

无锡智慧城市环境监控预警平台

目 录

第一章 概述.....	3
第二章 数据探索和分析.....	4
一 数据来源.....	4
二 数据探索分析.....	4
第三章 可视化展示.....	8
一 实时城市动脉和呼气监控.....	8
二 污染源地理网络.....	9
三 超标企业和污染物监控.....	10
四 环境影响因素分析.....	11
五 实时预测.....	17
第四章 总结.....	25
第五章 展望.....	26
第六章 团队.....	27

第一章 概述

本平台主要面向无锡市政府对城市的动态感知，通过对水和空气质量的监测与分析，了解无锡环境质量的基本指标。通过对历史数据的整理分析和建模，以及实时监控数据的引入，实时监控城市环境质量，了解城市现状，并具备对未来一段时间的环境污染进行预警的能力。

智慧城市环境监测预警平台基于 SAS Viya 云就绪(cloud-ready)开放式综合分析平台搭建，主要使用了 SAS 可视化分析(SAS Visual Analytics)，SAS 可视化建模统计(SAS Visual Statistics)和 SAS 可视化建模统计(SAS Visual Statistics)等模块来探索数据和高级分析建模，同时可接入无锡市环保局和统计局的实时数据 API 进行动态可视化展示。该系统可以提供总体环保状态感知，实时环保指标可视化，环保状态评估和预测的功能，同时也可可为无锡市市政府环保制定，检验，评估，修正提供依据，并能做到进一步探究环境问题成因，其主要实现以下目标:

构建全面的空气水体监测智慧型地理网络，实现无锡市废弃废水等主要污染源的全市尺度的展现。

建设覆盖全市的规划统一的空气水体在线监测系统，实现超标污染物质的及时预警。

利用 SAS Viya 机器学习和数据挖掘网络对历史数据进行高级分析建模，为无锡市政府提供战略尺度上的环保建议。

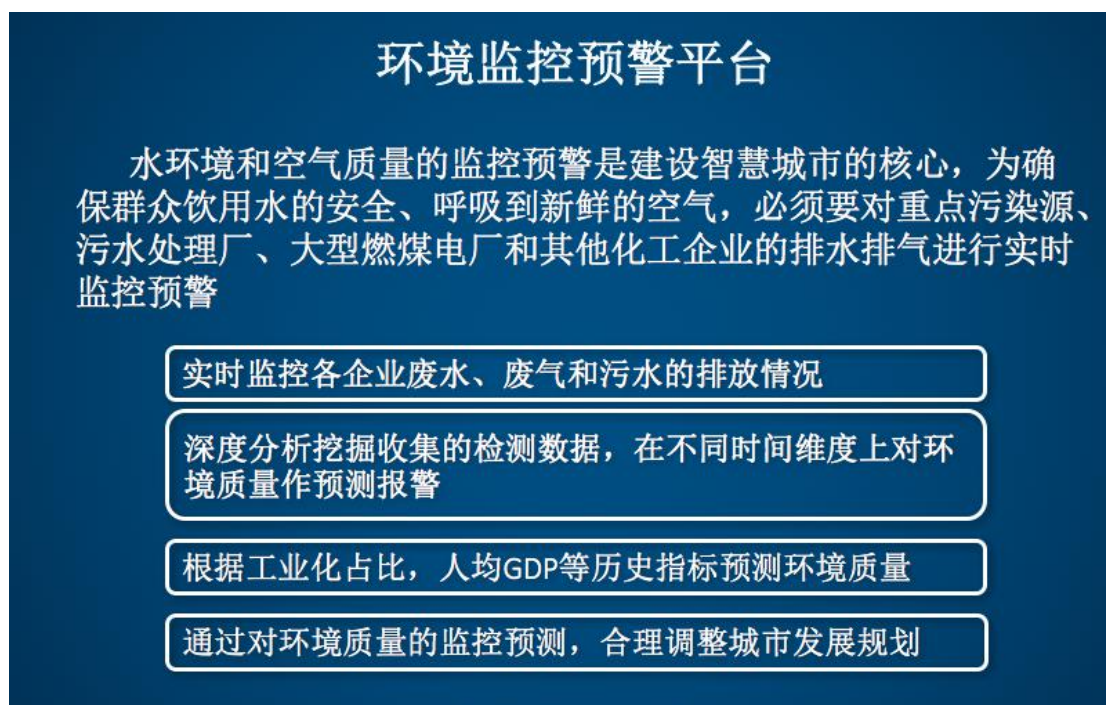


图 1.1 无锡环境监控预警平台

第二章 数据探索和分析

一 数据来源

1、大赛提供数据集

通过此次大赛提供的给定数据集，筛选提取出可用数据，整理成数据集。如：从“无锡市公开数据 2”中，提取出“水环境质量指标、空气环境质量指标、废水排放指标、废气排放指标”等数据。从“无锡市统计年鉴”数据中分析得到无锡市 GDP 产值，人均绿地面积，汽车保有量和市区人口等独立变量。

2、无锡政府官方网站

通过查询无锡政府官方网站上的公示信息，筛选提取出可用数据，整合了无锡市人民政府、无锡市环境保护局和无锡市统计局等相关公开数据，并用 SAS base 整理成“无锡市污染源超标月数据”和“无锡市水体质量数据”等数据集。

3、中国环保局官网

利用 python 从中国环保局官网高效爬取并提取出结构化的无锡市日均空气监测数据，使用 python pandas 和 csv 等库对其进行初步的清洗工作

二 数据探索分析

一个完整的数据分析流程包括：数据准备与取样、数据探索和分析、数据调整、模型研发、模型和知识的解释与评价。

数据准备取样阶段，是要从数据的海洋中提取出与本问题有关系的数据，并且尽量保证提取数据的高质量，以期所提取的数据能够准确高效的反应问题的规律。在此阶段要注意数据的完整性和有效性，并且需要有足够的数据量能够支撑起探索问题的规律。

数据探索分析阶段，是对前一阶段的数据集进行进一步的挖掘和探索，在数据的基础上找出规律，解决问题。在这一阶段，SAS 的可视化分析工具是分析人员的强有力的支撑，通过不同维度不同方法的可视化手段分析数据，免除了复杂的数学运算过程和编程过程，让分析人员能够将思维聚焦在数据呈现的关系上，提高思维集中性和灵敏性。

数据调整阶段是在对数据的初步探索和分析之后，找出了一些数据规律，进一步明确问题细化问题，然后对现有数据进行调整和改进。

明确问题并且有足够数量的高质量数据之后，就可以针对本问题进行数据挖掘模型化分析。对 SAS 来说，这一步有着非常多而广的手段来满足分析人员的需求，并且相应的有足够的可视化工具来展示分析成果。

最后需要对模型进行分析解释和评价，完善模型，丰富经验。

图 2.2.1 展示了 SAS 的数据分析流程，图 2.2.2 展示了本平台的设计架构和数据流向。

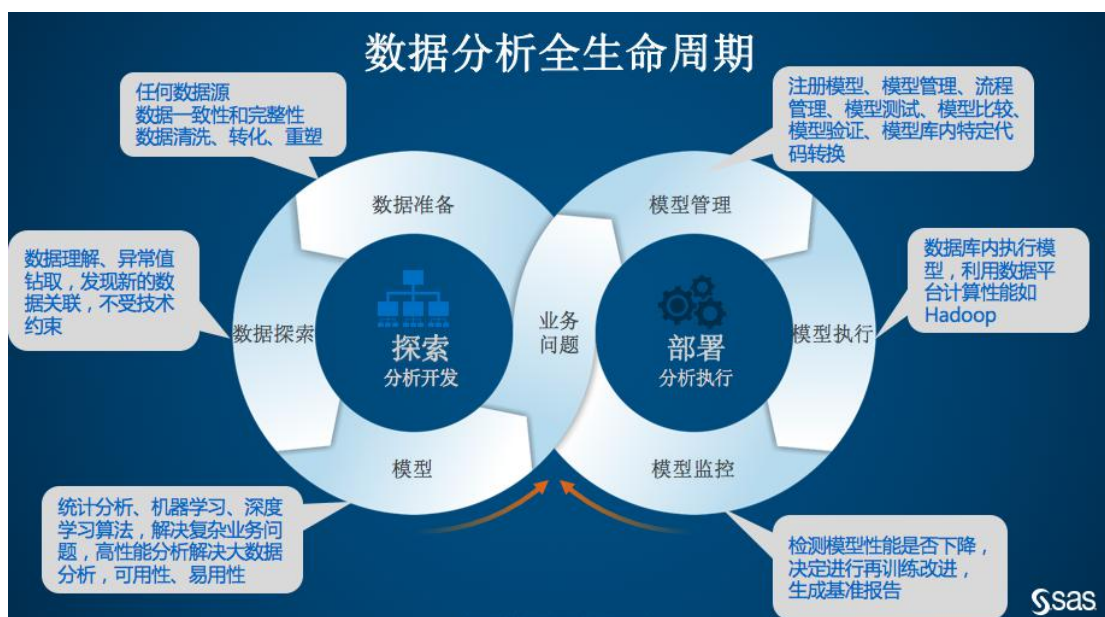


图 2.2.1 数据分析生命周期

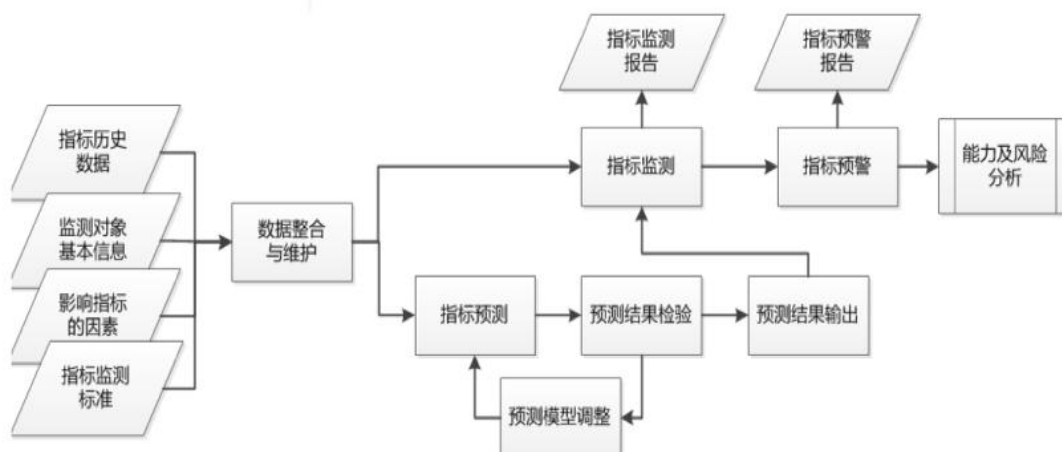


图 2.2.2 表结构和指标设计架构

1.数据准备

使用 SAS base 语言，基于空气综合质量和水体质量等设计指标进行数据字段的映射和翻译；同时设计数据挖掘宽表或者交易表进行数据提取和处理，如一些异常值，空值(miss 值)，错误数值。同时整理已有的数据集，对数据进行单位转化和标准化，完成数据的一致性和完整性。如：对现有数据进行整理，将数据按照日期间隔，以年为单位的数据整合成一张表，以月为单位的数据整合成另一张表。经过整合优化的表结构及实例数据如下：

