



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

No.21104870023

1.

2.

3.

中国研究生数学建模竞赛论文标题

WRF-CMAQ

WRF-CMAO

□□□□□□□□□□□□□□□□ AQI □□□□□□□□□□□□ AQI □□□□□□□□
 □ Python □□□ pandas □□□□□□□□□□□□□□□□□□□□□□□□ 2020 □ 8
 □ 25 □□ 2020 □ 8 □ 28 □□□□□□ AOI □□□□□□□□□□□□□□□□□□

Python 1 “A
”
AQI
AP AP
SPSS AQI Pearson
Critic AP
AP AP AP
Pearson AP
5

□ □ □ □ □ □ □ □ □ □ □ □ □ □ □

[illegible][illegible]

□ □ □ □ □ □

1

目录

1. 问题重述	3
1.1 问题背景	3
1.2 问题提出	4
2. 背景阐述、模型合理假设以及符号说明	5
2.1 背景阐述	5
2.2 模型合理假设	5
2.3 符号说明	6
3. 问题的分析	6
3.1 问题一：通过给定污染物数据计算当日的 AQI 指数	6
3.1.1 AQI 指数计算方式与相关背景	6
3.1.2 快速求解 AQI 指数算法及对应结果	7
3.2 问题二根据对污染物浓度的影响程度，对气象条件进行分类	7
3.2.1 问题分析与数据预处理	7
3.2.2 模型建立、求解与验证	9
3.3 问题三依据一次预报数据和实测数据进行气象二次预报	11
3.3.1 问题描述和分析	11
3.3.2 模型建立与求解	11
3.4 问题四 xxx	12
3.4.1 问题描述和分析	12
3.4.2 模型建立与求解	12
4. 模型的评价	12
4.1 模型的优点	12
4.2 模型的缺点	12
5. 写作参考格式	13
6. 参考文献	13
附录 A 程序代码	14

1. 问题重述

1.1 ☐ ☐ ☐ ☐

WRF-CMAQ “ ”
WRF
WRF CMAQ 1-1 1-2^[1]

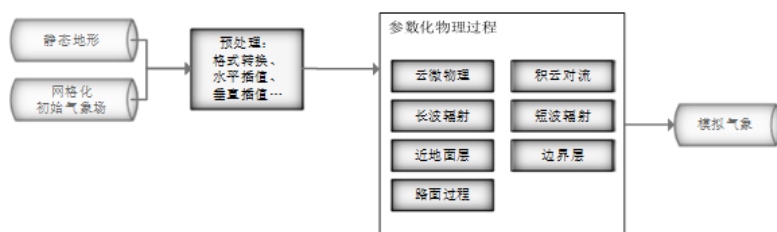
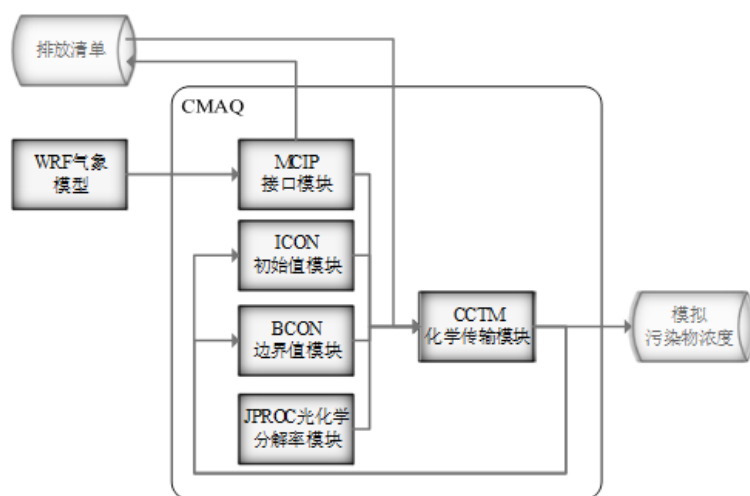


图 1-1 中尺度数值天气预报系统 WRF 结构^[2]

图 1-2 空气质量预测与评估系统 CMAQ 结构^[3]

The diagram illustrates the nesting of four nested boxes, each representing a different model domain. The boxes are labeled as follows from outermost to innermost:

- CMAQ**: The outermost box, representing the largest domain.
- WRF-CMAQ**: The second box from the outside, representing a nested domain.
- WRF-CMAQ**: The third box from the outside, representing a further nested domain.
- WRF-CMAQ 1-3**: The innermost box, representing the smallest domain.

