Supporting Information

Land use regression modelling to estimate historic (1962-1991) exposures to black smoke and sulphur dioxide for Great Britain

John Gulliver[1], Chloe Morris[1], Kayoung Lee[1], Danielle Vienneau[1], David Briggs[1], and Anna Hansell[1].

1. MRC-HPA Centre for Environment and Health, Department of Epidemiology & Biostatistics, Imperial College London, London, UK.

Contents

Tables:

S1. Mean and variability of monitored concentrations ($\mu g/m^3$) across Great Britain.

S2. Predictor variables offered in the development of LUR models.

S3. Development of LUR models for Black smoke by year

S4. Development of LUR models for SO_2 by year

S5. Correlation (Pearson's R) between monitoring sites common to each pair of years.

S6. Comparison between the range of R^2 from the 'leave-one-out' analysis and overall R^2 from model evaluation.

Figures:

S1. Model evaluation for black smoke: predicted versus observed concentrations ($\mu g/m^3$)

S2. Model evaluation for SO_2: predicted versus observed concentrations in ($\mu g/m^3$)

As can be seen in Figure 1, BS models for 1962 and 1971 show some heteroscedacticity and the overall model under-prediction can be seen to be related to higher concentrations. For 1981 and 1991 the heteroscedacticity is less obvious due to the large under-prediction for about 10% of the evaluation sites. The scatter-plot for the 1962 SO_2 model (Figure 2) is homoscedastic, with only a slight under-prediction at higher concentrations. For SO_2 in 1971 and 1981 there is slight heteroscadacticity but overall only slight under-prediction, and in 1981 the dominance of poor predictions for two monitoring sites, as mentioned above, is evident. The weaker result for the 1991 SO_2 model is exemplified by the large amount of scatter between the monitoring sites.

S3. Modelled black smoke concentrations for 1962, 1971, 1981, and 1991

S4. Modelled sulphur dioxide concentrations for 1962, 1971, 1981, and 1991

Table S1. Mean and variability of monitored concentrations ($\mu g/m^3$) across Great Britain

Pollutant	Year	Min	5%	Mean	95%	Max	S.D.	Skewness	N
BS	1962	2.8	47.0	143.8	314.1	575.0	84.6	1.2	667
	1971	1.2	14.3	52.2	114.7	280.5	33.3	1.5	958
	1981	0.9	7.0	19.5	41.3	105.5	11.6	2.3	964
	1991	2.5	5.2	15.9	30.2	48.2	7.8	1.0	192
SO2	1962	3.7	49.8	160.5	301.9	449.7	77.8	0.7	603
	1971	10.2	31.7	95.3	179.9	288.8	47.0	0.8	916
	1981	2.5	17.5	47.5	84.5	190.9	22.0	1.2	945
	1991	11.9	15.6	34.9	62.4	90.2	14.5	0.8	191

Annual averages were computed from daily averages for monitoring stations with at least 75% data capture across the year

Table S2. Predictor variables offered in the development of LUR models.

