Science of the Total Environment 580 (2017) 719–733

Contents lists available at ScienceDirect

Science of the Total Environment

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

A novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine

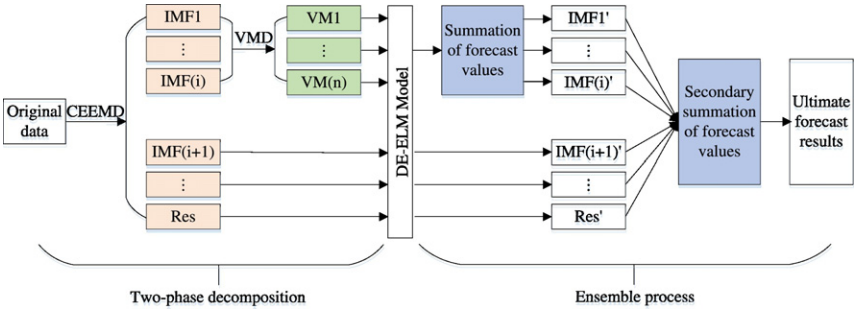
Deyun Wang [a,b,c,*], Shuai Wei [a,b], Hongyuan Luo [a,b], Chenqiang Yue [a,b], Olivier Grunder [c]

[a] School of Economics and Management, China University of Geosciences, Wuhan 430074, China
[b] Mineral Resource Strategy and Policy Research Center, China University of Geosciences, Wuhan 430074, China
[c] Université de Bourgogne Franche-Comté, UTBM, IRTES, Rue Thierry Mieg, 90010 Belfort cedex, France

HIGHLIGHTS

- A novel two-phase decomposition technique is proposed for the AQI series decomposition.
- VMD is employed to conduct the secondary decomposition of IMFs with high frequencies.
- The ELM model optimized by DE algorithm has strong function approximation ability.
- The proposed model is applied to AQI forecasting.
- The proposed model outperforms other considered models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 2 August 2016
Received in revised form 12 November 2016
Accepted 2 December 2016
Available online 16 December 2016

Keywords:
Air quality index (AQI)
Complementary ensemble empirical mode decomposition (CEEMD)
Variational mode decomposition (VMD)
Differential evolution (DE)
Extreme learning machine (ELM)

ABSTRACT

The randomness, non-stationarity and irregularity of air quality index (AQI) series bring the difficulty of AQI forecasting. To enhance forecast accuracy, a novel hybrid forecasting model combining two-phase decomposition technique and extreme learning machine (ELM) optimized by differential evolution (DE) algorithm is developed for AQI forecasting in this paper. In phase I, the complementary ensemble empirical mode decomposition (CEEMD) is utilized to decompose the AQI series into a set of intrinsic mode functions (IMFs) with different frequencies; in phase II, in order to further handle the high frequency IMFs which will increase the forecast difficulty, variational mode decomposition (VMD) is employed to decompose the high frequency IMFs into a number of variational modes (VMs). Then, the ELM model optimized by DE algorithm is applied to forecast all the IMFs and VMs. Finally, the forecast value of each high frequency IMF is obtained through adding up the forecast results of all corresponding VMs, and the forecast series of AQI is obtained by aggregating the forecast results of all IMFs. To verify and validate the proposed model, two daily AQI series from July 1, 2014 to June 30, 2016 collected from Beijing and Shanghai located in China are taken as the test cases to conduct the empirical study. The experimental results show that the proposed hybrid model based on two-phase decomposition technique is remarkably superior to all other considered models for its higher forecast accuracy.

1. Introduction

With the rapid development of urbanization and industrialization, traffic scale and energy demand continue to expand. According to

* Corresponding author at: School of Economics and Management, China University of Geosciences, Wuhan 430074, China.
E-mail address: wang.deyun@hotmail.com (D. Wang).

http://dx.doi.org/10.1016/j.scitotenv.2016.12.018
0048-9697/

2035 BP World Energy Outlook (British Petroleum (BP), 2015), carbon emissions from energy consumption in 2035 will increase by 60% than 2013. Large numbers of contaminants, such as CO, CO_2, SO_2, CH_4, NO_x and atmospheric particulate, are released into the atmosphere increasingly, which results in serious air pollution problems. During 2015, China suffered from eleven widespread and persistent hazes, especially the one from 27th November to 1st December making a significant impact on North China, Henan Province and Shandong Province (China Meteorological Bureau, 2016). In recent years, air pollution gradually spreads from a single city to the contiguous region, presenting an obvious regional feature. AQI is an important indicator to measure how the air quality is, which is closely bound up with human health (Harrison and Yin, 2000; Ribeiro et al., 2016). Thus, how to forecast the AQI precisely plays an important role in both controlling air pollution and promoting the sustainable development of human society.

Air quality forecasting has aroused great social attention, and great efforts have been made by scholars in air quality forecasting in recent years, which mainly covers the concentrations forecasting of PM_{10}, $PM_{2.5}$, SO_2, NO_x, CO and O_3. The related forecasting methods can be classified into three main categories: deterministic models, statistical models and hybrid models. The deterministic methods can simulate the process of discharge, accumulation, diffusion and transfer of a pollutant by employing meteorological, emission and chemistry models (Chen et al., 2013). However, the scale and quality of the emission data adopted make great influence on the forecast accuracy of deterministic methods (Zhang et al., 2012). Moreover, the deterministic methods need much computational time to accomplish the forecasting process (Niu et al., 2016). Therefore, many researchers integrated the deterministic methods with intelligent or statistical methods in order to enhance the forecast accuracy. For instance, Konovalov et al. (2009) proposed a hybrid model for PM_{10} forecasting based on the deterministic model and statistical model, where the statistical model was utilized to correct predictions made by the deterministic model. Song et al. (2015) developed an adaptive neuro-fuzzy model to implement deterministic forecasting of PM based on the data of hourly and 12 h averaged air pollutants within Yangtze River Delta region of China.

