

# 基于空间统计和多元统计的耕地 影响因素及回归模型研究

——以重庆市石柱县为例

何昌华<sup>1</sup>, 陈丹<sup>2</sup>, 李天国<sup>1</sup>, 李辉<sup>2</sup>, 徐晓军<sup>1</sup>

(1. 昆明理工大学 环境科学与工程学院, 云南 昆明 650500; 2. 重庆工商大学 融智学院, 重庆 400067)

**摘要:** [目的] 通过空间自相关和多元回归分析, 揭示耕地空间分布规律, 为土地开发复垦及整理提供快速的评价方法。[方法] 以耕地面积占比为空间变量, 运用空间自相关及马塞克图分析耕地分布整体特征, 通过距离、地形、NDWI 和人口密度共 9 个因素对耕地空间分布进行多元回归分析, 模拟耕地分布适宜性并进行了检验。[结果] 空间自相关分析结果表明, 距离和地形因素对耕地空间分布具有显著影响, 空间自相关分析 Moran's  $I$  值为 0.701 5, 研究区耕地分布主要为不显著、LL(低空间自相关)和 HH(高空间自相关)类型, 其中不显著类型占研究区总面积的 65% 以上; 基于多远回归分析结果表明: 回归模型具有较高拟合优度和可靠性( $R^2 = 0.846$ ), 模拟得到的耕地分布适宜性图与现有耕地分布基本吻合。[结论] 研究区耕地空间分布总体上呈现较强的正相关关系, 且受距离、地形因素影响明显; 回归模型能够较好地揭示研究区耕地空间分布规律; 研究区具有一定耕地补充潜力; 将回归模型应用于土地开发复垦以及整理工作中, 有利于提高补充耕地质量, 减弱水土流失以及优化区域土地结构。

**关键词:** 土地开发整理; 空间自相关; 最大信息系数(MIC); 多元回归

**文献标识码:** A

**文章编号:** 1000-288X(2017)02-0199-08

**中图分类号:** F301.21

**文献参数:** 何昌华, 陈丹, 李天国, 等. 基于空间统计和多元统计的耕地影响因素及回归模型研究[J]. 水土保持通报, 2017, 37(2): 199-206. DOI:10.13961/j.cnki.stbctb.2017.02.031; He Changhua, Chen Dan, Li Tianguo, et al. A Study on Influencing Factors of Cultivated Land Based on Multivariate Regression and Spatial Statistics[J]. Bulletin of Soil and Water Conservation, 2017, 37(2): 199-206. DOI:10.13961/j.cnki.stbctb.2017.02.031

## A Study on Influencing Factors of Cultivated Land Based on Multivariate Regression and Spatial Statistics

—A Case Study of Shizhu County, Chongqing City

HE Changhua<sup>1</sup>, CHEN Dan<sup>2</sup>, LI Tianguo<sup>1</sup>, LI Hui<sup>2</sup>, XU Xiaojun<sup>1</sup>

(1. Faculty of Environmental Science and Engineering, Kunming University of Science and Technology, Kunming, Yunnan 650500, China; 2. Rongzhi College of Chongqing Technology and Business University, Chongqing 400067, China)

**Abstract:** [Objective] We aimed to reveal the regularity of spatial distribution of cultivated land by spatial autocorrelation analysis and multivariable regression, in order to provide a rapid evaluation method for land development, reclamation, and consolidation. [Methods] Coverage ratio by cultivated land was as response variable, and methods of spatial autocorrelation and mosaic plot were utilized to demonstrate its spatial pattern. Nine factors such as Euclidean distances, terrain, NDWI, population density, simulated cultivated land distribution suitability, etc. were used as independent variables, and multivariate regression of them with the

收稿日期: 2016-07-15

修回日期: 2016-08-13

资助项目: 重庆工商大学融智学院科研项目“重庆市耕地影响因素分析及土地开发整理可行性快速评价模型研究”(20157002)

第一作者: 何昌华(1993—), 男(汉族), 四川省资阳市人, 硕士研究生, 研究方向为环境工程和 3S 理论与应用。E-mail: hechanghua01@126.com。

通讯作者: 徐晓军(1960—), 男(汉族), 浙江省杭州市人, 博士, 教授, 主要从事 3S 理论与应用与矿山环境影响评价研究。E-mail: xuxiaojun88@sina.com。

response variable was conducted to test the distributional suitability of cultivated land. [Results] The Euclidean distances and terrain have significant impacts on the spatial distribution of cultivated land, and the Moran's  $I$  index is 0.701 5. In addition, the main types of local indicators of spatial association(LISA) distribution are not significant. L-L(low spatial autocorrelations) and H-H(high spatial autocorrelations) and insignificant types are three of the main types, especially the third type covered over 65% of study area. Multivariate regression behaved well in the distribution suitability simulation of cultivated land, it was remarkably coincided with the present distribution of cultivated land. The regression model was testified reliable and had goodness of fit ( $R^2=0.846$ ). [Conclusion] (1) The spatial distribution of cultivated land in the study area generally exhibits a strong positive correlation. And the distribution of cultivated land is affected by distance and terrain significantly. (2) The regression model can well reveal the spatial distribution of cultivated land in the study area, showing that the study area has a potential for cultivated land supplement. (3) We can improve the quality of additional cultivated land, reduce soil erosion, and optimize the land utilization structure if under the guidance of the regression model for land development, reclamation and consolidation.

