International Conference on Geospatial Theory, Processing, Modelling and Applications (Toronto, Ontario, Canada)

Spatially Constrained Geospatial Data Clustering for Multilayer Sensor-Based Measurements

N.M. Dhawale
V.I. Adamchuk
S. O. Prasher
Department of Bioresource Engineering
McGill University

P. R. L. Dutilleul
Department of Plant Science
McGill University

R.B. Ferguson
Department of Agronomy and Horticulture
University of Nebraska-Lincoln

September 7, 2014

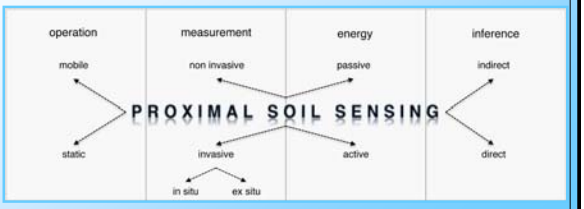
McGill Nebraska Lincoln

Precision Agriculture and Sensor Systems Team

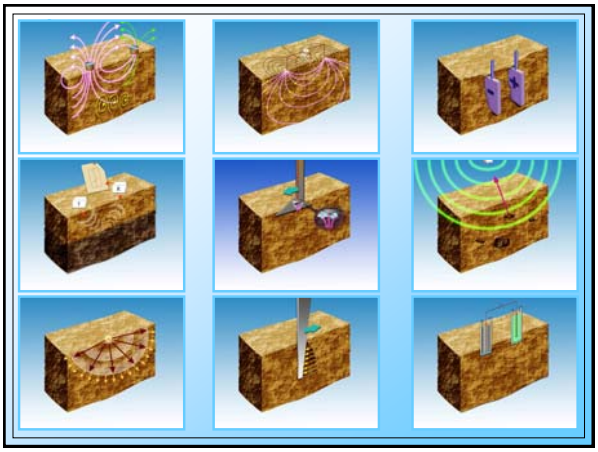
- Development of Proximal Soil and Plant Sensing Systems
- Geospatial Data Processing and Management
- Practical Implementation of Precision Agriculture

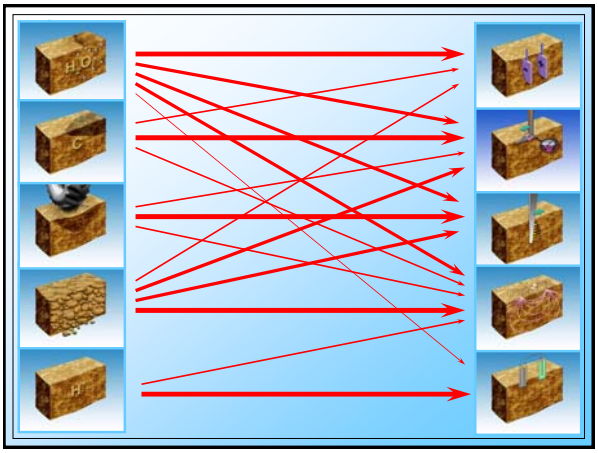
Proximal Soil Sensing

Proximal Soil Sensing (PSS) is a set of technologies developed to measure the physical, chemical and biological properties of soil when placing the sensor in contact with, or at a proximal distance (less than 2 m) to, the soil being characterized

