

# 玉米种植面积空间抽样调查方案优化设计

王 迪<sup>1,2</sup>, 周清波<sup>1,2\*</sup>, 陈仲新<sup>1,2</sup>, 刘 佳<sup>1,2</sup>

(1. 农业部农业信息技术重点实验室, 北京 100081; 2. 中国农业科学院农业资源与农业区划研究所, 北京 100081)

**摘 要:** 抽样比、样本空间布局及抽样单元尺度是组成空间抽样调查方案的基础要素。为进一步改善现行农作物种植面积空间抽样调查效率, 该文以吉林省德惠市为研究区, 以玉米种植面积为研究对象, 选取正方形网格作为抽样单元, 通过空间分析、“3S”技术与传统抽样方法相结合进行农作物种植面积空间抽样方案优化设计试验研究。结果表明, 抽样单元间空间自相关性随单元尺度的增大而增大, 两者间呈线性正相关关系。当抽样单元尺度为 500 m×500 m 时, 抽样单元间空间自相关性几乎不存在。遵循传统抽样理论要求样本间相互独立原则, 选取 500 m×500 m 作为最优抽样单元尺度; 对抽样单元内玉米种植面积与耕地面积进行相关分析发现, 两者间存在极显著线性正相关关系。为提高玉米种植面积空间分层抽样效率, 可选取耕地面积作为分层标志; 以抽样外推总体相对误差( $r$ )和变异系数(coefficient of variation, CV)为空间抽样效率评价指标, 在 4 种(简单随机、系统等距、分层随机及分层系统等距)样本空间布局方式中, 选取分层系统等距抽样作为最优样本布局方式; 在 7 种抽样比(0.5%、1.0%、1.5%、2.0%、2.5%、3.0%、3.5%)设计水平中, 选取 1%作为最优抽样比。该文可为提高农作物面积空间抽样调查效率提供试验依据。

**关键词:** 作物; 外推总体; 误差分析; 空间抽样; 玉米; 种植面积; 抽样单元尺度; 样本空间布局

doi: 10.3969/j.issn.1002-6819.2014.08.014

中图分类号: S127

文献标识码: A

文章编号: 1002-6819(2014)-08-0117-09

王 迪, 周清波, 陈仲新, 等. 玉米种植面积空间抽样调查方案优化设计[J]. 农业工程学报, 2014, 30(8): 117—125.

Wang Di, Zhou Qingbo, Chen Zhongxin, et al. Optimization of spatial sampling schemes for maize planting acreage estimation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(8): 117—125. (in Chinese with English abstract)

## 0 引 言

农作物种植面积信息是国家制定粮食政策和经济计划的重要依据<sup>[1-2]</sup>, 及时准确掌握农作物种植面积信息对于加强农业生产管理、确保中国粮食安全具有重要意义<sup>[3-4]</sup>。目前, 国内外主要通过“3S”技术(遥感技术、地理信息技术、全球定位技术)与传统抽样方法相结合构建空间抽样调查方案进行大区域农作物种植面积估计<sup>[5-9]</sup>。由于抽样比、样本空间布局和抽样单元尺度(分别反映了样本的数量多少、样本的空间布设方式及样本的空间尺寸大小)是组成空间抽样调查方案的 3 个基础要素<sup>[10]</sup>,

因此, 优化设计这 3 个基础要素对于提高空间抽样调查效率显得尤为重要。以往国内外在农作物种植面积空间抽样调查方案设计研究方面, 大多侧重于采用单一的空间抽样方案进行区域农作物种植面积估算, 如美国的大面积农作物估产计划(LACIE 计划, Large Area Crop Inventory and Experiment)和欧盟的农业遥感监测计划(MARS 计划, Monitoring Agriculture with Remote Sensing)均采用分层抽样方案对本区域的农作物种植面积进行监测<sup>[11-13]</sup>。国内在农作物面积空间抽样方面的研究主要集中在如何通过“3S”技术与传统抽样方法相结合来提高抽样调查效率, 采用的抽样方法主要包括分层抽样<sup>[14-17]</sup>和整群抽样<sup>[18]</sup>, 涉及的农作物类型有小麦、水稻和棉花等粮食和经济作物。如陈仲新等<sup>[14]</sup>采用分层抽样与遥感技术相结合, 建立了一套全国冬小麦面积变化抽样外推方法。该法以全国分县冬小麦面积为分层标志, 采用 TM 影像获取样本单元内冬小麦面积作为样本观测值进行全国冬小麦面积变化率的估计。吴炳方和李强子<sup>[18]</sup>提出基于作物种植结构区划, 采用整群抽样和样条采样技术结合的思路估算全国农作物面积。而对空间抽样方

收稿日期: 2013-10-30 修订日期: 2014-03-18

基金项目: 欧盟第 7 框架项目(270351); 国家自然科学基金(青年)资助项目(41001247); 农业部农业信息技术重点实验室开放基金(2013011); 中央级公益性科研院所基本科研业务费专项资金(IARRP-2014-17)

作者简介: 王 迪(1977—), 男, 副研究员, 工学博士, 主要研究方向为作物面积空间抽样方法。北京 中国农业科学院农业资源与农业区划研究所, 100081。Email: wangdicaas@126.com

\*通信作者: 周清波(1965—), 男, 研究员, 博士、博士生导师, 主要从事农情遥感领域的基础研究和应用基础研究。北京 中国农业科学院农业资源与农业区划研究所, 100081。Email: zhouqingbo@caas.cn

