

基于方差四叉树法的滨海盐土电导率采样布局研究*

史舟 李艳 金辉明

(浙江大学环境与资源学院农业遥感和信息技术应用研究所, 杭州 310029)

摘要 利用土壤空间变异特性和空间分布特征进行采样设计是当前土壤采样研究的重要内容。采用方差四叉树法(Variance quad tree method, 简称 VQT), 结合半方差函数, 设计滨海盐土采样的最优布局。并利用普通克里格法对传统网格采样法与方差四叉树采样法所得到的不同的样点数目进行插值, 计算估值误差并进行精度比较。结果发现, 同样的样本数目, 利用方差四叉树法得到的克里格估值误差明显地较利用网格采样法得到的克里格估值误差小, 其采样效率提高约 16% ~ 25%。该方法的优势在于, 可设计在土壤特性变异大的区域密集采样而在变异较小的区域稀疏地采样, 从而在有效表达土壤空间变异性的同时, 提高了采样效率, 减少了采样成本。

关键词 方差四叉树法; 采样; 空间变异; 精确农业; 克里格插值

中图分类号 S159 文献标识码 A

合理的土壤取样设计和精确的室内分析是土壤养分精确管理的前提。如何在满足需求精度的前提下, 用最少的采样点来设计最佳的布点方案, 从而节省野外采样和室内分析的高成本, 是当前土壤取样设计研究的主要目的。目前, 国内外很多研究采用土壤空间变异特性和空间分布特征进行这方面的工作^[1]。如 McBratney 和 Webster 早在 1983 年进行的研究结果就表明运用地统计学的块状克里格最小估计方差所得出的土壤采样效率较传统统计方法高出 3~9 倍^[2]。最近, Ferreyra 等利用地统计学的半方差函数和空间退火模拟算法来设计土壤水的采样策略, 与传统采样模式对比, 这种优化的采样模式可以以较少的样品获得较高的预测精度^[3]。国内也陆续有将地统计学和地理信息系统技术结合起来对一定条件下土壤合理采样数量进行研究^[4~7]。

本研究所采用的方差四叉树法(Variance quad tree method, 简称 VQT)在计算机图像处理方面经常应用, 主要包括图像分割、压缩或编码等^[8,9]。而根据该算法原理, 同时结合地统计的变异函数改进的 VQT 方法, 可以很好地应用于采样布局研究。国外,

就有人曾采用此方法来进行精确农业采样模式的设计^[10]。本研究尝试采用此方法对滨海盐土电导率的采样设计进行应用研究, 并将结果与传统采样方法进行比较, 来检验 VQT 方法设计的采样效率。

1 材料与方法

1.1 研究区域

研究区位于浙江省上虞市西北地区、杭州湾南岸的海涂实验农场。该地属于亚热带区域, 自然植被为常绿阔叶林, 年平均温度 16.5 °C, 年均降水量为 1 300 mm。土壤主要来自于河海相沉积物, 土壤质地以轻壤土或砂壤土为主。土壤类型是粗松咸砂土, 剖面发育不明显, 同一剖面上下层质地较为均匀。近年来该区已被连续垦种。由于围垦措施和种植利用的不同, 反映土壤盐分含量的指标(土壤电导率)的变异相当显著。本研究的地块围垦于 1996 年, 2000 年开垦为棉田, 面积为 10.5 hm²(见图 1)。

1.2 数据获取和采样方法

运用 20 m 网格采样法, 通过一个便携式电导率

* 国家自然科学基金项目(40001008, 40571066)和中德合作项目 BMBF(No. AZ39742)资助

作者简介: 史舟(1970-), 副教授, 博士, 主要从事农业遥感和信息技术、土壤特性空间变异及精确农业的研究。E-mail: shizhou@zju.edu.cn

