

# 短叶松同龄纯林林分平均直径的时序分析

范小莉<sup>1</sup>, 杨 华<sup>1\*</sup>, 亢新刚<sup>1</sup>, 高 延<sup>2</sup>, 冯启祥<sup>2</sup>

(1. 北京林业大学 林学院, 北京 100083; 2. 汪清县林业局, 吉林 汪清 133200)

**摘 要:**根据 Box-Jenkins 建模原理, 将 ARIMA 模型拟合时间序列的方法用于同龄纯林林分平均直径生长的拟合及预测研究。利用 Eviews 软件分别对短叶松同龄纯林样地 1 号、2 号、3 号的林分平均直径序列建模和预测, 选出最优模型形式。然后用样地 4 的林分平均直径序列对所得到的模型形式进行适应性验证。通过用这 4 块样地数据对最优模型形式的拟合和预测, 结果显示, 最优模型形式剔除 AR(2)项的 AR(3)模型能够较好的拟合平均直径序列, 决定系数( $R^2$ )分别为 0.907 5、0.909 7、0.910 3、0.949 8, 并且预测相对误差(Bias)很小( $<1\%$ )。在直径生长序列的近期预测中应用时间序列的 ARIMA 模型方法较为合适。

**关键词:**短叶松; 平均直径; 时间序列; ARIMA 模型

**中图分类号:**S758.55      **文献标志码:**A      **文章编号:**1001-7461(2011)04-0162-05

## Time Series Analysis of Mean Jack Pine Even-aged Pure Stand Diameter

FAN Xiao-li, YANG Hua<sup>1\*</sup>, KANG Xin-gang, GAO Yan, FENG Qi-xiang

(1. College of Forestry, Beijing Forestry University, Beijing 100083, China; 2. Wangqing Forestry Bureau, Wangqing, Jilin 133200, China)

**Abstract:** According to Box-Jenkins principle and ARIMA model, the time series of the average diameter of Jack pine even-aged pure stand was fitted and predicted. A model was constructed and predicted for the data of sample plot 1, 2, 3 by Eviews software to select out the form of the optimal model. The data from plot 4 were used for model feasibility verifying. After fitting and prediction for four sample plots, the results showed that the AR(3) model rejecting AR(2) term better fit the series of the average diameter,  $R_1^2 = 0.907\ 5$ ,  $R_2^2 = 0.909\ 7$ ,  $R_3^2 = 0.910\ 3$ ,  $R_4^2 = 0.949\ 8$ , and bias was very small( $<1\%$ ). ARIMA model was suitable for short-term prediction of the series of diameter.

**Key words:** Jack pine; average diameter; time series; ARIMA model

时间序列分析已广泛应用于农、林业生产的生长、病虫害、土壤性质变化等领域。张荷观用时间序列的时域分析方法建立了单株木生长量预测模型, 较好地解决了由一株立木现时几年的总生长量值预测这棵树未来几年的总生长量值问题<sup>[1]</sup>。邱学清等用时间序列分析方法对黑荆树林分平均直径、平均高和蓄积量的月总生长量的趋势项和随机项进行了研究, 建立了这两项的叠合模型, 准确的描述并预测黑荆树的月生长情况<sup>[2]</sup>。陈广胜等采用时间序列分析法, 分析了人工林落叶松木材生长轮密度的变异

规律并建立了变异规律模型和预测模型, 为制定人工林木材培育措施时提供理论依据<sup>[3]</sup>。谭伟等以树木生长机理为依据, 与 ARMA 模型相结合, 采用 Kalman 滤波器的最小二乘估计方法, 建立了 AR-MAT 趋势预测模型, 应用该模型对单株树木生长过程描述和预报预估取得良好效果<sup>[4]</sup>。

林分直径生长模型的研究早在 20 世纪 80 年代就在我国兴起, 多年来, 对国内许多树种的林分直径生长均建立了模拟预测模型。周元满等以 Richards 生长函数为模型拟合了不同造林密度条件下桉树无

收稿日期: 2010-08-03    修回日期: 2010-12-27  
基金项目: 林业公益性行业专项课题(200804027)  
作者简介: 范小莉, 女, 硕士生, 主要研究林木生长过程。E-mail: lizhi-0527@163.com  
\* 通讯作者: 杨华, 女, 副教授, 硕士生导师。E-mail: huayang8747@163.com

