

# 基于数据驱动模型的辣木全球环境适宜性分布评估

范沛薇<sup>1</sup>, 郝蒙蒙<sup>1,2</sup>, 丁方宇<sup>1,2\*</sup>, 王倩<sup>1,2</sup>, 董东林<sup>3</sup>, 马志民<sup>4</sup>, 钱军<sup>4</sup>, 李新华<sup>4</sup>, 陈静<sup>4</sup>, 万文宏<sup>4</sup>, 赵艳<sup>4</sup>, 张小波<sup>5\*</sup>

1. 中国科学院地理科学与资源研究所资源利用与环境修复重点实验室, 北京 100101
2. 中国科学院大学资源与环境学院, 北京 100049
3. 中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083
4. 江西省信息中心, 南昌 330036
5. 中国中医科学院中药资源中心, 道地药材国家重点实验室培育基地, 北京 100700

**摘要** 辣木作为新中药资源, 被广泛应用于植物医药。收集了全球 960 条辣木分布记录, 结合气候、土壤、地形和太阳辐射空间协变量, 采用空间信息技术与数据驱动模型, 评估了辣木全球环境适宜性空间分布状况。结果表明, 全球适宜种植辣木的土地资源总量约为  $3.61 \times 10^9$   $\text{hm}^2$ , 主要分布在北美洲南部、南美洲东部和北部、非洲中部、亚洲的东南部及大洋洲北部地区。

**关键词** 辣木; 空间协变量; 空间信息技术; 数据驱动模型; 环境适宜性

辣木为热带、亚热带乔木, 原产于印度、巴基斯坦、孟加拉和阿富汗<sup>[1]</sup>, 现广泛存在于非洲、东南亚、太平洋和南美洲等地区。辣木是一种多功能植物, 其叶、果、根、种子和树皮等部位有着不同的利用价

值<sup>[2-3]</sup>。例如, 嫩叶和果荚可以食用且富含较多营养物质<sup>[4]</sup>, 而根、树皮和种子通常被用作药物, 具有抗癌、抗菌、降血压、抗氧化等功效<sup>[5-6]</sup>。

由于辣木的多功能性, 其经济价值和商业价值

收稿日期: 2020-12-24; 修回日期: 2022-03-12

基金项目: 云南省科技计划项目(2017ZF004); 江西省“生态云”大数据平台总体框架和指标体系规划设计采购项目(GXHT-19YYYB-0903); 国家重点研发计划项目(2017YFC1700701, 2017YFC1701603); 中国中医科学院中药资源中心自主选题科研项目(ZZXT201804)

作者简介: 范沛薇, 硕士研究生, 研究方向为资源环境遥感应用, 电子信箱: peiwei\_fan@163.com; 丁方宇(通信作者), 特别研究助理, 研究方向为资源环境遥感应用, 电子信箱: dingfy.17b@igsrr.ac.cn; 张小波(共同通信作者), 副研究员, 研究方向为中药资源, 电子信箱: jack110007@163.com

引用格式: 范沛薇, 郝蒙蒙, 丁方宇, 等. 基于数据驱动模型的辣木全球环境适宜性分布评估[J]. 科技导报, 2022, 40(7): 94-99; doi:10.3981/j.issn.1000-7857.2022.07.010

逐步受到全球重视。许多国家争相投入大量资金研发和推广相关产品,如辣木籽食用油<sup>[7]</sup>、辣木叶抗菌药物<sup>[8]</sup>、降血糖辅助药物<sup>[9]</sup>及抗肿瘤药物<sup>[10]</sup>等。在医学领域,辣木的相关产品已被证实可用于100多种疾病的治疗<sup>[5,11]</sup>。各国开始扩大辣木种植规模,完善相关产业链建设<sup>[12]</sup>。以往有关辣木的研究主要集中在化学成分的提取<sup>[13]</sup>、食品科学技术<sup>[14]</sup>及其药理效果<sup>[5]</sup>等方面,而有关辣木的全球资源分布及适宜性方面则鲜有研究。近年来,中国在云南、广东、广西、福建、海南等地区开展了辣木的引种培育工作<sup>[15]</sup>。为了减少盲目种植造成的经济损失,推动辣木相关产业链科学规划,本研究以已知的辣木空间分布数据为基础,结合气象、太阳辐射、土壤和地形4类空间协变量,使用数据驱动模型和空间信息技术,在全球尺度上绘制辣木环境适宜性分布图,估算了各地区适宜种植辣木的土地资源量。

