

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

齐文虎 谢高地 丁贤忠

(中国科学院地理科学与资源研究所 北京 100101)

摘要 通过土壤密集采样识别农田土壤物理生化性状的空间差异,构建土壤与农作物产量关系模型,用间接方法提高土壤性状信息的分辨率并定量评价这些信息的可靠性非常必要。在评价多种插值方法的基础上,对中国科学院上海精准农业示范基地每公顷 3 个样本、等距离系统抽样的 280 个土壤有机质样本采用克里格方法进行插值,估算样本之间土壤有机质含量和分析估算的可靠性,并与距离倒数函数插值法进行对比。对几个减少样本数量的方案进行分析,提出在满足精准水稻种植管理分区要求前提下可减少 1/2 原有土壤采样样本,并给出适于水稻精准种植土壤采样设计的逐步优化方法。

关键词 土壤采样 克里格插值 优化设计

Study on soil sampling for precision agriculture-A case study from Shanghai Experimental Demonstration Base. QI Wen-Hu, XIE Gao-Di, DING Xian-Zhong(Institute of Geographical Sciences and Natural Resources, Chinese Academy of Sciences, Beijing 100101), *CJEA*, 2003, 11(1):48~52

Abstract It is significant to explore the spatial variation of soil biochemistry characteristic by densely soil sampling, to construct the mathematical models for revealing the relation between crop yield and its environment for precision agriculture, to raise the resolution of soil maps through indirect methods and evaluate its reliability. Based on the evaluation of interpolation methods, this paper adopts Kriging interpolator to produce soil map of organic matter in Shanghai Base, using 280 soil samples of systematic sampling of 3 samples per hectare. The comparison between Kriging and inverse distance function interpolating is carried out by using the same data set, showing that the Kriging interpolator is better than the latter in this situation. This study also analyzes the scenario of producing soil maps by reducing the amount of samples and gives the relevant evaluations. Based on these analyses, the authors present the procedure of optimal design of soil sampling for precision agriculture of rice growing.

Key words Soil sampling, Kriging interpolation, Optimal design

通过以采样点为基础信息源的空间统计分析,提出理论的优化土壤采样密度,再结合经济合理的考虑确定可行的土壤采样密度,是设计土壤采样方案的行之有效方法之一。

本研究以中国科学院上海精准农业试验示范基地 2 年水稻精准种植土壤采样和分析实践,在分析和评价形成反映主要土壤性状空间分布成图和插值方法的基础上,探讨优化土壤采样设计的有效途径。

1 土壤采样设计原则与样点布设

研究在中国科学院上海精准农业试验示范基地进行,将水稻田块划分成 30m×300m 条田,基本南北向,中间以灌、排水渠和农机行道隔开。为能显现出土壤性状的空间不均匀性,尽可能采用密集采样,并将其作为土壤的本底调查,分析土壤有机质、pH、全 N、速效磷、速效钾等。

在无产量图的情况下,精准农业土壤采样设计原则是根据对采样田块土壤、地貌、植被、土地利用历史等了解,第 1 次采取较稀疏的样本间距,随着精准农业的实施以及对土壤性状空间变化的了解,增加或减小样本的间距,样点选择要考虑使其落在条田横向的不同部位。

样点布设至少有 3 种方案可选择,即随机抽样、分层抽样和机械或系统抽样。本研究采用机械抽样方法,参照仅有的数据,第 1 次本底调查确定为 3 个样点/hm²,样本间距约 57.7m,用 GPS 定位。于 2000 年 5 月采样,为减少偶然性,在每样点混合采样,以样点为中心,在中心和以中心为圆心、5m 为半径的圆周上随机

