

# 基于可拓数据挖掘的 天然异龄林起测直径研究

林进添 陈平留\*

(福建农林大学 林学院 福建 福州 350002)

**摘要:**从森林资源资产评估资产核查的角度出发,为准确表述林分质量、客观反映出材情况和有效降低外业工作量,运用可拓数据挖掘方法,选择林分蓄积量减幅、商品材出材量减幅以及株数减幅3个评价特征,计算不同起测直径水平下的关联函数值,分析提高起测直径带来的样地调查影响和变化规律,进而确定天然异龄林合理起测直径水平。结果表明,将天然异龄林起测直径从5 cm提高到9 cm,即将起测径阶从6径阶调整为10径阶,可以满足森林资源资产核查调查精度的要求。

**关键词:**天然异龄林;起测直径;数据挖掘;可拓学

中图分类号:S758.5<sup>+</sup>5 文献标志码:A 文章编号:1000-2286(2012)05-0965-06

## A Study on Minimum Measured Diameter of Natural Uneven-aged Forests Based on Extension Data Mining

LIN Jin-tian, CHEN Ping-liu\*

(College of Forest, Fujian Agriculture and Forestry University, Fuzhou 350002, China)

**Abstract:** With the purpose of accurately describing stand quality, objectively reflecting merchantable volume yielding and effectively reducing outside workload, by using Extension Data Mining three evaluation characteristics are reasonably determined, and the correlation function values under different minimum measured diameters are calculated in this paper. Then the paper investigates the direct impact and variation of sample plots' evaluation characteristics, brought by raising minimum measured diameter, to determine the reasonable level of minimum measured diameter for natural uneven-aged forests. Finally the paper proposes to raise the minimum measured diameter of natural uneven-aged forest from 5 cm to 9 cm, i. e. to adjust the diameter class from the diameter class of 6 cm to the diameter class of 10 cm.

**Key words:** natural uneven-aged forest; minimum measured diameter; data mining; extension theory

天然异龄林林分结构复杂,具有复林层、多树种、年龄差异大等特点,其生境和结构的复杂多样,给森林资源资产评估资产核查工作在调查精度和工作量两方面带来了难题与挑战。天然异龄林资产核查的调查精度和工作量除受调查方法影响,在一定程度上还与起测直径(Minimum Measured Diameter, MMD)有关。鉴于此,本文从森林资源资产评估资产核查视角,引入可拓数据挖掘的技术方法,以准确表述林分质量、客观反映出材情况和有效降低外业工作量为出发点,研究和确定天然异龄林合理起测直径。

收稿日期:2012-03-19 修回日期:2012-05-12

基金项目:国家林业局林业公益性行业科研项目(200704010)

作者简介:林进添(1986—),男,博士生,主要从事森林资源资产评估和森林认证研究, E-mail: fafafa810@163.com;

\* 通讯作者:陈平留,教授,博士生导师, E-mail: fjcpl@126.com。

## 1 数据来源与预处理

所用数据资料为福建省森林资源连续清查天然异龄林样地,共 40 块,均为阔叶林地。考虑到我国现行一、二类调查技术规定都将乔木树种的起测直径确定为 5 cm,且将 14 cm 以上划为分中、大径级 1 组<sup>[3-2]</sup>。因此将各样地分别在起测直径 5,6,7,8,9,10,11,12,13 cm 下,按 2 cm 为一个径级、分径阶统计株数,其中胸径在 35 cm 及以上的归入 36 cm 径阶。并按照不同起测直径情况,分别计算各样地的平均胸径、蓄积量、商品材出材量等。以起测直径 5 cm 时的原始数据作为其他起测直径下评价信息元的计算基础。

## 2 研究思路与方法

将准确表述林分质量、客观反映出材情况和有效降低外业工作量作为对天然异龄林资产核查的要求,并选取合适的评价特征、建立关联函数将其分别定量化表达。在某个起测直径下,根据关联函数的取值情况,可将样地划为正域、负域和零界<sup>[3-5]</sup>。正域即表示在该起测直径下符合资产核查要求。实施起测直径变换,对林分质量、出材情况和工作量的影响具有同时性、不同方向性。因此,可以根据不同起测直径情况下,正域样地数量的变化规律,来确定合理的天然异龄林起测直径。

