Supporting Information

Authors: Jeff D. Yanosky*, Cathryn C. Tonne, Sean D. Beevers, Paul Wilkinson, Frank J. Kelly

Title: "Modeling Exposures to the Oxidative Potential of PM_{10}"

Number of pages: 12

Number of Tables: 2

Number of Figures: 5

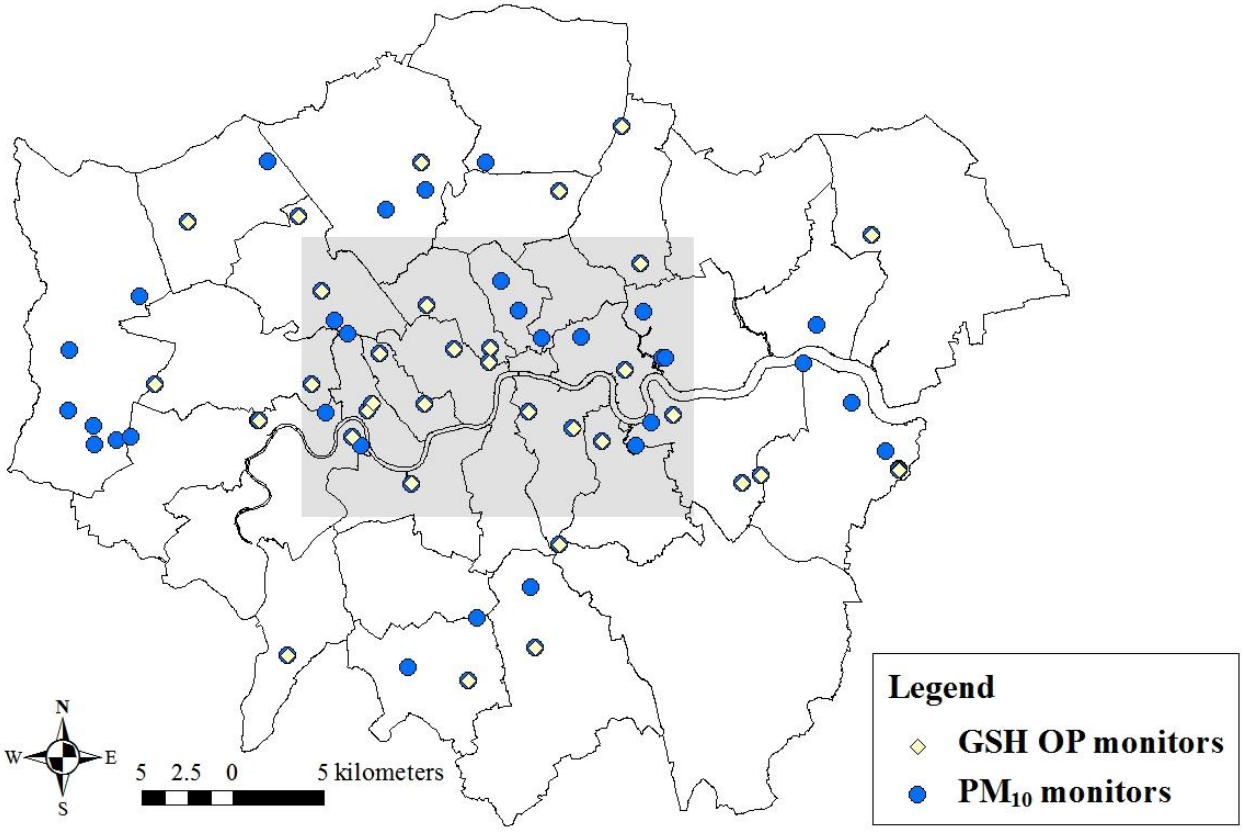


Figure S1. Map of the study domain of greater London, U.K. and GSH OP and PM_{10} monitoring locations; lines are the London borough boundaries; shaded region is the area shown in Figures 1 and S3-S5.

Table S1. Summary of model fit and key weekly cross-validation (CV) results from the covariate selection process for GSH OP and PM_{10} models.

Pollutant	Model[A]	GIS covariates	Model R^2	Cross-validation $R^{2\,B}$
GSH OP	Referent model	None	0.28	0.21
	Univariate models	Urban volumetric density within 100 m	0.39	0.29
		Green coverage ratio within 250 m	0.30	0.21
		Cumulative traffic of heavy-goods vehicles within 50 m	0.30	0.22
		Cumulative traffic of heavy-goods	0.30	0.21

		vehicles within 100 m		
		Gridded NO_X tailpipe emissions from buses, cars, motorcycles, and taxis	0.31	0.22
		NO_X tailpipe emissions from heavy-goods vehicles within 50 m	0.30	0.20
		PM_{10} tailpipe emissions from buses, cars, motorcycles, and taxis within 50 m	0.46	0.39
		Gridded PM_{10} brake and tire wear emissions from buses, cars, motorcycles, and taxis	0.32	0.23
		PM_{10} brake and tire wear emissions from all vehicles within 50 m	0.44	0.36
	Multivariable model[C]	PM_{10} brake and tire wear emissions from all vehicles within 50 m and NO_X tailpipe emissions from heavy-goods vehicles within 100 m[D]	0.50	0.42
	Seasonal GIS effects	As above, but seasonal slope for PM_{10} brake and tire wear emissions from all vehicles within 50 m	0.52	0.44
PM_{10}	Referent model	None	0.74	0.74
	Univariate model	Mean of 2003, 2004, and 2006 annual-average PM_{10} predictions from hybrid emissions/dispersion model	0.84	0.83

[A]: Each model included a smooth function for time trend, *h(t)*. Also, as discussed in the main text, GSH OP models assumed a compound-symmetric covariance structure among the within-site errors whereas PM_{10} models did not assume a non-zero covariance structure.

[B]: Across weekly averages (2,118 at 34 locations for GSH OP and 12,041 at 66 locations for PM_{10}). For PM_{10}, two high values at two sites were removed as outliers.

[C]: After removing non-significant terms and those that did not increase the CV R^2.

[D]: 50 m buffer used initially though 100 m buffer performed slightly better in multivariable models.

Table S2. Summary of prediction model parameters for GSH OP and PM_{10} (Equations 2 and 3 in the main text, respectively).

Pollutant	Model terms	Mean (SE[A])	p-value[B]
GSH OP	α, intercept (autumn season is the reference category)	0.751 (0.052)	<2e-16
	β_1: slope for NOx tailpipe emissions within 100 m from heavy-goods vehicles[C]	-0.123 (0.030)	4.8e-5
	β_2; slope for x_{Spring}, an indicator for spring season	-0.227	2.5e-5

		(0.054)	
	β_3; slope for x_{Summer}, an indicator for summer season	-0.194 (0.048)	4.9e-5
	β_4; slope for x_{Winter}, an indicator for winter season	-0.122 (0.050)	0.0146
	β_5; slope for z_{2i}: PM_{10} brake and tire wear emissions within 50 m from all vehicles[C] (autumn season is the reference category)	5.55 (1.549)	3.5e-4
	β_6; slope for $z_{2i}x_{Spring}$; z_{2i} and x_{Spring} as above	5.45 (0.701)	1.2e-14
	β_7; slope for $z_{2i}x_{Summer}$; z_{2i} and x_{Summer} as above	5.75 (0.732)	6.5e-15
	β_8; slope for $z_{2i}x_{Winter}$; z_{2i} and x_{Winter} as above	1.70 (0.716)	0.0178
	$h(t)$; smooth function for time trend (40.5 estimated df)	See Figure S2A	<2e-16
	e_{it}; errors	21.4% of total variance is within-site	
PM_{10}	α; intercept	1.77 (0.323)	4.5e-08
	β_1; slope for z_{1i}: Mean of 2003, 2004, and 2006 predicted annual-average PM_{10}[D]	1.01 (0.012)	<2e-16
	$h(t)$; smooth function for time trend (220.0 estimated df)	See Figure S2B	<2e-16

[A]: SE is standard error; these values are model-based and do not include uncertainty from model specification.

