Environmental Modelling & Software 88 (2017) 84–92

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A satellite-based model for estimating $PM_{2.5}$ concentration in a sparsely populated environment using soft computing techniques

Bijan Yeganeh [a, b], Michael G. Hewson [c], Samuel Clifford [b, d], Luke D. Knibbs [e], 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] Mathematical Sciences School, Queensland University of Technology, Brisbane, Queensland, 4000, Australia
[e] School of Public Health, The University of Queensland, Herston, Queensland 4006, Australia

ARTICLE INFO

Article history:
Received 29 April 2016
Received in revised form
13 November 2016
Accepted 14 November 2016

Keywords:
$PM_{2.5}$
Aerosol optical depth
ANFIS
SVM
BPANN
Australia

ABSTRACT

We applied three soft computing methods including adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM) and back-propagation artificial neural network (BPANN) algorithms for estimating the ground-level $PM_{2.5}$ concentration. These models were trained by comprehensive satellite-based, meteorological, and geographical data. A 10-fold cross-validation (CV) technique was used to identify the optimal predictive model. Results showed that ANFIS was the best-performing model for predicting the variations in $PM_{2.5}$ concentration. Our findings demonstrated that the CV-R^2 of the ANFIS (0.81) is greater than that of the SVM (0.67) and BPANN (0.54) model. The results suggested that soft computing methods like ANFIS, in combination with spatiotemporal data from satellites, meteorological data and geographical information improve the estimate of $PM_{2.5}$ concentration in sparsely populated areas.

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 Jan 2016)	Data format	Software required	Data availability
OMI Near-UV AOD	Aura OMI AOD 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	" "	" "	" "	" "
Industrial point source $PM_{2.5}$ 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	" "

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

http://dx.doi.org/10.1016/j.envsoft.2016.11.017
1364-8152/

(*continued*)

Name of the data set	Data source (Developer) (All websites accessed on Jan 2016)	Data format	Software required	Data availability
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	" "	" "

Software availability

The following software has been used in this study for statistical analysis, spatial data processing and map creation:

- R v.3.2.3 (R Foundation for Statistical Computing, Vienna, Austria)
- MATLAB R2014b (MathWorks Inc., Natick, USA)
- ArcGIS version 10.2 (ESRI Inc., Redlands, USA)

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

1. Introduction

Exposure to fine particulate matter ($PM_{2.5}$, particles with aerodynamic diameter less than 2.5 μm) is a leading environmental risk factor associated with respiratory and cardiovascular morbidity and mortality (Franklin et al., 2007) and it is the twelfth-ranked contributor to the global burden of diseases (Forouzanfar et al., 2015).

Urbanisation increases the risk of being exposed to $PM_{2.5}$ (Han et al., 2015), and Australia, as one of the most urbanised countries in the world, is faced with adverse health effects of $PM_{2.5}$. To date, very little attention has been paid to the health effect of exposure to $PM_{2.5}$ in Australia. Some studies consistently suggest that $PM_{2.5}$ is associated with respiratory diseases and has significant effects on mortality (Barnett et al., 2005; Simpson et al., 2005), while conflicting results have been reported on cardiovascular health effects (Hinwood et al., 2006). These inconsistent results could be due to difficulties in assessing the Australian population exposure to $PM_{2.5}$.

Ground level aerosol measurement has been historically provided by ground monitoring networks, but there are high establishing and maintaining expenses associated with these measurements (Wu et al., 2012). The sparse ground $PM_{2.5}$ measurement network in Australia makes it difficult to evaluate the spatiotemporal variability of $PM_{2.5}$ and has significantly restrained the epidemiological studies on $PM_{2.5}$ health effects. Australia is the sixth largest country in the world by area while its population is quite small compared to the land size (Australian Government, 2015). Australia is one of the 10 least dense populated countries in the world (United Nations, 2015). The majority of the Australian population is living in the east and west coasts (Lunn et al., 2002). The population within these areas is concentrated in urban centres, particularly the capital cities (Australian Bureau of Statistics, 2012a,b; Lunn et al., 2002). Therefore, limited monitoring stations were established only in populated areas due to population distribution in Australia. Had such monitoring networks existed, there would have been no guarantee of an effective measurement of the spatiotemporal variation of $PM_{2.5}$, since it is changing on scales much smaller than monitoring networks density.

Estimates of air pollution exposure have been traditionally provided by assigning measurements derived from one (Chen et al., 2006) or several air pollution monitors (Barnett et al., 2005; Brook et al., 2010; Chan et al., 2006), allocating exposure using the nearest monitoring station (Lee et al., 2014) or using different proxies to estimate a local population's exposure (Hoffmann et al., 2007; Salam et al., 2008; Samet, 2007). There is potential for over-smoothing the exposure estimation and the results are likely to be biased with all these approaches (Jerrett et al., 2005a).

Satellite imagery is another important tool rapidly gaining interest in air pollution monitoring as it provides sequential observations over a broad area. Satellite sensors can be coupled with ground-based sensors at different spatiotemporal scales to reduce the limitations of surface monitoring station (Reis et al., 2015). Aerosol Optical Depth (AOD) is the most common parameter derived from satellite observations and applied to estimate $PM_{2.5}$. AOD describes the level of which aerosols attenuate the electromagnetic radiation at a given wavelength by absorption or scattering in an atmospheric column (Chudnovsky et al., 2012; Kaufman et al., 2002; NASA, 2013). The availability of satellite-derived AOD has helped to overcome the problems associated with sparse monitoring networks by providing observations where previously there were none (Hoff and Christopher, 2009; Reis et al., 2015).

A variety of methods have been used to investigate the quantitative relationship between satellite-derived AOD and ground-level $PM_{2.5}$ measurements. These studies mainly fall into two major classes: numerical-based methods and empirical observation-based methods (Lin et al., 2014).

Numerical-based models, including dispersion and chemical transport models, are still under development due to the uncertainties regarding the definition of source inventories, and chemical and dynamical processes of aerosols in atmosphere (Gupta and Christopher, 2009b; Kondragunta et al., 2008). Empirical observation-based methods rely on the relationship between air quality measurements and different observations (Maciejewska et al., 2015). Several techniques have been used to describe this relationship including simple regression (Chu et al., 2003), multiple regression (Dirgawati et al., 2015; Gupta and Christopher, 2009b; Li et al., 2011), geostatistical methods (Jerrett et al., 2005b; Kunzli et al., 2005), generalized additive models (GAM) (Strawa et al., 2013), land use regression (Henderson et al., 2007; Kloog et al., 2011; Knibbs et al., 2014; Liu et al., 2009), and hybrid approaches (Beckerman et al., 2013b; Lindstrom et al., 2011). Soft computing refers to computational techniques which are able to achieve optimal solutions for analysing complicated phenomena at reasonable costs (Carnevale et al., 2016; Kruse et al., 2013; Ovaska, 2004). In recent years, soft computing techniques such as support