

杉木人工林林分断面积生长模型的贝叶斯法估计

张雄清^{1,2}, 张建国^{1,2*}, 段爱国^{1,2}

(1. 中国林业科学研究院林业研究所, 国家林业局林木培育重点实验室, 北京 100091;

2. 南京林业大学南方现代林业协同创新中心, 江苏 南京 210037)

摘要:以江西杉木人工林为例,以 Korf 型、Richards 型和 Hossfeld 型 3 种模型为基础,通过广义代数差分法(GADA)分别建立杉木林分断面积生长模型。结果表明:以 Richards 型为基础的杉木林分断面积预测精度最高,以 Richards 型模型为最优模型,分别基于贝叶斯法和传统法(非线性最小二乘法)估计杉木林分断面积生长模型。研究发现,利用贝叶斯法估计杉木林分断面积生长模型,预测精度相当且预测值的可靠性比传统法好。

关键词:贝叶斯法;传统法;林分断面积;杉木

中图分类号:S711

文献标识码:A

Application of Bayesian Method in Stand Basal Area Prediction of Chinese Fir Plantation

ZHANG Xiong-qing^{1,2}, ZHANG Jian-guo^{1,2}, DUAN Ai-guo^{1,2}

(1. Key Laboratory of Tree Breeding and Cultivation, State Forestry Administration Research Institute of Forestry, Chinese Academy of Forestry;

Beijing 100091, China; 2. Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University,

Nanjing 210037, Jiangsu, China)

Abstract: Chinese fir (*Cunninghamia lanceolata* (Lamb.) Hook.), an endemic tree species in China's subtropical area, is one of the most important fast-growing tree species for timber production in southern China. Based on the periodic data of the Chinese fir in Jiangxi province, three stand basal area models (Korf-based model, Richards-based model, and Hossfeld-based model) were developed using generalized algebraic difference approach (GADA). The results showed that Richards-based model was the best for modeling the stand basal area of Chinese fir in the study. Additionally, Bayesian method and Classical method (nonlinear least squares method) were used to estimate the Richards-based model. Although the precision of Bayesian method was nearly equal to that of the classical method, the model reliability using Bayesian method was better than using classical method.

Key words: Bayesian method, stand basal area, Chinese fir, *Cunninghamia lanceolata*

杉木(*Cunninghamia lanceolata* (Lamb.) Hook.)是我国亚热带地区特有的优良用材树种,也是我国南方主要的造林树种。第七次全国森林资源清查表明,全国杉木人工林面积为 853.86 万 hm^2 , 占全国造林面积的 21.35%, 在我国森林资源中占有重要的地位。

在林分生长和收获预估体系中,林分断面积既是用来预估材积收获的重要变量,又是被估计的主要因子。林分断面积具有较高的稳定性和预估性及在林业调查和生产实践中的易测定性^[1], 因此,杉木林分断面积是杉木人工林林分测算因子建模中的主要对象。

收稿日期:2013-10-17

项目基金:国家自然科学基金(31300537);中国林业科学研究院林业研究所科研院所基本科研业务资金项目(RIF2013-09)和中国林业科学研究院科研院所基本科研业务资金项目(CAFYBB2014QB002)

