

基于蓄积的森林生物量估算方法的对比分析

尹惠妍, 李海奎

(中国林业科学研究院资源信息研究所,北京 100091)

关键词:一类清查;样地生物量;估算;模型

中图分类号:S718.5

文献标识码:A

Comparison of the Methods Estimating Forest Biomass Based on Stock Volume

YIN Hui-yan, LI Hai-kui

(Research Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: Based on national forest inventory data of Guangdong province, and selecting 176 sample plots dominated by *Pinus massoniana*, the biomass of the sample plots was estimated with different methods, which taking the keep sample stock volume as the independent variable and the sample plot biomass as the dependent variable. By comparing the calculation principle and process characteristics of these methods, the appropriate process was chosen to estimate the sample biomass. The results showed that the effect of continuous function for biomass expansion factor was poor, while that of the other methods was similar. For estimating the biomass of *Pinus massoniana*, the method of simultaneous equations with intercept was the best.

Key words: forestry inventory; sample plot biomass; estimation; models

森林生物量是指一个森林群落一定时间内积累的有机质总量,是森林生态系统重要的特征数据,因此世界各国越来越重视对森林生物量的监测与研究^[1-3],建立的生物量模型众多^[4-6]。大尺度森林生物量监测,是以省、流域、国家乃至全球为对象,在估算方法一致的前提下,对多个时间点的森林生物量进行估测和评价。开展全国森林生物量监测与评估,建立适合大区域的森林生物量的估算方法是一项重要的基础工作^[7]。

基于森林清查资料是估计区域、国家尺度乔木生物量的主要途径之一,森林资源清查的方法是在森林调查的基础上,对每个调查样地进行每木检尺,获取森林调查的详细数据^[8]。目前国际上对基于森

林连续清查资料的森林生物量估测主要有3种方法^[9]:IPCC法、生物量转换因子连续函数法和生物量回归模型法。3种方法都属于材积源生物量法,也叫生物量转换因子法,是利用林分生物量与木材材积比值的平均值,乘以该森林类型的总蓄积量,得到该森林类型的总生物量的方法。

我国对于森林生物量的研究从20世纪70年代开始发展,大多数学者对森林生物量都是在生态系统层次上的研究,虽然模型众多^[10-11],但实用性不强。在国家水平上树种复杂,对大尺度生物量的估计,不同的估计方法差别很大,对森林资源数据的充分利用与深度挖掘还不够。传统的林分生物量测定的方法主要为:皆伐实测法、标准木法与回归估计

收稿日期:2014-01-21

基金项目:国家自然科学基金“基于森林清查数据的大尺度森林碳储量监测方法研究”(31370634);国家高技术研究发展计划(863计划)“多尺度森林生物量建模及产品验证技术研究”(2012AA12A306)

作者简介:尹惠妍,女,中国林科院资源信息所硕士研究生,研究方向为生物量建模产品研究。E-mail:huiyanyin@163.com

* 通讯作者:李海奎,E-mail:hk_li@163.com

法,针对前两种方法,皆伐实测法对林木资源的浪费与破坏严重,标准木法的野外调查工作量太大,对于较大区域的森林生物量的估算适用性不强。本文主要以国家森林资源连续清查中广东省的数据为基础,以马尾松(*Pinus massoniana* Lamb.)为优势树种的样地为例,在已知样地蓄积情况下,探索样地生物量的估算方法,为区域尺度森林生物量的估算研究提供理论基础。

1 数据来源与处理

1.1 数据来源

本文主要利用第七次全国森林资源连续清查中广东省的样木表^[12]。样本数据会因季节、立地条件、采集过程等的差异而产生扰动数据^[13],若扰动数据参与建模,会引起建模结果的异常,本研究应用残差分析方法剔除异常值。最初样本为182个样地,剔除扰动数据之后为176个样地。文中原始数据的处理以及方程的拟合与验证都是借助forstat^[14]软件进行的。表1是马尾松建模样本统计量。

表1 生物量建模样本统计量

树种	样本量	平均数	标准差	最大值	最小值	
马尾松	176	样地地上生物量/t	1.606 0	1.529 6	6.835 9	0.007 2
		样地地下生物量/t	0.272 2	0.258 9	1.262 4	0.001 2
		样地总生物量/t	1.878 1	1.787 5	8.098 3	0.008 3
		样地蓄积/m ³	2.435 0	2.301 7	12.366 0	0.010 0
		优势树种蓄积比例	0.877 8	0.112 1	1.000 0	0.651 0

