

北部湾沿海 5 种红树林群落生物量的遥感估算^{*}

Biomass Estimation of Five Kinds of Mangrove Community in Beibu Gulf Based on Remote Sensing

曹庆先^{1,2,3}, 徐太平², 鞠洪波³

CAO Qing-xian^{1,2,3}, XU Da-ping², JU Hong-bo³

(1. 广西红树林研究中心, 广西红树林保护重点实验室, 广西北海 536000; 2. 中国林业科学研究院热带林业研究所, 广东广州 510520; 3. 中国林业科学院资源信息研究所, 北京 100091)

(1. Guangxi Mangrove Research Center, Guangxi Key Lab for Mangrove Conservation, Beihai, Guangxi, 536000, China; 2. The Institute of Tropical Forest, Chinese Academy of Forestry, Guangzhou, Guangdong, 510520, China; 3. The Institute of Forest Resource Techniques, Chinese Academy of Forestry, Beijing, 100091, China)

摘要: 采用广西海南两地 2004 年红树林 TM 影像和红树林群落样地调查数据, 应用多元逐步回归分析方法, 选取影像光谱特征及纹理特征等作为自变量, 对红海榄 (*Rhizophora stylosa*)、木榄 (*Bruguiera gymnohiza*)、白骨壤 (*Vicennia mariana*)、桐花树 (*Aegiceras corniculatum*)、秋茄 (*Kandelia candel*) 以及混合(不分树种)红树林生物量遥感估算建模, 研究红树林生物量的遥感估算方法。结果, 生物量估算模型拟合效果红海榄> 木榄> 白骨壤> 桐花树> 混合, 秋茄模型无法拟合。回归过程中采用稳健诊断方法去除影响点, 并应用聚类分析和因子分析的方法排除多重共线性。得到的生物量估算模型通过相关检验, 可以高效、快速地进行红树林生物量的估测。

关键词: 生物量 估算 纹理特征 多元回归分析

中图法分类号: Q948, S757.2 文献标识码: A 文章编号: 1005-9164(2011)03-0289-05

Abstract: In order to evaluate the biomass of mangrove based on remote sensing, the biomass model of mangrove is estimated with the multiple regression analysis, by extracting spectral information and textural features from TM images, and combining the field survey biomass data. The results show that the accuracy order of biomass estimation model is *Rhizophora stylosa* > *Bruguiera gymnohiza* > *Vicennia mariana* > *Aegiceras corniculatum* > mixed species, and that *Kandelia candel* cannot be modeled. During the process of multiple regression analysis, the effect point is removed by ellipsoidal multivariate trimming and the multicollinearity is eliminated by cluster analysis and factor analysis. All the biomass estimation models got through relevant test and can be applied on the biomass estimation of mangrove.

Key words: biomass, estimation, texture feature, multiple regression analysis

红树林生态系统是地球上生产力最高的生态系统之一^[1], 它在碳吸收和储备方面的作用不容忽视。以前的研究发现, 红树林和热带雨林一样高产, 是地

球碳循环的一个至关重要的部分, 尽管它们只覆盖了地球表面 0.1% 的面积, 却固定了相当于大气中 5% 的碳^[2]。

目前, 基于遥感影像的红树林生物量估测研究较少, 一方面红树林属于潮汐类沼泽, 并且其分布区一般多云、雨、雾等, 较难获得大面积红树林的清晰遥感影像, 且在研究过程中, 要排除海水上升和下降的影响; 另一方面, 红树林湿地分布于淤泥滩地, 地面调查极其艰难, 而建模又需要大量实测样地数据的支持;

收稿日期: 2011-05-25

修回日期: 2011-06-27

作者简介: 曹庆先(1981-), 女, 博士研究生, 主要从事生态学 3S 技术应用研究。

^{*} 广西科学院科研专项(08YJ16HS01), 广西壮族自治区科学技术厅项目(桂科攻 1140002-2-3)资助。

再者,红树林生境随潮汐变化,又只能在河口淡水流经区生长,难以测得稳定的生境特征值;其分布区地形平坦开阔,没有较大变化,地形因子等又不适合作为研究变量。黎夏、叶嘉安等人^[3]研究了红树林湿地植被生物量的雷达估算方法,但目前这一技术的应用有其局限性,在我国红树林研究工作中所占的比例极少。

