Environmental Pollution 231 (2017) 1232–1244

Contents lists available at ScienceDirect

Environmental Pollution

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

Daily air quality index forecasting with hybrid models: A case in China☆

Suling Zhu [a], Xiuyuan Lian [b, *], Haixia Liu [b], Jianming Hu [b], Yuanyuan Wang [b], Jinxing Che [c]

[a] School of Public Health, Lanzhou University, Lanzhou 730000, Gansu, China
[b] School of Mathematics & Statistics, Lanzhou University, Tianshuinanlu 222, Lanzhou, China
[c] School of Science, Nanchang Institute of Technology, Nanchang 330099, JiangXi, China

A R T I C L E I N F O

Article history:
Received 18 April 2017
Received in revised form
15 August 2017
Accepted 18 August 2017
Available online 19 September 2017

Keywords:
Hybrid model
Air pollution indexes
Forecasting

A B S T R A C T

Air quality is closely related to quality of life. Air pollution forecasting plays a vital role in air pollution warnings and controlling. However, it is difficult to attain accurate forecasts for air pollution indexes because the original data are non-stationary and chaotic. The existing forecasting methods, such as multiple linear models, autoregressive integrated moving average (ARIMA) and support vector regression (SVR), cannot fully capture the information from series of pollution indexes. Therefore, new effective techniques need to be proposed to forecast air pollution indexes. The main purpose of this research is to develop effective forecasting models for regional air quality indexes (AQI) to address the problems above and enhance forecasting accuracy. Therefore, two hybrid models (EMD-SVR-Hybrid and EMD-IMFs-Hybrid) are proposed to forecast AQI data. The main steps of the EMD-SVR-Hybrid model are as follows: the data preprocessing technique EMD (empirical mode decomposition) is utilized to sift the original AQI data to obtain one group of smoother IMFs (intrinsic mode functions) and a noise series, where the IMFs contain the important information (level, fluctuations and others) from the original AQI series. LS-SVR is applied to forecast the sum of the IMFs, and then, S-ARIMA (seasonal ARIMA) is employed to forecast the residual sequence of LS-SVR. In addition, EMD-IMFs-Hybrid first separately forecasts the IMFs via statistical models and sums the forecasting results of the IMFs as EMD-IMFs. Then, S-ARIMA is employed to forecast the residuals of EMD-IMFs. To certify the proposed hybrid model, AQI data from June 2014 to August 2015 collected from Xingtai in China are utilized as a test case to investigate the empirical research. In terms of some of the forecasting assessment measures, the AQI forecasting results of Xingtai show that the two proposed hybrid models are superior to ARIMA, SVR, GRNN, EMD-GRNN, Wavelet-GRNN and Wavelet-SVR. Therefore, the proposed hybrid models can be used as effective and simple tools for air pollution forecasting and warning as well as for management.

1. Introduction

Air pollution is an important environmental problem in many parts of the world (Kurt and Oktay, 2010). On the one hand, along with rapid development, the emission of substandard industrial pollutants is a major cause of severe air pollution in industrialized areas. On the other hand, there has been a dramatic increase in population coupled with rapid economic development. Atmospheric pollution in China is very serious and is mainly reflected as a high concentration of suspended particles in the urban atmospheric environment. For example, there is a high level of sulfur dioxide (SO_2) and particulate matter (PM), along with a rapid increase in automotive exhaust emissions, and an aggravating tendency in nitrogen oxides (NO_x) pollution. In recent years, air pollution has led to increasingly hazy weather in China.

The public has become increasingly concerned about the atmospheric environment because air quality affects everyone (Zhang et al., 2012). Furthermore, air pollution may have serious impacts on human health, including asthma, impaired lung function, cardiopulmonary illnesses (Yahya et al., 2014), obstacles to physiological functions, and increased mortality rates (Mindell and Joffe, 2004). Therefore, there are many suggestions to protect the population from heavy pollution, for example: i) the general public should reduce outdoor activities; ii) some vulnerable populations,

☆ This paper has been recommended for acceptance by Dr. Hageman Kimberly Jill.
* Corresponding author.
E-mail address: 819612640@qq.com (X. Lian).

http://dx.doi.org/10.1016/j.envpol.2017.08.069
0269-7491/

including children, pregnant women, the elderly, and patients with respiratory disease and cardiovascular diseases, should stay indoors as much as possible and stop outdoor activities; and iii) primary schools and kindergartens should reduce outdoor physical education and outdoor activities (Sheng and Tang, 2015). Therefore, air pollution forecasting plays a vital role in people's daily life, as well as in warning and controlling air pollution.

