延河流域典型物种分布预测模型比较研究

曹剑侠[1], 温仲明[1,2], 李 锐[1,2]
(1. 西北农林科技大学 资源环境学院, 陕西 杨凌 712100; 2. 中国科学院 水利部 水土保持研究所, 陕西 杨凌 712100)

摘 要: 物种分布预测一直以来都是生态学研究的重要内容之一。应用生态学的发展为物种分布预测提供了众多强有力的模型,在推进物种分布预测进展的同时,也增加了合适模型选择的难度。评价和比较不同模型的预测效果,对于模型的选择和应用具有非常重要的意义。以黄土丘陵区延河流域为研究区,采用 R 语言和 BIOMOD 程序包为平台,选择人工神经网络(artificial neural networks, ANN)等 9 个较常用的物种分布模型,比较它们在物种分布预测精度上的差异,为物种分布预测模型的选择提供依据,也为进一步预测未来气候变化情景下物种空间分布的变化奠定基础。研究结果表明,不同模型对不同物种的模拟精度差异明显。根据 Kappa, TSS 和 Roc 评价方法, 9 个模型对百里香(*Thymus mongolicus*)分布的预测精度最高; 对铁杆蒿(*Artemisia gmelinii*)分布的模拟精度最差; 而对其余物种分布的模拟精度均比较理想,其中以随机树 RF 模型最好。

关键词: 物种分布预测; R—BIOMOD; 最优模型
文献标识码: A 文章编号: 1000-288X(2010)03-0134-06 中图分类号: Q141, X171.4

Comparison of Predictive Models for Representative Species Distribution in Yanhe River Basin

CAO Jian-xia[1], WEN Zhong-ming[1,2], LI Rui[1,2]
(1. *College of Resources and Environment, Northwest A & F University, Yangling* 712100, *Shaanxi, China*; 2. *Institute of Soil and Water Conservation, Chinese Academy of Sciences and Ministry of Water Resources, Yangling, Shaanxi* 712100, *China*)

Abstract: The spatial distribution of species is one of the important research projects in ecology. In recent years, much progress has been made in predictive models of species with the development of applied ecology, but on the other hand, it makes more difficult to use the species distribution models because the number of available techniques or models is large and is increasing steadily, making it confused for users to select the most appropriate methodology for their needs. So evaluating and comparing the predictive accuracy of different distribution models has great significances for selecting and using these models. BIOMOD (BIOdiversity MODelling), a new computation framework based on R language, is presented. Yanhe River basin was selected as the study area. In order to select a suitable model for simulating and predicting the typical species distributions in the study area, nine widely used modeling techniques in species predictions, such as artificial neural networks (ANN), were compared to predict the representative species distribution. The comparison of their differences in prediction accuracy can not only provide evidence for selecting species distribution models, but also lay a foundation for projecting the species spatial distributions into different environmental conditions (e.g. climate change scenarios). Three available techniques of Roc, Kappa, and TSS were used to assess each model's performance. Results from the model comparison showed that the relative model performance and simulation accuracy of different models were quite different across species. The evaluation indicated that the nine models had the highest predictive accuracy for the *Thymus mongolicus* distribution, yet the predictive accuracy of the *Artemisia gmelinii* distribution was the lowest. Furthermore, the nine models can predict the distribution of the rest species very well. It is concluded that the RF model which has the highest predictive

收稿日期: 2010-01-07 修回日期: 2010-03-19
资助项目: 国家自然科学基金项目"黄土丘陵区潜在植被格局及其对未来气候变化的响"(40871246); 国家科技支撑课题"植被优化配置与可持续建设技术"(2006BAD09B03)
作者简介: 曹剑侠(1984—), 女(汉族), 河南省南阳市人, 硕士研究生, 研究方向为资源环境遥感监测与评价。E-mail: cjxbaby1984@163.com。
通信作者: 温仲明(1969—), 男(汉族), 陕西省定边县人, 博士, 副研究员, 主要从事植被生态恢复重建等方面的研究工作。E-mail: zmwen@ms.iswc.ac.cn。

accuracy using the three methods is the best one among the nine models.