注:上表中地上生物量、地下生物量、样地蓄积均指以马尾松为优势树种的176样地的统计信息。

1.2 数据处理

根据广东省样木表,依据前人已建立的生物量模型^[15],计算样地的单木地上生物量与地下生物量,将单木生物量相加,得到各样地地上生物量与地下生物量以及样地总生物量。森林资源连续清查中统一使用各省根据部颁LY208-77二元立木材积表导算的一元立木材积表^[16]来估算单木材积,本研究直接使用连续清查样木表中的单木材积数据^[17],单株材积累加和得到样地蓄积,根据样地内各树种的蓄积比例,确定样地的优势树种。建模样本选择以马尾松为优势树种的样地数据。

2 研究方法

2.1 模型选择

以样地生物量为因变量,以样地蓄积为自变量,建立不同的生物量预估模型。本研究选取方精

云^[18]的转换因子连续函数法、一元线性截距式方程、一元线性无截距方程^[19]、联立方程组截距式与联立方程组无截距式^[20]共5种方法进行生物量估算。方精云的方法作为一种标准或对照,一元线性方程是最基本的数学方程形式,也是生物数学建模中常用的模型。由于本文的自变量只有样地蓄积一个,样地生物量可分为地上生物量与地下生物量,采用联立方程组模型的形式可以充分考虑地上生物量与地下生物量两者的相互制约关系,即包含根茎比的问题。本文分含截距与不含截距两种情况,以便观察在样地蓄积很小(甚至为0)时,样地生物量的情况。

$$\text{模型(1): } B = a \times V + b$$

式中: B 为单位面积生物量(t), V 为单位面积蓄积(m^3), a, b 为参数。其参数值为固定值,此方法为方精云的转换因子连续函数法,主要用于与其他模型进行对比,此方法基本统计单元为公顷,而本文树种的基本单位为 0.0667 hm^2 ,需进行单位转换。

$$\text{模型(2): } B = a \times V + b$$

式中: B 为样地生物量(t), V 为样地蓄积(m^3), a, b 为参数。

$$\text{模型(3): } B = a \times V$$

式中: B 为样地生物量(t), V 为样地蓄积(m^3), a 为参数。

$$\text{模型(4): } \begin{cases} Bt = a \times V + b \\ Br = a \times Vt \end{cases}$$

式中: Bt 为地上生物量(t), Br 为地下生物量(kg), V 为样地蓄积(m^3), a, b, c 为参数。

$$\text{模型(5): } \begin{cases} Bt = a \times V \\ Br = a \times Vt \end{cases}$$

式中: Bt 为地上生物量(t), Br 为地下生物量(kg), V 为样地蓄积(m^3), a, c 为参数。

2.2 模型评价指标

采用以下6项指标对生物量模型进行选型与评价,分别为决定系数^[21](R^2)、估计值的标准误差(SEE)、总相对误差(TRE)、平均相对误差(MSE)、平均相对绝对误差(RAM)和预估精度 P ^[22]。计算公式分别为:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$SEE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{(n - p)}}$$

$$TRE = \frac{\sum (y_i - \hat{y}_i)}{\sum \hat{y}_i} \times 100$$

$$MSE = \frac{1}{n} \sum (y_i - \hat{y}_i) / \hat{y}_i \times 100$$

$$RAM = \sum |(y_i - \hat{y}_i) / \hat{y}_i| / n \times 100$$

$$P = [1 - t_{\alpha} \cdot (SEE/\bar{y}) / \sqrt{n}] \times 100$$

式中: y_i 为实际观测值, \hat{y}_i 为模型估计值, \bar{y} 为样本的平均值, n 为样本单元数, p 为模型参数个数, t_{α} 为置信水平为 α 时的 t 值, 一般 α 取值为 0.05。在以上 6 项指标中, R^2 与 SEE 是回归模型的最常用指标^[23], 这两项指标可以反映模型的拟合优度, 也可以反映自变量的贡献率和因变量的离差状况。 TRE 与 MSE 是反映拟合效果的重要指标, 可以反映

