

285种鸟类在中国生境适宜性与丰富度数据集研发

陆慧缘^{1,2}, 张锐^{1*}, 姜琳琳^{2,3}

- 中国科学院西北生态环境资源研究院冰冻圈科学与冻土工程全国重点实验室, 兰州 730000;
- 苏州科技大学地理科学与测绘工程学院, 苏州 215009;
- 安徽师范大学地理与旅游学院, 芜湖 241002

摘要: 中国是世界鸟类种群数量最丰富的国家之一。作者基于 eBird 与 GBIF 平台的鸟类观测数据、数字高程模型 (DEM)、年降雨量、年均温与蒸散发数据, 利用 Biomod2 平台的 GLM、MaxEnt、RF 及综合模型, 研发了 285 种鸟类在中国生境适宜性与丰富度数据集, 以为生物多样性研究、环境评估等研究提供数据支持。该数据集汇集的内容包括苍鹰、八哥、小云雀等共计 285 种鸟类在 2000、2005、2010、2015、2020 年的生境适宜性、物种丰富度数据。模型验证结果表明数据集具有较高的精度, 测试集 AUC 均值为 0.991, TSS 均值为 0.923。数据集空间分辨率为 0.05°, 时间分辨率为 5 年。数据集存储为 .img、.tif 格式, 由 1,430 个数据文件组成, 数据量为 5.66 GB (压缩为 1 个文件, 37 MB)。

关键词: 生物多样性; 物种分布模型; 鸟类; 遥感监测

DOI: <https://doi.org/10.3974/geodp.2025.02.06>

CSTR: <https://cstr.science.org.cn/CSTR:20146.14.2025.02.06>

数据可用性声明:

本文关联实体数据集已在《全球变化数据仓储电子杂志 (中英文)》出版, 可获取:

<https://doi.org/10.3974/geodb.2025.01.08.V1> 或 <https://cstr.science.org.cn/CSTR:20146.11.2025.01.08.V1>。

1 前言

鸟类作为世界上分布范围最广的物种之一, 不仅是生态系统中不可缺少的一员^[1], 还可以缓解人类心理焦虑以促进身心健康^[2]。然而世界自然保护联盟 (International Union for Conservation of Nature, IUCN) 濒危物种红色名录显示, 在全球有上百种鸟类处于灭绝或野外灭绝状态, 超过 15% 的鸟种处于极危 (CR)、濒危 (EN) 与易危 (VU) 等受威胁状态^[3]。城市化进程导致大量林地、湿地等生物多样性丰富的土地类型转化为建设用地, 使得鸟类生存空间被压缩^[4,5]。与此同时, 气候变化对鸟类多样性的影响正进一步加剧。自 19 世纪工业革命以来, 城市化进程不断推进, 全球气温稳步上升, 部分鸟类由于无法适应

收稿日期: 2025-02-07; 修订日期: 2025-05-22; 出版日期: 2025-6-25

基金项目: 江苏省教育厅 (202210332083Y)

*通讯作者: 张锐, 中国科学院西北生态环境资源研究院, zhangrui@radi.ac.cn

数据引用方式: [1] 陆慧缘, 张锐, 姜琳琳. 285 种鸟类在中国生境适宜性与丰富度数据集研发[J]. 全球变化数据学报, 2025, 9(2): 189–195. <https://doi.org/10.3974/geodb.2025.02.06>. <https://cstr.science.org.cn/CSTR:20146.14.2025.02.06>.

[2] 陆慧缘, 张锐, 姜琳琳. 285 种鸟类在中国生境适宜性与丰富度数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.01.08.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2025.01.08.V1>.

气候变化, 其种群数量已呈现出下降趋势^[6,7]。鸟类不得已自发向气温较低的两极与高海拔地区迁移以应对气候变化^[8,9]。

物种分布模型 (Species Distribution Models, SDMs) 也称生态位模型, 是一种利用物种观测数据和环境数据模拟物种潜在分布的方法^[10]。SDMs 模型能够在仅依靠环境数据与少量物种观测样本的情况下, 模拟物种生存适宜性, 已经成为鸟类保护不可或缺的方法之一^[11,12]。作为主流物种分布建模平台的 Biomod2 不仅集成了多种主流的物种分布模型, 如广义线性模型 (Generalized Linear Model, GLM)、最大熵模型 (Maximum Entropy Model, MaxEnt)、随机森林 (Random Forest, RF) 模型等, 还提供了集成模型 (Ensemble Model, EM) 用于整合不同模型以提高预测精确度^[13]。