Keywords: predictive distribution of species; R-BIOMOD; best model

随着全球变化对物种多样性的影响, 准确预测物种的空间分布及其对气候变化的响应, 成为近年来生态学和生物保护研究的热点[1-3], 并开展了大量的研究工作, 建立了很多新的模型方法, 如广义相加模型(generalized additive model, GAM)[4]、分类回归树(classification and regression tree analysis, CTA)[5-6]、神经网络系统[7]、基于规则的遗传算法预测(genetic algorithm for rule set prediction, GARP)[8] 等, 为生物多样性保护和生态恢复决策及评估等提供了重要技术支持。但对于具体的物种分布预测, 选择什么样的模型方法, 往往因模型的背景、原理和适用条件等的差异而不同, 选择什么样的模型方法比较合适对于应用者具有一定的难度[9]。因此, 对目前主要的物种分布模型进行比较研究, 并对其预测精度进行评估, 具有重要的指导意义。

黄土高原地区丘陵起伏, 地形复杂, 物种与环境关系复杂[10]。温仲明、赫晓慧等[1,11-12] 利用广义相加模型研究了延河流域自然植被分布与环境的关系, 初步表明, 利用非线性建模途径的预测结果要好于线性模型, 但除此之外, 是否还有更合适的模型来研究该区的物种空间分布, 目前尚未有深入探讨。为此, 本文选择延河流域为研究区, 针对黄土高原特殊的自然地理环境, 综合考虑影响植被分布的主要因素, 对目前物种分布研究中比较常用的 9 个模型进行比较研究, 以确定适宜的物种分布预测模型, 为该区域植被恢复的规划、保护和管理等提供科学依据, 亦希望为我国植被—环境研究提供借鉴。

1　研究区概况

延河流域位于 36°23′—37°17′ N, 108°45′—110°28′ E 之间, 流域全长 286.9 km, 总面积 7 687 km^2, 平均坡度为 4.3‰, 河网密度约为 4.7 km/ km^2。本流域属大陆性气候, 年降雨量 500 mm 左右, 年平均气温 9 ℃, 立地环境多变, 影响物种空间分布的因素众多, 涉及温度、降雨、蒸发及地形、土壤等因素。从东南向西北, 降雨、温度具有明显的梯度变化特征, 植被分布也呈现渐次变化, 从南向北分为南部辽东栎(*Quercus liaotungensis*)、刺槐(*Robinia pseudoacacia*)、油松(*Pinus tabulaeformis* Carr)、阔叶—针叶混交林带, 中部延安到安塞之间为柠条(*Caragana korshinskii* Kom)、白羊草(*Bothriochloa ischaemun*) 等草灌过渡带, 安塞以北为百里香、长芒草(*Stipa bungeana*) 草原带, 随环境梯度的变化明显。该区植物资源较丰富, 植物物种总数大体为 589 种, 分属 81 科[13]。

2　研究方法

2.1　模型介绍

根据目前进展, 非参数模型在建立物种分布模型方面较线性模型有较大的优势, 本文选择的 9 个模型, 除广义线性模型[14] 外(generalized linear model, GLM), 也大多为非参数模型, 如广义相加模型[14]、分类回归树分析[15]、人工神经网络、面域包络模型(surface range envelope, SRE)、广义增强模型(generalized boosting model, GBM)、随机树(breiman and cutler's random forest for classification and regression/ Random Forest, RF)、混合判别式分析(mixture discriminant analysis, MDA)、多元自适应回归样条(multiple adaptive regression splines, MARS)。这 9 个模型作为利用点数据进行空间预测的重要方法, 可分别在一系列响应变量和环境因子之间建立统计关系, 对于预测单个物种对多个环境变量的综合响应非常方便且具有深远的意义。

