*Ninth International Conference on Precision Agriculture*
*July 21, 2008*

# Enhancement of On-the-Go Soil Sensor Data Using Guided Sampling

*Viacheslav I. Adamchuk*
*Biological Systems Engineering*
*University of Nebraska - Lincoln*

*Raphael A. Viscarra Rossel*
*CSIRO Land and Water*
*Canberra, Australia*

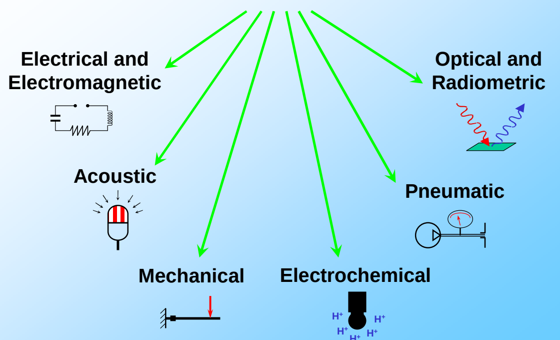
*David B. Marx*
*Statistics*
*University of Nebraska - Lincoln*

*Ashok K. Samal*
*Computer Science and Engineering*
*University of Nebraska - Lincoln*

## Background

- Assessment of soil variability is essential in site-specific management
- Variable rate application requires accurate information about soil spatial structure
- Obtaining adequate spatial information for a field is expensive using conventional soil sampling and analysis methods
- Accurate mapping of soil attributes requires high density on-the-go sampling
- Recent on-the-go sensors can reveal spatial variation in soils, but improved prescription algorithm are needed

## On-the-go Soil Sensors




## Applicability of On-the-Go Soil Sensors

| Soil property | | | | | |
|---|---|---|---|---|---|
| Soil texture (clay, silt and sand) | Good | OK | | Some | |
| Soil organic matter or total carbon | Some | Good | | | |
| Soil water (moisture) | Good | Good | | | |
| Soil salinity (sodium) | OK | | | | Some |
| Soil compaction (bulk density) | | | Good | Some | |
| Depth variability (hard pan) | Some | | OK | Some | |
| Soil pH | | Some | | | Good |
| Residual nitrate (total nitrogen) | Some | Some | | | OK |
| Other nutrients (potassium) | | Some | | | OK |
| CEC (other buffer indicators) | OK | OK | | | |

## Guided (Targeted) Sampling

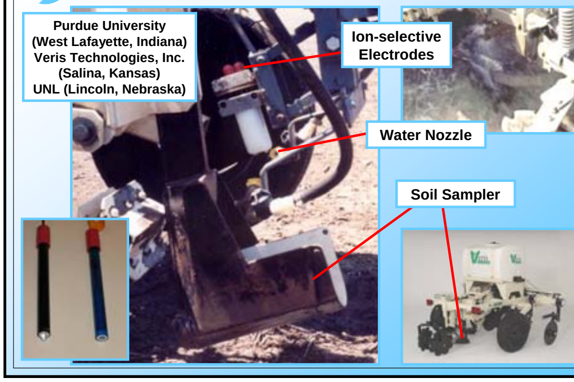
- Prescription rules:
  1) Cover the entire range of data from each source
  2) Avoid field boundaries and other transition zones
  3) Spread samples over the entire field
- Current difficulties:
  1) Poor ability to simultaneously consider multiple data layers
  2) Uncertain number of needed guided samples
  3) Difficult validation and comparison of a sampling scheme with alternatives

## Objectives

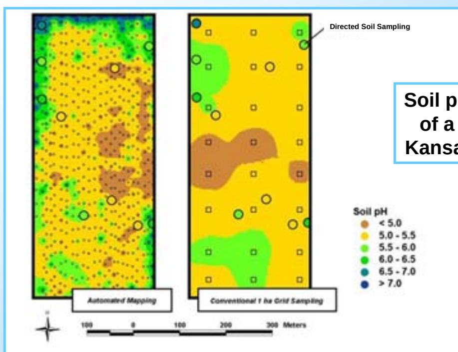
- Develop a set of criteria that may be used to compare alternative sampling schemes
- Evaluate a number of different sampling schemes
- Two different sensors with relatively low correlation between their outputs have been used to test the proposed methodology