* 中国科学院知识创新工程重要方向性项目“长江三角洲水稻精准种植技术体系的研究、集成与示范”(KZCX2-412)资助

收稿日期:2002-06-26 改回日期:2002-07-15

取 5 个样本混合后形成 1 个样本,土样采自 20cm 的耕层,样本数量共计 280 个,样点布设见图 1。

2 土壤性状空间插值方法

为能使土壤性状图和产量图匹配,必须采用插值方法估算未采样点的土壤性状值。有多种插值方法可以选择,用 $Z(\mathbf{X})$ 表示在坐标为 \mathbf{X} 的地点某个土壤性状值,如土壤有机质含量值,其中 $\mathbf{X} = (X, Y)$ 是一个矢量;用 $\hat{Z}(\mathbf{X}_0)$ 表示未被采样点的土壤性状值, $Z(\mathbf{X}_i)$ 是采样点的已知土壤性状值。插值是用已知采样点性状值的加权平均计算未被采样点的值,一般可表示为:

$$\hat{Z}(\mathbf{X}_0) = \sum_{i=1}^N \lambda_i Z(\mathbf{X}_i) \quad (1)$$

其中, λ_i 是赋予第 i 个采样点值的权重。常被采用的插值方法有希多森多边形法、三角形法、希泊松自然邻域法、距离倒数函数法、样条函数法等,它们的共同特点是估算为“局地”的,只有与待估算点邻近的采样点才起作用。而趋势面法是用多元回归法构造一个趋势面函数,区域内任何一点的土壤性状值均可由此函数表达,与其他方法不同,每个待估算点的值由所有采样点的值确定。但以上方法无法给出插值误差的估计,而精准农业土壤性状空间差异的反映是小尺度的,在同一田块要逐年连续采样,需精确地估算研究区域各点性状值以及插值估算的误差,以便改进采样方案设计。

本研究采用克立格插值法,克立格插值法是用已知采样点性状值的加权平均估算或预测待估算点的性状值,估计的基本表达式与式(1)相同,此外为保证估算无偏,规定权重必须满足:

$$\sum_{i=1}^N \lambda_i = 1 \quad (2)$$

与估算误差的均值等 0,即:

$$E[\hat{Z}(\mathbf{X}_0) - Z(\mathbf{X}_0)] = 0 \quad (3)$$

克立格方法用 $\hat{Z}(\mathbf{X}_0) - Z(\mathbf{X}_0)$ 之差的方差达到极小来确定权重值,即:

$$\min Var = E[\{\hat{Z}(\mathbf{X}_0) - Z(\mathbf{X}_0)\}^2] = 2 \sum_{i=1}^N \lambda_i \gamma(\mathbf{X}_i, \mathbf{X}_0) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j \gamma(\mathbf{X}_i, \mathbf{X}_j) \quad (4)$$

其中:

$$\gamma(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{2} E[\{\hat{Z}(\mathbf{X}_i) - Z(\mathbf{X}_j)\}^2] \quad (5)$$

一般称之为半方差函数, $\gamma(\mathbf{X}_i, \mathbf{X}_j)$ 的值在各向同性的假定下,只取决于 \mathbf{X}_i 点和 \mathbf{X}_j 点之间的距离 $d = |\mathbf{X}_i - \mathbf{X}_j|$,而与它们的绝对空间位置无关。研究区域根据已有采样点的值和它们在区域内坐标位置,绘出 $\gamma(\mathbf{X}_i, \mathbf{X}_j)$ 随 \mathbf{X}_i 点和 \mathbf{X}_j 点之间距离 d 变化的曲线(见图 2), \mathbf{X}_i 点和 \mathbf{X}_j 点之间距离越大,其半方差的值就越大,说明它们的距离越远,其相关程度就越低,这为只用与未知点邻近的少数采样点的插值方法提供了依据。据式(4)求解权重时,需要计算 $\gamma(\mathbf{X}_i, \mathbf{X}_j)$ 和 $\gamma(\mathbf{X}_i, \mathbf{X}_0)$,其中 \mathbf{X}_0 是待估算点的坐标,而此处的 $Z(\mathbf{X}_0)$ 是未知的,因此需通过已知的 $\gamma(\mathbf{X}_i, \mathbf{X}_j)$, $i = 1, 2, \dots, N$, 构建半方差函数 $\gamma(d)$ 。图 2 中方块是已知的半方差值,曲线为拟合结果。根据已知半方差值随 d 的变化,选用指数函数进行拟合:

$$\gamma(d) = C_0 + C[1 - \exp(-d/A_0)] \quad (6)$$

拟合的残差平方和 $RSS = 9.5E - 05$,是待选的线性、球面、圆等

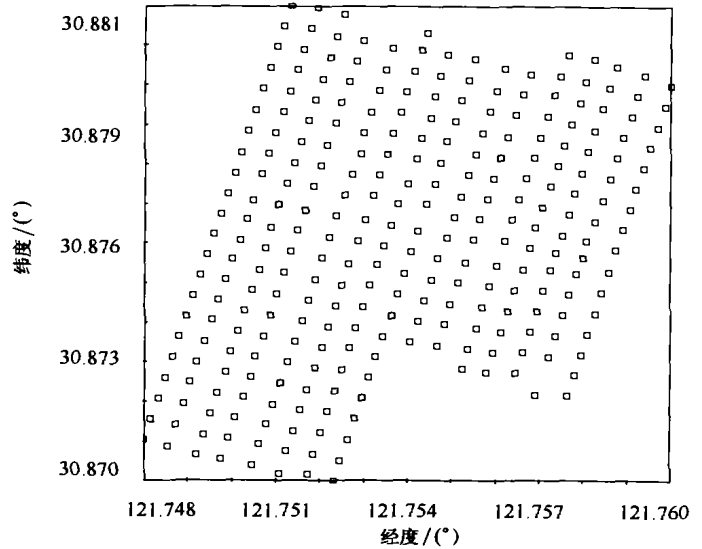


图 1 土壤采样点分布

Fig.1 The distribution of soil sampling

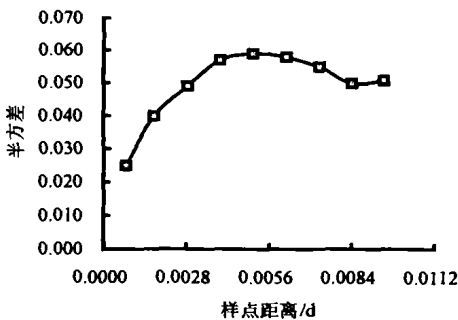


图 2 半方差与 d 之间关系曲线

Fig.2 The relation between $\gamma(\mathbf{X}_i, \mathbf{X}_j)$ and d

模型中最小的,即根据实际数据选用指数函数拟合的效果最好。

3 用不同插值方法分析土壤采样密度

3.1 不同插值方法成图效果的比较

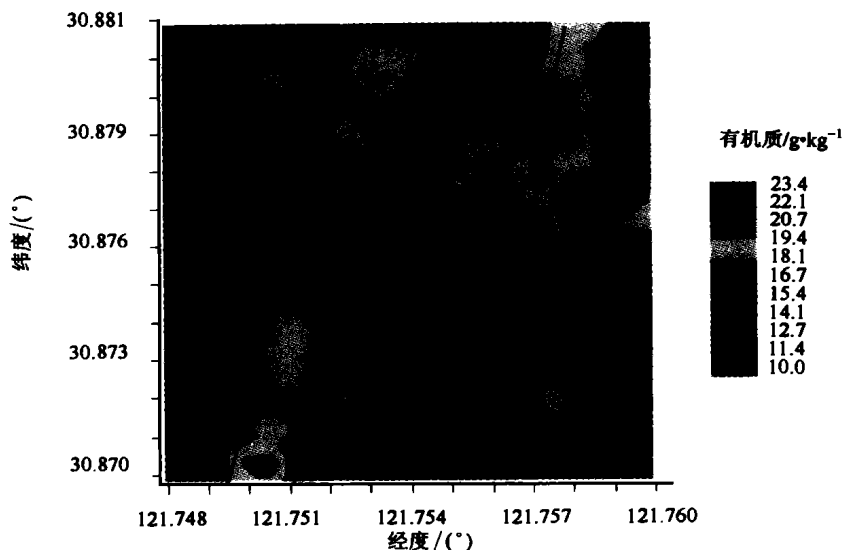


图 3 土壤有机质含量图(2000 年)

