

基于典型相关分析法的县域土地利用结构研究

张晶^{1,2}, 王景平, 刘书忠 (1. 德州学院, 山东德州 253023; 2. 中国科学院地理科学与资源研究所, 北京 100101)

摘要 基于土地利用详查数据, 以全国 2 475 个县域为基本研究单元, 运用典型相关分析法, 对我国县域土地利用结构进行研究, 并运用典型相关系数和冗余度分析进行结果检验, 根据研究结果概述县域土地利用结构的主要特征及其分布规律。

关键词 县域; 土地利用结构; 典型相关分析

中图分类号 F301.2 文献标识码 A 文章编号 0517-6611(2008)01-00257-03

Research on the Land Use Structure of Counties Based on Canonical Correlation Analysis

ZHANG Jing et al (Dezhou University in Shandong Province, Dezhou, Shandong 253023)

Abstract Based on the survey data of land utilization, with 2 475 counties in China as basic research units, the canonical correlation analysis was used to study the land use structure of counties in China, and the canonical correlation coefficient and redundancy analyses were used in result-checkup. The principle characters and distribution law of land use structure of counties were summarized on the basis of research results.

Key words County area; Land use structure; Canonical correlation analysis

近年来, 随着多元统计方法在 LUCC 研究中的普遍应用^[1-2], 典型相关分析作为一种经典的多元统计分析方法, 被广泛应用于区域土地利用结构研究^[3]。典型相关分析是揭示两组多元随机变量之间相关关系的分析技术^[4]。笔者以全国 2 475 个县域为基本研究单元, 以 1996 年土地利用详查数据为基础, 试图通过定量分析我国县域土地利用结构的特征及其影响因子, 为县域土地利用结构的稳定性乃至 LUCC 的驱动力研究提供科学依据。

1 资料来源与处理

我国地域辽阔, 土地资源的利用方式、土地利用结构、土地利用程度具有明显的区域特点, 将全国划分为 12 个大区^[5], 由于全国园地、居民点及工矿用地和交通用地占土地总面积的比例均很小, 在此, 将耕地和园地合并成一类、居民点及工矿用地和交通用地合并成一类, 最终形成 6 个土地利用类别, 分别是: 耕地(包括园地)、林地、牧草地、建设用地(包括居民点及工矿用地和交通用地)、水域、未利用地。

土地利用结构的影响因子, 主要涉及自然和社会经济因子两个方面。在自然因子方面所选取的指标主要包含各个县域单元多年平均降水、温度、积温、地形等 10 项指标, 其中, 降水、温度、活动积温的数据由中国气象局数据中心提供的气象站点数据经插值生成, 地形数据由中国 1 400 万地形图经 ARC/INFO 处理获得。自然要素的名称及代码如下: A1 多年平均降水、A2 多年平均温度、A3 无霜期、A4 > 5 积温、A5 > 10 积温、B1 < 1 000 m 地貌所占比例、B2 1 000 ~ 2 000 m 地貌所占比例、B3 2 000 ~ 4 000 m 地貌所占比例、B4 4 000 ~ 6 000 m 地貌所占比例、B5 > 6 000 m 地貌所占比例。社会经济因子主要选取了人口、产值、产量、投入等 27 项因子, 主要来源于中国自然资源数据库以及各省的统计年鉴, 与土地利用结构数据相匹配, 其时间段也为 1996 年。社会经济因子的名称及代码如下: D1 总人口、D2 城市人口、D3 乡村人口、D4 劳动力人口、D5 第一产业人口比例、D6 第二产业人口比例、D7 第三产业人口比例、D8 户规模、D9 人口密度、E1 国内生产总值、E2 第一产业总值、E3 第二产业总值、E4 第三

产业总值、E5 林业总产值、E6 牧业总值、E7 渔业总值、E8 种植业产值、E9 农民人均纯收入、F1 粮食总产、F2 肉类总产、F3 猪牛羊肉总产、F4 牛奶产量、F5 禽蛋总产、G1 农机投入量、G2 化肥用量、G3 农村用电量、G4 农药用量。