## Example: Mobil Sensor Platform (MSP)


## Direct Soil Measurement

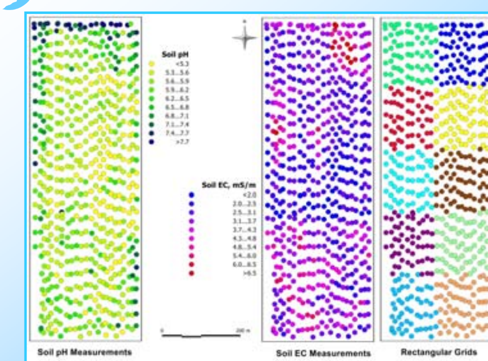



## Soil pH Maps




Soil pH Maps of a 23-ha Kansas Field

## Maps of Soil pH and EC




## Spatial Separation (S-optimality)

$$S_{opt} = \frac{N(N-1)}{2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\frac{1}{\sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}}$$

- N is the number of guided samples (N = 10)
- x and y are the spatial coordinates for $i^{th}$ and $j^{th}$ locations

## Data Spread (D-optimality)

$$D_{opt} = |Z'Z|$$

$$Z = \begin{bmatrix} 1 & z_1 \\ 1 & z_2 \\ \vdots & \vdots \\ 1 & z_N \end{bmatrix}$$

- $z_i$ is the value of pH or EC for $i^{th}$ measurement

## Local Homogeneity (H-criterion)

$$H_{cr} = 1 - \frac{\sum_{i=1}^{N} \sum_{j=1}^{n_i} \left(z_i - z_j\right)^2}{\sum_{i=1}^{N} n_i \cdot H_{\max}}$$

- $n_i$ is the number of existing nearest neighbors for $i^{th}$ location ($n_i$ = 2 to 4)
- $H_{max}$ is the maximum value of 1-$H_{cr}$ for the given dataset, obtained using ten points with the greatest mean squared difference with neighbors

## Objective Function

$$OF = \sqrt[5]{S_{opt} \cdot D_{opt-pH} \cdot D_{opt-EC} \cdot H_{cr-pH} \cdot H_{cr-EC}}$$

- S-optimality
- D-optimality (soil pH)
- D-optimality (soil EC)
- H-criteria (soil pH)
- H-criteria (soil EC)

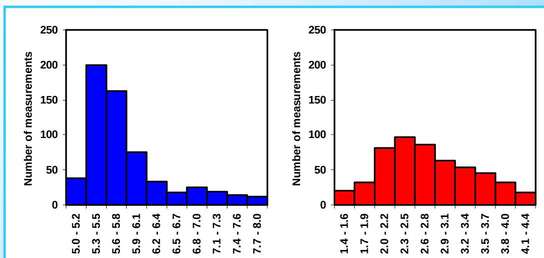
## Scaling

1) Dividing individual values by the median of corresponding criterion estimate for 10,000 random selections
2) Scaling (0 to 1) with respect to the range of corresponding criterion estimates for 10,000 random selections
3) Scaling (0 to 1) of the rank obtained for all selections considered (including 10,000 random and 100 prescribed selections)

## Evaluated Strategies

1. Ten completely random locations
2. Ten random locations within 25% of total possibilities with the lowest local variability in soil pH
3. Ten locations randomly selected from ten equal intervals of soil pH
4. Ten random locations within 25% of total possibilities with the lowest local variability in soil EC
5. Ten locations randomly selected from ten equal intervals of soil EC
6. Ten locations randomly selected from ten rectangular grids
7. Ten locations randomly selected according to categorical separation procedure (three categorized levels of soil pH and EC)
8. Ten locations randomly selected according to categorical separation procedure with local homogeneity constrains
9. Ten locations randomly selected according to a Latin hypercube sampling (LHS) procedure
10. Ten locations randomly selected according to a Latin hypercube sampling (LHS) procedure with S-optimality and H-criteria constrains

