Computers and Mathematics with Applications 62 (2011) 2016–2028

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Predicting daily ozone concentration maxima using fuzzy time series based on a two-stage linguistic partition method

Ching-Hsue Cheng [a], Sue-Fen Huang [a,*], Hia-Jong Teoh [b]

[a] *Department of Information Management, National Yunlin University of Science & Technology, YunLin, Taiwan*

[b] *Department of Accounting and Information Technology, Ling Tung University, Taiwan*

ARTICLE INFO

Article history:
Received 7 January 2010
Received in revised form 21 June 2011
Accepted 22 June 2011

Keywords:
Air quality
Pollutant standards index
Fuzzy time series
Uniform discretion method
Cumulative probability distribution approach

ABSTRACT

Air pollution is a result of global warming, greenhouse effects, and acid rain. Especially in highly industrialization areas, air pollution has become a major environmental issue. Poor air quality has both acute and chronic effects on human health. The detrimental effects of ambient ozone on human health and the Earth's ecosystem continue to be a national concern in Taiwan. The pollutant standard index (PSI) has been adopted to assess the degree of air pollution in Taiwan. The standardized daily air quality report provides a simple number on a scale of 0 to 500 related to the health effects of air quality levels. The report focuses on health and the current PSI subindices to reflect measured ozone (O_3) concentrations. Therefore, this study uses the O_3 attribute to evaluate air quality. In an effort to forecast daily maximum ozone concentrations, many researchers have developed daily ozone forecasting models. However, this continuing worldwide environmental problem suggests the need for more accurate models. This paper proposes two new fuzzy time series based on a two-stage linguistic partition method to predict air quality with daily maximum O_3 concentration: Stage 1, use the fuzzy time series based on the cumulative probability distribution approach (CPDA) to partition the universe of discourse into seven intervals; Stage 2, use two linguistic partition methods, the CPDA and the uniform discretion method (UDM), to repartition each interval into three subintervals. To verify the forecasting performance of the proposed methods in detail, the practical collected data is used as and evaluating dataset; five other methodologies (AR, MA, ARMA, Chen's and Yu's) are used as comparison models. The proposed methods both show a greatly improved performance in daily maximal ozone concentration prediction accuracy compared with the other models.

1. Introduction

Air pollution is the introduction of chemicals, particulate matter, or biological materials that cause harm or discomfort to humans or other living organisms, or damage the natural environment, into the atmosphere. As one of major tropospheric photochemical oxidants, ozone depletion due to air pollution has long been recognized as a threat to human health as well as to the Earth's ecosystems. The World Health Organization states that 2.4 million people die each year from causes directly attributable to air pollution. About 4% of deaths in the United States can be attributed to air pollution, according to the Environmental Science Engineering Program at the Harvard School of Public Health. Health effects of air pollution

* Corresponding author. Tel.: +886 923602532.
E-mail address: g9623801@yuntech.edu.tw (S.-F. Huang).

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.06.044

Table 1
PSI pollutant subindices and influence on health.

PSI	PM_{10} (μg/m3)	SO_2 (ppb)	CO (ppm)	O_3 (ppb)	NO_2 (ppb)	Category
50	50	30	4.5	60	–	Good
100	150	140	9	120	–	Moderate
200	350	300	15	200	600	Unhealthy
300	420	600	30	400	1200	Very unhealthy
400	500	800	40	500	1600	Hazardous
500	600	1000	50	600	2000	

have been studied in many different parts of the world. Thus, monitoring and forecasting the parameters of air quality are popular and important topics. Accurate models for prediction are needed because such models would allow compliance and noncompliance in both short-term and long-term aspects.

Air quality forecasting is among the most common environmental forecasting applications, and is usually performed by air quality agencies or authorities responsible for the monitoring and management of the atmospheric environment in urban agglomerations. As photochemical air pollution is one of the most pronounced air quality problems of developed countries, ozone forecasting is usually part of the core of every air quality forecasting system or application. The basic technology for analyzing air pollution is through the use of a variety of mathematical models for predicting the transport of air pollutants in the lower atmosphere.

