

参赛密码 \_\_\_\_\_  
(由组委会填写)

第十届华为杯全国研究生数学建模竞赛

学 校      三峡大学

参赛队号      11075008

队员姓名	1.尹辉
	2.雷林
	3.姚远



## 第十届华为杯全国研究生数学建模竞赛

题 目 空气中 PM2.5 问题的研究

### 摘 要：

本文建立了相关性分析模型，灰色关联度模型，混合回归模型，高斯烟羽模型，分期治理最优化模型等模型，通过定量与定性分析的方法，从相关因素、分布与演变、控制管理三个方面，对 PM2.5 进行了深入的研究与探讨。

针对问题一：

1、以六种污染物为相关量，建立了相关性分析模型。将附件 1 的数据代入模型中，求得的结果表明：相关性最高的指标组是 PM2.5 和 CO，其相关系数为 0.82，相关性最低的指标组是 NO<sub>2</sub> 和 O<sub>3</sub>，其相关系数为 -0.063，即独立性最强。

2、以 PM2.5 为参考数列，其它 5 种污染物为比较数列，建立了灰色关联度分析模型，将附件 1 的数据代入模型中，求得的结果表明：PM2.5 与其它五种污染物的平均关联度为 0.80，可见相关性较高。以 PM2.5 为因变量，其它五种污染物为自变量，先后建立多元线性回归模型和混合回归模型，模型结果表明：混合回归模型更优（相关系数由 0.85 增加为 0.89）。

3、利用互联网收集到全国 76 个城市 AQI 的 6 个监测指标和湿度数据，以 PM2.5 为因变量，其它五种污染物为自变量，建立了线性回归模型。将湿度指标也考虑为自变量后，回归模型的相关系数得到明显提升（由 0.88 提升到 0.92），表明湿度与 PM2.5 存在较强的相关性。

针对问题二：

1、通过伽马分布预测出 2013 年西安市 13 个地区 PM2.5 的全部数据，利用 MATLAB 画出了 PM2.5 时空分布图，并得出了三种分布规律。考虑到各地区“污染程度”为较模糊的概念，因此建立了模糊综合评价模型，对每个地区的污染程度进行了综合评价，模型结果表明：高压开关厂地区污染指数最高（94.39），阎良区地区污染指数最低（75.27）。

2、对 PM2.5 受风力影响在大气中扩散的问题，建立了高斯烟羽模型进行分析。假设风向为正北向，风速为  $40\text{ km/h}$ ，扩散系数  $\sigma$  为 0.00001，排放源有效高度为  $50\text{ m}$ ，初始浓度为各监测站点的最高值，对模型进行求解，得到 13 个地区的扩散数据（仅列出高压开关厂地区的上风处扩散数据）：

距离 (km)	0	2	4	6	8	10	11	12
PM2.5 浓度 ( $\mu\text{g}/\text{m}^3$ )	1000	850	703	480	292	108	21	0
时间 (min)	0	20	36	58	82	106	115	121

3、将 PM2.5 污染程度划分为重度污染、中度污染和轻度污染（安全）三个级别，同样假设风向为正北向，风速为  $40\text{ km/h}$ ，扩散系数  $\sigma$  为 0.00001，排放源有效高度为  $50\text{ m}$ ，初始浓度为某站点最高值 2 倍，利用高斯烟羽模型求出 13 个地区的扩散数据，结合各个地区之间的距离，得到了各地区的污染程度。以高压开关厂为例，得到结果如下：

轻度污染（安全）	中度污染	重度污染
阎良区 临潼区 广运潭 纺织城 长安区	市体育馆 曲江文化集团 兴庆小区	其它 地区

4、利用互联网收集到了福岛第一核电站的放射性物质扩散数据，将放射性物质与 PM2.5 扩散数据进行对比，发现两者的扩散规律总体一致，从而验证了模型的合理性。利用物质的自身沉降作用和雨水吸附作用对高斯烟羽模型进行了修正，得到了修正后更为一般性的扩散模型。

针对问题三：

1、根据以往空气质量的变化趋势及 PM2.5 当前年平均浓度 ( $280\text{ }\mu\text{g}/\text{m}^3$ )，预测出在不治理的情况下，五年后 PM2.5 年平均浓度为  $324\text{ }\mu\text{g}/\text{m}^3$ 。然后采用分期治理的思想，将五年的治理时期分为前期、中期和后期。考虑到实际治理进度的变化规律与柯西分布函数相似，通过计算机模拟找出了最理想的柯西分布函数，由此确定了 PM2.5 的分期治理计划：

时期	前期（准备期）	中期（治理期）			后期（稳固期）
年次	第一年	第二年	第三年	第四年	第五年
治理量百分比 (%)	9.6	32.7	38.8	15.3	3.6
治理量	31.1	105.94	125.71	49.57	11.66

2、以专项治理总费用最小为目标，建立了最优化模型。然后同样采用分期治理的思想，利用柯西分析函数对最优化模型进行修正，得到修正后的分期治理最优化模型。以数据 1 中 PM2.5 年平均浓度 ( $82\text{ }\mu\text{g}/\text{m}^3$ ) 为初始浓度，假设最终治理目标为  $30\text{ }\mu\text{g}/\text{m}^3$ ，对模型进行求解，得到总费用为 3.38（百万），逐年治理计划如下表：

时期	前期（准备期）	中期（治理期）			后期（稳固期）
年次	第一年	第二年	第三年	第四年	第五年
治理量 $\mu\text{g}/\text{m}^3$	4.15	13.65	19.1	9.85	3.25
费用（百万元）	0.086	0.932	1.824	0.485	0.053

将模型得到的治理计划与实际环境治理计划进行对比，发现两者的治理进度变化规律总体一致，从而验证了模型的合理性。

关键字：相关性分析模型 灰色关联度模型 高斯烟羽模型 柯西分布函数

# 1 问题重述

## 1.1 问题背景

“2013 年初以来，中国发生大范围持续雾霾天气。据统计，受影响面积约占国土面积的 1/4，受影响人口约 6 亿人”。对空气质量监测，预报和控制等问题，国家和地方政府均制定了相应政策、法规和管理办法。调整了环境空气功能区分类等。AQI 是无量纲指数，它的分项监测指标为 6 个基本监测指标（二氧化硫  $\text{SO}_2$ 、二氧化氮  $\text{NO}_2$ 、可吸入颗粒物  $\text{PM}_{10}$ 、细颗粒物  $\text{PM}_{2.5}$ 、臭氧  $\text{O}_3$  和一氧化碳  $\text{CO}$  等 6 项）。新标准中，首次将产生灰霾的主要因素——对人类健康危害极大的细颗粒物  $\text{PM}_{2.5}$  的浓度指标作为空气质量监测指标。

## 1.2 问题的提出

对细颗粒物  $\text{PM}_{2.5}$  及相关的因素的统计数据还太少，对  $\text{PM}_{2.5}$  的客观规律也了解得很不够。我们必须千方百计利用现有的数据开展研究，同时新课题、探索性研究、“灰箱问题”也有可能成为数学建模爱好者的用武之地。

## 1.3 本文所需解决的问题

### 一、 $\text{PM}_{2.5}$ 的相关因素分析

请依据附件 1 或附件 2 中的数据或自行采集数据，利用或建立适当的数学模型，对 AQI 中 6 个基本监测指标的相关与独立性进行定量分析，尤其是对其中  $\text{PM}_{2.5}$ （含量）与其它 5 项分指标及其对应污染物（含量）之间的相关性及其关系进行分析。如果你们进而发现 AQI 基本监测指标以外的、与  $\text{PM}_{2.5}$  强相关的（可监测的）成分要素，请陈述你们的方法、定量分析结果、数据及来源。

### 二、 $\text{PM}_{2.5}$ 的分布与演变及应急处理

请依据附件 2、附件 3 中的数据或自行采集某地区的数据，通过数学建模探索完成以下研究：

1、描述该地区内  $\text{PM}_{2.5}$  的时空分布及其规律，并结合环境保护部新修订的《环境空气质量标准》地区进行污染评估。

2、建立能够刻画该地区  $\text{PM}_{2.5}$  的发生和演变（扩散与衰减等）规律的数学模型，合理考虑风力、湿度等天气和季节因素的影响，并利用该地区的数据进行定量与定性分析。

3、假设该地区某监测点处的  $\text{PM}_{2.5}$  的浓度突然增至数倍，且延续数小时，请建立针对这种突发情形的污染扩散预测与评估方法。并以该地区  $\text{PM}_{2.5}$  监测数据最高的一天为例，在全地区  $\text{PM}_{2.5}$  浓度最高点处的浓度增至 2 倍，持续 2 小时，利用你们的模型进行预测评估，给出重度污染和可能安全地区。

4、采用适当方法检验你们模型和方法的合理性，并根据已有研究成果探索  $\text{PM}_{2.5}$  的成因、演变等一般性规律。

### 三、空气质量的控制管理

数据 1 所在地区的环境保护部门考虑治污达标的紧迫性和可行性，在未来五年内，拟采取综合治理和专项治理相结合的逐年达到治理目标的方案。请考虑以下问题：

1、该地区目前 PM2.5 的年平均浓度估计为 280 (单位为  $\mu g / m^3$ )，要未来五年内逐年减少 PM2.5 的年平均浓度，最终达到年终平均浓度统计指标 35 (单位为  $\mu g / m^3$ )，请给出合理的治理计划，即给出每年的全年年终平均治理指标。

2、据估算，综合治理费用，每减少一个 PM2.5 浓度单位，当年需投入一个费用单位 (百万元)，专项治理投入费用是当年所减少 PM2.5 浓度平方的 0.005 倍 (百万元)。请你为数据 1 所在地区设计有效的专项治理计划，使得既达到预定 PM2.5 减排计划，同时使经费投入较为合理，要求你给出五年投入总经费和逐年轻费投入预算计划，并论述该方案的合理性。

## 2 模型假设与符号说明

### 2.1 模型假设

- 假设 1：连续排放时 PM2.5 排放的速率恒定；
- 假设 2：PM2.5 在平整、无障碍的地面上空扩散；
- 假设 3：地面及地标地物对 PM2.5 无吸收；
- 假设 4：风向水平，风速和风向恒定；
- 假设 5：风速大于无风情况下放射性气体扩散的速度；
- 假设 6：排放源的源强是连续且均匀的，初始时刻 PM2.5 的浓度、温度呈均匀分布。

### 2.2 符号说明

表 2-1 符号说明

参数符号	符号说明
$X_i(i=1,2,\cdots,6)$	分别代表 SO <sub>2</sub> 、NO <sub>2</sub> 、PM10、CO、O <sub>3</sub> 、PM2.5 的监测指标
$r_i(i=1,2,\cdots,5)$	分别代表 SO <sub>2</sub> 、NO <sub>2</sub> 、PM10、CO、O <sub>3</sub> 对 PM2.5 的关联度
$x_i$	PM2.5 第 $i$ 年减少 (治理) 量
$y$	代表样本 PM2.5 的含量
$\bar{u}$	平均风速
$d$	PM2.5 物质粒子直径
$\sigma_i$	扩散系数
$\alpha$	PM2.5 年平均增长率
$c_0$	PM2.5 初始浓度

$c_j$	第 $j$ 年 PM2.5 浓度
$Z$	五年专项治理投入总经费
$S_i$	每年专项治理投入费用
$k_i$	第一年到第 $i$ 年 PM2.5 的累计减少量与 PM2.5 初始量之比

### 3 问题分析

#### 3.1 针对问题一

1、对 AQI 中 6 个基本监测指标的相关与独立性进行定量分析：这个问题是典型的相关性分析问题，因此可以建立相关性分析模型来求解。

2、对 PM2.5 含量与其它 5 项污染物含量之间的相关性及其关系进行分析：首先可以利用灰色关联度模型分析 PM2.5 含量与其它 5 项污染物含量的关联度，然后可以以 PM2.5 含量为因变量，其它污染物含量为自变量，先后建立多元线性回归方程和混合回归模型，从而确定 PM2.5 与其它污染物的定量关系。

3、PM2.5 其它相关因素的探寻：题目中提到了二氧化硫（SO<sub>2</sub>），二氧化氮（NO<sub>2</sub>），一氧化碳（CO）是在一定环境条件下形成 PM2.5 前的主要气态物体，可以认为这些污染物在空气中通过物理和化学反应转化为了 PM2.5，而水一般都是化学反应不可缺少的物质，因此有理由相信空气湿度与 PM2.5 有关。首先可以通过互联网收集到湿度数据，然后对数据进行分析，说明湿度与 PM2.5 的相关性。

