

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

周国模¹,刘恩斌¹,刘安兴²,周宇峰¹

(1. 浙江林学院环境科技学院, 临安 311300; 2. 浙江省森林资源监测中心, 杭州 310020)

摘要:非线性模型迭代参数初始值的选取及拟合的结果能否可用非常重要且一直是一个难题。传统的方法只是凭感觉尝试不进行定量评估,对于用阻尼最小二乘进行拟合只注重精度高低,没有从理论上研究和分析评价得出的结果能否可用。为了克服这一不足,使参数选取较好,阻尼最小二乘得出结果可用。利用非线性回归原理在这两方面做了探讨,提出了二个判别指标: $\beta^T < 1, \beta^N < 1$,分析了每一指标在参数不同状态的控制作用,并在 Weibull 分布函数参数辨识中做了应用。针对目前毛竹林林分结构规律研究的不足,利用浙江省 230 个毛竹林样地资料研究了省域尺度的毛竹林胸径分布规律,经 SPSS 软件中的 P-P 图检验与柯尔莫哥洛夫检验,表明毛竹林胸径结构服从 Weibull 分布;再对 Weibull 分布模型拟合求解与作图对比,结果显示 Weibull 分布能很好地描述毛竹林胸径分布规律。利用 Weibull 分布三参数与林分特征因子间的相关关系,经分析研究,导出了一个改进的二元分布模型,用该二元分布模型很好地测度了浙江省毛竹林胸径和年龄分布规律,最后得到了浙江省毛竹林株数按胸径年龄分布的理论频数。

关键词:Weibull 分布;参数辨识;毛竹林;直径分布;二元分布模型

文章编号:1000-0933(2006)09-2918-09 **中图分类号:**Q14, S718. 5, S758 **文献标识码:**A

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, China

ZHOU Guo-Mo¹, LIU En-Bin¹, LIU An-Xing², ZHOU Yu-Feng¹ (1. School of Environment Sciences, Zhejiang Forestry University, Lin'an 311300, China; 2. Monitoring Center for Forest Resources, Zhejiang Province, Hangzhou 310020, China). Acta Ecologica Sinica, 2006, 26(9): 2918 ~ 2926.

Abstract: It is very important and difficult to select initial parametric value and to evaluate the result which is applicable or not of nonlinear model. To solve nonlinear equation with Levenberg-marquart iteration method, the traditional approach to select initial parametric value is based on the sense and multiple testing, and only focuses on the accuracy of the fitted value while ignores the applicability of the result theoretically. These shortcomings were overcome in this study through. In order to conquer these shortcomings, this paper makes the Nonlinear Regression Statistics and two key estimation indices $\beta^T < 1, \beta^N < 1$. The control function of $\beta^T < 1, \beta^N < 1$ in the various parametric states were analyzed and used to solve parameters of Weibull Distribution function as well. About 230 samples datum of Moso bamboo (*Phyllostachys pubescens*) in Zhejiang Province were applied to study the distribution pattern of diameter of Moso bamboo forests. The results show that the diameter distribution pattern of Moso bamboo forest is conformed to Weibull Distribution through the P-P chart and Kolmogorov test. Using of the relationship with Weibull's three parameters and characteristic factors of forests, two-dimensional distribution function is induced. In the presented report, we

基金项目:国家自然科学基金资助项目(30271072);浙江省科技厅资助项目(2003C32029)

收稿日期:2006-01-09; **修订日期:**2006-07-11

作者简介:周国模(1961~),男,浙江省诸暨人,博士,教授,主要从事森林资源与环境管理研究。E-mail:zhougm@zjfc.edu.cn

Foundation item: The project was supported by National Natural Science Foundation of China (No. 30271072); the Department of Science and Technology of Zhejiang Province (No. 2003C32029)

Received date: 2006-01-09; **Accepted date:** 2006-07-11

Biography: ZHOU Guo-Mo, Ph.D., Professor, mainly engaged in forest resources and environment management. E-mail: zhougm@zjfc.edu.cn

also demonstrated that two-dimensional distribution function had high accuracy in the measurement on the distribution patterns of diameter and age of Moso bamboo forests. Finally, the theoretical probability value of diameter and age distribution of Zhejiang Province were presented.

Key words: Weibull Distribution; parametric identification; Moso bamboo (*Phyllostachys pubescens*) forest; diameter distribution; two-dimensional Distributing Model

浙江省位于我国东南沿海,地处中亚热带,其地理坐标为东经 118°01′~123°10′,北纬 27°06′~31°11′。是我国毛竹 (*Phyllostachys pubescens*) 分布最丰富的省份之一。据 2005 年统计资料,全省有毛竹林面积 57.53 万 hm^2 , 占全省森林面积的 9.84%, 竹林产值 180 亿元, 是浙江省山区农民经济收入的重要来源。据研究毛竹林还是重要的森林碳汇, 其年固碳能力位居亚热带区域树种之最^[1], 对全球气候变化具有不可低估的作用。而目前在有关毛竹林的研究方面, 还未见有毛竹林胸径年龄分布规律的研究报道。毛竹林的胸径年龄分布状况与毛竹林群落的产量结构及机能有着密切的关系, 合理的毛竹林分结构是充分发挥毛竹林效能的基础, 研究毛竹林胸径年龄分布规律可预估毛竹林各径级年龄株数, 进而为计算毛竹林生物量及生物碳库提供了科学依据。

