative for: Urbgr, Nat, Tree, Talt and Ycoord for both pollutants; and Tsea for PM₁₀
- b. "Buffer" zone distances (m): 0; 100; 200; 300; 400; 500; 600; 700; 800; 1000; 1200; 1500; 1800; 2000; 2500; 3000; 3500; 4000; 5000; 6000; 7000; 8000; 10000
- c. Original Corine classes: Hdr: class 111; Ldr: class 112; Ind: class 121; Port: class 123; Urbgr: class 141-142; Tbu: class 111-133; Nat: class 311-423; Res: class 111-112
- d. Transformed Altitude is calculated as $\sqrt{\text{nalt}/\text{max}(\text{nalt})}$, where $\text{nalt}=\text{altitude}-\text{min}(\text{altitude})$
- e. Transformed Distance to sea is calculated as $\sqrt{(\text{minimum distance}/\text{max}(\text{minimum distance}))}$

Table 3. Comparison of all models

Year	Model	With Satellite							Without Satellite							
		Model ^a		Evaluation ^b					Model ^a		Evaluation ^b					
		Variables ^c	Adj-R ²	R ²	ME	MAE	MB	MAB	Variables ^c	Adj-R ²	R ²	ME	MAE	MB	MAB	
NO₂ Models																
2005	Corine	SNO2-05, Minrd-1800, Nat-600, Majrd-100, Tbu-300, Minrd-1800-10000	0.58	0.56	-1.8	8.1	13	37	Tbu-2000, Minrd-400-10000, Nat_600, Majrd-100, Minrd-400	0.55	0.54	-1.7	8.8	16	42	
	Global	SNO2-05, Minrd-400-1800, Majrd-100, Tree-300, Minrd-400, Isurf-800	0.56	0.57	-1.6	8.0	13	37	Minrd-500-2500, Majrd-100, Tree-700, Minrd-2500-10000, Minrd-500	0.51	0.51	-1.6	9.1	18	44	
2006	Corine	SNO2-06, Minrd-1500, Majrd-100, Nat-500, Minrd-1500-10000, Tbu-300	0.55	0.50	-1.5	8.3	11	35	Tbu-1500, Minrd-100-10000, Nat-500, Majrd-100, Minrd-100, Pop-1000	0.53	0.47	-1.6	8.5	11	36	
	Global	SNO2-06, Minrd-200-1500, Majrd-100, Isurf-500, Minrd-200, Tree-700	0.54	0.50	-1.0	8.3	13	36	Majrd-100, Minrd-200, Minrd-200-10000, Tree-700, Isurf-500, Tsea	0.49	0.46	-1.1	8.9	14	39	
2007	Corine	SNO2-07, Minrd-1500, Majrd-100, Nat-600, Tbu-300, Minrd-1500-10000,	0.55	0.50	-1.3	8.5	18	41	Tbu-1200, Minrd-200-10000, Nat-600, Majrd-100, Minrd-200	0.51	0.48	-1.6	8.8	19	43	
	Global	SNO2-07, Minrd-300-1500, Majrd-100, Tree-500, Minrd-300, Isurf-700	0.54	0.54	-1.4	8.1	20	43	Minrd-400-2500, Majrd-100, Tree-800, Minrd-2500-10000, Minrd-400, Talt, Tsea	0.48	0.46	-1.8	9.1	23	49	
2005-2007	Corine	SNO2-05-07, Minrd-1500, Nat-600, Majrd-100, Minrd-1500-10000, Tbu-300	0.60	0.46	-1.8	8.4	8	34	Tbu-2000, Minrd-200-10000, Nat-600, Majrd-100, Minrd-200	0.56	0.48	-1.8	8.3	10	35	
	Global	SNO2-05-07, Minrd-400-1500, Majrd-100, Tree-700, Isurf-500, Minrd-400	0.58	0.50	-1.5	8.0	11	34	Minrd-500-2500, Majrd-100, Tree-800, Minrd-2500-10000, Minrd-500	0.52	0.41	-1.2	8.6	14	38	
PM₁₀ Models																
2005	Corine	SPM, Ycoord, Hdr-1200, Nat-500, Ind-100, Tsea	0.35	0.37	-0.7	5.7	5	22	Ycoord, Tbu-10000, Nat-1000	0.22	0.25	-1.1	6.1	4	23	
	Global	SPM, Ycoord, Isurf-1000, Tree-500, Tsea, Majrd	0.35	0.44	-0.8	5.4	4	21	Tree-800, Ycoord, Isurf-1000-10000, Isurf-1000	0.22	0.25	-1.1	6.0	4	23	
2006	Corine	SPM, Ycoord, Tbu-600, Pop-1800, Tsea, Majrd-100, Nat 600	0.37	0.36	-1.2	6.0	3	22	Pop-1800, Ycoord, Nat-1000, Talt, Majrd-100, Ldr-10000	0.25	0.25	-1.9	6.4	1	23	
	Global	SPM, Ycoord, Isurf-800, Tree-100, Majrd-100	0.36	0.35	-1.5	6.0	2	22	Tree-1000, Ycoord, Isurf-1000, Majrd-100, Talt	0.24	0.20	-1.8	6.6	2	24	
2007	Corine	SPM, Ycoord, Minrd-2500, Talt, Tbu-100	0.49	0.45	-0.6	4.8	3	18	Ycoord, Nat-2000, Minrd-10000, Tbu-100, Talt, Majrd	0.42	0.36	-0.5	5.2	4	19	
	Global	SPM, Ycoord, Minrd-200-2500, Talt, Minrd-200, Majrd, Tree-100	0.50	0.47	-0.5	4.6	4	17	Tree-1200, Ycoord, Minrd-200-6000, Talt, Minrd-200, Majrd-100	0.40	0.37	-0.8	5.1	3	19	
2005-2007	Corine	SPM, Ycoord, Nat-1200, Tsea, Pop-1800, Tbu-400, Majrd-100	0.48	0.48	-0.2	4.4	3	17	Nat-2000, Ycoord, Tbu-300-10000, Talt, Tbu-300	0.36	0.31	-0.2	4.4	3	17	
	Global	SPM, Ycoord, Isurf-200, Tree-600, Tsea, Majrd-100	0.48	0.48	-0.3	4.4	3	17	Tree-1000, Ycoord, Minrd-10000, Talt, Majrd-100	0.36	0.34	-0.8	4.9	3	18	

Best models shaded grey

a. Model building using natural logarithm of concentration (LN concentration)

b. Model evaluation using concentration ($\mu\text{g}/\text{m}^3$): ME = mean error ($\mu\text{g}/\text{m}^3$); MAE = mean absolute error ($\mu\text{g}/\text{m}^3$); MB = mean bias (%); MAB = mean absolute bias (%)

c. Variables listed by order of entry into models, with satellite forced into the model as the first variable

