

Urban Air Pollution Research and Its Cause Analysis Based on Gray Theory

Weiqliang Wang, Xibin Zhang

School of Mathematics and Statistics of Central China Normal University, Wuhan Hubei
Email: wwq2015212826@mails.ccnu.edu.cn, 305170350@qq.com

Received: Mar. 31st, 2018; accepted: Apr. 19th, 2018; published: Apr. 26th, 2018

Abstract

In this paper, a gray prediction GM (1,1) model and a BP neural network model are respectively used to predict the air quality status of all first-tier cities and new-tier cities in China. From the better predictions, six cities that cannot meet the air quality standards set by the 13th Five-Year Plan of China by 2030. They are Beijing, Wuhan, Tianjin, Chongqing, Xi'an and Changsha. Then, the air quality status and air pollutants of the above six cities were analyzed by gray relational analysis. Therefore, we identify the main air pollutants in each city. Finally, the main pollutants are integrated using principal component analysis and multiple linear regression to find out the main causes of the major pollutants in each city and provide an important reference for the environmental governance in each city.

Keywords

Gray Forecast, BP Neural Network, Gray Relational Degree, Principal Component Analysis, Multiple Linear Regression

基于灰色理论下的城市空气污染状况研究及成因分析

王维强, 张席斌

华中师范大学数学与统计学学院, 湖北 武汉
Email: wwq2015212826@mails.ccnu.edu.cn, 305170350@qq.com

收稿日期: 2018年3月31日; 录用日期: 2018年4月19日; 发布日期: 2018年4月26日

摘要

本文先分别运用灰色预测GM(1,1)模型和BP神经网络模型分别对我国所有一线城市和新一线城市的空气

质量状况进行了一个预测, 并对比预测结果, 从更优的预测结果中挑出了到2030年不能达到我国十三五计划规定的空气质量标准的六座城市——北京, 武汉, 天津, 重庆, 西安, 长沙。接着对上述六座城市的空气质量状况与各空气污染物进行了灰色关联度分析, 找出了各城市的主要空气污染物。针对每个城市的主要污染物综合运用了主成分分析和多元线性回归等方法找出了各城市主要污染物的主要成因, 对各城市的环境治理提供了重要的参考。

关键词

灰色预测, BP神经网络, 灰色关联度, 主成分分析, 多元线性回归

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

“十一五”期间, 我国环境保护取得了令人可喜的成绩, 特别是空气污染物这项指标超额完成规划目标。但是我国环境形势任然很严峻, 根据耶鲁大学公布的全球环境绩效指数(EPI)报告显示, 中国 GDP 总量从 2006 年的第四名上升到目前第二名并稳居第二名, 然而环境绩效指数却从 2006 年的第 94 名(参评 133 个国家)下降到 2016 年的倒数第 2 名, 其中空气质量是拖后中国排名的最主要的原因[1] [2]。中国用 40 年的时间基本上完成了发达国家 100 多年的城市化进程, 却也只用 30 多年的时间“集聚”和“爆发”了发达国家 100 多年的环境问题。如果中国解决不好当前我国的环境问题, 不但会严重影响社会经济的可持续发展, 还可能影响到政府执政的公信力。

在国家环保局发布的文件《国家环境保护“十三五”规划基本思路》中就初步提出了 2020 年及 2030 年两个阶段性目标——到 2020 年, PM2.5 超标 30%以内城市必须实现 PM2.5 年均浓度达标。到 2030 年, 要实现全国城市环境空气质量基本达标, 生态环境质量全面改善, 生态文明水平全面提高[2]。

因此, 了解并掌握空气污染物背后的成因, 计算出每种污染物对于空气质量的贡献以及分析各污染物背后成因, 不仅方便城市进行针对性整治, 也能促进每座城市的生态文明建设。进一步, 计算出每座城市能否完成以及该如何调整发展模式才能完成“十三五”规划的目标, 不仅能方便各座城市及时调整城市规划的目标, 还能加快中国环境治理和改善的进程, 具有重大的现实指导意义。