选择一个合适的概率密度函数, 对毛竹林直径分布模型的建立和使用起着至关重要的作用, 研究认为灵活性大、适应性强且积分形式简单的威布尔分布 (Weibull Distribution) 能很好地描述林木的直径分布状态^[2], 它同样可用来描述毛竹林直径分布规律。目前, 威布尔分布已应用于生态领域^[3,4], 在林业上应用最广泛的还在于研究林木直径分布方面^[2,5~9]。

有关 Weibull 分布参数辨识的方法最常用的是最小二乘法, 但大多数研究都普遍注重参数拟合精度^[6,7,9], 而对参数初始值选取好坏评价及拟合得到的结果能否可用还未有研究。

本文试图用 Weibull 分布来研究浙江省毛竹林的直径年龄分布规律, 并对 Weibull 分布参数辨识改进方面做一些探索性工作。

1 资料来源

浙江省于 1979 年建立了森林资源连续清查体系, 以 5a 为一个复查周期。共设置固定样地 4250 个, 样点格网为 4km × 6km, 样地形状为正方形, 边长 28.28m, 面积 800m², 其中属于毛竹纯林的样地共有 230 个, 本研究即取此 230 个样地资料。样地内 5cm 以上的竹子均要调查记载, 调查内容主要有量测胸径, 测定年龄 (当年生新竹记为 0 度竹, 仅记录株数, 不测胸径; 1~2 年生竹记为 1 度竹; 3~4 年生竹记为 2 度竹, 依此类推)。

230 个大小为 28.28m × 28.28m 的毛竹样地资料的基本统计数据见表 1。其胸径从 5~15cm, 年龄为 1~4 度以上, 样地毛竹株数从 18~461 株不等。

表 1 浙江省毛竹林胸径年龄样地统计 (株)
Table 1 Sample data (No.) on diameter and age of Moso bamboo (*Phyllostachys pubescens*) in Zhejiang Province

龄阶 (度 °) Age class (du °)	胸径 Diameter (cm)										
	5	6	7	8	9	10	11	12	13	14	15
1	1047	1238	1747	1817	1698	1292	637	295	86	23	3
2	898	1285	1783	1986	1845	1386	768	384	109	31	11
3	275	426	624	814	693	590	322	169	49	7	2
4 及以上 4 above	125	191	319	422	462	394	266	145	40	10	4
总数 Total	2345	3140	4473	5039	4698	3662	1993	993	284	71	20

*度是毛竹的年龄单位, 1 度包含 2 年 du is a age unit for bamboo, one du contains two years

2 毛竹林直径一元分布模型及求解

2.1 威布尔分布函数

威布尔分布的一般形式为:

$$f(x, a, b, c) = \left(\frac{x-a}{b} \right)^c \frac{c}{x-a} e^{-\left(\frac{x-a}{b} \right)^c}$$

式中, a, b, c 为威布尔分布的 3 个参数。 a 为位置参数, $b > 0$ 时, 为尺度参数, 是疏散性指标, $c > 0$ 时, 为形状参数; 当 $c < 1$ 时, 威布尔分布为倒 T 型分布; 当 $c = 1$ 时, 威布尔分布为指数分布, 当 $c = 2$ 时, 威布尔分布为卡方(χ^2)分布, 当 $c = 3.6$ 时, 威布尔分布近似正态分布; 当 $c \rightarrow \infty$ 时, 威布尔分布化为单点分布。

2.2 毛竹林胸径服从威布尔分布的假设检验

2.2.1 P-P 图检验 对于已有数据服从哪种分布还不能确定时, 首先要对数据做分布预检验, 以保证后续工作的有效性, 也为确定应采用的统计检验方法提供依据。而 SPSS 中的 P-P 图提供了这种功能^[10]。

SPSS 中的 P-P 概率图是用于检验样本的概率分布是否服从某种理论分布的图形, 检验是根据某变量概率累积比与指定分布概率累积比之间的关系。可以利用 P-P 概率图检验实际数据是否符合指定的分布, 当符合指定分布时, 数据对应的散点位于图中右斜对角位置近于直线分布, 而在无趋势 P-P 图中则呈离散分布, 并对称于以 0 为水平轴的带内。如果不成直线, 但有一定规律, 可以对变量数据按照某种转换方式进行转化, 使转化后的数据更接近指定的分布。

对浙江省实测的毛竹株数按径阶的概率累积分布数据进行 Weibull 分布检验, 生成的检验图形见图 1 和图 2。结果显示: 图 1 所示的 Weibull P-P 图的散点近似呈一条直线; 而无趋势 Weibull P-P 图(图 2)的散点均匀分布在纵坐标为 0 的直线上下, 故可以认为浙江省毛竹株数的径阶分布服从威布尔分布。

2.2.2 柯尔莫哥洛夫检验 上述用 SPSS 软件中的 P-P 概率图较为直观形象地验证了毛竹林胸径服从威布尔分布, 在此进一步对分布函数型式进行理论检验。常用的分布拟合检验方法有: χ^2 检验与柯尔莫哥洛夫检验^[11]。 χ^2 检验法是将总体各单元可能的取值范围分成若干区间, 根据每一区间上理论频数与样本的实际频数值之差对总体分布作出的检验。柯尔莫哥洛夫检验是对每一样本单元值均考虑与其相应理论值的偏差, 而且还是样本累积频率与实测累积频率的比较, 因此它要比 χ^2 检验要精确些。

