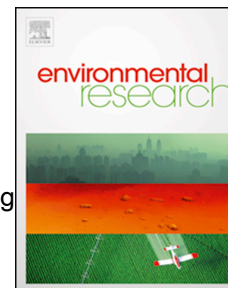# Accepted Manuscript

Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO_2 for a birth cohort in Sydney, Australia

Christine T. Cowie, Frances Garden, Edward Jegasothy, Luke D. Knibbs, Ivan Hanigan, David Morley, Anna Hansell, Gerard Hoek, Guy B. Marks

Please cite this article as: Cowie, C.T., Garden, F., Jegasothy, E., Knibbs, L.D., Hanigan, I., Morley, D., Hansell, A., Hoek, G., Marks, G.B., Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO_2 for a birth cohort in Sydney, Australia, *Environmental Research* (2019), doi: https://doi.org/10.1016/j.envres.2019.03.068.

Title:

Comparison of model estimates from an intra-city land use regression model with a national satellite-LUR and a regional Bayesian Maximum Entropy model, in estimating NO$_2$ for a birth cohort in Sydney, Australia

Authors:

Christine T Cowie [1,2,3,4], Frances Garden[1,2], Edward Jegasothy[5], Luke D Knibbs[3,6], Ivan Hanigan[3,7], David Morley[8], Anna Hansell[8], Gerard Hoek[9], Guy B Marks[1,2,3,4]

Affiliations:

1. South West Sydney Clinical School, University of New South Wales, Sydney, Australia

2. Ingham Institute of Medical Research, Sydney, Australia

3. Centre for Air Pollution, Energy & Health Research (CAR)

4. Woolcock Institute of Medical Research, The University of Sydney, Australia

5. Sydney School of Public Health, University of Sydney, Sydney, Australia

6. School of Public Health, The University of Queensland, Brisbane, Australia

7. University of Canberra, Canberra

8. MRC-PHE Centre for Environment & Health, Imperial College London, London, UK

9. Institute of Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, The Netherlands

Corresponding author:

Christine T Cowie; c.cowie@unsw.edu.au; +61 411 738 528

Declaration of interest:

1

Abstract

Background

Methods for estimating air pollutant exposures for epidemiological studies are becoming more complex in an effort to minimise exposure error and its associated bias. While land use regression (LUR) modelling is now an established method, there has been little comparison between LUR and other recent, more complex estimation methods. Our aim was to develop a LUR model to estimate intra-city exposures to nitrogen dioxide (NO_2) for a Sydney cohort, and to compare those with estimates from a national satellite-based LUR model (Sat-LUR) and a regional Bayesian Maximum Entropy (BME) model.

Methods

Satellite-based LUR and BME estimates were obtained using existing models. We used methods consistent with the European Study of Cohorts for Air Pollution Effects (ESCAPE) methodology to develop LUR models for NO_2 and NOx. We deployed 46 Ogawa passive samplers across western Sydney during 2013/2014 and acquired data on land use, population density, and traffic volumes for the study area. Annual average NO_2 concentrations for 2013 were estimated for 947 addresses in the study area using the three models: standard LUR, Sat-LUR and a BME model. Agreement between the estimates from the three models was assessed using interclass correlation coefficient (ICC), Bland-Altman methods and correlation analysis (CC).

Results

The NO_2 LUR model predicted 84% of spatial variability in annual mean NO_2 (RMSE: 1.2 ppb; cross-validated R^2: 0.82) with predictors of major roads, population and dwelling density, heavy traffic and commercial land use. A separate model was developed that captured 92% of variability in NOx (RMSE 2.3 ppb; cross-validated R^2: 0.90). The annual average NO_2 concentrations were 7.31 ppb (SD: 1.91), 7.01 ppb (SD: 1.92) and 7.90 ppb (SD: 1.85), for the LUR, Sat-LUR and BME models respectively. Comparing the standard LUR with Sat-LUR NO_2 cohort estimates, the mean estimates from the LUR were 4% higher than the Sat-LUR estimates, and the ICC was 0.73. The Pearson's correlation coefficients (CC) for the LUR vs Sat-LUR values were r=0.73(log-transformed data) and r=0.69 (untransformed data). Comparison of the NO_2 cohort estimates from the LUR model with the BME blended model indicated that the LUR mean estimates were 8% lower than the BME estimates. The ICC for the LUR vs BME estimates was 0.73. The CC for the logged LUR vs BME estimates was r=0.73 and for the unlogged estimates was r=0.69.

Conclusions

Our LUR models explained a high degree of spatial variability in annual mean NO_2 and NOx in western Sydney. The results indicate very good agreement between the standard LUR, national-scale sat-LUR, and regional BME models for estimating NO_2 for a cohort of children residing in Sydney, despite the different data inputs and differences in spatial scales of the models, providing confidence in their use in epidemiological studies.

Keywords

Land use regression; Bayesian; satellite; air pollution; exposure assessment; comparison

Background

Methods for estimating air pollution exposures in epidemiological studies have evolved substantially over the last 15 years, from simple measures such as assigning concentrations from the nearest air quality monitor, to models with increasing complexity and computational requirements. This has been driven by the need to reduce exposure misclassification and hence improve the validity of the estimation of air pollution – health response functions. There are several factors that influence the selection of the optimal approach to exposure modelling. These include the nature of the pollutant, its temporal and spatial distribution and the nature of the health effect being assessed. These factors also influence the choice of appropriate epidemiological study design (Baxter et al. 2013). Notwithstanding all these considerations, it will always be important to select exposure estimation methods that are most reliable and accurate (Baxter et al. 2013; Brauer et al. 2008; Sellier et al. 2014).

Land use regression (LUR) has been widely used to estimate concentrations of NO_2, NOx, PM_{10} and $PM_{2.5}$ (Hoek et al. 2008; Ozkaynak et al. 2013; Ryan and LeMasters 2007). LUR modelling gained popularity for estimating NO_2 and NOx because of the availability of relevant land use data and the ease with which empirical data for NO_2 and NOx could be collected from multiple locations using passive samplers. It has been used to estimate pollutant concentrations in various geographic settings ranging from very local scales, including a previous study in a 50 km^2 area of Sydney, Australia (Rose et al. 2011), to intra-city scales in Canada (Crouse et al. 2009; Henderson et al. 2007; Jerrett et al. 2007), the US (Gonzales et al. 2012; Jerrett et al. 2013), Europe (Beelen et al. 2013), the UK (Briggs et al. 2000), and Asian cities (Chen et al. 2010; Choi et al. 2017; Gurung et al. 2017).

Recent enhancements to LUR modelling include the incorporation of data obtained by satellite monitoring, which allows pollutant estimation over very large areas, for instance, national or continental scales (Bechle et al. 2015; de Hoogh et al. 2016; Hoek et al. 2015; Hystad et al. 2011; Knibbs et al. 2014; Novotny et al. 2011; Vienneau et al. 2013; Young et al. 2016), and even globally (Larkin et al. 2017). As LUR modelling alone does not incorporate data on chemical transformation, such as the formation of secondary and tertiary pollutant formation (de Hoogh et al. 2016), and is considered less suitable for modelling background concentrations at large scale, the incorporation of satellite data mitigates this limitation of LUR models. However, the disadvantages of relying on satellite data is the need for high-level technical expertise for data manipulation and the need to deal with missing satellite data. Furthermore, satellite-based models in the absence of other predictor data, may underestimate variability of ground based measurements as they are less able

to capture small scale spatial variation especially near point or linear sources of pollution (Geddes et al. 2016; Kharol et al. 2015; van Donkelaar et al. 2015).

Dispersion and chemical transport models (CTMs) are other common methods for assigning air pollution exposures (HEI 2010; Ozkaynak et al. 2013). Dispersion modelling utilises data from emissions inventories, meteorological data, knowledge of air chemistry and relatively complex mathematical modelling. While these models can provide high temporal resolution they may lack the fine spatial resolution needed to describe the distribution of exposure to NO_2 and NOx, which have steep spatial gradients.

In an effort to reduce the uncertainty associated with these individual approaches, recent work has sought to combine or "blend" separate exposure model estimates with varying spatial and temporal resolutions (Akita et al. 2014; Buteau et al. 2017; Hanigan et al. 2017). Hanigan et al (2017) used a Bayesian Maximum Entropy (BME) model to blend estimates from an Australian LUR model using satellite data sat-LUR (Knibbs et al. 2014), a CTM (Cope et al. 2014), and measurements from fixed site regulatory monitors, to produce NO_2 estimates for Sydney. The BME model resulted in a 6% improvement in Root Mean Square Error (RMSE) compared to the sat-LUR model and 16% improvement compared to the CTM model. Hence, this ensemble approach to estimation offers promise of incremental improvements in model precision. However, the improved predictive power gained through the Bayesian approach comes at the cost of substantially greater complexity and, hence, the need for more computer and human resources.

The aims of this study were two-fold: 1) to develop a standard LUR model to estimate NO_2 at the residential addresses of children in the Childhood Asthma Prevention Study (CAPS) cohort in Sydney (Garden et al. 2018); and 2) to compare those LUR model estimates with estimates from a national satellite-based LUR model (Sat-LUR) and a regional Bayesian Maximum Entropy (BME) model.

