Environmental Pollution xxx (2017) 1–14

Contents lists available at ScienceDirect

Environmental Pollution

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

Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: A case study in China☆

Yunzhen Xu [a], Pei Du [b,*], Jianzhou Wang [b]

[a] School of Basic Medical Science, Lanzhou University, Lanzhou, 730000, Gansu Province, China
[b] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

A R T I C L E I N F O

Article history:
Received 7 September 2016
Received in revised form
6 November 2016
Accepted 16 January 2017
Available online xxx

Keywords:
Forecasting and early warning system
Fuzzy synthetic evaluation
Hybrid forecasting model
Emissions distribution evaluation
Interval forecasting

A B S T R A C T

As the atmospheric environment pollution has been becoming more and more serious in China, it is highly desirable to develop a scientific and effective early warning system that plays a great significant role in analyzing and monitoring air quality. However, establishing a robust early warning system for warning the public in advance and ameliorating air quality is not only an extremely challenging task but also a public concerned problem for human health. Most previous studies are focused on improving the prediction accuracy, which usually ignore the significance of uncertainty information and comprehensive evaluation concerning air pollutants. Therefore, in this paper a novel robust early warning system was successfully developed, which consists of three modules: evaluation module, forecasting module and characteristics estimating module. In this system, a new dynamic fuzzy synthetic evaluation is proposed and applied to determine air quality levels and primary pollutants, which can be regarded as the research objectives; Moreover, to further mine and analyze the characteristics of air pollutants, four different distribution functions and interval forecasting method are also employed that can not only provide predictive range, confidence level and the other uncertain information of the pollutants future values, but also assist decision-makers in reducing and controlling the emissions of atmospheric pollutants. Case studies utilizing hourly $PM_{2.5}$, PM_{10} and SO_2 data collected from Tianjin and Shanghai in China are applied as illustrative examples to estimate the effectiveness and efficiency of the proposed system. Experimental results obviously indicated that the developed novel early warning system is much suitable for analyzing and monitoring air pollution, which can also add a novel viable option for decision-makers.

1. Introduction

With the rapid development of urbanization and industrialization, as well as the rapid growth of energy consumption, atmospheric environment pollution has been becoming more and more serious. The dominant pollutant, Particulate matter (PM) contains a lot of poisonous and harmful substances, which can induce many diseases, including asthma, respiratory disease, lung cancer, etc. (Qin et al., 2014). These related issues have already been paid high attention to by public and the corresponding departments in China. In recent years, a great mass of areas frequently encounter serious smog, haze and PM, in the light of figures from China's Ministry of Environmental Protection (CMEP), involving Jing-Jin-Ji region, northeastern China, south and central China, and other areas. These regions account for 25% of China's area, where some six hundred million people are affected (Niu et al., 2016).

With the aim of improving and protecting air quality, a sequence of relative initiatives have already been implemented or are in the process by the environmental managing departments. For example, about 946 air pollution monitoring stations distributed in 190 cities across the country, have been established for obtaining the real-time data, air quality evaluation standards are constantly improved by the CMEP, and the new ambient air quality standard (AAQS) (GB 3095-2012) was published in 2012, in which the annual average values of $PM_{2.5}$, PM_{10}, SO_2, and NO_2 are 15 $\mu g/m^3$, 40 $\mu g/m^3$, 20 $\mu g/m^3$ and 40 $\mu g/m^3$, respectively (Chen et al., 2013). Additionally, the State Council of China announces that by 2017, the

☆ This paper has been recommended for acceptance by B. Nowack.

* Corresponding author.
E-mail addresses: xuyzh15@lzu.edu.cn (Y. Xu), renshengdp@126.com (P. Du), wjz@lzu.edu.cn (J. Wang).

http://dx.doi.org/10.1016/j.envpol.2017.01.043
0269-7491/

Please cite this article in press as: Xu, Y., et al., Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: A case study in China, Environmental Pollution (2017), http://dx.doi.org/10.1016/j.envpol.2017.01.043

