

文章编号:1671-6833(2011)01-0125-04

基于主成分分析的农业旱灾等级区划研究

马细霞, 李艳, 刘磊

(郑州大学 水利与环境学院, 河南 郑州 450001)

摘要:旱灾是制约一个地区经济发展的主要因素之一,旱灾的等级区划又是农业抗旱能力综合评价的基础,目前对旱灾等级区划的方法比较主观且单一化.笔者在分析主成分分析法的基础上,建立了基于主成分分析的农业综合干旱及旱灾区划模型,并将其应用于东北四个地区的农业旱灾区划,评价出了四个地区的区划等级都属于中度脆弱区.用层次分析法的评价结果与之相比较,两种评价结果一致,且与实际情况相符,表明主成分分析法是一种有效的区划农业干旱及旱灾的方法.

关键词:农业综合干旱;主成分分析;旱情

中图分类号:S271 **文献标志码:**A

0 引言

我国处于东亚大陆的东南部,是典型的季风气候国家.夏季受季风和台风的影响,冬季受西伯利亚干冷空气的影响,降雨年际、年内变化异常强烈,一年内有60%的雨量集中在夏秋两季的100天左右,加之我国地域辽阔,地形复杂,降雨地域分布极不均衡^[1],容易出现覆盖面广、持续时间长的旱情.由于受气候和人类活动影响,近年来,干旱发生日益频繁,严重以上干旱出现的次数也有所增加.2009年河南遭受1951年以来的特大旱情,持续受旱时间3个多月,受灾面积达63.1%.2010年云南、贵州、广西、重庆、四川发生严重干旱,局部地区持续受旱时间超过5个月,其中云南大部、贵州西部和广西西北部已达特大干旱等级.另外全球气候变暖出现的各种迹象显示,目前很可能进入一个新的百年尺度的干旱期,干旱发生的频率和强度必然增加,甚至有可能发生特大干旱灾害^[2].为了保障城乡供水安全、粮食安全和生态与环境安全,逐步提高我国抗旱减灾能力和管理水平,主动应对日益严重的干旱灾害,最大可能减轻旱灾损失,为经济社会全面协调可持续发展提供有力支撑,迫切需要对旱灾及其相关问题进行深入研究.

2008年12月,水利部办公厅发出了《关于开

展抗旱规划编制工作的通知》(办规计函[2008]542号),在全国范围内部署开展抗旱规划编制工作,并印发了《抗旱规划工作大纲》.但是,目前国内对旱情预警、评估、旱灾等级区划等方面的研究还处于起步阶段,少有文献见诸报端.为了弄清干旱及农业由于干旱而造成损失的地区分布情况,倪深海等人曾应用层次分析法绘制了中国农业干旱脆弱性分区图,确定了各地区的农业综合干旱区划等级^[3].对于灾害脆弱性分区的问题,武玉艳等人用熵值法对灾害脆弱性进行了评价^[4].但是层次分析法的评价结果主观性强,易与实际情况不一致,熵值法虽客观但不能有效的将具有相关性的指标分开.笔者针对灾害脆弱性区划目前存在的一些问题引入主成分分析的方法,这种方法能客观地确定权重,并且能在保证原始数据信息损失最小的情况下,用少数的综合变量来代替原有的多维变量,使数据结构简化,得到的主成分是原始变量的线性组合且两两不相关^[5].笔者利用主成分分析方法对东北4个地区的农业综合干旱进行评价,并与层次分析法评价结果进行对比分析,验证了本文方法的可行性,为农业旱灾等级区划提供了一种新途径.

1 主成分分析方法

主成分分析最早是由 Pearson 在 1901 年的生

收稿日期:2010-08-20;修订日期:2010-10-08

基金项目:河南省自然科学基金项目(0411050800)

作者简介:马细霞(1963-),女,河南叶县人,郑州大学教授,主要从事水文学及水资源方面的教学与科研工作, E-mail: maxx@zzu.edu.cn.