纹理是复杂的视觉实体或者子模式的组合,有亮度、色彩、陡度、大小等特征^[4],可以帮助抑制异物同谱、同物异谱现象的发生^[5]。2003年 Dengsheng Lu 应用 TM 影像研究热带森林发现,应用纹理特征值结合波段值或植被指数来估测森林生物量时,其精度比单独应用遥感影像的波段值或植被指数精度有较大提高,取得较好的效果^[6]。2006年,李明诗等^[7]对 SPOT5 HRG 影像进行纹理分析,结合研究区地形特征,分别对 5 个森林类型实现其生物量估算模型辨识与验证,结果发现纹理特征值在回归方程中有较大贡献。

Haralick^[8]定义了 14 种纹理特征。最常用的用于提取遥感图像中纹理信息的特征统计量主要有:均值(Mean)、方差(Var)、均匀性(Hom)、对比度(Contrast)、相异性(Dis)、熵(Entropy)、二阶矩(Second Moment)、灰度相关(Correlation)等。纹理特征值应用于生物量估测的研究却很少。应用其进行红树林湿地生物量的研究目前还没有文献报道。本研究即是基于 TM 影像纹理与光谱特征进行红树林生物量的估算。

1 研究方法

1.1 地面调查

研究样地分布于广西铁山港湾、钦州湾、珍珠湾、廉州湾、海南东寨港。共有 126 个样地,其中木榄 [*Bruguiera gymnorhiza* (Linn.) Savigny] 群落(即木榄为优势种,下同)样地 21 个,红海榄 (*Rhizophora stylosa* Griff.) 群落样地 16 个,秋茄 [*Kandelia candel* (Linn.) Druce] 群落样地 23 个,桐花树 [*Aegiceras corniculatum* (Linn.) Blanco] 群落样地 38 个,白骨壤 [*Avicennia marina* (Forsk.) Vierh.] 群落样地 28 个。所有样地进行每木检尺,并测得覆盖度、叶面积指数、土壤盐度等,用 GPS 记录样地中心点坐标。样地调查时间为 2004~2008 年。红树林叶生物量及地上部分总生物量的计算,采用广西大学林学院温远光老师的广西英罗港 5 种红树植物群落的生物量方程^[9]。

1.2 数据处理与信息提取

1.2.1 数据预处理

广西西段、广西东段、海南东寨港 3 副 TM 影像均为 2004 年 10 月拍摄。TM 影像以 WGS84 地理坐标的 1:10000 矢量地形图为基准进行校正,校正模型采用三次多项式,校正误差控制在 0.5 个像元以内,采用最近邻插值方法^[10]保留原始影像像素值。应用红树林矢量分布图裁切遥感影像,得到研究区红树林分布范围内的遥感影像。切除被水淹没红树林区影像(红树林矢量分布范围内 $NDVI \leq 0$ 的区域)。把用来校正遥感影像的数据对 126 个样地点做仿射变换,减少样地点坐标与相应影像的匹配误差。

1.2.2 遥感影像信息提取

利用 ArcGIS 的 Sample 工具,提取样地点所在像元及其周围 4 个像元的像素值,取 5 个值的平均做为该样地遥感信息源,这样虽然会带来一定误差,但会使样地数据距离其实际值的波动更小。所提取的遥感信息主要包括:影像波段值、植被指数、纹理特征值三部分。

(1) 波段值提取。提取的 TM 影像单波段的数据和其派生数据有: TM 1、TM 2、TM 3、TM 4、TM 5、TM 6、TM 7、TM 1/TM 2、TM 1/TM 3、TM 2/TM 3、TM 1/TM 4、TM 3/TM 4、TM 2/TM 4、TM 1/TM 5、TM 2/TM 5、TM 3/TM 5、TM 4/TM 5、TM 1/TM 7、TM 2/TM 7、TM 3/TM 7、TM 4/TM 7、TM 5/TM 7、 $TM(4+5-2)/TM(4+5+2)$ 、 $TM(4 \times 3)/7$ 、 $TM 3/TM(1+2+3+4+5+7)$ 、 $1/TM 1$ 、 $1/TM 2$ 、 $1/TM 3$ 、 $1/TM 4$ 、 $1/TM 5$ 、 $1/TM 7$, 共 31 个因子。

