

# 基于贝叶斯累加回归树评估中国非洲猪瘟发生风险区域

王颢然, 肖建华, 欧阳茂霖, 高宏岩, 刘宇馨, 高利, 高翔, 王洪斌\*

(东北农业大学动物医学学院黑龙江省动物普通疾病防治重点实验室, 哈尔滨 155030)

**摘要:** 为探究中国非洲猪瘟发生的风险区域, 对疫情的风险评估提供决策参考。该研究基于贝叶斯累加回归树 (Bayesian Additive Regression Tree, BART) 模型, 应用 2018 年 8 月至 2021 年 8 月非洲猪瘟发生数据评估了中国非洲猪瘟的风险区域及相关的因素。结果表明: 1) 在影响非洲猪瘟风险区分布的环境因素中, 归一化城市土地指数 ( $0.213 \pm 0.026$ ) 贡献性最高, 其次是归一化差分植被指数 ( $0.207 \pm 0.028$ ), 年平均气温 ( $0.199 \pm 0.025$ ), 家猪的分布 ( $0.194 \pm 0.025$ ) 和最冷季降水量 ( $0.187 \pm 0.026$ )。2) 在评估模型中, 小型城市相比于其他类型的城市发生非洲猪瘟的风险更高。非洲猪瘟的发生风险随着年平均气温、归一化植被指数、家猪的数量的升高而升高。最冷季降水量的升高会降低非洲猪瘟发生的可能性。3) 非洲猪瘟的风险区域主要集中在中国的东部和西南。4) 在预测地图中, 中国东南地区的不确定性较高, 鉴于影响非洲猪瘟传播和发生的因素众多, 未来需要对此区域保持重点关注。在模型的预测精准性评估中, 模型曲线下面积 (Area Under Curve, AUC) 为 0.90, 证明风险地图的预测准确性较高。研究结果可为 BART 在动物传染病风险评估中的应用提供参考, 同时为了解中国非洲猪瘟发生主要风险区域和影响因素, 采取合理的预防和控制措施提供信息和建议。

**关键词:** 病害; 模型; 非洲猪瘟; 风险评估; 贝叶斯累加回归树; 影响因素

doi: 10.11975/j.issn.1002-6819.2022.11.020

中图分类号: S855

文献标志码: A

文章编号: 1002-6819(2022)-11-0180-08

王颢然, 肖建华, 欧阳茂霖, 等. 基于贝叶斯累加回归树评估中国非洲猪瘟发生风险区域[J]. 农业工程学报, 2022, 38(11): 180-187. doi: 10.11975/j.issn.1002-6819.2022.11.020 <http://www.tcsae.org>

Wang Haoran, Xiao Jianhua, Ouyang Maolin, et al. Assessing the risk areas for African swine fever in China using Bayesian additive regression trees[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(11): 180-187. (in Chinese with English abstract) doi: 10.11975/j.issn.1002-6819.2022.11.020 <http://www.tcsae.org>

## 0 引言

非洲猪瘟 (African Swine Fever, ASF) 是由非洲猪瘟病毒 (African Swine Fever Virus, ASFV) 引起猪的一种急性、热性、高度接触性传染病<sup>[1]</sup>。该病以呼吸障碍, 皮肤和内脏器官严重出血为主要特征<sup>[2]</sup>。根据患病的严重程度可分为最急性, 急性, 亚急性 and 慢性四种形式, 强毒株感染导致的猪发病率和死亡率可高达 100%<sup>[3]</sup>。鉴于该病的严重危害性, 世界动物卫生组织将其列为必须报告的动物疫病, 中国将其列为一类动物疫病<sup>[4]</sup>。

1921 年在肯尼亚首次报道了 ASF 疫情, 疫情起初主要发生在非洲的东部区域, 随后疫情不断在撒哈拉以南的国家爆发<sup>[5]</sup>。1957 年, ASF 首次在葡萄牙里斯本机场附近的猪场中发生, 经调查认为是饲喂病猪来自于安哥拉国际航班上受污染的食品垃圾而引起的<sup>[6]</sup>。随后, 西班牙 (1960 年)、意大利 (1967 年)、法国 (1977 年)、比利时 (1985 年)、荷兰 (1986 年) 等欧洲国家均出现 ASF 的疫情报道<sup>[3,6]</sup>。1971 年, ASF 侵入西半球的古巴

(1971 年)、巴西 (1978 年)、海地 (1978 年) 等国家, 这些国家通过对猪只的大规模扑杀从而得以根除 ASF<sup>[7]</sup>。2007 年格鲁吉亚黑海港口城市波季的家猪出现不明原因死亡, 经调查后格鲁吉亚农业部上报了该地区首起 ASF 疫情<sup>[8]</sup>。疫情迅速在高加索地区蔓延, 相继传入了亚美尼亚 (2007 年), 俄罗斯联邦 (2007 年) 和阿塞拜疆 (2008 年) 等地区。ASF 传入俄罗斯后, 疫情从南部地区逐年向西北的欧洲国家和俄罗斯中部地区进行扩张, 乌克兰 (2012 年)、白俄罗斯 (2013 年)、波兰 (2014 年) 等多个国家相继出现 ASF 疫情的报道。ASF 在俄罗斯境内经多次长距离的跳跃式传播后, 于 2017 年 3 月 27 日到达远东地区的伊尔库茨克州, 距离中国边境满洲里仅 1 000 km<sup>[3]</sup>。2018 年 8 月 1 日, 辽宁省沈阳市沈北新区某养殖户的生猪发生 ASF 疫情。2018 年 8 月 3 日, 经中国动物卫生与流行病学中心 (国家外来动物疫病研究中心) 确诊, 该起疫情为 ASF 疫情, 这是中国首起 ASF 疫情<sup>[9]</sup>。随即, ASF 疫情开始迅速在中国多个省市蔓延, 对中国生猪养殖业造成了重大的危害<sup>[10]</sup>。

传染病的发生与流行必须同时具备传染源、传播途径和易感动物这三个重要的环节。猪科动物是 ASFV 的易感动物, 携带 ASFV 的家猪和野猪是主要的传染源。软蜱是 ASFV 的主要生物媒介, 也是 ASFV 在饲养与野外环境中传播的重要因素<sup>[11]</sup>。ASFV 可以通过接触传播、食物传播、软蜱叮咬等多种传播途径使猪只感染与

收稿日期: 2021-09-18 修订日期: 2022-05-21

基金项目: 国家自然科学基金项目 (31802217)

作者简介: 王颢然, 博士生, 研究方向为兽医信息学。

Email: 617280311@qq.com

\*通信作者: 王洪斌, 博士, 教授, 博士生导师, 研究方向为兽医信息学、动物麻醉和动物比较医学。Email: hbwang1940@neau.edu.cn

发病。例如，通过接触受污染的饲料、水源、泔水、交通工具等。养殖管理、环境变化、交通运输等均是影响 ASFV 传播与发生的重要因素<sup>[12]</sup>。鉴于目前尚无有效的疫苗和治疗方法，ASF 的预防和控制主要是依靠严格的监测和卫生管理等措施，所以建立有效的风险评估模型是预防和控制 ASF 蔓延的重要手段<sup>[13-14]</sup>。