## 1 数据及来源

### 1.1 辣木分布数据

本研究使用的辣木分布数据共计960条。其中,942条分布记录来源于全球生物多样性信息数据库(<http://www.gbif.org/>)。其余18条分布记录,通过文献和实地调查获取。

### 1.2 空间协变量

以往研究表明,辣木喜温、耐旱且抗逆性强<sup>[6]</sup>,并多种植于热带以及亚热带地区。综合以上条件,本研究选择气象、太阳辐射、土壤和地形4个类型的环境要素构建空间协变量。第1类为气象要素,包括年累积降水量、年平均气温和年平均水气压。第2类为太阳辐射系数。第3类为土壤要素,包括土壤含水量、土壤类别以及土壤深度。第4类为地形要素,包括高程和坡度。具体数据名称和数据来源详见表1。

表1 空间协变量信息

类别	数据名称	数据来源
气象	年累积降水量/mm	WorldClim 数据库, 2.0 版: <a href="http://www.worldclim.org/version2">http://www.worldclim.org/version2</a>
	月平均温度/°C	
	月平均水汽压/kPa	
太阳辐射	平均太阳辐射/(kJ·m <sup>-2</sup> ·d <sup>-1</sup> )	WorldClim 数据库, 2.0 版: <a href="http://www.worldclim.org/version2">http://www.worldclim.org/version2</a>
	土壤含水量/%	
土壤	土壤类别	世界土壤数据库: <a href="http://www.isric.org/">http://www.isric.org/</a>
	土层深度/mm	
地形	高程/m	CGIAR 空间信息联盟: <a href="http://srtm.csi.cgiar.org/">http://srtm.csi.cgiar.org/</a>
	坡度/(°)	

### 1.3 背景数据

综合文献中已有辣木种植环境数据,以及生态作物数据库网站(<http://ecocrop.fao.org/>)获取的相关数据作为背景点的依据,以平均温度<17.5°C或年累积降雨量<250 mm作为筛选不适宜种植辣木的背景数据依据<sup>[17]</sup>。

## 2 模型及方法

本研究采用数据驱动模型——增强回归树模

型(boosted regression trees, BRT),该模型在传统的分类回归树算法基础之上,通过不断地随机选择和自学习方法产生多重回归树,以提高模型的稳定性和预测精度<sup>[18]</sup>。该模型包含基于树模型的优势,能处理不同类型的预测变量,并允许缺失值数据<sup>[19]</sup>。该模型通过拟合多个树的过程,能较好地弥补单一树模型预测缺点,比大多传统模型方法有更强的预测能力。该模型能较好地模拟不同影响因子对因变量的影响结果,在R3.3.3的统计编程设计环境下,结合扩展包(dismo、caret、gbm)构建增强回归树

模型,并评估模拟精度。

模型将收集到的辣木现有种植记录点根据地理坐标信息转换为 5 km×5 km 分辨率的网格单元。在本研究中得到了反映适宜环境条件的 479 个网格单元,同时随机选取了相同数量反映不适宜环境条件的背景点。为了减少背景点对模型模拟的影响,随机选取了 100 次背景点。在每一次实验中,共构建 958 个样本,并将其分为训练样本和验证样本。其中,训练样本和验证样本各占总样本的 50%。对 100 个增强回归树模型进行了拟合,并采用 10 倍交叉验证方法避免训练过程中过度拟合的出现。基于前人研究的相关经验<sup>[20-21]</sup>,设置了增强回归树模型的主要参数。本研究采用相对贡献(relative contribution, RC)指标量化各空间协变量对整体增强回归树模型的贡献。