作者简介:张雄清,男,副研究员。研究方向:人工林定向培育和生长模型模拟。Email:xqzhang85@caf.ac.cn

* 通讯作者:研究员,博士生导师,主要从事森林培育研究。

杉木林分断面积生长模型的研究,为林业工作者科学经营管理杉木,揭示杉木生长发育规律提供重要的参考依据。杜纪山等^[2]基于 Richards 型和 Schumacher 型分别建立了杉木林分断面积生长模型,研究发现 Richards 型优于 Schumacher 型。张建国等^[3]利用差分模型模拟杉木林分断面积生长,并通过传统法(最小二乘法)估计模型参数,但是,利用传统法很难对参数估计的不确定性和预测值的可靠性进行评价^[4]。近些年,根据文献报道,贝叶斯法是评价模型不确定性的一个很好的方法,已经在环境、生态、医疗、水文等研究领域得到广泛应用^[5-7];而且贝叶斯法综合利用了先验信息和样本信息,先验信息是在进行统计推断时不可缺少的一个因素,可以来自历史文献资料或者主观信念,这在林业研究工作中是很重要的。在森林的连续调查中,每一次的调查分析结果都是下一次调查分析的最合理的先验信息。在林业中,贝叶斯法也有了一定的研究,如:生物量模型^[8],直径分布模型^[9-10],直径生长模型^[11]等;但是,贝叶斯方法在杉木林分断面积生长模型中的研究未见报道。本文以江西杉木人工林为研究对象,利用贝叶斯法估计杉木断面积生长模型,并与传统推断法(最小二乘法)进行比较分析,为分析杉木人工林分断面积生长模型预测结果的可靠性提供参考。

1 试验地概况

试验地位于江西省分宜县大岗山年株林场场部后山,属于罗霄山脉北端的武功山支脉,114°30′~114°45′E,27°30′~27°50′N,海拔 250 m,母岩为砂页岩,年平均气温 16.8℃,降水量 1 656 mm,年蒸发量 1 503 mm,属南亚热带季风气候区。

该试验林建立于 1981 年,采用 1 年生苗随机区组试验设计,5 个密度:A 密度(2 m×3 m)、B 密度(2 m×1.5 m)、C 密度(2 m×1 m)、D 密度(1 m×1.5 m)、E 密度(1 m×1 m),每个密度 3 次重复,总共 15 个样地,每个样地面积为 600 m²。在每个样地周围各植 2 行同样密度的杉木作为保护带。杉木林分的主要生长指标见表 1。

表 1 杉木林分的主要生长指标

项目	最小值	最大值	均值	标准差
年龄/a	6	18	11.11	3.94
优势木平均高/m	5.78	16.43	10.26	2.62
林分密度/(株·hm ⁻²)	1 633.33	9 983.33	5 094.07	2 631.93
林分断面积/(m ² ·hm ⁻²)	1.35	63.50	29.84	15.58

1989 年前逐年进行样地调查,1989 年后隔年调查。对样地内每株树进行编号,并进行每木检尺。本研究数据来自 1985 年到 1997 年。

2 研究方法

2.1 杉木林分断面积生长模型的建立

以 Korf 型^[12]、Richards 型^[13]和 Hossfeld 型^[3]为基础利用广义代数差分法(GADA)建立杉木林分断面积动态生长模型:

$$\text{Korf 型: } B = \alpha_1 \left(\frac{B_0}{\alpha_1} \right)^{\left(\frac{t_0}{t_1} \right)^2} \quad (1)$$

$$\text{Richards 型: } B = \alpha_1 \left\{ 1 - \left[1 - \left(\frac{B_0}{\alpha_1} \right)^{1/\alpha_2} \right]^{1/t_0} \right\}^{\alpha_2} \quad (2)$$

$$\text{Hossfeld 型: } B = \frac{\alpha_1}{1 - (1 - \alpha_1/B_0)(t_0/t_1)^{\alpha_2}} \quad (3)$$

式中: B_0 为林分初始断面积(m²·hm⁻²); t_0 为林分初始年龄; t_1 为林分生长到某一阶段 B 时的年龄; α_1 、 α_2 为待估参数。

2.2 贝叶斯理论

贝叶斯推断的基本方法是将未知参数的先验信息与样本信息结合,再根据贝叶斯定理,得出后验信息,然后根据后验信息去推断未知参数的分布^[15]。令 $y = (y_1, y_2, y_3, \dots)$ 为数据向量, $\theta = (\theta_1, \theta_2, \theta_3, \dots)$ 为参数向量,则根据贝叶斯理论,其基本公式为:

$$p(y, \theta) = p(y | \theta)p(\theta) = p(\theta | y)p(y) \quad (4)$$

式中: p 为概率分布函数或者密度函数。在传统方法中,可以利用最小二乘法或者最大似然估计法进行参数 θ 估计;在贝叶斯方法中,通过概率分布来描述参数 θ 的不确定性,进而估计参数 θ 。根据贝叶斯条件概率, θ 的条件概率分布为:

$$p(\theta | y) = \frac{p(y | \theta)p(\theta)}{p(y)} \quad (5)$$

其中,对于连续型 θ , $p(y) = E_{\theta}(p(y | \theta)) = \int p(y | \theta)p(\theta) d(\theta)$ 。