[B]: p-values are approximate due to estimation of smoothing parameter[1].

[C]: Units are tonnes year^{-1}.

[D]: From hybrid emissions-dispersion/regression model[2-3]; units are μg m^{-3}.

Calculation of mean fractional bias and mean fractional error

We calculated mean fractional bias (MFB; %) and mean fractional error (MFE; %) for the N pairs of model predictions ($\hat{y}_i$) and measurements (y_i) using the following formulas:

$$MFB = \frac{1}{N}\sum_{i=1}^{N}\frac{(\hat{y}_i - y_i)}{(\frac{\hat{y}_i + y_i}{2})} \quad [1]$$

$$MFE = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{(\frac{\hat{y}_i + y_i}{2})} \quad [2]$$

Residual correlation and diagnostics

We evaluated assumptions about model residuals using the untransformed, original-scale data, as well as those using natural log and square root transformations. Because none of these

transformations markedly improved residual diagnostics or improved predictive accuracy in preliminary models, we modeled GSH OP without transformation. Despite our use of a compound symmetric covariance structure for the within-site errors (chosen based on the cross-validation (CV) R^2), we found evidence of low to moderate temporal autocorrelation in residuals from the final GSH OP prediction model, though little (Pearson's r <0.2) remained for lags greater than two weeks. The temporal alignment of the OP data resulted in several weekly values being repeated (those for which the monitoring period was longer than one week): these are displayed as successive horizontal points in Figure S2A. Though counterintuitive, temporal autocorrelation increased slightly when an $AR(1)$ covariance structure was evaluated. Residual diagnostics showed that other modeling assumptions were reasonable, with little to no spatial correlation or spatio-temporal interaction remaining (see below).

Spatio-temporal interaction and kriging

The model in Equation 1 includes a smooth spatial trend $g(s_i)$ that is assumed constant over time. We provided this function the greatest flexibility possible by using a basis dimension of $k = 0.9I$. The complexity of this function was determined during model fitting using bivariate penalized thin-plate splines. To evaluate whether any spatio-temporal interaction remained in the data and to potentially explain additional variability in the outcome, we added seasonal spatial terms, $g_{Season}(s_i)$, and weekly spatial terms, $g_t(s_i)$, separately, to the GSH OP and PM_{10} models. Time-varying spatial terms were also specified using bivariate penalized thin-plate splines. Weekly spatial models were fit in a back-fitting arrangement[4] in which weekly smooth spatial terms were fit to the model residuals, iterating between the non-spatial and spatial model components until convergence[5]. We also examined seasonal and weekly semivariograms of the model residuals to assess the extent of any remaining spatial variability, and examined the level

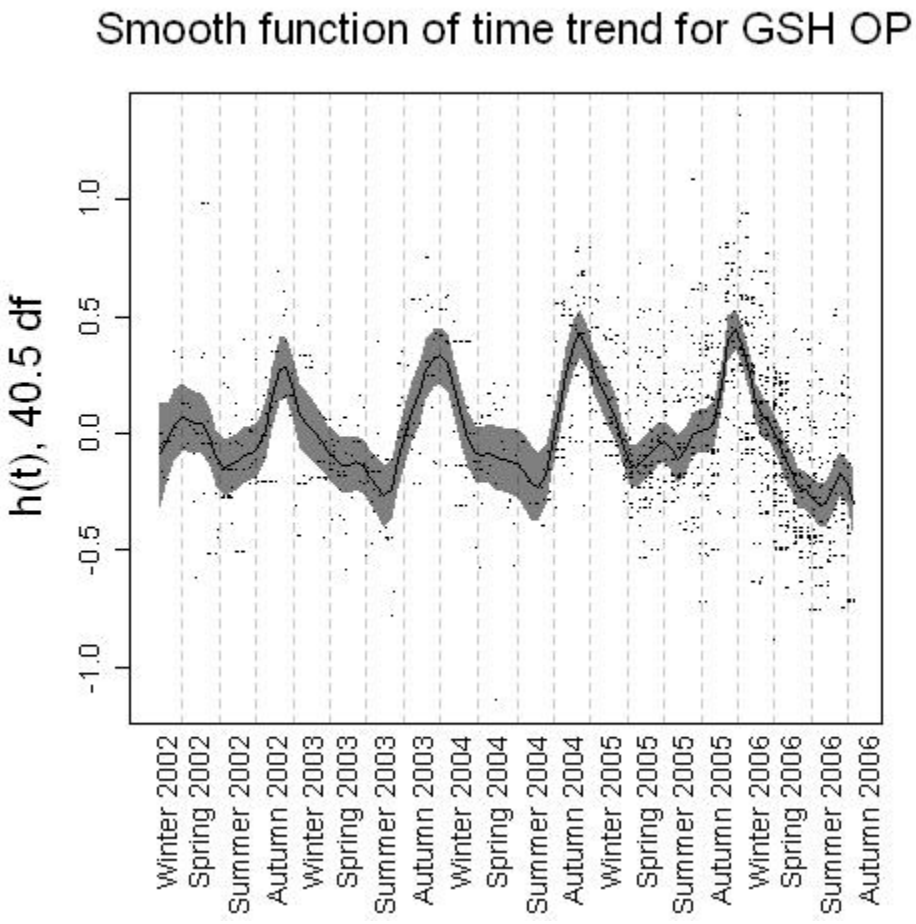
of serial autocorrelation in model residuals by creating autocorrelation plots for each monitoring site and summarizing these across sites.

To evaluate the sensitivity of model performance to different spatial modeling approaches, we evaluated ordinary and simple kriging models in which residual spatial variation was modeled as a mean-zero Gaussian stochastic process. These spatial models were embedded within the GAM structure, with the covariance of this process modeled as the sum of a nugget parameter (τ^2) plus a Matern covariance, parameterized by the partial sill (σ^2), range (φ, limited to 0-1000), and differentiability parameter (κ, fixed at 1). Ordinary and simple kriging models were fit using the `krige.conv()` function in the geoR package[6] for R, with variogram parameters estimated using maximum likelihood via the `likfit()` function. In the event that `likfit()` failed to converge, we used `variofit()` instead. To fit these models, we again iterated between the non-spatial and spatial model components until convergence. In addition to kriging models with a single spatial term, we evaluated kriging models that used seasonal and weekly spatial terms, as for the spatial smoothing spline models. For the single and seasonal kriging models, the intercept α was removed from the non-spatial component and so was estimated as the mean parameter using ordinary kriging. In weekly kriging models, the intercept α was included in the non-spatial component of the model, so simple kriging, with the beta parameter set to zero, was used to model residual weekly spatial variability.

Modeling residual spatial variability, either using smoothing splines or ordinary/simple kriging, at either the seasonal or weekly level (though insufficient data were available to fit weekly models before August 2004), decreased rather than increased the predictive accuracy of GSH OP models as assessed by the weekly CV R^2 (data not shown), even though the model R^2 increased in some cases. This decrease in predictive accuracy indicates that spatial variation in

GSH OP levels occurred on a spatial scale smaller than that able to be described by smoothing splines or kriging, given the density of available monitoring. Results for PM_{10}, similar to those for GSH OP, showed that weekly or seasonal spatial modeling, using either smoothing splines or kriging, decreased rather than increased predictive accuracy (data not shown).

