

基于数据挖掘方法的玉米和油菜土地适宜性分析

——以贵州省遵义市为例

张晨¹, 林艳², 周华^{1,3}

(1.贵州省自然资源勘测规划研究院, 贵州 贵阳 550004; 2.贵州大学 生命科学学院, 贵州 贵阳 550025; 3.自然资源部 土地利用重点实验室贵州科研基地, 贵阳 贵州 550004)

摘要: [目的] 开展多作物土地适宜性分析, 为农业生产的作物规划、土地利用规划和管理提供一种准确有效的方法。[方法] 基于数据挖掘方法, 综合运用随机森林算法、综合指数评价法、空间约束多元聚类和空间统计等多种方法相结合, 定量刻画分析了多作物种植土地综合适宜性。[结果] ①随机森林算法能定量精准实现多空间要素数据驱动下作物种植空间布局, 且能识别出对玉米和油菜种植土地适宜区选择具有关键影响的因子; ②研究区内玉米和油菜种植适宜区具有显著的空间异质性, 玉米适宜区占种植土地总面积的91.23%, 但油菜适宜区仅占69.64%; 同时适宜于二者种植的主要区域占13.08%, 主要分布在凤冈县中部及北部、湄潭县北部和余庆县西北部。[结论] 数据挖掘能够为种植土地利用方式选择提供最优方案的可能性, 这一方法用于种植土地适宜性评价具有很好的通用性。

关键词: 数据挖掘; 土地适宜性评价; 耕地资源; 随机森林算法; 作物; 遵义市

文献标识码: B

文章编号: 1000-288X(2022)03-0188-11

中图分类号: P968, TP391

文献参数: 张晨, 林艳, 周华. 基于数据挖掘方法的玉米和油菜土地适宜性分析[J]. 水土保持通报, 2022, 42(3): 188-198. DOI: 10.13961/j.cnki.stbctb.20220518.002; Zhang Chen, Lin Yan, Zhou Hua. Analysis on land suitability for maize and rapeseed production based on data mining method [J]. Bulletin of Soil and Water Conservation, 2022, 42(3): 188-198.

Analysis on Land Suitability for Maize and Rapeseed Production Based on Data Mining Method —A Case Study at Zunyi City, Guizhou Province

Zhang Chen¹, Lin Yan², Zhou Hua^{1,3}

(1. Guizhou Institute of Natural Resources Survey and Planning, Guiyang, Guizhou 550004, China; 2. College of Life Sciences, Guizhou University, Guiyang, Guizhou 550025, China; 3. Guizhou Research Base of Key Laboratory of Land Use, Ministry of Natural Resources, Guiyang, Guizhou 550004, China)

Abstract: [Objective] Conducting multi-crop land suitability analysis was carried out in order to provide an accurate and effective method for crop planning, land use planning, and management of agricultural production. [Methods] Based on data mining method, the comprehensive suitability of multi-crop land was quantitatively characterized by a random forest algorithm, a comprehensive index evaluation method, spatially constrained multivariate clustering, and spatial statistics. [Results] ① A random forest algorithm quantitatively and accurately produced the spatial layout of crop planting driven by multi-spatial elements, and identified factors that had a key influence on the selection of suitable planting land for maize and rapeseed. ② The suitable areas for maize and rapeseed planting exhibited significant spatial heterogeneity in the study area. The suitable areas for maize accounted for 91.23% of the total planting land area, but the suitable areas for rapeseed accounted for only 69.64%. The areas suitable for the planting of both crops was 13.08% of the total area, and were mainly located in the central and north parts of Fenggang County, the north part of

收稿日期: 2021-12-01

修回日期: 2021-12-05

资助项目: 自然资源部碳中和与国土空间优化重点实验室开放资助项目“喀斯特生态脆弱区碳源/汇时空变化特征及低碳目标下国土空间格局优化研究: 以贵州省为例”(2021CNT02004)

第一作者: 张晨(1993—), 男(苗族), 贵州省六盘水市人, 硕士, 研究方向为空间数据挖掘、空间分析与建模。Email: chzh1812zfly@163.com.

通信作者: 周华(1980—), 男(汉族), 贵州省贵阳市人, 博士, 主要从事国土空间规划、用途管制研究。Email: 34092739@qq.com.

Meitan County, and the northwest part of Yuqing County. [Conclusion] Data mining provides the possibility of selecting the optimal planting land use pattern. The method used in this study has good potential for evaluating the suitability of land in many locations for the production of various crops.

Keywords: data mining; land suitability evaluation; cultivated land resources; random forest; crops; Zunyi City

为了进一步落实国家粮食安全战略,可持续利用农业土地资源,综合考虑自然资源禀赋和了解当地自然环境条件的主要局限性可以帮助我们选择特定的作物种植,可以为决策者创建最佳的作物种植决策管理^[1-2]。气候、地形、土壤等自然环境条件和土地用途的不同,土地对特定作物的适宜性会呈现差异性,有必要根据最有效的自然环境条件因素对土地单元进行分类和管理,而土地适宜性评价(land suitability assessment, LSA)是促进农业可持续发展和进行科学的土地利用规划的关键途径^[3],科学开展 LSA 在全面推进乡村振兴加快农业农村现代化中起着十分重要的作用,是保障乡村振兴战略实施的重要举措。

近 10 a 来,LSA 在土地利用这一核心主题下围绕评价对象、评价理论、评价技术方法和评价数据等方面开展研究。在评价对象方面,开展蚕豆^[1]、小麦^[3-5]、油菜^[2]、烤烟^[6]、高粱^[7-8]、玉米^[8]等多种作物 LSA 研究。在 LSA 理论方面,大致可以分为空间叠加^[9]、多指标决策模型^[4]、人工智能^[5]等 3 种方法,其中,空间叠加使用较多的是加权叠加^[6,10],多准则决策模型最常用层次分析法^[11-13]、主成分分析法^[14]等,人工智能则尝试使用支持向量机和随机森林^[5]、人工神经网络^[15]等。在 LSA 技术方面,主要是借助遥感、地理信息系统技术手段开展评价研究工作,其在改善土地适宜性评价方面具有很大潜力^[3]。在 LSA 数据方面,涉及到矢量数据、栅格数据或矢量—栅格混合数据多种空间数据类型^[16-17]。综上所述,当前,最为常用的 LSA 理论技术方法是在遥感、地理信息系统技术手段支持下,使用表征气候、地形、土壤等自然环境条件多源多维空间数据和空间叠加、多准则决策模型等进行 LSA 建模分析,但在数据挖掘视角下开展多作物 LSA 的研究成果相对较少。此外,随着空间数据采集技术不断革新使得量化自然环境因素的有效数据大量存在,致使 LSA 越来越复杂,现有农业资源与农业生产活动不匹配可能导致粮食供求关系失衡,合理的农业生产作物规划在资源投入获取与分配中起着重要的作用^[7-8]。玉米和油菜分别是遵义市粮食产量第二和油料产量第一的作物,在连续种植玉米或玉米休耕的地区,油菜可以作为一种有价值的轮作作物出现^[18],探讨该地区主要作物玉米、油菜等农业生产作物规划,有利于提高玉米和油菜的产量以改善

