林业科学研究 2000, 13(1): 75～79
Forest Research

文章编号: 1001-1498(2000) 01-0075-05

非线性树高曲线模型的研究

王明亮, 李希菲
(中国林业科学研究院资源信息研究所, 北京 100091)

摘要: 引入失拟检验的理论和方法检验了树高曲线模型的适度, 应用全模型和选模型的检验理论和方法对 3 参数模型的某个参数取定值作了检验, 结果表明 2 参数模型描述树高曲线已经足够。比较了 6 个 2 参数树高模型, 常用的幂函数式 $H = aD^b$ 和双曲线式 $H = aD/(D+b)$ 以及本文提出的 Richards 式的特例即 $H = a(1-e^{-0.05D})^c$ 均表现良好, 推荐作为基本的树高曲线模型。

关键词: 树高曲线; 失拟检验; 全模型; 选模型
中图分类号: S758.5　　文献标识码: A

树高曲线模型在林业生产与实践中应用广泛, 在生长与收获模型研究中受到重视[1～4]。近期, 标准树高曲线模型研究方兴未艾[5～8]。作为该类研究基础的树高曲线基本形式不一而足, 有必要对树高曲线的基本形式作一比较和选择, 为标准树高曲线的研究提供基础。

关于树高曲线模型的研究, 文献[1]、[2]和[3]分别作过比较系统的研究。文献[1]比较了 13 个线性树高曲线模型, 文献[2]比较了 20 个非线性模型, 文献[3]则比较了 33 个线性和非线性模型。文献[1]的研究表明 Schumacher 式(线性形式)拟合效果最好。文献[2]的研究表明, 许多上凸函数和 S 型函数都可以用于描述树高曲线, 这些函数的拟合结果(均方误、相关指数等)很接近; 文献[3]得到类似结果。

文献[2]、[3]对树高曲线的比较是针对某树种树高曲线在地区的平均表现, 没有具体到林分。本文以杉木(*Cunninghamia lanceolata* (Lamb.) Hook.)材料将树高曲线的比较具体到林分上。模型的研究, 应首先检验模型本身适合与否, 即确定一个具体的回归函数是否适当地拟合了数据。在此基础上, 才谈得上模型之间的比较即模型的选优。本文引入失拟检验[9]的方法来检验模型的适度。

文献[2]、[3]的研究都表明 4 参数的树高曲线模型不足取, 3 参数的模型一般要好于 2 参数模型。但这种"好于"在统计上是否显著? 文献[10]指出, 任何模型都需要一定的弹性同时保持一定的稳定性来拟合数据或预测。模型的弹性和稳定性是对立的, 表现在模型参数上就是参数个数的多寡。一般地, 参数越多弹性越强但稳定性越弱。本文应用全模型和选模型的理论[9]来检验某些树高曲线模型的参数是否冗余。

下面列出了一些比较常用的树高曲线模型, 分别属于幂函数((1))、双曲线((2)、(4))、Schumacher 式((5)、(6)(7))、单分子((8))、Richards 式((9))、Gompertz 式((10))、Logistic 式((11))、Weibull 式((12))或其扩展形式的范畴。(13)、(14)、(15)分别为本文提出的(9)、

收稿日期: 1999-03-16
基金项目: 国家自然科学基金项目"林分生长的地理和种源变异及其模型的研究"(编号: 39670609)以及中国林科院基金课题"标准树高曲线模型机理及地区差异的研究"内容
作者简介: 王明亮(1970-), 男, 山东寿光人, 助理研究员.

(10)、(11) 式的特例。15 个非线性树高曲线模型如下:

$H = aD^{b}$	(1);	$H = aD/(D+b)$	(2);	$H = a[D/(D+1)]$	(3)
$H = a/(1+bD^{-c})$	(4);	$H = ae^{-b/D}$	(5);	$H = ae^{-bD^{-c}}$	(6)
$H = ae^{-b/(D+c)}$	(7);	$H = a(1-e^{-bD})$	(8);	$H = a(1-e^{-bD})^{c}$	(9)
$H = ae^{-be^{-cD}}$	(10);	$H = a/(1+be^{-cD})$	(11);	$H = a(1-e^{-bD^{c}})$	(12)
$H = a(1-e^{-0.05D^{c}})$	(13);	$H = ae^{-be^{-0.1D}}$	(14);	$H = a/(1+be^{-0.15D})$	(15)

1 数据资料

浙江开化地区 81 块杉木测高样地,样地面积 400～600 m^2, 年龄范围在 12～26 年生,密度范围在 1 000～4 500 株 · hm^{-2},立地指数范围在 8～18 m。每块样地内测高株数多于 30 株。

