

DOI:10.3969/j.issn.1004-6933.2013.04.010

K-L PPC 模型在水资源可持续利用评价中的应用

常 进,王卫平,陈 星,张其成,向 龙

(河海大学水文水资源学院,江苏 南京 210098)

摘要:基于水资源可持续利用的内涵和意义,从水资源、社会经济和生态环境 3 个子系统建立水资源可持续利用指标体系;针对水资源可持续利用评价中各指标的不确定性和模糊性,为提高评价模型的分辨率,引入基于实数编码的加速遗传算法的投影寻踪分类(K-L PPC)模型。以塔里木河源流区为例,通过优选指标、确立等级和构建模型,寻找最优的投影方向。评价结果基本符合当地实际情况,验证了 K-L PPC 模型在区域水资源可持续利用评价的实用性。

关键词:K-L PPC 模型;水资源可持续利用;RAGA;评价

中图分类号:TV213.4 **文献标志码:**A **文章编号:**1004-6933(2013)04-0050-04

Application of K-L PPC model to evaluation of sustainable utilization of water resources

CHANG Jin, WANG Weiping, CHEN Xing, ZHANG Qicheng, XIANG Long
(College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China)

Abstract: Based on the meaning and significance of the sustainable utilization of water resources, an index system was established by considering water resources conditions, social economy, and the eco-environment. Considering the uncertainties and vagueness of the evaluation indicators for the sustainable utilization of water resources, the K-L PPC model, based on the real-coded accelerating genetic algorithm, was developed in order to raise the resolution of the evaluation model. A case study was conducted in the source area of the Tarim River. The optimal projection direction was determined through index optimization, grade determination, and model construction. The evaluation results were in accordance with the actual local situation, verifying the practicability of the K-L PPC model in the evaluation of the sustainable utilization of water resources. This provides a scientific basis for the protection and sustainable utilization of water resources.

Key words: K-L PPC model; sustainable utilization of water resources; RAGA; evaluation

水资源可持续利用是指在维持水的持续性和生态系统整体性的条件下,支持人口、资源、环境与经济协调发展和满足代内和代际人用水需要的全部过程。实现水资源的可持续利用,发挥水资源的经济、社会和生态效益,支持经济社会的可持续发展,是今后长时期内的重要战略任务。关于水资源可持续利用评价的研究方法,以往大多采用模糊综合评价法、灰色聚类法、模糊神经网络法、物元分析法等,这些方法的模型计算大多存在人为赋权的干扰以及等级

分辨率较粗的不足,由于综合评价往往由多个非线性指标来决定,采用传统的数据分析方法建立聚类和评价模型时,受到数字化限制,难以找到数据的内在规律^[1]。鉴于此,笔者以塔里木河源流区为研究对象,从水资源、社会经济及生态环境 3 个子系统构建水资源可持续利用评价指标体系,将基于实数编码的加速遗传算法(real coded accelerating genetic algorithm,简称 RAGA)^[2]的 K-L 绝对信息散度投影寻踪分类模型(Kullback-Leibler absolute information

基金项目:国家自然科学基金(51209071)
作者简介:常 进(1988—),男,硕士研究生,研究方向为生态水文与水环境保护。E-mail:825855653@qq.com

divergence of projection pursuit classification model, 简称 K-L PPC) [3-4] 用于水资源可持续利用能力评价中, 该模型更好地揭示了高维数据结构特征, 将高维数据投影到低维子空间, 在低维数据上进行分析, 从而达到分析高维数据的目的, 为区域水土资源开发、社会经济发展和生态环境保护提供科学支持。

1 K-L PPC 模型介绍

1.1 加速遗传算法 (RAGA)

遗传算法 (GA) 是一种以达尔文自然进化论和孟德尔遗传变异理论为基础的全局随机搜索优化计算技术, 它同时考察多个候选解, 模拟生物遗传进化过程, 淘汰劣质解, 鼓励发展优质解, 逐步提高解群体的质量, 直至收敛, 获得 (准) 最优解, 主要包括选择、交叉和变异等操作 [5]。加速遗传算法是遗传算法的一种改进算法, 它利用在运行过程中产生的优秀个体逐步调整优化变量的初始变化区间, 形成加速运行 [6]。实践表明, 加速遗传算法的寻优效果要强于前者。

1.2 投影寻踪分类模型 (PPC)

