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, The 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)

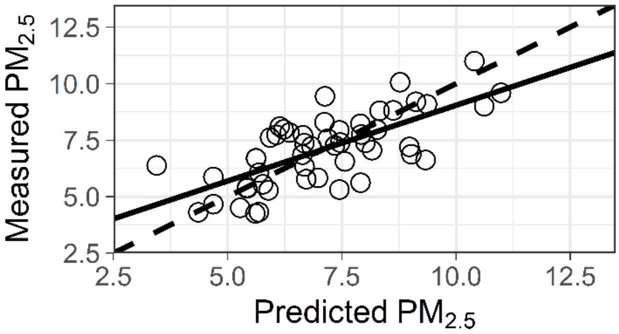
Word count (excl. title page, captions, references): 5,452

Small tables: 2

Small figures: 1

Large figures: 1

TOC Art



ABSTRACT

Australia has relatively diverse sources and low concentrations of ambient fine particulate matter (<2.5 µm, $PM_{2.5}$). Few comparable regions are available to evaluate the utility of continental-scale land-use regression (LUR) models including global geophysical estimates of $PM_{2.5}$, derived by relating satellite-observed aerosol optical depth to ground-level $PM_{2.5}$ ('SAT-$PM_{2.5}$'). We aimed to determine the validity of such satellite-based LUR models for $PM_{2.5}$ in Australia.

We used global SAT-$PM_{2.5}$ estimates (~10 km grid) and local land-use predictors to develop four LUR models for year-2015 (two satellite-based, two non-satellite-based). We evaluated model performance at 51 independent monitoring sites not used for model development. An LUR model that included the SAT-$PM_{2.5}$ predictor variable (and six others) explained the most spatial variability in $PM_{2.5}$ (adjusted $R^2 = 0.63$, RMSE (µg/m^3 [%]): 0.96 [14%]). Performance decreased modestly when evaluated (evaluation $R^2 = 0.52$, RMSE: 1.15 [16%]). The evaluation R^2 of the SAT-$PM_{2.5}$ estimate alone was 0.26 (RMSE: 3.97 [56%]). SAT-$PM_{2.5}$ estimates improved LUR model performance, while local land-use predictors increased the utility of global SAT-$PM_{2.5}$ estimates, including enhanced characterization of within-city gradients.

Our findings support the validity of continental-scale satellite-based LUR modeling for $PM_{2.5}$ exposure assessment in Australia.

INTRODUCTION

Long-term exposure (>1 year) to ambient particulate matter with an aerodynamic diameter <2.5 μm ($PM_{2.5}$) was the sixth-leading risk factor for global mortality in 2016.[1] Over the last decade, improvements in the spatio-temporal resolution of satellite-derived data have increased their utility for air pollution exposure assessment in epidemiological studies.[2-7] Satellites have enabled exposure assessment to be extended to regions with few or no ground-based air quality monitors.[8-10] Even in regions with relatively dense monitoring, satellites have, for example, enabled exposure assessment in urban-rural fringe and rural areas, allowing the study of more representative population-based cohorts than those selected based on their proximity to urban monitoring sites.[11-13]

However, despite these recent advances, the spatial resolution of most satellite instruments and processing algorithms may not fully capture local-scale, small-area (~1 km or less) exposure contrasts within cities, which may be of interest in epidemiological studies.[4] For example, although some global retrievals with higher resolution capability are emerging, such as the 1 × 1 km Multi-Angle Implementation of Atmospheric Correction (MAIAC),[14] most satellite retrievals of aerosol optical depth (AOD) are at around 10 × 10 km resolution. In turn, relating AOD to surface $PM_{2.5}$ employs information at coarser resolution than AOD. Among 3,646 global cities with a population >100,000 people, the median city was ~5 km wide (population-weighted value: ~13 km). Among megacities (population >10 million, representing ~15% of the total city-dwelling population), the median width was 48 km.[15] Those findings suggest that satellite data may only capture within-city $PM_{2.5}$ gradients in very large cities.

One method for potentially improving the spatial resolution of $PM_{2.5}$ estimates is to use geophysically-derived estimates, obtained by relating satellite AOD to surface $PM_{2.5}$ concentrations using chemical transport model (CTM) simulations, in land-use regression (LUR) models (e.g.,[16-18]). LUR, an empirical method, is often used to estimate within-city variability in pollutant concentrations using saturation sampling and spatial predictor variables.[5] Geophysical or other satellite-derived $PM_{2.5}$ estimates can be used as an independent (predictor) variable in an LUR model ('satellite-based LUR'), alongside other satellite and non-satellite-derived land-use predictors. This approach attempts to combine the information advantages of satellite-derived estimates (i.e., large spatial coverage, facilitating continental-scale models over long temporal scales) with those of traditional LUR (i.e., predictors that aim to capture local-scale sources of variability).

Satellite-based LUR, and related spatial modeling methods, have been used to estimate annual or daily $PM_{2.5}$ at national or continental scale in the Northern Hemisphere: Canada,[17] Western Europe,[18] China,[19-21] and the USA.[16,22,23] The models in those studies were developed using 177, 549, 835, 1185, 1479, 1464, 1462, and 1928 regulatory $PM_{2.5}$ monitoring sites, respectively, an approximate monitor density (monitors per million people) of: ~5.2 (Canada), ~1.3 (Western Europe), ~1.1 (China), and ~6.3 (USA). For comparison, Australia (year-2015: ~23.7 million people, 49 $PM_{2.5}$ monitors) has a monitor density of ~2.1; more dense than China and Western Europe, but substantially less dense than the USA and Canada.

Moreover, $PM_{2.5}$ in major Australian cities, which are all located on the coastline, has a different source profile compared with countries where previous modeling studies were performed. For example, the contribution of chloride-depleted, pollutant-aged sea salt to $PM_{2.5}$ mass is generally larger, while that of mixed or unspecified anthropogenic sources of secondary sulfates and nitrates is generally smaller.[24,25] Australia also has some of the lowest ambient $PM_{2.5}$ concentrations in the world, providing an opportunity to assess the utility of satellite-based LUR at low levels.[8]

Australia legislated an ambient $PM_{2.5}$ standard in 2016 (8 μg/m^3 annual average, 25 μg/m^3 daily average).[26] As such, limited historical $PM_{2.5}$ observations can make long-term exposure assessment challenging. In this study, we aimed to: (1) determine if satellite-based LUR for $PM_{2.5}$ could be applied to Australia, a Southern Hemisphere setting with a diverse mixture of natural and anthropogenic sources and low PM concentrations, and; (2) evaluate the extent to which inclusion of land-use predictor variables improved upon geophysical satellite-derived estimates, and vice-versa.

MATERIALS AND METHODS

$PM_{2.5}$ measurements

We obtained daily $PM_{2.5}$ measurements from ambient air monitoring agencies in Australia for the period 1 Jan 2000 through 31 Dec 2015, where: (a) measurements were collected for at least one year; (b) the latitude and longitude of the monitoring site (including any relocations) was known to five decimal places; (c) the data had undergone quality assurance (QA) procedures, and; (d) an offline or continuous gravimetric method for $PM_{2.5}$ mass concentration was used.[27] We obtained measurements collected at 70 sites that met these inclusion criteria (Figure S1). We classified each site's location based on census data, proximity to roads, and industrial $PM_{2.5}$ sources (supporting information [SI], page S4).

$PM_{2.5}$ measurements were performed using different methods, including offline gravimetric (low-volume samplers, including Partisol instruments) and continuous gravimetric techniques (tapered element oscillating microbalance [TEOM] and beta-attenuation monitor). We did not exclude sites on the basis of measurement method *a priori*, and assumed that sites using different methods could contribute useful information about the spatio-temporal variability of $PM_{2.5}$. TEOMs, which were used at the large majority of sites (~85%), can underestimate $PM_{2.5}$ because losses of semi-volatile and volatile compounds occur when the sample is heated to remove particle-bound water.[28] We first investigated the relationship between $PM_{2.5}$ measured using different methods (SI, pages S3-S5). We then calculated mean $PM_{2.5}$ in each year for sites with ≥70% non-missing daily measurements (for continuous measurements only) as a balance between maximizing sites and capturing potential seasonal variability. We used offline measurements (collected one-day-in-three) for subsequent model evaluation ('Independent evaluation' section).

