

生物量统一模型构建及非线性偏最小二乘辨识 ——以毛竹为例

刘恩斌, 周国模*, 姜培坤, 葛宏立, 杜华强

(浙江林学院环境科技学院, 临安 311300)

摘要: 树种生物量数据是研究许多林业与生态问题的基础, 如何准确估算树种的生物量十分重要, 建立生物量模型是生物量估计的主要手段。鉴于常用估算生物量模型具有多样性、经验性、在实际应用中有很限制等特点, 提出利用 Chebyshev 正交多项式系构建生物量估算统一模型的新思路, 这种构建模型的方法有完备的理论基础、统一的模型结构, 其实质是一种进行变量区间变换, 缩小变量定义域的数值插值方法, 模型阶数越高, 插入的点越多, 估算结果越符合实际, 这与树木的树干解析与树木生长原理是相一致的; 鉴于常用最小二乘具有的一些缺陷, 用适应范围更广的偏最小二乘对统一模型进行参数辨识; 指出毛竹生物量的研究与其它树种生物量的研究不同, 根据毛竹自身的生长特征, 把其生物量的研究分成 2 个阶段; 最后估算了毛竹第 1 阶段的生物量, 结果表明拟合精度很高: 当提取 55 个主成分时, $S_{\text{PRESS}, h}$ 达到最小, 其值为 31.4390, $R^2 = 0.9987$, 离差平方和为 0.2518, 比传统方法估计误差 5.623 提高了一个数量级。

关键词: Chebyshev 正交多项式; 偏最小二乘; 毛竹; 生物量统一模型

文章编号: 1000-0933(2009)10-5561-09 中图分类号: Q141, Q948 文献标识码: A

Construction of biomass unite model with nonlinear partial least squares regression: a case of study for bamboo biomass

LIU En-Bin, ZHOU Guo-Mo*, JIANG Pei-Kun, GE Hong-Li, DU Hua-Qiang

School of Environmental Sciences and Technology, Zhejiang Forestry University, Linan 311300, China

Acta Ecologica Sinica, 2009, 29(10): 5561 ~ 5569.

Abstract: Forest biomass is a basic quantity character of the forest ecosystems, and which indicates the structure and function of forest. To estimate accurately biomass, establishing biomass models is a major way. However, model of common estimated biomass is experience and its application is very limited. In this paper, we constructed a Chebyshev's orthogonal polynomial's property and constructed base of Continuous Functions's Hyperspaces, and uniform Model of forest biomass was expressed as linear combination of the basis, and provided a new method to tree genus's theoretic biomass model. This method of constructed model has complete theoretic base, so current Model of Forest Biomass was an especial example of uniform Model. Method of model's construction was essential numerical interpolation, This interpolation firstly processed interval transform and reduced independent variable's fields of definition. Model order was the higher, and its interpolation point was more. The estimated result was much according with practical, and also it was according with principium of stem analysis and wood growth. As disadvantages of common least squares, we applied partial least squares to calculate model of forest's biomass. This paper indicated Moso bamboo biomass's research was not same as other tree genus's. On the base of Moso bamboo's character, its biomass's was divided into 2 phases. Finally, this paper applied uniform Model of Forest Biomass to research Moso bamboo's biomass of the first phase, and result indicated fitted effect was better: $S_{\text{PRESS}, h} = 31.4390$, $R^2 = 0.9987$, minimum variance was 0.2518, The result shows that the improved method can improve more 10

基金项目: 国家自然科学基金资助项目(30771715); 浙江省科技厅资助项目(2008C12068); 国家自然科学基金资助项目(30700638)

收稿日期: 2008-06-18; 修订日期: 2008-11-21

* 通讯作者 Corresponding author. E-mail: zhougm@zjfu.edu.cn

percent estimative precision than that of old method.

Key Words: Chebyshev's orthogonal polynomial; Partial least squares; moso bamboo; unite model of biomass

树种生物量数据是研究许多林业问题和生态问题的基础,准确估算树种的生物量十分重要。建立生物量模型是生物量估测的主要手段,常用估算树种生物量的模型有:(1)多项式模型^[1~5]:这种形式的模型由于都是经验性的,其结构多样,没有统一的形式,在实际应用中不具有通用性;(2)山本式模型^[4~6]:目前在实践中用的最多,但是也是经验性的,在有些情况下,山本式模型对树种生物量的估算结果还没有多项式模型精度高^[1],这说明山本式模型存在缺陷;(3)相容性生物量模型^[7~9]:能克服多项式模型与山本式模型对树木各维量独立建模所造成的各维量生物量之和不等于树木生物量总量的缺陷,但是其各维量的建模仍是建立在经验基础上的。因此在理论上如何构建树种生物量估算的统一模型值得研究。

要对树种的生物量有一个比较准确的估计,其参数辨识方法也很重要,最常用的模型参数辨识方法是普通最小二乘与 Marquardt 算法^[6~9],但是它们本身有很多的缺陷,如:(1)变量之间出现多重相关性时,矩阵的逆不存在,此时普通最小二乘不能适用;(2)观测样本个数比模型变量个数少时,普通最小二乘也不能适用;(3)当建模样本存在异常数据时,先要进行异常数据的剔除,否则估计结果偏差较大;(4)Marquardt 非线性最小二乘算法需要给定模型的初始值,初值选的不好,迭代就不收敛。因此能克服以上 4 点缺陷的偏最小二乘算法为构建的统一模型提供一种新的参数辨识方法。