案的优化设计研究则较少。虽有部分学者对农作物种植面积空间抽样方案设计中涉及的抽样单元尺度、抽样方法进行了比较评价,如 Gallego 等<sup>[19]</sup>通过分析不同尺度下的抽样单元空间相关性,同时兼顾抽样调查成本,对欧盟 MARS 计划中的农作物种植面积年际变化快速估计行动采用的抽样单元尺度进行了优化设计研究。阳小琼等<sup>[20]</sup>以抽样精度、最小样本量和稳定性为评价指标,对水稻种植面积空间抽样方法进行了优选。张锦水等<sup>[21]</sup>则以规则网格为抽样单元,以抽样误差和抽样比为评价依据,对抽样网格大小、分层层数进行了优化设计研究。王迪等<sup>[22]</sup>采用两阶段抽样方法与遥感、地理信息系统技术相结合对山东省的冬小麦面积进行估算,并对两阶段抽样方案中的第 1 阶段的抽样方法进行了优选。鉴于抽样单元尺寸的合理制定对提高农作物面积空间抽样效率所具有的重要意义,王迪等<sup>[23]</sup>通过分析抽样单元尺寸与其对应的全局自相关指数的相关关系进行样方尺寸初选;在此基础上,以相对误差、变异系数和样本容量为抽样效率评价指标,通过定量评价不同样方尺寸下的冬小麦面积空间抽样效率最终实现对抽样单元尺寸的优选。然而,上述研究大多仅围绕空间抽样方案中的某一基础要素进行优选分析,缺乏对空间抽样方案设计所涉基础要素进行全面优化设计研究,从而限制了农作物种植面积空间抽样效率的进一步提高。

针对这一问题,本研究以吉林省德惠市为研究区,选取玉米种植面积为研究对象,通过空间分析、“3S”技术与传统抽样方法相结合,对农作物种植面积空间抽样方案中的 3 个基础要素(抽样比、样本空间布局 and 抽样单元尺度)进行优化设计试验研究,旨在为改善现行农作物面积抽样调查效率提供解决途径。

## 1 材料与方法

### 1.1 研究思路

农作物种植面积空间抽样调查方案优化设计的研究思路主要包括 5 个步骤:1) 抽样基础数据准备。包括研究区基础地理信息数据、研究区近期耕地和玉米空间分布数据的收集整理;2) 空间抽样方案设计。基于上述抽样基础数据,进行抽样单元尺度、分层标志、抽样比及样本空间布设方式设计;3) 抽样单元尺度优选。遵循传统抽样理论要求样本间相互独立原则,通过引入全局空间自相关指数 Moran's  $I$  进行抽样单元尺度优选;4) 样本抽选及总体推断。基于设计的空间抽样方案进行样本抽选、样本观测值获取、外推总体与误差估计;5) 样本空间布设方式和抽样比优选。以抽样外推总体

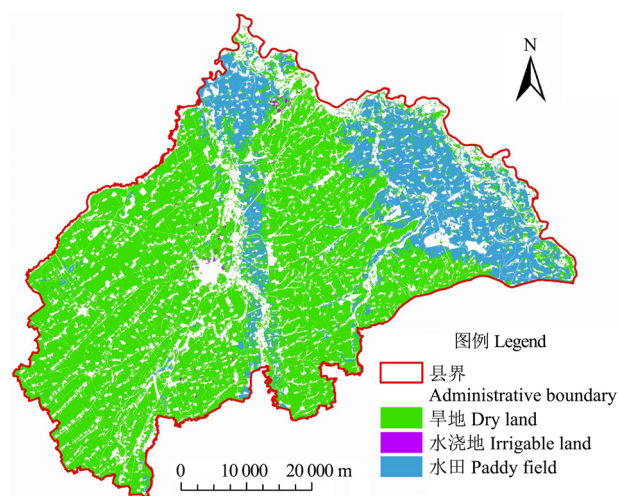
相对误差和变异系数为评价指标,进行样本空间布设方式优选。综合衡量抽样精度和抽样费用,进行抽样比优选。

### 1.2 研究区概况

德惠市位于吉林省中北部,属于松辽平原的主要农业自然经济类型区,地处  $44^{\circ}02' \sim 44^{\circ}53'N$ 、 $125^{\circ}14' \sim 126^{\circ}24'E$  之间,全市总面积  $3\,460\text{ km}^2$ ,其中耕地面积为  $2.54 \times 10^5\text{ hm}^2$ 。地形地貌类型为山前冲洪积台地,地势为南高北低,呈波状起伏,但相对落差不大,大部分海拔高度在  $149 \sim 241\text{ m}$  之间。土壤类型主要以黑土、草甸土、黑钙土和冲积土为主,占全市总土地面积的 90% 以上。气候属于半湿润带大陆性季风气候,四季分明,春季多风干旱,夏季炎热多雨,秋季昼暖夜爽,冬季寒冷漫长。多年平均气温  $4.4^{\circ}\text{C}$ ,年平均积温  $2\,851^{\circ}\text{C}$ ,年均日照时数为  $2\,695.2\text{ h}$ ,无霜期为  $139\text{ d}$ ,多年平均降雨量为  $520\text{ mm}$ ,主要集中在夏季。主要耕作制度为一年一熟制。农业生产以粮食作物为主,经济作物为辅。粮食作物主要有玉米、水稻、大豆、高粱,其中玉米是德惠市最重要的粮食种植作物。

### 1.3 基础数据

主要包括 3 部分:1) 基础地理信息数据。德惠市行政边界数据(比例尺  $1:1$  万,县界,矢量格式);2) 耕地空间分布数据。德惠市第二次全国土地大调查耕地图斑数据(比例尺  $1:1$  万,矢量格式);3) 农作物空间分布数据。德惠市 2009 年玉米空间分布数据(基于 SPOT5 影像提取,影像轨道号 296-260、296-261、297-260,获取日期:2009 年 8 月 1 日,空间分辨率  $10\text{ m}$ )。图 1 和图 2 分别给出了德惠市耕地和玉米空间分布情况。



注: 耕地空间分布数据提取自第二次全国土地大调查资料(2006 年)。  
Note: The spatial distribution of cultivated land is derived from the Second National Land Survey data(in 2006).

