

---

# 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：重庆大学-周艺璇

作品名称：全国大气污染可视分析系统

作品主题关键词：时空态势、时空分布、污染源识别、污染类型与差异对比分析、高维数据可视化

团队成员：周艺璇，重庆大学，543246329@qq.com，队长

林小嘉，重庆大学，351397085@qq.com

杨 彬，重庆大学，1144149233@qq.com

姜润枫，重庆大学，401186943@qq.com

唐 豪，重庆大学，1963024305@qq.com

胡海波，重庆大学，haibo.hu@cqu.edu.cn，指导老师

团队成员是否与报名表一致（是或否）：是

是否学生队（是或否）：是

使用的分析工具或开发工具：D3.js, Vue.js, ECharts, Tableau, git, python, Node.js, Excel

共计耗费时间（人天）：60 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

### 1.1 作品主题与问题场景

自 2013 年“大气十条”实施以来，全国各地狠抓大气污染治理，我国大气污染防治成效显著。近年来，全国多处空气质量监测站收集到大量具有高维、时序特点的空气质量数据；使用可视化与可视分析技术，对识别主要大气污染源、分析污染成因、分析污染时空分布模式、监测时空演变态势、比较各地大气污染差异、评估大气改善状况、识别特定事件等方面进行分析，对解读该类数据、提供知识见解具有重要意义。

本作品主要使用 2013 年-2018 年中国高分辨率大气污染再分析开放数据集，使用污染物特征值计算、污染物扩散模型与风廓线计算、对污染特征值聚类、分析污染指标与气象因素相关性等方法，设计并实现对六年间全国大气污染状况的展示性可视化系统；同时重点对全国各地大气污染成因与类型、主要污染源、大气污染时空分布模式、时空演变态势展开可视分析，并提供对任意两个地区不同或相同时间点的对比分析视图，比较其大气污染物差异；为大气污染防治工作的分析、指挥和决策提供有效手段和决策依据。

### 1.2 目标用户与应用价值

本作品目标用户广泛，包括关注大气污染情况的大众，分析及评估大气污染成效并制定防治措施的部门，以及大气污染数据分析的相关研究人员与学者。

基于大气污染数据的可视化使大众能直观与直接地感受数据所展示出的某地区某时地污染情况，同时直观地为相关部门与政府展示各地大气污染的情况，结合各地大气污染物差异的比较结果，对其相关措施的有效性与改进建议提供辅助支持；对于相关研究人员与学者，通过数据处理思路与可视化视图与交互的设计，提供大气污染数据分析的新思路。

---

## 二、数据介绍

### 2.1 数据来源与格式

结合设计主题与应用场景，本作品主要使用大会主办方提供的 2013-2018 年中国高分辨率大气污染再分析开放数据集，以及由阿里云 DataV 提供的来自高德开放平台 2020 年 4 月更新的地理数据（<http://datav.aliyun.com/tools/atlas>）。

通过 DataV Atlas 地图选择器，本作品获取了中国、全国 34 个省份、以及各省份下二级行政区划地理数据文件，以.json 格式存储，作为系统中地图地理数据的输入。

主办方提供的大气污染数据集由中国科学院大气物理研究所等单位发布，数据包括日均值数据和小时值数据，存储为.csv 格式，包括我国六项常规污染物、五个常用气象要素以及所在网格点的经纬度值。

### 2.2 数据预处理

结合可视分析需求，本作品对数据进行了不同方面的预处理。最终用于可视分析系统输入的数据均被转换为.json 格式，大小约为 118G，具体的数据预处理过程如下。

#### 2.2.1 逆地理编码与数据清洗

首先使用高德地图开放 API 对数据集中的经纬度做逆地理编码，获得每对经纬度坐标对应的行政区划信息到二级地级行政区。

在此过程中，删除了发现的 38 条不在高德地图查询范围内的坐标数据；对直辖市地区统一使用直辖市的名称。

获取到行政区划信息后，将该信息添加到源数据集的最后作为新的列，得到经过逆地理编码与数据清洗后的数据，该数据以.csv 格式存储。

#### 2.2.2 数据统计

对清洗后的数据，按照行政区划，分别统计某省份（省级行政区）和某城市（地级行政区）内所有经纬度点数据的平均值，得到按行政区划统计的日均值数据与小时值数据；在此数据的基础上，对各省份、城市的污染物变量及气象变量按小时、日、月、年做统计，得到各行政区划不同时间粒度的污染物变量与气象变量均值。在.csv 的基础上，将该数据转换为.json 格式。

#### 2.2.3 IAQI、AQI、污染等级计算

空气质量指数（AQI, Air Quality Index），是根据空气中的各种成分占比，将监测的空气浓度简化成为单一的概念性指数值形式。它将空气污染程度和空气质量状况分级表示，适合于表示城市的短期空气质量状况和变化趋势。空气质量分指数（IAQI, Individual Air Quality Index）是指根据单项污染物计算得出的空气质量指数。

根据 2012 年 2 月 29 日国务院发布新修订的《环境空气质量标准》（GB 3095-2012）中对环境空气质量标准分级、主要污染物项目和这些污染物在各个级别下的浓度限值，本作品对各项污染物的 IAQI 与 AQI 进行了计算，并判断得出最终的污染等级。

(1) 污染物浓度标准

本作品中按照以下空气质量分指数及对应的污染物项目浓度限值标准计算各污染物空气质量分指数（表 1）。

表 1 空气质量分指数及对应的污染物项目浓度限值

空气质量 分指数 (IAQI)	SO <sub>2</sub> 24h 平均 (μg/m <sup>3</sup> )	SO <sub>2</sub> 1h 平均 (μg/m <sup>3</sup> )	NO <sub>2</sub> 24h 平均 (μg/m <sup>3</sup> )	NO <sub>2</sub> 1h 平均 (μg/m <sup>3</sup> )	PM10 24h 平均 (μg/m <sup>3</sup> )	CO 24h 平均 (mg/m <sup>3</sup> )	CO 1h 平均 (mg/m <sup>3</sup> )	O <sub>3</sub> 8h 滑动平均 (μg/m <sup>3</sup> )	PM2.5 24h 平均 (μg/m <sup>3</sup> )
0	0	0	0	0	0	0	0	0	0
50	50	150	40	100	50	2	5	100	35
100	150	500	80	200	150	4	10	160	75
150	475	650	180	700	250	14	35	215	115
200	800	800	280	1200	350	24	60	265	150
300	1600	1600	565	2340	420	36	90	800	250
400	2100	2100	750	2090	500	48	120	1000	350
500	2620	2620	940	3840	600	60	150	1200	500

其中，SO<sub>2</sub> 1h 浓度高于 800μg/m<sup>3</sup> 的，不再进行其空气质量分指数的计算，按 24h 平均浓度计算分指数报告；当 O<sub>3</sub> 8h 浓度高于 800μg/m<sup>3</sup> 的，不再进行其空气质量分指数的计算，按 1h 平均浓度计算分指数报告。

(2) IAQI 与 AQI 计算公式

根据相关标准，本作品中使用如下公式计算 IAQI 与 AQI:

$$IAQI_p = \frac{IAQI_{Hi} - IAQI_{Lo}}{BP_{Hi} - BP_{Lo}}(C_p - BP_{Lo}) + IAQI_{Lo}$$

其中，BP<sub>Hi</sub>代表某特定污染物所在的最小区间的上限。BP<sub>Lo</sub>代表某特定污染物所在的最小区间的下限，BP<sub>Hi</sub> - BP<sub>Lo</sub>代表该区间的上下限差值，为一个固定值。IAQI<sub>Hi</sub>代表某特定污染物所在的最小区间对应的空气质量分指数(IAQI)的区间的上限，IAQI<sub>Lo</sub>代表某特定污染物所在的最小区间对应的空气质量分指数(IAQI)的区间的下限，IAQI<sub>Hi</sub> - IAQI<sub>Lo</sub>代表该区间的上下限差值，为一个固定值。C<sub>p</sub>代表实际测得的污染物浓度，C<sub>p</sub> - BP<sub>Lo</sub>代表污染物实际浓度与浓度指标的地位值之差，它不是一个固定值。整个乘积再加上各级别的起点指数即可算出该污染物的空气质量分指数。每种指标按照对应时间区间取值，在相同时间区间下的所有污染物的 IAQI 计算结果的最大值作为这个时间区间的 AQI 数值。

---

## 三、分析任务与可视分析总体流程

### 3.1 分析任务

#### 3.1.1 可视分析任务

结合本作品所使用大气污染再分析数据集高维、时序等特点，考虑大气污染分析场景，本作品实现以下分析任务：

- T1 展示不同时间粒度下各地区各项污染指标与气象数据时空分布模式及时空演变态势
- T2 分析识别各地主要大气污染物
- T3 分析各地不同污染类型与类型变化
- T4 展示并分析各地大气污染物差异、分析各地污染防治成效
- T5 对特定事件的发现与体现

#### 3.1.2 可视化设计目标

在上述可视分析任务指导下，本作品设计的可视分析系统视图与交互应满足以下设计目标：

##### G1 满足分析任务

针对可视分析任务做直接的视图设计，避免为了展示而展示；丰富视图内容与视图之间的上下文分析，确保可视分析任务的完成。

##### G2 视图准确直观

要求视图展示信息与真实数据一致，保证真实性与准确性；使用合乎认知的视觉编码，并在系统中做统一，避免歧义；采用易于理解的隐喻与设计，同时具有一定的创新性。

##### G3 交互灵活直接

视图中的交互应简单直接，用户易于上手；并足够灵活，满足用户常见与合理的分析需求。

##### G4 可拓展性

设计的可视化视图应能轻松地应用于其他类似结构或应用场景的数据，具有可扩展性。

### 3.2 可视分析流程

对上述分析任务，本作品提出以下可视分析流程（图 1）。

从数据出发，本系统主要使用的数据为两种粒度：日均数据与小时均数据。对两种粒度的数据分别展开由粗粒度到细粒度的可视化设计与分析过程。

视图设计上首先由多个概览视图出发，视图内的交互提供了灵活的数据更新与下钻功能；通过控制面板等视图间的交互，将分析视角由粗粒度转换到细粒度，实现围绕用户兴趣点的细粒度分析，过程中辅以丰富灵活的视图内和视图间交互手段，支持大气污染时空分布模式、时空演变态势、污染成因、污染类型等可视分析任务的实现。

再通过页面切换，由日均数据分析转到对小时数据的分析。除了采用类似想法的概览视图外，该页面主要采用对比分析的思想，允许用户通过控制面板选择两个感兴趣的地点与时间点，并对该条件下的不同情况进行多视图关联对比分析，在完成大气污染时空分布模式与时空演变态势分析的基础上，进行污染差异对比分析与污染改善情况分析。

