

基于随机森林的山洪灾害风险评估 ——以江西省为例

吴小君¹, 方秀琴¹, 任立良², 吴陶樱¹, 苗月鲜¹

(1. 河海大学 地球科学与工程学院, 南京 211100;

2. 河海大学 水文水资源与水利工程科学国家重点实验室, 南京 210098))

摘要:依据流域灾害系统理论,以江西省山洪灾害的特点为基础,从触发因子、下垫面孕灾环境和承灾体3个角度确定影响山洪灾害的9个评价指标。将历史发生山洪灾害点的正样本集和5组随机采样点的负样本集组成不同的总样本集,并进行训练样本和测试样本的随机分组。通过训练样本构建了基于随机森林的模型,并通过测试样本对模型进行检验,平均精度为86.26%,表明构建的随机森林模型具有较好的拟合效果。对山洪灾害风险图在行政分区上进行了风险高值区面积绝对值和面积占比的定量统计分析,结果发现景德镇和上饶市山洪灾害风险度较高和很高的区域不仅面积占比大,所占面积绝对值也很大,属于山洪灾害高发区域。

关键词:山洪灾害; 风险评估; 随机森林; GIS

中图分类号: X43; S422

文献标识码: A

文章编号: 1005-3409(2018)03-0142-08

Risk Assessment of Mountain Torrents Disaster Based on Random Forest —A Case Study in Jiangxi Province

WU Xiaojun¹, FANG Xiuqin¹, REN Liliang², WU Taoying¹, MIAO Yuexian¹

(1. College of Earth Science and Engineering, Hohai University, Nanjing 211100, China; 2. State Key

Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China)

Abstract: According to the theory of mountain torrent disaster system, based on the characteristics of mountain torrent disaster in Jiangxi Province, nine factors were selected to determine the impact of mountain torrent disaster from three aspects including trigger factors, the underlying surface of the disaster environment and the relief body. We took the historical data of the mountain torrent disaster as the positive sample set, five groups of random samples as a negative sample set to form the total sample set, and randomized the training samples and test samples, respectively. Through training samples, the estimation models were set up based on the random forest. Accuracy assessment showed an average accuracy of 86.26%, which indicated the satisfying performance of random forest model. Based on the mapping of mountain torrent disaster risk assessment, statistical analyses were carried out according to different administrative areas. The results showed that Jingdezhen and Shangrao were the most vulnerable to mountain flood disaster as they had both relative and absolute larger values in the higher or greater area of the risk degree of the mountain flood disaster.

Keywords: mountain torrent disaster; risk assessment; random forest; GIS technology

一般来说,山洪是由于降雨导致的具有突发性、流量大、破坏力强等特点的地表径流,它是洪水的一种表现形式^[1]。山洪灾害对房屋建筑、交通道路、水

利工程设施、农林畜牧业等造成破坏,极端情况下甚至会导致人员伤亡,给人口、社会和经济带来巨大损失^[2]。我国地形复杂,丘陵较多,2/3的土地面积几

收稿日期:2017-07-19

修回日期:2017-08-18

资助项目:“十三五”国家重点研发计划项目“陆地水循环演变及其在全球变化中的作用研究”(2016YFA0601500);中央高校基本科研业务费专项资金资助(2016B05414)

第一作者:吴小君(1991—),女,安徽阜阳人,研究生,从事遥感方面研究。E-mail:2360868477@qq.com

通信作者:方秀琴(1978—),女,安徽池州人,博士,从事地表参数的遥感反演及遥感、地理信息系统和水文模型的集成等方面的研究。

E-mail:kinkinfang@hhu.edu.cn

乎为丘陵。山区的面积远远超过世界平均水平^[3],是世界上受山洪影响最严重的地区之一^[4]。

评估山洪灾害风险的影响因素有很多,例如社会经济、自然和技术等,并且其评估过程需考虑多个指数,与此过程相关的主要困难是指数和风险水平之间的多变量和非线性关系。因此,山洪灾害的风险评估一直是人们研究的难点和重点^[5-7]。近些年,快速发展的人工智能技术促使很多研究学者在对象评价中引用机器学习算法。这些方法大大提高了计算量,能较好地解决非线性问题,但仍存在许多弱点。比如,人工神经网络具有收敛速度慢和局部极小的问题^[8],支持向量机数学函数复杂,使用不方便,解决多分类问题效果不佳^[9],决策树很容易陷入局部最优。最重要的是,这些智能算法无法估计每个指标对总风险的贡献。而随机森林(Random Forest,RF)是一种基于统计学习理论的组合分类方法,它的非线性特性使其适用于多变量预测,因此在很多领域都有广泛的应用^[10-14]。随机森林的智能学习机器可以在大型数据库上高效运行,并提供关于分类中特定变量重要性的估计,这使得 RF 在解决风险评估中固有的非线性问题以及估计每个指标的重要性程度方面具有相当大的优势。赖成光等通过构建随机森林模型对东江流域进行洪灾害风险评价,并在试验中对比了 SVM 方法,结果表明 RF 方法比 SVM 方法精度更高^[15];Zhaoli Wang 等根据洪水系统理论,以东江流域为例,选取十一个指标构建评价体系,提出了一种基于 RF 的评价模型^[16];Quanlong Feng 等提出了一种基于高分辨率无人机图像的随机森林与纹理分析相结合的城市洪水制图方法,并指出与其他分类器相比较,随机森林优于最大似然法和人工神经网络^[17]。但是目前随机森林在山洪灾害风险评估方面的研究比较匮乏,并且当前应用于洪灾风险的研究评价指标体系不够完善,对不同区域的洪灾风险等级也无法进行准确判断。

