Analysis of Soil Properties Predictability Using Different On-the-Go Soil Mapping Systems

H.-H. Huang[1], V. Adamchuk[1], A. Biswas[2], W. Ji[3], and S. Lauzon[1]

[1]Bioresource Engineering Dept., McGill University (Ste-Anne-de-Bellevue, QC, Canada)

[2]School of Environmental Sciences, University of Guelph (Guelph, ON, Canada)

[3]Soil and Environment Dept., Swedish University of Agricultural Sciences (Skara, Sweden)

A paper from the Proceedings of the
14th International Conference on Precision Agriculture
June 24 – 27, 2018
Montréal, Québec, Canada

Abstract. Understanding the spatial variability of soil chemical and physical attributes allows for the optimization of the profitability of nutrient and water management for crop development. Considering the advantages and accessibility of various types of multi-sensor platforms capable of acquiring large sensing data pertaining to soil information across a landscape, this study compares data obtained using four common soil mapping systems: 1) topography obtained using a real-time kinematic (RTK) global navigation satellite system (GNSS) receiver, 2) apparent soil electrical conductivity obtained using an electromagnetic induction instrument with topographic data, 3) combination of apparent soil electrical conductivity obtained using galvanic contact resistivity sensing, subsurface soil reflectance and direct soil pH measurements with topographic data, and 4) passive gamma-ray spectroscopy with topographic data with regards to their capability to predict six soil properties: clay content, cation exchange capacity (CEC), soil pH, soil organic matters (SOM) content, extractable potassium (P) and phosphorus (K) levels. These systems were used to map two agricultural fields: NX (45 ha) and ST (40 ha) in northeastern Ontario, Canada. It was shown that sensor combinations produced lower prediction errors as compared to individual sensors.

Keywords. *sensor fusion, sensor calibration, sampling optimization, regression modeling*

Introduction

Understanding the spatial variability of soil chemical and physical attributes optimizes the profitability of nutrient and water management for crop development. Soil mapping systems with various types of proximal soil sensors provide crop growers with a great opportunity to access soil heterogeneity at a sub-meter spatial scale in an efficient and less invasive manner. Studies suggest that sensing information linked to soil pH, electrical conductivity, organic matter content, soil moisture, etc., can be obtained in a relatively cost-effective manner (Adamchuk 2007; Lund 2011; Heege 2013). Given the easy accessibility to many options of PSS in recent decades, sensor-based soil characterization has gained in popularity (Viscarra Rossel et al. 2010). Various types of PSS capable of directly or indirectly estimating the magnitude of soil properties commonly quantified by conventional soil testing methods are available on the market. The majority of sensor deployment scenarios employ off-site interpretation of sensor-based soil maps to alter nutrient input (Viscarra Rossel et al. 2011). Although many sophisticated soil mapping systems can be used to detect specific soil properties, one single system capable of responding to all soil properties does not yet exist (Adamchuk et al. 2011; Mahmood et al., 2012).

An economical alternative incorporates the strengths of different sensors by assembling them on an on-the-go platform (Kuang et al. 2012; Mahmood et al. 2012; Adamchuk et al. 2011). Such platforms are typically equipped with a real-time kinematic (RTK) level global satellite system (GNSS) receiver to take into account the topographic-induced hydrological effects on the spatial distribution of agri-chemicals and yield. Elevation and its derivatives, such as slope, aspect, and topographic wetness index (TWI), flow direction, flow length, catchment area, etc. have been used as auxiliary data for approximate soil heterogeneity (De Benedetto et al. 2013; Piikki et al. 2013, 2015). Additionally, soil conductivity sensors, potable ion-selective sensors and soil reflectance sensors are the three other popular choices for precision farming applications (Srinivasan, 2006). Conductivity sensors measure the electrical conductivity of bulk soil to infer soil texture, salinity and moisture. On-the-go ion-selective sensing systems use electrodes, such as ion-selective electrode and ion-selective field effect transistor to detect ion activity associated with plant required nutrients in a soil solution (e.g., pH, nitrate-N, etc.) from soil while traveling. Soil reflectance sensors are based on the principle of spectroscopy to quantify soil properties associated with color changes, such as organic matter, moisture content and texture and other influential properties such as cation exchange capacity, soil pH and total nitrogen. Intensive research studies have been conducted to investigate the optimum wavelengths or wavelength ranges (e.g., visible, mid-infrared, near-infrared) for the soil properties of interest. Thus, the ability to take into account the effects of sensor responses from multiple soil attributes when translating sensing measurements into a specific attribute of interest has become an attractive field of research for sensor developers.

Considering the advantage and accessibility of various types of multi-sensor platforms capable of acquiring large sensing data pertaining to soil information across a landscape, this study compares the predictability of four common soil mapping systems: 1) topography only (RTK), 2) DUALEM-21S (Dualem, Inc., Milton, Ontario, Canada) with topography, 3) Veris MSP3 (Veris Technologies, Inc., Salina, Kansas, USA) with topography, and 4) SoilOptix (Practical Precision, Inc., Tavistock, Ontario, Canada) with topography for six soil properties: percent clay, cation exchange capacity (CEC), soil pH, soil organic matters (SOM), potassium (K) and phosphorus (P) levels.

