ABSTRACT

OHLETZ, JANEL LOUISE. Understanding Nutrient Status of High Yielding Corn Varieties Using Hyperspectral Imaging and Corn Yield Contest Data. (Under the direction of Dr. Jeffrey G. White).

One traditional approach for assessing crop nutrient status has been to use tissue and soil test results to determine crop macro- and micronutrient adequacy. These results would be used for nutrient management decisions referencing sufficiency ranges (SR) to determine crop nutrients needs and fertilizer rates (Campbell et al., 2000). Equal to the importance of assessing crop nutrient needs prior to the season is assessing the nutrient status within the growing season. Recently, the collection of high-resolution spectral data by remote sensing hyperspectral cameras has provided greater insight into chemical and biological processes in a non-invasive manner. We had three main objectives. First, to determine limiting factors among soil and plant tissue results of high-yielding corn to guide better agronomic recommendations. Second, assess which system better determined nutrients correlated with yield through comparing two multivariate and one machine learning (ML) approach for analyzing tissue and soil nutrient data. Third, to study the application of hyperspectral imaging (HSI) technology for detecting diurnal variation of canopy reflectance in a trial of two hybrid corn (*Zea mays* L.) varieties under two irrigation regimes (irrigated and non-irrigated). Two sets of study data were analyzed to achieve these objectives. One study evaluated the nutrient status of soil and plant tissue sampled at four growth stages (<V6, >V6, VT-R1, and >R6), management practices, and yield collected in 2017 and 2018 from growers intending to enter the NC Corn Yield Contest using multivariate statistics to determine relationships among study variables. The second study utilized site-specific, HSI data collected from a mast-mounted off-nadir view camera overlooking a corn field. For this study we used three ML algorithms: a support vector machine, an artificial neural network, and

a decision tree to analyze the diurnal changes in canopy reflectance to determine if the image data could reliably classify the experimental treatments. The highest yield in the NC Corn Yield Contest study occurred in 2017 (15.1 Mg ha^{-1}). Most individual nutrients among samples were not highly correlated to yield with the exception of tissue Mg (R^2=0.28). Tissue N, P, and Mg were below the SR in ≥ 50% of samples for at least one growth stage. With exception of S, the soil nutrient status in >90% of fields were at or above levels whereby a yield response to fertilizer would be low. Using the Diagnosis and Recommendation Integrated System (DRIS), the seasonal mean of the nutrient balance index for macronutrients had a slightly higher correlation to yield (R^2 = .32), but only in 2017. Conversely, a least absolute shrinkage and selection operator (LASSO) regression model resulted in a higher correlation to yield (R^2 = .48) using both tissue and soil macronutrients taken at tasseling; while the addition of soil parameters to LASSO or ML analyses improved model performance. Likewise, the addition of management practices greatly improved the performance of these models, indicating that relative to the soil and tissue parameters tested, management practices had the greatest influence on yield models in our study. For the hyperspectral imaging study, the ML algorithms were able to distinguish between hybrids but not irrigation regimes, although the yield differences were between irrigation regimes and not hybrids. Classification performance was highest among images of the NE field which received forward scattering light, but the performance varied by the time of image capture. The influence of environmental conditions, soil type, and location create a challenge for improving nutrient recommendations to reach yield potential. More robust field studies with greater participation from across NC are needed to gain a better insight into complex agricultural systems.

Understanding Nutrient Status of High Yielding Corn Varieties Using Hyperspectral Imaging and Corn Yield Contest Data.

by
Janel Louise Ohletz

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Soil Science

Raleigh, North Carolina
2020

APPROVED BY:

_____	_____
Jeffrey G. White. Committee Chair	Carl Crozier
_____	_____
Michael Kudenov	Cranos M. Williams

DEDICATION

This dissertation is dedicated to farmers the globe over and the soil over the globe; without which I would be hungry, naked, and cold. Apart from the food, fiber, and fuel that they provide, soil provides one additional vital component essential to life — it replenishes my soul when I dig my hands into the richness under my feet. Teaming with life and treasures yet discovered, soil is the linchpin of our survival.

In memory of Dr. James Brent Loy, an advisor, mentor and dear friend;

he seeded a legacy that will endure.

BIOGRAPHY

Janel was born and raised on a small farm in rural New Hampshire. Always up for a new experience, she was a high school exchange student for a year in Pitea, Sweden; a wrangler on a dude ranch in Allenspark, CO; and in Paris, France she broadened her love of all things food. Prior to returning to school for her bachelor's degree in Sustainable Agriculture and Food Systems from University of New Hampshire, she worked in the hospitality industry as a chef at numerous northern VA restaurants and was chef/owner of the Wakefield Inn and Restaurant, Wakefield, NH. She continued her studies at UNH and in 2015 earned a M.S. in Agricultural Science studying the effect of grafting melons to interspecific hybrid squash for season extension, disease resistance, and increased yields with Dr. J. Brent Loy. She is deeply committed to working toward a more sustainable and equitable food system. She believes in making an impact by being part of the conversation, whether that be to advocate for a change in our food and agricultural systems or speaking to new and emerging farmers about how good food starts by building from the ground up.

ACKNOWLEDGMENTS

First and foremost I want to acknowledge my advisor Dr. Jeffrey G. White for giving me this opportunity to work on a fascinating and truly interdisciplinary research project. Your support and understanding throughtout this process was a key ingredient in the final product. I also want to thank my committee members: Dr. Carl Crozier, Dr. Michael Kudenov, and Dr. Cranos M. Williams, for offering sage advice and words of encouragement along the way. I am thankful for all the numerous NCSU professors who challenged and guided me through the world of crop and soil science. I especially wish to thank Dr. Dean Hesterberg for taking the extra time to demystify chemistry, Dr. Randy Wells for a good dose of humor during my first semster at NCSU, and Dr. Bob Patterson who willingly engaged in thought-provoking conversation pertaining to food justice, sustainable agriculture, and equitable farming. I am also thankful for encouragment and support from friends: Caner Ferhatğolu, April Maxwell, and Donna Liebelt. Also a special thank you to the cardiologists at UNC, without whom I might not be writing this.

The person I am today was shaped many years ago on a farm in NH. We did not have much, but we had love. My parents taught me that anything can be achieved through hard work, perseverance, and pride in oneself. Thank you Ken and Joan Jackson, your love and support through all my adventures are truly a treasure to cherish. My daughters Amelia, Lauren, and Olivia who were my strength when the going got tough. Lastly, I could not have done this without the unconditional love and support from my husband Stefan who believed in me when I did not. Always and Forever.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

Investigating the Management Practices and Nutrient Status of High-Yielding Farmer-Grown Corn in North Carolina

Abstract

NC is ranked 18th in the US for area of field corn (*Zea mays* L.) planted, but it ranks lowest for yield regionally: an average of 8.9 Mg ha^{-1} (USDA, NASS, 2016). Traditionally, nutrient management decisions in NC use sufficiency ranges (SR) to determine crop nutrients needs and fertilizer rates (Campbell et al., 2000). Our objective was to determine limiting factors among soil and plant tissue results of high-yielding corn growers in NC to guide better agronomic recommendations. In 2017 and 2018, we evaluated the nutrient status of soil and plant tissue sampled at four growth stages (<V6, >V6, VT-R1, and >R6) from growers intending to enter the NC Corn Yield Contest. Grower management practices and yield were collected and analyzed using multivariate statistics to determine relationships among these and results from soil and tissue samples. The highest yield occurred in 2017 (15.1 Mg ha^{-1}). Most individual nutrients among samples were not highly correlated to yield with the exception of tissue Mg (R^2=0.28). Tissue N, P, and Mg were below the SR in ≥ 50% of samples for at least one growth stage. With exception of S, the soil nutrient status in >90% of fields where at or above levels whereby a yield response to fertilizer would be low. The influence of environmental conditions, soil type, and location create a challenge for improving nutrient recommendations to reach yield potential. More robust field studies with greater participation from across NC are needed to gain a better insight into this complex system.

Introduction

Field corn (Zea mays L.) is an important commodity crop and a significant contributor to both the US and NC economies. In 2019, the value of U.S. field corn production was US$52.9 billion (USDA, 2020). From 1986 to 2017, corn production in the United States has nearly doubled from 515 to 911 million Mg ha^{-1}. The higher U.S. production is due to an increase in average corn yields from 7.5 to 11.1 Mg ha^{-1} rather than increased acreage (USDA, 2016). Much of the historic yield increase can be attributed to advances in corn hybrids through breeding programs focused on higher yielding varieties Modern corn hybrids have been shown to produce significant yields at increased planting densities, mainly due to years of selective breeding to improve plant productivity by choosing phenotypes with greater leaf area index and radiation use efficiency (Hammer et al., 2009; Tollenaar & Aguilera, 1992). Corn hybrids introduced since the late 1980's have higher kernel density, weight, and N concentration (Elmore, Sawyer, & Boyer, 2019; Vyn & Tollenaar, 1998), potentially resulting in high-yielding corn having distinctive nutrient needs compared to older hybrids.

The majority of field corn is used in animal feed and ethanol production, and a small percentage for products and additives in the commercial food sector (USDA, 2018). North Carolina is among the nation's top pork and poultry producers, with an annual production value of almost US$6 billion, leading to a large demand for grain corn (USDA NASS, 2019). Corn demand is projected to increase due to the increased demand for animal products driven by population growth and development, and the desire for more renewable energy sources such as ethanol. The environmental impact of the food production system has become more apparent recently and has led to increased pressure on the agricultural industry to become more sustainable. One

option to improving agricultural sustainability in NC is to increase locally grown grain. However, current NC corn production is not able to meet demand, and much of the grain is imported from out of state. As of 2017, NC ranked 18th among the states in grain corn production with 840,000 production hectares yielding just over 3 million Mg. However, the state has one of the lowest average yields in the Southeast: 8.9 Mg ha^{-1} (USDA, 2019) (Table 1-1). There is a limit to the arable land suitable for corn production, thus in order to keep pace with demand, the bushels per acre would need to increase. This puts pressure on NC corn growers to improve yields while ensuring environmentally friendly farming practices. Intensive agricultural practices can have detrimental consequences on the health of the soil and the environment, and increasing agricultural inputs affects the profitability and the economic sustainability of NC farmers (Baumhardt, Stewart, & Sainju, 2015; Tonitto et al., 2006).

North Carolina is surrounded by states with higher yields; thus NC likely has the potential to increase yield. Corn growers in NC have shown that yields >12.5 Mg ha^{-1} are possible as demonstrated by the entrants in the NC Corn Yield Contest. The contest is held annually, and growers are awarded for the highest yield within their division. In recent years, contest yields have been over 18 Mg ha^{-1}. Comparing the NC Corn Yield Contest results to the statewide average yield shows there is a yield gap in NC of almost 6.2 Mg ha^{-1}. Many NC growers have already incorporated high-yielding corn varieties, yet there appear to be factors limiting yield.

Table 1-1. Southeastern US corn yield data: area harvested, yield per ha, and total production by state for 2017 and 2018.

	Area harvested		Corn yield (Mg ha^{-1})		Total production (Mg)	
	Year					
State	2017	2018	2017	2018	2017	2018
	————————————1,000s————————————					
	———ha———		———Mg ha^{-1}———		———Mg———	
Alabama	95	101	10.5	9.8	997	991
Georgia	99	115	11.0	11.0	1,095	1,274
N. Carolina	340	336	8.9	7.1	3,030	2,383
S. Carolina	132	125	8.5	8.0	1,123	1,000
Virginia	138	132	8.5	9.2	1,209	1,207

One way to increase in-state corn supply is to determine the barriers to higher yields. Additionally, understanding the nutrient and management decisions needed to increase yields, without increasing the potential environmental costs of pollution to ground- and surface waters, would help shift corn production in NC toward a more sustainable future. However, economic stability is also important for sustainability, and honing fertilizer rates to reduce overapplication of nutrients can increase profitability of corn production. Sustainability goals for corn growers are important for the ability to reach harvest goals and maintain best practices that help to improve soil health; water and air quality; and also return a good profit. To have the most efficient application, a nutrient recommendation plan should take into account multiple factors including the needs of the crop, soil type, potential weather impacts, and the stage of growth (Khiari, Parent, & Tremblay, 2001; Long, Assefa, Schwalbert, & Ciampitti, 2017); but must also consider the economic impact to the grower.

Tissue analysis can provide an idea of the nutrient status of the crop. Sufficiency ranges (SR) and critical values (CRV) for the southeastern United States are based on research conducted at multiple universities and research institutions to establish standard ranges of plant nutrient

concentrations at specific growth stages to help diagnose nutrient problems and maintain healthy crops (Baker et al., 2000). The premise of the SR and CRV plant analysis systems is that healthy crops contain nutrient elements in predictable concentrations. The crop tissue CRV has been defined as the level beyond which a 5 to 10% yield increase may be realized (Baker et al., 2000). The SR is generally above the critical value and below the point when nutrient concentration is in excess and yield may begin to decrease due to mild toxicity. These concentrations can then be used to establish benchmarks for excessive and deficient levels of nutrients that can help to manage fertilizer application. The SR and CRV values were extrapolated from research and established using a benchmark of the ideal nutrient levels from response curves (Campbell and Plank, 2000). These studies were not designed specifically to establish SR but rather to measure crop response to specific fertilizer application rates and calibrate soil tests (Campbell and Plank, 2000). Therefore, it is likely that the studies did not take into account other abiotic and biotic factors influencing plant growth. However, much of this research was done more than three decades ago using older hybrids with lower yield potential compared to today's hybrids (Jones et al., 1990).

The plant nutrient uptake system is very complex and compounding this are influences outside of agricultural management such as weather and genetics (Jones et al., 1990; Bruetsch and Estes, 1976; Woli et al., 2018). Additionally, adjusting nutrients based on a single deficient nutrient may not address the problem, as complex synergistic and antagonistic relationships can confound the interpretation (Rietra, Heinen, Dimkpa, & Bindraban, 2017). Sufficiency ranges are based on research performed in a specific geographical area that may not match the field conditions of most NC fields. The diversity of soil across NC increases the complexity in

developing nutrient recommendations that will work for all sets of circumstances. Different soil types have been shown to influence nutrient uptake (Fribourg, Bryan, Lessman, & Manning, 1976). For instance, in eastern NC, known as the Coastal Plain, there are many sandy soils along with considerable drained land. Drained lands can be highly productive but are susceptible to flooding from large rain events. Whereas the Piedmont soils, which are in the central part of the state, contain large amounts of clay and have very different nutrient and irrigation needs. Assessing the soil characteristics is an important component to a well-informed nutrient management strategy.

Soil analysis is a useful guide to determine the baseline of nutrient availability and physical and chemical characteristics for optimum crop growth. Generally taken prior to the start of the growing season, soil samples can allow for adjustments to be made to the soil prior to planting such as liming to increase pH, and fertilizer application to adjust nutrient deficiencies. However, fertilizer recommendations based solely on soil sample results may not be accurate as inconsistencies can occur depending on sampling time and technique (Sabbe and Marx, 1987).

It is important to understand environmental influences and nutrient deficiencies limiting corn yields to realize a resilient food system for the future demands of population growth. The timely diagnosis of crop nutrient deficiencies is vital for optimum yield outcomes. It can prove difficult, however, to pinpoint the best approach to obtain reliable results to enable management decisions to address the deficiency before there is a reduction in yield potential in that year (Jones et al., 1990).

The sustainability of field corn production needs solutions that account for variability within the agriculture production system and an understanding of the weak links that may be

limiting yield. There has been little investigation of on-farm agriculture management practices in high-yielding field corn along with soil and plant nutrient status throughout the growing season. A few studies utilized data collected in the National Corn Yield Contest to investigate the yield response to varying climactic and management variables across the Corn Belt (Carter, Riha, Melkonian, & Steinschneider, 2018; Long et al., 2017). However, we found no studies conducting soil and tissue sampling throughout the growing season of contest fields over a range of farmer-grown corn representing a wide range of management styles and growing conditions. Therefore, the goal of our study was to gain a better understanding of factors influencing yield outcomes for high-yield farmer-grown corn in NC. There were three main objectives. First, determine if tissue nutrient concentrations were correlated with yield and within established SRs. Second, determine if tissue nutrient levels and/or grain yield were related to North Carolina Department of Agriculture and Consumer Services (NCDA&CS) soil test values and/or management factors such as irrigation, tillage, hybrid, row width, population, and fertilization. Lastly, determine presence/absence of temporal variability in NCDA&CS soil test and if it was associated with temporal variability of tissue nutrients.

Materials and Methods

Data Collection

In 2017 and 2018, we conducted an observational field study with participating NC corn growers with high yield potential. Solicitation and selection of growers for the study were based on past participation in the NC Corn Yield Contest. The Contest awards high yielding participants in several categories: an irrigated division and six dryland divisions divided by similar climate and soil conditions (Heiniger, 2014). The assumption for selecting Contest entrants was that these

growers have management practices conducive to producing high yields. Criteria for inclusion of growers in the study were recent participation in the NC Corn Yield Contest, intention to enter the Contest in the study years, and field locations in North Carolina. In 2107, we mailed 510 postcards to recent contest entrants with information about the study and a link to the enrollment website. In addition, email invitations were sent to the 13 NCDA&CS regional agronomists, 110 NC Extension Service field agents, and 134 NC agricultural consultants gleaned from the tri-societies "Find a Professional Search". We created a WordPress website describing the study, directed interested readers to an enrollment survey, and provided instructions regarding soil and tissue sampling. These included pre-paid NCDA&CS sample submission forms. We offered to do the sampling but were never requested to do so. In 2018, a large mail campaign was initiated to increase awareness and participation in the study. Letters were mailed to 149 NC Corn Yield Contest entrants from 2017, and 470 letters were mailed to 2018 NC Commodity Conference attendees. Grower entrants, collaborating extension agents, and agricultural consultants self-enrolled via an online Qualtrics survey instrument (Qualtrics, Provo, UT) accessed through the link and QR-code provided on the solicitation postcards, letters, and contest website. Growers could enter more than one field into the study. The study fields were scattered among the three main regions of North Carolina: Coastal Plain, Piedmont, and Mountains. Most of the study entrants were located in the Piedmont and Coastal Plain (Table 1-2).

Table 1-2. Numbers of soil and tissue samples submitted for 2017 and 2018 by region and corn growth stages: early vegetative (< V6); late vegetative (> V6); tasseling (VT-R1); and maturity (>R6) (Hanway, 1966).

Year	Number of individual fields sampled					
	Region					
	Coastal Plain		Piedmont		Mountains	
	With[a]	Without	With	Without	With	Without
2017	22	13	1	13	1	0
2018	16	2	0	5	1	0
	Number of individual tissue and soil samples					
	Tissue	Soil	Tissue	Soil	Tissue	Soil
2017	118	98	26	17	3	3
2018	60	60	5	0	2	2
	Growth stage					
	Early (< V6)		Early (> V6)		Tasseling (VT-R1)		Maturity (> R6)	
	Tissue	Soil	Tissue	Soil	Tissue	Soil	Tissue	Soil
2017	23	31	53	30	38	28	33	29
2018	24	19	15	15	15	15	13	13

[a]Fields with and without associated yield from NC Corn Yield Contest entry.

Individual growers or NC extension agents and agronomists assisting with the study gathered soil and tissue samples from 0.4 hectare plots earmarked for entrance into the NC Corn Yield Contest in the respective year. Tissue and soil samples were collected at the same locations on the same day at four corn growth stages: seedling, <V6 (whole plant sampled); early, >V6 (most recent mature leaf); tasseling/silking, VT/R1 (earleaf); and maturity, R6 (earleaf) (Hanway, 1966). The four sampling stages corresponded to tissue nutrient SR recommended by the NCDA&CS laboratory (Baker et al., 2000). Sampling techniques were consistent with the NCDA&CS recommended methods (McGinnis, Stokes, & Cleveland, 2014). Approximately 20 subsamples of plow-layer soil were collected at each field location in the area allocated for the corn yield contest and bulked. There was a high degree of variability in the number of tissue and soil samples per growth stage and not all fields had samples at all four growth stages.

Samples were submitted to the NCDA&CS laboratory (4300 Reedy Creek Road, Raleigh, NC) for analysis. The NCDA&CS laboratory collaborated in the study by charging grower fees (soil: gratis; tissue $5.00) rather than research fees ($5 and $12, respectively). Soil and tissue samples were analyzed using established methods (Hardy et al. 2014). For soil, these were Mehlich 3: P, K, Ca, Mg, S, Cu, Zn, Mn, and Na (Mehlich, 1984a); cation exchange capacity (CEC) and base saturation (BS%) (Mehlich, Bowling, & Hatfield, 1976); pH, acidity (AC), and lime requirement (Mehlich, 1976); soil class (mineral, mineral-organic, organic); sieved weight-to-volume (w/v); and humic matter (HM%) (D.H. Hardy, 2014; Mehlich, 1984b). For tissue N, the laboratory used O_2 combustion gas chromatography; for the other nutrients, HNO_3 microwave digestion followed by inductively coupled plasma (ICP) spectrophotometry (McGinnis et al., 2014)

Sample analysis results were compiled for each study year, and NCDA&CS soil test index values were converted to quantitative equivalents (Hardy et al., 2014). Tissue results for macronutrients required conversion from percentage to g kg^{-1}; and soil P, K, S, Mn, Cu and Zn were converted to mg kg^{-1}. Soil Ca and Mg were converted from percentage of CEC occupied by the specific cation to mg kg^{-1}. Nutrient ranges from tissue samples were compared to the reference CRV and SR for the southeastern US (Baker et al., 2000). Yield data from the NC Corn Yield Contest were matched, when possible, to the specific field entered. Otherwise, the mean value of all fields for an individual grower was used for all fields associated with that grower. This occured for 40% of the field data in 2017 and none in 2018.

