

邓晰元, 郑锦涛, 周晓辉, 等. 基于 PCA-HCA 模型的我国节水分区[J]. 南水北调与水利科技(中英文), 2023, 21(1): 137-147. DENG X Y, ZHENG J T, ZHOU X H, et al. Water-saving regionalization in China based on PCA-HCA model[J]. South-to-North Water Transfers and Water Science & Technology, 2023, 21(1): 137-147. (in Chinese)

# 基于 PCA-HCA 模型的我国节水分区

邓晰元<sup>1,2</sup>, 郑锦涛<sup>1,2,3</sup>, 周晓辉<sup>4</sup>, 马涛<sup>1,2</sup>, 郑皓<sup>1,2</sup>, 梁秀<sup>5</sup>, 王国庆<sup>1,2</sup>

(1. 南京水利科学研究院, 南京 210029; 2. 水文水资源与水利工程科学国家重点实验室, 南京 210029;  
3. 河海大学水文水资源学院, 南京 210024; 4. 中国科学院南京地理与湖泊研究所, 南京 210008;  
5. 江西省水利科学院, 南昌 330029)

**摘要:**为因地制宜地推进我国节水工作,以我国 31 个省(自治区、直辖市)为基础单元,按照领域协同、统筹兼顾、区划完整、有效衔接的原则,从水资源状况、社会经济条件、农业用水特征、工业用水特征和生活用水特征 5 个方面选取 14 项指标,构建节水分区指标体系。采用主成分分析(principal component analysis, PCA)和系统聚类分析(hierarchical clustering analysis, HCA)相结合的方法,将 31 个省(自治区、直辖市)划分成为东北地区、华北地区、西北地区、东南沿海地区、华中地区和西南地区 6 大节水分区。节水分区特征分析结果表明,各分区间水资源、经济社会和分领域用水指标差异显著,水资源短缺和经济社会发展对节约用水具有显著促进作用。

**关键词:**节水分区;主成分分析;系统聚类分析;指标体系;特征分析

**中图分类号:**TV213 **文献标志码:**A **DOI:**10.13476/j.cnki.nsbdqk.2023.0015

我国为解决水资源短缺问题,1998 年成立全国节约用水办公室,大力推进节水工作,用水效率大幅提升。然而,由于我国地理和气候特征多样、水资源时空分布不均<sup>[1]</sup>,当前节水工作仍面临着区域领域不平衡、潜力挖掘不充分<sup>[2-3]</sup>、节水措施难落地等问题。科学合理划定全国节水分区是因地制宜推进我国节水工作的重要基础。在相继开展国家节水行动和全民节水行动等一系列重大节水战略的背景下,开展我国节水分区研究,对科学抓准分区节水重点、稳步推进节水政策实施具有重要意义<sup>[4]</sup>。以我国 31 个省(自治区、直辖市)(以下简称研究区)为单元,提出我国节水分区划定的原则,采用主成分分析与系统聚类相结合的方法,划定节水分区,并分析各分区的水资源、社会经济以及分领域用水特征,为科学推进我国节水事业提供有力支撑。

## 1 节水分区研究进展

农业、工业、城镇生活是节水的重点领域。从全国尺度看,在我国已出台的政策文件中,农业分

区较为类似,大致可分为东北、海河、淮河、黄河上中游、黄河中下游、西南内陆河、长江中下游和东南沿海等区域,见表 1,不同政策文件对以上部分区域加以合并,或根据区域特点进行细化,但基本遵循水资源分区套行政区划的原则。这种分区方式打破了传统行政区划的界限,增加了行政管理的难度。此外,我国目前还没有针对工业和城镇生活领域开展全国尺度分区。《节水型社会建设“十三五”规划》和《规划和建设项目节水评价技术要求》统筹规划节水型社会建设的重点方向和任务,将研究区划分为东北、华北、西北、西南、华中和东南沿海 6 大区域,这种分区方式未打破传统行政区划的界限,更适用于节水工作的推进、管理和监督。《“十四五”节水型社会建设规划》提出了“十四五”时期全面推进节水型社会建设的总体要求、主要任务、重点领域和保障措施,但没有对区域布局做进一步细化。

从省级尺度看,各省(自治区、直辖市)出台的农业用水定额标准在充分考虑本地水资源条件和经济社会发展水平等因素的条件下划定分区。其

收稿日期:2022-03-21 修回日期:2022-11-26 网络出版时间:2023-01-03

网络出版地址:https://kns.cnki.net/kcms/detail//13.1430.TV.20221230.1650.001.html

基金项目:“十四五”国家重点研发计划项目(2021YFC3201100);水利部水资源节约项目(126302001000210004);南京水利科学研究院中央级公益性科研院所基本科研业务费专项资金项目(Y522009)

作者简介:邓晰元(1992—),女,辽宁辽中人,工程师,主要从事节水评价、节水潜力、非常规水资源利用等研究。E-mail:xydeng@nhri.cn

中,大部分按照地形地貌或行政区划进行分区,也有一部分按照气候条件进行分区(如内蒙古自治区、贵州省等)。各省(自治区、直辖市)划定的农业分区空间单元较细,这有利于农业节水工作的开展,却不利于国家层面农业节水战略的推进和实施。各地出台的生活用水定额标准,主要采用 3 种方法

对生活用水定额进行分类:一是按照城市规模,分为特大城市(镇)、大城市(镇)、中城市(镇)、小城市(镇)等;二是按照排水、淋浴设施、热水供应等室内设施分类;三是按照地理位置及市级行政单元分类。其中,第二种分类方式过于微观,不适于在大、中尺度上进行分区推广。

表 1 全国尺度节水分区现状

Tab. 1 Status quo of national scale water saving regionalization

面向对象	政策文件	分区方式
节水型社会	《节水型社会建设“十三五”规划》《规划和建设项目节水评价技术要求》	东北、华北、西北、西南、华中、东南沿海
	《节水型社会建设“十一五”规划》《节水型社会建设“十二五”规划》	东北、黄淮海、长江中下游、华南、西南、西北
	《节水型社会评价指标体系和评价方法》 <sup>[5]</sup>	缺水发达、缺水欠发达、平水发达、平水欠发达、丰水发达、丰水欠发达
农业	《国家农业节水纲要(2012—2020年)》	东北、西北、黄淮海、南方
	《全国现代灌溉发展规划》	东北、黄淮海、长江中下游、华南沿海、西南、西北
	《全国节水灌溉规划》	东北区、海河区、黄河上中游区、黄河中下游区、淮河区、东南沿海区、长江中下游区、西南区、内陆河区
	《全国农业现代化规划(2016—2020年)》	优化发展区,包括东北区、华北区、长江中下游区、华南区;适度发展区,包括西北区、北方农牧交错区、西南区;保护发展区,包括青藏区、海洋渔业区
	《全国农业可持续发展规划(2015—2030年)》	优化发展区,包括东北区、黄淮海区、长江中下游区和华南区;适度发展区,包括西北及长城沿线区、西南区;保护发展区,包括青藏区和海洋渔业区

