

# 基于 BIOMOD 的青海湖流域 高原鼠兔分布模拟

吴艺楠<sup>②</sup> 马育军<sup>①②\*</sup> 刘文玲<sup>②</sup> 李小雁<sup>①②</sup> 王佩<sup>①②</sup>

① 北京师范大学地表过程与资源生态国家重点实验室 北京 100875; ② 北京师范大学资源学院 北京 100875

**摘要:** 随着统计模型及空间信息数据的不断发展和完善, 物种分布模型已经成为全球变化背景下研究大尺度物种分布情况的重要工具。高原鼠兔 (*Ochotona curzoniae*) 是青藏高原特有的关键物种, 在青藏高原生态系统中占有重要地位。通过采集高原鼠兔的分布点数据及环境变量数据, 基于 R 语言中 BIOMOD 包中的 7 个模型对其在青海湖流域的分布进行了模拟。结果表明, 高原鼠兔主要分布于青海湖西岸和北岸、天峻县周边及布哈河流域上游, 影响高原鼠兔分布的主要环境因子为距道路距离、距居民点距离、最暖月最高气温、NDVI 标准差、最冷季和最干季降水量。BIOMOD 组合模型中, 推进式回归树模型 (GBM) 和最大熵模型 (MAXENT) 的模拟效果最好, 广义线性回归模型 (GLM) 结果较差。而优化后的结果显示, 模拟结果的集成和筛选能有效提高模型的精度和效果。

**关键字:** 物种分布模型; BIOMOD; 高原鼠兔

**中图分类号:** Q958 **文献标识码:** A **文章编号:** 0250-3263 (2017) 03-390-13

## Modelling the Distribution of Plateau Pika (*Ochotona curzoniae*) in Qinghai Lake Basin Using BIOMOD

WU Yi-Nan<sup>②</sup> MA Yu-Jun<sup>①②\*</sup> LIU Wen-Ling<sup>②</sup> LI Xiao-Yan<sup>①②</sup> WANG Pei<sup>①②</sup>

① *State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875;*

② *School of Natural Resources, Beijing Normal University, Beijing 100875, China*

**Abstract:** Species distribution model has become an important tool to study the species distribution at large-scale in the context of global change due to the development and improvement of statistical models and spatial information data. Plateau Pika (*Ochotona curzoniae*) is a keystone species in the Qinghai-Tibet Plateau and plays an important role in the entire ecosystem. The Qinghai Lake Basin is located in the northeast of the Qinghai-Tibetan Plateau and is a typical closed inland basin with a watershed area of approximately 29 661 km<sup>2</sup> that (Fig. 1). This research aimed to model the distribution of Plateau Pika in the Qinghai Lake Basin using seven models from BIOMOD package in R with occurrence data and

**基金项目** 国家自然科学基金项目 (No. 41301013, 41130640);

\* 通讯作者, E-mail: myj3648@163.com;

**第一作者介绍** 吴艺楠, 男, 硕士; 研究方向: 生态系统模拟; E-mail: ggg\_0953@qq.com。

收稿日期: 2016-12-22, 修回日期: 2017-03-12 DOI: 10.13859/j.cjz.201703004

environmental variables. AUC (area under the curve) and TSS (true skill statistic) based on confusion matrix (Table 1) were chosen to evaluate the performance of different models. The results showed that the Plateau Pika mainly distributed in the west and north bank of Qinghai Lake, around Tianjun county and in the upstream of the Buha River (Fig. 3 and Fig. 4). The most important environmental factors affecting the distribution of Plateau Pika were the distance to road and, to the settlement of people, the air temperature of the warmest month, the NDVI standard deviation, and the precipitation of the coldest and driest season (Table 2). The Boost Regression Tree model (GBM) and Maximum Entropy model (MAXENT) make the best predictions, while the Generalized Linear Model (GLM) gives a poor result (Fig. 2). The optimized result shows that the integration and selection of can improve the accuracy and performance of the model effectively (Fig. 5 and Fig. 6).

**Key words:** Species distribution model; BIOMOD; Plateau Pika, *Ochotona curzoniae*

高原鼠兔 (*Ochotona curzoniae*) 是青藏高原的特有物种, 属兔形目鼠兔科, 多栖息在海拔 3 000 m 以上土壤较为疏松的坡地和河谷, 广泛分布于高山草甸、草甸草原和荒漠草原 (Smith et al. 2008), 具有植食、穴居、集群的习性。高原鼠兔是肉食性兽类和猛禽的主要食物来源, 在生态系统食物网结构中占有承上启下的地位 (周雪荣等 2010), 是青藏高原生态系统的物种 (Smith et al. 1999)。对于高原鼠兔在整个青藏高原生态系统中的作用始终存在争论, 一方面它们被认为是加速草原退化、威胁草原生态系统安全的有害物种 (韩天虎等 2008, 李建川等 2010), 另一方面也有研究认为适当密度的高原鼠兔活动可以加快养分循环, 改变土壤微环境, 促进牧草生长, 提高生物多样性, 改善生态系统功能 (Aho et al. 1998)。另有研究表明, 高原鼠兔是气候变化背景下生态系统响应的指示器, 能在一定程度上反映气候变化对生态环境的影响 (Smith et al. 2004)。目前已有大量关于高原鼠兔生态学特性 (殷宝法等 2004, 齐新章等 2008)、食性 (赵廷贵等 2002, 赵廷贵 2005)、高原鼠兔扰动后植被及土壤响应 (石红霄等 2010, 贾婷婷等 2014) 等方面的研究。在点尺度上, 也有学者对高原鼠兔生境选择进行了研究, 王涪等 (2004) 研究发现影响高原鼠兔生境选择的主要因子依次是生境位置、土壤质地、离水源

的距离、灌丛植物盖度和阔叶植物高度; 施银柱 (1983) 研究表明, 半垦地之类植被状况较差的地区有利于鼠兔生存。然而, 大尺度上高原鼠兔空间分布格局的研究非常缺乏。

近年来, 随着计算机技术和统计学工具的不断丰富, 物种分布模型 (species distribution modelling, SDM) 被国内外科学家广泛用于模拟研究动植物在大尺度上的分布情况, 其基本原理是利用物种的分布数据 (包括存在点和非存在点) 及不同的环境栅格层, 依据不同的算法估计物种的生态位, 并投影到景观中, 以概率的形式反映物种对生境的偏好程度, 结果可以解释为物种出现的概率、生境适宜度或物种丰富度等 (李国庆等 2013)。伴随物种分布模型数量的不断增加, 许多模型并未开发出可视化界面, 调用不同算法非常困难。另外, 利用不同模型模拟不同物种在不同环境下的精度差异可能非常大, 并且受到样本量的影响。因此, 法国的 Wilfried Thuiller 博士在 2003 年提出了一个新型计算机平台框架 BIOMOD, 该框架使用 10 种不同的建模技术, 利用参数估计、非参数估计及机器学习等不同原理, 预测当前物种分布并综合评估, 然后只选择准确度符合要求的模型对每个物种进行预测, 以便达到最大精度, 大大简化了不同模型的应用难度, 解决了预测不同物种分布时的模型选择问题 (Thuiller 2003)。特别是该模型在 2016 年的更新版