	序号	日期	行政区域	企业名称	超标因子	排放标准	当月超标最大值	超标率
1	1	2018/05	惠山区	无锡永达...	COD	50	196.81	293.62
2	2	2018/04	江阴市	江阴市碧...	COD	60	63.25	5.4166666667
3	3	2018/03	江阴市	江阴市华...	COD	60	61.07	1.7833333333
4	4	2018/02	江阴市	亚同环保...	COD	60	98.27	63.783333333
5	5	2018/01	江阴市	江阴市华...	COD	76	86.62	13.973684211
6	6	2017/12	江阴市	江阴市华...	COD	76	90.86	19.552631579
7	7	2017/10	江阴市	江阴市碧...	COD	60	71.17	18.616666667
8	8	2017/09	江阴市	亚同环保...	COD	60	116	93.333333333
9	9	2017/09	江阴市	江阴市碧...	COD	60	62.49	4.15
10	10	2017/09	江阴市	江阴港虹...	氨氮	5	5.43	8.6

图 2.2.3 无锡超标企业表结构

	监测日期	企业名称	经度	纬度	污染物类型	value
1	2018/02	盛隆资源再生（无锡）有限公司	120.466876	31.53521	PH 值	14.93
2	2018/02	盛隆资源再生（无锡）有限公司	120.466876	31.53521	氨氮(万吨)	0.676
3	2018/02	盛隆资源再生（无锡）有限公司	120.466876	31.53521	化学需氧量(万吨)	26
4	2018/02	盛隆资源再生（无锡）有限公司	120.466876	31.53521	总氮（mg/L）	4.8
5	2018/02	盛隆资源再生（无锡）有限公司	120.466876	31.53521	总磷（mg/L）	0.041
6	2018/02	无锡华润上华科技有限公司	120.40202	31.521518	PH 值	7.53
7	2018/02	无锡华润上华科技有限公司	120.40202	31.521518	氨氮(万吨)	17.2
8	2018/02	无锡华润上华科技有限公司	120.40202	31.521518	化学需氧量(万吨)	30
9	2018/02	无锡华润上华科技有限公司	120.40202	31.521518	总氮（mg/L）	36
10	2018/02	无锡华润上华科技有限公司	120.40202	31.521518	总磷（mg/L）	1.25

图 2.2.4 无锡废气废水地图网络表结构

	日期	总体水质标准	水体状态	总氮浓度标准	II 类标准河流	III 类标准河流	IV 类标准河流	V 类标准河流
1	2018/06	IV	轻度富营养	IV	1	1	9	2
2	2018/05	IV	轻度富营养	V	2	5	4	2
3	2018/04	IV	轻度富营养	V	1	4	5	3
4	2018/03	V	轻度富营养	V	2	3	5	3
5	2018/02	IV	轻度富营养	V	1	1	8	3
6	2018/01	IV	轻度富营养	V	0	4	3	6
7	2017/12	IV	轻度富营养	IV	0	6	5	2
8	2017/11	IV	轻度富营养	IV	1	8	4	0
9	2017/10	V	轻度富营养	IV	1	5	5	2
10	2017/09	V	轻度富营养	IV	0	3	10	0

图 2.2.5 无锡实时水体监控网络表结构

	date	type	_1188A	_1189A	_1190A	_1191A	_1192A	_1193A	_1194A	_1195A
1	2018/01/01	AQI	272	328	.	305	359	356	315	334
2	2018/01/01	PM2.5	218	240	228	240	253	231	240	255
3	2018/01/01	PM10	288	354	325	339	292	310	305	349
4	2018/01/01	SO2	18	15	21	18	23	15	18	20
5	2018/01/01	NO2	72	78	80	71	94	63	64	49
6	2018/01/01	O3	137	131	129	111	123	118	131	93
7	2018/01/01	CO	2.4	2.7	2.7	2.5	2.8	2.4	2.5	2.9
8	2018/01/02	AQI	68	67	66	70	55	67	67	61
9	2018/01/02	PM2.5	157	169	190	168	180	174	180	159
10	2018/01/02	PM10	251	282	300	293	230	257	264	250

图 2.2.6 无锡实时空气监控网络表结构

	时间	旅游业总收入(亿元)	汽车保有量(万辆)	用电量(亿千瓦时)	人均绿地(平方米)	建成区绿化覆盖率%	GDP(亿元)	人均GDP(万元)	市区人口(万人)	工业产业占比%	总氮(mg/L)	总磷(mg/L)	富营养化指数	二氧化碳(毫克/立方米)
1	2000	116.1	8.91	137.97	7.98	35.67	1201	2.76	434.61	56	6.61	0.199	74.8	0.0106
2	2001	111.5	11.18	158.44	7.82	29.69	1360	3.12	435.9	59.8	2	0.1	72	.
3	2002	150	14.53	198.66	5.84	34	1590	3.68	432.58	58.2	.	.	68	0.057
4	2003	180.51	18.5	238.88	8.2	37.5	1901.22	4.3	442.54	57.4	.	.	65	0.054
5	2004	229.68	23.89	289.7	10	40	2350	5.26	447.19	58.8	.	.	63	0.075
6	2005	301.6	29.39	337.05	10.8	44.1	2805	5.63	452.84	60.2	2	0.1	63	0.069
7	2006	376.72	35.73	333.1	11.34	43.6	3300	5.77	457.8	59.7	2	0.1	65	0.059
8	2007	445.76	43.23	326.75	11.87	43	3858	6.52	461.74	58.4	2	0.1	64	0.065
9	2008	520.3	49.94	349.07	12.54	43.2	4419.5	7.3	464.2	57.6	2	0.1	65	0.06
10	2009	618.5	60.07	364.8	13	43.1	4992	8.11	465.65	56.3	3	0.1	60	0.06

⑪ 二氧化氮 (毫克/立 方米)	⑫ 可吸入颗 粒物(毫 克/立方...	⑬ 综合空气 质量指标	⑭ 工业废水 排放总量 (亿吨)	⑮ 生活污水 排放总量 (亿吨)	⑯ 生活污水 化学需氧 量	⑰ 工业废气 排放总量 (亿标立...	⑱ 二氧化硫 排放总量 (万吨)	⑲ 工业排放 二氧化硫 (万吨)	⑳ 生活污染 源排放二 氧化硫(...	㉑ 烟排放总 量(万吨)	㉒ 工业排放 烟(万吨)	㉓ 生活污染 源排放烟 (万吨)
0.032	0.171	2.5405	1.79	.	.	770.32	8.176	.	.	2.291	.	.
.	0.156	2.08	2.762	1.041	2.463	1157.53	9.62	8.72	0.9	2.8	2.46	0.34
0.04	0.133	2.645833333	3.29	1.23	2.4476	1614.17	10.386	9.582	0.804	2.499	2.174	0.325
0.041	0.1	2.172333333	3.16	0.99	1.82	1412.35	9.72	8.95	0.77	2.43	2.2	0.23
0.042	0.118	2.678833333	4.58	1.7	2.86	2309.75	13.99	13.35	0.64	4.48	4.13	0.35
0.038	0.1	2.347833333	4.9	1.75	4.28	2631	21.4	15	0.6	5.2	4.8	0.36
0.037	0.094	2.138833333	5.375	1.24	4.42	2678	23.57	14.2	0.6	5.59	5.23	0.36
0.038	0.083	2.071166667	4.53	1.86	3.95	3595.67	30.19	11.66	0.53	6.25	5.95	0.3
0.04	0.1	2.243333333	4.33	2.08	3.29	3517.73	10.19	9.6	0.59	4.76	4.45	0.31
0.04	0.1	2.243333333	4.16	2.18	3.51	3727.66	10.48	9.92	0.56	4.44	4.09	0.35

图 2.2.7 无锡空气水体预测预警表结构

2.数据探索

一方面探索数据，分析数据，提取异常值，寻求新的数据关联，另一方面提取样本数据进行剖析，验证指标设计，探索不同的指标对环境污染的影响，分析环境污染可能的原因。

列	唯一	空	空白	模式计数	均值	中位数
⊕ PH 值	95% (23)	11% (3)			10.18	7.67
⚠ 企业名称	% (22)	0	0	12		
⊕ 化学需氧量	91% (22)	11% (3)			94.83	82.50
⊕ 总氮 (mg/l)	100% (22)	18% (5)			16.59	8.67
⊕ 总磷 (mg/l)	% (17)	18% (5)			0.79	0.22
⊕ 氨氮(万吨)	95% (23)	11% (3)	11% (计数: 3)		7.63	1.73
⊕ 监测日期	(...)	0			21,315...	21,326...
⊕ 纬度	% (22)	0			31.58	31.55

