

带截距的非线性方程与分段建模方法对立木生物量估计的比较

张连金¹, 曾伟生^{2*}, 唐守正²

(1. 中国林业科学研究院林业研究所, 北京 100091; 2. 中国林业科学研究院资源信息研究所, 北京 100091)

摘要:针对用单一的非线性模型拟合生物量方程会导致小径阶林木的估计明显有偏的问题,以东北落叶松和南方马尾松的地上生物量数据为例,提出采用带截距的非线性方程和分段建模方法来改进对立木生物量的估计,并对二者的预估效果进行了比较。结果表明:两种方法都能有效克服小径阶林木的有偏估计问题,同时对整个模型的预估效果也有一定程度的改善;从对比分析结果看,分段拟合方程的预估效果还要略好一些。

关键词:立木生物量;非线性方程;分段建模;截距;比较;落叶松;马尾松

中图分类号:S711

文献标识码:A

Comparison of Nonlinear Regression Equation with Intercept and Segmented Modeling Approach for Estimation of Single-Tree Biomass

ZHANG Lian-jin¹, ZENG Wei-sheng², TANG Shou-zheng²

(1. Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China;

2. Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: Single-tree biomass equations with constant parameters within the range of size classes may result in obvious biased estimation for small young trees. Based on the above-ground biomass data of larch (*Larix* spp.) of the north-east and Masson pine (*Pinus massoniana*) of the south in China, two methods were presented to improve the estimation of tree biomass, which were nonlinear regression equation with intercept and segmented modeling approach, and the fit statistics of the models were compared. The results showed: (i) the two methods not only were effective to solve the problem of biased estimation for small trees, but also improved the prediction of biomass model for all sample trees in some extents; (ii) the segmented regression models were slightly better than nonlinear regression equation with intercept for single-tree biomass estimation.

Key words: single-tree biomass; nonlinear equation; segmented modeling; intercept; comparison; larch; Masson pine

森林是陆地生态系统的主体,是维持生态平衡和改善生态环境的重要保障,在应对全球气候变化中发挥着不可替代的作用。因此,世界各国越来越重视对森林生物量的监测。而要开展全国森林生物量监测和评估,建立适合较大区域范围的通用性立木生物量模型将是必不可少的一项基础工作。常用

的非线性立木生物量模型的一般结构形式为^[1]:

$$y = \beta_0 x_1^{\beta_1} \cdots x_i^{\beta_i} \quad (1)$$

式中, y 为林木生物量, x_i 为反映林木大小的变量, β_i 为模型参数。当自变量 x_i 只考虑林木胸径 D 时,(1)式就是应用最广的幂函数公式 $M = aD^b$ (式中 M 为生物量, a 、 b 为参数)^[2-6]。如果建模对象的林木

收稿日期:2010-09-03

基金项目:国家自然科学基金(31070485)资助。

作者简介:张连金(1980—),女,福建三明人,博士生,主要研究方向:天然林经营模拟。E-mail:zlianjin1102@126.com

*通讯作者:教授级高工。E-mail:zengweisheng@sohu.com

胸径 D 跨度较大,尤其是当小的幼树也一并考虑时,单一的模型往往难以全面描述 M 随 D 的变化。Bond-Lamberty 等^[7]在研究建立加拿大马尼托巴省北方森林的6个树种的生物量模型时发现,生物量与胸径之间的关系在3 cm 以下发生了显著变化。根据作者对东北落叶松 (*Larix* spp.) 和南方马尾松 (*Pinus massoniana* Lamb.) 的立木地上生物量数据进行分析,如果将不够检尺直径(6 cm 径阶以下)的幼树也一并参与建模,则对小径阶林木的生物量估计会存在明显偏差。针对这一问题,本文将探讨利用带截距的非线性方程和分段建模两种方法,来改进对立木生物量的估计,并对这两种方法的估计效果进行对比。