图 1 德惠市耕地空间分布

Fig.1 Spatial distribution of cultivated land in Dehui County

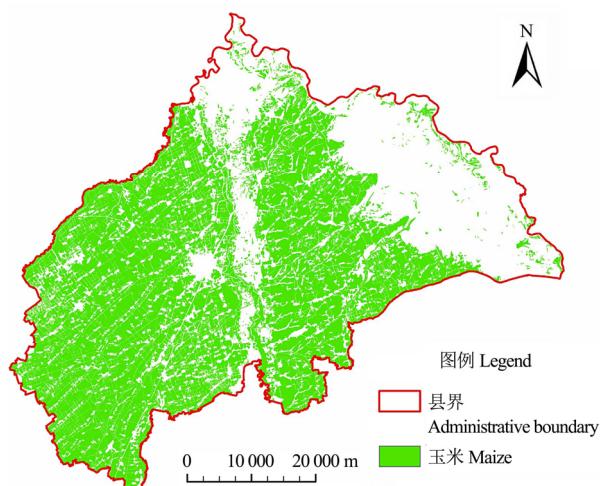


图 2 德惠市 2009 年玉米空间分布

Fig.2 Spatial distribution of maize in Dehui County in 2009

## 1.4 空间抽样方案设计

### 1.4.1 抽样单元尺度优选

遵循传统抽样理论要求样本间相互独立原则进行抽样单元尺度优选<sup>[24-25]</sup>。基于该原则，本研究选取空间分析中的全局空间自相关指数 Moran's  $I$  来评价不同尺度下的抽样单元空间自相关性。由于当 Moran's  $I$  等于或接近于 0 时，意味着抽样单元间的空间相关程度等于或近乎为 0，即相互独立。因此，通过分析多种尺度下的抽样单元空间自相关性，遵循样本间相互独立原则，进行抽样单元尺度的确定。为便于剖分研究区、构建抽样框，本研究选择正方形网格作为抽样单元形状。考虑样本单元调查时（地面调查或遥感调查）的可操作性和工作量大小，同时兼顾抽样调查的实际意义，以 500 m 为水平间隔，抽样单元尺度共设计 8 种水平，分别是 500 m×500 m，1 000 m×1 000 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，3 500 m×3 500 m，4 000 m×4 000 m。基于设计的抽样单元尺度离散研究区、构建抽样框。将研究区玉米空间分布数据与各种抽样框相叠加，获取抽样单元观测值。

全局空间自相关指数 Moran's  $I$  按式 (1) 计算<sup>[26]</sup>：

$$I = \frac{N}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_{i=1}^N \sum_{j=1}^N w_{ij}} \quad (1)$$

式中： $I$  是全局空间自相关指数 Moran's  $I$ ，取值范围为 $[-1, 1]$ ； $N$  是总体容量，即抽样框内全部总体单元数量； $Y_i$  是抽样框内第  $i$  个抽样单元内的玉米种植面积（抽样单元观测值）； $Y_j$  是抽样框内第  $j$  个

抽样单元内的玉米种植面积； $\bar{Y}$  是抽样框内各抽样单元玉米种植面积均值； $w_{ij}$  是空间邻接矩阵。

### 1.4.2 抽样比设计

参考以往相关文献报道<sup>[20, 27-28]</sup>中对抽样比的设定结果，结合考虑传统抽样方法对样本容量的要求（如分层抽样要求单层内样本容量  $n$  不得小于 2 等），本研究以 0.5% 为水平间隔，共设计 7 种水平的抽样比，分别是 0.5%、1.0%、1.5%、2.0%、2.5%、3.0% 和 3.5%。

### 1.4.3 样本空间布局设计

样本的空间布局采用 4 种设计方式：简单随机抽样、系统等距抽样、分层随机抽样和分层系统等距抽样。需要说明的是，为消除单套样本外推总体时产生的不稳定性，本研究针对每种抽样单元尺度的抽样框分别采用简单随机抽样、系统等距抽样、分层随机抽样和分层系统等距抽样方法抽选出 5 套样本。

简单随机抽样布局方式是指利用 ArcGIS9.3 软件对抽样框内的全部总体单元按照由 1~ $N$  顺序进行数字编码。采用 SPSS16.0 软件在 1~ $N$  个数字中生成与样本容量  $n$  相等数量的伪随机数，当总体单元的数字编码与生成的伪随机数相同时，该单元即被抽中，同时该样本单元的空间位置也被确定。基于上述过程即可完成对样本单元的空间布局设计。

系统等距抽样布设样本单元的过程如下：1) 利用 ArcGIS9.3 软件对抽样框内各总体单元的 ID 号按照由小到大顺序排列；2) 确定抽样间隔  $k$ 。通过总体容量  $N$  除以样本容量  $n$  并取整得到  $k$  值。将总体分为  $n$  段，则每段包含  $k$  个总体单元；3) 从第 1 段的  $k$  个总体单元中随机抽出 1 个单元作为起点；4) 对于其余的  $n-1$  段，每隔  $k$  个单元抽出 1 个样本单元，依此类推，直至抽出  $n$  个样本单元。

分层随机抽样布设样本单元的过程如下：1) 分层标志的设计。为提高分层抽样效率，在抽样框内选取与目标变量（玉米种植面积）高度相关的辅助变量作为分层标志；2) 层数设计。为降低层内总体单元方差，同时兼顾抽样调查成本，本研究设计分层层数为 5 层；3) 层限确定。采用累积频率等值平方根法确定各层层限；4) 样本容量分配。采用尼曼 (Neyman) 分配方式对样本容量在各层内进行再分配；5) 样本抽选。各层内采用简单随机抽样方式抽选样本单元。

分层系统等距抽样布设样本单元的过程基本同分层随机抽样，不同之处在于样本抽选，各层内采用系统等距方式抽选样本单元。

#### 1.4.4 样本观测值获取

由于本研究的目的是进行空间抽样方案优化设计,为定量评价各种空间抽样方案的效率优劣,基于研究区 2009 年玉米空间分布数据,采用 ArcGIS9.3 软件统计抽中样本单元内玉米种植面积作为样本观测值进行总体外推与误差估计。

