

doi: 10.7621/cjarrp.1005-9121.20190903

## · 技术方法 ·

# 农作物面积抽样单元空间变异性及其对抽样外推效率的影响\*

仲格吉, 周清波, 王迪<sup>\*</sup>

(中国农业科学院农业资源与农业区划研究所/农业农村部农业遥感重点实验室, 北京 100081)

**摘要** [目的] 联合遥感和传统抽样方法的空间抽样技术是进行大区域农作物面积监测的有效手段。但传统抽样要求抽样单元间相互独立, 并未考虑到区域农作物空间变异性存在对农作物面积空间抽样效率是否有影响及其影响程度如何, 从而限制了农作物面积监测的效率和精度。[方法] 文章以吉林省德惠市为研究区, 利用变异函数分别对10种抽样单元尺度内玉米和水稻种植面积的空间变异性进行定量评价。选择3种空间抽样方案(简单随机抽样、系统抽样、分层抽样), 分别计算不同空间变异程度下的样本容量( $n$ )、抽样相对误差( $Re$ )和总体总值估计量的变异系数( $CV(\hat{Y})$ )作为空间抽样效率的评价指标, 定量分析空间变异性对农作物面积空间抽样效率的影响, 提出适宜空间变异农作物的空间抽样优化方案。[结果] (1) 玉米和水稻面积的空间变异性(反映为基台值 $C_0 + C$ )均随着抽样单元尺度的增加逐渐而减小, 其中 $500m \times 500m$ 和 $2500m \times 2500m$ 抽样单元尺度下农作物面积的空间自相关程度相对最小。(2) 相同抽样比下, 农作物面积在3种抽样方法下的相对误差和变异系数均随着空间变异性减小(抽样单元尺度增加)逐渐增大, 其中玉米和水稻在分层抽样条件下的相对误差和变异系数具有相对最小值, 分别限制在10%和20%以内, 其次为系统抽样( $Re < 40\%$ ,  $CV(\hat{Y}) < 88\%$ )和简单随机抽样( $Re < 50\%$ ,  $CV(\hat{Y}) < 75\%$ )。在抽样方法被选定的情况下, 两种农作物的分层抽样相对误差变化率随着抽样比的增加迅速降低, 当抽样比增至5%时, 其相对误差和变异系数的变化率基本趋于稳定, 继续增加抽样比来提高抽样精度的效果甚微。在 $3500m \times 3500m$ 抽样单元尺度内, 按5%抽样比确定的样本容量对玉米和水稻面积进行分层抽样时, 可达到95%的抽样精度( $Re < 5\%$ )。综合农作物在不同抽样单元尺度下的空间变异特征和空间抽样效率, 基于 $2500m \times 2500m$ 的抽样单元尺度和5%抽样比进行分层抽样的空间变异性农作物空间抽样方案是一种高精度、低成本、代表性强的空间抽样方案。[结论] 该研究为区域农作物面积抽样调查方法的改善和抽样效率的提高提供技术支持和参考依据。

**关键词** 农作物面积 抽样单元尺度 空间变异性 空间抽样 抽样效率

**中图分类号:**S-3 **文献标识码:**A **文章编号:**1005-9121[2019]09020-17

## 0 引言

农作物面积是各级管理部门进行农业生产指导、国家制定粮食政策的重要依据<sup>[1]</sup>, 及时准确地了解农作物种植面积, 对于准确估计和预测农作物产量, 加强农作物生产管理<sup>[2-5]</sup>, 确保我国粮食安全具有重要意义<sup>[6,7]</sup>。抽样技术因具有低成本、高质量、高效率的优越性而被广泛应用于社会、经济、科技、自然等调查领域。随着“3S”技术的发展, 结合了遥感与传统抽样方法优势的空间抽样技术逐渐被应用到大区域农作物面积监测当中<sup>[8-12]</sup>。如美国的大面积农作物估产计划(Large Area Crop Inventory Experiment, LACIE)和欧盟的农业遥感监测计划(Monitoring Agriculture with Remote Sensing, MARS), 利用“3S”技

收稿日期: 2018-12-06

作者简介: 仲格吉(1991—), 女, 甘肃甘南人, 硕士研究生。研究方向: 农作物面积空间抽样

\*通讯作者: 王迪(1977—), 男, 辽宁鞍山人, 副研究员。研究方向: 农作物面积空间抽样理论与技术。Email: wangdicaas@126.com

\*资助项目: 国家自然科学基金重点项目“基于“三位一体”空间抽样理论研究及其二联查找表研建”(41531179)

术与分层抽样相结合的空间抽样调查方法有效提高了大区域农作物面积信息的获取精度和时效性<sup>[13-15]</sup>。

近年来,国内外学者利用空间抽样方法在农作物面积估计领域相继开展了研究<sup>[16-21]</sup>,选用的抽样方法包括整群抽样、分层抽样和多阶段抽样,涉及到的农作物类型有小麦、玉米、水稻及棉花等。但已有的这些研究所使用的传统抽样方法均建立在抽样单元相互独立的基础上<sup>[22-24]</sup>,较少针对样本点的空间分布特征开展特定的抽样调查,农作物由于受自然条件(气候、土壤类型、地形地貌等)和社会经济因素等影响,在一定尺度上存在不同程度的空间变异,这种空间变异性会导致样本容量和抽样方差的过高估计<sup>[25, 26]</sup>,从而限制空间抽样方案的优化设计。变异函数作为一种解释空间数据的方法,除了考虑样本单元的数字特征之外,更注重于研究区域化变量的空间分布特征,能够同时描述区域化变量的结构性变异和随机性变异<sup>[27-29]</sup>,是空间变异分析的有效工具<sup>[30]</sup>。利用变异函数计算区域农作物种植面积的空间变异性,将对空间变异性的考量纳入空间抽样单元尺度的优选和抽样效率的评估当中,可进一步改善农作物空间抽样调查方法在不同地区的适用性,从而提高空间抽样效率<sup>[47, 49]</sup>。

目前国内外关于抽样单元空间变异性研究主要集中于生态资源调查<sup>[31, 32]</sup>、森林群落结构<sup>[33]</sup>以及土地利用<sup>[34, 35]</sup>等领域,对农作物的空间抽样调查研究侧重于遥感与抽样相结合的方案设计<sup>[36-40]</sup>方面和传统抽样中各抽样要素对抽样效率的定量化影响方面<sup>[1-43]</sup>,关于农作物种植面积调查单元在空间分布上的空间变异性对空间抽样外推总体的效率(抽样精度和抽样费用)是否有影响、影响程度及如何影响的研究则较少被关注,尚不能科学解释当抽样单元存在空间变异性时,空间抽样调查效率的变化特征及规律,从而影响了空间抽样方法在国家尺度范围内的进一步推广与应用。因此,文章选取吉林省德惠市为研究区,通过设置10种抽样单元尺度,利用变异函数定量分析玉米和水稻空间变异性尺度变化特征,采用3种空间抽样方法分别对农作物面积进行总体估计,以样本容量( $n$ )、相对误差( $Re$ )和外推总体变异系数( $CV(\hat{Y})$ )作为空间抽样效率的评价指标,分析不同抽样单元尺度下农作物面积空间变异性对空间抽样效率的影响,通过抽样单元尺度、抽样方法和样本容量的优选,提出适用于空间变异农作物的空间抽样方案。

## 1 研究区域与研究方法

### 1.1 研究区域

德惠市位于吉林省中北部,松辽平原中部,地处 $125^{\circ}14'E \sim 126.28'E$ , $44^{\circ}32'N - 45^{\circ}51'N$ 区间,全市总面积 $3\,435\text{km}^2$ ,粮食播种面积 $20.762\,0\text{万 hm}^2$ 。气候属于中温带大陆季风气候,年平均气温 $4.4^{\circ}\text{C}$ ,年均日照时数为 $2\,695.2\text{h}$ ,年降水量为 $520\text{mm}$ 。受夏季风影响,全年降水量集中在夏季,雨热同季有效降水多,能够满足一年一熟农作物生产的需要,主要农作物为玉米、水稻,玉米播种面积 $15.746\,1\text{万 hm}^2$ ,水稻播种面积 $4.597\,4\text{万 hm}^2$ 。

### 1.2 数据来源

基础数据主要包括:(1)德惠市基础地理信息数据。2005年后德惠市行政边界数据(比例尺 $1:10\,000$ ,县界,矢量格式);(2)农作物空间分布数据。2007年德惠市玉米、水稻空间分布数据(源自SPOT5影像提取,获取日期:2007年8月1日,空间分辨率10m),图1表明了2007年德惠市玉米和水稻空间分布情况。

### 1.3 数据处理

#### 1.3.1 异常值处理

异常值的存在会造成连续表面的中断,直接影响变量的分布特征,掩盖变量固有的空间结构特征,因此该研究采用域值法识别处理异常值,即样本平均值 $\bar{A}$ 加减3倍标准差 $s$ ,在此区间 $\bar{A} \pm 3s$ 以外的数据均定为异常值,然后分别用最大和最小值代替异常值<sup>[44, 45]</sup>。

#### 1.3.2 正态分布检验

地统计学通常要求原始数据呈正态分布,否则会产生比例效应<sup>[46]</sup>。比例效应的存在会使实验变异函

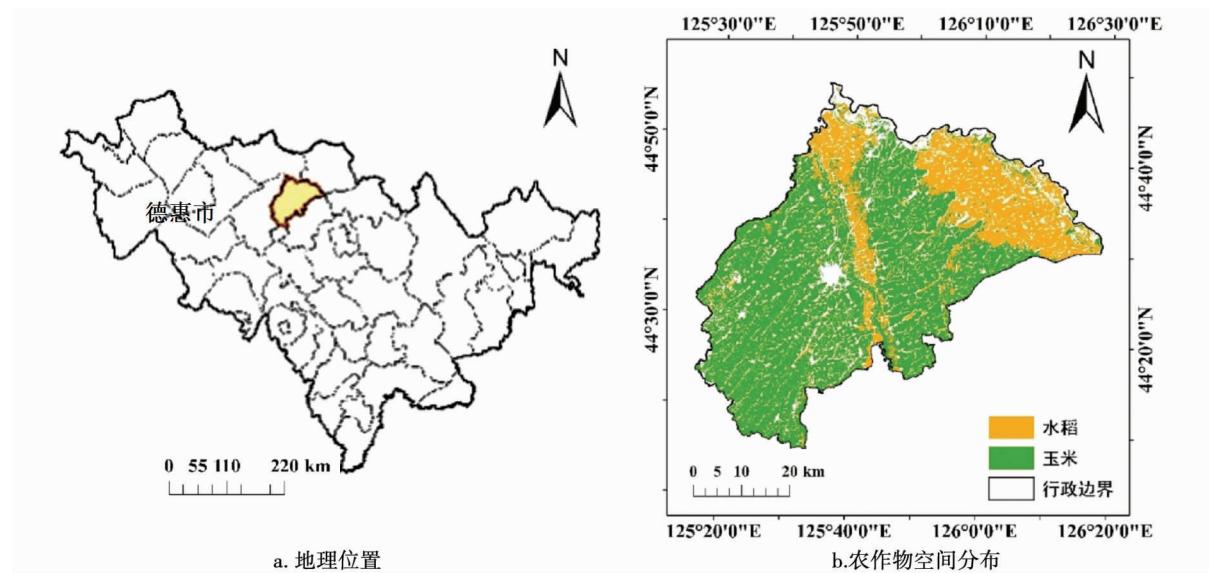


图 1 2007 年德惠市玉米和水稻空间分布

数值产生畸变，致使基台值和块金值增大，降低估计精度，导致某些结构特征不明显。因此在对数据进行地统计分析时，首先应对其正态分布性进行检验。

该文利用 SPSS20 和 Matlab2016 软件计算区域化变量的偏度系数 (*skewness*)、峰度系数 (*kurtosis*) 等参数，并以此检验数据是否服从正态分布（该文以 *skewness* = 0 作为正态分布基准）。