投影寻踪分类模型是投影寻踪模型的一种, 其基本思路是: 把高维数据投影到低维 (1/2 维) 子空间上, 并通过最优化某个投影指标, 寻找出能反映高维数据结构或特征的投影, 在低维空间上对数据结构进行分析, 以达到研究和分析高维数据的目的 [7]。它避免了专家打分的人为干扰因素, 在定量评价指标数据的处理上更具优势。

1.3 K-L PPC 模型构建

K-L 信息散度是由 Kullback 和 Leibler 在 Shannon 互信息的基础上提出的, 它很好地度量了两个分布之间的距离。一般认为服从正态分布的数据含有的有用信息最少, 因而笔者感兴趣的是与正态分布差别大的结构。多元正态分布的任何一维线性投影仍然服从正态分布, 因此如果一个数据在某个方向上的投影与正态分布差别较大, 那它就一定含有非正态的结构。高维数据在不同方向上的一维投影与正态分布的差别是不一样的, 它显示了在这一方向上所持有的有用信息的多少, 因此可以用投影数据的分布与正态分布的差别来作为投影指标 [8]。PPC 模型的准确性和稳定性决定于密度窗宽的确定科学与否, 而 K-L PPC 不依赖于任何参数的确定, 用投影数据的分布与正态分布的差别来作为投影指标, 相比于 PPC 更加稳定科学。

K-L PPC 建模过程包括如下 4 个步骤 [8-9]:

a. 样本评价指标集的归一化处理。设各指标值的样本集为 $\{x^*(i, j) | i = 1, 2, \dots, n, j = 1, 2, \dots,$

$p\}$, 其中 $x^*(i, j)$ 为第 i 个样本第 j 个指标值, n, p 分别为样本的个数 (样本容量) 和指标的数目。为消除各指标值的量纲和统一各指标值的变化范围, 可采用式 (1) 进行极值归一化处理:

$$x(i, j) = \begin{cases} \frac{x^*(i, j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} & \text{正向指标} \\ \frac{x_{\max}(j) - x^*(i, j)}{x_{\max}(j) - x_{\min}(j)} & \text{负向指标} \end{cases} \quad (1)$$

式中: $x(i, j)$ 为指标特征值归一化的序列; $x_{\max}(j)$ 、 $x_{\min}(j)$ 分别为第 j 个指标值的最大值和最小值。

b. 构造正态分布函数 $Q(k_3, k_4)$ 。把 p 维数据 $\{x(i, j) | j = 1, 2, \dots, p\}$ 综合成以 $A = \{A(1), A(2), A(3), \dots, A(p)\}$ 为投影方向的一维投影值 $Z(i)$ 。

$$Z(i) = \sum_{j=1}^p A(j)x(i, j) \quad (i = 1, 2, \dots, n) \quad (2)$$

基于正态分布函数的偏度 $k_3 = \frac{E[(Z - \bar{Z})^3]}{\sigma^3}$ 和

峰度 $k_4 = \frac{E[(Z - \bar{Z})^4]}{\sigma^4} - 3$ 都等于零, 因此在约束条件

$\sum_{j=1}^p A^2(j) = 1 (0 \leq A(j) \leq 1)$ 下, 最小化函数 $Q(k_3, k_4) = |k_3| + |k_4|$ 即可得到接近于正态分布的一维投影值 $Z(i)$ 。

c. 构造投影指标函数 $Q(z, Z)$ 。把 p 维数据 $\{x(i, j) | j = 1, 2, \dots, p\}$ 综合成以 $a = \{a(1), a(2), a(3), \dots, a(p)\}$ 为投影方向的一维投影值 $z(i)$ 。

$$z(i) = \sum_{j=1}^p a(j)x(i, j) \quad (i = 1, 2, \dots, n) \quad (3)$$

利用 K-L 信息散度作为投影指标, 目的是能够寻找投影矩阵, 使得投影以后的数据分布与正态分布有最大的差异。 $Z(i)$ 、 $z(i)$ 均是离散值向量, 最佳投影方向就是最大可能暴露高维数据某类特征结构的投影方向, 因此可以通过求解投影指标函数最大化问题来估计最佳投影方向, 即:

$$\max: Q(z, Z) = |L_D(z, Z)| + |L_D(Z, z)| \quad (4)$$

约束条件:

$$\sum_{j=1}^p a^2(j) = 1 (0 \leq a(j) \leq 1) \quad (5)$$

其中

$$L_D(z, Z) = \sum \left| z_i \log \frac{z_i}{Z_i} \right|, L_D(Z, z) = \sum \left| Z_i \log \frac{Z_i}{z_i} \right|$$

d. 分类评价 (优序排列)。把由步骤 c 求得的投影方向代入式 (3) 后可得各样本点的投影值 $Z^*(i)$ 。将 $Z^*(i)$ 与 $Z^*(j)$ 进行比较, 二者越接近, 表示样本 i 与 j 越倾向于同一类。若按 $Z^*(i)$ 值从大到小排序, 则可以将样本从优到劣进行排序。

