

广义模型及分类回归树在物种分布模拟中的应用与比较

曹铭昌^{1,2}, 周广胜^{1*}, 翁恩生¹

(1. 中国科学院植物研究所植被数量生态学重点实验室, 北京 100093; 2. 中国科学院研究生院, 北京 100039)

摘要:比较 3 个应用较广的模拟物种地理分布模型: 广义线性模型(GLM)、广义加法模型(GAM)与分类回归树(CART)对中国树种地理分布模拟的优劣, 以提出更为合适的模拟物种地理分布模型, 并用于预测气候变化对物种地理分布的影响。3 个模型对中国 15 种树种地理分布的模拟研究表明: 除对油松、辽东栎分布的模拟精度稍差外, 对其余树种分布的模拟精度均较高, 其中以 GAM 模型最好。结合地理信息系统(GIS), 比较分析了这 3 个模型对青冈、木荷、红松和油松 4 种树种的地理分布模拟效果, 结果亦表明: 这 3 个模型均能很好模拟青冈和木荷的地理分布, 而 GLM 模型对红松分布的模拟结果不太理想, 3 个模型对油松分布的模拟结果均不甚理想, 其中以 GLM 模型最差。基于 3 个模型对未来气候变化下青冈与蒙古栎地理分布的预测表明: GLM 模型与 GAM 模型对青冈分布的预测结果较为接近, 青冈在未来气候变化情景下向西和向北扩展, 而 CART 模型预测青冈在未来气候变化情景下除有向西、向北扩展趋势外, 广东和广西南部青冈分布区将消失; 3 个模型均预测蒙古栎在未来气候变化情景下向西扩展, 扩展面积的大小为: GAM 模型的模拟面积>GLM 模型>CART 模型。

关键词:物种分布; 广义线性模型; 广义加法模型; 分类回归树

文章编号: 1000-0933(2005)08-2031-10 中图分类号: Q141 文献标识码: A

Application and comparison of generalized models and classification and regression tree in simulating tree species distribution

CAO Ming-Chang^{1, 2}, ZHOU Guang-Sheng^{1*}, WENG En-Sheng¹ (1. Laboratory of Quantitative Vegetation Ecology, Institute of Botany, Chinese Academy of Science, Beijing 100093, China; 2. Graduate School of Chinese Academy of Sciences, Beijing 100039, China). *Acta Ecologica Sinica*, 2005, 25(8): 2031~2040.

Abstract: Comparison of three models: Generalized Linear Models (GLM), Generalized Additive Models (GAM) and Classification and Regression tree (CART) in simulating tree species distribution in China were done in this paper, in order to select a suitable model for simulating and predicting tree species distributions under climate change in the future. The simulating 15 tree species distributions by three models indicated that three models could simulate the tree species geographical distributions very well except *Pinus tabulaeformis* and *Quercus liaotungensis*. GAM model is the best one among them. Combining with geographical information system (GIS), the simulating effects for geographical distributions of *Cyclobalanopsis glauca*, *Schima superba*, *Pinus koraiensis*, and *Pinus tabulaeformis* were compared. The results showed that three models could simulate the geographical distributions of *Cyclobalanopsis glauca* and *Schima superba* very well; GLM model could not simulate the geographical distribution of *Pinus koraiensis* very well; and three models could not simulate the geographical distribution of *Pinus tabulaeformis*. The relative performance of different models was discrepant among species, suggesting that the most accurate model varies between species. Moreover, the geographical distributions of typical tree species: *Cyclobalanopsis glauca* and *Quercus mongolica* under climate change were studied based on GLM model, GAM model and

基金项目:国家自然科学基金资助项目 (49905005, 40231018)

收稿日期: 2004-11-12; **修订日期:** 2005-05-10

作者简介: 曹铭昌 (1979~), 男, 江西安福人, 硕士, 主要从事全球生态学研究. E-mail: caomingc@ibcas.ac.cn

* **通讯作者** Author of correspondence. E-mail: zhougs@pubilc2.bta.net.cn

Foundation item: National Natural Science Foundation of China (No. 49905005, 40231018)

Received date: 2004-11-12; **Accepted date:** 2005-05-10

Biography: CAO Ming-Chang, Master, mainly engaged in global ecology. E-mail: caomingc@ibcas.ac.cn

CART model. The simulated results of *Cyclobalanopsis glauca* geographical distributions by GLM model and GAM model were similar, and showed its geographical distribution would shift to the west and the north under climate change, however CART model indicated that the *Cyclobalanopsis glauca* geographical distributions in Southern Guangdong province and southern Guangxi province would disappear. The simulated results of *Quercus mongolica* geographical distributions by three models were similar, and they would shift to the west, and the expanding areas ranged from GAM model, GLM model to CART model.

Key words: species distribution; generalized linear models; generalized additive models; classification and regression tree

物种分布与环境的关系一直以来都是生态学研究中的重点问题。近年来,随着统计技术和地理信息系统(Geographic Information System,GIS)的发展,用于预测物种分布的模型技术也得到了迅速的发展和提高。模型技术已被广泛应用于生物地理分布^[1~9]、生物多样性^[10~14]和气候变化影响评估^[15~18]等方面。

在描述植物分布与环境关系的众多模型中,广义线性模型(generalized linear models,GLM))在预测物种分布中显示出较强的预测能力^[19,20],因而被广泛应用。然而,广义线性模型不能处理复杂的响应关系^[21],为此产生了广义加法模型(generalized additive models,GAM)^[22,23]。随着计算机技术的发展,基于规则的模型技术,如回归分类树(classification and regression tree,CART)^[24~26]和人工神经网络(artificial neural network,ANN)^[27,28]等方法也已经被应用于物种分布研究。

