Environmental Research 163 (2018) 16–25

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

Long-term nitrogen dioxide exposure assessment using back-extrapolation of satellite-based land-use regression models for Australia

Luke D. Knibbs[a,b,*], Craig.P. Coorey[a], Matthew J. Bechle[c], Julian D. Marshall[c], Michael G. Hewson[d], Bin Jalaludin[b,e,f,g], Geoff G. Morgan[b,h], Adrian G. Barnett[i]

[a] Faculty of Medicine, The University of Queensland, Herston, QLD 4006, Australia
[b] Centre for Air Quality and Health Research and Evaluation, Glebe, NSW 2037, Australia
[c] Department of Civil and Environmental Engineering, University of Washington, Seattle 98195, WA, USA
[d] School of Education and the Arts, Central Queensland University, Rockhampton, QLD 4700, Australia
[e] Population Health, South Western Sydney Local Health District, Liverpool, NSW 2170, Australia
[f] Ingham Institute for Applied Medical Research, Liverpool, NSW 2170, Australia
[g] School of Public Health and Community Medicine, The University of New South Wales, Kensington, NSW 2052, Australia
[h] University Centre for Rural Health, School of Public Health, The University of Sydney, Lismore, NSW 2480, Australia
[i] School of Public Health and Social Work, Queensland University of Technology, Kelvin Grove, QLD 4059, Australia

ARTICLE INFO

Keywords:
Exposure assessment
Back-extrapolation
Air pollution
Long-term
Epidemiology

ABSTRACT

Assessing historical exposure to air pollution in epidemiological studies is often problematic because of limited spatial and temporal measurement coverage. Several methods for modelling historical exposures have been described, including land-use regression (LUR). Satellite-based LUR is a recent technique that seeks to improve predictive ability and spatial coverage of traditional LUR models by using satellite observations of pollutants as inputs to LUR. Few studies have explored its validity for assessing historical exposures, reflecting the absence of historical observations from popular satellite platforms like Aura (launched mid-2004). We investigated whether contemporary satellite-based LUR models for Australia, developed longitudinally for 2006–2011, could capture nitrogen dioxide (NO_2) concentrations during 1990–2005 at 89 sites around the country. We assessed three methods to back-extrapolate year-2006 NO_2 predictions: (1) 'do nothing' (i.e., use the year-2006 estimates directly, for prior years); (2) change the independent variable 'year' in our LUR models to match the years of interest (i.e., assume a linear trend prior to year-2006, following national average patterns in 2006–2011), and; (3) adjust year-2006 predictions using selected historical measurements. We evaluated prediction error and bias, and the correlation and absolute agreement of measurements and predictions using R^2 and mean-square error R^2 (MSE-R^2), respectively. We found that changing the year variable led to best performance; predictions captured between 41% (1991; MSE-R^2 = 31%) and 80% (2003; MSE-R^2 = 78%) of spatial variability in NO_2 in a given year, and 76% (MSE-R^2 = 72%) averaged over 1990–2005. We conclude that simple methods for back-extrapolating prior to year-2006 yield valid historical NO_2 estimates for Australia during 1990–2005. These results suggest that for the time scales considered here, satellite-based LUR has a potential role to play in long-term exposure assessment, even in the absence of historical predictor data.

1. Introduction

Exposure assessment in studies of long-term health effects of air pollution is often hampered by sparse or missing measurements (Hart et al., 2009; Hystad et al., 2012). This challenge is most pronounced in studies of multi-decadal exposures, which is one reason why there are fewer studies focused on them compared with the relatively large body of evidence on shorter-term exposures (Hansell et al., 2016). One option for addressing these limitations is land-use regression (LUR) and other air pollution modelling techniques. LUR is a frequently used method for assigning exposures in epidemiological studies. It uses environmental predictors (such as nearby road length, traffic volume and land use categories) to capture variability in measured pollutant concentrations, and can then be applied to estimate concentrations at unmeasured locations (Hoek et al., 2008; Marhsall et al., 2008).

Traditionally, most LUR models were developed for specific cities and their applicability to other locations was limited, which constrained their use in national- or multi-national health studies (Allen

* Correspondence to: School of Public Health, Faculty of Medicine, The University of Queensland, Herston, QLD 400, Australia.
E-mail address: l.knibbs@uq.edu.au (L.D. Knibbs).

https://doi.org/10.1016/j.envres.2018.01.046
Received 1 November 2017; Received in revised form 5 January 2018; Accepted 30 January 2018

et al., 2011; Briggs, 2007; Poplawski et al., 2009). Recently, several studies have incorporated satellite-derived observations of pollutants and other predictors of ground-level pollutants (e.g., impervious surfaces and tree cover). These satellite-based LUR models can potentially serve the dual purpose of improving predictive ability and extending spatial coverage compared with traditional LUR (Jerrett et al., 2017), which has led to national-and multi-national models for nitrogen dioxide (NO_2), $PM_{2.5}$ ($< 2.5\,\mu m$) and PM_{10} ($< 10\,\mu m$) (Bechle et al., 2015; Beckerman et al., 2013; de Hoogh et al., 2016; Hoek et al., 2015; Hystad et al., 2011; Knibbs et al., 2014; Novotny et al., 2011; Vinneau et al., 2013; Young et al., 2016). Notably, the technique has recently been used to develop a global model for NO_2 that captured 54% of spatial variation in 2011 mean concentrations (Larkin et al., 2017).

Previous studies have demonstrated a role for LUR in historical NO_2 exposure assessment, either through development of models using historical predictor data (i.e., to match or approximate the year(s) of interest) or, when this is not feasible, via back-extrapolation of estimates from more recent models (e.g., Beelen et al., 2007; Cesaroni et al., 2012; Chen et al., 2010; Eeftens et al., 2011; Gulliver et al., 2013; Gulliver et al., 2016; Levy et al., 2015; Wang et al., 2013). However, despite the potential benefits of satellite-based LUR models their validity for historical exposure assessment has received limited attention (Hystad et al., 2012). This aspect of satellite-based LUR remains largely unexplored, perhaps reflecting the absence of historical, high spatial resolution satellite data. For example, the ozone monitoring instrument (OMI) aboard the Aura satellite is a popular source of NO_2 observations and was launched in mid-2004.

In this study, we sought to evaluate the ability of national satellite-based LUR models for Australia to capture historical levels of NO_2 using multiple back-extrapolation methods. We aimed to add to the limited literature on historical estimation of NO_2; most studies have been performed in North America and Western Europe using relatively dense monitoring networks, and only one study used satellite data (Hystad et al., 2012). Australia provides a useful contrast to these other locations because of its continental scale, highly urbanised and concentrated population, and relatively scant temporal and spatial coverage from the ground-based NO_2 monitoring network.

2. Methods

2.1. Overview of satellite-based LUR models

We previously developed satellite-based LUR models for annual mean NO_2 using generalised estimating equations (GEEs) fit to data from the 68 continuous regulatory chemiluminescence monitors operating throughout Australia during 2006–2011 (population = 24.5 million; area = 7.7 million km^2; ~0.3 NO_2 monitors/100,000 persons; ~0.9 monitors/100,000 km^2). The models were used to predict annual NO_2 for each year during that period; their development and validation are described in detail elsewhere (Knibbs et al., 2014, 2016). Briefly, we developed two models: one included the tropospheric column abundance of NO_2 molecules observed by the OMI spectrometer aboard the Aura satellite as a predictor (molecules per cm^2; 'column model'). The other model included the estimated ground-level NO_2 concentration (ppb; 'surface model'), based on also including a surface-to-column ratio from the Weather Research and Forecasting model coupled with Chemistry (WRF-Chem). We used five-fold cross-validation with five replications and found that our column and surface models, respectively, explained 81% (RMSE = 1.4 ppb) and 79% (RMSE = 1.4 ppb) of spatial variability in annual mean NO_2 across Australia during 2006–2011 (Knibbs et al., 2014).