(2) 植被指数、缨帽变换和主成分变换指数提取。比值植被指数 $RVI = TM 4 / TM 3$, 归一化植被指数 $NDVI = (TM 4 - TM 3) / (TM 4 + TM 3)$, 环境植被指数 $DVI = TM 4 - TM 3$, 调整土壤亮度的植被指数 $SAVI = (1 + 0.5) \times (TM 4 - TM 3) / (TM 4 + TM 3 + 0.5)$, 垂直植被指数 $PVI = 0.939(TM 4) - 0.344(TM 3) + 0.09$ 。TM 3、TM 4 分别表示 TM 影像第三、第四波段灰度值。缨帽变换(KT)提取亮度指数(BRIGHT)、绿度指数(GREEN)、湿度指数(WET)指数值,主成分变换(PC)前 3 个主成分 PC1、PC2、PC3 包含了 TM 影像 95% 以上的信息,取 PC1、PC2、PC3 作为研究变量。

(3) 纹理特征提取。采用基于灰度共生矩阵(GLCM)的方法分别提取 TM 影像 7 个波段常用的 8 种纹理特征值,选择步距为 1, 3×3 窗口, 0° 、 45° 、 90° 、 135° 4 个方向的平均值,共 56 个值作为研究变量。

1.3 模型构建

模型构建流程如图 1 所示。

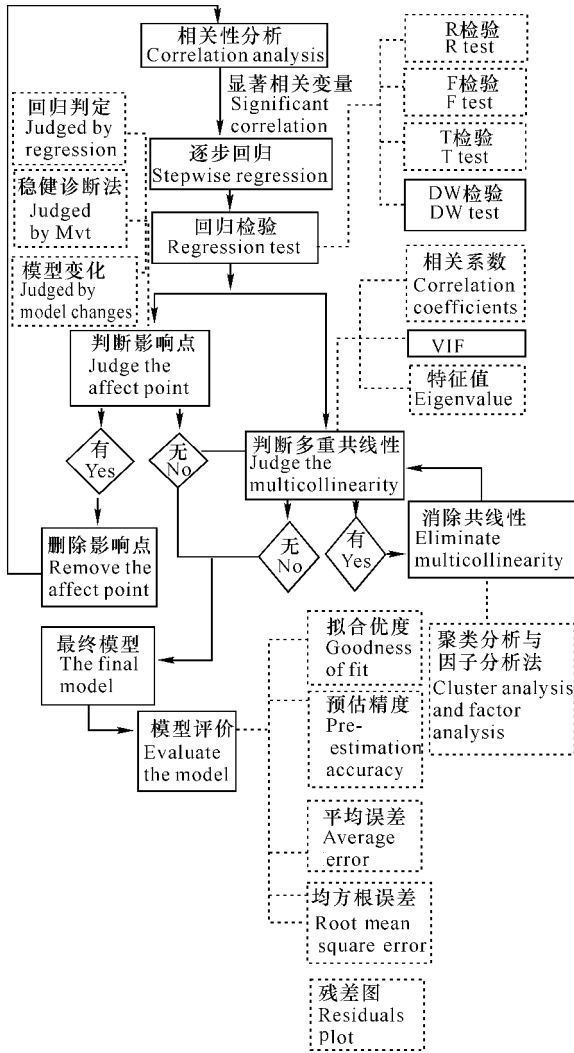


图 1 模型构建流程

Fig. 1 Flow chart of model building

1.3.1 变量筛选

信息提取的自变量多达 101 个, 经回归分析, 与因变量显著相关的变量才有资格进入模型。由于样地覆盖度是影响生物量与其它变量相关关系的主要因素, 所以, 相关分析是在控制覆盖度的前提下, 求算生物量与其它变量的相关系数, 即偏相关系数。

1.3.2 影响点判断

用回归判定、稳健诊断、模型变化检验 3 种方法进行影响点综合判定。

(1) 回归判定。在回归过程中, 识别影响点的方法有标准化残差、学生残差、学生删除残差、Mahalanobis 距离、中心点杠杆值、COOK 距离、协方差比等判断某一观测值是否为影响点, 往往需要使用多个指标进行综合判断, 只使用少数几个有可能得到矛盾的结果。

(2) 稳健诊断方法。当应用多个指标判定仍出现

矛盾结果时, 引入影响点的稳健诊断方法。即用 MVT (ellipsoidal multivariate trimming) 方法针对方差-协方差矩阵进行稳健诊断^[11, 12]。