我国植被茂盛,物种丰富,植被、物种分布与环境的关系一直是气候—植被关系研究的重点。关于中国植被、物种与环境的关系已有大量研究^[29~34],但应用广义线性模型、广义加法模型、分类回归树和人工神经网络等模型技术对我国物种分布及其对气候变化响应的研究还未见报道。这些模型技术能够应用环境变量预测物种分布,定量分析物种分布与环境因子的关系,预测环境变化对物种分布的潜在影响,有助于理解我国物种的分布及其未来发展趋势,对制定减缓气候变化影响的对策具有重要的指导意义。基于这一考虑,本文在全国尺度上,选取我国 15 种常见树种,比较分析广义线性模型、广义加法模型和分类回归树 3 种模型技术对中国树种地理分布模拟的优劣,以提出更为合适的模拟物种地理分布模型,用于预测气候变化对物种地理分布的影响;同时,亦希望借此为研究我国物种分布提供新思路。

1 材料和方法

1.1 树种分布数据

用于模型比较的 15 个树种为青冈(*Cyclobalanopsis glauca*)、水青冈(*Fagus longipetalata*)、甜槠(*Castanopsis eyrei*)、木荷(*Schima superba*)、石栎(*Lithocarpus glaber*)、光皮树(*Cornus wilsoniana*)、乌桕(*Sapium sebiferum*)、山桐子(*Idesia polycarpa*)、油松(*Pinus tabulaeformis*)、辽东栎(*Quercus liaotungensis*)、红松(*Pinus koraiensis*)、蒙古栎(*Quercus mongolica*)、胡桃楸(*Juglans mandshurica*)、白檀(*Symplocos paniculata*)和黄连木(*Pistacia chinensis*)。这些树种均具有较明显的分布区域,且分布在我国的南北各地,其中青冈、水青冈、甜槠、木荷、石栎、光皮树、乌桕、山桐子等主要分布在我国亚热带常绿阔叶林区域;油松、辽东栎主要分布在我国暖温带落叶阔叶林区域;红松、蒙古栎、胡桃楸主要分布在我国温带针阔混交林区域;白檀、黄连木则分布较广,在亚热带常绿阔叶林及暖温带落叶阔叶林区域均有分布。通过查询中国及各省的植物志、树木志、森林志、中国科学院植物所标本馆标本记录和相关文献等,广泛收集了每一树种分布数据,并具体到县一级行政单位。同时,利用 Arc GIS 8.1 版本将每一树种的存在与不存在(用 1 表示存在,0 表示不存在)分布数据作为一项属性添加到数字化的中国县级行政区划图(中国科学院地理研究所资源与环境信息系统国家重点实验室制作,1996)属性表上,然后利用这一属性将数字化的中国县级行政区划图转化为 3799 个 50 km×50km 栅格单元。

1.2 气候数据

Woodward^[35]和 Prentice^[36]等从影响植物生理生态特性的角度选取气候变量,以反映植物对冬季低温的忍受程度、对热量的需求和对湿度的需求。这些气候变量是限制植物生长的主要因子。据此,本文选取 Kira 寒冷指数(coldness index,CI)^[37]和 1 月份最低温度表示冬季低温,用 Kira 温暖指数(warmth index,WI)^[37]和 7 月份最高温度表示热量需求,用徐文铎提出的湿润指数(humidity index,HI)^[38]和年降水量表示湿度需求。

本文所用的气候数据为 1960~1990 年全国 647 个气象台站的气象资料,由此得到 Kira 温暖指数、寒冷指数、徐文铎的湿润指数,1 月份最低温度、7 月份最高温度和年降水量等 6 个气候指标。为使气候资料与树种分布数据相统一,利用 Kriging 插值法将气候数据插值到 3799 个 50 km×50 km 栅格单元。

未来中国的气候变化情景采用高学杰等^[39]利用区域气候模式(Regional Climate Model,RCM)嵌套全球气候模式获得的 CO₂ 浓度倍增下气候情景:年均温增加 2.5℃,年降水增加 12%。据此计算未来气候指标,以预测树种的未来分布。

1.3 统计模型

用于比较模拟树种分布与未来分布趋势的 3 个模型为:广义线性模型(GLM)^[40]、广义加法模型(GAM)^[41]和分类回归树

(CART)^[42]。

1.3.1 广义线性模型 广义线性模型为常规正态线性模型的直接推广。它不需要响应变量服从正态分布,响应变量可以服从指数型分布族中的任何概率分布,如二项分布、泊松分布、伽马分布及负二项分布等。响应变量的值通过预测变量的线性组合来预测得到,响应变量和预测变量之间通过一个联系函数连接起来,一般形式为:

$$g(\mu) = \alpha + \sum_{j=1}^p \beta_j x_j$$

式中, $g(\mu)$ 为一特定联系函数,当响应变量服从二项分布时, $g(\mu)=\log (\mu/(1-\mu))$; α 为常数, β 为回归系数, μ 为由预测变量线性结合预测得到的响应变量发生的概率。

1.3.2 广义加法模型 广义加法模型为广义线性模型非参数化的扩展,它用平滑函数 f_i 代替参数 β_i ,因而数据中的非线性关系,如双峰和不对称现象,就可以很容易被发现,因而它比广义线性模型更灵活,即该模型的预测结果不是来自于一个预先设定好的模型。它能应用非参数的方法检测数据的结构,并找出数据中的规律,从而得到更好的预测结果。其一般形式为:

$$g(\mu) = a + \sum_{j=1}^p f_j(x_j)$$

1.3.3 分类回归树 分类回归树是一种非参数化的回归及分类技术,不需要预先假设响应变量和预测变量之间的关系,而是根据响应变量,利用递归划分法,将由预测变量定义的空间划分为尽可能同质的类别。每一次划分都由预测变量的一个最佳划分值来完成,将数据分成两个部分,重复此过程,直到数据不可再分。分类回归树算法由树生长和树剪枝 2 个步骤组成^[24]。

1.3.4 模型运算 模型运算均在 S-PLUS 6.0 软件上进行。广义线性模型运算时定义方程式中每个预测变量与响应变量之间为线性关系;广义加法模型运算时运用样条平滑函数,每个预测变量的平滑自由度均限定为 4;分类及回归树运算结果应用 prune 函数剪枝,得到一个最佳分类树。模型结果应用 Arc GIS 8.1 版本以地图形式表示出来。