Variables		Direction of Effect	Point	Buffer (km)							
Type	Description			0.5	1	2	3	4	5	7.5	10
Roads	Length of motorways (m)	+		X	X	X	X				
	Length of A-roads (m)	+		X	X	X	X				
	Length of B-roads	+		X	X	X	X				
	Length of minor roads	+		X	X	X	X				
	Major roads: length of motorways + A-roads[a]	+		X	X	X	X				
	Length of B-roads + minor roads[a]	+		X	X	X	X				
Land cover	High density residential (m^2)	+		X	X	X	X	X	X	X	X
	Low density residential (m^2)	+		X	X	X	X	X	X	X	X
	Industry and commercial land (m^2)	+		X	X	X	X	X	X	X	X
	Ports (m^2)	+		X	X	X	X	X	X	X	X
	Other urban: land associated with road and rail networks, and mine, dump and construction sites (m^2)	+		X	X	X	X	X	X	X	X
	Urban green space: Green urban areas + sport and leisure facilities (m^2)[a]	-		X	X	X	X	X	X	X	X
	Forest (m^2)	-		X	X	X	X	X	X	X	X
	Agriculture (m^2)	-		X	X	X	X	X	X	X	X
	High density urban: *High density residential + industry*[a]	+		X	X	X	X	X	X	X	X
	Low density urban: *low density residential + green urban areas + sport and leisure facilities + mine, dump and construction sites + airports + land associated with road and rail networks*[a]	+		X	X	X	X	X	X	X	X
	Non-residential, low density urban: *green urban areas + sport and leisure facilities + mine, dump and construction sites + airports + land associated with road and rail networks*[a]	+		X	X	X	X	X	X	X	X
	Other natural areas: *scrub + open spaces + wetlands + waterways + coastal marshes and flats*[a]	-		X	X	X	X	X	X	X	X
	Other natural areas + agriculture[a]	-		X	X	X	X	X	X	X	X
	Other natural areas + forest[a]			X	X	X	X	X	X	X	X
Population	Estimated counts from areal weighting			X	X	X	X	X	X	X	X
Trend	Co-ordinates (x,y, x^2, y^2, xy)	+/-	X								
Altitude	Height above sea level (m)	-	X								
	Square of height above sea level (m)	-	X								

a. combined land cover or road variables.

Table S3. Development of LUR models for Black smoke by year

Year (number of sites)	Variable type	Variable description	β	P (sig)	Adj R^2	SEE
1962 (534)		(constant)	.970	.087		
	Trend surface	X	7.985e-006	.001		
		Y	9.890e-006	.000		
		X^2	-8.5610e-012	.001		
		Y^2	-8.903e-012	.000		
		XY	-2.478e-012	.080	.501	.440
	Land cover	Other natural areas and agriculture within 1km	-1.415e-007	.000	.609	.389
		Forest within 3 km	-6.391e-008	.000	.636	.376
	Roads	Minor road and 'B' Road length within 3 km	2.239e-006	.000	**.676**	**.354**
1971 (767)		(constant)	-1.965	.000		
	Trend surface	X	1.678e-005	.000		
		Y	7.835e-006	.000		
		X^2	-1.776e-011	.000		
		Y^2	-6.065e-012	.000		
		XY	-3.397e-012	.002	.503	.483
	Roads	Minor road and 'B' Road length within 1 km	1.749e-005	.000	.656	.402
	Land cover	Low density residential within 10 km	1.712e-009	.000	.670	.394
	Roads	Major road length within 1km	3.153e-005	.000	**.680**	**.388**
1981 (771)		(constant)	-1.040	.012		
	Trend surface	X	1.1662e-005	.000		
		Y	5.752e-006	.000		
		X^2	-1.134e-011	.000		
		Y^2	-3.402e-012	.000		
		XY	-5.757e-012	.000	.270	.479
	Land cover	Forest and other natural areas within 1km	-2.749e-007	.000	.330	.458
		Other urban areas within 10km	5.227e-008	.000	.371	.444
	Roads	Minor Road and 'B' Road length within 1 km	1.129e-005	.000	**.408**	**.431**
1991 (155)		(constant)	-5.765	.006		
	Trend surface	X	2.409e-005	.003		
		Y	1.748e-005	.000		
		X^2	-2.001e-011	.011		
		Y^2	-1.041e-011	.000		
		XY	-1.972e-011	.000	.345	.450
	Land Cover	Forest and other natural areas within 10 km	-2.287e-009	.016	.372	.440
		Other urban areas within 7.5 km	4.3238e-008	.022	**.390**	**.434**

N.B. Trend surface variables were entered in all cases as a group (e.g. X,Y,X^2, Y^2, XY)