2. 各大城市大气污染物浓度及其趋势预测

2.1. 各大城市大气污染物浓度值

因为大气污染浓度与工业产值, 经济发展具有很重要的关系, 因此本文主要以一线城市包括北京, 上海, 广州, 深圳, 以及新一线城市成都, 杭州, 武汉, 天津, 南京, 重庆, 西安, 长沙, 青岛, 沈阳, 大连, 厦门, 苏州, 宁波和无锡共十九个城市为代表进行研究。

首先从中国空气质量在线平台(<https://www.aqistudy.cn/>)搜集各大城市历史空气质量指数 AQI 以及各大空气污染物的每月数据。但为了保证预测精度, 我们以每季度的平均值进行分析, 因为每年相同季度的各污染物的值的波动范围大致是相同的。每季度的数据如表 1 所示(以北京市为例, 其余见附件一):

Table 1. The concentration of air pollutants in Beijing in recent years (excerpt)**表 1.** 北京市近年来各季度大气污染物浓度值(节选)

指标	2014 年 1 季度	2014 年 2 季度	2014 年 3 季度	2014 年 4 季度	2015 年 1 季度	2015 年 2 季度	2015 年 3 季度	2015 年 4 季度	2016 年 1 季度
AQI	146.3	116.7	119.7	119	126.7	107.7	103	148.7	100
PM2.5	111.9	68	72.3	88	91.7	62.2	52	114.8	68.1
PM10	138.7	112.2	92.1	123.4	126.7	94.8	66.2	114.4	94.5
SO ₂	47.2	12.5	5.9	17.5	26.9	7.9	4.2	12.4	17.8
CO	1.8	0.9	0.92	1.5	1.5	0.9	0.8	1.9	1.2

2.2. 各大城市按照其污染程度分类

AQI (Air Quality Index)是空气质量指数定量描述空气质量状况的无量纲指数。其数值越大、级别和类别越高,说明空气污染状况越严重,对人体的健康危害也就越大。

AQI 的计算公式如下:

$$AQI = \max \{IAQI_1, IAQI_2, \dots, IAQI_n\}$$

其中 IAQI——空气质量分指数, n——污染物项目。且

$$IAQI_p = (IAQI_{Hi} - IAQI_{Lo}) \times (C_p - BP_{Lo}) / (BP_{Hi} - BP_{Lo}) + IAQI_{Lo}$$

IAQI_p——污染物项目 P 的空气质量分指数;

C_p——污染物项目 P 的质量浓度值;

BP_{Hi}——相应地区的空气质量分指数及对应的污染物项目浓度指数表中与 C_p相近的污染物浓度限值的高位值;

BP_{Lo}——相应地区的空气质量分指数及对应的污染物项目浓度指数表中与 C_p相近的污染物浓度限值的低位值;

IAQI_{Hi}——相应地区的空气质量分指数及对应的污染物项目浓度指数表中与 BP_{Hi}对应的空气质量分指数;

IAQI_{Lo}——相应地区的空气质量分指数及对应的污染物项目浓度指数表中与 BP_{Lo}对应的空气质量分指数。

其数值与空气质量分类的关系如表 2 所示。

因此我们基于各大城市近年来各月份 AQI 的平均值,先大致对城市归为如下两类:

第一类: AQI 位于 50~100 (二级良)的城市:

上海 广州 深圳 杭州 南京 重庆 长沙 青岛 大连 厦门 苏州 宁波 无锡

第二类: AQI 位于 101~150 (三级轻度污染)的城市:

北京 武汉 成都 天津 沈阳 西安

从分类结果来看,这 19 个城市大都处于二级,三级之中。

2.3. 各大城市空气质量指数以及各大气污染物浓度的预测

2.3.1. 灰色预测

以下运用灰度预测的方法分别对各大城市空气质量指数和各大气污染物 PM2.5, PM10, SO₂, CO, NO₂, O₃ 进行预测(以北京市 AQI 预测为例,其余见附录附件二链接) [3] [4] [5]。

Table 2. AQI's relation to the classification of air quality
表 2. AQI 的数值与空气质量分类的关系

AQI 的数值	0~50	51~100	101~150	151~200	201~300	300 以上
污染程度	一级优	二级良	三级 轻度污染	四级 中度污染	五级 重度污染	六级 严重污染

