

基于投影寻踪的云南省农业水资源效率评价

殷欣, 刘小刚, 张彦, 耿宏焯, 杨启良, 杨具瑞

(昆明理工大学 现代农业工程学院, 云南 昆明 650500)

摘要: 为探讨云南省农业水资源利用效率, 选取与资源利用效率、经济效益、生态效益相关的 15 个指标, 将遗传算法与传统的优化方法结合, 利用投影寻踪技术的基本原理, 建立了农业水资源利用效率评价模型。利用该模型对云南省 16 个地区 2000、2005 及 2010 年的农业水资源利用效率方案进行综合评价, 优化农业水资源利用效率的状态变量指标的最佳投影方向, 确定各样本的投影值, 从而对各地区进行分类和评价。评价结果很好地反映了各评价指标的贡献率和方向性, 符合高原季节性干旱地区农业水资源利用的现状。

关键词: 遗传算法; 投影寻踪; 云南省; 农业水资源; 效率评价

文献标识码: B

文章编号: 1000-288X(2013)05-0271-05

中图分类号: TV213.9

Efficiency Evaluation of Agricultural Water Resources in Yunnan Province Based on Projection Pursuit Model

YIN Xin, LIU Xiao-gang, ZHANG Yan, GENG Hong-zhuo, YANG Qi-liang, YANG Ju-rui

(School of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract: To investigate the use efficiency of agricultural water resources in Yunnan Province, an evaluation model of agricultural water use efficiency was established by selecting the resource use efficiency, economic benefit, ecological benefit and other related 15 indexes, combining genetic algorithm with traditional optimization method, and using the principles of projection pursuit regression (PPR) technique. The model was used to comprehensively evaluate agricultural water scheme of 16 areas in Yunnan Province in 2000, 2005 and 2010, optimize the best projection direction of state variable index of the agricultural water resources use efficiency, and obtain each sample projection value, so as to classify and evaluate each area. The evaluation result best reflects the contribution and directivity of each index and accords with plateau seasonal drought region of the agricultural water resources status.

Keywords: genetic algorithm; projection pursuit; agricultural water resources; Yunnan Province; efficiency evaluation

云南省位于中国西南边陲, 雨量充沛, 河流众多, 水资源量虽然很大, 但在旱季则比较匮乏, 尤其是雨季开始前的 2—4 月是干旱最严重的时间, 季节性干旱呈频发与并发趋势。近年来, 随着社会经济的高速发展, 各行业用水量急剧增加, 云南省正面临着生态环境恶化和水资源短缺的严峻局面, 其中农业用水所占比重大是造成问题的根源之一。大量的农业用水引发了尖锐的水资源供需矛盾, 严重制约了经济社会的可持续发展。因此, 研究农业水资源利用效率方案的模型和方法具有重要意义。

农业水资源的利用效率主要是指农业用水的利用效率, 农业用水包括灌溉用水和林牧渔业用水两部分, 其中灌溉用水占农业用水比例的 90% 以上, 用水量主要与灌溉技术、灌溉面积和当年的降水量等因素有关。本文将农业水资源利用效率指标体系划分为 3 个层次, 第 1 层是目标层, 为农业水资源总体的利用效率水平; 第 2 层是准则层, 主要包括资源利用率、生态效益、经济效益 3 个方面; 第 3 层是指标层, 共选取了 15 项基础指标建立云南省农业水资源利用效率评价体系 (如表 1 所示)。

收稿日期: 2012-10-26

修回日期: 2012-12-16

资助项目: 国家自然科学基金项目“有限灌溉对云南干热河谷区小粒咖啡的水氮耦合效应”(51109102), “调亏灌溉条件下作物水氮高效利用机制和供水供氮模式”(50879073), “限量灌溉和施氮对小桐子产量和品质效应研究”(51009073); “十二五”国家科技支撑计划项目(2011AA100504); 云南省应用基础研究资助项目(2010ZC043); 云南省教育厅重点项目(2011Z035)

作者简介: 殷欣(1988—), 男(汉族), 湖北省广水县人, 硕士研究生, 主要从事计算机农业应用技术研究。E-mail: yinxin616@126.com。

通信作者: 刘小刚(1976—), 男(汉族), 甘肃省镇原县人, 博士, 副教授, 主要从事农业水土资源高效利用研究。E-mail: liuxiaogang666@qq.com。