1 研究方法与理论基础

1.1 总体思路

为了在理论上得到树种生物量估算的统一模型,并对其参数进行准确辨识,在此先对本文的研究方法做一扼要的概述:(1)把生物量统一模型看作是连续函数空间中的一个元素;(2)如果能找到这个空间的一组基,则树种生物量统一模型可表示为这组基的线性组合;(3)正交多项式系的任意不同项都是线性无关的,所以它可以作为连续函数空间的一组基,而实践证实 Chebyshev 正交多项式系有许多的优良特性,非常适合数据拟合^[10,11]。基于以上几方面考虑,本文选用 Chebyshev 正交多项式系构造了一组乘积型基作为连续函数空间的基底,这样就可以在理论上构造出树种生物量估算的统一模型;(4)最后用偏最小二乘对所构建的统一模型进行参数辨识。

1.2 理论基础

1.2.1 Chebyshev 正交多项式的正交性与递推性

Chebyshev 多项式系 $\{T_n(x) = \cos(n \arccos x), n = 0, 1, 2, \dots\}$ 是在区间 $[-1, 1]$ 上,以 $\rho(x) = \frac{1}{\sqrt{1-x^2}}$ 为权的正交多项式系,即

$$\int_{-1}^1 \frac{1}{\sqrt{1-x^2}} T_m(x) T_n(x) dx = \begin{cases} 0 & m \neq n \\ \pi/2 & m = n \neq 0 \\ \pi & m = n = 0 \end{cases} \quad (1)$$

相邻的 3 个 Chebyshev 多项式具有如下递推关系式:

$$\begin{cases} T_0(x) = 1, & T_1(x) = x \\ T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) (n = 1, 2, \dots) \end{cases} \quad (2)$$

1.2.2 生物量估算统一模型的构建

从以上 2 个性质可知 Chebyshev 多项式系 $\{T_n(x)\} (n = 0, 1, 2, \dots)$ 各项之间线性无关,且在 $[-1, 1]$ 上连续,则 $\{T_n(x)\} (n = 0, 1, 2, \dots)$ 可以作为连续函数空间 $C[-1, 1]$ 的一组基。因此 $C[-1, 1]$ 上的任意连续函数 $f(x)$ 都可表示为:

$$f(x) = \sum_{i=1}^{\infty} c_i T_i(x) \quad (3)$$

(3)式适应于一元的情况,对于定义在 p 维空间 $\{x_1^*, x_2^*, \dots, x_p^*\}$ 上的多元连续函数,由于 $T_{c_1}(x_1^*) T_{c_2}(x_2^*), \dots, T_{c_p}(x_p^*)$ ($c_1, c_2, \dots, c_p = 0, 1, 2, \dots$, 式中假定 $x_1^*, x_2^*, \dots, x_p^*$ 的定义域都是 $[-1, 1]$) 是线性无关的。所以它可以作为 p 维连续函数空间的一组基,称为乘积型基。假设影响树种生物量的因子经区间变换(见(13)式)后为 $x_1^*, x_2^*, \dots, x_p^*$, 则生物量估算的统一模型可表示为:

$$f(x_1^*, x_2^*, \dots, x_p^*) = \sum_{k=0}^{\infty} a_k T_{c_1}(x_1^*) T_{c_2}(x_2^*) \dots T_{c_p}(x_p^*) \quad (c_1, c_2, \dots, c_p = 0, 1, 2, \dots) \quad (4)$$

在实际应用时,当 k 等于某个数 M 即满足精度要求,因此(4)式可写为:

$$f(x_1^*, x_2^*, \dots, x_p^*) = \sum_{k=0}^M a_k T_{c_1}(x_1^*) T_{c_2}(x_2^*) \dots T_{c_p}(x_p^*) \quad (c_1, c_2, \dots, c_p = 0, 1, 2, \dots) \quad (5)$$

(5)式就是本文要构建的树种生物量估算统一模型。

当考虑较多影响生物量的因素且模型各变量取较高阶数时,(5)式右边参数个数比较多,变量间存在多重相关性且变量个数多于样本个数的可能性很大,因此要用偏最小二乘来拟合本文所构建的模型。

1.3 偏最小二乘算法

1.3.1 偏最小二乘算法的基本思想

偏最小二乘算法是集主成分分析、典型相关分析、回归分析为一体的现代多元统计分析。

现对它的基本建模思想概述如下:设 X 为所有自变量组成的矩阵; Y 为所有因变量组成的矩阵;由于偏最小二乘采用主成分提取的方法,因此在 X 中提取第 1 主成分,用 x_1 来表示,在 Y 中提取第 1 主成分,用 y_1 来表示。再结合典型相关分析的方法,得到下式:

$$\begin{aligned} \max \quad & r(x_1, y_1) \\ \begin{cases} x_1^T x_1 = 1 \\ y_1^T y_1 = 1 \end{cases} \end{aligned} \quad (6)$$

式中, $r(x_1, y_1)$ 表示 2 个成分的相关系数。

偏最小二乘还采用回归分析的方法,因此分别实施 X 对 x_1 的回归及 Y 对 y_1 的回归。如果回归方程已经达到满意的精度,则算法终止;否则将利用 X 被 x_1 解释后的残余信息及 Y 被 y_1 解释后的残余信息进行第 2 轮的成分提取,直到满足精度为止^[12~15]。具体实施步骤详见参考文献^[12]。