Materials and methods

To compare the soil predictability of the four selected soil mapping systems, the procedure illustrated in **Figure 1** was followed. Field data was collected at two agricultural fields, NX (2015) and ST (2016), located in Eastern Ontario, Canada (**Figure 2**). NX and ST are approximately 40 ha and 45 ha in area, respectively. Data collection was performed included soil sensing using the four soil mapping strategies and through soil sampling.

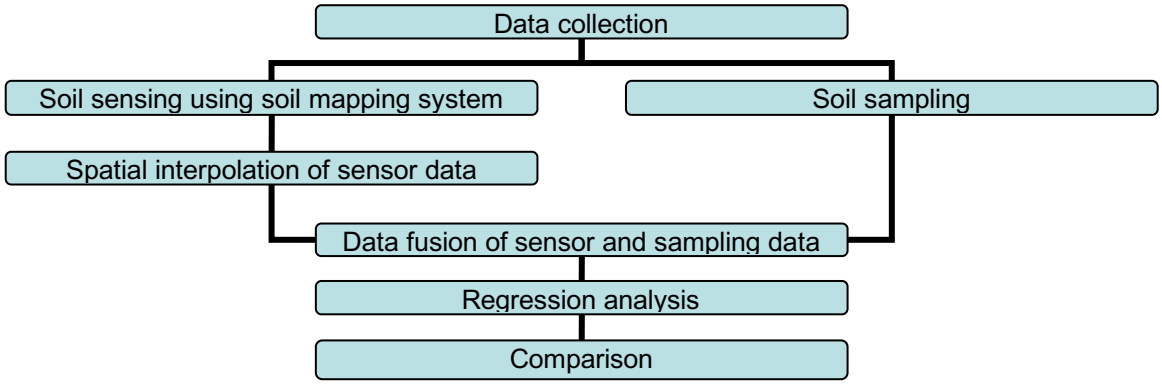


Figure 1. Flowchart illustration the procedure used to compare the predictability of the four soil mapping systems.

Figure 2. Geographic locations of the two studied fields: ST and NX.

Soil sensing

As shown in **Figure 3**, the four popular soil mapping systems: RTK, Dualem, MSP3 and SoilOptix were used to collect dense geo-referenced sensor measurements in this study. Their description and evaluated sensor variables are listed in **Table 1**. An RTK-level GNSS receiver was used to collect elevation measurements and to extract different topographic features: elevation, slope, aspect ratio index (ARI) and topographic wetness index that potentially affect agro-chemical movement across a field. ARI was calculated as *sin(aspect/2)*. TWI was calculated using SAGA GIS version 2.4 (University of Hamburg, Hamburg, Germany) developed by Conrad et al. (2015) to quantify topographic-induced hydrological effects. Dualem, pairing a horizontal co-planar geometry (HCP) and a perpendicular geometry (PRP) windings, was used to obtain EC_a measurements according to the signals accumulated from four depths [e.g., PRP1 (0-0.5 m), HCP1 (0-1.5 m), PRP2 (0-1.0 m) and HCP2 (0-3 m)]. MSP3 housing an EC surveyor (EC 3100 model) to collect EC_a measurements based on the signals from two depths [e.g., shallow (0-0.3 m) and deep (0-0.9 m)], a set of pH electrodes to estimate soil pH, and an dual-wavelength OpticMapper Module to obtain visible red (660 nm) and near-infrared (940 nm) soil

reflectance at a depth of approximately 5 cm below the surface. SoilOptix scanner was used to collect the selected attributes of gamma rays emitted from top soil (0-0.3 m) including total count, ^{40}K, ^{232}Th and ^{238}U. Raw sensor data was filtered to remove abnormal values. As listed in **Table 1**, a total of 17 sensor variables were used to predict soil properties; among them, four under RTK, four under Dualem, five under MSP3, and four under SoilOptix.

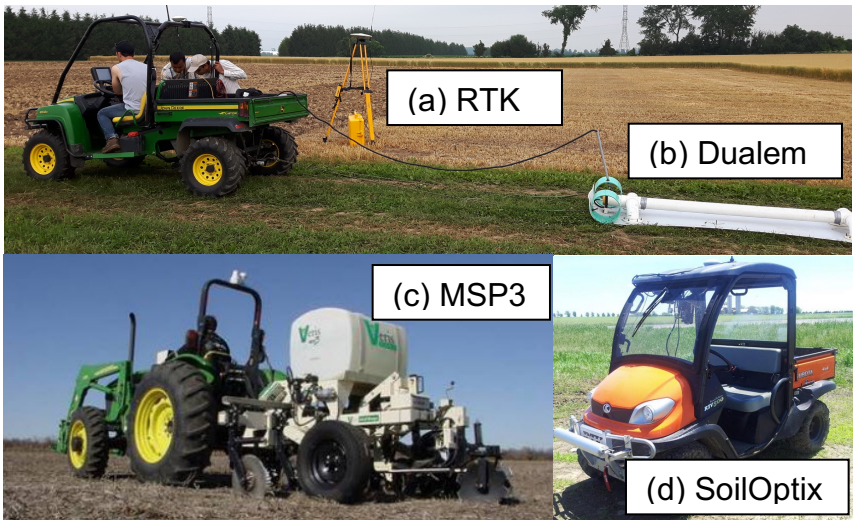


Figure 3. Four soil mapping systems: (a) RTK, (b) Dualem, (c) MSP3, and (d) SoilOptix used in this study.