Compared with deterministic models, statistical models are much easier, quicker and more efficient, which mainly involve autoregressive integrated moving average model (ARIMA), multiple linear regression (MLR), generalized autoregressive conditional heteroskedasticity (GARCH), grey model (GM), Markov models and artificial intelligence (AI) based models. For example, Jian et al. (2012) utilized the ARIMA model to investigate the effect of meteorological factors on submicron particle concentrations under busy traffic conditions. Agirre-Basurko et al. (2006) proposed two MLP based models and one MLR based model to forecast the O_3 and NO_2 levels in real-time up to 8 h ahead at four stations in Bilbao located in Spain. Stadlober et al. (2008) developed a MLR model combining information of the present day with meteorological forecasts of the next day to forecast daily PM_{10} concentrations for sites located in Bolzano, Klagenfurt and Graz. Kumar and Ridder (2010) developed a hybrid model integrated GARCH modelling technique and FFT-ARIMA to forecast daily maximum O_3 concentrations, and to make probabilistic forecasts of ozone episodes at four urban sites of two major European cities (London and Brussels), the results revealed the good forecasting performance of the proposed model. Pai et al. (2011) proposed seven types of first-order and one-variable grey differential equation model to predict hourly PM concentrations in Banciao city of Taiwan, the results showed that GM (1,1) is an efficient early-warning tool for providing PM information to the inhabitants. Sun et al. (2013) utilized hidden Markov models (HMMs) to forecast daily average $PM_{2.5}$ concentrations, the comparisons between different distributions used in HMMs showed that the closer the distribution employed in HMMs is to the observation sequence, the better the model prediction performance. Ordieres et al. (2005) compared the forecast capability of three different neural networks for $PM_{2.5}$ concentrations forecasting, including multilayer perceptron (MLP), radial basis function (RBF) and square multilayer perceptron (SMLP). Voukantsis et al. (2011) constructed an ANN (multi-layer perceptron) model to accomplish the forecasting of daily PM_{10} and $PM_{2.5}$ concentrations.

With consideration of forecast accuracy, hybrid models which combine some single models are widely used in air quality forecasting field. For example, Slini et al. (2006) proposed a hybrid system (PCA-MLP) to forecast PM_{10} concentrations in Thessaloniki. Díaz-Robles et al. (2008) presented a hybrid model combining ARIMA and ANN to improve the forecast accuracy of PM_{10} concentrations. Lin et al. (2011) forecasted concentrations of air pollutants including PM_{10}, NO_x and NO_2 by logarithm SVM with immune algorithms (IA). Perez (2012) developed a combination model of ANN and a nearest neighbor method to conduct PM_{10} forecasting. Antanasijević et al. (2013) applied the artificial intelligence model combining the genetic algorithm (GA) and ANN to forecast PM_{10} emission. The complexity and difficulty of air quality forecasting urge us to pursue better prediction methods. Fortunately, a promising idea of "decomposition and ensemble" has been developed gradually and applied to forecast time series in different areas (Yu et al., 2015). This "decomposition and ensemble" procedure mainly includes the following three steps: (1) decompose the original time series into a set of independent components; (2) predict each component; (3) aggregate the forecast values of all components. Some empirical researches have confirmed that the idea of decomposition and reconstruction can provide good prediction results for time series forecasting. For instance, Bai et al. (2016) utilized wavelet transform (WT) technique and back propagation neural network (BPNN) to forecast daily air pollutants (PM_{10}, SO_2 and NO_2) concentrations, and experiment results showed that the W-BPNN model has better forecasting performance than mono-BPNN model in terms of the statistics indexes and forecast accuracy. Zhou et al. (2014) developed a hybrid model based on EEMD and a general regression neural network (GRNN) to predict $PM_{2.5}$ concentrations, and obtained accurate results. Liu et al. (2015) presented four different hybrid models by combining four signal decomposing algorithms (e.g., wavelet decomposition (WD)/wavelet packet decomposition (WPD)/EMD/fast ensemble empirical mode decomposition (FEEMD)) and ELM model to complete the multi-step-ahead wind speed forecasting, the experiments indicated that all the proposed hybrid models have better performance than the single ELM model. Yu et al. (2016) built a novel decomposition and ensemble model integrating EEMD and extended extreme learning machine (EELM) for crude oil price forecasting, results indicated that the hybrid model is a promising tool for predicting complicated time series with high volatility and irregularity.

However, the deficiencies still exist in the single decomposition methods mentioned above. The single decomposition technique often cannot thoroughly handle non-stationarity of random and irregular data series. Thus, borrowing the idea of hybrid models which integrate the advantages of single models, this paper proposes a hybrid two-phase decomposition technique based on CEEMD and VMD for dealing with the non-stationary characteristics associated with AQI series in order to improve the forecast accuracy. In this proposed two-phase decomposition technique, VMD method, as a new technique of signal decomposition, is introduced to conduct the secondary decomposition of IMFs with high frequencies. Accordingly, a novel hybrid model based on the two-phase decomposition technique and ELM model optimized by DE algorithm is developed for AQI forecasting in this paper.

The remainder of this paper is organized as follows. Section 2 briefly describes the methodologies adopted in this paper, including CEEMD, VMD, ELM and DE algorithm. The hybrid forecasting model is proposed in Section 3 on the basis of Section 2. The case studies of AQI forecasting in Beijing and Shanghai are presented respectively in Section 4. Finally, Section 5 summarizes the main findings based on the forecast results.