已有学术研究中的农业节水分区大多以“归纳相似性,区别差异性,照顾行政区界”为原则,以促进区域农业(灌溉)水资源高效利用和有序管理<sup>[6-7]</sup>、节水农业<sup>[8]</sup>、可持续发展<sup>[9]</sup>为目标,以行政区划或水资源三级区套行政区为单元,采用叠置法、地理相关法、主导指标法和模糊数学法等方法进行分区<sup>[10-12]</sup>。具体操作中,大多采用构建农业节水分区指标体系的方法,定性与定量相结合,综合考虑各项与农业节水相关的指标,客观反映不同地区的节水重点与要求<sup>[13-15]</sup>。但受资料全面性、指标系统性以及主观干扰性的制约和影响,分区成果与实际情况可能存在一定偏差,在全国范围内的认可度较低。目前,工业节水空间分区相关研究较少,有的研究者<sup>[16]</sup>基于主导指标聚类分析法,综合考虑工业发展规模、资源环境、工业结构、技术投入、环境以及经济杠杆等指标,将研究区划分为 5 种类型:资源压力驱动型、发展需求驱动型、资源约束-发展需求制约型、资源富余型以及发展需求制约型。城镇生活节水分区基本均为节水型社会建设而服务<sup>[17]</sup>,主要依托层次分析法构建节水型社会评价指标体系,采用

聚类分析法和定性定量相结合的方法划定节水分区<sup>[18]</sup>。

理论上,节水分区问题可以简化为数据分类问题,即根据各省(自治区、直辖市)节水变量的特征,将节水特征较为一致的省(自治区、直辖市)划分为一类。数据分类算法模型可分为 2 种类型:一种是带有标签数据的监督式分类;另一种是没有标签数据的非监督式分类<sup>[19-20]</sup>。监督式分类的主要目的是利用分类模型对未知数据做出预测,即根据已有数据的分类成果预测未来。非监督式分类并不具有标签变量(先验知识),只能根据数据的自身特征判断类别<sup>[21]</sup>。鉴于节水分区没有“标准答案”或“明确答案”,因此节水分区问题显然是非监督式分类问题,只能从各省(自治区、直辖市)节水变量的自身特征出发判断节水分区类别。目前,常用的非监督式分类方法为聚类算法<sup>[10,18,22]</sup>,主要包括基于层次的聚类、基于分割的聚类、基于密度的聚类、基于网格的聚类以及基于模型的聚类 5 种类型<sup>[23]</sup>。其中,基于层次的聚类即系统聚类法,与其他算法相比,具有概念简单、计算过程直观、不需预先假设分类

个数、对距离度量标准的选择不敏感以及人为干扰少等优点,但计算效率相对较低。此外,由于选择的节水分区指标部分具有相关性,相互之间并不完全独立,需要先采用主成分分析法<sup>[24]</sup>对原始分区指标数据进行降维,以达到在不影响数据准确性的情况下,降低系统聚类法的计算复杂度、提高计算效率的目的。

## 2 节水分区研究方法

### 2.1 节水分区原则

科学合理划定节水分区,应有利于各项节水政策的有效落实,有利于水资源保护和集约高效利用,符合经济社会发展的历史规律,符合节约资源和保护环境的发展格局。其划定原则主要包括以下几个方面:领域协同,农业、工业和城镇生活三大领域统一分区,领域间协同节水;统筹兼顾,统筹考虑水资源条件、社会经济条件和农业、工业、城镇生活的用水需要;区划完整,分区时不打破省级行政区划,确保行政区划的完整性;有效衔接,分区成果应与经济社会发展、水资源、土地利用等已有规

划相衔接。

### 2.2 节水分区指标体系

根据上述原则,从水资源状况,社会经济条件,农业、工业及生活三大领域用水特征共5方面构建节水分区指标体系,共包括14项具体指标,见表2。其中,水资源状况包括水资源总量、跨省调水工程调入水量和非常规水源利用量3项指标,分别反映本地水资源状况、通过跨省调水工程主动调引的外域水资源量和非常规水资源利用情况;社会经济条件包括人口密度、人均地区生产总值和人均水资源量3项指标,分别反映地区人口分布、经济发展水平和人均水资源占有情况;农业用水特征包括农业用水占比、实际灌溉亩均用水量和农田灌溉水有效利用系数3项指标,分别反映农业用水结构、农业生产用水效率和农田灌溉用水效率;工业用水特征包括工业用水占比和万元工业增加值用水量2项指标,分别反映工业用水结构和工业生产用水效率;生活用水特征包括生活用水占比、城镇人均生活用水量和城镇公共供水管网综合漏损率3项指标,分别反映生活用水结构、城镇居民用水情况和供水系统供水效率。

