

空气中 PM2.5 问题的研究

——以蚌埠市为例

邵晴晴

(安徽财经大学, 安徽 蚌埠 233030)

【摘要】: 空气质量问题越来越引起各界关注, 依据蚌埠市 PM2.5 的各项监测数据, 参照我国新修订的《环境空气质量标准》的标准和要求, 对蚌埠市空气中各污染物的相关性进行定性和定量分析, 并对 PM2.5 与 PM10、氮氧化物等因素进行相关性分析和检验, 建立 PM2.5 与 CO、SO₂、NO₂、PM10 等因素之间的多元线性回归模型, 并建立高斯扩散模型, 刻画出蚌埠市空气中 PM2.5 的时空分布规律, 并对该地区进行污染的评估, 最终得出含有风力、湿度等因素的多元线性回归模型。最后运用对比法, 将实际检测出来的数据与模型计算出来的模拟值进行比较与检验, 并对模型的优缺点进行了评价。

【关键词】: AQI; Pearson 相关系数; 多元线性回归

【中图分类号】: TB

【文献标识码】: A

【doi】: 10.19311/j.cnki.1672-3198.2018.17.091

1、引言

随着我国 GDP 保持着高速稳定的增长, 另一方面, 环境污染问题也逐渐引入关注。这主要是因为经济增长方式出现问题, 传统工业发展方式较为守旧落后, 主要能源消费以煤炭石油等为主, 必然导致大量的废气污染物排放。这对建设绿色循环经济产生了一定破坏作用, 不利于生态良好发展, 不利于建设环境友好型社会。废气排放, 管控不当使空气污染的问题更加突出, 尤其是近几年频繁出现的雾霾天气, 在一定程度上对居民身体健康也造成了一定威胁。通过对 PM2.5 的相关影响因素分析和分布与演变规律分析可以找到治理 PM2.5 的更科学便捷的方式, 从而对其他地区的空气污染治理有前车之鉴, 提供了应用范例。除此之外, 社会各界对空气检测监管也十分关注, 相应的政策也得到出台建立, 相关的法律法规及污染管理方式也已明文规定, 特别是 2012 年新修订的《环境空气质量标准》, 首次采用空气质量指数这个概念, 将其作为污染治理中空气检测的指标, 主要包括对空气中的颗粒物、臭氧 O₃ 和一氧化碳 CO 等 6 项的检测。因此, 本文在政策及标准建立的基础上, 从监测指标角度着手进行分析。传统空气环境污染物问题治理对策有: 植树种草、火电厂污染源控制、农作物的控制燃烧处理、工业排放污染的控制、汽车污染源的控制等。三线城市在我国比重占多数, 蚌埠市作为安徽省这个人口大省的第三城市, 其污染雾霾等问题具有代表性、典型性, 对该地区的治理, 对其他三线城市有很好的应用和借鉴效果, 给同类城市提供了预防和处理环境问题的方式该地区空气中 PM2.5 问题的研究具有重要意义。

2、主要指标符号说明

x_i : 表示第 i 个污染物指标的含; y : PM2.5 含量模拟值; b_i : 第 i 个污染物的线性相关系数; R^2 : 拟合优度系数; F : F 检验值; σ_x : x 轴方向扩散系数; σ_y : y 轴方向扩散系数; U : 平均风速。

3、模型的构建

3.1 相关性模型

我们收集到蚌埠市空气质量指数 AQI 中 6 个基本监测指标，得到一天内各个时刻的平均值，利用 SPSS 软件对数据进行相关性处理，得到各指标因素之间的相关性成都，并根据相关性分析结果选取对应指标建立多元线性回归模型。

同时，在对空气质量指数基本监测指标的分析基础上，对于 PM2.5 具有高度相关的检测指标进一步做定量分析。由于选取的指标变量数据较少，我们明显发现，PM2.5 与相应的 5 个指之间含量的变化态势相似，存在一定的线性变化关系。

将原始数据进行标准归一化处理，用 Pearson 相关系数来衡量 AQI 中 6 个基本检测指标的相关性。通过软件运行得到结果，得到 PM2.5 含量与 PM10、NO₂、SO₂、CO 相关系数均大于 0.8，具有高度的相关关系，PM2.5 与臭氧相关系数极弱，因此本文将不考虑臭氧对 PM2.5 的影响，在建立模型的过程中忽略臭氧因素对模型结果的影响。