在影响因子统计分析中, 为消除数据之间量纲的差异, 对各类指标采用统一的标准化方法, 先对数据进行标准差标准化, 再采用极值标准化方法, 将标准化数据压缩到 [0, 1] 闭区间内。当自变量之间存在较高的线性相关时, 会导致回归系数的置信区间变宽, 系数变得不稳定, 使其不能反映自变量的独立作用。因此研究过程中, 通过社会经济统计软件包 SPSS 对自变量进行多重共线性分析与处理, 最终各区域分别选取最适的自变量进入统计分析。

2 运行结果及检验

2.1 运行结果 运用 SPSS 中的子程序 CANCORR 对上述数据进行分析, 即可得到县域土地利用结构与自然及社会经济因子的典型相关结果。由于同一随机变量组内各典型变量间的样本协方差为零, 不同组不对应的典型变量间的样本协方差也为零。因此, 变量组与标准变量组间的关系转化为相对应的典型变量间的关系。典型载荷的统计涵义就是这种转化关系的反映, 体现出两个变量组间的相关程度。表 1 为各个地区土地利用结构典型相关分析的载荷系数, 因变量组的 A、C、F、G、W、U, 依次代表耕地、建设用地、林地、草地、水域和未利用地。在此, 自变量组仅列出主要相关因子, 其符号意义同前。

2.2 运行结果精度检验 检验典型相关分析统计重要性的最一般方法是看其典型相关系数, 即典型变量间的相关系数。表 2 给出了各个区域 6 个典型变量的典型相关系数。可以看出, 这些系数普遍较高, 尤其是前两个典型相关变量对应的相关系数值均在 0.80 以上, 表明判别出的解释变量能清晰充分地解释相应标准变量的分布。

检验典型相关分析结果的另一个方法是进行冗余度分析。一是看被解释的标准变量组的相关性被其自身典型相关变量解释的百分比, 二是看被解释变量组的典型相关性被其对立的解释变量组的典型变量解释的百分比。将二者进行对比, 能清楚地揭示目标变量组被自变量组解释的程度。