Satellite-based $PM_{2.5}$ predictor variables

We used satellite-based estimates of surface $PM_{2.5}$ as an LUR predictor. Estimates for Australia (years-2000–2015) were subset from a global set of annual mean $PM_{2.5}$ estimates (years-1998–2015), described in detail by van Donkelaar et al.[29] Their approach used five daily AOD products, calibrated against ground-based sun photometer observations from the aerosol robotic network (AERONET), to determine the relative uncertainty of each AOD product. A CTM (GEOS-Chem) was used to simulate the relationship between AOD and ground-level $PM_{2.5}$, and provide an additional source of AOD data. The CTM simulations of AOD to near-surface extinction were adjusted using space-borne observations of aerosol light extinction from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP). The adjusted AOD to $PM_{2.5}$ relationship was then applied to each of the five AOD sources to estimate near-surface $PM_{2.5}$ on a 0.1° (~10 km) grid. Finally, each data source was combined, weighted according to its uncertainty, to produce global geophysical satellite-based estimates of annual mean $PM_{2.5}$ at 0.1° (SAT-$PM_{2.5}$; V4.GL.02.NoGWR).[29,30]

van Donkelaar et al.[29] describe the subsequent step of applying geographically-weighted regression (GWR) to adjust for residual bias between the SAT-$PM_{2.5}$ estimates and a global network of ground-based PM monitors (2008–2013), including regulatory monitors in Australia (n=32 for $PM_{2.5}$, n=20 for PM_{10}, total=52 sites), based on a spatially-varying relationship between the bias and simulated aerosol composition, urban land surfaces, and local topography.[29] The application of GWR primarily accounted for large-scale bias in the global $PM_{2.5}$ estimates, introduced by the CTM used to relate AOD to surface $PM_{2.5}$, since the GWR training was restricted to monitors >100 km away.[29]

We used the geophysical 0.1° annual SAT-$PM_{2.5}$ estimates (global cross-validation [CV] R^2 = 0.65) as an LUR predictor. Two global GWR-adjusted estimates were also available, gridded at 0.01° and 0.1°, respectively. We did not use these GWR-adjusted estimates as an LUR predictor because they already included information on measured $PM_{2.5}$ (i.e., because $PM_{2.5}$ monitors used to derive the GWR adjustment in Australia were the dependent variable in this study) and land-use variables. We performed sensitivity analyses to assess the effect of the decision to exclude the GWR estimates from the analysis.

Other LUR predictor variables

We built on methods used in our previous national satellite-based LUR for NO_2,[31] and expanded them to include predictors more relevant to $PM_{2.5}$, informed by Australia-specific $PM_{2.5}$ source apportionment studies.[24,32-36] We sought satellite- and non-satellite spatial predictor variables that could potentially capture natural and anthropogenic sources of variability in $PM_{2.5}$.[16-18,37,38]

The predictor variables we selected, their spatio-temporal resolution, source, and pre-processing are summarized in Table S1. Some variables were calculated at each monitoring site (e.g., point estimates of elevation). Other variables were calculated around each site in 22 circular buffers with radii of 100 m to 10 km as either an average (e.g., population density) or sum (e.g., length of major roads) within each buffer. Larger buffers were included for predictor variables that could potentially influence $PM_{2.5}$ over greater distances. For example, annual burned area and active fire (‘hotspot’) density were calculated in 10, 25, 50, 100, 250, and 500 km buffers only.[39-41]

LUR models assessed

We explored two LUR models: (1) offered no satellite-based $PM_{2.5}$ estimates ('NOSAT'), and; (2) offered geophysical ~10 km SAT-$PM_{2.5}$ estimates ('SAT'). All other variables offered to each model were identical, including the non-$PM_{2.5}$ satellite-derived variables described in Table S1 (i.e., the NOSAT model referred to no satellite-based $PM_{2.5}$ only, not the absence of all satellite predictors).

Using the NOSAT model, we sought to establish a non-satellite baseline to compare with the model offered the SAT-$PM_{2.5}$ predictor variable to assess their relative utility for predicting $PM_{2.5}$. We did not force SAT-$PM_{2.5}$ into the LUR model, as has been done in other studies to test its contribution.[18,38] We assumed that if SAT-$PM_{2.5}$ was an important predictor of $PM_{2.5}$, it would be selected during LUR model development (i.e., it was only added if it met the same inclusion criteria as other LUR predictors, described in the following section).

LUR model development

Of the 70 sites that measured for at least one year during 2000–2015, the greatest number of simultaneous measurements occurred in 2015 (n=49). We developed LUR models for year-2015, and used the remaining 21 sites as part of an independent evaluation of model predictions in earlier years (see 'Independent evaluation'). We explored using our previous method for NO_2 by fitting a longitudinal model to all years of data,[31] but the limitations that method imposed on variable selection, coupled with the small number of sites in earlier years, led us to use a different approach here.

We used a supervised forward addition linear regression method for LUR development, guided by the well-documented European Study of Cohorts for Air Pollution Effects (ESCAPE) protocol,[42] which has been widely implemented and validated for $PM_{2.5}$ at city- through to national- and continental-scale, including for satellite-based LUR models.[18,38,43,44]

Forward addition is a standard approach in LUR model development, the protocols for which generally emphasize the empirical plausibility of predictor variables and model parsimony.[5] However, variable selection using forward addition can be unstable.[48] We performed sensitivity analyses to assess the variables selected using resampling-based methods (random forest with 500-2000 trees; 'randomForest' package in R v3.4.4), compared with those selected using forward addition.

There were 313 potential predictor variables for year-2015 (291 buffer and 22 point variables, Table S1). We excluded variables, both time-varying and non-time-varying, with >75% zero or missing values (e.g., variables using buffers smaller than the resolution of satellite-derived data), leaving 274 variables. The expected effect direction of each variable was defined prior to model development, following standard LUR methods.[5] Variables that reflected known or potential $PM_{2.5}$ sources in Australia (e.g., landscape fires, vehicle emissions, coal-fired power generation) were expected to be positively associated with $PM_{2.5}$, while those that reduce $PM_{2.5}$ concentrations, or indicate an absence of anthropogenic sources, were expected to be negatively associated (e.g., rainfall, wind speed, tree cover).

Informed by the ESCAPE protocol,[42] we started the model with the variable that explained the most variability in measured $PM_{2.5}$ (based on adjusted R^2), provided its coefficient followed the pre-specified direction. Of the remaining variables, the one explaining the most variability in $PM_{2.5}$ was added if it: (1) was statistically significant at the 5% level, and (2) increased the adjusted R^2 of the model by >1%, and; (3) had a coefficient in the specified direction, and; (4) did not change the coefficient direction of a variable already in the model. The process was repeated until no variables remained that fulfilled these criteria.

LUR model diagnostics

Following the ESCAPE method, predictor variables that were initially included were checked for ongoing statistical significance at $p \leq 0.20$; any variables not meeting this criterion were removed.[42] The variance inflation factor (VIF) of all variables was assessed for evidence of collinearity (VIF >3), and potentially influential observations were evaluated using df-beta values ($>2/\sqrt{n}$) and Cook's distance ($>4/n$). Model residuals were checked for normality and constant variance, and residual spatial autocorrelation was tested with Moran's *I* in ArcGIS v.10.4.

Due to the relatively small number of sites, we investigated the sensitivity of the final LUR models to site selection. Following Hystad et al.,[17] we used a bootstrap method ('boot' package in R[45]) based on a random sample of year-2015 sites (n=49, with replacement) to obtain predictor coefficients and R^2 of LUR models; the process was repeated 10,000 times to determine 95% bootstrap confidence intervals. As we sought to apply LUR models to other years, the same method was used to indicate the temporal sensitivity of a year-2015 model to site selection in four randomly drawn years, one each from the 4 four-year periods in our study (excluding 2015): 2003, 2007, 2010, and 2014.

Independent evaluation

The model development R^2 can overestimate LUR performance at sites not used for development, and this has also been demonstrated for leave-one-out-cross validation when applied to LUR models.[46-50] We therefore sought additional, independent $PM_{2.5}$ measurements, which were not used for model development, in order to assess potential overfitting and evaluate the LUR models' robustness (i.e., the extent to which model fit and error changed when tested with new data).

The $PM_{2.5}$ sites not used for LUR model development in year-2015 (n=21) were used to calculate annual $PM_{2.5}$ for each year they had valid observations during 2000–2014. We then performed web searches and contacted investigators to locate additional $PM_{2.5}$ measurements (2000–2015), with an emphasis on publicly-available non-regulatory monitoring campaigns done by government agencies and non-regulatory sites operated privately (i.e., by companies, universities or research institutes). We applied similar inclusion criteria to these measurements as those for LUR model development (precise site co-ordinates, gravimetric measurements, ≥1 year monitoring), but were less stringent for some criteria (SI, pages S11-S12).

We identified 30 additional sites, for a total of 51 independent evaluation sites (=165 site-years of $PM_{2.5}$ observations, table S2, figure S2). None of the LUR model development sites were used for evaluation. LUR models predicted annual average $PM_{2.5}$ at each site for the years that it had valid data by matching any annual time-varying predictor variables, which were available for the entire period, to the same period and applying the year-2015 coefficients.

We used standard metrics to compare LUR model predictions against evaluation data; measured $PM_{2.5}$ regressed on predicted $PM_{2.5}$ (R^2, including bootstrap estimates), coefficients, root mean square error (RMSE, absolute and percentage), and bias (mean and fractional). In addition to R^2, which is the squared correlation of measurements and predictions, we calculated the mean square error R^2 (MSE-R^2) as an indicator of absolute agreement on a 1:1 line; MSE-R^2 can be negative if the MSE of model predictions is greater than the variance of evaluation measurements.[46,50-52] We compared overall results to those for year-2015 (47% of sites) and pre-2015 only (53% of sites) to assess the sensitivity of the evaluation to site selection in both periods. The same regression diagnostics used for LUR development were applied to evaluation.