Submitted samples consisted of 147 tissue and 118 soil in 2017, and 67 tissue and 60 soil in 2018. Not all fields had both tissue and soil samples. In 2017, 52 field samples lacked

associated yields. In 2018, 60 field samples were associated with a yield, however, not all sampled fields were entered into the NC Corn Yield Contest but representative yields were obtained. The Coastal Plain represented the highest number of samples followed by the Piedmont and Mountains (Table 1-2). In 2017, 80% of the tissue samples and 83% of the soil samples submitted were from the Coastal Plain. In 2018, submissions from the Coastal Plain increased to 90% of the tissue samples and 97% of the soil samples submitted. In 2018, there were few samples submitted from the Piedmont and one set of samples from the Mountains.

Grower information was collected from both the initial study entry form and from the NCDA&CS soil and tissue test submission forms. The information collected was comprised of grower address, county, planting date, and sample collection date. Certified yields and management practices were captured from the NC Corn Yield Contest entry forms when available. Management data included: planting density; row spacing; fertilizer types (N, P, K, S, and micronutrients), application rates, and timing; tillage practice; corn hybrid; previous crop; water regime; and herbicide and pesticide applications. Management practices varied by grower but included both conventional, conservation, and no-till soil preparation; and irrigated or rain-fed water regimes. In 2018, two locations representing 23 field samples were not entered into the contest. We received relative yield; however, the management practices were not recorded.

Data analysis

Due to insufficient field samples for the mountain region and all but one piedmont field without Corn Yield Contest entry, we were unable to analyze the data by region. The samples from the Piedmont and Mountains were included for general summary statistics on yield, tissue nutrient concentrations, soil characteristics, and soil nutrient content, but removed prior to

performing regression analysis. Summary statistics of management practices were separated by year, and tissue and soil nutrients are presented by year and by growth stage. The statistical analysis of tissue and soil samples was completed using JMP Pro 14 software (SAS Institute, Cary, NC). Analysis of variance (ANOVA) was used to determine effects on yield of irrigation, tillage, use of starter fertilizer, and manure application. Linear regression analysis was performed to investigate the relationships of yield to a number of variables: plant nutrient concentrations; soil nutrients and characteristics; seeding density; row width; and nutrient application rates of N, P, K, and S. Due to a large difference in the number of observations between years, as well as regions within 2018, comparative analyses of years and of regions within years were not possible. We characterized the strength of correlations as $R^2 \leq 0.10$,"very weak"; $0.10 < R^2 \leq 0.25$, "weak"; $0.25 < R^2 \leq 0.50$ "moderate"; $0.50 < R^2 \leq 0.75$, "strong"; and $R^2 > 0.75$ "very strong".

Results and Discussion

Yield

We were not able to obtain yield data for all fields entered in our study, as not all locations were entered into the Corn Yield Contest. In 2017, 59% of fields had associated yields; in 2018, 91% did. In 2017, the Piedmont (4% of fields) and the mountain region (4% of fields) represented a small percentage of the yield data. Similarly, most yield data in 2018 were from the Coastal Plain (94% of fields), and none from the Piedmont. One field located in the mountain region corresponded to the highest yield in both years, 17.1 and 16.7 Mg ha^{-1} (273 and 266 bu ac^{-1}) in 2017 and 2018, respectively. Overall, yield in 2017 was greater than in 2018 (Table 1-3). The lower yield in 2018 was mainly influenced by adverse weather in various parts of the state including: 1) excessive rain early in the season that required replanting in some fields; 2) high

wind mid-season that caused lodging; and 3) two hurricanes just prior to harvest that caused significant flooding in the eastern Piedmont and in the Coastal Plain.

Table 1-3. Main effects of management practices: starter fertilizer, irrigation, manure; and year (2017 and 2018), on mean (± SD) corn grain yield for NC Corn Yield Contest study participants. There were no Treatment X Year interactions (p = .23 to .98).

	Yield			
	Management practice			
Factor	Starter Fertilizer	Irrigation	Manure	Tillage
Treatment	——————————Mg ha^{-1}——————————			
No	14.6 (±1.2)a†	15.1 (±1.5)a	15.1 (±1.5)a	14.4 (±1.6)a
Yes	15.5 (±1.5)a	15.4 (±1.3)a	15.5 (±1.3)a	14.9 (±1.5)a
Year				
2017	15.1 (±1.2)a			
2018	13.8 (±1.3)b			

† Within columns within a factor, means followed by the same letter are not significantly different according to LSD (0.05).

In 2017 and 2018, the growers in our study achieved higher average yields (15.1 and 13.8 Mg ha-1, respectively) than the statewide averages those years (8.9 and 7.1 Mg ha-1, respectively). This equates to a yield gap of 6.2 to 6.7 Mg ha-1 between the average NC corn yield and the high-yielding populations. Similar to the yields in our study, the statewide average corn yield was higher in 2017 than 2018 (USDA, 2018). Over the past decade, both the highest total production and yield per acre of grain corn in NC was in the Coastal Plain (USDA, NASS, 2018). In 2018, nine of the ten highest producing counties were located in the Coastal Plain (USDA, NASS, 2018). The number of study growers in the Piedmont that entered the Corn Yield Contest (one in 2017 and none in 2018) was much lower than the initial number in our study (seven in 2017 and two in 2018); growers with compromised or lower projected yields were not likely to have entered the Contest. The final statistical analysis was performed only on data collected in the coastal plain region, as there were not enough fields from the mountain or piedmont regions for

statistical analysis. Therefore, the study findings focus on the Coastal Plain and conditions of that region. As of the last agriculture census, nine out of the top ten NC counties in total corn production were located in the Coastal Plain (NCDA&CS, 2017). Soils in the NC Coastal Plain are representative of millions of acres of highly productive agricultural land in the southeastern SE US Coastal Plain; thus our results are likely to be relevant to other parts of the southeastern Coastal Plain beyond the borders of NC.

Management practices

Combinations of management practices (tillage, fertilizer rates, preceding crop, irrigation, row spacing, and plant population density) were unique for each field across NC. In 2017 and 2018, respectively: starter fertilizer was applied to 65 and 40% of fields; 63 and 79% of fields were rain-fed; and manure was applied to 35 and 57% of the fields (data not shown). Conventional tillage was used in 53 and 32% of the fields in 2017 and 2018, respectively. Yield was not affected by starter fertilizer application, tillage practice, irrigation, or application of manure in either year (Table 1-3). Planting date and yield had a moderate positive correlation in 2018, but no correlation in 2017 (Figure 1-1).

Plant population densities ranged from 68,520 to 98,860 plants ha^{-1} across all fields and years; the mean was 82,000 plants ha^{-1} (Table 1-4). Row width ranged from 50.8 to 101.6 cm across fields and years; the mean row widths were 66.2 and 73.7 cm in 2017 and 2018, respectively. The application rates of commercial fertilizers and inclusion of micronutrients varied greatly among fields. In general, application rates did not correlate to yield with the exception of P applied in 2017 (R^2 = 0.29) (Figure 1-2).

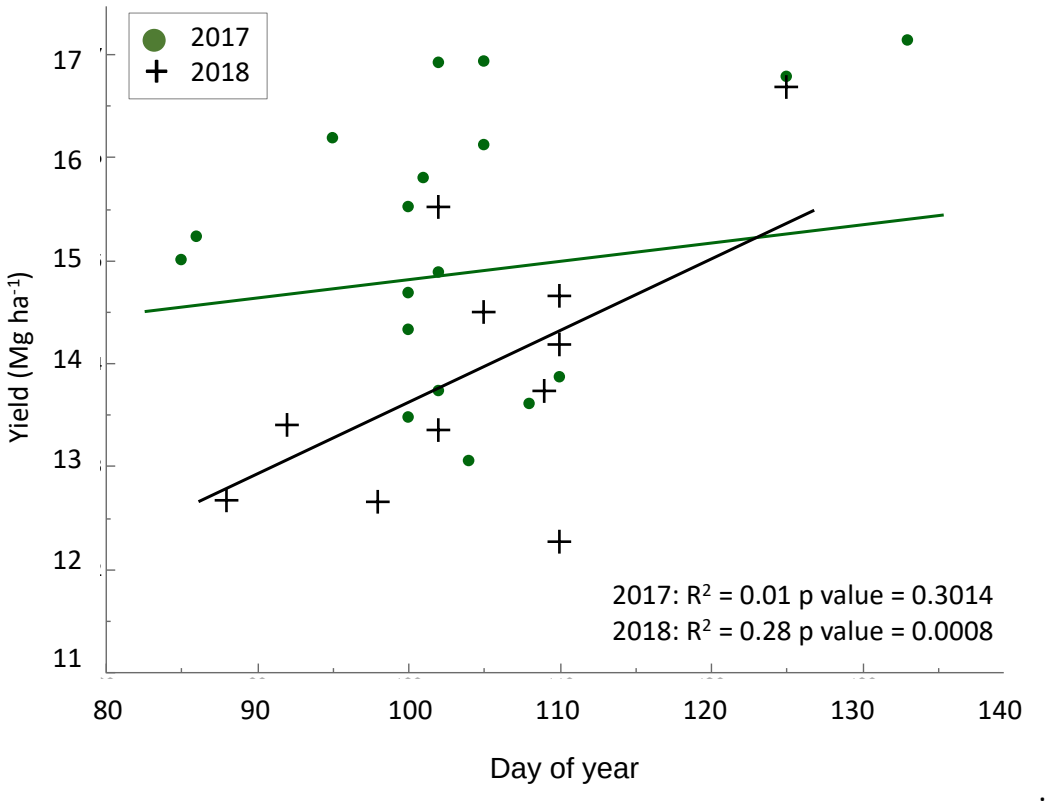

Figure 1-1. Relationship of corn yield (Mg ha-1) as a function of planting date (day of year) for the 2017 and 2018 seasons.

Table 1-4. Effects of year on management practices: mean (SD) for plant population density; row width; and fertilizer application rates of elemental N, P, K and S, used by study participants in 2017 and 2018.

	Management Practice			
Year	**Plant Population Density**		**Row width**	
	————plants ha^{-1}————		————cm————	
2017	82,481 (8919)a[a]		74 (11)a	
2018	80,898 (6726)a		66 (15)a	
	Applied nutrients			
	N	**P**	**K**	**S**
	————————————————kg ha^{-1}————————————————			
2017	229 (62)a	15 (17)b	75 (86)a	16 (19)a
2018	256 (82)a	48 (42)a	126 (83)a	24 (25)a

[a]Within columns, means followed by the same letter are not significantly different according to LSD (0.05).

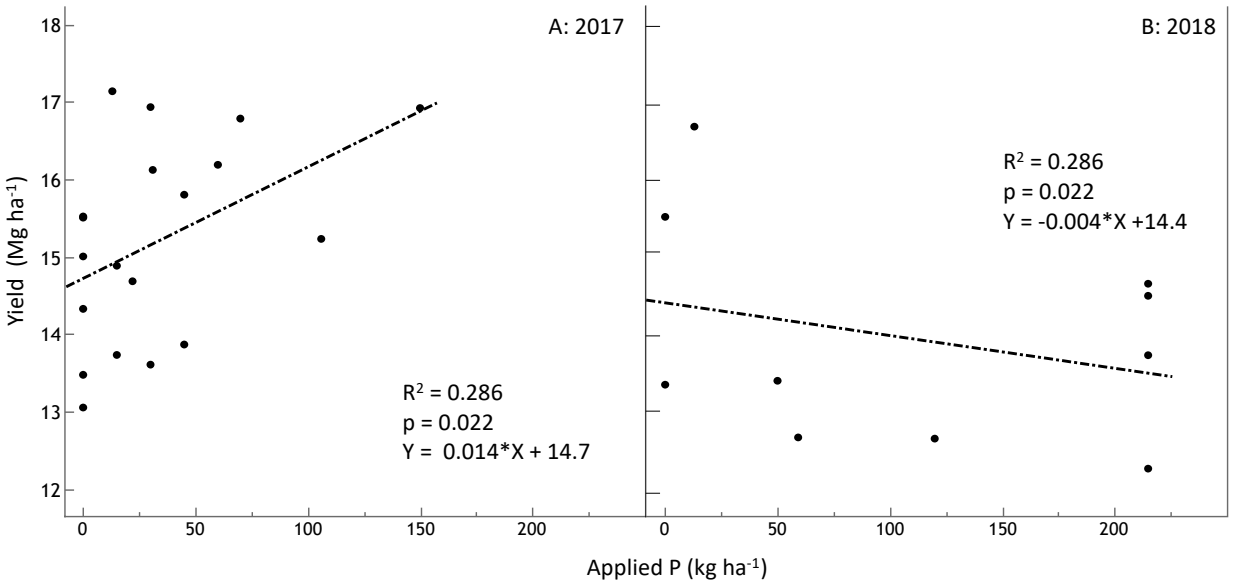

Figure 1-2. Correlation between yield and application rate of fertilizer P for 2017 and 2018 NC Corn Yield Contest study fields. Each black dot represents one field.

The wide range of management practices used by the growers demonstrates their adaptability to their environmental conditions in order to achieve high yields. There were no management practices strongly correlated with yield in either year. Much of this was likely due to confounding effects of weather, soils (Tolk, Howell, & Evett, 1999), and hybrid differences (Echarte et al., 2013), along with a small sample size with a high degree of variability among locations. In general, growers achieving high yields followed practices suggested in the NC Corn Production Guide (Heiniger et al., 1999).

In NC, the recommended plant population density for corn ranges from 56,000 to 81,500 plants ha^{-1} depending on soil type, soil water holding capacity, and desired row width (Heiniger, 2018). Most fields in our study had plant populations within the recommended plant densities; 16 fields had plant densities above the recommended range. Increased plant densities of 81,700 to 107,900 have been found to increase yields (van Roekel & Coulter, 2011). Newer corn hybrids have improved tolerance of physiological stresses at higher plant densities (> 54,000 plants ha^{-1})

(Tollenaar and Wu, 1999), and more upright leaves to allow more light interception deeper into the canopy (Duvick, 2005). Thus, choosing the recommended plant population for a specific hybrid is important. According to Assefa et al. (2017), per plant production has increased; however, much of the per hectare yield increases can be attributed to intensification and best management practices rather than just hybrid improvement. In addition to site-specific influences, the ideal plant density must also account for the irrigation and tillage practices being used. Well-drained soils with irrigation can support a higher plant population (Heiniger et al, 1999).

Increased plant population density may warrant higher nutrient application rates. Among growers in our study, rates for fertilizer N ranged from 132 to 386 kg ha-1 (Table 1-4). In NC currently, the recommended N rate is calculated using a database of soil-dependent realistic yield expectations (RYE) and N-use factors (NUF) that range from 14.3 to 17.9 kg N Mg-1 corn (Rajkovich, Crozier, Smyth, Crouse, & Osmond, 2015). By comparison, the NUF of the growers in our study ranged from 9.3 to 31.5 kg N Mg^{-1} corn, and the mean, 18.6 kg N Mg^{-1} corn, was higher than the suggested NUF for NC, indicating a lower N-use efficiency than predicted in the RYE database. This was not an unreasonable outcome — typically N-use efficiency decreases as N rates increase —and growers targeting maximum yield often apply relatively high rates of N. For corn in the coastal plain region, it has been shown that fertilizer N rates could be reduced by 25% below the standard N rate without significant yield reductions (Austin, Osmond, & Shelton, 2019). Furthermore, they found that environmental conditions had a greater influence on yield than fertilizer N.

The application of P showed a positive influence on yield only in 2017, when application rates were generally lower than in 2018 (Figure 1-2). Additions of fertilizer P may not have been needed, as many NC fields have high levels of soil P from years of animal manure application (Johnson, Osmond, & Hodges, 2005). Though our study was not able to discern the amount of P applied as a starter fertilizer, Cahill et al. (2008) found that in NC the addition of P in starter fertilizer did not lead to corn yield increases compared to N-only starter fertilizer. However, if soil P is not easily accessible to plants due to cool soil temperatures, soil moisture conditions, slow root growth, and soil pH (Clarke et al., 1990; Roberts and Israel, 2017), yields could be impacted by lack of starter P. Therefore, it is likely that yield responses in our study had more to do with the influence of environmental conditions affecting P availability than on P application rates.

Plant tissue nutrient concentrations

There was a high degree of variability in the tissue nutrient levels among fields, especially Mn (Figure 1-3). The variability of individual nutrients was generally similar both years with the exceptions of Ca, Mn, and B, which were more variable in 2017 than in 2018. The degree to which tissue nutrient levels varied among fields can be attributed to many factors including soil type, environment, genetics, and management (Bruetsch & Estes, 1976; Fribourg et al., 1976; D.J. Mulla & Schepers, 1997). It is likely that the high variability of these factors among growers contributed to the lack of strong correlations ($R^2 > .5$) between tissue nutrient concentrations and yield.

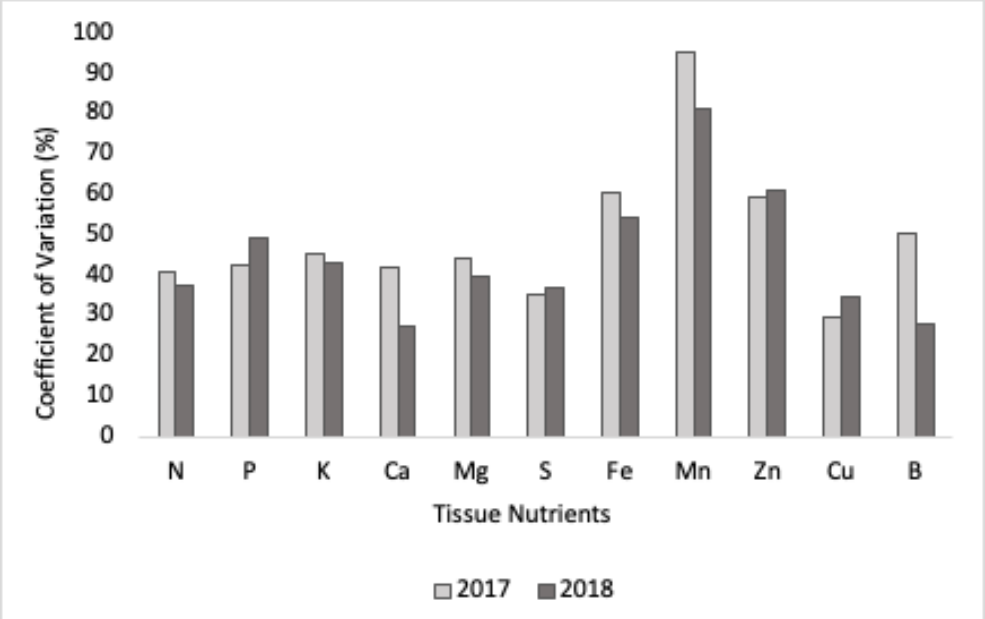

Figure 1-3. Coefficients of variation (CV) for plant tissue macro- and micronutrient concentrations including all four sampling stages in the 2017 and 2018 seasons. Samples taken from fields of NC Corn Yield Contest study participants.

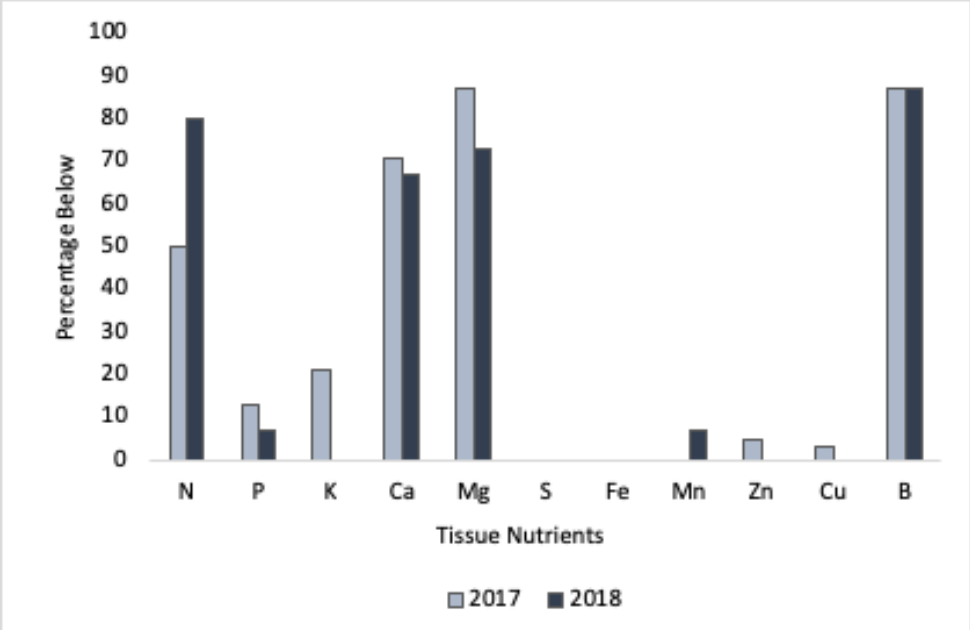

Figure 1-4. Percentage of plant tissue samples below macro- and micronutrient critical values for samples taken from fields of high-yielding corn at the tasseling (VT-R1) growth stage in 2017 and 2018.

