植 物 分 类 与 资 源 学 报 2013, **35** (5): 647 ~ 655
Plant Diversity and Resources DOI: 10.7677/ynzwyj201312127

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*

毕迎凤[1,2], 许建初[3], 李巧宏[1], Antoine Guisan[4], Wilfried Thuiller[5], Niklaus E. Zimmermann[4], 杨永平[1], 杨雪飞[1**]

(1 中国科学院昆明植物研究所资源植物与生物技术所级重点实验室, 云南 昆明 650201; 2 中国科学院大学, 北京 100049; 3 中国科学院昆明植物研究所山地生态系统研究中心, 云南 昆明 650201; 4 瑞士洛桑大学生物医学学院, 瑞士; 5 法国格勒诺布尔第一大学高山生态实验室, 法国)

摘要: 新型统计方法和多源、多尺度空间信息数据的产生促进了物种空间分布模型的快速发展。不同的物种空间分布模型在生态学理论的运用以及前提假设上存在差异。选用不同的模型方法和输入数据会带来预测结果的不确定性。对比并集成多个物种空间分布模型, 同时利用多组输入数据可降低预测的不确定性, 提高物种分布模拟的精度。本文以中国特有种铁杉 (*Tsuga chinensis*) 为例, 运用基于 R 语言开发的 BioMod 软件包对比 9 个物种空间分布模型对铁杉的模拟效果。最后以曲线下面积 (ROC) 为权重集成 9 个模型的模拟结果, 产生和筛选最佳的铁杉潜在空间分布图。研究发现随机森林模型 (RF) 的模拟效果最好, 其次是多元适应回归样条函数模型 (MARS) 和广义相加模型 (GAM), 模拟效果最差的是表面分布区分室模型 (SRE)。模型集成结果显示, 最适宜铁杉分布的区域集中在中国的西南及四川盆地周围, 其次零星分散于华南和台湾部分地区。这一结果与前人对铁杉自然分布的描述和研究结果较为吻合。研究进一步表明, 通过模型的集成能有效地降低由于单个模型所带来的模拟结果不确定性, 从而提高模拟的精度和效果。

关键词: 物种空间分布模型; 铁杉; 模型集成; 分布区; BioMod
中图分类号: Q 948.2 **文献标识码**: A **文章编号**: 2095-0845(2013)05-647-09

Applying BioMod for Model-Ensemble in Species Distributions: a Case Study for *Tsuga chinensis* in China

BI Ying-Feng[1,2], XU Jian-Chu[3], LI Qiao-Hong[1], Antoine Guisan[4], Wilfried Thuiller[5], Niklaus E. Zimmermann[4], YANG Yong-Ping[1], YANG Xue-Fei[1**]

(1 *Key Laboratory of Economic Plants and Biotechnology, Kunming Institute of Botany, Chinese Academy of Sciences*, Kunming 650201, China; 2 *University of Chinese Academy of Sciences*, Beijing 100049, China; 3 *Centre for Mountain Ecosystem Studies, Kunming Institute of Botany, Chinese Academy of Sciences*, Kunming 650201, China; 4 *Département d'Ecologie et Evolution Faculté de Biologie et Médecine Université*, de Lausanne CH-1015 Lausanne, Switzerland; 5 *Laboratoire d'Ecologie Alpine, University Grenoble 1-J.* Fourier, France)

Abstract: The integration of new statistical techniques and increasing availability of multi-sources and multi-scale data sets promote the development of species distribution modeling. Yet, choice of data sets, different model types and their underlying ecological theories and assumptions can cause uncertainty in model predictions. In order to decrease prediction uncertainty, studies using model ensemble are gaining in popularity. In this paper we apply the BioMod package developed under R environment to predict the spatial distribution of *Tsuga chinensis* using nine differ-

* 基金项目: 中国科学院知识创新工程重要方向项目——西南野生生物资源的挖掘与利用 (KSCX2-EW-J-24)
** 通讯作者: Author for correspondence; E-mail: xuefei@mail.kib.ac.cn
收稿日期: 2012-10-19, 2013-02-26 接受发表
作者简介: 毕迎凤 (1987-) 女, 硕士研究生, 研究方向为生态学和植物的气候变化响应。E-mail: biyingfeng@mail.kib.ac.cn

ent models. Our aims were to evaluate model performance, select explanatory variables, and assemble the best predictive output. Random Forest, MARS and GAM performed the best amongst the nine models compared, while SRE was the worst. The ensemble models predicted that the areas of high probability for *T. chinensis* presence lie mainly in Southwest China and the periphery of the Sichuan basin, and are also distributed sporadically in South China and Taiwan. These predictions reflect the actual distribution pattern of *T. chinensis*, and show high agreement with other analyses. The application of BioMod for model ensemble lowers uncertainty and improves the prediction performance.

Key words: Species distribution model; *Tsuga chinensis*; Model assembly; Biogeography; BioMod

物种空间分布模型是利用物种已知分布点、环境信息及二者之间的相互关系来模拟和预测物种在地理空间中的分布状况（Austin, 2002; Guisan 和 Zimmermann, 2000; Pulliam, 2000），是利用物种分布样本点来推测物种分布区的一种有效途径和方法（Marmion 等, 2009）。物种的空间分布信息不但有助于我们了解物种的历史演化过程，还可用于研究物种对环境的生态适应性及变化（Elith 等, 2006; McPherson 和 Jetz, 2007），为土地利用规划、资源管理、物种保护和引种驯化提供重要的科学依据。

20 世纪 90 年代初，物种空间分布模型的研究须从地形图或其他纸质环境图层的数字化入手。记载物种分布的标本信息也须通过手工查阅、整理和输入。这些工作不仅费时耗力，而且可用于建模的环境信息仅为地形地貌及其衍生的基本指标。随着数字化信息的爆炸式增长、计算科学的发展以及统计方法的创新，物种空间分布模型研究得以飞速发展。

