

生物量精确估算模型与参数辨识方法及应用

刘恩斌¹, 李永夫¹, 周国模^{1,*}, 施拥军¹, 莫路锋²

(1. 浙江林学院环境科技学院, 临安 311300; 2. 浙江林学院信息工程学院, 临安 311300)

摘要: 从生物量模型的构建与参数辨识方法的改进对生物量进行精确估算。用 Chebyshev 多项式系的组合构建了 p 维连续函数空间的一组乘积型基, 进而建立了生物量估算统一模型, 它具有如下特点: (1) 可以克服常用生物量估算模型的经验性、不稳定性、不通用性及对生物量影响因素适应性差的特点, (2) 它适合于影响生物量的任何因素, 故适应范围非常广且很稳定, (3) 可根据实际需要及估算精度确定影响生物量的因素及其阶数大小, (4) 模型对生物量的估算相当于在区间 $[-1, 1]$ 上进行的数值插值, 变量阶数越高, 所插入的点就越多, 估算结果越符合实际, 整个估算的插值过程与树木的树干解析与树木生长原理是相一致的。对所建模型的参数辨识方法做了探讨, 经典最小二乘算法是生物量估算的最常用参数辨识方法, 由于它本身固有的一些缺陷使常用最小二乘的估算精度与使用范围受到很大的限制, 现代多元统计分析的偏最小二乘算法可以克服常用最小二乘的缺陷, 但在提取成分时仍具有不足, 针对偏最小二乘的缺陷本文对它做了改进, 改进算法即能克服偏最小二乘的不足还能使估算精度大大提高。用 2 个案例对 3 种生物量估算方法做了对比分析, 结果表明生物量估算统一模型与偏最小二乘改进算法精度最高, 其生物量估计误差在零附近排成一条直线。

关键词: 生物量估算统一模型; 偏最小二乘改进算法; 生物量

Ascertainment and application of precision model in biomass estimate as well as parameter identification

LIU Enbin¹, LI Yongfu¹, ZHOU Guomo^{1,*}, SHI Yongjun¹, MO Lufeng²

1 School of Environment Technology, Zhejiang Forestry University, Linan 311300, China

2 School of Information Engineering, Zhejiang Forestry University, Linan 311300, China

Abstract: This paper explores further the precision estimate of biomass by making improvement on the way biomass models are constructed and parameters identified. To that end, a unified model is set up to estimate the biomass by constructing a group of product-type base of p -dimensional continuous function space through Chebyshev polynomial combination. This model has the following characteristics: (1) It can overcome the experientialism, instability, limited applicability and poor adaptability to biomass-affecting factors ridden in conventional models in biomass estimate. (2) It has wider and more stable applicability given that this model is applicable to any of the biomass-influencing factors. (3) It can determine on the factors that affect the biomass and the size of the order according to actual needs and required accuracy of estimate in different cases. (4) This model works in a way similar to number interpolation between the interval $[-1, 1]$, the higher order the variables go, the more points should be inserted, the more realistic the estimate becomes. The above interpolation process in estimation abides by the same principle by which tree growth are measured based on the tree trunk analysis. Moreover, it constitutes an equally important task to find best method in identifying parameters for each estimate model. So far, the most commonly used method in parameter identification in term of biomass estimate is classic least squares algorithm. However, because of its inherent defects, classic least squares algorithm is bounded in accuracy and application. Though the partial least-squares algorithm of modern multivariate statistical analysis can somewhat overcome the shortcomings of traditional least-square, it is still not perfect in abstracting constituents. In view of that problem, this thesis has made improvement on

基金项目: 国家自然科学基金资助项目(30771715, 30900190); 国家林业局 948 项目(2008-4-49)

收稿日期: 2009-03-07; **修订日期:** 2009-06-07

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

modern multivariate statistical analysis to increase the accuracy of calculation. As a result, not only are the shortcomings of the partial least-square overcome, but the accuracy in estimate is raised substantially. By comparing three methods in calculating biomass in two cases, it proved that the unified model in biomass estimate, together with improved partial least square algorithm can render the most accurate result so much so that the biomass-estimate error formed a straight line closely along the zero axis.

Key Words: unified model for biomass estimation; the improved partial Least-Squares Algorithm; 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,精确估算树种的生物量十分重要。生物量的精确估算包括 2 方面的内容。

第一,构建有完备理论基础的生物量模型。目前有关生物量估算模型可以分为以下 3 种类型:(1)多项式模型^[1-4],这类模型以孟宪宇^[4]、迈耶^[4]、赵克升^[4]等提出的最为典型,对影响生物量自变量的选择比较灵活,从这方面来说多项式模型的适应性比较强,但其结构多样,没有统一的形式,反应了多项式模型的构建带有很多人为主观因素,其构建理论科学依据不足,在实际应用中不具有通用性;(2)山本式模型及其变型^[3-8],日本山本和藏 1918 年提出山本式模型以来,目前在实践中已成为应用最多的生物量估算模型,但也是经验性的,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^[4],这说明山本式模型存在缺陷,至于变型出来的模型,一方面其结构带有很大的为主观性,使得不同的人可构造出不同形式的模型,这也从侧面反应了山本式模型的适应性差,人们试图想办法克服它本身固有的缺陷,但始终没有摆脱模型结构变化本身具有的主观随意性较大的特点;(3)相容性生物量模型^[9-13],这类模型是用建联立方程组的方法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,但其各维量模型的构建实质仍是建立在经验模型基础上的。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。

第二,估算方法的选择与改进。最小二乘是最为经典的参数辨识方法,在生物量估算的以上 3 类模型中用的最多,常用最小二乘存在一些缺陷:(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,需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实质上是对非线性模型采取的线性逼近(因取至泰勒逼近展开式的一阶项),在拟合模型时需要给定模型的初始值,如果初值选取不当,迭代往往不收敛。为了克服这些缺陷,许多学者对它做了非线性逼近与初值选取评价方面的研究^[14-15],但仍然需要选取迭代初值。常用最小二乘的这些缺陷使它在实际应用中受到很大的限制,因此选取能克服常用最小二乘算法以上 4 个缺陷的参数辨识方法显得非常重要。基于常用生物量估算模型及参数辨识方法的以上缺陷,本文试图在理论上寻找生物量统一模型的构建方法及参数辨识的改进算法。