当不知道总体服从那种类型的分布时, 可用柯尔莫哥洛夫统计量来检验总体是否服从指定的分布, 这里先假定总体服从 Weibull 分布。

对于浙江省毛竹林来说, 230 个样地不同胸径的毛竹总株数可以当作是样本, 根据柯尔莫哥洛夫检验原理得:

$$\lambda = \sqrt{n} \sup |F_n(x) - F(x)| = 0.379219$$

式中, n 为样本单元数, $F_n(x)$ 为实测样本累积频率分布函数, $F(x)$ 为理论累积分布函数, $\sup$ 为取上界。经查表得 $Q(\lambda) = Q(0.379219) = 0.001285$, 则 $1 - Q(\lambda) = 0.998715$ 是一个很大的概率, 所以原假设成立, 毛竹胸径服从 Weibull 分布。

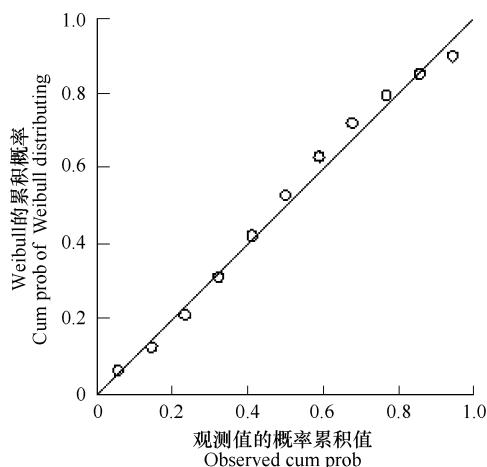


图1 毛竹胸径 Weibull P-P 散点图

Fig. 1 Weibull P-P Plot of Moso bamboo ' diameter

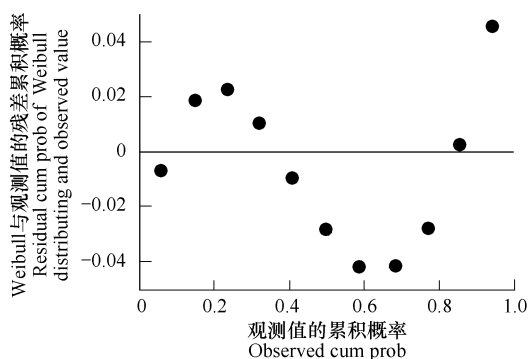


图2 毛竹胸径无趋势 Weibull P-P 散点图

Fig. 2 Detrended Weibull P-P Plot of Moso bamboo ' diameter

经上述两种方法的假设检验,说明浙江省毛竹林直径是服从 Weibull 分布的。

2.3 模型的拟合与检验

2.3.1 非线性模型参数的评价指标 由于各种方法需用数值方法进行迭代求解,在迭代求解过程中,若迭代初始点选得不好,迭代过程不一定收敛。因为有的非线性模型对参数的近似值十分敏感,只有在参数近似值的近似程度非常好的情况下,才能得到可用的结果,有的非线性模型在给定初值处即使收敛也不一定能很好地反映模型对客观实际的描述。这是因为非线性模型可能有几个收敛域,在某个收敛域内迭代得到的结果不一定可用^[12]。在此笔者提出了参数检验标准:

对于非线性模型
$$Y = f(X, \theta) + \varepsilon \quad (1)$$

经过一系列的推理证明得到如下两个指标(具体推理见附录):

$$\beta^N = \frac{W^N d\beta^2}{V^{\frac{2}{2}}} < 1, \beta^T = \frac{W^T d\beta^2}{V^{\frac{2}{2}}} < 1 \quad (2)$$

式中, β^N 是说明模型非线性强度的指标, $\beta^N < 1$ 说明模型非线性强度不大,可以取至泰勒公式展开一阶项进行线性化, β^N 越小则模型非线性强度越弱。 β^T 是说明模型参数选取好坏的指标,当 $\beta^T < 1$ 说明参数选取比较好, β^T 越小则说明模型参数选取的越好。而 β^N 与 β^T 又是紧密相关互相依赖的,模型非线性强度越强越不易线性化,对参数选择就越敏感,参数值越难选。

阻尼最小二乘拟合非线性模型时,其原理就是用泰勒公式逼近原有模型,取一阶项线性化处理。

从附录可以看出: β^N 与 β^T 这两个指标是针对阻尼最小二乘算法而提出的。

以下说明了这两个指标对模型参数状态的控制性评价:

(1) 在模型初值选取时,用 β^N 与 β^T 指标做一次检验,如果 $\beta^N < 1$ 且 $\beta^T < 1$ 说明非线性模型非线性强度不大且初值选取得较好。从下面检验可以看到,如果初值检验能满足 $\beta^N < 1$ 与 $\beta^T < 1$ 则迭代很快收敛,而且很稳定。但只满足 $\beta^N < 1$ 与 $\beta^T < 1$ 时,对有的非线性模型不一定能满足既定精度要求,所以要进行参数迭代。

(2) 当满足既定精度要求时,对有的非线性模型来说,不一定能满足 $\beta^N < 1$ 与 $\beta^T < 1$ ^[5],所以参数迭代终止时,再进行一次 $\beta^N < 1$ 与 $\beta^T < 1$ 指标的检验。如果 $\beta^N < 1$,则能满足阻尼最小二乘算法泰勒公式展开取至一阶项的线性化要求,如果 $\beta^T < 1$,则此时的参数取值也较好。因此要使阻尼最小二乘算法得出结果可用必须满足 3 个条件: $\beta^N < 1$, $\beta^T < 1$,及达到既定精度要求。否则 $\beta^N > 1$ 且即使迭代收敛也能满足既定精度要求,但此时用阻尼最小二乘算法得出的结果,由于不满足其本身所要求的条件,所以信息失真太大而不能用。