传染病研究专家越来越重视应用现代信息技术加强动物养殖管理和风险评估<sup>[15]</sup>。一些统计学和机器学习方法也逐渐被应用于评估动物传染病的风险区域及相关流行风险因素的变化趋势。例如，Ma 等基于最大熵模型结合气象、家畜分布等因素对中国小反刍兽疫的风险区域进行了探索<sup>[16]</sup>；Wang 等基于多准则决策支持分析通过专家评分确定主要的风险因素并对中国的口蹄疫风险区域进行了评估<sup>[17]</sup>；Assefa 等基于集合物种分布模型结合相关的风险因素对埃塞俄比亚阿姆哈拉州的炭疽病风险区域进行模拟<sup>[18]</sup>。在疾病生态学中，增强回归树（Boosted Regression Trees, BRTs）已成为研究人畜共患病和病媒分布的主要方法<sup>[19-21]</sup>。

大多数流行的方法在处理不确定性上都存在明显的局限性。在此背景下，2020 年 Carlson 开发了贝叶斯累加回归树（Bayesian Additive Regression Trees, BART）物种分布模型<sup>[22]</sup>。BART 是一种强大的机器学习方法，由 Chipman 等在 2008 年提出，并在 2010 年将其推广应用到分类问题上<sup>[23]</sup>。虽然 BART 本身不是一种新方法，但它们对于物种分布模型（Species Distribution Model, SDM）来说是非常新颖的。贝叶斯累加回归树带来了分类树方法的概念和优势，通过结合贝叶斯方法可以直观地处理模型中的不确定性，这可能使贝叶斯累加回归树成为一种替代 BRTs 的方法。此外，BART 拥有 BRTs 的许多优点，如易于应用和可视化的“黑箱”模型，并且在模型比较中优于其他分类回归树方法<sup>[23]</sup>。

本研究基于 BART 结合相关的影响因素对中国 ASF 的风险区域进行评估。同时，对 ASF 发生相关的主要影响因素及变化趋势进行分析。研究结果可为 BART 在动物传染病风险评估中的应用提供参考，同时为了解中国非洲猪瘟发生的主要风险区域及影响因素，采取合理的预防和控制措施提供信息和建议。

## 1 材料与方法

### 1.1 数据的获取

ASF 疫情数据来源于联合国粮食及农业组织（Food and Agriculture Organization of the United Nations, FAO）的跨界动植物病虫害紧急预防系统，疫情数据的时间范围为 2018 年 8 月—2021 年 8 月。为了消除建模过程中由于集群效应可能产生的偏差，应用 ArcGIS 10.2 将所收集的数据以 5 km 设置缓冲区，对缓冲区进行叠加分析，删除两相交的缓冲区对应分布点中的一个，直至每 5 km×5 km 网格中仅有一个 ASF 疫情点<sup>[24]</sup>。

### 1.2 BART 模型

BART 是一种将贝叶斯理论与累加树模型结合的集

成方法，主要包括累加树模型和累加树模型参数的先验信息两个部分，本质上是对累加树模型中的参数施加先验信息<sup>[25]</sup>。在易于理解的单树模型（单颗二叉回归树）中，树  $T$  是由一组内部节点决策规则和一组叶子节点组成的二叉树， $M=\{\mu_1, \mu_2, \dots, \mu_b\}$  表示一组与  $T$  的  $b$  个叶子节点都相关联的参数值。其中，非叶子节点的分割规则是对输入变量值域空间进行二元分割，分割的形式为  $\{x \in A\}$  和  $\{x \notin A\}$ ， $A$  为输入变量  $x$  值域的子集。通过由上至下的决策规则使每个  $x$  都与树  $T$  的一个叶子节点相关联，其关联的参数  $\mu_i$  即为输入的  $x$  的响应值。在给定  $T$  和  $M$  的前提下，单颗树的模型可以表示为

$$Y=g(x; T, M)+\varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (1)$$

式中  $g(x; T, M)$  就是把  $x$  映射到  $\mu_i \in M$  的函数， $\varepsilon$  是样本数的高斯噪声，并且  $\varepsilon$  服从  $N(0, \sigma^2)$ 。因此，累加树模型可以表示为

$$Y=\sum_{j=1}^m g(x; T_j, M_j)+\varepsilon, \varepsilon \sim N(0, \sigma^2) \quad (2)$$

式中  $T_j$  表示为每一颗二叉回归树， $M_j$  表示与  $T_j$  关联的叶子节点参数集， $x$  通过函数  $g(x; T_j, M_j)$  映射到  $\mu_{ij} \in M$ 。当树的数量  $m$  大于 1 时，每一个  $\mu_{ij}$  仅代表最终预测值的一部分。当  $g(x; T_j, M_j)$  只有一个变量  $x$  时，每个  $\mu_{ij}$  仅代表这个变量的主要影响结果。当  $g(x; T_j, M_j)$  依赖于多个变量  $x$  时， $\mu_{ij}$  则表示为多个变量的交互作用结果<sup>[26]</sup>。

由于参数  $(T_1, M_1)$ ， $(T_2, M_2)$ ， $\dots$ ， $(T_m, M_m)$  和  $\varepsilon$  决定了 BART 模型，因此需要对这些参数添加先验信息。对于一个给定的样本数据集，通过应用贝叶斯方法可以得到模型参数的联合后验分布，然后使用采样获得参数值。BART 模型参数训练使用的是贝叶斯 Backfitting MCMC（Markov Chain Monte Carlo）算法，是一种结合了贝叶斯 Backfitting 拟合策略和 MCMC 重复抽样的算法<sup>[27]</sup>。

前述的 BART 模型是基于连续型响应变量，而 BART 模型同样可以使用 logit link 处理二分类问题，其公式为

$$Y=\Phi[\sum_{j=1}^m g(x; T_j, M_j)] \quad (3)$$

式中  $\Phi$  是标准正态分布的累积分布函数<sup>[28]</sup>。

### 1.3 环境数据的处理及模型的构建

根据先前对 ASF 等传染病的风险评估研究，选择应用气候数据、土地利用类型、家猪分布数据、归一化差分植被指数（Normalized Difference Vegetation Index, NDVI）对 ASF 进行风险区模拟<sup>[19,29-32]</sup>。表 1 为最初考虑构建 BART 模型的 33 个环境变量。其中，19 个生物气候变量来源于 WorldClim 数据集（<https://www.worldclim.org/>）；从土地利用协调数据库（<https://luh.umd.edu/index.shtml>）中选择 12 个相关土地利用的环境变量。家猪密度的分布来源于 FAO（<https://www.fao.org/home/zh/>）。NDVI 对于测量无脊椎动物的分布或将地貌划分为不同的生物群落来说是一个重要的变量<sup>[22]</sup>。在缺失软蜱分布数据的情况下，选择应用 NDVI 数据进行补充，NDVI 数据来源于资源环境科学与数据中心（<https://www.resdc.cn/Default.aspx>）。在获得所有环境变量数据后，所有数据转为栅格格式，并裁剪至中国的地理区域范围，栅格的空间分辨率约为 5 km×5 km。