1.2 □ □ □ □

□□ 1 □□□□ 1 □□□□□□□□ A □ 2020 □ 8 □ 25 □□ 8 □ 28 □□□□□□ AQI
□□□□□□□□□□□□□□□□ 1 □□□□□□□□

The diagram illustrates the chemical reaction cycle in the atmosphere. At the top, $\text{RCHO}, \text{O}_3, \text{HONO}$ etc. are shown with a downward arrow labeled $h\nu$ pointing to the OH (RO) box. On the left, a cycle between VOCs and VOCs' is connected to the OH (RO) and $\text{HO}_2 (\text{RO}_2)$ boxes, labeled "自由基循环 '慢循环'" (Free radical cycle 'slow cycle'). In the center, a cycle between OH (RO) and $\text{HO}_2 (\text{RO}_2)$ is shown. To the right, a cycle between NO and NO_2 is connected to the $\text{HO}_2 (\text{RO}_2)$ and O_3 boxes, labeled "氮氧化物循环 '快循环'" (Nitrogen oxide cycle 'fast cycle'). An arrow labeled $h\nu$ points from O_3 to NO_2 . At the bottom, a downward arrow points from the $\text{HO}_2 (\text{RO}_2)$ box to $\text{H}_2\text{O}_2, \text{HNO}_3, \text{ROOH}$ etc.

注: VOCs' 代表区别于作为反应物 VOCs 的另一些挥发性有机物。

表 5 空气质量等级及对应空气质量指数（AQI）范围

优	良	轻度污染	中度污染	重度污染	严重污染	特严重污染
AQI 0~50	AQI 51~100	AQI 101~150	AQI 151~200	AQI 201~300	AQI 301~400	AQI 401~500

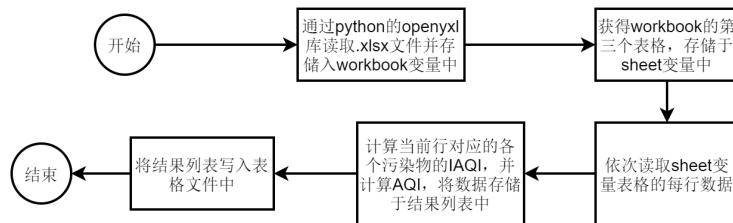


图 3-1 完全处理“监测点 A 逐日污染物浓度数据”表流程图

3.1.2 空气质量指数 AQI 计算结果表

空气质量指数 AQI 是根据空气中的主要污染物的浓度，按照一定的算法计算出来的。AQI 的取值范围是 0 到 500，其中 0 表示空气质量最好，500 表示空气质量最差。AQI 的计算公式如下：

$$AQI = \frac{I_H - I_L}{Q_H - Q_L} \times (Q - Q_L) + I_L$$

其中， I_H 和 I_L 分别表示 AQI 的上下限值， Q_H 和 Q_L 分别表示 AQI 的上下限对应的污染物浓度， Q 表示当前污染物的浓度， I_L 表示 AQI 的下限值。

表 6 问题一 AQI 计算结果表

日期	地点	AQI 值	
		AQI	污染物
2020/8/25	监测点 A	60	O ₃
2020/8/26	监测点 A	46	PM ₁₀
2020/8/27	监测点 A	109	O ₃
2020/8/28	监测点 A	138	O ₃