The CAPS cohort is a birth cohort (n=616), recruited from six maternity hospitals in south-west and western Sydney, between 1997 and 1999. The children underwent clinical and respiratory assessments at 1.5, 3, 5, 8, 11.5 years (Marks et al. 2006; Toelle et al. 2013) and 14 years. Previous analyses investigating the association between weighted road density, as a marker of traffic related air pollution, and respiratory and allergic outcomes in this cohort found that weighted road density within 100m of the home address, was associated with an increased risk of house dust mite allergy and allergic rhinitis (Hansell et al. 2014). For future exposure-response analyses of this cohort, we aimed to apply more accurate air pollution estimates by developing a standard NO_2/NOx LUR model.

Our second aim to compare NO_2 cohort estimates calculated from the three different spatial models is important, especially for settings where there may be limited capacity to develop more complex exposure models. To do this, we assessed the agreement between three estimates of NO_2 for each cohort address, derived from the standard LUR model and two alternative models: a national sat-LUR model (Knibbs et al. 2014); and an ensemble BME regional model (Hanigan et al. 2017). While a number of studies (Table S1) have compared pollutant estimates derived from different hybrid models, to our knowledge there are no studies that have examined the agreement between three such models with substantially varying spatial scales.

Methods

Study area and site selection

Sydney is a coastal city located on the eastern seaboard in the state of New South Wales, Australia. Greater Sydney (Greater Capital City Statistical Area as defined by the Australian Bureau of Statistics (ABS)) covers an area of 12,368 km^2, and is fringed by the Central Coast to the north, national parks to the south and the Blue Mountains to the west (City of Sydney 2018). In 2017 it had a resident population of 5.1 million people, and a population density of 407 persons/km^2, although the built urban area, estimated at 4,064 km^2, supports a density of 1,237 persons/km^2 (City of Sydney 2018). The study area for development of the LUR comprised the western half of the urban area of Sydney, covering 3122 km^2. The home addresses of the members of the CAPS cohort at time of recruitment were all within this study area. (Figure 1). Over the fourteen years of cohort follow-up, some cohort members moved to new addresses outside the study area, for which we did not attempt to estimate exposures.

We followed methods from the European Study of Cohorts for Air Pollution Effects (ESCAPE) study for site selection, measurement and development of the LUR models (Beelen et al. 2013; Eeftens et al. 2012). NO_2 and NOx were measured at 46 sites that were chosen to represent: 1) the range in pollutant concentrations likely to occur in the study area; 2) the likely range in predictor variables; and 3) the geographic extent of the study cohort addresses. Where possible, sites were located outside or close to cohort members' houses. Sites ranged from urban background sites with very low expected traffic counts and away from industry, to sites along major roads. The busiest of these roads recorded >73,000 vehicles per day (vpd) whereas quiet back streets were estimated to have less than 500 vpd. We categorised sites according to the ESCAPE protocol, but were mindful that Australian cities exhibit quite different geographical traits to European cities; they are generally larger in area and have lower population and dwelling densities, tend not to be subject to street canyon effects except for in central business districts, and are more reliant on personal vehicle use. We categorised the sites *a priori* based on local knowledge: 18 sites as ***traffic*** sites, 26 as ***urban background*** sites, and two sites as ***regional*** sites. Traffic sites were over-sampled given that traffic

is a major contributor to NO_2/NOx in the Sydney metropolitan region contributing 62% of NOx emissions (NSW EPA 2012). One site was co-located for comparative purposes with an Office of Environment & Heritage (OEH) regulatory monitor in Prospect, a western Sydney suburb.
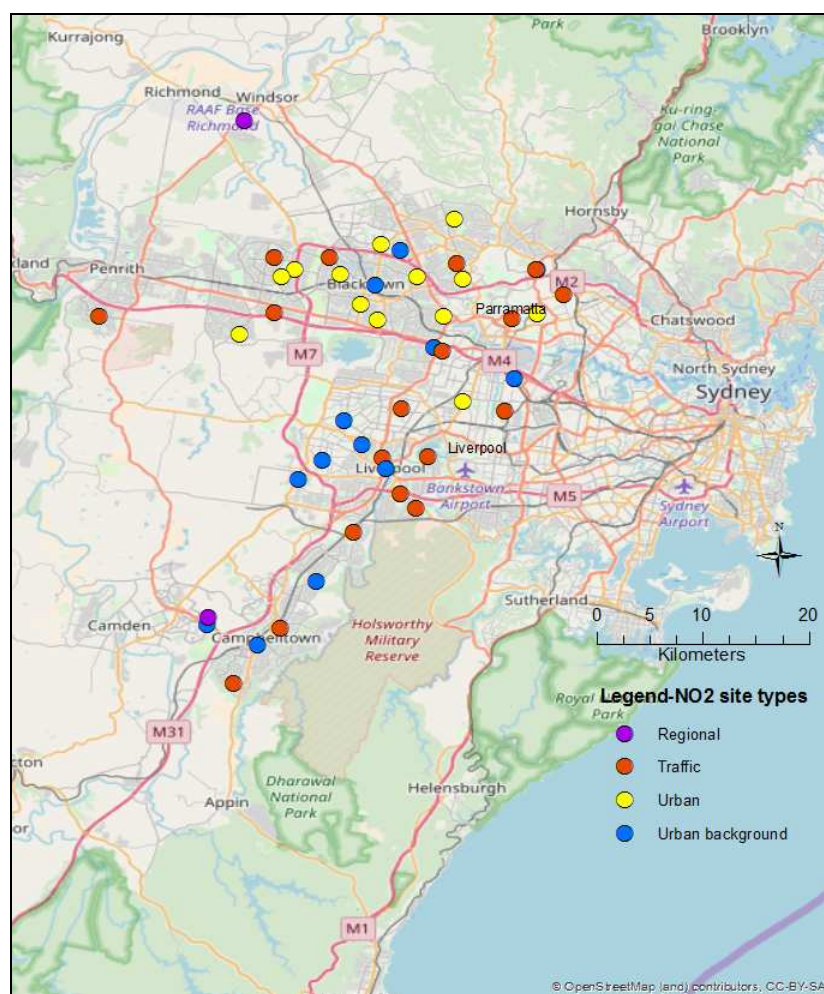

Figure 1. Study area, NO_2/NOx passive sampler sites by site type

NO_2 and NOx measurements

We used Ogawa passive diffusion samplers (Ogawa and Co. FL, USA) to measure NO_2 and NOx at the 46 sites. The samplers were placed *in situ* for two weeks in each of three seasonal measurement periods as per the ESCAPE study protocol: 22 July-5 August, 2013 (winter-cold season); 28 November-12 December 2013 (summer-hot season); and 18 March-1 April, 2014 (autumn-intermediate period). Duplicate samplers were randomly assigned and deployed at 11-13% of sites, varying by round. One pair of duplicate samplers was placed at the OEH fixed site monitor during each round of testing. Measurements were excluded when variability between duplicate samplers at the same sites exceeded 30%. Two field blanks and two laboratory blanks were also deployed per round. All samplers were simultaneously deployed and collected over an 8-10 hour

day across the study area to minimise the influence of meteorological impacts. Most samplers were placed on telegraph poles at a height of 2.2m to avoid tampering. However 3 samplers were placed under awnings at cohort dwellings. Samples were refrigerated until sent by courier to Edith Cowan University, Western Australia, for spectrophotometer analysis by the same laboratory used for the Perth LUR NO_2/NOx analysis (Dirgawati et al. 2015). The detection limits were 2.0 ppb (3.76 $\mu g/m^3$) for NO_2 and 3.4 ppb (6.39 $\mu g/m^3$) for NOx. ESCAPE Excel calculation sheets for NO_2 and NOx concentration calculations were used, taking into account local temperature, relative humidity and field blank results.

Annual average concentrations for NO_2 and NOx were calculated from the three measurement periods, after correcting for temporal variation, according to ESCAPE Study methods. For each site the measured concentration was adjusted by using data from a continuously operating regulatory monitoring site at Prospect (within the study area). In summary, the site concentrations were adjusted by subtracting the difference between the annual average concentration at the reference site and the measured concentration at the reference site for the relevant two week monitoring period (Cyrys et al. 2012; ESCAPE Study 2010). This allowed us to use data for all sites, even where data for only two of the three measurement periods were available.

Latitude and longitude coordinates for each site were assigned using Google maps.

Predictor variables

Data for the potential predictor variables were collated from various sources (Supplementary Table S2). Buffers of varying sizes were calculated to represent the density of specific classes of land use, population and dwelling density, and traffic counts on all roads within each buffer. ArcGIS 10.3 (ESRI, Redlands, CA, USA) was used to create the geographic variables and buffers and to calculate distance based variables (eg distance to main roads). The software R 3.2.1 (using the packages raster, rgeos, rgdal) was used to create and process some of the variables ie combined traffic variables, population and household densities, land use densities. All spatial datasets were converted to WGS 84/UTM zone 56S (EPSG:32756) for consistent projection in metres.

