

应用因子分析法对林分蓄积相关因子的分析

王京民, 李清顺

(国家林业局 西北林业调查规划设计院, 陕西 西安 710048)

摘要:利用 2006 年新疆阿勒泰山国有林管理局经营区森林资源连续清查 2 493 块样地数据, 从中提取落叶松样地 261 块, 通过因子分析法对与林分蓄积相关的因子进行分析。采用主成分分析法, 提取特征值大于 1 的 3 个公因子, 累积方差贡献率 64.38%。结果表明: 第一公因子为平均胸径、平均树高、平均年龄, 其定义为基本林分因子, 其方差贡献率为 32.89%; 第二公因子为郁闭度和样木总株数, 定义为林分密度因子, 其方差贡献率为 18.16%; 第三公因子为海拔和坡向, 定义为立地条件因子, 其方差贡献率为 13.33%。最后把样地分为建模样本和验证样本, 进行了多元回归并对其进行了验证, 结果与因子分析的结果一致。

关键词:林分蓄积; 森林资源连续清查; 因子分析; 多元线性回归

中图分类号:S758.5 **文献标志码:**A **文章编号:**1001-7461(2012)01-0169-04

Analysis of Associating Factor with Stand Volume Using Factor Analysis

WANG Jing-min, LI Qing-shun

(Northwest Institute of Forest Inventory Planning and Design, SFA, Xian, Shaanxi 710048, China)

Abstract: Data of 261 plarch plots were extracted from those of 2 493 fixed sample spots during the continuous forest inventory in the year of 2006 carried out by Altay Administration Bureau of State-owned Forests in Xinjiang, China. Factors related to stand volume were analyzed. Principal component analysis was adopted in the estimation of the volume. Three common factors with the eigenvalue more than 1 selected, and the cumulative variance contrubution was 64.38%. The first common factor was composed of the average breast diameter, average height and averge age, which was defined as the basic stand factor with variance contribution of 32.89%. The second common factor was composed of canopy density and the number of sampling trees, which was defined as stand density factor with the variance contribution of 18.16%. The third common factor was composed of altitute and slope, which was defined as the site condition factor with the variance contribution of 13.33%. Finally, the sample plots were divided into modeling and validation samples for the multiple regression and validation. The results of validation were consistent with the factor analysis.

Key words: stand volume; continuous forest inventory; factor analysis; multiple linear regression

蓄积是衡量森林数量的主要指标, 单位面积蓄积的大小标志着林地生产力的高低, 也反映出经营措施的效果。另外在森林资源中, 经济利用价值最大的仍然是木材资源。因此, 林分蓄积量的测定是林分调查的主要目的之一, 它为森林经营和采伐利用提供重要的数量依据。林分蓄积和单株材积一

样, 是由断面积、树高和形数三要素构成。但它与单株材积又有区别, 因为林分是由树木的群体组成, 它具有生长、积累的过程。因此, 它受林木直径、树高、形数和株数的制约, 并受树种、年龄、立地条件和经营措施的影响^[1]。本文利用 2006 年阿勒泰山国有林管理局经营区森林资源连续清查 2 493 块样地中

的 261 块落叶松样地,通过因子分析法^[2-10],定性的找出了与林分蓄积相关的因子,并对这些因子进行了分类及命名。

1 研究区概况与数据预处理

1.1 研究区概况

阿尔泰山国有林管理局经营区位于新疆维吾尔自治区北部,地处阿尔泰山中东段,北与俄罗斯接壤,西北、东北分别与哈萨克斯坦和蒙古毗邻。地理平均海拔介于 1 000~3 500 m 之间,主峰—友谊峰海拔 4 374 m。年降水量随海拔升高而增加,自然降水主要集中在 6—8 月份,降雪期 7—8 个月,积雪深厚。阿尔泰林区很多地方人迹罕至,人为干扰特别小,乔木林多为自然生长状态下的原始生态林。该地区属于寒温带针叶林区,森林呈块状、带状等不连续的间断分布。主要组成树种是西伯利亚落叶松,其次是云杉。

1.2 数据来源

数据采用 2006 年阿尔泰山国有林管理局经营区森林资源连续清查 2 493 块标准地数据,标准地面积为 0.08 hm²。为了研究方便,从中提取有林地 331 块(剔除了 5 块有明显记录错误的样地),其中落叶松(*Larix gmeini*)261 块,云杉(*Picea asperata*)38 块,桦木(*Betula* spp)12 块,冷杉(*Abies fabri*)10 块,杨树(*Populus*)2 块,红松(*Pinus koraiensis*)2 块以及山杨(*Populus davidiana*)2 块。331 块样地的林木平均年龄为 35~270 a,海拔 1 000~2 640 m,平均树高 1.6~36.2 m,林分蓄积 0.679~53.100 m³,林木总株数 1~241 株。

