

第四届数学中国数学建模网络挑战赛

地址：内蒙古数学会
电话：0471-5220129

邮编：010021

网址：www.tzmcm.cn
Email：2011@tzmcm.cn

第四届“互动出版杯”数学中国

数学建模网络挑战赛

承 诺 书

我们仔细阅读了第四届“互动出版杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：1387

参赛队员（签名）：

队员 1：罗时超

队员 2：赵珍丽

队员 3：赵娟

参赛队教练员（签名）：杜世平

参赛队伍组别：大学本科

第四届“互动出版杯”数学中国

第四届数学中国数学建模网络挑战赛

地址：内蒙古数学会

电话：0471-5220129

网址：www.tzmcm.cn

Email：2011@tzmcm.cn

邮编：010021

数学建模网络挑战赛

编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

1387

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

第四届数学中国数学建模网络挑战赛

地址: 内蒙古数学会
电话: 0471-5220129

邮编: 010021

网址: www.tzmcm.cn
Email: 2011@tzmcm.cn

2011 年第四届“互动出版杯”数学中国 数学建模网络挑战赛

题 目 : 生物多样性的评估

关 键 词 PSR模型 生物多样性等级聚类 模糊层次分析

灰色预测 熵权法 Q型聚类 模糊聚类

摘 要:

虽然某些地区的生物多样性较为丰富,但其整体处于退化的阶段,因此对生物多样性发展趋势的研究和保护区的优先级评定具有重大意义。

问题一中构建了压力—状态—响应(PSR)框架分类体系,总结出生态系统健康指数的层次结构,建立了PSR模型。利用模糊层次分析方法得到了各个可测量指标的权重,进而求出PSR模型的三大指标(综合状态指数CSI、综合压力指数CPI、综合响应指数CRI)的权重,再根据PSR模型计算出综合健康指数。在此基础上,根据某地区跨度5年的生态系统健康状况统计数据,建立了灰色预测模型,得到了下一年该地区的CSI、CPI、CRI的预测值,由此得到CHI的预测值。得出该地区的森林综合健康程度由原来的很健康变为一般健康。

对于问题二,首先运用层次分析法对特种多样性(包括植物丰富度、动物丰富度)生态系统类型、植被垂直层谱、中国特有种及外来入侵度这五个指标进行排序,考虑到层次分析得出指标的权系数具有一定的主观性,采用熵权法对系数进行了修正,然后进行加权求和,建立生物多样性指数BI的表达式,计算出各个省生物多样性指数,并基于生物多样性等级进行聚类。同时运用Q型聚类,模糊聚类求解,最终分析3类聚类方法的结果,发现Q型聚类和模糊聚类的结果完全一样,并且与基于生物多样性等级进行聚类的结果相差极小。得出生物多样性为优的省是:云南。生物多样性为差的也是优先保护的省是:天津、黑龙江、青海、宁夏、江苏、上海、辽宁、内蒙古、北京、河北、山西。最后我们基于此求解思想对保护区的优先级评定撰写了一份报告,供联合国环境规划署参考。

参赛队号 1387

参赛密码 _____
(由组委会填写)

所选题目 B

第四届数学中国数学建模网络挑战赛

地址：内蒙古数学会
电话：0471-5220129

邮编：010021

网址：www.tzmcm.cn
Email: 2011@tzmcm.cn

英文摘要（选填）

（此摘要非论文必须部分，选填可加分，加分不超过论文总分的 5%）

Abstract

Although the biological diversity of someplace is rich, the entirety of biological diversity remains regressing, so the study of the trends of development of biological diversity and the priority evaluation for every conservation area are important.

Question one, this paper establishes a pressure-state-response (PSR) frame classification, summarize the hierarchical structure of ecosystem health index, then establish the PSR model. Taking advantage of the indistinct administrative level analytic procedure to get the weight of every measurable index and then find out the weight of the three indexes which include composite state index, composite pressure index and composite response. Finally we calculate the composite health index according to the model of PSR. Getting the predicted value of CSI, CPI and CRI of next year is designed to calculate the value of CHI. We establish a model of the gray predict, based on the statistics data of ecosystem health in a region fly-pasting 5 years. At last we draw a conclusion that the forest composite health level becomes common healthy from quite healthy.

Question two, first of all, we apply the analytic hierarchy process to the sorting order of 5 indexes, the diversity of special type, which consist of botanical richness and animal richness, the type of ecosystem, the vertical sheaf of vegetation, the special of China and the invasion of foreign specials.

In consideration of the indicative weight coefficient which is calculated by AHP provided with a certain subjectivity, we change the way, and use the method of entropy method to revise the coefficient, next, summarize the coefficient by weighing, then establish the expression of biological diversity index BI to calculate the biological diversity index of all provinces and cluster the indexes by biological diversity rank. At the same time, we analyze the result of three methods comprising Q cluster and indistinct cluster, and find out the discrepancy is very small. A conclusion is reached: Yunnan province is the superior, some worse province are Tianjin Heilongjiang and so on.

In the end, we write a report to the UN about the conclusion of priority level evaluation.

#1387

目录

1、问题的提出.....	2
2、模型的假设与符号说明.....	2
2.1 模型的假设.....	2
2.2 符号说明.....	2
3、问题的分析.....	3
3.1 问题一的分析.....	3
3.2 问题二的分析.....	3
4、模型的建立与求解.....	3
4.1 问题一.....	3
4.1.1 模型的分析.....	3
4.1.2 指标权重的确定(模糊层次分析法).....	5
4.1.3 灰色预测模型的建立.....	7
4.2 问题二.....	11
4.2.1 问题思路及前期准备.....	11
4.2.2 模型 1 的建立与求解(基于生物多样性等级聚类).....	13
4.2.3 模型 2 的建立与求解(Q型聚类).....	18
4.2.4 模型 3 的建立与求解(模糊聚类).....	21
4.2.5 三个聚类方法的比较与结论.....	23
4.2.6 评定优先保护地区的报告.....	23
5、模型的优缺点评价.....	23
5.1 本文的优点.....	25
5.2 本文的缺点.....	25
6、模型的推广.....	25
7、参考文献.....	25
8、附录.....	26

#1387

1、问题的提出

问题一

设计一个合理的估计方法，依据可测量的指标，估计某个地区的生物多样性的发展情况。

问题二

- 依据合理的数学模型，设计一个指标，综合各方面因素，以便评定应该优先保护哪些地区。
- 撰写一份报告，提交联合国环境规划署。

2、模型的假设与符号说明

2.1 模型的假设

- 影响生物多样性的因素：物种丰富度、生态系统类型、多样性植被垂直层谱的完整性、物种特有性、外来物种入侵度这五个因素在近几年内没有变化。
- 不考虑生物生存环境受到重大天灾的影响；
- 生物多样性模型在相对理想化的环境、时间、控制因素下进行。
- 本文引用数据、资料均为可靠真实。
- 假定森林生态系统没有遭受自然灾害而引起的生物多样性剧减。
- 假定生态资源数据统计真实可靠，不会对模型的精度造成影响。
- 假定森林保护区的级别没有发生变化，将其设定为 1。
- 假设在统计数据期间，没有产生影响三大指标的新因素，指标的代表性不变。

2.2 符号说明

	符号	意义
第一问	CHI	生态系统健康指数
	CSI	综合状态指数
	CPI	综合压力指数
	CRI	综合响应指数
第二问	w_j	第j项指标的权重
	BI	生物多样性指标
	P_{ij}	第j项影响因素下第i个指标与该因素的比重
	X_{ij}	第i指标第j个指标与该因素的数值
	e_j	第j项影响因素的熵值

#1387

g_j	差异系数
a	归一化后的物种丰富度
b	归一化后的生态系统多样性
c	归一化后的植被垂直层谱的完整性
d	归一化后的物种特有性
e	归一化后的外来物种入侵度
n	评价对象
m	评价指标

3、问题的分析

3.1 问题一的分析

求解生物多样性的发展情况，通过将其转化为生态系统的健康指数来求解。根据网上相关资料，根据压力—状态—响应（*PSR*）框架模型分类，总结出生态系统的健康指数的层次结构表，运用模糊层次分析法求出可测量的指标的权重，将权重通过一系列的转化，求出 *PSR* 模型的三大指标（综合状态指数 *CSI*、综合压力指数 *CPI*、综合响应指数）*CRI* 的权重，再根据 *PSR* 模型中综合健康指数计算公式计算出 *CHI*。通过网上公布的权威数据，我们得到某地区最近 5 年的与生态系统健康相关的一些指数的值，运用上述方法，求出这 5 年的生态系统健康指数 *CHI* 值，运用灰色预测建模方法，预测出该地区生态系统健康状况的发展趋势，即求得此地区的生物多样性的发展情况。

