

---

(请勿改动此页内容和格式。此承诺书打印签名后作为纸质论文的封面，注意电子版论文中不得出现此页。以上内容请仔细核对，如填写错误，论文可能被取消评奖资格。)

## 摘要

近年来我国 GDP 快速增长，但空气污染的弊病日益突出。本文依据题目所给数据和信息，建立集成模型，研究江浙沪地区的空气质量问题，分析总结影响空气质量的关键参数，并提出对于环境保护工作的几点建议。

对于问题一，我们首先总结概括国标和美标对**空气质量指数 (AQI)** 计算的规定及衡量空气质量等级的标准，发现  $PM_{2.5}$  在现行标准 AQI 计算中起主要作用。为综合考虑各种污染物的影响，我们结合江浙沪地区实际情况，利用**层次分析法**为六种污染物设定权值，在此基础上建立综合评价空气质量情况的 **CAQI 模型** (Comprehensive AQI model, 综合 AQI 模型)。

对于问题二，我们首先收集江浙沪地区各城市的空气质量检测数据并进行预处理。之后一方面通过**主成分分析法**建立识别城市主要污染源的模型；另一方面在主成分分析法的基础上运用**层次聚类分析法**进行城市间污染特征相似性分析，将江浙沪地区复杂的空气污染情况，划分为**几种主要的污染类型**。进而综合分析影响空气质量的主要污染源的性质和种类。

对于问题三，我们首先根据单污染源和污染物为氮氧化物的前提条件，确定选取**高斯烟羽模型**。之后，我们搜集平均风速、太阳辐射强度等数据，确定**大气稳定度等级**等核心参数。最后，我们针对小风、静风、烟气抬升、取样时间较长等多种情况**引入修正**并结合实际不断优化数据选择，求解出污染源方圆 51 公里不同时间空气污染浓度分布和空气质量等级。

对于问题四，我们首先依据模型二分析影响空气质量的关键因素，如颗粒物污染、交通污染等；其次对模型三的参数针对 AQI 进行**灵敏度分析**，得出影响污染物排放的有效高度和排放速率等对空气质量影响较大。根据我们的模型和求解结果，得出针对不同城市、不同重点区域的建议报告。

最后，我们对模型进行了优缺点分析和推广。

**关键字：** 层次分析法 主成分分析法 层次聚类分析法 高斯烟羽模型

# 目录

<b>一、问题重述</b>	<b>5</b>
1.1 问题背景	5
1.2 问题要求	5
<b>二、问题分析</b>	<b>5</b>
2.1 问题一分析	5
2.2 问题二分析	6
2.3 问题三分析	6
2.4 问题四分析	6
<b>三、模型的假设</b>	<b>6</b>
<b>四、符号说明</b>	<b>7</b>
<b>五、模型的建立与求解</b>	<b>8</b>
5.1 问题一模型的建立及求解	8
5.1.1 模型的准备：总结 EPA AQI 和国标 AQI 计算过程	8
5.1.2 在上述总结基础上进行创新与改进	10
5.2 问题二模型的建立及求解	12
5.2.1 数据搜集与预处理	12
5.2.2 运用 PCA 模型 (主成分分析) <sup>[5;6]</sup> 提取主成分	13
5.2.3 运用 HCA(层次聚类分析) 模型 <sup>[6;7]</sup> 进行城市间特征相似性分析	14
5.2.4 数据概况与与处理结果	15
5.2.5 主成分分析 (PCA) 运行结果	15
5.2.6 层次聚类分析 (HCA) 运行结果	16
5.3 问题三模型的建立及求解	18
5.3.1 模型选取	18
5.3.2 模型建立	18
5.3.3 模型修正	21
5.3.4 问题三求解	25
5.4 问题四的求解	25
5.4.1 依据模型二分析影响空气质量的关键因素	25

5.4.2 模型三中各参数对 AQI 的灵敏度分析	26
5.4.3 建议报告	26
<b>六、模型的分析与检验</b>	<b>27</b>
6.1 模型一的分析与检验	27
6.1.1 一致性检验	27
6.1.2 数据规模与覆盖范围	27
6.1.3 CAQI 与传统 AQI 结果对比	28
6.2 模型二的分析与检验	29
6.2.1 PCA 模型有效性	29
6.2.2 HCA 模型稳定性	29
6.2.3 结果一致性检验	29
6.3 模型三的分析与检验	29
6.3.1 污染物扩散对空气质量的影响	29
6.3.2 污染物扩散的空间特征分析	30
<b>七、模型的评价</b>	<b>31</b>
7.1 模型的优点	31
7.1.1 模型一	31
7.1.2 模型二	32
7.1.3 模型三	32
7.2 模型的缺点	32
7.2.1 模型一	32
7.2.2 模型二	33
7.2.3 模型三	33
<b>八、模型的推广</b>	<b>33</b>
<b>参考文献</b>	<b>33</b>
<b>附录 A CAQI 算法</b>	<b>35</b>
<b>附录 B 高斯烟羽模型</b>	<b>36</b>
<b>附录 C 层次聚类分析</b>	<b>38</b>

## 一、问题重述

### 1.1 问题背景

随着我国经济高速发展，GDP 不断增长，工业化和城市化进程加快，化石燃料等能源消耗总量攀升，对自然环境和生态平衡产生一定负面影响，由 PM2.5、PM10、二氧化氮、二氧化硫、烟尘等因素造成的空气污染问题尤为突出，不仅影响市民生活质量，也对社会可持续发展带来更大挑战。为缓解空气污染问题，国家能源委员会出台《系能源产业振兴和发展规划》等“国家新能源发展战略”政策。如今我国发展正从高速发展阶段向高质量发展阶段转型，环境能源问题已上升至国家安全级别，强调实施节能减排措施、对能源利用方式进行创新以及大力推广新能源技术，体现了国家对于环境保护与可持续发展的高度重视。

### 1.2 问题要求

我们通过搜集并分析相关数据，运用数学思想，建立数学模型来研究下列以江浙沪地区为研究对象的空气污染相关问题：

(1) 结合我国《环境空气质量标准》与美国 EPA 标准，构建一套基于多种污染物浓度的空气质量等级评价模型。

(2) 通过查阅数据与资料，梳理江浙沪地区当前的主要空气污染来源，列出各类污染源的典型特征及相关排放参数（如污染物种类、排放强度、排放周期等），并进一步分析影响区域空气质量的主要污染源类型与性质。

(3) 建立单污染源空气扩散模型，用以描述对周围空气污染的动态影响规律。考虑废气排放烟囱高 50 米的一工厂，排放物以氮氧化物为主。早上 9 点至下午 3 点期间，排放浓度为  $406.92 \text{ mg/m}^3$ ，排放速度为  $1200 \text{ m}^3/\text{h}$ ；晚上 10 点至凌晨 4 点期间，排放浓度为  $1160 \text{ mg/m}^3$ ，排放速度为  $5700 \text{ m}^3/\text{h}$ 。利用所建立的扩散模型和空气质量等级评价模型求解该工厂方圆 51 公里分别在早上 8 点、中午 12 点、晚上 9 点空气污染浓度分布和空气质量等级。

(4) 根据上述模型和结果，分析影响空气质量的关键参数与主导因素。在此基础上，结合江浙沪地区实际情况，向环保部门撰写一份政策建议报告，提出切实可行的空气污染防控与区域治理措施建议。

## 二、问题分析

### 2.1 问题一分析

我们首先查阅相关资料，总结概括国标（GB 3095，HJ 633）和美标（EPA AQI）中对于空气质量指数（AQI）计算的规定，分析其共同点和差异。

针对国标和美标均直接采用空气质量分指数 (IAQI) 的最大值作为 AQI, 可能无法整体反映空气质量情况的问题, 我们创新性地提出用层次分析法为每种污染物设定权值, 取 IAQI 大于 50 的几种污染物 (主要污染物) 的 IAQI 求加权均值, 将该结果作为综合空气质量指数 (CAQI) (Comprehensive Air Quality Index)。

## 2.2 问题二分析

对于单个城市, 我们收集其不同日期和时间收集到的六种污染物 (污染物种类同国标和美标) 的浓度, 采用主成分分析法选定若干主成分, 之后通过主成分在各个污染物上的载荷倒推主成分为何种污染源。

对于江浙沪地区整体空气质量情况, 我们考虑采用层次聚类法将具有相似大气污染特征的城市聚类, 刻画每个“城市类”的典型画像, 进而更加整体、全面地考察影响空气质量的主要污染源。

## 2.3 问题三分析

氮氧化物为气态污染物, 因此采用气体扩散模型。

烟囱为点污染源, 考虑选用高斯烟羽模型。查阅江浙沪地区太阳辐射等级和地面风速, 确定大气稳定度, 并确定模型中其他参数。

之后, 代入数据进行计算, 分析结果并针对几种特殊情况进一步修正模型参数。

## 2.4 问题四分析

本问题要求在前三问模型构建与分析的基础上, 识别影响空气质量变化的关键因素, 针对江浙沪地区提出可行性强的治理建议。

为此, 我们首先基于问题二的 CAQI 模型和问题三的污染扩散仿真结果, 进行灵敏度分析, 探究排放速率、污染物浓度、大气稳定度、风速等关键参数对空气质量等级的影响程度。

最终, 结合模型及求解结果撰写一份面向环保部门的政策建议报告。

# 三、模型的假设

- 污染物选定  $PM_{2.5}$ ,  $PM_{10}$ ,  $O_3$ ,  $CO$ ,  $SO_2$ ,  $NO_2$ , 忽略其余少量污染物对空气质量的影响。
- 单污染物空气质量分指数 (IAQI) 计算中, 污染物浓度与健康危害在各个分段区间内呈线性关系。
- 不同污染物之间无显著非线性交互。
- 污染源 (本题即为烟囱) 均匀稳定连续排放污染物。

- 区域风向风速稳定。
- 污染物在水平和垂直方向上的浓度分布呈正态分布。
- 污染物在输送过程中质量守恒。(大气点源污染物的实际扩散中,由于大气污染物的密度很小,地面对其有很大的反射作用认为地面对污染物的反射系数为 1,即污染物没有耗散)。
- 烟囱的出口半径为 2.5 米,烟气出口温度为 150 摄氏度。

