The forecasting research of early warning systems for atmospheric pollutants: A case in Yangtze River Delta region

Yiliao Song\textsuperscript{a}, Shanshan Qin\textsuperscript{b,c,*}, Jiansheng Qu\textsuperscript{d}, Feng Liu\textsuperscript{a}

\textsuperscript{a} School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
\textsuperscript{b} MOE Key Laboratory of Western China’s Environmental Systems, College of Earth and Environmental Sciences, Lanzhou University, Lanzhou 730000, China
\textsuperscript{c} Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada
\textsuperscript{d} Information Center for Global Change Studies, Lanzhou Information Center, Chinese Academy of Sciences, Lanzhou 730000, China

HIGHLIGHTS

- The deficiency of forecasting and early warning systems is emphasized in this paper.
- Interval prediction is proposed for addressing the uncertainty of PMs.
- An artificial intelligence method is introduced to improve performance.
- The results are validated well in the Yangtze River Delta region in China.

ARTICLE INFO

Article history:
Received 31 January 2015
Accepted 17 June 2015
Available online 23 June 2015

Keywords:
Particle matter (PM)
Emissions distribution
Adaptive neuro-fuzzy (ANF) model
Dynamic interval forecasts
Forecasting and early warning systems

ABSTRACT

The issue of air quality regarding PM pollution levels in China is a focus of public attention. To address that issue, to date, a series of studies is in progress, including PM monitoring programs, PM source apportionment, and the enactment of new ambient air quality index standards. However, related research concerning computer modeling for PM future trends estimation is rare, despite its significance to forecasting and early warning systems. Thereby, a study regarding deterministic and interval forecasts of PM is performed. In this study, data on hourly and 12 h-averaged air pollutants are applied to forecast PM concentrations within the Yangtze River Delta (YRD) region of China. The characteristics of PM emissions have been primarily examined and analyzed using different distribution functions. To improve the distribution fitting that is crucial for estimating PM levels, an artificial intelligence algorithm is incorporated to select the optimal parameters. Following that step, an ANF model is used to conduct deterministic forecasts of PM. With the identified distributions and deterministic forecasts, different levels of PM intervals are estimated. The results indicate that the lognormal or gamma distributions are highly representative of the recorded PM data with a goodness-of-fit $R^2$ of approximately 0.998. Furthermore, the results of the evaluation metrics (MSE, MAPE and CP, AW) also show high accuracy within the deterministic and interval forecasts of PM, indicating that this method enables the informative and effective quantification of future PM trends.

© 2015 Published by Elsevier Ltd.

1. Introduction

The thick smog and haze along with particulate matter (PM) that frequently occurs and covers most cities of China is an issue of great focus by the environmental management ministry and public attention. Apart from sulfur dioxide and nitrogen oxides, which are widely known as major air pollutants, PM is more harmful and deserves more attention. Among them, PM$_{2.5}$ has the greatest effect on health; exposure to it has been associated with hospital admissions and several serious health effects, including premature death. People with asthma, cardiovascular or lung disease, as well as children and the elderly, are considered to be the populations most sensitive to PM$_{2.5}$ (MECC, 2012). A related published study has...
found that China’s poor air quality is reducing the lifespan of its citizens. Between 1981 and 2001, PM in China hovered approximately 400 μg/m². According to this study, every 100 μg of pollutants lowers the life expectancy at birth by three years (News, 2013). In addition, PM is also responsible for environmental effects such as corrosion, soil pollution, damage to vegetation and reduced visibility (MECC, 2012).

Adverse effects on health and on the natural environment have been associated with exposure to PM over both short periods (such as a day) and longer periods (a year or more). Thus, we desire for our air to be as clean as possible. The environment ministry works on reducing PM levels to protect and improve air quality through, for instance, new regulatory initiatives and targeted programs. China is progressing very quickly. Although some of China’s achievements in air quality have been attained, the continuing challenges of reducing PM concentrations still need to be addressed. The Ministry of Environmental Protection (MEP) of China promulgated new ambient air quality standards (AQS) (GB 3095-2012) in 2012 that established new standards for PM2.5 (e.g., annual average value of 35 μg/m²), among others, and tightened the annual average value for PM10 from 100 to 70 μg/m² (World Bank, 2012).

China’s environmental supervisors have recently issued the country’s most comprehensive and toughest plans to control and in some regions reduce air pollution by the year 2017, setting stricter limits on the levels of PM2.5. Many cities have already announced their updated targets to control PM2.5. The target for Beijing is to maintain the concentration of PM2.5 at approximately 60 μg/m³ by 2017, a level approximately 25 percent lower than 2012 levels. Officials from Shanghai and Tianjin have vowed that by 2017 their PM2.5 levels will be reduced by 20 percent compared with 2012 levels. Shijiazhuang, the capital of Hebei, has promised to reduce emissions by 30 percent from its 2012 levels (State Council, 2013).