We subsequently evaluated model performance using an independent data set of passive samplers deployed during 2006–2014. We found the column and surface models, respectively, captured 66% (RMSE = 2.0 ppb) and 69% (RMSE = 2.0 ppb) of spatial variability in annual NO_2 at 98 non-roadside sites (Knibbs et al., 2016). The present study builds on those analyses by exploring the models' ability to capture historical levels of annual NO_2, and determine their validity for assigning multi-decadal exposures in cohort studies of health effects.

2.2. Measurement data for historical validation

We contacted the eight agencies responsible for regulatory air quality monitoring across Australia's six states and two territories. We obtained daily NO_2 concentrations (ppb) from all monitoring sites during 1990–2005, provided: (a) measurements were performed for at least one calendar year; (b) a calibrated chemiluminescence monitor compliant with Australian Standard 3580.5.1–1993 was used (SAI Global, 2017); (c) data were subject to quality assurance (QA) procedures, and; (d) coordinates for the site location were known to at least five decimal places. Although NO_2 had been measured in some Australian capital cities as early as the 1960s, most cities had either no monitoring or only a single site throughout the 1970s and 1980s, and measurement techniques and frequency were inconsistent (Cleary, 1969; National Environment Protection Council, 2000). We therefore selected 1990 as our earliest year because Australia's NO_2 monitoring network underwent substantial expansion in the early-to-mid 1990s prior to the introduction of the first national air quality standards in 1998 (National Environment Protection Council, 1998). For the present study, we used 2005 as our last year because the models were developed using data from 2006 to 2011, and previously validated for 2006–2014 (Knibbs et al., 2014, 2016). That time frame allowed us to assess our models' historical performance over the 16-year period (1990–2005) prior to the 6-year period they were developed for.

We obtained data from 90 monitoring sites. To our knowledge, they represent all regulatory monitors that met our inclusion criteria. The sites spanned six of Australia's eight states and territories; no historical data were available for Tasmania or the Northern Territory, which are the smallest state and territory by population, respectively. Many of the sites had been used to develop our LUR models for 2006–2011 (Knibbs et al., 2014). Because of the sparse Australian monitoring network, we did not exclude these sites but instead undertook sensitivity analyses to assess the influence of model development sites and non-development sites on our validation results, which are described in Section 2.6.

Seven monitoring sites, all in major cities, had been relocated between 0.2 and 2.2 km from their original location during the study period, of which one site had been relocated twice (0.5 and 1.0 km, respectively). Because NO_2 can be spatially heterogeneous over such distances in urban areas, we treated the pre- and post-relocation measurements as being from different sites (Gilbert et al., 2003; Marshall et al., 2008; Pleijel et al., 2004; Roorda-Knape et al., 1999). This yielded 98 sites available for further analyses.

2.3. Processing of measurements

We sought to maximise inclusiveness while minimising the potential for seasonal bias due to missing data. We therefore included sites with 50% or greater non-missing daily NO_2 observations in a given year, provided there was at least one month of valid data per season (Hystad et al., 2011). As we were interested in assessing our LUR models' ability to capture long-term average concentrations, we also recorded sites that had 50% or greater non-missing data during 1990 through 2005, provided: (a) at least two years of valid data were collected in the first (1990–1997) and second (1998–2005) eight years of our sixteen-year study period, respectively, and; (b) of these, at least one month of data was collected per season per year. We used this approach as a balance between seeking to include a sufficiently large number of sites, but without compromising the ability to capture changes in NO_2 over the study period. We undertook sensitivity analyses to assess the stability of long-term NO_2 trends and the effects of using more stringent site inclusion criteria on our results (i.e., requiring 60%, 70%, or 80% of data to be non-missing).

2.4. Monitoring site classification

We first defined monitoring sites using standard Australian Bureau of Statistics (ABS) criteria; sites were classified as urban if they were located in the greater metropolitan area of a capital city or were located in an ABS-defined significant urban area with a population > 20,000 people in the 2011 census (Australian Bureau of Statistics, 2011). We classified sites that did not meet those criteria in a single category called 'rural and remote'. We then classified the remaining urban sites as either: (1) 'roadside' (≤15 m to a major road); (2) 'urban traffic' (not roadside, but ≤100 m to a major road), or; (3) 'urban background' (not roadside or urban traffic, and > 100 m to a major road). The distance thresholds and classifications were based on our previous validation study (Knibbs et al., 2016), which in turn was informed by the exponential decay in NO_2 observed near major roads, and which has a half-life of approximately 100 m (Gilbert et al., 2003; Karner et al., 2010; Pleijel et al., 2004; Roorda-Knape et al., 1999). We manually assessed all borderline sites close to each distance cut-off using Google Earth and Street View before classifying them. Our ninety-eight sites comprised four roadside sites (4%), nine urban traffic sites (9%), eighty-two urban background sites (84%) and three rural and remote sites (3%). We examined the sensitivity of our site classifications to both the distance thresholds for each category and the definition of major roads using our previous methods (Knibbs et al., 2016).

We used the Australian National Pollutant Inventory to assess sites' proximity to industrial emissions of nitrogen oxides (Department of the Environment and Energy, 2017a). None of the sites were located < 300 m from an industrial point source, and 93% of sites were located > 1 km based on point source data from 1999 to 2005; point source data prior to 1999 were not available.

2.5. LUR model predictions

We sought to use our surface and column LUR models to predict NO_2 at the measurement sites and assess the models' ability to capture spatial variability in annual NO_2 year-by-year and averaged over the whole 16-year study period (1990–2005). However, the large majority of our LUR predictors were not available historically. For example, satellite-derived tropospheric column observations and surface estimates of NO_2, which contribute the most and second-most predictive ability to our column and surface models, respectively, were unavailable because the Aura satellite was launched mid-2004 (Knibbs et al., 2014).

We instead employed the 'present-day' LUR predictors used to develop both LUR models for 2006–2011. Although the satellite NO_2 predictors corresponded to each year in that period, the specific year for other predictors was dependent on data availability. For example, roads data were from 2013, land-use categories were from 2011, impervious surfaces were from 2000 to 2001, and industrial NO_X emissions were from 2008 to 2009. Details on data sources and processing have been reported previously (Knibbs et al., 2014).

2.6. Back-extrapolation of predictions

We then explored three back-extrapolation approaches for estimating historical NO_2. First, we used predicted NO_2 for the earliest year in our models' development (2006) unmodified (i.e., 'do nothing' approach). Second, we set the independent variable 'year' in our LUR models to each year between 1990 and 2005. Because our LUR models were developed longitudinally for 2006–2011 using GEEs, we were able to assess how well they captured NO_2 prior to this period (Knibbs et al., 2014). Using this method, all other LUR predictors were the same as the 'do nothing' approach, but the temporal gradient of modestly decreasing NO_2 observed during 2006–2011 when developing our models was back-extrapolated to 1990. This approach corresponded to a 0.14 and 0.16 ppb change in estimated NO_2 (higher concentrations in earlier years) for the column and surface models, respectively (i.e., a 2.1 ppb [column model] and 2.4 ppb [surface model] increase in NO_2 from 2005 to 1990).

Our third method back-extrapolated predictions for year-2006 using NO_2 measurements at sites that monitored continuously during 1990–2005. We calculated the ratio of annual mean NO_2 in each of the sixteen years to that in 2006 and used it to adjust the year-2006 predictions using standard methods (Bechle et al., 2015; Gulliver et al., 2013). We based the ratio on four sites spanning three states (New South Wales, Victoria and Western Australia). Two sites were in urban background locations, one was at the urban-rural fringe and one was in a rural area. Our selection criteria for the sites are described in the supplement (page S1). We also calculated the absolute difference (in ppb) between NO_2 in 2006 and each year during 1990–2005, to assess whether using it to adjust the predictions yielded better results than the ratio method (Gulliver et al., 2013). For both methods, we tested whether four sites combined were preferable to two urban or two non-urban sites separately, following the approach used by Gulliver et al. (2013). The measurement sites used for adjustment of predictions were excluded from subsequent validation analyses of this back-extrapolation method.