#### 1.4 空间变异性特征确定方法

地统计学方法是基于区域化变量理论基础的一种空间分析方法，主要以变异函数作为工具来表达区域化变量随着空间位置和方向不同所产生的差异。关于其基本原理和方法，很多文献已经有很好的解析和综述<sup>[46-49]</sup>，该文作以简要说明。当区域化变量满足二阶平稳假设和本征假设时，变异函数可表示为：

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

式 (1) 中， $\gamma(h)$  半方差函数； $h$  为分隔两样点的矢量，称为步长； $N(h)$  为相距为  $h$  的样点对数目； $Z(x_i)$  和  $Z(x_i + h)$  分别为区域化变量  $Z(x)$  在位置  $x_i$  和  $x_i + h$  处的实测值。

该文中关于实验变异函数的计算、理论模型的拟合及变异函数图的绘制分别由软件 GS + 9.0 和 Excel2010 共同完成。

#### 1.5 空间抽样方法

(1) 抽样比设计。参考以往相关研究中对抽样比的设定结果，为了便于对比不同抽样方法的抽样外推精度和稳定性，该研究设计了 4 种抽样比 ( $f=1\%, 5\%, 10\%, 20\%$ ) 来确定不同空间抽样下的样本容量 ( $n$ )。

(2) 抽样方案。该文选用了 3 种常用的抽样方法：简单随机抽样、系统抽样和分层抽样来进行对比分析。以下 3 种抽样方法的实现均通过 Matlab2016 软件编程完成。

简单随机抽样是指从一个单元数为  $N$  的总体中无放回地逐个抽取单元，每次都在所有尚未进入样本的单元中等概率的抽取，直到  $n$  各单元被抽中。该文利用 ArcGIS10.2 软件对各抽样尺度下的全部总体单元按照由  $1 \sim N$  顺序进行数字编码（不包括农作物面积为 0 的抽样单元），当总体单元的数字编码与软件生成的伪随机数相同时，该单元即被抽中，同时该样本单元的位置也被确定。

系统等距抽样是指对研究总体按一定的顺序排列，每隔一定的间隔抽取一个单元的抽样方法。①样本单元布设过程同简单随机抽样；②抽样间隔  $k$  确定。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  $k$  个基础单元; ③从第一段的  $k$  个总体单元中随机抽出一个单元作为起点; ④对于其中的  $n-1$  段, 每隔  $k$  个单元抽出一个样本单元, 以此类推, 直至抽出  $n$  个样本单元。特别地, 当  $N$  不是  $n$  的整数倍时, 总是会出现样本容量不固定或者一些抽样单元被排除在间隔  $K$  外永远不会被抽中的情况, 为避免这一问题, 选择使用圆圈系统抽样方法对总体单元进行编号<sup>[50]</sup>。

分层随机抽样先按照某种规则把总体划分为不同的层, 然后在层内再进行随机抽样, 各层的抽样之间是独立进行的。①样本单元布设同简单随机抽样; ②分层标志的设计。为提高分层抽样效率, 分别提取抽样单元内玉米和水稻的种植面积比作为分层标志; ③层数设计。在兼顾层内方差尽量小原则和抽样调查成本的前提下, 该研究选择分层数为 5 层; ④样本容量分配。在已知抽样比的条件下, 采用比例分配法在各层内进行再分配。⑤样本抽选。各层内采用简单随机方式抽选样本单元。

## 1.6 抽样单元观测值获取

根据德惠市玉米和水稻的空间分布数据, 利用 ArcGIS10.2 软件, 建立  $500\text{m} \times 500\text{m}$ 、 $1\ 000\text{m} \times 1\ 000\text{m}$ 、 $1\ 500\text{m} \times 1\ 500\text{m}$ 、 $2\ 000\text{m} \times 2\ 000\text{m}$ 、 $\cdots$ 、 $5\ 000\text{m} \times 5\ 000\text{m}$  共 10 种尺度的正方形网格, 分别叠加玉米和水稻空间分布数据作为抽样框, 利用 dissolve 工具计算出每一抽样单元内的玉米和水稻种植面积。以每一抽样单元中心点作为采样点、抽样单元内农作物种植面积比作为区域化变量来描述研究区内农作物的空间变异特征, 以抽样单元内农作物种植面积比作为基础单元观测值进行抽样总体外推与误差估计。

## 1.7 抽样效率评价

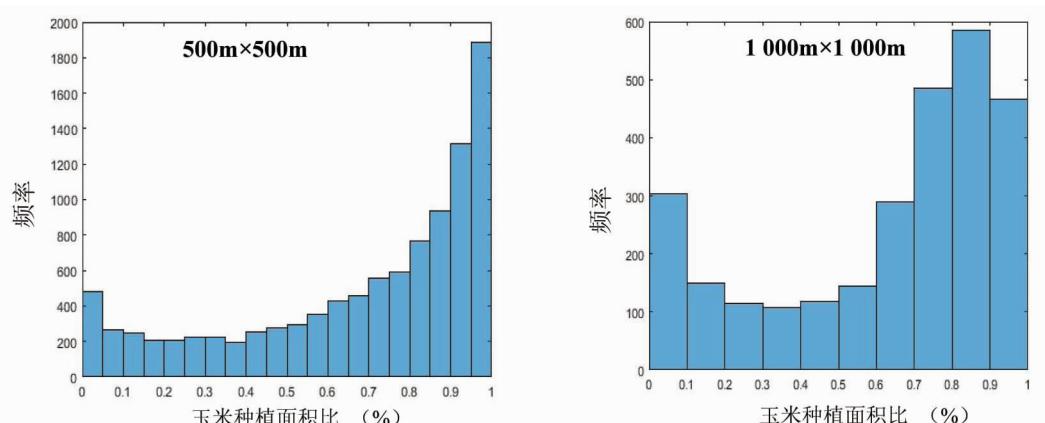
抽样效率指的是以最小费用达到抽样调查要求的估计精度<sup>[52]</sup>, 它包含抽样调查的两个重要方面: 费用和精度, 费用由样本容量 ( $n$ ) 反映, 抽样精度由抽样相对误差 ( $Re$ ) 反映。该文在抽样精度基础上, 加入总体总值估计量的变异系数  $CV(\hat{Y})$  作为抽样总体推断稳定性的评价指标, 结合  $n$  和  $Re$  共同作为空间抽样效率的评价指标。该文侧重点在于不同抽样方案的抽样效率比较, 因此采用简单估计量进行空间抽样的总体外推与误差估计, 估计量的公式计算参照文献<sup>[52,53]</sup>, 此处不再赘述。

# 2 结果分析

## 2.1 抽样单元内农作物面积比描述性统计分析与正态分布检验

在传统抽样条件下, 异常值一般是指数值明显偏离其所属的其余样本观测值的个别观测值。该文在利用域值法检验数据时, 所有尺度下的玉米和水稻种植面积比例均处于  $\bar{A} + 3s$  区间内, 说明玉米和水稻面积比例数据不存在异常值, 可直接进行正态检验。

在进行正态分布检验之前, 为了更直观地了解数据的分布情况, 该文以玉米为例在图 2 中展示了抽样单元内玉米种植面积比在 10 种抽样单元尺度下的直方图变化情况。可以看出, 各抽样单元尺度下的直方图均呈现出倾斜的分布态势, 数据不服从正态分布。进一步采用偏度系数法确定数据的不对称程度及方向, 从表 1 和表 3 可以看出, 不同尺度 ( $scale$ ) 下的玉米种植面积比具有较强的左偏效应 ( $skewness < 0$ )