图 2.2.8 样本数据剖析

第三章 可视化展示

在完成了数据的整理和初步探索之后，通过 SAS Viya 平台进行数据的分析和可视化展示。

一 实时城市动脉和呼气监控

1、无锡实时动脉监测

水脉是一个城市的动脉，水的质量与城市居民的生活质量息息相关。利用“无锡实时水体监控”数据，通过 VA 分析完成无锡市水体质量历史数据的可视化展示和实时数据的动态监控，截图如图 3.1.1 所示。

展示图的上半部分是无锡市 13 条河流质量的历史数据折线图，横坐标代表时间，纵坐标代表相应类别标准河流数量，不同的颜色代表不同类别标准。我们可以直观而清晰地看到不同类别标准的河流数量的变化曲线。

展示图的下半部分是对河流质量的动态监控，这一部分的可视化产品是动态并且可交互的。随着时间的推移，左侧的河流质量环状图在实时变化，同时右侧的总氮标准和总体水质的数据也会相应的变化。与此同时，我们可以通过点击某一个类别标准河流来突出重点监控该类河流的数量和质量，此时右侧的数据显示的将会是该类河流的质量数据。如图展示的是重点监控 II 类标准河流，此时左侧环状图代表 II 类标准河流的绿色色块会被突出展示，动态效果下可以更轻易地追踪 II 类河流的数量变化，而右侧展示的是此时 II 类河流的水质数据，而不是全市所有河流的总体水质数据。



图 3.1.1 无锡实时动脉监测

2、无锡实时呼气监测

空气质量对人民生活质量的重要性不言而喻。利用“无锡实时空气质量监测”数据，通过 VA 完成无锡市空气质量实时监测的可视化展示，截图如图 3.1.2 显

示。

空气质量实时监测部分是动态并且可交互的，上方地时间条显示出当前时间，随着时间的推移动态展示出空气质量各个指标的数值，如：AQI、PM2.5、二氧化氮浓度、二氧化硫浓度等指标，并且用绿、黄、红直观地显示出该指标是否超标以及超标严重程度。

由于此部分是实时监控，显示数据默认是最新一天的数据，也可以通过调节上方的时间条回看历史数据。

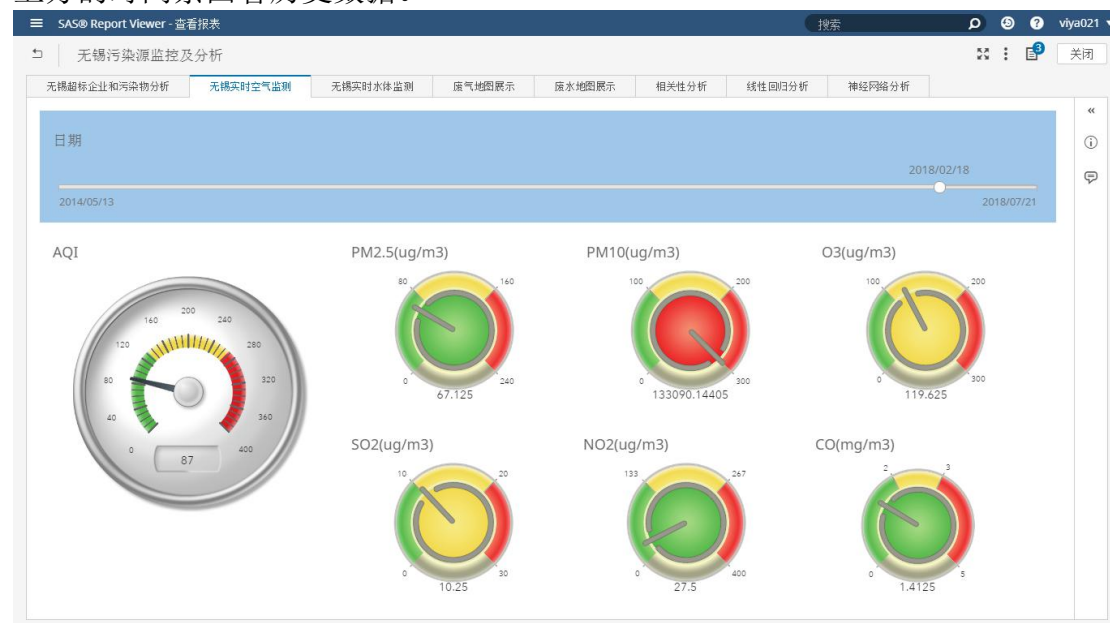


图 3.1.2 无锡实时呼气监测

二 污染源地理网络

各大企业的污染物排放也会对城市的环境质量造成巨大影响。利用“无锡废气废水地图表”数据，根据各个污染物排放企业的经纬度数据，在地图上绘制出各个污染物排放源。图 3.2.1 展示了空气污染物地理监控图，图 3.2.2 展示了水体污染物地理监控图。

污染源地理监控网络是一个动态可交互的展示产品，用不同颜色的色块区分不同的污染物，用户可以自主选择上方的时间区间，直观地查看该时间段内各个影响空气质量或者水体质量的主要污染物排放来源。

通过分析，我们能看到污染物的排放更加集中在哪里，这可以为无锡市接下来的企业整顿、相应企业建址和环境保护重点提供方向性的建议。

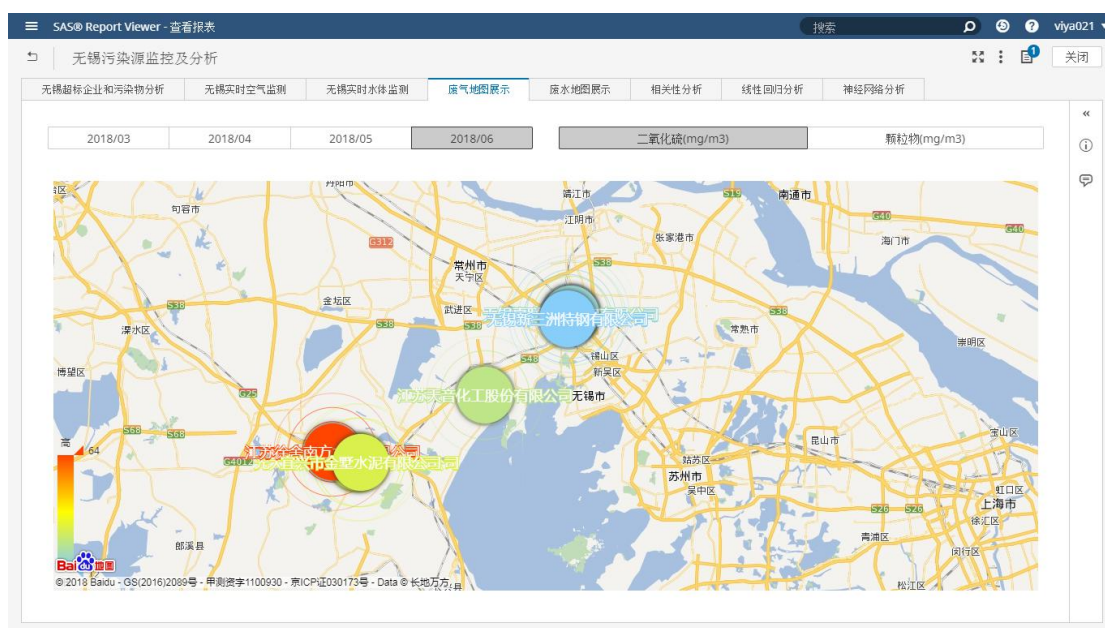


图 3.2.1 无锡空气污染物地理监控网络

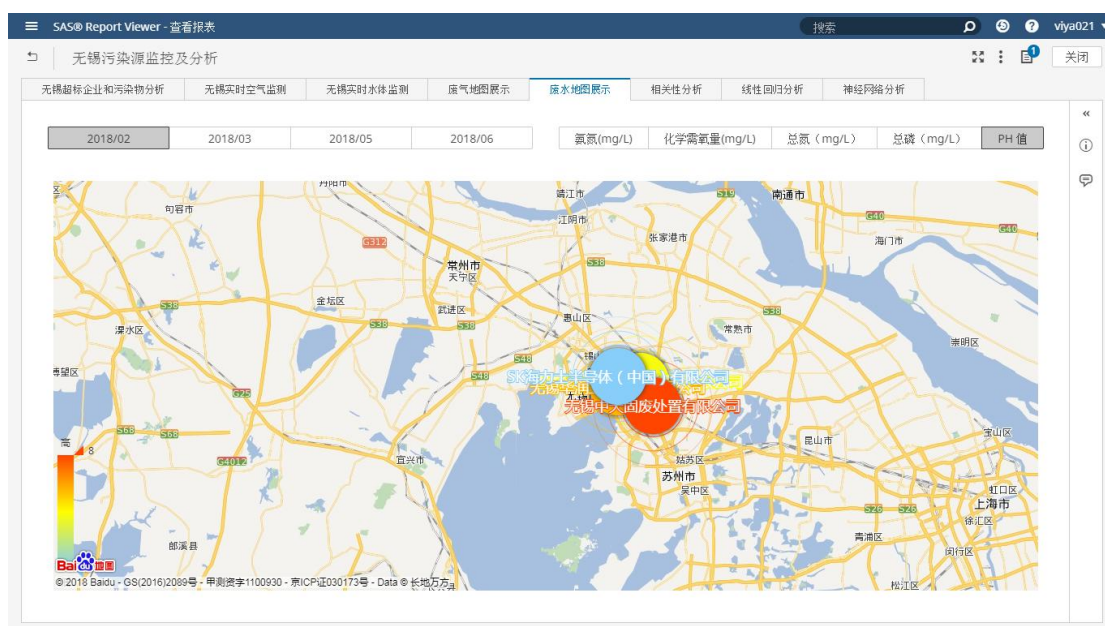


图 3.2.2 无锡水体污染物地理监控网络

三 超标企业和污染物监控

在污染物排放来源数据分析基础上，还需要了解有哪些企业污染物排放超标，其后可以有针对性地对相应企业进行整顿。利用“无锡市污染源超标月数据”，通过 VA 进行数据分析，截图如图 3.3.1 所示。