### 2.1 确定评价特征及其原特征

根据天然异龄林的生境结构特点和森林资源资产评估工作需要,本文用林分蓄积量减幅、商品材出材量减幅以及株数减幅 3 个评价特征来考核提高起测直径带来的样地调查影响和变化规律。3 个评价特征的原特征如表 1,表中权重是由决策者确定的权系数,表示各评价特征的重要程度。

表 1 评价特征的原特征

Tab.1 Primitive characteristics of evaluating characteristics

评价特征 Evaluating characteristic	代号 Code	原特征 Primitive characteristic	权重 Weight
蓄积量减幅 Proportion decrease of volume	d <sub>1</sub>	平均胸径/cm 样地总株数 单株材积/cm <sup>3</sup>	0.4
商品材出材减幅 Proportion decrease of commercial timber outturn	d <sub>2</sub>	平均胸径/cm 样地总株数 单株材积/cm <sup>3</sup>	0.4
株数减幅 Proportion decrease of tree number	d <sub>3</sub>	样地总株数	0.2

表 2 不同起测直径下的样地原始数据

Tab.2 Raw data of sample plots under the different minimum measured diameter

起测直径/cm MMD	样地号 Sample Plot NO.	样地总株数 Tree number	平均胸径/cm Average DBH	蓄积量(m <sup>3</sup> /666.7m <sup>2</sup> ) Volume	商品材出材量(m <sup>3</sup> /666.7m <sup>2</sup> ) Commercial timber outturn
6	1	52	20.0	13.5	11.3
	2	77	15.5	12.0	10.0
	.....	.....	.....	.....	.....
	20	18	12.3	2.0	1.6
7	1	43	22.9	13.4	11.2
	2	69	16.5	12.0	9.9
	.....	.....	.....	.....	.....
	20	14	14.0	1.9	1.6
	.....	.....	.....	.....	.....

2.2 列出原特征对应的原始数据表和评价信息元表

分别在不同起测直径下 统计出原特征对应的原始数据 ,计算并列出评价信息元表。以下列出部分原始数据( 表 2) 和评价信息元( 表 3) 。

表 3 不同起测直径下的评价信息元

Tab.3 Evaluation information element under the different minimum measured diameter

起测直径/cm MMD	样地号 Sam ple Plot NO.	蓄积减幅/% Proportion decrease of volume	商品材出材减幅/% Proportion decrease of commercial timber outturn	株数减幅/% Proportion decrease of tree number
6	1	0.706 8	0.618 9	14.754 1
	2	0.880 3	0.776 3	11.494 3
	.....	.....	.....	.....
	20	1.063 5	0.941 0	10.000 0
7	1	1.413 6	1.237 8	29.508 2
	2	1.584 5	1.397 4	20.689 7
	.....	.....	.....	.....
	20	3.190 4	2.822 9	30.000 0
.....	.....	.....	.....	.....

2.3 建立和计算关联函数<sup>[6-10]</sup>

建立最优点在  $x_0$  处达到时的初等关联函数: 设  $X_0 = \langle a \ b \rangle$  ,  $X = \langle c \ d \rangle$  ,  $x_0 \in \langle a \ b \rangle$  ,  $X_0 \subset X$  ,且无公共端点 ,则建立初等关联函数:  $k(x) = \frac{\rho(x, x_0, X_0)}{D(x, x_0, X)}$  ,当  $X_0$  与  $X$  有公共端点 ,对一切  $x \neq x_z$  ,则初等关联函数为:

$$k(x) = \begin{cases} \frac{\rho(x, x_0, X_0)}{D(x, x_0, X)} & \text{其他} \\ \frac{\rho(x, x_0, X_0)}{D(x, x_0, X)} - 1 & \rho(x, X) = \rho(X, x_0) \text{ 且 } x \neq X_0 \end{cases}$$

建立评价特征关联函数 ,首先得根据专业规范、专家意见 ,对 3 个评价特征最优区间和允许区间作如下取值:

表 4 评价特征基本区间

Tab.4 Basic interval of evaluating characteristics

评价特征 Evaluating characteristic	最优区间/ $X_0$ Optimum Interval		允许区间/ $X$ Allowable Interval		最优值 Optimal value $x_0$	权系数 Weight
	$a$	$b$	$c$	$d$		
蓄积减幅/% Proportion decrease of volume	0	10	0	20	0	0.4
出材量减幅/% Proportion decrease of commercial timber outturn	0	5	0	15	0	0.4
株数减幅/% Proportion decrease of tree number	20	40	0	40	40	0.2