性系林分直径生长模拟预测模型,利用所建模型可以对桉树无性系林分的生长动态进行预测<sup>[5]</sup>。杜纪山根据二类调查固定样地资料,采用合成分析法,建立了与距离无关的落叶松单木平方直径生长量模型,为二类调查森林资源的数据更新打下基础<sup>[6]</sup>。本文将用时间序列分析方法,拟用 ARIMA 模型分析短叶松同龄纯林林分平均直径生长变化趋势,并对未来几年的平均直径进行预测,为林分的经营管理提供有效的科学依据。

## 1 材料和方法

### 1.1 数据来源

研究样地为加拿大北方森林区的短叶松同龄纯林,数据资源引自姚晓红硕士论文<sup>[7]</sup>所收集的林分平均直径生长过程资料。选取 4 块短叶松同龄纯林样地资料,编号为样地 1 号到样地 4 号,对样地中每棵树在胸径处用生长锥钻取木芯,获得各株树的每年的直径生长量,然后对样地内所有树的直径做算术平均数,得到整个林分的平均直径生长序列,序列长度  $N$  表示林分从 1 年到  $N$  年的  $N$  个平均直径的个数,1 号样地  $N=48$ ,2 号样地  $N=60$ ,3 号样地  $N=60$ ,4 号样地  $N=60$ 。其中分别用 1 号、2 号、3 号样地前  $(N-2)$  年数据进行模型拟合,后两年数据进行验证,4 号样地进行模型的适应性验证。

### 1.2 方法

ARIMA 模型又称为 Box-Jenkins 模型,它由三个过程组成:自回归过程  $AR(p)$ 、单整过程  $I(d)$ 、移动平均过程  $MA(q)$ 。

- ①  $AR(p)$  模型形式为:
$$x_t = \theta_0 + \theta_1 x_{t-1} + \theta_2 x_{t-2} + \cdots + \theta_p x_{t-p} + \epsilon$$
- ②  $MA(q)$  是模型形式为:
$$x_t = \mu + \epsilon_1 - \alpha_1 \epsilon_{t-1} - \alpha_2 \epsilon_{t-2} - \cdots - \alpha_q \epsilon_{1-q}$$
- ③  $I(d)$  是指将一个非平稳时间序列转化为平稳序列所要进行差分的次数。

当  $p, d, q$  均不为 0, 此时的模型为完整的  $ARI-MA(p, d, q)$  模型,模型的公式为:

$$\theta(B)\Delta^d x_t = \alpha(B)\epsilon_t$$

模型的基本的思想是,对于非平稳的时间序列,用若干次的差分使其成为平稳序列,再进行拟合,将序列表示成关于序列的过去某些点的自回归和关于残差的移动平均的组合。

1.2.1 模型拟合 模型拟合之前,先用差分法或者中心化方法使数据平稳,然后用单位根检验(ADF)法检验其平稳性。

(1) 模型形式的选择 模型形式的选择通过相

关分析实现,有以下原则:

- ① 自相关系数拖尾,偏自相关系数  $p$  阶截尾,选用  $AR(p)$  模型;
- ② 自相关系数  $q$  阶截尾,偏自相关系数拖尾,选用  $MA(q)$  模型;
- ③ 自相关系数和偏自相关系数都为拖尾时,选用  $ARMA(p, q)$  模型。

由非零系数衰减为小值波动的过程非常迅速,视为截尾;由显著非零的系数衰减为小值波动的过程比较缓慢或者非常连续,视为拖尾。

(2) 模型的定阶 本文模型的定阶( $p, q$  取值)方法主要应用了偏相关函数法、AIC 和 BIC 准则法。

① 偏相关函数法 由于偏相关函数  $\varphi_{kk}$  是随机变量不可能具有严格的“截尾性”,只能呈现出在某步之后围绕零值上、下波动,因此,需要借助数理统计的方法来判断  $\varphi_{kk}$  的截尾性。对于每一个  $p > 0$  检查  $\varphi_{p+1, p+1}, \varphi_{p+2, p+2} \cdots \varphi_{p+m, p+m}$  中落入  $\varphi_{kk} \leq \frac{1}{\sqrt{N}}$  的比例是否占总数  $M(M = \sqrt{N})$  的 68.3%, 具体公式为:
$$P\left\{\varphi_{kk} \mid \leq \frac{1}{\sqrt{N}}\right\} = 0.683.$$
 即当  $p = p_0$  时,若满足上述条件,则判定  $\varphi_{kk}$  在  $p_0$  处截尾,此时的阶数  $p$  取值为  $p_0$ 。

