

# The Analysis of Influence Factors of AQI and Prediction Based on Partial Correlation and Stepwise Regression Methods

Menglian Dai<sup>1</sup>, Tong Wang<sup>1</sup>, Haiyin Deng<sup>1</sup>, Fang Wang<sup>1,2\*</sup>

<sup>1</sup>College of Science, Hunan Agricultural University, Changsha Hunan

<sup>2</sup>Agricultural Mathematical Model and Data Process Center, Hunan Agricultural University, Changsha Hunan

Email: \*topwang619@163.com

Received: May 19<sup>th</sup>, 2017; accepted: Jun. 5<sup>th</sup>, 2017; published: Jun. 8<sup>th</sup>, 2017

## Abstract

Air quality index (AQI) is a dimensionless index which describes air condition quantitatively. It can be used to evaluate the quality of atmospheric environment as well as control contamination. In this paper, to investigate the influence of atmospheric pollution indexes on AQI, three cities of China, namely, Beijing, Changsha and Haikou, are chosen for your consideration. Six kinds of atmospheric pollution index contains PM2.5 of daily date are used as observation. On the one hand, partial correlation analysis is employed for each index of the three cities respectively to determine the main pollutants affecting AQI. It shows the main factors affecting the air quality in Beijing, Changsha and Haikou are PM10, PM2.5, NO<sub>2</sub> and O<sub>3</sub>; PM10, PM2.5, NO<sub>2</sub> and CO; PM10, PM2.5 and SO<sub>2</sub>, respectively. Moreover, the most significant correlation between the six indicators and AQI is PM2.5 for both Beijing and Changsha, but that is PM10 for Haikou. On the other hand, we also use stepwise regression method to access the primary factors affecting the AQI, which is consistent with the conclusion obtained from partial correlation analysis. In addition, we obtain the AQI optimal linear regression prediction model through partial correlation analysis and stepwise regression analysis. Besides, by the model tests, we find the proposed models for the three cities are workable, which can be used to forecast short-term AQI. Finally, some suggestions are provided for improving the air quality according to the results.

## Keywords

Air Quality Index (AQI), Partial Correlation Analysis, Stepwise Regression Analysis

# 基于偏相关与逐步回归方法的AQI影响因素分析及预测

戴孟莲<sup>1</sup>, 王彤<sup>1</sup>, 邓海银<sup>1</sup>, 王访<sup>1,2\*</sup>

\*通讯作者。

<sup>1</sup>湖南农业大学理学院, 湖南 长沙

<sup>2</sup>湖南农业大学, 农业数学建模与数据处理研究中心, 湖南 长沙

Email: topwang619@163.com

收稿日期: 2017年5月19日; 录用日期: 2017年6月5日; 发布日期: 2017年6月8日

## 摘要

空气质量指数(AQI)是定量描述空气质量状况的无量纲指数, 可用于评价大气环境质量以及控制污染。本文以北京、长沙、海口AQI及PM<sub>2.5</sub>等6种大气污染指标的日均数据为研究对象。一方面, 分别对三个城市的各个指标进行偏相关分析, 结果表明影响北京空气质量的主要因素是PM<sub>10</sub>、PM<sub>2.5</sub>、NO<sub>2</sub>和O<sub>3</sub>; 影响长沙空气质量的主要指标是PM<sub>10</sub>、PM<sub>2.5</sub>、NO<sub>2</sub>和CO; 而影响海口空气质量的主要因素是PM<sub>10</sub>、PM<sub>2.5</sub>和SO<sub>2</sub>。北京、长沙对AQI影响最大的是PM<sub>2.5</sub>, 而海口的PM<sub>10</sub>与AQI的相关关系最为显著。另一方面, 应用逐步回归分析得到了影响上述三城市AQI的最显著的空气污染指标, 结果与偏相关分析的结果相吻合。此外, 通过偏相关分析和逐步回归分析, 获得AQI的最优线性回归预测模型, 模型的拟合效果显著。模型检验表明建立的预测方程能有效用于AQI的短期预测。最后针对分析结果提出一些解决对策, 为空气质量改善提供建议。

