

# 投影寻踪技术在松嫩平原西部水资源 可持续利用中的应用

卞建民, 张 芳

(吉林大学 环境与资源学院, 长春 吉林 130026)

**摘 要:** 根据松嫩平原西部水资源复合系统的特征和水资源条件, 建立了区域水资源可持续利用评价指标体系。应用投影寻踪技术, 采用实码遗传算法优化投影指标函数, 将方案多维评价指标值综合为一维投影值, 根据投影值的大小实现方案的优选。研究表明, 投影寻踪技术可以有效地解决权重的人为干扰, 通过优化排序实现对评价结果的定量化表征。

**关键词:** 松嫩平原西部; 水资源可持续利用; 投影寻踪; 遗传算法

文献标识码: B

文章编号: 1000—288X(2007)03—0093—04

中图分类号: TV 213. 9

## Application of Projection Pursuit Model in Sustainable Water Resource Utilization in the Western Songnen Plain

BIAN Jian-min, ZHANG Fang

(College of Environment and Resources, Jilin University, Changchun, Jilin 130026, China)

**Abstract:** In terms of the characteristics of water resource compound system and the conditions of water resources, an evaluation system of sustainable water resource utilization has been established according to the principles of sustainability. Projection pursuit model (PPM) and real coding based genetic algorithm (RGA) have been used in the evaluation, which can synthesize the multiple factors into a single pursuit value and attain an optimal scheme. The result shows that it is effective to solve the weighted human interference, and fulfill the evaluation of sustainable water resource utilization by an optimized sequence.

**Keywords:** Western Songnen Plain; sustainable water resource utilization; projection pursuit model (PPM); real coding based genetic algorithm(RGA)

水资源可持续利用是既要满足当代人的水需求, 又不危及子孙后代的水需求, 并满足其需求能力的变化和发展, 是社会、经济可持续利用的基础条件。近年来, 随着对水资源需求的不断增长, 有关水资源可持续利用的概念、评价指标和模型应用成为水科学和环境科学研究的热点问题, 很多学者对水资源可持续概念、评价指标及评价方法进行了广泛的探讨, 提出了综合评判法、灰色聚类法、层次分析法、模糊模式识别模型、人工神经网络模型和模糊物元模型等<sup>[1-7]</sup>。上述模型通过对高维数据(多个评价指标值)的处理, 结合专家给出的权重矩阵, 实现了在低维子空间完成降维评价的过程。但由专家凭经验给出的权重矩阵存在较大的主观性和人为干扰因素, 直接影响评价结果, 而且权重是否属于各项指标(多维)在低维子空间的最佳投影, 也无法确定。基于以上问题的存在, 本文将投影寻踪方法(PP)引入松嫩平原

西部水资源可持续利用研究中, 通过该模型的应用实现了该区水资源可持续利用评价。

## 1 基于实码遗传算法的投影寻踪技术

### 1.1 投影寻踪模型

简称 PP 模型(projection pursuit model)是一种新兴的适应于高维数据处理的探索性数据分析方法<sup>[8]</sup>, 通过高维数据经过某种组合投影到低维子空间上, 对投影的构形, 采用投影指标函数来衡量投影暴露某种结构的可能性大小, 寻找出使投影指标函数达到最优(即能反映高维数据结构或特征)的投影值, 然后根据该投影值来分析高维数据的结构特征。其中, 投影指标函数的构造及其优化问题是应用 PP 方法能否成功的关键<sup>[9-10]</sup>。

其应用步骤包括对评价指标的归一化处理、线性投影、构造投影指标函数和优化投影指标函数。其中

优化投影指标函数要找出能将数据清晰地散布为有意义结构的投影方向<sup>[11]</sup>, 可通过求解投影指标函数最大化问题获得。

$$\max Q(a) = S_z D_z, s. t. \sum_{j=1}^p a^2(j) = 1$$

这是一个以  $\{a(j) | j = 1 \sim p\}$  为优化变量的复杂非线性优化问题, 用常规优化方法处理较困难。模拟生物优胜劣汰规则与群体内部染色体信息交换机制的遗传算法(RGA) 是一种通用的全局优化方法, 用它来求解上述问题十分简便和有效。