② AIC 和 BIC 准则法 设定一个拟合模型的最高阶数为  $L$ , 则  $AR(k)$  模型 AIC 准则定阶公式为  $AIC(k) = \log \varphi_k^2 + \frac{2k}{N} (0 \leq k \leq L)$ , 在  $1 \leq k \leq L$  范围内,如果  $k = p$  时,  $AIC(k) = \min$ , 则适用的模型为  $AR(p)$ 。

BIC 准则函数的定义如下:  $BIC(p) = \log \varphi_k^2 + \frac{p \log N}{N}$  若某一阶数  $p_0$  满足  $BIC(p_0) = \min_{0 \leq p \leq L} BIC(p)$  则取  $p_0$  为模型的最佳阶数。

(3) 参数估计 应用了最小二乘法对模型参数进行估计。

#### 1.2.2 模型的验证

(1) 模型的检验 模型的检验主要有两部分即:参数的显著性检验和模型的显著性检验。

① 参数的显著性检验 就是要检验每一个未知参数是否显著为零,目的是为了模型最精简。本文采用  $T$  检验统计量来检验未知参数的显著性。

② 模型的显著性检验 一个显著有效的模型应该能够提取观察序列中几乎所有的相关信息,即拟合残差序列应该为白噪声序列。因此,模型的显著性检验就是对残差序列的白噪声检验。检验统计

量为  $LB$  统计量,当  $LB$  统计量的  $p$  值显著大于 0.05 时接受原假设,认为该拟合模型显著有效。

如果随机过程  $X_{(t)}(t=1,2,\cdots)$  是由一个不相关的随机变量序列构成,即对于所有  $s\neq t$ ,随机变量  $X_t$  和  $X_s$  的协方差均为零,即随机变量  $X_t$  和  $X_s$  互不相关,则称其为随机过程。对于一个纯随机过程来说,若其期望和方差均为常数,则称其为白噪声过程。白噪声过程的样本实现成为白噪声序列。

检验一个序列是否为白噪声序列的过程,称为白噪声检验。

(2)模型评价 拟合结果通过以下统计量和指标评价: $AIC$ 、 $BIC$  统计量、决定系数( $R^2$ )、相对误差( $Bias$ )。计算公式如下:

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}; Bias = \frac{|x_i - \hat{x}_i|}{x_i}$$

(3)模型的预测 预测原则是预测方差最小原则,即  $Var_{\hat{x}_i(l)}[e_i(l)]=\min\{Var[e_i(l)]\}$ 。  $AR(p)$  模型的  $l$  步预测式为:

$$\hat{x}_i(l) = \theta_1 \hat{x}_i(l-1) + \cdots \theta_p \hat{x}_i(1-p)^{[8-9]}$$

## 2 模型的构建

### 2.1 数据的平稳化

对 1 号样地短叶松 1~46 a 的平均直径序列数据( $x_t$ )作时序图(图 1),平均直径随时间的变化呈现增长趋势。

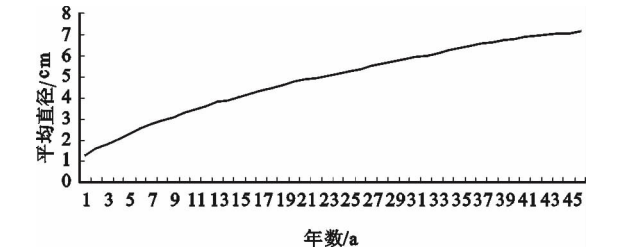


图 1 林分平均直径序列( $x_t$ )折线图

Fig. 1 Line of the stand average diameter series

因此,采用差分的方法消除增长趋势,经过一阶差分以后检验此时数据( $y_t$ )的平稳性,主要采用的是  $ADF$  检验。检验结果如表 1 所示:

表 1 林分平均直径一阶差分序列  $y_t$  的  $ADF$  检验结果

Table 1 ADF test of the stand average diameter first-order difference series		
	$t$ -统计量	概率
ADF 检验统计量	-2.640 6	0.009 4
检验临界值	$\alpha=0.01$ -2.616 2	
	$\alpha=0.05$ -1.948 1	
	$\alpha=0.10$ -1.612 3	

由表 1 可以看出, $p$  值为  $0.009\ 4<0.05$ ,序列通过了  $ADF$  检验,可见一阶差分生成的新序列是平稳的。

### 2.2 模型的识别和建立

为了确定使用  $AR(p)$  模型、 $MA(q)$  模型或  $ARMA(p,q)$  模型对平稳的时间序列进行估计,首先需根据差分后的平稳序列作自相关( $ACF$ )图和偏自相关( $PACF$ )图,结果如图 2。

