



中国研究生创新实践系列大赛
“华为杯”第十八届中国研究生
数学建模竞赛

学 校	西北工业大学
参赛队号	21106990088
队员姓名	1. 钟瑞国
	2. 古淳月
	3. 魏子彦

中国研究生创新实践系列大赛

“华为杯”第十八届中国研究生

数学建模竞赛

题 目

空气质量预报二次建模

摘

要：

随着社会的发展和公众环境意识的提高，空气污染越来越受到公众关注。准确的空气质量预测可以为政府决策和居民活动提供有用的信息。本文针对现有空气质量预测模型准确率低等问题，在 WRF-CMAQ 模型的基础上，结合实测数据对一次预报结果进行二次建模分析，提高了预测准确度。

对于问题一，分析了附件 1 中逐小时及逐日的各污染物实测数据误差，计算了监测点 A 从 2020 年 8 月 25 日到 8 月 28 日每天实测的各污染物 AQI 值，给出了该时间段内的空气质量指数（AQI）及首要污染物（结果见表 3.4）。

对于问题二，在污染物排放情况不变的前提下，根据对污染物浓度的影响对气象条件进行分类，并阐述各类气象条件的特征。为了对多变量数据进行高效利用，首先对数据进行挖掘分析，对相关指标进行融合，在尽可能多地减小原指标信息损失的基础上，实现了指标的优化。本文采用了主成分分析（PCA）法以及核主成分分析（KPCA）法对线性及非线性数据进行降维处理，计算出各个气象条件的相关性，得到大气压、风速、总温度、总湿度、降雨量五个综合气象条件指标及其特征信息；分析了气象条件对单一污染物浓度的影响，给出了气象条件对各个污染物浓度影响的排序及相关性（结果见表 4.1）。特别地，针对风速、总温度、总湿度对臭氧 O_3 浓度的影响展开了深入讨论；分析了综合气象条件的相互组合对多个污染物浓度的影响，给出了图 4.4 的气象条件分类结果，针对综合气象条件的典型天气做了说明。

对于问题三，基于 WRF-CMAQ 一次预报模型数据及实测数据，要求建立适用于监测点 A、B、C 的二次预报模型，并且 AQI 预报值的最大相对误差尽量小。本文首先引入了专家经验对数据进行处理分析，对异常值进行了剔除以及缺失值的补充，提取数据的特征信息。然后，采用了基于 Takagi-Sugeno 的模糊宽度学习模型（Fuzzy-BLS），构建了基于 Fuzzy-BLS 的二次预报模型，通过模型求解给出了监测点 A、B、C 在 2021 年 7 月 13 日至 15 日各个污染物的浓度及 AQI 的初步预测结果。从训练集及测试集两部分对三个监测点的 AQI 计算结果进行了检验及误差计算，并进行分析，发现模型对于 O_3 与 NO_2 污染物浓度的预测存在局限性。因此建立了基于种间竞争（Lotka-Volterra）的 Fuzzy-BLS 二次预报模型（LoVo-FBLS 模型），引入了“种间竞争”概念，通过具有时变参数的微分方程对 O_3 与 NO_2 浓度进行数据修正，种间竞争模型对 O_3 与 NO_2 预测近乎与实际数据曲线相重合，优化了 Fuzzy-BLS 的二次预报模型，经修正后， O_3 浓度的相对误差为 0.0379， NO_2 浓度的相对误差为 0.0227。

计算得到了 2021 年 7 月 13 日至 15 日这三天各个监测点的 AQI 及首要污染物，其中

A 点 AQI (首要污染物) 为 45 (无)、100 (O_3)、54 (O_3)，B 点为 44 (无)、42 (无)、30 (无)，C 点为 80 (O_3)、81 (O_3)、66 (O_3)，详细结果见表 5.4、表 5.5、表 5.6。

对于问题四，针对监测点 A 及临近区域内的监测点 A1、A2、A3，建立了区域协同预报模型。在前述 LoVo-FBLS 模型的基础上，引入空气流动模型对邻近区域进行协同二次预报，进一步提升 AQI 预测的准确度。首先在对原始数据中的风向进行坐标系转换，从平均风向及风向频率两个角度对风向进行分析，引入污染系数，对风向做加权平均处理，将监测点简化为污染的连续面源，采用虚拟点光源法的面源扩散模式，得到下风口的污染物浓度，给出污染物从扩散城市到受影响城市的浓度计算方法。通过求解模型，给出了污染物从某一城市扩散到临近城市的三维趋势图，得到了 2021 年 7 月 13 日至 15 日监测点 A 的 AQI (首要污染物) 为 48 (无)、102 (O_3)、55 (O_3)，详细结果见表 6.2。比对问题三、问题四模型计算出的监测点 A 的 AQI 以及首要污染物，并对两个模型求解结果进行误差对比分析，发现问题四模型明显减小了问题三模型求解结果的误差。

关键词：空气质量预报，AQI，模糊宽度学习，协同竞争，LoVo-FBLS 模型，空气流动模型

目录

一、问题背景与重述	4
1.1 问题背景	4
1.2 问题重述	4
二、模型假设及符号说明	5
2.1 模型假设	5
2.2 符号说明	5
三、问题一的求解	6
3.1 问题分析与计算过程	6
3.2 问题求解结果	7
四、问题二的建模与求解	8
4.1 问题分析	8
4.2 数据处理	8
4.2.1 PCA 分析	8
4.2.2 KPCA 分析	10
4.2.3 综合气象条件指标	12
4.3 气象条件分析	12
4.3.1. 对单一污染物浓度影响	12
4.3.2. 整体性分析	15
五、问题三的建模与求解	17
5.1 问题分析	17
5.2 基于 Fuzzy-BLS 的二次预报模型	17
5.3 基于种间竞争的 Fuzzy-BLS 二次预报模型	21
六、问题四的建模与求解	26
6.1 问题分析	26
6.2 基于空气流动的污染物扩散模型	26
6.3 模型求解及误差对比分析	30
6.3.1 模型求解	30
6.3.2 误差对比分析	33
七、模型评价与改进	34
7.1 模型的优点	34
7.2 模型的缺点	35
参考文献	36
附录	37

一、问题背景与重述

1.1 问题背景

城市化、工业化的快速发展以及人类活动的加剧使得环境问题日益突出，空气污染已成为世界上所有发达国家和发展中国家急需面临的重要问题。一般来说，空气污染是指空气中的固体颗粒(如 $\text{PM}_{2.5}$ 、 PM_{10})和特殊气体(如 SO_2 、 NO_2 、 O_3 和 CO)这些大气污染物存在的量、性质和时间对人类的健康、生活造成危害的现象。研究^[1]指出空气污染与人类各种疾病密切相关。长期暴露在污染的空气中，会损害人的心脏和肺的正常功能，导致严重的慢性疾病。此外，空气污染会导致极端天气，阻碍交通，扰乱日常生活，直接关系到人类的生活质量和健康安全。

为了保护人们免受环境空气污染，近年来，许多国家和非政府组织建立了空气质量监测站获取空气质量指数(Air Quality Index, AQI)。空气质量指数^[2]表示空气污染的程度，它受空气中各种污染物浓度的影响。影响空气质量的因素主要来自人为污染物的排放，包括机动车尾气、工厂垃圾、住宅供暖、垃圾焚烧等。准确的空气质量预测可以为政府决策和居民活动提供有用的信息。通过预测未来的空气质量状况及变化趋势^[3]，可以为环境监测部门合理控制、管理和有效预防大气污染提供科学的决策依据，有效减少空气污染造成的损害。

1.2 问题重述

由于模拟气象场、排放清单以及臭氧生成机理的不确定性，目前已有的 WRF-CMAQ 模型对空气质量预报结果不理想，因此需要在 WRF-CMAQ 等一次预报模型模拟结果的基础上，结合更多的数据源进行二次建模，提高预报的准确性。

本题给出了多个监测点逐小时、逐日污染物浓度和气象的实测数据及 WRF-CMAQ 模型运行产生的预报数据，基于这些数据解决以下问题：

问题一：针对附件 1 的数据，通过题目所给方法求解监测点 A 从 2020 年 8 月 25 日至 28 日每天实测的 AQI 和首要污染物，并给出计算结果表。

问题二：针对附件 1 的数据，根据气象条件对污染物浓度的影响对气象条件进行分类，并阐述各类气象条件的特征。

问题三：针对附件 1、2 的数据，建立适用于 A、B、C 三个监测点的二次预报数学模型，预测未来 3 天 6 种常规污染物的单日浓度值，结果需满足 AQI 预报值最大相对误差尽可能小、首要污染物准确度尽可能高。通过模型求解给出这三个监测点在指定日期的 6 种常规污染物的单日浓度值，计算 AQI 和首要污染物并给出计算结果表。

问题四：针对附件 1、3 的数据，建立 A、A1、A2、A3 四个监测点的区域协同预报模型，结果需满足 AQI 预报值最大相对误差尽可能小、首要污染物准确度尽可能高。通过模型求解给出这四个监测点在指定日期的 6 种常规污染物的单日浓度值，计算 AQI 和首要污染物并给出计算结果表。与问题三的模型进行对比，分析协同预报模型对监测点 A 的污染物浓度预报准确度提升情况及原因。

二、模型假设及符号说明

2.1 模型假设

- (1) 假设题目所给出的原始数据是真实有效的，相关异常值的剔除或缺失值的补充不会对结果产生重大影响；
- (2) 假设各监测点无极端天气的出现，自然灾害出现的罕见气象条件对预测污染物浓度的影响较小；
- (3) 假设 A、B、C 三个监测点间的污染物排放情况不相互影响，污染物浓度仅与自身气象条件有关；
- (4) 由于题目未给出具体地点，假设臭氧 O₃ 的浓度变化仅与气象条件有关，地形地貌、该地防治政策不对臭氧预测产生大的影响。

2.2 符号说明

本文涉及的符号由表 2.1 给出，未列出的符号均以出现处为准

表 2.1 符号说明	
符号	符号说明
IAQI	空气质量分指数
AQI	空气质量指数
\tilde{x}_i	标准化指标变量
Φ	非线性映射函数
p	权重向量
λ	特征值
K	核矩阵
p_s	线性相关系数
L	种群个体数
D_i^R, D_j^R	i, j 城市污染物 R 浓度
Df_{ij}^R	受影响城市污染物浓度
C	下风口地面浓度

三、问题一的求解

3.1 问题分析与计算过程

问题一需要计算监测点 A 从 2020 年 8 月 25 日到 8 月 28 日每天实测的 AQI 和首要污染物，并给出 AQI 计算结果表。AQI 的计算过程如下：

(1) 首先求解污染物 (SO_2 、 NO_2 、 CO 、 $\text{PM}_{2.5}$ 、 PM_{10}) 的 24 小时平均，计算臭氧 (O_3) 的最大 8 小时滑动平均，臭氧计算公式如下：

$$C_{\text{O}_3} = \max_{t=8,9,\dots,24} \left\{ \frac{1}{8} \sum_{i=t-7}^t c_i \right\}$$

其中 c_t 为臭氧在某日 $t-1$ 时至 t 时的平均污染物浓度。

(2) 计算各污染物空气质量分指数 (IAQI)，计算公式如下：

$$\text{IAQI}_P = \frac{\text{IAQI}_{\text{Hi}} - \text{IAQI}_{\text{Lo}}}{\text{BP}_{\text{Hi}} - \text{BP}_{\text{Lo}}} \cdot (C_P - \text{BP}_{\text{Lo}}) + \text{IAQI}_{\text{Lo}}$$

其中， IAQI_P 表示污染物 P 的空气质量分指数，结果进位取整数； C_P 为污染物 P 的质量浓度值； BP_{Hi} ， BP_{Lo} 分别是与 C_P 相近的污染物浓度限值的上位值与低位值； IAQI_{Hi} ， IAQI_{Lo} 分别是 BP_{Hi} ， BP_{Lo} 对应的空气质量分指数。

各项污染物项目浓度限值及对应的空气质量分指数级别需查表 3.1。

表 3.1 空气质量分指数 (IAQI) 及对应的污染物项目浓度限值

指数或污染物项目	空气质量分指数及对应污染物浓度限值									单位
空气质量分指数（IAQI）	0	50	100	150	200	300	400	500	-	
一氧化碳（CO）24小时平均	0	2	4	14	24	36	48	60	mg/m ³	
二氧化硫（SO ₂ ）24小时平均	0	50	150	475	800	1600	2100	2620		
二氧化氮（NO ₂ ）24小时平均	0	40	80	180	280	565	750	940		
臭氧（O ₃ ）最大8小时滑动平均	0	100	160	215	265	800	-	-	μg/m ³	
PM ₁₀ 24小时平均	0	50	150	250	350	420	500	600		
PM _{2.5} 24小时平均	0	35	75	115	150	250	350	500		

注：(a) 臭氧 (O_3) 最大 8 小时滑动平均浓度值高于 $800 \mu\text{g}/\text{m}^3$ 的，不再进行其空气质量分指数计算。

(b) 其余污染物浓度高于 $\text{IAQI}=500$ 对应限值时，不再进行其空气质量分指数计算。

(3) 空气质量指数 (AQI) 取各分指数中的最大值，即：

$$AQI = \max \{IAQI_1, IAQI_2, IAQI_3, \dots, IAQI_n\}$$

本题中，对于 AQI 的计算仅涉及 SO_2 、 NO_2 、 CO 、 $PM_{2.5}$ 、 PM_{10} 、 O_3 这六种污染物，其公式如下：

$$AQI = \max \{IAQI_{SO_2}, IAQI_{NO_2}, IAQI_{PM_{10}}, IAQI_{PM_{2.5}}, IAQI_{O_3}, IAQI_{CO}\}$$

空气质量等级范围根据 AQI 数值划分，AQI 取值对应的空气质量等级及首要污染物需查表 3.2。

表 3.2 AQI 取值对应的空气质量等级及首要污染物

空气质量指数 (AQI) 范围	空气质量等级	首要污染物
[0, 50]	优	无
[51, 100]	良	
[101, 150]	轻度污染	IAQI 最大的污染物
[151, 200]	中度污染	(IAQI > 100 的污染物为超标污染物)
[201, 300]	重度污染	
[301, +∞]	严重污染	

3.2 问题求解结果

根据附件 1 中监测点 A 逐小时污染物浓度与气象实测数据求解 SO_2 、 NO_2 、 CO 、 $PM_{2.5}$ 、 PM_{10} 的 24 小时平均以及 O_3 的最大八小时滑动平均监测浓度，对比逐小时与逐日污染物浓度与气象实测数据计算结果，核对数据后得到这 4 天监测点 A 的污染物数值如表 3.3：

表 3.3 监测点 A 各污染物 24 小时/最大 8 小时滑动平均浓度

日期	SO_2	NO_2	PM_{10}	$PM_{2.5}$	O_3	CO
	(单位: $\mu g/m^3$)					(单位: mg/m^3)
2020/8/25	8	12	27	11	112	0.5
2020/8/26	7	16	24	10	92	0.5
2020/8/27	7	31	37	23	169	0.6
2020/8/28	8	30	47	33	201	0.7

通过对表 3.3 中各污染物的求解以及查表 3.1、表 3.2 可得监测点 A 在 2020 年 8 月 25 日至 28 日这 4 天各个污染物的 IAQI，进而得到 AQI 值，其中 IAQI 最大的污染物即为首要污染物，求解结果如表 3.4 所示。