2 树高曲线模型的失拟检验

模型的比较,应首先检验模型本身适合与否,即确定一个具体的回归函数是否适当地拟合了数据或者说回归模型的形式是否正确。一般地,可以采用适应性检验的方法来判断模型有无系统误差。这里,我们引入另外一种检验方法——失拟检验。失拟检验假设在给定 x 时,观察值 y 独立、正态、等方差,显然这些假设都是很一般的假设。文献[9]给出了线性回归函数失拟检验的一般方法。通过分析线性回归函数失拟检验的推理过程,可以类似地应用到非线性回归函数。下面简单介绍失拟检验。

一般地,统计模型可表示为

$$y = E(y|x) + e = f(x) + e$$

y 为因变量观测值,x 为自变量观测值,$E(y|x)$ 表示回归函数,$f(x)$ 表示回归函数 $E(y|x)$ 的某种具体形式,e 为误差。

在重复观测条件下,将残差平方和 $SSE = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ 分解为两部分,即

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{j=1}^{C}\sum_{k=1}^{n_j}(y_{jk} - \hat{y}_j)^2 = \sum_{j=1}^{C}\sum_{k=1}^{n_j}(y_{jk} - \bar{y}_j)^2 + \sum_{j=1}^{C}n_j(\bar{y}_j - \hat{y}_j)^2$$

y_i 为第 i 个观测值,$\hat{y}_i$ 为拟合回归模型 $y=f(x)$ 得到的第 i 个拟合值,n 为样本个数,c 为按 x 的水平分组个数,n_j 为第 j 组样本个数、满足 $n = \sum_{j=1}^{C}n_j$,y_{jk}为第 j 组第 k 个观测值,$\hat{y}_j$ 为第 j 组观测的拟合值,由回归模型 $y = f(x)$ 得出,$\bar{y}_j$ 为第 j 组样本 y_i 的观测均值即 $\bar{y}_j = (1/n_j)\sum_{k=1}^{n_j}y_{jk}$ 。

SSE 的第一部分称为纯误差部分,记为 $SSPE = \sum_{j=1}^{C}\sum_{k=1}^{n_j}(y_{jk} - \hat{y}_j)^2$,其基本思路是建立在某些 x 水平有重复的事实之上,它反映观测值随机误差的影响。第二部分称为失拟部分,记为 $SSLF = \sum_{j=1}^{C}n_j(\bar{y}_j - \hat{y}_j)^2$,是对“真实的”回归函数 $E(y|x)$ (以不依赖于模型形式假设的组平均数 $\bar{y}_j$ 来表示) 与假定回归函数 $y=f(x)$ 偏离程度的度量,反映回归模型形式 $y=f(x)$ 是否为真正的回归函数形式的这样一种非随机误差的影响。

在 y_i 正态、独立、等方差以及 $E(y|x)=f(x)$ 的假设条件下,文献[9]指出,统计量 $F=(SSLF/c)/[SSPE/(n-c)]$ 遵从第一自由度为 c、第二自由度为 $n-c$ 的 F 分布。给定显著性

水平(例如 $\alpha= 0.05$) , 如果 $F \leq F(c, n-c)$ 则接受原假设 $H_o: E(y) = f(x)$, 表明给定的回归模型形式 $f(x)$ 正确; 如果 $F > F(c, n-c)$ 则接受备择假设 $H_a: E(y) \neq f(x)$, 表明给定的回归模型形式 $f(x)$ 不正确。

以本文数据, 胸径按 1 cm 分组取平均值, 在此基础上计算。

失拟检验的结果见表 1。可以看出, (1) ～(12) 模型中, (3) 式 $H = a[D/(D+1)]^b$ 和(5) 式 $H = ae^{-b/D}$ 较之其余 10 个模型不适于描述树高曲线, 其余 10 个模型更为适合。但需要指出, 以上 12 个模型对 81 个样地的适应性检验全部通过, 这表明在对模型形式的检验方面失拟检验要比适应性检验更加严格。

表 **1**　非线性树高曲线模型的失拟检验

模　型	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
通过检验样地数	77	76	60	76	57	72	78	74	77	78	76	77	76	73	77
通过率/ %	95.1	93.8	74.0	93.8	70.4	88.9	96.3	91.4	95.1	96.3	93.8	95.1	93.8	90.1	95.1

3　模型参数的约简

注意到, (2) 式为(4) 式参数 $c= 1$ 的特例; (5) 式为(6) 式参数 $c= 1$ 的特例或者为(7) 式参数 $c= 0$ 的特例; (8) 式为(9) 式参数 $c= 1$ 的特例或者为(12) 式参数 $c= 1$ 的特例。由上面的失拟检验结果可以看出, (2) 式作为(4) 式特例、(8) 式作为(9) 式或(12) 式的特例同样可以很好地拟合数据。考虑到模型参数在一定意义上表达了树种的生物学特性, 我们期望模型的参数个数尽可能地少以便于进一步的分析; 并且, 较多的模型参数(至少是部分地) 造成了非线性回归分析的不容易收敛或者不收敛。