从图 1 可以看出, 北京市的 AQI 有下降的趋势。但是, 从预测曲线来看, 灰色预测后的趋势几乎是线性的, 不太符合实际, 因此接下来我们采取 BP 神经网络再进行一次预测, 并最后将这两种方法进行误差分析选取最佳的一种的预测方法。

2.3.2. BP 神经网络预测

以下运用 BP 神经网络预测的方法分别对各大城市空气质量指数和各大气污染物 PM_{2.5}, PM₁₀, SO₂, CO, NO₂, O₃ 进行预测(以北京市 AQI 预测为例, 其余见附录附件二链接) [6], 预测结果如图 2 所示。

从图 2 可以看出, BP 神经网络预测结果是有一定的波动似乎是比较符合实际的。不过从 AQI 的预测结果可以看出。到 2030 年, 两个方法预测值仅相差 6 个浓度值左右, 因此有必要对这两种方法进行误差比较分析。

2.3.3. 灰色预测与 BP 神经网络预测误差比较分析

我们以前 10 个季度为样本, 以后四个数据作为检测。得到两种方法预测结果和误差如图 3, 图 4 所示。

从图 3 和图 4 可以看出, 虽然两者预测值相差不大, 但 BP 神经网络预测趋势与原数据大致相同, 而灰色预测则有点出入。

进一步计算出预测的平均残差可得: 灰色预测: 13.22, BP 神经网络: 6.34。

显然 BP 神经网络更加优于灰色预测。因此我们选用 BP 神经网络对城市的空气质量指数和各污染物进行预测。

2.3.4. 至 2030 年未能达标的城市

依据 BP 神经网络, 我们预测出到 2030 年 AQI 仍超出 100 的城市有以下 6 个北京, 武汉, 天津, 重庆, 西安, 长沙。以下我们将针对这 6 座城市分析其主要污染物及污染物成因。

2.4. 未达标城市的污染物与空气质量指数的灰色关联度分析

从 2.3.4 中我们已经得出了未达标的 6 座城市, 以下就 6 座城市的空气质量指数与各污染物的数据对其进行灰色关联度分析。

2.4.1. 灰色关联度模型

STEP1: 选取参考序列, 这里以 AQI 的数据为参考序列。

$$x_0 = \{x_0(k) | k=1, 2, \dots, n\} = (x_0(1), x_0(2), \dots, x_0(n)) \quad (k \text{ 表示时刻}, n=7) \quad (1)$$

$$x_i = \{x_i(k) | k=1, 2, \dots, n\} = (x_i(1), x_i(2), \dots, x_i(n)), i=1, 2, \dots, m \quad (m \text{ 为比较数列的个数}) \quad (2)$$

STEP2: 比较各污染物的数据对 AQI 在各时刻的关联系数

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \quad (3)$$

其中 $\rho \in [0,1]$ 为分辨系数, 一般为 0.5, $\min_s \min_t |x_0(t) - x_s(t)|$, $\max_s \max_t |x_0(t) - x_s(t)|$ 为两级最小差和最大差。

