

2021 年第八届中国可视化与可视分析大会

数据可视分析挑战赛

(ChinaVis Data Challenge 2021)

作品说明文档

参赛队名称： 爱好者队-金伟中

作品名称： BDVis:气象大数据可视化辅助科学决策支持系统

作品主题关键词： 大气污染源分析、大气污染时空态势分析

团队成员： 金伟中，爱好者队，kimjwz512@gmail.com，队长

李博，爱好者队，2524722791@qq.com

李浙川，爱好者队，ZChuan_Lee@126.com

陈红任，爱好者队，428447121@qq.com

赵柯存，爱好者队，173974319@qq.com

团队成员是否与报名表一致（是或否）： 是

是否学生队（是或否）： 否

使用的分析工具或开发工具（如果使用了自己研发的软件或工具请具体说明）： D3, Cubism.js, Kudu, Echarts, React, , BDVis (自研数据可视分析工具)

共计耗费时间（人天）： 80 人天

本次比赛结束后，我们是否可以在网络上公布该文档与相关视频（是或否）： 是

一、作品简介：请围绕作品主题、要解决的问题\场景、目标用户\读者、应用价值等方面简要介绍作品

1.1 作品主题以及要解决的问题

我们设计并实现了一个多途径多角度探索式的可视分析系统，旨在帮助分析者：

1. 明确空气污染发展阶段，识别重大污染事件，给予用户多种应对措施；
2. 从多个角度（政府政策/经济发展）和多个尺度（按照年/月/日）探究污染源转变的根本原因；
3. 挖掘污染演化模式的深层原因；

1.2 目标用户

我们的用户是任何有上述信息需求的分析者，既包括数据分析从业人员，又包括对气象污染关心的人民群众。几类典型的潜在用户陈列如下：

1. 期望了解气象污染各方面信息，如政府空气治理措施、各个省份气象污染现状的普通大众
2. 需要在气象污染与污染源发源地的动态变化找出信息，做出决策的政府职能部门
3. 需要回顾、总结、编撰气象污染物传播、发展史的媒体从业人员
4. 意图以从 2013-2018 年为例，开展气象污染防治、气象传播研究领域科学研究的科研人员

1.3 应用价值

本作品可用在各种非专业大气污染展示场景

二、数据介绍：请围绕数据来源、数据格式、数据严谨性、数据清洗等方面简要介绍

2.1 数据来源

鉴于大气污染的复杂性，我们尝试从多个方面收集本次比赛所需数据，并使用大数据相关技术进行数据全生命周期管理。除本次挑战赛官方提供的数据集外（约 3 亿条），我们还收集其原始数据（约 77 亿条）用于整体分析，收集中国省市区边界数据用于地区划分。此外，我们还收集了国家统计局开放的发电量、产煤量、高污染企业等数据用于辅助分析。

2.2 数据格式

大气污染数据全部存储到 Kudu 数据存储引擎中，提供了比肩 parquet 的性能，存储空间占用相差也不多，可以轻松地对快速变化的数据进行快速分析。对于其他 CSV 文件统一以 Text 形式存入 Hive 表。

2.3 数据严谨性

鉴于比赛集中于国内数据，因此我们在原始数据的基础上将中国地区分离出来，使用 146 个坐标点，包含中国几乎全部陆地以及大片海洋，过滤后共有 51004 个网格点，约 27 亿条数据。然后我们在 ChinaVis2021 给出的数据集上进行验证，最后匹配 41939 个，仅少了 310 个，覆盖率 99.27%，精度满足需求。

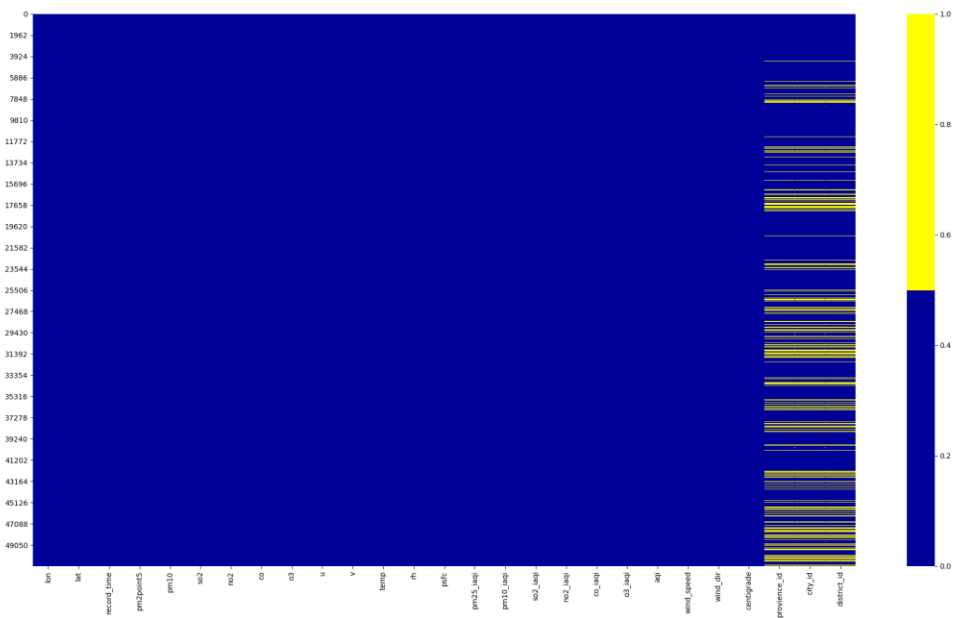


图 2-1 国内坐标覆盖情况（I、II为原始数据，III为官方数据）

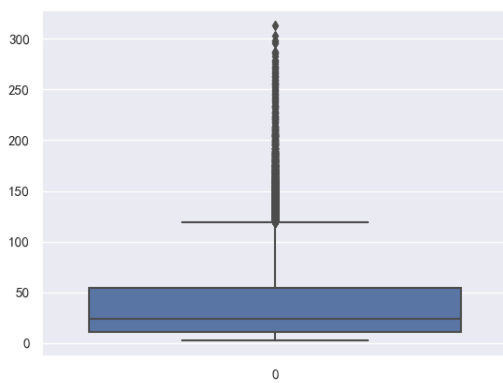
后续我们再根据省市区边界进行数据筛选，过滤后共有 42641 个网格点。然后我们在 ChinaVis2021 给出的数据上进行验证，最后匹配 41907 个，仅少了 342 个，覆盖率 99.19%，精度满足需求。

2.4 数据清洗

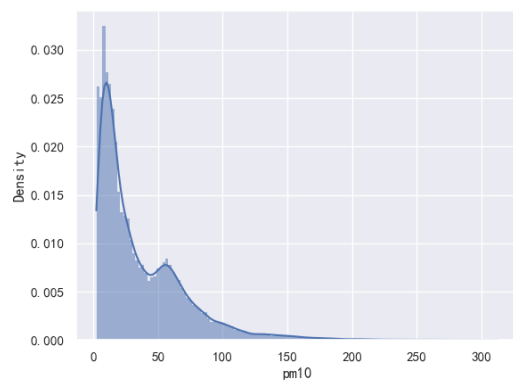
在获取原始数据后，需要进行数据质量检查，对不符合要求的数据需进行相应的清洗工作后才能投入使用。在使用 pandas 进行数据清洗的过程中发现，大气污染数据的质量是比较高的，无需多余处理，可直接进行运算。而其他采集到的 CSV 文件数据质量低，需要对重复数据进行去重，对残缺数据填充 null 值，并检查数据是否异常。



图I缺失数据热图



图II箱型图 (pm10)



图III列分布偏度

三、分析任务与可视分析总体流程

数据可视化的作用在于视物致知，本次可视化的统计分析即是从复杂的污染物气象数据中剥离出异常点以及异常点背后的逻辑脉络即是本次的分析任务；

表 3-1 项目目的是为了完成的五个分析任务

序号	简述	补充说明
1	帮助分析者快速建立对国内 13-18 年污染情况演变趋势理解的模型	让分析者快速理解污染的演变情况的原因，演变速度，波及范围，伤害程度
2	从多角度去深刻理解大气污染的深层次原因	多角度意味着是不单单从污染物数据演变，还要从政策、经济、人类活动来考虑
3	在多种层次的视觉下，观察各地区的大气污染的转变	多层次视觉，通过对污染原始数据的转换利用多种可视化方式进行数值映射
4	探究污染起止演变的关键事件	污染演变起止事件的起因与转折
5	向公众普及大气污染的危害与防治措施	通过移动端与 PC 端的结合让公众可以参与到其中

表 3-2 可视化系统总体流程

流程	标题	补充解释
1	明确分析方向	本系统意在让多用户多群体使用后获取有用的信息与知识
2	采集多种数据	根据选择主题“大气污染源时空态势分析”从不同角度不同层次收集各类不同结构不同方向的数据
3	数据处理与变换	数据可视化的基础是数据表示与变换，为了实现有效的可视化分析，输入数据需要从原始状态到一种便于处理的结构化数据

		表示形式。而且各种数据源获取的原始数据大多是“脏数据”，不能直接用于信息展示，需要进行数据清洗、数据集成与数据变换等处理，去除数据噪声，完善数据，最后将数据格式转换为图表需要的格式。
4	数据可视化映射	将原始数据处理完成并转换为结构化表示形式后，根据数据特征和要表达的信息选择多种图表、设计图表元素，将数据以一种直观，用户（交互）操作后可以获取相应的“知识”的方式呈现出来。
5	可视化系统交互设计	通过图表之间的交互，通过 PC 端与移动端的结合，能够让不同用户群体在我们的系统中找到自己的定位，提供观察数据的不同视角，方便查询和跟踪变化情况。
6	用户分析推理	分析推理挖掘出深度的信息是可视化系统的最终目标，但这不仅仅是用户的任务，也是可视化系统设计者的任务，这需要我们通力协作，实现一个让用户易于接受但又不至于难以上手的工具，同样的也有助于学者探索出污染防治的应对措施。