#### 3.2 针对问题二

1、描述该地区 PM2.5 的时空分布及其规律，并地区进行污染评估：首先可以利用 MATLAB 画出该地区 PM2.5 的时空分布图，并由时空分布图分析 PM2.5 的分布规律。由于“污染程度”属于比较模糊的概念，因此可以利用模糊综合评价模型对各站点进行综合污染评估。

2、建立能够刻画该地区 PM2.5 的扩散规律的数学模型，并利用该地区的数据进行定量与定性分析：考虑风速的情况下，空气中污染物的扩散，可以采用高斯烟羽模型进行分析，利用此模型求出西安市 13 个地区 PM2.5 的扩散数据，从而对该地区的污染情况进行定量与定性分析。

3、全地区 PM2.5 含量最高点处的含量增至 2 倍，持续 2 小时，利用模型进行预测评估，给出重度污染和可能安全地区：首先根据新修订的《环境空气质量标准》将 PM2.5 划分为重度污染，中度污染，轻度污染（安全）三个等级，然后利用模型求出各个地区的扩散数据，最后结合各地区两两距离，判断各个地区的污染情况。

4、采用适当方法检验模型和方法的合理性，并根据已有研究成果探索 PM2.5 的成因、演变等一般性规律：由于放射性物质在大气中的扩散规律与 PM2.5 的扩散规律相似，因此可以将模型求出的 PM2.5 的扩散数据与已有的放射性物质扩散数据进行对比，从而检验结果和模型的合理性。对 PM2.5 扩散

一般性规律的探索，可以利用 PM2.5 自身沉降作用、雨水吸附作用等一般性因素对高斯扩散模型进行修正，从而得出 PM2.5 更为一般性的扩散规律。

3.3 针对问题三

1、给出合理的 PM2.5 治理计划：考虑到实际环境治理过程中，往往是分期进行治理工作的，而且治理进度是先慢变快，再由快变慢（这种变化规律与柯西分布函数的变化规律很相似）。因此可以将 PM2.5 的治理计划分为前期（准备期），中期（治理期），后期（稳固期）三个时期，并由柯西分布函数对治理计划进行模拟，得出与实际环境治理过程相符的分期治理计划。

2、给出合理的 PM2.5 专项治理计划：首先可以以专项治理费用最小为目标，建立最优化模型，同样可以引入分期治理的思想，利用柯西分布函数对模型最优化模型进行修正，得到修正后的分期治理最优化模型。

4 数据分析

4.1 数据收集

利用互联网收集到 2013 年 9 月 20 日上午 10 点整，全国 76 个城市 AQI 的 6 种污染物浓度数据及湿度数据，如表 4-1（仅列出部分城市数据，详细结果见附录一）

表 4-1 76 个城市 AQI 和湿度数据

城市	PM2.5 ( $\mu\text{g}/\text{m}^3$ )	PM10 ( $\mu\text{g}/\text{m}^3$ )	CO ( $\text{mg}/\text{m}^3$ )	NO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	O <sub>3</sub> ( $\mu\text{g}/\text{m}^3$ )	SO <sub>2</sub> ( $\mu\text{g}/\text{m}^3$ )	湿度 (%)
海口	7	16	0.375	5	10	2	66
福州	12	32	0.467	20	58	6	28
成都	20	29	0.704	44	19	9	93
舟山	14	36	0.701	4	73	2	72
厦门	14	41	0.34	14	79	2	50
丽水	19	44	0.431	19	68	10	30
拉萨	16	33	1.624	16	35	13	16
惠州	28	52	0.707	22	69	18	66
上海	21	45	0.566	14	83	10	62
张家口	11	25	0.298	20	71	20	35
嘉兴	25	42	0.354	26	92	12	63
温州	24	47	0.624	22	68	8	68
中山	38	72	0.911	37	35	31	61
东莞	44	82	0.961	44	65	55	62
盐城	18	57	0.248	11	61	12	57
珠海	20	44	1.133	20	55	10	63
台州	16	47	0.531	21	61	9	79
广州	50	72	0.828	30	46	26	58
深圳	38	74	1.114	41	46	25	68
江门	32	59	1.065	22	56	27	57

## 4.2 数据补全

### 4.2.1 附件 1 数据补全

#### 1、利用回归方程补全数据

附件 1 中 2013 年 4 月 29 日和 6 月 7 日的 PM10 数据缺失。以 PM2.5 为自变量，其他五种污染物为因变量，建立了多元线性回归方程，并用附件 1 中未缺少的数据对模型进行求解，得到回归方程（ $R = 0.85$ ），利用此模型对缺失数据进行预测，结果如表 4-2：

表 4-2 PM10 数据预测

日期	PM10
2013 年 4 月 29 日	64
2013 年 6 月 7 日	42

#### 2、利用伽马分布补全数据

利用卡方拟合优度检验对附件 1 中 PM2.5 数据进行检测，发现数据服从伽马分布：

表 4-3 卡方拟合检验结果

附件 1 中 PM2.5 卡方拟合优度检验	
假设检验	
零假设	服从伽马分布
自由度	4
卡方统计量	1.8
P 值	0.6
显著性水平	0.1
结果	接受零假设

利用伽马分布对附件 1 中 2013 年 8 月 27 日至 2013 年 12 月 31 日的 PM2.5 数据进行预测，预测结果如表 4-4（仅列出部分数据，详细数据见附录二）：

表 4-4 附件 1 全年 PM2.5 数据

时期	PM2.5 含量
2013 年 4 月 27 日	144
2013 年 4 月 28 日	104
2013 年 4 月 29 日	49
2013 年 4 月 30 日	54
2013 年 5 月 1 日	197
2013 年 5 月 2 日	189
2013 年 5 月 3 日	61
2013 年 5 月 4 日	156
2013 年 5 月 5 日	109
2013 年 5 月 6 日	72
2013 年 5 月 7 日	50
2013 年 5 月 8 日	114
2013 年 5 月 9 日	19
2013 年 5 月 10 日	25



#### 4.2.2 附件 2 数据补全

同样利用卡方拟合优度检验对附件 2 中 13 个站点的 PM2.5 数据进行检测，发现数据也服从伽马分布：

表 4-5 卡方拟合优度检验

附件 2 中 PM2.5 卡方拟合优度检验	
假设检验	
零假设	服从伽马分布
自由度	4
卡方统计量	1.6
P 值	0.3
显著性水平	0.1
结果	接受零假设

利用伽马分布对附件 2 中全年的 PM2.5 数据进行预测，预测结果如表 4-6（仅高压开关厂的部分数据，详细数据见附录三）：

表 4-6 高压开关厂 PM2.5 全年数据

时期	PM2.5 含量
2013 年 4 月 27 日	80.6
2013 年 4 月 28 日	143.7
2013 年 4 月 29 日	181
2013 年 4 月 30 日	75.8
2013 年 5 月 1 日	69.4
2013 年 5 月 2 日	164
2013 年 5 月 3 日	79.8
2013 年 5 月 4 日	137.4
2013 年 5 月 5 日	179
2013 年 5 月 6 日	131.8
2013 年 5 月 7 日	331
2013 年 5 月 8 日	55
2013 年 5 月 9 日	286
2013 年 5 月 10 日	164

#### 4.3 数据处理

##### 4.3.1 附件 1 中 PM2.5 分指数数据转化为浓度数据

利用参考文献<sup>[2]</sup>给出的污染项目的空气质量分指数计算公式，将附件 1 中各种污染物的分指数数据转化为浓度数据，结果如表 4-7（仅列出部分数据，详细数据见附录二）：

表 4-7 附件 1 的浓度数据

时期	PM 分指数	PM 浓度
2013 年 1 月 1 日	90	67
2013 年 1 月 2 日	143	109.4
2013 年 1 月 3 日	58	41.4
2013 年 1 月 4 日	142	108.6
2013 年 1 月 5 日	175	132.5
2013 年 1 月 6 日	215	165
2013 年 1 月 7 日	250	200
2013 年 1 月 8 日	309	259
2013 年 1 月 9 日	273	223
2013 年 1 月 10 日	329	279
2013 年 1 月 11 日	299	249
2013 年 1 月 12 日	299	249
2013 年 1 月 13 日	246	196
2013 年 1 月 14 日	261	211
2013 年 1 月 15 日	260	210

#### 4.3.2 附件 2 中 PM<sub>2.5</sub> 等级划分及频数统计

根据新修订的《环境空气质量标准》将 PM<sub>2.5</sub> 划分为 5 个等级，如表 4-8：

表 4-8 PM<sub>2.5</sub> 等级划分

等级	一级	二级	三级	四级	五级
PM <sub>2.5</sub>	0-75	76-115	116-150	151-250	>250

根据 PM<sub>2.5</sub> 的 5 个等级，对附件 2 中 PM<sub>2.5</sub> 数据进行频数统计，统计结果为：

表 4-9 附件 2 中 PM<sub>2.5</sub> 等级频数统计

等级 监测站	一级	二级	三级	四级	五级
高压开关厂	15	54	51	114	108
兴庆小区	24	63	63	84	84
纺织城	21	66	78	75	93
小寨	42	72	69	69	81
市人民体育场	30	45	45	90	90
高新西区	27	51	48	93	114
经开区	27	57	57	81	81
长安区	36	72	66	75	84
阎良区	30	63	39	108	93
临潼区	39	51	57	78	72
曲江文化集团	42	66	72	63	87
广运潭	18	51	60	102	93
草滩	18	60	36	123	99
全市平均	27	57	69	93	102

### 4.3.3 附件 2 中 13 个站点的空间分布及距离确定

#### 1、13 个站点的空间分布图

利用 google 地图可以画出西安市 13 个站点的空间分布图，如图 4-1（A-M 分别高压开关厂、高新西区、小寨、市体育馆、曲江文化集团、兴庆小区、纺织城、广运潭、草滩、经开区、临潼区、阎良区、长安区）



图 4-1 13 个站点（地区）空间分布图

由空间分布图及西安市实际地图，可以发现 13 个站点均匀分布在西安市 9 个行政区中，因此可以利用这 13 个站点的数据分别对西安市 13 个地区进行分析。详细分布情况见表 4-10：

表 4-10 13 个站点行政区的分布

行政区	监测站点	地区面积（平方千米）
莲湖区	高压开关厂	38
雁塔区（西北）	高新西区	152
雁塔区（东南）	小寨	152
新城区（西北）	市体育馆	31
新城区（西南）	曲江文化集团	31
碑林区	兴庆小区	22
灞桥区（南区）	纺织城	322
灞桥区（北区）	广运潭	322
未央区（东北）	草滩	261
未央区（西南）	经开区	261
临潼区	临潼区	898
阎良区	阎良区	240
长安区	长安区	1583

2、13 个站点的距离确定。

利用 google 地图距离计算公式，结合 13 个站点的空间分布图，可以算出 13 个站点（地区）的两两距离，结果如表 4-11（以高压开关厂为例）：

表 4-11 各站点距高压开关厂的距离

地区	监测站点	距离（千米）
莲湖区	高压开关厂	0
雁塔区（西北）	高新西区	3.28
雁塔区（东南）	小寨	6.29
新城区（西北）	市体育馆	4.13
新城区（西南）	曲江文化集团	5.2
碑林区	兴庆小区	5.14
灞桥区（南区）	纺织城	14.47
灞桥区（北区）	广运潭	14.82
未央区（东北）	草滩	2.8
未央区（西南）	经开区	4.64
临潼区	临潼区	30.71
阎良区	阎良区	52.7
长安区	长安区	11.82

#### 4.3.4 浓度增加为 2 倍后污染程度等级划分

根据新修订的《环境空气质量标准》将 PM2.5 污染等级划分为三个等级，如表 4-12：

表 4-12 PM2.5 污染等级划分

污染程度	轻度污染（安全）	中度污染	重度污染
PM2.5 浓度（ $\mu\text{g} / \text{m}^3$ ）	0-350	350-500	500-1000

## 5 问题一的解答

### 5.1 AQI 中 6 个监测指标的相关与独立性分析

#### 5.1.1 相关性分析模型的建立

由协方差方程  $Cov(X, Y) = E\{[X - E(X)][Y - E(Y)]\}$  可知其相关系数：

$$\rho = \frac{Cov(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \quad (5-1)$$

进而可建立 6 个监测指标的相关性分析模型：

$$\rho = \frac{Cov(X_i, X_j)}{\sqrt{D(X_i)}\sqrt{D(X_j)}} \quad (5-2)$$

其中： $i$ 或 $j=1, 2, \dots, 6$

$X_1$  代表二氧化硫(SO<sub>2</sub>)的监测指标；

$X_2$  代表二氧化氮(NO<sub>2</sub>)的监测指标；

$X_3$  代表可吸入颗粒物(PM10)的监测指标；

$X_4$  代表一氧化碳(CO)的监测指标;

