

文章编号: 1007-6301 (2002) 01-0017-08

土地利用结构综合数值表征 ——以中国西部地区为例

王秀红^{1,2}, 何书金¹, 罗 明²

(1. 中国科学院地理科学与资源研究所, 北京 100101;

2. 北京大学城市与环境学系, 北京 100871)

摘要: 应用多元统计分析中的因子分析, 以中国西部地区各县市级行政单元为例, 阐述了土地利用结构类型之间的关系, 主要因子在表征土地利用集约性上的物理意义, 以及土地利用结构的综合数值表征方法。利用 GIS 的空间分析技术, 展示了因子分析综合得分值极为丰富的结构性信息, 并分析了这些信息在宏观分类、分区中的作用。最后, 结合土地利用率与土地利用结构的信息熵, 探讨了综合数值方法在分析土地利用问题以及防治土地退化方面的应用。

关 键 词: 中国西部; 土地利用结构; 因子分析; 信息熵; GIS

中图分类号: F301.24 **文献标识码:** A

1 引言

土地合理利用战略的制定和实施, 需要人们从不同方位和层次了解土地利用结构的空
间分异特征。在常规土地利用结构和程度的表征方法中 (诸如土地利用率、农业用地指数、
垦殖率、林地利用率、森林覆盖率、建设用地指数等), 单个指标所含的结构性信息很少,
多个指标的组合才能较多地反映土地利用的结构性特征; 而这给分类和分区等定量化研究
带来许多不便。这样, 从纷繁复杂的信息源中提取主要信息, 利用较少可比性指标, 直观
表征较多土地利用结构信息的研究显得十分必要。

由于计算机技术和数学方法的不断应用, 发掘和利用综合数值方法, 探索区域宏观分
异特征的研究已经取得许多进展^[1~7]。因子分析作为降维分析的一种手段, 实际上是建立一
种从高维空间到低维空间的映射, 这种映射能保持样本在高维空间的某种“结构”; 其中最
明显的是与“排序”有关的结构^[8,9]。因此, 因子分析不仅可以研究各个指标之间的关系, 进
而进行指标归类; 而且更重要的是可用较少的可比性指标, 从独特的角度反映较多的土地
利用结构性特征。

单纯的综合数值分析, 只能对评价单元或样本进行数值排序; GIS 技术的应用, 则可以

收稿日期: 2001-12; 修订日期: 2002-01

基金项目: 国土资源部 2000 年科技专项计划资助 (200010303); 国土资源部重点项目资助
(2000301)

作者简介: 王秀红 (1964-), 男, 博士, 副研究员。从事综合自然地理和土地资源研究, 发表论著近
40 余篇 (部)。

实现样本在空间上的数值排序。这样, GIS 技术成为表征含结构性映射综合数值的重要工具或语言。换言之, 没有 GIS 技术, 综合数值分析的充分表征是不可能的。本文的目的是探索土地利用结构的综合数值表征方法, 结合西部土地生态整理工作的蓬勃发展^[10-12], 为西部土地的合理利用提供新的定量化依据。

2 资料来源与研究方法

本研究区的西部具体指陕西、甘肃、宁夏、新疆、内蒙古、青海、西藏、广西、云南、贵州、四川、重庆 12 个省、自治区和直辖市, 评价单元是根据土地利用数据与图件资料匹配状况的 974 个县级行政区。土地利用数据源于权威性的全国土地利用详查资料 (1996 年), 采用一级土地利用结构类型的面积比重, 即耕地比重 (x_1)、园地比重 (x_2)、林地比重 (x_3)、牧草地比重 (x_4)、居民点及工矿用地比重 (x_5)、交通用地比重 (x_6)、水域比重 (x_7) 和未利用地 (x_8) 比重。