污染物排放超标企业按照市辖区分类，展示图左侧的环状图展示出各个市区超标企业的频数分析。如截图所示，我们可以清晰地看出江阴市的污染物排放超标企业最多，占据全无锡市超标企业的 45%。展示图右侧上方是超标企业名称的

词云，展示出各个区超标企业的名字。右侧下方是相应的超标污染物的标准值和当月超标值的统计数据。

通过分析，我们能直观的感受各个市区超标企业的频数比，相应的超标公司，以及各个污染物的排放超标情况。

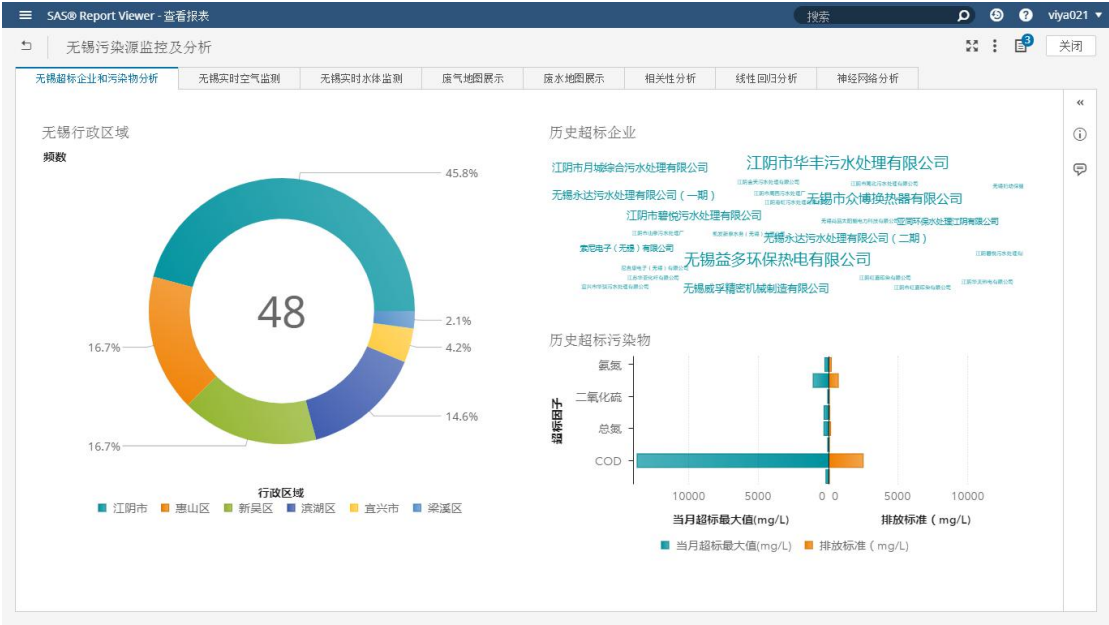


图 3.3.1 超标企业和污染物分析

四 环境影响因素分析

1、多元线性回归

1.1 相关性分析

为了分析变量之间的关系，找出影响综合空气质量指标和水体富营养化程度的关键因素，我们对各个变量之间的相关性进行了分析。相关系数的取值在-1和+1之间，其绝对值越大表示两个变量之间的关联性越强。相关性分析可以通过 SAS 的 PROC CORR 过程步来实现。

在分析之前，为了统一量纲，我们对部分指标数据使用了 z-score 标准化方法进行了标准化处理，即：新数据=（原数据-均值）/标准差。

指标间的相关性如下表所示：

Pearson 相关系数		
Prob > r under H0: Rho=0		
	富营养化指数	综合空气质量指标
旅游业总收入(亿元)	-0.84747	-0.79393
	<.0001	<.0001
汽车保有量(万辆)	-0.84547	-0.79498
	<.0001	<.0001
用电量(亿千瓦)	-0.90148	-0.72868
	<.0001	0.0006

人均绿地(平方)	-0.88859	-0.65591
	<.0001	0.0031
建成区绿化覆盖率%	-0.75951	-0.36059
	0.0003	0.1416
GDP(亿元)	-0.88756	-0.75984
	<.0001	0.0003
人均 GDP(万)	-0.89866	-0.7557
	<.0001	0.0003
市区人口(万)	-0.88631	-0.77764
	<.0001	0.0001
工业产业占比%	0.7078	0.74007
	0.001	0.0004
工业废水排放总量(亿吨)	0.09521	0.26361
	0.7258	0.3239
生活污水排放总量(亿吨)	-0.82567	-0.37381
	0.0001	0.1699
工业排放烟(万吨)	-0.66926	-0.57428
	0.0064	0.0252
生活污染源排放烟(万吨)	0.74273	0.56843
	0.0015	0.027

表 3.4.1 指标相关性矩阵

从相关性矩阵中可以看出，富营养化指数与旅游业总收入、人均绿地面积、建成区绿化覆盖率等指标存在很强的负相关性，与工业产业占比、生活污染源排放烟存在很强的正相关性。特别注意的是富营养化指数与工业废水不存在相关性，这是因为工业废水中含有较少的氮磷等有机元素，另外富营养化与生活污水呈现出很强的负相关性，可能是由于近几年污水处理系统的改造升级造成的，虽然生活污水的排放越来越多，但是对水污染的影响反而越来越小。

分析综合空气质量指标与其他解释变量的相关性可以发现，空气质量指标与工业产业占比具有最强的正相关性且显著，说明工业占比越高，排放的废气越多，对空气造成的影响越恶劣。

部分指标间的散点图如下所示：

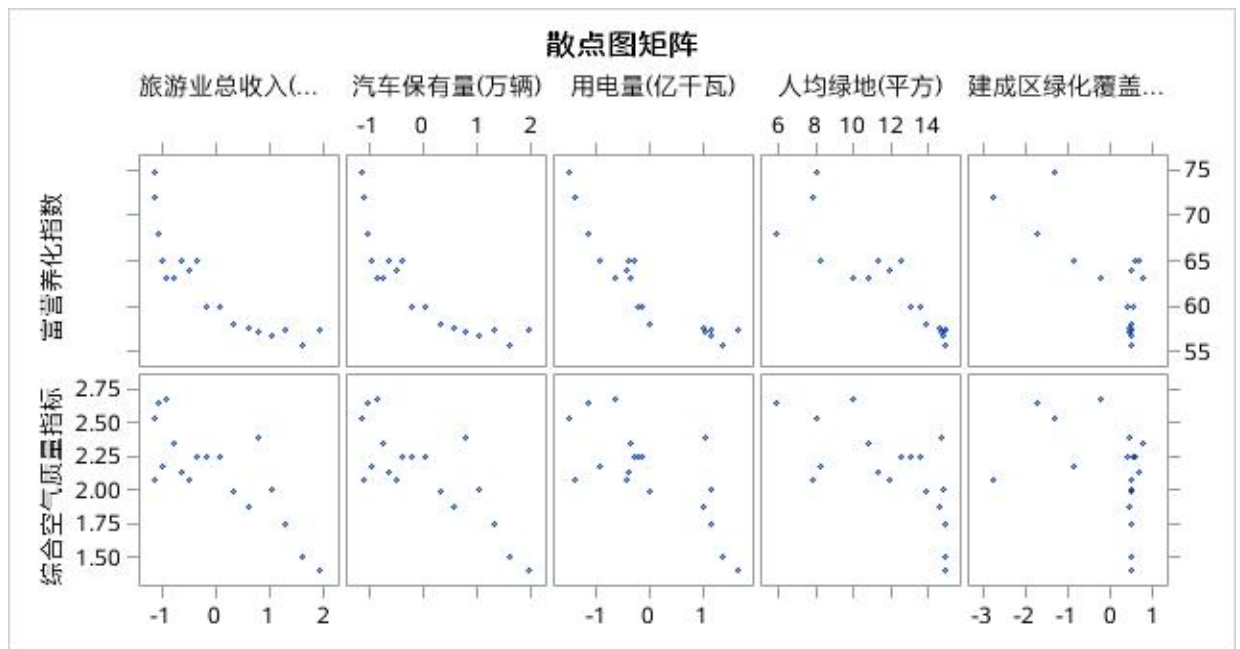


图 3.4.2 散点图矩阵

相关性矩阵和散点图同样可以在 SAS Viya 中展示，如下图所示，颜色越深代表这两个指标间的相关性越强，白色代表这两个指标间无相关性。



图 3.4.3 设计指标相关性分析

1.2 富营养化指数影响因素分析

以富营养化指数为因变量，其他指标为自变量，建立多元线性回归模型，即

$$Y = X\beta + \varepsilon$$

引入所有变量的回归结果如下所示：

参数估计						
变量	自由度	参数估计	标准误差	t 值	Pr > t	方差膨胀

Intercept	1	104.1836	40.59301	2.57	0.2365	0
人均绿地	1	-3.28714	2.88996	-1.14	0.4591	430.1769
旅游业总收入	1	80.08288	105.66092	0.76	0.5871	46022
汽车保有量	1	-62.01433	83.59774	-0.74	0.5937	28351
用电量	1	-1.15835	3.74448	-0.31	0.809	62.96094
GDP	1	-40.66561	49.93819	-0.81	0.5649	11494
建成区绿化覆盖率	1	-6.68258	3.63471	-1.84	0.3171	88.76846
人均 GDP	1	24.4739	37.52902	0.65	0.6321	6197.85478
工业排放烟	1	-1.34226	1.21011	-1.11	0.4671	41.2348
生活污染源排放烟	1	-25.5644	24.32101	-1.05	0.4841	76.84136
市区人口	1	20.83677	14.06589	1.48	0.378	853.29472
工业产业占比	1	10.16527	7.0819	1.44	0.3874	209.17003
工业废水排放总量	1	2.30889	3.46161	0.67	0.6255	90.0257
生活污水排放总量	1	0.88077	1.93671	0.45	0.7283	33.52344

