

福建省闽侯县区域农作物种植面积的空间抽样方案

刘金福^{1,2,3}, 林芳芳¹, 路春燕^{1,2,3}, 尤添革^{1,2,3}, 陈远丽¹, 吴碧致¹, 朱建平⁴

(1.福建农林大学计算机与信息学院,福建 福州 350002; 2.生态与资源统计福建省高校重点实验室,福建 福州 350002; 3.福建省资源环境空间信息统计研究中心; 福建 福州 350002; 4.厦门大学管理学院,福建 厦门 361005)

摘要: 基于国产 GF-1 号卫星影像农作物空间分布数据,以正方形网格作为抽样单元形状,综合应用遥感技术、空间抽样方法、模拟退火算法及空间自相关理论,对福建省闽侯县农作物种植面积进行估算。结果表明:根据全局自相关指数 Moran's I 及其显著性指标 Z Score,选取 $1\ 500\ \text{m} \times 1\ 500\ \text{m}$ 作为最优单元尺寸;对研究区进行剖分,构建抽样框,对比空间随机抽样、空间系统抽样及空间分层抽样的抽样结果,空间分层抽样方法的样本容量为 37,相对误差为 3.86%,抽样效率最高;运用模拟退火优化算法规划野外调查的最优路径,确定空间抽样调查方案。

关键词: 遥感技术; 分层抽样; 模拟退火优化算法; 农作物面积

中图分类号: S127 文献标识码: A 文章编号: 1671-5470(2018)02-0243-07

DOI: 10.13323/j.cnki.j.fafu(nat.sci.).2018.02.018

Spatial sampling plan for crop acreage in Minhou County, Fujian Province

LIU Jinfu^{1,2,3}, LIN Fangfang¹, LU Chunyan^{1,2,3}, YOU Tiange^{1,2,3}, CHEN Yuanli¹,
WU Bizhi¹, ZHU Jianping⁴

(1.College of Computer and Information Science, Fujian Agriculture and Forest University, Fuzhou, Fujian 350002, China;
2.Key Lab. of Fujian Universities for Ecology and Resources Statistics, Fuzhou, Fujian 350002, China;
3.Fujian Resources and Environment Spatial Information Statistics Research Center, Fuzhou, Fujian 350002, China; 4.School of Management, Xiamen University, Xiamen, Fujian 361005, China)

Abstract: Based on the spatial distribution data of crops of domestic GF-1 satellite image, the square grid was taken as the sampling unit shape, and the remote sensing technology, spatial sampling methods, simulated annealing algorithm and spatial autocorrelation theory were applied to this study for estimating the planting area of crops in Minhou County, Fujian Province. The results showed that according to global autocorrelation index Moran's I and its significant index Z Score, $1\ 500\ \text{m} \times 1\ 500\ \text{m}$ was selected as the optimal size unit. And the study area was divided to construct the sampling frame. Compared with the sampling results of spatial random sampling, spatial system sampling and spatial stratified sampling, the sampling efficiency of the spatial stratified sampling method was the highest because its number of samples was 37, the relative error was 3.86%. The space sampling survey scheme was determined because the optimal path for the field survey was found by simulated annealing algorithm.

Key words: remote sensing technology; stratified sampling; simulated annealing algorithm; crop acreage

近年来,由于城镇扩张、环境污染及耕地管理不善,我国农田面积不断减小,实现农作物种植面积精准调查对国家制定粮食政策与经济计划具有重要的指导意义^[1-3]。

农作物种植面积调查通常采用传统抽样调查方法,但存在调查基础资料时效性不高、野外调查工作量大等问题^[4]。近年来随着遥感技术的发展,因其高时效、宽范围、低成本等优点,常被用于农作物种植面积调查^[3-5],而遥感影像中同物异谱、同谱异物及混合像元的普遍存在,使得复杂景观及大范围观测中不同种类农作物种植面积的估算精度难以达到 90%。因此,许多专家考虑将遥感技术与抽样技术结合来调查农作物种植面积^[5-9]。朱爽等^[5-7]基于研究区遥感卫星影像及相关基础地理信息数据,采用分层抽样方法对水稻、玉米等农作物种植面积进行估计;美国国家农业统计署(NASS)采用分层两阶段空间抽样与遥感技