#### 四、符号说明

符号	意义	单位
$C_i$	第 i 种污染物的浓度	$\mu\text{g}/\text{m}^3$ 或 $\text{mg}/\text{m}^3$
$I_i$	第 i 种污染物的空气质量分指数 (IAQI)	
$C(x, y, z, H)$	位置 $(x, y, z)$ 的污染物浓度 (H 为参数)	$\text{mg}/\text{m}^3$
$Q$	污染物质量排放率	$\text{mg}/\text{s}$
$u$	风速	$\text{m}/\text{s}$
$H$	污染源有效高度	$\text{m}$
$\sigma_y(x)$	横向扩散参数	$\text{m}$
$\sigma_z(x)$	垂直扩散参数	$\text{m}$
$C_r$	距污染点远距离	$\text{m}$
$Q_h$	烟气燃释效率	$\text{KJ}/\text{s}$
$H_s$	烟气几何高度	$\text{m}$
$P_a$	大气压力	$\text{KPa}$
$T_s$	烟气出口温度	$\text{K}$
$T_a$	环境大气温度	$\text{K}$
$\Delta T$	烟气出口温度与环境大气温度差值	$\text{K}$
$Q_v$	实际排烟率	$\text{m}^3/\text{s}$
$u$	烟囱出风口出风速	$\text{m}/\text{s}$
$u_0$	烟囱临近气象站的平均风速	$\text{m}/\text{s}$
$\Delta H$	烟气抬升高度	$\text{m}$
$V_s$	烟囱出口处烟气排放速度	$\text{m}/\text{s}$
$D$	烟囱出口直径	$\text{m}$
$\frac{dT_a}{dZ}$	烟囱几何高度以上的大气温度梯度	$\text{K}/\text{m}$

## 五、模型的建立与求解

### 5.1 问题一模型的建立及求解

#### 5.1.1 模型的准备：总结 EPA AQI 和国标 AQI 计算过程

根据题目要求，为了建立空气质量评价模型，我们首先对国标和美标进行总结。国标参考 GB 3095—2012<sup>[1]</sup> 和 HJ 633—2012<sup>[2]</sup>(二者同步实施), 美标参考美国国家环保局 EPA-454/B-24-002<sup>[3]</sup>

##### 1. 选定污染物:

国标和美标均选定  $PM_{2.5}$ 、 $PM_{10}$ 、 $O_3$ 、 $CO$ 、 $SO_2$ 、 $NO_2$ 。

##### 2. 用线性插值公式计算 IAQI:

对于第  $i$  种污染物，其浓度为  $C_i$ ，确定其所在浓度区间  $[C_{i,low}, C_{i,high}]$  及对应 AQI 区间  $[I_{i,low}, I_{i,high}]$ （国家标准 GB 3095 和 EPA AQI 各有不同阈值）。

$$IAQI_i = \frac{I_{i,high} - I_{i,low}}{C_{i,high} - C_{i,low}} (C_i - C_{i,low}) + I_{i,low} \tag{1}$$

表 2 空气质量分指数及对应的污染物项目浓度限值

摘自 HJ 633-2012

空气质量分指数 (IAQI)	污染物项目浓度限值									
	二氧化硫 ( $SO_2$ ) 24 小时平均 ( $\mu g/m^3$ )	二氧化硫 ( $SO_2$ ) 1 小时平均 ( $\mu g/m^3$ ) (1)	二氧化氮 ( $NO_2$ ) 24 小时平均 ( $\mu g/m^3$ )	二氧化氮 ( $NO_2$ ) 1 小时平均 ( $\mu g/m^3$ ) (1)	颗粒物 (粒径小于等于 $10\mu m$ ) 24 小时平均 ( $\mu g/m^3$ )	一氧化碳 ( $CO$ ) 24 小时平均 ( $mg/m^3$ )	一氧化碳 ( $CO$ ) 1 小时平均 ( $mg/m^3$ ) (1)	臭氧 ( $O_3$ ) 1 小时平均 ( $\mu g/m^3$ )	臭氧 ( $O_3$ ) 8 小时平均 ( $\mu g/m^3$ )	颗粒物 (粒径小于等于 $2.5\mu m$ ) 24 小时平均 ( $\mu g/m^3$ )
0	0	0	0	0	0	0	0	0	0	0
50	50	150	40	100	50	2	5	160	100	35
100	150	500	80	200	150	4	10	200	160	75
150	475	650	180	700	250	14	35	300	215	115
200	800	800	280	1200	350	24	60	400	265	150
300	1600	(2)	565	2340	420	36	90	800	800	250
400	2100	(2)	750	3090	500	48	120	1000	(3)	350
500	2620	(2)	940	3840	600	60	150	1200	(3)	500
说明:	<p>(1) 二氧化硫 (<math>SO_2</math>)、二氧化氮 (<math>NO_2</math>) 和一氧化碳 (<math>CO</math>) 的 1 小时平均浓度限制仅用于实时报，在日报中需使用相应污染物的 24 小时平均浓度限值。</p> <p>(2) 二氧化硫 (<math>SO_2</math>) 1 小时平均浓度值高于 <math>800\mu g/m^3</math> 的，不再进行其空气质量分指数计算，二氧化硫 (<math>SO_2</math>) 空气质量分指数按 24 小时平均浓度计算的分指数报告。</p> <p>(2) 臭氧 (<math>O_3</math>) 8 小时平均浓度值高于 <math>800\mu g/m^3</math> 的，不再进行其空气质量分指数计算，臭氧 (<math>O_3</math>) 空气质量分指数按 1 小时平均浓度计算的分指数报告。</p>									



表 3 美国国家环保局 (EPA) 标准

These Breakpoints...							...equal this AQI	...and this category
$O_3$ (ppm) 8-hour	$O_3$ (ppm) 1-hour <sup>1</sup>	$PM_{2.5}$ ( $\mu g/m^3$ ) 24-hour	$PM_{10}$ ( $\mu g/m^3$ ) 24-hour	$CO$ (ppm) 8-hour	$SO_2$ (ppb) 1-hour	$NO_2$ (ppb) 1-hour	AQI	
0.000-0.054	-	0.0-9.0	0-54	0.0-4.4	0-35	0-53	0-50	Good
0.055-0.070	-	9.1-35.4	55-154	4.5-9.4	36-75	54-100	51-100	Moderate
0.071-0.085	0.125-0.164	35.5-55.4	155-254	9.5-12.4	76-185	101-360	101-150	Unhealthy for Sensitive Groups
0.086-0.105	0.165-0.204	(55.5-125.4) <sup>3</sup>	255-354	12.5-15.4	<sup>3</sup> 186-304	361-649	151-200	Unhealthy
0.106-0.200	0.205-0.404	(125.5-225.4) <sup>3</sup>	355-424	15.5-30.4	<sup>3</sup> 305-604	650-1249	201-300	Very unhealthy
0.201-( <sup>2</sup> )	0.405+	225.5+	425+	30.5+	<sup>3</sup> 605+	1250+	301+	Hazardous <sup>4</sup>

<sup>1</sup> Areas are generally required to report the AQI based on 8-hour  $O_3$  values. However, there are a small number of areas where an AQI based on 1-hour  $O_3$  values would be more precautionary. In these cases, in addition to calculating the 8-hour  $O_3$  index value, the 1-hour  $O_3$  value may be calculated, and the maximum of the two values reported.

<sup>2</sup> 9-hour  $O_3$  values do not define higher AQI values ( $\geq 301$ ). AQI values of 301 or higher are calculated with 1-hour  $O_3$  concentrations.

<sup>3</sup> 1-hr  $SO_2$  concentrations do not define higher AQI values ( $\geq 200$ ). AQI values of 200 or greater are calculated with 24-hour  $SO_2$  concentration.

<sup>4</sup> AQI values between breakpoints are calculated using equation 1 to this appendix. For AQI values in the hazardous category, AQI values greater than 500 should be calculated using equation 1 and the concentration specified for the AQI value of 500. The AQI value of 500 are as follows:  $O_3$  1-hour—0.604ppm;  $PM_{2.5}$  24-hour—325.4 $\mu g/m^3$ ;  $PM_{10}$  24-hour—604 $\mu g/m^3$ ;  $CO$  ppm—50.4ppm;  $SO_2$  1-hour—1004ppb; and  $NO_2$  1-hour—2049ppb.

### 3. 计算 AQI:

国标、美标均采用以下计算公式:

利用分项空气质量指数 (IAQI) 计算 AQI

$$AQI = \max(IAQI_i) \quad (2)$$

表 4 AQI 标准对照表

AQI 范围	美标描述	国标描述
0-50	Good	优
51-100	Moderate	良
101-150	Unhealthy for Sensitive Groups	轻度污染
151-200	Unhealthy	中度污染
201-300	Very Unhealthy	重度污染
301+	Hazardous	严重污染

### 5.1.2 在上述总结基础上进行创新与改进

国标和美标均直接采用  $\max(IAQI_i)$  作为最终 AQI，查阅江浙沪地区 AQI 资料，发现基本为  $PM_{2.5}$  的 IAQI，不能充分反映污染的程度。

因此，我们尝试用层次分析法 AHP<sup>[4]</sup> 优化计算，为便于区分，我们将这种做法计算所得的空气质量指数称为 CAQI。

## AHP

### 1. 原理:

通过构建污染物重要性判断比较矩阵，将专家知识转化为定量权重。进而综合反映各污染物对 AQI 的影响。

### 2. 构造判断矩阵:

专家对污染物两两比较重要性（1-9 标度法）。

表 5 AHP 判断矩阵标度含义

标度	含义
1	表示两个元素相比，具有同样的重要性
3	表示两个元素相比，前者比后者稍重要
5	表示两个元素相比，前者比后者明显重要
7	表示两个元素相比，前者比后者极其重要
9	表示两个元素相比，前者比后者强烈重要
2,4,6,8	表示上述相邻判断的中间值
1 ~ 9 的倒数	表示相应两因素交换次序比较的重要性

查阅资料，结合江浙沪地区实际情况，得出六种污染物的比较判断矩阵如下表：

表 6 六种污染物的比较判断矩阵

	$SO_2$	$CO$	$NO_2$	$O_3$	$PM_{2.5}$	$PM_{10}$
$SO_2$	1	4	1/2	1/3	1/3	1/2
$CO$	1/4	1	1/3	1/4	1/4	1/3
$NO_2$	2	3	1	1/2	1/2	1
$O_3$	3	4	2	1	1	2
$PM_{2.5}$	3	4	2	1	1	2
$PM_{10}$	2	3	1	1/2	1/2	1