步骤 1 应用回归模型中的变量为自变量, 计算除样本数据的均值 x^*

$$x^* = \frac{1}{n} \sum_{i=1}^n x_i$$

步骤 2 计算马氏距离平方和

$$d_i^2 = (x_i - x^*)' S^{-1} (x_i - x^*), i = 1, 2, \dots, n$$

步骤 3 d_i^2 值越大, 越有可能为影响点。如果具有最大 d_i^2 值的观测点, 是逐步回归中判定的奇异值, 则判定该点为影响点, 将其删除。

(3) 模型变化检验。前两种方法都无法确定影响点时, 就应用去掉影响点后模型会有较大变化这一特征进一步检验。

1.3.3 多重共线性的消除

在回归方程中, 虽然各自变量对因变量是有意义的, 但某些自变量彼此相关, 即存在共线性问题, 给评价自变量的贡献率带来困难。所以, 需要找到一种方法, 既能保留这些变量的大部分信息, 又能把这些引起多重共线性的变量转化为互相独立的变量的方法。聚类分析和因子分析相结合的方法正能满足要求^[13]。聚类分析首先对这些变量进行分类, 具有相似性的变量自成一类, 对于变量个数大于 1 的类再进行因子分析, 将多个变量转化为少数几个不相关的综合指标, 应用因子分析后的变量代替原始变量参与模型构建, 既不损失变量信息, 又能消除变量间多重共线性问题。

1.4 模型的评价方法

本研究仅对叶生物量遥感估算模型进行估测精度评价, 因为只有叶生物量模型是直接由遥感因子反演而得, 地上总生物量模型是根据与叶生物量内在的关系间接而得, 此处进行模型评价意义不大。

1.4.1 拟合优度

拟合优度是相关系数的平方 (R^2), 为回归平方和在总平方和中所占的比率, 体现了回归模型所能解释的因变量变异性的百分比。 R^2 越大, 说明模型拟和程度越好。

1.4.2 预估计精度

$$P = \left(1 - \frac{t_a \sqrt{\sum (w_i - \bar{w})^2}}{\bar{w} \sqrt{n(n-t)}}\right) \times 100\%$$

$$\bar{w} = \frac{1}{n} \sum w_i$$

w_i 为实测值, \hat{w}_i 为估计值, n 为样本容量, t_a 为置信水平为 $\alpha = 0.05$ 时的 t 分布值, t 为回归模型中参

数个数。预估计精度 P 越大, 说明模型净度越高。

1.4.3 均方根误差和平均误差

由于本研究样地较少, 精度评价在像元尺度采用交叉评价方法进行。应用基于遥感信息的叶生物量模型, 计算各样地预测生物量值, 根据生物量实测值和预测值, 计算均方根误差 ($RMSE$) 和平均误差 (e)。

$$RMSE = \frac{\sqrt{\sum_{i=1}^N (X_i - Y_i)^2}}{N} \times 100$$

$$e = \frac{\sum_{i=1}^N (X_i - Y_i)}{N}$$

X_i 为实测样地生物量, Y_i 为回归模型估测样地生物量。

2 结果与分析

2.1 生物量估测模型

根据以上步骤, 各树种分别以叶生物量、叶生物量平方根、叶生物量立方根、叶生物量常用对数、叶生物量自然对数、叶生物量倒数作为因变量, 和与其显著相关的遥感因子或环境变量进行逐步回归分析, 选取最优模型如表 1, 秋茄未得到模型。

表 1 红树林叶生物量遥感估测模型

Table 1 Biomass estimation model of mangrove leaves by remote sensing

树种 Specie	模型 Model	R	R^2	F	$Sig.$
红海榄 <i>R. stylosa</i>	$Bio_{\text{叶}} = (12.016 \times \text{dis}5 + 14.505)^2$	0.860	0.739	14.2000	0.005
木榄 <i>B. gym-norrhiza</i>	$Bio_{\text{叶}} = (144.560 \times \text{Hom}7 - 29.176)^2$	0.803	0.644	16.3000	0.003
桐花树 <i>A. corn-iculatum</i>	$Bio_{\text{叶}} = (-0.175 \times \text{Var}7 + 7.095)^3$	0.685	0.470	14.6200	0.000
秋茄 <i>K. candel</i>	未能得到 Failure				
混合 * Mixed Species	$Bio_{\text{叶}} = 90.410 \times \text{Band}6 + 78.819 \times \text{Fac}1 - 10969.704$ $\text{Fac}1 = -0.006 \times \text{Hom}2 + 0.069 \times \text{Mean}7 + 0.265 \times \text{Mean}3 + 0.680 \times \text{Mean}2$	0.664	0.441	31.1640	0.000

