

基于随机森林算法的农村居民点用地适宜性评价*

——以重庆市合川区为例

李林芝¹, 陈星宇¹, 张虹¹, 孙德亮¹, 文海家³, 陈爽³

(1. 重庆师范大学 GIS 应用研究重庆市高校重点实验室, 重庆 401331;

2. 重庆大学山地城镇建设与新技术教育部重点实验室, 重庆 400045; 3. 重庆市规划设计研究院, 重庆 401147)

摘要:【目的】为合理布局农村居民点,推进全域乡村规划发展,科学评价农村居民点用地适宜性提供参考。【方法】选取区位条件、生产属性、自然条件、社会经济等4类20个因子,构建农村居民点用地适宜性评价指标体系,选取重庆市合川区1000个农村居民点作为正样本,并按1:1的比例随机选取1000个非农村居民点作为负样本,基于随机森林算法,构建农村居民点用地适宜性评价模型,利用自然断点法将研究区内农村居民点用地适宜性等级划分为适宜、一般适宜、不适宜共3类,根据评价结果提出合川区农村居民点用地适宜性分区的优化方案。【结果】1) 验证模型的有效性,得到训练数据、测试数据、全体数据的曲线下面积值分别为0.999、0.996和0.944,模型测试集最高精度为0.991,表明模型预测准确性高且稳定;2) 评价因子中坡度、距农村道路距离和地形起伏度对农村居民点用地适宜性评价结果的影响较大;3) 适宜区位于合川区中部、北部地势平缓的丘陵盆地,建议保留发展;一般适宜区多位于西南、东南部,以及距嘉陵江、涪江1000m范围外,建议内部改造;不适宜区位于南部、东南部岭谷,分布于三江两侧和华蓥山支脉云雾山一带,建议搬迁撤并。【结论】利用随机森林模型进行农村居民点用地适宜性评价的精度较高,可为农村居民点管控提供参考。

关键词: 随机森林算法; 因子分类; 用地适宜性评价; 农村居民点; 合川

中图分类号: X24; K90

文献标志码: A

文章编号: 1672-6693(2022)03-0093-12

在乡村振兴战略的背景下,对农村居民点用地适宜性进行评价,能有效识别预测农村居民点的潜力空间,为村镇用地的选择与布局提供定量评价依据^[1]。2019年,自然资源部发布的《关于加强村庄规划促进乡村振兴的通知》中明确提出需要统筹农村住房布局,农村居民点用地适宜性问题成为农村土地资源整合的重要内容;2022年,《中共中央、国务院关于做好2022年全面推进乡村振兴重点工作的意见》中指出要扎实稳妥推进乡村建设;云南、四川、北京等省市先后出台了村庄规划的文件,明确了农村居民点布局的条件。农村居民点用地适宜性评价是农村居民点优化的重要依据,科学合理的评价方法与手段,对推动乡村振兴战略实施具有重大的现实意义^[2]。

近些年来,有关农村居民点用地适宜性评价的研究主要集中在研究方法、研究内容、指标体系等方面。如曲衍波等人^[3]选取生态、生产、生活共3个方面的指标,利用ArcGIS,构建生态适宜度模型,划定北京市平谷区的农村居民点为高、中、低、不适宜等4个等级并提出优化方案。梁照凤等人^[4]选择自然、经济和社会因子,构建农村居民点用地适宜性评价体系,利用ArcGIS,确定农村居民点用地适宜性级别。白雪等人^[5]选取自然环境、生产生活和社会经济等3个方面共12个指标构建适宜性评价体系,使用均方差决策法和TOPSIS模型评价山西农谷区农村居民点布局适宜性并提出优化策略。传统的研究方法多基于ArcGIS,利用层次分析法、多因素综合评价、聚类^[6-10]等方法评估;指标体系多选取自然、生态、社会经济、交通等因子^[11-13]。通过上述研究发现,农村居民点用地适宜性评价对社会经济因素考虑不足,指标体系的构建和权重值受到人为主观因素干扰,评价指标权重赋值不具有普适性^[14],农村居民点用地适宜性评价研究亟待从指标体系和研究方法上进行优化。随着人工智能的发展,越来越多领域运用机器学习进行研究。Xu等人^[15]利用机器学习判断山区中农村居民点存在的较高可

* 收稿日期:2021-10-14 修回日期:2022-01-07 网络出版时间:2022-05-18 15:24

资助项目:国家自然科学基金(No. 41807498; No. 42071217);重庆市自然科学基金(No. cstc2020jcyj-msxmX0841);重庆市人文社科重点研究基地项目(No. 19JD030);重庆市教育委员会科学技术研究项目(No. KJQN202000525);国家社会科学基金(No. 19XGL027)

第一作者简介:李林芝,女,讲师,博士研究生,研究方向为城镇规划、历史文化遗产保护,E-mail: sclilinzhi@163.com;通信作者:张虹,女,副教授,博士,E-mail: zh_ angh@cqnu.edu.cn