1 数据来源

本文所用数据为我国东北落叶松和南方马尾松

的地上生物量实测数据,每个树种各150株样木,其中东北落叶松的采集地点涉及内蒙古、黑龙江、吉林、辽宁4省区,南方马尾松的采集地点涉及江苏、浙江、安徽、福建、江西、湖南、广东、广西、贵州等9省区。样本单元数按各省资源多少分配,并兼顾天然和人工起源。每个树种的样木按2、4、6、8、12、16、20、26、32、38 cm 以上10个径阶均匀分布,每个径阶的样木按树高级也尽量均匀分布,在大尺度范围内具有广泛的代表性。全部样木都用钢围尺准确测量胸径,将样木伐倒后,用皮尺测量树干长度(树高)和活树冠长度(冠长),分干材、干皮、树枝、树叶称鲜质量,并分别抽取样品带回实验室,在85℃恒温下烘干至衡质量,根据样品鲜质量和干质量分别推算样木各部分干质量并汇总得到地上部分干质量。表1为两个树种立木地上生物量实测数据的基本情况。

表1 立木地上生物量实测数据的基本情况

树种	样本量	变量	平均值	最小值	最大值	标准差	变动系数/%
落叶松	150	胸径/cm	16.6	1.6	44.1	11.9	71.83
		树高/m	12.9	2.5	28.2	7.0	54.52
		冠幅/m	4.32	0.85	10.60	2.29	53.08
		冠长/m	8.35	1.25	23.60	4.57	54.77
		地上生物量/kg	177.040	0.614	1 160.740	234.024	132.19
马尾松	150	胸径/cm	16.6	1.5	47.2	12.1	73.13
		树高/m	12.0	2.0	27.6	7.2	60.20
		冠幅/m	4.47	0.60	12.00	2.55	57.12
		冠长/m	6.24	1.30	17.52	3.56	57.06
		地上生物量/kg	169.100	0.317	1 039.144	233.739	138.23

2 方法

为了具有更普遍的意义,本文以如下形式的二元立木生物量模型作为研究对比基础:

$$M = aD^b H^c \quad (2)$$

式中, M 为立木地上生物量(kg), D 为胸径(cm), H 为树高(m), a 、 b 、 c 为参数。显然,(2)式只要 D 或 H 为0,得到的生物量估计值就为0。根据专业知识,只要树高低于1.3 m,胸径 D 就为0,而树高刚好等于1.3 m的林木无疑具有一定的生物量。正是这一因素的影响,导致用(1)或(2)式得到的生物量估计值在小径阶会产生显著偏差。如果建模样本不涉及小径阶样木,则这一影响基本可以忽略不计。就如拟合树高-胸径曲线一样,一般情况下可以不考虑胸径等于0时树高等于1.3 m这一限制条件,而如果是针对幼龄林,则一定要设计为 $H = 1.3 + f(D)$

的形式,否则就会有显著影响。

基于以上考虑,在非线性模型(2)的基础上加上一个截距常数 a_0 ,变为以下形式:

$$M = a_0 + aD^b H^c \quad (3)$$

这里称(3)为带截距的非线性方程。类似(3)式的生物量方程尽管并不常见,但也有成功的应用,如在Zianis等^[8]汇编的欧洲各个树种的607个生物量方程(涉及39个树种)中,就有78个方程如(3)式一样带有截距常数(不含对数转换形式的方程)。

另外,根据对生物量数据的分析结果知,模型(2)除了导致2、4 cm等小径阶样木的估计结果出现偏差以外,还对相邻的6、8 cm径阶样木的估计结果产生一定影响。因此,解决小径阶样木估计偏差的另一途径是分段建模方法。