### 3. 计算权重向量:

- 求矩阵最大特征值  $\lambda_{\max}$  对应的特征向量  $W = [w_1, w_2, \dots, w_n]^T$ .
- 归一化:

$$w_i = \frac{w_i}{\sum w_i} \quad (3)$$

计算得  $\lambda_{\max} = 6.154$ ,  $W = [0.684, 0.335, 1, 1.780, 1.780, 1]^T$ , 归一化为  $[0.104, 0.051, 0.152, 0.271, 0.271, 0.152]^T$ 。

#### 4. 得出权重 $w_i$ :

表 7 六种污染物的权重

	$SO_2$	$CO$	$NO_2$	$O_3$	$PM_{2.5}$	$PM_{10}$
权重	0.104	0.051	0.152	0.271	0.271	0.152

#### 优化结果

##### 1. $IAQI$ 取主要污染物加权均值:

$$CAQI = \begin{cases} \frac{\sum_{i=1}^n w_i IAQI_i}{\sum_{j=1}^n w_j} & \text{if } \exists IAQI_i \geq 50 \\ \max(IAQI_1, IAQI_2, \dots, IAQI_n) & \text{otherwise} \end{cases} \quad (4)$$

##### 2. 重新划定 $CAQI$ 阈值如下:

表 8 Tok-k  $CAQI$  阈值

CAQI 范围	美国国家标准描述	中国国家标准描述
0-50	Good	优
51-90	Moderate	良
91-140	Unhealthy for Sensitive Groups	轻度污染
141-180	Unhealthy	中度污染
181-270	Very Unhealthy	重度污染
270+	Hazardous	严重污染

## 5.2 问题二模型的建立及求解

针对一个城市的污染源，我们选用主成分分析法 (PCA) 进行分析; 为从整体上把握江浙沪地区的污染源情况，我们选用层次聚类分析 (HCA) 对江浙沪城市进行分类。

### 5.2.1 数据搜集与预处理

1. 从生态环境部门、统计局、气象局等官方渠道获取公开数据，收集江浙沪地区（上海、江苏、浙江）各城市的空气质量监测数据（ $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ ,  $NO_2$ ,  $CO$ ,  $O_3$

等), 以及可能的污染源排放数据 (工业排放、交通流量、电厂燃煤量、居民生活燃烧、农业活动等)。

2. 数据清洗, 采用  $3\sigma$  原则处理异常值。
3. 对数据进行 Z-score 标准化。

对 (一个城市)  $m$  个样本的  $n$  种污染物浓度矩阵  $X_{m \times n}$  标准化:

$$y_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (5)$$

$$\mu_j = \frac{1}{m} \sum_{i=1}^m x_{ij} \quad (6)$$

$$\sigma_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (x_{ij} - \mu_j)^2} \quad (7)$$

得到标准化矩阵  $Y_{m \times n}$ . 其中  $x_{ij}$  为第  $i$  个样本的第  $j$  个污染物的原始浓度值,  $\mu_j$  为污染物  $j$  的平均值,  $\sigma_j$  为污染物  $j$  的标准差。

### 5.2.2 运用 PCA 模型 (主成分分析)<sup>[5;6]</sup> 提取主成分

#### PCA 模型原理:

1. 数据标准化 (5.2.1 已完成)。
2. 计算协方差矩阵:

$$R = \frac{1}{m-1} Y^T Y \quad (8)$$

反映污染物间线性关系。

3. 特征分解

求  $R$  的特征值  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  和对应特征向量  $v_1, v_2, \dots, v_n$

其中:

- $\lambda_k$ : 第  $k$  主成分解释的原始数据方差 (越大说明该方向越重要)。
- $v_k$ : 污染物在该主成分上的投影强度 (绝对值越大贡献越大)。

4. 选择主成分数量: 我们选择前  $k$  个起主要作用的主成分。前  $k$  个主成分的累计方差贡献率是  $\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i}$ , 我们选取  $k$  使得累计方差贡献率达到 80% 以上。

以主成分序号为横轴, 以特征值大小为纵轴绘制碎石图 (Scree Plot), 通过碎石图的拐点等特征验证主成分选取的合理性。

5. 代入数据计算后, 选定  $k$  个特征向量  $v_1, \dots, v_k$ , 得到载荷矩阵  $P = [v_1, v_2, \dots, v_k]$ , 行代表原始污染物变量, 列代表提取的主成分。
6. 载荷表示原始污染物与该主成分之间的相关程度。载荷的绝对值越大, 表示该污染物在该主成分中的贡献越大。

分析:

- 如果 PC1 在  $CO$ 、 $NO_2$ 、 $PM_{2.5}$  上的载荷较高，而  $SO_2$  载荷低，则 PC1 可能代表交通源。
- 如果 PC2 在  $SO_2$ 、 $PM_{2.5}$ 、 $NO_2$  上的载荷较高，则 PC2 可能代表工业燃烧或燃煤源。
- 如果 PC3 主要在  $PM_{10}$  上载荷高，可能代表扬尘源。
- 如果 PC4 主要在  $O_3$  上载荷高，其他污染物载荷低，可能代表光污染源。

由此我们就得到了某个城市（监测点）的  $k$  个主要污染源。

### 5.2.3 运用 HCA(层次聚类分析) 模型<sup>[6;7]</sup> 进行城市间特征相似性分析

我们采用凝聚型 HCA，将江浙沪地区的空气污染源状况划分为几个类型。

#### 数据准备：

1. 对各个城市计算其 6 种污染物年均值或季度均值，得到一个  $m \times 6$  矩阵（ $m$  为城市数量）。
2. 每一行表示一个城市在  $PM_{2.5}$ 、 $PM_{10}$ 、 $SO_2$ 、 $NO_2$ 、 $CO$ 、 $O_3$  上的特征。
3. 数据进行 Z-score 标准化。
4. PCA 计算主成分得分矩阵。

#### HCA 原理：

1. 使用主成分得分矩阵，每个城市在前  $k$  个主成分上的得分构成其特征向量  $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{ik}]$ 。将每个城市抽象为  $k$  维空间中的一个点考虑。
2. 用距离量化城市之间的“不相似程度”，本题选用欧几里得距离。

对于两个  $k$  维点  $A = (a_1, \dots, a_k)$  和  $B = (b_1, \dots, b_k)$ ：

$$d(A, B) = \sqrt{\sum_{j=1}^k (a_j - b_j)^2} \quad (9)$$

3. 选用链接准则刻画两个城市簇之间的距离。本题选用平均连接法 (Average Linkage)，用两个簇之间所有点对距离的平均值刻画城市簇之间距离。假设簇 A 包含  $n_A$  个点，簇 B 包含  $n_B$  个点。则：

$$D(A, B) = \frac{1}{n_A n_B} \sum_{x \in A} \sum_{y \in B} d(x, y) \quad (10)$$

4. 构建聚类树（谱系图 Dendrogram）

- 横轴表示城市，初始时每个城市自成一簇。
- 纵轴表示聚合过程中簇之间的“距离”。
- 重复：计算所有现有簇之间的距离，合并距离最近的两个簇，每次合并形成一个“树枝节点”；

- 直到：所有数据点合并成一个大簇。

观察谱系图，通过距离跳跃法设定聚类阈值，确定最优聚类数。

5. 解释聚类效果

再次查看每个簇在主成分得分上的平均值，描绘每个簇的“典型画像”。将江浙沪地区的复杂空气污染情况，简化为几种主要的污染类型区域。

5.2.4 数据概况与与处理结果

1. 数据质量控制:

本研究收集了江浙沪地区 18 个主要城市的空气质量数据，时间范围为 2024 年 7 月至 2025 年 6 月。通过 Z-score 标准化处理，确保各污染物指标具有可比性。

2. 污染物浓度分布特征:

各城市污染物年均浓度特征显示明显的区域差异，为后续聚类分析提供了基础。

5.2.5 主成分分析 (PCA) 运行结果

1. 主成分提取结果:

各城市主成分分析结果显示，大部分城市需要 4 个主成分达到 90% 以上的方差贡献率，表明污染源结构相对复杂。

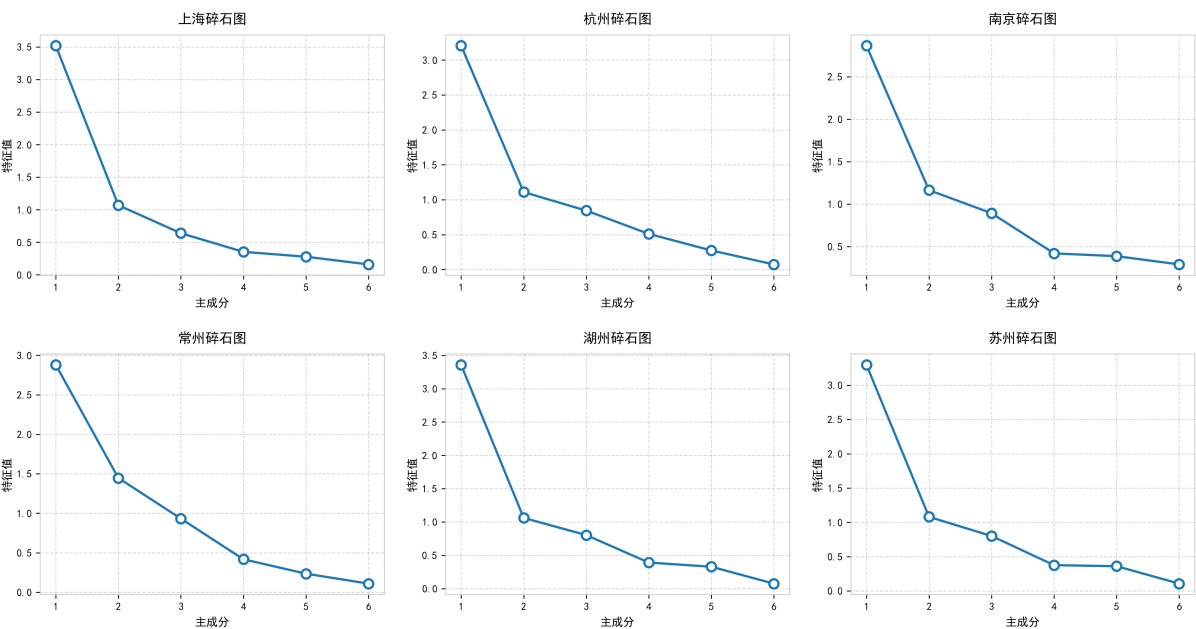


图 1 各城市主成分分析碎石图

2. 污染源识别结果:

表 9 主要污染物及污染源识别结果

城市	主成分	主要污染物 (载荷)	污染源类型
上海	PC1	$PM_{2.5}(0.481)$ , $PM_{10}(0.475)$	颗粒物源
	PC2	$O_3\text{-}8h(0.938)$	光化学源
	PC3	$SO_2(0.787)$ , $CO(0.512)$	燃烧源
	PC4	$CO(0.673)$ , $NO_2(0.425)$	机动车源
杭州	PC1	$PM_{2.5}(0.526)$ , $PM_{10}(0.504)$	颗粒物源
	PC2	$O_3\text{-}8h(0.868)$	光化学源
	PC3	$CO(0.716)$ , $SO_2(0.503)$	燃烧源
	PC4	$SO_2(0.647)$ , $CO(0.528)$	工业源
南京	PC1	$NO_2(0.506)$ , $CO(0.485)$	交通源
	PC2	$O_3\text{-}8h(0.827)$ , $PM_{2.5}(0.512)$	复合源
	PC3	$PM_{2.5}(0.693)$ , $SO_2(0.480)$	工业源
	PC4	$PM_{10}(0.769)$ , $CO(0.452)$	扬尘源
	PC5	$CO(0.712)$ , $SO_2(0.448)$	燃烧源

### 3. 关键发现:

- (a) 交通污染是区域主要污染源，多个城市的第一主成分为机动车排放。
- (b) 光化学污染普遍存在，所有城市都存在明显的光化学污染特征。
- (c) 工业污染存在区域分化，部分城市工业燃烧源突出。
- (d) 南京等城市污染源结构最为复杂，需要更多主成分解释。

## 5.2.6 层次聚类分析 (HCA) 运行结果

### 1. 聚类方法选择:

采用平均连接法进行层次聚类分析，使用欧几里得距离度量城市间污染特征相似性。通过距离跳跃法确定最优聚类数为 4 个簇。



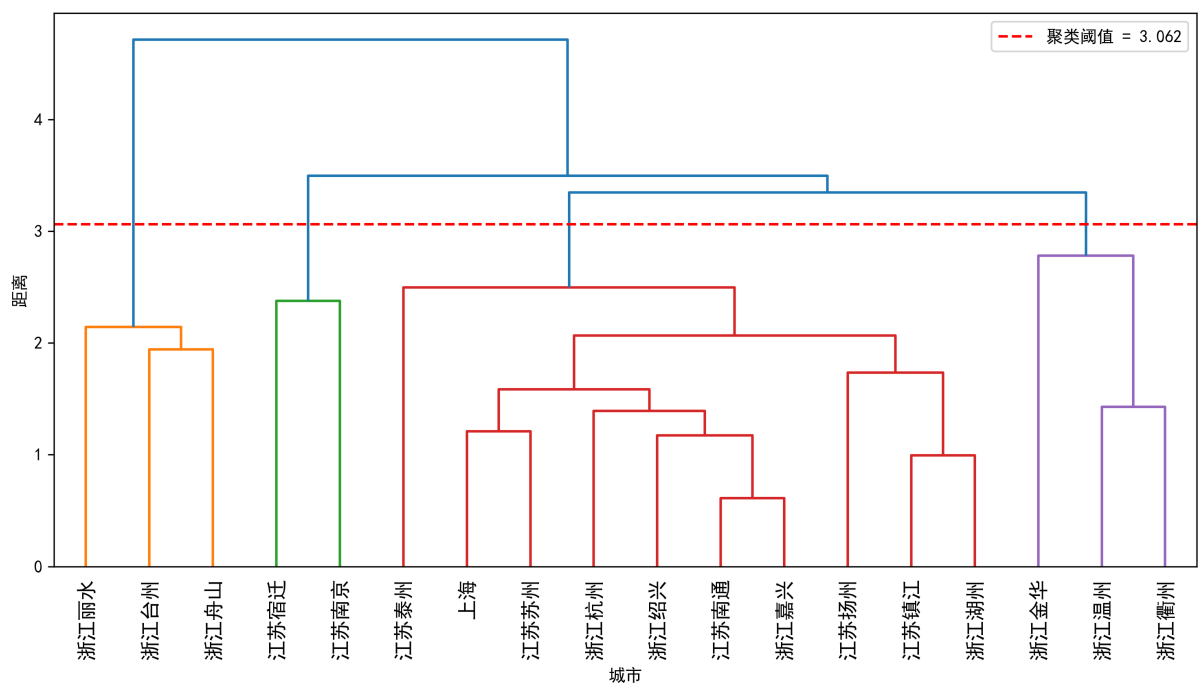


图 2 江浙沪部分城市空气污染特征层次聚类树状图

## 2. 聚类结果分析:

表 10 江浙沪部分城市空气污染特征层次聚类分析结果

簇号	包含城市	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub> -8h	污染特征
1	台州, 舟山, 丽水	-1.363	-1.621	-0.698	-1.848	-1.302	-1.585	NO <sub>2</sub> 相对较低
2	宿迁, 南京	2.095	1.094	-0.726	-0.671	-0.020	0.600	颗粒物污染较重
3	南通, 泰州, 扬州, 镇江, 绍兴, 嘉兴, 湖州, 杭州, 上海, 苏州	0.023	0.256	0.689	0.534	0.691	0.588	交通污染较重
4	温州, 金华, 衢州	-0.111	0.039	-1.115	0.513	-0.988	-0.776	SO <sub>2</sub> 相对较低

注：所有污染物浓度均为 Z-score 标准化处理后的结果。

### (a) 簇 1：空气质量优良城市

- 包含城市：浙江台州、浙江舟山、浙江丽水等。
- 污染特征：所有污染物指标都较低，空气质量最好。
- 地理特征：沿海城市，环境容量较大。

(b) 簇 2：颗粒物污染较重城市

- 包含城市：江苏宿迁、江苏南京等。
- 污染特征： $PM_{2.5}$  和  $PM_{10}$  浓度较高，颗粒物污染突出。
- 地理特征：内陆城市，受地形和气象条件影响。

(c) 簇 3：交通污染较重城市

- 包含城市：江苏南通、江苏泰州、江苏扬州、江苏镇江、浙江绍兴、浙江嘉兴、浙江湖州、浙江杭州、上海、江苏苏州等。
- 污染特征： $NO_2$  和  $CO$  浓度较高，机动车尾气排放特征明显。
- 地理特征：长三角核心区域，人口密集、车流量大。

(d) 簇 4： $SO_2$  相对较低城市

- 包含城市：浙江温州、浙江金华、浙江衢州等。
- 污染特征： $SO_2$  浓度较低，其他指标中等水平。
- 地理特征：浙江中西部地区，产业结构相对清洁。

3. 聚类特征深度分析:

(a) 簇间差异分析:

- 簇 1 与簇 3 差异最大，反映空气质量优良城市与工业污染城市的明显对比。
- 簇 2 具有独特的颗粒物污染特征，与其他簇区分明显。
- 簇 4 在  $SO_2$  指标上表现突出，体现了区域污染源结构的多样性。

(b) 地理分布特征:

聚类结果与城市地理位置高度相关，沿海城市倾向于聚为一类，内陆工业城市形成另一类，体现了地理环境对污染特征的重要影响。

## 5.3 问题三模型的建立及求解

### 5.3.1 模型选取

氮氧化物为气态污染物，烟囱可近似为点源污染物，故进行大气点源污染扩散模拟。

综合考虑选用高斯烟羽模型<sup>[8;9]</sup>进行扩散模拟。

### 5.3.2 模型建立

以污染点源在地面的投影为坐标原点， $x$  轴沿风向， $y, z$  分别为风向切向、地表水平面垂向，建立高斯模型坐标系。

$C(x, y, z, H)$ : 位置  $(x, y, z)$  的污染物浓度 ( $mg/m^3$ ) ( $H$  为参数);

$Q$ : 污染物质量排放率 ( $mg/s$ );

$u$ : 风速 ( $m/s$ );

$H$ : 污染源有效高度 ( $m$ );

$\sigma_y(x), \sigma_z(x)$ : 横向扩散参数和垂直扩散参数 ( $m$ ), 随下风距离  $x$  而变化。

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

由此我们初步建立本题求解所需模型:

1. 高架点源地面浓度公式, 令  $z = 0$ , 得:

$$C(x, y, 0, H) = \frac{Q}{\pi u \sigma_y(x) \sigma_z(x)} \exp\left[-\frac{1}{2}\left(\frac{y^2}{\sigma_y^2(x)} + \frac{H^2}{\sigma_z^2(x)}\right)\right] \quad (12)$$

2. 在Equation 12的基础上, 进一步令  $y = 0$ , 可得下风轴线 (沿  $x$  轴方向) 上的浓度分布公式:

$$C(x, 0, 0, H) = \frac{Q}{\pi u \sigma_y(x) \sigma_z(x)} \exp\left(-\frac{H^2}{2\sigma_z^2(x)}\right) \quad (13)$$

3. 扩散参数经验公式

用 *Pasquill-Gifford* 曲线拟合表达式 (单位:  $m$ ):

$$\sigma_y(x) = \gamma_1 x^{\alpha_1}, \quad \sigma_z(x) = \gamma_2 x^{\alpha_2}, \quad (14)$$

其中  $(\gamma_1, \alpha_1, \gamma_2, \alpha_2)$  依大气稳定度类别  $A-F$  选取。

中国国家标准 GB/T 3840-91<sup>[10]</sup> 中对大气稳定度划分的规定如下:

**表 11 大气稳定度的等级**

地面风速 <sup>1)</sup> $m \cdot s^{-1}$	太阳辐射等级					
	3	2	1	0	-1	-2
$\leq 1.9$	A	A~B	B	D	E	F
2~2.9	A~B	B	C	D	E	F
3~4.9	B	B~C	C	D	D	E
5~5.9	C	C~D	D	D	D	D
$\geq 6$	D	D	D	D	D	D