网络出版地址: <https://kns.cnki.net/kcms/detail/50.1165.N.20220517.0943.004.html>

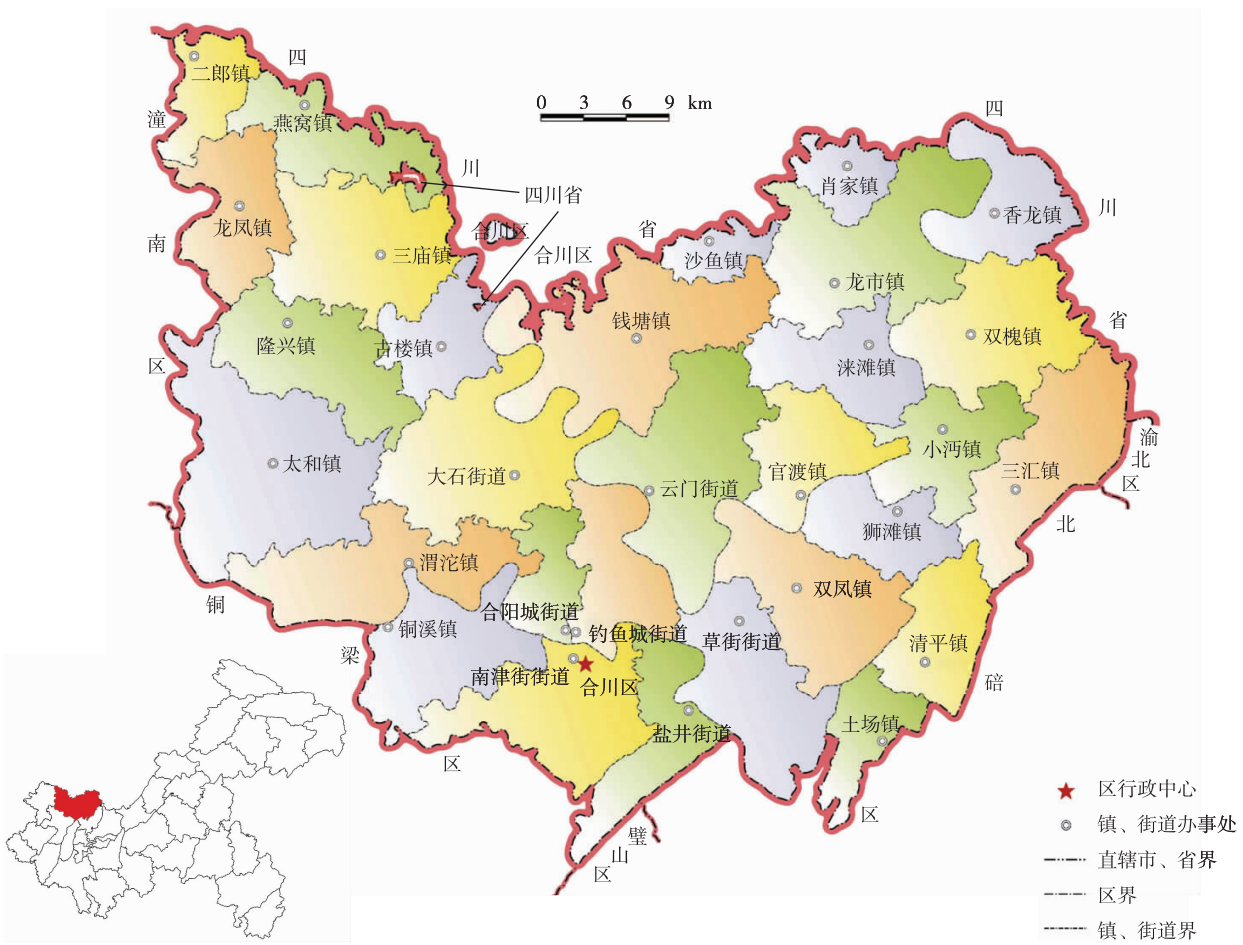
能被遗弃的土地,有效预测废弃土地发生的可能性。徐枫等人^[16]应用随机森林算法探讨农村居民点的适宜性,实现大规模未知数据的预测与评估。以上研究均证实机器学习对农村居民点用地适宜性研究具有较好的适用性。机器学习具有精度高、速度快等特点,且能够在不完整、大量数据集上有效运行,能科学评价因子的相对权重,减少人工误判^[17]。机器学习在评估、验证和预测方面较传统方法更具有优势。

指标体系中虽然涉及到社会经济因素,但评价因子无法体现劳动力和居民点规模对农村居民点用地适宜性的影响。基于此,本文以重庆市合川区农村居民点图斑为研究单元,选取区位条件、生产属性、自然条件、社会经济等 4 类共 20 个因子,利用随机森林算法构建农村居民点用地适宜性评价模型,将研究区内农村居民点用地适宜性等级利用自然断点方法将结果划分为适宜、一般适宜、不适宜 3 个等级,根据评价结果提出合川区农村居民点用地适宜性分区的优化方案。

1 研究区概况与数据来源

1.1 研究区概况

合川区位于重庆市西北部,地理范围在东经 105°58'37"~106°40'37"、北纬 29°51'02"~30°22'24"之间(图 1),嘉陵江、渠江、涪江三江横贯全境。合川区地处中丘陵和重庆平行岭谷的交接地带,区域内的丘陵占全区总面积的 90%,区域地貌呈南面低,东、北、西三面高。合川区幅员面积 2 344.07 km²,辖 23 个镇、7 个街道办事处、322 个村、97 个居民委员会,全区户籍总人口为 151.78 万人,其中主城区人口 56.61 万人。2021 年合川区的城镇化率为 70.59%,在全市区县中排名第 11 位,已进入城镇化“S”曲线的平缓的成熟阶段,城镇常住居民可支配收入为 37 927 元,在全市区县中排名第 20 位。农村常住居民可支配的收入为 18 850 元,在全市区县中排名第 13 位^[18]。



合川区在重庆市的位置

注:底图来源于重庆市勘测院(重庆市地图编制中心),审图号:渝 S(2016)029 号,下同

图 1 合川区地理位置及行政区划图

Fig. 1 Geographical location and administrative division map of Hechuan district

1.2 数据来源

本研究选取的数据涵盖土地利用数据、社会经济统计数据、遥感影像数据、数字高程数据、规划数据和兴趣点数据等,具体名称及来源见表1。其中:根据重庆市规划和自然资源局获取的农村居民点,提取得到38 474个农村居民点矢量图斑,再根据重庆市人均居民点用地面积和农村常住人口数量及农村居民点的分布情况,筛选出面积大于1 500 m²的29 842个农村居民点作为研究对象。

表1 农村居民点数据及数据来源
Tab.1 Rural settlement data and data sources

数据名称	数据来源	类型	精度
地质资料	国家地质资料数据中心	栅格	30 m×30 m
土地利用	地理国情监测云平台	矢量	1:100 000
行政区划	重庆市规划和自然资源局	矢量	1:100 000
河网	重庆市水利局	矢量	1:100 000
道路	中国科学院资源环境科学数据中心	矢量	1:100 000
卫星影像	地理空间数据云平台	栅格	30 m×30 m
POI数据	网络爬虫	数据表	
城镇建成区	地理国情监测云平台	栅格	30 m×30 m
社会经济	合川区统计年鉴	数据表	
农村居民点	重庆市规划和自然资源局	矢量	

2 评价指标体系建立

参考《国家乡村振兴战略规划(2018—2022年)》,借鉴农村居民点用地适宜性评价指标体系的已有经验与案例^[19-22]。从区位条件(距其他基础设施距离、距中心城区距离、距乡镇中心距离、距耕地距离)、生产属性(人均耕地面积、人均农田面积)、自然条件(距河流距离、地形起伏度、高程、坡度、坡向、降雨量、归一化植被指数(NDVI)、距县级以上道路距离、距农村道路距离)、社会经济(居民点规模指数、常住人口数、人口户数、人均GDP、劳动力指数)4个类型20种影响因子(变量)构建评价指标体系(表2),评价研究区农村居民点用地适宜性。其中:居民点规模指数为居民点面积与土地面积之比,劳动力指数为男人人数与劳动力人数之比。

农村居民点用地适宜性评价受到不同因素的影响,生产属性与区位条件决定了农村居民点的地理优势和生产优势。自然条件如高程、坡度、坡向等影响了农村居民点的分布和用地规模。农村居民以务农为主,耕地、农田、居民点的空间分布密度和耕地的服务范围与农村居民点之间存在较高的空间邻接性和体量差异,故采用人均耕地面积、人均农田面积等因子表征农村居民点的特征因子;区位条件与距中心城区、乡镇中心、耕地距离呈正比,受到各行政中心的影响;合川区高山阻隔,三江交汇的山水资源环境影响了道路的可达性。道路可达性越低,时间成本越高,农村居民点用地适宜性越差,保留必要性越低^[23];人口户数、常住人口数反映了人口的集聚度;人均GDP、劳动力指数和居民点规模指数体现了当地的发展水平。