## 1.2 基于实数编码的遗传算法

遗传算法(genetic algorithm, 简称 GA) 是基于生物进化过程中优胜劣汰规则与群体内部染色体信息交换机制的、处理复杂优化问题的一类通用性强的新方法。GA 利用简单的编码技术和算法机制来模拟复杂的优化过程, 只要求优化问题是可计算的, 而对目标函数和约束条件的具体形式、优化变量的类型和数目不作限制, 在搜索空间中进行自适应全局并行搜索, 运行过程简单, 计算结果丰富, 适合于处理常规优化方法棘手的复杂优化问题<sup>[12]</sup>。

国内外常用的 GA 实施方案使用二进制编码, 需要频繁的编码和解码, 计算量大, 而且只能产生有限的离散点阵, 还可能产生额外最优点。为此, 我们采用基于实数编码的遗传算法(Real coding based Genetic Algorithm, 简称 RGA)<sup>[12]</sup>。算法流程见图 1 所示, 本文采用 Visual C++ 编程实现。

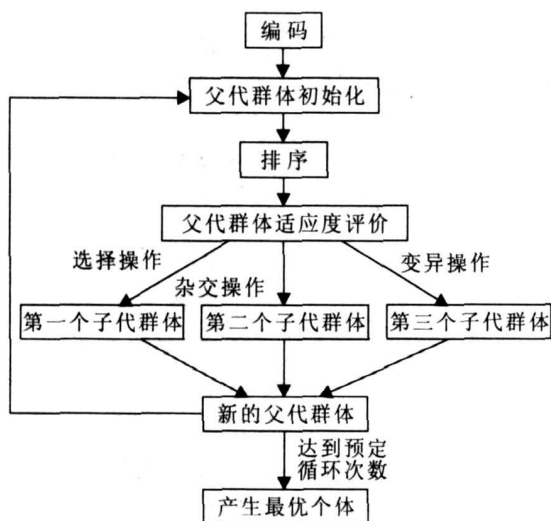


图 1 基于实数编码的遗传算法流程图

应用实码遗传算法优化投影指标函数, 该方法将求解问题转化为以下最大化问题:

$$\begin{cases} \max Q[a(1), a(2), \dots, a(p)] \\ \sum_{j=1}^p a(i) = 1 \end{cases}$$

## 2 松嫩平原西部水资源可持续利用评价

### 2.1 松嫩平原西部水资源环境背景

松嫩平原西部地处我国北方半干旱半湿润大陆性季风气候区, 冬季低温干燥, 降水量稀少; 夏季高温湿润, 降水量较多。多年平均降水量为 300~500 mm, 多年平均蒸发量为 1 500~1 900 mm, 是多年平均降水量的 3.50~4.75 倍左右。主要河流有松花江、嫩江、洮儿河、霍林河及季节性的地表径流, 多以客水流经该区, 利用程度较低。地下水是该区主要的水源, 地下水天然资源量为  $3.42 \times 10^9 \text{ m}^3$ , 最大开采量为  $2.51 \times 10^9 \text{ m}^3$ , 占地下水天然资源量的 73.46%。

该区农田灌溉用水量最大, 占总用水量的 60%; 林牧渔业的地表水用水量最大, 占地表水实际用水量的 53.46%。工业不发达, 工业用水量相对较少<sup>[13]</sup>。长期以来, 由于对水资源开发利用无序, 导致该区水资源严重不足, 直接影响了农牧业生产。据资料分析, 白城市因缺水造成的受灾面积, 1996 年为  $4.75 \times 10^5 \text{ hm}^2$ ; 1997 年为  $5.37 \times 10^5 \text{ hm}^2$ ; 1998 年为  $4.24 \times 10^5 \text{ hm}^2$ ; 1999 年为  $5.20 \times 10^5 \text{ hm}^2$ 。共计减产粮食近  $3.20 \times 10^9 \text{ kg}$ , 相当于该地 2 a 的粮食产量<sup>[14]</sup>。

