

引用格式:杨江燕,殷守强,张利,等. 基于空间聚类分层抽样的黄骅市县域耕地质量等别监测样点布设[J]. 资源科学, 2019, 41(2): 257-267. [Yang J Y, Yin S Q, Zhang L, et al. The layout of county area cultivated land quality monitoring samples in Huanghua City based on spatial clustering stratified sampling[J]. *Resources Science*, 2019, 41(2): 257-267.] DOI :10.18402/resci.2019.02.05

基于空间聚类分层抽样的黄骅市县域耕地质量等别监测样点布设

杨江燕¹, 殷守强², 张利¹, 门明新¹, 陈影¹

(1. 河北农业大学国土资源学院, 保定 071000;

2. 中国矿业大学地球科学与测绘工程学院, 北京 100083)

摘要:耕地质量监测是当前国家及时掌握耕地质量和产能变化的重要工作手段。合理布设耕地质量监测样点可以大幅提高耕地质量监测的效率。以河北省黄骅市县域耕地为研究对象,以耕地的质量属性和空间位置属性为耕地质量等别监测区的划定依据,采用空间聚类法生成20种备选监测区方案,通过综合比较方差、抽样误差、抽样效率和抽样弹性系数,选出初始监测区方案,并在此基础上局部优化,最终生成耕地质量等别监测区方案,最后以耕地质量等别监测区为分层依据,运用分层抽样法布设了耕地质量等别监测样点。研究表明:①在抽样误差为1%的要求下,分区数为65的备选耕地质量等别监测区方案的综合样本容量为77,抽样效率相对较高,被选定为初始监测区方案;②随着耕地质量等别监测区数目从5增加到100,相同监测区内耕地单元之间在空间位置和耕地质量等别上的差异程度由快速减少到趋于平稳;③在布设相同数量的监测样点时,本文提出的基于空间聚类分层抽样的样点布设方法在表土质地、剖面构型、盐渍化、有机质含量、排水条件、灌溉条件和国家耕地质量自然等指数等因素方面的抽样误差分别为0.37、1.02、1.39、0.91、0.31、1.53和1.27,均明显低于传统的等别分层抽样、简单随机抽样、网格分层抽样,具有较高的抽样效率。研究成果可为耕地质量等别监测样点布设相关工作和研究提供有效指导。

关键词:耕地质量等别;监测样点布设;抽样误差;抽样效率;空间聚类分层抽样;黄骅市

DOI :10.18402/resci.2019.02.05

1 引言

耕地是农业生产最重要的资源,其数量、质量与国家的粮食安全密切相关^[1]。近年来,耕地质量管理一直备受国内外学者和政府决策者关注^[2-4]。其中,耕地质量等别监测是及时摸清耕地质量及其变化情况和保护国家粮食安全的有效路径^[5]。而耕地质量等别监测样点布设是耕地质量等别监测的基础和关键^[6],样点布设的准确性与否将关系到整个研究区监测成果的精度与可信度^[7],因而选取具

有综合代表性、稳定性、适宜性的样点至关重要。

关于耕地质量的具体定义,农业部侧重于土壤肥力方面,没有考虑光照、温度等因素的影响。环保部倾向于从污染的角度来考虑,如砷、汞等土壤重金属污染。本文所采用的是国土部所进行的农用地分等中的耕地质量,包括国家耕地质量自然等别、国家耕地质量利用等别、国家耕地质量经济等别,是在气候、土壤、利用状况和经济效益等综合影响下的耕地综合生产能力。

收稿日期:2018-07-14;修订日期:2018-11-08

基金项目:国土资源部两套指标体系试点项目(20170411);河北省典型样带土地资源与生态环境监测评价项目(20170428)。

作者简介:杨江燕,女,安徽安庆人,硕士生,研究方向为土地资源管理。E-mail:15931895996@163.com

通讯作者:门明新, E-mail:menminn@sina.com

前人关于耕地质量等别监测样点布设的研究方法主要包括网格法^[8,9]、因素组合法^[10]、分层抽样法^[11,12]。也有部分学者对变异函数理论模型^[13]和基于Kriging估计误差^[14]，基于地统计学的变异函数的变异结果^[13,15,16]，基于空间平衡法^[17]等方法在耕地质量等别监测样点布设中作出了尝试。布点的依据也主要是国家耕地质量自然等别和国家耕地质量自然等指数^[6,7,12,18]。张玉臻^[19]和郭力娜^[11]等则分别从省级和国家级角度出发，在不同级别上^[11,20]进行了探索。相关学者虽已取得不错成果，但仍有不足：

(1)在监测样点布设过程中，当前研究多是在国家自然质量等指数的基础上进行类别划分，使用面积比例法分配各类型耕地中监测样点数目，而对表土质地、剖面构型、盐渍化程度、土壤有机质含量、灌溉保证率、排水条件等具体因素考虑较少，虽然耕地的国家耕地质量自然等指数的高低在一定程度上反映了以上指标的情况，但是却并不能具体反映每一个因素的变化情况，因此仅仅基于国家耕地质量自然等指数划分类别导致监测样点对各分等因素代表性较差；

(2)由于耕地不仅仅是在各项因素综合作用下形成的自然经济综合体，同时还是一种空间实体，因此，耕地质量等别监测样点布设需要兼顾耕地空间位置属性，而在现有的样点布设研究中较少考虑耕地单元的空间位置；

(3)现阶段主要是针对全国和全省的角度出发，对县域进行研究的还很少，大尺度的研究结果虽在较大范围有较好的可比性，但是难以兼顾到县域耕地质量等别监测样点布设的精度。

在已有研究成果基础之上，为保证监测样点的空间位置代表性和耕地质量代表性，本研究以河北省县域黄骅市为研究对象，以农用地分等成果为数据基础，综合考虑耕地单元的空间位置属性和质量属性，运用空间聚类方法生成备选监测区方案，通过综合比较方差、抽样误差和抽样效率选出初始监测区方案，并经局部优化，得到耕地监测区方案，最后以耕地质量等别监测区为分层依据，运用分层抽样法布设耕地质量等别监测样点，以期今后的耕地质量等别监测样点布设工作和研究提供一定的参考。