Nevertheless, before all Chinese cities attain air quality standards that comply with the AQS recommended by China and the World Health Organization (WHO), a series of work is needed, including a monitoring program for providing actual concentrations, air quality calculations and computer modeling for predicting future PM trends. Emissions inventories that describe the sources and categories of emissions for a specific pollutant, studies on control strategies for reducing emissions effectively, the formal adoption of measures to ensure that the reductions will be achieved, and periodic reviews to evaluate whether the required reductions or the expected results are achieved in reality (EPA, 2013). For now, the MEP of China has established 946 air monitoring stations covering 190 cities across China that monitor real-time air pollution data. This information is communicated to the public through China’s Air Quality Index (AQI) as hourly concentrations of each pollutant.

Forecasting has shown great significance on the managing and controlling system in other areas (Wang et al., 2009, 2010, 2014). However, the related work concerning computer modeling for predicting future PM trends is still poor in China, though very significant. First, if a high PM concentration level occurs, residents should be given as much notice as possible in advance, and the ability to warn the public can allow the initiation of regulatory actions to prevent pollution (Qin et al., 2014). Thereby, the related ministry should issue a short-term and long-term outlook or a smog prediction for several hours or days. Secondly, predicting the future trends in PM concentration is an effective way to provide essential information for improving air quality. Based on such forecasts, it is easy to identify whether an area meets the PM standards and, if not, the amount of reductions needed to meet those standards can be quantified. Therefore, to achieve the demanded reductions, some stricter and more enforceable measures to reduce emissions sources can be proposed and adopted by the environmental agencies that are responsible for the implementation plan (EPA, 2013). According to the aforementioned analysis, an effective dynamic forecasting model is urgently needed to perform predictions on future PM trends.

The major methods for PM prediction include chemical transport models (CTMs) and empirical statistical models which mainly focus on the sources and transport of chemical species. Different chemical mechanisms, chemical kinetic expressions, reaction rate coefficients, chemical species and gas phase reactions are usually incorporated into very complex models (Sun et al., 2013). The accuracy of CTMs is sensitive to the scale and quality of the emissions data used (Han et al., 2008), largely stemming from the incomplete knowledge on the sources, dispersion of PM, transport processes and atmospheric chemicals (Cobourn, 2010). The PM concentration prediction models presented in the literature are mostly statistical models that include autoregressive integrated moving average (ARIMA), multi-linear regression (MLR), artificial neural networks (ANNs), hidden Markov model, or hybrid models (Perez and Salini, 2008; Qin et al., 2014).

For linear models, ARIMA and MLR have been adopted to forecast air quality but with variable accuracy mainly owing to their linear mapping ability in non-linear processes (Stadlober et al., 2008; Akyüz and Çağabul, 2009). Pérez et al. (2000) used an ANN model to predict 1 h average PM2.5 concentrations over 1 h—24 h outlooks, where errors range from 30% to 60%. Sun et al. (Sun et al., 2013) proposed a hidden Markov model with different emissions distributions to predict 24 h-average PM2.5 concentrations in North California. A hybrid clustering algorithm (HCA) was used for forecasting PM10, and the results showed a 10%
improvement over ANN models (Vlachogiannis and Sfetsos, 2006). Additionally, Qin et al. (2014) analyzed and predicted PM concentration levels over four major cities in China using hybrid models, improving the forecasting results by 24%, 16%, 16% and 13% for different strategies.

According to the above literature, studies on PM predictions are primarily driven by individual or hybrid models and focused on generating deterministic forecasts, whereas none of them address the uncertainty of PM forecasts properly, which is critical for forecasting and early warning systems. Kani and Riahy (2008) noted that the purpose of most approaches is to make a deterministic forecast, but knowledge about uncertainty is not directly provided. From a practical view, an inherent and irreducible uncertainty exists in every forecast regardless of forecasting model type, rendering decision-making problematic or even prone to mistakes (El-Fouly et al., 2006). As for the decision-maker, overestimation and underestimation type, rendering decision-making problematic or even prone to mistakes (El-Fouly et al., 2006). As for the decision-maker, over- and gamma, are developed to forecast PM levels and evaluate the concentration levels over four major cities in China using hybrid forecasting model, and the evaluation metrics.

The remainder of this paper is organized as follows: section 2 presents the related methodology; section 3 describes the data sources and study areas; and a case study and analysis are given in section 4. In the next section, the conclusions are presented.

### 2. Methodology

In this section, the related approaches used in this study are introduced briefly, including PM emission distribution functions, the adaptive neuro-fuzzy (ANF) model, the proposed PM interval forecasting model, and the evaluation metrics.