分段建模方法在林业领域鲜见应用,但在其它领域则被广泛应用^[9-14]。如杨润书^[10]利用分段建

模方法对云南高山地区高等级公路 GPS 高程进行分段拟合,有效提高了 GPS 高程的拟合精度;王雨时^[11]以跨音速段的最大值为界分超音速段和亚音速段,利用 Logistic 曲线对弹丸战斗部及其破片的空气阻力系数进行了分段拟合,结果表明分段建模方法是可行的;张学军、夏晓东^[14]针对电路仿真领域中对具有复杂非线性特性的电子器件不能采用单一神经网络建模的问题,采用分段建模方法不仅降低了模型的复杂度,还提高了模型的有效性。

分段建模需要解决两个核心问题:一是找到合适的节点;二是确保节点处无缝对接。第一个问题通过数据分析并结合多方案预先建模对比就能得到解决。本研究经过对两个树种的数据进行分析,并通过设定 5 cm 和 6 cm 两个节点进行建模对比,表明选定 6 cm 作为节点是合适的。第二个问题的解决,要通过对模型参数设置一些限制条件来实现。在本例中,分两段来建立以下模型:

$$M_1 = a_1 D^{b_1} H^{c_1} (D \geq 6 \text{ cm}) \quad (4)$$

$$M_2 = a_2 D^{b_2} H^{c_2} (D < 6 \text{ cm}) \quad (5)$$

因为 6 cm 以上林木是森林生物量的主体,先利用 120 株 6 cm 以上样木拟合模型(4),然后以模型(4)为基础,通过对模型(5)的参数进行限制,实现与模型(4)在胸径 6 cm 处无缝对接。对于简单的一元模型,只需设定一个点 D_0 作为节点;而对于二元模型,则必须设定一条线才能将分段模型连接起来。假设这条线经过 (D_0, H) ,其中 $D_0 = 6 \text{ cm}$ 。要保证模型(4)、(5)在这条线上的生物量估计值相等,则其参数必须满足以下条件:

$$c_2 = c_1 \quad (6)$$

$$\ln a_2 = (\ln a_1 + b_1 \ln D_0) - b_2 \ln D_0 \quad (7)$$

对(5)式取对数后得到:

$$\ln M_2 = \ln a_2 + b_2 \ln D + c_2 \ln H \quad (8)$$

再将(6)、(7)两式代入(8)式,经过化简后,得到:

$$b_2 = (\ln a_1 + b_1 \ln D_0 + c_1 \ln H - \ln M_2) / (\ln D_0 - \ln D) \quad (9)$$

假设:

$$y = \ln a_1 + b_1 \ln D_0 + c_1 \ln H - \ln M_2 \quad (10)$$

$$x = \ln D_0 - \ln D \quad (11)$$

则(9)式变为:

$$y = b_2 x \quad (12)$$

因此,利用 6 cm 以下 2 个径阶的 30 株样木的 D 、 H 、 M (在上述分段拟合模型中表述为 M_2),先按(10)、(11)式计算出 y 和 x 值,再拟合截距为 0 的线性方程(12)式,就得到了 b_2 的估计值;然后代入(7)式,就可以算出 a_2 的估计值。

为了对带截距的非线性方程(3)和分段拟合方程(4)、(5)的预估效果进行对比,主要考虑总相对误差 TRE (Total Relative Error)和平均系统误差 MSE (Mean System Error)这 2 项指标^[15]:

$$TRE = \sum_{i=1}^n (y_i - \hat{y}_i) / \sum_{i=1}^n \hat{y}_i \times 100 \quad (13)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) / \hat{y}_i \times 100 \quad (14)$$

式中, y_i 为实际观测值, \hat{y}_i 为模型预估值。除了针对整个样本以外,重点是要分径阶检验,并同时与(2)式的拟合结果进行比较。另外,模型本身的统计指标,还考虑以下 4 项^[1,15]:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (15)$$

$$SEE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n - p)}} \quad (16)$$

$$MPE = t_\alpha \cdot (SEE / \bar{y}) / \sqrt{n} \times 100 \quad (17)$$

$$MPSE = \frac{1}{n} \sum_{i=1}^n |(y_i - \hat{y}_i) / \hat{y}_i| \times 100 \quad (18)$$