For all three back-extrapolation methods, we used year-2006 as the basis for our LUR predictions. We also investigated whether this was comparable to 2006–2011 overall (i.e., the average for the whole period covered by our models). The first approach used column and surface estimates of NO_2 for 2006 only, while the second used the average during 2006–11; all other LUR predictors were the same. Finally, to further assess LUR performance over time, we compared results for each back-extrapolation method in a given year to those in four 4-year blocks comprising our study period (1990–1993; 1994–1997; 1998–2001; 2002–2005).

2.7. Model development and out-of-range sites

Because sites used to develop the LUR models also comprised the majority of validation sites in most years, excluding them would have resulted in too few sites for analysis. In a sensitivity analysis, we instead excluded the sites not used for development as an indirect assessment of the influence of model development sites. We excluded sites with values of one or more LUR predictor outside the range of values observed at sites used to develop our models. We took that approach to prevent unrealistically high or low predictions, following methods used by Wang et al. (2012). We excluded 9 out of 98 sites (two roadside, four urban traffic, two urban background, and one rural and remote), leaving 89 available for validation provided they also met our inclusion criteria for NO_2 completeness. We assessed the effect of excluding these sites by comparing our results to those with the sites included.

2.8. Site representativeness

We used two approaches to assess the representativeness of our validation sites for estimating NO_2 exposures in Australia. We first compared the distribution of LUR predictors at the validation sites to those at ~345,000 census 'mesh block' centroids covering all of Australia. Mesh blocks are the smallest spatial unit used in the national census, and contain 62 people on average (Australian Bureau of Statistics, 2011). We then undertook a more rigorous assessment by comparing the validation sites with the geocoded national address file (G-NAF). The G-NAF contains ~14.1 million points, representing all addresses in Australia, by combining and comparing land, postal and electoral records (Public Sector Mapping Agencies, 2013). We restricted our comparison to the ~8.4 million address points that were matched in all three of these administrative databases, and which had geocoding reliability sufficient to assign a centroid within an address parcel boundary using standard G-NAF QA criteria (Public Sector Mapping Agencies, 2013).

2.9. Validation of predictions

We used established LUR validation methods, and regressed predicted NO_2 on measured NO_2, assessing model performance on the basis of R^2, RMSE (percentage and absolute), and bias (fractional and absolute) (Basagaña et al., 2012; Johnson et al., 2010; Gulliver et al., 2013; Wang et al., 2012, 2016). We evaluated model predictions for annual NO_2 in each year during 1990–2005, and also the average over the entire 16-year period. We assessed the normality and variance of residuals, and tested spatial correlation among residuals using Moran's *I*.

While R^2 is a standard metric for evaluating an LUR model's performance, it is based on the correlation between measurements and predictions rather than their absolute agreement. A high R^2 alone does not always imply the validity of model predictions in epidemiological studies (Szpiro et al., 2011; Wang et al., 2012). Therefore, we also evaluated absolute agreement using mean-square-error R^2 (MSE-R^2), which describes how well the relationship between measurements and predictions follows a 1:1 line (Basagaña et al., 2012; Gulliver et al., 2013; Szpiro et al., 2011; Wang et al., 2012). Unlike R^2, MSE-R^2 yields negative values when the average of measurements has a lower MSE than the predictions (Basagaña et al., 2012; Szpiro et al., 2011). We used R (v. 3.2.2) for our analyses (R Project for Statistical Computing, Vienna, Austria).

3. Results

3.1. NO_2 measurements

Fig. 1 shows the location of the 89 validation sites around Australia. Between 14 (16%, 1991) and 62 (70%, 2003) sites met our inclusion criteria for annual NO_2 in each year during 1990–2005; in total they collected 690 site-years of measurements over that period. Of the sites that met the inclusion criteria for a given year, between 6 (43%, 1991) and 51 (89%, 2005) had been used to develop our LUR models. Forty-five sites (51%) met our inclusion criteria for long-term average NO_2 over the 16-year study period, and they collected 524 site-years of measurements. Thirty-three of these (73%) had been used to develop our LUR models.

Descriptive statistics for measured NO_2 are presented in Table 1. We observed a modest but persistent decrease in NO_2, averaging ~0.22 ppb per year during the period with the greatest number of monitoring sites (1993–2005) (Figs. S1, S2). This value was consistent with the ~0.15 ppb annual decrease observed for 2006–2011 when developing our LUR models (Knibbs et al., 2014).

3.2. LUR model performance

Our unmodified LUR column model predictions for year-2006 (i.e., 'do nothing' approach) captured between 41% (1991, MSE $R^2 = -8\%$) and 80% (2003, MSE-$R^2 = 77\%$) of spatial variability in annual NO_2 (Fig. 2, Table S1). Prediction error was greatest for earlier years, with RMSE ranging from 5.4 ppb (42%) in 1990 to 1.6 ppb (20%) in 2004. On average, model predictions underestimated measured concentrations, and absolute bias ranged from −3.6 ppb in 1992 (fractional bias = −0.3) to −0.05 ppb in 2004 (fractional bias = −0.01). We observed very similar performance for the surface model (Fig. 2); the full results are presented in the supplement (Table S2). Column and surface models, respectively, captured 75% (MSE-R^2 = 61%) and 76% (MSE-R^2 = 57%) of spatial variation in annual NO_2 averaged over 1990–2005 (Fig. 3). Their prediction biases were −1.3 ppb and −1.5 ppb, respectively, and with RMSE of 2.4 ppb (26%) and 2.5 ppb (27%) (Table 3).

