

# 区域水资源潜力综合评价的遗传投影寻踪方法\*

杨晓华<sup>1</sup> 杨志峰<sup>1,2</sup> 郦建强<sup>3</sup>

1. 北京师范大学环境科学研究所, 水环境模拟国家重点实验室, 北京 100875;

2. 北京师范大学水科学研究所, 水沙科学教育部重点实验室, 北京 100875;

3. 水利部水利水电规划设计总院, 北京 100011

**摘要** 为解决单项水资源潜力评价指标评价结果的不相容问题, 提高水资源潜力综合评价各层次的分辨力和评价模型的精度, 采用大样本数据, 利用投影寻踪、遗传算法、阶梯型曲线和水资源潜力评价标准, 为水资源潜力综合评价提出了一种新的方法——遗传投影寻踪方法(GPPM). 文中给出了实施 GPPM 的详细步骤, 作为示例, 对黄河中游关中地区水资源潜力进行了综合评价. 结果表明, GPPM 的评价结果可靠、直观、适用性强, 不仅可以对水资源潜力进行综合评价, 同时可以科学地确定评价指标的权重和对评价指标的合理性进行检验. GPPM 可广泛应用于各种水资源的综合评价.

**关键词** 水资源潜力 综合评价 投影寻踪 遗传算法 精度

黄河以其占全国河川径流 2% 的有限水资源, 承担着占全国 15% 耕地面积和 12% 人口及 50 多座大中城市的供水任务, 同时还要向流域外部分地区远距离调水; 黄河流域关中地区人多、地多而水少, 水土资源不相匹配. 随着我国西部大开发战略的实施, 关中地区乃至整个黄河流域水资源的短缺还将趋于严重<sup>[1,2]</sup>. 为了以水资源的可持续利用保障经济社会的可持续发展, 水资源的开发必须考虑其潜力.

水资源潜力<sup>[3]</sup>是指一个地区在可以预见的期间, 以水资源开发利用不引起生态环境退化为前提, 可以开发利用的潜在水资源量. 所谓水资源潜力综合评价就是根据水资源潜力评价指标, 通过所建立的数学模型, 对一个地区的水资源潜力进行评价, 为水资源可持续利用提供科学依据和决策支持. 由于实际水资源潜力各项指标的评价结果常常是不相容的, 直接利用水资源潜力评价标准表进行水资源潜力等级评判缺乏实用性. 模糊综合评判方法<sup>[3,4]</sup>、层次分析法<sup>[5]</sup>、灰色关联法<sup>[6]</sup>、多目标决策-理想点法<sup>[7,8]</sup>和人工神经网络法<sup>[9]</sup>等也难以从各层次考察水资源的潜力. 20 世纪 70 年代, Fried-

man<sup>[10]</sup>给出了多元数据分析的投影寻踪 (projection pursuit, PP) 方法, 该算法能在一定程度上解决多指标样本分类等非线性问题. 最近, PP 方法已被用来对水质问题进行评价<sup>[11,12]</sup>, 但由于只采用 5 个样点, 所建立的数学模型不够准确<sup>[13]</sup>, 对于多元数据这种复杂的拓扑结构, 一般也很难找到最优的投影方向, 目前现代优化算法<sup>[14]</sup>中的遗传算法, 由于具有较强的全局搜索能力、信息处理的隐并行性、鲁棒性和可规模化等优点<sup>[15~17]</sup>, 是处理这类优化问题的较好的方法. 基于上述原因, 本文利用 PP 方法把多指标转化为单一的投影指标, 采用大量的样本, 利用全局收敛的 Gray 码加速遗传算法<sup>[18]</sup>来优化投影方向, 并根据最佳投影值与其对应等级之间所呈现的关系, 建立新的数学模型, 提出了一种适合于对水资源潜力进行综合评价的新方法——遗传投影寻踪方法 (genetic projection pursuit method, GPPM), 并对黄河中游关中地区的水资源潜力进行了综合评价.

## 1 GPPM 的计算技术

PP 是处理高维数据, 尤其是非正态数据的一

2002-11-20 收稿, 2003-01-13 收修改稿

\* 国家重点基础研究发展规划资助项目(批准号: 1999043605)

E-mail: y1x1h1@sohu.com

类新兴统计方法,它的基本方法是将高维数据投影到低维子空间上,使用者通过观察投影图像,采用投影目标函数来衡量投影揭示数据某种结构的可能性大小.其中投影目标函数的构造、优化方法的选择、数学模型的建立是应用PP解决实际问题的关键.传统的PP计算技术复杂,精度难以控制,在一定程度上限制了PP方法的深入研究和广泛应用.为此,本文提出一套基于Gray码加速遗传算法的PP建模方法用于水资源潜力的综合评价.GPPM的计算包括以下5步:

**步1** 建立投影数据.根据水资源潜力评价标准产生用于水资源潜力评价的原始数据,它包括水资源潜力指标  $x^*(i, j)$  及对应的评价分值  $y(i)$ ,  $i=1, \dots, n; j=1, \dots, np$ . 其中  $n, np$  分别为样品的个数和指标数.对  $x^*(i, j)$  进行归一化处理为  $x(i, j)$ , 即

$$x(i, j) = (x^*(i, j) - x_{\min}(j)) / (x_{\max}(j) - x_{\min}(j)), \quad i = 1, \dots, n; j = 1, \dots, np, \quad (1)$$

其中  $x_{\min}(j), x_{\max}(j)$  分别为第  $j$  个指标的最小值和最大值.

**步2** 计算投影值.设  $a = (a(1), a(2), \dots, a(np))$  为投影方向,PP方法就是把  $x(i, j)$  投影到  $a$  上,得到一维投影值  $z(i)$

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

**步3** 建立投影目标函数.在综合投影值时,要求投影值  $z(i)$  应尽可能大地提取  $x(i, j)$  中的变异信息,即  $z(i)$  的标准差  $S_Z$  尽可能大,同时要求  $z(i)$  与  $y(i)$  的相关系数  $R_{ZY}$  的绝对值  $|R_{ZY}|$  尽可能大<sup>[12]</sup>,本文取投影目标函数  $f(a)$  为

$$f(a) = S_Z |R_{ZY}|. \quad (3)$$

**步4** 用Gray码加速遗传算法来优化投影方向.投影目标函数  $f(a)$  随着投影方向  $a$  变化而变化,可通过求解投影目标函数最大值来估计最佳投影方向  $a$ , 即

$$\begin{aligned} \max f(a) &= S_Z |R_{ZY}|, \quad (4) \\ \text{s. t. } \sum_{j=1}^{np} a(j)^2 &= 1, -1.0 \leq a(j) \leq 1.0. \end{aligned}$$

这是一个以  $a = (a(1), a(2), \dots, a(np))$  为优化变量的非线性优化问题,用常规的方法处理较困难,为了增强GPPM的实际应用能力和计算效率,本文用目前较好的Gray码加速遗传算法给出最佳投影方向.其基本思想是,在单位超球面中随机抽取若干个初始投影方向作为初始群体,用Gray码加速遗传算法来优化投影目标函数,并建立与投影目标函数的大小相应的适应度函数,适应度大的个体被保留的机会大,通过选择、杂交、变异等遗传操作,最后最大适应度函数值对应的个体与最大投影目标函数对应的最佳投影方向  $a^*$  相对应.

**步5** 建立水资源潜力综合评价数学模型.由步4求得的最佳投影方向  $a^*$  代入式(2)后,得到第  $i$  个样品的投影值  $z^*(i)$ , 根据  $z^*(i)$  与  $y(i)$  的散点图可建立水资源潜力综合评价的数学模型.详见第2部分.

## 2 水资源潜力综合评价

本文以文献[3]数据为依据,鉴于水资源潜力的大小与水资源开发利用情况关系密切,将水资源潜力的评价分为2个层次,即水资源开发利用所处阶段及水资源潜力的大小.将水资源潜力分为3个等级,相应的水资源开发利用也分为3个阶段:(i)初始阶段  $V_1$ , 水资源开发潜力巨大;(ii)发展阶段  $V_2$ , 水资源进一步开发仍具有较大的潜力;(iii)饱和阶段  $V_3$ , 水资源进一步开发潜力较小.3个阶段  $V_1, V_2, V_3$  对应的水资源潜力的3个等级分值分别为0.95, 0.5, 0.05, 分值越大,表示水资源潜力越大.

另外,影响水资源潜力的因素有很多,根据区域水资源特征,我们选取的评价指标为<sup>[3]</sup>: (i) 水资源开发利用率  $U_1$ ; (ii) 灌溉率  $U_2$ ; (iii) 地表水控制率  $U_3$ ; (iv) 重复利用率  $U_4$ ; (v) 人均占有水量  $U_5$ ; (vi) 人均供水量  $U_6$ ; (vii) 平均渠系水利用系数  $U_7$ ; (viii) 客水利用率  $U_8$ ; (ix) 供水量模数  $U_9$ ; (x) 开发利用程度  $U_{10}$ .表1列出了10个指标的情况.

