

doi:10.16055/j.issn.1672-058X.2020.0004.016

# 基于 PCA-聚类分析的终端二氧化碳排放量 分布特征综合评价\*

吴凤华<sup>a</sup>, 胡澄<sup>b\*\*</sup>, 江泽标<sup>a</sup>, 田娟<sup>a</sup>  
(贵州大学 a. 矿业学院; b. 土木工程学院, 贵阳 550025)

**摘要:**针对高能源需求带来环境与减排压力这一问题,提出了对于终端控制二氧化碳排放量的评价方法;利用省际二氧化碳核算法得出 30 个省份终端二氧化碳排放量,并采用 PCA-聚类分析法对终端二氧化碳排放分布特征进行分析。聚类分析在不知先验知识的情况下对指标变量进行分类,结合主成分分析综合得分进行人工判别分析,能更好地对终端二氧化碳的分布特征进行分析,两方法结合归纳总结了各省终端二氧化碳排放污染程度。研究结论表明:空气质量最差的省份是隶属华北地区的河北省,其次是整个华中地区以及华东、华南、华北地区的一些省份,空气质量最好的为西南、西北地区,加大对重大污染区域整改力度,对未来的能源结构调整和低碳环保具有一定借鉴意义和参考价值。

**关键词:**主成分分析;系统聚类分析;终端二氧化碳排放分布特征;综合评分

**中图分类号:**X511 **文献标志码:**A **文章编号:**1672-058X(2020)04-0111-07

## 0 引言

国际经合组织(OECD)定义,二氧化碳排放是指由碳的燃烧和生物体呼吸形成,在特定区域和时间段内,将温室气体或其前体释放到大气中。二氧化碳排放造成的温室效应问题,已引起多数科学家们的广泛关注<sup>[1-4]</sup>。在气候变化的压力下,减少温室气体(GHG<sub>s</sub>)的排放是保持可持续发展的关键问题<sup>[5]</sup>。当前主要核算的方法都是以《2006 年 IPCC 国家温室气体清单指南》作为燃烧二氧化碳排放计算标准<sup>[6-8]</sup>,但其得到的计算往往都是化石燃料燃烧或某一省份企业的

碳排,由于省份产业结构和经济发展程度的差异性,不同指标得到环境污染结果也大不相同。在误差允许的范围内参考省际二氧化碳排放核算方法<sup>[9]</sup>,计算二氧化碳排放时,不考虑原材料使用和消耗(消耗原材料的过程未产生 CO<sub>2</sub>)数据,将终端二氧化碳排放量作为衡量环境污染的指标。主成分(PCA)聚类分析主要用于食品、农业科学研究领域,解释种群和揭示物质品质或结构之间的关系<sup>[10-12]</sup>,在环境科学方面涉猎尚不足,为了弥补中国低碳、安全、清洁的环境科学评价方法体系上的空缺,利用该方法定位分析终端二氧化碳分布规律特征。

收稿日期:2019-09-17;修回日期:2019-11-04.

\* 基金项目:贵州省科学技术基金项目(黔科合基础[2019]1102);贵州省科技计划课题(黔科合基础[2016]1082).

作者简介:吴凤华(1993—),女,硕士研究生,从事安全科学与技术研究.

\*\* 通讯作者:胡澄(1977—),男,贵州人,博士,讲师,从事安全工程与技术绿色建筑与节能技术研究. Email:1255321804@

# 1 分析方法概述

## 1.1 聚类分析方法

聚类分析是统计学上度量事物相似性的有效分析方法,能在无先验知识的情况下自动分类,其中把具备相似个体的集合称为一类,在分类过程中不同类之间有明显的区别,是一种探索性分析。聚类分析中最简单直观的是系统聚类分析方法,不存在多重共线性。类与类之间采用系统聚类中的类平均距离方法。根据数据将终端二氧化碳排放分类成不同群体,并评价这些群体的排放等级。

