

MaxEnt 模型国内外研究进展及其在不同气候背景 与区域尺度下的应用

麻雯迪¹, 李世宗²

(1. 西南林业大学生物多样性保护学院, 云南 昆明 650224; 2. 云南森林自然中心, 云南 昆明 650225)

摘要: 物种分布模型 (Species Distribution Models, SDMs) 作为生物地理学及宏观生态学的重要研究方法, 在生命科学及环境科学领域发挥重要作用, 在诸多物种分布模型中, 最大熵模型 (Maximum Entropy model; MaxEnt) 因其数据要求较低、预测精确率较高、使用步骤简便等优点, 在众多物种分布模型中脱颖而出, 也在不断发展过程中得以优化及改进。通过对目前公开发表的 MaxEnt 模型相关文献及著作进行分析, 尝试从 MaxEnt 模型国内外研究进展、不同气候背景下的应用、区域方面的应用三个角度叙述其发展历程及优化提升。在综合分析基础上, 采用扩大预测范围及预测数据样本量建立模型的方法策略, 从而解决物种数据量过少或评估调查区域生物多样性空白这一问题。

关键词: MaxEnt 模型; 研究进展; 不同气候背景; 区域应用

Development of maximum entropy model at home and abroad and its applications in different climatic backgrounds and regional scales

MA Wendi¹, LI Shizong²

(1. Faculty of Biodiversity Conservation, Southwest Forestry University, Kunming 650224;

2. Yunnan Forests and Nature Center, Kunming 650225)

Abstract: As an important research method in biogeography and macroecology, Species Distribution Models (SDMs) play an important role in the fields of life science and environmental science. Among many species distribution models, Maximum Entropy model (MaxEnt) stands out among them because of its advantages such as its low data requirements, high predicted accuracy, easy-to-use operational steps and so on. By reading the related literatures and works about Maximum Entropy model currently published, we tried to describe its development process and optimization from three perspectives: the development of the Maximum Entropy model at home and abroad, its applications in different climatic backgrounds and regional scales. On the basis of comprehensive analysis, we adopt an idea of expanding the prediction range and the sample size of prediction data to build a model to solve the problem of too little data and the investigation of biodiversity in blank areas.

Key words: maximum entropy model; research progress; different climatic backgrounds; the application in areas

物种分布模型 (Species Distribution Models, SDMs) 是一种基于物种的分布数据及可能与其相关的环境因子数据, 依据分布点所在的环境信息评估物种的生存需求, 再投影到选定的研究范围内, 从而推断物种适宜生境分布的数学模型^[1-4]。物种分布模型几乎被应用于生命科学及环境科学涵盖的所有分支学科中^[5]。依据 Thuiller 等^[5] 2009 年研究中的指标, 在最近更新的 ISI Web of Science^[6] 中使用 “species distribution models”、“niche models”、“habitat models”、“bioclimatic models” 等为关键词检索时发现, 迄今为止近 20

年的时间内, 物种分布模型被广泛应用于环境科学、生命科学等领域, 其中动物科学、生物多样性保护、分子生物学、遗传学、进化学、海洋科学、气象学等方面的研究不断涌现。这些模型在气候变化对物种分布的影响、野生动植物保护、病虫害防治、入侵物种监测、物种适宜性分析等研究领域发挥着重要的作用^[7]。

物种分布模型的发展, 始于 BIOCLIM 模型的开发和应用^[8-9]。作为生物地理学及宏观生态学研究的重要方法之一, 物种分布模型在国内外各研究领域得到了广泛的应用。针对不同的建

收稿日期: 2022-12-20

作者简介: 麻雯迪, 女, 本科, 研究方向为野生动物与自然保护区管理。

通讯作者: 李世宗, 男, 高级工程师, 主要从事野生动物的收容救护、疾病治疗、恢复饲养、人工繁育、资源调查、野外放归、跟踪监测、科普教育和专家咨询工作。

模目的、物种生态位特征以及建模数据基础, 研究者建立了不同算法的模型^[7]。如表面分布区
分室模型 (Surface Range Envelope, SRE)、分类树
分析 (Classification Tree Analysis, CTA)、随机森林
(Random Forest, RF)、广义线性模型 (Generalized
Linear Models, GLM)、广义加性模型 (Generalized
Additive Models, GAM)、基于规则集的遗传算法
(Genetic Algorithm for Rule-set Prediction, GARP)
等。Elith 等^[10]利用 16 种建模方法, 分别对来自
6 个地区的 226 个物种进行了物种分布预测, 并
对 GAM、GARP 和 BIOCLIM 模型进行了方法完善,
探索了当时尚未推广的机器学习方法——最大熵
模型 (Maximum Entropy Model, MaxEnt) 和增强回归
树 (Boosted Regression Trees, BRT)。在此之后, 物
种分布模型的发展有了较大的飞跃, 模型的类型
及运行方式不断提升优化, 模型的应用也从单一
模型的使用逐渐演变为多重模型组合运用。

1 最大熵理论

目前 MaxEnt 模型是诸多物种分布模型中公
认的性能较良好的模型之一, 在各类物种分布研
究中使用率较高^[11-12]。

熵是 19 世纪中叶德国物理学家 Clausius 作
为热力学的一个概念而提出来的, 用以描述能量
流动过程中转化的量和转化方向^[13]。之后, 熵
的概念又经推广, 在各个领域得到应用发展。
1957 年, Jaynes^[14]基于熵增原理提出了最大熵
理论。该理论认为: 在已知条件下, 熵最大的事
物接近该事物最真实的状态。其核心要点是: 最
大熵状态不排除初始数据约束条件下的任何可能
性, 最大熵分布是与给定信息兼容的最广泛的分
布。在生物学中, 一个实测存在的系统具有“耗
散”的特征。耗散使系统的熵不断增加, 直至该
生命系统与环境的熵最大, 此时系统与环境之
间的关系达到平衡。在物种潜在分布的相关研究
中, 物种与其生长环境即可视为一个系统。通过
计算系统具有最大熵时的状态参数确定物种和环
境之间的稳定关系, 并以此估计物种的分布, 这
就是 MaxEnt 模型的理论基础^[9,11]。