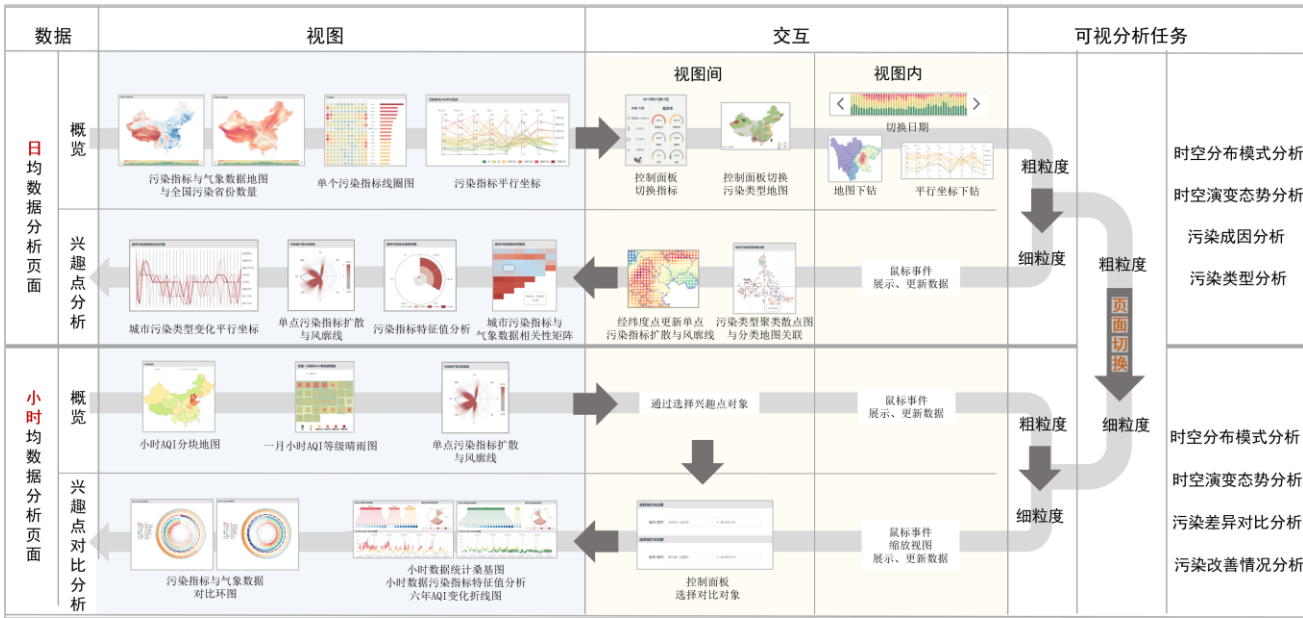


图 1 系统可视分析流程

通过两个层级的粗粒度到细粒度的分析，本作品竭力对不同粒度数据展示的信息与提供的见解分析提供支持。在数据处理的基础上，通过丰富直观准确的视图设计与灵活直接的交互设计，完成可视分析任务与可视化设计目标。

## 四、数据处理与算法模型

### 4.1 污染物特征值计算与大气污染类型（成因）分析

利用污染物数据分析某地区地污染类型与成因是大气污染数据分析的重要部分。按照《环境空气质量标准》要求，我国环境空气质量常规监测六项指标在浓度变化与不同污染物之间的浓度值上存在数量级的变化，直接使用比值法分析大气污染的特征时，不同指标的微小的特征变化被这种差距所掩盖，无法分辨污染特征在时间序列或空间序列上的差异。

本作品结合段菁春等人 2018 年提出的特征雷达图设计中数据处理的算法，对特定地区不同时间的污染物数据作归一化成分谱计算、污染物特征值计算、污染物特征标准值及上下限计算的处理；并将某时间该地区污染物真实数据做上述处理后与该地区当月特征标准值、上下限的比较关系，将该地区当日的大气污染类型分类，包括：标准型、偏二次型、偏沙尘型、偏机动车型、偏燃煤型、偏烟花型、偏钢铁型，以及其他型共八类。此过程中得到的比较关系与分类结果作为后续可视分析系统的输入，设计视图对城市某日的大气污染类型与污染成因做展示与分析。

#### 4.1.1 污染物特征值计算

本作品中对污染物特征值的计算限定于对某一特定地区（城市）某日的污染物数值，与该地当月平均特征值的比较。但该方法可推广至特定时间或地区的污染物数值，与该地区或一段时间内平均特征值的比较。

其中，污染物特征值计算主要包括四部分处理。

##### (1) 计算每个城市在所有日期的百分比成分谱

$$P_{ij} = \frac{c_{ij}}{\sum_{j=1}^n c_{ij}}$$

其中， $P_{ij}$ 为*i*时间第*j*种污染物百分比成分谱； $c_{ij}$ 为第*i*时间第*j*种污染物的原始数值，除 CO 浓度单位为  $\text{mg}/\text{m}^3$  外，其余浓度单位均为  $\mu\text{g}/\text{m}^3$ 。

##### (2) 计算每个城市在所有月份中每种污染物百分比成分谱的均值与标准差

$$\bar{P}_j = \frac{\sum_{i=1}^m P_{ij}}{m}$$
$$S_j = \sqrt{\frac{\sum_{i=1}^m (P_{ij} - \bar{P}_j)^2}{m - 1}}$$

其中， $\bar{P}_j$ 为第*j*种污染物在某个月份的百分比成分谱均值； $m$ 为该月天数。 $S_j$ 为第*j*种污染物在某个月份的百分比成分谱标准差。

##### (3) 根据平均值与标准差，计算出该城市当月各种污染物特征值标准值与上下限；

$$Max_j = \frac{\bar{P}_j + S_j}{\bar{P}_j}$$

$$Min_j = \frac{\bar{P}_j - S_j}{\bar{P}_j}$$

其中，标准值为一定时间该地区各项污染物百分比成分谱与平均污染特征成分谱的比值，值为 1；上（下）限为一定时间该地区各项污染物百分比成分谱加上（减去）其标准差与平均污染特征成分谱的比值。

#### (4) 特定场景下的应用

$$C_j = \frac{P_j}{\bar{P}_j}$$

对某一城市某特定日期的污染物数据，对每种污染物计算百分比成分谱后，除以对应的均值，得到每种污染物当日的特征值 $C_j$ ；并与其对应的上下限进行比较，得到污染类型。

#### 4.1.2 大气污染类型（成因）分析

根据特定地点特定日期每种污染物的特征值与其标准值、上下限的比较，结合实际经验，对污染类型进行分类。对 PM2.5, PM10, SO<sub>2</sub>, NO<sub>2</sub>, CO 这五种污染物的主要来源进行分析。PM2.5 与 PM10 主要来自扬尘；SO<sub>2</sub> 主要来自燃煤、钢铁等工业过程；NO<sub>2</sub> 主要来自机动车驾驶，燃煤、玻璃等工业过程；CO 主要来自燃煤、机动车、钢铁和生物质等的不完全燃烧。根据污染物特征值的不同组合，将大气污染类型分为以下几类。

- 标准型：各污染物特征值均未超出污染物特征值上下限，污染特征未发生显著变化。
- 偏二次型：PM2.5 超出上限，表明污染特征受二次颗粒物生成影响显著。
- 偏沙尘型：PM10 超出上限。
- 偏机动车型：NO<sub>2</sub> 与 CO 特征值超出上限，表明污染特征受机动车影响显著。
- 偏燃煤型：SO<sub>2</sub> 特征值明显超出上限，表明污染特征受燃煤排放影响显著。
- 偏烟花型：PM2.5 和 SO<sub>2</sub> 特征值明显超出上限，结合具体情况分析污染特征受烟花燃放过程影响。
- 偏钢铁型：SO<sub>2</sub>、NO<sub>2</sub>、与 CO 特征值超出上限，表明污染特征受工业排放过程影响。
- 其他型：其他情况。没有明显的污染物特征，结合具体情况加以分析。

#### 4.2 基于污染物特征值的聚类算法

聚类分析指将物理或抽象对象的集合分组为由类似的对象组成的多个类的分析过程，是一种探索性分析。在分类的过程中不必事先给出一个分类的标准，直接从样本数据出发进行分类。

由于天气数据复杂多变并且不同时间不同地区的污染物浓度以及大气情况数值差异显著，难以从中分析出空气污染和大气情况的变化特征。于是，本作品通过 4.1 章中介绍的方法，基于污染物特征值进行聚类，尽可能降低类内差异从而获得一个比较好的聚类结果。聚类的结果可以提供多个可能的解，选择最终解需要研究者的主观判断和后续的分析。



本作品采用了高斯混合模型（GMM）来对数据的特征值进行聚类。任意形状的概率分布都可以用多个高斯分布函数去近似，也就是说 GMM 就是由多个单高斯密度分布（Gaussian）组成的。对于天气数据这种类间数量不平衡的数据，采用数学模型对其进行拟合更加符合数据集的特征。对于聚类的结果采用 T-SNE 降维至二维空间进行输出，并采用合适的编码对信息进行映射，用来辅助分析城市的污染情况。

### 4.3 污染物扩散模型与风廓线

#### 4.3.1 污染物扩散模型

风的存在使污染物会从某一点开始扩散。假设每 15 公里的检测点作为污染物扩散的点源，并且污染物在地面进行扩散。采用高斯羽烟模型计算污染物的扩散情况，对于地面点源，有效源高度  $H = 0$ ，当污染物到达地面后背全部反射，可令  $H = 0$  得到地面连续点源的扩散公式：

$$C(x, y, z) = \frac{q}{\pi u \sigma_y \sigma_z} \exp \left[ -\frac{1}{2} \left( \frac{y^2}{\sigma_y^2} + \frac{z^2}{\sigma_z^2} \right) \right]$$

其中， $C$  为污染物浓度（单位： $kg/m^3$ ）； $q$  为源强（单位： $kg/s$ ）； $u$  为泄露高度的平均风速（单位： $m/s$ ）； $\sigma_y, \sigma_z$  分别用浓度标准偏差表示的  $y$  轴、 $z$  轴上的扩散参数； $H$  为泄露有效高度（单位： $m$ ）。

大气稳定度是指大气中的某一气团在垂直方向上的稳定程度。大气稳定度与天气现象、时空尺度及地理条件密切相关，其级别的准确划分非常困难。根据数据所包含的信息，通过该监测点的风速判断该点的大气稳定程度以此来确定  $\sigma_y = ax^b, \sigma_z = cx^d$ 。

将风速由 [0,2,3,4,5,6] 分成若干个区间将大气稳定度映射到 A,B,C,D,E,F 六个等级，分别对应不同的参数选择（表 2）。

表 2 大气稳定程度与参数选择

稳定度	a	b	c	d
A	0.527	0.865	0.28	0.90
B	0.371	0.866	0.23	0.85
C	0.209	0.897	0.22	0.80
D	0.123	0.905	0.20	0.76
E	0.098	0.902	0.15	0.73
F	0.065	0.902	0.12	0.67

对一段时间内的特定经纬度点进行设置来获取污染数据，计算污染物扩散的罗盘布点数值，将污染物浓度映射为颜色，将每一时间的污染物浓度映射的颜色叠加到罗盘方位图中形成污染物扩散散点图。

#### 4.3.2 风廓线

为了结合污染物扩散散点图的信息对污染扩散情况进行分析，作品中将风向风速向量分解到北、东北、东、东南、南、西南、西、西北 8 个方向的坐标轴上，将每一时间点的每个方向的风速进行叠加制作了风廓线雷达图，借此来展示这段时间内风向的大致情况；并以日为粒度和以小时为粒度的视图中风廓线来展示未来 7 天或 24 小时风向和风速的整体情况。方便研究人员对污染物传播的研究和预算，对环境预警具有明显的帮助。

#### 4.4 污染物之间、污染物与气象因素相关性

Pearson 相关系数表明了两个因素之间的线性相关关系。相关系数越接近 1 代表两者之间正相关性越强，越接近-1 代表两者间负相关性越强，越接近零则相关性越弱。表 3 是取值范围与对应的相关强度。

表 3 相关系数与相关强度对应表

相关系数	相关强度
1.0 - 0.8	极强正相关
0.8 - 0.6	强正相关
0.6 - 0.4	中等正相关
0.4 - 0.2	弱正相关
0.2 - 0.0	极弱正相关或不相关
0.0 - -0.2	极弱负相关或不相关
-0.2 - -0.4	弱负相关
-0.4 - -0.6	中等负相关
-0.6 - -0.8	强负相关
-0.8 - -1.0	极强负相关

本作品利用各个城市 2013 年 1 月 1 日到 2018 年 12 月 31 日的 PM2.5、PM10、SO<sub>2</sub>、NO<sub>2</sub>、CO、O<sub>3</sub> 六种污染物和温度（TEMP）、相对湿度（RH）、大气压强（PSFC）与风速（SPEED）（由 U、V 计算得出）五项气象因素的日均值计算得到了该城市污染物与污染物之间、污染物与气象因素之间的相关性。