2 研究区概况、数据来源与研究方法

2.1 研究区概况

黄骅市地处滨海平原区，界于38°08'N—38°42'N，117°04'E—117°51'E，总面积约为2391.46 km²。全市地势较低，自西南向东北逐渐下降，平均海拔只有3~5 m，属于暖温带半湿润季风气候区，全年无霜期约为210天，日照时数2700小时，年均气温为12.5℃左右，盛行西南风，年均降雨量约为626.6 mm，但降雨分配不均，主要集中在夏季，春、冬两季降水量较少，不能形成径流。该区域土壤比较贫瘠，土壤盐渍化严重，主要包括潮土、沼泽土和滨海盐土3种类型，其中潮土面积最大，土壤含盐量多在0.3%以上，pH值多呈碱性，肥力总体偏低。研究区耕地总面积约869.64 km²，约占研究区总面积的36.36%，耕地图斑共有12 619个，全部分布在黄骅市的西部和中部的11个乡镇，东北沿海地区的2个乡镇无耕地分布。国家耕地质量自然等指数处于1541~3355之间，国家耕地质量自然等别位于7~11等，其中10等地面积最大，高达耕地总面积的68.23%，9等地面积次之，占耕地总面积20.31%。

2.2 数据来源与处理

本研究所采用的国家耕地质量自然等指数、农用地分等因素值、标准耕作制度等耕地质量数据来源于黄骅市2014年1:10万国家耕地质量等别年度更新成果数据库，包括技术报告、矢量数据、表格和图件等。研究区土地利用类型数据和行政区划数据分别来源于2014年1:10万土地利用现状变更数据库中的地类图斑图层和行政区图层，均从黄骅市国土资源局获得。

2.3 研究方法

本文运用空间聚类法生成具有不同分区数的备选监测区方案；通过比较不同备选监测区方案的方差、抽样精度和抽样效率，从中选出初始监测区方案，通过局部优化，生成耕地质量等别监测区方案；最后以此为分层依据，最终建立了涵盖“耕地等别—分等因素—空间位置”差异的监测样点。由于研究区为县尺度，属于滨海平原一年两熟区，地貌和气候差异较小，因此并没有考虑这类因素对监测样点布设的影响。

2019年2月

2.3.1 备选监测区方案生成方法

该文首先综合考虑耕地地块的耕地质量属性和空间位置属性,采用空间聚类法^[21-23]生成备选监测区方案。主要包括以下几个关键步骤。

(1)空间聚类指标的选择。空间聚类指标代表性与否直接决定监测区方案划定的合理性,是保证样点布设准确性的前提。因此,为使生成的同一监测区内的耕地单元质量相似、位置相邻,本文从耕地单元的质量属性和空间位置属性^[24]构建空间聚类指标体系:其中,质量属性包含表土质地、剖面构型、盐渍化程度、土壤有机质含量、灌溉保证率、排水条件、国家自然等指数,空间位置属性包含X坐标值和Y坐标值。

(2)空间聚类指标的标准化。为避免指标单位差异对聚类结果产生的干扰,需对所选指标进行标准化处理。计算公式为:

$$Z_{ik} = 100 \times \frac{F_{ik} - \min(F_k)}{\max(F_k) - \min(F_k)} \quad (1)$$

式中: Z_{ik} 表示第 i 个耕地单元第 k 个指标的标准化值; F_{ik} 表示第 i 个耕地单元第 k 个指标的质量属性原始值; $\max(F_k)$ 表示所有耕地单元第 k 项指标的最大值; $\min(F_k)$ 表示所有耕地单元第 k 个指标的最小值。

(3)耕地单元相似度的计算。结合相关研究^[25,26],为保证耕地单元的空间位置属性和质量属性对聚类结果产生相同的作用,本文赋予两者同等权重,并用属性距离公式度量相似度。两个耕地单元间的距离越小,相似度越大,计算公式为:

$$d_{ij} = \frac{1}{6} \sqrt{\sum_{k=1}^6 (Z_{ik} - Z_{jk})^2} + \frac{1}{2} \sqrt{(Z_{ix} - Z_{jx})^2 + (Z_{iy} - Z_{jy})^2} \quad (2)$$

式中: d_{ij} 表示第 i 个耕地单元与第 j 个耕地单元的距离; Z_{ik} 、 Z_{jk} 分别表示第 i 个耕地单元和第 j 个耕地单元第 k 项分等因素标准值; Z_{ix} 、 Z_{jx} 分别表示第 i 个耕地单元和第 j 个耕地单元的 x 坐标标准值; Z_{iy} 、 Z_{jy} 分别表示第 i 个耕地单元和第 j 个耕地单元的 y 坐标标准值。

(4)空间聚类方法的具体实施。借助 matlab 软件,运用 K-means 聚类法和循环算法,根据所构建的

指标体系对研究区所有的耕地单元进行空间聚类分析,生成 20 个具有不同分区数的备选监测区方案。分区数 K 组成以 5 为首项,100 为末项,5 为公差等的等差数列。K-means 法的基本流程为^[27]:①随机选择 k 个初始聚类中心;②计算单元到各中心的距离,并将单元归到距离最小的类;③重新计算 k 个类的均值;④不断重复步骤②和③,如果标准测度函数收敛或达到最大迭代次数,则迭代结束,否则继续迭代,最后得到备选耕地质量等别监测区。

2.3.2 监测区方案的确定方法

为对备选监测区方案的分层效果进行比较分析,本研究引入方差、抽样误差、抽样效率指标对备选监测区方案的分区效果进行评价,从中选定初始监测区方案,最后将初始监测区方案经过局部优化,确定出最终的监测区方案。

(1)区间方差和区内方差的计算。区间方差,等于每一个分区的属性均值与总体属性均值差值的平方的和,反映不同分区之间属性值的差异程度;区内方差则为每个耕地单元属性值与所在分区的属性平均值的差值的平方和,反映的是相同分区属性值的差异程度。因此,属性值的区内方差越小,区间方差越大,表示该属性值在区内的差异程度越小、均一性越强,在区间的差异程度越大、非均质性越强,分区效果越理想。此处的组间方差和组内方差均借助 SPSS 19.0 软件计算完成。

(2)抽样误差的计算。本文引入统计学中的抽样理论,抽样误差是指因抽样的随机性所导致的样本统计量与总体真实值之间的差异,抽样误差值越大,则分区效果越差,反之则越好。在置信度为 95% 的水平下,抽样均值的最大相对误差(以下简称抽样误差)具体计算公式如下:

$$r(u_{\bar{y}}) = \frac{\sqrt{V} t_{0.025}(n-1)}{u_{\bar{y}}} \times 100\% \quad (3)$$

$$V = \sum W_h^2 (1 - f_h) \frac{S_h^2}{n_h} = \frac{(\sum W_h S_h)^2}{n} - \frac{1}{N} \sum W_h S_h^2 \quad (4)$$

$$n_h = n \times f_h = n \frac{W_h S_h}{\sum W_h S_h} \quad (5)$$

式中: $r(u_{\bar{y}})$ 为抽样均值的最大相对误差; V 为耕地单元某一因素的抽样均值的方差; $t_{0.025}(n-1)$ 为

自由度为 $n-1$, 置信度为 95% 时的概率度; $u_{\bar{y}}$ 为样本均值的抽样分布的均值; W_h 为 h 层耕地单元的数量比例; f_h 为 h 层的抽样比; S_h 为 h 层某属性值的标准差; n_h 为 h 层样本容量; n 为样本容量; N 为评价单元总数目; \bar{y} 为耕地单元某一因素的总体均值。

