12-2021

Demonstrating Spatial Patterns of Crop Productivity in a Minnesota Corn Field Using Hierarchical Multiple Regression Models and Ordination

Ndumezulu Mpofu

Demonstrating Spatial Patterns of Crop Productivity in a Minnesota Corn Field Using Hierarchical Multiple Regression Models and Ordination

by

Ndumezulu T Mpofu

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Geographic Information Science

December, 2021

Thesis Committee:
Kirk Stueve, Chairperson.
Jeffrey Torguson
Michner Bender
Eric Fuller

Abstract

Accurate and timely assessment of within-field crop vigor heterogeneity is essential for detecting field-wide crop productivity and yield, contributing to improvements in the management of corn fields. Yet, studies designed to explore the spatial heterogeneity of crop vigor in corn over different productivity zones, where soil nutrient characteristics are known to limit crop productivity during the growing season, as yet been reported. We assessed whether changes in temporal weather conditions within a growing season, contribute to crop vigor variability. Furthermore, we evaluated whether within-season changes in precipitation and temperature contribute to variable nutrient concentrations within different productivity zones. More so, we utilized random forest regression to calculate the relative importance of predictor variables to crop vigor variability. We then employed hierarchical multiple regression (HMR) to build several regression models to determine whether the collinearity of variables (soil characteristics) showed a significant improvement in the R^2 i.e., the proportion of explained variance in crop vigor response. The principal component analysis (PCA) was employed to find components that express as much of the inherent variability of the complete data set as possible as well as, to plot how variables map relative to field productivity or management zones. We inferred that, changes in precipitation and temperature during the growing season influence soil nutrient concentrations within productivity zones especially, potassium, calcium, nitrogen, phosphorus, and magnesium. We hypothesize that, significant and yet subtle crop vigor differences can be observed within field productivity zones attributed to the heterogeneity of soil macro nutrient concentrations within corn fields, using the combined utility of remote sensing and hybrid statistical approaches. Thus, aiding farmers to ascertain, early season, whether they will obtain a poor harvest or not, improve on soil nutrient use efficiency, and field management practices to ensure a bumper harvest.

Table of Contents

List of Tables

List of Figures

Figure Page

Chapter 1: Introduction

Across the globe, corn is characterized by high production and high exports concentrations. However, the world is experiencing an unprecedented, huge change in this regard. How to ensure a sustainable and consistent supply of corn, and thereby ensuring food security for all nations has become particularly important (Wang & Hu., 2021). The production of corn in the Midwest, Minnesota in particular, has dominated the agricultural landscape for almost a century. What remains of interest to both producers and consumers alike, is enhancing the sustainability of this vast agricultural crop (Fischer et al., 2014). Soil is one of the most important environmental factors and is considered as the main source in providing essential plant nutrients, water reserves and a medium for plant growth. Agricultural soils with a good physical quality provide favorable conditions for plant growth and development. Resilient, productive soils are necessary to sustainably intensify agriculture to increase yields. Quantitative models have been used to explore how soil organic matter relates to crop yield potential of corn and wheat in light of co-varying factors of management, soil type, and climate (Oldfield et al., 2018). The effects of a changing climate and enhanced climate variability are already being seen across the Midwest. Over the past century, temperatures have risen across all seasons, growing seasons have become longer, precipitation patterns have changed, and extreme precipitation events have increased in frequency and severity (Hatfield et al., 2011). Because of the sensitivity of agriculture to weather and climate conditions, these impacts can have substantial direct and indirect effects on farm production and profitability. Temperature effects on plant growth have been extensively studied and several studies suggest that temperature stresses may be extremely significant in terms of affecting crop growth and yield (Tebaldi & Lebell, 2008).

Changes in climate entail changes in the variability and frequency of extreme weather. To minimize the impact of climate variability and change on crop yield, there is need to develop and evaluate adaptation strategies (Basso et al., 2018). The more frequent occurrence of extreme weather events has caused declines in yield in grain crop across the Midwest (Lesk et al., 2016). Climate change, particularly the warming temperature and varying precipitation, has negative effect on grain yield and soil organic carbon. Southworth et al. (2000) argued that the increase in temperature under climate change leads to faster crop development which reflects in yield declines and quicker soil carbon mineralization. Undoubtedly, climate change and the increase of extreme weather events will have an adverse impact on agricultural production in the US Midwest (Rotter et al., 2018).

The spatial heterogeneity of crop growth within fields is rarely quantified but it is essential for estimating yield and optimizing crop management (Stadler et al., 2014). Relationships between crop growth and soil physical characteristics have been described before using a high number of invasive measurements (Van Ittersum et al., 2013). However, non-invasive methods are available for characterizing soil heterogeneity, investigate growth characteristics and yield estimation (Blasch et al., 2020). Previous studies suggest that the solution to reducing existing yield gaps within farms lies in understanding factors limiting yield in areas with agricultural intensification potential. Diagnostics on farm nutrient omission trials have been conducted to assess soil nutrient related constraints to corn yield and to quantify their variability. Results have indicated that the large variability of soil nutrient related constraints need to be accounted for to optimize corn yield (Alfonso et al., 2021). Large variability in crop responses to macronutrients at various spatial scales present challenges for developing effective

fertilizer recommendations for corn production. Thus, warranting the need for zone specific fertilizer recommendations using decision support tools that consider variable soil fertility conditions and yield responses obtained from past studies (Ge et al., 2011).

Not all zones of a grower's field are equally productive; some zones are often more productive relative to the rest of the field, others always less. Of interest though, is that other areas fluctuate in their production capacity year after year, attributed to the interaction between climate, soil, topography and management (Maestrini & Basso, 2018). Crop vigor and crop vigor response is frequently heterogenous within-fields and related to multiple factors that are spatio-temporally sensitive. Although predicting crop productivity and their spatio-temporal variability under a changing climate is challenging, it is an essential undertaking for crop management and policy making (Ahmad et al., 2020). Soil variability plays a significant role in crop productivity and crop performance can be influenced by the soil-nutrient spatial variability. Consequently, crop productivity is spatially variable within heterogeneous productivity zones. Of paramount importance, both from an economic and an environmental point of view, is understanding why crop productivity in certain portions of the field has high variability over a growing season (Hatfield & Prueger, 2015). Accurate crop productivity maps of a growing season are essential for effective agricultural monitoring. Less attention has been paid to the dynamics of composition and spatial extent of crop productivity within a season. However, understanding the dynamic progress of the composition and spatial structure of mosaicking crops is critical for a diversity of agricultural monitoring activities (Song et al., 2017).

Since productivity is influenced by soil characteristics, the spatial pattern of productivity could be caused by a corresponding variation in certain soil properties Mzuku et al. (2005).

Micronutrients are essential plant nutrients taken up by crops in very small amounts, but a deficiency can have profound effects on crop vigor because they perform important physiological functions. Studying the spatial variability in soil nutrients and/or moisture and their impact on crop productivity or yield is essential (Zhou et al., 2020). Traditionally, fertilizers have been applied in fields without considering within-field soil nutrient spatial variation. Such agricultural practices are inefficient due to under application or over application within management zones (Fu et al., 2009).

Field measurements of chlorophyll content, crop height, and crop vigor are good indicators of crop status. These can be related to soil properties to identify within-field management measures (Duncan et al. 2015). Crop vigor (productivity) and senescence are physiological processes that characterize plant growth and development phases. Crop vigor is an indicator of robustness and rapidness in the growth of plant canopy architecture, primarily influenced by its genetic makeup interacting with the environmental factors. Early crop vigor provides the horsepower that drives plant development and yield while senescence marks the final stage of organ development in plants, characterized by a series of degenerative programmed processes leading to plant death (Makanza et al., 2018).

Remote Sensing and Crop Vigor Indices

Remote sensing has been recognized as a cost-effective way to detect the spatial and temporal variability of crop growth and productivity. Over the last two decades, there has been significant advancement in the application of geospatial technologies in agriculture (Maestrini & Bruno, 2018). Improved resolutions of remotely sensed images coupled with more precise on-the ground systems have allowed farmers to become more sensitive about the spatio-temporal

variations of crop yields occurring in their fields. Aerial sensing technologies offer radically new perspectives for assessing crop growth (ground canopy cover) and leaf senescence at low cost, faster, and in a more objective manner.

Recent studies have shown that UAV-based aerial sensing platforms have great potential for monitoring the dynamics of crop canopy characteristics like crop productivity through ground canopy cover and canopy senescence. Proximal and remote sensor surveys have been employed in precision agriculture, to delineate and monitor within field variations in soil and crop attributes. Thus, guiding variable rate control of inputs so that in-season management can be responsive, matching strategic fertilizer application to site-specific (management zones) field conditions (Hedley, 2015).

Statistical Methods and Crop Vigor Variability

Several statistical approaches have been developed and utilized so far, based on the analysis of the within-field variability in crop and soil properties (Corti et al., 2020), but procedures were rarely suited and robust for operational conditions. Random Forest has been used in previous studies to predict crop yield responses to climate and biophysical variables at global scale and regional scales in wheat, maize, and potato. Results showed that Random Forest is an effective and versatile machine learning method for crop yield predictions for its high accuracy and precision, ease of use, and utility in data analysis (Jeong et al., 2016). A different study developed a random forest regression-based approach to model interactions between planting rate, topography, and soil characteristics and their effects on yield. With increasingly available topographical and soil information, there is a growing interest in developing variable rate strategies to exploit variation in the agricultural landscape in order to maximize production.

A Random Forest-based approach was developed to model the interactions between planting rate, topography, and soil characteristics and their effects on yield.

To develop effective agricultural and food policies at the regional and global scale, accurate predictions of crop yield are critical (Jeong et al., 2016). Jig Han Jeong et al. (2016) evaluated Random Forests (RF), a machine learning method, for its ability to predict crop yield responses to climate and biophysical variables at global and regional scales, using crop yield data from various sources and regions for model training and testing. Statistical modelling methods based on machine learning algorithms can provide alternatives to traditional regression approaches and overcome some of their limitations (Lawler et al., 2006). These are ideal for ecological systems because they are flexible enough to handle complex problems with multiple interacting elements and hold great promise for the advancement of understanding and predicting ecological phenomena (Olden et al. 2008). Due to its dependence on multiple factors such as crop genotype, environmental factors, management practices, and their interactions, crop yield prediction is extremely challenging (Khaki et al., 2019). By employing statistical methods such as multivariate regression models, and machine learning, an efficient empirical method for classification and prediction; remotely sensed data has been widely used for crop yield estimation (Kim & Lee, 2016). While several prediction models have been used such as, multiple linear regression (MLR), support vector machines (SVM) and stochastic gradient boosting (SGB), studies that have carried out model performance statistics have revealed that, machine learning techniques perform marginally better than the MLR but, the random forest regression model provides in most cases the highest accuracy (Forkuor et al., 2017). Spectral vegetation indices extracted from unmanned aerial vehicle images and machine learning algorithms have

been proven effective in assisting nutritional analysis in plants. Spectral vegetation indices processed with the random forest algorithm have proposed as a framework to infer the nitrogen content at canopy level (Osco et al., 2019).

Recent research has shown that machine learning can provide reasonable predictions faster and with higher flexibility compared to simulation crop modelling. A committee of models can reduce prediction bias, variance, and can better capture the underlying distribution of the data. In more recent studies the principal component analysis, stepwise linear regression models and hierarchical clustering have been used to evaluate the multivariate relationship between weather, environmental factors, grain quality and yield (Osco et al., 2018).

Some studies have used Multi-Temporal Yield Pattern Analysis to identify potential erroneous yield maps, Principal Component Analysis (PCA) to detect yield patterns, evaluated temporal yield pattern stability using per-pixel analysis, and generated productivity stability units based on k-means clustering and zonal statistics. Detected yield patterns were associated with varying production conditions. Multivariate statistical analysis (PCA, HCA, PLS-DA, amongst others) contribute to reduce the dimensionality of data, to extract important information from the entire spectrum, improving the reliability of the analyses and facilitating the interpretation of the results (Kuhnen et al., 2010).

Other studies have employed hierarchical linear modelling to solve maize yield prediction problems over years and regions. Results showed that HLM provided higher accuracy and outperformed linear regression and multiple linear regression methods (Zhu et al., 2021). One study applied an integrated analysis approach comprising Classified and Regression Tree (CART), Generalized Linear Mixed Model (GLMM), and Factor Analysis (FA), to explain soil

and management-related factors influencing corn yield gaps between two fields. Classification and Regression Tree models were used to explain variables and interactions influencing crop yields in Eastern India. Factor analysis was used to cluster variables into common and easily interpretable factors. FA revealed that soil was the main factor influencing maize yield gaps at both sites, rather than management related variables.