收稿日期: 2017-05-24 修回日期: 2017-09-15

基金项目: 全国统计科学研究资助项目(2015LZ18)。

作者简介: 刘金福(1966-),男,教授,博士生导师,研究方向: 资源环境统计。Email: fjlj@126.com。通信作者朱建平(1962-),男,教授,博士生导师,研究方向: 应用数理统计、数据挖掘。Email: xmjzhu@163.com。

术相结合的方法对美国大面积农作物面积进行监测和估算^[8];王劲峰等^[9]充分考虑样本间的空间关联性,在空间分层抽样的基础上,提出了“三明治”空间抽样模型。

近年来,抽样理论方法及遥感调查应用体系不断完善,被广泛应用于农作物种植面积调查中。而大多数农作物种植面积遥感调查应用研究区域主要集中在我国北方平原地区^[10-12],南方山地丘陵区相关研究并不多见,且实际遥感抽样方案设计中往往忽略抽样调查的成本效益。目前我国农作物种植面积抽样调查研究主要采用 Landsat、SPOT 和 QuickBird 等国外中高分辨率卫星数据,而较少应用 ZY1 号、ZY3 号、GF-1 号及 GF-2 号等国产新型卫星数据。为此,选取典型的南方山地丘陵区闽侯县作为研究区域,采用国产 GF-1 号卫星数据,提出一种遥感技术、抽样技术与模拟退火算法相结合的农作物抽样面积调查方法,以期对南方山地丘陵区农作物种植面积调查提供最优遥感抽样框,同时促进国产影像数据的应用推广。

1 研究区概况

闽侯县隶属福建省福州市,地处北纬 25°47′—26°37′,东经 118°51′—119°25′,素有“八闽首邑”之称。该县地处福建东南部,地形地貌错综复杂,县境内中山、低山、丘陵、盆谷地以及平原兼有,以中山为主,平原主要为河流冲积平原(主要分布于闽江两岸)。闽侯县下辖 15 个乡镇 313 个行政村,土地总面积 2 136 km²,北部中低山为林、茶区,中部低山丘陵为粮、渔、果、防护林综合区,南部冲积平原为粮、渔、牧多种经营区,西南部中低山丘陵为林、果混合区。

2 方法

以闽侯县为研究区域,基于 GF-1 号卫星影像农作物空间分布数据,考虑抽样调查成本及调查精度,设计空间抽样调查方案。首先对研究区进行剖分,应用空间自相关指数 Moran's I 确定抽样单元最优尺寸;再构建抽样框,对比分析空间随机抽样、空间系统抽样及空间分层抽样的抽样精度,筛选最优抽样方法进行农作物面积抽样估算;最后综合考虑抽样调查精度和可达性,利用模拟退火优化算法选择野外抽样调查的最优路径。

2.1 数据预处理

GF-1 号卫星影像因其具有幅宽大、重访周期短及分辨率高的特点,被广泛应用于农业遥感、环境监测、减灾应急等领域^[13]。因此,其数据可作为农作物空间分布数据提取的基础数据源。首先对 GF-1 号遥感影像进行辐射定标、大气校正和正射校正;再以地形图为参考,选择控制点,通过重采样纠正 GF-1 号影像的几何偏差,误差控制在 0.5 个像元以内;最后以 eCognition 9.0 软件为操作平台,采用面向对象遥感分类方法提取研究区农作物,结合目视解译及 Google earth 卫星影像对提取结果进行修正,获取闽侯县农作物遥感识别数据,结果见图 1。其中闽侯县 1:5 万行政区划图及交通路网数据被收集以建立最优抽样框模型。

