

## **Simultaneous measurement of multiple soil properties through proximal sensor data fusion: a case study**

### **ABSTRACT**

In this research, proximal soil sensor data fusion was defined as a multifaceted process which integrates geospatially correlated data, or information, from multiple proximal soil sensors to accurately characterize the spatial complexity of soils. This has capability of providing improved understanding of soil heterogeneity for potential applications associated with crop production and natural resource management. To assess the potential of data fusion for the purpose of improving thematic soil mapping over the single sensor approach, data from multiple proximal soil sensors were combined to develop and validate predictive relationships with laboratory-measured soil physical and chemical properties. The work was conducted in an agricultural field with both mineral and organic soils. The integrated data included: topography records obtained using a real-time kinetic (RTK) global navigation satellite system (GNSS) receiver, apparent soil electrical conductivity (ECa) obtained using an electromagnetic induction sensor, and content of several naturally occurring radioisotopes detected using a mobile gamma-ray spectrometer. In addition, the soil profile data were collected using a commercial ruggedized multi-sensor platform carrying a visible and near-infrared (vis-NIR) optical sensor and a galvanic contact soil ECa sensor. The measurements were carried out at predefined field locations covering the entire study area identified from sensor measured a priori information on field elevation, ECa and gamma-ray count. The information was used to predict: soil organic matter (SOM), pH, lime buffer capacity (LBC), as well as concentration of phosphorus (P), potassium (K), calcium (Ca), magnesium (Mg), and aluminum (Al). Partial least squares regressions (PLSRs) were used to predict soil properties from individual sensors and different sensor combinations (sensor data fusion). By integrating the data from all of the proximal soil sensors, SOM, pH, LBC, Ca, Mg, and Al were predicted simultaneously with  $R^2 > 0.5$  ( $RPD > 1.50$ ). Improved predictions were observed for most soil properties based on sensor data fusion than those based on individual sensors. After choosing the optimal sensor combination for each soil property, the predictive capability was compared using different data mining algorithms, including support vector machines (SVM), random forest (RF), multivariate adaptive regression splines (MARS), and regression trees (CART). Improved predictions for SOM, Ca, Mg, and Al were observed using SVM over PLSR. The predictive capability was followed by RF and MARS, with CART. Predictions of pH and LBC were only feasible using MARS and PLSR, respectively. In this field, it was not possible to predict extractable P and K using all tested sensor combinations or algorithms. With large variability in SOM, the field presents a special situation and thus, the result could be specific to the study site. Further research includes an extended number of experimental sites covering different geographic areas around Eastern Canada.

**Keyword:** Vis-NIR spectroscopy; Gamma-ray; Apparent electrical conductivity; Sensor fusion; Data fusion; Partial least square regression; Data mining algorithms