收稿日期: 2005-10-27; 收到修改稿日期: 2006-03-16

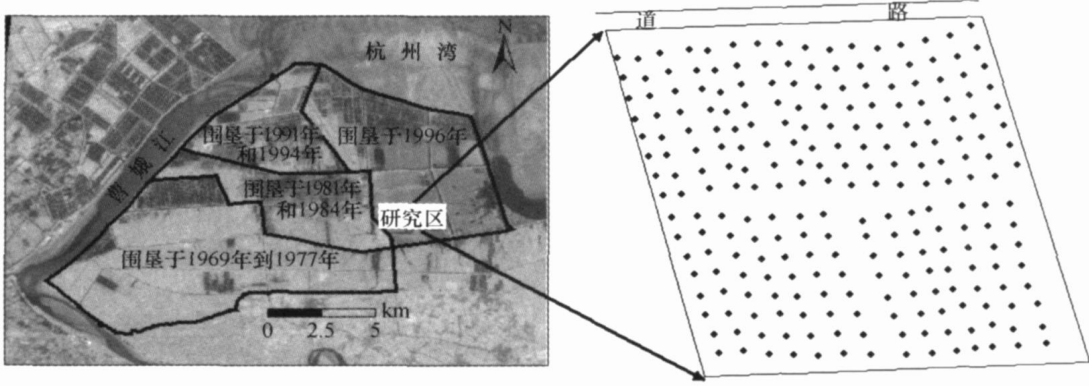


图1 研究区域位置和采样点位布局图

Fig. 1 Location of study site and distribution of sampling points

测定计(WET 土壤水分盐分温度三参数速测仪), 在表土(0~ 20 cm)共测得 240 个土壤体积电导率(Bulk electrical conductivity, 简称 EC_b) 数据。在每个采样点 1 m 的圆周范围内共测定 5 次, 取其平均值作为该样点的 EC_b 值。差分 GPS 被用来进行样点定位。采样日期为 2005 年 4 月 25 日。

1.3 研究方法

研究采用的 VQT 算法, 是对方差较大区域不断地进行四分, 直到方差较大值小于指定的阈值为止。其中的方差用 Φ_h 表示, 其计算公式见下:

$$\Phi_h = \sqrt{n_h \bar{r}(A_h, A_h)} \quad (1)$$

式中, n_h 为区层 h 内离散点对数目; $\bar{r}(A_h, A_h)$ 为区层 h 内半方差的平均值, 它可以利用公式(2)来求得:

$$\bar{r}(A_h, A_h) = \frac{1}{A_h^2} \iint_{A_h} \iint_{A_h} r(x-x') dx dx' \quad (2)$$

并可用公式(3)来近似表示:

$$\begin{aligned} \bar{r}(A_h, A_h) &\approx \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} r(x_i - x_j) \approx \\ &\frac{1}{2n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} [z(x_i) - z(x_j)]^2 \end{aligned} \quad (3)$$

式中, A_h 为区层 h 的面积; $\bar{r}(x_i - x_j)$ 为区层 h 内离散点对 x_i 和 x_j 之间的半方差值; $z(x_i)$ 和 $z(x_j)$ 为变量 x 的实测值。

根据以上推导可得出公式(4):

$$\Phi_h \approx \sqrt{\frac{1}{2} \sum_{i=1}^{n_h} \sum_{j=1}^{n_h} [z(x_i) - z(x_j)]^2} \quad (4)$$

其具体分割算法流程见图 2:

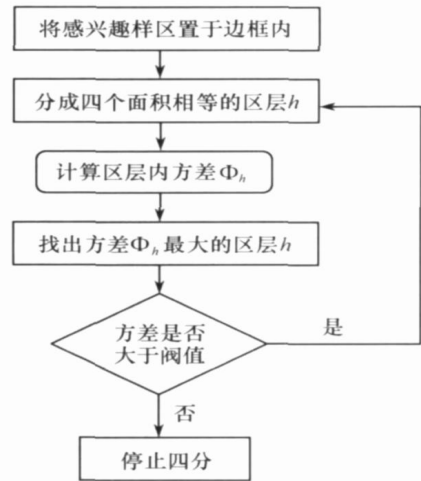


图2 四叉树分割流程示意图

Fig. 2 Flowchart of the variance quad tree method

该算法将感兴趣的研究区进行等方差分割。并认为在分层随机采样中, 如果下列公式成立, 那么对一总体为 n 的样本, 其采样方差为最小。

$$n_h = n \frac{A_h S_h}{\sum A_h S_h} \quad (5)$$

式中, n_h 为区层 h 内的样品数目; n 为总的样品数目; S_h 为区层 h 的标准差; A_h 为区层 h 的面积。

若要在每一区层选择同样数目的样本, 也就是说, 对任意 i 和 j , 都有 $n_i = n_j$ 。那么对于任意 i 和 j , $A_i S_i$ 应等于 $A_j S_j$ 。因而, 该算法是将那些 $A_h S_h$ 值最大的层不断地进行划分, 直到所有层的 A_h 和 S_h 的乘积基本相等为止。若用平均半方差值的方根取代 $A_h S_h$ 中的 S_h 则有:

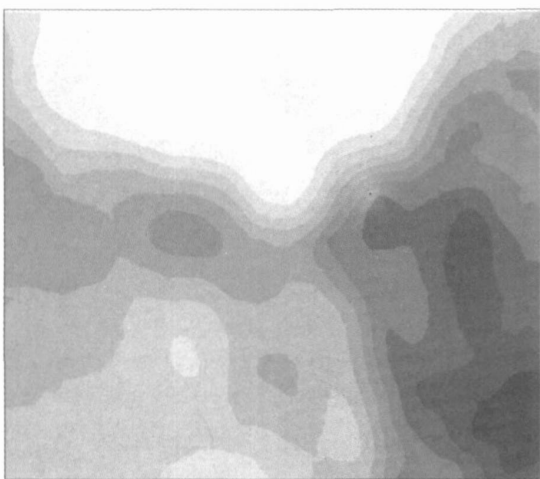
$$A_h S_h \equiv A_h \sqrt{\bar{r}(A_h, A_h)}$$

$$\begin{aligned}
 &= A_h \sqrt{\frac{1}{A_h^2} \iint_A r(x-x') dx dx'} \\
 &\approx n_h \sqrt{\frac{1}{n_h^2} \sum_{i=1}^n \sum_{j=1}^n r(x_i - x_j)} \quad (6) \\
 &= \sqrt{\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [z(x_i) - z(x_j)]^2} \\
 &= \Phi_h
 \end{aligned}$$

这样, 采样就只在每一层内抽取同样数目的随机样品。

2 结果与讨论

首先利用普通克立格法对 2005 年 4 月 25 日采集的土壤电导率数据 (EC_b) 进行插值计算和制图, 并得到其空间分布图(见图 3)。可以看出, 该区的盐分数据具有较大的极差和较高的变异性, 盐分的空间分布表现出东部区域高而北部和西南部区域低的趋势。产生这种现象的主要原因是由于田块围垦改良历史较短, 种植模式和管理措施不一。另外主要是受研究区东部大片鱼塘影响, 鱼塘中大量的地下水入渗到田块东部区, 使之受地下水返盐影响强烈, 即盐分在土壤剖面中随上升水流升到地面, 水分蒸发, 盐分积聚在表层土壤, 导致其表土盐分升高^[11]。西部和北部区较低的含盐量主要由土壤管理措施引起。



图例 EC_b ($mS\ m^{-1}$)

<30	70-90	130-150	190-210
30-50	90-110	150-170	210-230
50-70	110-130	170-190	>230

图 3 由普通克立格法生成的 EC_b 的空间分布图

Fig. 3 Smooth contour map of soil EC_b produced with the ordinary kriging

然后采用 VQT 方法对土壤电导率空间分布图(图 3)进行分层划分计算。图 4 显示随着 VQT 算法的迭代次数的增加, 所找出的具有最大预测方差的区层的方差值。最大方差随着分层迭代次数的增加(也即分层数目的增加)而迅速减少, 随后减少的幅度开始变小。对本研究, 最大的层内方差在达到 52 次划分后开始趋于稳定。该图可以被用来计算所要求的分层数目以确保每个层内包含有理想的方差。

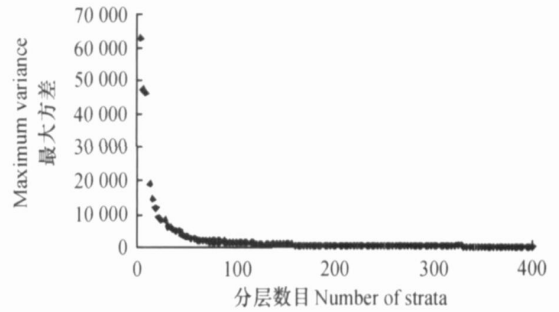


图 4 层内最大方差与分层数之间的关系

Fig.4 Maximum variance per strata as a function of the number of strata

图 5 是将研究区划分为 400 个层的采样设计图。可以看出, VQT 算法可找出那些田间盐分在较小范围上变异最大的区域。这些区域具有较大的局部方差和非连续性, 因而可以被视为独立的管理单元进行特定管理。在采样时, 可以对这些区域加密采样而对那些具有较好连续性的区域稀疏采样, 这样不但确保可以收集到典型的样本, 而且可以节省采样成本, 增加随后的克立格插值预测的精度。