本文利用因子分析方法, 对 974 个样本 (县级行政区) 所提供的前 7 大类土地利用结构类型 ($x_1 \sim x_7$) 进行分析。未利用地比重 (x_8) 的信息可由前 7 个指标计算出 (已利用土地比重隐含含有未利用土地比重的信息), 同时在变量中去除未利用地比重 (x_8), 可保证因子分析所用的相关系数矩阵行列式的值不为 0。最后, 每个样本均被赋予一个反映土地利用结构性特征的综合评价或综合得分值。

具体步骤: 在计算抽样适度测定值大于 0.5 的前提下, 利用主成分分析法提取公共因子, 根据方差累计贡献率提取特征根 (一般认为大于 70%), 选择方差最大正交旋转进行变换, 依据回归法计算因子得分, 根据特征根的贡献率大小进行加权平均计算综合得分^[8,9]。其公式为:

$$E = \lambda_1 f_1 + \lambda_2 f_2 + \dots + \lambda_n f_n$$

式中 E 为评价单元的综合得分值, f_n 为前第 n 个因子, λ_n 为前第 n 个特征根。

在计算结构性综合数值的基础上, 进一步计算土地利用结构的信息熵, 以反映土地利用的多样性^[5]。设评价单元的土地总面积为 S , 8 大类土地利用类型的面积分别为 S_1, S_2, \dots, S_8 , 则土地利用类型的面积比重 $x_i = S_i/S$, 土地利用结构的信息熵 H 为:

$$H = - \sum x_i \ln x_i$$

最后, 利用 GIS 技术对因子分析综合得分值 (分 A、B、C、D 四级)、土地利用率 (分 I、II、III 三级) 和信息熵 (分 1、2、3 三级) 3 种综合性指数的空间格局进行叠加, 并进行了类型区的划分, 以展示综合数值表征方法的实际应用。

3 结果与讨论

3.1 土地利用结构类型的归并

土地利用结构类型的归并是综合数值分析的基础。土地利用结构类型比重之间存在着程度不同的相关性, 通过因子分析, 可对结构类型做进一步的归并, 深化人们对土地利用结构类型之间“亲疏”关系的理解。因子分析所用的相关系数矩阵的行列式值为 0.050 26

(不等于 0), 说明所用矩阵不是退化的; 抽样适度测定值为 0.682 (大于 0.5), 故所用数据可以进行因子分析。由表 1 所示, 前 3 个特征根的累计贡献率已经达到 78.79% (大于 70%), 即 7 个变量反映的信息可由 3 个主成分 ($\lambda_1 + \lambda_2 + \lambda_3 = 5.5$ 个变量) 反映 78.79%; 因此, 在 7 个特征根中提取前 3 个特征根。前 3 个特征根对应的特征向量见表 2。

表 1 因子分析总方差解释
Tab. 1 Total Variance Explained

提取值	旋转提取值			因 子		
	特征根 (λ)	方差/%	累计方差/%	特征根 (λ)	方差/%	累计方差/%
1	3.054	43.629	43.629	2.733	39.042	39.042
2	1.612	23.032	66.662	1.648	23.536	62.577
3	0.849	12.128	78.790	1.135	16.213	78.790

提取方法: 主成分分析, 旋转方法: 方差最大正交旋转。

表 2 旋转后的因子成分矩阵和因子成分得分系数矩阵
Tab. 2 Rotated Component Matrix and Component Score Coefficient Matrix

变量		x_1	x_2	x_3	x_4	x_5	x_6	x_7
因子	1	0.894	0.485	- 0.223	- 0.447	0.848	0.846	0.122
	2	0.077	0.381	0.909	- 0.809	- 0.036	0.068	- 0.102
	3	- 0.034	0.336	- 0.164	- 0.043	0.261	0.124	0.953
系数	1	0.397	0.08	- 0.175	- 0.084	0.315	0.336	- 0.178
	2	- 0.07	0.224	0.597	- 0.471	- 0.102	- 0.05	0.034
	3	- 0.256	0.269	- 0.002	- 0.027	0.047	- 0.081	0.941

