Atmospheric Environment 187 (2018) 401–409

Contents lists available at ScienceDirect

Atmospheric Environment

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

A hybrid Grey-Markov/ LUR model for PM_{10} concentration prediction under future urban scenarios

Shan Xu[a], Bin Zou[a,*], Sedra Shafi[a], Troy Sternberg[b]

[a] School of Geosciences and Info-Physics, Central South University, Changsha, China
[b] School of Geography and Environment, University of Oxford, Oxford, OX1 2JD, UK

ARTICLE INFO

Keywords:
Urban sprawl
Land use simulation
Air pollution prediction
Grey-Markov
LUR

ABSTRACT

Exploring the spatial distribution of air pollutants under future urban planning scenarios is essential as urban sprawl increases in China. However, existing published prediction models usually forecast pollutant concentrations at the station level or estimate spatial distribution of pollutant in a historical perspective. This study has developed a hybrid Grey-Markov/land use regression (LUR) model (GMLUR) for PM_{10} concentration prediction under future urban scenarios by employing the forecast of Grey-Markov model as surrogate measurements to calibrate the spatial estimations of LUR model. Taking the agglomeration of Changsha-Zhuzhou-Xiangtan (CZT) in China as a case, the superiority of GMLUR was tested and spatial distribution of PM_{10} concentrations based on four potential land use scenarios for the year 2020 were predicted. Results show that GMLUR modelling outperforms LUR modelling with clear lower average relative percentage error (5.13% vs. 24.09%) and root-mean-square error (5.50 μg/m^3 vs. 21.31 μg/m^3). The economic interest scenario identifies the largest demands of future built-up (2 306.50 km^2) and bare (34.88 km^2) areas. Built-up area demands for the business as usual scenario, resource-conserving scenario, and ecological interest scenario are 362.67, 1 042.22, and 1 014.70 km^2, respectively. Correspondingly, the economic interest scenario identifies the severest PM_{10} pollution with the highest mean predicted concentration of 53.78 μg/m^3 and the largest percent (19.43%) of area exceeding the Level 2 value (70 μg/m^3) of Chinese National Ambient Air Quality Standard (CNAAQS); these are significantly higher than those of the business as usual scenario (49.63 μg/m^3, 6.28%). The resource-conserving scenario (46.79 μg/m^3) and ecological interest scenario (46.76 μg/m^3) are cleaner with no area exceeding the Level 2 value of CNAAQS. It can be concluded that GMLUR modelling provides a feasible way to evaluate the potential outcome of future urban planning strategies in the perspective of air pollution.

1. Introduction

Air pollution, especially particulate matter, has been associated with numerous adverse effects on human health in urban areas, including increased mortality and morbidity from respiratory, lung, and cardiopulmonary cancer (Zou et al., 2015; Loomis et al., 2013; Beelen et al., 2014; Wang et al., 2014; Kim et al., 2015). Though concentrations of urban PM_{10} generally have been declining since the turn of the 21st century (Cheng et al., 2013; Querol et al., 2014), the control of particulate matter remains an enormous challenge with increasing industrial production, travel behaviour and construction activities that follow decades of rapid urbanization (Feng et al., 2017). Understanding the spatial distribution of PM_{10} concentrations under varying future urban planning scenarios is a crucial challenge in designing urban development strategies and the prevention of air pollution exposure.

Various approaches to predict the concentrations of urban air pollutants have been tested. Efforts include temporal forecasting and spatial mapping. The temporal forecast makes predictions on historical relationships and trends from data on air pollutant observation. The commonly used methods are well-tested and have demonstrated promising forecasting accuracy. This is typified by neural networks, support vector machine learning, support vector regression, parametric and nonparametric regression, autoregressive moving (integrated) average modelling, grey system theory, etc. (Hooyberghs et al., 2005; Hrust et al., 2009; Kumar and Jain, 2010a; Lotfalipour et al., 2013; Qin et al., 2014; Wang et al., 2015; Donnelly et al., 2015; Hamzacebi and Karakurt, 2015). Among these, the grey method has been proved to be an effective way to make predictions of air pollutants at a relatively large scale (e.g. annual scale) under conditions with limited data. Yet forecasting precision of grey method might be affected by the random fluctuations of the data sequence. The Grey-Markov model, which introduced the Markov chain models to reduce the random fluctuation,