表 1 用于构建模型的环境变量  
Table 1 Variables used in the model

变量 Variable	描述 Description	变量 Variable	描述 Description
Bio1	年平均气温 Annual mean temperature/°C	Bio18	最暖季降水量 Precipitation of warmest quarter/mm
Bio2	平均气温日较差 Mean diurnal range/°C	Bio19	最冷季降水量 Precipitation of coldest quarter/mm
Bio3	等温性 Isothermally (Bio2/Bio7) ×100	Primf	归一化原始森林土地指数 Normalized forested primary land index
Bio4	温度季节性标准差 Standard deviation of temperature seasonality	Primn	归一化原始非森林土地指数 Normalized non-forested primary land index
Bio5	最暖月最高温度 Max temperature of warmest month/°C	Secdf	归一化潜在的森林次生土地指数 Normalized potentially forested secondary land index
Bio6	最冷月最低温度 Min temperature of coldest month/°C	Secdn	归一化潜在的非森林次生土地指数 Normalized potentially non-forested secondary land index
Bio7	年均温度变化范围 Temperature annual range (Bio5–Bio6)/°C	Pastr	归一化受管理的牧场指数 Normalized managed pasture index
Bio8	最湿季度平均温度 Mean temperature of wettest quarter/°C	NRI	归一化牧场指数 Normalized rangeland index
Bio9	最干季度平均温度 Mean temperature of driest quarter/°C	NULI	归一化城市土地指数 Normalized urban land index
Bio10	最暖季度平均温度 Mean temperature of warmest quarter/°C	C3ann	归一化 C3 年生作物指数 Normalized C3 annual crops index
Bio11	最冷季度平均温度 Mean temperature of coldest quarter/°C	C3per	归一化 C3 多年生作物指数 Normalized C3 perennial crops index
Bio12	年均降水量 Annual precipitation/mm	C3nfx	归一化 C3 固氮作物指数 Normalized C3 nitrogen-fixing crops index
Bio13	最湿月降水量 Precipitation of wettest month/mm	C4ann	归一化 C4 年生作物指数 Normalized C4 annual crops index
Bio14	最干月降水量 Precipitation of driest month/mm	C4per	归一化 C4 多年生作物指数 Normalized C4 perennial crops index
Bio15	降水量季节变化系数 Coefficient of variation of precipitation seasonality	PigN	家猪的数量 Number of domestic pigs
Bio16	最湿季降水量 Precipitation of wettest quarter/mm	NDVI	归一化差分植被指数 Normalized difference vegetation index
Bio17	最干季降水量 Precipitation of driest quarter/mm		

为了避免环境因子间高度的相关性影响预测结果的准确性,对环境因子进行预处理。首先,通过调用基于 R 语言开发的“car”包进行方差膨胀因子(Variance Inflation Factor, VIF)分析以初步消除高度相关的变量。在 VIF 分析中, VIF 值大于 10 的变量被认为存在多重共线性问题并选择剔除<sup>[33-34]</sup>。随后,将通过 VIF 分析后的环境变量纳入 BART 模型中,根据模型中变量重要性程度进行再次的筛选。BART 模型与其他模型(例如增强回归树和最大熵)一样需要使用缺失数据。根据 Carlson 的建议,应用 ArcGIS 10.2 在研究区域内随机生成与 ASF 疫情点相同数量的伪缺失数据用于构建模型<sup>[22]</sup>。最后,应用“embarcadero”R 包进行筛选和构建模型,规则如下:

- 1) 拟合一个包含所有预测变量和一个小树集合的完整模型(默认  $m=10$ ),固定次数(默认  $n=50$ )。
- 2) 移除所有 50 次运行中重要程度最低的变量。
- 3) 再次重新运行模型消除重要程度最低的变量( $n=50$  次),记录训练集上的均方根误差(Root Mean Square Error, RMSE)。
- 4) 重复步骤 2) 和 3),直到只剩下最重要的变量。
- 5) 最后,选择平均 RMSE 最低的模型。

平均 RMSE 最低的模型被用于评估 ASF 风险区域,对于变量重要性的评估通常在 BART 模型中通过计算指定变量在树的完整后验概率抽样中使用的次数来衡量(类似于 BRTs 中的变量重要性评估)。变量反应曲线是通过应用 BART 模型对每个变量在其值范围内预测而生成的。变量重要性的标准差和变量反应曲线的可信区间通过最低 RMSE 模型的 50 次重复运行计算而完成的。

#### 1.4 不确定性分析及 ROC 分析

预测结果的 95%可信区间用于可视化 BART 预测中的不确定性,通过计算受试者工作特征(Receiver

Operating Characteristic, ROC) 曲线下面积(Area Under Curve, AUC)值量化模型的预测准确性。一般而言,当  $AUC \leq 0.6$  时认为模型预测结果失败;  $0.6 < AUC \leq 0.7$  时,模型预测结果较差;  $0.7 < AUC \leq 0.8$  时,模型的预测结果一般;  $0.8 < AUC \leq 0.9$  时,模型预测结果良好;  $0.9 < AUC \leq 1.0$  时,模型预测结果优秀。

#### 1.5 统计学方法

使用 R 4.0.5 软件中的“car”包对 31 个环境变量进行 VIF 分析,应用“embarcadero”包对通过 VIF 分析的环境变量进行再次筛选并构建 BART 模型,并评估变量的重要性,构建变量反应曲线,二维部分依赖图,ASF 风险地图以及进行不确定性分析及 ROC 分析。

## 2 结果与分析

### 2.1 变量的初步筛选

VIF 值小于 10 的所有变量被用于构建 BART 模型,共计 16 个环境变量见表 2。

表 2 环境变量的 VIF 值  
Table 2 Variance Inflation Factor (VIF) values for environment variables

变量 Variable	VIF 值 VIF value	变量 Variable	VIF 值 VIF value
Bio1	5.334	C4ann	2.669
Bio2	3.654	Primf	4.016
Bio8	3.652	Primn	2.520
Bio15	3.086	NRI	2.508
Bio19	4.135	Secdf	5.259
C3ann	2.572	NULI	1.253
C3nfx	1.964	PigN	1.394
C3per	1.966	NDVI	1.499

### 2.2 变量的确定及变量重要性

分别应用不同数量的(10、20、50、100、150、200)

树构建 BART 模型, 当去掉 11 个变量后模型的 RMSE 最低 (图 1a), 最终被用于构建 ASF 风险评估模型的 5 个变量重要程度如下: 归一化城市土地指数的在所有变量中重要性最高,  $0.213 \pm 0.026$ 。其次是归一化差分植被指数,  $0.207 \pm 0.028$ ; 年平均气温,  $0.199 \pm 0.025$ ; 家猪的分布,  $0.194 \pm 0.025$ ; 最冷季降水量,  $0.187 \pm 0.026$  (图 1b)。