3.2 问题二的分析

题目要求综合各方面因素，以便评定应该优先保护哪些地区。首先提取影响生物多样性的重要指标，本文从横向和纵向考虑以及根据指标选取的科学性，代表性，实用性等方面，最终确定以特种多样性（包括动物丰富度和植物丰富度）、生态系统多样性、植被垂直层谱完整性、物种特有性及外来物种入侵度五个方面综合考虑。然后运用层次分析求解各指标的权重，由于层次分析带有一定的主观性，故建立熵权法修正权重，然后建立生物多样性 *BI* 的表达式，求解 31 个省的 *BI*，然后对 31 个省 *BI* 进行等级聚类，此外建立 *Q* 型聚类、模糊聚类，最终比较分析结果，得出优先保护的地区。

4、模型的建立与求解

4.1 问题一

4.1.1 模型的分析

这里对森林生物多样性进行动态分析，灰色理论系统是研究这种具有少数据和贫信息带来的灰色不确定性的工具。因此我们建立森林生物多样性的灰色预测模型。首先我们以 *PSR* 模型为基础，建立森林生态系统的健康评价体系，并以此来估计某地区的生态多样性发展状况。

PSR 模型是最常用，最有效的用于分析生态系统健康状况的模型。压力—状态—响应（*PSR*）框架模型：

#1387

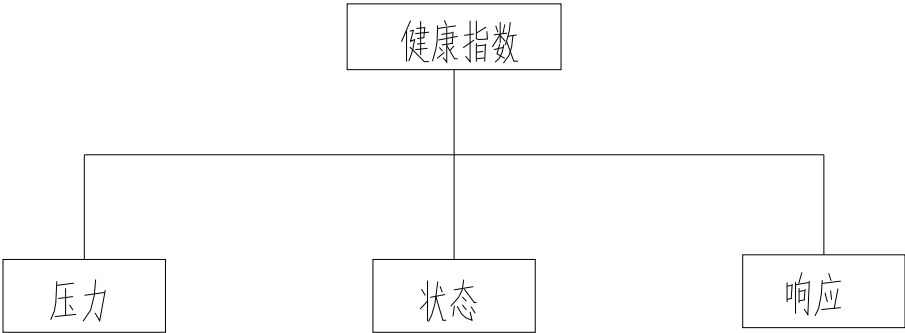


图1 压力—状态—响应（PSR）框架模型

针对某地区的森林，要建立森林生态系统评价指标体系，全面真实的反应森林生态系统的健康状况，综合各方面因素，我们得到森林生态系统健康指标体系层次结构表：

表1 森林生态系统健康指标体系层次结构表

目标层	项目层	要素层	指标层
	压力	人口状况	数量
			密度
		经济发展水平	GDP
			人均 GPD
			工业产值
			GDP 增长速度
		资源利用状况	耕地面积
		环境污染程度	环境污染
	状态	森林状况	面积
			覆盖度
			种类
			平均胸径
		森林其他生物状况	鱼类
			贝类
			鸟类
		森林环境状况	平均气温
			年降水量
	响应	大众意识	环保意识
			宣传教育
		保护区情况	级别
			人员数量
			年经费
			面积
		科研水平	国内外项目
			发表论文数
			科研考察

#1387

评价指标体系的选取必须具有相当的完备性和代表性,能够中反应影响森林生态系统健康的各种因素,依据评价指标体系设置的整体性原则、可操作性原则、层次性原则、动态性原则,从压力、状态、响应3个方面选取能切实反应森林生态系统健康状况的指标,评价指标体系归纳为:

- 目标层:以森林生态系统健康综合指数作为总目标层;
- 项目层:压力、状态、响应;
- 要素层:由构成压力、状态、响应项目层的要素组成;
- 指标层:由可直接度量的指标构成。4.1.2 指标权重的确定(模糊层次分析法)

4.1.2 指标权重的确定(模糊层次分析法)

模糊层次分析法克服了层次分析法中判断矩阵的一致性与人类思维一致性存在的显著差异,引入了模糊一致矩阵,无需再进行一致性检验,同时使用幂法来计算排序向量,可以减少迭代次数,提高收敛速度,满足计算精度的要求

系统的递阶层次结构建立,所建立的系统递阶层次结构与PSR模型做构建的指标体系结构相同。

①构建优先判断矩阵

根据下表所示的0.1-0.9的模糊判断尺度,建立优先判断矩阵 $F = (f_{ij})_{n \times n}$

表2 0.1-0.9的模糊判断尺度表

标度	定义	说明
0.5	同等重要	两个元素相比,同等重要
0.6	稍微重要	表示两个元素相比,一个元素比另一个元素稍微重要
0.7	明显重要	表示两个元素相比,一个元素比另一个元素明显重要
0.8	重要得多	表示两个元素相比,一个元素比另一个元素重要得多
0.9	极端重要	表示两个元素相比,一个元素比另一个元素极端重要
0.1, 0.2 0.3, 0.4	反比较	若元素 a_i 和元素 a_j 相比较得到判断 r_{ij} , 则元素 a_i 与元素 a_j 相比较得到的判断为 $r_{ji} = 1 - r_{ij}$ 。

②模糊一致性矩阵

再将优先关系矩阵转化为模糊一致性矩阵: $R = (r_{ij})_{n \times n}$

$$r_{ij} = (r_i - r_j) / 2n + 0.5, \text{其中 } r_i = \sum_{j=1}^n f_{ij}, i = 1, 2, \dots, n. \quad (1)$$

③计算排序向量(权重向量)

(1) 方根法

$$W^{(0)} = \left[\frac{\sum_{j=1}^n r_{1j}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}}, \frac{\sum_{j=1}^n r_{2j}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}}, \dots, \frac{\sum_{j=1}^n r_{nj}}{\sum_{i=1}^n \sum_{j=1}^n r_{ij}} \right]^T \quad (2)$$

#1387

(2) 利用幂法求解精度更高的排序向量(权重向量)

将互补判断矩阵 $R = (r_{ij})_{n \times n}$ 转化为互反判断矩阵 $E = (e_{ij})_{n \times n}$, 其中 $e_{ij} = r_{ij} / r_{ji}$ 排序向量 $W^{(0)}$ 作为初始向量 $V^{(0)}$, 用公式 $V^{(k+1)} = EY^{(k)}$, $Y^{(k)} = V^{(k)} / \|V^{(k)}\|_{\infty}$, $k = 1, 2, \dots$ 进行迭代. 若 $\|V^{(k+1)}\|_{\infty} - \|V^{(k)}\|_{\infty} < \varepsilon$, ε 为给定的误差, 则 $\|V^{(k+1)}\|_{\infty}$ 即为最大特征值,

$A = \left[V_{k+1,1} / \sum_{i=1}^n V_{k+1,i}, \dots, V_{k+1,n} / \sum_{i=1}^n V_{k+1,i} \right]^T$ 为排序向量, 否则, 继续迭代.

④求解指标的排序向量

根据网上相关数据, 得到生态系统健康指标的指标层的优先判断矩阵, 运用 matlab 软件(程序见附录 1)可求得指标层的权重:

表 3 生态系统健康指标的指标层的权重

指标层权重												
数量	密度	GDP	人均 GPD	工业 产值	GDP 增长 速度	耕地 面积	环境 污染	面积	覆盖 度	种类	平均 胸径	鱼类
0.04 4	0.03 9	0.03 3	0.04 1	0.03 7	0.04 0	0.04 0	0.03 9	0.04 0	0.03 8	0.03 9	0.03 9	0.04 2
贝类	鸟类	平均 气温	年降 水量	环保 意识	宣传 教育	级别	人员 数量	年经 费	面积	国内 外项 目	发表 论文 数	科研 考察
0.04 0	0.04 2	0.03 6	0.03 9	0.03 5	0.03 9	0.03 3	0.03 6	0.03 6	0.03 8	0.04 0	0.04 0	0.03 4

⑤指标的归一化

森林生态系统将抗指标体系指标数量较多, 包含生物指标, 物理指标, 化学指标以及经济指标等, 但大多指标能够定量, 对于少数定性的指标可采用文献分析等方法纳入整个指标体系中。对红树林生态系统健康评价需要进行归一化处理, 消除量纲的差异。

正向指标(当指标值越高表明生态系统越健康)归一化:

$$N = \frac{n - n_{\min}}{n_{\max} - n_{\min}} \quad (3)$$

逆向指标(当指标值越低表明生态系统越健康)的归一化:

$$N = \frac{n - n_{\max}}{n_{\min} - n_{\max}} \quad (4)$$

适度指标(指存在一个临界阈值的指标)的归一化:

#1387

$$N = \frac{n - n_{mid}}{n_{min} - n_{max}} \quad (5)$$

其中 n 代表原始值, n_{min} 代表所有原始值中的最小值, n_{max} 代表所有原始值中的最大值, n_{mid} 代表临界阈值。

⑥生态系统健康指数的计算

根据综合健康指数计算整个森林生态系统的总指数值,最后根据总指数分级数值范围确定湿地生态系统健康的等级。湿地综合健康指数的确定可根据以下的公式计算:

$$CHI = 100(CSI - 0.2CPI + 0.2CRI) \quad (6)$$

式中: CSI 为综合状态指数, CPI 为综合压力指数, CRI 为综合响应指数。

$$CSI = \sum_{i=1}^n W_i \cdot N_i \quad (7)$$

$$CPI = \sum_{i=1}^n W_i \cdot N_i \quad (8)$$

$$CRI = \sum_{i=1}^n W_i \cdot N_i \quad (9)$$

式中: n 为评价指标的个数; N_i 表示相对应的第 i 中指标的归一化值, $0 \leq N_i \leq 1$;

W_i 为指标 i 的权重。

4.1.3 灰色预测模型的建立

(1) 数据检验与处理

本文的原始数据采用某地林业局 5 次森林生态资源清查的统计数据,对数据进行相关性分析,由于清查期只有 5 次,具有少数数据和贫信息带来的灰色不确定性,因此选取灰色预测的方法对生物多样性的未来发展状况进行预测研究。为了保证建模方法的可行性,需要对已知数据列做必要的检验处理。设参考数据为:

$$x(0) = (x(0)(1), x(0)(2), \dots, x(0)(n)) \quad (10)$$

计算数列的级比:

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(0)} \quad (k = 2, 3, \dots, n) \quad (11)$$

如果所有的级比 $\lambda(k)$ 都落在可容覆盖 $X = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ 内,

#1387

则数列 $x^{(0)}$ 可以作为模型 $GM(1,1)$ 和进行数据灰色预测.

