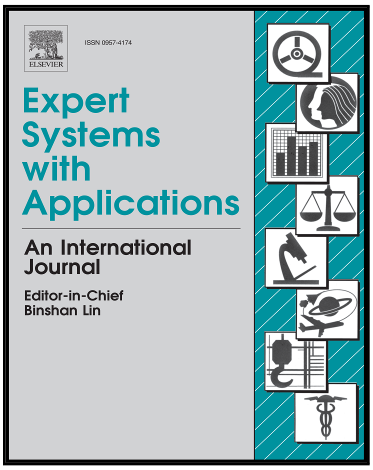

# Developing an early-warning system for air quality prediction and assessment of cities in China

Jianzhou Wang , Xiaobo Zhang , Zhenhai Guo , Haiyan Lu

PII: S0957-4174(17)30319-6
DOI: 10.1016/j.eswa.2017.04.059
Reference: ESWA 11295

To appear in: *Expert Systems With Applications*

Received date: 19 January 2017
Revised date: 24 April 2017
Accepted date: 25 April 2017

Please cite this article as: Jianzhou Wang , Xiaobo Zhang , Zhenhai Guo , Haiyan Lu , Developing an early-warning system for air quality prediction and assessment of cities in China, *Expert Systems With Applications* (2017), doi: 10.1016/j.eswa.2017.04.059

**Highlights**

- An early-warning system is developed for air quality.
- Pollutant emission characteristics are analyzed using distribution functions.
- Dynamic forecast intervals are constructed for addressing the uncertainty.
- Air quality is evaluated by integrating fuzzy set theory and AHP.
- The results show that the developed early-system is effective and reliable.

-

# Developing an early-warning system for air quality prediction and assessment of cities in China

Jianzhou Wang[a], Xiaobo Zhang[a,]*, Zhenhai Guo[b], Haiyan Lu[c]

[a] *School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*

[b] *State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China*

[c] *School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia*

** Corresponding author. Address: School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China*

*E-mail address: zxb100498@163.com (X. Zhang)*

Author #1:

Name: Jianzhou Wang

Affiliation: Dongbei University of Finance and Economics, Dalian, China

Email: wangjz@dufe.edu.cn

Author #2:

Name: Xiaobo Zhang

Affiliation: Dongbei University of Finance and Economics, Dalian, China

Email: zxb100498@163.com

Author #3:

Name: Zhenhai Guo

Affiliation: State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, China

Email: gzh@lasg.iap.ac.cn

Author #4:

Name: Haiyan Lu

Affiliation: School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia

Email: haiyan.lu@uts.edu.au

Dear reviewer,

Thank you very much for the positive and constructive comments on our manuscript entitled "*Developing an early-warning system for air quality prediction and assessment: A case study in China*". (No.: ESWA-D-17-00252R1). We have studied the comments and tried our best to revise the manuscript. Here below is our response to your comments and the revised parts in the revised manuscript are marked by highlight color,

Reviewer #1: Revise

It is important for the author(s) to put more efforts in rationalizing and generalizing the case study (an early warning system in China). There are a lot of results and discussion on the performance of the developed system. But it appears that there is also a need to recognize and describe the "boundary" between the "phenomenon" that the case study attempt to capture and the context (in which the developed model is to be applied). See many litertures on the research that is based on a case study. A lack of the discussion on the developed system's generalization is the remaining weakness of the manuscript.

Response: Thank you very much for your comment and it is positive and constructive for improving the quality of our paper. We agree that the case study research method is a very useful empirical inquiry that investigates a contemporary phenomenon within its real-life context; however, based on new techniques of data analytics, this paper develop an early warning system for air pollution, which tends to be a verification study. Therefore, in revised manuscript, we have changed the title and treat a case as verification of the new system. Besides, according to your comments, the further discussion on the developed system was also added in the revised manuscript. Please see lines 715-737.

Moreover, another two cities: Wuhan and Nanjing in China are randomly selected as illustrative examples to further verify the effectiveness of the developed early warning system. To facilitate the analysis of results, the sampling design in the original manuscript is maintained in the two cities. In addition, in the assessment module, the data from October 21, 2015 to October 30, 2015 are also selected as an illustrative case. Meanwhile, the assessment results at 8:00 AM daily for Nanjing and 12:00 PM for Wuhan are selected to clearly show the results. Table 1, Table 2, Table 3, Table 4 and Table 5 show the obtained results using the developed early-warning system. From these tables, it can be found that the developed system is effective and reliable for air pollution monitoring and management. In addition, the estimated distributions show significant difference on the performance evaluation indices. This is not only because of different pollutants, but also geographic locations reflecting different processes of industrialization and urbanization.

**Table 1**

$R^2$ and RMSE values of the distributions examined.

| Environmental parameter | Moving average window size | Method | Evaluation metrics | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Lognormal | | Rayleigh | | Gamma | | Weibull | |
| **Wuhan** | | | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** |
| $PM_{2.5}$ | 24 h | CSO | 0.995 | 0.005 | 0.935 | 0.017 | 0.981 | 0.009 | 0.947 | 0.016 |
| | | MLE | 0.993 | 0.006 | 0.904 | 0.021 | 0.968 | 0.012 | 0.905 | 0.021 |
| $PM_{10}$ | 24 h | CSO | 0.947 | 0.011 | 0.919 | 0.014 | 0.973 | 0.008 | 0.978 | 0.007 |
| | | MLE | 0.911 | 0.014 | 0.921 | 0.014 | 0.967 | 0.009 | 0.978 | 0.007 |
| $SO_2$ | 24 h | CSO | 0.985 | 0.010 | 0.953 | 0.018 | 0.979 | 0.012 | 0.962 | 0.016 |
| | | MLE | 0.987 | 0.009 | 0.896 | 0.026 | 0.977 | 0.012 | 0.941 | 0.020 |
| $NO_2$ | Hourly | CSO | 0.990 | 0.005 | 0.962 | 0.010 | 0.991 | 0.005 | 0.970 | 0.009 |
| | | MLE | 0.978 | 0.008 | 0.926 | 0.014 | 0.986 | 0.006 | 0.953 | 0.011 |
| $O_3$ | 8 h | CSO | 0.938 | 0.014 | 0.676 | 0.031 | 0.988 | 0.006 | 0.991 | 0.005 |
| | | MLE | 0.876 | 0.020 | 0.597 | 0.035 | 0.962 | 0.011 | 0.974 | 0.009 |
| CO | Hourly | CSO | 0.999 | 0.004 | 0.974 | 0.019 | 0.993 | 0.010 | 0.983 | 0.015 |
| | | MLE | 0.987 | 0.013 | 0.861 | 0.043 | 0.965 | 0.022 | 0.903 | 0.036 |
| **Nanjing** | | | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** | ***$R_2$*** | ***RMSE*** |
| $PM_{2.5}$ | 24 h | CSO | 0.987 | 0.007 | 0.959 | 0.012 | 0.985 | 0.007 | 0.961 | 0.012 |
| | | MLE | 0.983 | 0.008 | 0.933 | 0.015 | 0.981 | 0.008 | 0.936 | 0.015 |
| $PM_{10}$ | 24 h | CSO | 0.952 | 0.011 | 0.893 | 0.017 | 0.962 | 0.010 | 0.949 | 0.012 |
| | | MLE | 0.941 | 0.013 | 0.890 | 0.017 | 0.961 | 0.010 | 0.920 | 0.015 |
| $SO_2$ | 24 h | CSO | 0.943 | 0.013 | 0.924 | 0.015 | 0.975 | 0.009 | 0.977 | 0.008 |
| | | MLE | 0.915 | 0.016 | 0.885 | 0.019 | 0.968 | 0.010 | 0.969 | 0.010 |
| $NO_2$ | Hourly | CSO | 0.977 | 0.009 | 0.995 | 0.004 | 0.996 | 0.004 | 0.995 | 0.004 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MLE | 0.935 | 0.016 | 0.988 | 0.007 | 0.992 | 0.005 | 0.987 | 0.007 |
| $O_3$ | 8 h | CSO | 0.932 | 0.014 | 0.515 | 0.038 | 0.980 | 0.008 | 0.982 | 0.007 |
| | | MLE | 0.870 | 0.020 | 0.440 | 0.041 | 0.977 | 0.008 | 0.980 | 0.008 |
| CO | Hourly | CSO | 0.984 | 0.009 | 0.977 | 0.011 | 0.982 | 0.010 | 0.987 | 0.008 |
| | | MLE | 0.945 | 0.017 | 0.970 | 0.013 | 0.986 | 0.009 | 0.969 | 0.013 |

**Table 2**

Forecast results.

| Model | Time | Environment parameters | | | | | |
|---|---|---|---|---|---|---|---|
| | | $PM_{2.5}$ | $PM_{10}$ | $SO_2$ | $NO_2$ | $O_3$ | CO |

| **Wuhan** | | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ARIMA | Sep | 8.88 | 142.01 | 11.76 | 276.56 | 9.47 | 84.28 | 69.01 | 8550.41 | 39.49 | 5801.47 | 57.04 | 10.29 |
| | Oct | 8.80 | 162.24 | 10.92 | 438.26 | 10.82 | 102.65 | 54.50 | 12399.35 | 45.39 | 7417.37 | 58.09 | 37.58 |
| | Total | 8.84 | 152.36 | 11.33 | 359.26 | 10.16 | 93.68 | 61.59 | 10518.93 | 42.51 | 6627.92 | 57.58 | 24.24 |
| SVM | Sep | 5.30 | 19.74 | 6.16 | 63.35 | 5.21 | 10.14 | 70.07 | 797.05 | 33.83 | 1181.99 | 28.69 | 0.43 |
| | Oct | 5.64 | 43.93 | 7.49 | 141.37 | 6.08 | 7.93 | 37.96 | 1421.31 | 75.61 | 1263.35 | 33.81 | 1.63 |
| | Total | 5.48 | 32.11 | 6.84 | 103.25 | 5.66 | 9.01 | 53.65 | 1116.33 | 55.20 | 1223.60 | 31.31 | 1.05 |
| | Sep | 6.93 | 25.86 | 7.57 | 80.95 | 5.90 | 9.94 | 49.06 | 782.15 | 42.68 | 2522.20 | 33.11 | 0.46 |
| BPNN | Oct | 8.70 | 71.28 | 9.78 | 186.29 | 6.38 | 13.13 | 38.87 | 1347.17 | 56.56 | 2469.47 | 34.03 | 1.54 |
| | Total | 7.84 | 49.09 | 8.70 | 134.82 | 6.15 | 11.57 | 43.85 | 1071.12 | 49.78 | 2495.23 | 33.58 | 1.01 |
| | Sep | 2.03 | 2.84 | 2.58 | 10.93 | 2.46 | 1.71 | 27.63 | 225.12 | 13.58 | 141.42 | 26.86 | 0.32 |
| SSA-BPNN | Oct | 2.57 | 8.80 | 3.11 | 24.68 | 2.88 | 2.08 | 20.65 | 323.45 | 20.52 | 179.60 | 33.41 | 1.09 |
| | Total | 2.31 | 5.88 | 2.85 | 17.96 | 2.67 | 1.90 | 24.06 | 275.41 | 17.13 | 160.95 | 30.21 | 0.71 |
| **Nanjing** | | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE |
| ARIMA | Sep | 9.30 | 37.10 | 8.92 | 115.55 | 19.22 | 8.19 | 64.90 | 4960.85 | 18.19 | 183.97 | 27.24 | 0.22 |
| | Oct | 10.16 | 115.36 | 8.39 | 283.51 | 14.34 | 14.22 | 53.94 | 6774.98 | 24.54 | 220.90 | 29.13 | 0.35 |
| | Total | 9.74 | 77.45 | 8.65 | 202.16 | 16.70 | 11.30 | 59.25 | 5896.26 | 21.46 | 203.01 | 28.21 | 0.29 |
| SVM | Sep | 7.64 | 25.64 | 5.88 | 42.75 | 10.92 | 1.87 | 40.44 | 443.94 | 34.01 | 956.83 | 21.36 | 0.07 |
| | Oct | 6.36 | 34.22 | 5.25 | 82.26 | 12.45 | 13.16 | 32.79 | 699.81 | 41.70 | 641.86 | 24.55 | 0.15 |
| | Total | 6.98 | 30.06 | 5.55 | 63.14 | 11.71 | 7.70 | 36.49 | 575.96 | 37.98 | 794.31 | 23.01 | 0.11 |
| | Sep | 9.23 | 24.53 | 8.18 | 57.09 | 10.77 | 1.91 | 45.22 | 515.52 | 54.63 | 2195.68 | 19.11 | 0.06 |
| BPNN | Oct | 10.51 | 70.83 | 8.36 | 151.41 | 12.98 | 18.35 | 34.37 | 770.72 | 58.09 | 1474.72 | 20.72 | 0.11 |
| | Total | 9.89 | 48.42 | 8.27 | 105.75 | 11.91 | 10.39 | 39.63 | 647.20 | 56.42 | 1823.68 | 19.94 | 0.09 |
| | Sep | 2.97 | 4.29 | 2.25 | 5.74 | 4.37 | 0.31 | 24.79 | 125.28 | 16.93 | 100.30 | 10.06 | 0.02 |
| SSA-BPNN | Oct | 2.36 | 4.89 | 1.75 | 6.71 | 4.35 | 2.01 | 17.64 | 164.05 | 24.30 | 128.98 | 12.27 | 0.03 |
| | Total | 2.66 | 4.60 | 1.99 | 6.24 | 4.36 | 1.19 | 21.10 | 145.28 | 20.73 | 115.10 | 11.20 | 0.03 |

**Table 3**

Interval forecasting results for different significance levels.