1.4 模型验证和评估

为验证模型的模拟质量,把整个数据分为两个子集:训练和评估子集。通过随机取样取得总数据集的 70% 作为训练子集,用来训练模型;余下的 30% 作为评估子集,用来验证模型。同时,应用 Kappa 统计值^[43]和接收机工作特征曲线(receiver operating characteristic curve, ROC 曲线)^[44-45]来评估和比较模型模拟精度。Kappa 值法是通过模拟结果取不同的阈值,得到不同的 Kappa 值,此中最大的 Kappa 值被用来评估模型^[46],评估标准^[47]为:Kappa 值<0.4,失败(fail)、0.4~0.55,一般(fair);0.55~0.7,好(good);0.7~0.85,很好(very good);>0.85,非常好(excellent)。ROC 曲线则不需要通过对模拟结果选取固定的阈值来确定模型精度,它将不同阈值的正确模拟存在的百分率(精确性)和 1 减去正确模拟不存在的百分率(特异性)通过作图法表示在图上,通过比较曲线和 45°线之间的面积(area under curve, AUC)来确定模型的模拟精度。该模型的评估标准^[48]为:AUC 为 0.50~0.60,失败(fail);0.60~0.70,较差(poor);0.70~0.80,一般(fair);0.80~0.90,好(good);0.90~1.0,非常好(excellent)。

2 结果分析

2.1 模型评估

将评估子集的数据代入训练后的模型中,用其所得结果给出相应的 AUC 和 Kappa 值(表 1),以此来评估不同模型对树种分布的模拟精度。

由表 1 可知,GLM 模型、GAM 模型与 CART 模型 3 个模型的 AUC 平均值分别为 0.951、0.980 与 0.948,均大于 0.9,模型模拟精度均达到非常好程度;Kappa 平均值分别为 0.675、0.793 与 0.786,模拟精度分别为好、很好与非常好。无论从 AUC 和 Kappa 值来看模型的模拟效果,GAM 模型都优于 GLM 模型和 CART 模型;而从 AUC 来看,GLM 模型的模拟效果稍优于 CART 模型;从 Kappa 值来看,CART 模型的模拟效果好于 GLM 模型。

3 个模型对青冈、水青冈、甜槠、木荷、石栎、光皮树、乌桕、山桐子、蒙古栎、白檀、黄连木等树种的模拟均达到较高精度,AUC 均达到非常好的模拟效果;Kappa 值亦均达到好的效果。其中,3 个模型对黄连木分布的模拟效果最为突出,AUC 和 Kappa 值最高,均达到非常好的模拟效果。红松、胡桃楸分布的 GLM 模型模拟效果不太理想,Kappa 值分别为 0.464 和 0.340,其它两个模型的模拟效果则较好。油松、辽东栎分布的 3 个模型模拟的 AUC 和 Kappa 值均最低,以 Kappa 值为例,对油松和辽东栎分布的 GLM 模型模拟的 Kappa 值分别为 0.337 和 0.281,模拟效果均失败;而 GAM 模型和 CART 模型的模拟效果均只达到一般程度。

2.2 树种现实分布模拟比较

通过选取最大化 Kappa 值将模型模拟的树种现实分布借助于地理信息系统软件以地图方式显示,以比较模型模拟效果的优劣。选取模拟效果较好的青冈、木荷、红松及模拟效果较差的油松进行比较分析。图 1~图 4 分别为青冈、木荷、红松与油松的

现实分布和 3 个模型的模拟结果。可以看出,3 个模型均能很好地模拟青冈和木荷的现实分布,且模拟效果较为接近,其中 GAM 模型和 CART 模型的模拟效果较为理想,而 GLM 模型的模拟效果则稍差。GLM 模型对红松分布的模拟效果不太理想,而另 2 个模型的模拟效果则较好。3 个模型对于油松分布的模拟效果均不甚理想,其中以 GLM 模型的模拟效果稍好,与模型评估结果相吻合。模型模拟结果中,油松在我国新疆西北部、新疆与青海交界处及新疆与内蒙古交界处有大量分布,这显然与油松的实际分布不符,可能与油松分布区域狭长、零散,且为人工树种,受人为干扰较大有关。而青冈与木荷的模拟效果较为理想,可能是因为青冈、木荷主要分布在亚热带常绿阔叶林区域,基本处于自然分布状况,受人为干扰较小,且分布广泛。

表 1 3 个模型对 15 个树种分布模拟的 AUC 和 Kappa 值
Table 1 AUC and Kappa values for simulated 15 tree species distributions by GLM,GAM and CART

树种 Tree species	AUC			Kappa		
	GLM	GAM	CART	GLM	GAM	CART
青冈 <i>Cyclobalanopsis glauca</i>	0.985	0.992	0.977	0.855	0.892	0.898
水青冈 <i>Fagus longipetalata</i>	0.963	0.980	0.922	0.664	0.794	0.675
甜槠 <i>Castanopsis eyrei</i>	0.981	0.993	0.954	0.754	0.857	0.805
木荷 <i>Schima superba</i>	0.995	0.996	0.995	0.886	0.905	0.919
石栎 <i>Lithocarpus glaber</i>	0.964	0.968	0.948	0.669	0.645	0.664
光皮树 <i>Cornus wilsoniana</i>	0.985	0.987	0.965	0.802	0.801	0.799
乌桕 <i>Sapium sebiferum</i>	0.987	0.992	0.962	0.893	0.913	0.906
山桐子 <i>Idesia polycarpa</i>	0.973	0.978	0.953	0.741	0.741	0.793
油松 <i>Pinus tabulaeformis</i>	0.846	0.948	0.894	0.337	0.595	0.640
辽东栎 <i>Quercus liaotungensis</i>	0.843	0.932	0.747	0.281	0.501	0.434
红松 <i>Pinus koraiensis</i>	0.934	0.986	0.975	0.464	0.876	0.854
蒙古栎 <i>Quercus mongolica</i>	0.934	0.983	0.944	0.622	0.783	0.763
白檀 <i>Symplocos paniculata</i>	0.989	0.992	0.993	0.887	0.899	0.901
胡桃楸 <i>Juglans mandshurica</i>	0.892	0.986	0.995	0.340	0.764	0.809
黄连木 <i>Pistacia chinensis</i>	0.996	0.997	0.993	0.933	0.952	0.936
平均值 Mean	0.951	0.980	0.948	0.676	0.795	0.786