2.2　数据采集与处理

2.2.1　*植被数据的采集与处理*　取得具有代表性的空间样本数据, 是物种分布模型建立的基础。本文采用环境梯度分层采样技术进行植被样本数据的采集, 首先将延河流域划分为 17 个环境梯度单元; 然后在不同的梯度单元内, 根据每个梯度单元内的栅格数量, 确定相应的需采集的样点数量进行采样。采样时以环境梯度单元为基础, 考虑不同立地条件, 对处于稳定状态的或顶级演替阶段的自然植物群落进行调查。样地设置采用典型取样法, 对林地样方一般设为 10 m × 10 m, 草本样地为 2 m × 2 m。每个乔木群落内设置 1 个灌木样方(5 m × 5 m), 10 个草本样方, 每个灌木群落内设置 10 个草本样方。数据采集是逐次进行的, 时间为 2005 年, 2006 年及 2008 年。经过数据质量校验, 剔除部分研究区的误差样本(流域界外或数据记录缺失), 最后实际可用的样点数为 280 个。物种数据(响应变量) 整理为二元数据, 即物种存在用 1 表示, 不存在用 0 表示。

2.2.2　*环境因子的选择与获取*　在对影响植被分布的主要环境因子进行分析的基础上, 重点考虑了区域性的环境预测变量数据, 主要包括两类, 分别为气候因子和地形因子。

气候数据为 1980—2000 年延河流域及周边地区 57 个气象站点的气象资料, 来源于黄土高原各省、县属气象局。根据现有研究, 气候因子主要指温度和降雨[16-17]。根据对植被生长发育的影响, 选择了年均最

冷月气温、年均最热月气温、年均植被生长季节气温、年均气温、年均雨季降雨量、年均总降雨量、年均蒸发量、降雨季节变化和温度季节变化 9 个指标来表征气候对植被分布的影响。由于气象站提供的是逐年逐月的平均气温、极值温度与降雨, 因此各因子指标需要通过运算求得, 计算方法参见文献[11]。

地形主要指地形的起伏变化, 是影响局部植被空间分布的重要因素, 依据地形因子的重要性, 本研究选择坡度、坡向、高程与地形部位 4 个地形因子参与模型的建立。其中, 坡度、坡向可从 DEM 直接生成, 高程从 DEM 直接读取。地形部位本文采用高程残差分析, 并结合坡度, 将延河流域的地形部位划分为 7 类, 即河道与沟间平地、下坡位、中坡、上坡、峁顶、高平地和细小沟谷。

2.2.3 软件工具 本研究采用 R 语言和 BIOMOD 程序包构建模型。R 是近年来普及速度非常快的应用软件系统, 内含许多实用的函数, 资源丰富且可扩展性强[18-19]。BIOMOD 是基于 R 语言的一个程序包, 通过对模型中各个参数进行多次试验分析, 确定模型的所有参数, 并且提供了多种方法对模型进行检验。

2.2.4 模型验证和评估 在 BIOMOD 中对模型的评价有两类: 拟合度的评估和模型的精度评价。前者使用与每个模型相对应的标准方法, 如 ANOVA、AIC 等; 后者可用 3 种不同的技术方法(kappa, true skill statistics, Roc 曲线) 来评估模型的性能。本研究运用数据分割技术, 将整个原始数据分为 2 个子集, 通过比较不同比例随机分割数据的效果, 最终取总数据集的 80% 作为训练子集, 用来校正模型; 余下的 20% 作为评估子集, 用来验证模型。同时, 应用 Roc, Kappa 和 TSS3 种方法来评估和比较模型精度。