| Environ. | Significance | Wuhan | | | | | | Nanjing | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sep. | | Oct. | | Total | | Sep. | | Oct. | | Total | |
| **PM$_{2.5}$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.15 | 100% | 58.75 | 100% | 84.50 | 100% | 71.92 | 100% | 45.22 | 100% | 78.49 | 100% | 62.38 |
| | 0.25 | 100% | 37.16 | 100% | 53.45 | 100% | 45.49 | 99.85% | 28.43 | 100% | 49.35 | 99.93% | 39.23 |
| | 0.35 | 99.85% | 20.93 | 99% | 30.10 | 99.28% | 25.62 | 99.56% | 15.97 | 100% | 27.71 | 99.65% | 22.03 |
| **PM$_{10}$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.15 | 100% | 97.84 | 100% | 140.24 | 100% | 119.52 | 100% | 71.79 | 100% | 118.09 | 100% | 95.68 |
| | 0.25 | 99.56% | 62.12 | 99% | 89.05 | 99.50% | 75.90 | 100.% | 45.60 | 100% | 75.00 | 99.93% | 60.77 |
| | 0.35 | 98.68% | 35.06 | 97% | 50.25 | 97.78% | 42.83 | 99.71% | 25.73 | 100% | 42.33 | 99.72% | 34.30 |
| **SO$_2$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.15 | 100% | 48.72 | 100% | 46.11 | 100% | 47.39 | 100% | 18.03 | 100% | 35.71 | 100% | 27.15 |
| | 0.25 | 100% | 30.31 | 100% | 28.69 | 99.79% | 29.48 | 100% | 11.04 | 100% | 21.87 | 100% | 16.63 |
| | 0.35 | 99.85% | 16.93 | 99.30% | 16.02 | 99.57% | 16.46 | 100.00% | 6.12 | 98.90% | 12.12 | 99.43% | 9.21 |
| | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| **$NO_2$** | 0.15 | 92.39% | 60.95 | 95.52% | 104.82 | 93.99% | 83.39 | 92.82% | 56.70 | 96.84% | 84.19 | 94.89% | 70.89 |
| | 0.25 | 81.41% | 38.00 | 89.93% | 65.34 | 85.77% | 51.98 | 82.55% | 35.62 | 90.92% | 52.89 | 86.87% | 44.53 |
| | 0.35 | 58.57% | 21.24 | 74.13% | 36.53 | 66.52% | 29.06 | 62.46% | 19.99 | 76.07% | 29.69 | 69.48% | 25.00 |
| | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| **$O_3$** | 0.15 | 98.83% | 227.44 | 94% | 159.46 | 96.57% | 192.67 | 98.09% | 210.48 | 92.98% | 138.31 | 95.46% | 173.24 |
| | 0.25 | 97.80% | 132.80 | 91.05% | 93.11 | 94.35% | 112.50 | 94.28% | 119.80 | 88.86% | 78.72 | 91.48% | 98.60 |
| | 0.35 | 93.70% | 71.77 | 84.76% | 50.32 | 89.13% | 60.80 | 89.00% | 63.88 | 81.98% | 41.98 | 85.38% | 52.58 |
| | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| **CO** | 0.15 | 88.87% | 0.95 | 78.46% | 1.37 | 83.55% | 1.17 | 99.27% | 1.04 | 98.21% | 1.23 | 98.72% | 1.14 |
| | 0.25 | 75.26% | 0.60 | 66.15% | 0.87 | 70.60% | 0.74 | 96.63% | 0.66 | 94.91% | 0.78 | 95.74% | 0.72 |
| | 0.35 | 52.71% | 0.34 | 45.31% | 0.49 | 48.93% | 0.42 | 87.39% | 0.37 | 83.63% | 0.44 | 85.45% | 0.41 |

Table 4

Assessment results for Wuhan.

| Time | 21/10/2015 | 22/10/2015 | 23/10/2015 | 24/10/2015 | 25/10/2015 | 26/10/2015 | 27/10/2015 | 28/10/2015 | 29/10/2015 | 30/10/2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual | Ⅲ | Ⅲ | Ⅲ | Ⅲ | Ⅲ | Ⅲ | Ⅰ | Ⅰ | Ⅱ | Ⅱ |
| | | | | | Assessment based on lower bound | | | | | |
| Ⅰ | 0.324 | 0.333 | 0.330 | 0.333 | 0.333 | 0.333 | 0.818 | 0.760 | 0.755 | 0.529 |
| Ⅱ | 0.645 | 0.620 | 0.527 | 0.239 | 0.480 | 0.329 | 0.182 | 0.240 | 0.245 | 0.471 |
| Ⅲ | 0.029 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅳ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅱ | Ⅱ | Ⅱ | Ⅰ | Ⅱ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ |
| | | | | | Assessment based on upper bound | | | | | |
| Ⅰ | 0.080 | 0.210 | 0.251 | 0.211 | 0.188 | 0.291 | 0.269 | 0.332 | 0.323 | 0.315 |
| Ⅱ | 0.254 | 0.123 | 0.082 | 0.122 | 0.145 | 0.042 | 0.672 | 0.588 | 0.603 | 0.243 |
| Ⅲ | 0.473 | 0.437 | 0.297 | 0.071 | 0.267 | 0.151 | 0 | 0.080 | 0.030 | 0.402 |
| Ⅳ | 0.194 | 0.230 | 0.332 | 0.355 | 0.328 | 0.333 | 0 | 0 | 0 | 0.040 |
| Ⅴ | 0 | 0 | 0 | 0.241 | 0 | 0.182 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅲ | Ⅲ | Ⅳ | Ⅳ | Ⅳ | Ⅳ | Ⅱ | Ⅱ | Ⅱ | Ⅲ |
| | | | | | Assessment based on deterministic forecast | | | | | |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Ⅰ | 0.257 | 0.319 | 0.317 | 0.305 | 0.323 | 0.333 | 0.633 | 0.494 | 0.448 | 0.331 |
| Ⅱ | 0.372 | 0.287 | 0.178 | 0.028 | 0.151 | 0.057 | 0.367 | 0.457 | 0.456 | 0.526 |
| Ⅲ | 0.370 | 0.394 | 0.478 | 0.383 | 0.456 | 0.400 | 0 | 0 | 0 | 0 |
| Ⅳ | 0 | 0 | 0.027 | 0.284 | 0.071 | 0.209 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅱ | Ⅲ | Ⅲ | Ⅲ | Ⅲ | Ⅲ | Ⅰ | Ⅰ | Ⅱ | Ⅱ |

Table 5

Assessment results for Nanjing.

| Time | 21/10/2015 | 22/10/2015 | 23/10/2015 | 24/10/2015 | 25/10/2015 | 26/10/2015 | 27/10/2015 | 28/10/2015 | 29/10/2015 | 30/10/2015 |
|---|---|---|---|---|---|---|---|---|---|---|
| Actual | Ⅱ | Ⅱ | Ⅲ | Ⅲ | Ⅲ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ |
| Assessment based on lower bound | | | | | | | | | | |
| Ⅰ | 0.444 | 0.446 | 0.333 | 0.333 | 0.333 | 0.456 | 0.630 | 0.568 | 0.482 | 0.421 |
| Ⅱ | 0.483 | 0.468 | 0.460 | 0.621 | 0.552 | 0.443 | 0.370 | 0.322 | 0.454 | 0.546 |
| Ⅲ | 0 | 0 | 0.055 | 0.005 | 0.035 | 0 | 0 | 0 | 0 | 0 |
| Ⅳ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅱ |
| Assessment based on upper bound | | | | | | | | | | |
| Ⅰ | 0.289 | 0.283 | 0.253 | 0.317 | 0.200 | 0.322 | 0.328 | 0.314 | 0.279 | 0.254 |
| Ⅱ | 0.187 | 0.217 | 0.081 | 0.029 | 0.133 | 0.219 | 0.332 | 0.471 | 0.264 | 0.149 |
| Ⅲ | 0.524 | 0.499 | 0.250 | 0.333 | 0.288 | 0.459 | 0.340 | 0.152 | 0.458 | 0.562 |
| Ⅳ | 0 | 0 | 0.279 | 0.321 | 0.336 | 0 | 0 | 0 | 0 | 0.035 |
| Ⅴ | 0 | 0 | 0.138 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅲ | Ⅲ | Ⅳ | Ⅲ | Ⅳ | Ⅲ | Ⅲ | Ⅱ | Ⅲ | Ⅲ |
| Assessment based on deterministic forecast | | | | | | | | | | |
| Ⅰ | 0.332 | 0.322 | 0.333 | 0.329 | 0.333 | 0.333 | 0.333 | 0.440 | 0.323 | 0.326 |
| Ⅱ | 0.547 | 0.575 | 0.127 | 0.253 | 0.155 | 0.592 | 0.605 | 0.544 | 0.601 | 0.480 |
| Ⅲ | 0.054 | 0.041 | 0.421 | 0.417 | 0.509 | 0.025 | 0.000 | 0.016 | 0.062 | 0.096 |
| Ⅳ | 0 | 0 | 0.118 | 0 | 0.004 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅱ | Ⅱ | Ⅲ | Ⅲ | Ⅲ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ |

If you have any question about this paper, please don't hesitate to let me know.

Yours sincerely,

Xiaobo Zhang

## Abstract

Air quality has received continuous attention from both environmental managers and citizens. Accordingly, early-warning systems for air pollution are very useful tools to avoid negative health effects and develop effective prevention programs. However, developing robust early-warning systems is very challenging, as well as necessary. This paper develops a reliable and effective early-warning system that consists of air quality prediction and assessment modules. In the prediction module, a hybrid forecasting method is developed for predicting pollutant concentrations that effectively estimates future air quality conditions. In developing this proposed model, we suggest the use of a back propagation neural network algorithm, combined with a probabilistic parameter model and data preprocessing techniques, to address the uncertainties involved in future air quality prediction. Meanwhile, a pre-analysis is implemented, primarily by using optimized distribution functions to examine and analyze statistical characteristics and emission behaviors of air pollutants. The second method, which is developed as part of the second module, is based on fuzzy set theory and the Analytic Hierarchy Process, and it performs air quality assessments to provide a clear and intelligible description of air quality conditions. Using data from the Ministry of Environmental Protection of China and six stages of air quality classification levels, specifically *good*, *moderate*, *lightly polluted*, *moderately polluted*, *heavily polluted* and *severely polluted*, two cities in China, Chengdu and Hangzhou, are used as illustrative examples to verify the effectiveness of the developed early-warning system. The results demonstrate that the proposed methods are effective and reliable for use by environmental supervisors in air pollution monitoring and management.

**Key words:** *early-warning system; uncertainty problem*; *interval forecast*; *fuzzy set; weighted air quality assessment*

## *Introduction*

In recent years, air pollution has received increasing attention due to the negative effects, such as respiratory diseases, that it has on human health. Relevant research in China found that the lifespan of the citizens is decreasing because of the poor air conditions. Thus, China's environmental supervisors have issued some plans and programs, including EIA (Environmental Influence Assessment) and Emergency Response for reducing air pollution. For example, the Technical Regulation on Ambient Air Quality Index (AQI) is used in the evaluation of air quality, which is carried out to monitor the atmospheric environment and to provide the public with information on air pollution information. In China, the air quality index (AQI) is widely acknowledged as the metric of air quality that is used to provide the public with information on adverse air pollution conditions.

Over the past few years, many studies of air pollution have mainly focused on forecasting the concentrations of atmospheric pollutants and air quality assessment or evaluation. Compared with assessment, the forecasting of the concentrations of air pollutants, especially particulate matter (PM), appears to be more popular. Alternatively, many methods have been proposed for forecasting atmospheric pollutants using single or hybrid models, such as ARIMA (Samia et al., 2012), multiple linear regression (MLR) (Stadlober et al., 2008; Akyüz and Çabuk, 2009; Genc et al., 2010), artificial neural networks (ANN) (Song et al., 2015; Feng et al., 2015; Pauzi and Abdullah, 2015; Perez, 2016; Biancofiore et al., 2017), fuzzy time series models (Domańska & Wojtylak, 2012)support vector machines (Osowski and Garanty, 2007), gray models (Pai et al., 2013a), hidden Markov models (HMM) (Dong et al., 2009; Sun et al., 2013), Adaptive Neuro-Fuzzy inference systems(ANFIS) (Taylan, 2016) and adaptive Dynamic Programming (DP) (Ariyajunya et al., 2017) methods. Most of these models have obtained the expected results by successfully overcoming the nonlinear and chaotic characteristics of time series of the concentrations of air pollutants. In contrast with forecasting, air quality assessment needs to take into account the number and type of pollutants to the greatest possible degree so as to produce better assessments using proper methods. Moreover, air quality indices, as well as the evaluation of the adverse impacts of pollutants on the health as a whole, largely depends on the concentrations of pollutants, which lie within arbitrary ranges. The assessment of air pollution has recently become an important issue due to its significance, and several new methodologies have been developed for the evaluation of air quality, such as artificial neural networks (Feng et al., 2013; Mishra and Goyal, 2016), Bayesian models (Yong et al., 2008), fuzzy logic (Liu et al., 2009; Sowlat et al., 2011; Yadav et al., 2014; Sen et al. 2015; Xu et al., 2017), and fuzzy logic based on the Analytic Hierarchy Process (AHP) (Upadhyay & Dashore, 2011; Akkaya et al., 2015). Based on the literature on the assessment of air quality mentioned above, fuzzy logic appears to have become increasingly popular because of its ability handle uncertainty and subjectivity using FIS. However, fuzzy inference systems cannot measure the negative effects of individual pollutants on human health, since different air pollution parameters may cause different pathological responses within human bodies. The Analytic Hierarchy Process (AHP) provides a good solution to this problem. According to descriptions of the AHP methodology (Saaty, 1980, 1994), AHP can assign priorities through analysis of the effect of individual air pollutants on health. It thereby assigns different weights to different pollutants, which makes the assessment results more accurate. For example, Upadhyay et al. (2014) developed an AHP coupled fuzzy pattern recognition model to assess air quality in the district of Howrah. Olvera et al. (2016) developed a weighted Fuzzy Inference System to perform air quality assessment in the Mexico Valley area.

According to the literature mentioned above, studies on air pollution primarily emphasize the importance of obtaining deterministic forecasts or improving the assessment of air quality, but overlook the uncertainties in future air quality conditions, which is necessary for environmental influence assessment systems and early-warning system. In fact, all forecasting methods generate errors because of the models' systematic errors, resulting in inherent and irreducible uncertainties in forecasting (Pinson and Kariniotakis, 2010; El-Fouly et al., 2006). Therefore, in this paper, an early-warning system is developed for air quality prediction and assessment. In this system, an intelligent algorithm is employed to determine an optimal unimodal parameter distribution model for six air pollutants ($PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$, $O_3$ and CO), which help examine and analyze the statistical characteristics of air pollutant emissions. These models focus primarily on research on PM particles, though related work was also performed. Second, this paper emphasizes the integration of prediction and assessment in the developed early-warning system, which provides a trustworthy reference for both air pollution supervisors and members of the public. Finally, the developed early-warning system can quantify the uncertainties of air quality conditions to effectively predict future air pollution crises and then produces comprehensive and reliable analyses of air quality, thus overcoming the limitations of traditional method that produce deterministic prediction.

In general, the major contributions of this work are as follows:

● This paper describes a new early-warning system that consists of air quality prediction and assessment modules. In this system, the uncertainty of air quality predictions is emphasized to potentially extract information that can be used to improve the effectiveness of the early-warning system.

● This paper utilizes unimodal parameter distribution models to examine and analyze the statistical characteristic of air pollutant emissions. The emissions of different pollutants may differ statistically, requiring the selection of parameter models. Therefore, for specifying concrete pollutant emissions regimes, a stochastic heuristic optimization algorithm is employed to determine the choice and performance of the modeled parameter distributions.

● In the prediction module, a hybrid model is developed that integrates a back propagation neural network algorithm, an optimal parameter distribution model and data preprocessing techniques to generate deterministic and interval forecasts, thus overcoming the deficiencies of traditional air quality prediction methods.

● The air quality assessment module described in this paper emphasizes the understanding and analysis of the importance levels of different pollutants. A trapezoidal function is used to identify and define the negative impacts of individual pollutants. In addition, the importance levels of different pollutants differ greatly. Therefore, an analytic hierarchy process based on expert knowledge is employed to provide a reliable and intelligible description of air quality conditions.

● The early-warning system can predict air quality conditions to make a more reasonable and comprehensive analysis of air pollution and then provide a trustworthy reference for air pollution management and for providing the public with information on adverse air pollution conditions.

