浙 江 林 业 科 技，2020，40（2）：71－76
J Zhejiang For Sci Technol

doi:10.3969/j.issn.1001-3776.2020.02.011

马尾松地径与胸径关系模型的比选研究

李　辉

（湖南省东安县林业局，湖南　东安　425900）

摘要：为了研究马尾松 *Pinus massoniana* 地径与胸径的相关性，于 2019 年 12 月，以东安县马尾松 4 个近熟林样地的地径与胸径数据，先用 SPSS 曲线回归分析中 11 种数学模型进行拟合初判，初步选择相关系数 R^2 比较高的线性方程、二次曲线、三次曲线、幂函数 4 种函数模型，然后对这四种函数模型方程进行 SPSS 非线性回归分析。结果表明，采用高斯-牛顿迭代法推算的幂函数曲线方程 $y = 0.599\ 484x^{1.078\ 469}$ 为最优拟合模型，其调整相关系数（$\overline{R}^2 = 0.960\ 808$）最大，拟合度最好，精度相对较高。采用高斯-牛顿迭代法推算出的幂函数模型方程要优于采用曲线直线化法的原幂函数模型方程的拟合效果。经模型适用性检验，总相对差（*RS*）=－1.34%，－5% < *RS* < 5%，绝对平均百分比误差（*MAPE*）=7.6% < 10%，均在合理范围内。

关键词：马尾松；地径；胸径；幂函数模型；拟合

中图分类号：S758.62　　**文献标识码**：A　　**文章编号**：1001-3776（2020）02-0071-06

Comparison on Correlation Models for Ground Diameter and *DBH* of *Pinus massoniana*

LI　Hui

（Dong’an Forestry Bureau of Hunan, Dong’an 425900, China）

Abstract: In December 2019, 4 sample plots were established in near-maturing *Pinus massoniana* forest in Dong’an county, Hunan province for determination of ground diameter and DBH. 11 mathematical models in SPSS curve regression analysis were fitting to make compare. The result demonstrated that 4 models had higher R^2, like conic, cubic curve, power function. Nonlinear regression analysis of SPSS on these models showed that the power function curve equation by Gauss-Newton method was the optimal fitting model, $y = 0.599\ 484x^{1.078\ 469}$, with adjustment of the correlation coefficient ($\overline{R}^2$= 0.960808) the largest. Fitness test of the model indicated that the total relative difference was -1.34%, between -5% to 5%, with absolute mean percentage error of 7.6%.

Key words: *Pinus massoniana*; ground diameter; DBH; power function model; fitting

马尾松 *Pinus massoniana* 是我国南方重要的乡土树种，适应能力强，面积广。目前，滥伐马尾松林案件时有发生，由于林木采伐后原木被拉走，能测定的因子只有地径，且现在采伐林木都用电锯，锯口离地面都不会超过 5 cm，一般离地面 2 cm。目前，多数地区没有本地的地径材积表，只能通过地径推算采伐木胸径，然后通过二元材积表查出采伐木蓄积。为此，许多学者根据不同地区马尾松地径与胸径数据进行了回归方程拟合分析研究，分别认为拟合最好的是一元线性方程模型[1-3]、二项式方程模型[4-6]、三项式方程模型[7]、幂函数方程模型[8-10]。本研究根据东安县马尾松地径与胸径调查数据，对其进行了相关性研究和分析，为有关采伐马尾松林的