Tissue nutrient concentrations differed among growth stages, but not by year (Table 1-5). In the first three growth stages for N, P, and K, and in the first two growth stages for S, concentrations were higher compared to the >R6 growth stage (Table 1-5). Conversely, tissue Ca concentration was highest at maturity (> R6), and tissue Mg concentration did not change markedly over the growing season (Table 1-5). The critical value (CRV) in corn is evaluated at VT-R1. In both years, a high percentage of fields had tissue N, Ca, Mg, and B levels below the CRV (Figure 1-4). In 2017, half or more of the fields were below the CRV for: N (50% of fields), Ca (71%), Mg (87%), and B (87%) (Figure 1-4). In 2018, greater than 50% of fields were below the CRV for: N (80% of fields), Ca (67%), Mg (73%), and B (87%) (Figure 1-4).

The evaluation of nutrient SR is divided into the four growth stages. The status of tissue nutrient concentrations varied among growth stages when evaluated against the SR (Figure 1-5). A comparison of tissue nutrient concentrations to SR by growth stage showed that greater than 50% of samples were below the SR for N, P, and Mg at one or more growth stages in 2017 and 2018 (Figure 1-5). At the <V6 growth stage, over 70 and 50% of tissue samples in 2017 and 2018, respectively, were below the SR for Mg, however this percentage decreased markedly as the season progressed (Figure 1-5). Interestingly, tissue Mg was negatively correlated with yield in 2018 but not in 2017 (Figure 1-6). The concentrations of N, P, K and S decreased over the growing season, while Ca was highest at the >R6 stage (Table 1-5). Conversely, no changes in tissue Mg and micronutrient concentrations were detected over the growing season. Tissue nutrient concentrations can change due to soil conditions influencing uptake (Fribourg et al., 1976),

Table 1-5 . Main effects of year and growth stage on the least square mean and CV (%) of corn tissue macronutrients N, P, K, Ca, Mg, and S in samples taken at four growth stages in 2017 and 2018 from NC Corn Yield Contest study entrants. There were no Year X Growth stage interactions (p = 0.33 – 0.97) with the exception of tissue Ca (p = 0.02).

	Tissue macronutrient											
	N		P		K		Ca		Mg		S	
Factor	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Year	(g kg^{-1})	%	(g kg^{-1})	%	(g kg^{-1})	%	(g kg^{-1})	%	(g kg^{-1})	%	(g kg^{-1})	%
2017	47.6a[a]	42	4.0a	43	46.6a	46	4.3a	45	1.8a	46	2.6a	35
2018	43.4a	39	3.6a	46	46.0a	45	4.2a	27	2.1a	37	2.4a	33
Growth stage												
<V6	47.6a	17	4.0a	41	46.6a	24	4.3b	30	1.8ab	26	2.6a	20
>V7	34.4b	24	3.1b	23	32.4b	28	3.0c	32	1.5b	52	1.9b	32
VT-R1	30.0b	12	3.1b	16	23.8c	16	3.9bc	27	1.7b	38	1.8bc	19
>R6	14.9c	36	1.6c	55	15.7d	48	5.6a	37	2.4a	42	1.4c	38

[a]Within columns within a factor, means followed by the same letter are not significantly different according to LSD (0.05).

physiological processes at specific growth stages (Jones, Eck, & Voss, 1990), and diurnally (Mundorf et al., 2015). Woli, Sawyer, Boyer, Abendroth, & Elmore, (2017) found that P concentration was constant in leaves until the R5 stage, but that N and K concentrations decreased steadily after tasseling.

The use of SRs is an efficient method to determine the nutrient status of a crop, however, multiple factors may influence the accuracy. The concentrations of nutrients within plant organ tissues can change throughout the growing season, and are influenced by both genetics and agronomic conditions (Bender, Haegele, Ruffo, & Below, 2013). Newer-era hybrids compared to 1960-era hybrids have shown lower concentrations of N and P in leaf tissue at growth stages prior to R5, but no difference in K concentrations (Woli, Sawyer, Boyer, Abendroth, & Elmore, 2018). Changes in nutrient accumulation due in part to newer hybrids and management practices may warrant a revision of sufficiency levels. Ciampitti & Vyn, (2014) found that N, P, K, and S uptake rates varied with grain yield, and the dry matter to nutrient ratio increased with yield. This has implications for nutrient management planning: assuming a nutrient removal consistent across yield levels as opposed to a grain-yield-specific rate could lead to incorrect estimations of nutrient uptake in high-yielding corn environments and therefore an over-estimation of nutrient removal by the crop.

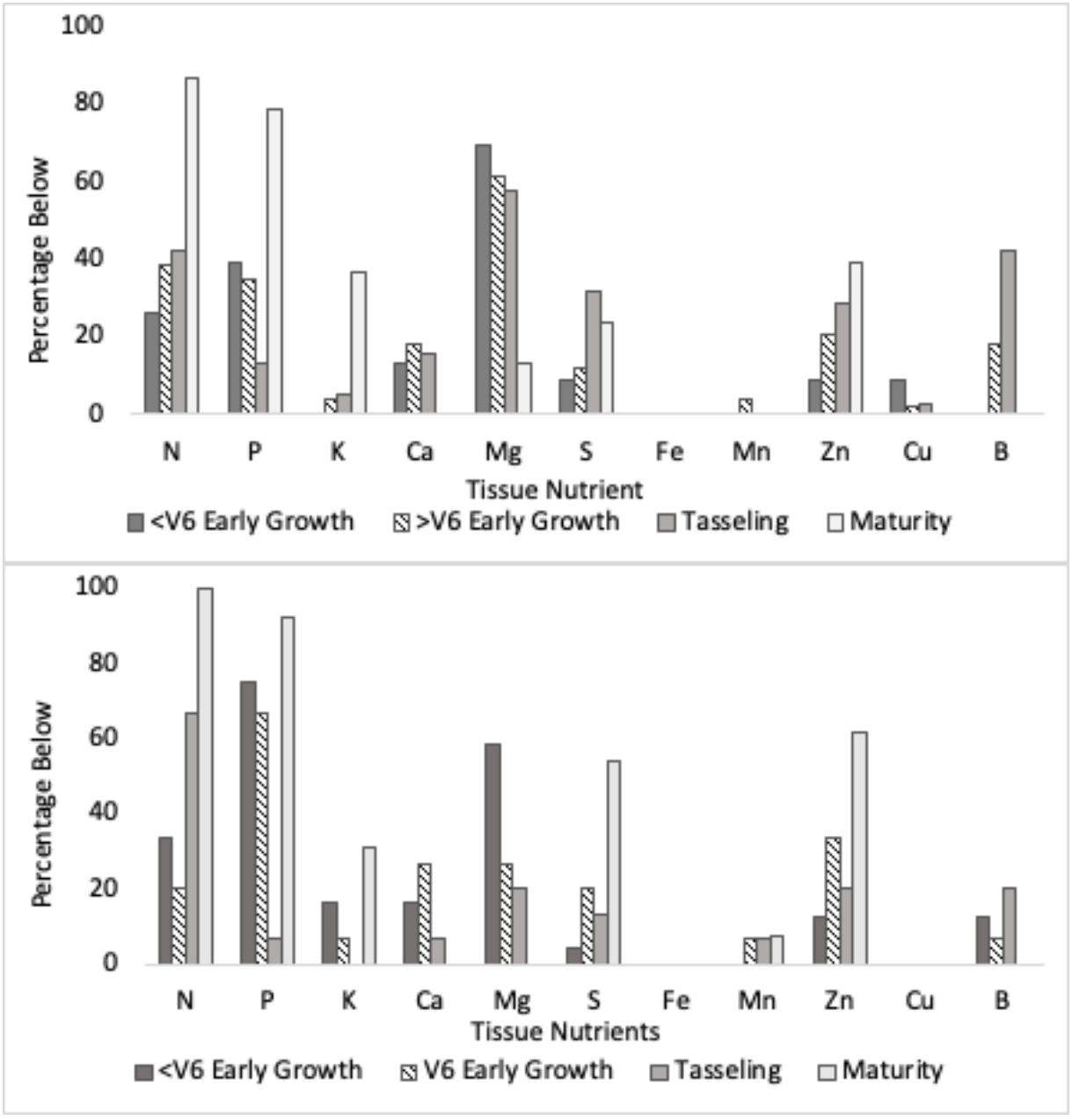

Figure 1-5 . Percentage of plant tissue samples below the lowest sufficiency range value at each of four growth stages (< V6, >V7, VT-R1, and >R6) for macro- and micronutrients in 2017(A) and 2018 (B). Samples taken from fields of NC Corn Yield Contest study participants.

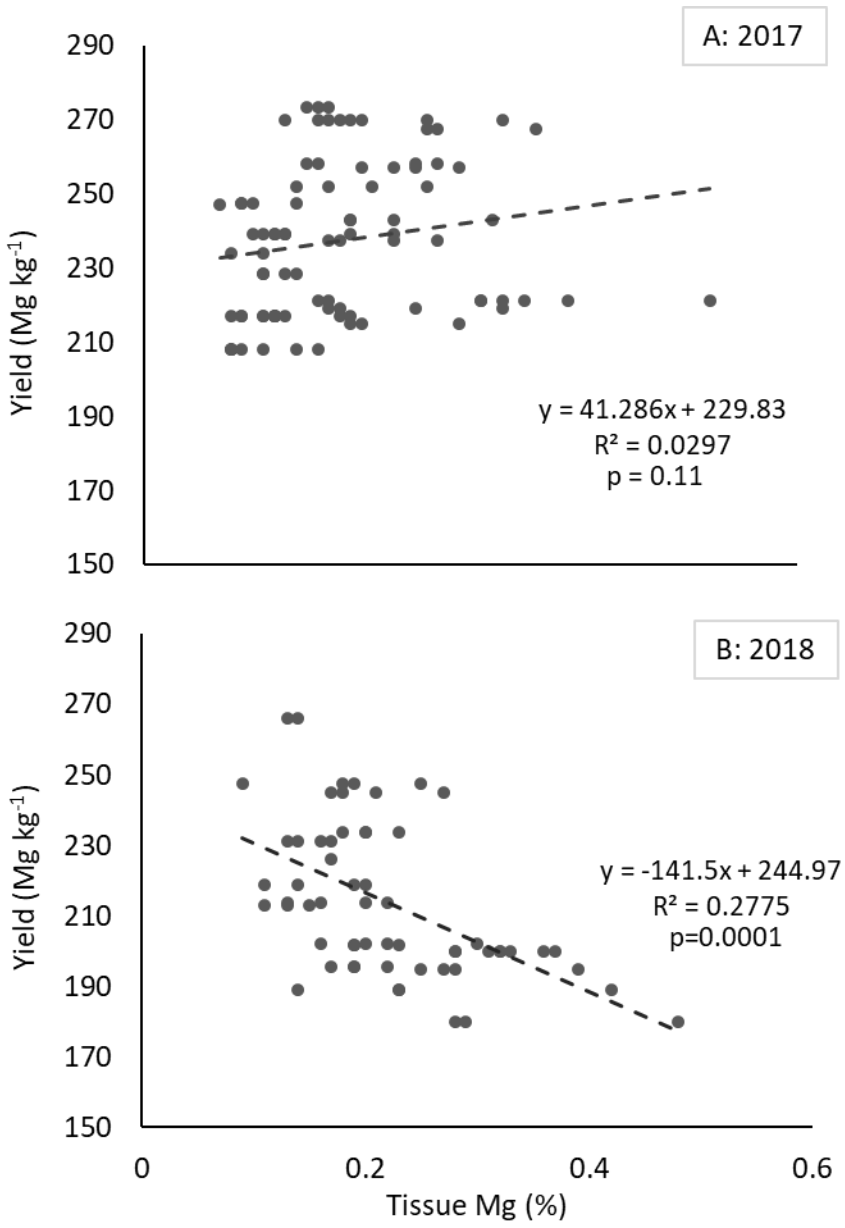

Figure 1-6. Corn yield as a function of plant tissue Mg concentration for samples taken from fields of NC Corn Yield Contest study participants in 2017 and 2018 at four growth stages (< V6, >V7, VT-R1, and >R6).

The tissue nutrient with the highest percentage of samples below both CRV and SR in 2017 and 2018 was Mg. The current CRV for Mg is 2.5 mg kg^{-1} and the SR is 2.0 to 6.0 mg kg^{-1} (Campbell, 2000), but studies have shown that similar yields can be obtained over a wide range of concentrations. Fox and Piekielek (1984) found that tissue Mg concentrations between 1.0 and

3.2 mg kg^{-1} had no effect on yield outcomes. It is likely that using the SR as a predictor of Mg status in the tissue may not be accurate. Perhaps the CRV and SR levels are too high: Walworth and Ceccotti (1990) showed that corn yields greater than 18 Mg ha^{-1} could be achieved with corn ear leaves at tasseling contained Mg levels as low as 1.2 mg kg^{-1}.

Tissue nutrient ratios for N/P, N/K, N/S, and Fe/Mn are used as a means to assess the balance between nutrients that have shown to have strong relationships within the plant. The NCDA&CS considers N/S ratios between 10:1 and 18:1 to be optimum (Campbell, 2000), while Stewart and Porter (1969) considered N/S ratios from 12:1 to 15:1 to be ideal. The N/S ratio in our study ranged from 7.5 to 25:1. Relative to the NCDA&CS bounds, seven percent of samples were below 10:1 and 41% above 18:1. The N/S ratio differed among the growth stages ($p < 0.0001$), but there was no difference between years ($p = .33$), nor a year by growth stage interaction ($p = .55$) (not shown). The mean N/S ratio was 19:1 for the <V6 and >V7 growth stages; 17:1 for VT-R1; and 11:1 at maturity, which was significantly lower than the first three growth stages (data not shown). In 2017, at the <V6 and > V7 growth stages, the N/S ratio was negatively correlated to yield, $R^2 = .55$ and .27, respectively (data not shown). There is a close association of N and S in plant tissue (Fageria, 2001; Terman, Allen, & Giordano, 1973), and both nutrients are required for protein formation. Without adequate S, N compounds will accumulate within the plant tissue until S becomes available.

In general, yield was not correlated with tissue nutrient concentrations, with a few exceptions (Table 1-6). There was a weak positive correlation with S ($R^2 = .09$) in 2017 (Table 1-6) and a negative correlation with Mg ($R^2 = .28$) in 2018 (Figure 1-6). Manganese had a weak negative correlation with yield in 2017 ($R^2 = .05$) and a weak positive correlation in 2018 ($R^2 =$

.07) (Table1- 6). There were pairwise correlations between N, P, K, S, Fe, and Zn in both years. The strongest correlations were all positive and between: N and S (R^2 = .66 and .70 in 2017 and 2018, respectively); N and K (R^2 = .70 and .64); and Fe with Al (R^2 = .74 and .80). Tissue Mg was negatively correlated with: K (R^2 = .12 and .14, in 2017 and 2018, respectively); N (R^2 = .10) and P (R^2 = .13) in 2017; and Mn (R^2 = .10) in 2018 (Table 1-6). Since both N and K are needed in the highest concentration relative to dry matter within the plant, it is understandable that they would be correlated. Potassium, though not a structural component of the plant, is required for regulating physiological processes within the plant such as photosynthesis and transport of assimilates (Dibb & Thompson, 1985; Stromberger, Tsai, & Huber, 1994). It is likely due to the role of K in the assimilation process that K and N are highly correlated (Blevins, 1985).

The negative correlation between tissue Mg and K is likely due to an antagonism between the two cations leading to reduced rates of Mg uptake if K is at high levels in the soil (Dibb & Thompson, 1985; Jakobsen, 1993). Izsáki (2017) found that high levels of applied K that resulted in soil K levels > 300 mg kg^{-1} reduced tissue Mg significantly in the <V6 and VT-R1 stages;

Table 1-6. Correlation matrix with R^2 values and whether the relationship was positive or negative (+/−) for relationships among yield and plant tissue nutrient concentrations for 2017 (upper right) and 2018 (lower left). Bolded values indicate moderate to strong correlations. Empty boxes indicate no correlation.

	Yield	Tissue N	Tissue P	Tissue K	Tissue Ca	Tissue Mg	Tissue S	Tissue Fe	Tissue Mn	Tissue Zn	Tissue Cu	Tissue B	Tissue Na	Tissue Al
Yield							0.09(+)		0.05(-)			0.05(-)		
Tissue N			**0.59(+)**	**0.70(+)**	0.08(-)	0.10(-)	**0.66(+)**	0.08(+)	0.05(-)	**0.30(+)**				
Tissue P		**0.49(+)**		**0.41(+)**	0.13(-)	0.13(-)	**0.34(+)**		0.09(-)	**0.37(+)**				
Tissue K		**0.64(+)**	**0.38(+)**		0.12(-)	0.12(-)	**0.40(+)**		0.13(-)	**0.31(+)**				
Tissue Ca						0.08(+)		0.08(+)	**0.38(+)**		**0.23(+)**			
Tissue Mg	**0.28(-)**			0.14(-)	0.16(+)									
Tissue S		**0.70(+)**	**0.42(+)**	**0.50(+)**				0.10(+)		0.17(+)				
Tissue Fe		**0.56(+)**	0.23(+)	**0.52(+)**			**0.51(+)**		0.08(+)					**0.74(+)**
Tissue Mn	0.07(+)					0.10(-)					0.05(+)			
Tissue Zn		**0.35(+)**	**0.29(+)**	**0.27(+)**			**0.32(+)**	**0.31(+)**				0.11(+)	0.09(+)	
Tissue Cu												0.08(+)		
Tissue B					0.07(+)								0.07(+)	
Tissue Na														
Tissue Al		**0.30(+)**		**0.33(+)**			**0.24(+)**	**0.80(+)**		0.19(+)				

however, this was not associated with a reduction in yield. Ohno and Grunes (1985) found that high soil K levels led to decreased translocation of Mg to the shoot. Jacobson (1999) showed that crop nutrient concentrations are a reflection of soil conditions and chemical activity ratios. Additionally, research by Ortas (2016) found genotype influenced the nutrient accumulation of Mg and K especially at high K fertilizer application rates. Changes in nutrient accumulation are not exclusive to Mg and K. Recent hybrids have resulted in greater accumulation of dry matter and N, P, and K, along with reduced nutrient concentrations, implying greater nutrient-use efficiency (Woli et al., 2017). Therefore, the SR and CRV used to make nutrient recommendations may no longer be relevant and could lead to over application of nutrients. This can increase nutrient losses to ground and surface waters and reduce crop profitability.

There is limited research to explain the strong correlation of Fe and Al within corn tissue. Though there was a positive correlation between Fe and Al in our study, it was not likely due to synergistic nutrient uptake. Lidon et al., (2000) found no increase in Fe uptake with increased Al availability. Low soil pH increases the availability of Fe and Al in soils, while soil pH > 5.5 reduces it, therefore reducing plant uptake and the possibility of toxicity. The SR for Fe is 40 to 250 mg kg-1 at the <V6 stage and 30 to 250 mg kg-1 at all other growth stages (Campbell and Plank, 2000). There is no established SR or toxicity level for tissue Al in corn, as genetic variability and physiological factors of Al accumulation and their effects on plant growth outcomes are not predictable. At the <V6 growth stage, the highest percentage of samples, 45 and 15% for Al and Fe, respectively, had tissue concentrations in excess of 250 mg kg^{-1}. However, of samples taken at <V6 growth stage from soils with pH > 5.5, 47% had Fe and 5% had Al tissue concentrations > 250 mg kg-1 with concentrations ranging from 252 to 524 mg kg-1 Fe and 274 to 1100 mg kg-1

Al. In contrast, there were no samples at the >V7 and VT-R1 stages with concentrations > 250 mg kg-1 and only one sample at >R6, which was where the soil pH was 6.1.

One possible explanation for the high levels of Al and Fe in some tissue samples is soil contamination. If a rainfall event occurred prior to sampling it is likely that soil contamination was present, especially at the <V6 stage when a corn plant is small and the whole plant is sampled; the highest number of samples with high concentrations of Fe and Al were from this earliest growth stage. Campbell and Plank (1998) stated that Fe, Al, and Si are the elements that can most often affect tissue samples due to soil and dust particles, and mainly in young plants and grasses. Jones (1963) showed a distinct reduction in Fe levels in corn tissue samples after washing.