2.2 空间抽样方案设计

2.2.1 抽样单元尺寸设计 以正方形网格为抽样基础单元形状,遵循“抽样单元间相互独立”原则^[14],设计 500 m×500 m、800 m×800 m、1 000 m×1 000 m、1 200m×1 200 m 和 1 500 m×1 500 m 等 5 种抽样基础单元尺寸方案;采用 Arcgis 10.2 对整个研究区进行剖分,计算各个抽样基础单元中农作物的面积;依据其空间分析模块分别计算对应的 Moran's I 指数,选择空

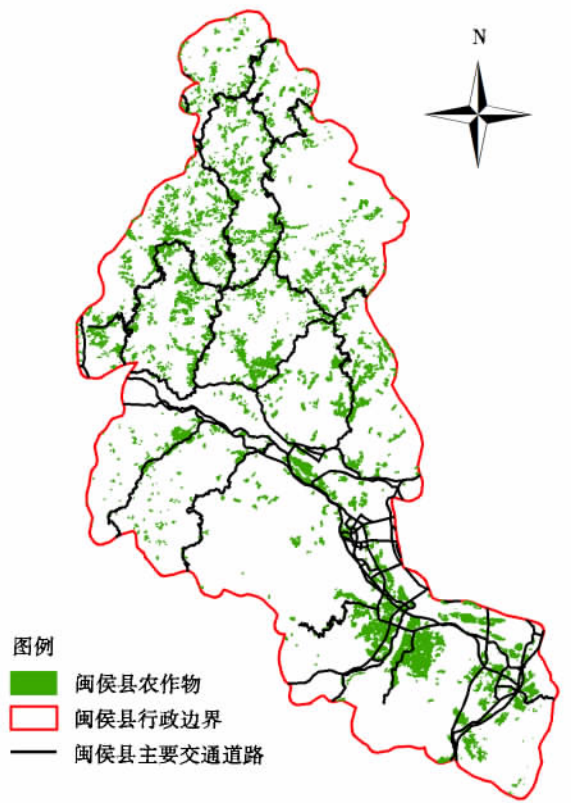


图 1 闽侯县农作物的空间分布

Fig.1 Spatial distribution of crops in Minhou County

间自相关指数最小的抽样基础单元尺寸方案作为最优抽样单元尺寸.全局空间自相关指数 Moran's I 模型^[15]表示如下:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \mu) (x_j - \mu)}{\sum_{i=1}^n (x_i - \mu)^2} \quad (1)$$

式中 I 是全局空间自相关指数; i, j 代表不同的空间单元编号; N 表示研究区所有空间基础单元的个数; x 表示研究区空间抽样基础单元的面积; m 为研究区所有抽样基础单元农作物面积的均值; w_{ij} 为空间权值矩阵, 反映空间抽样基础单元 i 与 j 的空间关系.

2.2.2 空间抽样方法 利用空间自相关指数 Moran's I 确定最优抽样单元尺寸; 采用空间随机抽样、空间系统抽样及空间分层抽样 3 种方法进行抽样估算, 筛选最优抽样方法.

(1) 简单随机抽样: 指从含有 N 个单元的总体中, 随机、独立抽取 n 个单元进行总体估算和误差计算, 样本容量按照式(2)-(5)计算^[16]:

$$n_0 = \left(\frac{t}{r} \right)^2 \frac{S^2}{\bar{Y}^2} \quad (2)$$

$$n = \frac{n_0}{1 + n_0/N} \quad (3)$$

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i \quad (4)$$

$$S^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2 \quad (5)$$

式中: n_0 为初始样本容量; t 为可靠性指标(按 95% 可靠性); r 为设计相对误差, 设为 10%; S^2 是总体单元的离散方差; \bar{Y} 为总体单元值; n 为样本容量, 当 $n_0/N > 0.05$ 时, 采用式(3)修正; N 是总体单元数.