#### 1.4.5 外推总体与误差估计

4 种样本单元布设方式均采用简单估计量进行总体外推与误差估计。其中简单随机抽样和系统等距抽样(由于在本研究中属无关标志排队等距抽样,因此可视为简单随机抽样)外推总体总值及其估计量方差的无偏估计参照文献[29]中的公式计算;分层随机抽样和分层系统等距抽样中外推总体总值及其估计量方差的无偏估计参照文献[30]中的公式计算。

为定量评价玉米种植面积空间抽样方案外推总体的误差和稳定性,本研究选取相对误差( $r$ )和总体总值估计量的变异系数(CV)2个指标<sup>[23]</sup>。

## 2 结果与分析

### 2.1 抽样单元尺度优选结果

为进行抽样单元尺度优选,图3绘出了8种抽样单元尺度下的全局空间自相关指数 Moran's  $I$  及相应的显著性检验指标 Z Score 的变化情况。由于 Moran's  $I$  反映了抽样框内各抽样单元间的空间自相关程度。在 $[0, 1]$ 区间内, Moran's  $I$  越接近 1, 说明抽样单元间的空间正相关性越强; Moran's  $I$  越接近 0, 说明抽样单元间的空间正相关性越弱, 当 Moran's  $I$  等于或几乎为 0 时, 说明抽样单元间的空间相关程度等于或近乎为 0, 即相互独立。显著性检验指标 Z Score 反映了抽样单元间存在空间自相关的显著性程度, Z Score 越大, 则拒绝原假设( $H_0$ : 抽样框内各抽样单元间不存在空间相关性)的概率越大; Z Score 越小, 则拒绝原假设的概率越小。从图3中可以看出, 随着抽样单元尺度的增大, 对应的 Moran's  $I$  和显著性检验指标 Z Score 也随着增大。这说明随着抽样单元尺度的增大, 抽样单元间的空间自相关性越强, 同时显著性程度也越高。当抽样单元尺度为  $500\text{ m} \times 500\text{ m}$  时, 此时对应的 Moran's  $I$  不足 0.1, Z Score 也最小, 说明抽样单元间的空间相关性几乎不存在, 即抽样单元间相互独立。另外, 通过试验分析还发现, 当抽样单元尺度从  $500\text{ m} \times 500\text{ m}$  处继续减小时, 对应的 Moran's  $I$  数值也会随之变小, 但变小的幅度很小。当抽样单元尺度为  $100\text{ m} \times 100\text{ m}$  时, 此时

的 Moran's  $I$  (0.042) 与  $500\text{ m} \times 500\text{ m}$  时的 Moran's  $I$  (0.060) 相差仅为 0.018, 两者均低于 0.1。参考以往关于空间自相关性评价方面的文献<sup>[31]</sup>表明, 当 Moran's  $I$  小于 0.1 时, 即可视为空间相关性几乎不存在, 也就是说当抽样单元尺度小于等于  $500\text{ m} \times 500\text{ m}$  时, 抽样单元间的空间相关性几乎都不存在。因此, 在满足传统抽样理论要求样本间相互独立条件下, 如果降低抽样单元尺度会产生 3 个弊端: 一是总体容量增大, 抽样单元变异系数增大, 样本容量增加从而导致抽样费用增加; 二是抽样单元内地物对总体的代表性降低, 进而降低抽样外推总体精度; 三是抽样调查工作量增加。因抽样单元变异系数增大导致样本容量增加, 使得样本单元在空间上的分布更为分散, 从而导致样本调查的工作量增大。综合考虑上述因素, 本研究选取  $500\text{ m} \times 500\text{ m}$  作为最优的抽样单元尺度。

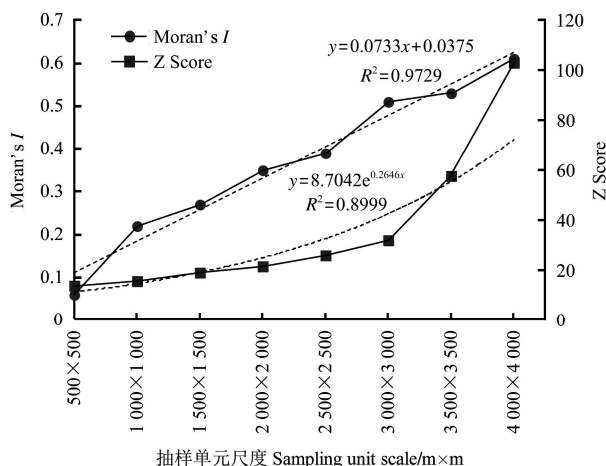


图3 全局空间自相关指数 Moran's  $I$  随抽样单元尺度的变化情况

Fig.3 Changes of global spatial autocorrelation index (Moran's  $I$ ) at different sampling unit scales

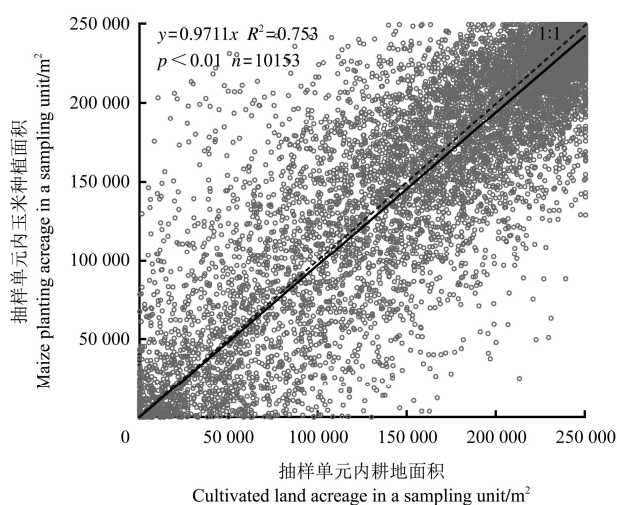
另外, 还需要说明的是, 本研究在进行抽样单元尺度分析时使用的是研究区历史近期玉米空间分布数据。利用该数据进行抽样单元尺度分析的原因具体包括 2 个方面: 一是数据的真实性。因为本文的研究对象为玉米种植面积, 抽样单元的属性值同样为玉米种植面积, 为真实反映抽样单元内的玉米种植面积空间相关特征, 因此, 本研究在进行抽样单元尺度分析时使用的是研究区玉米空间分布数据, 如果使用耕地空间分布数据, 由于耕地与玉米的空间分布不完全一致, 在确定抽样空间相关性时会存在一定误差; 二是数据的可获取性。为定量确定抽样单元空间相关程度, 需要获取研究区全部抽样单元的属性值。如果采用地面调查方式获取全