### 3 研究结果

#### 3.1 空间协变量相对重要性

表 2 显示了建模分析过程中相关空间协变量的相对贡献程度。在本研究中,年平均水汽压是最重要的空间协变量,占增强回归树模型总体解释度的 77.27%±5.29%。根据相对贡献率数值,增强回归树集成模型的其他预测因素依次为:年平均气温(RC 为 7.59%±0.72%)、年累计降水量(RC 为 4.69%±2.08%)、平均太阳辐射(RC 为 3.33%±

1.16%)、土壤含水量(RC 为 2.33%±1.10%)、高程(RC 为 1.81%±0.72%)、坡度(RC 1.80%±0.83%)、土壤类别(RC 为 0.76%±0.46%)和土壤深度(RC 为 0.41%±0.37%)。

表 2 相关空间协变量的相对贡献

空间协变量	平均相对贡献率/%	标准偏差/%
年平均水汽压	77.27	5.29
年平均气温	7.59	3.29
年累计降水量	4.69	2.08
平均太阳辐射	3.33	1.16
土壤含水量	2.33	1.10
高程	1.81	0.72
坡度	1.80	0.83
土壤类别	0.76	0.46
土壤深度	0.41	0.37

#### 3.2 辣木环境适应性分布

通过模型模拟结果,辣木的环境适宜区主要分布在北美洲东南部、南美洲东北部地区、非洲中南部、亚洲东南部以及大洋洲北部地区(图 1)。环境适宜性较高的地区,主要集中在印度、巴西、澳大利亚、印度尼西亚等国。适宜区所在的国家也主要集中在热带以及亚热带地区,如北美洲的墨西哥和古巴等;南美洲的巴西东部、委内瑞拉和哥伦比亚等;非洲的塞内加尔、马里南部、马达加斯加西部沿海及莫桑比克沿海等;亚洲的印度西部沿海、菲律宾北部、中国的南部等地区;大洋洲的印度尼西亚、巴布亚新几内亚、澳大利亚北部沿海等。

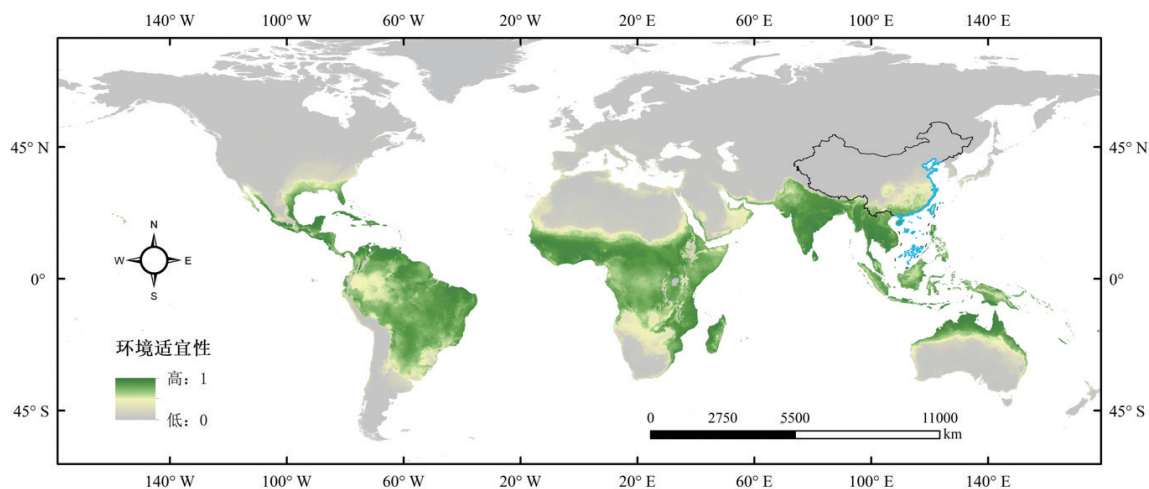


图 1 全球辣木环境适宜性分布评价

### 3.3 精度评定

已知的辣木种植记录包括 960 个点。从采样点分布情况来看,辣木基本分布在非洲中部、亚洲东南部、大洋洲北部、北美洲南部以及南美洲东北部,欧洲地区基本无分布。从视觉效果上看,发生点分布区域的环境适宜性相对较高,样本点分布与模型结果较为一致(图 2)。