表2 节水分区指标体系

Tab. 2 Water saving regionalization index system

一级指标	二级指标	三级指标	数据来源	指标解释
	水资源状况	水资源总量/亿m <sup>3</sup>	《中国水资源公报》	流域或区域内地表水资源量、地下水资源量以及两者的不重复计算量的总和
		跨省调水工程调入水量/亿m <sup>3</sup>	各省(区、市)水资源公报	通过跨省调水工程的入境水量
	综合性指标	非常规水源利用量/亿m <sup>3</sup>	《中国水资源公报》	处理后可以利用或在一定条件下可直接利用的再生水、集蓄雨水、淡化海水、微咸水、矿坑水等
	社会经济条件	人口密度/(人·km <sup>-2</sup> )	《中国统计年鉴》	单位土地面积上的人口数量
		人均地区生产总值/万元	《中国统计年鉴》	按人口平均计算的地区社会最终产品和劳务的总值
		人均水资源量/m <sup>3</sup>	《中国水资源公报》 《中国统计年鉴》	区域内水资源量与人口数的比值
		农业用水占比/%	《中国水资源公报》	农业用水占用水总量的比重
分领域指标	农业用水特征	实际灌溉亩均用水量/m <sup>3</sup>	《中国水资源公报》	农业实际灌溉面积上的亩均用水量
		农田灌溉水有效利用系数	《中国水资源公报》	作物生长实际需要水量占灌溉水量的比例
		工业用水占比/%	《中国水资源公报》	工业用水占用水总量的比重
分领域指标	工业用水特征	万元工业增加值用水量/m <sup>3</sup>	《中国水资源公报》	工业每产生1万元增加值的取水量
		生活用水占比/%	《中国水资源公报》	生活用水占用水总量的比重
		生活用水特征	城镇人均生活用水量/(L·d <sup>-1</sup> )	《中国水资源公报》
		城市公共供水管网综合漏损率/%	《中国城乡建设统计年鉴》	城市和县城公共供水管网漏损水量与供水总量的比值

## 2.3 节水分区技术方法

### 2.3.1 主成分分析法

主成分分析法(principal component analysis, PCA)是一种降低数据维度的方法,使用较少的几个综合指标代替原来较多的变量指标,使这些综合指标尽可能多地反映原始指标的信息,且彼此之间相互独立,从而达到简化和优化的目的。主要步骤如下:

步骤 1 获取各基本单元节水指标值。

步骤 2 对各个指标值进行 z-score 标准化变换。

步骤 3 对标准化指标值进行 KMO 统计量和 Bartlett 球体检验,验证数据是否适合进行主成分分析。

步骤 4 计算标准化指标值的协方差矩阵,设为  $R_{n \times p}$ ,其中:  $n$  为基本单元数,即  $n=31$ ;  $p$  为指标体系中最终层次的指标值个数。

步骤 5 计算协方差矩阵的特征值及相应的正交化单位特征向量  $\omega$ 。

步骤 6 将协方差矩阵的特征值按大小顺序排列,以计算特征值的累计贡献率

$$G(m) = \sum_{i=1}^m \lambda_i / \sum_{k=1}^n \lambda_k \quad (1)$$

式中,  $\lambda_i$  为协方差矩阵从大到小排列的第 1、第 2、...、第  $i$  个特征值,当累计贡献率  $G(m)$  大于 85% 时,则认为前  $m$  个主成分能够足够反映原来变量的信息。

### 2.3.2 系统聚类分析法

系统聚类分析(hierarchical cluster analysis, HCA)法通常称为层次聚类法或分层聚类法,是聚类分析中广泛使用的一种方法。先将聚类的样本或变量各自看成一类,然后确定类与类间的相似统计量;选择最接近的两类或若干个类合并成一个新类,计算新类与其他各类间的相似性统计量;再选择最接近的两类或若干个类合并成一个新类,直到所有的样本或变量都合并成一类为止。

主要步骤如下:

步骤 1 将 31 个样本各自看成一类。

步骤 2 采用欧氏距离平方法计算类之间的距离。

步骤 3 选择距离最小的两个类合并成一个新类。计算聚合系数,即各类组内平均距离。

步骤 4 重复步骤 2 和 3,直至将所有的样本都合并成一类。

步骤 5 绘制聚合系数随分类数变化的折线图,判断折线图拐点,确定最佳分类数及各省(自治区、直辖市)所属类别。

采用 R 软件进行数据前处理、主成分分析和系

统聚类<sup>[25]</sup>。

## 3 节水分区结果及特征分析

### 3.1 节水分区过程及结果

为提高分区结果的可靠性,分别选用研究区 14 项指标的 3 套数据集开展节水分区研究。3 套数据集分别为:数据集 1, 2020 年数据,反映现状年节水分区指标情况;数据集 2, 2018—2020 年平均值数据,反映 2018 年机构改革成立全国节约用水办公室以来的节水分区指标情况;数据集 3, 2012—2020 年平均值数据,反映近 10 年节水分区指标情况。其中,由于仅北京、天津、河北、山东、河南和甘肃 6 省(直辖市)存在跨省调水工程调入水量,因此考虑不将跨省调水工程调入水量纳入主成分分析和系统聚类计算,而将其作为辅助指标参与后续的节水分区调整过程。

#### 3.1.1 主成分分析

表 3 为 KMO 统计量和 Bartlett 球体检验结果,3 套数据集的 KMO 统计量均大于 0.5,且 Bartlett 球体检验显著,因此 3 套数据集均适合进行主成分分析。

表 3 KMO 统计量检验和 Bartlett 球体检验

Tab. 3 KMO statistic test and Bartlett sphere test

数据集	KMO值	Bartlett球体检验		
		卡方值	自由度	P值
数据集1	0.614	352.500	78	<0.001
数据集2	0.598	360.511	78	<0.001
数据集3	0.605	365.402	78	<0.001

表 4 为 3 套数据集主成分分析中各成分的贡献率和累计贡献率。对于数据集 1,特征值最大的前 5 个成分的累计贡献率为 87.3%;对于数据集 2,特征值最大的前 5 个成分的累计贡献率为 86.9%;对于数据集 3,特征值最大的前 5 个成分的累计贡献率为 86.8%,均保留了原指标因子 85% 以上的信息。因此,数据集 1、数据集 2 和数据集 3 均选取前 5 个成分作为主成分进行节水分区计算。

#### 3.1.2 系统聚类分析

针对主成分因子进行系统聚类,3 套数据集的聚合系数折线见图 1。通过目估,判断 3 套数据集的折线图拐点分别为 13、11 和 11,系统聚类分析结果见表 5。

#### 3.1.3 分区结果及调整

从表 5 可知:第 1 组为河北、山西、内蒙古、山东、河南和陕西,第 2 组为辽宁、吉林和黑龙江,第

3组为浙江、福建和广东,第4组为安徽、江西、湖北和湖南,第5组为四川、贵州和云南,第6组为广西和海南,第7组为甘肃、青海、宁夏和新疆,以上7组省(自治区、直辖市)能有效聚集、自动成类;北京、天津、上海和西藏单独成类,较难判断其所属类别。数据集2的聚类结果与数据集1相比,内蒙古、

安徽和四川的分组不同;数据集3的聚类结果与数据集2相比,辽宁、浙江、重庆和青海的分组不同。与数据集1的聚类结果相比,数据集2和数据集3倾向于将广西与海南聚集在一起;与数据集3的聚类结果相比,数据集1和数据集2倾向于将东南沿海省份(浙江、福建和广东)聚集在一起。