The rest of this paper is organized as follows. Section 2 introduces the relevant methodology. In the Section 3 the structure of the early-warning system is presented. Section 4 introduces the evaluation criteria that are used to assess the models. Section 5 presents the study area and describes the data. In section 6, a case study and analysis are given. Section 7 provides the conclusions of the study.

## *Methodology*

### *Method of estimating distribution functions*

The intent of fitting distribution functions is to characterize the statistical behavior of environmental parameters. As a result, it also helps facilitate interval estimation. In this paper, the *PDFs* of the lognormal, gamma, weibull and rayleigh were adopted to model the distributions of air pollutant concentrations (see **Table 1**). These distribution functions, which represent a popular way to reflect the characteristics of data, have been successfully used in some parts of literatures. For example, Sun et al. (2013) used a hidden Markov model with different emission distributions to predict 24-hour-average $PM_{2.5}$ concentrations. Song et al. (2015) applied the distributions of PM emission in the construction of PM interval forecasts. Wang et al. (2015) and Wu et al. (2013) used the probability distribution of wind speeds in wind energy and extreme wind speed estimation. Meanwhile, intelligent optimization method provides a more robust searching capability for the determination of parameters than conventional approaches such as graphical, minimum least square (MLS) and maximum likelihood estimation (MLE) methods. Therefore, this study utilizes the cuckoo search optimization (CSO) algorithm to search for the optimal distribution parameter value to determine which distribution is most suited to represent air pollutant emissions. A detailed review of CSO algorithm can be found in Zhang et al. (2017).

**Table 6**

Probability density functions.

| Distribution | Probability density function | Parameters |
|---|---|---|
| Lognormal | $f(\mathrm{x},\mu,\sigma)=\frac{1}{x\sqrt{2\pi\sigma}}\exp(-\frac{(\ln(\mathrm{x})-\mu)^2}{2\sigma^2})$ | $\mu,\sigma$ |
| Rayleigh | $f(\mathrm{x},\sigma)=\frac{1}{\sigma^2}\exp(-\frac{x^2}{2\sigma^2})$ | $\sigma$ |
| Gamma | $f(\mathrm{x},\xi,\theta)=\frac{x^{\xi-1}}{\theta^{\xi}\Gamma(\xi)}\exp(-\frac{x}{\theta})$ | $\xi,\theta$ |
| Weibull | $f(\mathrm{x},\mathrm{k},\mathrm{c})=\frac{k}{c}(\frac{x}{c})^{k-1}\exp(-(\frac{x}{c})^k)$ | $k,c$ |

***Air quality prediction method***

In this section, a back propagation neural network algorithm, combined with data preprocessing technique and probability-based method, is introduced to tackle the uncertainty in future air quality conditions.

### 1.1.1 *Back propagation neural network improved using singular spectrum analysis*

In this study, the back propagation neural network (BPNN) is employed for constructing air pollutant predictor. Within the structure of a BP neural network, each layer is composed of certain neuron nodes, and there are specific connection weights among the neuron nodes in different layers. Generally, BP neural network training is an error back-propagation process that provides a basis for modifying the network weights. As the repeated error back-propagation progresses, the error between the target value and the actual value gradually approaches an acceptable value. In addition, the chaotic nature of the original pollutant data, which has an important effect on accuracy of the resulting forecast, is the real problem, regardless of the forecasting methods used. To overcome this problem, this work makes use of the singular spectrum analysis (SSA) algorithm in pre-processing data series to improve forecasting performance of the BP neural network, resulting in the SSABPNN predictor. A brief review of SSA can be found in Ma et al. (2017).

### 1.1.2 *Interval estimation of air pollutant concentrations*

Quantifying potential uncertainties using interval forecasting has been applied to a number of problems, such as forecasting electric load, wind speed and air pollutant concentrations. The construction of forecasting intervals in this process is primarily discussed because of its complexity and subjectivity, and many methods of performing interval forecasting rely on special assumptions about the distribution of the historical data. Generally, given a confidence level of (1-α) %, the relationship between the predicted interval and the actual value can be represented by Eq. (1):

$$P\left(\mathrm{P}_{lower}\le Y_t\le P_{upper}\right)=1-\alpha \tag{1}$$

According to the above equation, if deterministic forecasting of future pollutant concentrations is implemented, the corresponding dynamic forecasting interval can be estimated properly.

Further, this paper assumes that the predicted values $\hat{y}$ , which are the expected values of future points, have the similar distribution *f* as the historical values. Assuming symmetrical probabilities, the dynamic forecasting intervals $\left(\hat{P}_{Lower}, \hat{\mathrm{P}}_{Upper}\right)$ can be estimated using Eq. (2):

$$\begin{cases} \left\{\left(\hat{P}_{Lower}, \hat{\mathrm{P}}_{Upper}\right) \middle| P(\hat{\mathrm{P}}_{lower} \le Y_t \le \hat{P}_{upper}) = 1 - \alpha\right\} \\ \int_{\hat{P}_{Lower}}^{\hat{y}} f(x)\,dx = F(\hat{y}) - \alpha / 2 \\ \int_{\hat{y}}^{\hat{P}_{Upper}} f(x)\,dx = 1 - (F(\hat{y}) + \alpha)) / \end{cases} \quad (2)$$

Given a confidence level and a deterministic forecast time series, the corresponding $\left(\hat{\boldsymbol{P}}_{Lower}, \hat{\boldsymbol{P}}_{Upper}\right)$ can be obtained.

### *Air quality assessment methods*

The air quality index equation (USEPA, 2006, 2009) is commonly used as a reference. However, in the AQI, the pollutant with the highest concentrations is identified as the reference basis for producing protective plans by environmental supervisors, which frequently results in irrational allocation of resources and misleading information being provided to the public. To overcome this problem, the Analytic Hierarchy Process (AHP) is employed to generate priority assignments through analyzing the effect of each air pollutant on health. To determine the classification level to which each set of air quality observations belongs, the fuzzy set theory is introduced in detail.

### 1.1.3 *Analytic Hierarchy Process*

In air quality assessment, different pollutants may generate different problems for sensitive groups. For this reason, the importance levels of environmental parameters must be quantified to produce deterministic weights. The weights are then attributed to the corresponding environmental parameters, making the assessment and analysis more sensible (Carbajal-Hernández, 2012; Chakraborty and Dey, 2006). An Analytic Hierarchy Process (AHP) has considerable advantages in performing prioritization. The procedure of applying AHP can be implemented using the following three steps. In the first step, based on the representation of the problem using a hierarchical structure, the scale of importance of the environmental parameters should be presented properly. **Table 2** shows the importance values used in this work.

The next step is to develop a consistent matrix through pairwise comparison. A consistent matrix is a positive reciprocal *n*×*n* matrix whose elements satisfy the relation $a_{ij} \times a_{jk} = a_{ik}$ for *i, j, k* = 1, …, *n*. These pairwise comparisons can also be described using the following matrix form:

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \tag{3}$$

where $w_i$ is the importance value of the *i*th air pollutant.

The vector $\bar{w} = [\bar{w}_1, \bar{w}_2, \cdots, \bar{w}_n]$ is then calculated by averaging the row elements as follows:

$$\bar{w}_i = \frac{1}{n}\sum_{j=1}^{n} w_i / w_j, \quad i = 1, \cdots, n. \tag{4}$$

Subsequently, the priority weight $w_i$ is obtained by normalizing $\bar{w}_i$ as follows:

$$w_i = \frac{\bar{w}_i}{\sum_{i=1}^{n} \bar{w}_i}, \forall i = 1, \cdots, n. \tag{5}$$

where the value of $\sum_{i=1}^{n} w_i$ is 1.

Finally, the consistency ratio (CR) for the pairwise comparison matrix must be calculated using:

$$CR = \frac{\lambda_{\max} - n}{(\mathrm{n}-1)\,\mathrm{RI}} \tag{6}$$

where $\lambda_{\max}$ represents the maximum eigenvalue of the matrix '***A***' and *n* and *RI* are the matrix size and the random index, respectively. **Table 3** presents the pairs (*n*, *CR*). In this paper, the importance level of each pollutant is specified by air quality experts. The results show that the CR value is 0, which is acceptable, according to (Saaty, 2004). The pairs (*importance value*, *priority weight*) are presented in **Fig. 1**.

**Table 2**

Scale of importance.

| Importance value | Definition |
|---|---|
| 1 | Equal importance |
| 2 | Weak or slight importance |
| 3 | Moderate importance |
| 4 | Moderate plus importance |
| 5 | Strong importance |

| | |
|---|---|
| 6 | Strong plus importance |
| 7 | Very strong or demonstrated importance |
| 8 | Quite strong importance |
| 9 | Extreme importance |

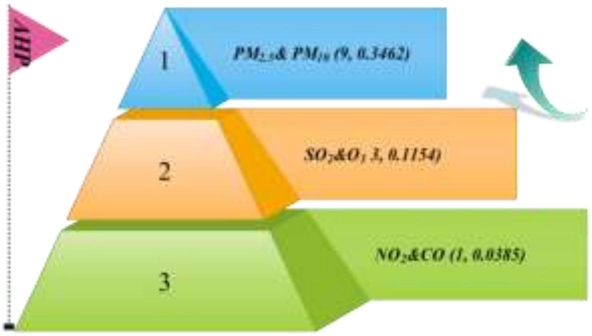

**Fig. 1. The pairs (*importance value, priority weight*) for the six pollutants.**

### 1.1.4 *Fuzzy sets*

In fuzzy inference systems, membership functions can transform uncertain information into values in the [0, 1] range. Given a pollutant concentration and classification level, membership functions can determine the specific range to which it belongs. Three membership functions (*triangular*, *trapezoidal* and *Gaussian*) are commonly used. In this paper, when concentrations corresponding to the classification levels *good* and *extremely high* were considered, the trapezoidal membership function was selected. For the intermediate level, the triangular function was selected. For quick computing, linear membership functions were used. The trapezoidal function can be defined as follows.

$$trapezoidal: \mu(x,a,b,c,d) = max\left\{ min\left( \frac{x-a}{b-a}, 1, \frac{d-x}{d-c} \right), 0 \right\} \tag{7}$$

where *x* represents the air quality parameter, and *a*, *b*, c and *d* are membership function parameters. The triangular function is obtained when *b*=*c*. **Table 4** displys the classification levels according to Technical Regulation on Ambient Air Quality Index of China. Based on **Table 4**, **Table 5** presents the values of the membership function parameters.

In this work, the negative effects of individual pollutants were evaluated by $\mu(x)$, which represents to what degree a concentration value belongs to a specific classification level. For the *i*th pollutant, the output membership set for the determined classification level can be represented as follows:

$$\mu_i = \left\{ \mu^G(x), \mu^M(x), \mu^{Lp}(x), \mu^{Mp}(x), \mu^{Hp}(x), \mu^{Sp}(x) \right\} \tag{8}$$

where *G*, *M*, *Lp*, *Mp*, *Hp*, and Sp∈{*"Good"*, *"Moderate"*, *"Lightly polluted"*, *"Moderately polluted"*, *"Heavily polluted"*, *"Severely polluted"*}.

### 1.1.1 *Weighted air quality assessment*

The air quality index (AQI) is determined using the highest individual AQI value for measured pollutants in China. In this work, to obtain an ensemble of air quality assessments, the membership matrix **R** can be defined as:

$$\boldsymbol{R} = \begin{bmatrix} \mu_{11} & \mu_{21} & \cdots & \mu_{m1} \\ \mu_{12} & \mu_{22} & \cdots & \mu_{m1} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{1n} & \mu_{2n} & \cdots & \mu_{mn} \end{bmatrix} \tag{9}$$

where $\mu_{mn}$ represents the membership value that reflects to what degree the *m*th pollutant concentration value belongs to the *n*th classification level. Based on the membership matrix **R** and the weight matrix **W**, the output ensemble membership can be obtained as follows:

$$\boldsymbol{\mu}_{out} = \boldsymbol{R} \times \boldsymbol{W} \tag{10}$$

Finally, the ensemble of air quality assessment results can be evaluated as follows:

$$\mu_{R\_out} = max\left\{\boldsymbol{\mu}_{out}^{p}\right\} \tag{11}$$

where p$\in${*"Good", "Moderate", "Lightly polluted", "Moderately polluted", "Heavily polluted", "Severely polluted"*}. $\mu_{R\_out}$ represents the membership value of the current air quality in the determined classification level and reflects the negative impact of air pollution by integrating the six pollutants using weighting coefficients.

**Table 3**

Random Index values.

| Matrix size(n) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Random Index (RI) | 0 | 0 | 0.52 | 0.89 | 1.11 | 1.25 | 1.35 | 1.4 | 1.45 | 1.49 |

**Table 4**

Air quality classification levels according to the Technical Regulation on Ambient Air Quality Index of China.

| Parameter | Level Ⅰ | Level Ⅱ | Level Ⅲ | Level Ⅳ | Level Ⅴ | Level Ⅵ |
|---|---|---|---|---|---|---|
| | "Good" | "Moderate" | "Lightly polluted" | "Moderately polluted" | "Heavily polluted" | "Severely polluted" |
| $PM_{2.5}$[μg/m$^3$] | 0-35 | 36-75 | 76-115 | 116-150 | 151-250 | >250 |
| $PM_{10}$[μg /m$^3$] | 0-50 | 51-150 | 151-250 | 251-350 | 351-420 | >420 |
| $SO_2$[μg/m$^3$] | 0-50 | 51-150 | 151-475 | 476-800 | 801-1600 | >1600 |
| $NO_2$[μg/m$^3$] | 0-100 | 101-200 | 201-700 | 701-1200 | 1201-2340 | >2340 |
| $O_3$[μg/m$^3$] | 0-100 | 101-160 | 161-215 | 216-265 | 266-800 | >800 |
| CO[mg/m$^3$] | 0-5 | 6-10 | 11-35 | 36-60 | 61-90 | >90 |

**Table 5**

Values of the membership function parameters.

| Classification | "Good" | | | "Moderate" | | | "Lightly polluted" | | | "Moderately polluted" | | | "Heavily polluted" | | | "Severely polluted" | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | a=b | c | d | a | b=c | d | a | b=c | d | a | b=c | d | a | b=c | d | a | b | c=d |
| $PM_{2.5}$[μg/m$^3$] | 0 | 17.5 | 55 | 17.5 | 55 | 95 | 55 | 95 | 132.5 | 95 | 132.5 | 200 | 132.5 | 200 | 300 | 200 | 300 | ∞ |
| $PM_{10}$[μg /m$^3$] | 0 | 25 | 100 | 25 | 100 | 200 | 100 | 200 | 300 | 200 | 300 | 385 | 300 | 385 | 460 | 385 | 460 | ∞ |
| $SO_2$[μg/m$^3$] | 0 | 25 | 75 | 25 | 75 | 313 | 75 | 313 | 638 | 313 | 638 | 1200 | 638 | 1200 | 1850 | 1200 | 1850 | ∞ |
| $NO_2$[μg/m$^3$] | 0 | 50 | 150 | 50 | 150 | 450 | 150 | 450 | 950 | 450 | 950 | 1770 | 950 | 1770 | 2910 | 1770 | 2910 | ∞ |
| $O_3$[μg/m$^3$] | 0 | 50 | 130 | 50 | 130 | 187.5 | 130 | 187.5 | 240 | 187.5 | 240 | 532.5 | 240 | 532.5 | 1067.5 | 532.5 | 1067.5 | ∞ |
| CO[mg/m$^3$] | 0 | 2.5 | 7.5 | 2.5 | 7.5 | 22.5 | 7.5 | 22.5 | 47.5 | 22.5 | 47.5 | 75 | 47.5 | 75 | 115 | 75 | 445 | ∞ |

***Framework of the developed early-warning system***

The framework of the developed early-warning system is presented in **Fig. 2**. It consists of pre-analysis of the data, the air quality prediction model and the air quality assessment method. In general, the developed early-warning system is divided into three stages; the main tasks of each stage are summarized as follows.

