SUPPORTING INFORMATION

Satellite-based land-use regression for continental-scale long-term ambient $PM_{2.5}$ exposure assessment in Australia

Luke D. Knibbs[1,2]*, Aaron van Donkelaar[3], Randall V. Martin[3,4], Matthew J. Bechle[5], Michael Brauer[6], David D. Cohen[7], Christine T. Cowie[2,8], Mila Dirgawati[9,10], Yuming Guo[2,11], Ivan C. Hanigan[2,12], Fay H. Johnston[2,13], Guy B. Marks[2,8], Julian D. Marshall[5], Gavin Pereira[14,15], Bin Jalaludin[2,16], Jane S. Heyworth[2,9,17], Geoffrey G. Morgan[2,12], Adrian G. Barnett[18]

1. Faculty of Medicine, School of Public Health, The University of Queensland, Herston, QLD 4006, Australia
2. Centre for Air Pollution, Energy and Health Research, Glebe, NSW 2037, Australia
3. Department of Physics and Atmospheric Science, Dalhousie University, Halifax, NS, B3H 4R2, Canada
4. Smithsonian Astrophysical Observatory, Harvard-Smithsonian Center for Astrophysics, Cambridge, MA, 02138, United States
5. Department of Civil and Environmental Engineering, University of Washington, Seattle, WA, 98195, United States
6. School of Population and Public Health, The University of British Columbia, Vancouver, BC, V6T 1Z3, Canada
7. Centre for Accelerator Science, Australian Nuclear Science and Technology Organisation, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia

8. South Western Sydney Clinical School, The University of New South Wales, Liverpool, NSW 2170, Australia
9. School of Population and Global Health, The University of Western Australia, Perth, WA 6009, Australia
10. Environmental Engineering, Institut Teknologi Nasional, Bandung, Jawa Barat 40213, Indonesia
11. Department of Epidemiology and Biostatistics, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC 3004, Australia
12. School of Public Health, The University of Sydney, Sydney, NSW 2006, Australia
13. Menzies Institute for Medical Research, University of Tasmania, Hobart, TAS 7000, Australia
14. School of Public Health, Curtin University, Bentley, WA 6102, Australia
15. Telethon Kids Institute, The University of Western Australia, Perth, WA 6008, Australia
16. Population Health, South Western Sydney Local Health District, Liverpool, NSW 2170, Australia
17. Clean Air and Urban Landscapes Hub, National Environmental Science Programme, Melbourne, VIC 3010, Australia
18. School of Public Health and Social Work, Queensland University of Technology, Kelvin Grove, QLD 4059, Australia

* Corresponding author (e: l.knibbs@uq.edu.au; ph: +61 7 3365 5409)

Number of pages: 46

Number of figures: 16

Number of tables: 12

METHODS

Measurement methods and correlation

The large majority (~85%) of measurements were performed using tapered element oscillating microbalance instruments (TEOM). If a site used both a TEOM and another method, we therefore used the TEOM data for consistency with the majority of sites.

We first investigated the relationship between TEOM, BAM, and offline gravimetric measurements using long-term co-located measurements performed by regulatory agencies. We used national campaign, where co-located Partisol and TEOM measurements were performed over 24 h (on a one-day-in-three basis), to investigate the validity of TEOMs compared with an offline gravimetric reference method in different Australian urban locations. TEOMs (and BAMs) are a US-EPA designated automatic (online) equivalence method for measuring ambient $PM_{2.5}$, while the Partisol is designated as a manual (offline) reference, or equivalence, method.[1] TEOM measurements in this campaign did not have an internal correction factor applied (NEPC, 2003). Partisol measurements were performed using manufacturer-specified equilibration conditions (20–23°C, 30–40% RH) before and after sampling.

We regressed daily Partisol on TEOM measurements collected at the three sites with the longest data records (Footscray, Victoria; Alphington, Victoria; Chullora, New South Wales), where each site had been monitored using both methods over six- to ten-years. The two methods showed very good correlation, with respective R^2 values of 0.85, 0.76, and 0.80 at each site, and regression coefficients (SE) of 1.15 (0.01), 1.05 (0.02), and 0.95 (0.02) μg/m^3. We assessed an additional three measurement sites with shorter data records (Brooklyn, Victoria; Rocklea, Queensland; Springwood, Queensland), for a total of six sites nationally

(~4,000 site-days of $PM_{2.5}$ measurement). Using all six sites, we observed good correlation (R^2 = 0.69), with a regression coefficient of 0.97 (0.01) μg/m^3.

Some sites (n=10) in New South Wales and Victoria switched from TEOMs to BAMs during the latter part of the study period (between 2012 and 2015). Three of the BAM sites (Footscray, Alphington, Chullora) were co-located with Partisol or other offline gravimetric sampler (for between 0.7 and 2 years at each site), and the two methods demonstrated very good correlation: R^2 = 0.89 (coefficient = 0.90 [0.03] μg/m^3), R^2 = 0.90 (coefficient = 0.87 [0.02] μg/m^3), and R^2 = 0.87 (coefficient = 0.97 [0.02] μg/m^3).

Based on the analyses above, we did not apply a correction factor to TEOM measurements because they were well-correlated with gravimetric Partisol measurements and our primary aim was to assess spatial contrasts in $PM_{2.5}$. For the same reason, we did not correct BAM measurements either.

Monitoring site classification

The 70 $PM_{2.5}$ monitoring sites were from all six states and two territories in Australia: three in the Australian Capital Territory, twenty in New South Wales, three in the Northern Territory, nineteen in Queensland, three in South Australia, four in Tasmania, twelve in Victoria, and six in Western Australia.

We classified the sites based on whether they were located in a significant urban area (i.e., population >10,000) using standard Australian Bureau of Statistics (ABS) criteria.[2] Sites that were not located in urban areas were classified in a combined category of rural and remote (n = 5/70). Because they were used for regulatory air monitoring, the urban and rural sites were largely in background locations free from nearby emissions. For example, the median distance to the nearest major road and industrial $PM_{2.5}$ point source was 290 m and 1.6 km, respectively.

We considered, but did not pursue, imputation of $PM_{2.5}$ at PM_{10}-only monitoring sites because it could introduce uncertainty due to variability in Australia's mixture of natural and anthropogenic PM sources, and because the yield of additional sites in both number and spatial distribution was relatively low (n ~26).[3]

LUR variables

Table S1 shows the LUR predictors we selected, their source, spatial resolution and any additional processing undertaken. This selection of variables was informed by Australian $PM_{2.5}$ source apportionment studies using positive matrix factorization (PMF; see references in following paragraph).

For example, applying PMF to fifteen years' data collected by the Australian Nuclear Science and Technology Organisation (ANSTO) at four sites in Sydney (population ~4.8 million) with varying source proximity showed that, on average, sources comprised: wind-blown soil (3-4% of $PM_{2.5}$ mass), sea salt (4-17%), mixed industrial and aged sulfate (9-23%, including industry and heavy oil combustion), secondary sulfate (20-30%, including coal-fired power generation and industry), mixed biomass and diesel combustion (12-32%, including landscape fires and wood heaters), and motor vehicles (20-26%).[4] During the winter months, the contribution to $PM_{2.5}$ mass from wood heater smoke can peak at 60-80%; up from 10% during summer.[4] Comparable results have been reported for other large Australian population centres as well as semi-rural locations, although some specific local sources vary (e.g.,[5-8]). Our selection of LUR predictors sought to capture these sources, which are relevant to $PM_{2.5}$ exposures within the Australian population.

Table S1. LUR predictor variables.

Variable (units)	**Approx. spatial resolution**	**Point or buffer^**	**Source** (all weblinks active at 05/04/2018)
Annual mean $PM_{2.5}$ (each year during 2000–2015) gridded at 0.1° (μg/m^3)*****	10 km	point	Global satellite-derived surface estimates ('SAT-$PM_{2.5}$') http://fizz.phys.dal.ca/~atmos/martin/?page_id=140 van Donkelaar et al. (2016)[9]
As above, but adjusted using geographically-weighted regression and gridded at 0.01° (μg/m^3)*****	1 km	point	As above
As above, but gridded at 0.1° (μg/m^3)*****	10 km	point	As above
Elevation (m)	30 m	point	Satellite-derived (SRTM) Geoscience Australia 1-second smoothed digital elevation model http://www.ga.gov.au/metadata-gateway/metadata/record/gcat_72759
Distance to coast (km)	n/a	point	Derived using 'Near' command in ArcGIS (excludes inland lakes)
Annual mean rainfall (mm), 1960-1991	2.5 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/rainfall/index.jsp
Annual mean daily average temperature (°C), 1960-1991	2.5 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/temperature/index.jsp
Annual total heating degree days (count, 18°C reference temperature), 1961-1990	10 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/degree-days/index.jsp?maptype=1&period=an&product=hdd18#maps
Annual total cooling degree days (count, 18°C reference temperature), 1961-1990	10 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/degree-days/index.jsp?maptype=3&period=an&product=cdd18#maps

Annual mean daily 3pm relative humidity (%), 1976-2005	10 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/relative-humidity/index.jsp?maptype=3&period=an#maps
Annual mean daily solar exposure (MJ/m^2), 1990-2011	5 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/solar-exposure/index.jsp
Annual mean wind speed (km/h), 2004-2008	12.5 km	point	Australian Bureau of Meteorology http://www.bom.gov.au/jsp/ncc/climate_averages/wind-velocity/index.jsp
Tree cover (%)*	250 m	buffer[a]	Satellite-derived vegetation continuous fields product http://www.landcover.org/data/vcf/ DiMiceli et al. (2011)[10]
Water cover (%)	250 m	buffer[a]	As above
Impervious surfaces (%)	1 km	buffer[a]	Satellite-derived NOAA constructed impervious surface area product 2000-2001 http://ngdc.noaa.gov/eog/dmsp/download_global_isa.html Elvidge et al. (2007)[11]
Land-use type (%)[c]	mesh block[d]	buffer[a]	Australian Bureau of Statistics http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Geography
Annual total burned vegetation area (%)*****	500 m	buffer[a]	Satellite-derived estimates of burned area (where QA values $\leq$3) http://modis-fire.umd.edu/pages/BurnedArea.php (Borschetti et al., 2013)[12]
Annual active fire density (fires/km^2)*****	1 km	buffer[a]	Satellite-derived estimates of active fires (where detection confidence $\geq$30%) http://modis-fire.umd.edu/pages/ActiveFire.php Giglio (2015)[13]
Population density (persons/km^2)	1 km	buffer[a]	Australian Bureau of Statistics (2011 census) http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.007
Major roads (km)[e,f]	n/a	buffer[b]	PSMA Australia Transport and Topography product https://www.psma.com.au/products/transport-topography
Minor roads (km)[e,g]	n/a	buffer[b]	As above
Total roads (km) (= major roads + minor	n/a	buffer[b]	As above

roads [excludes unsealed roads])			
Unsealed roads (km)[e,h]	n/a	buffer[b]	As above
Distance to nearest major road (m)	n/a	point	As above (derived using 'Near' command in ArcGIS)
Distance to nearest road (m)	n/a	point	As above
$PM_{2.5}$ point source emission density (kg/km^2)[i] *	n/a	buffer[a]	Australia National Pollutant Inventory http://www.npi.gov.au/
$PM_{2.5}$ point source site density ($sites/km^2$)[i] *	n/a	buffer[a]	As above
Distance to nearest $PM_{2.5}$ point source site (m)[i]	n/a	point	As above (derived using 'Near' command in ArcGIS)
Distance to nearest operational coal-fired power plant (km)	n/a	point	Geoscience Australia (derived using 'Near' command in ArcGIS) https://data.gov.au/dataset/power-stations
Distance to nearest operational natural gas-fired power plant (km)	n/a	point	As above
Households using wood as main energy source for heating (%)	Two values per state (capital city and the rest of state)	point	Australian Bureau of Statistics (Energy Use and Conservation Survey, 2011) http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4602.0.55.001Mar%202011?OpenDocument

Footnotes

[a] average of variable within buffer
[b] sum of variable within buffer

[c] four land use categories were examined; residential, commercial, industrial, and open space (which was the sum of water, parks and agricultural land).