表 3 是各个地区冗余度分析结果, 从表中可以看出, 各

作者简介 张晶(1972-), 女, 山东夏津人, 在读博士, 从事资源开发和区域发展方面的研究。

收稿日期 2007-06-13

表1 典型相关分析结果

Table 1 Results of canonical correlation analysis

	地区 Region	典型变量 Typical variables					
		1	2	3	4	5	6
土地利用结构 Land Use Structure	长江流域	0.751(C)	- 0.725(A)	0.758(F)	- 0.687(W)	0.679(U)	-
	东北区	- 0.892(C)	- 0.696(A)	0.680(F)	- 0.783(G)	- 0.708(W)	0.839(U)
	横断山区	- 0.997(C)	- 0.868(A)	0.918(F)	0.542(G)	- 0.865(W)	- 0.634(U)
	黄淮海平原	- 0.985(C)	- 0.830(A)	- 0.697(F)	- 0.839(U)	0.783(W)	-
	华南区	0.741(W)	0.723(C)	- 0.727(A)	0.603(U)	- 0.557(F)	-
	黄土高原	0.945(C)	0.745(A)	0.836(F)	0.776(U)	0.595(G)	0.697(W)
	江南区	- 0.966(C)	- 0.866(A)	- 0.818(F)	0.668(U)	0.587(W)	-
	内蒙- 长城区	0.923(A)	0.863(G)	- 0.587(F)	- 0.461(C)	0.834(U)	- 0.789(W)
	青藏高原	- 0.764(A)	0.805(W)	0.774(G)	- 0.867(F)	0.799(U)	- 0.940(C)
	四川盆地	- 0.981(C)	- 0.827(G)	0.607(F)	0.485(U)	- 0.245(A)	- 0.438(W)
	西北区	0.944(W)	- 0.852(G)	0.619(F)	0.709(A)	0.695(U)	0.857(C)
	云贵高原	- 0.940(C)	- 0.897(A)	- 0.537(U)	0.773(F)	0.747(W)	0.542(G)
	自然社会经济影响因子 Socio-economic impacts of natural factors	长江流域	0.928(E1)	- 0.736(F1)	0.622(E5)	- 0.438(E7)	- 0.356(E9)
东北区		- 0.847(D3)	0.702(A2)	0.476(E1)	0.602(A1)	- 0.550(E7)	- 0.426(G1)
横断山区		- 0.955(D1)	- 0.955(F1)	0.510(A2)	0.467(A4)	- 0.687(E9)	- 0.225(D1)
黄淮海平原		- 0.918(D1)	0.707(F1)	- 0.393(E5)	0.684(A2)	0.358(E7)	-
华南区		0.632(E2)	0.690(D1)	- 0.651(F1)	- 0.475(A2)	- 0.425(E9)	-
黄土高原		0.818(D1)	- 0.711(B1)	0.440(A1)	- 0.340(A4)	- 0.491(A1)	0.456(E7)
江南区		- 0.869(D1)	- 0.901(F1)	0.702(D9)	- 0.315(E9)	0.534(E7)	-
内蒙- 长城区		0.848(F1)	- 0.704(A2)	- 0.482(B1)	- 0.453(E1)	0.407(A2)	- 0.316(E8)
青藏高原		- 0.749(E1)	- 0.536(E2)	- 0.537(A2)	- 0.383(A1)	- 0.536(D3)	- 0.524(B4)
四川盆地		- 0.926(D1)	0.820(A2)	0.730(E2)	- 0.314(E9)	0.378(D5)	0.326(E1)
西北区		- 0.446(A2)	0.658(E2)	0.374(A1)	0.356(C3)	- 0.470(A1)	0.385(D9)
云贵高原		- 0.908(D1)	- 0.817(F1)	- 0.603(E9)	0.556(E6)	0.404(A1)	- 0.538(A2)

注: 长江流域、黄淮海平原、华南与江南区因牧草地比例低于2%, 未计算, 以“-”表示。下同。

Note: The proportion of grassland in Yangtze river valley, Huang-Hai-Hi Plain, Southern China and Jiangnan region are lower than 2%, so data in these area was not calculated, and marked with "-". The same as below.

表2 典型相关系数

Table 2 Canonical correlation coefficients

地区 Region	典型变量 Typical variables					
	1	2	3	4	5	6
长江流域 Yangtze river valley	0.949	0.882	0.730	0.625	0.436	-
东北区 Northeast	0.896	0.813	0.715	0.600	0.488	0.220
横断山区 Hengduan Mountains	0.992	0.938	0.936	0.833	0.777	0.602
黄淮海平原 Huang-Hai-Hi Plain	0.971	0.905	0.756	0.535	0.478	-
华南区 Southern China	0.909	0.815	0.742	0.619	0.536	-
黄土高原 Loess plateau	0.904	0.820	0.706	0.612	0.565	0.472
江南区 Jiangnan region	0.932	0.831	0.604	0.561	0.453	-
内蒙- 长城区 Inner Mongolia-the Great Wall area	0.909	0.856	0.749	0.701	0.554	0.360
青藏高原 Qinghai-Tibet Plateau	0.955	0.948	0.797	0.693	0.467	0.394
四川盆地 Sichuan Basin	0.980	0.881	0.797	0.693	0.530	0.368
西北区 Northwest	0.868	0.853	0.567	0.478	0.447	0.367
云贵高原 Yunnan-Guizhou Plateau	0.929	0.816	0.723	0.652	0.561	0.522

对土地利用结构进行解释。

3 分析与讨论

3.1 耕地 我国粮食产量与县域单元耕地数量密切相关。长江流域、黄淮海平原地区、横断山区等7大区域的粮食产量与耕地面积典型相关, 其典型载荷系数均在0.7以上。这些区域合计县域单元占全国县域单元总数的63.87%。由此可见, 耕地保护对粮食安全具有特殊重要的意义。值得一提的是, 东北地区是我国耕地和园地分布最为广泛的区域, 占全国总面积的15.69%, 其耕地面积的分布主要受温度的影响, 典型载荷系数为0.702。黄土高原地区作为我国生态建设的重点区域, 同时也是我国粮食生产水平较低的区域, 其耕地面积的发展主要受地形的影响, 其典型载荷系数为0.711。