具体实施过程可以用简单方便的矩估计^[13~15]大概确定一下模型的初始值或根据 Weibull 3 个参数取值与模型图像形状的关系,结合对应的散点图来取,在程序设计上可以加一个变量用来判断是否满足 $\beta^N < 1$, $\beta^T < 1$ 。

从以上可以看出参数状态可分为初值评价标准确定的参数状态(L_B),初选模型参数原始状态(L_0)与满足精度的参数真值最终近似状态(L_Z),从 L_B 到 L_Z 已证实能收敛,从 L_0 到 L_B 对初值的选取比直接从 L_0 到 L_Z 要容易的多。这就说明 $\beta^N < 1$, $\beta^T < 1$ 对模型初值选取提供了一个理论依据,使非线性模型初值选取更容易且更有目的性,这也是此标准的实际意义所在。

2.3.2 模型求解与检验

(1) Weibull 分布参数评价与相关系数检验 利用 Weibull 三参数取值与模型图像形状的关系,结合对应的散点图,取初值为: $a = 1.25$, $b = 5.72$, $c = 2.58$ 时,用非线性模型参数检验标准得: $\beta^N = 0.088280$, $\beta^T = 0.059859$,经过 15 次迭代就达到稳定,下面只列出了最后 5 次迭代结果(表 2)。这说明了 Weibull 分布的非线性强度是比较弱的,同时也说明了许多学者

表 2 最后 5 次迭代过程

迭代过程 Iteration process	离差平方和 Deviation sum of square	参数 Parameters		
		a	b	c
11	0.000448	1.337068	7.442110	3.605851
12	0.000448	1.336439	7.442711	3.606148
13	0.000448	1.336439	7.442711	3.606148
14	0.000448	1.336666	7.442494	3.606041
15	0.000448	1.336666	7.442494	3.606041

通常采用取常用对数进行线性化处理的原因。

保留前 5 位有效数字得拟合模型为：

$$f(x, \theta) = \left[\frac{x - 1.336666}{7.442494} \right]^{3.606041} \frac{3.606041}{x - 1.336666} e^{\left(-\frac{x - 1.336666}{7.442494} \right)^{3.606041}}$$

从以上参数迭代过程可以得出各参数取值稳定,迭代终止,此时的 $\beta^N = 1.0429 \times 10^{-7}$, $\beta^T = 2.7614 \times 10^{-8}$ 。决定系数为 $R^2 = 0.991220$,能满足精度要求。

根据拟合结果作模型曲线见图 3 所示:从图 3 可以看出毛竹胸径分布理论曲线与实测数据非常吻合。经峰度与偏度检验,得峰度值为 -1.633000,偏度值为 0.004000。

3 毛竹胸径年龄二元分布模型

前文的一元 Weibull 只能反映毛竹株数随径阶的分布规律,还不能反映毛竹株数随直径、年龄的二元分布规律,因此需要对一元 Weibull 分布进行改进。

3.1 Weibull 分布参数预估方程的建立

关于二元 Weibull 分布概率密度函数目前在理论上还不成熟,没有形成像正态分布那样具体的函数表达式,但许多学者研究认为 Weibull 3 个参数与林分特征

因子存在一定的回归关系^[8,9],有的是指数形式的,有的是多元线性回归关系的。现设 Weibull 3 个参数与林分年龄(A)和直径(D)有以下回归关系:

$$\begin{cases} c = c_0 + c_1^* A + c_2^* D \\ b = c_3 + c_4^* A + c_5^* D \\ a = c_6 + c_7^* A + c_8^* D \end{cases} \quad (3)$$

将上述(3)式代入 Weibull 分布模型,得改进的二元分布模型为:

$$f(x, y) = \frac{c_0 + c_1^* A + c_2^* D}{c_3 + c_4^* A + c_5^* D} \left(\frac{D - (c_6 + c_7^* A + c_8^* D)}{c_3 + c_4^* A + c_5^* D} \right)^{c_0 + c_1^* A + c_2^* D - 1} e^{-\left(\frac{D - (c_6 + c_7^* A + c_8^* D)}{c_3 + c_4^* A + c_5^* D} \right)^{c_0 + c_1^* A + c_2^* D}}$$

3.2 二元分布模型参数求解与检验

用阻尼最小二乘法对上述二元改进模型进行拟合得:

$$c_0 = -10.161000, c_1^* = -1.365000, c_2^* = 3.260521, c_3 = -75.972000, c_4^* = 15.700487, \\ c_5^* = 16.571509, c_6 = 96.377912, c_7^* = -17.273000, c_8^* = -17.779000$$

离差平方和为:0.002200,决定系数 $R^2 = 0.902000$,能满足精度要求的。

对拟合后的模型作如下立体图(图 4):由图 4 可以看出,曲线在概率-胸径投影面上还是 Weibull 形状,1 度,2 度胸径曲线投影之所以与二维胸径-概率分布曲线(图 3)最类似,是因为全省毛竹实测资料中 1 度,2 度的毛竹占了较大比重,这就说明了用此模型来描述毛竹胸径年龄分布规律时能充分反映调查样地的实际情况,也说明了该模型对毛竹林分特征反映较二维 Weibull 分布全面,以及年龄对毛竹林分布规律的影响。当然本文在这方面也只是一个初步尝试,有关反映胸径年龄性能更好的模型还有待进一步研究。