[d] a mesh block is the smallest spatial unit used in the Australian census and their size varies (median = 0.04 km^2); on average they contain 62 people.

[e] positional accuracy for roads ±2 m in urban areas, ±10 m in rural and remote areas. Attribute accuracy is 99.09% for key attributes (name and unique identifier).[14]

[f] major roads were defined as national/state highways, arterial roads (which are major connector roads for national and state highways) and sub-arterial roads (which are connectors between highways and/or arterial roads, or serve as an alternative for arterial roads).[14]

[g] minor roads were defined as collector roads (which are connectors between sub-arterial roads, and distribute traffic to local roads) and local roads (which provide property access).[14]

[h] unsealed roads were defined as those with a surface other than brick, concrete or tar.[14]

[i] total (fugitive + non-fugitive) estimated $PM_{2.5}$ emissions from all (n=2,003 in 2015) point sources around Australia (industry, commercial).

^ 22 circular buffers were created for each variable with radii of 100 m, 200 m, 300 m, 400 m, 500 m, 600 m, 700 m, 800 m, 1 km, 1.2 km, 1.5 km, 1.8 km, 2 km, 2.5 km, 3 km, 3.5 km, 4 km, 5 km, 6 km, 7 km, 8 km, and 10 km. An additional five buffers were determined: 25 km, 50 km, 100 km, 250 km, 500 km. These were used (along with the 10 km buffer) to calculate the burned area and active fire density variables only. Unsealed roads were calculated at the 22 smaller buffers as well as 25 km and 50km.

***** indicates an LUR predictor that varied annually over the study period (2000–2015).

Independent evaluation data

We used similar inclusion criteria for independent evaluation data as for model development sites, although numerical annual averages were acceptable in lieu of daily measurements, provided missing data were reported. We also included sites using one-day-in-three offline gravimetric methods if they operated for at least a full year. For example, we used long-term (2000–2015) measurements by the Australian Nuclear Science and Technology Organisation (ANSTO) that agree well with continuous methods and adequately capture temporal variability within a year.[4,6-7]

We identified 30 additional sites that met the inclusion criteria, for a total of 51 independent evaluation sites (Figure S1). New South Wales (n=18) and Queensland (n=23), the most- and third-most populous states, respectively, accounted for 41 sites. The remaining 10 sites were in the Australian Capital Territory (n=1), the Northern Territory (n=1), Victoria (n=5), South Australia (n=1), and Western Australia (n=2). All sites were located in an ABS-defined significant urban area (population >10,000).

LUR models were used to predict annual average $PM_{2.5}$ for the period that a site had valid data by matching any annual time-varying predictors (e.g., satellite-based $PM_{2.5}$) to the same period. Any models without time-varying predictors were applied without adjustment (i.e., year-2015 predictions). We assessed the global satellite-based $PM_{2.5}$ estimates (i.e., SAT-$PM_{2.5}$) as well as the two geographically-weighted regression (GWR)-adjusted estimates (gridded at 0.1° and 0.01°), because only 2/51 $PM_{2.5}$ measurement sites were used to inform the GWR adjustment.

Twelve sites had data for 2016 only; given the minimal year-to-year variability in annual $PM_{2.5}$ we included these sites but checked them for undue influence on the evaluation. Most sites (n=34) had data for a single year during 2000–2016 (Table S2, S3). Of the 17 sites with multiple years of measurement, 12 had data back to 2000 and one site had data for 2000 only; a total of 13 sites that year (Table S3).

Table S2. Number of evaluation sites per calendar year, and number of years of measurement per site. *see notes above.

Year	n sites
2000	13
2001	13
2002	13
2003	13
2004	10
2005	9
2006	8
2007	7
2008	6
2009	8
2010	7
2011	9
2012	8
2013	6
2014	6
2015	17
2016*	12
Total site-years	**165**

n years	n sites
1	34
2	3
3	1
4	2
5	3
6	2
7	1
16	5
Total site-years	**165**

Figure S1. Map showing 49 $PM_{2.5}$ sites used for LUR model development (i.e., sites that met the inclusion criteria in year-2015). A national population density grid ($1km^2$) is shown in the background.[15]

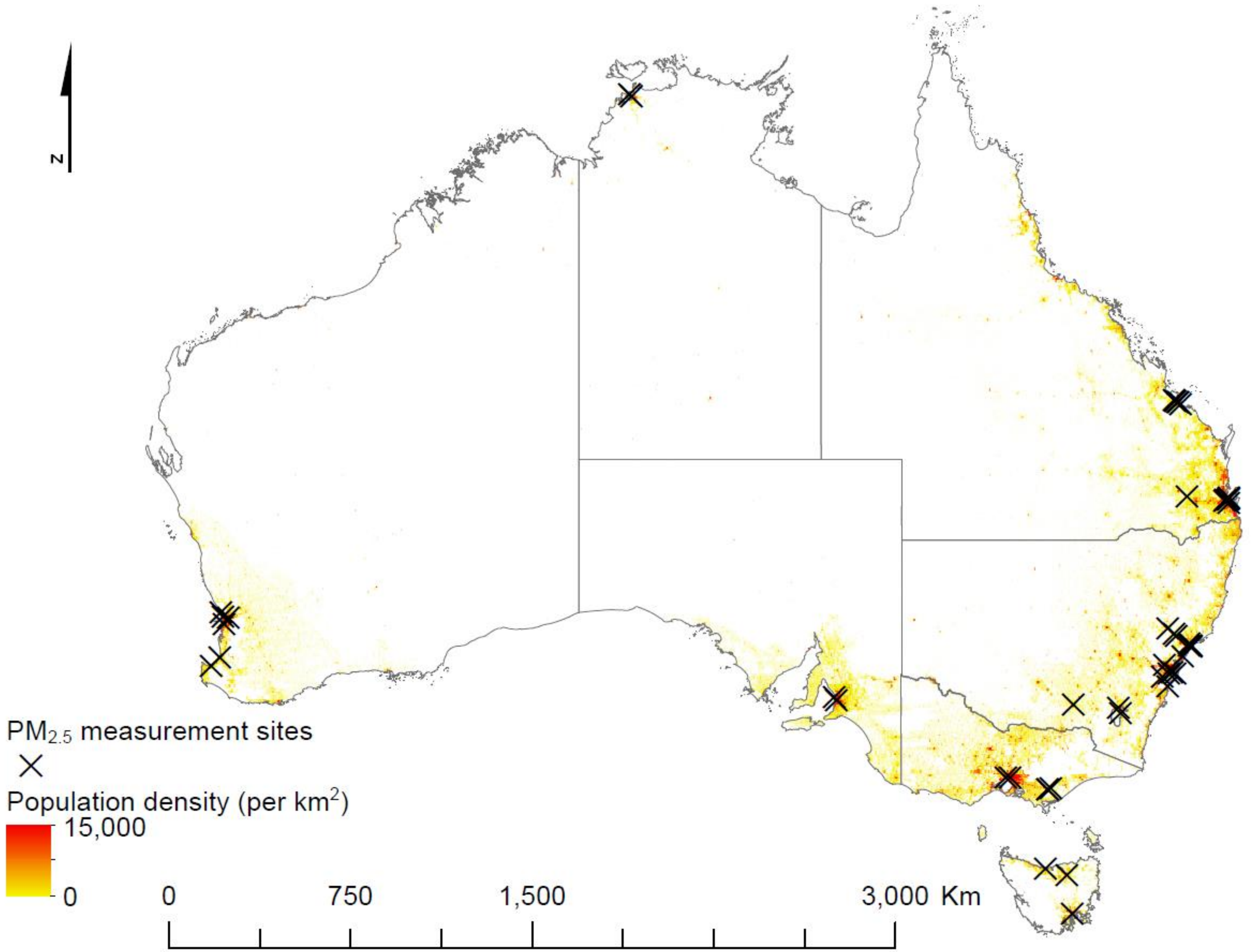
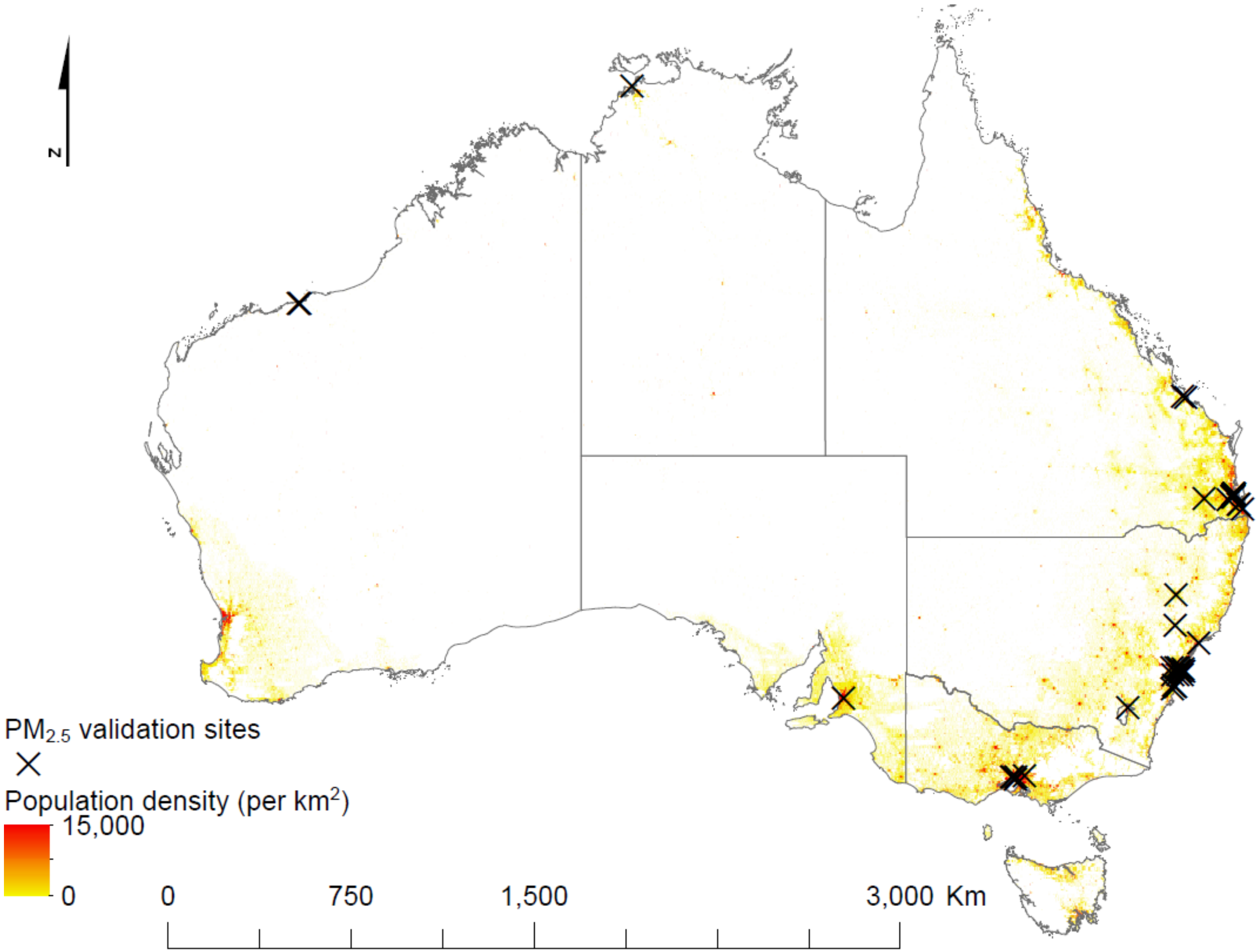


Figure S2. Map showing 51 $PM_{2.5}$ measurement sites used for independent evaluation. A national population density grid ($1km^2$) is shown in the background.[15]



RESULTS

Table S3. Descriptive statistics of annual mean $PM_{2.5}$ at model development and evaluation sites (units = µg/m^3).

Location	***n***	**Mean (SE)**	**SD**	**Min**	**5th**	**25th**	**50th**	**75th**	**95th**	**Max**
Development sites	49	6.9 (0.2)	1.6	3.4	3.9	5.8	7.4	8.1	8.8	9.1
Evaluation sites	51	7.1 (0.2)	1.6	4.3	4.3	5.8	7.2	8.0	9.8	11.0

Table S4. Tests of autocorrelation in LUR model residuals.

