

文章编号: 1007-4619 (2001) 04-0277-05

以遥感和地理信息系统为基础的森林蓄积 LS 估计 自变量选择研究

李崇贵, 赵宪文

(中国林业科学研究院 资源信息研究所, 北京 100091)

摘 要: 根据最小二乘(LS)原理, 借助遥感和地理信息系统, 研究森林蓄积定量估测最优自变量的选择方法。通过实例分析了 LS 估计确定森林蓄积自变量的选择规律, 并探讨了 LS 估计预报森林蓄积的潜力, 对实际工作具有一定的指导意义。

关键词: LS 估计; 平均残差平方和; 森林蓄积

中图分类号: TP79 **文献标识码:** A

借助遥感(RS)和地理信息系统(GIS)进行森林蓄积定量估测, 国内外学者作过不少探讨。当前较多采用地面样地所对应的 RS 图像各波段灰度值、灰度比值及样地 GIS 信息为自变量, 以样地蓄积观测值为因变量, 通过数量化方法, 建立蓄积和自变量之间的回归模型, 进行蓄积量估测^[1,2]。在所建方程中, 自变量个数的确定、自变量的选择对估测的可靠性及精度影响很大。有的学者趋向选择较多的 RS 波段信息^[1], 有的则趋向选择较多的样地 GIS 信息^[2]。究竟哪些信息与蓄积估测关系较大, 在所见的森林蓄积定量估测文献中, 尚未进行这方面的系统分析与仿真研究。本文的目的是采用 LS 估计, 以地面样地对应的 RS 和 GIS 信息为基础, 通过计算机仿真研究影响蓄积估测自变量的选择规律。

1 蓄积估计变量选择原则

设某林区地面样地总数为 n , 样地的蓄积观测值向量为 V_o , 在样地对应的 RS 和 GIS 信息中, 影响蓄积估测的全部自变量维数为 p (含常数项), 其观测矩阵为 X , 则该林区蓄积线性估测模型可表示为

$$V_o = XY + e, E(e) = 0, COV(e) = \sigma^2 I \quad (1)$$

式中, e 为 $n \times 1$ 随机误差向量; σ^2 为样地蓄积观测方差; γ 为 $p \times 1$ 待定参数向量; I 为 n 阶单位阵。

将 X 表示成分块形式 $X = (X_q : X_t)$, 相应将 γ 表示成 $\gamma' = (\gamma'_q : \gamma'_t)$, 则(1)式可变为:

$$V_o = X_q \gamma_q + X_t \gamma_t + e \quad (2)$$

式中 X_q 包含常数项, 且 X_q 和 X_t 分别为 q, t 列的列满秩阵。考虑以下两种情况:

1) 蓄积真实模型为 $V_o = XY + e$, 因丢掉了一些必要的自变量, 而被误认为

$$V_o = X_q \gamma_q + e \quad (3)$$

2) 蓄积真实模型为 $V_o = X_q \gamma_q + e$, 因引进了一些多余的自变量, 而被误认为

$$V_o = XY + e \quad (4)$$

设模型(3)和(4)两种模型待定参数的 LS 估计及方差估值分别为:

$$\tilde{\gamma}_q = (X'_q X_q)^{-1} X'_q V_o$$

$$\hat{\sigma}_q^2 = V'_o (I - X_q (X'_q X_q)^{-1} X'_q) V_o / (n - q)$$

$$\gamma = (X'X)^{-1} X'V_o = \begin{pmatrix} \gamma_q \\ \Lambda \\ \gamma_t \end{pmatrix}$$

$$\hat{\sigma}^2 = V'_o (I - X(X'X)^{-1} X') V_o / (n - p - 1)$$

则有以下结论^[3]

1) 若模型(4)正确, 则有

$$E(\tilde{\gamma}_q) = \gamma_q + (X'_q X_q)^{-1} X'_q X_t \gamma_t$$

$$COV(\tilde{\gamma}_q) \geq COV(\gamma_q)$$

要使 $\tilde{\gamma}_q$ 为 γ_q 的无偏估计, 必有 $(X'_q X_q)^{-1} X'_q X_t \gamma_t =$

收稿日期: 2000-09-01; 修订日期: 2000-12-14

基金项目: 遥感和 GIS 技术用于森林资源空间动态定量预测方法的研究。

作者简介: 李崇贵(1966—), 男, 四川省西充县人, 中国林业科学研究院林业遥感博士研究生。主要从事遥感、GIS 及 GPS 在林业资源及国土资源调查、定量监测等方面的研究与开发, 已发表论文多篇。