近 20 a 来, 由于气候日趋干旱化, 更加重了水资源的紧张的局面。1979—1999 年间, 松嫩平原西部地区地下水开采量由  $3.36 \times 10^8 \text{ m}^3$  (1979), 占可开采资源量的 13.38%, 增加到  $1.20 \times 10^9 \text{ m}^3$  (1999), 占可开采资源量的 47.78%, 水位下降了 9.4 m, 现有机井 20% 无水或间歇有水。若不及时采取有效的措施, 按照这样的发展态势, 预计到 2010 年将会有 70% 的井无水可抽。

### 2.2 评价指标体系的建立

目前, 关于水资源可持续利用评价指标体系已有了很多研究<sup>[15-17]</sup>。本文在借鉴已有研究成果的基础上<sup>[4-6, 18]</sup>, 根据该区水资源环境背景和社会经济发展, 遵循科学性、系统性与层次性相结合及可操作性的原则, 建立了水资源可持续利用评价指标体系, 选取对水资源可持续利用有较大影响、便于收集和计算, 同时有利于指标的纵向和横向比较。

水资源可持续利用主要受到水环境背景、社会经济条件和水资源工程的开发利用程度的影响, 如水资源总量、可开采地下水量、供需水量模数、水质、人口、工业需水量、农业需水量、水资源开发利用水平、水资源管理水平等。对水资源可持续利用的评价要求不仅要考虑水环境背景及其开发利用现状, 还要考虑它能否适应将来的自然生态和社会经济发展水平。只有当水资源开发利用状况满足现在和将来较长时期

内的需求,并未引起明显的或持久性的生态环境问题时,这种水资源利用才是可持续的<sup>[17]</sup>。

综合以上分析,本文将水资源可持续利用综合评价体系划分为水资源条件、水资源利用率、水资源管理及技术设施和水资源综合效益4个子系统,这4个方面基本反映了水资源可持续利用的现状与程度。水资源条件包括产水系数、产水模数、人均水量和地均水量。其中,产水系数主要反映气候条件对水资源本身的影响,产水模数则反映水资源的赋存条件,人均水量反映人口对于水资源的压力,地均水量反映单位面积水资源占有量,它们共同体现水资源的自然状况。水资源利用率包括工业用水利用率和农业用水利用率。水资源利用效率直接影响水资源的可持续利用。水资源管理水平包括水资源开采率、灌溉率和开采强度3项,用每1 km<sup>2</sup>含井数来近似表示水资源的开采强度。水资源综合效益主要用万元工农业产值需水量来表示。

### 2.3 模型应用及结果分析

将上述指标体系的因子值导入计算程序中运行,完成对指标的归一化处理,该步骤需要对指标进行正负作用方向的判定,对越大越优型指标按下式进行归一化处理:

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

对越小越优型指标按下式进行归一化处理:

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

式中:  $x^*(i, j)$  —— 第  $i$  个地区第  $j$  个评价指标;  $x_{\min}(j)$ ,  $x_{\max}(j)$  —— 分别为评价集中第  $j$  个评价指标的最小值和最大值。

以  $a = [a(1), a(2), \dots, a(p)]$  为投影方向将进行线性投影,得到一维投影值  $z(i)$ :

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

其中,  $a = [a(1), a(2), \dots, a(p)]$  为待定向量,通过构造投影指标函数并对其优化,选出最佳投影方向  $a$ ,求得  $z(i)$  的值,按值的大小进行评价。

为求得最佳投影方向  $a$ ,根据投影值  $z(i)$  的散布特征,构造投影指标函数为:

$$Q(a) = S_z D_z$$

式中:  $S_z$  —— 投影值的标准差;  $D_z$  —— 投影值  $z(i)$  的局部密度,即

$$S_z = \left\{ \sum_{i=1}^n \frac{[z(i) - E_z]^2}{n-1} \right\}^{0.5}$$

$$D_z = \sum_{i=1}^n \sum_{j=1}^n (R - r_{ij}) u(R - r_{ij})$$

式中:  $E_z$  —— 系统  $\{z(i) | i = 1 \sim n\}$  的均值;  $R$  —— 求局部密度的窗口半径; 距离  $r_{ij} = |z(i) - z(j)|$ ;  $u(t)$  —— 单位阶跃函数,当  $t \geq 0$  时其函数值为1,当  $t < 0$  时其函数值为0。