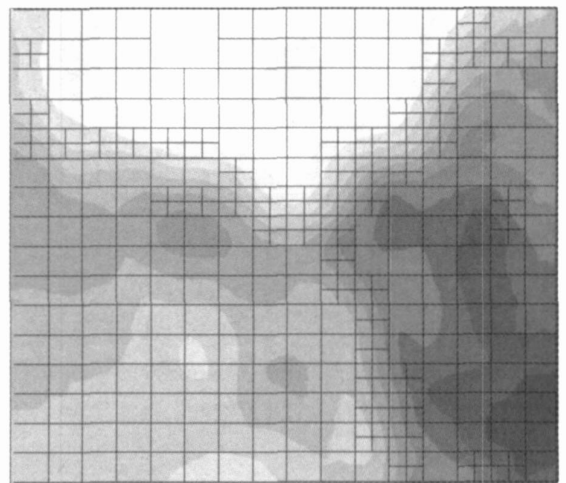


图 5 利用 VQT 算法获得的 400 个层的土壤 EC_b 采样模式

Fig. 5 Sampling schemes for soil EC_b generated with the VQT method of 400 soil strata

VQT 算法适用于对于连续的变量进行采样模式的设计。该算法还可以推广根据作物产量变异来设计作物产量的采样方案, 进而指导那些被认为对作物产量具有较大影响的土壤属性的采样。而且, 由于确保了每个样品都是具有统一方差的采样区块的典型代表样品, 因此所收集的信息较仅仅通过格网采样收集到的信息更有价值。

本研究最初设计的是 $20\text{ m} \times 20\text{ m}$ 的网格采样, 相当于在每 hm^2 采集 24 个样品。根据目前土壤采样和分析, 这样的一个采样模式是需要较高成本的。在国外, 如在澳大利亚, 一般设计的采样模式是在每 hm^2 田块

上收集 10 个样品, 但这样的模式并不适合推广应用。在美国, 目前土壤采样模式以最小 1 hm^2 的格网为单元进行设计, 这样一个方案也是非常粗略的。根据国外学者大量的研究调查, 要得到可信的土壤属性的点位信息, 样点距离应该在 30 m (相当于每 hm^2 10 个样点) 和 80 m (相当于每 hm^2 1.6 个样点) 之间^[12, 13]。这种采样模式在国外的大农场经营可能比较适用, 而对中国以家庭联产承包分散经营的体制, 未必是一种很好的选择。因此研究者必须根据自己的研究区的特点, 结合从研究区获取的关于研究对象空间变异性的一些先验知识来设计最为合理的采样策略。

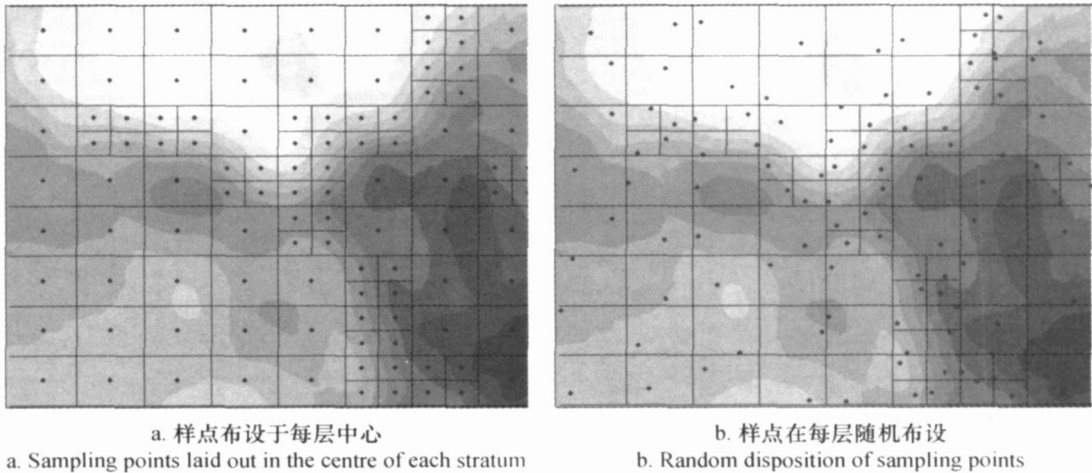


图 6 利用 VQT 算法获得的 105 个层的土壤 EC_b 采样模式

Fig. 6 Sampling schemes for soil EC_b generated with the VQT method of 105 strata