1 生物量统一模型的构建

把生物量统一模型看作是连续函数空间中的一个元素;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;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^[16-17]。本文按照这种思路用 Chebyshev 正交多项式系来构建生物量统一模型。

1.1 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:

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

相邻的三个 Chebyshev 多项式具有如下递推关系式:

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

从(2)式可以看出, Chebyshev 多项式系是可列的(与自然数相对应的无穷序列称为可列序列,当 $n=0$ 时不是自然数,但一个有限序列与可列序列的并仍是可列的),从而可以作为一维连续函数空间的一组基(一维连续函数空间基的个数是可列的,如一维连续函数空间的基 $\{1, x, x^2, \dots, x^m, \dots\}$ 是可列的)。

1.2 多元生物量统一模型的构建

由于 Chebyshev 多项式系只有一个自变量,而本文所构建的模型是多元生物量统一模型,故要能反映生物量与 p 个影响生物量因素的函数关系,需要用 p 个 Chebyshev 多项式系的组合来构建 p 维连续函数空间的基。

从以上 Chebyshev 多项式系的构成与性质可以看出,其自变量的定义域为 $[-1, 1]$,而影响生物量的因素其取值范围不完全在 $[-1, 1]$ 内,这就得进行变量的区间变换,其公式为:

$$x = \frac{2x_{\text{原}} - (a + b)}{b - a}, x_{\text{原}} \in [a, b] \quad (3)$$

式中, $x_{\text{原}}$ 为与生物量有关的因素如胸径、树高等。

设与生物量有关的 $p(p \geq 1)$ 个因素经(3)式变换后的变量记为 $x_1, x_2, \dots, x_p$, 则对于定义在 p 维空间 $\{x_1, x_2, \dots, x_p\}$ 上的多元生物量函数,由于 $T_{c_1}(x_1)T_{c_2}(x_2)\dots T_{c_p}(x_p) (c_1, c_2, \dots, c_p = 0, 1, 2, \dots)$ 是两两线性无关的。所以它可以构成 p 维连续函数空间的一组基(可列个可列序列的并仍是可列序列),称为乘积型基。这样树种生物量估算统一模型可表示为:

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

但在实际应用时,当 k 等于某个数 M 时既满足精度要求,因此(4)式可写为:

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

(5)式就是本文要构建的树种生物量估算统一模型。当估算出(5)式中的每个参数 $a_k (k = 0, 1, \dots, M)$ 后,再用 $[(b-a)x_i + a + b]/2 (i = 1, 2, \dots, p)$ 代替 x_i , 这样就可以把(5)式中的每个 $x_i (i = 1, 2, \dots, p)$ 还原为 p 个生物量的影响因素。

从(5)式可以看出:(1)树种生物量统一模型是一种拟线性模型(可以线性化的模型),有完备的理论基础,随着 M 的无限增大可以无限的逼近树种实际生物量模型。这就克服了常用生物量模型的经验性、主观性特点;(2)常用生物量模型对影响森林生物量的主导因素限制太死即只有胸径、树高等少数几个因素,其实树种生物量与许多生物学因素和非生物学因素密切相关。如果把年龄、气温和降水等许多生物和非生物学因素的协同作用考虑进去,则随着影响生物量因素的增加,估算树种生物量时需要对常用生物量模型进行改进,但目前的改进方法仍停留在主观修改参数的阶段^[7-8,15],这种修改缺乏科学依据且主观随意性较大。而(5)式表示的生物量模型适应于任意 p 个与生物量有关的影响因素,可以克服主观修改参数的缺陷;(3)树种生物量统一模型中包含多个生物量的影响因子且取较高阶数,这时对生物量的估算考虑比较全面且估算精度很高,从而便于找到各个影响因素与生物量之间以及各个影响因素之间的相互关系,但这样做使得统一模型中所含变量参数较多,这时可采用递阶偏最小二乘算法高度浓缩变量简化模型;(4)如果由于实验条件、人力及物力等限制很难获得某些影响生物量因素的实测数据,根据研究目的,结合实际,当不考虑这些影响因素时对生物量的研究不会产生过大的影响,此时 p 就可以取较小的值,统一模型中所含参数个数成倍减少。如果研究目的与实际需求对生物量估算精度要求不高,此时可以取较低的变量阶数,则统一模型中所含参数个数还能减少。因此(5)式所示的生物量统一模型可以根据研究目的、实际需要与精度要求,非常灵活地选取生物量影响因素个数及模型各变量所取阶数。

前面提到树种生物量的精确估算不久在于其理论模型的构建,还在于参数辨识方法的选择与改进。偏最小二乘算法是集主成分分析、典型相关分析、回归分析于一体的现代多元统计分析,以处理多变量、海量数据而著称,可以克服常用最小二乘的以上 4 个不足。

2 偏最小二乘算法的基本思想

现对它的基本建模思想概述如下:设 X 为所有的自变量组成的矩阵; Y 为所有的因变量组成的矩阵;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,这时 x_1 所包含的原数据的变异信息达到最大即 $\text{Var}(x_1) \rightarrow \max$;在 Y 中提取第 1 主成分,用 y_1 来表示,这时 y_1 所包含的原数据的变异信息达到最大即 $\text{Var}(y_1) \rightarrow \max$ 。再结合典型相关分析的方法,得到下式:

$$\max r(x_1, y_1) \text{ s. t } \begin{cases} x_1^T x_1 = 1 \\ y_1^T y_1 = 1 \end{cases} \quad (6)$$

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^[18-20]。具体实施步骤详见参考文献^[18]。

从以上偏最小二乘算法的基本思想可以看出,偏最小二乘回归中要求:

$$\text{Cov}(x_i, Y) = \sqrt{\text{Var}(x_i) \text{Var}(Y)} r(x_i, Y) \rightarrow \max \quad (7)$$