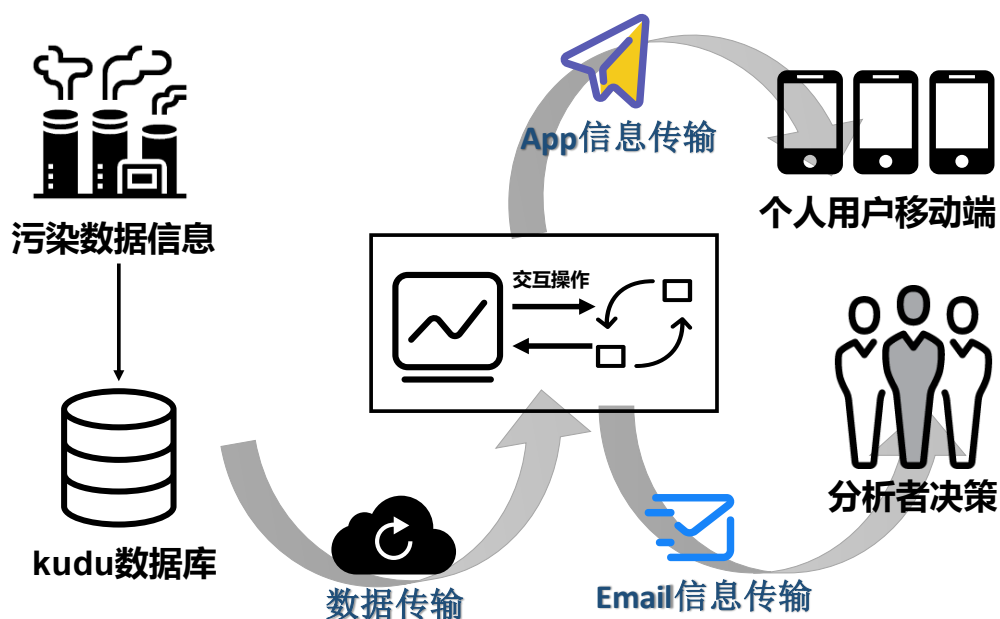


图 3-3 气象科学决策支持系统信息传输流程

技术架构设计:

平台前端使用的 React 技术中的 umi 框架实现页面逻辑与渲染, 使用 D3.js、百度地图 API、Echarts 实现地理信息可视化。

后端使用 Springboot 框架实现快速开发, Mybatis-plus 作为操作数据库的工具; 对于相应数据源的数据, 通过前端传来的表名和 sql, 找到数据源的对应数据连接, 再通过数据连接得到该数据源的驱动和对应的 url, 然后根据不同的驱动调用对应的类的数据查询方法, 在这个方法里面再根据不同的 url 返回不同的 jdbc 连接, 最终执行 sql 查询语句, 对于返回的数据类型, 则是通过 Map 实现字段的组装。

四、数据处理与算法模型

4.1 整体概览

在本次比赛前期调研中我们发现，如果要完整展示大气污染情况，那么数据量将达百亿级，传统数据库难以承担。并且，从实际生产的角度出发，构建一套大气污染数据仓库也是十分必要的。因此，我们保留主干，设计了如下图所示的数仓体系架构，其具有快速摄取、快速存储、BI 即时查询等关键功能点。

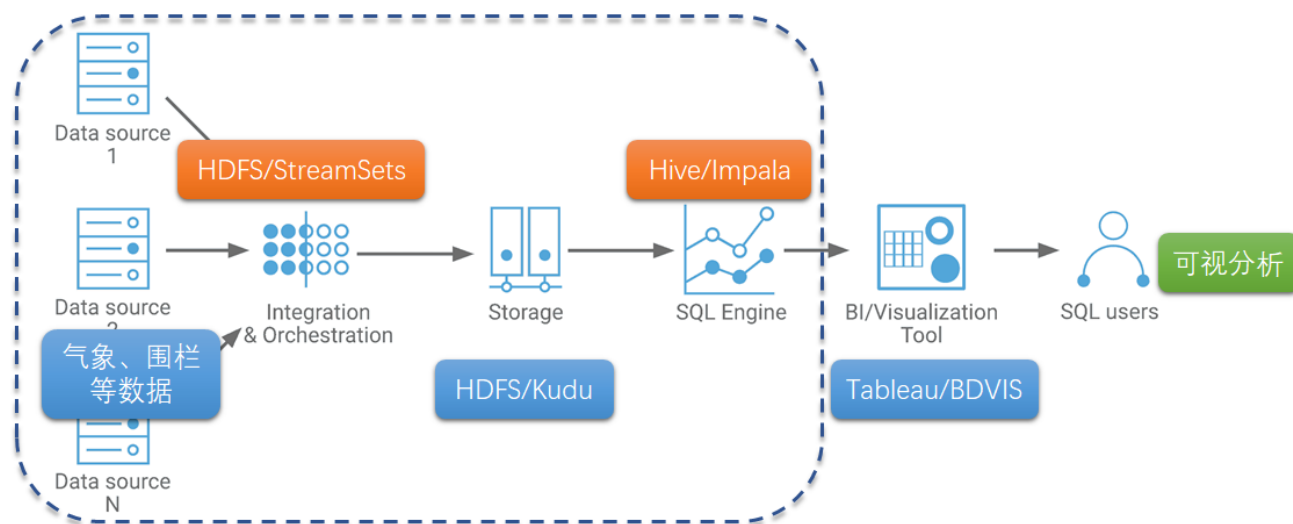


图 4-1 大气污染数据仓库体系架构图

接下来我将对上图虚线框内的数据部分进行介绍。

4.2 数据摄取

对于官方提供的 CSV 数据，我们首先编写脚本，根据文件名新增时间列。然后使用实时采集工具——Streamsets 对目录下的所有 CSV 文件进行数据采集，并存入 Kudu，用于 OLAP；对于其他 CSV 文件，我们先将格式化成，然后上传至 HDFS，最后载入进 Hive 表中。数据抽取流程如下图：

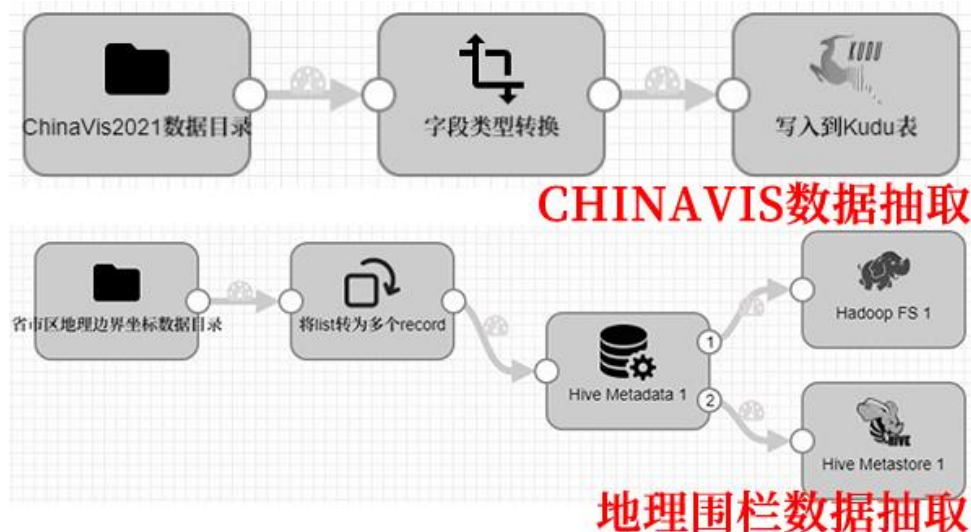


图 4-2 数据抽取流程

对于 NetCDF 格式的原始数据，我们首先使用 pandas 工具将 nc 文件转储为 parquet 文件。相比 CSV 格式，压缩率、处理速度均得到很大提高，仅为原始文件（380G）大小的 52%。由于数据量过大，我们采取离线批处理的方式进行 ELT 工作。先抽取加载（Extract-Load），再转换（Transform）的步骤帮助我们快速载入数据完成建表工作，后续的数据转换则借助数据查询引擎 Impala 完成。首先将 parquet 文件上传至 HDFS，然后载入进 Hive 表，最后通过 Impala 抽取到 Kudu 中并进行分区。优化之后，查询速度可达秒级。

4.3 编写 SQL UDF

为了方便用户使用，降低查询时延，提高数据新鲜度，我们以空间换时间，将数据进行处理转换。针对原始数据中的每一个字段都编写相应的 UDF 或者计算公式，并保留原始数据与转换后数据，形成一张宽表。根据前端展示的实际需求，提供 SQL 优化方案，极大地提高响应速度。

● AQI

根据中国的空气污染分指数计算表编写适用于 SQL 的 IAQI UDF，根据输入的污染物类型以及数值来查表，然后返回计算后的 IAQI 结果。

为方便非专业人士辨别空气质量好坏，我们采用空气质量分指数与污染物项目浓度限值表，分别计算各污染物对应的 IAQI，然后求最大值，获得此刻空气质量分指数。

● 区域

根据围栏的经纬度坐标，采用射线法过滤区域内坐标点，编写适用于 SQL 的区域过滤 UDF。

由此，我们根据事先获取到的省市区边界围栏数据，包含飞地，将国内 4w 多个坐标点绑定到省市区的行政区划代码上。实现查询的加速以及按省市区分析的可行性。

● UV 风

根据 uv 风速计算出实际风速以及风向，编写适用于 SQL 的 UV&SD 转换 UDF。

由于展示所需数据量大，因此在大数据部分完成 UV 风的转换操作，提高前端响应速度。

● 温度

根据开尔文与摄氏度转换公式，计算出我们常用的摄氏度温度。

根据这些转换函数和计算逻辑，我们创建了一张新表供前端使用。通过提供计算后字段、限定查询字段、缩小时间范围、缩小区域范围以及其他谓词下推到数据存储引擎的 SQL 优化方案，不仅减少返回前端的数据量，还提高响应速度。

4.4 节点的选择

由于所有的数据都建立于 42249 个经纬度坐标点之上，映射到地图之中对于前端在地图中选中位置时会造成一定的困扰，选中地点的经纬度大概率不会是我们所拥有的 42249 个坐标节点。对此，我们只能在我们所拥有的 42249 个节点中寻找距离选中位置的最近的一个节点，获取该节点的数据进行展示与分析。

对此我们采取牺牲空间来换取时间的形式，在项目初始化阶段根据已知的 42249 个节点，构建一颗 KD 树，当我们在地图中选中节点时便可以根据该 KD 树快速的获取到当前最近的一个包含数据的节点。便可以根据节点的坐标请求后端接口获取相关数据。