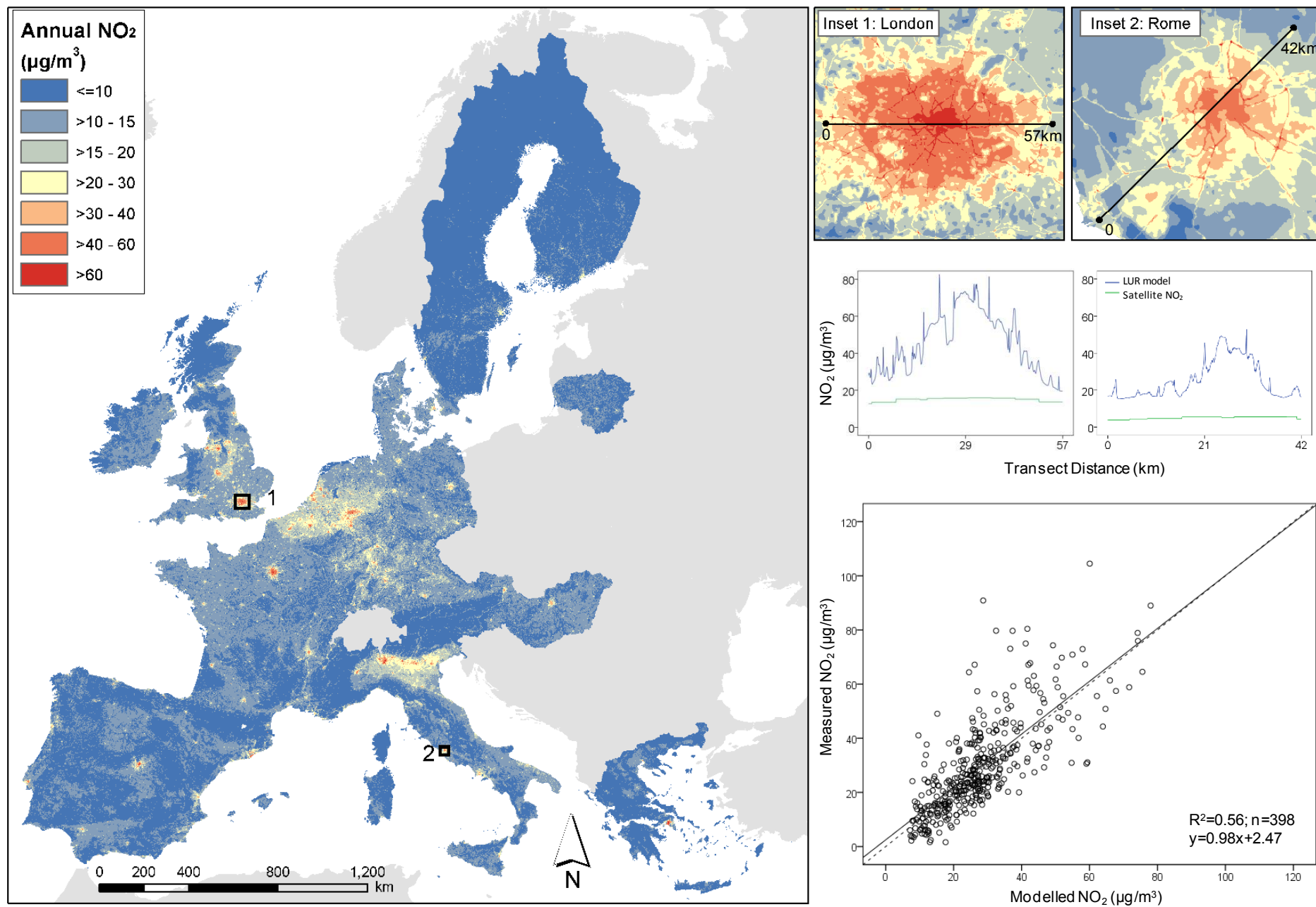


Figure 1. Map and profile plots of NO₂ concentration in 2005 using satellite data; scatterplot of modelled vs. measured NO₂ at evaluation sites

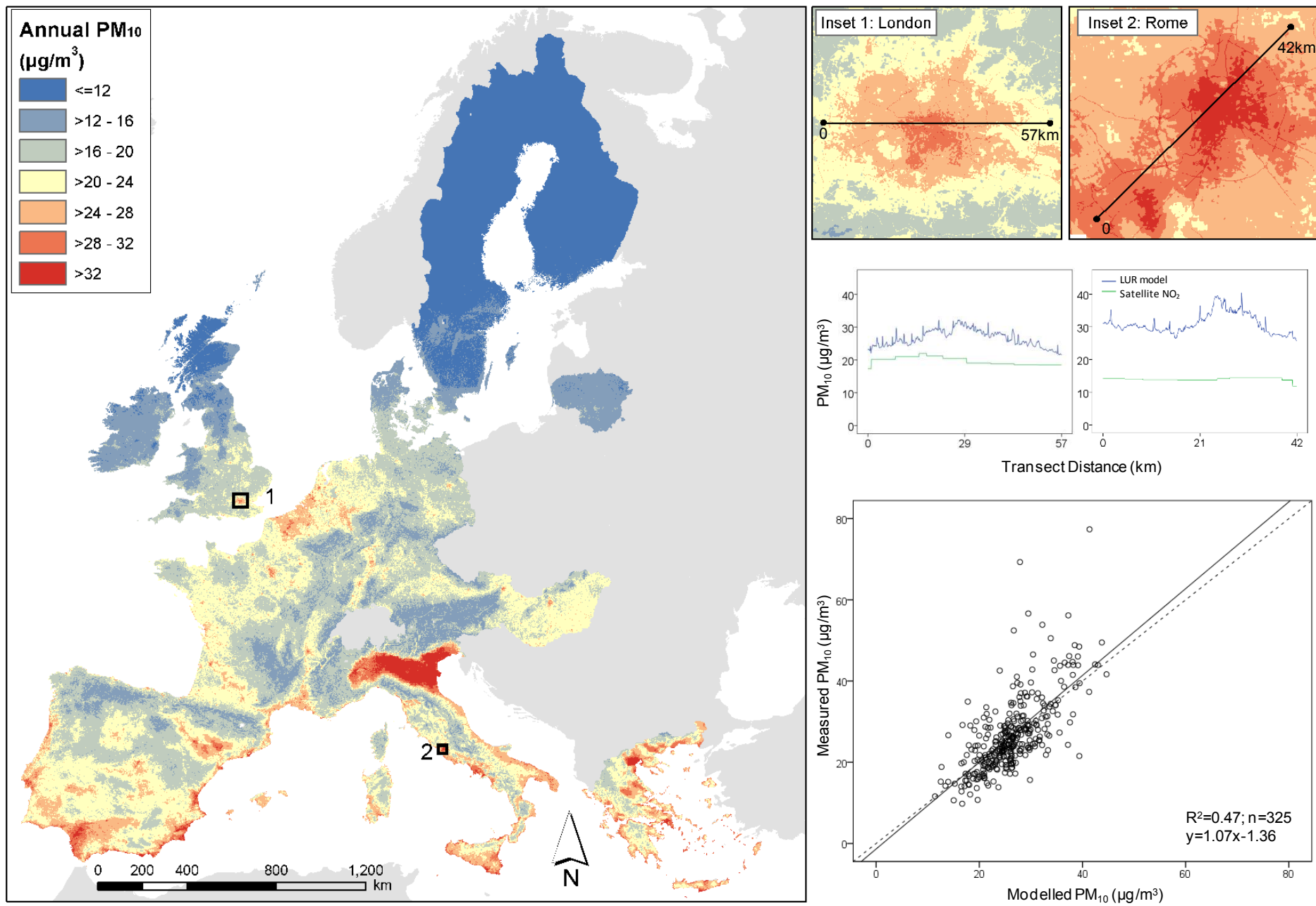


Figure 2. Map and profile plots of PM₁₀ concentration in 2007 using satellite data; scatterplot of modelled vs. measured PM₁₀ at evaluation sites

Supporting Information

Western European land use regression incorporating satellite- and ground-based measurements of NO₂ and PM₁₀

Danielle Vienneau^{1,2,3*}, Kees de Hoogh³, Matthew J. Bechle⁴, Rob Beelen⁵, Aaron van Donkelaar⁶, Randall V. Martin^{6,7}, Dylan B. Millet⁴, Gerard Hoek⁵, Julian D. Marshall⁴