模型精度评价采用曲线下面积(Area under the curve, AUC)进行模型精度评测<sup>[22]</sup>。AUC 值范围 0.5~1.0, AUC 值越接近 1.0 表示模型判断力越好, AUC>0.9 模型效果为很好。增强回归树模型对训练数据集(10 倍交叉验证下的 AUC=0.955±0.009)和测试数据集(AUC=0.955±0.010)均具有良好的模拟性能。

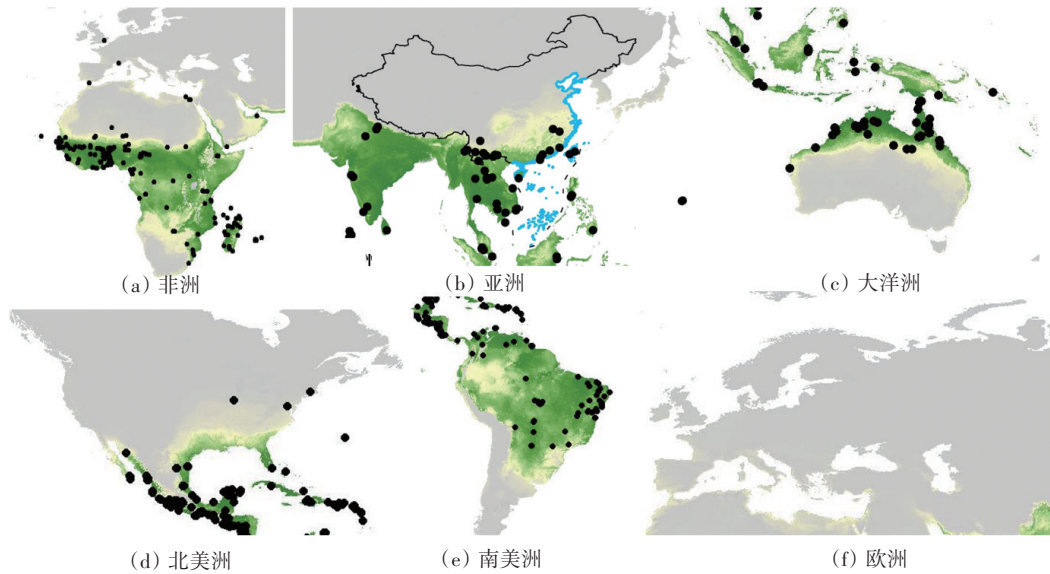


图 2 预测环境适宜性分布与实际辣木出现记录分布

### 3.4 辣木适宜种植的土地资源面积

以 0.5 为阈值,将环境适宜性分布图转化为二值图,对每个网格单元进行土地资源适宜和不适宜的区分。表 3 列出了预测的全球辣木适宜种植的面积,总量约为  $3.62 \times 10^9 \text{ hm}^2$ 。在涉及的 6 个大洲中,非洲的适宜面积最大,约为  $1.31 \times 10^9 \text{ hm}^2$ ,第二大的适宜面积在南美洲,约为  $1.10 \times 10^9 \text{ hm}^2$ ,第三大在亚洲约为  $7.86 \times 10^8 \text{ hm}^2$ ,其次分别为大洋洲  $2.14 \times 10^8 \text{ hm}^2$ 、北美洲  $2.05 \times 10^8 \text{ hm}^2$  以及欧洲  $6.70 \times 10^5 \text{ hm}^2$ 。