表1 评价标准<sup>[3]</sup>

指标	$U_1(1)$	$U_2(2)$	$U_3(3)$	$U_4(4)$	$U_5(5)$	$U_6(6)$	$U_7(7)$	$U_8(8)$	$U_9(9)$	$U_{10}(10)$
$V_1$	<50	<15	<5	<50	>345	>345	<0.55	<0.2	<10	<30
$V_2$	50~75	15~50	5~25	50~80	300~345	250~345	0.55~0.73	0.2~1	10~15	30~70
$V_3$	>75	>50	>25	>80	<300	<250	0.73	>1	>15	>70

在表1中各等级取值范围内均匀随机产生各10000个指标样本  $x^*(i, j)$ , 与对应的等级  $y(i)$  一起组成样本系列, 并对  $x^*(i, j)$  进行归一化处理为  $x(i, j)$ ,  $i = 1, \dots, 30000$ ;  $j = 1, \dots, 10$ , 用 GPPM 中的(1)~(4)式得最佳投影方向

$$a^* = (0.466, 0.285, 0.156, 0.380, -0.213, -0.454, 0.256, 0.322, 0.150, 0.373).$$

其中最佳投影方向各分量值代表了相应各指标的权重, 而由于指标1的权重最大, 表明其对应的水资源开发利用率指标对水资源潜力评价影响最大, 这与经验相符.

最大投影目标函数值  $f(a^*) = 0.75$ .

最佳投影值  $z^*(i)$  与等级  $y(i)$  的散点图见图1. 从图1可以看出:

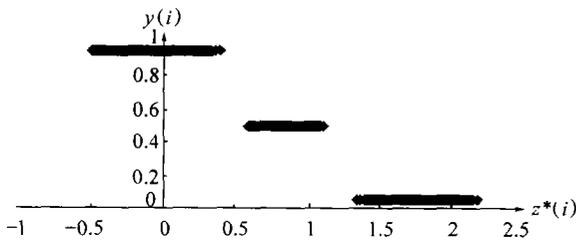


图1 投影值  $z^*(i)$  与经验等级  $y(i)$  的散点图

(1)  $z^*(i)$  与  $y(i)$  的图形为阶梯型、下降曲线, 可以用与该曲线对应的函数即(5)式建立水资源潜力综合评价的数学模型. 由于本文取的样本较多为10000, 所以投影值  $z^*(i)$  与经验等级  $y(i)$  的散点图中横坐标区间的端点易定出, 在端点处附近还可以通过比较投影值的大小, 来确定水资源潜力的大小. 用 GPPM 计算的水资源潜力评价模型为

$$y^*(i) = \begin{cases} 0.95 & z^*(i) \leq 0.5 \\ 0.50 & 0.5 < z^*(i) \leq 1.3 \\ 0.05 & z^*(i) > 1.3 \end{cases} \quad (5)$$

(2) 用 GPPM 计算的各个样本的最佳投影值  $z^*(i)$  与对应等级  $y(i)$  之间的误差为0, GPPM 在

一定条件下还是稳定的(另文讨论). 可见用 GPPM 对水资源潜力进行综合评价精度很高.

关中地区水资源潜力综合评价的各指标数值详见文献[3]. 于是用 GPPM 和文献[3]的数据, 可对关中地区水资源潜力作出综合评价, GPPM 对关中地区水资源潜力评价结果与文献[3]的模糊综合评判方法计算的结果对比详见表2. GPPM 计算结果表明, 整个关中地区水资源潜力处于发展阶段  $V_2$ , 水资源开发利用仍保持上升势头, 水资源开发利用潜力较大. 就5个地市的水资源开发潜力而言, 宝鸡处于初始阶段  $V_1$ , 水资源开发利用潜力巨大; 西安、铜川、咸阳、渭南处于发展阶段  $V_2$ , 水资源开发利用潜力较大. GPPM 即可以利用分值从整体上来判别各地区水资源开发利用所属阶段, 又可以利用投影值对属于同一阶段的各地区水资源潜力大小进行细致评价. 例如, 西安、铜川分值都为0.5, 而投影值分别为1.012, 1.271, 虽它们都处于发展阶段  $V_2$ , 但由于投影值与分值的图形为阶梯型、下降曲线, 可以得出西安比铜川水资源开发利用潜力更大一些. 模糊综合评判<sup>[3]</sup>的评价结果为西安比铜川水资源开发利用潜力大, 它们都介于初始阶段  $V_1$  和发展阶段  $V_2$  之间, 可见用模糊综合评判<sup>[3]</sup>来判别水资源开发利用所属阶段, 结果较为“模糊”.