Comparison of LUR models and satellite $PM_{2.5}$

We used the evaluation measurements to compare the performance of national LUR models, one of which was offered the SAT-$PM_{2.5}$ estimate, with the global SAT-$PM_{2.5}$ estimates used alone. We included the GWR-adjusted global estimates in the evaluation as a sensitivity analysis.

LUR model application and representativeness

We used the final LUR models to estimate annual mean $PM_{2.5}$ at the ~347,000 census mesh block centroids that cover Australia. Mesh blocks are the smallest spatial unit in the national census and have an average (± SD) population of 62 people (±59), while 25% have a population of zero. Mesh blocks have a large size range (0.0001 to 165,000 km^2; 25th percentile: 0.02 km^2, 50th: 0.04 km^2, 75th: 0.09 km^2).[53] We calculated the population-weighted mean $PM_{2.5}$ concentrations using both LUR models; if a model included a time-varying predictor variable, we estimated $PM_{2.5}$ for each year during 2000–2015. We interpolated between predicted $PM_{2.5}$ concentrations at each mesh block using ordinary kriging in ArcGIS to produce national maps at 1 × 1 km resolution for display purposes.

The correlation between mesh block-level predictions of different LUR models was calculated using Pearson's r, and Spearman's ρ to compare if predictions were relatively high or low in the same area.[54] Using mesh block predictions from our previous national NO_2 LUR model (2006–2011),[31] we assessed the correlation between annual mean $PM_{2.5}$ and NO_2 to compare the spatial variability of the two pollutants.

The percentiles of LUR predictor variables at model development and evaluation sites were compared to those at mesh blocks. We used this as a general indicator of the sites' validity for developing and evaluating, respectively, $PM_{2.5}$ exposure models for the Australian population.

RESULTS

The mean (± SD) $PM_{2.5}$ concentration at model development sites in year-2015 was 6.9 (± 1.6) μg/m^3. Additional descriptive statistics are in Table S3.

LUR models

The final LUR models are in Table 1. The LUR model offered the SAT-$PM_{2.5}$ predictor variable (SAT) included it and six other variables (adjusted R^2: 0.63, RMSE [%]: 0.96 μg/m^3 [14%]). The SAT-$PM_{2.5}$ variable contributed 0.10 to the adjusted R^2 of this model. The model not offered the satellite-$PM_{2.5}$ variable (NOSAT) had four predictors and an adjusted R^2 of 0.59 (RMSE: 1.01 μg/m^3 [16%]). Residential area (5 km) was in both models, contributing 0.17 to the adjusted R^2 value of each.

The proportion of households using wood heaters was in both models, contributing 0.03 and 0.22 to the adjusted R^2 of SAT and NOSAT, respectively. This wood heater variable, derived from a national energy-use survey,[55] had only two values for each of the six states in Australia; one for each state's capital city and one for the remainder of each state. Because of its coarse resolution, we explored the robustness of the LUR models to the exclusion of that predictor variable. Applying that approach, the final models (SAT minus wood ['SAT-W'] and NOSAT minus wood ['NOSAT-W']) had adjusted R^2 values of 0.51 (RMSE: 1.10 μg/m^3 [15%]) and 0.56 (RMSE: 1.05 μg/m^3 [15%]), respectively.

Table 1. LUR model summary. Predictor variables are listed in the order they were added to model. *time-varying (annual) predictor. ^cumulative total adjusted R^2 with the addition of each variable. SAT: LUR model offered the SAT-$PM_{2.5}$ variable; SAT-W: same as SAT, but wood heater predictor not offered; NOSAT: LUR model not offered SAT-$PM_{2.5}$ variable; NOSAT-W: same as NOSAT, but wood heater predictor not offered. VIF: variance inflation factor; RMSE: root mean square error (expressed in μg/m^3 and as a percentage of mean $PM_{2.5}$ at all measured sites).

LUR model	**Predictor variable (units)**	**β (SE)**	**P**	**Adj. R^{2}^**	**VIF**
SAT	Intercept	5.62 (0.74)	<0.001		
Adj. R^2 (R^2): 0.63 (0.68)	Residential area, 5 km (%)	0.03 (0.01)	0.001	0.17	1.3
RMSE (μg/m^3): 0.96 (14%)	Sat-$PM_{2.5}$ (μg/m^3)*	0.91 (0.14)	<0.001	0.26	1.3
	Annual average rainfall (mm)	-0.002 (0.001)	0.004	0.43	1.5
	Commercial area, 1.5 km (%)	0.04 (0.02)	0.116	0.51	1.4
	Households using wood heaters (%)	0.07 (0.02)	0.001	0.54	2.0
	Tree cover, 5 km (%)	-0.07 (0.03)	0.016	0.59	1.6
	Annual average wind speed (km/h)	-0.17 (0.07)	0.027	0.63	1.2
SAT-W	Intercept	5.64 (0.66)	<0.001		
Adj. R^2 (R^2): 0.51 (0.55)	Residential area, 5 km (%)	0.03 (0.01)	0.009	0.17	1.1
RMSE (μg/m^3): 1.10 (16%)	Sat-$PM_{2.5}$ (μg/m^3)*	0.83 (0.16)	<0.001	0.26	1.2
	Annual average rainfall (mm)	-0.003 (0.001)	<0.001	0.43	1.2
	Commercial area, 1.5 km (%)	0.08 (0.03)	0.005	0.51	1.2
NOSAT	Intercept	14.8 (2.1)	<0.001		
Adj. R^2 (R^2): 0.59 (0.63)	Residential area, 5 km (%)	0.02 (0.01)	0.052	0.17	2.0
RMSE (μg/ m^3): 1.01 (15%)	Annual average relative humidity (%)	-0.21 (0.04)	<0.001	0.24	1.5
	Households using wood heaters (%)	0.13 (0.02)	<0.001	0.46	1.8
	Major roads, 10 km (km)	0.005 (0.001)	<0.001	0.56	2.6

	Burned area, 25 km (%)*	0.11 (0.05)	0.04	0.59	1.3
NOSAT-W	Intercept	19.4 (2.5)	<0.001		
Adj. R^2 (R^2): 0.56 (0.60)	Residential area, 5 km (%)	0.02 (0.01)	0.011	0.17	1.1
RMSE (μg/m^3): 1.05 (15%)	Annual average relative humidity (%)	-0.27 (0.05)	<0.001	0.24	1.8
	Annual 18°C heating degree days (count)	0.002 (0.0003)	<0.001	0.32	1.6
	Elevation (m)	-0.007 (0.002)	<0.001	0.49	1.8
	Commercial area, 1 km (%)	0.05 (0.02)	0.007	0.56	1.1

Diagnostics and sensitivity analyses

The variables in the LUR models were mostly consistent with variables selected via resampling (Figure S3). LUR model residuals exhibited some spatial clustering, but did not show statistically significant evidence of autocorrelation (Table S4) and other assumptions of linear regression were met (e.g., linearity, homoscedasticity, normality of errors). There was no evidence of individual monitoring sites overtly influencing the models. The bootstrap analysis suggested sensitivity of the final LUR model coefficients and R^2 to site selection (Table S5). Coefficients and performance of the year-2015 models appeared particularly sensitive to site selection in four randomly selected prior years, although the small number of sites made this difficult to assess (Table S6). The percentiles of LUR predictor variables at model development and evaluation sites were similar to those at mesh block centroids (Table S7).

The two GWR-adjusted satellite $PM_{2.5}$ estimates were not included in the two final LUR models they were offered to, which resulted in both final models being identical to the NOSAT model.

Independent evaluation

LUR model evaluation results are in Table 2 and descriptive statistics for $PM_{2.5}$ at the 51 evaluation sites are in Table S3. The SAT model had an independent evaluation R^2 (RMSE) of 0.52 (1.15 µg/m^3 [16%]), a decrease of 0.11 from model development. The SAT-W model had an evaluation R^2 (RMSE) of 0.49 (1.24 µg/m^3 [17%]; 0.02 lower than development. The NOSAT model (R^2: 0.21; RMSE: 1.67 µg/m^3 [24%]) exhibited the greatest decrease from development to evaluation (by 0.38), while the decrease of the NOSAT-W model (R^2: 0.43; RMSE: 1.26 µg/m^3 [18%]) was more modest by comparison (by 0.13). Model evaluation plots are in the SI (Figures S4-S7). MSE-R^2 values for all models reflected the same trends observed for evaluation R^2 values, but were lower. On average, most models under-predicted $PM_{2.5}$.

Table 2. Independent evaluation of four LUR models and the global geophysical satellite $PM_{2.5}$ estimates (SAT-$PM_{2.5}$) at 51 measurement sites (n=165 site-years of measurements), measured $PM_{2.5}$ regressed on predicted $PM_{2.5}$, *change from model build adjusted R^2 to independent evaluation R^2, MSE-R^2: mean square error R^2 (^negative MSE-R^2 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/m^3 and as a percentage of mean $PM_{2.5}$ at all evaluation sites). SAT: LUR model offered the SAT-$PM_{2.5}$ variable; SAT-W: same as SAT, but wood heater predictor not offered; NOSAT: LUR model not offered SAT-$PM_{2.5}$ variable; NOSAT-W: same as NOSAT, but wood heater predictor not offered.

