林业科学研究 1997, 10(4): 420～424
Forest Research

年龄隐含的生长模型在森林资源连续清查中的应用*

葛宏立 项小强 何时珍 方陆明

摘要 根据森林资源监测体系的特点,将年龄隐含的生长模型应用于森林资源连续清查。这种模型不需要年龄,其误差与预测年限 ΔA 有关,当 $\Delta A=0$ 时,误差为零。用 Johnson-Schumacher 模型作为以年龄为自变量的模型,以此模型为基础模型导出的年龄隐含模型为例进行了试验,数据为浙江省的一类连续清查数据中的1989～1994年的马尾松复位样木数据。计算例子表明, 5 a 定期生长量的总和估计精度在 95% 的概率保证下可达 80% 以上。文章给出了模型的推导,模型的性质分析,误差及精度估计方法等。这种模型用复位样木建模,应用方便。可以以省为建模单元,一个树种(组)建一个模型,而不区分人工林、天然林,异龄林、同龄林,优势树种、组成树种,林木或散生木。

关键词 生长模型 Johnosn-Schumacher 模型 年龄隐含模型 森林资源连续清查

对于任何模型都应有两个必备条件:建立模型的条件与应用模型的条件。建模的条件是建什么样的模型,就应有什么样的建模数据,应用的条件就是对于自变量精度的要求至少不低于建模时对自变量的精度要求。如果应用时无法提供某些自变量,或是自变量精度不能保证,这时,宁可应用自变量少,对自变量要求低的较简单的模型。

森林资源监测体系,有其自身的很多特点。其一,它通常不能提供满足年龄显示模型要求的主要自变量,即年龄。其二,建模应考虑到所有树种(组),所有林相。其三,具有复测的成对数据。其四,定期复查,所以模型预测的年限一般不超过一个复查间隔期,目前为5 a,属于短期预测(或更新)。其五,数据多。一、二两点是不利因素,三、四、五是有利因素。为扬长避短,我们提出不含年龄的生长模型,即以前期观测值作为自变量。下面称年龄隐含的模型为年龄隐含模型,以年龄为自变量的模型为年龄模型。

1 模型推导

设
$$y+\epsilon=f(A) \tag{1}$$
y 可以是直径、材积、树高等因子, A 为年龄,这是通常的年龄模型的形式。设前期观测值为 (y_1, A_1),后期观测值为 (y_2, A_2), $A_2=A_1+\Delta A$,则
$$\begin{cases} y_1+\epsilon_{y_1}=f(A_1) & (2a) \\ y_2+\epsilon_{y_2}=f(A_1+\Delta A) & (2b) \end{cases}$$
从式(2a)中解出 A_1, $A_1=f^{-1}(y_1+\epsilon_{y_1})$,将 A_1 代入式(2b),有

1996-05-08 收稿。
葛宏立工程师,项小强,何时珍(林业部华东设计院 浙江金华 321001);方陆明(浙江林学院计算中心)。
* 本文为联合国援助项目 CPR/91/151"建立国家森林资源监测系统"的部分内容。

$$y_2+\epsilon_{y_2}=f(f^{-1}(y_1+\epsilon_{y_1})+\Delta A) \tag{3}$$

此式即为以前期观测值 y_1 为自变量的模型, 它不含实际年龄, 我们也可将 $A_1=f^{-1}(y_1+\epsilon_{y_1})$ 称为名义年龄(nominal age)。应用时的形式为:

$$\hat{y}_2=f(f^{-1}(y_1)+\Delta A) \tag{4}$$

若式(1) 为 Johnson-Schumacher 模型, 即:

$$y+\epsilon=a\mathrm{e}^{-b/A} \tag{5}$$

则相应地式(3) 为:

$$y_2+\epsilon_{y_2}=a\exp\left(\frac{-b}{\Delta A-\dfrac{b}{\ln(y_1+\epsilon_{y_1})-\ln a}}\right) \tag{6}$$

2　模型的性质

模型(3) 具有以下性质:

①模型左边含有误差 ϵ_{y_2}, 右边含有误差 ϵ_{y_1}, 即除了因变量 y_2 有误差外, 自变量 y_1 也有误差。对这种模型, 一般不能用普通最小二乘法估计参数, 而应用专门的方法, 算法可参考文献[1～3]。

②对于预测形式(4), 令 $\epsilon=\hat{y}_2-y_2$, 显然, ϵ 与 ΔA 有关, 当 ΔA 增大时, ϵ 的绝对值也增大, 当 $\Delta A=0$ 时, $\epsilon\equiv 0$, 即误差为零。设 ΔA_0 为检验或建模时的间隔期, $S_\epsilon(\Delta A_0)$ 为 ϵ 的检验或建模时的标准差, 则预测 ΔA 时, 对其标准差 $S_\epsilon(\Delta A)$ 可作线性假设

$$S_\epsilon(\Delta A)=\frac{\Delta A}{\Delta A_0}S_\epsilon(\Delta A_0) \tag{7}$$