但是使 $\text{Cov}(x_i, Y)$ 达到最大并不一定能使 $r(x_i, Y)$ 与 $\text{Var}(x_i)$ 同时达到最大。这时可能会出这样的情况:主成分 x_i 有 $\text{Var}(x_i) \rightarrow \max$,但其对应的 $r(x_i, Y)$ 却不是最大,而另一个 x_i^* 有较小的 $\text{Var}(x_i^*)$,但却有 $r(x_i^*, Y) \rightarrow \max$ 即 $r(x_i^*, Y) > r(x_i, Y)$,按照偏最小二乘的基本原理 x_i 会被选中而 x_i^* 会被错误地落选。这是偏最小二乘算法固有的缺陷,其缺陷产生的原因主要是解释矩阵中含有大量与响应向量无关的信息,这样入选的成分虽然具有大的协方差,但对响应变量缺乏解释能力。因此要使生物量得到精确估算,还必须对适应范围较广的偏最小二乘算法做进一步的改进,为了能在理论得出改进的算法,在此先介绍一个重要结论。

3 重要结论

假设 $X_{n \times p}$ 、 $Y_{n \times q}$ 是原始数据经过标准化后的矩阵, X_k 、 Y_k 是 X 、 Y 经 k 次信息提取后的残差矩阵,由于 $x_1 = Xw_1$ (w_1 是 X 的第一个轴,是单位向量), $y_1 = Yc_1$ (c_1 是 Y 的第一个轴,是单位向量),根据偏最小二乘的基本原理得:

$$\max < Xw_1, Yc_1 > \text{ s. t } \begin{cases} w_1^T w_1 = 1 \\ c_1^T c_1 = 1 \end{cases} \quad (8)$$

这就相当于在相同约束条件下求 $w_1^T X^T Y c_1$ 的最大值。采用拉格朗日算子经系列推导得:

$$\begin{cases} X^T Y Y^T X w_1 = \theta_1^2 w_1 \\ Y^T X X^T Y c_1 = \theta_1^2 c_1 \end{cases} \quad (9)$$

从(9)可以看出, w_1 、 c_1 分别为 $X^T Y Y^T X$ 与 $Y^T X X^T Y$ 的特征向量。

由于 $\theta_1^2 = \|X^T Y\|^2$, 所以 $w_1 = \frac{X^T Y}{\|X^T Y\|}$, 类似的有 $w_k = \frac{X_{k-1}^T Y_{k-1}}{\|X_{k-1}^T Y_{k-1}\|}$ 。

因此 $X^T Y Y^T X w_1 = X^T Y Y^T X \frac{X^T Y}{\|X^T Y\|} = Y^T X X^T Y w_1$, 这说明 $Y^T X X^T Y$ 是 $X^T Y Y^T X$ 的特征值,而且可以证明它是

唯一的非零特征值^[21]。这就是改进算法所需重要结论。

4 改进的偏最小二乘算法

改进算法首先把 $X^T Y Y^T X (X_{n \times p}, Y_{n \times q})$ 的 $p-1$ 个零特征根对应的特征向量 $w_2, w_3, \dots, w_p$ 组成一个新的矩阵 D , 可以看出 D 与 Y 是正交的;其次把与 Y 正交的所有向量写成 XD 的线性组合 $XD\beta$ ($\beta^T \beta = 1$);再次把使

$XD\beta$ 方差较大的成分去掉,这些成分相应于 $D^T X^T XD$ 的前 s 个最大特征值 $\lambda_1, \lambda_2, \dots, \lambda_s$ 对应的特征向量,这里 s 的选取就是要使剩下的成分不显著,可根据 $\sum_{i=1}^s \lambda_i / \sum_{i=1}^{p-1} \lambda_i \approx 0.99$,把这 s 个特征值对应的特征向量作为列组成一个新矩阵 F 并令 $U = XDF$,则 U 正交于 Y ,再令 $XG = X(I - DF(U^T U)^{-1} U^T X)$;最后把 XG 作为自变量矩阵应用偏最小二乘对因变量 Y 进行回归模拟。从这里可以看出,改进的算法是把自变量矩阵 X 变成 XG 的偏最小二乘回归。

现对改进算法能克服偏最小二乘缺陷的原因加以解释:要尽可能多的扣除解释矩阵中与响应变量无关的信息就是寻找能使 $XD\beta$ 的方差达到最大的 β ,而能使 $XD\beta$ 的方差达到最大的向量 β 就是矩阵 $D^T X^T XD$ 的最大的特征值对应的特征向量,因此,寻找那些具有大方差且与响应向量线性无关的信息就等价于寻找矩阵 $D^T X^T XD$ 的具有较大特征值的标准特征向量 $\beta_1, \beta_2, \dots, \beta_s$,由此可知矩阵 F 就是那些解释矩阵中与响应向量无关的信息。最后将解释矩阵 X 投影到这些标准特征向量所张成的正交补空间上去,投影后得到的矩阵就是扣除了那些与 Y 无关信息的解释矩阵,由正投影算子的定义与性质^[22]可得 XG 就是投影后得到的矩阵。

偏最小二乘改进算法不需要提取 XG 的所有成分,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改进算法回归建模时必须要进行交叉有效性检验,其实施过程见文献^[18]。

下面结合 2 个实例对本文提出的统一模型与改进偏最小二乘进行分析。

5 应用案例

5.1 案例 1

5.1.1 数据来源

调查样地设在福建省九龙江北溪中游、华安县北部的西陂国有林场,该林厂分为路亭、新建、西陂、洋坑四个工区。为使调查具有很好的代表性,在 4 个工区中选择不同立地、不同年龄、不同密度的杉木林分,设置 $25\text{m} \times 25\text{m}$ 9 块样地。

5.1.2 样木选择

选择样木的目的在于以较少的样本推断总体的特征。因此凡供选取的样木必须要能代表总体的特性。一般来说,取样的方法要根据总体的均质性、人力、财力、时间及允许误差来决定。为此本次调查依据典型选取方法,在 9 块样地中选择 26 株样木,量测胸径、树高、平均高、平均胸径并进行总生物量的测定。