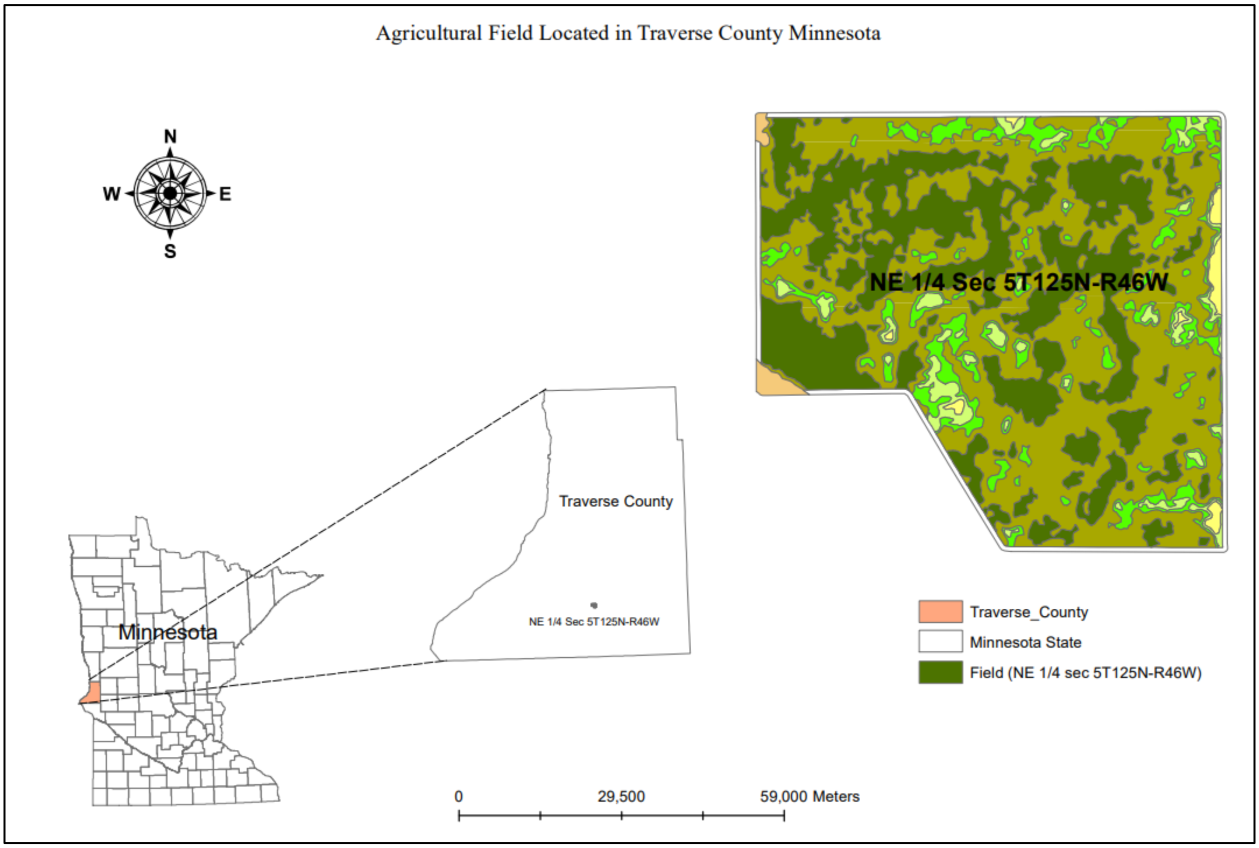
Pertinent to this study, the GLMM identified agroecology-specific factors influencing corn yield gaps as soil-available phosphorus and zinc. Evidently showing that there is a growing need to advance understanding of soil and management-related factors directly influencing within-season crop vigor response and yield. Whetton et al. (2021) carried out a study that introduced a new non-linear correlation analysis method based on a non-linear finite impulse response (NFIR) model to study and quantify the effects of ten soil properties on crop yield. Results of this study demonstrated that the individual and total contribution of the soil properties on crop yield varied throughout the different cropping seasons, weather conditions, and crops. (Whetton et al., 2021)

Previous research studies have extensively looked at spatial variability of crop yields at field scale. This study sort to explore the spatial heterogeneity of crop vigor in corn over different productivity zones, where soil nutrient characteristics are known to influence crop productivity during the growing season. The first objective was to map the spatio-temporal response of corn vigor between different image dates using the normalized difference red edge index (NDRE) as a proxy for crop productivity. The second objective was to establish key determinant soil characteristics influencing corn vigor response between different productivity zones. The other objective was to determine whether there is a correlation between corn vigor

response and variable soil characteristics. Last but not least, the study sort to establish whether there are significant differences in crop vigor response to variable soil characteristics between different productivity zones and different dates of the growing season.

Figure 1

The Study Site (NE ¼ Sec 5 T125N-R46W). Located in Traverse County, Minnesota



The study was conducted in a parcel of land located in Minnesota, 117.4 tillable acres. Most of the field used to be on the bottom of Lake Agassiz, but one of the old beachlines spans the southwestern parts of the corn field. The Midwest is home to some of the most productive agricultural soil in North America. It is home to the Corn Belt, the largest corn producing area in the US. The Midwest has a wide variety of soils and each type of soil has a background story to

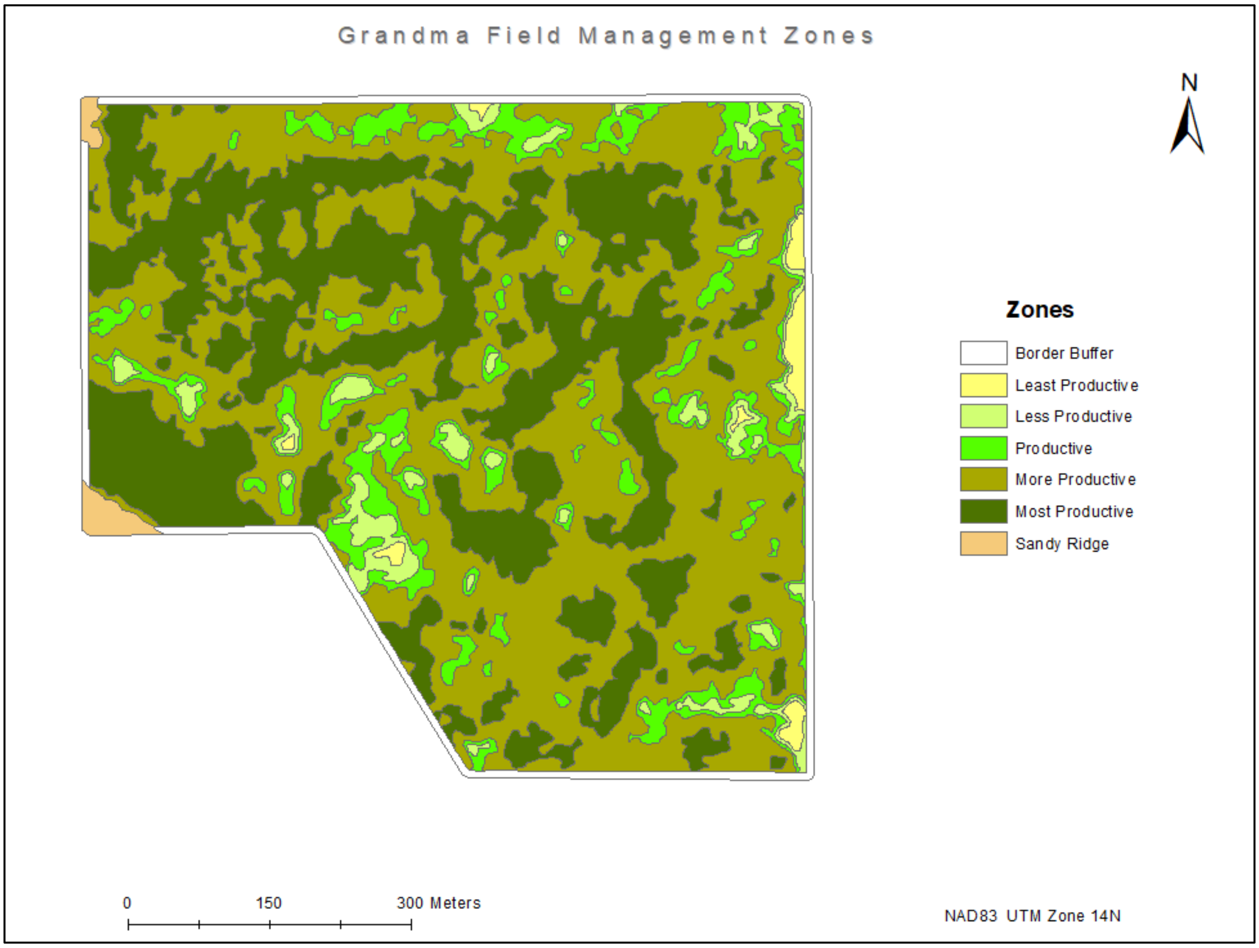
its origin. The region often obtains 40 inches per year and has a strong longitudinal gradient in precipitation. A stronger latitudinal gradient explains the continental climate. Minnesota has a continental climate, with cold, often frigid winters and warm summers. The growing season is 160 days or more in the south-central and southeastern regions, but 100 days or fewer in the northern counties. Normal daily mean temperatures range from 7°F (–14°C) in January to 66°F (19°C) in July for Duluth, and from 12°F (–11°C) in January to 74°F (23°C) in July for Minneapolis-St. Paul, often called the Twin Cities. The lowest temperature recorded in Minnesota was –60°F (–51°C), at Tower on 2 February 1996; the highest, 114°F (46°C), at Moorhead on 6 July 1936. Annual precipitation (1971-2000) averaged 31 in (79 cm) at Duluth and 29.4 in (75 cm) at Minneapolis-St. Paul. Precipitation is lightest in the northwest, where it averaged 19 in (48 cm) per year. Heavy snowfalls occur from November to April, averaging about 70 in (178 cm) annually in the northeast and 30 in (76 cm) in the southeast. Minnesota designated Lester soil as the official state soil in 2012. These soils are very productive and of significant importance to the economy in Minnesota. Lester soils are in 17 different counties in south-central Minnesota and total over 600,000 acres. These soils formed under woody vegetation that has been removed in most areas for agricultural production. Lester soils are well drained and formed in loamy, calcareous glacial till on ground moraines. They have a dark grayish brown loam surface with an eluvial horizon and an argillic horizon with clay loam and loam subsoils. Lester soils have properties developed from both grassland and forest environments and are primarily used for forage, corn, and soybean production.

Chapter 2: Methods and Materials

A range of approaches consisting of Digital Image processing, Geographic Information Techniques (GIS) and several statistical techniques, were used to demonstrate how corn vigor varies across different productivity zones of Grandma field. More so, demonstrate how variable soil macronutrient concentrations coupled with changes in in-season climatic conditions (i.e., temperature and precipitation) determine the subtle differences in crop vigor between sequential dates of a growing season. Also, to improve crop vigor variability mapping, a traditional remote sensing index (Normalized Difference Red Edge Index) was combined with statistical models, yielding a robust hybrid approach capable of flashing out these agronomically important differences that are occurring earlier in the season.

Figure 2

Map Showing Productivity Zones in Grandma Field (NE ¼ Sec 5 T125N-R46W)



High resolution drone imagery approximately 1m by 1m, geometrically and radiometrically corrected, were employed in this study. This was done with an intent to remove geometric distortions on the imagery so that individual pixels can be in their proper planimetric (x,y) position. Radiometric correction was done to correct data for sensor irregularities and unwanted atmospheric noise and convert data to represent reflected/emitted radiation. By so doing we made sure that actual radiant flux and recorded radiant flux are resolved. We also obtained soil zonation data for this particular study site (field). The soil zonation maps were

created with an objective to capture variability within the field in a typical good growing season using in-season Rapid Eye imagery post tessellation, but after crop canopy closure.

Different vegetation indices were calculated to delineate soil zones and minimum mapping units were applied for the initial iteration. Zones were then split based on known differences and soil substrates. The productivity or management zones are divided into five categories: least productive, less productive, productive, more productive, and most productive.

Normalized Difference Red Edge Index

Multiple platforms, from proximal to unmanned aerial systems based remote sensing, have been developed and applied to increase the throughput, efficiency, and objectivity during field phenotyping (Xie et al., 2018). Several studies have demonstrated the applications of satellite and drone imagery in agricultural production and crop phenotyping and suggest crop productivity as a plant trait that can be evaluated using high resolution satellite or UAV imagery (Hassan et al., 2018; Zhang et al., 2019). Hence, in-season high resolution imagery was used to calculate the crop vigor values (NDRE). The Normalized difference Red Edge Index (NDRE) is a spectral index that is built as a blend of several bands, i.e., Near InfraRed spectrum and a band that uses a narrow spectral range between visible Red and NIR. The index is very similar to the more common and powerful analog with a greater history, NDVI.

However, NDRE is more sensitive than NDVI for a certain period of crop maturation. NDRE is a better indicator of plant conditions than NDVI for middle and late season crops that have already accumulated a large amount of chlorophyll. It is more relevant for intensive use during the entire cultivation season, as NDVI often becomes inaccurate after plants accumulate a maximum amount of chlorophyll content (Index Saturation).

Zonal Statistics and Spatial Joins

Zonal statistics calculate a statistic for each zone defined by a zone dataset, based on values from another dataset (a value raster) (Lehmkuhl et al., 2021). A single value is computed for every zone in the input zone dataset. Pertinent to the data matrix for further statistical analysis was the mean or average of all cells in the value raster, i.e., NDRE corresponding to a specific productivity zone. Creating a zonal statistics table was vital. The next step was to use the zonal statistics table tool to calculate a subset of statistics on crop vigor values (NDRE) within the five management zones. The tool generated a table showing several statistical values (mean, median, standard deviation etc.).

The spatial join tool was then used to append attributes of the join features, i.e., the soil sample points and the productivity zones. Attributes of the soil sample feature dataset were copied over to the productivity zone output feature class.

Data Matrix

The most important matrix for any statistical procedure is the data matrix. In this study, a data matrix was created in Excel (comma delimited .csv file) with average values of both crop vigor NDRE values and mean values of variable soil nutrient content (held constant across the growing season) from the soil test results. The data matrix has two categorical and independent observations, i.e., management zones and day of measurement. It also has 18 predictor variables (different soil nutrient characteristics) accounting for one response variable, i.e., crop productivity values (NDRE values). The observations form the rows of the data matrix and the variables form the columns.

Random Forest Regression

Random forest is an ensemble of decision trees. Like other machine learning techniques, random forests use training data to learn to make predictions (Ishwaran & Lu, 2019). One advantage of Random Forest is its ability to compute predictor variable importance (Snider et al., 2021). Variables with high importance are drivers of the outcome and their values have a significant impact on the outcome values.

