

基于偏最小二乘法的平南县森林蓄积量 估测模型研究

洪奕丰, 林 辉, 严恩萍, 李永亮

(中南林业科技大学 林业遥感信息工程研究中心, 湖南 长沙 410004)

摘要: 以平南县两景 CBERS-02B 影像为遥感数据源, 提取 CBERS-02B 影像 5 个波段、植被指数的灰度值; 利用研究区 1:10 000 地形图提取海拔、坡度、坡向等 GIS 因子; 以各遥感因子和 GIS 因子为自变量, 研究区样地蓄积量为因变量, 采用偏最小二乘法建立了森林蓄积量估测模型。结果表明: (1) CBERS-02B 遥感影像 5 个波段, 植被指数, 郁闭度、海拔高、坡向、坡度等与蓄积量相关性不显著, 且各因子之间存在明显的多重相关性; (2) 运用偏最小二乘法建立的森林蓄积量估测模型, 精度为 85.44%, 相对误差仅为 -0.232%。研究结果为科学、准确、高效地估测森林蓄积量提供了理论依据和有效方法。

关键词: 森林蓄积量; 估测模型; 遥感信息; 3S 技术; CBERS-02B 影像; 偏最小二乘法

中图分类号: S 758.5⁺ 1

文献标志码: A

文章编号: 1673-923X(2011)05-0080-06

Forest volume estimation model based on partial least square in Pingnan County, Guangxi Zhuang Autonomous Region

HONG Yi-feng, LIN Hui, YAN En-ping, LI Yong-liang

(Research Center of Forest Remote Sensing & Information Engineering, Central South
University of Forestry & Technology, Changsha 410004, Hunan, China)

Abstract: By taking the CBERS-02B image maps as the remote sensing (RS) resources, the gray values of five wave band and vegetation indexes were extracted; and elevation, slope aspect and other GIS factors were collected from the topographic map of forest resource inventory sample plots of Pingnan County in 2009; by means of partial least square, and adopting the RS and GIS factors as independent variables, and setting forest volume of sample plots in study region as dependent variable, the forest volume estimation model was established. The results show that (1) Forest volume was not correlated significantly to the factors such as five multi-spectral bands of CBERS-02B remote sensing image, vegetation index, canopy density, elevation, slope, aspect and so on, and there were obvious multiple correlations among the five factors; (2) The overall precision of forest volume estimation model set by Partial Least Square was 85.44%, and the relative precision merely was -0.232%. The study results provides a theoretical foundation and effective method for estimating forest volume scientifically, accurately and efficiently.

Key words: forest volume; estimation model; remote sensing information; 3S technology; CBERS-02B image map; partial least square

收稿日期: 2011-03-11

基金项目: 林业公益性行业科研专项“林分结构与生长模拟技术研究(201104028)”；国家自然科学基金资助项目(30871962)“森林树种波谱特及生化成分相关性研究”；高等学校博士学科点专项科研基金(200805380001)“南方乔木树种高光谱遥感测定与分析”；湖南省自然科学基金(07JJ3060)“高分辨率遥感图像森林信息智能识别技术研究”

作者简介: 洪奕丰(1986-), 男, 福建漳州人, 硕士研究生, 研究方向: 林业遥感与地理信息系统

通讯作者: 林 辉(1965-), 女, 湖北黄冈人, 教授, 博士生导师, 研究方向: 林业遥感与地理信息系统

森林是地球上最大的生态系统,而森林蓄积是表征森林数量最重要的指标之一^[1]。传统方法测定森林蓄积虽有一定精度,但耗费人力物力,且周期较长。利用 3S(RS、GIS、GPS) 技术进行森林蓄积的定量估测一直是林业遥感研究的重点和难点。近年来,随着 3S 技术的飞速发展和日益完善,及其在森林资源调查管理中应用的不断深入和普及,国内外已有众多学者进行了相关研究并取得重要成果。研究表明,遥感和 GIS 因子与森林长势、蓄积量、郁闭度等都存在较大的相关性,能有效反映森林植被生长和分布特征,可作为相关因子直接参与模型估测森林蓄积量^[2-3]。

笔者采用 CBERS-02B 高分辨率影像作为遥感数据源,结合平南县 2009 年森林资源二类调查数据,采用偏最小二乘方法估测森林蓄积量,结果表明该方法拟合效果较好,为科学、准确、高效估测森林蓄积量提供了理论依据。