在适宜面积最大的前 10 位国家中,巴西的适宜种植辣木的土地资源面积最大,为  $6.74 \times 10^8 \text{ hm}^2$ ,其次是印度 ( $2.53 \times 10^8 \text{ hm}^2$ )、扎伊尔 ( $1.86 \times 10^8 \text{ hm}^2$ )、澳大利亚 ( $1.77 \times 10^8 \text{ hm}^2$ )、印度尼西亚 ( $1.38 \times 10^8 \text{ hm}^2$ ) 以及苏丹 ( $1.07 \times 10^8 \text{ hm}^2$ )。此外,其他 4 个国家的适宜土地资源面积均不足  $10^8 \text{ hm}^2$ ,分别为中国 ( $8.02 \times 10^7 \text{ hm}^2$ )、墨西哥 ( $7.72 \times 10^7 \text{ hm}^2$ )、哥伦

表 3 适合辣木生长的主要区域和国家的适宜土地资源面积

地区/国家	适宜土地资源面积/ $\text{hm}^2$
非洲	$1.31 \times 10^9$
扎伊尔	$1.86 \times 10^8$
苏丹	$1.07 \times 10^8$
尼日利亚	$7.43 \times 10^7$
南美洲	$1.10 \times 10^9$
巴西	$6.74 \times 10^8$
哥伦比亚	$7.60 \times 10^7$
亚洲	$7.86 \times 10^8$
印度	$2.53 \times 10^8$
印度尼西亚	$1.38 \times 10^8$
中国	$8.02 \times 10^7$
大洋洲	$2.14 \times 10^8$
澳大利亚	$1.77 \times 10^8$
北美洲	$2.05 \times 10^8$
墨西哥	$7.72 \times 10^7$
欧洲	$6.70 \times 10^5$
世界	$3.62 \times 10^9$

比亚 ( $7.60 \times 10^7 \text{ hm}^2$ ) 和尼日利亚 ( $7.43 \times 10^7 \text{ hm}^2$ )。除上述 10 个国家外,其余 137 个国家具有适宜辣木生长的土地,总面积共计约为  $1.77 \times 10^9 \text{ hm}^2$ 。

## 4 讨论

将现有辣木分布记录数据和收集整理的气象、太阳辐射、土地和地形环境要素进行关联,利用空间信息技术和数据驱动模型算法,模拟全球辣木种植适宜区的分布,并计算适宜地面积,弥补了当前对辣木全球分布认知不足的缺陷,增强了对全球辣木环境适宜程度的了解。

在使用 BRT 模型对辣木环境适宜性模拟过程中,相关空间协变量的相对贡献率显示出多个环境变量在区分适合辣木的土地方面发挥重要作用,年平均水汽压、年平均气温、年累计降水量、平均太阳辐射和土壤含水量的平均相对贡献率都高于 2.00%,是辣木生长的主要环境限制因素,这与 Tshabalala 等的研究结果相符<sup>[12]</sup>,且本次研究的模型模拟精度 95.5%,高于其研究所得 81% 的模拟精度。结果表明:辣木生长适宜区主要分布在中美洲、亚洲东南部、大洋洲北部、非洲中部以及南美洲东部和北部,与样本点分布情况以及《中国植物志》<sup>[23]</sup>中阐述范围基本一致,在中国的适宜种植区则主要集中在分布于南部沿海地区。全球适宜性土地资源面积统计结果表明,全球辣木适宜性面积总量约为  $3.62 \times 10^9 \text{ hm}^2$ ,非洲中部 ( $1.31 \times 10^9 \text{ hm}^2$ )、南美洲东部和北部 ( $1.10 \times 10^9 \text{ hm}^2$ )、亚洲东南部 ( $7.86 \times 10^8 \text{ hm}^2$ ) 适合种植辣木的面积较大。从单个国家来看,巴西 ( $6.74 \times 10^8 \text{ hm}^2$ ) 和印度 ( $2.53 \times 10^8 \text{ hm}^2$ ) 均具有良好的辣木种植环境。辣木的市场需求仍不断扩大,但盲目的引种、扩种不仅会造成生态环境破坏,也会影响辣木产量和质量。利用模型预测结果可为辣木适宜生长区的确定及合理人工培育种植提供参考,对缓解辣木过度采挖、解决辣木供需矛盾等问题有重要意义。

