Atmospheric Environment 122 (2015) 696–704

Contents lists available at ScienceDirect

Atmospheric Environment

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

Development and transferability of a nitrogen dioxide land use regression model within the Veneto region of Italy

Alessandro Marcon [a, b, *], Kees de Hoogh [b, c, d, 1], John Gulliver [b, 1], Rob Beelen [e, f], Anna L. Hansell [b, g]

[a] Unit of Epidemiology and Medical Statistics, Department of Diagnostics and Public Health, University of Verona, Verona, Italy
[b] MRC-PHE Centre for Environment and Health, Department of Epidemiology & Biostatistics, School of Public Health, Imperial College London, London, UK
[c] Swiss Tropical and Public Health Institute, Basel, Switzerland
[d] University of Basel, Basel, Switzerland
[e] Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands
[f] National Institute for Public Health and the Environment, Bilthoven, The Netherlands
[g] Imperial College Healthcare NHS Trust, London, UK

HIGHLIGHTS

- Land-use regression (LUR) is often used to estimate urban air pollution exposure.
- No studies have looked at transferability of LUR models from regions to cities.
- We developed a LUR model using NO_2 regulatory data for a region of Italy for 2010.
- When transferred to a inner city, the model was unable to capture NO_2 variability.
- LUR models should not be transferred to nested areas with different characteristics.

ARTICLE INFO

Article history:
Received 23 December 2014
Received in revised form
5 October 2015
Accepted 6 October 2015
Available online 19 October 2015

Keywords:
Air pollution
Ambient air
Environmental exposure
Exposure assessment
Geographic information systems
Land use regression

ABSTRACT

When measurements or other exposure models are unavailable, air pollution concentrations could be estimated by transferring land-use regression (LUR) models from other areas. No studies have looked at transferability of LUR models from regions to cities. We investigated model transferability issues. We developed a LUR model for 2010 using annual average nitrogen dioxide (NO_2) concentrations retrieved from 47 regulatory stations of the Veneto region, Northern Italy. We applied this model to 40 independent sites in Verona, a city inside the region, where NO_2 had been monitored in the European Study of Cohorts for Air Pollution Effects (ESCAPE) during 2010. We also used this model to estimate average NO_2 concentrations at the regulatory network in 2008, 2009 and 2011. Of 33 predictor variables offered, five were retained in the LUR model ($R^2 = 0.75$). The number of buildings in 5000 m buffers, industry surface area in 1000 m buffers and altitude, mainly representing large-scale air pollution dispersion patterns, explained most of the spatial variability in NO_2 concentrations ($R^2 = 0.68$), while two local traffic proxy indicators explained little of the variability ($R^2 = 0.07$). The performance of this model transferred to urban sites was poor overall ($R^2 = 0.18$), but it improved when only predicting inner-city background concentrations ($R^2 = 0.52$). Recalibration of LUR coefficients improved model performance when predicting NO_2 concentrations at the regulatory sites in 2008, 2009 and 2011 (R^2 between 0.67 and 0.80). Models developed for a region using NO_2 regulatory data are unable to capture small-scale

Abbreviations: aR^2, Adjusted R^2; ARPAV, Agenzia Regionale per la Prevenzione e Protenzione Ambientale del Veneto; ESCAPE, European Study of Cohorts for Air Pollution Effects; GEIRD, Gene Environment Interactions in Respiratory Diseases (study); GIS, Geographic information system; LOOCV, Leave-one-out cross-validation; LUR, Land-use regression; NO_2, Nitrogen dioxide; RMSE, Root mean square error.

* Corresponding author. Unit of Epidemiology and Medical Statistics, Department of Diagnostics and Public Health, University of Verona, Strada Le Grazie 8, 37134 Verona, Italy.

E-mail addresses: alessandro.marcon@univr.it (A. Marcon), c.dehoogh@unibas.ch (K. de Hoogh), j.gulliver@imperial.ac.uk (J. Gulliver), R.M.J.Beelen@uu.nl (R. Beelen), a.hansell@imperial.ac.uk (A.L. Hansell).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.atmosenv.2015.10.010

variability in NO_2 concentrations in urban traffic areas. Our study documents limitations in transferring a regional model to a city, even if it is nested within that region.

1. Introduction

Regression mapping, also called Land Use Regression (LUR) modelling, is one commonly used exposure assessment technique (Jerrett et al., 2005). LUR has been recently employed in the European Study of Cohorts for Air Pollution Effects (ESCAPE), where models for several air pollutants, including nitrogen dioxide (NO_2), have been developed based on dedicated monitoring campaigns (Beelen et al., 2013) and applied (Schikowski et al., 2014) to estimate residential outdoor air pollutant concentrations. With respect to interpolation methods like ordinary kriging, LUR is generally better in capturing small-scale variability in air pollution, especially in urban areas, where air pollution concentrations vary widely across short distances, and when the monitoring network is relatively sparse (Jerrett et al., 2005; Gulliver et al., 2011; Akita et al., 2014; Hoek et al., 2008). With respect to air dispersion models, LUR models have less demanding data, software and hardware requirements (Beelen et al., 2010; Wang et al., 2015).