粮食安全。因此,如何把握当前数据驱动的战略机遇,突破 LSA 多源多维属性数据认知与分析的困境,构建更加综合、科学与智能的多作物 LSA 是值得研究的。

鉴于此,本研究以科学布局作物种植空间,合理利用农业土地资源作为切入点,思考这种自然环境条件差异性对多作物种植空间布局的影响。在此基础上基于数据挖掘方法,引入机器学习算法随机森林构建气候、地形、土壤等多源多维属性数据与不同作物种植空间之间的关联模型,试图开展多作物土地适宜性分析,为农业生产的作物规划、土地利用规划和管理提供一种准确有效的方法。

1 研究区概况

研究区位于贵州省北部遵义市,东经 105°36′—108°12′,北纬 27°08′—29°14′,总面积 3.08×10^6 hm²,占贵州省总面积的 17.47%。东邻铜仁市,南抵黔东南州、黔南州和贵阳市,西南与毕节市接壤,西与四川省交界,北连重庆市,是大西南通江达海的重要通道。地处云贵高原向湖南丘陵和四川盆地过渡的斜坡地带,地形起伏较大,海拔 210~2 208.13 m,平均海拔 984.51 m。地貌类型复杂,境内大娄山山脉及其支脉呈东北向斜贯本市中部,是市内南北水系的分水岭,遵义市被划分为两大片,山南以低中山丘陵和宽谷盆地为主,一般耕地比较集中连片,土地利用率高。山北以中山峡谷为主,山高谷深,山地垂直差异明显,耕地比较分散。气候属于中亚热带季风湿润气候区,终年温凉湿润,气候温和,全年日均气温 4~32℃,日照充足,雨水充沛,年均降水量 1 097.8 mm。土壤类型多样,低山丘陵盆地区主要分布着以酸性黄壤为主,有机质含量高。地质地貌条件复杂,降雨分布不均,水分的区域性和时间性变化很大,对土地利用与土壤的形成均有较大的影响。

遵义市在贵州省国土空间总体规划(2021—2035 年)中被划定为 5 类农业现代化复合发展分区之一的黔北农业区,是贵州省重要农产品生产基地和粮食、油料作物的主要产地。科学合理布局作物种植空间是保障现代山地特色高效农业强省建设以及持续推进农业现代化当中不可或缺的关键环节。开展多作物生产的土地综合适宜性评价实地量化研究既是亟待解决的科学问题,也是落实乡村振兴重大国家战略的具体措施。

2 数据来源与研究方法

2.1 数据来源及预处理

本研究所用矢量数据来源包括遵义市自然资源局所提供的 2020 年耕地资源质量分类数据库成果, 2020 年贵州省地理国情监测成果地表覆盖数据集农作物种植样点专题图层 V_SPCP, 及市行政区划; 栅格数据来源包括数字高程模型采用中科图新 (<http://www.tuxingis.com/store.html>) 提供的 12.5 m 空间分辨率 DEM 数据, 国家青藏高原科学数据中心提供的 2017 年中国区域融合日照时数 10 km 高分辨率地表太阳辐射数据集, 国家地球系统科学数据中心提供的 2012 年中国 0.083 33° 分辨率逐月土壤有效水分数据集, 中国科学院资源环境科学与数据中心提供的 500 m 空间分辨率 ≥ 10 °C 活动积温 (经 DEM 校正) 数据集, 以及贵州省气象局提供的 30 a (1989—2018) 30 m 空间分辨率年均气温和年均降水量栅格数据。为便于叠加分析, 所有空间数据坐标系统统一转换成 CGCS2 000_3_Degree_GK_Zone_26。

提取耕地资源质量分类数据库成果中的耕地和种植属性标注为“即可恢复”和“工程恢复”的相关地类图斑作为评价单元, 共计 1 153 748 个, 涵盖种植土地范围。在作物种植 LSA 中, 表征地形特征的坡度、海拔由 DEM 数据生成, 其中坡度采用百分比坡度度量; 除了土壤有效水分之外, 表征土壤特性的土层厚度、土壤质地、有机质含量、土壤 pH 值取自耕地资源质量分类数据库成果, 其中土壤质地是非数值型数据, 这里依据国务院第三次全国土地调查领导小组办公室印发的《第三次全国土地调查耕地资源质量分类工作方案》进行数值化, 1, 2, 3 分别表示壤质、黏质、砂质; 表征气候条件的因素有活动积温、年均降水量、年均气温和日照时数。评价单元相关属性赋值计算时均将栅格数据重新采样至 5 m \times 5 m 空间分辨率, 然后使用分区统计叠加计算获得评价单元属性数据; 玉米和油菜种植采样点是采用国土调查云手机应用软件 App 据实调查, 分别有 171 个和 127 个。

2.2 研究方法

本研究总体研究思路是紧扣影响作物种植气候、地形和土壤 3 大类自然环境条件, 在作物适宜性生长因子约束下构建起作物种植空间适宜性评价因子特征数据集; 在既定的评价单元内使用随机森林算法建立作物与评价因子特征数据集之间的关联模型, 得出作物种植空间适宜性评价结果及其评价因子贡献度; 在此基础上, 利用 GIS 技术空间化作物种植空间适宜性评价结果以分析作物种植空间布局特征; 综合

运用综合指数评价法、空间约束多元聚类 and 空间统计研究作物适宜性空间等级特征及异质性, 进而使用空间叠加分析法实现多作物种植空间综合适宜性分析。研究思路详见图 1。

研究思路中所采用的随机森林算法是一种在 Bagging 策略的基础之上通过引入双随机性来实现次优分割的有监督机器学习算法^[19], 其对特征贡献度 I 评估是使用袋外数据计算每棵分类树的袋外误差 E , 然后对袋外数据第 j 个特征 X^j 随机加入噪声, 并计算新的袋外误差 $X^j(t)$, 则特征变量 X^j 的贡献度定义为^[20]:

$$I(X^j) = \frac{1}{n} \sum_{t=1}^n [E^j(t) - E_t] \quad (1)$$

若特征变量 X^j 的变化引起袋外误差增加越大, 模型准确度就会大幅度降低, 则说明这一特征变量对于样本分类结果影响就很大, 也就是说它的贡献度比较高^[21]。用于作物种植 LSA 的样本由两个部分组成: ①根据外业实地调查获得的标记样本; ②以作物种植土地适宜性限制性因素阈值作为依据^[2, 8] (表 1), 根据每一个限制性因素阈值构成条件语句筛选得到多组不适宜种植标记样本, 多组不适宜种植标记样本协同验证, 取其空间交集作为最终作物不适宜种植样本, 以提高不适宜样本标记的可靠性。

表 1 玉米和油菜种植土地适宜性限制性因素

土地适宜性限制性因素	玉米	油菜
土层厚度/cm	<25	—
土壤 pH 值	<5.5, >9.0	<5.0, >8.5
年均气温/°C	<15, >28	<10, >35
海拔高度/m	—	>1 000