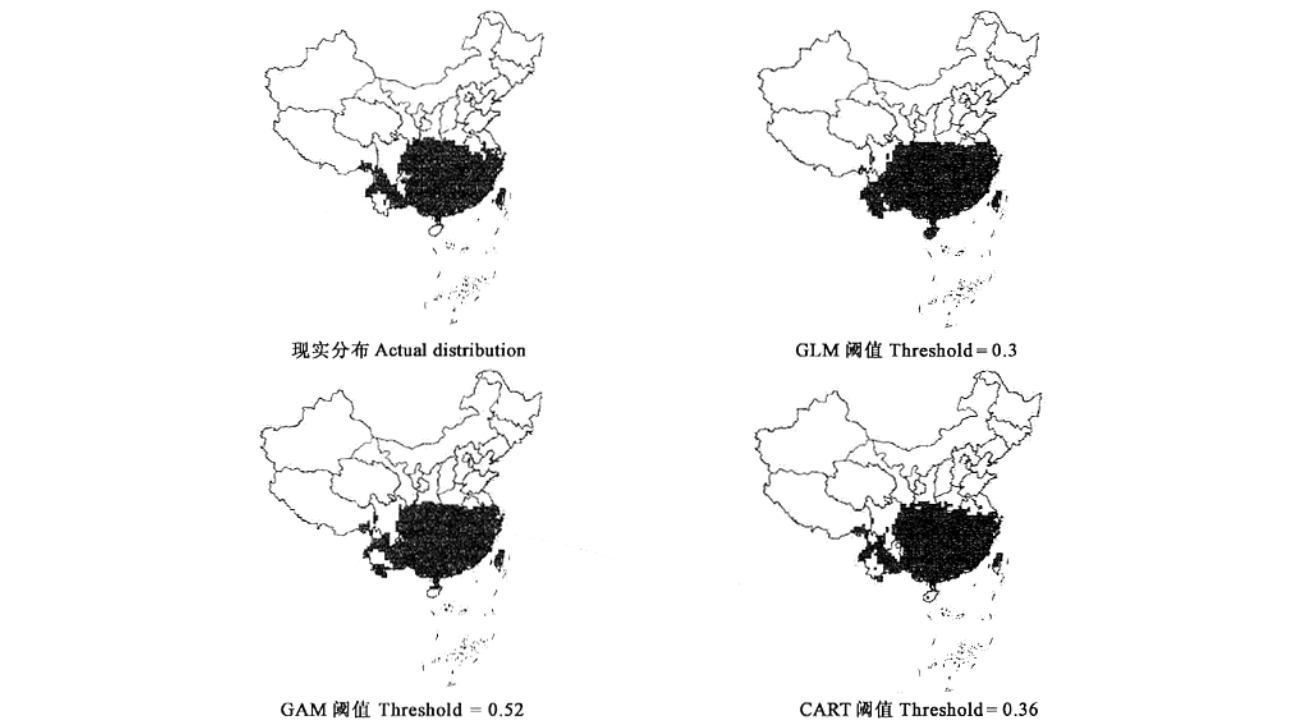


图 1 青冈的现实分布及 3 个模型模拟的分布
Fig. 1 Actual and simulated distribution for *Cyclobalanopsis glauca* according to GLM,GAM and CART
黑色代表有分布,白色代表无分布 Dark corresponds to distribution, white to no distribution

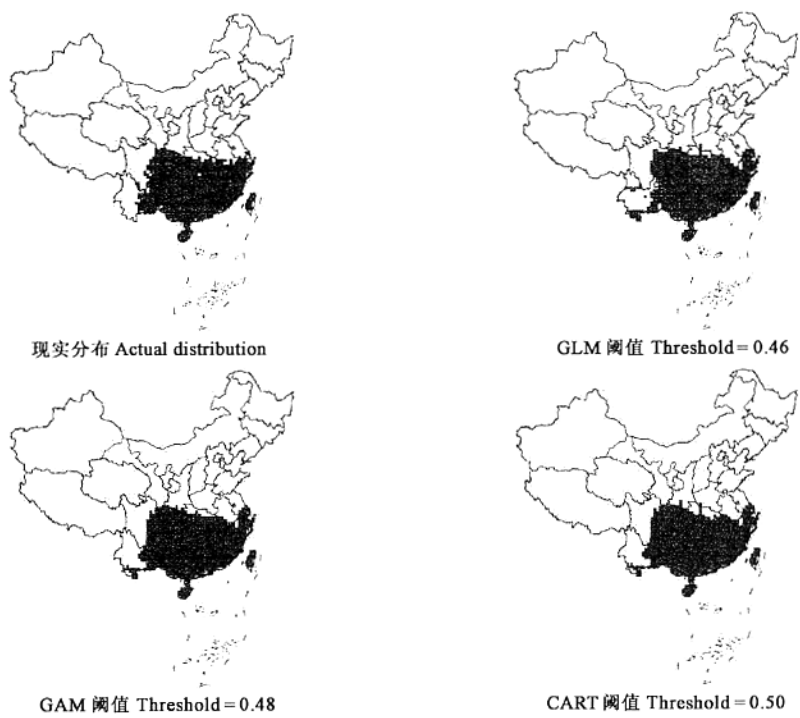


图 2 木荷的现实分布及 3 个模型模拟的分布

Fig. 2 Actual and simulated distribution for *Schima superba* according to GLM,GAM and CART
黑色代表有分布,白色代表无分布 Dark corresponds to distribution, white to no distribution

