IMPLEMENTATION OF PROXIMAL SOIL SENSING, DATA FUSION AND MACHINE LEARNING TO IMPROVE PHOSPHORUS MANAGEMENT AT A FIELD SCALE

A. Lachgar, D. J. Mulla and V. Adamchuk
University of Minnesota, Saint Paul, USA and
McGill University, Montreal, Canada
lachg001@umn.edu
Abdelkrim.LACHGAR@um6p.ma

ABSTRACT

In the context of a rapid increase in phosphorus (P) fertilizers prices, new techniques are needed for geospatial predictions of soil P for improved P fertilizer management, while increasing farmer profitability and reducing environmental concerns. One of the biggest issues in site-specific phosphorus management is the substantial spatial variability in plant available P across fields. This leads to an expensive and laborious process for accurate mapping soil P using a traditional soil sampling and laboratory analysis approach. To overcome this barrier, emerging sensing and data interpretation technologies should be employed to accurately assess spatial heterogeneity of P within fields, and to help farmers optimize mineral P fertilization recommendations.

In this study, we applied machine learning algorithms and novel data fusion concepts to analyze integrated high-density spatial data layers related to the potential P availability to plants. Machine learning algorithms were used to evaluate the relative importance of different auxiliary soil properties at predicting plant-available P. These variables were used for modeling P spatial distribution. High-density data mining techniques and various sensor data fusion algorithms and optimization techniques were used to predict P spatial distribution and identify site-specific management zones. Supervised machine learning algorithms such as Support Vector Machine was used, along with sensor data fusion.

Auxiliary data to predict available P included high-density apparent soil electrical conductivity (ECa), gamma-ray spectrometry, high-resolution topography as well as soil test data from a field in Southwestern Ontario, Canada. Spatial maps were prepared using elevation data based on a Real-Time Kinematic (RTK) global navigation satellite systems (GNSS) receiver, DUALEM-21S sensing, a gamma-ray (SoilOptix) spectrometer and laboratory analysis of soil samples. An ATV was used to collect on-the-go measurements using this proximal soil sensing equipment.

The machine learning models successfully predicted P and generated high resolution maps showing P status zones and recommendation maps at the local level. The results of supervised regressions and reclassifications modeling represent a robust decision tool for phosphorus precision nutrient management and variable rate technology. Regression machine learning models were turned into classification models. Results of this study show that radial basis kernel trick SVM regressor trained and validated with 10 fold repeated cross validation performed the best. In fact, a cost value equal to 1, sigma=0.08 and epsilon=0.1, provides an R-squared of 0.53 and 0.54, an RMSE of 8.79 and 8.98 for SVM K-fold cross validation and repeated cross-validation

respectively. Therefore, the choice of repeated cross-validation slightly improved our available P spatial predictions.

Environmental sensor covariate fusion combined with spatial machine learning algorithms is a useful tool to help farmers save P fertilizer and preserve environmental resources through understanding the available P spatial variability across the study area. Future research should focus on combining prediction results based on different machine learning algorithms and integrate them with traditional regression techniques.

Keywords: *Proximal soil sensing, Geospatial predictions, spatial variability, data mining, data fusion, modeling, supervised machine learning algorithms, support vector machine.*

INTRODUCTION

A world population expected to reach 9.1 billion by 2050 requires a rise in food production by 70% (FAO, 2009). However, soils are becoming degraded in many parts of the world, meaning that agriculture can only occur in a limited space. Thus, it is essential that this limited space be accompanied by an increase in soil productivity with respect to sustainable development and environmental protection. If soils at agricultural sites are deficient in P, it can lead to up to 15% yield losses (Mahdi et al., 2012). To overcome this issue, farmers will often add P fertilizers to adjust soil fertility and maximize yields. Similarly, yield can be improved in field' zones where fertilizer application is needed and reduce over-fertilization was beneficial in limited yield potential zones (Simard et al., 1998).

When applying P fertilizer on fields, farmers will typically apply a uniform P rate. However, soil P spatial variability in field scale tends to be high in most places (Hong-xia et al., 2010). Thus, applying a uniform P rate neglects P spatial variation within soils and can lead to over or under application of fertilizers. This increases production input costs, and excess P may runoff to surrounding hydrologic ecosystems (Fang et al., 2002). One solution to this issue is to adopt precision agriculture, which involves dividing fields into zones and delivering customized management to each zone (Mulla and Miao, 2016). The goal of precision nutrient management is to ensure that each zone in the field receives a rate of fertilizer that is specific to that zone's requirement (Mulla, 1991; 1993).