Model	**R^2**	**MSE-R^2^**	**R^2 change***	**β (SE)**	**Int.**	**RMSE (μg/m^3)**	**RMSE (%)**	**MB (μg/m^3)**	**FB#**
SAT	0.52	0.45	-0.11	0.73 (0.10)	1.93	1.15	16.3	-0.03	-0.004
SAT-W	0.49	0.37	-0.02	0.67 (0.10)	2.34	1.24	17.4	-0.001	-0.002
NOSAT	0.21	-0.15	-0.38	0.44 (0.12)	4.05	1.67	23.5	-0.09	-0.011
NOSAT-W	0.43	0.34	-0.13	0.80 (0.13)	1.09	1.26	17.8	0.38	0.06
SAT-$PM_{2.5}$	0.26	< -1	n/a	0.66 (0.16)	4.88	3.97	56.0	-3.72	-0.72

Bootstrap analyses indicated some sensitivity of evaluation R^2 values to site selection (Table S8). Evaluation R^2 values remained stable or decreased slightly when comparing year-2015 sites with pre-2015 sites for the three models with a time-varying predictor variable (Table S8).

The performance of the SAT-$PM_{2.5}$ estimates when applied to 51 evaluation sites is in Table 2. Results for the two GWR-adjusted $PM_{2.5}$ estimates are in the SI (Table S9); of the three $PM_{2.5}$ estimates, the highest R^2 observed was for SAT-$PM_{2.5}$ (R^2: 0.26; RMSE: 3.97 μg/m^3 [56%]). The SAT and SAT-W LUR models, which both included the SAT-$PM_{2.5}$ product, increased the evaluation R^2 by 0.26 and 0.23, respectively, and decreased prediction error compared with the SAT-$PM_{2.5}$ product alone.

Predicted $PM_{2.5}$

We estimated a national population-weighted average (± SD) $PM_{2.5}$ concentration of 6.6 (±1.7) μg/m^3 in 2015, using SAT LUR model predictions at ~347,000 census mesh block centroids (Tables S10-S11). Predictions for earlier years ranged from 5.9 (± 1.7) μg/m^3 in 2010 to 7.6 (± 2.2) μg/m^3 in 2003 and were mostly stable over 2000–2015. Measured $PM_{2.5}$ decreased slightly in some locations and increased slightly in others (Figure S8). Population-weighted average $PM_{2.5}$ predictions from the other models were similar to SAT (Table S10). SAT-$PM_{2.5}$ estimates (year-2015) had a population-weighted average of 3.3 (±0.8) μg/m^3.

Figure 1 shows the distribution of mesh block $PM_{2.5}$ estimates from the SAT model for the year it was developed (2015); results for other LUR models are in the SI and were mostly similar (Figures S9-S11). The spatial distribution of estimated $PM_{2.5}$ was more variable. Figure 2 shows the SAT model prediction map (other models in Figures S12-S14). In general, models predicted relatively high $PM_{2.5}$ concentrations in the sparsely populated central and north-western regions, consistent with the geophysical SAT-$PM_{2.5}$ estimates, and also in urban areas.

We estimated that Australian Bureau of Statistics (ABS)-defined significant urban areas (i.e., population >10,000)[53] accounted for 86.4% of national population-weighted $PM_{2.5}$ exposure in 2015. Figure 2 shows $PM_{2.5}$ gradients across the two most-populous capital cities, Sydney (~4.8 million) and Melbourne (~4.5 million), as well as the least-populous capital city, Darwin (~135,000). This figure highlights the enhanced within-city variability in SAT LUR-estimated $PM_{2.5}$, compared with SAT-$PM_{2.5}$ estimates used alone, in a sample of Australian urban areas. The enhanced concentration gradients were due to the LUR predictions and not an artefact of kriging (Figures S15-S16). Predictions from the SAT and SAT-W models had very good correlation over 2000–2015 (r = 0.84, ρ = 0.84), while the SAT model had moderately good correlation with the NOSAT model (r = 0.68, ρ = 0.67, Table S12). $PM_{2.5}$ estimates from all models had poor to fair correlation with NO_2 estimates at mesh block level (r from 0.09 to 0.50, ρ from 0.13 to 0.47).

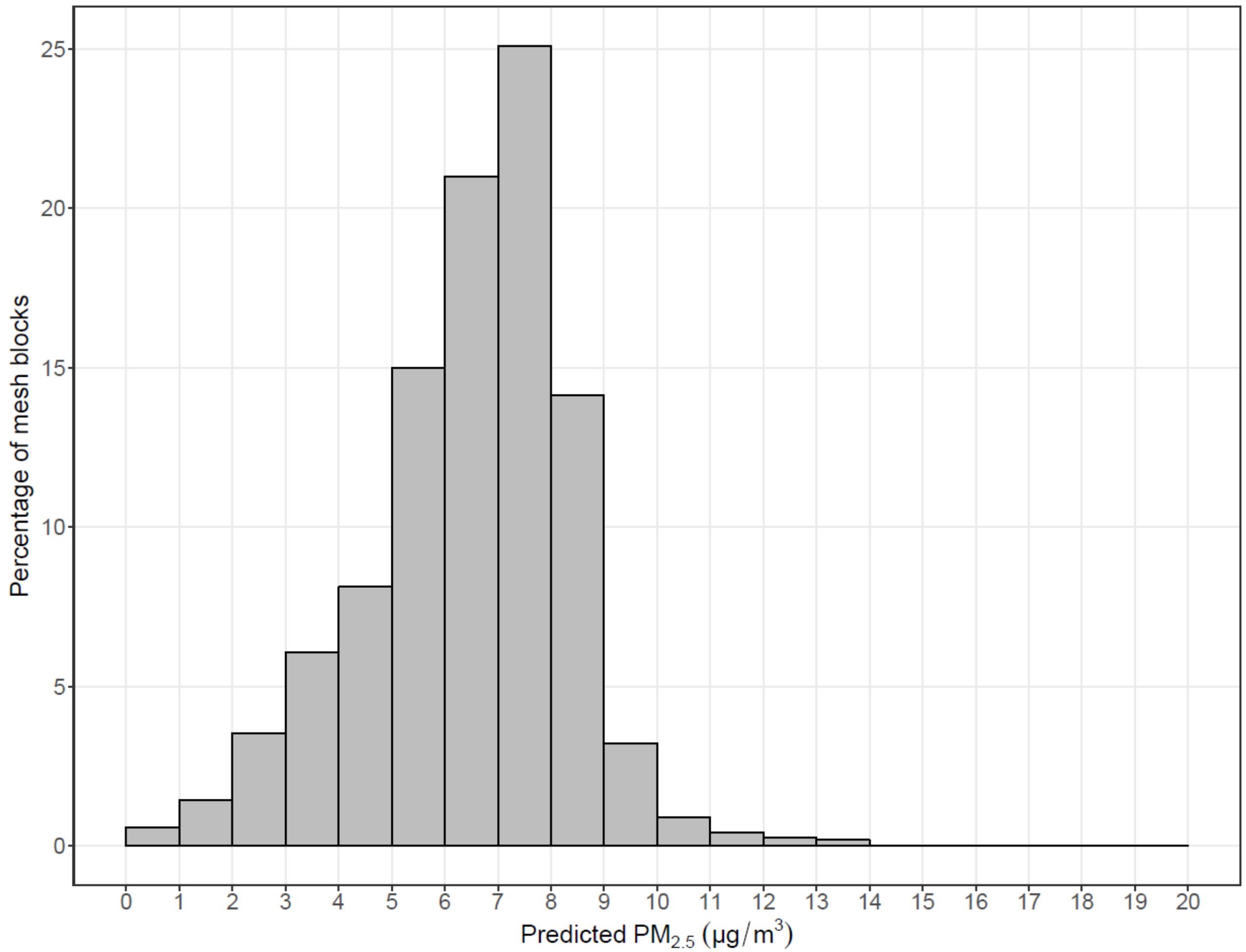


Figure 1. The distribution of annual mean $PM_{2.5}$ (year-2015) predicted by the SAT model at ~347,000 census mesh block centroids that cover Australia. Mesh blocks have an average (± SD) population of 62 people. Mesh block boundaries are selected to yield about 30 to 60 dwellings or no dwellings at all (25% of mesh blocks have a population of zero).[53]

