引文格式：雷济华,杨波. 利用机器学习算法评价湖南省生态旅游适宜性[J]. 地理空间信息, 2026, 24(1): 52-56. doi:10. 3969/j.issn. 1672-4623.2026.01.011

利用机器学习算法评价湖南省生态旅游适宜性

雷济华，杨　波*
（湖南师范大学 地理科学学院，湖南 长沙 410081）

摘　要：将机器学习算法引入生态旅游适宜性评价，基于多源地理空间数据，利用不同模型在同一应用进行横向对比，并以湖南省为例进行验证。结果表明：①在常用机器学习算法中，LGB的效果更好，准确率为87.82%，精确率为90.38%，F_1-score为83.19%，AUC值为0.937 2，整体分类性能优秀；②生态旅游适宜性高度适宜区、中度适宜区、边际适宜区和不适宜区的占比分别为60.2%、20.1%、12.7%和7.0%，说明湖南省开发潜力较大、资源禀赋较多，未来可按照“两板块一区域”的策略重点发展高度适宜区；③特征重要度排名前五的特征为NDVI、人口密度、距水域距离、距居民点距离和地形起伏度，生态旅游发展时应着重考虑这些因素。

关键词：生态旅游；机器学习；适宜性评价；湖南省；多源地理空间数据
中图分类号：TU982　　**文献标志码：**B　　**文章编号：**1672-4623（2026）01-0052-05

Ecotourism Suitability Evaluation of Hunan Province Based on Machine Learning Algorithm

LEI Jihua, YANG Bo
(School of Geographical Sciences, Hunan Normal University, Changsha 410081, China)

Abstract: In this paper, we introduced machine learning algorithms into ecotourism suitability evaluation. Based on multi-source geospatial data, we compared different models in the same application to provide a side-by-side comparison, and taking Hunan Province for example, carried out validation. The conclusions show that ①among the commonly used machine learning algorithms, LGB is more effective, with accuracy of 87.82%, precision of 90.38%, F_1-score of 83.19%, and AUC value of 0.937 2, demonstrating excellent overall classification performance. ②The results of ecotourism suitability classification are as follows, highly suitable areas, moderately suitable areas, marginally suitable areas, and unsuitable areas account for 60.2%, 20.1%, 12.7%, and 7.0% of the total number of suitable areas, respectively, which indicates that Hunan Province has greater development potential and more resource endowments. In the future, we can refer to the strategy of “two plates and one regional belt” and focus on the development of highly suitable zones. ③The top five importance features are NDVI, population density, distance from water, distance from settlements, and topographic relief, which should be emphasized in ecotourism development.

Key words: ecotourism, machine learning, suitability evaluation, Hunan Province, multi-source geospatial data

全球生态环境问题日益凸显，人们对环境保护和可持续发展的关注与日俱增，生态旅游作为一种可持续的旅游形式，受到广泛关注。传统旅游模式给自然环境带来了过度压力，阻碍了旅游业的长期可持续发展。生态旅游适宜性评价旨在评判特定区域是否适合开发生态旅游地和开展生态旅游活动，从而为开发规划和区域可持续发展提供科学理论依据，对于推动生态旅游的可持续发展、提升旅游体验质量具有重要意义。诸多学者[1-4]提出了适宜性评价方法。随着数据科学和科学技术的发展，AHP等定性决策方法在处理大规模和复杂数据时存在局限性，部分学者[5-10]将完全定量的机器学习算法引入适宜性评价领域，以提供更精确的预测和决策支持。目前对于生态旅游适宜性评价的研究较少，且在选择机器学习模型时未说明理由，未进行多模型横向对比，无法确定不同模型在同一应用中的优劣。鉴于此，本文提出了一种生态旅游适宜性评价的建模框架，以确定和预测湖南省内的生态旅游适宜地。基于多源空间数据，结合目前常用的机器学习算法对湖南省的生态旅游适宜性展开研究，从自然环境、资源禀赋和社会经济3个方面构建特征矩阵，综合比对各种指标以选择最优模型，计算适宜性结果，并对影响区域适宜性的因素进行深入分析，为湖南省生态旅游开发和可持续发展提供支持和建议。

项目来源：国家自然科学基金资助项目（41171342）；湖南省教育厅重点资助项目（17A127）。
第一作者简介：雷济华（2001—），硕士，研究方向为机器学习及生态旅游，E-mail：2748757031@qq.com。
通信作者：杨波（1974—），博士，教授，研究方向为资源环境遥感与旅游大数据，E-mail：yb@hunnu.edu.cn。