对于 SDMs 来说, 物种观测数据质量越高, 模拟结果往往越精确。eBird 提供的鸟类观测数据因为观测记录数量多、观测记录空间范围广、观测记录时间序列长, 已经成为公民科学平台中应用最广的鸟类观测平台^[14,15]。全球生物多样性信息机构 (Global Biodiversity Information Facility, GBIF) 网站作为全球首屈一指的物种观测记录平台, 收集了来自 eBird 网站的物种观测信息。为此本研究使用来自 eBird 与全球生物多样性信息机构的鸟类观测数据, 借助 Biomod2 构建了 3 种单个物种分布模型与集成模型, 用于模拟鸟类分布。

研究基于 eBird 与 GBIF 平台所提供的鸟类观测数据, 构建了一个包括 285 种鸟类的大尺度分布数据集。该数据集的研发旨在: (1) 为研究鸟类的空间分布格局提供依据; (2) 为研究全球气候变化背景下鸟类栖息地的变化提供参考; (3) 为生态保护政策的制定与生态红线保护区域的划定提供数据。

2 数据集元数据简介

《285 种鸟类在中国生境适宜性与丰富度数据集》^[16]的地理区域、数据年代、时间分辨率、空间分辨率、数据集组成、数据出版与共享服务平台、数据共享政策等信息见表 1。

3 数据研发方法

作者首先对鸟类观测数据进行时间序列处理, 根据数据观测的时间将鸟类观测数据拆分到不同年份。其次, 将不同年份的鸟类观测数据, 根据鸟种拆分。最后, 对拆分后的数据进行空间分析, 去除空间距离相隔较近的鸟类观测数据, 保证鸟类观测数据在空间上相对独立。此外, 为提高模拟的可靠性与准确性, 本研究仅选择在研究时段内, 每个年份的可用观测记录均大于 5 条的鸟种。最终选择了 285 种鸟类。

作者使用 R 语言中的 Biomod2 建模平台分析物种分布情况, 该平台能够调用多种单个物种分布模型, 并基于调用的单个物种分布模型构建综合模型。本研究中调用了广义线性模型 (GLM)、最大熵模型 (MaxEnt)、随机森林模型 (RF) 三种单个物种分布模型, 并基于 3 种模型构建综合模型 (EM) (图 1)。广义线性模型 (GLM) 可以处理不同类型的数据, 结果具有较好的解释性。最大熵模型 (MaxEnt) 基于信息熵的概念来预测物种分布, 是最受欢迎的物种分布模型。随机森林模型 (RF) 是一种机器学习方法, 通过构建多个决策树并将预测结果结合起来, 以提高模型的准确性和鲁棒性, 对于特征之间的复杂关系具

表 1 《285 种鸟类在中国生境适宜性与丰富度数据集》元数据简表

条 目	描 述
数据集名称	285 种鸟类在中国生境适宜性与丰富度数据集
数据集短名	HabitatSuitability&Richness2000-2020
作者信息	陆慧缘, 中国科学院西北生态环境资源研究院、苏州科技大学地理科学与测绘工程学院, 19200214139@post.usts.edu.cn 张锐, 中国科学院西北生态环境资源研究院, zhangrui@radi.ac.cn 姜琳琳, 安徽师范大学, jiangll@ahnu.edu.cn
地理区域	中国 (未包括面积较小的岛、礁)
数据年代	2000–2020 年
时间分辨率	5 年
空间分辨率	0.05°
数据格式	.img、.tif
数据量	37 MB (压缩后)
数据集组成	285 种鸟类在 2000、2005、2010、2015、2020 年的生境适宜性、物种丰富度数据
基金项目	江苏省教育厅 (202210332083Y)
数据计算环境	数据基于 R 语言与 ArcGIS 软件计算
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101, 中国科学院地理科学与资源研究所
数据共享政策	(1)“数据”以最便利的方式通过互联网系统免费向全社会开放, 用户免费浏览、免费下载; (2) 最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源; (3) 增值服务用户或以任何形式散发和传播 (包括通过计算机服务器)“数据”的用户需要与《全球变化数据学报 (中英文)》编辑部签署书面协议, 获得许可; (4) 摘取“数据”中的部分记录创作新数据的作者需要遵循 10% 引用原则, 即从本数据集中摘取的数据记录少于新数据集总记录量的 10%, 同时需要对摘取的数据记录标注数据来源 ^[17]
数据和论文检索系统	DOI, CSTR, Crossref, DCI, CSCD, CNKI, SciEngine, WDS, GEOSS, PubScholar, CKRSC