图 6 是利用 VQT 算法将分层数目设为 105 层 (相当于每 hm^2 采 10 点, 即样点间距为 30 m) 所得的采样设计模式。可以看出, VQT 算法在研究田块内找出好几个空间分布变化较大的区块, 从而保证了在这些区块的采样调查在一个较密的尺度上进行。而利用规则的格网采样就不能做到这一点。很明显, 利用一个较少的样品数目, VQT 算法就可以提供较为合理的采样模式, 该模式不像规则网格采样那样具有空间上的一致性, 而是在土壤特性变异大的区域密集采样而在变异较小的区域稀疏采样。图 6a 中, 每个样点的位置处于每层的中心, 当然也可以随机布设, 如图 6b。利用 VQT 算法对得到的 105 个土壤样点 (图 6a) 进行普通克立格插值得到土壤 EC_b 的分布图 (图 7), 可以看到与原始 240 个采样点所得到的土壤 EC_b 普通克立格插值图 (见图 3) 具有相似的空间分布特征, 很好的表现出东部区域 EC_b 值高而北部和西南部区域低的趋势, 而两者的样点数目相差 $1/2$ 多。

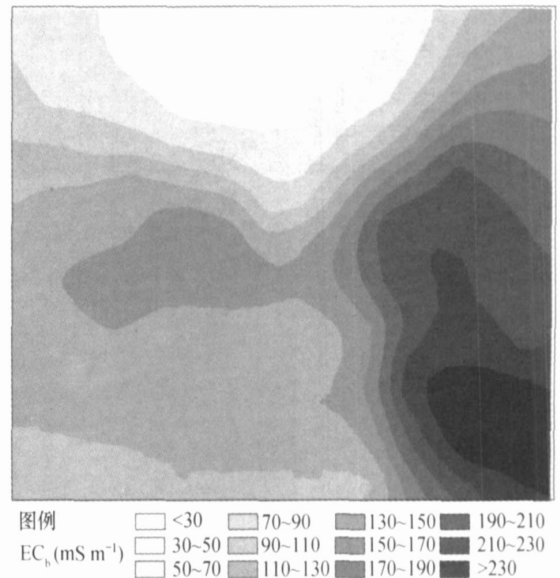


图 7 由 VQT 算法得到的 105 个土壤 EC_b 样点 (图 6a) 进行普通克立格插值

Fig. 7 Interpolation with the ordinary kriging method of soil EC_b of 105 sample points produced with the VQT method (see Fig. 6a)

因此,方差四叉树算法通过利用前一时段上采样得到的先验知识来指导下个时段的采样,进行最优样点的布局和设计,从而可更为有效地对土壤的变异性进行描述,并提高插值计算和预测的精度。

另外,本研究还进一步引入二叉树法。当找出4个面积相等的区层中 Φ_i 最大者,再将其分别纵向和横向分为面积相等的两个区层a和b,然后判断并找出纵向划分和横向划分得到的两个区层的方差之差较大者,对具有方差之差较大的区层再通过纵向和横向划为面积相等的两个区层。并重复该步骤直到达到某一给定的迭代阈值。该方法由于是先进行方差四叉树然后再进行二叉树分层,因此称之为方差四叉-二叉树法。图8给出了利用方差四叉-二叉树算法设计得到的土壤 EC_b 采样模式图。之所以要先进行方差四叉树分层再进行二叉树分层,主要是考虑到,当分层到一定程度时,进行二叉树法,可以减少分层的数目,并进而减少采样数目和成本。该方法在实践中需要进一步的验证。