(BIOMOD2) 中整合了当下物种分布研究中最热门的 Maxent 模型, 使其更加完善。

本文应用 BIOMOD 组合模型, 对高原鼠兔在青海湖流域的分布进行模拟, 探究高原鼠兔在大尺度空间的分布规律, 并揭示影响高原鼠兔分布的主要环境因子。

## 1 BIOMOD 基本原理

目前, BIOMOD2 提供了 10 种常见的物种分布模型供科研工作者选择。其所需输入数据为物种存在点数据、非存在点数据以及与该物种分布相关的环境变量, 如果非存在点数据缺失, BIOMOD 中也可以将背景点作为伪非存在点数据。输入环境变量为栅格图, 并且所有环境变量层需统一坐标系、范围和栅格大小。下面分别介绍 BIOMOD 中包含模型的基本原理。

### 1.1 表面分布区分室模型

表面分布区分室模型 (surface range envelope, SRE) 也称为 BIOCLIM 模型, 它汇总物种存在点每个环境变量的最大值和最小值, 预测的物种分布范围环境变量取值在这些最大值和最小值之间。此方法简单、直观, 无需考虑解释变量间的相互作用, 而且所有解释变量的权重统一, 得到适宜或不适宜分布的二元结果 (Busby 1991)。该模型的应用主要集中于气候因素对物种分布状况的影响及其对气候变化的响应, 例如 Beaumont 等 (2005) 利用 SRE 基于 35 种不同气候变量预测了澳大利亚 25 种蝴蝶的分布状况。

### 1.2 广义线性模型

广义线性模型 (generalized linear model, GLM) 是多元线性回归模型的推广, 是针对因变量为非正态分布和非线性时的改进模型, 它允许因变量为二项分布、泊松分布等离散型的分布, 因此比一般线性模型具有更广泛的应用 (Nelder et al. 1972)。BIOMOD 中集成的 GLM 模型应用自动逐步回归法, 可以通过选择模型拟合优度的统计量 Akaike Information Criterion (AIC) 或 Bayesian Information Criteria

(BIC) 去除冗余变量来筛选相对重要的环境因素, 从而提高模型精度 (Thuiller et al. 2009)。Guisan 等 (1999) 利用 GLM 模型预测了美国内华达州春山山脉中所有高山植被的分布, 结果表明应用回归分析剔除相关性低的环境因子后精度更高。

### 1.3 广义加性模型

广义加性模型 (generalized additive model, GAM) 是广义线性模型和可加模型的混合, 在数据格式复杂、难以用一般广义线性模型拟合时应用较多。其工作原理是将因变量对每个自变量绘制一条平滑曲线, 并将结果相加, 尽可能趋近于训练数据集的平滑曲线。该模型通过光滑样条函数进行局部优化, 因此残差相对较小, 更适宜处理因变量和自变量比较复杂的关系, 数据量越大精度越高 (Hastie et al. 1986)。Araújo 等 (2007) 利用 GAM 成功模拟了欧洲觅梦绢蝶 (*Parnassius mnemosyne*) 在不同气候和食物条件组合下的分布状况。

### 1.4 多元适应回归样条函数

多元适应回归样条函数 (multiple adaptive regression splines, MARS) 可看作扩展的线性模型, 可对变量之间的非线性及相互作用进行建模, 其假设模型的解释变量在不同等级具有不同的最优化参数 (Friedman 1991)。MARS 的优越之处在于样条函数结点是通过运算自动确定的, 同时可估计变量间的交互作用 (Friedman 1991)。Leathwick 等 (2006) 应用 GAM 和 MARS 模拟了新西兰 15 种淡水鱼的分布, 发现 MARS 在大样本情形下表现更好。

### 1.5 柔性判别分析

柔性判别分析 (flexible discriminant analysis, FDA) 是一种应用于多种元素的分类工具, 为广义线性判别分析的一种, 其采用非参数算法替代线性参数算法来改进参数估计过程 (Hastie et al. 1994)。Reiss 等 (2011) 应用包括 FDA 在内的 9 种不同模型成功模拟了 20 种北冰洋深海鱼类的分布状况。

### 1.6 分类树分析

分类树分析 (classification tree analysis, CTA) 通过对多个变量的分析, 按照区分的难易程度不断地将空间环境分为同质组, 从而进行物种分布预测, 分类树分析根据自变量的一致性, 对因变量进行重复分组, 类似树的枝叶形成过程。每次数据均被分为两组, 每一组尽量同质, 即组内方差最小 (Breiman et al. 1984)。BIOMOD 中集成的 CTA 模型应用 10 个检验点集交叉验证的方法来取得解释的变异和枝叶数间的平衡值, 并且可以设置分类的方法和节点数 (Thuiller et al. 2009)。Thuiller 等 (2003) 应用 GLM、GAM、CTA 三种模型, 模拟了 4 种地中海树种在 3 种不同尺度上的分布状况, 三种模型的预测精度差异不大。

### 1.7 推进式回归树

推进式回归树 (boosted regression tree, BRT, 也称 generalized boosted regression models, GBM) 运用推进方法对多个简单模型进行拟合, 可以提高运算速度, 然后综合各个结果形成最优化的响应预测 (Ridgeway 1999, Friedman 2001)。该方法运算过程是构建不同的回归树来代表物种分布和环境变量之间的最优关系, 新树的构建取决于其前一棵树的残差最小值, 最终对所有预测值进行加权平均。相比于传统回归树模型, 该模型可以在迭代过程中不断调整不同解释变量的权重和解释变量在不同区间的权重, 达到优化分类的目的 (Friedman 2001)。Pearman 等 (2008) 利用 BRT 成功模拟了欧洲 7 种乔木的分布, 发现模拟结果和现实结果非常一致。

### 1.8 人工神经网络

人工神经网络 (artificial neural network, ANN) 是一种模仿生物神经网络结构和功能的数学模型 (Ripley 2007), 常用于关系复杂的输入和输出变量间的建模。它的原理是在类似的神经网络系统中构建模型的多重组合, 通过定义节点权重的衰减程度和一定的隐元, 趋近训练集产生的平滑函数。人工神经网络具有通过调整模型参数进行学习的功能 (Ripley

2007), 但运算时间较长, 需要设置合适的权重避免多余的交叉验证。Pearson 等 (2004) 利用 ANN 模拟了英国 4 种植物的潜在分布区, 结果表明其分布与气候变化以及土地利用情况有密切关系。