Land use data from the Australian Bureau of Statistics (ABS) at mesh-block level was extracted (ABS 2011). A mesh-block (MB) is the smallest statistical area available in the hierarchical ABS Australian Statistical Geography Standard 1270.0.55.001 (ASGS 2011) and represents around 30-60 dwellings. The ABS assigns each MB a land use category based on the planning designation for the majority of land in that area. ABS land use data were collapsed from nine land use types to four (residential, industrial, commercial and open) for LUR model development.

8

Data on resident population and number of dwellings was obtained from the 2011 ABS Census (ABS 2011), and the population and number of dwellings in each MB were divided by the MB area to create density variables.

Elevation data were obtained from GeoScience Australia and comprised Digital Elevation Models (DEM) using SRTM 1 second data (approximately 30 m resolution) from the Shuttle Radio Topography Mission (SRTM) conducted by NASA with Space Shuttle Endeavour over 11 days in 2000, where 80% of the Earth's surface topography was mapped (Gallant et al. 2011). The DEM-S (smoothed digital elevation model) was used to extract the elevation at each of the sampling sites.

Traffic data were obtained from two sources: the Zenith traffic model and the NSW Lands and Property Information (LPI). The Zenith traffic model was built using all available traffic count data from Roads & Maritime Services (RMS), the NSW government authority with responsibility for all primary (highways, motorways, arterial/main, sub-arterial/main), and some secondary roads (distributor roads) together with data from local councils for local roads where available (Zenith Traffic Model 2014). The Zenith model combines the traffic count data with data from road and rail networks, land use data, household travel surveys (destination and purpose of journey), to estimate traffic counts for all of the Greater Sydney Metropolitan Area (GSMA) for non-local roads and a small proportion of local roads. Zenith modelled traffic data were used to calculate many of the traffic related variables (Table S2).

To capture local roads in our traffic GIS layer, we combined the Zenith traffic modelled data with road network data from the LPI to obtain a representative layer of traffic counts for all roads. These two sources of traffic data were combined in ArcGIS to assign a traffic count to all roads. Where counts were missing, local roads were assigned a count of 500 vehicles per day (vpd) and a count of 20 vpd for heavy vehicles. Major roads were defined as roads with counts >5000 vpd, consistent with the ESCAPE study. Where counts for heavy vehicles were missing for major roads, they were assigned a count of 500 heavy vpd.

LUR model development

Multiple regression modelling was conducted separately for NO_2 and NOx. The measured NO_2 and NOx readings were the outcome variables in the regression models. Models were built using a standardised stepwise forward selection procedure, in accordance with the ESCAPE methodology (Beelen et al. 2013). Univariate analyses were conducted to identify the variable contributing most (highest adjusted R^2) to the variation in the pollutant. The remaining variables were then separately added to the model to determine the model with the next highest adjusted R^2. Variables were added and retained in the model in accordance with ESCAPE criteria (Table 1 (footnote)). We excluded variables from the model when they had >75% zero values. Variables with different buffer sizes

were added to the model as indicated above. There were 135 potential predictor variables considered in the regression models.

Diagnostic tests used to select the final models included: 1) variance inflation factor (VIF) was required to be < 3, representing a lack of collinearity between the variables; 2) an examination of influential observations using Cook's distance (D) >0.8; and 3) an examination of Moran's I test for heteroscedasticity, normality and spatial autocorrelation of residuals (testing the assumption of independence). If the final models resulted in Cook's D values >0.8, the relevant sites were sequentially removed from the model and the relative changes in variable parameter estimates, the p values of the variables and the model adjusted R^2 were examined to determine the changes in the model structure. For each of the sites, we also considered type and nature of the site eg whether a site might represent heavy traffic because of its location, and thus the value of retaining the site in model development.

LUR model validation

We used "hold-out" validation to check the robustness of our models (Dirgawati et al. 2015; Gulliver et al. 2013; Johnson et al. 2010; Wang et al. 2012; Wang et al. 2016). Hold-out validation is considered to be an improvement on leave-one-out cross-validation (LOOCV) validation which has been found to overestimate the explanatory power of LUR models when the number of sample sites is small (Wang et al. 2012). Hold-out validation withholds specified randomly selected subsets of sampling sites as "training" datasets and develops models based on a smaller number of sites. Due to the relatively small number of sampling sites in our study, we ran three separate hold-out modelling procedures as a sensitivity analysis, first with-holding 10% of data points (10-fold), then 25% (4-fold) and finally 50% (2-fold) of data points. We used the R3.2.1 statistical package for this analysis. The RMSE represents the absolute difference between predicted and measured concentrations and so is an indicator of reliability of the LUR prediction model.

Back and forward extrapolation of annual average concentrations

LUR models were used to estimate annual average concentrations of NO_2 and NOx at the cohort addresses for the study period, July 2013-June 2014. To avoid over-extrapolation of pollutant estimates at the cohort addresses, we restricted or "truncated" values for the predictor variables to those observed at the field monitoring sites, when running the models (de Hoogh et al. 2014; Dirgawati et al. 2015; Wang et al. 2012).

These cohort estimates were then adjusted both backwards (pre-2013) and forwards (post-2013) in time using the mean of the average annual concentrations from five fixed site regulatory monitors located within the study area (Prospect, Liverpool, Chullora, St Mary's, Richmond), and using 2013

as the base year. The cohort estimates were adjusted by the difference in concentration between 2013 (base year) and the year of interest which was dependent on when the cohort participants were tested.

NO_2 predictions at the CAPS cohort addresses and comparison with sat-LUR estimates and BME model estimates

In brief, the national Sat-LUR used NO_2 data from fixed site regulatory monitors from around Australia (n=68) as the outcome variable, and satellite derived NO_2 estimates from the Ozone Monitoring Instrument (OMI) from the Aura satellite, land use, and traffic data as predictor variables (Knibbs et al. 2014). Generalised estimating equation (GEE) models were used to develop annual models for 2006-2011, but for this study were updated with 2013 data (satellite and fixed monitor data). The Sat-LUR model was validated using passive sampler NO_2 data (including data from this study) and was found to predict 58% of variability across all sites and 69% variability at the urban near-traffic and background sites in two capital cities (Sydney and Perth) (Knibbs et al. 2016).

Bayesian maximum entropy (BME) modelling uses Bayesian analyses to blend different sources of data with varying temporal and spatial resolutions, with the objective of garnering all data into a blended estimate which is improved and more precise (ie smaller uncertainties). The BME model used in this analysis combined NO_2 estimates from the Sat-LUR model, a Chemical Transport Model (CTM) and regulatory fixed site monitors. The predictive BME model was built using 2011 data, but validated using the NO_2 passive sampler data collected in 2013-14 for our standard LUR model, and then adjusted for seasonal differences. This was due to an absence of alternative independent data from 2011. Of note, NO_2 data from the fixed site monitors showed only minimal change during the period 2011-2014. The BME reported 6% and 16% improvements in root mean square error (RMSE) compared with the Sat-LUR and CTM models respectively, when compared using a Sydney wide study area for that analysis (Hanigan et al. 2017).

We estimated NO_2 concentrations for all CAPS cohort addresses within a 15 km radius of the passive sampler sites using the LUR model, the Sat-LUR and BME model. This study area of 3122 km^2 encompassed all of the CAPS cohort addresses at time of recruitment and represented an area for which we could confidently estimate exposures using the model. For the LUR model the explanatory regression variables were calculated for each geocoded address and the estimates made using the NO_2 and NOx regression equations. For the model agreement analysis, estimates using the Sat-LUR model and the BME model were calculated as follows: the Sat-LUR estimates were based on the centroids of each mesh-block (smallest geostatistical area used in the Australian census) and assigned to each cohort geocode; the BME model estimated NO_2 concentrations for centroids on a 100 x100 m grid. All three models were used to estimate NO_2 concentrations during 2013 to enable agreement comparison.

As the NO$_2$ concentrations were log normally distributed, we used the natural logarithm of the values to conduct the following agreement analyses: visual examination of agreement using scatter plots and calculated Pearson's correlation coefficients as commonly reported in the past; and Bland-Altman plots and calculation of the intra-class correlation coefficients (ICC) (Shrout and Fleiss inter-rater reliability test) to estimate relative and absolute agreement. The Bland-Altman plots and ICC calculations are the most appropriate statistical tools for assessing agreement between two methods (Giavarina 2015; Koo and Li 2016), but have been seldomly applied for comparing LUR and other model estimates (Buteau et al. 2017; de Hoogh et al. 2014).

SAS (version 9.4; SAS Institute Inc., Cary, NC, USA) was used to conduct the regression analyses, diagnostic tests, and ICC. The LUR model was used to produce NO$_2$ and NOx estimates for 250 x 250 m gridpoints across the study area, and these were mapped for visual presentation purposes using R software version 3.2.1.