表 3.4.4 全变量参数估计

一般从经验上来判断：当 $0 < VIF < 10$ 时（VIF，方差膨胀因子），表示不存在多重共线性；当 $10 \leq VIF < 100$ 时，表示存在较强的多重共线性；当 $VIF \geq 100$ 时，则意味着存在严重多重共线性。从上面的结果中我们可以看出，自变量之间存在严重的多重共线性。因此，我们采用变量逐步选择和人工干预的方式对自变量进行了筛选，最终选取了指标人均绿地面积和工业产业占比作为解释变量，对水体的富营养化指数进行建模，最终模型结果如下所示：

参数估计						
变量	自由度	参数估计	标准误差	t 值	Pr > t	方差膨胀
Intercept	1	73.14465	3.4063	21.47	<.0001	0
人均绿地	1	-0.97043	0.2699	-3.60	0.0037	2.26784
工业产业占比	1	2.05712	0.94848	2.17	0.0509	2.26784

表 3.4.5 水体质量模型参数估计表

方差分析					
源	自由度	平方和	均方	F 值	Pr > F
模型	2	241.56865	120.78432	33.22	<.0001
误差	12	43.63535	3.63628		
校正合计	14	285.204			

表 3.4.6 水体质量模型方差分析表

均方根误差	1.9069	R 方	0.847
因变量均值	62.12	调整 R 方	0.8215
变异系数	3.06971		

表 3.4.7 水体质量模型指标

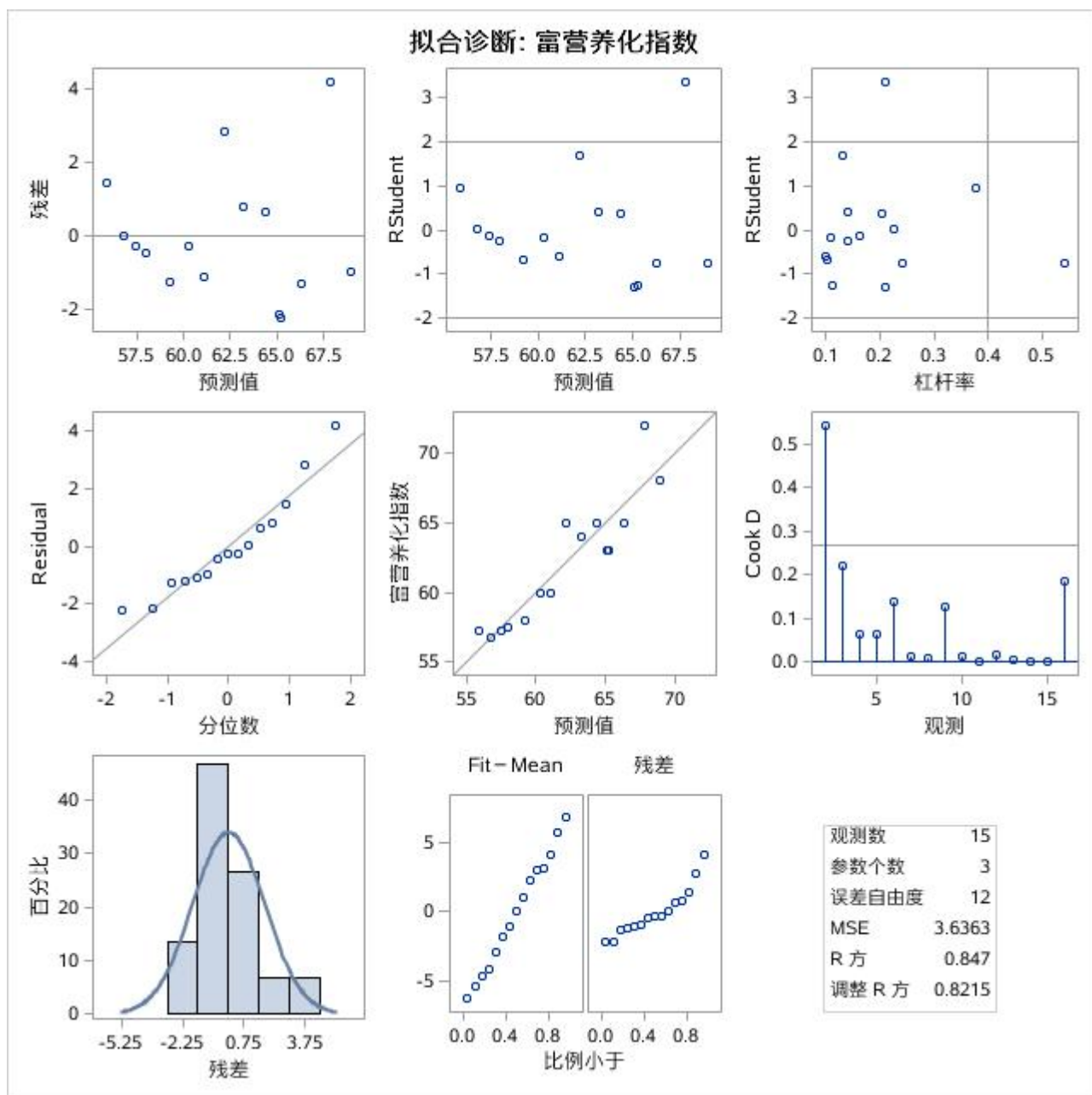


图 3.4.8 水体质量拟合诊断结果

1.3 空气综合指标影响因素分析

用上节的方法对空气综合指标进行分析。为了消除变量之间的多重共线性，同样采用逐步变量选择和人工干预的方式进行指标筛选，最终选取了指标“建成区绿化覆盖率”和“生活污染源排烟量”作为自变量，对空气综合指标进行建模，最终模型结果如下所示：

参数估计						
变量	自由度	参数估计	标准误差	t 值	Pr > t	方差膨胀
Intercept	1	1.96761	0.11781	16.7	<.0001	0
建成区绿化覆盖率	1	-0.02444	0.06172	-0.4	0.6991	1.08771
生活污染源排放烟	1	0.97541	0.4439	2.2	0.0484	1.08771

表 3.4.9 空气质量参数估计

方差分析					
源	自由度	平方和	均方	F 值	Pr > F
模型	2	0.30827	0.15414	2.98	0.089
误差	12	0.62069	0.05172		
校正合计	14	0.92896			

表 3.4.10 空气质量方差分析

均方根误差	0.22743	R 方	0.3318
因变量均值	2.19084	调整 R 方	0.2205
变异系数	10.38091		

表 3.4.11 空气质量模型指标

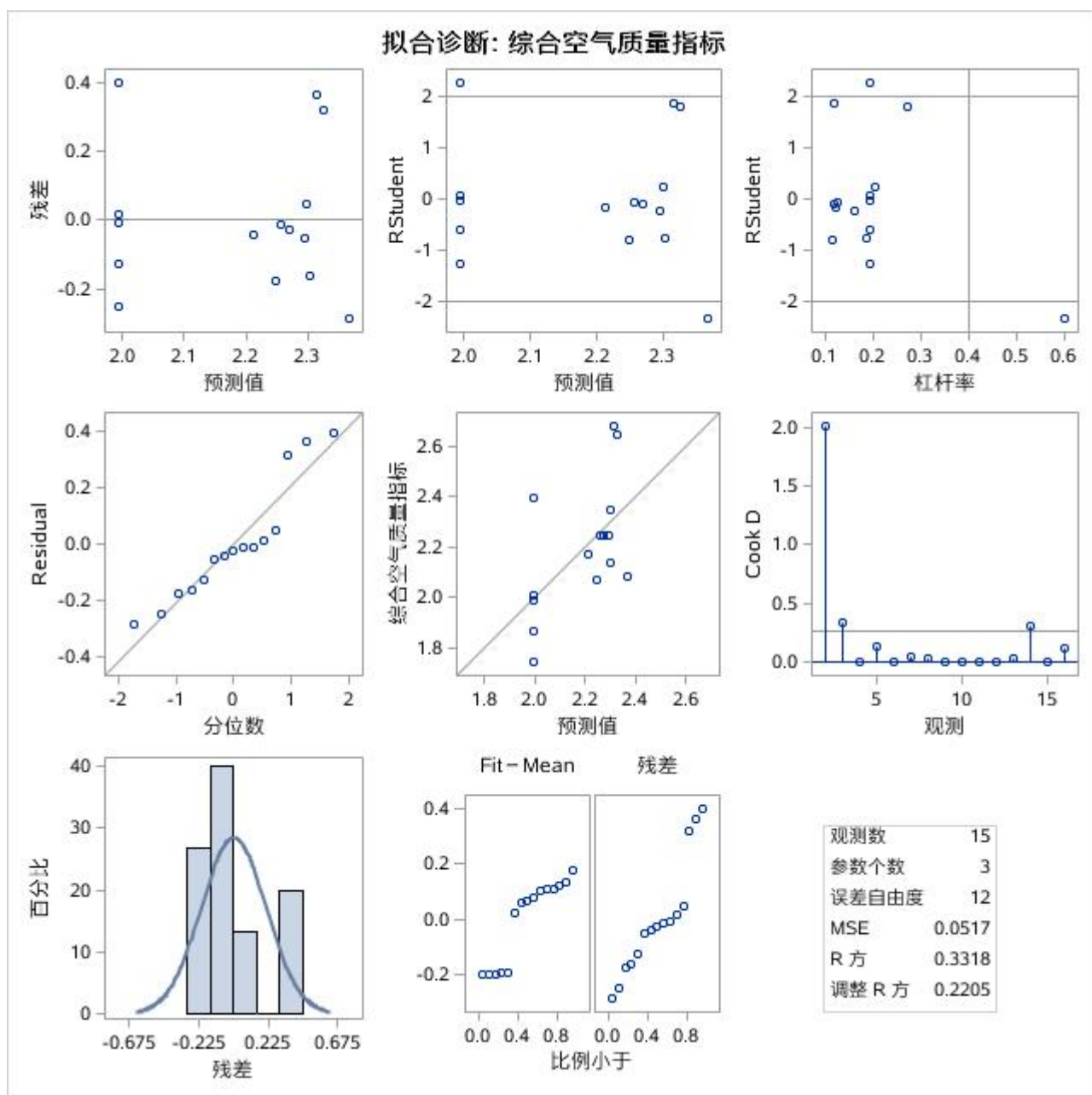


图 3.4.12 空气质量拟合诊断结果

2、神经网络模型

神经网络（Neural Networks, NN）是由大量的、简单的处理单元（称为神经元）广泛地互相连接而形成的复杂网络系统,它反映了人脑功能的许多基本特征,是一个高度复杂的非线性动力学习系统。它的每一个神经元是将输入数据经过激活函数的计算得到一个输出结果。基本模型将神经元排列在各层,第一层称为输入层,连接到一个称为隐藏层的神经元层,依次连接到称为目标或输出层的最终层。每层神经元之间通过一定的权重矩阵连接,形成一个神经网络模型。