## 关键词

空气质量指数, 偏相关分析, 逐步回归分析

Copyright © 2017 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

随着我国经济的飞速发展, 城市规模的不断扩大, 大气污染日益严重, 成为首要的社会问题之一, 环境空气质量和整个生态系统和人类健康息息相关[1]。对于空气污染问题不同学者从不同角度进行了研究[2]-[13]。例如, 王淑芸等[2]对河北沧州市空气质量及其污染物进行了统计分析, 发现沧州市空气质量优良的天数居多, 影响沧州市空气质量的首要污染物为PM<sub>10</sub>, 其次为SO<sub>2</sub>, 空气质量受季节影响较为严重。Jassim等[3]通过EPA方法识别对巴林岛空气质量有影响的最重要的污染物, 并将气象因素的影响纳入了考虑范围, 得出PM<sub>2.5</sub>浓度在一些工业城市更高的结论。Jarauta-Bragulat等[4]提到空气质量指数(AQI)与一些污染物相关, 而这些污染物又是各自浓度一种转化的计算, 第一次提出将整合合成数据应用到空气质量表达的方法。Gocheva-Ilieva等[5]应用因子分析和Box-Jenkins方法检查保加利亚主要污染物的浓度以及布拉戈耶夫格勒镇臭氧浓度, 用时间序列分析和ARIMA方法, 短期预测空气污染并对一些城区起到提醒作用。雷正翠等[6]对2001-2006年常州空气质量的变化特点进行了统计分析, 分析了气象条件对常州空气污染的影响和贡献程度。毕丽玫, 史建武等[8]用皮尔逊相关系数分析了温度、湿度、气压、风速、能见度对SO<sub>2</sub>, NO<sub>2</sub>, PM<sub>10</sub>, PM<sub>2.5</sub>的相关性, 并以此说明气象环境对污染物浓度的影响。Lu和Fang[9]使用三个不同分布(正态对数分布、威布尔分布、皮尔逊V型分布)分析PM<sub>2.5</sub>、PM<sub>10</sub>和风速之间的关系, 研究表明PM值与风速之间的反向相关关系可用某个关系式表示。朱奕[10]对污染物进

行了后向气团轨迹分析,发现气象条件对颗粒污染物的浓度有着显著影响,且除本地排放外,来自中东部污染严重城市的区域性气流输送是导致长沙市颗粒物污染的重要原因。徐小丽[11]运用主成分分析法(PCA)探讨西安市及伦敦市的空气质量监测指标中多种污染物与PM<sub>2.5</sub>的相关性,结果表明:PM<sub>10</sub>与PM<sub>2.5</sub>最相关,同时,利用多重分析发现西安市和伦敦市有着共同特点。Kumar A等[12]通过以PCA为基础的神经网络预测印度德里AQI的日均数据,结果显示PCA神经网络要优于神经网络预测。

然而,上述研究大多针对影响空气质量的特定两个变量的相互关系进行研究,所用的方法主要是简单的相关分析。这种分析只考虑两变量间的相互作用,没有考虑其他变量对它们的影响。然而,系统内所有变量之间的相互作用在一定程度上可能掩盖待研究的两变量间的关系。例如,当两变量的简单相关系数较大时,有时并不是这两个变量本身的内在联系所决定的,它可能是由另外一些变量对它们的共同作用造成的,因此简单的相关系数不能真正反映两个变量间的线性相关程度。大气污染是一个多因素控制的、复杂的系统问题,在研究多变量的情况下,一个要素的变化往往影响到其它要素的变化,即各个空气污染指标之间具有高相关性,互相影响,因此与AQI的相关程度会受到其他指标的影响。基于此文引进偏相关分析对AQI的影响因素进行研究。当考察两个变量之间的线性相关关系时,偏相关分析能控制其他变量对它们的影响,能有效揭示这两个变量之间的真实相关程度[14]。然而,在空气质量的研究中,利用偏相关分析的研究报道不多,且仅针对单一地区进行研究。例如,王涛等[15]利用偏相关分析描述PM<sub>2.5</sub>、PM<sub>10</sub>与NO<sub>2</sub>、SO<sub>2</sub>的关系,结果表明江苏省的二次PM<sub>2.5</sub>主要源自的SO<sub>2</sub>的气-粒转化过程,固定污染源的间接贡献较大;PM<sub>10</sub>更多源自固定污染源而非移动污染源。徐莲等[16]针对NO<sub>2</sub>、O<sub>3</sub>两种特征性大气污染物,在总结出一般性的污染变化规律的基础上,进一步采用偏相关分析方法,证实了NO<sub>2</sub>是生成O<sub>3</sub>的重要前体污染物。

当研究了AQI与各个影响因素的真实相关程度后,为其建立短期回归预测模型是另一个重要的工作。然而,利用多个自变量建立回归模型时,若试图将所有的自变量都引进回归模型,往往不能对其进行有效的解释,因此需要挑选出对因变量有显著影响的自变量。逐步回归分析是一个有效剔除冗余变量的方法,该方法通过将前进法和后退法结合,保证了最后所得的回归子集是最优回归子集[14]。

此外,现有的文献大多以空气污染指数(API)为标准,研究大气污染问题,随着2012年国家环保部门修订了空气质量的测量标准,首次将PM<sub>2.5</sub>、NO<sub>2</sub>等物质纳入到空气质量勘测的范畴,研究AQI将能更准确反映环境的实际状况,预测AQI成为现在空气质量研究领域的一个话题[17][18]。基于以上分析,本文试图利用偏相关分析与逐步回归来研究AQI与各空气污染指标的影响情况及其进行短期预测。