Precision P fertilization managements is still not considered in most arable land. As phosphorus fertilizer prices have been consistently increasing, and concerns over an efficient use of a limited resource are rising, spatial predictions methods and new innovative P sensing models within agricultural fields are required to optimize the use of P fertilizer with regards to spatial variability for some improved P variable rates prescriptions.

Data availability and collection represents major concerns over the past two decades to boost soil productivity, take rational management decisions, and reduce fertilizers' leaching and pollution. Yet, the development of sensor technology is making a great difference by offering voluminous and diverse data to farm practitioners and agricultural policy experts (Nawar et al., 2017). However, no commercial sensor exists for on-the-go available phosphorus measurements. Only reflectance techniques appear to be near this goal (Sinfield et al., 2010). On the other hand, Adamchuck et al. (2011) pointed out the plausible low accuracy of a single sensor and the need for sensor fusion. Integration of different proximal soil sensors could provide farmers with robust soil properties predictions (Adamchuck et al., 2011). Data integration plays a key role in producing effective machine learning models by incorporating data from a variety of sources (Dong., 2018). Therefore, fusion of suitable sensors data in the process of variable rates management decisions should be tested to determine whether it will improve farmer’s profitability and reduce environmental impacts.

Digital soil mapping is becoming affordable thanks to recent advances in machine learning during the last few decades (McBratney et al., 2003). Machine learning algorithms are increasingly being used in the precision agriculture community to map soil properties and generate spatial predictions (Brungard et al. 2015; Thorsten et al. 2018). Digital soil mapping with supervised or unsupervised machine learning is needed to understand spatial variability and improve models’ performances (Grunwald et al., 2006; Kim et al., 2012; Boruvka et al., 2008). Machine learning models have been used to decipher spatial patterns in a given dataset and produce useful predictions (Witten et al., 2011). They have been used to predict soil horizons and soil types (Jafari et al, 2012), soil types and topographic attributes (Silveira et al., 2013; Dobos et al., 2000), soil organic carbon (Nawar et al., 2020), soil cation exchange capacity by fusing PXRF and vis-NIR datasets (Mengxue et al., 2020). Moreover, spatial distribution of topographic attributes is inherently impacting landscapes and consequently soil properties. This allows soil scientists to characterize soils and produce pedological maps (Klingebiel et al., 1987; Moore et al., 1991).

Int this study, we aim to infer available phosphorus spatial pattern by fusing a variety of high-dense sensor data and using machine learning algorithms to predict phosphorus values in the soil of unsampled locations. The sensors we focused on include gamma-ray, Dualem-21s and topographic attributes. We explored the importance of these sensor’s variables using the supervised Support Vector Machine algorithm. Model predictions were then used to produce accurate digital soil maps portraying spatial predicted available phosphorus across the study site. Ultimately, the main goal of this research is to test the capability of machine learning algorithms to assist us in predicting phosphorus status in space and guide farmers in adopting precision phosphorus management.

MATERIALS AND METHODS

Data collection

Soil sampling and proximal soil sensing occurred in a field located at Ontario, Canada. Proximal soil sensors were used to map this field. The soil texture is mainly loam which allows suitable drainage conditions. Figure 1 illustrates the site location and boundaries, soil sampling locations, gamma-ray sensor and soil apparent electrical conductivity (ECa) readings.

Figure 1: Study area in Ontario, Canada: Soil sampling locations, Soil apparent electrical conductivity (ECa) with Dualem sensor, Gamma -ray sensor readings, Elevations (RTK GPS).

Geostatistical analysis and sensor data fusion

Empirical Bayesian kriging (EBK) was used to produce geostatistical raster layers with high-dense proximal soil sensor data (i.e. Dualem and Gamma) and topographic attributes (DEM, slope and aspect) (ESRI, ArcGIS Pro). Figure 2 illustrates examples of raster layers produced based on high-dense soil sensors readings.

Raster to points function in ArcGIS (version 10.8.1; ESRI, Redlands, California, USA) was used to extract and match locations between soil sampling and proximal soil sensing datasets based upon the geostatistical raster layers. This allowed us to proceed to sensor data fusion and integrate all collected data in space. The fused spatial dataset was then considered to be a ground truth dataset which we used during the training and validation of our machine learning spatial prediction models

Composite band rasters were produced by overlaying all geospatial covariates issued from each proximal soil sensor variables. This composite bands raster stack represents the predictors of available P spatial distribution across the study site, using machine learning models which were developed in this study. Table 1 describes sensor covariate raster map information; raster dimensions, resolution, extent and coordinate reference system.