Random forest was used to calculate the relative importance of predictor variables to crop vigor variability (feature selection). The objective was to ascertain which variables are pertinent to explaining the spatial heterogeneity in the response of crop vigor (NDRE values) during the growing season. The assumption is that there is value in describing variable crop productivity spatially and relative to soil profile characteristics within each management zone. Statistical analysis was done using R studio version 1.4.1 and data are presented as two plots, i.e., the Mean Decrease Accuracy (%Inc MSE) and the Mean Decrease Gini (Inc Node Purity).

Hierarchical Multiple Regression

Hierarchical regression is a way to show if variables of interest explain a statistically significant amount of variance in the dependent variable after accounting for all other variables (Zahedi et al., 2020). It is often used as a framework for model comparison rather than a statistical method (Lai et al. 2021, Mayer et al., 2020). However, in this study it was used to build several regression models by adding variables (soil characteristics) to previous models at each step. Later models included smaller models in previous steps. Our interest was to determine whether the newly added variables (soil characteristics) showed a significant improvement in the

R^2, i.e., the proportion of explained variance in crop vigor (dependent variable) productivity or response.

Principal Component Analysis

The principal aim of the principal component analysis (PCA) is dimension reduction (Kurita, 2019). It is based on the correlation or covariance matrix and is often the first step for further multivariate data analysis procedures, i.e., Cluster Analysis, Multiple Regression and Discriminant Analysis (Granato et al., 2018). When dealing with data sets that consist of several variables, it is difficult to graphically inspect the main data structure of a multivariate data set. Thus, it is required to find components that express as much of the inherent variability of the complete data set as possible. The data matrix comma delimited (.csv) file was imported into R Studio, version 1.4.1, and scaled to get the features with maximum variance (the variance is high for high magnitude features) skewing the PCA towards the high magnitude features.

Scaling was done using the scale() function which divides the columns of data by their standard deviation. In x argument data [, -1:2] was used to exclude the first two categorical variables, i.e., management zones and day of measurement. The second step was extracting the principal components using the eigen() function from the base package and prcomp() from Stats package. The eigen() function was applied to the covariance matrix to calculate the eigenvalues and eigenvectors. An eigenvector is a matrix of values that represents direction, and an eigenvalue is a value specifying the amount of variation in that direction. In most cases the first two components account for the most variation in the data set. Results are viewed as components from the result of eigen decomposition.

The third step was computing the proportion of variance. For this, we attached values component of eigen decomposition from scaled data to get variances. The total variation was consistent with the total number of components and each component had unit variance. So, each variance was divided by the total variation to get the proportion of each variance. The cumsum() function was applied to proportion of variance, obtained in the third step, to get the cumulative proportions. Results are presented as percentage variation explained by the first PC and the percentage cumulative proportion explained by the first three PCs.

The fourth step was displaying the variance explained by the PCs using the plot() and scree plot() functions. The plot() function displays the variances or eigenvalues accounted for each component. The scree plot was used to display the principal object created in preceding steps. Results are represented using a simple scree plot, a line scree plot and a variance bar plot.

In this study we were also interested in seeing the correlation of the original variables with the principal components (PCs). To do this, the first step was to get the mean deviations for each variable. The next step was combining the PCs with data variables using the cbind() function then, applying the cov() function, resulting in a correlation matrix of PCs and data variables in combination.

A biplot is a type of a plot that allows one to visualize how the samples relate one to another in a PCA (which samples are similar, and which are different) and will simultaneously reveal how each of variable contributes to each principal component (Solonechmyi et al., 2018; Yildrim, 2018; Zhang et al., 2021). The ggbiplot package was used to plot the PCA as a biplot which includes both the position of each sample in terms of PC1 and PC2 and also shows how the initial variables map relative to the sample positions.

Chapter 3: Results

NDRE indices (Figure 3) generated for the different image dates of the growing season show the crop vigor values from near canopy closure to senescence on an approximate bi-weekly sequence. The graduated color ramp represents variable crop vigor status across the field. The red areas correspond with low crop vigor values and represent areas of low to poor productivity. The bright yellow areas correspond with medium crop vigor values showing moderately high productivity. The light to dark green areas represents the most productive sections of the field. The NDRE values reflect a somewhat apparent change or variability in crop vigor across the growing season, however the values are not sufficient to capture subtle crop vigor variations between different management zone which is the key objective of this study. To capture the variability of crop vigor between different productivity zones, zonal statistics where employed (Figure 4).

Figure 3

Normalized Difference Red Edge Indices computed from In-season Imagery for the Year 2018 Growing Season, in Grandma Field. (Red indicates areas with low crop vigor values; bright yellow represents moderate crop vigor values and light to dark green areas represent high crop vigor values.)

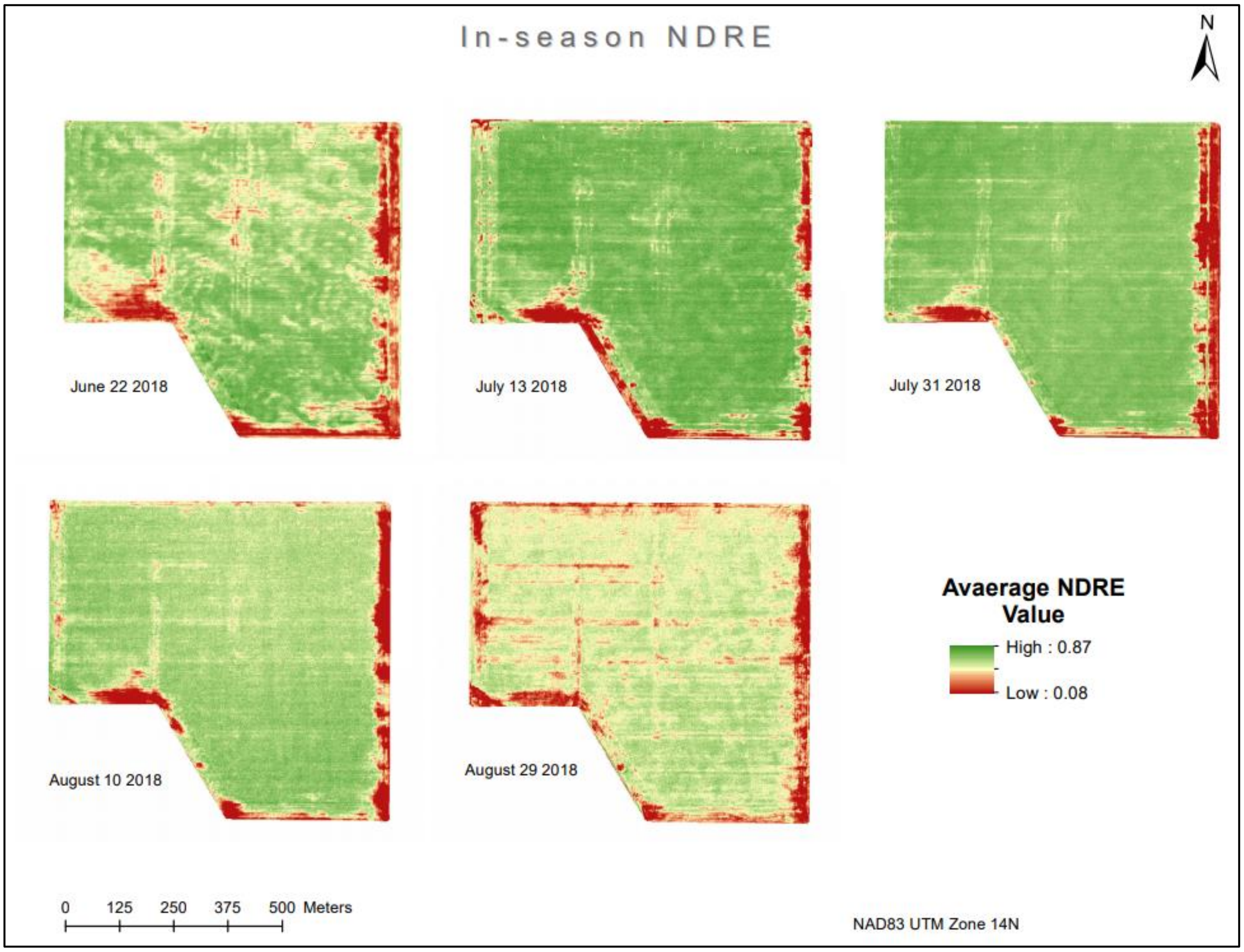


Figure 4 illustrates the variability of crop vigor by zone. Zonal statistics generated crop vigor variability maps using the mean (statistic type), which calculated the average crop vigor value corresponding to the different management zones. The red areas indicate low mean crop vigor values, bright yellow to orange areas indicate moderately high mean crop vigor values and bright green areas represent zones with high mean crop vigor values. Mean values show a

significant change in crop vigor between zones across the season however, they do not demonstrate a statistically significant difference in crop productivity between zones and across the season. More so, indices do not capture the influence of explanatory factors or variables underlying crop vigor variability between zones. To capture and demonstrate the influence of soil characteristics to variable crop vigor response between zones, random forest regression, multiple hierarchical partitioning, and ordination (Principal Component Analysis) were selected.

Figure 4

Average (mean) NDRE Values Summarized by Zone, Using Zonal Statistics Tool in ArcMap 10.6.1.

