

文章编号: 1000-5692(2003)02-0219-05

拟合闽北杉木干形曲线的三次样条函数的节点选择

刘 健, 郭建宏, 郭进辉, 郑淑娟

(福建农林大学 林学院, 福建 南平 353001)

摘要: 收集闽北杉木解析木材料, 运用统计学的知识并结合计算机技术, 用三次样条函数拟合杉木干形曲线, 以样本相关系数均值为主要评价指标, 筛选出拟合杉木干形曲线的三次样条函数的节点, 即从树干基部算起, 依次为树高的 0, 10%, 15%~20%, 50%~60%, 80%~85%和 100%处。经 3 处方案的拟合比较和实例验证, 结果表明所选节点的拟合效果明显优于用传统经验节点的拟合效果; 所选节点整体拟合精度高, 适应性好, 能够全面确切反映杉木完整形状。表 4 参 8

关键词: 样条函数; 杉木; 干形曲线; 节点选择

中图分类号: S758 **文献标识码:** A

树干纵断面轮廓的对称曲线称为干形曲线。该曲线自基部向梢端变化大体呈凹曲线、平行于 x 轴的直线、抛物线和相交于 x 轴的直线这 4 种类型。为了探索能够表达整株树干的形状, 大偶真一和门捷列夫等提出过许多干形曲线式^[1]。这些干形曲线式一般只能表示出中上部形状, 对基部适应性较差。样条函数是光滑连接的分段多项式, 具有很强的适应数据微小变化的能力, 并具有整体光滑性, 因而成为曲线修匀的理想工具^[2]。美国 NCHS 标准就是采用三次样条函数修匀的^[3]。近年来, 样条函数在医学和物理学等领域得到广泛的研究和应用^[3,4], 在林学和生态学方面的研究还不多见。薛亚琴等^[5]根据传统的经验数据作为节点, 用样条函数拟合林木干形曲线, 取得了较好的拟合效果^[5]。因受计算技术的限制, 早期的研究没有进行节点的选择, 且随着树种、气候等的差异, 林木的干曲线变化较大^[1,5]。为此, 本研究运用统计学的知识, 结合计算机技术, 用三次样条函数拟合杉木 *Cunninghamia lanceolata* 干形曲线, 筛选出整体拟合精度高, 适应性好, 能够全面确切反映杉木完整形状的三次样条函数节点, 并进行拟合计算及精度比较。

1 研究方法

1.1 三次样条函数简介^[4~9]

所谓样条函数, 从数学上说, 就是按一定光滑性要求“对接”起来的分段多项式。具体地说, 在区间 $[a, b]$ 上, 给定一个分划: $a = x_1 < x_2 < x_3 \cdots < x_n = b$, 这样我们在区间 $[x_i, x_{i+1}]$ 上可以找到近似的三次多项式 $p_i(x)$ ($i=1, 2, \dots, n-1$), 要求:

$$p_i(x_i) = f(x_i), p_i(x_{i+1}) = f(x_{i+1}), \quad (i = 1, 2, \dots, n-1), \quad (1)$$

且
$$p'_i(x_i) = p'_{i-1}(x_i), p''_i(x_i) = p''_{i-1}(x_i), \quad (i = 2, 3, \dots, n-1), \quad (2)$$

收稿日期: 2002-09-28; 修回日期: 2003-02-23

基金项目: 福建省教育厅科学基金资助项目(JB01165)

作者简介: 刘健(1963-), 男, 福建福州人, 副教授, 从事森林资源资产评估和森林可持续经营等研究。

则称这样的分段多项式函数为样条函数, 点 x_i 称为样条函数的节点。为使表达式简单, 记: $h_i = x_{i+1} - x_i$, $\phi_i = p''_{i+1}(x_i) = p''_i(x_i)$ 。由定义知, 每一个 $p''_i(x)$ 都是 x 的线性函数, 因此在区间 $[x_i, x_{i+1}]$ 上可由区间端点的值 ϕ_i 和 ϕ_{i+1} 用线性插值按 Langrange 公式直接写出: 即: $p''_i(x) = \frac{x_{i+1} - x}{h_i} \phi_i + \frac{x - x_i}{h_i} \phi_{i+1}$ 。将上式积分二次, 利用条件 (1) 得出:

$$p_i(x) = \frac{\phi_i}{6h_i}(x_{i+1} - x)^3 + \frac{\phi_{i+1}}{6h_i}(x - x_i)^3 + [\frac{f_{i+1}}{h_i} - \frac{h_i\phi_{i+1}}{6}](x - x_i) + [\frac{f_i}{h_i} - \frac{h_i\phi_i}{6}](x_{i+1} - x) \quad (3)$$