部抽样单元内的玉米种植面积存在可操作性难度较大、获取成本高等问题,为实现本文的研究目的,故选择使用基于遥感影像提取的研究区历史近期玉米空间分布数据进行抽样单元尺度分析。

## 2.2 分层标志选择

分层标志是影响分层抽样效率的关键因子。为提高分层抽样效率,需要选择与本研究中的目标变量(即玉米种植面积)高度相关的辅助变量作为分层标志。鉴于玉米种植于耕地之上(即有耕地的区域才有玉米种植),同时考虑到辅助变量信息的实际可获取性问题,本研究初步选择耕地面积作为分层标志进行玉米种植面积分层抽样(包括分层随机抽样和分层系统等距抽样)设计。为定量分析研究区耕地面积与玉米种植面积间的相关程度,图4绘出了抽样单元尺度为 $500\text{ m}\times 500\text{ m}$ 的抽样框内各抽样单元内耕地面积与玉米种植面积散点图。从图4中可以看出,抽样单元内的玉米种植面积与耕地面积呈线性正相关关系,对回归方程进行显著性检验, $F(1, 10\ 151)=3.181\times 10^4>F_{0.01}(1, 10\ 151)=6.63$ ,说明两者间存在极显著的线性正相关关系,这表明抽样单元内的耕地面积可以被选择作为本研究中的分层标志。



注:抽样单元尺度为 $500\text{ m}\times 500\text{ m}$ 。

Note: The sampling unit scale is  $500\text{ m}\times 500\text{ m}$ .

图4 抽样单元内耕地面积与玉米种植面积散点图

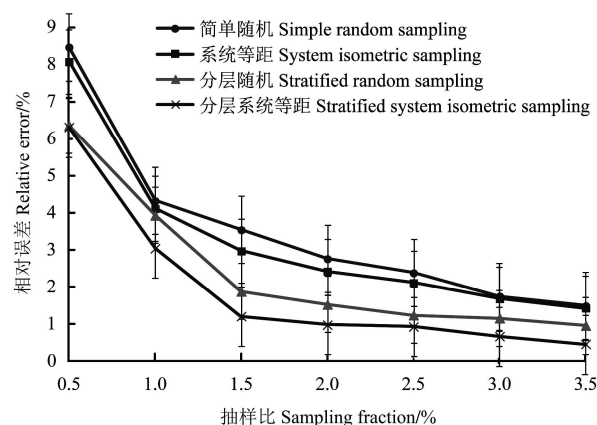
Fig.4 Scatter diagram of maize planting acreage and cropland area in sampling units

## 2.3 样本空间布设方式和抽样比优选

### 2.3.1 样本空间布设方式

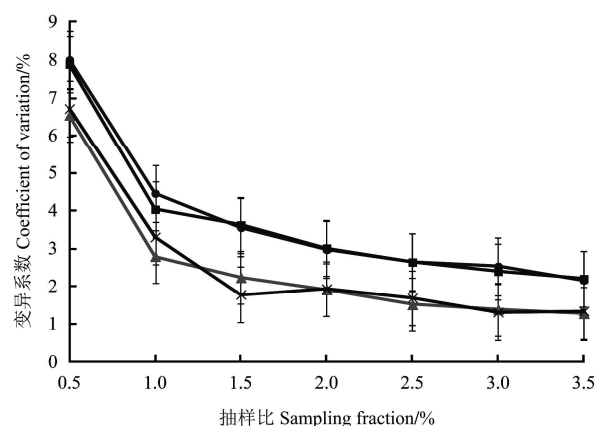
为了优选样本空间布局方式,图5a和图5b分别绘出了抽样外推总体相对误差 $r$ 和总体总值估计量的变异系数(coefficient of variation, CV)随样本空间布局方式的变化情况。从图5a中可以

看出,在抽样比相同条件下,简单随机抽样样本空间布局方式下的 $r$ 值最高,其余按照由高到低的顺序依次为系统等距、分层随机、分层系统等距的样本布局方式。4种样本布局方式中,简单随机和系统等距布局方式下的 $r$ 值比较接近,而分层随机和分层系统等距布局方式的 $r$ 值明显低于前二者。这说明采用分层系统等距的样本空间布局方式进行研究区玉米种植面积总体外推精度最高。从图5b中可以看出,在抽样比相同条件下,简单随机和系统等距样本布局方式下的研究区玉米种植面积总体总值估计量的CV值近乎相同;分层随机与分层系统等距样本布局方式下的CV值大体上比较接近,但后两者的CV值明显低于前两者。这说明在4种样本空间布局方式中,采用分层系统等距的样本空间布局方式进行总体估计时的稳定性最高。综合衡量抽样外推总体的精度与稳定性这2个指标,可以看出分层系统等距的样本空间布局方式效率最高。



a. 抽样外推总体相对误差

a. Relative errors of population extrapolation



b. 总体总值估计量的变异系数

b. CV of population total estimator

图5 抽样相对误差、变异系数随样本布局方式和抽样比的变化曲线

Fig.5 Relative error, CV curves of different spatial samples layouts and sampling fractions