系统聚类的步骤:(1)计算样品两两间的距离;(2)合并距离最近的为一新类;(3)计算新类与当前类的距离;(4)重复前面步骤(2)(3),合并距离最近的类为新类,直到所有的类并为一类后截止分类;(5)画出聚类系谱图,决定类的个数和类别。

## 1.2 主成分分析方法

主成分分析(PCA)在某种程度上可以说是因子分析的重要分支,其思想就是降维(线性变换),用较少新变量代替繁多复杂的旧变量,在不损失大部分主要信息的前提下尽可能多地反应原始变量信息<sup>[13]</sup>。设置  $N$  个样本,  $p$  个变量表示  $X_i, i=1, 2, \dots, p$ 。把  $p$  个变量削减为  $m$  个,构建主成分分析模型如下:

$$\begin{cases} F_1 = X_1 U_{11} + X_2 U_{12} + \dots + X_m U_{1m} \\ F_2 = X_1 U_{21} + X_2 U_{22} + \dots + X_m U_{2m} \\ \vdots \\ F_p = X_1 U_{p1} + X_2 U_{p2} + \dots + X_m U_{pm} \end{cases} \quad (1)$$

主成分荷载矩阵  $U$  与因子荷载矩阵  $A$  以及特征值  $\lambda$  的公式如下:

$$U_{ij} = A_{ij} / \sqrt{\lambda_i} \quad (2)$$

这里系数矩阵  $U_{ij}$  代表一个主成分作为原来变量线性组合的比例系数。 $i(1, 2, \dots, p), j(1, 2, \dots,$

$m)$ , 通常  $m \ll p$ ;  $F$  表示复合主成分分数,  $W$  表示权重,是分量的方差贡献率,进行权重归一化,公式如下:

$$\begin{aligned} F &= F_1 * (W_1 / W_1 + W_2 + W_3 + W_4) + \\ &F_2 * (W_2 / W_1 + W_2 + W_3 + W_4) + \\ &F_3 * (W_3 / W_1 + W_2 + W_3 + W_4) + \\ &F_4 * (W_4 / W_1 + W_2 + W_3 + W_4) \end{aligned} \quad (3)$$

PCA 分析方法的步骤:(1)定义要研究问题是否适合提取因子(主成分法);(2)确定因子个数;(3)诠释因子;(4)计算因子得分并评价。

## 2 确定指标变量

核算二氧化碳排放量的活动水平数据均来自 2018 中国能源统计年鉴《分地区能源平衡表(实物量)》终端能源消费量<sup>[14]</sup>,根据各种能源最低发热量分别对能源产品所产生的热值标准化,计算出 2017 年 30 个地区分部门二氧化碳排放量,如表 1 所示,农林牧渔业、工业、建筑业、交通运输存储和邮政业、批发零售业和住宿、餐饮业、其他、生活消费 7 个部门在制造能源产品活动中都会排放出二氧化碳,分别记为指标元素  $X_1 - X_7$ 。

## 3 聚类分析

利用 SPSS 软件将指标值导入变量,将地区导入统计个案,通过组间联接绘出树状图如图 1 所示。先进行聚类分析再人工进行甄别。从节点处划分变量的合并过程分为 3 类。先是合并贵州、上海、甘肃、天津、北京、青海、海南、黑龙江、云南、浙江、吉林、宁夏、陕西、安徽、内蒙古、新疆、广西、江西、重庆、福建为一类,接着合并广东、湖南、辽宁、湖北、山西、四川、河南、山东、江苏为一类,最后加上河北,30 个变量全部合并。初步判断河北和其他省市不同,污染最严重,单独为一类。