##### 4.4.1 计算相关性

###### (1) 计算每个城市的日均风速值

$$speed = \sqrt{v^2 + u^2}$$

其中，*speed*为某城市某天的风速日均值，*v*为某城市某天的经向风速日均值，*u*为某城市某天的纬向风速日均值，单位*m/s*。

###### (2) 计算污染物间、污染物与气象因素间的相关性

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

其中， $r$ 为某个城市因素 $X$ 与因素 $Y$ 的皮尔森相关性， $X_i$ 为因素 $X$ 在时间 $i$ 的日均值， $Y_i$ 为因素 $Y$ 在时间 $i$ 的日均值， $n$ 为 2013 年 1 月 1 日到 2018 年 12 月 31 日的天数。

4.2.2 污染物间、污染物与气象因素间相关性分析

表 4 为普遍情况下污染物间、污染物与气象要素间相关强度。

表 4 各项指标相关性分析结果

	PM2.5	PM10	SO <sub>2</sub>	NO <sub>2</sub>	CO	O <sub>3</sub>
PM10	极强正相关					
SO <sub>2</sub>	强正相关	强正相关				
NO <sub>2</sub>	中等正相关	强正相关	中等正相关			
CO	强正相关	强正相关	强正相关	强正相关		
O <sub>3</sub>	中等正相关 (低纬度地区或西藏) / 弱负相关	中等正(低 纬度地区或 西藏) /弱负 相关	中等正(低 纬度地区或 西藏) /负相 关	中等负相关	中等负相关	
TEMP	中等负相关	中等负相关	中等负相关	中等负相关	中等负相关	强/中等正相 关/弱负相关 (低纬度地 区或西藏)
RH	中等负相关	中等负相关	中等负相关	弱负相关/弱 正相关	不相关	中等负相关
PSFC	中等正相关	中等正相关	中等正相关	中等正相关	中等正相关	中等负相关
SPEED	不相关	不相关	不相关	极弱负相关	极弱负相关	不相关

## 五、可视化与交互设计

### 5.1 系统总体介绍

本作品系统主要分为两个页面（图 2），分别主要以日均值数据与小时值数据作为可视分析系统的输入，通过切换按钮实现页面切换。作品中设计实现了多种视图，辅以合理的视觉编码，同时支持灵活直观的交互设计，支持用户开展对大气污染时空分布模式、时空演变态势、污染成因分析、污染类型分析、污染差异对比分析以及污染改善情况分析等分析任务。



图 2 系统概览图

本可视分析系统主要采用浅色为底，以 AQI 标准等级颜色为基础，根据实际展示效果对系统进行配色；使用颜色准确编码数值，并辅以灰白边框区分系统中不同视图。

#### 5.1.2 日均数据分析页面概览

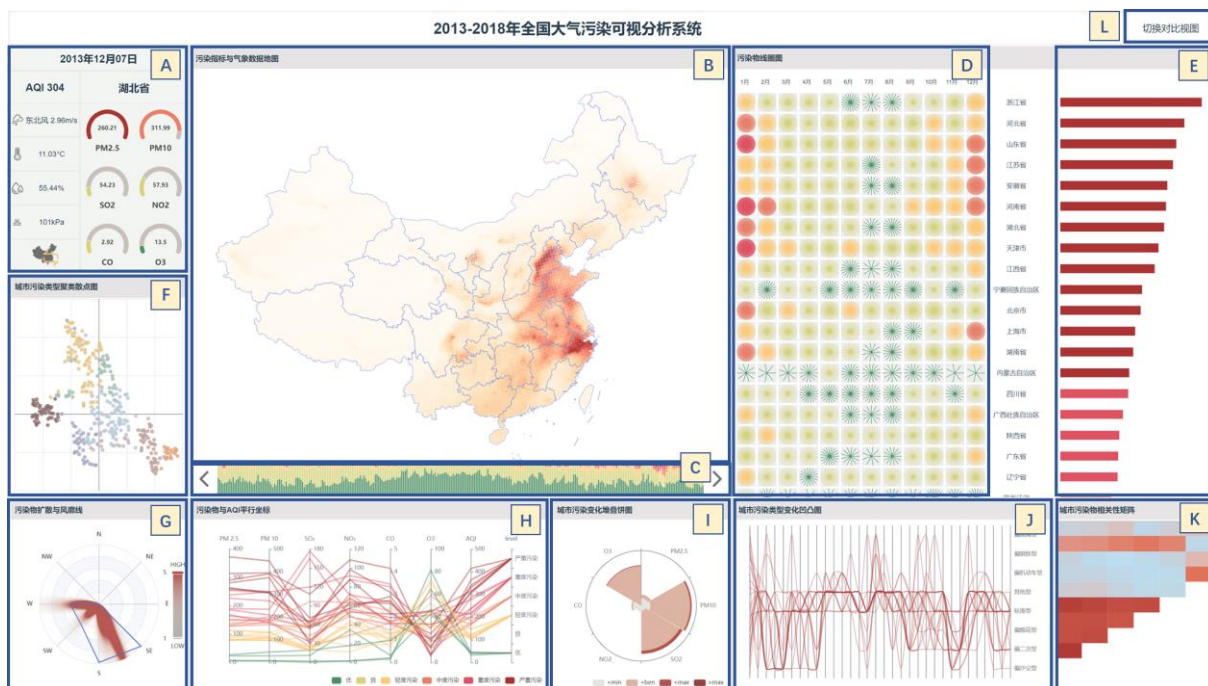


图 3 日均数据分析页面概览

日均数据分析页面中，主要包括控制区域与 11 个视图区域。

控制区域包括页面控制台（图 3A）与页面切换按钮（图 3L）。页面控制台作为当前页面的主要交互入口之一，页面切换按钮实现对两个页面的切换。

视图区域包括污染指标与气象因素地图（图 3B）、全国全年污染等级统计条（图 3C）、污染物线圈图（图 3D）、AQI 排行条形图（图 3E）、城市污染类型聚类散点图（图 3F）、污染物扩散与风廓线（图 3G）、污染物与 AQI 平行坐标（图 3H）、城市污染变化堆叠饼图（图 3I）、城市污染变化类型凹凸图（图 3J）、以及城市污染物相关性矩阵（图 3K）。

本页面主要基于日均值数据，对大气污染时空分布模式、时空演变态势、污染成因与污染类型分析、以及污染物扩散、污染指标相关等分析任务提供支持。

### 5.1.2 小时数据分析页面概览

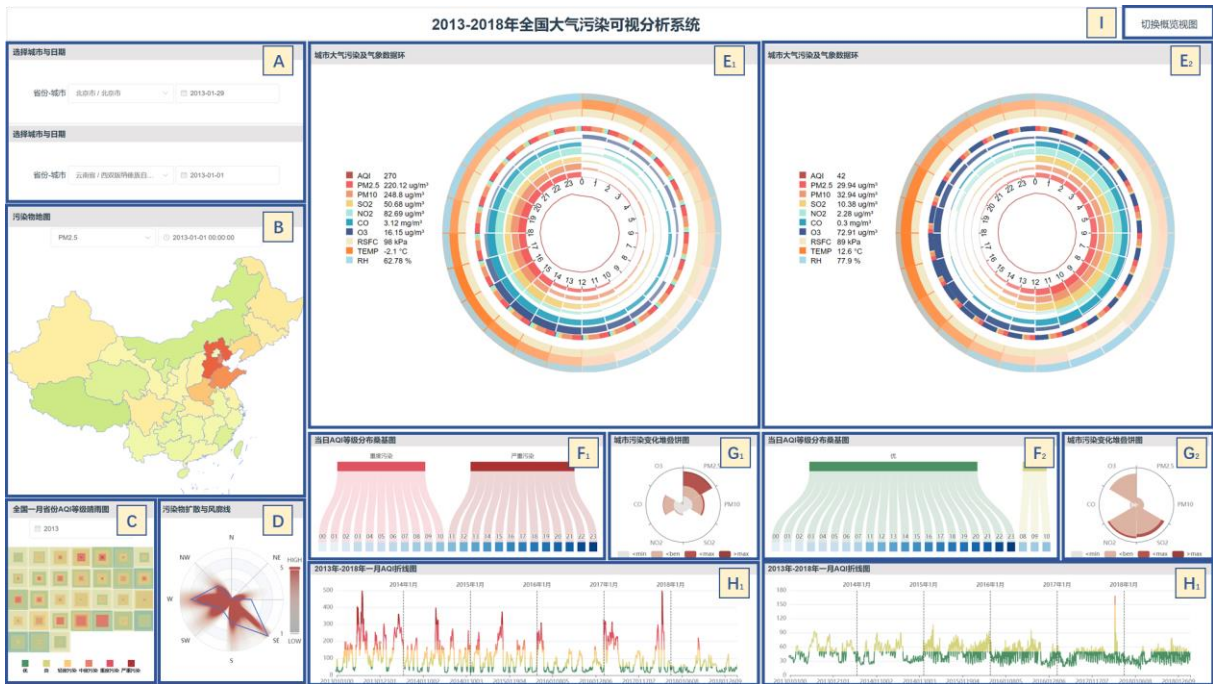


图 4 小时数据分析页面概览

小时数据分析页面中，以两个模块的视图组合形成对比视图，辅以控制区域与普通视图，对小时数据做更全面详细的分析。

此页面上，除了同样用于页面切换的控制按钮（图 4I）外，本页面还包含对对比视图信息选择的选择面板（图 4A），对两个比较视图的数据做配置。

对每个选定的城市或日期，使用四个视图对其数据进行刻画。分别是城市大气污染气象数据环（图 4E）、当日 AQI 等级分布桑基图（图 4F）、城市污染变化堆叠饼图（图 4G）、与 2013-2018 年一月 AQI 折线图（图 4H）。

除了对比视图，本页面还包括使用小时数据的污染物地图（图 4B）、全国一月省份 AQI 等级晴雨图（图 4C）、与污染物扩散与风廓线（图 4D）。

本页面作为第二页面，在日均数据分析页面的基础上，转向对细粒度的小时数据的分析，进一步实现大气污染是时空分布模式、时空演变态势、污染差异对比分析与污染改善情况分析等分析任务。

## 5.2 详细视图与交互设计

本节将详细介绍系统中各视图与交互的设计。

### 5.2.1 日均数据分析页面控制台



图 5 日均数据分析页面控制台

#### (1) 视图设计

日均数据分析页面控制台对当前日期、用户在地图上所指向的地区的名称、AQI 指数、四项气象因素数值信息（图 5A、B）使用文本展示；对六项污染指标数值使用仪表盘展示；同时包含地图视图切换按钮（图 5D）。

#### (2) 交互设计

此控制台的数据，随用户鼠标在地图上的指向动态更新。

其次，用户可以通过点击六项污染物指标或四项大气因素（图 5B、C），切换其他视图（如地图）展示的数据。点击地图切换按钮（图 5D），将地图切换为城市污染类型分类地图。

### 5.2.2 污染指标与气象因素地图、污染类型分类地图

#### (1) 视图设计

本系统中地图根据输入数据类型不同，分为三种地图：污染指标地图、气象因素地图与污染类型分类地图。其中，污染指标地图与气象因素（除风速）地图（图 6）均采用散点的形式，将数据中每个坐标点的数据进行渲染，根据不同的颜色映射，反映数值大小。风速地图使用颜色与线段长度映射风速大小，线段的方向映射风的方向。污染类型分类地图使用分块填充颜色，映射不同的污染类型。