### 1.9 随机森林

随机森林 (random forest, RF) 是一种机器学习的分类方法, 它集成了回归和分类的双重思想, 通过构建不同决策树来进行单个树的拟合。BIOMOD 中集成的 RF 模型采用 Breiman 和 Cutler 用于分类和回归的随机森林代码。整个方法包括树的构建阶段、打分阶段和选择阶段, 其工作原理相对独特, 在构建每个分类树时考虑因素更多, 因此误差相对较低 (Breiman 2001)。BIOMOD 中可以设置构建分类树的个数 (默认为 500)、节点大小 (默认为 5)、最大节点数 (默认为无限) 等, 分类树个数较多可以提高运行精度, 但同时会增加运行时间, 因此在设置时需要进行取舍。Cutler 等 (2007) 利用 RF 模拟了美国加州 4 种入侵植物的分布状况, 结果表明, 相比于传统分类树模型及回归模型, RF 具有更好的模拟精度。

### 1.10 最大熵模型

最大熵模型 (maximum entropy, MAXENT) 基于热力学第二定律。按照该定律, 一个非均衡的生命系统通过与环境的物质和能量交换以保持其存在, 也就是说, 一个实际存在的系统具有“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, 而此时系统与环境之间的关系达到平衡状态。在物种潜在分布的相关研究中, 可将物种与其生长环境视为一个系统, 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 (Phillips et al. 2006)。Peterson 等 (2007) 利用 MAXENT 及 GARP 模拟了 3 种北美鸟类的分布, 结果表明, MAXENT 模拟策略比 GARP 保守, 预测精度受参数值的影响较大。

## 2 高原鼠兔分布模拟方法

### 2.1 研究区域

研究区域选择在青藏高原东北部的青海湖流域,地理位置介于 36°32'~37°15'N, 99°36'~100°47'E 之间(图 1)。流域地形西北高、东南低,海拔在 3 194 ~ 5 174 m 之间,总面积为 29 661 km<sup>2</sup>。流域内植被类型包括灌丛植被(温性河谷灌丛与高寒灌丛)、草原植被(温性草原和高寒草原)、高山流石植被、草甸和沼泽植被(高寒草甸、盐生草甸和沼泽草甸)以及栽培植被和沙生植被等(侯佩玲等 2004)。

### 2.2 野外调查

高原鼠兔分布数据采集于 2015 年 8 月及 2016 年 6 月的野外调查,共取得 53 个存在点和 41 个不存在点的数据(图 1)。高原鼠兔分布调查采用主观抽样的方法,在研究区域内不同地貌类型区随机寻找高原鼠兔的踪迹,用 GPS 记录鼠兔存在点和不存在点的位置。所有调查点经过筛选,距离相隔 2 km 以上,以排除

空间自相关的影响。

### 2.3 环境数据

建模共使用 35 种不同环境数据,主要包括以下几类。

(1) 地形数据:地形因素和生物分布密切相关,并且和其他生物因子相比,地形因子相当稳定(Rosenzweig 1995)。本次研究所用的数字高程数据(digital elevation model, DEM)来自中国国家测绘局,空间分辨率为 25 m,坡度和坡向根据 DEM 在 ArcGIS 10.2(ESRI, Inc., Redlands, CA, USA)中提取获得。将所有数据在 ENVI 5.1 中进行重采样,转换成 asc 格式分辨率为 1 km × 1 km 的栅格图层,共 3 个环境层。

(2) 生物气候数据:生物气候数据从 WorldClim 数据库中获得,该数据库根据 1950~2000 年超过 4 000 个气象站的观测数据生成了全球生物气候数据层,包括连续年变化的降水和温度数据(Hijmans et al. 2005)。本研究利

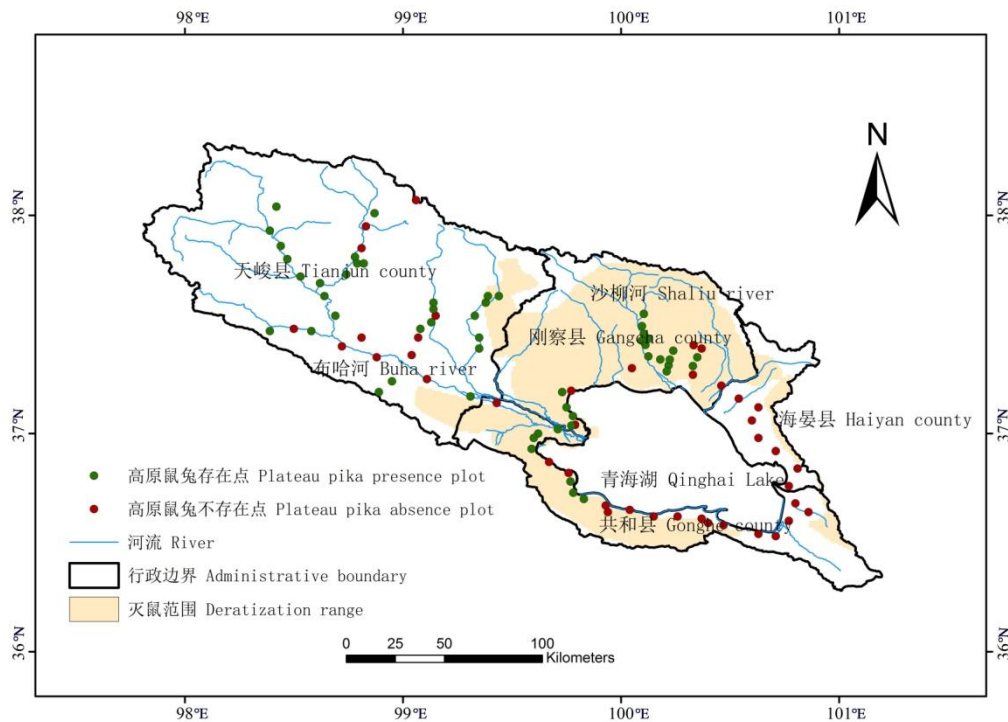


图 1 青海湖流域地理位置及高原鼠兔分布采样点

Fig. 1 The location of Qinghai Lake Basin and the sampling points for Plateau Pika

用其中的 19 个环境数据层，包括年平均气温、气温月较差、气温的等温性、气温的季节性、最暖月最高气温、最冷月最低气温、气温年较差、最湿季平均气温、最干季平均气温、最暖季平均气温、最冷季平均气温、年降水量、最湿季降水量、最干季降水量、降水的季节性、最湿季降水量、最干季降水量、最暖季降水量和最冷季降水量，空间分辨率为 1 km。