出回归模型系统偏差的情况, 二者一般控制于一定的数值范围内(如 $\pm 5\%$), 趋于 0 时模型效果最好。 RAM 是反映单株生物量估计值的精度指标, 即每个样本点距回归曲线的相对平均误差。预估精度 P 是反映平均生物量估计值的精度指标。除此以外, 还可考虑参数的稳定性和残差的随机性。

3 结果与分析

3.1 生物量模型

各生物量模型的参数估计以及各相关评价标准如表 2 所示, 表 2 中给出了部分模型参数的 t 估计值。

表 2 模型参数估计与评价指标

模型	参数估计(t 估计值)			评价指标					
	参数 a	参数 b	参数 c	R^2	SEE	$TRE/\%$	$MSE/\%$	$RAM/\%$	$P\%$
(1)	0.510 1	1.045 1	-	0.742 8	909.162 7	43.176 4	33.312 1	39.804 2	92.80
(2)	0.753 9(53.265 9)	42.482 5(0.897 1)	-	0.942 3	430.698 7	-0.004 7	-3.054 4	15.732 8	96.58
(3)	0.763 1(78.654 8)	-	-	0.942 0	431.694 6	1.075 9	3.144 3	15.540 1	96.53
(4)	0.644 8	36.482 7	0.169 1	0.942 3	430.698 7	-0.004 2	-3.068 2	15.737 9	96.58
(5)	0.652 7	-	0.169 1	0.942 0	431.694 6	1.075 6	3.143 9	15.540 0	96.53

注: 上表中部分模型参数后括号中的数值为该参数的 t 估计值。

从表 2 中可以看出, 5 种方法中转换因子连续函数法的决定系数最小为 0.742 8, 其预估精度也较小为 92.80, 其他反映系统误差的指标值则比较大, 相比之下, 其他 4 种方法预估效果较此方法好。其他模型的决定系数均达 0.94, 其中最高值为一元截距式的 0.942 3。从决定系数与估计值的标准误差来看, 含截距式的一元函数与联立方程组形式预估效果相当, 均优于无截距式的模型形式。从总相对误差看, 联立方程组截距式模型的数值最小, 其次为一元截距方程, 其余两种模型则相近。总相对误差与平均相对误差都是反映回归模型系统偏差情况的指标, 从这两项指标看, 含截距的联立方程组形式较为合适。平均相对绝对误差是反映每个样点距回归曲线的相对平均误差, 从此项指标来看 4 种模型的数值基本相同, 各模型的预估精度均较高。

3.2 常数项显著性检验

由于本研究中用于建立回归模型的自变量只有样地蓄积量一个, 不对自变量的参数值做显著性检验。

一元线性截距式与一元线性无截距式存在常数项的差异, 现对常数项进行显著性检验。模型(2)中常数项的 t 值为 0.897 1, 查 t 分布表的, 当 $\alpha = 0.05$ 时, $t_{\alpha/2}(200) = 1.972$, $t_{\alpha/2}(100) = 1.984$, 所以

$1.972 < t_{\alpha/2}(176) < 1.984$, $t = 0.891 < 1.972$, 所以 $\alpha = 0.05$ 的水平下, 模型一中的常数项与 0 无显著性差异。

在联立方程组的两种估计方法中, 两种模型的区别在于是否含有常数项, 模型的总平方和(TSS)与残差平方和(RSS)之差剩余平方和(GSS), 是由常数项引起的误差项。现将两种模型的残差平方和(RSS)进行比较, 模型四为 RSS_1 , 模型五为 RSS_2 , 看模型中常数项的显著性, 可用 $F = \frac{RSS_2 - RSS_1}{GSS/(n-2)} \sim F(1, n-2)$, 给定显著水平 α , 若 $F > F_{\alpha}$, 则表明 b 显著; 若 $F < F_{\alpha}$ 则, 表明 b 不显著。经计算得, $F = 0.002 317$, 常数项与 0 不具有显著性。

3.3 模型综合评价

从表 2 中各模型的评价指标来看, 各评价指标数值相近, 难以确定模型的优劣程度。现对各模型进行综合评价^[24], 其中决定系数(R^2)与预估精度 P 为极大型指标, 即总是期望指标的取值越大越好。而估计值的标准误(SEE)、总相对误差(TRE)、平均相对误差(MSE)与平均相对绝对误差(RAM)的绝对值为极小型指标, 即总是期望指标的取值越小越好。先对各指标值进行一致化处理^[25], 将极小型数值通过求倒数转化为极大型数值, 即对各模型的