因此,本文尝试选择随机森林算法,结合江西省山洪灾害的特点和历史山洪灾害调查数据,构建指标体系,建立基于随机森林算法的山洪灾害风险评估模型。本文的目的是计算不同地区山洪灾害的风险程度,然后对风险等级进行分类,绘制江西省山洪灾害风险区划图,并对该地区进行统计分析。

1 研究区与指标体系

1.1 研究区简介

本文的研究区域为江西省境域,其位于长江中下

游南岸,土地总面积 16.7 万 km²,占全国国土面积的 1.7%。江西省地理位置险要,以山地丘陵居多,北部则面朝鄱阳湖^[18],全省东、南、西三面的边缘山岭构成省际天然界线,形成一个自然的分水岭。

江西省属中热带温暖湿润季风气候,气候湿润,降雨较多,年降水量 1 400~1 900 mm^[19]。省内丘陵山地众多,水网稠密,河湖众多。独特的气候特征及地形地貌,导致了省内山洪灾害的频繁发生^[20-22]。因此,为了加强对山洪灾害的风险管理,制定相应的防灾减灾策略,在江西省开展山洪灾害风险评估具有十分重要的意义。

1.2 指标体系和数据

依据山洪灾害体系理论^[23],综合考虑研究区山洪灾害的特征,参考相关研究,结合实地考察情况以及数据可获取性,我们从致灾因子、孕灾环境和承灾体 3 个方面选取了 9 个影响因素。具体选取的影响因素及其数据来源具体见表 1。

为了数据空间尺度的统一,首先应结合实际情况进行相关的数据预处理。利用 ArcGIS 10.2 软件中的栅格计算器对每个指标进行极差标准化,得到 9 个尺度一致的栅格数据,且数据的每个栅格单元都是 1 km×1 km,各个指标见图 1。

表 1 选取的 9 个影响因素及其数据来源

风险因子	影响因素	数据来源
致灾因子	年平均降雨量	江西省地图集
	年暴雨天数	
	年最大 10 min 暴雨量均值	
孕灾环境	年最大 1 h 暴雨量均值	江西省暴雨洪水查算手册
	年最大 6 h 暴雨量均值	
	地形指数	由汇流面积和坡度计算得来
承灾体	高程	SRTM
	人口密度	
	GDP 密度	中国科学院资源环境科学数据中心

2 山洪灾害风险评估模型

2.1 研究方法

(1) 随机森林算法概述。随机森林包含两个重要参数,即预选变量的个数和树的个数,这两个参数是决定随机森林预测能力的两个重要参数^[24],预选变量的数目决定了单个决策树的情况,随机森林中的树数决定了整个随机森林的整体大小。

本文中随机森林算法使用了 R 软件中的 randomForest 包^[25],具体实现时需预先设置两个重要的参数:mtree 和 ntry,这两个参数所决定的值对应于上述随机森林的两个重要参数,mtry 为每棵分类

