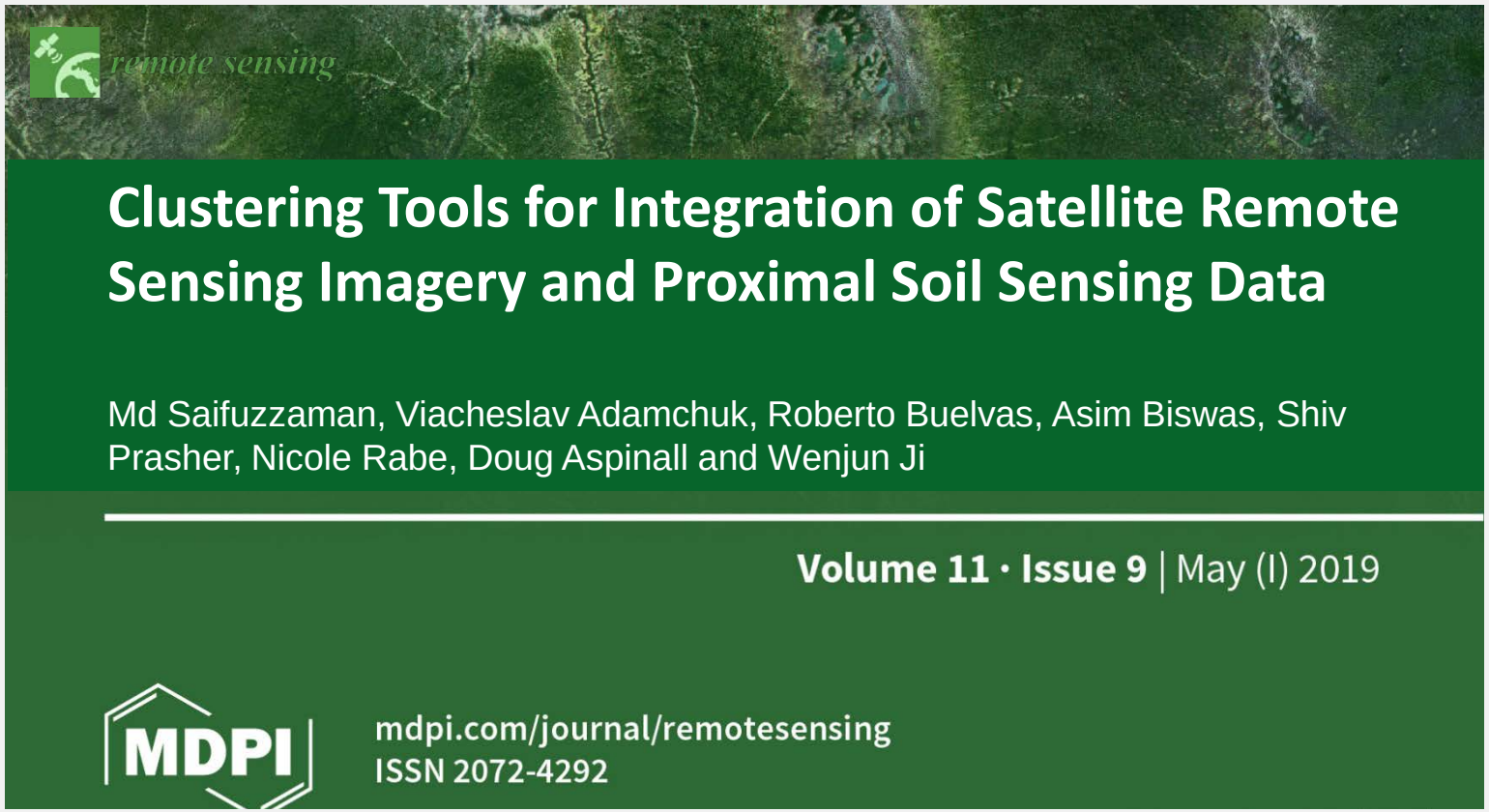
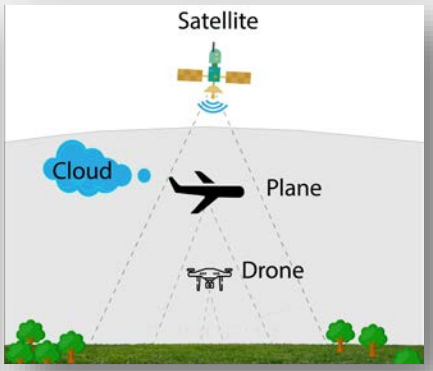
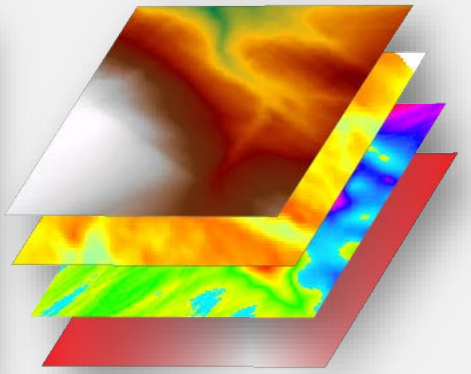
Study I: Published