Model	**Moran's *I***	**p value**
SAT	0.19	0.14
SAT-W	0.32	0.10
NOSAT	0.04	0.64
NOSAT-W	0.16	0.20

Table S5. Bootstrap estimates of confidence intervals for LUR model R^2 values and coefficients (year-2015, n sites = 49).

Model	**Predictor (units)**	**β**	**β 95% CI (lower)**	**β 95% CI (upper)**
SAT	Intercept	5.619	3.758	6.548
R^2 (95% CI): 0.68 (0.43, 0.76)	Residential area, 5 km (%)	0.03	0.011	0.05
	Sat-$PM_{2.5}$ (μg/m^3)	0.905	0.639	1.862
	Annual average rainfall (mm)	-0.002	-0.003	-0.0004
	Commercial area, 1.5 km (%)	0.039	0.002	0.089
	Households using wood heaters (%)	0.072	0.03	0.111
	Tree cover, 5 km (%)	-0.073	-0.125	-0.021
	Annual average wind speed (km/h)	-0.165	-0.355	0.021
SAT-W	Intercept	5.644	3.905	6.604
R^2 (95% CI): 0.55 (0.31, 0.68)	Residential area, 5 km (%)	0.025	0.005	0.041
	Satellite $PM_{2.5}$ (μg/m^3)	0.826	0.535	1.589
	Annual average rainfall (mm)	-0.003	-0.004	-0.001
	Commercial area, 1.5 km (%)	0.075	0.032	0.145
NOSAT	Intercept	14.8	10.485	21.097
R^2 (95% CI): 0.63 (0.38, 0.75)	Residential area, 5 km (%)	0.022	0.002	0.041
	Annual average relative humidity (%)	-0.209	-0.306	-0.138
	Households using wood heaters (%)	0.126	0.079	0.173
	Major roads, 10 km (km)	0.005	0.003	0.008
	Burned area, 25 km (%)	0.105	0.034	0.206

NOSAT-W	Intercept	19.377	13.191	26.965
R^2 (95% CI): 0.60 (0.34, 0.73)	Residential area, 5 km (%)	0.023	0.008	0.035
	Annual average relative humidity (%)	-0.265	-0.37	-0.183
	Annual 18°C heating degree days (count)	0.002	0.001	0.002
	Elevation (m)	-0.007	-0.012	-0.005
	Commercial area, 1 km (%)	0.046	0.017	0.096

Table S6. Bootstrap estimates of confidence intervals for LUR model R^2 values and coefficients (year-2003; n sites = 18). Note: we randomly selected four years to assess sensitivity to year; year-2003 results are presented as an illustrative example.

Model	**Predictor (units)**	**β**	**β 95% CI (lower)**	**β 95% CI (upper)**
SAT	Intercept	9.341	-11.616	70.560
R^2 (95% CI): 0.74 (0.41, 0.76)	Residential area, 5 km (%)	0.043	-0.551	0.362
	Sat-$PM_{2.5}$ (µg/m^3)	0.125	-40.364	6.001
	Annual average rainfall (mm)	-0.001	-0.016	0.066
	Commercial area, 1.5 km (%)	0.063	-12.948	1.997
	Households using wood heaters (%)	0.024	-0.966	5.727
	Tree cover, 5 km (%)	-0.148	-1.808	2.091
	Annual average wind speed (km/h)	-0.296	-6.867	2.091
SAT-W	Intercept	7.843	0.412	22.842
R^2 (95% CI): 0.52 (0.02, 0.68)	Residential area, 5 km (%)	0.048	-0.023	0.106
	Satellite $PM_{2.5}$ (µg/m^3)	-0.044	-1.553	2.397
	Annual average rainfall (mm)	-0.002	-0.006	0.024
	Commercial area, 1.5 km (%)	0.019	-1.281	0.155
NOSAT	Intercept	12.824	-44.710	63.480
R^2 (95% CI): 0.67 (0.41, 0.76)	Residential area, 5 km (%)	0.087	-0.018	0.360
	Annual average relative humidity (%)	-0.114	-1.031	0.935
	Households using wood heaters (%)	-0.011	-0.508	1.499
	Major roads, 10 km (km)	-0.005	-0.017	0.017
	Burned area, 25 km (%)	-0.572	-13.764	36.1

NOSAT-W	Intercept	21.703	-35.660	90.140
R^2 (95% CI): 0.63 (0.27, 0.67)	Residential area, 5 km (%)	0.054	-0.062	0.167
	Annual average relative humidity (%)	-0.322	-2.797	0.223
	Annual 18°C heating degree days (count)	0.001	-0.006	0.017
	Elevation (m)	0.017	-1.049	0.169
	Commercial area, 1 km (%)	0.054	-1.340	0.154

Table S7. Percentiles of LUR model predictors at development, evaluation, and mesh block centroid sites. * indicates time-varying predictor (year-2015 values presented).

Model	**Predictor (units)**	**Min**	**5th**	**25th**	**50th**	**75th**	**95th**	**Max**
SAT: development sites	Residential area, 5 km (%)	0	0.6	18.2	37.4	48.7	64.6	66.0
n=49	Sat-$PM_{2.5}$ ($\mu g/m^3$)*	2.2	2.3	2.7	2.9	3.3	6.0	7.9
	Annual average rainfall (mm)	341	499	708	910	1139	1483	1707
	Commercial area, 1.5 km (%)	0	0	0.1	3.5	6.6	19.8	34.3
	Households using wood heaters (%)	2.3	2.4	4.0	8.5	21.3	32.6	33.2
	Tree cover, 5 km (%)	6.3	8.1	9.8	12.7	18.1	28.0	29.9
	Annual average wind speed (km/h)	1.8	2.1	4.6	6.3	7.6	8.9	11.0
SAT: evaluation sites	Residential area, 5 km (%)	0	4.2	29.6	50.4	64.0	70.2	83.8
n=51	Sat-$PM_{2.5}$ ($\mu g/m^3$)*	2.3	2.5	2.6	3.1	3.4	7.0	8.2
	Annual average rainfall (mm)	325	434	892	1151	1200	1641	1819
	Commercial area, 1.5 km (%)	0	0	1.3	5.1	12.9	36.0	44.5
	Households using wood heaters (%)	2.3	3.1	4.0	4.0	7.5	25.6	32.0
	Tree cover, 5 km (%)	5.3	6.7	8.3	12.4	16.2	27.3	40.5
	Annual average wind speed (km/h)	1.8	1.9	3.8	5.0	5.5	10.2	11.0
SAT: mesh block centroids	Residential area, 5 km (%)	0	0	15.8	40.4	57.8	70.7	100
n=344,954	Sat-$PM_{2.5}$ ($\mu g/m^3$)*	1.8	2.4	2.8	3.2	3.5	4.0	10.9
	Annual average rainfall (mm)	124	437	668	854	1180	1680	3661
	Commercial area, 1.5 km (%)	0	0	0	1.3	4.8	16.1	100
	Households using wood heaters (%)	2.3	3.5	4.0	5.0	21.3	32.0	33.2
	Tree cover, 5 km (%)	0	5.4	10.5	14.8	24.3	44.6	81.7

	Annual average wind speed (km/h)	0.8	2.7	4.5	5.5	7.0	9.9	29.4
SAT-W: development sites	Residential area, 5 km (%)	0	0.6	18.2	37.4	48.7	64.6	66.0
n=49	Satellite $PM_{2.5}$ ($\mu g/m^3$)*	2.2	2.3	2.7	2.9	3.3	6.0	7.9
	Annual average rainfall (mm)	341	499	708	910	1139	1483	1707
	Commercial area, 1.5 km (%)	0	0	0.1	3.5	6.6	19.8	34.3
SAT-W: evaluation sites	Residential area, 5 km (%)	0	4.2	29.6	50.4	64.0	70.2	83.8
n=51	Satellite $PM_{2.5}$ ($\mu g/m^3$)*	2.3	2.5	2.6	3.1	3.4	7.0	8.2
	Annual average rainfall (mm)	325	434	892	1151	1200	1641	1819
	Commercial area, 1.5 km (%)	0	0	1.3	5.1	12.9	36.0	44.5
SAT-W: mesh block centroids	Residential area, 5 km (%)	0	0	15.6	40.2	57.8	70.7	100
n=345,878	Satellite $PM_{2.5}$ ($\mu g/m^3$)*	1.8	2.4	2.8	3.2	3.5	4	10.9
	Annual average rainfall (mm)	124	437	669	856	1183	1686	3113
	Commercial area, 1.5 km (%)	0	0	0	1.3	4.7	16.1	100
NOSAT: development sites	Residential area, 5 km (%)	0	0.6	18.2	37.4	48.7	64.6	66.0
n=49	Annual average relative humidity (%)	44.8	46.9	52.6	56.6	57.8	62.8	66.9
	Households using wood heaters (%)	2.3	2.4	4.0	8.5	21.3	32.6	33.2
	Major roads, 10 km (km)	17.6	24.9	133.6	191.6	356.9	610.9	733.9
	Burned area, 25 km (%)*	0	0	0	0.01	0.5	10.6	17.8
NOSAT: evaluation sites	Residential area, 5 km (%)	0	4.2	29.6	50.4	64.0	70.2	83.8
n=51	Annual average relative humidity (%)	33.3	44.2	55.2	56.6	57.6	59.5	60.2
	Households using wood heaters (%)	2.3	3.1	4.0	4.0	7.5	25.6	32.0
	Major roads, 10 km (km)	20.9	47.3	149.7	393.8	465.9	630.7	772.2

	Burned area, 25 km (%)*	0	0	0	0	0.03	2.6	9.8
NOSAT: mesh block centroids *n*=347,453	Residential area, 5 km (%)	0	0	15.3	40.0	57.7	70.7	100
	Annual average relative humidity (%)	21.2	43.2	52.7	56.4	59.5	63.3	81.8
	Households using wood heaters (%)	2.3	3.5	4.0	7.5	21.3	32.0	33.2
	Major roads, 10 km (km)	0	20.8	82.0	186.1	384.2	666.6	889.6
	Burned area, 25 km (%)*	0	0	0	0	0.1	4.3	81.6
NOSAT-W: development sites *n*=49	Residential area, 5 km (%)	0	0.6	18.2	37.4	48.7	64.6	66.0
	Annual average relative humidity (%)	44.8	46.9	52.6	56.6	57.8	62.8	66.9
	Annual 18°C heating degree days (count)	1	75	466	930	1260	2305	2418
	Elevation (m)	1	2	7	18	39	473	587
	Commercial area, 1 km (%)	0	0	0	2.9	6.9	27.6	49.7
NOSAT-W: evaluation sites *n*=51	Residential area, 5 km (%)	0	4.2	29.6	50.4	64.0	70.2	83.8
	Annual average relative humidity (%)	33.3	44.2	55.2	56.6	57.6	59.5	60.2
	Annual 18°C heating degree days (count)	1	5	454	564	1210	1735	1793
	Elevation (m)	0	3	13	20	42	469	586
	Commercial area, 1 km (%)	0	0	0.3	3.8	13.2	36.7	68.5
NOSAT-W: mesh block centroids *n*=345,796	Residential area, 5 km (%)	0	0	15.7	40.3	57.8	70.7	100
	Annual average relative humidity (%)	21.2	43.2	52.7	56.4	59.5	63.2	74.8
	Annual 18°C heating degree days (count)	0	121	747	1124	1728	1824	2832
	Elevation (m)	0	3	16	41	112	568	1659
	Commercial area, 1 km (%)	0	0	0	0.3	4.6	19.0	100

Table S8. Bootstrap estimates of confidence intervals for LUR model independent evaluation R^2 values, including overall, year-2015 and pre-year-2015.