回归树构建时节点拆分的次数, $ntree$ 为随机森林中分类回归树的个数, 这两个参数是依据随机森林构建

时产生的袋外误差确定的^[26]。根据训练样本集和 $randomForest$ 函数, 构建风险评估模型^[27-28]。

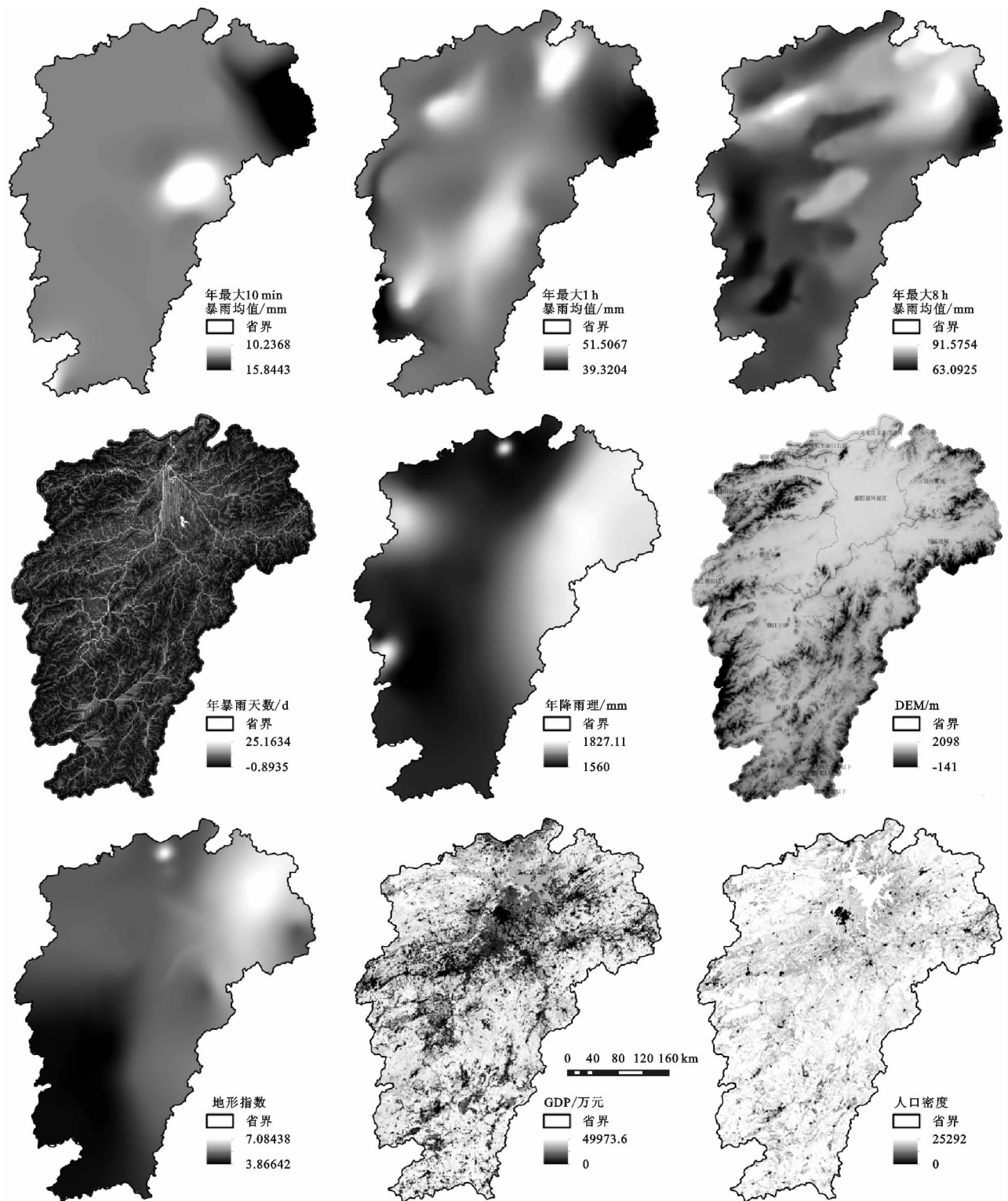


图 1 9 个影响因素空间分布

(2) 评价步骤。①根据江西省山洪特点选取风险指标;②选取历史发生山洪灾害的为正样本,未发生过灾害的为负样本,同时把正负总样本集分为 70% 的测试样本和 30% 训练样本;③将训练样本输入随机森林算法,建立山洪风险评估模型,并通过测

试样本对模型进行检验;④将江西省数据输入到模型中得到山洪灾害风险度,利用 ArcGIS 制成山洪灾害风险评价图。

由图 2 可以看出,该模型的核心部分是训练样本。随机森林算法可通过输入的样本建立山洪风险度与指

标数据之间的关系,从而建立相应的分类规则,从而对待测数据进行分类得到不同区域的山洪风险度。

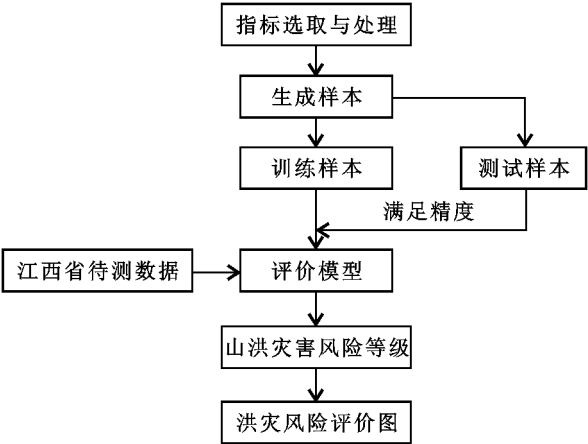
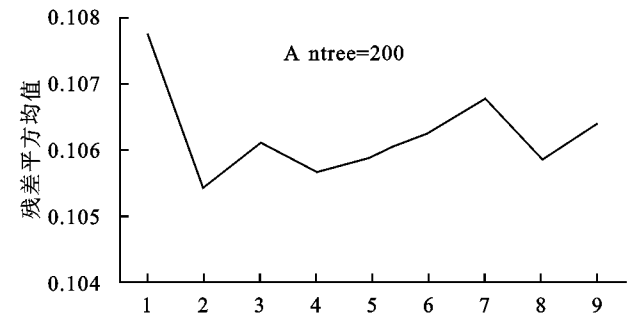


图 2 评价流程

2.2 样本选取及数据预处理

本研究以 2 009 个历史发生山洪灾害点构成正样本数据集,为了更加科学地构建随机森林模型,需要采集一定数量的未发生山洪灾害的点来组成负样本数据集。考虑到地理相似性,以历史发生山洪灾害点为中心做 3 km 范围的缓冲区,缓冲区外的区域即为负样本可采样的区域。在负样本可采样区域随机采集与正样本数据集同等数量的负样本,从而形成负样本数据集,将采集得到的负样本数据集与正样本数据集组合,组成基于随机森林的山洪灾害评估模型的总样本数据集。

