

基于遥感的福建闽侯丘陵区农作物 种植面积空间抽样方法

林芳芳¹, 刘金福^{1,2,3}, 路春燕^{1,2,3}, 朱建平⁴, 陈远丽¹, 洪伟²

(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: 以福建闽侯县作为研究区, 采用传统抽样(简单随机抽样、系统抽样、分层抽样)、空间抽样(空间随机抽样、空间系统抽样、空间分层抽样)等方法对研究区农作物种植面积进行样本抽选、总体推算及误差估计。结果表明: 1 500 m×1 500 m 格网为最优抽样单元尺寸, 空间分层抽样方法的相对误差为 3.86%, 变异系数为 6.03%, 抽样成本为 6.03, 抽样效率高。与传统抽样方法相比, 空间抽样方法显著减少样本容量, 节约调查成本。

关键词: 福建; 闽侯丘陵区; 农作物种植面积; 空间抽样方法; 遥感监测

中图分类号: S127 **文献标识码:** A **文章编号:** 1671-5470(2017)06-0678-07

DOI: 10.13323/j.cnki.j.fafu(nat.sci.).2017.06.013

Spatial sampling method for crop acreage based on remote sensing in hilly area in Minhou County, Fujian Province

LIN Fangfang¹, LIU Jinfu^{1,2,3}, LU Chunyan^{1,2,3}, ZHU Jianping⁴, CHEN Yuanli¹, HONG Wei²

(1.College of Computer and Information Science, Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China;
2.Key Laboratory 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: Traditional sampling methods of simple random sampling, systematic sampling, stratified sampling, and spatial sampling methods of spatial random sampling, spatial system sampling and spatial stratified sampling were applied to conduct sample extraction, population extrapolation and error estimation in crop acreage in Minhou County. The results showed that spatial stratified sampling had the highest sampling efficiency with 1 500 m×1 500 m square grid as the optimal sampling unit size. And relative error, variation coefficient and sampling cost was 3.86%, 6.03% and 6.03, respectively. Compared to traditional sampling method, spatial sampling method significantly reduced sampling size and saved cost.

Key words: Fujian Province; hilly area of Minhou County; crop acreage; spatial sampling methods; remote sensing monitoring

对农作物种植面积进行遥感监测,可精准、及时地提供农作物种植面积,是了解农业生产、资源以及进行农业结构调整的前提,对国家制定粮食政策与经济计划具有重要的指导意义^[1]。将遥感技术与抽样技术相结合来监测农作物种植面积的研究备受关注^[2-11]。基于遥感技术,采用分层两阶段空间抽样技术监测与估算美国大尺度农作物面积^[2];结合遥感面积抽样框与系统等距抽样方法对伊朗哈马丹省农作物开展空间抽样调查^[3];采用系统抽样方法估算土地覆盖,研究结果表明变化估算效果好,而估算土地覆盖面积并不理想^[4]。然而,大多研究都采用 Landsat、HJ 星、MODIS 等中低分辨率遥感数据,难以保证小区域农作物种植面积的高精度提取^[5-9]。且农作物种植面积监测研究的区域主要局限于地形相对简单的北方平原地区,地形复杂多样的山地丘陵区的相关研究并不多见^[5-7,10]。

本研究以福建闽侯县为研究区,基于分辨率为 2 m 的国产 GF-1 号卫星影像,采用简单随机抽样、系统

收稿日期: 2017-05-14 修回日期: 2017-09-12

基金项目: 全国统计科学研究项目(2015LZ18)。

作者简介: 林芳芳(1991-)女,硕士研究生,研究方向:资源环境统计.Email: fafu_linfangfang@163.com. 通讯作者刘金福(1966-)男,教授,博士生导师,研究方向:资源环境统计.Email: fjljf@126.com.