5.1.3 结果与对比分析

由于实验条件限制与实际需求,本案例只考虑胸径(D)与树高(H)2 个影响生物量的因素且要求有较高的估算精度。

(1) 山本式二元生物量模型 $M = c_0 D^{c_1} H^{c_2}$ 与 Marquardt 算法结果:离差平方和为 5434.51108, $R^2 = 0.9770$ 。最后 5 步迭代结果见表 1。

表 1 最后 5 次迭代过程
Table 1 Ultimate five iteration process

迭代数 Iteration number	离差平方和 Residual sum of squares	参数 Parameter		
		c_0	c_1	c_2
1	5434.5113	0.1359	0.9684	1.3633
2	5434.5113	0.1359	0.9684	1.3633
3	5434.5111	0.1359	0.9685	1.3631
4	5434.5111	0.1359	0.9685	1.3631
5	5434.5111	0.1359	0.9685	1.3630

从表 1 可以看出,参数迭代已经稳定。

(2) 生物量估算统一模型 $f(H^*, D^*) = \sum_{k=0}^M a_k T_{c_1}(H^*) T_{c_2}(D^*)$ (H^* 为胸径 H 经区间变换后的变量, D^*

树高 D 经区间变换后的变量,以下类似)与偏最小二乘算法结果:当 $c_1, c_2 = 0, 1, \dots, 5, M = 35$ (共有 36 个量,其中第一个是常量),根据交叉有效性检验从 35 个成分中提取了 23 个主成分,这时离差平方和为:111.2942, $R^2 = 0.9920$ 。模型各参数见表 2。

表 2 杉木生物量统一模型参数
Table 2 Parameters of Taxodiaceae's Biomass unite 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
a_0	-329.9801	a_{12}	15.2843	a_{24}	-92.8709
a_1	-113.8759	a_{13}	24.6859	a_{25}	-393.1794
a_2	-83.4227	a_{14}	727.6378	a_{26}	-4.0632
a_3	-346.8658	a_{15}	153.1785	a_{27}	178.0785
a_4	-275.9666	a_{16}	32.9323	a_{28}	27.9816
a_5	-122.3140	a_{17}	18.8429	a_{29}	13.3669
a_6	-259.6467	a_{18}	259.0292	a_{30}	-25.2752
a_7	240.2070	a_{19}	-20.8958	a_{31}	-80.2427
a_8	-308.6638	a_{20}	-520.3732	a_{32}	-617.9830
a_9	269.2178	a_{21}	-18.0157	a_{33}	-142.6310
a_{10}	89.7553	a_{22}	-139.6355	a_{34}	-215.7154
a_{11}	-331.3210	a_{23}	-189.7664	a_{35}	-130.7727

(3)生物量估算统一模型 $f(H^*, D^*) = \sum_{k=0}^M a_k T_{c_1}(H^*) T_{c_2}(D^*)$ 与偏最小二乘改进算法结果:当 $c_1, c_2 = 0, 1, \dots, 5, M = 35$,根据交叉有效性检验从 35 个成分中提取了 12 个主成分,这时离差平方和为:11.2547, $R^2 = 0.9996$,模型各参数见表 3,实测值与估计值的对比见表 4。

表 3 杉木生物量统一模型参数
Table 3 Parameters of Taxodiaceae's Biomass unite 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
a_0	-0.0007	a_{12}	362.5327	a_{24}	-369.2062
a_1	-8.0762	a_{13}	140.4158	a_{25}	-447.8451
a_2	-183.5355	a_{14}	-55.4136	a_{26}	37.4171
a_3	-467.2217	a_{15}	139.3366	a_{27}	192.5231
a_4	-764.0033	a_{16}	-194.4644	a_{28}	187.9717
a_5	-62.8948	a_{17}	132.9404	a_{29}	-53.6732
a_6	7.3664	a_{18}	72.6844	a_{30}	141.8118
a_7	-462.3769	a_{19}	78.1244	a_{31}	47.7751
a_8	323.2425	a_{20}	-772.8236	a_{32}	-690.5563
a_9	809.9172	a_{21}	-195.1425	a_{33}	-253.5562
a_{10}	286.9931	a_{22}	125.4327	a_{34}	-240.6766
a_{11}	-434.2631	a_{23}	-118.8844	a_{35}	-276.9517

(4)为了更直观的比较说明 3 种估算方法的拟合精度,把 3 种方法对每株样本的估算误差绘成图,见图 1。

从图 1 可以看出,生物量估算统一模型与改进偏最小二乘算法估算精度最高,其误差在零附近几乎排成一直线,估算精度比生物量估算统一模型与偏最小二乘算法有了很大的提高。

表 4 杉木生物量估测值与实测值

Table 4 Observed values and Fitted values of Taxodiaceae's Biomass

样本号 Sample number	实测值 Observed value	估计值 Fitted value	样本号 Sample number	实测值 Observed value	估计值 Fitted value
1	102.5486	102.5898	14	326.0703	326.0734
2	207.8506	207.8195	15	291.4030	291.4010
3	119.5171	119.3864	16	253.0200	253.0211
4	114.7233	114.7134	17	270.0450	270.0456
5	155.4738	155.5154	18	8.1360	8.0815
6	18.4923	18.2692	19	3.1140	3.1003
7	5.1160	5.2115	20	77.9120	78.3498
8	28.6673	28.6593	21	101.5105	101.4258
9	13.8945	13.7536	22	30.6720	30.6699
10	38.8085	38.5291	23	23.8534	21.6660
11	42.3077	42.6645	24	26.0428	28.4452
12	105.1782	105.2038	25	29.4236	29.4740
13	15.3782	15.5035	26	95.3402	94.9260

5.2 案例 2

5.2.1 数据来源

2005 年 4 月,毛竹笋高度为 30cm 左右的出笋期,在地形条件基本一致的毛竹林中,选择地径为 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及高度的毛竹(笋),作为样木株,将样木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)梢部均匀取样带回实验室,测定其样品鲜质量后,将其在 105℃ 杀青 30min,后在 70℃ 下烘干,测定干物质量。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。