1　研究区概况与数据来源

湖南省位于我国中部、长江中游地区，地理范围为24°38′~30°08′N、108°47′~114°15′E；属亚热带季风气候，四季分明、光热充足、降水丰沛、雨热同期，气候条件优越，四季变化较明显，适宜人居和农作物、绿色植物生长；地貌类型以山地、丘陵居多，总面积为21.18万km^2。湖南省具有得天独厚的地理位置、优越的气候条件、丰富的自然资源和悠久的历史文化，为生态旅游的发展奠定了坚实基础。

本文选取多源异构地理空间数据进行生态旅游适宜性分析，具体数据见表1。

表1　数据来源

数据类型	数据来源	数据年份
POI数据	高德地图API数据开放接口	2022
土地利用数据	ESRI 10 m土地覆盖数据（https://livingatlas.arcgis.com）	2022
生态系统服务价值数据	中国陆地生态系统服务价值空间分布数据集(http://www.resdc.cn/DOI)	2020
行政区划数据	中国科学院资源环境科学数据中心（http://www.resdc.cn/）	2022
DEM数据	STRM数字高程数据库（http://srtm.csi.cgiar.org）	2019
统计数据	湖南省统计局（https://tjj.hunan.gov.cn/）	2022
路网数据	OSM(https://www.openstreetmap.org)	2022
遥感数据	欧空局（https://dataspace.copernicus.eu/）	2022
气象数据	国家地球系统科学数据中心-黄土高原分中心(http://loess.geodata.cn)	2022
人口数据	WorldPop人口栅格数据集(https://www.worldpop.org)	2020
河流水系数据	全国地理信息资源目录服务系统（http://www.webmap.cn）	2020

2　研究方法与技术路线

2.1　理论分析

机器学习是一种强大的数据分析和预测工具，理论基础包括监督学习、无监督学习和特征工程等[11]。机器学习的研究思路与指标赋值方法相似，都是通过建立关系模型进行预测，因此将机器学习算法引入适宜性评价具有一定的可行性。适宜性评价结果可被映射为机器学习中已知样本的标签或未知数据的预测分类结果，而生态旅游适宜性影响因子则被映射为机器学习中的特征变量。这种映射关系为机器学习算法在生态旅游适宜性评价中的应用提供了基础。

2.2　机器学习算法

RF算法由Leo B于2001年提出。其核心思想是在bagging过程中首先对训练数据子集进行随机抽样[12]；再通过有放回抽样重复抽取同一数据；最后在分类和回归问题中通过投票和平均来组合多个决策树进行预测。ADB算法则是基于boosting思想，首先对所有样本赋予相同的权重；然后根据当前误差率最小的基分类器计算并更新样本权值，重复训练不断更新迭代直至满足条件，得到分类模型[13]。XGB由Chen等[14]于2016年提出，优势为在所有场景中的可伸缩性。其基本原理是通过迭代计算每个弱分类器形成一个强分类器，并对整个目标函数进行优化；再累加所有决策树的预测结果，构建一个增强的决策树[10]。轻量梯度提升机（LGB）是对GBDT的优化，通过保留梯度较大的样本来加速训练过程。该算法包括基于梯度的单边采样和独占特征捆绑两种新技术，分别用于处理大量的数据实例和特征[15]。LGB可得到更高的准确率，但内部结构较复杂，可解释能力较差。SVM是一种基于统计学习理论的机器学习算法，其目标是找到一个由支持向量确定的超平面[16]，能在理论上实现对线性可分类数据的最优分类。

2.3　技术路线与特征矩阵构建

本文利用经典机器学习算法对湖南省进行生态旅游适宜性评价。具体技术路线包括：①首先获取各生态旅游影响因子数据，并统一投影至同一坐标系，再利用研究区矢量边界进行掩膜提取得到研究区范围，并将特征因子栅格重采样为100 m×100 m分辨率，然后将各特征因子栅格图导入Python中作为机器学习模型的特征因子；②按照1∶2的比例选取生态旅游景点正、负样本点，通过样本点对应的特征因子值训练各机器学习模型，并选出最优模型；③将湖南省特征矩阵导入训练后的各机器学习模型进行分类预测，得到每个格网的适宜性概率以及各模型的特征因子重要性排序。