收稿日期：2019-11-11；修回日期：2020-02-20

作者简介：李辉，林业工程师，从事森林资源保护与调查规划设计；E-mail:76721777@qq.com。

林业案件技术鉴定及编制本地区马尾松地径材积表提供参考。

1 研究材料与方法

1.1 研究地概况

东安县位于湖南省西南边陲，湘江上游，地处南岭山系向湘中丘陵过渡地带，地理坐标范围在110°34′10″～110°59′33″E，26°07′04″～26°52′29″N之间。年均气温为17.6℃，年均降水量为1 338 mm，年均相对湿度为80%，平均风力在2～3级，年均无霜期为298 d。林业用地面积达144 606.2 hm^2，占土地总面积的65.54%，森林覆盖率为58.1%，林木绿化率为63.0%，马尾松林面积达19 454.3 hm^2，占有林地面积的25.49%。属于中亚热带常绿阔叶林区，森林植物资源丰富，典型地带性植被为常绿阔叶林，兼有以杉木 *Cunninghamia lanceolata*，马尾松为主要成分的针叶林、竹林和经济林。

1.2 材料来源

2019年12月，在东安县不同地点的马尾松林内随机设立4块调查样地。样地距离林缘10 m以上，大小为20 m×25 m。4块样地均为天然的近熟林，土壤类型为红壤，调查样地基本情况见表1。调查方法为用围尺绕测样地内马尾松的带皮地径及对应的带皮胸径，地径为距上坡地面2 cm处的树木直径，胸径为距上坡地面1.3 m处的树木直径，测量样地内所有样株。本研究共测量马尾松207株，测量所得数据地径在7.1 ~ 48.2 cm，胸径在5.0～38.0 cm，马尾松样木株数、胸径径阶分布情况见表2，其中样地1～3共计177株用于拟合建模，样地4共计30株用于验模。

表1 调查样地基本情况
Table 1 Information of sample plots

样地号	地点	坡度	坡向	坡位	海拔/m	土层厚度/cm
1	白牙市镇柳溪村	平坡	西坡	中坡	170	85
2	紫溪市镇五里牌村	缓坡	南坡	中坡	175	60
3	白牙市镇杳脚村	缓坡	南坡	上坡	205	60
4	井头圩镇井头圩村	平坡	东坡	下坡	180	80

表2 马尾松样木株数及其胸径径阶分布情况
Table 2 Distribution of diameter class of sample trees

径阶/cm	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34	36	38	合计
建模株数/株	11	18	17	18	16	21	11	12	7	14	10	4	7	6	3	0	2	177
验模株数/株	1	3	3	3	1	5	2	2	2	3	1	3	2					30

1.3 研究方法

1.3.1 异常数据处理 将207株马尾松的地径及对应的胸径数据输入Excel表格中，并以地径为x坐标，胸径为y坐标显示散点图，观察是否有样本偏离其它绝大多数样本。如没有，说明该样本无异常数据，样本有效。如有，剔除该异常样本，剩余的样本作研究样本。

1.3.2 模型方程初选 采用SPSS曲线回归分析，通过选取线性、二次、三次、指数、幂、对数、S（S形曲线）、增长、复合、倒数、Logistic（逻辑斯蒂）共11个函数模型，分别拟合马尾松地径与胸径的相关关系进行初选，选出几个相关系数R^2高的模型。采用SPSS软件曲线估计回归分析中曲线模型是用曲线函数转换成直线方程后的新数据进行拟合的，即曲线直线化[11]拟合，得到的相关系数R^2为转换成的直线方程相关系数，非原曲线函数的相关系数。此模型的最佳参数应用高斯-牛顿迭代法求出[12-13]。

1.3.3 模型方程选取

1.3.3.1 模型方程选取 SPSS非线性回归分析是采用高斯-牛顿迭代法进行曲线函数参数估计的，选定模型按此分析进行拟合，并以模型方程最后一次线性化的F统计量判断模型参数的显著性。按模型评价指标选出拟合度

最好和精度最高的模型方程作为马尾松地径与胸径最优拟合模型。

1.3.3.2 模型评价指标 相关系数 R^2 衡量的是回归方程整体的拟合度，但由于 R^2 有当变量数目增加而增大的缺点，特别是多项式，当添加更多项时，会增加决定系数 R^2，虽可获得更接近数据的拟合，但代价是模型更为复杂，R^2 无法解释，而调整后的相关系数（$\overline{R}^2$）包括了一项对模型中项数的惩罚，是扣除了回归方程中所受到的包含项数影响的相关系数，更适合比较不同的模型对同一数据的拟合程度，可更准确地反映模型的好坏，因此以 $\overline{R}^2$ 作拟合优度评价。马尾松地径与胸径关系模型主要评价指标有：$\overline{R}^2$、均方根误差（$RMSE$）、平均绝对误差（MAE）、绝对平均百分比误差（$MAPE$）4 个指标，其相应公式如下：

