

林业自然地理的模糊聚类区划

刘 羿^{1,2}, 余光辉¹, 刘安兴³, 张国江³

(1. 南京林业大学 森林资源与环境学院, 江苏 南京 210037; 2. 国家林业局 调查规划设计院, 北京 100714;
3. 浙江省森林资源监测中心, 浙江 杭州 310020)

摘要: 依照相关区划原则, 选取林业数据中的有林地、灌木林地、无林地和非林地等 4 个指标, 采用模糊聚类法对研究地浙江省范围内 67 个县级行政单位实施区划分类, 其目的是用数学方法对林业自然地理的量化因子进行数量化研究分析。根据聚类结果将全省分为 5 个类型, 其结果与研究地的自然地理状况基本相符。研究表明: 基于模糊聚类的林业自然地理区划是可行的, 进一步根据国民经济与社会发展各部门的要求, 加入相关的专业数据, 可以得到更为综合的区划结果。以行政单位作为区划单元得到的结果, 有利于各级政府的决策和对自然资源的总体把握。表 1 参 17

关键词: 森林经理学; 区划; 模糊聚类; 林地数据; 浙江省

中图分类号: S757 文献标志码: A 文章编号: 1000-5692(2008)04-0422-05

Fuzzy cluster based on nature geographic regionalization of forest

LIU Yi^{1,2}, SHE Guang-hui¹, LIU An-xing³, ZHANG Guo-jiang³

(1. College of Forest Resources and Environment, Nanjing Forestry University, Nanjing 210037, Jiangsu, China; 2. Academy of Forest Inventory and Planning, State Forestry Administration, Beijing 100714, China;
3. Zhejiang Monitoring Center of Forest Resources, Hangzhou 310020, Zhejiang, China)

Abstract: According to related regionalization principles, this research selected four indices including the area of forest land, shrubbery, non-stocked land and non-forestry land. Based on the principle of fuzzy cluster analysis, 67 counties in Zhejiang Province were clustered into 5 groups. The results basically accorded with the actual condition. The results indicated that it was feasible to regionalize by fuzzy cluster. And other data were added to accomplish some more complicated regionalization. The regionalization results were meaningful for governments to make decision and manage the nature resource when it was partitioned by administrative unit. [Ch, 1 tab. 17 ref.]

Key words: forest management; regionalization; fuzzy cluster; forest data; Zhejiang Province

自然地理区划是从地域角度出发, 研究地球表层综合体, 揭示地域分异规律, 探讨不同尺度地域划分的科学。依据自然地理地带性和区域分异规律进行的自然地域区划历来是地理学研究的重要内容。它有助于人们认识区域自然地理的规律, 而且可为拟订国家和地区的经济建设发展规划提供必要的科学依据^[1]。随着我国经济水平的迅猛发展, 可持续发展概念的深入人心, 出于经济发展、生态建设等工作的需要, 自然地理的区划显得尤为重要。对区划研究的不断深入, 区划方法也逐渐完善, 新方法与新技术在生态区划过程中得到了越来越广泛的应用^[2-4]。利用模糊聚类方法, 以林业自然资源数据探讨林业自然地理区划, 对自然地理区划作专业性的研讨。从林业专业入手, 研究自然地理因子的数量化分析, 既有理论意义, 又有实践价值。

收稿日期: 2007-08-31; 修回日期: 2007-11-06

基金项目: 国家自然科学基金资助项目(30571491)

作者简介: 刘羿, 硕士研究生, 从事森林经理研究。E-mail: liuyi_afip@163.com。通信作者: 余光辉, 教授, 博士生导师, 从事森林经理研究。E-mail: ghshe@njfu.com.cn

1 自然地理区划原则

自然地理区划的原则是自然地理的核心问题之一。不同区划因其区划方法与区划对象和目标不同, 区划原则也不同^[5]。区划原则一般依研究性质与地域状况设定。笔者采取了以下几条原则。

1.1 综合性原则和主导因素原则

陆地表面的任何自然区域都是地带性因素与非地带性因素共同作用的产物, 在自然界没有纯粹的地带性因素形成的自然区域, 也没有纯粹的非地带性因素形成的自然区域。因此, 在自然地理区划中, 应将 2 类因素结合起来进行分析。同时 2 类因素的作用常有主次之分, 在进行区划时, 应该查明自然区域形成或分异的主导因素, 以及反映此主导因素的主导指标, 科学、客观地进行自然地理区划。