Table 1. The descriptions and the evaluated sensor variables of the four soil mapping platforms.

Systems	Soil characteristics	Sensing variables
RTK	Landscape	Elevation
		TWI – Topographic wetness index
		Slope
		ARI – Aspect ratio index
Dualem	Soil apparent electrical conductivity (EC_a)	PRP1
		HCP1
		PRP2
		HCP2
MSP3	Soil apparent electrical conductivity (EC_a)	EC.sl – Shallow EC_a
	Soil acidity	EC.dp – Deep EC_a
	Reflectance of soil color	pH.s – Veris soil pH
		RED – Red reflectance
		NIR – Near-infrared reflectance
SoilOptix	Top soil gamma rays	r.TC – Total count of gamma rays
		r.40K – Gamma rays ^{40}K
		232Th – Gamma rays ^{232}Th
		238U – Gamma rays ^{238}U

Soil sampling

Soil sampling was performed to investigate the predictability of the four mapping systems and to develop further strategies for soil mapping optimization. Three sampling strategies: grid, target and validation samplings were deployed at the two selected fields. Grid sampling was performed using 1-ha square grids. With each grid, one composite soil sample was formed by mixing multiple soil cores scattered within the grid cell at a fixed depth of 0.15 m. A total of 35 and 46 grid samples were collected for the ST and NX fields, respectively. Target sampling locations for both fields were modeled using a newly developed algorithm, Neighborhood Search Algorithm

(Dhawale et al., 2016). This approach determined 20 optimal locations at each site by delineating spatial-clusters based on the spatial variability of multiple sensor layers. Validation samples were collected as a separated dataset for validation purposes. 10 random samples were obtained for each field. The total number of soil samples was 65 and 76 for ST and NX, respectively. All sampling locations were geo-referenced using a Garmin handheld GNSS/GPS device with WASS correction. The collected soil samples were sent to certified laboratories [i.e., A &L Canada Laboratories Inc. (London, Ontario, Canada) for chemical analysis and Agro-Enviro-Lab (La Pocatière, Québec, Canada) for textural analysis and soil organic matter] to obtain soil property content. In this study, six measured soil properties of interest were clay content (%), SOM (%), pH, CEC (meq hg^{-1}), P (ppm), K (ppm). **Figure 4** illustrates the spatial distribution of data points collected using soil mapping systems and soil sampling.

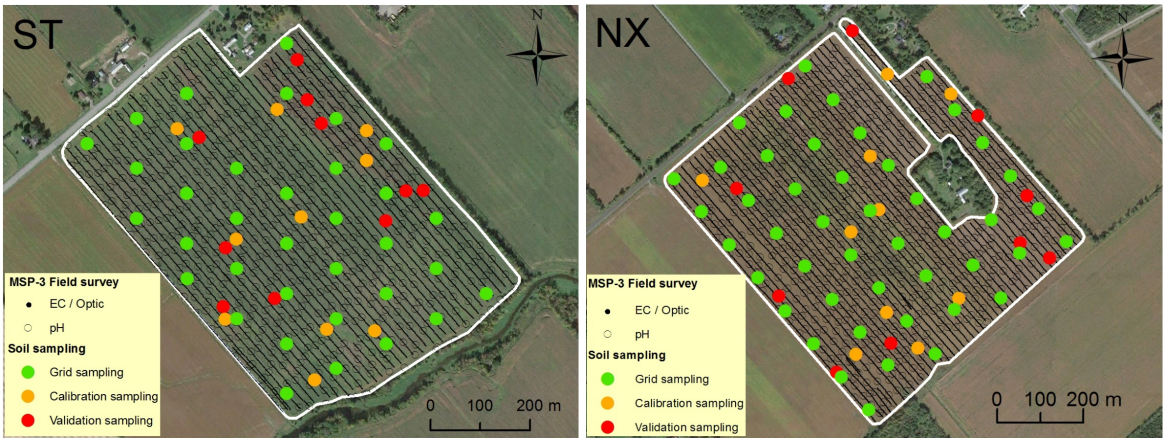


Figure 4. Illustration of the spatial distribution of sensor measurements and soil sampling locations at the two studied fields ST and NX.

Spatial interpolation and extraction

Ordinary kriging (OK) interpolation was used to create the continuous sensor data surface as sensor measurements were distributed along multiple travelling passes with different fixed intervals. The interpolated surface allowed further extraction of sensor data at soil sampling locations. The commercial GIS software ArcGIS™ developed by ESRI (2015) was used to execute this task. In its Geostatistical Analyst tool, a 2nd degree polynomial surface trend removal function was used to ensure the autocorrelation of sensor data following a spherical fitting model. All point based sensor data layers were interpolated through this process and stored as a raster format with 5m x 5 m grid cells. Furthermore, ArcGIS was also used to perform spatial extraction of sensor data (i.e., raster surface) at each soil sampling location. A tabular file showing the coordinates of soil sampling locations and corresponding soil analysis data and multiple types of sensor measurements was prepared for regression analysis.