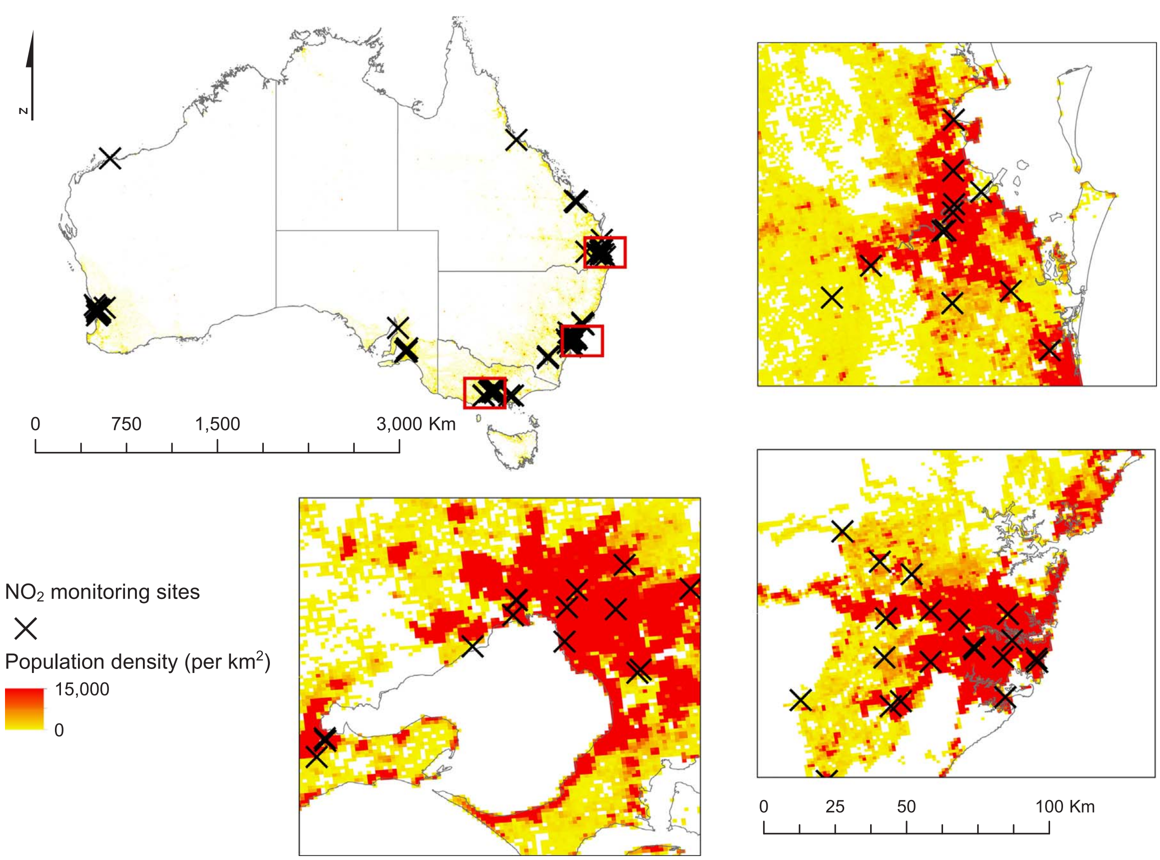

Fig. 1. The 89 NO_2 monitoring sites around Australia used for historical validation (cross symbols). The sites are shown against a 1 km × 1 km national grid of population density produced by the Australian Bureau of Statistics using data from the 2011 census (Australian Bureau of Statistics, 2014). The three insets show (clockwise) Brisbane, Sydney, and Melbourne; the three most populous cites in Australia.

Table 1
Descriptive statistics for measured daily NO_2 (ppb) by year and overall.

Year	*n* sites	Mean	S.D.	Min.*	5th	25th	50th	75th	95th	Max.
1990	15	13.2	8.6	0.1	2.0	6.8	11.7	17.9	30.0	57.0
1991	14	12.8	8.1	0.1	2.6	6.8	11.3	17.0	29.0	54.0
1992	15	13.1	8.4	0.1	2.7	6.8	11.5	18.0	29.0	63.1
1993	33	10.3	7.7	0.1	1.6	5.0	8.7	13.9	24.9	71.6
1994	38	9.9	6.8	0.1	1.8	5.0	8.4	13.5	23.2	53.0
1995	41	10.3	6.3	0.1	2.0	5.4	9.1	14.0	22.1	46.7
1996	41	8.9	5.6	0.1	1.7	4.7	8.0	12.2	19.3	46.0
1997	49	9.5	6.4	0.1	1.7	4.7	8.1	13.3	21.7	46.0
1998	48	8.9	6.1	0.1	1.2	4.1	7.8	12.4	20.6	52.0
1999	54	8.5	5.8	0.1	1.0	4.0	7.4	12.0	19.4	36.9
2000	51	8.2	5.6	0.1	1.0	4.0	7.0	11.4	18.8	38.1
2001	56	8.5	5.7	0.1	1.0	4.0	7.4	12.0	19.4	37.7
2002	57	8.2	5.9	0.1	1.0	3.9	7.0	11.3	19.9	39.0
2003	62	8.1	5.6	0.1	1.0	4.0	7.0	11.1	19.0	40.9
2004	59	7.7	5.3	0.1	1.1	3.9	6.7	10.5	18.0	39.0
2005	57	7.8	5.3	0.1	1.3	4.0	6.6	10.8	18.2	36.0
1990–2005	45	9.3	6.3	0.1	1.6	4.6	8.0	12.9	21.1	83.5

* Minimum refers to lowest positive value.

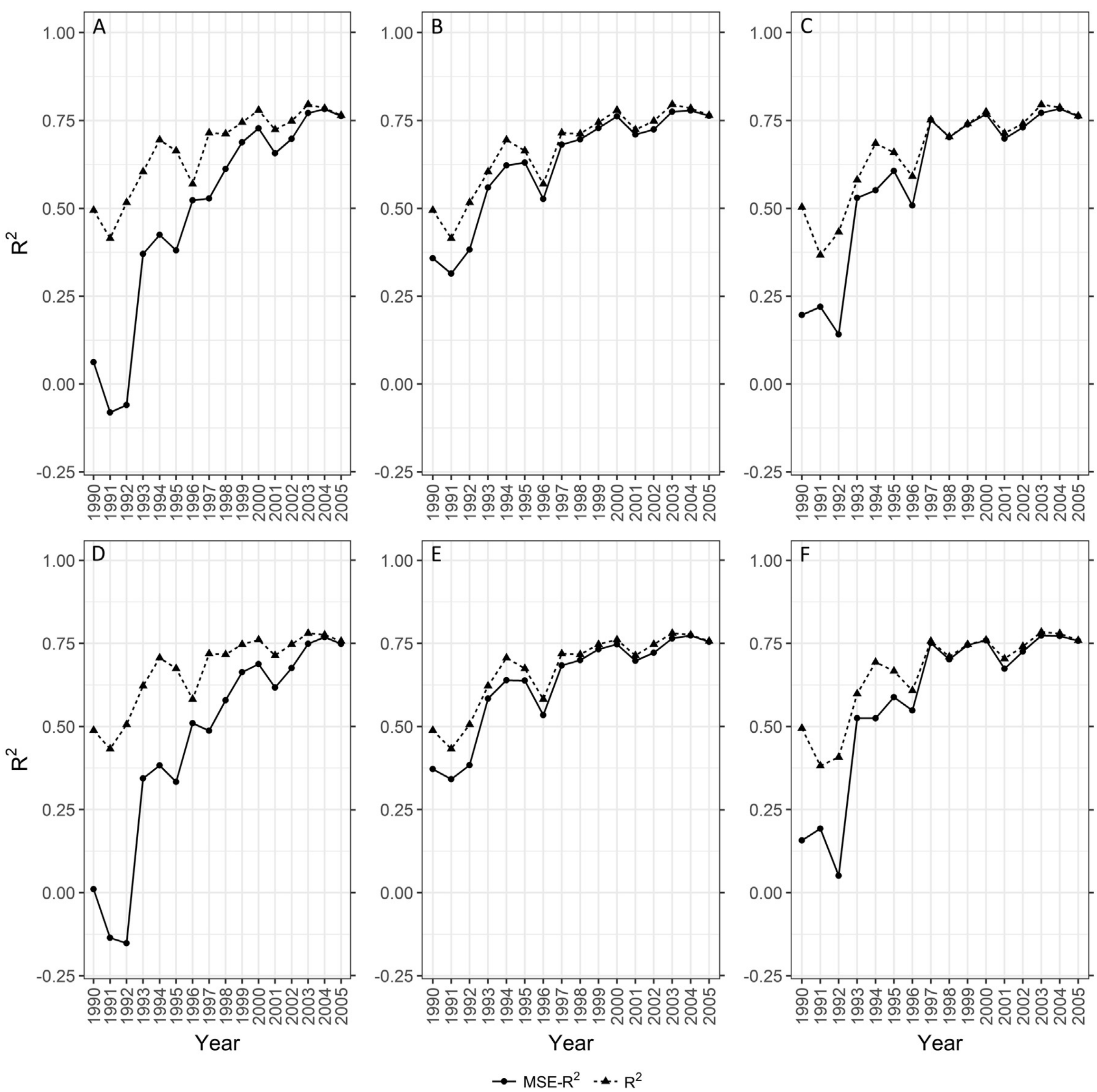

Fig. 2. LUR model performance by year measured by R^2 and MSE-R^2. The top row shows column model predictions using the 'do nothing' (A), change year predictor (B), and adjust using measurement (C) approaches. The bottom row shows the same results for the surface model (D, E, and F). Full validation statistics are in Table 2 and the supplement.