1.2 资源开发和生态建设方向一致性原则

自然地理区划不仅要正确阐明自然地域分异规律, 而且要为发挥地区自然条件优势, 合理开发自然资源和进行生态建设指明方向和拟订措施。因此, 所划分的自然地理区域, 其自然资源开发和生态建设方向等应存在一定程度上的一致性, 以便于实施。

1.3 地域连续性原则

以往自然地理区划要求空间连续性原则, 这通常是指所区划的地域在空间上应是连续的, 而不应割裂开来。因而, 经常以地形地貌等的区别作为区划的主要依据之一^[6]。我国的行政体制是按省、市等管理的, 而因为上述空间连续性的原则不便于各行政单位实施。文章就是以行政单位为区划单元, 确保了行政区域的连续性。

2 自然地理区划方法

伴随科学技术的发展, 区划方法也产生了重大的转变。在先进软件的支持下, 对大量数据的定量分析成为可能, 区划方法由原先以专家分析为主的人工处理逐渐走向定量分析。如聚类分析法和地理信息系统等方法在区划过程中的单独或者综合应用^[7-9]。文章采用模糊聚类的方法, 对浙江省县级单位依据林地类型面积进行区划。为了避免各单位因地域面积不同对结果造成影响, 将林地类型转换为地域内的比重, 因而可以采用模糊聚类的方法处理数据^[10]。这里介绍模糊聚类分析的一般步骤^[11]。

2.1 数据标准化

设论域 $U = \{x_1, x_2, \dots, x_n\}$ 为被分类的对象, 每个对象又由 m 个指标表示其性状, 即:

$$x_i = (x_{i1}, x_{i2}, \dots, x_{im}) \quad (i = 1, 2, \dots, n) \quad (1)$$

于是得到原始数据矩阵为:

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix}。$$

不同指标通常具有不同量纲, 为了不同量纲下的量进行比较, 通常需要对数据作变换。数据标准化的结果应该使得到的数据矩阵中的数据均在区间 $[0, 1]$ 上, 即满足模糊矩阵的要求。

2.2 建立模糊相似矩阵

设论域 $U = \{x_1, x_2, \dots, x_n\}$, $x_i = (x_{i1}, x_{i2}, \dots, x_{im})$, 依照传统聚类方法确定相似系数。 x_i 与 x_j 的相似程度 $r_{ij} = R(x_i, x_j)$ 。可以选择的方法有很多, 文章采用的是相关系数法, 得到了较好的聚类结果。

$$r_{ij} = \frac{\sum_{k=1}^m |x_{ik} - \bar{x}_i| |x_{jk} - \bar{x}_j|}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (2)$$

式(2)中, $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$, $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$ 。

2.3 动态聚类

数据标准化得到了模糊相似矩阵 R 。为了得到具有传递性的模糊等价矩阵, 可以通过传递闭包法由 R 得到模糊等价矩阵 R' 。得到了 R 的传递闭包 $t(R)$, 即模糊等价矩阵 R' 后, 通过从大到小选取的值, 可以形成动态聚类图^[12]。

2.4 确定最佳阈值

对于不同的阈值, 可以得到不同的分类, 从而形成动态聚类图。对于如何选取阈值, 从数理角度看, 一般用 F 统计量来确定 的最佳值。设论域 $U = \{x_1, x_2, \dots, x_n\}$ 为样本空间(样本总数为 n), 而每个样本 x_i 有 m 个特征: $x_i = (x_{i1}, x_{i2}, \dots, x_{im}) (i = 1, 2, \dots, n)$ 。

记 \bar{x}_k 为总体样本的中心向量:

$$\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}, \quad k = 1, 2, \dots, m. \quad (3)$$

设对应于 值得分类数为 r , 第 j 类的样本数为 n_j , 第 j 类的样本记为 $x_1^{(j)}, x_2^{(j)}, \dots, x_{n_j}^{(j)}$, 第 j 类的聚类中心为向量 $\bar{x}_1^{(j)}, \bar{x}_2^{(j)}, \dots, \bar{x}_m^{(j)}$, 其中 $\bar{x}_k^{(j)}$ 为第 k 个特征值的平均值:

$$\bar{x}_k^{(j)} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ik}^{(j)}, \quad k = 1, 2, \dots, m. \quad (4)$$