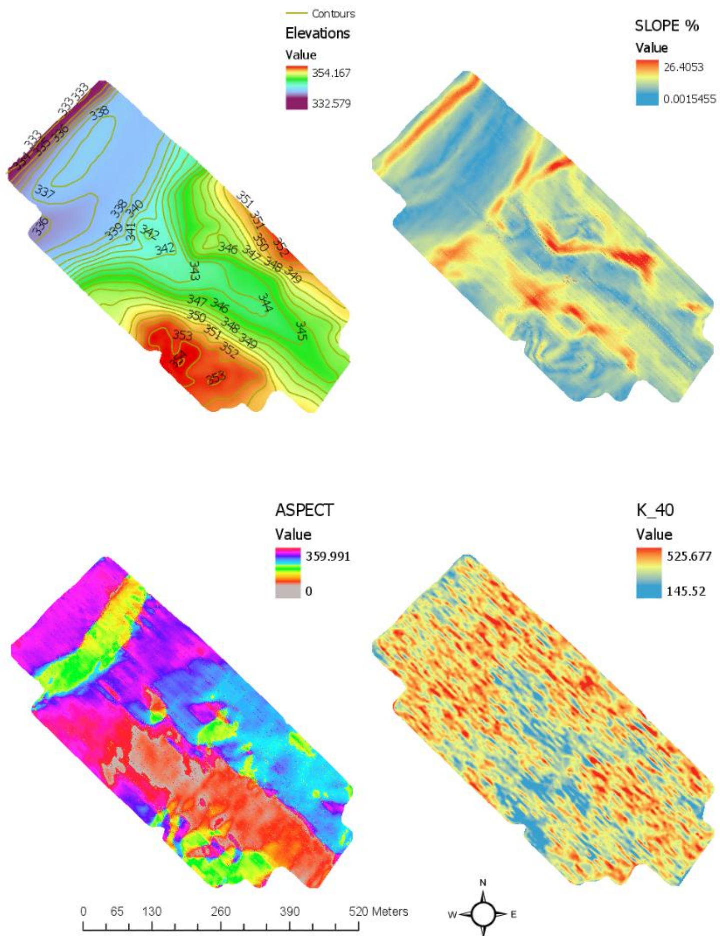


Figure 2: Empirical Bayesian kriging geostatistical raster layers of topographic attributes (DEM, Aspect and slope), and Gamma-ray (K_40)

Table 1: Predicted available P raster information

Dimensions	3129, 3227, 10097283 (number of rows, number of columns, number of cells)
Spatial resolution	0.2*0.2 (X, Y)
Extent	566130, 566775.4, 4824739, 4825364 (Xmin, Xmax, Ymin, Ymax)
coordinate reference system	UTM NAD83 Zone 17

RESULTS AND DISCUSSION

Machine learning and model's selection

Predictor raster bricks are made of environmental sensor covariates. These raster bricks are used to predict available P in the same extent and at the level of each cell center with Support

Vector Machine algorithms we trained and validated using cross validation and repeated cross validation. Results of spatial predictions demonstrated that the north west part of the field is deficient in available P, and the rest of the field does not need P fertilizer.

Reclassification of P spatial predictions

The results of the reclassifications represent a useful decision tool for precision P nutrient management. Validated machine learning models were turned into classifications models. We split the raster cells values based upon predicted available P across the study area and the official P fertilization requirements in Canada (Reid, K., 2006) which requires P fertilization when it is below 30ppm (Figure 3). These classification models should be useful in terms of adopting P variable rates technology. In this study, farmers should be able to decide whether or not P fertilizer is required. Also, farmers could navigate over the study site using GPS/RTK equipment and determine which specific areas need to be fertilized, and avoid fertilizing when it is not needed.