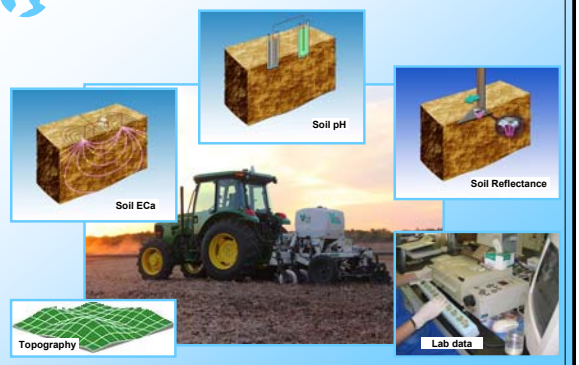


Advances in Agronomy 113: 237-283





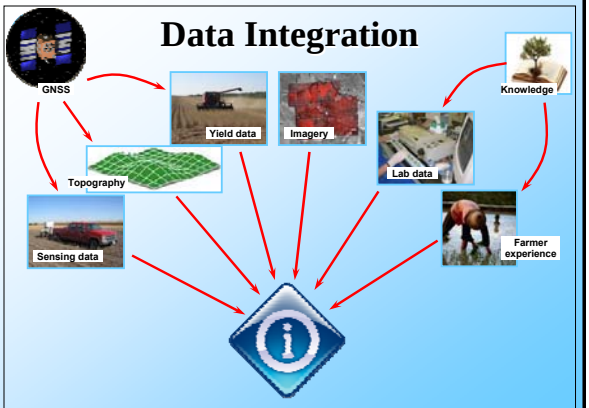
Sensor Fusion



Veris® MSP3

Soil EC and Field Elevation Mapping

Data Integration

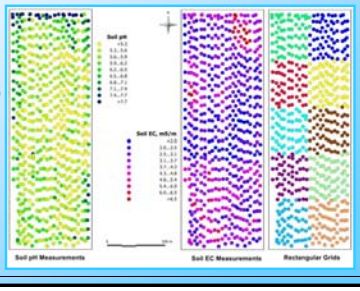


Aerial Imagery System

Site Specific Data Calibration

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

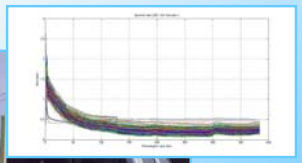
- S-optimality
- D-optimality (soil pH)
- D-optimality (soil EC)
- H-criteria (soil pH)
- H-criteria (soil EC)



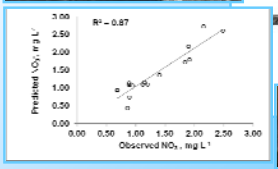
Geoderma 163:63-73

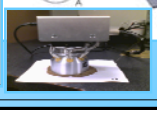
Semi-Automatic Soil Sampling

Soil Profiling Sensors

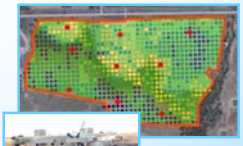


On-the-Spot Analyzer



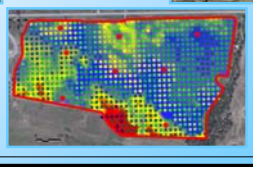


Wireless Sensor Networks



+

=



Map of Real-Time Water Stress Impact Assessment

Present Tools

- K-means clustering
- Distance matrix is based on data
- No account for spatial distances
- The results depends on the selection of initial centroids
- No repeatability
- Need for cross-validation
- Complexity
- Frequently occurring discontinuities

Neighbourhood Search Analysis

- Every delineated area must be contiguous
- Not every part of the field needs to be delineated
- The rest of the field is the rest of the field
- Small areas should not be considered
- Minimum variability within delineated areas
- Multiple layers with no artificial weighting

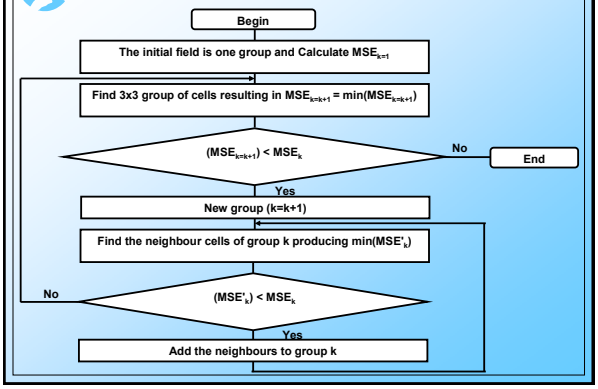
Performance Indicators

$$MSE = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij} - \dot{X}_j\right)^2}{N} \qquad R^2 = 1 - \frac{MSE}{MSE_{k=1}}$$

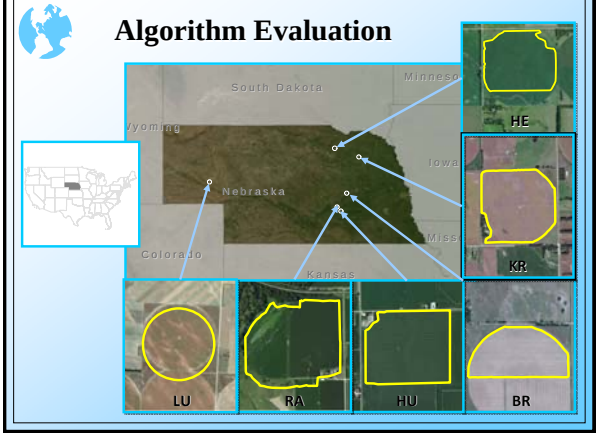
X_{ij} = sensor-value for i^{th} grid cell within j^{th} group
$\dot{X}_j$ = sensor-value average for j^{th} group
k = the number of grid cell groups
n_j = the number of grid cells within j^{th} group
N = the total number of non-zero grid cells

$$R^2_{Product} = R^2_{ShallowEC_a} \cdot R^2_{DeepEC_a} \cdot \left(R^2_{Elevation}\right)^2$$