建立 KD 树主要步骤：

- (1) 建立一维数组，存储每一个点的索引，并进行随机打乱。
- (2) 编写分割函数以及选择分割的节点的函数。
- (3) 定义 KD 树递归函数，实现选择分割维度，选择分割节点，将左边数据递归建立左子树，右边数据递归建立右子树。
- (4) 实际搜寻过程中，需要注意的是，先出队列的是叶子节点，距离查找点比较近，最先加入最大堆，从而堆顶距离比较小，在最大堆不满时，进行距离判断，可能会将在 k 近邻范围内的节点排除掉，因此预先加入一个极大距离节点，可避免最大堆不满时，排除掉正确的节点。

五、可视化与交互设计

5.1 系统总体介绍

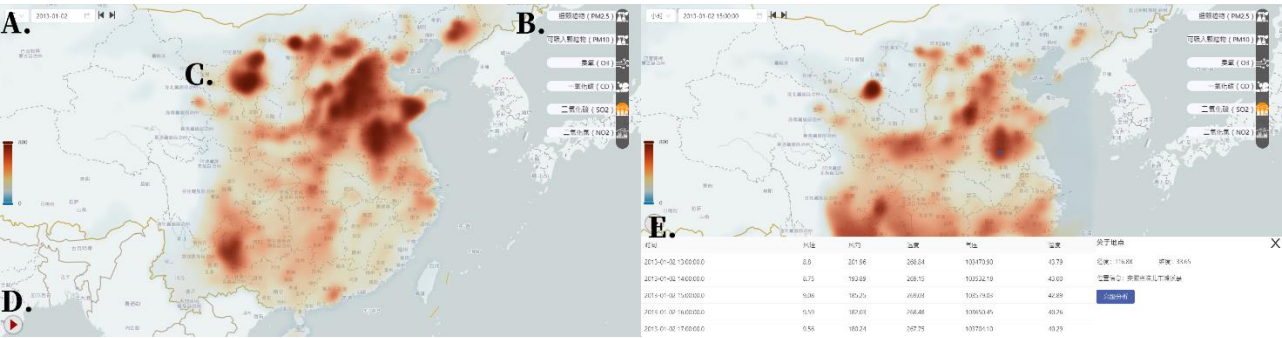


图 5-1 气象大数据可视化辅助科学决策支持系统界面

我们的系统由五部分组成，其中分别的功能为：

表 5-2 系统构成

序号	名称	补充说明
A	时间选择面板	让分析者快速选中她所感兴趣的时间段与时间粒度
B	大气污染整体概览面板	结合时间选择面板可以快速展示污染源在国内的分布情况
C	精确地区污染演变面板	用户点击感兴趣的地点可以展示她所感兴趣的地区及其周边范围的相关具体数据演变情况
D	整体概览面板操控栏	用于操控大气污染整体概览面板
E	污染事件分析弹窗	通过用户选择使用弹窗的形式，展示多种结合算法分析的污染事件

5.2 时间选择面板

系统时间选择面板为分析者提供往期污染事件回顾查询服务，分析者可以快速定位浏览和回顾 2013-2018 年五年所发生的污染传播情况，如下图所示，时间选择面板包括三个部分：“时间粒度选择”、“准确时间选择”、“时间切换组件”。

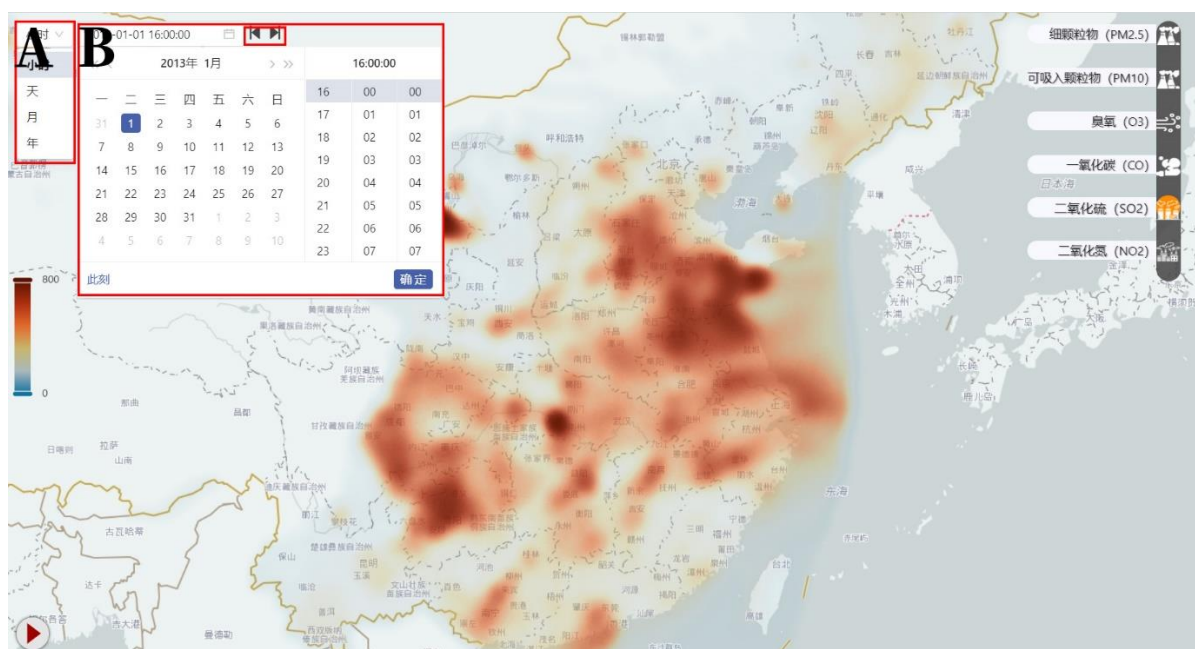


图 5-3 时间选择面板

系统时间选择面板为分析者提供历史数据精确回顾查询,其中图 5-2 可以看到 A 部分用于选中整体展示的时间粒度分别由, B 部分用于选中用户感兴趣的时间范围, 左右切换按钮用于快速切换。

5.3 大气污染整体概览面板

这个面板为分析者提供了大气污染的整体概览功能，我们可以通过第四节介绍的数据处理与算法模型，让分析者能够通过热力图与地区信息聚类快速获取污染事件与传播的有效信息，图中我们可以看到，2013 年在中国中部发生了二氧化氮严重的超标事件。

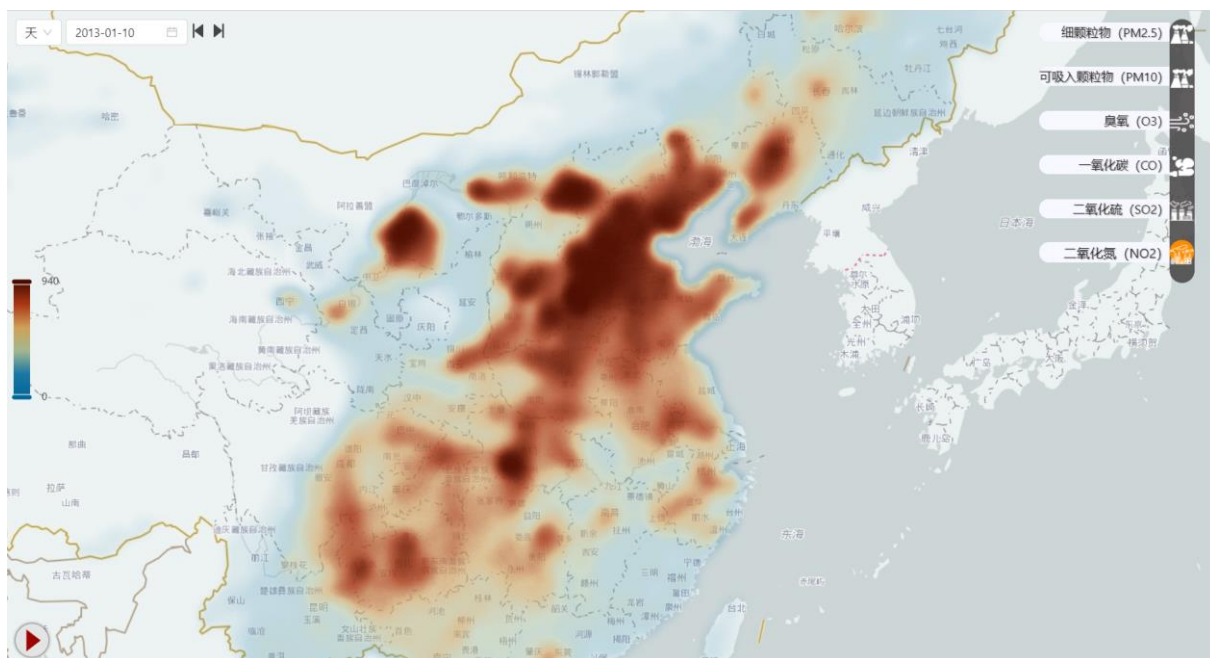


图 5-4 大气污染整体概览热力图

如图表 5-3 的右侧选择栏目，我们可以选择多种污染源以及污染类型，如果选择了时间段模式，在下方我们提供了播放历史数据演变的功能，可以分析者拥有一种更加好的沉浸式体验，可以更容易结合多种数据类型得到想要的分析结论。

5.4 精确地区大气污染演变展示面板

在此面板中，我们使用 4.2 节中介绍的技术路线，为分析者提供一个历史污染传播回顾功能。左下方底部有一个视频播放按钮，用于展示用户感兴趣的时间点两端的污染事件演变情况，支持用户自定义多种内容，也大大提高了用户的自主操作性；其中播放按钮展示以小时为时间粒度下的那一时刻左右两小时的污染变化情况。