MaxEnt 模型的使用通过结合物种已知的地
理分布信息 (经度、纬度数据) 和相应的环境因
子数据, 将研究区域内物种分布点所在的气候数
值与该区域的其他地区的气候数值进行相似度比

较, 从而预测该区域内物种可能分布的范围及其
适生度^[4,11]。相比于与 MaxEnt 模型相似性较强
的一些建模方法, 特别是 GLM、GAM、人工神
经网络 (Artificial Neural Networks, ANN) 等建模
方法, MaxEnt 模型的诸多优点使其得以迅速推广:
(1) MaxEnt 模型运行要求数据仅为物种存在数据
及整个研究区域的环境信息, 当训练数据量有限
时, 该建模方法的优势凸显; (2) 该模型利用连
续数据和分类数据, 合并不同变量之间的交互,
综合环境变量对物种分布进行预测, 而 GLM 和
GAM 则是在变量之间没有交互作用的情况下, 利
用模型的可加性解释每个变量的适用性; (3) 为
避免模型过度拟合, 该模型应用了正则化 (regu
larization) 训练增益, 以确定环境变量对物种分
布增益的贡献大小; (4) 模型结果输出是连续
的, 可以对不同区域的建模适用性进行精细区
分, 如果需要二元预测, 在选择阈值时具有很
大的灵活性, 若应用于保护规划, 则预测的相
对环境适用性的细微差别对于保护区规划将会
具有较高的价值; (5) 该模型概率分布具有简
明的数学定义, 易于分析; (6) MaxEnt 模型目
前已开发了有效的确定性算法, 保证收敛到最
优 (最大熵) 概率分布^[11]。2013 年, Merow 等^[15]
系统地整理并介绍了 MaxEnt 模型的工作原理
及建模参数设置的意义, 为进一步认识、优化、
提升 MaxEnt 模型奠定了基础。

2 MaxEnt 模型应用

2.1 MaxEnt 模型国内外研究进展

MaxEnt 模型相较于其他模型的预测更加准
确, 操作简单、适用性强, 因此其具有较大的
提升空间^[4,11,16]。故而, MaxEnt 模型在不
断地推广中得以优化, 并展现出其优势。

MaxEnt 模型应用推广初期, 利用物种已知
分布数据对其潜在分布进行预测即展现了良好
的预测效果。Phillips 等^[11]2006 年第一次
使用并推行此模型时, 以褐喉三趾树懒 (*Bradypus*
variegatus) 和森林小稻鼠 (*Microrhizomys*
minutus) 为研究对象, 通过 MaxEnt 模型与
GARP 模型建立比较, 展现了 MaxEnt 模型较
GARP 模型更高的预测精确度。Peterson 等^[17]
于 2007 年基于 3 个北美鸟种 (*Caprimulgus*
vociferus; *Coccyzus americanus*; *Zenaida*
macroura) 的已知分布信息, 确定出三物种
分布区域的中值经纬度, 以此为原

点建立二维坐标系,划分出对角线(第二、四象限)与非对角线(第一、三象限)象限,以对角线象限内数据点为训练数据,作为空间分层子样本,非对角线象限内数据点为测试与评估数据,分别建立 MaxEnt 模型与 GARP 模型,结果再次印证了 MaxEnt 模型较 GARP 模型更高的预测精确度,以及预测分布在未采样区域的可转移性。

国内的学者则通过类比的方法,对 MaxEnt 模型的性能进行了评估。王运生等^[18]以相似穿孔线虫(*Radopholus similis*)为研究对象,应用受试者工作特征曲线(Receiver Operating Characteristic curve, ROC)对 CLIMEX 模型(Climex dymex, Climex)、邻域模型(Domain Model, DOMAIN)、GARP 模型、MaxEnt 模型、BIOCLIM 模型在内的 5 种模型的预测结果进行了比较分析,受试者工作特征曲线下面积(Area Under Curve, AUC)以 MaxEnt 模型输出结果的值最大。由此表明,5 种模型中 MaxEnt 模型预测效果最优。李明阳等^[19]通过逻辑斯蒂回归(Logistic Regression, LR)、分类与回归树模型(Classification and Regression Tree, CART)、GARP 模型、MaxEnt 模型 4 种途径建立美国外来入侵物种斑马纹贻贝(*Dreissena polymorpha*)潜在生境预测模型,从 AUC 值、Pearson 相关系数、Kappa 值 3 个方面检验模型预测精度,结果显示,4 个生态位模型预测精度均达到优良水平,而其中 MaxEnt 模型在物种现实生境模拟、主要生态环境因子筛选、环境因子对物种生境影响的定量描述方面都表现出了优越的性能。后续的研究工作中,学者们囊括了更为广泛的物种分布模型,如生态位因子分析(Ecological Niche Factor Analysis, ENFA)、马氏典型模型(Mahalanobis typicality, MAHAL)、支持向量机(Support Vector Machines, SVM)等,对各模型的预测精确度进行对比,综合实际应用结果得出,机器学习模型较非机器学习模型的置信区间较小,对物种分布的预测更加集中,数据更加整齐,受随机变量的影响更小,更加稳定,其中 MaxEnt 模型表达出更为优异的预测精确度与稳定性^[20-23]。

在 MaxEnt 模型应用发展阶段,物种已知分布数据的样本存在偏差的问题与模型参数设置受到诸多研究者的关注。Phillips & Dudík^[24]基于 Elith 等^[10]2006 年的研究,运用 6 个地区的 226 个物种对 MaxEnt 模型参数设置进行调整优化,