1. Swiss Tropical and Public Health Institute, Basel, Switzerland
2. University of Basel, Basel, Switzerland
3. MRC-PHE Centre for Environment and Health, Department of Epidemiology and Biostatistics, Imperial College London, United Kingdom
4. Department of Civil Engineering, University of Minnesota, Minneapolis, USA
5. Institute for Risk Assessment Sciences, Division Environmental Epidemiology, Utrecht University, Utrecht, the Netherlands
6. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Canada
7. Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts, USA

Contents:

Table S1. Number of monitoring sites by year	2
Table S2. Evaluation statistics by regions for best models (based on concentration (µg/m ³))	3
Table S3. Sensitivity analysis - rotating 20% evaluation subsets	4
Table S4. Sensitivity analysis - NO ₂ models based on all monitoring sites.....	5
Table S5. Sensitivity analysis - PM ₁₀ models based on all monitoring sites	6
Table S6. Final NO ₂ models by year.....	7
Table S7. Final PM ₁₀ models by year	8
Table S8. Summary of model building and evaluation statistics	9
Figure S1. Measured ground-based NO ₂ vs. satellite-derived NO ₂ , at all monitoring sites.....	10
Figure S2. Measured ground-based PM ₁₀ vs. mean 2001-2006 satellite-derived PM _{2.5} , at all monitoring sites.....	11
Figure S3. Map and profile plots of NO ₂ concentration in 2005 without satellite data; scatterplot of modelled vs. measured NO ₂ at evaluation sites	12
Figure S4. Map and profile plots of PM ₁₀ concentration in 2007 without satellite data; scatterplot of modelled vs. measured PM ₁₀ at evaluation sites.....	13
Figure S5. Modelled vs. measured NO ₂ concentration (µg/m ³) at evaluation sites for final models shown in Tables S3 and S5	14
Figure S6. Modelled vs. measured PM ₁₀ concentration (µg/m ³) at evaluation sites for final models shown in Tables S4 and S5.....	15

Table S1. Number of monitoring sites by year

Country	NO ₂ Sites				PM ₁₀ Sites			
	2005	2006	2007	2005-2007	2005	2006	2007	2005-2007
AT	146	145	151	138	107	109	127	98
BE	56	57	62	48	39	39	44	35
DE	394	409	419	358	371	400	417	329
DK	12	12	12	12	10	6	7	4
ES	352	321	372	261	260	236	241	171
FI	26	29	25	20	31	29	27	20
FR	456	463	456	399	305	299	289	211
GB	92	97	38	31	64	67	47	39
GR	16	20	23	15	6	7	13	4
HU	23	23	21	21	19	23	23	18
IE	6	8	7	4	8	8	11	5
IT	309	379	509	259	159	248	296	127
LT	8	12	12	6	12	12	13	10
LU	5	5	6	4	1	1	3	0
NL	42	50	51	37	36	38	36	33
PT	56	55	56	48	45	42	45	37
SE	11	14	16	9	14	20	25	10
Total	2010	2099	2236	1670	1487	1584	1664	1151

Countries: Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Ireland (IE), Italy (IT) Lithuania (LT), Luxembourg (LU), the Netherlands (NL), Portugal (PT), Spain (ES), Sweden (SE), United Kingdom (GB)

Table S2. Evaluation statistics by regions for best models (based on concentration ($\mu\text{g}/\text{m}^3$))A. Final NO_2 model, year 2005 (Corine + satellite vs. Corine only)

Regions ^a	At Evaluation Sites ^b				N	At All Sites				N
	With Satellite		Without Satellite			With Satellite		Without Satellite		
	R ²	RMSE	R ²	RMSE		R ²	RMSE	R ²	RMSE	
Overall	0.56	11.54	0.54	11.82	398	0.51	11.80	0.49	12.04	2010
DK-FI-SE-LT	0.73	8.63	0.75	8.47	11	0.62	8.28	0.58	9.51	57
BE-LU-NL	0.66	7.50	0.67	7.82	25	0.53	10.08	0.54	9.16	103
GB-IE	0.44	16.41	0.40	16.86	19	0.64	12.05	0.61	12.15	98
DE	0.60	8.55	0.61	8.68	69	0.58	9.79	0.65	8.99	394
FR	0.58	9.58	0.59	9.88	90	0.50	10.15	0.48	11.07	456
HU-AT	0.61	9.42	0.60	9.04	26	0.43	11.70	0.46	11.13	169
PT-ES	0.63	9.71	0.61	9.95	93	0.67	10.25	0.64	10.65	408
IT-GR	0.52	17.67	0.50	18.22	65	0.43	17.57	0.41	18.25	325

B. Final PM_{10} model, year 2007 (Global + satellite vs. Global only)

Regions ^a	At Evaluation Sites ^b				N	At All Sites				N
	With Satellite		Without Satellite			With Satellite		Without Satellite		
	R ²	RMSE	R ²	RMSE		R ²	RMSE	R ²	RMSE	
Overall	0.47	6.74	0.37	7.40	325	0.49	6.26	0.35	7.07	1664
DK-FI-SE-LT	0.30	7.29	0.31	5.73	15	0.38	6.64	0.44	5.49	72
BE-LU-NL	0.38	3.38	0.48	3.64	16	0.32	4.28	0.34	4.32	83
GB-IE	0.00	5.20	0.03	5.97	10	0.57	4.40	0.53	5.25	58
DE	0.58	4.03	0.54	4.10	76	0.50	4.12	0.48	4.18	417
FR	0.32	5.54	0.26	5.82	55	0.23	5.27	0.16	5.66	289
HU-AT	0.06	5.16	0.10	5.04	27	0.26	4.94	0.35	4.97	150
PT-ES	0.39	9.67	0.33	9.93	66	0.32	8.52	0.29	8.43	286
IT-GR	0.54	7.86	0.19	10.35	60	0.45	8.02	0.07	11.00	309

a. Regions listed north to south

b. Evaluation sites refers to the 20% sites not used in model building

Countries: Austria (AT), Belgium (BE), Denmark (DK), Finland (FI), France (FR), Germany (DE), Greece (GR), Hungary (HU), Ireland (IE), Italy (IT) Lithuania (LT), Luxembourg (LU), the Netherlands (NL), Portugal (PT), Spain (ES), Sweden (SE), United Kingdom (GB)

Table S3. Sensitivity analysis - rotating 20% evaluation subsets

Year	Model	Subset ^b	Model Building ^a	Model Evaluation ^b		Decision
				LN Concentration	Concentration (µg/m ³)	
				Adj-R ²	R ²	
NO₂						
2005	Corine with satellite	1	0.58	0.58	0.56	final model
		2	0.58	0.58	0.39	
		3	0.59	0.53	0.45	
		4	0.58	0.60	0.54	
		5	0.58	0.60	0.52	
2006	Corine with satellite	1	0.56	0.52	0.41	reject: spatial autocorrelation
		2	0.55	0.55	0.50	final model
		3	0.56	0.52	0.42	
		4	0.55	0.54	0.47	
		5	0.54	0.60	0.50	
2007	Corine with satellite	1	0.55	0.59	0.50	final model
		2	0.56	0.56	0.41	
		3	0.57	0.50	0.43	
		4	0.56	0.54	0.43	
		5	0.55	0.58	0.52	
PM₁₀						
2005	Global with satellite	1	0.35	0.41	0.44	final model
		2	0.36	0.34	0.36	
		3	0.36	0.37	0.38	
		4	0.36	0.33	0.35	
		5	0.36	0.29	0.30	
2006	Corine with satellite	1	0.35	0.40	0.38	reject: spatial autocorrelation
		2	0.36	0.32	0.34	final model
		3	0.38	0.30	0.30	
		4	0.34	0.44	0.43	
		5	0.38	0.29	0.32	
2007	Global with satellite	1	0.50	0.50	0.47	final model
		2	0.50	0.53	0.48	
		3	0.50	0.53	0.52	
		4	0.50	0.53	0.52	
		5	0.53	0.41	0.41	