否则取适当的常数 c 做变换: $y^{(0)}(k) = x^{(0)}(k) + c \quad (k=1,3,\dots,n)$ 使得数列

$y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$ 的级比:

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in X \quad (k=2,3,\dots,n) \quad (12)$$

(2) 建立模型 $GM(1,1)$

所采用的 $GM(1,1)$ 模型的灰色微分方程为:

$$x^{(0)}(k) + az^{(1)}(k) = b \quad (13)$$

其时间响应序列为:

$$\hat{x}^{(1)}(k+1) = \left(x^{(1)}(0) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad k=1,2,\dots,n \quad (14)$$

参数 a, b 根据 $u = [a, b]^T = (B^T B)^{-1} B^T Y$ 求解, 其中:

$$B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \quad Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(n) \end{bmatrix},$$

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad (15)$$

最终解得微分方程预测模型为:

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k); \quad k=1,2,\dots,n \quad (16)$$

根据上述方法求得 CRI , CSI , CPI 的预测值, 计算出综合健康指数 CHI 的预测值。然后利用模型计算第 $k+1$ 个 CHI 清查期内对应指数的预测值。

(3) 模型的求解

灰色模型的参数 a, b 的值如表所示:

表 4 灰色模型的参数

指数	a	b	$GM(1,1)$
综合状态指数	-0.1422	98.2467	$\hat{x}_1^{(0)}(k+1) = 104.7842e^{0.1422k}$
综合压力指数	-0.3389	127.7861	$\hat{x}_2^{(0)}(k+1) = 123.7842e^{0.3389k}$

#1387

综合响应指数	0.0893	172.2031	$x_3^{(0)}(k+1) = 170.7842e^{-0.0893k}$
综合健康指数	-0.089	165.2071	$x_4^{(0)}(k+1) = 163.7842e^{0.089k}$

某地区连续五次清查的指标数据如表所示：

表 5 某地区连续五年的生态健康调查表

某地区连续五年的生态健康调查表					
数量	162	185	193	217	235
密度	570	650.9259	679.0741	763.5185	826.8519
GDP	134	189	232	254	280
人均 GDP	976.3	1021.622	1202.073	1850.599	2040.03
工业产值	132	152	168	183	209
GDP 增长速度	10.3	14.10448	17.83284	19.52388	21.52239
耕地面积	133	123	116	109	97
环境污染	35	38	40	39	36
面积	1388	1296	1257	1247	1231
覆盖度	99.7	93.09164	90.29027	89.57197	88.42269
种类	28	25	24	20	18
平均胸径	9.71	9.12	8.89	8.56	8.13
鱼类	127	116	102	108	104
贝类	110	103	89	82	78
鸟类	192	179	154	150	145
平均气温	23.2	24.6	25.1	25.8	26.4
年降水量	1725	1701	1693	1699	1713
环保意识	95.8	96.3	97.2	97.7	98.2
宣传教育	29	33	36	39	42
级别	1	1	1	1	1
人员数量	28	31	36	40	46
年经费	200	209	221	233	246
面积	3902	3924	3957	3968	3975
国内外项目	35	36	39	43	49
发表论文数	39	41	48	52	58
科研考察	32	37	43	47	53

由统计数据可以计算出 4 个生态健康指数序列，包括 CSI 为综合状态指数，CPI 为综合压力指数，CRI 为综合响应指数，综合健康指数。如表 6 所示：

表 6 各次生态健康指数

指数	第一次	第二次	第三次	第四次	第五次
综合状态指数	82.337	80.527	79.23	78.9236	78.524
综合压力指数	76.6248	81.0273	84.576	85.98	87.204
综合响应指数	73.7061	73.9603	74.785	74.238	75.0182
综合健康指数	80.54417	79.1136	77.2718	76.5752	76.08684

#1387

利用作 *MATLAB* 作图分析，分别得到综合状态指数、综合压力指数、综合响应指数、综合健康指数的发展趋势图（见图 1）；

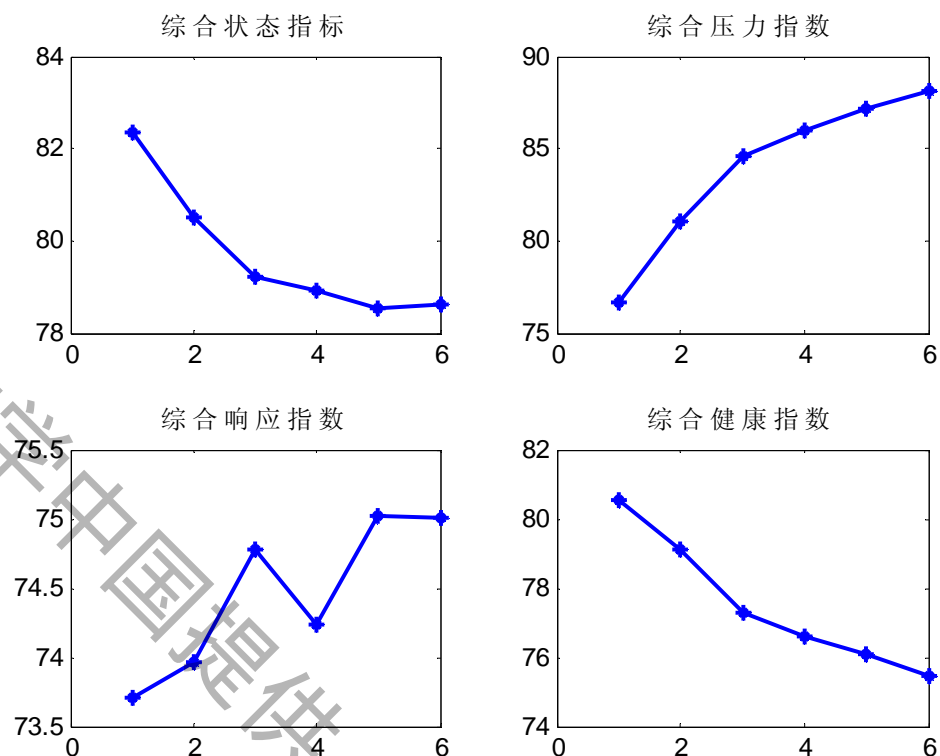


图 1 4 个生态指数发展趋势图

将上表数据带入微分方程模型 $x^{\Lambda(0)}(k+1) = x^{\Lambda(1)}(k+1) - x^{\Lambda(1)}(k)$; 得到下一次的生态系统综合健康指数为 75.4369。

(4) 模拟精度检验

本文采用准光滑度检验，平均相对误差 E （越小越好）、绝对关联度 d （越大越好）、均方差比值 C （越小越好）和小误差概率 p （越大越好）等进行精度检验。

表 7 模拟精度检验

模拟精度检验					
指数	平均绝对误差	绝对关联度	均方差比值	小误差概率	精度级别
综合状态指数	0.0436	0.9867	0.2801	1	2
综合压力指数	0.0651	0.9734	0.8013	0.2	3
综合相应指数	0.009	0.9978	0.0513	1	1
综合健康指数	0.0213	0.9932	0.1203	1	1

(5) 预测精度检验

采用原始数据 7 维新陈代谢子序列作滚动检验（邓聚龙，2002），求解模型的预测精度。

#1387

表 8 预测精度检验

指数	固定预测	精度	滚动预测	精度
综合状态指数	213.43	98.33	213.43	100
综合压力指数	101.97	97.99	101.97	100
综合相应指数	218.76	96.68	218.75	100
综合健康指数	256.56	98.64	256.56	100

(6) 模型结果分析

根据综合健康指数的数值结合森林的实际情况，依据 CHI 值可划分为：很健康、一般健康、亚健康。其中 $CHI \geq 80$ 为很健康， $60 \leq CHI \leq 80$ 为一般健康， $CHI \leq 60$ 为亚健康。