1.3 数据预处理

与蓄积相关的因子包括优势树种、平均胸径、平均树高、郁闭度、样木总株数、平均年龄(龄组)、地类、立地条件(海拔、坡向、坡位、土壤厚度)等^[11]。优势树种类型对森林蓄积影响较大,但都是一个定性指标,因此只能按优势树种进行分类,然后再分别进行因子分析。由于其它树种样地数量的限制,只能选择落叶松作为因子分析的目标,也就是说研究对象为 261 块落叶松样地。对于平均年龄和龄组,由于其内涵是一致的,为了更为准确的反应与蓄积的关系,选择平均年龄作为因子分析的载荷。

根据国家森林资源连续清查技术规程,坡向为样地范围内的地面朝向,共分为 9 类,具体分类及各类的方位角如表 1 所示。在进行因子分析时,将坡向分为阴坡、阳坡和无坡向三类。将东坡、北坡、东北坡及西北坡划归阴坡,将西坡、南坡、东南坡及西

南坡划归为阳坡,无坡向单独一类,分别赋值为 2、1、0,将其量化。

研究因子有平均胸径、平均树高、郁闭度、样木总株数、平均年龄、海拔、坡位、坡向、土层厚度共 9 个因子。

表 1 森林资源调查样地坡向分类
Table 1 Category of slope in forest inventory

编号	坡向	方位角/°	估测坡向
1	北坡	337.5~22.5	阴坡
2	东北坡	22.5~67.5	阴坡
3	东坡	67.5~112.5	阴坡
4	东南坡	112.5~157.5	阳坡
5	南坡	112.5~202.5	阳坡
6	西南坡	112.5~247.5	阳坡
7	西坡	247.5~292.5	阳坡
8	西北坡	292.5~337.5	阴坡
9	无坡向	坡度<5	无坡向

2 利用因子分析法进行分析

2.1 因子分析法的数学模型^[12]

因子分析法的核心是用较少的相互独立的因子反映原有变量的绝大部分信息,可以将这一思想用数学模型来表示。设原有 p 个变量 $X=(x_1, x_2, \dots, x_p)$,其均值为 $E(X)=0$ 。将每个原有变量用 k ($k < p$)个因子 f_1, f_2, \dots, f_k 的线性组合来表示,即

$$\begin{cases} x_1=a_{11}f_1+a_{12}f_2+a_{13}f_3+\dots+a_{1k}f_k+\epsilon_1 \\ x_2=a_{21}f_1+a_{22}f_2+a_{23}f_3+\dots+a_{2k}f_k+\epsilon_2 \\ \dots\dots\dots \\ x_p=a_{p1}f_1+a_{p2}f_2+a_{p3}f_3+\dots+a_{pk}f_k+\epsilon_p \end{cases}$$

上式就是因子分析的数学模型,用矩阵的形式表示为 $X=AF+\epsilon$ 。其中 F 称为公共因子, A 为因子载荷阵, $a_{ij}=(i=1,2,\dots,p;j=1,2,\dots,k)$ 称为因子载荷; ϵ 为特殊因子(随机误差),表示原有变量不能被解释的部分,相互独立,服从正态分布,均值为 0。

2.2 因子分析

在对原始数据进行标准化处理的基础上,利用 SPSS 软件作因子分析。采用主成分分析法选取特征值大于 1 的 3 个因子作公因子,公因子的 KMO 检验值为 0.673,累积方差贡献率为 64.38%。由于公因子含义不清,运用最大方差法对因子载荷矩阵进行正交旋转,得到公共因子的载荷矩阵,并采用回归法得到因子得分系数矩阵(表 2)。

第一公因子对方差的贡献率为 32.89%,它在平均胸径、平均树高、平均年龄(带方框的数据)上有较大载荷。林分蓄积是由断面面积、树高和形数三要素来决定的,而断面面积与平均胸径成正比关系,因此该因子包含了林分蓄积三要素中的两个,定义为基

本林分因子;第二公因子在郁闭度和样木总株数上有较大载荷,其方差贡献率为 18.16%,定义为林分密度因子;第三公因子在海拔、坡向上有较大载荷,其方差贡献率为 13.33%,定义为立地条件因子。

