Identifying Key Factors Influencing Yield Spatial Pattern and Temporal Stability for Management Zone Delineation

Lorena N Lacerda[1], Yuxin Miao[1], Katsutoshi Mizuta[1], Kirk Stueve[2]

[1]Soil, Water, and Climate Department, University of Minnesota, Saint Paul, MN, USA

[2]Ceres Imaging, Inc., Oakland, CA, USA

A paper from the Proceedings of the
15th International Conference on Precision Agriculture
June 26-29, 2022
Minneapolis, Minnesota, United States

Abstract. *The objective of this study is to use machine learning models to identify key soil and landscape properties affecting yield spatial patterns and yield temporal stability for management zone delineation and to evaluate the consistence of these factors in different prediction models. The study was carried out in a 44 ha corn-soybean rotation field in western Minnesota, USA. Yield maps from 7 years collected from 2014 to 2020 were used to create yield spatial trend (YST; average normalized yield map) and yield temporal stability maps (YTS; coefficient of variation map). In the complete dataset, 29 different soil and landscape properties were used as input in the machine learning models including relative elevation, slope, curvature and aspect, calculated from LiDAR elevation data at 1 m resolution downloaded from the MN TOPO website; topographic wetness index and soil brightness index calculated from PlanetScope images at 3 m spatial resolution; soil physical properties, and macro and micronutrients collected with SoilOptix, a high-resolution soil mapping system; and shallow and deep electrical conductivity. A farmer-friendly dataset was also tested using mostly variables that are available online and that can be easily accessed by farmers. All maps were interpolated to a 3 m grid using kriging. Prediction models for YST and YTS were created using random forest, support vector machine and XGBoost algorithms. To identify features that were relevant for the models, Boruta algorithm was used for feature selection. Once features were selected based on importance, Spearman correlation was used to exclude features that were highly correlated to each other to avoid redundance. Results showed that while all features were deemed important, relative elevation was the most relevant factor influencing both YST and YTS. In the farmer-friendly dataset brightness index was the most important feature for YST, and relative elevation was the most important for YTS. Other attributes such as slope, iron, sulfur, potassium and calcium soil concentrations and soil organic matter were also among the most important factors for both YST and YTS. Random forest (RF) was the best performing model among all models and test sets for both response variables.*

Lacerda, L.N., Miao, Y., Mizuta, K., & Stueve K. (2022). Identifying key factors influencing yield spatial pattern and temporal stability for management zone delineation. In Proceedings of the 15th International Conference on Precision Agriculture (online). Minneapolis, MN: International Society of Precision Agriculture.

Keywords. *Yield Spatial Trend, Yield Temporal Stability, Machine Learning, Soil Landscape Properties, Feature Selection, Management Zone Delineation.*

Introduction

The identification and division of homogenous subareas within a field for site specific management is the basic premise of precision agriculture (Brock et al. 2005; Doerge 1999). These homogenous areas are commonly known as management zones and are used to indicate where different rates of inputs may be applied. The delineation of management zones has been extensively explored, but there is still a posing challenge to its use in precision agriculture applications due to the numerous factors involved in limiting crop yield such as soil and landscape properties, biotic factors, and weather conditions. Thus, identifying and selecting the key properties driving crop yield and using them as base for management zone delineation is of the utmost importance (Sanches et al. 2019; Chang et al. 2014).

The most common approaches for management zone delineation are the use of individual or multiple soil properties (Wang et al. 2021; Scudiero et al 2013; Shaner et al. 2008), crop performance (Blackmore et al 2003), and a combination of soil and crop attributes (Miao et al. 2018). Another important step in delineating management zones is the selection of the statistical methods to group these homogenous sub-areas in groups or clusters. Fridgen et al. (2004) pointed out the need for an easy-to-use software to help in the decision-making process and developed the management zone analyst software (MZA). Due to its user-friendly interface, its capabilities to classify data into zones and to inform the user the optimal number of potential zones, MZA became widely used. Brock et al. (2005) used MZA to delineate zones based on multiple years of yield data from four different corn-soybean rotation systems. MZA was also used to delineated zones for smallholder farms in Brazil using remote sensed data (Breunig et al. 2020). Jiang et al. (2011) used MZA to delineate management zones based on multiple measured soil properties in China.