表 3.4 监测点 A 的 AQI 及首要污染物

日期	IAQI						AQI	空气质量	首要污染物
	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃	CO			
2020/8/25	8	15	27	16	60	13	60	良	O ₃
2020/8/26	7	20	24	15	46	13	46	优	无
2020/8/27	7	39	37	33	109	15	109	轻度污染	O ₃ (超标污染物)
2020/8/28	8	38	47	48	138	18	138	轻度污染	O ₃ (超标污染物)

四、问题二的建模与求解

4.1 问题分析

问题二基于题目所给的附件 1 原始数据，需要根据气象条件对污染物浓度的影响对气象条件进行分类，并阐述各类气象条件的特征。本文首先对原始数据进行处理，采用主成分分析（PCA）法以及核主成分分析（KPCA）法对数据进行降维处理，并计算各个气象条件的相关性，得到综合气象条件及其特征信息；从单一、整体两个角度对气象条件对单一污染物浓度影响、综合气象条件的相互组合对多个污染物浓度的影响进行了分析，特别地，针对气象条件对臭氧 O₃ 浓度的影响展开了深入讨论。

4.2 数据处理

本题提供了多变量大数据集，但多个变量使得气象条件的分类变得复杂，同时，温度、地表温度等变量之间具有相关性，孤立的分析各变量或盲目减少指标会造成有效信息的丢失，从而降低结果的准确度。为达到对多变量大数据集进行全面分析的目的，对相关指标进行融合、减少需要分析的指标、同时尽可能减少原指标包含信息的损失尤为重要。

4.2.1 PCA 分析

PCA(Principal Component Analysis)，即主成分分析方法，是一种使用最广泛的数据降维算法^[4]。PCA 的主要思想是将 n 维特征映射到 k 维上，这 k 维是全新的正交特征也被称为主成分，是在原有 n 维特征的基础上重新构造出来的 k 维特征。其基本思想是将原先众多具有一定相关性的指标，更新组合为一组新的互相独立的综合指标，并代替原先的指标。步骤如下：

(1) 对原始数据进行标准化处理

假设进行主成分分析的指标变量有 m 个： x_1, x_2, \dots, x_m ，共有 n 个评价对象，第 i 个评价对象的第 j 个指标的取值为 x_{ij} 。将各指标值 x_{ij} 转换成标准化指标 \tilde{x}_{ij} ，

$$\tilde{x}_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

其中， \bar{x}_j 为第 j 个指标的样本均值：

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

s_j 为第 j 个指标的样本标准差：

$$s_j = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}, (j=1, 2, \dots, m)$$

对应地， \tilde{x}_i 为标准化指标变量：

$$\tilde{x}_i = \frac{x_i - \bar{x}_i}{s_i}, (i=1, 2, \dots, m)$$

(2) 计算相关系数矩阵 R_e

相关系数矩阵 $R_e = (r_{ij})_{m \times m}$

$$r_{ij} = \frac{\sum_{k=1}^n \tilde{x}_{ki} \cdot \tilde{x}_{kj}}{n-1}, (i, j=1, 2, \dots, m)$$

式中 $r_{ij}=1$, $r_{ij} = r_{ji}$, 是第 i 个指标与第 j 个指标的相关系数

(3) 计算特征值和特征向量

特征值由大到小排序

计算相关系数矩阵 R_e 的特征值 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m > 0$ 及对应的特征向量 u_1, u_2, \dots, u_m , 其中 $u_j = (u_{1j}, u_{2j}, \dots, u_{mj})^T$, 由特征向量组成 m 个新的指标变量

$$\begin{cases} y_1 = u_{11}\tilde{x}_1 + u_{21}\tilde{x}_2 + \dots + u_{m1}\tilde{x}_m \\ y_2 = u_{12}\tilde{x}_1 + u_{22}\tilde{x}_2 + \dots + u_{m2}\tilde{x}_m \\ \dots\dots\dots \\ y_m = u_{1m}\tilde{x}_1 + u_{2m}\tilde{x}_2 + \dots + u_{mm}\tilde{x}_m \end{cases}$$

式中 y_1 是第 1 主成分, y_2 是第 2 主成分, \dots , y_m 是第 m 主成分。

(4) 选择 p ($p \leq m$) 个主成分, 计算综合评价

(a) 计算特征值 λ_j ($j=1, 2, \dots, m$) 的信息贡献率和累积贡献率。

主成分的信息贡献率 b_j 如下：

$$b_j = \frac{\lambda_j}{\sum_{k=1}^m \lambda_k} (j=1, 2, \dots, m)$$

主成分的累积贡献率 α_p 如下：

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_k}$$

当 α_p 接近于 1 ($\alpha_p = 0.85, 0.90, 0.95$) 时, 则选择前 p 个指标变量 y_1, y_2, \dots, y_p 作为 p 个主成分, 代替原来 m 个指标变量, 从而可对 p 个主成分进行综合分析。

(b) 计算综合得分 Z

$$Z = \sum_{j=1}^p b_j y_j$$

其中 b_j 为第 j 个主成分的信息贡献率，根据综合得分值可进行评价。

4.2.2 KPCA 分析

PCA 适用于数据的线性降维，而核主成分分析(Kernel PCA, KPCA)可实现数据的非线性降维，用于处理线性不可分的数据集。PCA 是线性的，其对于非线性数据往往显得无能为力，KPCA 是对 PCA 算法的非线性扩展，KPCA^[5]能够挖掘到数据集中蕴含的非线性信息。

KPCA 为了更好地处理非线性数据，引入非线性映射函数 Φ ，将原空间的数据映射到高维空间，其基本思想如下：

(1) 求解特征空间的特征值

用 X 的每一列来表示一个样本，即

$$X = [x_1, x_2, \dots, x_N]$$

其中，每个样本点 x_i 为 M 维列向量， X 中共 N 个样本。 $M \times N$ 维矩阵 X 所在的空间称为输入空间(Input space)。

用一个非线性映射 Φ 将 X 中的向量 x 映射到高维空间(记为 D 维)：

$$\Phi(x): R^M \rightarrow R^D, D \gg M$$

高维空间称为特征空间(Feature space)，记为 F 。将矩阵 X 中所有样本都映射到特征空间 F 中，得到一个 $D \times N$ 的新矩阵 $\Phi(X)$ ：

$$\Phi(X) = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)]$$

对矩阵 $\Phi(X)$ 进行中心化处理后得到：

$$\sum_{i=1}^N \Phi(x_i) = 0$$

特征空间 F 中的协方差矩阵为：

$$C_F = \frac{1}{N} \Phi(X) [\Phi(X)]^T = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \Phi(x_i)^T$$

其中 C_F 是 $D \times D$ 的矩阵。

求解协方差矩阵的特征值问题：

$$\Phi(X) [\Phi(X)]^T p = \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \Phi(x_i)^T p = \lambda p$$

其中 p 为特征空间中的权重向量，即主成分的方向。

(2) 通过核函数计算矩阵 K

当 λ 不为 0 时，对应的特征向量 p 可以表示为所有 $\Phi(x_i)$ 的线性组合，即：

$$p = \sum_{i=1}^N \alpha_i \Phi(x_i) = \Phi(X) \alpha$$

其中 N 维列向量 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$ 。

综合上式推导得：

$$[\Phi(X)]^T \Phi(X) [\Phi(X)]^T \Phi(X) \alpha = \lambda [\Phi(X)]^T \Phi(X) \alpha$$

定义矩阵 $K = [\Phi(X)]^T \Phi(X)$ ，则 K 为 $N \times N$ 的对称半正定矩阵，其 i 行 j 列的元素为 $K_{ij} = [\Phi(x_i)]^T \Phi(x_j)$ 。则求解以下特征值问题的非零特征值：

$$K\alpha = \lambda\alpha$$

因为矩阵 K 中的元素可由 K_{ij} 计算得到，不需要显式地定义映射 $\Phi(x)$ ，而只需定义特征空间中向量的点积即可，这就是核技巧。

定义函数 $k(x, y) = \Phi(x)^T \Phi(y)$ ，则 $k(x, y)$ 称为核函数。通过引入核函数，可以用输入空间中的低维向量 x 和 y 来计算其映射到特征空间 F 中后，高维向量 $\Phi(x)$ 和 $\Phi(y)$ 的点积。

通过核函数计算矩阵 K ，然后计算其较大的特征值对应的特征向量 α ，进而得到相应的权重向量 p 。常用核函数有：

(a) 多项式核：

$$k(x, y) = (x^T y + c)^d$$

(b) 高斯核/径向基函数核：

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) = \exp\left(-\gamma\|x - y\|^2\right)$$

(c) Sigmoid 核：

$$k(x, y) = \tanh(ax^T y + r)$$

(3) p 的归一化及 $\Phi(x)$ 的中心化

特征空间的主成分为，需要对 p 进行归一化。将 $\Phi(x_i)$ 向第 a 个坐标轴 p_a 投影可由下式计算：

$$t_{ai} = \Phi(x_i)^T p_a = \sum_{j=1}^N \alpha_{aj} K(x_i, x_j) = \sum_{j=1}^N \alpha_{aj} K_{ij}$$

将整个 $\Phi(x)$ 向 p_a 投影：

$$t_a = \Phi(x)^T p_a = [\Phi(X)]^T \Phi(X) \alpha_a = K \alpha_a = \lambda \alpha_a$$

矩阵 $\Phi(x)$ 中所有向量中心化得到新的矩阵为：

$$\tilde{\Phi}(X) = \Phi(X) - \Phi(X) \mathbf{1}_N$$

其中 $\mathbf{1}_N = \frac{1}{N} \mathbf{1}_{N \times 1} [\mathbf{1}_{N \times 1}]^T$ ，表示一个 $N \times N$ 的矩阵，其每个元素都为 $\frac{1}{N}$ 。

中心化后的核矩阵为：

$$\tilde{K} = K - K \cdot \mathbf{1}_N - \mathbf{1}_N \cdot K + \mathbf{1}_N \cdot K \cdot \mathbf{1}_N$$

由原数据 X 计算出 K 后，再根据上式得到中心化后的核矩阵。

4.2.3 综合气象条件指标

根据上述主成分分析（PCA）法及核主成分分析（KPCA）法的模型，对原始的多变量大数据集进行处理降维，并对气象条件及各个污染物浓度、AQI 等相关指标进行融合、减少需要分析的指标，有利于数据处理更为准确、指标分析更为全面。

基于气象学和气候学相关知识^[6]，本题将接近地 2 米温度、地表温度、感热通量、潜热通量、长波辐射、短波辐射、地面太阳能辐射、边界层高度归为“总温度”下的分支，比湿、湿度归为“总湿度”下的分支，采用 PCA 以及 KPCA 算法计算并选取最优解，最终得到总温度的相关特征信息（由 PCA 得到，为 PCA 第一分量，包含原有 90% 信息特征）以及总湿度的相关特征信息（由 KPCA 线性核得到，为 KPCA 第一分量，包含原有 92% 信息特征）。

综上，我们将多天气变量转化为大气压、风速、总温度、总湿度、降雨量五个综合气象条件指标及其特征信息。

4.3 气象条件分析

在污染源排放情况不变的情况下，气象条件是影响空气质量的主要原因，因此对气象条件进行分析有利于进一步研究空气质量。根据上文确定了五项综合气象条件指标及各指标的特征信息，本节展开分析气象条件对污染物浓度、AQI 的影响。

4.3.1. 对单一污染物浓度影响

采用斯皮尔曼（Spearman）相关系数对上文求解得到的五个综合气象条件（大气压、风速、总温度、总湿度、降雨量）和 6 个污染物浓度采用斯皮尔曼（Spearman）相关系数进行相关性计算。这里仅计算综合气象条件对单一污染物浓度的相关性，多个气象条件对多个污染物浓度的影响在下文详细分析。

Spearman 秩相关系数通常被认为是排列后的变量之间的 Pearson 线性相关系数 p_s ，在实际计算中，有更简单计算 p_s 的方法。假设原始的两组数据 x_i, y_i 已经按从大到小的顺序排列，记 x'_i, y'_i 为原在排列后数据的所在位置，则 x'_i, y'_i 称为变量的 x_i, y_i 秩次，则 $d_i = x'_i - y'_i$ 为 x_i, y_i 的秩次之差。

如果没有相同的秩次，则 p_s 可由下式计算：

$$p_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

其中， n 为观察值的数量。

如果有相同的秩次，那么就需要计算秩次之间的 p_s ：

$$p_s = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$

题目中指出“污染物排放情况不变”，污染物排放情况主要分为当地排放及外来污染物侵入两个方面。而风向影响着外来污染物的侵入及自身污染物的消散（大气污染状况与风向风速的关系，董志根），会使污染物排放情况出现波动，因此本节在气象条件中不考虑风向因素，但第三问及第四问会对其进行分析。

以总温度、总湿度、风速、大气压、AQI、6 中污染物分别的 IAQI 值共 11 项作为分

析变量，根据 Spearman 相关系数的计算可得这 11 个指标的相关性如图 4.1。

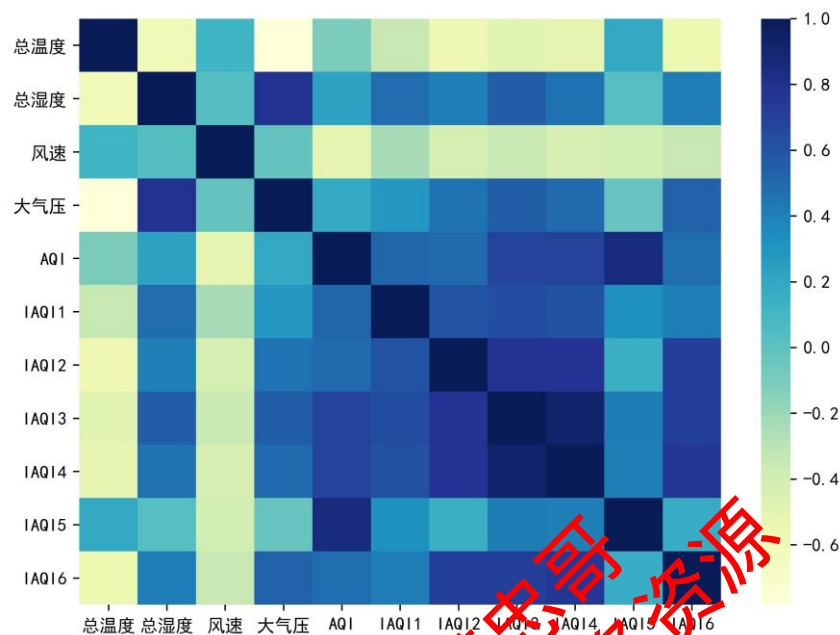


图 4.1 各指标相关性分析

图 4.1 中，横纵坐标为各个指标，其中 IAQI1、IAQI2、IAQI3、IAQI4、IAQI5、IAQI6 分别代表 SO_2 、 NO_2 、 $\text{PM}_{2.5}$ 、 PM_{10} 、 CO 、 O_3 的 IAQI 值，颜色代表相关性。通过观察图 4.1 可以发现，各个气象条件指标对单一污染物浓度及 AQI 造成影响的同时，不同气象条件的相互组合也会对相关污染物浓度及 AQI 造成不同的影响。

根据大量文献及常识可知风速和降雨量会对污染物浓度的扩散和沉降产生显著的影响^[7]，在单一污染物浓度影响分析中，风速和降雨量不作为重点因素考量，这两个综合气象条件会在下文整体性分析中加以详细阐述。因此，单一污染物浓度影响主要考虑的综合指标为大气压、总温度、总湿度三项。

根据上述 Spearman 相关系数的计算可得大气压、总温度、总湿度对分别 6 种污染物浓度的影响程度，表 4.1 给出了这 3 项综合气象条件对单一污染物浓度及 AQI 的影响排序及相关性。