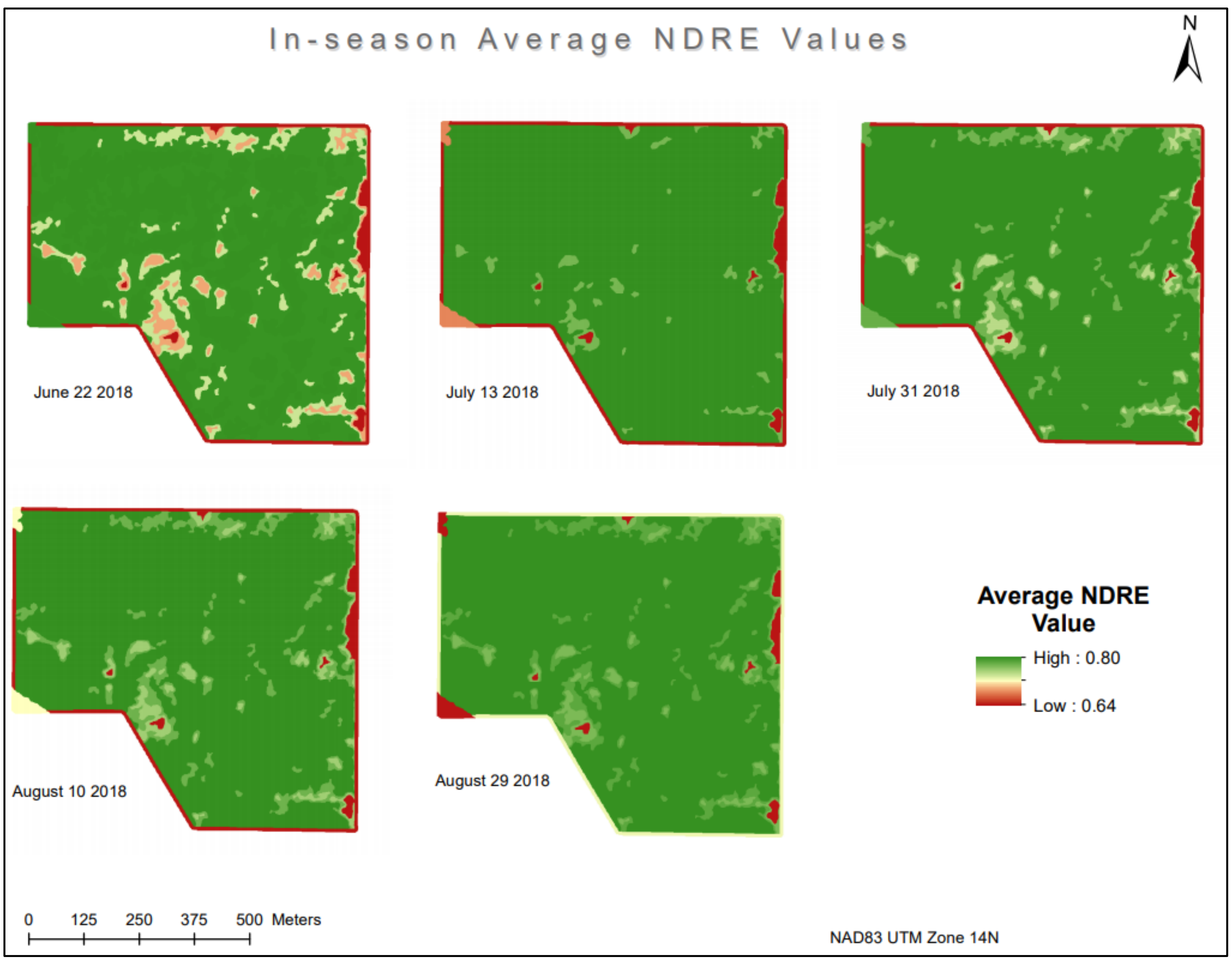


Figure 5

Graphical Representation of Average NDRE Values for the Month of June 2018

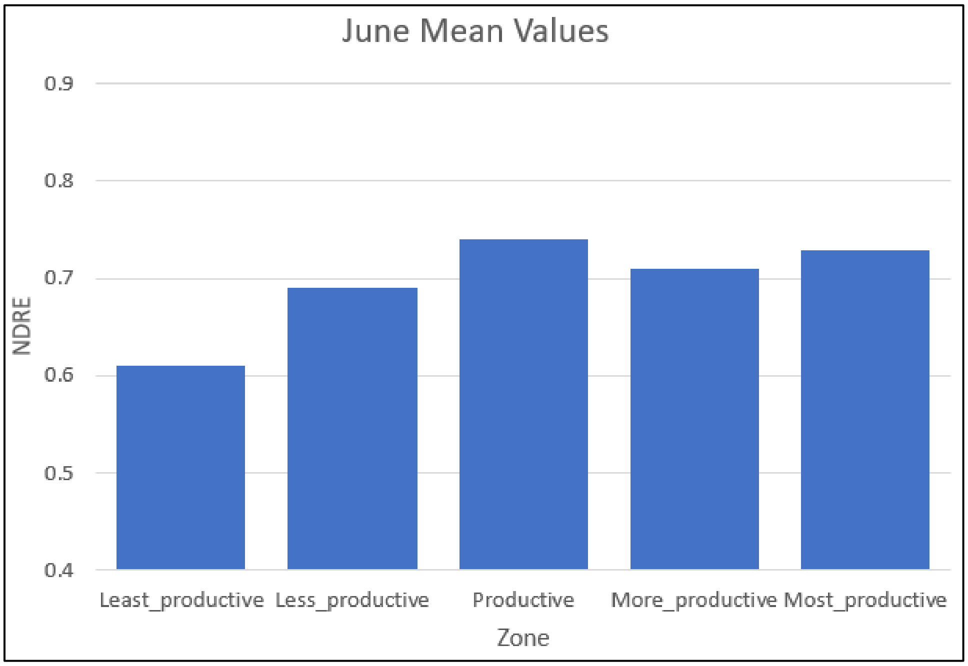
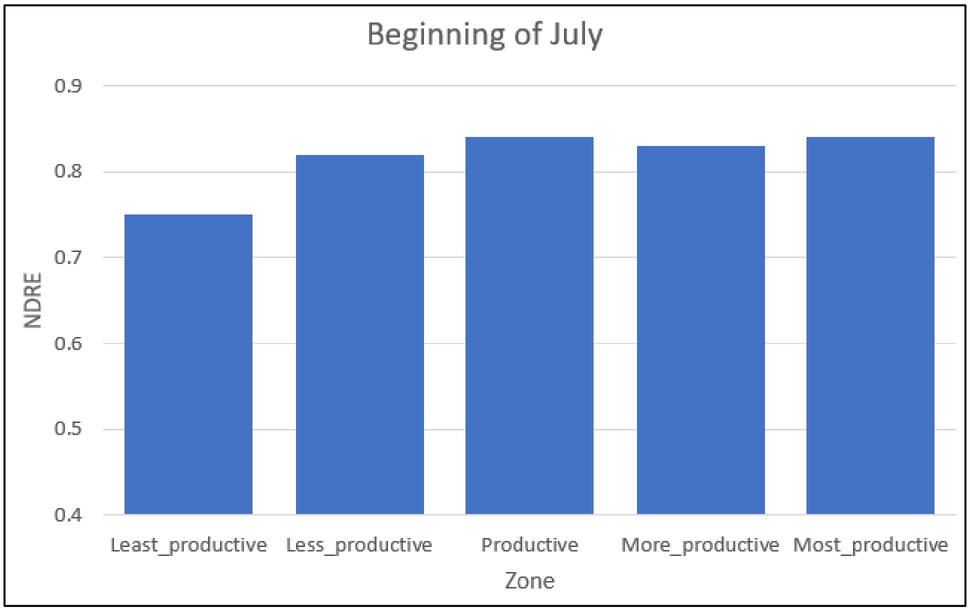


Figure 5 shows results of crop vigor (NDRE) values at the beginning of the growing season. The results indicate significant differences in crop vigor between the least productive zones of the field. As expected, the least productive zone of the field had the lowest crop vigor value (0.62), significantly lower than the less productive zone (0.68). However, the productive zones of the field did not appear to vary much in crop vigor. Of agronomic interest to note, is how the productive zone appears to have higher productivity than the more and most productive zones of the field. More so, the relatively low crop vigor values observed within the more productive zone, prompted further investigation in the study.

Figure 6

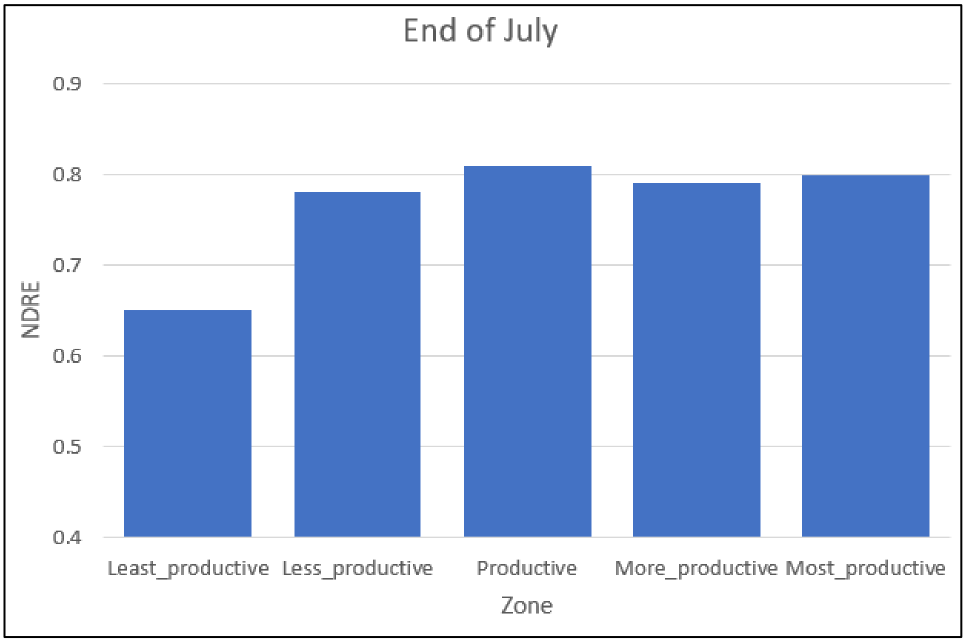
Graphical Representation of Average NDRE Values at the Beginning of July 2018



Results indicate a significant change in crop vigor values at the beginning of July. Crop vigor significantly changed (increased) especially in the less productive zones of the field. The least productive zone had a change of .15 while the less productive zone had a .12 change. Agronomically, these changes imply a significant difference in crop productivity. There were significant changes in crop productivity within the often-productive zones of the field that is, the productive (.9), more productive (.8) and most productive (.11) zone.

Figure 7

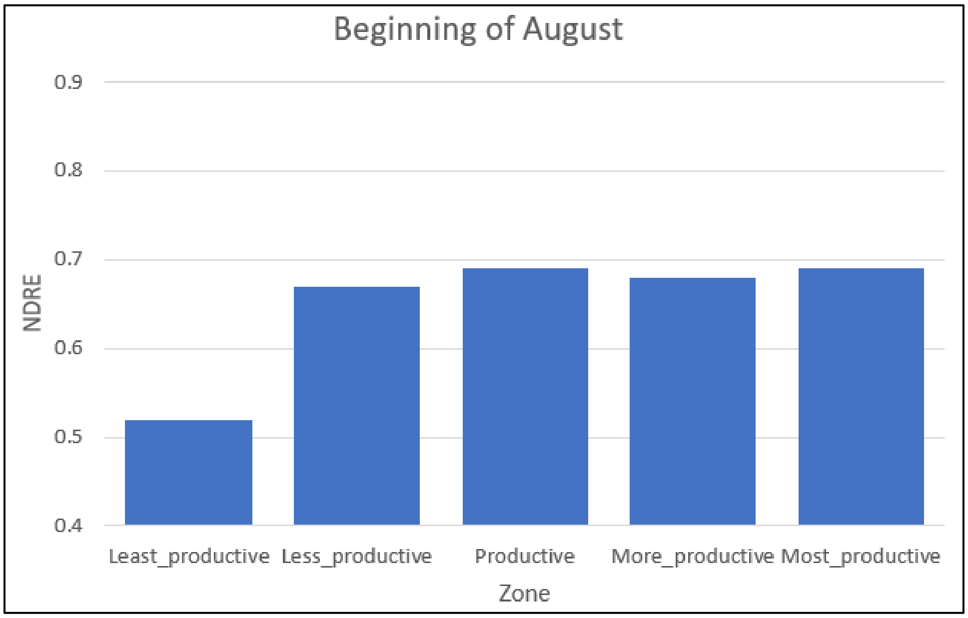
Graphical Representation of Average NDRE at the End of July 2018



Results show a significant change (drop) in corn vigor in all productivity zones of the field. The least productive zones of the field suffered a more significant decline in corn vigor. However, the observed decline in crop vigor, within the productive zones of the field, is notably a less significant difference compared to the least productive zones.

Figure 8

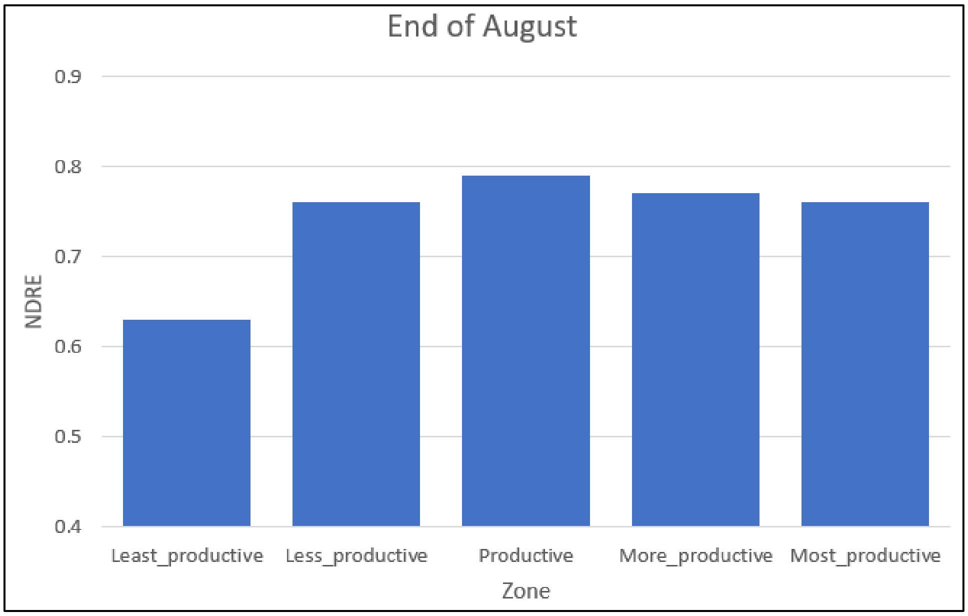
Graphical Representation of Average NDRE Values at the Beginning of August 2018



Results show a further decline in crop vigor at the beginning of August compared to the most recent observations at the end of July. The trend so far shows that, the most significant difference is observed within the least productive zones of the field, i.e., a .13 decline in the least productive zone and a .10 decline in the less productive zone. However, at the beginning of August, even the consistently productive zones of the field suffered a significant decline in crop vigor. With a .12, .10 and .11 decline in crop vigor within the productive, more productive, and most productive zones, respectively.

Figure 9

Graphical Presentation of Average NDRE Values at the End of August 2018

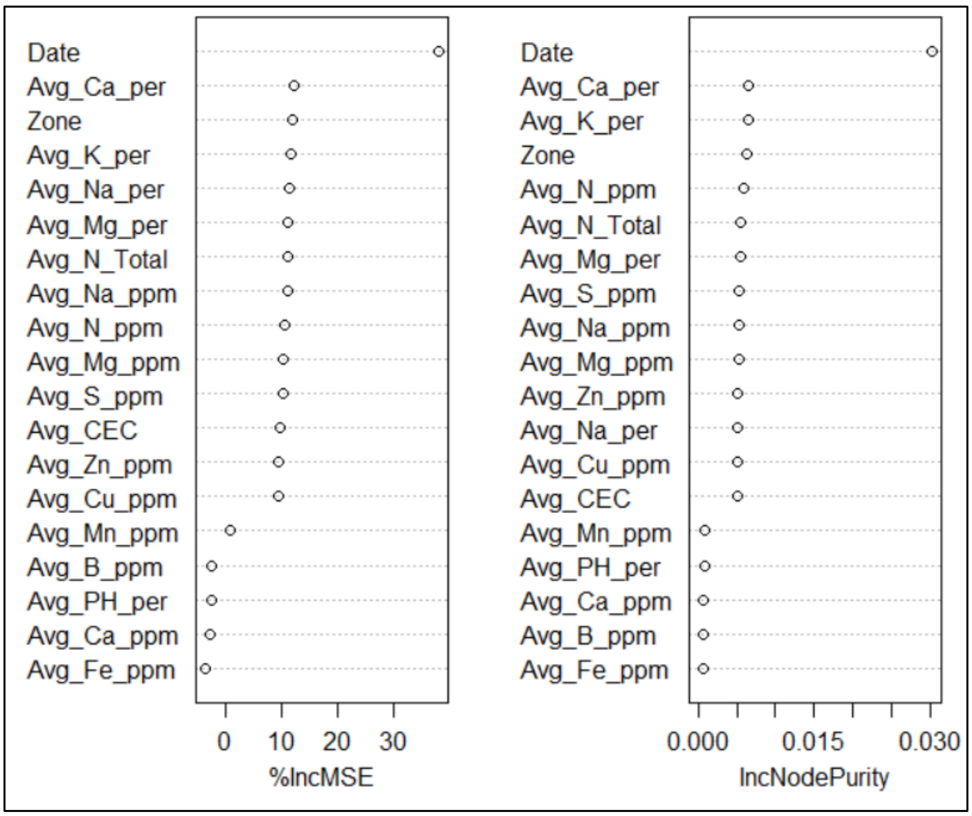


The end of August shows a significant increase in crop vigor compared to observations at the beginning of the month, within all productivity zones.