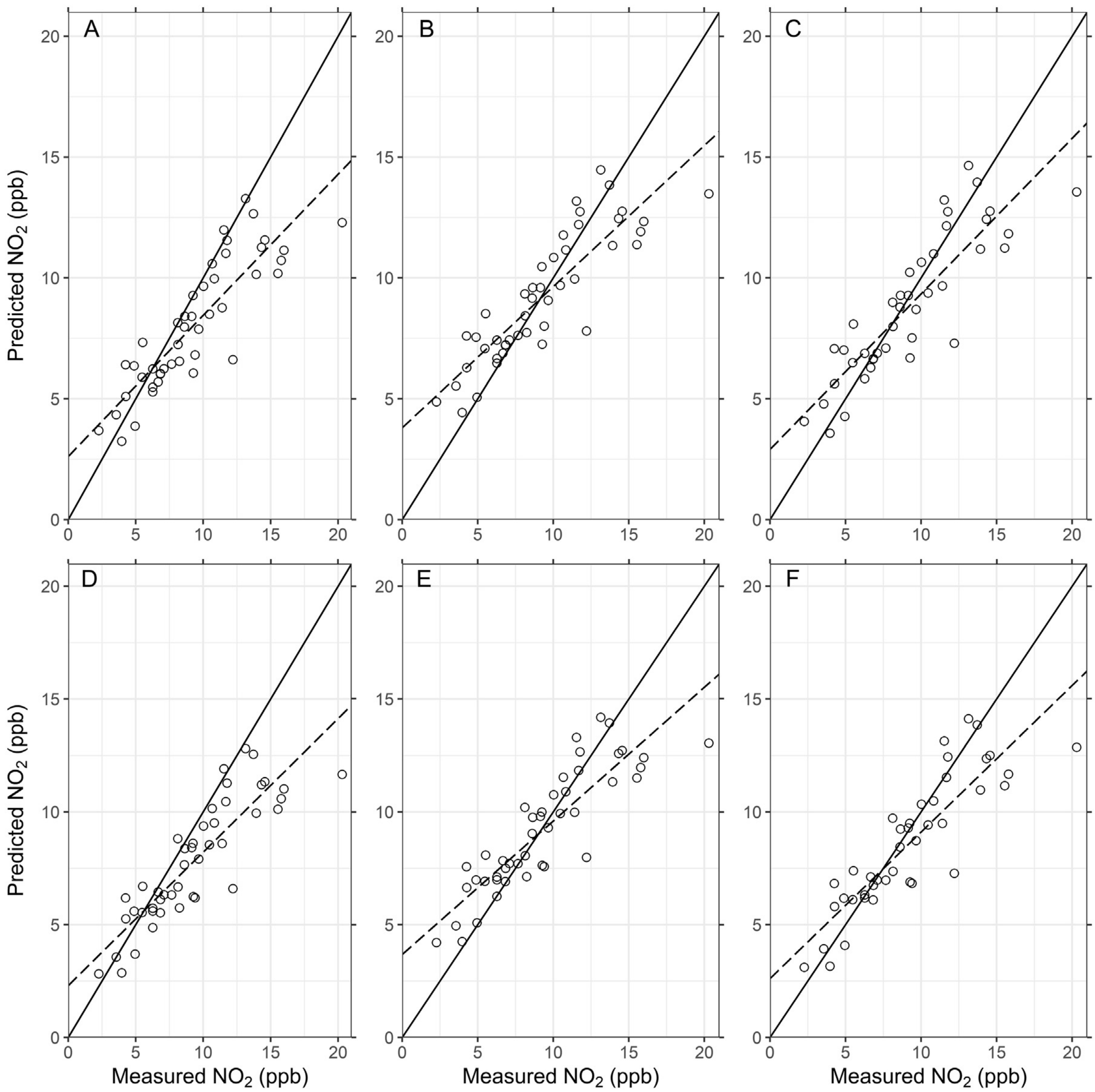

Fig. 3. Measured vs. LUR-predicted annual NO_2 averaged over 1990–2005 at 45 sites. The top row shows column model predictions using the 'do nothing' (A), change year predictor (B), and adjust using measurement (C) approaches. The bottom row shows the same results for the surface model (D, E and F). The dashed line is the regression line and the solid line is the line of agreement. Full validation statistics are in Table 3.

Table 2
Validation statistics by year for column model predictions after changing the 'year' predictor.

Year	*n* sites	R^2	β (95% CI)	Int.	MSE-R^2	RMSE (ppb)	RMSE (%)	Bias (ppb)	FB (–)
1990	15	0.49	0.26 (0.13, 0.39)	8.6	0.36	4.5	34.6	−0.9	−0.08
1991	14	0.41	0.30 (0.12, 0.48)	7.9	0.31	4.1	31.7	−1.3	−0.10
1992	15	0.52	0.41 (0.21, 0.60)	6.2	0.38	3.8	28.5	−1.6	−0.13
1993	33	0.60	0.45 (0.34, 0.56)	5.4	0.56	3.2	31.0	−0.3	−0.03
1994	38	0.69	0.48 (0.39, 0.57)	4.8	0.62	2.8	27.9	−0.4	−0.04
1995	41	0.66	0.54 (0.44, 0.65)	4.3	0.63	2.3	22.7	−0.4	−0.04
1996	41	0.57	0.60 (0.46. 0.74)	4.3	0.53	2.3	25.9	0.7	0.07
1997	49	0.71	0.58 (0.49, 0.67)	3.6	0.68	2.3	24.3	−0.4	−0.04
1998	48	0.71	0.61 (0.51, 0.70)	3.5	0.70	2.2	24.7	−0.1	−0.01
1999	54	0.74	0.64 (0.55, 0.73)	3.2	0.73	2.0	23.8	0.2	0.02
2000	51	0.78	0.67 (0.58, 0.75)	2.9	0.76	1.8	22.5	0.1	0.02
2001	56	0.72	0.63 (0.54, 0.72)	2.9	0.71	2.0	23.8	−0.2	−0.02
2002	57	0.75	0.61 (0.53, 0.69)	3.1	0.72	2.1	25.2	−0.1	−0.01
2003	62	0.80	0.67 (0.60, 0.75)	2.8	0.78	1.8	22.0	0.1	0.02
2004	59	0.78	0.74 (0.66, 0.83)	2.2	0.78	1.6	20.4	0.2	0.03
2005	57	0.76	0.73 (0.64, 0.82)	2.2	0.76	1.6	20.7	0.1	0.01

Note FB = fractional bias. The equivalent results for the surface model and other back-extrapolation methods are presented in the supplement.

Table 3
Validation statistics for model predictions of annual NO_2 averaged over 1990–2005.

Model	Back-extrapolation	*n* sites	R^2	β (95% CI)	Int.	MSE-R^2	RMSE (ppb)	RMSE (%)	Bias (ppb)	FB (–)
Column	Do nothing	45	0.75	0.58 (0.50, 0.67)	2.6	0.61	2.4	26.1	−1.3	−0.15
Surface	Do nothing	45	0.76	0.59 (0.51, 0.68)	2.3	0.57	2.5	27.3	−1.5	−0.18
Column	Change year	45	0.75	0.58 (0.50, 0.67)	3.8	0.72	2.1	22.3	−0.1	−0.01
Surface	Change year	45	0.76	0.59 (0.51, 0.68)	3.7	0.72	2.1	22.0	−0.1	−0.01
Column	Adjust with ratio	41	0.75	0.64 (0.54, 0.74)	2.9	0.73	2.0	22.1	−0.4	−0.04
Surface	Adjust with ratio	41	0.76	0.65 (0.55, 0.75)	2.6	0.72	2.1	22.5	−0.6	−0.07

Note FB = fractional bias.