5.2.2 结果分析

从实测数据(表 6),毛竹在这 1 个月的时间生物量随时间变化很大,故在这一个月的生长阶段毛竹生物量与时间有关,这样毛竹生物量的影响因子就有地径 D (不同的毛竹其地径不同,故把地径作为毛竹生物量的影响因素)、竹高 H 与时间 t (t 取值为第 1 次实测时 $t=0$, t 以后的取值为每次实测时间与第 1 次实测时间的间隔天数)。而到目前为止还没有生物量与时间有关的模型,这时常用的方法是人为主观修改模型参数。

(1) 山本式模型变型与 Marquardt 算法:由表 6 实测数据可以得出,毛竹在这一个月的时间内生长速度是“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。所以在山本式模型 $M = c_0 D^{c_1} H_{c_2}$ 的基础上令 $c_1 = c_{11} e^{c_{12} t}$,修改参数后的模型用 Marquardt 算法估算生物量,此时离差平方和为 5.623,确定系数 $R^2 = 0.971$ 。

(2) 生物量估算统一模型 $f(H^*, D^*, t^*) = \sum_{k=0}^M a_k T_{c_1}(H^*) T_{c_2}(D^*) T_{c_3}(t^*)$ 与偏最小二乘算法结果:当 $c_1, c_2, c_3 = 0, 1, \dots, 5, M = 215$,根据交叉有效性检验从 215 个成分中提取了 55 个主成分,此时离差平方和为

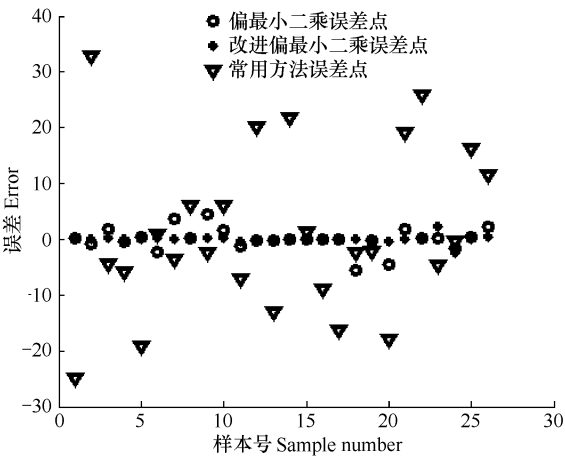


图 1 误差分布散点图
Fig. 1 Scatter plot of error's distribution

0.2518, $R^2 = 0.9987$ 。

(3) 生物量估算统一模型 $f(H^*, D^*, t^*) = \sum_{k=0}^M a_k T_{c_1}(H^*) T_{c_2}(D^*) T_{c_3}(t^*)$ 与偏最小二乘改进算法结果: 当 $c_1, c_2, c_3 = 0, 1, \dots, 5, M = 215$, 根据交叉有效性检验从 215 个成分中提取了 54 个主成分, 此时离差平方和为 1.0679×10^{-4} , $R^2 = 0.9999$, 模型各参数见表 5, 实测值与估计值的比较见表 6。

表 5 毛竹生物量统一模型参数
Table 5 Parameters of Moso Bamboo's Biomass unite 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	2.0380×10^{-7}	a_{54}	1.0894	a_{108}	0.4161	a_{162}	0.2060
a_1	0.0434	a_{55}	3.7016	a_{109}	0.2632	a_{163}	-1.3699
a_2	0.5975	a_{56}	-1.4447	a_{110}	-0.0767	a_{164}	1.1440
a_3	1.1194	a_{57}	-0.2201	a_{111}	-1.1890	a_{165}	0.7666
a_4	1.6592	a_{58}	2.2262	a_{112}	-0.7178	a_{166}	-1.5659
a_5	-0.4862	a_{59}	-0.2220	a_{113}	1.2773	a_{167}	-0.0496
a_6	0.1895	a_{60}	-1.8798	a_{114}	-0.2714	a_{168}	-0.9340
a_7	0.4301	a_{61}	-0.0128	a_{115}	0.1497	a_{169}	0.1263
a_8	0.5089	a_{62}	2.8736	a_{116}	-0.4659	a_{170}	-0.9503
a_9	-0.1565	a_{63}	-0.0760	a_{117}	0.4532	a_{171}	-0.0039
a_{10}	-0.1782	a_{64}	-0.3745	a_{118}	-0.5165	a_{172}	1.7845
a_{11}	0.7158	a_{65}	1.0576	a_{119}	-0.5860	a_{173}	0.3343
a_{12}	0.2816	a_{66}	-0.2882	a_{120}	0.0501	a_{174}	-0.1563
a_{13}	-0.0019	a_{67}	-0.3469	a_{121}	-0.7577	a_{175}	-0.7882
a_{14}	-0.3265	a_{68}	-0.7579	a_{122}	-0.0487	a_{176}	0.7527
a_{15}	0.0838	a_{69}	1.6733	a_{123}	1.4425	a_{177}	-1.8140
a_{16}	-0.2738	a_{70}	-0.6827	a_{124}	1.5955	a_{178}	0.1758
a_{17}	1.4670	a_{71}	-0.8441	a_{125}	-2.4031	a_{179}	-0.7068
a_{18}	-0.8820	a_{72}	0.2344	a_{126}	0.2724	a_{180}	-0.8516
a_{19}	-0.5246	a_{73}	0.9576	a_{127}	0.3605	a_{181}	-0.4693
a_{20}	0.4012	a_{74}	-0.1842	a_{128}	-0.0559	a_{182}	-0.1866
a_{21}	0.1280	a_{75}	-0.2272	a_{129}	0.0996	a_{183}	-0.2409
a_{22}	0.8264	a_{76}	0.0504	a_{130}	-0.0448	a_{184}	-0.1480
a_{23}	-1.2947	a_{77}	-1.0858	a_{131}	1.4532	a_{185}	1.0691
a_{24}	-0.1204	a_{78}	1.0675	a_{132}	-0.1421	a_{186}	-0.3334
a_{25}	0.4733	a_{79}	-0.4129	a_{133}	-0.6384	a_{187}	-0.2454
a_{26}	-1.1540	a_{80}	-0.8788	a_{134}	1.5936	a_{188}	0.6142
a_{27}	-0.9120	a_{81}	0.2040	a_{135}	0.5750	a_{189}	-1.1083
a_{28}	-0.1099	a_{82}	1.6203	a_{136}	-0.8280	a_{190}	-1.0650
a_{29}	0.0990	a_{83}	1.0852	a_{137}	0.5357	a_{191}	-0.3481
a_{30}	1.6763	a_{84}	-0.0400	a_{138}	-0.8001	a_{192}	1.3071
a_{31}	-1.2274	a_{85}	-0.2148	a_{139}	0.8357	a_{193}	1.1809
a_{32}	-0.1139	a_{86}	-0.0280	a_{140}	0.8828	a_{194}	-1.0708
a_{33}	-0.6740	a_{87}	0.8675	a_{141}	-0.4797	a_{195}	0.5272
a_{34}	-0.5548	a_{88}	-0.3090	a_{142}	0.8360	a_{196}	0.3956
a_{35}	-0.0926	a_{89}	-0.1421	a_{143}	0.5717	a_{197}	0.0092
a_{36}	0.0621	a_{90}	-0.7335	a_{144}	0.0926	a_{198}	0.3069
a_{37}	0.5517	a_{91}	0.8059	a_{145}	-2.3379	a_{199}	0.1807
a_{38}	-0.6189	a_{92}	1.5623	a_{146}	0.0384	a_{200}	0.0250
a_{39}	-1.7815	a_{93}	-0.8905	a_{147}	0.8368	a_{201}	0.4529
a_{40}	2.4891	a_{94}	-0.5311	a_{148}	-0.8066	a_{202}	1.0356
a_{41}	1.5123	a_{95}	0.6702	a_{149}	0.0583	a_{203}	1.5179
a_{42}	-0.3222	a_{96}	0.6042	a_{150}	-0.9380	a_{204}	-0.5810
a_{43}	-0.4188	a_{97}	-0.4048	a_{151}	1.6014	a_{205}	-0.7825