**Keywords:** land development and consolidation; spatial autocorrelation; maximal information coefficient; multiple regression

土地是人类生存之本,耕地更是社会经济发展最重要的基础资源之一<sup>[1]</sup>,耕地质量管理是耕地保护的重要内容关系到国家粮食安全<sup>[2-3]</sup>,为此中国相继出台了相关法律、法规和规划<sup>[4]</sup>。但由于人类活动和自然条件变化等原因,导致我国耕地面积大大减少<sup>[5-6]</sup>;而土地开发整理的目的在于扩大耕地空间分布面积,提高土地利用集约利用率,改善区域生态条件<sup>[7-8]</sup>。但是,目前中国土地开发整理工作中更注重耕地数量上的平衡,而缺少对耕地质量的考虑,导致部分补充耕地不具有实际使用价值<sup>[9]</sup>,因此,为土地开发整理提供合理可行的科学方法从而保质保量地达到耕地保护的目地具有非常重要的意义。赵登辉等<sup>[10]</sup>从耕地的自然属性出发,认为耕地质量是一个综合属性,主要由土壤肥力和空间位置决定,因此在耕地质量管理中,优化耕地空间分布结构是管理的重要组成部分,是耕地保护得以实现的重要保证<sup>[11]</sup>。耕地保护不仅仅要求在数量上做到占补平衡,而且在耕地质量上也要具有实际价值;因此在耕地保护中除了考虑补充耕地的土壤肥力这一本质属性之外必不可少地应该考虑补充耕地的空间分布,进而有利于区域土地利用结构调整和格局优化实践<sup>[12]</sup>。本研究针对耕地空间位置分布这一自然属性,以耕地分布占比为空间变量,采用空间自相关分析法<sup>[13]</sup>,探索耕地分布在 500 m×500 m 尺度上的空间聚集规律及其受地形因素的影响,进而选用合理的影响因素结合多元回归分析得出耕地空间适宜性空间分布图<sup>[14-15]</sup>,旨在为耕地补充选址和土地利用结构优化提供依据。

## 1 数据源、方法与文章结构

主要采用土地变更数据库、DEM 数字高程和

TM 遥感影像数据,运用 ArcGIS 平台、GeoDa 软件以及 R 语言进行数据分析及可视化。

### 1.1 研究区概况

石柱土家族自治县位于长江上游地区、重庆东部,三峡库区腹心(107°59'22"—108°34'28"E,29°39'15"—30°33'02"N),截止到 2013 年,石柱县国土总面积 3 012 km<sup>2</sup>,下辖 32 个乡镇。2010 年第六次全国人口普查,石柱县常住人口为 41.5 万。七曜山、方斗山两大山脉平行排列斜贯石柱县全境,形成两道天然屏障,横亘南北,形成“两山夹一槽”的特殊地貌。全县地势东南高西北低,呈起伏状下降。最高海拔 1 934.1 m,最低海拔 119 m,境内以中山、低山为主,兼有山原、丘陵,是具有代表性的典型西南山区地貌<sup>[16]</sup>。

### 1.2 数据来源及数据预处理

选取的数据源为重庆市石柱县 2013 年度土地变更调查数据库和基本农田数据库,运用 ArcGIS 平台提取水域、村庄、农村道路和公路数据,并运用 Euclidean Distance 工具箱计算得到 30 m×30 m 精度的距离栅格数据,即因子  $X_1-X_4$ ;获取中国科学院地理空间数据云平台的研究区 DEM 数字高程数据( $X_5$ ),分辨率为 30 m×30 m,并计算坡度( $X_6$ )、坡向( $X_7$ )以及地形位指数数据;采用 Land SAT 2013 年遥感影像数据(4,5 波段)计算得到归一化水指数(NDWI)( $X_8$ );根据统计年鉴及其他政府公布数据,统计并计算研究区各区县的人口密度( $X_9$ )。技术路线图如图 1 所示。本研究栅格数据分辨率均为 30 m×30 m,因此运用 ArcGIS 平台的 Extract Multi Values To Points 工具把多个因素数据整合于点集数据用于分析。