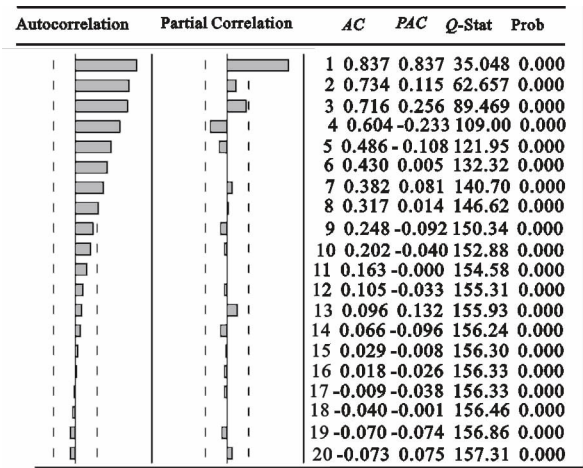


图 2  $y_t$  的  $ACF$  和  $PACF$  图

Fig. 2 Autocorrelation function and partial correlation function of  $y_t$  (偏自相关  $AC$ -自相关系数,  $PAC$ -偏自相关系数,  $Q$ -Stat 统计量,  $Prob$ -伴随概率,下同。)

通过图 2 检验出此时的序列为非白噪声序列 ( $p<0.05$ ),并且,  $ACF$  图显示自相关系数缓慢的接近 0,呈现明显的拖尾现象,而  $PACF$  图显示偏自相关系数在  $k=1$  后迅速的落入置信区间内,呈现明显的截尾现象,由此,根据模型识别的规则,在  $ACF$  图拖尾,  $PACF$  图截尾的情况下,应采用  $AR(p)$  模型,其中  $p$  为待定系数。根据模型的定阶方法,确定阶数  $p=3$ ,然后采用对  $p$  从低阶到高阶逐个进行试验,根据模型的  $R^2$ 、 $AIC$  和  $BIC$  值、以及系数是否具有统计学意义进行综合判断。

综合上述判断标准,初步判定选用  $AR(3)$  模型。利用最小二乘法对备选模型进行参数估计,结果如表 2。

表 2  $AR(3)$  回归结果

Table 2 Regression result of $AR(3)$				
变量	系数	标准差	$t$ -统计量	概率
常数	0.067 2	0.030 8	2.181 9	0.035 4
$AR(1)$	0.449 4	0.136 7	3.286 6	0.002 2
$AR(2)$	0.008 8	0.151 0	0.058 3	0.953 8
$AR(3)$	0.422 1	0.129 6	3.256 5	0.002 4

如表 2 所示,  $AR(2)$  项无统计学意义 ( $p = 0.953\ 8$ ), 因此剔除  $AR(2)$  项, 然后重新进行检验, 结果如表 3。

表 3  $AR(3)$ 剔除  $AR(2)$ 项后的回归结果

Table 3 Regression result of  $AR(3)$  culling  $AR(2)$

变量	系数	标准差	$t$ 统计量	概率
常数	0.067 3	0.030 3	2.219 2	0.032 4
$AR(1)$	0.453 4	0.116 8	3.881 1	0.000 4
$AR(3)$	0.426 6	0.101 7	4.194 4	0.000 2
决定系数	$R^2=0.907\ 5$		因变量均值=0.123 0	

此时模型参数均有统计学意义,  $AIC = -5.541\ 0$ ,  $SBC = -5.416\ 8$  为所检验的模型中的最小值,  $R^2 = 0.907\ 5$  模型的拟合效果好, 因此最后选用剔除  $AR(2)$  项的  $AR(3)$  模型对此平稳序列进行拟合估计。

模型的数学表达式为:

$$y_t = 0.067\ 3 + u_t$$
$$u_t = 0.453\ 4\ u_{t-1} + 0.4266\ u_{t-3} + e_t$$
$$y_t = 0.067\ 3 + 0.453\ 4(y_{t-1} - 0.067\ 3) + 0.426\ 6(y_{t-3} - 0.067\ 3) + e_t$$
$$x_t = x_{t-1} + 0.067\ 3 + 0.453\ 4(y_{t-1} - 0.067\ 3) + 0.426\ 6(y_{t-3} - 0.067\ 3) + e_t$$