由于 ϵ 与 ΔA 有关, 所以在建模时, ΔA 应是一个常数, 如果不是常数, 则在拟合时, 应根据不同的 ΔA 而给不同的权。

设 n 为样地数, m_i 为第 i 块样地的样木数, 若 ϵ 相互独立, 则

$$\mathrm{Var}(\Sigma y_2)=\Sigma m_i\mathrm{Var}(\epsilon)\approx S_\epsilon^2(\Delta A)\Sigma m_i \tag{8}$$

即 $S_\epsilon^2(\Delta A)\Sigma m_i$ 可作为总量 Σy_2 的方差, 这时, 以 $\Sigma\hat{y}_2$ 作为 Σy_2 的估计值, 其误差限为

$$\Delta=u_\alpha S_\epsilon(\Delta A)\sqrt{\Sigma m_i} \tag{9}$$

u_α 为标准正态分布时与概率为 α 相对应的双侧分位数。

但因模型中没有立地质量, 竞争等指标, 所以对样地来说, 估计值可能存在系统偏差, 对于立地质量好的来说, 估计值可能系统偏低, 对于立地质量差的来说, 可能系统偏高。所以, 在样地内, ϵ 可能不相互独立, 这时, 上面的误差限的算法不一定正确。为此, 令 $\bar{\epsilon}=\overline{\hat{y}}_2-\overline{y}_2$, $\overline{\hat{y}}_2$ 和 $\overline{y}_2$ 为样地的后期的估计值的平均数和观测值的平均数。可以认为, $\bar{\epsilon}$ 是相互独立的。令:

$$S_{\bar{\epsilon}}(\Delta A_0)=\sqrt{\Sigma m_i\bar{\epsilon}^2/\Sigma m_i} \tag{10}$$

$$S_{\bar{\epsilon}}(\Delta A)=\frac{\Delta A}{\Delta A_0}S_{\bar{\epsilon}}(\Delta A_0) \tag{11}$$

这时, Σy_2 的方差为

$$\mathrm{Var}(\Sigma y_2)=\mathrm{Var}(\Sigma m_i y_2)=S_{\bar{\epsilon}}^2(\Delta A)\Sigma m_i^2 \tag{12}$$

这时, 误差限为

$$\Delta = u_{\alpha} S_{\bar{\epsilon}}(\Delta A) \sqrt{\sum m_i^2} \tag{13}$$

当 ϵ 相互独立时, 式(9) 与式(13) 是近似相等的, 令 $u_{\alpha} S_{\epsilon}(\Delta A) \sqrt{\sum m_i} = u_{\alpha} S_{\bar{\epsilon}}(\Delta A) \sqrt{\sum m_i^2}$ 可解得

$$S_{\bar{\epsilon}}(\Delta A) = S_{\epsilon}(\Delta A) \sqrt{\sum m_i / \sum m_i^2} \tag{14}$$

当 ϵ 不相互独立时, 式(14) 并不成立, 且总有

$$S_{\bar{\epsilon}}(\Delta A) > S_{\epsilon}(\Delta A) \sqrt{\sum m_i / \sum m_i^2} \tag{15}$$

所以, 当 ϵ 不相互独立时, 用式(13) 算得的误差限要比用式(9) 算得的误差限大, 一般应该用式(13) 计算误差限, 因为, 即使立地质量等因子都考虑了, ϵ 也不可能完全相互独立。上面的推导是对单木模型而言的, 如是平均因子模型, 则不涉及(9) 式的算法, 只用(13) 式。

这里要指出, $\mathrm{Var}(\epsilon)$ 与 $\mathrm{Var}(\epsilon_{y_2})$ 是不一样的, 且总有 $\mathrm{Var}(\epsilon) > \mathrm{Var}(\epsilon_{y_2})$。因为 ϵ_{y_2} 是在考虑 ϵ_{y_1} 的情况下计算的, 而 ϵ 是在模型建立后, 认为 ϵ_{y_1} 为零的情况下计算的, 所以, ϵ 的绝对值要比 ϵ_{y_2} 大。

3 计算例子

数据取自浙江省监测体系的 1989、1994 两期数据的马尾松复测样木数据, $\Delta A_0 = 5$, 模型分人工林、天然林、人工林和天然林合并的年龄隐含的单木模型, 变量为材积, 共提取复位样木 3 755 株, 剔除了严重异常的样木 72 株, 参加建模的 3 683 株。株数、样地数及计算精度的株数见表 1, 计算精度的株数及 $\sum m_i^2$ 均包括建模时被剔除的株数, 年龄隐含模型为(6) 式, 拟合的结果见表 2。