A)



B)

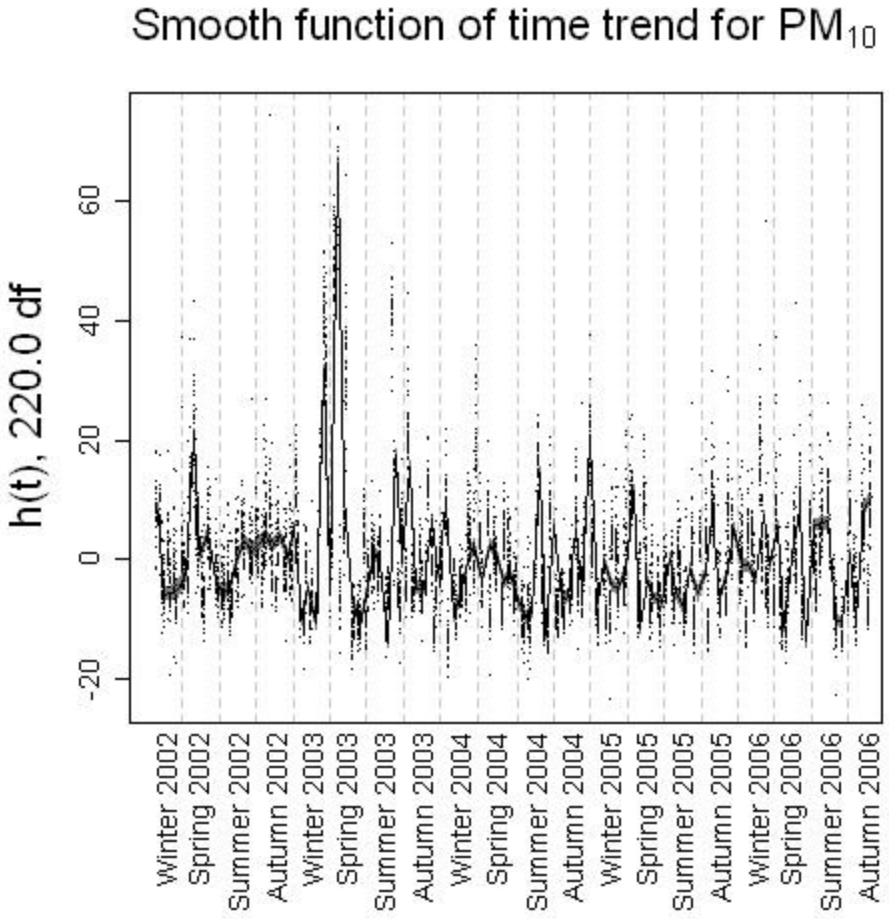
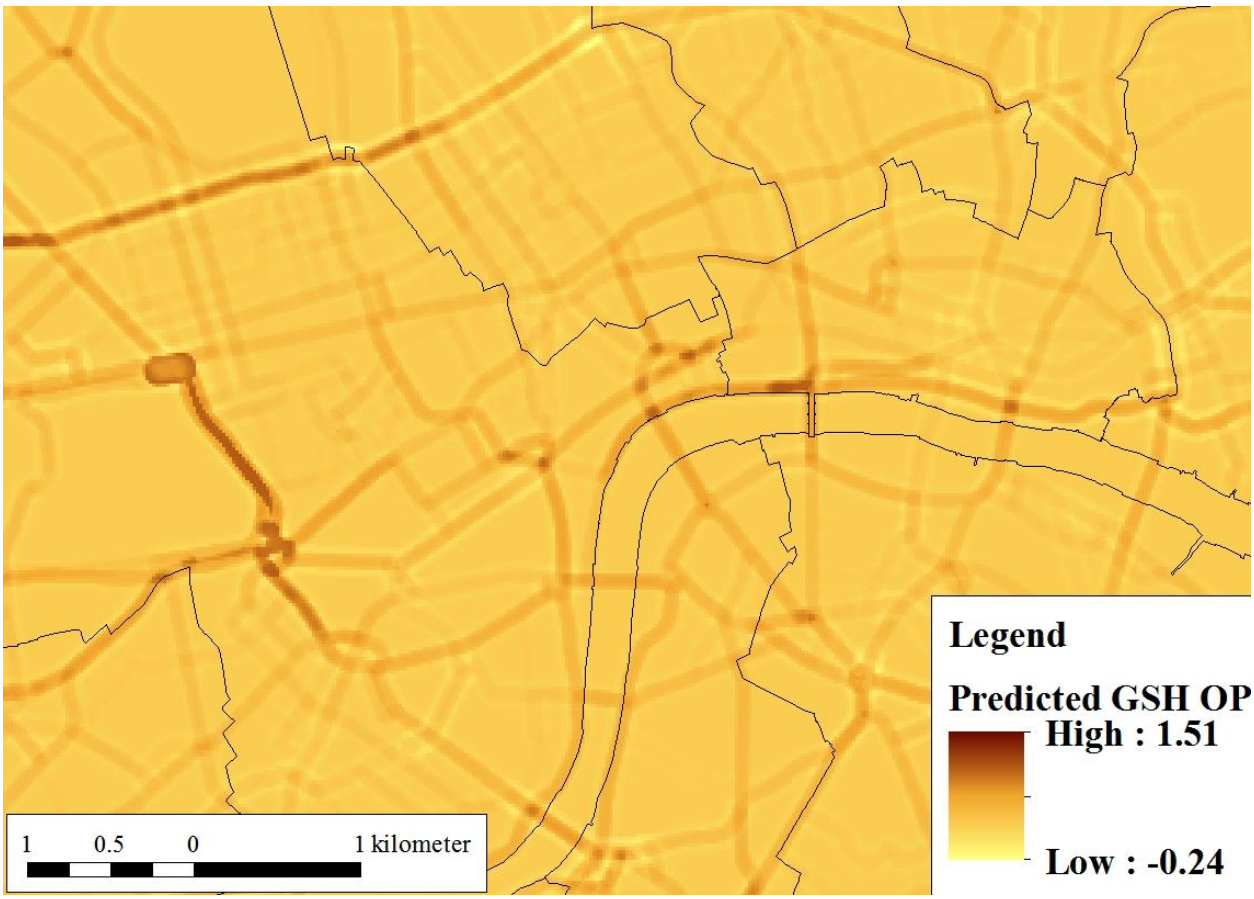
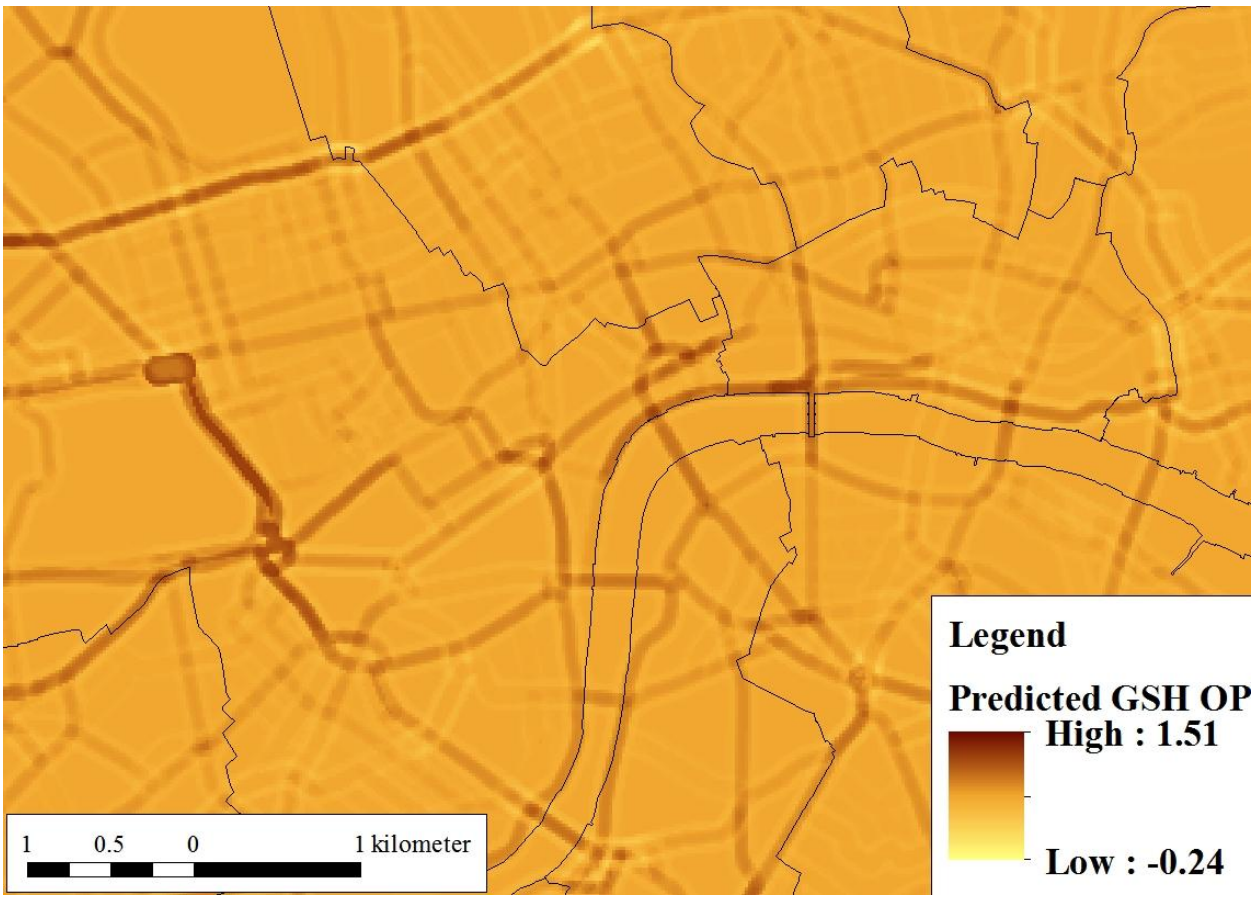