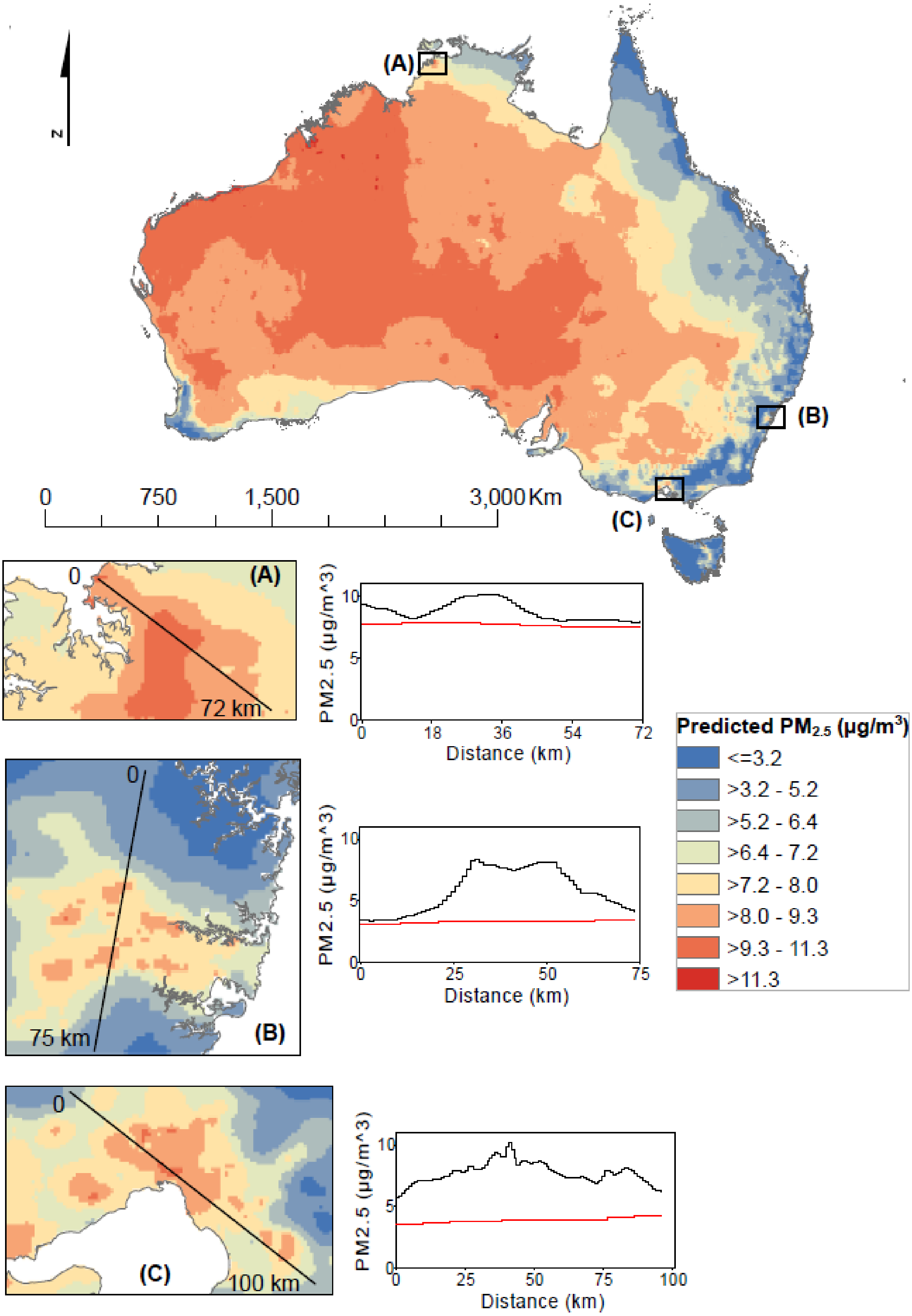

N
(A)
(B)
(C)
0
750
1,500
3,000Km
0
72 km
PM2.5 (µg/m^3)
0
5
10
0
18
36
54
72
Distance (km)
0
75 km
0
25
50
75
Distance (km)
0
100 km
0
25
50
75
100
Distance (km)
Predicted PM2.5 (µg/m3)
<=3.2
>3.2 - 5.2
>5.2 - 6.4
>6.4 - 7.2
>7.2 - 8.0
>8.0 - 9.3
>9.3 - 11.3
>11.3

Figure 2. Annual mean $PM_{2.5}$ (year-2015) predicted by the SAT LUR model (displayed on a 1 × 1 km grid). Insets A (Darwin; pop. ~135,000), B (Sydney; ~4.8 million) and C (Melbourne; ~4.5 million) show a selection of Australian capital cities. Inset graphs show SAT LUR model predictions (black line) and geophysical ~10 km SAT-$PM_{2.5}$ estimates (red line) across transects within each city. Note: Darwin (tropical monsoon climate) has few sources of traffic or industrial emissions but is frequently affected by landscape fires during the 'dry' season (April through October).

DISCUSSION

Overview

We used publicly available data to develop continental-scale Australian LUR models for annual average $PM_{2.5}$ concentrations (year-2015), including two models with a satellite-based $PM_{2.5}$ predictor variable. The LUR model development R^2 values ranged from 0.51 to 0.63. We evaluated these models with independent data (51 sites, years-2000–2015) and used them to estimate the spatial distribution of $PM_{2.5}$ at ~347,000 census mesh blocks. Based on the independent evaluation, we concluded that the two models including the satellite-based $PM_{2.5}$ predictor had the greatest robustness; R^2 of 0.52 (RMSE: 1.15 μg/m^3 [16%]) and 0.49 (RMSE: 1.24 μg/m^3 [17%]), respectively.

To our knowledge, these are the first national $PM_{2.5}$ models for Australia. Our findings add to the growing international literature showing national- and continental-scale $PM_{2.5}$ exposure models, both with and without satellite data.[16-23, 56-58]

Comparison to other studies

Our study used a relatively small number of ground monitors for LUR development (n=49) compared with other satellite-based $PM_{2.5}$ modelling studies using regulatory data performed at national- or continental-scale (from 177 to 1,928 monitors). [16-23] The independent evaluation R^2 of our best model (0.52) was consistent with results from the two studies most methodologically comparable to ours: de Hoogh et al.[18] in Western Europe (evaluation R^2: 0.51 to 0.58), and Hystad et al.[17] in Canada (development R^2: 0.46).

One explanation for these findings is that a relatively small number of sites can yield valid LUR models, at least in Australia, provided they capture a range of $PM_{2.5}$ and predictor variable values that represent the locations to which the models are applied.[5,50,59,60] The consistency between $PM_{2.5}$ predictors at model development sites and mesh blocks centroids (Table S7) suggests that the sites were appropriate for estimating population exposures to $PM_{2.5}$, although predictions at the extremes (<5th and >95th $PM_{2.5}$ percentiles) are less reliable due to out-of-sample values.[50] We did observe sensitivity of LUR coefficients and R^2 to site selection in the bootstrap analysis, in keeping with the Canadian study.[17] We will use this information to estimate uncertainty when assigning $PM_{2.5}$ exposures, and encourage other users to consider this uncertainty in their analyses.

The role of satellite $PM_{2.5}$ estimates in LUR models

Our results indicate that SAT-$PM_{2.5}$ estimates improve continental LUR models for Australia, but also that the LUR method can enhance the national utility of a global SAT-$PM_{2.5}$ product. Incorporating SAT-$PM_{2.5}$ in an LUR model yielded a doubling, compared with SAT-$PM_{2.5}$ used alone, in the evaluation R^2 (0.52 *vs.* 0.26, respectively) and population-weighted average $PM_{2.5}$ (6.6 [±1.7] *vs.* 3.3 [±0.8] μg/m^3, respectively). We also observed markedly enhanced within-city spatial variability in LUR-predicted $PM_{2.5}$ compared with SAT-$PM_{2.5}$ used alone (Figure 2), which highlights the potential of the LUR predictor variables to capture additional spatial variability in urban $PM_{2.5}$ concentrations.[18,38] Our results demonstrate additional empirical evidence to support the incorporation of satellite observations of $PM_{2.5}$ in large-scale

LUR models. One possible implication of these collective results for epidemiological studies was highlighted by Jerrett et al.,[4] who assessed seven exposure models and showed that those using both satellite- and ground-based $PM_{2.5}$ information generally had larger relative risks for mortality than satellite estimates alone.

The inclusion of geophysical satellite-derived estimates of $PM_{2.5}$ from a global 0.1° resolution product (SAT-$PM_{2.5}$)[29] contributed 0.10 to the adjusted R^2 of LUR models. Unlike most previous studies, the SAT-$PM_{2.5}$ estimate was not the largest contributor to LUR model R^2 in this study.[16-18,38] However, the SAT and SAT-W models had a smaller decrease from model R^2 to evaluation R^2 than the corresponding models without SAT-$PM_{2.5}$; a difference of 0.31 and 0.06, respectively. This finding suggests less overfitting and greater robustness of models that included geophysical SAT-$PM_{2.5}$ estimates. The role of selected other LUR predictor variables is described in the SI (page S43).

An attractive feature of the SAT-$PM_{2.5}$ product is its ability to capture time-varying regional events that affect annual $PM_{2.5}$ across Australia (e.g., landscape fires). This is a potential explanation for the greater robustness observed in the two models that included SAT-$PM_{2.5}$. SAT-W is parsimonious (four predictors), showed the smallest decrease in R^2 from development to evaluation, and has relatively low prediction error. The SAT model performed best at model development, had comparable evaluation statistics to SAT-W and required seven predictors, one of which was a coarsely-resolved wood heater variable. Wood heater emissions can contribute up to 60-80% of ambient $PM_{2.5}$ mass during winter in urban and rural areas in southern Australia.[24,32-35] However, despite this empirical basis for its inclusion, it is possible that the wood heater variable is a proxy for capital cities *vs*. other locations in each state, or some other source of variability in $PM_{2.5}$. Moreover, error associated with estimating LUR predictors for exposure assessment can attenuate effects in epidemiological studies.[46,47,52] Models with fewer predictors that offer comparable performance may be preferable, provided they can be accurately estimated for the target population. We plan to use both the SAT and SAT-W models in epidemiological studies.