投影指标函数随着投影方向  $a$  的变化而变化,不同投影方向反映不同的数据结构特征。通过求解投影指标函数最大化问题可以获得能够将数据清晰地散布为有意义结构的投影方向,即最佳投影方向<sup>[10]</sup>。

$$\max Q(a) = S_z D_z, s. t. \sum_{j=1}^p a^2(j) = 1$$

这是一个以  $\{a(j) | j = 1 \sim p\}$  为优化变量的复杂非线性优化问题,用常规优化方法处理较困难。模拟生物优胜劣汰规则与群体内部染色体信息交换机制的遗传算法(RGA)是一种通用的全局优化方法,用它来求解上述问题十分简便和有效。部分计算代码如下:

/\* 生成第1个子代群体 \*/

for( $k = 1; k \leq N - 5; k++$ )

/\*  $N$  为群体规模 \*/

{  $u[k] = \text{rand}() / 32767.0$ ;

/\* 生成  $n - 5$  个随机数 \*/

$m = 0$ ;

if( $u[k] = 0$ )

for( $j = 1; j \leq M; j++$ )

$y1[k][j] = y[1][j]$ ; /\*  $y$  为父代群体,  $y1$

为第1个子代群体 \*/

else for( $i = 1; i \leq N; i++$ ) /\* 若  $u[k]$  在

( $p[i-1], p[i]$ ) 中,则第  $i$  个个体被选中 \*/

{ if( $(u[k] > p[i-1]) \& \& (u[k] \leq p[i])$ )

{ for( $j = 1; j \leq M; j++$ )

$y1[k][j] = y[i][j]$ ;

$m = 1$ ;

}

if( $m = 1$ ) break; }

}

for( $i = 1; i \leq 5; i++$ ) /\* 移民操作,将最优秀的5个父代个体直接进入子代群体中 \*/

for( $j = 1; j \leq M; j++$ )

$y1[N-5+i][j] = y[i][j]$ ;

最后将求得的最佳投影方向  $a$  代入式子  $z(i) =$

$\sum_{j=1}^p a(j) x(i, j)$  即得各投影值  $z(i)$ 。 $z(i)$  值越大对应的区域可持续利用程度越高。本文所选取2000年的实测数据进行了松嫩平原西部水资源可持续利用评价,结果见表1。

表 1 松嫩平原西部水资源可持续利用评价结果

县市	宁江	扶余	前郭	长岭	乾安	洮北	洮南	大安	通榆	镇赉
$\alpha$	0.342 4	0.413 7	0.176 5	0.546 1	0.218 7	0.161 6	0.264 1	0.469 6	0.094 5	0.095 4
$Z$	0.805 4	0.253 1	0.470 4	0.244 0	0.380 3	2.538 0	1.225 6	0.798 4	0.807 0	0.806 3
排序	5	9	7	10	8	1	2	6	3	4
$Z$	0.865 1	0.322 1	0.524 7	0.320 2	0.324 4	2.477 2	1.364 7	1.232 0	0.859 6	0.865 6
排序	5	9	7	10	8	1	2	3	6	4
$Z$	0.842 7	0.340 1	0.525 9	0.306 0	0.340 4	2.550 3	1.347 6	1.211 0	0.843 0	0.842 8
排序	6	9	7	10	8	1	2	3	4	5

注： $\alpha$  为最佳投影方向； $Z$  为投影值。

根据排序结果将该区水资源可持续利用程度划分为可持续、较持续、不可持续、极不可持续 4 个级别。其中 $< 0.6$  为极不可持续,  $0.6\sim 1.0$  为不可持续,  $1.0\sim 2.0$  为较可持续,  $> 2.0$  为可持续。