Regression analysis

In this study, the analyses were conducted using R version 3.4 (R Development Core Team, 2013). Exploratory analyses: histogram, boxplot and correlation matrix were used to investigate data distribution, quality, and pattern providing the basis of classical multivariate regression. Both histogram and boxplot were used to guide the process of outlier removal to optimize regression performance. Pearson correlation matrix was used to provide rapid observation of the senor responses to soil properties. Partial least squares regression (PLSR) analysis was used to evaluate the strength of soil predictability using the four mapping systems under multiple linear models. As topographic measurements providing substantial soil characteristics are easily available, this study compared the soil prediction using 1) RTK only, 2) Dealem + RTK,

3) MSP + RTK and 4) SoilOptic + RTK. The interaction terms of predictor variables were added to regression analysis. To obtain a global comparison of the prediction quality, all soil samples (i.e., 76 at NX; 65 at ST) were used in the regression analysis regardless of the different sampling strategies. As a rule of thumb, the predictor variables were limited to 7, which was 10% of the number of samples, in the model selection of PLSR analysis. Regression coefficient (R^2) and root mean squared error of prediction model (RMSEP) were used as quality indicators.

Results and discussion

Figure 5 is the histogram of sensor data showing the density distribution plot, mean (μ), standard deviation (*s*) and number of measurements (n) for both fields. In general, Dualem PRP1 EC_a was lower (μ = 15.67 mSm^{-1} at NX and μ = 15.67 mSm^{-1} at ST) than Veris shallow EC_a with a smaller standard deviation (s = 15.67 mSm^{-1} at NX and μ = 15.67 mSm^{-1}). Both EC_a instruments revealed two spatial clusters with two data peaks across ST. In terms of topology, NX was situated at a higher elevation and was relatively flatter than ST. The total count of gamma rays was very similar in terms of mean, standard deviation and range. Red reflectance was relatively smoother at ST with a narrower range. Both pH data layers presented a similar range (roughly 5 ≤ pHs ≤ 8; s < 1.0). Veris pH sensor also reveals two clusters similar to EC_a instruments at ST.

Figure 6 is the histogram of measured soil properties. For the selected properties, the measurements obtained using grid and target sampling strategies show similar distribution patterns. Both fields have a similar range of clay content, yet, NX contained the higher clay percentage (20% – 25%) than the one of ST (15% – 20%). The range of SOM was also similar with slightly higher variations at NX. Although the range at both fields was similar, the acidity was stronger at ST. Similar to Veris pH sensor data, soil sampling of pH also revealed two data clusters suggesting potential subdivisions across the field. The level of CEC was slightly higher at ST than NX, which was consistent with the higher SOM level at St. Potassium level and variation were much higher at ST than NX (μ = 141.56 ppm and s = 47.78 ppm at ST; μ = 83.34 ppm and s = 18.45 ppm at NX). Phosphorus level was also higher and more varied at ST than NX with (μ = 40.34 ppm and s = 19.65 ppm at ST; μ = 33.60 ppm and s = 15.67 ppm sat NX).

Figure 7 shows the correlation matrixes of the measured soil properties and the sensor measurements for both NX and ST, illustrated by using Pearson correlation coefficient (*r*). The increased color intensity represented a stronger correlation with blue showing positive correlation and red showing negative correlation. Among the RTK variables, clay content was highly correlated with elevation at NX (*r* = -0.85) but weaker at ST (*r* = -0.62). TWI was another topographic variable showing good correlation but only presented at ST (*r* = 0.62). The four Dualem sensor variables all highly correlated with clay content, in particular, PRP1 (*r* = 0.79) for both fields. MSP3 variables were strongly correlated with clay content except RED reflectance for the NX field. Among them, EC.sl (the shallowest EC_a measurements of MSP3) showed the strongest correlation with clay content (*r* = 0.89 at NX and *r* = 0.8 at ST), similar to PRP1 (Dualem's shallowest EC_a measurements). Two variables of SoilOptix: r.232Th and r.TC showed a strong correlation with clay content. For SOM correlation, all four mapping systems had a weak to moderate correlation (-0.45 < *r* < 0.45) with elevation and EC_a associated variables performed slightly better at NX. Such correlation patterns did repeat at ST and only pH showed a moderate correlation (*r* = -52). The correlation with measured pH was very different for the two fields. RTK variables (except ARI), Dualem variable, and MSP3 variables (except optic reflectance) were well correlated with soil pH and CEC at ST. However, all mapping systems performed poorly at correlating soil pH, except MSP3 variables at NX. At least two variables from each mapping systems had good correlation with CEC at ST. As for P and K, all four systems provided poor correlations. **Figure 8** is the correlation matrix of different sensor measurements to investigate the potential collinearity. Visually, the two matrixes were similar

with slight differences. Among them, ARI, RED, r.40K, r.238U were not correlated with any other variables at both fields. Slope and TWI were not correlated with the rest of variables at NX field.