表1 部门终端二氧化碳排放总量

Table 1 Total amount of CO<sub>2</sub> emissions from the sector

地区	农、林、牧、渔业 $X_1/(ton^4)$	工业 $X_2/(ton^4)$	建筑业 $X_3/(万吨)$	交通运输、 存储和邮政 业 $X_4/(ton^4)$	批发、零售业 和住宿、餐饮 业 $X_5/(ton^4)$	其他 $X_6/(ton^4)$	生活消费 $X_7/(ton^4)$	综合得分 排序
北京	34.02	270.76	67.00	1 182.16	101.98	224.22	855.90	26
天津	83.77	2 853.80	205.42	413.10	55.08	230.78	418.65	27
河北	410.78	29 456.05	265.38	811.63	372.31	349.19	3 246.55	1
山西	396.67	11 694.33	108.62	952.32	395.36	471.90	1 929.28	6
内蒙古	1365.19	8 624.12	399.74	968.82	742.83	597.44	1 025.03	12
辽宁	269.13	12 590.81	36.91	1872.12	167.09	715.30	1 379.01	5
吉林	287.11	4 412.57	106.52	1 210.71	339.05	687.72	448.80	20
黑龙江	1 101.14	6 000.36	29.36	1453.41	1719.30	1 749.91	786.00	13
上海	42.71	2 213.63	107.85	1 466.96	1 282.75	406.26	499.28	24
江苏	425.74	16 401.71	30.54	1 802.92	172.76	28.45	695.14	7
浙江	382.75	5 388.78	293.44	1 527.33	214.20	461.45	554.70	18
安徽	248.62	9 319.27	195.64	1 028.84	144.81	256.66	642.71	16
福建	174.79	6 296.71	125.24	1 093.04	27.64	54.57	282.30	22
江西	133.61	7 991.42	38.77	730.19	120.96	117.13	605.18	17
山东	482.08	17 977.96	748.86	2 131.20	889.13	741.42	2 067.37	2
河南	337.94	11 466.46	230.48	1 206.52	365.12	141.06	1 253.60	9
湖北	811.71	10 668.05	197.29	1 759.14	682.16	681.52	1 533.51	8
湖南	1027.80	12 198.31	542.95	1767.09	1147.56	1 282.27	1 481.05	3
广东	308.05	9 686.01	113.61	3 289.78	662.39	149.99	1 813.16	4
广西	149.11	7 582.86	3.64	988.99	82.21	45.79	253.73	30
海南	78.53	514.60	24.21	284.47	13.41	72.51	61.94	19
重庆	145.67	6 377.37	117.46	930.65	105.07	100.58	261.94	23
四川	250.12	11 306.06	188.78	1 347.52	407.72	321.95	834.95	11
贵州	419.04	3 584.93	111.16	665.47	2 355.59	2036.17	1 842.74	10
云南	451.93	6 194.68	2166.26	1 082.36	298.29	156.36	905.77	15
陕西	114.52	5 151.94	70.46	628.35	207.65	244.32	697.62	21
甘肃	173.19	3 282.00	75.25	484.40	81.39	163.16	689.10	25
青海	19.13	1 254.65	35.26	174.55	51.76	68.34	245.90	29
宁夏	20.28	4 835.62	116.20	46.37	158.43	42.86	133.39	28
新疆	542.96	7 619.06	96.65	974.48	208.47	229.26	1 221.38	14

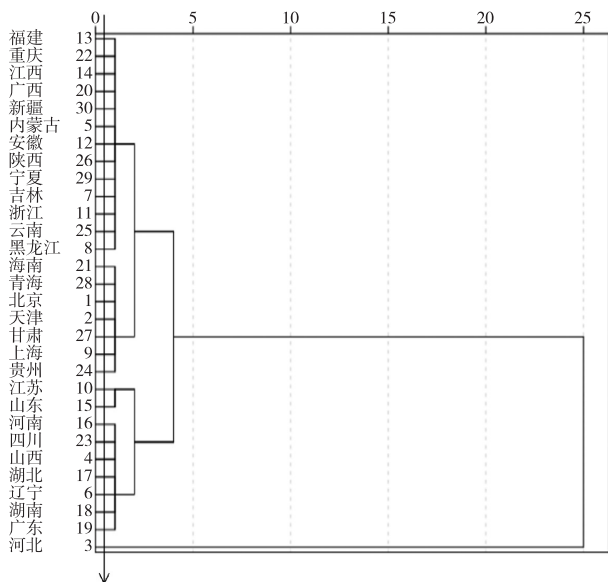


图 1 使用组间联结的树状图

Fig. 1 Tree diagram using inter-group connections

## 4 主成分分析模型构建

### 4.1 描述性统计

要确定变量是否适合主成分分析,则需要对变量进行描述性统计分析处理。表 2 显示 KMO 为  $0.619 > 0.6$ , 变量中有许多共同因素; Bartlett 的球形测试具有  $0.000 < 0.05$  的显著  $P$  值, 说明相关系数矩阵非单位矩阵, 则表明该数据适合于主成分分析。