3 结果与分析

3.1 影响物种分布的主要环境要素比较

哪些因素会影响到物种的空间分布, 不同的模型有不同的结果。BIOMOD 能够计算提取出各个变量的相对重要性值。其值一般在 0~ 1 之间, 值越高意味着变量越重要, 0 意味着变量根本不重要, 当重要值大于 1 时表明该变量非常重要。本文仅以长芒草和虎榛子(*Ostryopsis davidiana*) 为例, 具体说明不同模型在分析物种分布影响因素方面的差异(表 1)。

表 **1** 各环境因子的重要值

物种	模型	坡度	坡向	地形部位	高程	年均最冷月气温	年均植被生长季节气温	年均气温	年均最热月气温	年均雨季降雨量	年均总降雨量	年均蒸发量	降雨季节变化	温度季节变化
虎榛子	ANN	0.097	0.016	0.003	0.695	0.000	0.000	0.000	0.000	0.129	0.367	0.510	0.000	0.000
	CTA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.977
	GAM	0.000	0.318	0.000	0.190	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.852
	GBM	0.002	0.013	0.001	0.002	0.002	0.000	0.000	0.001	0.003	0.074	0.001	0.003	0.532
	GLM	0.000	0.365	0.000	0.901	0.000	0.000	0.000	0.000	0.000	0.000	0.872	0.000	0.000
	MARS	0.000	0.154	0.000	0.000	0.000	0.000	0.000	1.319	0.000	1.013	1.194	0.000	1.283
	MDA	0.000	0.000	0.000	0.084	0.000	0.000	0.284	0.018	0.235	0.372	0.000	0.000	0.000
	RF	0.020	0.045	0.013	0.007	0.007	0.008	0.008	0.011	0.050	0.102	0.042	0.016	0.220
	SRE	0.190	0.087	0.049	0.182	0.074	0.087	0.060	0.145	—	0.479	0.270	0.365	0.577
长芒草	ANN	0.220	0.074	0.048	0.457	0.000	0.014	0.006	0.018	0.544	0.395	0.581	0.000	0.000
	CTA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.770	0.000	0.273	0.000
	GAM	0.131	0.163	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.733	0.000	0.000	0.000
	GBM	0.005	0.000	0.000	0.002	0.001	0.000	0.001	0.001	0.064	0.272	0.012	0.023	0.029
	GLM	0.000	0.000	0.000	0.000	0.147	0.000	0.000	0.000	0.307	1.331	0.000	0.000	0.000
	MARS	0.000	0.000	0.000	0.248	0.000	0.000	1.362	0.511	0.000	0.991	0.356	0.000	0.000
	MDA	0.021	0.000	0.000	0.453	0.000	0.000	0.122	0.000	0.772	0.540	0.000	0.324	0.000
	RF	0.016	0.006	0.000	0.019	0.016	0.014	0.017	0.013	0.066	0.072	0.104	0.051	0.035
	SRE	0.046	0.004	0.000	0.011	0.016	0.017	0.007	0.016	0.016	0.025	0.012	0.012	0.002

由表 1 可知, 各环境因子在不同物种不同模型中的重要性差别很大。如年均温在预测长芒草分布的 MARS 模型中非常重要, 而在其它模型中则根本不重要; 年均最热月气温在预测虎榛子的 MARS 模型中非常重要, 而在预测其它物种的 MARS 模型中(其在预测百里香、铁杆蒿、白羊草分布的 MARS 的重要值均为 0) 则根本不重要。通过模型比较, 可以清楚地看出影响物种分布的主要环境因素, 如长芒草在延河流域的

空间分布主要受年均降雨量的影响, 坡度、年均雨季降雨量、年均温度、年均蒸发量、高程也会对其有所影响, 而地形坡位对其几乎没有影响。同时, 所选的 9 个模型, 由于其本身的侧重点不同, 故选取的环境变量也不同并且数目差异较大, 如 RF 模型选取的环境变量比较多, 而 CTA 和 GLM 模型选取的环境因子都很少。