首先，大量多源多尺度地理信息和数字化资源可以免费获取和利用（Brotons 等, 2004; Elith 等, 2006），如世界和区域范围的数字化地形数据、具有时间序列的遥感植被指数、Worldclim 的气候数据以及数字化的标本数据；其次，高性能计算机的普及使复杂和海量数据运算成为可能；再次，各种新型和高级统计方法的开发和应用及其与最新生态学理论的有效整合（Austin, 2007），促进了模型方法的多样化。上述发展使我们突破过去利用单一方法或基于有限的环境因子进行物种空间分布预测的局限，促进多种方法的运用和对比以提高模拟效果，并为检验不同的假设、回答不同的科学问题创造条件。当前关于物种空间分布模型的研究热点主要集中于：研发新的模型（Phillips 等, 2006; Prasad 等, 2006）；比较多种模型的模拟效果（Brotons 等, 2004; Elith 等, 2006; Leathwick 等, 2006）；模型集成以提高模拟效果（Araújo 和 New, 2007）；研究不同分类群在演化过程中的环境和地理空间分化过程（Heibl 和 Renner, 2012; Smith 和 Donoghue, 2010）；研究生物多样性空间分布格局（Canhos 等, 2004）；模拟未来气候变化情景下物种的分布区变化（Bellard 等, 2012; Maiorano 等, 2012; Thuiller, 2004）；比较入侵物种在原生地和入侵地间的生态位差异（Gallien 等, 2010; Petitpierre 等, 2012; Václavík 和 Meentemeyer, 2012）。

尽管物种空间分布模拟的方法和技术取得快速发展，但如何从众多的模型中选择最佳者仍是个难题（Austin, 2007; Austin 和 Van Niel, 2011）。同时，不同的模型在构建过程中所基于的生态学理论和前提假设不尽相同（Guisan 和 Thuiller, 2005），模拟过程和算法有所差异，造成模拟结果的不确定性（Barry 和 Elith, 2006; Buisson 等, 2010; Naimi 等, 2011; Thuiller, 2004; van Horssen 等, 2002; Wiens 等, 2009）。另外，模型的初始设置和参数设置也会带来一定的预测差异。模型评估中，或因评估数据的独立性和代表性程度不同，产生评判结果的不确定性。因此目前有很多研究正围绕降低模型的不确定性而开展，BioMod 就是为解决这个问题而开发的研究平台。BioMod 由法国格勒诺布尔第一大学高山生态实验室、瑞士洛桑大学生物医学学院、西班牙马德里国家自然博物馆生物多样性与进化生物学部以及葡萄牙艾武拉大学共同研发。它是基于 R 开发的免费和公开的软件包，可在 http://r-forge.r-project.org/projects/BioMod/进行免费下载。其优点是能够处理由不同模型方法和非独立评估样本所带来的不确定性。BioMod 采用了 9 种可选的物种分布模型，并通过集合解决模型间差异的问题。与单一模型相比，BioMod 可运用不同种类的模型并设置不同的初始条件、参数和

限制性条件，进行大量的运算，综合分析所有运算结果的共性、差异和不确定性。其结果涵盖各种条件和不同情况下预测的可能性，具综合性、总结性和可靠性的特点（Thuiller 等，2009）。

本文以中国特有种铁杉（*Tsuga chinensis*）为研究对象，应用 BioMod 对比和集成多个模型模拟其潜在空间分布，同时检验不同的气候因子组合对铁杉空间分布模拟效果的差异。

1　研究材料和方法

1.1　研究物种

铁杉为常绿乔木，喜酸性土壤，多生长于多雨多雾、湿度较大、气候凉润的山地环境。标本记载产地主要包括河南、陕西、甘肃、湖北、四川、贵州等地。常在海拔 2 000 ~ 3 000 m 之间与云南铁杉（*T. dumosa*）、麦吊云杉（*Picea brachytyla*）、油麦吊云杉（*P. brachytyla* var. *complanata*）、冷杉（*Abies fabri*）等组成针叶树混交林，少数成纯林（郑万钧和傅立国，1978）。

1.2　物种分布数据

从标本记录上提取铁杉分布的样本点，标本资料来源于中国数字植物标本馆（http://www.cvh.org.cn/cms/）。共获取标本记录 761 份，实体标本分别馆藏于中国科学院植物研究所（PE），江苏植物研究所（NAS），中国科学院西北高原生物研究所（HNWP），西北农林科技大学（WUK），广西植物研究所（IBK），庐山植物园（LBG），中国科学院华南植物研究所（IBSC）和中国科学院成都生物研究所（CDBI）。首先对 761 份标本数据进行初步筛选，去除空间信息及其他信息不明确的记录，以及同一采集地多份标本的重复记录，最终保留 237 条有效记录（图 1）。接着检查每条包含经纬度坐标信息的记录，如核对无误则直接采用，否则予以校正。对于没有经纬度的标本采集记录，通过其记载的省、县和小地名信息，在 Google Earth 上查询并重建其坐标信息。应该注意的是重建的坐标数据不完全准确代表真实的标本采集坐标信息，是其近似估计值。尽管如此，根据所使用栅格数据的空间分辨率大小，在一定范围内偏离真实值并不会降低模拟效果。利用标本信息重建的空间数据已经被大量的运用到物种的空间分布模型当中，并被证明有效可行（Loiselle 等，2008；Schmidt 等，2005）。在本研究的实际操作中，结合卫星影像，在小地名分布范围内根据标本记录中关于生境和海拔高度的描述，主观选择一个符合上述标准的空间点作为近似的标本采集点。由于所采用的气候栅格数据空间分辨率为 0.0083°，标本采集点坐标精确到小数点后 4 位数即可。本研究只有物种的实际分布点记录，没有物种的非分布点记录，故运用 BioMod 的功能生成拟非分布点（Pseudo absence），具体详见本文 1.5 节。