3.2 空气质量指数 AQI 计算结果表

3.2.1 空气质量指数 AQI 计算结果表

















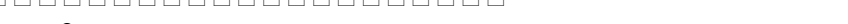







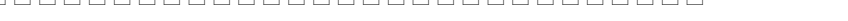


























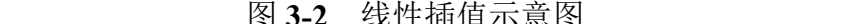
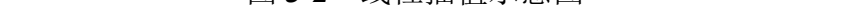

空气质量指数 AQI 是根据空气中的主要污染物的浓度，按照一定的算法计算出来的。AQI 的取值范围是 0 到 500，其中 0 表示空气质量最好，500 表示空气质量最差。AQI 的计算公式如下：

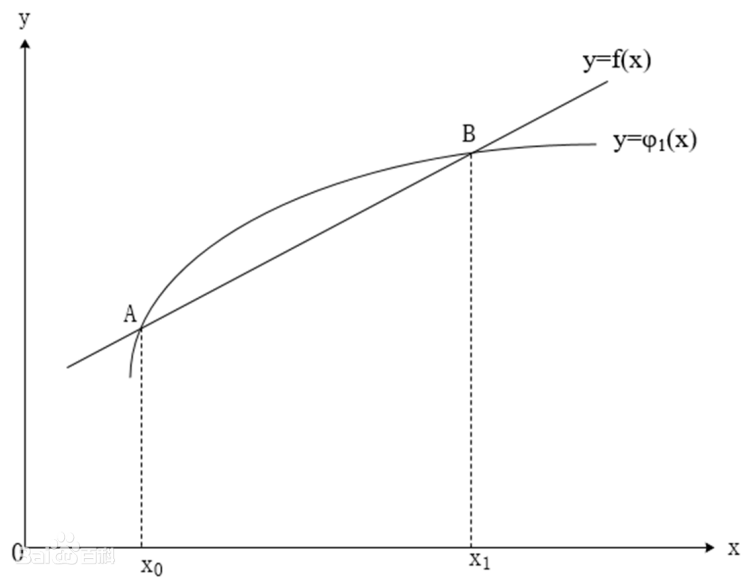
$$AQI = \frac{I_H - I_L}{Q_H - Q_L} \times (Q - Q_L) + I_L$$

其中， I_H 和 I_L 分别表示 AQI 的上下限值， Q_H 和 Q_L 分别表示 AQI 的上下限对应的污染物浓度， Q 表示当前污染物的浓度， I_L 表示 AQI 的下限值。

根据表 6 中的数据，我们可以计算出 AQI 的计算结果表如下：

日期	地点	AQI	污染物
2020/8/25	监测点 A	60	O ₃
2020/8/26	监测点 A	46	PM ₁₀
2020/8/27	监测点 A	109	O ₃
2020/8/28	监测点 A	138	O ₃

-  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**  **AP**

[illegible]

3-2 $y = f(x)$ x_0 x_1 y_0 y_1

333

--	--	--	--	--	--	--	--

$$y = \varphi_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

2. 计算 AQI 值。AQI 值由下式计算：

$$AQI = \frac{I - I_1}{I_2 - I_1} \times (Q_2 - Q_1) + Q_1$$
 其中， I 为待测污染物的分指数， I_1 和 I_2 为 I 所在区间的上下限值， Q_1 和 Q_2 为 I_1 和 I_2 对应的 AQI 值。

$$Z X_i = \frac{X_i - \mu}{\delta}, \mu = \frac{1}{n} \sum_{i=1}^n X_i, \delta = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \mu)^2} \quad (3)$$

3. 计算 AQI, SO₃, NO₂, O₃, PM₁₀, PM_{2.5}, CO 的 Z 值。ZAQI, ZSO₃, ZNO₂, ZO₃, ZPM₁₀, ZPM_{2.5}, ZCO, ZT, ZH, ZP, ZWS, ZWD

3.2.2 计算 Pearson 相关系数

1. Person 相关系数。Person 相关系数由下式计算：

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (4)$$
 其中， N 为样本容量， x_i 和 y_i 为第 i 个样本的 X 和 Y 值。

$$r = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - (\sum x_i)^2} \sqrt{N \sum y_i^2 - (\sum y_i)^2}} \quad (4)$$