特征矩阵主要从自然环境、资源禀赋和社会经济3个方面选取特征因子。自然生态环境是生态旅游的前提条件，限制着旅游开发方向与承载力水平[17]。资源禀赋是一个地区进行生态旅游的基础，制约着生态旅游的发展[18]。社会经济的发展程度对生态旅游的开发建设具有决定作用，是影响旅游景区空间格局演化的主要推动力[19]。湖南省生态旅游适宜性特征矩阵的特征因子见表2。

3　研究结果与分析

3.1　模型评估

本文采用混淆矩阵、F_1-score和ROC曲线评估模型性能。混淆矩阵是一种将模型预测结果与真实情况

表2 机器学习特征因子

影响因素	特征选取	计算方法
自然环境	NDVI	NDVI=(NIR−R)/(NIR+R)， NIR为近红外波段， R为红光波段
	地形起伏度/m	$R=H_{max}-H_{min}$， H_{max}、H_{min}为单位面积内最大、最小高程
	坡度/(°)	$I=(H/L)\times100\%$， H为高程， L为水平距离
	坡向/(°)	在ArcGIS中通过坡度计算工具处理得到
	高程/m	矢量底图掩膜提取DEM
	距水域距离/m	在ArcGIS中采用欧氏距离工具得到
	水系密度/ km·km^{-2}	$P=L/S$， L为水系长度， S为区域面积
	平均气温/0.1 ℃	提取多年平均气温栅格值，计算平均气温，插值至30 m分辨率，再经掩膜提取
	年总降水量/0.1mm	提取每年降水量栅格值，计算年总降水量，插值至100 m分辨率，再经掩膜提取
资源禀赋	土地利用类型	矢量底图掩膜提取地表覆被状况
	生物多样性价值/(万元/km^2)	从已有数据集中获取
	美学景观价值/(万元/km^2)	从已有数据集中获取
	生态旅游资源密度/(个/km^2)	在ArcGIS中采用密度计算
社会经济	路网密度/km·km^{-2}	$K=(L/S)\times100\%$， L为道路长度
	人口密度	经矢量底图掩膜提取
	GDP/万元	在ArcGIS中采用密度计算
	距道路距离/m	在ArcGIS中采用欧氏距离工具得到
	距居民点距离/m	在ArcGIS中采用欧氏距离工具得到
	距景点距离/m	在ArcGIS中采用欧氏距离工具得到

进行比较的表格，可直观反映模型在各类别上的表现。本文选择准确率、精确率、召回率和F_1−score来评估模型性能。精确率是指模型预测的正类别中有多少是真正的正类别，召回率是指真正的正类别中有多少被模型预测为正类别，而F_1−score是精确率和召回率的调和平均值，值越接近1，说明模型越稳定。ROC曲线也是衡量二分类模型性能的重要指标。AUC是ROC曲线下方的面积，反映模型预测的正、负样本之间的排列顺序的好坏程度。AUC的取值范围为0~1，值越大表示模型性能越好。本文将AUC值分为5类，即不及格（0.5~0.6）、差（0.6~0.7）、一般（0.7~0.8）、好（0.8~0.9）和优（0.9~1）[20]。

为使模型更加可靠，本文采用十折交叉验证方法，每次将其中一份作为训练集，其余9份作为验证集，既可避免小数据样本的过拟合，又可通过网格搜索遍历参数组合选择最优参数。通过多次迭代对模型进行评估和验证，减少了因某次划分导致的模型性能过度依赖特定的训练集或测试集，提高了模型的泛化评估能力，且在最大程度利用数据的前提下降低了过拟合风险。本文比较了RF、ADB、XGB、LGB和SVM五种机器学习算法，各机器学习算法需要统一设置随机种子，结果见图1、2和表3，可以看出，LGB和RF的AUC值分别为0.937 2和0.941 0，说明这两个分类器可以准确预测生态旅游的区域适宜性；LGB和RF的指标得分相差无几，结合图1可知，LGB在雷达图中面积明显大于RF，说明LGB的综合性能优于RF。综上分析，本文选择LGB作为生态旅游适宜性评价的判断模型。