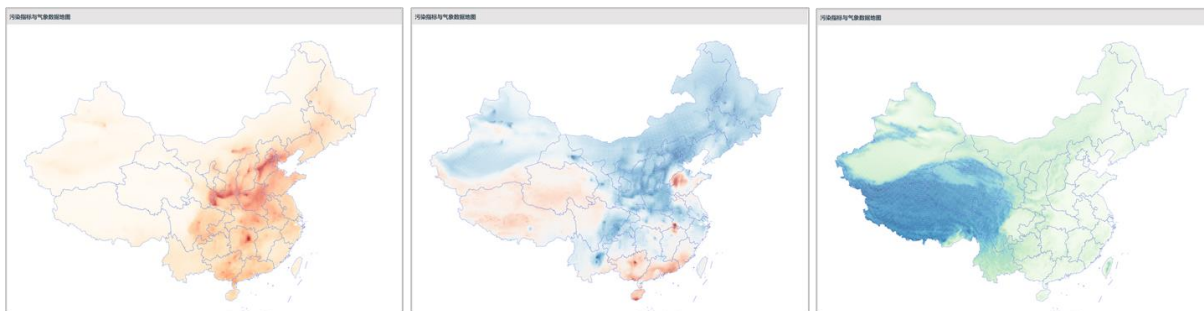


图 6 污染指标与气象因素地图（左：PM2.5 中：O<sub>3</sub> 右：PSFC）

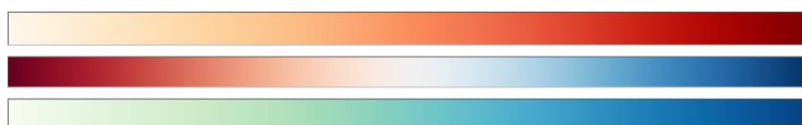


图 7 污染指标与气象因素使用的颜色映射

根据污染指标与气象因素的类型，地图中分别使用 D3.js 提供的三种不同的颜色方案：对 PM2.5、PM10、SO<sub>2</sub>、NO<sub>2</sub>、CO，使用红色单色调渐变（图 7 上）；对 O<sub>3</sub>、气温（TEMP）使用由蓝色到红色渐变（图 7 中），分别映射数值由低到高；对相对湿度（RH）、大气压强（PSFC）使用绿色到蓝色渐变（图 7 下），分别映射数值由低到高。

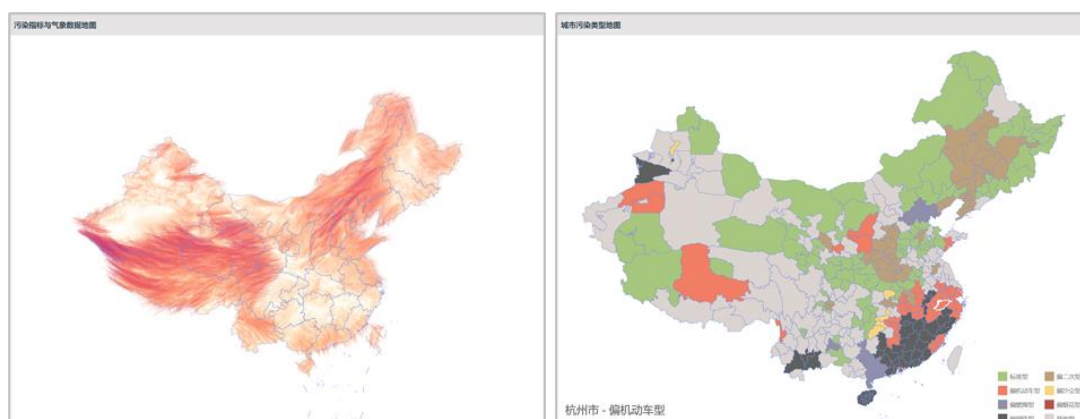


图 8 风地图（左）与城市分类地图（右）

风地图（图 8 左）中，使用线段展示当前坐标点的风：线段颜色与长度映射风速大小，直线方程映射风的方向。

城市分类地图（图 8 右）中，以城市为单位，用颜色分块填充，映射城市当日的污染类型分类结果。

## (2) 交互设计

视图内的交互（图 9）。每个地图均设计并实现了鼠标滑过更新控制面板数据、鼠标点击视图下钻、点击空白处返回等交互操作；同时支持拖拽、大小缩放等鼠标事件。

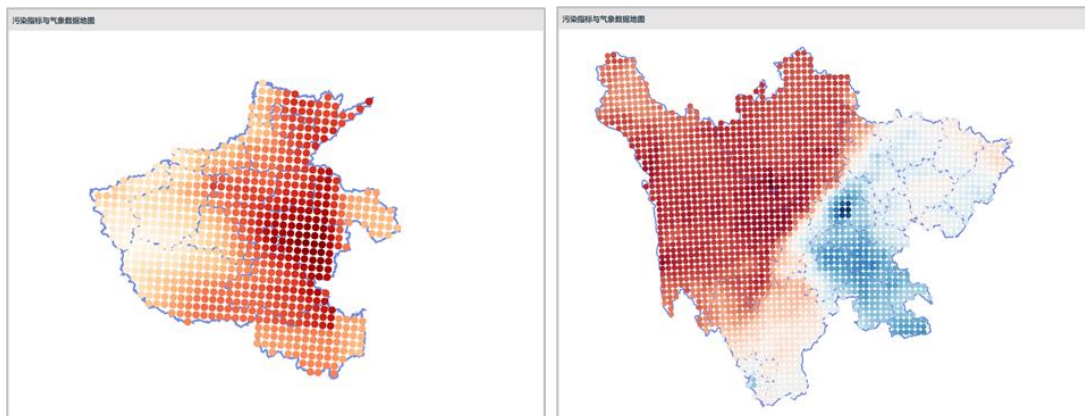


图 9 地图下钻及缩放交互

视图间的交互。地图作为该页面上切换行政区划或坐标点的主要入口。页面中其他视图，包括污染物线圈图与 AQI 排行条形图、污染物扩散与风廓线、污染物与 AQI 平行坐标、城市污染变化堆叠饼图、城市污染物相关性矩阵的数据均与其关联并实时更新。

### 5.2.3 全国全年污染等级统计条

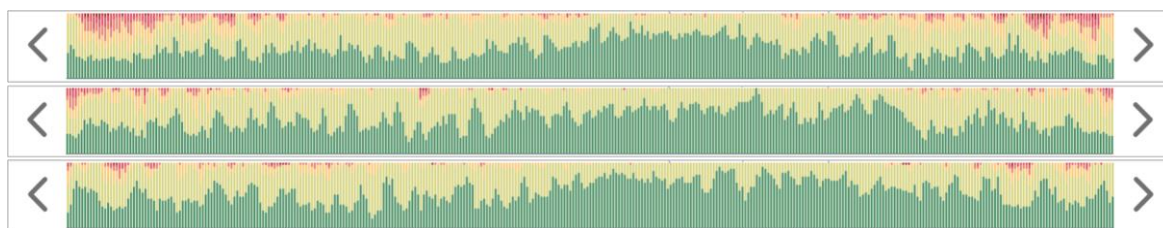


图 10 全年污染等级统计条（上：2013 年 中：2016 年 下：2018 年）

#### (1) 视图设计

全年污染等级统计条（图 10）使用堆叠柱状图展示某一年全国所有日期中全国属于六种污染等级的省份的数量，使用不同的颜色映射不同污染等级，从而展示全年全国省份污染等级的统计情况及全年整体变化。

#### (2) 交互设计

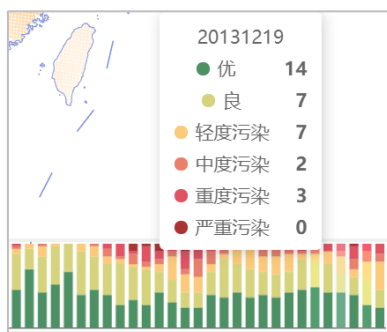


图 11 全年污染等级统计条缩放与信息展示



该视图作为页面中时间切换的主要入口，设计并实现了点击事件：点击每一个柱形图，页面中其他视图会根据对应的日期重新请求数据并更新视图。鼠标悬浮时展示日期与当日全国大气污染属于不同等级的省份数量。同时该视图支持鼠标滚轮缩放、拖拽等操作（图 11）。

#### 5.2.4 污染物线圈图与 AQI 排行条形图

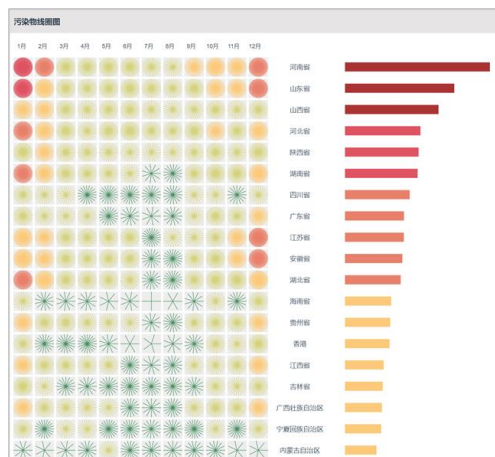


图 12 污染物线圈图（左）与 AQI 排行条形图（右）

##### (1) 视图设计

污染物线圈图（图 12 左）用于展示某一年全国某污染指标月均值的变化情况。以大气中颗粒物为圆形、以颗粒物直径衡量颗粒物大小的背景为灵感，线圈图中，使用线段的条数映射该污染指标的数值，使用线段的颜色映射当月平均 AQI 对应的污染等级；从而展示全国各省份（或全省各城市）在当年不同月份当前污染指标的变化情况。线圈中线段数量越多，表示污染的浓度越高。

AQI 排行条形图（图 12 右）使用矩形长度映射当前日期全国各省份（或全省各城市）的 AQI，使用颜色映射对应的 AQI 等级，并按照 AQI 从高到底排列。

线圈图中展示的顺序与 AQI 条形图的顺序一致。

##### (2) 交互设计



图 13 污染物线圈图交互（左：PM2.5 中：NO<sub>2</sub> 右：O<sub>3</sub>）

该视图本身提供了鼠标悬浮展示信息、滚轮上下滑动等交互。此视图随页面中其他视图更新年份、地区、污染物而实时更新（图 13）。

### 5.2.5 城市污染类型聚类散点图

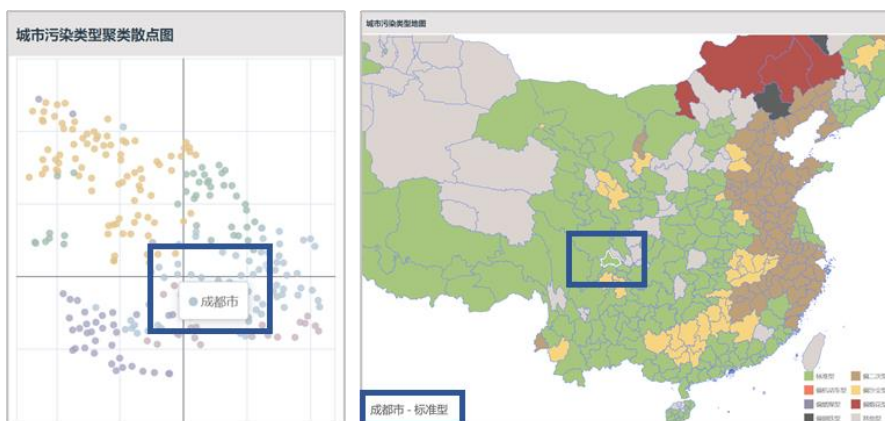


图 14 城市污染类型聚类散点图与交互

#### (1) 视图设计

城市污染类型聚类散点图使用基于污染指标每日特征值数据进行聚类结果、并使用 T-SNE 降维后得到的结果作为输入，将所有城市作为散点映射到坐标轴中，使用颜色映射不同的类别。

#### (2) 交互设计

该视图随全年污染等级统计条中切换日期而更新；点击视图中某一城市对应的散点，与污染类型地图上该点进行对应（图 14）。用户可以查看两种不同的对城市污染类型进行分类的结果的异同，寻找污染类型相似的城市，结合其他视图分析污染成因。