3.2 林地 我国县域单元林地比重和影响区域之间差异显著。东北地区是林地分布最为广泛的区域, 其林地面积占全国林地总面积的20.89%。该区林地分布与国民生产总值成正相关, 典型载荷系数为0.476, 与第一产业人口成负相关, 其典型载荷系数为-0.449, 这一方面说明东北地区林业产值在国民生产总值中占有重要地位, 但同时由于人口数量的增加, 将导致耕地、交通用地等与林地发生矛盾。此外, 云贵高原、华南区等5个区域的林业产值与农民收入水平、县域单元林地所占比例显著相关, 表明林地与区域经济因子相互作用较大。而黄土高原、西北地区、横断山区等地林业发展主要受区域温度、降水等自然因子的影响, 自然条件构成其主要影响因子。

个区域典型变量1与典型变量2的被解释变量组75%以上可以由解释变量组解释, 说明自变量组的自然-社会因子与标准变量组对应的土地利用类型存在较高的相关性, 可以用其

表3 冗余指数分析

Table 3 Analysis of redundancy index

地区 Region	解释信息 Explanatory information		被解释信息 Interpreted information		被解释百分比 Percentage of being interpreted	
	1	2	1	2	1	2
	长江流域 Yangze river valley	0.155	0.210	0.140	0.164	90.323
东北区 Northeast	0.276	0.163	0.235	0.134	85.145	82.209
横断山区 Hengduan Mountains	0.360	0.171	0.354	0.151	98.333	88.304
黄淮海平原 Huang Hai- Hi Plain	0.432	0.120	0.407	0.098	94.213	81.667
华南区 Southern China	0.320	0.299	0.264	0.235	82.500	78.595
黄土高原 Loess plateau	0.260	0.231	0.212	0.187	81.538	80.952
江南区 Jiangnan region	0.437	0.182	0.380	0.152	86.957	83.516
内蒙- 长城区 Inner Mongolia-the Great Wall area	0.318	0.185	0.263	0.146	82.704	78.919
青藏高原 Qinghai- Tibet Plateau	0.186	0.343	0.169	0.308	90.860	89.796
四川盆地 Sichuan Basin	0.500	0.252	0.481	0.195	96.200	77.381
西北区 Northwest	0.200	0.139	0.170	0.109	85.000	78.417
云贵高原 Yunnan Guizhou Plateau	0.343	0.104	0.297	0.083	86.589	79.808

3.3 牧草地 中国县域单元牧草地所占比例主要受自然因子的作用, 温度与降水构成主要影响因子。青藏高原地区是我国牧草地分布最为广泛的区域, 其草地面积占全国草地总面积的36.23%。该区草地分布主要受温度和降水的影响, 其典型载荷系数分为- 0.537、- 0.505, 可见, 自然条件是区域牧草地发展的主要限制因子。从表1可以看出, 横断山区、内蒙古及长城沿线、四川盆地、云贵高原区的牧草地所占比例均受到区域温度的影响, 而东北与黄土高原地区则主要受降水的制约。

3.4 建设用地 人口和经济增长两个因子共同构成我国县域居民和交通用地发展的基本影响因子。黄淮海平原地区是中国居民和交通建设用地分布最为广泛的区域, 占全国总面积的23.53%, 从表1可以看出, 其主要影响因子为人口, 相应典型载荷系数为- 0.918。此外, 东北地区、横断山区等8个大区居民和交通用地的面积也均是受到区域人口发展的影响, 这些区域合计的县域单元占全国县域单元总数的

80.74%, 因此可以认为, 人口是中国县域单元居民和交通用地发展的主要影响因子。长江流域和内蒙古及长城沿线的居民和交通建设用地主要受区域国民生产总值的影响, 特别是长江流域, 其典型载荷系数达到0.928, 居交用地与经济发展水平相关显著。青藏高原地区因其特殊的地形地貌特征, 人口与居民用地分布于海拔相对较低的区域。