计算关联函数值 ,下面以起测直径为 6 cm 时的 1 号样地为例作说明:

( i ) 评价特征  $d_1$  的关联函数  $k_1(x)$  在  $x_0$  处达到最大值。计算距值:

$$\rho(0.706\ 8, X) = |0.706\ 8 - \frac{0+20}{2}| - \frac{20-0}{2} = 0.706\ 8, \rho(0.706\ 8, X_0) = |0.706\ 8 - \frac{0+10}{2}| - \frac{10-0}{2}$$

=0.706 8 ,因为  $\rho(0.706 8, X) = \rho(0.706 8, X_0)$  ,所以  $D(x, X_0, X) = 0 - 10 = -10$ 。

最优值  $\rho(x, x_0, X_0)$  在区间  $X_0$  中点的左侧 ,选择左侧距公式 ,且  $x = 0.706 8 > x_0$  ,故  $\rho(x, x_0, X_0) = 0.706 8 - 10 = -9.293 2$ 。

代入关联函数  $k_1(x_{11}) = \frac{-9.293 2}{-10} = 0.929 3$  ,即起测直径 6 cm 时 1 号样地关于评价特征  $d_1$  的关联函数值为 0.929 3。

(ii) 评价特征  $d_2$  的关联函数  $k_2(x)$  在  $x_0 = 0$  处达到最大值。计算距值:

$$\rho(0.618 9, X) = |0.618 9 - \frac{0+15}{2}| - \frac{15-0}{2} = 0.618 9 ,$$

$\rho(0.618 9, X_0) = |0.618 9 - \frac{0+5}{2}| - \frac{5-0}{2} = 0.618 9$  ,因为  $\rho(0.618 9, X) = \rho(0.618 9, X_0)$  ,所以  $D(x, X_0, X) = 0 - 5 = -5$  ,又最优值  $X_0 = 0$  在区间  $X_0$  中点的左侧 ,选择左侧距公式 ,且  $x = 0.618 9$  ,故:  $\rho(x, x_0, X_0) = 0.618 9 - 5 = -4.381 1$ 。

代入关联函数公式:  $k_2(x_{12}) = \frac{-4.381 1}{-5} = 0.876 2$  ,即起测直径 6 cm 时 1 号样地关于评价特征  $d_2$  的关联函数值为 0.876 2。

(iii) 评价特征  $d_3$  的关联函数  $k_3(x)$  在  $x_0 = 40$  处达到最大值 ,表示在允许的变幅内 ,尽可能的减少调查株数 ,即减少外业工作量。计算距值:

$$\rho(14.754 1, X) = |14.754 1 - \frac{0+40}{2}| - \frac{40-0}{2} \quad \rho(14.754 1, X_0) = |14.754 1 - \frac{20+40}{2}| - \frac{40-20}{2}$$

因为  $\rho(14.754 1, X) \neq \rho(14.754 1, X_0)$  且  $x \neq X_0$  ,所以:  $D(x, X_0, X) = -14.754 1 - 5.245 9 = -20$ 。最优值  $x_0 = 40$  在区间  $X_0$  中点的  $y$  右侧 ,选择右侧距 ,且  $X = 14.754 1 < x_0$  ,故  $\rho(x, x_0, X_0) = 20 - 14.754 1 = 5.245 9$ 。

代入关联函数公式:  $k_3(x_{13}) = \frac{5.245 9}{-20} = -0.262 3$  ,即起测直径 6 cm 时 1 号样地关于评价特征  $d_3$  的关联函数值为 -0.262 3。

同理 ,可求得起测直径为 6 cm 时 2~40 号样地关于 3 个评价特征的关联函数值  $k_i(x_{ni})$  ,其中  $i = 1, 2, 3$  ,代表 3 个评价特征;  $n = 1, 2, \dots, 40$  ,代表 40 个样地号。重复以上步骤 ,可继续求取起测直径为 7~13 cm 时各样地关于 3 个评价特征的关联函数值(表 5)。

表 5 不同起测直径下的关联函数值(部分)

Tab.5 Correlation function values under the different minimum measured diameter (Part)