物理学理论研究中引入,后来在多元统计中被广泛应用.多元统计中的评价指标可以看成是指标变量,每一变量都不同程度地反映了该类指标的信息,变量之间难免存在重叠、相关的关系.在用现代多元统计方法研究多变量问题时,变量太多会增大计算量和增加问题的复杂性,人们自然希望在进行定量分析的过程中涉及的变量应尽可能少,而信息量尽可能多.主成分分析是解决这一问题的理想工具,其思想就是在保证原有信息损失很少的前提下,通过降维将原有的多个指标转化为一个或几个综合指标,即主成分.每个主成分都是原始变量的线性组合,并且各个主成分之间互不相关^[6].

主成分分析法主要步骤^[7]如下.

步骤1 数据指标矩阵. 设农业综合干旱及早灾区划的评价体系指标个数为 n 个,区划的地区个数为 m 个,数据矩阵为 x , x_{ij} 代表第 i 个地区,第 j 个指标的数值.

$$x = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$$

步骤2 指标的标准化处理. 由于各指标数据量纲不一致,数量间的差异很大,需要将不同度量的指标转化为同度量的指标,使各指标具有可比性.在标准化处理时没有采用数学统计方法中的归一化处理,而是与区划标准联系起来,将各个指标与区划标准进行比较,从而能更好地把指标与实际情况联系起来.小于一级标准的取 1,大于五级标准的取 5,其间的由内插得到.

步骤3 相关系数矩阵的计算. 设 r_{ij} 为标准后第 i 个指标与第 j 个指标之间的相关系数,其计算公式为:

$$r_{ij} = \frac{\sum_{k=1}^n |(x_{ki} - \bar{x}_i)| |(x_{kj} - \bar{x}_j)|}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}} \quad (1)$$

可得相关系数矩阵 R .

步骤4 特征值和特征向量的计算. 利用上述相关关系矩阵 R , 求出矩阵的 n 个特征值 λ_i ($i = 1, 2, \dots, n$), 按其大小顺序排列, 即 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$, 然后分别求出特征向量 e_i ($i = 1, 2, \dots, n$). 方差大的新变量对模型的贡献大, 方差小的新变量对模型贡献小. 新变量 y_1, y_2, \dots, y_n 分别称为第一主成分, 第二主成分, \dots , 第 n 主成分.

$$\begin{cases} y_1 = e_1^{(1)} x_1 + e_2^{(1)} x_2 + \cdots + e_n^{(1)} x_n \\ y_2 = e_1^{(2)} x_1 + e_2^{(2)} x_2 + \cdots + e_n^{(2)} x_n \\ \vdots \\ y_n = e_1^{(n)} x_1 + e_2^{(n)} x_2 + \cdots + e_n^{(n)} x_n \end{cases} \quad (2)$$

步骤5 计算主成分贡献率及累计贡献率.

主成分贡献率:

$$H_i = \lambda_i / \sum_{i=1}^n \lambda_i \quad (i = 1, 2, \dots, n) \quad (3)$$

累积贡献率:

$$TH_i = \sum_{i=1}^q \lambda_i / \sum_{i=1}^n \lambda_i \quad (i = 1, 2, \dots, n) \quad (4)$$

步骤6 主成分的选取. 一般若前 q 个主成分提供了 85% 以上的信息量, 则取特征值 $\lambda_1, \lambda_2, \dots, \lambda_q$ 为对应的第一、第二、 \dots 、第 q 个主成分, 变量数由 n 个减至 q 个, 产生了降维的效果 (其中 $q < n$)^[5].

步骤7 综合评价. 根据 n 个指标在第 i 个公共因子上的荷载向量 $e_i = (e_i^{(1)}, e_i^{(2)}, \dots, e_i^{(n)})$, 得第 j 个指标在全部 q 个公共因子上荷载的平方和 (指标的公共性方根).

$$h_j = \frac{\sqrt{\lambda_i} e_j^{(i)}}{\sum_{i=1}^q (e_j^{(i)} \sqrt{\lambda_i})^2} \quad (5)$$

假如选取了两个主成分, 则第一个指标在这两个主成分上的公共性方根即为

$$h_1 = \frac{\sqrt{\lambda_1} e_1^{(1)}}{\sqrt{(\sqrt{\lambda_1} e_1^{(1)})^2 + (\sqrt{\lambda_2} e_1^{(2)})^2}}$$

它的大小反映了第 j 个指标在公共性部分的作用或重要性程度. 比较 n 个指标的公共性, 可知哪一个指标在公共性方面起的作用大. 所以把归一化后每个指标的公共性方根 h_j 作为该变量的权重, 构成一个度量农业综合干旱及早灾区划的综合指标:

$$Y = \sum_{i=1}^n h_i x_i \quad (6)$$

式中: h_i 为归一化后的各指标的权重; x_i 为各个指标数据标准化后的值.