Table S4. Development of LUR models for SO2 by year

Year (number of sites)	Variable type	Variable description	β	P (sig)	Adj R^2	SEE
1962 (482)		(constant)	-0.048	.937		
	Trend surface	X	9.713e-006	.000		
		Y	1.238e-005	.000		
		X2	-7.295e-012	.009		
		Y2	-1.055e-011	.000		
		XY	-9.170e-012	.000	.279	.473
	Roads	Minor road length within 3 km	1.943e-006	.000	.552	.373
	Land cover	Low density urban within 10km	2.564e-009	.000	.573	.364
		High density urban within 1km	1.535e-007	.000	**.605**	**.350**
1971 (733)		(constant)	0.290	.306		
	Roads	Minor road length within 3km	1.937e-006	.000	.390	.444
	Trend surface	X	1.156e-005	.000		
		Y	4.772e-006	.000		
		X2	-1.092e-011	.000		
		Y2	-3.767e-012	.000		
		XY	-3.543e-012	.001	.594	.362
	Land cover	Industrial and commercial land within 10km [compound ring variable: (<=3km * 0.84) + (> 3km & <=10km * 0.16]	4.107e-008	.000	.620	.351
	Roads	Major road length within 3km	7.108e-006	.000	.633	.344
	Land cover	Low density urban within 10km	1.839e-009	.000	**.649**	**.337**
1981 (756)		(constant)	0.975	.006		
	Trend surface	X	8.284e-006	.000		
		Y	3.701e-006	.000		
		X2	-7.643e-012	.000		
		Y2	-2.984e-012	.000		
		XY	-3.406e-012	.002	.244	.446
	Land cover	Low density urban within 10 km	2.095e-009	.000	.330	.420
		Other urban within 10km	3.637e-008	.000	.352	.413
		High density urban within 2 km	3.946e-008	.000	**.378**	**.404**
1991 (153)		(Constant)	2.495	.000		
	Trend surface	Y	4.050e-006	.000		
		Y2	-4.619e-012	.000	.168	.373
	Land cover	Other urban within 7.5km	4.569e-008	.001	.218	.362
		Low density urban within 10 km	1.387e-009	.016	**.243**	**.356**

N.B. Trend surface variables were entered in all cases as a group (e.g. X,Y,X^2, Y^2, XY)

Table S5. Correlation (Pearson's R) between monitoring sites common to each pair of years.

Pollutant	Year	R	p	Number of sites
Black smoke	62-91	0.53	.005	27
	71-91	0.60	.000	52
	81-91	0.68	.000	26
	62-71	0.76	.000	335
	62-81	0.41	.000	174
	71-81	0.56	.000	149
Sulphur dioxide	62-91	0.26	.201	25
	71-91	0.61	.000	51
	81-91	0.70	.000	106
	62-71	0.83	.000	305
	62-81	0.31	.000	155
	71-81	0.58	.000	380

Table S6. Comparison between the range of R^2 from the 'leave-one-out' analysis and overall R^2 from model evaluation.

Pollutant	Year	Model performance against 20% retained sites (as in Table 2, main paper) Adj R^2	Model performance using 'leave-one-out' analysis Range of R^2
Black smoke	1962	.56	0.54-0.58
	1971	.41	0.40-0.45
	1981	.38	0.36-0.40
	1991	.34	0.30-0.37
Sulphur dioxide	1962	.71	0.69-0.74
	1971	.57	0.55-0.60
	1981	.26	0.25-0.31
	1991	.31	0.28-0.37