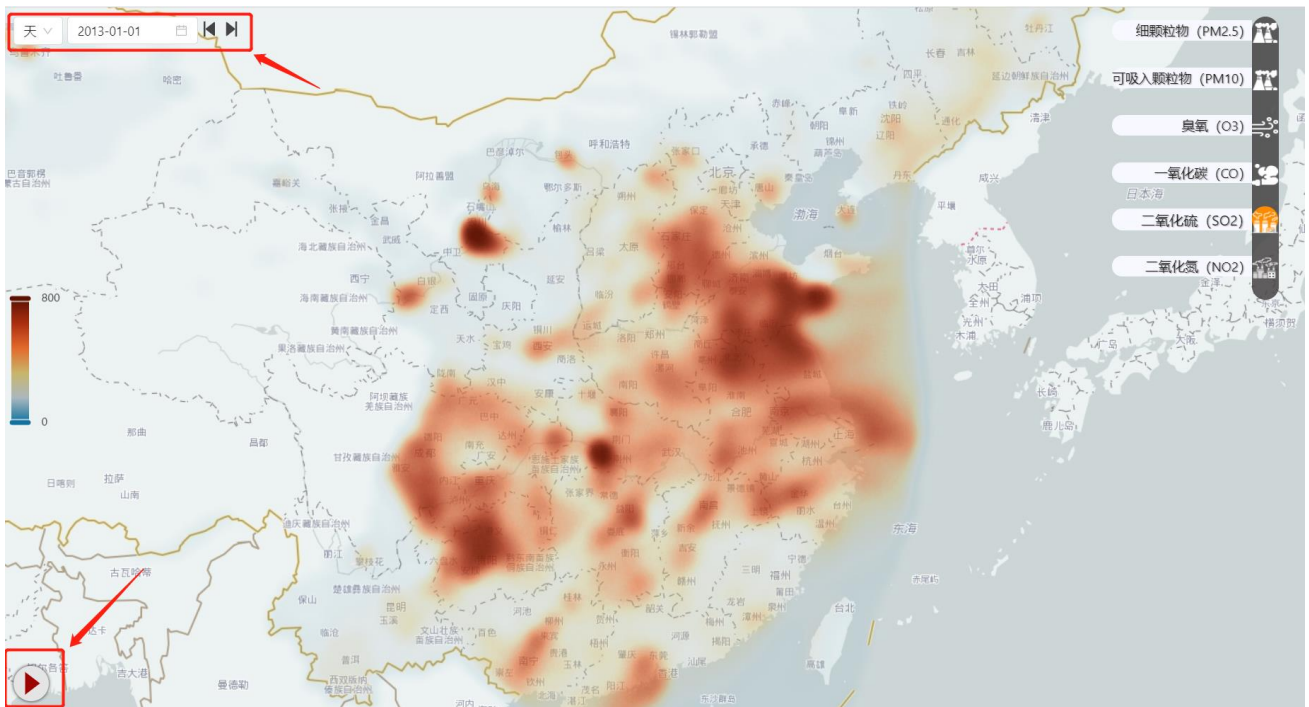


图 5-5 大气污染传播演变播放功能

5.5 气象数据挖掘面板

在此面板中，我们使用 4.2 节中介绍的地图坐标聚合技术，为分析者提供一个精确定位的信息展示功能。

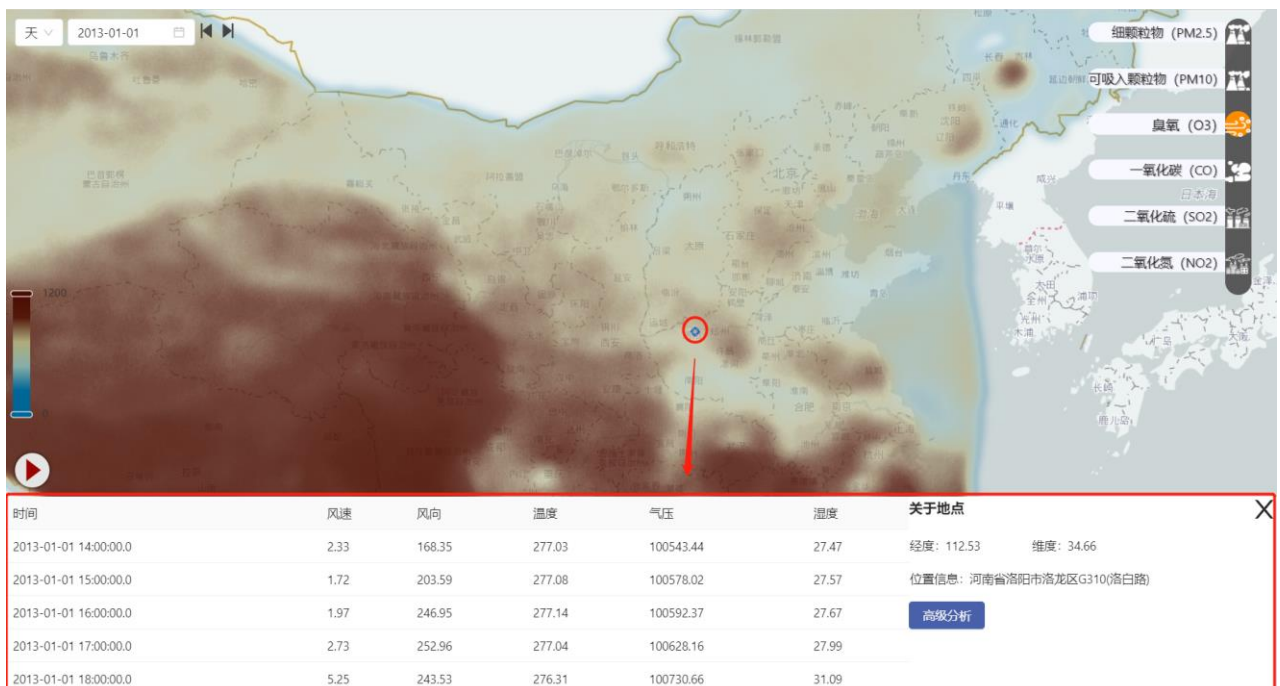


图 5-6 精确地点的信息展示

如图 5-5 底部侧，我们在底部弹窗中提供了多种信息展示组件，其中包括**基本信息**、**风速信息**、**复合图表信息**等等。下面我们来一一介绍每个组件的情况：

时间	风速	风向	温度	气压	湿度	关于地点	
2013-01-01 14:00:00.0	6.98	↓	-0.95	101900.07	44.30	经度：115.38	纬度：34.88
2013-01-01 15:00:00.0	7.7	↓	-1.09	101965.15	46.82	位置信息：河南省商丘市民权县	
2013-01-01 16:00:00.0	8.49	↓	-1.27	102020.49	44.96		
2013-01-01 17:00:00.0	8.57	↓	-2.10	102074.43	45.25	高级分析	
2013-01-01 18:00:00.0	9.0	↓	-3.12	102183.59	43.78		

图 5-7 基本信息展示

基本信息：基本信息这里能够展示分析人员选中的地点五小时内，按照先后顺序排序的风速、风向、温度、地表气压、地表湿度等等信息，有助于对该地区的一个情况有一个整体的把握。

时间	风速	风向	温度	气压	湿度	关于地点	
2013-01-01 14:00:00.0	4.32	↘	-0.12	102025.88	62.81	经度：118.13	纬度：34.35
2013-01-01 15:00:00.0	5.84	↘	0.34	102083.60	57.01	位置信息：江苏省徐州市邳州市	
2013-01-01 16:00:00.0	6.52	↘	0.44	102122.02	51.72		
2013-01-01 17:00:00.0	7.05	↘	-0.25	102177.79	51.06	高级分析	
2013-01-01 18:00:00.0	7.97	↘	-1.30	102308.64	51.31		

图 5-8 风速信息展示

风速信息：因为风速信息这里不单单是只有风速，还有风向，并且是有八个多风向，于是乎我们给风速这里提供了一个详细信息展示的窗口，向比于直接书写文字来展示，图表展示更加易于理解。



图 5-9 复合信息展示

复合图表信息：因为污染源的传播与产生不是简简单单一方面的，而是湿度、气温、风向以及人员活动与密集程度导致的，所以高级图表信息展示按钮用于展示这些特定的信息，联合分析更加直观与易于理解；其中高级图表包含污染物与影响污染物因素关系图、分析雷达图、该地区各类污染物时序变化情况堆叠图以及 13-18 五年全国发电量变化情况。

六、实验\案例\场景分析

6.1 案例 1：回溯 2013-2018 年各地时空方向的污染演变情况

首先，我们使用时间选择器的时间粒度选择功能，以天为粒度展示各地区整体演变情况，如分析者对某一天或者某一段时间的演变感兴趣，它可以切换为以小时为最小粒度进行精确的展示，如图 6-1 所示，我们可以利用 A 模块的时间粒度选择按钮进行粒度切换，同样的我们可以使用 B 模块利用播放组件对感兴趣的时间段内的污染源进行有效播放，让分析者动态化分析污染的传播途径与流程。

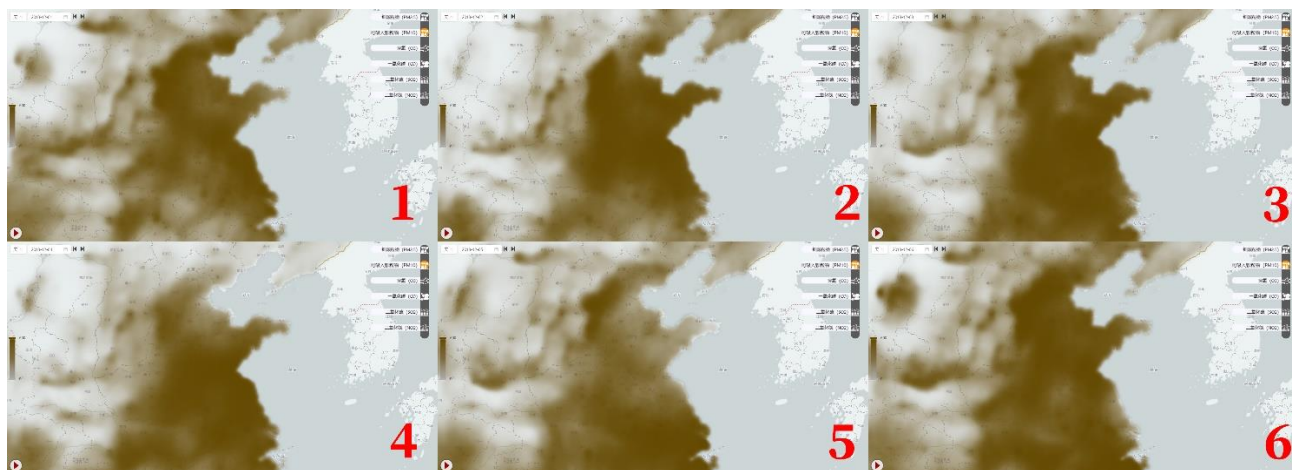


图 6-1 利用播放按钮对 PM10 污染物进行播放的切片效果

6.1.1 2013 年中国中东部严重雾霾污染特征分析

2013 年 12 月中国中东部严重雾霾事件，是指起始于 2013 年 12 月 2 日至 12 月 10 日的重度雾霾事件，是中华人民共和国 2013 年入冬后最大范围的雾霾污染，几乎涉及中东部所有地区。天津、河北、山东、江苏、安徽、河南、浙江、上海等多地空气质量指数达到六级严重污染级别，使得京津冀与长三角雾霾连成片。首要污染物 PM2.5 浓度，日平均值超过 150 微克/立方米，部分地区达到 300 至 500 微克/立方米，其中上海市在 12 月 6 日污染达到 600 微克/立方米以上，局部至 700 微克/立方米以上，从图 6-2 可以看出，中东部地区雾霾演变从 2 号开始到 10 号逐渐消退。