a. Model building based on natural logarithm of concentration (LN concentration) using 80% of monitoring sites

b. Model evaluation using 20% reserved monitoring sites

Table S4. Sensitivity analysis - NO₂ models based on all monitoring sites

Variables	Model building ^a				
	β^b	IQR	β^* IQR	VIF	Partial Adj-R ²
2005 - Corine with satellite					
Constant	2.245				
Minor roads 1500m	4.37E-06	56158	0.25	2.6	0.37
Satellite-derived surface NO ₂ 2005	6.46E-02	3.0	0.19	1.3	0.49
Major roads 100m	6.02E-04	0.0	0.00	1.1	0.53
Total built up land 300m	3.31E-03	55.2	0.18	2.2	0.56
Minor roads 1500-10000m	1.19E-07	981014	0.12	1.9	0.57
Semi-natural land 600m	-4.03E-03	4.4	-0.02	1.6	0.58
2006 - Corine with satellite					
Constant	2.35				
Minor roads 2000m	2.49E-06	88596	0.22	2.7	0.34
Satellite-derived surface NO ₂ 2006	4.30E-02	3.8	0.17	1.2	0.44
Semi-natural land 500m	-4.85E-03	1.2	-0.01	1.6	0.48
Major roads 100m	6.55E-04	0.0	0.00	1.1	0.53
Total built up land 400m	3.44E-03	53.1	0.18	2.2	0.54
Minor roads 2000-10000m	1.07E-07	939846	0.10	1.9	0.55
2007 - Corine with satellite					
Constant	2.28				
Minor roads 1500m	4.35E-06	55676	0.24	2.5	0.33
Satellite-derived surface NO ₂ 2007	6.51E-02	3.02	0.20	1.3	0.46
Major roads 100m	6.21E-04	0.00	0.00	1.0	0.50
Semi-natural land 600m	-4.35E-03	4.42	-0.02	1.6	0.53
Total built up land 300m	3.20E-03	55.17	0.18	2.2	0.55
Minor roads 1500-10000m	1.02E-07	917706.00	0.09	1.8	0.56

a. Model building based on natural logarithm of concentration (LN concentration) using 100% of monitoring sites

b. All p-values < 0.000

Adj-R² including country dummy variables: 0.62 (year-2005), 0.61 (year-2006) and 0.62 (year-2007)

Table S5. Sensitivity analysis - PM₁₀ models based on all monitoring sites

Variables	Model building ^a				
	β^b	IQR	$\beta^* \text{ IQR}$	VIF	Partial Adj-R ²
2005 - Global with satellite					
Constant	3.36				
Tree canopy 500m	-3.45E-03	7.5	-0.03	1.2	.12
Satellite-derived surface PM _{2.5} 2001-6	2.10E-02	7.1	0.15	1.1	.20
Y coordinate	-1.93E-07	775844	-0.15	1.1	.30
Impervious surface 800m	2.71E-03	40.5	0.11	1.2	.34
2006 - Corine with satellite					
Constant	3.47				
Satellite-derived surface PM _{2.5} 2001-6	2.23E-02	6.9	0.15	1.1	.13
Y coordinate	-1.82E-07	778267	-0.14	1.2	.25
Semi-natural land 1000m	-3.04E-03	10.4	-0.03	1.2	.31
High density residential 1500m	3.04E-03	13.1	0.04	1.2	.34
Major roads 100m	2.05E-04	0.0	0.00	1.1	.35
Distance to sea	-1.87E-01	0.4	-0.07	1.2	.36
2007 - Global with satellite					
Constant	3.65				
Y coordinate	-2.85E-07	780423	-0.22	1.2	.14
Satellite-derived surface PM _{2.5} 2001-6	2.00E-02	7.1	0.14	1.1	.30
Impervious surface 1000m	2.32E-03	36.4	0.08	1.4	.42
Altitude	-7.37E-01	0.2	-0.12	1.2	.47
Minor roads 200m	4.59E-05	1564	0.07	1.4	.49
Major roads	4.94E-04	0.0	0.00	1.1	.50

a. Model building based on natural logarithm of concentration (LN concentration) using 100% of monitoring sites

b. All p-values < 0.000

Adj-R² including country dummy variables: 0.54 (year-2005), 0.53 (year-2006) and 0.54 (year-2007)

Table S6. Final NO₂ models by year

Variables	Model building ^a				
	β^b	IQR	$\beta^* \text{ IQR}$	VIF	Partial Adj-R ²
2005 - Corine with satellite					
Constant	2.31				
Minor roads 1800m	3.22E-06	74727	0.24	2.6	0.38
Satellite-derived surface NO ₂ 2005	6.13E-02	3.0	0.18	1.3	0.48
Semi-natural land 600m	-4.84E-03	4.4	-0.02	1.6	0.52
Major roads 100m	5.91E-04	0.0	0.00	1.1	0.56
Total built up land 300m	3.15E-03	55.2	0.17	2.1	0.57
Minor roads 1800-10000m	1.04E-07	978059	0.10	2.0	0.58
2006 - Corine with satellite					
Constant	2.35				
Minor roads 1500m	3.96E-06	54683	0.22	2.5	0.33
Satellite-derived surface NO ₂ 2006	4.30E-02	4.0	0.17	1.3	0.43
Major roads 100m	6.49E-04	0.0	0.00	1.1	0.48
Semi-natural land 500m	-5.19E-03	1.2	-0.01	1.6	0.52
Minor roads 1500-10000m	1.22E-07	965644	0.12	1.7	0.54
Total built up land 300m	3.10E-03	51.7	0.16	2.2	0.55
2007 - Corine with satellite					
Constant	2.3				
Minor roads 1500m	4.21E-06	55549	0.23	2.6	0.32
Satellite-derived surface NO ₂ 2007	6.37E-02	3.1	0.20	1.3	0.45
Major roads 100m	6.33E-04	0.0	0.00	1.0	0.49
Semi-natural land 600m	-4.30E-03	5.3	-0.02	1.6	0.53
Total built up land 300m	3.13E-03	58.6	0.18	2.2	0.54
Minor roads 1500-10000m	1.01E-07	912789	0.09	1.8	0.55
2005-2007 - Corine with satellite					
Constant	2.3				
Minor roads 1500m	4.09E-06	54754	0.22	2.5	0.37
Satellite-derived surface NO ₂ 2005-2007 ^c	5.29E-02	3.2	0.17	1.3	0.48
Semi-natural land 600m	-4.79E-03	4.4	-0.02	1.6	0.53
Major roads 100m	5.93E-04	0.0	0.00	1.1	0.57
Minor roads 1500-10000m	1.22E-07	961813	0.12	1.8	0.58
Total built up land 300m	3.16E-03	48.3	0.15	2.3	0.60

a. Model building based on natural logarithm of concentration (LN concentration) using 80% of monitoring sites