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

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \times \frac{n-1}{n-p-1} \tag{2}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - \hat{y}_i)^2}{n}} \tag{3}$$

$$MAE = \frac{1}{n}\sum_{n=i}^{n}\left|y_i - \hat{y}_i\right| \tag{4}$$

$$MAPE = \frac{1}{n}\sum_{n=i}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\% \tag{5}$$

上述各式中，$\overline{y}_i$ 为第 i 株样木的胸径实测值；$\hat{y}_i$ 为整个样木的胸径平均值；$\hat{y}_i$ 为第 i 株样木的胸径估计值；n 为样本数；p 为解释变量个数或多项式的次数。

$\overline{R}^2$ 表示方程的拟合效果，$\overline{R}^2$ 越大，方程拟合效果越好。$RMSE$ 是用来衡量观测值同真实值之间的偏差，其值越小越好。MAE 更能反映预测值误差的实际情况，其值越少，模型精确度越高。$MAPE$ 表示模型精度的大小，是衡量模型精度高低的常用指标，排除了样本个体间正负误差的相互抵消，可反映胸径估计的误差水平，其值越小，胸径估计的误差越小，模型精度越高。

1.3.3.3 模型的适应性检验 检验数据为独立采集的检验样方样本，当检验样方样本 $-5\% <$ 总相对差(RS) $< 5\%$，$MAPE<10\%$，同时通过 F 检验（$F \leqslant F_{0.05}$），则所建立并选择的地径与胸径数学模型适用。否则，应重新建模或者选择其它模型检验[15]。相关公式如下：

$$RS = \sum_{i=1}^{n}(y_i - \hat{y}_i)\bigg/\sum_{i=1}^{n}\hat{y}_i \times 100\% \tag{6}$$

式中，$\hat{y}_i$ 为第 i 株样木的胸径估计值；y_i 为第 i 株样木的胸径实测值。

$$F = \frac{\frac{1}{2}\left[\left(a\sum_{i=1}^{n}y_i + b\sum_{i=1}^{n}x_i y_i\right) - \left(2\sum_{i=1}^{n}x_i y_i - \sum_{i=1}^{n}x_i^2\right)\right]}{\frac{1}{n-1}\left[\sum_{i=1}^{n}y_i^2 - \left(a\sum_{i=1}^{n}y_i + b\sum_{i=1}^{n}x_i y_i\right)\right]} \tag{7}$$

式中，a，b 为回归直线的参数；n 为样木数；x_i 为第 i 株样木的胸径估计值；y_i 为第 i 株样木的胸径实测值。

2 结果与分析

2.1 异常数据处理

将 207 株马尾松的地径及对应的胸径数据输入 Excel 表格中，显示散点图，没有发现样本数据显著偏离其

它样本，说明该批样本无异常数据，建模样本和验模样本都有效。

2.2 模型方程初选

在SPSS软件中，以地径为自变量x，胸径为因变量y，将测量用于建模的177株马尾松地径及对应的胸径数据录入SPSS“数据视图”中，在11种函数模型方程进行拟合，结果见表3。线性（$R^2=0.960$）、二次（$R^2=0.961$）、三次（$R^2=0.960$）、幂（$R^2=0.969$）4种函数模型的R^2都较高，说明这4种函数模型是极度正相关，拟合度都较好。表3中除线性模型外，其它模型R^2都为变换成直线方程的相关系数。在原始数据相关系数R^2的计算中，必须用未经变量代换的原始数据计算[16]。经各模型方程变换化为直线得到的参数代入原方程后推算，线性模型$R^2=0.960\ 466$、二次原模型方程$R^2=0.961\ 109$、三次原模型方程$R^2=0.961\ 162$、幂原模型方程$R^2=0.960\ 788$。经过曲线估计回归分析拟合，初选出了4个拟合较好的模型方程，公式如下：

