

基于信息熵的自然保护区定量评价^{*}

王本洋¹, 罗富和^{1,2}, 余世孝³

(1. 华南农业大学林学院, 广东 广州 510642 2. 广东省农业科学研究院,
广东 广州 510640 3. 中山大学生命科学学院, 广东 广州 510275)

摘 要: 针对自然保护区评价因指标众多、标准不一且难以量化表达而存在实际操作上的困难, 提出利用信息熵量化评价指标的新方法。定义自然保护区综合熵指数 GEI (Gross Entropy Index) 的计算模型为 $GEI = \frac{1}{k} \sum_{i=1}^k CE_i$ 和 $CE_i = -w_i p_i \log p_i$, 其中 CE_i (Criterion Entropy) 为第 i 个指标的熵, w_i 为第 i 个指标的权重, $p_i = x_i/X$ 为第 i 个指标的概率值, k 为选取的指标个数。自然保护区的 GEI 值越大, 则其不确定度越大, 生态复杂性程度越高, 其自然保护的价值也越大, 因而越需要得到有效的保护。对海南省 3 处典型自然保护区进行的实例研究表明, 基于信息熵的保护区定量评价方法, 充分利用了保护区的本底数据资料, 能较好地解决指标众多且标准不一、难以量化表达的问题, 分析结果具有较好的可比性, 是保护区综合定量评价的一种有效方法。

关键词: 自然保护区; 生态评价; 信息熵; 海南省

中图分类号: X826 Q149 **文献标识码:** A **文章编号:** 0529-6579 (2006) 06-0083-04

自然保护区是指将具有典型代表性的景观地理单元以及其它为科研、教育、文化娱乐目的而划分出的保护地域的总称^[1]。建立自然保护区对稀有濒危动植物种类的就地保护具有特别重要的意义^[2], 是保护自然生态系统和野生动植物物种多样性的主要方法和重要手段。

自然保护区建立之后, 应该及时进行评估以估计现有状况^[3]; 而保护区生态评价的关键是在统一的评价原则的基础上, 对不同区域尺度上同一类型生态系统, 建立生态评价的指标体系, 确定各指标的重要性并使之数量化^[4-5]。然而, 由于评价标准的多样化、相对性及其难以量化的表达, 根据自然保护的观点很难确定保护区的优劣^[3]。

国内外相关研究报道^[3-8]表明, 尽管自然保护区评价指标的标准不一, 但使用频次较高的包括多样性、自然性、稀有性、代表性、面积适宜性、稳定性和人类干扰等指标; 而且普遍使用基于专家打分的层次分析 (Analytic Hierarchy Process, AHP) 方法对其作数量化处理^[4-5, 8]。尽管 AHP 方法是大家普遍接受和认可的一种计算因子权重的数学方法, 并被广泛应用于许多行业和领域^[9-10], 但由于权重受专家的主观影响很大, 因而从评价结果的可比性来看, AHP 量化方法仍然缺乏严密性和完整性。

本文将在前人工作基础上利用信息熵实现有关评价指标的量化表达, 得到保护区综合评价指数, 并利用该方法进行保护区综合评价的实例研究, 以期对保护区的规划和管理提供相关理论参考。

1 基于信息熵的保护区定量评价指数

1.1 自然保护区评价的不确定性

Sutter^[11]把生态学分析中不确定性的来源分为内在随机性或变异性、知识不完备和评价活动中的谬误三类。Rowe^[12]认为, 不确定性源于或可纠正的信息缺乏; 并把不确定性分为过去或将来状态的不确定 (或时间不确定性)、复杂性带来的不确定性和测度不确定性三类。类似地, 保护区生态评价也表现出一定程度的不确定性。

信息论是解决不确定性问题的有效工具。根据信息论, 熵用于定量描述事件的不确定性, 熵越大表示不确定性越大。因此, 基于信息论的分析方法在生态学中获得广泛应用, 如 Shannon 的信息测度公式由 Margle 首次引入生态学用于测度多样性以后, 已与物种丰富度 S 和 Simpson 指数 D 一起成为少数几个应用最为广泛的多样性指数^[13]。基于 Shannon 信息熵 $H(X)$ 的多样性测度公式还被用于群落的复杂性研究中^[14-15]。

* 收稿日期: 2006-01-20

基金项目: 国家自然科学基金重点资助项目 (30430570); 华南农业大学校长基金资助项目 (4400-K05024)

作者简介: 王本洋 (1973 年生) 男, 博士, 讲师; 通讯联系人: 余世孝; E-mail: jssyx@mail.sysu.edu.cn

1.2 信息熵

Shannon (1948) 把统计熵作为基本组成部分推广用于信息理论中, 给熵以新的定义, 以表示系统的不确定性^[16]。作为一个随机事件的不确定性或信息量的量度, 熵与信息量建立了联系。