注: 地面风速 ( $m \cdot s^{-1}$ ) 系指离地面 10m 高度处 10 分钟平均风速, 如使用气象台 (站) 资料, 其观测规则见《地面气象观测规范》。

表 12 横向扩散参数幂函数表达式系数值  $\sigma_y = \gamma_1 x^{\alpha_1}$

(取样时间 0.5h)

稳定度	$\alpha_1$	$\gamma_1$	下风距离, m
A	0.901074	0.425809	0~1000
	0.850934	0.602052	1000
B	0.91437	0.281846	0~1000
	0.865014	0.396353	1000
B-C	0.919325	0.2295	0~1000
	0.875086	0.314238	1000
C	0.924279	0.177154	1~1000
	0.885157	0.232123	1000
C-D	0.926849	0.14394	1~1000
	0.88694	0.189396	1000
D	0.929418	0.110726	1~1000
	0.888723	0.146669	1000
D-E	0.925118	0.0985631	1~1000
	0.892794	0.124308	1000
E	0.920818	0.0864001	1~1000
	0.896864	0.101947	1000
F	0.929418	0.0553634	0~1000
	0.88723	0.0733348	1000

表 13 垂直扩散参数幂函数表达式系数值  $\sigma_Z = \gamma_2 x^{\alpha_2}$

稳度度	$\alpha_2$	$\gamma_2$	下风距离, m
A	1.12154	0.0799904	0~300
	1.5136	0.00854771	300~500
	2.10881	0.000211545	500
B	0.964435	0.12719	0~500
	1.09356	0.057025	500
B-C	0.941015	0.114682	0~500
	1.0077	0.0757182	500
C	0.917595	0.106803	0~0
C-D	0.838628	0.126152	0~2000
	0.75641	0.235667	2000~10000
	0.815575	0.136659	10000
D	0.826212	0.104634	1~1000
	0.632023	0.400167	1000~10000
	0.55536	0.810763	10000
D-E	0.776864	0.111771	0~2000
	0.572347	0.528992	2000~10000
	0.499149	1.0381	10000
E	0.78837	0.0927529	0~1000
	0.565188	0.433384	1000~10000
	0.414743	1.73421	10000
F	0.7844	0.0620765	0~1000
	0.525969	0.370015	1000~10000
	0.322659	2.40691	10000

分析查得参数应用于高斯烟羽模型。

### 5.3.3 模型修正

#### 1. 风速较小时修正：

以上分析在风速  $u > 1.5m/s$  时能得到较好的结果,但在风速  $u < 1.5m/s$  时没有明显的风向,而且烟囱出风口处的风速可能大于  $u$ ,与模型的假设不符。

表 14 小风、静风扩散参数表达式系数值 (取样时间 0.5h)

扩散参数	大气稳定度 ( <i>P.S</i> )	<i>y</i>	$\alpha$	下风距离 ( <i>m</i> )
$\sigma_z$	A	0.228205	1.16593	0-500
		0.049064	1.41327	500-2000
		0.017258	1.55074	2000
	B	0.0360763	1.01128	0-1000
		0.192024	1.110256	1000
	C	0.426406	0.912511	0
	D	0.446905	0.855756	0-1000
		1.30023	0.701154	1000
	E	0.523275	0.77422	0-1000
		1.408	0.63929	0-3000
		4.09832	0.497485	3000
	F	0.64	0.69897	0-1000
		1.024	0.630929	0-3000
		4.65031	0.441928	3000

此时有经验公式：

$$C_r = \sqrt{\frac{2}{\pi}} \frac{q}{2\pi ur\sigma_z(x)} \exp\left(-\frac{H^2}{2\sigma_z^2(x)}\right) \quad (15)$$

其中  $C_r$  表示距污染源距离为  $r$ (单位:  $m$ ) 处的污染物浓度。此时要修正  $\sigma_z$ ：

## 2. 烟气抬升高度修正：

烟囱排出的烟气一开始主要受到热力和动力作用上的抬升，然后才会以扩散作用为主。烟气到达一定高度后趋于稳定，此高度为烟气抬升高度，用  $\Delta H$  表示。

$Q_h$ : 烟气热释效率 ( $KJ/s$ );

$H_s$ : 烟气几何高度 ( $m$ );

$P_a$ : 大气压力 ( $KPa$ )

$T_s$ : 烟气出口温度 ( $K$ );

$T_a$ : 环境大气温度 ( $K$ );

$\Delta T = T_s - T_a$ ;

$Q_v$  实际排烟率 ( $m^3/s$ );

$u$ : 烟囱出风口出风速 ( $m/s$ )。

取临近气象站平均风速  $u_0(m/s)$  进行修正, 修正方法:

$$u = \begin{cases} u_0 \left( \frac{H_s}{10} \right)^m & \text{当 } H_s \leq 200\text{m} \\ u_0 \left( \frac{200}{10} \right)^m & \text{当 } H_s > 200\text{m} \end{cases} \quad (16)$$

$$Q_h = 3.5 P_a Q_v \Delta T / T_s \quad (17)$$

其中  $m$  为风廓线幂指数值:

**表 15 风廓线幂指数值**

地区	A	B	C	D	E	F
城市	0.10	0.15	0.20	0.25	0.30	0.30
农村	0.07	0.07	0.10	0.15	0.25	0.25

• 有风、中性或不稳定时:

(a) 当烟气热释放率  $Q_h \geq 2100 \text{ KJ/s}$  且  $\Delta T \geq 35\text{K}$  时:

$$\Delta H = n_0 Q_h^{n_1} H^{n_2} / u \quad (19)$$

$\Delta H$ : 烟气抬升高度 ( $m$ )

$n_0$ : 烟气及地表状况系数

$n_1$ : 烟气热释放率指数

$n_2$ : 烟气高度指数

**表 16  $n_0, n_1, n_2$  的选取**

$Q_h (KJ/S)$	地表状况 (平原)	$n_0$	$n_1$	$n_2$
$Q_h \leq 2100$	农村或城市远郊区	1.427	1/3	2/3
	城市及近郊区	1.303	1/3	2/3
$2100 \leq Q_h < 21000$ 且 $\Delta T \geq 35K$	农村或城市远郊区	0.332	3/5	2/5
	城市及近郊区	0.292	3/5	2/5

(b) 当  $1700\text{KJ/s} \leq Q_h \leq 2100\text{KJ/s}$  且  $\Delta T \geq 35\text{K}$  时:

$$\Delta H = \Delta H_1 + (\Delta H_2 - \Delta H_1) \frac{Q_h - 1700}{400} \quad (20)$$

其中:

$$\Delta H_1 = \frac{2(1.5V_s D + 0.01Q_h)}{u} - \frac{0.048(Q_h - 1700)}{u} \quad (21)$$

$V_s$ : 烟囱出口处烟气排放速度 ( $m/s$ )

$D$ : 烟囱出口直径 ( $m$ )

$\Delta H_2$ : 按Equation 19计算。

(c) 当  $Q_h \leq 1700\text{KJ/s}$  或  $\Delta T < 35\text{K}$  时:

$$\Delta H = \frac{3.0V_s D + 0.02Q_h}{u} \quad (22)$$

(d) 有风且稳定:

$$\Delta H = Q_h^{1/3} \left( \frac{dT_a}{dZ} + 0.0098 \right)^{-1/3} u^{1/3} \quad (23)$$

其中  $\frac{dT_a}{dZ}$ : 烟囱几何高度以上的大气温度梯度 ( $K/m$ )。

(e) 静风和小风时:

$$\Delta H = 5.50Q_h^{1/4} \left( \frac{dT_a}{dZ} + 0.0098 \right)^{-3/8} \quad (24)$$

### 3. 扩散系数修正:

当烟气热释放率大于  $2100\text{KJ/s}$  时, 烟气在巨大的热力作用下导致烟气抬升, 会加大  $\sigma_z$

$$\sigma'_z = \sigma_z + \left( \frac{\Delta H}{k} \right)^2 \quad (25)$$

其中  $k$  为无量纲的经验修正系数:

- 平坦地形:  $k \approx 1.4$
- 丘陵/复杂地形:  $k \approx 1.8$
- 通用建议值:  $k \in [1.0, 2.0]$

### 4. 取样时间修正:

取样时间大于  $0.5\text{h}$  时需要修正  $\sigma_y$

时间稀释公式:

$$\sigma_{y_{t_1}} = \sigma_{y_{t_2}} \left( \frac{t_1}{t_2} \right)^q \quad (26)$$

式中参数:

- $q$ : 时间稀释指数, 取值见下表
- $t_1, t_2$ : 烟气扩散时间



表 17 时间稀释指数  $q$  的取值

稳定度等级	0.5-2 小时	2-24 小时
B	0.27	0.36
B-C	0.29	0.39
C	0.31	0.42
C-D	0.33	0.45
D	0.35	0.48

5.3.4 问题三求解

通过分析过去一年（2024 年 7 月至 2025 年 6 月）早中晚的太阳辐射强度、平均气温、平均风速等气象因素，早中晚分别选择 C,B,D 作为大气稳定度，得出 51km 内的污染物浓度如下表所示：

表 18 不同大气稳定度下污染物浓度随下风向距离的变化

时间	大气稳定度等级	2km ( $\mu g/m^3$ )	5km ( $\mu g/m^3$ )	10km ( $\mu g/m^3$ )	51km ( $\mu g/m^3$ )
早上	C (中性)	0.884	0.142	0.046	0.00320
中午	B (不稳定)	0.031	0.017	0.005	0.00022
晚上	D (稳定)	0.861	0.222	0.078	0.00440