表 2 KMO 和 Bartlett 球形检验结果

Table 2 KMO and Bartlett spherical test results

取样足够度的 Kaiser-Meyer-Olkin 度量	近似卡方	Bartlett 的 球形度检验	
0.619	97.008	df	Sig.
		21	0.000

### 4.2 确定因子个数

解释的总方差由表 3 给出, 降维提取出一个主成份, 成分 1 特征根  $\lambda = 3.033 > 1$ , 贡献了方差的 43.333%, 成分 2 特征根  $\lambda = 1.534 > 1$ , 两个成分累计贡献率为 65.232%。为了能够反映大部分信息, 需要提取大于 85% 的贡献率, 第 3 成分特征根  $\lambda = 0.981$  接近于 1, 第 4 成分特征根  $\lambda = 0.705$  接近于 1, 4 个成分的累加贡献率为 89.330%, 显著反映原始变量 89.330% 的信息。这 4 个成分可概括为第一产业(农、林、牧、渔业)、第二产业(工业、建筑业)、第三产业(交通运输、存储、邮政业、批发、零售业和住宿、餐饮业)、其他(包括居民生活消费)终端所产生的二氧化碳排放量。随着社会和科学技术的进步发展, 第一产业和第二产业物质生产部门的比重都不断下降, 第三产业和其他方面产业持续上升, 但发展中国家第二产业的比重始终大于第一产业和第三产业, 因此第 1 主成分为第二产业, 第 2 主成分为第三产业, 第 3 主成分为其他, 第 4 主成分为第一产业。

表 3 基于相关矩阵分析的方差比例

Table 3 Proportion of variance based on correlation matrix analysis

成分	合计	初始特征值			提取平方和载入		旋转平方和载入	
		方差的贡献 百分比/%	累积贡献 百分比/%	特征 系数 $\lambda$	方差的贡献 百分比/%	累积贡献 百分比/%	方差的贡献 百分比/%	累积贡献 百分比/%
1	3.033	43.333	43.333	3.033	43.333	43.333	3.033	43.333
2	1.534	21.919	65.253	1.534	21.919	65.253	1.534	21.919
3	0.981	14.009	79.262	0.981	14.009	79.262	0.981	14.009
4	0.705	10.069	89.330	0.705	10.069	89.330	0.705	10.069
5	0.501	7.162	96.493				0.501	7.162
6	0.146	2.092	98.585				0.146	2.092
7	0.099	1.415	100.000				0.099	1.415

### 4.3 综合得分评价

根据成分矩阵系数表4可以写出因子的表达式,然后计算出因子得分。

由于变量在分析前已经进行了量纲统一,不需要进行标准化,因此根据式(1)和式(2)得到主成分的变形表达式如下:

首先计算出:

$$F_1 = 0.440X_1 + 0.450X_2 + 0.243X_3 + 0.675X_4 +$$

$$1.079X_5 + 1.994X_6 + 2.447X_7$$

$$F_2 = -0.074X_1 + 0.584X_2 + 0.248X_3 + 0.375X_4 -$$

$$0.786X_5 - 1.534X_6 + 1.360X_7$$

$$F_3 = 0.123X_1 - 0.164X_2 + 0.928X_3 - 0.119X_4 -$$

$$0.055X_5 - 0.160X_6 - 0.575X_7$$

$$F_4 = 0.010X_1 - 0.191X_2 - 0.036X_3 + 0.892X_4 -$$

$$0.013X_5 - 0.264X_6 - 0.880X_7$$

而后分别得出4个主成分的因子得分矩阵:

$$F_1 = [3602.57, 3193.87, 23090.62, 12194.71, 9733.68, 12037.98, 5790.45, 11440.64, 5447.26, 10736.62, 6204.22, 7285.83, 4508.1, 6002.17, 17419.35, 9921.8, 12240.65, 14685.39, 12193.06, 4947.23]^T$$

$$F_2 = [1355.44, 2038.32, 21129.12, 8773.37, 5291.59, 8691.17, 2325.04, 1008.26, 914.03, 10996.72, 3642.42, 6224.93, 4383.79, 5488.83, 12423.81, 8382.54, 7382.61, 6990.32, 8610.85, 4999.35, 369.87, 4211.09, 7457.57, -129.14, 5284.9, 3664.14, 2727.1, 994.42, 2859.84]^T$$

$$F_3 = [-652.35, -596.93, -6573.69, -3088.19, -1716.6, -3136.89, -1120.31, -1820.82, -854.91, -3237.48, -1150.69, -1857.25, -1197.58, -1718.44, -3804.03, -2532.09, -2704.3, -2700.42, -2939.67, -1497.31, -134.07, -1202.21, -2362.63, -2027.35, 358.91, -1291.87, -931.57, -346.63, -780.5, -1959.45]^T$$

$$F_4 = [186.99, -613.2, -7861.56, -3211.58, -1853.17, -2138.1, -344.72, -1015.68, 318.97, -2142.84, -286.36, -1502.04, -493.63, -1440.14, -3581.48, -2263.95, -2005.71, -2419.66, -560.35, -801.22,$$

$$81.55, -649.15, -1786.82, -2280.72, -1133.42, -1106.03, -846.3, -320.13, -1016.98, -1722.1]^T$$

表4 成分矩阵A

Table 4 Ingredient Matrix A

部 门	1	2	3	4
农、林、牧、渔业	0.766	-0.129	0.215	0.017
工业	0.557	0.723	-0.203	-0.236
建筑业	0.241	0.246	0.919	-0.036
交通运输、存储和邮政业	0.567	0.315	-0.100	0.749
批发、零售业和住宿、餐饮业	0.764	-0.556	-0.039	-0.009
其他	0.762	-0.586	-0.061	-0.101
生活消费	0.770	0.428	-0.181	-0.277

再根据式(3)得到主成分综合得分F的变形公式:

$$F = F_1 * (43.333/89.330) + F_2 * (21.918/89.330) + F_3 * (14.009/89.330) + F_4 * (10.069/89.330)$$

最后计算综合得分:  $F = [1998.93, 1886.72, 14468.44, 7221.95, 5542.02, 7239.11, 3164.84, 5397.09, 2768.57, 7157.25, 3690.61, 4601.13, 3019.04, 3826.57, 10498.13, 6217.51, 7099.11, 8142.72, 7503.41, 3301.42, 454.69, 2976.58, 5758.73, 5886.12, 4666.26, 3108.01, 2366.93, 877.91, 1836.18, 4847.35]^T$ ,对计算结果进行排序,如表1所示。

结合聚类分析与主成分分析进行人工判别分类来定义环境污染程度,河北为14468.44,评分最大,权重评分大于14000,定义为最高,单独为一类;广东、湖南、辽宁、湖北、山西、河南、山东、江苏为一类,权重得分均介于6000—11000,定义为较重;四川、贵州、上海、甘肃、天津、北京、青海、海南、黑龙江、云南、浙江、吉林、宁夏、陕西、安徽、内蒙古、新疆、广西、江西、重庆、福建为一类,权重得分均介于400—5900,定义为较轻。

## 5 结 论

(1) 从综合评分表1权重排序可以看出:当前终端二氧化碳分布特征中,西南、西北以及华东、华

南、华北部分地区环境污染情况较良好,华东地区山东、江苏,华南地区广东、华北地区山西以及整个华中地区环境污染形势较为严峻,华北地区中的河北环境污染最为严重。华东、华南、华中、华北、东北部分地区需要加大环境治理力度,从终端源头控制第二、三产业二氧化碳排放量,需要从节约化石能源的使用,增加核电、风电、太阳能、生物质能等清洁低碳的能源技术开发及应用上着手,改善环境进一步加剧恶化的状况,从而提高空气环境质量。