此外, 使用的综合指数评价法是在确定的评价指标基础之上, 将指标数据无量纲化, 并对每个指标赋予一定的权重, 然后进行无量纲化指标数值加权求和, 进而得到综合指数^[22], 其计算公式如下:

$$ECI = \sum_{i=1}^n w_i \text{Index}_i \quad (2)$$

式中: ECI 表示评价系统综合指数; i 表示评价指标; w_i 表示评价指标对应的权重; Index_i 表示评价指标数值。指标无量纲化处理采用最大最小值归一化法进行指标数据无量纲化处理^[23], 其计算公式如下:

$$r_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \quad (3)$$

式中: r_{ij} 表示指标数值无量纲化结果; x_{ij} 表示 i 项指标第 j 项数值。结合随机森林算法给出的每一个评价因子贡献度进而计算得到 LSA 综合指数, 以此来研究作物适宜性空间分异规律。

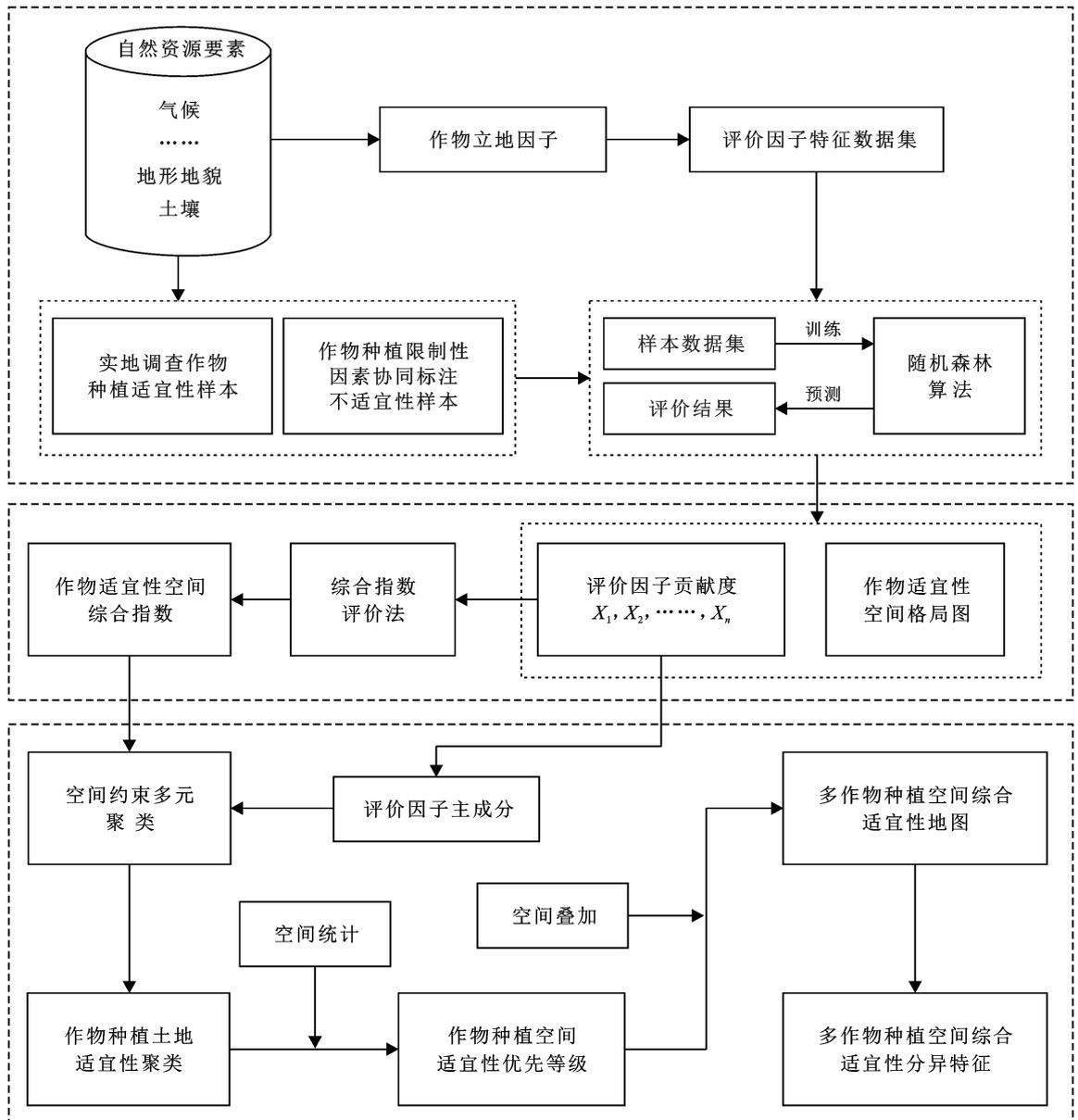


图 1 多作物种植空间适宜性评价技术路线图

2.3 作物种植土地适宜性评价因子

气候条件对于种植土地的开发和利用影响最大,气候影响着能够种植什么作物,一年能够种植几茬^[24],温度是作物生长发育的重要条件,需要在一定的温度条件下才能进行;活动积温为作物在其全生长期内提供热量需求,只有当温度累积到一定的总和时才能完成其发育周期;日照时数是太阳辐射因子当中影响作物光合作用的主要因素,光照时间长度对光合作用产物有着重要影响;此外,降水量作为作物生长发育所需水分主要来源,在相同的年均降水量下,降水变率越大,作物生产的稳定性越低。作物实际产量除了受气候条件的影响外,还受到地形和土壤等影响。地形影响着物质与能量的再分配,尤其山区地形对光照、热量、水分以及土壤肥力等起着再分配作用,

对作物的分布和生长有很大的影响^[2,25]。土壤为作物提供所必需的生存空间直接影响着作物的生长发育状况^[26],深厚土层提供作物扎根良好条件,对养分和水分的保蓄能力强;土壤质地则关系到土壤的物理性质以及蓄水保肥能力;作物养分供给的来源之一是土壤有机质,对于保持土壤良好的物理性质具有重要作用;土壤 pH 值对不同作物生长的影响程度不同,酸碱度过高或过低都会影响作物养分吸收,加速养分流失,使土壤失去耕种价值;土壤有效水分不足会抑制作物的光合作用,还会影响作物产量品质。顾及到数据的可获取性,选取气候、地形、土壤 3 个大类 11 个对作物种植具有重要影响因素构成 LSA 特征数据集(表 2)。

表 2 作物种植土地适宜性评价因子^[2,24-26]