Random Forest Regression

Figure 10

Random Forest Variable Importance Plot Showing Which (of the 19) Variables or Soil Characteristics Better Predict Crop Vigor in Increasing Order of Magnitude. (Feature selection used to avoid over fitting in the regression models and to find out the relative importance of variables.)



Random Forest regression results indicate that the date within the growing season plays a fundamental role in determining corn vigor during the growing season. More so, management zones and known macronutrients essential for crop growth and development. i.e., calcium, potassium, magnesium, and nitrogen, were selected as the most influential predictors of crop vigor within the field.

Hierarchical Multiple Regression

Table 1

Multiple Linear Regression Model for Crop Vigor Response to Different Acquisition Dates
(Explains 53% of the crop vigor response.)

```
Call:
lm(formula = NDRE_Mean ~ Date, data = data)

Residuals:
   Min     1Q Median     3Q    Max 
-0.130  0.004  0.020  0.034  0.048 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.65000    0.02673  24.319 2.51e-16 ***
DateEA       0.09200    0.03780   2.434 0.024437 *  
DateEJ       0.11600    0.03780   3.069 0.006058 ** 
DateJ        0.04600    0.03780   1.217 0.237790    
DateJB       0.16600    0.03780   4.392 0.000282 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.05977 on 20 degrees of freedom
Multiple R-squared:  0.5331,	Adjusted R-squared:  0.4397 
F-statistic: 5.708 on 4 and 20 DF,  p-value: 0.003119
```

Table 2

Multiple Linear Regression Model for Crop Vigor Response to Potassium and Magnesium
(Explains 42% of crop vigor response).

```
Call:
lm(formula = NDRE_Mean ~ Avg_K_per + Avg_Mg_per, data = data)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.122154 -0.041875  0.001019  0.038125  0.107846 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)   
(Intercept) -0.182219   0.490970  -0.371  0.71408   
Avg_K_per    0.525997   0.183756   2.862  0.00905 **
Avg_Mg_per   0.005404   0.007668   0.705  0.48832   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06372 on 22 degrees of freedom
Multiple R-squared:  0.4162,	Adjusted R-squared:  0.3631 
F-statistic: 7.841 on 2 and 22 DF,  p-value: 0.002687
```

Table 3

Multiple Linear Regression Model for Crop Vigor Response to Average Potassium, Magnesium, Calcium, and Sodium (Explains 44% of crop vigor response).

```
Call:
lm(formula = NDRE_Mean ~ Avg_K_per + Avg_Mg_per + Avg_Ca_per
    Avg_Na_per, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-0.112 -0.046  0.014  0.036  0.118

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.45049   16.91933   0.381    0.707
Avg_K_per   -0.07072    0.69504  -0.102    0.920
Avg_Mg_per  -0.04329    0.15885  -0.273    0.788
Avg_Ca_per  -0.06120    0.16702  -0.366    0.718
Avg_Na_per  -0.11619    0.20553  -0.565    0.578

Residual standard error: 0.06541 on 20 degrees of freedom
Multiple R-squared:  0.4408,    Adjusted R-squared:  0.3289
F-statistic: 3.941 on 4 and 20 DF,  p-value: 0.01617
```

Table 4

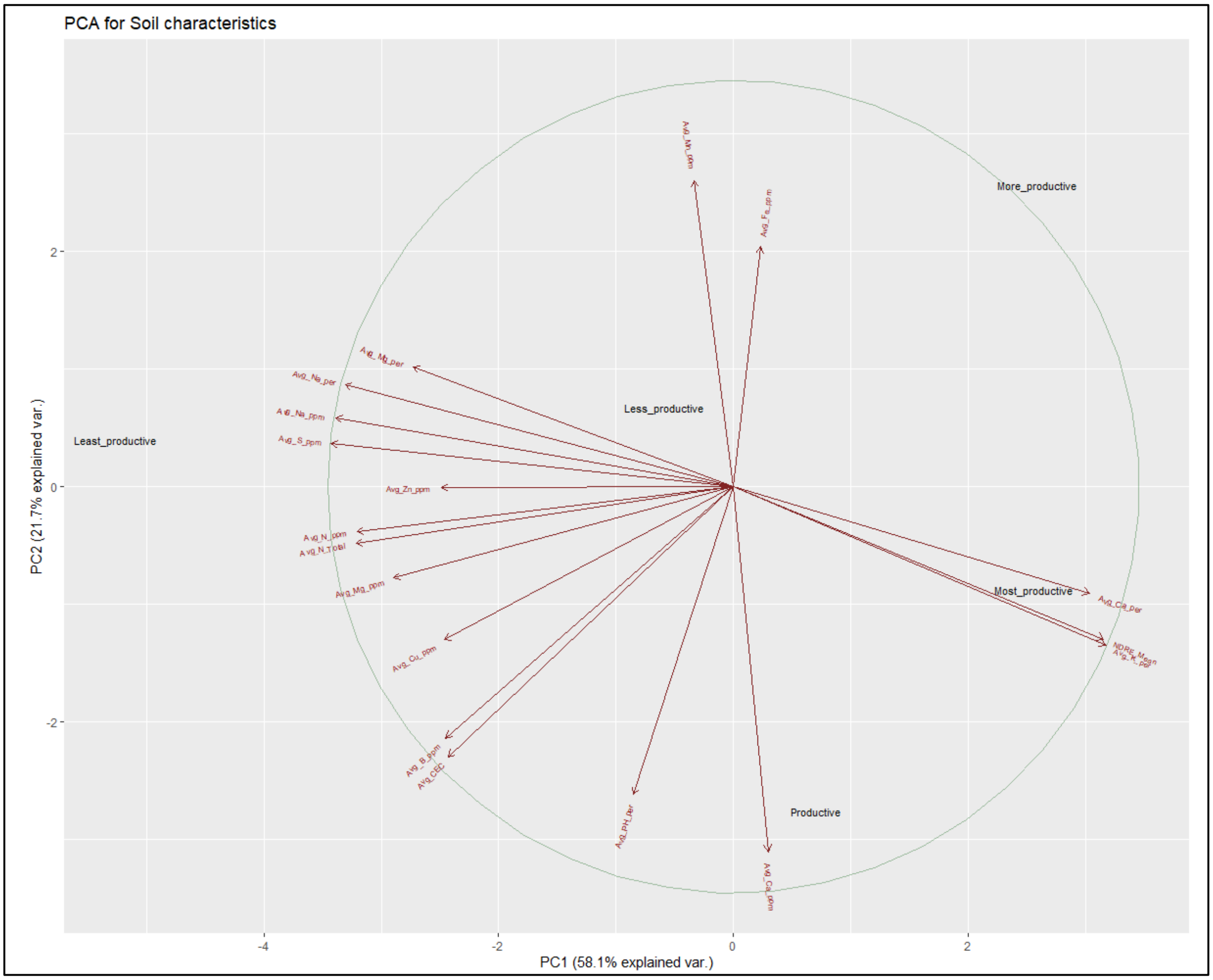
Multiple Linear Regression Model for Crop Vigor Response to Average Potassium, Magnesium, Calcium, Sodium, Sulfur, and Nitrogen. (Explains 44% of crop vigor response.)

```
Call:
lm(formula = NDRE_Mean ~ Avg_K_per + Avg_Mg_per + Avg_Ca_per +
    Avg_Na_per + Avg_S_ppm + Avg_N_ppm + Avg_N_Total, data = data)

Residuals:
   Min     1Q Median     3Q    Max
-0.112 -0.046  0.014  0.036  0.118

Coefficients: (3 not defined because of singularities)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.45049   16.91933   0.381    0.707
Avg_K_per   -0.07072    0.69504  -0.102    0.920
Avg_Mg_per  -0.04329    0.15885  -0.273    0.788
Avg_Ca_per  -0.06120    0.16702  -0.366    0.718
Avg_Na_per  -0.11619    0.20553  -0.565    0.578
Avg_S_ppm         NA         NA      NA       NA
Avg_N_ppm         NA         NA      NA       NA
Avg_N_Total       NA         NA      NA       NA

Residual standard error: 0.06541 on 20 degrees of freedom
Multiple R-squared:  0.4408,    Adjusted R-squared:  0.3289
F-statistic: 3.941 on 4 and 20 DF,  p-value: 0.01617
```

Principal Component Analysis

Figure 11

Principal Component ggbiplot with Objects (productivity zones) Displayed as Points and Variables Displayed as Vectors or Arrows



The arrows contain the information on loadings or represent the variable vectors. The length of the variable vector indicates how well the variables are represented by the graph (with a perfect fit if all vectors have the same length). The length of the arrows in the plot is directly proportional to the variability included in the first two principal components. The angle between

any two arrows represents the correlation between variables. If an angle between two variable vectors is zero, then the variables are collinear. If the angle is orthogonal both variables show a lack of correlation (the smaller the acute angle the stronger the positive correlation between two variables). Variable vectors with a greater obtuse angle are negatively correlated.

Figure 12

Variance Bar Plot Displaying Variances and Eigenvalues Accounting for Variances Explained by Principal Components (PC's)

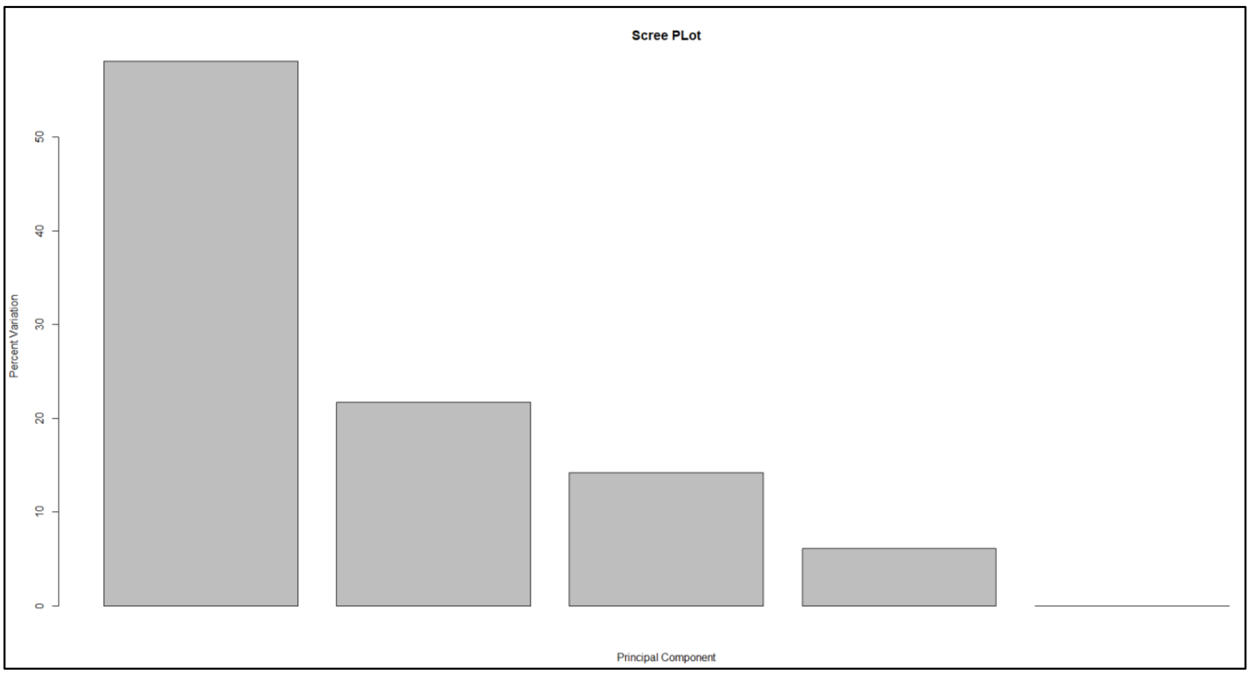
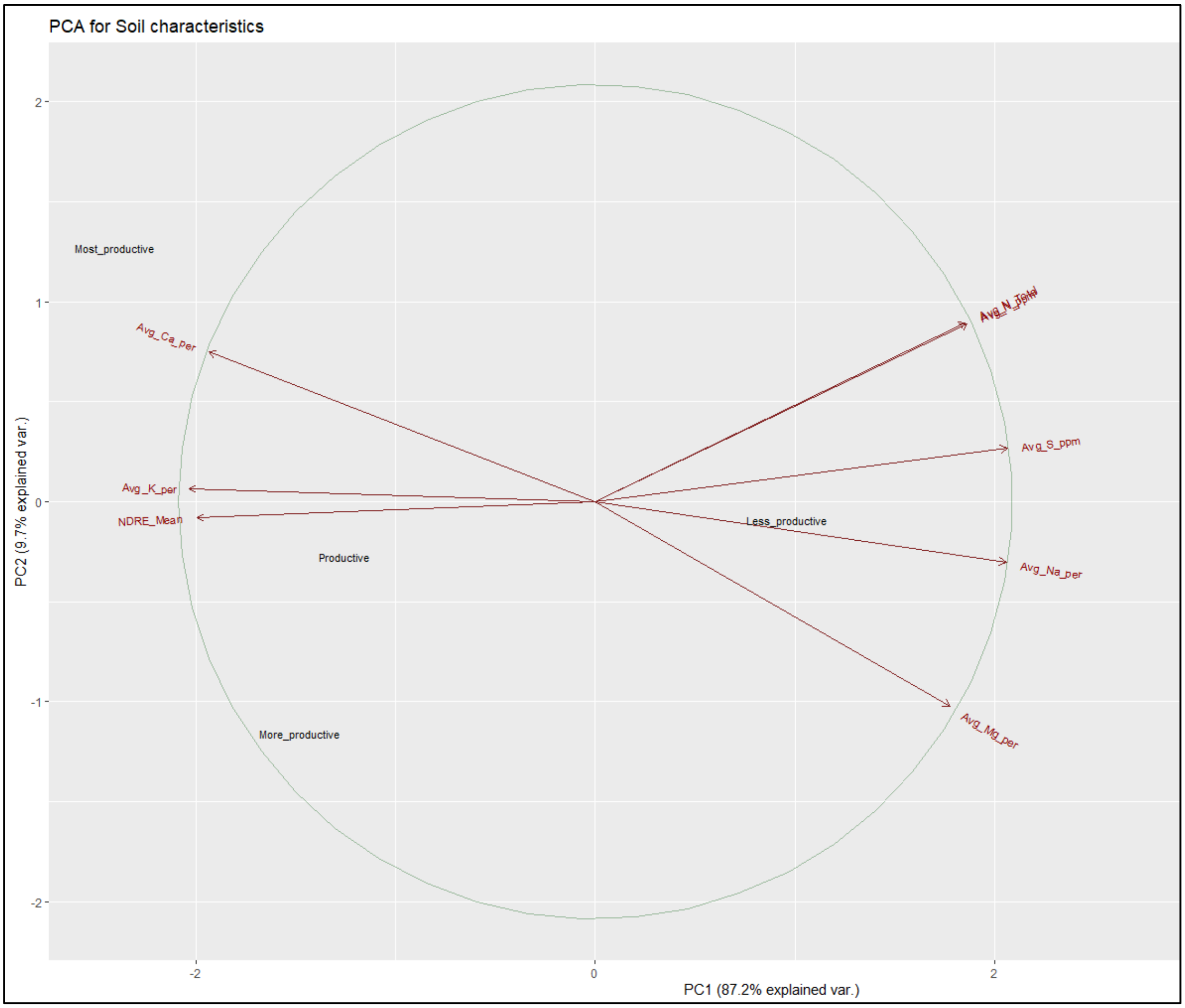


Figure 13

ggbiplot for the Month of June (Beginning of the growing season.)