神经网络模型和回归模型最主要的区别就是要估计的值的数量级以及权重之间的复杂关系。

本组用神经网络模型对综合空气质量指标的建模结果如下图所示:

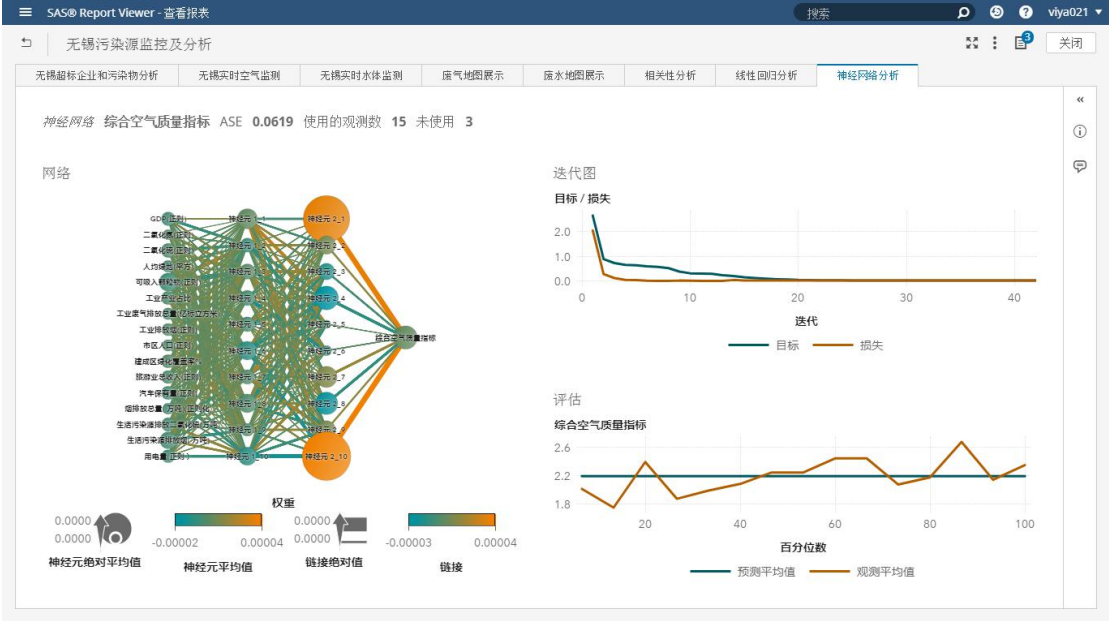


图 3.4.13 神经网络分析

五 实时预测

时间序列是指将同一统计指标的数值按其先后发生的时间顺序排列而成的数列。时间序列分析的主要目的是根据已有的历史数据对未来进行预测。这类模型不需要建立因果关系模型,仅需要其变量本身的数据就可以建模。本组用 SAS 的 PROC ARIMA 对空气的 AQI 指数和 O3 的 24 小时平均浓度进行分析预测。

考虑到获取的数据为 2014/05/13-2018/07/21 的日监测数据,我们对数据做了预处理,加工出月度数据和周数据,所以本组的思路是从长期(月)、中期(周)、短期(日)分别分析指标的历史变化趋势,并对应预测未来的走势。

以 AQI 指数和 O3 为例作简单的预测分析:

1、AQI 指数预测

AQI 指数的月度预测和周预测如下图所示:

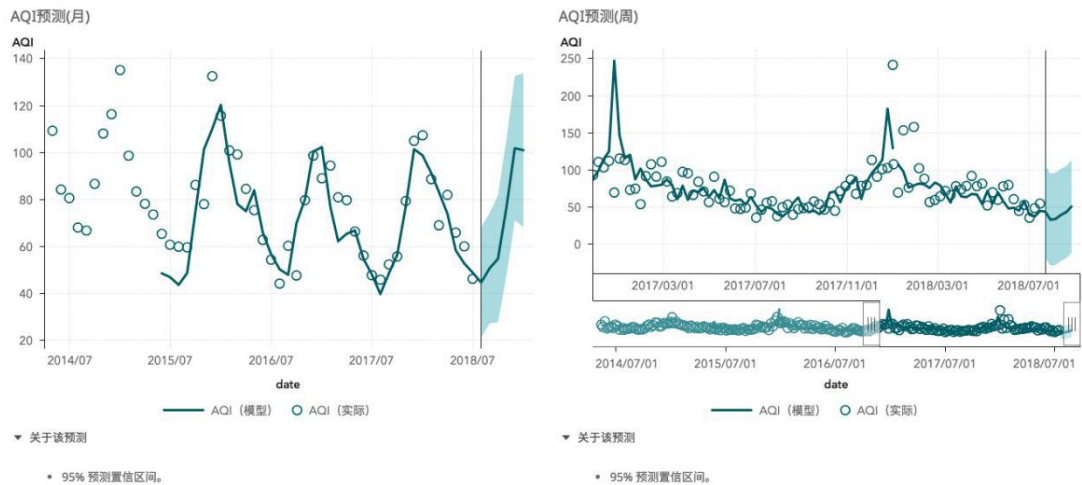


图 3.5.1 AQI 月数据和周数据预测

从月度数据的拟合预测曲线中可以看出，AQI 指数的变化呈现出明显的季节性变化，并且有着逐年降低的趋势。7 月份是变化曲线波谷，在未来的几个月内，AQI 指数将逐月递增。

AQI 的日预测数据如下图所示：

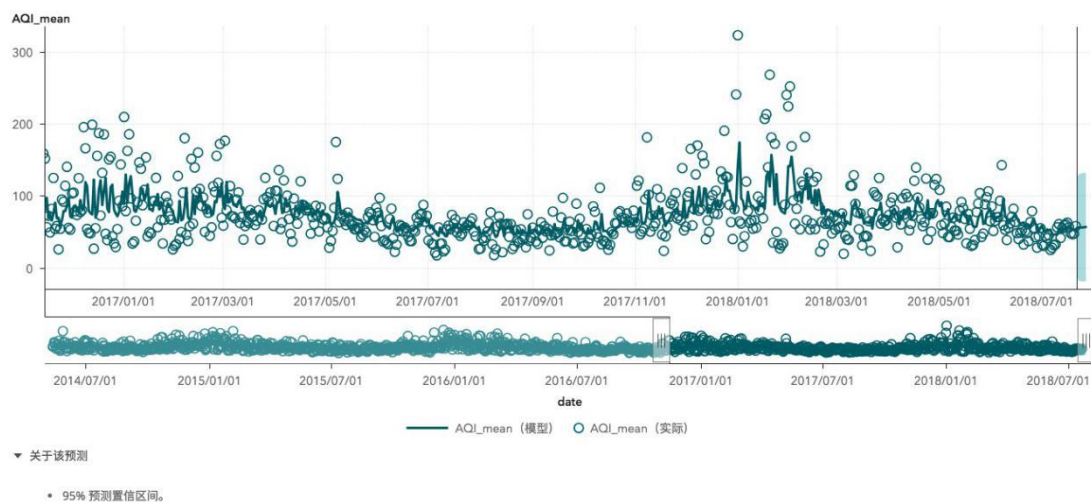


图 3.5.2 AQI 日数据预测

从月预测的结果图中我们得知，在未来的几个月内，AQI 指数有逐渐递增的趋势。但是分析日预测结果图可知，但这是一个缓慢渐变的过程，在未来的几日内，AQI 指数基本保持不变，并无骤升或骤降的异常变化。