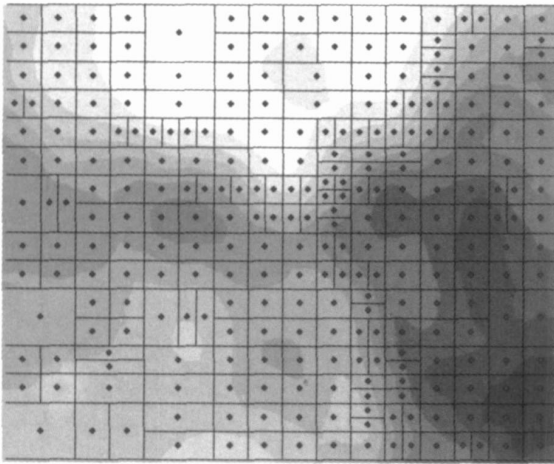


图8 利用方差四叉-二叉树算法设计的土壤 EC_b 采样模式
Fig. 8 Sampling schemes for soil EC_b designed with the integrated variance quad tree and two tree method

3 精度比较

利用传统的网格法和VQT算法对海涂土壤 EC_b 进行采样设计,对两种采样设计所得到的不同的样点数目,分别进行克里格插值并计算插值的标准误差,并将不同取样数目和估算的标准误差关系绘制成图(图9)。可以看出,利用方差四叉树设计得到的采样模式比利用传统的网格采样模式具有明显优势,对同样的样本数目,根据方差四叉树法得到的克

立格估算误差明显地比利用网格采样法得到的克里格估算误差小,也即要达到同样的精度,方差四叉树法需要的样本数量要小。假设图中的虚线是预测所能允许的最大标准误差(40 mS m^{-1}),则对四叉树采样法,达到该精度需要约90个样点,而对于网格法,则需要135个样点。在本研究中,应用VQT算法设计的采样方案进行克里格估算时,其采样效率比网格法提高约16%~25%。

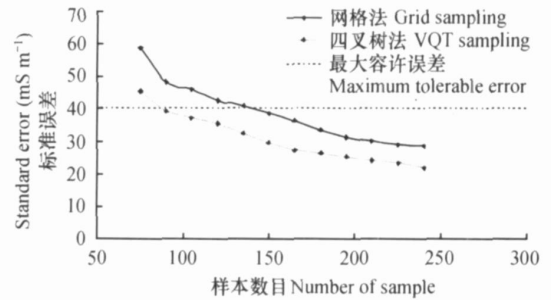


图9 VQT和网格采样的克里格估值标准误差与采样数目的关系

Fig. 9 Kriging standard errors of the VQT and grid sampling methods as a function of the number of samples

4 结论

方差四叉树算法(VQT)提供了一种方式来将一个连续的变量按方差分割从而得到具有等方差(等变异性)的区层。可通过对变异较大的区域增加采样密度,来提供比传统规则网格采样法更为高效的采样布点方案。本研究的结果表明应用VQT算法设计的采样方案进行克里格估算时,其采样效率比传统网格法提高约16%~25%。

在一定的采样精度要求下,该方法有望为多个变量设计一种较传统规则格网法更为有效的采样方案,从而可使研究变量的变异性得到更好的描述。同时它还可以推广应用于作物产量采样方案的设计,为精准农业的土壤和作物网格采样设计提供借鉴。

但是在实际应用中,该方法的最大限制在于它必须要有研究区土壤特性的先验空间分布信息,所以它不适用于空间分布特征未知的田块的首次采样设计。该方法适用于诸如实施精确农业的田块或长期监测的农田区等需要周期性采样的地方。这样随着田间土壤空间分布特性信息的不断积累,采用本方法进行采样设计,效率会越来越高,更能体现其优势。本研究对象为小范围田间尺度,如果应用于大范围采样设计,是否会受到其他要素如地形、水系等