抽样、分层抽样、空间随机抽样、空间系统抽样及空间分层抽样方法对农作物种植面积进行抽样估算, 比较不同抽样方法的优劣, 筛选最优抽样方法, 旨在为福建丘陵区农作物种植面积遥感监测提供方法。

1 材料与方法

1.1 研究区概况

福建省闽侯县(N25°47′—26°37′, E118°51′—119°25′) 隶属福建省福州市, 土地总面积 2 136 km²。全县地形地貌错综复杂, 中山、低山、丘陵、盆谷地以及平原兼有, 主要分布在县境北部和西南部。闽侯县境内年平均气温 14.8~19.5 ℃, 多年平均降水量为 1 673.9 mm, 属亚热带季风气候。闽侯县作为福建省主要产粮区之一, 粮食作物以水稻为主, 其次是甘薯、大豆、马铃薯等。

1.2 数据处理

以福建闽侯县作为研究区, 数据来源主要包括两部分: (1) 闽侯县 1 : 50 000 行政区划图矢量数据; (2) 覆盖研究区的 2016 年 5 月份的国产 GF-1 号遥感影像。

GF-1 号遥感影像数据处理主要包括全色与多光谱影像辐射校正、数据融合、几何校正等处理。具体步骤: (1) 采用 ENVI 5.1 软件对国产 GF-1 号遥感影像进行辐射定标、大气校正和正射校正; (2) 采用 ENVI 5.1 软件的 PC Spectral sharpening 融合功能模块对 PMS 传感器的多光谱和全色波段进行融合; (3) 以地形图为参考, 选择控制点, 通过重采样纠正 GF-1 号遥感影像的几何偏差, 误差控制在 0.5 个像元以内。GF-1 号影像处理效果见图 1。

面向对象的分类方法能够充分利用高、中分辨率遥感影像丰富的光谱、纹理细节信息, 通过更精确刻画影像地物的尺寸、形状、邻域地物的关系, 提高信息提取的精度^[10]。以 eCognition 9.0 软件为操作平台, 利用面向对象分类方法完成研究区农作物数据的提取。首先, 采用多尺度分割算法对处理后的 GF-1 号遥感影像进行分割, 根据影像分割结果及目视判断效果设置分割尺度为 15, 形状参数为 0.1, 紧凑度为 0.5; 然后, 利用归一化植被指数(NDVI) 将植被与非植被区域分开; 最后, 在植被区根据植被(农作物、林地与草地等)的光谱特征、NDVI 以及纹理信息, 分别建立不同类别的隶属度函数, 将农作物提取出来。

在采用面向对象分类法提取农作物的基础上, 结合野外调查、目视解译及 Google earth 卫星影像对分类结果进行修正, 获取闽侯县农作物遥感识别数据, 提取结果见图 2。将该提取结果与 2016 年闽侯县统计年鉴中农作物种植面积行比较, 分类精度达 91.28%。国产 GF-1 号卫星遥感影像分辨率达 2 m, 提取的农作物遥感识别数据可近似为地面值。



图 1 GF-1 号影像处理效果图

Fig.1 Effect diagram after GF-1 image processing

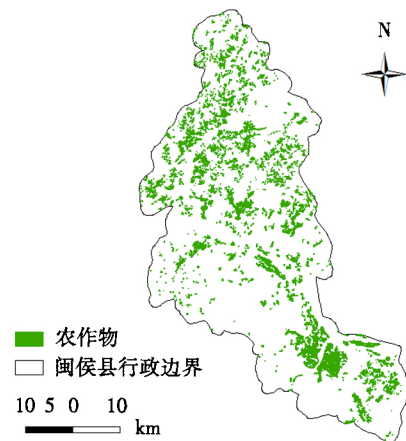


图 2 闽侯县农作物的空间分布

Fig.2 Spatial distribution of crops in Minhou County

1.3 研究方法

1.3.1 抽样单元尺寸优选 抽样单元尺寸是空间抽样方案设计过程的基础要素, 以正方形网格作为抽样单元基础形状, 考虑闽侯县县域面积及 GF-1 号卫星影像的空间分辨率, 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、2 000 m×2 000 m、2 500 m×2 500 m、3 000 m×

3 000 m等 8 种抽样基础单元尺寸方案^[12].利用 ArcGIS 10.2 软件 Fishnet 工具对研究区进行分割,得到 8 种抽样框.基于农作物遥感识别数据,统计各个抽样框中抽样基础单元内农作物面积,计算面积比(单个方格内农作物面积与方格面积的比值).计算 8 种抽样基础单元尺寸方案下农作物面积变动系数的变化情况,确定最优抽样单元尺寸.总体变动系数计算公式为^[13]:

$$c = \frac{\sqrt{\frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2}}{\bar{y}} \tag{1}$$

式中: c 为总体变动系数; N 为总体基础单元数; y_i 为第 i 个基础单元的面积比; \bar{y} 为总体相应农作物面积比.