Results

Figure 2 shows the distribution of NO$_2$ and NOx concentrations, as measured by the Ogawa passive samples, classified by site type. NO$_2$ concentrations ranged from 3.8 to 17.6 ppb (7.1 to 33.1 μg/m^3) and NOx ranged from 6.5 to 43.4 ppb (12.2 to 81.6 μg/m^3) (Figure 2; Supplementary Table S3).
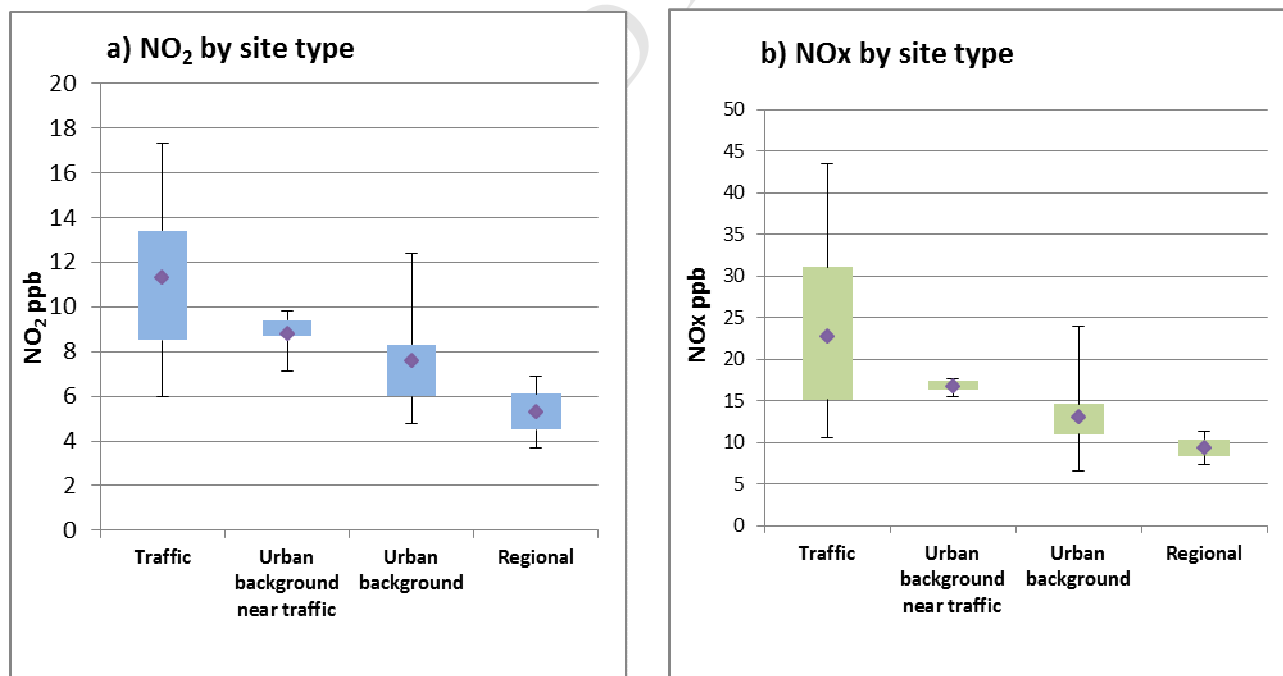

Figure 2. NO$_2$ and NOx concentrations by site type

All Ogawa samplers returned readings above the detection limit. Five pairs (29%) of duplicate NO$_2$ readings were excluded as they had > 30% variability. One of these pairs included a cracked

sampler and two other pairs included samplers that had been incorrectly installed. Supplementary Table S4 shows the NO_2 concentrations measured by the Ogawa passive samplers that were co-located with the OEH fixed site chemiluminesence monitor. The winter readings could not be assessed due to four days of missing data at the OEH sites. The summer readings were 3.0 ppb lower on the Ogawa passive sampler (5.5 ppb passive sampler vs 8.5 ppb regulatory monitor); and autumn readings were 1.2 ppb lower on the Ogawa sampler (9.2 ppb sampler vs 10.4 ppb monitor).

Traffic variables, particularly those related to heavy traffic and major roads, population density within 5 km, dwelling density within 100 m, and commercial land use were the strongest predictors of NO_2 and NOx in each LUR model (Table 1). The best NO_2 LUR model had an $R^2=0.84$ and RMSE= 2.35 $\mu g/m^3$. Moran's test (p=0.97) indicated that there was no spatial autocorrelation. The best LUR model for NOx had an $R^2=0.916$ and RMSE=4.35 $\mu g/m^3$. Moran's test (p=0.96) indicated that there was no spatial autocorrelation.

For the NO_2 and NOx models, one and four sites respectively, were found to have Cook's D values >0.8. These were sequentially removed from the model but the model structures and estimates did not change substantially (<10%) with exclusion of any of the sites. Thus, all sites were retained in the model.

Validation of the models

In all three "hold-out" scenarios the estimated R^2 and RMSE were similar to the estimates in the initial model (Table 2). This lends support to the reliability of the selected models.

Figure 3 (a, b) illustrates the NO_2 and NOx spatial predictions over the study area estimated by the LUR models. Expanded views of part of the study area incorporating the suburban centres of Parramatta and Liverpool are provided to illustrate the variation in pollutant estimates.

Comparison of standard LUR estimates with satellite-LUR estimates for the CAPS cohort addresses

Table 3 provides summary statistics for the estimated NO_2 concentration at the CAPS cohort addresses according to the LUR, the Sat-LUR and the BME models.

There was good agreement between the NO_2 LUR and both the Sat-LUR NO_2 and BME model estimates. The ICC for comparison of all three methods was 0.93. The Bland-Altman plots (Figure 4) and the scatterplots (Figure 5 and Supplementary Fig S1 (untransformed values)) illustrate that the three sets of model estimates showed very good agreement, although there was greater variability in the estimates at higher NO_2 concentrations. The scatter plots also showed that both the Sat-LUR and BME model estimates were higher than the LUR estimates at lower NO_2 concentrations and lower than the LUR estimates at high NO_2 concentrations.

13

Table 1. LUR models for NO₂ and NOx for Western Sydney, 2013-2014

Pollutant predictors**	Coefficient	Standard Error	Standardised coefficient*	P	VIF	Increm-ental R^2
NO$_2$						
Intercept	5.7295	1.0983	16.5454			
Product of traffic intensity on nearest major road & inverse distance to the road (vpd/metre) [a]	0.0015	4.079×10^{-4}	1.9094	0.039	2.31	0.49
Population density within 5000 m (number of people)[f]	3.91×10^{-5}	6.99×10^{-6}	2.0419	<0.0001	1.09	0.65
Dwelling density within 100 m (number)	0.0653	0.0173	1.4682	<0.0006	1.34	0.73
Heavy traffic intensity on nearest road (vpd) [b]	0.0012	2.401×10^{-4}	2.6114	0.001	2.26	0.80
Commercial land use within 700 m (proportion of land area within buffer distance)	9.96×10^{-6}	3.00×10^{-6}	1.3523	0.002	1.36	0.84
NOx						
Intercept	23.6982	3.554	30.9472	<0.0001	0	
Length of major roads within 75m buffer (metres) [c]	43.3518	6.9328	5.2753	<0.0001	1.69	0.58
Product of heavy traffic intensity on nearest road & inverse distance to nearest road (vpd/metre)[d]	0.0558	0.0073	7.0991	<0.0001	2.04	0.70
Dwelling density within 100 m (number)	0.1116	0.0329	2.6182	0.0016	1.42	0.79
Elevation (square root) (m)	-1.770	0.3326	-3.8857	<.0001	1.27	0.84
Heavy traffic intensity on nearest major road (vpd) [e]	0.0012	2.671×10^{-4}	3.3809	0.0001	1.46	0.88
Commercial land use within 500 m (proportion of land area within buffer distance)	3.42×10^{-5}	9.37×10^{-6}	2.7218	0.0008	1.32	0.90
Population density within 5000 m (number of people)	4.039×10^{-5}	1.344×10^{-5}	2.1078	0.0047	1.17	0.92

** Variables added and retained in the models when: 1) absolute increase in adjusted R^2 >1%; 2) coefficients in the pre-specified expected direction and, 3) direction of effect of the coefficients of the retained variables did not change with addition of the new variable.* Predictor variables in the final model were used to generate standardised predictor variables (and coefficients) by subtracting the mean for each variable from each value for that variable and dividing by the standard deviation (SD).

[a] Intmajorinvdist (ESCAPE name)
[b] Heavytrafnear
[c] Majroad75
[d] Heavyintinvdist
[e] Heavytrafmajor

Table 2. Results from hold-out cross validation of NO$_2$ and NOx LUR models

Hold out validation	% data points used for training dataset	Adjusted R^2	RMSE
NO$_2$ model		**0.84**	**2.35**
10-fold	10%	0.82	2.41
4-fold	25%	0.85	2.37
2-fold	50%	0.85	2.33
NOx model		**0.92**	**4.35**
10-fold	10%	0.90	4.51
4-fold	25%	0.90	4.68
2-fold	50%	0.85	6.56

Table 3. Summary statistics for NO$_2$ (ppb) estimated for the CAPS cohort addresses by the LUR, Sat-LUR and BME model

Model	N	Mean	Median	SD	Min	Max	25%	75%
LUR	947	7.3	7.3	1.9	3.1	17.7	6.1	8.2
Sat-LUR	947	7.0	6.7	1.9	3.6	18.0	5.8	7.9
BME	947	7.9	7.7	1.9	4.5	18.1	6.6	8.9

Mean estimates from the LUR were 4% higher than the Sat-LUR estimates (95%CI: 3% to 6%). The limits of agreement indicating where 95% of the values for the fold difference between the two estimates lie, were 0.7 (lower limit) and 1.52 (upper limit) (antilog values of Figure 4a). The ICC for the LUR vs Sat-LUR was 0.73. The correlation between the LUR and Sat-LUR (log values) was r=0.73 and for the unlogged estimates was r=0.68.