Utility of $PM_{2.5}$ models in epidemiological studies

The measured (± SD) annual mean $PM_{2.5}$ in 2015 was 6.9 (± 1.6) μg/m^3; our population-weighted estimates were also relatively low (6.6 ± 1.7 μg/m^3). However, similar levels in a recent Canadian cohort study (6.3 ± 2.5 μg/m^3) were associated with comparable or larger effects on non-accidental mortality per 10 μg/m^3 increase than those reported from countries with higher $PM_{2.5}$ concentrations, and no lower threshold was observed.[61] That finding highlights the utility of national $PM_{2.5}$ exposure models for cohorts in countries like Australia, which are relatively unique in their suitability for assessing long-term exposure to low concentrations, a topic for which empirical evidence is limited.[62,63]

Predicted $PM_{2.5}$ and NO_2 generally had low levels of correlation at mesh block centroids. The two pollutants share some common sources, but have different spatio-temporal patterns; no single widely-monitored air pollutant is an ideal surrogate for assessing exposure to traffic and other anthropogenic emissions.[64-66] These findings suggest the utility of pollutant-specific exposure assessment models.

Developing more finely resolved exposure models requires consideration of how they will be applied and the spatial variability of the pollutant being estimated. For example, $PM_{2.5}$ estimates on a 100 m or 1 km grid may, in some cases, be better suited to studies where both the home and work (or school) addresses are known and population mobility can be accounted for. By comparison, a coarser grid (e.g., 5 or 10 km) could include home and work in one grid cell for some individuals, regardless of whether the work address is known.[67-68]

Other applications of $PM_{2.5}$ models

Measured annual $PM_{2.5}$ has decreased modestly in some Australian cities since 2000, while levels in other cities have remained stable or increased slightly (Figure S8).[27,69] Because legislation and monitoring for $PM_{2.5}$ in Australia is largely focused on urban background locations,[26] regulatory monitoring may not capture potentially important spatial patterns in $PM_{2.5}$. For example, we estimated that ~17% (~3.73 million people) of the Australian population live in a mesh block with $PM_{2.5}$ above the annual standard of 8 $\mu g/m^3$ (SAT model, year-2015). They accounted for ~23% of national population-weighted exposure, of which ~4.2% was incurred by people living outside an ABS-defined significant urban areas (i.e., living in an area with a population <10,000).[53] Notwithstanding model prediction error, this simple example illustrates how a national exposure model, linked to routine census data, can identify communities that may require specific air quality investigations.

Limitations

A limitation of this study is the use of a single year. Because models include the time-varying SAT-$PM_{2.5}$ predictor, we sought to predict earlier years by assuming that year-2015 coefficients are applicable. Our bootstrap analysis, while limited by small site numbers, did not support this assumption. However, our sensitivity analysis using year-2015 and pre-2015 evaluation sites showed only minor differences in evaluation performance (Table S8). These mixed findings mean it is possible that model robustness pre-2015 may be reduced in some locations. Other methods could be explored, such as year-specific LUR models, back-extrapolation of year-2015 predictions using historical monitoring data, and/or inclusion of CTM simulations for Australian cities as an additional source of area-specific $PM_{2.5}$ information.[18,54,70,71]

The data used for independent evaluation were selected carefully, with criteria emphasizing quality and consistency with measurement methods used at model development sites. However, the identification of sites was driven by availability and accessibility of data; this approach could potentially bias our model evaluation results in either direction.[48] Because evaluation sites were located in urban areas (>10,000 people), results here might not inform model validity outside that context. The sites were comparable to model development sites and mesh block centroids in terms of LUR predictor distribution and $PM_{2.5}$ concentration (Table S7), supporting their use for evaluation. We plan to undertake ongoing evaluation as new data becomes available.

ACKNOWLEDGEMENTS

The predictions from our LUR models are freely available; please contact the corresponding author. We are grateful to all the agencies that supplied data used in this project. We thank Mr Solomon Woldeyohannes for his advice regarding the figures. L.D.K. acknowledges financial support from the Centre for Air Quality and Health Research and Evaluation, an NHMRC Centre of Research Excellence (APP1030259). G.P. is supported by NHMRC project grants (GNT1099655, GNT1141510). This article was developed in part under Assistance Agreement no. RD83587301 awarded by the US Environmental Protection Agency (EPA). This article has not been formally reviewed by the US EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the agency. The US EPA does not endorse any products or commercial services mentioned in this publication.

Supporting Information. $PM_{2.5}$ measurement methods and QA, LUR predictor selection and data sources, model development site location map, independent evaluation data sources and methods, additional results (descriptive statistics and figures for measured $PM_{2.5}$, QA results, evaluation tables and plots, LUR model predictions and maps, mesh blocks *vs*. kriging, correlation between $PM_{2.5}$ and NO_2), additional discussion of predictors in final LUR models.

REFERENCES

1. GBD 2016 Risk Factors Collaborators. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* **2017**, *390*, (10100), 1345-1422.
2. Paciorek, C. J.; Liu, Y., Limitations of Remotely Sensed Aerosol as a Spatial Proxy for Fine Particulate Matter. *Environ. Health Perspect*. **2009**, *117*, (6), 904-909.
3. van Donkelaar, A.; Martin, R. V.; Brauer, M.; Kahn, R.; Levy, R.; Verduzco, C.; Villeneuve, P. J., Global Estimates of Ambient Fine Particulate Matter Concentrations from Satellite-Based Aerosol Optical Depth: Development and Application. *Environ. Health Perspect.* **2010**, *118*, (6), 847-855.
4. Jerrett, M.; Turner, M. C.; Beckerman, B. S.; Pope, C. A.; van Donkelaar, A.; Martin, R. V.; Serre, M.; Crouse, D.; Gapstur, S. M.; Krewski, D.; Diver, W. R.; Coogan, P. F.; Thurston, G. D.; Burnett, R. T., Comparing the Health Effects of Ambient Particulate Matter Estimated Using Ground-Based versus Remote Sensing Exposure Estimates. *Environ. Health Perspect.* **2017**, *125*, (4), 552-559.
5. Hoek, G., Methods for Assessing Long-Term Exposures to Outdoor Air Pollutants. *Curr. Env. Health Rep*. **2017**, *4*, (4), 450-462.
6. Kloog, I.; Koutrakis, P.; Coull, B. A.; Lee, H. J.; Schwartz, J., Assessing temporally and spatially resolved PM2.5 exposures for epidemiological studies using satellite aerosol optical depth measurements. *Atmos. Environ*. **2011**, *45*, (35), 6267-6275.
7. Kloog, I.; Nordio, F.; Coull, B. A.; Schwartz, J., Incorporating Local Land Use Regression And Satellite Aerosol Optical Depth In A Hybrid Model Of Spatiotemporal PM2.5 Exposures In The Mid-Atlantic States. *Environ. Sci. Technol*. **2012**, *46*, (21), 11913-11921.
8. Brauer, M.; Freedman, G.; Frostad, J.; van Donkelaar, A.; Martin, R. V.; Dentener, F.; Dingenen, R. v.; Estep, K.; Amini, H.; Apte, J. S.; Balakrishnan, K.; Barregard, L.; Broday,

D.; Feigin, V.; Ghosh, S.; Hopke, P. K.; Knibbs, L. D.; Kokubo, Y.; Liu, Y.; Ma, S.; Morawska, L.; Sangrador, J. L. T.; Shaddick, G.; Anderson, H. R.; Vos, T.; Forouzanfar, M. H.; Burnett, R. T.; Cohen, A., Ambient Air Pollution Exposure Estimation for the Global Burden of Disease 2013. *Environ. Sci. Technol*. **2016**, *50,* (1), 79-88.

9. Larkin, A.; Geddes, J. A.; Martin, R. V.; Xiao, Q.; Liu, Y.; Marshall, J. D.; Brauer, M.; Hystad, P., Global Land Use Regression Model for Nitrogen Dioxide Air Pollution. *Environ. Sci. Technol*. **2017**, *51*, (12), 6957-6964.

10. Shaddick, G.; Thomas, M. L.; Green, A.; Brauer, M.; Donkelaar, A.; Burnett, R.; Chang, H. H.; Cohen, A.; Dingenen, R. V.; Dora, C.; Gumy, S.; Liu, Y.; Martin, R.; Waller, L. A.; West, J.; Zidek, J. V.; Prüss-Ustün, A., Data integration model for air quality: a hierarchical approach to the global estimation of exposures to ambient air pollution. *J. R. Stat. Soc. Ser. C Appl. Stat*. **2018**, *67*, (1), 231-253.