Evaluation of soil parameters

With the exceptions of w/v, BS, and pH, there was a high degree of variability (CV = 30 to 140%) for many soil parameters within years and growth stages (Tables 1-7 and 1-8). Soil w/v, BS, and pH had CVs < 20%, which indicates a lower deviation from the mean and shows that these variables did not differ greatly among fields compared to those soil parameters with much higher CVs. Regardless the variability, no differences in HM, w/v, CEC, BS, and Ac between years or among growth stages were detected (Table 1-7). No differences in the levels of macro- and micronutrients between years were detected (Table 1-8). With the exception of soil K, no differences among growth stages were detected (Table 1-8). However, the power to detect differences was likely hindered by the high variability in the data. Soil K tended to be the least variable among the soil nutrients and was higher at the <V6 stages than at the VT-R1 and > R6 growth stages (Table 1-8).

Table 1-7 . Main effects of year and growth stage on LS mean and CV (%) of soil: humic matter (HM[a]), sieved weight per volume (w/v), cation exchange capacity (CEC), base saturation (BS), active acidity (Ac), and pH, in samples taken at four growth stages in 2017 and 2018 from NC Corn Yield Contest study entrants. There were no year X growth-stage interactions (p ≥ 0.08 to 0.97).

	Soil parameter											
	HM		w/v		CEC		BS		Ac		pH	
Factor	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Year	(%)		(g cm^{-3})		(cmol kg^{-1})		(%)		(cmol kg^{-1})			
2017	2.0a[b]	127	1.1a	16	10.0a	74	77.6a	12	2.1a	87	5.6a	10
2018	1.3a	130	1.1a	11	8.5a	43	76.6a	13	1.8a	65	5.7a	10
Growth Stage												
<V6	1.9a	140	1.1a	15	10.0a	68	78.5a	12	1.9a	92	5.6a	11
>V7	1.7a	135	1.1a	15	9.0a	65	77.4a	12	2.0a	87	5.6a	9
VT-R1	1.8a	118	1.1a	12	9.3a	62	75.9a	13	2.0a	66	5.6a	10
>R6	1.5a	136	1.1a	14	9.3a	71	77.1a	13	1.9a	82	5.8a	10

[a]Humic matter as determined by the NCDA&CS method (Hardy et al., 2014) is strongly correlated with soil organic matter (SOM) (Weber and Peter, 1982; Blumhorst et al., 1990; Gonese and Weber, 1998). SOM (%) ≈ 1.3 HM + 0.9.

[b]Within columns within a factor, means followed by the same letter are not significantly different according to LSD (0.05).

Table 1-8 . Main effects of year and growth stage on the LS mean and CV (%) for soil nutrients in samples taken (n=149) at four growth stages in 2017 and 2018 from NC Corn Yield Contest study entrants. There were no year X growth stage interactions (p ≥ .70 to .92).

	Soil nutrient															
	P (mg kg^{-1})		K (mg kg^{-1})		Ca (mg kg^{-1})		Mg (mg kg^{-1})		S (mg kg^{-1})		Mn (mg kg^{-1})		Zn (mg kg^{-1})		Cu (mg kg^{-1})	
Factor	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
Year																
2017	273a[a]	72	226a	54	1112a	86	163a	79	38.1a	57	37.2a	132	11.8a	80	3.5a	101
2018	239a	71	196a	40	943a	53	146a	42	30.7a	42	37.2a	95	8.4a	75	2.3a	95
Growth Stage																
<V6	273a	72	226a	45	1112a	81	163a	67	38.1a	54	37.2a	95	11.8a	76	3.5a	98
>V7	266a	72	167ab	40	1040a	74	160a	70	36.3a	51	53.2a	126	10.8a	80	3.7a	101
VT-R1	298a	69	153b	54	1371a	82	204a	70	28.1a	42	28.8a	113	12.3a	86	4.3a	99
>R6	267a	75	160b	55	1143a	80	175a	76	26.9a	54	58.7a	136	12.0a	83	4.1a	106

[a]Within columns within a factor, means followed by the same letter are not significantly different according to LSD (0.5).

It is likely that the decrease in soil K was due to leaching in sandy soils such as those found in the Coastal Plain (Sharpley, 1990). Potassium in the soil is present in four pools: soil solution K that is immediately available for plant uptake; exchangeable K that is readily available for release into the soil solution through ion exchange; slow and non-exchangeable K that is fixed and slowly available; and structural K that is part of the clay lattice structure and very slowly available through weathering processes. Soil conditions greatly influence K cycling between the exchangeable and non-exchangeable fractions. It has even been suggested that current methods of estimating soil K availability are unreliable considering the soil dynamics of clay mineralogy, climactic conditions, crop species, and genetic variability in root distribution (Römheld and Kirkby, 2010).

Most fields were of the mineral soil class (MIN), however two fields in 2017 and one field in 2018 were classified as organic (ORG) (not shown). Soil mineral class, as reported by the NCDA&CS laboratory, is based on humic matter content and sieved weight/volume (w/v) (D. Hardy & Tucker, 2013). In NC, the recommended pH for corn is 6.0 for mineral soils and slightly lower in organic soils, however 85% of the mineral soil field samples in our study were below pH 6.0, and the mean of the organic soils was pH 5.0.

With respect to temporal changes in manageable soil parameters such as nutrient concentrations, pH, and Ac, the primary concern is whether these change within individual fields such that sampling time is an important factor to consider. When looking at the change in levels over the season within individual fields, more than 50% of fields had soil P, Ca, S, Mn, Zn, and Cu levels increase from the <V6 to the >V7 growth stages (Figures 1-7 and 1-8). More than 50% of fields had soil nutrient levels for S, Mn, and Zn decrease between the >V7 and VT-R1 growth

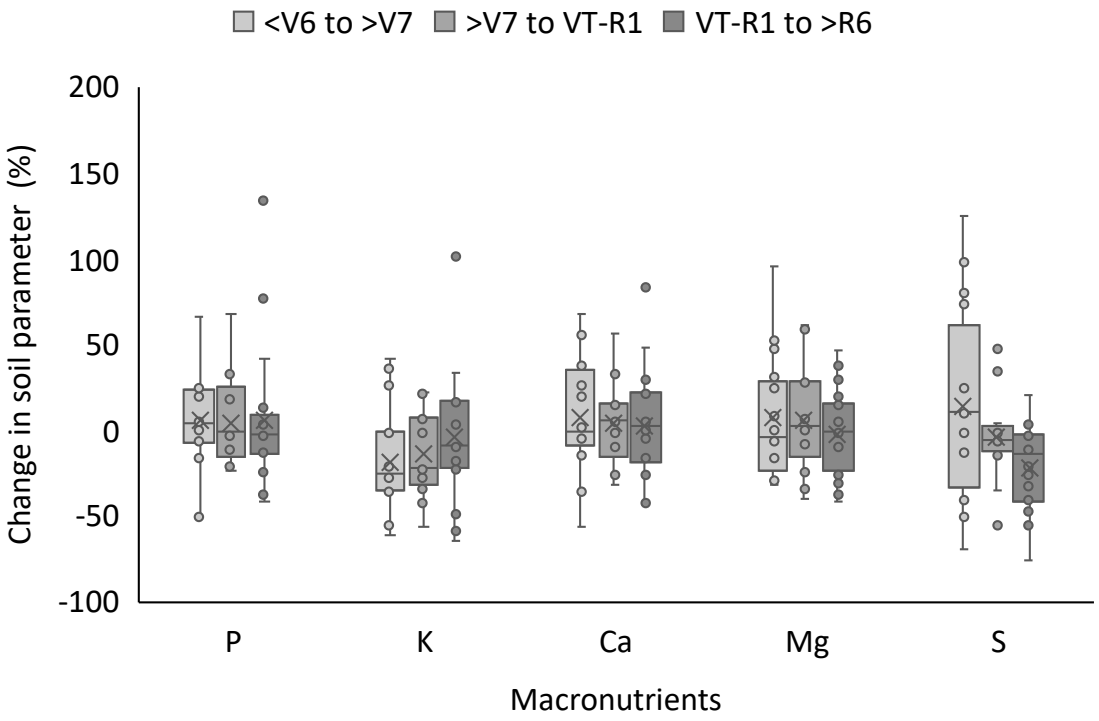

Figure 1-7. Distribution of percent change in soil macronutrients between growth stages in individual fields for samples taken at four growth stages (< V6, > V7. VT-R1, and > R6) from NC Corn Yield Contest study fields combined over 2017 and 2018. The box and whisker plots show the extreme upper and lower values excluding outliers and a box representing the inner 50% of the values. The “X” in the box is the mean and the mid-line the median; the upper and lower parts of the box are 25% of the values above and below the median.

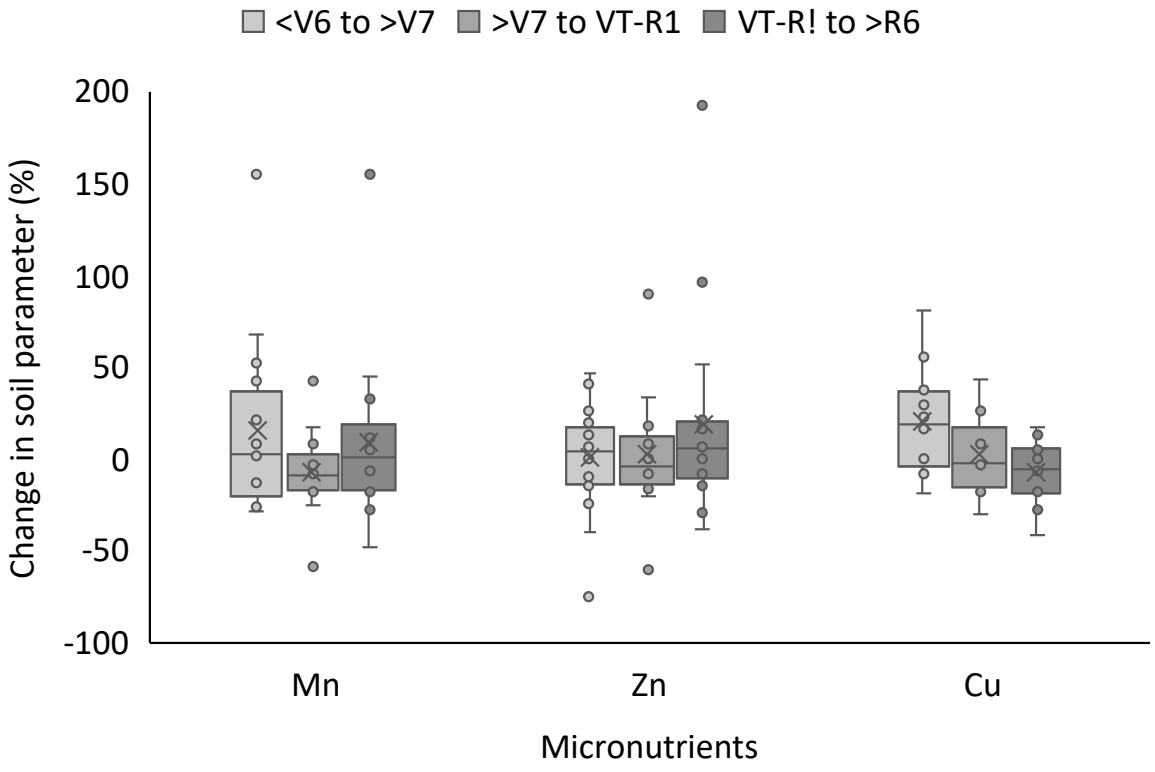

Figure 1-8. Distribution of percent change in soil micronutrient concentrations within individual fields between four growth stages (< V6, > V7. VT-R1, and > R6) for samples taken from NC Corn Yield Contest study fields in 2017 and 2018. The box and whisker plots show the extreme upper and lower values excluding outliers and a box representing the inner 50% of the values. The "X" in the box is the mean and the mid-line the median; the upper and lower parts of the box are 25% of the values above and below the median.

stages. The level of K decreased in most fields between all growth stages, and some fields decreased by > 25% below the previous level. Between <V6 and >V7, soil S and Mn increased >40% in > 25% of fields (Figures 1-7 and 1-8). In addition to the soil nutrients, the levels for soil parameters for HM, CEC, and Ac increased in most fields between each of the three growth stages, with exception for HM which decreased between the VT-R1 and >V6 growth stages (Figure 1-9).

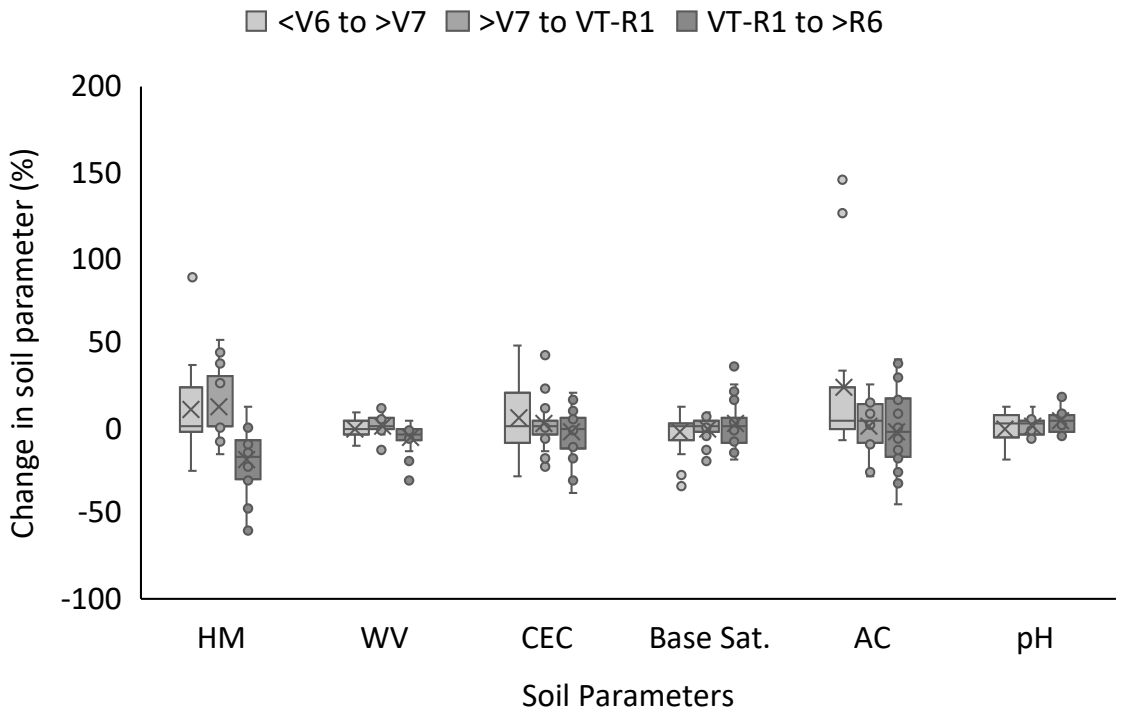

Figure 1-9. Distribution of percent change in soil parameters: percent humic matter (HM), weight/volume ratio (w/v), cation exchange capacity (CEC), base saturation (BS), exchangeable acidity (AC), and pH within individual fields between four growth stages (< V6, > V7. VT-R1, and > R6) for samples taken from NC Corn Yield Contest study fields in 2017 and 2018. The box and whisker plots show the extreme upper and lower values excluding outliers and a box representing the inner 50% of the values. The "X" in the box is the mean and the mid-line the median; the upper and lower parts of the box are 25% of the values above and below the median.

The high degree of variability by which fields changed between growth stages along with the magnitude and direction of the changes indicates how dynamic the fields were within a growing season. It is likely that the substantial changes in some of the soil parameters throughout the season were influenced by environmental conditions. Large changes in soil moisture along with early season nutrient applications likely contributed to the high percentage of change in between the first three growth stages. However, we had at our disposal information only for total nutrients added and whether started fertilizer had been applied, but not the specific breakdown of application rate and timing.

There were few correlations of yield with soil nutrients in either year (Table 1-9). Soil K was weakly and positively correlated with yield in both years, and Mn had a weak negative correlation with yield in 2018 (Table 1-9). Correlations among soil nutrients were generally similar in 2017 and 2018. Phosphorus had a moderate positive correlation with Zn in both years and a strong positive correlation with Cu in 2017 but a weak positive correlation in 2018. In both years, soil K had a weak positive correlation with S and a moderate positive correlation with Mn; K had a weak positive correlation to Ca in 2017. Soil S was also weakly correlated with Ca and Mg in both years. The correlation between Zn and Cu was very strong in 2017 and moderate in 2018.

There were several correlations among fertilizer application rates and soil nutrients. These correlations were mostly weak to moderate and either positive or negative (Table 1-9). In 2018, however, there were two exceptions: a strong positive correlation between fertilizer N rate and soil Mn, and a strong negative correlation between fertilizer P rate and soil Cu. In 2017, there was a moderate correlation between fertilizer N rate and soil Ca. In 2018, there were more soil nutrients with moderate correlations to fertilizer application rates: the fertilizer N rate was negatively correlated to soil Cu; the fertilizer P rate was positively correlated with soil Mn; the fertilizer K rate was negatively correlated with soil P, Zn, and Cu; and fertilizer S rate was positively correlated to soil Ca, Mg, and Mn. In addition, there were correlations among the fertilizer rates themselves (Table 1-9). In 2017, the fertilizer K rate had moderate positive correlations with fertilizer N (R^2 = .34) and P (R^2 = .47) rates. In 2018, fertilizer N rate had strong positive correlations with fertilizer P (R^2 = .64) and K (R^2 = .67) rates, and fertilizer P rate had a moderate positive relationship with fertilizer K (R^2 = .35) rate.

Table 1-9. Correlation matrix with R^2 values for correlations among yield, soil macro- and micronutrients, and fertilizer applications of N, P, K, and S for 2017 (upper right with yellow shading) and 2018 (lower left with green shading). (+) and (-) indicate direction of the relationship. Bolded values indicate moderate to strong correlations.

	Yield	P mg/kg	K mg/kg	Ca mg/kg	Mg mg/kg	S mg/kg	Mn mg/kg	Zn mg/kg	Cu mg/kg	Applied N	Applied P	Applied K	Applied S
Yield			.05(+)								.20(+)	.09(+)	.05(+)
P mg/kg								**.50(+)**	**.75(+)**	**.07(-)**	.17(-)	.13(-)	.07(+)
K mg/kg	.16(+)			.19(+)		.08(+)	.08(+)			.11(-)		.06(-)	
Ca mg/kg						.18(+)				**.25(-)**		.07(-)	.07(+)
Mg mg/kg						.17(+)				.11(-)	.08(+)		.09(+)
S mg/kg			.14(+)	.14(+)	.21(+)						.08(+)	.07(+)	
Mn mg/kg	.16(-)		**.28(+)**										
Zn mg/kg		**.42(+)**							**.80(+)**	.07(-)	.07(-)		
Cu mg/kg		.17(+)						**.46(+)**		.09(-)	.10(-)	.09(-)	
Applied N		**.19(-)**					**.50(+)**		**.27(-)**			**.34(+)**	
Applied P		.21(-)		0.18(-)			**.28(+)**		**.52(-)**	**.64(+)**		**.47(+)**	.06(+)
Applied K		**.39(-)**		0.16(-)			.11(+)	**.41(-)**	**.38(-)**	**.67(+)**	**.35(+)**		
Applied S			.12(+)	**.25(+)**	**.32(+)**		**.25(+)**	.13(-)	.24(-)	.14(+)	.11(+)		

There were correlations between some of the soil acidity parameters (Table 1-10), but none were correlated with yield (data not shown). Strong positive correlations existed between HM and Ac, and between BS and pH (Table 1-10). There were strong negative correlations between CEC and w/v, as well as Ac with w/v, BS, and pH (Table 1-10). Humic matter was positively correlated with CEC (Table 1-10) and this relationship was influenced by the soil class (Figure 1-10). In organic class soils, there was a moderate correlation between HM and CEC (R^2 = .33) which was greater than in mineral soils (R^2 = .21). These relationships between HM, which is a measure of soil organic matter content, and CEC are reasonable: typically, the greater the organic matter content the higher the CEC. Exchangeable acidity had a strong positive correlation with HM, a weak positive correlation with CEC, and moderate negative correlations with base saturation and pH.

Unsurprisingly, BS and CEC were both positively correlated with soil K, Ca, and Mg. However, the correlations of Ca and Mg with CEC were strong compared to weak correlations with BS, and both BS and CEC were weakly correlated with K. These correlations were to be expected, as CEC is a measure of the total negative exchange sites in the soil that attract cations such as Ca2+, Mg2+, and K+; and the percentage of cations occupying the CEC is known as the base saturation. There were also weak positive correlations of CEC with soil P, S, and Zn (Table 1-10). Among the soil nutrients, Ca, Mg, and S were moderately positively correlated with HM%; Ca and Mg were moderately negatively correlated with w/v, while K and S were weakly negatively correlated; K, Ca, Mg, and Mn were weakly positively correlated with pH; and Mn was very weakly negatively correlated with HM%, exchangeable acidity, and pH (Table 1-10). These relationships between HM and soil nutrients are expected since increased organic matter improves nutrient

cycling due to the positive relationship between OM and the soil microbial community. The complex dynamics between soil acidity, organic matter content, and soil texture influence the plant available nutrients. Soil with higher clay content and/or organic matter will have a higher CEC and have more labile nutrients to support agronomic crops.