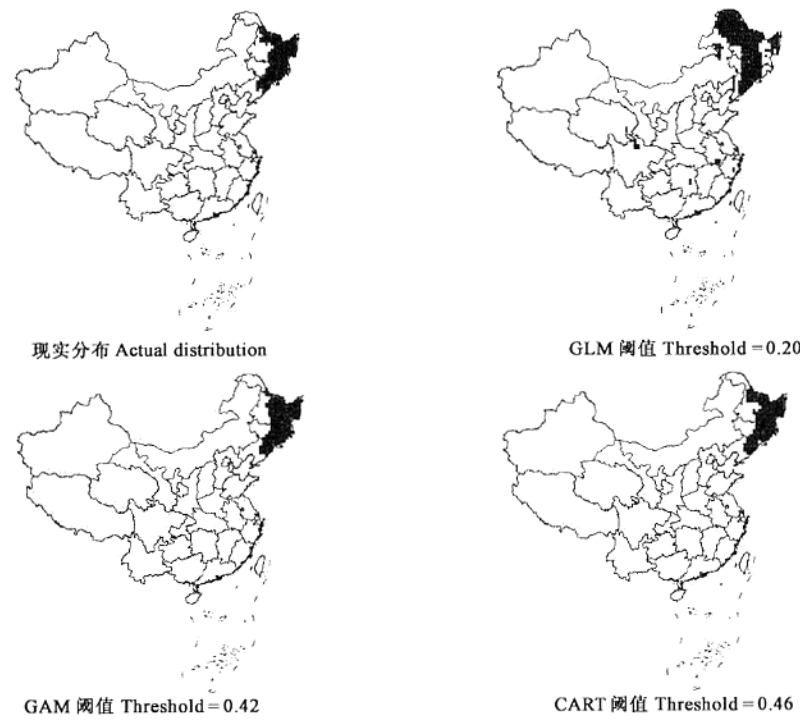


图 3 红松的现实分布及 3 个模型模拟的分布

Fig. 3 Actual and simulated distribution for *Pinus koraiensis* according to GLM,GAM and CART
黑色代表有分布,白色代表无分布 Dark corresponds to distribution, white to no distribution

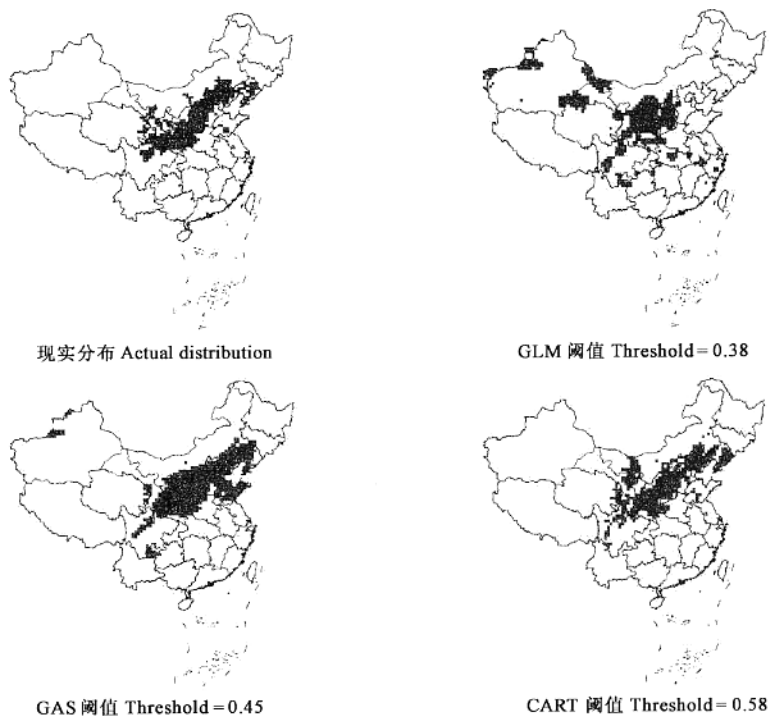


图 4 油松的现实分布及 3 个模型模拟的分布

Fig. 4 Actual and simulated distribution for *Pinus tabulaeformis* according to GLM,GAM and CART
黑色代表有分布,白色代表无分布 Dark corresponds to distribution, white to no distribution

2.3 未来气候条件下模型预测结果分析

为比较模型模拟效果的优劣,利用地理信息系统软件将未来气候情景下模型预测的树种未来潜在分布结果和树种现实分布叠加显示,并以栅格单元为基础计算每一树种在气候变化下分布区面积的具体变化。选取现实分布中预测结果较好的青冈(图 5)和蒙古栎(图 6)一南一北 2 种树种进行比较分析。可以看出,在未来气候变化情景下,GLM 模型和 GAM 模型预测的青冈分布结果较为接近,除基本保持原有分布区域外(分别为 94%和 96%),均向西和向北扩展,而 CART 模型模拟结果除向西向北扩展外,青冈在广东和广西南部分布区域将消失。总之,3 个模型关于未来气候变化情景下青冈分布的预测均表明,青冈在基本保持原有分布的基础上,其未来潜在分布区域均有所扩大,不过从 Kappa 值来看,CART 模型预测结果的可信度更大。

同样,3 个模型关于未来气候变化情景下蒙古栎分布的预测均表明,蒙古栎有明显的向西扩展趋势,扩展面积的大小为:

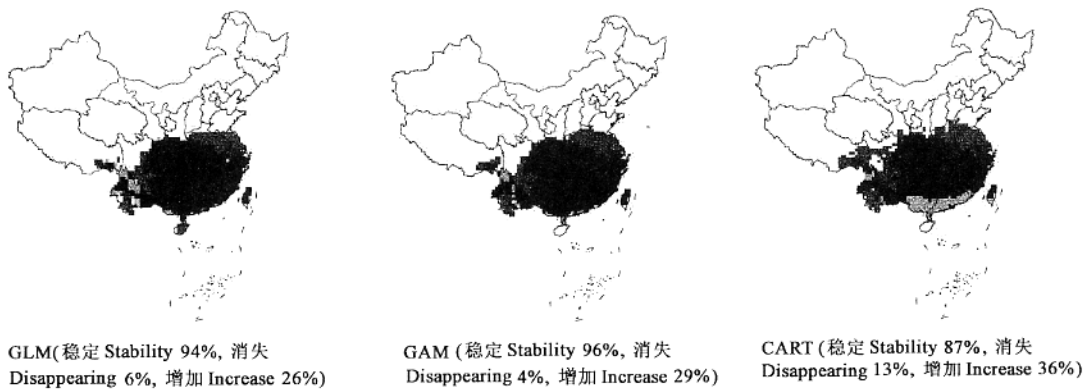


图 5 青冈的现实分布及 3 个模型预测的未来分布比较

Fig. 5 Comparison of actual and simulated distribution for *Cyclobalanopsis glauca* under climate change according to GLM,GAM and CART
黑色代表有分布,白色代表无分布 Dark colour corresponds to stable distribution, dark grey to new distribution, clear grey to lost distribution