* Corresponding author. Department of Geosciences and Remote Sensing, Central South University, Changsha, 410083, China.
E-mail addresses: shan_xu@csu.edu.cn (S. Xu), 210010@csu.edu.cn (B. Zou), sedrashafich@gmail.com (S. Shafi), troy.sternberg@geog.ox.ac.uk (T. Sternberg).

https://doi.org/10.1016/j.atmosenv.2018.06.014
Received 27 November 2017; Received in revised form 8 June 2018; Accepted 10 June 2018
Available online 15 June 2018
1352-2310/

has been successfully applied in forecasting the electric-power demand, energy consumption and foreign tourist arrivals (Huang et al., 2007; Kumar and Jain, 2010b; Sun et al., 2016). It could be an effective way to improve the accuracy of the grey method for fluctuated air pollution data sequences. However, an important mutual limitation is that all those methods are usually conducted at the station level, which is not able to characterize the patterns of air pollution under future urban scenario spatially and to evaluate the validity of varied urban planning strategies.

Spatial mapping, on the other hand, attempts to retrieve the spatial distribution characteristics of air pollution concentrations based on physiochemical processes or the relationship between air pollutants and their potential predictors. Satellite remote sensing (RS), air dispersion modelling, spatial interpolation, and land use regression (LUR) models are practical, frequently used solutions (Fang et al., 2016; Zou et al., 2016a, 2016b, 2017; Apte et al., 2017; Zhai et al., 2018). LUR, which predicts concentration of air pollutant at a given site based on surrounding land use, meteorology factors and other variables obtained through geographic information system (GIS), is now a popular method for air pollution estimation at fine spatial resolutions because of its low requirement of intensive computations and easy availability of related data (Henderson et al., 2007; Ross et al., 2007; Hoek et al., 2010; Zou et al., 2015; Jedynska et al., 2017). Due to the follow-up long-term epidemiological studies usually have longer periods than monitoring data used in LUR modelling; few investigations have tried to transfer the LUR model across time (Mölter et al., 2010; Marcon et al., 2015; Meng et al., 2015; Zou et al., 2016c). Although the results indicate that the LUR models were reasonably stable over time and it was possible to transfer them to different years, these reported studies were performed retrospectively. The fundamental reason why LUR is seldom used in future urban scenarios may be the lack of essential inputs of both air pollution observations and land use distributions.

Fortunately, the simulation of urban land-use change dynamics is well developed and comprehensive. The methods can be put into the following groups: cellular automata model, agent-based model, empirical statistical model, and hybrid models (Verburg et al., 2004; Chen et al., 2008; Santé et al., 2010; Zhang et al., 2013; Basse et al., 2014; Groeneveld et al., 2017; Liu et al., 2017). Among them, the conversion of land use and its effects at the small regional extent (CLUE-S) model can derive empirically quantitative relations between land use change and driving factors from cross-sectional analysis at multiple scales. This simulates possible changes under land use scenarios spatially explicit in small regions at a fine spatial resolution. It has been introduced with notable accuracy globally, which provides a reliable foundation for analysis of future urban planning strategies (Verburg et al., 2002).

In this study we developed a hybrid Grey-Markov/LUR (GMLUR) model to explore the spatial patterns of PM_{10} concentrations under future urban scenarios. Research integrates the prediction of station-based Grey-Markov modelling with the spatial mapping of LUR. The overall objective is to extend two-dimensional spatial mapping of LUR into the three-dimensional spatial prediction of GMLUR to achieve the area-based forecast of PM_{10} concentrations and to illustrate the potential effect of urban planning scenarios on the temporal evaluation of the spatial distribution of PM_{10} concentrations.