表 1 云南省农业水资源利用效率评价体系

目标层	准则层	指标层	
		指标名称	指标表达式
农业水资源总体利用效率水平	资源利用效率	水资源开发利用率	供水量/总水资源量
		农业用水比例	农业用水量/总用水量
		农业水资源利用效率系数	粮食产量/农业用水量
		农业灌溉地均用水量	农业灌溉用水量/有效灌溉面积
		人均农业用水量	农业用水量/总人口
		人均水资源占有量	水资源总量/总人口
		地均水资源占有量	水资源总量/有效灌溉面积
		农田旱涝保收率	旱涝保收面积/总耕地面积
		有效灌溉率	有效灌溉面积/总耕地面积
		水资源重复利用率	重复利用水量总和/计算时段内供水量
	经济效益	农民人均纯收入	农民总收入/农民总人口
		每 1 m ³ 用水的农业生产总值	农业生产总值/农业用水量
		人均粮食产量	粮食总产量/总人口
	生态效益	生态用水比例	生态用水量/总用水量
		单位面积化肥施用量	总化肥施用量/总耕地面积

1 农业水资源利用效率方案评估的投影寻踪模型

1.1 PP 模型简介

投影寻踪(projection pursuit model, 简称 PP 模型)是一种可用于高维数据分析,既可以作探索性分析,又可以作确定性分析的方法。PP 方法的特点主要有:(1)能成功克服高维数据所带来的严重困难;(2)可以排除与数据结构和特征无关的或者关系很小的变量的干扰;(3)为使用一维统计方法解决高维问题开辟了途径;(4)与其他非参数方法一样可以用来解决某种非线性问题^[1]。

1.2 PPE 模型建模

(1)评价指标集及归一化处理。假设各个指标集 $\{x^*(i,j) | i=1,2,\dots,n; j=1,2,\dots,p\}$, n, p 分别为样本的个数和农业水资源利用效率评价的指标数。为消除量纲,对式子进行归一化处理:

对于越大越优的指标:

$$x(i,j) = \frac{x^*(i,j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad (1)$$

对于越小越优的指标:

$$x(i,j) = \frac{x_{\max}(j) - x^*(i,j)}{x_{\max}(j) - x_{\min}(j)} \quad (2)$$

式中: $x^*(i,j)$ ——第 i 个样本的第 j 个指标值; $x(i,j)$ ——指标特征值归一化后的序列; $x_{\max}(j)$, $x_{\min}(j)$ ——第 j 个指标值的最大值和最小值^[2-3]。

(2)构造投影指标函数 $Q(a)$ 。PPE 就是把 p 维数据 $\{x^*(i,j) | i=1,2,\dots,n; j=1,2,\dots,p\}$ 合成

以 $a = \{a(1), a(2), a(3), \dots, a(p)\}$ 为投影方向上的一维投影值 $z(i)$:

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

式中: $a(j)$ ——第 j 个指标值的单位长度向量; $x(i,j)$ ——指标特征值归一化后的序列; $z(i)$ ——第 i 个样本的投影值。投影值 $z(i)$ 的散布特征为:整体上投影点之间尽可能散开,而在局部投影点尽可能密集,最好凝聚成若干个点团。因此,投影指标函数可表达为:

$$Q(a) = S_z * D_z \quad (4)$$

式中: S_z ——投影值 $z(i)$ 的标准差; D_z ——投影值 $z(i)$ 的局部密度; $Q(a)$ ——投影指标函数,即:

$$S_z = \frac{\sqrt{\sum_{i=1}^n [z(i) - E(z)]^2}}{\sqrt{n-1}} \quad (5)$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^n [R - r(i,j)] * u[R - r(i,j)] \quad (6)$$

式中: $E(z)$ ——序列 $\{z(i) | i=1,2,\dots,n\}$ 的平均值; R ——局部密度的窗口半径,可以根据试验来确定,一般可取 0.1; $r(i,j)$ ——样本之间的距离, $r(i,j) = |z(i) - z(j)|$; $u(t)$ ——单位阶跃函数,当 $t < 0$, 其函数值为 0, 当 $t \geq 0$, 其值为 1。

(3)优化投影目标函数。当指标值的样本集合确定后,指标函数 $Q(a)$ 只随着 a 的变化而变化。最佳投影方向最大可能地反映了高维数据特征结构的投影方向,可通过求解指标函数最大化来确定最佳投影方向。