续表							
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_{44}	0.2389	a_{98}	-0.3571	a_{152}	-0.8575	a_{206}	1.5700
a_{45}	3.2088	a_{99}	1.9786	a_{153}	-1.3185	a_{207}	-0.0417
a_{46}	-6.7578	a_{100}	0.2629	a_{154}	0.4494	a_{208}	-0.6741
a_{47}	-1.7336	a_{101}	-1.3409	a_{155}	-0.6210	a_{209}	0.1057
a_{48}	0.1338	a_{102}	-0.0167	a_{156}	1.0164	a_{210}	-0.4759
a_{49}	-0.3896	a_{103}	-0.4802	a_{157}	1.2138	a_{211}	0.3238
a_{50}	1.2510	a_{104}	0.6784	a_{158}	0.4709	a_{212}	-0.7287
a_{51}	0.3363	a_{105}	0.3261	a_{159}	-1.1743	a_{213}	-0.5469
a_{52}	-1.6413	a_{106}	0.5611	a_{160}	-0.4962	a_{214}	1.1341
a_{53}	-0.4149	a_{107}	0.2522	a_{161}	-2.1470	a_{215}	-0.2055

从表 5 可以看出,模型参数个数比样本个数要多,故常用最小二乘不适合于本案例模型的参数估计。

表 6 毛竹生物量估测值与实测值

Table 6 Observed values and Fitted values of Moso bamboo's Biomass

地径/cm Diameter	时间 Time									
	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	0.0800	0.1765	0.1763	0.4156	0.4165	1.0146	1.0125	1.3824	1.3829
7.3	0.0317	0.0320	0.0774	0.0774	0.2735	0.2736	0.4380	0.4379	0.5893	0.5893
9.4	0.0530	0.0501	0.1080	0.1095	0.4149	0.4136	0.7340	0.7355	1.1198	1.1181
9.0	0.0767	0.0786	0.1884	0.1864	0.4336	0.4358	0.6773	0.6743	0.9169	0.9203
13.6	0.1231	0.1231	0.2734	0.2734	0.7671	0.7671	1.2914	1.2914	1.9711	1.9711
11.1	0.0943	0.0938	0.2196	0.2208	0.3655	0.3649	0.7780	0.7796	1.1893	1.1891
8.4	0.0437	0.0467	0.1075	0.1082	0.3305	0.3300	0.5554	0.5564	0.8416	0.8406
8.1	0.0413	0.0384	0.0987	0.0986	0.1929	0.1929	0.4261	0.4258	0.6552	0.6555
12.3	0.1475	0.1483	0.2291	0.2292	0.6521	0.6515	1.0462	1.0478	1.8984	1.8980
10.9	0.1010	0.1024	0.2375	0.2361	0.3532	0.3539	0.6297	0.6284	0.9064	0.9067

地径/cm Diameter	时间 Time									
	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	1.7578	2.3234	2.3241	2.8887	2.8876	3.7262	3.7263	4.6170	4.6170
7.3	0.8925	0.8925	1.3115	1.3115	1.6129	1.6130	1.7718	1.7718	2.0138	2.0138
9.4	1.5146	1.5146	1.9417	1.9424	2.3576	2.3562	2.8579	2.8582	3.3769	3.3767
9.0	1.1974	1.1965	1.5672	1.5671	2.2889	2.2905	3.0379	3.0375	3.7982	3.7984
13.6	2.6185	2.6185	3.7956	3.7956	4.4025	4.4025	5.4692	5.4692	6.5940	6.5940
11.1	1.7332	1.7335	2.4657	2.4652	2.8771	2.8791	3.6290	3.6290	4.2210	4.2209
8.4	1.1216	1.1218	1.6371	1.6371	1.8987	1.8979	1.9622	1.9625	2.0260	2.0259
8.1	0.8449	0.8446	1.0996	1.0997	1.3935	1.3936	1.6883	1.6882	2.0297	2.0297
12.3	2.7890	2.7886	3.3292	3.3288	3.6659	3.6665	4.1293	4.1292	4.6085	4.6085
10.9	1.4886	1.4883	2.0883	2.0884	2.4150	2.4136	2.7466	2.7465	3.0565	3.0566