**Stage I**: Pre-analysis. Four unimodal parameter distributions, specifically the lognormal, gamma, weibull and Rayleigh distributions, are utilized to examine and analyze the characteristics of pollutant emission. In particular, when the four parameter models are used to model time series of six pollutants, the fitting performance depends on the estimation of the model parameters. As previously mentioned, artificial intelligent optimization algorithms have a robust capacity to determine value of the model parameters and thus the CS optimization algorithm is used to determine the optimal parameter values. Finally, the probability distribution with the best fitting performance can be used to represent the pollution regime of the studied area and reflects the statistical characteristics of pollutant emissions.

**Stage II**: Forecasting. The neural network predictor is employed to forecast the pollutant concentrations using the information from the historical data. To improve the forecasting performance, powerful signal processing techniques are integrated into the input processing of the neural network. By applying the improved predictor to pollutants concentrations, the deterministic forecasts can be obtained. Further, probability-based methods are also integrated into the improved predictor to generate forecast intervals, solving the uncertainty problems associated with air quality forecasting.

**Stage III**: Air quality assessment. Fuzzy set theory and the Analytic Hierarchy Process are used to construct the air quality index. To overcome the defects of the conventional air quality index (AQI), this paper first employs the trapezoidal function to identify and define the negative impacts of individual pollutants, resulting in membership degrees for the individual pollutants. Second, to obtain ensembles of air quality assessments, each pollutant is assigned a priority using Analytic Hierarchy Process-based expert knowledge.

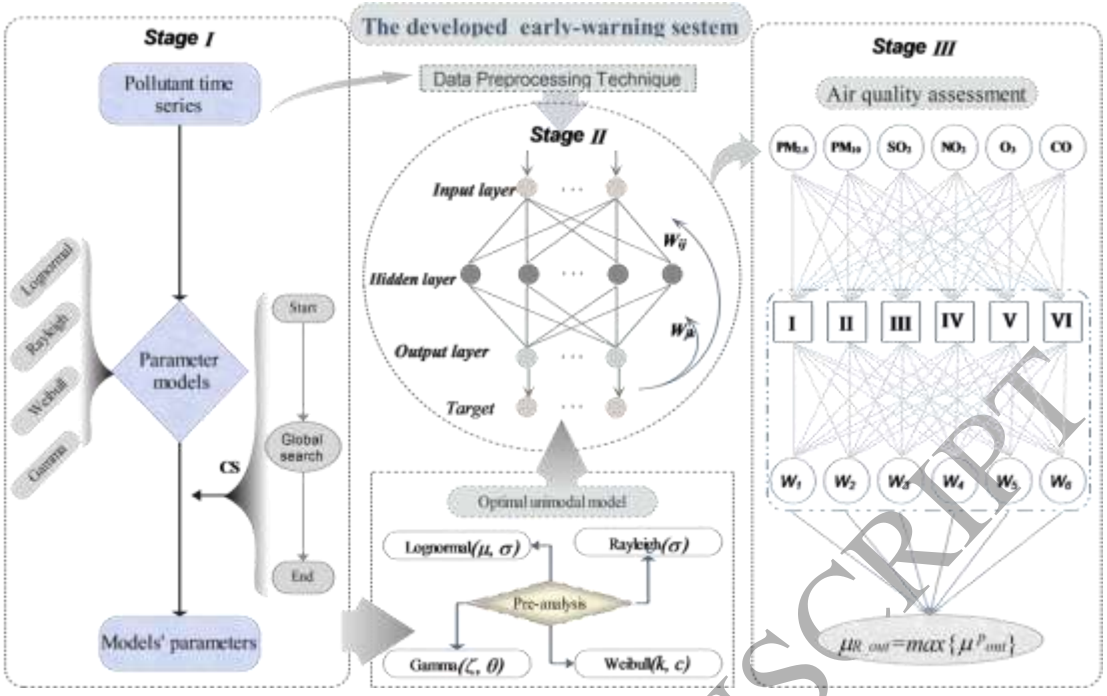

**Fig. 2. The structure of the developed early-warning system**

***Evaluation criteria***

To assess the distributions of air pollutants concentrations, $R^2$ and the root mean squared error (*RMSE*) are adopted to evaluate the fitting performance of each distribution to the original data. The metrics $R^2$ and *RMSE* are defined as:

$$R^2 = 1 - \sum_{i=1}^{N}(F_i - \hat{F}_i)^2 \Big/ \sum_{i=1}^{N}(F_i - \bar{F})^2 \tag{12}$$

$$RMSE = \sqrt{\sum_{i=1}^{N}(F_i - \hat{F}_i)^2 \Big/ N} \tag{13}$$

where $N$, $F_i$ , $\hat{F}_i$ and $\bar{F}$ denote the number of groups, the observed cumulative probability, the cumulative probability that is estimated with the theoretical distributions and the average of $F_i$ , respectively.

The mean square error (MSE) and the mean absolute percent error (MAPE) are used to evaluate the effectiveness of deterministic forecasting. These two indices are given by:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}\left(\hat{Y}_t - Y_t\right)^2 \tag{14}$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{T}\left|\frac{\hat{Y}_t - Y_t}{Y_t}\right| \times 100\% \tag{15}$$

where $\hat{Y}_t$ and $Y_t$ are the predicted and actual values, respectively.

Another two criteria that are adopted to evaluate the quality of the forecast intervals are the interval coverage probability (*ICP*) and the interval average width (*IAW*), which are defined as follows:

$$ICP = \frac{1}{T}\sum_{t=1}^{T}\rho_t \tag{16}$$

$$IAW = \frac{1}{T}\sum_{t=1}^{T}(\mathrm{U}_t - L_t) \tag{17}$$

where $\rho_t = 1$ if the target value is covered by $[\mathrm{U}_t, \mathrm{L}_t]$; otherwise $\rho_t = 0$. $U_t$ and $L_t$ represent the upper and lower bounds of the *t*th forecast interval, respectively.

***Study area and data description***

***Study area***

Chengdu, which is located in the southwestern part of China, is the capital of Sichuan Province as well as a state Historical and Cultural City, with its more than 3000 years of history. The city contains a large area of hills and mountains, has an area of 12,390 km$^2$, and has a maximum altitude of 5364 m above sea level. Due to its particular geographic location, climate variables, such as air temperature, humidity and wind speed, vary considerably within its boundaries. In addition, Chengdu is the high-tech industry base, trade center and transportation hub of southwestern China.

Hangzhou, which is the capital of Zhejiang Province, is located along southeastern coast of China. Given that it combines history and culture with a natural environment that integrates rivers, lakes and hills, Hangzhou has become an important national tourist city. Located at a geographic position of latitude 30° 16' North and longitude 120° 12' East, Hangzhou experiences a subtropical monsoon climate with four distinct seasons. Benefiting from its geographic conditions, Hangzhou enjoys a warm and humid climate with sufficient sunshine and plentiful rainfall. In addition, Hangzhou is renowned the “Paradise on Earth”, the “Cultural State”, the “Home of Silk”, the “Tea Capital” and the “Town of Fish and Rice” in China.

***Data description***

The air quality data used in this paper were obtained from the Ministry of Environmental Protection of China (http://113.108.142.147:20035/emcpublish/). The data report the concentrations of six key pollutants: Carbon monoxide and nitrogen dioxide were reported using a moving average window size of 1 h, sulfur dioxide and particulate matter smaller than 10 and 2.5 $\mu m$ were reported using a moving average window size of 24 h and ozone was reported using a moving average window size of 8 h. The data extend from November 1, 2014 to October 31, 2015 in Chengdu and Hangzhou. In addition, pollutant concentration levels are shown in **Fig. 3**.

In this paper, to verify the proposed approach, we assume that the concentrations six pollutants are random variables, and the data sample is split into two parts, which are the training subset and the testing subset. For each city, the air quality data extending from November 1, 2014 to August 31, 2015 are used in the training process to construct the model, and the data extending from September 1, 2015 to October 31, 2015 are used in the testing process to verify the performance of the developed system. In addition, in this study, each BP predictor is constructed with twelve input nodes, twenty-five hidden nodes and six output nodes. In this sense, the proposed method is able to forecast a series of future values of each parameter $x(t+1)$, $x(t+2),\ldots,x(t+6)$ using a series of past values x($t$), $x(t-1),\ldots,x(t-11)$. More importantly, driven by the different initial weights and thresholds, the prediction performance of neural network is always unstable. To overcome this shortcoming, each predictor is evaluated 100 times and then the average value is used.

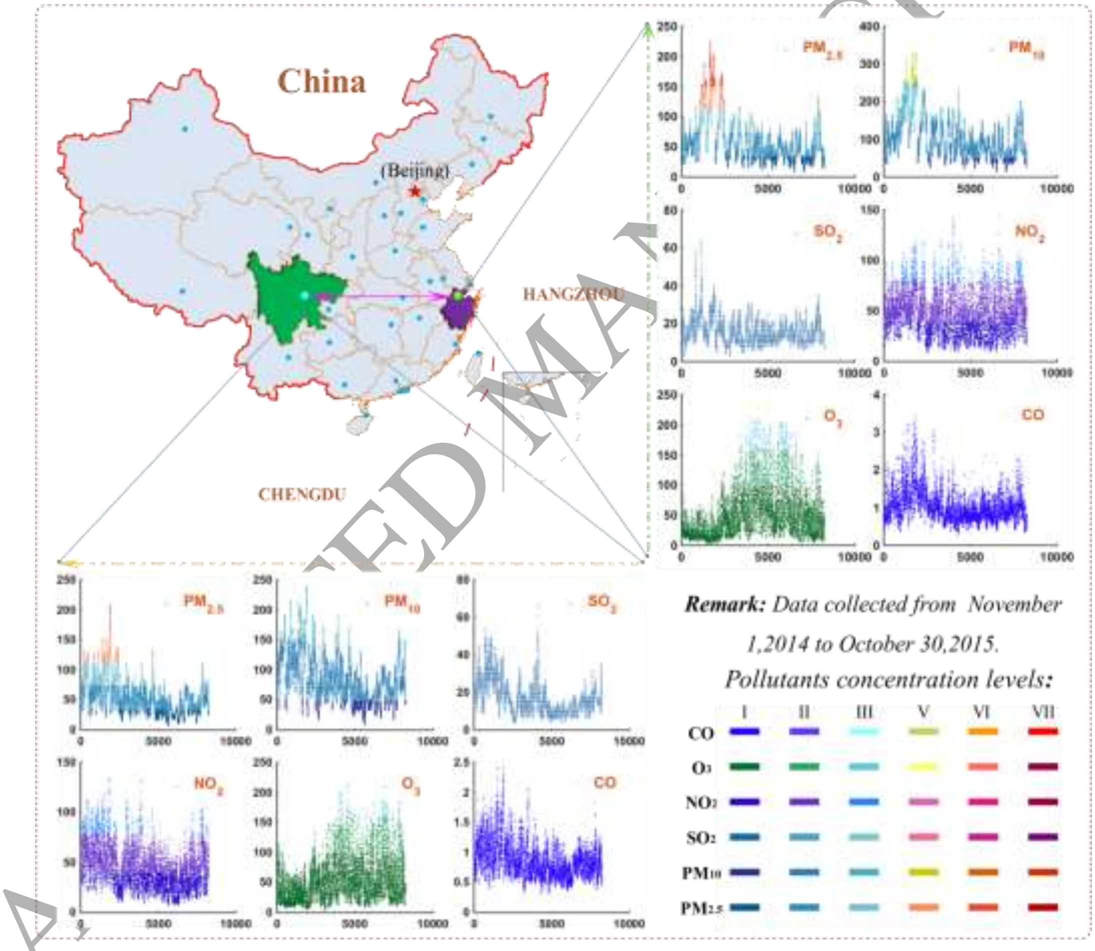

**Fig. 3. Pollutant concentration levels for Chengdu and Hangzhou.**

***Case study and analysis***

In this section, pollutant distributions are primarily estimated by using probability density functions optimized using the CSO algorithm. To obtain the forecast intervals, deterministic forecasting is performed with the improved neural network model. Using these deterministic forecasts and the optimal distribution parameters, interval forecasts are presented in this paper. Based on deterministic and interval forecast results, the corresponding air quality assessment index values can be calculated, according to the classification levels for the air quality parameters.

***Estimation of pollutant distributions***

The air pollutant distributions are estimated using four probability density functions, *i.e.*, the lognormal, rayleigh, gamma and weibull distributions. The best distribution function representing the pollution regime is determined in terms of the performance of the distributions according to two evaluation criteria. In estimating the parameters, maximum likelihood estimation (MLE), which can provide asymptotically centered estimators, serves as a basis for comparison to confirm the effectiveness of the cuckoo search optimization algorithm. **Table 6** presents the estimated parameter values of the four distribution functions for the two cities. **Table** 7 shows the evaluation indices of the performances of the distributions. As **Table 6** shows, the parameter values reflect the scale and translational transformations of these distributions. At Chengdu, the parameter values of $PM_{2.5}$, $PM_{10}$, $SO_2$, $NO_2$ and $O_3$ are close to those of Hangzhou, respectively, whereas the values for CO value reflects substantial differences between the two cities, which gives an indication of how pollutant characteristics change between different geographic locations. It should be noted that the value of the location parameter of the lognormal distribution is negative because CO occurs at smaller concentrations than the other pollutants; for example, the maximum value and the minimum value for hourly CO measurements in Chengdu are 3.454 (mg/m$^3$) and 0.278 (mg/m$^3$), respectively. When the evaluation criteria are used to evaluate the performance of the distributions, as can be seen from **Table 7**, the values of $R^2$ and RMSE vary together; thus, the $R^2$ values can be considered as the primary indicator of the performance of the distributions. In all of the estimated distributions, the maximum and minimum $R^2$ values are obtained from CSO-Gamma-$NO_2$ and MLE-Rayleigh-CO at Chengdu, but it cannot be concluded that the gamma distribution provides the best distribution of all of the environmental parameters. According to the **Table 7**, CSO-Lognormal yields the best performance for $PM_{2.5}$, $PM_{10}$, $O_3$ and CO in Chengdu and $SO_2$, $O_3$ and CO in Hangzhou, CSO-Gamma yields the best performance for $SO_2$ and $NO_2$ in Chengdu and $PM_{10}$ and $NO_2$ in Hangzhou; and MLE-Gamma yields the best performance for $PM_{2.5}$ in Hangzhou. It can be concluded that lognormal and gamma distribution functions seem to have more powerful capabilities to fit the empirical pollutant emissions than the Rayleigh and Weibull distributions, and CSO algorithm has a distinct advantage in helping the distribution functions perform better. **Fig. 4** shows the values of the evaluation indices $R^2$ and RMSE.