Pearson 相关系数 $r=0$ 表示 x 与 y 不相关， $r=1$ 表示完全正相关， $r=-1$ 表示完全负相关。
 ① 用 Python 或 SPSS 计算 ZAQI, ZNO₂, ZSO₃, ZO₃, ZCO, ZPM₁₀, ZPM_{2.5} 与 Person 的相关系数。
 ② 用 SPSS 计算 Excel 数据。
 ③ 用 SPSS 计算 \rightarrow 计算 \rightarrow 计算 AQI 12 个 Pearson 相关系数。
 ④ 用 SPSS 计算 \rightarrow 计算 ZAQI, ZCO 6 个 Pearson 相关系数。

3-3. 计算 Critic 权重。AP 权重。AHP 权重。Critic 权重。Critic 权重。

$$\sum_{i=1}^n (1 - r_{ij})$$
 其中， r_{ij} 为第 i 个样本的 x_i 与第 j 个样本的 y_j 的 Pearson 相关系数。

$$C_j = \delta_j \sum_{i=1}^n 1 - r_{ij}$$
 其中， n 为样本容量， C_j 为第 j 个样本的 Critic 权重。

$$C_j = \delta_j \sum_{i=1}^n 1 - r_{ij}$$

4. 计算 n 个样本的 Critic 权重 C_j 。

		相关性						
		ZAQI	ZSO2	ZNO2	ZPM10	ZPM2.5	ZO3	ZCO
ZAQI	皮尔逊相关性	1	.447**	.574**	.757**	.747**	.500**	.502**
	显著性（双尾）		.000	.000	.000	.000	.000	.000
	个案数	19388	19388	19388	19388	19388	19388	19388
ZSO2	皮尔逊相关性	.447**	1	.409**	.561**	.542**	.182**	.350**
	显著性（双尾）	.000		.000	.000	.000	<.001	.000
	个案数	19388	19388	19388	19388	19388	19388	19388
ZNO2	皮尔逊相关性	.574**	.409**	1	.668**	.669**	-.298**	.681**
	显著性（双尾）	.000	.000		.000	.000	.000	.000
	个案数	19388	19388	19388	19388	19388	19388	19388
ZPM10	皮尔逊相关性	.757**	.561**	.668**	1	.883**	.237**	.609**
	显著性（双尾）	.000	.000	.000		.000	<.001	.000
	个案数	19388	19388	19388	19388	19388	19388	19388
ZPM2.5	皮尔逊相关性	.747**	.542**	.669**	.883**	1	.203**	.661**
	显著性（双尾）	.000	.000	.000	.000		<.001	.000
	个案数	19388	19388	19388	19388	19388	19388	19388
ZO3	皮尔逊相关性	.500**	.182**	-.298**	.237**	.203**	1	-.123**
	显著性（双尾）	.000	<.001	.000	<.001	<.001		<.001
	个案数	19388	19388	19388	19388	19388	19388	19388
ZCO	皮尔逊相关性	.502**	.350**	.681**	.609**	.661**	-.123**	1
	显著性（双尾）	.000	.000	.000	.000	.000	<.001	
	个案数	19388	19388	19388	19388	19388	19388	19388

**在 0.01 级别（双尾），相关性显著。

图 3-3 ZAQI 变量 Person 相关矩阵示意图

$$W_j = \frac{C_j}{\sum_{i=1}^m C_i}$$
 Critic 权重计算结果表

$$AP = \sum_{i=1}^n W_i Z X_i \quad (5)$$

AQI 6 权重计算结果表
 AQI 权重计算结果表
 Python 权重计算结果表

表 7 AP 权重计算结果表

变量	权重	变量	权重	变量	权重
ZSO2	0.127	ZNO2	0.163	ZPM10	0.215
ZPM2.5	0.212	ZO3	0.142	ZCO	0.142