表4 各成分的特征值、贡献率及累计贡献率

Tab. 4 Characteristic value, contribution rate and cumulative contribution rate of each component

成分	数据集1			数据集2			数据集3		
	特征值 $\lambda$	贡献率/%	累计贡献率/%	特征值 $\lambda$	贡献率/%	累计贡献率/%	特征值 $\lambda$	贡献率/%	累计贡献率/%
PC <sub>1</sub>	4.992	38.4	38.4	5.116	39.4	39.4	5.158	39.7	39.7
PC <sub>2</sub>	2.915	22.4	60.8	2.922	22.5	61.8	3.014	23.2	62.9
PC <sub>3</sub>	1.444	11.1	71.9	1.301	10.0	71.8	1.201	9.2	72.1
PC <sub>4</sub>	1.089	8.4	80.3	1.136	8.7	80.6	1.080	8.3	80.4
PC <sub>5</sub>	0.904	7.0	<b>87.3</b>	0.827	6.4	<b>86.9</b>	0.833	6.4	<b>86.8</b>
PC <sub>6</sub>	0.717	5.5	92.8	0.734	5.6	92.6	0.766	5.9	92.7
PC <sub>7</sub>	0.384	3.0	95.7	0.366	2.8	95.4	0.323	2.5	95.2
PC <sub>8</sub>	0.183	1.4	97.1	0.210	1.6	97.0	0.235	1.8	97.0
PC <sub>9</sub>	0.144	1.1	98.2	0.179	1.4	98.4	0.178	1.4	98.4
PC <sub>10</sub>	0.113	0.9	99.1	0.099	0.8	99.2	0.104	0.8	99.2
PC <sub>11</sub>	0.062	0.5	99.6	0.062	0.5	99.6	0.061	0.5	99.6
PC <sub>12</sub>	0.034	0.3	99.9	0.036	0.3	99.9	0.038	0.3	99.9
PC <sub>13</sub>	0.019	0.1	100.0	0.012	0.1	100.0	0.010	0.1	100.0

基于数学方法聚类,虽具有分类意义,但仍有部分地区无法明确类别,也存在某些地区的分区结果与实际情况不协调的情况(如:数据集1和数据集2将浙江、福建、广东和重庆聚类成组,但重庆与浙江、福建和广东并不接壤;数据集1将海南、甘肃、青海、宁夏和新疆聚集成组,但海南与西北地区相距甚远),故本研究根据分区结果中多数地域上相对集中、连片的特点,对部分省区进行微调。此外,作为节水分区,为了服务于未来用水定额制定和修订、水资源论证和节水评价等工作,需要开展水资源状况、社会经济条件和分领域用水效率等指标比对,因此每个节水分区应具有多个省份,考虑将单独成类的省(自治区、直辖市)纳入周边地区集群。

具体调整如下:

北京、天津。北京地处华北平原北部,东与天津毗连,其余均与河北相邻,是我国的政治中心、文化中心,具有特殊的地位。京津冀地缘相接、人缘相亲,地域一体、文化一脉,相互融合、协同发展。由表5可以看出,北京和天津各单独成类,依据分区有效衔接性原则,将北京和天津纳入河北、山西、内

蒙古、山东、河南和陕西集群。根据跨省调水工程调入水量指标,南水北调中线和东线主要受水区所在省份,即北京、天津、河北、山东和河南5省(直辖市)均在此分区中,侧面体现了该分区的合理性。

重庆。由表5可以看出,重庆倾向于与浙江、福建、广东聚类成组,但重庆与浙江、福建、广东并不接壤,因此将浙江、福建、广东与重庆拆解,将重庆划入四川、贵州和云南集群。

西藏。北邻新疆,东接四川,东北紧靠青海,东南连接云南,周边与国外接壤,地域辽阔,是中国西南边陲的重要门户。由表5可以看出,西藏自成一类,反映了西藏水资源条件和社会经济发展的特殊性。依据分区有效衔接性原则,将西藏纳入周边地区集群,将西藏与四川、贵州、云南和重庆划为一类。

广西目前有2种聚类方式,一种是与贵州和湖南等内陆省份聚集成组,另一种是与海南聚集成组;海南目前有2种聚类方式,一种是与甘肃和宁夏等省份聚集成组,另一种是与广西聚集成组;浙江目前有2种聚类方式,一种是与江苏聚集成组,另一种是与广东和福建聚集成组。考虑沿海自然条件和

沿海开放政策,将沿海省份聚为一类,即江苏、浙江、福建、广东、广西和海南划为一类。

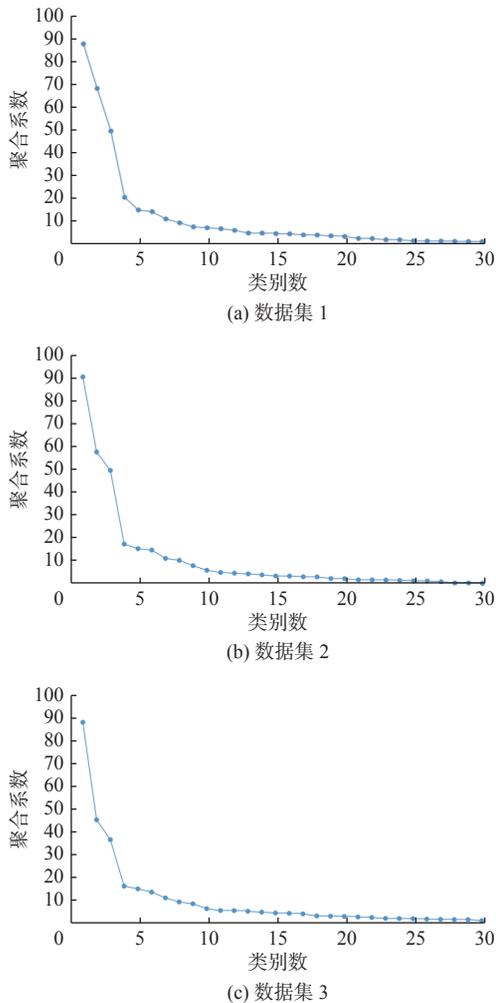


图 1 聚合系数图

Fig. 1 Aggregation coefficient diagram

上海地处中国东部、长江入海口,东临东海,北、西与江苏、浙江两省相接。由表 5 可以看出,上海单独成类,反映了中国国际经济、金融中心的特殊性。综合考虑地理位置、经济发展和沿海城市开放等因素,将上海与江苏、浙江、福建、广东、广西和海南划为一类。