$X_5$  代表臭氧 ( $O_3$ ) 的监测指标;

$X_6$  代表 PM2.5 的监测指标。

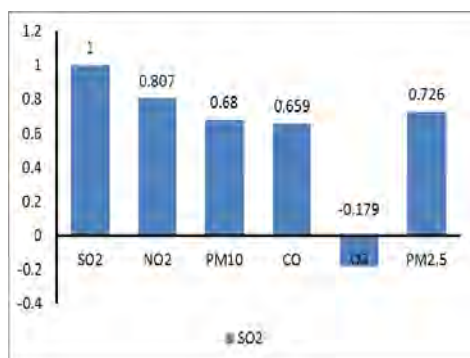
### 5.1.2 相关性分析模型的求解

用 MATLAB 对模型进行求解, 得出结果如表 5-1:

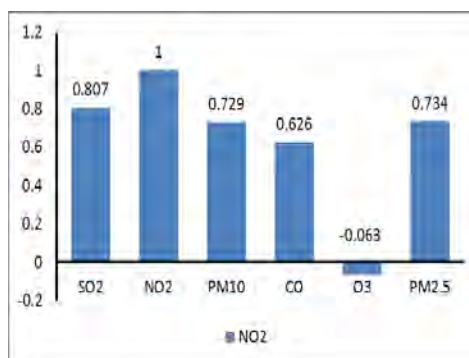
表 5-1 相关性模型求解结果

指标	SO <sub>2</sub>	NO <sub>2</sub>	PM10	CO	O <sub>3</sub>	PM2.5
SO <sub>2</sub>	1	0.807	0.680	0.659	-0.179	0.726
NO <sub>2</sub>	0.807	1	0.729	0.626	-0.063	0.734
PM10	0.680	0.729	1	0.588	-0.066	0.780
CO	0.659	0.626	0.588	1	0.587	0.822
O <sub>3</sub>	-0.179	-0.063	-0.067	-0.381	1	-0.352
PM2.5	0.726	0.734	0.780	0.822	-0.352	1

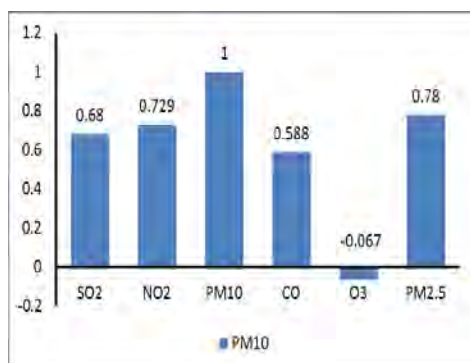
由表 5-1 得出各污染指标相关性的柱状示意图 5-1:



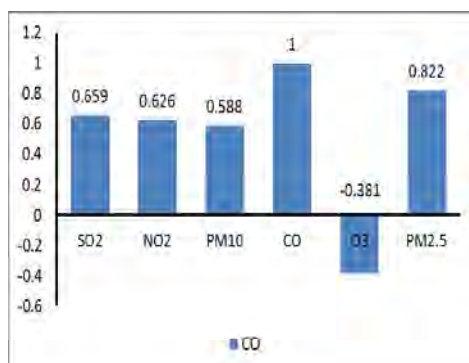
(a)



(b)



(c)



(d)

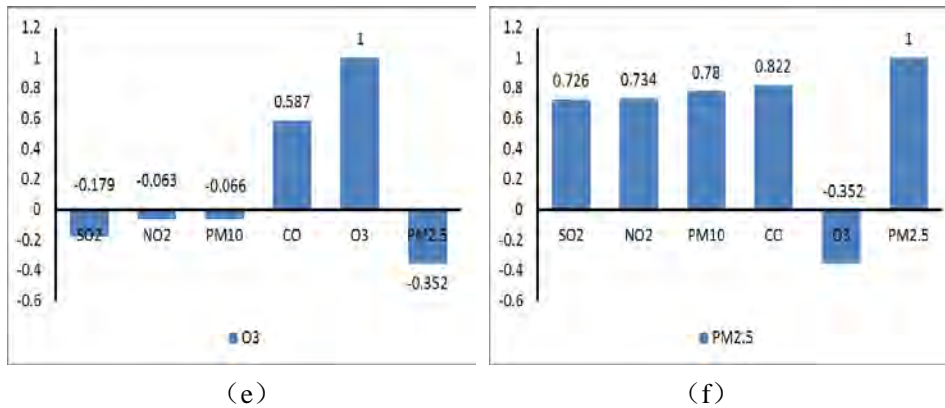


图 5-1 (a)-(f) 各污染指标相关性的柱状示意图

### 5.1.3 结果分析

由相关性分析模型的结果可以得到如下结论：

#### (1) 相关性高的指标组

联系最为紧密的监测指标组依次是：(PM2.5, CO), (SO2, NO2), (PM2.5, PM10), (PM2.5, NO2), (PM2.5, SO2), 相关系数均在 0.7 以上。

#### (2) PM2.5 与其它指标相关性最高

PM2.5 监测指标与 CO, PM10, NO2, SO2 监测指标的相关均较高, 相关系数均在 0.7 以上, 表明 PM2.5 与这些指标的监测但与 O3 的相关系数只有 -0.352。

#### (3) O3 与其它指标相关性低

由结果表格及示意图可以看出, O3 与其它 5 项监测指标的相关性较差且均为负相关, 相关系数均在 -0.4 以下。

## 5.2 PM2.5 含量与其它 5 项污染物含量的相关性分析

### 5.2.1 灰色关联度模型的建立

作为一个发展变化的系统, 关联分析实际上是动态过程发展态势的量化比较分析, 所谓发展态势比较, 也就是系统各时期有关统计数据的几何关系的比较。

定义一 设有序列

$$x = (x(1), x(2), \dots, x(n))$$

则称映射

$$f: x \rightarrow y$$

$$f(x(k)) = y(k), \quad k = 1, 2, \dots, n \quad (5-3)$$

为序列  $x$  到序列  $y$  的数据变换。

1) 当

$$f(x(k)) = \frac{x(k)}{x(1)} = y(k), \quad x(1) \neq 0 \quad (5-4)$$

称  $f$  是初始化变换。

2) 当

$$f(x(k)) = \frac{x(k)}{\bar{x}} = y(k), \quad \bar{x} = \frac{1}{n} \sum_{k=1}^n x(k) \quad (5-5)$$

称  $f$  是均值化变换。

3) 当

$$f(x(k)) = \frac{x(k)}{\max_k x(k)} = y(k) \quad (5-6)$$

称  $f$  是百分比变换。

4) 当

$$f(x(k)) = \frac{x(k)}{\min_k x(k)} = y(k), \quad \min_k x(k) \neq 0 \quad (5-7)$$

称  $f$  是倍数变换。

5) 当

$$f(x(k)) = \frac{x(k)}{x_0} = y(k) \quad (5-8)$$

其中  $x_0$  为大于零的某个值，称  $f$  是归一化变换。

6) 当

$$f(x(k)) = \frac{x(k) - \min_k x(k)}{\max_k x(k) - \min_k x(k)} = y(k) \quad (5-9)$$

称  $f$  是极差最大值化变换。

## 2、关联分析

定义 2 选取参考数列

$$x_0 = \{x_0(k) | k=1, 2, \dots, n\} = (x_0(1), x_0(2), \dots, x_0(n))$$

其中  $k$  表示时刻。假设有  $m$  个比较数列

$$x_i = \{x_i(k) | k=1, 2, \dots, n\} = (x_i(1), x_i(2), \dots, x_i(n)), \quad i=1, 2, \dots, m$$

则称

$$\xi_i(k) = \frac{\min_s \min_t |x_0(t) - x_s(t)| + \rho \max_s \max_t |x_0(t) - x_s(t)|}{|x_0(k) - x_i(k)| + \rho \max_s \max_t |x_0(t) - x_s(t)|} \quad (5-10)$$

为比较数列  $x_i$  对参考数列  $x_0$  在  $k$  时刻的关联系数，其中  $\rho \in [0, 1]$  为分辨系数。

称 (5-10) 式中  $\min_s \min_t |x_0(t) - x_s(t)|$ 、 $\max_s \max_t |x_0(t) - x_s(t)|$  分别为两级最小差及两级最大差。

一般来讲，分辨系数  $\rho$  越大，分辨率越大； $\rho$  越小，分辨率越小。

(5-10) 式定义的关联系数是描述比较数列与参考数列在某时刻关联程度的一种指标，由于各个时刻都有一个关联数，因此信息显得过于分散，不便于比较，为此我们给出

定义 3 称：

$$r_i = \frac{1}{n} \sum_{k=1}^n \xi_i(k) \quad (5-11)$$

为数列  $x_i$  对参考数列  $x_0$  的关联度。

其中：  $r_1$  :SO<sub>2</sub> 对 PM2.5 的关联度；

$r_2$  :NO<sub>2</sub> 对 PM2.5 的关联度；

$r_3$  :PM10 对 PM2.5 的关联度；

$r_4$  :CO 对 PM2.5 的关联度；

$r_5$  :O<sub>3</sub> 对 PM2.5 的关联度。

### 5.2.2 灰色关联度模型的求解

在利用 (5-10) 式及 (5-11) 式计算关联度之前，本文需要对数据的进行初始化处理。一般来讲，实际问题中的不同数列往往具有不同的量纲，而本文在计算关联系数时，要求量纲要相同。因此，需要首先对各种数据进行无量纲化。另外，为了易于比较，要求所有数列有公共的交点。为了解决上述问题，本文对给定数列进行变换

定义 4 给定数列  $x = (x(1), x(2), \dots, x(n))$ ，称

$$\bar{x} = \left( 1, \frac{x(2)}{x(1)}, \dots, \frac{x(n)}{x(1)} \right) \quad (5-12)$$

为原始数据  $X$  的初始化数列。

由于 O<sub>3</sub> 指标与 PM2.5 指标成负相关性，即 O<sub>3</sub> 含量越高，PM2.5 含量越低，故采取以下公式进行无量纲转化

$$x_i = \left( 1, \frac{x_i(1)}{x_i(2)}, \frac{x_i(1)}{x_i(3)}, \frac{x_i(1)}{x_i(4)}, \frac{x_i(1)}{x_i(5)} \right), i = 6 \quad (5-13)$$

利用 MATLAB 编程对模型进求解 ( $\sigma=0.5$ )，得到结果如图 5-2:

表 5-2 灰色关联度模型的求解结果

(PM2.5,SO <sub>2</sub> )	(PM2.5,NO <sub>2</sub> )	(PM2.5,PM10)	(PM2.5,CO)	(PM2.5,O <sub>3</sub> 1 小时)
$r_1$	$r_2$	$r_3$	$r_4$	$r_5$
0.785	0.837	0.858	0.872	0.645

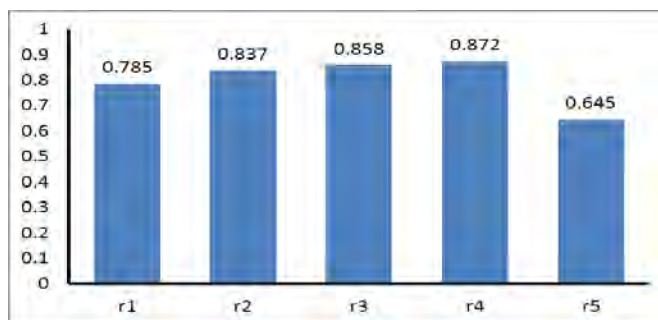


图 5-2: 关联度柱状图



### 5.2.3 结果分析

由关联度模型的结果可得出如下结论：PM2.5 含量与其它五项污染物含量的关联度均较高，因此可以然后以 PM2.5 为因变量，其它五种污染物为自变量，建立了线性回归模型。

## 5.3 PM2.5 含量与其它 5 项污染物含量回归方程

### 5.3.1 多元线性回归模型的建立

以  $m$  个自变量  $x_1, x_2, \dots, x_m$ ，因变量  $y$ ，建立多元线性回归方程：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (5-14)$$

式中  $\beta_0, \beta_1, \dots, \beta_m, \sigma^2$  都是与  $x_1, x_2, \dots, x_5$  无关的未知参数，其中  $\beta_0, \beta_1, \dots, \beta_5$  称为回归系数。

现得到  $n$  个独立观测数据  $(y_i, x_{i1}, \dots, x_{i5})$ ， $i = 1, \dots, n, n > 5$ ，由(5-14)式得

$$\begin{cases} y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_m x_{im} + \varepsilon_i \\ \varepsilon_i \sim N(0, \sigma^2), i = 1, \dots, n \end{cases} \quad (5-15)$$

