

文章编号: 2096-1618(2023)01-0116-07

基于机器学习对重庆西南部空气质量的综合分析

范颖¹, 吉莉²

(1. 重庆市南川区生态环境监测站, 重庆, 400700; 2. 重庆市北碚区气象局, 重庆 400700)

摘要:利用重庆市南川区 2017–2020 年环境监测数据, 综合采用数理统计方法对空气质量及其大气污染物浓度的特征进行分析, 并基于灰色关联度和 PCA 主成分对空气质量进行分析。结果表明: 南川空气质量及大气污染浓度季节变化明显, 均呈下降趋势; 南川以优良天数为主, 无严重污染过程, 轻度污染占比多于中度和重度污染天数; O_3 对空气质量影响最大是出现在夏季, 其余季节不明显, $PM_{2.5}$ 和 PM_{10} 是除夏季外, 对空气质量影响最大的要素, 特别是冬季; 基于 PCA 主成分分析, PM_{10} 与 $PM_{2.5}$ 是主要污染物, O_3 主要是夏季仅次于 $PM_{2.5}$ 、 PM_{10} 的污染物。

关键词:空气质量; 灰色关联度; PCA 主成分

中图分类号: X823

文献标志码: A

doi: 10.16836/j.cnki.jcuit.2023.01.017

0 引言

随着社会经济的快速发展, 中国大部分城市的大气污染从单一形式向复合污染快速转变。区域性的复合空气污染是以 $PM_{2.5}$ 、 O_3 为主要特征^[1], 对人们的生活、健康都产生了较大的影响^[2–4]。对于这些污染物, 国家环境空气质量标准空气质量指数(AQI)将 6 项污染物(PM_{10} 、 SO_2 、 NO_2 、 $PM_{2.5}$ 、 O_3 、CO)用统一的评价标准呈现。根据中国《环境空气质量指数(AQI)技术规范(试行)》^[5], 将 AQI 对应于空气质量的 6 个级别, AQI 越大, 级别越高, 说明空气污染状况越严重, 对人体健康的影响也越明显^[6]。

国内外学者对空气质量开展了大量的研究^[7–11], 赵颖等^[12]对连云港市空气质量现状及其变化趋势进行了系统分析; 李刚^[13]对 2015 年克拉玛依市空气质量特征进行分析, 结果表明克拉玛依市空气质量整体较为清洁, 优良天数占有效监测天数的 91.5%; 许海超等^[14]根据山东省各地级市的地理位置和不同的季节气候, 分析了空气质量指数 AQI 特征; 陈颖等^[15]以山西省 11 个地级市为研究区域, 针对其空气质量问题, 采用聚类分析和主成分分析方法, 分析影响山西省空气质量的主要因素; 王宝英^[16]选取多个气象因素和历史 AQI 作为 AQI 的影响因素对南昌、南京和合肥的 AQI 进行预测。

近年来, 机器学习作为人工智能领域的重要分支, 越来越多的学者将机器学习技术应用于大气污染方面的研究并取得了较好的成果^[17–20]。主成分分析法(PCA)^[21]和灰色关联度法^[22]是机器学习方法中对多指标综合分析的方法。这些分析方法能够从多水平、多因素着手, 综合分析各指标的整体效应, 使筛选出的结果更具有科学性。目前关于重庆空气质量的研究主

要是针对主城区的单个大气污染物的特征分析^[23]、大气污染的气象因素分析^[24–25]、气象因子对大气污染影响预报^[26–27]等。针对重庆空气质量的综合研究较少。特别是重庆市南川区位于重庆西南部, 属主城都市区, 是渝南黔北的重要交通门户, 因此, 以南川为重庆东南部代表, 分析其空气质量特征并利用主成分和灰色关联度法分析影响空气质量的因素, 为空气质量改善提供科学依据。

1 资料与方法

1.1 资料来源

重庆市南川区位于重庆市西南部, 处于重庆主城区与渝东南武陵山区城镇群、黔北地区的连接带, 是东向、南向出渝大通道的重要节点(图 1)。采用的空气质量数据为南川城区空气自动监测国控点一天监测数据的均值, 即 2017–2020 年 PM_{10} 、 SO_2 、 NO_2 、 $PM_{2.5}$ 、 O_3 、CO、AQI 的日均值。