3.2　预测结果图比较

R—BIOMOD 不仅可以分析各模型中影响物种分布的主要环境因素, 并可以绘制模型的预测结果图(仅以长芒草为例, 如图 1), 图中越接近黑色部分表示物种的出现概率越大。尽管 8 个模型的环境因子、算法、假设、计算内容和预测性能各不相同, 但从分布图上看, 就预测的总体趋势而言, 其几乎在全流域都有分布, 且主要分布在延河流域中北部, 与温仲明、郝晓慧[1, 12] 等人的研究相符。但不同模型的预测结果图是有所差别的, 如在流域南部, 8 个模型的预测结果都不相同。这需要进一步通过模型精度比较, 选择出最优的模型并结合实际情况加以判断。

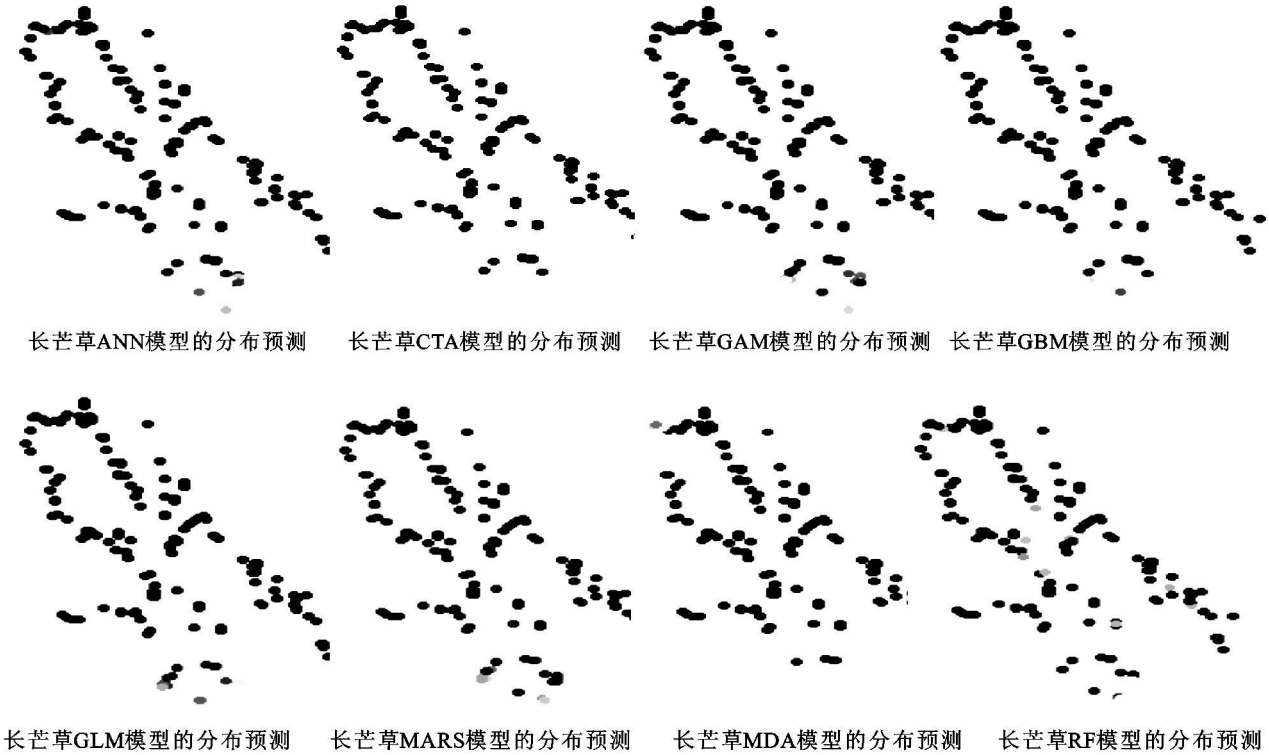


图 **1**　长芒草(***Stipa bungeana***) 各个模型的预测结果

3.3　模型预测精度比较

常用的模型评价指标有总体准确度(overall accuracy)、灵敏度(sensitivity)、特异度(specificity)、Kappa 统计量[20]、TSS[20] (true skill statistic) 和 AUC[20]。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 3 种方法相应的结果, 以此来评估不同模型对所选 8 个物种分布的模拟精度。