(3) 综合样本量的计算方法。在实际的样点布设过程中, 在一定的抽样误差前提下, 为了节省物力人力, 应优先选择综合样本量最小的监测区方案。在某个备选方案中, 当抽取 m 个样本时, 各项因素的抽样误差均不大于 r 。因此, m 可看作在该备选方案中, 各项因素满足抽样误差小于 r 的要求时抽样的最少样本量。根据公式(3)、(4)和(5)可推导出在某一因素抽样误差为 $r(u_{\bar{y}})$ 时的样本容量计算公式(6)和(7), 如下:

$$n = \frac{(\sum W_h S_h)^2}{V + 1/N \sum W_h S_h^2} = \frac{(\sum W_h S_h)^2}{\left(\frac{r(u_{\bar{y}})}{t_{0.025}(n-1)}\right)^2 + 1/N \sum W_h S_h^2} \quad (6)$$

$$n_h = n f_h = n \frac{w_h S_h}{\sum w_h S_h} = \frac{(w_h S_h) \sum w_h S_h}{\left(\frac{r(u_{\bar{y}})}{t_{0.025}(n-1)}\right)^2 + 1/N \sum W_h S_h^2} \quad (7)$$

通过公式(6), 可以求得当分层数为 h 时, 在某一项因素的抽样误差为 r 时各层的样本量, 进而可以推导出在所有因素的抽样误差均为 r 时各层的综合样本量。在某一分区方案下, 如果把第 i 项因素抽样误差为 r 时的第 h 层样本量设为 $n(h, i)$ 。那么在所有因素抽样误差均为 r 时, 第 h 层最大样本量为:

$$m_h = \max\{n(h, i)\} \quad (i = 1, 2, \dots, 6) \quad (8)$$

因此, 在所有因素抽样误差为 r 时, 各层最大样本量的和, 即综合样本量为:

$$m = \sum m_h \quad (9)$$

(4) 抽样效率计算方法。在相同样本容量要求下, 抽样方法的抽样误差越低, 抽样效率越高; 在相同抽样误差要求下, 抽样方法的样本容量越少, 抽样效率越高。因此, 抽样效率分别与抽样误差和样本容量呈反向变动关系, 可看作是抽样误差和综合样本量综合作用的函数。通过比较不同抽样方法

的抽样效率, 可权衡它们在样本容量和抽样误差两个方面的综合优势。某种抽样方法的抽样效率高, 表示该抽样方法在样本容量和抽样误差 2 个方面的综合优势越大, 反之则综合优势越低。抽样效率函数计算公式为:

$$E = \frac{1}{mr} \quad (10)$$

式中: E 表示抽样方法的抽样效率; m 表示样本容量; r 表示抽样误差。

(5) 抽样弹性系数计算方法。为进一步深入测度抽样误差随着综合样本量变动程度的大小, 揭示抽样效率的变动情况, 本文借鉴经济学中的弹性系数^[28, 29], 提出抽样误差弹性系数, 计算公式如下:

$$\beta_n = \frac{|r_m - r_n| / (r_m + r_n)}{|m - n| / (m + n)} \quad (11)$$

式中: m 和 n 分别表示抽样误差对应的综合样本量, $n < m$; r_m 和 r_n 分别表示样本容量为 m 和 n 时的抽样误差; β_n 表示综合样本量为 n 时的抽样误差弹性系数。 $\beta_n > 1$, 表示样本容量由 n 变为 m 时, 抽样误差的相对变化率大于综合样本量的相对变化率, 抽样效率增大; $\beta_n = 1$, 表示综合样本量由 n 变为 m 时, 抽样误差的相对变化率等于综合样本量的相对变化率, 抽样效率不变; $\beta_n < 1$, 表示综合样本量由 n 变为 m 时, 抽样误差的相对变化率小于综合样本量的相对变化率, 抽样效率减少。

2.3.3 基于监测区方案耕地质量等别监测样点的分层抽样

耕地质量等别监测样点需要满足下面的条件:

①要保证代表性, 每种耕地类别至少布设一个监测样点, 实现全面监控, 所布设耕地监测样点能够对所有质量因素具有代表性; ②要保持稳定性, 便于长期监测; ③数量不宜过多, 样点数量过多会造成人力物力等资源的浪费, 因此, 需要综合权衡监测精度和监测成本, 制定效率较高的监测方案。

把每个监测区耕地总面积最大的组合类型称为该监测区的主导因素组合类型, 为增强耕地质量等别监测单元的代表性程度, 首先选择主导因素组合耕地单元作为各耕地质量等别监测区的备选耕地质量等别监测单元; 为保证耕地监测单元的稳定性, 便于进行长期的监测, 优先选择其中受到国家

2019年2月

严格保护的基本农田作为最终的备选耕地质量等别监测单元。

3 结果及分析

3.1 备选监测区方案的生成与比较

从图1可以看出,各因素间的区内、区间方差有各自的变化特征。其中,盐渍化程度、剖面构型和土壤有机质随着分区数的不断增大,区内方差逐渐减少、区间方差逐渐增大,但是当分区数增大达到一定程度后,区内、区间方差趋于稳定。表明这3项因素随着监测区数目的不断增大,开始区内变异程度逐渐减少、区间变异程度逐渐增大,分区效果越来越好,但是当分区数达到一定程度时,分区效果变化不明显。灌溉保证率在监测区数目小于65时,区内、区间方差存在较大的波动,当监测区数目大于65后趋于稳定。国家耕地质量自然等指数、表土质地、排水条件从分区数为5到分区数为100的范围内,区间、区内方差均较为稳定,表明监测区数目的增加对改善这3项因素的分区效果并不明显。

通过SPSS 19.0软件得到不同监测区数目下耕地单元质心坐标的区内、区间方差,用来表示不同分区数下耕地单元各项因素的空间分区效果。由图2可知,耕地单元质心坐标的区内方差明显小于区间方差,空间分区效果较好。随着监测区数目的增加,区间方差逐渐增加,区内方差逐渐减少,两者差距逐渐扩大,变化速度由快变慢。表明随着监测