表 4.1 综合气象条件对单一污染物浓度及 AQI 的影响分析

污染物浓度及 AQI	综合气象条件					
	大气压		总温度		总湿度	
	影响排序	相关性	影响排序	相关性	影响排序	相关性
SO_2	3	正相关	2	负相关	1	正相关
NO_2	2	正相关	1	负相关	3	正相关
PM_{10}	2	正相关	3	负相关	1	正相关
$\text{PM}_{2.5}$	1	正相关	2	负相关	3	正相关

O ₃	-	-	1	正相关	2	负相关
CO	2	正相关	1	负相关	3	正相关
AQI	2	正相关	3	负相关	1	正相关

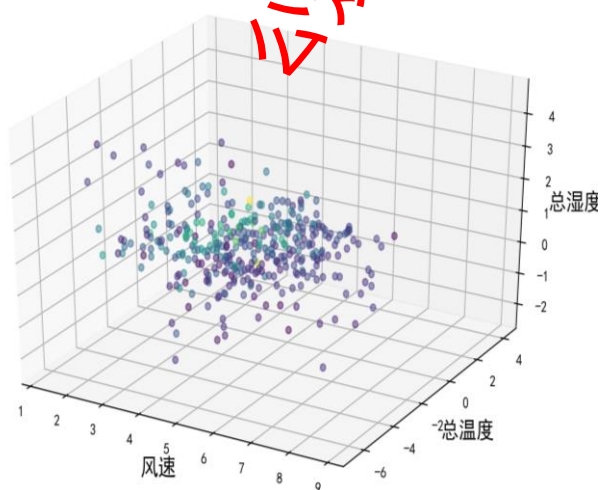
上表中的影响排序 1、2、3 表示综合气象条件对污染物浓度/AQI 的影响程度由高到低排序；相关性中正相关表示两变量一个随另一个的增大（减小）而增大（减小），变化趋势相同；负相关表示即一个随另一个的增大（减小）而减小（增大），变化趋势相反。

根据表 4.1 可以发现大气压、总温度、总湿度都对 SO₂、NO₂、PM_{2.5}、PM₁₀、CO 浓度和 AQI 值造成不同程度的影响，而臭氧 O₃ 作为本文中唯一分析的二次污染物，其变化过程为复杂的物理化学知识。影响臭氧浓度的因素主要有两项，一个是前体物 NO_x 和 VOCs 浓度，另一个就是风速风向、日照、温度、湿度等气象条件^[8]

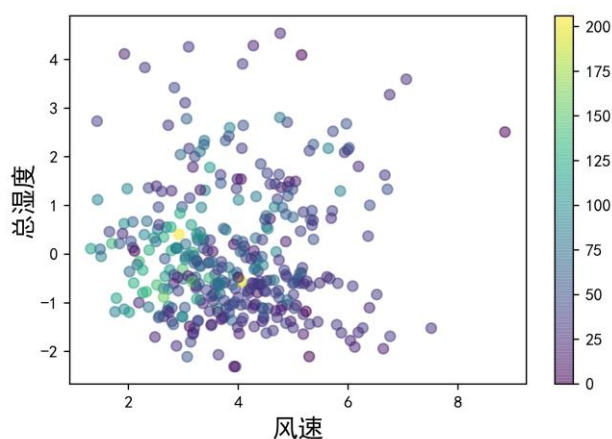
不同地区的 O₃ 污染浓度特征不同，例如我国京津冀、珠三角和长三角都出现过较为严重的区域性光化学烟雾^[7]，第二十四届大气污染防治技术研讨会就曾针对 2017 年珠三角地区就臭氧严重污染事件进行深入研讨^[10]，而中小城市也存在着受 O₃ 危害蔓延的趋势。

相同地区不同时间、不同气象条件下 O₃ 浓度也会发生变化。太阳紫外辐射是形成 O₃ 的光化学反应中一个重要条件，此外风速也会造成臭氧浓度的变化。夏季太阳紫外辐射强烈，温度相对较高，因此 O₃ 浓度污染最为严重，而冬季最弱。通过观察，表 4.1 可以看出 O₃ 浓度与总温度为正相关，与总湿度为负相关。

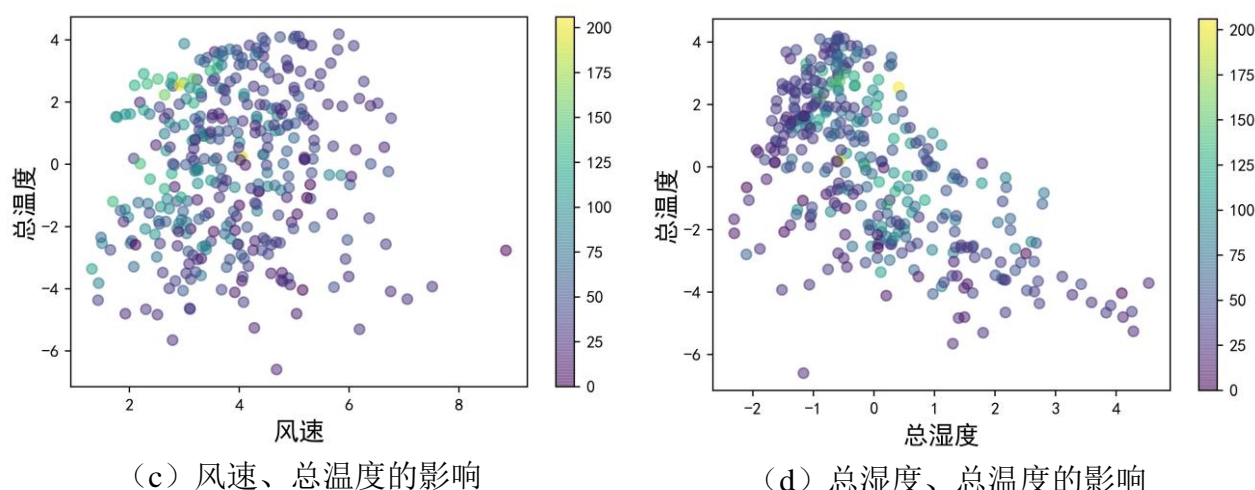
图 4.2 给出了风速、总温度、总湿度对 O₃ 的 IAQI 值大小的影响，图中横纵坐标为气象条件，右侧颜色变化为臭氧的 IAQI 值大小。由图可以发现总温度越大，臭氧 IAQI 越大；风速和总湿度越大，臭氧 IAQI 值越小。当风速十分大时，臭氧 IAQI 基本小于 50。相对总温度和风速，总湿度对 O₃ 浓度影响程度很小。



(a) 风速、温度、总湿度的影响

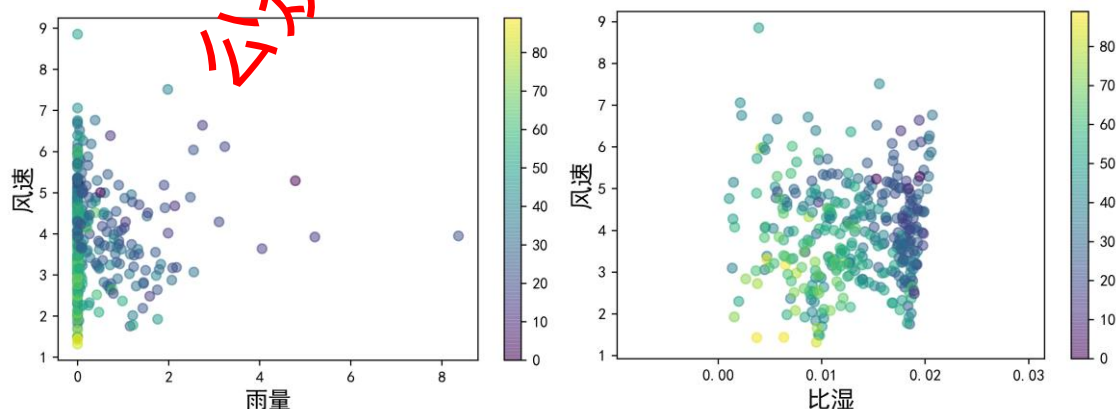


(b) 风速、总湿度的影响

图 4.2 风速、总温度、总湿度对 O_3 的 IAQI 影响

4.3.2. 整体性分析

由于降雨量对 AQI 存在显著影响，因此我们把是否降雨作为最主要分类条件，在此基础上加入其他综合气象条件的影响。风速有利于污染物的扩散，所以在有风情况下当地的 AQI 会下降，但还必须考虑大风天气下带来的扬尘可能导致 PM_{10} 、 $PM_{2.5}$ 浓度值的异常上升，从而导致 AQI 升高。根据上文分析，湿度是又一个重要的影响因素。图 4.3 给出了雨量和风速、比湿和风速对 PM_{10} 的 IAQI 值影响，可以看到当降雨/风速越大（小）， PM_{10} 的 IAQI 值越小（大）；当湿度大时，风速较大的 PM_{10} 的 IAQI 值较小，即湿度大风速大有利于污染物扩散，又不会带来较大扬尘，反之，当湿度较小时，风速较大时会使得 PM_{10} 的 IAQI 上升，导致扬尘。

图 4.3 风速、雨量、比湿对 PM_{10} 的 IAQI 影响

例如附件 1 所给的监测点 A 逐小时污染物浓度与气象实测数据中，2020 年 12 月 30 日 10:00 时的数据与 2020 年 8 月 19 日 11:00 时数据风速同为 4.8，但 12 月 30 日湿度为 41， PM_{10} 为 64，8 月 19 日湿度为 86、 PM_{10} 为 14。

上文中我们主要对大气压、总湿度、总温度三个综合气象条件引入气象条件变化度的

概念，以大气压为例，当大气压相较于前一天变化特别明显时，称为高大气压气象条件，故三个综合条件共可以形成八种天气分类，例如高大气压高温高湿度、低大气压高温低湿度等。

大气压、总温度、总湿度三个综合气象条件对 AQI 的影响存在正负, 且不同综合气象条件之间存在一定的联系与干扰, 所以未必会存在理想条件下的八种气象分类, 为此引入 AQI 差值校验, 将当日 AQI 与上日 AQI 作差, 在不考虑其他因素影响下, AQI 变化显著 (± 20) 的日期其气象条件一定存在显著变化影响, 将这样的天气与八种气象分类进行对应, 可完成最终分类。

气象条件分类结果如图 4.4，图中分类优先级表示一般情况同等条件下，气象条件变化相同数值时的相对分类优先级；AQI 影响程度也是一般情况同等条件下，气象条件变化相同数值时的相对影响程度；图中大小指的是数值大小，高低指的是数值变化高低。

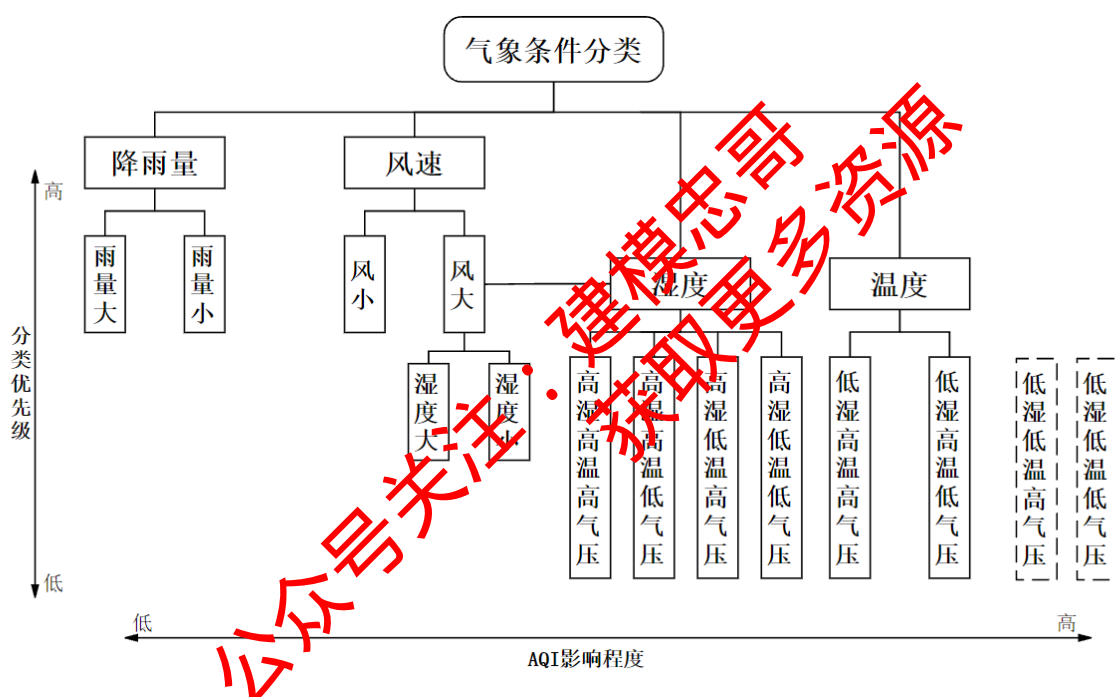


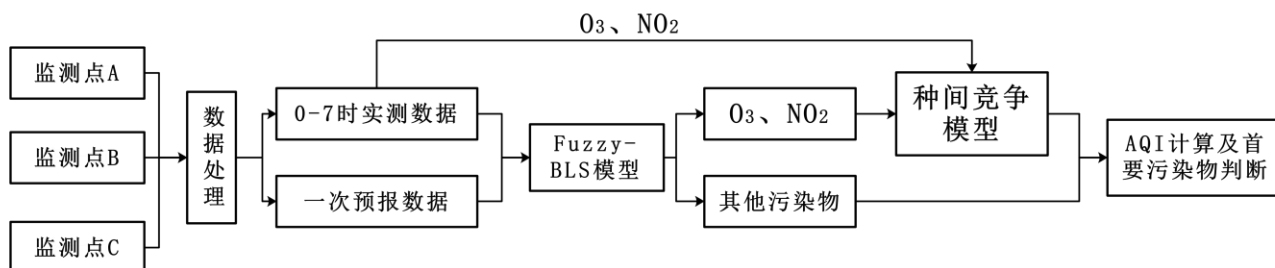
图 4.4 气象条件分类结果

当湿度发生显著变化时，即气象条件分类为高湿度高大气压高温、高湿度高大气压低温、高湿度低大气压高温、高湿度低大气压低温四种，一般地，AQI 升降与湿度升降有关，或者受湿度影响最大，典型天气有 2020 年 11 月 04 日、2021 年 04 月 05 日、2021 年 07 月 08 日等。

当湿度未发生显著变化时,温度成为主要影响因素(大气压由于同一地区海拔相同时,大气压几乎一致,但会受到温度等因素影响,不是单独变量,因此不针对性考虑)。低温高大气压低湿度比较少见,本题也不予考虑。即气象条件分类为高温高大气压低温度、高温低大气压低湿度,温度对 AQI 影响最大,典型天气有 2020 年 12 月 26 日、2021 年 03 月 25 日、2021 年 05 月 07 日、2021 年 06 月 08 日等。低温低大气压低湿度不会对 AQI 产生影响,不作重点分析。