Roc 方法是使用 ROC 曲线来表示模型敏感度与特异性之间的平衡度, ROC 曲线是以 1– 特异度为横坐标, 以灵敏度为纵坐标绘制而成, 曲线下面积(AUC)的大小作为模型预测准确度的衡量指标。其值一般介于 0.5~ 1 之间, 值愈大, 表明模型愈优。由表 2 可知, 用同一模型模拟不同物种的分布, 用此方法评估 ANN 等 8 个模型对 8 个物种预测精度的平均值分别从 0.7~ 1 不等, 表明各个模型模拟精度有所差别, 以 RF 最优, GBM 和 ANN 效果良好, 而其余 4 个模型的模拟效果均不理想。同时, 就某个特定的物种来说, 不同模型对其分布的模拟效果也差别较大, 如铁杆蒿、茭蒿和白羊草, 只有 RF 和 GBM 模型模拟成功, 其它 6 个模型均模拟失败; 而所有的模型对百里香和大针茅的模拟效果都比较好。

Kappa 统计量综合考虑了物种分布率、灵敏度、特异度, 其值介于 0~ 1 之间, 值愈大, 表明模型愈优。由表 3 可知, 用此方法评估 ANN 等 9 个模型的平均值从 0.282 到 0.991, 模拟精度差别很大。除 SRE 模型精度不合格外, 其它 8 个模型的模拟精度都可以。模型的模拟效果从优到劣依次为: RF> GBM> ANN > MARS> GAM> CTA > MDA> GLM > SRE。

表 2 用 Roc 方法评估 9 个模型对 8 个物种分布模拟的值

物 种	ANN	CTA	GAM	GBM	GLM	MARS	MDA	RF	SRE
长芒草	0.893	0.728	0.808	0.878	0.744	0.695	0.806	1.000	—
百里香	0.944	0.822	0.942	0.953	0.929	0.937	0.921	1.000	—
铁杆蒿	0.696	0.675	0.720	0.838	0.635	0.693	0.610	1.000	—
茭 蒿	0.814	0.731	0.725	0.877	0.681	0.733	0.712	1.000	—
白羊草	0.791	0.659	0.713	0.809	0.669	0.793	0.616	1.000	—
大针茅	0.875	0.812	0.841	0.895	0.824	0.860	0.825	1.000	—
虎榛子	0.916	0.790	0.918	0.970	0.904	0.864	0.795	1.000	—
狼牙刺	0.819	0.701	0.777	0.850	0.737	0.808	0.772	1.000	—

表 3 用 Kappa 方法评估 9 个模型对 8 个物种分布模拟的值

物 种	ANN	CTA	GAM	GBM	GLM	MARS	MDA	RF	SRE
长芒草	0.566	0.499	0.492	0.511	0.423	0.491	0.482	0.985	0.241
百里香	0.666	0.570	0.674	0.681	0.619	0.622	0.643	1.000	0.523
铁杆蒿	0.277	0.335	0.280	0.413	0.204	0.184	0.202	0.985	0.043
茭 蒿	0.460	0.408	0.322	0.613	0.339	0.470	0.350	0.984	0.248
白羊草	0.479	0.292	0.322	0.485	0.276	0.456	0.182	0.993	0.126
大针茅	0.630	0.591	0.580	0.627	0.563	0.620	0.555	0.984	0.310
虎榛子	0.556	0.493	0.543	0.582	0.562	0.700	0.556	1.000	0.541
狼牙刺	0.469	0.331	0.399	0.537	0.333	0.440	0.373	1.000	0.226