区数目的增加,区内单元的空间聚合度和区间单元的空间分离度逐渐增加,即监测区数目越多,空间分区效果越好,并且可以假设当监测区数目增加到一定值时,区内、区间方差趋于平稳,分区效果达到最好。

综合分析可知,虽然空间聚类方法能够将空间位置属性和质量属性均相对一致的耕地单元划分在一个备选耕地质量等别监测区内,但在同种备选方案中,不同因素的分区效果不同;不同因素分区效果趋于稳定时对应的最小分区数也不同。因此,必须综合考虑20个备选方案中各项因素的分区效果,选择合适的耕地质量等别监测区方案。

3.2 初始监测区方案的选定

通过以上公式(3)—公式(10),可以得到不同备选监测区方案中各抽样误差对应的综合样本量,并用下划线标注出每一列的最小值,作为该抽样误差对应的最小综合样本量,具体见表1。从结果可看出,相同备选方案中,随着抽样误差的增加,综合样本量随之减少,减小速度逐渐变慢,最后逐渐接近备选方案的分区数。当抽样误差在0.1%~1.0%的范围内,抽样误差相同时,最小样本量随分区数的增加整体上逐渐减小,但幅度越来越小;当抽样误差在1.5%~10.0%的范围内,抽样误差一定时,综合样本量随着分区数增加先减小后增加,逐渐与分区数相同。

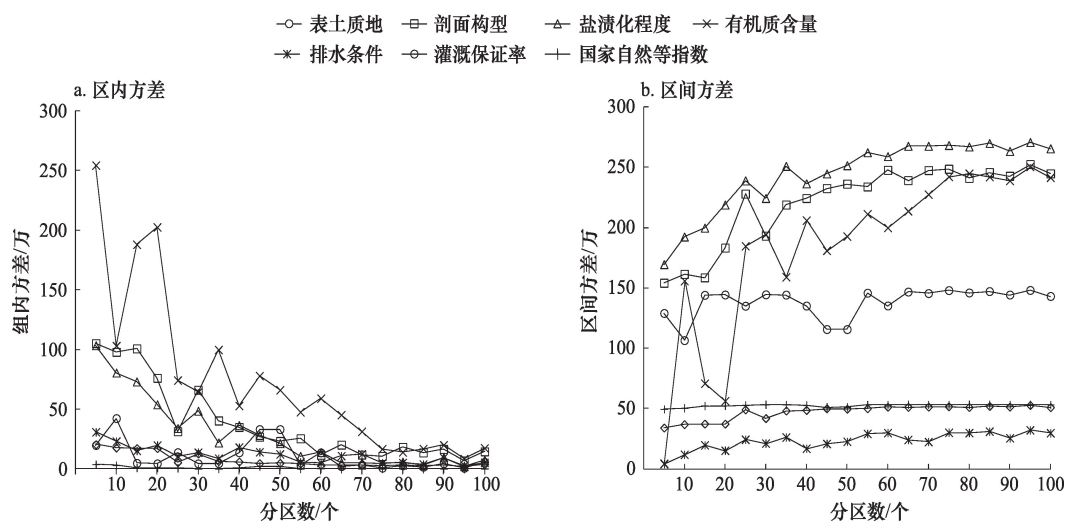


图1 不同监测区数目下各因素的的区内方差和区间方差

Figure 1 Intra-zone and inter-zone variance of each factor in different numbers of monitoring zones

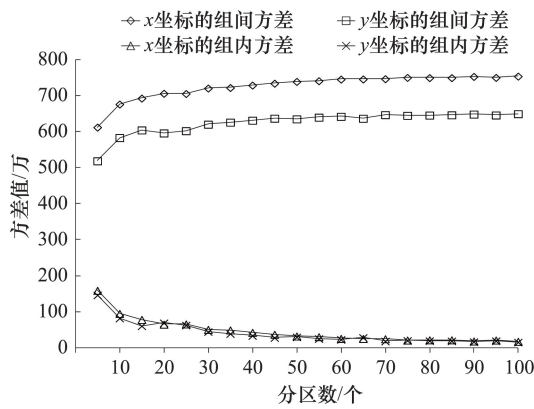


图2 耕地单元质心坐标的区内方差和区间方差的变化曲线

Figure 2 Changes of the intra-zone variance and inter-class variance of space coordinates of cultivated land in different layers

基于结果分析可知,分区数、抽样误差和综合样本量三者相互影响,确定最合适的监测区方案,需综合考虑它们之间的变动关系。使用公式(10)、公式(11)可进一步计算出抽样效率和抽样弹性系

数随着最小综合样本量的变动曲线。从图3结果可以看出,当综合样本量 ≤ 77 时,抽样误差弹性系数绝大部分大于1,抽样误差的相对变化率大于样本容量的相对变化率,抽样效率增大;同理,当综合样本量在77~137之间时,抽样效率先增加后减少;综合样本量 ≥ 137 时,抽样误差弹性系数小于1,抽样效率减小。由分析可知,当抽样误差弹性系数为1时,抽样误差的相对变化率等于样本容量的相对变化率,此时该抽样方案的抽样效率最高,与此相对应,综合样本量在77~137之间,分区数在65~90之间,抽样误差在1.0%和1.5%之间。

综合比较不同备选监测区方案的分区效果,以及整体权衡最小综合样本容量、抽样误差、抽样效率和抽样弹性系数等3项因素,选定分区数为65、抽样误差为1.0%、对应的综合样本容量为77的备选监测区方案为初始监测区方案。

表1 黄骅市不同备选监测区方案中各抽样误差对应的综合样本量

Table 1 Comprehensive sample sizes corresponding to different sampling errors in different alternative schemes of monitoring zones in Huanghua

