

空气中 PM2.5 的评价预测模型

王艳艳, 段红梅

(空军勤务学院基础部)

- 5 **摘要:** 采用灰色关联分析和 MATLAB 软件对空气质量指数 AQI 中 6 个基本监测指标的相关与独立性进行定量分析, 使用逐步回归建立了 PM2.5 粒子浓度与其它 5 项分指标统计相关拟合模型, 并对影响大气污染的气象因子进行了综合分析, 发现 PM2.5 粒子含量与气体污染物浓度存在不同程度的相关性, 且与气象条件亦存在显著的相关关系。通过确定气象要素场和 PM2.5 粒子浓度的关系可进一步研究不同地区的污染物输送和污染源扩散影响的问题。
- 10 **关键词:** 空气质量指数, 灰色关联分析, 逐步回归
- 中图分类号:** X820.2

Prediction and evaluation model of PM2.5 in the air

WANG Yanyan, DUAN Hongmei

(Department of Basic Courses, Air Force Logistics College)

- 15 **Abstract:** Using Grey Relational Analysis and MATLAB software, we did the quantitative analysis of six basic monitoring index about the correlation and independence in the AQI of air quality index. Further, using stepwise regression we established the statistical correlation fitting model of concentrations of PM2.5 particles and other five indicators, and made a comprehensive analysis of the meteorological factors of atmospheric pollution. Finally, we found that PM2.5 particles content and gas pollutant concentrations were different degree of correlation, and the relation with meteorological condition was also significant. By determining the relationship between the meteorological element field and the concentration of PM2.5 particles we can study pollutants transmission and spread of pollution sources in different parts.
- 20 **Key words:** Air quality index, Grey Relational Analysis; Stepwise regression
- 25

0 引言

- 空气质量指数 AQI^[1]的分项监测指标为二氧化硫 SO_2 、二氧化氮 NO_2 、可吸入颗粒物 PM10、细颗粒物 PM2.5、臭氧 O_3 和一氧化碳 CO , 其中 PM2.5 的形成机理和过程比较复杂。结合 2013 年 1 至 8 月份武汉市空气主要污染物的数据, 使用灰色关联分析和 MATLAB
- 30 软件对 AQI 中 6 个基本监测指标的相关与独立性进行定量分析, 并使用逐步回归建立了 PM2.5 粒子浓度与其它 5 项分指标统计相关拟合模型, 目的在于发现 PM2.5 粒子含量与气体污染物浓度之间的相关性, 进而为研究不同地区的污染物输送和污染源扩散影响提供理论依据。

1. AQI 中 6 个基本监测指标的灰色关联分析

- 35 数理统计中的, 因素分析主要采用回归分析、方差分析等。这些分析一般适用于因素较小的线性系统。对于多因素、非线性系统, 则难以处理。但这些方法存在以下问题:
- (1) 要求有大量数据, 数据量少就很难找出统计规律。
- (2) 要求样本服从某个典型的概率分布, 要求各因素数据与系统特征数据间呈线性关系且各因素之间彼此无关。这种要求往往难以满足。
- 40 (3) 可能出现量化结果与定性分析结果不符的现象, 导致系统的关系和规律遭到歪曲和颠倒, 尤其是我国统计数据十分有限, 而且现有数据灰度较大, 再加上人为的原因,

基金项目: 空军勤务学院青年科研基金项目 (BS006)

作者简介: 王艳艳 (1980-), 女, 讲师, 主要研究方向: 随机过程. E-mail: wywith@163.com

许多数据都出现几次大起大落，没有什么典型的分布规律。因此，采用数理统计方法往往难以奏效。

灰色关联分析方法^[2]弥补了采用数理统计方法所导致的缺憾。它对样本量的多少和样本有无规律都同样适用，而且计算量小，十分方便，更不会出现量化结果与定性分析结果不符的情况。

灰色关联分析的基本思想是根据序列曲线几何形状的相似程度来判断其联系是否紧密。曲线越接近，相应序列之间的关联度就越大，反之就越小。

1.1 数据的无量纲化处理

各因素组成的序列，一般来说取值单位不尽相同，而单位不同的数据是无法进行比较的，因此必须把原始数据进行无量纲化处理。无量纲化的方法很多，一般采用数据均值化。数据均值化是用这个数据列中每行数据的平均值去除此行所有数据，以得到新序列。

1.2 确定评估矩阵

关联系数是考虑序列曲线间几何形状的差别，用曲线之间差值的大小作为衡量关联系数的依据。6 项指标集既是参考数列又是比较数列，由以下公式计算第 i 行比较数列与第 j 行参考数列在 k 点的灰色关联系数。

$$L_{ij}(k) = \frac{\min_i \min_k |x_j^{(1)}(k) - x_i^{(1)}(k)| + \zeta \max_i \max_k |x_j^{(1)}(k) - x_i^{(1)}(k)|}{|x_j^{(1)}(k) - x_i^{(1)}(k)| + \zeta \max_i \max_k |x_j^{(1)}(k) - x_i^{(1)}(k)|} \quad (1)$$