#### 2.1. Distribution functions

The probabilistic distribution was adopted to estimate the PM emissions characteristics and to perform interval forecasts further. For that reason, we adopted several distribution functions, i.e., Weibull, Rayleigh, Lognormal and Gamma, to represent the statistical characteristics of the PM. In particular, Lognormal and Gamma distributions were used by Sun et al. (Sun et al., 2013) to model PM$_{2.5}$ time series. Additionally, the Lognormal function was found to be the best distribution for representing the observed PM$_{2.5}$. Table 1 gives the probability density function (PDF) and the cumulative distribution function (CDF) of the distributions mentioned above.

#### 2.2. Adaptive neuro-fuzzy (ANF) model

Neuro-fuzzy (NF) is a new terminology stemming from neural networks (NNs) engaged with fuzzy logic (FL), which enables a system to manage cognitive uncertainties in a manner more similar to humans. While FL performs an inference mechanism under cognitive uncertainty, computational NNs offer exciting advantages, such as learning, adaptation, fault tolerance, parallelism and generalization (Jang and Sun, 1996). The ANF model was created by combining the adaptive capacities of the neural networks with the fuzzy qualitative method to facilitate learning and adaptation. This fuzzy model is the realization of the functionality of fuzzy systems using neural techniques that can be trained with no need for expert knowledge to achieve the desired results (Keles et al., 2008). A basic ANF structure is shown in Fig. 1.

For the ANF model, two main advantages are gained. The first is the ability to model the characteristics of a given problem using a high-level linguistic model instead of low-level complex mathematical expressions. The second is that the embedded fuzzy system in a neural fuzzy network can self-adjust the parameters of the fuzzy rules using neural network learning algorithms to achieve the desired results (Keles et al., 2008).

<table>
<thead>
<tr>
<th>Distribution</th>
<th>PDF/CDF</th>
<th>Parameters</th>
</tr>
</thead>
</table>
| Weibull       | $\begin{align*}
  f(x; \lambda, k) &= \frac{2}{\lambda} \left( \frac{x}{k} \right)^{k-1} \exp\left( -\left( \frac{x}{k} \right) \right), \ x \geq 0 \\
  F(x; \lambda, k) &= 1 - \exp\left( -\left( \frac{x}{k} \right) \right)
\end{align*}$ | $k$=0 shape parameter, $\lambda$=0 scale parameter |
| Rayleigh      | $\begin{align*}
  f(x; \sigma) &= \frac{1}{\sigma \sqrt{2\pi}} \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right), \ x \geq 0 \\
  F(x; \sigma) &= 1 - \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right)
\end{align*}$ | $\sigma$=0 scale parameter |
| Lognormal     | $\begin{align*}
  f(x; \mu, \sigma) &= \frac{1}{\sigma \sqrt{2\pi}} \exp\left( -\frac{(x-\mu)^2}{2\sigma^2} \right), \ x > 0 \\
  F(x; \mu, \sigma) &= \frac{1}{2} \left( \operatorname{erf}\left( \frac{x-\mu}{\sqrt{2}\sigma} \right) \right)
\end{align*}$ | $\mu$=0 location parameter, $\sigma$=0 scale parameter |
| Gamma         | $\begin{align*}
  f(x; k, \theta) &= x^{k-1} \exp\left( -\frac{1}{2} \theta^2 \right), \ x > 0 \\
  F(x; k, \theta) &= \frac{1}{\theta^k} \int_0^x t^{k-1} \exp\left( -\frac{1}{2} \frac{t}{\theta^2} \right) dt, \ x > 0
\end{align*}$ | $k$=0 shape parameter, $\theta$=0 scale parameter |
2.3. Proposed PM interval forecasting method

Information on uncertainties is expressed by interval forecasts (Qin et al., 2015). Given a significance level of $2\alpha$, the relationship among the upper limit, lower limit and real value can be denoted by Eq. (1):

$$P(L_{\text{lower}} \leq Y_t \leq L_{\text{upper}}) = 1 - 2\alpha$$

A dynamic interval forecast method is proposed to give the uncertain information of future PM values by updating the expectation of the following point with the forecasting value. For example, the predicted value for the next point is $\hat{Y}$, and $\hat{Y}$ nearly reaches the maximum level of historical data. If the forecasting result is reliable and precise, then the lower and upper limits of the interval for the next point are both very large values.