表 **1** 样木株数、样地数

森林类型	建模株数	样地数	计算精度株数	$\sum m_i^2$
人工林+ 天然林	3 683	384	3 755	75 123
人工林	728	71	738	15 538
天然林	2 955	313	3 017	59 585

表 **2** 年龄隐含模型的拟合结果

森林类型	a	b	Q_1	Q_2	Q_{ϵ}	S_{ϵ}	r_{ϵ}	$Q_{\bar{\epsilon}}$	$S_{\bar{\epsilon}}$	λ
人工林+ 天然林	13.636 1	366.553	9.717 2 $\times 10^{-2}$	0.156 567	1.302 32	1.862 32 $\times 10^{-2}$	0.984 14	0.598 159	1.262 13 $\times 10^{-2}$	0.620 873
人工林	342.566	926.668	1.566 85 $\times 10^{-2}$	2.939 57 $\times 10^{-2}$	0.155 888	1.453 38 $\times 10^{-2}$	0.956 35	6.504 14 $\times 10^{-2}$	9.387 86 $\times 10^{-3}$	0.530 584
天然林	13.183 6	363.431	8.043 93 $\times 10^{-2}$	0.127 525	1.137 80	1.941 99 $\times 10^{-2}$	0.985 451	0.533 467	1.329 74 $\times 10^{-2}$	0.630 911

表 2 中的 Q_1 和 Q_2 为(6) 式中的 ϵ_{y_1}(即 $\epsilon_{v_{89}}$) 和 ϵ_{y_2}(即 $\epsilon_{v_{94}}$) 的平方和, Q_{ϵ} 为 $\epsilon = \hat{V}_{94} - V_{94}$ 的平方和, S_{ϵ} 和 r_{ϵ} 为根据 Q_{ϵ} 算得的标准差和相关指数, $Q_{\bar{\epsilon}} = \sum m_i \bar{\epsilon}^2 = \sum m_i (\overline{\hat{V}}_{94} - \overline{V}_{94})_i^2$, $S_{\bar{\epsilon}} = \sqrt{Q_{\bar{\epsilon}} / \sum m_i}$, $\hat{\lambda}$ 为前期方差与后期方差的比值的估计值, $\hat{\lambda} = S_{y_1}^2 / S_{y_2}^2 = Q_1 / Q_2$。计算 Q_{ϵ}、S_{ϵ}、r_{ϵ}、$Q_{\bar{\epsilon}}$、$S_{\bar{\epsilon}}$ 时都用到了被剔除的数据。这些值都是 $\Delta A_0 = 5$ 时的值, 为简便起见, 这里未加注明。

精度分析情况见表 3, 被剔除的数据也参加计算。蓄积的单位为 m^3。ΔM 为实际 5 a 生长

量, $\Delta M=\sum V_{94}-\sum V_{89}$; $\Delta\hat{M}$ 为模型 5 a 生长量, $\Delta\hat{M}=\sum\hat{V}_{94}-\sum\hat{V}_{89}$; $\Delta_{理}$ 为理论误差限, 根据(13)式计算, u_α 为 1.96, 保证概率为 95%, $\sum m_i^2$ 见表 1, $S_{\bar{\epsilon}}(\Delta A)$ 见表 2 中的 $S_{\bar{\epsilon}}$; $E_{理生}$为生长量的理论相对误差, $E_{理生}=\Delta_{理}/\Delta\hat{M}\times100\%$, $E_{理总}$为总生长量的理论相对误差, $E_{理总}=\Delta_{理}/\sum\hat{V}_{94}\times100\%$; $P_{理生}$为生长量的理论相对精度, $P_{理生}=1-E_{理生}$, $P_{理总}$为总生长量的理论相对精度, $P_{理总}=1-E_{理总}$; $\Delta_{实}$ 为实际误差, $\Delta_{实}=\sum V_{94}-\sum\hat{V}_{94}=\Delta M-\Delta\hat{M}$; $E_{实生}$、$E_{实总}$、$P_{实生}$和 $P_{实总}$为实际的相对误差和相对精度。表中没有* 的栏目为用对应的数据建模, 用对应的数据(包括被剔除样木)计算精度, 有* 的为用人工林+ 天然林的模型单独用人工林或天然林数据的精度。精度特别分定期生长量及总生长量, 而且重要的还是定期生长量的精度, 因为预测的年限一般不超过一个复查间隔期, 如果估计的误差还大于间隔期内的生长量, 那么这种预测就没有什么意义了, 从表 3 中可以看出, 生长量的理论相对精度大多在 80% 以上, 总生长量的理论精度在 93% 以上。