(4) 为了更直观的说明生物量估算统一模型与偏最小二乘算法及生物量估算统一模型与偏最小二乘改进算法的估算精度,绘残差散点图,见图 2。

从图 2 可以看出,改进算法估算误差在零附近排成一直线,比偏最小二乘算法估算精度有很大改善。

5.3 案例总结

从以上 2 个案例可以看出,案例 2 对生物量影响因子考虑的比案例 1 多,案例 2 的样本量要多于案例 1 且案例 2 样本生物量的变化幅度要比案例 1 小得多。现对两案例做如下总结:(1)山本式二元生物量模型(变型模型)与 Marquardt 算法对案例 2 的模拟结果应该比案例 1 要好,但从 R^2 这个评价指标来看并非如此,原因是山本式二元生物量模型的变型带有太大主观性,修改模型参数缺乏科学依据。故主观修改模型参数的方法具有很大的局限性,同时也说明山本式二元生物量模型适应性较差。(2)生物量估算统一模型不论是对案例 1 的模拟还是对案例 2 的模拟都比山本式二元生物量模

型(变型模型)要稳定的多,原因在于它的构建有完备的理论基础。有理论基础的模型采用性能好的参数辨识方法(偏最小二乘改进算法)精度就会得到很大提高。(3)案例 1 比案例 2 少考虑了一个影响因子,在变量取相同阶数时,案例 1 中模型参数比案例 2 少了 180 个。(4)如果对生物量估算精度要求不高时,模型中变量可取较低阶数,此时 2 案例生物量统一模型中所含参数个数就会减少。

6 结论与讨论

本文在前人研究的基础上对生物量精确估算做了进一步探讨,并通过 2 个案例对 3 种估算方法进行了对比分析。综合前面的分析,得出如下结论:

(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,该模型适应于与生物量有关的任意多个因素。

(2) 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。

(3) 生物量估算统一模型可以看作是实际生物量模型的逼近。

(4) 改进算法克服了偏最小二乘在提取成分时固有的缺陷,能使模型估算精度得到很大提高,在拟合模型时不需要给定参数初值。

(5) 2 个案例都表明生物量估算统一模型与偏最小二乘改进算法对树种生物量的估计精度最高,其估计残差在零附近排成一条直线。

(6) 本文构建生物量统一模型的方法为理论生物量模型的建立提供了一种新的思路。连续函数空间的基可以根据需要取,因此如果用傅立叶级数、正交小波基构建生物量统一模型,此时模型的性能如何还有待于进一步的研究。

(7) 当树种生物量统一模型中含有较多影响生物量的因素且取较高阶数时,模型所含参数比较多,可以采用递阶偏最小二乘算法高度浓缩变量简化模型,关于这方面的研究将在以后进行。

(8) 常用估算生物量的模型还有幂函数形式 $M = c_0(DH)^{c_1}$, 由于 $M = c_0(DH)^{c_1} = c_0 D^{c_1} H^{c_1}$, 故这种形式的生物量模型是二元山本式模型的特例,所以本文没有讨论生物量模型的幂函数形式。

(9) 本文构建的生物量统一模型,其变量阶数的选取也带有一定的主观性,但这种主观性是以函数逼近理论为基础与满足精度要求为前提的。

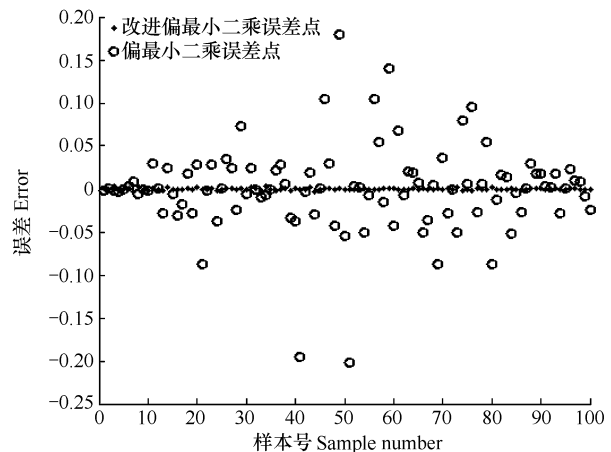


图 2 误差分布散点图

Fig. 2 Scatter plot of error's distribution

References:

- [1] Yang C H, Zhou X P, Wang X M. Regression models of above-ground biomass of *fargesia nitida* in dengsheng area in wolong nature reserve. *Scientia Silvae Sinicae*, 2008, 44(3): 113-123.
- [2] Wei X P, Zhao C M, Wang G X, Chen B M, Cheng D L. Estimation of above-and below-ground biomass of dominant desert plant species in an oasis-desert ecotone of minqin. *Acta Phytocologica Sinica*, 2005, 29(6): 878-883.
- [3] Zheng R M, Zheng Y S, Ding M F, Lin 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] Meng X Y, She G H. *Forest Measurement*. Beijing: China Forestry Publishing House, 2004: 131.
- [5] He Y H, Tian Y L, Ye D M, Qin J Q, Guo L S. Model of aboveground biomass of *nitraria tangutorum* and relationship between biomass and leaf area. *Journal of Desert Research*, 2005, 25(4): 541-546.
- [6] 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.
- [7] Luo Q B, Ning H, He D B, Jiang J S, Wu Z D, Zeng W S, Zhang S G. A study on standard volume dynamic model. *Forest Research*, 1992, 5(3): 263-270.
- [8] She G H, Liu E B, Ye J S, Lin S M. The improved algorithm and application of parametric identification for nonlinear tree volume model. *Journal of Nanjing Forestry University*, 2004, 28(3): 1-4.
- [9] Xing Y Q, Wang L H. Compatible biomass estimation models of natural forests in changbai mountains based on forest inventory. *Chinese Journal of Applied Ecology*, 2007, 18(1): 1-8.
- [10] Luo Q B, Zeng W S, He D B, Bao T H, Lin W R. Establishment and application of compatible tree above ground biomass models. *Journal of Natural Resources*, 1999, 14(3): 271-277.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] She G H, Liu E B, Ye J S, Lin S M. The least-squares minimization algorithm of resisting dimension error on application of nonlinear model. *Journal of Nanjing Forestry University*, 2005, 29(3): 9-13.
- [15] Zhou G Mo, Liu E Bin, 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): 2918-2925.
- [16] Lin C S. *Numerical Analysis*. Beijing: Science Press, 2006: 197-202.
- [17] Yan Q J. *Numerical Analysis*. Beijing: Bei Hang University Press, 2001: 122-125.
- [18] Wang H W, Wu Z B, Meng J. *Partial Least-Squares Regression-Linear and Nonlinear Methods*. Beijing: National Defence Industry Press, 2006: 275-276.
- [19] Walter L, Jan A P, Wold S. Partial least squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate. *Anal Chem*, 1983, 631-643.
- [20] Malthouse E C, Thmhana A C, Mah R S H. Nonlinear partial least squares. *Comput Chem Engng*, 1997, 21(8): 875-890.
- [21] Cheng B, Wu X Z. Assessing local influence in PLS regression by the second order approach. *Statistics & Probability Letters*, 2001, 53: 113-121.
- [22] Wang S W, Zheng W X. *The Summary of Real Variable Function and Functional Analysis (Second volume)*. Beijing: Higher Education Press, 2005: 305-310.