Figure S2. Plots of fitted smooth functions of time trend in prediction models for A) GSH OP and B) PM_{10}. The shaded regions and points display point-wise standard errors (including uncertainty about the overall mean) and partial residuals, respectively.

A)



B)



C)

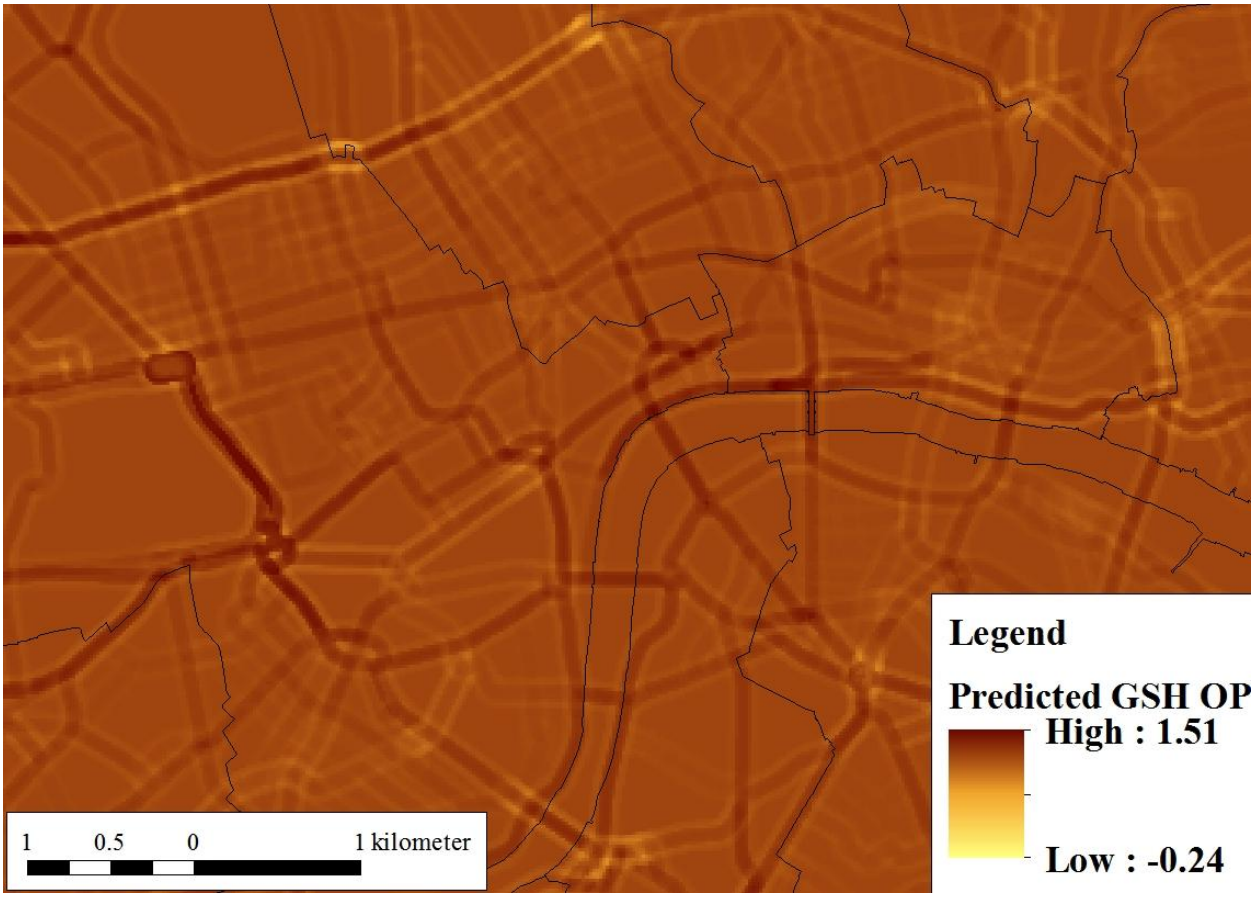


Figure S3. Maps of predicted GSH OP levels in OP μg^{-1} in a selected area of central London, U.K.; lines show the borough boundaries. A) Lowest week (beginning July 3, 2006, corresponding to Summer 2006 in Figure S3A; values ranged from -0.24 to 1.05 OP μg^{-1}); B) Mean across 2002-2006 (values ranged from 0.11 to 1.21 OP μg^{-1}); C) Highest week (beginning November 8, 2004, corresponding to Autumn 2004 in Figure S3A; values ranged from 0.71 to 1.51 OP μg^{-1}).

Predictive accuracy by site descriptors/categories

The GSH OP model performed well across years and in highly urban of greater London, though slightly less well in less urban areas. Predictive accuracy was also comparable across seasons, though performance in the summer season was slightly better than in other seasons.