In addition to the statistical methods and software used to divide data variables into zones, other mathematical principals are often used as the first step of management zone delineation to help with selection of key variables before performing clustering analysis. Principal component analysis is a commonly used mathematical tool to identify important factors in a larger dataset (Shukla et al. 2016; Yao et al. 2014). In more recent years, machine learning models such as random forest have been used to determine the relationship between response variable and predicting variables (Chen et al. 2017, Wang et al. 2021). Hence, the objective of this study was to use machine learning models to identify key soil and landscape properties affecting yield spatial pattern and yield temporal stability for management zone delineation and to evaluate the consistence of these factors in different prediction models.

Material and Methods

The study was carried out in a 44 ha rainfed corn-soybean rotation field (45º39’39.12” N, 96 º 18’34.93” W) in the Traverse County located in the western region of the state of Minnesota, USA (Figure 1). The majority of the field area has a silty clay loam soil with 0 to 1% slope and a silt loam with 0 to 2% slope soil types (USDA-NRCS 2022). The region is characterized by a snow climate with no defined dry season and hot summers with the temperature of the hottest month being equal or above 22 ºC (Kottek et al. 2006; Beck et al. 2018). Minimum and maximum temperature averages for the past 20 years from April to October were 3º and 31 ºC, and average total precipitation measure was 571.5 mm (NOAA 2022).

Yield maps from seven years were collected from 2014 to 2020 using a yield monitor. In the seven growing seasons between 2014 and 2020, soybean was planted for 4 years in 2015, 2016, 2018 and 2020, while corn was planted for 3 years in 2014, 2017 and 2019. A data cleaning protocol was developed to clean the yield data in order to mitigate machine errors and errors introduced by field conditions. Yield data points that exceed 33% or fell below 10% moisture were excluded

(Luck et al. 2015), as well as points with travel speed slower than 1.13 km/h and faster than 13.20 km/h. Sudden combine speed change was also used as a criterion. Points in which the speed increased or decreased by 15% compared to the previous point were also excluded. The last step of the cleaning protocol consisted of normalizing the moisture to 15% (for corn) and 13% (for soybean) and exclude any points that fell below or above 3 standard deviations from the mean. Once yield maps from all years were cleaned, the final wet yield mass adjusted to the outlined moisture was normalized by dividing each data point by the field average yield for a given year and multiplying by 100%. Normalized yield maps were then interpolated to a 3 m grid using kriging and used to create yield spatial trend (YST; average normalized yield map across 7 years) and yield temporal stability maps (YTS; coefficient of variation map) following Blackmore (2000).

Fig. 1. Location and boundary of rainfed corn-soybean rotation field used for the study. The field has an area of 44 ha and is located in the Traverse County, MN, USA.

Yield Spatial Trend (YST) and Yield Temporal Stability (YTS) prediction models

Predication models were classified in two types: complete dataset and farmer-friendly dataset. In the complete dataset, 29 different soil and landscape properties were used as input in the machine learning models including relative elevation, slope, curvature, aspect and topographic wetness index calculated from LiDAR elevation data at 3 m resolution downloaded from the MN TOPO website (http://arcgis.dnr.state.mn.us/maps/mntopo/); soil brightness index calculated from PlanetScope images at 3 m spatial resolution; soil physical properties, and macro and micronutrients collected with SoilOptix, a high-resolution soil mapping system; and shallow and deep soil electrical conductivity (Table 1). In the farmer-friendly dataset, only nine attributes were used (indicated in bold letter in table 1). Most of these attributes are either easily available to farmers and are offered for free in the sources cited above, or are attributes commonly collected by farmers in their fields, such as soil electrical conductivity.

PlanetScope satellite constellation is operated by Planet Labs (Planet Labs 2021) and provides images in the blue (455-515 nm), green (500-590 nm), red (590-670 nm) and NIR (780-860 nm) regions of the spectrum at a 3 m spatial resolution. The images downloaded from planet were part of the surface reflectance product offered by Planet, and a factor of 10,000 was applied to the individual bands to obtain correct reflectance data prior to soil index calculation (Planet 2019). All maps with different variables were interpolated to a 3 m grid using kriging tool in ArcGIS (ESRI, Redlands, CA USA).