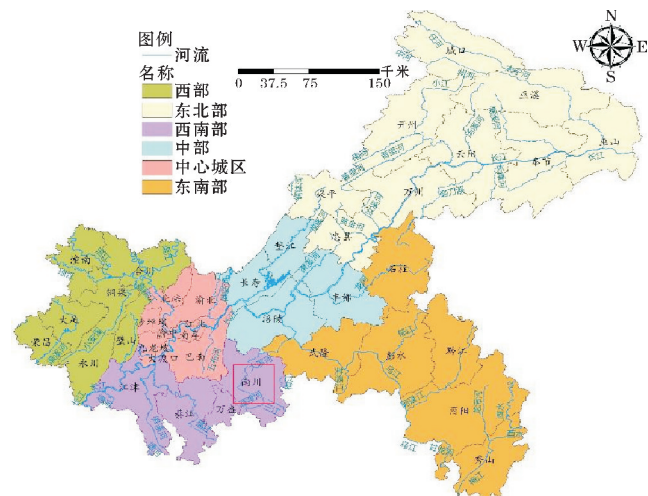


图 1 重庆市南川区区位图

收稿日期: 2021-10-26

基金项目: 重庆市气象局业务技术攻关资助项目(YWJSGG-202134)

1.2 研究方法

1.2.1 空气质量指数

空气质量指数等级如表 1 所示。

表 1 空气质量指数等级

空气质量指数	空气质量指数 级别(状况)	对健康影响情况
0~50	一级(优)	空气质量令人满意,基本无空气污染
51~100	二级(良)	空气质量可接受,但某些污染物可能对极少数异常敏感人群健康有较弱影响
101~150	三级 (轻度污染)	易感人群症状有轻度加剧,健康人群出现刺激症状
151~200	四级 (中度污染)	进一步加剧易感人群症状,可能对健康人群心脏、呼吸系统有影响
201~300	五级 (重度污染)	心脏病和肺病患者症状显著加剧,运动耐受力降低,健康人群普遍出现症状
300+	六级 (严重污染)	健康人群运动耐受力降低,有明显强烈症状,提前出现某些疾病

1.2.2 主成分分析

主成分分析(PCA)^[21]是利用数据降维的思想,在损失较少数据信息的前提下,把多个指标转化成几个为数较少的综合指标的多元分析方法,各个主成分是原始变量的线性组合,彼此互不相关。主成分分析以方差作为信息量的测度,取累计贡献率大的几个成分作为主成分。

首先对原始数据进行标准化,其次建立 R 矩阵。

$$X_i^* = \frac{X_i - E(X_i)}{\sqrt{D(X_i)}} \quad i = 1, 2, \cdots, p$$

其中, X_i 表示第 i 个变量。

计算 R 的特征根 λ_i 以及相应的特征向量 T_i , 然后确定累计方差贡献率决定主成分的个数。

其中方差贡献率为

$$\varphi_i = \frac{\lambda_i}{\sum_1^p \lambda_i}$$

累计方差贡献率为

$$\varphi_k = \frac{\sum_1^k \lambda_i}{\sum_1^p \lambda_i} \quad k < p$$

当累计方差贡献率的百分数较高时,方能确定分析中有 k 个主成分,通常有多少个特征值大于 1 的个数即为主成分个数^[15]。

最后主成分表达式解释了主成分的含义,加权综合主成分表达式按照每个主成分对实际问题的各个样品进行排序及评价。

主成分:

$$W_i = T_i X^*$$

加权综合主成分:

$$W = \sum_{i=1}^k \frac{\lambda_i}{\sum_{m=1}^p \lambda_m} W_i$$

1.2.3 灰色关联度

灰色关联度分析法^[22]是灰色系统理论中一种定量描述因素间发展势态的相似或相异程度的量化比较方法。如果曲线越接近,相应序列间关联度就越大,反之就越小。

(1)建立原始数据的因变量和自变量,并建立相应序列^[28]。

因变量序列: $Y_0 = \{y_0(k), k = 1, 2, \cdots, n\}$

自变量序列: $Y_i = \{y_i(k), k = 1, 2, \cdots, n\}, (i = 1, 2, \cdots, m)$

(2)对各序列进行无量纲化处理,进行变量对比。

$$Y'_0 = \left\{ \frac{y_0(k)}{y_0(1)}, k = 1, 2, \cdots, n \right\}$$

$$Y'_i = \left\{ \frac{y_i(k)}{y_i(1)}, k = 1, 2, \cdots, n \right\}$$

(3)每个被评价对象的自变量与因变量的的绝对差值进行计算。

即 $\Delta i(k) = |Y'_0(k) - Y'_i(k)| \quad k = 1, 2, \cdots, n, ; i = 1, 2, \cdots, m$