的影响, 需要进一步的研究。

参 考 文 献

- [1] 李艳, 史舟, 徐建明, 等. 地统计学在土壤科学中的应用及展望. 水土保持学报, 2003, 17(1): 178~ 182. Li Y, Shi Z, Xu J M, *et al.* Utilization and perspective of geostatistics in soil science (In Chinese). Journal of Soil and Water Conservation, 2003, 17(1): 178~ 182
- [2] McBratney A B, Webster R. How many observations are needed for regional estimation of soil properties? Soil Science, 1983, 135: 177~ 183
- [3] Ferreira R A, Apeztegui A H P, Sereno R, *et al.* Reduction of soil water spatial sampling density using scaled semivariograms and simulated annealing. Geoderma, 2002, 110: 265~ 289
- [4] 姜城, 杨俐苹, 金继运, 等. 土壤养分变异与合理取样数量. 植物营养与肥料学报, 2001, 7(3): 262~ 270. Jiang C, Yang L P, Jin J Y, *et al.* Soil nutrients variability and rational sampling quantity (In Chinese). Plant Nutrient and Fertilizer Science, 2001, 7(3): 262~ 270
- [5] 齐文虎, 谢高地, 丁贤忠. 精准农业土壤采样密度研究——以上海精准农业试验示范基地为例. 中国生态农业学报, 2003, 11(1): 48~ 52. Qi W H, Xie G D, Ding X Z. Study on soil sampling for precision agriculture—A case study from Shanghai Experimental Demonstration Base (In Chinese). Chinese Journal of Eco Agriculture, 2003, 11(1): 48~ 52
- [6] 王秀, 赵春江, 孟志军, 等. 精确农业土壤采样栅格划分方法的研究. 土壤学报, 2005, 42(2): 199~ 205. Wang X, Zhao C J, Meng Z J, *et al.* Filed soil sampling grids for precision agriculture (In Chinese). Acta Pedologica Sinica, 2005, 42(2): 199~ 205
- [7] Shi Z, Wang K, Baily J S. Sampling strategies for soil available K and P at field scale. Pedosphere, 2000, 20(1): 73~ 76
- [8] 庞全, 苏佳, 段会龙. 基于四叉树和交叉熵的面向对象图像分割方法. 浙江大学学报(工学版), 2004, 38(2): 1 616~ 1 618. Pang Q, Su J, Duan H L. Object oriented image segmentation method based on cross entropy and quad tree (In Chinese). Journal of Zhejiang University (Engineering Science Version), 2004, 38(2): 1 616~ 1 618
- [9] 吴盼密, 龙鹏飞. 基于四叉树和进化算法的分形图像压缩. 微计算机应用, 2005, 26(2): 198~ 200. Wu P M, Long P F. A fractal image compression algorithm based on quad trees and evolution algorithm (In Chinese). Microcomputer Applications, 2005, 26(2): 198~ 200
- [10] McBratney A B, Whelan B M, Walvoort D J J, *et al.* A purposive sampling scheme for precision agriculture. In: Stafford J V. ed. Precision Agriculture '99: Proceedings of the 2nd European Conference on Precision Agriculture. Odense Congress Centre, Denmark, UK: Sheffield Academic Press, 1999. 101~ 110
- [11] 丁能飞, 厉仁安, 董炳荣, 等. 新围砂涂土壤盐分和养分的定位观测及研究. 土壤通报, 2001, 32(2): 57~ 59. Ding N F, Li R A, Dong B R, *et al.* Long-term observation and study on salinity and nutrients of coastal saline soils (In Chinese). Chinese Journal of Soil Science, 2001, 32(2): 57~ 59
- [12] Haneklaus S, Knudsen L, Schnug E. Relationship between potassium and sodium in sugar beet. Communications in Soil Science and Plant Analysis, 1998, 29(11/14): 1 793~ 1 798
- [13] McBratney A B, Odeh I O A. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. Geoderma, 1997, 77(2/4): 85~ 113

SPATIAL SAMPLING SCHEME FOR COASTAL SALINE SOIL BASED ON THE VARIANCE QUAD-TREE (VQT) METHOD

Shi Zhou Li Yan Jin Huiming

(Institute of Agricultural Remote Sensing and Information Technology, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310029, China)

Abstract Sampling designing based on spatial distribution and variability of soil properties is an important issue in recent studies on soil sampling. In the present study, the variance quad-tree method (VQT), combined with the semivariogram function, was used to design an optimal sampling scheme in a field of coastal saline soils. The ordinary kriging method was adopted to interpolate the sampling points fixed with the regular grid method and the VQT method, respectively, and their prediction errors were subsequently calculated for accuracy comparison. Results reveal that with the same size of samples, the VQT sampling scheme was much smaller than the grid one in prediction error, and hence its sampling efficiency was improved by 16% to 25%. The advantage of the VQT method is that this scheme samples sparsely in fields relatively even in soil properties and more intensively in fields varying sharply. Thus it not only characterizes variability within the sampling area, but also improves sampling efficiency, and hence saves sampling cost.

Key words Variance quad-tree method (VQT); Sampling; Spatial variability; Precision agriculture; Kriging interpolation