综上所述,将研究区划分为东北、华北、西北、东南沿海、华中和西南 6 大分区,具体如下:

东北地区:包括辽宁、吉林和黑龙江。

华北地区:包括北京、天津、河北、山西、内蒙古、山东、河南和陕西。

西北地区:包括甘肃、青海、宁夏和新疆。

东南沿海地区:包括上海、江苏、浙江、福建、广东、广西和海南。

华中地区:包括安徽、江西、湖北和湖南。

西南地区:包括重庆、四川、贵州、云南和西藏。

### 3.2 节水分区特征分析

6 大节水分区的 2020 年、2018—2020 年和 2012—2020 年的水资源、经济社会和各领域用水特征见图 2。东北地区的人口密度、人均地区生产总值、工业用水占比和生活用水占比较低,农业用水占比较高,农田灌溉水有效利用系数较高,农田实际灌溉亩均用水量和城镇人均生活用水量较低,城镇公共供水管网综合漏损率最高。华北地区的水资源总量和人均水资源量最低,跨省调水工程调入水量、非常规水源利用量、生活用水占比最高,人口密度和人均地区生产总值较高,农业用水占比较低,农田灌溉水有效利用系数最高,农田实际灌溉亩均用水量、万元工业增加值用水量和城镇人均生活用水量最低。西北地区的水资源总量较低,存在少量跨省调水工程调入水量,人均水资源量较高,非常规水源利用量较低,人口密度、人均地区生产总值、工业用水占比和生活用水占比最低,农业用水占比最高,农田实际灌溉亩均用水量较高,万元工业增加值用水量较低,城镇公共供水管网综合漏损率最低。东南沿海地区的人口密度、人均地区生产总值、工业用水占比最高和人均水资源量较低,非常规水源利用量较高,生活用水占比较高,农业用水占比最低,农田实际灌溉亩均用水量最高,城镇人均生活用水量较高,城镇公共供水管网综合漏损率较低。华中地区的水资源总量和工业用水占比较高,农田灌溉水有效利用系数较低,城镇公共供水管网综合漏损率较高,万元工业增加值用水量最高。西南地区的水资源总量和人均水资源量最高,非常规水源利用量最低,农田灌溉水有效利用系数最低,万元工业增加值用水量较高,城镇人均生活用水量最高。

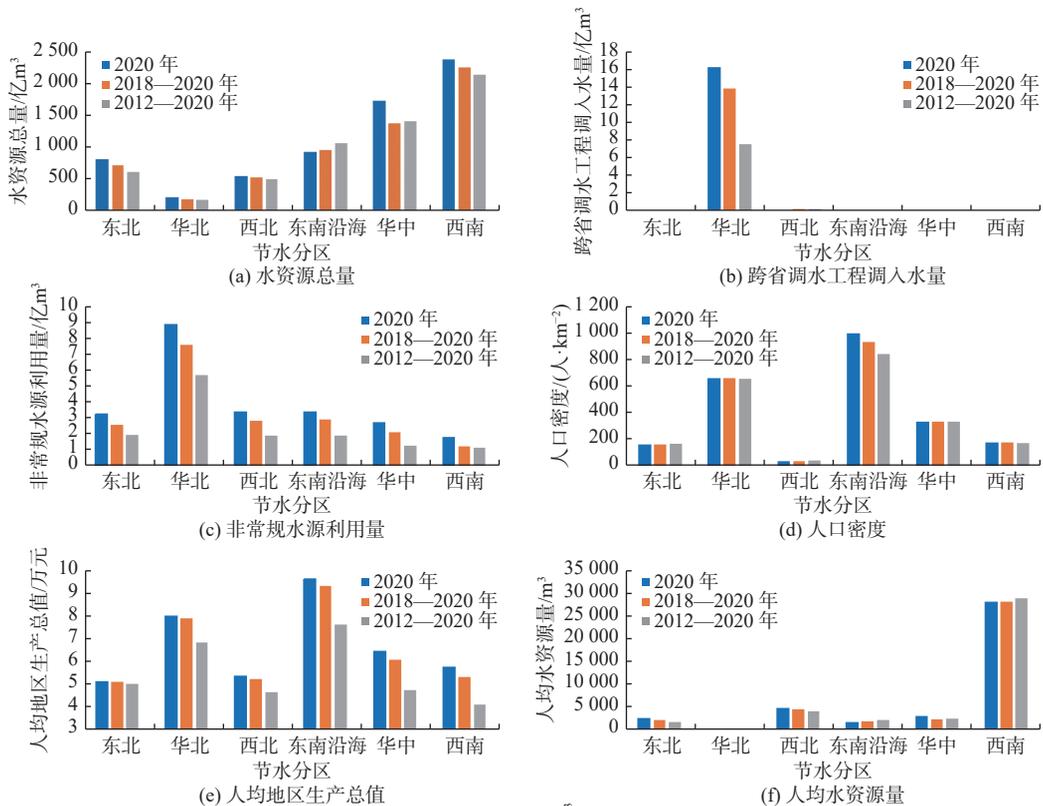
从分领域用水效率上看,华北地区和东北地区的农业用水效率最高,华北地区、西北地区和东北地区的工业用水效率最高,华北地区和西北地区的生活用水效率最高。从总体上看,南方地区(包括东南沿海地区、华中地区和西南地区)的水资源总量高于北方地区(包括东北地区、华北地区和西北地区),但北方地区整体用水效率好于南方地区,说明水资源短缺对于节约用水具有显著促进作用。北方地区中,华北地区的人均地区生产总值和整体用水效率好于东北地区和西北地区,说明除水资源短缺这一强影响因素外,社会经济条件对节约用水也具有促进作用。从分领域用水特征上看,工业用水