在许多情况下,偏最小二乘回归并不需要采用全部的成分 $t_1, t_2, \dots, t_A$ 进行建模,因为采用过多的成分会破坏对统计趋势的认识,因此在应用偏最小二乘回归建模时必须要进行交叉有效性检验,以确定提取成分数 m ($m < A$)。

1.3.2 偏最小二乘算法的交叉有效性

偏最小二乘算法的交叉有效性检验是通过考察增加 1 个新的成分后,能否对模型的预测功能有明显的改进来考虑。其实施过程如下:(1)把原来的 n 个样本点分成 2 部分,第 1 部分是除去某个样本点 i 的所有样本点的集合,第 2 部分是第 i 个样本点;(2)用 $n-1$ 个样本点并使用 h 个成分拟合偏最小二乘回归方程;(3)把刚才除去的样本点 i 代入前面拟合的回归方程,得到 y_j 在样本点 i 上的拟合值 $\hat{y}_{hj(-i)}$, 对于每一个 $i = 1, 2, \dots, n$ 重复上面的过程;(4)计算 y_j 的预测误差平方和 $S_{\text{PRESS},hi}$:

$$S_{\text{PRESS},hj} = \sum_{i=1}^n (y_{ij} - \hat{y}_{hj(-i)})^2 \text{ 与 } Y \text{ 的预测误差平方和为 } S_{\text{PRESS},h} = \sum_{j=1}^q S_{\text{PRESS},hj}; (5) \text{ 主成分数的确定:选取}$$

的成分数 m 使 $S_{\text{PRESS},h}$ 最小为宜^[12,14,15]。

1.4 树种生物量估算统一模型的非线性偏最小二乘辨识实施步骤

由于 Chebyshev 正交多项式是定义在区间 $[-1, 1]$ 上的,而树种生物量模型自变量的取值不在区间 $[-1,$

1] 上, 因此首先要对变量进行区间变换, 其公式如下:

$$x^* = \frac{2x - (a_1 + b_1)}{b_1 - a_1}, x \in [a_1, b_1] \quad (13)$$

这样就可以把模型自变量的定义域变换到 $[-1, 1]$ 上。综上所述, 树种生物量估算统一模型的非线性偏最小二乘算法步骤如下:

第一 进行变量的区间变换。设影响树种生物量的因子为 $x_1, x_2, \dots, x_p$, 经过区间变换后分别记为 $x_1^*, x_2^*, \dots, x_p^*$ 。

第二 应用 Chebyshev 正交多项式的递推性, 分别求 $T_{c_1}(x_1^*), T_{c_2}(x_2^*), \dots, T_{c_p}(x_p^*), c_1 = 0, 1, 2, \dots, k_1, c_2 = 0, 1, 2, \dots, k_2, \dots, c_p = 0, 1, 2, \dots, k_p$ 。

第三 求 $T_{c_1}(x_1^*)T_{c_2}(x_2^*)\dots T_{c_p}(x_p^*)$ 。

第四 对 $f(x_1^*, x_2^*, \dots, x_p^*) = \sum_{k=0}^M a_k T_{c_1}(x_1^*)T_{c_2}(x_2^*)\dots T_{c_p}(x_p^*)$ 应用偏最小二乘求参数 a_k 。

第五 把 $x_1^*, x_2^*, \dots, x_p^*$ 代入 $f(x_1^*, x_2^*, \dots, x_p^*) = \sum_{k=0}^M a_k T_{c_1}(x_1^*)T_{c_2}(x_2^*)\dots T_{c_p}(x_p^*)$, 使统一模型的自变量还原成影响生物量的原始变量。

为了说明以上方法的实用性、广泛性及估算精度的精确性, 本文用所构建的模型来估算毛竹生物量, 并用偏最小二乘进行参数辨识。

2 生物量统一模型的应用

毛竹生长快, 利用周期短, 有很强的吸收 CO_2 固定碳素的功能^[16], 如何准确估算毛竹生物量, 尤其是估算区域范围内毛竹林生物量, 是正确评价毛竹林对全球变化影响的关键。目前为止, 有关毛竹生物量的研究模型都是以胸径、年龄与竹高为自变量的^[17~20], 其实毛竹无次生形成层组织, 幼竹形成后, 在以后的生命活动中, 其高度不再增长, 因此对毛竹生物量的研究与其它树种是有区别的, 应该分两个阶段: 第 1 阶段为毛竹笋刚出地面开始到其枝叶完全展开终止, 在这个阶段毛竹处于高生长期, 竹高增长很快, 其碳素的积累也比较快, 而且随着时间的增加, 碳素的积累量明显增加, 故毛竹在这个阶段的生物量与地径、竹高、时间有关; 第 2 阶段为幼竹形成后开始到采伐利用终止, 在这个阶段, 竹高不在发生变化, 因此毛竹这个阶段生物量的影响因子与第 1 个阶段有所不同。从一些学者对毛竹生物量的研究成果看, 主要集中在第 2 阶段的生物量^[17~20], 而对毛竹第 1 阶段生物量及毛竹生物量与时间有关的研究很少。故应用本文所构建的生物量估算统一模型来研究毛竹第 1 阶段的生物量显得非常必要。