为了避免偶然现象的发生,对以上步骤重复 5 次,即在负样本可采样区域内重复随机采样 5 次,并



分别与正样本集组合,从而得到 5 组不同的总样本数据集,以此来反映山洪灾害评估的一般规律。

另外,为了方便构建模型和模型的验证,分别对 5 组总样本数据集进行训练样本集和测试样本集的划分,方法是运用 R 语言对总样本数据集进行随机划分,将 5 组总样本数据集中 70% 的样本数据作为训练样本,用于随机森林模型的构建,其余 30% 的样本数据作为测试样本,用于对模型的测试和精度验证,保证研究的顺利进行。通过 ArcGIS 软件提取样本点上的 9 个影响因素的数据,为之后随机森林模型的构建基础。

2.3 随机森林参数的选择

在用 R 软件中的函数 random Forest 时,函数会存在默认的决策树数量值以及决策树节点分支所选变量个数。但是,在我们实际操作中,系统计算的默认值不一定是最适用的,因此我们在实际应用中需要通过不断地尝试计算出最优值。

(1) mtry 值的确定。在构建模型的时候,一定要通过逐次计算来挑选最佳的 mtry 值。参数 mtry 的默认值在分类模型中是变量个数的二次根式,在回归模型中则是变量个数的 1/3。

本研究为构建模型选取的影响因素为 9 个,为了得到最佳的变量个数,具体试验步骤为:(1) 设定决策树数量为 200,mtry 的值从 1 取到 9,进行 9 次试验,得到 9 个模型的残差平方和以及拟合优度值;(2) 设定决策树数量为 1 000,再进行上述重复的试验步骤,对比并分析两组试验结果,结果见图 3。

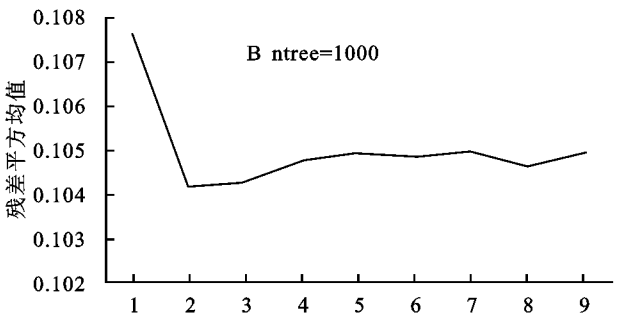


图 3 不同变量个数随机森林的残差平方均值

由图 3 看出,不管 ntree 值是 200 或者 1 000,mtry 值为 2 的时候残差平方和最小,即 mtry 值为 2 时模型为最优。

(2) ntree 值的确定。在确定了 mtry 值之后,下一步则要确定最优 ntree 值。一般 ntree 设置成一个较大的值即可,也可以不设置,随机森林模型会自动根据袋外误差计算出适宜的 ntree 值。一般该参数值设为五百或者一千,但也不是完全固定的,还需要结合实际情况进行选择。

mtry 值为 2 时模型为最优,所以接下来进行构建相应的随机森林模型,设定模型中决策树数量为 200,500,1 000,分别进行可视化分析,以此来确定决策树数量。模型误差与决策树数量关系见图 4。

从图 4 看出,当 ntree 值约大于 1 000 之后,模型的误差开始变得平稳,因此本文 ntree 值设为 1 000,以此来构建出最优的模型。

2.4 随机森林模型的建立

通过以上分析,确定了构建随机森林的最优模型的参数值,其中最佳变量个数为 2 个,最优的决策树数量

从上一步的分析来看,最佳变量个数为 2,即

为 1 000 棵。利用 5 组训练样本数据和参数值,在 R 中分别构建出 5 个基于随机森林的山洪灾害评估模型。研究表明,构建的 5 个模型中决策树的节点数最少的为 260 个,而决策树节点数最多的有 360 个。

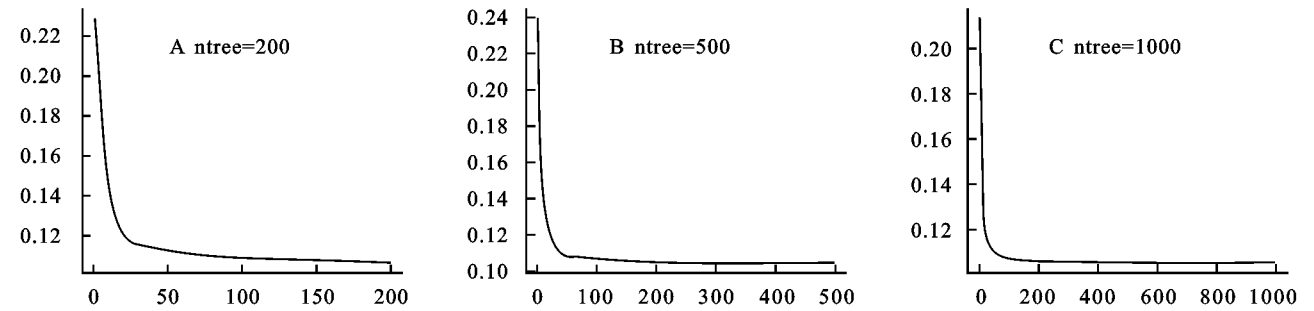


图 4 不同决策树数量情况下的模型误差

3 结果与分析

3.1 指标重要性分析

随机森林模型可以提取出模型中各个变量的重要性度量结果,可以通过 R 软件的随机森林包计算出各自变量对于模型判别效果的重要程度,这个功能帮助决策者认识并评估每个指标对总风险的贡献。这也是随机森林模型的一大特点。更确切地说,通过随机森林模型的重要性分析计算出各个变量对模型分类的影响程度,可以看出具体是哪些变量比较重要,在模型判别中具有重大的影响。