分区数/个	0.1%	0.2%	0.3%	0.4%	0.5%	1.0%	1.5%	2.0%	2.5%	3.0%	3.5%	4.0%	4.5%	5.0%	6.0%	7.0%	8.0%	9.0%	10.0%
5	14 001	11 422	9 312	7 554	6 271	2 596	1 314	777	511	361	267	206	164	133	94	70	55	43	36
10	12 086	9 634	8 051	6 643	5 441	2 195	1 102	651	425	299	222	170	135	110	75	56	44	34	28
15	11 890	9 649	8 142	6 779	5 590	2 282	1 151	680	447	313	230	180	142	116	78	60	47	38	29
20	9 116	5 764	3 881	2 813	2 118	699	332	192	127	88	69	53	49	38	31	<u>25</u>	<u>23</u>	<u>22</u>	<u>21</u>
25	8 960	5 717	3 679	2 491	1 777	530	245	141	93	68	51	41	36	31	<u>29</u>	26	26	25	25
30	5 775	2 981	1 677	1 041	703	194	89	54	<u>42</u>	<u>34</u>	<u>32</u>	<u>30</u>	<u>30</u>	<u>30</u>	30	30	30	30	30
35	5 602	2 964	1 686	1 058	718	199	97	61	47	38	37	35	35	35	35	35	35	35	35
40	4 272	2 065	1 138	702	475	140	73	<u>52</u>	44	42	40	40	40	40	40	40	40	40	40
45	3 670	1 783	972	600	408	124	<u>56</u>	55	49	46	45	45	45	45	45	45	45	45	45
50	2 798	1 319	719	446	305	105	65	55	51	51	50	50	50	50	50	50	50	50	50
55	2 599	1 215	665	416	287	100	66	59	56	55	55	55	55	55	55	55	55	55	55
60	2 354	1 022	543	337	236	90	68	61	60	60	60	60	60	60	60	60	60	60	60
65	1 465	609	327	213	155	<u>77</u>	67	65	65	65	65	65	65	65	65	65	65	65	65
70	2 180	833	425	261	188	83	72	70	70	70	70	70	70	70	70	70	70	70	70
75	1 517	599	323	211	155	83	75	75	75	75	75	75	75	75	75	75	75	75	75
80	1 728	710	383	250	182	91	82	80	80	80	80	80	80	80	80	80	80	80	80
85	1 413	562	309	208	157	92	85	85	85	85	85	85	85	85	85	85	85	85	85
90	<u>1 026</u>	<u>425</u>	<u>246</u>	<u>173</u>	<u>137</u>	94	91	90	90	90	90	90	90	90	90	90	90	90	90
95	1 601	729	412	275	206	112	97	95	95	95	95	95	95	95	95	95	95	95	95
100	1 265	509	289	201	158	107	100	100	100	100	100	100	100	100	100	100	100	100	100

注:每列带下划线加粗数字是在该误差范围下的最小综合样本量,如第一列带下划线加粗的1026表示在各因素的抽样误差为0.1%时的最小综合样本量为1026,由于随着抽样误差的不断增大,与之相对应的综合样本量及其变化均越来越小,因此抽样误差的节点选取0.1%~0.5%以0.1%为间隔,0.5%~5.0%以0.5%为间隔,5.0%~10.0%以1%为间隔。

2019年2月

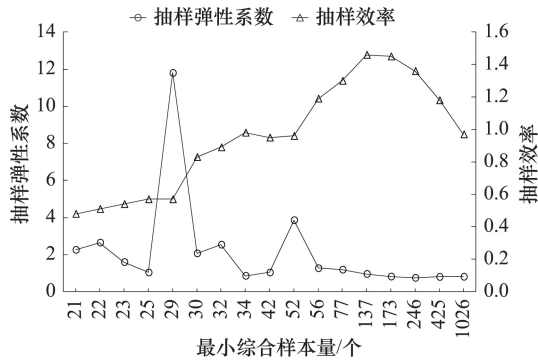


图3 抽样效率和抽样误差弹性系数随着最小综合样本容量变化曲线

Figure 3 Sampling efficiency and change in elastic coefficient of sampling error corresponding under different minimum comprehensive sample sizes

3.3 初始监测区方案的优化

为了增强样点布设的代表性,把各耕地单元分等因素值组合在一起,形成耕地单元的分等因素组合类型。在初始监测区方案中,虽然大部分监测区内组合类型基本相同,且连片分布,但仍有部分监测区内耕地单元的组合类型存在较大差异或空间距离较远的问题,以下针对几种主要不同的情况提出相应的优化方案:

(1)部分同一初始监测区内的耕地单元仍处于相距较远的几片区域,区内耕地单元的空间聚集性较差。针对这种情况,把处于同一个初始监测区内相隔较远的耕地单元划为多个监测区;

(2)部分初始监测区内的耕地单元尽管在空间上连片分布,但分属不同的分等因素组合类型,且面积均较大。针对这种情况,把处于同一个初始监

测区内不同因素组合类型的耕地划分为分等因素组合类型基本相同的多个监测区;

(3)将零散分布或面积较小的组合类型耕地监测区与邻近的监测区合并。这虽会在一定程度上增加抽样误差,但改善了耕地质量等别监测区的空间连续性,并极大地减少了耕地质量等别监测区的数量,节约监测成本。

基于以上方法在初始监测区上优化得到了最终的黄骅市监测区方案,运用SPSS 19.0软件计算优化前后监测区方案中各因素的区内方差及其占总方差的比例,从图4结果可以发现,绝大部分因素的区内方差在经过优化后小于优化前,尽管在优化监测区中某些因素的分区效果变差,但变化幅度不大,各因素的区内方差占总方差的比例均在15%以下,表明优化方案比初始方案在确保相同监测区内耕地单元的质量状况一致性和空间分布连续性两个方面更具综合优势,分区效果更好。

3.4 耕地质量等别监测样点布设

3.4.1 与其他监测样点布设方法的比较

文章以优化监测区为分层依据,采用分层抽样的方法抽取监测单元,共布设了81个监测样点。由于在以往的样点布设研究中,经常采用简单随机抽样方法、等别分层抽样方法和网格分层抽样等3种传统抽样方法布设样点,为了与空间聚类分层抽样结果比较,此处同样采用这3种抽样方法进行样点布设,分别计算它们的抽样误差。等别分层抽样和网格分层抽样误差计算方法工时参考公式(3)-(5),

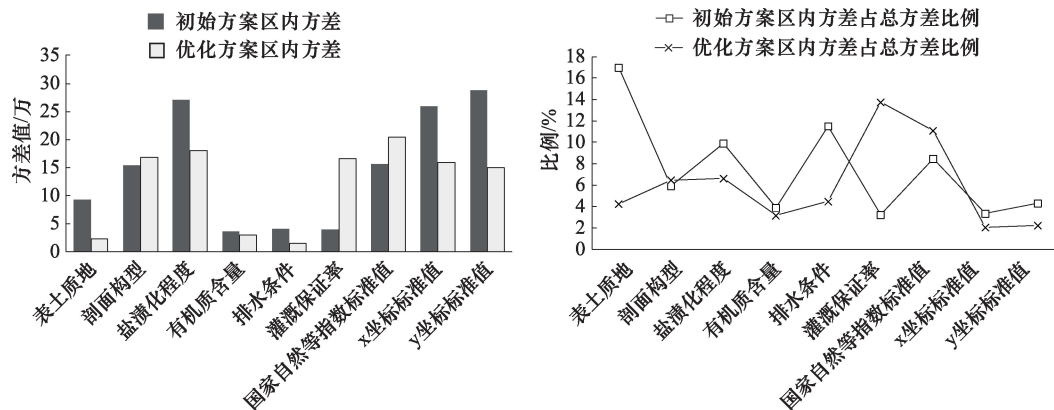


图4 黄骅市不同监测区方案的区内方差及其所占总方差的比例

Figure 4 Intra-class variance and the proportion accounting for the total variance of different factors in two monitoring zoning schemes in Huanghua