In this paper, the PM values are random variables, and estimations are deemed as the expectations for future points. Eq. (1) can be written into Eq. (2):

$$P(L_{\text{lower}} \leq Y_t \leq L_{\text{upper}}) = P(L_{\text{lower}} \leq Y_t \leq L_{\text{upper}} | E(Y_t) = \hat{Y}) \times P(E(Y_t) = \hat{Y})$$

Meanwhile, we assume that the predicted value has a similar form with the historical distribution $f$ and uses the historical variance $S^2$ as the variance of the unknown points. Then, the conditional probability is equal to Eq. (3):

$$P(L_{\text{lower}} \leq Y_t \leq L_{\text{upper}} | E(Y_t) = \hat{Y}) = \int_{L_{\text{lower}}}^{\hat{Y}} f(x|\hat{Y})dx + \int_{\hat{Y}}^{L_{\text{upper}}} f(x|\hat{Y})dx$$

$$\int_{x} (x - E(x))^2 f(x|\hat{Y})dx = S^2$$

$$\int_{x} f(x|\hat{Y})dx = \hat{Y}$$

(3)

Given a specified confidence level, pairwise lower and upper limits can be computed by Eq. (4). Through transformation, the interval can be deemed as an extension of the predicted value and, $\alpha$ signifies the degree of amplification or reduction.

$$\left\{ \begin{array}{l}
L_{\text{lower}} \leq Y_t \\
L_{\text{upper}} \leq \int_{L_{\text{lower}}} f(x|\hat{Y})dx \\
\int_{\hat{Y}}^{L_{\text{upper}}} f(x|\hat{Y})dx = 1 - 2\alpha
\end{array} \right.$$ (4)

Given the many coupled $(L_{\text{lower}}, L_{\text{upper}})$ satisfying Eq.(4), this paper uses the pair with symmetrical probabilities, namely:

$$\left\{ \begin{array}{l}
L_{\text{lower}} \leq Y_t \\
\int_{L_{\text{lower}}} f(x|\hat{Y})dx = F(\hat{Y}) - \alpha,
\int_{\hat{Y}}^{L_{\text{upper}}} f(x|\hat{Y})dx = (1 - F(\hat{Y})) - \alpha
\end{array} \right.$$ (5)

2.4. Evaluation metrics

For estimating the PM emissions distribution, the coefficient of determination $R^2$ is adopted to evaluate the fitness performance of each of the characteristics to the recorded data. $R^2$ is expressed as follows (Wang et al., 2015; Carta and Velázquez, 2011):

$$R^2 = \frac{\sum_{i=1}^{n} (\hat{F}_i - F)^2}{\sum_{i=1}^{n} (F_i - \bar{F})^2 + \sum_{i=1}^{n} (\hat{F}_i - \bar{F})^2}$$ (6)

where $F_i$ and $\hat{F}$ denote the observed and estimated cumulative probability, respectively. This metric evaluates the correlation between the recorded and the estimated cumulative probability.

Two metrics that are used to evaluate the effectiveness of deterministic forecasting are the mean squared error (MSE) and the mean absolute percent error (MAPE), which are defined as:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2$$ (7)

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|y_t - \hat{y}_t|}{y_t} \times 100\%$$ (8)

where $y_t$ and $\hat{y}_t$ are the actual recorded and estimated PM values at time $t$ ($t = 1, 2, \ldots, T$ prediction horizon), respectively.

Another two metrics that are used to measure the interval forecasting performance are the coverage probability (CP) and average width (AW), which are defined as:

$$\text{CP} = \frac{1}{T} \sum_{t=1}^{T} c_t$$ (9)

$$\text{AW} = \frac{1}{T} \sum_{t=1}^{T} (U_t - L_t)$$ (10)
where \( c_t = 1 \) if the target value \( y_t \in [L_t, U_t] \); otherwise, \( c_t = 0 \). \( L_t \) and \( U_t \) are the lower and upper bounds of the \( t \)th interval forecast, separately.

3. Data sources and study areas

The Ministry of Environmental Protection of China posts air quality data on a website (http://113.108.142.147:20035/emcpublish) hourly on major pollutants including ozone (O3), particulate matters (PM10 and PM2.5), sulfur dioxide (SO2), nitrogen dioxide (NO2) and carbon monoxide (CO). The new ambient air quality standards expanded focus on PM2.5 concentration. The referenced AQI can provide the public with health guidance and identify whether an area meets the air quality standard. In this study, hourly AQI data from December 2013 to November 2014 were collected from the three core cities in the Yangtze River Delta (YRD) region of China. Fig. 2 presents the specific locations of the study cities: Shanghai, Hangzhou, and Nanjing. The PM levels are also marked in this figure.

The YRD region is located in the eastern China, generally comprising the triangle-shaped territory of Shanghai, Jiangsu and Zhejiang province. The delta is one of the most densely populated zones in the world, includes some of the fastest-growing economies, and covers the financial center of China. This delta enjoys a marine monsoon subtropical climate with warm springs and falls, cool and dry winters, and hot and humid summers. Li et al. (2011) found that air quality in winter in the YRD region was generally worse than in summer, mainly stemming from unfavorable meteorological dispersion conditions. Furthermore, the pollution caused by local emissions was reinforced by the pollution transport from North China to the delta.