Table 1-10 . Correlation matrix with R2 values for soil parameters: humic matter (HM), weight to volume ratio (w/v), cation exchange capacity (CEC), base saturation (BS), exchangeable acidity (Ac), and pH (all determined by NCDA&CS methods [Hardy et al., 2014]) from soil samples collected at four growth stages (<V6, >V6, VT-R1, and >R6) from fields entered into the NC Corn Yield Contest study in 2017 and 2018. Bolded values indicate a moderate or strong correlation between the parameters.

	Soil parameter					
Soil parameter	HM (%)	w/v (g cm^{-3})	CEC (cmol kg^{-1})	BS (%)	Ac (cmol kg^{-1})	pH
HM (%)	——					
w/v (g cm^{-3})	**.49(-)**[a]	——				
CEC (cmol kg^{-1})	**.48(+)**	**.44(-)**	——			
BS (%)			.22(+)	——		
Ac (cmol kg^{-1})	**.72(+)**	**.29(-)**	.14(+)	**.31(-)**	——	
pH	.09(-)		.10(+)	**.76(+)**	**.45(-)**	**——**
Soil P			.05(+)			
Soil K		.06(-)	.14(+)	.11(+)		.10(+)
Soil Ca	**.48(+)**	**.41(-)**	**.78(+)**	.15(+)	.14(+)	.06(+)
Soil Mg	**.48(+)**	**.54(-)**	**.68(+)**	.14(+)	.16(+)	.05(+)
Soil S	**.29(+)**	.19(-)	.18(+)		.12(+)	
Soil Mn	.04(-)				.04(-)	.04(+)
Soil Zn			.03(+)			
Soil Cu						

[a]Values are R^2 with (+) or (-) indicating direction of relationship and bold indicating significant correlations ($p < 0.05$).

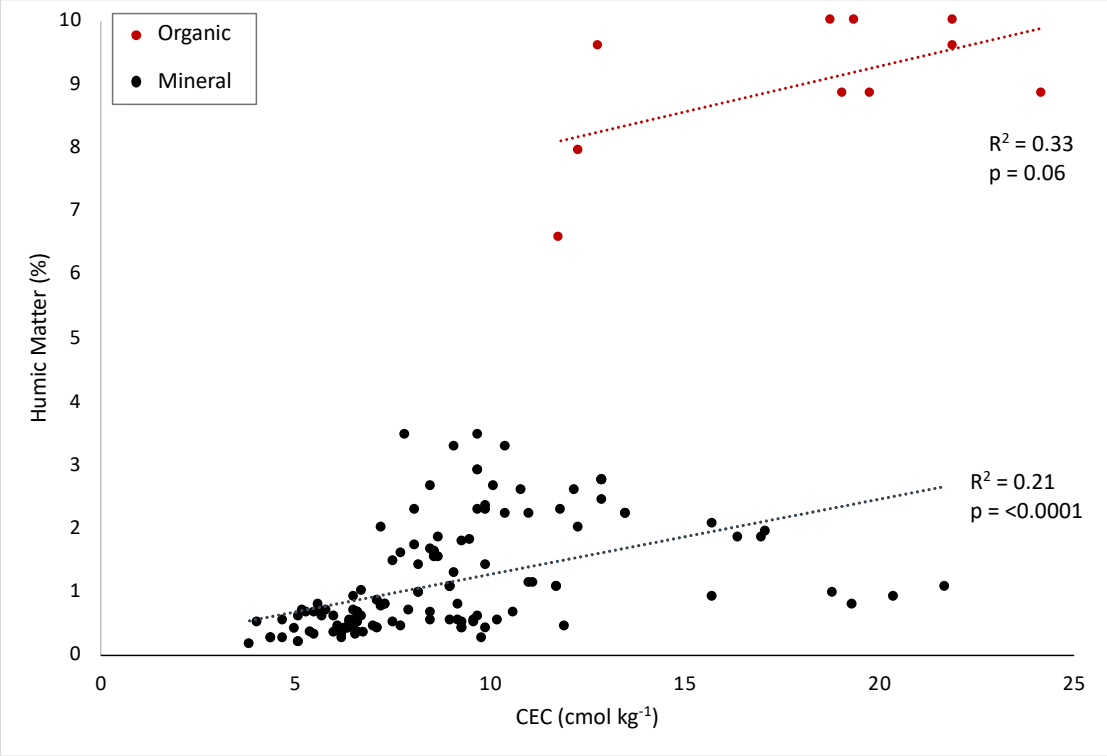

Figure 1-10 . . Relationship of soil humic matter (%) and CEC colored by soil class for 2017 and 2018 NC Corn Yield Contest study fields at all four growth stages (<V6, >V6, VT-R1, and >R6). Humic matter, CEC, and soil class as determined by NCDA&CS methods (Hardy et al., 2014).

Soil Index Values

Created to facilitate comparisons among nutrients, the NCDA&CS soil nutrient index system converts most macro- and micronutrient soil-test levels to a relative scale of 0 to 100+ based on the likelihood of a crop responding to fertilization (Hatfield, 1972). The index system was developed to allow growers a method of simple interpretation for soil test results and was based on correlations between soil nutrient test values and crop yield response curves. Indices are used for P, K, S, Mn, Zn, and Cu. Index values ≤ 10 are considered "very low" and would have a very high crop response to fertilizer application. Values between 11 and 25 are considered "low" and would be expected to have a high fertilizer response. For P, K, and S, index values

between 26 and 50 (“medium”) are expected to respond moderately, though the response would decrease as soil test values increase. For Mn, Zn, and Cu, indices ≥ 26 are considered unlikely to respond to additional fertilizer and indices > 50 for P and K are unlikely to respond. Indices above 100 are considered “very high”.

The nutrient statuses for P, K, Mn, Zn, Cu, and S of all soil samples were at or above the medium fertility level (>25) whereby the expectation of a crop yield response to fertilizer would be low to none (Figure 1-11A and B). A low percentage (< 10%) of soil samples were within the index range of 26 to 50 (medium) for all nutrients except for S, where 25 and 35% of samples were within this range in 2017 and 2018, respectively (Figure 1-11A and B). Thus, for all nutrients except S, the indices of > 90% of samples were > 51 (Figure 1-11B). Further, a high percentage of samples were in the excess (>100) index range for P (88%), Mn (65%), and Zn (77%) in both years, and for S (60%) in 2017. Otherwise, according to the soil index levels, there was ample soil P, K, S, Mn, Zn, and Cu available in most fields. Although about 25% of fields had a soil index < 50 for Cu, nutrient application would not be advised as no yield response would be expected. In the case of S, plow-layer soil test index values may not be an accurate assessment of availability in Coastal Plain sandy soils, which made up the bulk of the fields in our study. Due to leaching in sandy soils, subsoil sampling, which we did not perform, would be recommended to obtain an accurate assessment of plant-available soil S (David H Hardy, Tucker, & Stokes, 2014).

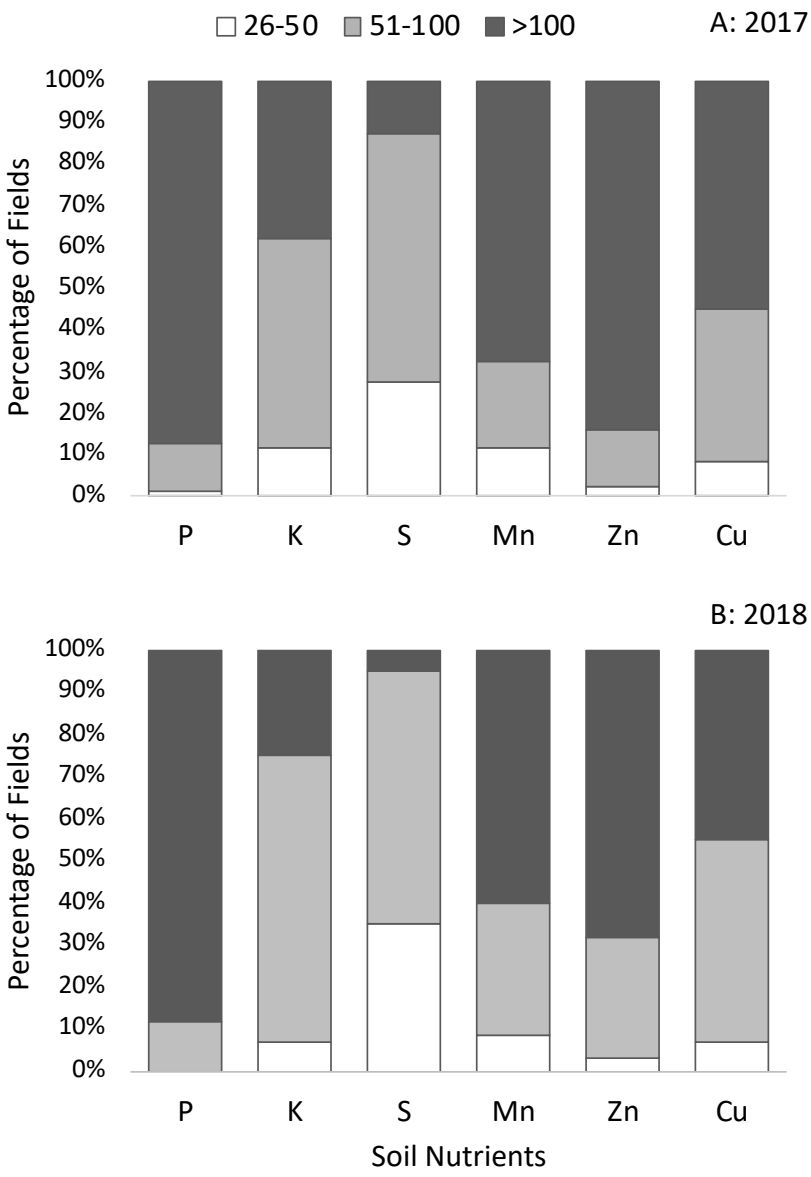

Figure 1-11. Percentage of fields with soil P, K, S, Mn, Zn, and Cu indices in the medium (26-50), high (51-100), and very high (>100) ranges for samples taken from NC Corn Yield Contest study fields in 2017 (A) and 2018 (B). None of the samples in our study had soil index levels below 26.

In NC, availability of soil Mg is assessed in relation to soil CEC. Soils are considered deficient and likely to respond to Mg applications when: 1) the Mg saturation of the CEC is <10% and this percentage times the CEC is < 0.5 $cmol_c\ kg^{-1}$, or 2) if the latter is ≤ 0.25 $cmol_c\ kg^{-1}$. In cases of deficiency, an application of 22 to 37 kg ha^{-1} Mg would be recommended. Only 4% of soil samples comprising three fields were deficient in 2017 and none in 2018 (Figures 1-12A and B), indicating that no yield response would be expected from Mg application in nearly all fields. Since Ca is a component of lime and likely present in high enough quantities within the soil, no a recommendation is provided on a soil report from the NCDA&CS. Considering the soil test results, growers would not be advised to apply Mg other than to replace nutrients removed from the crop.

Relationships between soil and tissue nutrients

There were few correlations between plant tissue and soil macronutrients beyond correlations of like nutrients in both soil and tissue (Table 1-11). There were positive weak correlations between tissue P and soil P, tissue S and soil S, tissue K and soil K, and tissue Mg and soil Mg (Table 1-11). There were a few correlations that varied between years. The correlations between tissue and soil K and between tissue and soil S were weaker in 2017 (R^2 = 0.16 and 0.07, respectively) than in 2018 (R^2 = 0.35 and 0.21, respectively) (data not shown). Conversely, the correlation of tissue Mg and soil Mg was stronger in 2017 (R^2 = 0.42) than in 2018 (R^2 = 0.25) (data not shown). The tissue and soil Mg correlation was the only one that was highly influenced

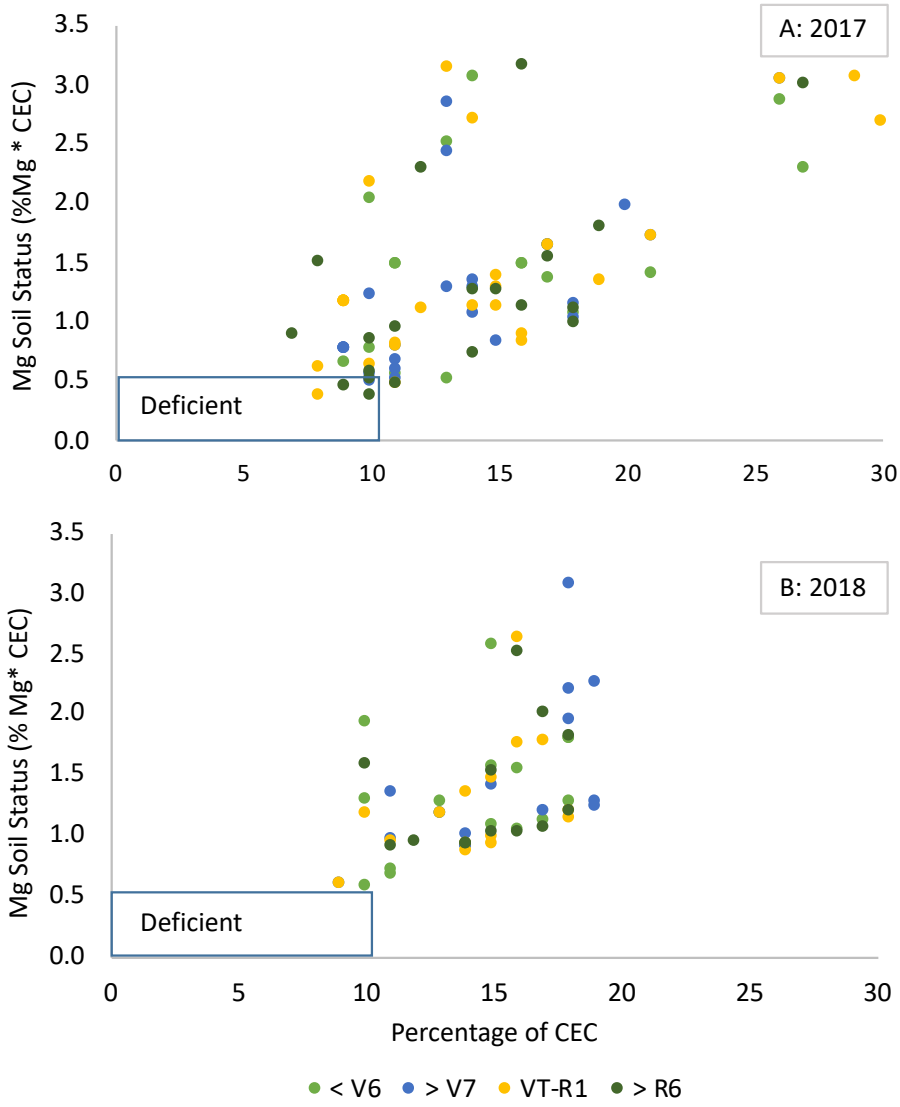

Figure 1-12. Status of soil Mg expressed as its cation exchange charge versus the proportion of the CEC occupied by Mg for soil samples taken at four growth stages stages (< V6, > V7. VT-R1, and > R6) from NC Corn Yield Contest study fields in 2017 (A) and 2018 (B). The interior of the blue box indicates deficiency according to the NCDA&CS soil test (Hardy et al, 2014).

by the growth stage (Figure 1-14), with the stronger correlations occurring at the >V7 stage (R2 = .58 and .28, in 2017 and 2018, respectively); at the VT-R1 stage (R2 = .62 and .32); and at the >R6 stage (R2 = .59) in 2017. At the <V6 stage, there was no correlation between tissue Mg and soil Mg in either year. The strengths of these relationships at the >V7 and VT-R1 growth stages suggested that there was a connection between the concentration of the Mg in tissue with the

content of Mg in the soil. There was a very weak negative correlation between tissue Ca and soil S. Tissue Mg had a weak positive correlation with soil Ca and a weak negative correlation with soil K. Tissue S had a moderate positive correlation with soil K (Table 1-11). There were few tissue macronutrients with correlations to soil parameters related to w/v and acidity. The most notable exceptions were that tissue Mg was weakly correlated with all non-nutrient soil parameters. There were positive correlations of tissue Mg with HM (%), CEC, BS, active acidity, and pH; and a weak negative correlation with w/v. In addition, there were very weak positive correlations between tissue P and w/v; tissue Ca with BS and pH; and tissue S with BS (Table 1-11).

Table 1-11. Correlations of plant tissue macronutrients as a function of soil parameters. Tissue and soil samples were collected at four growth stages (<V6, >V6, VT-R1, and >R6) in the 2017 and 2018 growing season from fields of NC Corn Yield Contest study entrants.

	Tissue macronutrient					
Soil parameter	N	P	K	Ca	Mg	S
P		.09(+)[a]		.03(+)		
K	.13(+)	.05(+)	.22(+)		.06(-)	.24(+)
Ca					.11(+)	
Mg					**.29(+)**	
S	.09(+)		.05(+)	.04(-)		.11(+)
Mn				.04(+)	.06(-)	
Zn		.08(+)	.04(+)	.06(+)		
Cu		0.06(+)		.05(+)		
HM					.09(+)	
w/v		.04(+)			.08(-)	
CEC					.14(+)	
BS				.03(+)	.05(+)	.03(+)
Ac					.03(+)	
pH				.05(+)	.04(+)	

[a] Values are R^2 with (+) or (-) indicating direction of relationship and bold indicating significant relationships ($p < 0.05$).

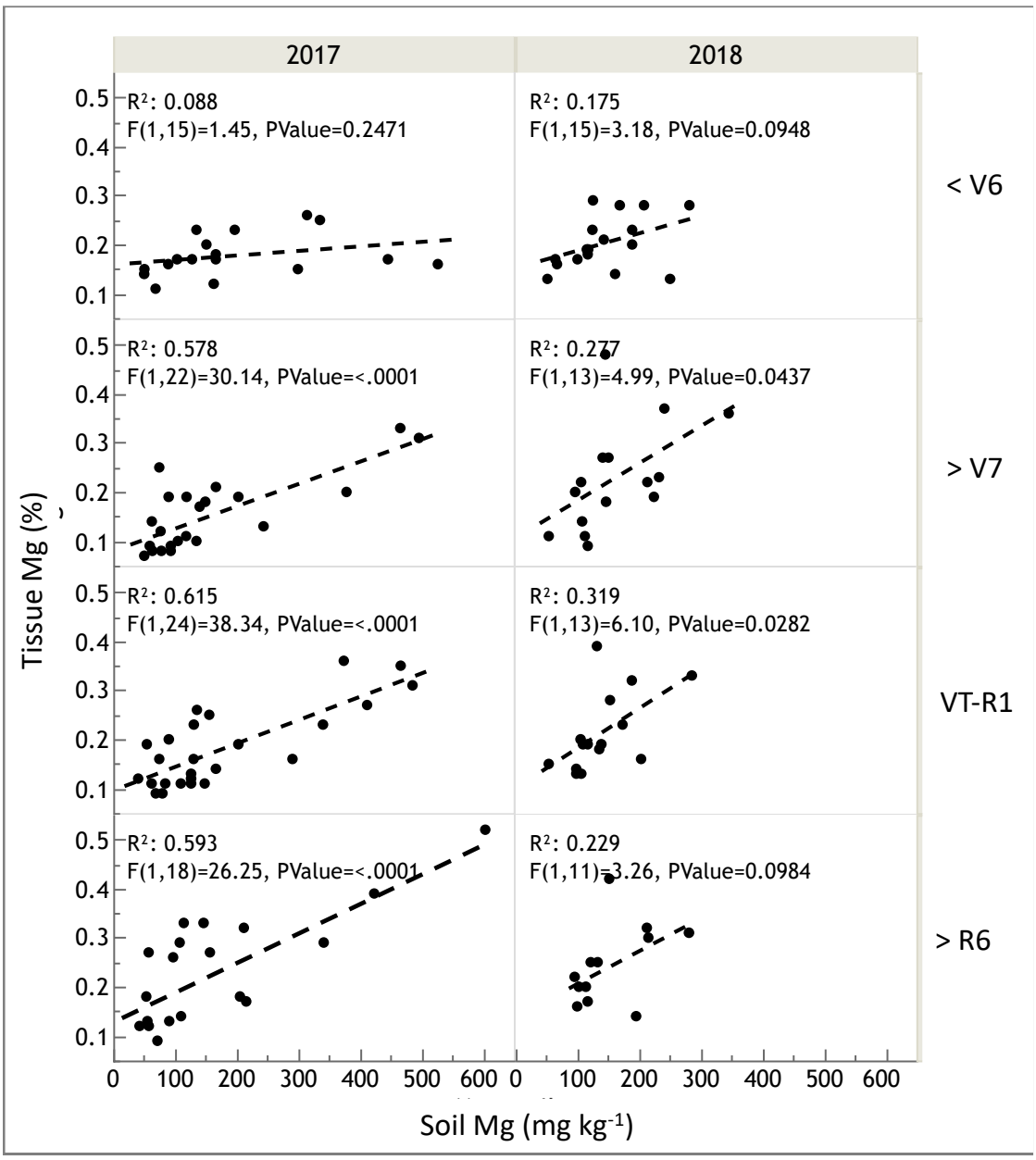

Figure 1-13. Linear regression of tissue Mg concentration (%) as a function of soil Mg (mg kg^{-1}) for samples taken at four growth stages (< V6, > V7. VT-R1, and > R6) from NC Corn Yield Contest study fields in 2017 and 2018. Correlation probability $p \leq 0.05$ considered significant.