(2) 系统抽样: 又称等距抽样, 从含有 N 个单元的总体中, 随机确定起点后, 按照预先规定的间隔抽取 n 个单元组成样本, 用于总体估算和误差估计. 系统抽样也是等概抽样方法之一, 其样本容量 n 的确定、总体估算和误差估计公式同随机抽样^[17].

(3) 分层抽样: 是在各层内独立、随机地进行抽样^[17]. 分层标志设计: 在研究区抽样基础单元的基础上, 以面积比(单个格网内农作物面积与格网面积的比值)为分层标志, 采用戴伦纽斯(Daleniuss)和霍捷斯(Hodges)提出的累计等值平方根法^[12], 以农作物面积比为分层变量, 根据最优分层原则, 得到层数和分层界限. 样本容量计算: 根据最优分配原则计算分层抽样样本容量及各层样本容量, 具体计算公式见文献[12, 18].

(4) 空间自相关: 在传统抽样方法的基础上, 考虑样本间的空间自相关可有效减少样本冗余及减少抽样成本. 计算空间自相关指数后, 利用标准化统计量 Z Score 对空间自相关的显著性水平进行检验^[19], 以确定样本单元间的空间自相关性.

与传统抽样方法相比, 在空间抽样调查中, 均值估计方差随空间抽样对象的相关程度而变化^[19].

空间抽样中均值的方差 V 表示为:

$$V = \text{var}\{\bar{Z} - Z(A) \mid Z\} = \frac{1}{n} [\sigma^2 - E\{C(X, Y)\}] = \frac{\sigma^2}{n} - \frac{E\{C(X, Y)\}}{n} \quad (6)$$

式中 σ^2 为总体方差; X, Y 为在研究区 A 中服从均匀分布的随机变量; $C(X, Y)$ 是变量 X, Y 的协方差. 空间抽样方法比传统抽样方法的均值方差减少 $E\{C(X, Y)\}$, 因此, 空间抽样样本容量为:

$$n' = \frac{\sigma^2 - R}{V} = n(1 - r') \quad (7)$$

式中: 令 $R = E\{C(X, Y)\}$, $r' = R/\sigma^2$ 为总体相关系数, n' 为空间抽样方法样本量.

2.2.3 样本值的获取与估算 基于研究区农作物面积空间分布数据, 采用 Arcgis 10.2 软件统计样本单元内农作物种植面积值, 并以此为样本值, 采用简单估计量进行总体估算及误差计算, 总体相对误差^[6]表示

如下:

$$r = \frac{|\hat{Y} - Y|}{Y} \times 100\% \tag{8}$$

$$\hat{Y} = N\bar{y} \tag{9}$$

式中: r 为相对误差; Y 为总体总值的真值; \hat{Y} 为总体总值的估计量; \bar{y} 为样本均值.

2.3 通过模拟退火算法选择最优路径

我国南方山地丘陵区,耕地景观异常破碎,间作套种普遍,在实际遥感抽样框设计中往往忽略抽样单元调查成本.因此,综合考虑影响调查成本的因素,利用智能优化算法建立最优抽样框选择模型,以优化抽样框设计方案.

模拟退火算法起源于物理中固体物质的退火过程,是一种通用的优化算法.其实现过程如图 2 所示,详细过程^[20]表示如下:

(1) 初始化: 栅格式环境信息用 0 和 1 组成的矩阵表示, 0 表示可通过栅格, 即有道路通过的单元格; 1 表示障碍物占用栅格, 即无道路通过的单元格. 初始化可选路径节点 $D = \{0, 1, \dots, n\}$; 取初始温度 T_0 足够大, 令 $T = T_0$, 任取初始解 S_1 , 确定每个 T 的迭代次数, 即 Metropolis 链长 L .

(2) 对当前温度 T 和 $k = 1, 2, 3, \dots, L$, 重复步骤(3)~(6).

(3) 对当前解 S_1 随机扰动产生一个新解 S_2 .

(4) 计算 S_2 的增量 $df = f(S_2) - f(S_1)$, 其中 $f(S_1)$ 为 S_1 的代价函数.