记

$$X = \begin{bmatrix} 1 & x_{11} & \dots & x_{15} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & \dots & x_{n5} \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (5-16)$$

$$\varepsilon = [\varepsilon_1 \dots \varepsilon_n]^T, \quad \beta = [\beta_0 \quad \beta_1 \dots \beta_5]^T$$

(5-14)可以表示为：

$$\begin{cases} Y = X\beta + \varepsilon \\ \varepsilon \sim N(0, \sigma^2 E_n) \end{cases} \quad (5-17)$$

其中  $E_n$  为  $n$  阶单位矩阵。

结合上述建模过程及本文自变量与因变量关系，得到多元线性回归方程：

$$\begin{cases} y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_5 x_5 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (5-18)$$

其中  $y$  代表样本 PM2.5 含量， $x_i (i=1, 2, \dots, 5)$  分别代表  $\text{SO}_2$ 、 $\text{NO}_2$ 、PM10、CO、 $\text{O}_3$  的含量， $\beta_i (i=0, 1, 2, \dots, 5)$  代表回归系数， $\varepsilon$  表示随机误差。

### 5.3.2 多元线性回归模型的求解

利用 MATLAB 工具箱函数 regress，直接对多元线性回归方程进行求解，得到的结果如表 5-3：

表 5-3 多元线性回归模型的求解结果

参数	参数估计值	置信区间
$\beta_0$	-5.299	[-18.130, 7.531]
$\beta_1$	0.116	[-0.156, 0.388]
$\beta_2$	0.270	[-0.009, 0.549]
$\beta_3$	0.484	[0.396, 0.572]
$\beta_4$	39.229	[30.546, 47.911]
$\beta_5$	-0.163	[-0.220, -0.106]
$R^2 = 0.853, F = 269.263, p = 0.0000$		

表 5-3 得出,  $R^2 = 0.8009$  指因变量  $y$  (蛋白质含量) 的 85.3% 可由模型确定,  $F$  值远远超过了  $F$  检验的临界值,  $p$  远远小于显著性水平 0.05, 因而问题一的模型整体上来看是可行的。但是考虑到实际生活中, PM2.5 的含量和监测关系到人们的工作生活和生命健康, 因此  $R^2 = 0.85$  的准确性肯定还达不到要求, 模型需要优化。

### 5.3.3 交叉回归模型的建立

由相关性模型可知: 相关性较高的含量指标组为  $(SO_2, NO_2)$ 、 $(NO_2, PM10)$ 、 $(SO_2, PM10)$ 、 $(SO_2, CO)$ 、 $(NO_2, CO)$ 、 $(PM10, CO)$  (相关系数均在 0.6 以上), 因此我们有理由相信这六组变量的交互作用会对 PM2.5 的含量产生影响。

在多元线性回归方程的基础上, 本文引入了  $x_1x_2, x_2x_3, \dots$  等六项交叉项, 建立如下交叉回归模型:

$$\begin{cases} y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_5x_5 + \beta_6x_1x_2 + \dots + \beta_{12}x_3x_5 + \varepsilon \\ \varepsilon \sim N(0, \sigma^2) \end{cases} \quad (5-19)$$

其中  $y$  代表样本 PM2.5 含量,  $x_i (i=1, 2, \dots, 5)$  分别代表  $SO_2$ 、 $NO_2$ 、 $PM10$ 、 $CO$ 、 $O_3$  的含量,  $\beta_i (i=0, 1, 2, \dots, 50)$  代表回归系数,  $\varepsilon$  表示随机误差。

### 5.3.4 交叉回归模型的求解

利用 MATLAB 工具箱函数 regress, 直接对交叉回归方程进行求解, 得到的结果如表 5-4:

表 5-4 交叉回归模型的求解

参数	参数估计值	置信区间
$\beta_0$	9.749	[-8.916, 28.414]
$\beta_1$	0.656	[-0.047, 1.359]
$\beta_2$	0.738	[0.121, 1.355]
$\beta_3$	-0.009	[-0.177, 0.159]
$\beta_4$	23.346	[-1.369, 48.062]
$R^2 = 0.893, F = 170.708, p = 0.0000$		

表 5-4 得出,  $R^2 = 0.893$  指因变量  $y$  (蛋白质含量) 的 89.3% 可由模型确定,  $F$  值远远超过了  $F$  检验的临界值,  $p$  远远小于显著性水平 0.05, 因而模型整体上看是可行的。

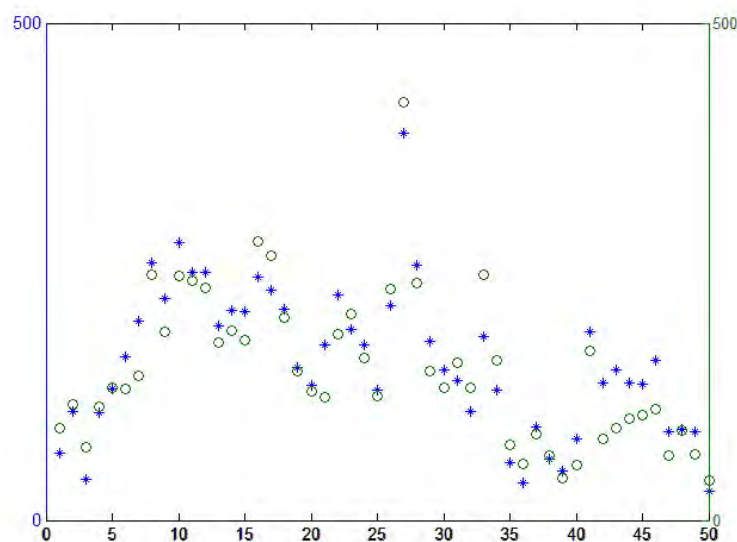
交叉回归方程为:

$$y = 9.749 + 0.656x_1 + 0.738x_2 + \dots - 0.152x_5 + 0.198x_1x_2 + \dots + 0.491x_3x_5 \quad (5-20)$$

### 5.3.5 交叉回归模型的检验

将原始数据代入到交叉回归方程中, 得到 PM2.5 的实验数据。将实验数据与实际 PM2.5 的数据进行残差及剩余标准差检验。

由 MATLAB 可求出交叉回归方程的残差图及剩余标准差:

图 5-3 交叉模型的残差图 (剩余标准差  $S = 20.4$ )

由残差图及剩余标准差可以看出, 交叉回归模型与数据拟合程度较好, 准确度较高, 这说明模型是完全可行的。

5.3.6 结果分析

在多元回归模型的基础上，增加了相关性较高的 6 项交互项，使回归模型的准确率得到了明显的提高（由  $R$  值由 0.85 增加到 0.89），使得模型更加优化，PM2.5 预测更加准确。

5.4 PM2.5 相关因素的探讨—空气湿度

5.4.1 空气湿度的说明

空气湿度是表示空气中水汽含量和湿润程度的气象要素，即代表空气中的含水量。研究表明：二氧化硫（SO<sub>2</sub>），二氧化氮（NO<sub>2</sub>），一氧化碳（CO）是在一定环境条件下形成 PM2.5 前的主要气态物体，可以认为这些污染物在空气中通过物理和化学反应转化为了 PM2.5，而水一般都是化学反应不可缺少的物质，因此有理由相信空气湿度与 PM2.5 有关。

5.4.2 空气湿度与 PM2.5 的关系探讨

为了说明空气湿度对 PM2.5 是否有关，本文首先根据在互联网上收集到的湿度及 PM2.5 数据建立关于不含空气湿度变量的线性回归模型一，然后建立含有空气湿度变量的线性回归模型二，通过两种线性回归方程一、二的结果比较，分析说明 PM2.5 与空气湿度的关系。

利用 MATLAB 求出模型一、二的结果如表 5-5：

表 5-5 模型一二结果对比

参数	模型一	模型二
$R^2$	0.876	0.916
$F$	81.088	68.815
$P$	0	0

5.4.3 结果分析

由上述结果可知，增加空气湿度指标后，线性回归方程的相关系数得到了明显提升（由 0.876 增加到 0.916），这表明空气湿度对 PM2.5 存在着较强的影响，引入 PM2.5 指标后，有利于用提高 PM2.5 预测的准确度。

6 问题二的解答

6.1 西安 PM2.5 的时空分布及规律探讨

6.1.1 PM2.5 时空分布图及规律探讨

1、PM2.5 时空分布图

结合附件 2 西安市 13 个监测站点的 PM2.5 指标及其空间分布，利用 MATLAB 画出 13 个地区的 PM2.5 时空分布图，如图 6-1（由于篇幅所限，给出部分地区时空分布图，详细结果见附录四）：

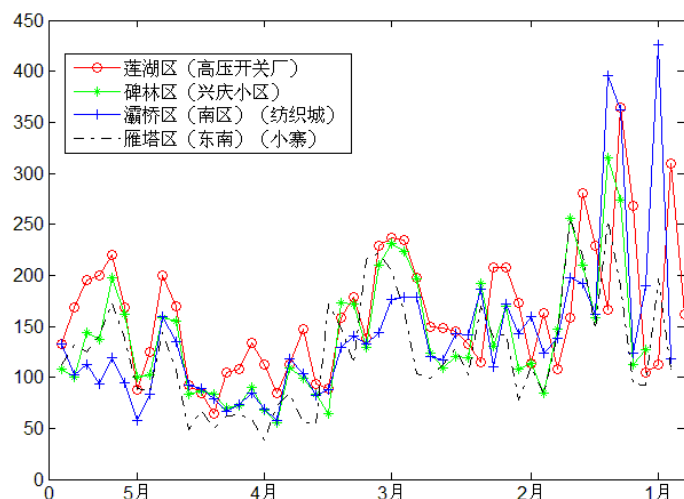


图 6-1 13 个地区中 4 个的时空分布图

另外，莲湖区（高压开关厂）全年的 PM2.5 时空分布图如下图所示：

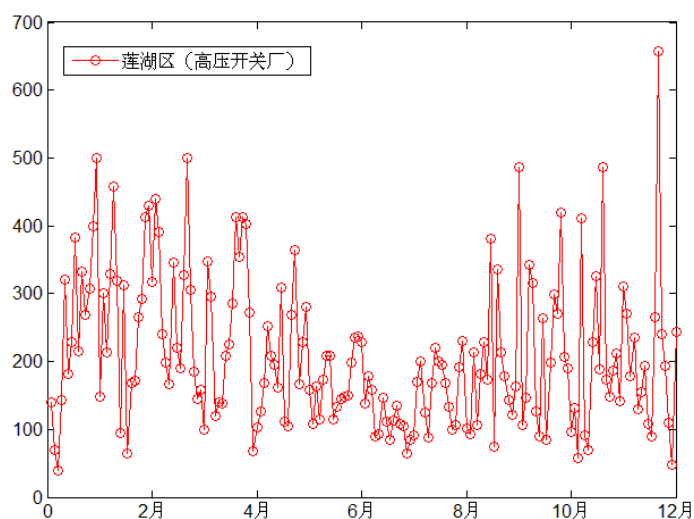


图 6-2 莲湖区（高压开关厂）全年的 PM2.5 时空分布图

## 2、PM2.5 的分布规律

由 PM2.5 的时空分布图及附件数据可以得到如下规律：

（1）3 月份 PM2.5 浓度最低，9 月份 PM2.5 浓度最高。

对西安市 13 个区的全年数据进行统计，发现 3 月份 PM2.5 平均浓度最低（108），9 月份 PM2.5 浓度最高（155）。

（2）市区 PM2.5 浓度明显大于郊区浓度。

13 个站点（地区）中长安区、阎良区、临潼区为郊区，其余十个区为市区。统计附件数据可得出：市区 PM2.5 浓度平均值约为 157，郊区 PM2.5 浓度平均值约为 113，市区 PM2.5 污染程度明显大于郊区。

（3）市区西部 PM2.5 浓度明显大于东部浓度。

市区西部 PM2.5 浓度平均值约为 162，市区东部 PM2.5 浓度平均值约为 150。

## 6.2 PM2.5 地区污染评估

### 6.2.1 模糊综合评价模型的建立

考虑到各个地区的 PM2.5 “污染程度”属于较为“模糊”的概念，而且评价因素较多，因此本文决定使用模糊评价模型对地区污染进行综合评价。

#### 1、模糊综合评价模型的建立

模糊综合评价方法，是应用模糊关系合成的原理，从多个因素（指标）对被评价事物隶属等级状况进行综合性评判的一种方法。其具体的步骤为：

- (1) 确定评判对象的因素论域为  $U$ ， $U = (u_1, u_2, \dots, u_n)$ ；
- (2) 确定评语等级论域  $V$ ， $V = (v_1, v_2, \dots, v_m)$ 。通常评语有  $V = (\text{很高, 高, 较高, } \dots, \text{较低, 低, 很低})$ ；
- (3) 进行单因素评判，建立模糊关系矩阵  $R$