### 2.3 变量的反应曲线

通过绘制环境变量的反应曲线, 可以更清晰的描述 ASF 发生的可能性大小与环境因素的关系。如图 2 所示, ASF 发生的可能性先是随着归一化城市土地指数升高而升高, 达到最高点后风险系数开始降低, 此结果表明小型城市相比于其他类型的城市发生 ASF 的风险更高。此外, ASF 的发生风险随着年平均气温, 归一化差分植被

指数, 家猪的数量的升高而升高。最冷季降水量的升高会降低 ASF 发生的可能性。

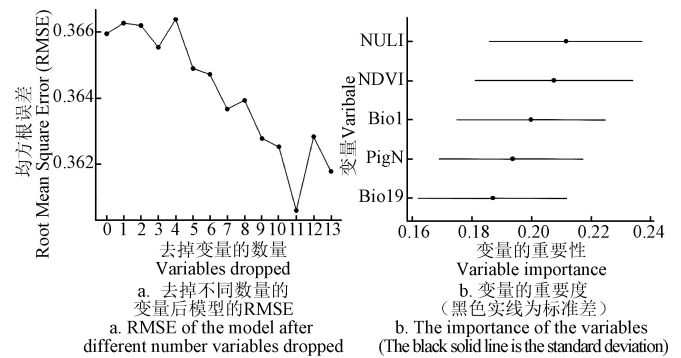
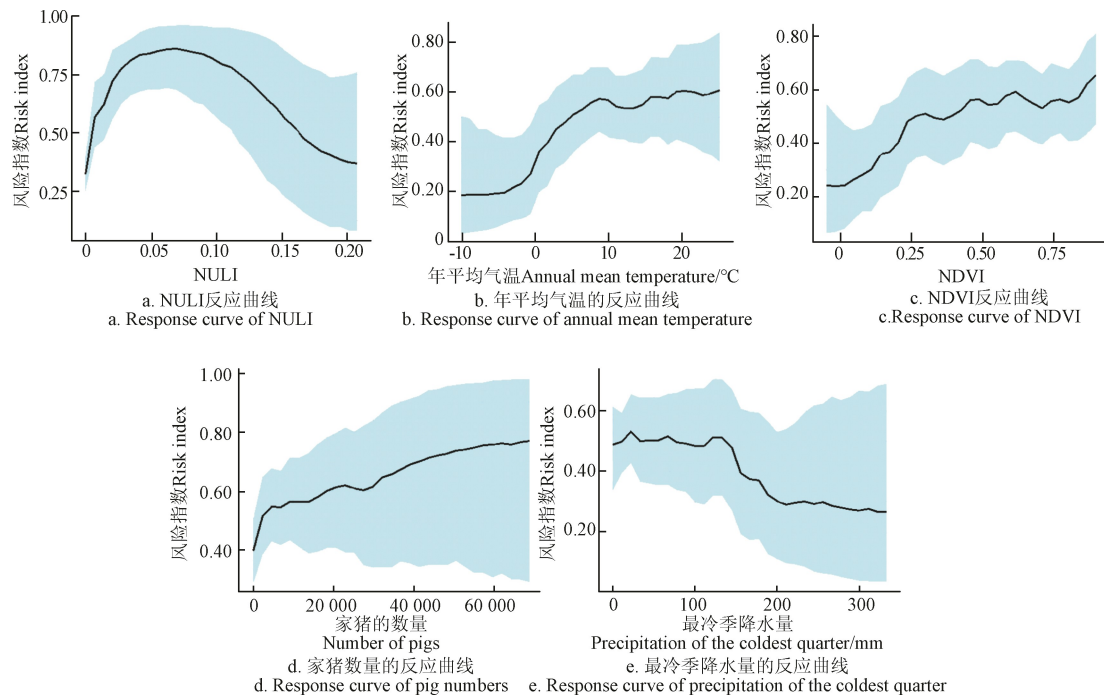


图 1 变量的筛选与重要程度

Fig.1 Selection and importance of variables



注: 阴影区域为 95%可信区间。

Note: The shaded area is 95% confidence interval.

图 2 环境变量的反应曲线

Fig.2 Response curve of environmental variables

通过二维部分依赖图可以更好的诠释两个因素之间的依赖关系。如图 3 所示, 以家猪的分布与年平均气温, 最冷季降水量和归一化城市土地指数之间的依赖关系进行诠释。当年平均气温大于  $20^{\circ}\text{C}$  时, 猪只数量的增加同时会增加 ASF 的发生可能性。猪只数量大于约 25 000 头, 最冷季降水量大于约 150 mm 时, ASF 的发生可能性会相对降低。城市土地的规模大于约 0.06 时, 猪只数量的增加同时会增加 ASF 的发生可能性。

### 2.4 ROC 曲线及二分类阈值

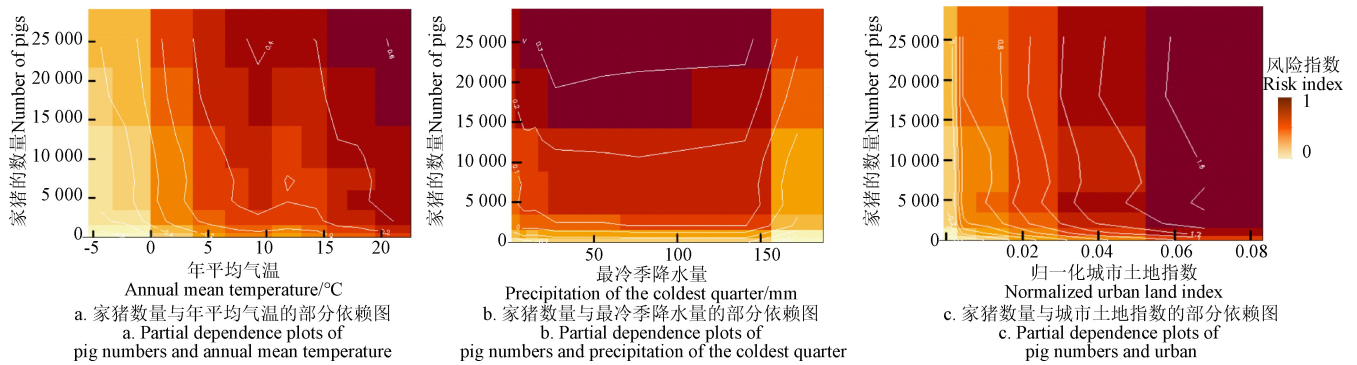
图 4a 展示了 BART 模型的 ROC 曲线, 模型曲线下面积 (AUC) 为 0.90, 证明模型表现优秀。图 4b 展示了训练集数据在模型中的预测概率分布。图 4c 显示了 ASF 是否发生的二分类阈值, 结果为 0.45, 应用此值构建 ASF 二分类风险地图。通过二分类阈值将训练点进行分类,