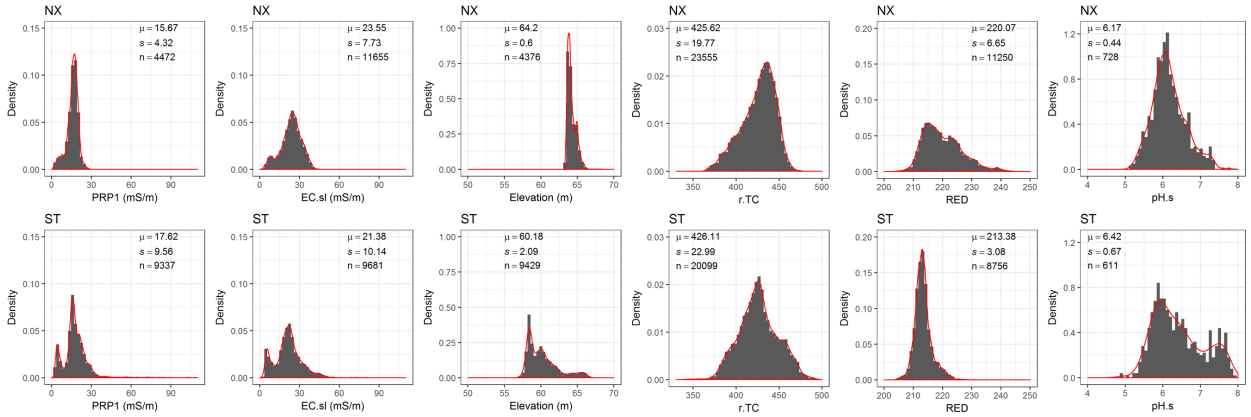


Figure 5. Histograms showing the distribution of sensing measurements collected at NX and ST fields.

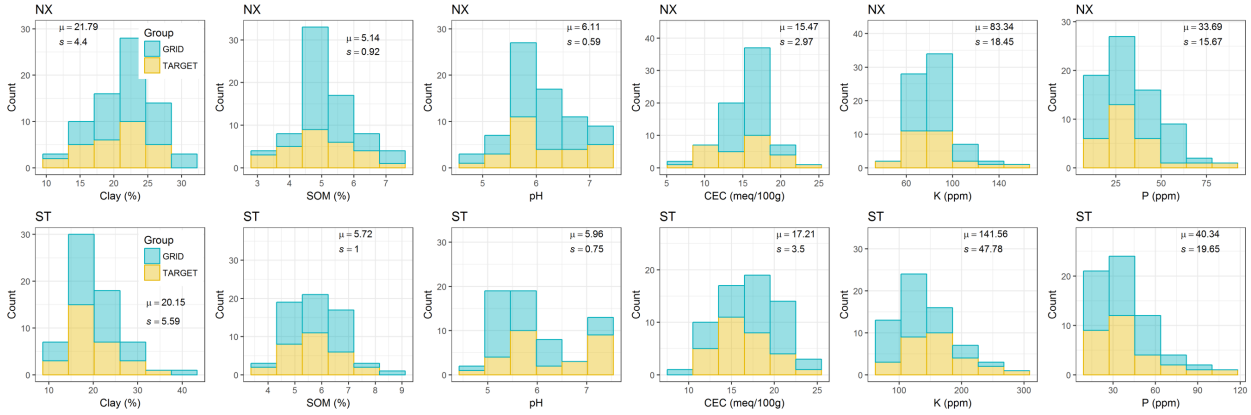


Figure 6. Histograms showing the distribution of soil property values obtained through grid and target samplings.