It is likely that environmental influences account for the high degree of variability among soil nutrient availability between fields and temporally within fields. This was likely a key factor in the weak relationships between like nutrients of tissue and soil when analyzed as a whole dataset or

when only broken out by year. This points to how much the effect of soil and climatic conditions can influence nutrient availability.

Conclusion

The growers in our study achieved higher average yields than the statewide averages in 2017 and 2018, (15.1 and 13.8 Mg ha-1, respectively) which equates to a yield gap of 6.2 to 6.7 Mg ha-1 between the average NC corn yield and the high-yielding populations in our study. The growers used a wide range of management practices which were generally within recommended guideline for obtaining high yields. Though management practices were no strongly correlated with yield in either year, this was likely due to the strong influence of environmental conditions between the years rather than a lack of benefit from following recommended growing practices. There are, however, recommendations that warrant further investigation. Changes in nutrient accumulation due in part to newer hybrids and management practices may necessitate a revision of SR to improve nutrient balance and achieve higher yields. It is possible that SR for some nutrient like N and P may need to be revised downward, and research has shown that newer hybrids have lower concentrations while still achieving high yields.

While some nutrients were likely not limiting as indicated by the tissue sample results, others like tissue Mg were may be needed in higher amounts by the plant. Tissue Mg was likely a limiting nutrient, especially in 2017 when approximately 60% of fields had tissue samples with Mg concentrations below the SR at the first three growth stages. Though few fields were determined to be deficient in soil Mg, many were near the lower levels for Mg and may benefit from Mg application to the soil. In addition, there were high rates of applied K among some

growers and many soils had high K levels, contributing to a high K/Mg ratio in the soil solution which can lead to reduced Mg uptake from an antagonistic relationship between K and Mg.

The availability and uptake of nutrients is governed by the biogeochemistry of the soil. There were other tissue nutrients that were below desired SR in >50% of the fields; but, with exception for S, the nutrient status of the soil for >90% of fields did not indicate a need for nutrient application due to a deficiency. One explanation for this may be a lower than ideal pH; < 85% of the mineral soil field samples in our study were below pH 6 which could reduce the availability of nutrients in the soil solution. Other influences to plant nutrient uptake are root morphology and environmental conditions which are in constant flux throughout the growing season. Our study only focused on what could be elucidated from soil and tissue test results across the growing season coupled with management practices. Soil texture and moisture, weather conditions, and other soil measurements such as hydraulic conductivity, and particle density were not determined. These factors are expected to vary in a complex fashion depending on depth and time of sampling. Indeed, more studies similar to ours involving multiple growing conditions with inclusion of environmental variables could lead to a better understanding of the complex dynamics between plant and soil but was beyond the scope of this study.

CHAPTER 2

Sufficiency Ranges, Diagnosis Recommendation and Integrated System, or Machine Learning: Comparing Systems for Nutrient Recommendation in North Carolina

Abstract

Currently ranked 18th in grain corn (*Zea mays* L.) production, NC had 340,200 ha planted to corn in 2017. Despite harvesting two and a half times more corn, 3.03 Tg, than the next largest southeastern producer, NC has one of the lowest average yields: 8.9 Mg ha^{-1}. One traditional approach for assessing crop nutrient status and associated management decisions has been to use tissue and soil test results to determine crop macro- and micronutrient adequacy. Our objective was to compare three techniques for analyzing tissue and soil nutrient data: two multivariate and one machine learning approach. The first was the diagnosis and recommendation integrated system (DRIS), a method for finding the most limiting nutrient based on tissue concentrations. The second was the least absolute shrinkage and selection operator (LASSO), a generalized regression method that performs variable selection and regularization. The third was ML using support vector machines, artificial neural networks, and decision tree algorithms that perform supervised classification. The goal was to assess which system better determined nutrients most correlated with yield. Across the 2017 and 2018 growing seasons, we evaluated the nutrient status of soil and tissue samples from high-yielding NC growers. Linear regression was performed relating yield to tissue nutrient concentrations and available soil nutrients at four growth stages. The correlations between yield and individual tissue and soil nutrients were mostly weak except for a moderate negative correlation with tissue Mg (R^2 = .28) in 2018. Using DRIS, the seasonal mean of the nutrient balance index for macronutrients had a

slightly higher correlation to yield (R^2 = .32), but only in 2017. Conversely, the LASSO regression model resulted in a higher correlation to yield (R^2 = .48) using both tissue and soil macronutrients taken at tasseling. The addition of soil parameters to LASSO or ML analyses greatly improved model performance. The addition of management practices greatly improved the performance of the LASSO and ML models, indicating that relative to the soil and tissue parameters tested, management practices had the greatest influence on yield models in our study.

Introduction

Demand for grain corn is driven by population growth, the increased demand for biofuels, and demand for feed grain. North Carolina has a large animal production industry valued close to US$6 billion which is supplied by corn produced largely outside of NC (USDA, NASS, 2018). North Carolina is currently ranked 18th in grain corn production nationwide owing to a high amount of production acreage (USDA, 2019); despite the high production, NC has one of lowest yields per hectare (8.9 Mg ha^{-1}) of the southeastern states (Table 2-1). Without additional land area to expand production, growers are faced with increasing the productivity of the current land inventory. However, intensive agricultural practices can have detrimental consequences to the health of the soil and environment.

Agriculture is often blamed for the excess nutrients that leach into groundwater or enter surface waters, potentially contaminating the water supply. There is a desire to increase sustainability within the agricultural industry; but for this to succeed, an understanding of the complex nutrient dynamics is needed to create nutrient management plans that foster crop productivity and avoid concurrent nutrient losses. Additionally, a greater understanding of factors limiting yield and of management practices that improve profitability without increasing

the associated environmental costs of production will further increase sustainability. Sustainable nutrient management plans should aim to achieve harvest goals, improve soil health and water and air quality, and return a profit.

Table 2-1. Southeastern US corn yield data: area harvested, yield per ha, and total production by state for 2017 and 2018.

	Hectares harvested		Corn yield (Mg ha^{-1})		Total production (Mg)	
			Year			
State	2017	2018	2017	2018	2017	2018
	———1,000s———		———1,000s———		———1,000s———	
Alabama	95	101	10.5	9.8	997	991
Georgia	99	115	11.0	11.0	1,095	1,274
N. Carolina	340	336	8.9	7.1	3,030	2,383
S. Carolina	132	125	8.5	8.0	1,123	1,000
Virginia	138	132	8.5	9.2	1,209	1,207

Currently, soil and tissue test results are the main method used for guiding nutrient management decisions in NC. Soil and tissue analysis are important tools for understanding the nutrient availability and uptake within the cropping system. Soil test results are used to guide the nutrient recommendations for the following season and tissue testing is used to assess the nutrient status within the growing season. Tissue testing is often used for diagnostics prior to applying corrective measures to address within-season nutrient deficiencies. There are optimum corn tissue nutrient concentrations that are typically assessed at multiple growth stages for optimum crop productivity at key points in the lifecycle (Campbell, 2000). For optimum growth, ample nutrients need to be available to the plant at every stage; deficiencies that are not addressed can result in decreased yields (Hanway, 1966). Diagnosing deficiencies in a timely manner is vital for optimum outcomes. Nutrient uptake, however, can also be affected by climatic, biotic, and geochemical processes that are often difficult to interpret and address with

a nutrient management plan (Binkley & Vitousek, 1989; Hussain et al., 2019; Marschner & Dell, 1994).

The tissue analysis approach utilizes nutrient critical values (CRV) and sufficiency ranges (SR) to determine the plant nutrient status (Campbell, 2000). Critical values were established to determine the point at which tissue nutrient concentrations were slightly below levels for optimum yield such that nutrient application would elicit a 5 to 10% yield response (Melsted, Motto, & Peck, 1969). The SR analysis system established benchmarks for excessive and deficient levels of nutrients based on predictable quantities of nutrient elements contained in healthy corn plants based on empirical research (Campbell, 2000). These quantities, along with critical values, are used to guide fertilizer application rates. Sufficiency ranges to diagnose nutrient problems within the southeastern US were established based on studies conducted at multiple research institutions (Campbell, 2000). One drawback to using CRV and SR is that they do not account for nutrient interactions that may influence the concentration of nutrients within the tissue (Hallmark et al., 1987).

Nutrient Balance Standards

In the past several decades, there have been several methods proposed utilizing a more balanced approach to nutrient analysis based on the dynamic nature of nutrients within plant tissue. Most formulas pair nutrients together in ratios and compare the ratios to established norms from high-yielding populations. Methods such as plant analysis and standardize scores (PASS) and compositional nutrient diagnosis (CND) differ in ease of interpretation and implementation, but require advanced statistical knowledge to compute (Baldock & Schulte, 1996; Parent & Dafir, 1992). Modesto et al. (2014) proposed a different method to establish foliar

nutrient balance standards; however, their approach did not incorporate S into the balance equation, and thus may potentially have biased the results by ignoring an important macronutrient. Additionally, their approach utilized a complex group of formulas that required a high degree of statistical knowledge to interpret nutrient deficiencies.

An older and more established nutrient balance approach developed by Beaufils (1973) is the Diagnosis and Recommendation Integrated System (DRIS), a method that utilizes the relationship between nutrient pairs among all nutrients of interest and aims to account for the more complex relationship of nutrients within the plant rather than individual nutrients. The DRIS is an evaluation method used to determine crop nutrient status by comparing ratios of the tissue concentrations of nutrient pairs with established norms from high-yielding crop varieties (Elwali, Gascho, & Sumner, 1985; Sumner, 1977; Walworth, Woodward, & Sumner, 1988). These norms were derived from a database of ca. 10,000 observations of leaf tissue data and corresponding yields collected from various states and countries, including 1500 obtained from the Tifton, GA experiment station (Elwali et al., 1985). A yield threshold of 10.0 Mg ha^{-1} was used to separate high-yielding from low-yielding populations (Elwali et al., 1985). The DRIS was among the first methods to incorporate the use of nutrient ratios to compare multiple nutrients in determining the most limiting to optimum growth. The established ratio norms for corn include the 11 macro- and micronutrients considered most important and included in tissue test reports such as the NCDA&CS tissue sample test (McGinnis et al., 2014). The DRIS has an advantage over CRV and SR due to its ability to take a more holistic view of the nutrient status in the plant, but some studies have shown that it cannot replace the SR approach but rather augment it (Soltanpour, Malakouti, & Ronaghi, 1995).

It is well known that nutrient concentrations differ within plant parts depending on the age and location of the tissue. Nitrogen concentrations are high in young leaf tissue during the vegetative stage and begin to decrease as the plant utilizes this resource for ear development, and therefore the N level is low comparatively in older tissue (Karlen, Flannery, & Sadler, 1988). Conversely, the tissue concentrations of other nutrients such as Ca and B increase as the plant ages (Karlen et al., 1988). There are SR established for four different corn growth stages, however, the broad range of each growth stage may not always allow for accurate assessment of crop deficiency. Walworth and Sumner (1987) showed that regardless of tissue age, DRIS had greater accuracy in diagnosing the most limiting nutrient compared to the sufficiency range approach. However, recent research has shown that the nutrient use efficiency of newer hybrids and their accumulation of nutrients within plant organs differ from older hybrids (Bender et al., 2013; Woli, Sawyer, & Boyer, 2018). Thus, the established norms may prove less reliable with current corn hybrids and the accuracy of diagnosis may be lowered as a result.

Machine Learning Techniques

The nutrient uptake system is complex and to make sense of the nutrient dynamics requires an interpretation system that is able to work in a multi-dimensional way to bring clarity to the complexity. Machine learning has recently gained attention as a computational method for data analysis, especially in the fields of genomics and biomedicine. The method is based on statistics and utilizes computer software to find hidden patterns in datasets which would otherwise be missed. Additionally, machine learning is useful when dealing with large datasets, or when the number of variables is large in comparison to the number of observations.

The least absolute shrinkage and selection operator (LASSO) is a regularized regression model that performs variable selection and shrinkage to enhance a model's accuracy of prediction and interpretability. The regularization refers to the shrinkage of predictor coefficients toward zero and reducing the size of the model and improving interpretation. The variable selection is useful for highly correlated inputs and helps to prevent overfitting of the model. This prevents the creation of a model with a high degree of specificity that would not be well suited to biological systems (Gábor & Banga, 2015). Biological systems tend to be dynamic systems with a high degree of variability, and models trained under specific conditions may not fit future data well, indicating the model was overfit to the training data.

Supervised machine learning (ML) is an approach that recently has been gaining favor for agriculture applications. Among the most useful applications for ML are: analysis of remote sensing data (Maxwell, Warner, & Fang, 2018); in the agri- and biotechnology domain; and also to process and interpret large amounts of data now collected in precision agriculture settings (Liakos, Busato, Moshou, Pearson, & Bochtis, 2018). Machine learning algorithms such as support vector machines (SVM), artificial neural networks (ANN) and decisions trees (DT) have proven useful for pattern recognition within large datasets, which can then be used for data classification and outcome prediction. Traditional methods using inferential statistics are not well suited for large heterogenous and unstructured datasets. Machine learning offers computational solutions needed to gain an understanding of the dynamic changes influencing yield within complex agricultural systems. Understanding the factors in the agriculture system most likely connected to improved yield outcomes for NC growers would progress the sustainability of NC agriculture.

Many studies have indicated that yield increases are likely due, in part, to increased nutrient uptake and use efficiency. Many growers have already incorporated high-yielding corn varieties; however, multiple factors may be preventing higher yields in NC. Our objective was to compare several analytical approaches to determining the soil and tissue nutrients and management practices most associated with yield of newer high-yielding corn varieties. This approach will allow for a better understanding of the interplay between the complex soil and plant systems and help improve agronomic recommendations for sustainable crop management. We aimed to better understand how agriculture management practices and seasonal variability in nutrient demands influence productivity of high yielding field corn in NC. The specific objectives of this study were to: 1) compare and contrast sufficiency ranges and the DRIS nutrient diagnostic methods to generalized regression and machine learning techniques; 2) determine the degree of correlation of the aforementioned techniques to yield outcomes from the NC Corn Yield Contest; and 3) determine the most likely factors influencing yield based on tissue and soil analysis using DRIS and machine learning techniques.

Materials and Methods

Data Collection

In 2017 and 2018, NC corn growers with potential for high yields were solicited to take part in an observational field study. The selection of growers for participation in the Corn Yield Contest study was partly based on their past entry in the NC Corn Yield Contest. Awards in the NC Corn Yield Contest are given to high-yielding participants in several categories: six dryland divisions representing areas of NC with similar soils and climate conditions; and an irrigated division (Heiniger, 2014). The targeted solicitation of growers with past participation in the NC

Corn Yield Contest was based on the assumption that these growers have management practices that foster high yields. Additional criteria for study inclusion were that growers: 1) had an intention of entering the NC Corn Yield Contest in the study years; 2) committed to harvest and submit a Contest entry form; 3) would allow soil and tissue sampling from the Contest field; and 4) had Contest field locations in North Carolina.

A website was created using WordPress that contained the study description and details. Included on the website was a link to an online Qualtrics survey platform (Qualtrics Co., Provo, UT) used for self-enrollment. The website provided soil and tissue sampling instructions along with prepaid NCDA&CS sample submission information forms which included pre-populated information specific to the study.

In 2017, grower recruitment for participation in the study began by mailing 510 postcards to recent contest entrants. The postcard contained a link to the enrollment website and the study information as well as a QR-code link to the Qualtrics survey. Additional recruitment efforts included sending email invitations to the 13 NCDA&CS regional agronomists, 110 field agents for the NC Agriculture Extension Service, and 134 agricultural consultants of NC collected using the "Find a Professional" search tool on the TriSocieties webpage. In 2018, the solicitation was expanded to increase study enrollment. Letters were used instead of postcards. The mail campaign included 2017 NC Corn Yield Contest entrants and 2018 NC Commodity Conference attendees (149 and 470 letters, respectively).

The sampling was performed by the entrant, consultant, or extension agent assisting with the study. We offered to do the sampling, but no-one requested us to do so. Soil and tissue samples were gathered on the same day from one-acre field plots designated by the grower for

entry into the NC Corn Yield Contest. The tissue sampling technique followed the recommended NCDA&CS sampling protocol (McGinnis et al., 2014). This protocol included proper handling of the tissue samples by placing air dried or refrigerated samples into a paper bag for transport to the NCDA&CS lab to reduce the possibility of sample degradation. The tissue samples were collected as close to the following four stages as possible: <V6 (whole plant sampled); >V7 (most recent mature leaf); VT-R1 (earleaf); and >R6 (earleaf) (Hanway, 1966). The four growth stages were chosen to correspond with the tissue nutrient sufficiency ranges recommended by the NCDA&CS laboratory (McGinnis et al. 2014). Soil samples were collected from within field locations allocated for the NC Corn Yield Contest entry. Twenty plow-layer soil samples were taken within the vicinity of the tissue sampling and bulked.

The samples were submitted for analysis by the NCDA&CS laboratory. The NCDA&CS laboratory collaborated in the study and reduced the associated fees for analysis by the laboratory from research fees (soil: $5.00; tissue $12.00) to grower fees (soil: gratis; tissue $5.00). Soil and tissue samples were analyzed using established methods (Hardy et al. 2014). For tissue N, oxygen combustion gas chromatography was used for analysis; for other nutrients HNO_3 microwave digestion followed by inductively coupled plasma (ICP) spectrophotometry (McGinnis et al., 2014). The soil analysis procedure used was Mehlich 3: P, K, Ca, Mg, S, Cu, Zn, Mn, and Na (Mehlich, 1984a); cation exchange capacity (CEC) and base saturation (BS%) (Mehlich et al., 1976); pH/acidity (Ac)/lime requirement (Mehlich, 1976); soil class (mineral, mineral-organic, organic); sieved weight-to-volume (W/V); and humic matter (HM%) (D.H. Hardy, 2014; Mehlich, 1984b). The analysis results were retrieved from the NCDA&CS laboratory's secure online portal and compiled into a spreadsheet. Additional information provided from analysis results were:

address, county, planting date, and sample collection date. The tissue macronutrient concentrations were converted from percentage to g kg^{-1}, the tissue micronutrients were converted from ppm to g kg^{-1}, and the soil nutrients P, K, S, Mn, Cu, and Zn were converted to mg kg^{-1}. Soil Ca and Mg were reported as a percentage of CEC and were converted to mg kg^{-1} as well.

The total number of samples for 2017 and 2018 were 147 and 67, respectively. However, the locations did not have consistent submissions for each growth stage, region, or year, resulting in incomplete sets of soil and tissue samples per field location (Table 2-2). Additionally, not all field sampled had associated yield, as not all growers entered the NC Corn Yield Contest in the respective year. This resulted in 24 fields in 2017 and 17 fields in 2018 with associated yields (Table 2-2).

The NC Corn Yield Contest data including certified yield was received for each study year after data was compiled from contest entry forms. Data included plant population density; row spacing; nutrient application (N, P, K, S, and micronutrients); manure or starter fertilizer use; tillage practice; hybrid; previous crop; with/without irrigation; and herbicide and pesticide applications. There was variation in management practices among growers which included: conventional, conservation, and no-till soil preparation; rain-fed and irrigated water management; and whether starter fertilizer was applied. In 2018, two study locations representing 23 field samples were not entered into the contest. The management practices for these 23 field were not recorded however, the relative yields were obtained.

Table 2-2. Numbers of fields sampled divided by fields with and without associated yields, and soil and tissue samples submitted for 2017 and 2018 by region and corn growth stages: early vegetative (< V6); late vegetative (> V6); tasseling (VT-R1); and maturity (>R6) (Ritchie, Hanway, & Benson, 1986).

Year	Number of individual fields sampled					
	Region					
	Coastal Plain		Piedmont		Mountains	
	With[a]	Without	With	Without	With	Without
2017	22	13	1	13	1	0
2018	16	2	0	5	1	0
	Number of individual tissue and soil samples					
	Tissue	Soil	Tissue	Soil	Tissue	Soil
2017	118	98	26	17	3	3
2018	60	60	5	0	2	2
	Growth stage					
	Early (< V6)		Early (> V6)		Tasseling (VT-R1)		Maturity (> R6)	
	Tissue	Soil	Tissue	Soil	Tissue	Soil	Tissue	Soil
2017	23	31	53	30	38	28	33	29
2018	24	19	15	15	15	15	13	13

[a]Fields with and without associated yield from NC Corn Yield Contest entry.

Data Analysis

Statistical analyses of tissue and soil data were performed using the JMP Pro 14 statistical software package (SAS Institute, Cary, NC). The sample data were compiled and matched to the corresponding yield from the NC Corn Yield Contest when possible. Field sites with incomplete data were utilized for analysis of nutrient trends but excluded when establishing correlation with yield. One yield was extremely low and considered an outlier and was therefore removed from the observations. Linear regression analysis was used to determine collinearity among variables.