Figure S1. Model evaluation for black smoke: predicted versus observed concentrations (μg/m^3) by year (N.B. scales vary by plots)
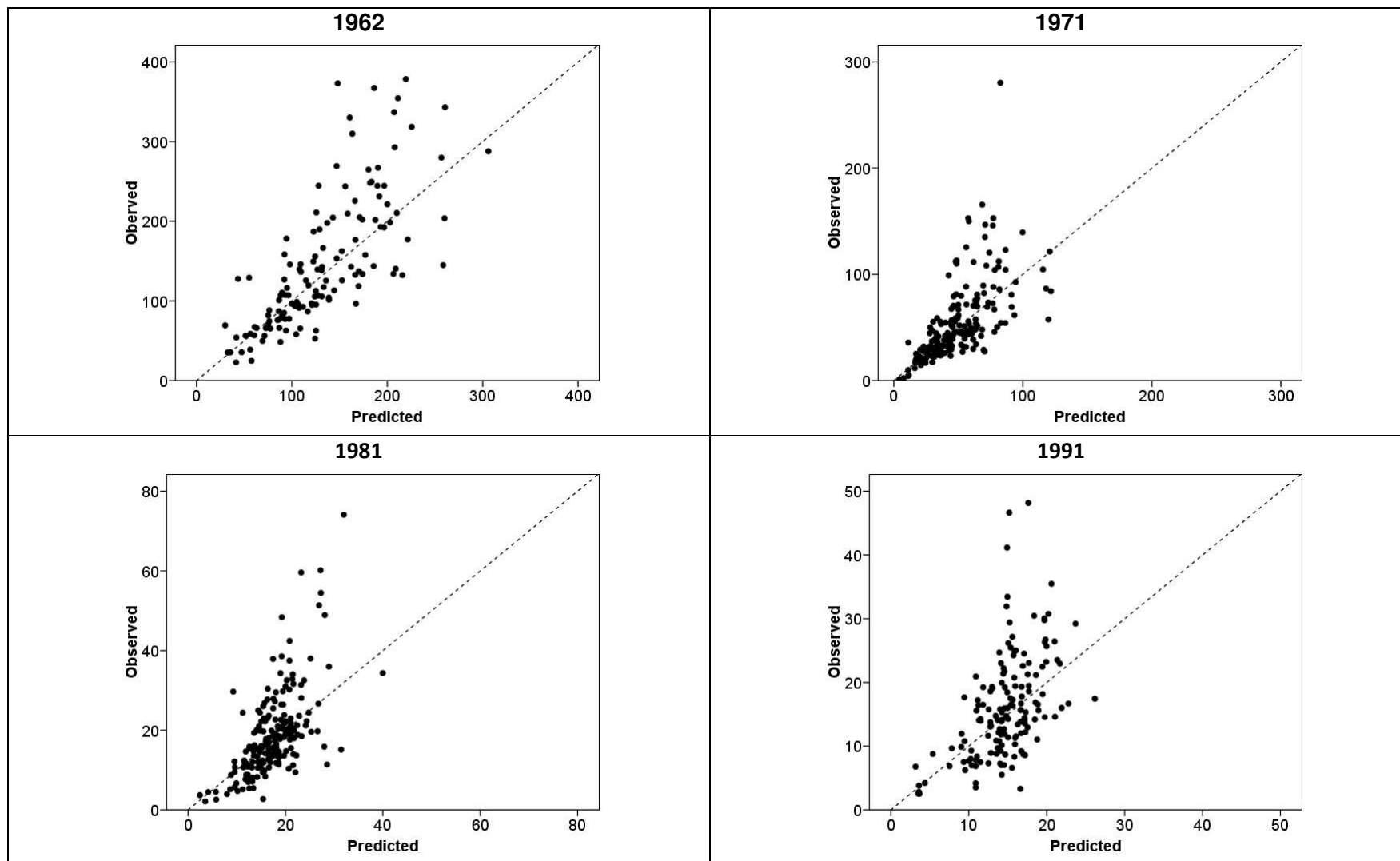
Figure S2. Model evaluation for SO₂: predicted versus observed concentrations in (μg/m^3) by year (N.B. scales vary by plots)