表 2 旋转后的因子载荷、特征值、贡献率、累积贡献率和因子得分系数

变量	因子载荷			因子得分		
	F_1	F_2	F_3	F_1	F_2	F_3
平均胸径	0.80	-0.41	-0.06	0.30	-0.09	0.00
平均树高	0.82	0.06	-0.21	0.39	0.18	-0.07
郁闭度	-0.22	0.84	-0.03	0.05	0.45	0.02
样木总株数	-0.45	0.79	-0.03	-0.07	0.38	-0.01
海拔	-0.05	-0.12	0.75	0.02	-0.03	0.52
坡位	0.09	0.01	-0.67	-0.02	-0.02	-0.47
坡向	0.15	0.44	0.50	0.20	0.32	0.40
平均年龄	0.81	-0.20	0.14	0.37	0.04	0.16
土层厚度	0.37	0.45	-0.34	0.24	0.31	-0.17
特征值	2.96	1.63	1.20			
方差贡献率/%	32.89	18.16	13.33			
累积贡献率/%	32.89	51.05	64.38			

3 分析结果的验证

3.1 样本分类与多元线性回归

把 261 块落叶松样地分为建模样本和验证样本两部分。把 231 块落叶松样地作为建模样本,余 30 块作为验证样本。以林分蓄积为因变量,公因子为自变量,采用多元线性回归方法^[13],分别对第一公因子、第一、二公因子、第一、二、三公因子建模。在对第一、二、三公因子建模时,发现坡向的 t 检验值为 0.343,在任何水平下都不显著,因此剔除坡向。这一点从坡向的因子载荷为 0.5 上也可以看出。最后得到的结果列表 3。

表 3 多元线性回归结果

参加建模因子	残差平方和	决定系数	F 检验值
第一公因子	12 056.65	0.30	32.09
第一二公因子	7 887.71	0.54	52.95
第一二三公因子	7 618.48	0.56	46.80

从表 3 可以看出,从第一公因子到第一二公因子,残差平方和明显下降,决定系数、 F 检验值都明显提高,而从第一二公因子到第一二三公因子,残差平方和小幅下降,决定系数少量提高, F 检验值小幅下降。这与因子分析的结果也是一致的。

3.2 回归模型的验证

在上述多元回归中,得到三个回归方程:
 $V_1 = -0.144D + 0.833H + 0.050A - 1.66$ (1)
 $V_2 = 0.108D + 0.479H + 0.0548A + 0.0191N + 25.6Y - 15.8$ (2)
 $V_3 = 0.113D + 0.528H + 0.0519A + 0.0252N + 24.9Y + 0.00379A_1 - 23.9$ (3)

其中 V 、 D 、 H 、 A 、 N 、 Y 、 A_1 分别代表林分蓄积、平均胸径、平均树高、平均年龄、郁闭度、样木总株数和海拔(坡向已经剔除)。

采用成对 t 检验法^[14]对上述 3 个方程得到的结果进行验证,结果如表 4:

表 4 成对 t 检验结果

检验对	均值	标准差	t 值
林分蓄积- V_1	3.58	7.71	2.54*
林分蓄积- V_2	2.97	7.45	2.18*
林分蓄积- V_3	2.33	7.09	1.80

查 t 分布表得 $t_{0.05}(29) = 2.045$,从以上的结果看,方程(1)、(2)得到的结果经成对检验后 t 值均大于 2.045,说明回归值与实际调查值之间存在差异(95%),但是由表 3 可以看出第一公因子的决定系数仅有 0.3,第一二公因子的决定系数为 0.54,说明林分蓄积与第一公因子有关,随着林分密度因子的加入,相关性不断提高,第一二公因子无法很好的估计林分蓄积,正是林分蓄积复杂性的体现。方程(3)得到的 t 值小于 2.045,说明回归值与实际调查值无明显差异,但是提高并不很大。进一步看,均值、标准差和 t 检验值随着预报因子的增加而不断减小,说明显著性在减小,即预测值和实际调查值之间的差异越来越小,这些都与因子分析的结果相一致。

另外,考察方程(1),发现平均胸径的系数出现了负值,这与现实是不符的,原因是因子间出现了多重共线性^[15]。平均胸径、平均树高和平均年龄的相关系数得到 $R_{DY} = 0.685$, $R_{HY} = 0.573$,由此可以看出出现负值的原因是因为平均胸径和平均树高都与平均年龄有较强的相关性,平均胸径对材积的作用

通过了平均树高或平均年龄这 2 个因子去影响材积。所以有可能在回归方程中平均胸径的系数会出现负值。关于多重共线性及其处理方法不在本文的讨论范围,有兴趣的可参考相关资料和文献。