效率与生活用水效率具有较高的一致性,即工业用水效率越高,生活用水效率也越高,这与城镇化等经济社会发展水平有关。

表 5 系统聚类分析结果

Tab. 5 Results of hierarchical cluster analysis

分类	数据集1	数据集2	数据集3
I	北京	北京	北京
II	天津	天津	天津
III	河北、山西、山东、河南、陕西	河北、山西、内蒙古、山东、河南、陕西	河北、山西、内蒙古、山东、河南、陕西
IV	内蒙古、辽宁	辽宁	辽宁、吉林、黑龙江
V	吉林、黑龙江	吉林、黑龙江	上海
VI	上海	上海	江苏、浙江
VII	江苏	江苏	安徽、福建、江西、湖北、湖南、广东、四川、贵州、云南
VIII	浙江、福建、广东、重庆	浙江、福建、广东、重庆	广西、海南
IX	安徽	安徽、江西、湖北、湖南、广西、海南、四川、贵州、云南、青海	重庆
X	江西、湖北、湖南、广西、贵州	西藏	西藏
XI	海南、云南、甘肃、青海、宁夏、新疆	甘肃、宁夏、新疆	甘肃、青海、宁夏、新疆
XII	四川		
XIII	西藏		



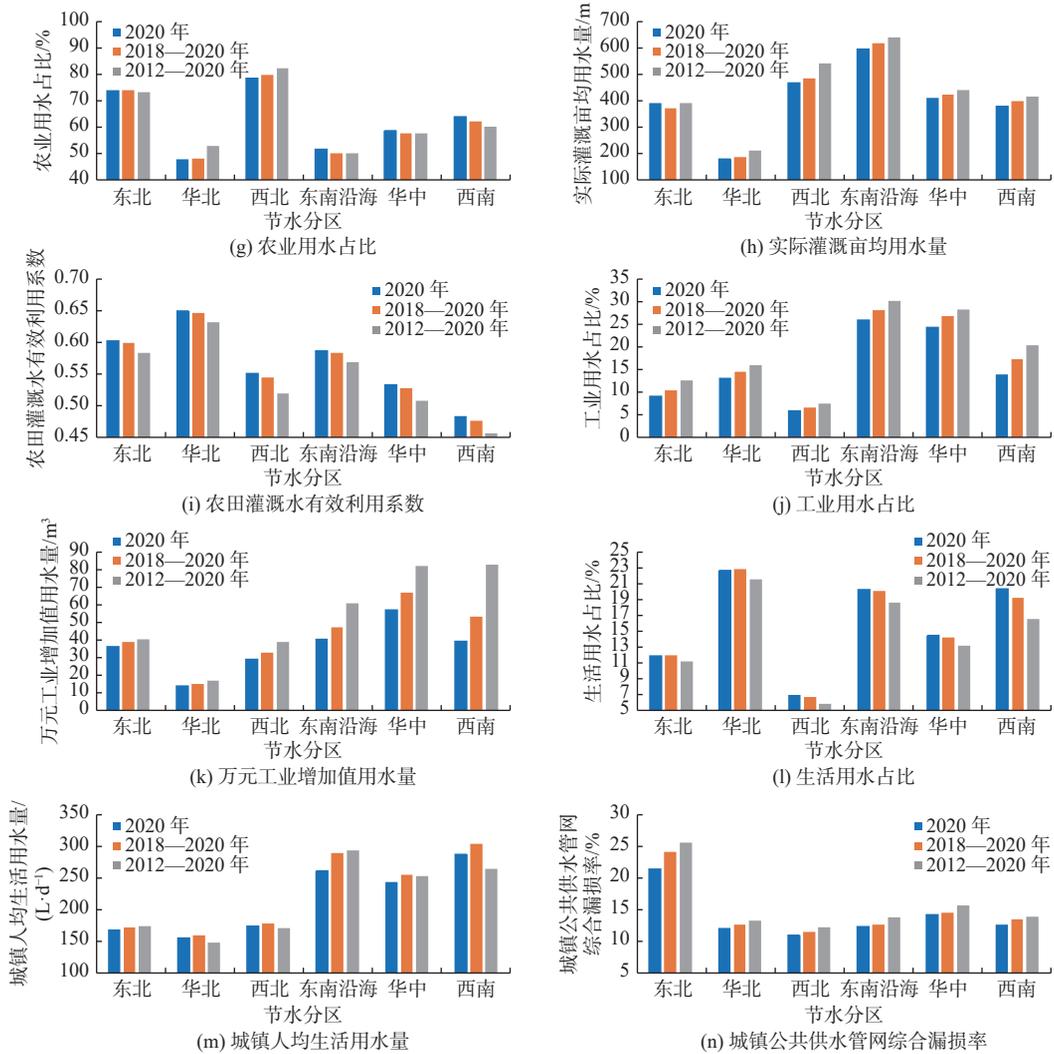


图 2 6 大节水分区水资源、经济社会、用水特征

Fig. 2 Characteristics of the six water-saving regions

为进一步说明分区指标对分区结果的影响,采用单因素方差分析法探索分区指标在 6 大分区的差异性。对显著性水平  $\alpha=0.05$ , 当  $F$  统计值大于  $F_{0.95}(5,25)=2.605$  时,认为不同分区下分区指标差异显著,且  $F$  统计值越大,显著性越明显。表 6 展示了参与计算的数据集 1~3 中 13 项分区指标的单因素方差分析结果。

由表 6 结果可知,除人均水资源量、人口密度、生活用水占比和人均地区生产总值对分区结果无

显著影响外,其他 9 个因素都对分区结果具有显著影响。其中,城镇公共供水管网综合漏损率、实际灌溉亩均用水量、农田灌溉水有效利用系数、水资源总量、非常规水源利用量、城镇人均生活用水量和万元工业增加值用水量对分区结果影响最为显著。总体上看,各分区水资源、经济社会条件以及分领域用水特征鲜明,分区间特征差异显著,也说明了分区方法和分区结果的可靠性。

表 6 单因素方差分析结果

Tab. 6 Results of one way ANOVA

因素	数据集1		数据集2		数据集3	
	$F$	Sig	$F$	Sig	$F$	Sig
水资源总量	5.586	0.001***	4.444	0.005**	4.350	0.006**
非常规水源利用量	4.792	0.003**	4.894	0.003**	4.399	0.005**
人口密度	1.441	0.244	1.469	0.235	1.508	0.223
人均地区生产总值	2.225	0.083	2.402	0.066	2.561	0.053

表 6(续)  
Tab. 6 (Continued)