该地区压力指数相对较高，而且呈逐年上升的趋势，这说明该区周围的经济发展和人类经济活动与当地生态环境出现了不协调的趋势，人类活动对于森林生态系统的健康造成较大的压力。该地的状态指数下降，说明该区生态系统中人与自然之间的协调度下降，而响应指数上升，说明当地的保护区的工作做得越来越好，人类对自然的积极响应对森林的生态健康指数带来了很大的影响，但综合健康指数却在下降。结合模型结果，我们建议：当地政府能够积极的协调好社会经济发展与自然环境之间的关系；在保持已有资源的基础上加大对森林保护区的建设，使得生态健康指数有所提高；建议保护区多与国际上的优秀保护区进行交流，加强对外合作；建议政府加强对周边民众的环保教育，提高民众的环保意识。

4.2 问题二

4.2.1 问题思路及前期准备

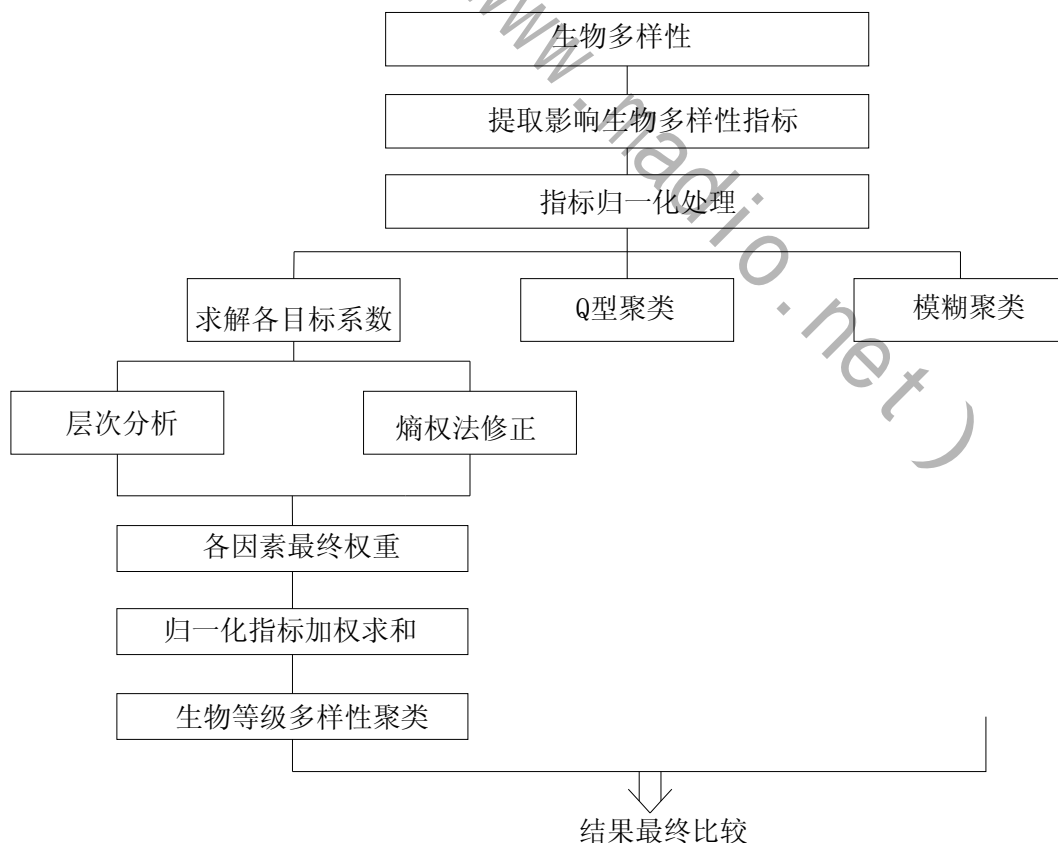


图 2 问题 2 的思路图

#1387

针对我国生物多样性受到严重威胁的现实，基于遗传多样性、物种多样性和生态系统多样性三个层次水平，本文从横向和纵向考虑以及根据指标选取的科学性，代表性，实用性等方面，最终确定以特种多样性（包括动物丰富度和植物丰富度）、生态系统多样性、植被垂直层谱完整性、物种特有性及外来物种入侵度五个方面综合考虑。其中前四种属于正向指标，第五种属于负向指标。

表9 省域生物多样性评价指标值原始数据

序号	省份	物种丰富度	生态系统类型多样性	植被垂直带谱的完整性	种子植物中国特有属数	外来入侵物种种数
1	北京	2528	95	60	6	13
2	天津	1630	16	60	5	10
3	河北	3610	110	60	13	61
4	山西	3188	111	60	16	35
5	内蒙古	3493	53	60	9	29
6	辽宁	3307	85	60	7	70
7	吉林	3145	94	80	2	45
8	黑龙江	2272	47	60	1	43
9	上海	2798	71	60	3	23
10	江苏	3424	79	60	18	97
11	浙江	6137	89	100	40	73
12	安徽	4375	97	100	26	84
13	福建	6515	140	100	31	73
14	江西	5083	130	100	46	64
15	山东	3921	110	80	15	65
16	河南	4499	106	80	36	53
17	湖北	6907	140	80	79	49
18	湖南	5658	146	100	61	43
19	广东	8726	90	100	49	71
20	广西	9913	140	100	75	47
21	海南	6568	70	100	22	50
22	重庆	6600	68	100	66	9
23	四川	11248	128	100	116	79
24	贵州	6747	146	100	80	54
25	云南	18317	174	100	116	145
26	西藏	7345	40	100	37	23
27	陕西	4566	89	80	47	47
28	甘肃	5370	85	80	39	38
29	青海	2566	50	60	13	27
30	宁夏	2306	69	60	4	21
31	新疆	4854	168	60	1	35

为了简化计算，即将有量纲的表达式，经过变换，化为无量纲的表达式，成为纯量。我们将原始数据进行归一化处理，

#1387

$$S_{ij} = \frac{Q_{ij}}{\text{MAX}(Q_{ij})} * 100 \quad (17)$$

归一化处理的结果见表 10

表 10 省域生物多样性评价指标归一化值

序号	省份	物种丰富度	生态生态系统类型多样性	植被垂直带谱的完整性	物种特有性	外来物种入侵度
1	北京	13.8	54.6	60	5.17	18.15
2	天津	8.9	9.2	60	4.31	21.66
3	河北	19.71	63.22	60	11.21	59.65
4	山西	17.4	63.79	60	13.79	38.75
5	内蒙古	19.07	30.46	60	7.76	29.31
6	辽宁	18.05	48.85	60	6.03	74.72
7	吉林	17.17	54.02	80	1.72	50.51
8	黑龙江	15.13	27.01	60	0.86	54.76
9	上海	15.28	40.8	60	2.59	29.02
10	江苏	18.69	45.4	60	15.52	100
11	浙江	33.5	51.15	100	34.48	41.99
12	安徽	23.88	55.75	100	22.41	67.78
13	福建	35.57	80.46	100	26.72	39.55
14	江西	27.75	74.71	100	39.66	44.45
15	山东	21.41	63.22	80	12.93	58.52
16	河南	24.56	60.92	80	31.03	41.58
17	湖北	37.71	80.46	80	68.1	25.04
18	湖南	30.89	83.91	100	52.59	26.83
19	广东	47.64	51.72	100	42.24	28.72
20	广西	54.12	80.46	100	64.66	16.74
21	海南	35.86	40.23	100	18.97	26.87
22	重庆	36.03	39.08	100	56.9	4.81
23	四川	61.41	73.56	100	100	24.79
24	贵州	36.83	83.91	100	68.97	28.25
25	云南	100	100	100	100	27.94
26	西藏	40.1	22.99	100	31.9	11.05
27	陕西	24.93	51.15	80	40.52	36.34
28	甘肃	29.32	48.85	80	33.62	24.98
29	青海	14.01	28.74	60	11.21	37.14
30	宁夏	12.59	39.66	60	3.45	32.15
31	新疆	26.5	96.55	60	0.86	25.45

4.2.2 模型 1 的建立与求解

(1) 层次分析法求解各指标权重

通过层次分析法首先对指标进行分析处理，以生物多样性指标为目标层，直接指标与间接指标为准则层，将动物丰富度、植物丰富度及植被垂直层谱归为直

#1387

接指标，生态系统多样性，外来物种入侵度及物种特有性归为间接指标，放入方案层，利用层次分析理论对各个指标进行赋权处理，最终得到所选取指标的层次总排序，分清孰重孰轻，为后序工作的开展有重大意义