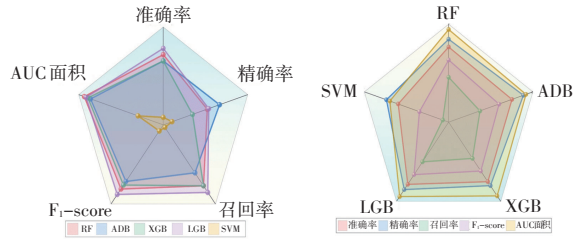
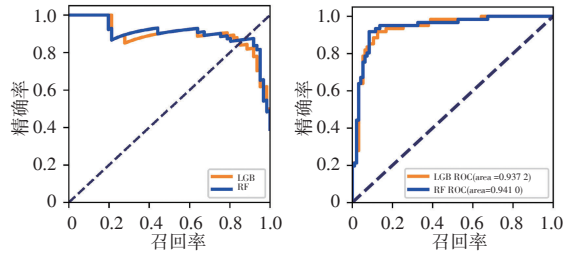


图1 机器学习模型性能比较



P−R曲线 ROC曲线

图2 LGB与RF的ROC曲线和P−R曲线对比

表3 机器学习分类器十折交叉验证性能比较

分类器	准确率/%	精确率/%	召回率/%	F_1−score/%	AUC面积
RF	87.18	90.20	75.41	82.14	0.941 0
ADB	86.54	91.67	72.13	80.73	0.929 1
XGB	86.54	88.46	75.41	81.42	0.933 7
LGB	87.82	90.38	77.05	83.19	0.937 2
SVM	80.77	86.05	60.66	71.15	0.844 7

模型返回的特征重要性排序见图3，可以看出，湖南省生态旅游适宜性的主要影响因子为NDVI、人口密度、距水域距离、距居民点距离和地形起伏度。对于生态旅游而言，较高的NDVI意味着更好的环境条件，是将区域设为生态旅游区的基本条件；人口密度影响着旅游市场的消费规模和生态旅游的开发强度，适当的人口密度有助于增加消费规模；距水域越近，有利于缓解温差，构造适宜的环境，提供丰富的生态体验和休闲娱乐活动，增加生态旅游的吸引力和可持续发展的潜力；距居民点距离与人口密度相辅相成，远离居民点的地区生态环境相对原始自然，而与居民点距离近的地区则能提供更好的物资补给和区位依托；地形起伏度较大的地区通常具有多样化的地貌景观和丰富的生物多样性，能丰富生态旅游的内容和体验。这些特征反映了自然环境、资源禀赋和社会经济对生态旅游开发的重要影响。

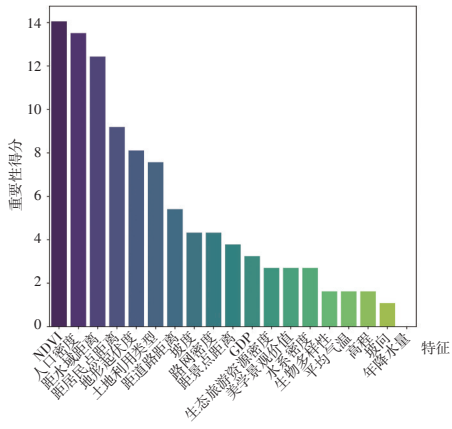


图3　LGB模型特征重要性

3.2　生态旅游适宜性分析

本文通过训练后的LGB模型对湖南省进行生态旅游适宜性预测，输出各格网的生态旅游适宜性概率值。首先利用Matlab将模型预测的各像元概率值进行栅格转换导入ArcGIS Pro中；再采用自然断点法[21-22]将概率分布分为4类，即高度适宜区（0.85~1）、中度适宜（0.59~0.85）、边际适宜区（0.37~0.59）和不适宜区（0~0.37）。

1）高度适宜区的面积为12.76万km^2，占比为60.2%。高度适宜区主要分布在湖南省西部的湘西土家族苗族自治州、张家界市和怀化市的大部分地区；南部的郴州市中部市区和东部，永州市中部和南部，株洲市南部；东北部的岳阳市东部，长沙市东部的浏阳市，株洲市茶陵县、禄口区。从地形形貌上看，高度适宜区主要分布在山地和水系周边地区。本文研究的适宜性范围仅限于陆地部分，水域的生态旅游适宜性默认存在于水域周边地区。西部高度适宜区主要沿武陵山和雪峰山两大山系分布，南部高度适宜区分布于阳明山和南岭地带，东部高度适宜区主要包括罗霄山脉、幕阜山和连云山地区，中部高度适宜区分布较少，集中在湘中丘陵地带。总体而言，高度适宜区具有植被茂盛、气候宜人、降水充沛、环境承载力强等特点，在开发生态旅游时应与环境保护理念相结合，将其视为生态旅游重点发展区域，加速旅游基础设施建设，积极响应生态旅游促进乡村振兴政策，推动生态旅游与环境保护理念的有机融合。