The mean estimates for the LUR model were 8% lower than the BME estimates (95%CI: -9.5% to -7%). The lower and upper limits of agreement were between 0.64 and 1.31 fold difference, respectively (antilog values of Figure 4b). The ICC for the LUR vs BME estimates was 073. The correlation between the LUR and the BME estimates (log value) was r=0.73 and for the unlogged estimates was r=0.69.

Figure 3 Surface maps of LUR (a) NO₂ and (b) NOx concentrations, for western Sydney and expanded view (Parramatta and Liverpool), 2013-2014
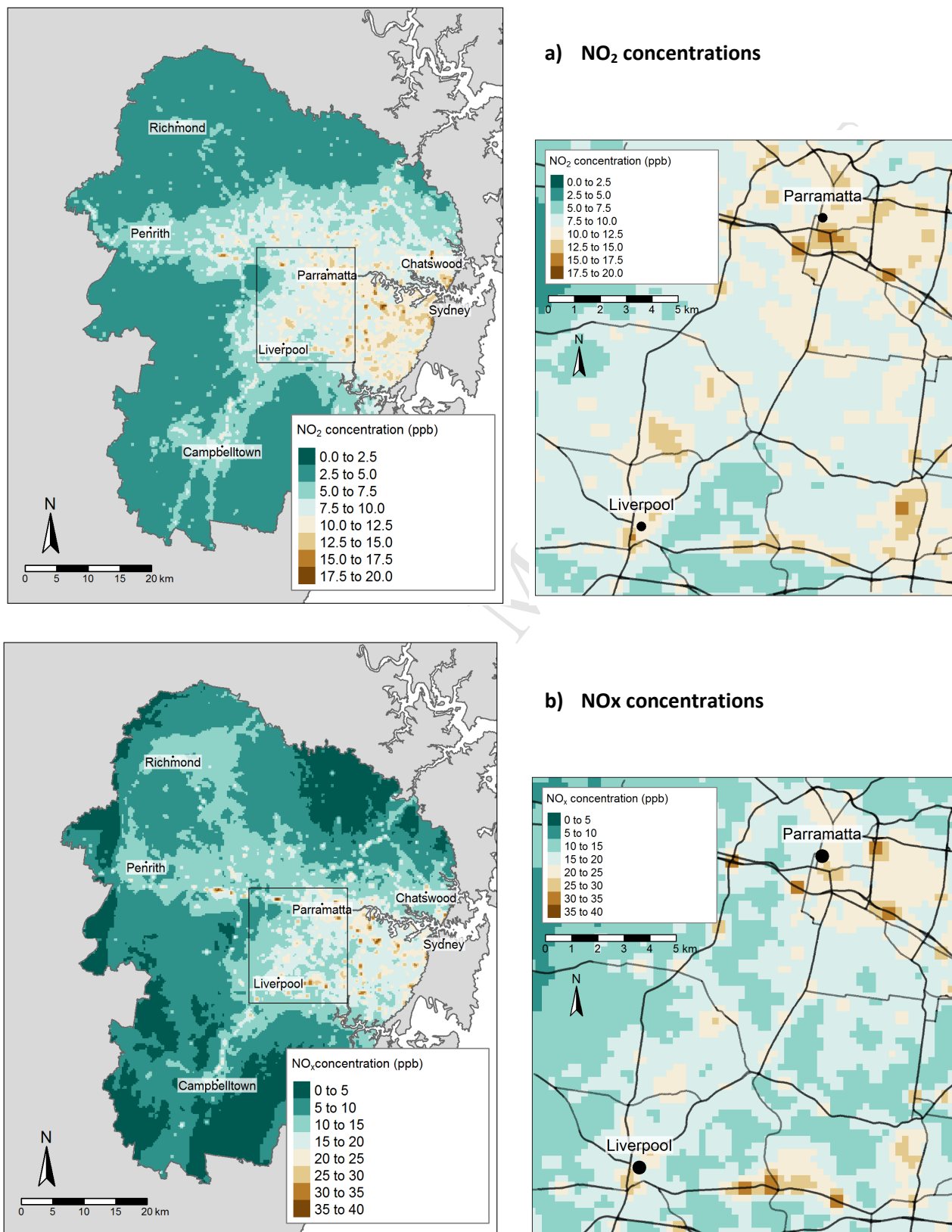

a) NO₂ concentrations

b) NOx concentrations

Figure 4. Bland-Altman plots of level of agreement (log ppb)

a)   Sat-LUR vs LUR NO$_2$ estimates (log ppb)   b) BME vs LUR NO$_2$ estimates (log ppb)
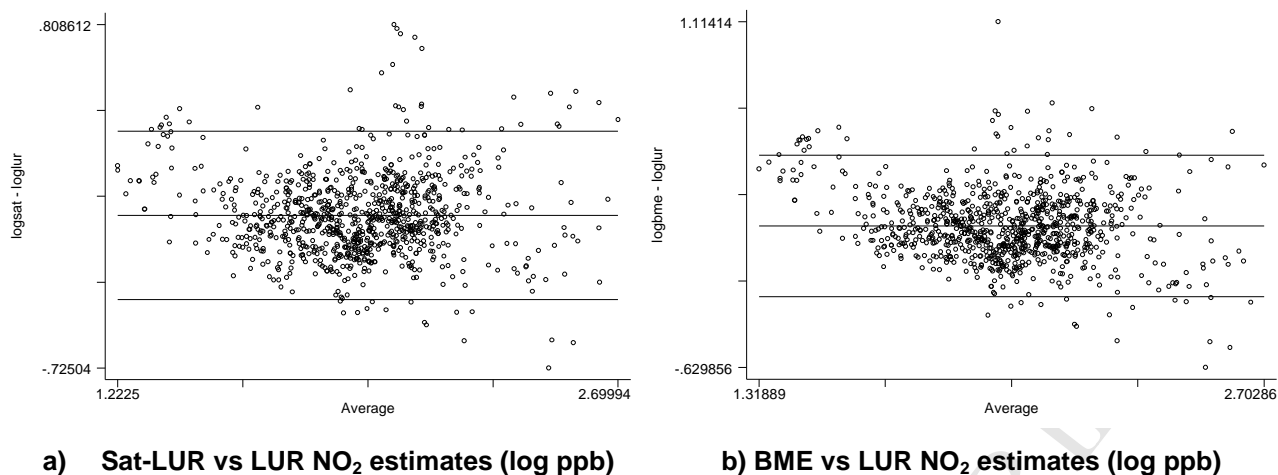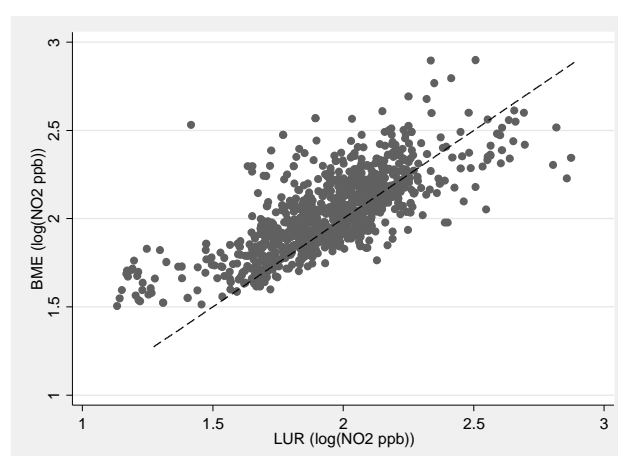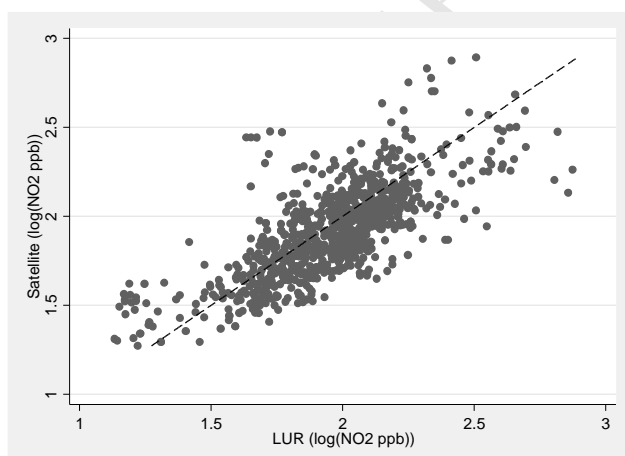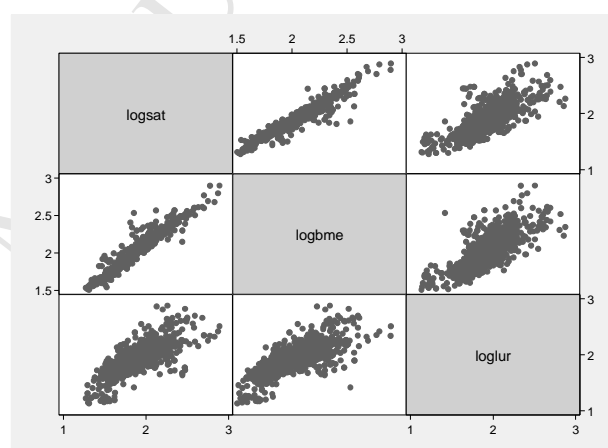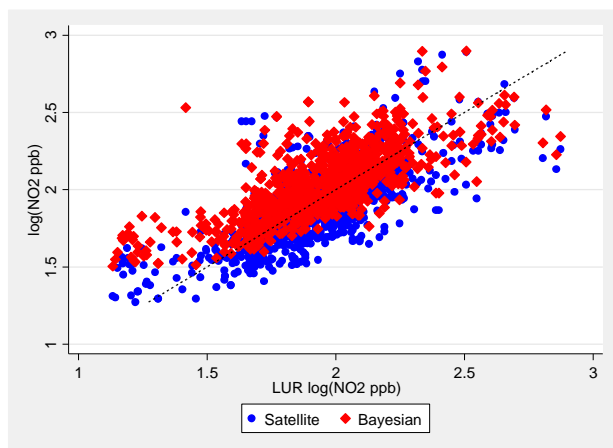
Figure 5. Scatter plots of estimates from a) LUR vs Sat-LUR (blue) and BME (red) NO$_2$; b) Scatter plot matrix; c) LUR vs Sat-LUR NO$_2$; d) LUR vs BME NO$_2$; (all log ppb)