最大化目标函数:

$$\max: Q(a) = S_z * D_z \quad (7)$$

约束条件:

$$s. t. \sum_{j=1}^p a^2(j) = 1, 0 \leq a(j) \leq 1 \quad (8)$$

这是一个以 $\{a(j) | j=1, 2, \dots, p\}$ 为优化变量的非线性优化问题, 本文采用加速遗传算法 (RAGA) 来解决此类全局优化问题^[4]。

(4) 方案优选。把最佳投影方向 a^* 带入式 (3) 后即得各方案的投影值 $z^*(i)$ ^[5]。对各项评价指标的各级评价标准区间取右端点值, 生成评价标准样本, 进行归一化处理, 利用 RAGA—PPE 模型进行综合投影, 得到标准样本的投影值。设定评价的最低等级为 1, 最高等级为 N, 则可得到标准样本的投影散

点图, 根据各状态划分值及其对应的投影值 $z^*(i)$ 建立投影寻踪评价模型 $y=f(z)$ 。按等级值从小到大排序, 得到最大的等级值, 进而求出最优的方案^[6-7]。

2 实例分析

2.1 评价方案指标计算

以云南省 16 个地区为单元, 对 2000, 2005 和 2010 年的实际水资源利用效率方案进行评价, 限于篇幅, 以 2010 年云南省农业实际水资源利用效率方案为例, 对云南省 16 个地区单元的各评价指标值进行计算。

表 2 为 2010 年云南省农业实际水资源利用效率方案的各评价指标的原始值。

表 2 2010 年云南省农业实际水资源利用效率方案评价指标原始值

指标名称	昆明	曲靖	玉溪	保山	昭通	丽江	普洱	临沧
水资源开发利用率/%	44.86	13.88	26.22	6.32	7.19	8.63	3.88	7.79
农业用水比例/%	36.99	51.88	46.68	78.81	65.25	83.3	88.9	84.94
农业水资源利用效率指数/(kg·m ⁻³)	1.40	3.83	1.32	1.42	3.47	0.85	0.98	0.95
农业灌溉地均用水量/(m ³ ·hm ⁻²)	5 970	3 695	4 201	7 228	3 996	7 429	8 037	9 496
人均农业用水量/m ³	120	114	149	317	97	409	364	343
人均水资源占有量/m ³	724	1 577	1 216	6 354	2 071	5 690	10 542	5 185
地均水资源占有量/(m ³ ·m ⁻²)	3.6	5.13	3.43	14.51	8.52	10.34	23.30	14.35
农田旱涝保收率/%	20.87	11.49	23.17	21.54	7.87	21.96	14.70	10.86
有效灌溉率/%	30.76	18.19	33.21	29.17	18.31	41.78	27.01	20.46
水资源重复利用率/%	63	218	144	600	551	444	1 009	445
农民人均全年纯收入/元	5 810	4 130	5 747	3 627	2 768	3 410	3 456	3 279
用水的农业生产总值/(元·m ⁻³)	14.03	20.2	18.23	8.17	10.5	4.11	5.2	8.5
人均粮食产量/kg	168	435	196	449	337	348	355	326
生态用水比例/%	14.24	1.23	0.57	0.23	0.47	0.33	1.14	0.55
化肥施用量/(kg·hm ⁻²)	411	308	338	301	189	458	146	359
指标名称	楚雄	红河	文山	版纳	大理	德宏	怒江	迪庆
水资源开发利用率/%	22.74	10.64	5.31	9.47	13.88	4.88	0.67	1.07
农业用水比例/%	77.87	72.47	72.44	86.03	75.16	82.75	80.26	76.46
农业水资源利用效率指数/(kg·m ⁻³)	1.37	1.24	2.68	0.63	1.29	1.14	1.16	1.18
农业灌溉地均用水量/(m ³ ·hm ⁻²)	5 918	6 553	4 401	12 496	6 829	8 445	10 428	6 338
人均农业用水量/m ³	261	254	135	516	284	436	287	311
人均水资源占有量/m ³	1 475	3 287	3 515	6 330	2 719	10 795	53 313	37 971
地均水资源占有量/(m ³ ·m ⁻²)	3.34	8.5	11.43	15.33	6.55	20.89	193.9	77.49
农田旱涝保收率/%	22.52	16.59	8.99	22.43	20.7	13.96	10.99	20.68
有效灌溉率/%	33.33	31.4	17.3	37.89	37.23	25.88	15.34	34.26
水资源重复利用率/%	145	320	601	517	251	746	4184	3418
农民人均全年纯收入/元	3 896	3 922	2 806	4 354	3 902	3 368	2 005	3 347
用水的农业生产总值/(元·m ⁻³)	10.55	7.48	12.75	4.49	9.9	6.58	3.12	4.28
人均粮食产量/kg	357	314	362	327	367	497	332	366
生态用水比例/%	0.66	1.2	0.53	0.28	0.95	15.64	0.32	0.32
化肥施用量/(kg·hm ⁻²)	354	370	209	348	425	265	63	158