式中， ζ ——分辨系数，在 $[0, 1]$ 中取值，一般取 0.5；

$\min_i \min_k |x_j^{(1)}(k) - x_i^{(1)}(k)|$ ——两级最小差； $\max_i \max_k |x_j^{(1)}(k) - x_i^{(1)}(k)|$ ——两级最大差。

1.3 计算关联度

因为关联系数列中的数据很多，信息过于分散，比较不便，所以有必要将各个时刻的关联系数集中为一个值。求关联系数列的平均值就是将这种信息集中处理的一种方法。这个平均值就是关联度。

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m L_{ij}(k) \quad (2)$$

1.4 优势分析

鉴于要分析 6 个指标相互之间关联度的大小，故灰色关联度分析的参考数列与比较数列都是 6 个指标的数据组成的数列。则参考数列和比较数列的关联度可构成关联矩阵。矩阵中的每一行表示同一母因素对不同子因素的影响；每一列表示不同母因素对同一子因素的影响。因此就可以根据矩阵中的各个行与各个关联度的大小来判断子因素与母因素的作用，分析哪些是主要影响因素，哪些是次要影响因素。起主要影响的因素称为优势因素，因此相应的就有优势母因素与优势子因素。

1.5 模型的求解

依据武汉市一个监测点 2013 年 1 月 1 日至 1 月 31 日的数据，利用 MATLAB 求出 6 项分指标之间的相关度，如表 1 所示。

表1 AQI 中 6 个基本监测指标的相关度

	PM2.5	SO ₂	NO ₂	PM10	CO	O ₃
PM2.5	1.0000	0.8396	0.8799	0.9311	0.8824	0.6837
SO ₂	0.8480	1.0000	0.9169	0.8614	0.8521	0.6893
NO ₂	0.8783	0.9106	1.0000	0.8989	0.8375	0.6864
PM10	0.9307	0.8527	0.8997	1.0000	0.8581	0.6880
CO	0.8877	0.8506	0.8465	0.8652	1.0000	0.6896
O ₃	0.6971	0.6893	0.7033	0.7026	0.6920	1.0000

80 由表 1 可以看出 6 个基本监测指标的相关与独立性，从中可以看出 O_3 独立性最高，与其他因素的关联度较小。从第二列可以看出，与 PM2.5 关联度由强至弱依次为 PM10、 NO_2 、 CO 、 SO_2 、 O_3 。同理可以分析出其他因素之间的关联强度。

2 .PM2.5 浓度综合相关逐步回归模型

85 根据表 1 中 PM2.5 含量与 SO_2 、 NO_2 、PM10、PM2.5、 O_3 和 CO 的相关特征，表明气体污染物与 PM2.5 浓度密切相关^[3]，我们建立逐步回归模型试图通过其他污染物拟合 PM2.5 浓度。自变量 SO_2 (x_1)、 NO_2 (x_2)、PM10 (x_3)、 CO (x_4) 和 O_3 (x_5)，变量长度为 2013 年 1 月 1 日至 1 月 31 日共 32 个样本，建立 PM2.5 (y) 的多元回归方程。运用逐步回归方法，从包含全部变量的回归方程中逐步剔除不显著的因子 x_5 ，最优回归方程为：

90
$$y = -39.3022 - 1.13019x_1 + 0.95817x_2 + 1.44782x_3 + 1.4289x_4 \tag{3}$$

复相关系数 R 为 0.920086，剩余标准差 $RMSE$ 为 22.1766，显著性检验 F 为 74.8371。



图 1 2013 年 1 月 1 日到 1 月 31 日逐步回归

95 为了检验回归方程，运用式 (3) 拟合了从 2013 年 1 月 1 日到 1 月 31 日的 PM2.5 浓度，从图 2 中可见，拟合效果不错（实线为实测数据，星点表示拟合值）。

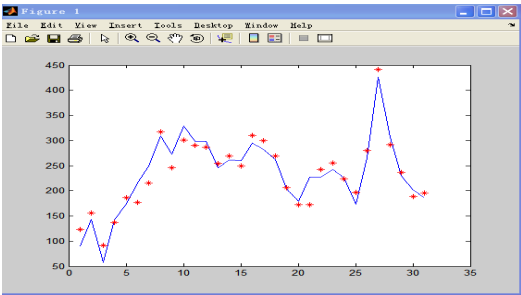


图 2 2013 年 1 月 1 日到 1 月 31 日的 PM2.5 观测及气体污染物相关模型拟合、实测曲线

从逐步回归结果来看，影响 PM2.5 的因子中权重较大的为 PM10、 CO ，其次为 SO_2 、 NO_2 。从图 2 可见，PM10， CO ， SO_2 ， NO_2 建立的回归模型拟合 PM2.5 的效果较好，

模拟的趋势基本与 PM2.5 相似，但在有些地方。例如 1 月 10 日的值与实测偏差较大，需要进一步增加统计样本和预测因子才能提高预报效果，而此时空气质量指数级别为六级，为严重污染^[4]。