A 权重计算结果表
 AQI 权重计算结果表
 CO 权重计算结果表
 SO2 权重计算结果表
 0.6 权重计算结果表
 0.4 权重计算结果表

$$newW_i = 0.6 * W_i + 0.4 * P_i, P_i = \frac{T_i}{\sum T_i}$$

T_i
 i

表 8 AP 权重计算结果表

ZSO2	0.076	ZNO2	0.260	ZPM10	0.226
ZPM2.5	0.165	ZO3	0.182	ZCO	0.091

AP

$$AP = 0.076 \times ZSO_2 + 0.260 \times ZNO_2 + 0.226 \times ZPM_{10} + 0.165 \times ZPM_{2.5} + 0.182 \times ZO_3 + 0.091 \times ZCO$$

3
 AP
 SPSS
 AP
 AP,GT,GH,GP,GWS,GWD
 3-4

		相关性					
		AP	ZT	ZH	ZP	ZWS	ZWD
AP	皮尔逊相关性	1	-.239**	-.407**	.373**	-.310**	-.050**
	显著性（双尾）		<.001	.000	.000	.000	<.001
	个案数	19388	19388	19388	19388	19388	19388
ZT	皮尔逊相关性	-.239**	1	.119**	-.823**	.087**	.117**
	显著性（双尾）	<.001		<.001	.000	<.001	<.001
	个案数	19388	19388	19388	19388	19388	19388
ZH	皮尔逊相关性	-.407**	.119**	1	-.397**	-.268**	.022**
	显著性（双尾）	.000	<.001		.000	.000	.003
	个案数	19388	19388	19388	19388	19388	19388
ZP	皮尔逊相关性	.373**	-.823**	-.397**	1	-.032**	-.165**
	显著性（双尾）	.000	.000	.000		<.001	<.001
	个案数	19388	19388	19388	19388	19388	19388
ZWS	皮尔逊相关性	-.310**	.087**	-.268**	-.032**	1	-.033**
	显著性（双尾）	.000	<.001	.000	<.001		<.001
	个案数	19388	19388	19388	19388	19388	19388
ZWD	皮尔逊相关性	-.050**	.117**	.022**	-.165**	-.033**	1
	显著性（双尾）	<.001	<.001	.003	<.001	<.001	
	个案数	19388	19388	19388	19388	19388	19388

**, 在 0.01 级别（双尾），相关性显著。

图 3-4 变量 AP 的 Person 相关系数矩阵

3-4
 Pearson
 ZWD
 AP
 AP
 ①
 AP
 WS
 “
 A
 WS
 AP
 WS
 AP
 ??