通过计算构建出的 5 个随机森林模型中各个影响因素对山洪灾害形成的重要程度,得到模型中山洪灾害影响因素的重要性排序图,见图 5。其中 IncMSE 代表的是精度平均减少值,从图 5 中可以看出,高程在构建出的 5 个随机模型中都是最重要的影响因素,表明高程对山洪灾害风险贡献程度是最大的,可知江西省的山洪灾害跟高程有着密不可分的关系,可能对最终的山洪灾害风险形成起到了决定性的作用;地形指数、年暴雨天数、人口密度和年最大 6 h 暴雨均值是对山洪灾害比较重要的 4 个影响因素;而年最大 10 min 暴雨均值、年最大 1 h 暴雨均值、年降雨量和 GDP 则是不太重要的 4 个影响因素,表明这 4 个影响因素对山洪灾害风险的贡献程度比较小,尤其是 GDP,在构建出的 5 个随机森林模型中均是最不重要的影响因素,说明 GDP 对最终的山洪灾害风险形成起到的作用微乎其微。

3.2 基于随机森林的山洪灾害风险评估

将整个研究区的影响因素数据输入到构建完成的 5 个随机森林模型中,推测出整个研究区的山洪灾害风险度值;然后进行山洪灾害风险区划,通过正态分布取值的方法来确定山洪灾害风险的等级阈值,进而依次划分出不同区域的山洪灾害风险等级^[29]。即:利用随机森林计算出的山洪灾害风险度值,将所

有栅格单元按大小进行排列,然后依次取单元总数的 10%,20%,40%,20%,10%作为 5 个风险等级的分级空间,分别对应山洪灾害风险很高、较高、中等、较低和很低 5 个风险等级。

3.3 精度分析

本研究从两个方面进行精度分析:其一,是评价随机森林模型的建模精度,即统计随机森林模型将训练样本集正确分类的比率,用来进行衡量模型对于训练样本集的拟合程度;其二,是评价山洪灾害风险评估结果的精度,即根据历史山洪灾害调查点数据与中等以上等级山洪灾害风险的空间分布匹配度,来定量评估结果的精度。

(1) 建模精度的检验。在这里设置了一个山洪灾害风险度的阈值(比如 0.5),如果样本所在位置的山洪灾害风险度推测结果大于这个阈值,则认为这个样本分类是正确的,否则为分类错误,如此就可以度量出测试样本集分类正确的比率(表 2)。

表 2 随机森林模型中样本分类正确的比率						
模型	样本数		误判数		精度/%	
	训练	测试	训练	测试	训练	测试
1	2814	1204	3	171	99.89	85.80
2	2814	1204	0	169	100.00	85.96
3	2814	1204	0	163	100.00	86.46
4	2814	1204	0	163	100.00	86.46
5	2814	1204	2	161	99.92	86.63
平均					99.96	86.26
精度						

从表 2 可以看出,根据 5 组不同的样本集构建的随机森林模型的训练样本平均精度为 99%以上,测试样本精度在 86%左右,测试平均精度为 86.26%。通过分析可知,构建的随机森林模型都比较满足精度要求。

(2) 山洪灾害风险评估结果的验证。将基于随机森林的山洪灾害风险评估模型的 5 次结果做如下处理:在整个研究区的每个栅格单元上,对构建的 5 个随机森林模型得到的山洪灾害风险度取平均值,计算出每个栅格单元的