(3) 植被数据：采用 MODIS/Terra-NDVI 产品，空间分辨率为 1 km，时间分辨率为 16 d。采用近 10 年的归一化差分植被指数 (normalized difference vegetation index, NDVI) 值代表植被生长平均状况，借助 upper-envelope 方法进行图像平滑处理。由于单独的 NDVI 值反映植被生长状况可能存在一定偏差，所以通过 ENVI 5.1 计算 NDVI 最大值、NDVI 平均值、NDVI 最小值以及 NDVI 标准差一起作为输入的环境因子，共 4 个环境层。

(4) 土壤数据：土壤数据包括土壤容重、黏粒含量、砂粒含量、有机碳含量，来自 SoilGrids 系统 (www.soilgrids.org)。它是由 International Soil Reference and Information Centre (ISRIC) 根据世界土壤信息产生的全球土壤数据产品，也是可更新的土壤属性和类别地图集合，最初在 1 km 分辨率下，使用基于模型的统计方法获得 (Hengl et al. 2014)，包含 0~2 m 的土壤信息。在本研究中，考虑到高原鼠兔仅分布于浅层土壤中，因此选用地表以下 0~5 cm 的土壤容重、黏粒含量、砂粒含量、粗颗粒体积占比、有机碳含量数据进行模拟，共 5 个环境层。

(5) 其他数据：研究表明距道路、水源及居民点的距离对鼠兔的分布均有一定影响 (王涓等 2004)，因此本研究采用的矢量数据包括青海湖流域的路网数据、河流分布数据及居民点分布数据，从寒区旱区科学数据中心 (http://westdc.westgis.ac.cn/) 获得，在 ArcGIS 10.2 中利用欧几里德距离功能计算得到距离栅格地图用于模拟。自 20 世纪 80 年代以来，当

地政府多次组织长时间大面积的灭鼠活动，对高原鼠兔的分布产生一定影响，因此，从青海省农牧信息网 (http://www.qhagri.gov.cn/sites/MainSite/) 查阅了近年来的灭鼠情况，并将有灭鼠措施的地区标记为 1、没有灭鼠措施的地区标记为 0，作为二值类型变量输入模型。

## 2.4 模型选择与精度评价

为了评估模型模拟质量，我们将数据库划分为两个子集：校准和评估，数据库 70% 的随机样本用于校准 (训练) 模型，剩余 30% 数据用于评估 (测试) 模型 (Fielding et al. 1997)。由于研究区面积相对较小且部分环境因子高度相关，GAM 模型获取估计参数时十分困难。同时，以 1 和 0 二值变量表征灭鼠措施影响，SRE 模型和 MARS 模型不能准确地估计参数。因此选用 BIOMOD 中剩余的 7 种模型进行模拟。在 MAXENT 模型运行时将距离道路 10 km 范围设置为背景点偏差校正区，以减小采样点在道路附近的采样误差。运行时随机选取训练点和测试点，重复运行 10 次，共 70 个模拟结果。BIOMOD2 中提供了多种精度评价指标，本次模拟采用当前广泛使用的 TSS 和 AUC 对不同模型的模拟结果进行评价 (Fielding et al. 1997, Allouche et al. 2006)。

根据物种分布模型评价误差矩阵 (表 1)，TSS 表示实测样本 (包括分布样本和未分布样

表 1 物种分布模型评价误差矩阵

Table 1 Confusion matrix of species distribution modelling evaluation

预测结果 Result	预测分布 Predicted	预测未分布 Predicted absence
实测分布 Investigated presence	a	b
实测未分布 Investigated absence	c	d

a 表示实测分布的预测正确值；b 表示实测分布的预测错误值；c 表示实测未分布的预测错误值；d 表示实测未分布的预测正确值。

a. True predicted presence; b. False predicted presence; c. False predicted absence; d. True predicted absence.

本)上的净预测成功率,  $TSS = \frac{a}{a+c} + \frac{b}{b+d} - 1$ ,  $a$ 、 $b$  分别表示实测分布的预测正确值和错误值,  $c$ 、 $d$  分别表示实测为分布的预测错误值和正确值。TSS 数值越大, 表明模型的预测结果越好。

AUC (area under the curve) 源于受试者工作特征曲线 (receiver operating characteristic curve, ROC), 该曲线根据一系列不同的二分类方式 (分界值或决定阈), 以真阳性率 (灵敏度, 即  $\frac{a}{a+c}$ ) 为纵坐标, 假阳性率 (1 - 特异度, 即  $\frac{b}{b+d}$ ) 为横坐标。在模型预测的概率分布图中, 通过取不同的阈值 (范围为 0 ~ 1), 可以得到不同的误差矩阵, 进而获得不同的横坐标和纵坐标, 用曲线拟合这些点, 得到的便是受试者工作特征曲线, 对该曲线下的面积进行积分, 则可得 AUC (Fielding et al. 1997)。AUC 的值域为 (0.5, 1), 0.5 代表完全随机的分类, 1 表示完全正确的分类, 因此, AUC 值越大, 表明模型的预测性能越好。

### 3 高原鼠兔分布模拟结果

#### 3.1 不同模型结果精度比较

对比 BIOMOD 中的 7 个模型, 两种模型评估方法表现结果基本一致 (图 2)。相比之下, GBM 和 MAXENT 模拟效果最好, AUC 的平均值分别为 0.886 和 0.889, TSS 平均值分别为 0.690 和 0.673; 其次为 FDA 和 RF, AUC 平均值大于 0.8, TSS 平均值大于 0.5; 而 GLM 模拟效果最差, AUC 仅为 0.669, TSS 仅为 0.330, 并且变化很大, 表明模型很不稳定, 而 CTA 表现稍好于 ANN。

#### 3.2 最优模型预测结果

综合以上比较, 最优模型结果为 GBM 和 MAXENT, 二者均能较为准确地预测高原鼠兔的分布。两个模型的模拟结果十分相似 (图 3, 4), 均表明高原鼠兔多分布于青海湖北部和西部, 以及布哈河上游汇流处。同时, MAXENT

模型在预测时更为保守和稳定, 模拟获得的适宜高原鼠兔分布面积相对较小, 而 GBM 模型模拟结果随机性更强, 运行不同次数得到的结果差异相对明显。

#### 3.3 影响高原鼠兔的环境因子

不同模型可能使用不同环境因子进行建模, BIOMOD 能够计算运行时各环境因子的重要性 (即对建模的贡献程度)。根据 10 次运行结果求得环境因子重要性的平均值 (表 2)。从中可以看出, 对于不同模型来说, 不同环境变量的影响程度差异明显, 因此, 在解释影响某物种的环境因素时, 必须选择精度较高的模型, 这样才能得到相对正确的影响因子。对于高原鼠兔的分布而言, 环境因子的评价应根据 GBM 和 MAXENT 的结果来进行, GBM 模拟结果显示, 距道路距离、距居民点距离、最暖月最高气温、NDVI 标准差较为重要, 而根据 MAXENT 模拟结果, 距居民点距离、最干季降水量、距道路距离、最冷季降水量较为重要。因此, 高原鼠兔的分布与温度、降水及人类活动均有密切关系。