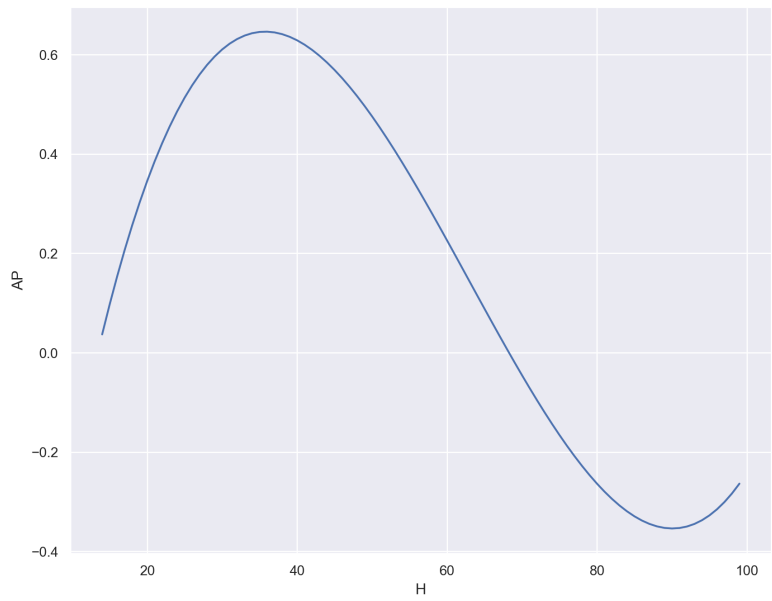


图 3-8 变量 **AP** 与湿度变量 **H** 的拟合曲线

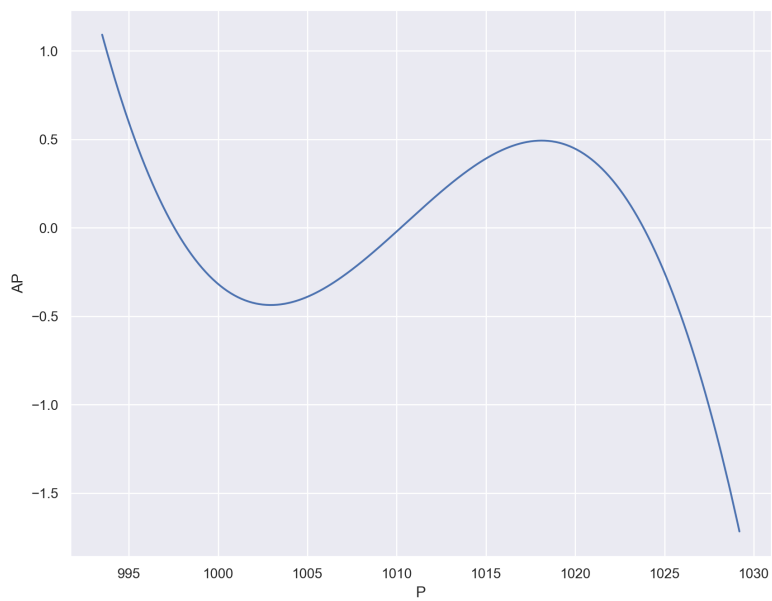


图 3-9 变量 **AP** 与气压变量 **P** 的拟合曲线

□□??□□□□ **WD** □□□□□□ **AP** □□□□□□□□□□ 0.04 □-0.06 □□□
 □□□□□□□□□□ **WD** □□“□□□□□□□□”□□□

3.2.3 □□□□□□

□□□□□□ 5 □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
 □□□□□□□□□□□□□□□□□□□□ **AP** □□□□□□□□□□□□□□□□□□□□
 □□□□□□□□□□□□

□1□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ **AP** □□□□□□□□
 □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□