Period	**Model**	**R^2**	**Lower**	**Upper**
All years	SAT	0.52	0.35	0.69
n=51	SAT-W	0.49	0.28	0.66
	NOSAT	0.21	0.03	0.47
	NOSAT-W	0.43	0.22	0.63
2015-only	SAT	0.48	0.23	0.67
n=24	SAT-W	0.42	0.10	0.70
	NOSAT	0.19	0.01	0.45
	NOSAT-W	0.43	0.16	0.61
Pre-2015-only	SAT	0.48	0.18	0.74
n=27	SAT-W	0.40	0.10	0.65
	NOSAT	0.23	0.004	0.63
	NOSAT-W	0.29	0.02	0.61

Table S9. Independent evaluation of two GWR-adjusted global satellite $PM_{2.5}$ estimates (gridded at 0.1° and 0.01°) at 51 measurement sites. MSE-R^2: mean square error R^2 ($^{\wedge}$negative MSE-R2 values mean MSE of model predictions was greater than the variance of the evaluation measurements, and values less than -1 represent prediction MSE more than double the variance of the evaluation measurements). FB: fractional bias ($^{\#}$dimensionless), MB: mean bias. RMSE: root mean square error (expressed in µg/m3 and as a percentage of mean $PM_{2.5}$ at all evaluation sites).

Model	**R^2**	**MSE-$R^{2\wedge}$**	**β (SE)**	**Int.**	**RMSE (µg/m^3)**	**RMSE (%)**	**MB (µg/m^3)**	**FB$^{\#}$**
GWR-0.1°	0.08	< -1	0.26 (0.13)	5.55	2.24	31.6	-1.14	-0.19
GWR-0.01°	0.08	< -1	0.24 (0.12)	5.60	2.20	31.0	-0.85	-0.15

Table S10. Population-weighted average $PM_{2.5}$ (SD) at ~347,000 mesh block centroids by year. * this model had no time-varying predictors and the year-2015 predictions were applied to previous years unchanged. Units = μg/m^3.

Year	**SAT**	**SD**	**SAT-W**	**SD**	**NOSAT**	**SD**	**NOSAT-W***	**SD**
2000	6.6	1.6	6.8	1.4	6.8	1.6	7.3	1.7
2001	6.7	1.6	6.9	1.4	6.8	1.6	7.3	1.7
2002	6.8	1.6	7.1	1.4	6.9	1.6	7.3	1.7
2003	7.6	2.2	7.7	2.2	6.8	1.6	7.3	1.7
2004	6.4	1.5	6.7	1.3	6.8	1.6	7.3	1.7
2005	6.3	1.5	6.6	1.3	6.7	1.6	7.3	1.7
2006	7.1	1.9	7.3	1.9	6.8	1.6	7.3	1.7
2007	6.5	1.6	6.8	1.5	6.8	1.6	7.3	1.7
2008	6.2	1.6	6.5	1.5	6.7	1.6	7.3	1.7
2009	6.9	1.6	7.2	1.3	6.8	1.6	7.3	1.7
2010	5.9	1.6	6.2	1.5	6.7	1.6	7.3	1.7
2011	6.2	1.5	6.5	1.4	6.8	1.6	7.3	1.7
2012	6.7	1.6	7.0	1.4	6.7	1.6	7.3	1.7
2013	6.3	1.6	6.6	1.4	6.8	1.6	7.3	1.7
2014	6.4	1.7	6.7	1.5	6.8	1.6	7.3	1.7
2015	6.6	1.7	6.9	1.6	6.8	1.6	7.3	1.7

Table S11. Percentiles for year-2015 predictions at ~347,000 mesh blocks. Units = μg/m^3. Minimum values are not zero but are rounded to 1 decimal place.

Model	**Min**	**5th**	**25th**	**50th**	**75th**	**95th**	**Max**
SAT	0.0	2.9	5.4	6.7	7.8	9.0	16.0
SAT-W	0.0	3.8	5.7	6.8	7.9	9.1	17.6
NOSAT	0.4	4.2	5.5	6.7	7.8	9.8	19.1
NOSAT-W	0.0	4.0	6.1	7.5	8.3	9.8	15.5

Table S12. Correlation between mesh block predictions from different models (average over 2000–2015).

	SAT	**SAT-W**	**NOSAT**	**NOSAT-W**
Pearson				
SAT	1.00	0.84	0.68	0.68
SAT-W		1.00	0.50	0.71
NOSAT			1.00	0.65
NOSAT-W				1.00
Spearman				
SAT	1.00	0.84	0.67	0.70
SAT-W		1.00	0.50	0.74
NOSAT			1.00	0.66
NOSAT-W				1.00

Figure S3. A variable importance plot based on a random forest model of year-2015 $PM_{2.5}$ (n=500 trees). Predictor variables appearing higher in the plot lead to a greater increase in model MSE when they are excluded than variables lower in the list. Note 1: the variables listed here cannot be directly compared to those in the final LUR models because of the entry and retention criteria specified for LUR predictors (for example, heating and cooling degree days both appear high in this list, but because of their strong negative correlation neither would be likely to be added to a model already containing the other). Of the 13 unique predictor variables in the four final LUR models, 9 appear in the list.

Note 2: other tree sizes between 500 and 2000 were tested, and other methods ('cforest' in R package 'party') were assessed to check sensitivity to correlated predictors; results were similar.

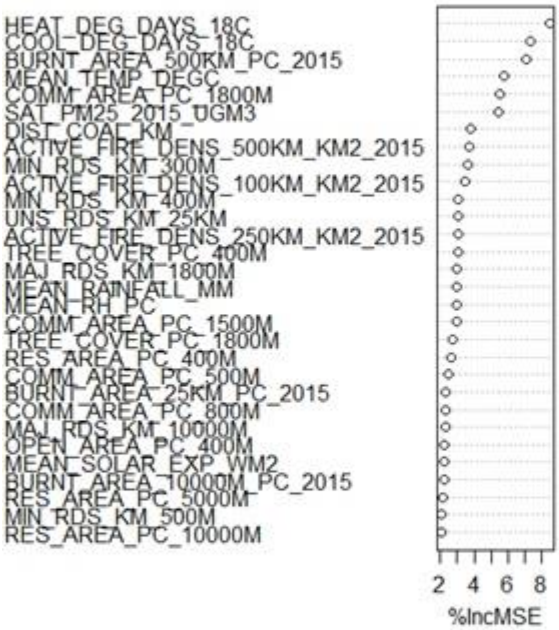


Figure S4. Independent evaluation plot (SAT model). Solid line = regression line, dashed line = line of equality. Evaluation statistics are in Table 2.

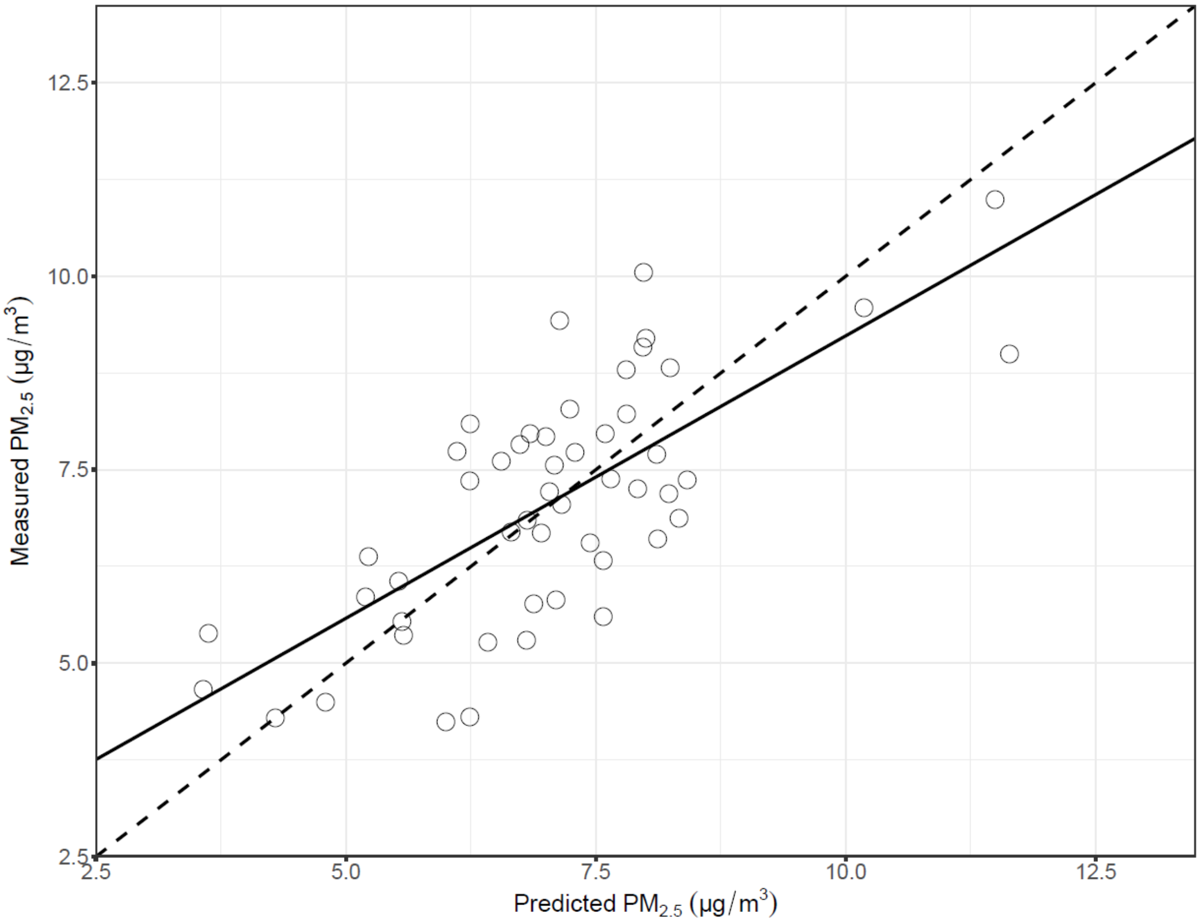


Figure S5. Independent evaluation plot (SAT-W model). Solid line = regression line, dashed line = line of equality. Evaluation statistics are in Table 2.

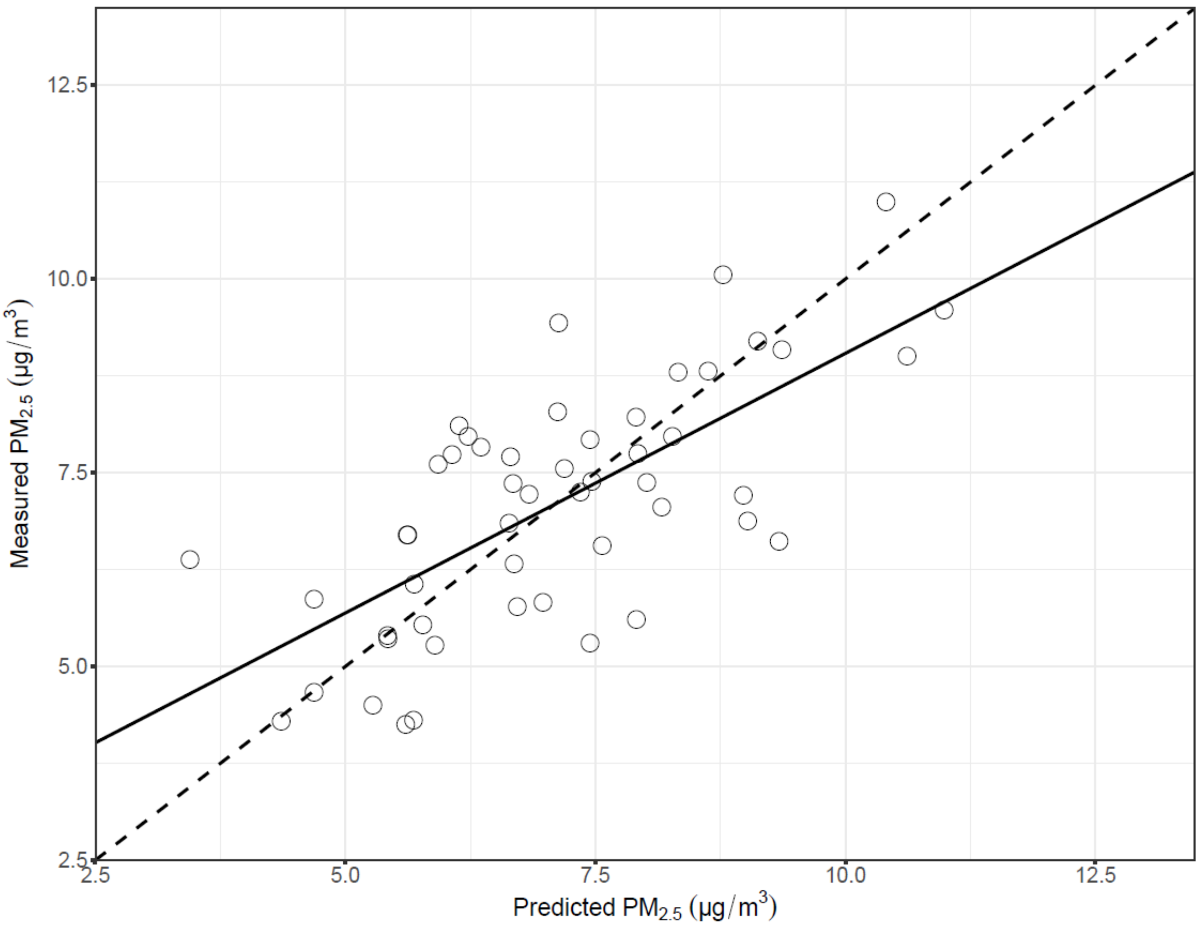


Figure S6. Independent evaluation plot (NOSAT model). Solid line = regression line, dashed line = line of equality. Evaluation statistics are in Table 2.