2）中度适宜区的面积为4.25万km^2，占比为20.1%。中度适宜区围绕高度适宜区以缓冲区样式分布，且与边际适宜区和不适宜区交错分布。中度适宜区主要分布于湖南省中部，包括围绕衡山周围部分以及阳明山和南岭的中间部分。在行政区划上，中度适宜区主要分布于常德市中部、益阳市北部、岳阳市中西部、湘潭市及其周围、邵阳市中东部、衡阳市、株洲市中西部、永州市北部和中部、郴州市中部。中度适宜区的植被覆盖度较高，生态旅游资源相对丰富，同时具有便利的交通条件和城镇依托，但相较于高度适宜区存在地形地貌复杂、生态系统脆弱等自然因素限制。因此，在中度适宜区的生态旅游开发中，政府和相关机构应加强对资源的保护和管理，推动生态旅游产业的可持续发展，最大程度地发挥区域旅游潜力，实现生态旅游与经济发展的双赢。

3）边际适宜区的面积为2.7万km^2，占比为12.7%。边际适宜区主要分布于湖南省北部、中部，与中度适宜区和不适宜区交错分布。该区域的植物覆盖度较低、气温较高、降水较少，生态旅游资源较少，分布分散，并不完全适合生态旅游开发。边际适宜区仍具有一定的旅游潜力，政府和相关部门应加强对其保护和管理，坚持保护优先，采取科学合理的措施，将其作为高度适宜区和中度适宜区的缓冲区或一些特定的旅游活动的补充和延伸。

4）不适宜区的面积为1.48万km^2，占比为7.0%。不适宜区主要分布于湖南省北部和中部，且分布极为分散和稀疏，包括岳阳市的洞庭湖生态旅游保护区、岳阳楼区，常德市的武陵区周围，长沙市中部，湘潭市岳塘区和雨湖区，株洲市的西北角，邵阳市双清区和衡阳市市区，娄底市娄星区和新化县。这些区域主要为城市化程度较高的建成区、生态环境保护区，或者存在一定的生态环境问题，如温度较高、生态旅游资源稀薄或交通不便等。政府和相关部门应加强对其环境检测和治理，积极推动生态恢复和修复工作，为可持续发展创造更加良好的生态环境基础。

根据湖南省生态旅游适宜区分布情况和《湖南省“十四五”旅游业发展规划》，确定湖南省各地区生态旅游发展定位。本文提出“两板块一区域”的生态旅游发展策略：以张家界市、怀化市、湘西土家族自治区、邵阳市西部和益阳市西部5个区域为核心构成的武陵—雪峰生态旅游板块，包含丰富的自然景观和人文景观。其中，张家界市的武陵源风景名胜区、天门山国家森林公园等自然景观以及湘西土家族苗族自治州的八面山、凤凰古城等人文景观，是武陵山生态旅游区的重要组成部分。怀化市的万佛山、邵阳市城步苗族自治县和罗溪国家森林湿地公园、娄底市新宁县的紫鹊界梯田、益阳市安化县和桃江县的万福山等景点，则为雪峰山生态旅游区增添了独特魅力。将武陵山生态旅游区和雪峰山生态旅游区相结合，不仅能共享基础设施、拓展旅游线路、提升景区知名度、保护环境，还能促进湖南省西部地区的经济和生态旅游产

业发展。以永州市、郴州市和株洲市构成的阳明山—南岭—罗霄山脉生态旅游板块，拥有丰富的自然和人文景观资源。永州市的阳明山国家森林湿地公园，江华瑶族自治县的瑶族古镇，郴州市的东湖、通天山以及株洲市茶陵县的云阳山、炎陵县的炎陵桃花岛，是该板块的重要景点。第二届湖南旅游发展大会的成功举办后，郴州市的旅游接待人次和旅游收入均居全省第二，增速全省第一。《湖南省“十四五”旅游业发展规划》明确指出，将以郴州为湘南门户，全力推动永州打造文化生态旅游名城。因此，阳明山—南岭—罗霄山脉生态旅游板块将作为重点发展区域，利用其丰富的自然和人文资源，大力发展生态旅游事业。在发展过程中，将依托生态旅游高度适宜区的发展，构建良好的交通网络和基础设施，实行高度适宜区优先发展，其他适宜区联动发展的良性循环，有助于提升当地旅游业的整体水平，促进经济社会的可持续发展。以岳阳市东部、长沙市东部构成幕阜山—连云山生态旅游区作为湖南省生态旅游的次要区域。临湘市的五尖山国家森林湿地公园、岳阳县的大云山国家森林公园、平江县的天岳幕阜山国家森林湿地公园以及浏阳市的大围山国家森林湿地公园，是其重要景点。