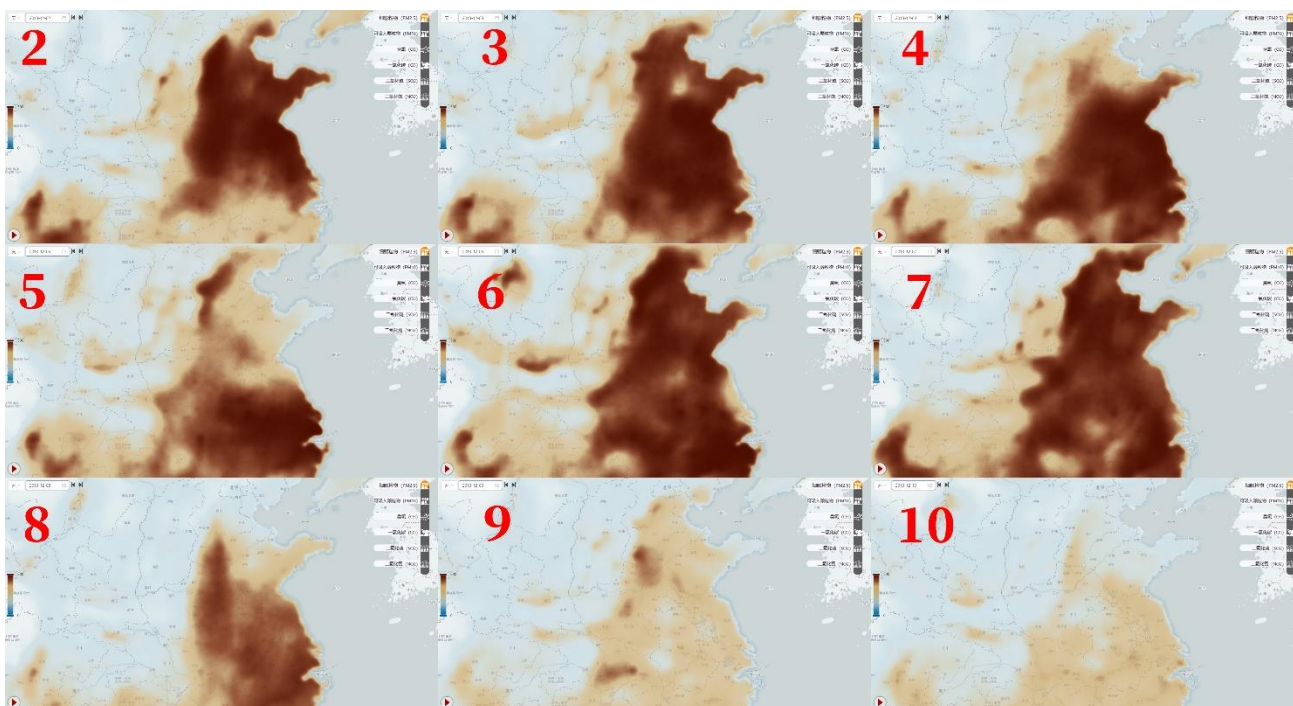


图 6-2 中东部地区 PM_{2.5} 污染物分布情况（初始图时间：2013 年 12 月 2 日）

6.1.2 长三角地区 2014-2018 年冬季强霾污染特征分析

长三角地区 PM_{2.5} 浓度的空间分布为北部浓度高，南部浓度低，其中看图 6-3 就能看出沿海城市的 PM_{2.5} 浓度相对内陆城市较低；

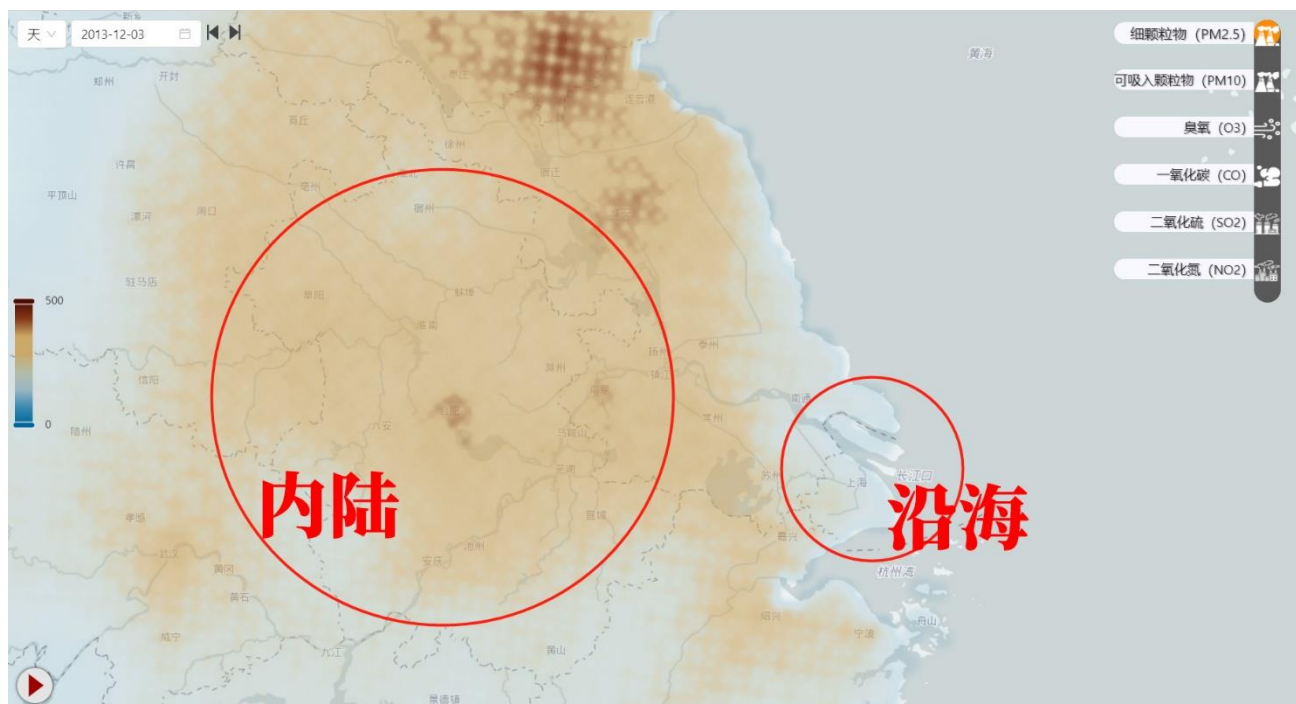


图 6-3 长三角地区 2014-2018 年 PM_{2.5} 内陆分布

从污染输送与扩散方面分析发现，长三角地区冬季主导风向为西北风，受北方气团长距离输送影响较大，导致冬季 PM2.5 浓度受我国北部污染物输送影响从而升，而沿海城市，受海陆风影响，有利于污染物的扩散，因此沿海城市 PM2.5 浓度相对偏低。冬季上海地区 2013 年 11 月到 12 月初，多天 AQI 超标。

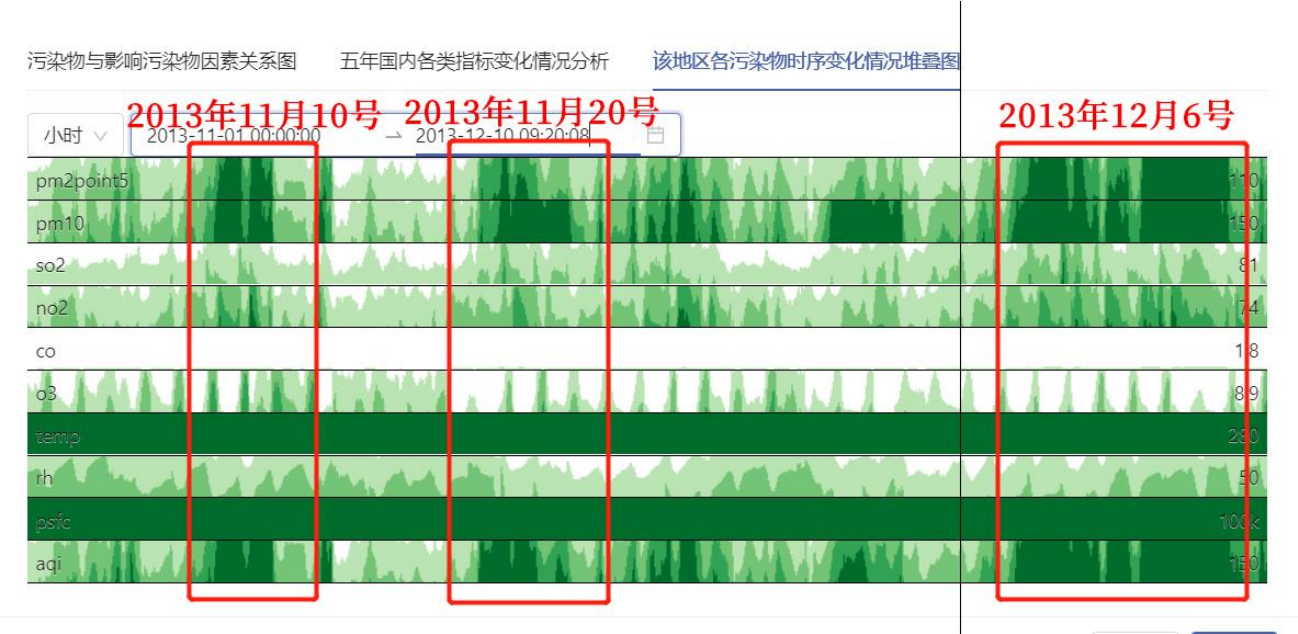


图 6-4 上海地区 2013 年冬季 11 到 12 月份三次严重雾霾污染

综上所述，长三角地区 PM2.5 浓度基本呈现一种由北向南逐渐好转的趋势，即冬季长三角地区发生强霾污染事件时，长三角北部地区污染更为严重。根据可视化分析与调查，导致长三角地区 PM2.5 浓度在空间上呈现北高南低的原因可能为：