TSS 方法考虑了冗余和替代性误差, 其值介于 0 ~ 1 之间, 值愈大, 表明模型愈优。由此方法评估 ANN 等 9 个模型的模拟精度, 其结果也有所差别, 其中除 RF 模拟精度非常好和 SRE 模拟精度非常差之外, 其它 7 个模型的模拟精度都一般。

所选的 9 个模型对长芒草等 8 个物种分布的模拟精度 3 种方法有所差异: Roc 方法均达到很好的模拟效果; TSS 和 Kappa 值亦取得不错的效果。其中, 9 个模型对百里香和虎榛子的模拟效果最为突出, 3 种方法评估值都较高, 均达到非常好的模拟效果。对长芒草、大针茅分布预测的 SRE 模型模拟失败, Kappa 值分别为 0.241, 0.310, 其它模型的模拟效果则较好。对铁杆蒿分布预测的 9 个模型评估的 3 种方法评估值均最低, 以 Kappa 值为例, 对铁杆蒿分布的 SRE 模型、MARS 模型、MDA 模型模拟的 Kappa 值分别为 0.043, 0.184, 0.202, 模拟效果失败, 而其他的模型除 RF 和 GBM 外, 模拟效果全部失败。但是无论从 3 种方法的 Kappa 值、TSS 值还是 Roc 方法的 AUC 值来看模型的模拟效果, RF 模型、GBM 模型和 ANN 模型都优于其他模型, SRE 模型模拟精度最差。

3.4 最优模型选择

由于不同模型关于物种分布和环境关系的假设和运算法则各不相同, 故很难对不同模型的预测进行比较。但 BIOMOD 则可以在过程中对模型进行直接比较, 其专门的函数将迭代在模型的每一次运算中, 根据所选的模型评价方法确定哪个模型具有最高的预测精度(表 4)。由表 4 可知, 不管使用 Roc, Kappa 或 TSS, 对研究区所选的 8 个物种, RF 都具有较好的预测性。

4 讨 论

本文针对模型众多且研究者难以选择的问题, 选择 ANN 等 9 个近年来在物种分布中应用的模型, 利用 BIOMOD 模型工具, 对延河流域长芒草等 8 种物种的地理分布进行比较研究。结果表明, 9 个模型对不同物种的模拟效果有所差异, 对百里香和虎榛子的模拟效果较好, 对铁杆蒿和白羊草的预测效果较差。用 Roc 曲线评价的 9 个模型对研究区 8 种物种的分布预测, 均达到较高的精度, 而另外两种指标评价的 9 个模型的模拟预测效果都在一般。总体而言, 所选的 9 个模型对长芒草等 8 种物种的分布预测各有千秋, 其中以 RF 模型的模拟效果最优, 模拟精度最高; GBM 模型、ANN 模型、MARS 模型也比较理想; GLM 模型、MDA 模型很差。

表 4　各个物种的最优模型

物种	Roc				Kappa				TSS			
	最优模型	交叉验证	灵敏度	特异度	最优模型	交叉验证	灵敏度	特异度	最优模型	交叉验证	灵敏度	特异度
长芒草	RF	0.731	0.996	1.000	RF	0.561	0.996	1.000	RF	0.510	0.996	1.000
百里香	RF	0.873	1.000	1.000	RF	0.576	1.000	1.000	RF	0.781	1.000	1.000
铁杆蒿	RF	0.471	0.996	1.000	RF	0.115	0.996	1.000	RF	0.156	0.996	1.000
茭蒿	RF	0.721	0.995	0.989	RF	0.450	0.989	1.000	RF	0.443	0.989	1.000
白羊草	RF	0.634	0.993	1.000	RF	0.286	0.993	1.000	RF	0.286	0.993	1.000
大针茅	RF	0.826	0.990	0.995	RF	0.624	1.000	0.989	RF	0.615	1.000	0.989
虎榛子	RF	0.881	1.000	1.000	RF	0.515	1.000	1.000	RF	0.720	1.000	1.000
狼牙刺	RF	0.743	1.000	1.000	RF	0.477	1.000	1.000	RF	0.420	1.000	1.000