(4)确定 $m = \min_{i=1}^m \min_{k=1}^n |y'_0(k) - y'_i(k)|$

$$M = \max_{i=1}^m \max_{k=1}^n |y'_0(k) - y'_i(k)|$$

(5)计算关联系数。

$$r(y'_0(k), y'_i(k)) = \frac{m + \xi \times M}{\Delta i(k) + \xi \times M}, k = 1, 2, \cdots, n$$

其中, ξ 为分辨系数,取值范围(0,1),取值越小,关联系数间的差异越大,区分能力越强,一般取 $\xi = 0.5$ 。

(6)计算关联度。

$$r(Y_0(k), Y_i(k)) = \frac{1}{n} \sum_{k=1}^n r_{oi}(y'_0(k) - y'_i(k))$$

(7)根据关联度得出最终评价结果。

2 结果分析

2.1 大气污染浓度变化特征

2.1.1 大气污染浓度年际变化

图 2 为南川 2017-2020 年大气污染浓度年际变化箱线图,从图 2 可知,污染物的浓度整体呈缓慢下降的趋势,其中 SO_2 下降最为明显,年均值、中位数分别下

降了 $22\text{ }\mu\text{g}/\text{m}^3$ 、 $17\text{ }\mu\text{g}/\text{m}^3$;各个大气污染物浓度在每年的整体趋势基本一致,臭氧浓度的整体箱线均高于其他污染物浓度。2017年的大气污染浓度离群值明显多于其他年份。

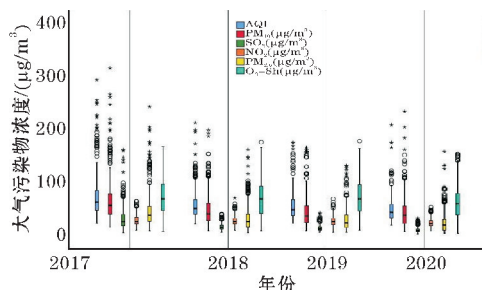


图2 南川2017–2020年大气污染浓度年际变化箱线图

(每个箱线的上缘、上边缘、中线、和下缘分别代表上限、上四分位、中位数、下四分位和下限,○为离群值,*为离群值极值)

2.1.2 大气污染浓度季节变化

图3是重庆市南川区大气污染物浓度季节变化图,

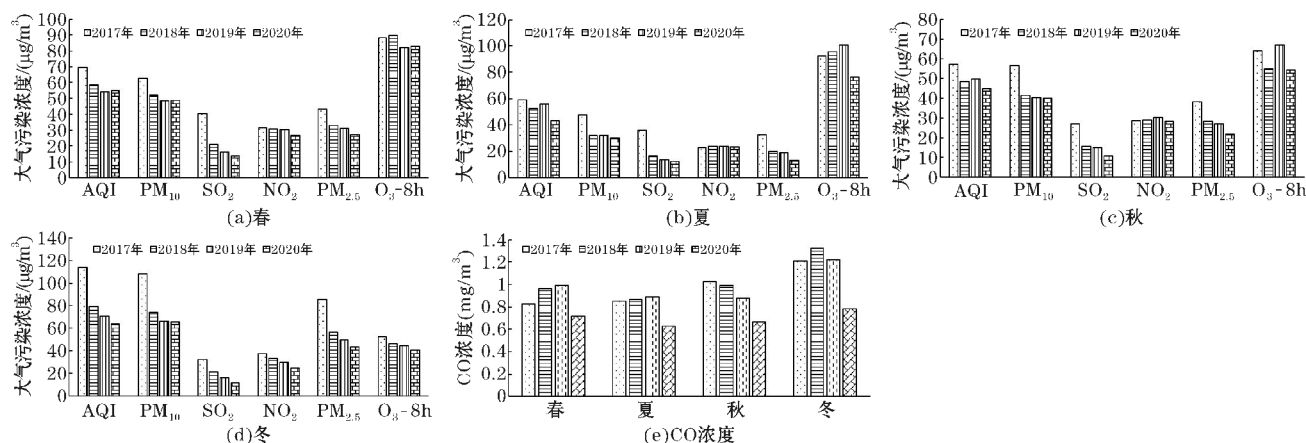


图3 2017–2020年大气污染浓度季节变化趋势图

2.1.3 大气污染月变化

图4是大气污染浓度月变化图,可见 PM_{10} 、 $\text{PM}_{2.5}$ 、 NO_2 月变化呈“U”型,7月值最低,高值出现在两端, PM_{10} 和 NO_2 的高值在12月, $\text{PM}_{2.5}$ 的高值在1月; O_3 则与 PM_{10} 、 $\text{PM}_{2.5}$ 、 NO_2 的月变化相反, SO_2 月变化不明显,12个月的变化相对平缓;AQI和CO的变化较一致,1月和12月为最高值,分别在年中(8月和6月)有一个次高点,最低点都出现在7月。