有较强的捕捉能力。在建模过程中, 抽取与数量相同的点位创建伪随机点, 随机抽取 80% 的数据作为训练集, 20% 的数据作为测试集, 每次建模中重复运行 5 次。共创建 3 次伪随机点并运行模型。

本研究选取了 4 个可能影响物种分布的环境变量, 包括高程、温度、降水和蒸发强度。高程数据来源于 WorldClim 数据集 2.1, 温度、降水与蒸发强度数据来源于国家青藏高原数据中心 (表 2), 数据集基于气象站点数据进行插值与降尺度得出^[18]。

4 数据结果与验证

4.1 数据集组成

《285 种鸟类在中国生境适宜性与丰富度数据集》, 包括多个生境适宜性子文件夹, 子文件夹名称为年份, 每个子文件夹中存储 285 个 .img 文件, 文件命名方式为“XXXX.img”, 其中 XXXX 为物种名称。此外, 每个年份包含 1 个 .tif 格式的物种丰富度数据, 物种丰富度由 285 种鸟类的适宜生境叠加而来。img 与 .tif 文件可以在 GIS 软件与相应的编程环境中处理。数据的时间分辨率为 5 年一期, 一共 5 期, 每期由 286 个文件组成。

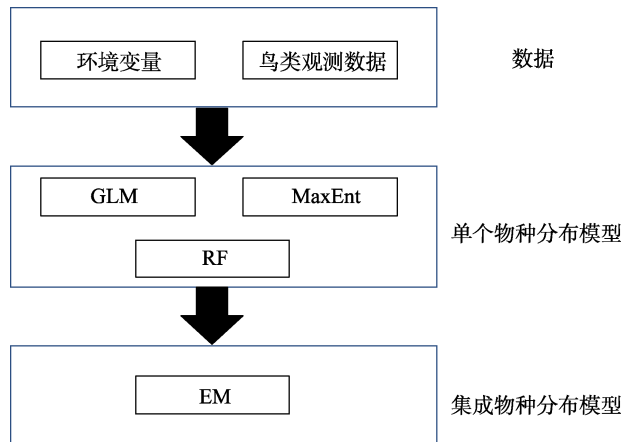


图1 技术路线

表2 环境变量数据来源表

变量名称	数据来源
高程	WorldClim 数据集 ^[19]
平均温度	中国 1-km 分辨率逐月平均气温数据集 (1901–2023) ^[20]
平均降水	中国 1-km 分辨率逐月降水量数据集 (1901–2023) ^[21]
蒸发强度	中国 1-km 逐月潜在蒸散发数据集 (1901–2023) ^[22]

4.2 数据结果

285 种鸟类的物种丰富度呈现出显著的空间异质性。整体表现为由东南沿海向西北内陆递减的空间分布格局 (图 2)。这种分布特征主要与区域生态环境差异密切相关。沿海地区温暖而湿润的季风型气候为植被生长提供了有力条件,为鸟类生长繁衍提供了更多食物。沿海地区还拥有大量的滩涂、湿地、红树林等丰富的生态系统类型,为更多鸟类提供了栖息地。

青藏高原地区由于其独特的地理位置,区域内部鸟种丰富度差异大,青藏高原南面拥有更优越的水热条件、更丰富的植被覆盖并提供了丰富的食物,更有利于鸟类生存繁衍。青藏高原北部受到地形的阻隔,水热条件不如南方地区,鸟类丰富度低。

新疆、青海等存在大量荒漠与半干旱草原,植被覆盖率低,且气候干燥,缺乏水资源,植被、气候的双重影响导致鸟类丰富度较低。东北区域尤其是黑龙江省气温低且低气温持续时间长,这种气候特征不仅直接限制了鸟类的生存,还通过影响植被生长和昆虫活动间接减少了鸟类食物资源的可获得性,导致鸟类丰富度低。

4.3 数据验证

曲线下面积 (AUC) 是评估物种分布模型性能的重要方法,其范围为 0 到 1,用于评估模型的预测性能,数值越大,表明模型的预测能力越强^[23]。The True Skill Statistic (TSS = 敏感性+特异性-1) 经常被用于衡量物种分布模型性能,其范围为-1 到 1,值越高模型性能越好^[24]。由于仅使用 AUC 容易出现过拟合现象,错误的评估模型精度。为此,在本研究