(5) 若 $df < 0$, 则接受 S_2 作为新的当前解, 即 $S_1 = S_2$; 否则计算 S_2 的接受概率 $\exp(-df/T)$, 即随机产生 $(0, 1)$ 区间上均匀分布的随机数 rand . 若 $\exp(-df/T) > \text{rand}$, 也接受 S_2 作为新的当前解, $S_1 = S_2$, 否则保留当前解 S_1 .

(6) 如果满足终止条件 Stop , 则输出当前解 S_1 为最优解, 结束程序.

终止条件 Stop 通常为: 在连续若干 Metropolis 链中新解 S_2 都没有被接受时终止算法, 或是设定结束温度. 否则按衰减函数衰减 T 后返回步骤(2). 该过程即为 Metropolis 过程. 逐渐降低控制温度, 重复 Metropolis 过程, 直到满足结束条件 Stop , 求出最优解.

3 结果与分析

3.1 抽样单元尺寸

利用 Arcgis 10.2 软件得到 5 种抽样单元尺寸的全局自相关指数 Moran's I 和其相应的显著性指标 Z Score 的变化情况(图 3). Moran's I 取值范围近似为 $[-1, 1]$, 越接近 -1 则代表单元间的分布越不集中, 越接近 1 则代表单元间的分布越集中, 接近 0 则代表单元间不相关. 对于全局自相关指数, 利用标准化统计量 Z Score 对空间自相关的显著性水平进行检验^[14]. 由图 3 可知 5 种抽样单元尺寸方案的变化规律, 即随着抽样单元尺寸增大, 对应的 Moran's I 和 Z Score 则相应减小, 表明抽样单元间的空间自相关性越弱, 显著性程度也越低. 当抽样单元尺寸为 $1\ 500\ \text{m} \times 1\ 500\ \text{m}$ 时, 对应的 Moran's I 最小, Z Score 也最小, 说明其空间自相关性较弱. 因此, 选取 $1\ 500\ \text{m} \times 1\ 500\ \text{m}$ 作为最优的抽样单元尺寸.