提取方法: 主成分分析, 旋转方法: 方差最大正交旋转

由表 2 可知, 第一因子在 x_1 、 x_2 、 x_5 、 x_6 的载荷较高, 主要是从耕地比重 (垦殖率)、园地比重、居民点与工矿用地比重以及交通用地比重 (x_5 与 x_6 之和即非农建设用地指数) 上反映土地利用的结构状况, 而且这 4 个变量对第一因子的影响都是正向的。说明现阶段总体而言, 垦殖率、园地比重和建设用地指数是同步增减的; 当然, 超过一定时间范围或具体到现阶段某地区, 一种土地利用结构类型的增加, 会制约其它土地利用结构类型的增加。第二因子在 x_3 、 x_4 的载荷较高 (以数值绝对值的大小确定载荷高低, 以正负号确定影响的方向), 主要从林地比重和牧草地比重方面反映土地利用的结构状况, 而且两个变量对第二因子的影响是反向的。第三因子在 x_7 的载荷较高, 主要从水域比重上反映土地利用结构特征。

综合比较 3 个因子的特征, 联系土地利用的生态地理条件和社会经济状况^[13, 14], 第一因子可解释为土地利用集约程度较高的因子, 反映在相对优良的生态地理条件下, 人口密度较高, 种植业、园地经营以及非农用地的建设业都得到发展, 土地利用的水平和强度较高, 相应的土地利用的经济效益较高。第二因子为土地利用集约程度较低的结构因子, 反映在较差的生态地理条件下, 人口密度较小, 种植业、园地经营以及非农用地建设业的发展受到极大的抑制, 土地利用的水平和强度较低, 但适合于林牧业的发展, 相应的土地利用的经济效益较低。第三因子可解释为土地利用集约程度很低的结构因子, 主要反映在土

地利用更为艰难的生态地理条件下, 人口密度极低, 其土地利用的经济效益极低, 如在西部占有很大面积比重的青藏高原和新疆等地区, 其水域多为现阶段难以直接利用的冰川及永久性积雪、咸水湖、沼泽等。

上述分析只表征宏观综合情况, 有三点值得重视: 第一, 就西部地区总体而言, 人类对水域的利用强度较弱, 但具体到其东南向某些人口稠密地区, 人类对水域的利用水平和强度较大。如四川的某些地区, 其生态地理条件较好, 人口密度较高, 水域利用率和利用效益也较高。第二, 3 个因子反映的生态地理条件是相互依存的, 同时有些土地利用类型的直接经济效益较低, 但其生态价值很高, 如林草地和水域的生态价值很高, 这就是现阶段提倡防止片面追求直接经济效益, 注重退耕还林、还草、还湿(地)的意义所在。第三, 这 3 个因子的物理意义可用上述方法解释, 但它们并不是几个变量的简单组合, 而是剔除重叠信息后的“提纯”, 是从 7 维空间到 3 维空间降维处理的结果, 也是数值综合的一种形式。进一步分析表 2 可知, 林地比重和牧草地比重对第一因子具有一定的反向作用; 在第一因子各变量中, 居民点与工矿用地比重以及交通用地比重与耕地比重之间存在更直接的空间相关性, 相比之下, 园地比重与耕地比重之间的关系较弱; 园地比重在 3 个因子中均有一定的载荷。

3.2 土地利用结构的综合数值表征

土地利用率, 即已利用土地占土地总面积的比重, 隐含有未利用土地比重的信息, 可视为所含结构性信息很少的综合性指数。从土地利用率值的空间分布格局可以看出, 青海西北部、新疆南部和东部以及甘肃的西北部是土地利用率低值地区, 从此低值区向北、东和南 3 个方向, 土地利用率逐渐提高。