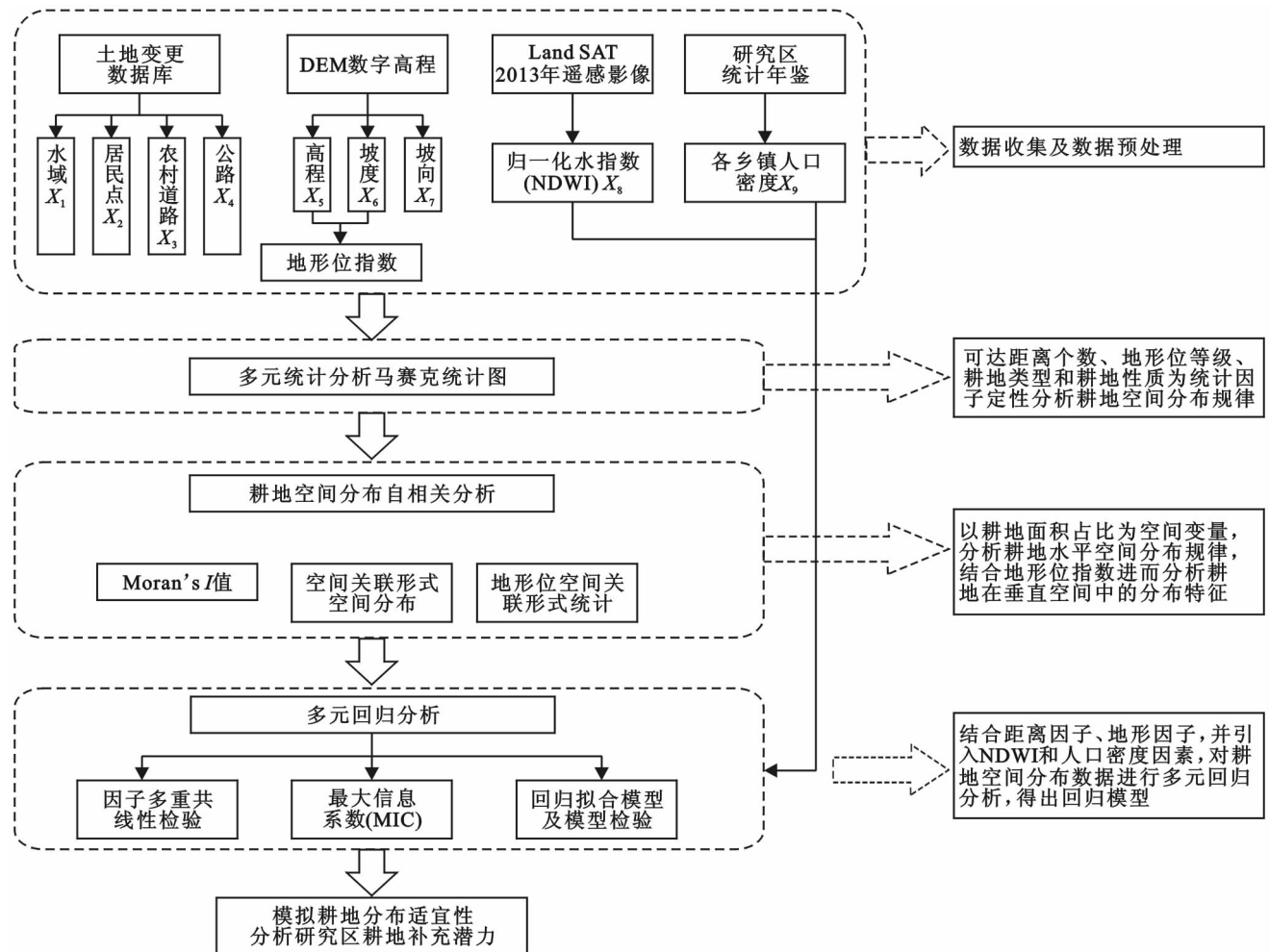


图 1 耕地影响因素回归模型研究技术路线

### 1.3 分析方法

本研究采用从定性到定量循序渐进的方式分析研究区耕地影响因素及其空间分布特征,并建立耕地空间分布回归模型。(1)运用马赛克统计图对耕地影响因素进行定性分析,为后文的空间自相关和模型建立提供支撑;(2)通过耕地空间自相关分析对耕地空间分布规律进行半定量分析,并结合地形位指数分析地形因素对耕地空间分布的影响;(3)引入归一化水指数和人口密度,通过多元回归建立研究区耕地空间分布模型,并对研究区耕地分布适宜性进行定量模拟。

1.3.1 地形位指数 地形位属于复合地形因子,可反映高程和坡度的综合影响,本文中的马赛克统计图和结合地形因素的空间自相关分析中因单个地形因子不足以反映地形条件,所以采用地形位指数,其公式为:

$$T = \ln \left[ \left( \frac{E}{E_0} + 1 \right) \times \left( \frac{S}{S_0} + 1 \right) \right] \quad (1)$$

式中: $T$ ——地形位指数; $E, S$ ——高程(m)和坡度( $^{\circ}$ ); $E_0, S_0$ ——研究区的平均值。地形位指数数值越大表明高程越高、坡度越大,反之亦然<sup>[17]</sup>。

1.3.2 马赛克统计图 本研究统计变量为满足可达距离个数、地形位等级、耕地性质和耕地类型,均为类别型的数据而非连续型定量数据,因此运用 R 语言绘制马赛克统计图进行多元数据的统计分析,通过矩形面积和颜色分别表示数据大小和残差值<sup>[18-19]</sup>,马赛克图反映的各因素对耕地空间分布的影响规律是后文进行空间自相关分析和回归分析的前提。

1.3.3 空间自相关分析 早在 1969 年 Tobler 提出了地理学第一定律,任何事物都与其他事物相联系,但邻近的事物比较远事物联系更为紧密<sup>[20]</sup>。本研究采用 Moran's  $I$  指数进行研究区空间自相关分析,包括全局空间自相关和局部空间自相关,用于空间数据的统计分析,说明耕地空间分布同其邻接空间单元的相关程度<sup>[21]</sup>。并利用 Moran's  $I$  散点图和 LISA 图进行空间格局的可视化<sup>[22]</sup>。