GAM 模型的预测扩展面积>GLM 模型>CART 模型。其中, GLM 模型预测的蒙古栎消失与新出现的分布区面积相同, 均为 33%, 而 GAM 模型和 CART 模型预测的蒙古栎分布区面积均为不同程度的扩展, 扩展面积分别为原有面积的 37% 和 15%。从 Kappa 值来看, GAM 模型的预测结果更为可信。

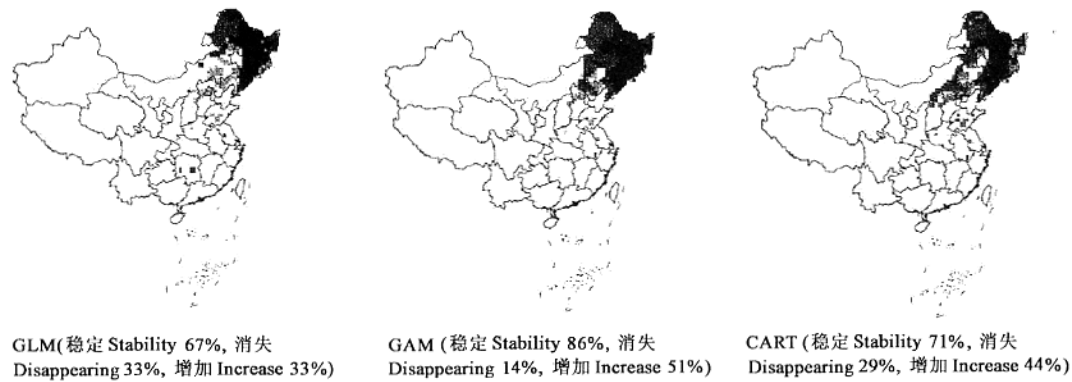


图 6 蒙古栎的现实分布及 3 个模型预测的未来分布比较

Fig. 6 Comparison of actual and simulated distribution for *Quercus mongolica* under climate change according to GLM, GAM and CART 黑色代表有分布, 白色代表无分布 Dark colour corresponds to stable distribution, dark grey to new distribution, clear grey to losed distribution

3 讨论

基于我国 15 种树种分布对模型模拟效果的评估表明, 3 个模型均有较高的模拟精度。在 3 个模型中, GAM 模型的模拟效果最好, CART 模型次之, GLM 模型最差。值得注意的是, 这并不意味着 GAM 模型对所有树种的模拟效果最好。当研究的对象和研究的方法不同时, 模型的模拟效果也不一致。例如, 关于油松、白檀、胡桃楸的分布, CART 模型的模拟效果就好于 GAM 模型, 其它一些研究也得到类似的结果, 如 Thuiller^[49]利用欧洲 61 个树种对 GLM 模型、GAM 模型和 CART 模型以及人工神经网络 4 个模型进行比较指出, 4 个模型的模拟效果没有显著差异, 应根据不同树种分布选择最优模型, 以达到最佳模拟效果。因此, 研究者在研究中应慎重选择模型, 依照研究对象、目的和要求选择合适的模型。

本论文在构造 GLM 模型时, 只应用了线性项; 在构造 GAM 模型时, 选择了样条平滑函数; 在构造 CART 模型时, 为两个子集。然而, 构造模型的方法很多, 如在构造 GLM 模型时, 可以考虑添加二次项、三次项、多次项及交互作用; 在构造 GAM 模型时, 可以选择其他平滑函数, 如 lowess 平滑函数; 在构造 CART 模型时, 可将数据分为 10 个子集, 利用 10 次交叉验证来运算模型^[49]。同时, 也可以应用其他模型技术, 如人工神经网络^[27,28]和遗传算法^[51~53]来构造模型。不同的模型构造方法将对模型的模拟效果产生一定的影响。本论文中 GLM 模型的模拟效果较差, 可能与其不能处理复杂曲线有关^[21], 也可能与构造模型的方法有关。Vatssuerès^[24]曾通过添加二次项、三次项和交互项构造 GLM 模型, 并利用逐步回归选择显著项, 使得 GLM 模型的模拟效果显著提高。

在区域和洲际尺度, 气候是控制植物分布的主要环境因子^[35], 许多研究在此基础上利用气候变量建立物种分布模型, 并用来预测气候变化对物种分布的潜在影响^[15~18]。然而, 必须指出的是, 随着空间尺度的减小, 其它的一些因素, 如土壤、地形、生物之间的交互作用, 物种扩散能力将对物种的分布产生重大影响^[54]。在小尺度上, 已有研究将土壤、地形和海拔等环境变量引入到物种分布模型中来, 取得了很好的模拟效果^[5,26]。本论文研究的尺度建立在全国尺度水平, 因而只选用了气候变量作为预测变量。结果表明, 3 个模型对大部分树种分布的模拟效果均能达到较好程度, 尤其是黄连木, 无论从 AUC 和 Kappa 值来看, 模拟效果均达到非常好的程度。较高精度的物种气候模型无疑可以用作研究气候变化下物种分布的潜在变化。