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix}, \quad 0 \leq r_{ij} \leq 1$$

其中  $r_{ij}$  为  $U$  中因素  $u_i$  对于  $V$  中等级  $v_j$  的隶属关系；

(4) 确定评判因素权向量  $A = (a_1, a_2, \dots, a_n)$ ， $A$  是  $U$  中各因素对被评事物的隶属关系，它取决于人们进行模糊综合评判是的着眼点，即根据评判时各因素的重要性分配权重；

(5) 选择评价的合成算子，将  $A$  与  $R$  合成得到  $B = (b_1, b_2, \dots, b_m)$ 。

综上所述，模糊综合评价模型为：

$$B = AOR = (a_1, a_2, \dots, a_n)O \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ r_{21} & r_{22} & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{n1} & r_{n2} & \cdots & r_{nm} \end{pmatrix} \quad (6-1)$$

$$b_j = (a_1 \bullet_* r_{1j}) + \cdots + (a_n \bullet_* r_{nj}), \quad j = 1, 2, \dots, m$$

其中： $B$  为 PM2.5 污染程度综合评价指数， $\bullet_*$  为模糊算子。

### 6.2.2 模糊综合评价模型的求解

#### 1、PM2.5 浓度分级及频数统计

首先根据新修订的《环境空气质量标准》将 PM2.5 浓度划分为 5 个等级，然后对附件 2 中 13 个监测站点 PM2.5 浓度的进行统计，得到各个站点的 PM2.5 对应的等级频数，如表 6-1（详细过程见数据分析）：

表 6-1 各个站点的 PM2.5 指数的频数

等级 监测站	一级	二级	三级	四级	五级
高压开关厂	15	54	51	114	108
兴庆小区	24	63	63	84	84
纺织城	21	66	78	75	93
小寨	42	72	69	69	81
市人民体育场	30	45	45	90	90
高新西区	27	51	48	93	114
经开区	27	57	57	81	81
长安区	36	72	66	75	84
阎良区	30	63	39	108	93
临潼区	39	51	57	78	72
曲江文化集团	42	66	72	63	87
广运潭	18	51	60	102	93
草滩	18	60	36	123	99

## 2、模型求解

将附件 2 统计的分级频数数据代入模型中，得到各个监测站点的 PM2.5 污染指数，结果如表 6-2：

表 6-2 模糊综合评价模型的结果

监测站	开关厂	广运潭	纺织城	小寨	市体育场	高新西区	经开区
污染指数	94.39	92.41	89.19	84.50	91.00	92.97	88.71
监测站	长安区	阎良区	临潼区	草滩	兴庆小区	曲江集团	全市平均
污染指数	83.95	75.27	81.26	93.39	88.87	85.27	87.78

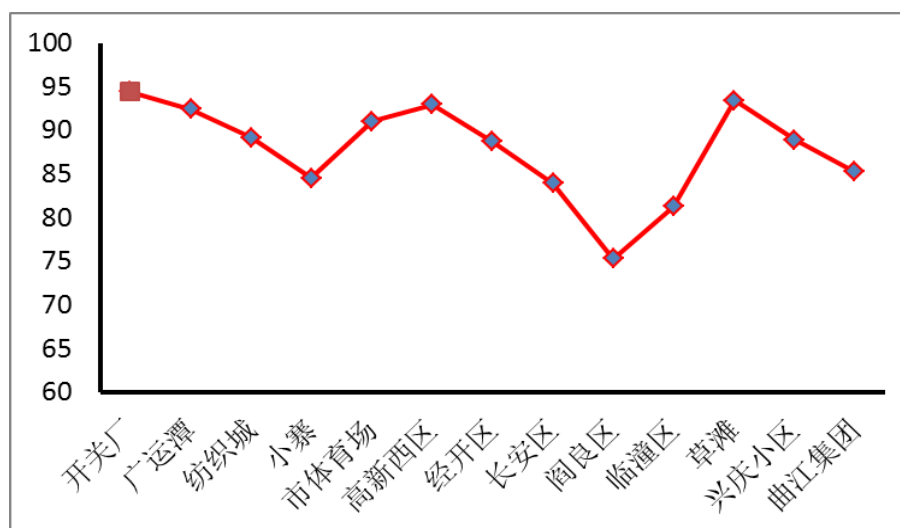


图 6-3 污染指数曲线分布图

由结果可得到如下结论：

① 市区 PM2.5 污染程度明显大于郊区污染程度。

由模型结果可得，市区 PM2.5 平均污染指数 89.67，郊区（长安区，临潼

区，阎良区）平均污染指数为 80.16，显然市区污染程度明显大于郊区污染程度。

② 市区西部 PM2.5 污染程度明显大于东部污染程度。

由模型结果可得，市区西部（高压开关厂，高新区）PM2.5 平均污染指数为 93.68，市区东部 PM2.5 为 89.16，显然市区西部 PM2.5 污染程度明显大于东部污染程度。

### 6.3 PM2.5 的扩散与衰减规律模型探讨

#### 6.3.1 PM2.5 高斯烟羽模型的建立

##### 1、高斯烟羽模型的建立

PM2.5 在大气中迁移和扩散的数值计算基本上可分为二步。第一步根据大气动力学理论进行所关心地区中风场的计算，其理论基础是大气运动方程、连续性方程、状态方程、热力学方程和水汽方程构成的基本方程组。在大气科学研究领域中，已有多个实用的大气环流模式。第二步进行已知风场中 PM2.5 迁移和扩散的计算，可采用类似于处理大气污染的方法，假设 PM2.5 不影响大气流体速度和温度，求解 PM2.5 的连续性方程，PM2.5 模拟扩散图如图 6-4。

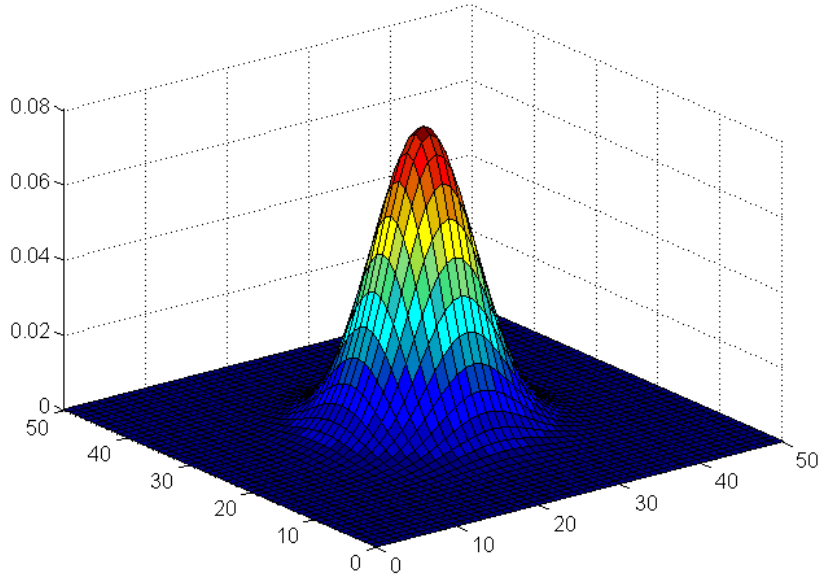


图 6-4 PM2.5 模拟扩散图

当风速为  $k \text{ m/s}$  时，利用连续点源高斯扩散模型分析监测站周边 PM2.5 浓度的变化情况。此排放点源是有边界点源，排放点源的实际高度为  $H$ 。以排放点源在地面的投影点为坐标原点，以风向方向为  $x$  轴，铅直方向为  $z$  轴，与  $x$  轴水平面垂直方向为  $y$  轴建立三维坐标系，由于扩散过程中浓度在  $y$ 、 $z$  轴上的变化分布符合高斯分布，所以下风向的任意一点  $C(x, y, z)$  的浓度函数为：

$$C(x, y, z) = A(x)e^{-ay^2}e^{-bz^2} \quad (6-2)$$

根据概率统计我们可以得出方差的表达式为：

$$\left\{ \sigma_y^2 = \frac{\int_0^\infty y^2 C(x, y, z) dy}{\int_0^\infty C(x, y, z) dy} \quad \sigma_z^2 = \frac{\int_0^\infty z^2 C(x, y, z) dz}{\int_0^\infty C(x, y, z) dz} \right\} \quad (6-3)$$



进而源强的积分公式可以根据假设得出：

$$Q = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \bar{u} C(x, y, z) dy dz \quad (6-4)$$

把（6-2）式代入（6-3）积分可以得出：

$$\begin{cases} a = \frac{1}{2\sigma_y^2} \\ b = \frac{1}{2\sigma_z^2} \end{cases} \quad (6-5)$$

将(6-2)式和(6-5)式代入(6-4)式可以得出：

$$A(x) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \quad (6-6)$$

最后再将把（6-5）、（6-6）式代入（6-2）式可以得出：

$$C(x, y, z) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left[-\left(\frac{y^2}{2\sigma_y^2} + \frac{z^2}{2\sigma_z^2}\right)\right] \quad (6-7)$$

上式为无界空间连续点源扩散的高斯模型，然而在实际中，由于地面的影响，烟羽是有界的。根据假设可以把地面看作一个镜面，对排放的气体起反射作用，同样我们可以利用“像源法”进行处理，原理和示意图在模型一的修正中提到，因此我们得出：

实源的贡献为：

$$C_1(x, y, z) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{1}{2} \frac{y^2}{\sigma_y^2}\right) \exp\left(-\frac{1}{2} \frac{(z-H)^2}{\sigma_z^2}\right) \quad (6-8)$$

像源的贡献为：

$$C_2(x, y, z) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{1}{2} \frac{y^2}{\sigma_y^2}\right) \exp\left(-\frac{1}{2} \frac{(z+H)^2}{\sigma_z^2}\right) \quad (6-9)$$

则该处的实际浓度为：

$$C(x, y, z) = C_1(x, y, z) + C_2(x, y, z) \quad (6-10)$$

综合上面的公式得到高斯烟羽扩散模型：

$$C(x, y, z, t, H) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left[-\frac{(z-H)^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+H)^2}{2\sigma_z^2}\right] \right\} \quad (6-11)$$

其中： $C(x, y, z, t, H)$ 代表排放源高度为 $H$ ，时间为 $t$ 时，坐标为 $(x, y, z)$ 处污染物浓度， $\bar{u}$ 代表平均风速， $Q$ 代表排放源污染物初始浓度， $\sigma_i$ 代表扩散系数。

## 2、高斯烟羽模型系数的确定

### 1)、扩散系数 $\sigma_y$ ， $\sigma_z$ 的确定过程：

中国国家标准规定的方法：根据时间、地理位置确定日倾角、太阳高度角，利用天气条件确定辐射等级，然后利用辐射等级和风速确定大气稳定度，最后查扩散参数幂函数表，确定扩散参数。

根据国家标准(GB/ T 13201--1991) 制定地方大气污染物排放标准的技术方法的规定, 划分大气稳定度的级别, 共分为6级A - F, A为极不稳定; F为极稳定。首先, 根据释放源所在地的经度和纬度以及排放的日期和时间计算当时的太阳高度角 $h_0$ ; 然后, 由太阳高度角 $h_0$ 和云量查出太阳辐射等级; 最后, 再根据地面风速确定当时的大气稳定度, 计算细节可参考文献。

表6-3: 大气稳定度的级别参考表

地面风速 ( $m/s$ )	白天太阳辐射			阴天的白天或夜间	有云的夜晚	
	强	中	弱		薄云遮天或低云 $\geq 0.5$	云量 $\leq 0.4$
<3	A	A-B	B	D	-	-
2-3	B	B	C	D	E	F
3-5	B-C	B-C	C	D	D	E
5-6	C-D	C-D	D	D	D	D
>6	C	D	D	D	D	D

有风时的扩散参数根据计算公式 (6-11), 并联合采用 *Briggs* 给出一套扩散参数, 如表 6-4 和表 6-5 所示。

表 6-4: *Briggs* 扩散参数 (开阔平原田野)

大气稳定度	$\sigma_y$	$\sigma_z$
A	$0.22x(1+0.0001x)^{1/2}$	$0.2x$
B	$0.16x(1+0.0001x)^{1/2}$	$0.12x$
C	$0.11x(1+0.0001x)^{1/2}$	$0.08x(1+0.0002x)^{1/2}$
D	$0.08x(1+0.0001x)^{1/2}$	$0.06x(1+0.0015x)^{1/2}$
E	$0.06x(1+0.0001x)^{1/2}$	$0.03x(1+0.0003x)$
F	$0.04x(1+0.0001x)^{1/2}$	$0.016x(1+0.0003x)$