本文选取中国三个城市作为研究对象,分别是首都北京,内陆城市长沙,沿海城市海口,以三个城市的AQI及PM<sub>2.5</sub>等6种污染指标数据为基础,利用偏相关分析研究6种污染指标对AQI的影响情况,并通过逐步回归分析获得上述三个城市的AQI回归方程,最后有针对性地提出改善城市的空气质量的建议。

## 2. 方法与数据

### 2.1. 偏相关分析

偏相关分析又称净相关分析,能够有效地控制其他变量影响条件下揭示两变量之间净相关的强弱程度。偏相关系数绝对值越接近1,表示两变量关系越密切,影响程度越大;越接近0,相关性越小[14]。因此偏相关系数能定量描述AQI与每一种空气污染指标的真实相关程度。通过求出AQI与各影响因素间的偏相关系数并进行检验,即可以得到影响AQI的主要因素。偏相关系数计算过程[19]如下:控制了变量 $x_2$ ,变量 $x_1$ 和 $y$ 的偏相关系数为:

$$r_{x_1 y \cdot x_2} = \frac{r_{y \cdot x_1} - r_{y \cdot x_2} r_{x_1 \cdot x_2}}{\sqrt{(1 - r_{y \cdot x_2}^2)(1 - r_{x_1 \cdot x_2}^2)}} \quad (1)$$

式中  $r_{y \cdot x_1}, r_{y \cdot x_2}, r_{x_1 \cdot x_2}$  分别表示  $y$  和  $x_1$ ,  $y$  和  $x_2$  及  $x_1$  和  $x_2$  的相关系数。

控制两个变量  $x_2, x_3$  时, 变量  $x_1, y$  之间的偏相关系数为

$$r_{x_1 y \cdot x_2 \cdot x_3} = \frac{r_{y x_1 \cdot x_2} - r_{x_1 x_3 \cdot x_2} r_{y x_3 \cdot x_2}}{\sqrt{(1 - r_{x_1 x_3 \cdot x_2}^2)(1 - r_{y x_3 \cdot x_2}^2)}} \quad (2)$$

式中  $r_{y x_1 \cdot x_2}$  为控制变量  $x_2, x_3$  时, 变量  $x_1$  与  $y$  的偏相关系数;  $r_{y x_3 \cdot x_2}$  为控制变量  $x_2$  时, 变量  $x_1$  与  $y$  的偏相关系数, 依此类推。

当控制变量有 3 个或 3 个以上时, 计算变量之间的偏相关系数依照式(1)、(2)类推。

## 2.2. 偏相关系数的检验

当要素之间的相关系数求出之后, 还需要对其进行显著性检验。由于抽样误差的存在, 样本中控制了其它变量的影响, 两个变量间偏相关系数不为 0, 不能说明在总体中这两个变量间便相关, 此时须进行进一步检验, 一般使用  $t$  检验方法[20]。

$$t = r \sqrt{\frac{n - k - 2}{\sqrt{1 - r^2}}} \quad (3)$$

式中:  $r$  是相应的偏相关系数,  $n$  是观测量数,  $k$  是控制变量的数量, 自由度  $n - k - 2$ 。选定显著性 0.05, 当  $t > t_{0.05}(n - k - 2)$  时拒绝原假设, 认为变量间有显著的线性关系。

## 2.3. 逐步回归分析

逐步回归法是在前进法和后退法基础上, 进行双向筛选的一种方法, 它的基本思想是“有进有出”。具体做法是: 将变量一个一个引入, 每引入一个变量后, 对已选入的变量进行逐个检验, 当原引入的变量由于后面变量的引入而变得不再显著时, 将其剔除。引入一个变量或从回归方程中剔除一个变量, 为逐步回归的一步, 每一步都需进行  $F$  检验, 直到既无显著的自变量选入回归方程, 也无不显著的自变量从回归方程剔除为止。这样就弥补了前进法和后退法的缺陷, 保证了最后所得的回归子集是最优回归子集[21]。

## 2.4. 数据来源

本文选取北京、长沙、海口的 AQI 和 PM2.5、PM10、SO<sub>2</sub>、CO、NO<sub>2</sub>、O<sub>3</sub> 浓度数据来源于空气质量历史数据网站(<https://www.aqistudy.cn/historydata/daydata.php>)。选取 2013 年 12 月~2016 年 11 月 AQI 与 PM2.5、PM10、SO<sub>2</sub>、CO、NO<sub>2</sub>、O<sub>3</sub> 6 种参评污染物的日均浓度, 每个城市都收集了共 1095 天的数据。在下面的分析中, AQI 为因变量, 6 种大气污染指标为自变量。

