

北大西洋大型常见冷水柳珊瑚潜在分布建模

佟瑞菊, 袁玉珠

(福建工程学院 交通运输系, 福建福州 350118)

摘要: *Paragorgia arborea* 及 *Primnoa resedaeformis* 是北大西洋最常见的两类大型冷水柳珊瑚, 该两种柳珊瑚显著增加了底栖环境的生境复杂度。其分布信息对资源管理及保护极其重要, 但难以获取。本文基于随机森林模型及最大熵模型预测其在北大西洋 Traena 区域的潜在分布。预测精度评价表明随机森林模型较最大熵模型具有更好的预测性能。平均曲率(90 m 尺度)是随机森林模型预测两种柳珊瑚分布的主导因子, 表明其与两种柳珊瑚的分布具有较强生态相关性。预测结果显示两种柳珊瑚趋向于分布在珊瑚礁体上。本预测成果可为该区域冷水珊瑚保护提供决策辅助信息。模型对比研究可为大型底栖动物分布建模的模型选择提供依据。

关键词: 冷水柳珊瑚; 潜在分布建模; 随机森林模型; 最大熵模型

中图分类号: Q141

文献标志码: A

文章编号: 1672-4348(2017)04-0393-06

Distribution modelling of large common cold-water gorgonians in North Atlantic

Tong Ruiju, Yuan Yuzhu

(School of Transportation, Fujian University of Technology, Fuzhou 350118, China)

Abstract: *Paragorgia arborea* and *Primnoa resedaeformis* are two large cold-water coral species commonly occurring in North Atlantic, which significantly increase the complexity of the benthic habitat. Mapping their distribution is fundamental for resource management and conservation, but is difficult (given their remoteness). In this study, their potential distribution at Traena Reef complex of the North Atlantic were predicted based on random forest (RF) and MaxEnt models, respectively. The RF prediction was shown to outperform the MaxEnt prediction. Mean curvature at an analysis scale of 90 m is the most useful terrain variable in RF prediction, which is ecologically correlated to the distribution of the two species. The two species were predicted to occur on the reefs. The result can contribute to cold-water coral protection at the study area. Comparing model performances is useful to provide for distribution modelling of large benthic species.

Keywords: cold-water gorgonians; potential distribution modelling; random forest (RF) model; MaxEnt

深海柳珊瑚 *Paragorgia arborea* 和 *Primnoa resedaeformis* 是北大西洋海域常见的两类最大型柳珊瑚物种, 其复杂的树状结构可为固着动物幼虫提供附着基质, 为多种底栖动物包括鱼类资源提供寄生、繁殖、索饵及躲避被捕食的场所^[1-2], 极

大地增加了深海生境复杂度, 具有极其重要的生态意义。深海柳珊瑚生长缓慢, 其树状结构易碎, 极易受底拖网捕鱼、延绳捕鱼、刺网捕鱼、石油开采等人类活动的影响。加强冷水珊瑚管理与保护刻不容缓^[3-4], 有效的物种分布信息是管理与保护

收稿日期: 2017-07-01

基金项目: 福建省中青年教育科研项目(JAT160323); 福建工程学院科研启动基金(GY-Z15123)

通讯作者简介: 佟瑞菊(1982-), 女, 山东莒县人, 讲师, 博士, 研究方向: GIS技术应用、物种分布建模。

的前提,然而深海观测数据获取成本高,该方面数据资源匮乏。基于有限观测数据进行冷水珊瑚生境适宜度预测,利用物种潜在分布表征信息反演物种实际分布信息^[3, 5-6],可为有关部门提供优化海洋监测及海洋资源管理的决策辅助信息^[7],对开展深海生态保护具有重要意义。此外,通过评价和比较不同模型的预测精度可为海洋底栖动物潜在分布建模研究的模型选择提供依据。

本研究以北大西洋 Treana 礁区两类主要的大型冷水柳珊瑚 *P. arborea* 及 *P. resedaeformis* 为例,基于 GIS (Geographic Information System, 地理信息系统) 技术、最大熵模型及随机森林模型预测其潜在分布,分析建模关键影响因子,研究冷水珊瑚的分布受环境影响的机理,通过精度评价分析比较两种模型的预测性能。