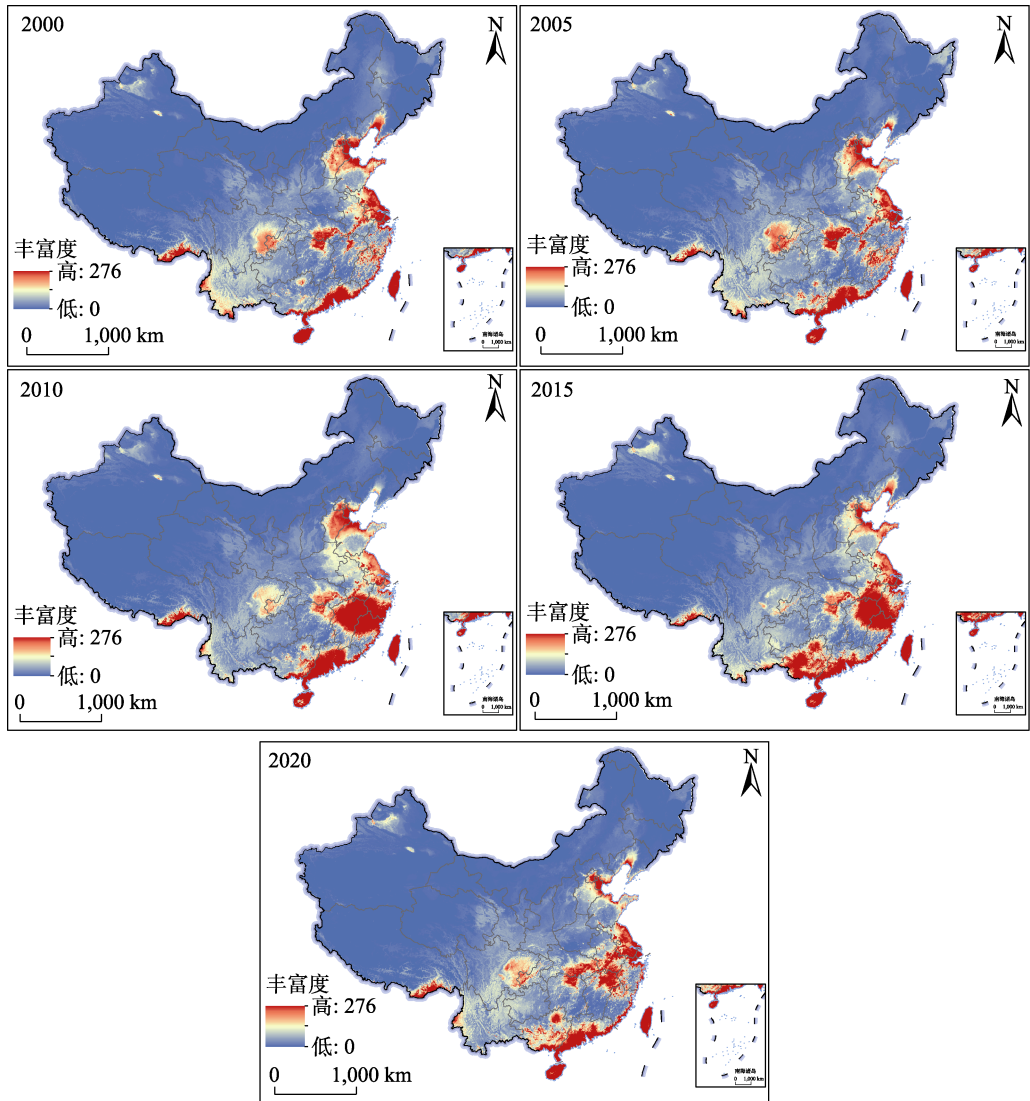


图 2 各年份全国 285 种鸟类种群多样性丰富度分布图

(该图基于国家测绘地理信息局标准地图服务网站下载的审图号为 GS(2019)1823 号标准地图制作, 底图无修改)

中使用 AUC 和 TSS 两项指标评估物种分布模型的预测性能^[25,26]。

如图 3 所示, 广义线性模型 (GLM) 中, 测试集 AUC 均值为 0.953, 平均 TSS 值为 0.849。最大熵模型 (MaxEnt) 模型中, 测试集 AUC 均值为 0.930, 平均 TSS 值为 0.822。随机森林模型 (RF) 中, 测试集 AUC 均值为 0.982, 平均 TSS 值为 0.920。在 EM 模型中, 测试集 AUC 均值为 0.991, 平均 TSS 值为 0.923。4 种模型测试集平均 AUC 均超过 0.9, 平均 TSS 值均超过 0.8, 所有的物种分布模型均表现出优越的性能。在单个物种模型中, RF 模型表现出高于 MaxEnt 模型与 GLM 模型的精度, 集成模型精度高于单个物种分布模型。为此, 本研究使用 EM 模型作为最终结果。