0。此表明要么 $\gamma_t=0$, 即后 t 个自变量与蓄积无关, 模型(3)正确; 要么矩阵 $X=(X_q; X_t)$ 的两部分正交, 即后 t 个自变量对蓄积估测不起任何作用。另外, 即使模型(4)正确, 丢掉一部分自变量后, 总使剩下的那部分自变量回归系数的 LS 估计的方差减小。

2) 若模型(4)正确, 当 $COV(\gamma_t) \geq \gamma_t \gamma_t'$ 时, 则有

$$COV(\gamma_q) \geq MSEM(\tilde{\gamma}_q)$$

$$Var(\Delta v_o) \geq MSEP(\tilde{v}_o = E(\Delta v_{oq}^2))$$

式中, $MSEM(\tilde{\gamma}_q)$ 为待估参数的平均平方误差矩阵; Δv_{oq} 、 Δv_o 分别为模型(3)和(4)的蓄积预测偏差; $MSEP(\tilde{v}_o)$ 为模型(3)的蓄积预测均方误差。

此表明: 若丢掉的自变量影响确实存在 ($\gamma_t \neq 0$), 但 γ_t 难以准确估计(用 $COV(\gamma_t) \geq \gamma_t \gamma_t'$ 刻画), 丢掉这样的变量可使预测精度提高; 相反, 若模型包含了不必要的自变量 ($\gamma_t=0$), 则使预测精度降低。

因此, 森林蓄积估测自变量的选择应遵循少而精的原则, 尽量丢掉那些影响很小或难以准确估计的变量。

2 自变量选择

2.1 自变量设置

以 RS 和 GIS 为基础, 建立森林蓄积估测模型, 自变量应是通过 RS 和 GIS 所能获取的地面样地信息。RS 所能提供的信息包括: 各波段的灰度值、灰度比值及色彩等。森林资源调查样地所建 GIS 系统能提供的信息包括: 样地的纵横坐标、海拔、地类、坡度、坡向、郁闭度等。当以 TM 为遥感信息时, 根据已有文献[1, 2], 本文设置如表 1 可供选择的自变量。

设置依据可概括如下:

(1) 考虑 TM 遥感数据各波段的特点及植物的波谱反射特征与植物分布密度呈线性相关, 可设置 $TM_1, TM_2, TM_3, TM_4, TM_5, TM_7$ 及 $TM_4/TM_3, (TM_4 - TM_3)/(TM_4 + TM_3)$ 为变量。

(2) 考虑森林在自然状态下或一定区域范围内呈连续分布, 蓄积与地理位置有一定的相关关系, 可设置样地的纵、横坐标及海拔高为变量。

(3) 地类包括有林地、疏林地、宜林地及其它类型, 按数量化方法处理后, 作为影响蓄积估测的变量。

(4) 郁闭度直接反映森林生长状况, 一般情况下, 郁闭度越大, 森林蓄积量可能就会越大, 因此, 可

作为影响蓄积估测的变量。但由于它在航天遥感图像上所判读郁闭度等级太少, 需先通过样地对应的 RS 和 GIS 信息确定。

表 1 蓄积估测方程中可供选择的变量

Table 1 The selectable arguments in the equation of forest volume estimation

编号	变量名称	对应信息
1	α	常数项
2	x_1	TM_1
3	x_2	TM_2
4	x_3	TM_3
5	x_4	TM_4
6	x_5	TM_5
7	x_6	TM_7
8	x_7	TM_4/TM_3
9	x_8	$(TM_4 - TM_3)/(TM_4 + TM_3)$
10	x_9	纵坐标/km
11	x_{10}	横坐标/km
12	x_{11}	有林地
13	x_{12}	疏林地
14	x_{13}	宜林地
15	x_{14}	海拔高/km
16	x_{15}	郁闭度

2.2 自变量选择准则^[3]

LS 估计有多种自变量选择准则^[4], 考虑其有效性和便于计算机仿真, 本文采用平均残差平方和准则, 按回归树的原则, 从所有可能的自变量子集中选择对蓄积估计最优的自变量子集。当自变量个数为 q 时, 平均残差平方和 (RMS_q) 为

$$RMS_q = \frac{V_o'(I - X_q(X_q'X_q)^{-1}X_q')V_o}{n - q} \quad (3)$$

LS 估计蓄积, 残差平方和随自变量个数的增加而减小, 其最小值在回归方程包含了对蓄积估计有影响的全部自变量时达到, 按(3)式选择自变量, 随着自变量个数 q 的增加, $(n - q)^{-1}$ 将增大, 因分子在不断减小, 开始时分子减少更多, 故 RMS_q 将逐渐减小, 到达最小后, 又将逐渐增大。其最小值对应的变量子集即为影响蓄积估测的最优变量子集。

