浙 江 林 学 院 学 报 1993，10(2)：184～188
Journal of Zhejiang Forestry College

杉木马尾松几个模型的预报有效性的计算机模拟试验*

林思祖　黄青峥**　吴旺民***

（福建林学院，南平 353001）

摘　要　用计算机模拟试验的方法，对杉木和马尾松平均树高、平均胸径及最大密度的数学模型的预报有效性进行检验。结果表明：各模型的模拟检验精度均达到99.0%以上，并且 $P\{R>R^*\}>85.0\%$。可认为各模型的总体相关系数及其预报效果是理想的。

关键词　计算机模拟；模拟试验；数学模型；杉木；马尾松
中图分类号　S711

在森林经营活动中，许多指标都需要以平均胸径、平均树高及最大密度模型为基础。这些模型不仅是制定产量预测的依据和经营数表及林分收获表的基础，而且是计算机辅助造林设计系统中的重要组成部分。这些数学模型即“理想”的模型是来源于“现实”系统中的调查观察，而这些模型的可靠性、实用性和精确性及稳定性决定了计算机辅助造林设计系统能否准确可靠稳定地反映“现实”系统。一般人们判断数学模型优劣的指标是相关系数的大小，这是正确的，但仅仅视相关系数的大小还是不够的。

建立这些模型的资料是从该总体中随机抽取的1个相对于总体单元数来说是很小容量的样本。它们是否有足够的代表性和可靠性反映林分生长过程，是值得我们进一步探讨的。在建立数学模型的实践中可以看到来源于同一地区的样本，各个学者用同一数学方程所建立的数学模型的相关系数是有差异的。理论上研究也证实相关系数是有差异的[1,2]。如果用同一总体的部分资料建立的数学模型的相关系数很大，但重演时相关系数若有很大变化的话，那么应用该模型预测同一总体的另一部分资料必定是不够准确的。所以必须对上述几个模型的预报效果的可靠性和稳定性进行统计检验。以往人们的一般做法是把收集样本分成两部分：一部分用来建立模型，另一部分用来检验模型预报效果的有效性。这是困难的。因此，为了克服这种困难，笔者提出利用统计模拟方法用计算机伪造调查观测数据中的预报量来检验上述几个模型预报的有效性[3]。

收稿日期：1992-03-15

*福建省林业厅资助课题的子课题之一；**现在寿宁县林业局工作；***现在尤溪县中仙林业站

1　模型和计算机模拟的原理

从总体中抽取若干组实测数据(x_1，y_1，x_2，y_2，…，x_i，y_i)$i=1$，2，…，n。假定预测量 y_i 满足线性回归方程：

$$y_i=f(x_i)+\varepsilon_i \tag{1}$$

式中，x_i 为自变量，y_i 为预报量，ε_i 是相互独立并且遵从$N(0,\sigma)$分布的随机变量。应用回归分析法得到回归方程的估计式，并可求得相关系数的估计值：

$$R^*=\sqrt{1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{2}$$

式中 $\hat{y}_i$ 表示回归方程中给出的预报值即理论值，$\bar{y}=\frac{1}{n}\sum_{i=1}^{n}y_i$ 是预报量 y 的平均值。

为了检验模型（1）和随机预报差异是否显著，由计算机产生遵从$N(0,\sigma)$的伪随机数ε_i，从而伪造了几个因变量 y_i，计算伪造因变量 y_i 的回归方程并求得相关系数 R 的抽样值[4,5]。

重复上述 Monte-Carlo 模型，得到R的平均值，并寻找R的经验分布，从而判断模型的优劣。

为了得到正态分布的伪随机数 ε_i，应用下列近似正态抽样方法抽样[6]。

假定 ξ_i ($i=1$, 2, …, m) 是相互独立并且遵从 $N(0,1)$ 上的均匀分布的随机变量，而 $\xi_i-\frac{1}{2}$就是相互独立遵从在$\left(-\frac{1}{2},\ \frac{1}{2}\right)$上的均匀分布的随机变量。

由(ЛЯИУИОВ)中心极限定理推知，当m充分大时，随机变量

$$\begin{aligned}\xi&=\left[\left(\xi_1-\frac{1}{2}\right)+\left(\xi_2-\frac{1}{2}\right)+\cdots+\left(\xi_m-\frac{1}{2}\right)\right]\cdot\sqrt{\frac{12}{m}}\\&=\left[(\xi_1+\xi_2+\cdots+\xi_m)-\frac{m}{2}\right]\cdot\sqrt{\frac{12}{m}}\end{aligned} \tag{3}$$