### 5.2.6 污染物扩散与风廓线

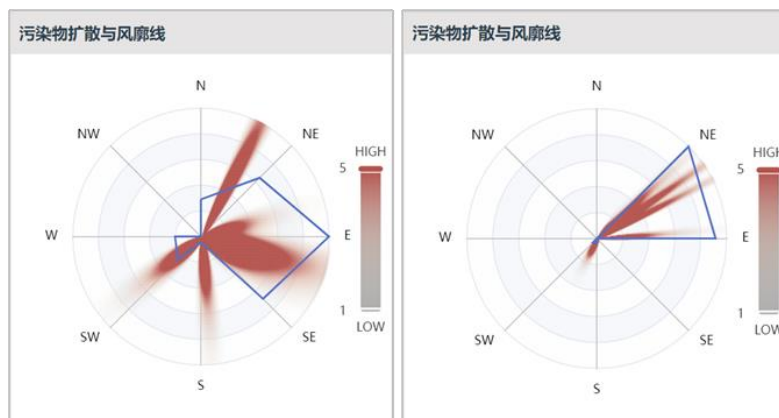


图 15 污染物扩散与风廓线

## (1) 视图设计

污染物扩散与风廓线图展示某坐标点从某时间点开始一段时间内（日均数据分析页面取未来 7 天，小时数据分析页面取未来 24 小时）的污染物扩散情况与风廓线。对用户选择的特定左边点，通过散点的颜色映射污染物浓度大小。在每个方位轴上标定对应方位的风速大小，并使用线段顺序相连构成风廓线。可以看出在一定风与污染物浓度的条件下污染物扩散的情况。

## (2) 交互设计

该视图通过用户点击地图下钻后的坐标点、结合页面当前时间动态更新（图 9）。

### 5.2.7 污染物与 AQI 平行坐标

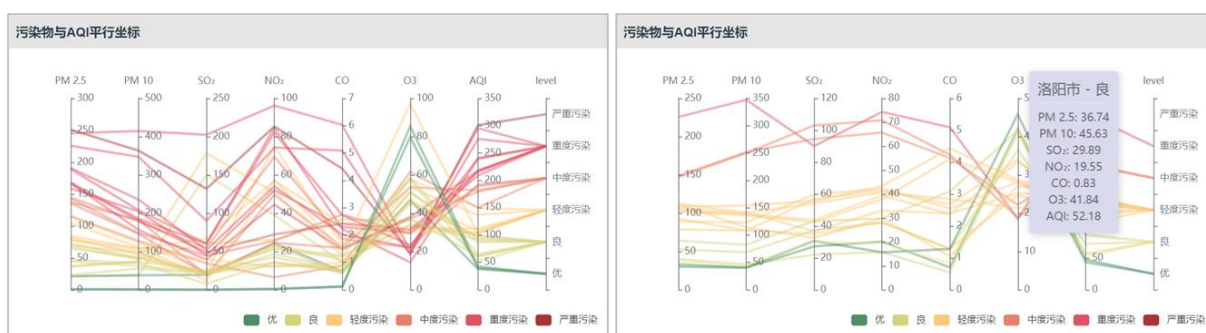


图 16 污染物与 AQI 平行坐标

## (1) 视图设计

污染物与 AQI 平行坐标展示全国各省（或全省各市）的各项污染指标与 AQI、AQI 等级情况。使用颜色映射不同的 AQI 等级。

## (2) 交互设计

该视图在展示全国各省数据时设计并实现了地区下钻操作。点击某省份的折线更新视图为该省份的所有城市的数据。同时支持鼠标悬浮展示信息操作。

### 5.2.8 城市污染变化堆叠饼图

## (1) 视图设计

城市污染变化堆叠饼图（图 17）展示某个城市当天六项污染指标特征值与所在月份特征值的标准值、上限、下限的比较关系。使用 4.1.1 节中计算污染物特征值的方法，将得到的结果与对应标准值、上限、下限进行比较，并映射到不同的数值范围上，使用颜色映射该数值所在的范围。

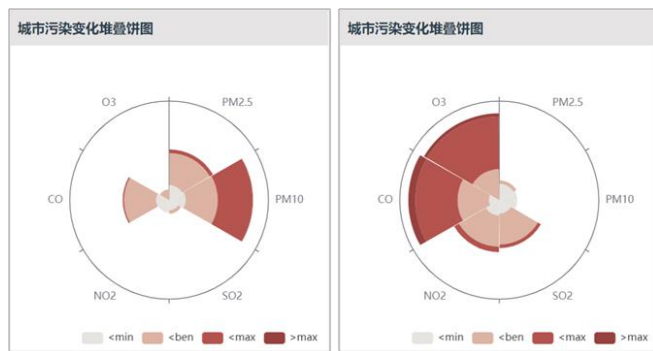


图 17 城市污染变化堆叠饼图

## (2) 交互设计

该视图通过页面其他视图切换城市与日期来更新数据。

### 5.2.9 城市污染类型变化凹凸图

#### (1) 视图设计

城市类型变化凹凸图（图 18）展示一个省份的所有城市在一个月内的污染类型变化情况。使用平行坐标为该视图的原型，每个坐标轴上对应每个日期的八种污染类型。

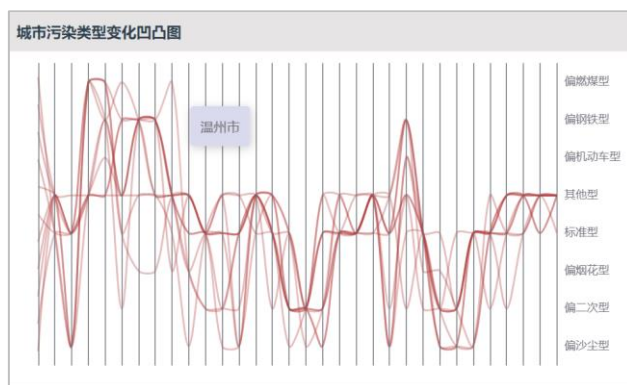


图 18 城市类型变化凹凸图

## (2) 交互设计

该视图通过页面其他视图切换省份与日期来更新数据，并提供鼠标悬浮展示信息。用户可以根据此视图提供的信息做进一步的分析。

### 5.2.10 城市污染物相关性矩阵

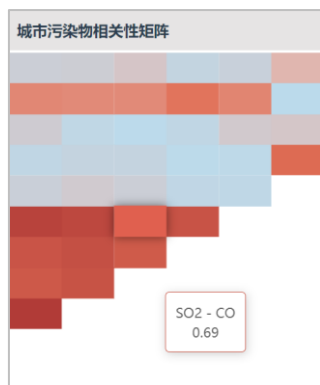


图 19 城市污染物相关性矩阵

城市污染物相关性矩阵（图 19）以城市为单位，展示该城市六项污染指标之间、六项污染指标与三项气象因素之间的相关性。使用颜色编码相关性系数，红色表示正相关、蓝色表示负相关。

该视图通过页面其他视图对城市的切换更新数据，并提供鼠标悬浮展示信息。

### 5.2.11 城市大气污染气象数据环

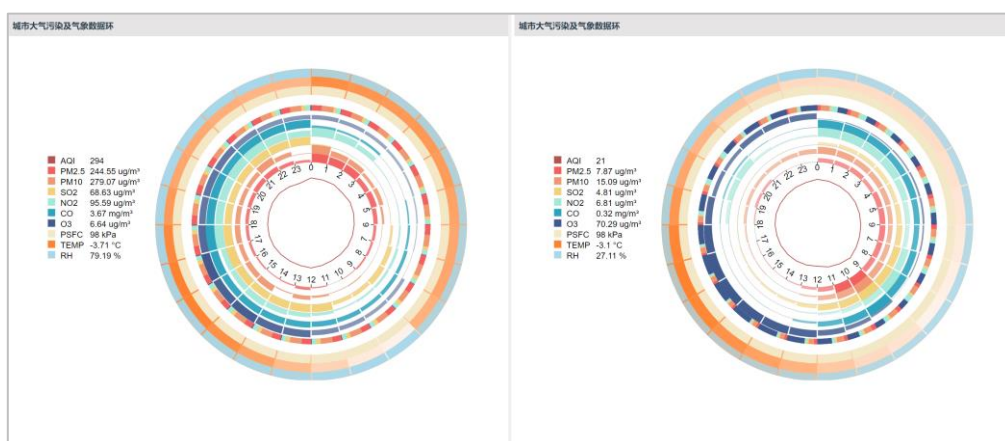


图 20 城市大气污染气象数据环

#### (1) 视图设计

城市大气污染气象数据环（图 20）用于展示某一个城市在某天 24 小时中的六项污染指标、三项气象数据与 AQI 的变化情况。



图 21 城市大气污染气象数据环视觉编码设计



本视图借用时钟的隐喻，采用径向布局，展示 24 小时内数据的数值与变化情况。

环上被平均分为 24 格，每格分别代表该日 0 点-23 点。

视图主要分为四层（图 21），分别采用不同的视觉编码设计。

- 从里向外第一层使用折线图展示每小时 AQI 数值。
- 第二层展示六项污染指标。使用（图 21 左）的设计方案，从里到外一共六层，每层分别对应 PM2.5、PM10、SO<sub>2</sub>、NO<sub>2</sub>、CO 与 O<sub>3</sub>。使用柱状图为基础视图，用颜色映射不同种类的污染物；用矩形高度与颜色透明度映射数值大小。每层的层高一致，将该种污染物当日的最大值与 0 映射到层高与 0，即矩形占扇环 100% 的高度时，代表此时为全天该种污染物的最高浓度。
- 第三层使用堆叠柱状图构成一个环（图 21 右上），每一个区域的弧长代表该种污染物占该小时总污染物的百分比，为此视图补全数据的绝对大小信息。
- 第四层展示气象数据（图 21 右下），分别从内向外、从外向内映射气压（PSFC）与相对湿度（RH），使用颜色透明度与高度映射数值大小；同时使用该区域的背景颜色映射气温（TEMP）。

## (2) 交互设计



图 22 对比视图配置控制台

用户通过对比视图配置控制台（图 22）选择进行比较分析的两个城市与对应的时间，更新对比板块中的多个视图。

环图中支持鼠标指向环上的区域，动态更新展示的数据。点击换上的时间点，更新该页面的其他视图。

### 5.2.12 当日 AQI 等级分布桑基图

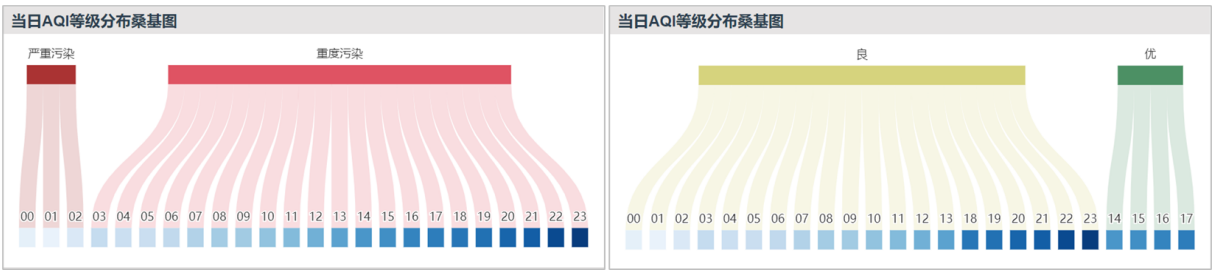


图 23 当日 AQI 等级分布桑基图

当日 AQI 等级分布桑基图（图 23）展示选定的城市在该日期的 24 小时中，每小时 AQI 等级分布与统计情况。使用连续颜色映射小时数由 0 点到 23 点；使用多种颜色映射 AQI 等级。