### 2.3.2 抽样比优选

以分层系统等距样本布局为例,从图 5a 中可以看出,随着抽样比的增大, $r$  值逐渐减小。在 [0.5%, 1.0%] 区间, $r$  值从 6.32% 降至 3.05%,降低速度较快;而在 [1.0%, 3.5%] 区间, $r$  值从 3.05% 降至 0.47%,降低速度则相对较为平缓。另外,从图 5b 中可以看出,随着抽样比的增大,CV 值也呈逐渐减小趋势。在 [0.5%, 1.0%] 区间,CV 值从 6.72% 降至 3.31%,同样减小速度较快;而在 [1.0%, 3.5%] 区间,CV 值从 3.31% 降至 1.35%,减小速度明显降低。从  $r$  值和 CV 值随抽样比的变化曲线中可以发现,在抽样比等于 1.0% 处是曲线变化的一个拐点位置。在该点之前增大抽样比可以显著改善抽样外推总体精度和稳定性;而在该点之后增大抽样比,则对于改善抽样外推总体精度和稳定性的作用不明显。另外,当抽样比为 1% 时,采用分层系统等距的样本空间布局方式进行总体外推的  $r$  值和 CV 值分别为 3.05% 和 3.31%,均满足国家统计局部门对农作物种植面积抽样调查精度和稳定性的要求,即  $r$  值和 CV 值分别不得高于 5%<sup>[32]</sup>。综合考虑抽样调查费用和抽样误差,选择 1% 作为本研究中的最优抽样比。

## 3 讨 论

基于德惠市玉米种植面积空间抽样调查方案优化设计研究结果,为进一步改善农作物种植面积空间抽样效率提供了一定的参考依据,但在本文的研究过程中仍存在一些有待进一步解决的问题。如农作物面积空间抽样调查方案优化设计涉及不同抽样方法、不同抽样框、作物的动态性及精度等多个问题,而本研究仅从抽样比、样本空间布局和抽样单元尺度 3 个基础要素角度对农作物面积空间抽样方案进行优化设计,尚缺乏针对不同抽样方法、抽样框、作物动态性及抽样精度对农作物空间抽样调查效率影响的定量研究。另外,如何在更大区域尺度(如省级研究区)对本方法进行推广应用,还需要进一步验证研究。针对于此,本文拟在以下 4 个方面开展后续研究,一是通过比较不同抽样方法(简单随机抽样、系统抽样、分层抽样、多阶段抽样等)的农作物面积抽样调查效率,进行空间抽样方法优选研究;二是针对不同类型抽样框(如圆形抽样框、六边形抽样框、以自然地块边界为抽样框等)的农作物面积空间抽样方案优化设计研究;三是进行农作物种植动态性对空间抽样效率的影响研究,定量确定农作物种植的动态变化特征,深入

分析农作物在区域空间上的分布特征与空间异质性,为空间抽样方案的优化设计提供参考依据;四是选取多种估计量外推农作物面积总体与误差估计,优选适合农作物面积空间抽样调查的估计量,进一步改善空间抽样外推农作物面积总体的精度。

还需要说明的是,本文采取研究区第二次全国土地大调查耕地图斑数据作为辅助变量进行空间抽样方案设计,以玉米种植面积遥感识别结果作为抽样外推总体精度的检验依据,最终实现对空间抽样方案的优化设计研究。在此过程中,本研究仅从影响抽样误差自身的各因素(抽样比、样本布局及抽样单元尺度)角度进行空间抽样方案优化设计研究,而就非抽样误差(如遥感分类误差)对方案设计结果的影响尚未考虑。实际上,遥感分类误差对抽样方案设计结果的影响体现为两个方面:一是从样本观测值角度而言,当遥感分类误差较大时,基于样本观测值外推总体的误差也成倍增大;二是从总体空间分布角度而言,如遥感分类误差较大,则设计变量(玉米种植面积)与辅助变量(耕地面积)间的相关性也随之降低,从而导致利用耕地作为辅助变量推断玉米面积的误差也相应变大。

## 4 结 论

为进一步提高农作物种植面积空间抽样调查效率,本研究以吉林省德惠市为研究区,选取玉米种植面积为研究对象,通过空间分析、“3S”技术与传统统计抽样方法相结合,分别设计 7 种抽样比水平、4 种样本空间布局方式和 8 种抽样单元尺度,对农作物种植面积空间抽样调查方案进行了优化设计试验研究,结果表明:

1) 抽样单元间空间自相关性随单元尺度的增大而增大,两者间呈线性正相关关系。当抽样单元尺度为 500 m×500 m 时, Moran's  $I$  值接近为 0,即在该尺度下,抽样单元间空间自相关性几乎不存在。遵循传统抽样理论要求样本间相互独立原则,选取 500 m×500 m 作为最优抽样单元尺度。

2) 通过对抽样单元内的玉米种植面积与耕地面积进行相关分析发现,两者间存在极显著的线性正相关关系。为提高农作物种植面积分层抽样效率,可选取抽样单元内的耕地面积作为分层标志。

3) 在抽样比相同条件下,以抽样外推总体相对误差和变异系数为抽样效率评价指标,在 4 种(简单随机、系统等距、分层随机、分层系统等距)样本空间布局方式中,选取分层系统等距作为最优样本布局方式。

4) 通过分析抽样外推总体相对误差和变异系数随抽样比变化曲线发现, 在 7 种 (0.5%、1.0%、1.5%、2.0%、2.5%、3.0%、3.5%) 抽样比设计水平下, 当抽样比为 1% 时的研究区玉米种植面积空间抽样效率最高, 因此, 选取 1% 作为最优抽样比。