The DRIS was calculated following the protocol by Walworth and Sumner (1987). The DRIS formula follows a multi-step process involving ratios of nutrients from the study population compared to nutrient ratios from a high-yielding population. The formula involves quotients for each pair of individual nutrients of interest, X/A, in which X and A are the tissue concentrations

for nutrients X and A, respectively. The other component (*x/a*) were the norms established from high-yielding populations (Elwali et al., 1985; Sumner, 1977). The first function proceeds through either Eq. [1] or [2] depending on whether the ratio X/A of the study population was greater or less than the ratio (*x/a*) of the high-yielding population norms. The CV was the coefficient of variation associated with the norm.

When X/A > *x/a*

$$\boldsymbol{f}(\mathbf{X/A}) = [(\mathbf{X/A})/(\boldsymbol{x/a}) - \mathbf{1}]/(\mathbf{0.01}\ \boldsymbol{CV}) \quad [1]$$

And when X/A < *x/a*

$$\boldsymbol{f}(\mathbf{X/A}) = [\mathbf{1} - (\boldsymbol{x/a})/(\mathbf{X/A})]/(\mathbf{0.01}\ \boldsymbol{CV}) \quad [2]$$

After calculating the ratios for each nutrient of interest, the second step was to determine the index value for each nutrient. The DRIS index formula for nutrient X (D_x) is given in Eq. [3],

$$D_x = \left\{ \frac{\Sigma \left[f\left(\frac{X}{A}\right) + f\left(\frac{X}{B}\right) \ldots\ldots - f\left(\frac{M}{X}\right) - f\left(\frac{N}{X}\right) \ldots\ldots \right]}{n} \right\} \quad [3]$$

where X is the nutrient of interest; A,B,…M,N are other nutrients of interest that are included to determine the index of X; and *n* is the number of individual nutrients being investigated inclusive of X. A positive DRIS index indicates either sufficient or excessive nutrients, and a negative DRIS index indicates nutrient deficiency. The closer the DRIS index is to zero the greater the nutrient balance.

The final step was to determine the Nutrient Balance Index (NBI) by summing the absolute value of all individual DRIS indices. The ideal NBI value is zero. The further NBI is from zero, the greater the nutrient imbalance (Walworth & Sumner, 1987).

Generalized Regression

The generalized regression formula least absolute shrinkage and selection operator (LASSO) is a regularized regression model. A regularized regression constrains the coefficient estimates to prevent overfitting the model. The LASSO performs variable selection and shrinkage to enhance a model's accuracy of prediction and interpretability by performing subset selection to reduce model size. Shrinkage refers to the removal of a variable from the model when constrained to zero and indicating no relevance to the predictor model. The LASSO method minimizes the sum of squares to determine the unknown coefficients related to the predictor value. The LASSO was created as an extension of classic regression with the addition of constraints applied to regression coefficients by the sum of the absolute values of the coefficients. The LASSO is constrained by *t*, a parameter that determines the amount of regularization. A selection operator helps to prevent overfitting of the model through automatic variable selection which reduces the model size and is useful for highly correlated variables. The LASSO estimate in the constrained form for a coefficient $\hat{\beta}$ is in Eq. [4] and is subject to the tuning parameter in Eq. [5] (Emmert-Streib, Dehmer, Emmert-Streib, & Dehmer, 2019; Tibshirani, 1996).

$$\hat{\beta} = arg\min\left\{\frac{1}{2n}\sum_{i=1}^{n}\left(y_i - \sum_j \beta_j x_{ij}\right)^2\right\} \quad [4]$$

$$\text{Subject to: } \|\beta\|_1 \leq t \quad [5]$$

where $\hat{\beta}$ is the coefficient estimate; β is the coefficient; *n* is the number of predictor variables, *x* is the predictor variable, and *y* is the dependent variable. The *i* and *j* are mathematical

conventions the refer to the row (*i*) and column (*j*) of a matrix, which in this case the rows are observations and the columns are variables.

The function inputs are the regression coefficient (β_j), the predictor variables (x_{ij}), the number of predictor variables (n), and the response variable (y_i). Additionally, the argument of the minimum (*arg* min) attains the minimum value to optimize the model and reduce the number of predictor variables. The tuning parameter *t* used was calculated within the JMP generalized regression platform. Models were assessed using the Akaike information criterion with a correction for small sample sizes (AICc) (Forster & Sober, 2011). The final model was chosen by selecting the one with the fewest predictor variables within the AICc indicated as best fit within the JMP software output solution path.

Machine Learning Techniques

The next phase of analysis involved supervised machine learning (ML) techniques to determine the most likely variables associated with yield through pattern recognition to classify the data. The ML was performed with Waikato for Knowledge Analysis (WEKA) Version 3.8.2 machine learning software (Department of Computer Science, University of Waikato, New Zealand). Three different machine learning model algorithms were used to select variables for the predictive model: a support vector machine (SVM) known as sequential minimal optimization; an artificial neural network (ANN) known as multilayer perceptron; and a decision tree (DT) known as J48 (Witten, Elbe, Hall, & Pal, 2011). The algorithms were subject to cross-validation to determine the model's prediction accuracy. Two cross-validation approaches, percent-split and stratified hold-out (K-fold), were used to improve orthogonal (independent) confirmation and overall outcome. The less conservative cross-validation method, percent-split,

divides the training and validation data by a set percentage. The more conservative cross-validation technique, stratified hold-out, divides the data randomly into equal groups and holds out each group as a validation set in each iteration. The percent split used 66% of randomly selected data as the training set and 34% as the test set. The stratified hold-out method used a K-fold technique with *n* = 6-folds, where one-fold was used for cross-validation and *n*-1 folds of randomly ordered data were used as the training set. Feature selection using a single-attribute evaluator function known as 'InfoGainAttributeEval:Ranker' in the WEKA software was used to eliminate irrelevant or redundant variables, reduce model size through ML iterations, and rank variable importance in best-fit models.

Yield, the dependent variable, was binned into three classes: low (< 13.5 Mg ha^{-1}); medium (13.5 to 15 Mg ha^{-1}); and high (> 15 Mg ha^{-1}). There were 49 explanatory variables considered for scenario inclusion: tissue nutrient concentrations (N, P, K, Ca, Mg, S, Fe, Mn, Zn, Cu, and B); additional tissue sample measurements (Al, Na, N:P, N:S, N:K, Fe:Mn); DRIS NBI for macronutrients and for all macro- and micronutrients; soil nutrients (P, K, Ca, Mg, S, Mn, Zn, Cu, and Na); non-nutrient soil characteristics (soil class, HM%, W/V, BS%, Ac, CEC, and pH); grower management practices (plant population density, planting date, row width, previous crop, tillage practice, irrigation use, application of starter fertilizer and/or manure, and fertilizer application rates of N, P, K, and S); and year and growth stage of sampling. The dataset was partitioned into four different dataset scenarios: the complete scenario which included 2017 and 2018 data comprised of all tissue and soil parameters; 2017 tissue and soil scenario comprised of 2017 data and all tissue and soil parameters; 2017 tissue, soil, and management scenario which contained all 49 variables; and the 2017 management only scenario comprised of only management

practice data. All observations included for analysis were from fields with corresponding yields and at least two in-season field samples.

The predictive models were then evaluated using the following parameters: percent of correct classification is the proportion of the observations correctly classified into low, medium and high yield classes; the Kappa statistic, which is related to the true positive rate; and the Area Under the Receiver Operating Characteristic Curve (AUROC) weighted average, which is related to the true negative rate and is a performance metric for how well the model distinguishes between classes. The percent of correct classification is based on the ability of algorithms to correctly classify between designated classes. The Kappa statistic is a measure of model reliability and takes into account that the algorithm results, or agreement of the model to the data, did not occur just by chance (Vieira, Kaymak, & Sousa, 2010). The Kappa statistic ranges from zero to one, where zero is agreement by chance and one is perfect agreement. Therefore, a Kappa statistic of > 0.51 means greater than 50% chance of the correct classification between classes. The AUROC measures the classification performance for a model. The values range from zero to one, and the higher the AUROC, the better the model is able to distinguish between classes.

Results and Discussion

Diagnosis and Recommendation Integrated System - DRIS

There was a significant interaction between year and growth stage for the N and K DRIS indices (Table 2-3). The simple effects for these and the other nutrients are illustrated in Figure 2-1. In 2017, the N index was above zero in tissue samples for all but the maturity growth stage (>R6) (Figure 2-1). However, the N DRIS index was closer to zero in 2018 than in 2017 at the

tasseling growth stage (VT-R1). The K DRIS index was closer to zero in 2018 than in 2017 at the first two growth stages (< V6 and > V7), but at similar levels in both years at the last two growth stages (VT-R1 and >R6).

The Mg and S DRIS indices were more negative in 2017 than in 2018 (Table 2-3). Whereas, we were unable to detect a year effect on the P and Ca DRIS indices. In 2017 and 2018, the mean P DRIS index was ± 10 prior to maturity. On the other hand, Ca and Mg DRIS indices indicated that most tissues samples were < -10 for the <V6, >V7, and VT-R1 growth stages (Figure 2-1A-C) but above zero at maturity. In all but a few fields, the S DRIS index was below zero at all growth stages for both years.

There were no interactions of year and growth stage for the DRIS micronutrients (Table 2-4). The DRIS indices for micronutrients (Table2-4; Figure 2-2) indicated that most tissue samples were at ideal levels for the first three growth stages except for an Fe excess in 2017 and a mild Mn deficiency in 2018, both at the <V6 growth stage (Figure 2-2A). However, by the >R6 stage, the Fe and Mn DRIS indices in 2017 were within ±10 index units of the zero line for most samples, but in 2018 the Fe and Mn indices of nearly half of the tissue samples were 50 index units above the line, indicating a possible excess (Figure 2-2D).

Table 2-3. Least-square means and CV (%) for the diagnosis and recommendation integrated system (DRIS) tissue nutrient indices of N, P, K, Ca, Mg, and S. Means shown are for main effects.

	DRIS macronutrient index											
	N		P		K		Ca		Mg		S	
Year (Y)	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
2017	27.3a[a]	271	-3.8a	376	59.8a	78	-26.6a	655	-25.9b	385	-30.9b	52
2018	12.2b	706	-10.2a	240	34.4b	74	-14.8a	820	-8.3a	342	-13.3a	58
Growth Stage (GS)												
<V6	27.3a	62	-3.8a	238	59.8a	46	-26.6b	75	-25.9b	97	-30.9ab	58
>V7	27.4a	73	2.3a	1326	56.1a	55	-29.9b	70	-23.2b	210	-32.7b	59
VT-R1	19.2a	87	4.2a	299	35.3b	55	-12.2b	138	-13.0b	297	-33.6b	59
>R6	-30.2b	92	-29.6b	118	13.3c	238	43.0a	128	25.8a	120	-22.2a	70
	ANOVA											
Y	***		ns[†]		**		ns		*		***	
GS	***		***		***		***		***		*	
Y x GS	***		ns		***		ns		ns		ns	

[a] Within columns and within years or growth stages, means followed by the same letter are not significantly different according to LSD (.05).

* Significant at the .05 probability level.

** Significant at the .01 probability level.

*** Significant at the .001 probability level.

† ns, non-significant at the .05 probability level.

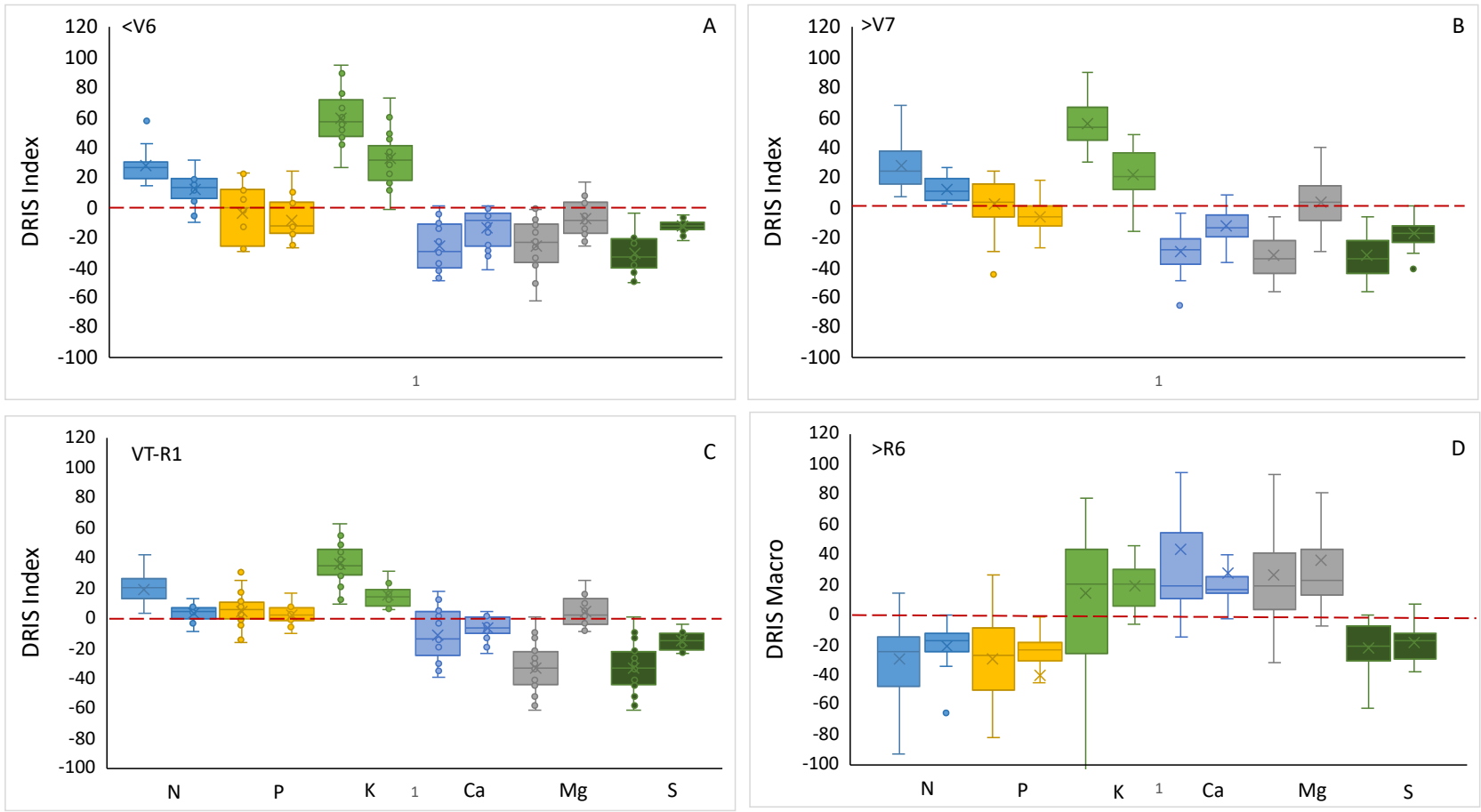

Figure 2-1. DRIS indices for macronutrients N, P, K, Ca, Mg, and S shown for four growth stages <V6 (A); >V7 (B); VT-R1 (C); and >R6 (D). Like-colored box and whiskers are for a single nutrient shown for 2017 (left bar) and 2018 (right bar). Dashed horizontal line indicates the point where nutrients are considered balanced. The box and whisker plots show the extreme upper and lower values excluding outliers and a box representing the inner 50% of the values. The "X" in the box is the mean and the mid-line the median; the upper and lower parts of the box are 25% of the values above and below the median.

Table 2-4. Least-square means and CV (%) for the diagnosis and recommendation integrated system (DRIS) tissue nutrient indices of Fe, Mn, Zn, Cu and B. Means shown are for main effects.

	DRIS micronutrient index									
	Fe		Mn		Zn		Cu		B	
Year (Y)	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
2017	42.2a[a]	195	-32.6b	220	16.1a	286	-23.4a	137	-2.4a	992
2018	29.6a	190	-8.3a	713	9.1a	4298	-16.4a	326	-14.0b	324
Growth Stage (GS)										
<V6	42.2a	109	-32.6b	154	16.1a	156	-23.4b	75	-2.4b	140
>V7	8.9b	198	-9.6ab	326	13.9a	140	-12.6ab	130	-0.6b	1339
VT-R1	13.3b	146	-14.1ab	245	10.7a	186	-8.4a	168	-1.5b	392
>R6	5.5b	1228	2.7a	1116	-9.9b	244	-6.8a	2324	8.5a	230
	ANOVA									
Y	ns[†]		*		ns		ns		**	
GS	***		**		***		**		**	
Y x GS	ns		ns		ns		ns		ns	

[a] Within columns and within years or growth stages, means followed by the same letter are not significantly different according to LSD (0.5).

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

*** Significant at the 0.001 probability level.

† NS, non-significant at the 0.05 probability level.

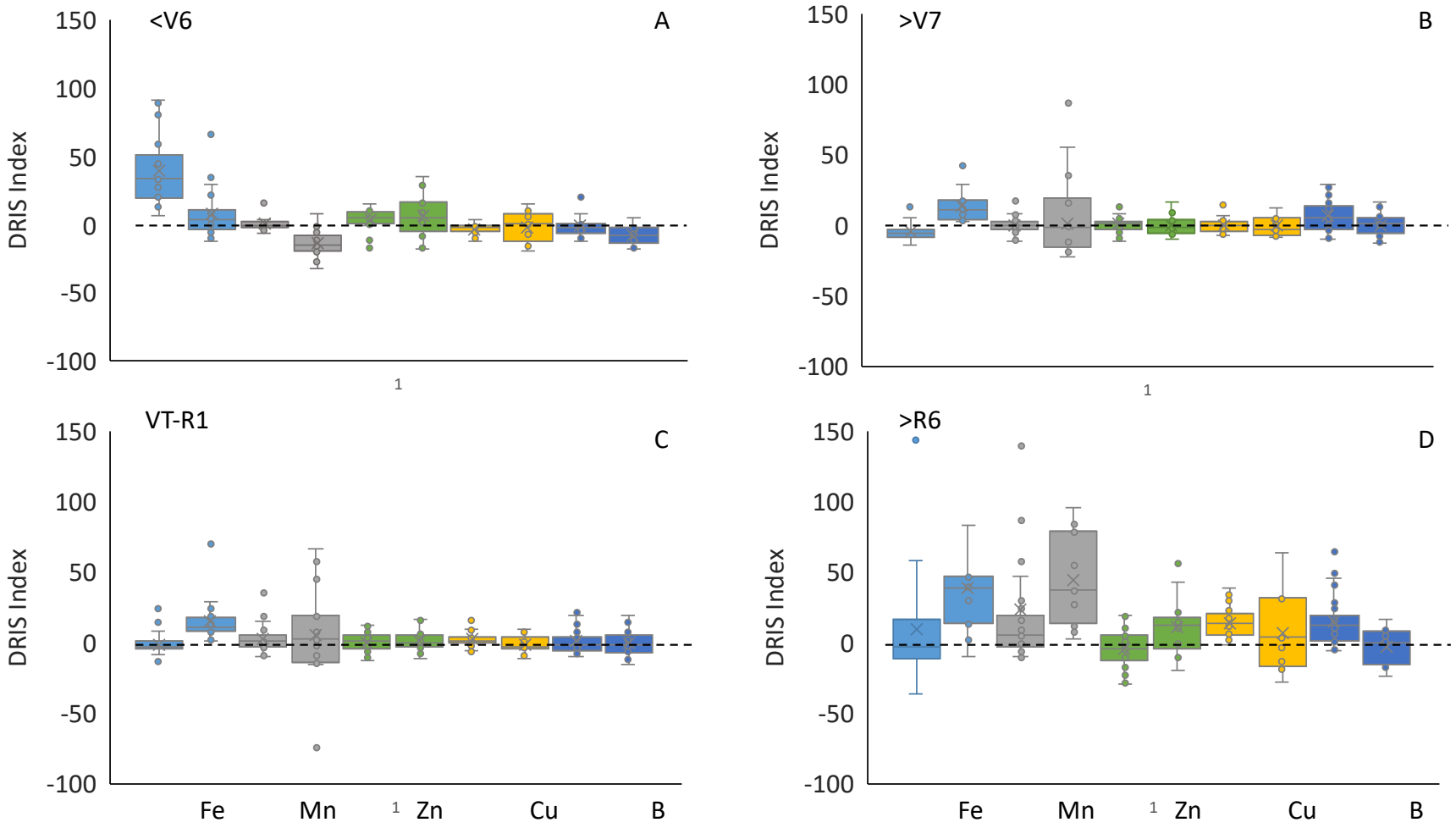

Figure 2-2. The DRIS indices for the micronutrients Fe, Mn, Zn, Cu, and B grouped by nutrient for the growth stages: <V6 (A); >V7 (B); VT-R1 (C); and >R6 (D). Like-colored box and whiskers are for a single nutrient shown for 2017 (left bar) and 2018 (right bar). Dashed horizontal line indicates the point where nutrients are considered balanced.