式中: $i=1, 2, \dots, n-1$ 。由 (3) 式可以看出, 只要求出每个 ϕ_i (待估参数), 就可由 (3) 式求出所有的三次插值多项式。通常采用追赶法求解, 树干顶部边界条件采用自由边界, 根径处边界采用固定边界, 一般取 -2.5 左右为宜。

1.2 节点选择方法

用样条函数方法模拟生物体形状, 应遵循客观规律, 了解其生物背景。根据林木干形的生物学特性, 用样条函数拟合林木干形曲线时, 一般取 6 个节点为宜 (包括根径处和梢顶处)。用杉木解析材料, 从树干基部算起, 把树高按 5% 间隔区分, 计算各相对高处断面高 $x_i (i = 0, 1, 2, \dots, 20)$, 利用插值法求各断面高处的断面半径 $y_i (i = 0, 1, 2, \dots, 20)$, 然后以组合的方式, 选取 6 个点 $P(x_i, y_i)$ 作为三次样条函数一组节点, 6 个点中要求必须选根径处 $P(x_0, y_0)$ 和梢顶处 $P(x_{20}, y_{20})$ 。这样的组合共有 $C_1^1 C_{19}^4 C_1^1$ 个, 即有 3 876 个节点组合。

从闽北地区收集的杉木解析木材料中随机等概地抽取 100 株样木, 样木树高变化幅度为 4~24 m, 胸径变化幅度为 6.0~32.5 cm。每组合用三次样条函数对 100 株样木进行拟合, 计算其相关系数 (r) 和残差平方和 (Q), 且进行相关系数的显著性 t 检验。进而计算每组合样木的相关系数均值 (\bar{r}) 和残差平方和均值 (\bar{Q}), 最后以相关系数均值 (\bar{r}) 作为评价节点组合优劣的指标。另外, 考虑到各组合节点对样木的整体适应性, 还对各组合中 100 株样木的相关系数 (r) 的分级情况及在整体中所占的比例进行了统计。

2 结果与分析

2.1 节点选择

各组合对 100 株样木拟合且通过相关系数显著性 t 检验后, 求算出相关系数均值 (\bar{r}), 按 \bar{r} 数值从高到低排序取前 10 位节点组合列于表 1 和表 2。

表 1 节点分布情况
Table 1 The distribution of nodes

序号	节点组/%				\bar{r}	\bar{Q}	相关系数分级/%				
	第 2 点	第 3 点	第 4 点	第 5 点			0.99 以上	0.99~0.98	0.98~0.97	0.97~0.96	0.96~0.95
1	10	15	60	85	0.993 1	3.343 3	78	16	5	1	0
2	10	15	55	85	0.993 0	3.181 6	75	19	5	1	0
3	10	20	55	85	0.993 0	3.462 2	76	17	6	1	0
4	10	20	60	85	0.993 0	3.707 9	79	13	6	2	0
5	10	15	50	80	0.992 9	3.190 1	77	17	5	1	0
6	10	15	55	80	0.992 9	3.314 6	78	15	6	1	0
7	10	15	65	85	0.992 8	3.590 6	74	18	7	1	0
8	10	20	50	80	0.992 7	3.431 8	75	18	7	0	0
9	10	15	45	80	0.992 7	3.158 5	76	17	6	0	1
10	10	20	55	80	0.992 7	3.608 6	78	15	6	1	0

说明: 1. 表中相关系数分级区间采用上限排除法; 2. 第 1 点 (根径处) 和第 6 点 (梢顶处) 表中不列出 (下同)