## 3. AQI 与 6 种污染指标的偏相关分析

利用 SPSS23.0 对 AQI 及 6 种污染物做偏相关分析, 控制其他 5 种变量, 计算出 AQI 与每种污染指标两两间的偏相关系数。以海口市为例, 结果如表 1~6 所示。

由表 7 可知, 北京的 AQI 与 PM2.5、PM10、NO<sub>2</sub>、O<sub>3</sub> 的偏相关系数通过检验, 即北京 AQI 的主要污染源为 PM10、PM2.5、NO<sub>2</sub> 和 O<sub>3</sub>; 长沙 AQI 的主要污染源为 PM10、PM2.5、NO<sub>2</sub> 和 CO。而海口 AQI 的主要污染源为 PM10、PM2.5 和 SO<sub>2</sub>。由此可看出, 这三个城市影响空气质量的大气污染源存在差异,

**Table 1.** Partial correlation between the AQI and PM2.5**表 1.** AQI 与 PM2.5 偏相关性

控制变量			AQI	PM2.5
PM10 & SO <sub>2</sub> & CO & NO <sub>2</sub> & O <sub>3</sub>	AQI	相关性	1.000	0.564
		显著性(双尾)	.	0.000
		自由度	0	1088
	PM2.5	相关性	0.564	1.000
		显著性(双尾)	0.000	.
		自由度	1088	0

**Table 2.** Partial correlation between the AQI and PM10**表 2.** AQI 与 PM10 偏相关性

控制变量			AQI	PM10
SO <sub>2</sub> & CO & NO <sub>2</sub> & O <sub>3</sub> & PM2.5	AQI	相关性	1.000	0.636
		显著性(双尾)	.	0.000
		自由度	0	1088
	PM10	相关性	0.636	1.000
		显著性(双尾)	0.000	.
		自由度	1088	0

**Table 3.** Partial correlation between the AQI and SO<sub>2</sub>**表 3.** AQI 与 SO<sub>2</sub> 偏相关性

控制变量			AQI	SO <sub>2</sub>
CO & NO <sub>2</sub> & O <sub>3</sub> & PM2.5 & PM10	AQI	相关性	1.000	-0.098
		显著性(双尾)	.	0.001
		自由度	0	1088
	SO <sub>2</sub>	相关性	-0.098	1.000
		显著性(双尾)	0.001	.
		自由度	1088	0

**Table 4.** Partial correlation between the AQI and CO**表 4.** AQI 与 CO 偏相关性

控制变量			AQI	CO
NO <sub>2</sub> & O <sub>3</sub> & PM2.5 & PM10 & SO <sub>2</sub>	AQI	相关性	1.000	0.054
		显著性(双尾)	.	0.074
		自由度	0	1088
	CO	相关性	0.054	1.000
		显著性(双尾)	0.074	.
		自由度	1088	0

**Table 5.** Partial correlation between the AQI and NO<sub>2</sub>  
**表 5.** AQI 与 NO<sub>2</sub> 偏相关性

控制变量		AQI	NO <sub>2</sub>
O <sub>3</sub> & PM2.5 & PM10 & SO <sub>2</sub> & CO	相关性	1.000	0.004
	AQI		
	显著性(双尾)	.	0.906
	自由度	0	1088
	NO <sub>2</sub>		
	显著性(双尾)	0.906	.
	自由度	1088	0

**Table 6.** Partial correlation between the AQI and O<sub>3</sub>  
**表 6.** AQI 与 O<sub>3</sub> 偏相关性

控制变量		AQI	O <sub>3</sub>
PM2.5 & PM10 & SO <sub>2</sub> & CO & NO <sub>2</sub>	相关性	1.000	-0.014
	AQI		
	显著性(双尾)	.	0.644
	自由度	0	1088
	O <sub>3</sub>		
	显著性(双尾)	0.644	.
	自由度	1088	0

**Table 7.** Results of the partial correlation analysis for the AQI and six factors in the three cities  
**表 7.** 三个城市AQI与6种大气污染物的偏相关分析综合结果

地区	项目	PM2.5	PM10	NO <sub>2</sub>	SO <sub>2</sub>	CO	O <sub>3</sub>
北京	偏相关系数	0.887	0.533	-0.177	-0.004	-0.04	0.107
	显著性 <i>p</i> 值	0.000	0.000	0.000	0.884	0.182	0.000
	结论	正相关	正相关	负相关	不显著	不显著	正相关
长沙	偏相关系数	0.967	0.156	0.061	0.043	-0.109	0.049
	显著性 <i>p</i> 值	0.000	0.000	0.043	0.158	0.000	0.103
	结论	正相关	正相关	正相关	不显著	负相关	不显著
海口	偏相关系数	0.564	0.636	0.004	-0.098	0.054	-0.014
	显著性 <i>p</i> 值	0.000	0.000	0.906	0.001	0.074	0.644
	结论	正相关	正相关	不显著	负相关	不显著	不显著