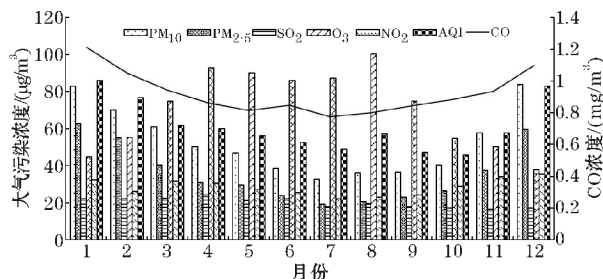


图4 2017–2020年大气污染浓度月变化图

从图3(a)可知6种大气污染浓度总体呈下降趋势,除 NO_2 和 O_3 外,其余4种大气污染的浓度在2018年变化最显著,明显低于2017年;而2018年 NO_2 浓度略低于2017年, O_3 浓度则高于2017年;2020年 SO_2 和 NO_2 的浓度明显低于2019年; SO_2 浓度变化较为显著,2020年的 PM_{10} 、 O_3 浓度和AQI指数比2019年略有回升,但均低于2018年。由夏季趋势变化图(图3b)可知,6种大气污染浓度变化较不规律,其中 O_3 平均8小时最大浓度从2017年到2019年均呈上升趋势,2020年浓度比2019年有所下降; PM_{10} 、 $\text{PM}_{2.5}$ 、 SO_2 则是呈逐年下降趋势; NO_2 的变化趋势不明显。秋季趋势变化图(图3c)与春季变化趋势基本一致。冬季变化趋势(图3d)可知,6种大气污染浓度变化相对规律,均呈逐年下降趋势。由CO的四季逐年变化图(图3e)可见,春夏两季在2017–2019年呈上升趋势,到2020年开始下降,秋季呈逐年的下降趋势,冬季在2018年达到高峰后,逐年下降。

2.2 AQI变化特征

2.2.1 污染天数变化特征

根据《环境空气质量指数》(AQI)技术规定,统计了南川6个等级出现的天数比例(图5),从图5可看出,南川无6级严重污染天数,5个等级占比中以优良天数为主,其中一级优的天数,2020年217 d占全年的59.8%,占比较多;2017年占比最少,只有26.6%;二级良的天数占比对多的2017年占全年58.5%,2020年占比最少,有36.9%;污染等级中,轻度污染2018年最多,有27 d,占全年的7.4%,2020年占比最少,只有2.2%;中度和重度污染天数在南川不明显,其中中度污染天数占比2017年最高,有4.9%,2020年较2017年减少了4.1%;重度污染天数占比最高是出现在2017年,有2.7%,2019–2020年无重度污染天数。总之,2017–2020年重庆南川的空气质量不断改善,优良天数逐渐增加。

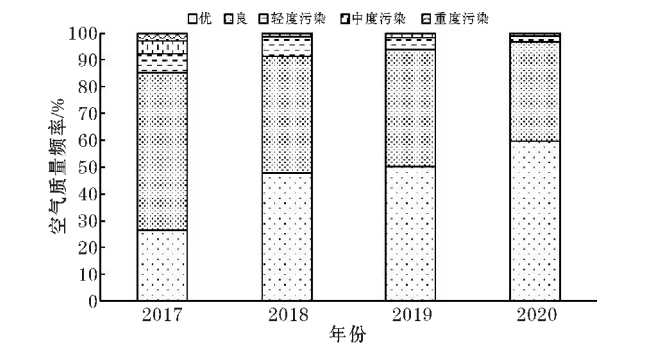


图5 重庆南川污染等级比例

2.2.2 污染过程变化特征

表2是重庆南川2017–2020年AQI>150持续污染过程的统计情况,2017年南川AQI>150的次数明显多于之后的3年,且污染持续天数最长达5 d,2018年和2020年AQI>150的次数最少,只发生了4次,但2019年大气污染平均持续天数最长,平均有2 d,2020年大气污染平均持续天数最短,平均1 d;总的来说,随

表2 2017–2020 AQI>150 的持续污染过程统计情况

年份	大气污染发生次数	大气污染平均持续天数/d	污染发展至峰值平均所需天数/d	大气污染最长持续天数/d
2017	16	1.6	1.3	5
2018	4	1.3	1	2
2019	5	2	1.7	4
2020	4	1	1	1
平均	7.25	1.475	1.25	3