4. Case study and analysis

In this section, we first estimated the distributions of PM emissions with parameters optimized by the cuckoo search optimization (CSO) algorithm. Following use of the fitted PM emission distribution and the ANF model, PM interval forecasts were performed over the subsequent 1 h- and 12 h-periods.

4.1. Different distributions used for PM emissions estimation

Various distribution functions have been adopted to represent the PM regimes at a specific study area. Sun et al. (2013) used Gaussian, lognormal, gamma and GEV distribution functions to mimic real PM2.5 data. The most appropriate PDFs could be identified by the goodness-of-fit measure: determination coefficient \( R^2 \). A large \( R^2 \) value indicates a better mimic of the fitted cumulative probability to the empirical cumulative probability (Wang et al., 2015).

In this study, the PM emissions distribution is constructed using the four PDFs of the Weibull, Rayleigh, Lognormal and Gamma
distributions. In the literature, the parameters of the above mentioned distributions are usually estimated by minimum least square (MLS) or maximum likelihood estimation (MLE) methods (Wang et al., 2015; Wu et al., 2013). Herein, those parameters are initially estimated by the MLE method and are further optimized by an artificial intelligent optimization (AIO) algorithm to maximize the objective function value \( R^2 \). This optimization approach was successfully adopted by Wang et al. (2015) and Wu et al. (2013) in wind energy and extreme wind speed estimation. A recently developed AIO algorithm, the cuckoo search optimization (CSO), was used, and the optimal parameters were confirmed through \( R^2 \) comparison. Table 2 presents the estimated parameters using the MLE and CSO methods. With reference to the estimation of PM density functions, Figs. 3–4 show the \( R^2 \) values obtained using the MLE and CSO methods and the models used in the estimation of the PM emission distributions. Given that the higher values of \( R^2 \) occur in the fit of the PM distributions to the recorded data, the lower values are the errors which are not reasonable for quantile estimation at a further step. It can likewise be observed that in the case of \( R^2 \) for the best fits representing the hourly or 12 h-averaged PM data, the PDF estimated using the CSO algorithm provided better estimation in all the cases analyzed.

In Shanghai (see Fig. 5), regardless of the time interval, the maximum frequency of PM10 and PM2.5 values are located in the interval [37.5, 52.5] \( \mu g/m^3 \) and [22.5, 37.5] \( \mu g/m^3 \), respectively. As in Hangzhou, the most frequent PM10 value is located in the range [52.5, 67.5] \( \mu g/m^3 \) for the hourly basis (see Fig. 6a) and [67.5, 82.5] \( \mu g/m^3 \) for the 12 h-averaged PM (see Fig. 6b). The most frequent PM2.5 value for both time intervals ranges from 37.5 to 52.5 \( \mu g/m^3 \) (see Fig. 7c–d). Meanwhile, in Nanjing, the maximum frequency for hourly and 12 h-averaged PM10 is located in the interval [97.5, 112.5] \( \mu g/m^3 \) (see Fig. 7a) and [82.5, 97.5] \( \mu g/m^3 \) (see Fig. 7b), respectively. The most frequent PM2.5 value is in the range of 37.5–52.5 \( \mu g/m^3 \) (see Fig. 7c–d). According to the most frequent PM values, it can be concluded that Nanjing suffered from the highest levels of PM pollution, followed by Hangzhou. The frequency of extreme PM values in the form of long thin tails shown in each subplot are the smallest, but they occurred, which implies that some extreme weather such as haze occurred in all of the cities during the study period.

From Figs. 5–7, it can also be observed that the distributions used in this study fitted well with the recorded PM values in a notably high percentage of the cases analyzed. However, an analysis of the hourly and 12 h-averaged PM data recorded in the three different cities showed that the distribution which best mimics the data was different. The best fits in the majority of cases came from the Lognormal distribution. It should be mentioned that in some cases the best fits came from the Gamma distribution (see Fig. 6c, Fig. 7c–d) with the corresponding \( R^2 \) slightly larger than that of the other distributions.

The intention of PM emissions distribution fits is to assist in the further step of constructing PM interval forecasts.

### 4.2. PM dynamic interval forecasts

In this paper, the values of PM2.5 and PM10 are considered to be random variables, and the 1 h-interval data from October to November in 2014 are the target points. This section will give detailed explanations of the results regarding the PM deterministic and interval forecasts. In the part on forecasting values, predicted accuracy is discussed the most while the interval width and the coverage probability that interval covers the actual recorded values are mainly discussed in the part on intervals. For each output of a specified dependent variable such as PM2.5, the independent variables are the values of CO, NO2, O3, and PM2.5 for the past two moments.