## 5 结论

合理扩大辣木种植面积是一个复杂的规划,其

中辣木的选种、育苗、栽培以及采摘各项环节中环境变化都对辣木的产量和质量有很大影响<sup>[24]</sup>,直接影响辣木中各类成分的含量。本研究仅从宏观空间环境的角度来讨论适合辣木种植的区域,对辣木的产量和各类成分估算未作过多涉及。今后的研究中,将结合基于地理信息系统的环境政策综合气候模型 (GIS-environmental policy-integrated climate model, GEPIC),探索经济效益良好且能够与生态环境协调发展的辣木引种、扩种新模式。

## 参考文献 (References)

- [1] Mulugeta G, Fekadu A. Industrial and agricultural potentials of Moringa[J]. Journal of Natural Sciences Research, 2014, 4(14): 2224-3186.
- [2] Eilert U, Wolters B, Nahrstedt A. The antibiotic principle of seeds of *Moringa oleifera* and *Moringa stenopetala*[J]. Planta Medica, 1981, 42(5): 55-61.
- [3] Al-Said M S, Mothana R A, Al-Yahya M A, et al. Edible oils for liver protection: Hepatoprotective potentiality of *Moringa oleifera* seed oil against chemical-induced hepatitis in rats[J]. Journal of Food Science, 2012, 77(7): T124-T130.
- [4] 初雅洁, 符史关, 龚加顺. 云南不同产地辣木叶成分的分析比较[J]. 食品科学, 2016, 37(2): 160-164.
- [5] Anwar F, Latif S, Ashraf M, et al. *Moringa oleifera*: A food plant with multiple medicinal uses[J]. Phytotherapy Research, 2007, 21(1): 17-25.
- [6] 许敏, 赵三军, 宋晖, 等. 辣木的研究进展[J]. 食品科学, 2016(23): 299-309.
- [7] Busari M B, Muhammad H L, Ogbadoyi E O, et al. Hypoglycaemic properties of *Moringa oleifera* Lam seed oil in Normoglycaemic rats[J]. IOSR Journal of Pharmacy & Biological Sciences, 2014, 9(6): 23-27.
- [8] Jha N, Mohanka R, Azad R. Antifungal investigation of the constituents of *Moringa oleifera* Lamk. root bark extract[J]. Asian Journal of Chemistry, 2009, 21(9): 7437-7439.
- [9] Momoh M A, Chime S A, Kenekukwu F C. Novel drug delivery system of plant extract for the management of diabetes: An antidiabetic study[J]. Journal of Nutraceuticals Functional & Medical Foods, 2013, 10(3): 252-263.
- [10] Guevara A P, Vargas C, Sakurai H, et al. An antitumor promoter from *Moringa oleifera* Lam[J]. Mutation Research, 1999, 440(2): 181-188.
- [11] Ganguly S. Indian ayurvedic and traditional medicinal implications of indigenously available plants, herbs and fruits: A review[J]. International Journal of Research in Ayurveda and Pharmacy, 2013, 4(4): 623-625.
- [12] Tshabalala T, Ncube B, Moyo H P, et al. Predicting the