注:表中数据来源于云南省水资源公报和云南统计年鉴。

2.2 指标评价标准

在选取农业水资源利用效率方案健康评价指标的基础上,建立其评价标准,本文将农业水资源利用效率健康划分为病态、不健康、亚健康、较健康、很健康 5 个等级^[8-9]。指标标准值确定方面,以理论或者目前现实所能达到的最高或最低极限值为健康的级别标准,以本区及类似地区最小值为病态的限定值,在病态和很健康之间平均划分为 3 个等级,作为不健康、亚健康、较健康的标准。各标准的区间范围详见表 3。表 3 中, j_1-j_{15} 分别表示农业水资源利用效率的 15 个指标, Y_1-Y_5 分别表示病态、不健康、亚健康、较健康和很健康等 5 个标准。

表 3 云南省农业水资源健康指标评价标准值

指标	Y_1	Y_2	Y_3	Y_4	Y_5
j_1	<1	1~15	15~30	30~45	>45
j_2	>90	75~90	60~75	45~60	<45
j_3	<1	1~2	2~3	3~4	>4
j_4	<4 500	4 500~6 000	6 000~7 500	7 500~9 000	>9 000
j_5	<100	100~200	200~300	300~400	>400
j_6	<1 000	1 000~2 000	2 000~3 000	3 000~4 000	>4 000
j_7	<4.5	4.5~9	9~13.5	13.5~18	>18
j_8	<10	10~15	15~20	20~25	>25
j_9	<18	18~24	24~30	30~36	>36
j_{10}	<100	100~300	300~900	900~2 700	>2 700
j_{11}	<3 000	3 000~4 000	4 000~5 000	5 000~6 000	>6 000
j_{12}	<4	4~8	8~12	12~16	>16
j_{13}	<200	200~300	300~400	400~500	>500
j_{14}	<0.5	0.5~1	1~2	2~4	>4
j_{15}	>500	400~500	300~400	200~300	<200

依据步骤 4 对各项评价指标的各级评价标准区间取右端点值,确定病态的左端点值和很健康的右端点值,这里分别取病态的右端点值的 0.5 倍和很健康的左端点值的 3 倍,然后利用均匀随机数在每个等级区间内随机产生 5 个值,总共构成 30 个指标样本^[10]。采用 Matlab 语言,编写相应的目标函数和约束指标函数。应用 RAGA 进行综合投影,得出最佳投影方向向量为: $a = [0.965\ 6, 0.238\ 9, 0.163\ 9, 0.000\ 1, 0.029\ 8, 0.018\ 3, 0.014\ 2, 0.012\ 3, 0.006\ 9, 0.003\ 3, 0.004\ 4, 0.005\ 1, 0.018\ 7, 0.001\ 6, 0.002\ 1]$; 对应的投影值为: 2.522 4, 3.118 1, 2.799 2, 3.071 3, 3.685 1, 0.091 0, -0.054 0, 0.006 1, -0.026 1, 0.001 4, 0.063 3, -2.299 3, -0.444 3, -0.384 2, -0.416 4, -0.389 0, -0.327 0, -0.689 7, -0.834 7, -0.774 6, -0.806 8, -0.779 3, -0.717 4, -1.080 0, -1.225 1, -1.165 0, -1.197 1, -1.169 7, -1.107 8 和