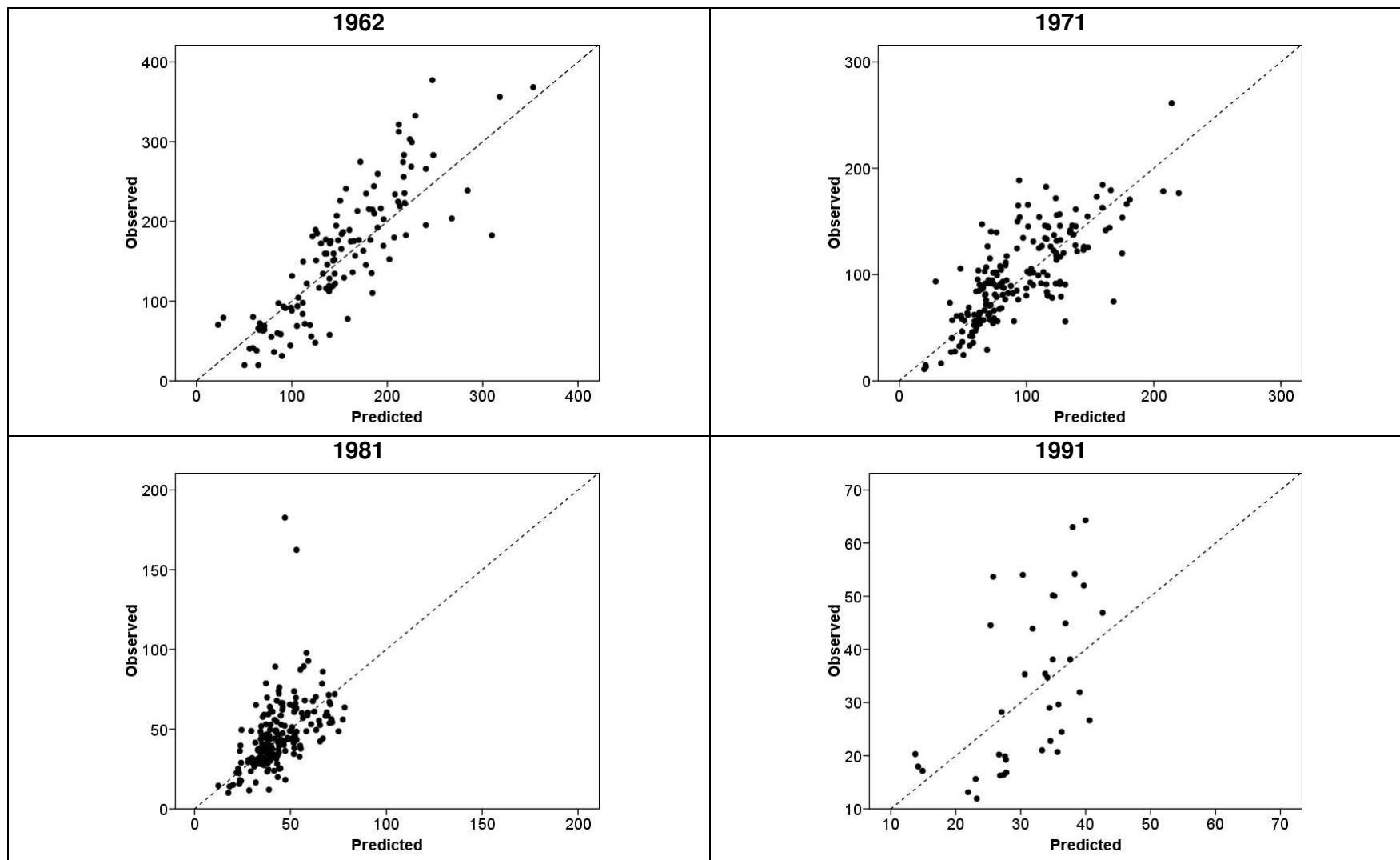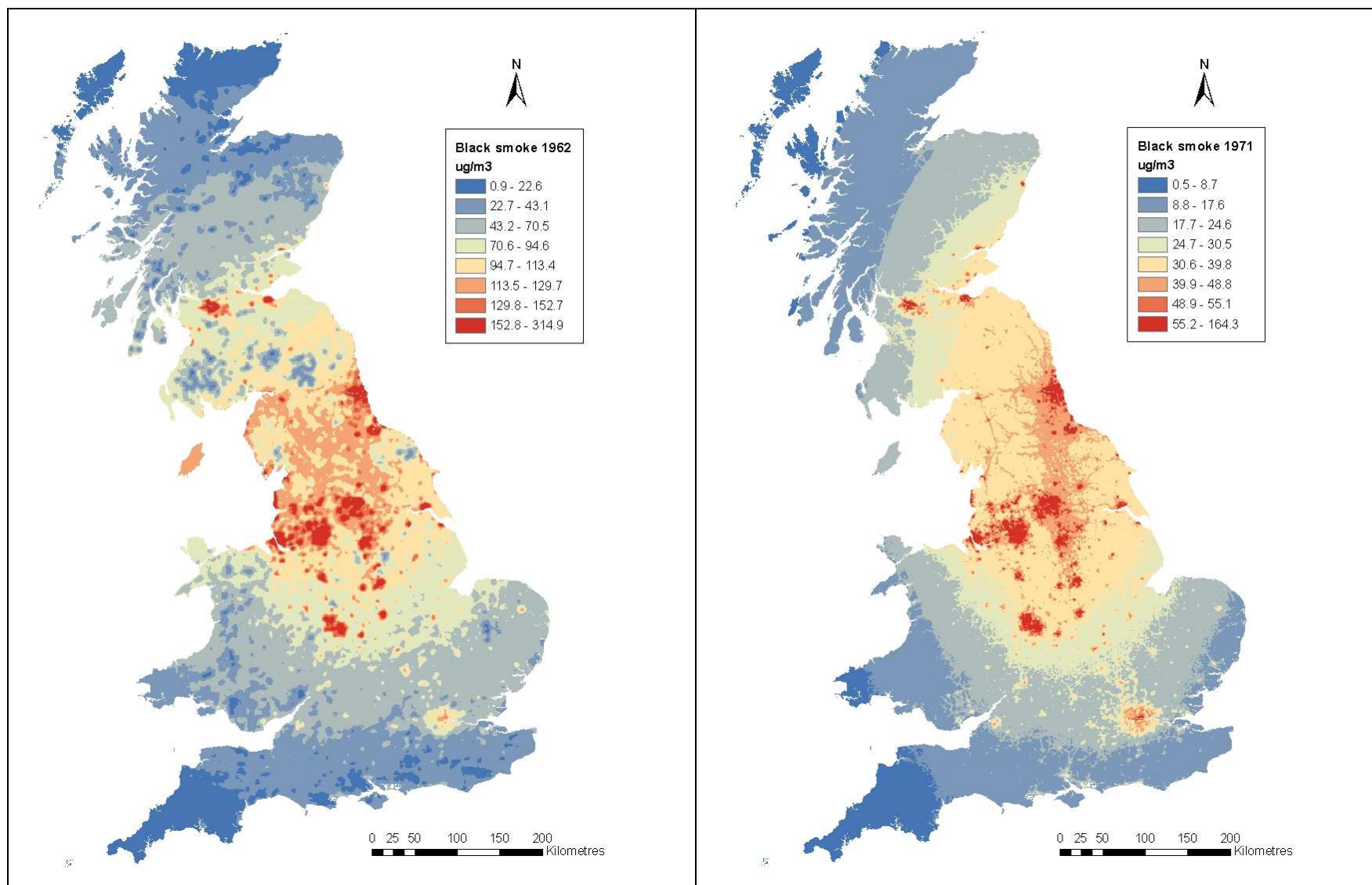**Figure S3. Modelled black smoke concentrations for 1962, 1971, 1981, and 1991**