3.3 浙江省毛竹胸径年龄实测频数与理论频数对比

根据上述求得的二元分布模型,即可得全省毛竹林分胸径年龄的理论频数,与全省毛竹林的实际频数对比(表 4)。

用表 4 中各径级与各龄级的概率估计值,乘以浙江省毛竹总株数,就可得浙江省毛竹株数按胸径年龄分布的估算值。

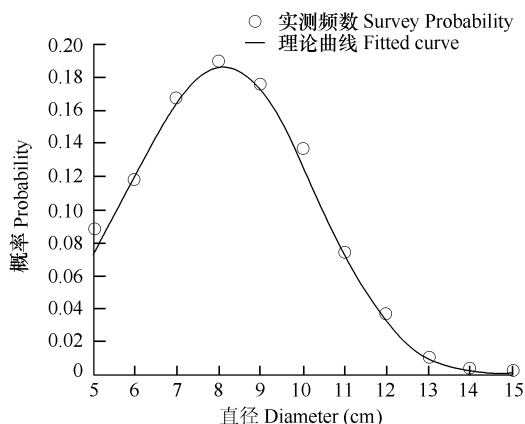


图 3 全省毛竹林胸径分布理论频数曲线图与实测频数散点图

Fig. 3 The data and the fitted Weibull Distribution curve of Moso bamboo's diameter

4 结论与讨论

本文利用浙江省森林资源连续清查样地资料,首次用 Weibull 分布函数来研究全省尺度的毛竹林分的胸径年龄结构规律。基于非线性模型迭代参数值选取的适当与否较难判断,且直接影响到模型估算结果对实际情况的反映,影响到迭代过程的稳定性;同时人们长期只关心用阻尼最小二乘算法得出结果的精度高低,不太关注其结果能否可用这一核心问题,为此对非线性模型的参数求解和评价方法也做了深入探讨,得到了以下结论:

(1) 利用非线性回归原理对参数值选取及阻尼最小二乘法得出结果能否可用进行了探讨。提出了 2 个评价指标,即: $\beta^T < 1, \beta^N < 1$,这样就能保证拟合参数结果较好而且可用,并在一元 Weibull 分布模型参数辨识中做了应用与检验。

(2) 用 Weibull 分布函数能很好地拟合浙江省毛竹林分的胸径结构,拟合模型的精度很高,据此可得到全省毛竹林株数按胸径分布的估算值。进一步表明了 Weibull 分布函数在描述林分直径分布方面具有的广泛适用性。

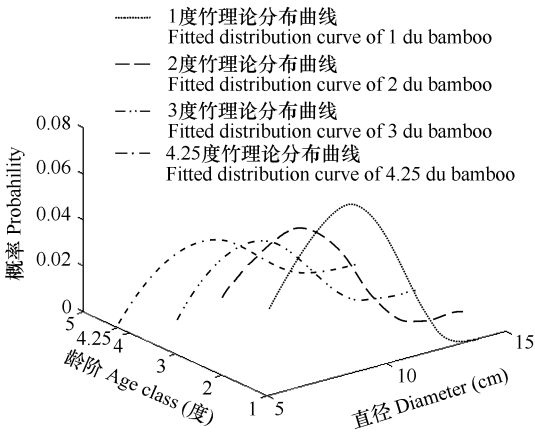


图4 浙江省毛竹林胸径年龄三维理论分布
Fig.4 Three-dimension of the fitted distribution on Moso bamboo's Diameter and age

说明:4度及以上竹是按4度的权0.8,5度的权0.15,6度的权0.05计算而得,其加权平均值是4.25。Note:Moso Bamboo for 4 du and above contained 4,5 and 6 du; here the weighted value is 4.25, which were weighted to 0.8,0.15 and 0.05 for 4 du,5 du and 6 du respectively

表4 浙江省毛竹林胸径年龄实测频数与理论频数表

直径 (cm) Diameter		龄阶 Age class							
		1		2		3		4 以上 above	
		实测值 Observed value	理论值 Fitted value	实测值 Observed value	理论值 Fitted value	实测值 Observed value	理论值 Fitted value	实测值 Observed value	理论值 Fitted value
5		0.039187	0.037433	0.033610	0.033649	0.010293	0.015017	0.004679	0.001763
6		0.046336	0.051950	0.048095	0.043271	0.015944	0.028194	0.007149	0.014763
7		0.065387	0.064981	0.066734	0.050938	0.023355	0.036262	0.011940	0.023120
8		0.068007	0.073341	0.074332	0.054806	0.030466	0.040093	0.015795	0.027656
9		0.063553	0.071277	0.069055	0.051401	0.025938	0.038059	0.017292	0.027542
10		0.048357	0.055318	0.051875	0.039093	0.022083	0.029401	0.014747	0.022303
11		0.023842	0.030574	0.028745	0.021675	0.012052	0.016857	0.009956	0.013665
12		0.011041	0.010123	0.014372	0.007483	0.006325	0.006233	0.005427	0.005625
13		0.003219	0.001547	0.004080	0.001272	0.001834	0.001206	0.001497	0.001300
14		0.000861	0.000074	0.001160	0.000075	0.000262	0.000089	0.000374	0.000129
15		0.000112	0.000001	0.000412	0.000001	0.000075	0.000002	0.000150	0.000004