作 F 统计量:

$$F = \frac{\sum_{j=1}^r n_j (\bar{x}^{(j)} - \bar{x})^2 / (r - 1)}{\sum_{j=1}^r \sum_{i=1}^{n_j} (x_i^{(j)} - \bar{x}^{(j)})^2 / (n - r)}, \quad (5)$$

式(5)中, $\bar{x}^{(j)} - \bar{x} = \sqrt{\sum_{k=1}^m (\bar{x}_k^{(j)} - \bar{x}_k)^2}$ 为 $\bar{x}^{(j)}$ 与 \bar{x} 的距离; $x_i^{(j)} - \bar{x}^{(j)}$ 为第 j 类中样本 $x_i^{(j)}$ 与中心 $\bar{x}^{(j)}$ 的距离。式(5)称为 F 统计量, $F \sim F(r - 1, n - r)$ 。如果 $F > F_{\alpha}(r - 1, n - r)$ ($\alpha = 0.05$), 根据数理统计理论可判断类与类之间的差异是显著的。如果满足条件的 F 不止一个, 可以选取其中较大的 F 值。

3 研究地概况与数据来源

3.1 研究地概况

浙江省地处我国东南沿海, 长江三角洲南翼, 介于 $27^{\circ}06' \sim 31^{\circ}11' N$, $118^{\circ}01' \sim 123^{\circ}10' E$ 。北部以杭嘉湖平原与太湖南岸、江苏、上海相接, 西北以白际山、天目山与安徽接壤, 西南部又分别以怀玉山、仙霞岭、洞宫山与江西和福建毗邻, 东部濒临东海。全省陆域面积为 10.18 万 km^2 , 约占我国国土面积的 1% 。东西宽与南北长相仿, 都在 450 km 左右。

浙江省地形地势为西南高, 东北低, 自西南向东北倾斜, 呈阶梯状下降。在全国地貌区划上属东南沿海低山丘陵、华中华东低山丘陵及江浙冲积平原的一部分, 以分割破碎的低山和丘陵为主。全省海拔 1000 m 以上的高山多分布在西南部和南部, 主要山峰海拔在 1500 m 以上。而中部则以丘陵为主, 其间大小盆地错落分布, 河谷较宽, 大小不等的河谷平原沿河分布。北部则是以钱塘江下游冲积平原与太湖流域的湖积平原、滨湖平原所组成的杭嘉湖平原为主, 地势平坦, 海拔为 $3 \sim 5$ m , 其间河网密布。另外, 在东部滨海的狭长地带, 零星分布着一些滨海平原, 常分布在入海河口的两侧及海湾内。沿海岛屿众多, 大于 500 m^2 的岛屿有 3061 个。全省以山地丘陵为主, 200 m 以上的山地丘陵总面积占全省面积的 70.4% , 平原面积占 23.2% , 河湖水域占 6.4% , 有“七山一水二分田”之说。

3.2 数据来源与处理

文章采用浙江省森林资源“九五”清查统计数据^[13], 对全省各县级单位进行区划。在数据处理过程中, 由于部分行政区划的调整以及部分县的处理便利, 对数据作了一些处理。其中, 萧山区和余杭区从杭州市区中分出; 鄞州区从宁波市区中分离; 玉环县并入温岭处理; 舟山以地级市处理。

通常原始数据不能满足数据分析方法的要求, 需要进行处理后才可以使用。在处理过程中, 考虑到乔木林、竹林和经济林均属于有林地, 将该 3 项合并成为 1 项指标, 与灌木林地、无林地、非林地等 3 项作为模糊聚类的指标。又模糊聚类方法中的模糊矩阵要求各元素均落在区间, 将有林地、灌木林地、无林地和非林地等的面积换算成为在各自县级单位总面积中的比例, 以其占总面积中的比例数值作为模糊聚类的指标。