3 随机森林算法

随机森林算法是一种由Breiman等人提出的机器学习算法,可做聚类、判别、回归、生存分析以及评估变量^[24-25]。基于训练样本和特征集的决策树是随机森林算法的基本分类器,决策树结合原始数据,构建不同子数据集的决策树模型 $y_1(x), \dots, y_k(x)$,根据投票结果筛选出精确度最高的一组构建模型,从而判断和预测整个区域的研究结果。具体公式为:

$$H(x) = \arg \max_Z \sum_{i=1}^k I(h_i(x) = Z),$$

其中: $H(x)$ 表示为模型输出结果, $h_i(x)$ 表示单棵决策树模型, Z 表示输出变量, $h_i(x) = Z$ 是使用变量 x 的第 i 颗树对变量 Z 的预测, $I(\cdot)$ 是一个显式函数。随机森林的分类方式主要是通过投票方式确定。

表 2 农村居民点用地适宜性评价的指标体系

Tab.2 An index system for evaluation of land use suitability in rural residential areas

准则层	指标层	分类阈值或标准
区位条件	距其他基础设施距离(POI 数据)/km	1. <0.5;2. 0.5~1;3. >1~2;4. >2~3;5. >3~10;6. >10
	距中心城区距离/km	1. <3;2. 3~5;3. >5~8;4. >8~12;5. >12
	距乡镇中心距离/km	1. <1;2. 1~2.5;3. >2.5~6;4. >4~6;5. >6
	距耕地距离/km	1. 0~0.2;2. >0.2~0.4;3. >0.4~0.6;4. >0.6~1.0;5. >1.0
生产属性	人均耕地面积/(km ² ·人 ⁻¹)	1. <1.5;2. 1.5~2;3. >2~2.4;4. >2.4~2.8;5. >2.8
	人均农田面积/(km ² ·人 ⁻¹)	1. <0.8;2. 0.8~0.95;3. >0.95~1.2;4. >1.2~1.5;5. >1.5
自然条件	距河流距离/km	1. 0~0.5;2. >0.5~1.0;3. >1.0~1.5;4. >1.5~2.0;5. >2.0
	地形起伏度/m	1. <10;2. 10~20;3. >20~30;4. >30~50;5. >50
	高程/m	1. <250;2. 250~350;3. >350~450;4. >450~700;5. >700
	坡度/(°)	1. <3;2. 3~8;3. >8~15;4. >15~25;5. >25
	坡向	1. 平地、南;2. 东南、西南;3. 东、西;4. 东北、西北;5. 北坡
	降雨量/mm	1. <1 131;2. 1 131~1 160;3. >1 160~1 186;4. >1 186~1 210; 5. >1 210~1 233;6. >1 233~1 266;7. >1 266~1 316;8. >1 316
	NDVI	1. >0.30;2. >0.20~0.30;3. >0.15~0.20;4. 0.10~0.15;5. <0.10
	距县级以上道路距离/km	1. 0~1.0;2. >1.0~3.0;3. >3.0~5.0;4. >5.0~7.0;5. >7.0
	距农村道路距离/km	1. 0~0.3;2. >0.3~0.5;3. >0.5~0.7;4. >0.8~1.0;5. >1.0
	社会经济	居民点规模指数
常住人口数/人		1. <22 000;2. 22 000~30 000;3. >30 000~50 000;4. >50 000~ 100 000;5. >100 000
人口户数/户		1. <12 000;2. 12 000~15 000;3. >15 000~25 000;4. >25 000~ 35 000;5. >35 000
人均 GDP/(万元·人 ⁻¹)		1. <2;2. 2~3;3. >3~5;4. >5~10;5. >10
劳动力指数		1. <0.2;2. 0.2~0.5;3. >0.5~0.53;4. >0.53~0.56;5. >0.56

模型预测的错误率可以通过全体样本集中剩余样本识别,规避单个决策树造成的误差。随机森林流程图见图 2。

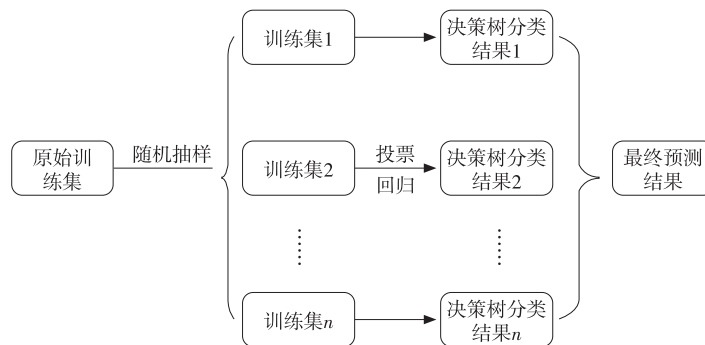


图 2 随机森林流程图

Fig.2 Random forest flow chart

随机森林的特征之一是 Gini 系数可计算出农村居民点用地适宜性评价因子的相对权重,在随机森林分类树中,Gini 系数法计算出的不纯度可衡量随机森林分类树中的最佳分割^[26]。不纯度是从一个数据集中随机选取子项,度量它被错误地划分到其他组里的概率。评估因子的重要性通过评价因子的平均 Gini 减少量占有所有评价因

素的平均 Gini 减少量之和的百分比来衡量^[27]。具体为:

$$G = \sum_{i \neq j}^J p(i|t)p(j|t),$$

其中: G 表示 Gini 系数, t 表示节点, 在节点 t 处依据概率 $p(i|t)$ 将随机选择的对象分配给类 i , 对象实际属于类 j 的估计概率是 $p(j|t)$ 。

4 模型训练

4.1 模型精度

去除原始农村居民点 500 m 缓冲区和水系作为不适宜性区, 以 1:1 的比例提取 1 000 个非农村居民点作为负样本, 并和原始 1 000 个正样本形成基础样本点数据, 采用十折交叉验证法进行随机森林模型的训练。十折交叉验证是将全体样本集(正负样本各 1 000 个)平均分成 10 个子集, 每次选取 9 个子集训练并生成 n 棵决策树, 1 个子集进行验证, 轮流选取, 并计算错误率, 得到每次试验的测试集精度(表 3)^[28]。

表 3 十折交叉验证的准确性

Tab. 3 Accuracy of 10-fold cross-validation

样本	精确度		样本	精确度	
	训练数据	测试数据		训练数据	测试数据
1	0.780	0.992	6	0.890	0.989
2	0.840	0.990	7	0.860	0.991
3	0.840	0.987	8	0.890	0.986
4	0.850	0.991	9	0.860	0.989
5	0.880	0.990	10	0.900	0.989