Predictive accuracy was lower at kerbside sites (those a few meters from the roadway) for weekly data (Table 1; CV R^2=0.13), though data were only available at only four sites. Thus, limited spatial variability was present and the model did not represent temporal variability well at these sites, perhaps because of the strong effects of very near traffic which may have varied in intensity during the week, or because of micro-scale (0-100 m) meteorological impacts such as changes in wind direction or effects of street canyons. This suggests that there is substantial spatio-temporal interaction occurring at kerbside locations, but that this variability is occurring on too small a spatial scale to be captured by the density of the monitors. However, predictive accuracy increased markedly among kerbside sites when comparing only spatial variability (spatial CV R^2=0.71). In addition, predictive accuracy for weekly levels was higher at roadside sites (those more than a few meters from the road) (Table 1; CV R^2=0.36) also highly influenced by nearby traffic. Model predictions of weekly GSH OP levels were also slightly less precise at kerbside and roadside sites compared to urban background sites (Table 1; RMSPE=0.31 and 0.32 vs. 0.25, respectively).

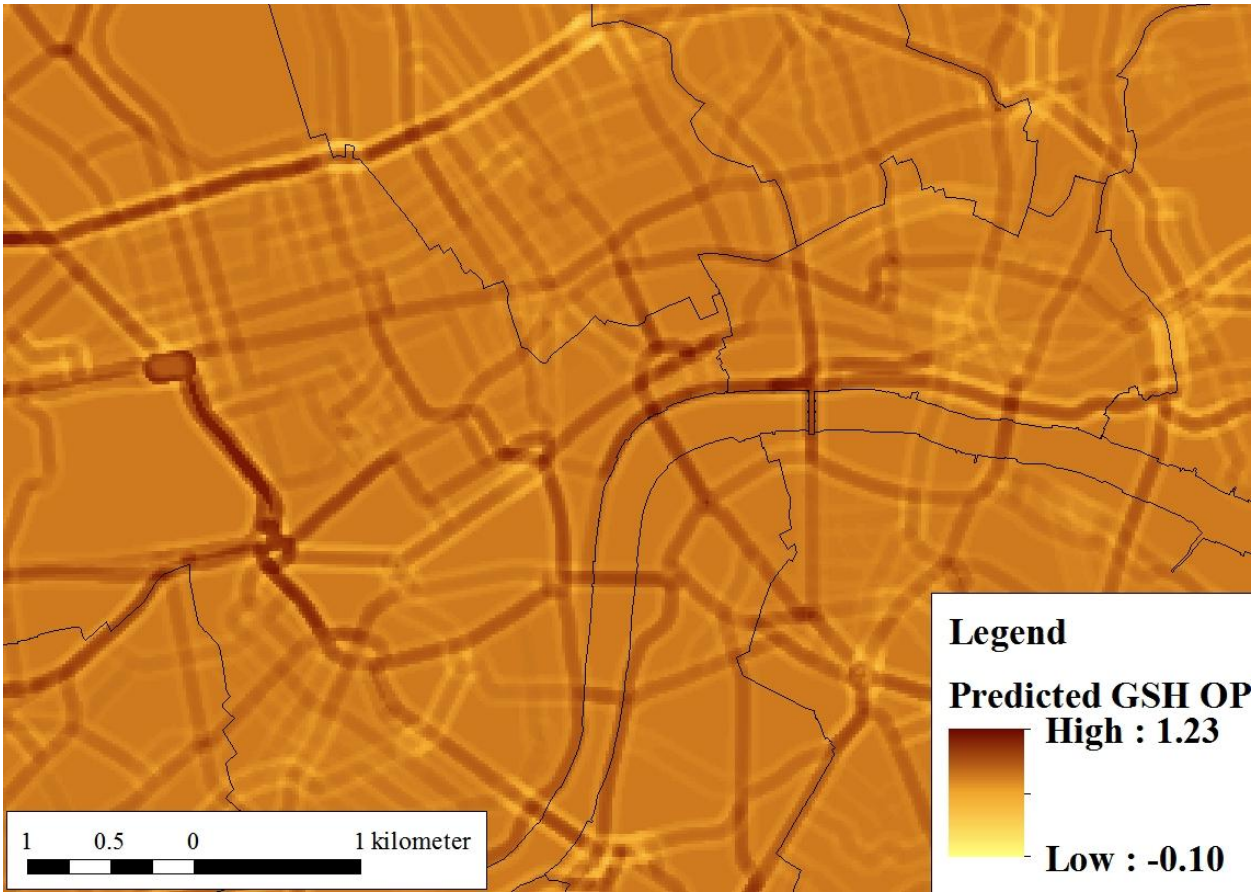
For the PM_{10} model, predictive accuracy increased slightly for the year 2003 (Table 1; CV R^2=0.92) due to the higher levels (and therefore larger range) of PM_{10} in late March 2003 (Figure S2B, in Spring 2003). Model predictions were less precise at kerbside locations (Table 1; RMSPE=5.74 μg m^{-3} at kerbside sites vs. 4.99 at roadside, 3.43 at suburban, and 3.64 at urban background sites).

Our results also demonstrate the ability of CV to select appropriate predictor variables and model forms without leading to over-fitting of the data (*i.e.*, modeling noise in the measurements rather than signal). Though intuitive, we found that spatial modeling using smooth spatial terms or kriging (see below) decreased predictive accuracy and led to increased model over-fitting for

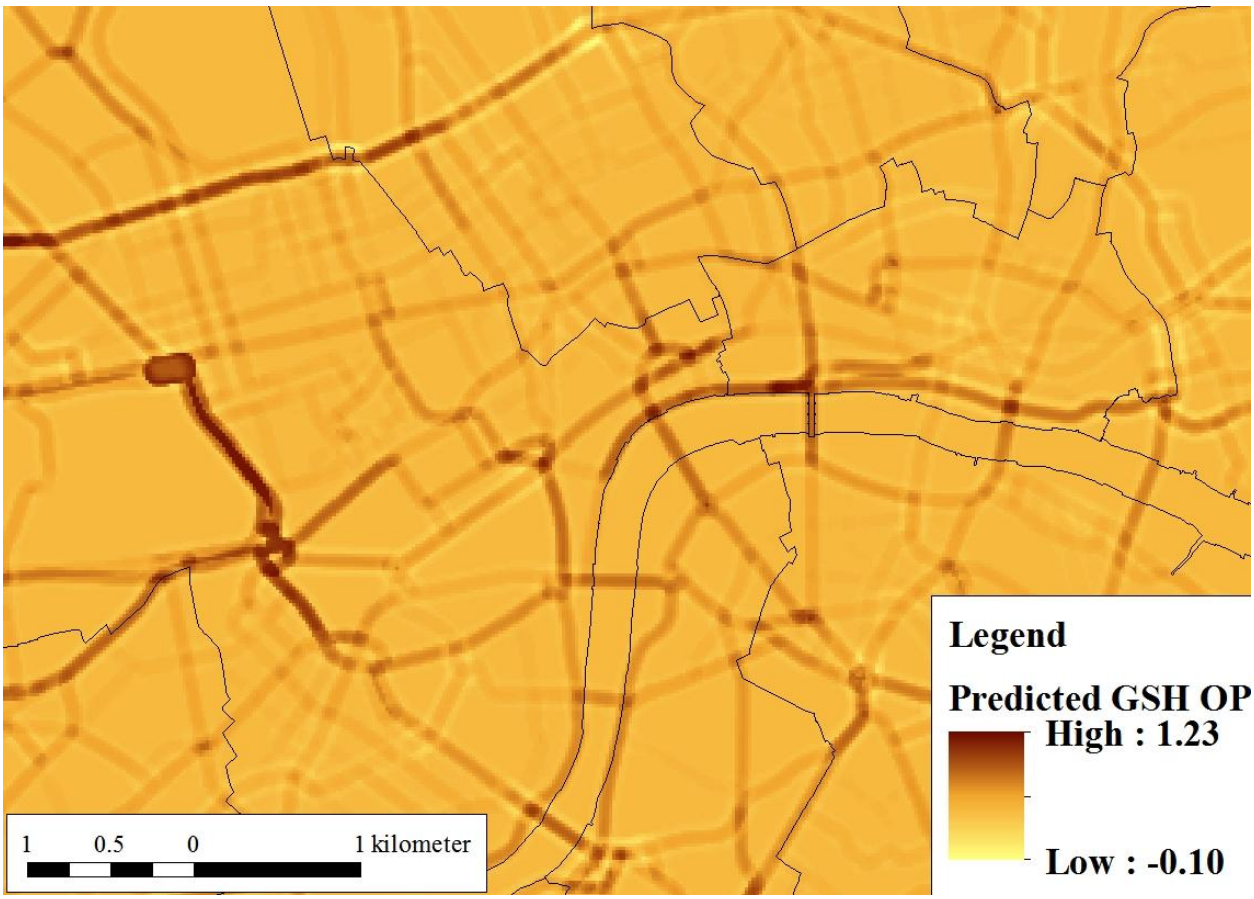
GSH OP and PM_{10}, even as it generally improved model R^2 values. Thus, our results highlight the importance of using CV for model selection in predictive models rather than the model R^2, AIC, or other model-fit based measures.