1 研究区概况

平南县位于广西壮族自治区东南部,地处东经 110°03′54″~110°39′42″,北纬 23°02′19″~24°02′19″。东连藤县,南接容县,西邻桂平市,北与金秀、蒙山两县接壤。东西宽 30 多 km,南北长 110 km。属南亚热带季风气候区,夏长冬短,高温多雨,年平均气温 20.3℃,年平均降雨量 1 585.5 mm。土地总面积 298 800 hm²,其中有林地 145 786.2 hm²,占土地总面积的 57.77%。

2 数据源

(1) 研究的遥感数据采用两景中巴资源卫星(CBERS-02B)影像。获取时间分别为 2008 年 7 月 10 日和 2009 年 8 月 12 日,数据包括 B1、B2、B3、B4、B5 五个波段,空间分辨率:19.5 m×19.5 m,轨道号:3-73、2-74,中心纬度:23.949 668°、23.265 267,中心经度:110.112 558°、110.413 42°。利用 1:10 000 地形图及其他资料对影像进行几何校正、增强、镶嵌、裁剪等处理,校正误差控制在一个像元内。

(2) 平南县 1967 年的 1:10 000 地形图,2009 年森林资源二类调查数据。

3 研究方法

3.1 偏最小二乘法

偏最小二乘法是一种新型的多元统计分析方法,集多元线性回归分析、典型相关分析和主成分分析的功能和优点于一体,将建模的数据分析与非模型的数据认知方法有机地结合起来,使模型的精度、稳健性、实用性得到提高,已广泛应用于各领域的研究中^[4-5]。

设有 p 个自变量 $X = \{x_1, x_2, \dots, x_p\}$, q 个因变量 $Y = \{y_1, y_2, \dots, y_q\}$ 和 n 个样本观测点。偏最小二乘法即在 X 与 Y 中分别提取出成分 t_1 和 u_1 (t_1 是 x_1, \dots, x_p 的线性组合, u_1 是 y_1, \dots, y_q 的线性组合),提取时,为了符合回归分析的要求,必须满足两个条件: (1) t_1 和 u_1 应尽可能大地携带各自数据表中自变量和因变量的变异信息; (2) t_1 与 u_1 的相关程度最好达到最大,从而尽可能好地代表 X 和 Y ,使得 t_1 对 u_1 的解释能力最强。

当第一个成分 t_1 和 u_1 被提取后,分别实施 X 对 t_1 和 Y 对 u_1 的回归。若回归方程已经达到满意精度,则算法终止;否则,将利用 X 被 t_1 和 Y 被 u_1 解释后的残余信息进行第二轮成分提取。如此往复,直到能达到满意精度为止。若最终对 X 共提取了 m 个成分 t_1, \dots, t_m ,偏最小二乘法将通过实施 y_k 对 t_1, \dots, t_m 的回归,然后再表达 y_k 关于原变量 x_1, \dots, x_m 的回归方程($k = 1, 2, \dots, q$)。

对结果进行交叉有效性检验,即对提取的每一个成分计算预测残差平方和 $S_{PRESS,h}$ (Predicted Residual Sum of Squares)。将所有 n 个样本点分成两部分:第一部分除去某个样本点 i 的所有样本点集合(共含 $n-1$ 个样本点),用这部分样本点并使用 h 个成分拟合一个回归方程;第二部分是把刚才被排除的样本点 i 代入前面拟合的回归方程,得到 y_i 在样本点 i 上的拟合值 $Y_{h(-i)}$ 。对于每一个 $i = 1, 2, \dots, n$,重复上述测试,定义 Y 的预测误差平方和为 $S_{PRESS,h}$,如式(1)所示

$$S_{PRESS,h} = \sum_{i=1}^n (y_i - Y_{h(-i)})^2 \quad (1)$$

当 $SPRESS_{h-1} - SPRESS_{h-2}$ 小于某个阈值时, 迭代结束; 否则继续迭代。

3.2 森林蓄积量估测

3.2.1 样地设置

结合平南县 2009 年森林资源二类调查数据, 设置 250 个样地, 如图 1 所示(其中 200 个用于建模, 50 个用于模型适用性检验)。样地设置需满足以下条件: (1) 样地森林类型为纯林, 在一定范围内连续分布, 且存在蓄积; (2) 遥感图像上样地的灰度值应均匀; (3) 样地应尽量均匀分布, 覆盖整个研究区。