3 实例分析

以云南思茅地区 1 幅 (6967 × 5965) 1992 年的 TM 图像和当年样地的一类调查资料为例进行分析。为保证遥感图像几何精校正精度, 本文采用美国 RSI 公司的 IDL 语言从 1 幅图像中分割出 1/4 的区域 (0:3483, 0:2982) 进行试验。利用该地区 1991 年

1:5万地形图,共选择了56个地面控制点,用ENVI遥感处理平台,按双线性内插重采样法进行了几何精校正,几何位置精度均在1个像元以内。根据校正后RS图像的平面坐标和地面样地的平面坐标,通过IDL语言编程,由计算机自动读取样地所对应像元的灰度值,1/4遥感图像区域内共包含129个样地。由各样地对应的RS图像灰度值及相应的一类调查资料,按表1的方式分别读取129个样地各

变量的值,并进行中心标准化。

根据表1的设置,包括常数项在内蓄积估测方程共有16个可供选择的变量。为分析随着 q 的增加,入选变量及 RMS_q 的变化规律,本文分析了不包括1个变量的所有可供选择的65519个子集的状况。不同变量个数 q 对应的可供选择的子集数 τ 、最优子集、相应的 RMS_q 及复相关系数如表2。

由表2可得以下结论:

表2 不同 q 值对应的最优自变量子集及相应的 RMS_q

Table 2 The optional argument subset and its RMS_q of different q value

q	不同 q 值对应的最优自变量子集	RMS_q	τ	复相关系数
2	α, x_{15}	12.1034	120	0.8276
3	α, x_9, x_{15}	11.9867	560	0.8309
4	α, x_1, x_9, x_{15}	11.9315	1820	0.8332
5	$\alpha, x_4, x_6, x_7, x_{15}$	11.8610	4368	0.8358
6	$\alpha, x_2, x_3, x_4, x_7, x_{15}$	11.8353	8008	0.8376
7	$\alpha, x_2, x_3, x_4, x_7, x_9, x_{15}$	11.7681	11440	0.8401
8	$\alpha, x_2, x_3, x_4, x_6, x_7, x_9, x_{15}$	11.7952	12870	0.8411
9	$\alpha, x_1, x_3, x_4, x_6, x_7, x_9, x_{10}, x_{15}$	11.8549	11440	0.8417
10	$\alpha, x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}, x_{15}$	11.9135	8008	0.8423
11	$\alpha, x_1, x_2, x_3, x_4, x_6, x_7, x_9, x_{10}, x_{11}, x_{15}$	12.0080	4368	0.8424
12	$\alpha, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{10}, x_{11}, x_{15}$	12.1075	1820	0.8424
13	$\alpha, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{15}$	12.2096	560	0.8244
14	$\alpha, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{14}, x_{15}$	12.1040	120	0.8246
15	$\alpha, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{14}, x_{15}$	12.1036	16	0.8246
16	$\alpha, x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}$	12.5318	1	0.8247

(1) 随着 q 的增加, RMS_q 逐渐减小,达到最小点后又逐渐增加,随后可能出现多个局部极小点,试验区蓄积估测最优方程在 $q=7$ 时达到,相应的最优变量子集为 $\alpha, x_2, x_3, x_4, x_7, x_9, x_{15}$ 。

(2) 随着 q 的增加,不同 q 值对应的最优变量子集与样地蓄积的复相关系数将逐渐增加,达到最大后又开始减小, $q=7$ 时的复相关系数与最大值相差很小。

(3) 不管蓄积估测方程中 q 值如何变化,常数项和样地郁闭度都必不可少。若方程中不包含郁闭度因子,平均残差平方和将会增大,蓄积估测精度也将显著降低,说明郁闭度对森林蓄积估测影响显著。

(4) 当 $q=7$ 时, RMS_q 达到最小,当 $q=8$ 时,其 RMS_q 和 $q=7$ 相差很小,且二者的最优自变量子集除新增变量外完全相同。因此,可得遥感图像的 TM_2, TM_3, TM_4, TM_7 及 TM_4/TM_3 与蓄积估测密切关系。在表征样地地理位置的3个变量中,代表样地纬度变化的纵坐标与蓄积估测有密切关系。