五、问题三的建模与求解

问题三的模型流程图如下：



5.1 问题分析

问题三要求在 WRF-CMAQ 一次预报模型数据及实测数据基础上建立二次预报模型，模型需适用于相隔较远的三个监测点 A、B、C，同时满足 AQI 预报值的最大相对误差尽量小。本文引入专家经验对数据进行处理分析，对异常值进行剔除以及缺失值进行补充，提取数据的特征信息。采用基于宽度学习的 Takagi-Sugeno 模糊模型（Fuzzy-BLS），构建基于 Fuzzy-BLS 的二次预报模型，计算模型给出监测点 A、B、C 的初步预测结果，并对计算结果进行检验及误差计算。为提高准确度，采用“种间竞争”模型对 O_3 与 NO_2 浓度进行进一步修正分析，建立基于“种间竞争”的 Fuzzy-BLS 二次预报模型，求解模型给出最终的污染物浓度及 AQI 预测结果。

5.2 基于 Fuzzy-BLS 的二次预报模型

WRF-CMAQ 方法通过构建三维欧拉大气化学与传输模拟系统为天气预报提供了可行的解决方案，但由于天气变化的机理过于复杂，国内外相关研究都缺少精准的定量天气变化模型，这一点尤其在臭氧的产生上表现得尤为明显，针对无法通过物理信息模型精准反映天气变化的情况，采用基于数据驱动模型对一次预报数据进行修正无疑是更为合理、有效的。

同时在天气预报中，专家经验是不容忽视的重要因素，中国古代就有“燕子低飞要下雨”“天上瓦块云，地上雨淋淋”等谚语，故而对于无法通过物理化学模型进行定量精准描述的天气变化，引入专家经验无疑会增强模型的合理性。

综上所述，引入模糊推理系统，通过 if-then 规则对气象数据和污染物数据中所蕴含的专家经验进行提取、表述，以达到更好的预测效果，这里针对缺少物理信息模型的情况，常采用自适应神经模糊推理系统（ANFIS），但该模型在特征维数以及输出结果等方面存在一定的局限性^[11]，为了保留模糊推理系统的优势，又使得模型便于本题目求解，引入基于宽度学习的 Takagi-Sugeno 模糊模型，称为 Fuzzy-BLS^[12]。构建了基于 Fuzzy-BLS 的二次预报模型。Fuzzy-BLS 模型如图 5.1 所示。

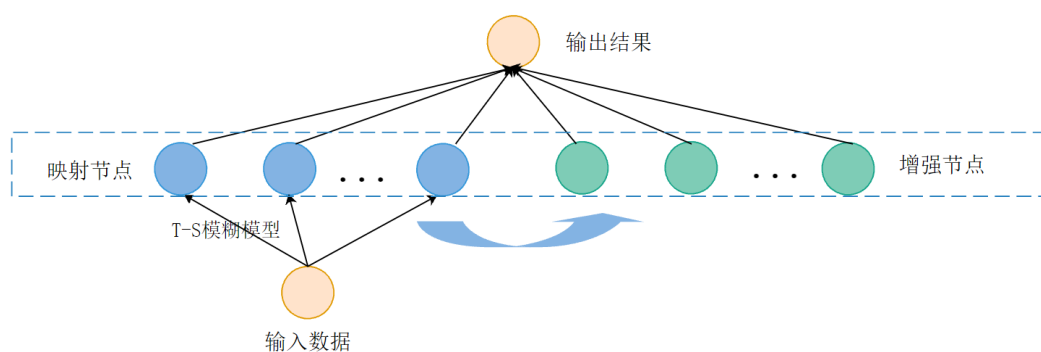
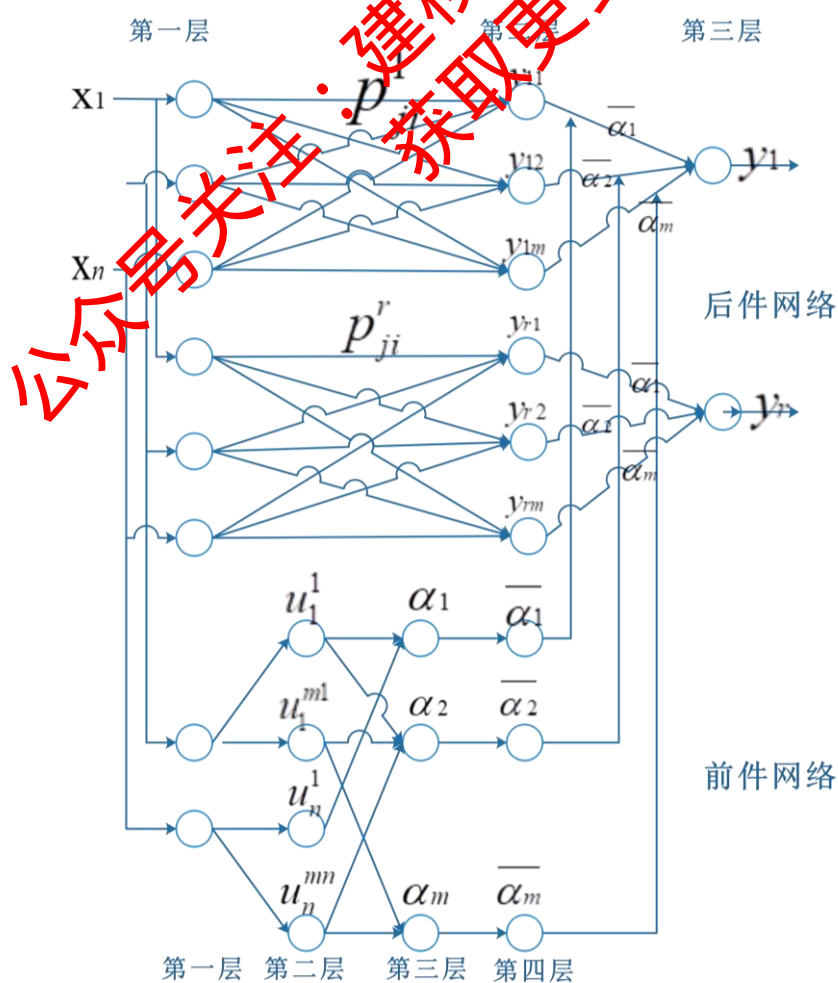


图 5.1 Fuzzy-BLS 模型

模型主要包含数据输入层，映射节点、增强节点层，输出结果层三层，数据输入后，通过 Takagi-Sugeno 模糊子系统进行初步计算，得到映射节点，同时将模糊子系统结果输入到增强层中，构建增强节点，将映射节点和增强节点通过岭回归或最小二乘法的方式确定每个节点的权重，最终输出结果，Takagi-Sugeno 模糊模型结构图如图 5.2 所示。

图 5.2 Takagi-Sugeno 模糊模型^[11]

具体流程如下：

在 Takagi-Sugeno 模糊模型^[12]中，通常使用有 *IF-Then* 规则，即

IF x_{s1} is A_{k1}^i and x_{s2} is A_{k2}^i and x_{sm} is A_{km}^i , Then $Z_{sk}^i = f_k^i(x_{s1}, x_{s2}, \dots, x_{sm})$
用公式表示为：

$$Z_{sk}^i = f_k^i(x_{s1}, x_{s2}, \dots, x_{sm}) = \sum_{t=1}^M \alpha_{kt}^i x_{st}$$

此处 α_{kt}^i 为一个系数，随机设为 (0, 1) 之间，后续通过岭回归算法求解伪逆得到，则：

$$Z_{si} = (w_{s1}^i Z_{s1}^i, w_{s2}^i Z_{s2}^i, \dots, w_{sk}^i Z_{sk}^i)$$

$$Z_i = (Z_{1i}, Z_{2i}, \dots, Z_{ni})$$

其中

$$w_{sk}^i = \frac{\tau_{sk}^i}{\sum_{k=1}^{ki} \tau_{sk}^i}$$

$$\tau_{sk}^i = \prod_{t=1}^M \mu_{kt}^i(x_{st})$$

$$\mu_{kt}^i(x) = \frac{1}{\sigma_{kt}^i} \exp\left(-\frac{(x - c_{kt}^i)^2}{2\sigma_{kt}^i}\right)$$

$\mu_{kt}^i(x)$ 为隶属度函数，使用高斯隶属度函数，对隶属度函数的参数，令 σ_{kt}^i 取 1， c_{kt}^i 通过 *k-means* 聚类算法求得。

再将此处得到的 Z_i 输入到增强层中：

$$H_j = \xi_j(Z^n W_{hj} + \beta_{hj})$$

$$H^m = (H_1, H_2, \dots, H_m)$$

得到增强节点输出，同时将 Z_i 在进行处理，得到映射节点输出：

$$F_{si} = \left(\sum_{k=1}^{ki} w_{sk}^i \left(\sum_{t=1}^m \delta_{kt}^i \alpha_{kt}^i x_{st} \right), \dots \right)$$

为便于求解引入 δ_{k1}^i ，作用和 α_{kt}^i 相似，则：

$$F_i = (F_{1i}, F_{2i}, \dots, F_{ni}) \triangleq D\Omega^i \delta^i$$

D 和 $\alpha_{kt}^i x_{st}$ 相关， Ω^i 和 w_{sk}^i 相关， δ^i 和 δ_{k1}^i 相关，那么总输出 Y 可以表示为：

$$Y = F^n + H^m W_e$$

$$Y = D\Omega\Delta + H^m W_e \triangleq (D\Omega, H^m)W$$

上式中需要求解的是 W ，通过岭回归算法求解伪逆得到

$$W = (D\Omega, H^m)^+ Y$$

$$(D\Omega, H^m)^+ = ((D\Omega, H^m)^T (D\Omega, H^m))^{-1} (D\Omega, H^m)^T$$

通过求解上述模型，得到监测点 A、B、C 在 2021 年 7 月 13 日至 15 日常规污染物 SO_2 、 NO_2 、 PM_{10} 、 $PM_{2.5}$ 、CO 的单日浓度值和 O_3 的最大八小时滑动平均，结果如表 5.1、表

5.2、表 5.3 所示（浓度计算结果保留两位小数）。

表 5.1 监测点 A 污染物浓度及 AQI 预测（初步）

二次模型日值预测									
预报日期	地点	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃ 最大八小时滑动平均	CO	AQI	首要污染物
		μg/m ³				mg/m ³			
2021/7/13	监测点 A	6.33	11.43	26.16	14.33	100.41	0.29	51	O ₃
2021/7/14	监测点 A	6.69	12.08	26.20	28.73	98.31	0.35	50	O ₃
2021/7/15	监测点 A	6.16	11.15	22.32	2.18	57.44	0.38	29	无

注：本表为初步结果，问题三最终结果为表 5.4、表 5.5、表 5.6

表 5.2 监测点 B 污染物浓度及 AQI 预测（初步）

二次模型日值预测									
预报日期	地点	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃ 最大八小时滑动平均	CO	AQI	首要污染物
		μg/m ³					mg/m ³		
2021/7/13	监测点 B	4.27	11.74	20.46	6.61	77.90	0.36	39	无
2021/7/14	监测点 B	3.24	9.93	19.13	9.44	66.15	0.41	34	无
2021/7/15	监测点 B	5.22	18.00	18.77	3.09	64.83	0.39	33	无

注：本表为初步结果，问题三最终结果为表 5.4、表 5.5、表 5.6

表 5.3 监测点 C 污染物浓度及 AQI 预测（初步）

二次模型日值预测									
预报日期	地点	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃ 最大八小时滑动平均	CO	AQI	首要污染物
		μg/m ³					mg/m ³		
2021/7/13	监测点 C	6.59	24.96	29.83	18.50	113.63	0.56	62	O ₃
2021/7/14	监测点 C	7.29	23.52	38.00	22.52	141.25	0.57	85	O ₃
2021/7/15	监测点 C	5.65	20.60	32.80	17.63	124.92	0.50	71	O ₃

注：本表为初步结果，问题三最终结果为表 5.4、表 5.5、表 5.6

为使分析更为准确，需对计算结果进行检验及误差计算。检验分为训练集以及测试集两部分，其均方误差分别为 $2.86\text{E-}17$ 、 0.01267 （零一归一化后的数据误差），通过观察发现误差十分小，进一步对 A、B、C 三个监测点的 AQI 的训练数据和测试数据结果及误差进行可视化，结果如图 5.3。

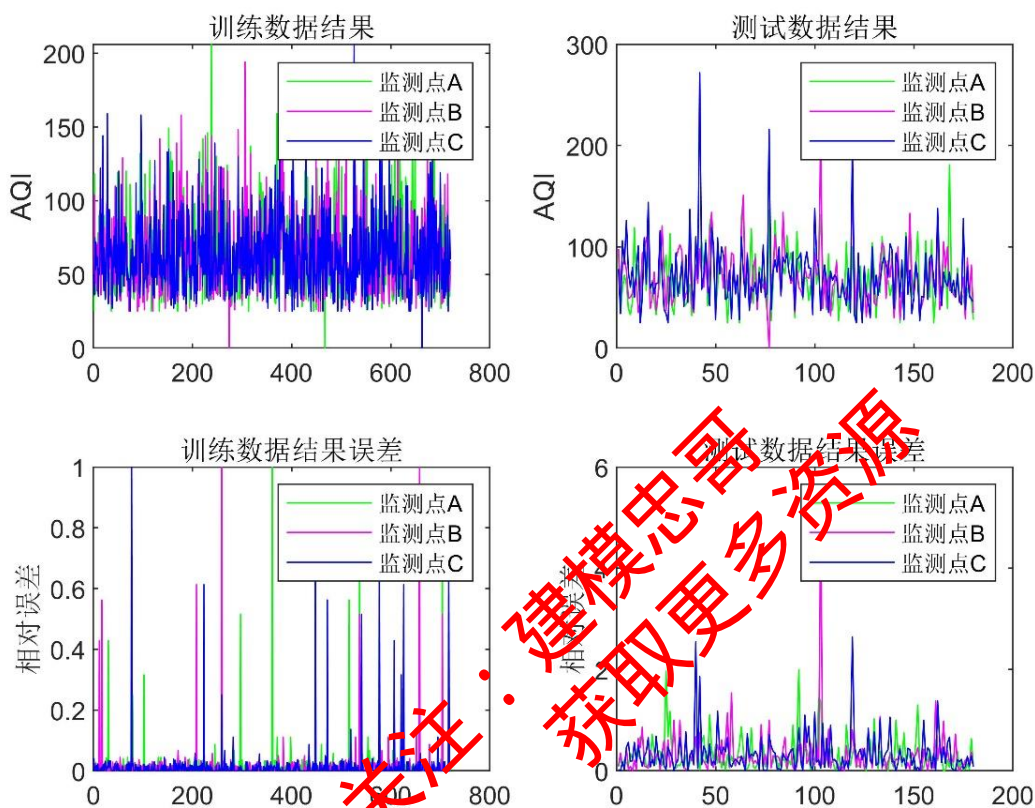


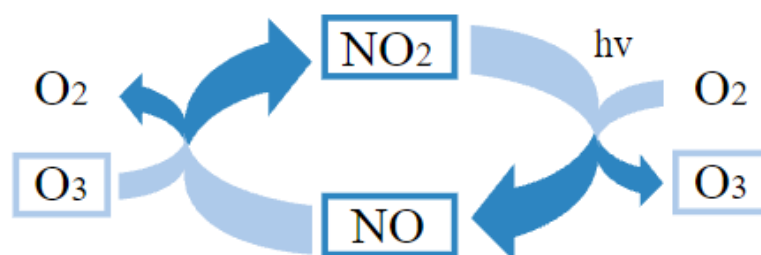
图 5.3 监测点 A、B、C 的 AQI 结果及误差可视化

图 5.3 中左右两列分别为 A、B、C 三个监测点的 AQI 训练数据或测试数据结果及相对误差，观察图 5.3 可以看出部分值还存在突变情况，相对误差存在特别大的现象。