STEP3: 计算各污染物的数据对 AQI 的关联度。

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (4)$$

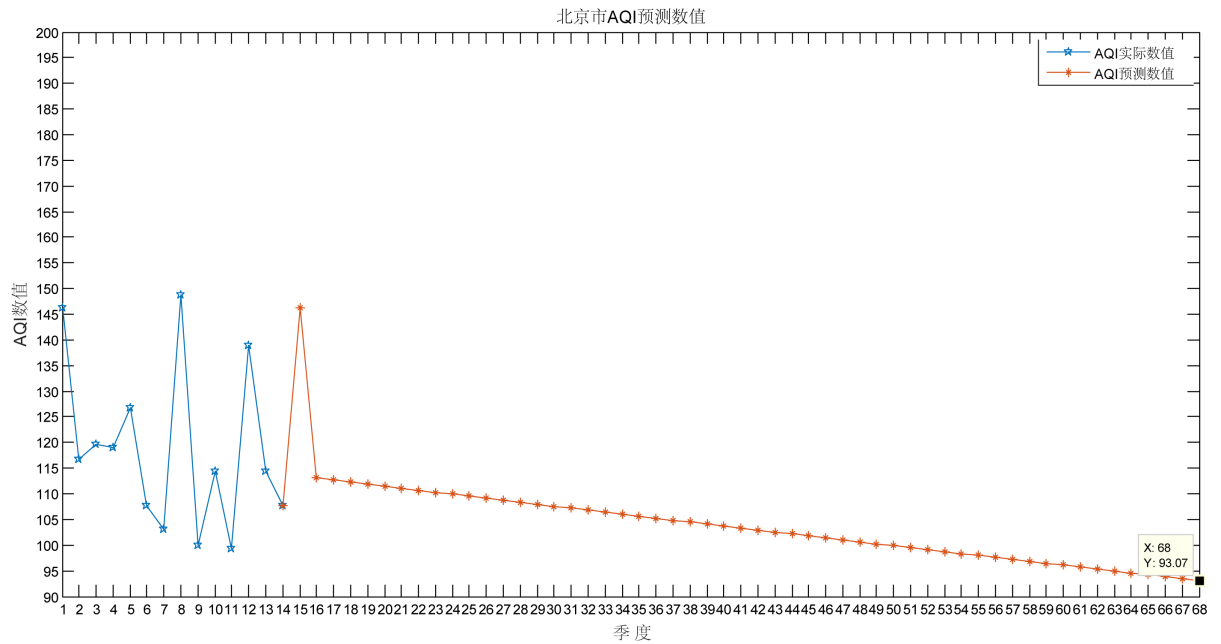


Figure 1. AQI grey prediction results of Beijing
图 1. 北京市 AQI 灰色预测结果

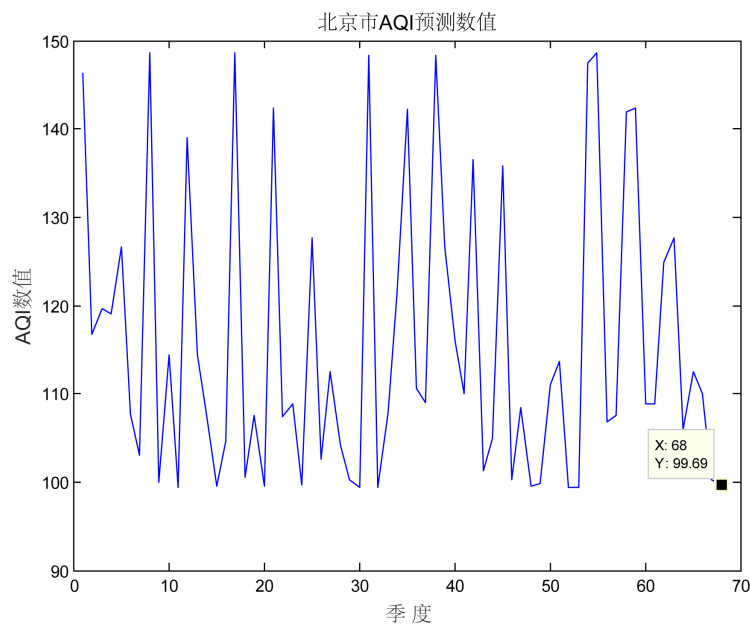


Figure 2. BP neural network prediction results of AQI in Beijing
图 2. 北京市 AQI 的 BP 神经网络预测结果