要素	评价因子
气候	年均气温、活动积温、日照时数、年均降水量
地形	海拔、坡度
土壤	土层厚度、土壤质地、有机质含量、土壤 pH 值、土壤有效水分

3 结果与分析

3.1 玉米和油菜种植适宜性空间评价

将玉米和油菜种植样点所在的图斑标注适宜性标签“1”，限制性因素协同标注图斑的空间交集标注不适宜性标签“0”，由此分别生成 LSA 样本集，按照 7:3 划分为训练数据集和测试数据集，其余图斑作为预测数据集。在使用随机森林算法进行作物 LSA 过程中，需要设置关键参数决策树棵数($n_{\text{estimator}}$)和保证模型达到最优时的最大特征数(\max_{feature})。本研究使用网格搜索法和交叉验证法进行参数寻优，通过反复迭代试验当 $n_{\text{estimator}}$ 分别为 63,50 时误差逐渐收敛并趋于稳定，这时的 \max_{feature} 设置为总特征数 n 的平方根，二者模型输出的泛化误差分别为 97.82% 和 98.14%。预测数据集分别放入训练得到的模型便可得到玉米和油菜种植空

间适宜性评价结果(图略)。

参与作物种植 LSA 区域占研究区总面积的 30.28%。研究区有很大一部分土地适合玉米生产，占作物种植 LSA 区域的 91.23%，而油菜仅占 69.64% (表 3)，在研究区范围内玉米适宜种植区远远大于油菜，其中面积较差最大的桐梓县相差 33 286.61 hm^2 ，平均相差 15 741.02 hm^2 ，最小的湄潭县相差 5 327.26 hm^2 ，但只有靠近研究区西北方向的赤水市范围内油菜适宜种植区大于玉米，相差 3 499.62 hm^2 ，这一定量研究结果表明并不是参与作物种植 LSA 单元都适宜种植玉米和油菜，相比之下，在相同的自然环境条件下，玉米种植土地适宜性远比油菜高得多，农业土地资源对特定作物种植适宜性具有差异性^[2]。

此外，玉米和油菜适宜性呈现出局部集中和零散分布两种模式，适宜种植区在空间上具有一定局部集聚特征，这样的区域有利于发展规模性农业产业，而零散分布的适宜种植区仍然可以用于玉米和油菜的生产，但会造成生产成本过高。由表 3 可知，整体而言，在研究区适宜种植范围内，不管玉米还是油菜，赤水市适宜种植区过于分散，集中度最低，不建议用于生产。

表 3 玉米和油菜适宜种植面积及集中度

县(市)	玉米			油菜		
	面积/ hm^2	比例/%	集中度	面积/ hm^2	比例/%	集中度
红花岗区	47 110.14	5.54	2.27	41 689.58	6.43	2.20
汇川区	45 493.16	5.35	2.38	31 134.31	4.80	2.09
播州区	91 411.90	10.75	2.74	76 875.27	11.85	2.69
桐梓县	84 070.30	9.89	2.36	50 783.70	7.83	2.28
绥阳县	73 461.70	8.64	2.63	48 622.76	7.49	2.46
正安县	75 765.04	8.92	2.15	49 448.18	7.62	2.09
道真县	52 292.98	6.15	1.99	41 123.32	6.34	2.17
务川县	78 964.72	9.29	2.22	55 890.99	8.61	2.26
凤冈县	66 474.54	7.82	1.84	61 096.88	9.42	1.95
湄潭县	51 228.41	6.04	1.56	45 901.15	7.07	1.61
余庆县	49 723.20	5.85	1.97	42 355.91	6.53	2.02
习水县	61 729.23	7.26	1.67	42 470.65	6.55	1.85
赤水市	16 270.65	1.92	0.85	19 770.27	3.04	0.93
仁怀市	55 957.89	6.58	2.43	41 657.27	6.42	2.32
合计	849 953.86	100.00	2.09	648 820.24	100.00	2.06

注:集中度=面积/评价单元数,这里的评价单元数是指经过空间融合处理以后的图斑个数。

3.2 评价因子对玉米和油菜种植适宜性影响

通过随机森林算法得出评价因子适宜性贡献度，分析了 11 个评价因子对玉米和油菜适宜性影响程度(图 2)。①土壤质地(40.35%)是玉米种植土地适宜性影响最显著的因子，其次是活动积温(36.19%)，另外，坡度、海拔、年均气温和有机质含量的影响最小。对于玉米而言，图 2 中土壤性质的权重与

Kazemi 等^[1]对蚕豆生产提出的权重相当，玉米种植适宜区壤黏土比例高达 91.71%，这可能与壤黏质土壤能够为玉米床层的生长提供丰富土壤养分含量有关^[27]。除土壤性质外，气候参数活动积温对玉米种植适宜区选择具有重要的影响，这不仅与 ≥ 10 $^{\circ}\text{C}$ 积温条件对玉米的阶段发育有显著影响有关，而且和 ≥ 10 $^{\circ}\text{C}$ 积温利用密切相关，还与活动积温能够为玉米群体生

物量、叶面积提供热量资源有关,综合表现为随着热量资源增加,生育期明显缩短,生物量增加,单位叶面积的籽粒生产效率更高,玉米阶段发育与区域热量资源有效匹配是玉米获得高产和热量资源高效利用的前提^[28]。②影响油菜种植土地适宜性最显著的因子是年均气温(45.99%),其次是活动积温(36.19%)。此外影响最小的因子有日照时数和有机质含量,这一定量结果表明年均气温和活动积温是油菜种植适宜区选择的主要限制因素,它们在油菜全生育期起着重要的作用^[29]。除气候参数外,对油菜种植适宜区选择影响较大的是地形因子海拔(2.64%)和坡度(2.21%),其中坡度权重与 Wanyama 等^[30]对玉米生产提出的权重相当,随着海拔增加加大了结霜的可能性,在油菜生长月份的播种和莲座期范围内发生霜冻对油菜危害极大,并会造成产量损失^[18]。不论是玉米还是油菜,都有自己独特的标准来决定一个地区是否适宜生产。因此,很大程度上土壤质地和活动积温决定着玉米适宜区的确定,而油菜适宜区的确定基本上取决于年均气温和活动积温,在建立良好玉米和油菜土地适宜性评价模型过程中土壤性质和气候条件具有巨大的参考价值。