**Fig. 5** and **Fig. 6** are provided to clearly show the performance of modeling the recorded data using the distributions optimized. The number of groups in the sub graph is 20. From the two figures, it can be seen that lognormal and gamma distributions yielded the best effectiveness after fitting, whereas the Rayleigh distribution obtained the worst performance. In general, if the slightly better performance of the gamma distribution compared to that of the lognormal distribution can be ignored, it is well known that the latter is the best distribution model.

**Table 6**

Parameter values of the four evaluated distribution functions.

| Environmental parameter | Moving average window size | Methods | Distribution function | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Lognormal | | Rayleigh | Gamma | | Weibull | |
| **Chengdu** | | | $\mu$ | $\sigma$ | $\sigma$ | $\xi$ | $\theta$ | $k$ | $c$ |
| $PM_{2.5}$ | 24 h | CSO | 3.92 | 0.57 | 41.50 | 3.54 | 15.36 | 58.38 | 2.05 |
| | | MLE | 3.95 | 0.58 | 51.05 | 3.10 | 19.84 | 69.56 | 1.76 |
| $PM_{10}$ | 24 h | CSO | 4.52 | 0.57 | 76.27 | 3.62 | 27.53 | 107.15 | 2.08 |
| | | MLE | 4.52 | 0.54 | 85.20 | 3.65 | 28.99 | 120.09 | 1.97 |
| $SO_2$ | 24 h | CSO | 2.69 | 0.41 | 11.98 | 6.38 | 2.42 | 16.38 | 2.84 |
| | | MLE | 2.67 | 0.41 | 12.14 | 6.05 | 2.60 | 17.76 | 2.38 |
| $NO_2$ | Hourly | CSO | 3.85 | 0.47 | 37.85 | 5.06 | 9.74 | 52.34 | 2.51 |
| | | MLE | 3.79 | 0.46 | 37.53 | 5.23 | 9.31 | 55.06 | 2.48 |
| $O_3$ | 8 h | CSO | 3.71 | 0.96 | 35.44 | 1.52 | 33.22 | 51.72 | 1.29 |
| | | MLE | 3.66 | 0.84 | 48.47 | 1.70 | 31.53 | 58.41 | 1.32 |
| CO | Hourly | CSO | -0.06 | 0.34 | 0.79 | 9.10 | 0.11 | 1.02 | 3.33 |
| | | MLE | -0.02 | 0.37 | 0.80 | 7.33 | 0.14 | 1.18 | 2.57 |
| **Hangzhou** | | | $\mu$ | $\sigma$ | $\sigma$ | $\xi$ | $\theta$ | $k$ | $c$ |
| $PM_{2.5}$ | 24 h | CSO | 3.91 | 0.55 | 41.00 | 4.36 | 12.41 | 57.33 | 2.18 |
| | | MLE | 3.87 | 0.53 | 43.83 | 3.88 | 14.18 | 62.36 | 2.05 |
| $PM_{10}$ | 24 h | CSO | 4.37 | 0.51 | 64.51 | 4.33 | 19.43 | 89.72 | 2.29 |
| | | MLE | 4.32 | 0.50 | 65.77 | 4.46 | 18.86 | 95.34 | 2.28 |
| $SO_2$ | 24 h | CSO | 2.69 | 0.54 | 12.15 | 4.04 | 3.91 | 16.78 | 2.24 |
| | | MLE | 2.69 | 0.51 | 13.49 | 3.98 | 4.23 | 19.11 | 2.01 |
| $NO_2$ | Hourly | CSO | 3.76 | 0.50 | 34.83 | 4.54 | 10.01 | 48.40 | 2.39 |
| | | MLE | 3.70 | 0.49 | 35.13 | 4.66 | 9.68 | 51.06 | 2.31 |
| $O_3$ | 8 h | CSO | 3.88 | 0.76 | 39.86 | 2.24 | 24.48 | 58.42 | 1.62 |
| | | MLE | 3.81 | 0.71 | 48.66 | 2.31 | 24.69 | 63.77 | 1.57 |
| CO | Hourly | CSO | -0.21 | 0.31 | 0.70 | 10.88 | 0.08 | 0.88 | 3.68 |
| | | MLE | -0.19 | 0.32 | 0.65 | 10.02 | 0.09 | 0.97 | 3.06 |

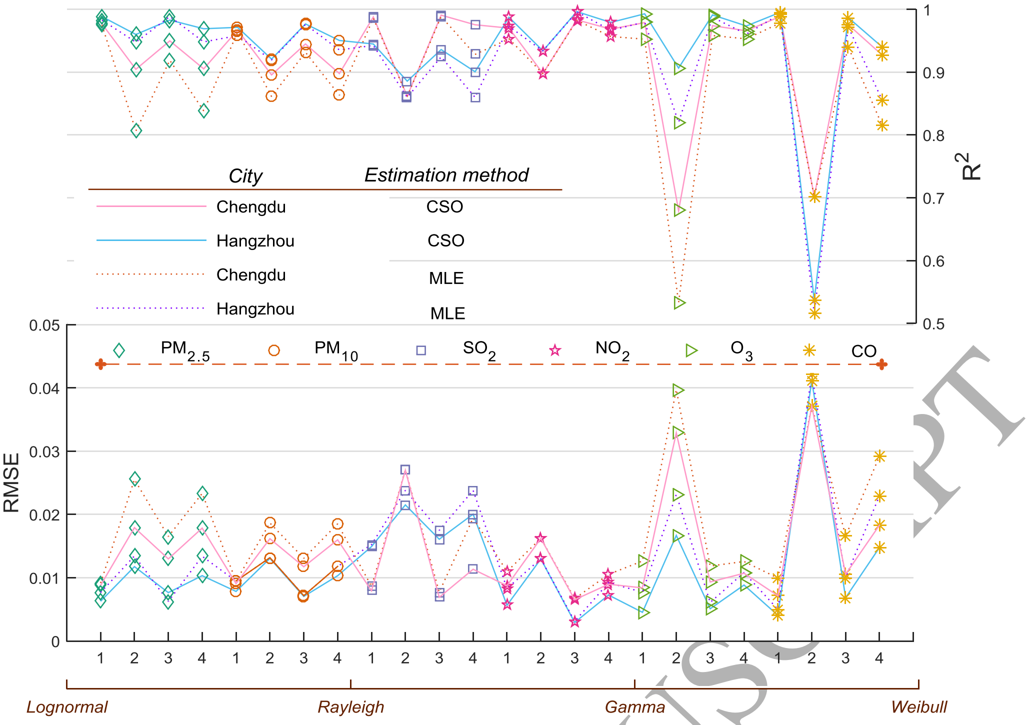

**Fig. 4. $R^2$ and RMSE values of the distributions examined.**

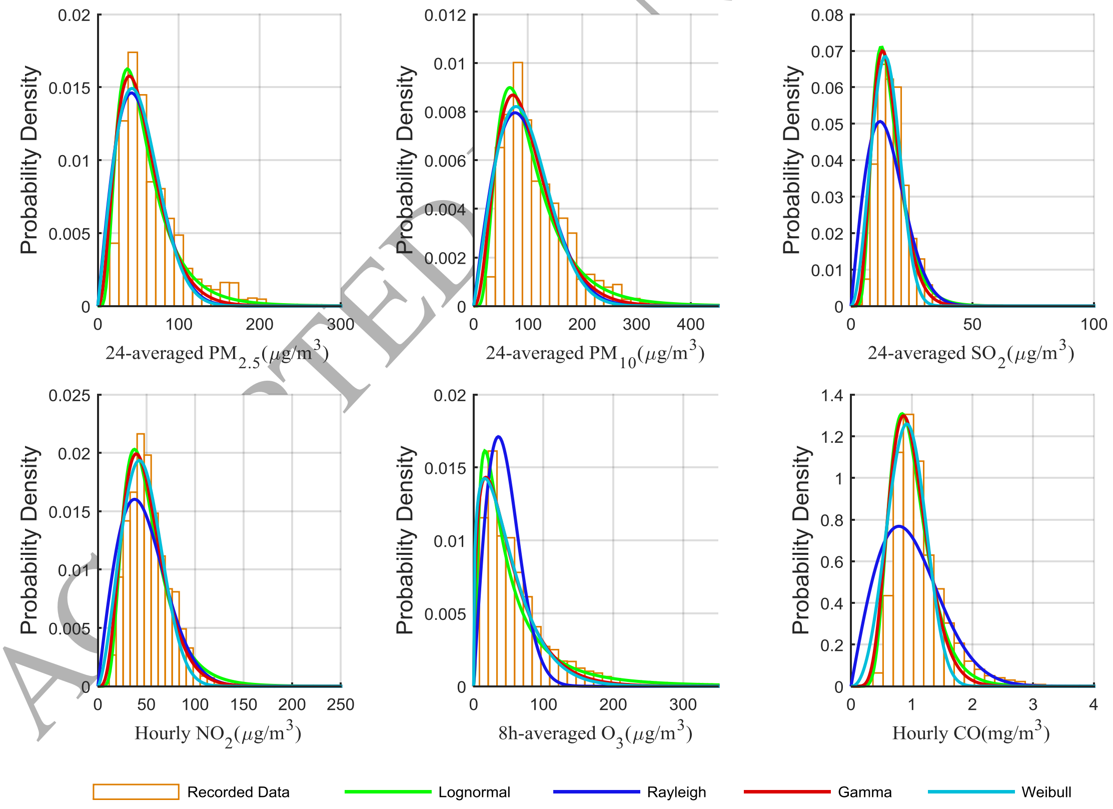

**Fig. 5. Fitted distributions and frequency histograms (Chengdu)**

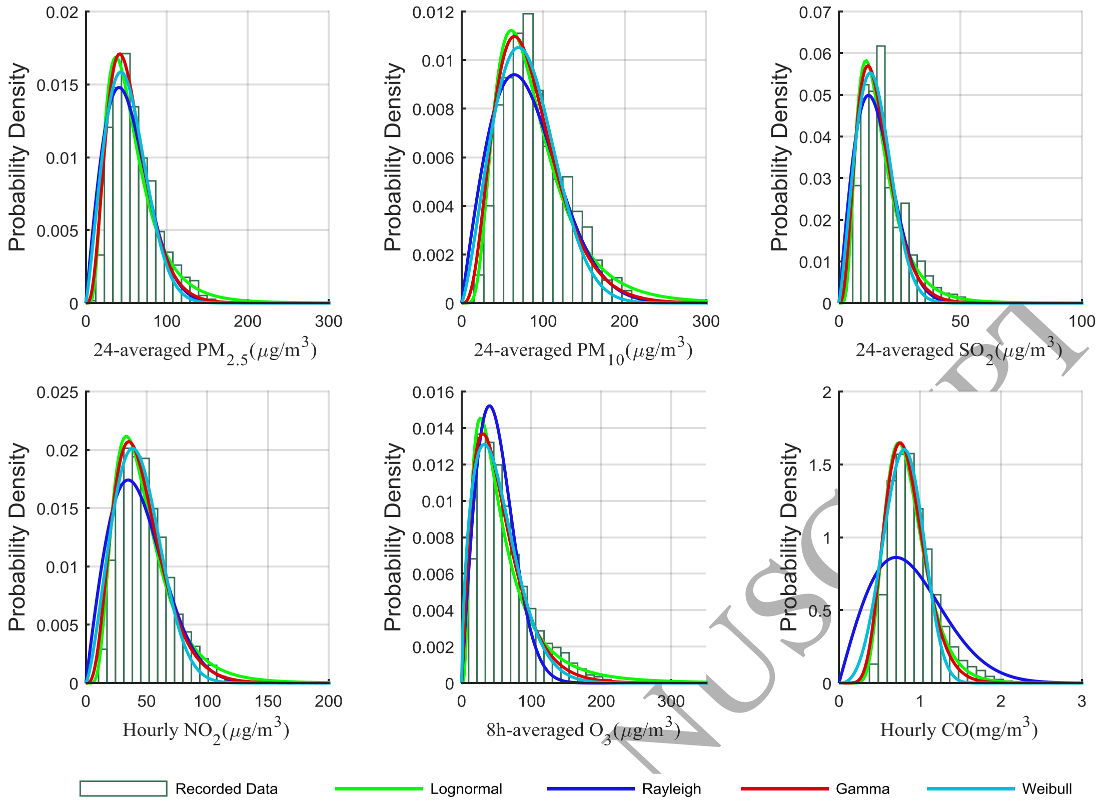

**Fig. 6. Fitted distributions and frequency histograms (Hangzhou).**

## Table 7

R$^2$ and RMSE values of the distributions examined.

| Environmental parameter | Moving average window size | Method | Evaluation metrics for distribution function | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Lognormal | | Rayleigh | | Gamma | | Weibull | |
| **Chengdu** | | | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** |
| PM$_{2.5}$ | 24 h | CSO | 0.977 | 0.009 | 0.905 | 0.018 | 0.950 | 0.013 | 0.905 | 0.018 |
| | | MLE | 0.975 | 0.009 | 0.807 | 0.026 | 0.919 | 0.017 | 0.839 | 0.023 |
| PM$_{10}$ | 24 h | CSO | 0.967 | 0.009 | 0.896 | 0.016 | 0.945 | 0.012 | 0.898 | 0.016 |
| | | MLE | 0.964 | 0.010 | 0.861 | 0.019 | 0.932 | 0.013 | 0.863 | 0.019 |
| SO$_2$ | 24 h | CSO | 0.988 | 0.008 | 0.862 | 0.027 | 0.991 | 0.007 | 0.976 | 0.011 |
| | | MLE | 0.986 | 0.009 | 0.862 | 0.027 | 0.989 | 0.008 | 0.928 | 0.019 |
| NO$_2$ | Hourly | CSO | 0.970 | 0.009 | 0.898 | 0.016 | 0.983 | 0.007 | 0.969 | 0.009 |
| | | MLE | 0.953 | 0.011 | 0.898 | 0.016 | 0.983 | 0.007 | 0.957 | 0.011 |
| O$_3$ | 8 h | CSO | 0.979 | 0.008 | 0.679 | 0.033 | 0.974 | 0.009 | 0.966 | 0.011 |
| | | MLE | 0.953 | 0.013 | 0.532 | 0.040 | 0.958 | 0.012 | 0.953 | 0.013 |
| CO | Hourly | CSO | 0.989 | 0.007 | 0.702 | 0.037 | 0.976 | 0.011 | 0.928 | 0.018 |
| | | MLE | 0.979 | 0.010 | 0.702 | 0.037 | 0.941 | 0.017 | 0.816 | 0.029 |
| **Hangzhou** | | | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** | ***R²*** | ***RMSE*** |
| PM$_{2.5}$ | 24 h | CSO | 0.989 | 0.006 | 0.960 | 0.012 | 0.983 | 0.008 | 0.969 | 0.010 |
| | | MLE | 0.983 | 0.008 | 0.948 | 0.014 | 0.989 | 0.006 | 0.948 | 0.013 |
| PM$_{10}$ | 24 h | CSO | 0.972 | 0.008 | 0.921 | 0.013 | 0.977 | 0.007 | 0.951 | 0.010 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | MLE | 0.958 | 0.010 | 0.920 | 0.013 | 0.976 | 0.007 | 0.936 | 0.012 |
| $SO_2$ | 24 h | CSO | 0.945 | 0.015 | 0.886 | 0.022 | 0.936 | 0.016 | 0.901 | 0.020 |
| | | MLE | 0.943 | 0.015 | 0.860 | 0.024 | 0.924 | 0.018 | 0.860 | 0.024 |
| $NO_2$ | Hourly | CSO | 0.988 | 0.006 | 0.934 | 0.013 | 0.997 | 0.003 | 0.979 | 0.007 |
| | | MLE | 0.973 | 0.008 | 0.934 | 0.013 | 0.996 | 0.003 | 0.967 | 0.009 |
| $O_3$ | 8 h | CSO | 0.993 | 0.005 | 0.907 | 0.017 | 0.991 | 0.005 | 0.974 | 0.009 |
| | | MLE | 0.980 | 0.008 | 0.820 | 0.023 | 0.988 | 0.006 | 0.961 | 0.011 |
| CO | Hourly | CSO | 0.996 | 0.004 | 0.537 | 0.041 | 0.987 | 0.007 | 0.940 | 0.015 |
| | | MLE | 0.993 | 0.005 | 0.515 | 0.042 | 0.973 | 0.010 | 0.856 | 0.023 |

In addition, the maximum likelihood estimation (MLE) method is inferior to the cuckoo search optimization (CSO) algorithm in terms of the resulting goodness of fit. This result implies that MLE cannot make full use of the information from the large samples, whereas the heuristic optimization algorithm should be properly utilized to search for the optimal parameter values of the distributions.