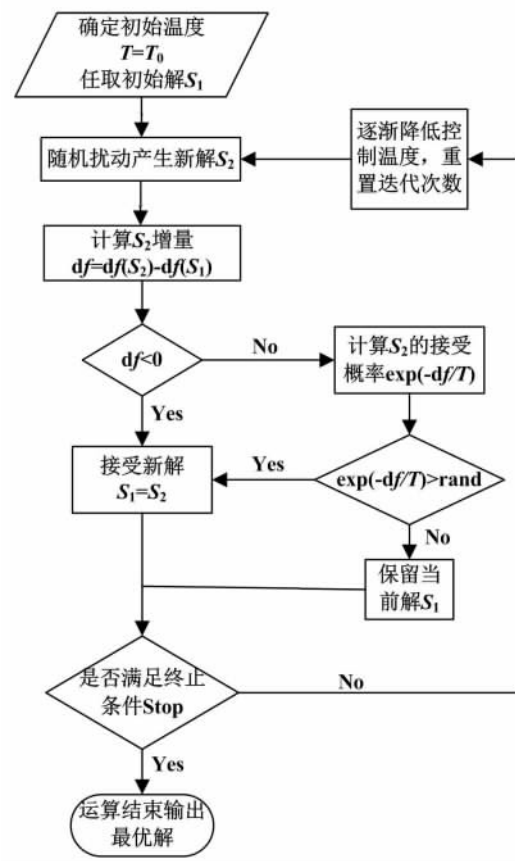


图 2 模拟退火算法的实现过程

Fig.2 The realization process of simulated annealing algorithm

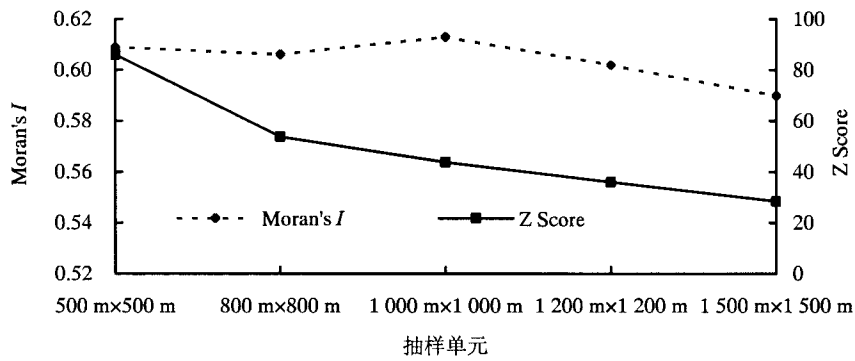


图 3 5 种抽样单元方案的空间自相关指数

Fig.3 Spatial autocorrelation index of five sampling unit schemes

3.2 空间抽样方案实证

为优选适宜南方山地丘陵区的农作物面积抽样方法,选择相对误差和样本容量作为评价指标.表 1 给出了在 95%置信水平下,3 种抽样方法达到 90%精度要求所需样本容量及抽样精度.由表 1 可知,3 种抽样方法中,空间分层抽样的抽样精度明显高于空间随机抽样及空间系统抽样,且空间分层抽样方法所用样本量最少.可见,空间分层抽样方法效率最高.

表 1 3 种抽样方法的统计结果

Table 1 Statistics of three sampling methods

抽样方法	置信度 /%	设计相对误差 /%	样本容量	相对误差 /%
空间随机抽样	95	10	252	6.53
空间系统抽样	95	10	252	4.70
空间分层抽样	95	10	37	3.86

依据图 3,以 1 500 m x 1 500 m 为最优抽样单元尺寸,采用 Arcgis 10.2 软件对闽侯县进行剖分,共得到 1 197 个抽样基础单元.表 2 给出了空间分层抽样方法的分层参数,采用戴伦纽斯(Dalenius)和霍捷斯(Hodges)提出的累计等值平方根法^[18].根据最优分层原则,将闽侯县抽样基础单元分为 3 层,各层的样本量分别为 26、9、2.在各层内随机抽取样本点,结果如图 4 所示.由表 2、图 4 可知,抽样基础单元主要集中在第 1 层与第 2 层(面积比 0%~20%),而第 3 层(面积比 20%~100%)的抽样基础单元数相对较少.可见,闽侯县地形较为复杂,农作物地块破碎程度较高,野外抽样调查难度较大.

表 2 分层参数表

Table 2 Parameter of stratified sampling

层号	累计平方根	各层单元数	分层标准 /%
1	28.88	834	0~5
2	57.38	294	5~20
3	82.77	69	20~100

为了降低野外调查难度,野外调查时考虑交通道路的可达性及调查难度,在进行分层抽样时,仅抽取有道路经过的基础抽样单元(410 个).在 95%置信水平、90%精度要求下,空间分层抽样方法的样本容量为 37,相对误差为 3.86%,适用于福建山地丘陵区农作物种植面积的调查.

3.3 野外调查最优路径选择

合理规划野外调查路径成为野外调查抽样的必要条件.模拟退火算法是一个用于组合优化问题的通用概率启发式算法,被广泛应用于路径优化问题^[20].在调查样本确定的基础上,结合闽侯县交通路网矢量图,以野外调查路径最短为目标函数,通过对模拟退火算法中重要参数的调整,设置初始值为 1,降温因子为 0.999,Metropolis 链迭代次数为 2×10^5 ,终止值为 10^{29} ,初始解随机产生.利用 Matlab 软件选择一条经过各个采样点的野外调查优化路线,以便在进行野外抽样调查时节省调查时间及降低调查成本.最优路径见图 4.