结果表明训练点大部分都落于 ASF 的发生风险区域中, 进一步证明了模型的预测效果较好, 见图 4d。

### 2.5 ASF 风险区域的分布及不确定分析

图 5a 和图 5c 展示了中国 ASF 风险区域的分布, 从图中可以看出中国 ASF 的风险区域主要集中在中国的东部和西南部, 主要包括东北三省 (黑龙江、辽宁、吉林)、河北、山东、河南、山西、陕西、安徽、江苏、湖北、湖南、广西、广东、海南、贵州、重庆、四川、云南等的部分区域。图 5b 中的不确定性地图展示了 ASF 风险预测地图 (图 5a) 中不确定性。在不确定性地图中, 中国东南部部分地区的预测不确定性较高, 说明此区域的预测结果可能会存在较高的偏差。鉴于影响 ASF 传播和发生的因素众多, 未来需要对此区域保持重点关注。而预测地图对于西部地区 (如西藏, 青海的部分区域) 的预测结果具有较高的确定性。





注：浅黄至深红表示 ASF 的风险指数。

Note: Light yellow to dark red indicates the risk index of ASF.

图 3 环境变量的二维部分依赖图

Fig.3 Two-dimensional partial dependence plots of environmental variables

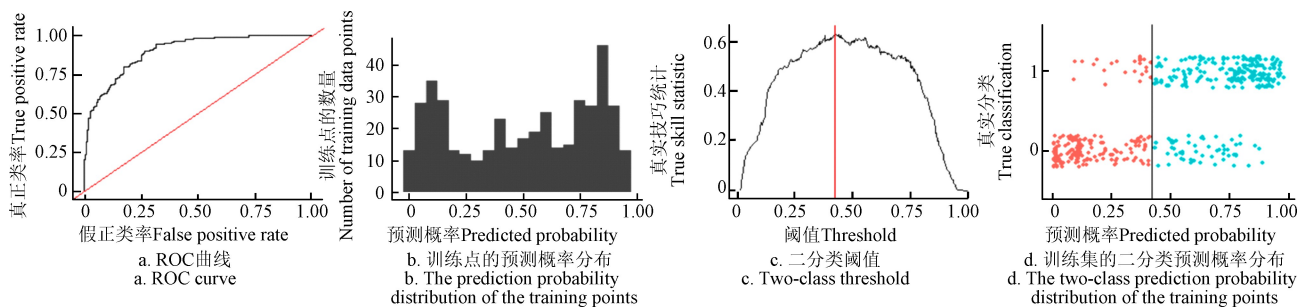
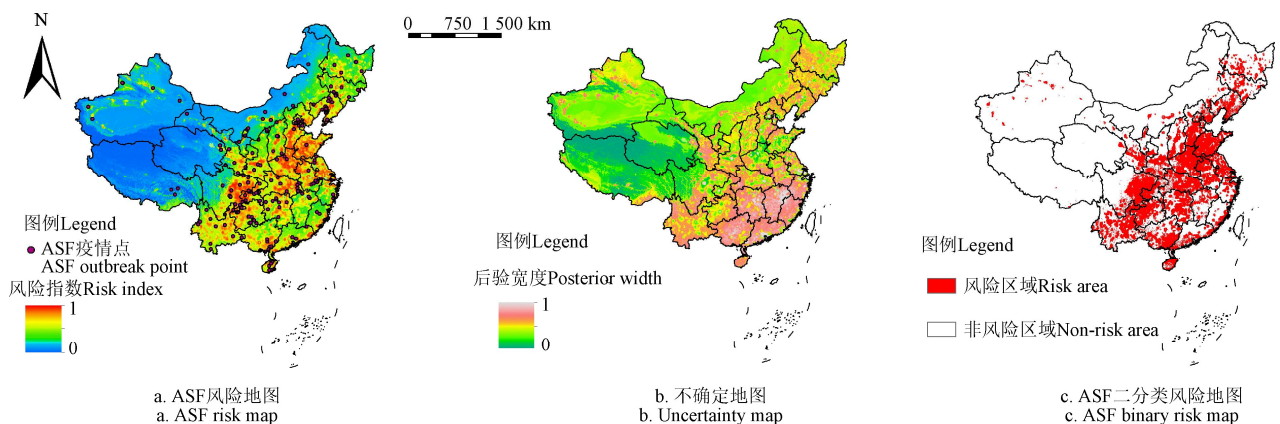


图 4 ROC 曲线及二分类阈值

Fig.4 Receiver Operating Characteristic (ROC) curve and two-class threshold



注：中国香港、中国澳门和中国台湾由于数据来源受限等因素，未包括于此研究内。

Note: Chinese Hong Kong, Chinese Macao and Chinese Taiwan are not included in this study due to limited data sources.

图 5 ASF 风险地图和不确定地图

Fig.5 African Swine Fever (ASF) risk map and uncertainty map

### 3 讨论

自 2018 年 8 月以来, ASF 在中国多个省市爆发并对中国的生猪市场和产业的发展造成了严重的影响<sup>[35]</sup>。ASFV 在外界环境中存活时间长, 传播途径多种多样, 可做长距离跨省市传播。同时此病毒基因型多且结构复杂, 目前尚未研发出有效的疫苗进行预防, 这给疫情的防控带来了极大的难度。针对此疫病最主要的预防措施就是消灭传染源和切断传播途径, 并保持长期的监测, 所以

建立有效的风险评估模型是预防和控制 ASF 蔓延的重要手段<sup>[14]</sup>。

在本研究中, 小型城市发生 ASF 风险可能性较高。小型城市的生猪养殖模式多以小规模养殖和散户养殖为主, 饲养管理和防控认知上的不足很有可能是 ASF 发生风险较高的原因之一。在周勋章等的研究中也证明了规模化养猪户在生物安全建设方面比散户户更加积极, 生物安全行为更加规范<sup>[36]</sup>。此外, 通过对 2018 年 8 月至 2021 年 8 月 ASF 疫情数据的观察也同样发现大部分的疫情点

发生在乡村和县镇。针对小型城市, 加强基层动物防疫体系建设对于 ASF 的防控至关重要。路平等在对评估基层兽医机构防控 ASF 能力和防控政策执行情况的研究中, 揭示了当前基层兽医机构部分关键能力存在不足, “人财物”资源配置好的地区防控 ASF 等重大动物疫病的能力较强<sup>[37]</sup>。综合上述, 小型城市更应该重点关注和加强对 ASF 的预防和控制。