Variable selection and machine learning prediction models

To identify features that were relevant for the models, Boruta algorithm using the Boruta package in R software (Ver. 4.1.1) was used for feature selection. The algorithm uses a random forest classifier to set a mean threshold value that will serve as a reference to classify feature importance (Liaw and Wiener 2002). Each decision tree uses an attribute for classification and its importance is calculated separately. Attribute importance is measured based on the loss of accuracy of classification, and the standard deviation and average of the accuracy loss are computed (Kursa and Rudnicki 2010). A Z score is also computed (average loss/ standard deviation), which is then used as the importance measure. Features that show importance value higher than shadow means are deemed important, and their importance increases with higher values.

Table 1. Complete list of variables used in the YTS and YST prediction models. Bold words represent the variables included in the farmer-friendly dataset.

Labels	Variables	Labels	Variables
BI	**Brightness Index**	**Ec Shallow**	**Electrical Conductivity (shallow layer)**
CEC	Cation Exchange Capacity	**Ec deep**	**Electrical Conductivity (deep layer)**
-	**Relative Elevation**	-	Loam
-	**Slope**	-	Sand
K	Potassium	Plant avail. water	Plant Available Water
S	Sulfur	-	Leakability
-	**Aspect**	**TWI**	**Topographic Wetness Index**
Fe	Iron	B	Boron
Ca	Calcium	Mn	Manganese
P	Phosphorus	Mg	Magnesium
Cu	Copper	Ca-Mg ratio	Calcium- Magnesium ratio
OM	**Organic Matter**	K-Mg ratio	Potassium- Magnesium ratio
Zn	Zinc	pH	pH
-	Clay	Silt	Silt
-	**Curvature**		

Once all features initially used were deemed important for the model, different models were tested by excluding features that were highly correlated to each other based on Spearman correlation. Data was divided into training, validation, and test sets (70, 20 and 10%, respectively). Three ML algorithms, including support vector machine (SVM), random forest (RF), and XGBoost were selected to construct prediction models for YST and YTS. Linear SVM is a Kernel-based technique that derives a linear hyperplane and separates data points into two classes, which often results in higher accuracy with non-normally distributed data than traditional stochastic models (Ivanciuc 2007). Data was preprocessed by centering and scaling the whole data before constructing models. RF consists of an arbitrary number of simple decision trees (mtry = 3 for present study) and can prevent overfit by producing a dataset with variables only important to the predicted variables for better accuracy (Were et al. 2015). XGBoost also creates trees but weights those trees higher if they achieve the better prediction based on root mean square error (RMSE) (Chen and Guestrin 2016). The algorithm provides built-in 5-fold cross-validation to avoid overfitting the model. The aforementioned algorithms were executed using the *caret* and *randomForest* package in R software (Ver.4.1.1).

Statistical analysis

The best prediction models were selected based on the highest R^2, lowest RMSE for each training, validation and test sets, and the computational time that took to run the training sets. The running time for each model was represented by the difference between the start time and the end time of each model. To calculate the start and end times the system's time was recorded immediately before and after the model was running.

Results and discussion

Yield Spatial Trend (YST)

The feature selection process results for the YST model using Boruta algorithm is shown in figure 2. Despite the different levels of importance, all attributes were deemed relevant for YST prediction. Among the top 10 most important attributes, landscape features and macro and micronutrients were the major factors for variability in spatial yield potential (Figure 2A). Relative elevation was the most important attribute with an importance score over 60, slope was the third most important attribute, while aspect was the tenth. Landscape properties such as elevation and slope affect yield spatial variability because they are determinant factors of water flow and accumulation in the soil (Pachepsky et al. 2001), which drives soil water availability.

Among the soil nutrients, iron (Fe) was the most important micronutrient and second most important attribute affecting yield spatial trend. Addressing field Fe deficiency is very challenging because of the influence of other soil conditions (Godsey et al. 2003) such as soil moisture and pH, in addition to the economically feasibility of applying Fe fertilizers for the whole field. However, site-specific Fe application can help increase yield in areas that suffer from a more limiting deficiency. Potassium (K) was the most important macronutrient and third among all nutrients available. Areas with higher K available in the soil can achieve higher corn yields than areas with K deficiency (Hussain et al. 2007). Organic matter (OM) was the only soil property within the first 10. The lowest importance scores were observed for the potassium-magnesium ratio, zinc, clay, silt, sand, and loam attributes. Similar to the complete dataset, all attributes in the farmer-friendly dataset were considered important (Figure 2B). The most important variables were BI, slope, and relative elevation, while the least important were soil Ec deep and shallow.