The air quality index (AQI) and $PM_{2.5}$ are two important indicators among pollution indexes, where $PM_{2.5}$ is particulate matter with a diameter of less than or equal to *2.5* μm. AQI is an indicator of air quality which reflects and evaluates the air quality status, which simplifies the concentrations of several pollutants into one single numerical form. The AQI is calculated with reference to the new ambient air quality standards (GB3095-2012), which covers six pollutants, including sulfur dioxide (SO_2), nitrogen dioxide (NO_2), $PM_{2.5}$, PM_{10} (particulate matter with a diameter less than or equal to 10 μm), ozone (O_3), and carbon monoxide (CO) (Sheng and Tang, 2015). There existed different definitions of AQI, such as fuzzy-based air quality index (FAQI, Sowlat et al., 2011) which defined the AQI by weighting the CO, PM_{10}, SO_2, NO_2, O_3 with fuzzy criterion, air pollution index (API, Yuan and Liu, 2014; Chen et al., 2016) which only covered CO, PM_{10}, SO_2, NO_2 and O_3. The popular pollution index AQI covers six pollutants ($PM_{2.5}$, CO, PM_{10}, SO_2, NO_2, O_3), so it is analyzed in this research. The definition of the used AQI is presented in the appendix (Yuan and Liu, 2014). Intuitively, it can be seen that the daily AQI time series shows the average pollution trend changes along with days. For the general public, the AQI is an important index that can be used to easily understand whether the air quality is bad or good. It is also helpful in data interpretation for decision making processes related to pollution mitigation measures and air quality management (Kumar and Goyal, 2011). According to air quality standards (GB3095-2012) and various impacts on human health, the AQI is divided into six classes (Fig. 1). Corresponding to the six classes of air quality, as the AQI increases, the level of pollution increases. In terms of different physical qualities and AQI, there are different suggestions for different people (Fig. 1).

Over the past few decades, air pollution warnings have caught the world's attention. Currently, many people make decisions on outdoor actions according to the air pollution forecasts to avoid the effects of atmospheric pollution on human health. Therefore, developing an accurate and effective AQI forecasting model is an important topic. The air and pollutants move in different ways and directions, mainly through natural causes and atmospheric phenomena (Kurt and Oktay, 2010). In fact, the atmosphere is an intricate dynamic system that is quite difficult to model (Kurt and Oktay, 2010). Therefore, forecasting the AQI or any other pollution index is not easy.

Faced with such a problem, some researchers have proposed techniques to forecast PM_{10}, $PM_{2.5}$ and other pollution indicators. According to the pattern of data processing, the related forecasting models can be classified into two classes: empirical models and chemical transport models (CTMs) (Cobourn, 2010). Konovalov et al. (2009) noted that chemical transport models are worse than empirical models for forecasting the air index PM_{10} because CTMs can simulate some components reasonably well, but cumulative errors from poorly modeled or missing components lead to relatively large errors in simulated $PM_{2.5}$ concentrations (Cobourn, 2010). Therefore, the forecasting models for AQI and other air pollution indexes should take full advantages of individual models or revise the results of CTMs according to the residual sequence of forecasting (Cobourn, 2010).

In terms of empirical models, statistical models are widely applied to forecast various air pollution indexes, for example, the ARIMA (autoregressive integrated moving average) model, multiple linear model (MLR), artificial neural networks (ANNs), support vector regressions (SVRs) and hybrid models. The ARIMA model is a classical statistical modeling technique for analyzing nonlinear

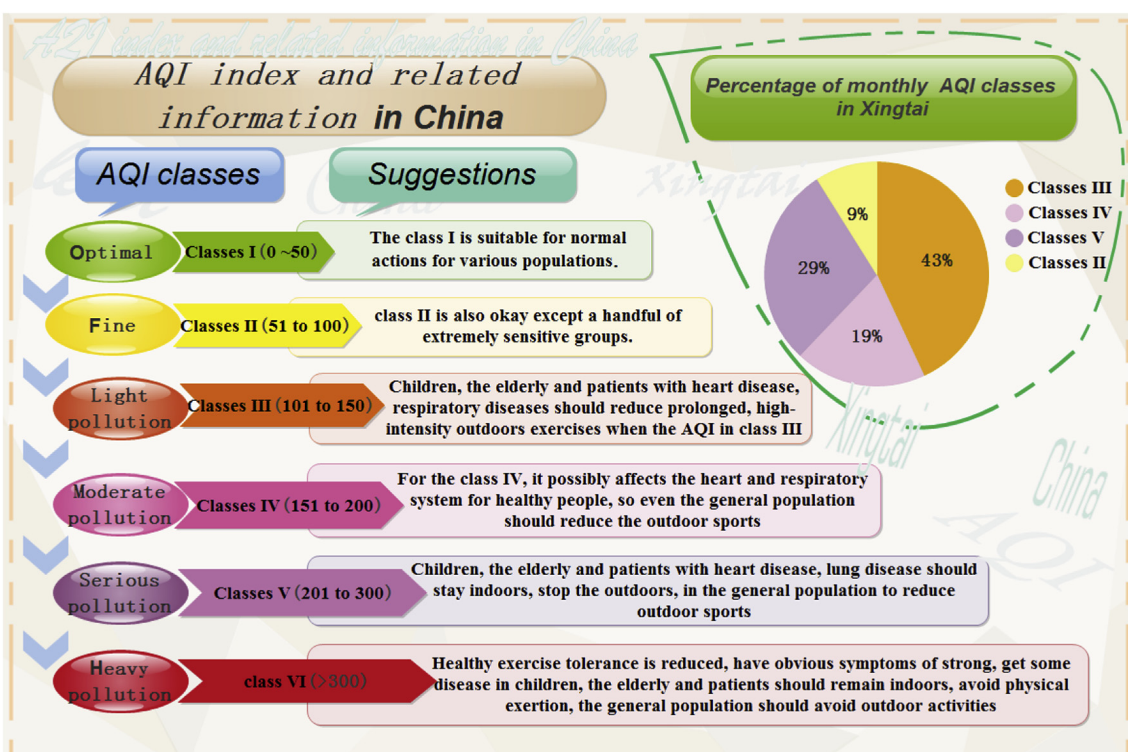

Fig. 1. The AQI index and related information in China.