### 5.2.13 2013-2018 年一月 AQI 折线图

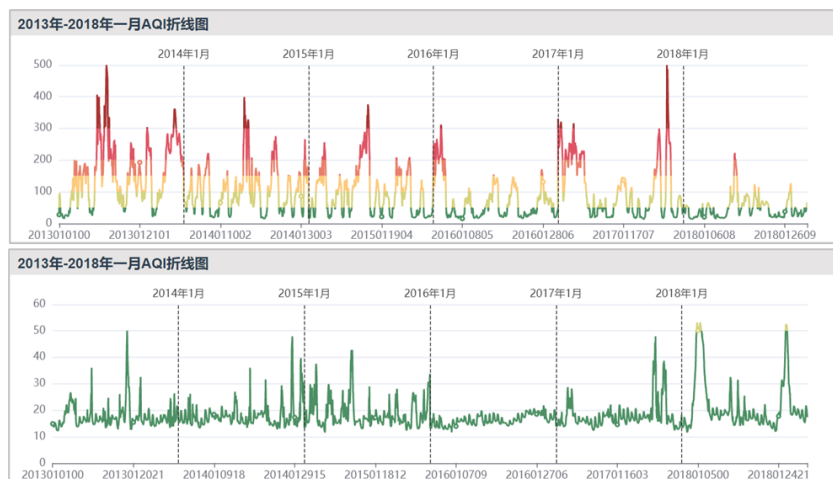


图 24 2013-2018 年一月 AQI 折线图

2013-2018 年一月 AQI 折线图（图 24）展示选定的两个城市在 2013 年-2018 年每年一月按小时计算的 AQI 数值的情况，并展示其变化（改善）情况。使用颜色映射对应的污染等级。

### 5.2.14 全国一月省份 AQI 等级晴雨图

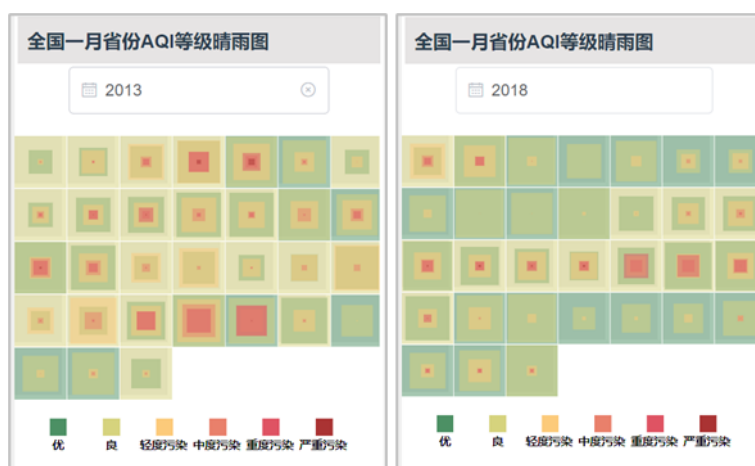


图 25 全国一月省份 AQI 等级晴雨图

全国一月省份 AQI 等级晴雨图（图 25）由北京大学可视化与可视分析实验室提出的疫情可视化晴雨表为设计灵感，并做出改进。使用不同颜色映射不同的污染物等级。使用矩形边长映射全国为该污染等级的省份数量。

用户可以使用下拉框选择对应年份更新视图。

## 六、实验\案例\场景分析

### 6.1 大气污染时空分布与演变态势特点分析

#### 6.1.1 大气污染与季节关联特点

分析发现，全国各省大气污染情况在每年年初与年底较为严重，年中（6-10月）污染情况相对较好：年底与年初全国污染等级为严重污染、重度污染、中度污染的省份数量较多，年中大部分省份为优、良等级（图 26）。

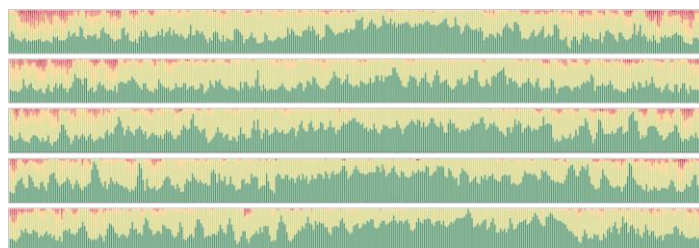


图 26 6 年污染等级统计图

对于单个污染指标而言，该特点仍然存在（图 27）。由污染物线圈图可以看出，对单个省份，在 12 月、1 月、2 月污染物平均值较高（线圈较为密集），并且 IAQI 对应的污染等级较严重（颜色为橘、红）；到一年的中间，如 6-10 月，污染物平均值降低（线圈较为稀疏），污染等级也减轻（颜色出现绿色）。



图 27 PM2.5、PM10、SO<sub>2</sub>2013 线圈图

考虑是由于冬季暖气使用频繁，导致 PM<sub>2.5</sub>、SO<sub>2</sub> 等污染物排放明显。

#### 6.1.2 大气污染与空间分布关联特点

##### (1) 污染指标与气象因素相关性与地域关联特点

通过城市污染物相关性矩阵发现，PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO 之间在绝大部分城市中互为强正相关，在极少的污染很低的地区为弱负相关，如：西藏自治区山南市（相关系数-0.16 左右），西藏自治区日喀则市（相关系数-0.02 左右）等。

O<sub>3</sub> 与其他污染指标与气象因素的相关性随地域不同变化较大。



$O_3$  与温度在全国南方部分地区（海南省、广东省、香港、台湾省、云南省部分城市共 52 个城市）为负相关，在其余 332 个城市均为正相关，且由南往北正相关强度由弱正相关到强正相关（相关系数 0.81 左右）（图 28）。

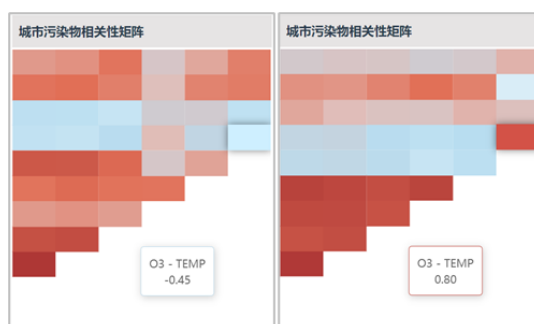


图 28 海南省保亭黎族苗族自治县（左）与河北省衡水市  $O_3$  与温度相关性



图 29 线圈图中污染指标与温度关系（左： $O_3$  中： $SO_2$  右： $PM_{2.5}$ ）

$O_3$  浓度随温度的变化通过线圈图也有所体现（图 29）。随着气温升高， $O_3$  的线圈变得更加密集，浓度增加；随气温降低， $O_3$  的线圈变得稀疏，浓度降低。而对其他污染物有着相反的情况。

全国大部分地区污染指标与气象因素的相关性情况相似，如  $PM_{2.5}$  与温度在不同城市均为正相关；而对于部分北方的地区，包括新疆维吾尔自治区、甘肃省、内蒙古自治区、黑龙江省、吉林省、辽宁省、北京市、天津市、河北省、山西省、山东省，这样的规律不再存在。这些省份中污染指标与气象因素的相关性情况不尽相同。

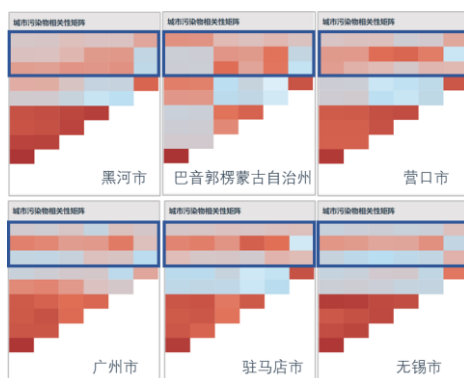


图 30 北方地区与其他地区污染指标与气象因素相关性差异

## (2) 污染指标在风的影响下的扩散

结合气象因素随时间对污染物浓度的变化，可以发现污染物扩散的特点。

河南省在 2013 年 12 月 8 日，许昌市、漯河市、平顶山市相接的地区  $PM_{2.5}$  的浓度较高（图 31 左），呈现出中心到四周、浓度由高到低的状态。观察当天风速图发现，由洛阳、郑州等地的西北风风速较大（图 31 中）；到 12 月 9 日，河南省东部大面积覆盖浓度较高的  $PM_{2.5}$ ，而西部  $PM_{2.5}$  的浓度相对较低（图 31 右）。可以分析是由于 12 月 8 日河南省西部大面积风速较大的西北风导致污染物扩散方向统一，污染物对相邻地区影响明显。

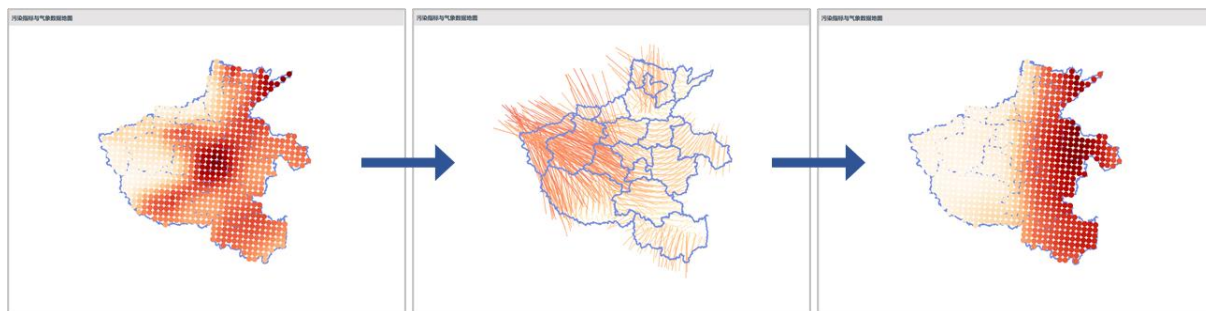


图 31 河南省 2013 年 12 月 8 日-9 日污染物扩散情况

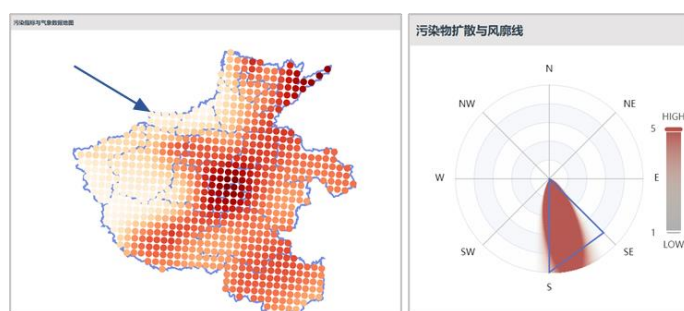


图 32 河南省洛阳市 2013 年 12 月 8 日某坐标点污染物扩散与风廓线

同时，结合污染物扩散与风廓线图可以看到，位于许昌市西北方向的洛阳多处的风均为西北风，导致污染物朝东南方向扩散情况明显。

## 6.2 主要污染源分析

从 AQI 排行榜发现每年河北省空气质量较差（图 33），结合地图可以看到河北省石家庄等市每年冬季  $PM_{2.5}$  污染都较为严重。查阅资料得知， $PM_{2.5}$  主要来源为汽车尾气、冬季烧煤产生的废气以及工业生产产生的废气等。并且，河北省的产业主要以炼钢、石油化工等重工业为主，因此，该地区的  $PM_{2.5}$  相较全国指数偏高。同时考虑到地形因素，河北省中部海拔较低，又紧邻燕山、太行山脉，所以冬季鲜有风，大气污染物难以扩散。