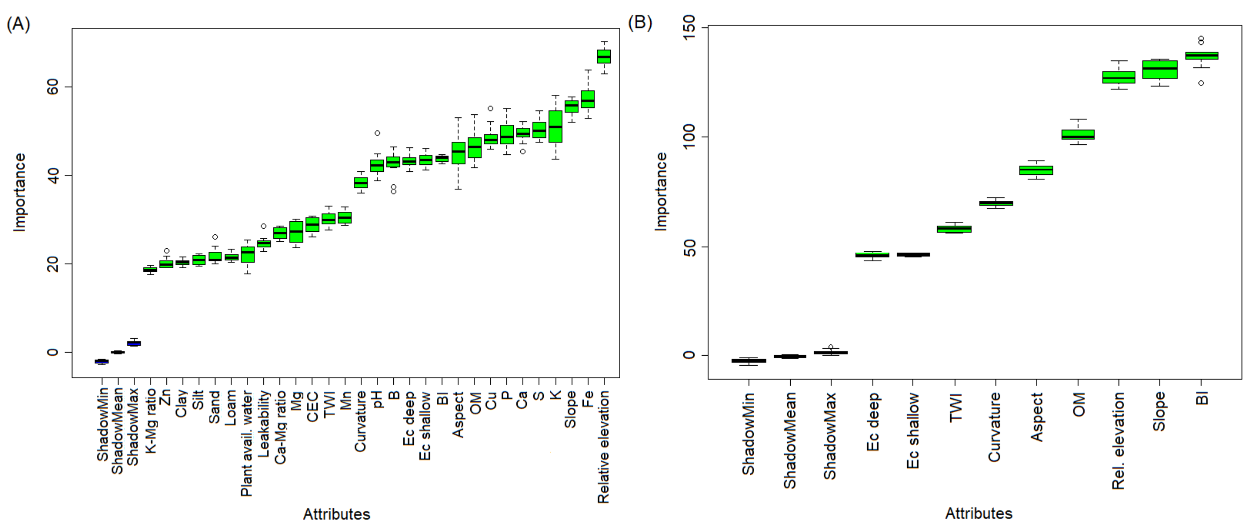

Figure 2. Boruta algorithm feature selection results for yield spatial trend (YST) prediction using the complete dataset (A), and the farmer-friendly dataset (B).

Table 2 shows the complete dataset results of the different machine learning YST prediction models testing different combinations of attributes. The random forest model showed the best performance among the three models in all tests. The XGBoost model also showed a good performance with the R^2 values in all tests above or equal to 0.68, while SVM models showed a high R^2 value only in test 1, in which all attributes were used. Test 1 had the highest R^2 values and lowest RMSE among all tests. Computational times for RF and XGBoost were slightly higher in this test due to the significantly bigger dataset, and significantly higher time for SVM.

To minimize the number of attributes used in the prediction model, attributes were excluded based on their importance and correlation to other attributes. Results showed that that R^2 values and running time decreased as more attributes were excluded, while RMSE increased. The random forest model on test 2 showed similar performance as test 1, while having only one third of the number of attributes. Using the top 10 most important variables in the RF model resulted in R^2 values of 0.92, 0.93 and 0.93, and RMSE values of 3.03, 3.07 and 2.98 for training, validation and test sets, respectively.

Table 2. Machine learning YST prediction models using support vector machine (SVM), random forest (RF) and eXtreme gradient boosting (XGBoost) models for the complete dataset. Different attribute combinations were tested to select the best performing model.