Fig. 3 Spatial distribution of organic matter content of soil in 2000

图 3 是利用 2000 年 280 个样点数据进行克里格(Kriging)插值的结果。精准农业土壤图的作用是根据土壤性状空间差异划分变量管理小区,小区划分的多少取决于按小区施肥等管理策略,本研究将土壤有机质含量图分为 11 个等级(它们可以合并成更少的等级或划分更细)。图 3 中用直线连成的线框表明试验田块东北角有机质含量最高,这可能是因靠近猪场,土壤中有有机肥料多的缘故;地处偏西部、沿条田方向的区域有机质含量亦较高,这可能与这些条田倒茬种植蔬菜有关,中间有一土壤有机质含量最低地块,在农田管理决策时需特别注意。在土壤图中有些小的斑块有机质含量较高,易被掩盖在某个小区内,或

对这些孤立斑块作特殊处理,这取决于管理决策系统。

图 4 是 2001 年在与 2000 年相同区域、相同数量样本和采样点进行重复采样分析数据所作土壤有机质含量图,插值方法亦相同。图 4 表明,2001 年与 2000 年基本结构框架类似,其主要不同点是在东北部沿条田方向出现 1 条土壤有机质含量最低小区,且 2001 年整体数值比 2000 年偏高,这可能是因 2 年土壤样本在不同实验室的分析方法不同所致。新出现的低有机质含量小区需要进一步分析其产生的原因,同时在实施变量农田管理决策时应特别注意。2 年土壤图对比表明,土壤有机质的空间分布基本稳定,土壤有机质含量的改变尚需较长的时间过程。但土壤有机质含量的空间分布有细微差异,除因施肥和作物生长吸收土壤养分的必然因素外,还有土壤自身生化过程的不确定性因素和采样随机性因素,加之作物产量与土壤养分之间关系模型尚未达到精确模型的程度,这些在利用土壤图进行管理小区划分时均应考虑。

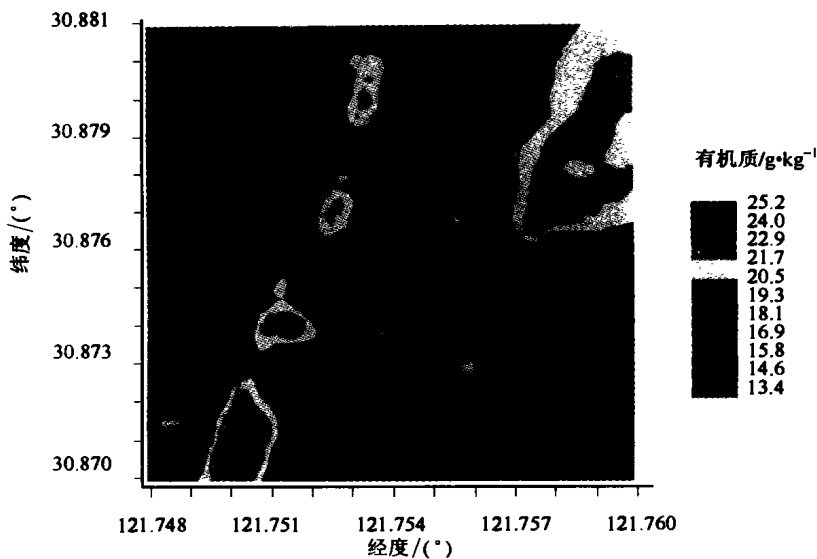


图 4 土壤有机质含量图(2001 年)