4 结论

1) 通过对阿勒泰山国有林管理局 2006 年森林资源连续清查样地数据进行因子分析,找出了与林分蓄积相关的三类因子,并把其定义为基本林分因子(平均胸径、平均树高和平均年龄)、林分密度因子(郁闭度和样木总株数)和立地条件因子(海拔和坡向)。这三类因子的提取与林分蓄积的三要素(断面积、平均树高、形数)是基本一致的。

2) 把 261 块落叶松样地分为建模样本(231 块)和验证样本(30 块)两类。对 231 块建模样本,把第一因子,第一、二公因子,第一、二、三公因子分别作为自变量,林分蓄积作为因变量分别进行多元回归建模(由于坡向的 t 值偏小被剔除)。结果显示从第一公因子到第一、二公因子,残差平方和明显下降,决定系数、 F 检验值都明显的提高,而从第一、二公因子到第一、二、三公因子,残差平方和小幅下降,决定系数少量提高, F 检验值小幅下降,这与因子分析的结果是一致的。对 30 块验证样本,我们采用成对 t 检验方法对回归方程进行了验证。结果显示均值、标准差和 t 检验值都随因子的增加而减小,这也与因子分析的结果相一致。综合以上结果,与林分蓄积相关的因子包括平均胸径、平均树高、平均年龄、郁闭度、样木总株数和海拔等。

3) 从以上的因子分析及其验证中找出了与林分蓄积相关的因子,这也与林分蓄积的三要素基本吻合。但是由于阿勒泰地区地理位置的特殊性(人迹罕至,人为干扰特别小),无需考虑经营类型及人为因素的影响(采伐、抚育等),但在其他地区,这些因素的影响都是显著的,必须加以考虑。

参考文献:

[1] 孟宪宇,余光辉、李凤日,等. 测树学[M]. 北京:中国林业出版

社,2005.

[2] 李卫东. 应用多元统计分析[M]. 北京:北京大学出版社,2008.

[3] RICHARD A J,DEAN W W. Applied multivariate statistical analysis[M]. 北京:清华大学出版社,2008.

[4] REYNOLDS C R,BIGLER E D. Standardized factor scores for use with the test of memory and learning, archive of clinical neuropsychology[J]. Elsevier Science,1995;10(4):382-383.

[5] YIN S C,KENNY P. Experiments in speaker adaption for factor analysis based speaker verification [J]. IEEE,2006;1-4244-0469-X;897-900.

[6] 吴铁群. 基于因子分析法的民营企业绩效体系构建及指标矩阵研究. [D]. 北京:华北电力大学,2008.

[7] 王增民. 因子分析法在企业经济效益的综合分析与评价中的应用. [J]. 数理统计与管理,2002,21(1):10-23.

WANG Z M. A synthetic evaluation on the economic benefit of enterprises -the application of main composition method[J]. Application of Statistics and Management, 2002,21(1):10-23.

[8] 翁海龙,贾红亮,陈宏伟,等. 思茅松高产脂优树产脂量相关因子分析[J]. 东北林业大学学报,2008,36 (11):69-71.

WENG H L,JIA H L,CHEN H W. Factors influencing resin yield of high-yield rosin Pinus kesiya var. langbianensis[J]. Journal of Northeast Forestry University,2008,36(11):69-71.

[9] 毛惠媛,赵新良. 基于因子分析的企业家选拔评价创新研究 [J]. 东北大学学报,2009,30(3):450-452.

MAO H Y,ZHAO X L. Innovative selection/evaluation of entrepreneurs in accordance to factor analysis[J]. Journal of Northeastern University(Natural Science),2009,30(3):450-452.

[10] 王长新. 马尾松采脂量的相关因子分析[J]. 河南科技大学学报,2004,24(3):22-25.

WANG C X. Correlative analysis between the factors and oleoresin output of pinus massoniana[J]. Journal of Henan University of Science and Technology,2009,30(3):450-452.

[11] 李崇贵,赵宪文、李春干,等. 森林蓄积量遥感估测理论与实现 [M]. 北京:科学出版社,2006.

[12] 周瑾. 因子得分的岭估计和主成分分析 [D]. 武汉:武汉理工大学,2005.

[13] 王黎明,陈颖,杨楠. 应用回归分析[M]. 上海:复旦大学出版社,2008.

[14] RICHARD A J. Probablity and statistics for engineers [M]. 北京:电子工业出版社,2005.

[15] 付小勇. 云南松林分生长模型研究[D]. 云南:西南林学院,2006.