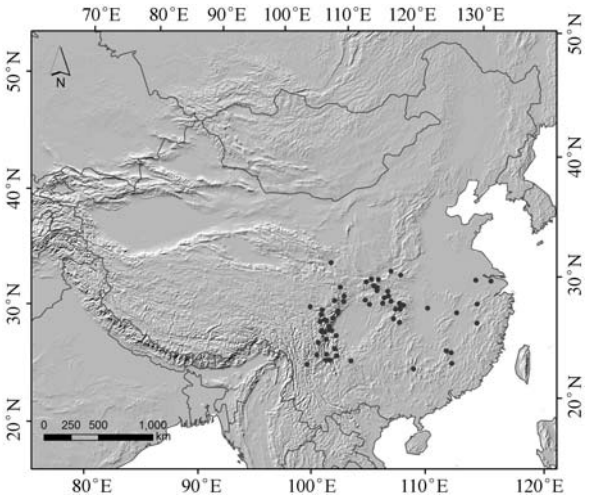

图 1　用于建模的铁杉标本点分布示意图

Fig. 2　Location map of herbarium collection used for species distribution models

1.3　气候变量

气候是决定物种分布的主要环境因素，且该数据易于获得。本文采用 Worldclim 全球气候数据库中空间分辨率为（0.0083°×0.0083°）的 19 个（Bio1-19）气候变量来研究铁杉的气候适宜分布区。包括年均温（Bio1）、平均月较差（Bio2）、等温性（Bio3）、气温的季节性（Bio4）、最暖月最高温（Bio5）、最冷月最低温（Bio6）、气温年较差（Bio7）、最湿季平均温（Bio8）、最干季平均温（Bio9）、最暖季平均温（Bio10）、最冷季平均温（Bio11）、年降水量（Bio12）、年最湿月降水量（Bio13）、年最干月降水量（Bio14）、降水的季节性（Bio15）、最湿季降水量（Bio16）、最干季降水量（Bio17）、最暖季降水量（Bio18）、最冷季降水量（Bio19）。该数据为 1950 ~ 2000 年的平均值，坐标投影系统为 UTM-WGS84，详情可参考 http://www.worldclim.org/current。由于未考虑其他可能的相关因子，如海拔、坡向、坡度、太阳辐射、植被指数等，本研究所预测的铁杉分布范围为其潜在分布区，而不代表其实际分布区。

除了模型的选择外，如何甄别好的解释性变量也是进行物种空间分布模拟研究的一个难点（Austin 和 Van Niel，2011）。如果将所有能获取的环境因子作为模型输入将使运算冗长且会降低模拟准确度（Williams 等，2012）。为达到最佳运算效果，在多数情况下，研究者会通过自己的经验来筛选输入模型的环境因子（Franklin，1998）。这种方法虽然简单而易于操作，但主观性较强。在事先不知道由哪些环境因子主导物种的空间分布格局的情况下，研究者可通过统计手段来寻求解决方案

(Williams 等, 2012), 使不同模型表现最优的环境因子组合也可能存在模型种类的差异。为降低由环境因子组合带来的不确定性, 本研究采用重复随机抽取环境因子的方法, 即在 N 个气候因子中随机选择 n 个作为解释性变量, 重复 m 次, 即随机生成 m 个组气候因子组合。针对每个气候因子组合, 用 9 种空间分布模型分别进行模拟, 以寻找对不同模型具有普适意义的环境因子组合。本文对 19 个气候因子, 每次随机抽取 5 个, 重复 30 次, 并按顺序给每个环境因子组合编号。

1.4 BioMod 所采用的 9 种空间分布模型介绍

1.4.1 *广义线性模型* (GLM) GLM 是线性模型的扩展, 是针对应变量为非正态分布和非线性的统计学方法, 算法包括简单线性项、二次项和多项式项。当应变量为二元变量时, 需要通过 Logit 转换 (McCullagh 和 Nelder, 1989)。

1.4.2 *广义相加模型* (GAM) GAM 为 GLM 的半参数性扩展 (Hastie 和 Tibshirani, 1990), 通常适用于数据格式复杂、难以用标准线性或非线性模型拟合的情况。响应曲线的形状由数据本身决定, 而非事先指定的参数模型。一般通过统计平滑方法 (smoother) 来实现。其工作原理是将应变量对某个自变量作图, 在简约的前提下尽可能拟合趋近训练数集的平滑曲线。该算法对每个变量绘制一条平滑曲线, 并将结果相加。

1.4.3 *多元适应回归样条函数* (MARS) MARS 是一种非参数的回归技术, 其假设模型的解释变量在不同等级有不同的最优化参数 (Friedman, 1991)。因此根据解释变量的等级, 可分段进行回归模拟并确认各分段的参数。参数的临界点或阈值取决于样条函数结点, 样条函数结点通过运算自动确定。

1.4.4 *柔性判别分析* (FDA) FDA 分析为广义的线性判别分析, 不同的是它采用非参数算法替代线性参数算法 (Hastie 等, 1994)。FDA 假设每个环境变量的不同级别呈高斯分布。与线性判别分析相比, 它能针对不同标准的类别 (如混合高斯) 进行分类。环境参数构成初级类别, 进而分成亚类别, 分类结果由亚类别产生。亚类别数量取决于训练样本的变异程度。

1.4.5 *人工神经元网络* (ANN) ANN 为模仿生物神经网络结构和功能的数学模型或计算模型 (Ripley, 1996)。神经网络通过大量的人工神经元联结进行计算。现代神经网络是一种非线性统计建模工具, 常用于关系复杂的输入和输出变量间的建模, 或用来探索数据的模式。

1.4.6 *分类树分析* (CTA) CTA 通过对应变量的分析, 将由环境变量所确定的空间递归划分为尽量同质的类别 (Breiman 等, 1999)。建树的过程中, 采用一个简单规则, 基于环境变量, 不断将数据分组。每次分为两组, 每一组内尽量同质。每个分组节点的异质性通过偏差来表示。最优分类树的产生是平衡偏差最小及叶数量最少的结果。