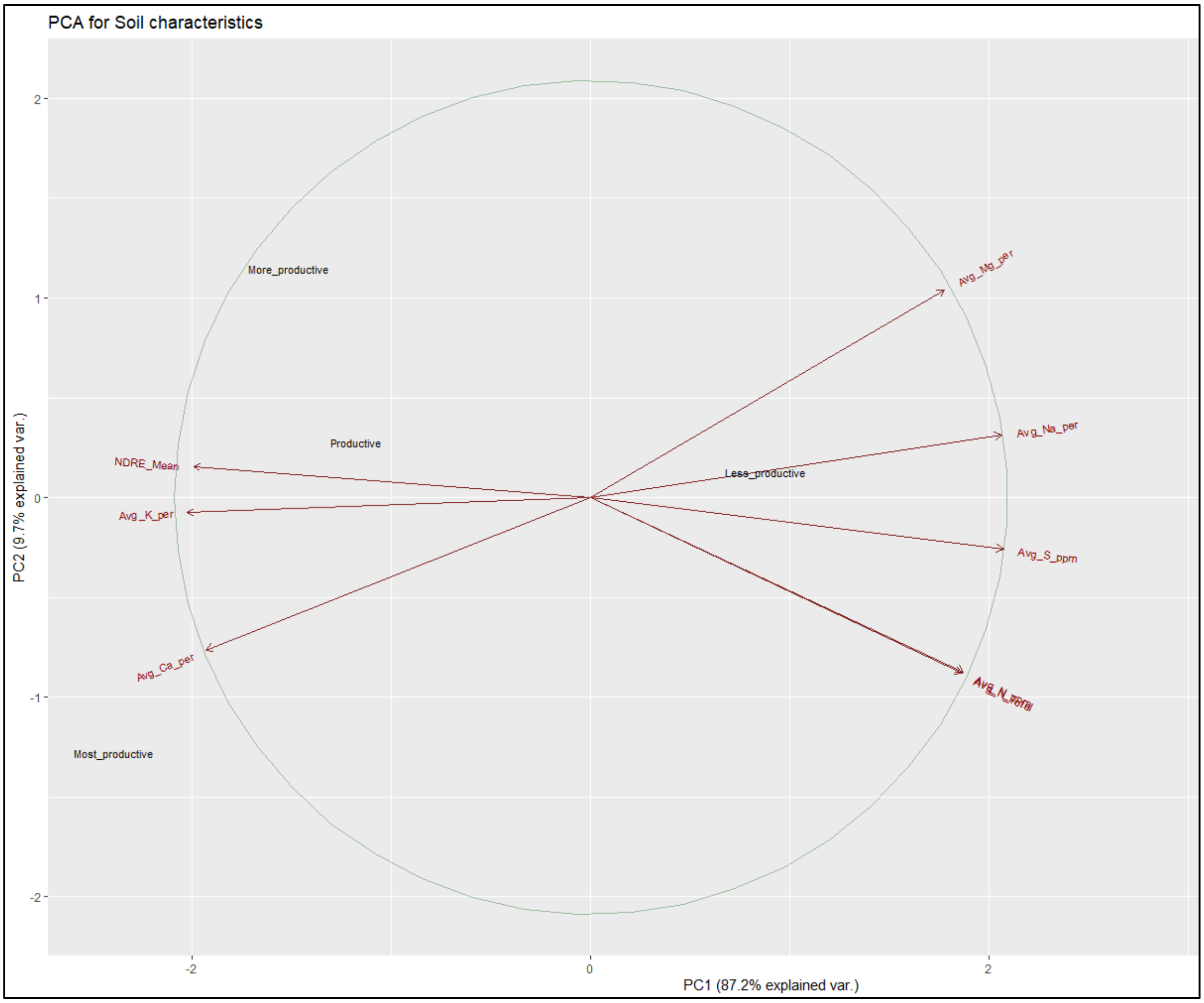


The observed vectors show that, the productive and most productive zones of Grandma field are characterized by high values of calcium and potassium. However, calcium is a limiting factor within the most productive zone. As observed there is a wider angular distance between NDRE and calcium, indicating a less positive correlation between the two variables. The acute angle between the NDRE mean and potassium is smaller. This indicates a strong positive

correlation between crop vigor response (NDRE mean value) and potassium. Model 2 (Table 2) explains 42% of the variation in crop vigor response with a p-value of 0.002 (*95% confidence interval, $p < 0.05$*). Potassium has a 0.009 (*95% confidence interval, $p < 0.05$*) level of significance in influencing crop vigor response within the productive zone (Table 2). The less productive zone is characterized by high nitrogen, sulfur, sodium, and magnesium. The obtuse angles between these soil characteristics (*explanatory variables*) and crop vigor values (*response variable*) indicate a strong negative correlation, with crop vigor response values (NDRE). The less productive and least productive zones are rich in nitrogen and sulfur while, magnesium and sodium are limiting factors.

Figure 14

ggbiplot for the Beginning of July (Beginning of the growing season.)

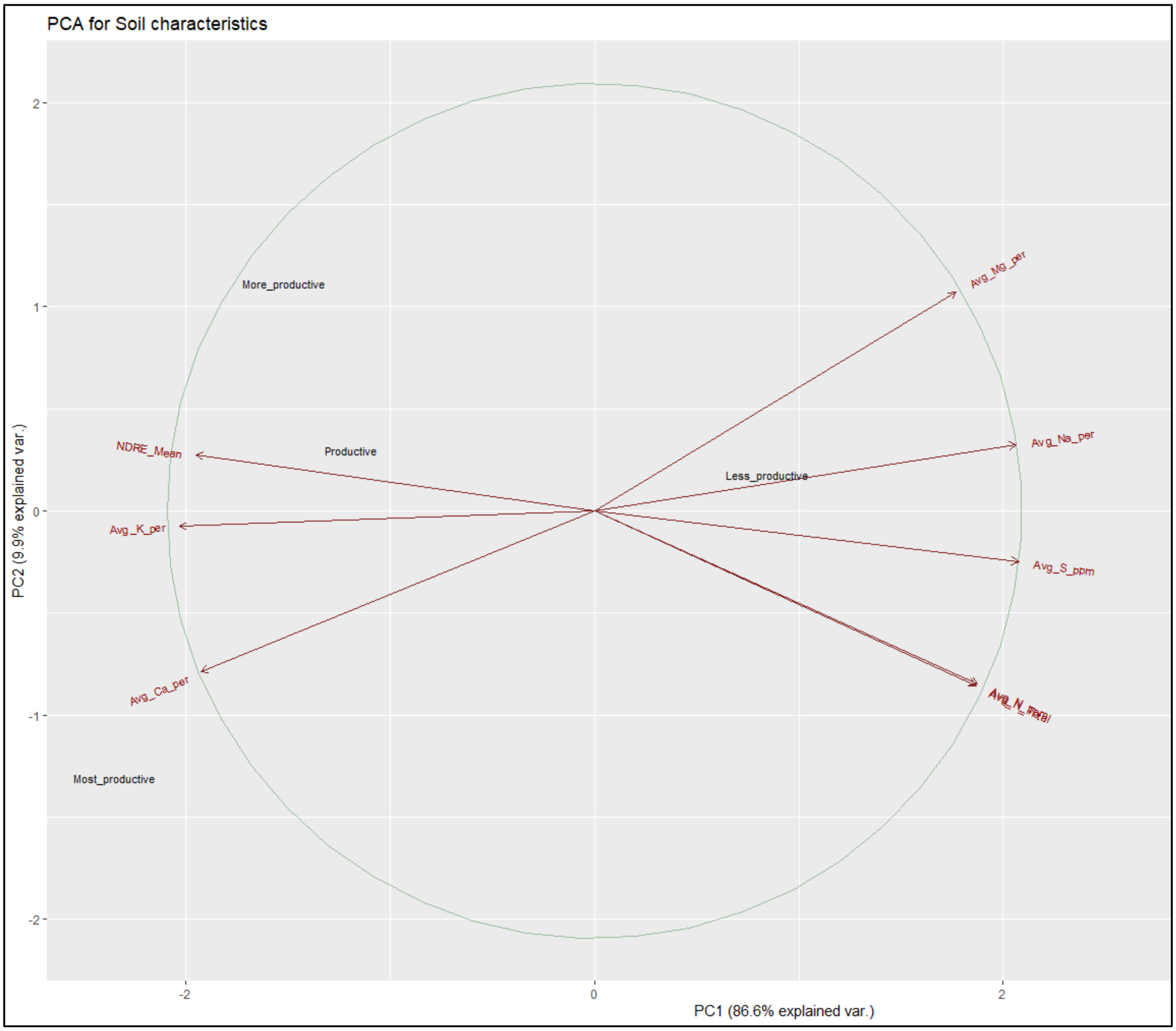


The observed vectors show that, the productive and most productive zones of the field are characterized by high concentrations of potassium. However, calcium is a limiting factor within the most productive zone. The acute angle between the NDRE mean and potassium is smaller. This indicates a strong positive correlation between crop vigor response (NDRE mean value) and potassium. The less productive zone is characterized by high nitrogen, sulfur, sodium, and

magnesium. The obtuse angles between these soil characteristics (*explanatory variables*) and crop vigor values (*response variable*) indicate a strong negative correlation, with crop vigor response values (NDRE). There is an observed difference in the limiting soil nutrients in the less productive and least productive zones, for the beginning of July as compared to the month of June. Nitrogen and sulfur are the limiting nutrients whilst sodium and magnesium appear to have been abundant and the most influential factors within these two zones.

Figure 15

ggbiplot for the End of July (Mid-season.)

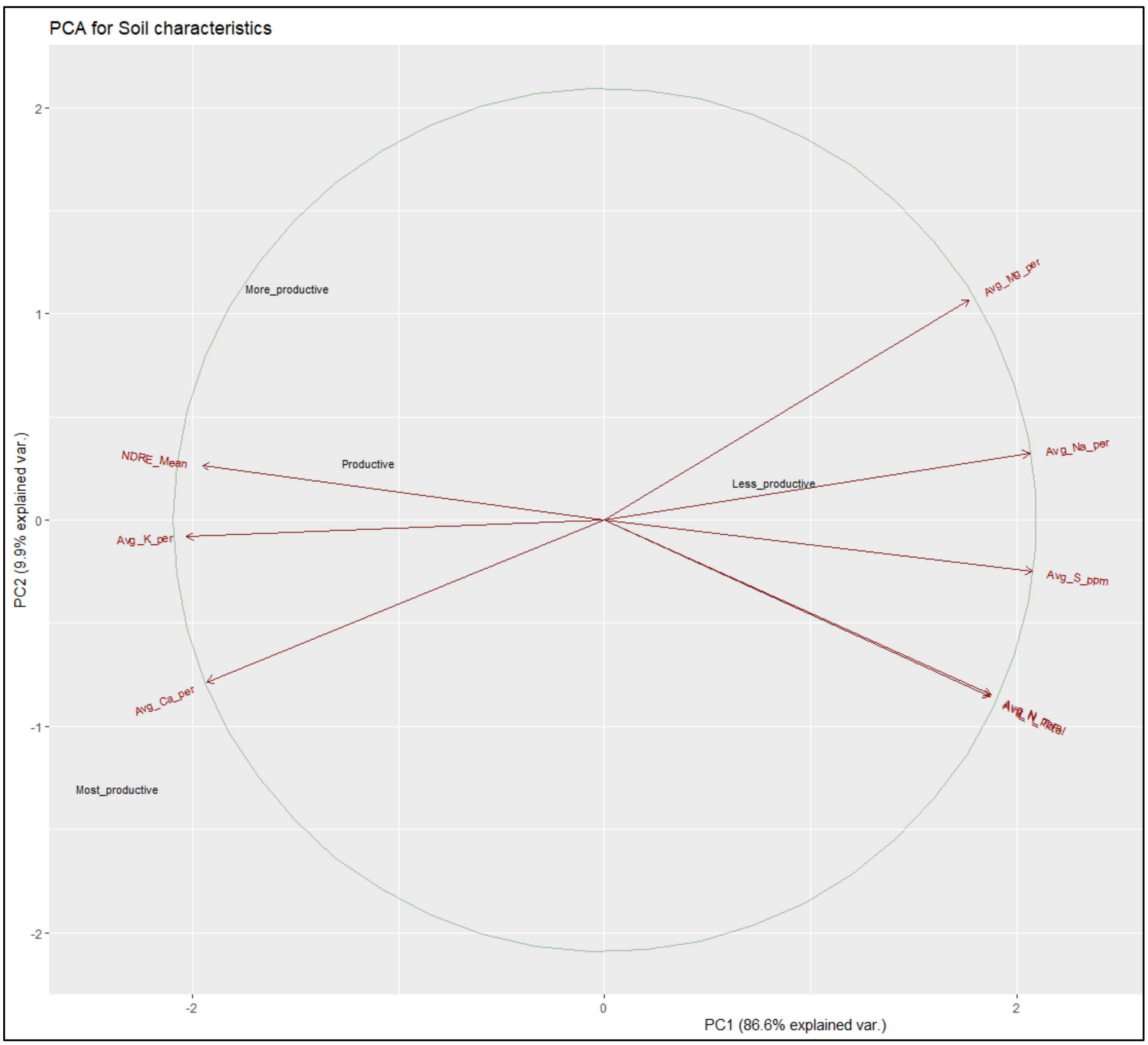


The observed vectors show that, the productive and most productive zones of Grandma field are characterized by high values of potassium. However, calcium is a limiting factor within the most productive zone. More so, there is a slight increase observed in the acute angle between the crop vigor values and potassium. This indicates a strong positive correlation between crop vigor response (NDRE mean value) and potassium, however there was a significant difference