将农村居民点适宜性评价近似看为二分模型, 采用混淆矩阵对模型精度进行评价。从表 4 可看出, 混淆矩阵显示精度均较高, 说明随机森林模型精度高, 具有很高的预测价值(表 4)。

表 4 随机森林模型混淆矩阵

Tab. 4 Random forest model confusion matrix

随机森林预测值	真实值		精度
	适宜	不适宜	
适宜	490	7	0.986
不适宜	10	493	0.980
召回率	0.980	0.986	0.983

4.2 模型检验

模型的有效性可以利用受试者工作特征(Receiver operating characteristic, ROC)曲线衡量, AUC 值与预测精度密切相关, 该值越接近 1 表明精确度越高。本研究的训练数据、测试数据、全体数据的 AUC 值分别为 0.999, 0.996 和 0.944, 再次表明模型预测准确性较高且稳定(图 3)。

5 农村居民点用地适宜性评价结果

将完成训练后的随机森林模型应用于整个研究区, 划定农村居民点用地斑块的适宜性等级: 适宜区、一般适宜区和不适宜区(图 4、图 5、表 5)。从图和表中可知, 3 个区的斑块比例是 62.44%, 21.00%, 16.56%, 其中适宜区域的居民点斑块最多, 占区域的大部分面积, 主要分布于合川区中部、北部地势平缓的丘陵盆地; 一般适宜区域主要分布于西南、东南部, 以及距嘉陵江、涪江 1 000 m 范围外; 不适宜区域斑块最少, 面积最小, 位于南部、东南部岭谷, 分布于嘉陵江、涪江、渠江两侧区域和华蓥山支脉云雾山一带(双凤至盐井一线和三汇-清平-土场以东)。随着农村居民点用地适宜性的降低, 各级面积占比呈递减趋势, 农村居民点的个数逐级降低。

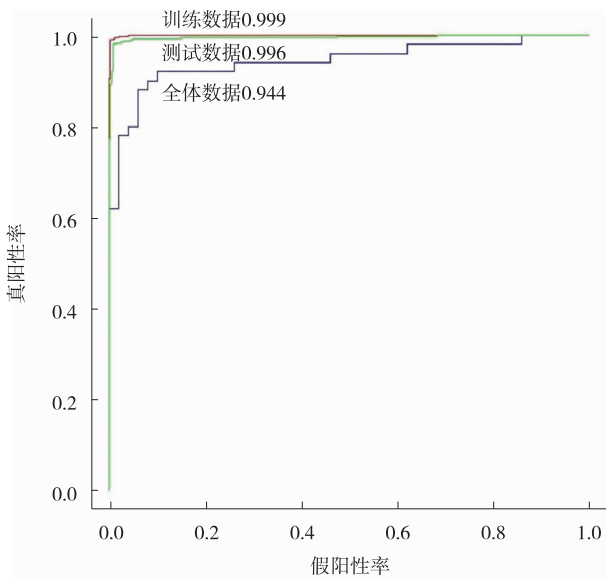


图 3 ROC 曲线及 AUC 值

Fig. 3 ROC curve and AUC value

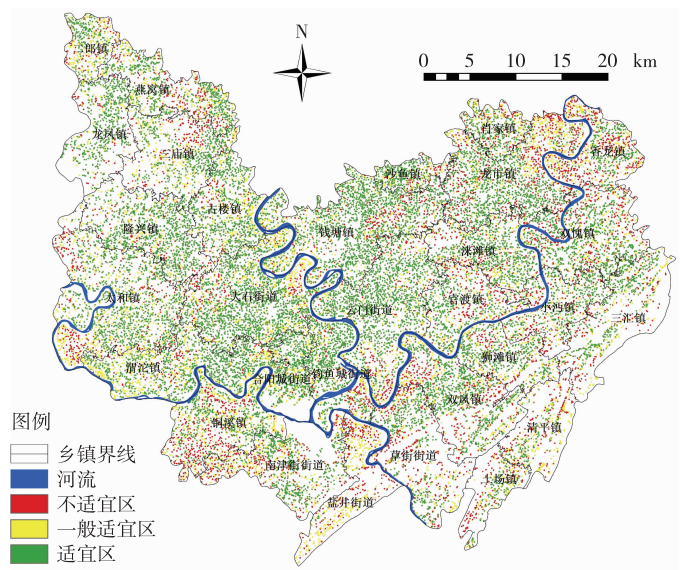


图 4 合川区农村居民点用地适宜性评价图

Fig. 4 Appraisal map of land use suitability for rural residential areas in Hechuan district

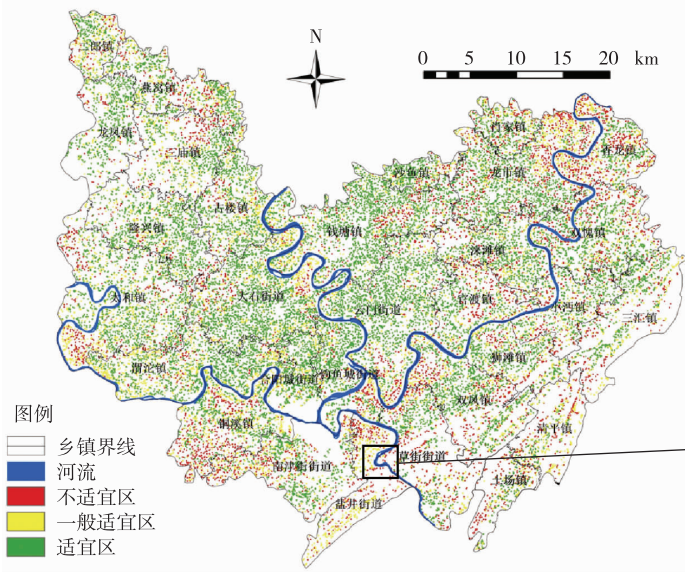


图 5 合川区农村居民点用地适宜性评价节点放大图

Fig. 5 Enlarged view of the evaluation node of land suitability for rural residential areas in Hechuan district



1) 适宜区。农村居民点适宜区域的斑块面积和斑块个数占比分别为 62.44% 和 59.60%，涉及农村居民点图斑 17 787 个，面积 79.40 hm²。该类型农村居民点平均斑块大小(即斑块总面积除以斑块个数)为 44.64 m² · 个⁻¹，主要分布在钱塘镇、大石街道、云门街道、渭沱镇、龙市镇、太和镇等区域(表 6)。适宜区与评价指标体系中各因子的作用程度相一致(图 6)，区位优势 and 自然条件的优越将提升农村居民点用地适宜性。

2) 一般适宜区。农村居民点一般适宜区域的斑块面积和斑块个数占比分别为 21.01% 和 22.37%，涉及农村居民点图斑 6 674 个，面积 26.70 hm²，该类型农村居民点平均斑块大小为 40.01 m² · 个⁻¹，主要分布在大石街道、龙市镇、三汇镇、双槐镇、太和镇、铜溪镇等区域，临近涪江、嘉陵江、渠江，地势略有起伏(表 6)。一般适宜区受到了三江的影响，与水资源关系密切。