b. All p-values < 0.000

c. Average of annual satellite-derived surface NO₂ for the three year period

Table S7. Final PM₁₀ models by year

Variables	Model building ^a				
	β^b	IQR	β^* IQR	VIF	Partial Adj-R ²
2005 - Global with satellite					
Constant	3.42				
Y coordinate	-1.87E-07	756842	-0.14	1.1	0.13
Satellite-derived surface PM _{2.5} 2001-6	2.26E-02	6.9	0.16	1.1	0.24
Impervious surface 1000m	2.46E-03	39.1	0.10	1.2	0.31
Tree canopy 500m	-2.86E-03	7.7	-0.02	1.3	0.33
Distance to sea	-2.20E-01	0.4	-0.08	1.1	0.34
Major roads	4.47E-04	0.0	0.00	1.1	0.35
2006 - Corine with satellite					
Constant	3.471				
Satellite-derived surface PM _{2.5} 2001-6	2.19E-02	6.9	0.15	1.1	0.13
Y coordinate	-2.00E-07	780076	-0.16	1.1	0.26
Total built up land 600m	8.86E-04	52.2	0.05	1.9	0.31
Population 1800m	1.04E-06	38155	0.04	1.3	0.33
Distance to sea	-2.42E-01	0.4	-0.09	1.1	0.35
Major roads 100m	1.97E-04	0.0	0.00	1.1	0.36
Semi-natural land 1000m	-1.77E-03	1.8	0.00	1.6	0.37
2007 - Global with satellite					
Constant	3.67				
Satellite-derived surface PM _{2.5} 2001-6	1.93E-02	7.1	0.14	1.1	0.16
Y coordinate	-2.83E-07	778777	-0.22	1.2	0.31
Minor roads 200-2500m	6.03E-07	119820	0.07	1.7	0.42
Altitude	-6.93E-01	0.2	-0.11	1.2	0.47
Minor roads 200m	3.52E-05	1538.0	0.05	1.6	0.48
Major roads	5.07E-04	0.0	0.00	1.1	0.49
Tree canopy 100m	-1.86E-03	8.2	-0.02	1.3	0.50
2005-2007 - Corine with satellite					
Constant	3.61				
Y coordinate	-2.40E-07	683381	-0.16	1.1	0.17
Satellite-derived surface PM _{2.5} 2001-6	2.24E-02	7.1	0.16	1.1	0.31
Semi-natural land 1200m	-2.30E-03	12.9	-0.03	1.7	0.40
Distance to sea	-3.45E-01	0.3	-0.11	1.1	0.43
Population 1800m	7.48E-07	37142	0.03	1.3	0.46
Total built up 400m	1.18E-03	51.0	0.06	1.7	0.47
Major roads 100m	1.42E-04	0.0	0.00	1.1	0.48

a. Model building based on natural logarithm of concentration (LN concentration) using 80% of monitoring sites

b. All p-values < 0.000

Table S8. Summary of model building and evaluation statistics

Year	Model	Model building ^a			Model Evaluation ^b										N
		Adj-R ²	SEE	N	LN concentration		Concentration (µg/m ³)							Regression line ^c	
					R ²	SEE	R ²	RMSE	ME	MAE	MB	MAB			
NO₂															
2005	Corine with satellite	0.58	0.42	1612	0.58	0.44	0.56	11.54	-1.8	8.1	13	37	y=0.98x+2.47	398	
2006	Corine with satellite	0.55	0.43	1674	0.55	0.42	0.50	11.35	-1.5	8.3	11	35	y=0.90x+4.15	425	
2007	Corine with satellite	0.55	0.42	1786	0.59	0.42	0.50	11.80	-1.3	8.5	18	41	y=0.91x+3.76	450	
2005-2007	Corine with satellite	0.60	0.39	1330	0.56	0.42	0.46	11.66	-1.8	8.4	8	34	y=0.84x+6.07	340	
PM₁₀															
2005	Global with satellite	0.35	0.27	1184	0.41	0.26	0.44	7.07	-0.8	5.4	4	21	y=1.13x-2.53	303	
2006	Corine with satellite	0.37	0.26	1263	0.35	0.27	0.36	7.56	-1.2	6.0	3	22	y=1.04x+0.22	321	
2007	Global with satellite	0.50	0.23	1339	0.50	0.23	0.47	6.74	-0.5	4.6	4	17	y=1.07x-1.36	325	
2005-2007	Corine with satellite	0.48	0.22	895	0.46	0.21	0.48	5.99	-0.2	4.4	3	17	y=1.00x+0.39	256	

a. Model building based on natural logarithm of concentration (LN concentration) using 80% of monitoring sites

b. Model evaluation using 20% reserved monitoring sites

c. See Figures S5 and S5 for scatterplots

ME = mean error (µg/m³); MAE = mean absolute error (µg/m³); MB = mean bias (%); MAB = mean absolute bias (%)

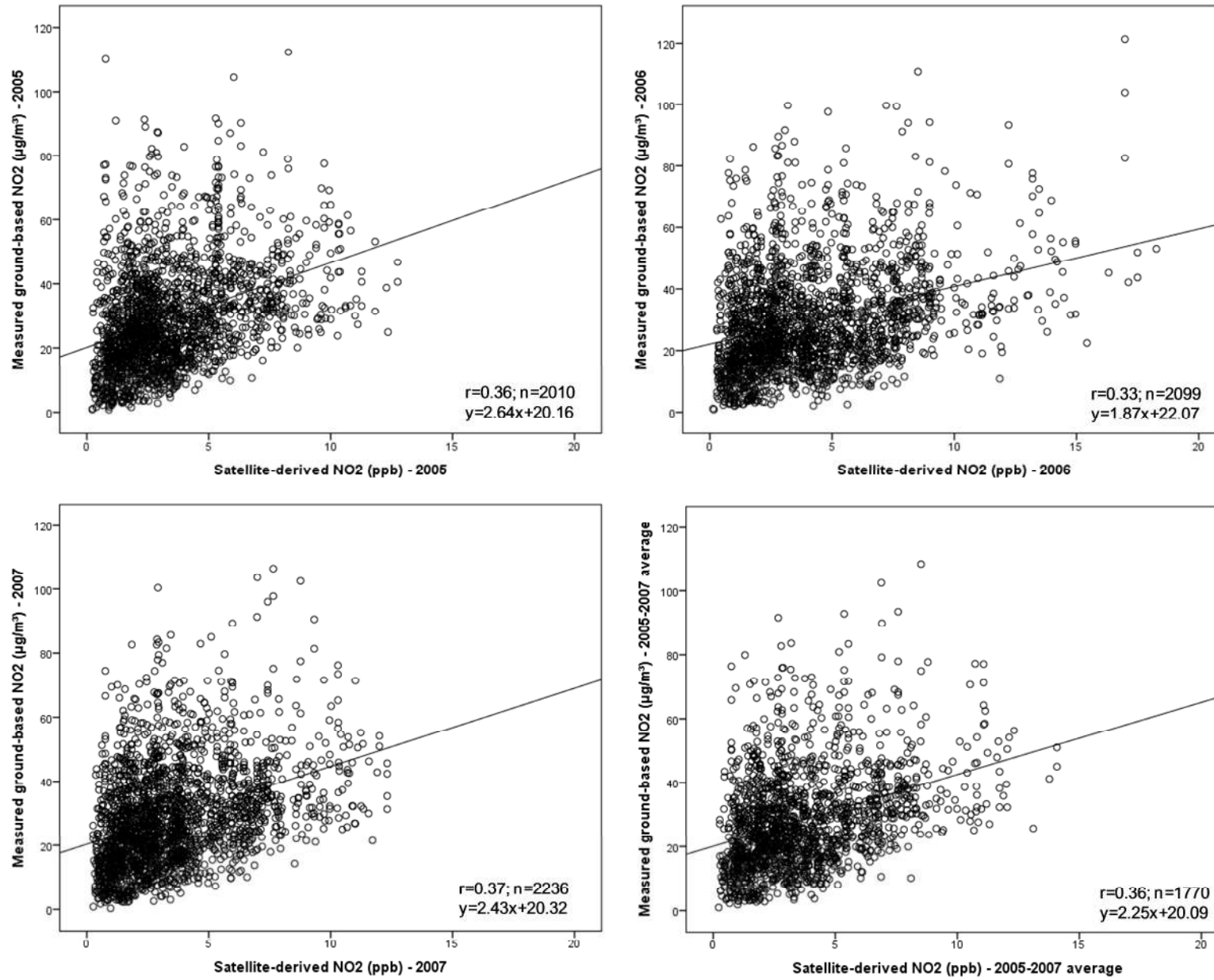


Figure S1. Measured ground-based NO₂ vs. satellite-derived NO₂, at all monitoring sites