2、O₃ 浓度预测

O₃ 浓度的月度预测和周预测如下图所示：

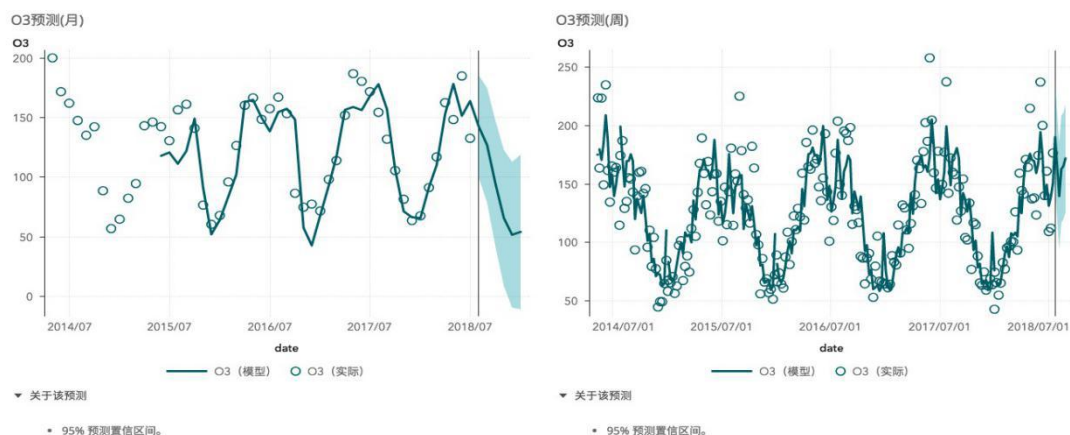


图 3.5.3 O3 月数据和周数据预测

从上图可以看出，O₃ 浓度同样呈现出明显的季节性变化，3-7 月份的 O₃ 浓度处于波峰状态，浓度较高，8-11 月份 O₃ 的浓度处于波谷状态，浓度较低。而且从最近 3 年的数据来看，O₃ 浓度只表现出周期性化，并无增高或降低的趋势。从月度数据的预测图中可以看出，未来的几个月内 O₃ 的浓度会逐月降低，从周度数据的预测图中可以看出，虽然 O₃ 浓度在会不断下降，但是就最近几周而言其浓度是不断上下波动变化的。