利用因子分析则可计算出含更多结构性信息的综合数值。第一因子得分值的空间格局, 其高值区主要分布于四川盆地和渭河谷地。某些林草地比重较低的样本, 其第一因子得分值较高, 说明耕地、园地和建设用地的同步增长会占用林地和牧草地。第二因子得分值的空间格局, 其高值区主要是森林比重较高的地区, 而其低值区主要是牧草地比重较高的地区。第三因子得分值主要反映水域比重值的分布格局。其高值区只是零星分布, 中低值区主要分布于西藏, 新疆西南部, 青海湖附近, 宁夏和内蒙古境内的黄河灌区, 四川盆地以及广西东南部, 其它为大范围存在的低值区。

因子分析综合得分值的计算, 从某种意义上是做进一步的降维分析。利用 GIS 空间分析技术中的自然分割法(natural breaks)展示的综合得分值(E)空间格局, 主要反映农林牧以及未利用土地的分布状况。 E 高值区($E > 6.26$), 主要是垦殖率高值区, 相应的建设用地指数和园地比重较高; E 低值区($E < -7.31$), 主要是牧草地比重的高值区。牧草地比重高值区以东的 E 中值区, 主要是森林以及农林、农牧和林牧业混合分布区; 而牧草地比重高值区往西北向的 E 中值区, 是未利用土地比重高值区(如新疆南部中段)。由此可见, 因子分析综合得分值含有土地利用结构的主要信息, 利用 GIS 的空间分析功能, 其丰富的结构性可得到充分的展示(见图版 1 中图 1a)。

信息熵的分析则是利用综合数值表征土地利用结构类型的多样性。理论上当 8 大类土地利用结构类型的面积相等时, 土地利用的信息熵达到理论最大值。信息熵的低值区($H < 0.57$)主要出现在新疆塔里木盆地、青藏高原西北部和内蒙古东北草原区(从西部乌拉特后旗, 到东部东乌珠穆沁旗), 前者是因为土地利用率过低, 而后者是因为很高的土地利用

率主要由牧草地的比重很高造成, 两区土地利用结构类型的面积比重相差很大。信息熵很低的内蒙古东北部草原区, 也是中度水蚀风蚀区^[13~16], 是值得关注的生态脆弱区。信息熵的高值区大致与垦殖率或第一因子得分值的高值区一致 (见图版 1 中图 1b)。

3.3 综合得分值与土地利用率及信息熵之间的关系

因子分析综合得分值 (E) 随土地利用率的增大而变异增大。土地利用率在 50% 以下的 E 值较低、变幅小, 大致变幅在- 5~ 0 之间, 说明未利用土地比重很大, 样本的土地利用结构比较简单, 造成其 E 值较低、变幅小。随着土地利用率的增大, E 值的变幅增大, 当土地利用率接近 100% 时, E 值变幅于- 9~ 20 之间, 说明具有很高土地利用率的样本的土地利用结构趋向多样化。如果较高的土地利用率主要由牧草地比重较高所致, 则 E 值降低, 其它情况可能造成 E 值增高。

信息熵 (H) 与土地利用率之间的关系, 属准抛物线型。当土地利用率很低时, H 也很低, 随着土地利用率的提高, H 值逐渐增高、变幅也增大。当土地利用率达到 85% 左右时, 有些样本的 H 值很高, 接近 1.8, 而且 H 值主要变幅于 0.8~ 1.6。随着土地利用率的进一步增大, H 值有逐渐下降的总趋势, 当土地利用率接近 100% 时, 有些样本的 H 值降至最低。

H 与 E 基本上是同步增减的。 H 值在 0~ 1 范围内, E 随着 H 的增高而平缓增高, H 值在 1~ 2 范围内, E 值随着 H 值增高的速率 (斜率) 增加。上述关系说明土地利用信息熵的增高, 有利于综合得分值的提高, 或者表明因子分析综合得分值中包含土地利用结构信息熵的部分信息 (图 2)。

3.4 综合得分值的应用

因子分析综合得分值、土地利用率和信息熵 3 种综合性指数的计算, 都涉及到一级土地利用的结构类型, 它们之间存在一定的相关性, 但又不尽相同。利用 GIS 技术对 3 种综合数值的空间格局进行叠加, 可以完成关于土地利用程度及结构的类型区划分 (图 3), 有助于土地利用问题的分析。