假设 X 是一个离散随机变量, 即它的取值范围 $R = \{x_1, x_2, \dots\}$ 是有限或可数的。设 $P_i = P\{X = x_i\}$, 则 X 的熵 (或称为信息熵^[17]) 定义^[18]为:

$$H(X) = - \sum_{i=1}^n P_i \log P_i \quad (1)$$

其中, 若令

$$I(x_i) = - \log P_i \quad (2)$$

则 $I(x_i)$ 表示 x_i 所含有的信息量, 并称为 x_i 的自信息量, 而信息熵 $H(X)$ 为其平均自信息量^[17]。

由于 Shannon 信息熵具有对称性, 即它只与随机变量的总体结构有关^[17], 所以它无法描述主观意义上对事件重要性判断的差别, 从而会淹没个别事件的重要性。加权熵既依赖于事件出现的概率, 又依赖于事件的权重系数 w_i , 其定义式^[17]为:

$$H_w(X) = - \sum_{i=1}^n w_i P_i \log P_i \quad (3)$$

1.3 综合熵指数

根据信息论原理和前述公式 (1)、(2)、(3), 自然保护区综合熵指数 (Gross Entropy Index GEI) 的计算模型为:

$$CEI = \frac{1}{k} \sum_{i=1}^k CE_i \quad (4)$$

$$CE_i = - w_i P_i \log P_i \quad (5)$$

其中 CE_i (Criterion Entropy) 为第 i 个指标的熵 (简称指标熵), w_i 为第 i 个指标的权重, $P_i = x_i/X$ 或 $P_i = 1 - (x_i/X)$ 为第 i 个指标的概率值 (计算时根据研究目的或生态意义取其中的一种形式)。k 为选取的指标个数, 需根据具体情况确定其大小。

对该计算模型的几点说明:

1) 指标熵 CE_i 为 Shannon 信息熵的特例, 即式 (1) 中 $i=1$ 和式 (3) 中 $q=1$ 时的情形; 由熵的可加性有 (5) 式。

2) 对数运算的底可取 2 或 10 或直接以自然对数 (底为 e) 进行运算, 可分别表示为 GE_2 、 GE_{10} 和 GE_e , 运算结果的单位分别为比特 (bits)、底特 (dets) 和纳特 (nats)。

3) 如果 $P_i=0$ 则规定 $CE_i=0$ 或该指标不参与计算 GEI 值。

4) 若所有指标的权重相等且为 w 则 CE 增大 w 倍; 若 w 等于 1, 则与原 CE 相等。

1.4 综合熵指数的生态学意义

根据熵理论, 一个系统的熵值越大, 则其不确定度也越大。一个自然保护区的综合熵指数越大, 则其不确定度越大, 生态复杂性程度越高, 其自然保护的价值也越大, 越需要得到合理的保护。

基于熵的评价, 不会受专家的主观影响, 因此其评价结果更具可比性。在实际操作过程中, 只需选取若干评价指标即可实现对不同保护区的综合定量评价, 并为其调整经营管理策略提供相关依据。

2 应用实例

2.1 数据来源

李景文等^[19]曾采用直接比较分类单位的大小和构建分类多样性指数 T_d 作综合比较两种方法, 从物种分类角度对海南省尖峰岭、五指山和霸王岭等 3 处典型自然保护区的生物多样性状况进行了探讨。本文以该文中数据为参考, 提取上述 3 个自然保护区和海南省的面积数据和高等植物、高等动物、保护植物、保护动物、特有植物的物种数量数据 (表 1), 共 6 个指标, 利用综合熵指数对 3 个自然保护区进行综合评价。

由于本文不考虑生物群落进化和分析地质时间尺度上的多样性^[19], 同时也为便于与李景文等^[19]的研究结果作对比, 因此, 没有包括科、属等高级分类单位的数据。

2.2 指标熵的计算方法

根据式 (3), 取各权重 $w_i=1$, 计算各指标熵 ($k=1$); 然后根据式 (5), 分别计算: 1) 取 $k=5$ (不包括面积指标) 和 $k=6$ (包括面积指标) 时; 2) 取 $k=2$ 或 3 (针对有关的行和列) 时, 3 个自然保护区的 GEI 值 (表 2)。其中, 取面积指标熵的概率计算公式为 $P_i = 1 - (x_i/X)$, 其中 x_i 为各保护区的面积, X 为海南省陆地面积。