表 6-5: *Briggs* 扩散参数 (城市)

大气稳定度	$\sigma_y$	$\sigma_z$
A-B	$0.32x(1+0.0004x)^{-1/2}$	$0.24x(1+0.0001x)^{-1/2}$
C	$0.22x(1+0.0004x)^{-1/2}$	$0.20x$
D	$0.16x(1+0.0004x)^{-1/2}$	$0.14x(1+0.0003x)^{-1/2}$
E-F	$0.11x(1+0.0004x)^{-1/2}$	$0.08x(1+0.0015x)^{-1/2}$

## 2)、有效高度 $H$ 的确定过程:

影响排放源抬升高度的因素很多, 主要包括: 扩散气体的初始速度和方向、初始温度、排放口直径、环境风速及风速随高度的变化率、环境温度以及大气温度。目前大多采用半经验公式计算抬升高度。此处采用20世纪80年代初Wilson根据管道破裂排放。

实验所得到的经验公式:

$$H = 2.4s \frac{d}{k} \quad (6-12)$$

此公式适合于出口喷射方向竖直向上，喷射路径上无障碍物的情况，这与我们的假设相同，所以在本题中可以用此方法确定放射源的有效高度。

### 6.3.2 PM2.5 高斯烟羽模型的求解及分析

假设风向正北方，风速为 40km/h，衰减系数为 0.00001，有效高度为 50m，PM2.5 初始浓度为各站点 PM2.5 浓度的最高值，将附件 2 及附件 3 中 PM2.5 数据代入高斯烟羽模型中，利用 MATLAB 模拟，得到如下 PM2.5 扩散模拟如图 6-5，及扩散数据如表 6-6、表 6-7（以高压开关厂为例）：

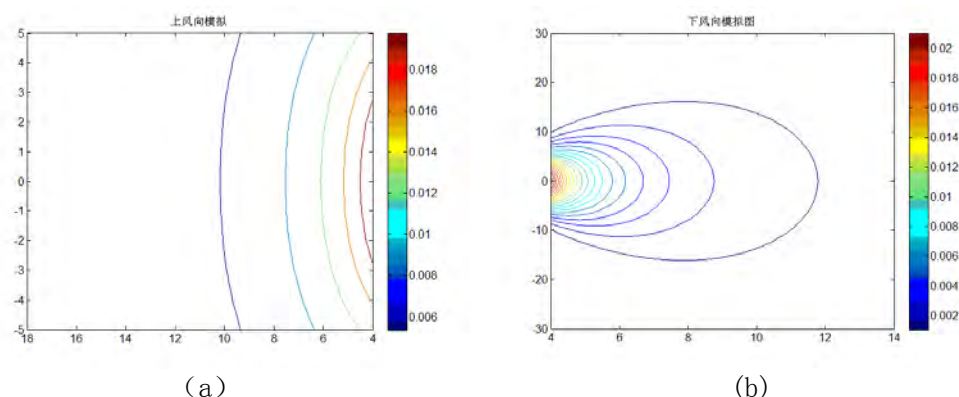


图 6-5 PM2.5 扩散模拟图，其中图（a）、（b）分别为上风和下风 L 公里处预测结果

表 6-6 上风处 PM2.5 浓度扩散

距离(km)	0	2	4	6	8	10	11	12
PM2.5 浓( $\mu\text{g} / \text{m}^3$ )	500	427	389	287	115	31	0	0
时间(min)	0	10	20	30	40	50	60	70

表 6-7 下风处 PM2.5 浓度扩散

距离(km)	0	2	4	6	8	10	11	12
PM2.5 浓( $\mu\text{g} / \text{m}^3$ )	500	472	389	313	292	135	72	0
时间(min)	0	20	40	60	80	100	120	140

由高斯烟羽模型的求解结果可以得到如下结论：

（1）PM2.5 衰减速度随距离的增大先变快，然后变慢。

由模型结果统计出距离监测点的各个距离段的 PM2.5 平均减少量，结果如表 6-8：

表 6-8 各个距离段的 PM2.5 平均减少量

距离段(km)	0—2	2—4	4—6	6—8	8—10	10—11
PM2.5 平均减少( $\mu\text{g} / \text{m}^3$ )	18.3	21.4	76.8	137	76	63

由表 6-8 可以清楚的看到，PM2.5 衰减速度随距监测站点距离的先增大变快，然后变慢。

（2）上风处 PM2.5 的衰减速度远大于下风处。

由模型结果可以发现（仅用高压开关厂数据说明），在距监测站的距离从 0 到 12 千米的过程中，PM2.5 气体的浓度从 500 降低到 0，上风处所需时间为

60min，而下风处所需时间为 140min。

6.4 PM2.5 突增及其应急处理

6.4.1 PM2.5 突增及扩散

假设该站点 PM2.5 浓度突然增加至最高浓度的 2 倍 ( $1000\mu g/m^3$ )，并持续 2 小时，假设风向为正北向，风速为  $40km/h$ ，扩散系数  $\sigma$  为 0.00001，排放源有效高度为  $50m$ ，将数据代入前面求出的高斯烟羽扩散模型中，得到上风处及下风处的结果如表 6-9 及表 6-10，并确定了模拟扩散图 6-6：

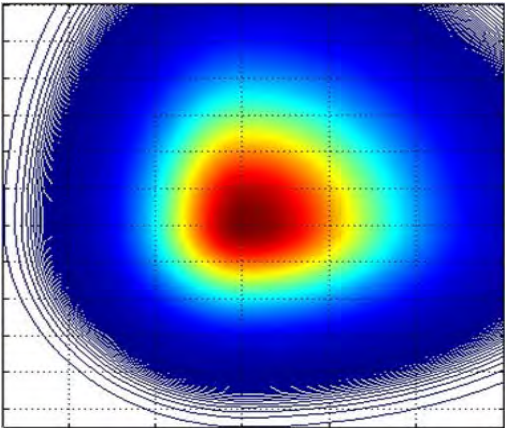


图 6-6 PM2.5 模拟扩散图

表 6-9 上风处 PM2.5 浓度扩散

距离 (km)	0	2	4	6	8	10	11	12
PM2.5 浓度 ( $\mu g/m^3$ )	1000	850	703	480	292	108	21	0
时间 (min)	0	20	36	58	82	106	115	121

表 6-10 下风处 PM2.5 浓度扩散

距离 (km)	0	2	4	6	8	10	11	12
PM2.5 浓度 ( $\mu g/m^3$ )	1000	827	671	438	223	31	0	0
时间 (min)	0	20	42	68	90	125	131	-

6.4.2 重度污染和可能安全的地区分析

1、污染等级划分

参考新修订的《环境空气质量标准》将 PM2.5 污染程度三个等级，如表 6-11。

表 6-11 PM2.5 污染程度分级

污染程度	轻度污染（安全）	中度污染	重度污染
PM2.5 浓度 ( $\mu g/m^3$ )	0-350	350-500	500-1000

由于轻度污染区的 PM2.5 ( $\mu g/m^3$ ) 相对较低，对人们的日常生活及生命健康不会带来太大影响，因此将轻度污染区考虑为安全区。

2、污染地区分布

由西安市的地图，可以确定各站点的空间分布，如图 6-7：



图 6-7 13 个站点空间分布

由求解结果及等级划分，结合各站点与高压开关厂的距离， $\text{PM}_{2.5}$  突增为  $1000 \mu\text{g}/\text{m}^3$  后，并持续 2 小时后，可以确定各站点的污染程度，如表 6-12：

表 6-12 各地区污染程度分析

轻度污染（安全）	中度污染	重度污染
阎良区 临潼区 广运潭 纺织城 长安区	市体育馆 曲江文化集团 兴庆小区	其它 区域

### 6.4.3 结果分析及应急处理

由污染地区划分图可以清楚得看出，在  $\text{PM}_{2.5}$ （高压开关厂）突增为  $1000 \mu\text{g}/\text{m}^3$  后，并持续 2 小时后，轻度污染地区（安全区）有纺织城，长安区，阎良区，临潼区，广运潭六个站点（地区），因此人们应尽量向这些安全地区撤离。

## 6.5 模型检验与一般性规律探讨

### 6.5.1 模型检验

根据参考文献<sup>[6][7]</sup>可以得到日本福岛第一核电站放射性物质的扩散衰减数据，将放射性物质和  $\text{PM}_{2.5}$  扩散衰减数据进行对比，从而检验模型的合理性。

表 6-12 放射性物质浓度扩散

距核电站距离 (km)	0	250	500	750	1000
放射性物质浓度 ( $\text{Bq}/\text{m}^3$ )	300	257	132	75	32
时间 (h)	0	24	48	72	96

表 6-13 PM2.5 浓度扩散

距离 (km)	0	2	4	6	8	10	11	12
PM2.5 浓度 ( $\mu\text{g}/\text{m}^3$ )	1000	827	671	438	223	31	0	0
时间 (min)	0	20	42	68	90	125	131	-

从表 6-12 及表 6-13 可以看出, 两种物质衰减速度都是随距离的增大先变快, 然后变慢。表明 PM2.5 浓度扩散这也验证了高斯烟羽模型的合理性。

### 6.5.2 PM2.5 一般性规律探讨

为了探讨 PM2.5 扩散衰减更为一般的规律, 本文考虑了 PM2.5 物质的自身沉降作用和雨水吸附作用两个更为一般的因素对模型的影响, 利用这两个因素对模型进行修正。

#### 1、高斯烟羽模型的修正

连续点源高斯扩散模型虽然能分析风力对浓度的影响情况, 但为了能更准确的探究风力对 PM2.5 物质浓度的影响情况, 我们将考虑排放源有效高度、PM2.5 物质自身的沉降作用和雨水吸附作用对 PM2.5 浓度的影响, 进而对连续点源高斯扩散模型进行修正。

##### 1) 考虑放射性物质自身的沉降作用对模型的修正

放射性物质的沉降速度取决于空气阻力和自身重力, 利用斯托克斯公式表示沉降速度:

$$V_s = \frac{\rho g d^2}{18\alpha} \quad (6-13)$$

$\rho$ : PM2.5 粒子密度,  $\text{kg}/\text{m}^3$ ;

$g$ : 重力加速度,  $9.8065 \text{ m/s}^2$ ;

$d$ : PM2.5 物质粒子直径  $m$ ;

$\alpha$ : 空气的动力粘性系数, 可取  $1.8 \times 10^{-5} \text{ kg}/(\text{m} \cdot \text{s})$ ;

$V_s$ : 沉降速度,  $\text{m/s}$

在扩散过程中重力沉降的位移叠加在羽流中心线上, 使中心线向下倾斜, 放射性物质粒子则相当于在下倾的中心线上扩散, 放射性物质的扩散与沉降的叠加

可认为是放射源以  $V_s$  的速度向下移动。在  $x$  处向下移动的垂直距离为  $V_s t = \frac{V_s x}{\alpha}$ ,

即排放源的有效高度  $h$  下降了  $V_s t = \frac{V_s x}{\alpha}$ , 排放源的有效高度成为  $h - V_s t = h - \frac{V_s x}{\alpha}$ ,

考虑到地面的全反射作用, 反射项的有效高度也变成了  $h - V_s t = h - \frac{V_s x}{\alpha}$ 。则修正

后的连续点源高斯扩散模型为:

$$C(x, y, z, h) = \frac{Q}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left[-\frac{(z-h+\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+h-\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] \right\} \quad (6-14)$$

##### 2) 考虑雨水的吸附作用对模型的修正

降雨对 PM2.5 的浓度有一定影响, 即雨水对 PM2.5 有一定的吸附作用。以吸附系数  $\beta$  来表示雨水对 PM2.5 吸附作用的大小,  $\beta$  与降雨强度的关系为:

$\beta = aI^b$ , 式中  $I$  为降雨强度,  $a, b$  为经验系数。雨水的吸附作用导致的 PM2.5 浓

度的减小，可对源强进行修正：

$$Q(x) = Q \exp\left(-\frac{\beta x}{\alpha}\right) \quad (6-15)$$

则进一步修正的连续点源高斯扩散模型为：

$$C(x, y, z, h) = \frac{Q(x)}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left[-\frac{(z-h+\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+h-\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] \right\} \quad (6-16)$$