1 物种分布预测模型

根据是否需要物种非分布点数据,可将物种分布模型大致分为两类:(1) 仅需物种分布数据的模型,主要包括 BIOCLIM (geographic information system, 地理信息系统)、DOMAIN, 生态位因子分析模型 ENFA, 限制变量与环境适宜性模型 LIVES 等;(2) 需要物种分布/非分布数据的模型,包括统计学模型广义线性模型 GLM (generalized linear model), 广义加法模型 GAM (generalized additive model), 多元自适应回归样条 MARS (multivariate adaptive regression splines) 等,机器学习模型人工神经网络 ANN (artificial neural network), 分类回归树 CART (classification and regression tree), 推进式回归树 GBM (generalized boosted regression models), 基于规则集的遗传算法模型 GARP (genetic algorithm for rule-set production), 最大熵模型 (maximum entropy, MaxEnt), 支持向量机模型 SVM, 随机森林模型 (random forest, RF) 等^[8-9]。基于物种分布/非分布数据的模型预测性能一般优于仅需物种分布数据的模型。MARS 和 GAM 原理相似,两者性能优于 GLM,三者适用于数值型环境变量;ANN、CART、GBM、MaxEnt、SVM 及 RF 这些复杂模型对数值型和分类型变量均适用,能够准确提取输入数据的隐藏特性,捕捉细节部分,但可导致过度拟合问题^[10-11]。这些模型已被应用于当前一系列研究中,其中 RF 模型表现出更优的预测性能^[12-14]。

随机森林模型 RF 是一种组合分类模型,由大量决策树分类模型 $\{h(X, \lambda_k), k = 1, \dots\}$ 组成,其中,参数集 $\{\lambda_k\}$ 是独立同分布的随机向量,对于给定自变量 X ,每个决策树分类模型均具有一票投票权,从而最终选择最优的分类结果^[15]。RF 首先通过自助 (bootstrap) 抽样从原始训练集中抽取 k 个样本,对 k 个样本分别建立 k 个决策树模型,从而得到 k 种分类结果,进而基于这 k 种分类结果对每个记录进行投票表决从而决定其最终分类,通过对大量分类树的汇总提高了模型的预测精度^[16]。RF 模型可用于分类、聚类、回归和生存分析、评估变量的重要性、检测数据中的奇异值、对缺失数据进行插补等,其对离群值不敏感^[17]。随机森林优于计算变量间的非线性关系,可反应变量间的交互作用^[18],其运算速度很快,在处理大数据时表现出优异的计算性能^[19],是取代神经网络等传统机器学习方法的新模型。因此,适用于模拟物种分布和环境变量之间复杂的相互关系。

目前一系列底栖动物分布建模研究中,通常采用生态位因子分析模型 ENFA、基于规则集的遗传算法模型 GARP 或最大熵模型 MaxEnt 3 种常规方法,其中以 MaxEnt 预测性能最优,究其原因可能是模型理论及应用的复杂性不太容易被海洋生态科学及海洋生物地理科学研究者所接受和采纳。最大熵原理认为推算某一未知事件发生的概率时,应采用的概率分布必须是在满足所有已知信息的前提下能够使得信息熵取得最大值的那个概率分布,只有这样才能做出无偏推测。最大熵模型在物种分布建模中的应用是假设多维自然环境空间中存在某一物种的一定数量的分布点,这些分布点与多个环境变量相关联,通过提取与分布点所关联的环境变量,基于最大熵算法推算该物种的生态需求,并将计算结果投射到不同的空间和时间的地理区域中,从而预测该物种在某一区域的潜在分布范围。

本研究拟采用最大熵模型 MaxEnt,并引入随机森林模型 RF,预测 *P. arborea* 及 *P. resedaeformis* 的潜在分布,评价比较二者的预测性能。Kappa 统计量 (Kappa statistics, 卡帕统计量)、TSS (true skill statistics, 真实技巧统计量) (true skill 统计) 及 ROC 曲线 (receiver operating characteristic curve, 接受者操作特征曲线) 是评价模型预测精