以同样方法对样地 2 号和样地 3 号的前  $(N-2)$  年平均直径序列分别进行拟合, 与样地 1 拟合结果相同, 剔除  $AR(2)$  项的  $AR(3)$  模型为最优模型,  $R^2$  分别为 0.909 7、0.910 3。

2.3 模型的验证

2.3.1 模型的白噪声检验 对所求得的模型的残差进行白噪声检验, 检验结果如图 3 所示。

$LB$  统计量的  $p$  值大于 0.05, 通过检验认定残差是一个白噪声过程, 所以接受该模型。



Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.171	0.171	1.3211	
		2	0.022	-0.007	1.3440	
		3	-0.200	-0.208	3.2339	0.072
		4	-0.118	-0.053	3.9164	0.141
		5	-0.107	-0.074	4.4914	0.213
		6	-0.075	-0.089	4.7787	0.311
		7	-0.080	-0.092	5.1136	0.402
		8	0.071	0.063	5.3851	0.495
		9	-0.085	-0.172	5.8894	0.553
		10	-0.004	-0.026	5.8902	0.660

图 3 模型残差的白噪声检验结果

Fig. 3 Result of the white noise test of the model residual

2.3.2 模型的预测 拟合模型的目的是对未来趋势进行预测, 利用所得模型对 47、48 a 的短叶松林分平均直径进行动态预测, 模型的预测值、实际值和

预测误差见表 4。

表 4 模型动态预测的实际值、预测值、相对误差

Table 4 Actual value, predictive value and relative error of dynamic prediction

年数/a	实际值	预测值	相对误差
47	7.280	7.270 5	0.001 30
48	7.322	7.342 8	0.002 98

预测相对误差非常小, 分别为 0.130% 和 0.298%, 说明该模型预测的效果非常好。

用上面同样的方法对样地 2 号 49、50 a 和样地 3 号 49、50 a 的短叶松林分平均直径进行动态预测, 预测的相对误差分别为 0.092%、0.159%、0.005 5%、0.061 6% 相对误差都很小, 模型预测效果好。

3 模型的适应性验证

用 4 号样地的短叶松平均直径数据对以上所得模型形式验证其可行性。以剔除  $AR(2)$  项的  $AR(3)$  模型形式对 1~56 a 的平均直径数据  $Z_t$  进行拟合, 然后对 57~60 a 的林木平均直径进行预测。拟合结果如图 4。

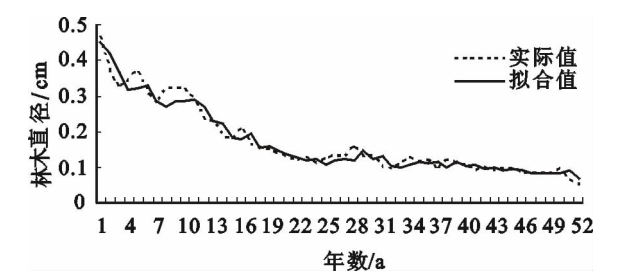


图 4  $z_t$  的一阶差分序列的拟合图

Fig. 4 Fitting graph of  $z_t$  first-order difference series

从图 4 看出, 模型预测的实际值与拟合值趋势一致, 并且  $R^2 = 0.949\ 8$ , 精度较高。用建立的模型对未来 4 a 即 57~60 a 的林木平均直径进行动态预测, 结果如表 5 和图 5。

表 5 模型动态预测年的真实值、预测值、误差

Table 5 Actual value, predictive value and relative error of dynamic prediction

年数/a	57	58	59	60
实际值	11.419	11.484	11.582	11.671
预测值	11.419	11.486	11.552	11.621
相对误差	0.000 026	0.000 148	0.002 556	0.004 267

如表 5 所示, 预测相对误差非常小 ( $< 1\%$ ), 因此, 通过验证可得此模型的拟合和预测精度高, 符合要求。

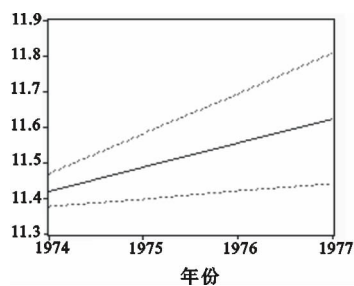


图 5 模型的动态预测图

Fig. 5 Dynamic prediction graph of the model

从图 5 看出,随着预测时间的延长,预测区间越来越宽,即随着预测年份的加长预测的误差越大,因为对于长期预测,在事物以后的发展过程中不可控制的因素较多,会较大幅度的影响预测的结果。