1.4.7 *随机森林* (RF) RF 是一种新的通过机器学习和集成的分类方法, 是包含很多决策树的分类器。BioMod 采用 Breiman 和 Culter 用于分类和回归的随机森林代码 (Fortran) (Breiman, 2001)。该方法运用 Bagging 和随机选择的概念, 通过大量分类树运算得到最终结果。Bagging 即是 Bootstrap Aggregations, 是对样本进行多次重复 Bootstrap 取样的方法。若原始数据集包括 M 个变量和 N 条观察记录, 每次随机抽取含 m 个变量的 n 个随机样本 (同时进行样本回置 replacement) 作为每棵分类树的训练数集。变量分组的最佳阈值将作为划分每棵分类树节点的阈值, 且根据每次取样的训练数集构建一棵带评分的分类树。综合评估所有通过 Bootstrap 取样构建的分类树, 取评分最高的分类树及其标准为最终结果。在参数调试中, 注意 RF 对 m 敏感, 在 BioMod 中通常将 m 设置为 1/2M。

1.4.8 *推进式回归树* (GBM/BRT) 如果说 GLM 是在物种分布和环境因子之间寻找一个最简约的拟合模型, Boosting 则是用多个简单模型进行拟合, 最终综合各个结果形成最为优化的响应预测。在 BioMod 中采用推进式回归树算法 (Boosted Regression Tree) (Friedman, 2001; Ridgeway, 1999), 是一种在回归树上运用 Boosting 的方法。具体运算过程是构建一系列简单而有序的回归树以代表物种分布和环境变量之间的最优关系, 每一棵树的构建取决于其前一颗树的残差。最终结果为所有预测的加权平均值。

1.4.9 *表面分布区分室模型* (SRE) SRE 以物种存在点环境信息的最大和最小值来确定物种“信封”状的生态位。此方法简单、直观, 无需考虑解释性变量间的相互作用, 且所有解释性变量的权重一致, 结果为二元, 但不可进行外差推值 (Busby, 1991)。

1.5 BioMod 的运行、模型评估和集成

在理想的状况下, 进行物种空间分布建模和评估都需要物种已知分布点和已知非分布点。仅少部分空间分布模型如 Bioclim, DOMAIN, Habitat, ENFA, PCA species 可在已知非分布点缺失的情况下使用。然而在通常情况下, 人们缺乏对已知非分布点信息的记录和掌握 (Brotons 等, 2004), 因此研究者针对这个情况开发出了一些能够在已知分布点的基础上生成拟非分布点的方法 (Barbet-Massin 等, 2012; Phillips 等, 2009)。在 BioMod 中可设置需要产生的拟非分布点套数、每套数据的样本量、生成方法 (包括 circles、squares、per、random 和 sre) 以及与已知分布点之间的最短距离等 (详情参考 Thuiller 等, 2009 和 Barbet-Massin 等, 2012)。本研究采用 Circles 的方法随机产生 2 套 (PA1 和 PA2) 500 个拟非分布点, Circle 的最短距离设置为 0.5°。

准确率的评估应采用在统计学上具有独立性的样本。但为节约操作成本，可将样本一分为二进行拆分，一部分作为训练数集用于建模，另一部分用于模型评估。在 BioMod 中，不仅可设置样本拆分的比例，还可进行反复多次独立的样本拆分，最终的评估结果为多次拆分评估的平均值。这样就有效地避免了仅由一次随机样本拆分就得出评估结论的不确定性。本研究中，将 80% 的样本用于模型的训练，其余 20% 用于预测结果的精度评估。我们采用三种当前最为广泛使用的模型评估指标，分别是 Kappa，TSS 和 AUC（Allouche 等，2006；Fielding 和 Bell，1997）。Kappa 用于评估样本数据与模拟结果之间的一致性。AUC（area under the curve）用来评估模型对分布和非分布进行区分的能力。TSS（true skill statistics）为基于 Kappa 改良的方法，既保留了 Kappa 的优点，也校正了 Kappa 受物种分布广泛程度影响的缺点（Allouche 等，2006）。

本研究中将随机生成的 2 套拟非分布点（PA1 和 PA2）和已知分布点进行 2 次随机样本分割产生训练数集和评估数集，并针对每个随机选择的气候因子组合，进行 9 种物种空间分布模型的模拟运算。针对一组气候因子和一组拟非分布点产生 18 个模拟结果（9 个模型×2 次随机样本分割），并对 18 个模拟结果进行集成。方法是以 ROC 为权重，用 18 个模拟结果综合计算每个栅格所代表的空间位置上铁杉的分布概率。每组气候因子通过 2 套拟非分布点共产生 36 个模拟运算结果和 2 个集成结果。30 个环境因子组合共计产生 1 080 个模拟结果和 60 个集成结果（图 1）。选择表现最好的 5 个环境因子组合的模型集成作为最终产出。

2　结果

2.1　不同模型预测结果的精度比较

对比 BioMod 中的 9 个模型，3 种模型评估方法表现结果基本一致（图 2）。相比之下，RF 的模拟效果最好，平均 Kappa，TSS 和 AUC 分别达到 0.76，0.78，0.95。其次是 MARS 和 GAM，均能达到 Kappa>0.7，TSS>0.7，AUC>0.9。表现最差的是 SRE，KAPP 为 0.56，TSS 为 0.60，AUC 为 0.80。其他 4 种模型的表现介于上述几种模型之间。

2.2　不同气候因子组合对模型预测精度的影响

不同的气候因子组合，模拟能力和预测效果存在较大差异（图 3）。对于绝大部分的气候因子组合，模拟能力较为相似，平均模拟精度为 Kappa=0.65，TSS=0.7，AUC=0.8。其中有一些气候因子组合的预测精度显著高于其他组合，例如组合序号 1、2、11、21 和 22（表 1），能达