2 实例分析

东北地区大部分地处中温带, 只有最北部属于寒温带, 属寒温带过渡的大陆性季风气候区, 一般特点是冬季干燥酷寒, 春季大风雨少. 全境日气温由南向北递减, 年降水量由东向北递减, 大部分地区属于半湿润半干旱地区. 现采用笔者方法对东北地区的农业综合干旱及早灾进行区划.

水资源承载能力、抗旱能力、农业旱灾情况等直接影响着该地区农业综合干旱及旱灾的区划。经综合分析,笔者选取人均水资源量、亩均水资源量、耕地灌溉率、水利工程调蓄率、旱灾频率、受旱率和成灾率作为分区指标,东北四个地区的农业

综合干旱及旱灾区划各项指标值见表1,各分区指标等级标准^[8]见表2。其中一级为极度脆弱区;二级为强度脆弱区;三级为中度脆弱区;四级为轻度脆弱区;五级为微度脆弱区。

表1 农业旱灾区划各项指标值

Tab.1 Index system digital about agricultural drought regionalization

地区	人均水资源量/ ($\text{m}^3 \cdot \text{人}^{-1}$)	亩均水资源量/ ($\text{m}^3 \cdot \text{亩}$)	耕地灌溉率/ %	水利工程蓄 水率/%	干旱频率/ %	受旱率/ %	成灾率/ %
A	340.43	142.09	0.188 5	0.188 1	0.461 5	0.152 0	0.073 0
B	508.37	178.38	0.199 9	0.276 4	0.519 2	0.101 9	0.061 7
C	380.82	257.09	0.319 7	0.430 0	0.153 8	0.057 3	0.039 4
D	643.71	641.16	0.282 3	0.642 0	0.480 7	0.124 5	0.079 0

表2 农业旱灾区划指标等级标准

Tab.2 Classification standard of index of agricultural drought regionalization

指标		等级				
		I	II	III	IV	V
水资源承载能力	人均水资源量/ $(\text{m}^3 \cdot \text{人}^{-1})$	<500	500~1 000	1 000~1 670	1 670~10 000	>10 000
	亩均水资源量/ $(\text{m}^3 \cdot \text{亩}^{-1})$	<200	200~500	500~800	800~1 100	>1 100
抗旱能力	耕地灌溉率/%	<10	10~30	30~60	60~80	>80
	水利工程调蓄率/%	<10	10~30	30~60	60~80	>80
农业旱灾	旱灾频率/%	>80	80~60	60~40	40~20	<20
	受旱率/%	>40	40~30	30~20	20~10	<10
	成灾率/%	>20	20~15	15~10	10~5	<5

按照本文前述方法对表1指标进行标准化处理,计算其相关系数矩阵:

$R =$

1.000 0	0.941 5	0.829 9	0.962 6	0.390 3	0.839 6	0.949 9
0.941 5	1.000 0	0.676 9	0.986 2	0.282 4	0.737 9	0.968 5
0.829 9	0.676 9	1.000 0	0.747 4	0.778 7	0.860 9	0.728 5
0.962 6	0.986 2	0.747 4	1.000 0	0.311 3	0.838 7	0.940 4
0.390 3	0.282 4	0.778 7	0.311 3	1.000 0	0.408 2	0.408 2
0.839 6	0.737 9	0.860 9	0.838 7	0.408 2	1.000 0	0.666 7
0.949 9	0.968 5	0.728 5	0.940 4	0.408 2	0.666 7	1.000 0