#### [参 考 文 献]

- [1] 陈水森, 柳钦火, 陈良富, 等. 粮食作物播种面积遥感监测研究进展[J]. 农业工程学报, 2005, 21(6): 166—170.  
Chen Shuisen, Liu Qinhua, Chen Liangfu, et al. Review of research advances in remote sensing monitoring of grain crop area[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2005, 21(6): 166—170. (in Chinese with English abstract)
- [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.  
Qian Yonglan, Yang Bangjie, Jiao Xianfeng. Accuracy assessment on the crop area estimating method based on RS sampling at national scale[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2007, 23(11): 180—187. (in Chinese with English abstract)
- [3] 杨邦杰, 裴志远, 周清波, 等. 我国农情遥感监测关键技术研究进展[J], 农业工程学报, 2002, 18(3): 191—194.  
Yang Bangjie, Pei Zhiyuan, Zhou Qingbo, et al. Key technologies of crop monitoring using remote sensing at a national scale: Progress and problems[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2002, 18(3): 191—194. (in Chinese with English abstract)
- [4] 周清波. 国内外农情遥感现状与发展趋势[J]. 中国农业资源与区划, 2004, 25(5): 9—14.  
Zhou Qingbo. Status and tendency for development in remotesensing of agriculture situation[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2004, 25(5): 9—14. (in Chinese with English abstract)
- [5] Tsiligrades T A. Remote sensing as a tool for agricultural statistics: A case study of area frame sampling methodology in Hellas[J]. Computers and Electronics in Agriculture, 1998, 20(1): 45—77.
- [6] Delince J. A European approach to area frame survey[C]// Processing of the Conference on Agricultural and Environmental Statistical Applications in Rome(CAESAR), Rome: European Commission GI & GIS, 2001, 2: 463—472.
- [7] Pradhan S. Crop area estimation using GIS, remote sensing and area frame sampling[J]. International Journal of Applied Earth Observation, 2001, 3(1): 86—92.
- [8] 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): 3306—3315.  
Zhang Jinshui, Shen Kejian, Pan Yaozhong, et al. HJ-1 remotely sensed data and sampling method for wheat area estimation [J]. Scientia Agricultura Sinica, 2010, 43(16): 3306—3315. (in Chinese with English abstract)
- [9] 郭明权, 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.  
Wu Mingquan, Yang Liangchuang, Yu Bo, et al. Mapping crops acreages based on remote sensing and sampling investigation by multivariate probability proportional to size[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(2): 146—152. (in Chinese with English abstract)
- [10] 王迪, 周清波, 刘佳. 作物面积空间抽样研究进展[J]. 中国农业资源与区划, 2012, 33(2): 9—14.  
Wang Di, Zhou Qingbo, Liu Jia. A review on spatial sampling for crop area estimation[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2012, 33(2): 9—14. (in Chinese with English abstract)
- [11] Gallego F J. Crop area estimation in the MARS project[C]// Conference on ten years of the MARS Project, Brussels: European Commission GI & GIS, 1999: 1—11.
- [12] Carfagna E, Gallego F J. Using remote sensing for agricultural statistics [J]. International Statistical Review. 2005, 73(3): 389—404.
- [13] 申克建, 何浩, 蒙红卫, 等. 农作物面积空间抽样调查研究进展[J]. 中国农业资源与区划, 2012, 33(4): 11—16.  
Shen Kejian, He Hao, Meng Hongwei, et al. Review on spatial sampling survey in crop area estimation[J]. Chinese Journal of Agricultural Resources and Regional Planning, 2012, 33(4): 11—16. (in Chinese with English abstract)
- [14] 陈仲新, 刘海启, 周清波, 等. 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.  
Chen Zhongxin, Liu Haiqi, Zhou Qingbo, et al. Sampling and scaling scheme for monitoring the change of winter wheat acreage in China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2000, 16(5): 126—129. (in Chinese with English abstract)
- [15] 焦险峰, 杨邦杰. 全国棉花种植面积遥感检测抽样方法设计[J]. 农业工程学报, 2002, 18(4): 168—171.  
Jiao Xianfeng, Yang Bangjie. Design of sampling method for cotton area estimation using remote sensing at a national level[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2002, 18(4): 168—171. (in Chinese with English abstract)
- [16] 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.
- Jiao Xianfeng, Yang Bangjie, Pei Zhiyuan. Paddy rice area estimation using a stratified sampling method with remote sensing in China[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2006, 22(5): 105—110. (in Chinese with English abstract)
- [17] 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.
- Zhu Shuang, Zhang Jinshui. Provincial agricultural stratification method for crop area estimation by remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(2): 184—191. (in Chinese with English abstract)
- [18] 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.
- Wu Bingfang, Li Qiangzi. Crop acreage estimation using two individual sampling frameworks with stratification[J]. Journal of Remote Sensing, 2004, 8(6): 551—569. (in Chinese with English abstract)
- [19] Gallego F J, Bamps C. Using CORINE Land cover and the point survey LUCAS for area estimation[J]. International Journal of Applied Earth Observation and Geoinformation, 2008, 10(4): 467—475.
- [20] 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.
- Yang Xiaoqiong, Zhu Wenquan, Pan Yaozhong, et al. Spatial sampling design for crop acreage estimation[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2007, 23(12): 150—155. (in Chinese with English abstract)
- [21] 张锦水, 潘耀忠, 胡潭高, 等. 冬小麦种植面积空间抽样效率影响因子分析[J]. 农业工程学报, 2009, 25(8): 169—173.
- Zhang Jinshui, Pan Yaozhong, Hu Tangao, et al. Analysis of influence factors about space sampling efficiency of winter wheat planting area[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2009, 25(8): 169—173. (in Chinese with English abstract)
- [22] 王迪, 周清波, 陈仲新, 等. 空间抽样方法估算冬小麦播种面积[J]. 农业工程学报, 2012, 28(10): 177—184.
- Wang Di, Zhou Qingbo, Chen Zhongxin, et al. Spatial sampling method for estimating winter wheat sown area[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(10): 177—184. (in Chinese with English abstract)
- [23] 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): 1232—1242.
- Wang Di, Zhou Qingbo, Chen Zhongxin, et al. Optimizing the size of spatial sampling units for estimating winter wheat sown acreage[J]. Journal of Natural Resources, 2013, 28(7): 1232—1242. (in Chinese with English abstract)
- [24] 曹志冬, 王劲峰, 李连发, 等. 地理空间中不同分层抽样方式的效率与优化策略[J]. 地理科学进展, 2008, 27(3): 152—160.
- Cao Zhidong, Wang Jinfeng, Li Lianfa, et al. Strata efficiency and optimization strategy of stratified sampling on spatial population[J]. Progress in Geography, 2008, 27(3): 152—160. (in Chinese with English abstract)
- [25] 姜成晟, 王劲峰, 曹志冬. 地理空间抽样理论研究综述[J]. 地理学报, 2009, 64(3): 368—380.
- Jiang Chengsheng, Wang Jinfeng, Cao Zhidong. A review of geo-spatial sampling theory[J]. Acta Geographica Sinica, 2009, 64(3): 368—380. (in Chinese with English abstract)
- [26] 王远飞, 何洪林. 空间数据分析方法[M]. 北京: 科学出版社, 2007: 110—112.
- [27] 胡潭高, 张锦水, 潘耀忠, 等. 基于不同抽样方法的遥感面积测量方法研究[J]. 国土资源遥感, 2008, 19(3): 37—41.
- Hu Tangao, Zhang Jinshui, Pan Yaozhong, et al. Researches on remote sensing area measurement based on different sampling methods[J]. Remote Sensing for Land & Resources, 2008, 19(3): 37—41. (in Chinese with English abstract)
- [28] 谭建光, 张锦水, 高晨雪, 等. 基于结构规模的冬小麦种植面积遥感抽样估算[J]. 农业工程学报, 2012, 28(23): 114—122.
- Tan Janguang, Zhang Jinshui, Gao Chenxue, et al. Winter wheat area estimation based on structure and scale using remote sensing[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2012, 28(23): 114—122. (in Chinese with English abstract)
- [29] 李金昌. 应用抽样技术[M]. 北京: 科学出版社, 2006: 45—51.
- [30] 杜子芳. 抽样技术及其应用[M]. 北京: 清华大学出版社, 2005: 126—132.
- [31] 李明阳, 姜文倩, 徐婷, 等. 基于总体表面属性特征的森林资源抽样调查方法比较[J]. 东北林业大学学报, 2011, 39(9): 49—52.
- [32] 国家统计局农调总队. 农村统计调查业务工作手册[Z]. 北京: 中国统计出版社, 2002: 46—51.