简单随机抽样又称纯随机抽样,是从总体中随机抽取某一数量的样本,每个样本被抽中的概率相等。其方差计算公式^[8]为:

$$V(u_{\bar{y}}) = \frac{1-f}{n} \times S^2 \tag{12}$$

式中: $u_{\bar{y}}$ 为抽样均值; V 为 $u_{\bar{y}}$ 的方差; f 为抽样比; n 样本容量; S^2 为总体方差。

从计算结果可以看出(表2),当样本容量均为81时,采用空间聚类分层抽样法布设样点的各质量属性抽样误差均明显小于其他3项,使样点在分等因素和耕地自然等指数方面更具有代表性,具有较好的布样效果。网格分层抽样抽样误差较大是因为该方法仅考虑了耕地单元的空间位置,对耕地单元的其他属性考虑较少,导致抽取的监测单元的质量属性代表性较差;国家耕地质量自然等别指数的等别分层抽样误差比其他抽样方法低,虽然抽取的监测单元对耕地自然等指数的代表性较好,但在分区时没有考虑各项分等因素,导致各项分等因素的抽样误差较大;等别分层抽样和简单随机抽样由于缺乏对耕地单元的空间位置属性的考虑,导致抽取的监测单元的空间代表性较差。

由于文章选取的研究区耕地等别为7、8、9、10、11等,其中面积以10等分布最广。从表3结果中可

以看出,空间聚类分层抽样布设的样点分布涵盖了所有的耕地等别,并且整体上数量和相应的等别面积成比例,较好的考虑到了耕地的等别控制因素;简单随机抽样缺乏对8等别耕地的考虑,等别分层抽样和网格分层抽样虽在样点布设上兼顾到了各等别的耕地,但在抽样误差、空间位置以及其他属性等方面的考虑却不够全面。

3.4.2 耕地质量等别监测样点布设结果

按照监测样点选取的标准以及出于监测单元稳定性的考虑,最终在研究区共有7518个满足条件的耕地单元数。在81个耕地质量等别监测区内,只有1个监测区没有满足条件的耕地,需把普通耕地作为备选耕地质量等别监测单元,最终确定7521个耕地单元作为备选耕地质量等别监测单元。借助SPSS 19.0软件,在筛选出来的7521个备选耕地质量等别监测单元中,以耕地质量等别监测区作为分区依据,运用分层抽样方法,在每一区内随机抽取1个耕地单元作为耕地质量等别监测单元。监测样点位于监测单元的质心位置,最终共布设81个监测样点。其中,80个位于基本农田上,1个位于普通耕地上。耕地质量等别监测样点空间分布情况如图5所示。

表2 黄骅市不同抽样方法各因素的抽样误差的比较

Table 2 Comparison of the sampling error of each factor between different sampling methods in Huanghua

抽样方法	表土质地/%	剖面构型/%	盐渍化/%	有机质含量/%	排水条件/%	灌溉条件/%	耕地质量国家自然等指数/%
空间聚类分层抽样	0.37	1.02	1.39	0.91	0.31	1.53	1.27
等别分层抽样	0.97	2.41	4.28	4.54	1.28	2.04	1.06
简单随机抽样	1.45	3.00	4.23	4.16	1.17	6.01	1.89
网格分层抽样	1.23	2.73	4.69	3.06	0.49	7.69	2.16

表3 黄骅市不同抽样方法布设耕地质量监测样点数量在各耕地等别中的分布

Table 3 Distribution of cultivated land quality monitoring sample number by different sampling methods in different cultivated fields in Huanghua

耕地等别	耕地面积/km ²	空间聚类分层抽样样点 布设数量/个	等别分层抽样样点 布设数量/个	简单随机抽样样点 布设数量/个	网格分层抽样样点 布设数量/个
10	593.45	55	42	63	50
11	69.94	10	13	12	10
7	21.71	5	5	1	1
8	8.14	1	3	0	4
9	176.41	10	8	5	16
总计	869.64	81	81	81	81

2019年2月

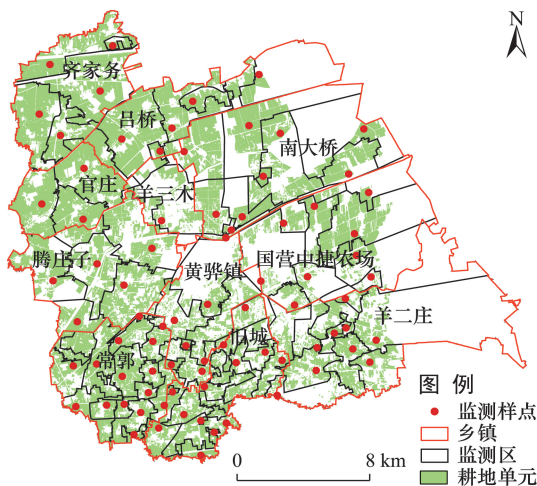


图5 黄骅市耕地质量等别监测优化监测区及耕地质量监测样点分布

Figure 5 Distribution of the cultivated land quality monitoring optimized monitoring zones and points in Huanghua

4 结论

本文以滨海平原区的典型地区——河北省黄骅市为例,提出了一种基于空间聚类分层抽样的耕地质量等别监测样点布设方法,并与传统样点布设方法比较了抽样效率。主要结论为:

(1)由于综合考虑了耕地各项质量属性和空间位置属性,本文采用的空间聚类算法对耕地质量等别监测区的划定效果较好。随着耕地质量等别监测区数目从5增加到100,相同监测区内耕地单元之间在空间位置和耕地质量等别上的差异程度先快速减少,最后保持平稳。

(2)分区数为65的备选耕地质量等别监测区的抽样效率相对较高,被选定为初始监测区方案,在抽样误差为1%的要求下,综合样本容量为77。与初始方案相比,优化后的耕地质量等别监测区方案在确保相同监测区内耕地质量一致性和空间分布连续性更具综合优势。

(3)本文提出的监测样点布设方法与等别分层抽样、简单随机抽样、网格分层抽样等传统抽样方法相比,具有较高的抽样效率,在布设相同数量的监测样点时,各项耕地质量因素的抽样误差均明显较低。

参考文献(References):

[1] 杨建宇,岳彦利,宋海荣,等.基于空间模拟退火算法的耕地质

量布样及优化方法[J]. 农业工程学报, 2015, 31(20): 253-261.