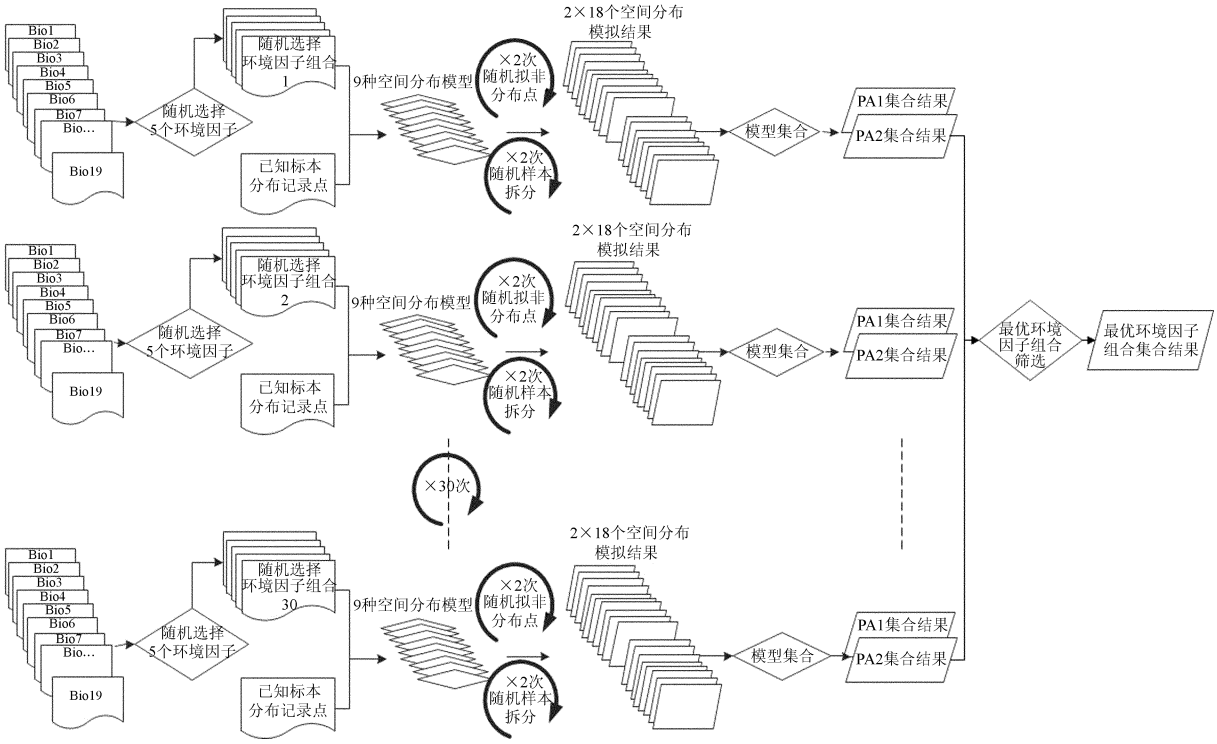

图 2　研究框架和技术线路

Fig. 2　Framework of the study

到 Kappa>0.77，TSS>0.70，AUC>0.91。同样，也有部分因子组合，如组合序号23和25，表现较差，其 Kappa 和 TSS 仅为0.55左右，AUC0.83 左右。表现较优的气候因子组合及其评估指标如表1所列。

2.3 模型集合结果

图4为基于表1中5个最佳气候因子组合，运用 BioMod 生成的两组拟非分布点 PA1 和 PA2，通过9个物种空间分布模型运算和集成产生的铁杉空间分布图。颜色及其数值大小代表铁杉在该

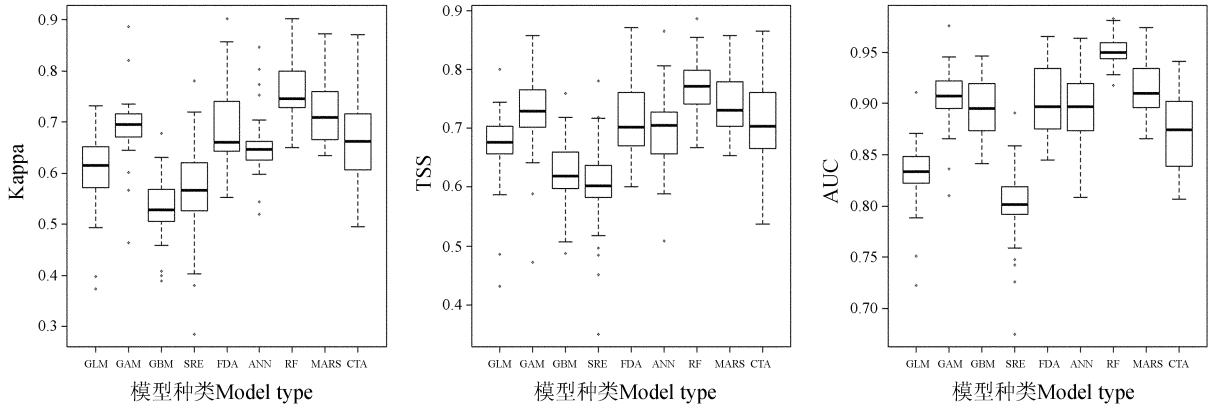

图3 9个模型的 Kappa，TSS 和 AUC 评估比较

Fig. 3 Kappa, TSS and AUC comparisons across nine models

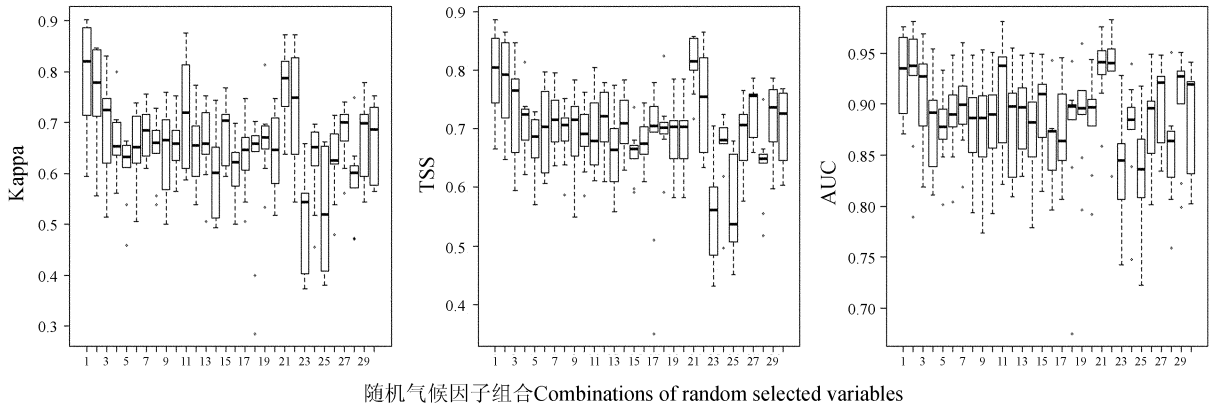

图4 30个随机选择产生环境因子组合的 Kappa，TSS，AUC 评估比较

Fig. 4 Kappa, TSS and AUC comparisons across 30 combinations of variables using random selection