度常用的3种统计方法。TSS统计具有Kappa统计的所有优点,同时在独立于出现率方面优于Kappa统计方法。基于ROC曲线可计算AUC值(area under curve, ROC曲线线下面积),该值独立于任何阈值,作为单一值用于评价模型预测性能。本文采用上述2种方法对预测结果进行评价。

2 Traena 礁区 *P. aborea* 及 *P. resedaeformis* 的潜在分布预测

Traena 礁区位于挪威大陆架中部,水深大约300 m的区域。该区域有一定数量的礁体沿东西方向略偏北延伸分布,礁体长度可达150 m(图1)。经载人潜水器探测,发现该区域有大量大型冷水柳珊瑚 *P. aborea* 及 *P. resedaeformis* 分布(图1)。本文所采用的数据主要包括多波束测深数据,潜水器搭载影像数据及相应的卫星定位数据。这些数据主要来自于挪威国家海洋研究所及欧盟

通过对影像数据进行后处理,沿载人潜水器的行进路线,采集了冷水珊瑚的出现点数据,共记录有58个 *P. aborea* 出现点及83个 *P. resedaeformis* 出现点(图1)。分别随机选取58及83个背景数据点作为两类珊瑚未出现点数据,连同出现点数据共同组成了两类珊瑚的样本数据。在建模过程中,随机选取80%的样本数据作为训练数据用于建模,另外20%样本数据用于评价模型预测精度。模型评价结果如表1所示。

预测精度评价结果表明最大熵模型 MaxEnt 及随机森林模型 RF 的预测精度均高于随机预测值0.5($P < 0.0001$)。根据表1,由TSS及ROC两种统计方法评价模型预测精度,结果均显示RF模型预测精度高于MaxEnt模型,RF模型具有更优预测性能。

表1 模型预测精度评价结果

Tab.1 Precision evaluation of predictive models for two species

项目	TSS_PA	ROC_PA	TSS_PR	ROC_PR
最大熵模型 MaxEnt	0.833	0.917	0.706	0.853
随机森林 模型 RF	1	1	0.941	0.997

基于随机森林模型 RF 及最大熵模型 MaxEnt 给出的预测结果均为连续数值,结果如图2所示。由图2可见,基于MaxEnt模型预测的 *P. aborea*

HERMES项目(hotspot ecosystem research on the margins of European seas,欧洲边缘海域生态系统热点研究项目)。

海底地形特征可通过控制洋流机制来影响有机质颗粒及珊瑚虫分布,并通过影响海底沉积物分布(与珊瑚初始附着关系密切),进而间接影响珊瑚分布。海底地形的变量,包括坡度、曲率(含平面曲率、剖面曲率及平均曲率)、水深位置指数、粗糙度、地形粗糙指数等,可作为海洋底层洋流速度或底质类型的指示参数。本研究采用地形参数坡度、平均曲率、剖面曲率、平面曲率、粗糙度、地形粗糙指数来描述地形特征。根据研究区域内主要地形特征的空间尺度,确定在30、50及90 m三个尺度上进行地形分析,并在这三个尺度上计算了坡度及曲率参数,在30 m尺度上计算了粗糙度、地形粗糙度指数,连同水深参数计算在内,共有15个参数作为因变量用于建模。

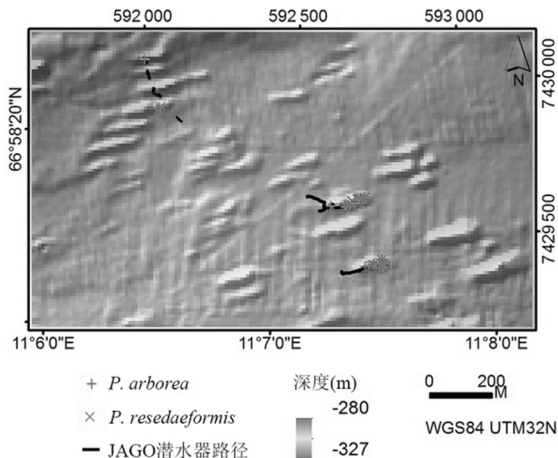


图1 北大西洋挪威海域 Treana 礁区水深数据、JAGO 载人潜水器潜水路径、两种柳珊瑚 *P. aborea*、*P. resedaeformis* 的出现点数据(单位:m)