从表 1 中前 10 位节点组合的相关系数均值 (\bar{r}) 来看, 最小值为 0.992 7, 说明这些节点组合的

拟合效果都是比较理想的。从相关系数分级情况看, 各组合中有 92% 以上的样木的相关系数在 0.98 以上 (包括 0.98), 且有 75% 以上的样木的相关系数在 0.99 以上, 说明这些节点组合对 100 株样木的整体适应性较好; 依据相关系数显著性 t 检验计算式 [即 $t = r \cdot (n - 2)^{0.5} \cdot (1 - r^2)^{-0.5}$], 对表 1 中各组合 100 株样木拟合的相关系数进行显著性检验, 计算其统计量 t 值均大于 $t_{0.01} (18) = 2.878$, 表明拟合的相关性显著。另外, 通过对各组合的相关系数均值 (\bar{r}) 与相关系数 (r) 在 0.98 以上的株数间的拟合, 两者之间有显著的正相关, 说明用相关系数均值 (\bar{r}) 评价各组节点拟合效果是比较合理的, 既考虑了个体的拟合效果, 又考虑了整体拟合效果。

从表 2 中 10 个组合的节点的分布来看, 第 2 节点在 10% 处, 第 3 节点在 15%, 20%, 25% 处有分布, 根据各点出现的频数, 选择第 3 节点在 15%

表 2 节点分布统计表

Table 2 The statistical table of node distribution

节点 相对高/%	第 2 点		第 3 点			第 4 点					第 5 点		
	10	15	20	25	45	50	55	60	65	80	85	90	
频 数	10	6	4	0	1	2	4	2	1	5	5	0	
频 率	100	60	40	0	10	20	40	20	10	50	50	0	

~20% 处。同样的原理, 选择第 4 节点在 50% ~ 60% 处, 第 5 节点在 80% ~ 85% 处。

2.2 精度比较

表 3 各方案拟合比较

Table 3 The comparison of fitting in some projects

从模外解析木材料中随机等概地抽取 20 株样木。模外样木树高变化幅度为 6.3 ~ 21.5 m, 胸径变化幅度为 6.5 ~ 28.4 cm。分别对 3 种节点选择方案进行拟合比较。其中方案 1 为本文用三次样条函数拟合筛选的节点取频数最大的点 (0, 10%, 15%, 55%, 85%, 100%)。方案 2 在常规经验节点^[5] 范围内选取, 取 (0, 5%, 15%, 40%, 95%, 100%)。另外, 由于胸径是测树学上比较重要的测树因子, 故本文在方案 1 的基础上, 考虑胸径测树因子的重要性, 采用了树高 10 m 以下的样木第 2 节点选取断面高 0.5 m 处, 第 3 节点选取断面高 1.3 m (胸径) 处, 而 10 m (包括 10 m) 以上的样木, 考虑到其适应简, 则第 2 节点选取 1.3 m 处, 其他节点按方案 1 选取, 简称方案 3。20 株样木分别各方案的拟合比较如表 3。