4 结论

通过对短叶松林分生长序列做时间序列建模分析,得到以下几点结论:

(1)根据 Box-Jenkins 建模原理,基于 Eviews 软件,经过反复选择、检验,最后选出了能够反映短叶松同龄纯林林分平均直径的变化规律,预测林分平均直径增长的趋势的模型形式  $AR(p)$ 模型。

(2)综合应用残差平方和法、偏相关函数法、AIC 和 BIC 准则法对模型定阶,最后确定阶数  $p=3$ ,然后经过参数的显著性检验和模型的有效性检验,确定最优模型形式剔除  $AR(2)$ 项的  $AR(3)$ 模型。

(3)用拟合模型对林分平均直径做动态预测,从预测结果来看,预测相对误差很小( $<1\%$ ),预测值比较接近实际,随着预测期的延长,预测误差会逐渐增大,使得预测值与现实值之间会有一定的差距。但其在短期预测中该模型的预测准确度很高。本文仅就短叶松林分平均直径生长模型进行了研究和探讨,所使用的方法和提出的模型形式可以为森林中其他树种的同类研究提供一定的借鉴,可以为制定森林经营措施以及收获预估计划提供科学的指导依据。

(4)通过对林木直径时间分析验证了时间序列分析这种动态数据处理的方法实践性强,用于预测没有固定的模式,并且实际事件本身的发展具有不平稳性,且随着时间延伸序列将不断地加入新的观

测值,所以一次分析所得到的 ARIMA 模型并不一定始终能作为最好的预测依据,因此,可通过逐年数据的补充,对模型不断进行修正。

参考文献:

[1] 张荷观. 时间序列的 ARIMA 模型在预测林木生长量中的应用[J]. 林业科学, 1986, 22(1): 94-100.  
ZHANG H G. Application of ARIMA time series model in tree growing forecast [J]. Scientia Silvae Sinicae, 1986, 22(1): 94-100.

[2] 邱学清, 江希钊, 黄健儿. 用时间序列叠合模型预测黑荆树林分月总生长量的研究[J]. 经济林研究, 1989, 7 (1): 62-69.  
QIU X Q, JIANG X D, HUANG J E. Forecasting the total monthly increment on wattle stand using the time series fold models[J]. Economic Forest Researches, 1989, 7 (1): 62-69.

[3] 陈广胜, 葛利, 郭仲凯. 人工林落叶松木材生长轮密度时间序列分析[J]. 东北林业大学学报, 2005, 33(2): 22-23.  
CHEN G S, GE L, GUO Z K. Time series analysis of the density of wood growth ring for larch plantations[J]. Journal of Northeast Forestry University, 2005, 33(2): 22-23.

[4] 谭伟, 余光辉, 彭世揆. ARMA 模型在单木生长预测中的应用[J]. 山地农业生物学报, 1998, 17 (6): 327-331.  
TAN W, SHE G H, PENG S K. Application of the ARMA model in prediction of individual tree's growth[J]. Journal of Mountain Agriculture and Biology, 1998, 17 (6): 327-331.

[5] 周元满, 谢正生, 刘新田. Richards 函数在桉树无性系林分生长预测上的应用研究[J]. 西南农业大学学报: 自然科学版, 2005, 27 (2): 240-243.  
ZHOU Y M, XIE Z S, LIU X T. Application of Richards growth function in prediction of the growth of eucalyptus clone stands[J]. Journal of Southwest Agricultural University: Natural Science Edition, 2005, 27 (2): 240-243.

[6] 杜纪山. 用二类调查样地建立落叶松单木直径生长模型[J]. 林业科学研究, 1999, 12(2): 160-164.  
DU J S. Modeling individual tree growth of Larix by using forest management inventory plots[J]. Forest Research, 1999, 12 (2): 160-164.

[7] 姚晓红. 林分生长数据的时间分析探讨[J]. 北京林业大学学报, 1990, 12(4): 10-16.  
YAO X H. Time series analysis of stand growth data[J]. Journal of Beijing forestry University, 1990, 12 (4) : 10-16.

[8] BOX G E P, JENKIS G M. Time series analysis: forecasting and control[M]. San Francisco: San Francisco Press, 1978.

[9] WILLIAM W S W. Time series analysis-univariate and multivariate methods(Second Edition)[M]. 北京: 中国人民大学出版社, 2009.