- spatial suitability distribution of *Moringa oleifera* cultivation using analytical hierarchical process modelling[J]. South African Journal of Botany, 2020, 129: 161–168.
- [13] Siddhuraju P, Becker K. Antioxidant properties of various solvent extracts of total phenolic constituents from three different agroclimatic origins of drumstick tree (*Moringa oleifera* Lam.) leaves[J]. Journal of Agricultural and Food Chemistry, 2003, 51(8): 2144–2155.
- [14] Ndabigengesere A, Narasiah K S. Quality of water treated by coagulation using *Moringa oleifera* seeds[J]. Water Research, 1998, 32(3): 781–791.
- [15] 刘子记, 孙继华, 刘昭华, 等. 特色植物辣木的应用价值及发展前景分析[J]. 热带作物学报, 2014, 35(9): 1871–1878.
- [16] 吕晓静. 环境因素对辣木光合作用及与其相关生理指标的影响[D]. 呼和浩特: 内蒙古农业大学, 2009.
- [17] 马春花, 邵建辉, 田洋, 等. 云南辣木种植区气候条件适宜性初步评价[J]. 热带农业科学, 2016, 36(6): 27–28.
- [18] Elith J, Leathwick J R, Hastie T. A Working guide to boosted regression trees[J]. Journal of Animal Ecology, 2008, 77(4): 802–813.
- [19] Hao M M, Chen S, Qian Y S, et al. Using machine learning to identify the potential Marginal Land suitable for giant silvergrass (*Miscanthus giganteus*) [J]. Energies, 2022, 15(2): 591–604.
- [20] Zheng C J, Jang D, Ding F Y, et al. Spatiotemporal patterns and risk factors for Scrub Typhus from 2007 to 2017 in Southern China[J]. Clinical Infectious Diseases, 2019, 69(7): 1205–1211.
- [21] Messina J P, Kraemer Moritz U G, Brady O J, et al. Mapping global environmental suitability for Zika virus [J]. eLife, 2016, 5(15272): 1–22.
- [22] 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, 15(4): 365–372.
- [23] 中国科学院中国植物志编辑委员. 中国植物志·第三十四卷·第一分册[M]. 北京: 科学出版社, 1984.
- [24] Alatar A A. Effect of temperature and salinity on germination of *Achillea fragrantissima* and *Moringa peregrina* from Saudi Arabia[J]. African Journal of Biotechnology, 2011, 10(17): 3393–3398.

## Estimating global environmental suitability distribution of *Moringa oleifera* Lam. using data-driven model

FAN Peiwei<sup>1</sup>, HAO Mengmeng<sup>1,2</sup>, DING Fangyu<sup>1,2\*</sup>, WANG Qian<sup>1,2</sup>, DONG donglin<sup>3</sup>, MA Zhimin<sup>4</sup>, QIAN Jun<sup>4</sup>, LI Xinhua<sup>4</sup>, CHEN Jing<sup>4</sup>, WAN Wenhong<sup>4</sup>, ZHAO Yan<sup>4</sup>, ZHANG Xiaobo<sup>5\*</sup>

1. Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
2. College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
3. Department of Geological Engineering and Environment, China University of Mining and Technology(Beijing), Beijing 100083, China
4. Information Center of Jiangxi Province, Nanchang 330036, China
5. State Key Laboratory Breeding Base of Dao-di Herbs, National Resource Center for Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing 100700, China

**Abstract** As a new Chinese medicine resource, *Moringa oleifera* Lam. is widely used in botanical medicine. To avoid economic losses caused by blind expanded introduction and cultivation it is urgent to evaluate the spatial distribution of land resources suitable for *Moringa oleifera* Lam. In this study, 960 global distribution records of *Moringa oleifera* Lam. are collected to assess the spatial distribution of global environmental suitability of *Moringa oleifera* Lam. by using spatial information technology and a data-driven model of spatial co-variables of climate, soil, topography and solar radiation. The results show that the land resources suitable for the cultivation of *Moringa oleifera* Lam. are about 3615.96 million hm<sup>2</sup> in the world, which are mainly distributed in Southern North America, Eastern and Northern South America, Central Africa, Southeast Asia and Northern Oceania. This study provides a solid foundation for the cultivation partition and promotion program of *Moringa oleifera* Lam. and helps to alleviate the over-mining of moringa and solve the contradiction between supply and demand of *Moringa oleifera* Lam.

**Keywords** *Moringa oleifera* Lam.; spatial covariates; spatial information technology; data-driven model; environmental suitability ●



(责任编辑 徐丽娇)