1. 长三角地区南部的人类活动较少，降低了大气污染物的排放；
2. 并且南部森林覆盖率较高，可以有效减少地面气溶胶进入大气，并吸附空气中的颗粒物；
3. 长三角地区北部植被覆盖率相对较低，人为活动相对强烈，导致大气污染物浓度比南部地区高；

6.1.3 上海地区大气污染特征分析

使用我们的系统针对分析 2013-2018 年上海地区冬夏大气污染情况：

夏季：由于江苏南部、安徽中部以及浙江北部为经济发达的工业区，PM2.5 排放量较大，经过上述区域的气团携带高浓度 PM2.5 向上海地区输送，传输作用非常明显；

冬季影响：由于河北南部、河南中东部、山东、安徽、河北中东部、江苏以及浙江中北部的气团均可对上海形成污染物的有效输送。

其中，来自京津冀以及周边的产生高污染区的气团在偏北气流的作用下，对上海的影响不可忽视。

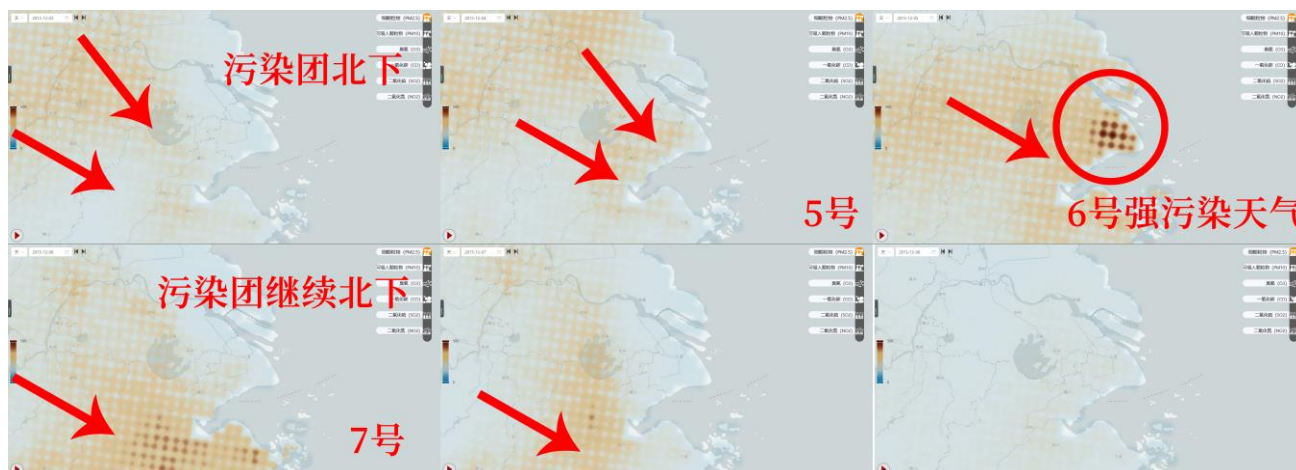


图 6-5 上海地区 2013 年冬季 12 月份大气污染（污染气流从北方伴随着偏北气流吹来）

6.2 案例 2：分析各地工业变化与污染源源头变化之间的关系

来到相关数据分析，我们使用国家统计局开放的发电量、产煤量、高污染企业等数据，分析从中国五年整体工业变化与污染源变化来考虑污染治理与污染演变之间的关系；例如上到政府再到私有企业，下至平民百姓都需要电，那么各个省份之间发电多少与该地的大气污染有着一定的关系，以图 6-6 所示，我们可以看到 13-18 年以广东、江苏、内蒙古、山东、河南、浙江为首的几个省份用电量居高不下，这与这几个省份含有大量工厂与工业设施有很大关系，同时分析图 6-7 的全国的用电构成我们可以看到，大气污染物 CO₂、SO₂ 甚至是 PM_{2.5}、PM₁₀ 颗粒物的产生原因来源于火力发电，其中火力发电量最高的省份也是江苏、山东、内蒙古、广东、河南这几个用电大省；

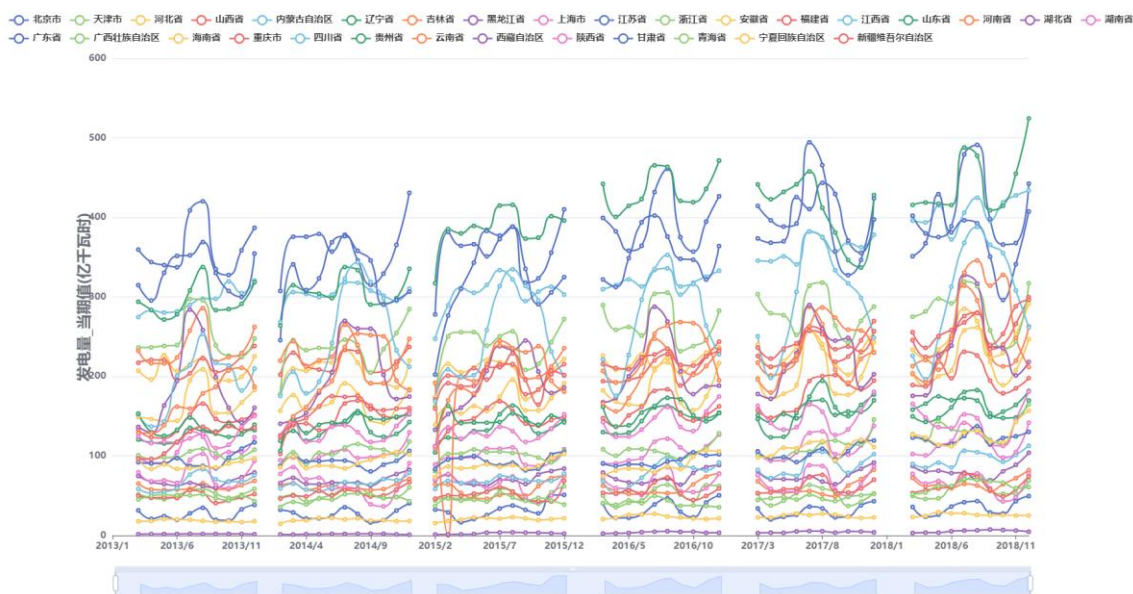


图 6-6 全国地区各省份发电量情况

再根据国家统计局的工业增加值增长速度数据（判断短期工业经济的运行走势，判断经济的景气程度），分析出有不少省份在治理大气污染过程中出现了产业转型的阵痛，工业增加值增长速度多年处于负值，其中有西藏、辽宁、山西、吉林、黑龙江、甘肃、内蒙古、天津、河北、海南；其中以辽宁与山西多年处于零值以下，从图 6-7 中可以看出这个结论，我们再利用我们的整体概览系统面板可以综合看出在从 2013 年污染最严重开始，国家开始综合治理与产业转型过程中，以东三省与山西等内陆省份最为剧烈，由于多年依赖于重工业发展，大排放、重污染等企业的相继关闭与治理，这些省份工业增加值增长速度多年处于负值；

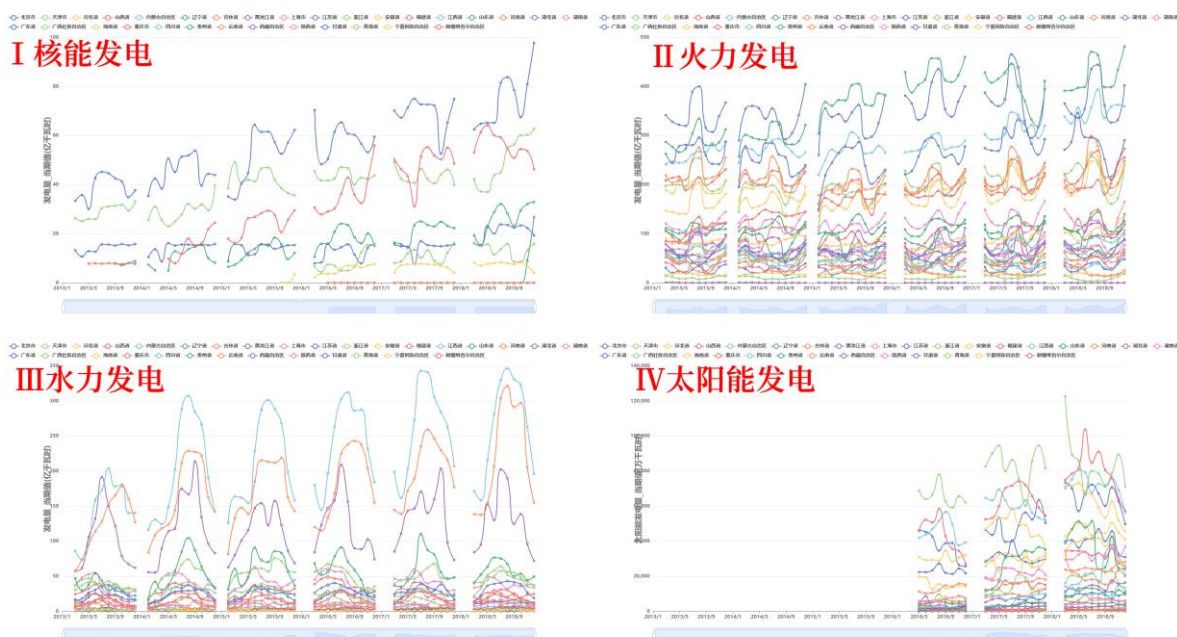


图 6-7 各个省份发电来源组成情况

再根据图 6-8 可以看到山西与辽宁省经过重工业转型的三年工业增加值增长速度变化呈现负数，但同时一年当中的好天气也在不断变多。



图 6-8 五年各个省份工业增加值增长速度变化情况（右图红色是山西，绿色是辽宁）

6.3 场景分析 3：分析中东部城市一年四季污染变化情况

首先，由于一年四季气温的不同，那么污染物的演变情况也是不一样的，根据气温湿度与地表气压之间都与污染物演变具有关系，从下图可以看到冬春两季污染扩散相对较慢，污染指数容易飙升，而夏秋两季则相反，污染扩散相对较快，污染聚集程度不大，同样通过图 6-10 也能看出，在 14 到 16 年演变基本上都遵循冬春季火力发电与地表气压低污染不易扩散导致污染程度大，夏秋两季地表污染扩散相对较快导致污染程度低。