是近似遵从$N(0,1)$，所以

$$\eta=\sigma\cdot\varepsilon=\left\{\xi_1+\xi_2+\cdots+\xi_m-\frac{m}{2}\right\}\cdot\sqrt{\frac{12}{m}}\cdot\sigma \tag{4}$$

是遵从$N(0,\sigma)$的随机变量，Shreider(1964)指出 $m=6$ 时，(4)式已足够近似正态分布[3]。此处$m=6$，因此产生均值为 0，方差为 σ^2 的正态分布随机变量公式为[6]：

$$\eta=\sigma\cdot\sqrt{2}\{\xi_1+\xi_2+\xi_3+\xi_4+\xi_5+\xi_6-3\} \tag{5}$$

2　数据来源和计算机模拟

笔者于1984年在福建省闽北地区 8 个县市收集了339块样地。样地进行每木检尺，分别径阶测树高，调查大于和等于平均木的林木冠幅，实测2 664株树冠，建立了计算机辅助造林设计系统[6]。其中平均高，平均胸径及最大密度模型分别为：

$$\overline{H}=a+bH_{优} \tag{6}$$

$$\overline{D}=a+b\overline{H} \tag{7}$$

$$N_{max}=a+b/\overline{D} \tag{8}$$

表1 树种平均高数学模型*

Table 1 $\overline{H}$-Hd models of *Cunninghamia lanceolata* and *Pinus massoniana*

树种	平均高模型	相关系数	回归剩余标准差
杉木	$\overline{H}=0.8978H_{优}-0.5510$	0.98401	0.7596
马尾松	$\overline{H}=0.9656H_{优}-1.0467$	0.98917	0.6575

表2 树种平均胸径数学模型

Table 2 $\overline{D}$-$\overline{H}$ models of *Cunninghamia lanceolata* and *Pinus massoniana*

树种	平均胸径模型	相关系数	回归剩余标准差
杉木	$\overline{D}=1.1928\overline{H}+1.3980$	0.93015	2.0161
马尾松	$\overline{D}=1.1670\overline{H}+0.3177$	0.95440	1.6415

表3 树种最大密度数学模型**

Table 3 Maximum density-$\overline{D}$ models of *Cunninghamia lanceolata* and *Pinus massoniana*

树种	最大密度模型	相关系数	回归剩余标准差
杉木	$N_{max}=2.6892+2346.995/\overline{D}$	0.99998	0.6714
马尾松	$N_{max}=12.0997+1410.918/\overline{D}$	0.99998	0.5929

注：*、**文献[6]中的杉木平均高模型$R=0.97$，马尾松最大密度模型$R=0.96$为笔误，谨此致歉

式中：$H_{优}$为优势高，$\overline{H}$为平均高，$\overline{D}$为平均胸径，N_{max}为最大密度，a，b为参数。回归分析结果列于表1，2，3。

根据上述原理，用上面模型的基础数据，对6个模型的预报有效性分别进行模拟试验。每个模型共计模拟100次。模拟第K次前相关系数平均值记为：

$$\overline{R}_K=\frac{1}{K}\sum_{i=1}^{K}R_i \tag{9}$$

模拟100次后相关系数的平均值见表4。模拟80次以后相关系数的平均值$\overline{R}_i$（$i=80,\cdots,$

表4 各模型模拟检验结果表

Table 4 Computer simulation tests on the models

模型	R^*	$\overline{R}n$	Pc(%)	$P(R>R^*)$(%)	α(%)
杉木平均高	0.98401	0.98893	99.95	96.0	85
马尾松平均高	0.98917	0.99270	99.97	98.0	85
杉木平均胸径	0.93015	0.94449	99.76	91.0	85
马尾松平均胸径	0.95440	0.96498	99.86	92.0	85
杉木最大密度	0.99998	0.99999	99.99	87.1	85
马尾松最大密度	0.99998	0.99999	99.99	89.1	85

100)变化不大已经达到稳定状态，并且已足够用来寻求R的经验分布。

当总体相关系数$R \neq 0$时，可以95.5%的可靠性估计总体相关系数R的估计值，用下式计算：$R = \frac{1}{100}\sum_{i=1}^{100} R_i$。其结果列于表4。

从上表可以看出，模拟检验的相关系数平均值与模型相关系数极为相近，并且各模型的模拟检验精度均大于99.0%以上，故说明模型的总体相关系数是理想的。

3　各模型预报有效性检验

反复用Monte-Carlo方法，在保留相同预报因子x_i以及样地数量的条件下，得到伪造因变量y_i的回归方程，求得相关系数R的一系列抽样值，并寻求R的经验分布。