(2) PCA-聚类分析法不但可解决不知先验类别的行分类分析,而且还能通过量化综合评分进行排序判断重要类别程度,从而提出改进措施。

(3) 利用省级核算能更好地对各省终端二氧化碳排放量分布情况作对比分析,且解决了量纲不统一的问题,节省了作主成分分析时的标准化步骤。

#### 参考文献(References):

- [1] 钱杰,俞立中.上海市化石燃料排放二氧化碳贡献量的研究[J].上海环境科学,2003,22(11):836—839  
QIAN J, YU L Z. Study on the Contribution of Carbon Dioxide Emission from Fossil Fuels in Shanghai [J]. Shanghai Environmental Science, 2003, 22(11): 836—839(in Chinese)
- [2] 欧国立,王妍.交通运输业技术进步对二氧化碳排放的影响——基于中国省级面板数据的实证研究[J].生态经济,2018,34(11):64—71  
OU G L, WANG Y. The Impact of Technical Progress of Transportation Industry on Carbon Dioxide Emission: An Empirical Study Based on Chinese Provincial Panel Data [J]. Ecology, 2018, 34(11): 64—71(in Chinese)
- [3] 孙建.中国区域技术创新的二氧化碳减排效应——基于宏观计量经济模型模拟分析[J].技术经济,2018,37(10):107—116  
SUN J. China's Regional Technology Innovation Carbon Dioxide Emission Reduction Effect: Simulation Analysis Based on Macro-Econometric Model [J]. Technical Economy, 2018, 37(10): 107—116(in Chinese)
- [4] 何建坤,卢兰兰,王海林.经济增长与二氧化碳减排的双赢路径分析[J].中国人口·资源与环境,2018,28(10):9—17  
HE J K, LU L L, WANG H L. A Win-Win Path Analysis of Economic Growth and CO<sub>2</sub> Emission Reduction [J]. China Population, Resources and Environment, 2018, 28(10): 9—17(in Chinese)

- [5] LI H, DONG L. Low-Carbon Benefit of Industrial Symbiosis From a Scope-3 Perspective: A Case Study In China [J]. Applied Ecology and Environmental Research, 2017, 15(3): 135—153
- [6] 龙芸.燃煤电厂CO<sub>2</sub>排放计算模型与方法研究[D].重庆:重庆大学,2016  
LONG Y. Research on Calculation Model and Method of CO<sub>2</sub> Emission in Coal-fired Power Plants [D]. Chongqing: Chongqing University, 2016(in Chinese)
- [7] 吴国华.化石能源消费的二氧化碳排放量计算与分析——以济南市为例[J].理论学刊,2012(3):61—65  
WU G H. Calculation and Analysis of Carbon Dioxide Emissions from Fossil Energy Consumption: Taking Jinan City as an Example [J]. Journal of Theory, 2012(3): 61—65(in Chinese)
- [8] 朱悦,周昊,郝晓雯.辽宁省二氧化碳排放现状调查及减排措施研究[J].安徽农业科学,2010,38(23):12357—12360  
ZHU Y, ZHOU H, HAO X W. Survey of CO<sub>2</sub> Emission Status and Research on Emission Reduction Measures in Liaoning Province [J]. Journal of Anhui Agricultural Sciences, 2010, 38(23): 12357—12360(in Chinese)
- [9] 黎水宝,程志,王伟,等.基于能源平衡表的宁夏二氧化碳排放核算研究[J].环境工程,2015,33(12):130—137  
LI S B, CHENG Z, WANG W, et al. Study on Ningxia CO<sub>2</sub> Emission Accounting Based on Energy Balance Sheet [J]. Environmental Engineering, 2015, 33(12): 130—137(in Chinese)
- [10] 廖兰,陈林萍,倪莉.基于主成分分析和聚类分析研究改性小麦蛋白的结构和功能特性[J].中国食品学报,2018,18(7):263—271  
LIAO L, CHEN L P, NI L. Study on the Structure and Functional Properties of Modified Wheat Protein Based on Principal Component Analysis and Cluster Analysis [J]. Journal of Chinese Institute of Food Science and Technology, 2018, 18(7): 263—271(in Chinese)
- [11] 陈思奇,孟满,杜勃峰,等.基于主成分分析与聚类分析综合评价不同菌种发酵刺梨果渣的香气品质[J].中国酿造,2019,38(6):152—159  
CHEN S Q, MENG M, DU B F, et al. Comprehensive Evaluation of Aroma Quality of Fermented Thorn Pear Pomace Based on Principal Component Analysis and Cluster Analysis [J]. China Brewing, 2019, 38(6): 152—159(in Chinese)