Black smoke 1981
ug/m3
- 0.8 - 4.8
- 4.8 - 7.8
- 7.9 - 10.5
- 10.6 - 12.6
- 12.7 - 14.7
- 14.8 - 17.0
- 17.1 - 21.4
- 21.5 - 40.0

Black smoke 1991
ug/m3
- 0.0 - 3.3
- 3.4 - 6.7
- 6.8 - 9.5
- 9.6 - 11.8
- 11.9 - 14.1
- 14.2 - 16.6
- 16.7 - 20.8
- 20.9 - 27.9
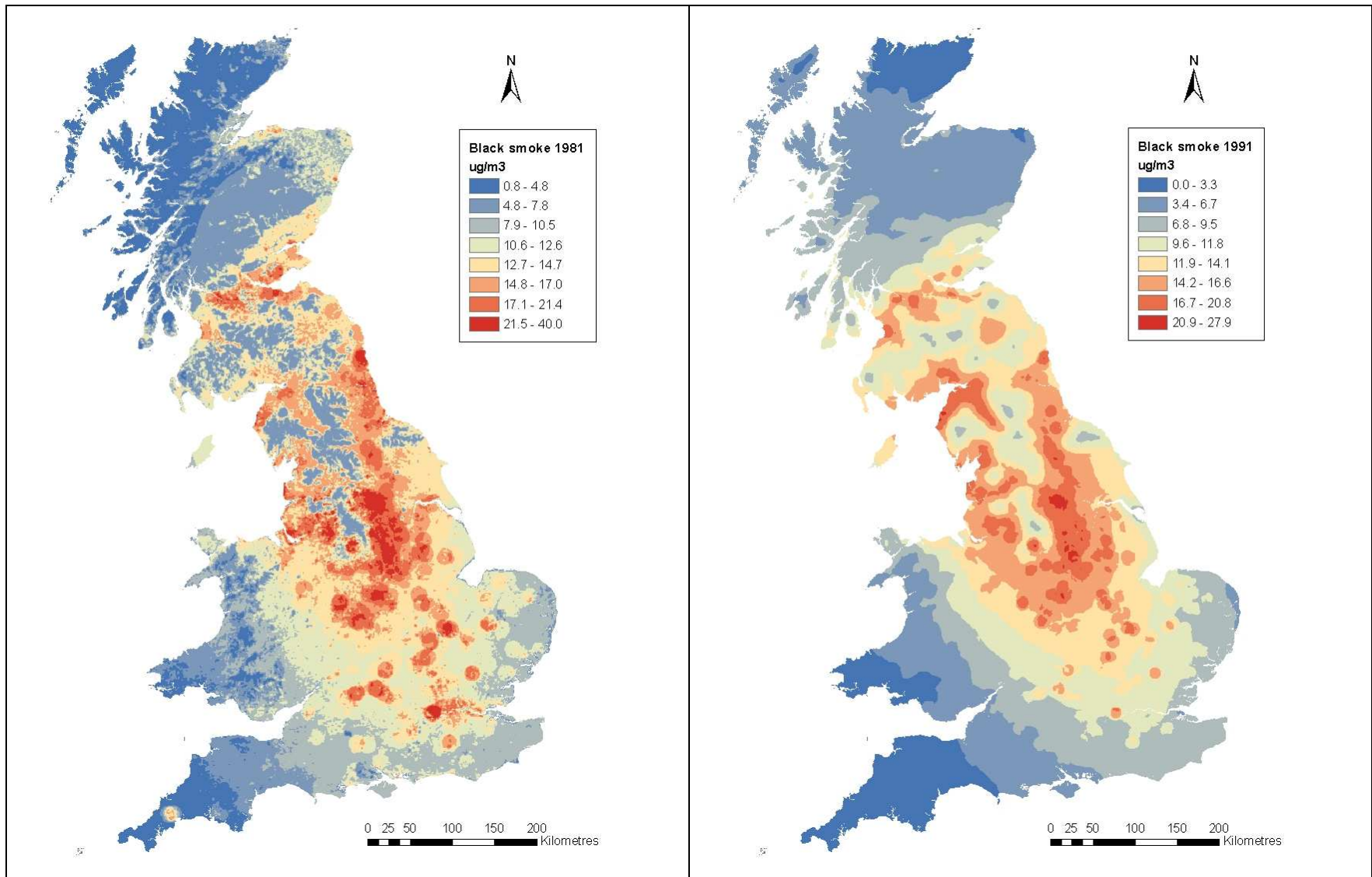
0 25 50 100 150 200 Kilometres

S10

Figure S4. Modelled sulphur dioxide concentrations for 1962, 1971, 1981, and 1991

Sulphur dioxide 1981
ug/m3
- 3.7 - 10.5
- 10.6 - 17.3
- 17.4 - 24.7
- 24.8 - 29.5
- 29.6 - 35.3
- 35.4 - 45.2
- 45.3 - 55.1
- 55.2 - 99.3

Sulphur dioxide 1991
ug/m3
- 1.7 - 9.6
- 9.7 - 14.7
- 14.8 - 19.8
- 19.9 - 24.7
- 24.8 - 29.5
- 29.6 - 32.6
- 32.7 - 39.8
- 39.9 - 64.2

N

0 25 50 100 150 200
Kilometres