1.3.2 抽样方法 采用简单随机抽样、系统抽样、分层抽样、空间随机抽样、空间系统抽样及空间分层抽样方法对研究区农作物种植面积进行样本抽选、总体推算及误差估计,筛选最优抽样方法.

传统抽样方法主要有简单随机抽样、系统抽样和分层抽样方法.简单随机抽样是从含有 N 个单元的总体中随机、独立地抽取 n 个单元组成样本.简单随机抽样样本容量 n 的计算公式见文献[12,14].系统抽样又称等距抽样,在进行抽样时先依样本容量确定抽样间隔,然后对总体抽样单元按一定顺序进行编号,随机确定抽样起始点;最后每隔一定的间隔抽取一个样本单元.系统抽样的样本容量确定、总体推算与误差估计公式与随机抽样相同.分层抽样以面积比(单个方格内农作物面积与方格面积的比值)为分层标志,采用累计等值平方根法获得最优分层^[16].以单个方格内农作物面积比为分层变量,根据最优分层原则,按 5% 的组距进行分组,计算各组频率和频率平方根累加值,再合并样本单元小于 2 的分组后进行重新分层,得到层数和分层界限;根据最优分配原则计算样本容量及各层样本容量,具体计算公式见文献[11,16].

空间抽样方法考虑抽样样本间的空间自相关,可有效减少样本冗余,降低调查成本^[17,18].Moran's I 指数是一种通用的空间自相关测量方法,其模型定义为^[19]:

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

式中: I 为 Moran's I 指数; N 为研究区总体单元数; x_i, x_j 为空间抽样基础单元的面积值; μ 为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 间的空间关系.

计算出空间相关性指数后,样本单元间空间自相关显著性检验步骤见文献[20].与传统抽样方法相比,空间抽样方法是在样本单元间自相关性显著的前提下,计算样本量.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^[21].空间抽样中均值的方差 V 表示为:

$$V = \text{var}\{\bar{Z} - Z(A) | Z\} = \frac{1}{n} [\sigma^2 - E\{C(X, Y)\}] = \frac{\sigma^2}{n} - \frac{E\{C(X, Y)\}}{n} \tag{3}$$

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.空间抽样方法的均值方差比传统抽样方法减少 $E\{C(X, Y)\}$,因此,空间抽样样本容量为:

$$n' = \frac{\sigma^2 - R}{V} = n(1 - r) \tag{4}$$

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.

1.3.3 抽样方法评价 为定量评价 6 种抽样方法对农作物种植面积估算的优劣,根据样本统计量推算总体特征,从相对误差、抽样稳定性及抽样成本三方面对抽样结果进行比较分析^[22],筛选最优抽样方法.为避免抽样偶然性,相对误差和变异系数结果均为连续抽取 20 套样本的平均值.

采用简单估计量进行样本统计量推算与误差估计.总体相对误差按式(5)计算,总体总值估计量的变异系数按式(6)计算^[10].

$$r = \frac{|\hat{Y} - Y|}{Y} \times 100\% \tag{5}$$

$$CV(\hat{Y}) = \frac{\sqrt{v(\hat{Y})}}{\hat{Y}} \times 100\% \tag{6}$$

式中: r 为总体相对误差; CV 为总体总值估计量的变异系数; Y 为总体总值的真值; \hat{Y} 为总体总值的无偏估计.