引用全模型和选模型(无限制模型和限制模型) 的假设检验理论和方法(见文献[9]) 检验(4) 式参数 $c= 1$、(9) 式 $c= 1$ 以及(12) 式 $c= 1$。记全模型为 $y = f(x; \beta) + e$, 用最小二乘法拟合这个全模型并求误差平方和 SSE, 记为 $SSE(F)$。考虑假设条件 $H_o: H\beta = C$, 此假设条件下的模型称为选模型或限制模型。用最小二乘法拟合这个选模型并求误差平方和, 记为 $SSE(R)$。比较 $SSE(F)$ 和 $SSE(R)$, 一般地总会有 $SSE(F) \leq SSE(R)$, 这是因为全模型中有较多的参数, 可以更好地拟合数据, 因此围绕拟合回归线的离差小。如果 $SSE(F)$ 接近于 $SSE(R)$, 全模型几乎与选模型一样反映了观察值围绕回归线的变差, 在全模型中增加一个参数实际上不能减少 y 的变差, 因此 $SSE(R) - SSE(F)$ 小, 表明假设 H_o 成立; 另一方面, 如果变差大, 表明备择假设成立, 因为模型中所加的参数大大地帮助减少了观察值 y 围绕回归线的变差。实际使用的检验统计量是 $SSE(R) - SSE(F)$ 的函数, 为

$$F^* = \frac{(SSE(R) - SSE(F))/(df_R - df_F)}{SSE(F)/df_F}$$

df_F 和 df_R 分别为对应于全模型和选模型的误差平方和 $SSE(F)$ 与 $SSE(R)$ 的自由度, 如果 H_o 成立, 对线性模型而言, 它服从 F 分布, 即 $F^* \sim (df_R - df_F, df_F)$; 对非线性模型而言, 在大样本情况下, 它近似服从 F 分布。

这里, 我们检验(4) 式参数 $c= 1$、(6) 式 $c= 1$、(7) 式 $c= 0$、(9) 式 $c= 1$ 以及(12) 式 $c= 1$; 另一方面, 研究发现, Richards 式(9) 参数 b 取 0.05, Gompertz 式(10) 参数 c 取 0.1, Logistic 式

(11)参数c取0.15的拟合效果良好，因此也检验了(9)式$b=0.05$、(10)式$c=0.1$、(11)式$c=0.15$并同时作了失拟检验(结果见表1，即模型(13)、(14)、(15))。全模型和选模型检验的结果见表2。

表**2** 非线性树高曲线模型的全模型、选模型检验

模 型	(4)	(6)	(7)	(9)	(12)	(13)	(14)	(15)
通过检验样地数	77	36	36	71	72	78	63	69
通过率/%	95.1	44.4	44.4	87.7	88.9	96.3	77.8	85.2

注：模型(4)、(6)、(9)、(12)分别表示对各自参数$c=1$的检验；(7)表示对参数$c=0$的检验；(13)、(14)、(15)分别表示对(9)、(10)、(11)式参数$c=0.05$、$c=0.1$、$c=0.15$的检验。

由表2可以认为，(4)式$c=1$、(9)式$c=1$或$b=0.05$、(10)式$c=0.1$、(11)式$c=0.15$以及(12)式$c=1$的检验得以通过，而(6)式$c=1$、(7)式$c=0$的检验则不通过。

考虑前面失拟检验结果，可以得出结论，2参数的双曲线式(2)和单分子式(8)以及(13)、(14)、(15)描述树高曲线已经满足要求，无须采用相应的3参数双曲线式(4)、Richards式(9)、Weibull式(12)以及Gompertz式(10)和Logistic式(11)。并且模型(2)、(13)、(14)、(15)均可线性化，便于应用线性模型的理论和方法进行研究。

4 树高曲线的比较

由前面结果，选择2参数模型(1)、(2)、(8)、(13)、(14)、(15)共6个模型进行比较。比较指标为残差平方和(SSE)。以各模型分别拟合81块样地，求出任一模型(与其它模型比较)拟合的优先数；采用符号检验，比较任一模型与其它模型拟合效果的优劣。符号检验原理[6]如下：令$n=n_1+n_2$(n_1表示某模型优于其它任何一个模型的优先数，n_2表示非优先数)，则当$n_1>(n+1)/2+0.98\sqrt{n+1}$，有95%可靠性断言某模型优于其它某个模型。依本文数据，$n_1\approx49.87$。表3列出了各模型的优先数。