### Table 2

<table>
<thead>
<tr>
<th>PM</th>
<th>Data type</th>
<th>Methods</th>
<th>Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Weibull</td>
</tr>
<tr>
<td>Shanghai</td>
<td>Hourly</td>
<td>MLE</td>
<td>185.5304</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>79.9225</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>85.3584</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>77.7896</td>
</tr>
<tr>
<td>PM2.5</td>
<td>Hourly</td>
<td>MLE</td>
<td>63.7827</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>58.9444</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>64.5654</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>58.3348</td>
</tr>
<tr>
<td>Hangzhou</td>
<td>Hourly</td>
<td>MLE</td>
<td>113.1978</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>107.4755</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>113.608</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>106.623</td>
</tr>
<tr>
<td>PM2.5</td>
<td>Hourly</td>
<td>MLE</td>
<td>73.9765</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>67.9989</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>74.3546</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>66.8092</td>
</tr>
<tr>
<td>Nanjing</td>
<td>Hourly</td>
<td>MLE</td>
<td>148.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>140.7347</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>148.8114</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>139.5600</td>
</tr>
<tr>
<td>PM2.5</td>
<td>Hourly</td>
<td>MLE</td>
<td>90.6706</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>84.5970</td>
</tr>
<tr>
<td></td>
<td>12 h-averaged</td>
<td>MLE</td>
<td>91.6255</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CS</td>
<td>83.9167</td>
</tr>
</tbody>
</table>
4.2.1. The deterministic forecasts of PM2.5 and PM10

To provide forecasting intervals for PMs, the individual prediction is estimated first. Table 3 gives the MSE and MAPE of PM2.5 or PM10 for October, November and the two-month period (denoted by Total in Table 3). Comparing the total MSE values of PM2.5 between cities, Shanghai (noted by SH) has the smallest total MSE of 59.39, and Nanjing (NJ) has the largest value of 92.28. However, when considering the total MAPE criteria, Shanghai, which should have the smallest MAPE as its best performance on MSE, has the largest MAPE. Meanwhile, the total MAPE of Hangzhou is the best with 8.4% among these three cities although Hangzhou’s total MSE of 61.31 is not the smallest. This could happen because in computing the mean PM2.5 values from October to November of these cities, the results are 67.99 for Hangzhou, 73.72 for Nanjing and 46.76 for Shanghai. The good performance of MSE in Shanghai is counteracted by a small dividend through the calculation of percentage errors. Therefore, multi-indicators should be used to judge the forecasting effect comprehensively and reliably. Due to a strongly positive correlation between PM2.5 and PM10, PM10 has similar results with PM2.5. The average PM10 level from October to November of the cities are 103.98 (Hangzhou), 125.50 (Nanjing) and 71.47 (Shanghai). Still, Shanghai has the smallest MSE of 136 and the largest MAPE of 11.55%, and the MSE of Nanjing, which is 184.27, is bigger than the other two cities. In contrast with the results of PM2.5, Nanjing rather than Hangzhou had the smallest MAPE of 8.43%.

Table 3 also provides the results for each month of October and November. Generally, the values in the October column are better than in the November column regardless of whether PM2.5 or PM10 and MSE or MAPE are considered, which indicates that PM values may have been more regular in October 2014.

To compare the results of October and November in detail, the boxplot of absolute error of PM2.5 values for three cities are given in Fig. 8. Due to the impact of extreme points on observation (large values of extreme points makes box parts too short to distinguish from each other), the outliers are separated and placed in the

---

![Fig. 3.](image1) $R^2$ metrics of the distributions used in estimation of the hourly PM characteristics of different cities in China. I: Weibull; II: Rayleigh; III: Lognormal; and IV: Gamma.

![Fig. 4.](image2) $R^2$ metrics of the distributions used in estimation of the 12 h-averaged PM characteristics of different cities in China. I: Weibull; II: Rayleigh; III: Lognormal; and IV: Gamma.
It can be clearly seen in Fig. 8 that the medians of these two months are similar while outliers vary greatly. For example, the median absolute error is 3.5 for October and 4.3 for November in Nanjing, but the smallest mild outlier of November is even bigger than the maximum one for October. Therefore, when using simple average of points as the criterion, the MSE for November will be magnified substantially by these outliers although 50% results for this month are similar to those for October.

The trend of a predominance in October is strongly shown in Nanjing and Shanghai, so we computed the mean real values of PM2.5 at the outliers and normal points. The averaged PM2.5 values for normal absolute error and outliers are 75.15 and 138.43 in Nanjing and 47.03 and 104.32 in Shanghai, respectively. Fig. 9 shows the absolute error of PM10 which has similar results to PM2.5. Therefore, it can be concluded that the point with a larger value of particulate matter may cause a decline in forecasting accuracy.