(3) 首次用改进 Weibull 分布函数得到了一个二元分布模型,并用该二元分布模型研究了浙江省毛竹林的胸径年龄分布规律。拟合的二元分布模型能适应于毛竹林具有不同偏度和峭度变化的情况,并得到了浙江省毛竹株数随胸径年龄分布的理论频数。

(4) 本文采用浙江省 230 个毛竹林样地资料中所有相同胸径或年龄的总株数来研究毛竹林分结构,故其结果只反映了省域尺度的毛竹林胸径年龄分布规律,并不能用来测度具体某个毛竹林林分的胸径年龄分布结构。

(5) 用 Weibull 分布来研究某一树种的林分测度时,最多的是研究胸径这样的一元测度,用 Weibull 分布来研究二元测度文献很少,至今其二元分布函数型式尚属空白。在目前还未有二元 Weibull 分布函数表达式的情况下,本文对 Weibull 分布函数做了改进,拟合结果虽然能反映实际情况,但由于这方面研究成果很少,还有

待于从数学理论上做进一步研究。

References :

- [1] Zhou G M, Jiang P K. Density, storage and spatial distribution of carbon in *Phyllostachypubescens* forest. *Scientific Silvae Sinicae*, 2004, 40(6) :20 ~ 24.
- [2] Zhou G M. Study on diameter distribution for *Cunninghamia lanceolata Fortunei* plantation. *Journal of Fujian College of Forestry*, 1992, 12(4) :399 ~ 405.
- [3] Yang X W, Zhang X X. Reproductive characteristics of the green peach aphid from different regions of China. *Acta Ecologica Sinica*, 2000, 20(1) :139 ~ 144.
- [4] Ren A Z, Gao Y B, Wang J L. Root distribution and canopy structure of *Salix gordejvii* in different sandy land habitats. *Acta Ecologica Sinica*, 2001, 21(3) :399 ~ 404.
- [5] Meng X Y. The study on diameter distribution of *Pinus tabulaeformis* plantation with Weibull function. *Journal of Beijing Forestry College*, 1985, 7(1) :30 ~ 40.
- [6] Hong L X, Du G J, Zhang Q R. Weibull D. B. H. distributions and their dynamic predictions in the natural uneven-aged evergreen broad-leaf juvenile forest. *Acta Phytocologica Sinica*, 1995, 19(1) :29 ~ 42.
- [7] Xing L F, Liang Y T, Ding X T. Computer program design for weibull diameter distribution model in parameter recovery method by c language. *Journal of Shandong Agricultural University*, 1995, 26(4) :425 ~ 434.
- [8] Liu J R, Zhao D F. Weibull distribution parameters and stand factor model of *Larix* plantation. *Scientific Silvae Sinicae*, 1997, 33(5) :412 ~ 417.
- [9] Wu C Z, Hong W. A study on Weibull D. B. H. distribution of Chinese fir plantation and its optimal fitting. *Acta Agriculturae Universitatis Jiangxiensis*, 1998, 20(1) :86 ~ 90.
- [10] Yu J Y, He X H. Statistics and analysis method of data and SPSS application. Beijing: Post & Telecommunications Press, 2003.
- [11] Sheng Z, Xie S Q, Pan C Y. Probability and statistics. Beijing: Higher Education Press, 2003.
- [12] Wang X Z. Theory and application of nonlinear model's parameter estimation. Wuhan: Wuhan University Press, 2002.
- [13] Jorgenson D W. Econometric methods for applied General Equilibrium Analysis. In H. E. Scarf and J. B. Shoven Eds. *Applied General Equilibrium Models*, Cambridge: Cambridge University Press, 1984. 139 ~ 203.
- [14] Deng J, Gu D S, Li X B. Parameters and quantile estimation for fatigue life distribution using probability weighted moments. *Chinese Journal of Computational Mechanics*, 2004, 21(5) :609 ~ 613.
- [15] Hosking J R M, Wallis J R. Parameter and quantile estimation for the generalized Pareto distribution, *Technometrics*, 1987, 29(3) : 339 ~ 349.
- [16] Wei B C. Neoteric nonlinear regression analysis. Nanjing: Southeast University Press, 1989.

参考文献:

- [1] 周国模, 姜培坤. 毛竹林的碳密度和碳贮量及其空间分布. *林业科学*, 2004, 40(6) :20 ~ 24.
- [2] 周国模. 杉木人工林直径分布的研究. *福建林学院学报*, 1992, 12(4) :399 ~ 405.
- [3] 杨效文, 张孝羲. 我国不同地区烟蚜种群生殖特征研究. *生态学报*, 2000, 20(1) :139 ~ 144.
- [4] 任安芝, 高玉葆, 王金龙. 不同沙地生境下黄柳(*Salix gordejvii*) 的根系分布和冠层结构特征. *生态学报*, 2001, 21(3) :399 ~ 404.
- [5] 孟宪宇. 使用 Weibull 分布对人工油松林直径分布的研究. *北京林学院学报*, 1985, 7(1) :30 ~ 40.
- [6] 洪利兴, 杜国坚, 张庆荣. 天然常绿阔叶异龄幼林胸径的 Weibull 分布及动态预测. *植物生态学报*, 1995, 19(1) :29 ~ 42.
- [7] 邢黎峰, 梁玉堂, 丁修堂. Weibull 径级分布模型参数回收算法的 C 程序设计. *山东农业大学学报*, 1995, 26(4) :425 ~ 434.
- [8] 刘君然, 赵东方. 落叶松人工林威布尔分布参数与林分因子模型的研究. *林业科学*, 1997, 33(5) :412 ~ 417.
- [9] 吴承祯, 洪伟. 杉木人工林胸径的 Weibull 分布及其最优拟合研究. *江西农业大学学报*, 1998, 20(1) :86 ~ 90.
- [10] 余建英, 何旭宏编著. 数据统计分析与 SPSS 应用. 北京: 人民邮电出版社, 2003.
- [11] 盛骤, 谢式千, 潘承毅 编. 概率论与数理统计. 北京: 高等教育出版社, 2003.
- [12] 王新洲著. 非线性模型参数估计理论与应用. 武汉: 武汉大学出版社, 2002.
- [14] 邓建, 古德生, 李夕兵. 确定可靠性分析 Weibull 分布参数的概率加权矩法. *计算力学学报*, 2004, 21(5) :609 ~ 613.
- [16] 韦博成, 著. 近代非线性回归分析. 南京: 东南大学出版社, 1989.