表**3** **2**参数非线性树高模型比较

模 型	(1)	(2)	(8)	(13)	(14)	(15)	模 型	(1)	(2)	(8)	(13)	(14)	(15)
(1)	0	35	39	34	47	56*	(13)	47	40	51*	0	52*	57*
(2)	46	0	53*	38	47	51*	(14)	33	34	33	29	0	60*
(8)	42	28	0	29	48	50*	(15)	23	30	30	24	21	0

注：带*者表示对应的横列模型优于对应的竖列模型。

由表3可见，模型(1)、(2)、(8)、(13)、(14)均优于(15)，模型(2)、(13)均优于(8)，模型(13)优于(14)。应该指出，模型(8)(单分子式)、(13)都作为Richards(9)式的特例，本文的数据表明这两个特例中以(13)式为好。

显然，模型(1)、(2)、(13)属于'最好'的树高曲线模型，(1)、(2)分别是常用的幂函数式和双曲线式。本文推荐的Richards式的特例(13)式即$H=a(1-e^{-0.05D})^c$可作为树高曲线模型。

5 结论和讨论

(1)引入了失拟检验的理论和方法检验了树高曲线模型的适度，应用全模型和选模型的检验理论和方法对3参数模型的某个参数取定值作了检验，从而对某些3参数模型作了参数的

约简, 结果表明 2 参数模型描述林分树高曲线已经足够。

(2) 比较了 6 个 2 参数树高模型, 常用的幂函数式 $H = aD^b$ 和双曲线式 $H = aD/(D+b)$ 以及本文提出的 Richards 式的特例即 $H = a(1-e^{-0.05D})^c$ 均表现良好, 推荐作为基本的树高曲线模型。

(3) 模型参数在一定意义上体现了树种的生物学特性, 尽可能地约简模型参数为研究参数的生物学意义提供了一种手段。本文将 Richards(9) 式的 b 参数取为 0.05, 以浙江开化地区的杉木材料进行验证得到良好效果, 该数值是否因地区和树种变化需要补充材料加以验证。

(4) 失拟检验要求有(同一实验水平上) 重复观测数据, 这使它在应用上受到限制。

参考文献:

[1] Curtis R O. Height-diameter and height-diameter-age eqations for second-growth Douglas-fir[J]. For Sci, 1967, 13: 365～375.

[2] Sweda T, Umemura T. A theoretical height-diameter curve(Ⅰ): Derivation and Characteristics [J]. J Jap For Soc, 1980, 62: 459～464.

[3] Huang S, Titus S J, Comparison of nonlinear height-diameter functions for major Alberta tree species [J]. Can J For Res, 1992, 22: 1297～1304.

[4] Fang Z, Bailey R L. Height-diameter models for tropical forests on Hainan Island in southern China[J]. For Ecol Manage, 1998, 110: 315～327.

[5] 惠刚盈, 盛炜彤, Gadow K V, 等. 杉木人工林收获模型系统的研究[J]. 林业科学研究, 1994, 7(4): 353～358.

[6] 李希菲, 唐守正, 袁国仁, 等. 自动调控树高曲线和一元立木材积模型[J]. 林业科学研究, 1994, 7(5): 512～518.

[7] 王明亮, 唐守正. 标准树高曲线的研制[J]. 林业科学研究, 1997, 10(3): 259～264.

[8] Lappi J. A longitudinal analysis of height/diameter curves [J]. For Sci, 1997, 43: 555～570.

[9] 约翰·内特, 威廉·沃塞曼, 迈克尔·H·库特纳. 应用线性回归模型[M]. 张勇, 王国明, 赵秀珍译. 北京: 中国统计出版社, 1990. 98～99, 533, 128～136.

[10] Zeide B. Analysis of growth equations [J]. For Sci, 1993, 39: 594～616.

Research on Nonlinear Height-diameter Models

WANG Ming-liang, LI Xi-fei

(The Research Institute of Forest Resources Information Techniques, CAF, Beijing 100091, China)

Abstract: The lack of fit test was introduced to test aptness of some height-diameter(H-D) models. The full model and reduced model test were employed to test whether it was reasonable to set some parameter of three-parameter H-D models a fixed value, with the result that two-parameter models were enough to describe H-D relationship. With the comparison of six two-parameter H-D models, the commonly used power function $H = aD^b$, the hyperbolic function $H = aD/(D+b)$ and the specail case of Richards function suggested by this paper $H = a(1-e^{-0.05D})^c$ gave the most satisfactory results, thus were recommended as the basic H-D models.

Key words: height-diameter models; lack of fit test; full model; reduced model