#### 3.4 模型集合优化结果

BIOMOD 的优势在于可以将所有 70 个模拟结果根据条件进行筛选, 根据评价指标的数值大小计算权重, 最后生成最优的模拟结果, 模型设置中可选择是否进行上述权重计算。根据模型评价指标, 分别选取 AUC 大于 0.9 及 TSS 大于 0.6 作为优秀模型的筛选条件, 得到 2 个最优模型 (图 5, 6)。

根据集合优化后的模型结果重新计算评价指标, 经过筛选后得到的模型精度有了明显提高。根据 AUC 筛选后, 优化模型 TSS 值为 0.934, AUC 值为 0.988; 而根据 TSS 筛选后, 优化模型 TSS 值为 0.924, AUC 值为 0.981。TSS 值均超过 0.9, AUC 值超过 0.98, 达到了非常优秀的程度。模拟结果显示, 高原鼠兔在整个流域内呈现线状分布的特征, 其在青海湖北岸刚察县、海晏县北部、西岸石乃亥乡、布哈河上游汇水区及天峻县东北部分布概率较

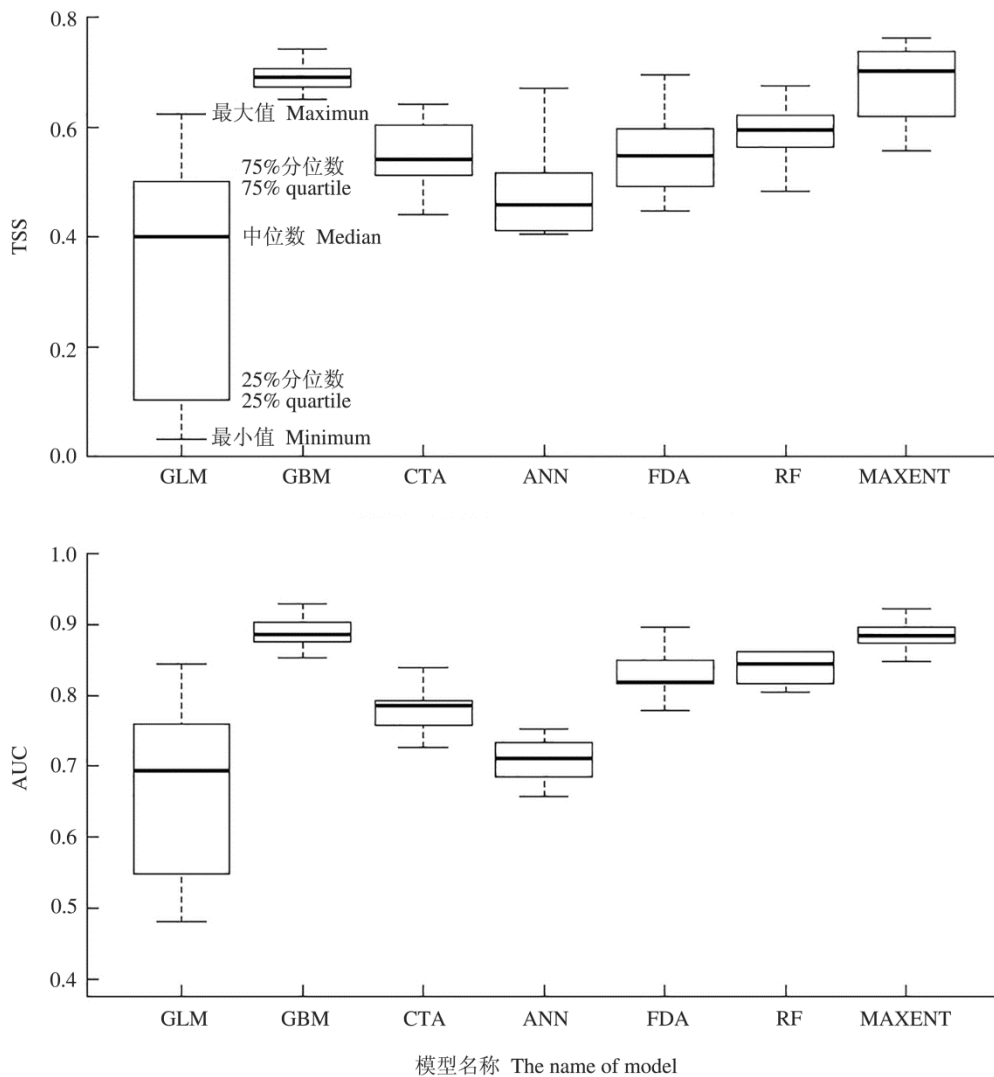


图 2 青海湖流域高原鼠兔分布模拟 7 种模型 TSS 和 AUC 比较

Fig. 2 Comparison of TSS and AUC of seven different models for the simulating of Plateau Pikas' distribution in Qinghai Lake Basin

高，而在青海湖东岸及南岸分布较少。

#### 4 讨论

本研究综合考虑地形、植被、土壤、气候和人为因素等环境条件模拟了青海湖流域的高原鼠兔分布，从结果来看，距居民点距离及距道路距离是影响高原鼠兔分布的重要因素，这有可能是因为居民点附近及道路两侧放牧强度高，导致靠近道路的草场退化程度较为严重，而高原鼠兔更易分布于轻度退化的草地（方毅

才等 1998），NDVI 标准差影响较大也一定程度反映了高原鼠兔对草地的选择偏好，这说明人类活动可能是影响高原鼠兔分布的重要因素。目前，人类灭鼠活动仍在进行，然而本文模拟结果显示，灭鼠因子的贡献度并不在前列，这可能是由于二值类型变量未能完全准确地反映灭鼠强度，而野外调查也发现即使灭鼠达到 95% 以上的区域仍可发现高原鼠兔频繁活动的踪迹，因此，人类活动对高原鼠兔分布的综合影响需要进一步研究。模拟结果还表明，最热