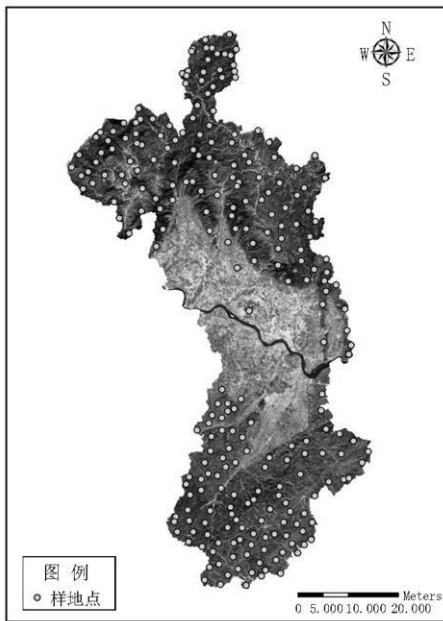


图 1 样地分布

Fig. 1 Distribution of sample plots

3.2.2 样地筛选

为减少样地数据对模型预测精度的影响, 本文运用标准差分析法剔除样本中离群值较大的数据, 即剔除样本中 $|x_{ij} - \bar{x}_j| > 2\sigma_j$ 的数据, 其中 σ_j 为变量 j 的样本标准差。重复上述步骤, 直到选满 250 个样本数据为止。

考虑到遥感因子和 GIS 因子之间量纲不同的影响, 将所有样本数据进行中心标准化, 统一量纲^[6], 如式(2)所示,

$$x'_{ij} = (x_{ij} - \bar{x}_j) / s_j \quad (2)$$

式(2)中: x'_{ij} 为中心标准化观测值, x_{ij} 为原始观测值, \bar{x}_j 为第 j 变量的均值, s_j 为第 j 变量的标准差。

3.2.3 变量选择

在估测模型中, 自变量的选取对模型可靠性、估测精度具有较大的影响, 故选取合适的自变量显得尤为重要。借助遥感手段估测森林蓄积量, 国内外学者已做过大量研究^[7-10]。本文以遥感和 GIS 为基础信息源估测森林蓄积量。

3.2.3.1 遥感因子提取

以样地 GPS 实测坐标为准, 在 ENVI 4.5 下提取 CBERS-02B 影像 5 个波段的灰度值, 作为森林蓄积量估测的可选因子。考虑到不同光谱通道所提供的遥感信息也不尽相同, 仅利用个别波段来提取植被信息是相当有限的, 故常将多光谱数据经线性或非线性组合, 产生对森林长势、分布情况有一定指示意义的植被指数^[11-12]。笔者利用 CBERS-02B 的 5 个光谱通道提取 4 种植被指数作为森林蓄积量估测的可选因子。

- 1) 差值植被指数: $I_{DWI} = B_4 - B_3$;
- 2) 比值植被指数: $I_{RVI} = B_4 / B_3$;
- 3) 归一化植被指数: $I_{NRVI} = (B_4 - B_3) / (B_4 + B_3)$;
- 4) 平方根植被指数: $I_{SQRT(RVI)} = \text{sqrt}(B_4 / B_3)$ 。

各指数计算式中, B_3 、 B_4 分别为 CBERS-02B 的红光、近红外波段的灰度值。

3.2.3.2 GIS 因子提取

利用 GIS 技术提取与森林蓄积量相关的 GIS 因子, 参与模型建立估测蓄积量是提高估测精度的有效方法。结合研究区 1:10 000 地形图, 在 ARC-MAP 下建立 DEM 模型, 生成基于栅格的海拔、坡向、坡度图层。根据样地 GPS 坐标, 分别提取对应的海拔、坡度、坡向值(其中: 阴坡与无坡向取值 0, 阳坡取值 1), 并将其设为定量因子。

3.2.3.3 其它因子提取

土地种类对森林蓄积量估测影响较大。二类调查中土地种类(只考虑一级子类)分有林地、疏林地、灌木林地、未成林造林地、苗圃地、无立木林地、宜林地、辅助生产林地、非林地等。考虑到模型自身特点, 为保证估测精度, 在 ENVI4.5 下, 将样地分有林地、非林地两类, 作为定性因子设置为可选变量。