2.3 结果与分析

计算结果如表 2 所示。当 $k=5$ (不包括面积指标熵) 时, 3 个自然保护区的 GEI 顺序为, 五指山 (0.470 0) > 霸王岭 (0.461 5) > 尖峰岭 (0.445 6); 当 $k=6$ (包括面积指标熵) 时为五指山 (0.553 1) > 霸王岭 (0.548 3) > 尖峰岭 (0.530 7)。因此, 综合来看, 在 3 个自然保护区中, 五指山自然保护区生态系统的复杂性程度最高, 具有最高的保护价值。

表 1 海南省 3 个自然保护区的基本数据集¹⁾
Tab 1 Basic data set of 3 nature reserves in Hainan, China

地区	高等植物种数	高等动物种数	重点保护植物种数	重点保护动物种数	特有植物	面积 /hm ²
尖峰岭 JFL	2220	372	27	43	238	10922.7
五指山 WZS	2146	289	25	43	223	13435.9
五指山 BWL	2194	363	30	18	140	5639
海南 HN	3150	561	93	73	505	3392000

1) 数据来源文献 [19]

表 2 海南省 3 个自然保护区的指标生态弱势
度指数和总生态弱势度指数¹⁾

Tab 2 Criterion, gross ecological disadvantage values
of 3 nature reserves in Hainan, China bits

指标	保护区	指标		k 值
		植物	动物	k=2
高等	JFL	0.3558	0.3930	0.3744
	WZS	0.3772	0.4930	0.4351
	BWL	0.3634	0.4064	0.3849
保护	JFL	0.5180	0.4498	0.4839
	WZS	0.5095	0.4498	0.4796
	BWL	0.5265	0.4981	0.5123
特有	JFL	0.5115		0.5115*
	WZS	0.5207		0.5207*
	BWL	0.5131		0.5131*
保护区	JFL	0.4618	0.4214	0.4456
	WZS	0.4691	0.4714	0.4700
	BWL	0.4677	0.4522	0.4615
k 值		k=3	k=2	k=5
面积	JFL	0.9560	0.5307	
	WZS	0.9684	0.5531	
	BWL	0.9824	0.5483	
k 值		k=6		

1) JFL, 尖峰岭; WZS, 五指山; BWL, 霸王岭; *号表示 k=1

当 k=3 (只考虑植物指标) 时, 3 个自然保护区的 GE 顺序为, 五指山 (0.4691) > 霸王岭 (0.4677) > 尖峰岭 (0.4618); 去掉特有植物指标 (k=2) 后则为, 霸王岭 (0.4450) > 五指山 (0.4434) > 尖峰岭 (0.4369)。五指山与霸王岭的顺序虽有变动, 但两者的指数值却十分接近。当 k=2 (只考虑动物指标) 时, 3 个自然保护区的 GE 顺序为, 五指山 (0.4714) > 霸王岭 (0.4522) > 尖峰岭 (0.4214)。因此, 从物种水平来看, 五指山自然保护区生态系统的动物多样性最丰富, 但其植物多样性与霸王岭接近。

3 讨论与结论

3.1 实例研究的结果及其意义

如果单从保护植物和保护动物的指标熵 (k=

2) 考虑, 霸王岭 (0.5123) > 尖峰岭 (0.4839) > 五指山 (0.4796); 但考虑面积指标 (k=3) 时则为, 霸王岭 (0.6690) > 五指山 (0.6426) > 尖峰岭 (0.6413) (这 3 个数据表中未列出)。五指山和尖峰岭 GE 大小顺序的变动, 表明面积指标在评价综合熵指数时是一个不可忽视的重要因素。同时也表明, 如果从保护珍稀濒危物种角度考虑, 则霸王岭具有最高的保护价值。李景文等^[19]主要是从物种分类角度在种、属、科等多个不同分类单位上对 3 个自然保护区的生物多样性状况进行研究; 在其综合多样性指数 T_d 中没有考虑面积指标对分析结果的影响。由于信息论在研究不确定性问题中的广泛有效性, 因此, 有理由认为, 本文基于信息熵分析的结果更具合理性。从综合角度考虑, 本文也认为五指山具有最高的保护价值。

李景文等^[19]的研究还发现, 五指山分类多样性在各分类单位上都较低, 这与实际预想有一定差距。本文研究结果表明, 从物种水平来看, 五指山自然保护区生态系统的动物多样性最丰富, 但其植物多样性与霸王岭接近。

3.2 关于综合熵指数

层次分析法在自然保护区综合评价中应用广泛, 但其基于专家打分的评价指标量化过程, 使评价结果带有一定的主观性。而且, 也没有充分利用保护区调查研究获得的大量本底数据。基于信息熵的综合定量评价方法, 能较好地解决在自然保护区评价过程中, 指标众多且标准不一、难以量化表达的问题; 同时避免了层次分析法评价过程中专家打分带来的主观影响, 因此其评价结果更可信。