2 应用分析

2.1 研究区概况

塔里木河地处我国西部干旱区,流域总面积 102 万 km²,其中干流长 1 321 km,是我国最长的内陆河。本文中所指的塔里木河三源流为和田河、叶尔羌河和阿克苏河,它们分别位于和田、喀什和阿克苏地区,是目前塔里木河干流的主要补给河流。塔里木河源流区水资源总量为 263.9 亿 m³,其中,地表水资源量为 246.4 亿 m³,地下水资源量为 193.2 亿 m³。塔里木河流域气候干旱,降雨极少,蒸发强烈,生态环境十分脆弱^[10-11]。在过去 50 a 中,塔里木河在以水资源开发利用为核心的大强度人类经济、社会活动的作用下,流域自然生态过程发生了显著变化,经济与生态矛盾日趋突出。上游三源流向干流输送的水量逐年减少,水质不断恶化,下游 320 km 的河道断流^[12]。为此,深刻了解源流区水资源状况,科学评价流域水资源可持续利用,对流域水资源科学管理和社会经济健康发展有着重要意义。

2.2 评价体系的建立及分级标准制定

评价指标体系的建立是水资源可持续评价的难点之一,随着水资源可持续利用研究的不断深入,已有很多关于评价指标体系建立的研究。笔者结合研究区的实际情况,从水资源、社会经济和生态环境 3 个子系统选取 10 项评价指标,制定了 4 级评价标准。水资源子系统选取 6 个评价指标,分别为人均水资源量 C_1 、人均供水量 C_2 、人均日生活水量 C_3 、供水模数 C_4 、水资源开发利用率 C_5 、耕地灌溉率 C_6 ;社会经济子系统选取两个指标,分别为 GDP 增长率 C_7 、人均 GDP C_8 ;生态环境子系统选取两个指标,分别为生态用水率 C_9 、干旱指数 C_{10} 。上述各因素反映了影响区域水资源可持续利用的供需水、社会经济以及环境保护等 3 个主要方面的情况,这些

因素对塔里木河水资源影响较大,是水资源可持续利用评价中较重要的因素。根据各评价指标的量值范围,将区域水资源可持续利用度划分为 4 个等级:Ⅰ级为水资源可持续利用好——基本未开发状态,Ⅱ级为水资源可持续利用较好——可持续利用状态,Ⅲ级为水资源可持续利用一般——弱可持续状态,Ⅳ级为水资源可持续利用较差——维持可持续利用困难。塔里木河源头区水资源可持续利用评价指标体系与评价标准及评价指标数据分别见表 1 和表 2。

2.3 模型的建立与结果分析

表 1 中各等级取值范围内均匀随机产生各 5 个样本 $x^*(i,j)$,与对应的等级一起组成样本系列,并对 $x^*(i,j)$ 进行归一化处理为 $x(i,j)$, ($i=1, 2, \cdots, 20; j=1, 2, \cdots, 13$),利用步骤 b 计算出 $A=(0.1211, 0.5180, 0.6431, 0.0790, 0.3399, 0.3072, 0.0815, 0.0185, 0.1965, 0.2042)$,将投影方向 A 代入式 (3) 即得一维投影值 $Z(i)$ 。利用步骤 c 计算出最大投影指标函数值 $Q(z,Z)=31.1561$,最佳投影方向 $a=(0.0378, 0.0373, 0.0351, 0.4910, 0.2594, 0.3829, 0.0916, 0.0483, 0.0541, 0.7262)$,最佳投影值 $z(i)=(2.1637, 1.1660, 0.7924, 0.6213, 0.5270, 0.4228, 0.3642, 0.3225, 0.2912, 0.2645, 0.2422, 0.2199, 0.2026, 0.1886, 0.1730, 0.1530, 0.1299, 0.1119, 0.0953, 0.0796)$ 。从而得出塔里木河源流区Ⅰ级水资源可持续利用能力投影值大致范围为 0.5270~2.1637,Ⅱ级投影值大致范围为 0.2645~0.4228,Ⅲ级投影值大致范围为 0.1730~0.2422,Ⅳ级大致投影值范围为 0.0796~0.1530。