1.4 技术流程

基于国产 GF-1 号遥感数据,以遥感识别数据作为地面抽样真值数据,采用简单随机抽样、系统抽样、分层抽样、空间随机抽样、空间系统抽样及空间分层抽样方法等 6 种方法对研究区农作物种植面积进行抽样估算,具体技术流程见图 3:(1) 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、2 000 m×2 000 m、2 500 m×2 500 m、3 000 m×3 000 m 等 8 种抽样基础单元尺寸方案^[12],分别计算各方案的总体变动系数,通过分析总体变动系数的变化情况确定最优抽样单元尺寸;(2) 基于最优抽样单元尺寸,构建抽样框;(3) 采用传统抽样方法和空间抽样方法对抽样框内总体单元进行抽样,确定样本容量并进行总体推算及误差计算;(4) 对抽样结果进行对比分析,筛选最优抽样方法.

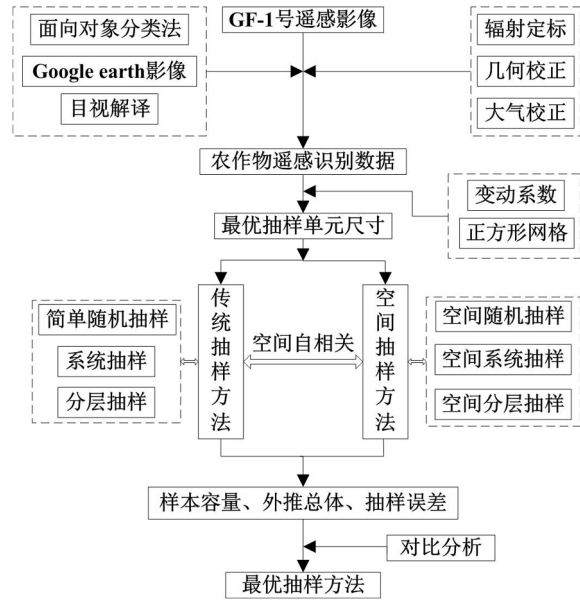


图 3 技术流程

Fig.3 Technical process

2 结果与分析

2.1 抽样单元尺寸优选

从表 1 可知:8 种抽样单元尺寸方案下,随着抽样单元尺寸的不断增大,闽侯县农作物种植面积变动系数的变化幅度逐渐减小;抽样单元尺寸增至 1 500 m×1 500 m 后,变动系数的变化情况逐渐趋于稳定(小于 2%)。因此,综合比较 8 种抽样单元尺寸方案变动系数的变化情况,选取 1 500 m×1 500 m 作为最优抽样单元尺寸.

2.2 抽样方法对比

以 1 500 m×1 500 m 为抽样单元尺寸构建抽样框,采用传统抽样方法和空间抽样方法对福建闽侯县农作物种植面积进行样本抽样,根据样本统计量推算总体特征,得到农作物种植面积空间抽样的相对误差、变异系数及抽样成本(表 2)。为便于对比、分析福建丘陵区农作物面积空间抽样方法的优劣,图 4、5 描绘了 6 种抽样方法的相对误差及变异系数,图 6 描绘了 6 种抽样方法的抽样成本.

根据表 2、图 4~6,对比 6 种抽样方法的估算结果,其相对误差及变异系数分别为 3.86%~9.20%和 5.21%~10.25%,说明 6 种抽样方法的相对误差较小,抽样稳定性较好.分别对比传统抽样方法和空间抽样方法的相对误差及变异系数,结果表明在样本容量相同的情况下,系统抽样优于简单随机抽样,空间系统抽样优于空间随机抽样.在样本容量不相同的情况下,

表 1 农作物种植面积变动系数随抽样单元尺寸的变化情况

Table 1 Changes in variation coefficient of crop acreage with different sampling unit sizes

抽样单元尺寸	变动系数	变化情况/%
500 m×500 m	2.354 9	
800 m×800 m	2.081 1	-11.63
1 000 m×1 000 m	1.954 6	-6.08
1 200 m×1 200 m	1.884 6	-3.58
1 500 m×1 500 m	1.810 4	-3.93
2 000 m×2 000 m	1.775 9	-1.91
2 500 m×2 500 m	1.739 7	-2.04
3 000 m×3 000 m	1.714 1	-1.47

表 2 研究区农作物种植面积 6 种抽样方法总体外推与误差估计结果

Table 2 Population extrapolation and error estimation for crop acreage in Minhou County via 6 sampling methods