因素	数据集1		数据集2		数据集3	
	F	Sig	F	Sig	F	Sig
人均水资源量	1.155	0.359	1.137	0.367	1.127	0.372
农业用水占比	2.466	0.060	2.626	0.048 <sup>*</sup>	2.675	0.045 <sup>*</sup>
实际灌溉亩均用水量	9.488	0 <sup>***</sup>	8.947	0 <sup>***</sup>	8.279	0 <sup>***</sup>
农田灌溉水有效利用系数	5.384	0.002 <sup>**</sup>	5.338	0.002 <sup>**</sup>	5.532	0.001 <sup>***</sup>
工业用水占比	3.142	0.025 <sup>*</sup>	3.624	0.013 <sup>*</sup>	3.577	0.014 <sup>*</sup>
万元工业增加值用水量	3.916	0.009 <sup>**</sup>	4.934	0.003 <sup>**</sup>	4.817	0.003 <sup>**</sup>
生活用水占比	2.365	0.069	2.274	0.078	2.166	0.090
城镇人均生活用水量	2.895	0.034 <sup>*</sup>	3.500	0.016 <sup>*</sup>	7.621	0 <sup>***</sup>
城市公共供水管网综合漏损率	13.895	0 <sup>***</sup>	24.658	0 <sup>***</sup>	26.576	0 <sup>***</sup>

注:\*\*\*表示显著性达到0.001,\*\*表示显著性达到0.01,\*表示显著性达到0.05。

#### 4 结语

目前,我国节水分区研究主要集中于农业领域,工业和城镇节水领域的分区研究较少,主要原因在于我国农业节水发展较早,且受地域和环境因素影响较大,在大尺度和中尺度上的区域内具有规律性、区域间具有分异性;工业节水主要受产业结构、技术条件和地方政策等因素影响,空间结构分散,很难在大尺度、中尺度,甚至小尺度上找到空间分异规律;城镇节水主要受水资源禀赋条件、经济社会发展水平、生活习惯等因素影响,从大尺度和中尺度上看,区域内具有规律性、区域间具有分异性,但由于城镇生活用水占比较少,其节水分区研究并未得到足够重视。

因此,以领域协同、统筹兼顾、区划完整和有效衔接为原则,从水资源状况、社会经济条件、农业用水特征、工业用水特征和生活用水特征 5 个方面选取 14 项指标,构建节水分区指标体系,采用主成分分析与系统聚类分析法相结合的方法,划定节水分区。根据分区结果中多数地域上相对集中、连片的特点,对部分省区进行微调,将研究区划分为东北地区(辽宁、吉林和黑龙江)、华北地区(北京、天津、河北、山西、内蒙古、山东、河南和陕西)、西北地区(甘肃、青海、宁夏和新疆)、东南沿海地区(上海、江苏、浙江、福建、广东、广西和海南)、华中地区(安徽、江西、湖北和湖南)以及西南地区(重庆、四川、贵州、云南和西藏)共 6 大节水分区。

总体上看,各分区水资源和经济社会条件以及分领域用水特征鲜明,分区间特征差异明显,其中城镇公共供水管网综合漏损率、实际灌溉亩均用水

量、农田灌溉水有效利用系数、水资源总量、非常规水源利用量、城镇人均生活用水量和万元工业增加值用水量 7 项指标对分区结果影响最为显著。今后将进一步从指标体系构建和数据获取等方面强化节水分区的合理性,为我国合理确定节水战略布局、科学制定节水推进策略等工作提供更有有力支撑。

#### 参考文献:

- [1] 张建云,王国庆. 气候变化对水文水资源影响研究[M]. 北京: 科学出版社, 2007.
- [2] 赵志博,赵领娣,王亚薇,等. 不同情景模式下雄安新区的水资源利用效率和节水潜力分析[J]. 自然资源学报, 2019, 34(12): 2629-2642. DOI: 10.31497/zrzyxb.20191212.
- [3] 代小平,周雯晶,韩宇平,等. 作物相对水分利用效率与理论节水潜力[J]. 水科学进展, 2019, 30(5): 682-690. DOI: 10.14042/j.cnki.32.1309.2019.05.008.
- [4] 浦承松,梅伟,张玉蓉. 云南省节水型社会建设分区发展方向研究[J]. 节水灌溉, 2010(8): 77-80. DOI: CNKI:SUN:JSGU.0.2010-08-022.
- [5] 中华人民共和国国家质量监督检验检疫总局. 节水型社会评价指标体系和评价方法: GB/T 28284—2012[S]. 北京: 中国标准出版社, 2012.
- [6] 郑和祥,李和平,郭克贞,等. 基于信息熵和模糊物元模型的牧区节水灌溉项目后评价[J]. 水利学报, 2013, 44(S1): 57-65. DOI: 10.13243/j.cnki.slxb.2013.s1.028.
- [7] 操信春,崔思梦,吴梦洋,等. 水足迹框架下稻田水资源利用效率综合评价[J]. 水利学报, 2020, 51(10): 1189-1198. DOI: 10.13243/j.cnki.slxb.20200260.
- [8] 裴源生,张金萍,赵勇. 宁夏灌区节水潜力的研究[J].