(1) 全局空间自相关。全局空间自相关能够描述某种地理现象或某一属性的整体分布,判断此现象或属性在空间上是否有聚集特性存在<sup>[23]</sup>。Moran's  $I$  计算公式为:

$$I = \frac{n}{s_0} \times \frac{\sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2} \quad (2)$$

其中,  $s_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$  是空间权重矩阵。

式中:  $n$ ——研究对象总个数;  $X_i, X_j$ ——空间单元观测值, 本文空间单元观测值即为耕地面积占比;  $\bar{X}$ ——所有观测值的平均值;  $w_{ij}$ ——研究对象  $i, j$  之间的空间权重。

(2) 局部空间自相关。针对研究区内部区域的空间自相关分析将采用局部空间自相关, 反映研究区各个空间单元上的某一属性与相邻局部单元上同一属性的相关程度。其中空间关联局域指标(local indicators of spatial association, LISA)用于揭示观测单元与周边单元的自相关性<sup>[24]</sup>。其计算公式为:

$$I_i = x_i' \sum_{j=1}^n w_{ij} x_j' \quad (3)$$

式中:  $x_i', x_j'$ ——标准化的单元观测值;  $w_{ij}$ ——研究对象  $i, j$  之间的空间权重。

运用 GeoDa 软件进行空间格局分析, 用于研究地理现象的空间分布规律。

1.3.4 最大信息系数(MIC) 2011年, 来自哈佛大学的 David 等提出针对相关系数的新算法〔最大信息系数(maximal information coefficient)〕, 它可以对数据的各种相关性进行快速估计, 发现一般的函数关系。其算法核心理念是, 如果两个变量之间存在着某种关系, 那么就应该能够通过变量散点图上切分网格, 并且使得大多数的数据点集中在该网格的部分单元格中<sup>[25-26]</sup>。本文通过 MIC 作者提供的 R 语言的扩展包进行计算 MIC 值。

1.3.5 多元回归分析 回归分析是一种用于定量分析事物之间的相互关系的统计方法, 反映 2 个或者多个变量之间的数量响应关系, 并通过回归方程的形式反映和描述这种关系, 进而准确把握变量受其他变量影响效应, 为控制和预测提供科学依据<sup>[27-28]</sup>。其中多元线性回归数学模型为:

$$y = b_0 + b_1 x_1 + b_2 x_2 + L + b_n x_n \quad (4)$$

式中:  $y$ ——因变量;  $x_n$ ——自变量;  $n$ ——自变量的个数。

本研究通过对数据进行适当的变形, 使其适用于多元线性回归形式, 基于 R 语言对数据进行回归分析, 得出回归模型, 进而对耕地空间分布进行模拟。

## 2 结果与分析

### 2.1 耕地空间分布特征及影响因素分析

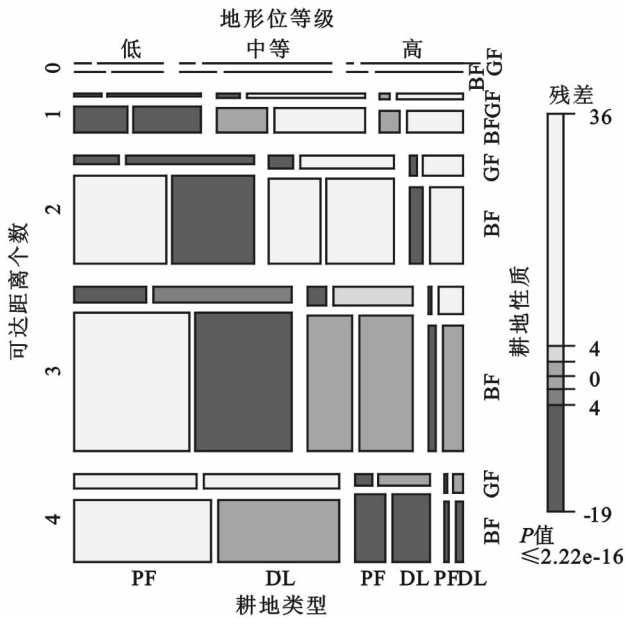
结合作者前期研究成果<sup>[15]</sup>, 设置居民点、公路、农村道路和水域对耕地分布影响的可达距离分别为