Machine learning algorithm	Test 1[a]							Test 2[b]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.68	6.05	0.69	6.17	0.67	6.10	135	0.23	9.50	0.24	9.74	0.23	9.44	23.6
RF	0.94	2.53	0.95	2.54	0.95	2.52	4.9	0.92	3.03	0.93	3.07	0.93	2.98	4.1
XGBoost	0.92	2.99	0.93	2.83	0.93	2.84	15.8	0.89	3.56	0.91	3.39	0.91	3.33	13.0

Machine learning algorithm	Test 3[c]							Test 4[d]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.19	9.68	0.2	9.95	0.19	9.72	11.5	0.23	9.44	0.23	9.71	0.23	9.44	16.5
RF	0.76	5.2	0.78	5.30	0.77	9.72	4.0	0.89	3.55	0.90	3.61	0.90	3.58	3.91
XGBoost	0.68	6.04	0.7	5.99	0.69	5.9	7.6	0.79	4.88	0.81	4.85	0.79	4.85	3.4

[a]Test 1: all attributes.
[b]Test 2: top 10 most important attributes.
[c]Test 3: top 5 most important attributes.
[d]Test 4: top 7 most important attributes.

The prediction model results for the second dataset are shown in table 3. The overall performance of all models was lower than the ones from the complete dataset, most likely due to the lower number of variables. However, accuracy was still high in tests 1 and 2. The best overall model was the RF in test 2, which showed the highest R^2 and lowest RMSE than any other model. Model accuracy was R^2= 0.84, 0.85 and 0.84 for training, validation, and test sets, respectively. The RMSE observed for the three sets were all equal or below 4.32. The RF model also had the fastest running time with 3.9 minutes. The lowest accuracies were seen in the fourth test, in which only the three top attributes were used.

Table 3. Machine learning YST prediction models using support vector machine (SVM), random forest (RF) and eXtreme gradient boosting (XGBoost) models for the farmer-friendly dataset. Different attribute combinations were tested to select the best performing model.

Machine learning algorithm	Test 1[a]							Test 2[b]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.63	6.54	0.64	6.65	0.62	6.58	23.5	0.62	6.57	0.64	6.68	0.62	6.6	18.2
RF	0.83	4.47	0.83	4.53	0.83	4.48	3.99	0.84	4.29	0.85	4.32	0.84	4.30	3.9
XGBoost	0.77	5.1	0.79	5.02	0.78	4.97	6.98	0.78	5.02	0.80	4.94	0.80	4.81	11.6

Machine learning algorithm	Test 3[c]							Test 4[d]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.33	8.92	0.33	9.19	0.33	8.87	13.6	0.32	8.95	0.32	9.25	0.32	8.93	7.97
RF	0.55	7.15	0.57	7.24	0.56	7.1	3.83	0.25	9.23	0.29	9.4	0.31	9.01	3.36
XGBoost	0.46	7.89	0.49	7.89	0.48	7.69	6.9	0.38	8.41	0.38	8.62	0.39	8.32	1.79

[a]Test 1: all attributes.
[b]Test 2: relative elevation, aspect, OM, slope, BI, curvature and Ec shallow.
[c]Test 3: relative elevation, aspect, OM, slope, BI, and curvature.
[d]Test 4: relative elevation, slope and BI.

Yield Temporal Stability (YTS)

Figure 3 shows the Boruta algorithm feature selection results for both the complete and farmer-friendly datasets. Similar to the yield spatial trend, relative elevation was also the most important factor affecting yield temporal stability (Figure 3A). Macro and micronutrients such as sulfur, copper, iron and calcium completed the list of the five most important attributes. Soil properties such as TWI, loam, sand, and clay contents, and curvature were the least relevant features. In the smaller dataset, relative elevation, organic matter, and brightness index were the most important attributes, while TWI and curvature had the lowest importance to the variability in temporal stability of the yield data. A study conducted in 2018 using artificial neural network-based approach to determine relevance of different attributes in corn and soybean yield also found that slope consistently showed high importance among the 20 predictor variables tested (Kross et al. 2018).

Prediction model results are shown in table 4 and 5 for complete and farmer-friendly datasets, respectively. The SVM models showed the lowest R^2 values, highest errors and longest running times among the models in all tests of both datasets. Despite the overall good prediction capabilities of the XGBoost models, random forest showed the best performance with the highest accuracy among the models. In the complete dataset, the R^2 values for training, validation and test sets for all models were above 0.79. These results indicated that five most important attributes would be enough to predict YTS in this field.