between the beginning and the end of July. The less productive zone is still characterized by high nitrogen, sulfur, sodium, and magnesium, with no notable significant difference. The obtuse angles between these soil characteristics (*explanatory variables*) and crop vigor values (*response variable*) indicate a strong negative correlation, with crop vigor response values (NDRE).

Figure 16

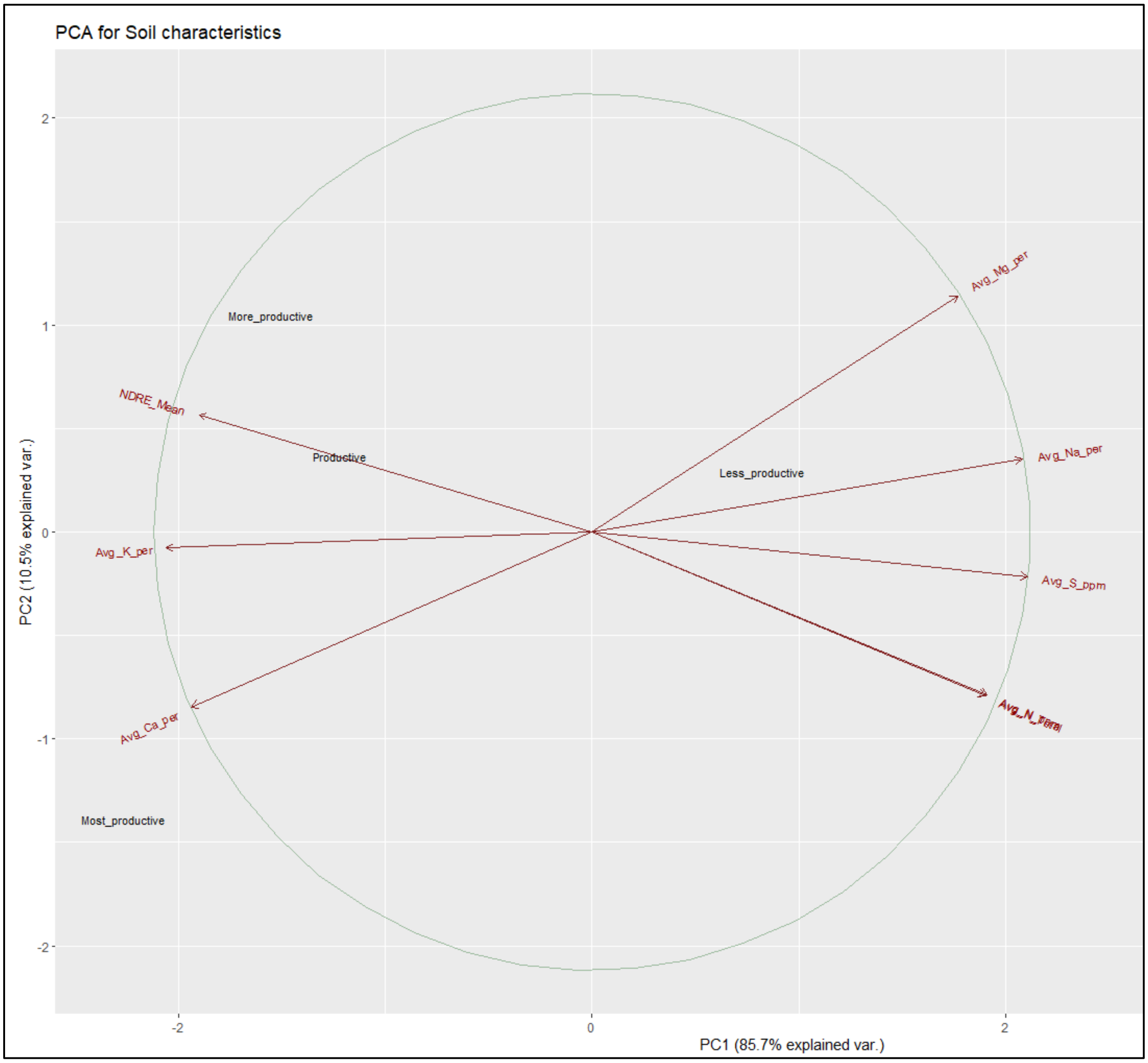
ggbiplot for the Beginning of August



The observed vectors show that, the productive and most productive zones of Grandma field are characterized by high values of potassium. However, calcium is a limiting factor within the most productive zone. More so, there is a slight increase observed in the acute angle between the crop vigor values and potassium. This indicates a strong positive correlation between crop vigor response (NDRE mean value) and potassium, however there was a significant difference between the end of July and the beginning of August. The less productive zone is still characterized by high nitrogen, sulfur, sodium, and magnesium, with no notable significant difference. The obtuse angles between these soil characteristics (*explanatory variables*) and crop vigor values (*response variable*) indicate a strong negative correlation, with crop vigor response values (NDRE).

Figure 17

ggbiplot for the End of August (End of season crop senescence.)



The observed vectors show that, the productive and most productive zones of Grandma field are characterized by high values of potassium. However, calcium is a limiting factor within the most productive zone. More so, there is a significant increase observed in the acute angle between the crop vigor values and potassium. Although there is a strong positive correlation

between crop vigor response (NDRE mean value) and potassium, however there was a significant decrease in the collinearity between crop vigor and potassium observed at the end of August. The less productive zone is still characterized by high nitrogen, sulfur, sodium, and magnesium, with no notable significant difference. The obtuse angles between these soil characteristics (*explanatory variables*) and crop vigor values (*response variable*) indicate a strong negative correlation, with crop vigor response values (NDRE).

Chapter 4: Discussion and Conclusion

From onset, expected results of the study will exhibit striking differences in crop vigor productivity between the top three productivity zones, i.e., productive, more productive, and most productive zones, and the bottom two zones, i.e., the least and less productive zones of the field. As expected, there were no statistically significant differences observed early season between the top three productivity zones. However, for a more rigorous agronomic observation, results from individual flights revealed a statistically significant difference between crop vigor between the least productive zones of the field and the more productive zones, early season. A finding consistent with studies reporting that, not all areas of a farmer's field are equal (Maestrini & Basso et al., 2018). Some always produce more relative to the rest of the field, others always less, while still other areas fluctuate in their production capacity from one year to the next, depending on the interaction between climate, soil, topography, and management. Understanding why the yield in certain portions of a field, has a high variability over time, is of paramount importance both from an economic and an environmental point of view. Through the better management of these areas, we can improve yields or reduce input costs and environmental impact (BAsso at al., 2018). As the season progressed, and mostly by end of season, results showed no statistically significant differences in crop vigor across the field attributed to the field crop being more homogenous, an outcome that is often expected. To understand the interaction between plants and environmental conditions, it is important to gather information on the spatial and temporal variability of crop growth status (Shang et al. 2015). The implementation of precision agriculture requires sub-field level information to critically detect and combat within-field yield limiting factors (Sozzi et al., 2018). The spectral resolution of satellite sensors, allow

to detect and monitor several characteristics of crops and soil, during the whole production process (Shang et al., 2015).

Of agronomic importance, observations from the results show that early season some zones of the field i.e., the productive and most productive zones evidently are lagging behind. Given that this was considerably a good year, considering precipitation and temperature, we infer that the lag was a result of different concentrations of soil nutrient characteristics. Knowledge of spatial variability of soil fertility and plant nutrition is critical for planning and implementing site specific crop field management. Site specific field management could therefore be considered for variable rate application of foliar fertilizers to increase yield in areas of low productivity or soil fertilizers to improve the quality of crop in some areas of the field (Jeong et al. 2016).

Different macro-nutrient concentrations influence the variability in crop vigor response between productivity zones. More so, the more productive zone evidently has low crop vigor values, on average, throughout the season attributed to leaching of salts decreasing crop vigor values. Our findings appear to corroborate the hypothesis that corn is sensitive to soil salinity. Soil salinity is a common cause of poor crop yield throughout North Dakota. More so, soil salinity is a global issue threatening land productivity, and estimates predict that 50% of all arable land will become impacted by salinity by 2050 (Butcher et al., 2016). Agronomically, a .15 change in average crop vigor value in the productive zones and .09 in the lower zones mid-season shows a considerably significant variation in crop vigor productivity, at the beginning to mid-season. Being able to characterize differences between productivity zones and understanding how ultimately yield is affected by management decisions implemented in the

field, and/or understanding the effects of variations in the field's biophysical characteristics, e.g., soil nutrients could improve crop growth and vigor (Falcon et al., 2021).

Our results reveal that different soil characteristics (nutrient content) within different productivity zones and temporal dates can facilitate crop vigor variability across the growing season. Different dates of the growing season explain 53% of the variability in crop vigor response with a 0.003 level of significance. Crop vigor response increased significantly with an increase in potassium, a key macronutrient that strengthens crops by contributing to early growth and assists crops to retain water. More so, it also keeps the plants from contracting diseases and insects. The increased crop vigor response was related to high concentrations of potassium in the productive and more productive zones of the field, which had consistently higher concentrations of macro-nutrients. We showed that crop vigor values vary across the field, with clusters of high crop vigor in the productive and more productive zones, and low crop vigor values in the less productive zones of the field.

However, results also confirmed that calcium is a limiting factor within the consistently most productive zone of the field. Calcium contributes to soil fertility by helping maintain a flocculated clay and therefore with good aeration. Because calcium has a stronger affinity for the exchange sites than sodium, added calcium a can improve soil structure by displacing sodium, which allows the negatively charged clay particles to aggregate. Plants require calcium to develop strong cell walls and membranes. Conversely, insufficient calcium in plants leads to a breakdown of cell walls and membranes, susceptibility to a variety of diseases and post-harvest problems particularly in fresh produce. There are soil conditions where calcium applications are very helpful. Sandy soils and crops irrigated with low calcium water may be particularly

vulnerable to low calcium availability. Soils with low pH generally have low calcium availability. Lower pH affects nutrient availability to plants. Many macro- and micro-nutrients are less available at lower pH, and calcium is commonly used to raise soil pH.

Our results also showed that the most productive parts of the field have high concentrations of potassium. Potassium is an important macro nutrient needed in the second highest amount by plants (behind nitrogen), underpinning crop yield production and quality determination (Williams, 2007). While involved in many physiological processes, potassium has important functions in plant water relations. Potassium's impact on water relations, photosynthesis, assimilate transport and enzyme activation can have direct consequences on crop productivity. Potassium is one of the major nutrients considered essential for crop growth and yield development, although not an integral component of any cellular organelle or structural part of the plant (Bruns & Ebelhar, 2006).

More so, nitrogen plays a pivotal role in the plant life cycle, a nutrient required for chlorophyll production and other plant cell components. The nitrogen status affects crop yield and biomass (Munoz-Huerta et al., 2013). The nitrogen status of crops is a key indicator of crop growth, yield production, and grain quality (Zhang & Tian, 2009). Nitrogen is the main limiting nutrient for plant growth and thus agricultural production (Berger et al., 2020).

Understanding the interactions among agricultural processes, soil characteristics, and plants is necessary for optimizing crop yield and productivity. Such plant characterization is extremely important for the identification of anomalous areas, extracting critical information for management (Falcon et al., 2021). To reduce existing yield gaps in farms, the solution lies in understanding factors limiting yield in areas with agricultural intensification potential. Through

this integrated approach, it was possible to identify both consistent and agronomically-specific factors limiting crop yield. Yield gaps in corn in small holder farms are outcomes of a complex interplay of climatic variations, soil fertility gradients, socio-economic factors, and differential management intensities. Dutta et al. (2020) utilized several machine learning approaches to investigate the relative influences of multiple biophysical and crop management features in determining maize yield variability. The random forest partial dependence plots revealed a positive association between farm size and maize productivity. Nonlinear support vector machine boundary analysis for the eight top important variables revealed complex interactions underpinning maize yield response (Dutta et al. 2020).