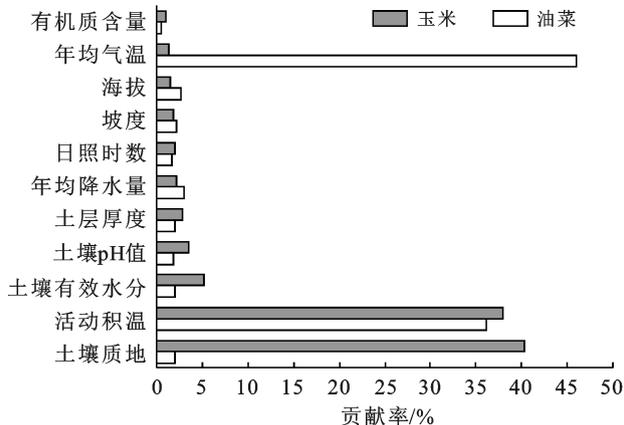


图2 玉米和油菜种植空间适宜性评价因子权重贡献度

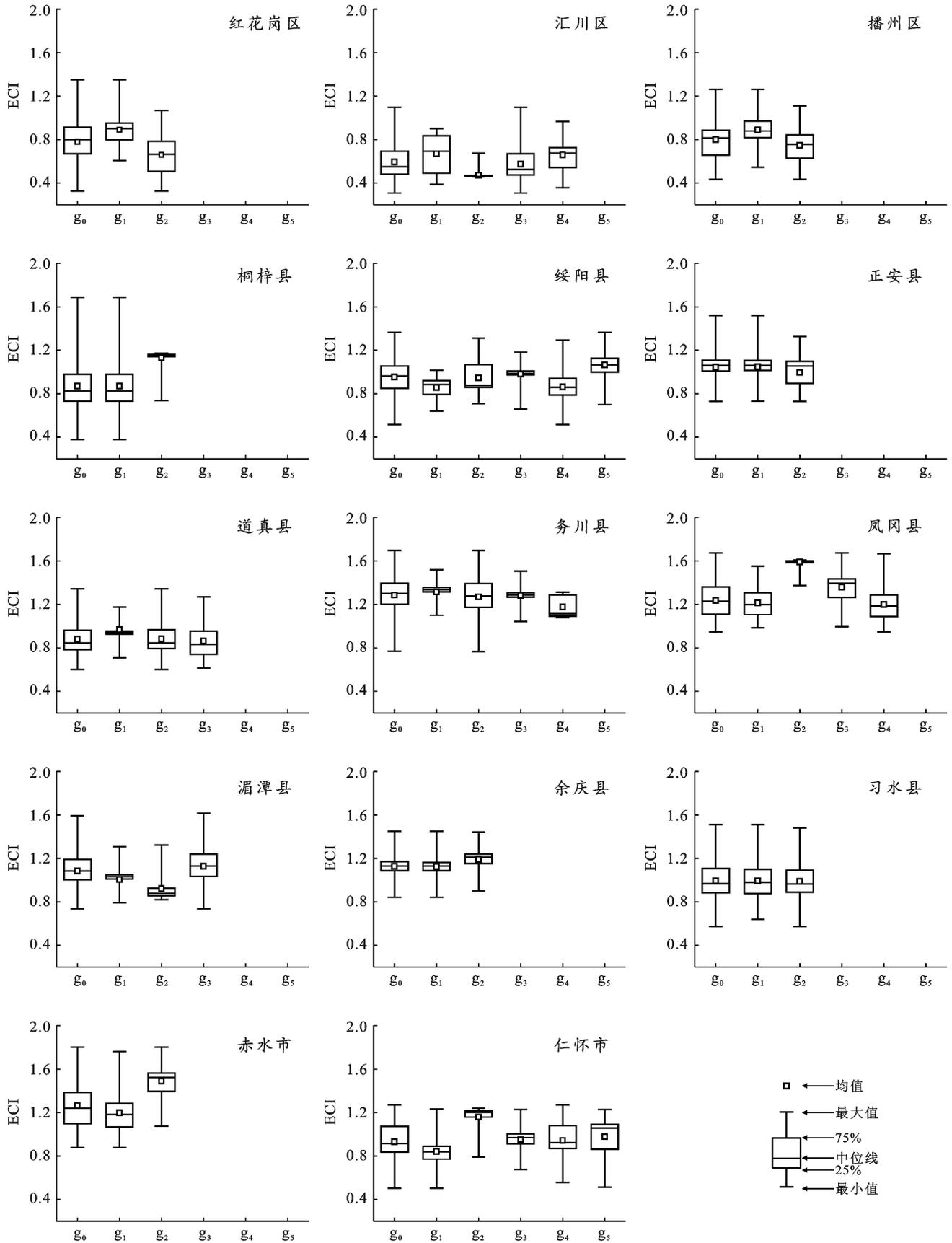
3.3 玉米和油菜种植适宜性空间等级划分

虽然有较大部分种植土地适合玉米和油菜的生产,但必须注意到这是较大尺度上的适宜,实际上局部地区适宜程度是存在差异的,不同适宜区生产玉米和油菜的产量是不同的。为了进一步揭示在更小尺度上玉米和油菜种植适宜程度的空间异质性,本文以 Delaunay 三角测量自然邻域关系作为空间约束,结合前文优选对玉米和油菜种植适宜性贡献度较大的主要评价因子和适宜性综合指数作为属性约束以增加空间聚类分组的确定性,使用空间约束多元聚类算法进行聚类分析,聚类有效性通过 Calinski-Harabasz 伪

F 统计量度量,探索评估最佳聚类数^[31],使用箱型图对聚类划分结果适宜性综合指数进行统计分析(如图3,4所示)。图3—4中的符号 g_{1-6} 分别代表聚类类别, g_0 代表县域全局。

由图3可以观察到,县域尺度上有3个地区的玉米种植适宜程度在空间上特别突出,分别是凤冈县、赤水市和务川县,其中赤水市综合指数差异最大,数值分布很广,绝大部分具有较低值,相反凤冈县大部分具有较高值。从种植适宜程度和空间上来看:①玉米种植适宜程度最高的区域位于研究区东部凤冈县(g_2),但仅占玉米总适宜面积的0.01%(48.39 hm^2),它们几乎看不见,这也意味着这个区域很小。②位于研究区西北部赤水市(g_2)是玉米种植高度适宜区,占玉米总适宜面积的0.40%(3 395.01 hm^2),虽然这一区域的玉米种植适宜程度较高,但是适宜区分布过于分散。③玉米种植适宜程度较高的区域位于研究区东北部务川县(g_1),占玉米总适宜面积的0.07%(608.58 hm^2)。结果表明确实是存在高度适宜玉米种植的区域,从经济上讲,受到集聚程度较小和适宜区分布较为分散的制约,用于玉米的生产将会造成成本过高,产量也很难保障,这样的区域不建议用于玉米生产。除此之外,紧随其后的分别是凤冈县(g_3)、赤水市(g_1)和务川县(g_2),占玉米总适宜面积2.21%(18 806.8 hm^2),1.51%(12 875.64 hm^2)和9.04%(76 842.49 hm^2),在空间上风冈县(g_3)适宜区呈现带状空间分布,务川县(g_2)分布较为密集,而赤水市(g_1)呈现出东北方向局部集聚和西北方向分散两种形态,相对之下,建议优先考虑务川县(g_2)用于玉米生产规划。玉米种植适宜程度最差的区域位于研究区中部汇川区,占玉米种植适宜总面积5.35%(45 493.16 hm^2)。

对比图3和图4可以观察到,在县域尺度上油菜种植适宜程度在空间上特别突出3个地区与玉米相同,这表明确实是存在高度适合玉米生长的地区也适合油菜生长,仅仅是适宜程度不同,这与 Muzira 等^[24]对玉米和高粱两种作物土地综合适宜性评价观点是一致的。从种植适宜程度和空间上来看:①最适宜油菜种植的区域赤水市(g_2)位于研究区西北部,仅仅占油菜总适宜面积的1.27%(8 233.44 hm^2),虽然这一区域的最适合油菜种植,但是同样存在空间分布过于分散的特征。②位于研究区东北部务川县(g_1)和务川县(g_1)油菜种植适宜程度相当,占油菜总适宜面积的3.93%(25 488.01 hm^2),适宜区域空间分布较为密集。③油菜种植适宜程度较高的区域是位于研究区东部凤冈县(g_2),占油菜总适宜面积的6.65%(43 171.50 hm^2),集中连片程度较好。