当然利用 BIOMOD 构建物种分布模型,由于模型函数是以程序包的形式被用户调用,会在模型的选取、构造和运算时,存在方法比较单一、参数形式简单化的不足,然而,作为研究物种—环境关系的新平台,R 语言和 BIOMOD 提供了较多的构建模型的方法和参数形式,能够针对同一套数据构建不同的模型关系,并对预测结果进行优劣评估,对于模型选择具有非常重要的意义,也为进一步深入研究物种分布模型奠定了基础。

[参 考 文 献]

[1] 温仲明, 焦峰, 焦菊英. 黄土丘陵区延河流域潜在植被分布预测与制图[J]. 应用生态学报, 2008, 19(9): 1897-1904.

[2] Zaniewski A E, Lehmann A, Overton J M. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns[J]. Ecological Modelling, 2002, 32 (4): 261-280.

[3] Remm Kalle. Case-based predictions for species and habitat mapping[J]. Ecological Modelling, 2004, 177 (3/4): 259.

[4] Guisan A, Edwards J T C, Hastie T. Generalized linear and generalized additive models in studies of species distributions: Setting the scene[J]. Ecoogical Mcoloical, 2002, 157: 89-100.

[5] 温小霓, 蔡汝骏. 分类与回归树及其应用研究[J]. 统计与决策, 2007(23): 14-16.

[6] Moisen G G, Freeman E A, Blackard J A. Predicting tree species presence and basal area in Utah: A comparison of stochastic gradient boosting, generalized additive models, and tree-based methods[J]. Ecological Modelling, 2006, 199: 176-187.

[7] 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[J]. Ecological Modelling, 1999, 120: 337-347.

[8] Anderson R P, Lew D, Peterson A T. Evaluating predictive models of species distributions: Criteria for seleeting optimal models[J]. Ecological Molelling, 2003, 162: 211-232.

[9] Leathwicka J R, Elithb J, Hastiec T. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions[J]. Ecologica Modelling, 2006, 199: 188-196.

[10] 余卫东, 闵庆文, 李湘阁. 黄土高原地区降水资源特征及其对植被分布的可能影响[J]. 资源科学, 2002, 24(6): 55-60.

[11] 温仲明, 赫晓慧, 焦峰, 等. 延河流域本氏针茅 (*Stipa bungeana*) 分布预测: 广义相加模型及其应用[J]. 生态学报, 2008, 28(1): 192-201.

[12] 赫晓慧, 温仲明, 王金鑫. 基于 GAM 模型的延河流域主要草地物种空间分布及其与环境的关系[J]. 生态学杂志, 2008, 27(10): 1718-1724.

[13] 付坤俊. 黄土高原植物志: 第 5 卷.[M]. 北京: 科学文献出版社, 1989: 10-200.

[14] Hastie T J, Tibshirani R J. Generalised Additive Models[M]. London: Chapman and Hall, 1990: 3-10.

[15] Vayssieres M P, Plant R E, Allen-Diaz B H. Classification trees: an alternative non-parametric approach for predicting species distributions[J]. Journal of Vegetation Science, 2000, 11: 679-694.

[16] 李斌, 张金屯. 黄土高原地区植被与气候变量的关系分析[J]. 生态学报, 2003, 23(1): 82-89.

[17] Heinrich Walter. Vegetation of the Earth[M]. Beijing: Science Press, 1984: 5-10.

[18] Venables W N, Smith D M. The R Development Core Team. An Introduction to R [M]. 2009: 34-98.

[19] Emmanuel Paradis. R for Beginners[M]. 2005: 37-71.

[20] Wilfried Thuiller, Bruno Lafourcade, Miguel Araujo. ModOperating Manual for BIOMOD[M]. 2009: 10-90.