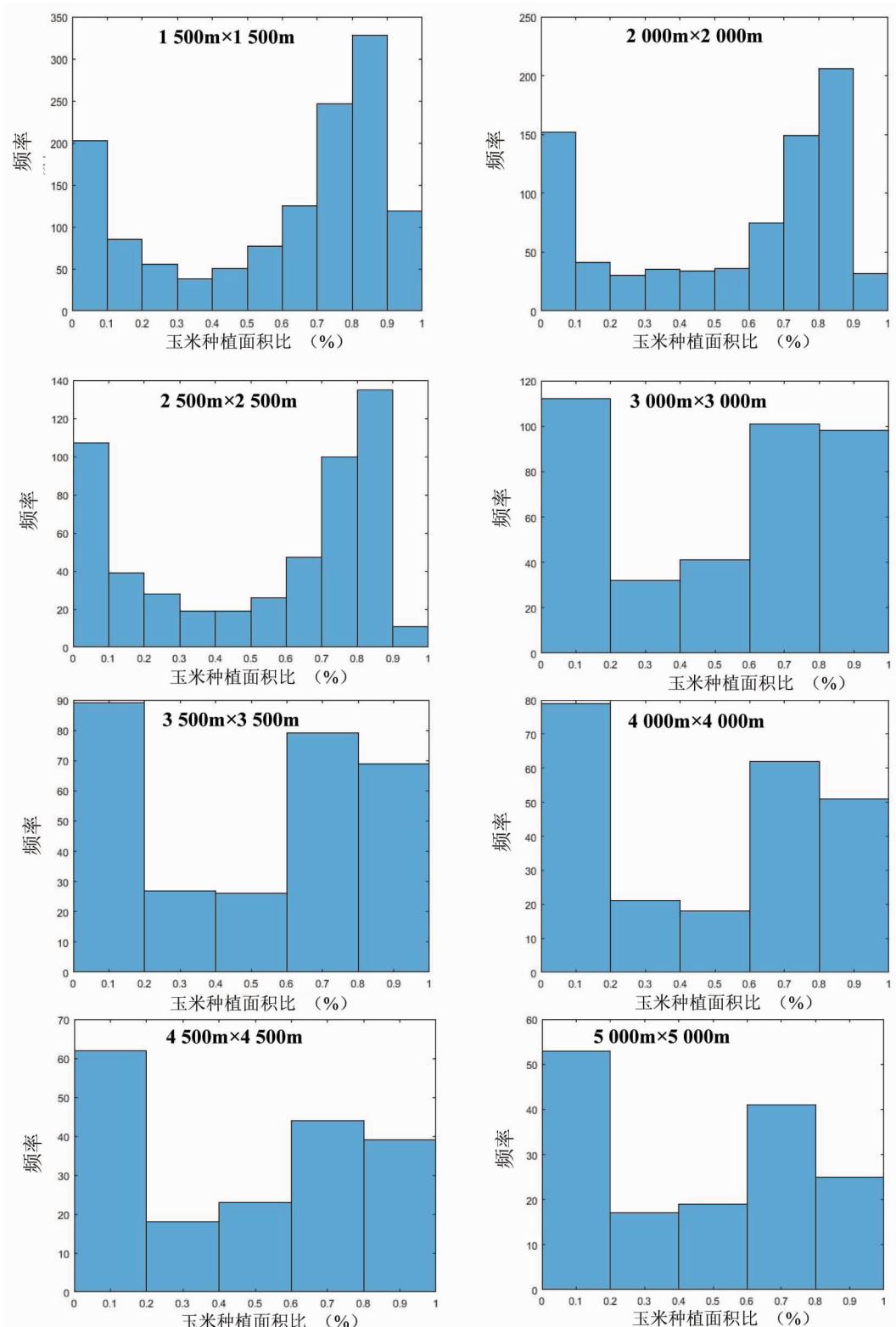


图 2 抽样单元内玉米种植面积比例的直方图

水稻种植面积比表现出较强的右偏效应 ( $skewness > 0$ )，这种偏态分布将导致地统计分析过程中计算结果的偏差。因此，在变异函数分析之前必须对其进行正态转换。通过表 2 和表 4，发现经过反正弦转换和平

方根转换过后的玉米和水稻种植面积比的偏度系数明显降低, 基本符合正态分布。

变异系数反映的是变量的相对变异程度, 当 $/CV/ < 10\%$ 时为弱变异性,  $/CV/ > 100\%$ 为强变异性,  $10\% < /CV/ < 10\%$ 为中等变异性。从表1和表3可以看出, 玉米和水稻种植面积比的变异系数变化范围分别为 $45\% \sim 77\%$ 、 $65\% \sim 135\%$ , 表明研究区内玉米和水稻的变异程度分别处于中等变异和强变异水平, 这是由于农作物的种植区域及该区域内土地利用状况、地形等状况不同导致的; 10种尺度下玉米和水稻种植面积比的变化范围(最小值min, 最大值max)均处于 $0\% \sim 100\%$ 之间, 说明在同一抽样单元尺度下农作物种植面积比的数值差异较大, 平均值分别在 $44\% \sim 85\%$ ,  $20\% \sim 55\%$ 之间, 说明德惠市内玉米的种植相对较为广泛, 水稻的种植则相对较少。

对于农作物种植面积比例(玉米和水稻)的常规统计分析只在一定程度上反映样本全体, 而不能定量地刻画农作物种植面积比例的随机性和结构性<sup>[51]</sup>, 因此该研究采用地统计学方法进一步对玉米和水稻种植面积比例空间变异性进行分析。

表1 玉米种植面积比常规统计特征值

| scale (m) | N      | skewness | kurtosis | min (%) | max (%) | $\bar{A}$ (%) | s       | CV (%)  |
|-----------|--------|----------|----------|---------|---------|---------------|---------|---------|
| 500       | 10 171 | -0.881   | -0.458   | 0       | 1       | 0.679 6       | 0.300 3 | 0.441 9 |
| 1 000     | 2793   | -0.816   | -0.727   | 0       | 1       | 0.618 7       | 0.308 6 | 0.498 8 |
| 1 500     | 1328   | -0.703   | -1.034   | 0       | 0.986 6 | 0.578 3       | 0.317 6 | 0.549 3 |
| 2 000     | 790    | -0.587   | -1.261   | 0       | 0.960 3 | 0.546 8       | 0.324 1 | 0.592 7 |
| 2 500     | 531    | -0.459   | -1.438   | 0       | 0.938 5 | 0.520 7       | 0.328 5 | 0.630 9 |
| 3 000     | 384    | -0.292   | -1.752   | 0       | 1       | 0.569 7       | 0.442 5 | 0.776 6 |
| 3 500     | 290    | -0.315   | -1.591   | 0.000 8 | 0.907 7 | 0.486 4       | 0.333 3 | 0.685 2 |
| 4 000     | 231    | -0.223   | -1.661   | 0.000 2 | 0.898 1 | 0.467 5       | 0.333 9 | 0.714 1 |
| 4 500     | 186    | -0.177   | -1.628   | 0.000 2 | 0.907 7 | 0.458 8       | 0.328 8 | 0.716 7 |
| 5 000     | 155    | -0.122   | -1.651   | 0.000 1 | 0.898 9 | 0.445 9       | 0.324 8 | 0.728 3 |

表2 玉米种植面积比正态转换后的统计特征值

| Scale (m) | N      | NT    | skewness | kurtosis | min (%) | max (%) | $\bar{A}$ (%) | CV (%)  |
|-----------|--------|-------|----------|----------|---------|---------|---------------|---------|
| 500       | 10 171 | Arsin | -0.276   | -0.893   | 0       | 1.570 8 | 0.844 1       | 0.522 8 |
| 1 000     | 2 793  | Arsin | -0.430   | -0.905   | 0       | 1.570 8 | 0.731 2       | 0.549 6 |
| 1 500     | 1 328  | Arsin | -0.456   | -1.143   | 0       | 1.406 9 | 0.671 3       | 0.587 9 |
| 2 000     | 790    | Arsin | -0.411   | -1.338   | 0       | 1.288 1 | 0.628 7       | 0.625 9 |
| 2 500     | 531    | Arsin | -0.309   | -1.477   | 0       | 1.218 1 | 0.595 6       | 0.663 6 |
| 3 000     | 384    | Arsin | -0.247   | -1.527   | 0.000 1 | 1.153 4 | 0.569 5       | 0.694 4 |
| 3 500     | 290    | Arsin | -0.187   | -1.593   | 0.000 8 | 1.137 8 | 0.552 9       | 0.716 7 |
| 4 000     | 231    | Arsin | -0.104   | -1.642   | 0.000 2 | 1.115 3 | 0.529 6       | 0.746 1 |
| 4 500     | 186    | Arsin | -0.049   | -1.612   | 0.000 2 | 1.137 7 | 0.517 8       | 0.751 1 |
| 5 000     | 155    | Arsin | 0.005    | -1.61    | 0.000 1 | 1.117 2 | 0.501 0       | 0.763 2 |

## 2.2 抽样单元尺度对空间变异性的影响

基于不同抽样单元尺度内玉米和水稻种植面积比剔除异常值并作正态转换后的数据, 分别建立理论变异函数模型、计算实际变异函数值 $\gamma(h)$ 并绘制变异函数曲线图, 由于不同的变异函数模型对于真实空间变异结构的拟合精度存在差异, 该文根据残差平方和(RSS)尽可能接近于0; 决定系数( $R^2$ )尽可能接近于1的标准来选择具有最佳拟合效果的农作物种植面积空间变异结构的变异函数模型。为了从整体上了解研究区内农作物种植面积的空间变异特征, 该研究利用玉米和水稻种植面积比的各向同性变异函数进行分析研究。利用GS+9.0对玉米和水稻种植面积比进行空间变异模型的拟合, 得到不同抽样单元尺度

表 3 水稻种植面积比常规统计特征值

| scale (m) | N     | skewness | kurtosis | min (%) | max (%) | $\bar{A}$ (%) | s       | CV (%) |
|-----------|-------|----------|----------|---------|---------|---------------|---------|--------|
| 500       | 4 825 | -0.16    | -1.478   | 0       | 1       | 0.539 2       | 0.353 9 | 65.62  |
| 1 000     | 1 480 | 0.143    | -1.565   | 0       | 1       | 0.439 5       | 0.346 3 | 78.81  |
| 1 500     | 770   | 0.375    | -1.465   | 0       | 9.885   | 0.375 4       | 0.339 2 | 90.34  |
| 2 000     | 494   | 0.591    | -1.232   | 0       | 0.964 9 | 0.329 2       | 0.326 9 | 99.31  |
| 2 500     | 345   | 0.707    | -1.09    | 0       | 0.937 9 | 0.301 7       | 0.319 6 | 105.96 |
| 3 000     | 262   | 0.85     | -0.86    | 0.000 1 | 0.937 2 | 0.275 9       | 0.310 4 | 112.53 |
| 3 500     | 208   | 0.962    | -0.601   | 0.000 1 | 0.901 4 | 0.255 3       | 0.300 7 | 117.81 |
| 4 000     | 175   | 1.087    | -0.348   | 0       | 0.901 4 | 0.232 3       | 0.290 5 | 125.07 |
| 4 500     | 146   | 1.175    | -0.003   | 0.000 1 | 0.893 3 | 0.220 0       | 0.281 4 | 127.93 |
| 5 000     | 125   | 1.235    | 0.072    | 0.000 2 | 0.876 9 | 0.208 1       | 0.276 8 | 132.99 |

表 4 水稻种植面积比正态转换后的统计特征值

| scale (m) | N     | NT    | skewness | kurtosis | min (%) | max (%) | $\bar{A}$ (%) | CV (%) |
|-----------|-------|-------|----------|----------|---------|---------|---------------|--------|
| 500       | 4 825 | OR    | -0.160   | -1.478   | 0       | 1       | 0.539 2       | 65.62  |
| 1 000     | 1 480 | OR    | 0.143    | -1.565   | 0       | 1       | 0.439 5       | 78.81  |
| 1 500     | 770   | SQRT  | -0.033   | -1.514   | 0.004 7 | 0.994 2 | 0.522 2       | 61.40  |
| 2 000     | 494   | SQRT  | 0.153    | -1.470   | 0.003 2 | 0.982 3 | 0.478 2       | 66.38  |
| 2 500     | 345   | 2SQRT | -0.149   | -1.282   | 0.028 3 | 0.984 1 | 0.616 0       | 42.56  |
| 3 000     | 262   | 2SQRT | -0.006   | -1.305   | 0.091 5 | 0.983 9 | 0.597 6       | 43.02  |
| 3 500     | 208   | 2SQRT | 0.066    | -1.302   | 0.096 3 | 0.974 4 | 0.578 3       | 44.69  |
| 4 000     | 175   | 2SQRT | 0.104    | -1.230   | 0.056 2 | 0.974 4 | 0.551 3       | 48.04  |
| 4 500     | 146   | 2SQRT | 0.208    | -1.291   | 0.109 7 | 0.972 2 | 0.542 8       | 47.74  |
| 5 000     | 125   | 2SQRT | 0.316    | -1.235   | 0.121 1 | 0.967 7 | 0.532 6       | 47.82  |

