Environmental Modelling & Software 100 (2018) 222–235

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Estimating the spatiotemporal variation of NO_2 concentration using an adaptive neuro-fuzzy inference system

Bijan Yeganeh [a, b], Michael G. Hewson [c], Samuel Clifford [b, d], Ahmad Tavassoli [e], Luke D. Knibbs [f], Lidia Morawska [a, *]

[a] International Laboratory for Air Quality and Health, Queensland University of Technology, Brisbane, Queensland 4001, Australia
[b] Centre for Air Quality and Health Research and Evaluation, Glebe, New South Wales 2037, Australia
[c] School of Education and the Arts, Central Queensland University, North Rockhampton, Queensland 4702, Australia
[d] ARC Centre of Excellence for Mathematical and Statistical Frontiers, Queensland University of Technology, Brisbane, Queensland 4001, Australia
[e] School of Civil Engineering, The University of Queensland, St Lucia, Queensland 4072, Australia
[f] School of Public Health, The University of Queensland, Herston, Queensland 4006, Australia

ARTICLE INFO

Article history:
Received 1 December 2016
Received in revised form
29 September 2017
Accepted 20 November 2017

Keywords:
NO_2
Satellite data
ANFIS
Spatiotemporal
Transport model
Australia

ABSTRACT

Statistical modelling has been successfully used to estimate the variations of NO_2 concentration, but employing new modelling techniques can make these estimations far more accurate. To do so, for the first time in application to spatiotemporal air pollution modelling, we employed a soft computing algorithm called adaptive neuro-fuzzy inference system (ANFIS) to estimate the NO_2 variations. Comprehensive data sets were investigated to determine the most effective predictors for the modelling process, including land use, meteorological, satellite, and traffic variables. We have demonstrated that using selected satellite, traffic, meteorological, and land use predictors in modelling increased the R^2 by 21%, and decreased the root mean square error (RMSE) by 47% compared with the model only trained by land use and meteorological predictors. The ANFIS model found to have better performance and higher accuracy than the multiple regression model. Our best model, captures 91% of the spatiotemporal variability of monthly mean NO_2 concentrations at 1 km spatial resolution (RMSE 1.49 ppb) in a selected area of Australia.

Data availability

The type and source of the data set considered in this study.

Name of the data set	Data source (Developer) (All websites accessed on April 2016)	Data format	Software required	Data availability
OMI tropospheric NO_2 column density (molecules $\times$ 10^{15}/cm^2)	Aura OMI tropospheric NO_2 column density product via NASA Giovanni interface http://giovanni.sci.gsfc.nasa.gov/giovanni/?instance_id=omil2g	HDF/NetCDF files	ArcGIS	Freely available
Major road	PSMA Australia Transport and Topography product https://www.psma.com.au/products/transport-topography	ESRI shape files	" "	Price depends on the area of interest
Minor road	" "	" "	" "	" "

* Corresponding author.
E-mail address: l.morawska@qut.edu.au (L. Morawska).

https://doi.org/10.1016/j.envsoft.2017.11.031
1364-8152/

(*continued*)

Name of the data set	Data source (Developer) (All websites accessed on April 2016)	Data format	Software required	Data availability
Industrial point source NOx emissions	Australia National Pollutant Inventory http://www.npi.gov.au/reporting/industry-reporting-materials	xml files	Microsoft Excel/R	Freely available
Australia population density	Australian Bureau of Statistics http://www.abs.gov.au/ausstats/abs@.nsf/mf/1270.0.55.007	PNG ESRI Grid GeoTIFF	ArcGIS	" "
Australia land use classification	Australian Bureau of Statistics http://www.abs.gov.au/websitedbs/censushome.nsf/home/meshblockcounts	Excel spreadsheets/CSV files	Microsoft Excel/R/ArcGIS	" "
Elevation	U.S. Geological Survey https://www.usgs.gov/products/maps/topo-maps	PNG GeoTIFF	ArcGIS	" "
Normalized difference vegetation index	Terrestrial Ecosystem Research Network http://www.auscover.org.au/node/9	NetCDF files	" "	" "
Temperature Rainfall Humidity Solar exposure	Australian Bureau of Meteorology http://www.bom.gov.au/climate/maps/#tabs=Maps	ESRI Grid GIF	" "	" "
Traffic data	Department of Transport and Main Roads http://www.tmr.qld.gov.au/Travel-and-transport/Road-and-traffic-info/Traffic-reports-and-road-conditions	ESRI shape files	" "	Price depends on the area of interest