表3 重庆南川空气质量与大气污染物灰色关联度

AQI	PM _{2.5}	PM ₁₀	NO ₂	CO	SO ₂	O ₃
年	0.7833	0.65817	0.78947	0.62124	0.63595	0.55245
春	0.65393	0.66381	0.57739	0.57176	0.54880	0.51102
夏	0.55881	0.55447	0.46502	0.48198	0.53640	0.67415
秋	0.73397	0.72646	0.57721	0.54824	0.54747	0.51555
冬	0.93657	0.85254	0.60825	0.54658	0.64129	0.45594

2.3.2 主成分对空气质量的综合分析

主成分分析是以降维方式提取主成分,提取出的主成分是原始变量的综合考量,可简化数据的复杂度。利用SPSS软件将PM₁₀、SO₂、NO₂、PM_{2.5}、O₃、CO作为原始输入变量,通过PCA算法进行空气质量的年、季节的主成分分析(表4)。分别对6项影响因素进行特征值和特征向量计算,从表4可知,年、季节的前2个主成分均>1,累积贡献率分别为73.447%、68.804%、68.256%、75.441%、72.915%,对比旋转平方和的结果,两者累计贡献率一致,因此两个主成分的重要性未发生变化。同时对比累积贡献率可知,秋季和年的贡献率最高,夏季贡献率最低。

对比主成分特征向量(表5),年变化的第一成分

着多项污染治理措施的实施,南川中度及以上的污染过程逐渐在减少,且污染过程持续时间也在逐渐缩短。

2.3 空气质量综合分析

2.3.1 灰色关联度分析对空气质量的综合分析

根据灰色关联度构建南川空气质量的评价模型,对空气质量进行综合分析,从表3可看出南川年空气质量与各大气污染物的关联度基本都大于0.6,这说明它们之间有较高的关系,要素之间耦合作用较强。其中和各大气污染物的关联度排序为:NO₂>PM_{2.5}>PM₁₀>SO₂>CO>O₃。说明在各指标中,NO₂、PM_{2.5}、PM₁₀对空气质量的影响程度最大,O₃的影响程度最小。春季空气质量与各大气污染物的关联度大于0.6的有PM_{2.5}和PM₁₀,系数最小的是O₃,只有0.51102;夏季空气质量与各大气污染物的关联度除了O₃系数大于0.6关系显著外,其余都小于0.6,最小的是NO₂只有0.46502;秋季空气质量与各大气污染物的关联度与春季相当,但是PM_{2.5}和PM₁₀与空气质量的关系更为显著都大于0.7,其余4个大气污染物的系数都小于0.6,说明秋季与空气质量关系最紧密的是PM_{2.5}和PM₁₀;冬季PM_{2.5}和PM₁₀的系数最大,超过了0.8,特别是PM_{2.5}达0.93657,说明影响冬季的空气质量主要是PM_{2.5}和PM₁₀,影响最小的则是O₃。总之,O₃对空气质量影响最大是出现在夏季,其余季节不明显,PM_{2.5}和PM₁₀是除夏季外,对空气质量影响最大的要素,特别是冬季。

中PM_{2.5}、PM₁₀、NO₂、CO、SO₂与空气质量呈正相关且影响较大,O₃与空气质量呈负相关,结果表明PM_{2.5}、PM₁₀、NO₂是第一主成分的主要污染物。第二主成分与PM_{2.5}、PM₁₀、NO₂、SO₂、O₃呈正相关,其中O₃、SO₂特征向量值较大,为次要污染物。计算主成分各因子的权重依次为PM₁₀>PM_{2.5}>CO、NO₂>SO₂>O₃。

春季第一成分与年变化一致,PM_{2.5}、PM₁₀、NO₂、CO、SO₂与空气质量呈正相关,O₃与空气质量呈负相关,PM_{2.5}、PM₁₀、NO₂特征向量值大于0.8,表明PM_{2.5}、PM₁₀、NO₂是第一主成分的主要污染物。第二主成分与第一成分相反,O₃呈正相关,且特征向量值最大,有0.877,NO₂、CO与空气质量呈反关系,说明O₃为第二主成分的污染物。计算主成分各因子的权重与年变化一致。