11. Crouse, D. L.; Peters, P. A.; van Donkelaar, A.; Goldberg, M. S.; Villeneuve, P. J.; Brion, O.; Khan, S.; Atari, D. O.; Jerrett, M.; Pope, C. A.; Brauer, M.; Brook, J. R.; Martin, R. V.; Stieb, D.; Burnett, R. T., Risk of nonaccidental and cardiovascular mortality in relation to long-term exposure to low concentrations of fine particulate matter: a Canadian national-level cohort study. *Environ. Health Perspect.* **2012**, *120,* (5), 708-714.

12. Kloog, I.; Ridgway, B.; Koutrakis, P.; Coull, B. A.; Schwartz, J. D., Long- and short-term exposure to PM2.5 and mortality: using novel exposure models. *Epidemiology*. **2013**, *24*, (4), 555-561.

13. Bechle, M. J.; Millet, D. B.; Marshall, J. D., Remote sensing of exposure to NO2: Satellite versus ground-based measurement in a large urban area. *Atmos. Environ*. **2013**, *69,* (0), 345-353.

14. Lyapustin, A.; Wang, Y.; Laszlo, I.; Kahn, R.; Korkin, S.; Remer, L.; Levy, R.; Reid, J. S., Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm. *J. Geophys. Res. Atmos*. **2011**, *116,* D03211.

15. Apte, J. S.; Bombrun, E.; Marshall, J. D.; Nazaroff, W. W., Global Intraurban Intake Fractions for Primary Air Pollutants from Vehicles and Other Distributed Sources. *Environ. Sci. Technol*. **2012**, *46*, (6), 3415-3423.

16. Beckerman, B. S.; Jerrett, M.; Serre, M.; Martin, R. V.; Lee, S.-J.; van Donkelaar, A.; Ross, Z.; Su, J.; Burnett, R. T., A Hybrid Approach to Estimating National Scale Spatiotemporal Variability of PM2.5 in the Contiguous United States. *Environ. Sci. Technol*. **2013**, *47*, (13), 7233-7241.

17. Hystad, P.; Setton, E.; Cervantes, A.; Poplawski, K.; Deschenes, S.; Brauer, M.; van Donkelaar, A.; Lamsal, L.; Martin, R.; Jerrett, M.; Demers, P., Creating National Air Pollution Models for Population Exposure Assessment in Canada. *Environ. Health Perspect.* **2011**, *119*, (8), 1123-1129.

18. de Hoogh, K.; Gulliver, J.; Donkelaar, A. V.; Martin, R. V.; Marshall, J. D.; Bechle, M. J.; Cesaroni, G.; Pradas, M. C.; Dedele, A.; Eeftens, M.; Forsberg, B.; Galassi, C.; Heinrich, J.; Hoffmann, B.; Jacquemin, B.; Katsouyanni, K.; Korek, M.; Kunzli, N.; Lindley, S. J.; Lepeule, J.; Meleux, F.; de Nazelle, A.; Nieuwenhuijsen, M.; Nystad, W.; Raaschou-Nielsen, O.; Peters, A.; Peuch, V. H.; Rouil, L.; Udvardy, O.; Slama, R.; Stempfelet, M.; Stephanou, E. G.; Tsai, M. Y.; Yli-Tuomi, T.; Weinmayr, G.; Brunekreef, B.; Vienneau, D.; Hoek, G., Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. *Environ. Res*. **2016**, *151*, 1-10.

19. Ma, Z.; Hu, X.; Huang, L.; Bi, J.; Liu, Y., Estimating Ground-Level PM2.5 in China Using Satellite Remote Sensing. *Environ. Sci. Technol*. **2014**, *48*, (13), 7436-7444.

20. Ma, Z.; Hu, X.; Sayer, A. M.; Levy, R.; Zhang, Q.; Xue, Y.; Tong, S.; Bi, J.; Huang, L.; Liu, Y., Satellite-Based Spatiotemporal Trends in PM2.5 Concentrations: China, 2004-2013. *Environ. Health Perspect.* **2016**, *124*, (2), 184-192.

21. Chen, G.; Li, S.; Knibbs, L. D.; Hamm, N. A. S.; Cao, W.; Li, T.; Guo, J.; Ren, H.; Abramson, M. J.; Guo, Y., A machine learning method to estimate PM2.5 concentrations

across China with remote sensing, meteorological and land use information. *Sci. Total Environ.* **2018**, *636*, 52-60.

22. Lee, S. J.; Serre, M. L.; van Donkelaar, A.; Martin, R. V.; Burnett, R. T.; Jerrett, M., Comparison of geostatistical interpolation and remote sensing techniques for estimating long-term exposure to ambient PM2.5 concentrations across the continental United States. *Environ. Health Perspect.* **2012**, *120*, (12), 1727-32.
23. Di, Q.; Kloog, I.; Koutrakis, P.; Lyapustin, A.; Wang, Y.; Schwartz, J., Assessing PM2.5 Exposures with High Spatiotemporal Resolution across the Continental United States. *Environ. Sci. Technol*. **2016**, *50*, (9), 4712-21.
24. 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.
25. Karagulian, F.; Belis, C.A.; Dora, C.F.C.; Prüss-Ustün, A.M.; Bonjour, S.; Adair-Rohani, H.; Amann, M., Contributions to cities' ambient particulate matter (PM): A systematic review of local source contributions at global level. *Atmos. Environ*. **2015**, *120*, 475-483.
26. National Environment Protection (Ambient Air Quality) Measure, 2016. https://www.legislation.gov.au/Details/F2016C00215 (accessed 30 April 2018).
27. National Environment Protection Council, 2014. Draft Variation to the National Environment protection (Ambient Air Quality) Measure Impact Statement. http://www.environment.gov.au/protection/nepc/nepms/ambient-air-quality/variation-2014/impact-statement (accessed 10 April 2018).
28. Ayers, G. P.; Keywood, M. D.; Gras, J. L., TEOM vs. manual gravimetric methods for determination of PM2.5 aerosol mass concentrations. *Atmos. Environ.* **1999**, *33*, (22), 3717-3721.

29. 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. *Environ. Sci. Technol*. **2016**, *50*, (7), 3762-3772.

30. Atmospheric Composition Analysis Group, Dalhousie University http://fizz.phys.dal.ca/~atmos/martin/?page_id=140 (accessed 30 April 2018).

31. Knibbs, L. D.; Hewson, M. G.; Bechle, M. J.; Marshall, J. D.; Barnett, A. G., A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environ. Res*. **2014**, *135*, 204-211.

32. 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).

33. Cope, M, Keywood, M, Emmerson, K, Galbally, I, et al. Sydney Particle Study- Stage-II. The Centre for Australian Weather and Climate Research, 2014. http://www.environment.nsw.gov.au/~/media/82F9E2CEF1F54FA2B0AC65EA6FC4D542.ashx (accessed 30 April 2018)

34. 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).

35. 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).

36. Friend, A. J.; Ayoko, G. A.; Stelcer, E.; Cohen, D., Source apportionment of PM2.5 at two receptor sites in Brisbane, Australia. *Environ. Chem*. **2011**, *8*, (6), 569-580.

37. Hart, J. E.; Yanosky, J. D.; Puett, R. C.; Ryan, L.; Dockery, D. W.; Smith, T. J.; Garshick, E.; Laden, F., Spatial modeling of PM10 and NO2 in the continental United States, 1985-2000. *Environ. Health Perspect*. **2009**, *117*, (11), 1690-1696.

38. Vienneau, D.; de Hoogh, K.; Bechle, M. J.; Beelen, R.; van Donkelaar, A.; Martin, R. V.; Millet, D. B.; Hoek, G.; Marshall, J. D., Western European Land Use Regression Incorporating Satellite- and Ground-Based Measurements of NO2 and PM10. *Environ. Sci. Technol*. **2013**, *47*, (23), 13555-13564.

39. Johnston, F. H.; Hanigan, I. C.; Henderson, S. B.; Morgan, G. G.; Portner, T.; Williamson, G. J.; Bowman, D. M. J. S., Creating an Integrated Historical Record of Extreme Particulate Air Pollution Events in Australian Cities from 1994 to 2007. *J. Air Waste Manag. Assoc*. **2011**, *61*, (4), 390-398.

40. Price, O. F.; Williamson, G. J.; Henderson, S. B.; Johnston, F.; Bowman, D. M. J. S., The Relationship between Particulate Pollution Levels in Australian Cities, Meteorology, and Landscape Fire Activity Detected from MODIS Hotspots. *PLoS ONE* **2012**, *7*, (10), e47327.

41. 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.

42. ESCAPE Study Manual, ENV.2007.1.2.2.2. European cohort on air pollution, 2008. http://www.escapeproject.eu/manuals/ESCAPEStudy-manual_x007E_final.pdf (accessed July 2016)

43. Eeftens, M.; Beelen, R.; de Hoogh, K.; Bellander, T.; Cesaroni, G.; Cirach, M.; Declercq, C.; Dėdelė, A.; Dons, E.; de Nazelle, A.; Dimakopoulou, K.; Eriksen, K.; Falq, G.; Fischer, P.;

Galassi, C.; Gražulevičienė, R.; Heinrich, J.; Hoffmann, B.; Jerrett, M.; Keidel, D.; Korek, M.; Lanki, T.; Lindley, S.; Madsen, C.; Mölter, A.; Nádor, G.; Nieuwenhuijsen, M.; Nonnemacher, M.; Pedeli, X.; Raaschou-Nielsen, O.; Patelarou, E.; Quass, U.; Ranzi, A.; Schindler, C.; Stempfelet, M.; Stephanou, E.; Sugiri, D.; Tsai, M.-Y.; Yli-Tuomi, T.; Varró, M. J.; Vienneau, D.; Klot, S. v.; Wolf, K.; Brunekreef, B.; Hoek, G., Development of Land Use Regression Models for PM2.5, PM2.5 Absorbance, PM10 and PMcoarse in 20 European Study Areas; Results of the ESCAPE Project. *Environ. Sci. Technol*. **2012**, *46*, (20), 11195-11205.