土地利用率高值区: IA1 类型区主要分布于四川盆地西北部和关中盆地, 以种植业为主。土地利用的集约程度和利用效益也很高, 土地利用结构的多样性很大。突出的问题是由于城市化的发展和种植业的比较效益较低, 造成耕地数量和质量的下降。应走内涵挖潜的道路, 提高非农建设用地利用率以利于保护优质耕地, 同时注重防止耕地污染, 维持较高的土地利用结构信息熵。B1 类型区主要分布于广西东南和贵州东部, 以林业和农业发展为主; B2 类型区大面积分布于东部季风影响区, 主要包括林区, 农林、农牧和林牧交错区。这两类型区土地利用的集约程度和利用效益中等, 土地利用结构的多样性较大, 主要问题是森林过度砍伐和坡耕地较多, 造成严重的水土流失和生态环境恶化。应加强坡地退耕还林还草, 在保障生态安全的前提下, 提高土地利用的集约程度和利用效益。IC3 类型区主要

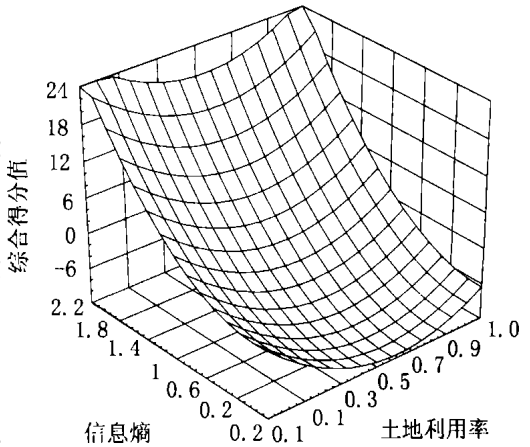


图 2 土地利用率、信息熵与因子分析综合得分值之间的关系

Fig. 2 Relationships among land use degree, information entropy and synthetic value of factor analysis

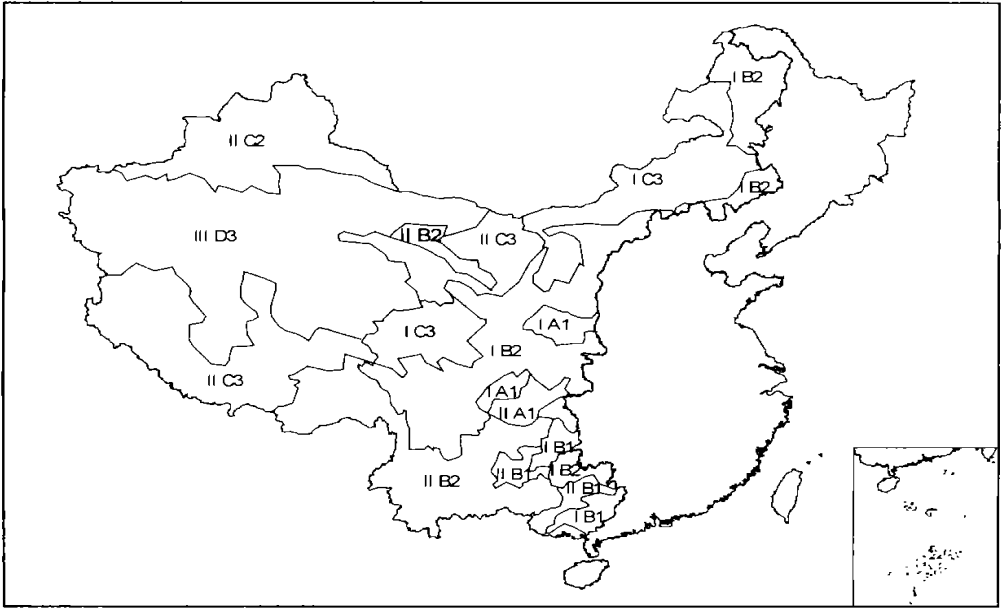


图 3 基于综合数值的土地利用程度及结构的类型区划分

Fig. 3 Division of land use degree and structure based on synthetic values

土地利用度: I (0.80~1), II (0.37~0.80), III (0.03~0.37); 综合得分值: A (6.26~20.68), B (-4.25~6.26), C (-9.18~-7.31), D (-7.31~-4.25); 信息熵: 1 (1.39~1.81), 2 (0.57~1.39), 3 (0.09~0.57)