北京、长沙的 PM2.5 对 AQI 的影响最大，这是因为 PM2.5 粒径小，富含大量的有毒、有害物质且在大气中的停留时间长、输送距离远，因而对人体健康和大气环境质量的影响更大。而海口的 PM10 对 AQI 的影响最大，文献[22]通过对海口市大气颗粒物来源进行源解析，发现 PM10 的主要源类为机动车尾气尘、城市扬尘、建筑水泥尘，这主要与海口市无取暖季、机动车保有量及机动车类型有关。大部分沿海地区受到大气环流的影响，对海陆空气污染的形成和运输具有重要的影响，可以混合不同的气团[21]，其污染的来源构成和形成机制可能不同于内陆地区。

#### 4. AQI 与 6 种大气污染物的逐步回归分析

本节通过逐步回归分析对这三个城市分别建立 AQI 的预测方程。在如何挑选出对 AQI 有显著影响的自变量的问题上，由于逐步回归吸收了前进法和后退法的优点，克服了它们的不足，保证了最后所得的回归子集是最优回归子集。

空气质量评价的主要污染物为 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、CO、NO<sub>2</sub>、O<sub>3</sub>，AQI 由其含量计算所得。下面运用 SPSS23.0 统计软件对 AQI 与上述 6 种污染物进行逐步回归，取引入自变量的显著性水平  $\alpha_{entry} = 0.05$ ，剔除自变量的显著性水平  $\alpha_{removal} = 0.1$ ，海口逐步回归的结果如表 8 所示。

由表 8 可知，逐步回归的最优子集为模型 3，即 PM<sub>10</sub>、PM<sub>2.5</sub>、SO<sub>2</sub>。仅 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub> 对 AQI 有重要的影响进入方程，其余 3 个自变量对 AQI 影响均不显著被剔除。即 PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub> 对 AQI 的影响显著，具有统计学意义，这个结果与偏相关的结果一致。且决定系数  $R^2 = 0.970$ ，说明在逐步拟合的线性回归方程的因变量 AQI 能被自变量解释的占 97%。

由表 9 可知，通过逐步回归分析，海口 AQI 的最优多元线性回归方程为：

$$\hat{y} = 8.849 + 0.507PM_{10} + 0.572PM_{2.5} - 0.202SO_2 \quad (4)$$

$P = 0.000 < 0.05$ ，说明在显著水平为 0.05 的情况下回归系数通过  $t$  检验，PM<sub>10</sub>、PM<sub>2.5</sub>、SO<sub>2</sub> 对 AQI 的线性影响均显著。

由表 10 可知， $F = 11706.377$ ， $p = 0.000 < 0.05$ ，说明在显著水平为 0.05 的情况下回归方程通过  $F$  检验，线性回归效果显著，逐步回归方程具有统计意义。该方程可用于海口 AQI 的预测。用同样的方法对北京和长沙的因变量 AQI 和自变量进行逐步回归，北京 AQI 的最优子集为 PM<sub>2.5</sub>、PM<sub>10</sub>、NO<sub>2</sub> 和 O<sub>3</sub>，最优多元线性方程为

Table 8. Summary of model

表 8. 模型摘要

模型	R	R方	调整后R方	标准估算的误差
1	0.976 <sup>a</sup>	0.952	0.952	4.4118
2	0.985 <sup>b</sup>	0.970	0.969	3.5246
3	0.985 <sup>c</sup>	0.970	0.970	3.5061

注：a. 预测变量：(常量)，PM<sub>10</sub>；b. 预测变量：(常量)，PM<sub>10</sub>，PM<sub>2.5</sub>；c. 预测变量：(常量)，PM<sub>10</sub>，PM<sub>2.5</sub>，SO<sub>2</sub>

Table 9. Regress coefficient

表 9. 回归系数表

模型		未标准化系数		标准化系数	$t$ 统计量	显著性
		B	标准误差	Beta		
1	(常量)	3.533	0.286		12.366	0.000
	PM <sub>10</sub>	0.930	0.006	0.976	147.569	0.000
	(常量)	8.363	0.300		27.924	0.000
2	PM <sub>10</sub>	0.502	0.018	0.527	28.040	0.000
	PM <sub>2.5</sub>	0.551	0.022	0.468	24.910	0.000
	(常量)	8.849	0.328		26.969	0.000
3	PM <sub>10</sub>	0.507	0.018	0.532	28.377	0.000
	PM <sub>2.5</sub>	0.572	0.023	0.486	25.093	0.000
	SO <sub>2</sub>	-0.202	0.057	-0.030	-3.536	0.000