***Forecast results***

Compared with deterministic forecasts, interval forecasts can properly overcome the uncertainty problems associated with pollutant forecasts. However, obtaining the high quality of forecast intervals depends on better deterministic forecasts. Thus, the accuracy of deterministic forecasts generated by a single model should be improved using other techniques or methods. In this paper, the singular spectrum analysis (SSA) algorithm is used to improve the performance of BP neural networks.

### 1.1.2 *Deterministic forecast results*

In this paper, the SSABPNN and BPNN models with 12 inputs, 25 hidden and 6 output layer nodes for the six air pollutants are built. In addition, the maximum training time is 500, the learning velocity 0.1, and the required training precision is 0.00004. Necessarily, the separately constructed BP neural network serves as the standard to verify the effectiveness of the SSA algorithm. Meanwhile, to further verify the prediction performance, ARIMA and support vector machine (SVM) models were established for use in the comparison. **Table 8** presents the resulting deterministic forecast for Chengdu and Hangzhou. When comparing the total MAPE values generated by the SSABPNN and BPNN models, the results for Chengdu show that BPNN yielded the largest MAPE value of 23.066% for $NO_2$, while SSABPNN has the largest MAPE value of 8.870% for $PM_{10}$. Considering the MSE values, the results of forecast for the same pollutant should be regarded as the reference to evaluate the two models because of the different concentrations associated with the different pollutants. From the MSE values presented in **Table 8**, it can be concluded that SSABPNN outperforms BPNN. For example, the total MSE value for $O_3$ shows that the MSE values of 5.5285 and 92.405 are obtained by SSABPNN and BPNN, respectively. It is quite obvious that the SSA algorithm can improve the forecasting performance of BPNN significantly. Analysis of the other pollutants and the results for Hangzhou lead to similar conclusions. More importantly, from Table 8, it can be seen that SSABPNN obtains lower MAPE values compared with ARIMA and SVM, which means that the developed model can generate effective and reliable results. To facilitate the discussion of the results, the following analyses focus on the results generated by SSABPNN.

**Table 8**

**Forecast results.**

| Model | Time | $PM_{2.5}$ | | $PM_{10}$ | | $SO_2$ | | $NO_2$ | | $O_3$ | | CO | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chengdu** | | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE |
| ARIMA | Sept | 6.799 | 8.132 | 5.647 | 20.127 | 13.232 | 8.421 | 50.199 | 865.580 | 52.010 | 1393.887 | 26.665 | 0.163 |
| | Oct | 4.541 | 11.629 | 3.888 | 28.290 | 10.125 | 5.765 | 43.620 | 1010.103 | 56.873 | 870.798 | 29.489 | 0.205 |
| | Total | 5.633 | 9.937 | 4.739 | 24.342 | 11.628 | 7.050 | 46.802 | 940.199 | 54.521 | 1123.810 | 28.123 | 0.185 |
| SVM | Sept | 7.109 | 6.626 | 7.858 | 22.099 | 8.360 | 2.447 | 23.200 | 140.445 | 18.488 | 126.351 | 11.749 | 0.028 |
| | Oct | 4.524 | 10.443 | 3.918 | 27.087 | 6.548 | 2.303 | 21.213 | 190.567 | 23.807 | 109.135 | 17.690 | 0.059 |
| | Total | 5.775 | 8.597 | 5.824 | 24.674 | 7.425 | 2.373 | 22.174 | 166.323 | 21.234 | 117.462 | 14.816 | 0.044 |
| BPNN | Sept | 5.519 | 5.306 | 5.238 | 16.199 | 8.123 | 2.423 | 24.381 | 140.662 | 15.746 | 93.464 | 11.922 | 0.027 |
| | Oct | 4.039 | 9.265 | 3.475 | 23.562 | 6.429 | 2.160 | 21.840 | 198.355 | 21.136 | 91.419 | 16.871 | 0.047 |
| | Total | 4.753 | 7.356 | 4.325 | 20.010 | 7.246 | 2.287 | 23.066 | 170.529 | 18.537 | 92.405 | 14.484 | 0.037 |
| SSABPNN | Sept | 1.714 | 0.803 | 1.504 | 1.943 | 3.192 | 0.333 | 9.304 | 25.346 | 5.395 | 5.255 | 5.111 | 0.006 |
| | Oct | 1.170 | 0.675 | 0.997 | 1.328 | 2.283 | 0.221 | 8.466 | 34.394 | 6.161 | 5.313 | 5.992 | 0.006 |
| | Total | 1.432 | 0.737 | 1.242 | 1.625 | 2.721 | 0.275 | 8.870 | 30.030 | 5.791 | 5.285 | 5.567 | 0.006 |
| **Hangzhou** | | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE |
| ARIMA | Sept | 4.247 | 7.887 | 4.095 | 15.165 | 10.096 | 4.415 | 45.008 | 490.582 | 39.991 | 1693.641 | 20.488 | 0.057 |
| | Oct | 4.387 | 9.345 | 3.914 | 21.545 | 10.445 | 9.205 | 41.900 | 704.203 | 37.452 | 1150.612 | 21.991 | 0.070 |
| | Total | 4.319 | 8.639 | 4.001 | 18.459 | 10.276 | 6.888 | 43.403 | 600.877 | 38.680 | 1413.268 | 21.264 | 0.064 |
| SVM | Sept | 4.110 | 6.685 | 3.774 | 15.020 | 6.534 | 1.856 | 25.283 | 120.393 | 13.818 | 247.244 | 11.027 | 0.013 |
| | Oct | 3.502 | 5.332 | 3.243 | 13.453 | 7.075 | 6.279 | 24.542 | 209.787 | 15.428 | 160.536 | 12.626 | 0.021 |
| | Total | 3.796 | 5.987 | 3.500 | 14.211 | 6.813 | 4.140 | 24.901 | 166.548 | 14.649 | 202.476 | 11.852 | 0.017 |
| BPNN | Sept | 3.765 | 6.348 | 3.603 | 14.283 | 6.514 | 1.843 | 26.881 | 129.084 | 14.014 | 204.959 | 10.819 | 0.013 |
| | Oct | 3.249 | 4.913 | 3.116 | 13.427 | 7.106 | 6.031 | 24.808 | 203.544 | 14.384 | 131.294 | 12.545 | 0.020 |
| | Total | 3.498 | 5.605 | 3.351 | 13.839 | 6.820 | 4.012 | 25.808 | 167.656 | 14.205 | 166.799 | 11.713 | 0.016 |
| SSABPNN | Sept | 1.371 | 1.153 | 1.186 | 2.096 | 2.243 | 0.217 | 8.681 | 17.597 | 4.249 | 12.906 | 4.364 | 0.002 |
| | Oct | 1.147 | 0.608 | 1.033 | 0.990 | 2.675 | 0.582 | 8.893 | 27.874 | 4.546 | 8.444 | 5.227 | 0.004 |
| | Total | 1.255 | 0.871 | 1.106 | 1.523 | 2.467 | 0.406 | 8.791 | 22.921 | 4.403 | 10.595 | 4.811 | 0.003 |

Considering the total MAPE values for Chengdu and Hangzhou, it can be seen that all of the MAPE values for the six air pollutants in Hangzhou, specifically 1.255% ($PM_{2.5}$), 1,106% ($PM_{10}$), 2.467% ($SO_2$), 8.791% ($NO_2$), 4.403% ($O_3$) and 4.811% (CO), are smaller than those of Chengdu. However, when considering the total MSE values, smaller MSE values are obtained for Chengdu for $PM_{2.5,}$ $SO_2$ and $O_3$ (0.737, 0.275 and 5.285, respectively). This result indicates that multiple metrics, instead of individual evaluation indices, should be used to evaluate the performance of forecasts comprehensively, so that problematic uncertainties are reduced in decision-making. In the real world, due to the existence of positive relationship between $PM_{2.5}$ and $PM_{10}$, their forecasting results perform similarity. In addition, the nature of moving average calculations seems to make the concentrations of the pollutants more regular, resulting in the relatively low MAPE value obtained. For example, the total MAPE of Chengdu for $PM_{2.5}$, $PM_{10}$ and $SO_2$ are the relatively low values of 1.432%, 1.242% and 2.721%, respectively. When comparing the results for September and October, it is obvious that the relatively good performance for $PM_{2.5}$, $PM_{10}$, $O_3$ and CO is obtained in October and September, in terms of the MAPE values. Meanwhile, the $PM_{2.5}$, $PM_{10}$, $SO_2$ and $NO_2$ in Chengdu have lower MAPE values in September, and the $SO_2$, $NO_2$, $O_3$ and CO in Hangzhou have the lower MAPE values in October, indicating that air pollutant values can reflect significant regularity, regardless of geographic location or period is considered.

To clearly express the forecasting effectiveness, **Fig. 7** and **Fig. 8** depict the relationships between the actual data and the predictions for the two cities. Due to the impact of the number of observed target points, the forecasting values are sorted by their actuals value before plotting. Clearly, the better forecast points agree well with the actual data, while the points representing worse predictions lie at a larger distance from the actual points. For example, **Fig. 8** (**d**, **f**) shows that $NO_2$ and CO in Hangzhou, which are associated with the MAPE values of 8.819% and 4.8%, have worse forecasting performance.

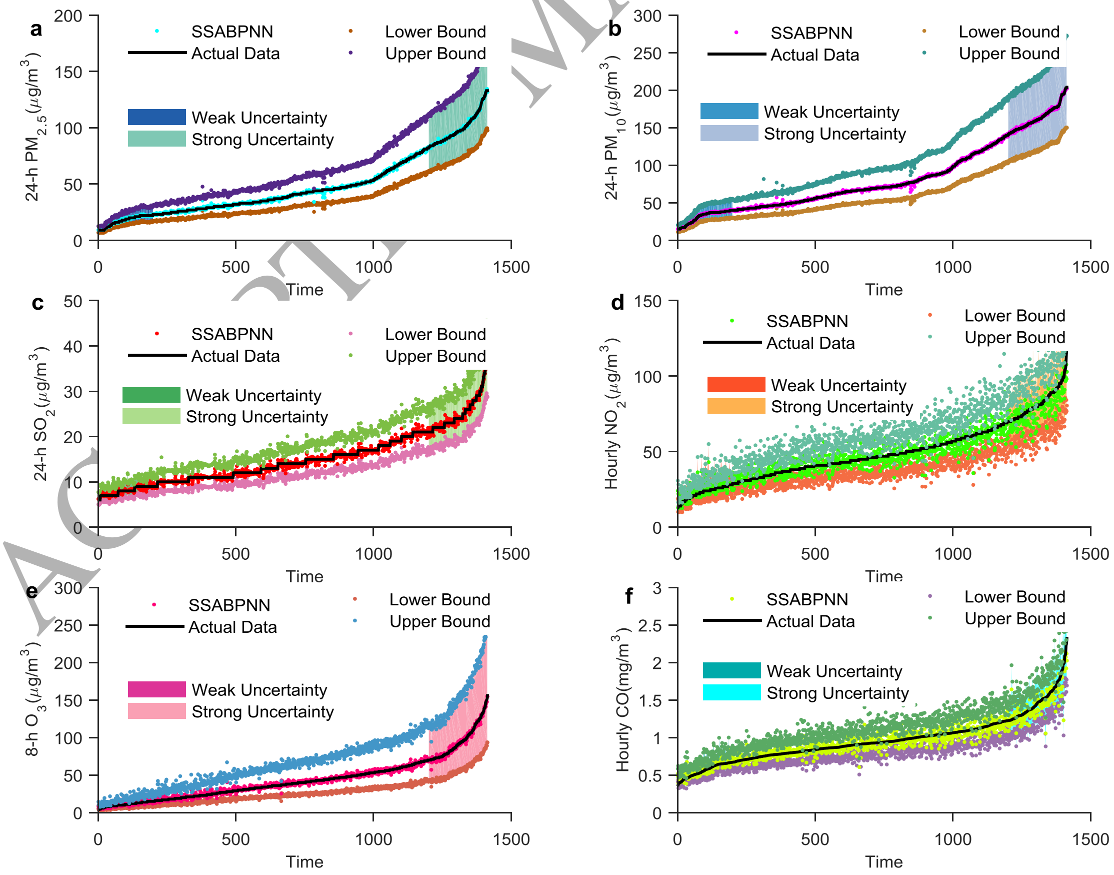

**Fig. 7. Interval forecast results for Chengdu.**

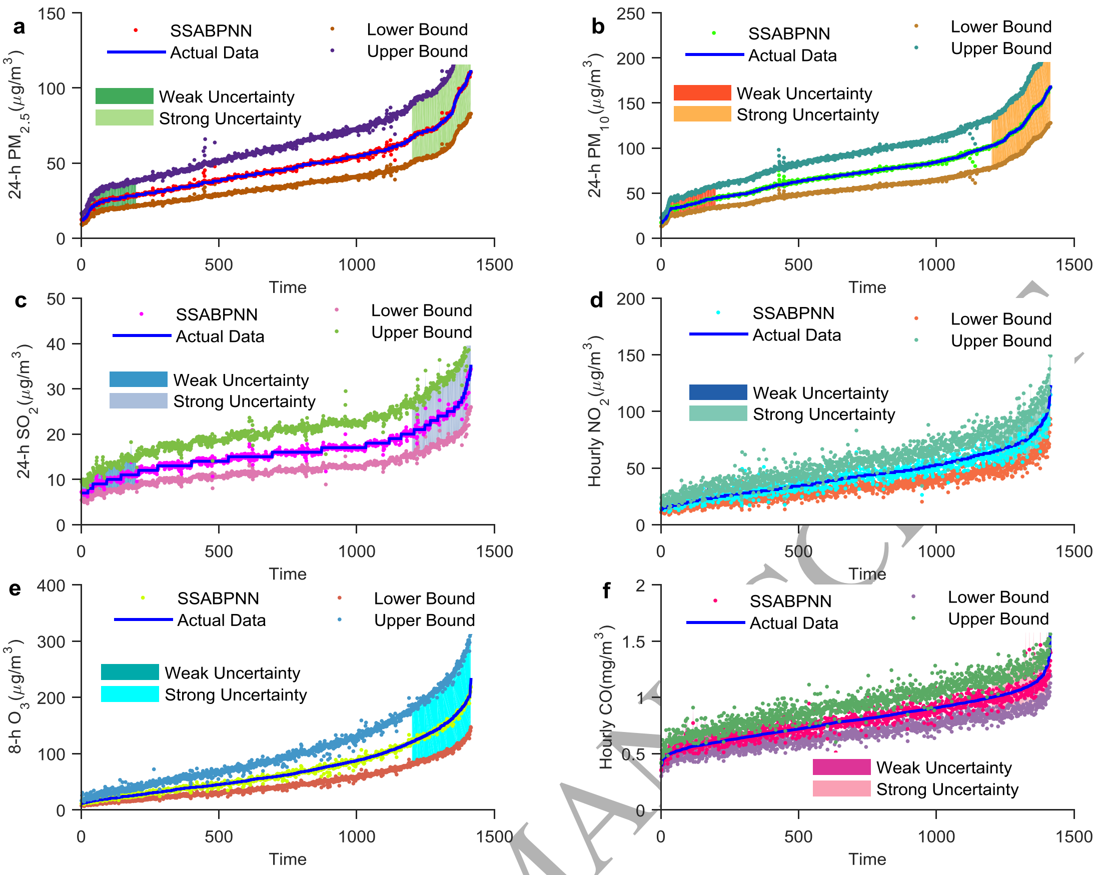

**Fig. 8. Interval forecast results for Hangzhou.**

## Table 9

Interval forecasting results for different significance levels.