Fig. 4 Spatial distribution of organic matter content of soil in 2001

图 5 是利用同样的样本数据,采用距离倒数函数法插值所得土壤有机质含量图(与图 3 和图 4 同样均只用待估算点邻近的 16 个样本点数据),用克里格法能较恰当地反映出土壤性状空间分布的结构和局地变化的特征,而用距离倒数函数法插值时其结果对函数中距离的方指数 p 的选择很敏感, p 小时插值结果过于粗糙, p 大时则过分反映细微变化, p 值的确定主要是经验性的,还难以从理论上确定 p 值大小。

3.2 减少采样密度后成图效果的比较

为探讨土壤采样密度的合理性和采样设计优化过程,本研究以 2000 年土壤有机质含量图为例,首先将

样本数量减少,隔列去掉样本,使样本横向(与条田方向垂直)间隔变为 $2\text{m} \times 57.5\text{m}$,纵向间隔不变,即只有 147 个样本参与插值,仍用克里格插值法,在形成半方差函数 $\gamma(d)$ 后同样用待估算点邻近的 16 个样本点插值,形成土壤有机质含量图(见图 6),它基本保持用 280 个样本插值的土壤图空间结构,仅在细微尺度上有所变化,用此土壤图作划分变量管理小区的依据与用全部样本所作土壤图(见图 3)为依据的结果基本类似。图 7 是在进一步减少样本数量(在隔列去掉所有样本基础上再在每列中间隔去掉样本),只剩下 74

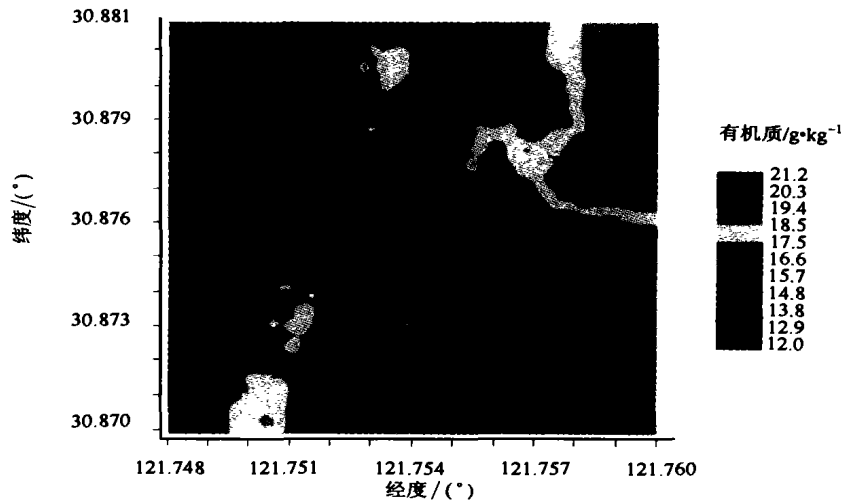


图 5 距离倒数函数插值法土壤有机质含量图(2000 年)

Fig. 5 Map of organic matter content of soil with IDW in 2000

个样本参与插值,这时样本的间距成为横纵向均为 $2\text{m} \times 57.7\text{m}$,该图已失去土壤有机质含量空间分布以条田方向为主线的结构特征,已完全不适用于作为农作管理小区划分依据。究其原因一是样本总数偏少,因此形成的 $\gamma(d)$ 函数与实际值差异过大;二是样本间隔过大,失去土壤有机质含量空间分布的横向变化突出的结构特征。

基于上述 2 次减少用于插值样本的结果分析表明,土壤有机质含量在整个研究区域内宏观结构是按条田方向分布,次一级的结构则呈条田内从南至北阶段性变化,再次一级结构则是个别的小范围的变化,这些细微变化在划分