由公式(2)、(3)和(4)计算特征值、主成分贡献率和累计贡献率,得 $\lambda_1 = 5.48, \lambda_2 = 1.05, \lambda_3 = 0.40, \lambda_4 = 0.07, \lambda_5 = 0, \lambda_6 = 0, \lambda_7 = 0$,第一主成分的贡献率为78%,第二主成分的贡献率为15%,两个主成分的累计贡献率为93% > 85%。由公式(5)计算得到各指标的权重为: $w = (0.146 2, 0.145 5, 0.146 0, 0.147 2, 0.143 9, 0.129 9, 0.141 0)$ 。由权重计算得到东北4个地区的农业综合干旱及旱灾区划等级见表3。为验证区划结果的合理性,笔者同时采用层次分析法进行了计算,其区划结果一并列入表3中。

对表3中两种方法区划结果进行对比分析,

发现两种方法评价结果基本一致,4个地区都属于中度脆弱区,与实际情况相符。东北地区水资源少,但是其水利工程的利用率较高,易发生季节性干旱及旱灾。

表3 两种方法得出的区划等级对照表

Tab.3 Comparison table about two methods of division classification

地区	A	B	C	D
主成分分析法计算结果	2.69	2.77	2.78	3.00
主成分分析法区划结果	III	III	III	III
层次分析法区划结果	II	III	III	III

3 结论

(1) 主成分分析能够将具有相关性的指标降维,简化指标体系。笔者基于主成分分析的理论方法,综合考虑农业受旱频率、农田受旱率、成灾率、耕地灌溉率以及当地的水资源状况等因素,建立了农业综合干旱及旱灾的区划模型,区划结果是东北地区的四个地方都属于中度脆弱区,这与用层次分析法得到的区划结果基本一致,表明主成分分析法是一种行之有效的办法。

(2)主成分分析法在标准化时与农业综合干旱及旱灾的等级标准结合了起来,与以往单一的用数学统计中的标准化不同,这样能更好的把指标与实际情况联系起来。

(3)在全球变暖的趋势下,干旱频率呈上升的趋势,影响干旱的因素及程度也愈渐复杂,文中所用的评价体系仍需要进一步的完善。但是提出用主成分分析的方法进行区划,仍有助于有关部门方案的决策,进而提升抗旱效率。

参考文献:

- [1] 姜德贵. 旱情评价及抗旱减灾研究[D]. 合肥工业大学水工结构工程系, 2002.
- [2] 李伟娟. 基于GIS的黑龙省甘南县旱情等级研究[D]. 东北农业大学农业水土工程系, 2008.
- [3] 倪深海, 顾颖, 王会容. 中国农业干旱脆弱性分区研究[J]. 水科学进展, 2005, 16(5): 705-709.
- [4] 武玉艳, 葛兆帅, 蒲英磊, 等. 基于熵值法的农业洪涝灾害脆弱性评价[J]. 安徽农业科学, 2009, 37(4): 1681-1682.
- [5] 万新宇, 包为民, 荆艳东, 等. 基于主成分分析的洪水相似性研究[J]. 水电能源科学, 2007, 25(5): 36-39.
- [6] 刘春龙. 改进的主成分分析法在水质评价中的应用[J]. 安徽农业科学, 2009, 37(22): 10642-10643.
- [7] 李祚泳, 丁晶, 彭荔红. 环境质量评价原理与方法[M]. 北京: 化学工业出版社, 2004.
- [8] 水利部水利水电规划设计总院. 中国抗旱战略研究[M]. 北京: 中国水利水电出版社, 2008.

Study on Regionalization of Agricultural Drought Based on the Principal Component Analysis

MA Xi-xia, LI Yan, LIU Lei

(School of Water Conservancy and Environment, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Drought is a major contributing factor to regional economic development, and its grading is the basis of comprehensive evaluation of drought resistance capacity. However, the drought grading method is fairly subjective and single. This article introduces the method and thought about principal components analysis. In this foundation, the classification of agricultural combined drought model based on PCA is founded. The model is applied to evaluating and ranking four areas drought in Northeast China area and they are compared with each other, then we came to the conclusion that the four areas all fall into medium fragile areas. This is the same as the result using the analytic hierarchy process. Furthermore, it corresponds to the practical situation. The conclusion shows that the principal components analysis is an effective method for grading drought.

Key words: agriculture combined drought; principal components analysis; drought