When we changed the 'year' independent variable in our LUR models to match each year being assessed, we observed identical R^2 statistics to the 'do nothing' scenario because all predictions changed by the same amount of NO_2 in a given year (i.e., 0.14 and 0.16 ppb were added to year-2006 column and surface model predictions for each successively earlier year from 2005 back to 1990). However, we found improvements in the absolute agreement of measurements and predictions for both column and surface models that were more pronounced in earlier years (1990–1995), and which were associated with reduced prediction error and bias (Fig. 2, Table 2, Table S3). The largest prediction bias and RMSE were −1.6 ppb (1992, fractional bias = −0.1) and 4.5 ppb (1990, 35%), respectively, both of which were markedly better than under the 'do nothing' scenario. We also observed improved absolute agreement for annual NO_2 averaged over 1990–2005, with MSE-R^2 of 72%, bias of −0.1 ppb (fractional bias = −0.01), and RMSE of 2.1 ppb (22%) for both models (Fig. 3, Table 3).

When we applied an adjustment to the year-2006 'do nothing' LUR predictions using the ratio of NO_2 measured at four sites in each year to that in 2006, we found improved R^2 and MSE-R^2 values and reduced prediction error and bias compared with the 'do nothing' approach for both models (Fig. 2, Table S4, Table S5). However, the improvement in validation statistics for a given year was less marked than that observed when we changed the 'year' independent variable. The prediction of annual NO_2 averaged over 1990–2005 was substantially better than under the 'do nothing' scenario, and similar to the results when changing 'year' (Fig. 3, Table 3). Both the ratio- and difference-based adjustments yielded consistent results (Tables S6 and S7), and using four sites to derive a combined adjustment led to improved performance compared with two urban or two non-urban sites (not shown).

3.3. Sensitivity analyses

The results we present are based on year-2006 satellite predictors and sites with 50% or greater data, including both model development and non-development sites, but excluding out-of-range sites. This selection was informed by our sensitivity analyses, which are presented in the supplement. Briefly, our site classifications were robust to changing the definition of major roads and a halving of road distance thresholds (Table S8). Doubling the distance thresholds resulted in an increase in the number of sites classified as urban traffic and a corresponding reduction in sites classified as urban background. Adopting more stringent inclusion criteria for data completeness (i.e., from 50% to 60%, 70%, or 80%) reduced the number of sites available for validation in most years, but did not substantially change the validation statistics (Tables S9-S12). Using year-2006–2011 mean values of satellite predictors in the column and surface models yielded validation results that were similar to those when using year-2006 values (Tables S13-S14). We observed an improvement in validation statistics compared with our main analysis that included both development and non-development sites (Tables S15-S16). Excluding nine sites with predictors outside the range used to develop our LUR models led to better validation statistics (Tables S17-S18). Finally, model performance in four 4-year time blocks (1990–1993; 1994–1997; 1998–2001; 2002–2005) was comparable to that in a given year (Table S19).

3.4. Validation QA and representativeness

The residuals in our validation analyses had approximately normal distributions and constant variance (Figs. S3-S19). There was no evidence of overt spatial correlation (Table S20). We found that the percentiles of LUR predictors at the historical validation sites were comparable to those at both ~345,000 census block centroids and ~8.4 million Australian addresses in the G-NAF (Table S21).

4. Discussion

4.1. Overall findings

We evaluated the ability of national satellite-based LUR models, developed for 2006–2011, to capture spatial variability in historic annual NO_2 concentrations across Australia during 1990–2005. Because of the absence of historical satellite and other LUR predictor data, we explored three approaches for back-extrapolating LUR estimates for 2006. We observed the best performance when we changed the 'year' independent variable in our LUR models to match the year being estimated, which extrapolated the linear decrease in NO_2 during 2006–11 to 1990–2005. We suspect the reason why this method performed best is because of the consistency with which measured NO_2 concentrations have declined in Australia; the average annual decrease in NO_2 during the models' development (2006–2011) was similar to that during the historical validation period (1990–2005; see Fig. S1).

The next-best method was back-extrapolation of 2006 predictions using adjustment based on historical NO_2 measurements. We found that for both methods, the correlation (R^2) and absolute agreement (MSE-R^2) of measurements and predictions were similar, which is an important consideration when using LUR in epidemiological studies (Basagaña et al., 2012). Our column- (simpler) and surface-based (more complex) satellite LUR models yielded comparable results, which is in keeping with our previous results and supports the use of the simpler column approach (Knibbs et al., 2014, 2016; Bechle et al., 2015). The prediction error and bias we observed here was generally consistent with a previous independent validation for 2006–2014 (Knibbs et al., 2016).

4.2. Approaches to historical estimation

Despite the large body of literature documenting adverse health effects of outdoor air pollution, relatively few studies have assessed outcomes following long-term exposures of 10 years' duration or more. This gap in the literature is often attributed to limited exposure data (Crouse et al., 2015; Hansell et al., 2016; Hystad et al., 2013). Historical estimation of air pollutants, most frequently NO_2 (but also $PM_{2.5}$, PM_{10}, black smoke, SO_2, and O_3), is an attractive option for addressing this limitation in retrospective cohort and case-control studies. A number of methods have been reported in the literature. Traditional LUR models are a popular approach, and studies with access to historical predictor and measurement data have developed models for specific periods of interest as far back as 1962 (e.g., Gulliver et al., 2011). In addition, two types of back-extrapolation have been employed; either using historical predictors in contemporary models (e.g., Levy et al., 2015; back to 1961), or contemporary predictions adjusted using historical measurements when historical predictors are not available (e.g., Gulliver et al., 2016; back to 1991). Traditional LUR models have been used for historical estimation at city- through to national-scale (Beelen et al., 2007; Cesaroni et al., 2012; Levy et al., 2015).

Spatio-temporal models of varying complexity, including kriging, inverse distance-weighting and generalised additive models (GAMs), have also been used for regional- and national-scale estimates of historical pollutant concentrations back to 1980 (Dadvand et al., 2011; Hart et al., 2009; Keller et al., 2015; Kim et al., 2017; Yanosky et al., 2009). Dispersion models (back to 1960) and Chemical Transport Models (CTMs; back to 1988), parameterised with estimated historical emission data, have been used at city- through national-scale (Bellander et al., 2001; Hogrefe et al., 2009). Other methods include adjustment of historical PM_{10} and total suspended particulate (TSP) measurements using ratios to estimate $PM_{2.5}$ across the USA (Lall et al., 2004; back to 1972), as well as near-road NO_2 and benzene models based on street canyon geometry and traffic flow in Denmark (Raaschou-Nielsen et al., 2001; back to 1968).

Despite the diversity of methods used to estimate historical air pollution exposures, few have focused on satellite-based LUR, and few have been applied outside of North America and Europe (Hystad et al., 2012). This method has a number of appealing aspects, including spatial coverage and ability to be used at national scales. However, the limited availability of historical satellite observations, as was the case in this study, is likely to have constrained its more widespread application. Our approach to address this constraint used predictions from the first year for which our LUR model was developed (2006), and then three alternatives for adjusting the predictions for use during 1990–2005.

4.3. Comparison to other studies

The study most comparable to ours was performed by Hystad et al. (2012). Their Canadian national models used monitor-derived back-extrapolation of year-2001–2006 satellite-observed surfaces for $PM_{2.5}$, year-2005–2007 surfaces for NO_2, and a CTM surface for O_3, to estimate historical exposures during 1975–1994. Their models explained up to 38%, 51% and 56% of spatial variability in annual mean NO_2, $PM_{2.5}$ and O_3, respectively, over their study period when validated against a 10% sample of withheld data. While they used satellite surfaces alone in that study, rather than as an independent variable in an LUR model, in subsequent work they described the application of a national year-2006 satellite-based LUR for NO_2 (Hystad et al., 2011) to historical estimates during 1975–1994 (Hystad et al., 2015), 1984–2006 (Crouse et al., 2015) and 1999–2008 (Stieb et al., 2016). Direct comparison with our study is difficult because historical validation data for their satellite-LUR estimates were not presented. However, a consistent finding in their work is that applying satellite-based LUR historically yields similar results when used alongside other long-term NO_2 exposure assessment methods in case-control and cohort studies (Hystad et al., 2015; Stieb et al., 2016), and captures spatial contrasts in exposure when used alone in large cohort studies (Crouse et al., 2015). Taken together with our findings, this suggests that satellite-based LUR has a potential role to play in long-term exposure assessment, even in the absence of historical predictor data.