表4 主成分特征值和方差贡献率

AQI	成分	初始特征值			提取载荷平方和			旋转载荷平方和		
		总计	方差百分比/%	累积/%	总计	方差百分比/%	累积/%	总计	方差百分比/%	累积/%
年	1	3.395	56.588	56.588	3.395	56.588	56.588	3.311	55.175	55.175
	2	1.012	16.859	73.447	1.012	16.859	73.447	1.096	18.272	73.447
	3	0.758	12.640	86.087						
	4	0.433	7.219	93.306						
	5	0.383	6.390	99.696						
	6	0.018	0.304	100.000						
春	1	3.042	50.707	50.707	3.042	50.707	50.707	2.980	49.661	49.661
	2	1.086	18.097	68.804	1.086	18.097	68.804	1.149	19.143	68.804
	3	0.934	15.564	84.368						
	4	0.554	9.231	93.600						
	5	0.340	5.659	99.259						
	6	0.044	0.741	100.000						
夏	1	3.037	50.611	50.611	3.037	50.611	50.611	2.756	45.939	45.939
	2	1.059	17.645	68.256	1.059	17.645	68.256	1.339	22.317	68.256
	3	0.814	13.574	81.830						
	4	0.624	10.401	92.231						
	5	0.416	6.928	99.159						
	6	0.050	.841	100.000						
秋	1	3.358	55.967	55.967	3.358	55.967	55.967	3.357	55.946	55.946
	2	1.168	19.474	75.441	1.168	19.474	75.441	1.170	19.495	75.441
	3	0.886	14.762	90.203						
	4	0.358	5.965	96.168						
	5	0.155	2.585	98.753						
	6	0.075	1.247	100.000						
冬	1	3.203	53.380	53.380	3.203	53.380	53.380	3.167	52.786	52.786
	2	1.172	19.535	72.915	1.172	19.535	72.915	1.208	20.129	72.915
	3	0.844	14.075	86.990						
	4	0.481	8.024	95.013						
	5	0.285	4.752	99.766						
	6	0.014	0.234	100.000						

夏季第一成分与空气质量均呈正相关,且 $PM_{2.5}$ 、 PM_{10} 的特征向量值最大,超过0.9,特征向量值最小的仍为 O_3 ,只有0.380,说明虽然在夏季空气质量指数会随着 O_3 浓度升高而升高,但是其主要因素还是 $PM_{2.5}$ 、 PM_{10} 。第二成分中 O_3 的特征向量值则明显高于第二高的 PM_{10} (0.134),说明 O_3 是夏季次要污染物。计算主成分各因子的权重依次为 $PM_{10}>PM_{2.5}>O_3>SO_2>NO_2>CO$ 。

秋季第一成分与夏季成分一样,均与空气质量呈正相关, PM_{10} 、 SO_2 、 O_3 特征向量值最大,第二成分中除 $PM_{2.5}$ 、

NO_2 、 O_3 与空气质量呈负相关外,其余都呈正相关,且 CO 的值最大,是第二成分的主要污染物。计算主成分各因子的权重依次为 $PM_{10}>SO_2>O_3>CO$ 、 $PM_{2.5}>NO_2$ 。

冬季第一成分 PM_{10} 、 $PM_{2.5}$ 、 NO_2 是主要污染物,第二成分 CO 为主要污染物。计算主成分各因子的权重依次为 $PM_{2.5}>PM_{10}>NO_2>SO_2>O_3>CO$ 。

总的来说,影响南川空气质量的主要污染物是 PM_{10} 与 $PM_{2.5}$, O_3 主要是夏季仅次 $PM_{2.5}$ 、 PM_{10} 的污染物。

表 5 主成分特征向量

AQI	成分	PM _{2.5}	PM ₁₀	NO ₂	CO	SO ₂	O ₃
年	主成分 1	0.941	0.932	0.822	0.786	0.504	-0.307
	主成分 2	0.074	0.126	0.012	-0.206	0.455	0.861
	综合得分	0.410	0.418	0.346	0.376	0.315	0.324
	权重	0.187	0.191	0.158	0.172	0.144	0.148
春	主成分 1	0.930	0.913	0.834	0.631	0.499	-0.022
	主成分 2	0.075	0.182	-0.055	-0.463	0.244	0.877
	综合得分	0.231	0.432	0.412	0.383	0.272	0.231
	权重	0.196	0.206	0.175	0.183	0.130	0.110
夏	主成分 1	0.935	0.920	0.591	0.668	0.614	0.380
	主成分 2	0.043	0.134	-0.284	-0.078	-0.446	0.868
	综合得分	0.409	0.425	0.323	0.304	0.373	0.380
	权重	0.185	0.192	0.146	0.138	0.169	0.172
秋	主成分 1	0.747	0.951	0.397	0.222	0.954	0.883
	主成分 2	-0.068	0.180	-0.593	0.871	0.048	-0.139
	综合得分	0.287	0.428	0.019	0.298	0.398	0.324
	权重	0.163	0.244	0.011	0.170	0.227	0.185
冬	主成分 1	0.947	0.951	0.866	-0.015	0.750	-0.300
	主成分 2	-0.115	-0.108	-0.127	0.847	0.200	-0.612
	综合得分	0.422	0.422	0.391	0.219	0.361	0.278
	权重	0.202	0.201	0.187	0.105	0.173	0.133