综上所述：修正的高斯烟羽扩散模型为：

$$C(x, y, z, h) = \frac{Q(x)}{2\pi\bar{u}\sigma_y\sigma_z} \exp\left(-\frac{y^2}{2\sigma_y^2}\right) \left\{ \exp\left[-\frac{(z-h+\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] + \exp\left[-\frac{(z+h-\frac{V_s x}{\alpha})^2}{2\sigma_z^2}\right] \right\} \quad (6-17)$$

$$\begin{cases} Q(x) = Q \exp\left(-\frac{\beta x}{\alpha}\right), \beta = al^b \\ V_s = \frac{\rho g d^2}{18\alpha} \end{cases}$$

其中： $C(x, y, z, t, H)$ 代表排放源高度为 $H$ ，时间为 $t$ 时，坐标为 $(x, y, z)$ 处污染物浓度， $\bar{u}$ 代表平均风速， $Q$ 代表排放源污染物初始浓度， $\sigma_i$ 代表扩散系数， $\beta$ 代表雨水吸附系数， $\alpha$ 代表动力粘性系数， $V_s$ 代表粒子沉降速度， $d$ 代表粒子直径。

## 2、修正后模型的检验

假设风向为正北向，风速为 $40\text{km/h}$ ，排放源有效高度为 $50\text{m}$ ， $\text{PM}_{2.5}$ 粒子密度为 $0.0000001\text{kg}/\text{m}^3$ ，直径为 $2.5\mu\text{m}$ ，沉降速度为 $1.1\text{cm/s}$ ，空气的动力粘性系数 $1.8\times 10^{-5}\text{kg}/(\text{m}\cdot\text{s})$ ；雨水对 $\text{PM}_{2.5}$ 物质的吸附作用为 $8\times 10^{-5}\text{kg}/(\text{m}\cdot\text{s})$ 。

以高压开关厂 $\text{PM}_{2.5}$ 最高浓度作为初始浓度，代入到修正后的高斯烟羽扩散模型中，利用MATLAB编程得到结果，如表6-14及表6-15。

表 6-14 上风处  $\text{PM}_{2.5}$  浓度扩散

距离 (km)	0	2	4	6	8	10	11	12
$\text{PM}_{2.5}$ 浓度 ( $\mu\text{g}/\text{m}^3$ )	500	427	389	295	125	32	0	0
时间 (min)	0	18	35	57	82	113	128	—

表 6-15 下风处  $\text{PM}_{2.5}$  浓度扩散

距离 (km)	0	2	4	6	8	10	11	12
$\text{PM}_{2.5}$ 浓度 ( $\mu\text{g}/\text{m}^3$ )	500	472	425	359	278	127	75	0
时间 (min)	0	20	36	58	82	106	115	121

## 6.6 本问小结

本问主要解决了如下问题：

1、利用MATLAB画出了2013年全年 $\text{PM}_{2.5}$ 的时空分布图，并得出了三种分布规律，建立了模糊综合评价模型，对每个地区的进行污染综合评价，得到了每个站点（地区）的 $\text{PM}_{2.5}$ 污染指数。

2、考虑风速因素的影响，建立了关于 $\text{PM}_{2.5}$ 扩散的高斯烟羽模型，并利

用此模型得到了各站点的扩散数据。

3、当 PM2.5 浓度突增为 2 倍时，利用高斯烟羽模型求出个站点扩散数据，并结合各个站点（地区）的两两距离，得到了各站点污染情况

4、根据参考文献<sup>[6-7]</sup>得到了福岛第一核电站的放射性物质扩散数据，将放射性物质与 PM2.5 扩散数据进行对比，发现两者的扩散规律大体一致，从而验证了模型的合理性。利用物质的自身沉降作用和雨水吸附作用对高斯烟羽模型进行了修正，得到了修正后更为一般性的扩散模型，并对模型进行检验。

## 7 问题三的解答

### 7.1 PM2.5 治理计划探讨

#### 7.1.1 5 年后 PM2.5 浓度预测（不治理的情况）

假设不对 PM2.5 进行治理，根据以往空气质量的变化趋势，PM2.5 的浓度必然会增加。因此在制定 PM2.5 治理计划时，应当考虑 PM2.5 在未来 5 年内增加量。

假设未来 5 年内，PM2.5 平均年增长率为  $\alpha$ ，当前 PM2.5 年平均浓度为  $C_0$ ，则五年后，PM2.5 浓度为：

$$C = C_0(1 + \alpha)^5 \quad (7-1)$$

根据以往空气质量的变化趋势，本文假设年平均增长率为 3%，则根据式（7-1）及当前浓度值  $280 (\mu\text{g} / \text{m}^3)$ ，可以预测出五年后 PM2.5 年平均浓度为： $324 \mu\text{g} / \text{m}^3$ 。因此，为了考虑 PM2.5 在未来 5 年内增加量，本文将  $324 \mu\text{g} / \text{m}^3$  作为 PM2.5 的初始浓度。

#### 7.1.2 PM2.5 治理计划探讨

##### 1、PM2.5 分期治理计划分析

考虑到实际环境治理过程中，往往是分期进行治理工作的，而且治理进度是先慢变快，再由快变慢，治理前期：需要进行很多的准备工作，治理过程遇到的困难也最多，进度自然很慢。治理中期：有了前期的准备工作及治理经验作基础，较前期而言，治理进度会大幅提升。治理后期：需要对前期及中期的治理成果进行稳固，而且由于治理量已经减少到一个相对较小的水平，因此如果需要使 PM2.5 继续减少，要付出的努力必然更大，因此在治理后期，治理进度会很低。

为了与实际治理过程相符，本文决定对 PM2.5 的治理采用分期治理的方法，将未来 5 年分为三个时期：

表 7-1 PM2.5 的治理分期表

时期	前期（准备期）	中期（治理期）			后期（稳固期）
年次	第一年	第二年	第三年	第四年	第五年



## 2、柯西分布函数确定分期计划

考虑到分期治理计划的治理进度首先由慢变快，然后在再由快变慢。这种变化规律与柯西分布函数的斜率（ $y$  的变化快慢）变化规律相似，因此可以考虑使用柯西分布函数来确定每个时期 PM2.5 的治理计划。

柯西分布函数为：

$$y = \begin{cases} 1 & x \leq a \\ \frac{1}{1 + \alpha(x-a)^\beta} & x > a \end{cases} \quad (7-2)$$

其中： $\alpha > 0, \beta > 0$ 。

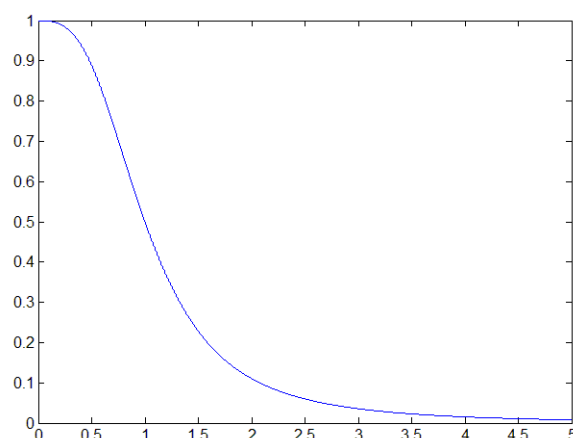


图 7-1 柯西分布函数图形

令  $a=0$ ，通过计算机模拟，找出了与分期治理计划进度变化规律最为相似的柯西函数（ $\alpha = 0.12, \beta = 3.05$ ）：

$$k_i = \begin{cases} 1 & i < 0 \\ \frac{1}{1 + 0.12i^{3.05}} & 0 \leq i \leq 5 \end{cases} \quad (7-3)$$

其中： $k_i$  为第一年到第  $i$  年 PM2.5 的累计减少量与 PM2.5 初始量之比。

由得到的柯西分布函数可以确定分期治理计划，如表 7-2 所示：

表 7-2 PM2.5 分期治理计划表

时期	前期（准备期）	中期（治理期）			后期（稳固期）
年次	第一年	第二年	第三年	第四年	第五年
治理量百分比（%）	9.6	32.7	38.8	15.3	3.6
治理量	31.1	105.94	125.71	49.57	11.66

由上表可以看出：

① 治理前期：治理进度较慢。这是因为前期需要进行很多的准备工作，又由于缺少 PM2.5 的治理经验，治理过程遇到的困难也最多。

② 治理中期：治理进度最快。这是因为有了前期的准备工作及治理经验作基础，治理进度自然会大幅提升。

③ 治理后期：治理进度最慢。这是因为需要对前期及中期的治理成果进行稳固，而且由于 PM2.5 的已经减少到一个相对降低的水平，因此如果需要使

PM2.5 继续减少，需要付出的努力必然更大，因此在治理后期，治理进度会很慢。

## 7.2 武汉市某地区专项治理计划探讨

### 7.2.1 武汉市某地区专项治理最优化模型建立

#### 1、初始 PM2.5 浓度的确定

首先根据附件 1 中 2013 年 1 月 1 日至 2013 年 8 月 26 日的 PM2.5 分指数数据，预测出 2013 年 8 月 26 日至 2013 年 12 月 31 日的 PM2.5 分指数数据，然后将 2013 全年分指数数据转化为浓度值，继而得到 2013 年全年平均 PM2.5 浓度为： $82 \mu\text{g}/\text{m}^3$ （详细过程见数据分析）。

#### 2、关于 PM2.5 专项治理费用的最优化模型确定

1) 假设未来五年里，PM2.5 每年的减少量为  $x_i$  ( $i=1, \dots, 5$ )，则有：

$$C_j - C_0 = \sum_{i=1}^j x_i \quad i, j = (1, 2, \dots, 5) \quad (7-4)$$

其中： $C_0$  代表 PM2.5 初始浓度， $C_j$  代表第  $j$  年 PM2.5 浓度。

2) 未来五年里，每年 PM2.5 专项治理的费用为：

$$S_i = 0.005x_i^2 \quad i = (1, 2, \dots, 5) \quad (7-5)$$

3) 未来五年里，PM2.5 专项治理投入总费用：

$$Z = \sum_{i=1}^5 S_i \quad (7-6)$$

综上所述，关于 PM2.5 专项治理费用的最优化模型为：

$$\begin{cases} \min Z = \sum_{i=1}^5 S_i \\ C_j - C_0 = \sum_{i=1}^j x_i \\ S_i = 0.005x_i^2 \quad i = (1, 2, \dots, 5) \\ 0 \leq x_i \leq C_0 \\ i, j = (1, 2, \dots, 5) \end{cases} \quad (7-7)$$

#### 3、分期治理对最优化模型的修正

考虑到实际环境治理过程中，往往是分期进行治理工作的，而且治理进度是先慢变快，再由快变慢，治理前期：需要进行很多的准备工作，治理过程遇到的困难也最多，进度自然很慢。治理中期：有了前期的准备工作及治理经验作基础，较前期而言，治理进度会大幅提升。治理后期：需要对前期及中期的治理成果进行稳固，而且由于治理量已经减少到一个相对较小的水平，因此如果需要使 PM2.5 继续减少，要付出的努力必然更大，因此在治理后期，治理进度会很低。

考虑到分期治理计划的治理进度首先由慢变快，然后在再由快变慢。这种变化规律与柯西分布函数的斜率（ $y$  的变化快慢）变化规律相似，因此可以考

考虑使用柯西分布函数来确定每个时期 PM2.5 的治理计划。

柯西分布函数为：

$$k_i = \begin{cases} 1 & i \leq 0 \\ \frac{1}{1 + \alpha i^\beta} & 1 \leq i \leq 5 \end{cases} \quad (7-8)$$

其中：  $k_i$  为第一年到第  $i$  年 PM2.5 的累计减少量与 PM2.5 初始量之比。

因此修正后最优化模型为：

$$\begin{cases} \min Z = \sum_{i=1}^5 S_i \\ C_j - C_0 = \sum_{i=1}^j x_i \\ S_i = 0.005 x_i^2 \quad i = (1, 2, \dots, 5) \\ k_i = \frac{1}{1 + \alpha i^\beta} \\ k_j = \frac{\sum_{i=1}^j x_i}{C_0} \\ 0 \leq x_i \leq C_0 \\ i, j = (1, 2, \dots, 5) \end{cases} \quad (7-9)$$

其中：  $Z$  代表五年专项治理投入总经费；

$S_i$  代表每年专项治理投入费用；

$C_0$  代表 PM2.5 初始浓度，  $C_j$  代表第  $j$  年 PM2.5 浓度；

$x_i$  代表第  $i$  年减少（治理）量；

$\alpha, \beta$  代表柯西分布函数系数；

$k_i$  为第一年到第  $i$  年 PM2.5 的累计减少量与 PM2.5 初始量之比。

### 7.2.2 分期治理最优化模型求解

假设经过五年的 PM2.5 专项治理，PM2.5 年平均浓度要求达到  $30 \mu\text{g} / \text{m}^3$ ，即  $C_5 = 30 \mu\text{g} / \text{m}^3$ ，通过分析可知初始年平均浓度为  $C_0 = 82 \mu\text{g} / \text{m}^3$ ，将数据代入到最优化模型中，利用 MATLAB 求解，求出结果如下：