森林郁闭度是反映森林生长状况的直接指标, 应作为重要因子考虑, 从地面调查数据中获取郁闭

度值, 设置为可选变量。

3.2.3.4 变量设置

本研究选取 200 个样本对应的遥感因子和 GIS

因子进行分析, 计算因子之间相关系数、各因子与蓄积量之间的相关性, 信息量及方差扩大因子, 计算结果如表 1、2、3 所示。

表 1 各因子与蓄积量之间的相关性检验

Table 1 Correlation inspection between factors and forest volume

蓄积量	海拔高	坡度	郁闭度	B_1	B_2	B_3	B_4	B_5	$(b_4 - b_3) / (b_4 + b_3)$	$(b_4 - b_3)$	(b_4 / b_3)	$\text{sqrt}(b_4 / b_3)$
1.000	-0.040	-0.065	0.031	-0.002	0.205	-0.020	0.089	0.042	0.207	0.215	0.211	0.211

表 2 各因子间系数的相关性检验

Table 2 Correlation coefficient inspection between factors

各因子	$\text{sqrt}(b_4 / b_3)$	坡度	郁闭度	海拔高	B_1	B_3	B_5	B_2	B_4	$(b_4 - b_3) / (b_4 + b_3)$	(b_4 / b_3)	$(b_4 - b_3)$
$\text{sqrt}(b_4 / b_3)$	1.000	-0.040	-0.056	0.127	0.238	0.038	0.103	-0.114	-0.017	-0.180	-0.132	-0.186
坡度		1.000	-0.047	-0.110	0.043	0.162	-0.014	0.139	-0.068	0.111	0.032	-0.064
郁闭度			1.000	0.059	0.036	0.017	0.233	-0.052	-0.115	-0.103	0.105	0.031
海拔高				1.000	0.203	-0.173	0.628	0.106	-0.102	0.059	-0.109	-0.003
B_1					1.000	-0.076	0.041	-0.024	-0.138	-0.146	0.142	-0.119
B_3						1.000	-0.143	-0.170	-0.570	0.548	0.546	-0.498
B_5							1.000	0.036	-0.303	0.114	0.050	-0.076
B_2								1.000	0.109	0.071	0.020	-0.257
B_4									1.000	-0.169	-0.026	-0.103
$(b_4 - b_3) / (b_4 + b_3)$										1.000	0.352	-0.583
(b_4 / b_3)											1.000	-0.826
$(b_4 - b_3)$												1.000

表 3 各因子信息量及方差扩大因子的统计

Table 3 Statistics of information content and variance expand factors

各因子	均值	方差	标准差	信息熵	V_{VIF}
海拔高	276.55	175.974	30966.679		2.724
坡度	29.92	10.286	105.806		1.402
郁闭度	0.491	0.1932	0.037		1.226
B_1	46.231	3.2474	10.546	-220.061	2.936
B_2	150.859	25.6064	655.687	37.987	11.272
B_3	58.9808	4.45011	19.804	-214.106	19.814
B_4	66.957	3.9231	15.390	-225.603	15.371
B_5	76.333	4.1534	17.250	-226.443	7.899
$(b_4 - b_3) / (b_4 + b_3)$	0.428067	0.0746149	0.006	24.050	30.435
$(b_4 - b_3)$	92.517333	23.4736744	551.013	37.665	169.126
(b_4 / b_3)	2.576667	0.3987256	0.159	49.694	95.273
$\text{sqrt}(b_4 / b_3)$	1.598667	0.1346069	0.018	45.456	26.416

分析表 1 可知, 遥感和 GIS 因子与蓄积量之间的相关程度不高, 表明前者对后者的独立解释能力不强; 由表 2 可知, 因子间均存在不同程度的相关性, 其中, I_{DVI} 与 I_{RVI} 、 I_{NRVI} 与 I_{DVI} 、海拔高与 B_5 、 B_3 与其他因子的相关性尤为显著; 表 3 可知, B_2 、 B_3 、 B_4 、 I_{NRVI} 、 I_{DVI} 、 I_{RVI} 、 $I_{SQRT}(RV)$ 的 V_{VIF} 值均大于 10, 表明这些因子间存在严重的多重相关性, 其他因子间也存在不同程度的多重相关性。综上, 一般回归模

型已不适用。

本文中, 采用偏最小二乘法建立森林蓄积量估测模型, 考虑到该模型自变量的选择门槛不高, 无需最优, 且较多变量有利于对提取的主成分进行累计解释能力分析。因此, 本文将所有变量引入, 参与森林蓄积量估测模型的建立。