3.2 多元线性回归模型

多元线性回归模型即含有多个变量的线性回归模型。在理论形式下，可以认为因变量 Y 与自变量 $x_1, x_2, \dots, x_k$ 得到多组观测值中的数据满足： $y_j = b_0 + b_1x_{1j} + b_2x_{2j} + \dots + b_px_{pj} + \varepsilon_j (j = 1, 2, 3, \dots, n)$ ，式中的 ε_j 相互独立且都服零均值同方差的正态分布随机变量。

由 PM2.5 与其他因子的相关性分析结果得到的结果，利用 MATLAB 软件进行多元线性回归拟合，其中自变量为 $x_1, x_2, \dots, x_4$ (PM10, NO₂, SO₂, CO)，因变量为 y，对剔除异常预处理后的数据进行回归拟合，可以得出： $R^2=0.9602$ 、 $F=42.2231$ 、 $p=0.0001$ 。

本文用复相关系数来衡量回归方程与观测值拟合的优良程度，结果中 $R^2>0.85$ ，F 的值较大， $p=0.0001<0.05$ ，以上三点的结果说明该曲线的线性关系显著，该回归模型具有统计学意义，模型成立。

得到该多元线性回归模型的结果如下，从模型中可以看出 PM2.5 与其他四个变量之间的定量分析结果为：每增加 1 单位的 PM10，PM2.5 增加 0.7627 个单位；每增加 1 单位的二氧化氮，PM2.5 增加 0.0819 个单位；每增加 1 单位的二氧化硫，PM2.5 减少 0.8173 个单位；每增加 1 单位的 CO，PM2.5 减少 45.8211 个单位。

$$y = 27.8126 + 0.7627x_1 + 0.0819x_2 - 0.8173x_3 - 45.8211x_4$$

3.3 局斯扩散模型

该理论假定污染物是从某区域附近泄漏，且泄露的排放量 Q (单位 ug/s)，风速大小 u 不随时间、地点和高度的变化而变化。由气体扩散空间任意一点 (x, y, 0)，得到污染源扩散气体的浓度。结合实际，在风力、湿度等天气条件下，合理考虑这些因素对蚌埠市 PM2.5 成因等规律的影响。

查阅资料可得，蚌埠市的共 6 个监测点（蚌埠学院、淮上区政府、百货大楼、工人疗养院、高新区、二水厂），监测搜集到每天 4 个不同时段（上午、中午、晚上、凌晨）的 PM2.5 日均含量分布数据。分别得到蚌埠 PM2.5 的时空分布和规律图，蚌埠市平均湿度、风力和 PM2.5 分别随时间变化规律图，见图 1、图 2。