44. Vienneau, D.; de Hoogh, K.; Beelen, R.; Fischer, P.; Hoek, G.; Briggs, D., Comparison of land-use regression models between Great Britain and the Netherlands. *Atmos. Environ.* **2010**, *44*, (5), 688-696.

45. Canty, A, Ripley, B. (2017). boot: Bootstrap R (S-Plus) Functions. R package version 1.3. https://cran.r-project.org/web/packages/boot/index.html (accessed July 25, 2018)

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

47. Basagana, X.; Aguilera, I.; Rivera, M.; Agis, D.; Foraster, M.; Marrugat, J.; Elosua, R.; Kunzli, N., Measurement error in epidemiologic studies of air pollution based on land-use regression models. *Am. J. Epidemiol*. **2013**, *178*, (8), 1342-1346.

48. Hoek, G, Beelen, R, Brunekreef B. Land use regression models for outdoor air pollution. In: Nieuwenhuijsen MJ (ed.) *Exposure assessment in environmental epidemiology* (2nd ed), Oxford University Press, UK, 2015, pp. 271-294.

49. Johnson, M.; Isakov, V.; Touma, J. S.; Mukerjee, S.; Özkaynak, H., Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmos. Environ.* **2010**, *44*, (30), 3660-3668.

50. Wang, M.; Beelen, R.; Eeftens, M.; Meliefste, K.; Hoek, G.; Brunekreef, B., Systematic Evaluation of Land Use Regression Models for NO2. *Environ. Sci. Technol*. **2012**, *46*, (8), 4481-4489.

51. Gulliver, J.; de Hoogh, K.; Hansell, A.; Vienneau, D., Development and Back-Extrapolation of NO2 Land Use Regression Models for Historic Exposure Assessment in Great Britain. *Environ. Sci. Technol*. **2013**, *47*, (14), 7804-7811.

52. Szpiro, A. A.; Paciorek, C. J.; Sheppard, L., Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology.* **2011**, *22*, (5), 680-685.

53. 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)

54. Gulliver, J.; de Hoogh, K.; Hoek, G.; Vienneau, D.; Fecht, D.; Hansell, A., Back-extrapolated and year-specific NO2 land use regression models for Great Britain - Do they yield different exposure assessment? *Environ. Int.* **2016**, *92-93*, 202-209.

55. Australian Bureau of Statistics, 2011 Energy Use and Conservation Survey, http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4602.0.55.001Mar%202011?OpenDocument (accessed 30 April 2018).

56. Kim, S. Y.; Olives, C.; Sheppard, L.; Sampson, P. D.; Larson, T. V.; Keller, J. P.; Kaufman, J. D., Historical Prediction Modeling Approach for Estimating Long-Term Concentrations of PM2.5 in Cohort Studies before the 1999 Implementation of Widespread Monitoring. *Environ. Health Perspect.* **2017**, *125*, (1), 38-46.

57. Yanosky, J. D.; Paciorek, C. J.; Laden, F.; Hart, J. E.; Puett, R. C.; Liao, D.; Suh, H. H., Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. *Environ. Health*, **2014**, *13*, (1), 63.

58. de Hoogh, K.; Heritier, H.; Stafoggia, M.; Kunzli, N.; Kloog, I., Modelling daily PM2.5 concentrations at high spatio-temporal resolution across Switzerland. *Environ. Pollut*. **2018**, *233*, 1147-1154.

59. Bechle, M. J.; Millet, D. B.; Marshall, J. D., National Spatiotemporal Exposure Surface for NO2: Monthly Scaling of a Satellite-Derived Land-Use Regression, 2000–2010. *Environ. Sci. Technol*. **2015**, *49*, (20), 12297-12305.

60. Novotny, E. V.; Bechle, M. J.; Millet, D. B.; Marshall, J. D., National Satellite-Based Land-Use Regression: NO2 in the United States. *Environ. Sci. Technol*. **2011**, *45*, (10), 4407-4414.

61. Pinault, L.; Tjepkema, M.; Crouse, D. L.; Weichenthal, S.; van Donkelaar, A.; Martin, R. V.; Brauer, M.; Chen, H.; Burnett, R. T., Risk estimates of mortality attributed to low concentrations of ambient fine particulate matter in the Canadian community health survey cohort. *Environ. Health*. **2016**, 15, 18.

62. Cohen, A. J.; Brauer, M.; Burnett, R.; Anderson, H. R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; Feigin, V.; Freedman, G.; Hubbell, B.; Jobling, A.; Kan, H.; Knibbs, L.; Liu, Y.; Martin, R.; Morawska, L.; Pope, C. A., 3rd; Shin, H.; Straif, K.; Shaddick, G.; Thomas, M.; van Dingenen, R.; van Donkelaar, A.; Vos, T.; Murray, C. J. L.; Forouzanfar, M. H., Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study 2015. *Lancet*. **2017**, *389*, (10082), 1907-1918.

63. Burnett, R. T.; Pope, C. A.; Ezzati, M.; Olives, C.; Lim, S. S.; Mehta, S.; Shin, H. H.; Singh, G.; Hubbell, B.; Brauer, M.; Anderson, H. R.; Smith, K. R.; Balmes, J. R.; Bruce, N. G.; Kan, H.; Laden, F.; Prüss-Ustün, A.; Turner, M. C.; Gapstur, S. M.; Diver, W. R.; Cohen, A., An Integrated Risk Function for Estimating the Global Burden of Disease Attributable to Ambient Fine Particulate Matter Exposure. *Environ. Health Perspect*. **2014**, *122*, (4), 397-403.

64. Health Effects Institute, 2010. Traffic-Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. Special Report 17. https://www.healtheffects.org/publication/traffic-related-air-pollution-critical-review-literature-emissions-exposure-and-health (accessed Feb 12 2018)

65. Jerrett, M.; Burnett, R. T.; Beckerman, B. S.; Turner, M. C.; Krewski, D.; Thurston, G.; Martin, R. V.; Donkelaar, A. v.; Hughes, E.; Shi, Y.; Gapstur, S. M.; Thun, M. J.; III, C. A. P., Spatial Analysis of Air Pollution and Mortality in California. *Am. J. Respir. Crit. Care Med*. **2013**, *188*, (5), 593-599.

66. Crouse, D. L.; Peters, P. A.; Hystad, P.; Brook, J. R.; van Donkelaar, A.; Martin, R. V.; Villeneuve, P. J.; Jerrett, M.; Goldberg, M. S.; Pope, C. A., 3rd; Brauer, M.; Brook, R. D.; Robichaud, A.; Menard, R.; Burnett, R. T., Ambient PM2.5, O(3), and NO(2) Exposures and Associations with Mortality over 16 Years of Follow-Up in the Canadian Census Health and Environment Cohort (CanCHEC). *Environ. Health Perspect*. **2015**, *123*, (11), 1180-1186.

67. Setton, E.; Marshall, J. D.; Brauer, M.; Lundquist, K. R.; Hystad, P.; Keller, P.; Cloutier-Fisher, D., The impact of daily mobility on exposure to traffic-related air pollution and health effect estimates. *J. Expo. Sci. Env. Epidemiol*. **2011**, *21*, (1), 42-48.

68. Ragettli, M. S.; Phuleria, H. C.; Tsai, M. Y.; Schindler, C.; de Nazelle, A.; Ducret-Stich, R. E.; Ineichen, A.; Perez, L.; Braun-Fahrlander, C.; Probst-Hensch, N.; Kunzli, N., The relevance of commuter and work/school exposure in an epidemiological study on traffic-related air pollution. *J. Expo. Sci. Env. Epidemiol*. **2015**, 25, (5), 474-481.

69. Australian Government, Department of the Environment and Energy. Australia: State of the Environment, 2016. https://soe.environment.gov.au/theme/ambient-air-quality/topic/2016/fine-particulate-matter-pm25 (accessed 30 April 2018)

70. Hystad, P.; Demers, P. A.; Johnson, K. C.; Brook, J.; van Donkelaar, A.; Lamsal, L.; Martin, R.; Brauer, M., Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. *Environ. Health* **2012**, 11, 22.

71. Knibbs, L. D.; Coorey, C. P.; Bechle, M. J.; Marshall, J. D.; Hewson, M. G.; Jalaludin, B.; Morgan, G. G.; Barnett, A. G., Long-term nitrogen dioxide exposure assessment using back-extrapolation of satellite-based land-use regression models for Australia. *Environ. Res*. **2018**, *163*, 16-25.