2.1 数据来源

2005 年 4 月, 毛竹笋高度为 30cm 左右 (笋高度小于 30cm 时其地径不稳定, 选取样本株、控制生物量难度比较大) 的出笋期, 在地形条件基本一致的毛竹林中, 选择地径 (带笋衣) 为 7~14cm 的毛竹笋 10 株作为标准株, 以后每隔 3d 测定标准株的地径及高度, 并在其周围立地条件基本相同的地方选择相同地径 (带笋衣) 及高度的毛竹 (笋), 作为样本株 (全部是新竹), 将样本株在地面处连株伐倒, 测定笋 (竹) 秆的质量, 从竹 (笋) 底部到竹 (笋) 梢部均匀取样, 测定其样品鲜质量后带回实验室, 将样品在 105℃ 杀青 30min, 后在 70℃ 下烘干, 测定干物质量, 从而推算出样本株的干生物量。这样的样株 (品) 采集过程, 一直持续到所选的标准株枝叶完全展开, 高生长停止为止。因此, 本次调查过程一直持续到 2005 年 5 月 10 日, 共伐倒样株 100 株, 取回其样品供实验室分析之用。

2.2 结果与分析

2.2.1 生物量统一模型偏最小二乘辨识分析

毛竹笋高度为 30cm 左右的出笋期开始到其枝叶完全展开终止这段时间, 毛竹处于高生长期: 竹高增长很快, 且随着时间的增加其生物量的积累量明显增加, 因此毛竹在这个阶段的生物量与地径 D (因为不同的毛竹其地径不同且地径是影响毛竹生物量非常重要的因素, 因此把地径作为模型的一个变量)、竹高 h 、时间 t

(本文中时间 t 取值为第一次实测时 $t=0$, t 以后取值为每次实测时间与第一次实测时间的间隔天数) 有关, D, h, t 经区间变换后的变量记为 D^*, h^*, t^* , 则这个阶段毛竹生物量估算的统一模型为:

$$f(D^*, h^*, t^*) = \sum_{k=0}^M a_k T_{c_1}(D^*) T_{c_2}(h^*) T_{c_3}(t^*)$$
, 当 $c_1, c_2, c_3 = 0, 1, \dots, 5, , M = 216$ 时, 应用偏最小二乘进行拟合即达到精度要求, 根据偏最小二乘交叉有效性原理, 从 215 个成分中提取了 55 个主成分, 为了说明问题的方便, 这里列出了前 60 个成分, 其计算结果见表 1。

表 1 偏最小二乘交叉有效性提取主成分个数
Table 1 Principal component's number of PLS's crossover effectiveness

提取的主成分个数 Number of principal extracting component'	R^2	$S_{\text{PRESS},h}$	提取的主成分个数 Number of principal extracting component'	R^2	$S_{\text{PRESS},h}$
1	0.7941	99.9440	31	0.9977	39.6310
2	0.9298	74.6430	32	0.9977	39.0190
3	0.9666	62.0160	33	0.9978	38.3860
4	0.9784	55.8930	34	0.9980	38.1040
5	0.9836	52.6100	35	0.9980	37.8820
6	0.9850	51.4010	36	0.9981	38.0100
7	0.9865	50.5850	37	0.9981	38.1640
8	0.9874	50.1550	38	0.9982	38.2880
9	0.9885	49.8820	39	0.9982	38.2600
10	0.9893	49.7770	40	0.9983	38.0930
11	0.9902	49.8560	41	0.9983	37.8500
12	0.9911	49.7270	42	0.9983	37.6960
13	0.9919	49.6880	43	0.9983	37.0760
14	0.9928	50.4100	44	0.9984	36.2610
15	0.9934	50.3790	45	0.9984	35.1740
16	0.9937	50.0270	46	0.9985	34.2390
17	0.9940	49.6950	47	0.9985	33.5660
18	0.9944	49.1360	48	0.9985	33.0530
19	0.9946	48.6850	49	0.9985	32.5410
20	0.9948	48.0490	50	0.9986	32.0820
21	0.9950	47.2640	51	0.9986	31.6440
22	0.9953	46.3850	52	0.9986	31.5890
23	0.9954	45.4600	53	0.9987	31.6190
24	0.9957	44.5310	54	0.9987	31.5490
25	0.9960	43.1470	55	0.9987	31.4390
26	0.9965	41.5940	56	0.9988	31.4990
27	0.9970	40.9000	57	0.9988	32.1520
28	0.9973	40.7490	58	0.9988	32.1240
29	0.9975	40.4930	59	0.9988	32.3070
30	0.9976	40.0370	60	0.9989	32.4770

从上表可以看出, 当提取 55 个主成分时 $S_{\text{PRESS},h}$ 达到最小, 其值为 31.4390, 此时确定系数 R^2 为 0.9987, 因此本文就用提取的这 55 个成分进行偏最小二乘回归建模。模型各系数的值见表 2。

从表 2 可以看出, 变量的个数 215 (其中第 1 项是常量) 比样本的个数 100 要多, 所以常用最小二乘不能适用于本文所建的模型, 这就充分体现了偏最小二乘的优越性。表 3 给出了毛竹生物量的估测值与实测值。由于数据列数比较多, 所以表 3 中未列出实测生物量对应的竹高值。