a)   b)

c)   d)

Discussion

17

We developed standard LUR models with high spatial predictive capability of 84% and 91% for NO_2 and NOx respectively, for an area covering around 75% of the built urban area of Sydney. We also show that estimates from the standard LUR models that we developed specifically for the study area agree well with estimates from more complex models incorporating both satellite and CTM data developed for much larger areas.

The validity of our models, 84% and 91% for NO_2 and NOx, respectively, compare well with the better predictive models from the European ESCAPE study (Beelen et al. 2013), North America (HEI 2010) and those reported in reviews of LUR modelling (Hoek et al. 2008; Ryan and LeMasters 2007). For example, the ESCAPE NO_2 models explained 55 to 92% (median 82%) of the variation in NO_2 and 49 to 91% (median 78%) in NOx, across 36 study areas in Europe (Beelen et al. 2013).

In Australia, only three previous LUR NO_2 models have used empirical ground based measurements; one in a small study area in Sydney (Rose et al. 2011), one for Perth, a metropolitan city on Australia's west coast (Dirgawati et al. 2015); and a recent model for Brisbane, a tropical city in northern Australia ((Rahman et al. 2017). The earlier Sydney model, incorporating fixed site monitored data and a simpler traffic measure (weighted road density), developed for a much smaller area (50 km^2) in northern Sydney, explained 80% of the overall variation in NO_2 (Rose et al. 2011). However, it was uncertain whether this model could be extrapolated across Sydney due to its small scale. We also updated the current study using traffic counts as a model input. The Perth LUR model reported predictions of 69% and 75% in NO_2 and NOx variability, respectively, across the metropolitan Perth area (Dirgawati et al. 2015). The Brisbane models explained 64% and 70% of variability in NO_2 and NOx concentrations respectively, but are not directly comparable to the Sydney models as they estimated daily pollutant concentrations using short-term and long-term monitoring campaigns, and sample sites were restricted to schools and EPA monitoring sites which may not represent the full range of expected concentrations (Rahman et al. 2017).

Annual mean NO_2 and NOx concentrations in Sydney (9.2 ppb NO_2, 17.1 ppb NOx) were almost twice as high as for Perth (5.3 ppb NO_2 and 9.9 ppb NOx). Our mean concentrations for NO_2 were similar to those in ESCAPE models from Scandinavia and Hungary and Erfurt, Germany, but substantially lower than for England and other European countries, albeit sampling for the ESCAPE models occurred between 2008-2011 (Beelen et al. 2013).

Slightly different predictors were chosen for our NO_2 model compared with the NOx model, which is expected given that NOx is a primary pollutant and NO_2 a secondary pollutant formed by reaction of NOx with oxygen in the atmosphere. As a consequence these two pollutants have different scaled spatial contrasts (Karner et al. 2010). The slightly different predictors for each model are consistent with the ESCAPE models for various cities (Beelen et al. 2013). Nevertheless, the nature of the predictors was similar for both pollutant models in that the main predictors related to: heavy traffic,

18

population density, dwelling density, and commercial land use, and each model included at least two traffic variables. We found commonality in the final variables selected in our model and the previous small-area Sydney LUR model (Rose et al. 2011), which included two traffic density variables, population density (within 500 m) and commercial land use (within 750 m). The Perth LUR models reported traffic intensity on the nearest road, household density, industrial land use, and road length within 50 m the main predictor variables for both NO_2 and NOx (Dirgawati et al. 2015). In Europe, the most consistently chosen explanatory variables across the 36 ESCAPE study areas included small-scale traffic and population and household density (Beelen et al. 2013). Our models demonstrate that variables reflecting traffic density (including heavy traffic on main roads) and distance to roads, are the strongest predictor variables for NO_2 and NOx pollution in Sydney. This is not surprising, given that traffic contributes approximately 62% of NOx emissions in Sydney (NSW EPA 2016). Related to this finding, the results from the passive samplers substantiated the a priori categorisation of sampling sites and demonstrated that NO_2 and NOx concentrations were greatest at the traffic sites, lowest at regional sites representative of the urban-rural fringe, and midway between the two categories at the urban background sites which were representative of suburban backstreets.

We report good agreement between the NO_2 cohort exposure estimates derived from our LUR model and the estimates derived from both the national Sat-LUR model and the BME model for our study area. There was little difference between the mean estimates for the three methods, with this difference being much lower than the precision for passive sampler analysis. The Sat-LUR model however, resulted in slightly lower mean and median estimates than did the LUR model, which may be a function of the location of the regulatory measurement sites used as the outcome variable for Sat-LUR model building, which in Sydney are at "background" locations. For all summary measures the two methods were highly comparable. Both models included traffic variables as major predictors of NO_2. We conclude that, at least in the present setting, the choice among the alternative models should be made on the basis of availability of data and feasibility, as all three models yield similar findings. From a broader perspective, the choice of application of a specific exposure model for epidemiological purposes will also be guided by the spatial extent of participants and study design.

The BME model estimated slightly higher mean, minimum and maximum NO_2 concentrations than the LUR model, however the mean difference between estimates from the two models was very small. The small differences might be due to the different spatial resolution for pollutant estimation: the Sat-LUR method was based on centroids of each mesh-block whereas the BME model estimated concentrations for centroids of 100 x 100 m grids. The BME model also incorporates information from the CTM model, which is reliant on pollutant inventory data, and thus likely to represent industrial source contribution of NO_2 differently to the industrial land use density data used for LUR and Sat-LUR model development.

One key feature of our study is that it compares estimates from models applied at different spatial scales. Even when using just one model method such as LUR, use of varying spatial scales for the same area can reveal different patterns with sometimes poor transferability (Marcon et al. 2015), particularly for pollutants with strong spatial gradients. Our results indicate that the variation in spatial scale of the three models does not appear to be a substantial limitation in the context of this study, seen by the good model(s) agreement and similar summary statistics for the estimates. However, our study found that the models with lower spatial resolution (Sat-LUR and BME), tended to predict higher concentrations at background sites compared with the LUR and the difference in spatial gradients may have contributed to the small variation in NO_2 estimates. The higher concentrations estimated by the Sat-LUR and BME models might be explained by the fact that the standard LUR relies on local knowledge of the study area characteristics for site selection so as to maximise the range of measured pollutant concentrations. Hence, theoretically it is better placed to capture minimum and maximum concentrations. Understanding how the potential biases at the lower and upper ends of our NO_2 exposure range might impact on the exposure-response coefficients in epidemiological analyses is important. We will test this by applying all three estimates to the CAPS cohort.

To date we have found no other studies of NO_2 estimation that have compared agreement, using appropriate agreement statistics, between LUR models and separately developed models using satellite data or Bayesian methods, that might indicate whether this finding is generalizable to other countries or areas. In Supplementary Table S1 we summarise results from studies (Beelen et al. 2010; Buteau et al. 2017; de Hoogh et al. 2014; Hennig et al. 2016; Marshall et al. 2008; Sellier et al. 2014; Wang et al. 2015; Wu et al. 2011) that have compared LUR estimates of NO_2 with other modelled NO_2 estimates (dispersion modelled including CTMs), or with measured concentrations (nearest monitor measurement; inverse distance weighted measurement). It reports correlations for the various comparisons which range from R=0.19 to 0.89, and includes at least two previous LUR comparison studies that underscore the influence of local characteristics in model development. Comparison of LUR vs CTM NO_2 estimates in Germany reported better R^2 when the CTM method was restricted to local traffic areas only (Hennig et al. 2016). A Dutch study reported good agreement at the mid-range of concentrations estimated by LUR versus dispersion modelling and versus validation sites, but larger differences at the ends of the concentration range, suggesting the differences might be due to the coarse categorisation for the industrial land use variable used in the LUR (Beelen et al. 2010).