Suggested Algorithm



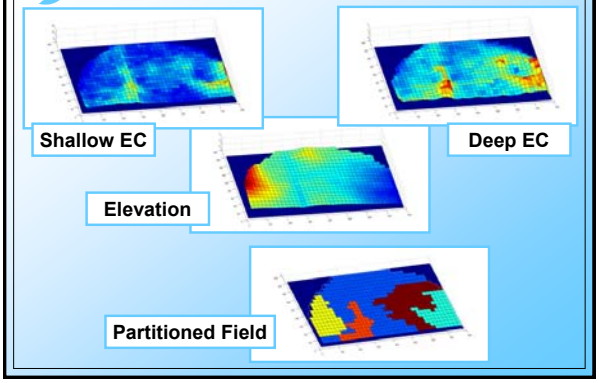
Algorithm Evaluation



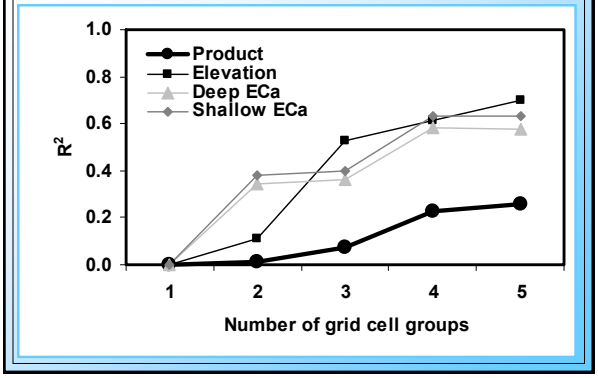
Summary of Sensor Data

Field ID	Area, ha	Mean	Range	SD
		Field elevation, m		
1	25	1.50	3.20	0.53
2	46	4.95	17.82	3.99
3	50	7.10	11.54	2.09
4	55	8.07	27.44	5.68
5	67	4.22	8.09	1.60
6	44	6.15	10.59	2.15
		Shallow ECa, mSm^{-1}		
1	25	0.73	1.58	0.28
2	46	3.99	13.14	1.67
3	50	6.21	11.64	1.84
4	55	2.44	9.04	1.72
5	67	7.25	9.32	1.88
6	44	2.29	7.42	0.82
		Deep ECa, mSm^{-1}		
1	25	7.62	27.66	3.76
2	46	30.24	86.90	14.39
3	50	4.10	8.68	1.71
4	55	16.31	61.97	12.06
5	67	51.01	80.77	14.07
6	44	25.72	81.74	14.36

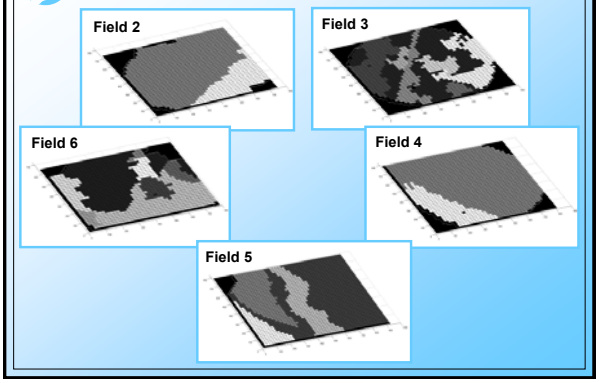
Neighbourhood Search Clustering



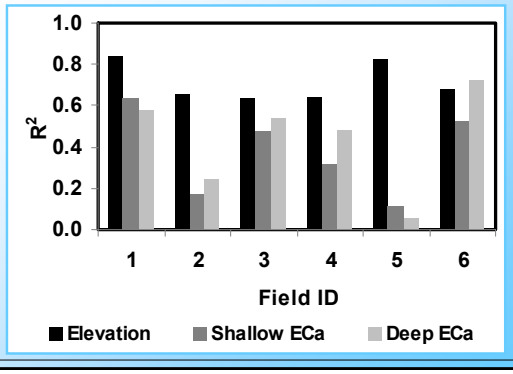
R^2 Dynamics



Other Fields

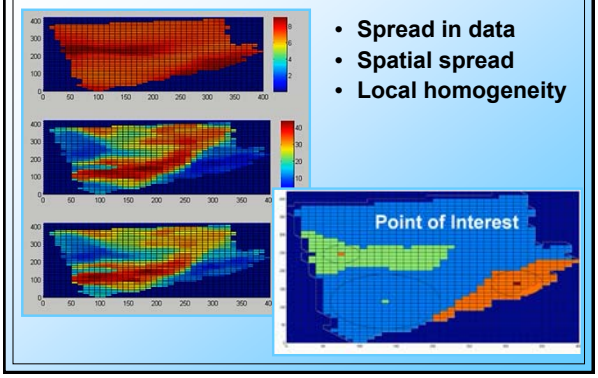


Overall Performance



Algorithm Performance

Anchor Locations



- Spread in data
- Spatial spread
- Local homogeneity

Summary

The spatial clustering algorithm developed in this study is based on a neighbourhood search method and seeks to minimize variance inside each group of interpolated grid pixels corresponding to an unlimited number of sensor-based data layers.

Preliminary tests of the algorithm using six production fields illustrated algorithm robustness when delineating field areas with different field elevations and soil EC_a measurements.

Each spatially constrained group of grid cells with the exception of the first group designated as "the rest of the field" emerged in response to every unique combination of data values relatively constant within each group.

http://adamchukpa.mcgill.ca
E:mail: viacheslav.adamchuk@mcgill.ca