引入了铰链特征(Hinge features),在训练数据集中建立更复杂的关系,并采用了 Logistic 输出结果格式,给出预测物种存在概率的估计,最后探索了应对样本选择偏差及减少模型构建时间的背景抽样策略。其后,Phillips 等^[25]又基于 226 个物种分布数据以 GAM、BRT、多元自适应回归(Multivariate Adaptive Regression Splines, MARS)、MaxEnt 模型 4 种建模方法研究采样偏差问题,在已知样本分布数据情况下,展示了如何通过使用与发生数据具有相同偏差的背景数据来处理样本选择偏差,以及当抽样分布未知时,可以通过组合所有使用相同方法收集或观察的目标物种组的发生记录进行近似抽样分布。最终得出,在使用物种已知分布数据模拟物种分布时,背景数据的选择与建模方法的选择同样重要,提高调查中对空间偏见影响的认识以及可能的建模补救措施,将极大地改善对物种分布的预测。

Anderson & Gonzalez^[26]2011 年的研究指出:用于建立物种分布模型的位点数据因调查人员、技术、强度、时间段不一,而对整合数据建立模型所产生的采样偏差不可避免,对于有偏差或较少位点信息的数据集更容易出现过度拟合的情况。因此 Anderson & Gonzalez 以委内瑞拉的一种特有鼯鼠(*Cryptotis meridensis*)为研究对象,对偏差数据建模产生的模型性能差异进行探究。结果显示,当采用适当的正则化并确定最佳模型复杂性时,有偏差和少数的局部数据集构成的模型,实现了高预测能力。与使用默认设置建模相比,模型参数设置调整为与特定物种的实际状况相一致时,将会优化模型的运行结果。同年,Elith et al.^[12]以一种热带树种(*Banksia prionotes*)和双棘鳉鱼(*Gadopsis bispinosus*)为研究对象,基于两物种的习性及其生境特征,运用 MaxEnt 模型预测气候变化下两物种的潜在分布区,从统计的角度描述了 MaxEnt 模型,表明该模型最小化了协变量空间中定义的两个概率密度之间的相对熵,并提出了 MaxEnt 模型优化建议:提供合适背景下的样本、合理处理样本偏差以及通过特征类型选择和正则化设置进行模型调整。刘勇涛等人^[27]参考 Anderson & Gonzalez^[26]的研究,通过设置 MaxEnt 模型训练过程中使用的特征和 β 乘数两个参数来调节模型复杂度,依据 AUC 值/小样本下校正的赤池信息标准(Akaike information criterion corrected for small sample size, AICc)2 个

指标选择 MaxEnt 模型的最佳设置, 预测中国入侵物种南美蟛蜞菊 (*Wedelia trilobata*) 在中国当前及未来气候变化情形下的适生区。该设置优化过程解决了 MaxEnt 模型预测入侵植物适生区的模型转移性问题, 更在保证预测结果准确性的前提下简化了模型复杂程度, 大大提高了其可利用性。张天蛟和刘刚^[28]在探究提高生态位模型时间转移能力的方法时, 通过 Spearman's rho 相关性系数、方差膨胀因子筛选环境变量, 和调节“Regularization Multiplier”参数 3 种方法探讨模型时间转移能力的影响, 利用美国阿肯色河流域鲤鱼 (*Notropis girardi*) 2 个时期的分布数据与环境数据, 验证了通过降低模型维度和复杂程度提高模型时间转移能力的有效性, 并提出在保证模型自适应的前提下, 通过变量相关性分析与模型参数调节相结合来提高模型的转移能力, 将为模型的精度验证和时间转移能力的提高提供重要参考。Syfert 等^[29]则利用 MaxEnt 模型与新西兰树蕨的总体数据评估了校正地理采样偏差与复杂的特征类型对拟合优度的影响。结果表明, 对地理采样偏差的校正使得拟合优度显著提高, 当采样偏差得到合理校正时, 特征类型的选择对预测性能的影响可以忽略不计, 即简单的特征类型即可支持 MaxEnt 模型预测结果的有效性。该研究也支持了 Phillips 等^[25]提出的“校正采样偏差的方法极大地提高了物种分布模型的预测性能”这一理论, 强调了物种分布模型的预测能力可能会因缺乏该物种出现的环境范围内的代表性发生数据而受到严重限制, 也指出了需谨慎对待处于边缘或外推超出与训练数据相关的环境范围的位点数据。Kramer-Schadt 等^[30]在 Phillips 等^[25]和 Syfert 等^[29]研究的基础上, 采用空间剔除和使用偏差栅格文件调整背景位点两个策略处理采样偏差的问题, 并比较了两个策略的效果。结果表明, 针对某些区域或环境特征有强烈采样偏差的情况, 应降低空间聚集数据的记录。在数据充足的情况下, 空间剔除及使用偏差栅格文件调整背景位点两个策略都是解决采样偏差的优选; 若样本大小不足以使用空间剔除法时, 使用偏差栅格文件调整背景位点比不校正采样偏差更可取, 但需注意错分误差 (commission errors) 和漏分误差 (omission errors) 的发生。