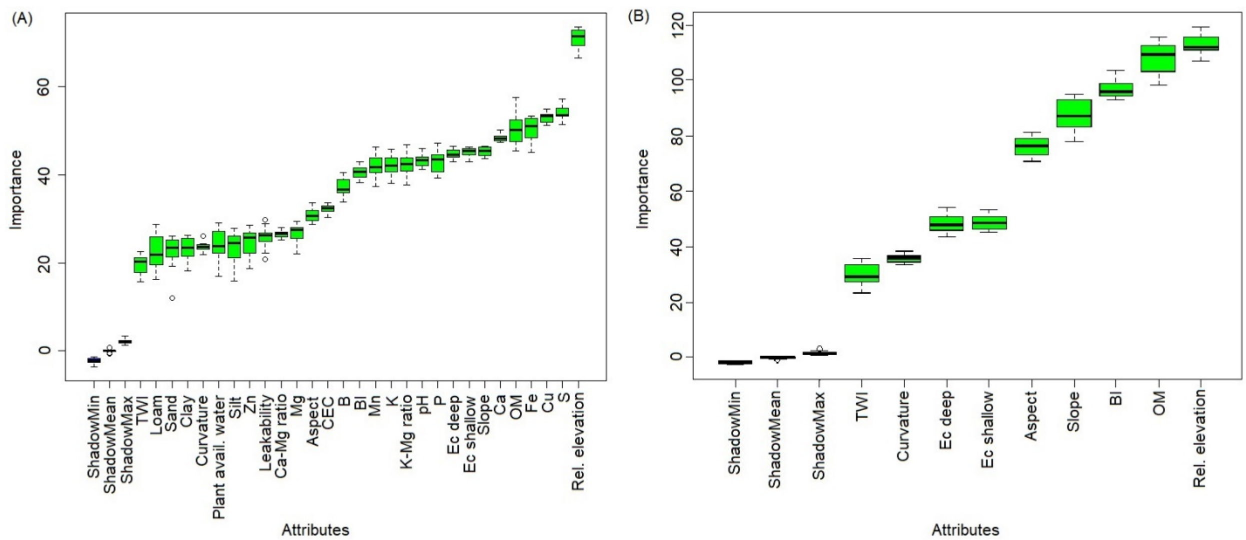

Figure 3. Boruta algorithm feature selection results for yield temporal stability (YTS) prediction using the complete dataset (A), and the farmer-friendly dataset (B).

Table 4. Machine learning YTS prediction models using support vector machine (SVM), random forest (RF) and eXtreme gradient boosting (XGBoost) models for the complete dataset. Different attribute combinations were tested to select the best performing model.

Machine learning algorithm	Test 1[a]							Test 2[b]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.29	4.47	0.31	4.59	0.31	4.55	110	0.23	4.68	0.26	4.8	0.24	4.78	20.4
RF	0.86	1.96	0.88	2.02	0.89	1.97	5.0	0.86	1.95	0.88	2.00	0.88	1.97	4.2
XGBoost	0.80	2.31	0.82	2.28	0.83	2.25	33.1	0.83	2.12	0.86	2.00	0.86	2.01	0.8
	Test 3[c]							Test 4[d]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.1	5.07	0.11	5.24	0.12	5.17	13.2	0.10	5.07	0.11	5.24	0.11	5.18	10.9
RF	0.83	2.15	0.84	2.24	0.85	2.19	3.9	0.79	2.37	0.80	2.51	0.79	2.49	3.9
XGBoost	0.65	3.13	0.66	3.17	0.66	3.16	3.9	0.69	2.91	0.72	2.89	0.70	2.91	9.1

[a]Test 1: all attributes.
[b]Test 2: top 10 most important attributes.
[c]Test 3: top 7 most important attributes.
[d]Test 4: top 5 most important attributes.

The accuracy of YTS prediction using a smaller dataset with less variables significantly decreased (Table 5). The best performing model was the random forest in test 2, using the top seven most important attributes. The R^2 values were 0.66, 0.67 and 0.68, while errors were 3.04, 3.13, and 3.07 for training, validation and test, respectively. In addition to the lower number of attributes, one possible cause of the decreased prediction accuracy is the absence of macro and micronutrients. Based on the feature selection results, the nutrient spatial variability in the soil has a great effect on the crop final yield.