ASF 作为一种动物传染病, 其传播和流行与环境的变化存在相关的联系。一些学者也相继开展了对 ASF 疫情与气候、地理等环境因素之间关系的研究。例如, Liang 等应用随机森林算法结合气候变量构建了 ASF 预测系统<sup>[29]</sup>。陈志杰等应用灰色关联分析法探究了海拔等因素与 ASF 疫情发病率和死亡率之间的关系<sup>[31]</sup>。在本研究中, 最冷季降水量的增加可能会降低 ASF 的发生风险, 此环境变量的变化趋势与 Ma 等应用最大熵模型对 ASF 疫情的预测结果基本保持一致<sup>[30]</sup>。此外, 在缺失软蜱分布数据的条件下, 应用归一化差分植被指数作为软蜱分布数据的替代用于模型的构建中, 作为测量无脊椎动物的分布或将地貌划分为不同的生物群落来说, 归一化差分植被指数是一个重要的变量<sup>[22]</sup>。归一化差分植被指数的变量反应曲线也从侧面反映出软蜱分布可能性的增加可能会增加 ASF 的发生风险。软蜱作为 ASFV 的主要生物媒介, 其对 ASFV 的传播能力同样会受到环境温度和湿度等因素的影响, 且 ASFV 在短距离内可以通过空气进行传播<sup>[3,32]</sup>。变量反应曲线的结果也表明了区域内猪只数量的增加可能会增加 ASF 的发生风险。猪作为 ASFV 的重要宿主和传染源, 一定范围内猪只数量的增加会提高接触 ASFV 感染源的可能性。石国宁等的研究也表明中国生猪养殖业空间格局决定了 ASF 疫情总体分布格局, 这与我们的研究结果相一致<sup>[32]</sup>。

在风险区域的评估中, 当前中国的西南和东部为 ASF 发生的主要风险区域。根据一些对 ASF 的时空分析结果, ASF 疫情的传播方向趋向于从中国的东北至东部和西南<sup>[13,31,38]</sup>。疫情的扩散可能与中国“南猪北养, 东猪西调”的生猪产业布局密切相关。随着中国交通运输的发展, 生猪的长途调运距离也越来越长, 交通的便利促使了中国 ASF 疫情的大面积扩散。此外, 在风险评估中中国的东南区域不确定性较高, 鉴于影响 ASF 传播和发生的因素众多, 未来也需要对此区域保持重点关注。

本研究存在一定的局限性, 一些在 ASF 发生和传播中发挥重要作用的风险因素由于数据获取等因素未被纳入其中, 例如, 养殖管理策略、饲养生产环境、生猪贸易往来、外来人员流动等因素。这些因素往往在 ASF 的发生和传播中起着关键作用。在后续的研究中, 会对这些因素进行着重的关注与探讨。

本研究以 ASF 为研究对象, 应用 BART 结合相关的影响因素对中国 ASF 的风险区域进行评估。同时, 对 ASF 发生相关的主要影响因素及变化趋势进行分析。研究结果可为 BART 在动物传染病风险评估中的应用提供参考, 同时为了解中国非洲猪瘟发生主要风险区域及影响因素, 采取合理的预防和控制措施提供信息和建议。

## 4 结 论

1) 在影响非洲猪瘟 (African Swine Fever, ASF) 风险区分布的环境因素中, 归一化城市土地指数 ( $0.213 \pm 0.026$ ) 贡献性最高, 其次是归一化差分植被指数 ( $0.207 \pm 0.028$ ), 年平均气温 ( $0.199 \pm 0.025$ ), 家猪的分布 ( $0.194 \pm 0.025$ ) 和最冷季降水量 ( $0.187 \pm 0.026$ )。

2) 小型城市相比于其他类型的城市发生 ASF 的风险更高。ASF 的发生风险随着年平均气温, 归一化植被指数, 家猪的数量的升高而升高。最冷季降水量的升高会降低 ASF 发生的可能性。

3) ASF 的风险区域主要集中在中国的东部和西南部, 主要包括东北三省 (黑龙江、辽宁、吉林)、河北、山东、河南、山西、陕西、安徽, 江苏, 湖北、湖南、广西、广东、海南、贵州、重庆、四川、云南等的部分区域。

4) 预测地图中中国东南地区的不确定性较高, 鉴于影响 ASF 传播和发生的因素众多, 未来需要对此区域保持重点关注。

## [参 考 文 献]

- [1] 王君玮, 张玲, 王志亮, 等. 非洲猪瘟传入我国危害风险分析[J]. 中国动物检疫, 2009, 26(3): 63-66.  
Wang Junwei, Zhang Ling, Wang Zhiliang, et al. Risk analysis on hazard of the introduction of Africa swine fever to China[J]. Chinese Journal of Animal Health Inspection, 2009, 26(3): 63-66. (in Chinese with English abstract)
- [2] 张焰洪, 尹子凤, 周先文. 非洲猪瘟的流行传播特点及防控策略[J]. 现代畜牧科技, 2019(11): 68-69.  
Zhang Yanhong, Yin Zifeng, Zhou Xianwen. Epidemiological characteristics and prevention and control strategies of African swine fever[J]. Modern Animal Husbandry Science & Technology, 2019(11): 68-69. (in Chinese with English abstract)
- [3] 张睿, 黄旌童, 鲍晨沂, 等. 非洲猪瘟流行病学及其在中国扩散的因素分析[J]. 病毒学报, 2019, 35(3): 512-522.  
Zhang Rui, Huang Yitong, Bao Chenyi, et al. Epidemiology of African swine fever and analysis of risk factors of its spread in China: An overview[J]. Chinese Journal of Virology, 2019, 35(3): 512-522. (in Chinese with English abstract)
- [4] 陈腾, 张守峰, 周鑫韬, 等. 我国首次非洲猪瘟疫情的发现和流行分析[J]. 中国兽医学报, 2018, 38(9): 1831-1832.  
Chen Teng, Zhang Shoufeng, Zhou Xintao, et al. The discovery and epidemic analysis of the first African swine fever epidemic in China[J]. Chinese Journal of Veterinary Science, 2018, 38(9): 1831-1832. (in Chinese with English abstract)
- [5] Montgomery R. On a form of swine fever occurring in British East Africa(Kenya Colony)[J]. Journal of Comparative Pathology and Therapeutics, 1921, 34: 159-191.
- [6] 王凯, 范志新, 田昊伦, 等. 非洲猪瘟的流行概况及防控策略[J]. 中国兽医学报, 2019, 39(5): 1027-1034.  
Wang Kai, Fan Zhixin, Tian Haolun, et al. Epidemic situation and prevention strategies of African server[J]. Chinese Journal of Veterinary Science, 2019, 39(5): 1027-1034. (in Chinese with English abstract)
- [7] Geering W, Penrith M, Nyakahuma D. Manual on the preparation of African swine fever contingency plans[R]. Rome: Food and Agriculture Organization of the United Nations, 2001.
- [8] 戈胜强, 李金明, 任炜杰, 等. 非洲猪瘟在俄罗斯的流行与研究现状[J]. 微生物学通报, 2017, 44(12): 3067-3076.