The dividend in MAPE’s computation strongly affects its value; thus, the MAPE given in Table 4 are classified into 6 ranks based on real values. The results of the classified MAPE are stochastic in finding a recapitulative regulation, which means the forecasting algorithm retrieves the information relatively completely. Therefore, the predictions are effective and reliable for both a result and a reference standard for PM forecasting.

Sometimes the predicted value of PM2.5 or PM10 over a period of time is needed to estimate the average air quality during this period. In this part, a half-day (12 h-) prediction is computed after the hourly data are divided into disjoint intervals with 12 points (or less than 12 if there are missing values), and the mean of these 12 points is the interval value. Through averaging every 12 points, the MSE and MAPE were augmented in multiples. Fig. 10 shows the relationships between the real values and predictions of PM2.5 and PM10 for three cities. To
Table 3
The average MSE and MAPE for PM$_{2.5}$ or PM$_{10}$ for October and November.

<table>
<thead>
<tr>
<th></th>
<th>PM$_{2.5}$ Oct</th>
<th>Nov</th>
<th>Total</th>
<th>PM$_{10}$ Oct</th>
<th>Nov</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>HZ 50.17</td>
<td>72.81</td>
<td>61.31</td>
<td>HZ 155.57</td>
<td>205.04</td>
<td>179.91</td>
</tr>
<tr>
<td></td>
<td>NJ 66.94</td>
<td>119.08</td>
<td>92.28</td>
<td>NJ 167.99</td>
<td>201.50</td>
<td>184.27</td>
</tr>
<tr>
<td></td>
<td>SH 30.81</td>
<td>88.23</td>
<td>59.39</td>
<td>SH 103.89</td>
<td>168.40</td>
<td>136.00</td>
</tr>
<tr>
<td>MAPE (%)</td>
<td>HZ 8.74</td>
<td>8.05</td>
<td>8.40</td>
<td>HZ 9.30</td>
<td>11.68</td>
<td>10.47</td>
</tr>
<tr>
<td></td>
<td>NJ 8.48</td>
<td>10.17</td>
<td>9.30</td>
<td>NJ 7.62</td>
<td>9.29</td>
<td>8.43</td>
</tr>
<tr>
<td></td>
<td>SH 9.00</td>
<td>10.40</td>
<td>9.69</td>
<td>SH 10.15</td>
<td>12.96</td>
<td>11.55</td>
</tr>
</tbody>
</table>

Fig. 7. Frequency histogram and the fitted distributions of the PM values (Nanjing).

Fig. 8. Boxplots of absolute error of PM$_{2.5}$ values for three cities.
make the comparisons clearly, the diagonal is drawn by a black dotted line. Obviously, better forecasting points have less distance from the diagonal. It can be seen in Fig. 10 that the prediction of PM\textsubscript{2.5} precedes the results of PM\textsubscript{10} because the points in the PM\textsubscript{10} subplot present a dispersive layout, and this predominance is also shown in the table where the MSE of PM\textsubscript{2.5} is much smaller than that of PM\textsubscript{10}. In addition, the points of Shanghai are closer to zero or to the diagonal than the other two cities, indicating that the average PM values of Shanghai as well as its forecasting error are smaller. From Fig. 10, the differences between the real values and predictions for Nanjing are generally larger than the other places, and the MSEs for Nanjing are 978.63 for PM\textsubscript{2.5} and 1883.40 for PM\textsubscript{10}. Therefore, it can be concluded that the absolute error is positively increased with the real value, which is a similar result in 12 h forecasting. For MAPE, a similar result with random 1 h data is shown in 12 h forecasting. The smallest MAPE of PM\textsubscript{10} was found in Shanghai, which also has the best MSE of PM\textsubscript{10}. Although Hangzhou does not have the best MSE of PM\textsubscript{2.5}, its relative error of PM\textsubscript{2.5} is the smallest among these three cities.

4.2.2. The PM dynamic interval forecasts

Interval width and coverage probability are influenced by the level of significance, and the optimal forecasting interval depends on the reality. Therefore, this section gives the interval results for different significance levels to make comparisons among each other. To express the results clearly, the interval width and cover probability are replaced by the notations AW and CP.

In this paper, future PM values are considered as a random
variable with a different distribution for each point in time. Given a specified distribution to a target point, the interval of this point can be limited by two symmetrical quantiles, and the interval is non-unique because we can use different quantiles. Obviously, the interval will be smaller as the value of significance level (\(\alpha\)) increases. A smaller \(\alpha\) indicates that the real value is more likely to be located in this interval. However, a large interval implies a weak capacity to accurately position the future values of particulate matter. It is absolutely correct that the future values of \(PM_{2.5}\) must be in the interval from zero to infinity, but this confirmation is meaningless for application.