226, 660, 290 和 311 m, 进而结合反映高程和坡度综合影响的地形位等级、耕地性质和耕地类型 3 个因素统计耕地的分布情况(图 2), 图中的 4 个因子分别表示满足可达距离的个数(0, 1, 2, 3, 4)、地形位等级(低、中和高等级)、耕地性质(基本农田和一般农田)和耕地类型(水田和旱地)。图 2 中, 嵌套矩形的面积正比于耕地各分类单元的实际面积, 即多维列表中的耕地实际面积。由图 2 可知, 从满足 0 个可达距离至满足 4 个可达距离, 耕地面积呈先增加后减少的趋势, 大部分耕地满足 2 个及其以上的可达距离, 其中最大值出现在满足 3 个可达距离的情况; 表明研究区大部分耕地分布在 2 个及其以上距离因子的可达距离之内, 且同时满足 3 个可达距离的耕地最多。就地形位而言, 随着满足的可达距离数的增加, 低等级地形位中的耕地比例不断增加, 中、高等级的耕地比例减少; 由可达距离的物理意义可知, 可达距离之内为耕地的优势分布区域, 且地形位指数越低表明坡度和高程越低, 即为耕地分布的适宜区域, 图中满足可达距离的数量较多的区域与地形位指数较小的区域变化规律一致, 因此可达距离数能够间接反映耕地质量的好坏即满足的可达距离数越多耕地地形条件越好, 更适合作为耕地。就耕地性质和类型来看, 基本农田在不同质量的耕地中都占有较大的比例, 而基本农田中的水田和旱地在低、中等级地形位中基本处于均分的状态, 而在高等级地形位和一般耕地中旱地占比具有绝对优势, 原因在于低地形位指数区域能够同时满足作为旱地和水田的要求, 而高地形位指数区域地形条件较差则只能作为旱地。图 2 中不同色阶可以表示拟合模型的残差值。浅色表明在假定不同耕地类型的比例与可达距离、地形位和耕地性质无关的条件下, 该耕地类型的占比通常超过预期值, 深色则具有相反的意义。由图 2 综合分析可知, 距离和地形条件对研究区耕地空间分布具有较强的相关性, 因此后文回归分析中距离和地形因子将成为主要的因素纳入到回归方程中。

### 2.2 耕地分布自相关分析

在研究耕地自相关分析之前首先对耕地数据进行预处理, 将研究区划分为  $500 \text{ m} \times 500 \text{ m}$  的面积单元, 并统计各单元中的耕地面积占比作为空间自相关分析的空间变量。研究采用 GeoDa 软件进行全局和局部空间自相关分析, 首先利用该软件计算邻接关系(rook, queen 和 bishop)的空间权重, 对比发现 Rook 邻接关系的空间邻接性更加符合正太分布特征, 因此选用 Rook 邻接原则计算确定研究区的空间权重, 计

算得到研究区耕地空间关联形式空间分布图(图 3b)及各单元的局部空间自相关 Moran's  $I$  值,最后统计不同的地形位指数在各关联形式中的分布情况及其核密度估计。



注:BF 为基本农田;GF 为一般农田;PF 为水田;DL 为旱地。

图 2 耕地空间分布马赛克统计结果

如研究区耕地空间分布图(图 3)所示,耕地主要集中于西南—东北走向的三星乡—河嘴乡沿线及长江沿岸的西沱镇—王场镇沿线的条带状地区;而研究区东南片区海拔较高且地形较为复杂,耕地分布较少且相对较为零散;整体而言耕地分布集中分布于海拔较低的地区。

由研究区空间关联形式的空间分布图(图 4)可知。图中空间相关性不显著表示该区域内耕地随机分布,HH 型表示高水平的区域被其他高水平区域包围,LL 型表示低水平的区域被低水平的其他区域包围,同理解释 LH 和 HL 两种空间关联形式;其中 HH 和 LL 型表示空间正相关性,表明区域的聚集性,而 HL 和 LH 型表示空间负相关性,表明区域的异质性。

从图 4 可知,研究区空间关联局部指标(LISA)分布主要为 HH,LL 和不显著类型,统计得知其中不显著类型占研究区总面积的 65%以上。HH 类型与耕地分布现状基本吻合,而 LL 类型与非耕地区域吻合,结合地形位指数亦可知 HH 类型较 LL 类型具有更低的地形位指数,表明研究区 HH 类型区域因海拔较低,地势较为平坦且靠近河流,耕地分布较为集中,LL 类型区域主要为山地,没有耕地分布,研究区耕地分布整体较集中,两极分化明显。研究区不显著类型区域表示不具有空间相关性,耕地为随机分布。因此在实际的耕地保护实践过程中因根据耕地分布

不同的空间关联类型采取相应的保护措施。如 HH 型,现有的耕地分布较为良好,应维持现有水平耕地自然条件,禁止非农业建设;LL 型,应根据不同的自然条件等制定适合的非农化开发,如旅游开发等。在实际的工作中作为补充耕地的地块须适应研究区耕地分布的现有规律,避免出现补充耕地不具备实际价值导致土地荒废和水土流失的情况。