3) 不适宜区。合川区农村居民点适宜区域的斑块面积和斑块个数占比分别为 16.56% 和 18.03%，涉及农村居民点图斑 5 381 个，面积 21.05 hm²。该类型农村居民点平均斑块大小为 39.12 m² · 个⁻¹，主要分布在草街街道、盐井街道、铜溪镇、香龙镇、双凤镇、土场镇等区域，该区域位于海拔较高、坡度较陡的地区，或者位于海拔

较低的三江交汇处。该区域地灾频发,交通不便、基础设施薄弱、用地条件恶劣,农村居民点零星分布,规模较小(表6)。

结果表明,合川区农村居民点用地适宜性评价等级的平均斑块大小与适宜性高低呈正比,应以保护耕地为原则进行优化。

表5 农村居民点用地适宜性评价

Tab. 5 Appraisal of the suitability of land use in rural residential areas

适宜等级	面积/hm ²	面积占比/%	分布区域	农村居民点个数
适宜区	79.40	62.44	合川区中部、北部地势平缓的丘陵盆地	17 787
一般适宜区	26.70	21.00	西南、东南部,以及距嘉陵江、涪江 1 000 m 范围外	6 674
不适宜区	21.05	16.56	南部、东南部岭谷,分布于嘉陵江、涪江、渠江两侧区域和华蓥山支脉云雾山一带(双凤至盐井一线和三汇-清平-土场以东)	5 381

表6 各乡镇农村居民点用地适宜性评价

Tab. 6 Evaluation of the suitability of land use for rural residential areas in various towns

hm²

序号	镇(街道)名	不适宜	一般适宜	适宜	序号	镇(街道)名	不适宜	一般适宜	适宜
1	草街街道	1.23	0.80	2.15	16	三庙镇	0.54	0.67	2.05
2	大石街道	0.52	1.36	6.52	17	沙鱼镇	0.17	0.33	1.15
3	钓鱼城街道	0.51	0.62	3.3	18	狮滩镇	0.66	0.73	2.12
4	二郎镇	0.23	0.56	0.81	19	双凤溪	1.29	1.08	3.69
5	古楼镇	0.23	0.07	2.56	20	双槐镇	1.32	1.12	3.43
6	官渡镇	0.58	0.59	1.51	21	太和镇	1.01	1.81	4.04
7	合阳城街道	0.09	0.57	2.36	22	铜溪镇	0.94	1.27	1.57
8	涑滩镇	0.51	0.68	2.43	23	土场镇	0.43	0.40	0.51
9	龙凤镇	0.17	0.31	1.87	24	渭沱镇	0.31	1.63	4.04
10	龙市镇	2.11	1.71	4.99	25	香龙镇	1.45	1.23	1.49
11	隆兴镇	0.36	0.72	3.35	26	肖家镇	0.47	0.57	1.36
12	南津街街道	0.96	0.83	2.23	27	小沔镇	0.61	0.61	1.67
13	钱塘镇	0.96	1.06	6.92	28	盐井街	0.88	1.38	0.09
14	清平镇	0.57	0.86	0.65	29	燕窝镇	0.58	0.84	2.06
15	三汇镇	0.65	1.19	2.34	30	云门街道	0.71	1.10	6.14

6 讨论与结论

6.1 讨论

6.1.1 因子重要性排序 农村居民点用地适宜性受到多种因素的影响,本研究在已有研究基础上,将劳动力指数、居民点规模指数等社会经济因素纳入农村居民点用地适宜性评价,构建了由区位条件、生产属性、自然条件、社会经济组成的共计4类20种因子的评价体系。本研究完善了评价体系,提高了评价结果的合理性和科学性,可为其他农村居民点用地适宜性评价提供参考。

农村居民点用地适宜性评价分级受到不同因素的影响。本文通过随机森林模型的平均 Gini 指数降低度指标对研究中20种影响因子进行了重要性评价,结果如图6。农村居民点与城镇、交通道路的关系影响农村居民点的建设状态。距离越近,越容易受到经济发展的影响;距离越远,则主要受到地貌形态的影响。农村居民点多为自发性建设^[29]。根据观察结果发现合川区的坡度、距农村道路距离、地形起伏度、降雨量对农村居民点的影响

最大。农村居民点零星分布在缓坡区域,坡度越大的区域越不适宜居住,农民日常活动影响了农村居民点与外界的交流,步行的可达性影响耕作的效率^[30]。降雨对农村居民点的影响主要取决于降雨量等对农村居民点的生存环境的影响^[31]。距县级以上道路距离、劳动力指数、人均 GDP 对农村居民点用地适宜性有一定影响。县级道路的等级、宽度、道路网密度等影响农村居民点的发展,道路对经济、产业、人口的影响较大^[32],道路的可达性反映了农村居民点的区位优势。劳动力指数和人均 GDP 反映了社会经济因素对农村居民点用地适宜性的影响。通过因子的重要性排序可以看出自然条件对合川区农村居民点用地适宜性的影响最大,其次是区位条件,社会经济和生产属性影响最小。合川区由于地处中丘陵和重庆平行岭谷之间,农村居民点的形成和发展受到嘉陵江、渠江、涪江和高山的阻隔,影响了农村居民点的可达性和规模。所以,自然条件和区位条件在农村居民点中的影响较大,这也和其他山区、山水地区的农村居民点研究结论吻合。