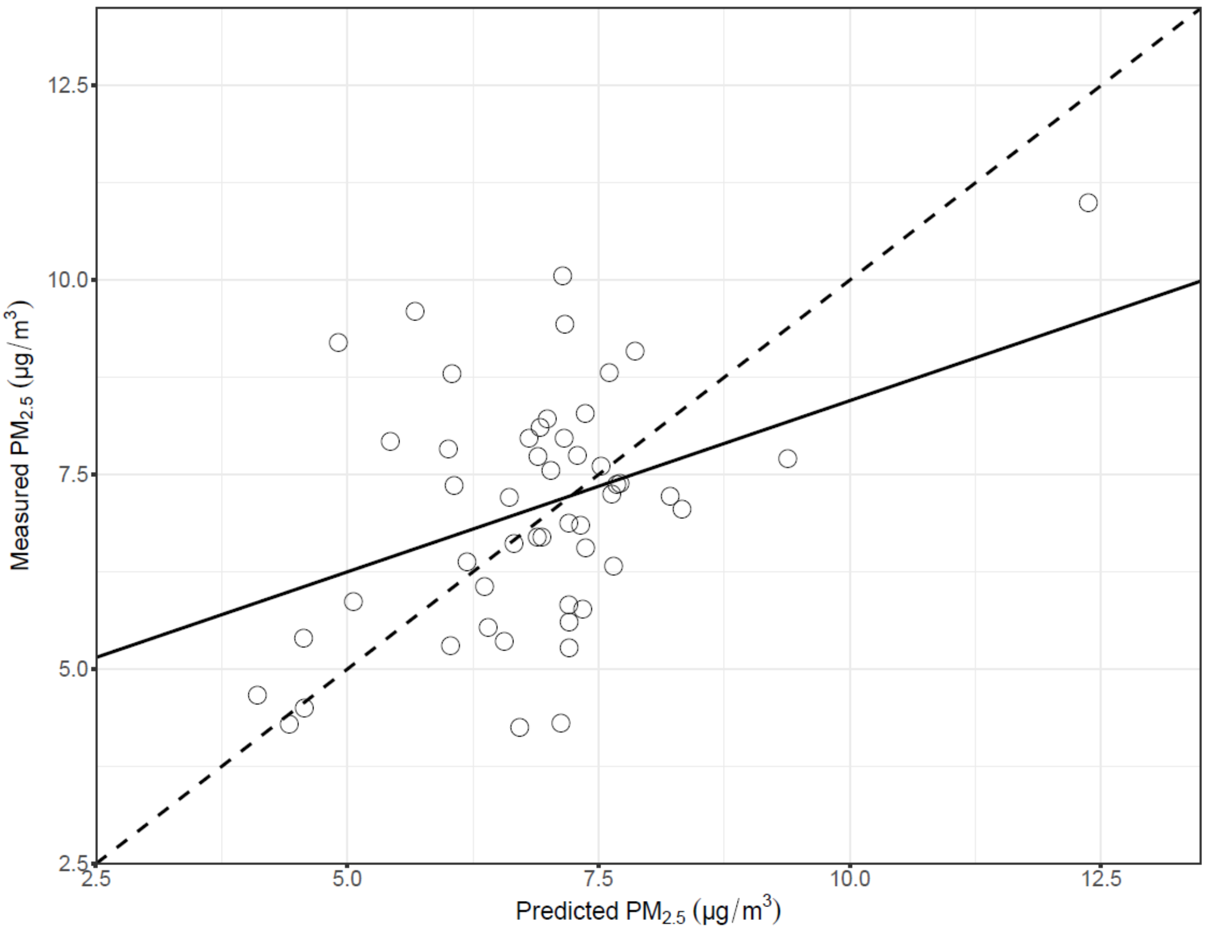


Figure S7. Independent evaluation plot (NOSAT-W model). Solid line = regression line, dashed line = line of equality. Evaluation statistics are in Table 2.

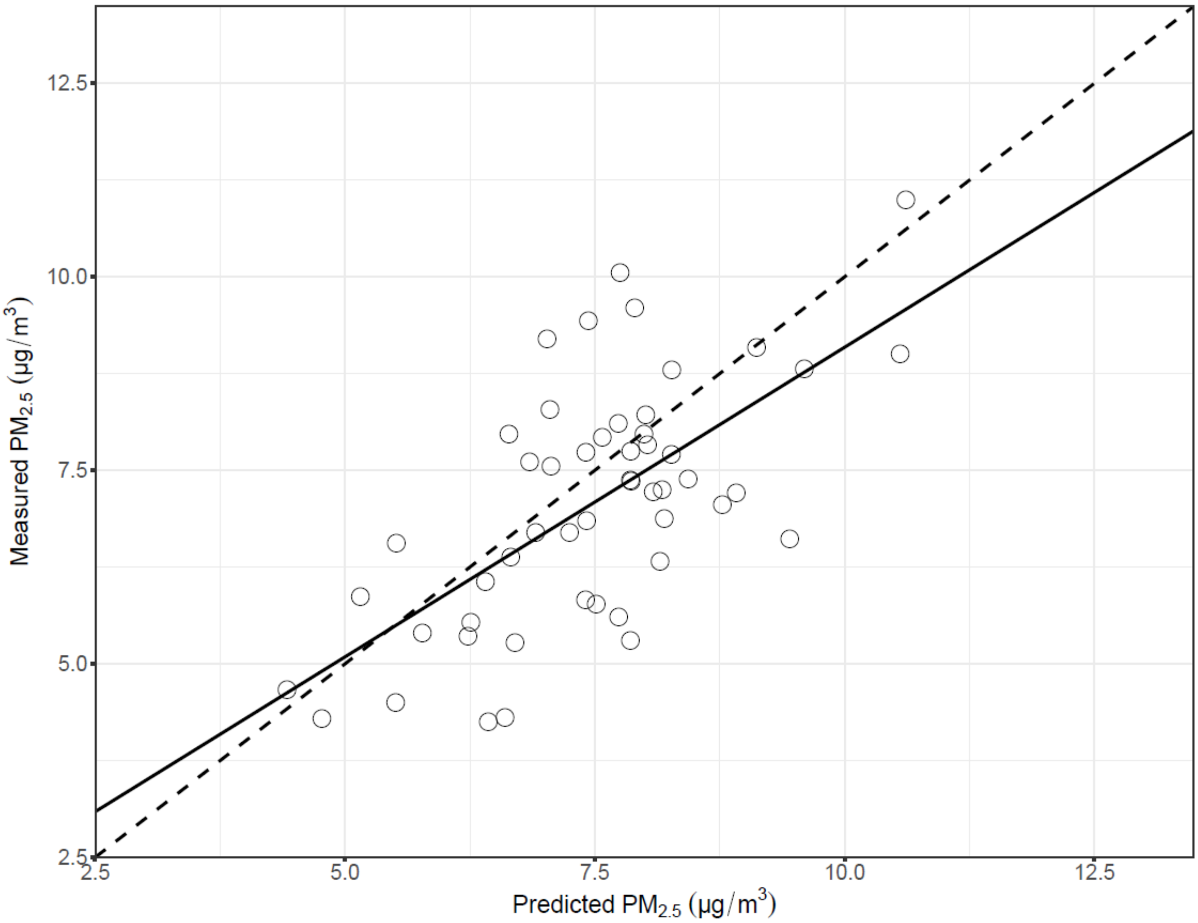


Figure S8. Measured annual mean $PM_{2.5}$ at regulatory monitoring sites during 2000–2015. Note: as described on page S4, some sites changed measurement technique in the latter part of this period.

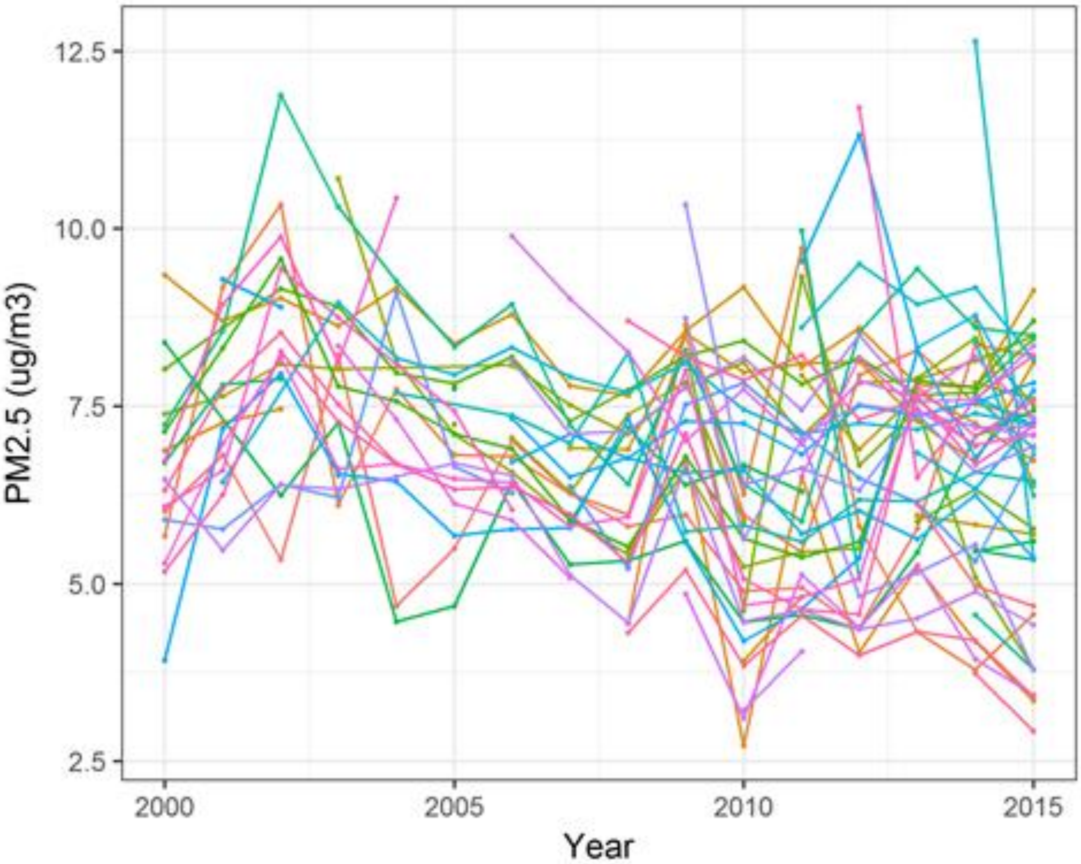


Figure S9. Distribution of SAT-W model predictions (year-2015) at ~347,000 mesh blocks.