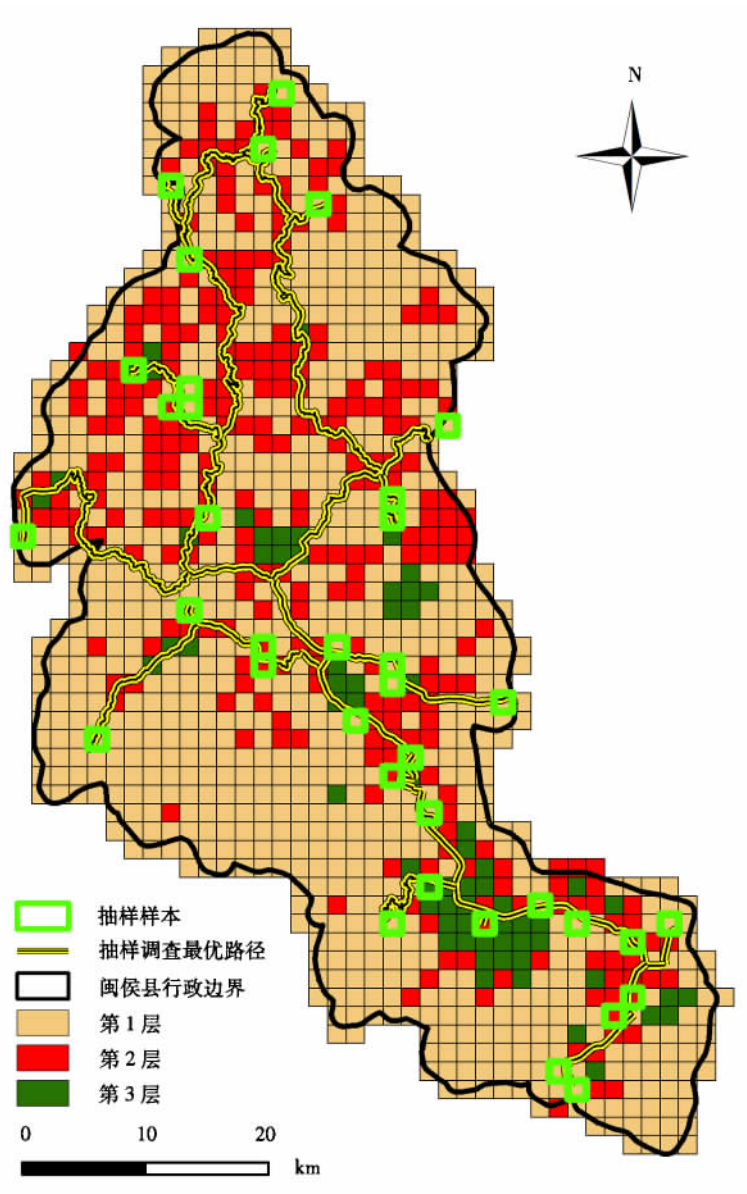


图 4 空间抽样方案设计

Fig.4 Design of spatial sampling scheme

4 小结

为估算南方丘陵山区农作物种植面积,以福建省闽侯县为研究区域,在国产 GF-1 号卫星影像农作物空间分布数据的基础上,以正方形网格作为抽样基础单元形状,综合应用遥感技术、抽样调查方法、空间自相关理论及智能优化算法,充分考虑抽样调查的估算精度和可达性及调查成本对空间抽样调查方案进行设计,估算农作物种植面积.结果表明:

(1) 遵循“抽样单元间相互独立”的原则,设计 500 m×500 m、800 m×800 m、1 000 m×1 000 m、1 200 m×1 200 m、1 500 m×1 500 m 等 5 种基础抽样单元尺寸方案.根据 5 种基础抽样单元尺寸方案对应的全局自相关指数 Moran's *I* 及其显著性指标 *Z* Score,选取 1 500 m×1 500 m 作为抽样单元最优尺寸.