It is known that ozone (O_3) is an effective anti-greenhouse gas, particularly in the upper troposphere, thus playing a direct role in climate change. Ozone regulates the oxidizing capacity of the atmosphere via production of the OH radical that acts as the principal cleaning agent in the atmosphere. In the lower atmosphere, elevated ozone is a pollutant, and it has adverse effects on both human health and foliage [1–3]. For example, elevated ozone levels may cause eye irritation, coughs, reduced athletic performance, and possible chromosome damage [4–6]. In Taiwan, the pollutant standard index (PSI) has been adopted to assess air pollution. This is a standardized daily air quality reporting tool that provides a simple number on scale of 0 to 500 related to the health effects of air quality levels. A PSI 24 h monitored value less 100 means that the air quality will be acceptable (see Table 1). A number of clinical studies have focused on multi-day exposure (100–800 lg/m^3) to ozone, which shows that, on repeated daily exposures to ozone, the determinant of lung function will occur after the first exposures, followed by a decrease on subsequent exposures; since ozone is a very strong oxidizing agent, even at low density it can also affect the eyes, nose, and throat. At high density it will increase the risk of respiratory throat problems [7]. Thus, the prediction of ozone concentrations is a fundamental tool for the anticipated implementation of strategies in public health.

The development of tools for predicting ozone concentrations has been drawing the attention of many scientists all over the world during the past decade [8–21,1,22–28,2,3,29,30]. The prediction of the ozone level for the next day is done on the basis of information including (1) past historical data, (2) meteorological parameters, such as the speed and direction of the wind, the temperature, humidity, and pressure, (3) environmental pollutant concentrations of particulate material (PM_{10}), ozone (O_3), sulfur dioxide (SO_2), carbon monoxide (CO), and nitrogen dioxide (NO_2). There are numerous reports on various statistical modeling methods to improve air quality forecasting, such as multiple linear regression (MLR), nonlinear regression (NLR), neural networks (NNs), and fuzzy systems. Chelani and Devotta [31] used a combination of an autoregressive integrated moving average model, which deals with linear patterns, and a nonlinear dynamical model, to approximate the linear relationships between ozone and its predictor variables. The performance is compared for one-step-ahead and multi-step-ahead forecasts using the error statistics (mean absolute percentage error and relative error). It is observed that the hybrid model outperforms the individual linear and nonlinear models. Alawi et al. [23] combine principal component regression (PCR) and an artificial neural network (ANN) to relate the lower ozone concentration to both meteorological variables as well as to concentrations of primary pollutants. Hence, the forecasting accuracy of combining PCR and an ANN is higher than that of individual forecasts. Lu et al. [12] developed a two-stage neural network to predict ozone concentrations from meteorological conditions. The two-stage neural network first utilized an unsupervised neural network (two-level clustering approach: SOM followed by *K*-means clustering) to cluster meteorological conditions into different meteorological regimes. Then a supervised multi-layer perceptron (MLP) neural network was used to simulate the nonlinear ozone–meteorology relationship within each meteorological regime. The results showed that meteorological conditions can explain at least 60% variance of ozone concentrations obtained by the two-stage neural network. Sousa et al. [1] predicted next-day hourly ozone concentrations through a new methodology based on feed-forward artificial neural networks using principal components as inputs; the results showed that the use of principal components as inputs improved the prediction of both models by reducing their complexity and eliminating data collinearity. Mintz et al. [32] used an automated fuzzy logic method, modified learning from examples (MLFE), whose computation is simple, and which analyzes training data one-by-one in order to generate a series of rules describing the system. Lin and Cobourn [30] developed Takagi–Sugeno fuzzy system ozone forecast models, which used the same input variables as existing NLR models, but the maximum temperature, wind speed, and special relative humidity were input as a resultant parameter, or nonlinear term extracted from the NLR model. They demonstrated that the fuzzy system model could provide better predictions of ozone than an MLR model.