注: NF 表示正太转换; OR 表示原始正态分布, SQRT 表示一次平方根正态分布, 2SQRT 表示两次平方根正态分布, scale 表示抽样单元尺度, N 表示抽样单元总体, skewness 和 kurtosis 分别表示偏度和峰度, max 和 min 分别表示最大值和最小值,  $\bar{A}$  表示标准差, CV 表示变异系数

下的变异函数、最优拟合的理论模型及其参数。

图 3 和图 4 分别为玉米和水稻种植面积比在各向同性条件下实际变异函数曲线图, 表 5 和表 6 为变异函数的理论模型及空间变异性的参数。各向同性变异函数图展示了很好的空间结构, 不同抽样单元尺度下变异函数模型拟合度基本都在 0.85 以上, 属于理想的模型拟合度, 说明选取的变异函数模型能够很好地体现其空间变异性。整体而言, 玉米和水稻的空间变异性表现出一定的相似性, 空间变异函数值在抽样单元尺度上表现出一定的下降趋势; 在变程范围内的空间变异性变化较强, 空间变异与距离 (Separation distance) 呈正相关, 当距离增加超过变程时, 农作物面积比在空间上的差异不明显, 此时空间自相关性可以忽略。

基台值  $C_0 + C$  表示区域化变量的最大变异<sup>[49, 54]</sup>,  $C_0 + C$  越大表明总的空间变异程度越高; 块金值  $C_0$  表示样点间距为零时的变异函数值, 代表随机变异的量, 表现在最小距离内由实验误差和灌溉、耕作措施等随机因素引起的变异; 块金值与基台值之比  $C_0/C_0 + C$  即块金系数, 该比值表示由随机因素引起的变异占系统内总变异的百分比, 当块金系数小于 25% 说明变量有强烈的空间相关性, 且空间变异以结构性变异为主; 25%~75% 时变量有中等空间相关性; 大于 75% 时变量空间相关性弱, 空间变异主要由随机因素引起<sup>[31, 46]</sup>。

就研究区整体而言, 抽样单元尺度的变化导致农作物面积空间结构发生变化, 直接影响着对其空间变异性的分析。通常小尺度上能够探测到更多的细节特征, 而大尺度容易掩盖小尺度上的空间变化, 即空间

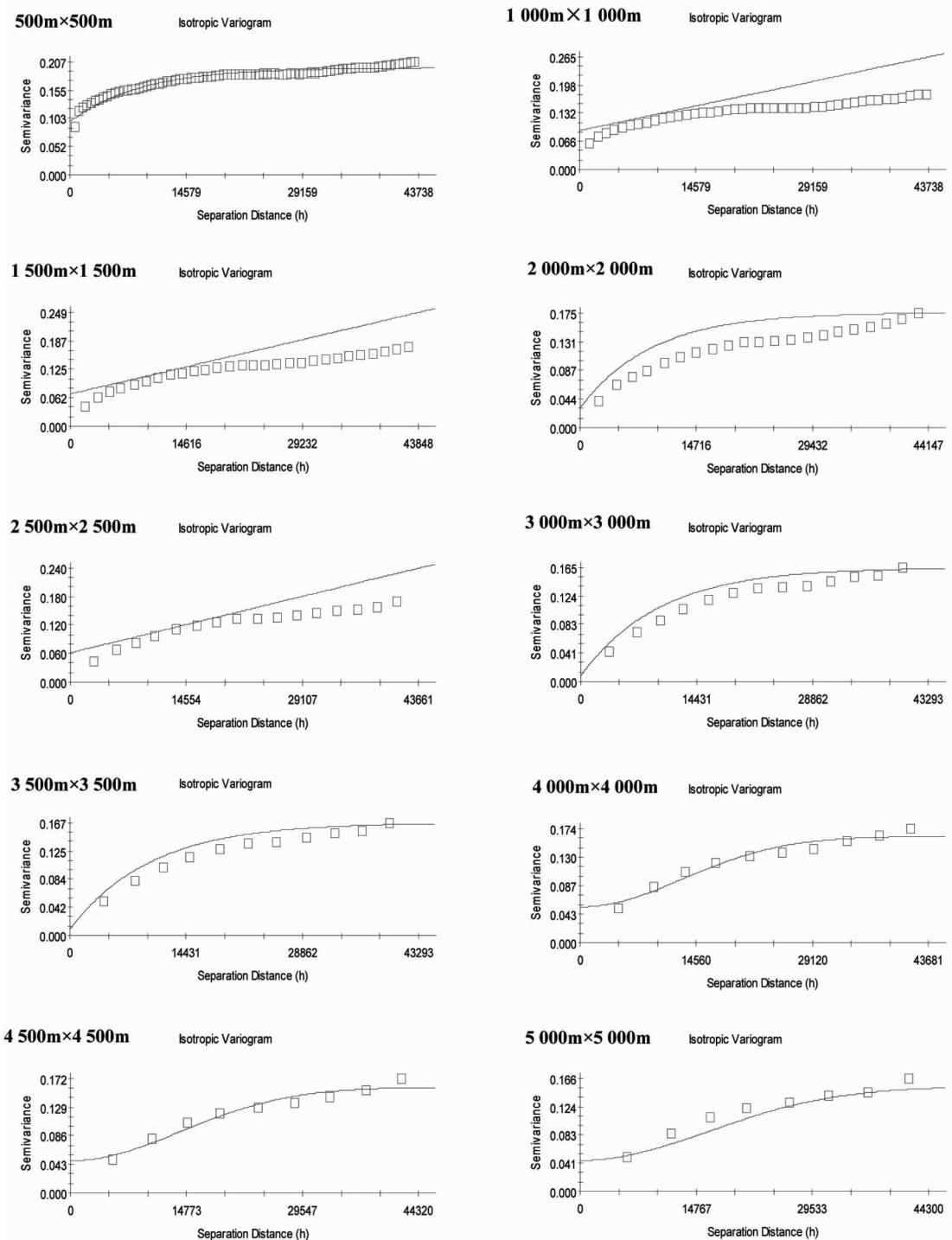


图3 不同抽样单元尺度下玉米种植面积比的变异函数曲线

注：Semivariance 表示变异函数值，Separation Distance 表示变程，下同

变异性会随着抽样单元尺度的增加而降低，但它们之间的相互影响关系难以界定<sup>[55,56]</sup>，因此适宜尺度的选取可以帮助探索作物面积的空间变化规律。

由表5和表6可以看出，不同网格尺度下玉米和水稻种植面积比在研究区范围内的总变异强度（基台值  $C_0 + C$ ）随着网格尺度的增加呈现逐渐减小的趋势，说明抽样单元尺度与研究区内作物的空间总变异程

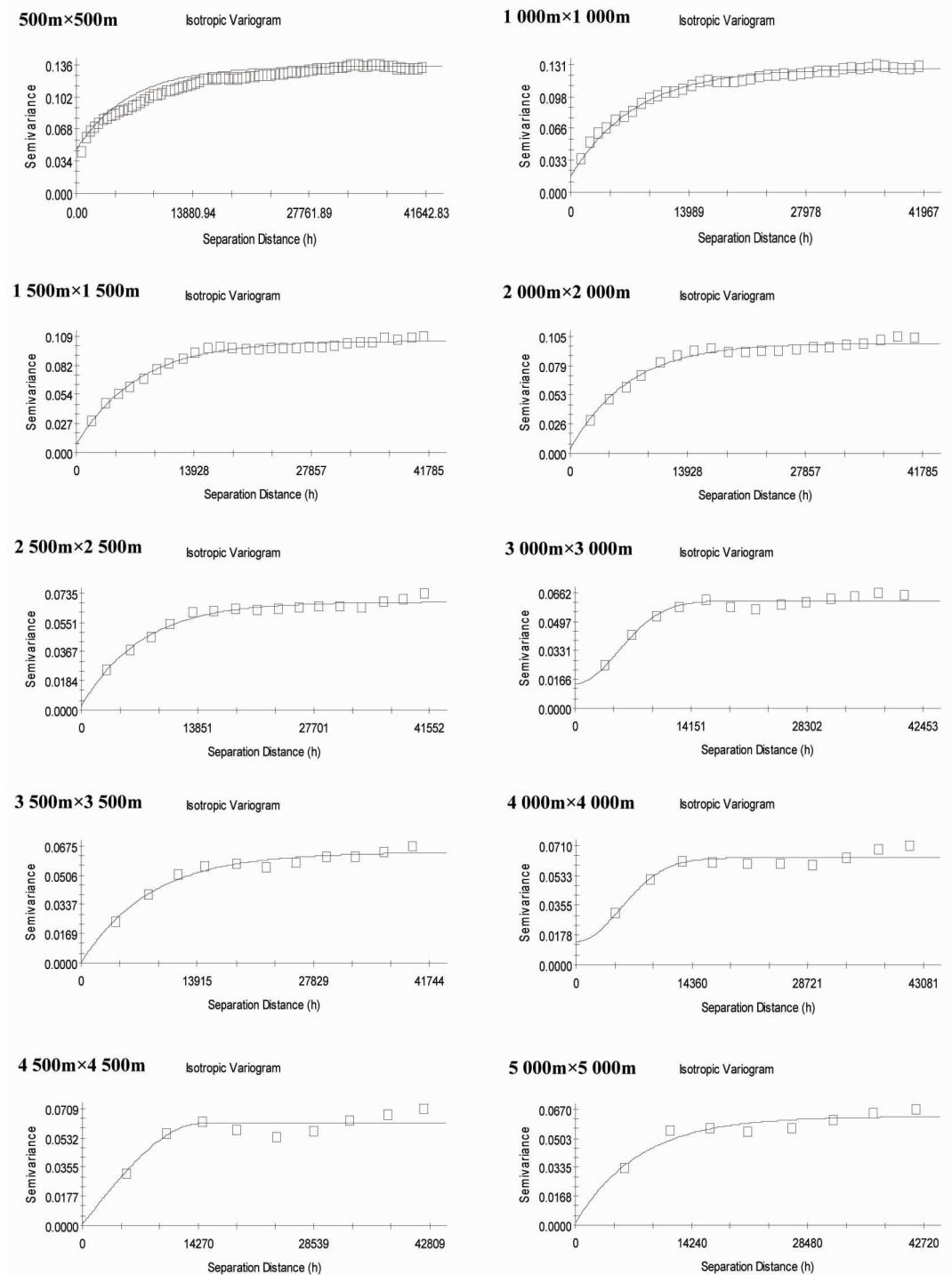


图 4 不同抽样单元尺度下水稻种植面积比的变异函数曲线

度为负相关关系。由表 5 中玉米的空间变异特征参数得出：基台值在 500m、1 000 ~ 2 500m、3 000 ~ 5 000m 幅度范围内的基台值表现出阶段递减趋势，每一阶段内变化相对平稳，即玉米面积主要表现出 3 种空间变异性程度，其中 500m 尺度下的基台值最大 (0.196 0)，1 000 ~ 2 500m 尺度范围内基台值集中在 0.17 左右，3 000 ~ 5 000m 范围内基台值集中在 0.15 左右；由表 6 中水稻的空间变异特征参数得出：基台值在 500 ~ 2 000m、2 500 ~ 5 000m 幅度范围内的基台值表现出阶段递减趋势，500 ~ 2 000m 尺度范围内基