[Yang J Y, Yue Y L, Song H R, et al. Sampling and optimizing methods of cultivated land quality based on spatial simulated annealing algorithm[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(20): 253-261.]

[2] Huffman E, Eilers R G, Padbury G, et al. Canadian agri-environmental indicators related to land quality: Integrating census and biophysical data to estimate soil cover, wind erosion and soil salinity[J]. *Agriculture, Ecosystems and Environment*, 2000, 81(2): 113-123.

[3] Doran J W. Soil health and global sustainability: Translating science into practice[J]. *Agriculture, Ecosystems and Environment*, 2002, 88(2): 119-127.

[4] 姜广辉,赵婷婷,段增强,等.北京山区耕地质量变化及未来趋势模拟[J]. 农业工程学报, 2010, 26(10): 304-311.[Jiang G H, Zhao T T, Duan Z Q, et al. Cultivated land quality change and its future trend modeling in Beijing mountainous area[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2010, 26(10): 304-311.]

[5] 郭文华.耕地保护向数量质量生态并重转变[J]. 国土资源情报, 2012, (12): 35-38.[Guo W H. Cultivated land protection shifts to quantity, quality and ecology[J]. *Land and Resources Information*, 2012, (12): 35-38.]

[6] 吴学敏.武穴市耕地质量监控样点布设研究[D]. 武汉: 华中师范大学, 2015. [Wu X M. Layout Study of the Monitor Samples Points of Land Quality in Wuxue[D]. Huazhong: Huazhong Normal University, 2015.]

[7] 孙亚彬,吴克宁,胡晓涛,等.基于潜力指数组合的耕地质量等级监测布点方法[J]. 农业工程学报, 2013, 29(4): 245-254. [Sun Y B, Wu K N, Hu X T, et al. Layout method for monitoring quality level of arable land based on combination of potential index[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, 29(4): 245-254.]

[8] 殷守强,王鑫,贺文龙,等.耕地质量监测中不同抽样方法的精度比较——以河北省黄骅市为例[J]. 资源科学, 2016, 38(11): 2049-2057. [Yin S Q, Wang X, He W L, et al. Comparison of precision among different sampling methods in cultivated land quality monitoring: A case study of Huanghua City in Hebei Province, China[J]. *Resources Science*, 2016, 38(11): 2049-2057.]

[9] 韩宗伟,黄魏,罗云,等.基于路网的土壤采样布局优化-模拟退火神经网络算法[J]. 应用生态学报, 2015, 26(3): 891-900. [Han Z W, Huang W, Luo Y, et al. Application of simulated annealing method and neural network on optimizing soil sampling schemes based on road distribution[J]. *Chinese Journal of Applied Ecology*, 2015, 26(3): 891-900.]

[10] 余述琼,张蚌蚌,相慧,等.基于因素组合的耕地质量等级监测样点布控方法[J]. 农业工程学报, 2014, 30(24): 288-297. [Yu S Q, Zhang B B, Xiang H, et al. Layout method for monitoring sample point of arable land quality level based on combination of factors[J]. *Transactions of the Chinese Society of Agricultural Engineering*