序号	树高	胸径	方案 1		方案 2		方案 3	
			r_1	Q_1	r_2	Q_2	r_3	Q_3
1	6.3	8.9	0.988 9	1.197 8	0.981 1	2.035 2	0.987 6	1.341 4
2	6.9	6.5	0.995 5	0.213 6	0.998 5	0.071 1	0.990 2	0.459 9
3	7.1	7.2	0.994 4	0.317 3	0.903 0	5.218 4	0.990 2	0.552 3
4	7.4	8.3	0.995 9	0.385 7	0.985 4	1.356 3	0.996 3	0.3496
5	8.6	12.	0.990 3	1.209 2	0.958 3	5.115 1	0.990 0	1.242 0
6	10.6	20.0	0.998 4	0.674 5	0.998 6	0.578 2	0.996 7	1.388 5
7	11.3	18.0	0.996 1	1.282 3	0.971 1	9.352 6	0.996 2	1.247 6
8	12.0	16.1	0.997 5	0.689 1	0.988 2	3.251 9	0.997 7	0.647 6
9	12.9	18.0	0.995 9	1.490 4	0.990 3	3.494 2	0.995 8	1.525 3
10	13.1	14.0	0.987 2	2.234 0	0.775 1	35.701 8	0.987 2	2.232 0
11	13.8	13.9	0.988 0	2.175 4	0.797 6	33.233 3	0.988 0	2.182 8
12	14.3	12.3	0.985 0	1.743 4	0.753 9	25.237 1	0.984 5	1.798 8
13	14.9	16.2	0.996 6	0.771 7	0.953 8	10.393 9	0.996 6	0.779 1
14	15.1	20.4	0.997 2	1.215 0	0.997 8	0.958 7	0.997 2	1.231 4
15	16.0	22.0	0.993 0	2.775 0	0.876 4	46.049 6	0.993 0	2.781 4
16	16.5	22.0	0.994 9	2.509 8	0.976 2	11.641 5	0.993 0	3.424 2
17	17.0	20.4	0.990 6	4.530 7	0.991 1	4.289 7	0.988 7	5.446 2
18	17.7	26.0	0.996 3	2.057 4	0.948 5	27.845 2	0.994 4	3.066 5
19	19.7	24.4	0.999 4	0.366 7	0.997 5	1.508 1	0.999 0	0.637 2
20	21.5	28.4	0.992 9	5.934 5	0.987 7	10.358 7	0.982 9	14.312 0
均值	13.1	21.5	0.993 7	1.688 7	0.941 5	11.853 0	0.992 3	2.332 3

薛亚琴等^[5] 的研究表明, 用方案 2 的节点拟合林木干形曲线, 其精度明显优于用多元回归等方法拟合的精度。通过对各方案拟合效果分析得知, 方案 1 的相关系数均值要比方案 2 的相关系数均值明显大, 而方案 1 的残差平方和均值要比方案 2 的残差平方和均值小, 且方案 1 的残差平方和均值要比方案 2 的残差平方和均值的一半还要小, 说明本文筛选的节点拟合的效果明显优于用常规经验节点拟合的效果。另外, 从相关系数均值和残差平方和均值来看, 方案 3 的拟合效果要明显优于方案 2 的拟

合效果, 而方案3的拟合效果要比方案1的拟合效果差, 但是不显著, 所以为了考虑胸径因子的重要性, 在实际应用过程中也可以选用方案3。

3 实例验证

随机抽取20株模外样木中的第13号样木进行拟合检验, 结果详见表4。

表4 拟合值与实际值的比较

Table 4 The comparison of fitted value with actual value

相对高/%	断面高/m	实际值/cm	估计值/cm	误差/cm	相对高/%	断面高/m	实际值/cm	估计值/cm	误差/cm	相对高/%	断面高/m	实际值/cm	估计值/cm	误差/cm
0	0.0	9.9	9.9	0.0	35	5.2	6.8	6.9	-0.1	70	10.4	4.9	4.9	0.0
5	0.7	8.4	8.7	-0.3	40	6.0	6.5	6.6	-0.1	75	11.2	4.5	4.6	-0.1
10	1.5	8.1	8.1	0.0	45	6.7	6.3	6.2	0.1	80	11.9	3.9	4.0	-0.1
15	2.2	7.9	7.9	0.0	50	7.5	6.0	5.9	0.1	85	12.7	3.3	3.3	0.0
20	3.0	7.7	7.7	0.0	55	8.2	5.7	5.7	0.0	90	13.4	2.6	2.3	0.3
25	3.7	7.4	7.5	-0.1	60	8.9	5.3	5.4	-0.1	95	14.2	1.8	1.1	0.7
30	4.5	7.2	7.2	0.0	65	9.7	5.0	5.2	-0.2	100	14.9	0.0	0.0	0.0

说明: 相关系数0.9966, 剩余标准差0.7717

用 F 检验^[8]对第13号模外样木拟合效果进行检验, 即在估计值与实测值之间建立一个线性回归方程, 用方程 $F = \frac{(N-2)[N(a-0)^2 + 2a(b-1)\sum x_i + (b-1)^2\sum x_i^2]}{2\sum (y_i - a - bx_i)^2}$ 来判断数据点对线性回归方程的偏差, 即检验 a 与0, b 与1之间的差异显著性。若 $F < F_{\alpha}(1, n-2)$, 说明估计值与实际值差异不显著, 拟合效果理想; 若 $F > F_{\alpha}(1, n-2)$, 说明估计值与实际值差异显著, 拟合效果不理想。经计算, 其结果为 $F = 1.21 < F_{0.10}(1, 18) = 3.01$, 即估计值与实际值差异不显著, 表明其拟合效果是理想的。