表5 不同尺度下抽样单元内玉米种植面积比例的变异函数模型及其参数

| 网格尺度<br>scale (m) | 理论模型        | 块金值<br>$C_0$ | 基台值<br>$C_0 + C$ | 块金系数<br>$C_0 / (C_0 + C)$ | 变程<br>Range | 决定系数<br>$R^2$ | 残差<br>RSS |
|-------------------|-------------|--------------|------------------|---------------------------|-------------|---------------|-----------|
| 500               | Exponential | 0.097 5      | 0.196 0          | 0.497                     | 20 206      | 0.930         | 0.004 8   |
| 1 000             | Linear      | 0.091 7      | 0.179 2          | 0.312                     | 22 804      | 0.894         | 0.096 8   |
| 1 500             | Linear      | 0.061 7      | 0.175 6          | 0.353                     | 25 844      | 0.912         | 0.049 1   |
| 2 000             | Exponential | 0.029 9      | 0.175 8          | 0.170                     | 27 147      | 0.929         | 0.016 2   |
| 2 500             | Linear      | 0.070 5      | 0.175 0          | 0.402                     | 21 697      | 0.903         | 0.019 1   |
| 3 000             | Exponential | 0.007 6      | 0.165 2          | 0.046                     | 29 283      | 0.977         | 0.002 6   |
| 3 500             | Exponential | 0.009 4      | 0.167 8          | 0.056                     | 30 144      | 0.979         | 0.002 0   |
| 4 000             | Gaussian    | 0.054 3      | 0.161 6          | 0.336                     | 24 194      | 0.949         | 0.000 6   |
| 4 500             | Gaussian    | 0.048 7      | 0.159 4          | 0.306                     | 23 208      | 0.950         | 0.000 5   |
| 5 000             | Gaussian    | 0.045 5      | 0.154 0          | 0.295                     | 28 007      | 0.948         | 0.001 2   |

表6 不同尺度下抽样单元内水稻种植面积比例的变异函数模型及其参数

| 网格尺度<br>scale (m) | 理论模型        | 块金值<br>$C_0$ | 基台值<br>$C_0 + C$ | 块金系数<br>$C_0 / (C_0 + C)$ | 变程<br>Range | 决定系数<br>$R^2$ | 残差<br>RSS |
|-------------------|-------------|--------------|------------------|---------------------------|-------------|---------------|-----------|
| 500               | Exponential | 0.046 2      | 0.134 4          | 0.344                     | 19 238      | 0.964         | 0.003 5   |
| 1 000             | Exponential | 0.016 3      | 0.127 6          | 0.128                     | 20 040      | 0.990         | 0.002 7   |
| 1 500             | Exponential | 0.007 7      | 0.104 4          | 0.074                     | 21 810      | 0.982         | 0.001 9   |
| 2 000             | Exponential | 0.003 3      | 0.099 6          | 0.033                     | 20 700      | 0.971         | 0.002 1   |
| 2 500             | Exponential | 0.002 5      | 0.068 1          | 0.037                     | 21 360      | 0.966         | 8.3E - 05 |
| 3 000             | Exponential | 0.000 1      | 0.063 5          | 0.002                     | 24 024      | 0.923         | 2.2E - 04 |
| 3 500             | Exponential | 0.000 1      | 0.063 8          | 0.002                     | 22 630      | 0.957         | 6.6E - 05 |
| 4 000             | Spherical   | 0.000 1      | 0.063 4          | 0.016                     | 23 423      | 0.883         | 1.5E - 04 |
| 4 500             | Spherical   | 0.000 7      | 0.062 0          | 0.011                     | 24 760      | 0.792         | 2.1E - 04 |
| 5 000             | Spherical   | 0.001 2      | 0.063 1          | 0.012                     | 25 760      | 0.854         | 1.3E - 04 |

注: Exponential 表示指数函数模型, Spherical 表示球状模型

台值集中在 0.11 左右, 2 500 ~ 5 000m 范围内基台值集中在 0.06 左右, 两阶段的基台值差异较大, 说明在不同抽样单元尺度下水稻的空间变异特征突出。

由于块金系数  $C_0 / (C_0 + C)$  能够反映作物面积空间变异的内在结构特征, 比值越低说明由系统自身 的空间自相关因素 (土壤、气候等自然因素) 引起的空间变异程度高, 即空间自相关越明显, 比值越高则说明空间自相关性越小, 变异主要由随机因素引起。因此, 以空间变异性的研究尺度为横坐标, 块金系数为纵坐标, 绘出德惠市玉米和水稻在连续尺度上的空间变异结构特征折线图, 参考图 5。

从图 5 可以看出, 玉米在 500m 处空间变异性最强 ( $C_0 / (C_0 + C) = 0.497$ ), 在第二阶段内空间变异性最强的尺度为 2 500m ( $C_0 / (C_0 + C) = 0.402$ ), 在第三阶段内空间变异性最强的尺度为 4 000m ( $C_0 / (C_0 + C) = 0.402$ ), 3 种尺度下的块金系数均大于 40% 为中等强度的空间相关, 说明在这 3 种抽样单元尺度下的空间变异性主要源自研究区内部生态系统自身的结构因素作用的结果, 即玉米面积的种植结构相对比较稳定; 对水稻而言, 两阶段内空间变异性最强的

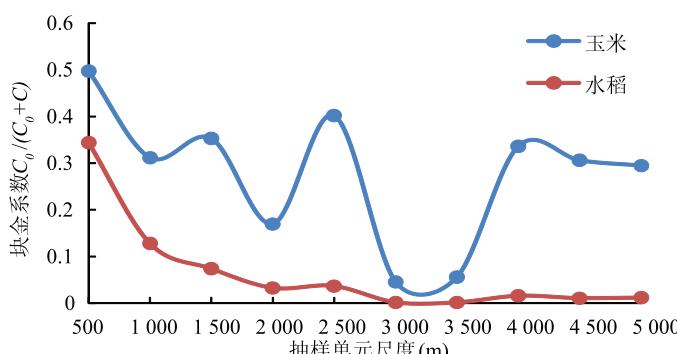


图 5 不同尺度下抽样单元内农作物面积空间变异结构变化

尺度分别为 500m ( $C_0/C + C_0 = 0.344$ ) 和 2 500m ( $C_0/C + C_0 = 0.037$ )，500m 处块金系数大于 25% 为中等自相关，结合表 6 发现自 1 500m 之后水稻的空间块金系数出现急剧下降的趋势，甚至在部分较大尺度下有  $C_0 = 0.0001$ ，说明自 1 500m 之后，水稻的空间变异性基本完全来自系统的空间自相关过程（结构因素），且在部分较大尺度上具有恒定空间变异（系统空间自相关引起的变异）。变程 Range 与块金系数呈现出相反的变化特征，随抽样单元尺度的增加而增大，即农作物空间依赖的范围也随着增大，特别是水稻在大尺度上具有明显空间自相关特征，但其空间变化的差异程度明显减小。

从玉米和水稻的空间变异特征来看，块金系数在整体上随着抽样单元尺度的增加而减小，即空间自相关对作物空间变异性的影响逐渐增大，表明在较大尺度上抽样单元内农作物面积比的相似性更强，具有相似农作物面积的抽样单元空间聚集作用明显。

## 2.3 空间变异性对农作物面积空间抽样效率的影响

### 2.3.1 不同抽样单元尺度下的抽样外推总体估计

为了对比 10 种抽样单元尺度下不同抽样方案的抽样外推精度和稳定性，分别计算玉米和水稻的抽样外推相对误差和变异系数，分析其在不同抽样单元尺度下的变化趋势。图 6 和图 7 给出了采用 3 种抽样方法（简单随机抽样 SR，系统抽样 SY，分层抽样 SS）对玉米和水稻进行抽样时抽样外推总体相对误差 ( $Re$ ) 和外推总体变异系数 ( $CV(\hat{Y})$ ) 的变化情况。可以看出，无论选择任何一种抽样比，通过 3 种抽样方法外推总体得到的  $Re$  及  $CV(\hat{Y})$  与抽样单元尺度之间具有非常明显的相关性。由图 6 可以看出，随