- neering, 2014, 30(24):288-297.]
- [11] 郭力娜, 张凤荣, 马仁会, 等. 基于标准样地的国家级农用地等别质量监测点设置方法探讨-以冀豫鄂三省为例[J]. 资源科学, 2009, 31(11): 1957-1966. [Guo L N, Zhang F R, Ma R H, et al. Methods for setting up farmland grade monitoring points from standard plots in Hebei, Henan and Hubei provinces[J]. *Resources Science*, 2009, 31(11): 1957-1966.]
- [12] 郭硕, 魏明欢, 简卿, 等. 县域耕地质量监测样点布设研究-以河北省昌黎县为例[J]. 江苏农业科学, 2018, 46(12): 227-232. [Guo S, Wei M H, Jian Q, et al. Study on the distribution of farmland quality monitoring samples in counties-Taking Changli County, Hebei province as an example[J]. *Jiangsu Agricultural Science*, 2018, 46(12): 227-232.]
- [13] 胡晓涛, 吴克宁, 马建辉, 等. 北京市大兴区耕地质量等级监测控制点布设[J]. 资源科学, 2012, 34(10): 1891-1897. [Hu X T, Wu K N, Ma J H, et al. Control points for quality of arable land quality monitoring in Daxing District, Beijing[J]. *Resources Science*, 2012, 34(10): 1891-1897.]
- [14] 杨建宇, 汤赛, 郎文聚, 等. 基于Kriging估计误差的县域耕地等级监测布样方法[J]. 农业工程学报, 2013, (9): 223-230. [Yang J Y, Tang S, Yun W J, et al. Sampling method for monitoring classification of cultivated land in county area based on Kriging estimation error[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2013, (9): 223-230.]
- [15] 王倩, 尚月敏, 冯锐, 等. 基于变异函数的耕地质量等别监测点布设分析-以四川省中江县和北京市大兴区为例[J]. 中国土地科学, 2012, 26(8): 80-86. [Wang Q, Shang Y M, Feng R, et al. Study on location sampling for monitoring the quality of arable land based on variation functions: Case studies in Zhongjiang County and Daxing District[J]. *China Land Sciences*, 2012, 26(8): 80-86.]
- [16] 马建辉. 基于变异函数和等别类型的耕地质量监测样点布设研究[J]. 世界农业, 2015, (5): 111-116. [Ma J H. Based on variograms and other types study on the layout of cultivated land quality monitoring samples[J]. *World Agriculture*, 2015, (5): 111-116.]
- [17] 杨建宇, 岳彦利, 宋海荣, 等. 基于空间平衡法的县域耕地质量监测布样方法[J]. 农业工程学报, 2015, 31(24): 274-280. [Yang J Y, Yue Y L, Song H R, et al. Sampling distribution method for monitoring quality of arable land in county area based on spatial balanced[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(24): 274-280.]
- [18] 祝锦霞, 徐保根, 章琳云. 基于半方差函数与等别的耕地质量监测样点优化布设方法[J]. 农业工程学报, 2015, 31(19): 254-261. [Zhu J X, Xu G B, Zhang L Y. Optimization layout method of monitoring sample points of cultivated land quality based on semi-variance analysis and grade combination[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(19): 254-261.]
- [19] 张玉臻, 孔祥斌, 刘炎, 等. 基于标准样地的省级耕地质量监测样点布设方法-以内蒙古自治区为例[J]. 资源科学, 2016, 38(11): 2037-2048. [Zhang Y Z, Kong X B, Liu Y, et al. Layout method of monitoring samples of cultivated land at the provincial level based on standard plots in Inner Mongolia[J]. *Resources Science*, 2016, 38(11): 2037-2048.]
- [20] 郭力娜, 马仁会, 徐东瑞, 等. 农用地等别质量监测带布设方法探讨-以冀豫鄂三省为例[J]. 资源科学, 2008, 30(8): 1199-1205. [Guo L N, Ma H R, Xu D R, et al. Method of setting up farmland grade monitoring belts in Hebei, Henan and Hubei[J]. *Resources Science*, 2008, 30(8): 1199-1205.]
- [21] 聂艳, 吴西子, 于婧, 等. 基于土地评价和空间聚类的基本农田划定方法研究-以湖北省鹤峰县为例[J]. 中国土地科学, 2013, (12): 39-45. [Nie Y, Wu X Z, Yu J, et al. Primary farmland zoning based on land evaluation and spatial clustering: A case of Hefeng County [J]. *China Land Sciences*, 2013, (12): 39-45.]
- [22] 聂艳, 吴学敏, 何颖, 等. 基于产能核算和空间聚类的县级基本农田划定研究[J]. 长江流域资源与环境, 2014, 23(6): 809-815. [Wu Y, Wu X M, He Y, et al. Demarcation method of prime farmland based on productivity and spatial clustering[J]. *Resources and Environment in the Yangtze Basin*, 2014, 23(6): 809-815.]
- [23] 郝星耀, 潘瑜春, 唐秀美, 等. 基于空间聚类的平原旱作农区土地平整单元区划分方法[J]. 农业工程学报, 2015, 31(5): 301-307. [Hao X Y, Pan Y C, Tang X M, et al. Land leveling partitioning of farming area in arid plain based on spatial clustering[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2015, 31(5): 301-307.]
- [24] 高飞, 樊明辉, 陈崇成, 等. 基于空间实体约束的空间聚类算法研究[J]. 测绘科学, 2008, 33(1): 45-47. [Gao F, Fan M H, Chen C C, et al. The determination of lunar gravity field by mass-point model and its character analysis[J]. *Science of Surveying and Mapping*, 2008, 33(1): 45-47.]
- [25] 李光强, 邓敏, 程涛, 等. 一种基于双重距离的空间聚类方法[J]. 测绘学报, 2008, 37(4): 482-488. [Li G Q, Deng M, Cheng T, et al. A dual distance based spatial clustering method[J]. *Science of Surveying and Mapping*, 2008, 37(4): 482-488.]
- [26] 周翠竹, 朱建军, 石岩. 一种基于双重距离约束的多层次空间聚类方法[J]. 测绘科学, 2014, 39(10): 98-101. [Zhou C Z, Zhu J J, Shi Y. A multi-level spatial clustering method based on dual distance constraints[J]. *Science of Surveying and Mapping*, 2014, 39(10): 98-101.]
- [27] 车丽美, 肖洋, 王甦易, 等. Kmeans聚类分析在形音字表音度中的应用[J]. 计算机技术与发展, 2011, 21(2): 223-225. [Che L M, Xiao Y, Wang S Y, et al. Application of Kmeans clustering analysis in Chinese pronunciation degree[J]. *Computer Technology and Development*, 2011, 21(2): 223-225.]
- [28] 刘卫东, 仲伟周, 石清. 2020年中国能源消费总量预测-基于定基能源消费弹性系数法[J]. 资源科学, 2016, 38(4): 658-664. [Liu W D, Zhong W Z, Shi Q. Forecast of China's total energy consumption in 2020 based on method of fixed based energy consumption elasticity coefficient[J]. *Resources Science*, 2016, 38(4): 658-

664.]

- [29] 杨足膺, 赵媛, 付伍明. 基于弹性系数的江苏省能源生态足迹影响因素分析[J]. 生态学报, 2010, 30(24): 6741–6748. [Yang Z Y,

Zhao Y, Fu W M. Influencing factors' analysis of Jiangsu's energy ecological footprint upon coefficient elasticity[J]. *Acta Ecologica Sinica*, 2010, 30(24): 6741–6748.]

The layout of county area cultivated land quality monitoring samples in Huanghua City based on spatial clustering stratified sampling

YANG Jiangyan¹, YIN Shouqiang², ZHANG Li¹, MEN Mingxin¹, CHEN Ying¹

(1. Institute of Land and Resources, Agricultural University of Hebei, Baoding 071000, China;

2. China University of Mining, Beijing 100083, China)

Abstract: Cultivated land quality monitoring is an important strategy to grasp the changes in cultivated land quality and productivity for the country. Arranging cultivated land quality monitoring samples in a reasonable way can greatly improve the efficiency of cultivated land quality monitoring. This study took the typical area of the coastal plain, Huanghua City as an example, generated 20 alternative monitoring area scenarios with the method of spatial clustering. The initial monitoring area plan was selected by comprehensive comparison of variance, sampling error, sampling efficiency, and sampling elastic coefficient, and by partial optimization, the project of the cultivated land quality monitoring area was finally generated. Finally, based on the stratified sampling area, the stratified sampling method was used to lay out the monitoring records of cultivated land quality. The results showed that under the requirement of 1% sampling error, the alternative cultivated land quality monitoring area with a partition number of 65 did exhibit a comprehensive sample capacity of 77 and a relatively high sampling efficiency, which was selected as the initial monitoring area. The difference in spatial position and cultivated land quality between cultivated land units in the same monitoring area decreased rapidly first, and finally remained stable, with the number of cultivated land quality monitoring areas increases from 5 to 100. By laying the same number of monitoring samples, the sampling error in terms of topsoil texture, section configuration, salinization, organic matter content, drainage conditions, irrigation conditions, and national natural index were 0.37, 1.02, 1.39, 0.91, 0.31, 1.53, and 1.27 respectively. Based on spatial clustering stratified sampling proposed in this paper, the observations were lower than traditional stratified sampling, simple random sampling, grid stratified sampling, and had higher sampling efficiency. The results are intended to provide effective guidance for the deployment of relevant work and research on the quality monitoring samples of cultivated land.

Key words: cultivated land quality; monitoring sample layout; sampling error; sampling efficiency; spatial clustering stratified sampling; Huanghua City