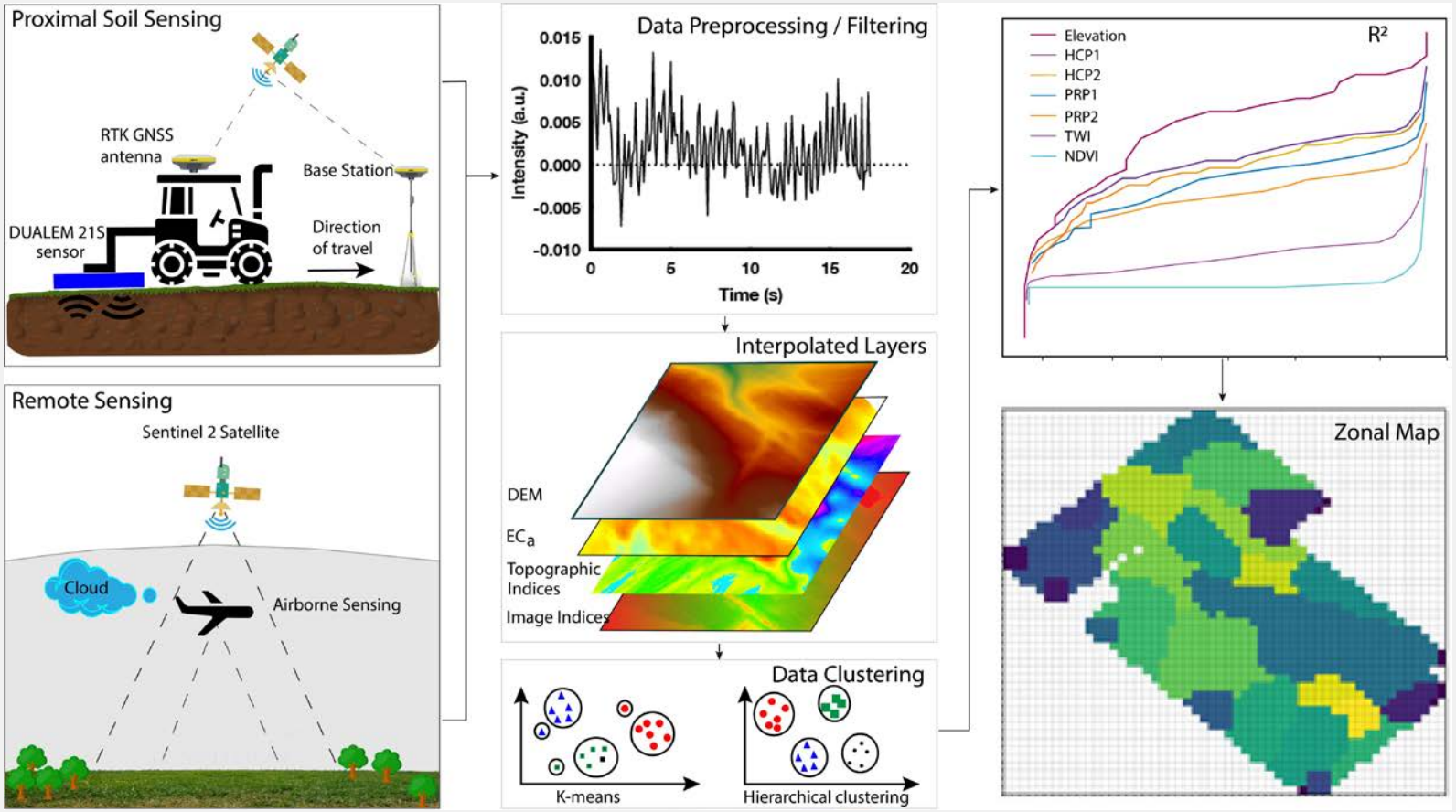
Data Clustering

- A large amount of geospatial data provide essential information for developing thematic maps.
- Hierarchical data clustering techniques are crucial for data mining, and high-density data analysis is important for field management.
- This study presents the process used to develop a new and enhanced clustering technique to better understand field variability.