表2 毛竹生物量统一模型参数表

Table 2 Parameters of moso bamboo's unite biomass model

参数 Parameter	参数值 Value of Parameter	参数 Parameter	参数值 Value of parameter	参数 Parameter	参数值 Value of parameter	参数 Parameter	参数值 Value of parameter
a_0	1.7854	a_{54}	-0.1816	a_{108}	-0.0961	a_{162}	0.2165
a_1	0.3288	a_{55}	-0.1134	a_{109}	0.0311	a_{163}	0.2746
a_2	-0.0266	a_{56}	0.1755	a_{110}	0.0865	a_{164}	0.0034
a_3	0.0279	a_{57}	0.1902	a_{111}	-0.1719	a_{165}	0.1470
a_4	0.1949	a_{58}	0.1102	a_{112}	-0.2091	a_{166}	0.2762
a_5	-0.0506	a_{59}	0.0041	a_{113}	0.1812	a_{167}	-0.0427
a_6	0.5016	a_{60}	-0.0183	a_{114}	-0.1172	a_{168}	-0.0580
a_7	-0.1476	a_{61}	-0.2931	a_{115}	-0.1682	a_{169}	-0.0010
a_8	0.0235	a_{62}	-0.1773	a_{116}	0.0679	a_{170}	0.1663
a_9	0.0457	a_{63}	0.4276	a_{117}	0.0697	a_{171}	-0.1384
a_{10}	0.0830	a_{64}	0.5719	a_{118}	-0.1051	a_{172}	-0.1271
a_{11}	0.1882	a_{65}	-0.2172	a_{119}	-0.2249	a_{173}	0.4862
a_{12}	-0.1807	a_{66}	0.1899	a_{120}	0.0259	a_{174}	-0.5332
a_{13}	0.4631	a_{67}	-0.0518	a_{121}	-0.2089	a_{175}	-0.4227
a_{14}	0.0603	a_{68}	-0.3389	a_{122}	-0.2151	a_{176}	0.1209
a_{15}	-0.0764	a_{69}	-0.0037	a_{123}	0.3228	a_{177}	-0.0852
a_{16}	-0.1592	a_{70}	0.2348	a_{124}	0.3356	a_{178}	-0.2715
a_{17}	0.1651	a_{71}	-0.1557	a_{125}	-0.5009	a_{179}	0.2895
a_{18}	-0.2291	a_{72}	-0.0328	a_{126}	0.1794	a_{180}	-0.0250
a_{19}	-0.2939	a_{73}	-0.0843	a_{127}	0.0885	a_{181}	-0.0352
a_{20}	0.4245	a_{74}	-0.0101	a_{128}	-0.2422	a_{182}	-0.1018
a_{21}	0.1518	a_{75}	-0.0944	a_{129}	-0.1627	a_{183}	-0.1114
a_{22}	-0.1030	a_{76}	-0.0920	a_{130}	0.0675	a_{184}	-0.0328
a_{23}	0.0397	a_{77}	0.0322	a_{131}	0.1058	a_{185}	0.1019
a_{24}	-0.1039	a_{78}	-0.1542	a_{132}	0.0346	a_{186}	0.0969
a_{25}	-0.2794	a_{79}	-0.0936	a_{133}	0.3345	a_{187}	0.0423
a_{26}	-0.0725	a_{80}	-0.0035	a_{134}	0.1946	a_{188}	-0.0696
a_{27}	0.6554	a_{81}	0.0043	a_{135}	-0.4695	a_{189}	-0.1377
a_{28}	0.3877	a_{82}	-0.1758	a_{136}	-0.4857	a_{190}	-0.0427
a_{29}	-0.1847	a_{83}	-0.1254	a_{137}	0.4541	a_{191}	0.2205
a_{30}	0.0725	a_{84}	0.0024	a_{138}	-0.1342	a_{192}	0.1998
a_{31}	0.2170	a_{85}	-0.0807	a_{139}	0.1507	a_{193}	0.4192
a_{32}	-0.0533	a_{86}	-0.0733	a_{140}	0.3859	a_{194}	0.2136
a_{33}	-0.0769	a_{87}	0.0207	a_{141}	0.1290	a_{195}	0.2138
a_{34}	0.0803	a_{88}	0.0952	a_{142}	-0.2601	a_{196}	0.0633
a_{35}	-0.2204	a_{89}	-0.1565	a_{143}	0.2585	a_{197}	-0.0950
a_{36}	0.2334	a_{90}	0.1479	a_{144}	-0.0932	a_{198}	0.0801
a_{37}	0.0949	a_{91}	0.0821	a_{145}	0.0053	a_{199}	0.3255
a_{38}	-0.1784	a_{92}	-0.1196	a_{146}	0.0369	a_{200}	0.2991
a_{39}	0.1251	a_{93}	-0.1253	a_{147}	-0.2217	a_{201}	0.3125
a_{40}	0.2545	a_{94}	0.0491	a_{148}	-0.3297	a_{202}	0.1923
a_{41}	-0.1550	a_{95}	0.0319	a_{149}	0.2151	a_{203}	-0.1837
a_{42}	0.1419	a_{96}	0.0914	a_{150}	-0.0178	a_{204}	-0.0666
a_{43}	0.3072	a_{97}	0.2890	a_{151}	-0.0303	a_{205}	-0.1176
a_{44}	0.1040	a_{98}	-0.0607	a_{152}	0.0795	a_{206}	0.2696
a_{45}	-0.1057	a_{99}	-0.1467	a_{153}	-0.1131	a_{207}	0.3388
a_{46}	0.0383	a_{100}	-0.1362	a_{154}	-0.1453	a_{208}	0.2084
a_{47}	0.1789	a_{101}	-0.0817	a_{155}	0.0349	a_{209}	0.0462
a_{48}	-0.1095	a_{102}	0.0846	a_{156}	0.2186	a_{210}	-0.3448
a_{49}	0.1183	a_{103}	-0.1310	a_{157}	0.2118	a_{211}	-0.5069
a_{50}	0.3644	a_{104}	0.1254	a_{158}	-0.0301	a_{212}	-0.2206
a_{51}	-0.0918	a_{105}	0.1598	a_{159}	0.4931	a_{213}	-0.0199
a_{52}	-0.4193	a_{106}	0.2022	a_{160}	0.4376	a_{214}	-0.0348
a_{53}	0.3826	a_{107}	0.1816	a_{161}	-0.6143	a_{215}	0.0287