表2 关中地区水资源潜力综合评价结果比较

地区	投影值	模糊综合评判 <sup>[3]</sup>		GPPM 综合评判	
		分值	阶段	分值	阶段
关中	0.780	0.532	$V_1 \sim V_2$	0.500	$V_2$
西安	1.012	0.475	$V_2 \sim V_3$	0.500	$V_2$
铜川	1.271	0.337	$V_2 \sim V_3$	0.500	$V_2$
宝鸡	0.376	0.631	$V_1 \sim V_2$	0.950	$V_1$
咸阳	0.992	0.389	$V_2 \sim V_3$	0.500	$V_2$
渭南	1.092	0.330	$V_2 \sim V_3$	0.500	$V_2$

### 3 结论

(1) GPPM 的评价结果直观、可靠. 利用 GPPM 可把水资源潜力多维评价指标综合成一维投影指标, 根据投影指标值的大小就可以对水资源潜

力进行统一评价, 解决了各单项水资源潜力评价指标评价结果的不相容问题。

(2) GPPM 不仅可以对水资源潜力进行综合评价, 同时可利用最佳投影方向科学地确定权重和对水资源潜力评价指标进行检验。

(3) GPPM 可提高水资源潜力综合评价问题各层次的分辨力。

### 参 考 文 献

- 1 王 颖, 等. 人类活动与黄河断流及海岸环境影响. 南京大学学报(自然科学), 1998, 34(3): 257
- 2 王爱军, 等. 黄河断流对全球气候变化的响应. 自然灾害学报, 2002, 11(2): 103
- 3 贾 嵘, 等. 区域水资源潜力综合评判. 西北工业大学学报, 1999, 27(5): 44
- 4 杨纶标, 等. 模糊数学原理及其应用. 广州: 华南理工大学出版社, 2001. 129~146
- 5 马云东, 等. 改进的 AHP 法及其在多目标决策中的应用. 系统工程理论与实践, 1997, 17(6): 40
- 6 吕 锋, 等. 多目标决策灰色关联投影法及其应用. 系统工程理论与实践, 2002, 22(1): 103
- 7 陈 武, 等. 应用多目标决策-理想点法综合评价水环境质量. 环境工程, 2002, 20(3): 64
- 8 钱颂迪, 等. 运筹学. 北京: 清华大学出版社, 2002. 451~453
- 9 薛建军, 等. 人工神经网络在水质评价中的应用. 水文, 1997: 37
- 10 Friedman J H, et al. A projection pursuit algorithm for exploratory data analysis. IEEE Trans on Computer, 1974, 23(9): 881
- 11 李祚泳. 投影寻踪技术及其应用进展. 自然杂志, 1997, 19(4): 224
- 12 金菊良, 等. 水质综合评价的投影寻踪模型. 环境科学学报, 2001, 21(4): 431
- 13 韩 明. 无失效数据可靠性进展. 数学进展, 2002, 31(1): 7
- 14 邢文训, 等. 现代优化计算方法. 北京: 清华大学出版社, 2000. 140~148
- 15 Holland J H. 基因算法. 科学(中译本), 1992, (11): 24
- 16 Peck C C, et al. Genetic algorithms as global random search methods: An alternative perspective. Evolutionary Computation, 1995, 3: 39
- 17 李敏强, 等. 遗传算法的模式欺骗性分析. 中国科学, 2002, 32(1): 95
- 18 杨晓华. 参数优选算法研究及其在水文模型中的应用. 南京: 河海大学出版社, 2002. 44~55



### 国家自然科学基金研究专著

《非流态化气固两相流——理论及应用》张文尝等 编著  
科学出版社 定价: 23.00 元

本书由 12 章组成. 第 1 章介绍了将散料视为连续介质的散料力学的基础理论; 第 2 章简要介绍了将气固两相均视为流体的双流体动力学模型的基础知识; 第 3 章阐明了非流态化气固两相流的基本特征及量化判据, 提出了表示各种流型及其存在与相互转化条件的相图; 第 4 章较完整地提出了非流态化气固两相流动力学理论, 其中包括固相动量守恒方程、气体压降方程、空隙率方程组成的微分方程组的推导、方程组的数值解法及简化的解析解法; 从第 5 章到第 11 章是应用非流态化气固两相流的基本理论解决固定床、移动床、料腿、V 型排料阀、L 型排料阀、气动式排料阀、孔口落料、架拱等特定对象的定量计算与设计问题, 并结合具体实验予以验证; 第 12 章介绍了非流态化气固两相流研究中特有的测量技术: 颗粒对器壁压力的测量及空隙率微小变化的测量。

本书力求通俗易懂, 图文并茂, 理论与实际应用相结合, 可作为从事颗粒流体两相流、流态化技术、其他颗粒学分支的科研人员及相关的工程技术人员的专业学习参考书, 也可作为大专院校教师、研究生的教学参考书。