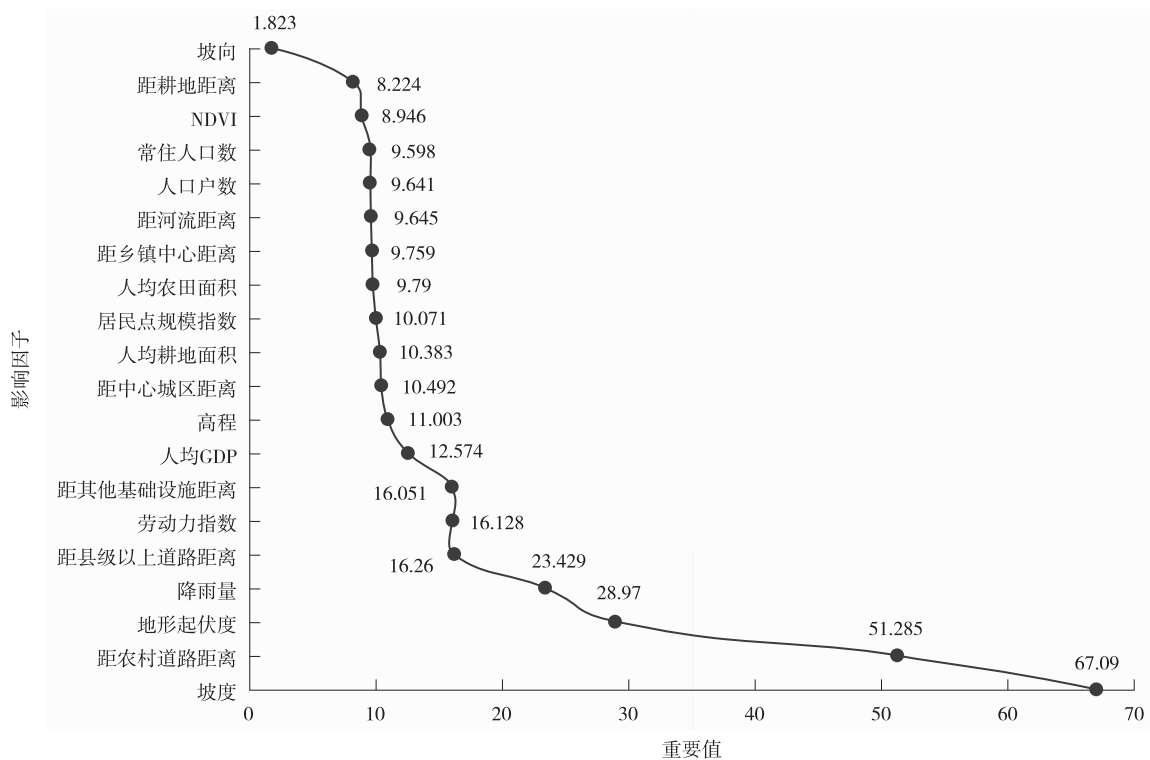


图 6 因子的重要性评价

Fig. 6 Factor importance evaluation

6.1.2 农村居民点用地适宜性分区特征与优化方案 根据农村居民点用地适宜性评价结果,并结合研究区自身特征和《国家乡村振兴战略规划(2018—2022年)》的要求,在合川区各农村居民点用地适宜性的评价约束下提出优化方案(表7、表8):

1) 适宜区。适宜区农村居民点的地理位置优越,人口较为集中,辐射周边若干居民点;紧邻城区的农村居民点,如大石街道、云门街道、钓鱼城街道部分农村居民点等应融入城区发展;农村居民点集聚且对周边农村居民点有辐射和带动作用的农村居民点应逐渐往中心村发展,在原有规模基础上有序推进改造提升,激活产业、优化环境、提高农村居民点的人居环境品质^[33]。

2) 一般适宜区。一般适宜区的农村居民点具有一定的发展潜力,该区应依托自身优势,进行内部改造,同时发展生态农业、文化旅游和特色工业等,以渠江画廊-涪滩禅宗文化旅游线路、城区-古楼-龙多山乡村旅游线路、钓鱼城-草街老街历史教育旅游线路等为依托,存留提升,提高环境品质,构建农村生态空间格局。

3) 不适宜区。不适宜的农村居民点用地规模小、布局分散,社会经济发展水平不高,城市辐射效应低下,应向邻近的中心村和就近耕作的适宜区搬迁。嘉陵江、涪江、渠江两侧区域和华蓥山支脉云雾山一带的不适宜区分布较多,由于生存条件恶劣、生态环境脆弱,必须搬迁,搬迁过程中统筹考虑农民的生计问题。同时,利用自然资源条件,发展现代农业,打造乡村生态旅游,闲置的农村居民点进行整合优化,退耕还林,提高环境的生态效益。

表7 合川区农村居民点优化思路

Tab.7 Ideas on optimizing rural residential areas in Hechuan district

适宜性分区	优化措施
适宜区	保留发展、产业带动、城镇化
一般适宜区	内部改造
不适宜区	整合优化、搬迁撤并

表8 农村居民点优化方案^[5]

Tab.8 Optimization plan for rural residential areas

功能特征	特征	发展方向	农村居民点优化要求
适宜区	地理位置优越,人口较为集中,交通便捷,区域社会经济发展水平高,与城市联系紧密,承接城市疏散功能和辐射带功能。	在原有规模基础上有序推进改造提升,激活产业、优化环境,提高农村居民点的人居环境品质。	鼓励发挥自身比较优势,强化主导产业支撑,改善生产生活条件。
一般适宜区	距离城市中心一般,农业经济与农业就业占主导地位,受城市辐射效应影响小。	完善配套功能,加快发展现代服务,引导农村土地有效流转,发展适度规模经营。	加快城乡产业融合发展,依托自身特色农业、工业、文化和旅游资源,发展生态农业、文化旅游。减少农民耕作半径,扩大耕地规模。
不适宜区	生存条件恶劣、生态环境脆弱,社会经济发展水平不高,城市辐射效应低。	实施农村居民点搬迁撤并。发展现代农业,形成生态旅游产业链。	限制农村居民点的新建、扩建活动,提高农村居民点的集约利用水平,退耕还林,提高生态效益。

6.2 结论

本文以重庆市合川区为研究区,运用随机森林算法,以农村居民点为切入点,实现了农村居民点用地适宜性研究方法的优化。主要结论如下:

1) 劳动力指数、人均 GDP 对农村居民点用地适宜性有一定影响,这符合当地的地理环境特征与发展需求。以往研究中考虑到社会经济因素忽略劳动力指数、居民点规模指数对农村居民点用地适宜性的影响。劳动力指数可定量表征区域内劳动力的性别差异,居民点规模指数可确定居民点规模与耕地面积的匹配度。这些社会经济因素可完善农村居民点用地适宜性的指标体系。

2) 随机森林模型准确率高达 0.991,且精确度均在 0.980 以上,AUC 值均较高,说明构建的随机森林模型适用于农村居民点用地适宜性的预测研究,且具有较高的准确性。

3) 从空间分布和数量来看,适宜性农村居民点分布符合山区发展特征,一般适宜和不适宜区域的农村居民点分布临近水资源丰富、地势平坦之处,这与研究区域的地质条件、地理环境有关,也证明了水源与农村居民点选址的空间关系。

4) 由于自然资源和环境条件的制约,合川区农村居民点用地适宜度空间分布差异较大,在用地适宜分级的基础上,提出各类农村居民点的优化措施,为推进全域乡村规划发展提供参考。

参考文献:

- [1] WAN A, LIU Y, XIE X, et al. Study on spatial layout optimization of Linpan settlements based on point pattern analysis[J]. Environment Development and Sustainability, 2020(17):1-19.
- [2] 张天柱,张凤荣,谢臻,等.精准扶贫背景下云南少数民族山区农村居民点空间格局演变[J].农业工程学报,2019,35(9):246-254.
ZHANG T Z, ZHANG F R, XIE Z, et al. Spatial pattern evolution of rural settlements in mountainous areas of ethnic minorities in Yunnan Province under the background of precise poverty alleviation[J]. Transactions of the Chinese Society of Agricultural Engineering, 2019, 35(9):246-254.
- [3] 曲衍波,张凤荣,姜广辉,等.基于生态位的农村居民点用地适宜性评价与分区调控[J].农业工程学报,2010,26(11):290-296.