Fig.1 Bathymetry data of the study area with the occurrence data of *P. aborea* and *P. resedaeformis*. The lines are the transect sections of the three JAGO dives with useful video data (unit: m)

及 *P. resedaeformis* 的潜在分布具有相似的特征,同样,基于RF模型预测的两种柳珊瑚的潜在分布也具有相似趋势。对比不同模型间预测结果,MaxEnt模型预测结果显示两种深海柳珊瑚具有相对较广的潜在分布区域,包含整个突起的礁体,而RF预测的潜在分布区范围较窄,主要是沿礁体的上部高处分布。此外,MaxEnt预测结果中两

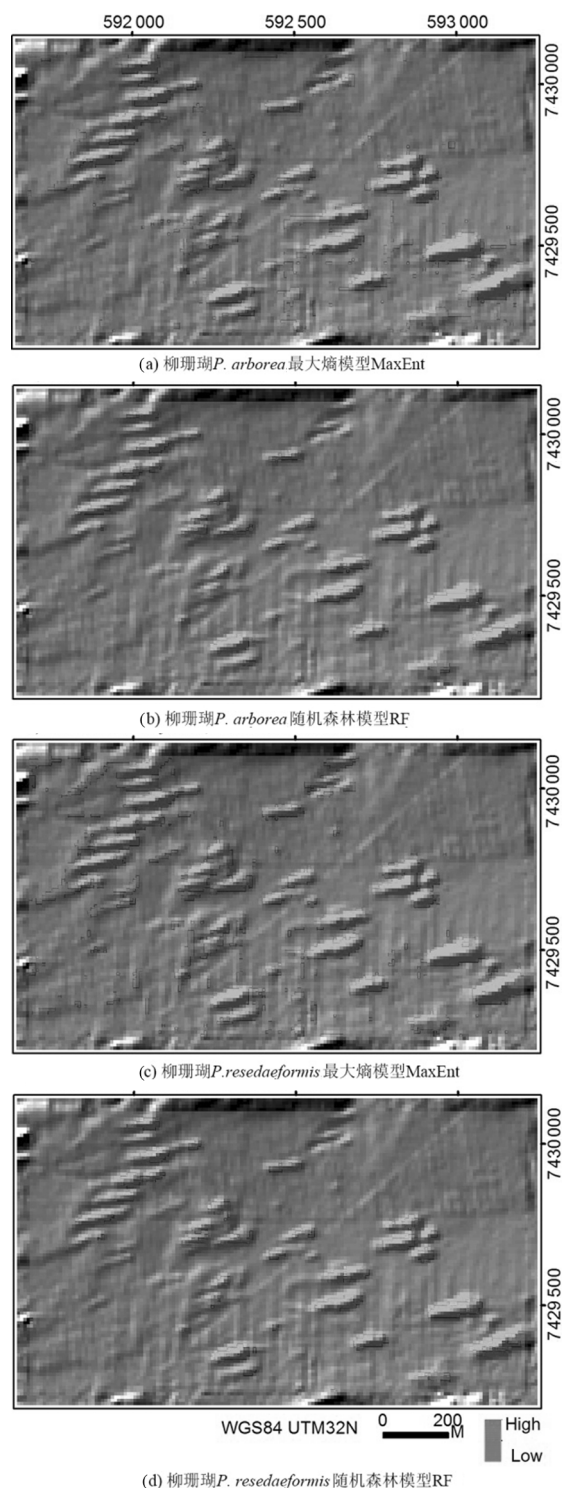


图 2 基于最大熵模型 MaxEnt 及随机森林模型 RF 预测的两种柳珊瑚的潜在分布 (单位: m)

Fig.2 Potential distributions of *P. arborea* and *P. resedaeformis* predicted based on MaxEnt and RF models, respectively (unit: m)

