Guided Soil Sampling for Enhanced Analysis of Georeferenced Sensor-Based Data

V.I. Adamchuk[1], D.B. Marx[2], A.T. Kerby[2], A.K. Samal[3], L.K. Soh[3]
R.B Ferguson[4], C.S. Wortmann[4]

[1]Department of Biological Systems Engineering, University of Nebraska-Lincoln
203 LW Chase Hall, Lincoln, NE 68583-0726, USA
Telephone: (+1 402) 472-8431
Fax: (+1 402) 472-6338
Email: vadamchuk2@unl.edu

[2]Department of Statistics, University of Nebraska-Lincoln

[3]Department of Computer Science and Engineering, University of Nebraska-Lincoln

[4]Department of Agronomy and Horticulture, University of Nebraska-Lincoln

1. Introduction

In recent years, increasingly popular precision agriculture technology has become one of primary beneficiaries of geocomputation methods. Modern farmers and their advisors have to deal with multiple field-scale data layers representing spatially variable growing environments. Varying the rate of fertilizers and other agricultural inputs according to local needs is a promising step toward sustainability of profitable and environmentally protective crop production. However, inappropriate prescribed variable rates have diminished the effectiveness and, therefore, the profitability of this management strategy. In part, this is because low-density (usually 1 ha) soil sampling and laboratory analysis has been the most widely used method to determine local needs throughout a field. This method is insufficient to accurately describe the spatial variability of yield limiting soil properties. Increasing the density of conventional soil sampling is expensive. However, numerous high-density data layers (e.g., yield maps, aerial images, sensor data, etc.) can now be obtained at relatively low cost (McBratney et al., 2005). Several studies have verified that proper analysis of these spatial data layers can improve the accuracy of prescription maps while reducing the need for laboratory analyses of the extracted soil samples (Thompson et al., 2004). Unfortunately, existing commercial software applications do not facilitate a straight-forward process of multi-layer data analysis and interpretation.

Therefore, the ultimate goal of our work is to formulate a comprehensive methodology for delineation of field areas with the greatest potential for differentiated management using on-the-go soil sensor technology. Using high-density georeferenced measurements obtained while moving across a field can reveal the spatial variability of soil characteristics in great detail (Adamchuk et al., 2004). However, most of the measured properties do not directly relate to a particular agronomic characteristic commonly used to prescribe the management of agricultural inputs. Thus, many service providers started using sensor-based maps to determine field areas significantly different from the rest of the field. They then sample these areas to determine whether a differentiated soil treatment could be beneficial.

Based on common practices used today, the process of locating these guided samples is subjective. Service providers simply display the maps and manually identify points for future sampling. Obviously, inexperienced professionals frequently fail to succeed while using this approach. As a result, systematic grid sampling remains dominant.

The objective of this publication is to initiate discussion on the proper prescription of guided soil sampling to supplement sensor-based soil maps with a limited number of verifiable laboratory measurements.

2. Materials and Methods

While working with leading experts that practice guided soil sampling, the following three rules have become apparent. Guided soil samples should: 1) be collected from relatively homogeneous field areas (away from the field boundary and away from locations where sensor data changes significantly over short distance intervals), 2) uniformly cover the entire range of sensor-based measurements (especially the highest and the lowest spots), and 3) be spread across the entire field to assure representation of different soil conditions unaccounted for by sensor measurements (e.g., soil types, terrain, profile structure, past management, etc.). Mathematically, these three criteria can be represented using the following three statistics respectively: 1) neighbourhood variability, 2) D-optimality criteria for a potential regression analysis, and 3) spatial predictability criterion.

A map of 598 soil pH measurements obtained using Mobile Sensor Platform (Veris Technologies, Inc., Salina, Kansas, USA) from a 23-ha field in Kansas was used to prescribe 10 guided soil sampling locations. Each sample was assumed to be co-located with a sensor-based measurement. A total of 63 different sets of guided soil samples were compared. The sets included: 1) 20 different combinations of 10 randomly selected points, 2) 19 different combinations of points randomly selected from 10 rectangular sections of the field (grid cells), 3) 20 different combination of points randomly selected from 10 equal intervals of on-the-go soil pH measurements (Figure 1), 4) the set of points with the highest similarity to the nearest existing neighbours in each of four directions, 5) the set of points located in the centre of each grid cell (20th combination of points evenly spread across the field), and 6) two subjective sets of guided soil samples selected using the general rules outlined above. Examples of guided sample combinations obtained according to these strategies are shown in Figure 2.