□2□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□ **P** □□□□□□□□□□□□□□



图 3-10 变量 AP 与风向变量 WD 的拟合曲线

1 AP 数据集的预处理
 2 数据集的预处理
 3 数据集的预处理
 4 AP 数据集的预处理
 5 数据集的预处理

表 9 气象类型分类表

气象类型	气象类型
气象类型	气象类型 WS
气象类型	气象类型 T 气象类型 H 气象类型 P
气象类型	气象类型 WD

3.3 气象类型分类

3.3.1 气象类型分类

1 气象类型分类 (O3) 数据集的预处理
 2 数据集的预处理
 3 数据集的预处理
 4 AQI 数据集的预处理
 5 LSTM 数据集的预处理

3.3.2 气象类型分类

数据集的预处理

- 数据集/数据集 pandas 数据集 pandas DataFrame 数据集 Ex-

Exploratory Data Analysis EDA pandas

-
-

3.3.3

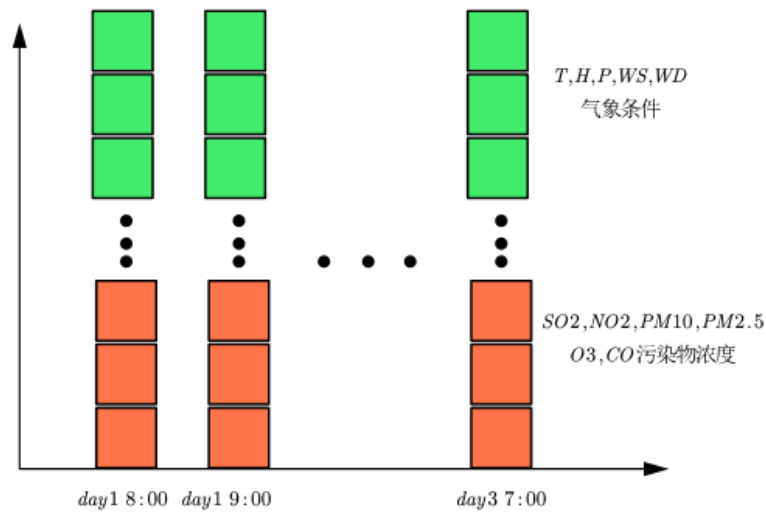


图 3-11 LSTM 模型输入。我们按图中所示的方式把三天（72 小时）的实测数据输入到 LSTM 模型中进行训练。

LSTM 6 + 5 72 11 2020 7 23 23 24 25 11 2020-07-23 LSTM 18 6

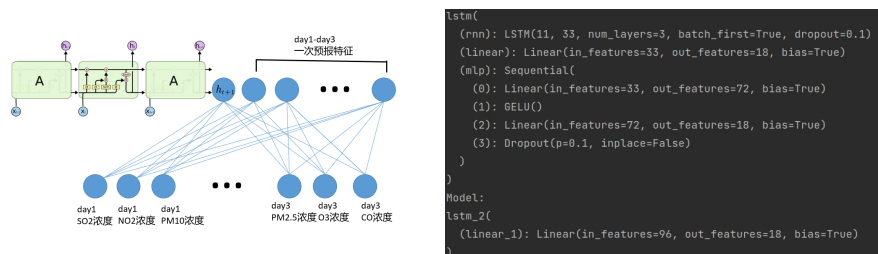


图 3-12 二次预测模型。通过 LSTM 对实测数据进行学习后接入一次模型预测的数据来提升预测准确度。最后直接输出三天的污染物浓度。

表 10 污染物浓度及 AQI 预测结果

日期	地点	污染物浓度 (μg/m³)							
		SO2	NO2	PM10	PM2.5	O3	CO	AQI	污染物
2020/7/13	A	6.06	43.60	41.28	35.92	86.76	0.73	55	NO2
2020/7/14	A	7.90	41.13	53.02	22.39	46.77	0.61	52	NO2&PM10
2020/7/15	A	9.80	42.16	54.02	48.52	48.61	1.10	67	PM2.5
2020/7/13	B	2.09	38.69	92.57	11.58	106.18	0.52	72	PM10
2020/7/14	B	21.52	19.93	41.09	16.18	49.90	0.53	42	PM10
2020/7/15	B	13.48	27.40	64.16	19.99	23.26	0.47	58	PM10
2020/7/13	C	9.03	31.59	72.48	17.95	-1.87	0.82	62	PM10
2020/7/14	C	9.12	40.59	49.47	30.22	49.57	1.03	51	NO2
2020/7/15	C	4.62	16.72	50.68	15.79	20.01	0.80	51	PM10

表 11 各污染物浓度 MSE

污染物	SO2(μg/m³)	NO2(μg/m³)	PM10(μg/m³)	PM2.5(μg/m³)	O3(μg/m³)	CO(μg/m³)
模型	21.41	2221.62	2105.39	1135.74	2728.31	0.32
实测	13.21	584.92	705.37	629.26	1828.01	0.16

3.4 模型性能评价

3.4.1 模型性能评价

模型性能评价主要关注模型的预测精度和稳定性。评价指标包括均方误差 (MSE)、平均绝对误差 (MAE) 和决定系数 (R²)。在模型训练过程中，我们使用交叉验证来评估模型的性能。通过比较不同模型的预测结果，我们可以选择出最优的模型。在模型应用过程中，我们使用实时监测数据来验证模型的预测能力。通过对比模型预测结果与实际监测数据，我们可以发现模型的不足之处，并对其进行优化。