| Environmental Parameter | Significance | Chengdu | | | | | | Hangzhou | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Sept | | Oct | | Total | | Sept | | Oct | | Total | |
| **$PM_{2.5}$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 1.00 | 33.97 | 1.00 | 60.70 | 1.00 | 47.81 | 1.00 | 39.89 | 1.00 | 50.89 | 1.00 | 45.59 |
| | 0.25 | 1.00 | 26.86 | 1.00 | 47.99 | 1.00 | 37.80 | 1.00 | 31.57 | 1.00 | 40.27 | 1.00 | 36.08 |
| | 0.3 | 1.00 | 20.68 | 1.00 | 36.95 | 1.00 | 29.10 | 1.00 | 24.33 | 1.00 | 31.03 | 1.00 | 27.80 |
| | 0.35 | 1.00 | 15.09 | 1.00 | 26.96 | 1.00 | 21.24 | 1.00 | 17.76 | 1.00 | 22.66 | 1.00 | 20.30 |
| | 0.4 | 0.99 | 9.88 | 1.00 | 17.65 | 1.00 | 13.90 | 0.99 | 11.63 | 1.00 | 14.83 | 1.00 | 13.29 |
| **$PM_{10}$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 1.00 | 57.13 | 1.00 | 103.70 | 1.00 | 81.24 | 1.00 | 56.89 | 1.00 | 75.87 | 1.00 | 66.73 |
| | 0.25 | 1.00 | 45.17 | 1.00 | 82.00 | 1.00 | 64.24 | 1.00 | 45.10 | 1.00 | 60.14 | 1.00 | 52.89 |
| | 0.3 | 1.00 | 34.79 | 1.00 | 63.14 | 1.00 | 49.46 | 1.00 | 34.79 | 1.00 | 46.39 | 1.00 | 40.80 |
| | 0.35 | 1.00 | 25.39 | 1.00 | 46.08 | 1.00 | 36.10 | 1.00 | 25.42 | 1.00 | 33.90 | 1.00 | 29.81 |
| | 0.4 | 1.00 | 16.62 | 1.00 | 30.16 | 1.00 | 23.63 | 1.00 | 16.65 | 1.00 | 22.21 | 1.00 | 19.53 |
| **$SO_2$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 1.00 | 9.31 | 1.00 | 11.92 | 1.00 | 10.66 | 1.00 | 13.76 | 1.00 | 16.01 | 1.00 | 14.92 |
| | 0.25 | 1.00 | 7.41 | 1.00 | 9.48 | 1.00 | 8.48 | 1.00 | 10.90 | 1.00 | 12.68 | 1.00 | 11.82 |
| | 0.3 | 1.00 | 5.73 | 1.00 | 7.34 | 1.00 | 6.56 | 1.00 | 8.40 | 1.00 | 9.77 | 1.00 | 9.11 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.35 | 0.99 | 4.20 | 1.00 | 5.37 | 0.99 | 4.80 | 1.00 | 6.14 | 0.99 | 7.14 | 0.99 | 6.65 |
| | 0.4 | 0.95 | 2.75 | 0.99 | 3.52 | 0.97 | 3.15 | 0.99 | 4.02 | 0.98 | 4.67 | 0.98 | 4.36 |
| **$NO_2$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 0.99 | 34.80 | 1.00 | 43.44 | 0.99 | 39.27 | 1.00 | 33.98 | 0.99 | 41.90 | 0.99 | 38.08 |
| | 0.25 | 0.98 | 27.63 | 0.98 | 34.49 | 0.98 | 31.18 | 0.99 | 26.95 | 0.98 | 33.23 | 0.98 | 30.20 |
| | 0.3 | 0.95 | 21.34 | 0.96 | 26.64 | 0.96 | 24.09 | 0.97 | 20.80 | 0.95 | 25.65 | 0.96 | 23.31 |
| | 0.35 | 0.89 | 15.61 | 0.90 | 19.49 | 0.89 | 17.62 | 0.90 | 15.20 | 0.89 | 18.75 | 0.90 | 17.04 |
| | 0.4 | 0.73 | 10.23 | 0.75 | 12.77 | 0.74 | 11.55 | 0.77 | 9.96 | 0.77 | 12.28 | 0.77 | 11.16 |
| **$O_3$** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 1.00 | 81.16 | 1.00 | 76.74 | 1.00 | 78.87 | 1.00 | 107.15 | 1.00 | 87.71 | 1.00 | 97.08 |
| | 0.25 | 1.00 | 62.64 | 1.00 | 59.22 | 1.00 | 60.87 | 1.00 | 83.84 | 1.00 | 68.63 | 1.00 | 75.96 |
| | 0.3 | 1.00 | 47.40 | 1.00 | 44.81 | 1.00 | 46.06 | 1.00 | 64.08 | 1.00 | 52.46 | 1.00 | 58.06 |
| | 0.35 | 0.99 | 34.16 | 0.99 | 32.30 | 0.99 | 33.20 | 0.99 | 46.52 | 0.99 | 38.08 | 0.99 | 42.15 |
| | 0.4 | 0.97 | 22.17 | 0.97 | 20.97 | 0.97 | 21.55 | 0.98 | 30.34 | 0.96 | 24.84 | 0.97 | 27.49 |
| **CO** | α | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW | ICP | IAW |
| | 0.2 | 1.00 | 0.56 | 0.99 | 0.57 | 1.00 | 0.56 | 1.00 | 0.42 | 1.00 | 0.44 | 1.00 | 0.43 |
| | 0.25 | 0.99 | 0.45 | 0.98 | 0.45 | 0.99 | 0.45 | 1.00 | 0.34 | 0.98 | 0.35 | 0.99 | 0.34 |
| | 0.3 | 0.98 | 0.35 | 0.96 | 0.35 | 0.97 | 0.35 | 0.99 | 0.26 | 0.95 | 0.27 | 0.97 | 0.27 |
| | 0.35 | 0.95 | 0.25 | 0.90 | 0.26 | 0.92 | 0.26 | 0.96 | 0.19 | 0.90 | 0.20 | 0.93 | 0.20 |
| | 0.4 | 0.82 | 0.17 | 0.77 | 0.17 | 0.79 | 0.17 | 0.86 | 0.13 | 0.80 | 0.13 | 0.82 | 0.13 |

### 1.1.3 *Interval forecast results*

Forecast intervals in this paper are applied to air quality for handling uncertainties and reducing their negative impact on the decision-making process. In this sense, appropriate and reliable decisions are based on high-quality of forecast intervals. Therefore, the interval coverage probability (ICP) and average width (IAW) are used to evaluate the performance of interval forecasting. Another important aspect, forecast interval validity, is also discussed in this work. Theoretically, if the ICP for a significance level ($\alpha$) is greater than or equal to the confidence level (100(1-$\alpha$) %), the corresponding forecast interval is satisfactory. However, there are many confidence levels corresponding to a given ICP, given that a smaller confidence level reflects the expectation that the narrower averaged width has a higher probability of containing the true value. To numerically demonstrate the analysis, forecast intervals for different confidence levels are given in this section.

In this work, the estimated distributions of different air pollutants were discussed in detail in section **4.1**, where we hypothesized that the distribution of forecast points is similar to that of the historical data. Forecast intervals can be obtained when deterministic forecasting points are obtained. According to the results in **Table** 7, the lognormal distribution function has the best performance in terms of the general estimated results. Therefore, the parameters of the lognormal distribution are utilized to estimate the forecast intervals in this section.

**Table 9** shows the interval forecasting results for different confidence levels. Considering the validity of forecast intervals, it is obvious that all of the ICPs are valid. Further, comparison of the results in **Table 9** and **Table 8** shows that the better deterministic forecasting can result in forecast interval with higher quality. For example, when the confidence level is 75%, the $PM_{2.5}$, $NO_2$ and CO in Chengdu, which are associated with total MAPE value of 1.432%, 8.870% and 5.791%, respectively, have ICP values of 1, 0.98 and 0.99, respectively. In addition, considering different confidence levels, it can be seen that the ICP value decreases as the value of the confidence level decreases. For example, the ICP of CO in Hangzhou yields the values of 1, 0.97 and 0.82 for the confidence level 80%, 70% and 60%, respectively. However, $PM_{2.5}$ and $PM_{10}$ in both Chengdu and Hangzhou yield an ICP value of 100% when either the total MAPE values of the deterministic forecasts or the confidence levels are considered, which implies that a single ICP value fails to assess the quality of forecast intervals. Therefore, IAW should be applied for evaluating the quality of forecast intervals. Comparing IAW among different confidence levels, it can be clearly found that the IAW decreases as the value of the significance level increases.

As mentioned above, there are many forecast intervals that correspond to different confidence levels. As an illustrative example, the deterministic forecast points, the actual data and the forecast intervals corresponding to a confidence level of 70% are shown in **Fig. 7** and **Fig. 8**. **Fig. 7** and **Fig. 8** show the effectiveness of the interval forecast, along with the evaluation criteria (ICP and IAW). This can be found in comparisons among the results shown in the two figures. In addition, it can be seen that the small and large points have narrow and wide forecast intervals, which indicates that the uncertainty tends to increase as the real value increases. However, outliers always occur as small or large points within data series in the real world, leading to reduction in forecast accuracy. This can be seen in **Fig. 7** and **Fig. 8**.

In general, interval forecast methods are applied for handling uncertainties, and then the ICP and IAW are used as criteria to judge the performance of the interval forecasting. For a given ICP value, an interval with a smaller IAW should be chosen, whereas a larger ICP is preferred when a smaller IAW is considered. In this paper, a significance level of 0.3 is used to construct intervals for air quality assessment.

### 1.1.4 *Air quality assessment results*

In this section, the air quality of the Chengdu and Hangzhou cities will be assessed using the proposed assessment methods and deterministic and interval forecast results covering the period from September 1, 2015 to October 31, 2015. To clearly describe the reliability and effectiveness of the assessment results, the data from October 21, 2015 to October 30, 2015 are randomly selected as an illustrative case. Meanwhile, the assessment results at 8:00 AM daily for Chengdu and 12:00 PM for Hangzhou are randomly selected to facilitate the analysis and discussion of the proposed methods. In addition, it is necessary to note that the assessment results are based on the forecasting results.

The final assessment results for Chengdu and Hangzhou are shown in **Table 10** and **Table 11**. Meanwhile, for the purposes of comparison, the results based on actual data are also shown in the corresponding tables. According to these tables, it can be seen that the assessments based on deterministic forecasts and the actual values yield identical results, indicating that the assessment results based on the forecasting results are effective and reliable and provide trustworthy reference data. However, as previously mentioned, the deterministic assessment could not address the uncertainty in future air quality because of the deficiency of deterministic forecasts. To overcome this problem, assessment based on lower and upper forecasting bounds is performed in this work. According to **Table 10** and **Table 11**, it can be seen that some of the levels based on upper bounds are substantially higher than those based deterministic forecasts, which implies that air quality conditions will be worse in probabilistic sense. For example, the result for October 21, 2015 in Chengdu shows that the assessment based on deterministic forecasts results in level Ⅲ, indicating the *"Lightly polluted"* conditions, whereas that based on the upper bound results in level Ⅳ, implying the "*Moderately polluted*". The upper bound thus provides a trustworthy reference to help develop effective prevention programs to avoid potential risks from air pollution. In general, the proposed method can obtain effective and reliable estimates of future air quality conditions and should be used as a tool for air pollution monitoring and providing advance warnings of adverse pollution conditions.