抽样方法	样本容量	相对误差 %	变异系数 %	抽样成本
简单随机抽样	614	9.20	5.21	100.00
系统抽样	614	7.25	5.21	100.00
分层抽样	88	5.05	7.07	14.33
空间随机抽样	252	6.53	9.80	41.04
空间系统抽样	252	4.70	10.25	41.04
空间分层抽样	37	3.86	6.03	6.03

由于简单随机抽样与系统抽样的样本容量较大,故变异系数相对较小.而分层抽样的样本容量明显降低且相对误差及变异系数较小,说明通过划分分层,增大各类型抽样单元间的共同性,抽出具有代表性的调查样本,可显著改善总体误差及抽样稳定性.

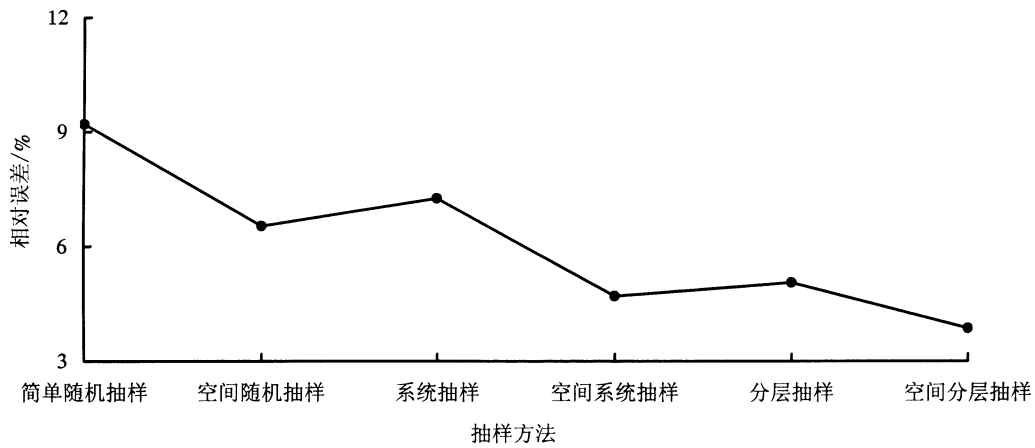


图 4 6 种抽样方法的相对误差

Fig.4 The relative error of variation for 6 sampling methods

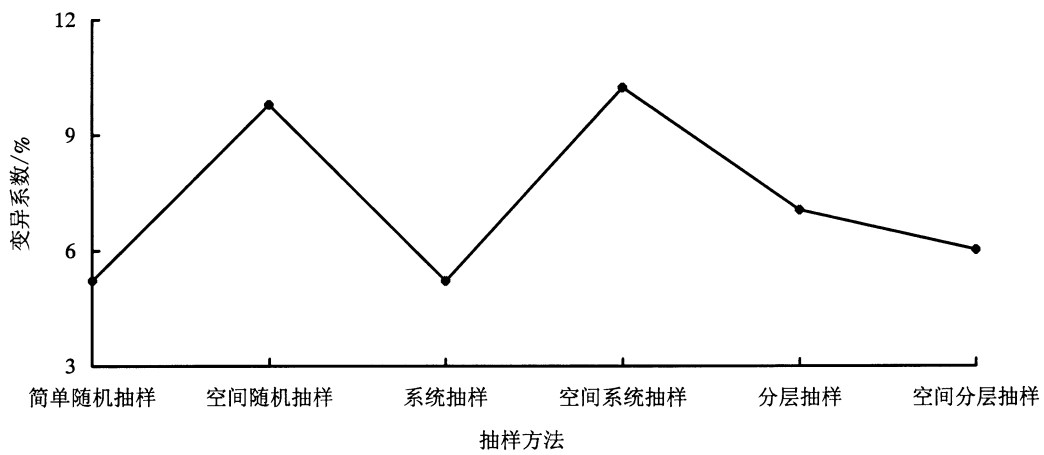


图 5 6 种抽样方法的变异系数

Fig.5 The variation coefficient of variation for 6 sampling methods

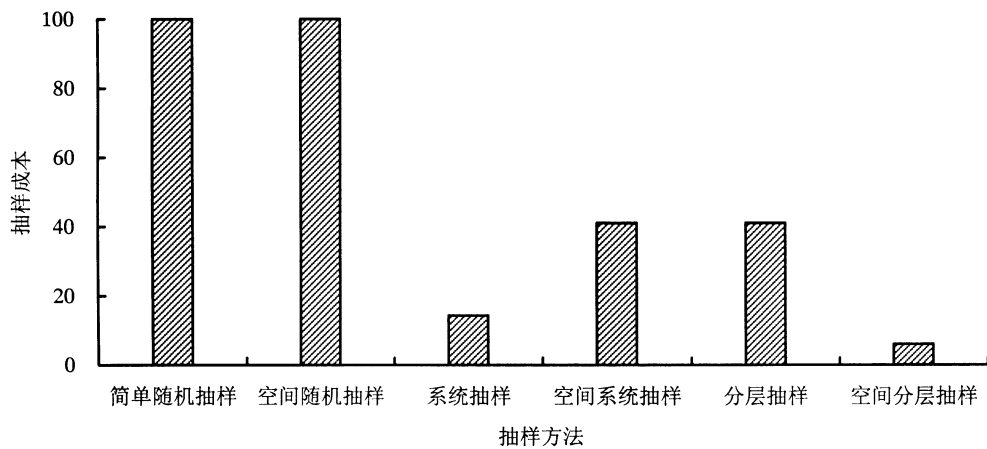


图 6 6 种抽样方法的抽样成本

Fig.6 The sampling cost of 6 sampling methods

由图 4~6 可知,与传统抽样方法相比,考虑抽样单元间空间自相关性时,空间抽样方法虽不能明显改善总体误差及抽样稳定性,但可显著降低样本容量,减少样本冗余,节约抽样成本.特别是空间分层抽样的

抽样成本仅为简单随机抽样的 6.03% , 显著降低抽样成本。