附录

设非线性模型为:

$$Y = f(X, \theta) + \varepsilon \quad (1)$$

式中: Y 为 $n \times 1$ 的观测向量, $f(X, \theta) = (f_1(X, \theta), f_2(X, \theta), \dots, f_n(X, \theta))$ 是由 n 个 θ 的非线性函数组成的向量, ε 为 $n \times 1$ 的随机误差向量, X 为相应指标向量, θ 为 $t \times 1$ 的参数向量 ($t < n$); $f_i(X, \theta)$ 为 θ 的非线性函数, 假设模型 (1) 关于 θ 存在二阶以上连续导数, 则模型 (1) 线性化后的误差方程为: $V = B\alpha + f(X, \theta) - Y$, 式中,

$$B = \begin{pmatrix} \frac{\partial f_1(X, \theta)}{\partial \theta_1} & \frac{\partial f_1(X, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_1(X, \theta)}{\partial \theta_t} \\ \frac{\partial f_2(X, \theta)}{\partial \theta_1} & \frac{\partial f_2(X, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_2(X, \theta)}{\partial \theta_t} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_n(X, \theta)}{\partial \theta_1} & \frac{\partial f_n(X, \theta)}{\partial \theta_2} & \dots & \frac{\partial f_n(X, \theta)}{\partial \theta_t} \end{pmatrix}$$

$$\alpha = (\alpha_1, \alpha_2, \dots, \alpha_t)'$$

$f(X, \theta)$ 在 θ_0 处泰勒展开为:

$$f(X, \theta) = f(X, \theta_0) + B\alpha + \frac{1}{2} \alpha' C \alpha + \varepsilon \quad (2)$$

式中: ε 为略去三次及三次以上各项所产生的误差向量, $C = (C_{kij})$,

$$C_k = \begin{pmatrix} \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1^2} & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1 \partial \theta_t} \\ \frac{\partial^2 f_k(X, \theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_2^2} & \dots & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_2 \partial \theta_t} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t \partial \theta_1} & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t \partial \theta_2} & \dots & \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t^2} \end{pmatrix}$$

是 C 的第 k 层矩阵 (关于三维数组的详细内容见参考文献^[16])。

而三维数组的运算比较复杂容易出错, 因此在实际计算时有必要进行简化, 设法避免三维数组的运算, 为此将 $f(X, \theta)$ 对 θ 的二阶偏导数重新排列, 而得一个 $n \times a$ 的矩阵 W , $a = \frac{1}{2} t(t+1)$ 与一个 a 的向量 β

$$\beta = \beta' = (\beta_1^2, \beta_2^2, \dots, \beta_t^2, \beta_1 \beta_2, \dots, \beta_1 \beta_t, \beta_2 \beta_3, \dots, \beta_{t-1} \beta_t)'$$

$$W = \begin{bmatrix} \frac{\partial^2 f_1(X, \theta)}{\partial \theta_1^2} & \dots & \frac{\partial^2 f_1(X, \theta)}{\partial \theta_t^2} & 2 \frac{\partial^2 f_1(X, \theta)}{\partial \theta_1 \partial \theta_2} & \dots & 2 \frac{\partial^2 f_1(X, \theta)}{\partial \theta_1 \partial \theta_t} & 2 \frac{\partial^2 f_1(X, \theta)}{\partial \theta_2 \partial \theta_3} & \dots & 2 \frac{\partial^2 f_1(X, \theta)}{\partial \theta_{t-1} \partial \theta_t} \\ \frac{\partial^2 f_2(X, \theta)}{\partial \theta_1^2} & \dots & \frac{\partial^2 f_2(X, \theta)}{\partial \theta_t^2} & 2 \frac{\partial^2 f_2(X, \theta)}{\partial \theta_1 \partial \theta_2} & \dots & 2 \frac{\partial^2 f_2(X, \theta)}{\partial \theta_1 \partial \theta_t} & 2 \frac{\partial^2 f_2(X, \theta)}{\partial \theta_2 \partial \theta_3} & \dots & 2 \frac{\partial^2 f_2(X, \theta)}{\partial \theta_{t-1} \partial \theta_t} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial^2 f_n(X, \theta)}{\partial \theta_1^2} & \dots & \frac{\partial^2 f_n(X, \theta)}{\partial \theta_t^2} & 2 \frac{\partial^2 f_n(X, \theta)}{\partial \theta_1 \partial \theta_2} & \dots & 2 \frac{\partial^2 f_n(X, \theta)}{\partial \theta_1 \partial \theta_t} & 2 \frac{\partial^2 f_n(X, \theta)}{\partial \theta_2 \partial \theta_3} & \dots & 2 \frac{\partial^2 f_n(X, \theta)}{\partial \theta_{t-1} \partial \theta_t} \end{bmatrix},$$