种珊瑚沿南北方向上的条状潜在分布区域与原始数据中可见的沿南北方向上的条状分布系统误差的方向位置一致,因此这部分区域的预测值很可能是受原始数据的误差影响所致。由此可见,MaxEnt 模型的预测结果对于原始数据的误差较敏感,而 RF 模型的预测结果则相对稳定。综合图 2 的预测结果,可以看出两类柳珊瑚的潜在分布区均沿东西偏北延伸的礁体分布。该预测结果与从 JAGO 潜水器三次潜水的影像资料中观察到的相应礁体上的情况是一致的(图 1)。Treana 礁区的这类细长形态的礁体在挪威海域具有典型性。这些礁体通常由该区域的主要冷水造礁珊瑚种 *Lophelia pertusa* 生长造礁而成。在迎向底层洋流一端,珊瑚构架上生长有半月形的活体珊瑚群,可以看作是礁体的活的头部,后面是新近死亡的较完整的珊瑚构架,再后面是成块团状分布的较早死亡的珊瑚构架,最后面是呈狭长分布覆盖有珊瑚碎块碎屑的尾巴。在 *Lophelia* 礁的头部有活体珊瑚覆盖的部分通常有较少柳珊瑚附着分布,在死亡的珊瑚构架上则多发现有柳珊瑚及其他海洋底栖动物比如贝壳类、鱼类、虾类等的分布,在狭长分布的珊瑚碎屑上则更多见 *P. resedaeformis* 的分布。

预测模型给出了各地形变量的重要性评价数据(图 3),由图 3 可见,对于 *P. arborea* 的分布预测中,平均曲率在 90 m 的分析尺度上对两种模型预测均含有最多有用信息,此外,平面曲率在 90 m 的尺度上对 MaxEnt 预测也含有较多有用信息。对于 *P. resedaeformis* 的分布预测中,平均曲率 90 m 对于 RF 模型,地形粗糙指数 30 m 及剖面曲率 50 m 对于 MaxEnt 预测具有最多有用信息。平均曲率 90 m 及平面曲率 90 m 描述抓取了珊瑚礁尺度的地形特征,前者突出显示珊瑚礁礁体的上部,后者突出描述珊瑚礁的最高处山脊部(图 4(a), 4(b))。剖面曲率 50 m 抓取了珊瑚礁较细节的局部变化特征,突出显示了珊瑚礁沿下坡方向坡度增加的部位(通常位于斜坡上部)(图 4(c))。地形粗糙指数 30 m 突出描述了珊瑚礁高度变化大的部位,即珊瑚礁的边坡(图 4(d))。由图 2 可见,两种深海柳珊瑚的分布与珊瑚礁地形特征是密切相关的,包括珊瑚礁的边坡及顶部。一方面,柳珊瑚是底栖附着动物,需要硬质底质进

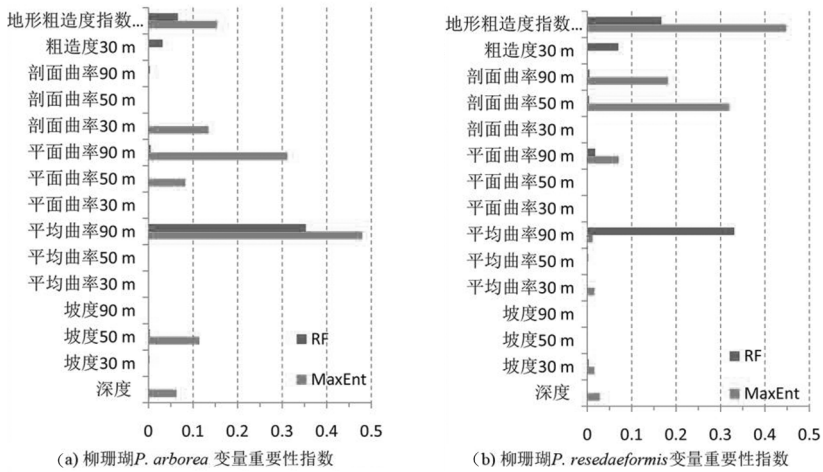


图 3 各地形变量的重要性评价

Fig.3 Index of terrain variable importance

行附着,比如珊瑚礁构架、大块石头等,相对于珊瑚礁周围遍布的松软沉积物,珊瑚礁给柳珊瑚提供了良好的附着基质;另一方面由于珊瑚礁生态系统主要的食物颗粒来源于底层洋流所携带的有