#### 1、柯西分布函数的确定

模型求解结果确定了柯西函数分布系数  $\alpha = 0.48$ ， $\beta = 3.12$ ，故柯西分布函数及图像如下：

$$k_i = \begin{cases} 1 & i < 0 \\ \frac{1}{1 + 0.48 i^{3.12}} & 0 \leq i \leq 5 \end{cases} \quad (7-10)$$

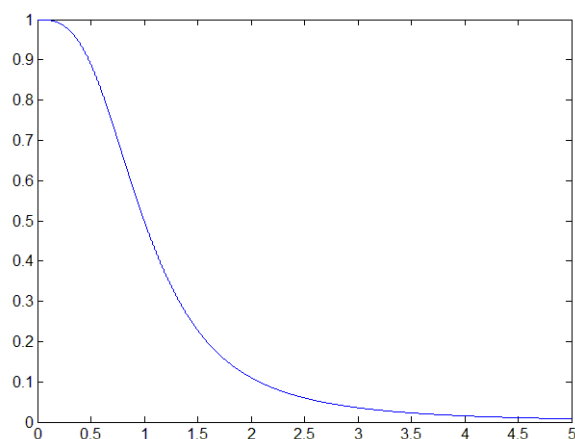


图 7-2 柯西函数分布图

## 2、分期专项治理计划的确定

表 7-3 分期专项治理计划的确定

时期	前期（准备期）	中期（治理期）			后期（稳固期）
年次	第一年	第二年	第三年	第四年	第五年
治理量百分比（%）	8.3	27.3	38.2	19.7	6.5
治理量	4.15	13.65	19.1	9.85	3.25
费用（百万元）	0.086	0.932	1.824	0.485	0.053
总费用（百万元）	3.38				

### 7.2.3 方案合理性分析

由模型结果可以看出，前期治理进度较慢（治理量百分比 8.3%），中期的治理进度最快（平均年治理百分比为 28.7%），后期治理进度最慢（年治理百分比 6.5%）。

考虑到实际环境治理过程中，往往是分期进行治理工作的，而且治理进度是先慢变快，再由快变慢，治理前期：需要进行很多的准备工作，治理过程遇到的困难也最多，进度自然很慢。治理中期：有了前期的准备工作及治理经验作基础，较前期而言，治理进度会大幅提升。治理后期：需要对前期及中期的治理成果进行稳固，而且由于治理量已经减少到一个相对较小的水平，因此如果需要使 PM2.5 继续减少，要付出的努力必然更大，因此在治理后期，治理进度会很低。

综上所述，模型的求解的治理计划与实际环境治理过程相符，因此，可以初步断定分期治理最优化模型是合理的。

## 8 模型评价与推广

### 8.1 模型评价

#### 8.1.1 模型的优点

（1）问题一中采用了相关性分析模型，灰色关联度模型，混合回归模型对 PM2.5 与其它污染物的相关性及其定量关系进行了深入探讨，并得出了令人满意

的结果。此外，还讨论了湿度指标与 PM2.5 的相关性，湿度指标的引入，使得 PM2.5 的预测准确度得到明显提升，对 PM2.5 的预测具有重要意义。

(2) 问题二中利用高斯烟羽模型对 PM2.5 的扩散分布进行分析，对西安市各地区的污染情况进行了定量与定性分析，模拟了 PM2.5 突增及其应急处理方案，得到了合理有效的结果。将模型结果与日本福岛第一核电站的放射性物质扩散数据进行对比，验证了模型的合理性。利用 PM2.5 自身沉降作用、雨水吸附作用等一般性因素对高斯扩散模型进行修正，从而得出 PM2.5 更为一般性的扩散规律。

(3) 问题三中采用了分期治理的思想将治理期分为三个时期，并利用柯西分布函数模拟出合理的治理计划。考虑费用时，同样引入了分期治理的思想，以专项治理费用最小为目标，建立了分期治理最优化模型，并得到了合理有效的结果。

### 8.1.2 模型的缺点

(1) 问题一中 PM2.5 的预测准确度为 0.91，还需要继续提高。

(2) 问题二中高斯烟羽模型的求解结果不够稳定，需要对模型进行优化。

(3) 问题三中分期治理的规划不够细化，可以将分期治理计划细化到每季度，每月，每星期。

## 8.2 模型改进

问题一中，除了引入湿度指标，还可以引入风力，温度，降雨量等更为一般性的指标，可以使得 PM2.5 的预测结果更准确。问题二中，可以考虑更多影响因素对高斯烟羽模型的影响，从而使得结果更加稳定。问题三中，可以将分期治理计划更加细化，治理计划细化到每季度，每月，每星期。

## 8.3 模型推广

本文是对 PM2.5 的相关因素、分布与演变、控制管理三个方面进行研究，研究的模型和方法，同样适用于二氧化硫 (SO<sub>2</sub>)，二氧化氮 (NO<sub>2</sub>)，一氧化碳 (CO) 等污染物。

## 9 参考文献

- [1] 环境空气质量标准, 中华人民共和国国家标准, GB 3095-2012。
- [2] 环境空气质量指数 (AQI) 技术规定 (试行): 中华人民共和国国家环境保护标准, HJ 633-2012。
- [3] 中华人民共和国环境保护部, 环境保护标准网站: <http://bz.mep.gov.cn/>。
- [4] 西安市环境监测站, 官方网站: <http://www.xianemc.gov.cn>。
- [5] PM25.in, 全国空气质量指数 (AQI) 排行榜, 官方网站: <http://www.pm25.in/rank>。
- [6] 王鹏飞, 费建芳, 程小平, 黄小刚. 福岛核泄漏物质在大气中输送扩散的数值模拟[J]. 第 28 届中国气象学会年会—S17 第三届研究生年会, 2011。
- [7] 岑况, 陈媛, 刘舒波, 于扬. 日本福岛第一核电站核泄漏后放射性物质运动轨迹[J]. 地学前缘 (中国地质大学 (北京); 北京大学). 2012. 3。
- [8] 欧阳坤. 基于三维和时态 GIS 的大气污染扩散模拟系统研究与实现[D]. 清华大学. 2011。
- [9] 云慧, 何凌燕, 黄晓锋, 兰紫娟, 李响, 曾立武. 深圳市 PM<sub>2.5</sub> 化学组成与时空分布特征[J]. 环境科学. 2013. 4。
- [10] 段国霞, 张承中, 周变红等. 西安南郊冬季 PM<sub>2.5</sub> 中重金属污染与危害分析[J]. 农业灾害研究. 2012。
- [11] 张俊波. 驻马店市空气质量时空分布特征及影响因素分析[J]. 大气科学研究与应用. 2011. 2。
- [12] 朱倩茹, 刘永红, 徐伟嘉, 黄敏. 广州 PM<sub>2.5</sub> 污染特征及影响因素分析[J]. 中国环境监测. 2012. 2。
- [13] 李芳. 西安市大气颗粒物 PM<sub>2.5</sub> 污染特征及其与降水关系研究[D]. 西安建筑科技大学. 2012. 5。

## 10 附录

部分主要程序如下：

数据分析：

1) 指数转换为浓度的函数

```
function y=IAQI2C(x,item)
m=length(x);
y=zeros(m,1);
IAQI=[0 50 100 150 200 300 400 500];
switch item
    case 'SO2'
        C=[0 50 150 475 800 1600 2100 2620];
    case 'NO2'
        C=[0 40 80 180 280 565 750 940];
    case 'PM10'
        C=[0 50 150 250 350 420 500 600];
    case 'CO'
        C=[0 2 4 14 24 36 48 60];
    case 'O3'
        C=[0 160 200 300 400 800 1000 1200];
    case 'O3_8'
        C=[0 100 160 215 165 800 inf inf];
    case 'PM2.5'
        C=[0 35 75 115 150 250 350 500];
end
n=length(IAQI)-1;
for i=1:m
    if isnan(x(i))
        y(i)=nan;
    else
        for j=1:n
            if x(i)>=IAQI(j) && x(i)<IAQI(j+1)
                IAQI_Hi=IAQI(j+1);
                IAQI_Lo=IAQI(j);
                BP_Hi=C(j+1);
                BP_Lo=C(j);
                break
            end
        end
        y(i)=(x(i)-IAQI_Lo)*(BP_Hi-BP_Lo)/(IAQI_Hi-IAQI_Lo)+BP_Lo;
    end
end
end
```

问题一：

### 1) 灰色关联度求解

```

clc,clear
load x.txt %把原始数据存放在纯文本文件 x.txt 中
for i=1:6
x(i,:)=x(i,:)/x(i,1); %标准化数据
end
data=x;
n=size(data,1);
ck=data(1,:);m1=size(ck,1);
bj=data(2:n,:);m2=size(bj,1);
for i=1:m1
for j=1:m2
t(j,:)=bj(j,:)-ck(i,:);
end
jc1=min(min(abs(t')));jc2=max(max(abs(t')));
rho=0.5;
ksi=(jc1+rho*jc2)./(abs(t)+rho*jc2);
rt=sum(ksi')/size(ksi,2);
r(i,:)=rt;
end
r
[rs,rind]=sort(r,'descend') %对关联度进行排序

```

### 问题二:

#### 1) 模拟上下风向

```

%% 上风向
Q=2; %输入源强
k=2; %输入风速
s=0.5;
u=k-s;%上风处
d=1; %步长
Z=0.4;
x=-200:d:-60; %下风向距离
y=-5:d:5; %横风向距离
[x,y]=meshgrid(x,y);
by0=0.08*x.*(1+0.0001*x).^(-1/2);
bz0=0.06*x.*(1+0.0015*x).^(-1/2);
tempy1=-y.*y./by0./by0./2;
tempy2=2.718282.^(tempy1);
c=Q/pi/u*((by0.*bz0).^(-1)).*tempy2
Cs=5; %输入求解的条数
contour(x,y,c,Cs);
shading interp;
colorbar;

```



```

% grid;
title('上风向模拟')

%% 下风向
Q=2; %输入源强
k=2; %输入风速
s=0.5;
u=k+s;
d=1; %步长
Z=0.4;
x=50:d:300; %下风向距离
y=-30:d:30; %横风向距离
[x,y]=meshgrid(x,y);
by0=0.08*x.*(1+0.0001*x).^(-1/2);
bz0=0.06*x.*(1+0.0015*x).^(-1/2);
tempy1=-y.*y./by0./by0./2;
tempy2=2.718282.^(tempy1);
c=Q/pi/u*((by0.*bz0).^(-1)).*tempy2
Cs=20; %输入求解的条数
contour(x,y,c,Cs);
shading interp;
colorbar;
% grid;
title('下风向模拟图')

```

问题三：

1) 求解问题三

%% 问题 3\_1 的求解

```

clc;
clear all;
close all;
% sysm a b
i=1:5;
k(i)=[0.9 0.6 0.2 0.05 0.0];
for a=0:100
    for b=0:100
        for i=1:5
            if i<0
                k(i)=1;
            else
                k(i)=1/(1+a*i.^b);
            end
        end
    end
end
end

```

```

end
a=0.1243;b=3.2631;
disp('问题一柯西分布函数系数为: ');
a,b
%_____ %
%% 问题 3_2 的求解
tic;
s0=0;
c0=82;
sum_xi=0;
x=[82 52 42 33 31];
for i = 1:5
    if(x(i)>=0&&x(i)<=c0)
        s(i) = 0.005*x(i).^2;
        z = s0+s(i);
        sum_xi = sum_xi+x(i);
    end
end
sum_xj0=0;
x=[4.1514 13.6485 19.104 9.8535 3.2490];
for j=1:5
    for i=1:j
        sum_xj0=sum_xj0+x(i);
        c(j)=sum_xj0+c0;
    end
end
s=[0.0863 0.9324 1.8241 0.4848 0.0527];
% sysm a b
a=0.4835;
b=3.1242;
for i=1:1:5
    k(i)=1/(1+a*i^b);
end
sum_xj1=0;
z=3.3793;
for j=1:5
    for i=1:j
        sum_xj1=sum_xj1+x(i);
        k(j)=sum_xj1/c0;
    end
end
disp('输出结果为: ');
disp('问题二柯西分布函数系数为: ');
a,b

```

```
disp('治理量 x(i)分别为: ');
for i=1:5
    x(i)
end
disp('治理费用 s(i)分别为: ');
for i=1:5
    s(i)
end
disp('总费用（百万元）Z 为: ');
z
toc;
%_____%
```