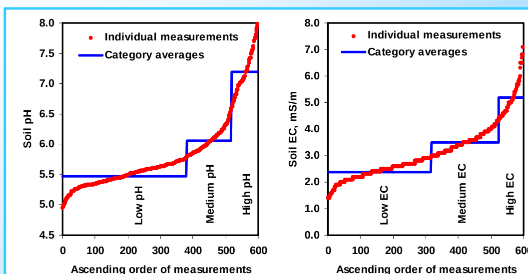
## Measurement Distributions




Soil pH

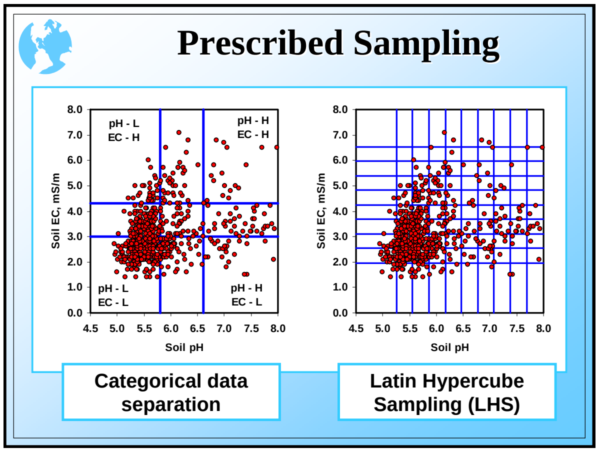
Soil EC, mS/m

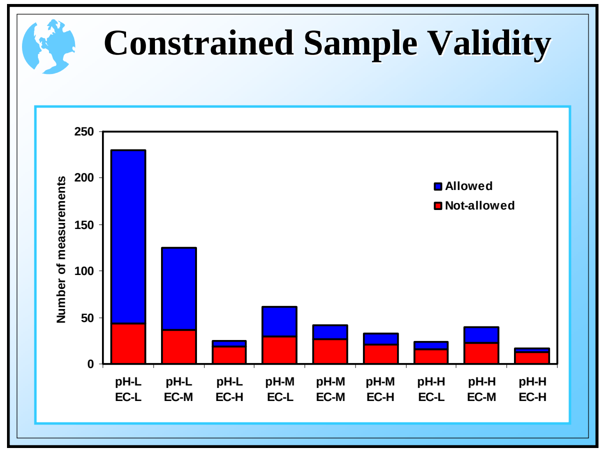
## Measurement Categorization

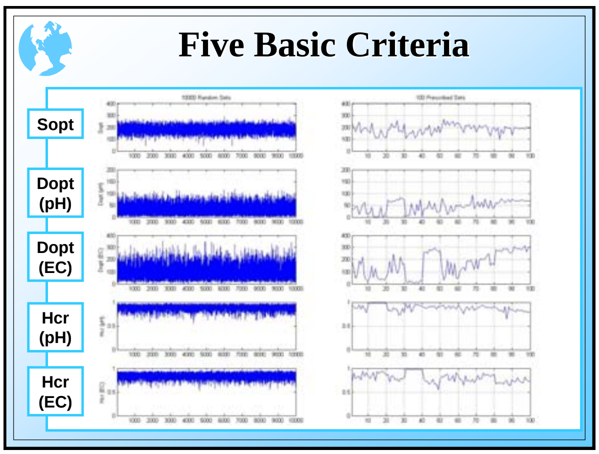


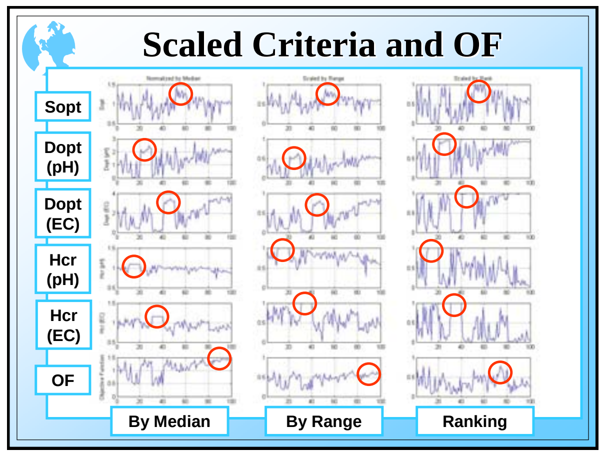

Soil pH

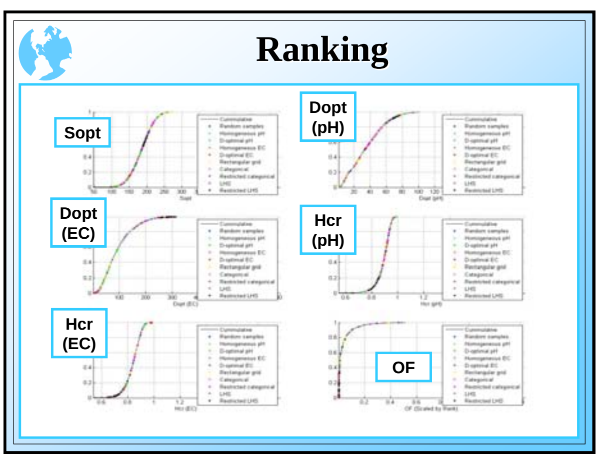
Soil EC

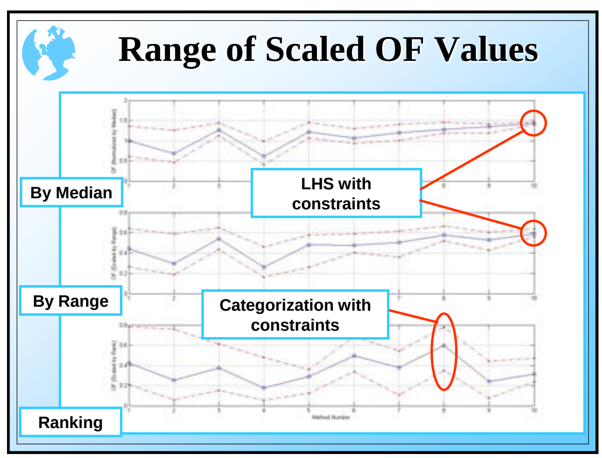

Prescribed Sampling
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
Soil EC, mS/m
pH - L
EC - H
pH - H
EC - H
pH - L
EC - L
pH - H
EC - L
4.5 5.0 5.5 6.0 6.5 7.0 7.5 8.0
Soil pH
Categorical data separation
Latin Hypercube Sampling (LHS)


Constrained Sample Validity
250
200
150
100
50
0
Number of measurements
Allowed
Not-allowed
pH-L EC-L
pH-L EC-M
pH-L EC-H
pH-M EC-L
pH-M EC-M
pH-M EC-H
pH-H EC-L
pH-H EC-M
pH-H EC-H


Five Basic Criteria
Sopt
Dopt (pH)
Dopt (EC)
Hcr (pH)
Hcr (EC)


Scaled Criteria and OF
Sopt
Dopt (pH)
Dopt (EC)
Hcr (pH)
Hcr (EC)
OF
By Median
By Range
Ranking


Ranking
Sopt
Dopt (pH)
Dopt (EC)
Hcr (pH)
Hcr (EC)
OF


Range of Scaled OF Values
By Median
LHS with constraints
By Range
Categorization with constraints
Ranking

# Summary

- An objective function accounts for representing the entire range of sensor data (Dopt), spreading around the field (Sopt) and local homogeneity (Hcr)
- Constrained categorical separation and Latin hypercube sampling were used to simultaneously address all established criteria
- Normalization by median for a large number of random sets appeared to be the most robust method from those considered to precondition estimates of each criterion prior to obtaining their geometrical mean (objective function)
- As long as the formulation of established criteria remains unchanged, this method prevents the subjectivity in setting the weights for individual criteria
- Further optimization of the number of guided sampling locations and the selection process in general is needed

***http://bse.unl.edu/adamchuk***
***E:mail: vadamchuk2@.unl.edu***