- Ge Shengqiang, Li Jinming, Ren Weijie, et al. Prevalence and research progress of African swine fever in Russia[J]. Microbiology China, 2017, 44(12): 3067-3076. (in Chinese with English abstract)
- [9] 农业农村部新闻办公室. 农业农村部发布非洲猪瘟Ⅱ级疫情预警[EB/OL]. (2018-08-03)[2021-08-27]. [http://www.moa.gov.cn/xw/zwdt/201808/t20180803\\_6155300.htm](http://www.moa.gov.cn/xw/zwdt/201808/t20180803_6155300.htm).
- [10] 田爱爱, 张慧敏, 孙虎强, 等. 我国 2018—2020 年非洲猪瘟疫情数据分析研究[J]. 南方农机, 2021, 52(15): 24-26, 31.
- Tian Aiai, Zhang Huimin, Sun Huqiang, et al. Data analysis and research of African swine fever epidemic in China from 2018 to 2020[J]. China Southern Agricultural Machinery, 2021, 52(15): 24-26, 31. (in Chinese with English abstract)
- [11] 汪葆玥, 刘玉良, 马静, 等. 非洲猪瘟: 传染源和传播途径研究进展与分析[J]. 中国动物传染病学报, 2020, 28(5): 103-110.
- Wang Baoyue, Liu Yuliang, Ma Jing, et al. African swine fever: Recent progress on transmission source and routes[J]. Chinese Journal of Animal Infectious Diseases, 2020, 28(5): 103-110. (in Chinese with English abstract)
- [12] 张志, 康京丽, 李晓成. 非洲猪瘟的流行特征和传播路线[J]. 中国动物检疫, 2018, 35(11): 48-51.
- Zhang Zhi, Kang Jingli, Li Xiaocheng. The epidemiological characteristics of African swine fever and its transmission routes[J]. China Animal Health Inspection, 2018, 35(11): 48-51. (in Chinese with English abstract)
- [13] 卢易, 王烁, 易敬涵, 等. 基于 GIS 的中国非洲猪瘟疫情风险分析[J]. 中国兽医学报, 2019, 39(1): 8-13, 20.
- Lu Yi, Wang Shuo, Yi Jinghan, et al. Risk analysis of African swine fever in China based on GIS[J]. Chinese Journal of Veterinary Science, 2019, 39(1): 8-13, 20. (in Chinese with English abstract)
- [14] 任钰峰, 许志国, 赵怡雯, 等. 建立养殖区域内非洲猪瘟风险评估模型的必要性和可行性[J]. 兽医导刊, 2018(19): 24-25.
- Ren Yufeng, Xu Zhiguo, Zhao Yiwen, et al. Necessity and feasibility of establishing an African swine fever risk assessment model in breeding areas[J]. Veterinary Orientation, 2018(19): 24-25. (in Chinese with English abstract)
- [15] Bayot B, Sonnenholzner S, Ochoa X, et al. An online operational alert system for the early detection of shrimp epidemics at the regional level based on real-time production[J]. Aquaculture, 2008, 277(3/4): 164-173.
- [16] Ma J, Gao X, Liu B, et al. Peste des petits ruminants in China: Spatial risk analysis[J]. Transboundary and Emerging Diseases, 2019, 66(4): 1784-1788.
- [17] Wang H, Xiao J, Ouyang M, et al. Assessment of foot-and-mouth disease risk areas in mainland China based spatial multi-criteria decision analysis[J]. BMC Veterinary Research, 2021, 17(1): 1-12.
- [18] Assefa A, Bihon A, Tibebu A. Anthrax in the Amhara regional state of Ethiopia: spatiotemporal analysis and environmental suitability modeling with an ensemble approach[J]. Preventive Veterinary Medicine, 2020, 184: 105155.
- [19] Carlson C, Kracalik I, Ross N, et al. The global distribution of Bacillus anthracis and associated anthrax risk to humans, livestock and wildlife[J]. Nature Microbiology, 2019, 4(8): 1337-1343.
- [20] Messina J, Kraemer M, Brady O, et al. Mapping global environmental suitability for Zika virus[J]. eLife, 2016, 5: e15272.
- [21] Bhatt S, Gething P, Brady O, et al. The global distribution and burden of dengue[J]. Nature, 2013, 496(7446): 504-507.
- [22] Carlson C. Embarcadero: Species distribution modelling with Bayesian additive regression trees in R[J]. Methods in Ecology and Evolution, 2020, 11(7): 850-858.
- [23] Chipman H, George E, McCulloch R. BART: Bayesian additive regression trees[J]. The Annals of Applied Statistics, 2010, 4(1): 266-298.
- [24] 马奔, 马晓羽, 陈会波, 等. 环境因子对我国卵形硬蜱适生区分布的影响研究[J]. 中国血吸虫病防治杂志, 2021, 33(3): 281-286.
- Ma Ben, Ma Xiaoyu, Chen Huibo, et al. Effects of environmental factors on the distribution of suitable habitats of Ixodes ovatus in China[J]. Chinese Journal of Schistosomiasis Control, 2021, 33(3): 281-286. (in Chinese with English abstract)
- [25] 邓兴文. 基于 BART 算法的分类问题研究[D]. 广州: 华南理工大学, 2014.
- Deng Xingwen. Research of Classification Problem Based on BART Algorithm[D]. Guangzhou: South China University of Technology, 2014. (in Chinese with English abstract)
- [26] 路媛. 基于贝叶斯累加回归树和随机森林的癫痫个性化治疗策略研究[D]. 济南: 山东大学, 2020.
- Lu Yuan. Study on Personalized Treatment Strategy of Epilepsy Based on BART and Random Forest[D]. Jinan: Shandong University, 2020. (in Chinese with English abstract)
- [27] 赵红玉, 冯一平, 马晓天, 等. 贝叶斯累加回归树模型在高血压药物个性化疗效评价中的应用[J]. 中华疾病控制杂志, 2021, 25(6): 637-643.
- Zhao Hongyu, Feng Yiping, Ma Xiaotian, et al. Application of Bayesian additive regression tree model in the evaluation of individualized efficacy of hypertension drugs[J]. Chinese Journal of Disease Control & Prevention, 2021, 25(6): 637-643. (in Chinese with English abstract)
- [28] 印明辉. 贝叶斯累加回归树扩展研究及在生存分析中的模型构建[D]. 武汉: 华中科技大学, 2017.
- Yin Minghui. Extensions of Bayesian Additive Regression Trees and Modelling in Survival Analysis[D]. Wuhan: Huazhong University of Science and Technology, 2017. (in Chinese with English abstract)
- [29] Liang R, Lu Y, Qu X, et al. Prediction for global African swine fever outbreaks based on a combination of random forest algorithms and meteorological data[J]. Transboundary and Emerging Diseases, 2020, 67(2): 935-946.
- [30] Ma J, Chen H, Gao X, et al. African swine fever emerging in China: Distribution characteristics and high-risk areas[J]. Preventive Veterinary Medicine, 2020, 175: 104861.
- [31] 陈志杰, 王新生, 汪权方. 我国非洲猪瘟时空特征及影响因素[J]. 中国兽医学报, 2020, 40(12): 2457-2464.
- Chen Zhijie, Wang Xinsheng, Wang Quanfang, et al. Spatio-temporal characteristics and influencing factors of African swine fever epidemic in China[J]. Chinese Journal of Veterinary Science, 2020, 40(12): 2457-2464. (in Chinese with English abstract)
- [32] 石国宁, 张涛, 王无为. 中国非洲猪瘟疫情的时空演化特征及影响因素[J]. 干旱区资源与环境, 2020, 34(3): 137-142.
- Shi Guoning, Zhang Tao, Wang Wuwei, et al. Spatio-temporal evolution characteristics and causes of African swine fever in China[J]. Journal of Arid Land Resources and Environment, 2020, 34(3): 137-142. (in Chinese with English abstract)
- [33] Liu B, Gao X, Ma J, et al. Modeling the present and future distribution of arbovirus vectors Aedes aegypti and Aedes albopictus under climate change scenarios in Mainland China[J]. Science of The Total Environment, 2019, 664: 203-214.
- [34] Chen Q, Yin Y, Zhao R, et al. Incorporating local adaptation into species distribution modeling of Paeonia mairei, an endemic plant to China[J]. Frontiers in Plant Science, 2020, 10: 1717.