分布于内蒙古中东部和青海省东南部,以牧业为主。土地利用的集约程度和利用效益低,土地利用结构的多样性也很小。主要问题是生态环境脆弱、草地过度放牧以及鼠虫危害,造成严重的草地退化、风蚀沙漠化和水土流失。应以草定畜,加强机械和生物灭鼠灭虫技术的应用,重视人工草地的建设,开展季节性畜牧业生产;土地利用多样性不易提高,特别不能以牺牲草地为代价增加其它土地结构类型,而应以保护生态环境,防沙治沙,维持草地的持续利用为主攻方向,生态退化严重的地段应退牧还草^[17~19]。

土地利用度中低值区: IIA 1 类型区主要分布于四川盆地东南部。主要问题类似于 IA 1 区,但土地利用度可以进一步提高。IB 1 类型区主要分布于广西中部和贵州西南部,以农业和林业为主; IB 2 主要分布于云南、西藏东南(以林业为主)和甘肃河西走廊(以农业和牧业为主)。这两个区土地利用集约型程度、利用效益以及土地利用多样性都较高,突出的问题是为扩大耕地面积而盲目劈林毁草,由此造成了严重的土地荒漠化和水土流失。应加强退耕还林还草的步伐,保护生态环境;加强农田基础设施建设,增加种植业的科技含量,提高粮食单产;在适宜的条件下,开发利用土地的后备资源。IC 2 类型区主要分布于新疆北部,以牧业为主; IC 3 类型区主要分布于内蒙古西部、青海东北部和西藏中西部,以牧业为主。两区的主要问题是生态环境脆弱,草场退化,土地沙漠化。应建立和完善牧业、农业和林业之间和谐的生态关系,在防风治沙、保护生态环境的前提下,积极发展绿洲农业和高原山地农业,缓解人口增长对草地过大的压力。IID 3 类型区主要分布于新疆南部和西藏中北部,以未利用地为主,有少许牧业和农业。主要问题是生态环境极度脆弱,土地荒漠化严重。应以保护生态环境为主,不能盲目开发未利用土地资源,特别是草地资源不

宜过度利用; 同时积极发展集约型农业, 解决人们的温饱问题^[17~19]。

4 结语

因子分析不仅可用于对土地利用结构类型进行合理归并, 而且可以提供含较多结构性映射的综合性数值。结合土地利用信息熵分析和 GIS 空间表达, 有益于土地利用结构及其表征的研究。

因子分析表明, 在我国西部土地利用结构中, 第一因子在耕地比重、园地比重、居民点及工矿用地比重和交通用地比重 4 个指标上有较大的载荷, 第二因子在林地比重和牧草地比重上有较大的载荷, 第三因子在水域比重上有较大的载荷。因子分析综合得分值的空间分布, 密切相关于土地利用集约化程度和利用效益的总体格局。

结合土地利用率和土地利用结构信息熵的空间格局, 因子分析综合得分值可以应用于土地利用结构和利用程度类型区的划分, 以及服务于土地利用问题的分析与土地退化防治对策的提出。

参考文献:

- [1] 李元 中国土地资源[M]. 北京: 中国大地出版社, 2000
- [2] 唐华俊, 陈佑启, Eric Van Ranst 中国土地资源可持续利用的理论与实践[M]. 北京: 中国农业科技出版社, 2000
- [3] 徐继填, 陈百明, 张雪芹 中国生态系统生产力区划[J]. 地理学报, 2001, 56(4): 1-8
- [4] 庄大方, 刘纪远 中国土地利用程度的区域分异模型研究[J]. 自然资源学报, 1997, 12(2): 105-111.
- [5] 陈彦光, 刘继生 城市土地利用结构和形态的定量描述: 从信息熵到分数维[J]. 地理研究, 2001, 20(2): 146-152
- [6] 龙开元, 谢炳庚, 谢光辉 景观生态破坏评价指标体系的建立方法 and 应用[J]. 山地学报, 2001, 19(1): 64-68
- [7] 胡孟春, 马荣华, 吴焕忠 海南省生态环境综合评价制图方法[J]. 地理学报, 2000, 55(4): 467-474
- [8] 于秀林, 任雪松 多元统计分析[M]. 北京: 中国统计出版社, 1999
- [9] 胡永宏, 贺思辉 综合评价方法[M]. 北京: 科学出版社, 2000
- [10] 罗明, 王军 中国土地整理的区域差异和对策[J]. 地理科学进展, 2001, 20(2): 97-103
- [11] 吴绍洪, 郑度, 杨勤业 我国西北地区生态地理区域系统与生态建设战略初步研究[J]. 地理科学进展, 2000, 20(1): 10-20
- [12] 方创琳, 徐建华 西北干旱区生态重建与人地系统优化的宏观背景及理论基础[J]. 地理科学进展, 2001, 20(1): 21-28
- [13] 毕于运 主编 中国耕地[M]. 北京: 中国农业科技出版社, 1995
- [14] 吴传钧, 郭焕成 中国土地利用[M]. 北京: 科学出版社, 1994
- [15] 全国农业区划委员会 中国农业自然资源和农业区划[M]. 北京: 农业出版社, 1989
- [16] 刘明光 中国自然地理图集[M]. 北京: 中国地图出版社 1998
- [17] Wang Xiuhong, Chapter 15: High-cold scrubs and meadow zone In: Mountain Geocology and Sustainable Development of the Tibetan Plateau[M]. Zheng Du (eds). The GeoJournal Library 57. 2000 Kluwer Academic Publishers Printed in the Netherlands, 303-325
- [18] 王秀红 青藏高原高寒草甸层带[J]. 山地学报, 1997, 15(2): 67-72
- [19] 王秀红, 申元村 柴达木盆地荒漠化及其防治[J]. 中国沙漠, 2001, 21(Supp.): 43-47.

On the Indicating Method of Synthetic Value for Land Use Structure ——a Case Study in West China

WANG Xiu-hong^{1,2}, HE Shu-jin¹, LUO Ming²

(1. Institute of Geographical Sciences and Natural Resources Research, C.A.S., Beijing 100101 China;

2. Department of Urban and Environment, Peking University, Beijing 100871 China)

Abstract: Based on detailed investigation data of land use structure in 1996, taking 974 counties or county-level cities in West China as examples, indicating method of synthetic value for land use structure was studied by using factor analysis and information entropy analysis

Factor analysis not only could be used to classify the types of land use structure, but also provided synthetic values on land use structure and information entropy. Factor analysis showed that factor 1 was closely related to the proportions of arable land, garden land, residential and factory-mining land, and traffic used land; factor 2 was determined by the proportions of forestry and grassland; while factor 3 had close relations to the proportion of water area. Containing more information, factor 1 showed that proportions of arable land, garden, residential and factory-mining land, and traffic used land were changing synchronously. However, the mentioned situation would be changed in future or in some local places at present. In other words, high proportions of garden land and construction land would reduce the proportion of arable land.

After analyzing the relationship among synthetic value of factor analysis, land use degree and information entropy, division on land use structure and degree was made, which is helpful to analyze the problems in regional land use.