Seasonal spatial variation in GSH OP

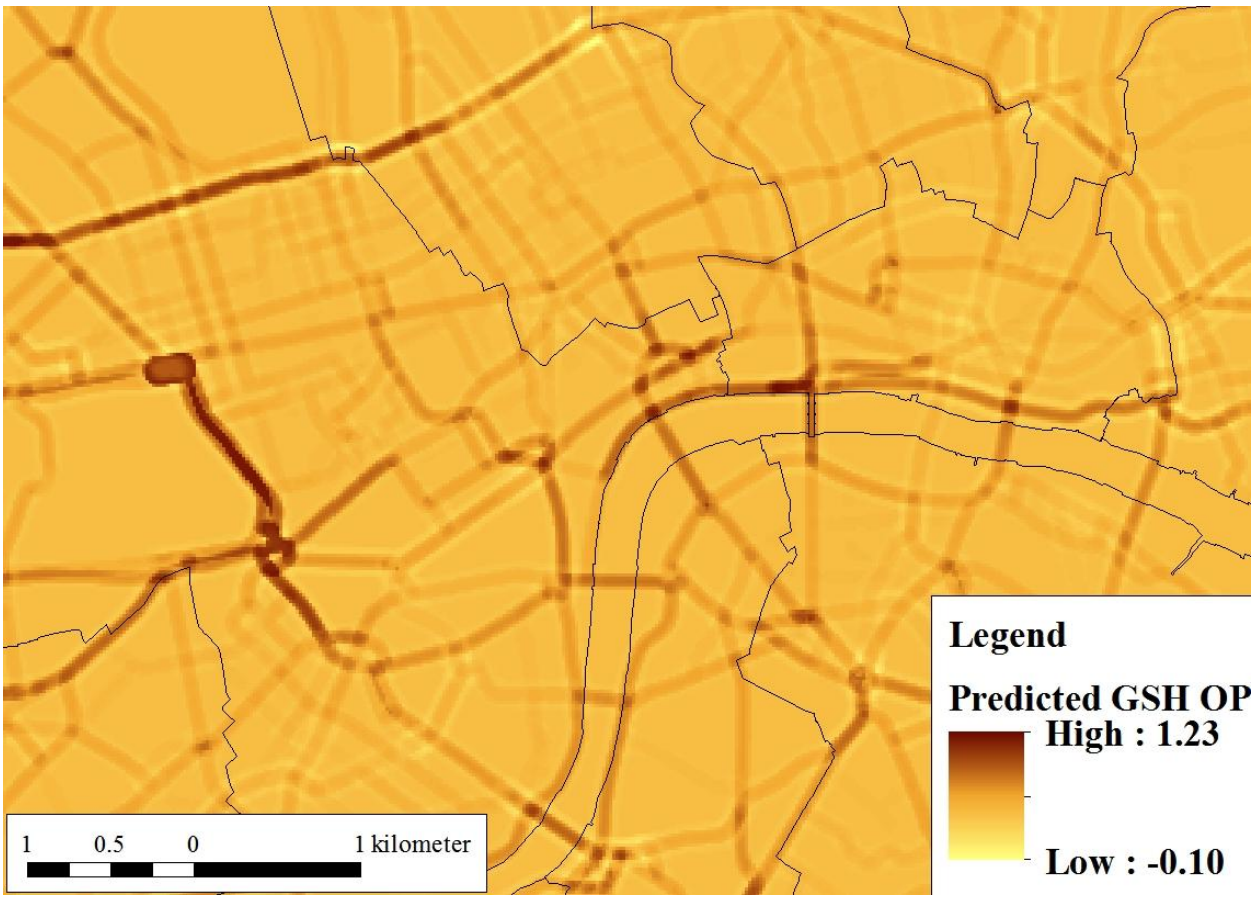
A)



B)



C)



D)