- [12] 于秋竹,孔宇,陈东升,等. 基于主成分及聚类分析的黑龙江省水稻产量与品质的综合评价[J]. 沈阳农业大学学报,2015,46(2):219—224  
YU Q Z, KONG Y, CHEN D S, et al. Comprehensive Evaluation of Rice Yield and Quality in Heilongjiang Province Based on Principal Component Analysis and Cluster Analysis [J]. Journal of Shenyang Agricultural University,2015,46(2):219—224 (in Chinese)
- [13] 陈艳. 基于主成分分析的江西低碳经济发展评价分析[J]. 南昌大学学报(人文社会科学版), 2012, 43(6): 63—90

- CHEN Y. Evaluation of Jiangxi Low Carbon Economy Development Based on Principal Component Analysis [J]. Journal of Nanchang University (Humanities and Social Sciences Edition), 2012, 43(6): 63—90 (in Chinese)
- [14] 国家统计局能源统计司. 2018 中国能源统计年鉴 [M]. 北京:中国统计出版社,2019  
Department of Energy Statistics. National Bureau of Statistics, 2018 China Energy Statistics Yearbook [M]. Beijing:China Statistics Press, 2019(in Chinese)

## Comprehensive Evaluation on Distribution Characteristics of Terminal Carbon Dioxide Emission Based on PCA-cluster Analysis

WU Feng-hua<sup>a</sup>, HU Cheng<sup>b</sup>, JIANG Ze-biao<sup>a</sup>, TIAN Juan<sup>a</sup>

(a. Mining School; b. Civil Engineering School, Guizhou University, Guiyang 550025, China)

**Abstract:** According to the problem in environment and emission reduction brought by high energy demand, an evaluation method for terminal control of carbon dioxide emission is proposed. By using inter-provincial carbon dioxide calculation method to obtain the terminal carbon dioxide emissions in 30 provinces and by using PCA-cluster analysis method, the distribution characteristics of terminal carbon dioxide emissions are analyzed. Cluster analysis analyzes the index variables without knowing the prior knowledge and combines principal component analysis and comprehensive scores for manual discriminant analysis, which can better analyze the distribution characteristics of terminal carbon dioxide. The terminal carbon dioxide pollution degree of each province is generalized and summarized. Research conclusions show that the province with the worst air quality is Hebei Province of North China, followed by entire Central China region and some provinces in East China, South China and North China, and the provinces with the best air quality are in southwest region and in northwest region. The intensification of the pollution treatment in polluted areas is of certain reference significance and reference value to energy structure adjustment and low carbon environment protection.

**Key words:** principal component analysis; systematic cluster analysis; distribution characteristics of terminal carbon dioxide emission; comprehensive score

责任编辑:李翠微

引用本文/Cite this paper:

吴凤华,胡澄,江泽标,等. 基于PCA-聚类分析的终端二氧化碳排放量分布特征综合评价[J]. 重庆工商大学学报(自然科学版),2020,37(4):98—104

WU F H, HU C, JIANG Z B, et al. Comprehensive Evaluation on Distribution Characteristics of Terminal Carbon Dioxide Emission Based on PCA-cluster Analysis [J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2020, 37(4): 98—104