Gulliver et al. (2016) compared year-1991 non-satellite LUR models for NO_2 across Great Britain, developed using historical predictor data, with estimates from a year-2009 model that was back-extrapolated using historical measurements. When validated, the back-extrapolated model: (1) explained up to 8% less variability in measured NO_2 (R^2 = 62% *vs*. 56%), and; (2) had a small increase in RMSE, compared with the year-specific 1991 model. Using both models to estimate exposure at the 1.3 million postcodes in Great Britain showed high levels of correlation and absolute agreement. Unlike Gulliver et al. (2016), we did not have historical predictor data to develop year-specific LUR models, and instead relied on back-extrapolation of year-2006 estimates. However, the consistency they observed between the two approaches suggests the latter may be well-justified in the absence of data to develop the former.

4.4. Limitations

Our study has multiple limitations. We used a relatively small number of monitoring sites to assess historical performance; between 43% and 89% of such monitors were also used to develop our LUR models, depending on year. This reflected the scarcity of long-term NO_2 monitoring in Australia. It also likely means the model performance we observed is overly optimistic (Basagaña et al., 2012; Cesaroni et al., 2012; Eeftens et al., 2011; Johnson et al., 2010; Wang et al., 2012). We previously reported a 10 to 15 percentage point decrease in our models' R^2 from development to validation when evaluated against an independent set of 98 passive sampler measurements spanning non-roadside sites in two Australian cities during 2006–2014 (Knibbs et al., 2016). We have probably over-estimated model performance in the current study by a similar amount. A more conservative estimate of the models' performance may be that they capture 60–65% of spatial variation in annual NO_2 averaged over 1990–2005.

Because the large majority of sites used to develop and historically validate our models were located in urban background areas, we were unable to assess how well they capture NO_2 hot-spots. This aspect is a consequence of using regulatory ambient monitoring data, which is high-quality but mostly confined to areas free from strong nearby NO_2 sources. We previously used independent passive sampler data to evaluate model predictions at roadside locations in 2006–2014, and observed reasonable correlation but poor absolute agreement (R^2 = 36%, MSE-R^2 = −18%). However, we also showed that a typical residential address in Australia is unlikely to be located at roadside and immediately next to a major road (Knibbs et al., 2016). Urban background sites are more representative of a typical address in Australia, and this is further supported here by the consistency observed between historical validation sites and the G-NAF database of all address centroids in the country.

We did not have complete historical LUR predictor data, which motivated our study of back-extrapolation of NO_2 in the first instance because we could not develop historical year-specific LUR models (e.g., Gulliver et al., 2016). The three approaches we used to back-extrapolate year-2006 NO_2 (i.e., 'do nothing', changing the 'year' predictor, adjustment using monitors) all assume that the relationships between LUR predictors and NO_2 observed when developing our models for 2006–11 apply back to 1990. The extent to which this holds true is difficult to assess due to the lack of detailed historical data, but other proxies are available; for example, the national population increased by 1.2% per year on average during our study period (Australian Bureau of Statistics, 2010). Although national-scale land-use change was mostly minor, increases in intensive urban and agricultural land use from 23,000 to 31,000 km^2 were reported (Department of Agriculture, Fisheries and Forestry, 2011). The number of vehicle kilometres travelled (VKT) increased by 10–20% during the study period in some cities and remained stable in others, but vehicle combustion technology also improved contemporaneously (Department of the Environment and Energy, 2017b). While we cannot exclude the possibility of changes in the LUR predictor/NO_2 relationship in some parts of Australia, which would lead to reduced predictive performance and validity for exposure assessment, we suspect that the magnitude of change has been relatively small and spatially consistent. Finally, it is possible that the correlation between NO_2 and other pollutants may have changed over time and/or space, which could affect the validity of NO_2 as a proxy for exposure to other pollutants in epidemiological studies.

4.5. Conclusions

We found that simple back-extrapolation methods for year-2006 predictions from national satellite-based LUR models captured up to 75% of spatial variation in annual NO_2 averaged over 1990–2005 at 45 sites around Australia. Prediction error and bias were generally consistent with a previous validation for 2006–2014, and the absolute agreement of measurements and model predictions (measured by MSE-R^2) was similar to their correlation (measured by R^2). We observed that the validation sites appear suitable for assessing NO_2 exposures of the Australian population, based on comparison with ~8.4 million addresses in Australia. Historical estimation of air pollution exposure using LUR can facilitate studies of long-term health effects, but is often hampered by predictor data availability, especially in satellite-based LUR. Our approach for national-scale historical exposure assessment could potentially be applied in other locations with scarce monitoring and predictor data. We plan to use it in a cohort study of 265,000 older Australians.

Acknowledgements

L.D.K. acknowledges financial support from the Centre for Air Quality and Health Research and Evaluation, an NHMRC Centre of Research Excellence. We thank Mr Jason Rossiter for his assistance with the G-NAF predictions. We are grateful to the state and territory departments that performed the NO_2 measurements and provided them to us.

Conflicts of interest

None.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.envres.2018.01.046.

References

Allen, R.W., et al., 2011. The transferability of NO and NO_2 land use regression models between cities and pollutants. Atmos. Environ. 45, 369–378.

Australian Bureau of Statistics, 2010. Population growth: Past, Present and Future. ⟨http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4102.0Main+Features10Jun+2010⟩ (Accessed 31 October 2017).

Australian Bureau of Statistics, 2011. Australian Statistical Geography Standard (ASGS): Volume 1 – Main Structure and Greater Capital City Statistical Areas, Australia. ⟨http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/D3DC26F35A8AF579CA257801000DCD7D/$File/1270055001_july%202011.pdf⟩ (Accessed 31 October 2017).

Australian Bureau of Statistics, 2014. Australian Population Grid, 2011. ⟨http://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/1270.0.55.007Main+Features12011?OpenDocument⟩ (Accessed 30 October 2017).

Basagaña, X., et al., 2012. Effect of the number of measurement sites on land use regression models in estimating local air pollution. Atmos. Environ. 54, 634–642.

Bechle, M.J., et al., 2015. National Spatiotemporal exposure surface for NO_2: monthly scaling of a satellite-derived land-use regression, 2000–2010. Environ. Sci. Technol. 49, 12297–12305.

Beckerman, B.S., et al., 2013. A hybrid approach to Estimating National scale Spatiotemporal variability of PM2.5 in the contiguous United States. Environ. Sci. Technol. 47, 7233–7241.

Beelen, R., et al., 2007. Estimated long-term outdoor air pollution concentrations in a cohort study. Atmos. Environ. 41, 1343–1358.

Bellander, T., et al., 2001. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. Environ. Health Perspect. 109, 633–639.

Briggs, D.J., 2007. The use of GIS to evaluate traffic-related pollution. Occup. Environ. Med. 64, 1–2.

Cesaroni, G., et al., 2012. Nitrogen dioxide levels estimated from land use regression models several years apart and association with mortality in a large cohort study. Environ. Health 11, 48.

Chen, H., et al., 2010. Back-extrapolation of estimates of exposure from current land-use regression models. Atmos. Environ. 44, 4346–4354.

Cleary, G.J., 1969. A status report: air pollution control in Australia. J. Air Pollut. Control Assoc. 19, 490–496.

Crouse, D.L., et al., 2015. 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. 123, 1180–1186.