洪灾风险度平均值,作为基于随机森林方法的山洪灾害风险评估的最终结果。因此,山洪灾害风险等级图见图 6。

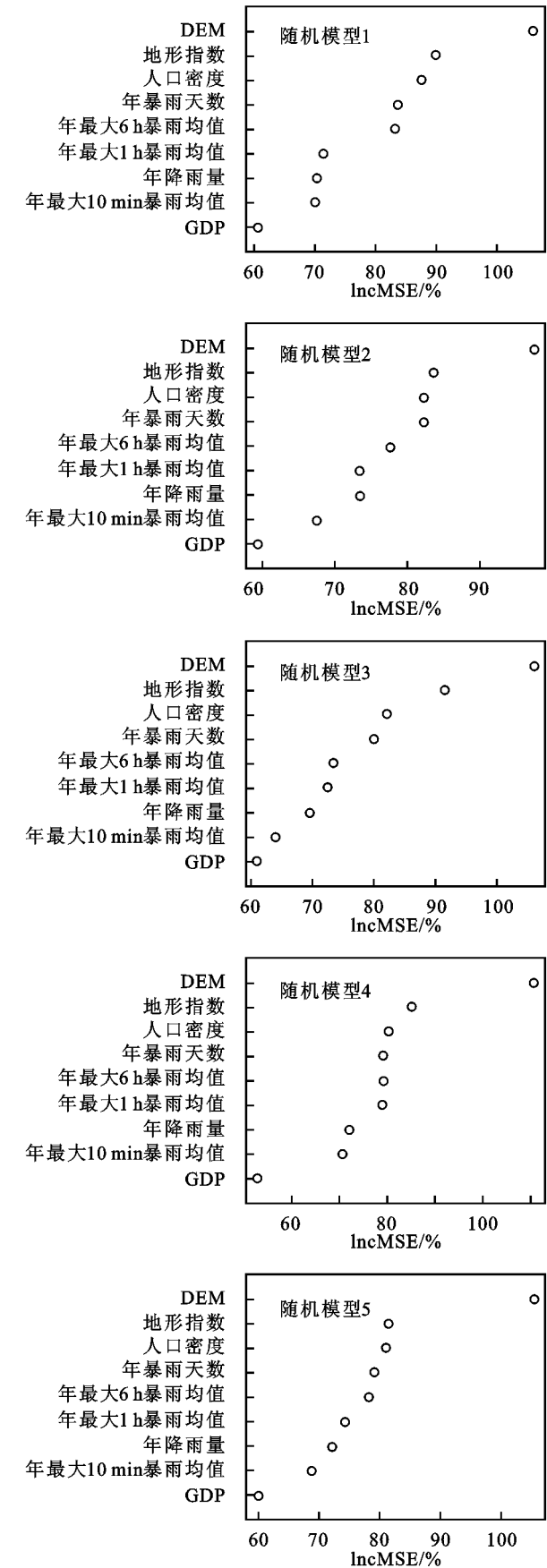


图 5 5 个随机森林模型中指标重要性

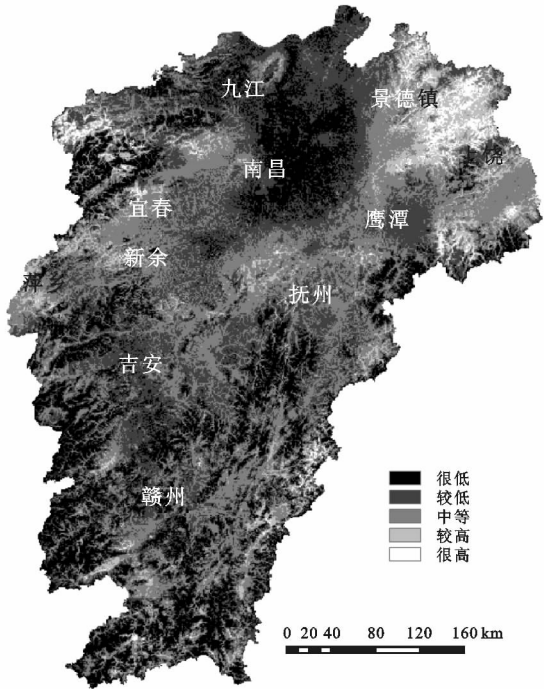


图 6 基于随机森林的江西省山洪灾害风险分区

把山洪灾害风险区划图和历史发生山洪点进行叠加,最终在每个风险等级上与若干灾害统计数据发生重合,计算各个风险等级上重合的点数相对于总点数的占比,见表 3。从表 3 中可以看出,与风险等级中等及以上的区域重合的历史发生山洪灾害点数占总点数的 86.96%。而与风险等级很低的区域重合的历史发生山洪灾害点多数处于河流发源地,人烟稀少。验证结果表明本研究最终得到的山洪灾害风险等级区划精度较高。

表 3 历史发生山洪灾害点分布统计表

风险等级	历史灾害点数	所占百分比/%
很低	257	12.79
较低	5	0.25
中等	104	5.18
较高	435	21.65
很高	1208	60.13

综合分析,我们发现随机森林在山洪灾害的评估方面的精度完全满足我们的研究需要,为下一步对江西省山洪灾害等级分区统计提供了理论支持。

3.4 山洪灾害风险等级分区统计与分析

将基于随机森林得出的山洪灾害风险等级图进行地级市分区的综合统计分析。分别统计出在不同的随机森林模型中,各地级市分区的 5 个山洪灾害风险等级下所占面积的百分比和面积绝对值。

从图 7 中可以发现,南昌市中山洪风险等级为很低的面积占 50%左右,风险等级为较低的面积也占了将近 40%,而风险等级为较高和很高所占的面积几乎为 0,所以在南昌市范围内可能发生山洪灾害的区域几乎没有;在赣州、九江、鹰潭和吉安等市,山洪灾害风险等级为很低和较低所占的面积大概在 60%~70%,而风险等级为