采用独立样本估测检验，即用同样的逐步回归方法建立的回归方程拟 2013 年 7 月 1 日到 7 月 31 日的 PM2.5 的含量可见，逐步回归模型计算估测值和观测实况的变化趋势差距较大。

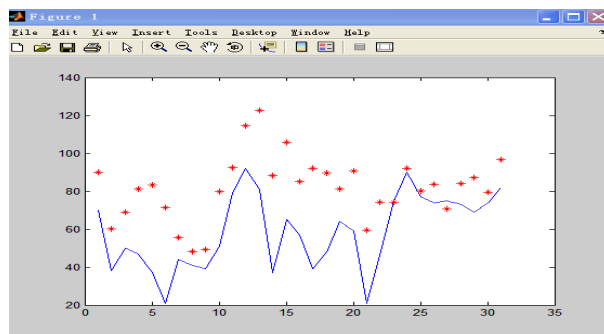


图 3 运用 1 月份模型拟合 7 月 1 日到 7 月 31 日的数据

于是我们使用 2013 年 7 月 1 日到 7 月 31 日的 6 种实测数据指标，采用逐步回归，剔除不显著的因子 x_1 、 x_2 、 x_5 ，得出最优回归方程为：

$$y = -29.6533 + 0.642647x_3 + 2.68394x_4 \quad (4)$$

复相关系数 R 为 0.830825，剩余标准差 $RMSE$ 为 4.86712，显著性检验 F 为 68.7544。

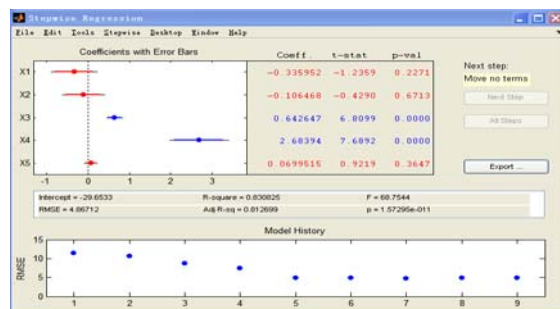


图 4 2013 年 7 月 1 日到 7 月 31 日逐步回归

为了检验回归方程，运用式 (4) 拟合了从 2013 年 7 月 1 日到 7 月 31 日的 PM2.5 浓度，从图 5 中可见，拟合效果不错（实线为实测数据，星点表示拟合值）。

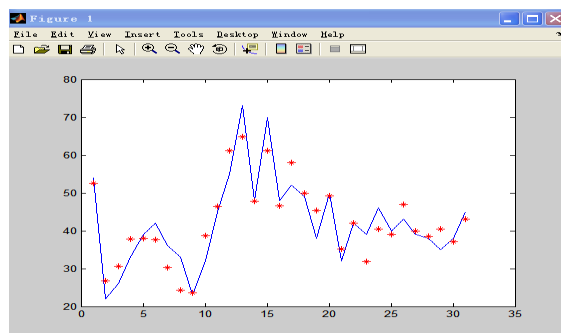


图 5 2013 年 7 月 1 日到 7 月 31 日的 PM2.5 观测及气体污染物相关模型拟合、实测曲线

3. 结论

基于对 2013 年 1 月 1 日到 1 月 31 日的 PM2.5 含量与 2013 年 7 月 1 日到 7 月 31 日的 PM2.5 含量的分析，可以得出 PM2.5 粒子随着季节的变化，变化趋势差距较大，一月

份的空气质量指标几乎全为重度污染，而七月份空气质量指标大多为良好，此时 PM2.5 含量主要与 PM10 和 CO 相关，与 SO_2 ， NO_2 ， O_3 相关性不强，两个逐步回归模型不同。究其原因，由于风力、湿度等天气和季节因素的影响^[5]，PM2.5 的发生和演变规律是不一样的。通过针对不同季节建立逐步回归模型，我们证明了用 PM10，CO， SO_2 ， NO_2 污染物拟合的 PM2.5 含量具有一定的可信度，基本能估算 PM2.5 粒子含量的变化趋势。

[参考文献] (References)

- [1] 环境保护部, HJ 633-2012.环境空气质量指数 (AQI) 技术规定 (试行) [S]: 北京: 中华人民共和国国家环境保护标准, 2012.
- [2] 徐卫国, 田伟利, 张清宇, 丁淑英, 郭慧. 灰色关联分析模型在环境空气质量评价中的修正及应用研究[J]. 中国环境监测, 2006.22 (3): 63-66.
- [3] 周丽, 徐祥德, 丁国安. 北京地区气溶胶 PM2.5 粒子浓度的相关因子及其估算模型[J]. 气象学报, 2003.12(6):761-767.
- [4] 蒲维维, 赵秀娟, 张小玲. 北京地区夏末秋初气象要素对 PM2.5 污染的影响[J]. 应用气象学报, 2011.12(6):716-722.
- [5] 高健, 柴发合. 我国大气颗粒物污染研究及其对控制对策的支撑[J]. 环境保护, 2014, 11: 32-34.