3.2.4 模型构建

从筛选后的 200 个样地数据中提取相关遥感和

GIS 因子, 采用偏最小二乘法建立蓄积量估测模型 (该过程运用自编程序在 Matlab 中完成)。在进行模型交叉有效性检验时, 将迭代结束的阈值设定为 0.05 ($r = 0.05$), 即表明当 h 大于某个阈值时, 新增加的成分 t_h 对减少模型误差已无明显改善作用。最后得到的蓄积量估测模型如式 (3) 所示:

$$\begin{aligned}
 V = & 63.4595 - 0.0007x_1 - 0.0127x_2 + \\
 & 0.4713x_3 + 1.4274x_4 - 0.0508x_5 + \\
 & 0.0069x_6 - 0.0368x_7 - 0.0280x_8 + \\
 & 0.0375x_9 + 2.5340x_{10} + 0.0082x_{11} + \\
 & 0.5342x_{12} - 1.5466x_{13}. \tag{3}
 \end{aligned}$$

式(3)中: V 为蓄积量, x_1 为海拔高, x_2 为坡度, x_3 为坡向, x_4 为郁闭度, x_5 为 B_1 , x_6 为 B_2 , x_7 为 B_3 , x_8 为 B_4 , x_9 为 B_5 , x_{10} 为 $INRVI$, x_{11} 为 $IDVI$, x_{12} 为 $IRVI$, x_{13} 为 $ISQRT(RVD)$ 。

3.2.5 适用性评价

将未参与建模的 50 个样地数据代入估测模型, 计算蓄积量估测值, 将预测值与实测值进行对比 (如图 2 所示), 结果表明该模型拟合效果较好, 精度达 85.44%; 进一步分析计算其残差, 具体分布如图 3 所示, 分析该图可知, 残差值分布均匀, 平均相对误差为 -0.232%, 表明适用性检验效果较好。

$$A_{mf} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n \bar{y}_i^2} \times 100\% \tag{4}$$

式(4)中: A_{mf} 为模型拟合精度; y_i 为实际测量值, \hat{y}_i 为模型预估值, \bar{y}_i 为模型预估平均值, n 为自由度, T 为模型参数个数, t_{α} 为置信水平为 α 时的 t 分布值。

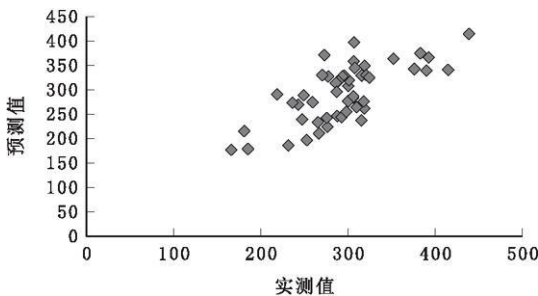


图 2 估测值与实测值对比

Fig. 2 Comparison of estimated values and measured values

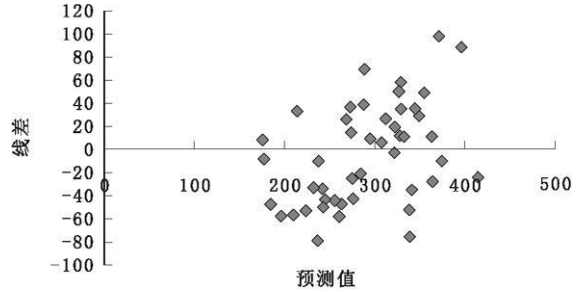


图 3 残差分布

Fig. 3 Distribution of residual errors

4 结论与讨论

综上所述, 本研究取得了如下结论:

(1) 通过分析遥感和 GIS 因子与森林蓄积量之间的相关性可知, 前者与后者之间的相关程度不高, 且前者之间存在明显的多重相关性, 一般回归模型不适用。

(2) 通过采用偏最小二乘法对研究区遥感与 GIS 因子进行拟合, 建立森林蓄积量估测模型, 结果表明: 1) 该模型精度、稳健性、适用性等均优于一般回归模型; 2) 在样本数据较少或因因子间相关性较高时, 该模型依然适用, 且精度较高。3) 模型自变量的解释能力高于一般回归模型; 4) 模型自变量的选择门槛不高, 无需最优, 且变量较多有利于对提取的主成分进行累计解释能力分析。5) 在多对多情况下, 该模型同样适用。

(3) 建立了森林蓄积量估测模型:

$$\begin{aligned}
 V = & 63.4595 - 0.0007x_1 - 0.0127x_2 + \\
 & 0.4713x_3 + 1.4274x_4 - 0.0508x_5 + \\
 & 0.0069x_6 - 0.0368x_7 - 0.0280x_8 + \\
 & 0.0375x_9 + 2.5340x_{10} + 0.0082x_{11} + \\
 & 0.5342x_{12} - 1.5466x_{13}.
 \end{aligned}$$