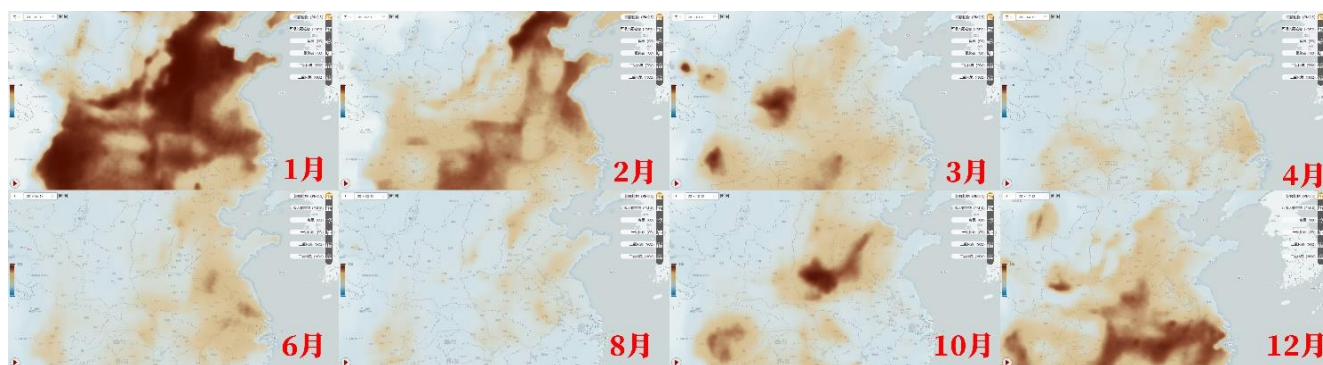


图 6-9 一年四季中东部地区 PM2.5 对比情况

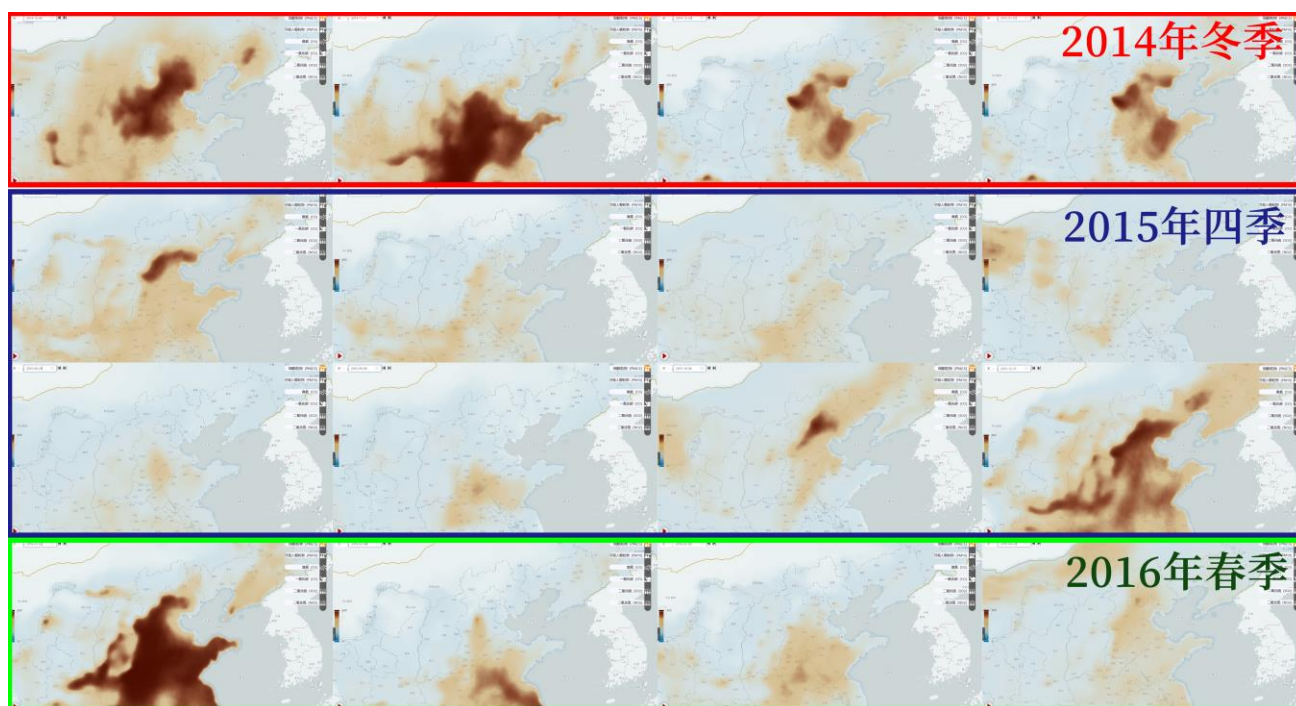


图 6-10 三年跨度中东部地区 PM2.5 对比情况

6.4 场景分析 4：场景分析精确位置的污染物指数与其他因素之间的关系

首先，由于六类污染物与风速，风向，气温，地表气压，湿度存在一定的关系，为了准确分析各地区的相关情况，我们提供了高级分析的选项，用于分析各地区所在地点的六类污染物的演变情况与气温，地表气压，湿度之间是否存在相关情况。由于第 6.1.2 节得出的结论是内陆比沿海城市重污染比例相对较多，那么使用我们“该地区各污染物时序变化情况堆叠图”功能也能看出这一结论，如图 6-12 所示。

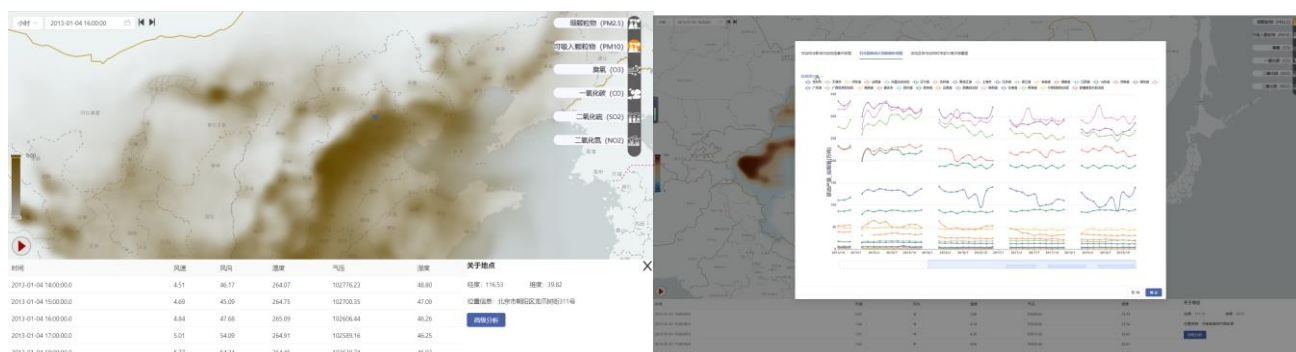


图 6-11 点击感兴趣的坐标位置然后展示该位置信息的相关细节信息

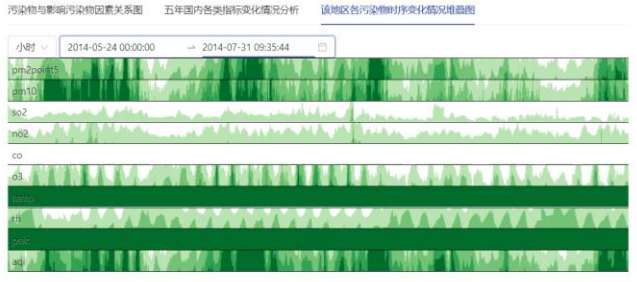
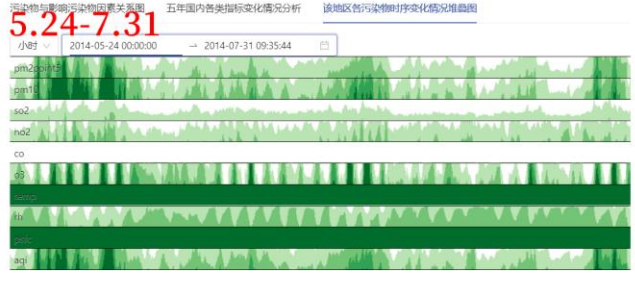
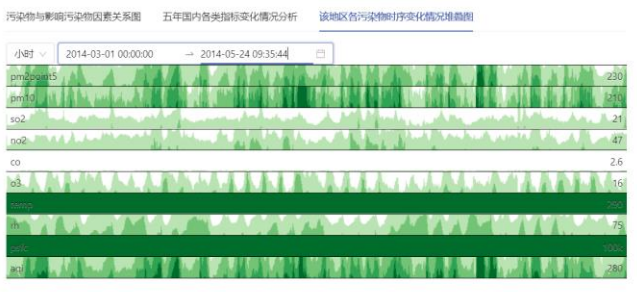
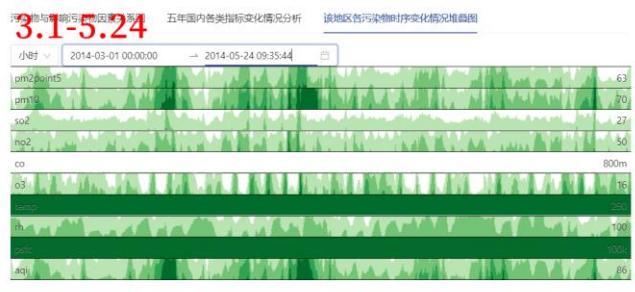
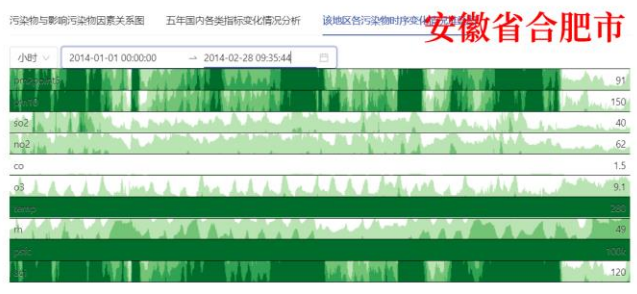
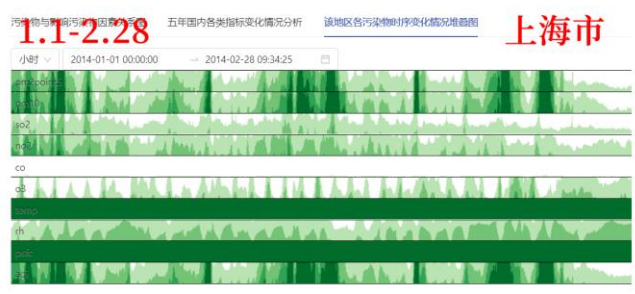