- [35] 朱增勇, 李梦希, 张学彪. 非洲猪瘟对中国生猪市场和产业发展影响分析[J]. 农业工程学报, 2019, 35(18): 205-210. Zhu Zengyong, Li Mengxi, Zhang Xuebiao, et al. Analysis on effects of African swine fever on China's pig market and industry development[J]. Transactions of The Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 205-210. (in Chinese with English abstract)
- [36] 周勋章, 李广东, 孟宪华, 等. 非洲猪瘟背景下不同规模养猪户生物安全行为及其影响因素[J]. 畜牧与兽医, 2020, 52(2): 133-141. Zhou Xunzhang, Li Guangdong, Meng Xianhua, et al. Biosafety behavior of farmers of different scale pig farms and the influencing factors of their behavior under the background of African swine fever[J]. Animal Husbandry & Veterinary Medicine, 2020, 52(2): 133-141. (in Chinese with English abstract)
- [37] 路平, 代德华, 李鹏, 等. 基层兽医机构防控非洲猪瘟效能评估与分析[J]. 中国动物检疫, 2020, 37(11): 31-39, 90. Lu Ping, Dai Dehua, Li Peng, et al. Performance evaluation and analysis on grassroots veterinary institutions for prevention and control of African swine fever[J]. China Animal Health Inspection, 2020, 37(11): 31-39, 90. (in Chinese with English abstract)
- [38] 王鑫, 何忠伟, 刘芳, 等. 中国非洲猪瘟疫情的时空演化分析[J]. 中国畜牧杂志, 2021, 57(9): 241-248. Wang Xin, He Zhongwei, Liu Fang, et al. Analysis on the spatio-temporal evolution of African swine fever in China[J]. Chinese Journal of Animal Science, 2021, 57(9): 241-248. (in Chinese with English abstract)

## Assessing the risk areas for African swine fever in China using Bayesian additive regression trees

Wang Haoran, Xiao Jianhua, Ouyang Maolin, Gao Hongyan, Liu Yuxin,  
Gao Li, Gao Xiang, Wang Hongbin<sup>\*</sup>

(Key Laboratory of the Provincial Education Department of Heilongjiang for Common Animal Disease Prevention and Treatment, College of Veterinary Medicine, Northeast Agricultural University, Harbin 155030, China)

**Abstract:** African Swine Fever (ASF) has been one type of acute, febrile, and highly contagious infectious disease during pig breeding. This infectious disease can be characterized by respiratory impairment and severe bleeding of the skin and internal organs. Four levels can be classified for the severity of the ASF disease, including the most acute, acute, subacute, and chronic. Once the infection by the strong strains can cause up to 100% morbidity and mortality in pigs. As such, the ASF has been mandated as the reporting animal disease by the World Organization for Animal Health (OIE), where China has been classified as a risk area for Class I animal disease. Given that there is no effective vaccine or treatment available so far, the prevention and control of ASF rely mainly on some protective measures, such as strict monitoring and health management. Therefore, it is a high demand for an effective risk assessment model to deal with the spread of ASF. This study aims to determine the risk areas of ASF in China, particularly for the decision-making on the risk assessment of the epidemic. A Bayesian Additive Regression Tree (BART) model was also constructed to assess the risk areas and influencing factors using the ASF outbreak data and 33 environmental factors from August 2018 to August 2021. Some buffers were set at every 5 km for the collected points of the ASF outbreak, in order to eliminate the possible bias from the clustering during modeling. One distribution point of two intersecting buffers was removed using overlay analysis, where there was only one ASF outbreak point per 5 km × 5 km grid. A Variance Inflation Factor (VIF) analysis was performed to initially eliminate the highly correlated variables. The high correlation among environmental factors was avoided for the high accuracy of prediction. Specifically, the factors with the VIF values greater than 10 were considered as the multi-collinearity for the removal in the VIF analysis. Subsequently, the environmental variables were verified, according to the importance level of the factors in the BART model. Finally, five optimal environmental variables were identified to establish the BART model for the assessment of the ASF risk areas, including the distribution of normalized urban land index, the normalized differential vegetation index, the mean annual temperature, the distribution of domestic pigs, and the coldest season precipitation. The results showed that: 1) There was the highest contribution of the normalized urban land index distribution ( $0.213 \pm 0.026$ ), followed by the normalized differential vegetation index ( $0.207 \pm 0.028$ ), the mean annual temperature ( $0.199 \pm 0.025$ ), the distribution of domestic pigs ( $0.194 \pm 0.025$ ), and the coldest season precipitation ( $0.187 \pm 0.026$ ). 2) There was a relatively higher ASF risk in the small cities, compared with the other types of cities. The ASF risk also increased with the average annual temperature, the normalized difference vegetation index, and the number of pigs. The high precipitation during the coldest season reduced the likelihood of ASF. 3) The risk areas of ASF were concentrated in eastern and southwestern China, including the parts of three northeastern provinces (Heilongjiang, Liaoning, and Jilin), Hebei, Shandong, Henan, Shanxi, Shaanxi, Anhui, Jiangsu, Hubei, Hunan, Guangxi, Guangdong, Hainan, Guizhou, Sichuan, Yunnan, and Chongqing. 4) In the risk map, there was a higher uncertainty in southeastern China. Therefore, much attention can be paid to this area in the future, due to various influencing factors for the spread and occurrence of ASF. Additionally, the Area Under the Curve (AUC) was 0.90 in the accuracy assessment of the model, indicating the high predictive accuracy of the risk map. The findings can provide a strong reference to determine the risk areas and influencing factors for the occurrence of ASF in China, thus formulating the prevention and control plans.

**Keywords:** disease; model; african swine fever; risk assessment; bayesian additive regression trees; influencing factors