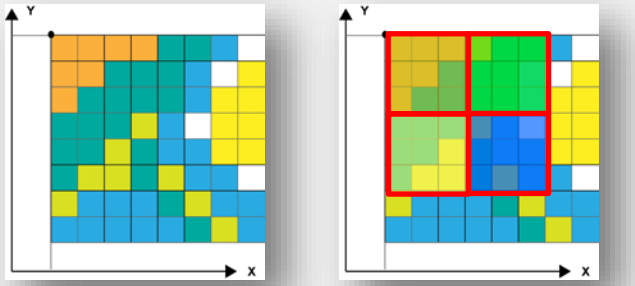


Sensor Data > Data processing > Clustering > Comparison



Neighborhood Search Analyst Tool

- NSA developed on open-source programming platform to enable hierarchical spatial clustering of multilayer data.
- The algorithm evaluates data structure on location-based information of the multiple layers to be used for delineating spatially constrained groups of high-density sensor-based measurements.
- The algorithm seeks to minimize the mean squared error (MSE) during the interactive grouping of spatially adjacent measurements similar to each other and different from other parts of the field.



Data and Study Sites

PSS and RS data

- Apparent Electrical Conductivity (EC_a)
- Elevation, slope, aspect ratio & topographic wetness index (TWI)
- Normalized difference vegetation index (NDVI), Normalized difference red-edge index(NDRE)

Field	Area (ha)	Soil Class	Target crops
WH	39.60	Loam	Soybean/Wheat
LD	21.00	Sandy loam - SL	Soybean
RB	75.00	Fine SL	Soybean/Wheat