Ideally, developing a LUR model would need a dense network of monitors, possibly providing an independent test set for model validation. Monitor locations should be selected that appropriately reflect the variation in air pollution concentrations at the residential addresses of the study population (using e.g. saturation sampling, Shmool et al., 2014). In practice, researchers are often relying on available data from regulatory monitoring networks. Although these networks may provide a wide temporal coverage, they usually have a sparse spatial coverage that may be poorly representative of residential areas. The alternative is undertaking expensive *ad hoc* monitoring, which might have limited temporal coverage (Jerrett et al., 2005).

The Genes Environment Interaction in Respiratory Diseases (GEIRD) population study is carried out in 8 centres in Italy, aimed at investigating environmental and genetic determinants of respiratory chronic inflammatory diseases (de Marco et al., 2010). Clinical examinations started in 2008 and are ongoing (Marcon et al., 2013). For three centres, including Verona (Northern Italy), LUR models developed in ESCAPE could be used to estimate air pollution exposure of study participants (Beelen et al., 2013). For centres where neither exposure models, nor dedicated monitoring campaigns, nor regulatory networks of adequate density are available, one option could be to develop LUR models using available regulatory data for a larger area or region. However, transferring models from regions to urban (although inner) areas requires a preliminary evaluation.

A number of studies have transferred LUR models between cities (Briggs et al., 2000; Jerrett et al., 2005; Poplawski et al., 2009; Allen et al., 2011) or countries (Vienneau et al., 2010; Wang et al., 2014). In some cases, both predictor variables and coefficients of LUR models were transferred (Allen et al., 2011; Jerrett et al., 2005), while in others regression coefficients were recalibrated using air pollution measurements from the target area (Briggs et al., 2000; Poplawski et al., 2009; Vienneau et al., 2010). The performance of transferred models was highly variable, ranging from very poor (Jerrett et al., 2005) to satisfactory (Briggs et al., 2000), but it was generally lower than that of models developed for a specific area (Poplawski et al., 2009). To our knowledge, few studies have attempted to transfer LUR models from a region to a smaller, nested area (Wang et al., 2014), and none of these have examined transferability from regions to cities.

In this study, we developed a LUR model for NO_2 using regulatory monitoring data for 2010 from the Veneto region, a large administrative area including the municipality of Verona (Fig. 1), and we studied its performance on the set of measurements carried out in Verona during the same year in ESCAPE. We also studied whether this model could be used to estimate NO_2 concentrations in other years of the 2008–2011 period.

2. Methods

2.1. Study area

The Veneto region has a surface of 18407 km^2 and about 4.8 million inhabitants (data on the 2011 census of the Italian population, available at http://dati-censimentopopolazione.istat.it/, last accessed on 11 June 2015). It is divided into 7 provinces. The Verona municipality, inside the Verona province, is 198.9 km^2 large and has 252520 inhabitants (Fig. 1).

2.2. Air pollution data

Daily and annual average NO_2 concentrations for years 2008–2011 measured by 53 regulatory stations in Veneto were retrieved from the European air quality database website (http://www.eea.europa.eu/). Of these, 41 were active throughout the whole period. The stations are operated by the regional environmental agency (Agenzia Regionale per la Prevenzione e Protenzione Ambientale del Veneto, [ARPAV]) and detect NO_2 by chemiluminescence. They are classified into regional background, urban (including suburban) background, industrial, and street stations. Stations with $\leq$75% daily average concentrations were not considered. Since most of the geographic information system (GIS) data were available only within the region and the largest buffer for calculation of predictors was 5 km, stations that were at <5 km from the regional border were not considered to avoid missing data in these buffers. Coordinates were retrieved from ARPAV and checked on satellite maps.

NO_2 concentrations measured in ESCAPE are described elsewhere (Cyrys et al., 2012). Briefly, measurements had been conducted by passive sampling (Ogawa badges) in Verona, at 40 sites classified as regional background, urban background and street sites, between January and June 2010, during three 14-day periods. Annual average concentrations had been calculated for all sites correcting for temporal variation, using measurements obtained from a background regulatory station that was operated in Verona year-round (Cyrys et al., 2012). This improves comparability between annual average concentrations measured by passive samplers and regulatory stations.

For each regulatory station, annual average concentrations were also estimated using daily regulatory data for the same 3 ESCAPE sampling periods, applying the ESCAPE temporal adjustment procedure (Cyrys et al., 2012).

2.3. Potential predictor variables

A GIS (Arc Map 10.1 software, ESRI, Redlands, California), set to