As mentioned in section 4.1, the majority of distributions for PM values are Lognormal, which signifies that the logarithm of the variable is normally distributed. For quick computation, the interval limits are estimated with normal distribution followed by exponential function. Table 5 shows the interval \(AW\) and \(CP\) of 1 h and 12 h predictions for different significance levels. To obtain reliable conclusions, each point is the average difference between the upper and lower limits. For example, when \(\alpha=0.3\), the coverage probability is 1033/1053, and the interval width is the average of all 1053 differences between upper and lower limits.

Because the distribution is based on forecasting values, the precision of point prediction strongly affects the accuracy of the forecasting intervals. This can be shown on the larger \(CP\) of 1 h values than the 12 h results. For example, given the significance level of 0.35, the average \(CP\)s of the hourly results are 95.16% for \(PM_{2.5}\) and 92.99% for \(PM_{10}\); for 12 h, the average \(CP\)s are 45.94% for \(PM_{2.5}\) and 48.74% for \(PM_{10}\). In addition, comparing the \(AW\) between 1 h and 12 h results, the 12 h \(AW\) of a given \(CP\) level is wider than the hourly width with a similar \(CP\) level. For example, given \(\alpha=0.45\) and the \(CP\) of 60.50% when \(\alpha=0.3\). However, the corresponding \(AW\) is 35.8, which is triple the value of 11. Therefore, the forecasting accuracy of points has influenced the effectiveness of interval for two parts, the probability of covering the real values and the width of predicted intervals.

Unlike forecasting for the values of target points, there are many forecasting intervals with different confidence levels. An interval with a high \(CP\) is not always the best interval in practice. A smaller \(\alpha\) is preferred when \(CP\) is mainly considered, and a larger \(\alpha\) will be chosen when a smaller \(AW\) is required. Meanwhile, the forecasting precision should be determined before interval forecasting, which can be realized by using historical information or testing a few points; the level of \(\alpha\) should be decreased if the predicted values are

![Fig. 11. Results of interval forecasts.](image-url)
inaccurate.

Combining the information from Table 5, we chose a significance level of 0.35 for 1 h PM and 0.25 for 12 h PM, and the results of the interval forecasts are displayed in Fig. 11. There are twelve subplots in that figure, and three types of basic information are given by numbers or letters. The number at the upper left corner is the subplot shows the interval predictions of hourly PM$_{2.5}$ in Hangzhou with a maximum value of 200. In addition, the purple (in the web version) points are for pairwise real and predicted values, and yellow (in the web version) shadows are drawn for the intervals with edges of orange (in the web version) points (pairwise real values and upper or lower limits). Furthermore, the forecasting results are sorted by real values before plotting. It can be seen from Fig. 11 that the intervals are universally wider when the real values are large. This is normal because large points are hard to predict. It can be observed that the majority of real values are located inside the forecasting interval, and the effects are predicted remarkably well.

5. Conclusions

From the perspective of epidemiological and exposure estimation, the future trend in PM in a city is of high importance for forecasting and early warning systems. With precise predictions, some regulatory actions could be initiated in advance to reduce the degree of negative effects posed by PM, such as reducing exposure to high levels of PM through limiting outside activity, staying indoors or wearing a mask. Furthermore, obtaining informative future PM levels would be an effective way to improve air quality. Specifically, based on that information, we can assess whether a city meets the air quality standards and how much reduction should be retained. Towards this end, some temporary but mandatory forces could be implemented to address that potential PM condition.

In this paper, based on hourly and 12 h-averaged air pollutant data, an adaptive neuro-fuzzy (ANF) model coupled with emissions distributions was used to perform deterministic and interval forecasts for PM within the three core cities of the YRD region of China, including Shanghai, Hangzhou and Nanjing. The main goal was to assess PM trends in a future period. We firstly examined and analyzed PM characteristics using distribution functions. It was found that the Lognormal or Gamma distributions fitted the recorded PM data best with a goodness-of-fit $R^2$ of approximately 0.998. Those different distributions mainly resulted from the different geographical conditions, meteorological conditions or emission sources. Afterward, an ANF model was applied for PM deterministic forecasts, which exhibited a good forecasting performance. Coupling the ANF model and the identified PM emission distributions, dynamic PM interval forecasts were achieved to handle the uncertainties in future PM values. All of the applicable case studies conducted show that the estimated intervals could cover a large majority of the recorded data, offering a great advantage for constructing and implementing forecasting and early warning systems.

Acknowledgments

The work was supported by the National Natural Science Foundation of China (Grant No. 71171102).

References


Colbourne, W.G., 2010. Air quality PM$_{2.5}$ air quality forecast model based on nonlinear regression and back-trajectory concentrations. Atmos. Environ. 44 (25), 3015–3023.