**Table 10**

Assessment results for Chengdu.

| **Time** | 21/10/15 | 22/10/15 | 23/10/15 | 24/10/15 | 25/10/15 | 26/10/15 | 27/10/15 | 28/10/15 | 29/10/15 | 30/10/15 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Actual** | Ⅲ | Ⅲ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅱ | Ⅰ |
| Levels | **Assessment based on lower bound** | | | | | | | | | |
| Ⅰ | 0.308 | 0.308 | 0.788 | 0.715 | 0.719 | 0.721 | 0.993 | 0.754 | 0.722 | 0.782 |
| Ⅱ | 0.461 | 0.588 | 0.212 | 0.285 | 0.281 | 0.279 | 0.007 | 0.246 | 0.278 | 0.218 |
| Ⅲ | 0.078 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅳ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Results** | Ⅱ | Ⅱ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ |
| Levels | **Assessment based on upper bound** | | | | | | | | | |
| Ⅰ | 0.286 | 0.285 | 0.278 | 0.270 | 0.305 | 0.304 | 0.835 | 0.340 | 0.305 | 0.351 |
| Ⅱ | 0.022 | 0.121 | 0.598 | 0.579 | 0.590 | 0.606 | 0.165 | 0.643 | 0.661 | 0.627 |
| Ⅲ | 0.268 | 0.392 | 0 | 0 | 0 | 0 | 0 | 0.017 | 0.016 | 0 |
| Ⅳ | 0.424 | 0.203 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Results | Ⅳ | Ⅲ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅰ | Ⅱ | Ⅱ | Ⅱ |
| Levels | **Assessment based on deterministic forecast** | | | | | | | | | |
| Ⅰ | 0.302 | 0.308 | 0.617 | 0.520 | 0.525 | 0.528 | 0.951 | 0.467 | 0.426 | 0.610 |
| Ⅱ | 0.126 | 0.323 | 0.383 | 0.480 | 0.475 | 0.472 | 0.049 | 0.430 | 0.472 | 0.390 |
| Ⅲ | 0.542 | 0.369 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅳ | 0.029 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Results** | Ⅲ | Ⅲ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅰ | Ⅱ | Ⅰ |

**Table 11**

Assessment results for Hangzhou.

| **Time** | 21/10/15 | 22/10/15 | 23/10/15 | 24/10/15 | 25/10/15 | 26/10/15 | 27/10/15 | 28/10/15 | 29/10/15 | 30/10/15 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Actual** | Ⅰ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅲ | Ⅰ |
| Levels | **Assessment based on lower bound** | | | | | | | | | |
| Ⅰ | 0.743 | 0.626 | 0.308 | 0.377 | 0.685 | 0.614 | 0.651 | 0.421 | 0.306 | 0.745 |
| Ⅱ | 0.257 | 0.374 | 0.593 | 0.508 | 0.315 | 0.386 | 0.349 | 0.471 | 0.609 | 0.255 |
| Ⅲ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.033 | 0 |
| Ⅳ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Results** | Ⅰ | Ⅰ | Ⅱ | Ⅱ | Ⅰ | Ⅰ | Ⅰ | Ⅱ | Ⅱ | Ⅰ |
| Levels | **Assessment based on upper bound** | | | | | | | | | |
| Ⅰ | 0.297 | 0.287 | 0.298 | 0.303 | 0.267 | 0.266 | 0.264 | 0.295 | 0.257 | 0.300 |
| Ⅱ | 0.612 | 0.557 | 0.206 | 0.344 | 0.645 | 0.560 | 0.619 | 0.416 | 0.094 | 0.603 |
| Ⅲ | 0 | 0.010 | 0.425 | 0.353 | 0 | 0.019 | 0.002 | 0.289 | 0.525 | 0 |
| Ⅳ | 0 | 0 | 0.071 | 0 | 0 | 0 | 0 | 0 | 0.124 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Results** | Ⅱ | Ⅱ | Ⅲ | Ⅲ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅲ | Ⅱ |
| Levels | **Assessment based on deterministic forecast** | | | | | | | | | |
| Ⅰ | 0.571 | 0.318 | 0.307 | 0.308 | 0.491 | 0.306 | 0.337 | 0.306 | 0.288 | 0.574 |
| Ⅱ | 0.429 | 0.584 | 0.470 | 0.612 | 0.509 | 0.604 | 0.556 | 0.676 | 0.334 | 0.426 |
| Ⅲ | 0 | 0 | 0.223 | 0 | 0 | 0 | 0 | 0 | 0.378 | 0 |
| Ⅳ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅴ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ⅵ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Results** | Ⅰ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅱ | Ⅲ | Ⅰ |

### *Conclusion and discussion*

This work developed an air quality early-warning system that can forecast future air pollutant concentrations using the combination of a back propagation neural network algorithm, an optimal parameter distribution model and data preprocessing techniques. Moreover, air quality can be evaluated by using fuzzy sets and the AHP. For a better understanding of air pollutant behavior, the characteristics of six pollutants were examined using four distribution functions. It was found that the lognormal or gamma distributions yielded the best fits to the data. Additionally, comparisons of the results between two cities showed that the air pollutants displayed different behaviors because of the different geographic locations. After estimating the parameters of the distribution functions, forecasting of air pollutant concentrations was implemented using the proposed methods. The process of forecasting was carried out in several phrases: first, deterministic forecasting was performed by using the single BPNN, ARIMA, SVM and SSABPNN models and the forecasting performance of SSABPNN was verified based on a comparison of the results. Second, interval forecast with different confidence levels were constructed using the estimated parameters of the probability distributions and deterministic forecasting results. The interval coverage probabilities (ICPs) and interval average widths (IAWs) showed that the forecast intervals were valid. Based on the forecasting results, we finally assessed the air quality conditions of Chengdu and Hangzhou, using a combination of fuzzy set theory and the Analytic Hierarchy Process. The analysis and discussion of the assessment results demonstrated that our methods are reliable and suitable for use by environmental supervisors in air pollution monitoring and management.

The results of the case in this paper verify the effectiveness of the developed early-warning system. Meanwhile, the main contributions provide important implications for the air pollution monitoring and management.

When examining and analyzing the statistical characteristic of air pollutant emissions, four unimodal distribution models are used to confirm pollutant emissions regimes. In the **Table 7**, the estimated distributions show significant difference on the performance evaluation indices. This is not only because of different pollutants, but also geographic locations. For different cities and pollutants, due to the process of industrialization and urbanization, the emission regimes may differ statistically. Therefore, different distribution models and techniques should be used to facilitate the analysis of statistical characteristics of pollutant emissions.

The developed early-warning system provides the prediction and assessment modules for future air quality conditions. First, the prediction for air pollutant concentrations seems to be the focus of many studies. In the real world, air pollutant concentrations series always show the complex nonlinearity, adding the uncertainty to the forecasting methods and showing the necessity of developing air quality early-warning. In this paper, this problem was effectively alleviated by using data preprocessing and interval forecasting, implying that data processing techniques and probabilistic methods should be used to overcome the uncertainty in future air quality. Second, the assessment module based on the forecasting results is performed to provide much more information about future air quality and assessment results focus on the intelligible description of air pollution. This is not only important for monitoring air pollution, but also guiding the daily activities of people.

In summary, the early-warning system presented in this paper consists of prediction and assessment modules, and its effectiveness is verified by performing a case study. Accordingly, our early-warning system can produce more reasonable and comprehensive analyses of air pollution, thus providing trustworthy reference data for use by environmental supervisors in air pollution monitoring and management and providing the public with more information about the negative effect of air pollution. Moreover, the developed early-warning system can be used as a monitoring tool in severely polluted areas to avoid the potential risks produced by the pollutants discharged from industrial sites and vehicle emissions. In addition, the pollutant classification levels used in early-warning systems should be modified to be suitable for different countries, such as China, and it is hoped that they can then guide and improve the daily lives of people.

***Limitations and future work***

While the developed early-warning can generate the desired results in term of the estimation of future air quality conditions, this study also includes several limitations. First, we consider only six air pollutants and ignore other environmental parameters that may affect air quality conditions. Second, only pollutant concentrations over time are considered, and other factors correlated with forecasting effectiveness are not taken into account.

Air quality has received global attention because of its important role in human daily activities. Therefore, developing reliable and effective air quality methods for air quality forecasting and assessment is extremely urgent in air pollution monitoring and management. Based on the work presented in this paper, we propose four future directions for further research.

- Pollutant concentrations time series show complex nonlinear characteristics. Developing a more precise hybrid model would be very useful for improving the early-warning system.
- High pollutant concentrations are the key factors that result in potential risks. Therefore, developing a model for forecasting peak air pollutant concentrations is necessary and significant.
- This paper adopts symmetrical probabilities to address the uncertainties in predicting air quality conditions. In addition, nonparametric methods can be employed to construct prediction intervals in further research.
- Different pollutants could cause different problems in various sensitive groups. Therefore, based on medical knowledge, more detailed work, such as early-warning systems for certain pollutants, should be performed to a greater degree to guide the daily activities of specific sensitive groups.

***Acknowledgment***

This study was supported by the National Natural Science Foundation of China (Grants No. 41475013).

***References***

Samia, A., Kaouther, N., Abdelwahed, T., (2012). A hybrid ARIMA and artificial neural networks model to forecast air quality in urban areas: case of Tunisia. Adv. Mater. Res. 518-523, 2969-2979.

Stadlober, E., H€ormann, S., Pfeiler, B., (2008). Quality and performance of a $PM_{10}$ daily forecasting model. Atmos. Environ. 42 (6), 1098-1109.

Akyüz, M., Çabuk, H., (2009). Meteorological variations of $PM_{2.5}/PM_{10}$ concentrations and particle-associated polycyclic aromatic hydrocarbons in the atmospheric environment of Zonguldak, Turkey. J. Hazard. Mater. 170(1), 13-21.

Genc, D.D., Yesilyurt, C., Tuncel, G., 2010. Air pollution forecasting in Ankara, Turkey using air pollution index and its relation to assimilative capacity of the atmosphere. Environ. Monit. Assess. 166, 11-27.

Song, Y., Qin, S., Qu, J., Liu, F., (2015). The forecasting research of early warning systems for atmospheric pollutants: a case in Yangtze River Delta region. Atmos. Environ. 118, 58-69.

Feng, X., Li, Q., Zhu, Y., Hou, J., Jin, L., Wang, J., 2015. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmos. Environ. 107, 118-128.

Pauzi, H.M., Abdullah, L., 2015. Neural network training algorithm for carbon dioxide emissions forecast: a performance comparison. Lect. Notes Electr. Eng. 315, 717-726.

Perez P, Gramsch E. (2016). Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes[J]. Atmospheric Environment, 124:22-27.

Biancofiore, F., et al., (2017). Recursive neural network model for analysis and forecast of PM10 and PM2.5, Atmospheric Pollution Research, http://dx.doi.org/10.1016/j.apr.2016.12.014.

D. Domańska, & M. Wojtylak. (2012). Application of fuzzy time series models for forecasting pollution concentrations. Expert Systems with Applications, 39(9), 7673-7679.

Osowski, S., Garanty, K., 2007. Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng. Appl. Artif. Intell. 20, 745-755.

Pai, T.Y., Hanaki, K., Chiou, R.J., 2013a. Forecasting hourly roadside particulate matter in Taipei county of Taiwan based on first-order and one-variable grey model. Clean e Soil Air Water 41, 737-742.

Dong, M., Yang, D., Kuang, Y., He, D., Erdal, S., & Kenski, D. (2009). Pm 2.5, concentration prediction using hidden semi-markov model-based times series data mining. Expert Systems with Applications, 36(5), 9046-9055.

Sun, W., Zhang, H., Palazoglu, A., Singh, A., Zhang, W., Liu, S., (2013). Prediction of 24-hour-average PM2.5 concentrations using a hidden Markov model with different mission distributions in Northern California. Sci. Total Environ. 443, 93-03.

Taylan, O. (2016). Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality. Atmospheric Environment.

Bancha Ariyajunya, Ying Chen, Victoria C.P. Chen, & Seoung Bum Kim. (2017). Data mining for state space orthogonalization in adaptive dynamic programming. Expert Systems With Applications, 76, 49-58.

Feng, Q., Wu, S., Du, Y., Xue, H., Xiao, F., Ban, X., Li, X., (2013). Improving neural network prediction accuracy for pm10 individual air quality index pollution levels. Environ. Eng. Sci. 30(12), 725–732.

Mishra, D., Goyal, P., (2016). Neuro-fuzzy approach to forecast $NO_2$ pollutants addressed to air quality dispersion model over Delhi, India. Aerosol Air Qual. Res. 16, 166–174.

Yong, L., Huaicheng, G., Guozhu, M., Pingjian, Y., (2008). A Bayesian hierarchical model for urban air quality prediction under uncertainty. Atmos. Environ. 42, 8464–8469.

Liu, K., Liang, H., Yeh, K., Chen, C., (2009). A qualitative decision support for environmental impact assessment using fuzzy logic. J. Environ. Inf. 13(2), 93–103.

Sowlat, M., Gharibi, H., Yunesian, M.,Mahmoudi, T., Lotfi, S., (2011). A novel, fuzzy-based air quality index (FAQI) for air quality assessment. Atmos. Environ. 45, 2050–2059.

Yadav, J., Kharat, V., & Deshpande, A. (2014). Fuzzy description of air quality using fuzzy inference system with degree of match via computing with words: a case study. Air Quality, Atmosphere & Health, 7(3), 325-334.

Sen, A., Lal, B., Tripathy, S.S., (2015). Determination of Air Quality Index Using Fuzzy Logic-Based Model. Proceedings of the International Conference on Electrical, Electronics, Signals, Communication and Optimization, IEEE, 1-4.

Xu Y, Yang W, Wang J, (2017) Air quality early-warning system for cities in China[J]. Atmospheric Environment, 148, 239–257.

Upadhyaya, G., Dashore, N., 2011. Fuzzy logic based model for monitoring air quality index. Indian J. Sci. Technol. 4 (3), 15–218.

Akkaya, G., Turanoğlu, B., Öztaş, S., (2015). An integrated fuzzy AHP and fuzzy MOORA approach to the problem of industrial engineering sector choosing. Expert Systems with Applications 42 (24), 9565–9573.

Saaty, T.L. (1980). The Analytic Hierarchy Process, McGraw Hill International Publication.

Saaty, T.L. (1994). How to Make a Decision: The Analytic Hierarchy Process. Interfaces. 24: 19–43.

Upadhyay, A., Kanchan, Goyal, P., Yerramilli, Gorai, A., (2014). Development of a fuzzy pattern recognition model for air quality assessment of Howrah City. Aerosol Air Qual. Res. 14, 1639–1652.

Olvera, M., Carbajal, J., Sánchez, L., Hernández, I., (2016). Air quality assessment using a weighted Fuzzy Inference System. Ecol. Inform, 33, 57–74.

P. Pinson and G. Kariniotakis, (2010). Conditional prediction intervals of wind power generation. IEEE Trans.Power Syst., 25: 1845–1856.

El-Fouly, T.H.M., El-Saadany, E.F., Salama, M.M.A., (2006). One day ahead prediction of wind speed using annual trends. In: Power Engineering Society General Meeting, 2006. IEEE, p.7.

Wang, J., Qin, S., Jin, S., Wu, J., (2015). Estimation methods review and analysis of offshore extreme wind speeds and wind energy resources. Renew. Sustain. Energy Rev. 42, 26-42.

Wu, J., Wang, J., Chi, D., (2013).Wind energy potential assessment for the site of inner Mongolia in China. Renew. Sustain. Energy Rev. 21, 215-228.

Zhang, X., Wang, J., & Zhang, K., (2017). Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by cuckoo search algorithm. Electric Power Systems Research, 270–285.

Ma, X., Jin, Y., & Dong, Q., (2017). A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. Applied Soft Computing, 296–312.

USEPA, (2006). Guidelines for the Reporting of Daily Air Quality – the Air Quality Index (AQI).

USEPA, (2009). Air Quality Index, A Guide to Air Quality and Your Health.

Carbajal-Hernández, J. J., Sánchez-Fernández, L. P., Carrasco-Ochoa, J. A., & Martínez-Trinidad, J. F. (2012). Assessment and prediction of air quality using fuzzy logic and autoregressive models. Atmospheric Environment, 60(6), 37-50.

Chakraborty, S., Dey, S., (2006). Design of an analytic-hierarchy-process-based expert system for non-traditional machining process selection. Int. J. Adv. Manuf. Technol. 31,490–500

Saaty, T., (2004). Decision making—the analytic hierarchy and network processes (AHP/ANP). J. Syst. Sci. Syst. Eng. 13(1), 1–35.