机质颗粒。突起的珊瑚礁结构使得迎面而来的底层洋流速度加速,从而在单位时间内为珊瑚礁上的底栖动物提供更多的与食物颗粒相遇,捕捉食物颗粒的机会。

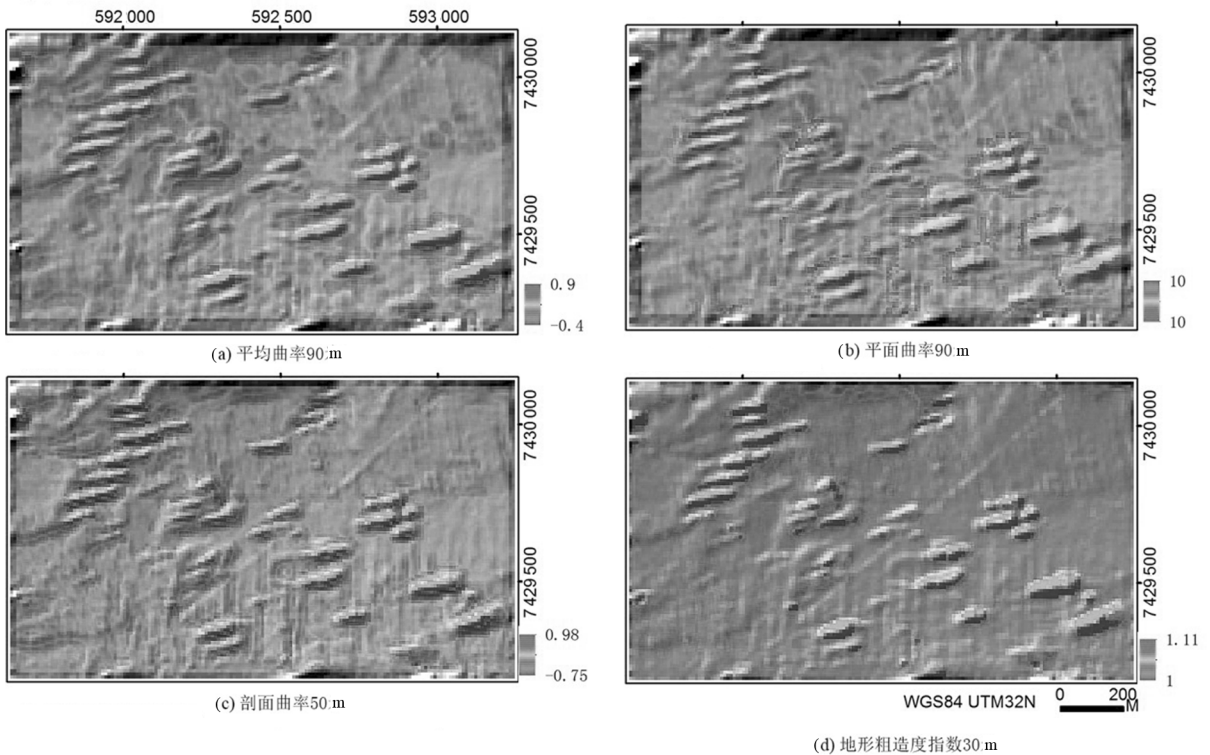


图 4 关键地形变量

Fig.4 Key terrain variables

3 结论

本研究采用物种分布模型最大熵模型 MaxEnt 及随机森林模型 RF, 基于多尺度地形变量及 GIS 技术, 预测北大西洋海域常见的两类深海柳珊瑚的潜在分布。模型精度评价结果显示相较于 MaxEnt 模型, RF 模型具有更高的预测精度,