孔维尧等^[31]2019 年汇总了国内外 MaxEnt 模型在取样偏差修正、模型复杂性调整、物种

分布判定阈值选择以及模型检验过程中的若干优化方法。研究指出, 在取样偏差的修正中, 空间剔除法的修正效果最好, 而背景限制法表现不佳。模型复杂性则受建模变量的数量、函数模式和调控系数三因素的影响, 在样本量小于建模变量数量时需依据其生态学意义进行变量筛选, 而函数模式对模型表现的影响不大, 因此一般选择简单模型, 同时注意调整调控系数以控制过度拟合。在判定物种出现阈值时, 应遵从客观性、等效性和判别力 3 个原则, 敏感度和特异性加和最大 (Maximum Sensitivity Plus Specificity, MSS) 是良好的阈值判定标准。模型检验方面又可分为不依赖阈值的检验 (Threshold Independent Model Evaluation) 和依赖阈值的检验 (Threshold Dependent Model Evaluation), 在不依赖阈值的模型评估方法中, 基于 AICc 和贝叶斯信息准则 (Bayesian Information Criteria, BIC) 选择的最优模型的表现优于基于 AUC 值或相关系数 (Correlation Coefficient, COR) 选择的模型; 在基于阈值的模型评估方法中, 真实技能统计 (True Skill Statistics) 能够兼顾模型漏分误差和错分误差, 不受假设缺失的影响, 且受物种流行度的影响较小。

以上即为 MaxEnt 模型不断优化发展的过程, 在这期间, MaxEnt 模型的应用不再仅局限于单一物种的潜在分布预测, 而是扩大到了多物种综合分布区预测, 探究种间相互作用及生物与环境的相互作用。如 Volkov 等^[32]运用 MaxEnt 模型与 RF 两种方法独立推断巴拿马巴罗科罗拉多岛 50 hm² 土地中丰富度最高的前 20 个热带树木群落的种间有效相互作用强度。研究表明, 种间相互作用较种内相互作用强度弱, 很好地解释了无相互作用的中性理论 (基于同一营养水平内的所有物种在人口统计学上处于相同地位), 但忽略了实际个体间的相互作用。因此, Volkov 等开发了一种将种间相互作用系统地纳入理论发展的方法, 即应用拉格朗日乘子法 (Lagrange Multipliers) 做约束条件, 强化“熵必须是包含尽可能多的可用信息的熵”的概念。2010 年, Williams 等^[33]基于最大熵理论的食物网度分布模型, 解释了度分布背后的机理, 提出了关于食物网中消费者、资源及无向网络的度分布的简单零模型建设的可能性。同时, MaxEnt 模型在人为干扰下的珍稀物种恢复保护工作中发挥了重要作用, 如颜明艳等^[35]通过构建北部湾中国海域中国鲎 (*Tachypleus*

tridentatus) 和圆尾鲨 (*Carcinoscorpius rotundicauda*) 大尺度潮间带环境空间分布和大跨度范围内的栖息地适应性模型, 提出在两种稚鲨适生区建立保护区, 并进行人工放流的建议, 以加强对海草和红树林的生态建设, 进而促进鲨资源种群恢复和发展。

2.2 MaxEnt 模型不同气候背景下的应用

在 MaxEnt 模型建立推广的 16 年间, 该模型的应用已遍及多个领域。近年来, 在特有物种及国家重点保护物种种群恢复^[36-38]、良种引种^[39]、资源利用与生产^[40-41]、病原体传播^[42-43]、虫害防治^[44]、保护区管理与建设^[45-46]等方面收获颇丰的研究成果。

依据其应用的不同气候背景, 将 MaxEnt 模型应用分为两类: 一是基于现今气候变化背景, 利用物种历史分布点对物种现生及未来潜在分布区进行推断预测; 二是基于 WorldClim 历史气候数据, 利用物种历史分布点对物种历史、现生潜在分布区进行推断预测。两者的不同之处在于所参照的气候背景不同, 而产生了对应不同时期的潜在分布区预测结果。

前者的应用偏重于对历史、现代、未来气候下的物种分布进行预测, 为深入理解物种进化历程及未来迁移轨迹提供重要的参考依据^[47-48]。联合国政府间气候变化专门委员会 (Intergovernmental Panel On Climate Change, IPCC) 第 6 次评估周期报告^[49]表明: 近几个世纪以来, 人为活动引发温室气体增加致使的全球变暖已经扭转了缓慢、长期的降温过程, 大气层、海洋生态、冰冻圈、生物圈等因此发生了广泛而迅速的变化, 各种极端天气及自然灾害出现的频率与强度也随之升高, 而全球变暖的加剧也对海平面上升、极圈冰盖融化等产生长期且不可逆转的影响。因此针对生态系统已受影响与脆弱性, 物种历史、现今、未来气候下的潜在分布预测将会为生物圈未来变化趋势提供科学依据。张天蛟^[50]基于 MaxEnt 模型, 对以阿肯色河流域鲤鱼 (*Notropis girardi*) 为代表的产漂流性卵的小型鱼类进行生态位建模及分析研究, 针对不同时期阿肯色河流域鲤鱼的分布情况建立模型, 研究自然环境因子和人为活动因子对阿肯色河流域鲤鱼潜在分布的影响, 并运用 IPCC 第 5 次评估周期报告发布的气候情景数据建立气候因子模型, 预测阿肯色河流域中点的物种分布

的变化趋势, 针对此提出相应保护政策; 张立娟等^[51]利用 Maxent 模型预测气候变化背景下青冈 (*Cyclobalanopsis glauca*) 空间分布格局的变化, 并用分布区质心移动分析适生区变化的方向和幅度, 探讨青冈分布格局的变化对中国亚热带北界移动的指示意义。在 IPCC 第 6 次评估周期报告推出后, 未来共享社会经济路径 (Shared Socioeconomic Pathway, SSP) 进入到研究者的视野中, 研究者以新形式的气候模型为背景开展了广泛的物种潜在分布区预测。例如, 刘超等^[48]的研究运用 MaxEnt 模型及 ArcGIS 将不同时期适宜生境分布范围缩小为单个中心点, 计算不同时期分布中心的位置与变化, 分析川梨 (*Pyrus pashia*) 的迁移趋势和历史迁移轨迹; Guo 等^[52]选择来自中国南方 (800 mm 等降雨线以南地区) 的 56 种常见栽培树种, 结合了 10 种物种分布模型, 分析在未来不同气候变化情景下潜在种植面积的变化。结果表明, 气候变量的变化, 特别是日平均气温变化幅度大于 $\pm 0.25^{\circ}\text{C}$ 时, 将导致中国南方大部分树种分布面积减少。随着温度的持续上升, 栖息地丧失的物种种数及其分布区域丧失的速度将显著增加, 也将导致树木间隙增加、环境退化和外来物种入侵的风险。在全球变暖的背景下, 中国南方的山区可能会演变为避难所。