其中 R^2 和 SEE 是回归模型的常用指标, MPE 是反映平均生物量估计值的精度指标, $MPSE$ 是反映单株生物量估计值的精度指标。

3 结果

首先利用全部 150 株样木,按(2)式分别两个树种拟合地上生物量模型,结果见表 2。因为生物量数据存在异方差性^[1,15-17],因此采用加权回归估计方法求解模型参数。

表 2 落叶松和马尾松地上生物量模型(2)的拟合结果

树种	参数估计值			统计指标					
	a	b	c	R^2	SEE	MPE	$MPSE$	TRE	MSE
落叶松	0.083 047	2.092 69	0.442 31	0.968 0	42.15	3.84%	17.06%	0.19%	2.64%
马尾松	0.072 234	2.079 53	0.491 13	0.965 3	43.84	4.19%	19.88%	1.28%	3.43%

从表2看总体拟合效果都不错,其中落叶松的统计指标均比马尾松要好。但是若按径阶来分析,则发现对小径阶林木的生物量估计存在明显偏差(见图1)。因为小径阶林木的生物量少,这种偏差的存在对整个森林生物量估计值的影响并不大,但

若要提高小径阶林木生物量的预估精度,则必须采取有关改进措施,采用带截距的非线性方程或分段建模方法。

按(3)式分别两个树种拟合地上生物量模型,结果见表3。

表3 落叶松和马尾松地上生物量模型(3)的拟合结果

树种	参数估计值				统计指标					
	a_0	a	b	c	R^2	SEE	MPE	$MPSE$	TRE	MSE
落叶松	0.330 33	0.063 068	2.006 84	0.630 90	0.968 8	41.76	3.81%	15.07%	-0.02%	0.05%
马尾松	0.276 59	0.057 693	2.132 54	0.509 74	0.965 6	43.79	4.19%	17.78%	0.00%	1.12%

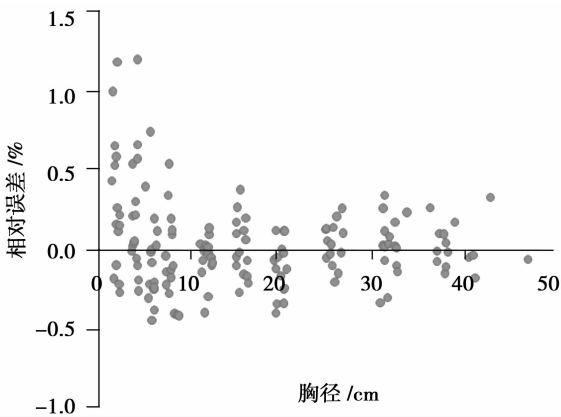


图1 马尾松生物量模型(2)的相对误差分布

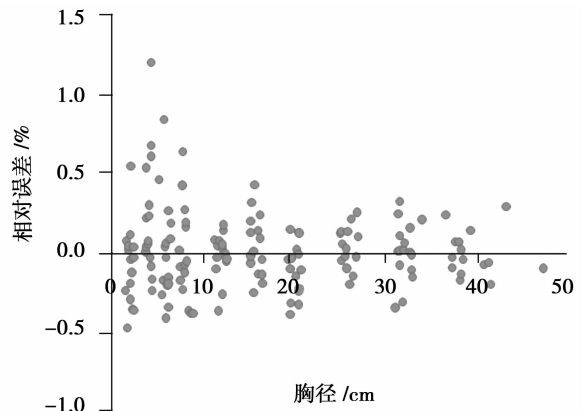


图2 马尾松生物量模型(3)的相对误差分布

从表3与表2的对比可以看出,模型(3)的各项统计指标都要好于模型(2),说明增加截距常数 a_0 在一定程度上改进了整个模型的预估效果。更重要的是,小径阶林木估计有偏的问题可以得到明显改