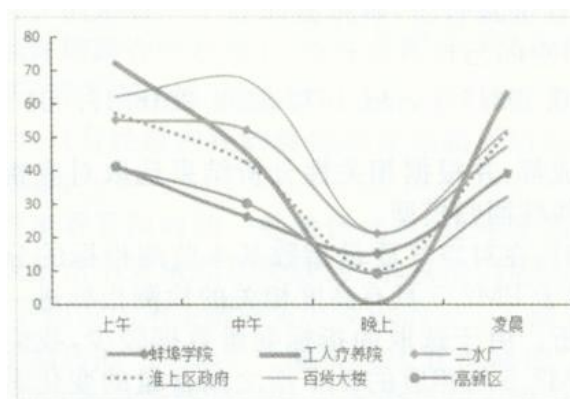


图 1 PM2.5 时空分布图

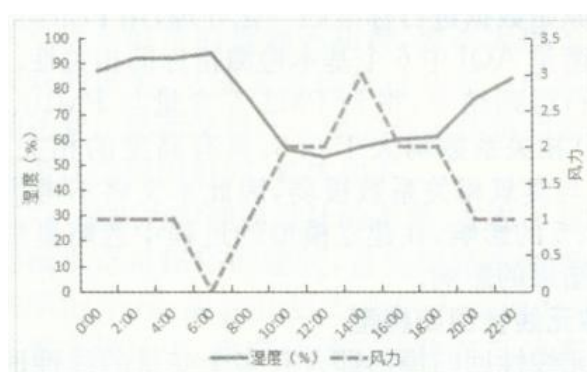


图 2 平均湿度、风力随时间变化规律

通过蚌埠市 PM2.5 时空分布图可见，蚌埠市 6 个环境监测点 PM2.5 含量有基本一致的变化幅度和趋势，随着时间的推移变化（上午—中午—晚上—凌晨），蚌埠市 PM2.5 的时空分布呈现波动下降趋势，从上午至晚上期间，6 个环境监测点的 PM2.5 浓度逐渐降低，至晚上是达到该日最低含量，随后空气中 PM2.5 浓度开始上升。其中，百货大楼、二水厂、高新区的 PM2.5 浓度总体较高，工人疗养院的 PM2.5 浓度变化范围最广，波动最高和最低浓度值也是最显著地，淮上区政府和蚌埠学院的两个监测点 PM2.5 浓度整体较低，每天的波动范围相对较小。结合《环境空气质量标准》，根据环境空气功能区质量要求得到环境空气污染基本项目浓度限值，可以推断出蚌埠市 6 个监测点 24 小时平均 PM2.5 浓度都属于二类区，均在 $35\text{--}75\mu\text{g}/\text{m}^3$ 内分布，满足二级空气质量要求。综合来看，时间和天气情况的变化也会对 PM2.5 浓度产生一定程度的影响，以此类推，季节变化也应是影响空气质量的重要因素，也就是说风力、湿度成为影响 PM2.5 含量的不可忽略因素。

在主要考虑风力和湿度的影响下，本文继续对蚌埠市的 PM2.5 浓度进行双因素分析。结合前面建立的多元线性回归模型，增加变量 x_5 , x_6 (湿度, 风力)，利用 MATLAB 软件求解得到新的多元线性回归模型。得到多元回归曲线的拟合结果： $b_0=38.1092$, $b_1=0.6498$, $b_2=0.1842$, $b_3=0.8166$, $b_4=-0.0307$, $b_5=-0.1443$, $b_6=-2.0151$, $R^2=0.9641$, $F=22.3930$, $p=0.0018$ 。

可见，蚌埠市空气中 PM2.5 浓度与 NO_2 、PM10 浓度有正比关系，与空气中 CO、 SO_2 含量以及风力、湿度大小成反比，且风力对 PM2.5 的含量影响比湿度的影响大。因此，本文建立的多元线性回归模型符合统计学规律，综合最终的模型为：

$$y = 38.1092 + 0.6498x_1 + 0.1842x_2 - 0.8166x_3 - 0.0307x_4 - 0.1443x_5 - 2.0151x_6$$

4、模型的检验

为了检验所建立的模型和运用方法的合理性与科学性，本文运用对比法，将实际检测出来的数据与模型计算出来的模拟值进行比较。得到以下是重新随机采取的多组实际数据，以及对应实际值与模拟值。由于篇幅限制本文仅给出 4 组随机实验数据，具体见表 1。

表 1 模拟值与实际值

x1	x2	x3	x4	x5	x6	模拟值	实际值
25	22	10	450	60	2	23.7374	24
32	23	10	470	71	1	28.284	30
80	22	7	660	75	3	51.2996	53
71	29	6	660	96	1	48.5573	49
...	...	...	...	...	...	...	...

通过与实际值进行比较发现，这四组数据的模拟值与实际值在一定程度上是高度拟合的，拟合精度分别为 99%、94%、97%、99%，均保持在 94%以上的准确率，具有较高的拟合精度，因此该模型是合理的。

5、模型的评价

(1) 优点：利用相关数据软件对数据进行处理并作出各种平面图，使相关描述结果更加直观，更易于理解，运用 MATLAB、SPSS 等数学软件依据模型进行计算，提高工作效率。

(2) 缺点：数据选取相对较少，代表性不够强烈，误差大，对于污染物的选取只选择了 6 项主要的污染因子，对于其他部分（季节因素等）的污染物采取了忽略处理，对最后的模型建立结果有一定的影响因素。在模型的建立中，我们的考虑并不全面，这样无疑会造成计算结果不准确。对于问题一，我们针对蚌埠市 PM2.5 的含量以及其他 5 个因素建立了多元线性回归模型，得出了 PM2.5 的含量与主要指标因素的线性关系，过程中我们并没有考虑到各指标因素之间的内部的相关影响，因此所得到的模型也不够精确，需要进一步改进。

[参考文献]：

[1]张恒斌，杨敏，魏换其. 空气中 PM2.5 问题的研究[J]. 广西师范学院学报：自然科学报，2014，31（4）：97-99.

[2]李勇，宋慧，李玉珍，丁伟利. 西安市空气 PM2.5 问题研究[J]. 黑龙江大学自然科学学报，2014，31（2）：236-237.

[3]王玮，汤大钢，刘红杰. 中国 PM2.5 污染状况和污染特征的研究[J]. 环境科学研究，2000，13（1）：1-5.