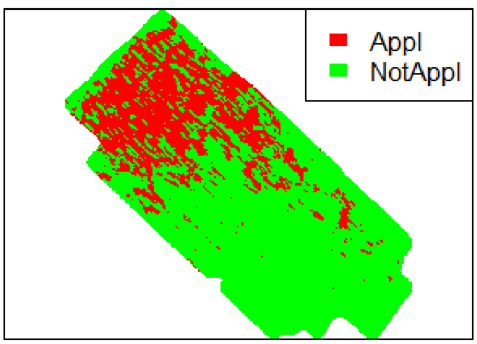
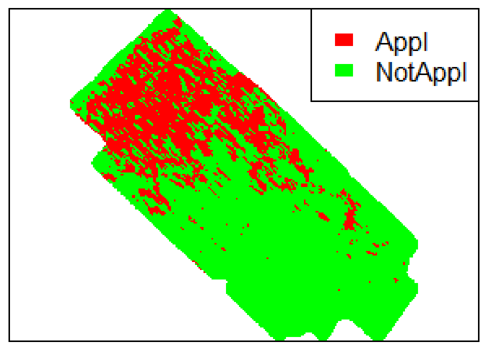
<table>
<tr><td>SVM regressor / K-fold cross validation</td><td>SVM regressor / Repeated cross validation</td></tr>
<tr><td>
</td><td>
</td></tr>
</table>

Figure 3: Re-classification of the predicted available P cell values based on two classes; "Appl": P amendments are required and "NotAppl": P amendments are not needed.

SVM is known to handle both linear and no-linear dataset and separations boundaries (support vectors). The cost function controls errors and margins, while creating the optimum hyperplane to separate data observations. In this study, radial kernel trick SVM regressor trained and validated with repeated cross-validation performed well in terms of spatial predictions of available P. A cost value equal to 1, sigma=0.08 and epsilon=0.1, provides an R-squared of 0.53 and 0.54, an RMSE of 8.79 and 8.98 for SVM K-fold cross validation and repeated cross-validation respectively. Therefore, the choice of repeated cross-validation improved slightly our available P spatial predictions (Table 2).

Table 2: Synthesis of the validated models and hyperparameter tuning

Models	Train control strategies	Hyperparameter tuning	RMSE	R-squared	MAE
Support Vector Machine	K-fold cross-validation	Cost= 1 Sigma = 0.08 Epsilon= 0.1	8.79	0.53	7.17
	Repeated cross-validation	Cost= 1 Sigma = 0.08 Epsilon= 0.1	8.98	0.54	7.20

Adoption of precision P nutrient management has the potential to be cost effective because farmers will greatly reduce their use of P fertilizer tremendously, especially in locations where P amendments are not needed. Further, variable rate technology should be environmentally friendly by reducing over-fertilization and surface water eutrophication. Thus, these strategies should improve soil health and reduce environmental concerns.

REFERENCES

Adamchuck, V. I., Viscarra Rossel, R. A., Sudduth, K. A., & Lammers, P. S. (2011). Sensor fusion for precision agriculture. In C. Thomas (Ed.), Sensor fusion—foundation and applications (pp. 27–40).

Behrens, Thorsten, Karsten Schmidt, Robert A MacMillan, and Raphael A Viscarra Rossel. 2018. "Multi-Scale Digital Soil Mapping with Deep Learning." *Scientific Reports* 8 (October). Nature Publishing Group:15244. https://doi.org/10.1038/s41598-018-33516-6.

Boruvka, L., Pavlu, L., Vasat, R., Penizek, V., Drabek, O., 2008. Delineating acidified soils in the Jizera Mountains region using fuzzy classification. In: Hartemink, A.E., McBratney, A., Mendonça-Santos, M. de L. (Eds.), Digital Soil Mapping With Limited Data. Springer, Netherlands, pp. 303–309.

Brungard, Colby W, Janis L Boettinger, Michael C Duniway, Skye A Wills, and Thomas C Edwards Jr. 2015. "Machine Learning for Predicting Soil Classes in Three Semi-Arid Landscapes." *Geoderma* 239. Elsevier:68–83.

Dobos, Endre, Micheli, Erika, Baumgardner, Marion F, Biehl, Larry, & Helt, Todd. (2000). Use of combined digital elevation model and satellite radiometric data for regional soil mapping. Geoderma, 97(3-4), 367 391.

Dong, X.L. and Rekatsinas, T., 2018, May. Data integration and machine learning: A natural synergy. In *Proceedings of the 2018 international conference on management of data* (pp. 1645-1650).

ESRI, ArcGIS Pro. What is empirical Bayesian kriging?
URL https://pro.arcgis.com/en/pro-app/help/analysis/geostatistical-analyst/what-is-empirical-bayesian-kriging-.htm

Fang, F, Brezonik, P. L, Mulla, D. J, & Hatch, L. K. (2002). Estimating Runoff Phosphorus Losses from Calcareous Soils in the Minnesota River Basin. Journal of Environmental Quality, 31(6), 1918-1929.

FAO. (2009). Feeding the world in 2050 - FAO. Retrieved from ftp://ftp.fao.org/docrep/fao/meeting/018/k6021e.pdf

Grunwald, S. Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics (Ed.); CRC Press: Boca Raton, 2006; https://doi.org/10.1201/9781420028188.