通过对模型的分析，不难发现，该模型虽然均方误差比较小，但对于某些典型气象条件下预测的 AQI 相对误差较大，说明采用数据驱动的机器学习算法所得到的空气质量二次预报模型存在很大的局限性，因此需对模型进行进一步改进。

5.3 基于种间竞争的 Fuzzy-BLS 二次预报模型

上文所建立的 Fuzzy-BLS 二次预报模型中 AQI 存在部分值误差较大的情况，对二次污染物 O_3 的处理效果仍不理想， O_3 浓度除了受气象条件影响外，其形成属于复杂的化学反应，与氮氧化物 NO_x (NO 、 NO_2) 和挥发性有机物 VOCs 发生光化学循环，反应过程中 NO_x 逐渐过渡成为 O_3 主控因子^[8]。氮氧化物 NO_x (NO 、 NO_2) 和 O_3 的光化学循环如下图所示，图中 $h\nu$ 代表光解。

图 5.4 NO_x 和 O_3 的光化学循环

由图可知当 NO_2 与 O_2 在一定条件可与 NO 和 O_3 互相转换， NO_2 与 O_3 此消彼长，是一个动态转换过程。因此，引入“种间竞争”概念对 NO_2 与 O_3 浓度进行数据修正。

“种间竞争”是一种不同物种间为争夺生活空间、资源、食物而产生的一种直接或间接抑制对方的现象^[13]。将 NO_2 与 O_3 分别视作两个种群， NO_2 与 O_3 浓度之间的转换属于种间竞争中的干扰性竞争，即通过竞争个体间的直接相互作用开展竞争。当 O_3 经过化学反应转换成 NO_2 时势必会导致 O_3 的减少与 NO_2 的增加，反之亦然， O_3 与 NO_2 存在高度非线性关系。种间竞争模型建立了一个积分-微分方程组来模拟两物种群落的竞争动态^[14]，其主要思路如下：

(1) Logistic 增长

Logistic 增长模型又称自我抑制性方程，其曲线根据种群个体数共分为开始、加速、转折、减速、饱和五个时期，其微分形式如下：

$$\frac{dL}{dt} = rL\left(1 - \frac{L}{L_{\max}}\right)$$

其中 $\frac{dL}{dt}$ 是单位时间内种群个体数的数量变化， r 是比增长率， L 是种群个体的数量， L_{\max} 是种群个体数的上限。

(2) 种间竞争模型

种间竞争模型是对 Logistic 增长模型的延伸，假设有不同的两个种群，它们独自生存时数量变化均服从 Logistic 规律：

$$\begin{cases} \frac{dL_1}{dt} = r_1 L_1 \left(1 - \frac{L_1}{L_{1\max}}\right) \\ \frac{dL_2}{dt} = r_2 L_2 \left(1 - \frac{L_2}{L_{2\max}}\right) \end{cases}$$

两种群间存在竞争关系时，即一个种群个体数的增长与另一个种群个体数的减少成正比时，则有：

$$\begin{cases} \frac{dL_1}{dt} = r_1 L_1 \left(1 - \frac{L_1}{L_{1\max}} - q_1 \frac{L_2}{L_{2\max}}\right) \\ \frac{dL_2}{dt} = r_2 L_2 \left(1 - \frac{L_2}{L_{2\max}} - q_2 \frac{L_1}{L_{1\max}}\right) \end{cases}$$

其中 L_1 、 L_2 分别为这两个种群个体的数量， r_1 、 r_2 为它们的固有增长率， $L_{1\max}$ 、 $L_{2\max}$ 为两种群的最大个体数（即种群个体数的上限 L_{\max} ）， q_1 是指单位数量的种群 2（相对 L_2 ）对提供给种群 1 的资源的消耗量为单位数量种群 1（相对 L_1 ）消耗量的 q_1 倍， q_2 同理。

本题中， L_1 、 L_2 分别代表 O_3 和 NO_2 的浓度， $L_{1\max}(t)$ 、 $L_{2\max}(t)$ 是 O_3 和 NO_2 的最大浓度，为时变项，公式如下：

$$\begin{cases} L_{1\max}(t) = \omega c_1(t) + \beta \frac{c_{1\max}(t)}{T_{t\max}} \cdot T_t \\ L_{2\max}(t) = \xi c_2(t) - \eta \frac{c_{2\max}(t)}{T_{t\max}} \cdot T_t \end{cases}$$

其中， $c_1(t)$ 、 $c_2(t)$ 为 O_3 和 NO_2 的二次预报浓度， $c_{1\max}(t)$ 、 $c_{2\max}(t)$ 为 O_3 和 NO_2 的二次预报最大浓度， T_t 是总温度， $T_{t\max}$ 是总温度最大值， ω 、 β 、 ξ 、 η 分别取 1，0.1，1.2，0.1。对于微分方程， r_1 、 r_2 分别取 0.8 和 1.2， q_1 、 q_2 分别取 0.03 和 0.08。

根据问题二的数据处理，近地 2 米温度、地表温度、感热通量、潜热通量、长波辐射、短波辐射、地面太阳能辐射、边界层高度归为“总温度”下的分支。由表 4.1 可知， O_3 与总温度正相关， NO_2 与总温度为负相关。当总温度高时， O_3 竞争比 NO_2 竞争的优势大，所以 O_3 浓度升高， NO_2 浓度降低；当总温度低时， O_3 竞争比 NO_2 竞争的优势小，所以 O_3 浓度降低， NO_2 浓度升高。

采用种间竞争模型对 O_3 、 NO_2 的 IAQI 值进行拟合，结果如图 5.5 所示。图中横坐标是指一天中的某一时刻，纵坐标是污染物浓度大小，单位为 $\mu g/m^3$ ，曲线说明在图中给出。特别地，平均值指经过数据处理后，该污染物浓度大小在某一时刻的平均值。例如某一污染物早晨 8 点的浓度为污染物所有早晨 8 点平均浓度。

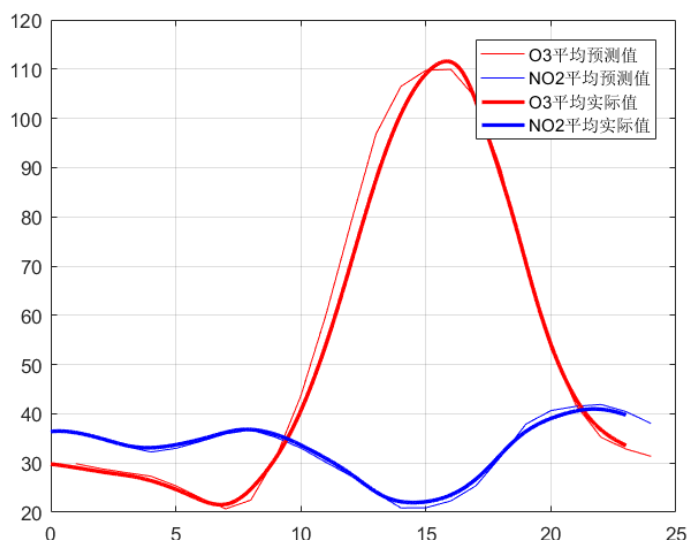
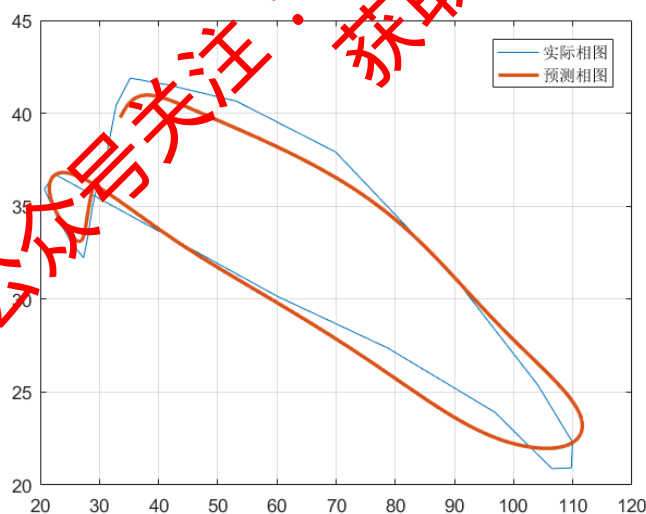
图 5.5 O_3 、 NO_2 平均 IAQI 实测值与预测值拟合

图 5.5 中, O_3 和 NO_2 的初始浓度为前八小时实测的平均浓度, 数值为 29.8 和 $36.4\mu g/m^3$ 。根据图 5.5 可以明显发现 NO_2 与 O_3 的竞争关系, 同时真实值曲线与预测值曲线近乎重合。对 NO_2 与 O_3 整体做一个实际相图和预测相图, 得到如图 5.6 所示的结果。

图 5.6 O_3 、 NO_2 实际相图和预测相图

结合图 5.5 及图 5.6 可以发现, 对 O_3 及 NO_2 预测近乎与实际数据曲线相重合, 采用种间竞争模型对 O_3 及 NO_2 的数据修正与预测十分有效, 且具有极高的准确度。Fuzzy-BLS 模型对 O_3 浓度预测的相对误差为 0.511 , NO_2 浓度预测的相对误差为 0.2283 , 经过种间竞争模型修正后, O_3 浓度的相对误差为 0.0379 , NO_2 浓度的相对误差为 0.0227 。

对二次预报模型求解, 得到监测点 A、B、C 在 2021 年 7 月 13 日至 15 日常规污染物 SO_2 、 NO_2 、 PM_{10} 、 $PM_{2.5}$ 、CO 的单日浓度值和 O_3 的最大八小时滑动平均, 最终结果如表 5.4、表 5.5、表 5.6 所示 (浓度计算结果保留两位小数)。

表 5.4 监测点 A 的 AQI 值及首要污染物（问题三答案）

预报日期	地点	二次模型日值预测							AQI	首要污染物
		SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃ 最大八小时滑动平均	CO			
		μg/m ³					mg/m ³			
2021/7/13	监测点 A	6.33	14.77	26.16	14.33	88.87	0.29	45	无	
2021/7/14	监测点 A	6.69	17.36	26.20	28.73	159.53	0.35	100	O ₃	
2021/7/15	监测点 A	6.16	14.64	22.32	2.18	103.68	0.38	54	O ₃	

表 5.5 监测点 B 的 AQI 值及首要污染物（问题三答案）

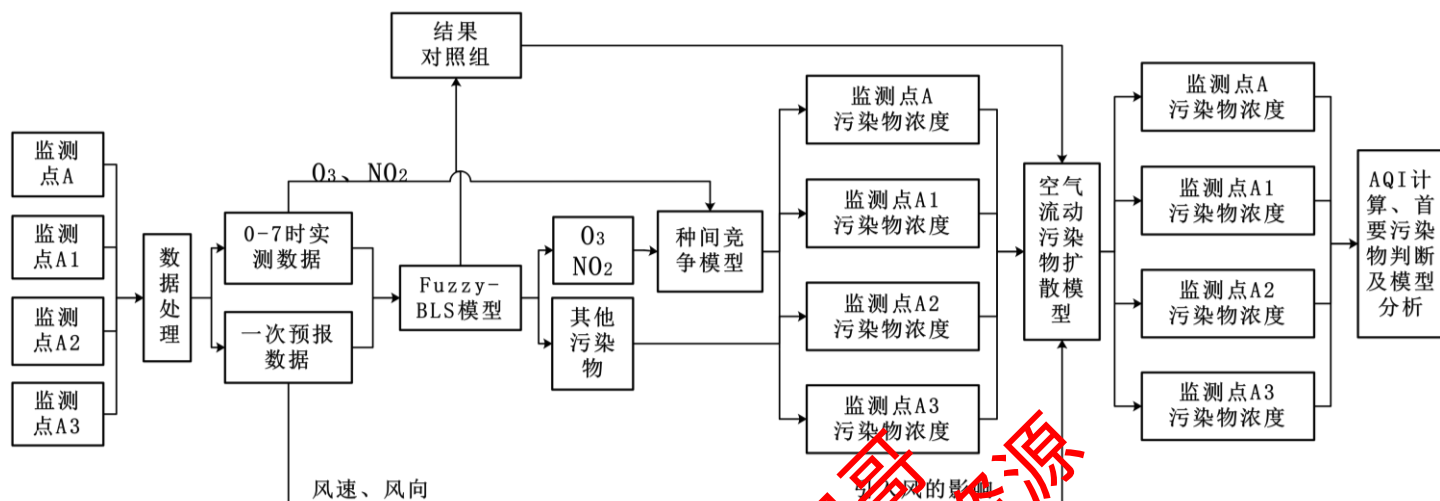
预报日期	地点	二次模型日值预测						AQI	首要污染物
		SO ₂	NO ₂	PM ₁₀	PM _{2.5}	最大八小时滑动平均	CO		
		μg/m ³				mg/m ³			
2021/7/13	监测点 B	4.27	9.95	20.46	6.61	87.75	0.36	44	无
2021/7/14	监测点 B	3.24	11.03	19.13	9.44	82.26	0.41	42	无
2021/7/15	监测点 B	5.22	9.95	18.77	3.09	59.90	0.39	30	无

表 5.6 监测点 C 的 AQI 值及首要污染物（问题三答案）

二次模型日值预测									
预报日期	地点	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃ 最大八小时滑动平均	CO	AQI	首要污染物
		μg/m ³					mg/m ³		
2021/7/13	监测点 C	6.59	18.28	29.83	18.50	135.78	0.56	80	O ₃
2021/7/14	监测点 C	7.29	15.52	38.00	22.52	136.29	0.57	81	O ₃
2021/7/15	监测点 C	5.65	19.77	32.80	17.63	118.86	0.50	66	O ₃

六、问题四的建模与求解

问题四模型流程如下



6.1 问题分析

问题四针对附件 1、3 的数据，要求对监测点 A 及临近区域内的监测点 A1、A2、A3 建立区域协同预报模型。本文在问题三所建立的基于种间竞争的 Fuzzy-BLS 二次预报模型基础上，引入空气流动模型对临近区域进行协同二次预报，进一步提升 AQI 预测的准确度。首先在对原始数据中的风向进行坐标系转换，从平均风向及风向频率两个角度对风向进行分析，引入与风向频率成正比、与平均风速成反比的污染系数。对平均风向做加权处理，采用虚拟点光源法的面源扩散模式，给出污染物从扩散城市到受影响城市的浓度计算方法。通过求解模型，给出了污染物从某一城市扩散到临近城市的三维趋势图，给出最终的污染物浓度及 AQI 预测结果，并对问题三、问题四的模型求解结果进行误差对比分析。

6.2 基于空气流动的污染物扩散模型

相邻区域的污染物浓度会因区域间的空气流动而改变，影响空气流动的原因主要包括风速及风向。低层大气中的空气流动因风向及风速大小的改变而改变，我们对监测点 A 及其相邻的监测点 A1、A2、A3 的风速和风向进行分析。

(1) 风向分析

首先对题目所给的风向度数进行坐标系转换，转换公式为：

$$wind_v = 180^\circ - (wind_v' + 90^\circ) = 90^\circ - wind_v'$$

其中， $wind_v'$ 为题目所给风向度数，范围为 $0 \sim 360^\circ$ ； $wind_v$ 为 x-y 坐标系下的角度，转换示意图如图 6.1 所示。