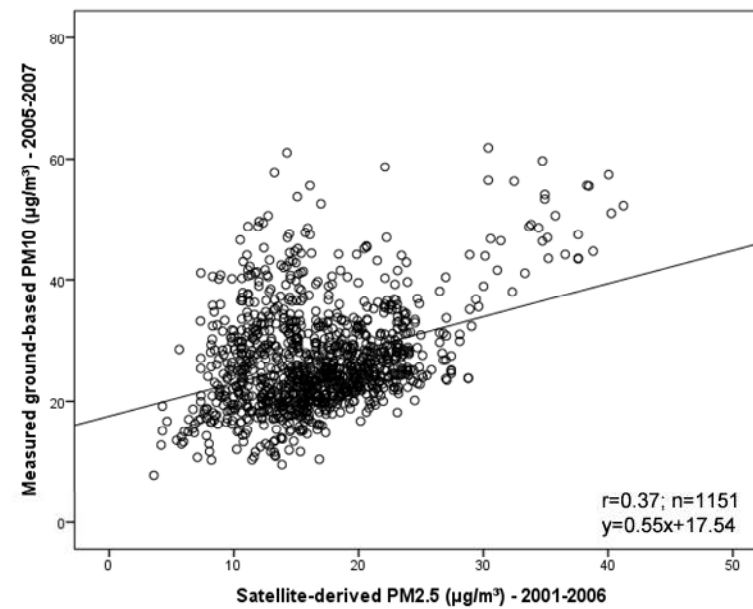
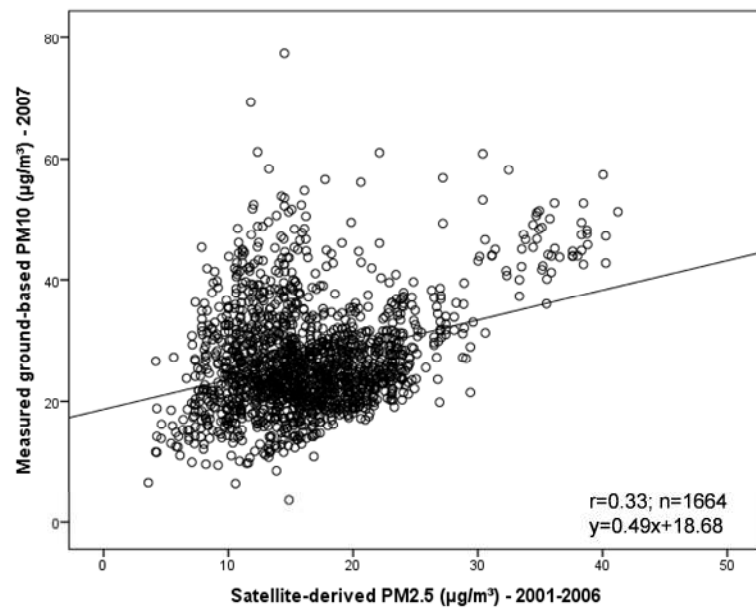
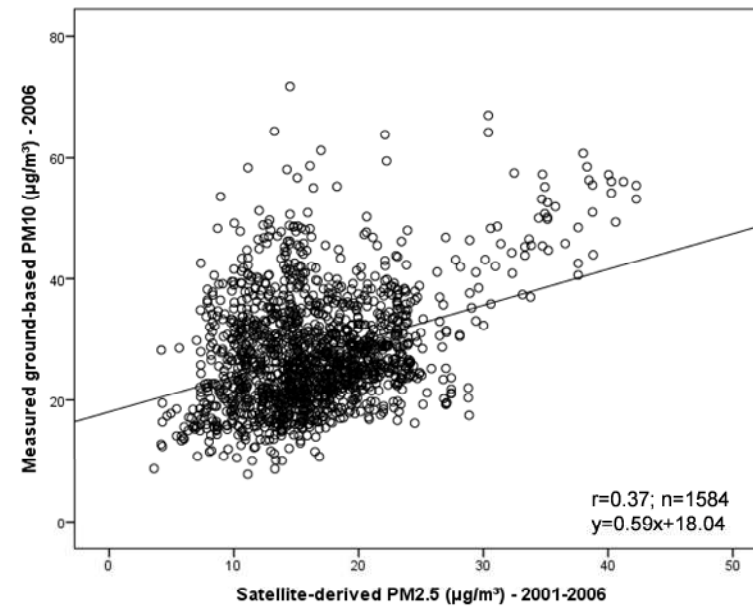
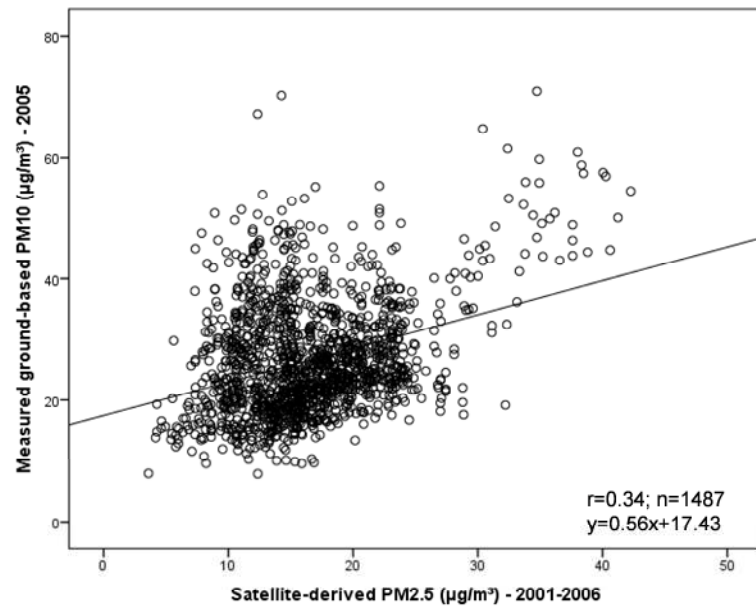


Figure S2. Measured ground-based PM₁₀ vs. mean 2001-2006 satellite-derived PM_{2.5}, at all monitoring sites