善(见图2)。

最后按分段建模方法,分别6 cm径阶以上和以下来拟合模型,结果见表4。有关统计指标是针对整个样本而言的,计算 SEE 指标时独立参数个数取5个。

表4 落叶松和马尾松地上生物量分段建模拟合结果

树种	参数估计值						统计指标					
	a_1	b_1	c_1	a_2	b_2	c_2	R^2	SEE	MPE	$MPSE$	TRE	MSE
落叶松	0.064 155	2.063 31	0.562 87	0.138 20	1.635 01	0.562 87	0.968 2	42.29	3.86%	14.67%	-0.31%	0.56%
马尾松	0.058 146	2.131 74	0.508 05	0.117 43	1.739 47	0.508 05	0.965 6	43.93	4.20%	18.09%	0.11%	2.00%

从表4看,整个模型的总体统计指标比模型(2)也有改进,而与模型(3)相差不大。如果进一步考虑各径阶的估计结果,则可以发现,采用分段建模方法有效地解决了小径阶的有偏估计问题。表5列出了3种估计方法对各个径阶的总相对误差 TRE 和平均系统误差 MSE 指标值。

从表5知,落叶松模型(2)对2 cm径阶的预估偏差达到30%以上,马尾松模型(2)对2、4 cm径阶的预估偏差都达到20%左右。采用带截距的非线性方程(3)和分段拟合方法,可以有效消除小径阶存在的系统偏差。如落叶松采用分段建模方法,各

个径阶的总体相对误差 TRE 基本都不超出 $\pm 5\%$ 的范围,仅16 cm径阶达到5.28%。马尾松除了4、20 cm这两个径阶的数据有所异常外,其它径阶的 TRE 和 MSE 指标基本也在 $\pm 5\%$ 以内。从表5的对比结果看,分段建模方法要略好于带截距的非线性方程。

4 结论

本文以东北落叶松和南方马尾松的地上生物量数据为例,针对用单一的非线性模型(2)拟合生物量方程会导致小径阶林木的估计明显有偏的问题,提出采用带截距的非线性方程(3)和分段建模方法

表5 落叶松和马尾松不同地上生物量模型分径阶对比结果

D/cm	落 叶 松						马 尾 松					
	总相对误差 TRE/%			平均系统误差 MSE/%			总相对误差 TRE/%			平均系统误差 MSE/%		
	模型(2)	模型(3)	模型(4/5)	模型(2)	模型(3)	模型(4/5)	模型(2)	模型(3)	模型(4/5)	模型(2)	模型(3)	模型(4/5)
2	30.96	2.56	1.34	39.08	3.83	5.22	21.20	-9.58	-5.99	30.45	-8.35	-0.72
4	4.48	3.84	-0.54	5.69	4.68	0.01	18.40	20.11	14.88	21.71	23.10	18.24
6	-9.39	-5.95	-1.86	-8.57	-4.29	-0.32	-5.85	-0.57	2.44	-4.04	1.26	4.18
8	-8.62	-6.94	-3.68	-8.25	-6.04	-2.93	-9.87	-4.60	-2.85	-6.46	-0.89	1.07
12	-6.28	-4.50	-2.91	-5.18	-2.93	-1.47	-6.44	-2.45	-1.92	-7.28	-3.29	-2.76
16	2.47	4.65	5.28	2.24	4.85	5.29	1.28	4.01	4.26	2.10	4.98	5.24
20	-1.97	-1.99	-1.27	-2.63	-2.50	-1.86	-10.40	-9.15	-9.04	-11.31	-10.01	-9.90
26	2.59	2.35	2.42	2.56	2.48	2.47	4.38	4.12	4.20	4.00	3.77	3.84
32	4.19	3.38	3.08	4.05	3.29	2.97	2.76	1.08	1.16	2.73	1.10	1.17
≥38	-2.44	-2.79	-3.74	-2.56	-2.84	-3.80	2.16	-0.68	-0.61	2.49	-0.28	-0.20