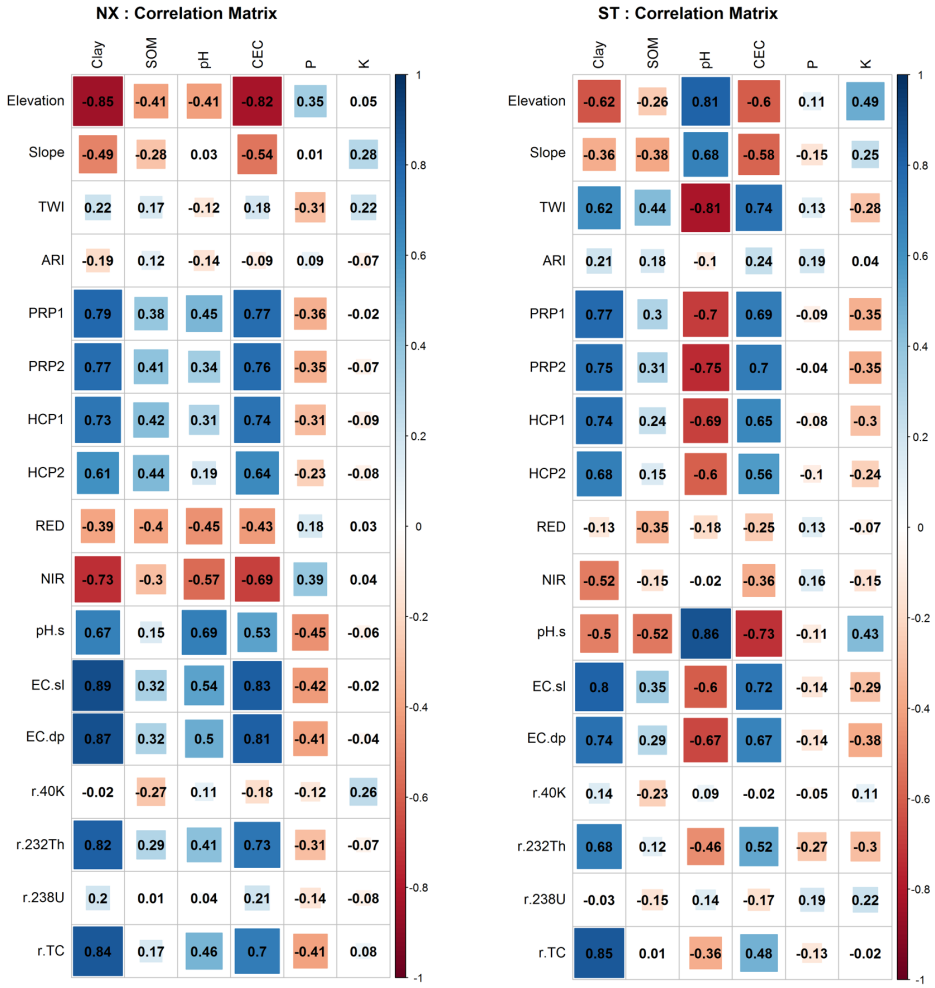


Figure 7. Correlation matrix showing the correlation between sensor measurements and soil properties for NX and ST fields.

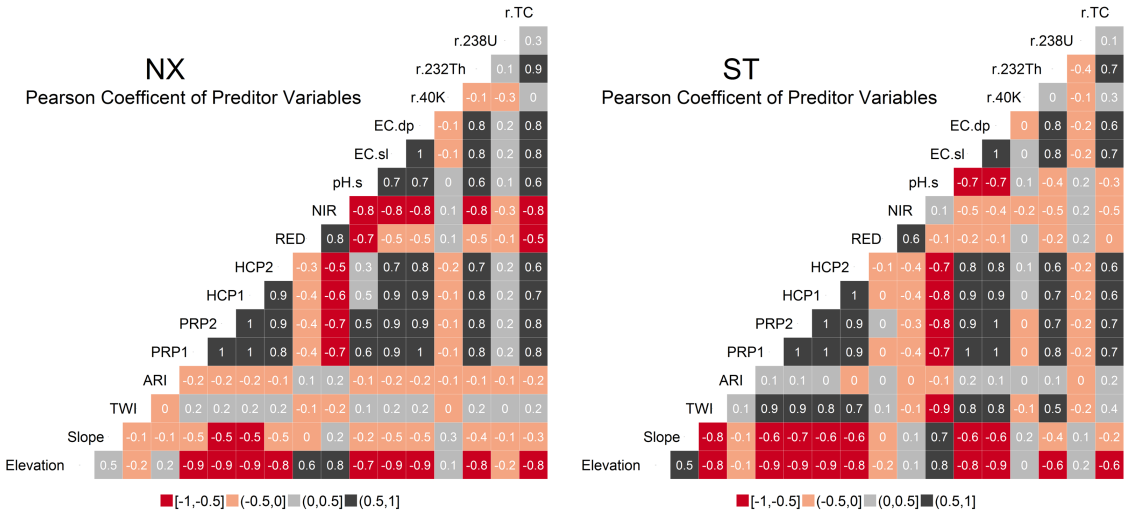


Figure 8. Correlation matrix showing the collinearity among sensor variables.

Figure 9 illustrates the spatial autocorrelation of different sensor measurements using an experimental semivariogram. All plots showed a tendency of spherically-modelled curve with various values of nugget, sill and range. These inconsistent spatial structures of senor data across the two evaluated fields suggest the uniqueness of soil composition at each site. **Figure 10** shows the data distribution patterns of the extracted sensor measurements at sampling locations (illustrated in red) with reference to the distribution of the entire sensor data (illustrated in black). The similarity in the two colored patterns in each plot suggests that the extracted sensor measurements were representative of further PLSR regression.

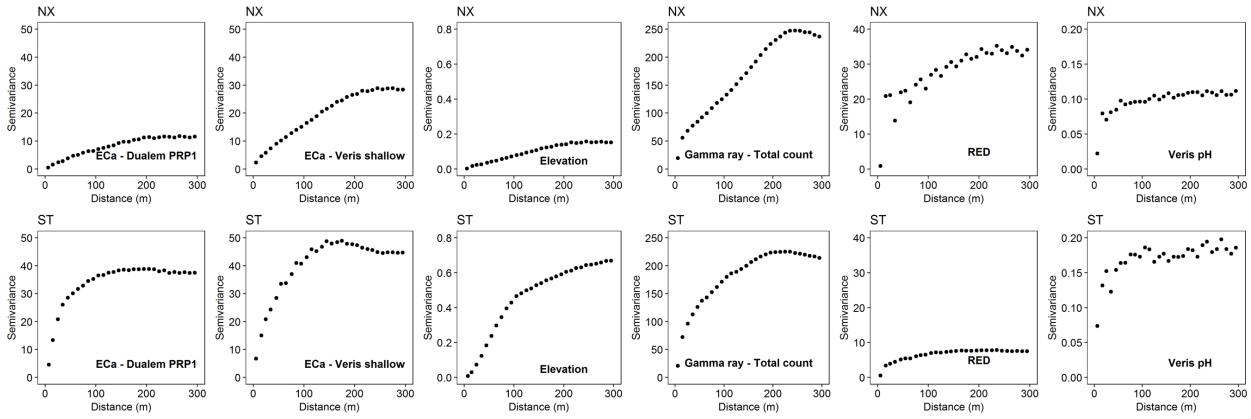


Figure 9. Semivariograms of different sensor measurements at NX and ST fields.