于是下列等式成立

$$\alpha' C \alpha = W \beta \quad (3)$$

证明: 设 $\mu = \alpha' C \alpha$, $v = W \beta$

则有

$$\mu = \left(\sum_{i=1}^t \sum_{j=1}^t C_{1ij} \theta_i \theta_j \quad \sum_{i=1}^t \sum_{j=1}^t C_{2ij} \theta_i \theta_j \quad \dots \quad \sum_{i=1}^t \sum_{j=1}^t C_{nij} \theta_i \theta_j \right)',$$

则

$$\mu_k = \sum_{i=1}^t \sum_{j=1}^t C_{kij} \theta_i \theta_j$$

再由三维数组 C 的定义得

$$\mu_k = \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1^2} \theta_1^2 + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1 \partial \theta_2} \theta_1 \theta_2 + \dots + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_1 \partial \theta_t} \theta_1 \theta_t + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_2 \partial \theta_1} \theta_2 \theta_1 + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_2^2} \theta_2^2 + \dots$$

$$\frac{\partial^2 f_k(X, \theta)}{\partial \theta_2 \partial \theta_t} \theta_2 \theta_t + \dots + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t \partial \theta_1} \theta_t \theta_1 + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t \partial \theta_2} \theta_t \theta_2 + \dots + \frac{\partial^2 f_k(X, \theta)}{\partial \theta_t^2} \theta_t^2$$

因为 $\frac{\partial^2 f_k(X, \theta)}{\partial \theta_i \partial \theta_j} \theta_i \theta_j = \frac{\partial^2 f_k(X, \theta)}{\partial \theta_j \partial \theta_i} \theta_j \theta_i$, 所以

$$\mu_k = \sum_{i=1}^t \frac{\partial^2 f_k(X, \theta)}{\partial \theta_i^2} \theta_i^2 + 2 \sum_{i=1}^{t-1} \sum_{j=i+1}^t \frac{\partial^2 f_k(X, \theta)}{\partial \theta_i \partial \theta_j} \theta_i \theta_j$$

再根据矩阵 $\mathbf{d}\beta$ 和 W 有 $v_k = \sum_{i=1}^t \frac{\partial^2 f_k(X, \theta)}{\partial \theta_i^2} \theta_i^2 + 2 \sum_{i=1}^{t-1} \sum_{j=i+1}^t \frac{\partial^2 f_k(X, \theta)}{\partial \theta_i \partial \theta_j} \theta_i \theta_j = \mu_k$

所以(3)式成立。将(3)式代入(2)式得

$$f(X, \theta) = f(X, \theta_0) + B\mathbf{d}\theta + \frac{1}{2} W\mathbf{d}\beta + \varepsilon \quad (4)$$

当 $f(X, \theta)$ 为线性模型时, 则有: $W\mathbf{d}\beta = \varepsilon = \Theta$, Θ 这里是 $n \times 1$ 的各元素为 0 的矩阵。

对于非线性模型有:

$$f(X, \theta) - f(X, \theta_0) - B\mathbf{d}\theta - \frac{1}{2} W\mathbf{d}\beta - \frac{1}{2} \varepsilon > 0 \quad (5)$$

显然 $W\mathbf{d}\beta - \frac{1}{2} \varepsilon$ 越大, 模型的非线性强度越强, $W\mathbf{d}\beta - \frac{1}{2} \varepsilon$ 越小, 模型的非线性强度越弱, 故可用 $W\mathbf{d}\beta - \frac{1}{2} \varepsilon$ 来判断模型的非线性强度。

根据(5)式有 $-V - \frac{1}{2} \varepsilon \geq W\mathbf{d}\beta - \frac{1}{2} \varepsilon$, 即 $V \geq W\mathbf{d}\beta - \frac{1}{2} \varepsilon$ (6)

将矩阵 $W\mathbf{d}\beta$ 投影到解迹线在 $\theta = \theta_0$ 处的切空间和法空间, 则有

$$W\mathbf{d}\beta = W^N \mathbf{d}\beta + W^T \mathbf{d}\beta \quad (7)$$

式中, $W^N = S_N W$, $W^T = S_T W$, S_T 为 B 生成的投影矩阵: $S_T = (B B^T)^{-1} B^T$, $S_N = I - S_T$, 下面对(7)式进行证明。

证明: $W^N \mathbf{d}\beta + W^T \mathbf{d}\beta = (W^N + W^T) \mathbf{d}\beta = (S_N + S_T) W\mathbf{d}\beta = W\mathbf{d}\beta$, 故(7)式成立。

由于 $W\mathbf{d}\beta - \frac{1}{2} \varepsilon \geq W^N \mathbf{d}\beta - \frac{1}{2} \varepsilon$, $W\mathbf{d}\beta - \frac{1}{2} \varepsilon \geq W^T \mathbf{d}\beta - \frac{1}{2} \varepsilon$, 将它代入(6)式得:

$$\beta^N = \frac{W^N \mathbf{d}\beta - \frac{1}{2} \varepsilon}{V - \frac{1}{2} \varepsilon} < 1, \quad \beta^T = \frac{W^T \mathbf{d}\beta - \frac{1}{2} \varepsilon}{V - \frac{1}{2} \varepsilon} < 1 \quad (8)$$