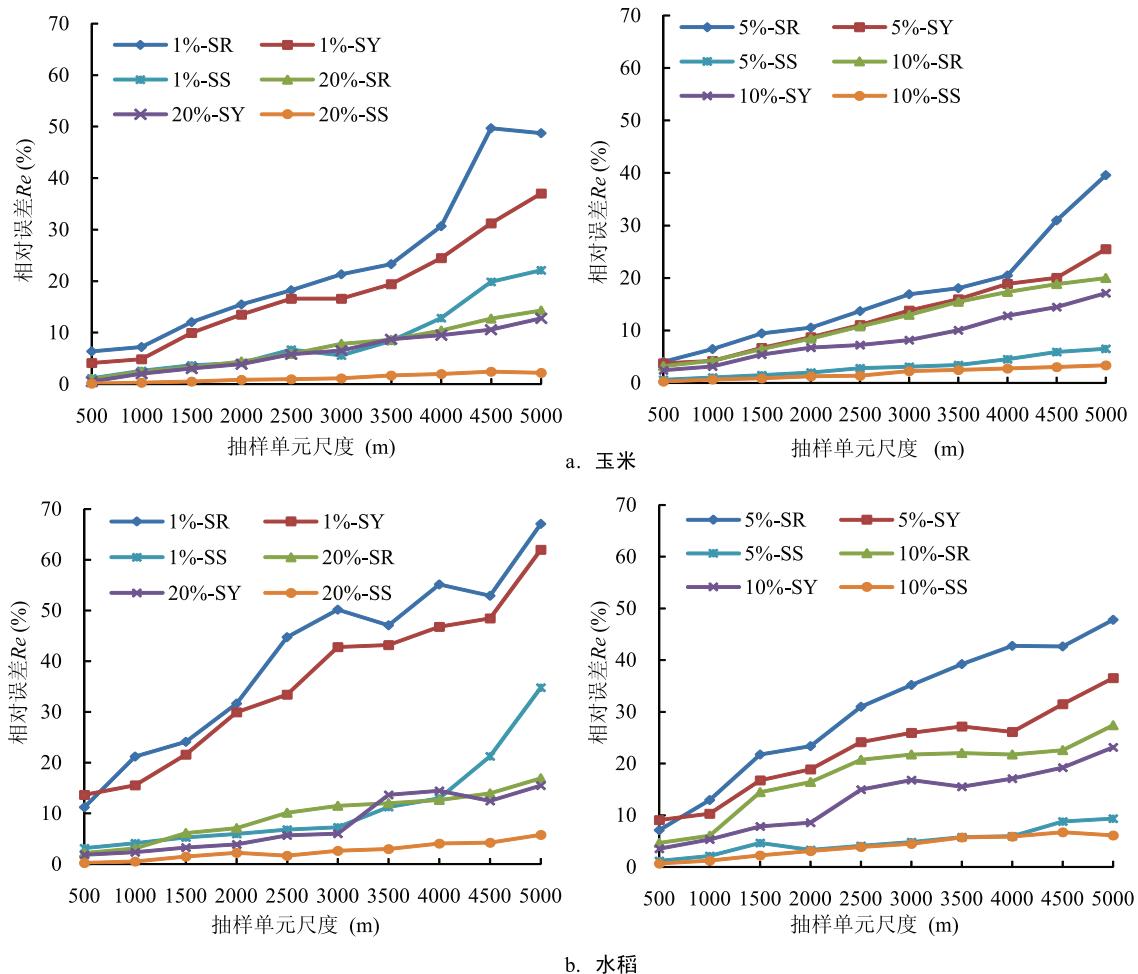


图 6 农作物面积抽样外推总体相对误差随抽样单元尺度的变化

注：SR 表示简单随机抽样，SY 表示系统抽样，SS 表示分层抽样

着抽样单元尺度的增加, 玉米和水稻种植面积估计的  $Re$  均呈逐渐增加的趋势, 说明随着系统内结构因素引起的空间变异性逐渐增加, 样本单元受空间自相关性的影响越强, 从而导致抽样误差逐渐增大; 对比图 7a 和图 7b 发现, 采用简单随机和系统抽样方法外推估计作物面积时,  $CV(\hat{Y})$  变化曲线的变化幅度随着抽样单元尺度的增加逐渐增大, 而分层抽样方法的  $CV(\hat{Y})$  虽然随着抽样单元尺度的增加而增加, 但其变化幅度非常小, 说明采用分层抽样方法时的外推稳定性优于简单随机和系统抽样方法。

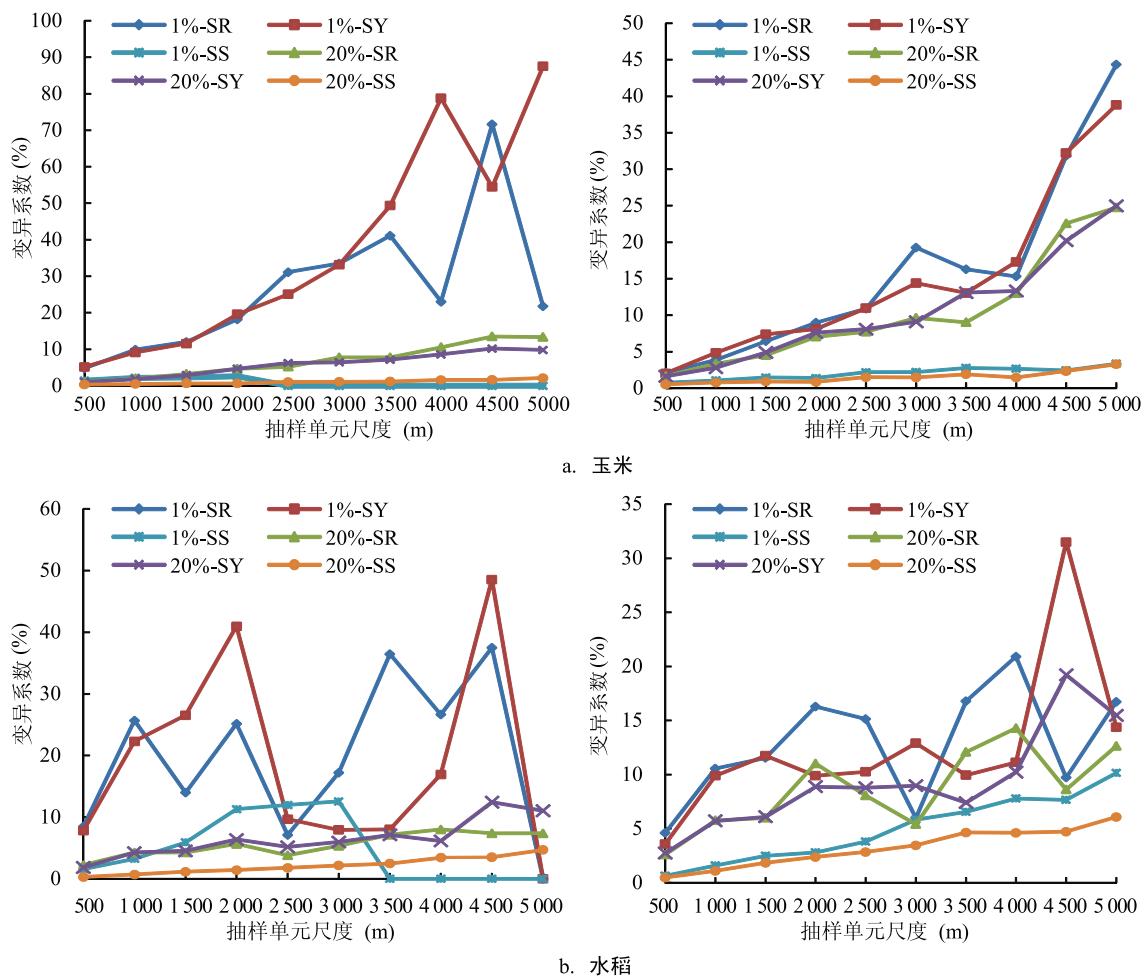


图 7 农作物面积抽样外推总体变异系数随抽样单元尺度的变化

### 2.3.2 抽样方法对空间抽样效率的影响

分别计算 4 种抽样比下玉米和水稻面积  $Re$  和  $CV(\hat{Y})$  的平均值, 以平均相对误差和平均变异系数代表当前抽样方法在不同抽样单元尺度下的抽样精度和外推稳定性。参考图 8 中 2 种农作物的  $Re$  在 10 中抽样单元尺度下的变化范围, 发现简单随机抽样的平均相对误差最大, 系统抽样次之, 分层抽样具有最小平均相对误差, 对应的, 当采用分层抽样方法时其平均变异系数的变化波动最小, 外推稳定性最高, 变化范围基本控制在 3% 以内。综合比较图 6、图 7 和图 8 种 3 种抽样方法的抽样精度 ( $Re$ ) 和稳定性 ( $CV(\hat{Y})$ ), 分层抽样为最理想的抽样方法。

### 2.3.3 抽样比对空间抽样效率的影响

为了分析不同抽样单元尺度下的相对误差  $Re$ 、变异系数  $CV(\hat{Y})$  与抽样比  $f$  之间的关系, 根据图 6~8 的结论, 综合考虑抽样相对误差及变异系数的影响因素, 选择分层抽样为最优抽样方法, 对比玉米和水稻的分层抽样相对误差在不同抽样比下的变化趋势。图 9 给出了玉米和水稻的分层抽样相对误差的变化曲