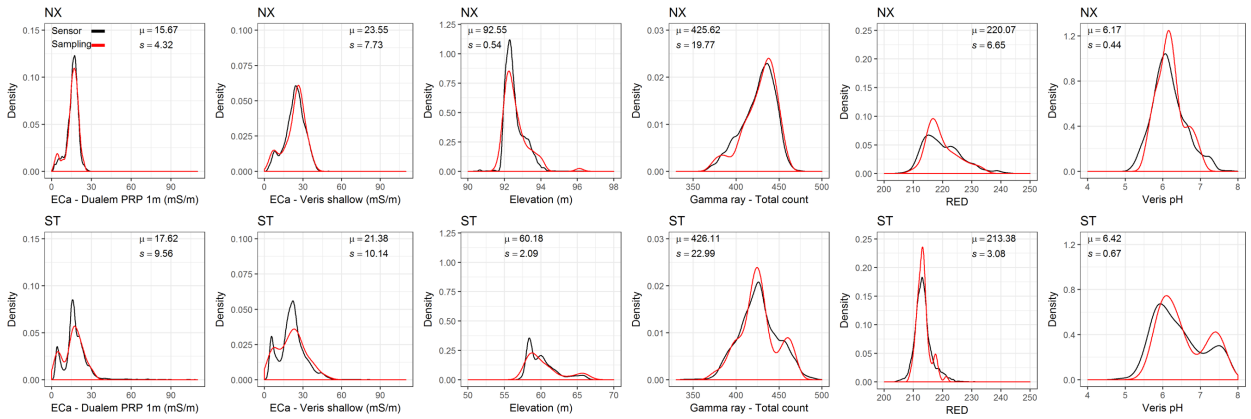
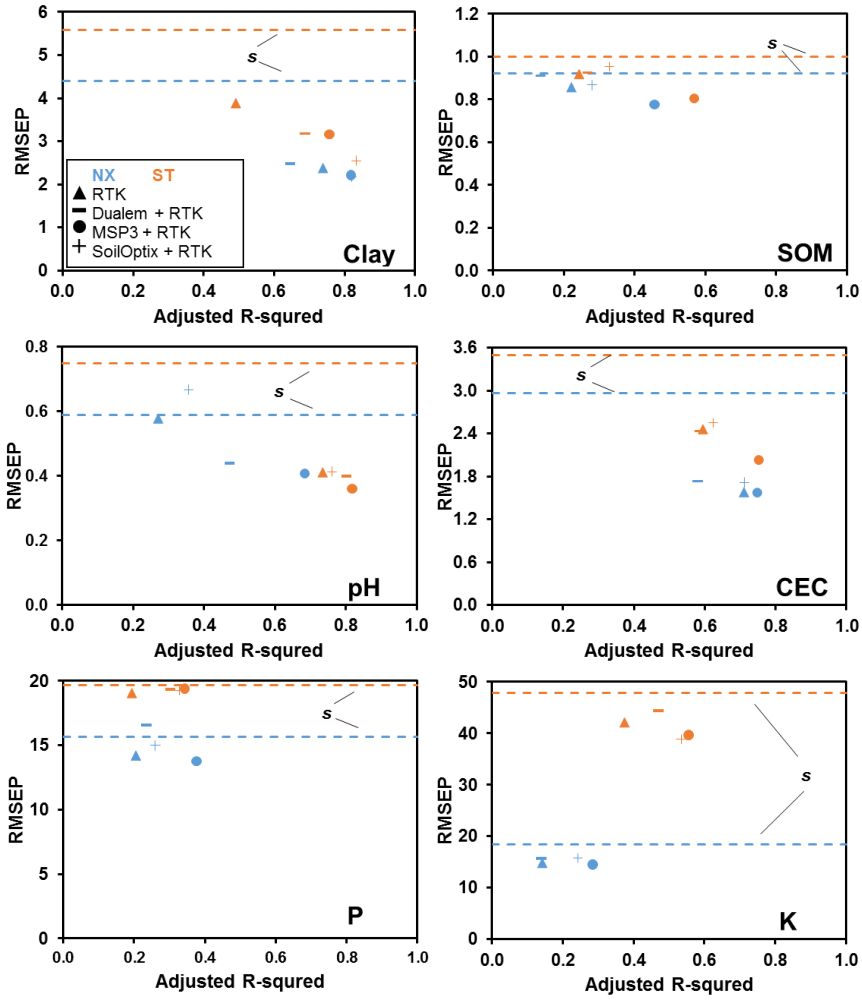


Figure 10. Histograms showing the patterns of sensor measurements at sampled locations with respect to the entire field for different sensor variables at NX and ST.