图 6-12 2014 年上半年 24 小时上海与安徽省合肥市内陆与沿海城市污染情况比较

6.5 场景分析 5：人类节假日活动与污染之间的关系

6.5.1 春节各地政策与当地污染之间的联系

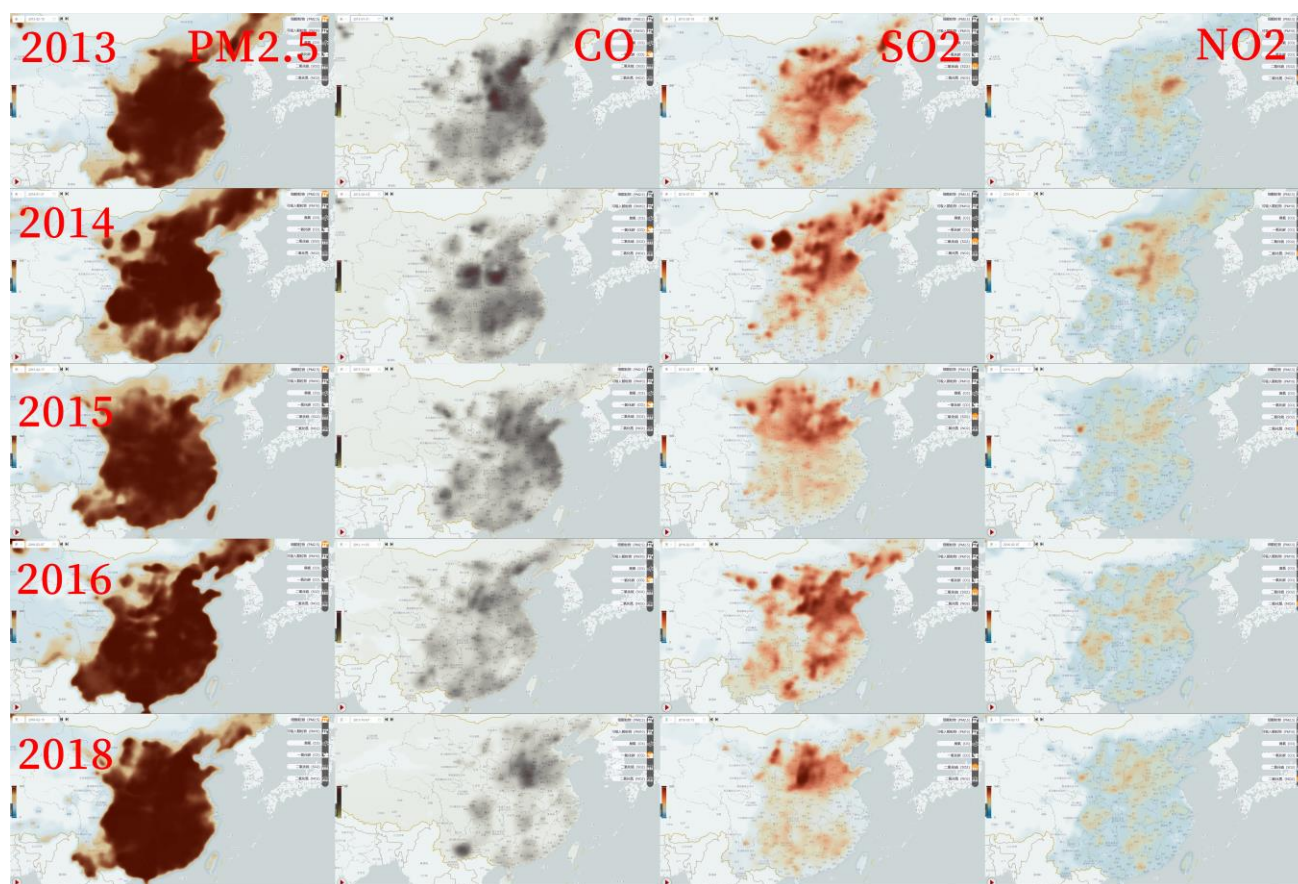


图 6-13 国内除夕当天四类污染物整体分布情况

首先，从 2013 年随着严格的清洁空气行动的实施，我国空气质量发生了显著变化，在春节期间最明显的就是多地大城市禁止燃放鞭炮烟花，这在一定程度上大大减少了污染物的产生，以四川省为例，成都从 2015 年 1 月 1 号开始在市区内禁止燃放烟花，在 2 月 18 号除夕当天，内江、自贡、宜宾、乐山、眉山等暂未燃放或者村庄并未遵守相关规定的地区出现了重度污染等现象，如图 6-14 圆圈框出的成都市与下方重度污染地区有明显的区别。

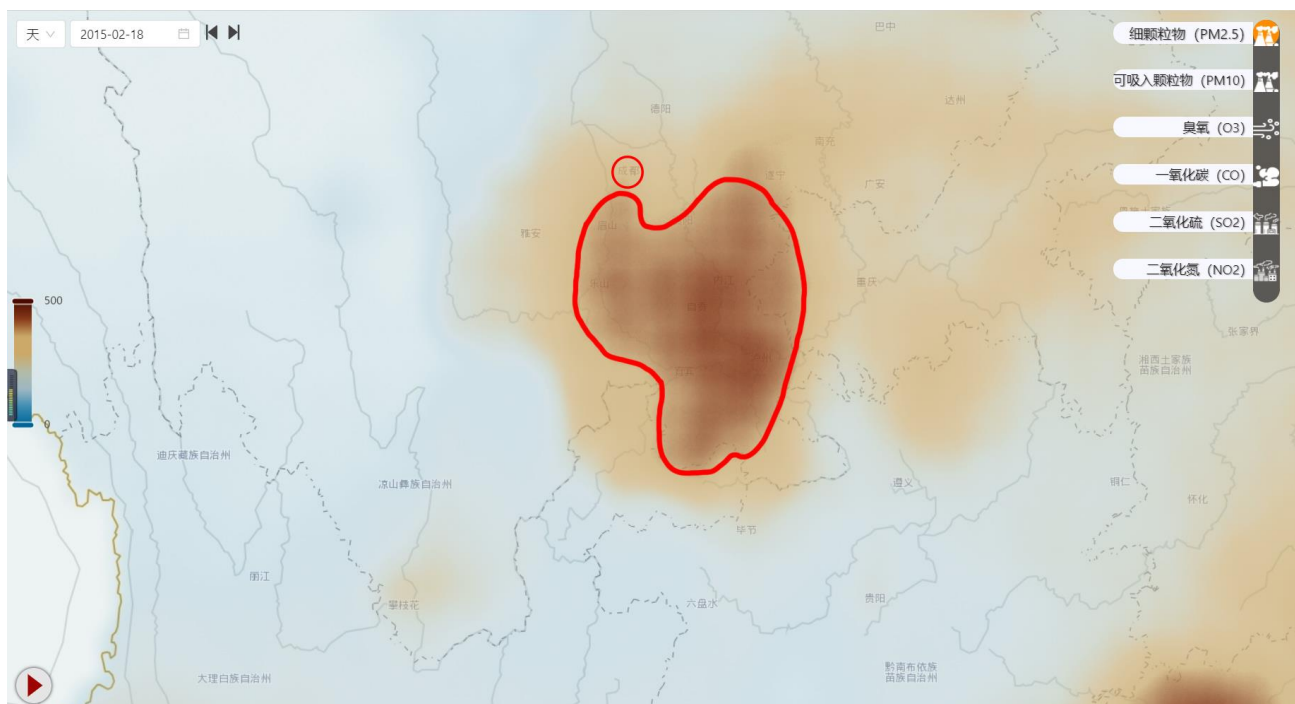


图 6-14 成都禁放烟花爆竹与周边市区形成明显区别（红色圆圈是成都市）

七、讨论与总结

7.1 系统功能点讨论

数据处理：本次大气污染可视化系统大数据端，采取的是离线批处理模式，对于历史数据的处理是十分有效的。我们观察到时间粒度最小是 1 小时，不清楚实际是不是也是这么粗粒度。因此后续我们考虑直接接入，物联网（如：MQTT）数据，然后通过流计算对数据进行处理，之后存储到 Kudu 等 OLAP 型组件上，历史数据则定期转储到 HDFS。不同的存储组件搭配不同粒度的预测、预警等算法。

数据分析：本系统前端较好地完成了数据展示的功能，但由于成员工作原因，数据分析部分显得薄弱。后续我们考虑首先增加更多地交互功能，比如：框选、联动、组合等效果，以及提供更多分析型图表，全面具体地分析大气污染。

7.2 针对 2013-2018 年国内污染治理建议

第一，重点监控主要污染行业。从经济发展阶段来看，很多重工业地区处于工业化中后期，面临着资源与环境的双重约束。鉴于大气污染的行业差异，有必要通过重点监控大气排放的主要行业。对于二氧化硫的排放，需要重点监控电力、热力的生产供应、非金属矿物制品业、化学原料及化学制品制造业、黑色金属冶炼及压延加工、纺织业、化学纤维制造业等行业。对于烟尘排放，需要重点监控电力、热力的生产供应、非金属矿物制品业、化学原料及化学制品制造业、纺织业、黑色金属冶炼及压延加工、化学纤维制造业等行业。对于粉尘排放，需要重点监控非金属矿物制品业与黑色金属冶炼及压延加工业。同时通过产业、环境等政策进行调控，引导资源的合理配置，大力发展高新技术产业与服务业，减少环境污染。

第二，继续提高排污收费标准。排污收费制是环境外部成本内在化的一种有效手段，是基于市场的环境政策工具之一。

第三，环保部门对于排污收费要实行专款专用，收之于企业，用之于企业。将这笔资金用于帮助排污重点企业开展排污治理工程，购置排污治理设施并使之正常运行，实行废气的达标排放。

7.3 总结

我们认为我们的系统很好地完成了第一章列出的分析任务，基本达到了设计目标，它可用在对于专业分析人员结合具体政策或者相关数据辅助分析污染源传播背后的原因，同样的也帮助了那些传统非气象工作人员的容易了解气象演变背后的故事。此外，我们还引入大数据技术，轻松应对海量数据。还可实现实时数据采集，提供各种 API 接口，可直接在生产环境中投入使用。