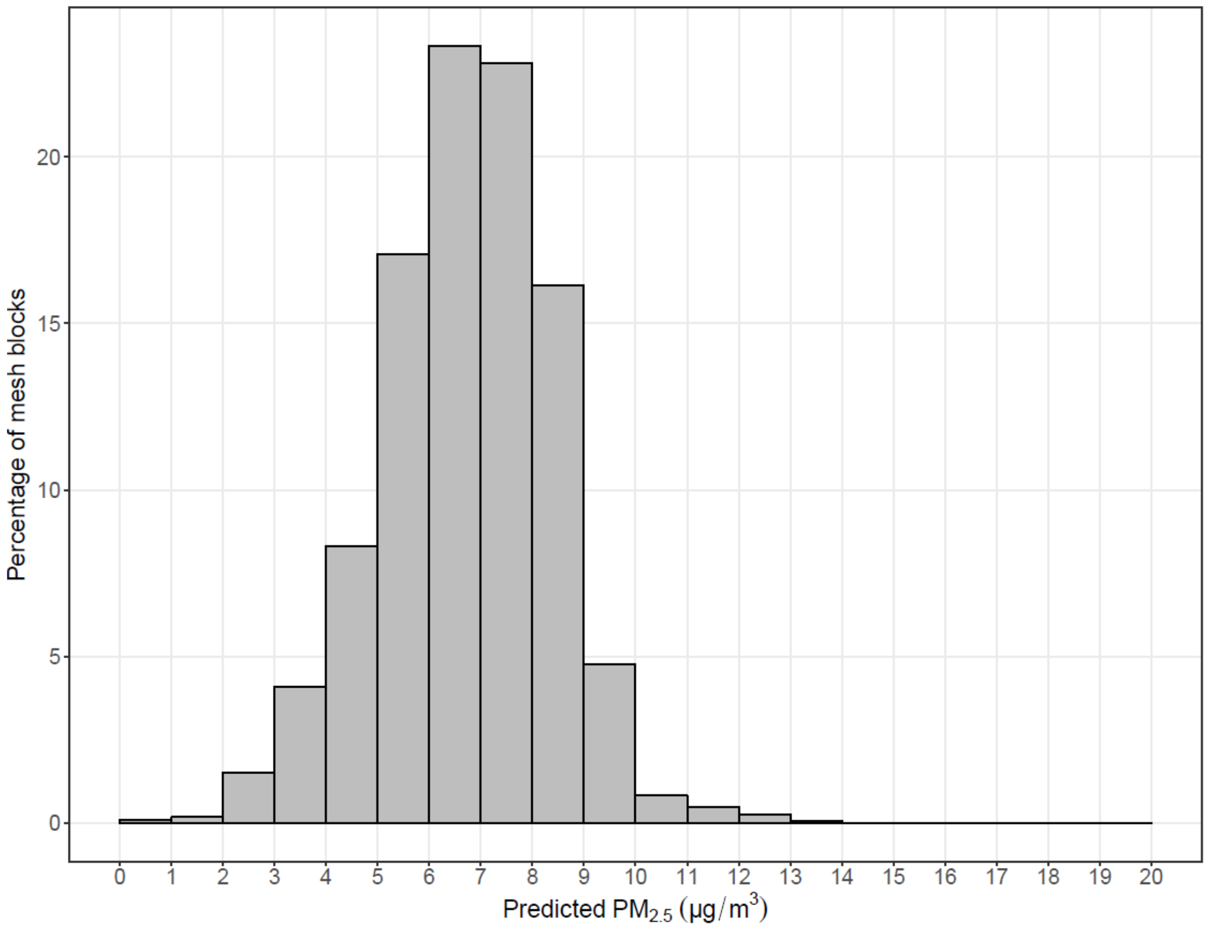


Figure S10. Distribution of NOSAT model predictions (year-2015) at ~347,000 mesh blocks.

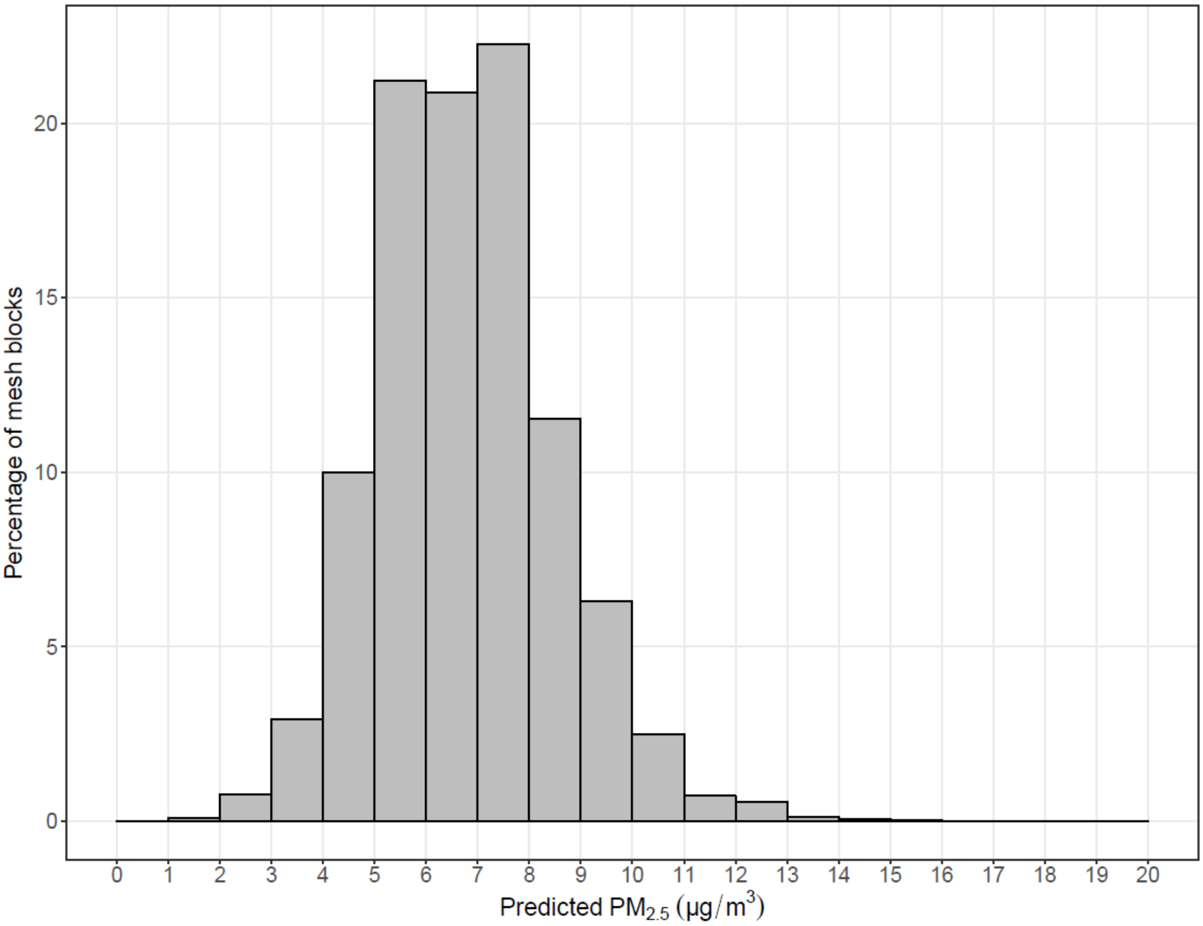


Figure S11. Distribution of NOSAT-W model predictions (year-2015) at ~347,000 mesh blocks.

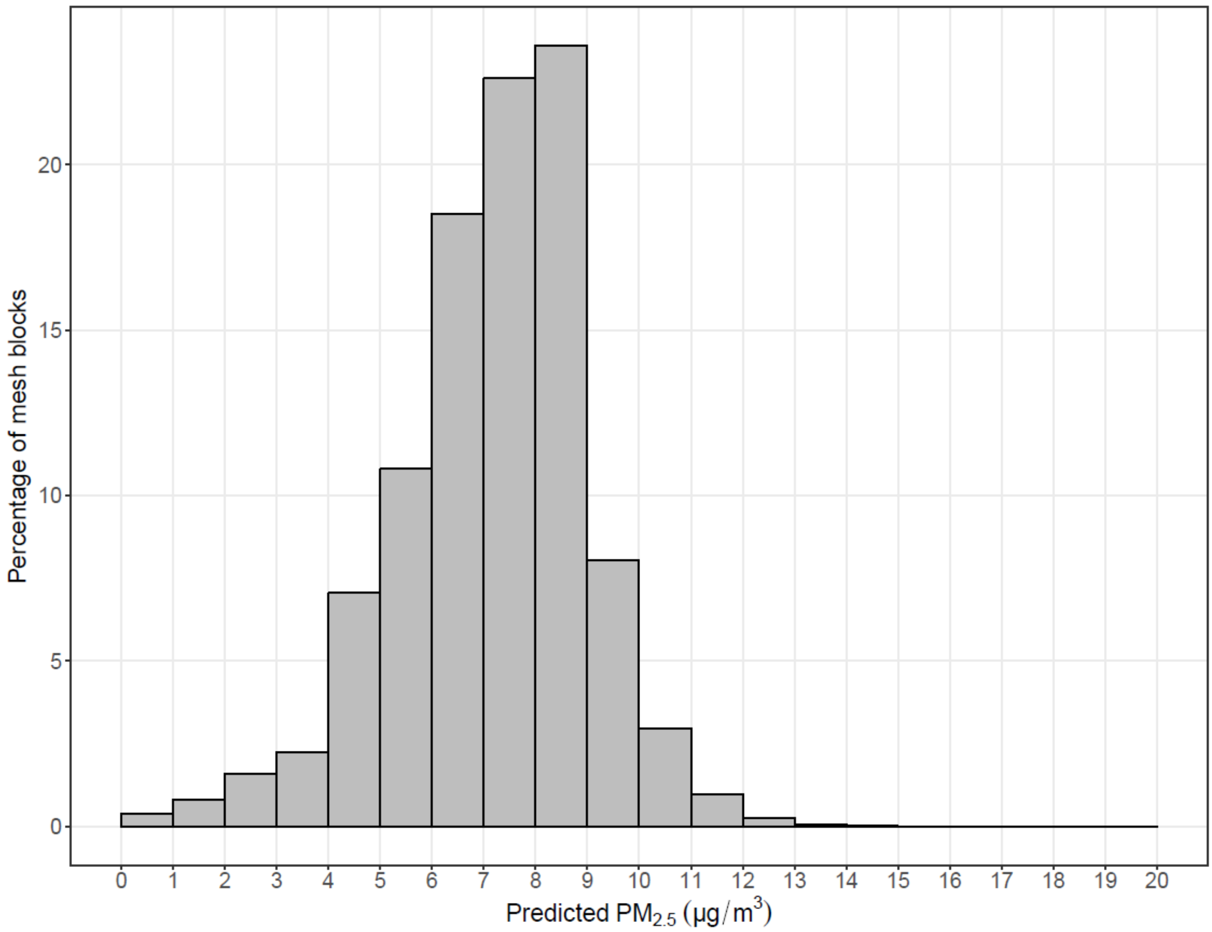


Figure S12. Map of SAT-W model predictions (year-2015).