Figure 11 shows the comparison of predictability of RTK only, Dualem + RTK, MSP3 + RTK and SoilOptix + RTK for clay content, SOM, soil pH, CEC, P and K for NX and ST fields. In general, MSP3 + RTK performed the best in terms of lower RMSEP and higher adjusted R-squared. The only exception was the clay prediction at ST where SoilOptix + RTK performed the best. All prediction models resulted in model error (i.e., RMSEP) lower than standard deviation (denoted as *s*) of measurements, except the prediction of soil pH using SoilOptix + RTK for NX and the prediction of P using Dualem + RTK for NX field. In most cases, RTK alone as well as SoilOptix + RTK showed the poorest performance. As for the prediction quality, clay, pH and CEC predictions were excellent with the highest adjusted $R^2 \approx 0.8$. Yet, the four mapping systems poorly estimated P and K content, in particular for P with adjusted R-squared < 0.4. The quality of K prediction using Dualem + RTK or SoilOptix + RTK for ST was moderated with adjusted $R^2 \approx 0.6$. Except MSP3 + RTK, the quality of SOM prediction using the rest of systems was not great.



s: standard deviation

Figure 11. Comparing the predictability of the four mapping systems 1) RTK, 2) Dualem + RTK, 3) MSP3 + RTK and 4) SoilOptix + RTK for the six measured soil properties: Clay, SOM, pH, P and K.

Conclusion

According to the results, MSP3 with topographic attributes from RTK provided very promising predictability for several soil properties of interest. Although the EC_a related measurements of both mapping systems were strongly correlated, the Dualem system with RTK did not perform as well as MSP3. This result suggests the advantage of fusing various types of sensor measurements to optimize soil characterization. Therefore, further study will focus on developing the calibration strategies for a MSP3 and RTK combined system to improve the quality of soil thematic maps.

Acknowledgement

This study was funded by the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA) New Directions Research Program (NDRP). Special thanks to Mr. Paul Hermans of DuPont Pioneer (Richmond, Ontario, Canada) for conducting MSP3 mapping and to Mr. Paul Raymer of Practical Precision, Inc. (Tavistock, Ontario, Canada) for conducting SoilOptix mapping. Great

appreciation are given to the cooperating farmers Jeremy Nixon of J & H Nixon Farms and Michael Schouten of Schouten Dairy Farms.

References

Adamchuk, V.I., Lund, E.D., Reed, T.M., & R.B. Ferguson. (2007). Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture*, 8(3), 139–149.

Adamchuk, V.I., Viscarra Rossel, R.A., Marx, D.B., & Samal, A.K. (2011). Using targeted sampling to process multivariate soil sensing data. Geoderma, 163(1): 63–73.

De Benedetto, D., Castrignanò, A., Rinaldi, M., Ruggieri, S., Santoro, F., Figorito, B., Gualano, S., Diacono, M., & Tamborrino, R. (2013). An approach for delineating homogeneous zones by using multi-sensor data. *Geoderma*, 199, 117–127.

Dhawale, N., Adamchuk, V.I., Huang, H-.H., Ji, W., Lauzon, S., Biswas, A. & Dutilleul, P. (2016). Integrated analysis of multilayer proximal soil sensing data. Paper presented at the *13th International Conference on Precision Agriculture*, July 31-August 4, St. Louis, Missouri, USA. International Society of Precision Agriculture (ISPA), 2016.

Heege, H.J. (2013). Sensing of natural soil properties. In H.J. Heege (Ed.) *Precision in crop farming* (pp. 51–99). Netherlands: Springer.

Kuang, B., Mahmood, H.S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & Van Henten, E. J. (2012). Sensing soil properties in the laboratory, in situ, and on-line. *Advances in Agronomy,* 114, 155–223.

Lund, E.D. (2011). Proximal sensing of soil organic matter using the Veris OpticMapper. In: *Proceedings of the 2nd Global Workshop on Proximal Soil Sensing*. Montreal, Quebec, Canada. 2011.

Mahmood, H.S., Hoogmoed, W.B., & Henten, E.J. (2012). Sensor data fusion to predict multiple soil properties. *Precision Agriculture*, 13, 628–645.

Piikki, K., Söderström, M., & Stenberg, B. (2013). Sensor data fusion for topsoil clay mapping. *Geoderma*, 199, 106–116.

Piikki, K., Wetterlind, J., Söderström, M., & Stenberg, B. (2015). Three-dimensional digital soil mapping of agricultural fields by integration of multiple proximal sensor data obtained from different sensing methods. *Precision Agriculture*, 16(1), 29–45.

Srinivasan, A. (Ed.) (2006). *Handbook of precision agriculture – Principles and applications*. New York: The Haworth Press Inc.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., & Lobsey, C. (2011). Proximal soil sensing: an effective approach for soil measurements in space and time. *Advances in Agronomy*, 113, 243–91. Elsevier.

Viscarra Rossel, R.A., McBratney, A.B., & Minasny, B., (2010). *Proximal soil sensing*. Netherlands: Springer.