以总体相对误差、抽样稳定性及抽样成本作为抽样方法的评价指标, 综合比较 6 种抽样方法, 结果表明空间分层抽样方法的农作物种植面积空间调查效率最高。

2.3 空间分层抽样

通过对比 6 种抽样方法的相对误差、抽样稳定性及抽样成本, 空间分层抽样方法的抽样效率最高。对采用传统抽样方法获得的 87 个抽样单元进行空间自相关显著性检验, 求得其 Moran's I 为 0.58, 标准化 Z 值为 28.48, 远大于置信度水平为 99% 时的临界值(2.58)。可见, 福建省闽侯县农作物种植面积抽样单元空间自相关性较强。考虑抽样单元间的空间自相关性, 空间分层抽样方法的样本容量为 37。表 3 给出了空间分层抽样方法的分层参数, 采用空间分层抽样方法将闽侯县抽样基础单元分为 3 层, 根据各层层权比例, 分配到各层的样本量分别为 26、9、2, 在各层内随机抽取样本点(图 7)。

由表 3 及图 7 可知, 抽样基础单元主要集中在第 1 层(面积比 0% ~ 5%), 第 3 层(面积比 20% ~ 100%) 的抽样基础单元数则较少。说明闽侯县农作物种植地块较为破碎, 特别是闽侯县北部丘陵地区地形复杂, 农作物地块破碎程度高。根据空间分层抽样方法, 37 个样本点随机分布于全县, 其相对误差为 3.86% , 变异系数为 6.03% , 抽样成本为 6.03。在考虑空间自相关的情况下, 采用空间分层抽样方法可获得较高的估算精度, 表现出高度的稳定性, 样本量较少, 从而降低调查成本。

3 小结

基于遥感技术, 以福建丘陵区闽侯县作为研究区, 采用国产 GF-1 号高分辨率遥感影像对研究区农作物空间分布数据进行提取, 考虑抽样单元间空间自相关性, 采用传统抽样方法和空间抽样方法对农作物面积进行抽样估算, 对比分析 6 种抽样方法在总体相对误差、抽样稳定性及抽样成本三方面的优劣, 结果表明:

(1) 8 种抽样单元尺寸方案下, 抽样单元尺寸增至 1 500 m×1 500 m 后, 变动系数的变化情况逐渐趋于稳定(小于 2%)。因此, 1 500 m×1 500 m 作为最优抽样单元尺寸。

(2) 传统抽样方法及空间抽样方法的相对误差及变异系数分别为 3.86% ~ 9.20% 和 5.21% ~ 10.25% , 相对误差较小, 抽样稳定性较好。与传统抽样方法相比, 空间抽样方法考虑了抽样单元间的空间自相关性, 显著降低抽样调查的样本容量, 减少样本冗余, 节约抽样成本。

(3) 福建省闽侯县耕地结构复杂, 内部差异较大, 农作物调查目标空间自相关性显著性较强。因此, 综合比较抽样结果, 空间分层抽样方法在各方面均优于其他抽样方法, 适宜于福建山地丘陵区农作物种植面积遥感监测。

参考文献

- [1] 吴炳方. 中国农情遥感速报系统 [J]. 遥感学报, 2004, 8(6): 481-497.
- [2] 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.
- [3] PRADHAN S. Crop area estimation using GIS, remote sensing and area frame sampling [J]. International Journal of Applied Earth Observation, 2001, 3(1): 6-92.

表 3 分层参数表

Table 3 Parameter of stratified sampling

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

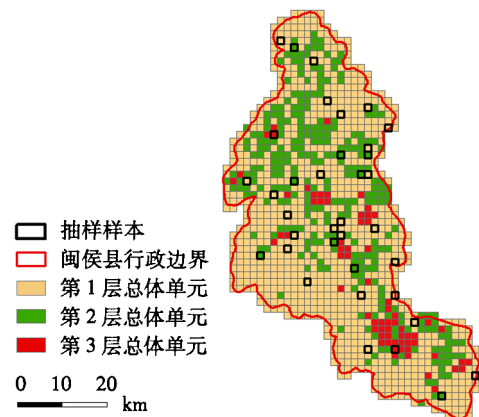


图 7 闽侯县农作物面积空间分层抽样样本的分布
Fig.7 Distribution of stratified sampling samples of crop acreage in Minhou County

- [4] GALLEGO F J. The efficiency of sampling very high resolution images for area estimation in the European Union[J]. *International Journal of Remote Sensing*, 2012, 33(6): 1 868-1 880.
- [5] 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.
- [6] 吴炳方,李强子.基于两个独立抽样框架的农作物种植面积遥感估算方法[J].*遥感学报*, 2004, 8(6): 551-569.
- [7] 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.
- [8] 黄冬梅,陈珂,王振华,等.利用空间抽样理论的遥感影像分类结果精度评价方法[J].*计算机应用与软件*, 2016, 33(7): 190-194.
- [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] 张建国,李宪文,吴延磊.面向对象的冬小麦种植面积遥感估算研究[J].*农业工程学报*, 2008, 24(5): 156-160.
- [11] 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.
- [12] 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.
- [13] 王海宾,邓华锋,程志楚,等.基于 3S 的森林植被面积空间抽样方法[J].*森林与环境学报*, 2015, 35(1): 74-80.
- [14] 宋新民,李新良.抽样调查技术[M].北京:中国林业出版社, 2007.
- [15] 张小伟.森林覆盖面积遥感监测中空间抽样技术及应用[D].南京:南京林业大学, 2011.
- [16] 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.
- [17] STEHMAN S V, SOHL T L, LOVELAND T R. Statistical sampling to characterize recent United States land-cover change[J]. *Remote Sensing of Environment*, 2003, 86(4): 517-529.
- [18] HAINING, R. *Spatial Data Analysis: Theory and Practice*[D]. Cambridge: Cambridge University Press, 2003.
- [19] 孙皓晗.空间分层抽样在森林覆盖监测中的应用研究[D].南京:南京林业大学, 2013.
- [20] 黄冬梅,陈珂,王振华,等.利用空间抽样理论的遥感影像分类结果精度评价方法[J].*计算机应用与软件*, 2016, 33(7): 190-242.
- [21] 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.
- [22] 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.

(责任编辑:叶济蓉)