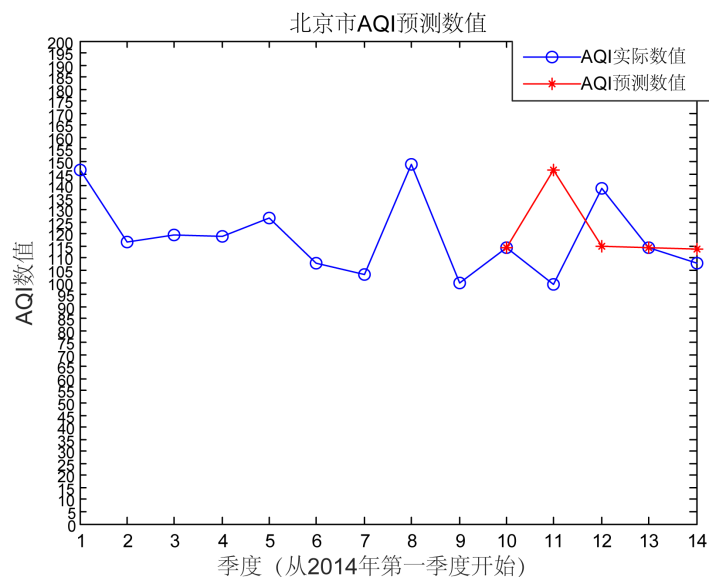


Figure 3. The error test of grey prediction

图 3. 灰色预测的误差检验

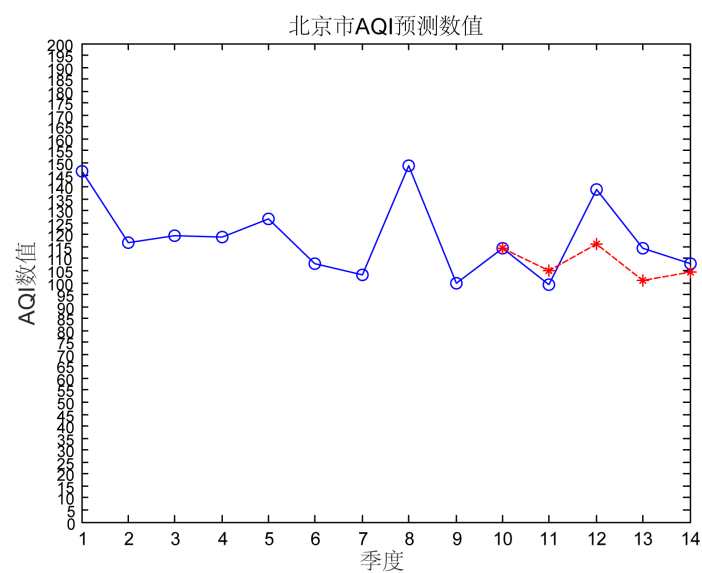


Figure 4. Error test of BP neural network

图 4. BP 神经网络的误差检验

2.4.2. 未达标 6 座城市各污染物浓度与 AQI 的关联度

通过 MATLAB2014a 编程得到, 未达标 6 座城市各污染物浓度与 AQI 的关联度的表格如表 3。

从表 3 可以看出, 北京, 天津, 重庆, 西安主要污染物是 PM10, 而武汉主要污染物浓度是 CO, 长沙主要污染物浓度是 SO₂。

3. 未达标城市的主要污染物的成因分析

3.1. 污染物成因的指标的确定

为了研究不能达标的六座城市的主要污染物的成因, 以便针对主要成因对点治理, 现对空气污染物

Table 3. Correlation between the concentration of pollutants and AQI in 6 cities

表 3. 未达标 6 座城市各污染物浓度与 AQI 的关联度

城市	PM2.5	PM10	SO ₂	CO	NO ₂	O ₃
北京	0.90104	0.93385	0.68212	0.87912	0.92136	0.58402
武汉	0.89782	0.87991	0.78718	0.91686	0.86016	0.57307
天津	0.93435	0.95769	0.72350	0.91399	0.9367	0.61620
重庆	0.93428	0.95097	0.80522	0.91755	0.74259	0.61282
西安	0.93748	0.96056	0.86441	0.93291	0.88404	0.58041
长沙	0.92611	0.80710	0.94790	0.89892	0.88000	0.57252

提取一个总的成因指标。查阅文献[7]知, 影响空气污染物原因主要可分为如下几种: 规模以上能源消费总量(万吨); 各城市的 GDP(亿元); 各城市的机动车保有量(辆); 绿化城市的覆盖率(公顷/立方千米); 工业废气排放量(亿立方米); 烟粉尘排放量(万吨); 查阅各城市的统计年鉴, 便能得到上述指标的值, 见附录附件 3。