表 3 毛竹生物量估值值与实测值的对比
Table 3 Observed and fitted values of moso bamboo's biomass

地径 (cm) Diameter	月-日 Month-day									
	04-13		04-16		04-19		04-22		04-25	
	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value
12.1	0.0812	0.0828	0.1765	0.1466	0.4156	0.5031	1.0146	0.9901	1.3824	1.5775
7.3	0.0317	0.0306	0.0774	0.0768	0.2735	0.2754	0.4380	0.4390	0.5893	0.5924
9.4	0.0530	0.0548	0.1080	0.1370	0.4149	0.3869	0.7340	0.7443	1.1198	1.1010
9.0	0.0767	0.0801	0.1884	0.1636	0.4336	0.4715	0.6773	0.6841	0.9169	0.9463
13.6	0.1231	0.1239	0.2734	0.2789	0.7671	0.7667	1.2914	1.2920	1.9711	1.9708
11.1	0.0943	0.0909	0.2196	0.2504	0.3655	0.3306	0.7780	0.7568	1.1893	1.0848
8.4	0.0437	0.0346	0.1075	0.1248	0.3305	0.3061	0.5554	0.5275	0.8416	0.8119
8.1	0.0413	0.0473	0.0987	0.0808	0.1929	0.2177	0.4261	0.4199	0.6552	0.6976
12.3	0.1475	0.1475	0.2291	0.2573	0.6521	0.5797	1.0462	1.0803	1.8984	1.7183
10.9	0.1010	0.1025	0.2375	0.2096	0.3532	0.3596	0.6297	0.6667	0.9064	0.9615

地径 (cm) Diameter	月-日 Month-day									
	04-28		05-02		05-04		05-07		05-10	
	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value	实测值 Observed value	估计值 Fitted value
12.1	1.7576	1.9599	2.3234	2.2550	2.8887	2.9165	3.7262	3.7393	4.6170	4.6134
7.3	0.8925	0.8890	1.3115	1.3193	1.6129	1.6130	1.7718	1.7560	2.0138	2.0115
9.4	1.5146	1.5124	1.9417	1.9212	2.3576	2.4078	2.8579	2.8435	3.3769	3.3597
9.0	1.1974	1.2485	1.5672	1.5483	2.2889	2.2091	3.0379	3.0896	3.7982	3.8264
13.6	2.6185	2.6263	3.7956	3.7884	4.4025	4.3969	5.4692	5.4742	6.5940	6.5937
11.1	1.7332	1.6281	2.4657	2.5169	2.8771	2.7823	3.6290	3.6563	4.2210	4.1976
8.4	1.1216	1.0673	1.6371	1.6732	1.8987	1.9251	1.9622	1.9609	2.0260	2.0166
8.1	0.8449	0.8605	1.0996	1.0945	1.3935	1.3875	1.6883	1.6584	2.0297	2.0215
12.3	2.7890	2.6494	3.3292	3.4163	3.6659	3.6116	4.1293	4.1112	4.6085	4.6177
10.9	1.4886	1.5310	2.0883	2.0524	2.4150	2.5028	2.7466	2.7283	3.0565	3.0804

从表 3 可以得出,模型估计结果非常好,误差离差平方和为 0.2518。自竹笋至幼竹,毛竹生长速度是“慢-快-慢”,呈现“S”形曲线,不同地径毛竹(笋)高生长曲线表现出相同的趋势,在这段时间里,毛竹生物量随时间推移增加非常快。其误差分布,见图 1。

从图 1 可以看出,除了个别样本点即异常数据的误差偏离较大外,大部分误差分布非常集中,而且分布幅度也比较小。随着 $T_{c_1}(D^*)$, $T_{c_2}(h^*)$, $T_{c_3}(t^*)$ 阶数的提高,估算结果还会更好,但本文所选的阶数已经能够满足精度要求。

2.2.2 与传统方法的比较

由于对毛竹生物量的研究主要集中在第 2 阶段,第 1 阶段生物量与时间有关的研究很少,因此估算毛竹第 1 阶段生物量的传统方法是修改常用模型参数^[21-22]且这种修改带有很大的为主观性。

山本式 $M = c_0 D^{c_1} H^{c_2}$ 是常用生物量估算模型,从 2.2.1 节的分析可见,毛竹第 1 阶段生物量生长呈“S”形

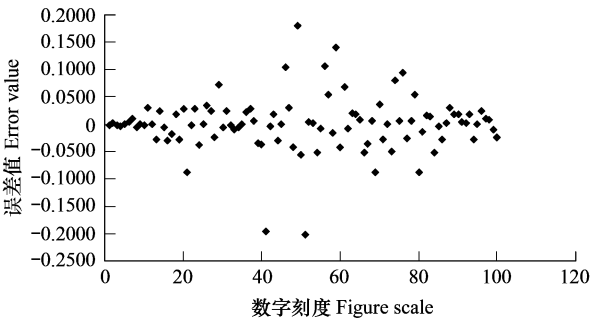


图 1 误差分布散点图
Fig. 1 Scatter plot of error's dstrbution

且与时间有关,故设 $c_1 = c_{11}e^{c_{12}t}$, $c_2 = c_{21}e^{c_{22}t}$,这时修改参数后的模型,用非线性最小二乘估算生物量的离差平方和为 5.623,确定系数 $R^2 = 0.971$ 。由此可见,本文所用方法的估算误差比传统方法提高了一个数量级。