3 结论与讨论

基于灰色关联度和主成分对重庆南川区 2017–2020 年空气质量及其相关物质浓度环境监测数据进行特征分析,得到以下结论:

(1)南川空气质量及大气污染浓度季节变化明显,春、秋两季 6 种大气污染浓度总体呈下降趋势,夏季 6 种大气污染浓度变化较不规律,其中 O₃ 平均 8 小时最大浓度 2017–2019 年均呈上升趋势,2020 年浓度比 2019 年有所下降;PM₁₀、PM_{2.5}、SO₂ 则是呈逐年下降趋势;NO₂ 的变化趋势不明显。冬季 6 种大气污染浓度变化相对规律,均呈逐年下降趋势。CO 春、夏两季在 2017–2019 年呈上升趋势,到 2020 年开始下降,秋季呈逐年下降趋势,冬季在 2018 年达到高峰后,逐年下降。PM₁₀、PM_{2.5}、NO₂ 在月变化中呈“U”型,7 月值最低;O₃ 则与 PM₁₀、PM_{2.5}、NO₂ 的月变化相反,SO₂ 月变化不明显,12 个月的变化相对平缓;AQI 和 CO 的变化较一致,1 月和 12 月为最高值。

(2)南川无 6 级严重污染天数,5 个等级占比中以优良天数为主,其中一级优的天数,2020 年占比最多,2017 年占比最少;二级良的天数 2017 年最多,2020 年占比最少;污染等级中,轻度污染占比多于中度和重度污染天数。其中 2019–2020 年无重度污染天数。2017 年南川 AQI>150 的次数最多,且污染持续天数最长达 5 d,2020 年大气污染平均持续天数最短,平均 1 d。

(3)基于灰色关联度方法,年空气质量与各大气

污染物的关联度从高到低排序为:NO₂>PM_{2.5}>PM₁₀>SO₂>CO>O₃>NO₂>PM_{2.5}>PM₁₀对空气质量的影响程度最大,O₃的影响程度最小。春季关联度高的是 PM_{2.5}和 PM₁₀;夏季与 O₃的关联度最高;秋、冬季与空气质量关系最紧密的是 PM_{2.5}和 PM₁₀;影响最小的则是 O₃。

(4)基于主成分分析,年、春季的主要成分是 PM₁₀、PM_{2.5}、CO,夏季主要污染物是 PM₁₀、PM_{2.5},次要污染物是 O₃。秋季 PM₁₀、SO₂、O₃ 的权重最大。冬季主要成分是 PM_{2.5}、PM₁₀、NO₂。

通过两种机器学习算法分析空气质量及其相关物质浓度的相关性,虽然结果大致相同,但是两种结果还是存有偏差,造成这一问题的原因主要是主成分分析法主要是利用降维,把多个指标转化为少数几个综合指标,每个综合指标都能够反映原始变量的大部分信息,且所含信息互不重复,使得到的结果更加科学有效。灰色关联度分析则是将 6 个指标视为一个灰色系统,在所有被评价对象中选出各指标的最优值,组成序列,构成最优样本。综上所述,主成分分析法和灰色关联度分析法均可以综合考虑各指标,因此在大气污染研究中,两种方法可在综合评价时相互检验。

参考文献:

[1] 腾浩宇. 基于常规大气监测数据对 PM_{2.5} 评估和预测[D]. 哈尔滨: 东北林业大学, 2015.