注：因变量：AQI，B 指回归系数和截距，左边对应常数代表截距，对应变量代表回归系数。

**Table 10.** Analysis of variance  
**表 10.** 方差分析

	模型	平方和	自由度	均方	F统计量	显著性
1	回归	423,857.104	1	423,857.104	21,776.660	0.000 <sup>b</sup>
	残差	21,273.961	1093	19.464		
	总计	445,131.065	1094			
2	回归	431,565.681	2	215,782.841	17,370.306	0.000 <sup>c</sup>
	残差	13,565.383	1092	12.423		
	总计	445,131.065	1094			
3	回归	431,719.404	3	143,906.468	11,706.377	0.000 <sup>d</sup>
	残差	13,411.661	1091	12.293		
	总计	445,131.065	1094			

注: a. 因变量: AQI; b. 预测变量: (常量), PM10; c. 预测变量: (常量), PM10, PM2.5; d. 预测变量: (常量), PM10, PM2.5, SO<sub>2</sub>。

$$\hat{y} = 21.698 + 0.931PM2.5 + 0.239PM10 - 0.218NO_2 + 0.027O_3 \quad (5)$$

长沙 AQI 的最优子集为 PM2.5、PM10、CO 和 NO<sub>2</sub>, 最优多元线性方程为

$$\hat{y} = 12.710 + 1.160PM2.5 + 0.05PM10 - 4.224CO + 0.055NO_2 \quad (6)$$

三个城市回归方程检验  $p$  值都为  $0.000 < 0.05$ , 线性回归效果显著, 且拟合优度好, 回归系数均通过检验。综合三个城市的偏相关分析, 逐步回归选取的最优子集与偏相关系数显著的自变量吻合。

## 5 模型检验

### 5.1. 回归模型精度分析

为了检验建立的上述三个城市回归模型的有效性, 我们首先来考察模型精度, 利用相对根误差 (RRMSE) 来描述模型精度, 如式(7)所示。利用式(4)~(6)求出训练样本中 1095 天(2013 年 12 月~2016 年 11 月)AQI 的预测值, 将其与 AQI 真实值比较, 如图 1 所示。从图中可以看出: 三个城市的 AQI 训练数据的真实值与预测值决定系数都达 0.97 以上, 模型的拟合优度较高, 其预测值基本分布在真实值周围, 且相对均方根误差均小于 12%, 预测精度较高, 尤以长沙的回归效果最佳。

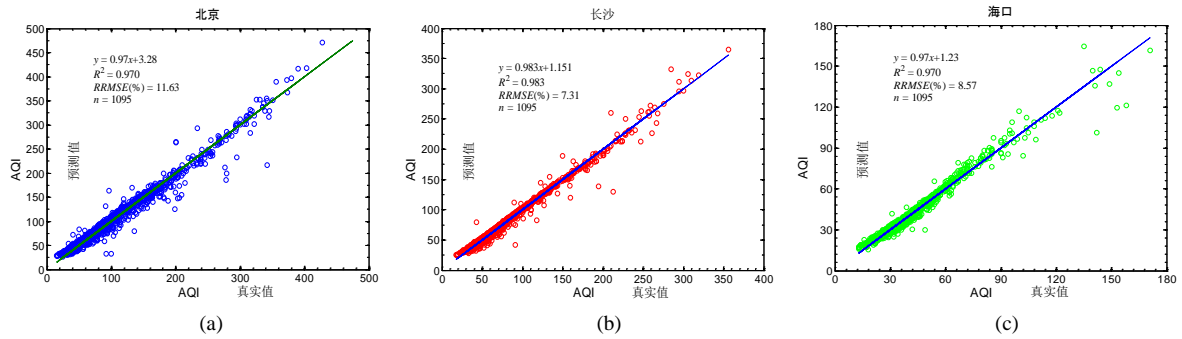
$$RRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - Q_i)^2 / \overline{Q_i}} \times 100\% \quad (7)$$