较高和很高的面积总共占了不到10%。

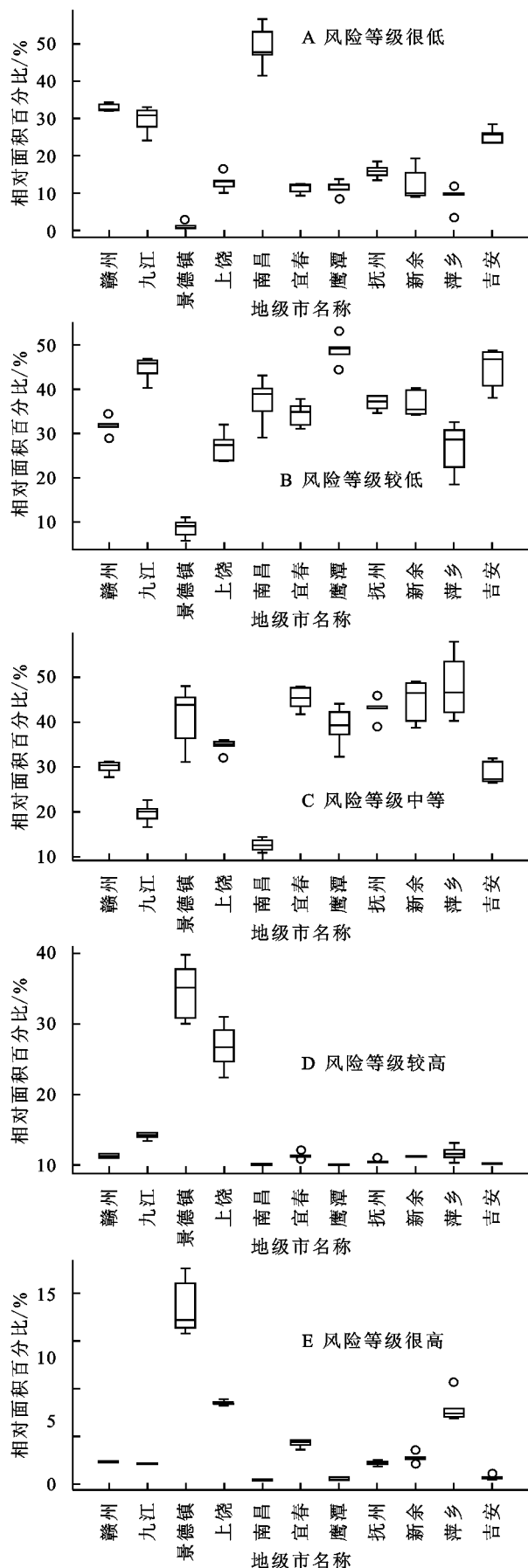


图7 各地级市中不同风险等级下所占面积百分比

因此在这些地级市只有小部分区域可能会有山洪灾害的发生;在宜春、抚州、新余和萍乡市,虽然山洪灾害风险等级为很低和较低所占的面积在40%左右,且风险等级为较高和很高的面积只占了10%左右,但这些地级市的大部分区域处在的风险等级为中等,因此大部分区域还是有山洪灾害发生的可能的;而在景德镇市和上饶市,山洪灾害风险等级为较高和很高所占的面积比值是所有地级市中最高的,并且风险等级为中等的面积也占了30%以上,尤其是景德镇市,风险等级为较高和很高的面积占了40%以上,风险等级为很低和较低的面积占了不到15%,所以绝大部分的区域都是有山洪灾害发生的可能。

山洪灾害风险等级为很高的区域面积绝对值中,上饶市是最大的,大概在1 500 km²以上,而山洪灾害风险等级为较高的区域面积绝对值中,上饶市也是最大的,接近4 000 km²;在景德镇市,山洪灾害风险等级为较高和很高的面积绝对值也比较大,很高的面积绝对值在500 km²以上,较高的面积绝对值在2 000 km²左右;与其他市相比,九江市风险等级很高的面积绝对值比其他市的稍微高一点,赣州市、宜春市的风险等级为较高的面积绝对值是比较大的,接近2 000 km²,但赣州市风险等级很低、较低和中等的区域面积绝对值都在12 000 km²左右;风险等级为中等的面积绝对值中,上饶市、宜春市、抚州市和吉安市是比较大的,都在8 000 km²左右;而南昌市、鹰潭市、新余市和萍乡市可能因为市面积小,基本各个风险等级上的面积绝对值都不到2 000 km²。

综合分析发现,在景德镇市和上饶市,山洪灾害风险等级为很高和较高的区域不仅面积占比大,而且所占面积绝对值也很大;在赣州市和宜春市,山洪灾害风险等级为很高和较高的面积占比虽然比较小,但由于市面积很大,所以风险等级为很高和较高的面积绝对值都还是比较大的;而在南昌市和鹰潭市,山洪灾害风险等级为很高和较高的面积百分比很小,面积绝对值也是同样的很小的。

4 结论

(1) 随机森林算法大大提升了计算速度,较好地解决非线性问题,基于随机森林的山洪灾害模型为山洪灾害的智能化评价提供了一种有效的途径;(2) 将历史山洪灾害点叠加到基于随机森林构建的风险评估模型得出的风险分区图上,统计每个风险等级上历史发生山洪灾害点的个数,验证随机森林模型的精度