注:横坐标中 g_1 、 g_2 、 g_3 、 g_4 、 g_5 和 g_6 分别代表按照县域尺度聚类类别。 g_0 表示县域全局。下同。

图 3 玉米种植适宜性空间等级划分

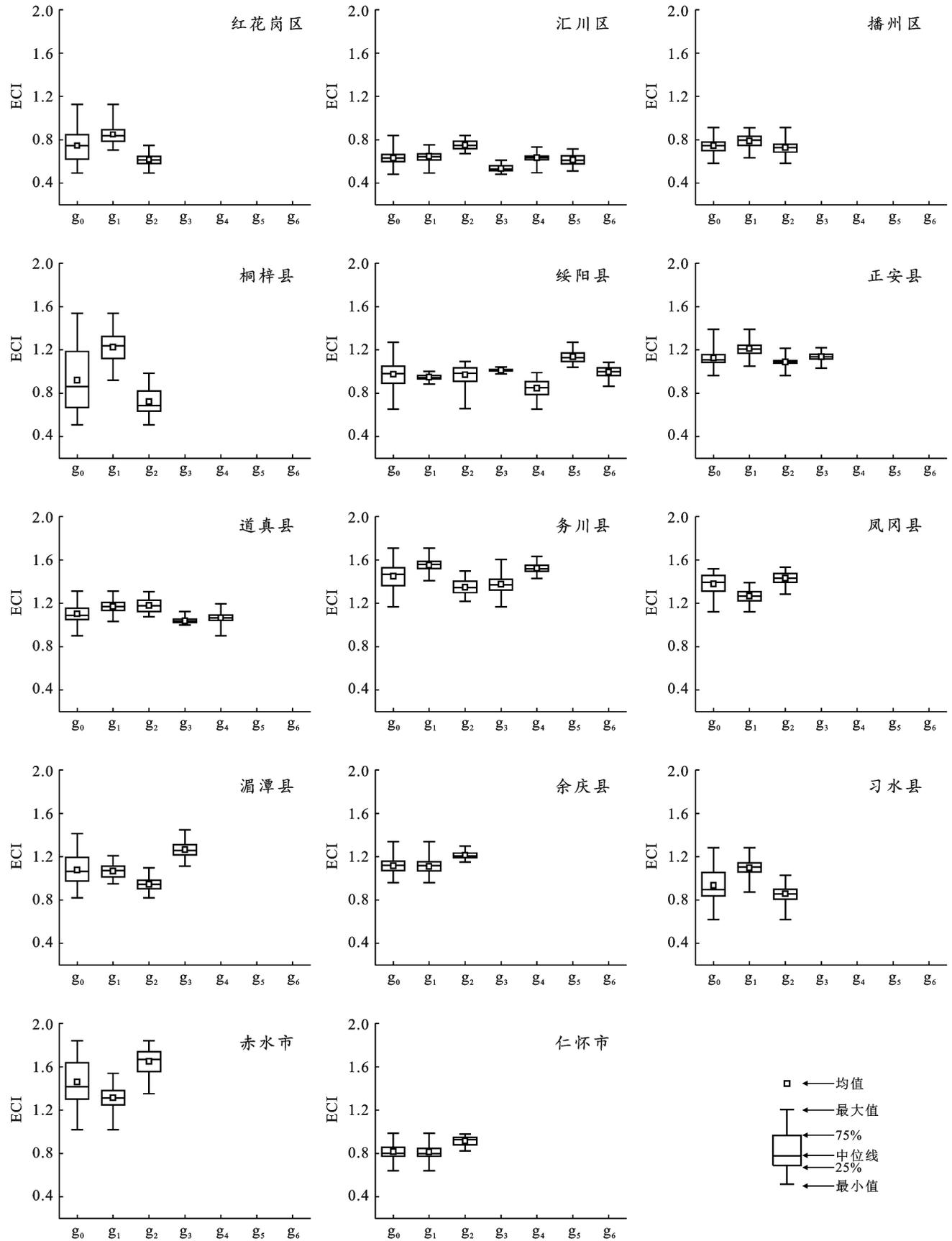


图 4 油菜种植适宜性空间等级划分

除此之外,紧随其后的分别是务川县(g_3)、务川县(g_2)和赤水市(g_1),占油菜总适宜面积 2.81% (18 259.16 hm^2), 1.87% (12 143.82 hm^2) 和 1.78% (11 536.83 hm^2), 同样地, 油菜种植适宜程度最差的也是汇川区, 占油菜总适宜面积的 4.80% (31 134.31 hm^2)。总体而言, 油菜高度适宜区空间分布比较密集, 务川县适宜性整体水平较高。

3.4 玉米和油菜种植空间综合适宜性

与 Muzira 等^[8]研究类似由于缺乏对作物适宜性水平进行分层的先验知识, 为了进一步了解玉米和油菜在种植空间上的综合适宜性, 这里在县域尺度上玉米和油菜种植空间等级划分结果之上, 采用 ArcGIS 10.2.2 数据分层自然断裂点法划分为 4 个等级, 进而采用空间叠加得到二者种植空间综合适宜性评价结果。在 4 个划分等级基础上对二者适宜性评价因子进行描述性统计(详见表 4)。从单一作物适宜程度

和空间分布来看: ①最适宜玉米种植的区域比例最小, 仅占总适宜面积的 2.59%, 主要集中分布于研究东部凤冈县境内; 而其中比例最大是中度适宜区, 比例达总适宜面积的一半以上, 空间分布基本上遍布整个研究区, 但主要分布在道真、桐梓、习水、正安县和仁怀市境内; 其次是高度适宜区, 比例达总适宜面积的 23.37%, 主要分布在绥阳、务川、凤冈、湄潭和余庆县境内; 轻微适宜区占总适宜面积的 12.99%, 主要分布在红花岗、汇川和播州区境内。②最适宜油菜种植的区域占总适宜面积的 13.46%, 主要分布在凤冈和务川县境内; 与玉米不同, 其中比例最大是高度适宜区, 占总适宜面积的 24.37%, 主要分布在正安、道真、湄潭、余庆县境内; 其次是中度适宜区, 占总适宜面积的 19.55%, 主要分布在播州区、绥阳县、习水县和仁怀市境内; 轻微适宜区占适宜面积的 13.25%, 主要分布在汇川区、播州区和桐梓县境内。