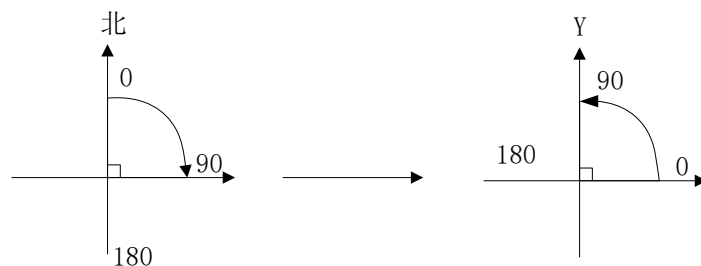


图 6.1 风向坐标转换示意图

假设一个质点运动轨迹的长度与方向仅与风向维持在某一角度的时间和该角度大小有关，对监测点 A、A1、A2、A3 的一天中各个时刻的风向进行可视化，结果如图 6.2 所示。根据观察可以发现各个时刻的风向轨迹杂乱无章，不能直接对其进行分析，必须对风向角度做相关处理。

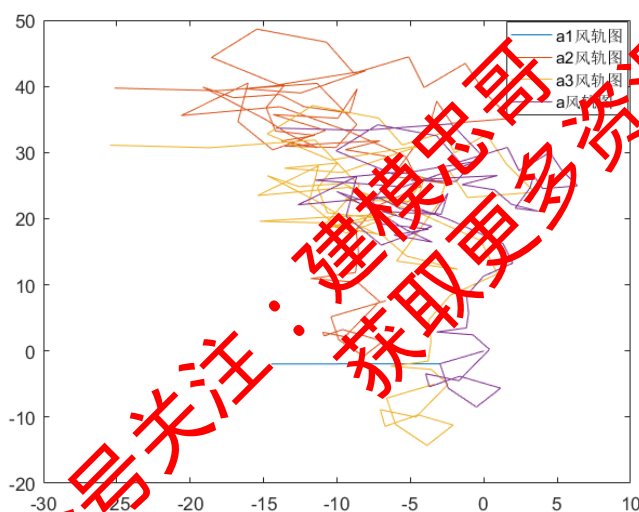


图 6.2 监测点 A、A1、A2、A3 的风向轨迹图

取风向角度的平均及风向频率，对其进行可视化，结果如图 6.3 所示，根据图 6.3 可以看到风向取平均值及风向频率后，可分析性大大提升。

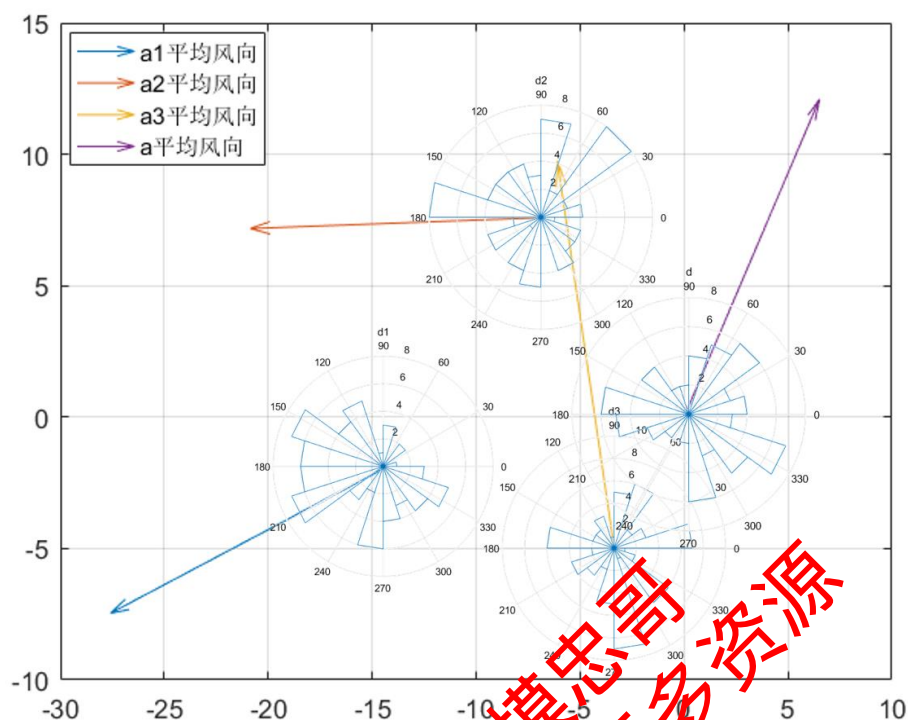


图 6.3 监测点 A、A1、A2、A3 的平均风向及风向频率

(2) 污染系数确定

综合考虑风向频率和平均风速，设定污染系数 ε ：

$$\varepsilon = \frac{f}{u}$$

其中， f 是风向频率， u 为平均风速。

上式表明某方位的污染系数与风向频率成正比，与平均风速成反比。如果该方位风速大、风向频率小，那么该方位的污染系数小，其下风向受到的影响就小。

(3) 加权平均风向

对平均风向进行加权处理：

$$wind_{oa} = \sum_{t=0}^{23} \frac{wind_s}{wind_{ns}} wind_o$$

其中， $wind_{oa}$ 表示加权平均风向， $wind_s$ 为当前风速， $wind_{ns}$ 是总风速之和， $wind_o$ 表示风向， t 指该天的某一时刻。

基于上文相关指标的分析，对空气流动模型进行简化，只考虑平均风向及其风速对相邻监测点的影响。将监测点简化为污染连续面源^[15]，使用虚拟点光源法的面源扩散模式，将城市简化为正方形，边长取为 0.5km，城市风向对应关系示意图如图 6.4 所示。

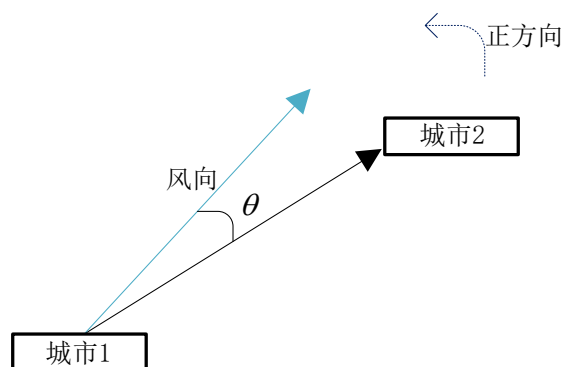


图 6.4 城市风向对应关系

简化模型，我们做如下假设：

(1) 假设有一距离为 x_0 的点源在面源的上风口，则它在面源单元中心线产生的污染物烟流宽度为 $2y=B$ ，其中， B 为面源宽度。

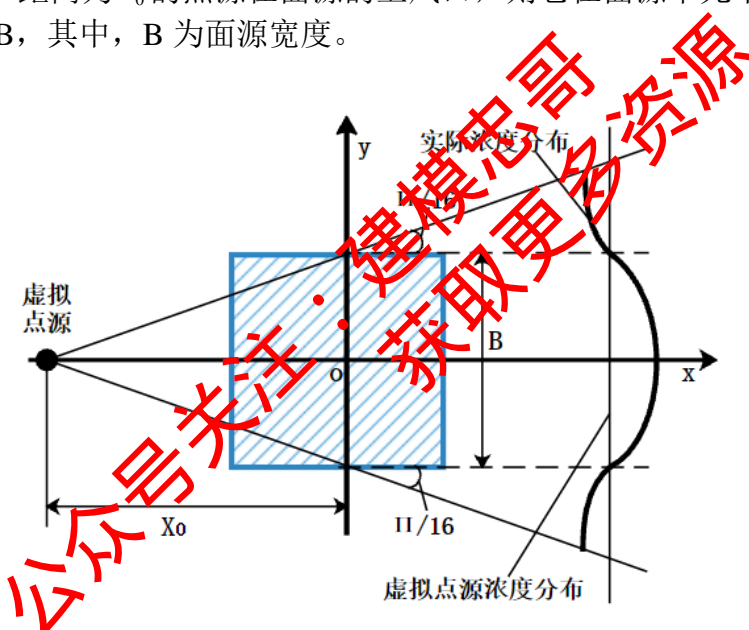


图 6.5 面源示意图

(2) 假设污染物均匀的分布在面源的 y 方向，且扩散后的污染物均匀分布，弧的长度为 $\pi(x_0 + x)/8$ 。

(3) 假设面源只对下风向起作用，不对上风向起作用，即取 θ ($-90^\circ < \theta < 90^\circ$) 为城市连线与风向夹角，与加权平均风向 $wind_{oa}$ 有关。

则下风口任一点的地面浓度为：

$$C = \sqrt{\frac{2}{\pi}} \frac{q\varepsilon}{\sigma_z \pi(x_0 + x)/8} \exp\left(-\frac{H^2}{2\sigma_z^2}\right) \cos \theta, (-90^\circ < \theta < 90^\circ)$$

以监测点 A、A3 为例，上式中 x_0 取 1000 米； x 为 6033 米，根据计算 A (0, 0) 与 A3 (-3.3543, -5.0138) 的坐标之间的距离得到； H 根据原始数据在一次预报中给出的风速高度取值 10 米； q 为单位时间内排放的污染物浓度，单位 $\mu\text{g}/\text{m}^3$ ； σ_z 根据图 6.6 P-G 扩

散曲线进行查找计算。

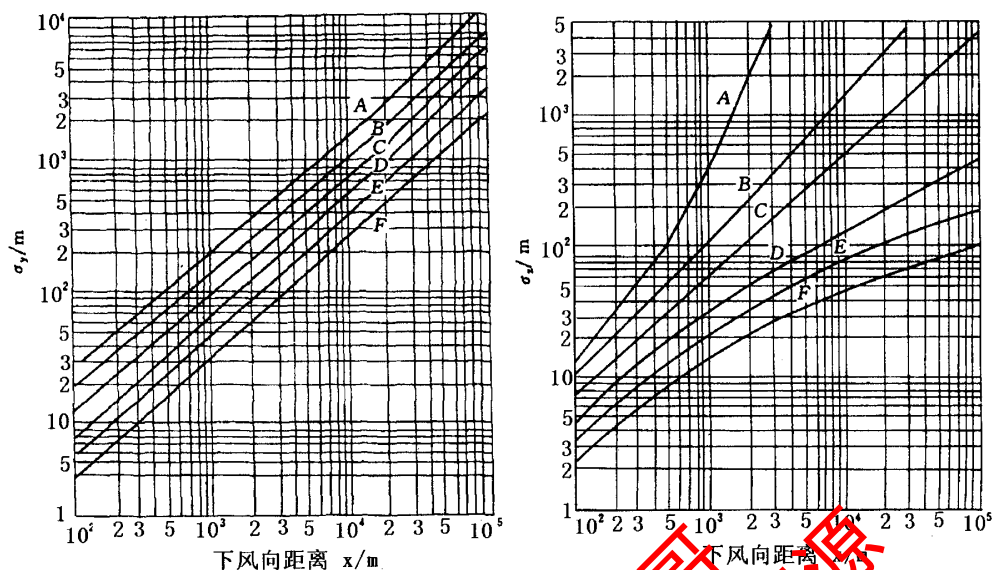


图 6.6 P-G 扩散曲线[51]

扩散城市为 i ，受影响城市为 j ， i 城市污染物的浓度为 D_i^R 。

$$\Delta D_{ij}^R = D_i^R - D_j^R$$

由上述模型得到污染物 R 从城市 i 扩散到受影响城市 j 的浓度值 Df_{ij}^R 后，由第三问模型 j 地点污染物 R 的浓度二次预测值 D_{pre2j} ，结合上述因素通过下式计算得到最终的预测值：

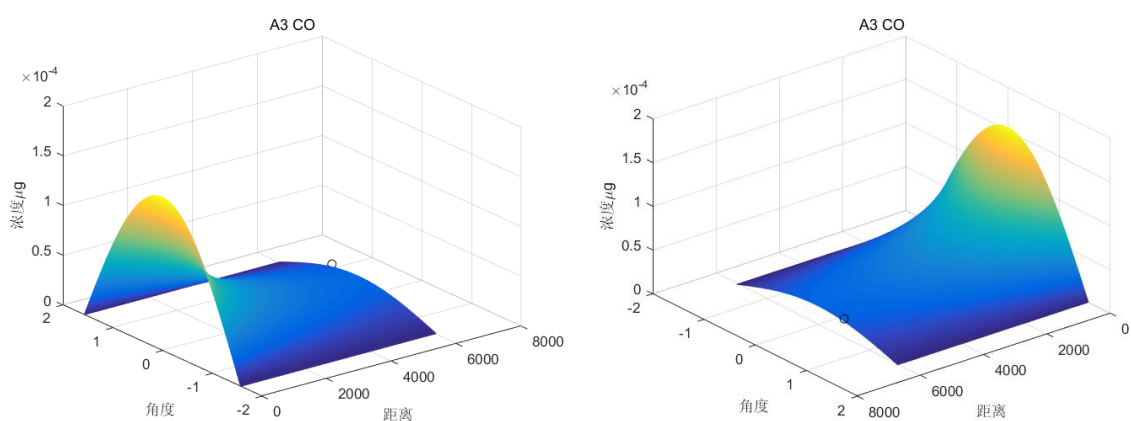
$$D_{pre,j} = D_{pre2j} + \Delta D_{ij}^R \frac{Df_{ij}^R}{D_i^R}$$

如果城市 i 污染物浓度小于城市 j 的污染物浓度时， $\Delta D_{ij}^R < 0$ ，表明空气流动导致受影响城市的污染物浓度降低，反之，增加。

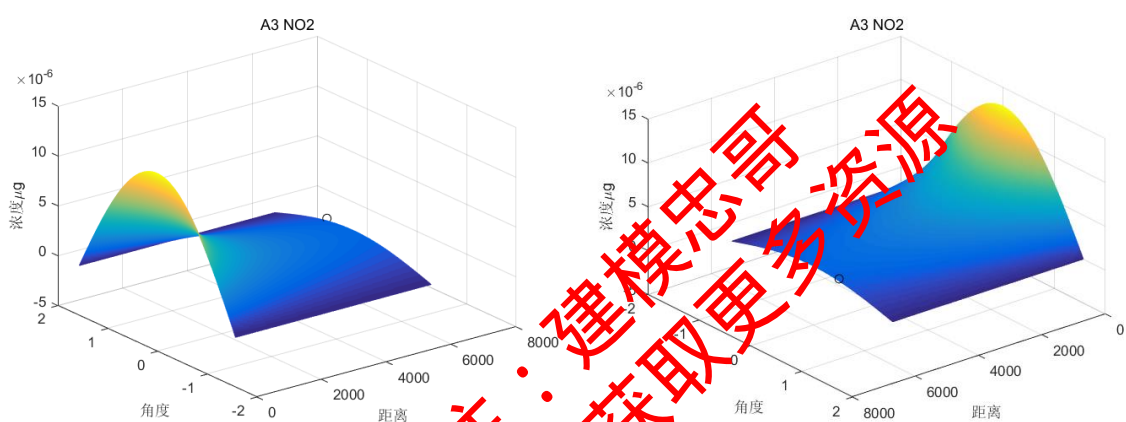
6.3 模型求解及误差对比分析

6.3.1 模型求解

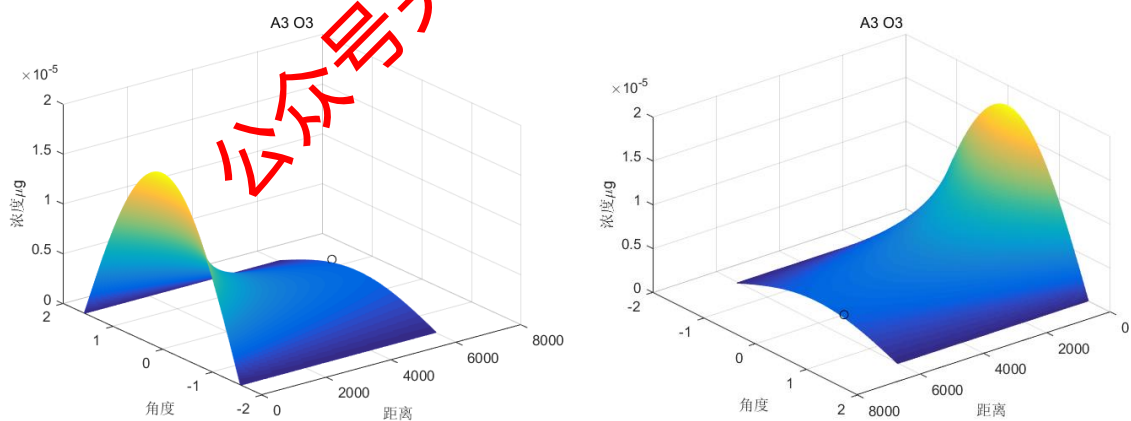
仍以 A 与 A3 为例，对模型进行求解，得到 A3 的 6 个污染物扩散到 A 的情况。三维可视化结果如图 6.7 所示。图中三个坐标轴分别为监测点 A 与 A3 之间的距离、角度 θ 、污染物浓度，角度与距离平面上的圆圈为 A 点位置，黄色最严重的地方为 A3 的位置，即扩散起点。通过观察可以发现扩散点 A3 的污染物起始浓度不同，扩散到 A 的趋势相似。