SEE、TRE、MSE、RAM 的绝对值求倒数,如公式(1)所示。

由于各项指标都有不同的单位与数量级,使得各指标之间存在着不可公度性。为避免在模型的综合评价过程中出现“大数吃小数”的错误结果,现将一致化后的各评价指标数值进行无量纲化处理^[26],即标准化处理,如公式(2)所示。对各评价模型进行预处理之后,根据6种模型,各指标值的标准差占所有指标标准差的权重确定各指标值得权重系数。从而得到各模型的综合评价结果。计算公式如下所示:

对于 n 个评价对象 X_{ij} ,每个评价对象有 m 个评价指标,其观测值分别为

$$X_{ij}(i = 1, 2, \dots, n; j = 1, 2, \dots, m)$$

一致化处理公式如(1)所示:

$$X'_{ij} = 1/|X_{ij}| \tag{1}$$

无量纲化处理方法如公式(2)所示。

标准样本变换法为:

$$X''_{ij} = \frac{X'_{ij} - \bar{X}_j}{S_j} \tag{2}$$

权重系数的计算方法如公式(3)所示

第 j 项指标的样本均值与样本均方差依次为

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X'_{ij},$$

$$S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (X'_{ij} - \bar{X}_j)^2}$$

则第 j 项指标的权重系数为

$$W_j = \frac{S_j}{\sum_{i=1}^n S_j} \tag{3}$$

综合评价模型计算公式如公式(4)所示。

$$y_j = \sum_{i=1}^n W_j \times X_{ij} \tag{4}$$

y_i 即为第 i 个评价对象的综合评价价值。

以上4个公式中: $i=1,2,3,4,5$ 代表模型数目, $j=1,2,3,4,5,6$ 代表指标数目。样本均值为 $\bar{X}_j =$

$$\frac{1}{n} \sum_{i=1}^n X'_{ij}, \text{样本方差为 } S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (X'_{ij} - \bar{X}_j)^2}。$$

得从表3中可以看出,模型4的预估效果最好,其次为模型2,模型3与模型5的预估效果相当。

表3 变换后的各模型评价指标的值

指标	模型1	模型2	模型3	模型4	模型5
R^2	-0.000 6	0.000 2	0.000 2	0.000 2	0.000 2
SEE	-1.467 2	0.370 8	0.362 8	0.370 8	0.362 8
TRE	-0.053 7	0.072 5	-0.053 1	0.087 5	-0.053 1
MSE	-0.104 3	0.028 3	0.024 1	0.027 7	0.024 2
RAM	-0.074 2	0.017 6	0.019 5	0.017 6	0.019 5
P	-0.011 5	0.003 0	0.002 8	0.003 0	0.002 8
总和	-1.711 5	0.492 4	0.356 2	0.506 7	0.356 2

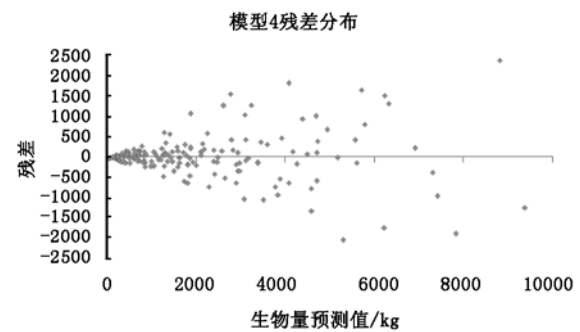
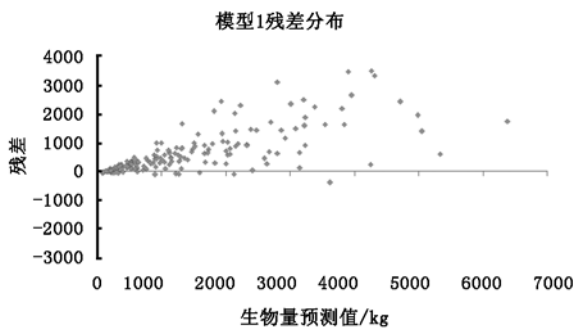