4 聚类结果

数据处理后得到的数值为各县级行政单位林地类型的比例值, 并依此得到待分析的模糊矩阵。采用相关系数法得到动态聚类结果。文中选择 $r = 0.99573$, 经检验, 取 $r = 0.99573$ 时, 类间差异是显著的。如果 r 取值更小, 分类数将激增, 结果意义就不大了。文章将全部单位聚类成为 5 个类。类型一: 建德市, 富阳市, 临安市, 桐庐县, 淳安县, 永嘉县, 文成县, 泰顺县, 安吉县, 东阳市, 武义县, 浦江县, 磐安县, 衢州市区, 江山市, 常山县, 开化县, 丽水市区, 龙泉市, 青田县, 庆元县, 缙云县, 遂昌县, 松阳县, 云和县, 景宁县, 仙居县, 天台县。类型二: 诸暨市, 奉化市, 嵊州市, 新昌县, 临海市, 三门县。类型三: 宁海县, 乐清市, 平阳县, 苍南县, 金华市区, 兰溪市, 永康市, 龙游县, 台州市区。类型四: 余杭区, 余姚市, 象山县, 鄞州区, 温州市区, 瑞安市, 德清县, 长兴县, 绍兴市区, 上虞市, 义乌市, 舟山市, 温岭市。类型五: 杭州市区(不含萧山区、余杭区), 萧山区, 宁波市区(不含鄞州区), 慈溪市, 嘉兴市区, 海宁市, 平湖市, 桐乡市, 海盐县, 嘉善县, 湖州市区。

对各类类中数据进行平均, 结果可见表 1。为了能够对该聚类结果有地理位置上直观的显示, 可将聚类结果反映在行政区域图上。

表 1 聚类类型权重数据

Table 1 The weights of clustering types

聚类类型	有林地	灌木林地	无林地	非林地
类型一	0.677 468	0.057 819	0.042 253	0.222 460
类型二	0.591 595	0.026 649	0.027 769	0.353 987
类型三	0.510 379	0.022 612	0.030 968	0.436 041
类型四	0.409 884	0.023 413	0.025 172	0.541 531
类型五	0.175 947	0.009 909	0.006 155	0.807 989

5 讨论

5.1 指标选取合理性的讨论

文章对浙江省全省进行了数量化自然地理区划的分析, 选取的指标为该省森林资源“九五”清查统计数据中林地面积, 指标分别为有林地、灌木林地、无林地和非林地。选取其为指标的原因有三^[14-16]。第一, 森林作为陆上最大的生态系统, 林地面积状况基本能够反映地区自然地理状况。第二, 森林多位于山区地带, 而平原地区多为经济发达地区, 该指标一定程度上能做到与经济发展方向一致。第三, 森林系统在生态建设和环境保护中的作用日渐突出, 以该指标进行自然地理区划也能为该省生态建设指明方向。而将省内各县级单位作为区划的单元, 不同于以往按地块类型的区划。按地块类型的区划固然合理, 以行政单位区划也有有利的一面。由于我国的行政体制是按各级行政单位管理的, 因此, 以行政单位区划对于各单位的决策与全局的控制都能起到积极的影响。

5.2 聚类结果与数据相关性分析

结果显示, 浙江省各县(市、区)的自然地理概况, 以林业专业的土地利用现状, 可以分成五大类。由表 1 的土地和用地比重清晰可见, 从类型一至类型五, 有林地面积权重由大至小, 而非林地面积权重由小趋大, 两者成负相关; 其中的灌木林地和无林地, 其基本变化趋势与有林地相似, 但其内部有跳跃现象, 即类型三的灌木林地权重小于类型四, 类型二的无林地权重小于类型三。由于灌木林地和无林地的权重数量级小于有林地和非林地, 因此其变化基本上从属于有林地权重的变化。这主要与山区林业用地的相关性或一致性有关, 同时启示我们, 在数量化分析方面数据量纲应基本保持相近。

浙江省 5 个类型区的面积权重与聚类图的权重对照可见, 在有林地面积权重中, 5 个类型区的类型一至类型四, 其相邻之间的差值大致为 0.08 ~ 0.09, 而类型四与类型五之间的差值达 0.23 以上,

显示在聚类图上的结果是类型一至类型四归为一大类后再与类型五相聚;同理,类型一与类型二的聚类结果显示,类型一与类型二的差异除了与其他类型的差异外,其非林地的面积权重达到0.12,因此,在聚类图上也有显现。从数据的平均权重和聚类图的相关对应分析,可进一步揭示其各类型区乃至类型区内部相关规律。