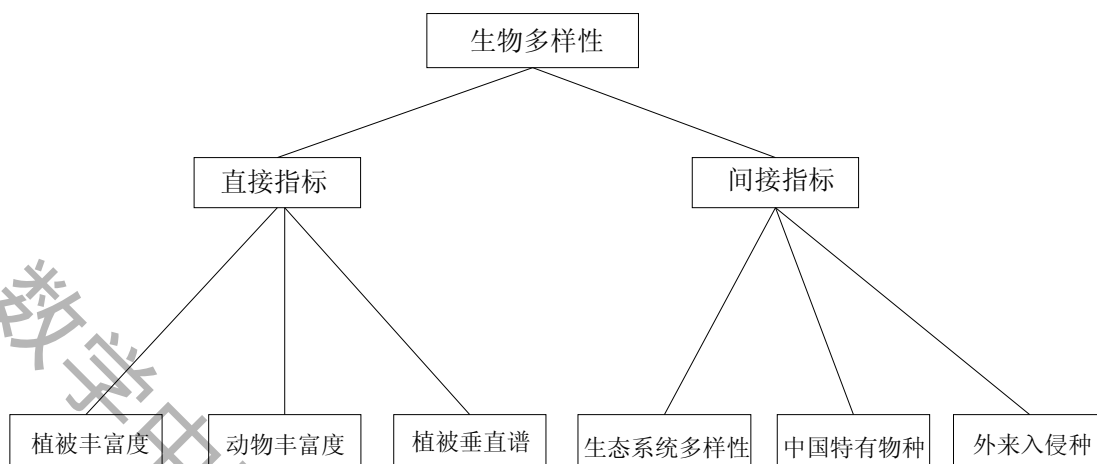


图3 层次分析结构图

1. 层次分析法概述

层次分析法是由美国著名的运筹学专家Saaty 首先提出的，它合理的将定性与定量的决策结合起来，按照思维和心理的规律将决策的过程透明化。层次结构如图3所示，通过将指标两两比较的方式建立判断矩阵，通常使用9 标度法（见图表11）。构造的矩阵 $A = (a_{ij})$ 在理论上应该具有以下一致性： $a_{ik} * a_{kj} = a_{ij}$ 因此构造出的判断矩阵应该进行一致性检验。当矩阵A 为一致性矩阵时，其最大特征值所对应的特征向量归一化后即成为排序权向量。

2. 层次分析法求解准则层一级指标权重

设所要构造的判断矩阵为 $A = (a_{ij})$ 其中元素 a_{ij} 的设定根据下表所示的标度法

表 11：1-9 标度含义表

标度	含义
1	表示两个因素相比，具有相同重要性
3	表示两个因素相比，前者比后者稍重要
5	表示两个因素相比，前者比后者明显重要
7	表示两个因素相比，前者比后者强烈重要
9	表示两个因素相比，前者比后者极端重要
2, 4, 6, 8	表示上述相邻判断的中间值
1, 1/2, 1/3, ..., 1/9	若因素 i 与因素 j 的重要性之比为 a_{ij} ，那么 因素 j 与因素 i 重要性之比为 $a_{ji} = 1/a_{ij}$

#1387

构造的判断矩阵为：

$$A = \begin{bmatrix} 1 & 3 \\ \frac{1}{3} & 1 \end{bmatrix}$$

求解得

$$\lambda = 2.0 \quad CI = 0 \quad CR = 0$$

权向量

$$w = [0.75 \quad 0.25]^T$$

3. 层次分析法求子准则层二级指标权重

直接指标和间接指标下的判断矩阵分别为

$$A_1 = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ \frac{1}{3} & \frac{1}{3} & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 5 & 7 \\ \frac{1}{5} & 1 & 3 \\ \frac{1}{7} & \frac{1}{3} & 1 \end{bmatrix}$$

直接指标下 $\lambda = 3.0 \quad CI = 0 \quad CR = 0$

$$w_1 = [0.4286 \quad 0.4286 \quad 0.1429 \quad 0 \quad 0 \quad 0]^T$$

间接指标下 $\lambda_2 = 3.06489$, $CI = 0.032445$, $CR = 0.005594$

$$w_2 = [0 \quad 0 \quad 0 \quad 0.7034 \quad 0.1839 \quad 0.08096]^T$$

以上矩阵均有 $CR < 0.01$ ，均通过了一致性检验

$$w_z = w^*[w_1, w_2] = [0.3214 \quad 0.3214 \quad 0.1072 \quad 0.1827 \quad 0.0471 \quad 0.0202]$$

将动物丰富度和植物丰富度归为一类表示特种多样性，得到特种多样性、生态系统多样性、植被垂直层谱完整性、物种特有性及外来物种入侵度这5个指标的权系数为

$$w_{z'} = [0.6428 \quad 0.1827 \quad 0.1072 \quad 0.0471 \quad 0.0202]$$

由以上可知在生物多样性评价体系中，动物丰富度与植物丰富度应该占最大比例，接下来依次是生态系统类型，植被垂直层谱，中国特有种，外来入侵度最轻。

(2) 熵权法求解权重

考虑到层次分析法的实质是一种主观赋权法，而熵权法是一种客观赋权法。在实际应用中，由于客观因素的复杂性、不确定性，对因素的相对重要性程度不容易做出精确判断，并且由于没有充分考虑各指标提供的信息量，人为给出的权重难免缺乏客观的科学依据。由熵权法得出的权重独立于人的主观偏好以及经

验，克服了主观赋权法的不足。下面用熵权法修正层次分析所得的权系数。

1.熵值法基本原理

在信息论中，熵是对不确定性的一种度量。信息量越大，不确定性就越小，熵也就越小；信息量越小，不确定性越大，熵也越大。根据熵的特性，我们可以通过计算熵值来判断一个事件的随机性及无序程度，也可以用熵值来判断某个指标的离散程度，指标的离散程度越大，该指标对综合评价的影响越大，以此来定各个影响因素的权重。

2.熵值法步骤

a) 选取 n ($n=31$) 个指标， m ($m=5$) 个影响因素，则 x_{ij} 为第 i 个指标的

第 j 个影响因素的数值，($i=1,2,\dots,n; j=1,2,\dots,m$)

b) 指标的标准化处理:异质指标同质化

由于各项数值的计量单位并不统一，因此在使用它们计算综合影响因素前，我们先要对它们进行标准化处理，即把指标的绝对值转化为相对值，并令 $x_{ij} = |x_{ij}|$ ，从而解决各项不同质指标值的同质化问题。而且，由于正向指标和负向指标数值代表的含义不同(高优指标数值越高越好，低优指标数值越低越好)，因此，对于高低优指标我们用不同的算法进行数据标准化处理。其具体方法如下：

$$\text{高优指标: } x'_{ij} = \left[\frac{x_{ij} - \min(x_{1j}, x_{2j}, \dots, x_{nj})}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} \right] \times 100$$

$$\text{低优指标: } x'_{ij} = \left[\frac{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - x_{ij}}{\max(x_{1j}, x_{2j}, \dots, x_{nj}) - \min(x_{1j}, x_{2j}, \dots, x_{nj})} \right] \times 100$$

则 x'_{ij} 为第 i 个指标的第 j 个影响因素的数值($i=1,2,\dots,n; j=1,2,\dots,m$)。为了方便起见，仍记数据 $x'_{ij} = x_{ij}$ (x_{ij} 是标准化后的数据)

1) 计算第 j 项影响因素下第 i 个指标占该因素的比重：

$$p_{ij} = \frac{X_{ij}}{\sum_{i=1}^n X_{ij}}, (i=1,2,\dots,n, j=1,2,\dots,m) \quad (18)$$

2) 计算第 j 项影响因素的熵值。

$$e_j = -k \sum_{i=1}^n p_{ij} \ln(p_{ij}), \text{ 其中, } k > 0, k = 1/\ln(n), e_j \geq 0 \quad (19)$$

3) 计算第 j 项影响因素的差异系数。对第 j 项影响因素，其值的差异越大，对方案评价的左右就越大，熵值就越小，定义差异系数：

#1387

$$g_j = \frac{1-e_j}{m-E_e}, \text{ 式中 } E_e = \sum_{j=1}^m e_j, \quad 0 \leq g_i \leq 1, \quad \sum_{j=1}^m g_j = 1$$

4) 求权值:

$$w_j = \frac{g_j}{\sum_{j=1}^m g_j} \quad (1 \leq j \leq m) \quad (20)$$

3.熵值法权系数(程序见附录2)求解结果为

$$w_j = [0.5 \quad 0.15 \quad 0.1 \quad 0.15 \quad 0.1]$$

(3) 最终权重的确定

表 12.层次分析法求解权重与熵值法求解权重的比较及最终权重的确定

	物种多样性	生态多样性	植被完整性	物种特有性	外来物种入侵
层次分析法权重	6428	0.1827	0.1072	0.0471	0.202
熵权法权重	0.5	0.15	0.1	0.15	0.1
最终权重	0.5714	0.1664	0.1036	0.0986	0.0601

(4) 生物多样性 BI 的计算方法

$$BI = 0.5714 * a + 0.1664 * b + 0.1036 * c + 0.0986 * d + 0.0601 * (1 - e) \quad (21)$$

其中 a : 归一化后物种丰富度 b : 归一化后生态系统多样性 c : 归一化后植被垂直层谱的完整性 d : 归一化后物种特有性 e : 归一化后外来特种入侵度