Irfan, M., Aziz, T., Maqsood, M.A. et al. Phosphorus (P) use efficiency in rice is linked to tissue-specific biomass and P allocation patterns. Sci Rep 10, 4278 (2020). https://doi.org/10.1038/s41598-020-61147-3

Jafari, A., Finke, P.A., Van de Wauw, J., Ayoubi, S., Khademi, H., 2012. Spatial prediction of USDA-great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. Eur. J. Soil Sci. 63, 284–298. http://dx.doi.org/10.1111/j.1365-2389.2012.01425.x.

Kim, J., Grunwald, S., Rivero, R.G., Robbins, R., 2012. Multi-scale modeling of soil series using remote sensing in a wetland ecosystem. Soil Sci. Soc. Am. J. 76, 2327–2341. http://dx.doi.org/10.2136/sssaj2012.0043.

Klingebiel, A.A., E.H. Horvarth, D.G. Moore, W.U. Reybold. (1987. Use of slope, aspect, and elevation maps derived from digital elevation model data in making soil surveys. Soil Science Society of America. SSSA Special Publication, 20 (1987), pp. 77-90

Kuhn, M. (2008). Caret package. Journal of Statistical Software, 28(5)

Liaw A, Wiener M (2002). "Classification and Regression by randomForest." R News, 2(3), 18-22. URL http://CRAN.R-project.org/doc/Rnews/.

Mahdi S, Talat M, Dar M, Hamid A, Ahmad L. 2012. Soil phosphorus fixation chemistry and role of phosphate solubilizing bacteria in enhancing its efficiency for sustainable cropping - a review. Journal of Pure and Applied Microbiology. 6(4):1905-1911.

McBratney, A.B, Mendonça Santos, M.L, & Minasny, B. (2003). On digital soil mapping. Geoderma, 117(1-2), 3-52.

Moore, I. D., G.A. Gessler, G.A. Peterson. 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal, 57 (1993), pp. 443-452

Mulla, D. J. 1991. Using geostatistics and GIS to manage spatial pat- terns in soil fertility. In: Kranzler, G. (ed.), Automated Agriculture for the 21st Century. ASAE, St. Joseph, MI, pp. 336–345.

Mulla, D. J. 1993. Mapping and managing spatial patterns in soil fertility and crop yield. In: Robert, P., W. Larson, and R. Rust (eds.), Soil Specific Crop Management. ASA, Madison, WI, pp.15-26.

Mulla, D. J. and Miao, Y. (2016) 'Near infrared Variable rate technology'.

Nawar, S. et al. (2017) Delineation of Soil Management Zones for Variable-Rate Fertilization: A Review. 1st edn, Advances in Agronomy. 1st edn. Elsevier Inc. doi: 10.1016/bs.agron.2017.01.003.

Nawar, Said, Munnaf, Muhammad Abdul, & Mouazen, Abdul Mounem. (2020). Machine Learning Based On-Line Prediction of Soil Organic Carbon after Removal of Soil Moisture Effect. Remote Sensing (Basel, Switzerland), 12(8), 1308.

Reid, K., 2006. Soil fertility handbook, publication 611. Toronto, Ontario. Retrieved November, 20, p.2018.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Silveira, C. T. et al. (2013) 'Soil prediction using artificial neural networks and topographic attributes', Geoderma. Elsevier B.V., 195–196, pp. 165–172. doi: 10.1016/j.geoderma.2012.11.016.

Simard RR, Nolin MC, Cambouris AN (1998) Application of precision farming to potato production in Québec. Better Crops 82(2):22–24

Sinfield, J. V., Fagerman, D. and Colic, O. (2010) 'Evaluation of sensing technologies for on-the-go detection of macro-nutrients in cultivated soils', Computers and Electronics in Agriculture, 70(1), pp. 1–18. doi: 10.1016/j.compag.2009.09.017.

Wan, Mengxue, Hu, Wenyou, Qu, Mingkai, Li, Weidong, Zhang, Chuanrong, Kang, Junfeng, . . . Huang, Biao. (2020). Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. Geoderma, 363, 114163.

Witten, I.H., Frank, E., Hall, M.A., 2011. Data mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann, Burlington.

Zhu Hong-xia, & Chen Xiao-min. (2010). Spatial Variability of Soil Phosphorus Based on Geostatistics. *2010 International Conference on Multimedia Technology*, 1-5.