MaxEnt 模型应用的后者, 则依据 WorldClim 历史气候数据, 对物种历史、现生潜在分布区进行推断预测, 经比较不同历史时期下的物种潜在分布区, 以得出物种历史迁移轨迹及未来分布变化趋势。例如, Alabar 等^[53]运用 MaxEnt 模型评估阿根廷西北部一森林 (Piedmont Forest) 8 种木材树种幼苗和成体的潜在分布, 及幼苗与成体分布区重叠的百分比。该研究结果表明, 幼苗的潜在分布存在纬度分布的变化, 有向南移动的趋势, 同时也印证了全球气候变暖的趋势。Aidoo 等^[54]在两种不同气候变化情景下, 建立 MaxEnt 模型预测当前面临风险的区域以及未来可能被亚洲柑橘木虱 (*Diaphorina citri*) 入侵的区域, 研究结果表明, 直至 2050 年, 中国、巴西和美国等主要的柑橘出口国家, 将持续受到这种害虫的威胁。

2.3 MaxEnt 模型在区域方面的应用

MaxEnt 模型的应用依托于被预测物种的已知分布点, 当该物种或物种集合 (属、科、目、纲阶元的种群集合) 分布广泛且已知分布数据量丰富时, 模型预测的范围也可相应扩大。如, 在

国家、世界等大尺度范围上建立预测模型。刘志远等^[36]利用 MaxEnt 模型对蓝马鸡 (*Crossoptilon auritum*) 栖息地适宜性进行研究, 了解蓝马鸡在中国的分布格局和限制其分布的主要环境因子, 从而为该物种有效保护和管理提供科学依据; Mu 等^[55]使用 MaxEnt 模型在当前及未来气候背景下, 预测了温室蟾 (*Eleutherodactylus planirostris*) 在全球和中国的潜在分布区, 结果表明, 该物种在中国的潜在分布区比目前所反映的区域要大得多, 且很可能从珠江三角洲扩散到周边地区, 在气候持续变暖的情况下, 其入侵范围将由南向北扩大, 为评估该物种入侵的风险提供了科学依据; Liu & Mai^[56]假设麻风树 (*Jatropha curcas*) 对气候变化的响应会呈现出因纬度而异的情形, 利用 MaxEnt 模型来说明 2021—2060 年在 SSP 各模式下气候变化对亚太地区麻风树栖息地的影响, 及其适宜栖息地边界的未来移动趋势, 为东南亚地区、太平洋岛屿应对不同气候变化对生物能源的影响作出建议; 杨启杰^[34]利用 MaxEnt、ArcGIS 等工具, 研究全球尺度上当前气候下桫欏 (*Alisophila spinulosa*) 的潜在适生区, 分析影响其分布的主要生态因子, 并模拟不同时期不同气候路径下桫欏的分布及适生区变化; 马世炎等人^[57]借助 MaxEnt、ArcGIS 等工具, 预测大豆蚜 (*Aphis glycines*) 的全球潜在地理分布区, 推测环境变量对大豆蚜分布的影响, 并指出可着重研究大豆蚜集中分布区的小范围生境, 有助于揭示大豆蚜的生态特征。

而以科学考察、定期监测等为目的的研究, 其调查范围往往是国家划分的行政区、各保护区与保护地、各水域或河流河段、生物多样性热点地区等特定区域。不仅如此, 部分目标研究物种因实地调查条件有限, 或濒危物种难以监测, 而获得的数据点有限, 通常将会利用全球生物多样性信息服务网络平台 (Global Biodiversity Information Facility, GBIF)、中国国家标本资源共享平台 (National Specimen Information Infrastructure, NSII)、公开发表的文献、书籍等渠道补充数据点。正如前文所言, 这类收据点的获得是基于不同人员、不同时期、不同地点的采集记录, 因而存在有一定的集中性及目的性。例如, 胡政香等^[42]利用 Figshare 在线数据库 (<https://figshare.com/>) 收集草原革蜱 (*Dermacentor nuttalli*) 的分布位点, 对草原革蜱在中国新疆维吾尔自治

区的分布进行了预测, 为进一步推进新疆地区蜱类种群分布研究、蜱传病原体检测以及蜱源性疾病的风险评估分析及防控提供研究数据; 高浩翔等^[45]综合考虑调查所涉及区域的面积及物种活动特性等因素, 制定了合理方案对甘肃兴隆山保护区野生马麝 (*Moschus chrysogaster*) 实地调查, 利用 MaxEnt 模型定量分析该保护区野生马麝的夏季潜在分布与各个环境因子之间关系, 识别并预测整个保护区野生马麝夏季适宜生境的空间分布格局, 为濒危野生马麝保护及自然保护区管理提供科学依据; 文世荣等人^[46]则采用红外相机技术对西双版纳保护区曼稿片区及其周边区域的印度野牛 (*Bos gaurus*) 的活动进行了记录, 通过建立 MaxEnt 模型对生境适宜性进行分析, 并提出了保护区管理和廊道建设等方面的建议; 王波等人^[58]通过收集各标本馆收录标本数据及各文献著作记录数据, 建立 MaxEnt 模型, 对中国西南喀斯特地貌区两栖物种丰富度分布格局与环境因子间的关系进行了探究; 赵梓伊等人^[59]推测“横断山区可能是齿突蟾属的起源中心和分化中心”, 利用 GBIF、公开发表的文献著作记录及横断山南段生物多样性保护优先区域 (横断山南) 的实地调查数据, 对包括西藏齿突蟾 (*Scutiger boulengeri*)、刺胸齿突蟾 (*Scutiger mammatus*)、胸腺齿突蟾 (*Scutiger glandulatus*)、贡山齿突蟾 (*Scutiger gongshanensis*)、圆疣齿突蟾 (*Scutiger tuberculatus*) 在内的 5 种齿突蟾属物种进行了潜在地理分布预测, 分析了影响五种齿突蟾属物种栖息地适宜性的主要环境因子, 为横断山南齿突蟾属物种种群动态监测、资源管理以及保护工作提供了科学参考。