洮北区水资源条件较好(产水系数、产水模数、地均水量较高), 水资源管理水平较高(水资源开采程度、灌溉率及水资源技术设施最高), 单位水资源工业产值高, 其评价结果为可持续级。大安和洮南水资源条件相对较好, 为比较可持续级。宁江、镇赉、通榆处于不可持续利用状态。扶余、长岭、前郭和乾安水资源利用十分不合理, 为极不可持续级。通榆、镇赉由于其水资源条件较差(产水系数、产水模数最低), 水资源开采程度较低, 并且利用率不高, 尤应注意的是该区水资源管理及技术设施较低, 耗水量很大, 使得水资源综合效益不高。所以该区应发挥部分地区水资源开采潜力较大的优势, 加大水利设施建设, 发展节水农业, 提高水资源的综合效益。

洮南、大安水资源条件相对较好, 但其工业用水利用效率较低, 水资源开采程度较高, 水资源综合效益一般。这说明水资源利用不合理, 存在水资源浪费的问题。因此, 上述地区应着重于提高水资源的利用率, 发展节水农业。

3 结论

应用本文所建立的水资源可持续利用评价指标体系, 将投影寻踪技术应用于水资源可持续利用评价, 获得了该区水资源可持续利用程度的评价结果。该结果表明, 区域水资源可持续利用是水资源复合系统综合指征的表现, 常规评价模型需要对评价因子给定基准, 目前国内外有关水资源可持续利用评价尚无统一的标准可以参照, 基于实数编码的遗传算法, 通过优化投影指标函数, 把方案多维评价指标值综合为一维投影值, 根据投影值的大小实现方案的优选, 有效解决权重的人为干扰, 通过优化排序实现对评价结

果的量化研究, 为区域水资源合理开发利用方案的制定提供了科学依据。另外在模型应用方面, 利用 PP 模型可以避免权重矩阵的人为干扰, 通过构造投影指标函数, 并对其优化, 选取最佳投影方向, 把方案多维评价指标值综合成一维投影值。投影值越大表示该方案越优。根据投影值的大小就可对方案集进行优选, 较其它方法具有明显的优势。

[ 参 考 文 献 ]

[ 1 ] 姚荣, 唐德善, 张娜. 区域水资源可持续发展评价模型的建立与应用[ J ]. 水电能源科学, 2004, 22( 1 ): 16—18.

[ 2 ] 肖长来, 林学钰, 梁秀娟, 等. 吉林省西部水资源可持续开发利用研究[ J ]. 长春科技大学学报, 2001, 21( 4 ): 354—358.

[ 3 ] 王锦国, 周志芳, 金忠青. 我国北方水资源持续发展状况的灰色聚类评价[ J ]. 水电能源科学, 2000, 18( 9 ): 10—12.

[ 4 ] 金菊良, 张礼兵, 魏一鸣. 水资源可持续利用评价的改进层次分析法[ J ]. 水科学进展, 2004, 15( 2 ): 227—232.

[ 5 ] 盖美, 田成诗. 大连市水资源可持续利用的模糊模式识别及对策研究[ J ]. 资源科学, 2003, 25( 2 ): 44—50.

[ 6 ] 宋松柏, 蔡焕杰. 区域水资源可持续利用评价的人工神经网络模型[ J ]. 农业工程学报, 2004, 20( 6 ): 89—92.

[ 7 ] 潘峰, 梁川, 王志良, 等. 模糊物元模型在区域水资源可持续利用综合评价中的应用[ J ]. 水科学进展, 2003, 14( 3 ): 271—275.

[ 8 ] Friedman J H, Turkey J W. A projection pursuit algorithm for exploratory data analysis[ J ]. IEEE Trans On Computer, 1974, 23( 9 ): 881—890.

[ 9 ] 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. 233—253.

[ 10 ] 金菊良, 刘永芳, 丁晶, 等. 投影寻踪模型在水资源工程方案优选中的应用[ J ]. 系统工程理论方法应用, 2004, 13( 1 ): 81—84.