起测直径/cm MMD	样地号 Sample plot NO.	$d_1$	$d_2$	$d_3$
6	1	0.929 3	0.876 2	-0.262 3
	2	0.912 0	0.844 7	-0.425 3
	.....	.....	.....	.....
	20	0.893 7	0.8118	-0.5000
	.....	.....	.....	.....
	$\max_{1 \leq n \leq 40}  k_i(x_{ni}) $	0.980 0	0.964 8	0.870 1
7	1	0.858 6	0.752 4	0.453 1
	2	0.841 6	0.720 5	0.017 9
	.....	.....	.....	.....
	20	0.681 0	0.385 6	0.500 0
	.....	.....	.....	.....
	$\max_{1 \leq n \leq 40}  k_i(x_{ni}) $	1.393 0	1.714 1	2.000 0

### 2.4 计算规范关联数

在不同起测直径情况下, 将上述关联函数值用公式:

$$k_{ni} = \frac{k_i(x_{ni})}{\max_{1 \leq n \leq n} |k_i(x_{ni})|} \quad n = 1, 2, \dots, K; i = 1, 2, 3 \quad (1)$$

分别进行规范化, 形成规范关联数表 6。

表 6 不同起测直径下规范关联函数值(部分)

Tab.6 Standardized correlation function values under the different minimum measured diameter (Part)

起测直径/cm MMD	样地号 Sample plot NO.	$d_1$	$d_2$	$d_3$
6	1	0.948 2	0.908 2	-0.301 4
	2	0.930 5	0.875 6	-0.488 8
	.....	.....	.....	.....
	20	0.911 9	0.841 4	-0.574 6
7	1	0.616 4	0.439 0	0.226 6
	2	0.604 1	0.420 3	0.008 9
	.....	.....	.....	.....
	20	0.488 8	0.225 0	0.250 0
.....	.....	.....	.....	.....

### 2.5 计算综合关联度和确定可拓分类

在不同起测直径情况下, 用公式  $K(n) = \sum_{i=1}^3 \alpha_j k_{ni}$  分别计算综合关联度。当起测直径为 6 cm 时, 各样地综合关联度为:

$$K(1) = 0.4 \times 0.948 2 + 0.4 \times 0.908 2 + 0.2 \times (-0.301 4) = 0.682 3$$

$$K(2) = 0.4 \times 0.930 5 + 0.4 \times 0.875 6 + 0.2 \times (-0.488 8) = 0.624 7$$

.....

$$K(20) = 0.4 \times 0.911 9 + 0.4 \times 0.841 4 + 0.2 \times (-0.574 6) = 0.586 4。$$

同理, 可推算其他起测直径下的样地综合关联度。

## 3 结果与分析

不同起测直径情况下各样地的综合关联度计算结果如表 7 所示。当起测直径提高到 6 cm、7 cm 时, 绝大部分样地的综合关联度均大于 0, 代表着起测直径提高到该水平时, 样地调查符合 3 个评价特征

表 7 不同起测直径下各样地的综合关联度(部分)

Tab.7 Comprehensive correlation degree of sample plots under the different minimum measured diameter (Part)

样地号 Sample plot NO.	起测直径/cm MMD							
	6	7	8	9	10	11	12	13
1	0.682 3	0.467 5	0.278 1	0.218 5	0.080 0	0.038 4	-0.014 3	-0.041 7
2	0.624 7	0.411 6	0.221 3	0.159 5	0.123 3	-0.074 4	-0.114 5	-0.150 5
.....	.....	.....	.....	.....	.....	.....	.....	.....
20	0.586 4	0.335 5	-0.112 5	-0.200 3	-0.256 9	-0.298 1	-0.250 8	-0.228 6
.....	.....	.....	.....	.....	.....	.....	.....	.....
39	0.600 0	0.441 2	0.245 7	0.093 3	-0.214 3	-0.310 1	-0.380 6	-0.356 4
40	0.607 7	0.614 2	0.340 8	-0.134 5	-0.242 2	-0.322 5	-0.253 3	-0.214 4