Table 5. Machine learning YTS prediction models using support vector machine (SVM), random forest (RF) and eXtreme gradient boosting (XGBoost) models for the farmer-friendly dataset. Different attribute combinations were tested to select the best performing model.

Machine learning algorithm	Test 1[a]							Test 2[b]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.21	4.75	0.24	4.88	0.22	4.87	17.6	0.22	4.74	0.24	4.88	0.22	4.87	12.9
RF	0.65	3.08	0.67	3.17	0.68	3.11	4.15	0.66	3.04	0.67	3.13	0.68	3.07	4.09
XGBoost	0.57	3.44	0.60	3.40	0.60	3.38	0.51	0.56	3.46	0.59	3.44	0.58	3.47	9.49
	Test 3[c]							Test 4[d]						
	Training		Validation		Test			Training		Validation		Test		
	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)	R^2	RMSE	R^2	RMSE	R^2	RMSE	Time (min)
SVM	0.07	5.16	0.08	5.33	0.09	5.24	9.74	0.21	4.79	0.23	4.9	0.21	0.88	11.2
RF	0.43	3.91	0.45	4.04	0.47	3.92	3.88	0.69	2.88	0.71	2.95	0.72	2.91	3.97
XGBoost	0.34	4.24	0.34	4.38	0.37	4.26	9.46	0.56	3.49	0.6	3.41	0.61	3.35	6.69

[a]Test 1: all attributes.
[b]Test 2: relative elevation, aspect, OM, slope, BI, curvature and Ec shallow.
[c]Test 3: top 5 most important properties.
[d]Test 4: relative elevation, aspect, OM, slope, BI and Ec shallow.

Conclusion

The identification of key attributes that affect yield spatial and temporal variability in a field can greatly contribute to the delineation of representative management zones for site-specific application. Results showed that different soil and landscape attributes had varying roles in predicting crop yield, and that field data easily available could be used to predict crop yield and delineate management zones. Despite the promising results more research is required to test the relevance of different attributes across multiple fields and conditions. The next steps for this research will be to delineate management zones based on the prediction model results and analyze the yield variability within each zone. In the future, this pipeline for management zone delineation will be tested for nitrogen management, and potential economic and agronomic benefits will be analyzed.

Acknowledgements

This project was funded by Minnesota Department of Agriculture/Agricultural Fertilizer Research and Education Council (AFREC) (R2020-32, R2021-32), the Minnesota Robotics Institute Seed Grant (MNR-12), and the USDA National Institute of Food and Agriculture (State project 1016571).

References

Beck, H., Zimmermann, N., McVicar, T., Vergopolan, N., Berg, A., & Wood, E. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific Data, 5, 180214.

Blackmore S. (2000). The interpretation of trends from multiple yield map. Computers and Electronics in Agriculture, 26(1): 37–51.

Breunig, F.M., Galvão, L.S., Dalagnol, R., Santi, A.L., Flora, D.P.D., & Chen, S. (2020). Assessing the effect of spatial resolution on the delineation of management zones for smallholder farming in southern Brazil. Remote Sensing Applications: Society and Environment, 19, 100325.

Brock A., Brouder, S.M., Blumhoff, G., & Hofmann, B.S. (2005). Defining yield-based management zones for corn-soybean rotations. Agronomy Journal 97(4), 1115-1128.

Chang, D., Zhang, J., Zhu, L., Ge, S. H., Li, P. Y., & Liu, G. S. (2014). Delineation of management zones using an active canopy sensor for a tobacco field. Computers and Electronics in Agriculture, 109, 172–178.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. New York, NY, USA: Association for Computing Machinery.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Tien Bui, D., et al. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. Catena, 151, 147-160.

Doerge, T. (1999). Management zone concepts. In S. Clay, et al. (Eds.) SSMG#2. Site specific management guidelines. Norcross, GA, USA: Potash and Phosphate Institute. Retrieved April 1, 2022

Fridgen J.J., Kitchen, N.R., & Sudduth, K.A. (2004). Management zone analyst (MZA): software for subfield management zone delineation. Agronomy Journal 96:100–108.