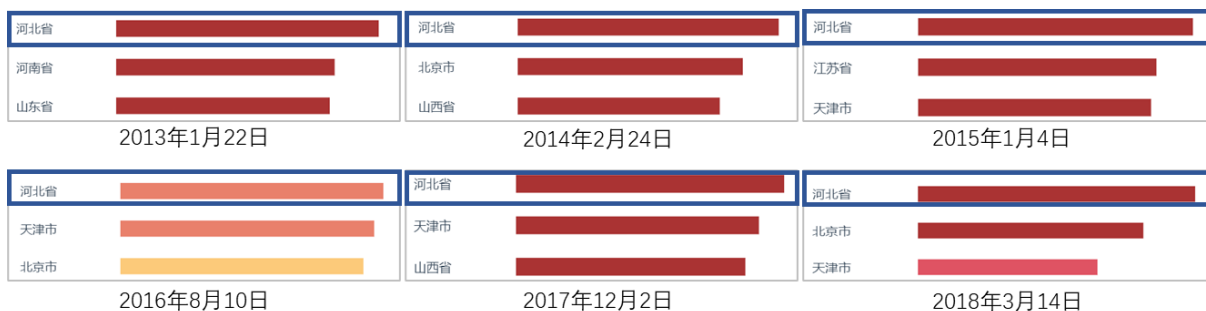


图 33 河北省 AQI 排行榜

同理，通过观察  $\text{SO}_2$  地图发现山西常年  $\text{SO}_2$  污染指数较其他地区偏高。经查阅，山西位于太行山左侧，其阻挡了西北风，使大气污染难以扩散。该省又以煤炭工业为主要产业。因此，该地区  $\text{SO}_2$  指数居高不下。

### 6.3 污染类型分析

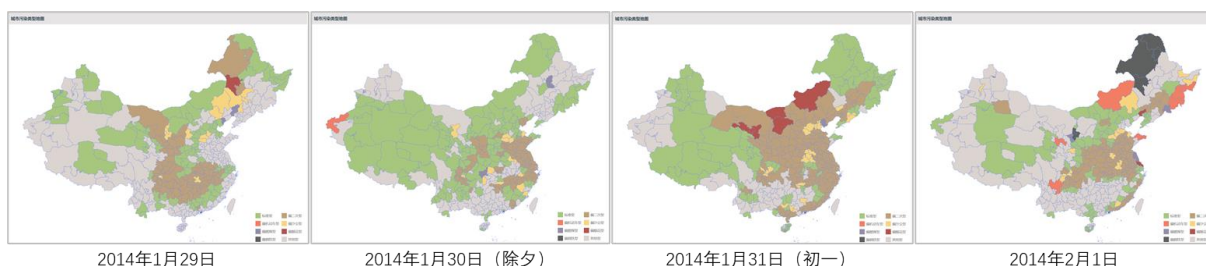


图 34 2014 年春节期间全国城市污染类型变化

由污染物类型地图，在 2014 年 1 月 30 日（除夕），全国各城市类型是以标准型的城市居多；但在 2014 年 1 月 31 日，偏二次型与偏烟花型的城市突然增多，其中以偏二次型为主（图 34）。4.1.2 节中介绍到，受二次颗粒物影响显著为主的污染为偏二次型，偏烟花型主要是指  $\text{SO}_2$  和  $\text{PM}_{2.5}$  指标都超出上限的类别城市，受烟花影响较为严重。结合实际情况分析，国内春节期间会燃放大量的烟花爆竹，导致该时段的偏二次型城市和偏烟花型的城市数量突增，尤其是春节当天达到峰值，在春节过后，这两种污染类型的城市数量逐渐减少。

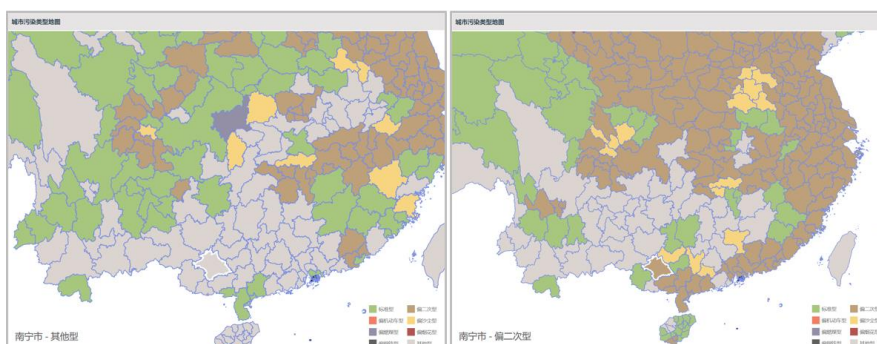


图 35 2014 年 1 月 30 日-31 日南宁市污染类型变化



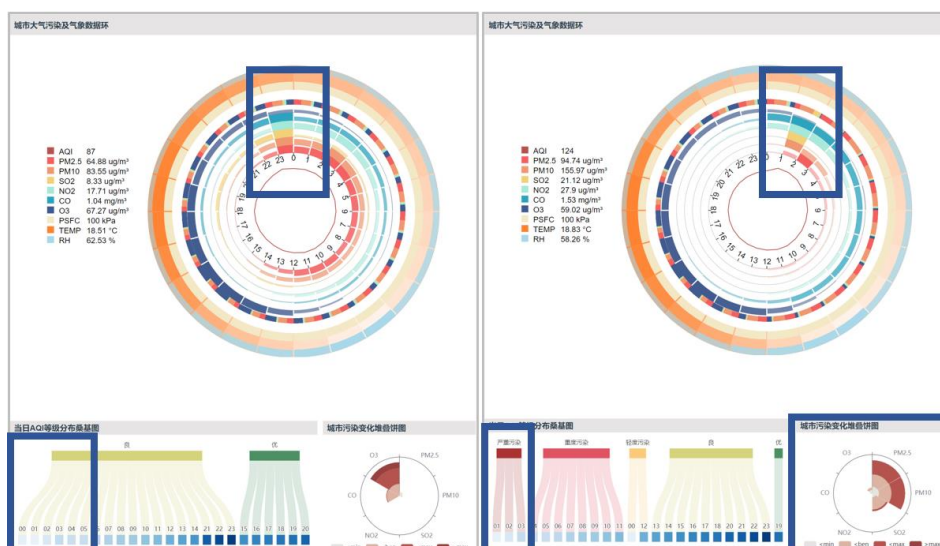


图 36 南宁市 2014 年 1 月 30 日与 1 月 31 日对比分析

进一步分析，选取广西南宁市作为观察城市进行深入研究。由污染类型地图可以看出，在 2014 年 1 月 30 日到 2014 年 1 月 31 日，该时段内南宁市由其他型城市转变为偏二次型城市（图 35），结合小时数据分析页面，对该城市不同小时进行了对比分析。由城市大气污染及气象数据环发现（图 36），在 30 日 23 点及 31 日 1 点到 3 点这段时间内，PM2.5 浓度增大，通过 AQI 等级分布桑基图发现，30 日 21、22、23 时，空气质量由“优”转为“良”，并在 31 日凌晨污染加重，成为“严重污染”，到 31 日白天，污染逐渐减轻。由城市污染变化堆叠图，当日 PM2.5 与 PM10 特征值接近上限，是当时的主要污染物。由于粗颗粒物排放较多，该城市由污染类型由其他型转化为了偏二次型（图 35）。

## 6.4 污染差异与污染改善分析

### 6.4.1 各项污染指标浓度改善：以 PM2.5 为例

通过污染物线圈图（图 37），全国各省份月均 PM2.5 浓度逐年下降（线圈变得稀疏），各省份污染情况也有所改善（绿色居多）。

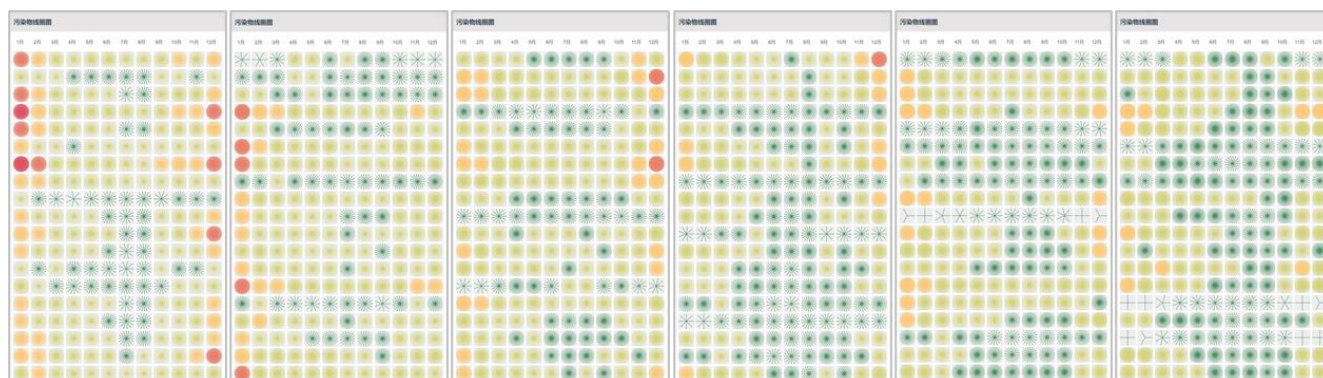


图 37 PM2.5 连续六年省份月均变化线圈图

## 6.4.2 城市 AQI 数值下降

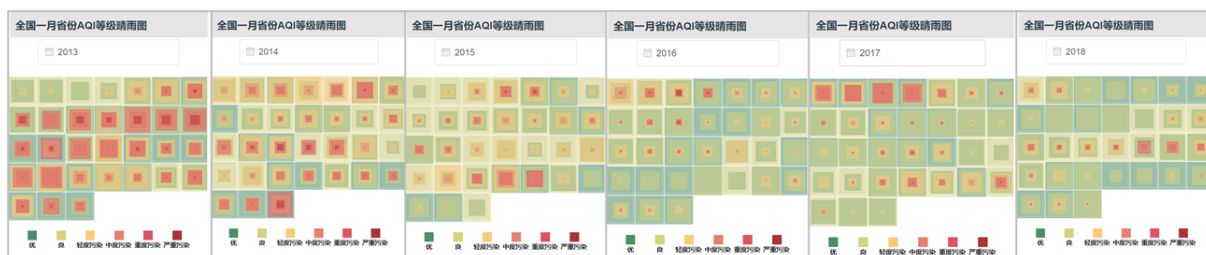


图 38 2013-2018 年 1 月全国省份 AQI 等级晴雨图变化

在 2013-2018 年的一月，从全国范围内省份的 AQI 等级统计情况来看，空气质量有明显改善；更多省份处于“优”“良”等级。

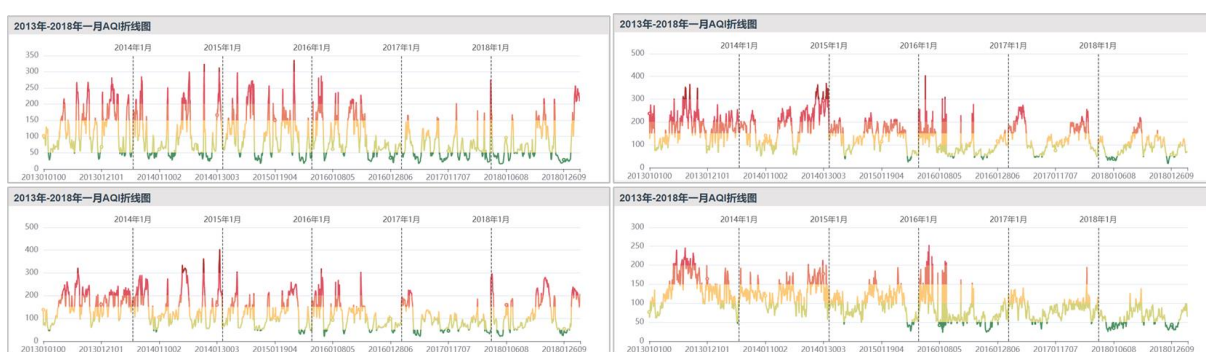


图 39 部分城市每年一月 AQI 折线图（左上：上海市 右上：成都市 左下：南京市 右下：遵义市）

同时，通过 2013-2018 年一月 AQI 折线图（图 39）可以看到，虽然一月为大气污染较为严重的月份，但大部分城市在 2013-2018 年间 AQI 数值有明显的降低；在一月期间处于“优”等级的小时数也明显变多，污染治理成效明显。