综合熵指数的计算模型中不限定指标数量的多少, 解决了指标数量不统一带来的问题。对不同保护区综合评价的过程中, 若评价指标一致, 则可保证评价结果间的良好可比性。

经过多年的建设, 许多自然保护区都积累了大量的本底数据 (比如珍稀植物种数等)。本文提出基于信息熵的计算模型可充分利用前期调查得到的

本底数据, 从中挖掘更多信息, 实现保护区的综合定量评价, 为自然保护区的管理者提供决策参考。

参考文献:

- [1] 金鉴明, 王礼婧, 薛大元. 自然保护概论[M]. 北京: 中国环境科学出版社, 1991: 231—271.
- [2] 贾小容, 苏志尧, 陈北光, 等. 广东省自然保护区 DCA 排序与 UPGMA 聚类研究[J]. 华南农业大学学报: 自然科学版, 2004, 25(2): 75—79.
- [3] 马建章. 自然保护区学[M]. 哈尔滨: 东北林业大学出版社, 1992: 107—109.
- [4] 史作民, 程瑞海, 陈力, 等. 区域生态系统多样性评价方法[J]. 农村生态环境, 1996, 12(2): 1—5.
- [5] 徐慧, 钱谊, 彭补拙, 等. 鹞落坪国家级自然保护区生态评价研究[J]. 农业环境保护, 2002, 21(4): 360—364.
- [6] MARGUIES C R, USHER M B. Criteria used in assessing wildlife conservation potential: a review[J]. Biological Conservation, 1981, 21: 79—109.
- [7] SMITH P G R, THEBERGE J B. A review of criterion for evaluation natural areas[J]. Environmental Management, 1986, 10: 715—734.
- [8] 郑允文, 薛达元, 张更生. 我国自然保护区生态评价指标和评价标准[J]. 农业生态环境, 1994, 10(3): 22—25.
- [9] 赵焕臣, 许树柏, 和金生. 层次分析法[J]. 北京: 科学出版社, 1986.
- [10] KUUSIPALO J. Managing biodiversity in a forestry environment[J]. Conservation Biology, 1994, 8: 450—460.
- [11] SUTTER G W. Ecological Risk Assessment[M]. Boca Raton, FL: Lewis Publishers, 1993.
- [12] ROWE W D. Understanding uncertainty[J]. Risk Analysis, 1994, 14: 743—750.
- [13] MAGURRAN A E. Ecological Diversity and Its Measurement[M]. New Jersey: Princeton University Press, 1988.
- [14] MCINTOSH R P. An index of diversity and the relation of certain concepts to diversity[J]. Ecology, 1967, 48: 392—404.
- [15] 覃林, 余世孝. 森林群落复杂性分析: 以广东黑石顶森林为例[J]. 生物多样性, 2004, 12(3): 354—360.
- [16] 邢修三. 物理熵、信息熵及其演化方程[J]. 中国科学: A 辑, 2001, 31(1): 77—84.
- [17] 傅祖云. 信息论——基础理论与应用[M]. 北京: 电子工业出版社, 2001: 20—34.
- [18] MCLELLER J. 信息论与编码理论[M]. 2 版. 李斗, 殷悦, 罗燕, 译. 北京: 电子工业出版社, 2004.
- [19] 李景文, 李俊清, 任艳林. 海南主要热带森林生态系统类型自然保护区物种分类多样性研究[J]. 北京林业大学学报, 2002, 24(5/6): 108—114.

Quantitative Evaluation of Nature Reserves Based on Information Entropy

Wang Ben yang¹, Luo Fuli he², Yu Shixiao³

(1. College of Forestry, South China Agricultural University, Guangzhou 510642, China;

2. Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China;

3. School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China)

Abstract: Quantitative evaluation of nature reserves is beset with numerous inconsistent and hard to be quantified criteria. This paper proposes a new approach based on information entropy and defines Gross Entropy Index (GEI) as $GEI = \frac{1}{k} \sum_{i=1}^k CE_i$, and $CE_i = -w_i p_i \log p_i$, where CE_i , w_i and $p_i = x_i/X$ are the criterion entropy, the weight and the probability of the i th criterion respectively and k is currently employed number of criteria. The higher the GEI, the greater the uncertainty of a nature reserve, and thus the greater its protection value, the more urgent need of effective protection action. Three typical nature reserves in Hainan province were evaluated as a case study. The results showed that quantitative evaluation of nature reserves based on information entropy can resolve the problems induced by numerous inconsistent criteria which hard to be quantified. can be applied in the practices of nature reserve management and even regional sustainable development, and is another effective approach to quantitative ecological evaluation of nature reserves.

Key words: nature reserve, quantitative evaluation, information entropy, Hainan province