图 3 研究区耕地空间分布

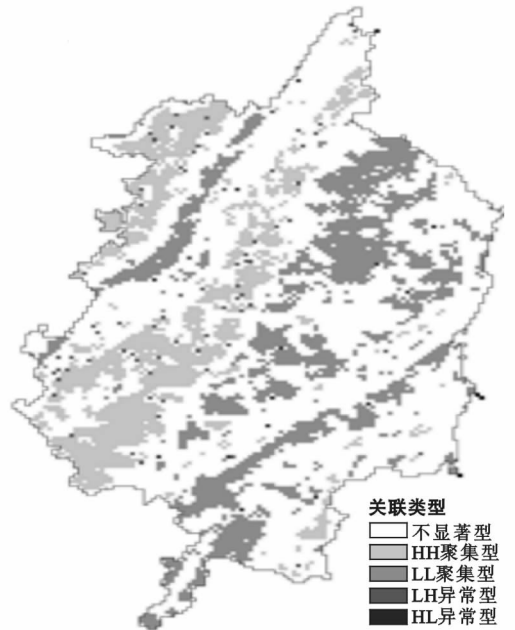


图 4 研究区耕地局部空间关联聚集状况

本研究中各分析单元数据主要呈 HH,LL 线性分布,且较多的点集中于 LL 象限内表明这些空间单元的空间相关性较类似,HH 象限的数据点则较为分散,表明 HH 类型的耕地分布数据内部具有一定的

差异性; Moran's  $I$  值为 0.701 5, 定量反映出研究区耕地在空间上呈较强正相关性。

通过统计地形位指数在各关联形式中分布情况可知, 各关联形式中地形位指数分布都较为集中, 其中 HH 类型较其他形式处于较低地形位指数的区域, 实际意义表明高水平的耕地分布都集中处于相对地形条件较好的区域, LL 类型较其他形式处于较高地形位指数的区域, 实际意义为低水平的耕地分布主要分布于地形条件较差的区域。

综上所述, 研究区的耕地分布主要表现为正相关性的聚集型, 且地形因素差异明显; 结合图 2 分析, 研究区乃至西南山区在土地开发整理过程中应该适应耕地聚集分布的自然规律以及充分考虑地形条件和距离水域、居民点等因子的距离因素。

### 2.3 耕地分布空间数据多元回归分析

结合前文分析可知, 耕地空间分布情况和地形因素及距离因素具有较强相关性, 本文除水域( $X_1$ )、居民点( $X_2$ )、农村道路( $X_3$ )和公路( $X_4$ )4 个距离因子以外进一步引入高程( $X_5$ )、坡度( $X_6$ )、坡向( $X_7$ )、归一化水指数(NDWI)( $X_8$ )以及研究区人口密度( $X_9$ )共 9 个因素对耕地空间分布进行空间数据回归分析。

在进行回归分析之前对数据进行预处理, 首先计算研究区范围内的 9 个因子的栅格数据, 再提取至距离为 30 m $\times$ 30 m 的样本点中, 并由耕地范围确定样本点中的耕地分布数据。

为避免出现过度拟合的情况, 要求先确定各自变

量间是否应互相独立。为此, 对各自变量进行多重共线性诊断, 结果容忍度在 0.35~0.58, 表明本次研究的 9 个自变量间多重共线性较弱, 适合全部纳入回归分析模型, 因此通过单因素数据构型和计算最大信息系数(MIC)对 9 个单因素进行分析以选择回归模型形式(表 1)。

表 1 各因素信息汇总表

影响因素( $X$ )	耕地比例( $Y$ )	最大信息系数 MIC	选用模型
离水域距离( $X_1$ )	耕地比例	0.7000 0	双曲线
离居民点距离( $X_2$ )	耕地比例	0.934 1	双曲线
离农村道路距离( $X_3$ )	耕地比例	0.998 0	双曲线
离公路距离( $X_4$ )	耕地比例	0.934 1	双曲线
高程( $X_5$ )	耕地比例	1	线性
坡度( $X_6$ )	耕地比例	1	双曲线
坡向( $X_7$ )	耕地比例	0.665 5	三角函数 COS
归一化水指数( $X_8$ )	耕地比例	0.839 0	高次
人口密度( $X_9$ )	耕地比例	0.756 1	高次

由表 1 可知,  $X_1 - X_9$  各因素对耕地占比都具有较大的 MIC 值, 表明各因素和耕地空间分布相关性较强, 适宜纳入到后面的回归模型中。并且根据各因素的数据构型和 MIC 信息确定了相应的模型形式进行回归分析, 如高程数据主要采用线性函数, 归一化水指数采用高次函数, 坡向因素选用 COS 函数等。

本研究对各因素进行了相应的数据变换, 利用 R 语言中的 lm 函数进行回归并得到回归信息汇总表(表 2)。

表 2 耕地空间分布的回归模型相关系数

因子形式	系数	T 值	显著性	因子形式	系数	T 值	显著性
截距	195.20	$<2e-16$	***	$X_8^3$	1.69e-04	$<2e-16$	***
$1/X_3$	-2.249	$<2e-16$	***	$\cos(250X_7)$	-1.300 0	3.76e-05	***
$1/X_4$	-1.162	9.89e-16	***	$1/X_6$	-129.800	2.79e-12	***
$\lg X_2$	-0.659	$<2e-16$	***	$X_5$	-0.007 5	0.002 856	**
$(1/X_3)\lg(X_2)$	-350.2	1.28e-14	***	$X_5(1/X_6)$	-0.048 7	0.000 402	***
$X_1$	-0.004 6	9.00e-09	***	$X_9$	0.044 96	3.87e-12	***
$X_8$	-3.291 0	$<2e-16$	***	$X_9^2$	-6.55e-05	2.58e-14	***
$X_1 \cdot X_8$	4.35e-05	3.21e-05	***				