## 6.5 特定事件分析

### 6.5.1 2013 年 1 月 14 日内蒙古企业偷排致内蒙古与宁夏 $\text{SO}_2$ 水平异常

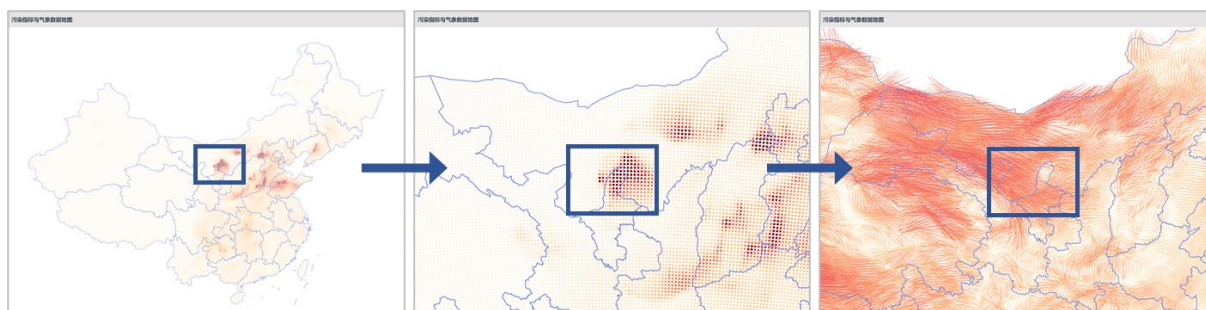


图 40 2013 年 1 月 14 日内蒙古与宁夏  $\text{SO}_2$  水平异常分析

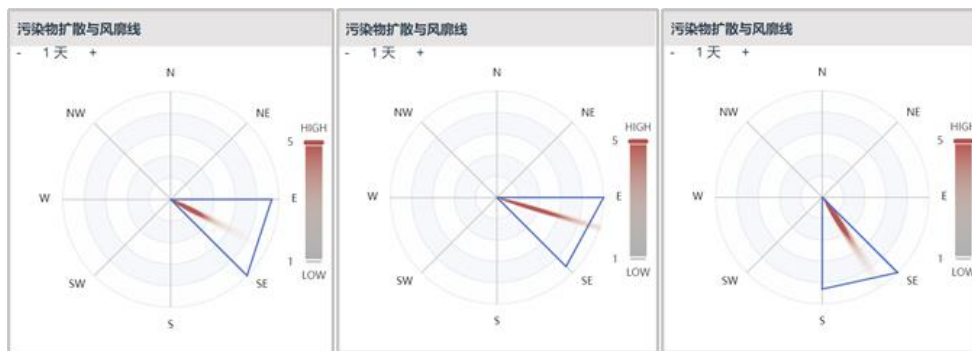


图 41 内蒙古部分点风向与污染物扩散情况

由 2013 年 1 月 14 日  $\text{SO}_2$  地图发现，内蒙古南边靠近宁夏的区域  $\text{SO}_2$  水平几乎为全国最突出的点。放大地图后发现，相比较于前一天，该地点  $\text{SO}_2$  突然升高，并且宁夏省银川市和石嘴山市的  $\text{SO}_2$  值也突然升高，超过标准线。结合污染物扩散图，观察出  $\text{SO}_2$  最高点处周围区域的污染物的扩散风向整体趋势偏东（图 41）。查阅资料得知，该  $\text{SO}_2$  最高点为内蒙古太西煤集团兴泰煤化工有限责任公司所在地，并且于 2020 年该公司被查出存在长期偷排废气行为，对周边大气环境造成严重污染，此外，该公司工厂地点较银川市和石嘴山市海拔较高，因此，12 月 7 日当天，受风向和地形的双重影响，银川和石嘴山  $\text{SO}_2$  指标也随之突增。

### 6.5.2 2015 年 11 月 6 日-8 日东北大范围大气污染



图 42 2013 年 11 月 6 日-8 日东三省污染严重

由 2015 年 11 月 6 日  $\text{PM}_{2.5}$  地图发现，东三省空气质量状况不佳。从 7 日开始辽宁省、黑龙江省与吉林省多个城市空气 AQI 指数超过 200。8 日沈阳等多个城市 AQI 达到超过 500。放大地图后结合污染物线圈图与排行榜发现辽阳市  $\text{PM}_{2.5}$  日均值超过 586（微克/立方米），远超  $\text{PM}_{2.5}$  严重污染标准值 250（微克/立方米）。至 11 月 9 日，空气重污染过程仍在持续，后续几日空气质量才逐渐改善。

经查证，环境保护部 2015 年 11 月 10 日向媒体报道，该部门于 11 月 9 日启动重污染天气督查工作。沈阳市环保局于 11 月 8 日下午 15 点 30 分启动了重污染天气一级红色预警，主要采取了重点工业企业减产 40%-50%、建筑工地全面停工、三环内货车禁行等措施。

### 6.5.3 2013 年 12 月 7 日合肥人工降雨影响污染指标



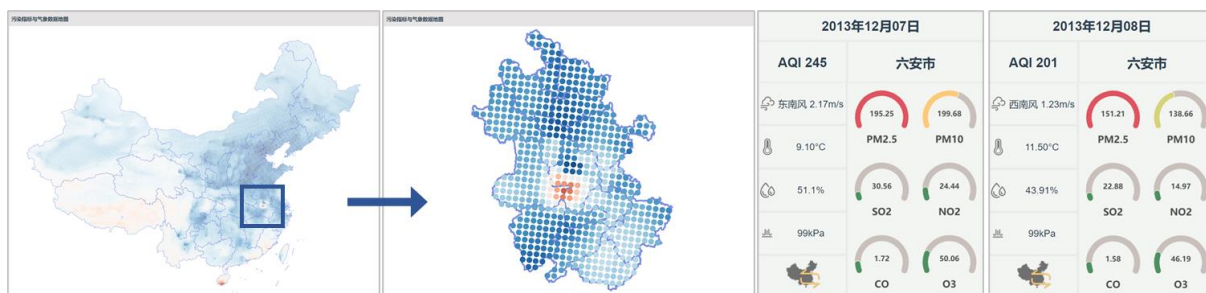


图 43 2013 年 12 月 7 日合肥 O<sub>3</sub> 与其他污染指标变化

在污染指标地图上发现，2013 年 12 月 7 日安徽省合肥市出现了与周围不太一致的较高浓度的 O<sub>3</sub>，并且在此后的一天，该地的多项污染指标都有所下降。结合当日天气预报发现，12 月 8 日与 12 月 9 日合肥市从多日连晴转为了雨。结合人工降雨技术对 O<sub>3</sub> 浓度的影响，以及雨后各项污染物指标都有下降的情况，推测为 12 月 7 日当地进行了人工降雨，导致当地 O<sub>3</sub> 的浓度较高。

### 6.5.3 2014 年 4 月 22 日-26 日新疆遭遇特大沙尘暴

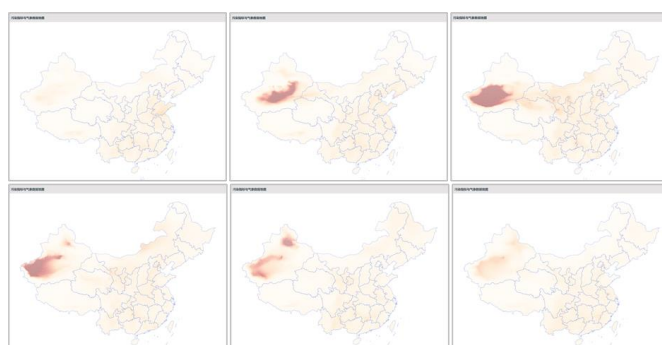


图 44 2014 年 4 月 22 日-26 日污染指标地图新疆 PM2.5 变化

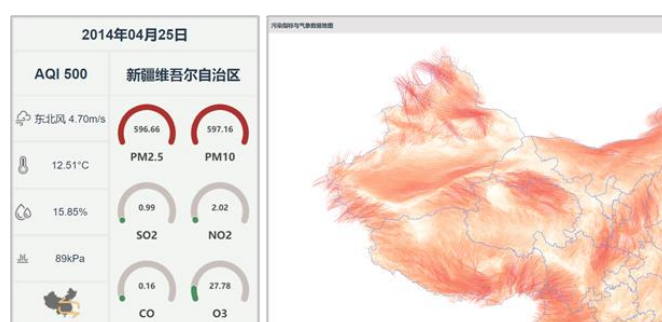


图 45 2014 年 4 月 25 日新疆风向与其他污染指标

观察污染指标与气象数据地图可发现，在 2014 年 4 月 22 日，新疆地区的 PM2.5 与 PM10 开始出现增加，并在 4 月 23 日，地图上出现明显的大面积“红斑”，且在 4 月 24 日达到顶峰，4 月 25 日随较大的东北风，“红斑”开始逐渐消散；一直持续至 4 月 27 日，“红斑”几乎完全消散。查阅该时段新闻发现，受一股较强冷空气自西向东的影响，新疆遭遇特大沙尘暴，并且，该沙尘暴辐射至了甘肃、内蒙古、宁夏等地区。

---

## 七、讨论与总结

本作品使用 2013 年-2018 年中国高分辨率大气污染再分析开放数据集，围绕分析大气污染时空分布模式、时空演变态势、识别主要污染源、分析污染成因、比较污染差异、评估大气环境状况、发现特定事件等分析任务，对数据进行多方面处理，并最终设计实现了全国大气污染可视分析系统，通过案例对系统进行评估。

首先对数据进行逆地理编码、数据清洗、数据统计、IAQI 与 AQI 等级计算等预处理操作。

为对地区污染物组成、主要污染源进行分析，对污染指标数据进行特征值计算，根据污染物产生原因对地区污染类型进行分类；同时基于特征值使用 GMM 聚类，T-SNE 对聚类结果降维，得到城市污染类型结果；应用污染物扩散模型模拟单点污染物扩散情况；并计算了污染指标与大气因素之间的相关性。在分析任务的指导下，基于不同的数据粒度，设计实现了双页面的可视分析系统。

日均数据分析页面针对日均数据做可视分析。使用污染指标与气象数据地图、污染指标线圈图、AQI 排行榜、污染指标与 AQI 平行坐标等视图提供对污染时空分布模式与时空演变态势的分析；并结合灵活丰富的交互，用户利用该页面其他视图做更细粒度的兴趣点分析。对线圈图、污染物扩散图与风廓线、污染变化堆叠饼图提出了创新的设计与实现。

小时数据分析页面主要通过对比模块展开除时空分布模式与演变态势外，对污染情况察隅与改善情况的分析，辅以分块 AQI 地图、小时 AQI 等级晴雨图、污染指标扩散与风廓线视图做兴趣点选取。设计实现了城市大气污染气象数据环，高效准确地表示一天内的多维数据，完成多个城市与日期的对比分析任务。

在使用系统展开案例分析的过程中，发现 2013 年-2018 年间，全国大气污染状况普遍得到改善，多项污染指标下降；全国污染通常在 12 月、1 月集中，并随着气温升高污染相对减轻，可能与冬季暖气使用等有关；同时对于大部分城市，PM<sub>2.5</sub>、PM<sub>10</sub>、SO<sub>2</sub>、NO<sub>2</sub>、CO 之间存在正相关性，而 O<sub>3</sub> 与其他污染指标多为负相关，且与气象因素的相关性在不同地区表现不同；通过视图可以分析出部分污染特点明显的地区，并结合实际情况分析其污染类型的变化与变化的原因。同时系统能够发现或体现大气污染相关的特定事件，例如企业偷拍废气事件、突发大面积污染、人工降雨对污染的影响、气象变化对污染的影响事件等。

本系统在能够完成可视分析任务的基础上，对视觉编码、视图创新与交互上做了相对细致的设计。但对于高维时序大气污染数据，还有众多可以切入的角度，以及与相关领域知识和数据结合，做更深入的分析。