表 4 玉米和油菜适宜性评价因子描述性统计

评价因子	作物类型	平均值	标准差	变异系数	最小值	中位数	最大值
年均气温/ $^{\circ}C$	玉米	15.92	0.43	0.03	14.83	15.89	16.90
	油菜	15.94	0.37	0.02	14.83	15.95	16.90
活动积温/ $^{\circ}C$	玉米	5 174.42	380.04	0.07	3 897.90	5 018.40	6 537.00
	油菜	5 265.16	423.67	0.08	3 848.70	5 048.90	6 368.30
日照时数/($W \cdot m^{-2}$)	玉米	181.41	18.02	0.10	139.02	178.98	225.22
	油菜	176.64	17.40	0.10	139.02	176.73	225.22
年均降水量/mm	玉米	1 130.58	35.95	0.03	1 022.20	1 128.03	1 209.79
	油菜	1 118.43	38.83	0.03	1 045.79	1 112.04	1 209.79
海拔/m	玉米	847.87	205.62	0.24	210.00	853.28	1704.24
	油菜	775.74	169.48	0.22	210.00	800.00	1 267.61
坡度/%	玉米	32.75	16.93	0.52	0.00	31.12	408.65
	油菜	30.96	16.65	0.54	0.00	29.38	226.74
土层厚度/cm	玉米	80.83	29.35	0.36	15.00	80.00	261.00
	油菜	75.84	28.20	0.37	2.00	75.00	225.00
有机质含量/($g \cdot kg^{-1}$)	玉米	29.43	8.14	0.28	1.00	30.23	99.20
	油菜	29.28	7.81	0.27	1.00	29.86	105.00
土壤 pH 值	玉米	6.98	0.76	0.11	5.54	6.91	9.00
	油菜	6.73	0.86	0.13	5.25	6.62	9.00
土壤有效水分/cm	玉米	12.74	3.32	0.26	6.66	11.28	17.78
	油菜	12.22	3.27	0.27	6.47	10.79	19.53

注: 土壤质地是非数值型变量, 经统计分析主要适宜于玉米和油菜种植的土壤质地类型相同, 以壤质和黏质为主。

从多作物综合适宜程度和空间分布来看: ①对于单独适宜区而言, 玉米最适宜区仅有 0.14%, 空间零散分布在凤冈县东部; 高度适宜区有 6.76%, 主要分布在务川县西部; 中度适宜区有 16.40%, 主要分布在桐梓县西部、正安县东南和西北部以及习水县中东部; 轻微适宜区有 3.23%, 主要分布在汇川区中部及西南部和播州区西北部。而油菜最适宜区仅有

0.92%, 空间零散分布在务川县、凤冈县和赤水市整个县域境内, 分布范围跨度较大; 高度、中度和轻微适宜分别为 1.37%, 1.33% 和 0.13%, 空间零散分布在桐梓县等 12 个县境内。②对于二者种植适宜程度相当区域而言, 同时最适宜二者种植区域有 2.14%, 主要分布在凤冈县中部及北部; 同时高度适宜二者种植区域有 10.94%, 主要分在湄潭县北部和余庆县

西北部;同时中度适宜二者种植区域有 16.73%,主要分在仁怀市境内;同时轻度适宜二者种植区域有 8.92%,主要分在汇川区西部及东部、红花岗区西部和播州区中部。③玉米种植适宜程度优于油菜区域,最适宜区凤冈县西南部有 0.31%,高度适宜区主要是绥阳县南部有 1.20%,中度适宜区主要有桐梓县中部及西南部有 3.40%。④油菜种植适宜程度优于玉米区域,最适宜区主要有凤冈县东部有 3.03%和务川县境内绝大多数地区有 5.98%,高度适宜区主要是正安县和道真县绝大多数地区有 9.65%,中度适宜区分布在红花岗区与播州区交界处以及汇川区最东边仅有 0.84%。

与以往 LSA 研究不同^[1-2,8-9,18,22],以往研究是从整个研究区出发对选取评价因子分层分级划分适宜等级进而图层综合叠加得到作物适宜区,评价单元是整个研究区,而本研究事先将评价单元界定在种植地图斑内进行作物种植 LSA,由于受地形制约评价单元空间分布遍及整个研究区,此外,本研究结果也表明经评价后适宜玉米和油菜种植区基本上遍及整个研究区。由表 4 可知年均气温和土壤质地对于二者种植适宜区选择几乎没有区别,似乎与本文研究结果相矛盾,但同时适宜二者种植重叠区有 69.71%,二者评价因子适宜范围信息是来源于整个研究区,这只是较大尺度的适宜性,而本文对评价因子贡献度是基于微观尺度评价单元进行,能够很好揭示出评价因子对二者种植适宜区选择差异性。此外,表 4 结果表明海拔大于 1 267.61 m 是油菜生产的主要制约因素,油菜种植 pH 值适宜范围 5.25~9.00 这结果与 Ostovari 等^[2]研究结果基本一致,玉米种植 pH 值适宜范围与 Muzira 等^[8]研究结果基本一致,不同的是:无论玉米还是油菜适宜种植坡度范围都远超 Ostovari 等^[2,8,30]不适宜坡度百分比最小下界值 30%的研究结果,这主要是因为研究区地处喀斯特山区耕地资源以坡耕地为主导的。

4 结论

(1) 随机森林算法能够定量精准识别出多种农作物适宜种植空间。这一算法能够整合多源多维属性数据信息对农作物种植土地适宜性进行综合分析,探索出多因素综合作用下农作物种植空间资源最优生态位,挖掘农作物潜在的种植适宜空间,且在多种影响因素中能够识别出对农作物种植适宜空间确定起主导作用的影响因素。

(2) 农作物种植适宜性空间具有显著的空间异质性。研究区内玉米和油菜种植适宜区具有显著的

空间异质性,玉米主要适宜种植区域占总适宜面积的 25.96%,主要分布在务川、凤冈、湄潭、余庆和绥阳县境内,油菜主要适宜种植区域占总适宜面积的 37.83%,主要分布在务川、凤冈、湄潭、余庆、正安和道真县境内,同时主要适宜二者种植区域有 13.08%,主要分布在凤冈县中部及北部、湄潭县北部和余庆县西北部。

(3) 数据挖掘能够为种植土地利用提供多方案决策。随机森林算法对不同农作物种植 LSA 结果表明对种植土地利用并不是唯一的,这不仅为区域种植土地利用提供了多方案选择,而且不同农作物样本对同一种种植土地适宜性程度评价结果可能相同,这也为选择适合种植土地利用的最优方案提供了可能,提高了这一方法的通用性。