5.4 问题四的求解

5.4.1 依据模型二分析影响空气质量的关键因素

在模型二中得出：江浙沪地区城市大致分为 4 类，其中空气污染较为严重的有簇 2（颗粒物污染较重城市）和簇 3（交通污染较重城市）。

1. 颗粒物污染较重城市：

以宿迁、南京等城市为代表。

颗粒物污染主要指  $PM_{2.5}$  和  $PM_{10}$  浓度较高，受工业燃烧源影响较大。

2. 交通污染较重城市：

以上海、苏州等城市为代表。

$NO_2$  和  $CO$  浓度较高，机动车尾气排放对空气质量影响大。

5.4.2 模型三中各参数对 AQI 的灵敏度分析

为了检验模型对实际情况中各中因素的敏感性，我们在保持污染物浓度不变的前提下，分别调整有效高度、排放速率等参数，观察 AQI 的相对变化趋势。

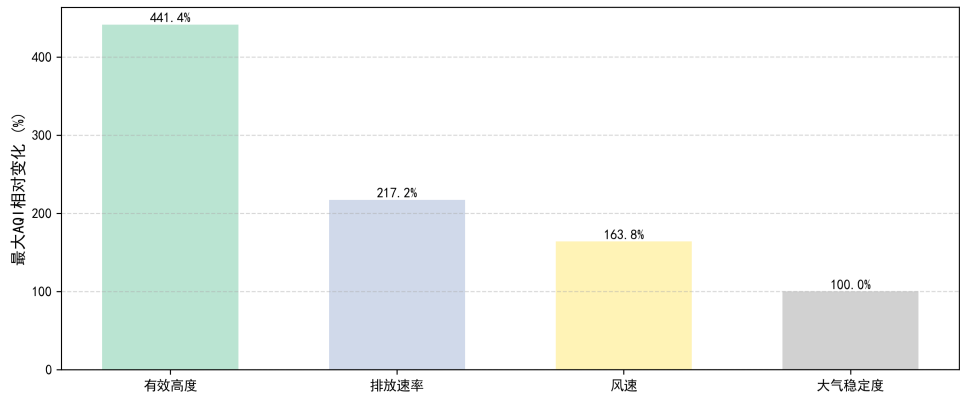


图 3 各参数对最大 AQI 的灵敏度分析

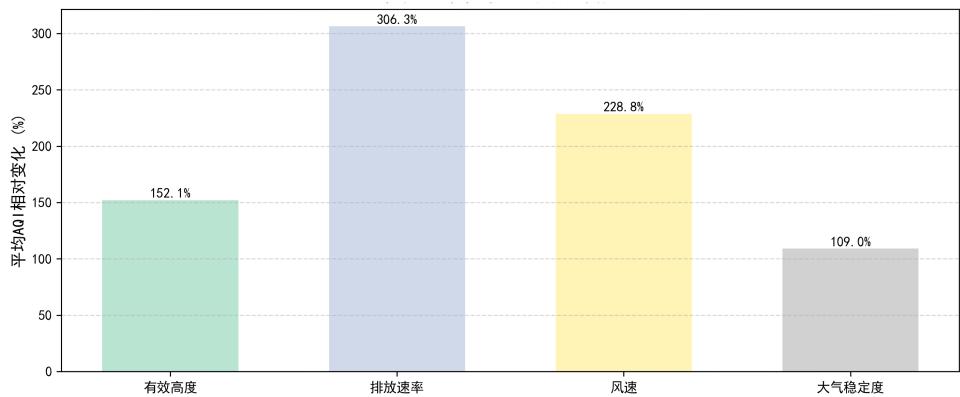


图 4 各参数对平均 AQI 的灵敏度分析

结果表明，在不同因素分别变化的情况下，CAQI 的整体评价结果会出现显著波动。其中有效高度对最大 AQI 相对变化影响最大，可达 441.4%，排放速率对平均 AQI 的影响最大，可达 306.3%。说明空气质量受污染源有效高度和排放速率影响较大。另外，污染源的有效高度与排放速率是影响空气质量波动的关键因素。

5.4.3 建议报告

根据我们构建的模型，对江浙沪地区部分城市进行了数据分析。结果显示，不同城市的主要空气污染来源具有显著差异，主要可分为颗粒物污染型（簇 2）与交通污染型（簇 3）。因此，我们提出以下分城市、分重点的治理方案。

- 针对颗粒物污染型城市

1. 工业源治理：
 

推进重点行业增效减排改造，严格控制工业燃煤、有色金属冶炼等排放源，不断推进污染物处理后排放的检查机制。
2. 扬尘控制：
 

实施精细化工地管理，增加对工地产生的颗粒物的沉降处理，如增加工厂周边水雾喷洒量。
3. 推动能源转型：
 

推广天然气、电力等清洁能源替代传统燃煤。
- 针对交通污染型城市
  1. 优化公共交通：
 

加大地铁和电动公交的投入力度，减少对私家车的依赖。
  2. 区域管控：
 

建立相关系统实时监测部分重点地区空气质量，对大流量交通进行调控。
- 针对关键灵敏参数
  1. 排放高度优化：
 

鼓励高污染行业提高排放点高度，合理设置烟囱高度以增强污染物扩散能力，减少低空空气污染。
  2. 放速率控制：
 

实施排放速率的上限管理，对于重点工厂的排放速率做出严格限制，推动企业减排。

## 六、模型的分析与检验

### 6.1 模型一的分析与检验

#### 6.1.1 一致性检验

- 计算一致性指标

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (27)$$

- 查随机一致性指标  $RI$ 。
- 要求  $CR = \frac{CI}{RI} < 0.1$ （否则调整判断矩阵）。

计算得  $CI = 0.038$ , 查表得  $RI(6) = 1.24$ , 则  $CR = 0.0306 < 0.1$  通过一致性检验。

#### 6.1.2 数据规模与覆盖范围

基于江浙沪地区 7 个城市的空气质量数据进行分析：

- 总数据量：2,554 条记录
- 数据时间跨度：约 1 年（每个城市 365-366 天）
- 覆盖城市：常州、合肥、杭州、上海、苏州、湖州、南京

### 6.1.3 CAQI 与传统 AQI 结果对比

将上述数据通过 CAQI 得出对应等级，与传统 AQI 结果比对，得出下列数表统计结果。基于上述分析与检验，我们认为 CAQI 模型具备一定的合理性，可以考虑将其与 AQI 搭配使用。

**表 19 CAQI 与传统 AQI 等级变化统计**

指标	数值	比例
总样本数	2,554	100%
等级变化次数	279	10.9%
首要污染物变化次数	181	7.1%

**表 20 等级变化类型统计**

等级变化类型	次数
良 → 轻度污染	82 次
轻度污染 → 良	113 次
中度污染 → 轻度污染	21 次
优 → 良	54 次
轻度污染 → 中度污染	7 次
重度污染 → 轻度污染	2 次

表 21 首要污染物分布对比

污染物	传统 AQI 次数	传统 AQI 比例	CAQI 次数	CAQI 比例
$O_3$ -8h	1,472	57.6%	1,511	59.2%
$PM_{2.5}$	602	23.6%	743	29.1%
$PM_{10}$	293	11.5%	181	7.1%
$NO_2$	187	7.3%	119	4.7%

## 6.2 模型二的分析与检验

### 6.2.1 PCA 模型有效性

- 各城市累计方差贡献率均超过 90%，模型解释能力良好。
- 载荷矩阵结构清晰，污染源识别结果符合实际情况。
- 主成分数量选择合理，避免了信息损失和过度拟合。

### 6.2.2 HCA 模型稳定性

- 采用平均连接法减少极值影响，提高聚类稳定性。
- 距离度量选择合适，标准化数据使用欧几里得距离有效。
- 聚类数确定方法科学，4 个簇的分类结果具有实际意义。

### 6.2.3 结果一致性检验

- PCA 识别的污染源类型与 HCA 聚类结果高度一致。
- 工业污染特征城市在两种方法中都得到突出体现。
- 交通污染城市聚为一类，符合区域污染特征预期。

## 6.3 模型三的分析与检验

### 6.3.1 污染物扩散对空气质量的影响

为了验证模型的有效性，我们模拟了在源强  $Q=6.0\text{g/s}$  风速为  $2.5\text{m/s}$ ，塔高为  $60\text{m}$ ，稳定性 D 级的单污染源影响下， $2\text{km}$  范围内空气质量的变化。可以看到，在污染源的作用下，污染最严重地区的 AQI 达到了 142，属于轻度污染；受到污染源影响的区域最远可延伸至下风向  $1\text{km}$  处。

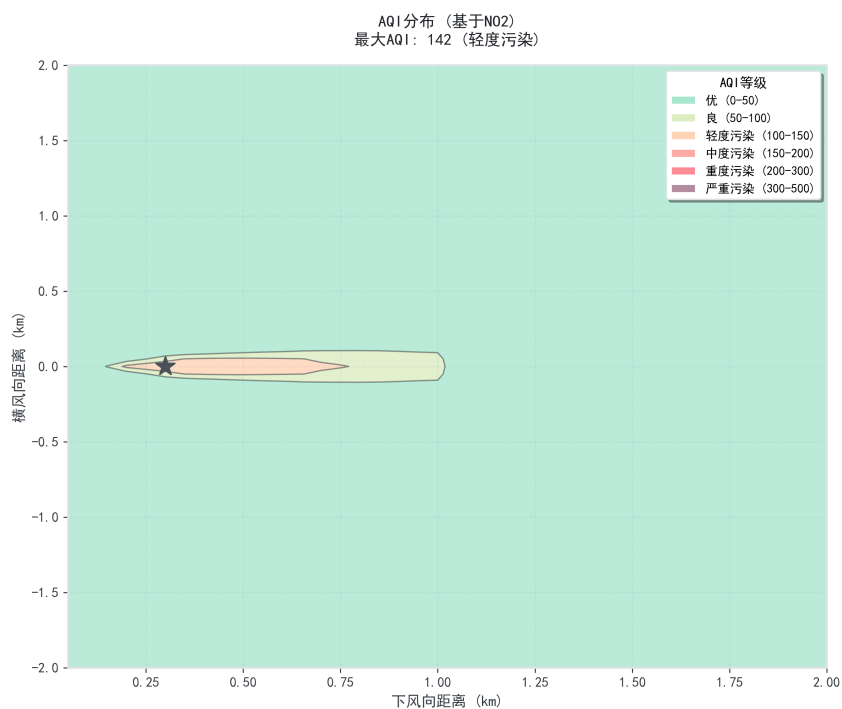


图 5 特定单一污染源作用下空气质量分布示意图

### 6.3.2 污染物扩散的空间特征分析

除了分析污染物扩散对空气质量的影响，我们还通过分析污染物浓度在空间上的分布特征来检验模型的合理性。我们生成了题目三所述条件下，中午 12 点（B 类稳定度）时地面污染物浓度的二维分布图，以直观展示污染物扩散的形态。