(5) 地类所属的有林地、疏林地、宜林地及其它与蓄积估测关系较小。随着自变量的增加,最先入选的地类因子为有林地,这与有林地在其它3种地类类型中对蓄积的影响最大相符,此表明 RMS_q 能有效筛选最优自变量子集。

(6) 以遥感和GIS为基础的森林蓄积估测,有的学者还考虑了样地土壤类型、土壤厚度、坡度、坡向、优势树种等因子^[2]。鉴于蓄积估测方程中所含变量的可获取性及方程的实用性,本文设置了表1的变量。但不管可选变量如何变化,按本文方法都能有效筛选最优变量子集。

(7) 当在表1中增加阴坡、阳坡及无坡向3种样地定性因子时,按相同的方法进行变量筛选,对蓄积估测起主要作用的因子将发生较大变化,预报精度也明显降低。此说明在云南思茅地区,坡向对蓄积估测的影响很小,考虑RS各波段灰度值已包含了坡向的影响,因此坡向不宜作为该地区蓄积估测可供选择的变量。

为研究所得最优变量子集的预报精度,在129个样地中,分3种方案进行预报试验,方案1为随机抽取的40个预报样地,方案2,3各包含顺序抽取的41预报样地,各方案样地分布如图1,2所示。建立蓄积估测方程时,预报样地不参加回归,预报样地的郁闭度使用其余样地的RS和GIS信息确定。 $q=7$ 对应的最优蓄积估测方程及蓄积预报偏差相对误差如表3。3种方案样地蓄积预报值与实测值的比较

如图3。表中数据表明,对40个左右样地,不论其分布状况如何,总蓄积预报平均精度将不低于90%,可满足生产的要求。3方案绝大部分样地蓄积预报值与实测值的偏差很小,仅在个别点处出现异常,其原因是由于在公里网交点处上的个别样地蓄积差异较大所致,当对大面积的连续森林资源蓄积进行预报时,将不会出现此现象,总蓄积预报精度会提高。

表3 不同预报点集对应的回归方程及预报精度

Table 3 The regression equation and forecast accuracy of different predicting point set

预报点集	最优蓄积预报方程	蓄积预报相对误差/%
方案1	$V_{\hat{0}i} = 20.8120 + 0.0052x_2 + 0.3473x_3 - 0.2337x_4 + 4.4892x_7 - 10.7551x_9 + 15.6538x_{15}$	8.15
方案2	$V_{\hat{0}i} = 56.5198 - 0.1450x_2 + 0.2206x_3 - 0.0499x_4 + 1.1152x_7 - 22.5390x_9 + 16.0215x_{15}$	6.52
方案3	$V_{\hat{0}i} = -28.7241 - 0.6589x_2 + 0.6746x_3 - 0.2042x_4 + 4.5360x_7 + 10.2731x_9 + 15.1868x_{15}$	10.22