Software availability

The following software has been used in this study for statistical analysis, spatial data processing, map creation, and calculating the meteorological and traffic-related parameters:

- R v. 3.2.3 (R Foundation for Statistical Computing, Vienna, Austria)
- MATLAB R2014b (MathWorks Inc., Natick, USA)
- ArcGIS v.10.2 (ESRI Inc., Redlands, USA)
- Weather Research and Forecasting v 3.8.1 (Powers et al., 2008)
- South-east Queensland Strategic Transport Model (Ryan et al., 2008)

Note: No specific software component has been developed for this study.

1. Introduction

Exposure to ambient air pollution is a major environmental risk factor associated with adverse health effects (Forouzanfar et al., 2015). Nitrogen dioxide (NO_2) is a major component of ambient air pollution and a strong marker of traffic-related emissions (Briggs et al., 1997; Richter et al., 2005). To date, epidemiological studies have demonstrated that there are adverse health effects associated with exposure to NO_2 (Crouse et al., 2010, 2015; Filleul et al., 2005; Mölter et al., 2014; Parent et al., 2013; Perez et al., 2012). In addition, NO_2 is recognized as a good proxy of particle number concentration in urban environments (Grundström et al., 2015). Hence more precise estimates of NO_2 concentration is needed to investigate its associated role on health effects.

Fossil fuel combustion including coal, gas and oil, are the major sources of NO_2 in Australia (Australian Government, 2010). As a subset of this, traffic related emissions are a major source of NO_2 in urban areas (Derwent and Hertel, 1998). About 80% of the NO_2 in Australian urban areas comes from motor vehicle exhaust (Australian Government, 2010). This indicates that traffic flow needs to be carefully investigated for estimating the NO_2 concentration in Australia as one of the most urbanized countries in the world.

Different approaches have been used to provide a proxy for traffic flow including calculating the length of the roads or road classification (Henderson et al., 2007; Knibbs et al., 2014; Sahsuvaroglu et al., 2006), and using nearest traffic count (Dirgawati et al., 2015; Ducret-Stich et al., 2013). Some studies used transportation models to obtain more accurate estimates of traffic flow, and this approach has been found to provide better results than previous approaches (Costabile and Allegrini, 2008; Kim and Guldmann, 2011, 2015; Shekarrizfard et al., 2015).

Moreover, traffic dynamics can also significantly affect the NO_2 emissions, as congested traffic (e.g. stop-and-go traffic) results in particulate matters and gaseous emissions peak beyond the free-flow traffic condition (Davis and Peckham, 2007; Giakoumis et al., 2012; Hagena et al., 2006; Keuken et al., 2010; Rakopoulos and Giakoumis, 2009). Consequently, traffic dynamics and condition plays a significant role in the emission of NO_2 from vehicles in urban areas and should be investigated during the NO_2 modelling process.

Estimates of air pollution concentration have been traditionally provided by ground monitoring networks. The sparse ground measurement network in many parts of the world, including Australia makes it difficult to evaluate the spatiotemporal variability of ambient air pollution. Even a dense network could not adequately monitor the spatiotemporal variability of ambient air pollution, since it is changing on scales much smaller than monitoring networks density. This represents a significant limitation on evaluating the adverse health effects associated with ambient air pollution.

Satellite imagery is another important tool rapidly gaining interest in air pollution monitoring as it provides sequential observations with extensive spatial coverage (Kloog et al., 2014). Data derived from satellite sensors can be combined with ground-based measurements at different spatiotemporal scales (Reis et al., 2015). The availability of satellite-derived data has helped to overcome the problems associated with sparse monitoring networks by providing observations where previously there were none (Martin, 2008; Reis et al., 2015).

Observation-based statistical methods have emerged as a powerful tool for exploring the quantitative relationship between ground-level NO_2 concentrations and satellite-derived data, and a variety of these methods have been used to quantify this relationship (Bechle et al., 2015; Carnevale et al., 2016; Hoek et al., 2015; Horsburgh et al., 2016; Hystad et al., 2011; Knibbs et al., 2014; Novotny et al., 2011; Reggente et al., 2014; Vienneau et al., 2013). Machine learning refers to computational techniques which are able to achieve optimal solutions for analyzing