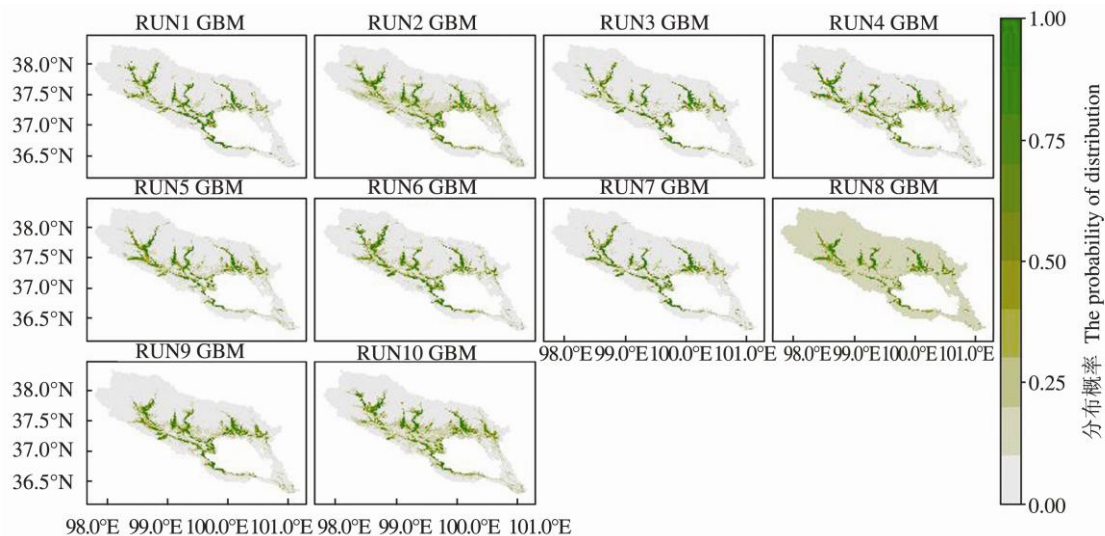


图 3 青海湖流域高原鼠兔分布 GBM 模型预测结果

Fig. 3 Results of GBM model for the distribution of Plateau Pika in Qinghai Lake Basin

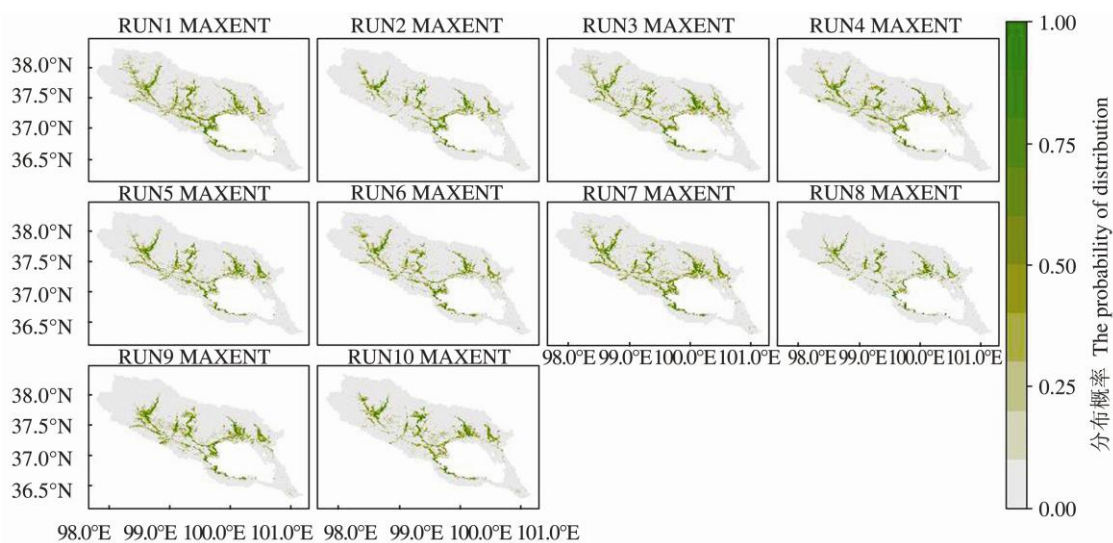


图 4 青海湖流域高原鼠兔分布 MAXENT 模型预测结果

Fig. 4 Results of MAXENT model for the distribution of Plateau Pika in Qinghai Lake Basin

月气温和干季降水量均与高原鼠兔分布有密切关系，高原鼠兔生活在高寒地带，对温度非常敏感 (Wang et al. 1999)，温度较高有利于其活动。由于目前缺乏针对高原鼠兔大尺度分布的野外调查，模拟结果并未能和其他野外调查结果进行详尽对比，但从模拟结果来看，高原鼠兔多分布于青海湖西岸的鸟岛景区、天峻县和刚察县，而东岸的沙地地区分布较少，这与前

人野外调查的鼠兔分布结果一致 (施银柱等 1994)。同时，植被、土壤等环境因子均能影响高原鼠兔分布，与前人研究较为一致 (马波等 2011, 张海娟等 2016)，说明 BIOMOD 不仅能够很好地模拟植物分布，而且能够模拟动物的分布状况。

以 GLM 和 CTA 为代表的传统回归模型和分类模型在模拟高原鼠兔的空间分布时效果不

表 2 青海湖流域高原鼠兔分布模拟 7 种模型的环境因子重要性

Table 2 The importance of environmental factors of seven different models simulating the distribution of Plateau Pika in Qinghai Lake Basin