Nomenclature

PM	Particulate matter
AQI	air quality index
AAQS	ambient air quality standard
AATs	average actual values
MLR	multi-linear regression
SVM	support vector machine
ANNs	artificial neural networks
SVR	support vector regression
FL	fuzzy logic
ALO	ant lion optimizer
SA	simulated annealing
CS	cuckoo search
KH	krill herd
FA	firefly algorithm
BA	bat algorithm
GWO	grey wolf optimizer
WT	wavelet transform
IMFs	Intrinsic Mode Functions
WPT	wavelet packet transform
MAE	mean absolute error
DFs	Distribution functions
IF	interval forecasting
R^2	determination coefficient
IFAW	IF average width
IFCP	IF coverage probability
CMEP	China's ministry of environmental protection
FSE	fuzzy synthetic evaluation
CTMs	chemical transport models
FSE	fuzzy synthetic evaluation
FNN	fuzzy neural network
MLE	maximum likelihood estimate
BPNN	back propagation neural network
WNN	wavelet neural network
PSO	particle swarm algorithm
ERNN	elman recurrent neural network
GA	genetic algorithm
GRNN	generalized regression neural network
ELM	extreme learning machine
ARIMA	autoregressive integrated moving average
CSO	chicken swarm algorithm
EEMD	ensemble empirical mode decomposition
SSA	singular spectrum analysis
EMD	empirical mode decomposition
CEEMD	complementary ensemble empirical mode decomposition
MAPE	mean absolute percentage error
RMSE	root average of the prediction error squares
CDF	cumulative distribution function
PDF	probability density function
y_i	observed value of the *i-th* datum
$\hat{y}_i$	predicted value of the *i-th* datum

respirable particulate matter (PM_{10}) concentration, in prefecture-level and above cities, will be lower over 10% than that in 2012; the number of good air quality days gradually increase and the concentration of fine particulate matter ($PM_{2.5}$) in Jing-Jin-Ji, Yangtze River Delta and Pearl River Delta areas decreased by 25%, 20%, 15%, respectively. Thereinto, Beijing's annual average concentration of $PM_{2.5}$ must be controlled within 60 μg/m^3; (State Council of China, 2013-09).

However, it's far from sufficient to improve air quality relying solely on these above mentioned measures. Fortunately, in recent years, many different approaches have been proposed for estimating, mining and forecasting PM levels that are very advantageous to establish the forecasting and early warning systems by applying monitoring data of air pollutants. However, the relevant researches and experiences of analyzing and predicting PM are still not comprehensive enough, although quite important. Generally speaking, the predictive styles of these methods can be summed up in two categories, i.e., deterministic prediction (Feng et al., 2015) and uncertainty prediction (Song et al., 2015) about PM and SO_2 trends in future, nevertheless none of them combine the advantages of both styles for analyzing and forecasting the characteristics and concentration of the coming PM and SO_2. What's more, few of them analyze or point out the reasons why they regard $PM_{2.5}$, PM_{10} or O_3 etc. as their study objects rather than the other pollutants. In order to avoid this blindness and improve air quality, this paper proposed a novel forecasting and early warning systems for improving the air quality, i.e., a dynamic fuzzy synthetic evaluation was applied to determine air quality levels and primary pollutants, which can be regarded as the research objectives, and then our research attempts from these two aspects (deterministic and uncertainty) mentioned above to forecasting $PM_{2.5}$, PM_{10} and SO_2 levels.

As we all known, the uncertain prediction, vital for forecasting PM and establishing the forecasting and early warning systems, which is based on the deterministic prediction, the better the prediction performance of the latter is, the higher the accuracy of the former is. Therefore, one of the main steps is to determine the appropriate forecasting method. According to the incalculable published related articles, the developed approaches for forecasting the concentration of atmospheric pollutants, to date, can also be divided into two classes: deterministic models, also called chemical transport models (CTMs) (e.g. CMAQ (Fernando et al., 2012), MOZART, CHIMERE (Dutot et al., 2007), CLaMS), and empirical statistical models, which chiefly including multi-linear regression (MLR) (Genc et al., 2010), fuzzy logic (FL) (Xiao et al., 2015), autoregressive integrated moving average (ARIMA) (Jian et al., 2012), support vector machine (SVM) (Luna et al., 2014; Wang et al., 2015a), artificial neural networks (ANNs) (Gennaro et al., 2013), etc.

Concentrating on the sources and transport of chemical species, many CTMs have been proposed for real-time air quality forecast guidance in North America and Europe (http://www.nws.noaa.gov/aq). Meanwhile, due to their significant approximations and uncertainties, great complexity of atmospheric chemical and transport processes generally lead to lower accuracy than well-developed. However, compared to CTMs, empirical statistical models apply multifarious statistical or machine learning techniques to build respective relationships between the observed