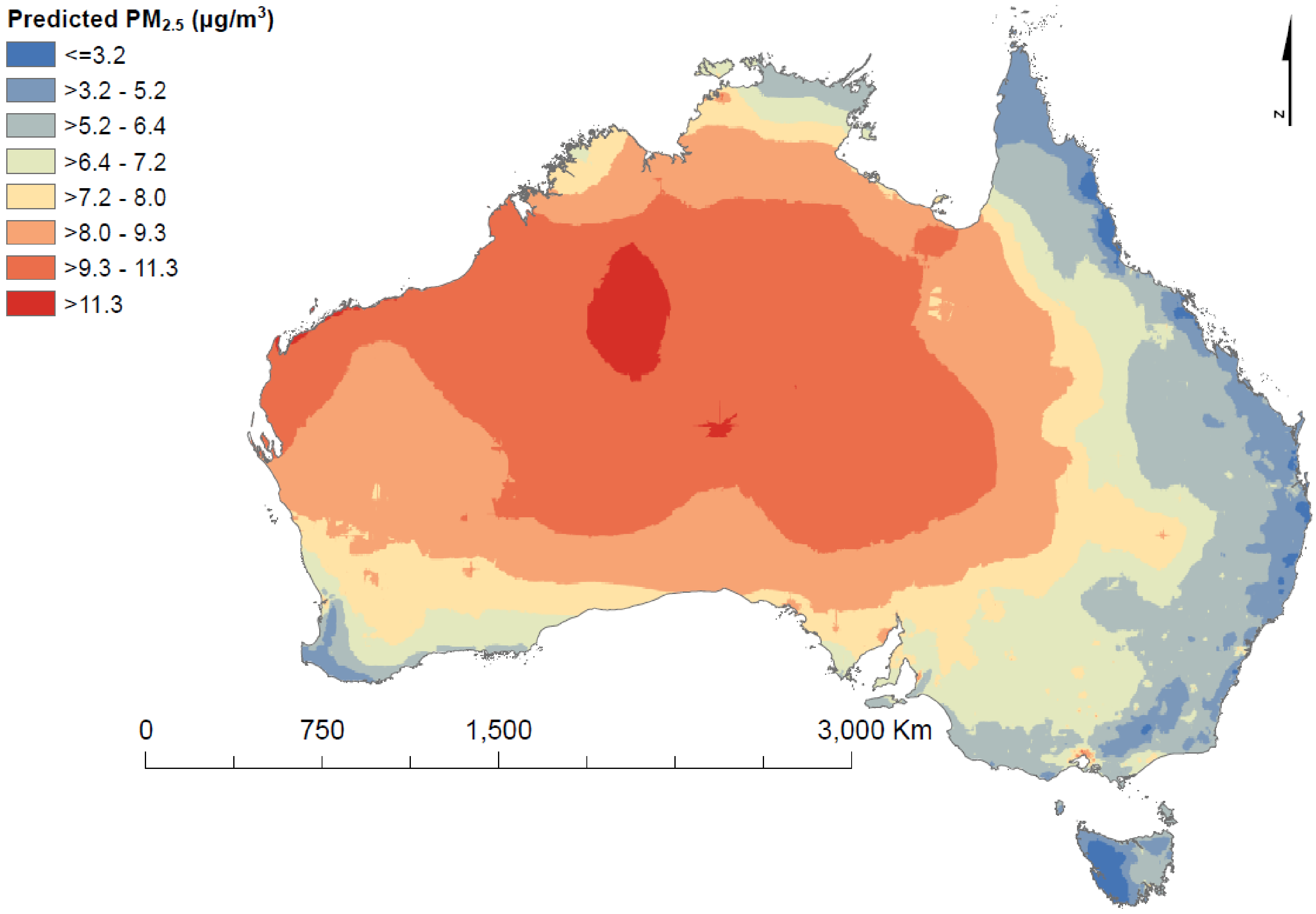


Figure S13. Map of NOSAT model predictions (year-2015).

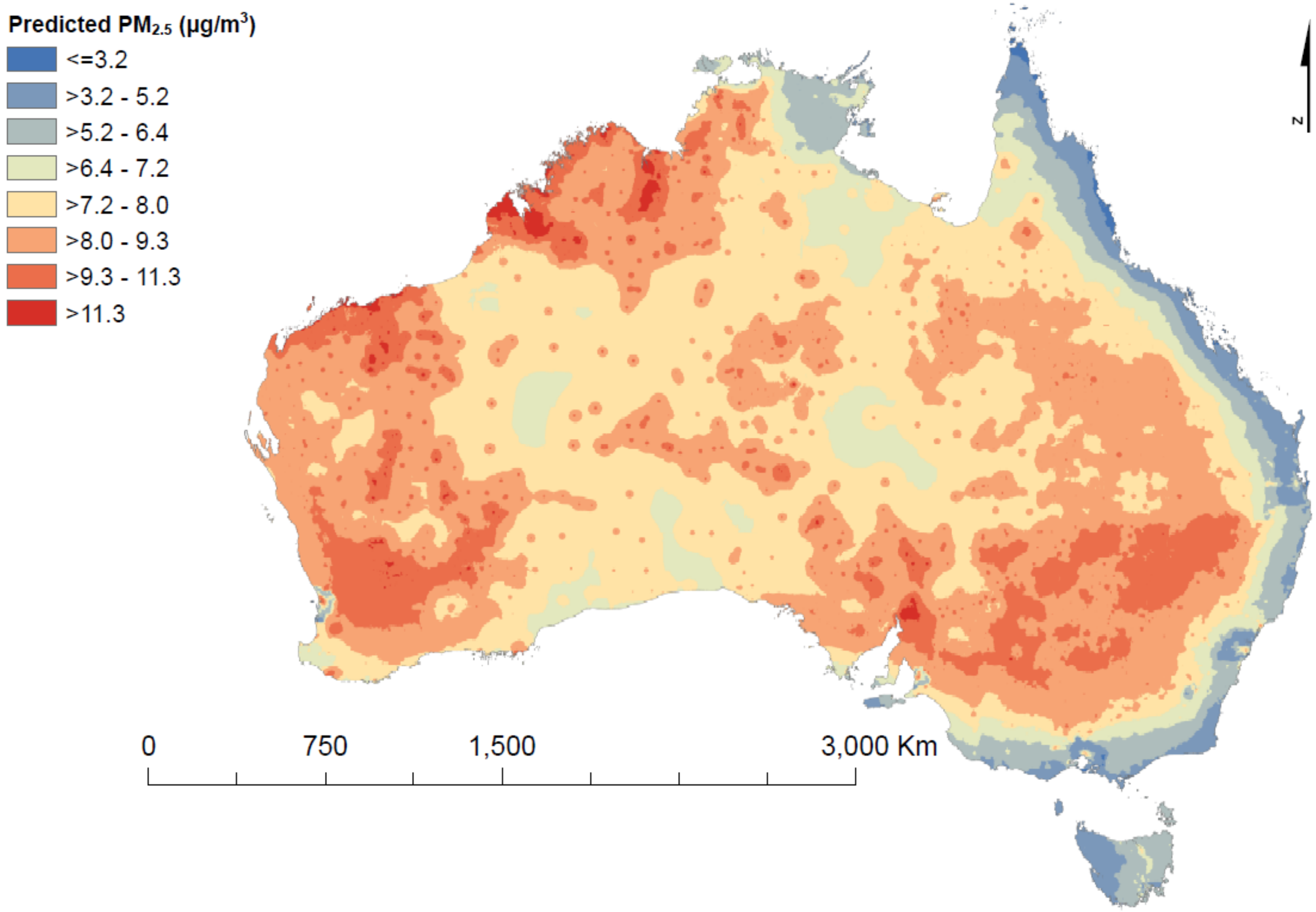


Figure S14. Map of NOSAT-W model predictions (year-2015).

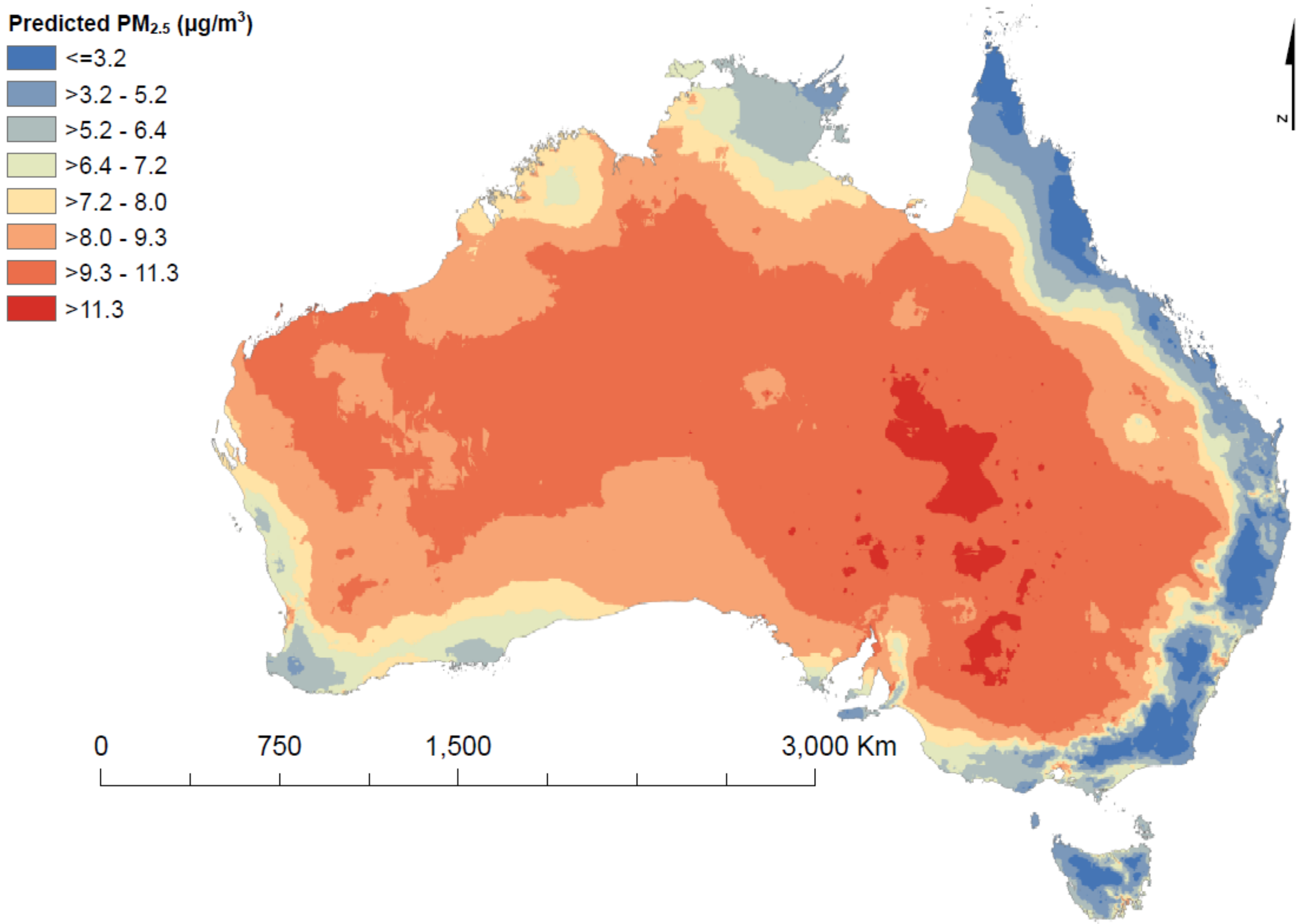


Figure S15. Correlation between LUR predictions of year-2015 $PM_{2.5}$ (x axis) and interpolated values (y-axis) from the kriging surface in Figure 2 (main text) at all ~347,000 mesh block centroids in Australia (Pearson's r = 0.996). The line of equality is shown in red.