表1 表现最优的5个气候因子组合及其通过9个模型模拟评估的 Kappa，TSS 和 AUC 平均值

Table 1 The five best performed variable-combinations and their average value of Kappa, TSS and AUC of model predictions across nine models

随机组合序号 No. of random selections	环境因子组合 Variable-combinations					Kappa	TSS	AUC
rdm1	Bio17 *	Bio15 **	Bio18 **	Bio11 **	Bio7	0.87	0.80	0.93
rdm2	Bio8	Bio11 **	Bio12	Bio16 *	Bio1 *	0.78	0.78	0.92
rdm11	Bio14	Bio16 *	Bio5 *	Bio18 **	Bio11 **	0.77	0.70	0.91
rdm21	Bio1 *	Bio15 **	Bio10	Bio17 *	Bio11 **	0.79	0.81	0.93
rdm22	Bio19	Bio15 **	Bio18 **	Bio5 *	Bio13	0.85	0.75	0.92

备注：** 标注的是在5个最优气候因子组合中出现3次的气候因子；* 为在5个最优气候因子组合中出现2次的气候因子

Note: ** denotes climatic variables occurred three times in the five best performed variable-combinations; * denotes two times occurred variables

栅格内的分布概率，数值越大表明铁杉的分布概率越高。从图中可看出，不同环境因子组合模拟得到的铁杉分布区存在一定差异，但大部分区域基本一致。总体而言，最适宜铁杉分布的区域集中在中国西南及四川盆地周围地区，零星分布于华南和台湾部分地区。

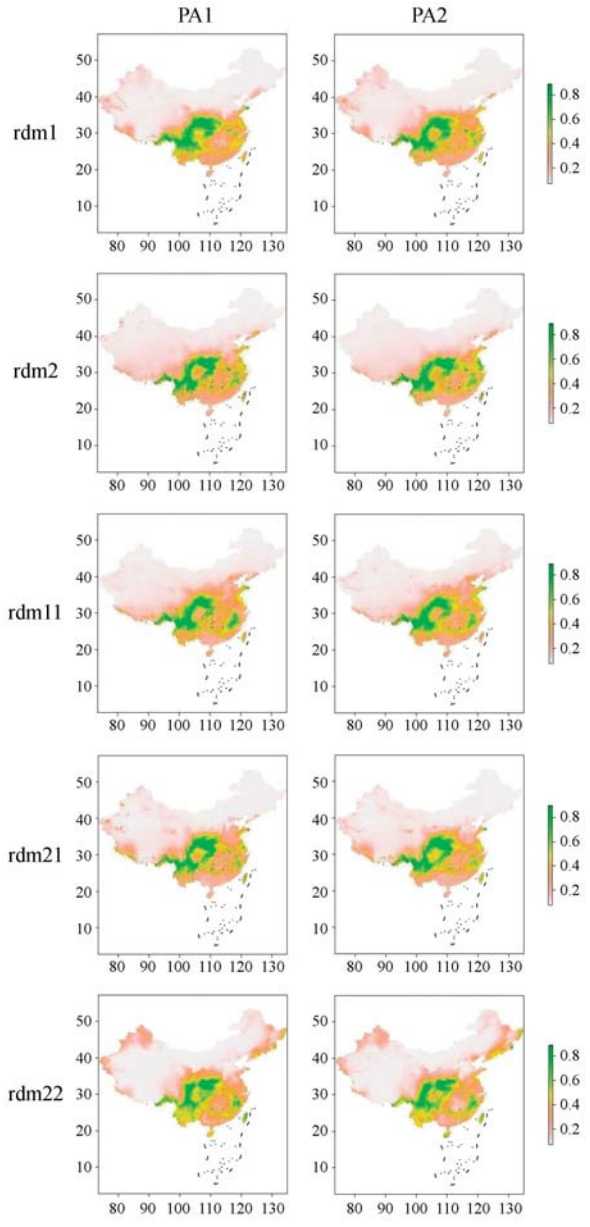

图 5　集成 9 个物种空间分布模型所生成的铁杉空间分布图。rm1，rm2，rm11，rm21 和 rm22 分别是 5 个最佳气候因子组合，PA1，PA2 为随机生成的两套拟非分布点

Fig. 5　Assembly maps of *T. chinensis* across nine models using the five best performed combinations of variables（rm1，rm2，rm11，rm21，rm22）and two sets of randomly produced Pseudo-absences（PA1 and PA2）

3　讨论

据文献记载（郑万钧和傅立国，1978），铁杉分布于甘肃白龙江流域、陕西南部、河南西部、湖北西部、四川东北部及岷江流域上游、大小金川流域、大渡河流域、青衣江流域、金沙江流域下游和贵州西北部海拔 1 200 ~ 3 200 m 地带，在河南、陕西、甘肃、湖北、四川东北部及贵州等地多呈星散分布，在四川西部峨边、泸定、天全等地尚有较大面积的森林。本文模拟的铁杉空间分布结果不仅能较好的反应文献记载的分布区，也和应俊生（1989）划分的铁杉属地理分布范围有较好的对应关系。