光照、温度、水分、土壤等自然条件变化是十分复杂,而 LSA 是针对特定用途对区域资源质量进行综合评定,在这些复杂因素的综合影响下 LSA 是一个复杂的科学问题,具有不断发展演化的动态特征,其研究属性亦呈现出跨学科、多元化特征,涉及较广泛的学科理论范畴。本文综合运用机器学习理论、生态学理论、空间统计及空间分析等理论、模型和技术手段在数据挖掘视角下对多作物土地适宜性进行初步探讨,可为区域农业用地适宜性评价及优化、明确种植土地利用优先级、建立科学合理耕地轮作休耕制度提供新的视角和思路。开展山区种植 LSA 涉及较多纷繁复杂的因素,未来需在评价指标体系的精细化、动态性等方面持续完善。例如,结合乡村振兴战略的现实要求,如何妥善考量社会经济要素对种植 LSA 的内在影响,值得深入研究。此外,遵义市作为典型的喀斯特生态环境脆弱区,除了喀斯特地区外还有其他地区,不同地区对特定作物土地适宜性评价因子选择上是否具有差别,需要进一步开展不同生态脆弱区下评价因子选择机制的研究,这也是构建科学合理 LSA 体系的关键环节之一。科学 LSA 评价体系对于优化国土空间格局,构建低碳排放的土地利用结构以及构建土地生态安全格局具有重要作用,这既是人与自然相互作用、相互影响这一地理学经典命题的应有之义,又是喀斯特生态环境脆弱区空间韧性研究等新领域的基础性科学问题。

[参 考 文 献]

- [1] Kazemi H, Sadeghi S, Akinci H. Developing a land evaluation model for faba bean cultivation using geographic information system and multi-criteria analysis: A case study: Gonbad-Kavous region, Iran [J]. Ecological Indicators, 2016,63:37-47.

- [2] Ostovari Y, Honarbakhsh A, Sangoony H, et al. GIS and multi-criteria decision-making analysis assessment of land suitability for rapeseed farming in calcareous soils of semi-arid regions [J]. *Ecological Indicators*, 2019, 103:479-487.
- [3] El Baroudy A A. Mapping and evaluating land suitability using a GIS-based model [J]. *Catena*, 2016, 140:96-104.
- [4] Keshavarzi A, Sarmadian F, Heidari A, et al. Land suitability evaluation using fuzzy continuous classification: A case study: Ziara region [J]. *Modern Applied Science*, 2010, 4(7): 72-81.
- [5] Taghizadeh-Mehrjardi R, Nabiollahi K, Rasoli L, et al. Land suitability assessment and agricultural production sustainability using machine learning models [J]. *Agronomy*, 2020, 10(4): 573.
- [6] Zhang Jiuquan, Su Yirong, Wu Jinshui, et al. GIS based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong Province of China [J]. *Computers and Electronics in Agriculture*, 2015, 114: 202-211.
- [7] Kahsay A, Haile M, Gebresamuel G, et al. Land suitability analysis for sorghum crop production in northern semi-arid Ethiopia: Application of GIS-based fuzzy AHP approach [J]. *Cogent Food & Agriculture*, 2018, 4(1): 1507184.
- [8] Muzira N M, Mushore T D, Wuta M, et al. Land suitability analysis of Zimbabwe for the production of sorghum (*Sorghum-bicolor*) and maize (*Zea mays*) using a remote sensing and GIS based approach [J]. *Remote Sensing Applications: Society and Environment*, 2021, 23: 100553.
- [9] 杨俊, 李雪铭, 曹永强, 等. 一种基于赋值技术与动态网格的空间数据组织策略及应用[J]. *国土资源遥感*, 2009, 21(3): 110-113.
- [10] Yin Shouqiang, Li Jing, Liang Jiabin, et al. Optimization of the weighted linear combination method for agricultural land suitability evaluation considering current land use and regional differences [J]. *Sustainability*, 2020, 12(23): 10134.
- [11] Pramanik M K. Site suitability analysis for agricultural land use of Darjeeling district using AHP and GIS techniques [J]. *Modeling Earth Systems and Environment*, 2016, 2(2): 1-22.
- [12] 何宗菲, 於家, 陈芸, 等. 城市建设用地的潜在生态适宜性评价: 以上海市宝山区为例[J]. *资源科学*, 2020, 42(3): 558-569.
- [13] 周文韬, 张文君, 邓云涛, 等. 基于 AHP-CV 组合赋权的低丘缓坡土地建设适宜性评价[J]. *西南师范大学学报(自然科学版)*, 2021, 46(6): 98-104.
- [14] Han Cheng, Chen Shengbo, Yu Yan, et al. Evaluation of agricultural land suitability based on RS, AHP, and MEA: A case study in Jilin Province, China [J]. *Agriculture*, 2021, 11(4): 370.
- [15] 邱炳文, 周勇, 李学垣. 地理信息系统支持下的区域土壤资源适宜性动态评价[J]. *土壤学报*, 2002, 39(3): 301-307.
- [16] Binte M R, Noguchi R, Ahamed T. Agricultural land suitability assessment using satellite remote sensing-derived soil-vegetation indices [J]. *Land*, 2021, 10(2): 223.
- [17] 程淑杰, 朱志玲, 白林波. 基于 GIS 的人居环境生态适宜性评价: 以宁夏中部干旱带为例[J]. *干旱区研究*, 2015, 32(1): 176-183.
- [18] Boyles M, Peeper T, Stamm M. Great Plains canola production handbook [J]. Manhattan, KS: Kansas State University Agricultural Experiment Station and Cooperative Extension Service, 2012: 6-18.
- [19] Breiman L. Random forests [J]. *Machine Learning*, 2001, 45(1): 5-32.
- [20] 林子聪, 任向宁, 朱阿兴, 等. 基于随机森林算法的耕地质量定级指标体系研究[J]. *华南农业大学学报*, 2020, 41(4): 38-48.
- [21] 马玥, 姜琦刚, 孟治国, 等. 基于随机森林算法的农耕地土地利用分类研究[J]. *农业机械学报*, 2016, 47(1): 297-303.
- [22] Van Beijma S, Comber A, Lamb A. Random forest classification of salt marsh vegetation habitats using quad-polarimetric airborne SAR, elevation and optical RS data [J]. *Remote Sensing of Environment*, 2014, 149: 118-129.
- [23] 李兴奇, 高晓红. 无量纲化方法的有效性评价[J]. *统计与决策*, 2021, 37(15): 24-28.
- [24] 张凤荣. 中国耕地[M]. 北京: 中国农业大学出版社, 2021.
- [25] 王秋兵. 土地资源学[M]. 2 版. 北京: 中国农业出版社, 2011.
- [26] 胡存智. 中国耕地质量等级调查与评定(全国卷)[M]. 北京: 中国大地出版社, 2013.
- [27] 赵晴月, 许世杰, 张务帅, 等. 中国玉米主产区土壤养分的空间变异及影响因素分析[J]. *中国农业科学*, 2020, 53(15): 3120-3133.
- [28] 于胜男, 高聚林, 明博, 等. 基于热量定量密植协同提升春玉米粒收品种产量及热量利用效率[J]. *中国生态农业学报(中英文)*, 2021, 29(12): 2046-2060.
- [29] 张佩, 高萃, 钱忠海, 等. 油菜花期物候主要限制因子分析及预报模型的构建[J]. *气象*, 2020, 46(2): 234-244.
- [30] Wanyama D, Mighty M, Sim S, et al. A spatial assessment of land suitability for maize farming in Kenya [J]. *Geocarto International*, 2021, 36(12): 1378-1395.
- [31] Vogel M A, Wong A K. PFS clustering method [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1979, 1(3): 237-245.