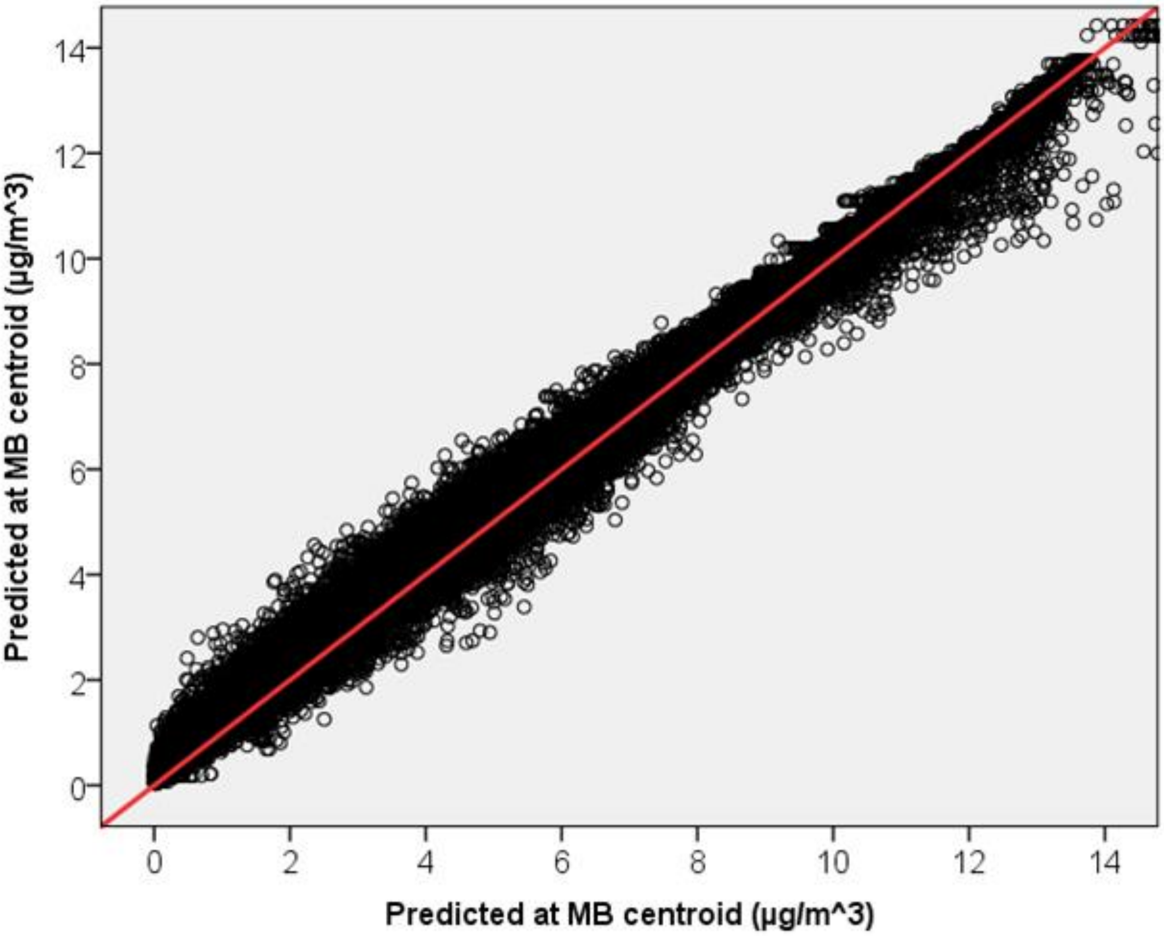
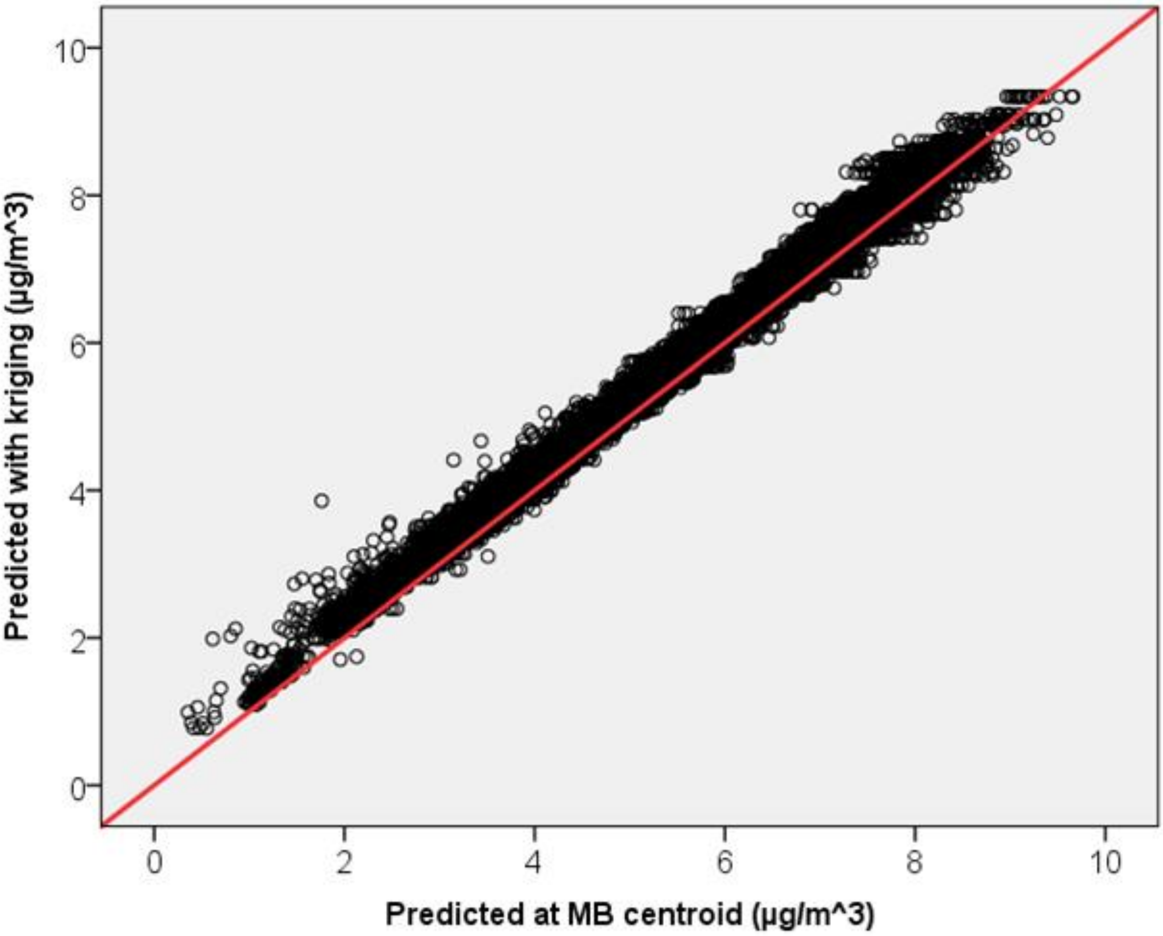


Figure S16. Correlation between LUR predictions of year-2015 $PM_{2.5}$ (x axis) and interpolated values (y-axis) from the kriging surface in Figure 2 (main text) at all ~56,000 mesh block centroids in Sydney (Pearson's $r = 0.994$). The line of equality is shown in red.

We also repeated this city-based analysis for the other capital cities shown in Figure 2 of the main text and the results were highly consistent with the national and Sydney-analyses: Darwin ($r = 0.967$); Melbourne ($r = 0.995$).



DISCUSSION

Other LUR predictors

Australian urban areas have relatively unique $PM_{2.5}$ source profiles, which differ from those in other countries where modeling studies have been performed. Our selection of LUR predictor variables was informed by Australian $PM_{2.5}$ source apportionment studies using positive matrix factorization (PMF). While receptor-based methods like PMF can lack the specificity to identify individual sources, they did suggest broad LUR predictor groups that may capture some of the complex sources and processes that underpin spatial variability of $PM_{2.5}$ in Australia.

The variables in the final models reflected most of the key $PM_{2.5}$ sources, or their proxies. For example, residential area (5 km) was positively associated with $PM_{2.5}$ in all four models, contributing 17% to adjusted R^2. Rather than a specific effect of residential area on $PM_{2.5}$, that variable may capture some traffic and other anthropogenic emissions associated with built-up residential areas (e.g., non-road combustion sources, construction, and cooking).[16] Other LUR predictors reflected meteorological and topographical factors that generally reduce $PM_{2.5}$, like higher rainfall (e.g., washout effect), wind speed, and elevation (e.g., dilution and dispersion). Tree cover (5 km) was negatively associated with $PM_{2.5}$ in the SAT model. Some trees are a biogenic source of secondary organic aerosol precursors, but in the largely urban areas and temporal scale (annual) we assessed, greater tree cover likely reflects an absence of anthropogenic sources.

REFERENCES

1. United States Environment Protection Agency, 2016. List of designated reference and equivalence methods. (https://www3.epa.gov/ttnamti1/files/ambient/criteria/AMTIC%20List%20Dec%202016-2.pdf) (accessed April 2018)
2. Australian Bureau of Statistics, 2011. Australian Statistical Geography Standard (ASGS): Volume 1, http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/D3DC26F35A8AF579CA257801000DCD7D/$File/1270055001_july%202011.pdf (accessed 1 Mar 2018)
3. Pereira, G.; Lee, H. J.; Bell, M.; Regan, A.; Malacova, E.; Mullins, B.; Knibbs, L. D., Development of a model for particulate matter pollution in Australia with implications for other satellite-based models. Environ. Res. 2017, 159, 9-15.
4. Cohen, DD, Antanacio, AJ, Stelcer, E, Garton, D. Sydney Particle Characterisation Study: $PM_{2.5}$ Source Apportionment in the Sydney Region between 2000 and 2014, 2016. New South Wales Office of Environment and Heritage. http://www.environment.nsw.gov.au/research-and-publications/publications-search/sydney-particle-characterisation-study-pm25-source-apportionment-sydney-region-2000-14 (accessed 10 Feb 2018).
5. Chan, Y.-C.; Cohen, D. D.; Hawas, O.; Stelcer, E.; Simpson, R.; Denison, L.; Wong, N.; Hodge, M.; Comino, E.; Carswell, S., Apportionment of sources of fine and coarse particles in four major Australian cities by positive matrix factorisation. Atmos. Environ. 2008, 42, (2), 374-389.
6. Hibberd M, Selleck, P, Keywood, M, Cohen, D, Stelcer E, Atanacio, A. Upper Hunter Valley Particle Characterization Study, 2013. CSIRO, Australia.

http://www.environment.nsw.gov.au/~/media/7947058EDB2E4FEFBD5598C23BB7B4FF.ashx (accessed 15 March 2018).

7. Hibberd M, Keywood, M, Selleck, P, Cohen, D, Stelcer E, Scorgie, Y, Chang, L. Lower Hunter Particle Characterisation Study, 2016. NSW Office of Environment and Heritage. http://www.environment.nsw.gov.au/~/media/1FA980C928AA4E9183DCF3B6861B48B7.ashx (accessed 30 April 2018).
8. Friend, A. J.; Ayoko, G. A.; Stelcer, E.; Cohen, D., Source apportionment of PM2.5 at two receptor sites in Brisbane, Australia. Environmental Chemistry 2011, 8, (6), 569-580.
9. van Donkelaar, A.; Martin, R. V.; Brauer, M.; Hsu, N. C.; Kahn, R. A.; Levy, R. C.; Lyapustin, A.; Sayer, A. M.; Winker, D. M., Global Estimates of Fine Particulate Matter using a Combined Geophysical-Statistical Method with Information from Satellites, Models, and Monitors. Environmental Science & Technology 2016, 50, (7), 3762-3772.
10. DiMiceli, C.M.; Carroll, M.L.; Sohlberg, R.A.; Huang, C.; Hansen, M.C.; Townshend , J.R.G. Annual Global Automated MODIS Vegetation Continuous Fields (MOD44B) at 250 m Spatial Resolution for Data Years Beginning Day 65, 2000 - 2010, Collection 5 Percent Tree Cover. Available online: http://www.landcover.org/data/vcf/ (accessed Feb 2018).
11. Elvidge, C. D.; Tuttle, B. T.; Sutton, P. C.; Howard, A. T.; Milesi, C. Global distribution and density of constructed impervious surfaces. Sensors. 7:1962–1979; 2007
12. Boschetti, L, Roy, D, Hoffmann, AA, Humber, M. MODIS Collection 5.1 Burned Area Product - MCD45 (Version 3.0.1), 2013. (http://modis-fire.umd.edu/pages/manuals.php) (accessed October 2017).
13. Giglio L. MODIS Collection 6 Active Fire Product User's Guide, Revision A, 2015. (http://modis-fire.umd.edu/pages/manuals.php) (accessed October 2017).

14. Public Sector Mapping Agencies (PSMA). Transport and Topography Data Product Description. 2013
15. Australian Bureau of Statistics, 2014. Australian Population Grid, 2011. (http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1270.0.55.007Main+Features12011?OpenDocument) (Accessed 30 October 2017).
16. Hystad, P.; Setton, E.; Carvantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M., et al. Creating National Air Pollution Models for Population Exposure Assessment in Canada. Environ Health Perspect. 119:1123-1129; 2011