	模型 Model						
	GLM	GBM	CTA	ANN	FDA	RF	MAXENT
海拔 Altitude	0.000	0.006	0.023	0.563	0.086	0.292	0.014
坡度 Slope	0.005	0.031	0.000	0.133	0.000	0.061	0.037
坡向 Aspect	0.001	0.004	0.066	0.247	0.000	0.082	0.010
年平均气温 Annual mean temperature	0.000	0.002	0.000	0.219	0.068	0.132	0.013
气温月较差 Mean of monthly (max temp - min temp)	0.028	0.000	0.000	0.128	0.000	0.053	0.069
气温的等温性 Isothermality	0.000	0.001	0.000	0.098	0.000	0.036	0.043
气温的季节性 Temperature seasonality	0.000	0.011	0.038	0.418	0.029	0.242	0.117
最暖月最高气温 Max temperature of warmest month	0.000	0.081	0.163	0.553	0.309	0.361	0.080
最冷月最低气温 Min temperature of coldest month	0.072	0.030	0.057	0.586	0.117	0.320	0.011
气温年较差 Temperature annual range	0.000	0.003	0.020	0.390	0.023	0.237	0.002
最湿季平均气温 Mean temperature of wettest quarter	0.000	0.001	0.000	0.325	0.000	0.262	0.001
最干季平均气温 Mean temperature of driest quarter	0.000	0.010	0.046	0.304	0.061	0.287	0.011
最暖季平均气温 Mean temperature of warmest quarter	0.000	0.003	0.033	0.336	0.264	0.252	0.018
最冷季平均气温 Mean temperature of coldest quarter	0.100	0.007	0.013	0.428	0.109	0.393	0.000
年降水量 Annual precipitation	0.000	0.034	0.029	0.685	0.252	0.243	0.012
最湿月降水量 Precipitation of wettest month	0.053	0.013	0.000	0.544	0.010	0.139	0.001
最干月降水量 Precipitation of driest month	0.008	0.001	0.000	0.130	0.000	0.033	0.000
降水的季节性 Precipitation seasonality	0.000	0.018	0.026	0.060	0.005	0.040	0.041
最湿季降水量 Precipitation of wettest quarter	0.000	0.004	0.000	0.411	0.062	0.101	0.012
最干季降水量 Precipitation of driest quarter	0.100	0.001	0.000	0.109	0.000	0.111	0.325
最暖季降水量 Precipitation of warmest quarter	0.091	0.000	0.000	0.386	0.021	0.113	0.131
最冷季降水量 Precipitation of coldest quarter	0.024	0.000	0.000	0.116	0.000	0.060	0.148
土壤容重 Soil bulk density	0.005	0.036	0.072	0.194	0.113	0.233	0.023
黏粒含量 Clay content	0.000	0.010	0.019	0.099	0.005	0.073	0.032
砂粒含量 Sand content	0.048	0.047	0.021	0.087	0.026	0.068	0.057
有机碳含量 Organic carbon content	0.000	0.029	0.028	0.095	0.010	0.110	0.058
粗颗粒体积比例 Proportion of coarse fragments volumetric	0.000	0.004	0.000	0.096	0.000	0.018	0.034
距道路距离 Distance to road	0.180	0.246	0.420	0.018	0.098	0.257	0.274
距居民点距离 Distance to settlement	0.000	0.165	0.269	0.008	0.126	0.304	0.360
距河流距离 Distance to river	0.031	0.042	0.049	0.001	0.019	0.141	0.066
NDVI 最大值 Maximum	0.000	0.002	0.026	0.006	0.000	0.085	0.028
NDVI 平均值 Mean	0.000	0.004	0.000	0.000	0.270	0.090	0.013
NDVI 最小值 Minimum	0.000	0.042	0.039	0.000	0.081	0.099	0.065
NDVI 标准差 Standard deviation	0.000	0.058	0.000	0.000	0.023	0.130	0.091
灭鼠因子 Deratization	0.399	0.012	0.008	0.010	0.007	0.036	0.008

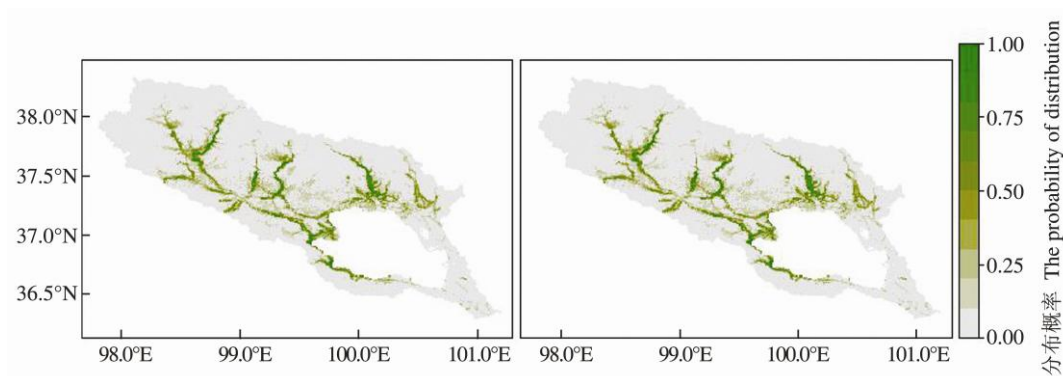


图 5 根据 AUC 筛选得到的最优模型（左为进行权重计算的，右为未经权重计算的）

Fig. 5 The optimal model filtrated by AUC (left for weighted calculation, right for unweighted calculation)

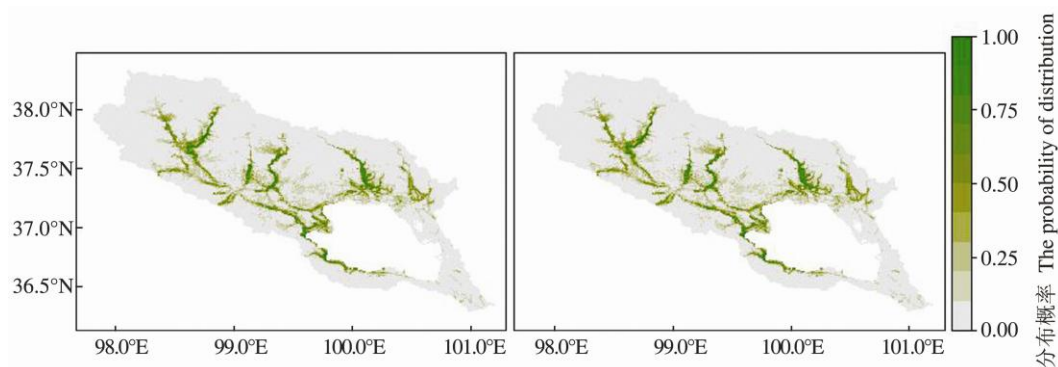


图 6 根据 TSS 筛选得到的最优模型（左为进行权重计算的，右为未经权重计算的）

Fig. 6 The optimal model filtrated by TSS (left for weighted calculation, right for unweighted calculation)

理想，而以 GBM、MAXENT 和 RF 为代表的机器学习模型的运用能带来更好的模拟效果。不同模型对不同环境因子的响应具有很大不同，本文中 GLM 模型表现出极大的不稳定性，这可能由于 GLM 模型仅使用线性项构建，导致其对环境因子的处理有很大差异。机器学习模型均表现出模拟的稳定性，说明其对环境因子的类型不敏感。不同模型模拟效果之间的差异，可能是由于各个模型理论假设不同，各个模型对识别基础生态位和现实生态位也有所偏重。基础生态位只能代表完全自然状态下物种可能的分布状况，而现实生态位则包括人为干扰（如道路、房屋的修建）、物种间的相互作用（竞争、捕食）的影响。目前，对于物种分布模型结果是否能反映现实生态位仍有争议

（Guisan et al. 2005），复杂模型如 GBM、MAXENT 和 RF 通过其复杂的运算技术能更好地代表物种现实生态位的情况。因此，BIOMOD 对于不同生态学研究目的来说，能更全面地反应物种可能分布和现实分布的状况。同时，BIOMOD 能同时研究多个物种的分布状况，大大方便了生态学家对某一大类物种（如鸟类）分布的对比分析。

## 参 考 文 献

- Aho K, Huntly N, Moen J, et al. 1998. Pikas (*Ochotona princeps*: Lagomorpha) as allogenic engineers in an alpine ecosystem. *Oecologia*, 114(3): 405–409.
- Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence kappa and the true skill