(a) CO



(b) NO_2



(c) O_3

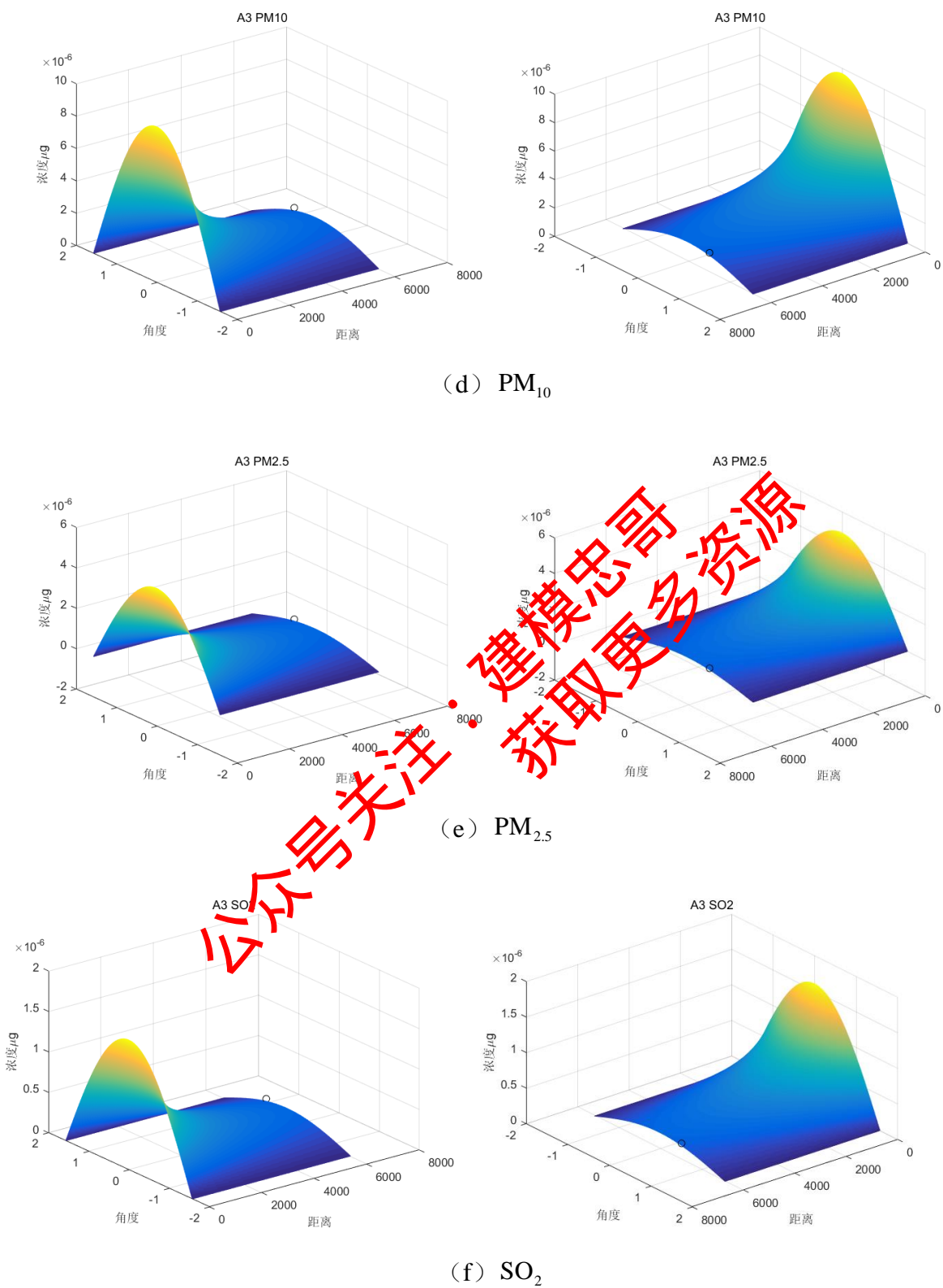


图 6.7 各污染物从 A3 扩散到 A 趋势图

根据问题三的模型，我们计算出 A、A1、A2、A3 各个污染物的浓度、AQI 以及首要污染物，在问题三的基础上我们引入基于空气流动的污染物扩散模型，进一步对监测点 A 进行修正，即为问题四的模型，对该模型进行求解得到监测点 A 的修正值。得到的结果如表 6.1 所示。

表 6.1 问题三、问题四模型计算结果

		二次模型日值预测							
预报日期	地点	SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃	CO	AQI	首要污 染物
						mg/m ³			
		μg/m ³							
2021/7/13	监测点 A	6.325079	14.77334	26.15912	14.32846	88.8687	0.292883	45	无
	监测点 A1	6.977533	14.29092	22.75845	8.109748	109.7299	0.390966	59	O ₃
	监测点 A2	5.956885	16.00487	29.76733	10.63463	120.8048	0.477755	68	O ₃
	监测点 A3	3.74213	5.767264	13.26247	5.892412	99.57139	0.344064	50	O ₃
	监测点 A 修正	6.325079	15.14789	26.15912	14.32846	94.57139	0.292883	48	无
2021/7/14	监测点 A	6.691752	17.35968	26.20493	28.73168	159.5271	0.35139	100	O ₃
	监测点 A1	8.65663	16.78211	29.84319	13.79615	176.5578	0.354742	116	O ₃
	监测点 A2	5.558868	19.53231	25.10382	8.878504	184.2708	0.413796	123	O ₃
	监测点 A3	5.170923	12.56483	14.44505	9.558678	164.9533	0.285112	105	O ₃
	监测点 A 修正	6.691752	18.1572	26.20493	28.73168	162.3125	0.35139	102	O ₃
2021/7/15	监测点 A	6.163751	14.63566	22.31957	2.175676	103.1774	0.376449	54	O ₃
	监测点 A1	8.465374	14.71777	27.69855	10.40853	120.5328	0.375758	67	O ₃
	监测点 A2	5.982272	19.12046	25.48548	6.729074	111.2994	0.496832	60	O ₃
	监测点 A3	3.987281	8.042781	17.55314	3.804669	97.42438	0.327107	49	无
	监测点 A 修正	6.163751	13.96033	22.31957	2.175676	105.4391	0.376449	55	O ₃

采用本问题的空气质量预报模型，对监测点 A 的 AQI 值及首要污染物进行计算，结果如下表：

表 6.2 监测点 A 的 AQI 值及首要污染物（问题四答案）

预报日期	地点	二次模型日值预测						AQI	首要污 染物
		SO ₂	NO ₂	PM ₁₀	PM _{2.5}	O ₃	CO		
		μg/m ³					mg/m ³		
2021/7/13	监测点 A	6.325079	15.14789	26.15912	14.32846	94.57139	0.292883	48	无
2021/7/14	监测点 A	6.691752	18.1572	26.20493	28.73168	162.3125	0.35139	102	O ₃
2021/7/15	监测点 A	6.163751	13.96033	22.31957	2.175676	105.4391	0.376449	55	O ₃

6.3.2 误差对比分析

通过对 2020-07 到 2021-07 实测数据的对比分析发现，在问题三、问题四建立的模型，均对空气质量有较好的预报，问题三的二次预报模型相对误差为 0.1086，问题四的二次预报模型相对误差为 0.0623，相比而言问题四引入 A1、A2、A3 的修正，能够进一步提升模型的准确性，而且引入 A1、A2、A3 的修正，在面对 A 点数据异常、缺失的情况时，有更可靠的预测结果。

对问题三、问题四的结果误差进行可视化，结果如图 6.8 所示

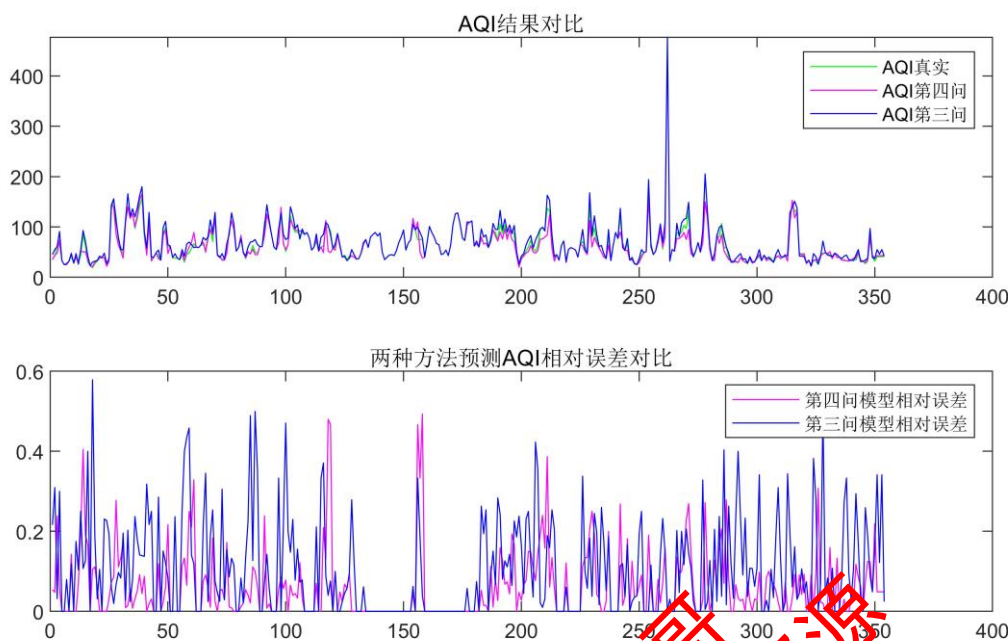


图 6.8 第三题、第四题误差对比

由图 6.8 可以发现，问题四模型引入了协同预测，建立空气流动的污染物扩散模型后明显减小了问题三模型求解结果的误差，使得空气质量预报更加准确，更接近真实值，可信度更高。

七、模型评价与改进

7.1 模型的优点

(1) 采用 PAC 及 KPCA 算法对数据进行降维，减轻了计算难度，减少了需要分析的指标，有利于数据处理更为准确、指标分析更为全面，提高了综合气象条件分类的合理性及准确度；

(2) 采用多种方法进行二次预报模型建立，本文在基于数据驱动的 Fuzzy-BLS 模型的基础上，先后引入了种间竞争模型、空气流动污染物扩散模型对结果进行修正，得到较为可信的空气质量二次预报模型；

(3) 充分考虑不同气象条件对空气质量的影响，重点对 NO_2 、 O_3 的预测进行改进，较好的解决了 O_3 预报不准确的问题，提升了模型的准确率；引入风速、风向的修正，进一步提高预测准确率，同时有效避免了 A 点数据丢失、异常值等情况所带来的预测误差，提升了模型的稳定性；

(4) 对数据进行充分的挖掘分析，较好的展示了不同空气污染物和气象条件之间的关联，例如，文献中提到 O_3 存在周末效应，而根据所给数据并未发现相关规律，同时在完成相关预测的同时，模型提出的相关优化方法，找到的相关规律联系可以应用在相关研究中，模型实用价值高。

7.2 模型的缺点

(1) 由于题目未给定监测点的真实地理位置，无法考虑当地地形因素、当地人口密集程度、经济发展程度、相关防治政策对臭氧浓度造成的影响；

(2) 第四问微分方程的计算某些参数需要查表获得超参的数值，没有程序化，带来使用不便；

(3) 微分方程对初始值的要求比较高，极端情况和异常突变等不可求导的值会导致模型产生错误；

(4) 微分方程使用 ODE45 进行数值求解，存在一定的误差。

公众号关注：建模忠哥
获取更多资源

参考文献

- [1] David Y.H. Pui, Sheng-Chieh Chen, Zhili Zuo. PM_{2.5} in China: Measurements, source s, visibility and health effects, and mitigation[J]. Particuology, 2014, 13(02): 1-26.
- [2] Karavas Zissis, Karayannis Vayos, Moustakas Konstantinos. Comparative study of air quality indices in the European Union towards adopting a common air quality index[J]. Energy & Environment, 2021, 32(6).
- [3] 卢亚灵, 李勃, 范朝阳, 王建童, 张鸿宇, 蒋洪强. 空气质量预测模拟技术演变与发展研究[J]. 中国环境管理, 2021, 13(04): 84-92.
- [4] 机器学习-数据降维之 PCA(SVD 奇异值分解&特征值分解), https://blog.csdn.net/qq_35912099/article/details/107244034
- [5] pca 和 KPCA 的详细概念, <https://max.book118.com/html/2016/0720/48715523.shtml>
- [6] Pulido-Velazquez D., Collados-Lara A.J., Pérez-Sánchez J., Segura Méndez Francisco J osé, Senent-Aparicio J.. Climate change impacts on the streamflow in Spanish basins monitored under near-natural conditions[J]. Journal of Hydrology: Regional Studies, 2021, 38.
- [7] Fu Zhongjie, Lin Haiping, Huang Bingrang, Yao Jiana. Research on air quality prediction method in Hangzhou based on machine learning[J]. Journal of Physics: Conference Series, 2021, 2101(1).
- [8] 赵秋月, 李荔, 李慧鹏. 国内外近地面臭氧污染研究进展[J]. 环境科技, 2018, 31(05): 72-76.
- [9] 符传博, 周航. 中国城市臭氧的形成机理及污染影响因素研究进展[J]. 中国环境监测, 2021, 37(02): 33-43.
- [10] 王浩霖, 吴锴, 刘一鸣, 范绍佳. 珠三角热浪期间持续臭氧污染的形成机制[A]. 中国环境科学学会. 第二十四届大气污染防治技术研讨会论文集[C]. 中国环境科学学会: 中国环境科学学会, 2020: 1.
- [11] 李国勇, 杨丽娟. 神经·模糊·预测控制及其 MATLAB 实现. 第3版[M]. 电子工业出版社, 187-287, 2013.
- [12] Shuang F, Chen C. Fuzzy Broad Learning System: A Novel Neuro-Fuzzy Model for Regression and Classification[J]. IEEE Transactions on Cybernetics, 2018, PP(99): 1-11.
- [13] 种间竞争, <https://baike.baidu.com/item/%E7%A7%8D%E9%97%B4%E7%AB%9E%E4%BA%89/2318027>
- [14] LIU Yan, HE Ze-rong. Stability results for a nonlinear two-species competition model with size-structure[J]. Applied Mathematics: A Journal of Chinese Universities, 2021, 36(01): 1-15.
- [15] 张耀, 王湛, NU YU. 修正高斯扩散模型对机场污染物浓度预测影响[J]. 环境保护科学, 2021, 47(03): 106-112.