来改进模型的预估效果。根据有关研究结果,可以得到以下结论:

(1)非线性模型(2)由于没有考虑树高等于1.3 m的林木必然具有一定的生物量这一基本前提条件,从而使模型的预估效果受到一定程度的影响,尤其是对小径阶林木的估计,会产生明显的系统偏差。从两个树种的结果看,2 cm径阶的预估偏差可以达到20%甚至30%以上。

(2)采用带截距的非线性方程(3)可以有效克服小径阶林木的有偏估计问题,同时对整个模型的预估效果也有一定程度的改善。从两个树种的结果看,模型(3)的截距常数大约在0.3左右,说明树高1.3 m的落叶松和马尾松林木,其生物量平均约为0.3 kg。

(3)利用分段建模方法同样可以解决小径阶林木的有偏估计问题,并在一定程度上改进整个模型的预估效果。从与带截距的非线性方程(3)的对比结果看,分段拟合方程的预估效果还要略好一些,该方法可以推广应用到林业数表领域材积方程等其它模型的研建。

参考文献:

- [1] Parresol B R. Assessing tree and stand biomass: a review with examples and critical comparisons[J]. For Sci, 1999, 45(4): 573-593
- [2] Ter-Mikaelian M T, Korzukhin M D. Biomass equations for sixty-five North American tree species[J]. Forest Ecology and Management, 1997, 97:1-24
- [3] Jenkins J C, Chojnacky D C, Heath L S, et al. National-scale biomass estimators for United States tree species[J]. For Sci, 2003, 49(1):12-35

- [4] Muukkonen P. Generalized allometric volume and biomass equations for some tree species in Europe[J]. Eur J Forest Res, 2007, 126: 157-166
- [5] Case B, Hall R J. Assessing prediction errors of generalized tree biomass and volume equations for the boreal forest region of west-central Canada[J]. Can J For Res, 2008, 38:878-889
- [6] Návar J. Allometric equations for tree species and carbon stocks for forests of northwestern Mexico[J]. Forest Ecology and Management, 2009, 257: 427-434
- [7] Bond-Lamberty B, Wang B C, Gower S T. Aboveground and below-ground biomass and sapwood area allometric equations for six boreal tree species of northern Manitoba[J]. Can J For Res, 2002, 32: 1441-1450
- [8] Zianis D, Muukkonen P, Mäkipää R, et al. Biomass and stem volume equations for tree species in Europe[J]. Silva Fennica, Monographs, 2005, 4:1-63
- [9] 魏绍凯,郑叔芳. 计算机辅助叶型曲线的最佳分段拟合[J]. 动力工程, 1993, 13(3): 48-52
- [10] 杨润书. 高山区高等级公路 GPS 测高的分段拟合及精度分析[J]. 测绘通报, 2005(12): 38-40
- [11] 王雨时. 弹丸战斗部及其破片空气阻力系数的 Logistic 曲线分段拟合[J]. 弹箭与制导学报, 2006, S1: 242-244
- [12] 李武选. 分段拟合技术在长期趋势建模过程中的具体应用[J]. 世界科技研究与发展, 2008, 30(3): 363-369
- [13] 朱家诚,汪进,吴天星,等. 凸轮从动件运动规律的分段建模方法及设计系统研究[J]. 机械设计, 2008, 25(7): 23-25
- [14] 张学军,夏晓东. 基于神经网络的电子器件分段建模方法[J]. 物理实验, 2009, 29(12): 20-22
- [15] 曾伟生,骆期邦,贺东北. 论加权回归与建模[J]. 林业科学, 1999, 35(5): 5-11
- [16] 张会儒,唐守正,胥辉. 关于生物量模型中的异方差问题[J]. 林业资源管理, 1999(1): 46-49
- [17] 胥辉. 生物量模型方差非齐性研究[J]. 西北林学院学报, 1999, 19(2): 73-77