线性方程：

$$y=\mathrm{a}+\mathrm{b}_1x \qquad (8)$$

二次曲线方程：

$$y=\mathrm{a}+\mathrm{b}_1x+\mathrm{b}_2x^2 \qquad (9)$$

三次曲线方程：

$$y=\mathrm{a}+\mathrm{b}_1x+\mathrm{b}_2x^2+\mathrm{b}_3x^3 \qquad (10)$$

幂函数方程：

$$y=\mathrm{a}x^{\mathrm{b}1} \qquad (11)$$

上述公式中，a，b_1，b_2，b_3均为方程参数。

表3 初选模型数据和参数估计值

Table 3 Primary model data and parameter estimate

方程	模型数据					参数估计值			
	R^2	F	df_1	df_2	Sig.	a	b_1	b_2	b_3
线性	0.96	4 251.568	1	175	0	－1.199	0.824		
对数	0.904	1 651.204	1	175	0	－33.11	16.684		
倒数	0.769	581.719	1	175	0	31.615	－267.30		
二次	0.961	2 149.995	2	174	0	－0.127	0.719	0.002	
三次	0.961	1 427.117	3	173	0	0.542	0.619	0.007	－5.76E-05
复合	0.927	2 213.075	1	175	0	5.062	1.051		
幂	0.969	5 505.792	1	175	0	0.62	1.067		
S	0.918	1 967.177	1	175	0	3.713	－18.046		
增长	0.927	2 213.075	1	175	0	1.622	0.05		
指数	0.927	2 213.075	1	175	0	5.062	0.05		
Logistic	0.927	2 213.075	1	175	0	0.198	0.951		

注：因变量为胸径（cm），自变量为地径（cm）。

2.3 模型方程选取

非线性回归分析采用高斯-牛顿迭代法进行曲线函数参数的估计，这种估计方法是对曲线函数作泰勒级数展开，以达到线性近似的目的，并反复迭代求解。用SPSS软件中“分析－回归－非线性”分别拟合初选方程（8）、（9）、（10）、（11），用模型初选的参数估计值作模型非线性回归分析的参数初值，分析得各模型参数，并用参数推算出模型的评价指标值，结果见表4。对于非线性回归模型的回归系数和回归方程显著性检验是对最后一次线性近似作F检验，这种检验也是合理的。由表4可知，4个模型Sig.值均为0，小于0.05，都通过了0.05水平检验，具有显著性，都有统计学意义。

由表4可知，各模型方程$\bar{R}^2$的大小顺序为，幂函数（$\bar{R}^2$=0.0.960 808）> 二次曲线（$\bar{R}^2$=0.960 662）> 三次

曲线（$\bar{R}^2$= 0.960 488）> 线性（$\bar{R}^2$= 0.960 240），说明幂函数拟合优度最好。另对比 SPSS 中用曲线回归分析的幂原模型方程 R^2 = 0.960 788，小于用非线性回归分析的幂原模型方程 R^2 = 0.961 03，表明用非线性回归分析模型的拟合效果要优于用曲线回归分析的模型。

表 4 马尾松地径与胸径相关关系模型方程参数和评价指标值

Table 4 Model equation parameters and evaluation index value of correlation between ground diameter and DBH of *P. massoniana*