*混合是指 5 种群落的样地数据放在一起进行生物量的估算。Bio 单位: 千克/像元。以下的表 2 和表 3 与此相同。Mixed species is to evaluate biomass by mixed sample data of five kinds of community types. The unit of biomass in the table is kg/pixel. The following Table 2 and Table 3 are the same.

根据地上部分总生物量与叶生物量样地数据进行逐步回归, 得到通过叶生物量求算地上总生物量的

模型, 见表 2。

表 2 红树地上总生物量与叶生物量关系模型

Table 2 Biomass estimation model of mangrove above ground from leaves biomass

树种 Specie	模型 Model	R^2	F	$Sig.$
红海榄 <i>R. stylosa</i>	$Bio_{\text{总}} = 357.308 + 7.675 \times Bio_{\text{叶}} + 0.001 \times Bio_{\text{叶}}^2$	0.977	149039.000	0.000
木榄 <i>B. gym-norrhiza</i>	$Bio_{\text{总}} = -539.925 + 10.767 \times Bio_{\text{叶}}$	0.999	117254.000	0.000
白骨壤 <i>A. marina</i>	$Bio_{\text{总}} = 11.53 \times Bio_{\text{叶}} - 345.325$	0.999	12671.000	0.000
桐花树 <i>A. corn-iculatum</i>	$Bio_{\text{总}} = 1.517 \times Bio_{\text{叶}}^{1.236}$	0.938	514.000	0.000
混合 Mixed species	$Bio_{\text{总}} = -634.687 + 10.946 \times Bio_{\text{叶}}$	0.993	13037.000	0.000

2.2 模型评价

基于像元尺度, 计算混合树种及四个树种叶生物量遥感估算模型的拟合优度、预估计精度、均方根误差和平均误差, 结果如表 3 所示。

表 3 各模型评价指标

Table 3 Accuracy evaluating indicators of models

树种 Specie	实测生物量 Measured biomass (t/hm ²)	预测生物量 Predicted biomass (t/hm ²)	平均误差 Average error (t/hm ²)	均方根误差 Root mean square error	拟合优度 Goodness of fit R^2	预估计精度 Prediction accuracy Pre-estimation accuracy
混合 Mixed species	18 223	18 730	0 510	0 710	0 438	0 780
桐花树 <i>A. corn-iculatum</i>	5 439	5 311	-0 130	0 590	0 470	0 790
白骨壤 <i>A. marina</i>	25 589	25 589	0 000	0 480	0 607	0 822
木榄 <i>B. gym-norrhiza</i>	49 683	48 267	-1 420	0 340	0 644	0 877
红海榄 <i>R. stylosa</i>	13 893	14 359	0 470	0 220	0 739	0 900

由表 3 分析可得: (1)应用多元逐步回归方法, 秋茄未能得到其生物量估算模型, 可能与其光谱反射特征有关, 秋茄在近红外波段起反射率明显低于其它红树^[14]。而 TM4 760~960nm 近红外波段对绿色植物最敏感, 为植物通用波段, 用于作物生物量测定、长势测量及水域判别等。另外也许和秋茄本身特性、样地的选择或者实地调查误差等其它因素有关。(2)5 个模型平均误差都较低, 最小的为 0t/hm², 最大的也仅为 1.42 t/hm²。说明 5 个模型的平均误差都很小, 模型拟合效果较好。(3)除秋茄外, 其它任何一种红

树的生物量遥感估算模型拟合优度和预估计精度都远高于混合树种生物量遥感估算模型,均方根误差远低于混合模型,这说明单树种生物量遥感反演效果要优于混合树种。红树群落的纯度是影响生物量遥感估算精度的主要因子。(4)从拟合优度 R^2 、均方根误差、估测精度三个指标判断,拟合效果优劣的排序是一致的,即红海榄>木榄>白骨壤>桐花树>混合。混合不分树种的生物量回归模型拟和效果最差,低于任何一个单纯树种,其拟合优度最小,估测精度最低、均方根误差最大。