- QU Y B,ZHANG F R,JIANG G H, et al. Suitability evaluation and subarea control and regulation of rural residential land based on niche[J]. Transactions of the Chinese Society of Agricultural Engineering, 2010, 26(11): 290-296.
- [4] 梁照凤,袁媛,陈文波,等. 基于加权 Voronoi 图的农村居民点用地适宜性评价与整治分区研究[J]. 江西农业大学学报, 2017, 39(6): 1244-1255.
- LIANG Z F, YUAN Y, CHEN W B, et al. A study on suitability evaluation of rural residential areas and its consolidation division based on weighted voronoi diagram[J]. Acta Agricultural Universitatis Jiangxiensis, 2017, 39(6): 1244-1255.
- [5] 白雪,刘立文,段永红,等. 基于 TOPSIS 模型的农村居民点布局优化研究:以山西农谷为例[J]. 湖北农业科学, 2021, 60(3): 60-65.
- BAI X, LIU L W, DUAN Y H, et al. Study on layout optimization of rural residential area based on TOPSIS model; taking agricultural valley in Shanxi as an example[J]. Hubei Agricultural Sciences, 2021, 60(3): 60-65.
- [6] 汤昇,于婧,陈艳红,等. 乡村振兴背景下农村居民点用地适宜性评价:以慈利县为例[J]. 湖北大学学报(自然科学版), 2020, 42(5): 531-538.
- TANG W, YU J, CHEN Y H, et al. Suitability evaluation of rural residential land under the background of rural revitalization; a case study of Cili county[J]. Journal of Hubei University (Natural Science), 2020, 42(5): 531-538.
- [7] 刘立文,段永红,李丽丽,等. 山西省农村居民点空间分布特征及其适宜性评价[J]. 中国农业资源与区划, 2022, 43(1): 100-109.
- LIU L W, DUAN Y H, LI L L, et al. Spatial distribution characteristics and suitability evaluation of rural residential areas in Shanxi province[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2022, 43(1): 100-109.
- [8] 文博,刘友兆,夏敏,等. 基于灰靶模型的宜兴市农村居民点布局适宜性评价及分类调控研究[J]. 地域研究与开发, 2016, 35(5): 153-157.
- WEN B, LIU Y Z, XIA M, et al. Suitability evaluation and regulation of rural residential land in Yixing city based on the gray target model[J]. Areal Research and Development, 2016, 35(5): 153-157.
- [9] 邵新娟. 基于生态安全格局的农村居民点布局优化研究[J]. 湖北农业科学, 2018, 57(17): 99-105.
- SHAO X J. Layout optimization research of rural residential land based on the ecological security pattern[J]. Hubei Agricultural Sciences, 2018, 57(17): 99-105.
- [10] 李冰清,王占岐,张利国,等. 基于集聚发展路径的农村居民点空间重构研究[J]. 中国人口·资源与环境, 2018, 28(11): 47-55.
- LI B Q, WANG Z Q, ZHANG L G, et al. Study on space reconstruction of rural settlement based on clustering development path[J]. China Population, Resources and Environment, 2018, 28(11): 47-55.
- [11] 唐秀美,刘玉,任艳敏. 基于适宜性与发展水平评价的村庄用地类型划分[J]. 农业工程学报, 2020, 36(1): 283-291.
- TANG X M, LIU Y, REN Y M. Classification of village land types based on suitability and development evaluation[J]. Transactions of the Chinese Society of Agricultural Engineering, 2020, 36(1): 283-291.
- [12] 晋蓓,陈琼,张颖,等. 基于 GIS 技术的杨陵区居民点用地适宜性评价[J]. 水土保持研究, 2018, 25(1): 397-402.
- JIN B, CHEN Q, ZHANG Y, et al. Study on suitability evaluation of residential area in Yangling district based on GIS technology[J]. Research of Soil and Water Conservation, 2018, 25(1): 397-402.
- [13] 邵红娟,张朝琼,蔡广鹏,等. 基于 GIS 的岩溶山区农村居民点用地适宜性评价[J]. 水土保持研究, 2015, 22(2): 200-203.
- GAO H J, ZHANG C Q, CAI G P, et al. Land suitability evaluation of rural settlements in Karst mountains based on GIS[J]. Research of Soil and Water Conservation, 2015, 22(2): 200-203.
- [14] ROY B. Optimum machine learning algorithm selection for forecasting vegetation indices: MODIS NDVI & EVI[J]. Remote Sensing Applications: Society and Environment, 2021, 23: 100582.
- [15] XU F, HO H, CHI G Q, et al. Abandoned rural residential land: using machine learning techniques to identify rural residential land vulnerable to be abandoned in mountainous areas[J]. Habitat International, 2019, 84: 43-56.
- [16] 徐枫,王占岐,张红伟,等. 随机森林算法在农村居民点适宜性评价中的应用[J]. 资源科学, 2018, 40(10): 2085-2098.
- XU F, WANG Z Q, ZHANG H W, et al. Application of random forest algorithm in suitability evaluation of rural residential land[J]. Resources Science, 2018, 40(10): 2085-2098.
- [17] SATHYADEVAN S, NAIR R R. Comparative analysis of decision tree algorithms: ID3, C4. 5 and random forest[M]. India: Springer, 2015.
- [18] 重庆市合川区统计局,国家统计局合川调查队. 合川区 2020 年统计年鉴[EB/OL]. (2021-02-22)[2021-10-14]. <http://www.hc.gov.cn/xxgk/qzfwj/tjxx/202102/W020210222385393306552.pdf>.
- Hechuan District Statistics Bureau Chongqing City, Hechuan Investigation Team of the National Bureau of Statistics. Hechuan district 2020 statistical yearbook[EB/OL]. (2021-02-22)[2021-10-14]. <http://www.hc.gov.cn/xxgk/qzfwj/tjxx/202102/>

W020210222385393306552.pdf.