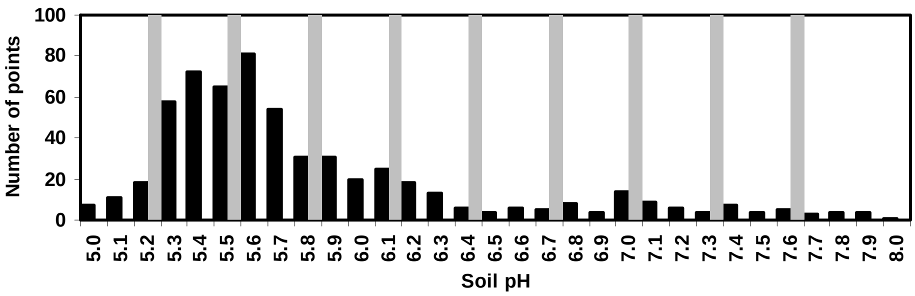

Figure 1. Distribution of on-the-go soil pH measurements separated in 10 intervals.

The quantitative evaluation criteria were: 1) mean squared difference (MSD) between pH measurement at a given point and the nearest existing neighbour in each of four directions (homogeneity criterion), 2) D-optimality of pH distribution corresponding to each subset of 10 points (D-optimality criterion), and 3) maximum standard error of a kriged map produced using each subset of 10 points (field spread criterion). To compare the overall performance of each set of 10 guided sampling locations, every criterion was standardized and the corresponding z-scores were added. The lower the score, the higher was the degree of a given criterion satisfaction.

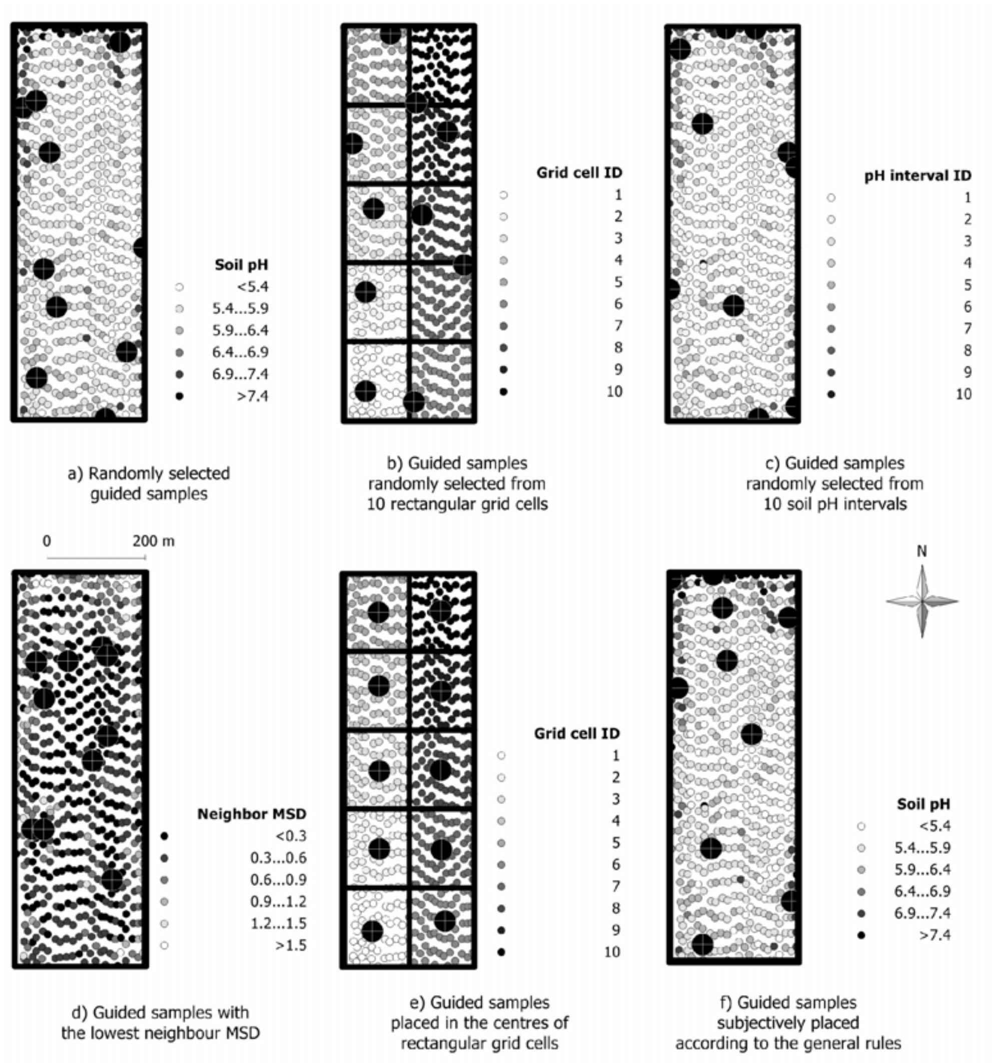

Figure 2. Example distributions of 10 guided samples identified according to six different specified rules.