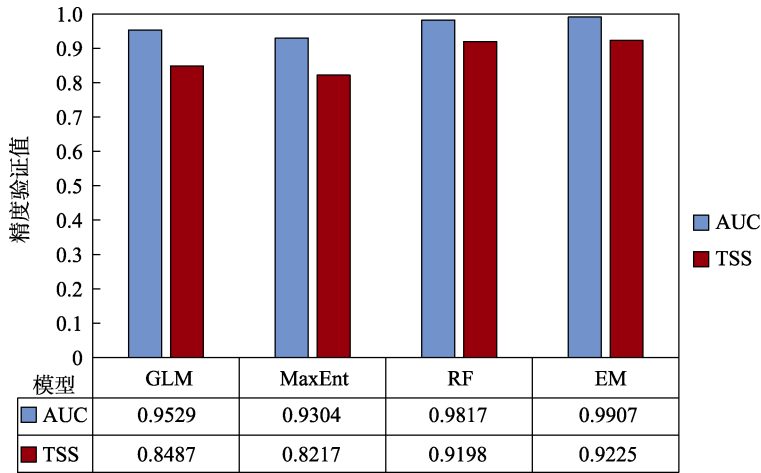


图3 物种分布模型精度验证结果统计图

5 讨论和总结

城市化进程的快速推进导致生态环境质量下降、全球气候变化、生物栖息地被破坏等一系列问题,进而导致鸟类生存受到威胁^[7,27]。中国鸟类生境适宜性数据集利用 eBird 与 GBIF 平台的鸟类观测数据,采用 R 语言与 ArcGIS 软件,根据气候数据与 Biomod2 编程平台,采用 3 种常见的物种分布模型与集成模型,计算了全国 2000–2020 年间 285 种鸟类的适宜生境数据。通过本数据集可以发现,285 种鸟类在中国的生境适宜性与物种丰富度呈现出显著的空间异质性,整体表现为由东南沿海向西北内陆递减的空间分布格局。

虽然中国鸟类生境适宜性数据集精度满足物种分布建模的精度要求,但在使用时依旧要注意以下问题:由于公民科学平台的物种观测数据的空间分布并不均衡、观测地点不固定,物种分布建模的结果难免受到影响。此外,虽然中国鸟类生境适宜性数据集使用集成模型以提高模型精度,但物种分布模型是模拟物种的潜在分布范围,而不是实际分布范围,与实际情景难免存在出入。

作者分工: 张锐对数据集的开发做了总体设计;陆慧缘、姜琳琳采集和处理了鸟类分布数据;陆慧缘设计了模型和算法并进行了数据验证;陆慧缘、张锐撰写了数据论文等。

利益冲突声明: 本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

- [1] Ding, Z. H., Cao, J. J., Wang, Y. The construction and optimization of habitat networks for urban-natural symbiosis: a case study of the main urban area of Nanjing [J]. *Forests*, 2023, 14(1): 18.
- [2] Buxton, R. T., Pearson, A. L., Lin, H. Y., *et al.* Exploring the relationship between bird diversity and anxiety and mood disorder hospitalisation rates [J]. *Geo-Geography and Environment*, 2023, 10(2): 8.
- [3] Lees, A. C., Haskell, L., Allinson, T., *et al.* State of the world's birds [J]. *Annual Review of Environment and Resources*, 2022, 47: 231–260.
- [4] Xu, C. L., Yu, Q., Wang, F., *et al.* Identifying and optimizing ecological spatial patterns based on the bird distribution in the Yellow River Basin, China [J]. *Journal of Environmental Management*, 2023, 348: 13.
- [5] Lu, H. Y., Shang, Z. Y., Ruan, Y. L., *et al.* Study on urban expansion and population density changes based