参考文献:

- [1] 杨春花, 周小平, 王小明. 卧龙自然保护区华西箭竹地上生物量回归模型. *林业科学*, 2008, 44(3): 113-123.
- [2] 魏小平, 赵长明, 王根轩, 陈宝明, 程栋梁. 民勤荒漠绿洲过渡带优势植物地上和地下生物量的估测模型. *植物生态学报*, 2005, 29(6): 878-883.
- [3] 郑容妹, 郑郁善, 丁闽锋, 林国新, 陈礼光, 王进丁. 苦竹生物量模型的研究. *福建林学院学报*, 2003, 23(1): 61-64.
- [4] 孟宪宇, 余光辉. *测树学*. 北京: 中国林业出版社, 2004: 131.
- [5] 何炎红, 田有亮, 叶冬梅, 秦佳琪, 郭连生. 白刺地上生物量关系模型及其与叶面积关系的研究. *中国沙漠*, 2005, 25(4): 541-546.

- [6] 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.
- [7] 骆期邦, 宁辉, 贺东北, 蒋菊生, 吴志德, 曾伟生, 张曙光. 二元立木材积动态模型研究. 林业科学研究, 1992, 5(3): 263-270.
- [8] 余光辉, 刘恩斌, 叶金盛, 林寿明. 非线性立木材积模型参数辨识的改进算法及其应用. 南京林业大学学报(自然科学版), 2004, 28(3): 1-4.
- [9] 邢艳秋, 王立海. 基于森林调查数据的长白山天然林森林生物量相容性模型. 应用生态学报, 2007, 18(1): 1-8.
- [10] 骆期邦, 曾伟生, 贺东北, 包拓华, 林文端. 立木地上部分生物量模型的建立及其应用研究. 自然资源学报, 1999, 14(3): 271-277.
- [11] 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.
- [12] 程堂仁, 冯菁, 马钦彦, 冯仲科. 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.
- [13] 张会儒, 唐守正, 王奉瑜. 与材积兼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.
- [14] 余光辉, 刘恩斌, 叶金盛, 林寿明. 非线性模型抗差最小二乘估计及其应用. 南京林业大学学报(自然科学版), 2005, 29(3): 9-13.
- [15] 周国模, 刘恩斌, 刘安兴, 周宇峰. 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.
- [16] 林成森. 数值分析. 北京: 科学出版社, 2006: 197-202.
- [17] 颜庆津. 数值分析. 北京: 北京航空航天大学出版社, 2001: 122-125.
- [18] 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.
- [22] 王声望, 郑维行. 实变函数与泛函分析概论(第二册). 北京: 高等教育出版社, 2005: 305-310.

附录: 偏最小二乘改进算法程序

偏最小二乘改进算法程序是用 Matlab 如件编写的, 其代码如下:

```
function XG = pianzuixiaoerchenggaijin(x, y)
```

```
[m, n] = size(x' * y * y' * x);
```

```
[v, d] = eig(x' * y * y' * x);
```

```
[m1, n1] = size(v);
```

```
j = 1;
```

```
b = ones(m1, n1 - 1);
```

```
for i = 1:m1
```

```
    if d(i, i)^2 <= e - 006
```

```
        b(:, j) = v(:, i);
```

```
        j = j + 1;
```

```
    end
```

```
end
```

```
[v1, d1] = eig(b' * x' * x * b);
```

```
[v2, aa] = sortpaixu(v1, d1);
```

```
[m2, n2] = size(aa);
```

```
mumval = sum(aa);
```

```
i = 1;
```

```
mumval1 = aa(1, 1);
```

```
while(mumval1/mumval <= 0.99)
```

```
    i = i + 1;
```

```
    mumval1 = mumval1 + aa(i, 1);
```

```
end
```

```
a = v2(:, 1:i);
```

```
u = x * b * a;
```

```
[m2, n2] = size(b);
```

```
I = eye(m2, m2);
```

$XG = x * (I - b * a * \text{inv}(u' * u) * u' * x);$

求得 XG 后,再做 Y 对 XG 的偏最小二乘回归。

改进算法要调用一个排序函数,其代码如下:

```
function [v,a] = sortpaixu(v,d)
[m,n] = size(d);
a = ones(m,1);
b = ones(m,1);
for i = 1:m
    a(i,1) = d(i,i);
end
for i = 1:m-1
    k = i;
    for j = i+1:m
        if(a(j,1) > a(k,1))
            k = j;
            if(k ~ = i)
                t = a(i,1);
                b = v(:,i);
                a(i,1) = a(k,1);
                v(:,i) = v(:,k);
                a(k,1) = t;
                v(:,k) = b;
            end
        end
    end
end
end
end
```