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

A R T I C L E I N F O

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

A B S T R A C T

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
Available online 22 February 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).