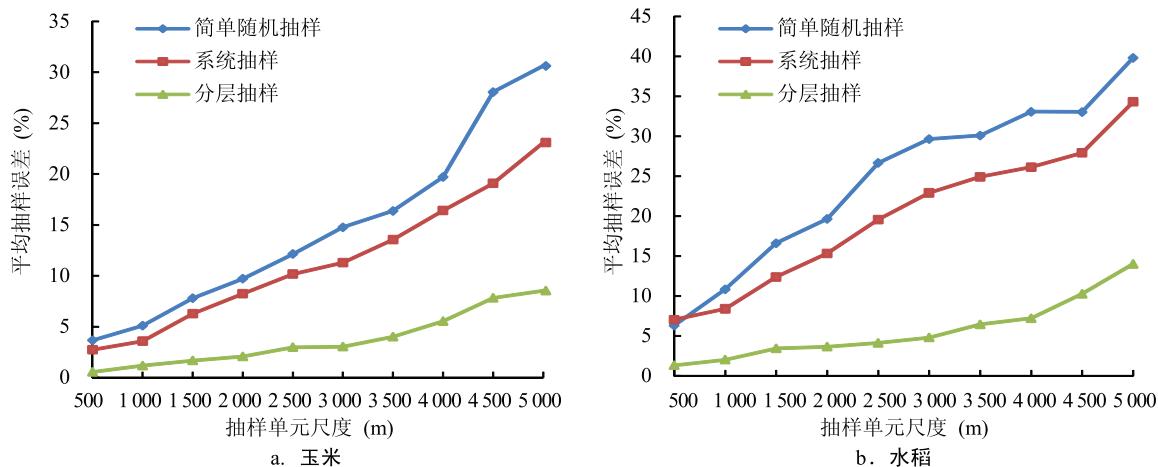


图 8 平均抽样相对误差随抽样单元尺度的变化曲线

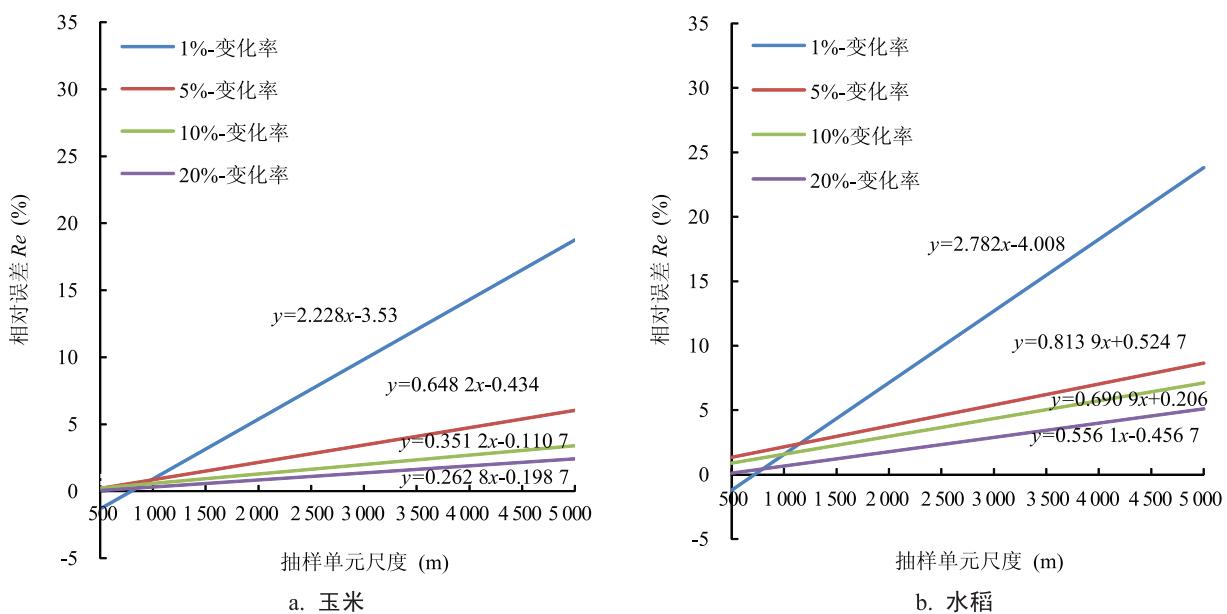


图 9 抽样比随抽样单元尺度的变化

线及其变化率，可以看出，随着  $f$  增加， $Re$  的变化趋势逐渐趋于水平（玉米： $R^2 > 0.85$ ，水稻： $R^2 > 0.73$ ），说明分层抽样的外推总体精度随着  $f$  的增加得到有效提升。对变化率曲线求一阶导数得到变化率曲线的斜率，通过斜率对比 4 种抽样比下分层抽样的  $Re$  随抽样单元尺度的变化情况。根据图 9 中分层抽样相对误差的变化率曲线可以看出，1% 抽样比下农作物相对误差变化率曲线的斜率 ( $y' = 2.23$ 、 $y' = 2.78$ ) 远远大于 5%、10% 和 20%，但是从 5% 开始，斜率之间的差异大幅减小，分别限制在 (0.08, 0.39) 和 (0.13, 0.26)，说明当抽样比达到 5% 以后，通过增加  $f$  来减小  $Re$  的效果甚微，即抽样精度基本趋于稳定，综合考虑图 8 中分层抽样  $CV(\bar{Y})$  的变化情况和抽样费用，选择 5% 的抽样比作为确定空间抽样样本容量的最优抽样比。

#### 2.3.4 抽样单元尺度对空间抽样效率的影响

由于研究目标有两种农作物，为了进一步选择最优抽样单元尺度，根据图 7~9 的结论，计算 4 种  $f$  下，玉米和水稻种植面积在各抽样单元尺度下的平均分层抽样相对误差、和平均分层变异系数，以分层平

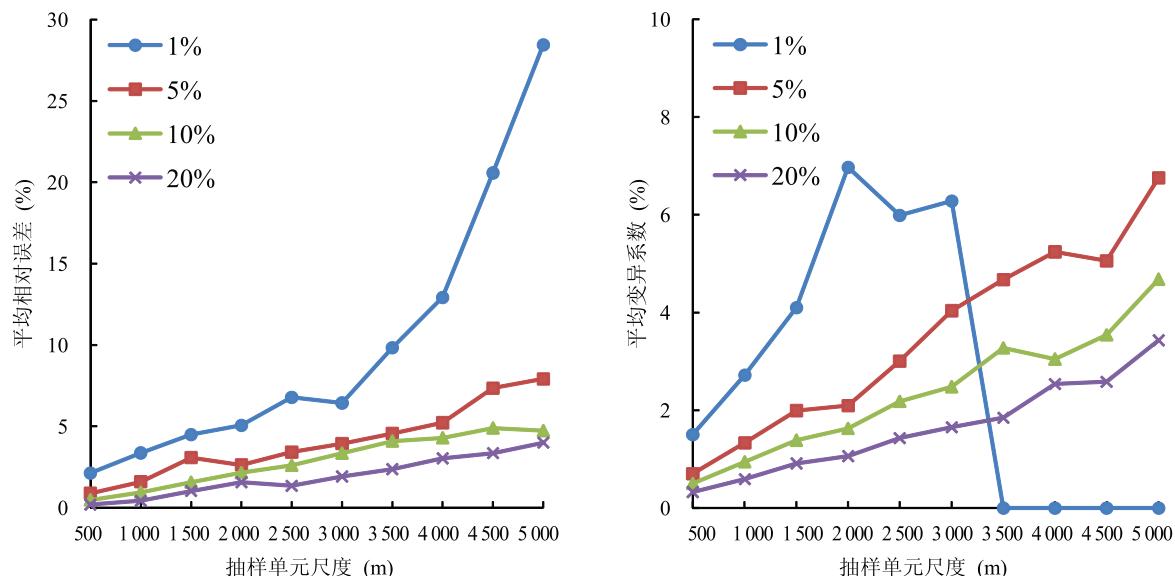


图 10 农作物在分层抽样条件下的空间抽样效率

注: 当抽样单元尺度大于  $3500\text{m} \times 3500\text{m}$  时, 由于层内方差为 0, 导致对应变异系数为 0

均抽样误差和分层平均变异系数作为德惠市农作物空间抽样精度和稳定性的评选标准。从图 10 可以看出, 4 种  $f$  下的平均相对误差和平均变异系数均随着抽样单元尺度的增加而增加; 在 5% 抽样比条件下, 当抽样单元尺度控制在  $3500\text{m} \times 3500\text{m}$  以内时, 农作物的分层平均相对误差小于 5%, 即面积估计精度大于 95%, 变异系数小于 7%。

综合德惠市作物空间变异特征, 玉米和水稻在  $3500\text{m}$  尺度内的最大空间变异对应抽样单元尺度分别为  $500\text{m}$  和  $2500\text{m}$ , 考虑到  $500\text{m}$  抽样单元尺度下的总体单元个数大于 1 万, 在节省抽样费用的前提下,  $2500\text{m} \times 2500\text{m}$  的抽样单元尺度为最适宜进行空间变异性农作物的抽样调查尺度。

### 3 结论与讨论

该文重点分析了农作物种植面积调查以具有空间变异性的玉米和水稻为调查对象时, 空间变异性对抽样外推总体精度和稳定性的影响, 通过对比不同抽样方案的抽样效率, 实现遥感空间抽样方案的优选。结果如下。

(1) 德惠市玉米和水稻面积的空间变异具有明显的尺度依赖特征, 由空间自相关(结构因素)引起的作物面积变异显著。两种作物在  $500\text{m} \times 500\text{m}$  和  $2500\text{m} \times 2500\text{m}$  的抽样单元尺度下的空间变异性较强, 空间自相关性相对最弱, 水稻在  $2500\text{m} \times 2500\text{m}$  的抽样单元尺度下基本具有恒定的空间变异性。

(2) 随着抽样单元尺度的增加, 3 种抽样方法的  $Re$  和  $CV(\hat{Y})$  逐渐增大, 简单随机抽样和系统抽样平均相对误差和平均变异系数的变化范围分别限制在 35% 和 25% 以内, 分层抽样的平均相对误差和平均变异系数的整体变化范围最低, 基本控制在 10% 以内, 说明采用分层抽样方法进行抽样时具有相对最优的抽样精度和外推稳定性;

(3) 随着  $f$  的增加, 作物面积估计的  $Re$  逐渐降低, 变化率逐渐减小。当  $f$  达到 5% 时,  $Re$  基本控制在 10% 以内,  $CV(\hat{Y})$  稳定在 12% 以内, 继续增大  $f$  对提高抽样精度和外推稳定性效果甚微, 综合考虑抽样费用, 选择 5% 作为最优抽样比。

(4) 在抽样方法和抽样比确定的情况下, 随着抽样单元尺度的增加, 作物面积外推估计的  $Re$  和  $CV(\hat{Y})$  也逐渐增加。当抽样单元控制在  $3500\text{m} \times 3500\text{m}$  以内时,  $Re$  限制在 5% 以内,  $CV(\hat{Y})$  稳定在 7% 以内。

综合德惠市玉米和水稻的空间变异性程度、样本容量和抽样方法 3 个评价指标, 认为以  $2\ 500\text{m} \times 2\ 500\text{m}$  构建抽样单元、采用分层抽样方法按 5% 抽样比进行抽样的空间抽样方案是适宜空间变异的农作物面积进行抽样外推调查的最高效方案。

该文在抽样单元的选取和抽样方案的设计上选取了具有空间变异性的调查对象, 但是在空间变异性分析的方法上, 仅采用了各向同性变异函数模型、基台值作为空间变异性的定义模式, 没有考虑到各向异性变异函数模参数可能会造成不同的变异性结果。在实际应用中, 不同的定义模式可能会影响抽样相对误差和变异系数的变化趋势, 因此在接下来的研究中可以考虑根据不同空间定义模式判定空间变异性, 然后利用分层抽样方法, 分析它们在样本容量、抽样精度和稳定性方面的差异。