Overall, our results suggest that for Midwest corn fields that are subdivided into respective productivity zones and where soil nutrient characteristics are known to limit crop productivity, on a good year, the inference would be, if there is more variability observed between zones that can be quantified statistically, the grower can most likely anticipate a poor harvest and/or yield. Conversely, if more uniformity is observed and/or homogeneity, there are not statistically significant differences between zones and that is more consistent with a bumper crop or high yield. However, the study has two key limitations which other studies or further research work could capitalize on. The first limitation of the study is that it was carried out in a single field, other studies could yield more in-depth results by comparing data sets from multiple fields. The other limitation is that the study focused on datasets from one growing season, while a comparison of results from several growing season (temporal data sets) could improve the quality of conclusions drawn from the study. Thus, yielding more informative and beneficial results to the stakeholders.

References

Ahmad, S. A., Fahad, M., & Waqas, M. M. (2020). remote sensing based framework to predict and assess the interannual variability of maize yields in pakistan using landsat imagery. *computers and electronics in agriculture*,178. 105732.

Alfonso, C., Hernamdez, M. D., Echarte, M., Cerrudo, A., & Echarte, L. (2021). Maize transpiration efficiency increases with N supply or higher plant densities. *Agricultural Water Management, 250.*

Basso, B., Dumont, B., Maestrini, B., Scherbak, I., Robertson, P., Porter, J. R., … Rosenzweig, C. (2018). Soil organic carbon and nitrogen feedbacks on crop yields under climate change. *Agricultural & Environmental Letters, 3*(1).

Berger, K., Verrelst, J., Feret, J.-B., Hank, J. Wocher, M., Mauser, W., & Camps-Valis, G. (2020). Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *International Journal of Applied Earth Observation and Geoinformation,* 92(10), 1016.

Blasch, G., Li, Z., & Taylor, J. A. (2020). Multi-temporal yield pattern analysis method for deriving yield zones in crop production systems. *Precision Agriculture, 21*.

Bruns, H. A., & Ebelhar, M. W. (2006). Nutrient Uptake of maize affected by nitrogen and potassium fertility in a humid subtropical environment. *Communications in Soil Science and Plant Analysis, 37.*

Butcher, K., Wick, A. F., DeSutter, T., Chatterjee, A., & Harmon, J. (2016). Soil salinity: A threat to global food security. *Agronomy Journal, 108*(6), 2189-2200.

Corti, M., Gallina, P. M., Cavalli, D., Ortuani, B., Cabassi, G., Cola, G., … Bregaglio, S. (2020). Evaluation of in-season management zones from high resolution soil and plant sensors. *Agronomy, 10*(8).

Duncan, J. M. A., Dash, J., & Atkinson, P. M. (2015). The potential of satellite-observed crop phenology to enhance yield gap assessments in smallholder landscapes. *Front Environ. Sci.*

Dutta, S. Chakraborty, S., Goswami, R., Banejee, H., Majumdar, K., Li, B., & Jat, M. L. (2020). Maize yield in smallholder agriculture system—an approach integrating socio-economic and crop management factors. P*one.*

Falcon, H., Wainwright, H. M., Dafflon, B., Ulrich, C., Soom, F., Peterson, J., . . . Hubard, S. S. (2021). Influence of soil heterogeneity on soybean plant development and crop yield evaluated using time-series of uav and ground-based geophysical imagery. S*cientific Report*, *11*(1), 7046.

Fischer, R. A., Byerlee, D., & Edmedes, G. O. (2014). Crop yields and global food securing: Will yield increase continuing to feed the world? *ACLAR Monograph No. 158*. Australian Centre for International Agricultural Research.

Forkuor, G., Hounkpatin, O. L., Welp, G., & Thiel, M. (2017). High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. *Plos One*, *12*(1), e0170478.

Fu, W., Tunney, H., & Chaosheng, Z. (2009). Spatial variation of soil nutrients in a dairy farm and its implications for site-specific fertilizer application. *Soil & Tillage Research, 106*(2), 185-193.

Ge, Y., Thomasso, J. A., & Sui, R. (2011). Remote sensing of soil properties in precison agriculture: a review. *Frontiers of Earth Science, 5*(3), 229-238.

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science and Technology*, *72*.

Hassan, M. A., Yang, M., Rasheed, A., Jin,S., Xia, S., Xiao, Y., & He, Z. (2018). Time series multispectral indices from unmanned aerial vehicle imagery reveal. senescence rate in bread wheat. *Remote Sensing.*

Hatfield, J., Boote, K. J., Kimball, B. A., & Ziska, L. (2011). Climate impacts on agriculture: Implications for crop production. *Agronomy Journal, 103*, 351-370.

Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes, 10*(A), 4-10.

Hedley, C. (2015). The role of precision agriculture for improved nutrient management on farms. *Journal of the Science of Food and Agriculture*, *25*(1), 12-19.

Ishwaran, H., & Lu, M. (2019). Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival. *Statistics in Medicine, 38*(4).

Jeong, J. H., Resop, J. P.., Mueller, N. D., Fleisher, D. H., Yun, K. Butler, E. E. … Kim, S.-H. (2016). Random forests for global and regional crop yield predictions. P*one.*

Khaki, W., Wang, L., & Archontoulis, S. V. (2019). A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science., 10*, 1750.

Kim, N., & Lee, Y.-N. (2016). Mmachine learning approaches to corn yield estimation using satellite images and climate data: A case of iowa state. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 34*(4), 383-390.

Kuhnen, S.,Ogliari, J. B., Dias, E., Boffo, E. F., Correia, I., Ferreira, A. G., . . . Maraschin, M. (2010). ATR-FTIR spectroscopy and chemometric analysis applied to discrimination of landrace maize flours produced in southern Brazil. *International Journal of Food Science and Technology, 45*(8), 1673-1681.

Lai, J., Zou, Y., Zhang, J., & Peres-Neto, P. (2021). Cold Spring Harbor Laboratory. rdacca.hp: An R package for generalizing hierarchical and variation partitioning in multiple regression and canonical analysis.

Lawler, J. J., White, D., Neilson, R. P., & Blaustein, A. R. (2006). Predicting climate induced range shifts: Model diffrences and model reliabilty. *Global Change Biology,* pp. 1365-2486.

Lehmkuhl, F., Nett, J. J., Potter, S., Schulte, P., Sprafke, T., Jary, Z., . . . Hambach, U. (2021). Loess landscapes of Europe-mapping, geomorphology, and zonal differentiation. *Earth Science Reviews, 215.*

Lesk, C. Rowhani, P., & Ramankutty, N. (2016). Influence of extreme weather disasters on global crop production. *Nature, 529*(7584), 84-87.

Maestrini, B., & Basso, B. (2018). Drivers of within-filed spatial and temporal variability of crop yield across the us midwest. *Scientific Reports, 8*(1).

Maestrini, B., & Bruno, B. (2018). Predicting patial patterns of within-field crop variability. *Field Crops Research, 219*(9), 106-112.

Makanza, R., Zaman-Allah, M., Cairns, J. E. Magorokosho, C., Trekegne, A., Olsen, M., & Prasanna, B. M. (2018). High-throughout phenotyping of canopy cover and senescence in maize field trials using aerial digital canopy imaging. *Remote Sensing, 10*(2), 330.

Mayer, B., Tadler, S., Rothenbacher, D., Seeger, J., & Wohrle, J. (2020). A hierarchical algorithm for multicentric matched cohort study designs. *Current Medical Research and Opinion, 36*(11), 1889-1896.

Munoz-Huerta, R., Guevara-Gonzalez, R. G., Contreas-Medina, L. M., Torres-Pacheco, I., Prado-Olivarez, J., & Ocampo-Velazquez, R. V. (2013). A review of methods for sensing the nitrogen status in plants: advantages, disadvantages and recent advances. *Sensors, 13*(8).

Mzuku, M., Khosla, R., Reich, R., Inman, D., Smith, F., & MacDonald L. (2005). Spatial variability of measured soil properties across site-specific management zones. *Soil Science Society of America Journal, 69*, 1572-1579.

Olden, J. D., Lawler, J. J., & Poff, N. L. (2008). Machine learning methods without tears: A primer for ecologists. *Quarterly Review of Biology, 83*(2),171-193.

Oldfield, E. E., Bradford, M. A., & Wood, S. A. (2018). Global meta-analysis of the relationship between soil organic matter and crop yields. *Soil, 5*, 15-32.

Osco, L. P., Kamos, A. P. M., Pereira, D. R., & Saito Moriya, E. A. (2019). Predicting canopy nitrogen content in citrus-trees using random forest algorithm associated to spectral vegetation indices from uav-imagery. *Remote Sensing, 11*(24), 2925.

Rotter, R. P., Hoffmann, M. P., Koch, M., & Muller, C. (2018). Progress in modelling agricultural impacts of and adaptation to climate changes. *Current Opinion in Plant Biology, 45*(B), 255-261.

Shang, J., Liu, J., Ma, B., Zhaoi, T., Jiao, X, Geng, X., . . . Walters, D. (2015). Mapping spatial variability of crop growth conditions using rapideye data in northern Ontario, Canada. *Remote Sensing of Environment, 168,* 113-125.

Snider, B., McBean, E. A., Yawney, J., Gadsden, S. A., & Patel, B. (2021). Identification of variable importance for predictions of mortality from covid-19 Using AI models for Ontario, Canada. *Front Public Health, 9*.

Solonechmyi, P., Kozachenko, M., Vasko, N., Gudzenko, V., Ishenko, V., Kozelets, G., . . . Vinyukov, A. (2018). AMMI and GGE biplot analysis of yield performance of spring barley (Hordeum vulgare L) varieties in multi environment trials. *Agriculture and Forestry, 64.*

Song, Q., Hu, Q., Zhou, Q., Hovis, C., Xiang, M., Tang, H., & Wu, W. (2017.) In-season crop mapping with GF-1/WFV data by combining object-based image analysis and random forest. *Remote Sensing in Agriculture and Vegetation, 9*(11), 1184.

Southworth, J., Randolph, J. C., Habeck, M., Doering, O. C., Pfeifer, R. A., & Johnston, J. J. (2000). Consequences of future climate change and changing climate variability on maize yields in the midwestern United States. *Agriculture, Ecosystems and Environment, 82*, 139-158.

Sozzi, M., Marinello, F., Pezzzuolo, A., & Sartori, L. (2018). *Benchmark of satellites image services for precision agricultural use.* Wageningen, DC: Department of lLnd, Environment, Agriculture and Forestry, University of Padova.

Stadler, A., Rudolph, S., Kupisch, M., Lagensiepen, M., Frank, M., & Frank, E. (2014, September). Qualifying and modeling crop growth heterogeneity at the field scale. [Conference paper]. Presented at Jahrestagung der Gesellschaft für Pflanzenbauwissenschaften, Vienna, Austria.

Tebaldi, C., & Lebell, D. B. (2008). Towards probabilistic projections of climate change impacts on global crop yields. *Geographical Research Letters, 35*, 1-6.

Van Ittersum, M. K., Cassman, K., Grassini, P., Wolf, J., Tittenell, P., & Hochman, Z (2013). Yield gap analysis with local to global relevance–a review. *Field Crop Research, 143,* 4-17.

Wang, J., & Hu, X. (2021, July 9). Research on corn production efficiency and influencing factors of typical farms: Based on data from 12 corn-producing countries from 2012 to 2019. *Plus One*, pp. 1-17

Whetton, R. L., Zhao, Y., Nuwai, S., & Mouazen, A. M. (2021). Modelling the influence of soil properties on crop yields using a non-linear nfir model and laboratory data. *Soil Systems, 5*(1).

Williams, K. P., Ali, M., Griffiths, A. J., & Jones, D. L. (2007). Evaluating the growth characteristics of lettuce in vermicompost and green waste compost. *European Journal of Soil Biology, 43.*

Xie, Q., Dash, J., Huang, W., Peng, D., Qin, Q., Mortimer, H., . . . Ye, H. (2018). Vegetation indices combining the red and red-edge spectral information for leaf area index retrieval. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 1482-1493.

Yildrim, J., Barutcular, C., Koc, M., Dizlek, H., Hossain, A., Islam, M. S., . . . El Sabagh, A. (2018). Assessment of the grain quality of wheat genotypes grown under multiple environments using GGE biplot analysis. *Fresenius Environmental Bulletin, 27*.

Zahedi, M., Delivan, F., Dalvand, S., & Gheshlagh, R. G. (2020). Examination of the relationship of knowledge of diabetes, attitude toward diabetes, and health literacy with diabetes management self-efficacy using hierarchical multiple regression modeling. *Clinical Diabetology, 9*(6), 394-399.

Zhang, H., & Srinivasan, R. (2021). A biplot-based PCA approach to study the relations between indoor and outdoor air pollutants using case study buildings. *Buildings, 11*.

Zhang, X.-H., & Tian, Q.-T. (2009). Hyperspectral evaluation of nitrogen accumulation in winter wheat leaves based on continuum-removed method. *The International Society for Optical Engineering, 10*(12), 1117.

Zhang, Y., Han, W., Niu, X., & Li, G. (2019). Maize crop coefficient estimated from uav-measured multispectral vegetation indices. *Sensors, 19*(1), 5250.

Zhou, Q., Zhang, B., Jin, J., & Li, F. (2020). Production limits analysis of rain-fed maize on the basis of spatial variability of soil factors in north china. *Precision Agriculture, 21*(6), 1187-1208.

Zhu, B., Chen, S., Cao, Y., Xu, Z., Tu, Y., & Han, C. (2021). A regional maize yield hierarchical linear model combining landsat 8 vegetative indices and meteorological data: case study in Jilin Province. *Remote Sensing, 13*(3), 356.