A limitation of all but two (Buteau et al. 2017; de Hoogh et al. 2014) of these comparative studies is that they have used correlation coefficients to describe the agreement between methods, rather than reporting agreement statistics such as Bland-Altman ICC statistics or Bland Altman plots.

Strong correlation does not inherently imply good agreement, as two methods might be highly correlated but concentrations may not agree well (Giavarina 2015; Koo and Li 2016).

Limitations of our LUR model development are the relatively small number of sample sites (n=46), compared to some studies. However, our separate hold-out validation analyses demonstrated the robustness of both the NO_2 and NOx models, with the adjusted R^2 found to be high and similar between different hold-out iterations of 10%, 25% and 50%, indicating that the smaller sample size is unlikely to have unduly influenced model estimates. It is also possible that the measurement of NO_2/NOx during three seasons using the passive samplers did not fully capture the annual variation in pollutant concentrations, however, we sought to adopt sampling methods which were consistent with most other LUR studies (Beelen et al. 2013; Hoek et al. 2008; Ryan and LeMasters 2007).

The number of variables in our final NOx model (n=7) is another potential weakness, given the number of sample sites is relatively low and the number of predictor variables is large. Our NOx model with seven predictor variables may be over-specified, however our models are consistent with ESCAPE models where the number of model variables ranged from 2-7. Furthermore, the incremental R^2 of the NOx model was 0.84 and 0.88 with four and five predictor variables respectively, reflecting good variance explained with a lower number of variables and variables were only retained in the model if VIF<3.

Strengths of our LUR model development include the use of a standardised and previously tested method (ESCAPE) that allows comparison of our results with a local Perth model and a large number of European models, whilst also acknowledging variation in local environmental and geographic conditions. Our hold-out validation is a second strength of the study, demonstrating robustness of the model irrespective of the number of samples used for model building and validation. A further strength is the use of agreement statistical analyses to report on comparison between the methods. In this study we used Bland-Altman statistics and determined the intra-class correlation coefficient whereas most previous studies have reported Pearson correlation coefficients, which are a measure of correlation rather than agreement (Giavarina 2015).

Conclusion

Our study is the first, as far as we are aware, to compare agreement between NO_2 estimates from a standard LUR to a national LUR model using satellite estimates and a Bayesian blended model. We reported strong agreement as well as small absolute differences between our LUR model and both of the other models, despite the different data inputs and differences in the spatial scales of the models. Our study reaffirms the place for standard LUR model methods, not only, but particularly, where it may be challenging to apply more complex exposure models because of data gaps and/or a lack of computational resources. Further work is required to determine whether this good

agreement holds true for other pollutants, especially for particulate matter concentrations. Our study cannot indicate which of the methods is the most valid as there is no gold standard method, however, it demonstrates that any of our tested exposure methods can be used to assign annual average NO_2 exposures for epidemiological analyses in Sydney. It also indicates that the LUR model might be generalisable to other areas within Australia, both urban and non-urban, although this will require validation in both rural settings as well as other cities. We will apply the three sets of model estimates for exposure-response analyses for the CAPS cohort to determine their impact on the effect estimates.

Acknowledgements

References

ABS. 2011. 1270.0.55.001 - Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2011. Available: http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/1270.0.55.001July%202011?OpenDocument; Accessed: 6 June 2015.

Akita Y, Baldasano JM, Beelen R, Cirach M, de Hoogh K, et al. 2014. Large scale air pollution estimation method combining land use regression and chemical transport modeling in a geostatistical framework. *Environ Sci Technol*; 48; 8:4452-4459.

Baxter LK, Dionisio KL, Burke J, Ebelt Sarnat S, Sarnat JA, et al. 2013. Exposure prediction approaches used in air pollution epidemiology studies: Key findings and future recommendations. *J Expos Sci Environ Epidemiol*; 23; 6.

Bechle M, Millet D, Marshall J. 2015. National spatiotemporal exposure surface for NO_2: Monthly scaling of a satellite-derived land-use regression, 2000-2010. *Environmental Science & Technology*; 49; 20:12297-12305.

Beelen R, Voogt M, Duyzer J, Zandveld P, Hoek G. 2010. Comparison of the performances of land use regression modelling and dispersion modelling in estimating small-scale variations in long-term air pollution concentrations in a Dutch urban area. *Atmospheric Environment*; 44; 36:4614-4621.

Beelen R, Hoek G, Vienneau D, Eeftens M, Dimakopoulou K, et al. 2013. Development of NO_2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe - The ESCAPE project. *Atmospheric Environment*; 72; Supplement C:10-23.

Brauer M, Lencar C, Tamburic L, Koehoorn M, Demers P, Karr C. 2008. A cohort study of traffic-related air pollution impacts on birth uutcomes. *Environmental Health Perspectives*; 116; 5:680-686.

Briggs DJ, de Hoogh C, Gulliver J, Wills J, Elliott P, et al. 2000. A regression-based method for mapping traffic-related air pollution: application and testing in four contrasting urban environments. *The Science of The Total Environment*; 253; 1-3:151-167.

Buteau S, Hatzopoulou M, Crouse DL, Smargiassi A, Burnett RT, et al. 2017. Comparison of spatiotemporal prediction models of daily exposure of individuals to ambient nitrogen dioxide and ozone in Montreal, Canada. *Environmental Research*; 156:201-230.

Chen L, Bai ZP, Kong SF, Han B, You Y, et al. 2010. A land use regression for predicting NO_2 and PM_{10} concentrations in different seasons in Tianjin region, China. *Journal of Environmental Sciences*; 22; 9:1364-1373.

Choi G, Bell M, Lee J. 2017. A study on modeling nitrogen dioxide concentrations using land-use regression and conventionally used exposure assessment methods. *Environmental Research Letters*; 12; 4.

City of Sydney. 2018. Greater Sydney. Geography. Available at:
http://www.cityofsydney.nsw.gov.au/learn/research-and-statistics/the-city-at-a-glance/greater-sydney.
Accessed: 3 May 2018.

Cope M, Keywood M, Emmerson K, Galbally I, Boast K, et al. 2014. The Centre for Australian Weather and
Climate Research Sydney Particle Study-Stage-II; CSIRO and Office of Environment and Heritage, Sydney,
Australia, 2014. ISBN: 978-1-4863-0359-5.

Crouse DL, Goldberg MS, Ross NA. 2009. A prediction-based approach to modelling temporal and spatial
variability of traffic-related air pollution in Montreal, Canada. *Atmospheric Environment*; 43; 32:5075-5084.

Cyrys J, Eeftens M, Heinrich J, Ampe C, Armengaud A, et al. 2012. Variation of NO_2 and NOx concentrations
between and within 36 European study areas: Results from the ESCAPE study. *Atmos Environ*; 62.

de Hoogh K, Korek M, Vienneau D, Keuken M, Kukkonen J, et al. 2014. Comparing land use regression and
dispersion modelling to assess residential exposure to ambient air pollution for epidemiological studies.
Environment International; 73:382-392.

de Hoogh K, Gulliver J, Donkelaar Av, Martin RV, Marshall JD, et al. 2016. Development of West-European
$PM_{2.5}$ and NO_2 land use regression models incorporating satellite-derived and chemical transport modelling
data. *Environmental Research*; 151:1-10.

Dirgawati M, Barnes R, Wheeler AJ, Arnold AL, McCaul KA, et al. 2015. Development of Land Use Regression
models for predicting exposure to NO_2 and NOx in Metropolitan Perth, Western Australia. *Environmental
Modelling & Software*; 74:258-267.

Eeftens M, Beelen R, de Hoogh K, Bellander T, Cesaroni G, et al. 2012. Development of Land Use Regression
Models for $PM_{2.5}$, $PM_{2.5}$ Absorbance, PM_{10} and PMcoarse in 20 European Study Areas; Results of the ESCAPE
Project. *Environmental Science & Technology*; 46; 20:11195-11205.

ESCAPE Study. 2010. ESCAPE Exposure Assessment Manual, Version 9. Available at: www.escapeproject.eu.

Gallant J, Dowling T, Read A, Wilson N, Tickle P, Inskeep C. 2011. 1 second SRTM Derived Digital Elevation
Models User Guide. Geoscience Australia. Available from: www.ga.gov.au/topographic-mapping/digital-
elevation-data.html. Accessed: 6 June 2015.

Garden FL, Toelle BG, Mihrshahi S, Webb KL, Almqvist C, et al. 2018. Cohort profile: The Childhood Asthma
Prevention Study (CAPS). *International Journal of Epidemiology*; 25 May 2018:doi.org/10.1093/ije/dyy1078.

Geddes JA, Martin RV, Boys BL, van Donkelaar A. 2016. Long-term trends worldwide in ambient NO_2
concentrations inferred from satellite observations. *Environmental Health Perspectives*; 124; 3:281-289.

Giavarina D. 2015. Understanding Bland Altman analysis. *Biochemia Medica*; 25; 2:141-151.