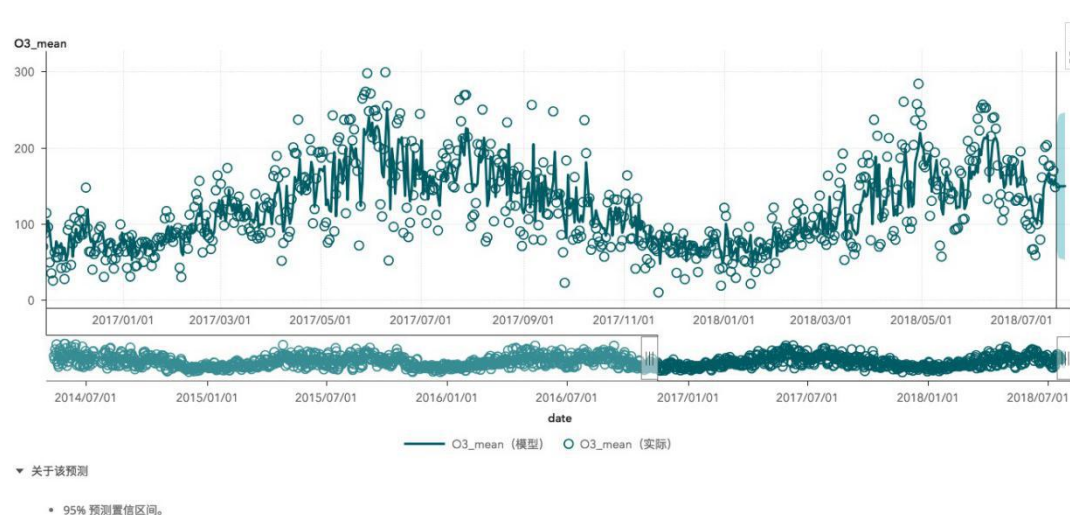


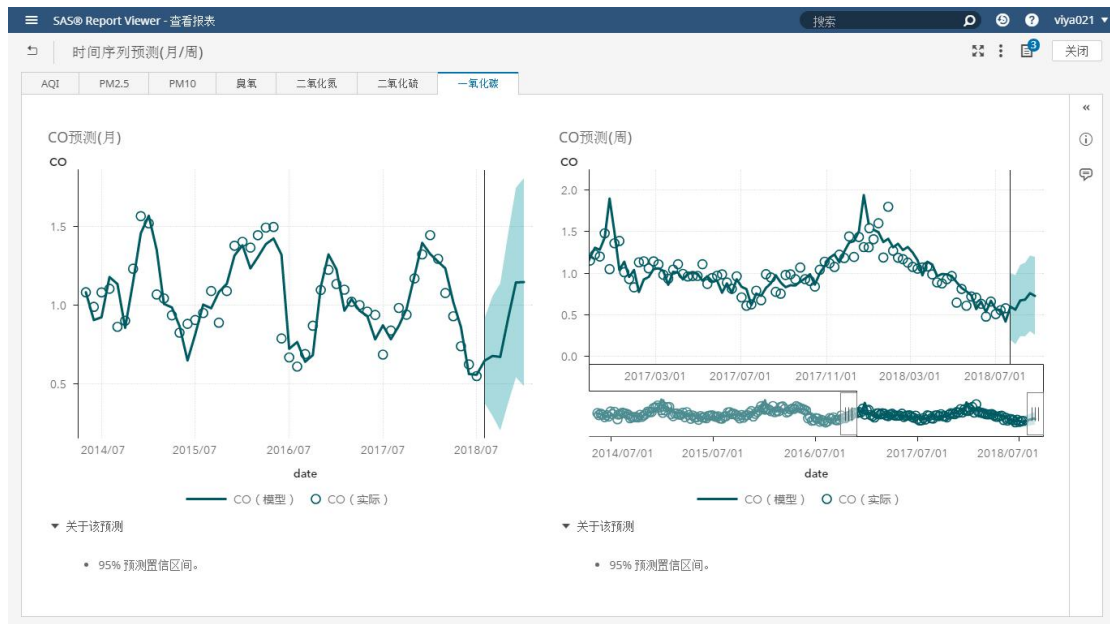
图 3.5.4 O3 日数据预测

从 O₃ 浓度的日数据预测图中可知，O₃ 浓度也是一个缓慢渐变的过程，在未来的几日内，其浓度基本保持不变，并无骤升或骤降的异常变化。

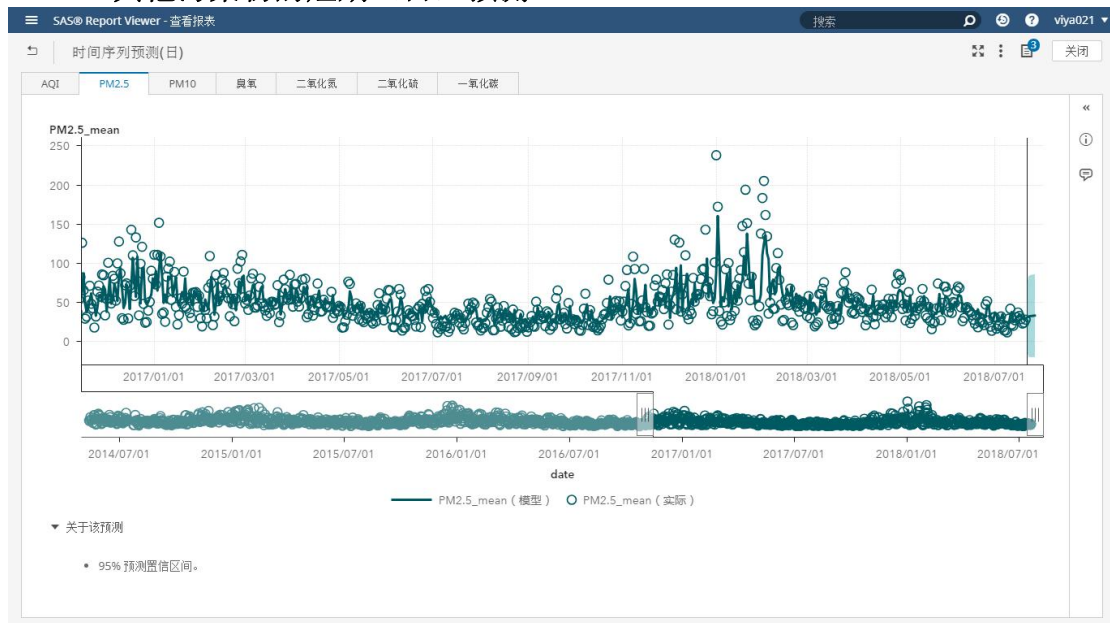
3、其他污染物的中长期（月/周）预测

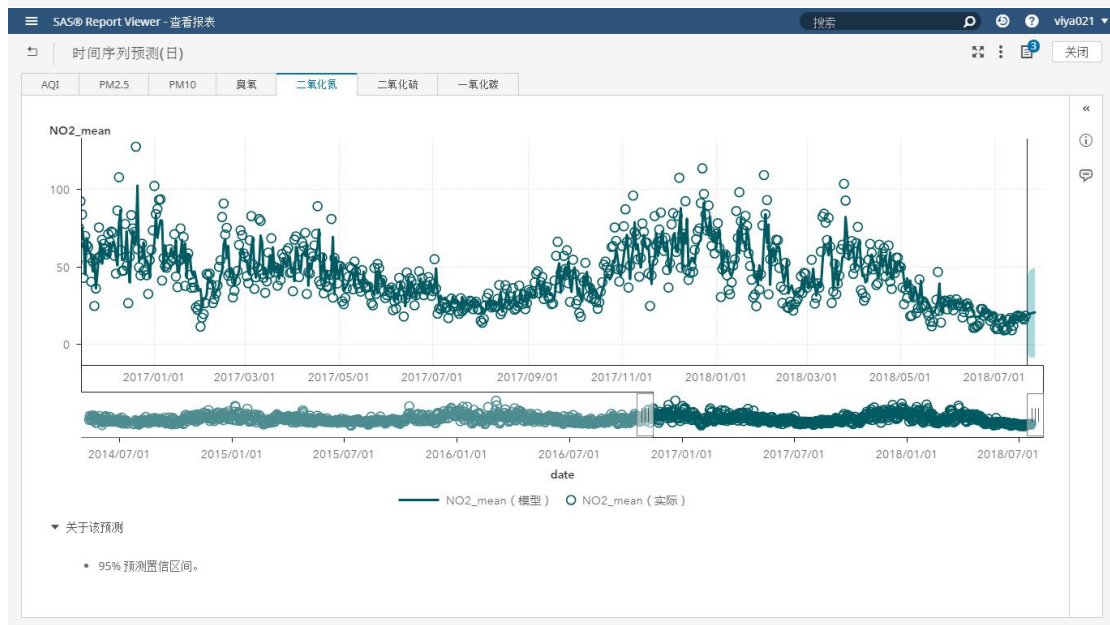
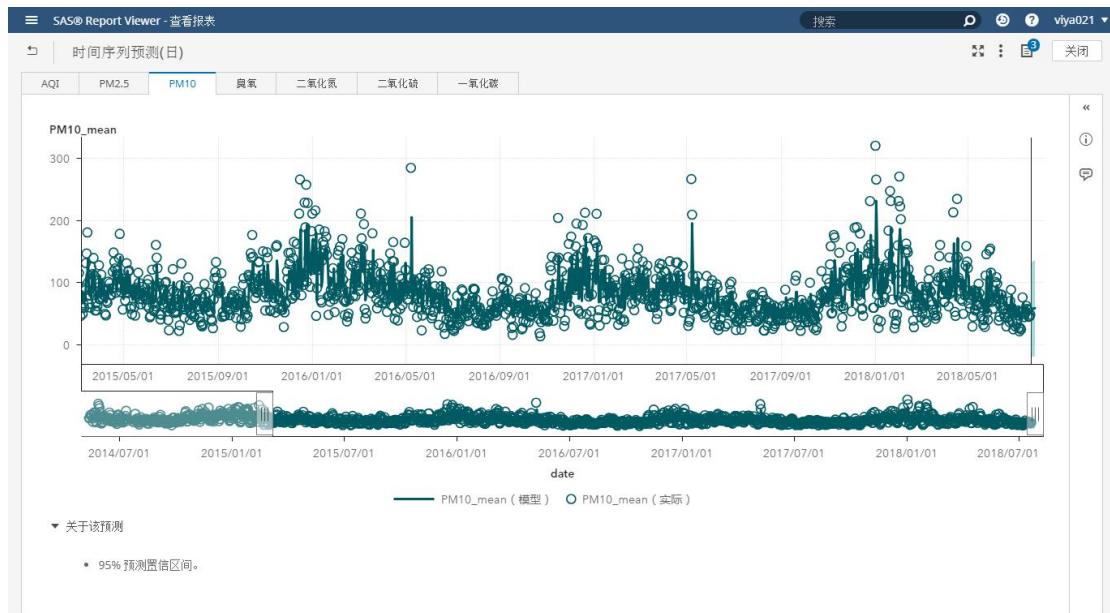


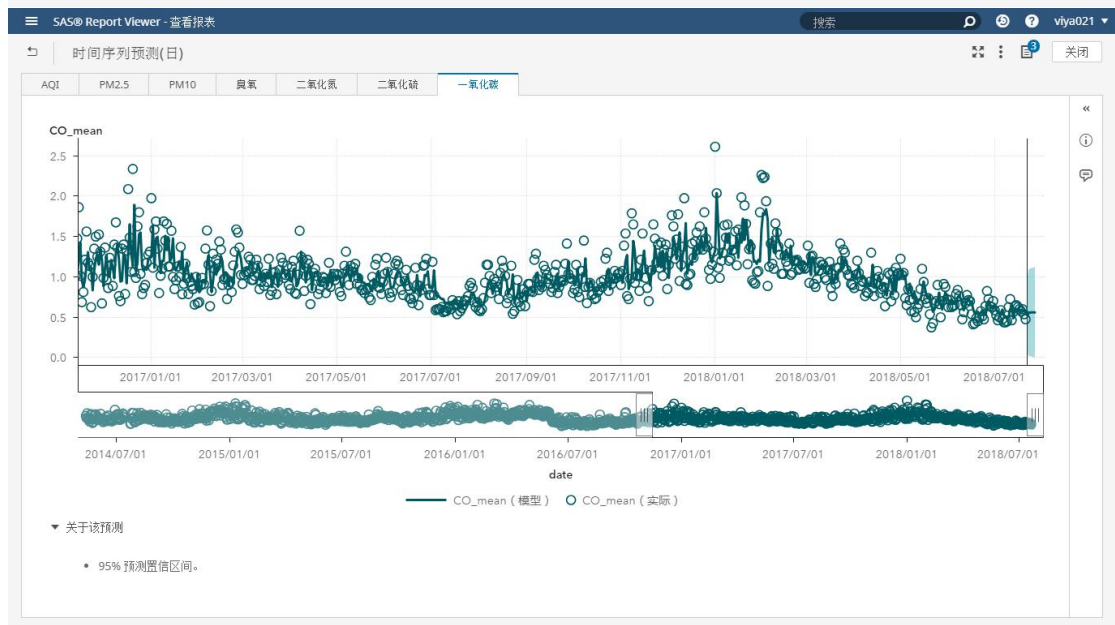
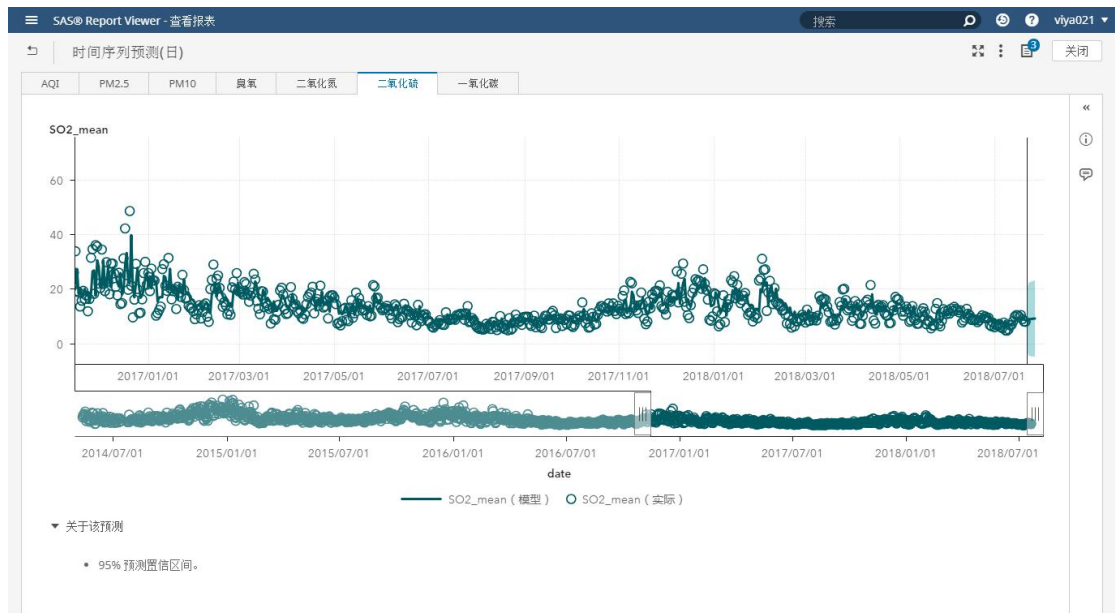




4、其他污染物的短期（日）预测







第四章 总结

本平台通过对多方数据的整理和清洗整合，通过 SAS Viya 平台进行数据的分析，完成了对无锡市水体和空气质量的历史数据分析和实时监控，无锡市污染源地理位置分析，了解无锡市空气和水体污染的来源，并对未来的空气和水体污染进行预测。

对无锡市 13 条主要河流质量的监控，我们发现河流质量呈现出明显的季节规律性，春夏季节河流质量有所上升，Ⅱ类标准河流和Ⅲ类标准河流数量有所增加，而秋冬季节的Ⅱ类和Ⅲ类标准河流数量明显减少。考虑到无锡的地理位置和天气原因，我们能发现天气和降雨量会对河流的质量有明显的影响。春夏季降雨量增大，污染物浓度被稀释，河流质量会因此上升。

对污染源进行地理位置分析，结合到地图上，我们能看出：空气质量相对较差的地区集中在工业化严重的地区，而市区内部的空气质量相对较好；水体污染物较多的地区集中在郊区，即工业化工厂较多的地区，工业废水排放较多，水体污染比较严重。

继续对影响空气和水体质量的因素进行相关性分析，我们能够看出一些比较符合平常认知的规律。比如水体的富营养化，其显著影响因素是人均绿地面积和工业产业占比。人均绿地面积越高，水体富营养化水平越低，水质越好，工业占比越高，则水体的富营养水平越高，水质越差。对空气综合指标影响最大的因素则是建成区绿化覆盖率和生活污染源排烟量。建成区绿化覆盖率越高，空气指标越低，空气质量越好，生活污染源排烟量越高，空气指标越高，空气质量越差。

总体来讲，目前一系列对无锡市水体和空气质量数据的分析，能够让无锡市政府从整体上把握无锡市各类污染的来源和分布，了解无锡市环境质量的历史情况，并对未来可见时间内进行一定程度的预测，能够为政府开展相关环保工作提供战略性建议。

第五章 展望

本平台旨在为无锡市建成智慧城市，在环境保护方向完成监控和预测，为无锡市智慧城市项目的推行和落地提供支撑。

在数据的选择上，我们选择了一部分大赛提供数据，并自主选择了一部分无锡市政府官方网站上公布的信息数据，但在数据量上还是不足以支撑更多更好的分析。比如对空气和水体质量的分析，在更多更详细的数据的支持下，可以继续其他的模型分析。

目前的项目产品在实时监测的部分，暂时只是利用历史数据模拟了一个实时监测，日后接入实时监测的 API 就可以轻松实现真正的实时监测。

本平台基于 SAS 的相关技术和产品，完成了一个环境监控预警平台的基本内容，实现了环境质量的分析和预测，并表现出了充分的稳定性和可扩展性，为无锡市智慧城市项目的落地提供了基础。

第六章 团队

成员	职责分工
李昌盛	数据清洗加工、分析建模
高陶有维	数据爬取清洗、分析建模
曹月	数据清洗、撰写文档