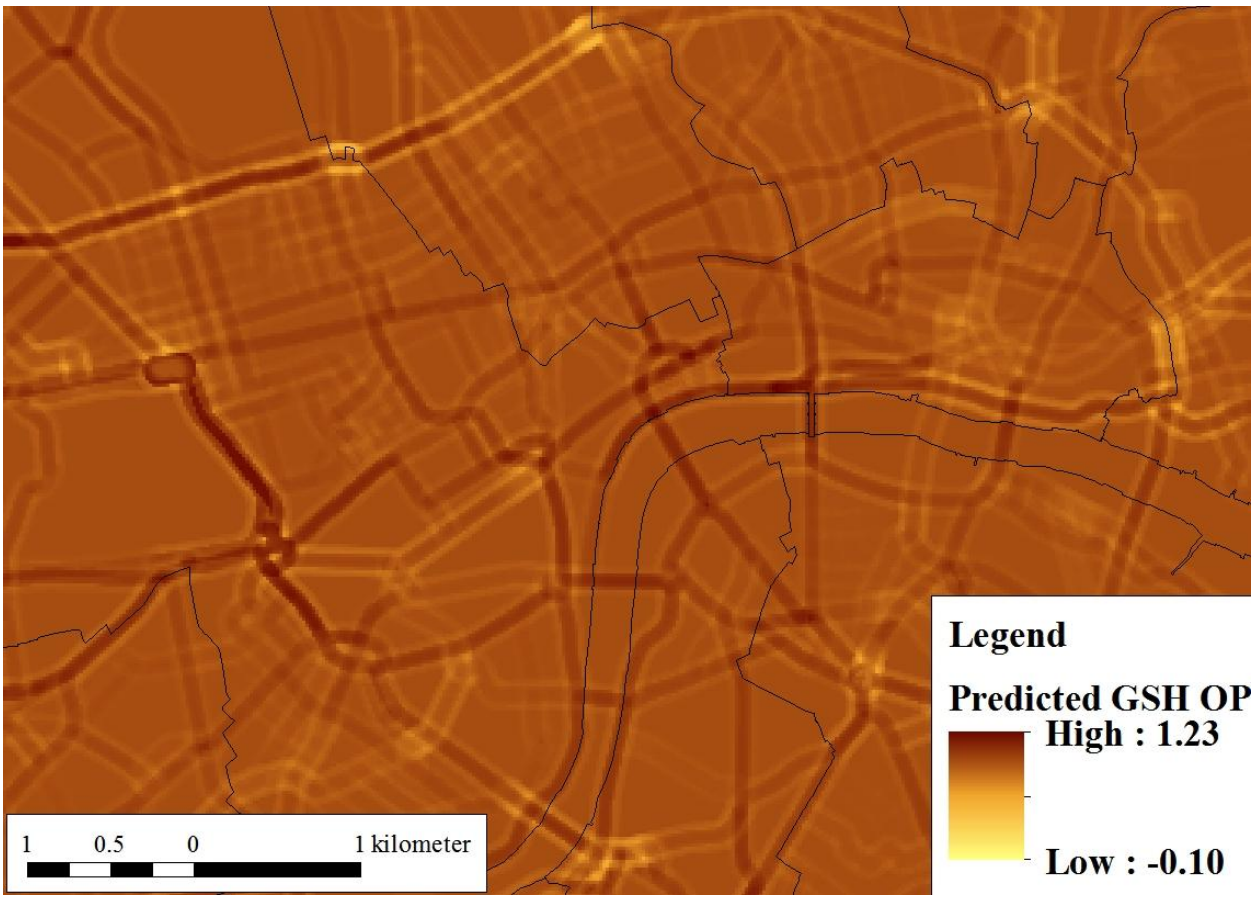
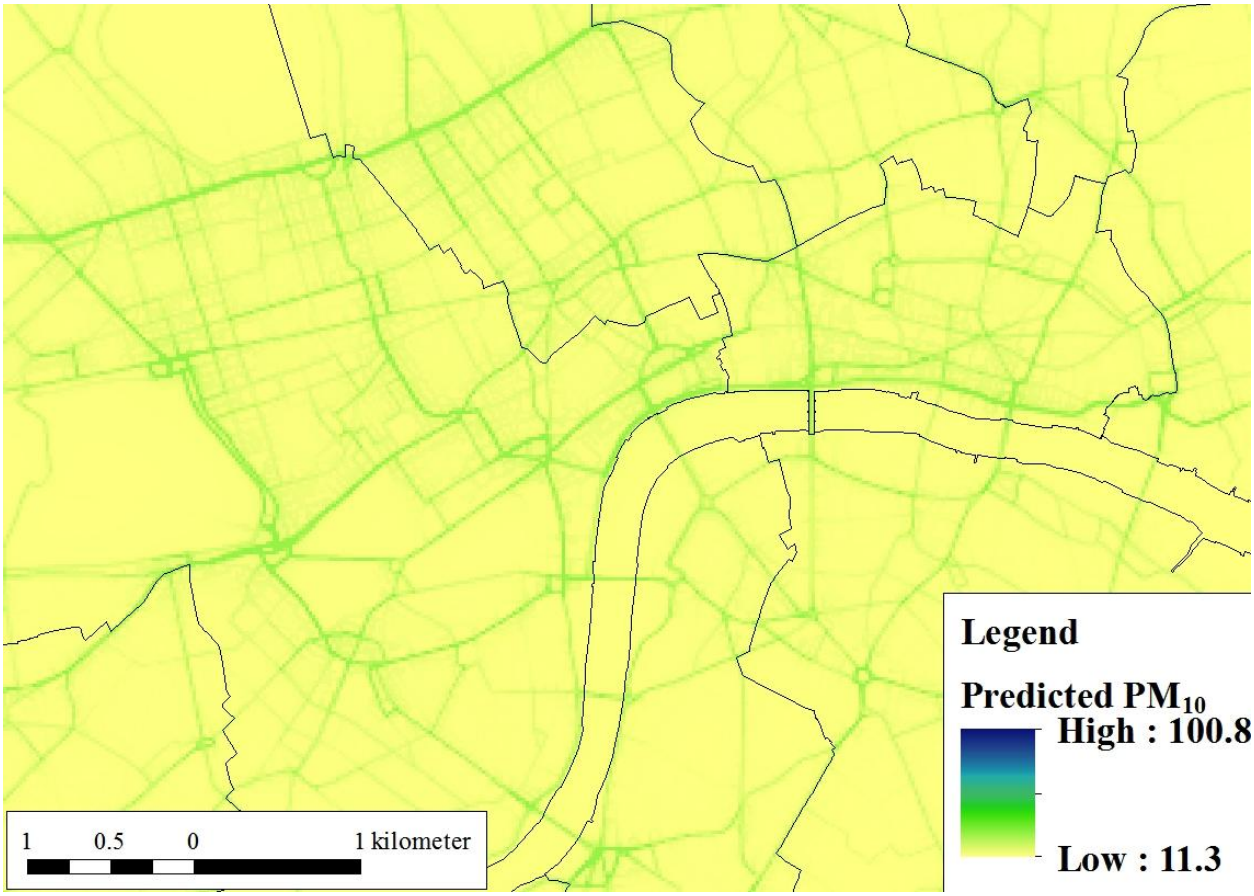


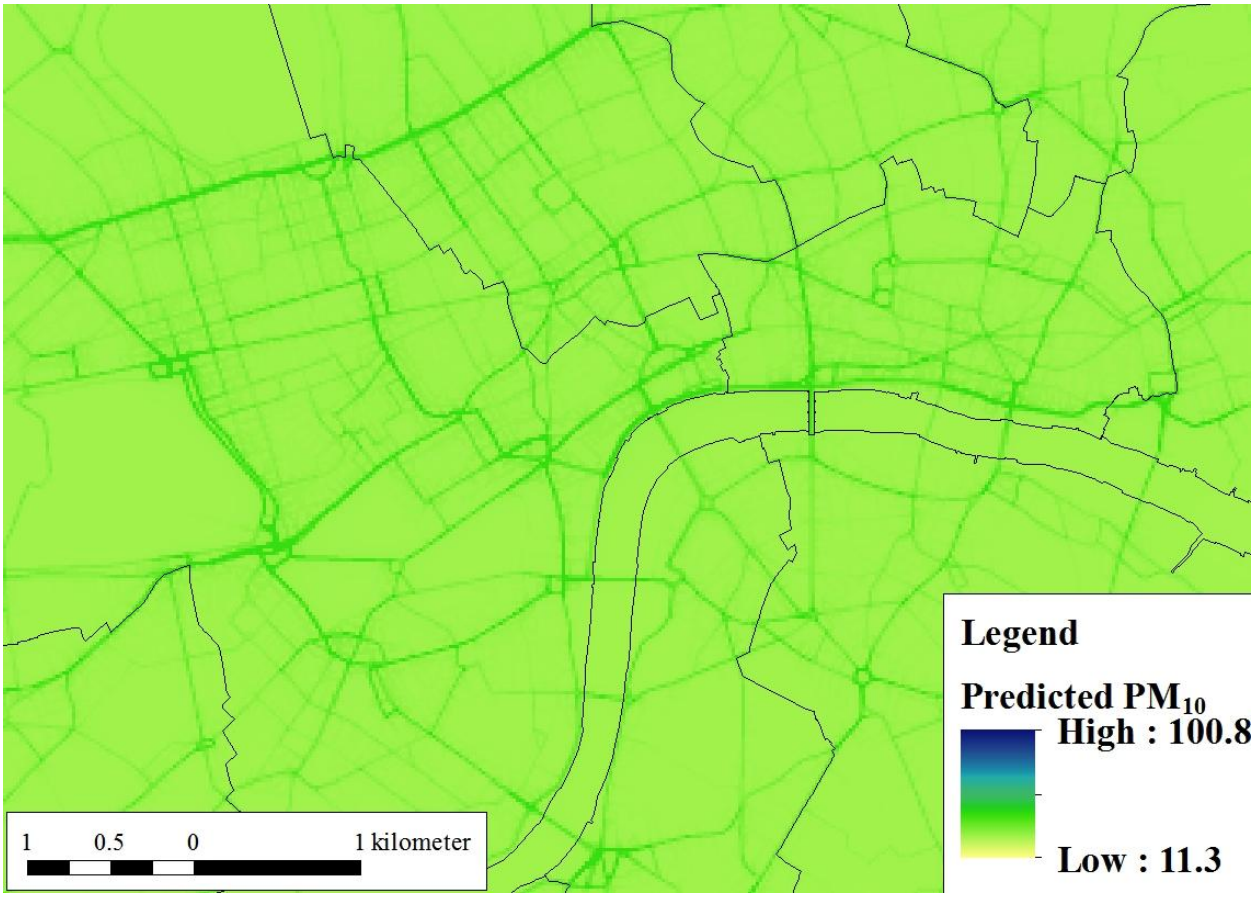
Figure S4. Maps showing spatial variation in seasonal mean predicted GSH OP levels in OP μg^{-1} in a selected area of central London; lines show the borough boundaries. A) winter, B) spring, C) summer, D) autumn.