(2) 以 1 500 m×1 500 m 作为基础抽样单元,对研究区进行剖分,构建抽样框.在 95% 置信水平、90% 精度要求下,采用简单随机抽样、系统抽样及分层抽样对研究区农作物面积进行抽样估算.综合比较抽样结果,空间分层抽样样本容量为 37,相对误差为 3.86%,优于简单随机抽样及系统抽样,适宜于福建山地丘陵区农作物种植面积的遥感监测.

(3) 在进行野外抽样调查时,耗费大量的人力物力.针对这一问题,在分层抽样的基础上,综合考虑抽

样调查的估算精度、可达性及调查成本,采用模拟退火算法找出一条最优路径,从而降低调查成本。

参考文献

- [1] 张焕雪,李强子,文宁,等.农作物种植面积遥感估算的影响因素研究[J].国土资源遥感,2015,27(4):54-61.
- [2] 钱永兰,杨邦杰,焦险峰.基于遥感抽样的国家尺度农作物面积统计方法评估[J].农业工程学报,2007,23(11):180-187.
- [3] 吴炳方.中国农情遥感速报系统[J].遥感学报,2004,8(6):481-497.
- [4] 邬明权,杨良闯,于博,等.基于遥感与多变量概率抽样调查的作物种植面积测量[J].农业工程学报,2014,30(2):146-152.
- [5] 朱爽,张锦水.面向省级农作物种植面积遥感估算的分层方法[J].农业工程学报,2013,29(2):184-191.
- [6] 王迪,陈仲新,周清波,等.冬小麦种植面积空间抽样样本布局的优化设计[J].中国农业科学,2014,47(18):3 545-3 556.
- [7] 焦险峰,杨邦杰,裴志远.基于分层抽样的中国水稻种植面积遥感调查方法研究[J].农业工程学报,2006,22(5):105-110.
- [8] BORYAN C, YANG Z W, MUELLER R, et al. Monitoring US agriculture: the US department of agriculture, national agricultural statistics service, cropland data layer program[J]. Geocarto International, 2011, 26(5):341-358.
- [9] WANG J F, ZHUANG D F, LI L F. Spatial sampling design for monitoring the area of cultivated land[J]. International Journal of Remote Sensing, 2002, 13(2):263-284.
- [10] 张锦水,申克建,潘耀忠,等.HJ-1号卫星数据与统计抽样相结合的冬小麦区域面积估算[J].中国农业科学,2010,43(16):3 306-3 315.
- [11] 潘学鹏,李改欣,刘峰贵,等.华北平原冬小麦面积遥感提取及时空变化研究[J].中国生态农业学报,2015,23(4):497-505.
- [12] 阳小琼,朱文泉,潘耀忠,等.作物种植面积空间对地抽样方法设计[J].农业工程学报,2007,23(12):150-155.
- [13] 刘国栋,邬明权,牛铮,等.基于GF-1卫星数据的农作物种植面积遥感抽样调查方法[J].农业工程学报,2015,31(5):160-166.
- [14] 王迪,周清波,陈仲新,等.冬小麦种植面积空间抽样单元尺寸优化设计[J].自然资源学报,2013,28(7):1 232-1 242.
- [15] 王雪青,陈媛,刘炳胜.中国区域房地产经济发展水平空间统计分析——全局 Moran's I, Moran 散点图与 LISA 集聚图的组合研究[J].数理统计与管理,2014,33(1):59-71.
- [16] 李宜展,朱秀芳,潘耀忠,等.农作物种植面积遥感估算优化研究——抽样单元[J].北京师范大学学报(自然科学版),2015,51(S1):119-126.
- [17] 宋新民,李新良.抽样调查技术[M].北京:中国林业出版社,2007.
- [18] 陈仲新,刘海启.全国冬小麦面积变化遥感监测抽样外推方法的研究[J].农业工程学报,2000,16(5):126-129.
- [19] 王劲峰,姜成晟,李连发,等.空间抽样与统计推断[M].北京:科学出版社,2009.
- [20] 庞峰.模拟退火算法的原理及算法在优化问题上的应用[D].长春:吉林大学,2006.

(责任编辑:叶济蓉)