3.2. 各城市主要污染物的主要成因分析

以北京为例, 对北京市的主要 PM10 进行成因分析。查阅文献[8]知, 可以采用最小二乘回归和主成分分析两种方法进行分析, 下文便分别采用这两种方法进行分析, 并最后通过误差检验来比较两种方法的优劣。

3.2.1. 最小二乘回归

最小二乘回归法又称最小平方法, 它是通过调整回归系数使得回归方程与实际数据之间误差的平方达到最小[9] [10]。其模型如下:

$$\min(\hat{Y} - Y)^2$$

其中 $\hat{Y} = \sum_{i=1}^7 \beta_i x_i + \alpha$ 为回归预测值, β_i 为回归系数, α 为回归方程的常数项。

采用最小二乘回归对北京主要污染物 PM10 进行成因分析得到回归方程如下:

$$y = 1175.97 + 0.018x_1 + 0.043x_2 + 0.13x_3 - 3488.32x_4 + 0.0004x_5 + 245.19x_6$$

从系数的绝对值的大小可以看出, 北京市主要污染物 PM10 的主要成因是城市绿化面积覆盖率和烟粉尘排放量。可见城市绿化面积覆盖率会降低北京市主要污染物 PM10 的浓度, 而工业烟粉尘排放量则会升高北京市主要污染物 PM10 的浓度。进一步求出均方误差如下:

$$Rmse_1 = 14.0524$$

3.2.2. 主成分分析

主成分分析也称主分量分析, 旨在利用降维的思想, 把多指标转化为少数几个综合指标(即主成分)。降维后的主成分能涵盖原信息的大部分信息[11] [12]。其步骤的程序框图如图 5 所示:

采用主成分分析法对北京主要污染物 PM10 进行成因分析得到回归方程如下:

$$y = 151.78 - 0.001x_1 - 0.002x_2 - 0.07x_3 - 260.63x_4 + 0.0006x_5 + 151.05x_6$$

其中选取的主成分个数为 4, 从系数的绝对值大小可以看出, 主要成因仍是北京市主要污染物 PM10 的

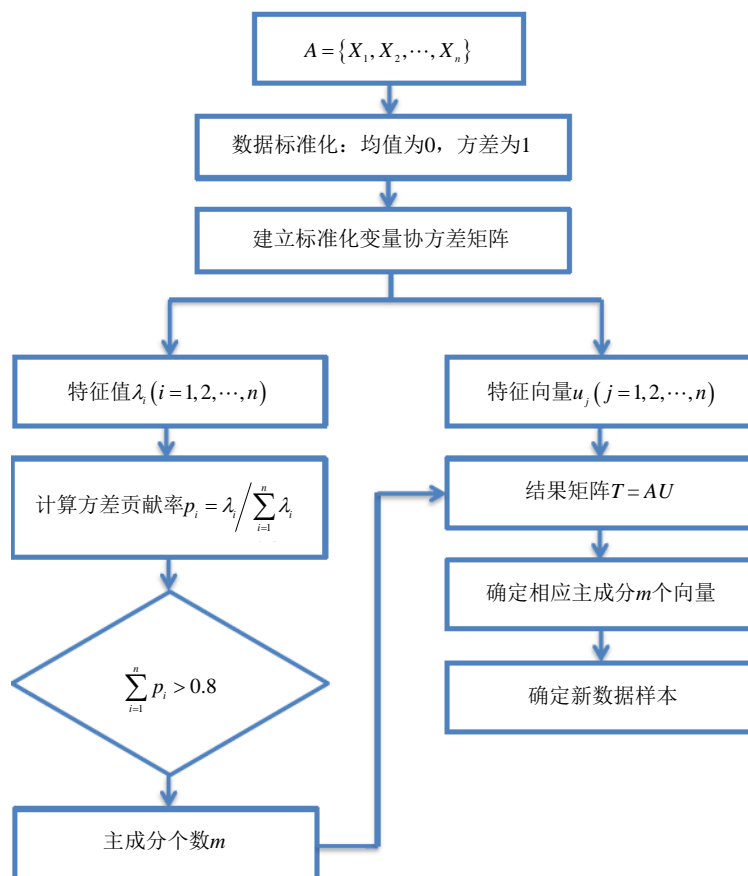


Figure 5. Flow diagram of principal component analysis
图 5. 主成分分析程序框图

主要成因是城市绿化面积覆盖率和烟粉尘排放量。进一步算出均方误差如下:

$$Rmse_2 = 13.3811$$

3.2.3. 综合分析

从均方误差可以看出。主成分分析的结果更加准确。因此对北京市的 PM10 的成因分析采用主成分分析法, 得出北京市主要污染物 PM10 的主要成因是城市绿化面积覆盖率和烟粉尘排放量。相似的, 对于其他城市, 可以得出:

天津市主要污染物 PM10 的成因分析采用最小二乘法回归更好, 其中回归方程为:

$$y = 1531.01 + 0.06x_1 - 0.06x_2 - 0.0001x_3 - 417.46x_4 - 0.0002x_5 - 0.0001x_6$$

因此天津主要污染物 PM10 最主要的影响因素是城市的绿化面积覆盖率。

重庆市主要污染物 PM10 的成因分析采用主成分分析法更好, 其中得到的回归方程:

$$y = 145.58 - 0.000435x_1 - 0.0015x_2 - 13.83x_4 - 0.0003x_5 + 0.2032x_6$$

其中可看出重庆市主要污染物的主要因素是城市绿化面积覆盖率和烟粉尘排放量。

西安市主要污染物 PM10 的成因分析采用最小二乘法回归更好, 其中得到的回归方程:

$$y = 699.45 + 0.25x_1 + 0.067x_2 + 1.33x_3 - 10.19x_4 - 0.03x_5 + 1.29x_6$$

其中可看出西安市主要污染物最主要的影响因素有三个: 机动车保有量, 城市绿化面积覆盖率和烟粉尘

排放量, 其中机动车保有量和烟粉尘排放量有助于 PM10 的增长, 而城市绿化面积覆盖率则能降低 PM10。

武汉市的主要污染物 CO 的成因分析采用最小二乘法回归更好, 其中得到的回归方程:

$$y = 312.29 + 0.004x_1 + 0.002x_2 - 6.64x_4 - 0.001x_5 + 2.14x_6 - 11.51x_7$$

注: x_7 代表的是固废综合利用率, 因武汉市统计年鉴新增了该数据, 因而多加了一个自变量。

通过回归方程可以看出影响武汉市的 CO 浓度最主要因素是城市绿化面积覆盖率, 烟粉尘排放量和固废综合利用率, 其中城市绿化面积覆盖率和固废综合利用率的提高都能显著降低 CO 的浓度, 而烟粉尘的排放量的增高则有助于提高 CO 浓度。

长沙市的主要污染物 SO₂ 的成因分析采用主成分分析法更好, 其中得到的回归方程:

$$y = 6.36 - 0.003x_1 - 140.72x_4 + 0.58x_6 - 19.31x_7$$

其中可以看出长沙市主要污染物 SO₂ 最主要的影响因素城市绿化面积覆盖率, 烟粉尘排放量和固废综合利用率, 其中城市绿化面积覆盖率和固废综合利用率的提高都能显著降低 SO₂ 的浓度, 而烟粉尘的排放量的增高则有助于提高 SO₂ 浓度。

从这上述七座城市可以看出城市绿化面积覆盖率都是影响各城市主要污染物浓度的最主要的因素, 这说明这七座城市绿化面积覆盖率的提高远不能和他们所排放的污染物浓度相匹配, 这七座城市还应在城市绿化面积上继续下功夫。

4. 结语

本文综合运用灰色预测, BP 神经网络的方法对全国的一线城市和新一线的城市的空气质量状况到 2030 年进行了一个预测, 并依照至 2030 年不能达到国家规定的标准挑选出了六座城市, 并对这六座城市运用灰色关联度的方法挑选出了其最主要的污染物, 并就其主要污染物运用主成分分析和最小二乘回归的方法找出了主要污染的主要成因和其中的量化表达式, 这将对每座城市的环境治理起重要的参考作用。本文所用的方法多样且进行了相互对比确保了所得结果和结论的准确性, 另外本文的一个最大的不足之处便是没有对污染物的主要成因落实到具体的行业或具体的公司、工厂上去, 这也将是今后一个主要的研究方向。