3 调查空白区域及样本量过少的方法策略

MaxEnt 模型在区域范围方面的应用, 不论是全球范围内的分布预测还是特定区域内的分布预测, 其研究范围皆受限于物种已知分布点及研究目的。但利用物种已知分布数据进行模型建立, 将面临一个问题, 某些濒危物种野外调查或公开记录得到的数据量过少, 将会致使物种预测结果精确度较低。Hernandez 等^[60]的研究表明, Maxent 模型在样本量保持于 25 及以上水平时呈现出较为稳定的预测分布区域。因此, 为确保 Maxent 模型特定研究区域内的预测精确度, 本文

采用如下方法策略：将原本的研究区域扩大至一定范围，使用该范围内的物种已知分布点建立模型^[16]，即扩大预测范围及预测数据样本量，以预测物种潜在分布区。

这种方法利用大范围区域预测物种潜在分布，而后取特定区域为研究范围进行分析。此方法对物种数据量过少及公开记录与采集空缺的地区的预测较为合适，且该方法的应用不会对特定研究区域内物种历史分布数据造成影响，也将会充分利用特定研究区域边缘数据，不会使得数据因边缘化被舍弃而以致预测的物种潜在分布区过小的情况。同时，此种策略也会将更多环境因子纳入 MaxEnt 模型使用变量中，通过 MaxEnt 模型自身配备的刀切法 (Jackknife) 及环境变量因子响应曲线 (Response curves)、其他类似于 ENMTools、SDMToolbox 等工具筛选环境变量，避免环境变量过度拟合对模型精度造成的影响，以此提高模型预测的精确率。

而对于因收集的数据信息来源不同、记录及解译方式不同造成的采样偏差无法避免，为此，采用此方法将更多的数据纳入建模的环境变量范围内，可降低研究区域边缘数据集中而导致的预测结果产生偏差的情况，进而再通过缓冲区筛选等方法，对存有偏差的集中数据位点进行处理，可进一步提高预测结果的精确率。

除此之外，大范围区域预测的物种潜在分布区也为现有保护区及保护地规划、管理与建设优化工作提供了科学依据，不同时期下物种分布面积的变化也预示着其未来分布区的扩张或缩减趋势。

4 讨 论

物种分布模型立足于物种与环境间的关系，依据分布点所在的环境信息评估物种的生存需求，建立物种适宜生境分布的数学模型。随着计算机科学、统计学习方法不断发展，物种分布模型相继涌现，与不同时代背景下科学问题的演变并行不悖，而逐渐渗透于生命科学、环境科学等诸多领域，用于解决生态学、生物地理学、保护生物学、外来入侵物种、全球气候变化等方面的实际问题。

在物种分布模型中，MaxEnt 模型是为诸多研究者广泛使用且性能较良好的模型之一。相较

于其他模型，MaxEnt 模型的优势在于，其依靠物种已知分布数据与研究区域的环境信息即可建立模型。并且依据最大熵理论，建立该模型使用已知信息做约束条件，取最大熵时，约束条件下随机量概率分布最符合客观情况，即预测分布状态最接近实际分布状态，将会得到较为准确的预测结果。

在 MaxEnt 模型应用推广的历程中，其主要经历了三个阶段。一是推广初期通过多模型类比展现 MaxEnt 模型良好性能与存在问题；二是样本偏差问题解决及参数优化阶段，研究者不断纳入新方法解决样本数据存在偏差、环境变量过度拟合等问题；三是在 MaxEnt 模型优化精进的同时，物种潜在分布区的预测也从单一物种逐渐演变为多个物种集合，单一模型的使用推广为多重模型组合运用。

在探索 MaxEnt 模型不同气候背景下的应用时，本文将对应的应用分为两类：一是基于气候变化背景，对物种现生及未来潜在分布区进行推断预测；二是基于 WorldClim 历史气候数据，对物种历史、现生潜在分布区进行推断预测。两者皆对应了不同气候背景下的物种潜在分布区域的变化趋势。

在 MaxEnt 模型区域应用方面，通过综合分析，物种研究区域受限于物种已知分布点及研究目的，当研究区域内物种已知样本量较少而不足以支撑 MaxEnt 模型达到理想结果时，即可采用扩大预测范围及预测数据样本量以预测其潜在分布区的方法策略，而后取特定区域为目标研究范围进行分析。此方法充分利用了目标研究区域内边缘化的数据，并通过筛选环境变量以避免空间自拟合等方法尽力提高模型预测精确度。

综上所述，MaxEnt 模型仍有着较大的提升空间。在普遍的推广使用中，各位研究者就不同研究对象及不同尺度范围的研究对 MaxEnt 模型参数设置等不断优化。对于研究区域范围内样本量不足及空缺的状况，值得多加关注并思考解决方案，使得 MaxEnt 模型能应用于更广泛的领域，服务于更多学科。

参考文献：

- [1] 王娟, 倪健. 植物种分布的模拟研究进展 [J]. 植物生态学报, 2006(6):1040-1053.