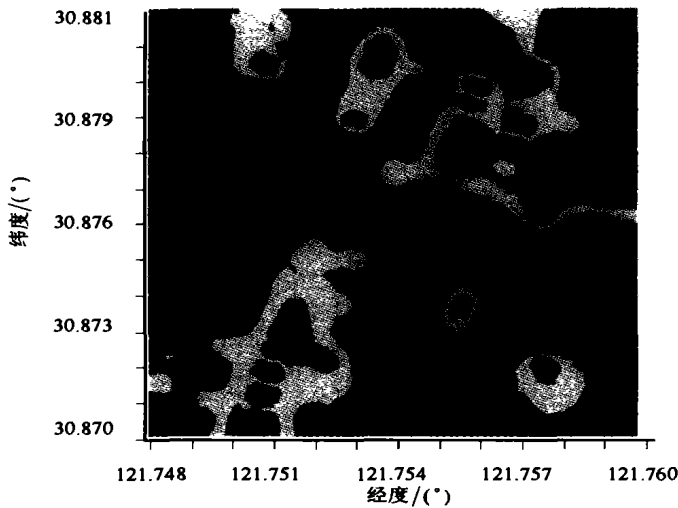


图 6 用隔列采样点插值土壤有机质图(2000 年)

Fig. 6 Map of organic matter content of soil with Kriging interpolation from 147 soil samples in 2000

农作管理小区时可以忽略,如有必要可采取特殊措施处理。

4 小 结

本研究样本布设方法理论正确,可在较大范围连片的水稻精准种植区应用。上海基地样点空间位置布设和样本数量的逐步优化可在现有基础上加大样点列(沿条田方向)之间的距离,其范围可以加大到 57.7m 至 $2\text{m} \times 57.5\text{m}$ 之间,样本列内样点间的距离保持仍为 57.7m 。经过几次土壤采样实践后,结合产量图分析和变量农作实施的分辨率要求,考虑可否适当加大样本列方向的样点间距,但应小于加大的列间距。该基地在经济成本许可条件下,最好保

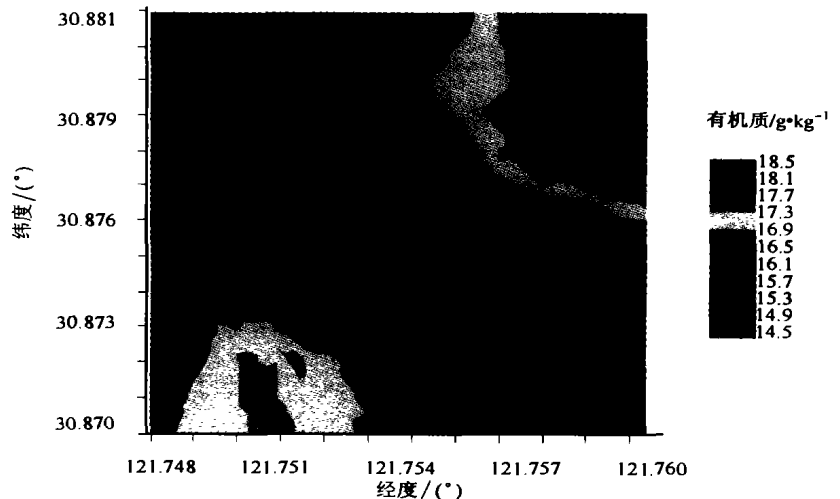


图 7 用隔行隔列采样点插值土壤有机质图(2000 年)

Fig. 7 Map of organic matter content of soil with Kriging interpolation from 74 soil samples in 2000

持至少 5 年 $57.7\text{m} \times 57.7\text{m}$ 的样本布设方案,以便于跟踪研究。为降低土壤采样成本,可减少不重要的要素分析。

参 考 文 献

- 1 齐文虎,石玉林. 计算机控制农业. 资源科学,1998 (3):23~27
- 2 Webster R., Oliver M. Geostatistics for Environmental Scientists, New York. 2001
- 3 Lamb J. A., Rehm G. W., *et al.* Grid Cell Size Needed for Sugar Beets Nitrogen Recommendations in Southern Minnesota: Nutrient Maps and Root Yields, in Proceedings of 5 th International Conference on Precision Agriculture, Minnesota, USA. 2000
- 4 Kravchenko A. N., Bullock D. G. Comparison of Interpolation Methods for Mapping Soil P and K Contents, in Proceedings of 4 th International Conference on Precision Agriculture, Minnesota, USA. 1998