Interpolated Data Layers and Variability

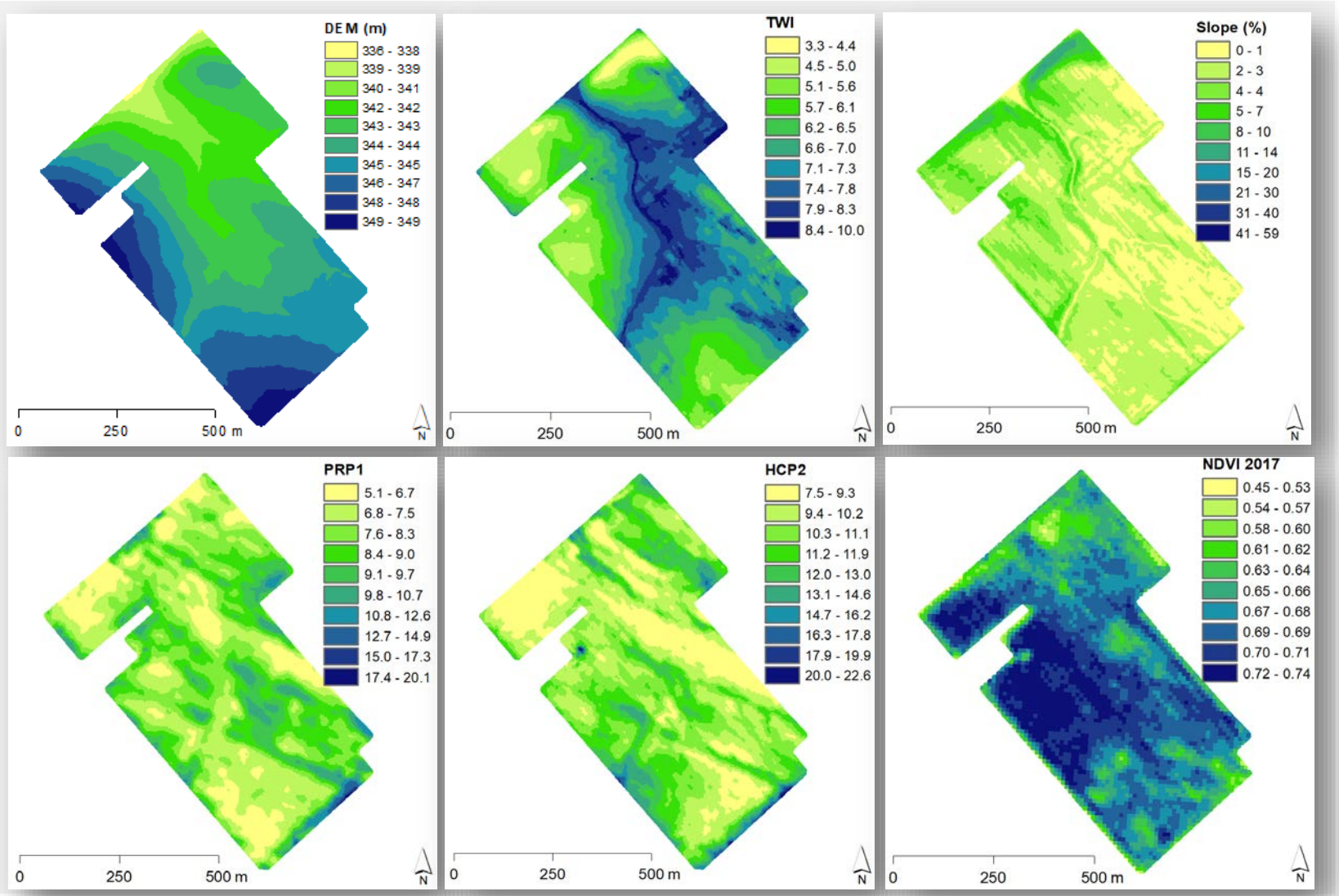
Chapter
3
4
5

Background

Methods

Results

Conclusion



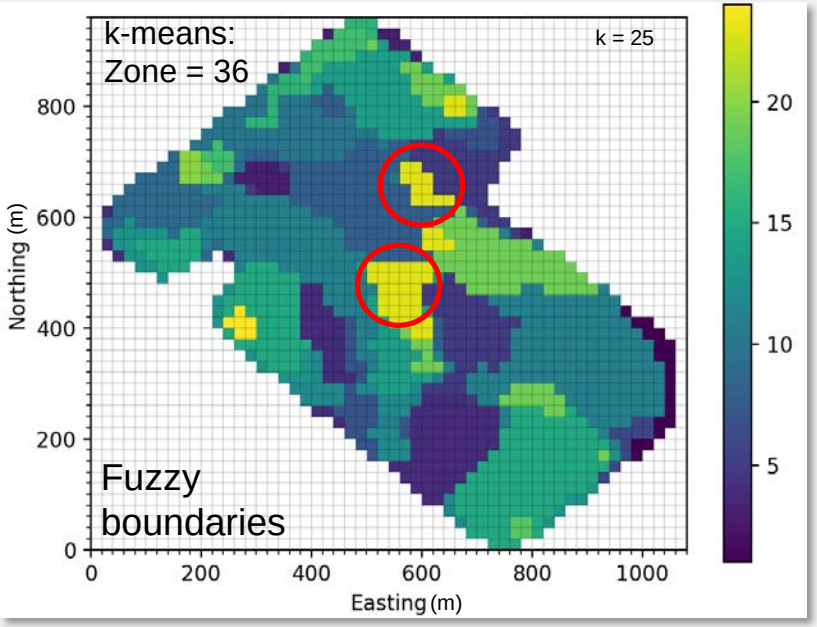
k-means vs. NSA Clustering

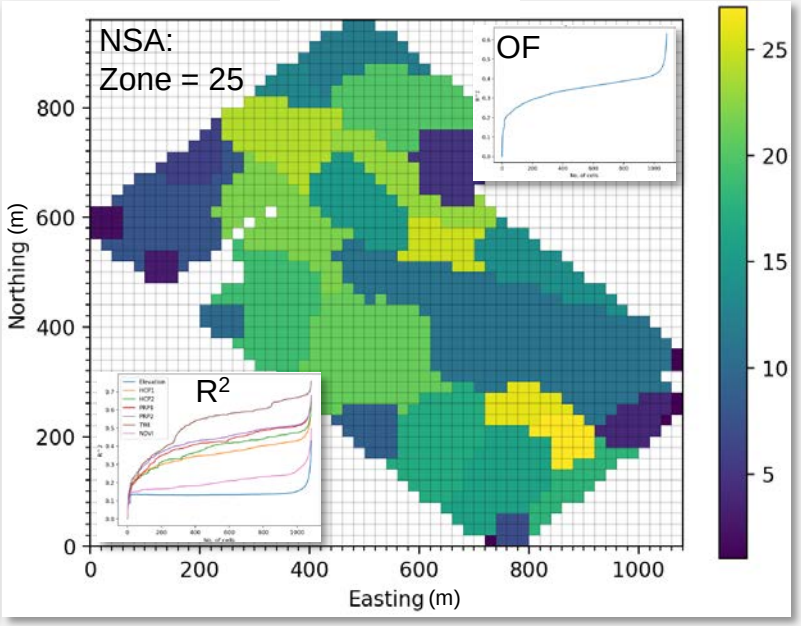
K-means

- R^2 and zones varied since there was a random component.
- User-defined clusters with isolated boundaries.

NSA

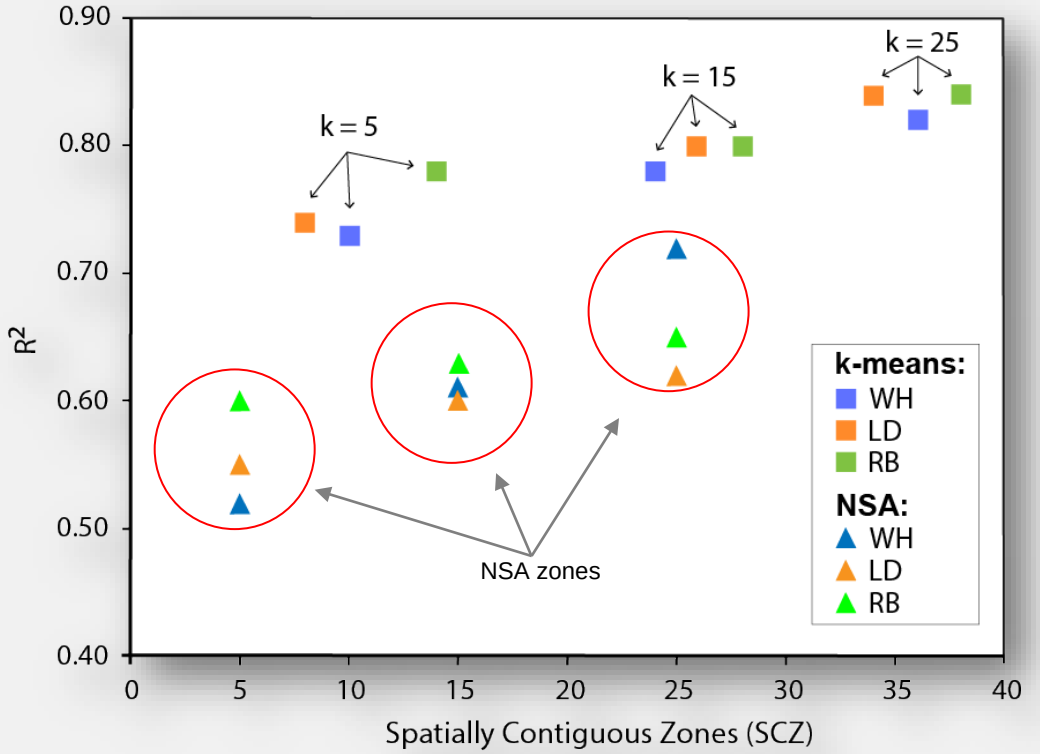
- Max. 70% field variance (R^2 = 0.70) achieved in NSA.
- Optimum number of zones without defining cluster centers.





Comparison: Two Clustering Methods

- **User defined clusters**: k = 5 / 15 / 25
- **k-means**: WH, LD, RB: 36, 34, 38 zones (k = 25)
- **NSA**: Stable and spatially contiguous zones



k-means: $R^2 = 0.80$

NSA: $R^2 = 0.60$

- The high-density and multivariate data clustering approach provided user-defined number of zones for the agricultural fields.
- Compared to fuzzy clustering algorithms, NSA has a unique capability for optimum zone separation.
- It proved to be capable of handling a significant number of variables and data layers for delineating the optimum number of zones in a more robust way.
- Sampling optimization and georeferenced thematic maps are useful for variable rate technologies and for other management purposes.

Acknowledgement

McGill University

Fonds de recherche sur la nature et les technologies Québec

Ontario Ministry of Agriculture, Food and Rural Affairs

Natural Resources Canada / Ressources naturelles Canada

Canada