- [2] Anderson R. P. A framework for using niche models to estimate impacts of climate change on species distributions [J] . *Annals of the New York Academy of Sciences*, 2013, 1297(1) : 8–28.
- [3] 郭彦龙, 赵泽芳, 乔慧捷, 等. 物种分布模型面临的挑战与发展趋势 [J] . *地球科学进展*, 2020, 35(12):1292–1305.
- [4] 欧阳泽怡, 欧阳硕龙, 吴际友, 等. 最大熵模型在植物适生区预测应用中的研究进展 [J] . *湖南林业科技*, 2022, 49(1):83–88.
- [5] Thuiller, W. , Lafourcade, B. , Engler, R. , et al. BIOMOD – A platform for ensemble forecasting of species distributions [J] . *Ecography*, 2009, 32: 369–373.
- [6] Clarivate. ISI Web of Science. [DB/OL] . (2022–01) [2022–06–17] . <https://www.webofscience.com/wos/alldb/basic-search>
- [7] 朱耿平, 刘国卿, 卜文俊, 等. 生态位模型的基本原理及其在生物多样性保护中的应用 [J] . *生物多样性*, 2013, 21(1):90–98.
- [8] Busby, J. BIOCLIM – A bioclimate analysis and prediction system [J] . *Plant protection quarterly*, 1991, 6: 8–9.
- [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] Elith, J. , Graham C. H. , Anderson R. P. , et al. Novel methods improve prediction of species' distributions from occurrence data [J] . *Ecography*, 2010, 29(2):129–151.
- [11] Phillips S. J. , Anderson R. P. , Schapire R. E. Maximum entropy modeling of species geographic distributions [J] . *Ecological Modelling*, 2006, 190(3): 231–259.
- [12] Elith, J. , Phillips, S. J. , Hastie, T. J. , et al. A statistical explanation of MaxEnt for ecologists [J] . *Diversity and Distributions*, 2010:1–15.
- [13] 张宁. 熵概念研究 [J] . *北京联合大学学报 (自然科学版)*, 2007(1):1–3.
- [14] Jaynes E. T. Information theory and statistical mechanics [J] . *Physical Review*, 1957, 106(4) :1–620.
- [15] Merow, C. , Smith, M. J. , Silander, J. A. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter [J] . *Ecography*, 2013, 36, 1058–1069.
- [16] 白君君, 侯鹏, 赵燕红, 等. 物种生境适宜性模型及验证的研究进展 [J/OL] . *生态学杂志*, 2022,41(7).
- [17] Peterson, A. T. , Papes, M. , Eaton, M. D. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent [J] . *Ecography*, 2007, 30, 550–560.
- [18] 王运生, 谢丙炎, 万方浩, 等. ROC 曲线分析在评价入侵物种分布模型中的应用 [J] . *生物多样性*, 2007(4):365–372.
- [19] 李明阳, 巨云为, Sunil Kumar, 等. 美国大陆外来入侵物种斑马纹贻贝 (*Dreissena polymorpha*) 潜在生境预测模型 [J] . *生态学报*, 2008(9):4253–4258.
- [20] 李白尼, 马骏, 侯柏华, 等. 基于生态位模型的番荔枝实蝇潜在适生性分布预测 [J] . *环境昆虫学报*, 2009,31(4):291–299.
- [21] 曹向锋, 钱国良, 胡白石, 等. 采用生态位模型预测黄顶菊在中国的潜在适生区 [J] . *应用生态学报*, 2010,21(12):3063–3069.
- [22] 郭乃嘉. 基于生态位模型的中间香型烤烟生态适宜区潜在分布预测 [D] . 重庆: 西南大学, 2013.
- [23] 孔晓泉. 物种分布模型的稳定性评估及应用软件 [D] . 安庆: 安庆师范学院, 2015.
- [24] Phillips, S. J. , Dud í k M. Modeling of species distributions with maxent: new extensions and a comprehensive evaluation [J] . *Ecography*, 2008, 31(2):161–175.
- [25] Phillips S. J. , Miroslav Dud í k, Elith J. , et al. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data [J] . *Ecological Applications*, 2009, 19(1):181–197.
- [26] Anderson, R. P. , Gonzalez, I. Species-specific tuning increases robustness to sampling bias in models of species distributions: An implementation with Maxent [J] . *Ecological Modelling*, 2011, 222(15): 2796–2811.
- [27] 刘勇涛, 戴志聪, 薛永来, 等. 外来入侵植物南美螞蟥菊在中国的适生区预测 [J] . *广东农业科学*, 2013,40(14):174–178.
- [28] 张天蛟, 刘刚. 提高生态位模型时间转移能力的方法研究 [J] . *中国农业大学学报*, 2017,22(2):98–105.
- [29] Syfert MM., Smith MJ., Coomes DA. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models [J] . *Plos One*. 2013, 8(2):e55158.
- [30] Kramer - Schadt, S. , Niedballa, J. , Pilgrim, J. D. , et al. The importance of correcting for sampling bias in MaxEnt species distribution models [J] . *Diversity and Distributions*, 2013, 19:1366–1379.
- [31] 孔维尧, 李欣海, 邹红菲. 最大熵模型在物种分布预测中的优化 [J] . *应用生态学报*, 2019,30(6):2116–2128.
- [32] Volkov, I. , Banavar, J. R. , Hubbell, S. P. , et al. Inferring species interactions in tropical forests [J] . *Proceedings of the National Academy of Sciences*, 2009, 106: 13854–13859.
- [33] Williams, R. J. Simple MaxEnt models explain food