Key words: West China; land use structure; factor analysis; information entropy; GIS

图版 1

王秀红：土地利用结构综合数值表征——以中国西部地区为例

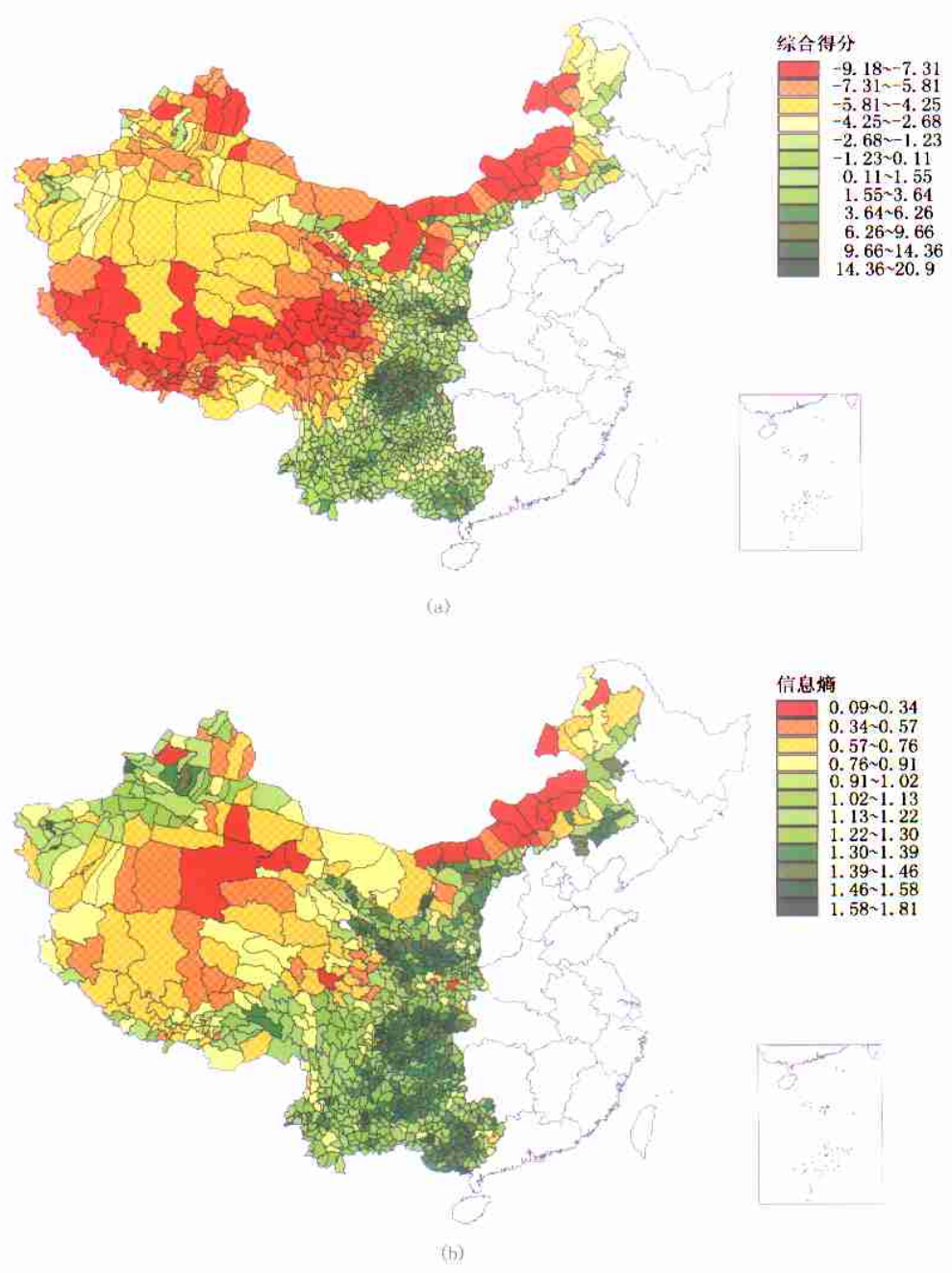


图 1 土地利用结构的综合数值表征（中国西部 1996 年）

Fig.1 Indicating method of synthetic value for land use structure (West China, 1996)