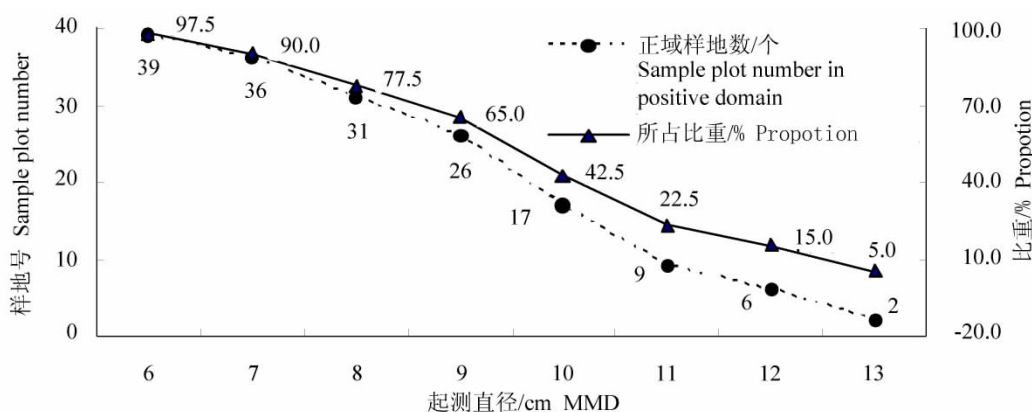


图 1 不同起测直径下正域样地数量变化

Fig. 1 Quantity variation of sample plots in the positive domain under the different minimum measured diameter 的目标性质, 满足资产核查要求。当起测直径继续提高时, 样地的综合关联度逐渐减小乃至变为负值。根据综合关联度值情况, 把综合关联度值大于 0 的样地划为正域、综合关联度值小于 0 的样地划为负域。从图 1 可以看出, 随着起测直径的提高, 正域样地数量先缓慢而后急剧下降; 说明起测直径适当提高, 一定程度上可满足调查精度又减少工作量; 但起测直径过高, 则有相反效果。正域样地数及其所占比重在起测直径 9 cm 时存在着一个拐点。鉴于此, 建议将天然异龄林起测直径从 5 cm 提高到 9 cm, 即将起测径阶从 6 径阶提高为 10 径阶。

### 4 讨论

仅从森林资源资产评估资产核查的角度出发, 得出天然异龄林资产核查可以适当提高其起测直径。一方面起测直径的提高, 一两个择伐周期内的可采伐资源不受起测直径的影响, 另一方面可以在准确描述林分质量水平、反映出材真实情况的基础上, 减少资源核查的外业工作量。未来研究天然异龄林合理起测直径, 应在评价特征的选取、区间取值和比重赋权上采取科学方法、凝聚更多领域专家的共识来确定; 此外, 由于天然异龄林资源少, 该类林分固定样地更是有限, 本文采用的样地资料有限, 未来研究应增加样地数量以提升信度和效率。

#### 参考文献:

[1]牛丽丽, 余新晓, 刘彦, 等. 不同起测胸径对判定油松分布格局的影响[J]. 北京林业大学学报, 2008, 30(2): 12-16.  
 [2]李丽, 惠淑荣, 惠刚盈, 等. 不同起测径对判定林木空间分布格局影响的研究[J]. 林业科学研究, 2007, 20(3): 334-337.  
 [3]杨春燕, 蔡文. 基于可拓集的可拓分类知识获取研究[J]. 数学的实践与认识, 2008, 38(16): 184-191.  
 [4]黄金才, 陈文伟. 可拓数据挖掘的概念与理论[J]. 计算机工程与应用, 2006, 14: 7-8.  
 [5]CAI Wen. Extension theory and its application[J]. Chinese Science Bulletin, 1999, 44(17): 1538-1548.  
 [6]李小妹. CPI 指数变换对产品销售影响的可拓数据挖掘[J]. 数学的实践与认识, 2009, 39(4): 178-183.  
 [7]方耀楣, 何万篷. 可拓数据挖掘在高校教学质量评价中的应用[J]. 数学的实践与认识, 2009, 39(4): 82-87.  
 [8]李兴森, 石勇. 基于可拓集的企业数据挖掘应用方案初探[J]. 哈尔滨工业大学学报, 2006(7): 1124-1128.  
 [9]杨春燕, 蔡文. 可拓工程[M]. 北京: 科学出版社, 2007: 82-91.  
 [10]蔡文, 杨春燕, 陈文伟, 等. 可拓集与可拓数据挖掘[M]. 北京: 科学出版社, 2008: 96-95.