式中:  $P_i$  为预测值,  $Q_i$  为真实值,  $n$  为样本个数。

### 5.2. 预测模型误差分析

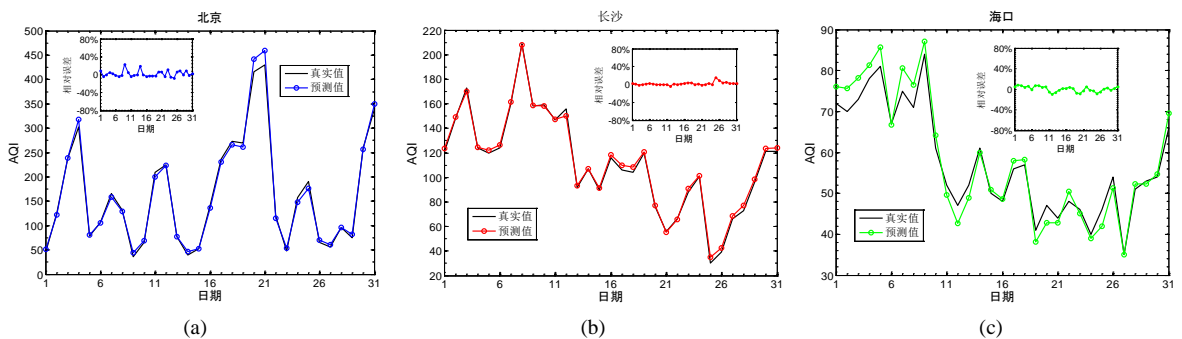
为了考察回归模型的预测效果, 利用求得的三个回归预测方程对三个城市的 2016 年 12 月的 AQI 进行预测, 并与真实数据作比较, 如图 2 所示。为了定量描述预测误差, 真实值与预测值的相对误差也显示在图 2 中。从图中可以看出, 除极少数点外, 北京和海口的预测相对误差都在 10% 以内, 而长沙的预测相对误差在 5% 以内, 说明上述回归方程的预测效果较好, 其相对误差较平稳, 波动较小。表明本文提出的最优线性回归方程用于预测短期 AQI 是可靠的。





**Figure 1.** The predictive results of AQI for (a) Beijing; (b) Changsha and (c) Haikou in December, 2013 to November, 2016. The solid lines denote the true value, beside them are the model evaluations

**图 1.** (a) 北京、(b) 长沙、(c) 海口三个城市 2013.12~2016.11 月 AQI 的预测结果，实线代表真实值，其旁边的参数表示模型评价结果



**Figure 2.** The predicative results of AQI for (a) Beijing; (b) Changsha and (c) Haikou in December, 2016. The main plot is true values and predicative values. Insect is relative error

**图 2.** (a) 北京、(b) 长沙、(c) 海口三个城市 2016 年 12 月 AQI 预测结果，主图是真实值与预测值，子图是相对误差

## 6. 结果分析及建议

在本文的研究中，我们发现北京、长沙、海口三个城市影响空气质量的大气污染源存在差异。对于北京和长沙两个城市，PM<sub>2.5</sub> 对 AQI 的影响最大；而海口的 PM<sub>10</sub> 对 AQI 的影响最大。这说明空气质量污染主要还是煤燃烧，工业排放的废气，城市扬尘，汽车尾气，建筑水泥尘等。三个城市的大气主要污染物都包含 PM<sub>2.5</sub> 和 PM<sub>10</sub>，且与 AQI 的关系密切，在控制大气污染上，防治细颗粒物污染尤为重要。提出的回归预测方程能用于 AQI 的短期预测，误差较小，模型可靠。

为了进一步地改善城市环境空气质量，提出以下几点建议：

1) 在本文的研究中，三个城市影响空气质量的大气污染源存在差异，即不同城市的大气污染影响因素不同，在控制大气污染上，不同城市的治理方案应该有所差异和侧重，有针对性地提出改善方案效果会更好。

2) 改变能源消耗结构，推动节能减排；强化工业污染源排放的监管工作，不达标的排放物不能轻易排放；提高工程项目的申报标准；强化城市建设道路施工和扬尘的控制，强化道路运输车辆管理，加强道路洒水降尘工作。

3) 优化产业结构和布局。将重工业行业向外转移，同时大力发展金融服务业、医疗服务业、教育产业和高新技术产业等，来达到既减少能源需求和污染物排放，又保证经济的持续增长和充分就业的目的，缓解环境保护与经济发展之间的矛盾。