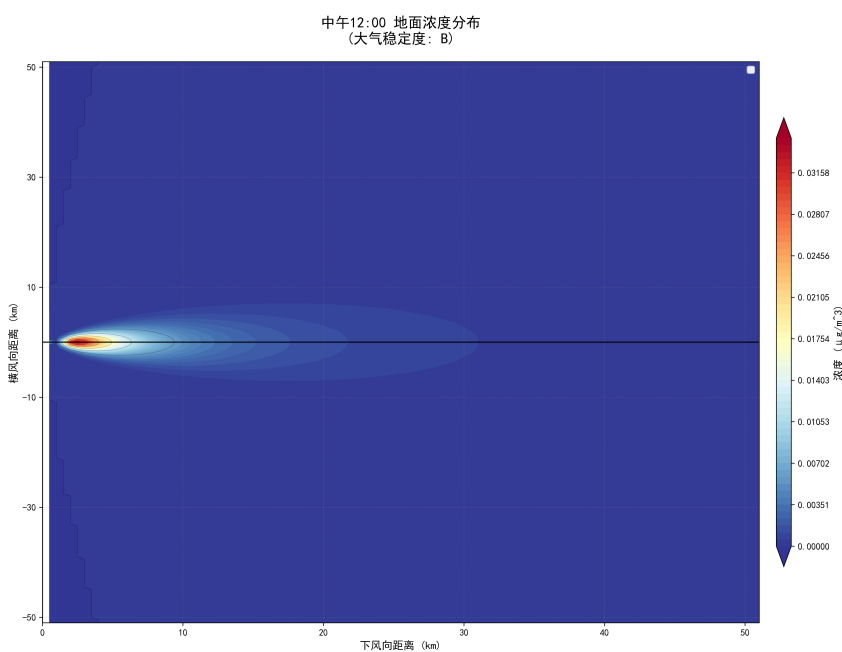


图 6 中午 12 点污染物浓度二维分布图

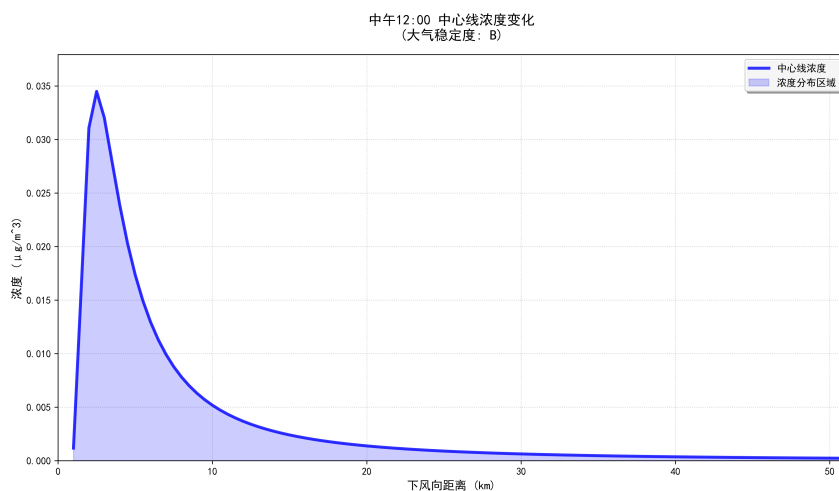


图7 中午12点污染物中心线浓度分布图

从图中可以清晰地观察到以下几点：

- **浓度沿主轴线分布：**

污染物的最高浓度始终出现在下风向的中心轴线上 ( $y = 0$ )，并在扩散距离足够大时随着与污染源距离 ( $x$ ) 的增加而逐渐降低。这符合污染物随风向下游输送和扩散的基本规律。

- **横向扩散特征：**

在垂直于风向的横向 ( $y$  轴方向)，污染物浓度以中心轴线为对称轴，向两侧迅速递减，呈现出经典的正态（高斯）分布形态。

- **影响范围的可视化：**

该图不仅展示了浓度的峰值，还清晰地勾勒出了污染物影响的主要区域。这种椭圆形的等浓度线分布是高斯烟羽模型的典型特征，为评估污染事件对周边环境的影响范围提供了直观依据。

## 七、模型的评价

### 7.1 模型的优点

#### 7.1.1 模型一

- **多指标综合权重化**

通过 AHP 引入专家判断，可根据江浙沪区域实际污染特征（如  $PM_{2.5}$  与  $O_3$  的季节差异）科学设定权重，相比现行 AQI 较单纯的  $\max(IAQI_i)$  更能反映多污染物的协同综合影响。

- **一致性检验保证可靠性**

模型分析与检验环节，判断矩阵需要通过  $CR < 0.1$  的一致性检验，这种检验方式客观

量化而又严格，充分确保权重分配的科学性与合理性。

- **可扩展性强,可调整性强**

专家比较矩阵与 CAQI 阈值区间均可结合新研究或政策修订，灵活调整，无需重构整体框架。

### 7.1.2 模型二

- **“点”、“面” 双层次分析**

既用 PCA 针对给定城市详细分析影响空气质量的污染源；同时又创新性地考虑江浙沪地区的整体特征，用 HCA 将江浙沪地区的情况划分为几个不同的类别。

- **降维与聚类相结合，通用性好**

收集数据首先用  $3\sigma$  原则进行清洗，之后用 PCA 有效提取主要污染源特征，减少数据冗余，准确识别城市主要污染源；城市划分中采用 PCA-HCA 联合，刻画区域相似性，直观划分“污染类型区”。

模型标准化流程清晰，适用于不同城市、不同季节、不同数据规模，可推广性强。

### 7.1.3 模型三

- **理论基础成熟,框架完善**

基于经典高斯烟羽公式，结合 *Pasquill-Gifford* 曲线，有牢固的理论基础，计算公式清晰，流程标准，易于实现。

- **多重修正提高精度,使模型更加完善可靠**

针对小风速、热力抬升、取样时间等因素引入公式修正，显著提升模型在复杂气象条件下的适用性。

- **可持续调整参数，进一步优化模型**

各种系数 ( $\gamma, \alpha, k, n_0, n_1, n_2$  等) 可依据现场监测数据或文献不断更新，从而优化模拟效果。

## 7.2 模型的缺点

### 7.2.1 模型一

- **计算工作量较大**

AHP 依赖专家经验，判断矩阵需要经过反复论证、多方研判之后才能确定，相比现行国标与美标更加复杂，工作量较大。

- **需重新定义模型阈值**

新定义的 CAQI 分级阈值尚无历史数据参考与学界研究支撑，可能需要一段时间的观测与论证才能确定阈值。



### 7.2.2 模型二

- **静态聚类无法反映动态演变**

HCA 分析阶段目前将城市污染物的静态均值输入模型，为静态聚类，无法反映不同时间城市聚类的动态演变。若要研究季节或年度趋势，需要多次重复分析或结合时序聚类等方法。

### 7.2.3 模型三

- **实际大气情况与理想化模型存在出入**

模型假设风速、风向、大气稳定度等参数不随时间变化且污染物排放连续稳定，实际情况往往动态多变，模型精度在强非稳态场景下降明显。

- **修正多采用半经验公式，需慎重确定数据**

小风、静风等条件下污染物扩散较为复杂，国内外尚未取得突破性研究；影响烟气抬升高度的因素过多、不确定性过大；这些情况的修正往往采用半经验公式，需要结合实际慎重确定数据。

## 八、模型的推广

本文模型首先创造性地提出了一种考虑各种污染物的影响，计算综合空气质量指数的方法。然后采用 PCA-HCA 联合模型构建分析特定城市和城市群大气污染源的框架，进而分析江浙沪地区的主要污染源。最后建立单污染源空气污染扩散模型，量化评判污染源对于空气质量的影响。三个模型层层深入，可利用数据分析总结影响空气质量的关键因素，在当今环境保护越来越受到重视的背景下具有应用价值。

本文模型框架清晰，未来可继续创新性改进、完善，如在模型一 CAQI 的计算中引入非线性项修正（不同污染物 IAQI 的乘积），考虑不同污染物的协同作用；在传统 PCA 基础上，引入动态时间因子模型，捕捉主要污染物随时间的变化特征；基于 GIS 与 Web 可视化技术，将模型成果以交互式地图形式展现等。

## 参考文献

- [1] 中华人民共和国环境保护部，国家质量监督检验检疫总局. GB 3095-2012 环境空气质量标准 [S]. 北京：中国标准出版社，2012：1-11.
- [2] 中华人民共和国环境保护部. HJ 633-2012 环境空气质量指数 (AQI) 技术规定 [S]. 北京：中国标准出版社，2012：1-11.

- [3] U.S. Environmental Protection Agency. Technical Assistance Document for the Reporting of Daily Air Quality - the Air Quality Index (AQI): EPA-454/B-24-002[R]. Research Triangle Park, NC: Office of Air Quality Planning and Standards, 2024.
- [4] SAATY T L. 层次分析法 [M]. 徐玖平, 罗利, 曹鸿翔, 译. 北京: 科学出版社, 2009.
- [5] 卢纹岱. 工业企业经济统计分析 [M]. 上海: 复旦大学出版社, 1988.
- [6] 何晓群. 多元统计分析 [M]. 6 版. 北京: 中国人民大学出版社, 2018.
- [7] EVERITT B S, LANDAU S, LEESE M, et al. Cluster Analysis[M]. 5th ed. Chichester: John Wiley & Sons, 2011.
- [8] 戴万瑞. 大气点源污染扩散模拟研究 [D]. 武汉: 武汉大学, 2018.
- [9] 孙志宽. 高斯烟羽扩散模型再研究 [J]. 环境与可持续发展, 2013(5): 107-109.
- [10] 中华人民共和国环境保护部. GB 3840-91 环境空气质量标准 [S]. 北京: 中国标准出版社, 1991.