贝叶斯方法中,在给定样本 y 下, θ 的条件分布 $p(\theta | y)$ 就是所求得的后验分布, $p(y | \theta)$ 是在给定 θ 下 y 的似然函数, $p(\theta)$ 是 θ 的先验分布。

2.3 先验分布

先验分布的选择在贝叶斯方法中非常重要^[16]。在上述杉木林分断面积生长模型中,需要为参数

α_1 、 α_2 选择合适的先验分布。许多学者选择利用无信息先验分布,该信息可以忽略不计,而且对参数估计的影响不大。对于无信息先验分布,一般选择均值为0,方差足够大的能够覆盖整个数据范围的正态分布^[6]。当然,也可以选择有信息先验分布作为贝叶斯方法中的先验分布,这些信息可以来自历史文献资料或者主观信念。本研究中,杉木林分断面积生长模型的2个参数的无信息先验分布分别为: $\alpha_1 \sim N(0, 1\ 000)$ 、 $\alpha_2 \sim N(0, 1\ 000)$ 。

对于贝叶斯参数估计,本研究中利用 WinBUGS 软件^[17]完成。该软件通过吉布斯抽样算法^[18]估计参数。在本研究中还通过 R2WinBUGS^[19]连接 R 软件和 WinBUGS 软件完成数据的输入和输出以及图形的生成。对于方差分析和传统推断法估计参数,在 SAS 软件中完成。在进行贝叶斯估计时,为了保证迭代收敛和得到稳定的参数后验概率值,迭代次数设为30万次,并去掉前面的50 000次退火迭代。

3 模型评价

在传统方法中,一般采用均方根误差(RMSE)作为模型的拟合统计量指标。

$$RMSE = \sqrt{\sum (y_i - \hat{y}_i)^2 / (n - 1)} \quad (6)$$

式中: y_i 为实际值(林分断面积); \hat{y}_i 为预测值; n 为观察个数。

对于贝叶斯模型,DIC 统计量是个非常好的拟合评价指标^[20]。与 RMSE 值一样,DIC 越小,说明模型拟合越好。因此,在贝叶斯模型中,通过 DIC 和 RMSE 2 个统计量选择最优模型。

$$DIC = Dbar + pD \quad (7)$$

式中: $Dbar = E_p(-2\log(p(y|\theta)))$,表示模型拟合数据的优劣, pD 则是模型中参数的有效个数, $pD = Dbar + 2\log(p(y|\bar{\theta}))$,表示模型的复杂度。

4 结果与分析

4.1 最优模型的选择

由表2可知:Richards 型模型模拟杉木林分断面积的 DIC 值最小,为362.95,其次是 Korf 型381.24, Hossfeld 型的最大,为410.27;而且,通过 RMSE 值的对比,也发现 Richards 型模型的 RMSE 值(1.070 6)最小。因此,结合这2个评价指标,选择 Richards 型模型作为最优模型预估杉木林分断面积。

4.2 贝叶斯估计法和传统估计法的比较

根据表2中 DIC 值的比较结果,以 Richards 型最优模型为基础,比较贝叶斯估计法和传统估计法。

根据贝叶斯法估计杉木人工林林分断面积生长模型参数,其参数的后验概率分布见图1。根据图1,可以发现杉木林分断面积生长模型的参数是随机变量,服从正态分布,这就更好地解释了林分断面积生长模型的不确定性。

表2 3个林分断面积生长模型的参数估计及评价

模型	参数估计		95% 置信区间		DIC	RMSE	
	估计值	标准差	2.5%值	97.5%值			
Korf 型	α_1	99.93	0.696 3	98.75	101.2	381.24	1.165 3
	α_2	0.90	0.024 2	0.85	0.94		
Richards 型	α_1	0.19	0.010 1	0.17	0.21	362.95	1.070 6
	α_2	3.20	0.226 6	2.76	3.66		
Hossfeld 型	α_1	95.00	2.425 0	90.95	99.05	410.27	1.310 7
	α_2	1.52	0.053 3	1.42	1.63		