对表 2 中的塔里木河源流区 4 个地区水资源可持续利用能力评价指标数据进行归一化,用最佳投影方向计算出 4 个地区水资源可持续利用能力,投影值 $Z^*(i)=(0.2299, 0.2667, 0.2314, 0.2315)$ 。

表 1 塔里木河源流区水资源可持续利用评价指标体系与评价标准

级别	$C_1/$ ($m^3 \cdot 人^{-1}$)	$C_2/$ ($m^3 \cdot 人^{-1}$)	$C_3/$ ($m^3 \cdot 人^{-1} \cdot d^{-1}$)	$C_4/(10^3$ $m^3 \cdot km^{-3})$	$C_5/\%$	$C_6/\%$	$C_7/\%$	$C_8/$ (元·人 ⁻¹)	$C_9/\%$	C_{10}
Ⅰ	(7 000,10 000)	(4 500,5 000]	(130,160]	[0,10)	[0,30)	[0,20)	(8,20]	(10 000,15 000]	(5,8]	[0,7)
Ⅱ	(5 000,7 000]	(3 500,4 500]	(100,130]	[10,35)	[30,50)	[20,40)	(6,8]	(6 000,10 000]	(3.5,5]	[7,25)
Ⅲ	(3 500,5 000]	(2 500,3 500]	(70,100]	[35,60)	[50,70)	[40,60)	(4,6]	(3 000,6 000]	(2,3.5]	[25,40)
Ⅳ	[1 750,3 500]	[1 250,2 500]	[35,70]	[60,120]	[70,140]	[60,120]	[2,4]	[1 500,3 000]	[0,2]	[40,80]

表 2 塔里木河源流区水资源可持续利用评价指标数据

地区	$C_1/$ ($m^3 \cdot 人^{-1}$)	$C_2/$ ($m^3 \cdot 人^{-1}$)	$C_3/$ ($m^3 \cdot 人^{-1} \cdot d^{-1}$)	$C_4/$ ($10^3 m^3 \cdot km^{-3}$)	$C_5/\%$	$C_6/\%$	$C_7/\%$	$C_8/$ (元·人 ⁻¹)	$C_9/\%$	C_{10}
喀什	2 180.75	2 834.25	57.56	56.83	133.72	100.25	19.40	4 508.14	1.15	21.04
和田	5 975.50	2 233.75	38.77	24.88	38.70	93.77	11.50	2 891.75	7.28	39.35
阿克苏	3 420.25	4 423.50	69.15	44.18	130.92	92.89	12.98	8 190.00	0.06	20.50
三源流	3 389.01	3 137.64	63.03	38.72	93.83	96.42	15.54	5 197.88	1.71	26.96

注:数据来源于 2005—2008 年新疆统计年鉴;蒸发量与降水量来自文献[13]。

将 $Z^*(i)$ 与各级水资源可持续利用能力投影值相比,从而得出 4 个地区水资源可持续利用能力等级:和田地区评价等级为Ⅱ级,喀什、阿克苏和三源流 3 个地区评价等级为Ⅲ级。可见,塔里木河源流区水资源可持续利用状况不容乐观,除和田地区外,其余基本呈现弱不可持续状态,具体为和田地区(Ⅱ)>三源流地区(Ⅲ)>阿克苏地区(Ⅲ)>喀什地区(Ⅲ),与实际情况相符。究其原因:西部干旱区经济以灌溉农业为典型特点的绿洲农业为主体,加上区域水利设施落后,耕地灌溉定额高,经济的发展需要耗费大量水资源,导致区域生态用水受到挤占。和田地区要努力改善现有的经济状况,提高水资源开发利用势在必行;喀什、阿克苏和三源流地区在保持现有的经济状况的基础上,进一步优化产业结构,发展生态农业,提高用水效率,加大生态需水和维护生态环境的投入,科学引导区域水土资源的开发。