2. Framework of GMLUR and supporting methods

The hybrid Grey-Markov/LUR modelling (GMLUR) is a calibrated LUR method which forecasts PM_{10} concentrations based on future land use and PM_{10} concentrations surrogates. The process can be divided into four steps. Step 1 develops and validates the base LUR model (LUR_H (t)) based on the historical observations of PM_{10} and various geographic elements. In Step 2 we obtain the predictions of PM_{10} concentration for the validation year (LUR_H predictions (t + n)) by apply the base LUR model to geographic elements in validation year (i.e. transfer LUR_H (t) temporally). The predictions of PM_{10}

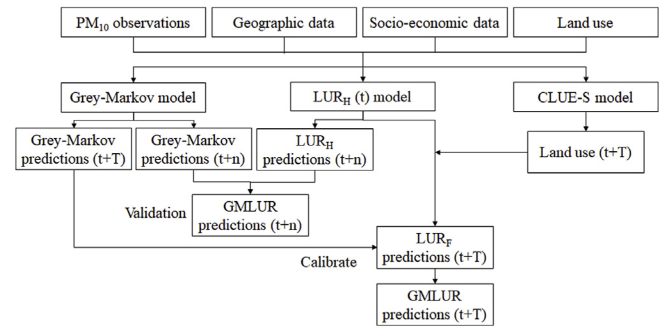

Fig. 1. The work flow of GMLUR.

concentrations based on Grey-Markov model (Grey-Markov predictions (t + n)) are then employed to calibrate the LUR_H predictions (t + n) through a ratio method. The results (i.e. GMLUR predictions (t + n)) are compared with the measurements to evaluate the reliability of GMLUR. Step 3 simulates the spatial patterns of land use for representative scenarios of the target year. In Step 4, to extrapolate PM_{10} concentrations in cells for the target year, we apply the GMLUR to future land use in Step 3 by calibrating predictions of LUR in the target year (LUR_F predictions (t + T)) with forecast level of PM_{10} from Grey-Markov model (Grey-Markov predictions (t + T)). Details of the entire procedure are illustrated in Fig. 1. The essential supporting methods include the following three parts.

2.1. Grey-Markov prediction

The hybrid Grey-Markov/LUR modelling (GMLUR) is a calibrated LUR method which The Grey-Markov model used here follows Huang et al. (2007) and can be described as follows:

Step I. The Grey forecasting model GM (1, 1) can be expressed as:

$$\hat{x}^{(0)}(k+1) = \left[x^{(0)}(1) - \frac{b}{a}\right]e^{-ak}(1-e^{a}),\ \ k=1,\ 2,\ \cdots,n \tag{1}$$

Where $\hat{x}^{(0)}(k+1)$ is the predictions of the original raw data series $x^{(0)}$, $x^{(0)} = (x^{(0)}(1),\ x^{(0)}(2),\ \cdots,\ x^{(0)}(n))$, a and b in Equation (1) are estimated using the least square method.

Step II. The values of $x^{(0)}$ are distributed in the region of the trend curve $\hat{y}(k) = \hat{x}^{(0)}(k+1)$ which may be divided into a convenient number of contiguous intervals. When $x^{(0)}$ falls in interval i, one of S such intervals, it may be regarded as corresponding to a state $Q_i = [Q_{1i},\ Q_{2i}]$ in an m order Markov unstable sequence, where $Q_{1i} = \hat{y}(k) + A_i$, $Q_{2i} = \hat{y}(k) + B_i$.

Step III. For Markov-chain series, the transition probability $P_{ij}(m)$ from state Q_i to Q_j can be established using the equation: $P_{ij}(m) = \frac{M_{ij}(m)}{M_i}$, (i, j= 1, 2, ⋯,S), where $M_{ij}(m)$ is the number of original data of state Q_j transferred from state Q_i for m-steps, M_i is the number of original data points in Q_i. These $P_{ij}(m)$ values can be arranged as a transition probability matrix:

$$R_{ij}(m) = \begin{bmatrix} P_{11}(m)\,P_{12}(m)\cdots P_{1j}(m) \\ P_{21}(m)\,P_{22}(m)\cdots P_{2j}(m) \\ \cdots\cdots\cdots \\ P_{i1}(m)\,P_{i2}(m)\cdots P_{ij}(m) \end{bmatrix}\cdot (i,\ j=1,\ 2,\ \cdots,S) \tag{2}$$

Step IV. After the determination of the future state transition of a system, i.e., the determination of Grey-elements Q_{1j}, Q_{2j} the changing interval of the forecast value is between Q_{1j}, Q_{2j}. The most probable forecast value $\hat{y}(k+1)$, is considered as the middle value of the determined state interval, that is

$$\hat{y}(k+1) = 1/2(Q_{1i} + Q_{2i}) = \hat{y}(k) + 1/2(A_i + B_i) \tag{3}$$