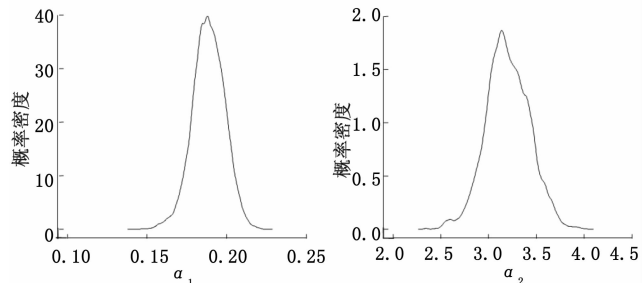


图1 杉木林分断面积生长模型的参数后验概率分布图

表3为基于杉木林分断面积数据建立的杉木林分断面积生长模型并分别通过传统法(非线性最小二乘法)和贝叶斯法估计的模型参数。从表3发现:贝叶斯法的参数估计置信区间与传统法估计的置信区间相近。通过比较 RMSE 值,发现这2种方法的预测精度相近。从残差图(图2)发现,这2种方法估计的林分断面积残差相对比较均匀地分布在“0”的上下部分,预测偏差都在-4到4之间;但从图3又发现,基于贝叶斯法预测杉木林分断面积,预测值的置信区间比传统法的置信区间小的多。这是因为贝叶斯法综合利用了样本信息和先验分布信息,而且在贝叶斯方法中,参数是随机变量,使贝叶斯法的预测值更可靠、稳定。

表3 基于传统法和贝叶斯法的杉木林分断面积生长模型参数估计值及模型评价

估计方法	参数估计值		95% 置信区间		RMSE	
	估计值	标准差	2.5%值	97.5%值		
传统法	α_1	0.19	0.010 0	0.17	0.21	1.070 6
	α_2	3.19	0.223 2	2.75	3.63	
贝叶斯法	α_1	0.19	0.010 1	0.17	0.21	1.0706
	α_2	3.20	0.226 6	2.76	3.66	