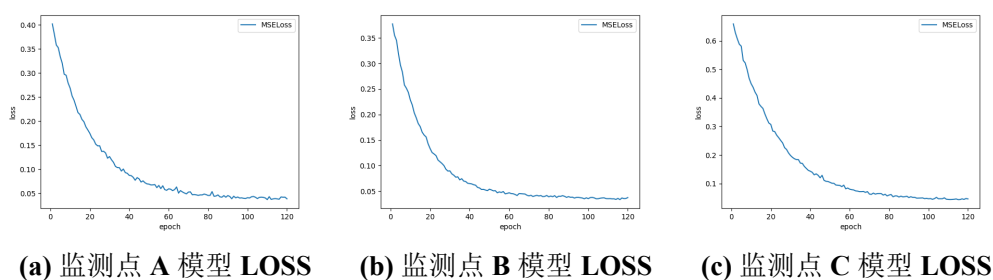
3.4.2 模型性能评价

模型性能评价主要关注模型的预测精度和稳定性。评价指标包括均方误差 (MSE)、平均绝对误差 (MAE) 和决定系数 (R²)。在模型训练过程中，我们使用交叉验证来评估模型的性能。通过比较不同模型的预测结果，我们可以选择出最优的模型。在模型应用过程中，我们使用实时监测数据来验证模型的预测能力。通过对比模型预测结果与实际监测数据，我们可以发现模型的不足之处，并对其进行优化。

4. 模型的评价

4.1 模型性能评价

模型性能评价主要关注模型的预测精度和稳定性。评价指标包括均方误差 (MSE)、平均绝对误差 (MAE) 和决定系数 (R²)。在模型训练过程中，我们使用交叉验证来评估模型的性能。通过比较不同模型的预测结果，我们可以选择出最优的模型。在模型应用过程中，我们使用实时监测数据来验证模型的预测能力。通过对比模型预测结果与实际监测数据，我们可以发现模型的不足之处，并对其进行优化。



□ □ □ □ □ □ □ □ □	AQI	□ □ □ □ □
□ □ □ □ □ A □	1682.42	24%
□ □ □ □ □ A □	897.86	26%
□ □ □ □ □ B □	612.49	39%
□ □ □ □ □ C □	217.69	33%

[illegible]

- [illegible]

[1] □□□, □□□, □□, □. □□□□□□□□□□□□□□□□ [J]. □□□□, 2019, 3.

[2] □□□. □□□□□□SMOKE□WRF□CMAQ □□□□□□□□□□ [M]. □□: □□□□□□□□, 2019.

[3] □□□. □□□□ [M]. □□: □□□□□□□□, 1997.

[4] □□□, □□, □□□. □□□□□□□□□□□□□□□□ [J]. □□□□, 2018, 31(05): 72-76.

[5] □ □ □ . □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ [J/OL] 2018, <https://max.book118.com/html/2018/0201/151478594.shtm>.

附录 A 程序代码

```
kk=2; [mdd, ndd]=size(dd);  
while ~isempty(V)  
    [tmpd, j]=min(W(i, V)); tmpj=V(j);  
    for k=2:ndd  
        [tmp1, jj]=min(dd(1, k)+W(dd(2, k), V));  
        tmp2=V(jj); tt(k-1, :)= [tmp1, tmp2, jj];  
    end  
    tmp=[tmpd, tmpj, j; tt]; [tmp3, tmp4]=min(tmp(:, 1));  
    if tmp3==tmpd, ss(1:2, kk)=[i; tmp(tmp4, 2)];  
    else, tmp5=find(ss(:, tmp4)~=0); tmp6=length(tmp5);  
    if dd(2, tmp4)==ss(tmp6, tmp4)  
        ss(1:tmp6+1, kk)=[ss(tmp5, tmp4); tmp(tmp4, 2)];  
    else, ss(1:3, kk)=[i; dd(2, tmp4); tmp(tmp4, 2)];  
    end; end  
    dd=[dd, [tmp3; tmp(tmp4, 2)]]; V(tmp(tmp4, 3))=[];  
    [mdd, ndd]=size(dd); kk=kk+1;  
end; S=ss; D=dd(1, :);
```