3.5 水域 我国县域单元水域面积所占比例与社会经济因子具有一定的相关性。长江流域是我国水域面积最为广泛的区域, 其面积占全国总面积的20.15%。该区渔业产值与水域面积的分布具一定相关性, 其典型载荷系数为- 0.438。此外, 东北地区、江南区渔业产值也与水域面积所占比例呈现一定的相关。

3.6 未利用地 我国县域单元未利用地所占比例受自然与社会因子的双重影响, 区域之间差异显著。西北地区是我国未利用地占土地面积比例最高的区域, 其未利用地的分布主要受降水的影响, 典型载荷系数为- 0.47。此外, 黄淮海平原、黄土高原、内蒙及长城沿线未利用地的发展主要受区域温度的限制。由此可见, 在上述区域, 未利用地的开发受自然条件的限制, 有一定难度。四川盆地、江南区、长江流域等地未利用地的分布与农民收入水平显著负相关, 可见, 未利用地的有效开垦是农民增收的一个重要途径。东北地区未利用地的分布则与农业机械化水平成反比, 说明提高区域农业机械化水平对该区未利用地的开垦有重要意义。

4 结语

我国县域土地利用结构与自然和社会因子存在相互影响。由于幅员辽阔, 各地区自然资源条件与经济发展水平存在较大差异, 因此, 土地利用结构及其对自然和经济的作用程度有所差别。定量分析县域土地利用结构与自然和社会因子的相互关系是土地合理开发利用和区域可持续发展的有效途径。

参考文献

- [1] 李秀彬. 全球环境变化研究的核心领域——土地利用 土地覆被变化的国际研究动向[J]. 地理学报, 1996, 51(6): 553- 558.
- [2] 李晓文, 方精云, 朴世龙. 近10年来长江下游土地利用变化及其生态环境效应[J]. 地理学报, 2003, 58(5): 659- 667.
- [3] 张明. 土地利用结构及其驱动因子的统计分析——以榆林地区为例[J]. 地理科学进展, 1997, 16(4): 19- 26.
- [4] 郭志刚. 社会统计分析方法—SPSS 软件应用[M]. 北京: 中国人民大学出版社, 1999: 307.
- [5] 封志明. 一个基于土地利用详查的中国土地资源利用区划新方案[J]. 自然资源学报, 2001, 16(4): 325- 333.

(上接第182页)

由于蕾期授粉需要去处部分花器官, 也会对结实造成影响。

参考文献

- [1] 杨明琪. 三色堇栽培与品种介绍[J]. 中国花卉园艺, 2002(8): 26.
- [2] 曹焯. 三色堇制种技术[J]. 中国花卉, 2003(6): 66- 68.
- [3] 鄂利锋, 高天丽. 三色堇杂交制种技术[J]. 中国花卉园艺, 2002(14): 26.
- [4] 葛蕴智, 葛红幸. 杂交三色堇制种技术操作规程[J]. 中国种业, 2005(4): 50- 51.
- [5] 廖晓虹. 不同去雄方式的杂交效果[J]. 云南农业科技, 1998(4): 18- 19.

[6] 齐鸣. 三色堇杂交一代优势研究简报[J]. 浙江农业科学, 2003(6): 309- 312.

[7] 王贵余. 三色堇保护地杂交制种技术[J]. 中国林业, 2004(3): 53- 54.

[8] 王军卫. 三色堇杂交制种技术[J]. 北方园艺, 2003(3): 63.

[9] 周希胜, 任红云, 王敏. 三色堇种子繁殖技术[J]. 吉林蔬菜, 2002(5): 28.

[10] 张咏梅, 孙鸿举, 李翠英. “三色堇”优质高产制种技术[J]. 内蒙古农业科技, 2001(10): 20- 21.

[11] 徐艳. 三色堇的商品化生产[J]. 中国花卉园艺, 2005(2): 20.

[12] 孙玉清. 台湾三色堇的栽培及杂交制种技术[J]. 瓜果菜药花, 2002(4): 26.