[2] 王胜杰,解淑艳,王军霞. 2016–2019 年汾渭平原城市空气质量状况分析[J]. 中国环境监测,

- 2020,36(6):57-65.
- [3] 刘丽,唐晓婵,马东卓.2013-2019年青岛市空气质量变化特征分析[J].中国环境监测,2020,36(6):57-65.
- [4] 张莹,宁贵财,康延臻,等.北京市大气污染物与呼吸系统疾病入院人数的时间序列[J].兰州大学学报(自然科学版),2015,51(1):87-92.
- [5] GB3095-2012,环境空气质量标准[S].2012.
- [6] 姜新华,薛河儒,张存厚,等.基于主成分分析的呼和浩特市空气质量影响因素研究[J].安全与环境工程,2016,23(1):75-79.
- [7] 喻义勇,谢放尖,陆晓波,等.南京亚青会环境空气质量状况及原因分析[J].环境监控与预警,2014(1):5-9.
- [8] 刘爱明.城市区域大气颗粒物的污染特征研究[J].安全与环境工程,2009,16(6):15-20.
- [9] 王玮,汤大钢,刘红杰,等.中国PM_{2.5}污染状况和污染特征的研究[J].环境科学研究.2000,13(1):1-5.
- [10] 于大江,吴艳玲.东北主要城市的空气质量长期变化特征[J].环境化学,2014(6):956-961.
- [11] 王翠红,梁新艳.山西省主要城市的空气质量长期变化特征研究[J].环境科学与管理,2016(1):94-97.
- [12] 赵颖,李婷婷.连云港市空气质量变化趋势分析[J].中国环境监测.2014,30(1):66-69.
- [13] 李刚.克拉玛依市空气质量特征分析[J].干旱环境监测,2017,31(2).
- [14] 许海超,李子君,姜爱霞,等.山东省空气质量指数的时空分布特征[J].济南大学学报,2017,31(2):168-175.
- [15] 陈颖,张仲伍.基于聚类分析和主成分分析的城市空气质量评价——以山西省11个地级市为例[J].山西师范大学学报(自然科学版),2020,34(4):72-78.
- [16] 王宝英.基于气象因素分析的城市AQI预测方法研究[D].南昌:南昌航空大学.2019.
- [17] 孙苏琪,王式功,罗彬,等.应用机器学习算法的成都市冬季空气污染预报研究[J].气象与环境学报,2020,36(2):98-104.
- [18] 赵域圻,杨婷,王自发,等.基于KZ滤波的京津冀2013-2018年大气污染治理效果分析[J].气候与环境研究,2020,25(5):499-509.
- [19] 蔡旺华.运用机器学习方法预测空气中臭氧浓度[J].中国环境管理,2018,2:78-84.
- [20] 陈俏.支持向量机应用于大气污染物浓度预测[D].西安:西安科技大学,2010.
- [21] 梁昱,刘禹含,王红丽,等.基于主成分分析的上海春季近地面臭氧污染区域性特征研究[J].环境科学学报2018,38(10):3807-3815.
- [22] 吴亚平,张琦,王炳赞,等.2020.四川雅安三种主要大气污染物浓度与气象条件的关系及其预测研究[J].高原气象,39(4):889-898.
- [23] 刘姣姣,叶堤,刘芮伶.重庆市臭氧污染特征研究[J].环境科学与管理,2020,45(9):59-62.
- [24] 吉莉,李强,司云燕,等.重庆北碚大气污染物的变化特征及其与气象要素的相关性[J].中国农学通报,2017,33(18):88-93.
- [25] 蒲茜,李振亮,张悦,等.重庆市O₃污染日的大气环流分型与传输特征[J].中国环境科学2021,41(1):18-27.
- [26] 刘姣姣,陈思,赵洁,等.重庆市PM_{2.5}污染特征分析及预报结果评估[J].环境科学与技术,2020,43(6):85-93.
- [27] 韩余,周国兵,陈道劲,等.重庆市臭氧污染及其气象因子预报方法对比研究[J].气象与环境学报,2020,36(4):59-66.
- [28] 李嗣同.我国大气污染的灰色关联分析[J].齐鲁工业大学学报,2014,28(4):55-58.

Comprehensive Analysis of Air Quality in Southwest Chongqing based on Machine Learning

FAN Ying¹, JI Li²

(1. Chongqing Nanchuan District Ecological Environment Monitoring Station, Chongqing 408400, China; 2. Weather Bureau in Beibei District of Chongqing City, Chongqing 400700, China)

Abstract: Based on the environmental monitoring data of Nanchuan from 2017 to 2020, the characteristics of air quality and air pollutant concentration were analyzed by using mathematical statistics method, and the air quality was analyzed based on grey correlation degree and PCA principal component analysis. The results show that: the seasonal variation of air quality and air pollution concentration in Nanchuan is obvious, and both of them show a downward trend; In Nanchuan, the number of days with good quality was the majority, and there was no serious pollution process, and the proportion of light pollution days was more than that of moderate and severe pollution days; O₃ has the greatest impact on air quality in summer, and it is not obvious in other seasons, and in addition to summer, PM_{2.5} and PM₁₀ have the greatest impact on air quality, especially in winter; Based on PCA principal component analysis, PM₁₀ and PM_{2.5} are the main pollutants, and O₃ is the third main pollutant which is next to PM_{2.5} and PM₁₀ in summer.

Keywords: air quality; grey correlation degree; principal component analysis