-1.470 4。将病态、不健康,亚健康、较健康和很健康 5 个状态分别对应 1, 2, 3, 4 和 5 级,可用逻辑斯蒂曲线来描述 $z^*(i)$ 与 $y(i)$ 之间的关系:

$$y^*(i) = \frac{N}{1 + e^{c(1) - c(2)z^*(i)}} \quad (9)$$

式中: $y^*(i)$ ——第 i 个样本等级的计算值; N ——最高等级值; $c(1)$ ——积分常数; $c(2)$ ——增长率; $z^*(i)$ ——第 i 个样本投影值的计算值,可通过求解如下优化问题来确定:

$$\min F[c(1), c(2)] = \sum_{i=1}^n [y^*(i) - y(i)]^2 \quad (10)$$

其中 $N=5$, $c(1)$, $c(2)$ 通过 AGA 优化估计得到 $c(1) = -6.256\ 7$, $c(2) = -9.200\ 3$, 这样得到的云南省农业用水等级的 PPGE 模型为:

$$y^*(i) = \frac{5}{1 + e^{-6.2567 + 9.2003z^*(i)}} \quad (11)$$

2.3 评价结果分析

利用上述方法,对云南省 16 地区的 15 个农业水资源利用效率健康性指标进行综合投影,得到最佳投影方向 a , 计算得到: $a = [0.999\ 0, 0.034\ 9, 0.014\ 2, 0.005\ 4, 0.016\ 6, 0.012\ 4, 0.007\ 6, 0.004\ 2, 0.002\ 0, 0.001\ 0, 0.000\ 2, 0.001\ 0, 0.000\ 1, 0.004\ 0, 0.000\ 6]$, 从最佳投影向量可以看出,对农业水资源利用效率方案优劣贡献较大的指标依次是:水资源开发利用效率,农业用水比例,人均农业用水量,农业水资源利用效率指数,人均水资源占有量。认为水资源开发利用效率和农业用水比例对农业用水健康的贡献最大,水资源开发利用效率排在指标权重的第 1 位,农业用水比例排在第 2 位,表明水资源开发利用效率和农业用水比例是影响农业用水健康性的关键因子,应当适当对其进行调整。将投影方向带入式(3)可分别得到不同年限不同地区的投影值 $z(i)$, 详见表 4 所示。投影值 $z(i)$ 是 15 个评价指标的标准化值在最佳投影方向上的综合投影值,反映了云南省不同地区农业水资源利用效率的健康性^[11]。

根据投影值 $z(i)$ 带入投影寻踪等级评价模型 $y = f(z)$ 中,最终得到云南省农业水资源利用效率方案的健康指数^[12](图 1)。从表 4 可得到如下结论: (1) 该健康指数反映了云南省不同地区的农业水资源利用效率相对健康性,健康等级较高区域主要为怒江、普洱、迪庆、昭通、文山等,这些地区多位于云南省边区,人口偏少但水资源量充足,光热资源丰富,属于农业水资源利用效率很健康的地区。(2) 根据 2000, 2005 及 2010 年的云南省水资源利用效率综合分析可知,农业用水健康指数随时间整体呈下滑态势,病态

及不健康地区数量及面积有所增加,由中部地区蔓延至西南部地区,由 5 个地区增大为 8 个。(3) 昆明、玉溪地区的水资源健康指数持续偏低,丽江、大理及版

纳的水资源健康呈下滑趋势,今后可适当增加这些地区的人均农业用水量及水资源占有量、水资源重复利用率,才能保证云南省农业水资源的健康。

表 4 云南省农业水资源利用效率方案的投影值及其健康指数

方案	昆明	曲靖	玉溪	保山	昭通	丽江	普洱	临沧	楚雄	红河	文山	版纳	大理	德宏	怒江	迪庆	
投影值	2000	1.00	0.98	1.02	0.62	0.28	0.55	0.05	0.25	0.72	0.66	0.51	0.12	0.73	0.43	0.01	0.16
	2005	1.04	0.71	1.06	0.67	0.37	1.00	0.19	0.47	0.94	0.83	0.44	0.49	0.93	0.72	0.01	0.10
	2010	0.96	0.68	0.76	0.67	0.35	1.10	0.23	0.84	0.81	0.85	0.41	0.79	1.00	0.56	0.01	0.26
健康指数	2000	0.30	0.30	0.20	3.20	4.90	3.80	4.90	4.90	2.00	2.70	4.10	4.90	1.90	4.50	4.90	4.90
	2005	0.20	2.20	0.10	2.60	4.70	0.30	4.90	4.40	0.40	1.00	4.50	4.30	0.50	2.00	4.90	4.80
	2010	0.30	2.50	1.60	2.70	4.70	0.10	4.90	1.00	1.20	0.80	4.60	1.30	0.20	3.70	4.90	4.80