适应性检验表明该估测模型拟合效果较好。

(3) 本研究同时存在着如下不足和需要改进的地方:

(a) 研究只考虑遥感因子和主要 GIS 因子, 对立地条件、环境因子等尚未分析, 今后应加强这方面因子的引用。

(b) 通过增加自变量来估测蓄积量, 虽能得到满意效果, 但不能从根本上解决问题, 今后应作进一

步探讨。

致谢: 衷心感谢中国资源卫星应用中心提供的两景研究区 CBERS-02B 遥感影像数据。

参考文献:

- [1] 程学武, 杨存建. 森林蓄积量遥感定量估测研究综述[J]. 安徽农业科学, 2009, 37(16): 7746-7750.
- [2] 赵宪文, 李崇贵, 斯林, 等. 森林资源遥感估测的重要进展[J]. 中国工程科学, 2001, (8): 15-28.
- [3] 李亦秋, 冯仲科, 邓欧, 等. 基于 3S 技术的山东省的森林蓄积量估测[J]. 林业科学, 2009, (9).
- [4] 陈楚, 关泽群, 张鹏林, 等. 利用 RS 和 GIS 的森林蓄积量偏最小二乘估测研究[J]. 湖北林业科技, 2004, (4): 25-28.
- [5] 罗批, 郭继昌, 李锐, 等. 基于偏最小二乘回归建模的探讨[J]. 天津大学学报, 2002; 35(6): 783-786.
- [6] 据存勇, 蔡体久. 用泛化改进的 BP 神经网络估测森林蓄积量

[J]. 林业科学, 2006, 12(42): 59-62.

- [7] 李崇贵, 赵宪文, 李春干, 等. 森林蓄积量遥感估测理论与实现[M]. 北京: 科学出版社, 2006: 15-26.
- [8] 朱海珍, 庞勇, 等. 基于 ENVISAT ASAR 数据的森林蓄积量估测研究[J]. 地理与地理信息科学, 2007, 23(12): 51-55.
- [9] 刘志华, 常禹, 陈宏伟. 基于遥感地理信息系统和人工神经网络的呼中林区森林蓄积量估测[J]. 应用生态学报, 2008, 19(9): 1891-1896.
- [10] 张宇, 谷建才, 等. 基于 RS 和 GIS 的径向神经网络模型对森林蓄积量的估测[J]. 浙江林业科技, 2009, 29(5): 51-54.
- [11] 杨存建, 刘纪远, 张增祥. 热带森林植被生物量遥感估算探讨[J]. 地理与地理信息科学, 2004, 20(6): 22-25.
- [12] 赵英时. 遥感应用分析原理与方法[M]. 北京: 科学出版社, 2003: 366-409.

[本文编校: 邱德勇]

(上接第 71 页)

法律规范中, 有关转基因安全的技术标准规范的内容几乎找不到, 有涉及到的往往也是蜻蜓点水一笔带过。转基因生物的安全性必须建立在转基因成分检测的结果上, 因此必须建立完善的转基因生物安全性检测标准, 包括制定转基因生物及其产品潜在风险的分析方法和技术体系, 并建立相应的风险管理技术规范、规程与方法, 这样才能保证各部门能够利用风险评估程序对转基因生物及其产品的运输、应用过程中可能产生的任何风险进行评估, 并根据评估结果采取相应的措施, 从而确保转基因生物及其产品的安全。

转基因生物技术是人类有史以来发展最快的技术, 它在带来巨大利益的同时可能也会存在潜在风险。在转基因生物安全监管的立法上, 我们必须结合国情将国外先进的生物安全法律法规的核心内容

选择性地融入到我国生物安全法规中, 顺应国际生物安全的发展形势, 使我国的生物安全管理有法可依, 保障我国生物技术的健康发展。

参考文献:

- [1] 罗怀熙. 关于美国农业转基因法律制度的考察报告[ED/OL]. 中国政府法制信息网, <http://www.chinalaw.gov.cn/article/dfx/x/zfzyj/200309/20030900060266.shtml>.
- [2] 于文轩, 王灿发. 国外生物安全立法及对中国立法的思考[J]. 科技与法律, 2005, (4): 11-12.
- [3] 国务院发展研究中心. 发达国家管理转基因生物安全的几种模式[J]. 中国农业信息快讯, 2002, (9): 16.
- [4] EC Directive 90/220 on the Deliberate Release of GMOs to the Environment.

[本文编校: 邱德勇]