为86.96%,说明本文使用随机森林评价江西省山洪灾害等级具有较高的准确性;(3)经过统计,发现上饶市和景德镇市山洪风险等级较高和很高的面积绝对值较大,说明该区域发生山洪可能性较高造成损失较大,须预先做好相应的防洪和抗洪措施。其次是赣州、九江和宜春等地区风险等级很高的面积绝对值也很大,政府亦需给予相应重视;(4)各地级市中不同风险等级下所占面积百分比来看,等级很高和较高所占面积百分比最多的依然是上饶市和景德镇市,其次是宜春、九江、赣州和萍乡等地区,其他区域山洪灾害风险水平大多处于风险较低的等级,应以灾害防治为主,同时也应加强灾害监测预警预报工作。

参考文献:

- [1] 管珉,陈兴旺. 江西省山洪灾害风险区划初步研究[J]. 暴雨灾害,2007,26(4):339-343.
- [2] 赵士鹏. 中国山洪灾害系统的整体特征及其危险度区划的初步研究[J]. 自然灾害学报,1996(3):93-99.
- [3] 马建华,胡维忠. 我国山洪灾害防治形势及防治对策[J]. 人民长江,2005,36(6):3-5.
- [4] 孟东勇,秦亚丽. 山洪灾害防治非工程措施项目建设探讨[J]. 江西建材,2014(6):82-82.
- [5] 李林涛,徐宗学,庞博,等. 中国洪灾风险区划研究[J]. 水利学报,2012,43(1):22-30.
- [6] Woodruff J D, Irish J L, Camargo S J. Coastal flooding by tropical cyclones and sea-level rise [J]. Nature, 2013,504(7478):44-52.
- [7] Hallegatte S, Green C, Nicholls R J, et al. Future flood losses in major coastal cities[J]. Nature Climate Change, 2013,3(9):802-806.
- [8] 刘小平,黎夏,叶嘉安,等. 利用蚁群智能挖掘地理元胞自动机的转换规则[J]. 中国科学:地球科学,2007,37(6):824-834.
- [9] Martens D, Backer M D, Haesen R, et al. Classification with ant colony optimization[J]. Ieee Transactions on Evolutionary Computation, 2007,11(5):651-665.
- [10] 赵钢铁,杨大文,蔡喜明,等. 基于随机森林模型的长江上游枯水期径流预报研究[J]. 水力发电学报,2012,31(3):18-24.
- [11] Dong L J, Xi-Bing L I, Peng K. Prediction of rock-burst classification using Random Forest[J]. Transactions of Nonferrous Metals Society of China, 2013,23(2):472-477.
- [12] Tesfamariam S, Liu Z. Earthquake induced damage classification for reinforced concrete buildings[J]. Structural Safety, 2010,32(2):154-164.
- [13] Chen X, Ishwaran H. Random forests for genomic data analysis[J]. Genomics, 2012,99(6):323-329.
- [14] Mihaillescu D M, Gui V, Toma C I, et al. Computer aided diagnosis method for steatosis rating in ultrasound images using random forests[J]. Medical Ultrasonography, 2013,15(15):184-190.
- [15] 赖成光,陈晓宏,赵仕威,等. 基于随机森林的洪灾风险评估模型及其应用[J]. 水利学报,2015,46(1):58-66.
- [16] Wang Z, Lai C, Chen X, et al. Flood hazard risk assessment model based on random forest[J]. Journal of Hydrology, 2015,527:1130-1141.
- [17] Feng Q, Liu J, Gong J. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier: A Case of Yuyao, China[J]. Water, 2015,7(4):1437-1455.
- [18] 刘筱琴,李昆. 江西省暴雨洪水地理变化规律研究[R]. 华东七省(市)水利学会第二十次学术研讨会,2007.
- [19] 李世勤,邱启勇,王述强. 江西山洪灾害防治实践及思考[J]. 中国水利,2012(3):51-54.
- [20] 李明辉,熊剑英. 江西省山洪灾害防治规划概述[J]. 江西水利科技,2005,31(2):73-77.
- [21] 林俊. 江西省山洪预警系统分析与设计[D]. 昆明:云南大学,2012.
- [22] 杨妩,柏林, YANG W,等. 浅谈江西省山洪灾害防治监测、通信和预警系统的规划[J]. 江西水利科技, 2006,32(2):83-85.
- [23] 程晓陶. 城市型水灾害及其综合治水方略[J]. 灾害学, 2010(S):10-15.
- [24] 王全才. 随机森林特征选择[D]. 辽宁大连:大连理工大学,2011.
- [25] 吴喜之. 应用回归及分类:基于 R[M]. 北京:中国人民大学出版社,2016.
- [26] 尼娜·朱梅尔,约翰·芒特, Nina Zumel,等. 数据科学:理论、方法与 R 语言实践[M]. 北京:机械工业出版社,2016.
- [27] 李欣海. 用 R 实现随机森林的分类与回归[R]. 北京:第五届 R 语言会议,2012.
- [28] 赵北庚. 基于 R 语言 randomForest 包的随机森林建模研究[J]. 计算机光盘软件与应用,2015(2):152-153.
- [29] 方秀琴,王凯,任立良,等. 基于 GIS 的江西省山洪灾害风险评估与分区[J]. 灾害学,2017,32(1):111-116.