3 结论

树种生物量模型的建立及其参数的求解算法,一直是生态工作者研究的焦点,但到目前为止,生物量的估算主要是建立在经验模型基础上,实践证明这些经验模型都有缺陷。

生物量的最常用估算方法是(非)线性最小二乘,前面提到它有许多不足,从上述估算实例发现偏最小二乘算法可以弥补这些不足。

综合本文的分析研究,得出如下的结论:

(1) 本文用 Chebyshev 正交多项式系构造了连续函数空间的一组乘积型基,从而得到树种生物量估算的统一模型,从模型的结构可以看出,常用的经验模型是本文所构建统一模型的特殊情形,而且该模型适应于与生物量有关的任意 p 个因素。

(2) 本文对所构建的生物量统一模型,应用偏最小二乘辨识方法提取了 55 个主成分,计算结果表明拟合精度很高: $S_{PRESS,h} = 31.4390$, $R^2 = 0.9987$,离差平方和为 0.2518。

(3) 偏最小二乘可以弥补(非)线性最小二乘所固有的不足,因此偏最小二乘回归确实是一种比较理想的建模方法,其适应范围非常广。

4 讨论

(1) 由于 Chebyshev 正交多项式 $T_{c_1}(x_1^*), T_{c_2}(x_2^*), \dots, T_{c_p}(x_p^*)$ 的定义域为 $[-1, 1]$, 因此随着 $T_{c_1}(x_1^*), T_{c_2}(x_2^*), \dots, T_{c_p}(x_p^*)$ 阶数的提高, $T_{c_1}(x_1^*), T_{c_2}(x_2^*), \dots, T_{c_p}(x_p^*)$ 的值就会减小,也就是树种统一生物量模型自变量的值在减小,这就相当于在区间 $[-1, 1]$ 上进行的数值插值,模型阶数越高,所插入的点就越多,估算结果越符合实际,这与树木的树干解析与树木生长原理是相一致的,如果这个插值点正好落在树木的年轮上,则本文所建的模型就非常真实的反映树木生长状况,对生物量的估计更加准确真实。

(2) 本文所用方法与其它插值方法的区别是:其它插值方法是在生物量模型自变量原始定义域上进行的,而本文所用的插值方法是先进行变量的区间变换,缩小模型自变量的定义域,然后在缩小后的定义域内进行插值。则当插入相等的点数时,用本文所述的方法要优于其它插值方法。

(3) 本文所构建的统一模型实际上是对树种理论生物量模型的无限逼近。

References:

- [1] Meng X Y. She guang hui. Forest Measurements. Beijing: China Forestry Publishing House, 2004. 136.
- [2] Jiang Z H, Fan S H, Feng H X. Biomass and Distribution Patterns of Populus x xiaohei Plantation in Sandy Land of North China. Scientia Silvae Sinicae, 2007, 43(11): 15-20.
- [3] Zheng R M, Zheng Y S, Ding M F, Ling G X, Chen L G, Wang J D. Study on the Biomass Model of Pleioblastus amarus. Journal of Fujian College of Forestry, 2003, 23(1): 61-64.
- [4] Zeng H Q, Liu Q J, Feng Z W. Estimation models of understory shrub biomass and their applications in red soil hilly region. Chinese Journal of Applied Ecology, 2007, 18(10): 2185-2190.
- [5] You Z D. A Study on Biomass Structure of Individual Plant of Shoot-used Dendrocalamus latiflorus Planted on Mountain. Acta Agriculturae Universitatis Jiangxiensis, 2002, 24(6): 806-808.
- [6] Li J H, Li C J, Peng S K. Study on the Biomass Expansion Factor of Poplar Plantation. Journal of Nanjing Forestry University (Natural Sciences Edition), 2007, 31(4): 36-40.
- [7] Tang S Z, Zhang H R, Xu H. study on establish and estimate method of compatible biomass model. Scientia Silvae Sinicae, 2006, 36: 19-27.
- [8] Cheng T R, Feng J, Ma Q Y, Feng Z K. Linear models with compatibility of stand biomass based on the forest resource inventory data. Journal of Beijing Forestry University, 2007, 29(5): 110-113.
- [9] Zhang H R, Tang S Z, Wang F Y. Study on Establish and Estimate Method of Biomass Model Compatible with Volume. Forest Research, 1999, 12(1): 53-59.