- 水利学报, 2007, 4(2): 239-243, 249.
- [9] JIANG Y, ZHANG Q, ZHAO X, et al. A geogrid-based framework of agricultural zoning for planning and management of water & land resources: A case study of northwest arid region of China[J]. *Ecological Indicators*, 2018, 89: 874-879. DOI: 10.1016/j.ecolind.2017.12.022.
- [10] 丁志宏, 郭兵托, 杨朝瀚. 基于模糊聚类的海河流域农业节水分区研究[J]. 人民黄河, 2018, 40(3): 44-48, 67.
- [11] 姜森严, 王铁良, 李波, 等. 辽宁省节水农业分区研究[J]. 中国农村水利水电, 2016(2): 14-16. DOI: 10.3969/j.issn.1007-2284.2016.02.004.
- [12] 王军德, 金彦兆, 程玉菲. 石羊河流域生态治理目标下节水灌溉技术分区[J]. 人民黄河, 2018, 40(9): 153-156. DOI: 10.3969/j.issn.1000-1379.2018.09.035.
- [13] 范海燕, 朱丹阳, 郝仲勇, 等. 基于AHP和ArcGIS的北京市农业节水区划研究[J]. 农业机械学报, 2017, 48(3): 288-293. DOI: 10.6041/j.issn.1000-1298.2017.03.036.
- [14] 王西琴, 吴若然, 李兆捷, 等. 我国农业用水安全的分区及发展对策[J]. 中国生态农业学报, 2016, 24(10): 1428-1434. DOI: 10.13930/j.cnki.cjea.160017.
- [15] 褚琳琳. 江苏省节水农业分区及发展模式[J]. 节水灌溉, 2014, 4(11): 91-95. DOI: 10.3969/j.issn.1007-4929.2014.11.025.
- [16] 姜蓓蕾, 耿雷华, 卞锦宇, 等. 中国工业用水效率水平驱动因素分析及区划研究[J]. 资源科学, 2014, 36(11): 2231-2239.
- [17] 张欣莹, 解建仓, 刘建林, 等. 基于熵权法的节水型社会建设区域类型分析[J]. 自然资源学报, 2017, 32(2): 301-309. DOI: 10.11849/zrzyxb.20160262.
- [18] 何力, 黄薇, 刘丹. 基于聚类分析方法的节水型社会分区研究[J]. 长江科学院院报, 2010, 27(4): 20-23. DOI: 10.3969/j.issn.1001-5485.2010.04.005.
- [19] BAO W, LIAN J N, YUE K. Integration of unsupervised and supervised machine learning algorithms for credit risk assessment[J]. *Expert Systems with Applications*, 2019, 128: 301-315. DOI: 10.1016/j.eswa.2019.02.033.
- [20] FRANKLIN J. The elements of statistical learning: Data mining, inference and prediction[J]. *The Mathematical Intelligencer*, 2005, 27(2): 83-85. DOI: 10.1007/BF02985802.
- [21] BERRY M W, MOHAMED A, YAP B W. Supervised and unsupervised learning for data science[M]. Switzerland: Springer, 2020.
- [22] AGARWAL A, MAHESWARAN R, SEHGAL V, et al. Hydrologic regionalization using wavelet-based multiscale entropy method[J]. *Journal of Hydrology*, 2016, 538: 22-32. DOI: 10.1016/j.jhydrol.2016.03.023.
- [23] SINAGA K P, YANG M S. Unsupervised *k*-means clustering algorithm[J]. *IEEE Access*, 2020, 8: 80716-80727. DOI: 10.1109/ACCESS.2020.2988796.
- [24] 杨蕊, 胡贤群, 王龙, 等. 基于主成分分析和模糊聚类的云南省农业节水分区[J]. 节水灌溉, 2021, 4(4): 92-97. DOI: 10.3969/j.issn.1007-4929.2021.04.016.
- [25] KASSAMBARA A. Practical guide to cluster analysis in R: Unsupervised machine learning[M]. California: CreateSpace Independent Publishing Platform, 2017.

## Water-saving regionalization in China based on PCA-HCA model

DENG Xiyuan<sup>1,2</sup>, ZHENG Jintao<sup>1,2,3</sup>, ZHOU Xiaohui<sup>4</sup>, MA Tao<sup>1,2</sup>,  
ZHENG Hao<sup>1,2</sup>, LIANG Xiu<sup>5</sup>, WANG Guoqing<sup>1,2</sup>

( 1. Nanjing Hydraulic Research Institute, Nanjing 210029, China; 2. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Nanjing 210029, China; 3. College of Hydrology and Water Resources, Hohai University, Nanjing 210024, China; 4. Nanjing Institute of Geography & Limnology, Chinese Academy of Sciences, Nanjing 210008, China; 5. Jiangxi Academy of Water Science and Engineering, Nanchang 330029, China )

**Abstract:** China has the highest degree of water resources stress. Due to large diverse geographical and climatic characteristics and uneven temporal and spatial distribution of water resources, current water-saving work is still facing problems such as regional imbalance, insufficient potential excavation, and difficult implementation of water-saving measures. Under the background of a series of major water-saving strategies such as the “National Water Conservation Action Plan”, researching water-saving zoning in China is of great significance to scientifically and

reasonably promote the implementation of water-saving policies.

The current situation of national water-saving zoning in the three major fields including agriculture, industry, and urban living in existing policy documents and academic papers was analyzed in detail through literature research. 31 provincial administrative regions were taken as the basic unit, and the principles of water-saving regionalization were proposed. Fourteen indicators from five aspects were selected to construct the index system of water-saving regionalization including water resources conditions, social and economic conditions, agricultural water use characteristics, industrial water use characteristics and domestic water use characteristics. The coupled method of principal component analysis (PCA) and hierarchical cluster analysis (HCA) was adopted to delimit water-saving zoning. The characteristics of water resources, economy and society, and water use of different fields in the six water-saving zones were analyzed.

Results showed that China's water-saving zoning research was mainly concentrated in the field of agriculture, and there were few zoning studies in the field of industrial and urban water-saving. The main reason is that China's agricultural water-saving development started earlier, and agricultural water-saving zoning is greatly affected by regional and environmental factors, which have regularity in large-scale and mesoscale, and differentiation between regions. Industrial water saving is mainly affected by factors including industrial structure, technical conditions, and local policies, whose spatial structure is scattered. It is difficult to find the pattern of spatial differentiation on large-scale, mesoscale, and even small-scale. Urban water saving is mainly affected by factors including water resources conditions, economic and social development level, and living habits. It is regular from the perspective of large-scale and mesoscale and differentiated between regions. However, due to the relatively small proportion of urban domestic water use, the research on its water-saving zoning has not been paid enough attention. By the coupled method of principal component analysis and hierarchical cluster analysis, and tuning small part of the provinces according to the principle of concentration and contiguity in zoning, the 31 provincial administrative regions in China were divided into six water-saving subregions: northeast China (Liaoning, Jilin, Heilongjiang), north China (Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia, Shandong, Henan, Shaanxi), northwest China (Gansu, Qinghai, Ningxia, Xinjiang), southeast coastal area of China (Shanghai, Jiangsu, Zhejiang, Fujian, Guangdong, Guangxi, Hainan), central China (Anhui, Jiangxi, Hubei, Hunan), and southwest China (Chongqing, Sichuan, Guizhou, Yunnan, Tibet). Feature analysis showed that there were significant differences among different regions, and the shortage of water resources and the development of the economy and society played a significant role in promoting water conservation.

The six water-saving subregions are consistent with the existing zoning pattern, which shows that the zoning method established has good reliability. The research results can provide strong support for the rational determination of water-saving strategic layout and scientific formulation of water-saving promotion strategies in China.

**Key words:** water-saving regionalization; principal component analysis; hierarchical cluster analysis; index system; feature analysis