5.3 聚类结果与研究地自然地理格局分析

浙江省5个类型区,其中3个类型区的有林地面积权重达到0.50以上,2个类型区在0.50以下,说明浙江省大多数县(市、区)属山地丘陵地形地貌,在林业专业用地分析上显示为林业用地,其结果与浙江省“七山一水二分田”的地貌格局相类似。因此,在类型一至类型三和县(市、区),为山区县;类型四的林业用地面积权重和非林业用地面积权重大致相等,该类型区为部分山区县和部分平原县,其中林业用地面积权重大于非林业地的称为部分平原县,而非林地面积权重大于林业用地的称为部分山区县;类型五是平原县。

5.4 进一步研究任务与方向

本研究主要采用林业专业的相关数据对自然地理作数量化的分析区划,其结果大致与研究地的自然地理特征一致,分区的结果是可行的^[17]。本研究主要在于方法的研讨,在此基础上,可根据国民经济和社会发展各相关部分的区划要求,加入相关专业的数据,形成不同专业的分区结果及综合性的分区结果,诸如我国目前正在展开的一些功能区区划和生态功能区划等,以便为相关部门的决策服务。

值得注意的是,基于模糊聚类的林业自然地理区划研究,仅仅是一种数量化方法的研究,由于侧重于量化数据,且这些数据的挖掘仅仅是有代表性的一部分而非全部,因此,其结果也有明显的局限性,诸如本文聚类结果中,同类中的“类型三”“类型四”等在地域上存在着交错与不相连的状况,与区划原则中的“地域连续性原则”不相符。类似的问题,需要进一步综合其他的数据加以研究。采用类似于这种数据挖掘的分析研究,可以克服研究人员在专业知识上的缺乏,或对专家决策系统加以辅佐,以使相关的区划研究成果更趋于科学性。

参考文献:

- [1] 倪绍祥. 中国综合自然地理区划新探[J]. 南京大学学报: 自然科学版, 1994, 30(4): 706 - 714.
- [2] 程叶青, 张平宇. 生态地理区划研究进展[J]. 生态学报, 2006, 26(10): 3 424 - 3 433.
- [3] KLIJIN F, UDO DE HASE H A. A hierarchical approach to ecosystem and its implications for ecological land classification[J]. Landscape Ecol, 1994, 9(2): 89 - 104.
- [4] HOST G E, POLZER P L, MLADENOFF D J, et al. A quantitative approach to developing regional ecosystem classifications[J]. Ecol Appl, 1996, 6(2): 608 - 618.
- [5] 杨勤业, 李双成. 中国生态地域划分的若干问题[J]. 生态学报, 1999, 19(5): 596 - 601.
- [6] 郑度, 葛全胜, 张雪芹, 等. 中国区划工作的回顾与展望[J]. 地理研究, 2005, 24(3): 330 - 337.
- [7] 王军辉, 顾万春, 夏良放, 等. 桉木种源的地理变异和种源区划[J]. 浙江林学院学报, 2005, 22(5): 502 - 506.
- [8] 李秀萍, 杨德刚, 韩剑萍, 等. 应用主成分分析、聚类分析划分新疆绿洲生态经济类型的初步研究[J]. 干旱区地理, 2002, 25(3): 264 - 270.
- [9] 洪军. 基于GIS的森林分类经营区划[J]. 东北林业大学学报, 2002, 30(4): 14 - 18.
- [10] 王志良, 邱林, 梁川, 等. 基于模糊聚类分析的水资源分区研究[J]. 华北水利水电学院学报, 2001, 22(2): 7 - 10.
- [11] 付强. 数据处理方法及其农业应用[M]. 北京: 科学出版社, 2006: 105 - 110.
- [12] 谢季坚, 刘承平. 模糊数学方法及其应用[M]. 3版. 武汉: 华中科技大学出版社, 2006: 85 - 88.
- [13] 浙江省林业局. 浙江省林业自然资源: 森林卷[M]. 北京: 中国农业科学技术出版社, 2002: 271 - 276.
- [14] 陈立人, 刘卫. 浙江省经济区划的综合性分析与协调发展研究[J]. 忻州师范学院学报, 2004, 20(5): 74 - 78.
- [15] 俞洁, 邵卫伟, 于海燕, 等. 浙江省生态功能区划研究[J]. 环境污染与防治, 2006, 28(8): 620 - 623.
- [16] 陶正明, 顾雪萍, 钱奇霞. 江西省吉水县天然林区域类型划分[J]. 浙江林学院学报, 2005, 22(2): 144 - 150.
- [17] 翟中齐. 中国林业地理概论: 布局与区划理论[M]. 北京: 中国林业出版社, 2003: 18 - 50.