We observed different slopes for PM_{10} brake and tire wear emissions from all vehicles within 50 m among the seasons. These result in different near-road gradients of predicted GSH OP in different seasons. In the autumn season, the slope is lowest and therefore the elevation in near-road GSH OP levels is less pronounced compared to the surrounding area. In contrast, in the summer the slope is highest and the increase in GSH OP in areas near roads is more evident. Also note the negative slope for NO_X tailpipe emissions from heavy-goods vehicles within 50 m remains constant across seasons, and therefore the areas between 50 and 100 m from roadways exhibit slightly lower predicted GSH OP levels than those further than 100 m.

A)



B)



C)

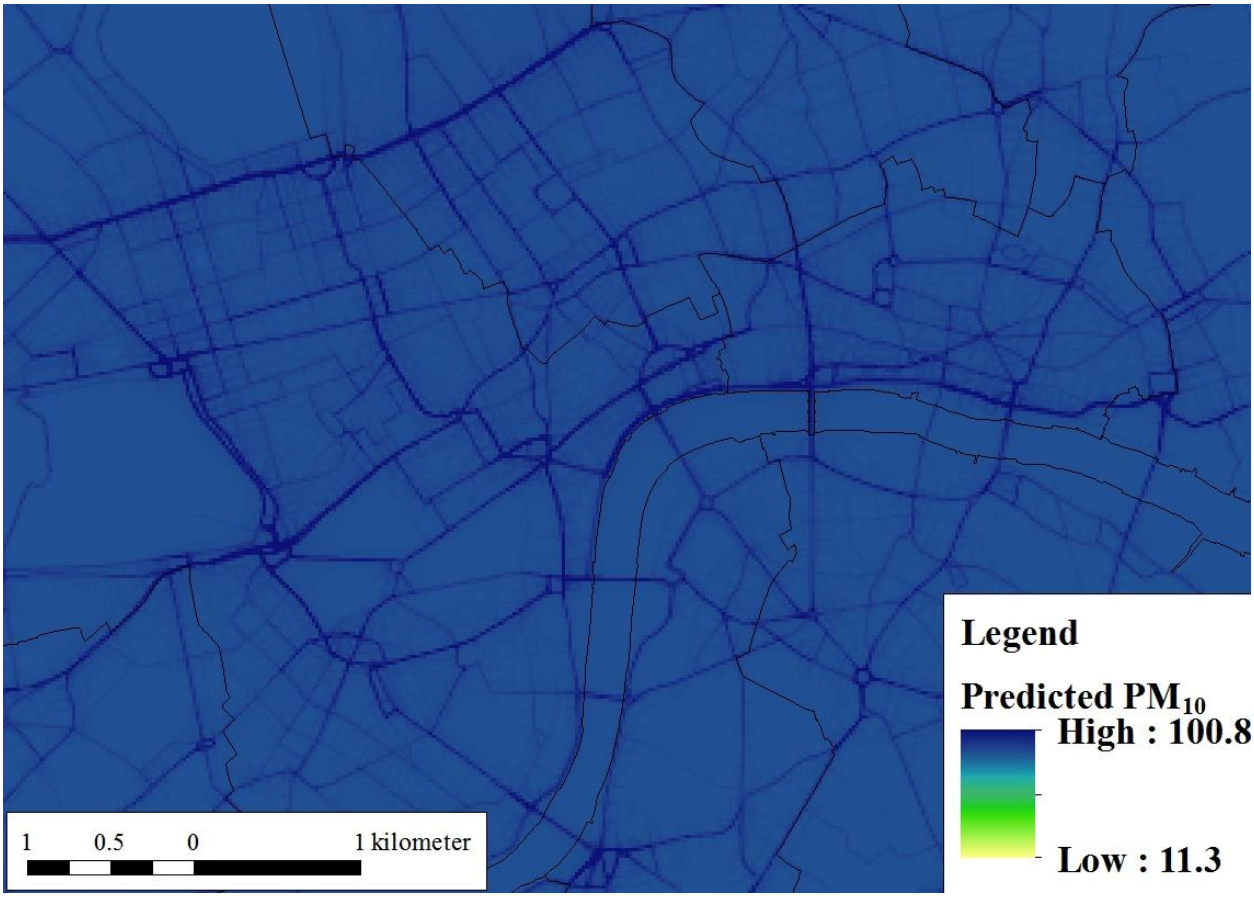


Figure S5. Maps of predicted PM_{10} levels in $\mu g\ m^{-3}$ in a selected area of central London, U.K.; lines show the borough boundaries. A) Lowest week (beginning May 22, 2006, corresponding to Spring 2006 in Figure S2B; values ranged from 11.3 to 27.7 $\mu g\ m^{-3}$); B) Mean across 2002-2006 (values ranged from 25.4 to 41.9 $\mu g\ m^{-3}$); C) Highest week (beginning March 24, 2003, corresponding to Spring 2003 in Figure S2B; values ranged from 84.4 to 100.8 $\mu g\ m^{-3}$).

Supporting Information References

1. Wood, S.N. *Generalized Additive Models: An Introduction with R*; Chapman and Hall/CRC: Boca Raton, Florida, U.S.A., 2006.

2. *The impact of the congestion charging scheme on air quality in London*; Health Effects Institute: Boston, MA, U.S.A., 2011; http://pubs.healtheffects.org/view.php?id=358.

3. Tonne, C.; Beevers, S.; Armstrong, B.G.; Kelly, F.; Wilkinson, P. Air pollution and mortality benefits of the London Congestion Charge: spatial and socioeconomic inequalities. *Occup. Environ. Med.* **2008**, 65 (9), 620-627.

4. Hastie T.; Tibshirani R. *Generalized Additive Models*; Chapman and Hall: New York, U.S.A., 1990.

5. Yanosky J.D.; Paciorek C.J.; Suh H.H. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. *Environ. Health Persp.* **2009**, 117 (4), 522-529.

6. Ribeiro Jr. P.J.; Diggle P.J. geoR: a package for geostatistical analysis. *R-NEWS* **2001**, 1 (2), 15-18.