(5) 各省生物多样性 BI 的值

表 13.各省生物多样性 BI 的值

编号	省份	BI	编号	省份	BI	编号	省份	BI
1	北京	28.62	11	浙江	44.9	21	海南	43.81
2	天津	17.97	12	安徽	37.43	22	重庆	48.78
3	河北	31.53	13	福建	50.34	23	四川	72.07
4	山西	31.81	14	江西	45.9	24	贵州	56.48
5	内蒙古	27.19	15	山东	34.81	25	云南	98.33
6	辽宁	26.77	16	河南	39.03	26	西藏	45.59
7	吉林	30.23	17	湖北	54.44	27	陕西	38.87
8	黑龙江	22.16	18	湖南	51.56	28	甘肃	40.99
9	上海	26.26	19	广东	54.64	29	青海	23.89
10	江苏	25.98	20	广西	66.05	30	宁夏	24.43
						31	新疆	41.99

#1387

(6) 各省基于生物多样性 BI 等级的聚类

表 14. 生物多样性状况的分级

生物多样性等级	生物多样性指数	生物多样性状况
优	$BI > 65$	物种高度丰富，特有属、种繁多，生态系统丰富多样
良	$40 < BI \leq 65$	物种较丰富，特有属、种较多，生态系统类型较多，局部地区生物多样性高度集中
一般	$30 < BI \leq 40$	物种较少，特有属、种不多，局部地区生物多样性较丰富，但生物多样性总体水平一般
差	$BI \leq 30$	物种贫乏，生态系统类型单一、脆弱，生物多样性极低

根据各省多样性 BI 的值，得出分类结果（见表 15）为：

表 15 省多样性 BI 的值，得出分类

编号	省份	BI	等级	编号	省份	BI	等级	编号	省份	BI	等级
25	云南	98.33	优	26	西藏	45.59	良	3	河北	31.53	中
23	四川	72.07	优	11	浙江	44.9	良	7	吉林	30.23	中
20	广西	66.05	优	21	海南	43.81	良	1	北京	28.62	差
24	贵州	56.48	良	31	新疆	41.99	良	5	内蒙古	27.19	差
19	广东	54.64	良	28	甘肃	40.99	良	6	辽宁	26.77	差
17	湖北	54.44	良	16	河南	39.03	中	9	上海	26.26	差
18	湖南	51.56	良	27	陕西	38.87	中	10	江苏	25.98	差
13	福建	50.34	良	12	安徽	37.43	中	30	宁夏	24.43	差
22	重庆	48.78	良	15	山东	34.81	中	29	青海	23.89	差
14	江西	45.9	良	4	山西	31.81	中	8	黑龙江	22.16	差
								2	天津	17.97	差

其中生物多样性为优的是：云南、四川、广西、

其中生物多样性为良的是：贵州、广东、湖北、湖南、福建、重庆、江西、西藏、浙江、海南、新疆、甘肃

其中生物多样性为一般的是：河南、陕西、安徽、山东、山西、河北、吉林

其中生物多样性为差的是：北京、内蒙古、辽宁、上海、江苏、宁夏、青海、黑龙江、天津

4.2.3 模型 2 的建立与求解（Q 型聚类）

(1) 方差分析

方差分析用于两个及两个以上样本差别的显著性检验。由于各种因素的影响，研究所得的数据呈现波动状。方差分析是在可比较的数组中，把数据间的总

#1387

的“变差”按各指定的变差来源进行分解的一种技术。对变差的度量，采用离差平方和。方差分析方法就是从总离差平方和分解出可追溯到指定来源的部分离差平方和，这是一个很重要的思想。

本文以中国省域生物多样性评价指标值为基础，根据对各个省份的生物多样性进行方差分析。在 Matlab 中，应用函数 $p = \text{anova1}(x)$ 进行平衡单因子方差分析，比较样本 $m \times n$ 的矩阵 X 中两列或多列数据的均值。其中，每一列包含有一个具有 m 个相互独立观测值的样本。它返回 X 中所有样本取自同一群体的零假设成立的概率 p 。若 p 值接近 0，则认为零假设可疑并认为列均值存在差异。本文方差分析的结果为

ANOVA Table					
Source	SS	df	MS	F	Prob>F
Columns	5.4807	30	0.18269	2.36	0.0005
Error	9.5879	124	0.07732		
Total	15.0686	154			

图4 中国各地生物多样性方差分析结果图

由 $p = 0.0005 < 0.01$ ，可知我国各省份的生物多样性存在显著差异，可以进行聚类分析。

(2) 对所 31 个省进行样本聚类

首先，将省域生物多样性评价指标归一化值标准化处理 以省作为样本，影响生态系统多样性的因素作为变量（指标）进行 Q 型聚类分析（对样本聚类），即每个样本都可以看成 p 维空间中的一点， n 个样本就是 p 维空间中的 n 个点。将 p 维空间中距离相近的点聚为一类，距离较远的点分成不同的类。

$$\text{标准化公式: } u_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (22)$$

表 16. 标准化处理结果

物种丰富度	生态系统类型多样性	植被垂直带谱的完整性	物种特有性	外来物种入侵度
0.0538	0.5	0	0.0435	0.1401
0	0	0	0.0348	0.177
0.1187	0.5949	0	0.1044	0.5761
0.0933	0.6012	0	0.1304	0.3566
0.1116	0.2341	0	0.0696	0.2574
0.1004	0.4367	0	0.0521	0.7344
0.0908	0.4936	0.5	0.0087	0.4801
0.0684	0.1961	0	0	0.5247
0.07	0.348	0	0.0175	0.2543
0.1075	0.3987	0	0.1479	1
0.27	0.462	1	0.3391	0.3906
0.1644	0.5127	1	0.2174	0.6615

#1387

0.2928	0.7848	1	0.2608	0.365
0.2069	0.7215	1	0.3914	0.4164
0.1373	0.5949	0.5	0.1217	0.5642
0.1719	0.5696	0.5	0.3043	0.3863
0.3162	0.7848	0.5	0.6782	0.2125
0.2414	0.8228	1	0.5218	0.2313
0.4252	0.4683	1	0.4174	0.2512
0.4964	0.7848	1	0.6435	0.1253
0.2959	0.3417	1	0.1827	0.2317
0.2978	0.3291	1	0.5653	0
0.5764	0.7088	1	1	0.2099
0.3066	0.8228	1	0.687	0.2462
1	1	1	1	0.243
0.3425	0.1519	1	0.3131	0.0656
0.176	0.462	0.5	0.4	0.3312
0.2241	0.4367	0.5	0.3304	0.2119
0.0561	0.2152	0	0.1044	0.3396
0.0405	0.3355	0	0.0261	0.2872
0.1932	0.962	0	0	0.2168

a) 原始数据的确定

由上表可知，资料矩阵为：

$$X = \begin{bmatrix} 0.538 & 0.5 & 0 & 0.0435 & 0.1401 \\ 0 & 0 & 0 & 0.0348 & 0.177 \\ 0.01187 & 0.5949 & 0 & 0.1044 & 0.5761 \\ \dots & \dots & \dots & \dots & \dots \\ 0.1932 & 0.962 & 0 & 0 & 0.2168 \end{bmatrix}$$