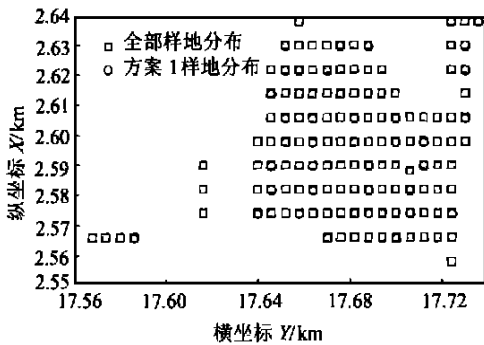


图1 方案1预报样地分布

Fig. 1 The forecast sample plot distribution of scheme 1

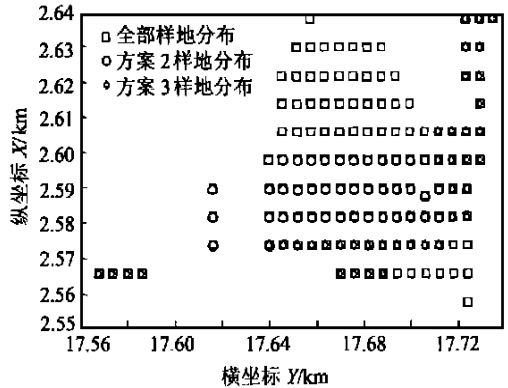


图2 方案2,3预报样地分布

Fig. 2 The forecast sample plot distribution of scheme 2 and scheme 3

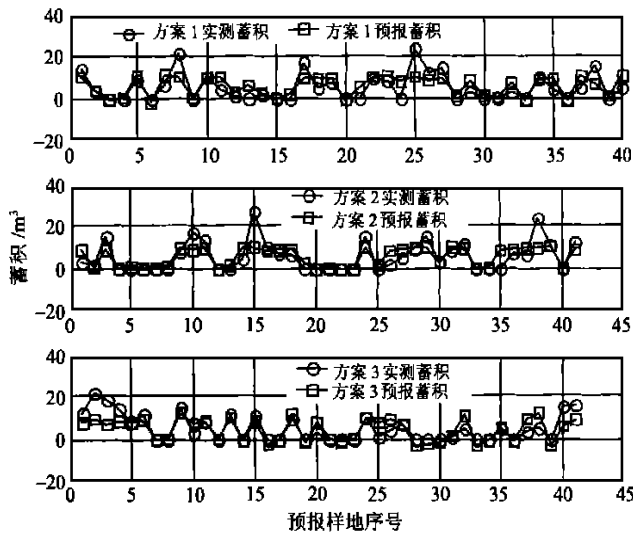


图3 3种方案样地蓄积预报值与实测值的比较

Fig. 3 The comparison of surveyed and predicted volume in 3 schemes

4 结 论

(1) 以遥感和GIS为基础,采用LS方法估测森林蓄积,利用 RMS_q 最小的准则可有效选取最优变量子集。利用30%—40%左右的样地建立蓄积估测方程,总蓄积预报精度可达90%以上,在实际工作中具有较好的应用前景。

(2) TM图像的 TM_2, TM_3, TM_4, TM_7 波段及波段比值 TM_4/TM_3 对蓄积估测有一定影响。这是因为这些波段反映了植被特点, TM_2 为绿色植物在可见光中的反射小高峰, TM_3 为反映植物吸收红光进行光合作用, TM_4 为植物高反射区, TM_7 也为植物水分反射特征波段,从而与蓄积量的估测相关紧密。在样地对应的GIS信息中,郁闭度对蓄积估测的影响较大,且可通过RS和GIS有效确定。表征样地纬度变化的纵坐标有一定的影响,而地类因子影响较小。因此以遥感和GIS为基础的森林蓄积估测,RS是基础,样地GIS信息不可缺少。

(3) 不带权LS估计要求各样地蓄积观测值的

方差相等。在实际状况下,此条件不一定满足,致使所建蓄积估测方程落实到林班的预报精度会出现一定的偏差,此可望通过带权估计或样地蓄积方差稳定化变换得到改善。

(4) LS估计变量选择很复杂。样地蓄积方差的非齐性、强影响点、高杠杆点及试验区范围大小对蓄积估测变量选择的影响规律还有待于进一步研究。

参 考 文 献 (References)

- [1] Zhao Xianwen. Quantitative estimation of forest remote sensing [M]. Beijing: China Forestry Press, 1997. [赵宪文. 林业遥感定量估测[M]. 北京: 中国林业出版社, 1997.]
- [2] You Xianxiang. Research on forest resource surveying and dynamic monitoring and information management system [M]. Beijing: China Forestry Press, 1995. [游先祥. 森林资源调查、动态监测、信息管理系统的研究[M]. 北京: 中国林业出版社, 1995.]
- [3] Chen Xiru, Wang Songgui. Modern regression analysis [M]. Hefei: Anhui Education Press, 1987. [陈希孺, 王松桂. 近代回归分析[M]. 合肥: 安徽教育出版社, 1987.]
- [4] Tang Shouzheng. Multivariate statistical analysis [M]. Beijing: China Forestry Press, 1989. [唐守正. 多元统计分析方法[M]. 北京: 中国林业出版社, 1989.]

A Research on the Variable Selection of Forest Volume Least Square Estimation Based on RS and GIS

LI Chong-gui, ZHAO Xian-wen

(The Research Institute of Forest Resources Information, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: The optimum variable selecting method on the forest volume quantitative estimation based on RS and GIS through the principle of least square is studied. The law of the variable selection on forest volume estimation is analyzed in detail by means of the real example, at the same time the predicting potentiality of forest volume through LS estimation is also discussed. The method has some guiding significances for the practical job.

Key words LS estimation; residual mean squares; forest volume