Godsey, C.B., Schmidt, J.P., Schlegel, A.J., Taylor, R.K., Thompson, C.R., & Gehl, R.J. (2003). Correcting iron deficiency in corn with seed row–applied iron sulfate. Agronomy Journal, 95, 160– 166.

Hussain, N., Khan, A.Z., Akbar, H., Bangash, N.G., Khan, H., & Idrees, M. (2007) Response of maize varieties to phosphorus and potassium levels. Sarhad Journal of Agriculture, 23(4), 881-888.

Ivanciuc, O. (2007). Applications of Support Vector Machines in Chemistry. In Reviews in Computational Chemistry, 291–400. John Wiley & Sons, Ltd.

Jiang, Q, Fu, Q., & Wang, Z. (2011). Delineating site-specific irrigation management zones. Irrigation and Drainage, 60(4), 464-472.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrift, 15, 259–263.

Kross, A., Znoj, E., Callegari, D., Kaur, G., Sunohara, M., & Van Vliet, L. (2018). Evaluation of an artificial neural network approach for prediction of corn and soybean yield. In: Proceeding of the 14th International Conference on Precision Agriculture, June 24- 27, Montreal, Quebec, Canada.

Kursa, M.B., & Rudnicki, W.R. (2010). Feature selection with the Boruta Package. Journal of Statistical Software, 36 (11), 1-13.

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

Luck, D. J., Mueller, N., & Fulton, P. J. (2015). Improving yield map quality by reducing errors through yield data file post-processing (Extension Publication). University of Nebraska-Lincoln Division of the Institute of Agriculture and Natural Resources, Lincoln, NE. Retrieved from https://extensionpublications.unl.edu/assets/pdf/ec2005.pdf

Miao, Y., Mulla, D.J., & Robert, P.C. (2018). An integrated approach to site-specific management zone delineation. Frontiers of Agricultural Science and Engineering, 5(4), 432-441.

NOAA National Centers for Environmental information, National Weather Service. Available online: https://www.weather.gov/wrh/climate (Accessed on 19 April 2022).

Pachepsky, Y. A., Timlin, D. J., & Rawls, W. J. (2001). Soil water retention as related to topographic variables. Soil Science Society of America Journal, 65(6), 1787–1795.

Planet (2019). Planet Surface Reflectance version 2.0. Accessed 03 March 2022. https://assets.planet.com/marketing/PDF/Planet_Surface_Reflectance_Technical_White_Paper.pdf

Planet (2021). Planet Imagery Product Specification. Accessed 03 November 2021.

Sanches, G. M., Graziano Magalhães, P. S., & Junqueir Franco, H. C. (2019). Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. Geoderma, 334, 90–98.

Scudiero E., P. Teatini, D.L. Corwin, R. Deiana, A. Berti, F. Morari. (2013). Delineation of site-specific management units in a saline region at the Venice Lagoon margin, Italy, using soil reflectance and apparent electrical conductivity. Computers and Electronics in Agriculture 99: 54-64.

Shaner, D. L., Khosla, R., Brodahl, M. K., Buchleiter, G. W., & Farahani, H. J. (2008). How well does zone sampling based on soil electrical conductivity maps represent soil variability? Agronomy Journal, 100(5), 1472–1480.

Shukla, A.K., Behera, S.K., Lenka, N.K., Tiwari, P.K., Prakash, C., Malik, R.S., et al. (2016). Spatial variability of soil micronutrients in the intensively cultivated Trans-Gangetic Plains of India. Soil & Tillage Research 163: 282– 289.

USDA- Natural Resources Conservation Service (2022). Web Soil Survey. https://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx (Accessed 03 March 2022).

Wang, M., Tavakkoli, E., Rengasamy, P., Huang, L., et al. (2021). Management zone delineation based on soil properties measure during the reproductive stage of rice in the field. Land Degradation & Development, 32(10), 3106-3121.

Were, K., Bui, D. T., Dick, Ø. B., & Singh, B. R. (2015). A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecological Indicators, 52, 394–403.

Yao, R. J., Yang, J. S., Zhang, T. J., Gao, P., Wang, X. P., Hong, L. Z., et al. (2014). Determination of site-specific management zones using soil physico-chemical properties and crop yields in coastal reclaimed farmland. Geoderma, 232–234, 381– 393.