- [19] 何翔宇,吴克宁,宋恒飞,等.基于图斑特征的农村居民点整理分区研究:以宜兴市为例[J].江西农业大学学报,2018,40(5):1092-1101.
HE X Y, WU K N, SONG H F, et al. A study on the rural residential area sorting based on the features of the pattern spot: a case study of Yixing[J]. *Acta Agriculturae Universitatis Jiangxiensis*, 2018, 40(5): 1092-1101.
- [20] 宋文,吴克宁,刘浩然,等.基于地理空间适宜性指数自相关的农村居民点整理分区[J].农业工程学报,2016,32(19):249-257.
SONG W, WU K N, LIU H R, et al. Zoning of rural residential land consolidation based on spatial autocorrelation of geographical spatial suitability index[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2016, 32(19): 249-257.
- [21] 杨学龙,叶秀英,赵小敏.鄱阳县农村居民点布局适宜性评价及其布局优化对策[J].中国农业大学学报,2015,20(1):245-255.
YANG X L, YE X Y, ZHAO X M. Rural residential layout suitability evaluation and optimization countermeasures in Poyang county[J]. *Journal of China Agricultural University*, 2015, 20(1): 245-255.
- [22] 刘玉娇,朱青,吕立刚,等.基于农户意愿分析的农村居民点整理时序分区[J].水土保持研究,2017,24(5):299-304.
LIU Y J, ZHU Q, LV L G, et al. Time-series zoning for rural residential land consolidation based on study of farm households' willingness[J]. *Research of Soil and Water Conservation*, 2017, 24(5): 299-304.
- [23] 魏璐瑶,陈娱,张正方,等.多情景视角下基于空间组合识别的农村居民点布局优化研究:以江苏省新沂市为例[J].地理研究,2021,40(4):977-993.
WEI L Y, CHEN Y, ZHANG Z F, et al. Rural settlements layout optimization based on spatial combination identification from a multi-scenario perspective: taking Xinyi city of Jiangsu province as an example[J]. *Geographical Research*, 2021, 40(4): 977-993.
- [24] SUN D L, XU J H, WEN H J, et al. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: a comparison between logistic regression and random forest[J]. *Engineering Geology*, 2020, 281: 105972.
- [25] 侯陈瑶,朱秀芳,孙丹一,等.基于随机森林的旱灾损失模型构建与应用:以河南省雨养小麦为例[J].灾害学,2019,34(4):210-215.
HOU C Y, ZHU X F, SUN D Y, et al. Construction and application of drought loss model based on random forest: Taking rain-fed wheat in Henan province as an example[J]. *Journal of Catastrophology*, 2019, 34(4): 210-215.
- [26] HO T K. The random subspace method for constructing decision forests[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998, 20(8): 832-844.
- [27] 孙德亮.基于机器学习的滑坡易发性区划与降雨诱发滑坡预报预警研究[D].上海:华东师范大学,2019.
SUN D L. Mapping landslide susceptibility based on machine learning and forecast warning of landslide induced by rainfall[D]. Shanghai: East China Normal University, 2019.
- [28] WU R, WANG J, ZHANG D, et al. Identifying different types of urban land use dynamics using Point-of-interest (POI) and random forest algorithm: the case of Huizhou, China[J]. *Cities*, 2021, 114(2021): 103202.
- [29] 张霞,魏朝富,倪九派,等.重庆市喀斯特槽谷地区农村居民点分布与地貌形态要素关系研究[J].中国岩溶,2012,31(1):59-66.
ZHANG X, WEI C F, NI J P, et al. Research on the distribution of rural settlements in relation to landform factors in karst trough valley of Chongqing municipality[J]. *Carsologica Sinica*, 2012, 31(1): 59-66.
- [30] 周雪,陈伟,王倩,等.考虑耕地利用条件的黄土丘陵沟壑区农村居民点整理分区及利用对策[J].水土保持研究,2019,26(5):312-320.
ZHOU X, CHEN W, WANG Q, et al. Zoning and strategy of rural residential land consolidation on loess hilly and gully regions with consideration of arable land use conditions[J]. *Research of Soil and Water Conservation*, 2019, 26(5): 312-320.
- [31] 张英杰,雷国平.地质灾害易发区农村居民点布局优化研究:以浙江洞头为例[J].生态与农村环境学报,2019,35(11):1387-1395.
ZHANG Y J, LEI G P. Analysis of spatial distribution and optimization of rural settlements in geological disaster-prone area: a case study of Dongtou county, Zhejiang province[J]. *Journal of Ecology and Rural Environment*, 2019, 35(11): 1387-1395.
- [32] 唐秀美,刘玉,任艳敏.基于适宜性与发展水平评价的村庄用地类型划分[J].农业工程学报,2020,36(1):283-291.
TANG X M, LIU Y, REN Y M. Classification of village land types based on suitability and development evaluation[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2020, 36(1): 283-291.
- [33] 赵智源,武国胜,林惠花.乡镇主导功能影响下的农村居民点整理研究:以舞钢市为例[J].福建师范大学学报(自然科学版),

2021, 37(3):96-105.

ZHAO Z Y, WU G S, LIN H H. A study on the arrangement of rural residential areas influenced by the leading functions of towns and villages: a case of Wugang city[J]. Journal of Fujian Normal University (Natural Science Edition), 2021, 37(3):96-105.

Resources, Environment and Ecology in Three Gorges Area

Evaluation of Suitability of Rural Residential Land Use Based on Random Forest Algorithm: Take Hechuan District, Chongqing City as an Example

LI Linzhi¹, CHEN Xingyu¹, ZHANG Hong¹, SUN Deliang¹, WEN Haijia², CHEN Shuang³

(1. The Key Laboratory of GIS Application Research, Chongqing Normal University, Chongqing 401331;

2. Key Laboratory of New Technology for Construction of Cities in Mountain Area, Ministry of Education,

Chongqing University, Chongqing 400045; 3. Chongqing Planning & Design Research Institute, Chongqing 401147, China)

Abstract: [Purposes] In order to rationally lay out rural residential areas and promote the development of rural planning in the whole region, it is very necessary to scientifically evaluate the suitability of rural residential areas. [Methods] It selects 20 factors in 4 aspects, including location conditions, production attributes, natural conditions, and social economy, to construct an evaluation index system for the suitability of rural residential areas, and selects 1 000 rural residential areas in the study area as a positive sample, a negative sample of 1 000 rural settlements is randomly selected in a 1 : 1 ratio. Based on the random forest algorithm, a land suitability evaluation model for rural settlements is constructed, and the suitability level of the rural settlements in the study area is divided into suitable ones using the natural breakpoint method. Generally suitable and not suitable for three grades, according to the evaluation results, the optimization plan for the suitability of rural residential land in Hechuan District is proposed. [Findings] 1) The effectiveness of the model was verified, and the AUC values of the training data, test data, and overall data were 0.999, 0.996, and 0.944, respectively, and the highest accuracy of the model test was 0.991, indicating that the model prediction accuracy is high and stable. 2) Among the evaluation factors, slope, distance from rural roads, and topographic undulation have a greater impact on the evaluation results of the suitability of rural residential sites. 3) The suitable area is located in the gentle hilly basin in the middle and north of Hechuan District, and it is recommended to keep and develop. Generally suitable areas are mostly located in the southwest and southeast, 1 000 meters away from Jialing River and Fujian River, and internal reconstruction is recommended; unsuitable areas are located in the south and southeast ridges and valleys and water systems, distributed on both sides of the Sanjiang River and the Huaying Mountain branch Yunwushan area. It is recommended to relocate and merge. [Conclusions] The use of random forest model to evaluate the suitability of rural residential areas has high accuracy, and provides a reference for the management and control of rural residential areas.

Keywords: random forest algorithm; factor classification; land use suitability evaluation; rural residential area; Hechuan

(责任编辑 黄 颖)