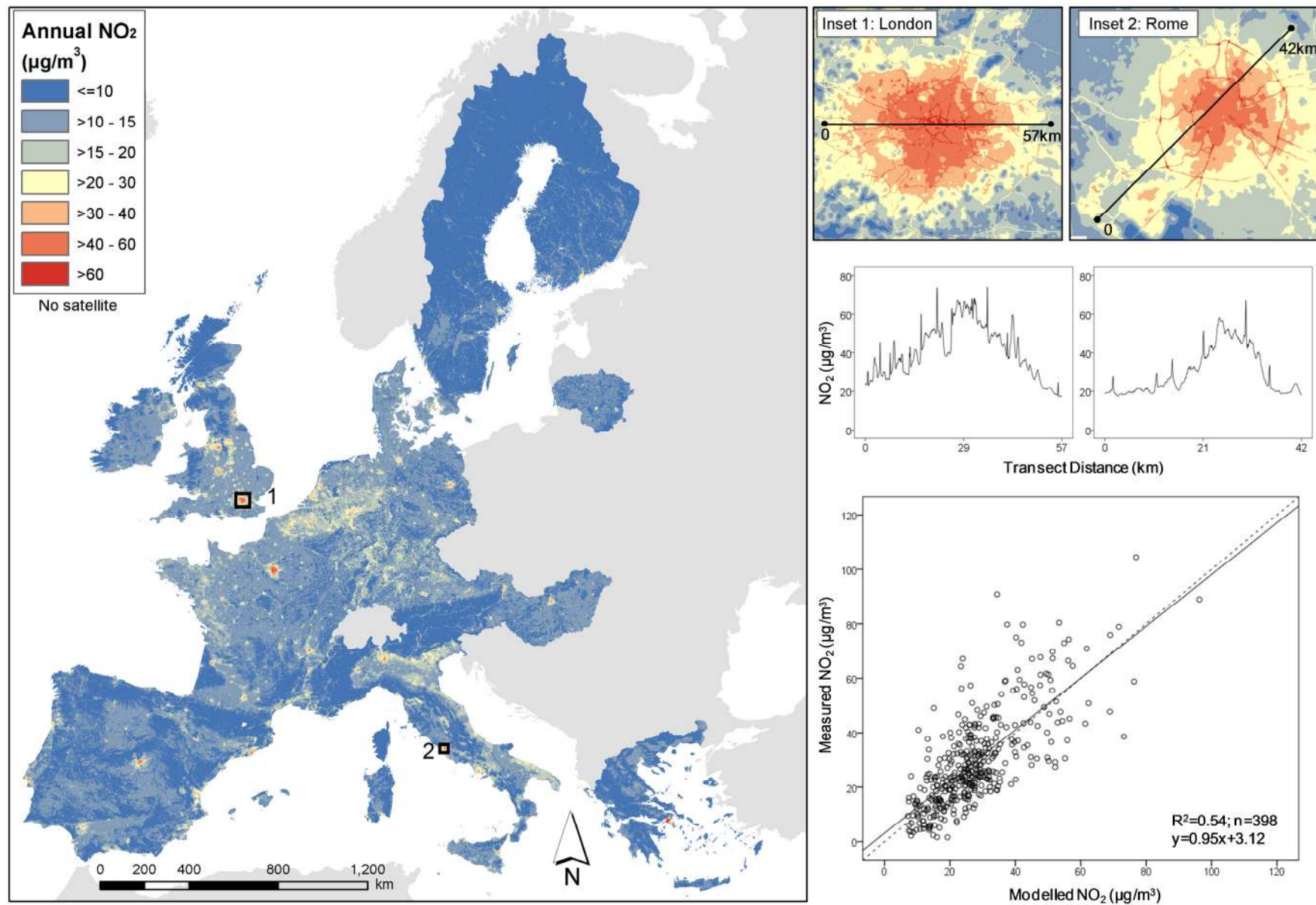


Figure S3. Map and profile plots of NO₂ concentration in 2005 without satellite data; scatterplot of modelled vs. measured NO₂ at evaluation sites

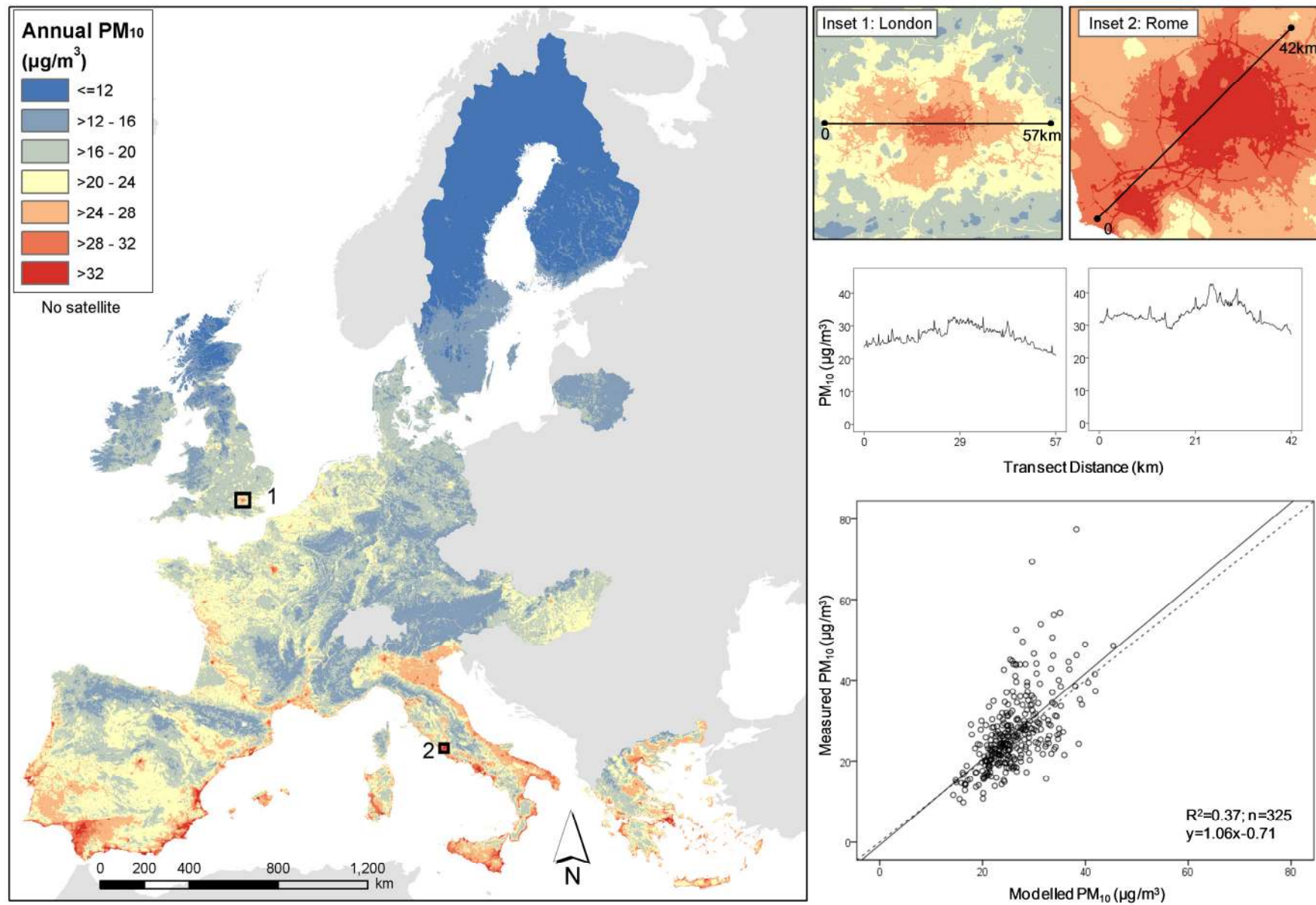


Figure S4. Map and profile plots of PM₁₀ concentration in 2007 without satellite data; scatterplot of modelled vs. measured PM₁₀ at evaluation sites