## Optimization of spatial sampling schemes for maize planting acreage estimation

Wang Di<sup>1,2</sup>, Zhou Qingbo<sup>1,2\*</sup>, Chen Zhongxin<sup>1,2</sup>, Liu Jia<sup>1,2</sup>

(1. Key Laboratory of Agri-informatics, Ministry of Agriculture, Beijing 100081, China;

2. Institute of Agriculture Resources and Regional Planning, Chinese Academy of Agriculture Sciences, Beijing 100081, China)

**Abstract:** Sampling fraction, sample layout, and sampling unit scale are three basic elements of a spatial sampling scheme. It plays an important role in optimizing these factors for decreasing the sampling cost and improving the extrapolation accuracy of survey sampling. In this study, spatial analysis, “3S” techniques, and traditional sampling methods were employed to optimize the three basic elements, aiming at the problem that the spatial sampling efficiency is still poor due to only one basic element (e.g. the sampling unit scale) being optimized in the existing spatial sampling studies for crop acreage estimation. DeHui County in Jilin Province was chosen as the study area, maize planting acreage as the study object, and square grids as the shape of the sampling units. First, the sampling unit scale, sampling fraction, and sample layout were formulated based on the Second National land survey data, and the spatial distribution data of maize in the study area in 2009 (derived from SPOT image that the spatial resolution is 10 m). In order to analyze the relationship between the scale and spatial correlation of sampling units, the sampling unit scales were designed to be 8 levels, that is 500 m × 500 m, 1 000 m × 1 000 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, 3 500 m × 3 500 m, and 4 000 m × 4 000 m, respectively, the sampling fractions were designed to be 7 levels, that is 0.5%, 1.0%, 1.5%, 2.0%, 2.5%, 3.0%, and 3.5%, and four patterns were selected as the samples layout, they are simple random sampling, system isometric sampling, stratified random sampling, and stratified system isometric sampling respectively to conduct the optimal design of the sampling fractions and samples layouts. Secondly, the optimal sampling unit scales were determined by introducing the global spatial autocorrelation index Moran’s *I*, following the traditional sampling principle that sampling units should be independent of each other; Thirdly, the samples were drawn, and the sample observations measured, population values were extrapolated, and the sampling errors were estimated using the designed spatial sampling scheme. Finally, the relative error of population extrapolation, coefficient of variation (CV) of the population total estimator, and sampling sizes were selected as the evaluation indices to find out the optimal sampling fraction and sample layout. The experimental results demonstrated that the spatial autocorrelation of sampling unit increases with its scales, and that there was a linear positive relationship between the sampling unit scale and spatial autocorrelation. While the sampling units scale was 500 m × 500 m, there was nearly no spatial autocorrelation among sampling units. Therefore, 500 m × 500 m was selected as the optimal sampling unit scale. In order to improve the stratified sampling efficiency, cropland area may be considered as a stratification index due to there being a very significant linear correlation between maize planting acreage and cropland area in sampling units; When the sampling fraction was the same level, the relative error and CV of population extrapolation which the sample layout was the stratified system isometric sampling were the lowest in four layout patterns; The efficiency of spatial sampling scheme in which the sampling fraction was 1% was the highest in the 7 designed levels. In this way, this research can provide a solution for improving the efficiency of a spatial sampling scheme to estimate crop acreage.

**Key words:** crops; extrapolation; error analysis; spatial sampling; maize; planting acreage; sampling unit scale; samples layout

(责任编辑：张俊芳)