3 讨论

本研究采用广西海南两地的 2004 年红树林 TM 影像和红树林群落样地调查数据,应用多元逐步回归分析方法,选取影像光谱特征及纹理特征等作为自变量,对红海榄、木榄、白骨壤、桐花树、秋茄以及混合不分树种红树林生物量进行反演,得到的结论是:模型拟合效果红海榄>木榄>白骨壤>桐花树>混合,秋茄未能得到拟合模型。虽然得到的 5 个模型之间存在一定的误差,但是 5 个模型通过相关检验,像元尺度上平均误差很小,预估计精度较高,不失为一种快速、高效估算红树林湿地生物量的一种手段。

5 个模型之间存在一定误差的原因可能在于:
(1)由于红树林分布区清晰的遥感影像难以获得,地面调查数据和遥感影像数据有一定的时间差,对回归结果有一定的误差影响。(2)红树林生长于泥滩地,地面调查比较艰难,样地面积小于一个像元,这也是误差存在的一个重要因素。(3)样地分布于广西海南两地,但地面生物量方程都是采用广西英罗港 5 种红树的生物量方程,对海南样地生物量计算有一定影响。个别树种因无法找到其地上生物量计算方程,只能采用与其相近树种的方程,使得地面生物量调查数据也存在一定误差。(4)遥感影像校正误差的存在也是影响预测精度的一个重要因子。

本研究先得到树叶生物量的遥感估测模型,再通过树叶与树木地上部分的内在关系进而推算地上部分总生物量的计算方法具有一定的现实意义。本研究采用的是多元逐步回归分析方法,下一步的工作将尝试应用神经网络、KNN 等非参数估计估计方法进行预测,以提高估测精度。

另外很少有人将纹理特征应用于生物量遥感估测,本研究应用了大量的纹理特征值,结果显示,生物

量与归一化植被指数 NDVI、LAI、TM3、TM4 等常用生物量遥感估测自变量基本不相关,与多个纹理特征存在较好的相关性,其内在的机理还有待于进一步研究。

参考文献:

- [1] 范航清.红树林——海岸环保卫士[M].南宁:广西科学技术出版社,2000.
- [2] 汤天波.红树林在碳循环方面的作用不容忽视[EB/O L].(2009-05-15)[2010-03-01].http://www.aims.gov.au/docs/media/news2009/20090422.html.
- [3] 黎夏,叶嘉安,王树功.红树林湿地植被生物量的雷达遥感估算[J].遥感学报,2006,10(3):5.
- [4] Rosenfeld A, Kak A. Digital picture processing, 2nd edition[M]. Academic Press, 1982.
- [5] Lee J, Philpot W. Spectral textures pattern matching: A classifier for digital imagery[J]. IEEE Transactions on Geosciences and Remote Sensing 1991, 29: 545-548.
- [6] Lu D S. Estimation of forest stand parameters using landsat TM images in the Brazilian Amazon Basin[M]. International Symposium on Remote Sensing of Environment; Information for Risk Management and Sustainable Development, Honolulu, HI, US, 2003.
- [7] 李明诗,谭莹,潘洁.结合光谱、纹理及地形特征的森林生物量建模研究[J].遥感信息,2006(6):6-9.
- [8] Haralick R M. Statistical and structural approaches to texture[J]. Proceeding of the IEEE, 1979, 67: 786-804.
- [9] 温远光.广西英罗港 5 种红树植物群落的生物量和生产力[J].广西科学,1999,6(2):142-147.
- [10] 方红亮,张健挺,刘卫国. ERDAS 遥感图像处理教程[M].北京:中国科学院地理研究所资源环境信息系统国家重点实验室,1998.
- [11] Devlin J S, Gnanadesikan R, Kettenring J R. Robust estimation of dispersion matrices and principal components[J]. Journal of the American Statistical Association, 1981, 76: 354-362.
- [12] Rousseeuw P J, Leroy A. Robust regression and outlier detection[M]. New York: John Wiley, 1981.
- [13] 林乐义,印凡成.基于聚类分析和因子分析消除多重共线性的方法[J].统计与决策,2008(8):153-155.
- [14] 翁强.红树植物光谱特征和红树植物的遥感图像分类研究[D].厦门:厦门大学,2006.

(责任编辑:邓大玉)