对于给定的显著水平α，若$P\{R>R^*\}>\alpha$成立，则认为该模型预报效果和随机预报效果相比差异不显著，即认为预报效果是有效的。

从各个系列R的抽样值频率分布来分析是近似正态分布。下面针对杉木平均高模型所进行的模拟检验的一系列R的抽样值遵从正态分布进行分布假设检验，计算见表5。

表5　杉木平均高模型R分布假设检验

Table 5　Hypothensis test on R distribution of $\overline{H}$-H_d model of *Cunninghamia lanceolata*

R值分组	实际频数 G_i	上限$-\overline{R}$	$U_i=\frac{上限-\overline{R}}{R'_R}$	$\Phi(U_i)$	理论概率值 P_i	理论频数 nP_i
0～0.984	4	−0.004 93	−2.062 76	0.019 70	0.019 70	1.970 00
0.984～0.985	3	−0.003 93	−1.644 35	0.050 50	0.030 80	3.080 00
0.985～0.986	7	−0.002 93	−1.225 94	0.109 30	0.058 80	5.880 00
0.986～0.987	5	−0.001 93	−0.807 53	0.209 00	0.099 70	9.970 00
0.987～0.988	12	−0.000 93	−0.389 12	0.352 00	0.143 00	14.300 00
0.988～0.989	11	0.000 07	0.029 29	0.512 00	0.160 00	16.000 00
0.989～0.990	21	0.001 07	0.447 69	0.673 60	0.161 60	16.160 00
0.990～0.991	21	0.002 07	0.866 11	0.807 80	0.134 20	13.420 00
0.991～0.991	8	0.003 07	1.284 52	0.899 70	0.091 90	9.190 00
0.992～0.993	6	0.004 07	1.702 93	0.955 43	0.055 74	5.574 00
0.993～1	2	0.011 07	4.631 80	0.999 99	0.044 56	4.456 00
	100	—	—	—	1	100

注：$S'_R=0.002\,39$；$\overline{R}=0.988\,93$

计算
$$\chi^2=\sum_{i=1}^{11}\frac{(G_i-nP_i)^2}{nP_i} \tag{10}$$

得χ^2值为14.011 13，查表得$\chi^2_{0.05}=15.507$，$\chi^2<\chi^2_{0.05}$，所以认为R的抽样值的经验分布为正态分布。

若给85.0%的显著水平，见表4，$P\{R>R^*\}$，各模型的$P\{R>R^*\}$均大于85.0%。

因此认为各个模型的预报效果与随机预报效果没有显著差异。分别用各模型来预报杉木、马尾松平均高、平均胸径、最大密度是有效的。

4 结论

4.1 依据各模型的线性回归方程 $y_i=f(x_i)+\varepsilon_i$，其误差项是遵从正态分布的原理，通过计算机模拟实验，解决了计算机辅助造林设计系统中的杉木、马尾松的平均高、平均胸径以及最大密度等几个数学模型的预报有效性的检验问题。

4.2 由检验可看出，对各模型模拟检验的相关系数平均值与模型相关系数极为相近，无显著性差异，并且各模型的模拟检验精度均达到99.0%以上，并且 $P\{R>R^*\}>85.0\%$，总体相关系数及其预报效果是理想的。

4.3 通过计算机模拟检验，表明各模型预报效果与随机预报效果没有显著差异，$P(R>R^*)>85.0\%$其预报效果良好。

参考文献

1 周华章．工业技术应用数理统计．北京：人民教育出版社，1964：126～167

2 〔瑞典〕Cramer H．统计学的数学方法．上海：上海科学技术出版社，1960：268～270

3 洪 伟，吴 敬．东北林学院学报，1983，11(2)：149～153

4 张建中．数学实践与认识，1974，(1～2)：28～40，43～56

5 中国科学院．概率统计计算．北京：科学出版社，1979：357～439

6 俞新妥，林思祖．林业科学，1986，22(4)：337～345

Lin Sizu (Fujian Forestry College, Nanping 353001, PRC), Huang Qingzheng, Wu Wangming. **A Computerized Simulation Test on the Effectiveness of Several Models of Chinese Fir and Masson Pine.** *J Zhejiang For Coll*, 1993, **10**(2): 184～188

Abstract: This paper deals with the effectiveness of $\overline{H}-H_d$ models, $\overline{D}-\overline{H}$ models and maximum density-$\overline{D}$ models of Chinese fir and Masson pine, making use of computerized simulation tests. The results showed that the simulation test precisions of the models reached up 99.0%, and $P\{R>R^*\}>85.0\%$. The population correlation coefficients and the forecasting effectiveness of the models were satisfactory.

Key words: computerized simulation; simulation tests; mathematical models; Chinese fir; Masson pine