且对于原数据的系统误差具有更好的稳定性。两类模型预测结果均显示在 Traena 礁区, 2 类柳珊瑚趋向于分布在珊瑚礁体上。RF 预测显示, 地形变量中的平均曲率在 90 m 分析尺度上与两类深海柳珊瑚的分布具有最强的生态相关性, 该结果可为后续类似尺度上的柳珊瑚或其他底栖动物预测的自变量参数的选择提供参考。

参考文献:

- [1] Kutti T, Bergstad O A, Fosså J H, et al. Cold-water coral mounds and sponge-beds as habitats for demersal fish on the Norwegian shelf[J]. Deep-Sea Res Pt II, 2014, 99: 122-133.
- [2] Freiwald A, Huehnerbach V, Lindberg B, et al. The Sula feef complex, Norwegian shelf[J]. Facies, 2002, 47: 179-200.
- [3] Fosså J H, Skjoldal H R. Conservation of cold-water coral reefs in Norway[M]. New York: Oxford University Press, 2010, 215-240.
- [4] Fabri M C, Pedel L, Beuck L, et al. Megafauna of vulnerable marine ecosystems in French mediterranean submarine canyons: Spatial distribution and anthropogenic impacts[J]. Deep-Sea Res Pt II, 2014, 104: 184-207.
- [5] Elith J, Phillips S J, Hastie T, et al. A statistical explanation of MaxEnt for ecologists[J]. Divers Distrib, 2011, 17: 43-57.
- [6] Huang Z, Brooke B, Li J. Performance of predictive models in marine benthic environments based on predictions of sponge distribution on the Australian continental shelf[J]. Ecol Inform, 2011(6): 205-216.
- [7] Rengstorf A M, Yesson C, Brown C, et al. High-resolution habitat suitability modelling can improve conservation of vulnerable marine ecosystems in the deep sea[J]. J Biogeogr, 2013, 40: 1702-1714.
- [8] 李国庆, 刘长成, 刘玉国, 等. 物种分布模型理论研究进展[J]. 生态学报, 2013, 33(16): 4827-4835.
- [9] 许仲林, 彭焕华, 彭守璋. 物种分布模型的发展及评价方法[J]. 生态学报, 2015, 35(2): 557-567.
- [10] 杨若男, 盛炎平. 应用于物种分布模型的多种算法[J]. 大学教育, 2016(5): 120-121.
- [11] 刘芳, 李晟, 李迪强. 利用分布有/无数据预测物种空间分布的研究方法综述[J]. 生态学报, 2013, 33(22): 7047-7057.
- [12] 桑满杰, 卫海燕, 毛亚娟, 等. 基于随机森林的我国荞麦适宜种植区划及评价[J]. 山东农业科学, 2015(7): 46-52.
- [13] 张雷, 王琳琳, 张旭东, 等. 随机森林算法基本思想及其在生态学中的应用——以云南松分布模拟为例[J]. 生态学报, 2014, 34(3): 650-659.
- [14] 毕迎风, 许建初, 李巧宏, 等. 应用 BioMod 集成多种模型研究物种的空间分布——以铁杉在中国的潜在分布为例[J]. 植物分类与资源学报, 2013, 5(5): 647-655.
- [15] 方匡南, 吴见彬, 朱建平, 等. 随机森林方法研究综述[J]. 统计与信息论坛, 2011, 26(3): 32-37.
- [16] Breiman L. Random forests[J]. Mach Learn, 2001, 45: 5-32.
- [17] 张雷, 王琳琳, 张旭东, 等. 随机森林算法基本思想及其在生态学中的应用——以云南松分布模拟为例[J]. 生态学报, 2014, 34(3): 650-659.
- [18] Peters J, Baets B D, Verhoest N E C, et al. Random forests as a tool for ecohydrological distribution modeling[J]. Ecol Model, 2007, 207: 304-318.
- [19] Cutler D R, Edwards T C Jr, Beard K H, et al. Random forests for classification in ecology[J]. Ecol, 2007, 88(11): 2783-2792.

(责任编辑: 肖锡湘)