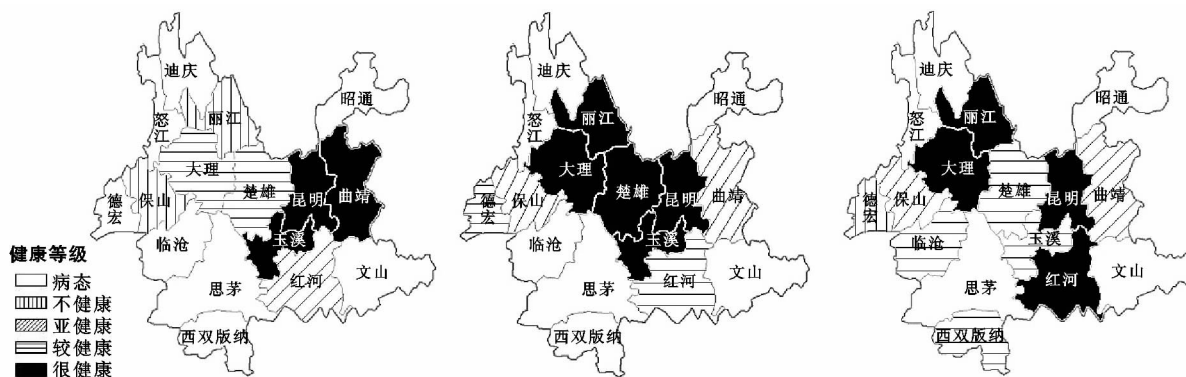


图 1 2000,2005 及 2010 年云南省农业水资源利用效率方案健康等级

3 结论

从资源利用效率、经济效益及生态效益等 3 个方面入手,对云南省农业水资源利用效率健康方案做了较为系统的研究,建立了农业水资源利用效率方案健康性评价遗传投影寻踪模型。对 2000,2005 及 2010 年云南省 16 个地区的农业水资源利用效率方案健康性进行了评价,并按农业水资源利用效率方案健康性高低划分为 5 级,将多维评价指标值组成的方案集综合成一维投影指标值,且综合考虑了指标的客观性和决策者的主观性。评价结果精确合理,在云南省农业水资源利用效率评价中具有较强的适应性,可供其他区域农业水资源配置方案评价参考。

[参 考 文 献]

[1] 付强,金菊良,梁川. 基于实码加速遗传算法的投影寻踪分类模型在水稻灌溉制度优化中的应用[J]. 水利学报, 2002(10):39-45.

[2] 金菊良,魏一鸣,丁晶. 水质综合评价的投影寻踪[J]. 环境科学学报, 2001,21(4):431-434.

[3] 孙琦伟,吴普特,王玉宝,等. 投影寻踪分类模型在黑河流域农业用水方案健康性评价中的应用[J]. 干旱地区

农业研究,2012,30(1):175-183.

[4] 金菊良,丁晶. 遗传算法及其在水科学中的应用[M]. 成都:四川大学出版社,2000.

[5] 张礼兵,程吉林,金菊良,等. 农业灌溉水质评价的投影寻踪模型[J]. 农业工程学报, 2006,22(4):15-18.

[6] 吴泽宁,左其亭,丁大发,等. 黄河流域水资源调控方案评价与优选模型[J]. 水科学进展, 2005,16(5):735-740.

[7] 王顺久,侯玉,张欣莉. 流域水资源承载能力的综合评价方法[J]. 水利学报,2003(1):88-92.

[8] 金菊良,周玉良,魏一鸣. 基于遗传程序设计的水质评价模型[J]. 水电能源科学,2004,22(2):1-5.

[9] 刘渝,杜江,张俊彪. 湖北省农业水资源利用效率评价[J]. 中国人口资源与环境,2007,17(6):60-65.

[10] Fu Qiang. Study on the PPE model based on RAGA to classify the county energy [J]. Journal of Systems Science and Information, 2004,2(1):73-82.

[11] Zhang Chi, Dong Sihui. A new water quality assessment model based on projection pursuit technique[J]. Environmental Science, 2009,21(S):154-157.

[12] Fu Qiang, Fu Hong. Applying PPE model based on RAGA in the investment decision marketing of water saving irrigation project[J]. Nature and Science, 2001 (1):57-61.