模型方程	a	b_1	b_2	b_3	R^2	$\bar{R}^2$	*RMSE*	*MAE*	*MAPE*	*F*	Sig.
y=a+b_1x	−1.198 899	0.824 416			0.960 466	0.960 240	1.525 664	1.110 774	0.066 926	4 251.57	0
y=a+b_1x+b_2x^2	−0.126 942	0.719 485 6	0.002 18		0.961 109	0.960 662	1.513 216	1.117 052	0.066 760	2 149.99	0
y=a+b_1x+b_2x^2+b_3x^3	0.541 517	0.619 351	0.006 56	−0.000 058	0.961 162	0.960 488	1.512 184	1.120 674	0.067 348	1 427.12	0
y=ax^b	0.599 484	1.078 469			0.961 030	0.960 808	1.514 739	1.112 453	0.066 528	4 315.66	0

各模型方程指标 *RMSE* 的大小顺序为，三次曲线（*RMSE* = 1.512 184）< 二次曲线（*RMSE* = 1.513 216）< 幂函数（*RMSE* = 1.514 739）< 线性（*RMSE* = 1.525 664），说明三次曲线的 *RMSE* 最小，偏差少，幂函数 *RMSE* 与其相差细微。指标 *MAE* 的大小顺序为，线性（*MAE*=1.110 774）< 幂函数（*MAE*=1.112 453）< 二次曲线（*MAE*=1.117 052）< 三次曲线（*MAE* =1.120 674），说明线性的 *MAE* 最小，精确度高，幂函数的精确度也比较高。4 个模型的 *MAPE* 均小于 10%，表明根据地径都可以利用这些模型精确地估计林木的胸径，幂函数（*MAPE* = 6.652 8%）< 二次曲线（*MAPE*=6.676%）< 线性（*MAPE*=6.692 6%）< 三次曲线（*MAPE* = 6.731 8%），幂函数的 *MAPE* 最小，说明幂函数方程对胸径估计的误差最小，估计精度最高。综合 4 个模型各评价指标值，各模型指标值差异不是很明显，其中幂函数模型方程的拟合优度最好，精度相对较高，东安县马尾松地径与胸径关系的最优数学模型公式为：

$$y = 0.599\,484x^{1.078\,469}$$

2.4 模型适用性检验

用测量验模 4 号样地的 30 株样木作最优数学模型的适用性检验，以地径为 x，用最优模型 $y = 0.599\,484x^{1.078\,469}$ 求出对应的检验样本胸径估计值。

经计算求得，$RS = -1.34\%$，$-5\% < RS < 5\%$，表明 RS 在合理范围内；$MAPE = 7.6\%$，$MAPE < 10\%$，表明 $MAPE$ 在合理范围内。

用检验样本胸径估计值 $\hat{y}_i$ 为 x，检验样本胸径实际值为 y 作回归方程，得：

$$y = -0.211\,594 + 0.998\,443x$$

在 95%的可靠性下对回归参数与理想直线参数 $a = 0$，$b = 1$ 进行 F 检验。经计算求得，$F = 0.243\,6$，按自由度 $df_1 = 2$，$df_2 = 29$，查 F 分布表，单侧界限值为 $F_{0.05} = 3.33$。很明显，$F = 0.243\,6 < F_{0.05} = 3.33$。

以上适用性检验，说明该模型实用性是很强的。

3 结论与讨论

（1）利用马尾松地径与胸径调查数据，采用 SPSS 曲线回归分析中 11 种数学模型进行拟合初判，其中，线性（R^2 = 0.960 466）、二次（R^2 = 0.961 109）、三次（R^2 = 0.961 162）、幂（R^2 = 0.960 788）4 种函数模型的 R^2 都较高，说明这 4 种函数模型拟合度都较好，但由于曲线回归分析是采用曲线直线化法对变换后的数据进行的，所得结果仅对变换后的数据来说是最佳拟合，当再变换回原数据坐标时，所得的回归曲线并不一定是最佳拟合。因此，我们用 SPSS 非线性回归分析，采用高斯-牛顿迭代法对初选的 4 种函数模型进行分析，其中，幂函数曲线方程 $y = 0.599\,484x^{1.078\,469}$ 为最优拟合模型，其调整相关系数（$\bar{R}^2$ = 0.960 808）最大，拟合度最好，精度相对较高。该结论与杨晓毅、秦鹏飞、刘健等[8-10]研究选定的最优模型方程一致，但选定的幂函数模型参数不