图1 模型(1)与模型(4)的残差分布图

3.4 各模型残差分布

除转换因子连续函数法的残差图外,各模型的残差分布图均为与 x 轴对称的图形,残差值在 x 轴上下随机散布,这说明回归直线对原始观测值的拟合情况良好。本文中模型(1)的残差图与其余模型的残差图差异较大,其余模型残差图分布相似,这里只给出模型(1)与最优模型(4)的残差分布图(图1)。残差呈现喇叭形分布,这说明模型存在异方差

性,可采用适当形式缩小预估变量的变动幅度,在一定程度上消除异方差^[27]。残差图表达各模型残差的分布情况,本文不再对异方差问题做深入探讨。

4 结论与讨论

4.1 结论

通过对5种模型的评价指标的比较分析及模型的综合评价,发现含截距的联立方程组模型(4)是

预测效果最好的,其次是含截距式的一元线性方程模型(2),不含截距式的一元线性方程与联立方程组模型预估效果相当。

样地生物量的测算方法的不同还与样地树种组成、立地条件、使用地域、样本单元等因素有关。本文就一类清查资料的样地水平生物量的估算方法进行研究,在此基础上可以按样地布设的比例直接扩展,得到省级水平该优势树种的生物量,实现不同尺度间的生物量扩展。

4.2 讨论

4.2.1 转换因子连续函数法 通过对模型各项评价指标及模型的综合分析得知,方精云的转换因子连续函数法预估效果不理想。可能的原因有以下两点,首先转换因子连续函数法的建模数据偏小,代表性不够,例如本文马尾松,其建模样本量只有12,适用的树种有马尾松(*Pinus massoniana* Lamb.)、云南松(*Pinus yunnanensis*)、思茅松(*Pinus kesiya* var. *langbianensis*)3种,单独估算一个树种的生物量偏差会增大。转换因子连续函数法是基于全国758组研究数据,并将中国的森林类型分为21类,实现了由样地调查向区域推算的尺度转换,其研究范围广泛,对于特定区域的预估结果存在差异。

该方法在某种程度上可以反映生物量随林龄的变化,但单位面积蓄积对生物量的影响较大,本文中的样地蓄积范围为 $0.01 \sim 12.37 \text{ m}^3$,扩展到公顷为 $0.15 \sim 185.55 \text{ m}^3$,转换因子分别为7.4774与0.5945,使得预估误差偏大。一类清查中某些样地蓄积较小,可能是由于林分内达到起测径阶的树木较少,也可能林分遭遇外界因素被毁,转换因子连续函数法对于较小的范围(如省级水平)不太适用。

4.2.2 常数项的意义 从表2以及各模型的残差分布图可以看出,其余4种模型的预测效果均较好,含截距式的模型(2)与不含截距模型(3),以及含截距的模型(4)与不含截距的模型(5)相比,常数项均与0无显著性差异,这是因为当样地蓄积趋近于0时,样地生物量也为0。

在实际应用中,线性回归模型中的常数项在很多情况下无实际意义。其具体含义为当所有解释变量的值为0的时候,被解释变量的数值。但是在林业研究中这是无意义的,就本文中的解释变量样地蓄积而言,其定义域并不含0,树木的起测胸径为5cm,样地蓄积就不可能为0。从线性回归的本质上来讲,所有参数的确定都为了使残差项的均值为0,

且残差项的平方和最小。因此,可以理解为常数项是对其他各个解释变量所留下的偏误的线性修正。本文中的模型(2)与模型(4)的预估效果均高于其对应的模型(3)与模型(5)。

4.2.3 联立方程组模型的分析 联立方程模型是指同时用若干个相互关联的方程,去表示一个被解释变量相互依存性的模型^[28],有些因变量可以以自变量的形式出现在另一个方程的右边,可以较深入的分析各变量的因果关系。如本文中的联立方程组,第一个方程的因变量样地地上生物量作为第二个方程的自变量,使两个变量相互制约。从表2中可以看出,模型(4)与模型(2)的预估效果相差不多,从模型的总体评价结果中显示出联立方程组模型的优势。

参考文献:

- [1] Kitterge J. Estimation of amount of foliage of trees and shrubs[J]. *J Forest*, 1944,42:905-912.
- [2] Ruard G A, martin G L, Bockheim J G. Comparison of constant and variable allometric ratios for estimating *Populus tremuloides* biomass [J]. *For Sci*, 1987, 33(2): 294-300.
- [3] Fehrmann L, Kleinn C. General consideration about the use of allometric equations for biomass estimation on the example of Norway spruce in central Europe [J]. *For Ecol Manag*, 2006, 236: 412-421.
- [4] Ter-Mikaelian M T, Korzukhin M D. Biomass equations for sixty-five North American tree species [J]. *Forest Ecology and Management*, 1997(97):1-24.
- [5] Zianis D, Muukkonen P, Makiapaa R, et al. Biomass and stem volume equations for tree species in Europe [J]. *Silva Fennica*, 2005(4):1-63.
- [6] Zianis D, Radoglou K. Comparison between Empirical and Theoretical Biomass Allometric Models and Statistical Implications for Stem Volume Predictions [J]. *Forestry*, 2006, 79(4): 477-487.
- [7] 曾伟生,唐守正. 东北落叶松和南方马尾松地下生物量模型研建 [J]. *北京林业大学学报*, 2011, 32(2):1-6.
- [8] 张志,田昕,陈尔学,等. 森林地上生物量估测方法研究综述 [J]. *北京林业大学学报*, 2011, 33(5):144-150.
- [9] 李海奎,赵鹏祥,雷渊才,等. 基于森林清查资料阔乔木林生物量估算方法的比较 [J]. *林业科学*, 2012, 48(5):44-52.
- [10] 冯宗炜,王效科,吴刚. 中国森林生态系统的生物量和生产力 [M]. 北京:科学出版社, 1999.
- [11] WANG C K. Biomass allometric equations for 10 co-occurring tree species in Chinese temperate forests [J]. *Forest Ecology and Management*, 2006(222): 9-16.
- [12] 李海奎,雷渊才,曾伟生. 基于森林清查资料的中国森林植被碳储量 [J]. *林业科学*, 2011, 47(7):7-12.
- [13] 陈文灿,张伟东,申屠惠良,等. 生物量建模过程中剔除扰动数

- 据方法探讨[J]. 华东森林经理,2012,26(1):61-62.
- [14]唐守正,郎奎建,李海奎. 统计和生物数学模型计算(forstat教程)[M]. 北京:科技出版社,2009:268-290.
- [15]李海奎,雷渊才. 中国森林植被生物量和碳储量评估[M]. 北京:中国林业出版社,2010.
- [16]常 昆. 导算一元立木材积表的技术方法[J]. 林业资源管理,1977(1):1-10.
- [17]孟宪宇等. 测树学[M]. 北京:中国林业出版社,2006:127-134.
- [18]Fang J ,Y, Chen A P, Peng C H, *et al.* Changes in forest biomass carbon storage in China between 1949 and 1998 [J]. Science, 2001. 292(5525):2320-2322.
- [19]郑立生,李海奎. 模型形式和地域对落叶松地上生物量预测的影响[J]. 林业资源管理,2013,2:83-88.
- [20]郎璞玫. 建立在固定样地上的林分生长线性联立方程组模型研究[J]. 北京林业大学学报,2007,29(1):37-41.
- [21]曾伟生. 全国立木生物量建模方法研究[D]. 北京:中国林科院,2011.
- [22]胥 辉. 一种生物量模型构建的新方法[J]. 西北农林科技大学学报:自然科学版. 2001,29(3):35-40.
- [23]曾伟生,张会儒,唐守正. 立木生物量建模方法[M]. 北京,中国林业出版社,2011:49-65.
- [24]金贞珍. 关于多指标综合评价方法及其权数问题的讨论[D]. 吉林:延边大学,2007.
- [25]韩中庚,陆宜清,周素静. 数学建模实用教程[M]. 北京,高等教育出版社,2012.
- [26]钟 霞,钟怀军. 多指标综合评价方法及应用[J]. 内蒙古大学学报:人文社会科学版,2004,36(4):107-111.
- [27]曾伟生,骆期邦,贺东北. 论加权回归与建模[J]. 林业科学,1999,35(5):5-11.
- [28]贾新明,刘 亮. 结构方程模型与联立方程模型比较[J]. 数理统计与管理,2008,27(3):439-446.