- statistic (TSS). *Journal of Applied Ecology*, 43(6): 1223–1232.
- Araújo M B, Luoto M. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16(6): 743–753.
- Beaumont L J, Hughes L, Poulsen M. 2005. Predicting species distributions: use of climatic parameters in BIOCLIM and its impact on predictions of species' current and future distributions. *Ecological Modelling*, 186(2): 251–270.
- Breiman L. 2001. Random forests. *Machine Learning*, 45(1): 5–32.
- Breiman L, Friedman J H, Stone C J, et al. 1984. Classification and Regression Trees. London: Chapman & Hall.
- Busby J. 1991. BIOCLIM - a bioclimate analysis and prediction system. *Plant Protection Quarterly*, 6(1): 8–9.
- Cutler D R, Edwards T C, Beard K H, et al. 2007. Random forests for classification in ecology. *Ecology*, 88(11): 2783–2792.
- Fielding A H, Bell J F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24(1): 38–49.
- Friedman J H. 1991. Multivariate adaptive regression splines. *The Annals of Statistics*, 19(1): 1–67.
- Friedman J H. 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5): 1189–1232.
- Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8(9): 993–1009.
- Guisan A, Weiss S B, Weiss A D. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143(1): 107–122.
- Hastie T J, Tibshirani R J. 1986. Generalized additive models. *Statistical Science*, 1(3): 297–310.
- Hastie T, Tibshirani R, Buja A. 1994. Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association*, 89(428): 1255–1270.
- Hengl T, de Jesus J M, MacMillan R A, et al. 2014. SoilGrids1km-global soil information based on automated mapping. *PLoS One*, 9(8): e105992.
- Hijmans R J, Cameron S E, Parra J L, et al. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25(15): 1965–1978.
- Leathwick J R, Elith J, Hastie T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. *Ecological Modelling*, 199(2): 188–196.
- Nelder J A, Wedderburn R W M. 1972. Generalized Linear Models. *Journal of the Royal Statistical Society Series A: General*, 135(3): 370–384.
- Pearman P B, Randin C F, Broennimann O, et al. 2008. Prediction of plant species distributions across six millennia. *Ecology Letters*, 11(4): 357–369.
- Pearson R G, Dawson T P, Liu C. 2004. Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, 27(3): 285–298.
- Peterson A T, Papeş M, Eaton M. 2007. Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30(4): 550–560.
- Phillips S J, Anderson R P, Schapire R E. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3): 231–259.
- Reiss H, Cunze S, König K, et al. 2011. Species distribution modelling of marine benthos: a North Sea case study. *Marine Ecology Progress Series*, 442(8): 71–86.
- Ridgeway G. 1999. The state of boosting. *Computing Science and Statistics*, 31: 172–181.
- Ripley B D. 2007. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press, 1–15.
- Rosenzweig M L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge University Press, 37–39.
- Smith A T, Foggini J M. 1999. The plateau pika (*Ochotona curzoniae*) is a keystone species for biodiversity on the Tibetan plateau. *Animal Conservation*, 2(4): 235–240.
- Smith A T, Weidong L, Hik D S. 2004. Pikas as harbingers of global warming. *Species*, 41(4): 5.
- Smith A T, Xie Y. 2008. *A Guide to the Mammals of China*. Princeton: Princeton University Press, 192–193.
- Thuiller W. 2003. BIOMOD - optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology*, 9(10): 1353–1362.
- Thuiller W, Araújo M B, Lavorel S. 2003. Generalized models vs.

- classification tree analysis: predicting spatial distributions of plant species at different scales. *Journal of Vegetation Science*, 14(5): 669–680.
- Thuiller W, Lafourcade B, Engler R, et al. 2009. BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, 32(3): 369–373.
- Wang D, Sun R, Wang Z, et al. 1999. Effects of temperature and photoperiod on thermogenesis in plateau pikas (*Ochotona curzoniae*) and root voles (*Microtus oeconomus*). *Journal of Comparative Physiology B*, 169(1): 77–83.
- 方毅才, 史青茂. 1998. 甘肃草地高原鼠兔危害与防治探讨. *甘肃农业*, 6(14): 41–42.
- 韩天虎, 花立民, 许国成. 2008. 高原鼠兔危害级别划分. *草业学报*, 17(5): 130–137.
- 侯佩玲, 沈小京. 2004. 浅谈青海湖流域草地资源的可持续利用. *青海环境*, 14(3): 119–122.
- 贾婷婷, 毛亮, 郭正刚. 2014. 高原鼠兔有效洞穴密度对青藏高原高寒草甸群落植物生态位的影响. *生态学报*, 34(4): 869–877.
- 李国庆, 刘长成, 刘玉国, 等. 2013. 物种分布模型理论研究进展. *生态学报*, 33(16): 4827–4835.
- 李建川, 杨乐. 2010. 高原鼠兔对藏北草原的危害及其主要天敌. *西藏科技*, 34(1): 65–67.
- 马波, 王小明, 刘晓庆, 等. 2011. 高原鼠兔洞穴数量与其栖息地植被分布格局的 GIS 分析. *生物多样性*, 19(1): 71–78.
- 齐新章, 王晓君, 朱世海, 等. 2008. 高原鼯鼠和高原鼠兔心脏对低氧环境的适应. *生理学报*, 60(3): 348–354.
- 施银柱. 1983. 草场植被影响高原鼠兔密度的探讨. *兽类学报*, 3(2): 181–187.
- 施银柱, 张道川, 边疆晖. 1994. 海北州、天峻县的高原鼠兔和高原鼯鼠的数量及其分布. *中国草地*, 15(6): 51–52.
- 石红霄, 于健龙. 2010. 高原鼠兔洞口密度对高寒嵩草草甸植被及土壤水分的影响. *中国草地学报*, 32(4): 109–112.
- 王滔, 王小明, 王正寰. 2004. 高原鼠兔生境选择的初步研究. *四川大学学报: 自然科学版*, 41(4): 1041–1045.
- 殷宝法, 王金龙, 魏万红, 等. 2004. 高寒草甸生态系统中高原鼠兔的繁殖特征. *兽类学报*, 24(3): 222–228.
- 张海娟, 谢久祥, 李希来. 2016. 青海河南县高原鼠兔的分布及其与土壤特性的相关性分析. *青海大学学报*, 34(1): 14–21.
- 赵廷贵. 2005. 高原鼠兔在冬春季对燕麦等三种饵料适口性及采食量的对比试验. *四川草原*, 25(7): 16–18.
- 赵廷贵, 潘桂兰, 骆海强. 2002. 高原鼠兔对燕麦等三种饵料适口性及采食量的对比试验. *青海草业*, 11(3): 8–9.
- 周雪荣, 郭正刚, 郭兴华. 2010. 高原鼠兔和高原鼯鼠在高寒草甸中的作用. *草业科学*, 27(5): 38–44.