- [10] Lin C S. Numerical Analysis. Beijing: Science Press, 2006. 197 — 202.
- [11] Yan Q J. Numerical Analysis. Beijing: , Beihang University Press, 2001. 122 — 125.
- [12] Wang H W, Wu A B, Meng J. Partial Least-Squares Regression — Linear and Nonlinear Methods. Beijing: National Defence Industry Press, 2006. 275 — 276.
- [13] Wei B C. Modern non-linear regression analysis. Nanjing: SouthEast University Press, 1989. 7 — 10.
- [14] Lindberg Walter, Persson J A, Wold S. Partial least squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. Anal Chem, 1983, 55: 643.
- [15] Malt house E C, Thmhana A C, Mah R S H. Nonlinear partial least squares. Comput Chem Engng, 1997, 21(8): 875 — 890.
- [16] Zhou G M, Jiang P K. Density, Storage and Spatial Distribution of Carbon in Phyllostachy pubescens Forest. Scientia Silvae Sinicae, 2004, 40(6): 20 — 24.
- [17] Chen S L, Wu B L, Wu M, Zhang D M, Cao Y H, Yang Q P. A Study on Aboveground Biomass of Young Bamboo Stands of Phyllostachys Pubescens in Degenerative Hill Soil Area. Acta Agriculturae Universitatis Jiangxiensis Aug. , 2006, 26(4): 527 — 531.
- [18] Peng Z Q, Lin Y M, Liu J B, Zhou X H. Biomass Structure and Energy Distribution of Phyllostachys heterocycla cv. pubescens Population. Journal of Xiamen University (Natural Science), 2002, 41(5): 579 — 583.
- [19] He D J, Hong W, Wu C Z, *et al.* Study on biomass and energy distribution of natural Phyllostachys heterocycla cv. Pubescens in Wuyi Mountains and its comparison with high-yield forest. Acta Botanica Boreali-occidentalia Sinica, 2003, 23(2): 291 — 296.
- [20] Chen Y M. Biomass Structure Comparison on Moso Bamboo Mixed Forest Communities. Central South Forest Inventory and Planning, 2007, 26(3): 18 — 23.
- [21] Cen J Y. Study on Two-Way Tree Volume Dynamic Model of Eucalyptus Plantations in Guangxi. Journal of South China Agricultural University, 2007, 28(1): 91 — 95
- [22] Zhou G M, Liu E B, Liu A X, Zhou Y F. The algorithm update of Weibull Distribution parametric identification and its application on measuring the distribution of diameter and age of Moso bamboo forests in Zhejiang Province. Acta Ecologica Sinica, 2006(26)9: 2918 — 2924.

参考文献:

- [1] 孟宪宇, 余光辉. 测树学. 北京: 中国林业出版社, 2004. 136.
- [2] 江泽慧, 范少辉, 冯慧想, 等. 华北沙地小黑杨人工林生物量及其分配规律. 林业科学, 2007, 43(11): 15 ~ 20.
- [3] 郑容妹, 郑郁善, 丁闽锋, 等. 苦竹生物量模型的研究. 福建林学院学报, 2003, 23(1): 61 ~ 64.
- [4] 曾慧卿, 刘琪璟, 冯宗炜, 等. 红壤丘陵区林下灌木生物量估算模型的建立及其应用. 应用生态学报, 2007, 18(10): 2185 ~ 2190.
- [5] 尤志达. 山地麻竹笋用林单株生物量结构研究. 江西农业大学学报(自然科学版), 2002, 24(6): 806 ~ 808.
- [6] 李建华, 李春静, 彭世揆. 杨树人工林生物量估计方法与应用. 南京林业大学学报(自然科学版), 2007, 31(4): 36 ~ 40.
- [7] 唐守正, 张会儒, 胥辉. 相容性生物量模型的建立及其估计方法研究. 林业科学, 2006, 36: 19 ~ 27
- [8] 程堂仁, 冯菁, 马钦彦等. 基于森林资源清查资料的林分生物量相容性线性模型. 北京林业大学学报, 2007, 29(5): 110 ~ 113.
- [9] 张会儒, 唐守正, 王奉瑜. 与材积相容的生物量模型的建立及其估计方法研究. 林业科学研究, 1999, 12(1): 53 ~ 59.
- [10] 林成森. 数值分析. 北京: 科学出版社, 2006. 197 ~ 202.
- [11] 颜庆津. 数值分析. 北京: 北京航空航天大学出版社, 2001. 122 ~ 125.
- [12] 王惠文, 吴载斌, 孟洁. 偏最小二乘回归的线性与非线性方法. 北京: 国防工业出版社, 2006. 275 ~ 276.
- [13] 韦博成著. 近代非线性回归分析. 南京: 东南大学出版社, 1989. 7 ~ 10.
- [16] 周国模, 姜培坤. 毛竹林的碳密度和碳贮量及其空间分布. 林业科学, 2004, 40(6): 20 ~ 24.
- [17] 陈双林, 吴柏林, 吴明, 等. 退化低丘红壤新造毛竹林地上部分生物量的研究. 江西农业大学学报, 2006, 26(4): 527 ~ 531.
- [18] 彭在清, 林益明, 刘建斌, 等. 福建永春毛竹种群生物量和能量研究. 厦门大学学报(自然科学版), 2002, 41(5): 579 ~ 583.
- [19] 何东进, 洪伟, 吴承祯, 等. 武夷山毛竹天然林生物量与能量分配规律及其与人工林的比较研究. 西北植物学报, 2003, 23(2): 291 ~ 296.
- [20] 陈岩明. 毛竹混交林群落生物量结构比较. 中南林业调查规划, 2007, 26(3): 18 ~ 23.
- [21] 岑巨延. 广西桉树人工林二元立木材积动态模型研究. 华南农业大学学报, 2007, 28(1): 91 ~ 95.
- [22] 周国模, 刘恩斌, 刘安兴, 等. Weibull 分布参数辨识改进及对浙江毛竹林胸径年龄分布的测度. 生态学报, 2006, 26(9): 2918 ~ 2925.