Dadvand, P., et al., 2011. Using spatio-temporal modeling to predict long-term exposure to black smoke at fine spatial and temporal scale. Atmos. Environ. 45, 659–664.

de Hoogh, K., et al., 2016. Development of West-European PM2.5 and NO2 land use regression models incorporating satellite-derived and chemical transport modelling data. Environ. Res. 151, 1–10.

Department of Agriculture, Fisheries and Forestry, 2011. Landscapes in Transition: Tracking Land use Change in Australia. ⟨http://www.agriculture.gov.au/abares/publications/display?Url=http://143.188.17.20/anrdl/DAFFService/display.php?Fid=pe_litlud9abll0790111a.xml⟩ (Accessed 31 October 2017).

Department of the Environment and Energy, 2017a. National Pollutant Inventory. ⟨http://www.npi.gov.au/⟩ (Accessed 30 October 2017).

Department of the Environment and Energy, 2017b. Australia State of the Environment 2016: Built Environment. ⟨https://soe.environment.gov.au/sites/g/files/net806/f/soe2016-built-launch-20feb.pdf?V=1488792899⟩ (Accessed 30 Octber 2017).

Eeftens, M., et al., 2011. Stability of measured and modelled spatial contrasts in NO(2) over time. Occup. Environ. Med. 68, 765–770.

Gilbert, N.L., et al., 2003. Ambient nitrogen dioxide and distance from a major highway. Sci. Total Environ. 312, 43–46.

Gulliver, J., et al., 2013. Development and back-extrapolation of NO_2 land use regression models for historic exposure assessment in Great Britain. Environ. Sci. Technol. 47, 7804–7811.

Gulliver, J., et al., 2016. Back-extrapolated and year-specific NO_2 land use regression models for Great Britain - Do they yield different exposure assessment? Environ. Int. 92–93, 202–209.

Gulliver, J., et al., 2011. Land use regression modeling to estimate historic (1962–1991) concentrations of black smoke and sulfur dioxide for great Britain. Environ. Sci. Technol. 45, 3526–3532.

Hansell, A., et al., 2016. Historic air pollution exposure and long-term mortality risks in England and Wales: prospective longitudinal cohort study. Thorax 71, 330–338.

Hart, J.E., et al., 2009. Spatial modeling of PM10 and NO_2 in the continental United States, 1985–2000. Environ. Health Perspect. 117, 1690–1696.

Hoek, G., et al., 2008. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmos. Environ. 42, 7561–7578.

Hoek, G., et al., 2015. Satellite NO_2 data improve national land use regression models for ambient NO_2 in a small densely populated country. Atmos. Environ. 105, 173–180.

Hogrefe, C., et al., 2009. A combined model–observation approach to estimate historic gridded fields of PM2.5 mass and species concentrations. Atmos. Environ. 43, 2561–2570.

Hystad, P., et al., 2012. Spatiotemporal air pollution exposure assessment for a Canadian population-based lung cancer case-control study. Environ. Health 11, 22.

Hystad, P., et al., 2013. Long-term residential exposure to air pollution and lung cancer risk. Epidemiology 24, 762–772.

Hystad, P., et al., 2011. Creating national air pollution models for population exposure assessment in Canada. Environ. Health Perspect. 119, 1123–1129.

Hystad, P., et al., 2015. Exposure to traffic-related air pollution and the risk of developing breast cancer among women in eight Canadian provinces: a case-control study. Environ. Int. 74, 240–248.

Jerrett, M., et al., 2017. Comparing the health effects of ambient particulate matter estimated using ground-based versus remote sensing exposure estimates. Environ. Health Perspect. 125, 552–559.

Johnson, M., et al., 2010. Evaluation of land-use regression models used to predict air quality concentrations in an urban area. Atmos. Environ. 44, 3660–3668.

Karner, A.A., et al., 2010. Near-roadway air quality: synthesizing the findings from real-world data. Environ. Sci. Technol. 44, 5334–5344.

Keller, J.P., et al., 2015. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. Environ. Health Perspect. 123, 301–309.

Kim, S.Y., et al., 2017. 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. 125, 38–46.

Knibbs, L.D., et al., 2016. Independent validation of national satellite-based land-use regression models for nitrogen dioxide using passive samplers. Environ. Sci. Technol. 50, 12331–12338.

Knibbs, L.D., et al., 2014. A national satellite-based land-use regression model for air pollution exposure assessment in Australia. Environ. Res. 135, 204–211.

Lall, R., et al., 2004. Estimation of historical annual PM2.5 exposures for health effects assessment. Atmos. Environ. 38, 5217–5226.

Larkin, A., et al., 2017. Global land use regression model for nitrogen dioxide air pollution. Environ. Sci. Technol. 51, 6957–6964.

Levy, I., et al., 2015. Back-extrapolating a land use regression model for estimating past exposures to traffic-related air pollution. Environ. Sci. Technol. 49, 3603–3610.

Marshall, J.D., et al., 2008. Within-urban variability in ambient air pollution: comparison of estimation methods. Atmos. Environ. 42, 1359–1369.

National Environment Protection Council, 1998. National Environment Protection (Ambient Air Quality) Measure. ⟨http://www.nepc.gov.au/nepms/ambient-air-quality⟩ (Accessed 31 October 2017).

National Environment Protection Council, 2000. Ambeint air Monitoring Programs around Australia. ⟨http://www.nepc.gov.au/system/files/resources/9947318f-af8c-0b24-d928-04e4d3a4b25c/files/aaq-ratf-rpt-appendix5-ambient-air-monitoring-programs.pdf⟩ (Accessed 31 October 2017).

Novotny, E.V., et al., 2011. National satellite-based land-use regression: NO_2 in the United States. Environ. Sci. Technol. 45, 4407–4414.

Pleijel, H., et al., 2004. On the logarithmic relationship between NO_2 concentration and

the distance from a highroad. Sci. Total Environ. 332, 261–264.
Poplawski, K., et al., 2009. Intercity transferability of land use regression models for estimating ambient concentrations of nitrogen dioxide. J. Expo. Sci. Environ. Epidemiol. 19, 107–117.
Public Sector Mapping Agencies, 2013. G-NAF Data Product Description Version 2.9.
Raaschou-Nielsen, O., et al., 2001. Air pollution from traffic at the residence of children with cancer. Am. J. Epidemiol. 153, 433–443.
Roorda-Knape, M.C., et al., 1999. Traffic related air pollution in city districts near motorways. Sci. Total Environ. 235, 339–341.
SAI Global, 2017. AS 3580.5.1-1993 Methods for Sampling and Analysis of Ambient Air Determination of Oxides of Nitrogen - Chemiluminescence Method. ⟨https://infostore.saiglobal.com/en-au/Standards/AS-3580-5-1-1993-297582/⟩ (Accessed 30 October 2017).
Stieb, D.M., et al., 2016. A national study of the association between traffic-related air pollution and adverse pregnancy outcomes in Canada, 1999–2008. Environ. Res. 148, 513–526.
Szpiro, A.A., et al., 2011. Does more accurate exposure prediction necessarily improve health effect estimates? Epidemiology 22, 680–685.
Vienneau, D., et al., 2013. Western European Land use regression incorporating satellite- and ground-based measurements of NO_2 and PM10. Environ. Sci. Technol. 47, 13555–13564.
Wang, M., et al., 2012. Systematic evaluation of land use regression models for NO_2. Environ. Sci. Technol. 46, 4481–4489.
Wang, M., et al., 2016. A new technique for evaluating land-use regression models and their impact on health effect estimates. Epidemiology 27, 51–56.
Wang, R., et al., 2013. Temporal stability of land use regression models for traffic-related air pollution. Atmos. Environ. 64, 312–319.
Yanosky, J.D., et al., 2009. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. Environ. Health Perspect. 117, 522–529.
Young, M.T., et al., 2016. Satellite-based NO_2 and model validation in a National prediction model based on Universal kriging and land-use regression. Environ. Sci. Technol. 50, 3686–3694.