Gonzales M, Myers O, Smith L, Olvera HA, Mukerjee S, et al. 2012. Evaluation of land use regression models for NO_2 in El Paso, Texas, USA. *Science of the Total Environment*; 432:135-142.

Gulliver J, de Hoogh K, Hansell A, Vienneau D. 2013. Development and back-extrapolation of NO_2 land use regression models for historic exposure assessment in Great Britain. *Environmental Science & Technology*; 47; 14:7804-7811.

Gurung A, Levy JI, Bell ML. 2017. Modeling the intraurban variation in nitrogen dioxide in urban areas in Kathmandu Valley, Nepal. *Environmental Research*; 155:42-48.

Hanigan IC, Williamson GJ, Knibbs LD, Horsley J, Rolfe MI, et al. 2017. Blending multiple nitrogen dioxide data sources for neighborhood estimates of long-term exposure for health research. *Environmental Science & Technology*; 51; 21:12473-12480.

Hansell AL, Rose N, Cowie CT, Belousova EG, Bakolis I, et al. 2014. Weighted road density and allergic disease in children at high risk of developing asthma. *PLoS ONE*; 9; 6.

HEI. 2010. Traffic Related Air Pollution: A Critical Review of the Literature on Emissions, Exposure, and Health Effects. HEI Special Report 17. Health Effects Institute, Boston MA. Panel on the Health Effects of Traffic-Related Air Pollution.

Henderson SB, Beckerman B, Jerrett M, Brauer M. 2007. Application of land use regression to estimate long-term concentrations of traffic-related nitrogen oxides and fine particulate matter. *Environmental Science & Technology*; 41; 7:2422-2428.

Hennig F, Sugiri D, Tzivian L, Fuks K, Moebus S, et al. 2016. Comparison of land-use regression modeling with dispersion and chemistry transport modeling to assign air pollution concentrations within the Ruhr area. *Atmosphere*; 7; 3.

Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, et al. 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment*; 42; 33:7561-7578.

Hoek G, Eeftens M, Beelen R, Fischer P, Brunekreef B, et al. 2015. Satellite NO_2 data improve national land use regression models for ambient NO_2 in a small densely populated country. *Atmospheric Environment*; 105:173-180.

Hystad P, Setton E, Cervantes A, Poplawski K, Deschenes S, et al. 2011. Creating national air pollution models for population exposure assessment in Canada. *Environ Health Perspect*; 119; 8.

Jerrett M, Arain MA, Kanaroglou P, Beckerman B, Crouse D, et al. 2007. Modeling the intraurban variability of ambient traffic pollution in Toronto, Canada. *Journal of Toxicology & Environmental Health Part A*; 70; 3-4:200-212.

Jerrett M, Burnett RT, Beckerman BS, Turner MC, Krewski D, et al. 2013. Spatial analysis of air pollution and mortality in California. *American Journal of Respiratory and Critical Care Medicine*; 188; 5:593-599.

Johnson M, Isakov V, Touma JS, Mukerjee S, Ozkaynak H. 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. *Atmospheric Environment*; 44; 30:3660-3668.

Karner AA, Eisinger DS, Niemeier DA. 2010. Near-roadway air quality: Synthesizing the findings from real-world data. *Environmental Science & Technology*; 44; 14:5334-5344.

Kharol SK, Martin RV, Philip S, Boys B, Lamsal LN, et al. 2015. Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America. *Atmospheric Environment*; 118; Supplement C:236-245.

Knibbs L, Hewson M, Bechle M, Marshall J, Barnett A. 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. *Environmental Research*; 135; 0:204-211.

Knibbs LD, Coorey CP, Bechle MJ, Cowie CT, Dirgawati M, et al. 2016. Independent validation of national satellite-based land-use regression models for nitrogen dioxide using passive samplers. *Environmental Science & Technology*; 50; 22:12331-12338.

Koo TK, Li MY. 2016. A guideline of selecting and reporting Intraclass Correlation Coefficients for reliability research. *Journal of Chiropractic Medicine*; 15; 2:155-163.

Larkin A, Geddes JA, Martin RV, Xiao Q, Liu Y, et al. 2017. Global land use regression model for nitrogen dioxide air pollution. *Environmental Science & Technology*; 51; 12:6957-6964.

Marcon A, de Hoogh K, Gulliver J, Beelen R, Hansell AL. 2015. Development and transferability of a nitrogen dioxide land use regression model within the Veneto region of Italy. *Atmospheric Environment*; 122:696-704.

Marks GB, Mihrshahi S, Kemp AS, Tovey ER, Webb K, et al. 2006. Prevention of asthma during the first 5 years of life: A randomised controlled trial. *Journal of Allergy & Clinical Immunology*; 118:53-61.

Marshall JD, Nethery E, Brauer M. 2008. Within-urban variability in ambient air pollution: Comparison of estimation methods. *Atmospheric Environment*; 42; 6:1359-1369.

Novotny EV, Bechle MJ, Millet DB, Marshall JD. 2011. National satellite-based land-use regression: NO_2 in the United States. *Environmental Science & Technology*; 45; 10:4407-4414.

NSW EPA. 2012. Air Emissions Inventory for the Greater Metropolitan Region in New South Wales. 2008 Calendar Year. Consolidated Natural and Human-Made Emissions: Result. NSW Environment Protection Authority. ISBN 978-1-74293-556-0. EPA 2012/0255.

NSW EPA. 2016. Consultation Paper. Clean Air for NSW. ISBN 978 1-76039-420-2. NSW Environment Protection Authority and Office of Environment and Heritage on behalf of the State of NSW. Sydney, Australia.

Ozkaynak H, Baxter L, Dionisio K, Burke J. 2013. Air pollution exposure prediction approaches used in air pollution epidemiology studies. *Journal of Exposure Science & Environmental Epidemiology*:1-7.

Rahman MM, Yeganeh B, Clifford S, Knibbs LD, Morawska L. 2017. Development of a land use regression model for daily NO_2 and NOx concentrations in the Brisbane metropolitan area, Australia. *Environmental Modelling & Software*; 95:168-179.

Rose N, Cowie C, Gillett R, Marks GB. 2011. Validation of a spatiotemporal land use regression model incorporating fixed site monitors. *Environmental Science & Technology*; 45; 1:294-299.

Ryan P, LeMasters G. 2007. A review of land-use regression models for characterizing intraurban air pollution exposure. *Inhal Toxicol*; 19 (Suppl 1):127-133.

Sellier Y, Galineau J, Hulin A, Caini F, Marquis N, et al. 2014. Health effects of ambient air pollution: Do different methods for estimating exposure lead to different results? *Environment International*; 66:165-173.

Toelle BG, Garden FL, Ng KKW, Belousova EG, Almqvist C, et al. 2013. Outcomes of the Childhood Asthma Prevention Study at 11.5 years. *Journal of Allergy and Clinical Immunology*; 132; 5:1220-1222.e1223.

van Donkelaar A, Martin RV, Brauer M, Boys BL. 2015. Use of satellite observations for long-term exposure assessment of global concentrations of fine particulate matter. *Environ Health Perspect*; 123; 2:135-143.

Vienneau D, de Hoogh K, Bechle MJ, Beelen R, van Donkelaar A, et al. 2013. Western European land use regression incorporating satellite- and ground-based measurements of NO_2 and PM_{10}. *Environmental Science & Technology*; 47; 23:13555-13564.

Wang M, Beelen R, Eeftens M, Meliefste K, Hoek G, Brunekreef B. 2012. Systematic evaluation of land use regression models for NO_2. *Environmental Science & Technology*; 46; 8:4481-4489.

Wang M, Gehring U, Hoek G, Keuken M, Jonkers S, et al. 2015. Air pollution and lung function in Dutch children: A comparison of exposure estimates and associations based on land use regression and dispersion exposure modeling approaches. *Environmental Health Perspectives*; 123; 8:847-851.

Wang M, Brunekreef B, Gehring U, Szpiro A, Hoek G, Beelen R. 2016. A new technique for evaluating land use regression models and their impact on health effect estimates. *Epidemiology*; 27; 1:51-56.

Wu J, Wilhelm M, Chung J, Ritz B. 2011. Comparing exposure assessment methods for traffic-related air pollution in an adverse pregnancy outcome study. *Environmental Research*; 111; 5:685-692.

Young MT, Bechle MJ, Sampson PD, Szpiro AA, Marshall JD, et al. 2016. Satellite-based NO_2 and model validation in a national prediction model based on universal kriging and land-use regression. *Environmental Science & Technology*; 50; 7:3686-3694.

Zenith Traffic Model. 2014. Zenith Traffic Estimates. Veitch-Lister Consulting Pty Ltd. *Available at: http://wwwveitchlistercomau/zenith/traffic-estimates [Accessed: 10 September 2014].*

Highlights

- We developed an intra-city LUR model for Sydney which predicted 84% and 91% of NO_2 and NOx variability.

- This study compared statistical agreement of NO_2 estimates from three models with varying spatial scales and inputs.

- The three models were an intra-city LUR, a Satellite-based LUR, and a Bayesian blended model.

- We report good agreement and small absolute differences between the three models.

- The models showed good agreement despite the varying spatial scales and inputs to the models.