注：由于此矩阵为 31 行 5 列，由于太大，故简记，具体数据上表

b) 距离的确定

利用欧式距离：

$$d_{ij} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (23)$$

c) 运用 SPSS15.0 进行 Q 型聚类分析得

#1387

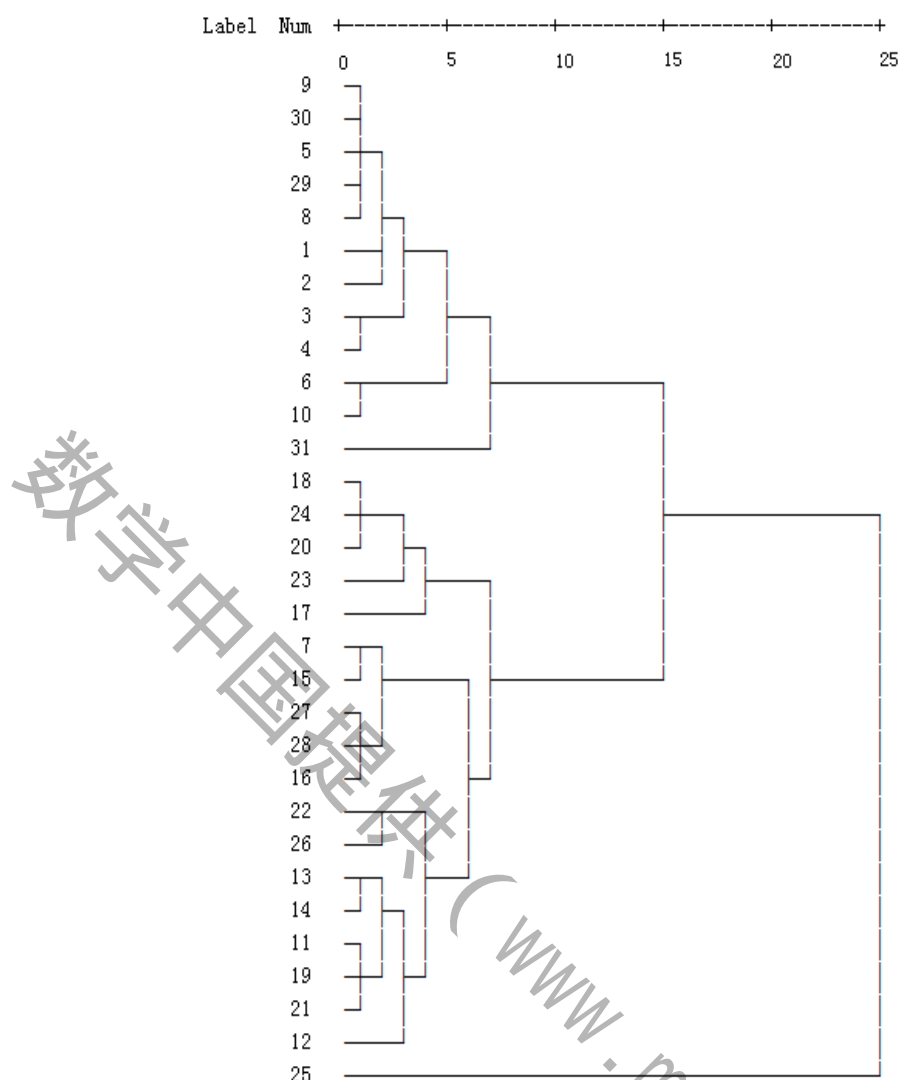


图 5 聚类结果图

由聚类结果可知，我国各地生物多样性可分为四类，分别为优、良、一般和差。

其中生物多样性为优的是：云南

其中生物多样性为良的是：四川、广西、贵州、广东、湖北、湖南、福建、重庆、江西、西藏、浙江、海南、安徽

其中生物多样性为一般的是：吉林、山东、河南、陕西以及甘肃

其中生物多样性为差的是：天津、黑龙江、青海、宁夏、江苏、上海、辽宁、内蒙古、北京、河北、山西

4.2.4 模型 3 的建立与求解（模糊聚类）

考虑样本与各因素间的模糊关系对 31 个省样本进行聚类分析，其样本集 $U=\{u_1, u_2, \dots, u_i, \dots\}$ ， u_i 表示第 i 个省域。每个样本有物种丰富度、生态系统类型、多样性植被垂直层谱的完整性、物种特有性、外来物种入侵度这五个因素这 5 个指标。所以这里可以用一个 5 维的行向量 $u_i=[u_{i1} \ u_{i2} \ u_{i3} \ u_{i4} \ u_{i5}]$ 来描述第 i 个样本。

#1387

1. 数据的标准化处理

首先，因为同一个样本有多个统计量，所以用这多个统计量的平均值来描述样本，这样处理后再对数据进行标准化处理。标准化公式同样为：

$$u_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (23)$$

得到结果如上表 16

2. 建立模糊相似矩阵

用绝对值减数法确定 r_{ij} ，其中 c 中取 0.25。：

$$r_{ij} = \begin{cases} 1 & (i = j) \\ 1 - c \sum_{k=1}^4 |u_{ik} - u_{jk}| & (i \neq j) \end{cases} \quad (i, j = 1, 2, \dots, 9) \quad (24)$$

通过在数学软件编写程序（见附录 3），得到了模糊相似矩阵 R ：

$$R = \begin{bmatrix} 1 & 0.8501 & 0.8358 & 0.889 & \dots & 0.8196 \\ 0.8501 & 1 & 0.7044 & 0.7576 & \dots & 0.6926 \\ 0.8358 & 0.7044 & 1 & 0.9307 & \dots & 0.7737 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.9144 & 0.8763 & 0.8238 & 0.877 & \dots & 0.7811 \\ 0.8196 & 0.6296 & 0.7737 & 0.8173 & \dots & 1 \end{bmatrix}$$

注上述矩阵为 31*31 的部分数据，具体数据见附表相似矩阵

3. 确定相似矩阵的等价矩阵

由于相似矩阵 R 已满足对称性及自反性，并不一定满足传递性，则变换过程主要进行对相似矩阵进行满足传递性的操作。使关系满足传递性的算法中，最出名的，就是 Washall 算法了，又称传递闭包法（它的思想在最短路的 Floyd 算法中亦被使用了）。在这里对相似矩阵 R 不断自乘（关系乘法），直至满足 $R^k = R^{2k}$ ，这时，称 R^k 为 R 的等价矩阵 R^* 。在 MATLAB

的程序中（见附录 3），编写了求解等价矩阵的程序，并求得等价矩阵 R^* 为：

$$R^* = \begin{bmatrix} 1 & 0.8848 & 0.889 & 0.889 & \dots & 0.8196 \\ 0.8848 & 1 & 0.8848 & 0.8848 & \dots & 0.8196 \\ 0.889 & 0.8848 & 1 & 0.9307 & \dots & 0.8196 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0.9229 & 0.8848 & 0.889 & 0.889 & \dots & 0.8196 \\ 0.8196 & 0.8196 & 0.8196 & 0.8196 & \dots & 1 \end{bmatrix}$$

注上述矩阵为 31*31 的部分数据，具体数据见附表等价矩阵

4. 聚类分析

对上问求出的 R^* 以参数 λ （ $\lambda \in (0,1)$ ）进行截取，当 λ 等于 0.865，时，进行聚类得到截矩阵：

#1387

$$R^*(0.865) = \begin{bmatrix} 1 & 1 & \dots & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 & \dots & 0 \\ 1 & 1 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & 1 & \dots & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}$$

注：图中两省略号中间的数据为 25 号云南的结果

此时将省域分成优、良、中、差 4 类，具体是：

其中生物多样性为优的是：云南

其中生物多样性为良的是：四川、广西、贵州、广东、湖北、湖南、福建、重庆、江西、西藏、浙江、海南、安徽

其中生物多样性为一般的是：吉林、山东、河南、陕西以及甘肃

其中生物多样性为差的是：天津、黑龙江、青海、宁夏、江苏、上海、辽宁、内蒙古、北京、河北、山西

4.2.5 三个聚类方法的比较与结论

在上述 3 种分类法中，发现在将 31 个省分为 4 类时，第 2 种（Q 型聚类）与第 3 种方法（模糊聚类）得到了完全相同的结果，且和第一种方法仅有很小很小的差别。由此看来，上述聚类法得到了较为合理的结果。模型 2 和模型 3 均略优于模型 1。

综上分析可以得出：优先保护的地区为天津、黑龙江、青海、宁夏、江苏、上海、辽宁、内蒙古、北京、河北、山西。

4.2.6 评定优先保护地区的报告

综合影响生物多样性的各方面因素，以便评定保护区的优先级。

1. 首先提取影响生物多样性的重要指标，本文从横向和纵向考虑以及根据指标选取的科学性，代表性，实用性等方面最终确定以特种多样性（包括动物丰富度和植物丰富度）、生态系统多样性、植被垂直层谱完整性、物种特有性及外来物种入侵度五个方面综合考虑。其中前四种属于正向指标，第五种属于负向指标

(1) 科学性原则：指标的选取应建立在对生物多样性充分认识，深入研究的基础上，选取的指标应目的明确，定义准确，必须能客观真实地反映生物多样性的基本特征变化规律和保护现状。

(2) 代表性原则：指标的选取应选取最能直接反映生物多样性本质特征的指标，排除一些与主要特征关系不密切的从属指标，使指标体系具有较高的代表性

(3) 实用性原则包括四层含义：一是所选指标的数据容易采集；二是便于更新；三是指标体系简明，综合性强；四是指标体系的应用具有较强的可操作性

2. 通过收集资料找到这五个指标对我国 31 个省市的生物多样性的影响。用归一化方法对所有数据进行归一化，归一化的公式

$$S_{ij} = \frac{Q_{ij}}{\text{MAX}(Q_{ij})} * 100$$

#1387

3. 运用层次分析求解各指标的权重，但在层次分析法中，人为的确定相对重要性，构造判断矩阵时带有一定的主观色彩，。虽然通过了一致性检验，但有时也难免造成误差较大的情况。而熵权法确定的指标的权重完全是根据数据间的关系来确定，但是这种客观权重值有时与实际情况相差较大，因此科学的权重应该为层次分析（主观权值）和熵权（客观权值）的综合度量相结合。达到主观与客观的完美结合。

4. 建立生物多样性 BI 系数的表达式，表达式为求解出的最终权重*归一化后的数据，并带入 31 个省归一化后的数据，得出 31 个省的 BI

$$BI = 0.5714 * a + 0.1664 * b + 0.1036 * c + 0.0986 * d + 0.0601 * (1 - e) \quad (25)$$

其中 a ：归一化后物种丰富度

b ：归一化后生态系统多样性

c ：归一化后植被垂直层谱的完整性

d ：归一化后物种特有性

e ：归一化后外来特种入侵度

5. 把各个省市的生物多样性分成了优，良，一般，差四个标准，然后对 31 个省 BI 进行等级聚类
生物多样性状况的分级

生物多样性等级	生物多样性指数	生物多样性状况
优	$BI > 65$	物种高度丰富，特有属、种繁多，生态系统丰富多样
良-	$40 < BI \leq 65$	物种较丰富，特有属、种较多，生态系统类型较多，局部地区生物多样性高度集中
一般	$30 < BI \leq 40$	物种较少，特有属、种不多，局部地区生物多样性较丰富，但生物多样性总体水平一般
差	$BI \leq 30$	物种贫乏，生态系统类型单一、脆弱，生物多样性极低