注: \*\*\* 表示在 1% 水平上显著, \*\* 表示在 5% 水平上显著, \* 表示在 10% 水平上显著。

因回归拟合过程中对因子进行了变换, 因子需再次对各自变量进行多重共线性诊断, 结果容忍度在 0.39~0.6 之间, 回归模型通过检验得  $R^2=0.845$  和  $p$  值为  $<2.2e-16$ , 表明本次回归模型具有很好的拟合优度。由表 2 可以看出, 根据显著水平可知, 该回归模型中所有因素回归系数都具有较高的显著性。其中 4 个距离因素及其交互项回归系数均为负数, 表明耕地的空间分布比例与距离因素呈负相关, 主要原

因是随着距离的增加, 农民耕作的往返时间增加, 运载和管理不便, 所以耕地分布将减少, 其中影响最显著的是距离居民点的距离, 在其他条件不变的情况下, 距离居民点的距离每增加 32.93 m, 耕地分布占比减少 1%; 坡度和高程的系数也为负数, 主要原因是随着坡度和高程的增加, 地形条件越来越差, 不适宜耕作, 所以耕地占比减少, 其中高程每增加 100 m 耕地分布占比减少 0.75%。

## 2.4 耕地分布空间适宜性模拟

由前文回归模型模拟计算的耕地空间分布情况(图 5)将研究区耕地划分为 5 个不同的耕地空间分布适宜性等级,对比图 3 可发现模拟的耕地空间分布与研究区 2013 年耕地空间分布现状趋势基本一致,验证了上文模型的合理性。

根据研究区模拟结果的 5 个不同等级结合土地利用类型的统计信息图(图 6)可知,研究区较适宜和适宜 2 个等级中耕地占有绝对的优势,占比均大于 50%,其次为林地,但同时也存在少量的裸地和工矿用地,以及 2% 以上的草地,在实际条件允许的情况下,这些可作为土地复垦的储备。因此在之后的土地复垦项目中因充分考虑各种影响因素保质保量地完成耕地保留。

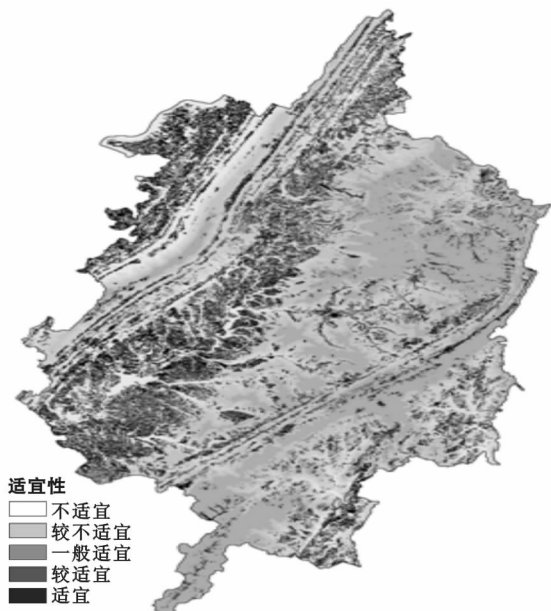


图 5 研究区耕地空间分布适宜性模拟结果

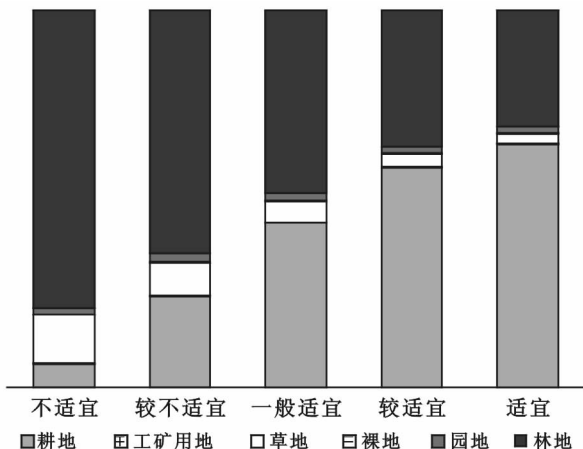


图 6 研究区耕地各适宜性等级中各地类占比统计结果

## 3 结论与讨论

(1) 马赛克图表明研究区耕地空间分布具有一定的规律;空间自相关分析得知研究区耕地空间分布总体上呈现较强的聚集式正相关关系,结合地形位指数统计分析可知,各关联形式中地形位指数分布具有明显的差异性;综合分析可知研究区耕地分布呈现聚集形式,并且受距离因素、地形因素的影响较为明显。

(2) 在前文统计分析的基础上收集距离因素、地形因素、NDWI 和人口密度共 9 个因子回归模拟得到拟合优度较好的回归模型,表明本文所选取的因子及回归模型能够较好地反映研究区耕地空间分布规律;进一步统计分析表明研究区仍具有较大质量较好的耕地补充潜力。

(3) 土地开发复垦以及整理过程中,因充分考虑影响耕地空间分布的各因素,以达到耕地保护中的质量平衡,进而减弱水土流失以及优化区域土地结构。

(4) 本研究较为全面地揭示耕地空间自相关性以及耕地空间分布与各影响因素之间的耦合规律。空间自相关及回归分析在现有技术条件下基本能够完成,其分析结果将有利于耕地空间分布调整和区域土地利用结构调整优化工作。