- on the inverse S-shaped function [J]. *Sustainability*, 2023, 15(13): 19.
- [6] Moller, A. P., Rubolini, D., Lehikoinen, E. Populations of migratory bird species that did not show a phenological response to climate change are declining [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2008, 105(42): 16195–16200.
- [7] Liu, Z. X., Zhang, W. W., Lu, H. Y., et al. Exploring evolution characteristics of eco-environment quality in the Yangtze River Basin based on remote sensing ecological index [J]. *Heliyon*, 2023, 9(12): 14.
- [8] Zhu, B. R., Verhoeven, M. A., Velasco, N., et al. Current breeding distributions and predicted range shifts under climate change in two subspecies of Black-tailed Godwits in Asia [J]. *Global Change Biology*, 2022, 28(18): 5416–5426.
- [9] Virkkala, R., Rajasärkkä, A., Heikkinen, R. K., et al. Birds in boreal protected areas shift northwards in the warming climate but show different rates of population decline [J]. *Biological Conservation*, 2018, 226: 271–279.
- [10] Rousseau, J. S., Betts, M. G. Factors influencing transferability in species distribution models [J]. *Ecography*, 2022, 2022(7): 13. DOI: 10.1111/ecog.06060.
- [11] Gaul, W., Sadykova, D., White, H. J., et al. Data quantity is more important than its spatial bias for predictive species distribution modelling [J]. *PeerJ*, 2020, 8: 27.
- [12] Yu, H., Cooper, A. R., Infante, D. M. Improving species distribution model predictive accuracy using species abundance: application with boosted regression trees [J]. *Ecological Modelling*, 2020, 432: 11.
- [13] Thuiller, W., Lafourcade, B., Engler, R., et al. BIOMOD—a platform for ensemble forecasting of species distributions [J]. *Ecography*, 2009, 32(3): 369–373.
- [14] Neate-Clegg, M. H. C., Horns, J. J., Adler, F. R., et al. Monitoring the world’s bird populations with community science data [J]. *Biological Conservation*, 2020, 248: 7.
- [15] Tejada, I., Medrano, F. eBird as a tool to improve the knowledge of Chilean birds [J]. *Revista Chilena de Ornitología*, 2018, 24(2): 85–94.
- [16] 陆慧缘, 张锐, 姜琳琳. 285 种鸟类在中国生境适宜性与丰富度数据集[J/DB/OL]. 全球变化数据仓储电子杂志, 2025. <https://doi.org/10.3974/geodb.2025.01.08.V1>. <https://cstr.science.org.cn/CSTR:20146.11.2025.01.08.V1>.
- [17] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. <https://doi.org/10.3974/dp.policy.2014.05> (2017 年更新).
- [18] Peng, S. Z., Ding, Y. X., Wen, Z. M., et al. Spatiotemporal change and trend analysis of potential evapotranspiration over the Loess Plateau of China during 2011–2100 [J]. *Agricultural and Forest Meteorology*, 2017, 233: 183–194.
- [19] Fick, S. E., Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas [J]. *International Journal of Climatology*, 2017, 37(12): 4302–4315.
- [20] 彭守璋. 中国 1 km 分辨率逐月平均气温数据集(1901–2023) [DB/OL]. 国家青藏高原数据中心, 2024. <https://doi.org/10.11888/Meteoro.tpd.270961>.
- [21] 彭守璋. 中国 1 km 分辨率逐月降水量数据集(1901–2023) [DB/OL]. 国家青藏高原数据中心, 2024. <https://doi.org/10.5281/zenodo.3114194>.
- [22] 彭守璋. 中国 1 km 逐月潜在蒸散发数据集(1901–2023) [DB/OL]. 国家青藏高原数据中心, 2024. <https://doi.org/10.11866/db.loess.2021.001>.
- [23] Phillips, S. J., Anderson, R. P., Schapire, R. E. Maximum entropy modeling of species geographic distributions [J]. *Ecological Modelling*, 2006, 190(3–4): 231–259.
- [24] Xu, L., Fan, Y., Zheng, J. H., et al. Impacts of climate change and human activity on the potential distribution of *Aconitum leucostomum* in China [J]. *Science of the Total Environment*, 2024, 912: 12.
- [25] Adeyemo, S. M., Granger, J. J. Habitat suitability model and range shift analysis for American Chestnut (*Castanea dentata*) in the United States [J]. *Trees Forests and People*, 2023, 11: 13.
- [26] Lobo, J. M., Jiménez-Valverde, A., Real, R. AUC: a misleading measure of the performance of predictive distribution models [J]. *Global Ecology and Biogeography*, 2008, 17(2): 145–151.
- [27] Zhang, W. W., Liu, Z. X., Qin, K., et al. Long-term dynamic monitoring and driving force analysis of eco-environmental quality in China [J]. *Remote Sensing*, 2024, 16(6): 22.