在模型研究中, 地理信息系统是一种很有力的辅助手段, 其强大的数据处理能力、空间分析功能极大地加强了模型模拟、预测和定量分析能力^[5,6,17,26,49]。在已有的关于中国物种分布与气候关系研究中, 在数据处理方面, 主要应用距离物种实际分布最近的气象站的气候资料代替物种实际分布地点的气候资料^[29,30,33], 由于气象站大部分位于城镇, 与物种的分布地点有一定距离, 不可避免地会影响到数据的精度。因而, 本论文利用地理信息系统分析和处理数据, 让树种分布数据和气候数据在每一个栅格单元上一一对应起来, 有效提高了数据精度。同时, 应用地理信息系统将模型模拟结果清晰和直观地表现在地图上, 可直接反映出模型模拟能力的差异, 尤其能定量分析环境变化下物种分布的潜在变化。例如, 未来气候变化情景下, 利用地理信息系统将模型预测的蒙古栎未来潜在分布与现实分布叠加显示表明, 蒙古栎有向西扩展的趋势, 其中 GAM 模型预测其分布区扩大的

面积最大,为 51%。需要注意的是,在分析不同模型预测的物种未来分布时,应结合其预测精度进行分析,较高预测精度模型的预测结果可信度较大。

4 结论

基于中国 15 种树种的地理分布比较分析了广义线性模型(GLM)、广义加法模型(GAM)和分类回归树(CART)的优劣。结果表明,3 个模型对于油松、辽东栎分布的模拟效果较差。以 Kappa 值为例,油松和辽东栎分布的 GLM 模型模拟的 Kappa 值分别为 0.337 和 0.281,模拟效果为失败,GAM 模型和 CART 模型的模拟效果则都在一般程度。3 个模型对其余树种分布的模拟均达到较高精度,其中对黄连木模拟效果最好,3 个模型模拟的 AUC 和 Kappa 值最高,均达到了非常好程度。总体而言,这 3 个模型模拟效果较高,其中以 GAM 模型的模拟效果最优。

以地理信息系统将青冈、木荷、红松、油松的模拟效果以地图方式显示出来表明,地理信息系统能直观、定量反映模型模拟效果与模型预测结果差异。3 个模型均能很好模拟青冈和木荷的现实分布,且模拟效果接近;红松分布的 GLM 模型模拟效果不太理想,油松分布的 3 个模型模拟效果均不甚理想,以 GLM 模型最差。这些结果与模型模拟效果评估相吻合。

在未来气候变化情景下,以 3 个模型模拟效果均较高的青冈、蒙古栎为研究对象分析基未来变化趋势。结果表明,GLM 模型与 GAM 模型预测的青冈未来分布结果较为接近,均向西和向北扩展,而 CART 模型预测的分布除向西向北扩展外,其在广东和广西南部分布区域将消失。3 个模型预测蒙古栎未来分布均将向西扩展,扩展面积的大小为:GAM 模型>GLM 模型>CART 模型。

References:

- [1] Austin M P. Models for the analysis of species' response to environmental gradients. *Vegetatio*, 1987,**69**:35~45.
- [2] Hill M O. Patterns of species distribution in Britain elucidated by canonical correspondence analysis. *Journal of Biogeography*, 1991,**18**:247~255.
- [3] Buckland S T, Elston A. Empirical models for the spatial distribution of wildlife. *Journal of Applied Ecology*, 1993,**30**:478~495.
- [4] Beerling D J, Huntley B, Bailey J P. Climate and the distribution of *Fallopia japonica*: use of an introduced species to test the predictive capacity of response surfaces. *Journal of Vegetation Science*, 1995,**6**:269~282.
- [5] Franklin J. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables. *Journal of Vegetation Science*, 1998,**9**:733~748.
- [6] Guisan A, Theurillat J P, Kienast F. Predicting the potential distribution of plant species in an alpine environment. *Journal of Vegetation Science*, 1998,**9**:65~74.
- [7] Guisan A, Weiss S B, Weiss A D. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 1999,**143**:107~122.
- [8] Corsi F, Duprè E, Boitani L. A large-scale model of wolf distribution in Italy for conservation planning. *Conservation Biology*, 1999,**13**:150~159.
- [9] Manel S, Dias J M, Ormerod S J. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological Modelling*, 1999,**120**:337~347.
- [10] Owen J G. Patterns of herpetofaunal species richness: relation to temperature, precipitation and variance in elevation. *Journal of Biogeography*, 1989,**16**:141~150.
- [11] Heikkinen R K. Predicting patterns of vascular plant species richness with composite variables: a meso-scale study in Finnish Lapland. *Vegetatio*, 1996,**126**:151~165.
- [12] Fraser R H. Vertebrate species richness at the mesoscale: relative roles of energy and heterogeneity. *Global Ecology and Biogeography Letter*, 1998,**7**:215~220.
- [13] Wohlgemuth T. Modeling floristic species richness on a regional scale: a case study in Switzerland. *Biodiversity Conservation*, 1998,**7**:159~177.
- [14] Midgley G F, Hannah L, Millar D, *et al.* Developing regional and species-level assessments of climate change impacts on biodiversity: a preliminary study in the Cape Floristic Region. *Biological Conservation*, 2003,**112**:87~97.
- [15] Huntley B, Berry P M, Cramer W, *et al.* Modeling present and potential future ranges of some European higher plants using climate response surfaces. *Journal of Biogeography* 1995,**22**:967~1001.
- [16] Sykes M T, Prentice I C. Climate change, tree species distributions and forest dynamics: a case study in the mixed conifer/northern hardwoods zone of northern Europe. *Climate Change*, 1996,**34**:161~177.
- [17] Bakkenes M, Alkemade R M, Ihle F, *et al.* Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biology*, 2002,**8**:390~407.