“两板块一区域”的发展策略为湖南省生态旅游开发确立了方向和目标，为政府部门制定区域生态旅游发展政策提供了科学依据。未来，可将湘西板块作为省域生态旅游的中心，利用张家界和湘西地区的传统优势推动生态旅游发展；同时以湘南板块的郴州市和永州市为发展重心，以东北部的幕阜山—连云山生态旅游区为长沙的生态旅游依托地，有助于促进湖南省生态旅游区的开发，也为中部邻近省份的生态旅游发展提供了理论指导和参考。通过这种方式，可以推动我国生态旅游的良性发展，实现生态保护和经济发展的双赢局面。

4 结 语

本文以湖南省为研究范围，结合多源空间数据，运用机器学习思维，采用经典机器学习算法进行生态旅游适宜性评价的实证研究，得出以下结论：

1）基于机器学习算法的生态旅游适宜性评价模型能在已有数据样本中识别出潜在模式或规律，避免了人为因素的介入，为选择适用于省域生态旅游适宜性评价方法提供了新思路和方案。

2）通过对比发现，LGB算法在湖南省生态旅游适宜性评价中的表现较好，准确率为87.82%，精确率为90.38%，召回率为77.05%，F_1-score为83.19%，AUC面积为0.937 2，整体性能优异，能准确反映湖南省的生态旅游适宜性情况。

3）湖南省生态旅游开发潜力巨大，高度适宜区面积占比为60.2%，但从四周到中部呈逐渐下降趋势，表明湖南省拥有丰富的自然、人文资源，有望成为重要的生态旅游目的地。生态旅游开发可参考“两板块一区域”策略，重点关注高度适宜区的开发。

4）对模型进行特征重要度排序发现，NDVI、人口密度、距水域距离、距居民点距离和地形起伏度对生态旅游具有较大影响。在实际应用中，应优先考虑这些因素。

然而，本文在理论、数据和应用层面存在一些不足，具体总结如下：在理论层面，以数据驱动代替了主观权重设置，但特征因子的构建并不系统，未来可综合多个学科的特征因子，总结出一个系统的生态旅游系数；在数据层面，缺少生态保护区、永久基本农田、饮用水源保护地等禁止开发区的矢量边界数据，未来可探索将更多环境保护区域和其他禁止开发地区的空间数据纳入模型中，以更全面地评估生态旅游适宜性，并与当地的乡村振兴和生态旅游发展战略相协调；在应用层面，本文中水域内部缺乏正样本点，导致模型在评估水域适宜性时存在偏差，未来可对重要的水域点位进行调整，确保模型能更准确地反映水域内部的适宜性情况，包括将重要水域点位调整至水域中间位置，以提高生态旅游适宜性评价的准确性。

参考文献

[1] Bunruamkaew K, Murayam Y. Site Suitability Evaluation for Ecotourism Using GIS & AHP: A Case Study of Surat Thani Province, Thailand[J]. Procedia-social and Behavioral Sciences, 2011, 21:269-278

[2] Chaudhary S, Kumar A, Pramanik M, et al. Land Evaluation and Sustainable Development of Ecotourism in the Garhwal Himalayan Region Using Geospatial Technology and Analytical Hierarchy Process[J]. Environment, Development and Sustainability, 2022, 25:1-42

[3] 石丹，刘晓霞. 基于模糊层次分析法的森林生态旅游景区开发潜力评价:以吉林省4个典型森林生态景区为例[J]. 桂林理工大学学报，2021，41(2):465-470

[4] 石丹，杨慧. 基于多层次灰色方法的生态旅游发展潜力评价及实证研究[J]. 中国农业资源与区划，2019，40(2):40-48

[5] Safitri S, Wikantika K, Riqqi A, et al. Spatial Allocation Based on Physiological Needs and Land Suitability Using the Combination of Ecological Footprint and SVM (Case

（下转第61页）