基金项目

- 1) 华中师范大学 2017 年度大学生创新创业训练计划 B 类项目;
- 2) 上海市青锐环境数据有限公司科研立项资助。

参考文献

- [1] Dennis, R.L., Byun, D.W., Novak, J.H., Galluppi, K.J., Coats, C.J. and Vouk, M.A. (1996) The Next Generation of Integrated Air Quality Modeling: EPA's Models-3. *Atmospheric Environment*, **30**, 1925-1938. [https://doi.org/10.1016/1352-2310\(95\)00174-3](https://doi.org/10.1016/1352-2310(95)00174-3)
- [2] 刘道文, 忽海娜. 非线性时间序列混沌特性分析及短期预测[J]. 计算机仿真, 2012, 29(10): 370-373.
- [3] 邓聚龙. 灰色系统基本方法[M]. 上海: 华中理工大学出版社, 1988: 58-76.
- [4] 孙中党, 赵勇, 田超, 李小浩. 郑州大气氮氧化物污染与灰色分析[J]. 重庆环境科学, 2003(1): 52-55, 62.
- [5] 王彩琴, 朱桦. 大气污染灰色预测的应用研究[J]. 西安建筑科技大学学报(自然科学版), 1998(1): 58-62.
- [6] 樊庆铨, 刘沙沙, 于淼. 灰色模型理论在哈尔滨市大气污染物 NO₂ 预测中的应用[J]. 环境保护科学, 2008(5): 1-3, 6.
- [7] 刘阳荷. 中国大气污染物排放量变动影响因素分析与精准减排[D]: [博士学位论文]. 济南: 山东大学, 2017.
- [8] 王海林. 北京市 PM2.5/PM10 的源解析[C]//中国化学会. 中国化学会第 26 届学术年会环境化学分会场论文集:

2008 年卷. 中国化学会, 2008: 1.

- [9] 石灵芝, 邓启红, 路婵, 刘蔚巍. 基于 BP 神经网络的大气颗粒物 PM₁₀质量浓度预测[J]. 中南大学学报(自然科学版), 2012, 43(5): 1969-1974.
- [10] 欧阳钧, 王爱枝. 基于 Matlab 的 BP 神经网络在大气污染物浓度预测中的应用[J]. 环境科学与管理, 2009, 34(11): 176-180.
- [11] Hedley, M.A. (1997) Evaluation of an Air Quality Simulation of the Lower Fraser Valle Meteorology Photochemistry. *Atmospheric Environment*, **31**, 1605-1630.
- [12] Zeng, B. and Li, C. (2016) Forecasting the Natural Gas Demand in China Using a Self-Adapting Intelligent Grey Model. *Energy*, 112. <https://doi.org/10.1016/j.energy.2016.06.090>

附 录

1、附件 1: 各一线城市和新一线城市各季度空气污染物数据统计表:

<https://pan.baidu.com/s/1oDAsS3t2YM0xT2nDQAQjTA>

2、附件 2: 各城市空气质量状况预测结果汇总(图像需用 MATLAB 打开):

<https://pan.baidu.com/s/1kAs2o4NEgtG7D7hOLTcXVQ>

3、附件 3: 6 座未达标城市的城市环境统计年鉴网址:

北京: <http://www.bjstats.gov.cn/nj/main/2016-tjn/jzk/indexch.htm>

天津: http://www.stats-tj.gov.cn/Article/List/List_29.html

西安: http://tjj.xa.gov.cn/ptl/def/def/index_923_6528.jsp

长沙: <http://www.cstj.gov.cn/tjn/jindex.jhtml>

重庆: <http://www.cqti.gov.cn/tjsj/shuju/tjn/j/>

武汉: <http://www.whjt.gov.cn/downlist.aspx>

Hans 汉斯

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>

下拉列表框选择: [ISSN], 输入期刊 ISSN: 2325-2251, 即可查询

2. 打开知网首页 <http://cnki.net/>

左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: sa@hanspub.org