[ 11 ] 张欣莉. 投影寻踪及其在水文水资源中的应用[ D ]. 四川大学博士学位论文, 2000.

4 3 种方法结果对比分析

为了检验 3 种分类方法的正确性,参考 2000 年的徐州市土地利用现状图,对假彩色合成的图像进行判读,用人工数字化的方法得到徐州市的水体面积为 9.734。通常情况下屏幕数字化的精度要明显高于计算机直接分类的精度,故把它作为对其它方法进行精度评判的依据(如表 2)<sup>[4]</sup>。

表 2 不同方法提取的水体面积及其误差大小

项 目	参考数据	监督分类法	综合阈值法
像元数	—	12 830	12 196
面积/ km <sup>2</sup>	9.73	11.54	10.97
绝对误差/ km <sup>2</sup>	0.00	1.80	1.23
相对误差/ %	0.00	18.50	12.60

通过上表,可以看到用综合阈值法能够有效地去除监督分类中多提取的水体,由于非监督分类在提取水体信息方面存在大量的误提现象,因而,在此不对其进行比较。

总之,综合阈值法应用于城市土地利用分类,能有效降低混合像元带来的影响,提高土地利用分类的精度,也将在利用遥感图像进行土地利用分类上产生重要的影响<sup>[2]</sup>。

5 结 论

用综合阈值法对遥感图像分类是一种比较理想的方法。综合各种文献资料,可以知道单纯依靠计算机自动分类来实现土地利用的分类其精度较低,为了获得较高的精度,有时不得不通过目视判读或直接手工勾划的方法。在计算机自动分类中,由于非监督分类方法无法有效地区分大多数地物,因此它是一种不可取的方法<sup>[2]</sup>。监督分类的客观性比较差,虽然精度可能较高,但不可能保证每一次都达到。而基于 NDVI, NDBI 等综合阈值法提取法,一旦波段和阈值确定,则它们的分类结果就是惟一的。

[参 考 文 献]

[ 1 ] 孙家柄,舒宁,关泽群. 遥感原理方法与应用[ M ]. 北京: 测绘出版社,1997. 324—333.

[ 2 ] 查勇,倪绍祥,杨山. 一种利用 TM 图像自动提取城镇用地信息的有效方法[ J ]. 遥感学报,2003( 1 ): 38—39.

[ 3 ] 杜培军,胡召玲,郭达志,等. 工矿区陆面演变监测分析与调控治理研究[ M ]. 北京: 地质出版社,2005. 52—54.

[ 4 ] 陈龙乾,郭达志,盛业华,等. 徐州市城区土地利用变化的卫星遥感动态监测[ J ]. 中国矿业大学学报,2004(9): 528—532.

( 上接第 96 页 )

[ 12 ] 金菊良,杨晓华,丁晶. 基于实数编码的加速遗传算法[ J ]. 四川大学学报( 工程科学版 ),2000,32(4): 20—24.

[ 13 ] 王天阳. 吉林西部水资源优化配置研究[ D ]. 吉林大学硕士学位论文,2003.

[ 14 ] 肖焕玉,赵国军. 吉林省西部干旱、半干旱生态区农业节水灌溉模式[ J ]. 吉林水利,2002( 11 ): 1—3.

[ 15 ] 邹积君,刘志文,张小芬,等. 区域水资源可持续利用指标体系的设计及评价方法研究[ J ]. 干旱区资源与环境,2003,17( 1 ): 37—40.

[ 16 ] 徐良芳,冯国章,刘俊民. 区域水资源可持续利用及其评价指标体系研究[ J ]. 西北农林科技大学学报( 自然科学版 ),2002,30(2): 119—122.

[ 17 ] 刘恒,耿雷华,陈晓燕. 区域水资源可持续利用评价指标体系的建立[ J ]. 水科学进展,2003,14( 3 ): 265—270.

[ 18 ] 卞建民,杨建强. 水资源可持续利用评价的指标体系研究[ J ]. 水土保持通报,2000,20(4): 43—45.