- [18] Berry P M, Dawson T E, Harrison P A, *et al.* Modelling potential impacts of climate change on the bioclimatic envelope of species in Britain and Ireland. *Global Ecology and Biogeography*, 2002, **11**:453~462.
- [19] Austin M P, Meyers J A. Current approaches to modelling the environmental niche of eucalypts; implications for management of forest biodiversity. *Forest Ecology and Management*, 1996, **85**:95~106.
- [20] Brito J C, Crespo E G, Paulo O S. Modelling wildlife distributions: logistic multiple regression vs overlap analysis. *Ecography*, 1999, **22**:251~260.
- [21] Yee T W, Mitchell N D. Generalized additive models in plant ecology. *Journal of Vegetation Science*, 1991, **2**:587~602.
- [22] Frescino T S, Edwards T C Jr, Moisen G G. Modeling spatially explicit forest structural attributes using Generalized Additive Models. *Journal of Vegetation Science*, 2001, **12**:15~26.
- [23] Lehmann A, Overton J M, Leathwick J R. GRASP: generalized regression analysis and spatial prediction. *Ecological Modelling*, 2003, **160**:165~183.
- [24] Vayssières M P, Plant R E, Allen-Diaz B H. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 2000, **11**:679~694.
- [25] Rouget M, Richardson D M, Lavorel S, *et al.* Determinants of distribution of six Pinus species in Catalonia, Spain. *Journal of Vegetation Science*, 2001, **12**:491~502.
- [26] Thuiller W, Araújo M B, Lavorel S. Generalised models versus classification tree analysis: a comparative study for predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, 2003, **14**:669~680.
- [27] Lek S, Delacoste M, Baran P, *et al.* Application of neural networks to modelling non-linear relationships in ecology. *Ecological Modelling*, 1996, **90**:39~52.
- [28] Pearson R G, Dawson T P, Berry P M, *et al.* Species: A Spatial Evaluation of Climate Impact on the Envelope of Species. *Ecological Modelling*, 2002, **154**:289~300.
- [29] Hong B G, Li S Z. The preliminary study of the correlations between the distribution of main evergreen broad-leaf tree species in Jiangsu and climates. *Acta Ecologica Sinica*, 1981, **1**(2): 105~111.
- [30] Xu W T. The relation between the zonal distribution of types of vegetation and the climate in northeast China. *Acta Phytocologica et Geobotanica Sinica*, 1986, **10**(4): 254~263.
- [31] Zhang X S. A vegetation-climate classification system for global changes studies in China. *Quaternary Sciences*, 1993, **2**:157~169.
- [32] Zhou G S, Zhang X S. Study on climate-vegetation classification for global change in China. *Acta Botanica Sinica*, 1996, **38**(1): 8~17.
- [33] Ni J, Song Y C. Relationships between geographical distribution of *Cyclobalanopsis glauca* and climate in China. *Acta Botanica Sinica*, 1997, **39**(5): 451~460.
- [34] Li B, Zhang J T. Analysis of relationships between vegetation and climate variables in Loess Plateau. *Acta Ecologica Sinica*, 2003, **23**(1): 82~89.
- [35] Woodward F I. *Climate and plant distribution*. Cambridge: Cambridge University Press, 1987.
- [36] Prentice I C, Cramer W, Harrison S P, *et al.* A global biome model based on plant physiology and dominance, soil properties and climate. *Journal of Biogeography*, 1992, **19**:117~134.
- [37] Kira T. On the altitudinal arrangement of climatic zone in Japan. *Kanti-Nougaku*. 1948, **2**:143~173.
- [38] Xu W T. Kira's warmth index and its application in the Chinese vegetation. *Chinese Journal of Ecology*, 1985, **3**:35~39.
- [39] Gao X J, Zhao Z C, Ding Y H, *et al.* Climate Change due to Greenhouse Effects in China as Simulated by a Regional Climate Model. *Advances in Atmospheric Sciences*, 2001, **18**:1224~1230.
- [40] McCullagh P, Nelder J A. *Generalized Linear Models*. second ed. London: Chapman and Hall, 1989.
- [41] Hastie T J, Tibshirani R. *Generalised Additive Models*. London: Chapman and Hall, 1990.
- [42] Breiman L, Friedman J H, Olshen R A, *et al.* *Classification and Regression Trees*. New York: Chapman and Hall, 1984.
- [43] Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, **20**:37~46.
- [44] Hanley J A, McNeil B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 1982, **143**:29~36.
- [45] Hanley J A, McNeil B J. Methods of comparing the areas under receiver operating characteristic (ROC) curve. *Radiology*, 1983, **148**:839~843.
- [46] Fieldings A H, Bell J F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 1997, **24**:38~49.
- [47] Monserud R A, Leemans R. Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling*, 1992, **62**:275~293

[48] Swets K A. Measuring the accuracy of diagnostic systems. *Science*, 1988,**240**:1285~1293.

[49] Thuiller W. BIOMOD-optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, 2003,**9**:1353~1363.

[50] Stockwell D R B, Peters D. The GARP modelling system: problems and solutions to automated spatial prediction. *International Journal Geographical Information Science*, 1999,**13**:143~158.

[51] Peterson A T, Sánchez-Cordero V, Soberón J, *et al.* Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling*, 2001,**144**:21~30.

[52] Peterson A T, Ortega-Huerta M A, Bartley J, *et al.* Future projections for Mexican faunas under global climate change scenarios. *Nature*, 2002,**416**:626~629.

[53] Aderson R P, Lew D, Peterson A T. Evaluating predictive models of species' distribution: criteria for selecting optimal models. *Ecological Modelling*, 2003,**162**:211~232.

[54] Pearson R G, Dawson T P. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 2003,**12**:361~371.

参考文献:

[29] 洪必恭,李绍珠. 江苏主要常绿阔叶树种的分布与热量关系的初步研究. *生态学报*,1981,**1**(2):105~111.

[30] 徐文铎. 中国东北主要植被类型的分布与气候的关系. *植物生态学与地植物学学报*,1986,**10**(4):254~263.

[31] 张新时. 研究全球变化的植被-气候分类系统. *第四纪研究*,1993,**2**:157~169.

[32] 周广胜,张新时. 全球变化的中国气候-植被分类研究. *植物学报*,1996,**38**(1):8~17.

[33] 倪健,宋永昌. 中国青冈的地理分布与气候的关系. *植物学报*,1997,**39**(5):451~460.

[34] 李斌,张金屯. 黄土高原地区植被与气候的关系. *生态学报*,2003,**23**(1):82~89.

[38] 徐文铎. 吉良的热量指数及其在中国植被中的应用. *生态学杂志*,1985,**3**:35~39.