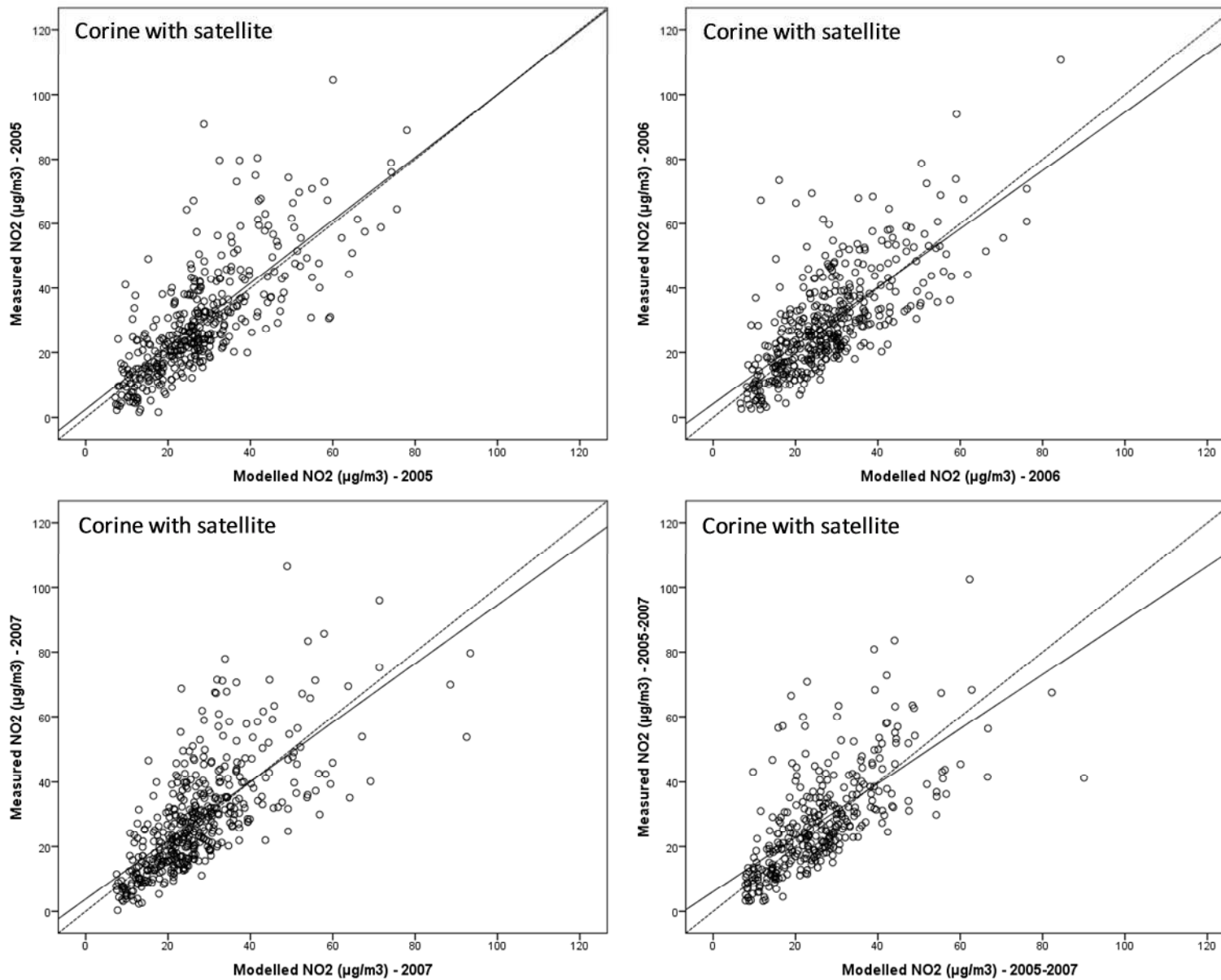


Figure S5. Modelled vs. measured NO₂ concentration (µg/m³) at evaluation sites for final models shown in Tables S3 and S5

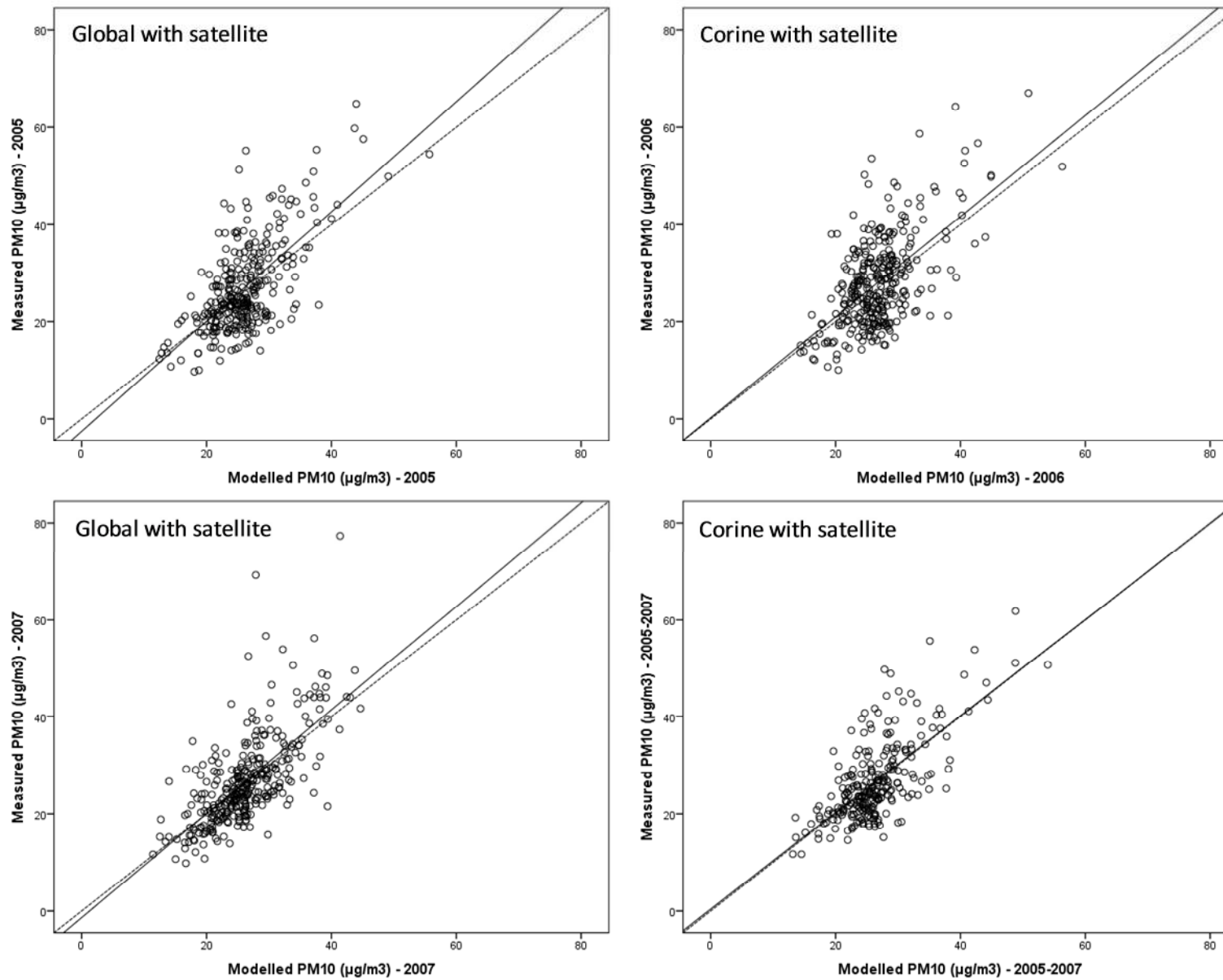


Figure S6. Modelled vs. measured PM₁₀ concentration ($\mu\text{g}/\text{m}^3$) at evaluation sites for final models shown in Tables S4 and S5