表 **3**　精度分析

森林类型	$\sum V_{89}$	$\sum V_{94}$	$\sum\hat{V}_{94}$	ΔM	$\Delta\hat{M}$	$\Delta_{理}$	$E_{理生}$	$E_{理总}$	$P_{理生}$	$P_{理总}$	$\Delta_{实}$	$E_{实生}$	$E_{实总}$	$P_{实生}$	$P_{实总}$
人工林+ 天然林	154.907	220.465	219.799	65.558	64.892	6.780	10.45	3.08	89.55	96.92	0.666	1.02	0.30	98.98	99.70
人工林	21.681	33.409	32.778	11.728	11.097	2.294	20.67	7.00	79.33	93.00	0.631	5.38	1.89	94.62	98.11
人工林*	21.681	33.409	33.109	11.728	11.428	2.265	19.82	6.84	80.18	93.16	0.30	2.56	0.90	97.44	99.10
天然林	133.226	187.056	186.423	53.830	53.197	6.362	11.96	3.41	88.04	96.59	0.633	1.18	0.34	98.82	99.66
天然林*	133.226	187.056	186.690	53.830	53.464	6.370	11.91	3.41	88.09	96.59	0.366	0.68	0.20	99.32	99.80

年龄隐含模型精度高的主要原因是误差只从调查年份(设这时树木的年龄为 A_1)开始积累, 在开始($\Delta A=0$)时不存在误差, 随着 ΔA 的增大, 误差逐渐增大, 只要 ΔA 不是太大, 总能有较高的精度。当 $y_1=0$, 即 $A_1=0$ 时, $A_2=A_1+\Delta A=\Delta A$, 这时, 年龄隐含模型退化为年龄模型, 从理论上讲, 只有在这种情况下, 年龄隐含模型的精度才会和年龄模型同样低。当 ΔA 很大时, 会接近这种情况。

对于精度分析, 最好是用连续三期的复位样木数据, 前二期数据用来建模, 第三期的数据用来检验, 但目前尚找不到连续三期的复位样木数据, 用建模的数据来作精度估计, 还是能基本说明问题的。

4　结论与建议

(1) 年龄隐含的生长模型具有较高的精度; 使用时只要知道上次的调查值 y_1 和 ΔA 就行, 使用方便, y_1 是实测值, 除了量测错误等异常情况外, 误差较小, 而 ΔA 则不存在误差, 所以模型的使用精度能得到保证; 森林资源连续清查体系能提供建模所需的成对复测的样木资料。所以, 年龄隐含的生长模型可以在连续清查体系中推广应用。

(2) 本文建议用单木模型, 因为单木模型适用范围广。以省为建模单元, 不区分人工林和天然林, 同龄林和异龄林, 纯林和混交林; 不区分优势树种和组成树种; 也可不区分林木和散生木。在应用时, 根据样木的树种用对应的模型, 这样可减少建模数量, 扩大建模样本容量。

(3) 年龄隐含模型是以年龄模型为基础的, 而前者在结构及建模上比后者复杂, 加上连续清查中 ΔA 一般不大的实际情况, 建议用简单的年龄模型作为基础模型, 而且这个基础模型要便于写成年龄的反函数形式。

参 考 文 献

1 葛宏立. 自变量也含误差的非线性模型拟合. 北京林业大学学报, 1996, (2): 73～77.

2 王柱. 最小平方距离法介绍. 数理统计与管理, 1986, (4) : 34～36.

3 唐守正. 利用对偶回归和结构关系建立林分优势高和平均高模型. 林业科学研究, 1991, 4(增刊) : 57～62.

Application of the Age-implicit Growth Model to Continuous Forest Inventory

Ge Hongli Xiang Xiaoqiang He Shizhen Fang Luming

Abstract This paper discusses the application of an age-implicit growth model to the National Continuous Forest Inventory (NCFI). The error of the model is related to the projection interval ΔA and will be zero when ΔA is zero. An example taken from that age-explicit Johnosn-Schumacher model is used as a base model from which an age-implicit form is derived. The experimental data are Masson pine's from Zhejiang Province's NCFI. The estimation precision of 5 years' total increment is greater than 80% for 95% confidence. The derivation, properties and error estimation of the model are presented. Data to be used for building age-implicit models must be those of being remeasured and matched. It is easy to use the model. Age-implicit models could be built at province level by tree species or group without considering the differences of plantation and natural, even-aged and uneven-aged, pure and mixed, dominant and nondominant, as well as stand trees and scattered trees.

Key words growth model age-implicit model Johnson-Schumacher model continuous forest inventory

Ge Hongli, Engineer, Xiang Xiaoqiang, He Shizhen (East China Forest Inventory Institute Jinhua, Zhejiang 321001); Fang Luming (The computer center of Zhejiang Forestry College).