4) 控制汽车增长率，大力发展新能源公用交通工具，提倡绿色出行；进一步完善地铁线路，错峰出

行,降低汽车尾气排放。

5) 加大城市绿化美化,增大绿地面积和绿化覆盖率。严格控制低质煤炭的销售和使用,特别是城市居民取暖用煤。

6) 污染物的区域传输也是影响城市空气质量的一个原因[23] [24],因此对于空气污染的治理,还要注意外源输送的影响。

## 基金项目

本文由湖南省社科基金项目(15YBA204)和湖南农业大学大学生科创项目(XCX16005)资助。

## 参考文献 (References)

- [1] 中国环境科学研究院,中国环境监测总站. GB3095-2012. 环境空气质量标准[S]. 北京:中国环境科学出版社,2012.
- [2] 王淑云,节江涛,熊险平,等. 城市空气质量与气象条件的关系及空气质量预报系统[J]. 气象科技. 2006, 34(6): 688-692.
- [3] Jassim, M.S. and Coskuner, G. (2017) Assessment of Spatial Variations of Particulate Matter (PM10 and PM2.5) in Bahrain Identified by Air Quality Index (AQI). *Arabian Journal of Geosciences*, **10**, 19-25.
- [4] Jarauta-Bragulat, E., Hervada-Sala, C. and Egozcue, J.J. (2016) Air Quality Index Revisited from a Compositional Point of View. *Mathematical Geosciences*, **48**, 1-13.
- [5] Gocheva-Ilieva, S.G., Ivanov, A.V., Voynikova, D.S., et al. (2014) Time Series Analysis and Forecasting for Air Pollution in Small Urban Area: An SARIMA and Factor Analysis Approach. *Stochastic Environmental Research and Risk Assessment*, **28**, 1045-1060.
- [6] 雷正翠,董芹,霍焱,等. 常州城市空气质量变化特征及其与气象条件的关系[C]//中国气象学会. 中国气象学会2008年会议论文集. 北京:中国气象学会,2008.
- [7] 杨钰,金永民,孙晓怡,等. 用因子分析法解析抚顺市大气颗粒物来源[J]. 环境保护与循环经济, 2005, 25(4): 7-9.
- [8] 毕丽玫,史建武,刘意,等. 昆明城区PM2.5与常规大气污染物及气象因素的相关性分析[C]//中国环境科学学会. 中国环境科学学会2013年学术年会论文集. 昆明:中国环境科学学会,2013.
- [9] Lu, H.C. and Fang, G.C. (2002) Estimating the Frequency Distributions of PM10 and PM2.5 by the Statistics of Wind Speed at Sha-Lu. *Science of the Total Environment*, **298**, 119-130.
- [10] 朱奕. 长沙城区PM10和PM2.5的污染特征及气象因素溯源分析研究[J]. 环境科学与管理, 2013, 38(11): 57-62.
- [11] 徐小丽. PM2.5变化趋势的多重分形分析研究[D]: [硕士学位论文]. 合肥:安徽大学,2014.
- [12] Kumar, A. and Goyal, P. (2013) Forecasting of Air Quality Index in Delhi Using Neural Network Based on Principal Component Analysis. *Pure and Applied Geophysics*, **170**, 711-722. <https://doi.org/10.1007/s00024-012-0583-4>
- [13] 李小飞,张明军,王圣杰,等. 中国空气污染指数变化特征及影响因素分析[J]. 环境科学, 2012, 33(6): 1936-1943.
- [14] 何晓群,刘文卿. 统计学[M]. 北京:中国人民大学出版社,2015.
- [15] 王涛,康准钰,刘梓雅,等. 基于比值和偏相关的大气污染物源识别——以2014年江苏省为例[J]. 复旦学报(自然科学版), 2016, 55(4): 538-542.
- [16] 徐莲,马民涛,金毓峯,等. 北京地区O<sub>3</sub>与NO<sub>2</sub>的偏相关分析[J]. 城市环境与城市生态, 2003(2): 67-68.
- [17] 黄世芹. 用卡尔曼滤波法和逐步回归法建立空气质量预报系统[J]. 贵州气象, 2002, 26(5): 20-22.
- [18] 刘天,刘瑛,钟杰. PM2.5成因的逐步回归分析及治理方案研究[J]. 数学的实践与认识, 2014, 44(15): 76-80.
- [19] 焦文玲,邹涛,巩晓峰,等. 燃气日负荷影响因素的相关分析和偏相关分析[J]. 煤气与热力, 2010, 30(5): 1-5.
- [20] 崔素芳,张振华,姚付启,等. 基于偏相关分析的烟台市土壤温度影响因素及预测模型研究[J]. 山东农业科学, 2010(1): 18-21.
- [21] Masiol, M., Rampazzo, G., Ceccato, D., et al. (2010) Characterization of PM10 Sources in a Coastal Area near Venice (Italy): An Application of Factor-Cluster Analysis. *Chemosphere*, **80**, 771-778. <https://doi.org/10.1016/j.chemosphere.2010.05.008>

- 
- [22] 宋娜, 徐虹, 毕晓辉, 等. 海口市 PM<sub>2.5</sub> 和 PM<sub>10</sub> 来源解析[J]. 环境科学研究, 2015, 28(10): 1501-1509.
- [23] 符传博, 唐家翔, 丹利, 等. 2014 年海口市大气污染物演变特征及典型污染个例分析[J]. 环境科学学报, 2016, 36(6): 2160-2169.
- [24] 李璇, 聂滕, 齐珺, 等. 2013 年 1 月北京市 PM<sub>2.5</sub> 区域来源解析[J]. 环境科学, 2015, 36(4): 1148-1153.

**期刊投稿者将享受如下服务:**

1. 投稿前咨询服务 (QQ、微信、邮箱皆可)
2. 为您匹配最合适的期刊
3. 24 小时以内解答您的所有疑问
4. 友好的在线投稿界面
5. 专业的同行评审
6. 知网检索
7. 全网络覆盖式推广您的研究

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

期刊邮箱: [aep@hanspub.org](mailto:aep@hanspub.org)