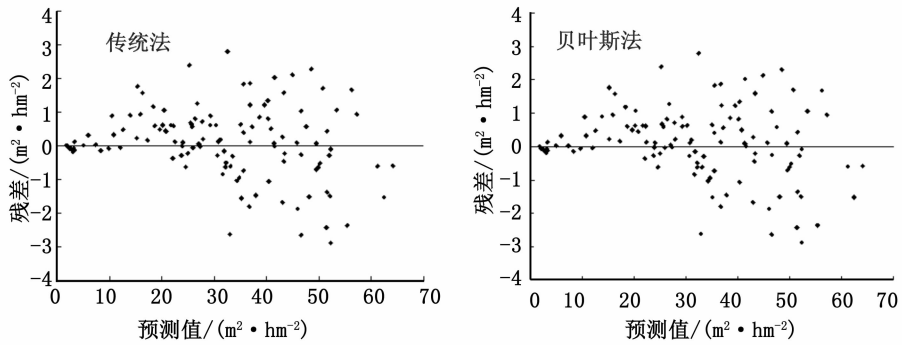


图2 杉木林分断面积生长模型的残差图

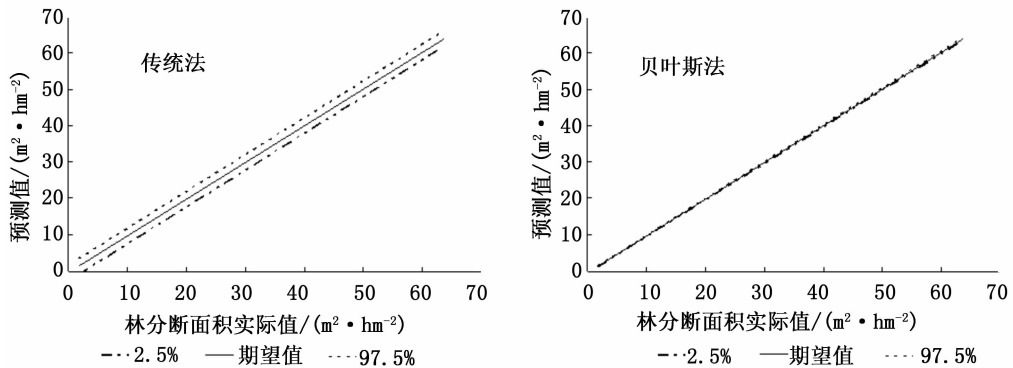


图3 杉木林分断面积生长模型的预测值(包括期望值和95%区间值)与实际值相关图

5 结论与讨论

在以往的研究中,林分断面积生长模型一般需要具备3个主要特征^[13, 21]:(1)生物学特性:当林分年龄趋近于无限大时,林分断面积模拟值必须有拐点和降近值;(2)预估步长一致性:林分断面积预测从年龄 t_1 到 t_3 与从 t_1 经 t_2 再到 t_3 预测的结果一样;(3)模型结构简易性:模型结构越复杂,意味着存在自变量间的共线性问题,导致模型预测结果的不稳定性。根据以上的3点特征要求,Bailey等^[22]提出利用代数差分法(ADA)建立生长模型。之后,Cieszewski等^[23]在ADA方法的基础上提出了广义代数差分法(GADA)分析动态模型。GADA法的主要优点在于可以根据林分的生长特性(降近值)对基础模型进行扩展,使模型弹性更好^[24-25]。本研究以Korf型、Richards型和Hossfeld型为基础,建立了杉木林分断面积生长模型,且以Richards型为基础建立的杉木林分断面积生长模型预测精度比较高,为本研究杉木林分断面积的最优模型。

本文分别利用贝叶斯法和传统法(非线性最小二乘法)估计Richards型的杉木林分断面积生长模

型,研究发现,通过贝叶斯法估计林分断面积生长模型,虽然预测精度相似,但其预测值可靠性、稳定性更好。这是因为在样本量比较大时,传统法和贝叶斯法的模拟精度比较接近,当样本量比较小时,贝叶斯法的模拟精度高于传统法^[8]。在统计推断的研究中,贝叶斯推断比传统推断法有很大的优势,主要表现在以下3方面:贝叶斯推断法综合利用了先验信息和样本信息,先验信息(分布)是在进行统计推断时不可缺少的一个因素,可以来自历史资料(文献)或者主观信念,而传统法仅利用了样本信息^[6, 26];贝叶斯法把样本和参数看作是随机变量,而传统法把未知参数估计值看作固定值^[4, 6];贝叶斯推断法并没有对参数或模型的构造加以限制,而传统法一般假设服从正态分布^[27]。当选择的先验分布为无信息先验分布时,贝叶斯估计的参数可信区间在数值上一般与传统法估计的置信区间相似^[26]。因此,在本研究中,利用无信息先验分布估计的参数与传统法估计值很接近。当在贝叶斯方法中利用有信息先验分布,预测的参数可信区间比无信息先验分布估计的可信区间窄,同样也比传统法估计的置信区间更精确可靠。

总之,利用贝叶斯法估计杉木林分断面面积生长模型,预测效果更可靠稳定。今后应该加强贝叶斯法在林业模型上的应用研究。当然,也可以在杉木林分断面面积生长模型中加入立地质量、林分密度等一些因素,建立分层贝叶斯模型,使先验分布更精确,进而提高杉木林分断面面积生长模型的精度。

参考文献:

- [1] 张雄清, 雷渊才, 陈新美, 等. 组合预测法在林分断面面积生长预估中的应用[J]. 北京林业大学学报, 2010, 32(4): 6-11.
- [2] 杜纪山, 唐守正. 杉木林分断面面积生长预估模型及其应用[J]. 北京林业大学学报, 1998, 20(4): 1-5.
- [3] 张建国, 孙洪刚. 杉木人工林断面面积生长规律及动态模拟[M]. 北京: 科学出版社, 2010: 58-75.
- [4] Li R, Stewart B, Weiskittel A. A Bayesian approach for modelling non-linear longitudinal/hierarchical data with random effects in forestry[J]. *Forestry*, 2012, 85(1): 17-25.
- [5] Clyde M. Model uncertainty and health effect studies for particulate matter[R]. Technical Report Series, NRCSE-TRS No. 027, 1999.
- [6] Ellison A M. Bayesian inference in ecology[J]. *Ecology letters*, 2004, 7: 509-520.
- [7] 李向阳. 水文模型参数优选及不确定性分析方法研究[D]. 大连: 大连理工大学, 2005.
- [8] Zhang X, Duan A, Zhang J. Tree biomass estimation of Chinese fir (*Cunninghamia lanceolata*) based on bayesian method[J]. *PLOS ONE*, 2013, 8(11): 1-7.
- [9] Green E J, Jr, Roesch FAA, Smith F M, *et al.* Bayesian estimation for the three-parameter Weibull distribution with tree diameter data[J]. *Biometrics*, 1994, 50(1): 254-269.
- [10] Bullock B P, Boone E L. Deriving tree diameter distributions using Bayesian model averaging[J]. *Forest Ecology and Management*, 2007, 242: 127-132.
- [11] Clark J S, Wolosin M, Dietze M, *et al.* Tree growth inference and prediction from diameter censuses and ring widths[J]. *Ecological Applications*, 2007, 17(7): 1942-1953.
- [12] Clutter J L, Jones E P. Prediction of growth after thinning in old-field slash pine plantations[R]. USDA Forest Service Research Paper, 1980: 217.
- [13] Anta M B, Dorado F C, Aranda U D, *et al.* Development of a basal area growth system for maritime pine in northwestern Spain using the generalized algebraic difference approach[J]. *Canadian Journal of Forest Research*, 2006, 36(6): 1461-1474.
- [14] Lindley D V. Bayesian thoughts[J]. *Significance*, 2000, 1(2): 73-75.
- [15] 茹诗松. 高等数理统计[M]. 北京: 高等教育出版社, 1998.
- [16] Gelman A, Carlin J B, Stern H S. *Bayesian Data Analysis*, 2nd edn[M]. USA: Chapman and Hall/CRC, 2004.
- [17] Spiegelhalter D J, Thomas A, Best N. *WinBUGS User Manual* [M]. Cambridge: MRC Biostatistics Unit, 2003.
- [18] Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm[J]. *The American Statistician*, 1995, 49(4): 327-335.
- [19] Sturtz S, Ligges U, Gelman A. R2WinBUGS: a package for running WinBUGS from R[J]. *Journal of Statistic Software*, 2005, 12(3): 1-16.
- [20] Spiegelhalter D J, Best N G, Carlin B P, *et al.* Bayesian measures of model complexity and fit (with discussion)[J]. *Journal of Royal Statistical Society, Series B*, 2002, 64: 583-616.
- [21] 杜纪山, 唐守正. 林分断面面积生长模型研究综述[J]. *林业科学研究*, 1997, 10(6): 599-606.
- [22] Bailey R L, Clutter J L. Base-age invariant polymorphic site curves[J]. *Forest Science*, 1974, 20(2): 155-159.
- [23] Cieszewski C J, Bailey R L. Generalized algebraic difference approach; theory based derivation of dynamic equations with polymorphism and variable asymptotes[J]. *Forest Science*, 2000, 46(1): 116-126.
- [24] Cieszewski C J. Three methods of deriving advanced dynamic site equations demonstrated on inland Douglas-fir site Curves[J]. *Canadian Journal of Forest Research*, 2001, 31(1): 165-173.
- [25] Cieszewski C J. Comparing fixed-and variable-base-age site equations having single versus multiple asymptotes[J]. *Forest Science*, 2002, 48(1): 7-23.
- [26] McCarthy M A. *Bayesian methods for ecology*[M]. Cambridge UK: Cambridge University Press, 2007.
- [27] Davidian M, Giltinan D M. *Nonlinear Models for Repeated Measurement Data*[M]. New York: Chapman and Hall, 1995.