4 讨论

用样条函数拟合杉木干形曲线的关键是节点的选择。本研究在节点选择问题上, 运用统计学的知识, 结合计算机技术, 通过大量的循环拟合计算, 选出了拟合精度高, 可靠性强的节点。考虑到实际适用性, 本文只给出了节点的范围, 即从树干基部算起, 依次为树高的0, 10%, 15%~20%, 50%~60%, 80%~85%, 100%处。在实际应用中如果所用的材料是树干解析材料, 节点可以直接选用节点范围内各相对高处的解析圆盘的数据^[5]。考虑胸径测树因子的重要性, 也可以选用方案3的结果。另外, 由于林木干形曲线受气候等因素影响比较大, 而本文用闽北地区解析木材料进行拟合, 故结果有一定的地域性, 如在其他地区应用时应进行拟合检验。

样条函数是一种分段多项式, 其构造特点之一是节点一定在曲线上, 具有插值灵活和最佳逼近等优良特性, 因而曲线精度高, 也可以通过节点的选择, 用于其他林木干形曲线的拟合。用样条函数拟合林木干形曲线, 可以用其旋转体的积分公式求其材积^[7]。该法拟合精度高, 节省野外测定工作量, 特别是在现行上部直径测定仪和全站仪的应用, 本法提供不伐倒树木而量测树木材积的方法。

参考文献:

- [1] 北京林业大学. 测树学 [M]. 北京: 中国林业出版社, 1986. 6-27.
- [2] 李岳生, 齐东旭. 样条函数方法 [M]. 北京: 科学出版社, 1977.
- [3] Hamill P V V, Drizd T A, Johnson, C L, et al. NCHS growth curves for children birth-18 years [R]. Washington DC: National Center for Health Statistics, 1979.
- [4] 尚磊, 徐勇勇, 侯茹兰, 等. 采用三次样条函数拟合体重百分位数曲线 [J]. 中国卫生统计, 2001, 18(5): 266-268.
- [5] 薛亚琴, 金敬子. 用样条函数方法拟合林木干形曲线的研究 [J]. 吉林林学院学报, 1987, 3(4): 42-47.

- [6] 陈小平, 于盛林, 刘文波. 基于遗传算法的三次样条函数拟合 [J]. 数据采集与处理, 2000, 15 (2): 138—141.
- [7] 郎奎健, 唐守正. IBMPC 系列程序集——数理统计, 调查规划, 经营管理 [M]. 北京: 中国林业出版社, 1987. 59—65.
- [8] 田有圳, 黄金桃, 林照授, 等. 凹叶厚朴一元立木材积方程的研究 [J]. 浙江林学院学报, 2002, 19 (3): 255—258.

A study of node selection of stem curves of *Cunninghamia lanceolata* in northern Fujian fitted with cubic spline functions

LIU jian, GUO Jian-hong, GUO Jin-hui, SHENG Shu-juan

(Forestry College of Fujian Agriculture and Forestry University, Nanping 353001, Fujian, China)

Abstract: On the basis of statistic knowledge and computer technology and with correlative coefficients fo samples as main evaluating indexes, the data from stem analysis of *Cunninghamia lanceolata* in the north of Fujian Province are used to fit the stem curves with cubic spline functions. From the base of the tree, the nodes are expressed by percentage of the tree height. The nodes selected in this paper are as follows: 0, 10%, 15%~20%, 50%~60%, 80%~85% and 100%. Compared with other methods, the fitting effects of selected nodes are better than those of traditional nodes. The tested sample shows that the nodes selected in this paper can precisely fit the stem curves and embody the integrated shape of *Cunninghamia lanceolata*. [Ch, 4 tab. 8 ref.]

Key words: spline functions; *Cunninghamia lanceolata*; stem curves; node selection