## 附录 A CAQI 算法

```
# 1. IAQI分级标准 (用于 calculate_iaqi 函数)
standards = {
    'PM2.5': [[0,50,0,35], [51,100,35,75], [101,150,75,115],
              [151,200,115,150], [201,300,150,250], [301,500,250,500]],
    'PM10': [[0,50,0,50], [51,100,50,150], [101,150,150,250],
              [151,200,250,350], [201,300,350,420], [301,500,420,600]],
    'SO2': [[0,50,0,50], [51,100,50,150], [101,150,150,475],
             [151,200,475,800], [201,300,800,1600], [301,500,1600,2620]],
    'NO2': [[0,50,0,40], [51,100,40,80], [101,150,80,180],
             [151,200,180,280], [201,300,280,565], [301,500,565,940]],
    'CO': [[0,50,0,2], [51,100,2,4], [101,150,4,14],
            [151,200,14,24], [201,300,24,36], [301,500,36,60]],
    'O3_8h': [[0,50,0,100], [51,100,100,160], [101,150,160,215],
               [151,200,215,265], [201,300,265,800], [301,500,800,1000]]
}

# 2. CAQI等级阈值 (用于 get_level 函数)
caqi_levels = {
    (0, 50): "优", (51, 90): "良", (91, 140): "轻度污染",
    (141, 180): "中度污染", (181, 270): "重度污染", (271, float('inf')): "严重污染"
}

def calculate_iaqi(pollutant, concentration):
    if pollutant not in standards: return 0
    for aqi_lo, aqi_hi, bp_lo, bp_hi in standards[pollutant]:
        if bp_lo <= concentration <= bp_hi:
            return round((aqi_hi - aqi_lo) / (bp_hi - bp_lo) * (concentration - bp_lo) + aqi_lo)
    aqi_lo, aqi_hi, bp_lo, bp_hi = standards[pollutant][-1]
    return round((aqi_hi - aqi_lo) / (bp_hi - bp_lo) * (concentration - bp_lo) + aqi_lo)

def get_level(caqi):
    for (low, high), level_desc in caqi_levels.items():
        if low <= caqi <= high:
            return level_desc
    return "严重污染" # 默认最高等级

def calculate_caqi_simple(pollutant_data, ahp_weights):
    iaqi_dict = {p: calculate_iaqi(p, c) for p, c in pollutant_data.items()}
    above_50 = {p: iaqi for p, iaqi in iaqi_dict.items() if iaqi >= 50}
    if above_50:
        total_w = sum(ahp_weights.get(p, 0) for p in above_50.keys())
        if total_w == 0:
            caqi = max(iaqi_dict.values()) if iaqi_dict else 0
        else:
            norm_weights = {p: ahp_weights[p] / total_w for p in above_50.keys()}
            caqi = sum(iaqi_dict[p] * norm_weights[p] for p in above_50.keys())
    else:
        caqi = max(iaqi_dict.values()) if iaqi_dict else 0
```

```

    return round(caqi, 1), get_level(caqi)
# AHP权重
ahp_weights = {
    'SO2': 0.104, 'CO': 0.051, 'NO2': 0.152,
    'O3_8h': 0.271, 'PM2.5': 0.271, 'PM10': 0.152
}
pollutant_data_example = {
    'PM2.5': 35, 'PM10': 56, 'CO': 1.2,
    'SO2': 16, 'NO2': 12, 'O3_8h': 87
}
caqi_value, pollution_level = calculate_caqi_simple(pollutant_data_example, ahp_weights)
print(f"CAQI值: {caqi_value}")
print(f"污染等级: {pollution_level}")

```

## 附录 B 高斯烟羽模型

```

import numpy as np
g = 9.8
hs = 50
D = 2.5
Ts_C = 150
Ts = Ts_C + 273.15
scenario = {'stability': 'B', 'Q': 406.92 * 1200 / (1000 * 3600), 'u_ref': 2.13, 'Ta_C':
    18.84, 'Vs': 15, 'p': 0.15, 'time_str': '12pm'}
def get_sigma(stability, x):
    x_km = np.asarray(x, dtype=float) / 1000.0
    x_km[x_km == 0] = 1e-9 # Prevent division by zero later if x is 0

    # Combined sigma_y and sigma_z parameters into a single dictionary for conciseness
    sigma_params = {
        'A': {'y': [(0.425809, 0.961071), (0.602052, 0.850934)], 'z_cond': [x_km <= 0.3, (x_km >
            0.3) & (x_km <= 0.5), x_km > 0.5], 'z_func': [lambda v: 0.0799501 * v**1.12151,
            lambda v: 0.0851771 * v**1.51350, lambda v: 0.090211545 * v**2.16881]},
        'B': {'y': [(0.281846, 0.914370), (0.396353, 0.865014)], 'z_cond': [x_km <= 0.5, x_km >
            0.5], 'z_func': [lambda v: 0.127190 * v**0.961435, lambda v: 0.057025 *
            v**1.09356]},
        'C': {'y': [(0.177154, 0.924279), (0.232123, 0.885157)], 'z_cond': [x_km <= 0.5, (x_km >
            0.5) & (x_km <= 2), x_km > 2], 'z_func': [lambda v: 0.106803 * v**0.917595, lambda
            v: 0.126152 * v**0.838628, lambda v: 0.235667 * v**0.756410]},
        'D': {'y': [(0.110726, 0.929418), (0.146659, 0.888723)], 'z_cond': [x_km <= 1, (x_km >
            1) & (x_km <= 10), x_km > 10], 'z_func': [lambda v: 0.104634 * v**0.826212, lambda
            v: 0.400167 * v**0.632023, lambda v: 0.810763 * v**0.55536]},
        'E': {'y': [(0.0864901, 0.920818), (0.101947, 0.896864)], 'z_cond': [x_km <= 1, (x_km >
            1) & (x_km <= 10), x_km > 10], 'z_func': [lambda v: 0.0927529 * v**0.788370, lambda
            v: 0.433384 * v**0.565188, lambda v: 1.73241 * v**0.414743]},
    }

```

```

        'F': {'y': [(0.0553634, 0.929418), (0.0732345, 0.888723)], 'z_cond': [x_km <= 1, (x_km >
            1) & (x_km <= 10), x_km > 10], 'z_func': [lambda v: 0.0620765 * v**0.784400, lambda
            v: 0.370015 * v**0.525969, lambda v: 2.40691 * v**0.322659]},
    }
    if stability not in sigma_params:
        raise ValueError(f"Invalid stability class: {stability}")
    coeff_y1, exp_y1 = sigma_params[stability]['y'][0]
    coeff_y2, exp_y2 = sigma_params[stability]['y'][1]
    sigma_y = np.piecewise(x_km, [x_km <= 1], [lambda v: coeff_y1 * v**exp_y1, lambda v:
        coeff_y2 * v**exp_y2])
    params_z = sigma_params[stability]
    sigma_z = np.piecewise(x_km, params_z['z_cond'], params_z['z_func'])
    return sigma_y * 1000, sigma_z * 1000
def calculate_H_u_iterative(scene_params, h_s, g, D, T_s, epsilon=0.1, max_iter=10):
    u_ref, Ta, Vs, p, stability = scene_params['u_ref'], scene_params['Ta_C'] + 273.15,
        scene_params['Vs'], scene_params['p'], scene_params['stability']
    F = g * Vs * (D / 2)**2 * (1 - Ta / T_s)
    H = float(h_s) # Ensure H is a float for iterative calculation
    for _ in range(max_iter):
        u_k = u_ref * (H / 10)**p
        # Ternary operator to condense the delta_h_k calculation
        delta_h_k = (38.71 * F**(3/5) / u_k) if stability in ['A', 'B', 'C', 'D'] else (2.6 * (F
            / (u_k * (0.02 if stability == 'E' else 0.035)))**(1/3))
        H_k = h_s + delta_h_k
        if abs(H_k - H) < epsilon:
            return H_k, u_k # Return directly when converged
        H, u = H_k, u_k # Update H and u for the next iteration
    return H, u
def gaussian_plume_model(Q, u, H, x, y, z, stability):
    sigma_y, sigma_z = get_sigma(stability, x)
    sigma_y = np.maximum(sigma_y, 1e-6)
    sigma_z = np.maximum(sigma_z, 1e-6)
    u = np.maximum(u, 1e-6)
    term1 = Q / (2 * np.pi * u * sigma_y * sigma_z)
    term_exp_y = np.exp(-y**2 / (2 * sigma_y**2))
    term_exp_z = np.exp(-(z - H)**2 / (2 * sigma_z**2)) + np.exp(-(z + H)**2 / (2 * sigma_z**2))
    concentration = term1 * term_exp_y * term_exp_z
    return concentration * 1e6 # Convert units to g/m³
H_final, u_final = calculate_H_u_iterative(scenario, hs, g, D, Ts)
target_x, target_y, target_z = 5000, 0, 0 #下风向5km处, 中心线, 地面
concentration = gaussian_plume_model(scenario['Q'], u_final, H_final, target_x, target_y,
    target_z, scenario['stability'])
print(f"有效烟囱高度 H_final: {H_final:.2f} m")
print(f"最终风速 u_final: {u_final:.2f} m/s")
print(f"目标点 ({target_x}m, {target_y}m, {target_z}m) 处的污染物浓度: {concentration:.4f}
    g/m³")

```

## 附录 C 层次聚类分析

```
import json, numpy as np, pandas as pd
from sklearn.preprocessing import StandardScaler
from scipy.cluster.hierarchy import linkage, fcluster
from scipy.spatial.distance import pdist
import warnings
warnings.filterwarnings('ignore')

with open('air_quality_data.json', 'r', encoding='utf-8') as f:
    data = json.load(f)

X, names, pollutants = [], [], ['PM2.5', 'PM10', 'SO2', 'NO2', 'CO', 'O3_8h']
for city, rec in data.items():
    if not rec: continue
    df = pd.DataFrame(rec)
    cols = [c for c in pollutants if c in df.columns]
    if len(cols) < 3: continue
    m = df[cols].mean()
    X.append(m.values)
    names.append(city)
used_cols = cols

X = StandardScaler().fit_transform(X)
Z = linkage(pdist(X), method='average')
labels = fcluster(Z, 4, criterion='maxclust')

for k in np.unique(labels):
    idx = np.where(labels == k)[0]
    avg = X[idx].mean(0)
    pol = used_cols[np.argmax(np.abs(avg))]
    if pol in ['PM2.5', 'PM10'] and avg.max() > 0: t = '颗粒物污染'
    elif pol in ['NO2', 'CO'] and avg.max() > 0: t = '交通污染'
    elif pol == 'SO2' and avg.max() > 0: t = '工业污染'
    elif pol == 'O3_8h' and avg.max() > 0: t = '光化学污染'
    else: t = f'{pol}较低'
    print(f"簇{k}: 城市: {' '.join([names[i] for i in idx])} | 特征: {t}")

print("\n最终聚类标签:")
for c, l in zip(names, labels):
    print(f"{c}: 簇{l}")
```