3. Results and Discussion

Different guided sampling definition strategies were superior when satisfying the corresponding criteria. Thus, guided samples with the lowest neighbour MSD (Figure 2d) produced the lowest z-score for the homogeneity criterion. D-optimality criterion was best satisfied when guided points were evenly spread across the entire range of soil pH measurements (example in Figure 2c). The even field spread through rectangular grid cells yielded the lowest z-scores for the field spread criterion (example in Figure 2b). However, this came at the expense of other criteria that were not purposely addressed, which affected the total z-score and caused it to be similar or higher than the total z-score for the random selection of guided soil samples (example in Figure 2a).

On the other hand, the comprehensive subjective strategy (example in Figure 2f) was aimed at reducing the total z-score, which appeared to be lower than the means corresponding to other strategies. However, in every case where randomization was involved, there was at least one combination resulting in yet a lower total z-score. Furthermore, placing sampling points in the centre of each grid (Figure 2e) also revealed a rather satisfactory performance, which is obviously misleading. It looks as if equal weighting of all three criteria may not be the most suitable. From a practical viewpoint, the D-optimality and homogeneity criteria tend to be more critical than the field spread criterion when it comes to the analysis of sensor data.

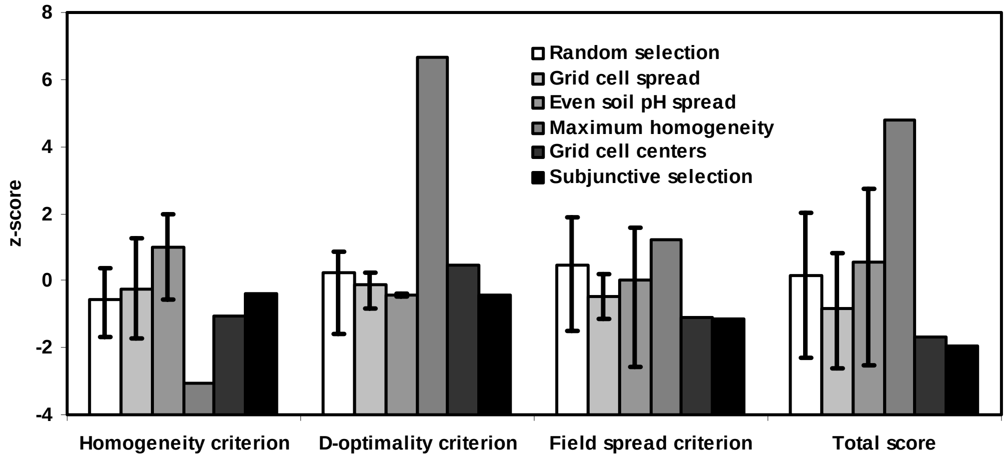

Figure 3. Comparing z-scores for six different guided sampling prescription strategies (error bars indicate extreme values observed through the randomization process).

In our continued work, we intend to consider a more involved analysis of the three criteria as they affect the ability of the guided samples to address the need for sensor calibration sustainable across the entire field. Also, it is necessary to question the number of guided sampling points desirable for a specific site. Industry's tendency to use multilayer sensor inputs brings an additional quest to engage two or more dense data layers when identifying the field locations most favourable for in-depth analysis.

4. Conclusions

With this work, we initiated the development of a comprehensive methodology for the analysis of georeferenced sensor-based soil measurements. As the first step, we quantified three different criteria commonly considered by practitioners when selecting points for additional investigations to supplement sensor-based measurements. These points should be spread across the entire field area, represent the entire range of sensor measurements, and be placed in the most homogeneous areas of the field. The fact that certain randomly selected sets of 10 guided samples produced lower z-scores than the subjective strategy indicates the possibility of using one of the optimization methods to identify guided sampling locations that would minimize an overall objective function.

5. Acknowledgement

This publication is a contribution of the University of Nebraska Agricultural Research Division, supported in part by funds provided through Hatch Act. Additional support was provided through the Channing B. and Katherine W. Baker Fund #3424 of the University of Nebraska-Lincoln (Lincoln, Nebraska, USA).

6. References

Adamchuk VI, Hummel JW, Morgan MT and Upadhyaya SK, 2004, On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, 44(1):71-91.

McBratney A, Whelan B and Ancev T, 2005, Future directions of precision agriculture. *Precision Agriculture*, 6:7-27.

Thompson AN, Shaw JN, Mask PL, Touchton JT and Rickman D, 2004, Soil sampling techniques for Alabama, USA grain fields. *Precision Agriculture*, 5:345-358.