## 参考文献

- [1] Chauhan H J, Arora M K, Agarwal A. Estimating land cover class area from remote sensing classification. *Journal of Applied Remote Sensing*, 2008, 2 (1): 183–198.
- [2] Quarmby N A, Milnes M, Hindle T L, et al. The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *International Journal of Remote Sensing*, 1993, 14 (2): 199–210.
- [3] 张养贞, 张树文. 县级玉米遥感估产实验及其效果研究. *地理科学*, 1995, 15 (2): 144–153.
- [4] Reynolds C A, Yitayew M, Slack D C, et al. Estimating crop yields and production by integrating the FAO crop specific water balance model with real-time satellite data and ground-based ancillary data. *Int. J. Remote Sens.*, 2000, 21 (18): 3487–3508.
- [5] Fulu T, Masayuki Y, Zhan Z. Remote sensing of crop production in China by production efficiency models: models comparisons estimates and uncertainties. *Ecology Modeling*, 2005, 183 (4): 385–396.
- [6] 陈沈斌. 小麦、玉米和水稻遥感估产技术试验研究文集. 北京: 中国科学技术出版社, 1993. 34–38.
- [7] 周清波. 国内外农情遥感现状与发展趋势. *中国农业资源与区划*, 2004, 25 (5): 9–14.
- [8] Tsiligirides TA. Remote sensing as a tool for agricultural statistics: a case study of area frame sampling methodology in Hellas. *computers and Electronics in Agriculture*, 1998, 20 (1): 45–77.
- [9] 王迪, 周清波, 刘佳. 作物面积空间抽样研究进展. *中国农业资源与区划*, 2012, 33 (2): 9–14.
- [10] 邬明权, 杨良闻, 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. *农业工程学报*, 2014, 30 (2): 146–152.
- [11] 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. *农业工程学报*, 2013, 29 (2): 184–191.
- [12] 王迪, 周清波, 陈仲新, 等. 空间抽样方法估算冬小麦播种面积. *农业工程学报*, 2012, 28 (10): 177–184.
- [13] Benedetti R, Bee M, Piersimoni F, et al. Agricultural survey methods. Wiley Online Library, 2010.
- [14] Gallego F J. Crop area estimation in the MARS project: Conference on ten years of the MARS Project, Brussels. 1999, 1–11.
- [15] Carfagna E, Gallego F J. Using remote sensing for agricultural statistics. *International statistical review*, 2005, 73 (3): 389–404.
- [16] 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. *农业工程学报*, 2012, 28 (23): 114–122.
- [17] 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. *遥感学报*, 2004 (06): 551–569.
- [18] Gallego F J, Stibig, Hans Jürgen. Area estimation from a sample of satellite images: The impact of stratification on the clustering efficiency. *International Journal of Applied Earth Observation and Geoinformation*, 2013, 22 (Complete): 139–146.
- [19] 焦险峰, 杨邦杰, 裴志远. 全国棉花种植面积遥感监测抽样方法设计. *农业工程学报*, 2002 (04): 159–162.
- [20] Zhao G, Hoffmann H, Yeluripati J B, et al. Evaluating the precision of eight spatial sampling schemes in estimating regional means of simulated yield for two crops. *Environmental Modelling & Software*, 2016, 80: 100–112.
- [21] Gallego F J, Bamps C. Using CORINE land cover and the point survey LUCAS for area estimation. *International Journal of Applied Earth Observations and Geoinformation*, 2008, 10 (4): 467–475.
- [22] Overmars K P, Koning G H, Veldkamp A. Spatial autocorrelation in multi-scale land use models. *Ecological Modelling*, 2003, 164 (2): 257–270.
- [23] Cliff A D, Ord J K. Spatial Processes, Models and Applications. *Journal of the Royal Statistical Society*, 1981, 147 (3): 238.
- [24] 王劲峰. 空间抽样与统计推断. 北京: 科学出版社, 2009.
- [25] Holmberg H, Lundqvist E H. A test for robust detection of residual spatial autocorrelation with application to mortality rates in Sweden. *Spatial Statistics*, 2015, 14: 365–381.
- [26] Curran P J. The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, 1988, 24: 493–507.
- [27] Woodcock C E, Strahler A H, Jupp D L B. The use of variograms in remote sensing: II. Real digital images. *Remote Sensing of Environment*, 1988, 25 (3): 349–379.

- [28] Oliver M A, Shine J A, Slocum K R. Using the variogram to explore imagery of two different spatial resolutions. International Journal of Remote Sensing, 2005, 26 (15): 3225–3240.
- [29] Garrigues S, Allard D, Baret F et al. Multivariate quantification of landscape spatial heterogeneity using variogram models. Remote sensing of environment, 2008, 112 (1): 216–230.
- [30] 万丽. 基于变异函数的空间异质性定量分析. 统计与决策, 2006, (2): 26–27.
- [31] 李灿, 张凤荣, 朱泰峰, 等. 大城市边缘区景观破碎化空间异质性——以北京市顺义区为例. 生态学报, 2013, 33 (17): 5363–5374.
- [32] 施黎莉. 复杂类型海洋环境监测数据的空间抽样方法优化. 上海海洋大学, 2016.
- [33] 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. 东北林业大学学报, 2011, 39 (9): 49–51.
- [34] Ping J L, Green C J, Zartman R E, et al. Exploring spatial dependence of cotton yield using global and local autocorrelation statistics. Field Crops Research, 2004, 89 (2): 219–236.
- [35] 李连发, 王劲峰, 刘纪远. 国土遥感调查的空间抽样优化决策. 中国科学D辑地球科学, 2004,
- [36] 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. 农业工程学报, 2000, 16 (5): 126–129.
- [37] 胡潭高, 张锦水, 潘耀忠, 等. 景观破碎度在冬小麦面积抽样设计中的应用研究. 遥感学报, 2010, 14 (6): 1127–1146.
- [38] 胡永森. 多尺度遥感数据结合空间抽样方法的大区域作物面积估算. 东华理工大学, 2016, 31–40.
- [39] 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. 农业工程学报, 2007, 23 (12): 150–155.
- [40] Wang J, Liu J, Zhuan D, et al. Spatial sampling design for monitoring the area of cultivated land. International Journal of Remote Sensing, 2002, 23 (2): 263–284.
- [41] 林芳芳, 刘金福, 路春燕, 等. 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. 福建农林大学学报(自然科学版), 2017, 46 (6): 678–684.
- [42] 张焕雪, 李强子, 文宁, 等. 农作物种植面积遥感抽样调查的误差影响因素分析. 农业工程学报, 2014, 30 (13): 176–184.
- [43] Wang D, Zhou Q B, Yang P, et al. Design of a spatial sampling scheme considering the spatial autocorrelation of crop acreage included in the sampling units. Journal of Integrative Agriculture, 2018, 17 (09): 2096–2106.
- [44] 刘付程, 史学正, 于东升, 等. 基于地统计学和GIS的太湖典型地区土壤属性制图研究——以土壤全氮制图为例. 土壤学报, 2004, 41 (1) 20–27.
- [45] 路鹏, 彭佩钦, 宋变兰, 等. 洞庭湖平原区土壤全磷含量地统计学和GIS分析. 中国农业科学, 2005, 38 (6): 1204–1212.
- [46] 王政权. 地统计学及其在生态学中的应用. 北京: 科学出版社, 1999, 35–149.
- [47] White J G, Welch R M, Norvell W A. Soil zinc map of the USA using GIS. Soil Sci. Soc. Am. J, 1997, 61: 185–194.
- [48] Gotway C A, Hergert G W. Incorporating spatial trends and anisotropy in geostatistical mapping of soil properties. Soil Sci. Soc. Am. J, 1997, 61: 298–309.
- [49] 温兆飞, 张树清, 白静, 等. 农田景观空间异质性分析及遥感监测最优尺度选择——以三江平原为例. 地理学报, 2012, 67 (3): 346–356.
- [50] 王迪, 陈仲新, 周清波, 等. 基于3种估计量的冬小麦播种面积总体外推效率比较. 中国农业资源与区划, 2013, 34 (6): 74–80.
- [51] 刘爱利, 王培法, 丁园圆. 地统计学概论. 北京: 科学出版社, 2012.
- [52] 杜子芳. 抽样技术及其应用. 北京: 清华大学出版社, 2005. 87–89.
- [53] 李金昌. 应用抽样技术. 北京: 科学出版社, 2006. 78–81.
- [54] 钱翌, 于洪, 王灵. 乌鲁木齐市米东区农田土壤重金属含量的空间分布特征. 干旱区地理, 2013, 36 (2): 303–310.
- [55] 陈利顶, 吕一河, 傅伯杰, 等. 基于模式识别的景观格局分析与尺度转换研究框架. 生态学报, 2006, 26 (3): 663–670.
- [56] 布仁仓, 李秀珍, 胡远满, 等. 尺度分析对景观格局指标的影响. 应用生态学报, 2003, 14 (12): 2181–2186.

## SPATIAL VARIABILITY OF SAMPLING UNIT OF CROP AREA AND ITS EFFECT ON SAMPLING EXTRAPOLATION EFFICIENCY<sup>\*</sup>

Zhong Geji, Zhou Qingbo, Wang Di<sup>\*</sup>

(Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences/Key Laboratory of Agricultural Remote Sensing, Ministry of Agricultural and Rural Affairs, Beijing 100081, China)

**Abstract** The spatial sampling technique combined with remote sensing and traditional sampling method is an effective method for monitoring crop area in large areas. However, traditional sampling requires that sampling units should be independent of each other, and does not consider whether the spatial variability of regional crops has an impact on the sampling efficiency of crop area and its degree of influence, thus limits the efficiency and accuracy of crop area monitoring. In view of the above problems, this study took Dehui city, Jilin province as the research area, the spatial variogram was used to quantitatively evaluate the spatial variability of crop area in 10 sampling units. To quantitatively evaluate the effect of spatial variation on sampling efficiency and to optimize sampling scheme for spatial variability crops, three sampling methods, i. e. , simple random (SR), systematic sampling (SY), and stratified sampling (ST), were selected to calculate sample size ( $n$ ), overall relative error of extrapolation (Re), and coefficient of variation (CV( $\hat{Y}$ )) as indices of sampling efficiency under different spatial variation. The results indicated that the spatial variability of the area of maize and rice (reflected as the base value  $C_0 + C$ ) decreased with the increase of the sampling unit scale, among them, the spatial autocorrelation of crop area was relatively small at the sampling scale of 500m × 500m and 2 500m × 2 500m. Under the same sampling ratio, the relative error and coefficient of variation of crop area under the three sampling methods all increased with the decrease of spatial variability (the sampling unit scale increases). The stratified sampling relative error and stratification coefficient of variation of maize and rice had relative minimum, which were limited to 10% and 20%, respectively, followed by systematic sampling (Re < 40%, CV( $\hat{Y}$ ) < 88%) and simple random sampling (Re < 50%, CV( $\hat{Y}$ ) < 75%). When the sampling method was selected, the relative error rate of stratified sampling of the two crops decreased rapidly with the increase of the sampling ratio, when the sampling ratio increased to 5%, the relative error and the coefficient of variation of the coefficient of variation tended to be stable. Within the sampling unit scale of 3 500m × 3 500m, the sampling precision (Re < 5%) could be achieved when stratified sampling was conducted on the area of maize and rice according to the sample size determined by 5% sampling ratio. Considering the spatial variation and spatial sampling efficiency of crops at different sampling unit scales, the stratified sampling method with 2 500m × 2 500m scale and 5% sampling ratio was a high-precision, low-cost and representative program for crop area with spatial variation. This study provides technical support and reference for the improvement of sampling method and sampling efficiency of regional crop area.

**Keywords** crop area; sampling unit scale; spatial variability; spatial sampling; sampling efficiency