尽管铁杉在中国分布面积较广，且广泛栽培，但其对环境需求的生态学研究还较为缺乏（Del Tredici 和 Kitajima，2004）。通过本研究筛选出来的环境因子，为进一步了解该物种的生态位特征和环境适宜条件提供基础信息。在所筛选出来的 5 个最佳气候因子组合中，出现频率最高的气候因子为最冷季平均温（Bio11）、降水的季节性（Bio15）和最暖季降水量（Bio18），分别出现了 3 次，而且共同出现的机率较大（表 1）。其次是年均温（Bio1）、最暖月最高温（Bio5）、最湿季降水量（Bio16）和最干季降水量（Bio17），分别出现了两次。由此可以判断，铁杉的空间分布主要受温度和降水的综合影响，特别是最冷季温度和最暖季的降水量。有研究表明，最冷季（或者说冬季）低温与物种对霜冻的响应较为密切，是决定针叶树分布的主要环境因子（Bannister 等，2001）。对铁杉标本分布点的气候特征研究发现适宜铁杉分布的最冷季平均温的平均值为 1.7℃（SD=4.3），最暖季降水量的平均值为 482 mm（SD=126）且具有明显的降水季节性。

如果采用自选的环境因子组合，不一定产生最佳的模拟效果，究其原因可能是各因子间存在共线性或一些重要的生态学本质未被发现。通过重复随机组合环境因子，使我们有机会掌握在随机条件下，模型本身能达到的模拟效果。分析发现，即使在没有任何生态学背景知识和统计筛选的协助下，利用任意随机环境因子组合也能取得可接受的模拟效果。研究还发现某些气候因子组合的模拟效果确实高于其他组合，预示着这些环境因子对物种空间分布的重要性。而表现较差的

环境因子组合正好说明其与物种的分布相关性不大。因此，在缺乏生态学背景知识的情况下，重复多次随机组合可作为一种筛选有效环境因子的替代办法，但须从生态学和生物学角度对其结果的合理性进行判断和解释。

从模型的表现来看，由简单到复杂，模拟效果越来越好。以 SRE 为代表的传统气候分室模型在一定程度上能够模拟铁杉的空间分布，而以 RF，GAM 和 MARS 为代表的新技术的运用能带来更好的模拟效果。不同模型模拟效果之间的差异，可能是各个模型对识别基础生态位和现实生态位的差别所致。气候分室模型的模拟结果可能更接近基础生态位，是在无其他因素干扰下，物种可占据的环境空间及其在地理空间中的映射。而在现实的自然生态系统演化过程，由于其他人为干扰（如土地利用转变）、非生物环境因素和生物因素（如竞争、排斥、传粉、种子散播和协同进化）的共同作用导致部分基础生态位未被目标物种所占据。RF，GAM 和 MARS 通过其复杂的运算技术可能或多或少的捕捉到了由上述因素干扰的结果。

对于既定的目标物种，通过对比能筛选出表现较优的模型。但通过一个物种筛选出来的最优模型，并不能推而广之，运用于其他物种。众多研究表明，预测效果受很多因素的影响，除模型类别和参数设置外，还有与物种相关的特征，如分布区环境特征、物种分布广度、分布的聚散模式以及稀有性等（Marmion 等，2009）。BioMod 研究平台的开发，能同时对比多个基于不同生态学原理的空间分布模型，使物种空间分布研究朝着更加高效和准确的方向发展。

致谢 感谢汪铁军博士对本文提出修改意见以及 Damien Georges 协助调试 BioMod 软件包。

〔参 考 文 献〕

郑万钧，傅立国，1978. 中国植物志（第7卷）[M]. 北京：科学出版社

Allouche O，Tsoar A，Kadmon R，2006. Assessing the accuracy of species distribution models：prevalence，kappa and the true skill statistic（TSS）[J]. *Journal of Applied Ecology*，**43**（6）：1223—1232

Araújo MB，New M，2007. Ensemble forecasting of species distributions [J]. *Trends in Ecology and Evolution*，**22**（1）：42—47

Austin M，2007. Species distribution models and ecological theory：a critical assessment and some possible new approaches [J]. *Ecological Modelling*，**200**（1-2）：1—19

Austin MP，2002. Spatial prediction of species distribution：an interface between ecological theory and statistical modeling [J]. *Ecological Modelling*，**157**（2-3）：101—118

Austin MP，Van Niel KP，2011. Improving species distribution models for climate change studies：variable selection and scale [J]. *Journal of Biogeography*，**38**（1）：1—8

Bannister P，Neuner G，Bigras F *et al.*，2001. *Frost Resistance and the Distribution of Conifers*，*Conifer Cold Hardiness* [M]. Spuiboulevard：Kluwer Academic Publishers，3—21

Barbet-Massin M，Jiguet F，Albert CH *et al.*，2012. Selecting pseudo-absences for species distribution models：how，where and how many? [J]. *Methods in Ecology and Evolution*，**3**（2）：327—338

Barry S，Elith J，2006. Error and uncertainty in habitat models [J]. *Journal of Applied Ecology*，**43**（3）：413—423

Bellard C，Bertelsmeier C，Leadley P *et al.*，2012. Impacts of climate change on the future of biodiversity [J]. *Ecology Letters*，**15**（4）：365—377

Breiman L，2001. Random forests [J]. *Machine Learning*，**45**（1）：5—32

Breiman L，Friedman J，Olshen R *et al.*，1999. *Classification and Regression Trees* [M]. New York：CRC Press

Brotons L，Thuiller W，Araújo MB *et al.*，2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability [J]. *Ecography*，**27**（4）：437—448

Buisson L，Thuiller W，Casajus N *et al.*，2010. Uncertainty in ensemble forecasting of species distribution [J]. *Global Change Biology*，**16**（4）：1145—1157

Busby JR，1991. BIOCLIM-a bioclimate analysis and prediction system [J]. *Plant Protection Quarterly*，**6**（1）：8—9

Canhos VP，Souza Sd，Giovanni RD *et al.*，2004. Global Biodiversity Informatics：setting the scene for a "new world" of ecological forecasting [J]. *Biodiversity Informatics*，**1**：1—13