附录

附录 1: AQI 计算代码

```

load shiche1.mat;
x=AS1;
IAQck=[0 50 100 150 200 300 400 500;
        0 50 150 475 800 1600 2100 2620;
        0 40 80 180 280 565 750 940;
        0 50 150 250 350 420 500 600;
        0 35 75 115 150 250 350 500;
        0 100 160 215 265 800 800 800;
        0 2 4 14 24 36 48 60];
[a b]=size(x);
AQI=zeros(a,2);
IAQ=zeros(a,2);
for i=1:1:a
    for j=1:1:6
        if x(i,5)>=800
            AQI(i)=-5;
            break;
        end
        for k=1:1:8
            if x(i,j)<=IAQck(j+1,k)
                Hi=k;
                IAQ(i,j)=
ceil(((IAQck(1,Hi)-IAQck(1,Hi-1))/(IAQck(j+1,Hi)-IAQck(j+1,Hi-1)))*(x(i,j)-IAQck(j+1,Hi-1))
+IAQck(1,Hi-1));
                break;
            end
        end
        if IAQ(i,j)>=500
            AQI(i)=-j;
            break;
        end
    end
    [AQI(i,1),AQI(i,2)]= max(IAQ(i,:));
end

```

附录 2: 数据处理代码

%缺省日期查找，数据直接由 matlab 导入附件


```
index=zeros(44390-44039,1);
S2=floor(S5);
for i=44039:1:44390
    index(i-44038)=length(find(S2(:,1)==i));
end
Z=find(index~=24);

%缺省数据补充及平均值计算
%0-7 实测数据
load matlabshuruA32.mat;
chushix=S5;
[a,b] = size(chushix);
temp = isnan(chushix);
result = chushix;
for i=1:a
    for j=1:b
        if temp(i,j) == 1
            result(i,j) = result(i-1,j);
        end
    end
end
x=result;
c=ceil(a/24);
shuru2=zeros(c,6);
for i=1:1:c
    for j=1:1:8
        for k=1:1:6
            shuru2(i,k)=shuru2(1,k)+x(((i-1)*24)+j,k);
        end
    end
end
shuru2=shuru2./8;
%输出数据
load matlabshuruA33.mat;
chushix=S8;
[a,b] = size(chushix);
temp = isnan(chushix);
result1 = chushix;
for i=1:a
    for j=1:b
        if temp(i,j) == 1
            result1(i,j) = result1(i-1,j);
        end
    end
end
```

```

end
shuchu=result1;
%输入数据
load matlabshuruA31.mat;
chushix=S2;
[a1,b] = size(chushix);
temp = isnan(chushix);
result2 = chushix;
for i=1:a1
    for j=1:b
        if temp(i,j) == 1
            result2(i,j) = result2(i-1,j);
        end
    end
end
x2=result2;
c2=ceil(a1/72);
shuru1=zeros(c2,b*3);
for i=1:1:c2
    for j=1:1:3
        for z=1:1:b
            for k=1:1:24
                shuru1(i,(j-1)*b+z)= shuru1(i,(j-1)*b+z)+ x2(((i-1)*72+(j-1)*24+k),z);
            end
        end
    end
end
shuru1=shuru1./24;
cc=zeros(c2,3);
for i=1:1:c2
    for j=1:1:3
        for k=8:1:24

ww=x2(((i-1)*72+(j-1)*24+k),20)+x2(((i-1)*72+(j-1)*24+k-1),20)+x2(((i-1)*72+(j-1)*24+k-1),
20)+x2(((i-1)*72+(j-1)*24+k-1),20)+x2(((i-1)*72+(j-1)*24+k-1),20)+x2(((i-1)*72+(j-1)*24+k-1
),20)+x2(((i-1)*72+(j-1)*24+k-1),20)+x2(((i-1)*72+(j-1)*24+k-1),20);
            if cc(i,j)<=ww
                cc(i,j)=ww;
            end
        end
    end
end
cc=cc./8;
shuru1(:,20)=cc(:,1);

```

```

shuru1(:,20+b)=cc(:,2);
shuru1(:,20+b+b)=cc(:,3);
[w1,h1]=size(shuru1);
[w2,h2]=size(shuru2);
[w3,h3]=size(shuchu);
SRtrainx=[shuru1(w1-302:1:w1-3,:) shuru2(w2-302:1:w2-3,:)];
SRtextx=[shuru1(w1,:) shuru2(w2,:)];
w3=w3-3;
SCtrainy=[shuchu(w3-301:1:w3-2,1:1:6) shuchu(w3-300:1:w3-1,1:1:6)
shuchu(w3-299:1:w3,1:1:6)];

```

附录 3：第二问：数据分析代码，以对逐小时实测数据分析为例

数据预处理

for indexs in df_D2_h.index:

for j in range(2,10):

if df_D2_h.loc[indexs].values[j] == '—':

df_D2_h.iloc[indexs,j] = df_D2_h.loc[indexs-1].values[j]

df_D2_h.dtypes

df_D2_h2=df_D2_h[["SO2 监测浓度($\mu\text{g}/\text{m}^3$)","NO2 监测浓度($\mu\text{g}/\text{m}^3$)","PM10 监测浓度($\mu\text{g}/\text{m}^3$)","PM2.5 监测浓度($\mu\text{g}/\text{m}^3$)","O3 监测浓度($\mu\text{g}/\text{m}^3$)","CO 监测浓度(mg/m^3)","温度(°C)"]].astype(dtype='float')

df_D2_h2.insert(0, '时', 0, allow_duplicates=False)

df_D2_h['时']=pd.to_datetime(df_D2_h['监测时间']).dt.hour

df_D2_h2.insert(0, '时', 0, allow_duplicates=False)

df_D2_h2['时']=df_D2_h['时']

df_D2_h2.insert(0, '温度', 0, allow_duplicates=False)

df_D2_h2['温度']=df_D2_h2['温度(°C)']

out_df_D2_h = df_D2_h2.groupby('时').agg(

SO2 = pd.NamedAgg(column="SO2 监测浓度($\mu\text{g}/\text{m}^3$)", aggfunc="mean"),

NO2 = pd.NamedAgg(column="NO2 监测浓度($\mu\text{g}/\text{m}^3$)", aggfunc="mean"),

PM10 = pd.NamedAgg(column="PM10 监测浓度($\mu\text{g}/\text{m}^3$)", aggfunc="mean"),

PM25 = pd.NamedAgg(column="PM2.5 监测浓度($\mu\text{g}/\text{m}^3$)", aggfunc="mean"),

O3 = pd.NamedAgg(column="O3 监测浓度($\mu\text{g}/\text{m}^3$)", aggfunc="mean"),

CO = pd.NamedAgg(column="CO 监测浓度(mg/m^3)", aggfunc="mean"),

wendu = pd.NamedAgg(column="温度", aggfunc="mean"),

)

#画图

out_df_D2_h[['PM10','PM25',"NO2"]].plot(kind='bar')

plt.savefig('小时'+'.jpg', dpi = 400 , bbox_inches = 'tight')

out_df_D2_h[['O3']].plot(kind='bar')

plt.savefig('小时 O3'+'.jpg', dpi = 400 , bbox_inches = 'tight')

out_df_D2_h[['CO']].plot(kind='bar')

```

plt.savefig('小时 CO'+'.jpg', dpi = 400 , bbox_inches = 'tight')
out_df_D2_h[['PM25']].plot(kind='bar')
plt.savefig('小时 PM25'+'.jpg', dpi = 400 , bbox_inches = 'tight')
out_df_D2_h[['PM10']].plot(kind='bar')
plt.savefig('小时 PM10'+'.jpg', dpi = 400 , bbox_inches = 'tight')
out_df_D2_h[['NO2']].plot(kind='bar')
plt.savefig('小时 NO2'+'.jpg', dpi = 400 , bbox_inches = 'tight')
out_df_D2_h[['SO2']].plot(kind='bar')
plt.savefig('小时 SO2'+'.jpg', dpi = 400 , bbox_inches = 'tight')
out_df_D2_h.plot(kind='area')
plt.savefig('小时 2'+'.jpg', dpi = 400 , bbox_inches = 'tight')

```

附录 4: Fuzzy-BLS 主函数^[12]

```

%load xinCSRtrainx.mat;
%load xinCSCtrainy.mat;
%load xinCSRtextx.mat;
%SR=[SRtrainx;SRtextx]';
%SRR=mapminmax(SR, 0, 1);
%x=SRR';
%SC=SCtrainy';
%[SCC,PS]=mapminmax(SC, 0, 1);
SR=[Untitled;S2]';
[SRR,PPS]=mapminmax(SR, 0, 1);
x=SRR';
SC=S1';
[SCC,PS]=mapminmax(SC, 0, 1);
y=SCC';
n=randperm(900);
train_x=x(n(1:720),:);
train_y=y(n(1:720),:);
test_x= x(n(721:900),:);
test_y= y(n(721:900),:);
Cc = 2^-30;
ss = .8;
be = 0.720;
NR=200
NF=100
NE=1000
rand('state',1)
for i=1:NF
    alp=rand(size(train_x,2),NR);
    Alp{i}=alp;
end

```

```

WE=rand(NF*NR+1,NE);
[NT,NTe,Training_time,Testing_time,TrainingAccuracy]
train(train_x,train_y,test_x,Alp,WE,ss,Cc,NR,NF);
zc=mapminmax('reverse',NTe',PS);
zz=abs(zc)';
zcy=mapminmax('reverse',NT',PS);
zzy=abs(zcy)';
testy=mapminmax('reverse',test_y',PS)';
trainy=mapminmax('reverse',train_y',PS)';
ty=testy';
ry=trainy';

%结果分析
testaqi=zeros(180,3);
trainaqi=zeros(720,3);
testaqick=zeros(180,3);
trainaqick=zeros(720,3);
errtest=zeros(180,3);
errtrain=zeros(720,3);
for i=1:1:3
testaqi(:,i)=aqi(ceil(zz(:,(i-1)*6+1:1:(i-1)*6+6))),);
testaqick(:,i)=aqi(ceil(testy(:,(i-1)*6+1:1:(i-1)*6+6))),);
trainaqi(:,i)=aqi(ceil(zzy(:,(i-1)*6+1:1:(i-1)*6+6))),);
trainaqick(:,i)=aqi(ceil(trainy(:,(i-1)*6+1:1:(i-1)*6+6))),);
end

for i=1:1:3
    for j=1:1:180
        if find(testaqi(j,i)>=300)~=0
            testaqi(j,i)=0;
        end
        if find(testaqick(j,i)>=300)~=0
            testaqick(j,i)=0;
        end
    end
end
for i=1:1:3
    for j=1:1:720
        if find(trainaqi(j,i)>=300)~=0
            trainaqi(j,i)=0;
        end
        if find(trainaqick(j,i)>=300)~=0
            trainaqick(j,i)=0;
        end
    end
end

```

```

end
end
for i=1:1:3
errtest(:,i)=abs(testaqick(:,i)-testaqi(:,i))./testaqick(:,i);
errtrain(:,i)=abs(trainaqick(:,i)-trainaqi(:,i))./trainaqick(:,i);
end
t=1:1:180;
tt=1:1:720;
subplot(2,2,1);
plot(tt,trainaqi(:,1),'g',tt,trainaqi(:,2),'m',tt,trainaqi(:,3),'b');
legend('监测点 A','监测点 B','监测点 C');
title('训练数据结果');
ylabel('AQI')
subplot(2,2,2);
plot(t,testaqi(:,1),'g',t,testaqi(:,2),'m',t,testaqi(:,3),'b');
legend('监测点 A','监测点 B','监测点 C');
title('测试数据结果');
ylabel('AQI')
subplot(2,2,3);
plot(tt,errtrain(:,1),'g',tt,errtrain(:,2),'m',tt,errtrain(:,3),'b');
legend('监测点 A','监测点 B','监测点 C');
title('训练数据结果误差');
ylabel('相对误差')
subplot(2,2,4);
plot(t,errtest(:,1),'g',t,errtest(:,2),'m',t,errtest(:,3),'b');
legend('监测点 A','监测点 B','监测点 C');
title('测试数据结果误差');
ylabel('相对误差')

```

附录 5：第三问微分方程组求解（带有时变项的微分方程）：

```

h=0.01;%所取时间点间隔
ts=[0:h:23];%时间区间
x0=[29.8,36.4];%初始条件
opt=odeset('reltol',1e-6,'abstol',1e-9);%相对误差 1e-6，绝对误差 1e-9
[t,x]=ode45(@fun,ts,x0,opt);%使用 5 级 4 阶龙格—库塔公式计算
figure
% plot([0:23],O3);hold on
% plot([0:23],NO2)
plot(O3,'r');;hold on
plot(NO2,'b');
plot(t,x(:,1),'r',t,x(:,2),'b','LineWidth',2),grid;
legend('O3 平均预测值','NO2 平均预测值','O3 平均实际值','NO2 平均实际值')
figure

```



```
plot(O3,NO2);hold on
```

```
plot(x(:,1),x(:,2),'LineWidth',2),grid %作相轨线
```

```
legend('实际相图','预测相图')
```

```
function dx=fun(t,x,r1,r2,n1,n2,s1,s2)
```

```
%%
```

```
NO2 = [];
```

```
O3 = [];
```

```
duanbo =
```

```
0];
```

```
NO2up = zeros(1,24);
```

```
for i = 1:23
```

```
NO2up(i) = NO2(i+1)-NO2(i);
```

```
end
```

```
NO2up(24) = NO2(24)-NO2(1);
```

```
O3up = zeros(1,24);
```

```
for i = 1:23
```

```
O3up(i) = O3(i+1)-O3(i);
```

```
end
```

```
O3up(24) = O3(24)-O3(1);
```

附录 6：第四问传播方程求解（以 20110713 数据为例）：

```
q = 5;
```

```
x0=1000;
```

```
x=[1:2:6033];
```

```
sigma = 10;
```

```
u = 6;
```

```
H = 10;
```

```
% theta = 0;
```

```
% C = sqrt(2/pi)*q./(u*sigma*pi*(x0+x))./8*exp(-H^2./(2*sigma^2)).*cos(theta);
```

```
Ctheta=[];
```

```
for theta = -pi/2:pi/200:pi/2
```

```
    C = sqrt(2/pi)*q./(u*sigma*pi*(x0+x))./8*exp(-H^2./(2*sigma^2)).*cos(abs(theta));
```

```
    Ctheta =[Ctheta,C'];
```

```
end
```

```
% plot(x,C);u'g
```

```
figure
```

```
mesh(x,-pi/2:pi/200:pi/2,Ctheta');
```

```
xlabel('距离')
```

```
ylabel('角度')
```

```
zlabel('浓度\mug')
```

```
%, '角度','浓度 ug'
```

```
hold on
scatter3(6033,pi/6,max(Ctheta(3017,:)), 'k')
title('A3 SO2')
figure
mesh(x,-pi/2:pi/200:pi/2,Ctheta');
xlabel('距离')
ylabel('角度')
zlabel('浓度\mug')
%,'角度','浓度 ug'
hold on
scatter3(6033,pi/6,max(Ctheta(3017,:)), 'k')
view(141,37)
title('A3 SO2')
```

公众号关注：建模忠哥
获取更多资源