3 结 语

- a. 区域水资源可持续利用评价是在一定时空范围内的水资源系统中、对多个因子及其协调状况的综合评价,是一个复杂大系统多目标发展的问题。笔者将 K-L PPC 模型应用到水资源可持续利用评价中,将高维数据映射到低维子空间,避免了人为因素等干扰,使模型更具科学性和稳定性。模型计算简便、适用性强,有利于投影寻踪方法的推广应用,给水资源可持续利用评价提供了一条值得探索的新途径。
- b. 塔里木河三源流区在满足生态需水、维护生态环境的前提下,为保持区域经济的稳定增长势头,应调整农业种植结构,推广节水技术,完善水利设施,提高水资源利用效率,适当增加地下水的开采,避免因水土开发而导致的水土流失。

参考文献:

[1] 付强,付红,王丽坤. 基于加速遗传算法的投影寻踪模型在水质评价中的应用研究[J]. 地理科学,2003,23 (2):236-239. (FU Qiang,FU Hong,WANG Likun. Study on the PPE model based on RAGA to evaluating the water quality[J]. Scientia Geographica Sinica,2003,23 (2):236-239. (in Chinese))

[2] 金菊良,杨晓华,丁晶. 标准遗传算法的改进方案:加速遗传算法[J]. 系统工程理论与实践,2000(4):8-13. (JIN Juliang,YANG Xiaohua,DING Jing. An improved simple genetic algorithm: accelerating genetic algorithm [J]. Systems Engineering-Theory & Practice,2000(4):8-13. (in Chinese))

[3] FRIEDMAN J H, TURKEY J W. A projection pursuit

algorithm for exploratory data analysis[J]. IEEE Trans On Computer,1974,23(9):881-890.

[4] FU Q, XIE Y G, WEI Z M. Application of projection pursuit evaluation model based on real-coded accelerating genetic algorithm in evaluating wetland soil quality variations in the Sanjiang Plain, China [J]. Pedosphere 2003,13(3):249-256.

[5] HOLLAND J H. Adaptation in natural and artificial systems [M]. Cambridge:The MIT Press,1975:70-116.

[6] SANNASIRAJ S A, SUNDAR V, SUNDARAVADIVELU R. The hydrodynamic behaviour of long floating structures indirectional seas [J]. Applied Ocean Research,1995,17:233-243.

[7] BLACK J L, MEI C C, BRAY C G. Radiation and scattering of water waves by rigid bodies [J]. J Fluid Mech,1971,46:151-164.

[8] 赵小勇,崔广柏,付强. K-L 绝对信息散度投影寻踪分类模型及其应用[J]. 人民长江,2010,41(15):91-93. (ZHAO Xiaoyong,CUI Guangbai,FU Qiang. K-L absolute information divergence of projection pursuit classification model and the application in irrigation engineering[J]. Yangze River,41(15):91-93. (in Chinese))

[9] 赵小勇,付强,邢贞相,等. 投影寻踪模型的改进及其在生态农业建设综合评价中的应用[J]. 农业工程学报,2006,22(5):222-225. (XIAO Xiaoyong,FU Qiang,XING Zhenxiang,et al. Improvement of projection pursuit model and its application in integrated evaluation of eco-agricultural construction[J]. Transactions of the CSAE,2006,22(5):222-225. (in Chinese))

[10] 王让会,樊自立. 塔里木河下游近 50 a 来沙质荒漠化演变规律[J]. 中国沙漠,2000,20(1):45-50. (WANG Ranghui, FAN Zili. Evolution patterns of sandy desertification in the lower reaches of Tarim River Basin in recent 5 decades[J]. Journal of Desert Research,2000,20 (1):45-50. (in Chinese))

[11] 赵振勇,王让会,张慧芝,等. 塔里木河下游荒漠生态系统退化机制分析[J]. 中国沙漠,2006,26(2):220-225. (ZHAO Zhenyong, WANG Ranghui, ZHANG Huizhi, et al. Degradation mechanism of desert ecosystem in lower reaches of Tarim River[J]. Journal of Desert Research,2006,26(2):220-225. (in Chinese))

[12] 贡璐,王宏卫,鲍平勇. 塔里木河上游典型绿洲景观变化及其生态效应分析[J]. 中国沙漠,2006,26(3):421-425. (GONG Lu, WANG Hongwei, BAO Pingyong, et al. Landscape change and ecological effect of typical oasis in the upper reaches of Tarim River[J]. Journal of Desert Research,2006,26(3):421-425. (in Chinese))

[13] 邓铭江. 中国塔里木河治水理论与实践[M]. 北京:科学出版社,2008:22-23.

(收稿日期:2013-02-28 编辑:徐 娟)