### [ 参 考 文 献 ]

- [1] 沈仁芳,陈美军,孔祥斌,等. 耕地质量的概念和评价与管理对策[J]. 土壤学报,2012,49(6):1210-1217.
- [2] 蔡海生,林建平,朱德海. 基于耕地质量评价的鄱阳湖区耕地整理规划[J]. 农业工程学报,2007,23(5):75-80.
- [3] 李子良,王树涛,张利,等. 经济快速发展地区耕地生产能力空间格局[J]. 农业工程学报,2010,26(11):323-331.
- [4] 韦仕川,熊昌盛,栾乔林,等. 基于耕地质量指数局部空间自相关的耕地保护分区[J]. 农业工程学报,2014,30(18):249-256.
- [5] 邵子南,王怀成,陈江龙,等. 中国农村居民点整理研究进展与展望[J]. 中国农业资源与区划,2013,34(3):10-15.
- [6] Coelho J C, Portela J, Pinto P A. A social approach to land consolidation schemes: A Portuguese case study: The valenca Project [J]. Land Use Policy, 1996,13(2):129-147.
- [7] 王军,严慎纯,白中科,等. 土地整理的景观格局与生态效应研究综述[J]. 中国土地科学,2012,26(9):87-94.
- [8] Petr Sklenicka. Applying evaluation criteria for the land consolidation effect to three contrasting study areas in Czech Republic [J]. Land Use Policy, 2006,23(4):502-510.

- [9] 张正峰, 赵伟. 土地整理的资源与经济效益评估方法[J]. 农业工程学报, 2011, 27(3): 295-299.
- [10] 赵登辉, 郭川. 对耕地定级与估价几个问题的思考[J]. 中国土地, 1997, 11(12): 18-19.
- [11] 奉婷, 张凤荣, 李灿, 等. 基于耕地质量综合评价的县域基本农田空间布局[J]. 农业工程学报, 2014, 30(1): 200-210.
- [12] 孙蕊, 孙萍, 吴金希, 等. 中国耕地占补平衡政策的成效与局限[J]. 中国人口·资源与环境, 2014, 24(3): 41-46.
- [13] Overmars K P, Koning G H J D, Veldkamp A. Spatial autocorrelation in multi-scale land use models [J]. Ecological Modelling, 2003, 164(2/3): 257-270.
- [14] 沈陈华. 丹阳市农村居民点空间分布尺度特征及影响因素分析[J]. 农业工程学报, 2012, 28(22): 261-268.
- [15] 陈丹, 周启刚, 何昌华, 等. 基于 MPI 的典型西南山区耕地空间分布影响因素研究: 以重庆石柱县为例[J]. 水土保持研究, 2014, 21(2): 228-233.
- [16] 杨沁汶, 安祺, 谢莹, 等. 基于 GIS 的重庆市石柱县土地利用现状分析[J]. 重庆工商大学学报: 自然科学版, 2012, 29(12): 87-94.
- [17] 郭洪峰, 许月卿, 吴艳芳. 基于地形梯度的土地利用格局与时空变化分析: 以北京市平谷区为例[J]. 经济地理, 2013, 33(1): 160-166.
- [18] Kabacoff R. R in Action: Data Analysis and Graphics with R [M]. Pearson Schweiz Ag: Manning Publications Co., 2015.
- [19] Hornik K, Zeileis A, Meyer D. The strucplot framework: Visualizing multi-way contingency tables with vcd [J]. Journal of Statistical Software, 2006, 17(3): 1-48.
- [20] Tobler W R. A computer movie simulating urban growth in the Detroit region [J]. Economic geography, 1970, 46(2): 234-240.
- [21] Verburg P H, Chen Y Q. Multiscale characterization of land-use patterns in China [J]. Ecosystems, 2000, 3(4): 369-385.
- [22] Anselin L, Syabri I, Kho Y. GeoDa: An introduction to spatial data analysis [J]. Geographical Analysis, 2006, 38(1): 5-22.
- [23] 孟斌, 王劲峰, 张文忠, 等. 基于空间分析方法的中国区域差异研究[J]. 地理科学, 2005, 25(4): 393-400.
- [24] Anselin L. Local indicators of spatial association: LISA [J]. Geographical Analysis, 1995, 27(2): 93-115.
- [25] Reshef D N, Reshef Y A, Finucane H K, et al. Detecting novel associations in large data sets [J]. Science, 2011, 334(6062): 1518-1524.
- [26] Zhang Yi, Jia Shili, Huang Haiyun, et al. A Novel algorithm for the precise calculation of the maximal information coefficient [J]. Scientific Reports, 2014, 4(4): 6662.
- [27] 项静恬, 郭世琪. 多元回归模型在实际应用中的几种推广[J]. 数理统计与管理, 1994, 13(4): 48-53.
- [28] 徐嘉兴, 李钢, 陈国良. 基于 Logistic 回归模型的矿区土地利用演变驱动力分析[J]. 农业工程学报, 2012, 28(20): 247-255.