- web degree distributions [J]. *Theoretical Ecology*, 2010, 3(1): 45–52.
- [34] 杨启杰. 基于 MaxEnt 模型的孑遗植物桫欏在不同时期的潜在适生区研究 [D]. 杭州: 浙江大学, 2021.
- [35] 颜明艳, 李琼珍, 宋洁, 等. 基于 MAXENT 模型评估北部湾潮间带中国鲎和圆尾鲎稚鲎的潜在地理分布及种群保育对策 [J]. *生态学报*, 2019, 39(9): 3100–3109.
- [36] 油志远, 王茗, 鲁碧耕, 等. 基于 MaxEnt 模型预测蓝马鸡的潜在分布区 [J]. *生态学杂志*, 2022, 41(11): 2271–2271.
- [37] 鲁碧耕, 骆念龙, 谢凡, 等. 黑颈鹤在四川省白玉县的繁殖栖息地质量评估 [J]. *西南民族大学学报 (自然科学版)*, 2022, 48(3): 237–244.
- [38] 李行, 黄允优, 阮涛, 等. 基于 MaxEnt 模型的邛崃山系中华小熊猫栖息地适宜性评价 [J]. *贵州师范大学学报 (自然科学版)*, 2022, 40(2): 34–39.
- [39] 胡超, 于静. 基于 MaxEnt 和 GIS 预测四川省杉木良种在湖北省同一适宜引种生态区 [J]. *四川林业科技*, 2022, 43(3): 85–93.
- [40] 林正雨, 陈强, 陈春燕, 等. 基于 MaxEnt 模型评估赤霞珠在川西横断山河谷的适生空间分布 [J]. *中国农业资源与区划*, 2020, 41(12): 144–155.
- [41] 叶雨, 樊丛照, 张际昭, 等. 基于 MaxEnt 和 GIS 的新疆紫草产地生态适宜性分析 [J]. *中国现代中药*, 2022, 24(5): 770–775.
- [42] 胡政香, 诺明达来, 余权威, 等. 新疆地区草原革蜚 MaxEnt 分布预测分析 [J]. *生态科学*, 2022, 41(4): 92–101.
- [43] 杨艺涛, 杜玉壮, 李乙冉, 等. 基于 MaxEnt 模型对我国长角涵螺潜在分布区的预测研究 [J]. *中国媒介生物学及控制杂志*, 2022, 33(3): 409–413.
- [44] 武万里, 刘垚, 马菁, 等. 基于 Maxent 模型的宁夏枸杞瘿螨适生性分析 [J]. *山地农业生物学报*, 2022, 41(2): 66–70.
- [45] 高浩翔, 申立泉, 刘瑞, 等. 基于 MAXENT 模型的野生马麝夏季生境适宜性研究 [JOL]. *生态学报*, 2022.
- [46] 文世荣, 方国, 张明霞, 等. 基于 MaxEnt 模型的西双版纳保护区曼稿片区印度野生牛生境适应性分析保护建议 [J]. *山东林业科技*, 2022, 52(02): 27–31+38.
- [47] Warren, R. F., VanDerWal, J., Price, J. T., et al. Quantifying the benefit of early climate change mitigation in avoiding biodiversity loss [J]. *Nature Climate Change*, 2013, 3(7): 678–682.
- [48] 刘超, 董星光, 田路明, 等. 基于 Maxent 模型的川梨不同时期潜在分布研究 [J]. *果树学报*, 2022, 39(11): 2056–2063.
- [49] IPCC, 2021: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC* [M]. Cambridge University Press, United Kingdom 2021.
- [50] 张天蛟. 产漂流性卵小型鱼类的生态位建模及分析 [D]. 北京: 中国农业大学, 2016.
- [51] 张立娟, 李艳红, 任涵, 等. 气候变化背景下青冈分布变化及其对中国亚热带北界的指示意义 [J]. *地理研究*, 2020, 39(4): 990–1001.
- [52] Guo Y., Zhao Z., Zhu F. X., et al. Climate change may cause distribution area loss for tree species in southern China [J]. *Forest Ecology and Management*, 2022, 511: 120134.
- [53] Alabar, F., Politi, N., N ú zaro, P., et al. Changes in the potential distribution of valuable tree species based on their regeneration in the Neotropical seasonal dry forest of north–western Argentina [J]. *Environmental Conservation*, 2022, 49(2): 83–89.
- [54] Aidoo O. F., Souza P. G. C., Silva R. S. da, et al. Climate - induced range shifts of invasive species (*Diaphorina citri* Kuwayama) [J]. *Pest management science*, 2002, 78: 2534–2549.
- [55] Mu, C., Guo, X., Chen, Y. Impact of Global Climate Change on the Distribution Range and Niche Dynamics of *Eleutherodactylus planirostris* in China [J]. *Biology*, 2022, 11(4): 588.
- [56] Liu G., Mai J.. Habitat shifts of *Jatropha curcas* L. in the Asia–Pacific region under climate change scenarios [J]. *Energy*, 2022, 251, 123885.
- [57] 马世炎, 于洪春, 赵奎军, 等. 基于 MaxEnt 模型的大豆蚜全球潜在地理分布分析 [J]. *昆虫学报*, 2022(5): 2–9.
- [58] 王波, 黄勇, 李家堂, 等. 西南喀斯特地貌区两栖动物丰富度分布格局与环境因子的关系 [J]. *生物多样性*, 2018, 26(9): 941–950.
- [59] 赵梓伊, 肖能文, 刘高慧, 等. 五种齿突蟾在横断山南潜在地理分布预测 [J]. *生态学报*, 2022, 42(7): 2636–2647.
- [60] Hernandez, P. A., Graham, C. H., Master, L. L., et al. The effect of sample size and species characteristics on performance of different species distribution modeling methods [J]. *Ecography*, 2006, 29, 773–785.
- [61] El–Gabbas A. and Dormann C. F. Improved species–occurrence predictions in data–poor regions: using large–scale data and bias correction with down–weighted poisson regression and maxent [J]. *Ecography*, 2018, 41(7): 1161–1172.