6. 建立 Q 型聚类、模糊聚类

1) 将省域生物多样性评价指标归一化值标准化处理 以省作为样本，影响生态系统多样性的因素作为变量（指标）进行 Q 型聚类分析（对样本聚类），即每个样本都可以看成 p 维空间中的一点， n 个样本就是 p 维空间中的 n 个点。将 p 维空间中距离相近的点聚为一类，距离较远的点分成不同的类。样品之间的距离越小，他们之间的相似程度越高；反之，距离越大，则表明两样品之间的相似程度越低。其中欧式距离：

$$d_{ij} = \left[\sum_{k=1}^p (x_{ik} - x_{jk})^2 \right]^{1/2} \quad (26)$$

2) 模糊聚类的实质是先找出标准化处理指标的相似矩阵，进而找出等价矩

#1387

阵，以参数 λ ($\lambda \in (0,1)$) 进行截取，求解结果

7. 最终比较 3 种聚类方法的结果，综合得出优先保护的地区为：优先保护的地区为天津、黑龙江、青海、宁夏、江苏、上海、辽宁、内蒙古、北京、河北、山西

5、模型的优缺点评价

5.1 本文的优点

(1) 将对生物多样性的发展情况研究转化为可测量指标(生态系统健康指数)变化趋势的考察，既综合考虑了环境、人类活动等对于生态系统健康指数的影响，又不失一般性的物种多样性研究。

(2) 层次分析得出指标的权系数具有一定的主观性，采用熵权法对系数进行了修正

(3) 对某地区的生态系统健康状况的发展趋势运用动态的灰色预测法，相对固定预测精度更高。

(4) 本文二问采用了基于生物多样性 BI 等级聚类，Q 型聚类，模型聚类，最后对结果进行了比较分析，得出优先保护的地区。

(5) *PSR* 模型的运用使得综合健康指数的计算简化，实用性更强。对指标权重的计算采用了模糊层次分析法，使得对指标权重的测定更加客观性。

5.2 本文的缺点

(1) 本文进行预测时由于数据有限，对模型的准确度造成一定的影响。

(2) 虽然采用相对客观的模型层次方法确定指标权重，但主观性依然较强。

(3) 本文没有考虑新的影响指标的加入，但使用的数据时间跨度较大，对预测的准确度有一定的影响。

6、模型的推广

层次分析与熵权法的结合还可以用于复杂系统的分析与决策，政策变动，自然灾害等对某地经济的影响大小模型。使结果达到更优。

7、参考文献

- [1] 尚天成, 高彬彬, 李翔鹏, 等. 基于层次分析法和熵权法的城市土地集约利用评价[J], 电子科技大学学报, 2009
- [2] 朱秀珍、崔远来、立远华等, 灌区运行状况综合评价权重系数的确定, 灌溉排水学报, 2004, 23
- [3] 许小勇. 模糊聚类分析算法的改进 Matlab 语言程序设计[J]. 云南民族大学学报. 15 (3): 196-199, 2006.
- [4] 宋丁全, 生物多样性基本概念及其数学方法, 金陵科技学院学报 2004 年第 2 期 20 卷
- [5] 武晟, 金苗. 熵权理论在水利工程投标中的应用[J]. 科技经济市场, 2011, (2): 28-30

#1387

[6] 白其峥, 数学建模案例分析, 北京, 海洋出版社, 2001. 1

[7] 姜启源. 数学模型 (第三版) [M]. 北京, 高等教育出版社, 2003

8、 附录

附录 1:

```

clear;
clc;
E=input('输入计算精度 e:')
Max=input('输入最大迭代次数 Max:')
F=input('输入优先关系矩阵 F:');
% 计算模糊一致矩阵
N=size(F);
r=sum(F');
for i=1:N(1)
    for j=1:N(2)
        R(i,j)=(r(i)-r(j))/(2*N(1))+0.5;
    end
end
E=R./R';
for i=1:N(1)
    S(i)=R(i,1);
    for j=2:N(2)
        S(i)=S(i)*R(i,j);
    end
end
S=S.^(1/N(1));
W=S./sum(S); % 方根法
a=input('参数 a=?');
W2=sum(R')/(N(1)*a)-1/(2*a)+1/N(1); % 排序法
V(:,1)=W'/max(abs(W)); % 利用幂法计算排序向量
for i=1:Max
    V(:,i+1)=E*V(:,i);
    V(:,i+1)=V(:,i+1)/max(abs(V(:,i+1)));
    if max(abs(V(:,i+1)-V(:,i)))<E
        k=i;
        A=V(:,i+1)./sum(V(:,i+1));
        break
    else
        end
end
end

```

2. 熵值法模型程序

x=zeros(1,m);

#1387

```

for j=1:m
for i=1:n
x(j)=X(i,j)+x(j);
end
end
for j=1:m
for i=1:n
P(i,j)=X(i,j)/x(j);
end
end
p=zeros(1,m);
for j=1:m
for i=1:n
p(j)=P(i,j).*log(P(i,j))+p(j);
end
end
E=0;
for j=1:m
e(j)=-1/log(n).*p(j);
E=e(j)+E;
end
G=0;
for j=1:m
g(j)=(1-e(j))/(m-E);
G=G+g(j);
end
for j=1:m
w(j)=g(j)/G;
end
S=zeros(1,n);
for i=1:n
for j=1:m
S(i)=S(i)+w(j)*P(i,j);
end
end
end

```

3. 模糊聚类相似矩阵与等价矩阵的求解程序

```

clear
clc
X=[ 0.0538    0.5000    0    0.0435    0.1401
    0         0        0    0.0348    0.1770
    0.1187    0.5949    0    0.1044    0.5761
    0.0933    0.6012    0    0.1304    0.3566

```

#1387

0.1116	0.2341	0	0.0696	0.2574
0.1004	0.4367	0	0.0521	0.7344
0.0908	0.4936	0.5000	0.0087	0.4801
0.0684	0.1961	0	0	0.5247
0.0700	0.3480	0	0.0175	0.2543
0.1075	0.3987	0	0.1479	1.0000
0.2700	0.4620	1.0000	0.3391	0.3906
0.1644	0.5127	1.0000	0.2174	0.6615
0.2928	0.7848	1.0000	0.2608	0.3650
0.2069	0.7215	1.0000	0.3914	0.4164
0.1373	0.5949	0.5000	0.1217	0.5642
0.1719	0.5696	0.5000	0.3043	0.3863
0.3162	0.7848	0.5000	0.6782	0.2125
0.2414	0.8228	1.0000	0.5218	0.2313
0.4252	0.4683	1.0000	0.4174	0.2512
0.4964	0.7848	1.0000	0.6435	0.1253
0.2959	0.3417	1.0000	0.1827	0.2317
0.2978	0.3291	1.0000	0.5653	0
0.5764	0.7088	1.0000	1.0000	0.2099
0.3066	0.8228	1.0000	0.6870	0.2462
1.0000	1.0000	1.0000	1.0000	0.2430
0.3425	0.1519	1.0000	0.3131	0.0656
0.1760	0.4620	0.5000	0.4000	0.3312
0.2241	0.4367	0.5000	0.3304	0.2119
0.0561	0.2152	0	0.1044	0.3396
0.0405	0.3355	0	0.0261	0.2872
0.1932	0.9620	0	0	0.2168]; %标准化后的矩阵

%-----计算模糊关系矩阵-----%

```

R=[];
a=size(X);
xlong=a(1);
for i=1:xlong
    for j=1:xlong
        if i==j
            R(i,j)=1;
        else
            temp=0;
            for k=1:5
                temp=temp+abs(X(i,k)-X(j,k));
            end
            R(i,j)=1-0.25*temp;
        end
    end
end
end

```


#1387

```

end
R
%-----计算模糊关系矩阵的等价矩阵-----%
unchanged=0;
cArr=R;
b=size(R);
cArrSize=b(1);
while unchanged==0
    unchanged=1;
    %--sigma:i=1:n(combine(conj(cArr(i,k),cArr(k,j))))
    for i=1:cArrSize
        for j=1:cArrSize
            mergeVal=0;
            for k=1:cArrSize
                if(cArr(i,k)<=cArr(k,j)&& cArr(i,k)>mergeVal)
                    mergeVal=cArr(i,k);
                elseif(cArr(i,k)>cArr(k,j)&& cArr(k,j)>mergeVal)
                    mergeVal=cArr(k,j);
                end
            end
            if(mergeVal>cArr(i,j))
                copyCArr(i,j)=mergeVal;
                unchanged=0;
            else
                copyCArr(i,j)=cArr(i,j);
            end
        end
    end
    %--copy back--%
    for i=1:cArrSize
        for j=1:cArrSize
            cArr(i,j)=copyCArr(i,j);
        end
    end
end
cArr

```