同，且所用研究方法和拟合优度评价指标不一样。与赵浩彦、卢昌泰、黎良财等[1-7]研究选定的最优模型则完全不同，这可能是地径所测量的离上坡地面距离不同引起的，更可能是研究地区不同，所在气候类型和立地条件不同，导致马尾松生长出现差异，从而使得不同地区马尾松地径与胸径关系模型的拟合效果出现差异[1]。用曲线回归分析的幂原模型方程 $R^2 = 0.960\ 788$，小于用非线性回归分析的幂原模型方程 $R^2 = 0.961\ 03$，说明采用高斯-牛顿迭代法推算出的幂函数模型方程要优于采用曲线直线化的原幂函数模型方程的拟合效果。地径与胸径的最优拟合模型经检验精度较高，具有较强的适用性。

（2）利用马尾松最优拟合模型方程 $y = 0.599\ 484x^{1.078\ 469}$，通过测量马尾松地径就可以推算出其胸径，再化成径阶数，用立木材积表查算出被伐木的材积，为林政资源管理及林业执法提供理论依据。

参考文献：

[1] 赵浩彦，张民侠，张洁，等. 南京马尾松根径与胸径多元混合效应模型研建[J]. 中南林业科技大学学报，2015，35（11）：43－48.
[2] 王华. 黔南地区马尾松地径材积式的建模与应用[J]. 贵州林业科技，2010，38（4）：4－7.
[3] 张宜香，唐初明. 以简化回归技术创建荔浦县伐区马尾松地径材积表[J]. 林业勘查设计，2011（3）：45－48.
[4] 卢昌泰，李吉跃，康强，等. 马尾松胸径与根径和冠径的关系研究[J]. 北京林业大学学报，2008，30（1）：58－63.
[5] 李宝银，朱德培，江正铨，等. 沙县杉木、马尾松、阔叶树地径一元材积表编制的研究[J]. 福建林业科技，1993，20（3）：24－28.
[6] 张金全. 建瓯市杉木与马尾松及阔叶树地径材积表的编制[J]. 现代农业科技，2012（19）：155－159，167.
[7] 黎良财，邓利. 柳州市马尾松地径一元材积表的编制[J]. 林业调查规划，2011，36（7）：1－3.
[8] 杨晓毅，赵浩彦. 马尾松地径与胸径相关关系研究[J]. 湖南林业科技，2014，41（5）：40－42.
[9] 秦鹏飞，罗文盛，陈旭. 宣汉县马尾松地径与胸径模型研究[J]. 四川林勘设计，2013（2）：44－46.
[10] 刘健，陈平留，陈昌雄. 闽北主要用材树种胸径与去皮地径关系的研究[J]. 福建林学院学报，1996，16（1）：45－48.
[11] 谢兰，高东红. 非线性回归方法的应用与比较[J]. 数学的实践与认识，2009，39（10）：117－121.
[12] 黄敬锋，王秀珍. 求解非线性回归方程及其统计检验方法初探[J]. 新疆气象，1990，13（10）：19－22.
[13] 韩兆洲，汪建华. 线性化最小二乘法的理论分析[J]. 统计与决策，2009（10）：29－30.
[14] 宋延山. 相关系数统计量的功能及其应用探讨——以 SPSS 为分析工具[J]. 统计教育，2008（11）：28－31.
[15] 国家林业局. 根径立木材积表编制技术规程：LY/T2103－2013[S]. 2013. http://www.doc88.com/p-5035001054007.html
[16] 王绍宗，郭光平，李恩来，等. 数理统计[M]. 北京：中国林业出版社，1992：200.
[17] 娄冬华. 曲线直线化与非线性回归[J]. 疾病控制杂志，1999，3（3）：224－225.