Del Tredici P，Kitajima A，2004. Introduction and cultivation of Chinese hemlock（*Tsuga chinensis*）and its resistance to hemlock woolly adelgid（*Adelges tsugae*）[J]. *Journal of Arboriculture*，**30**（5）：282—287

Elith J，Graham HC，Anderson PR *et al.*，2006. Novel methods improve prediction of species' distributions from occurrence data [J]. *Ecography*，**29**（2）：129—151

Fielding AH，Bell JF，1997. A review of methods for the assessment of prediction errors in conservation presence/absence models [J]. *Environmental Conservation*，**24**（1）：38—49

Franklin J，1998. Predicting the distribution of shrub species in southern California from climate and terrain-derived variables

(abstract GEO-BASE) [J]. *Journal of Vegetation Science*, **9** (5): 733—748

Friedman JH, 1991. Multivariate adaptive regression splines [J]. *The Annals of Statistics*, **19** (1): 1—67

Gallien L, Münkemüller T, Albert CH *et al.*, 2010. Predicting potential distributions of invasive species: where to go from here? [J]. *Diversity and Distributions*, **16** (3): 331—342

Guisan A, Thuiller W, 2005. Predicting species distribution: offering more than simple habitat models [J]. *Ecology Letters*, **8** (9): 993—1009

Guisan A, Zimmermann NE, 2000. Predictive habitat distribution models in ecology [J]. *Ecological Modelling*, **135** (2-3): 147—186

Hastie T, Tibshirani R, Buja A, 1994. Flexible discriminant analysis by optimal scoring [J]. *Journal of the American Statistical Association*, **89** (428): 1255—1270

Hastie TJ, Tibshirani R, 1990. *Generalized Additive Models* [M]. London: Chapman & Hall/CRC

Heibl C, Renner SS, 2012. Distribution models and a dated phylogeny for Chilean Oxalis species reveal occupation of new habitats by different lineages, not rapid adaptive radiation [J]. *Systematic Biology*, **61** (5): 823—834

Leathwick JR, Elith J, Hastie T, 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions [J]. *Ecological Modelling*, **199** (2): 188—196

Loiselle BA, Jørgensen PM, Consiglio T *et al.*, 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? [J]. *Journal of Biogeography*, **35** (1): 105—116

Maiorano L, Cheddadi R, Zimmermann NE *et al.*, 2012. Building the niche through time: using 13 000 years of data to predict the effects of climate change on three tree species in Europe [J]. *Global Ecology and Biogeography*: http://wileyonlinelibrary.com/journal/geb

Marmion M, Luoto M, Heikkinen RK, 2009. The performance of state-of-the-art modelling techniques depends on geographical distribution of species [J]. *Ecological Modelling*, **220** (24): 3512—3520

McCullagh P, Nelder JA, 1989. *Generalized Linear Models* [M]. London: Chapman & Hall/CRC, 35

McPherson JM, Jetz W, 2007. Effects of species' ecology on the accuracy of distribution models [J]. *Ecography*, **30** (1): 135—151

Naimi B, Skidmore AK, Groen TA *et al.*, 2011. Spatial autocorrelation in predictors reduces the impact of positional uncertainty in occurrence data on species distribution modeling [J]. *Journal of Biogeography*, **38** (8): 1497—1509

Petitpierre B, Kueffer C, Broennimann O *et al.*, 2012. Climatic niche shifts are rare among terrestrial plant invaders [J]. *Science*, **335** (6074): 1344—1348

Phillips SJ, Anderson RP, Schapire RE, 2006. Maximum entropy modeling of species geographic distributions [J]. *Ecological Modelling*, **190** (3-4): 231—259

Phillips SJ, Miroslav D, Jane E *et al.*, 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data [J]. *Ecological Applications*, **19** (1): 181—197

Prasad A, Iverson L, Liaw A, 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction [J]. *Ecosystems*, **9** (2): 181—199

Pulliam HR, 2000. On the relationship between niche and distribution [J]. *Ecology Letters*, **3** (4): 349—361

Ridgeway G, 1999. The state of boosting [J]. *Computing Science and Statistics*, **31**: 172—181

Ripley BD, 1996. *Pattern Recognition and Neural Networkds* [M]. London: Cambridge University Press

Schmidt M, Kreft H, Thiombiano A *et al.*, 2005. Herbarium collections and field data-based plant diversity maps for Burkina Faso [J]. *Diversity and Distributions*, **11** (6): 509—516

Smith SA, Donoghue MJ, 2010. Combining historical biogeography with niche modeling in the Caprifolium clade of *Lonicera* (Caprifoliaceae, Dipsacales) [J]. *Systematic Biology*, **59** (3): 322—341

Thuiller W, 2004. Patterns and uncertainties of species' range shifts under climate change [J]. *Global Change Biology*, **10** (12): 2020—2027

Thuiller W, Lafourcade B, Engler R *et al.*, 2009. BIOMOD-a platform for ensemble forecasting of species distributions [J]. *Ecography*, **32** (3): 369—373

Václavík T, Meentemeyer RK, 2012. Equilibrium or not? Modelling potential distribution of invasive species in different stages of invasion [J]. *Diversity and Distributions*, **18** (1): 73—83

Van Horssen PW, Pebesma EJ, Schot PP, 2002. Uncertainties in spatially aggregated predictions from a logistic regression model [J]. *Ecological Modelling*, **154** (1-2): 93—101

Wiens JA, Stralberg D, Jongsomjit D *et al.*, 2009. Niches, models, and climate change: Assessing the assumptions and uncertainties [J]. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (2): 19729—19736

Williams KJ, Belbin L, Austin MP *et al.*, 2012. Which environmental variables should I use in my biodiversity model? [J]. *International Journal of Geographical Information Science*, **26** (11): 2009—2047

Ying JS (应俊生), 1989. Areography of the gymnosperms of China (1)-distribution of the Pinaceae of China [J]. *Acta Phytotaxonomic Sinica* (植物分类学报), **27** (1): 27—38