In general, macronutrient DRIS indices were higher at growth stages prior to maturity for N, P, and K, and lower for Ca and Mg. Tissue nutrient concentrations change over the growing season due to availability, translocation, growth, and environmental factors. A study by Mundorf et al. (2015) found a three-way interaction between time of day, growth stage, and field location for the N, P, Ca, and Mg DRIS indices, and a two-way interaction of time of day and growth stage for the K DRIS index. In their study, N tissue concentrations of samples taken at V4 to V6 between 7 AM and 1 PM were ~10% higher than from samples taken at or later than 4 PM, but there was no difference at the R2 growth stage. Magnesium and P concentrations, on the other hand, varied by 24 to 32%, and 19 to 41%, respectively; but concentrations were not consistent by growth stage and time in both fields (Mundorf et al., 2015). This suggests that consistency in sampling time is an important factor in making accurate nutrient diagnosis when using tissue. In addition to growth stage and time of day differences in tissue concentrations, the plant part sampled can also influence the outcome of DRIS. Karlen et al. (1988) showed that DRIS indices calculated using the upper leaves were within ±18 index units of each other for all macro- and micronutrients with exception for Fe at the V4 and V8 stages and N near the R6 growth stage. However, Walworth and Sumner (1986) suggested that though DRIS indices varied depending on leaf position sampled, the nutrient diagnosed as most limiting did not.

When using DRIS to determine nutrient status, the ideal balance among plant tissue nutrients is achieved when all indices equal zero. The interpretation of DRIS results requires taking into account the status of tissue nutrients relative to one another within a tissue sample (Walworth & Sumner, 1987). In addition to indicating a limiting nutrient by the most negative value, it is possible for nutrient imbalances to occur in a sample when results show most nutrient

indices close to zero and one nutrient with a DRIS index well above zero. The latter could indicate limitations to growth due to toxicity of the nutrient; or antagonism between nutrients such as high K leading to a reduction in the uptake Mg (Terman, Allen, & Bradford, 1975). Taking this into account, the overall nutrient status of tissue samples in our study indicated that S was limiting throughout the season in both years (Figure 2-1; Table 2-3). There are a few possible explanations for this result. First, S deposition has been decreasing over the last few decades as efforts to clean powerplant emissions have increased. This has led to a reduction of S availability in the surface horizon, reducing S available for plant uptake. Second, the DRIS indices for S and other nutrients may not be accurate due to irrelevant or inaccurate norm values established over 30 years ago and not representative of today's hybrids. Lastly, the nutrient pairs used for the calculation of the S DRIS index might need to be reevaluated; perhaps S should be the denominator rather than the numerator in more nutrient pairs than just Ca/S. The nutrient pairs we used for the DRIS calculations were based on those established by Elwali et al. (1985). The form of the nutrient expression used in the DRIS calculation are the form with largest variance ratio in a comparison of variances from high-yielding and low-yielding segments of the population (Walworth & Sumner, 1987). It is likely that a more regionally focused set of norms would change the form of the expression. Walworth and Sumner (1987) stated that the form of the expression used for DRIS calculations can profoundly affect the outcome, making the case for a set of updated norms.

In addition to S, Mg and Ca were low in relation to other nutrients. This could have been due to the high level of K leading to antagonisms between Mg or Ca and reducing uptake. Research has shown that a high K/Mg ratio in the soil solution contributes to higher K uptake and

a reduction in Mg uptake (Ohno & Grunes, 1985; Rahmatullah & Baker, 1981). Another possible explanation is that our DRIS calculations used universal norms established by Elwali et al. (1985) and not regional norms for the Southeast. Walworth and Sumner (1986) compared the corn DRIS norms developed in different locations and found that Ca and Mg were two nutrients that varied significantly. The southeastern norms were lower compared to other locations, implying that Ca and Mg norm values would likely produce more accurate diagnosis if established locally. Local calibration of DRIS norms for corn was shown to improve accuracy of diagnosis (Escano et al. 1981).

A comparison of DRIS norms from four research groups revealed that DRIS indices varied greatly and resulted in differing diagnosis (dos Anjos Reis, 2002) . It was suggested that use of norms established from areas with similar climatic and soil conditions would lead to more accurate diagnosis of limiting nutrients. An evaluation by Beverly (1993) of the accuracy of sufficiency ranges and DRIS indices for N, P, K, and S in corn found that both systems were equally reliable in diagnosing P and K deficiencies, but that DRIS was less reliable than SR for N and S.

We examined relationships between an individual tissue macronutrient and the corresponding DRIS index in two ways —using DRIS indices calculated with only macronutrients or including micronutrients. Likewise, comparisons of individual tissue micronutrients and their corresponding DRIS indices were made using DRIS indices calculated using only micronutrients or including macronutrients. When DRIS indices were computed using only macronutrient tissue concentrations, P, K, Ca, and Mg were more strongly correlated with the matching DRIS index in 2017 than in 2018 (Table 2-5). Conversely, S tissue concentration was moderately correlated and more weakly in 2017 (R^2=0.24) than in 2018 (R^2=0.50) (Table 2-5). However, when micronutrients

were included in the DRIS calculation, the strength of correlations were either similar or increased for N, Ca, Mg and S in 2017; but in 2018, resulted in only a weak correlation between N tissue concentration and N DRIS. In 2017, micronutrients Fe and B had stronger correlations between their DRIS indices and tissue concentrations than in 2018 regardless of DRIS being calculated using only micronutrients or both macro- and micronutrients (Table 2-6). Conversely, correlations of Mn, Zn, and Cu DRIS indices with their tissue concentrations were stronger in 2018 if DRIS was calculated using only micronutrients; while in 2017 the correlations were stronger when all macro- and micronutrients were included in the analysis. The only correlation in 2018 from DRIS calculations made with macro-and micronutrients was a weak correlation between Mn DRIS and Mn tissue concentration. Environmental factors likely influenced the differences among the DRIS indices correlations with tissue nutrients. In 2018, there were several strong weather events, such as two hurricanes, that could have contributed to large changes in soil conditions and nutrient availability. It is likely that DRIS is less reliable when environmental conditions shift drastically within a growing season, as was the case in 2018.

Table 2-5. Correlations of plant tissue macronutrient concentrations with the DRIS indices calculated for macronutrients (N, P, K, Ca, Mg, and S) alone and with the micronutrients (Fe, Mn, Zn, Cu, and B). The DRIS indices were calculated using 2017 and 2018 tissue nutrient concentrations taken from fields in the NC Corn Yield Contest study at four growth stages (<V6, >V7, VT-R1, and >R6).

	Macronutrients					
Year & DRIS components	N	P	K	Ca	Mg	S
2017	R^2	R^2	R^2	R^2	R^2	R^2
Macronutrient	$.60^{***}$	$.54^{***}$	$.57^{***}$	$.77^{***}$	$.77^{***}$	$.24^{***}$
All nutrients	$.63^{***}$	$.44^{***}$	$.40^{***}$	$.79^{***}$	$.81^{***}$	$.50^{***}$
2018						
Macronutrients	$.64^{***}$	$.36^{***}$	$.51^{***}$	$.37^{***}$	$.48^{***}$	$.43^{***}$
All nutrients	$.07^{*}$	$.05^{ns\dagger}$	$.007^{ns}$	$.001^{ns}$	$.004^{ns}$	$.02^{ns}$

* Significant at the 0.05 probability level. ** Significant at the 0.01 probability level. *** Significant at the 0.001 probability level.

† ns, non-significant at the 0.05 probability level.

Table 2-6. Correlations of plant tissue micronutrient concentrations with the DRIS indices calculated for the micronutrients (Fe, Mn, Zn, Cu, and B) alone and with the macronutrients (N, P, K, Ca, Mg, and S). The DRIS indices were calculated using 2017 and 2018 tissue nutrient concentrations taken from fields in the NC Corn Yield Contest study at four growth stages (<V6, >V7, VT-R1, and >R6).

	Micronutrients				
Year & DRIS components	Fe	Mn	Zn	Cu	B
2017	R^2	R^2	R^2	R^2	R^2
Micronutrients	$.67^{***}$	$.62^{***}$	$.54^{***}$	$.16^{***}$	$.48^{***}$
All nutrients	$.74^{***}$	$.89^{***}$	$.70^{***}$	$.42^{***}$	$.64^{***}$
2018					
Micronutrients	$.32^{***}$	$.66^{***}$	$.64^{***}$	$.25^{***}$	$.32^{***}$
All nutrients	$.001^{NS}$	$.10^{*}$	$.001^{NS}$	$.001^{NS}$	$.05^{NS}$

*Significant at the .05 probability level. **Significant at the .01 probability level. ***Significant at the .001 probability level. † ns, nonsignificant at the .05 level.

A comparison of DRIS indices to the tissue nutrient CRV showed that the DRIS regression line and the CRV line did not intersect at the zero line for some nutrients, indicating lack of agreement between the two nutrient evaluation systems (Figures 2-3 to 8). The nutrients that showed agreement between the two systems were P and Ca in both years, and N in 2018. For P and Ca in both years and N in 2018, the assessments based on the CRV were nearly identical to those based on the DRIS zero line (Figures 2-4, 2-6, and 2-3B, respectively). In 2017, the comparison of DRIS N and tissue N showed levels above the zero line for the three early growth stages, whereas at the > R6 stage, most observations were below both the DRIS zero line and the critical value. Most tissue samples were below the critical value for Mg and below zero for the DRIS index, but some samples had Mg concentrations that would be considered ideal according to DRIS, but were below the critical value (Figure 2-7). Conversely, the S DRIS Index indicated that S was deficient, however, the S tissue concentration was above the CRV throughout most of the growing season (Figure 2-8).

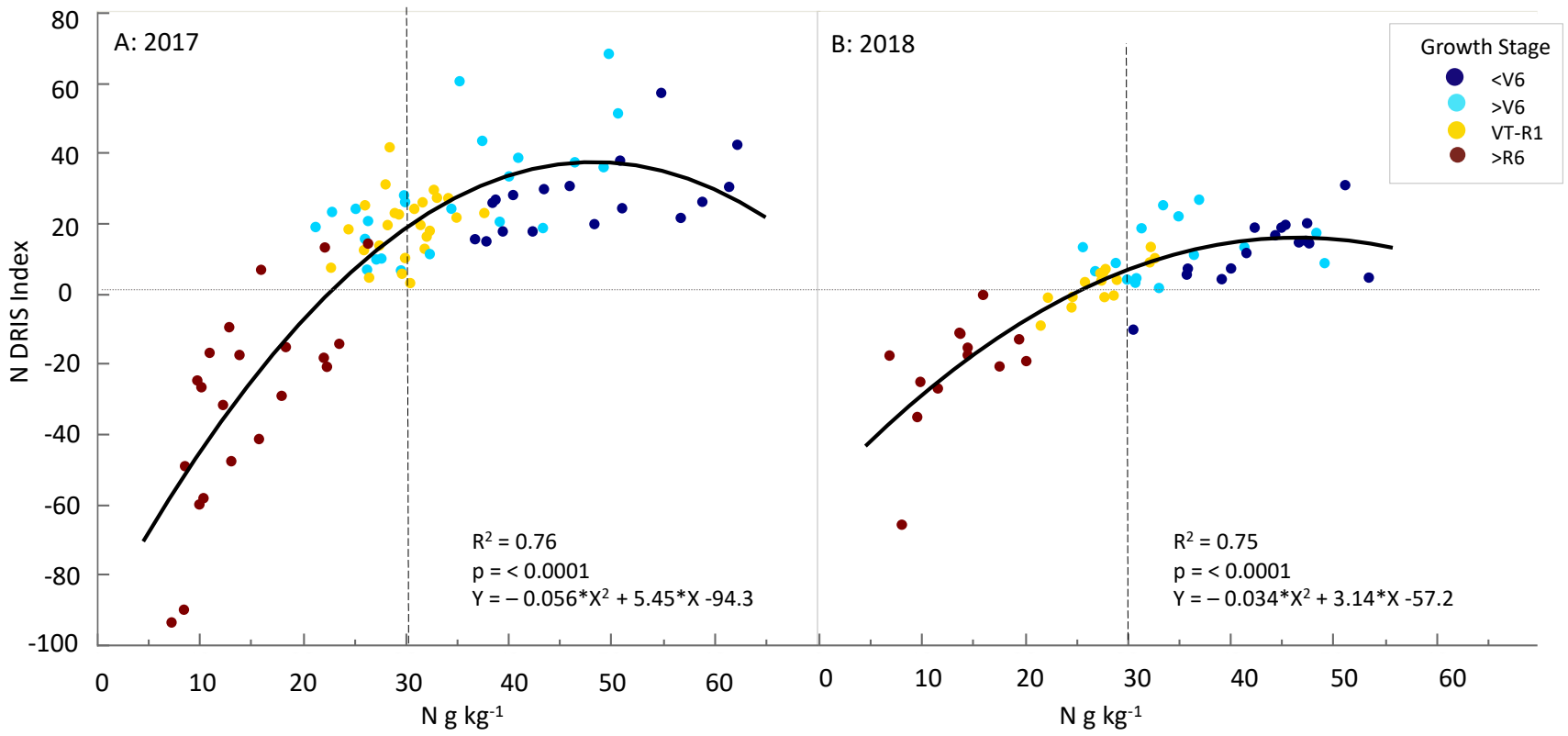

Figure 2-3. Relationships between the N DRIS Index using macronutrients and plant tissue N concentration for samples taken at four growth stages from fields of NC Corn Yield Contest participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

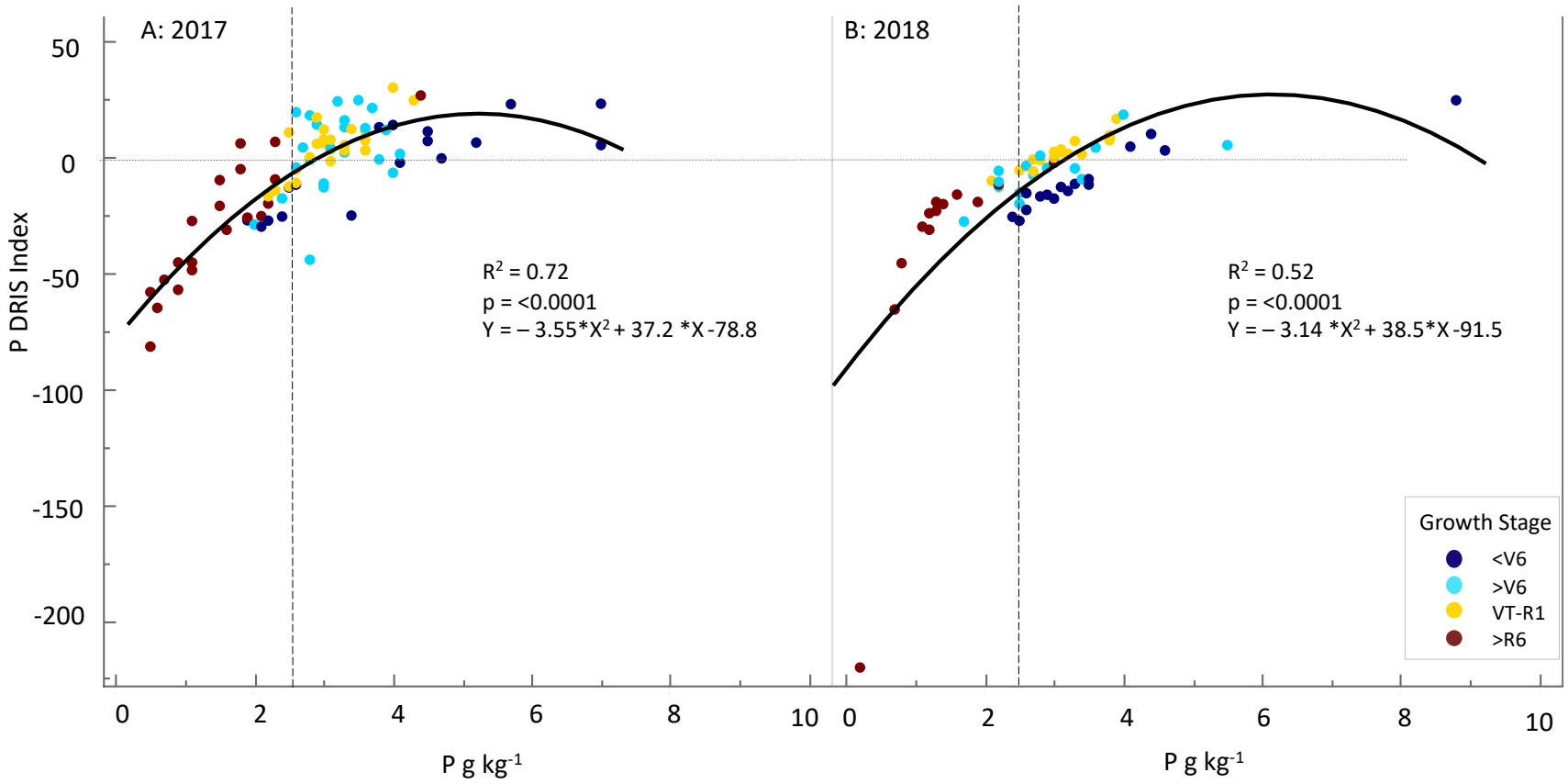

Figure 2-4 . Relationships between the P DRIS Index using macronutrients and plant tissue P concentration for samples taken at four growth stages from fields of NC Corn Yield Contest participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

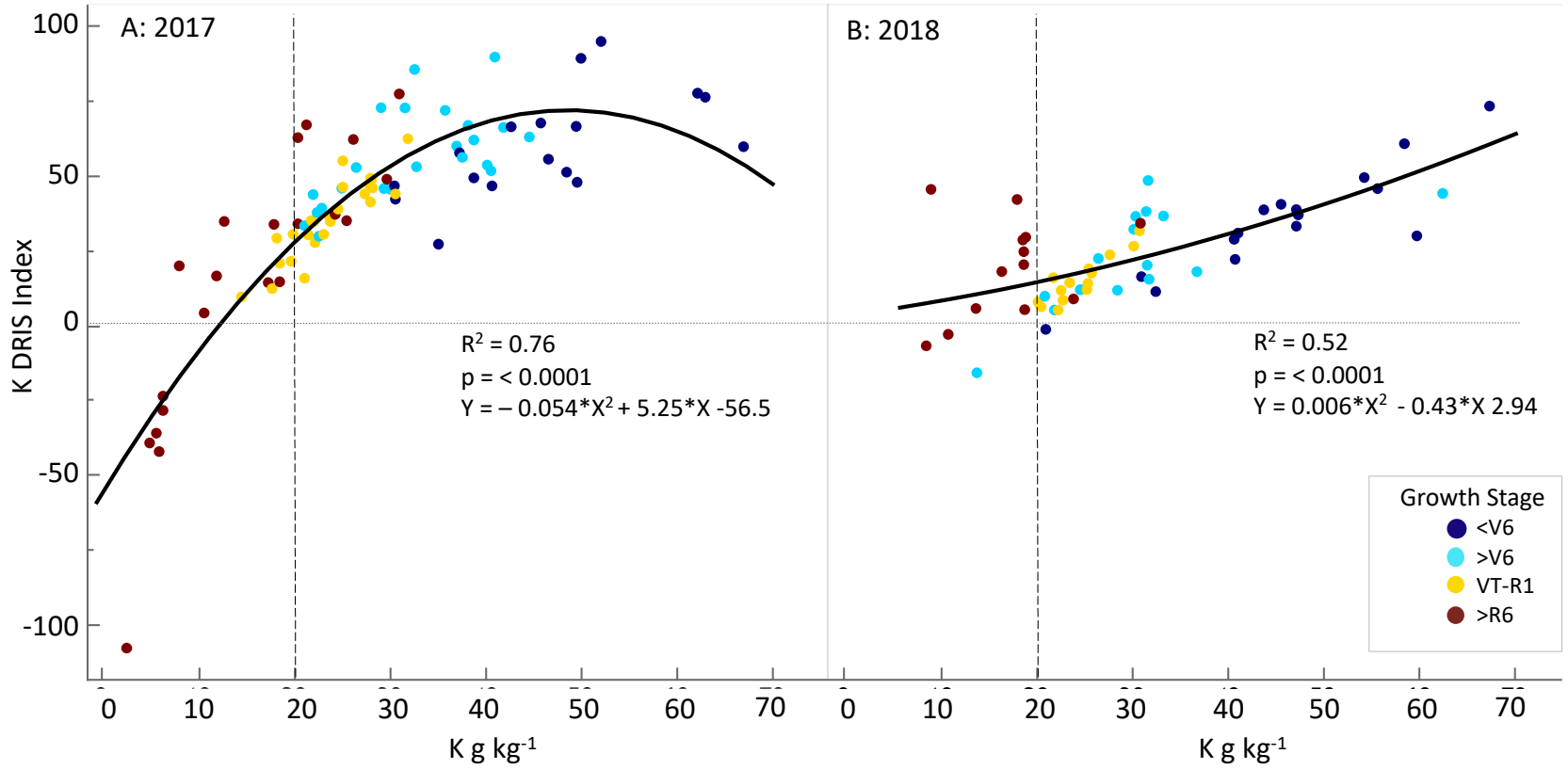

Figure 2-5. Relationship between K DRIS Index using macronutrients and plant tissue K concentration for samples taken at four growth stages from fields of NC Corn Yield Contest study participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

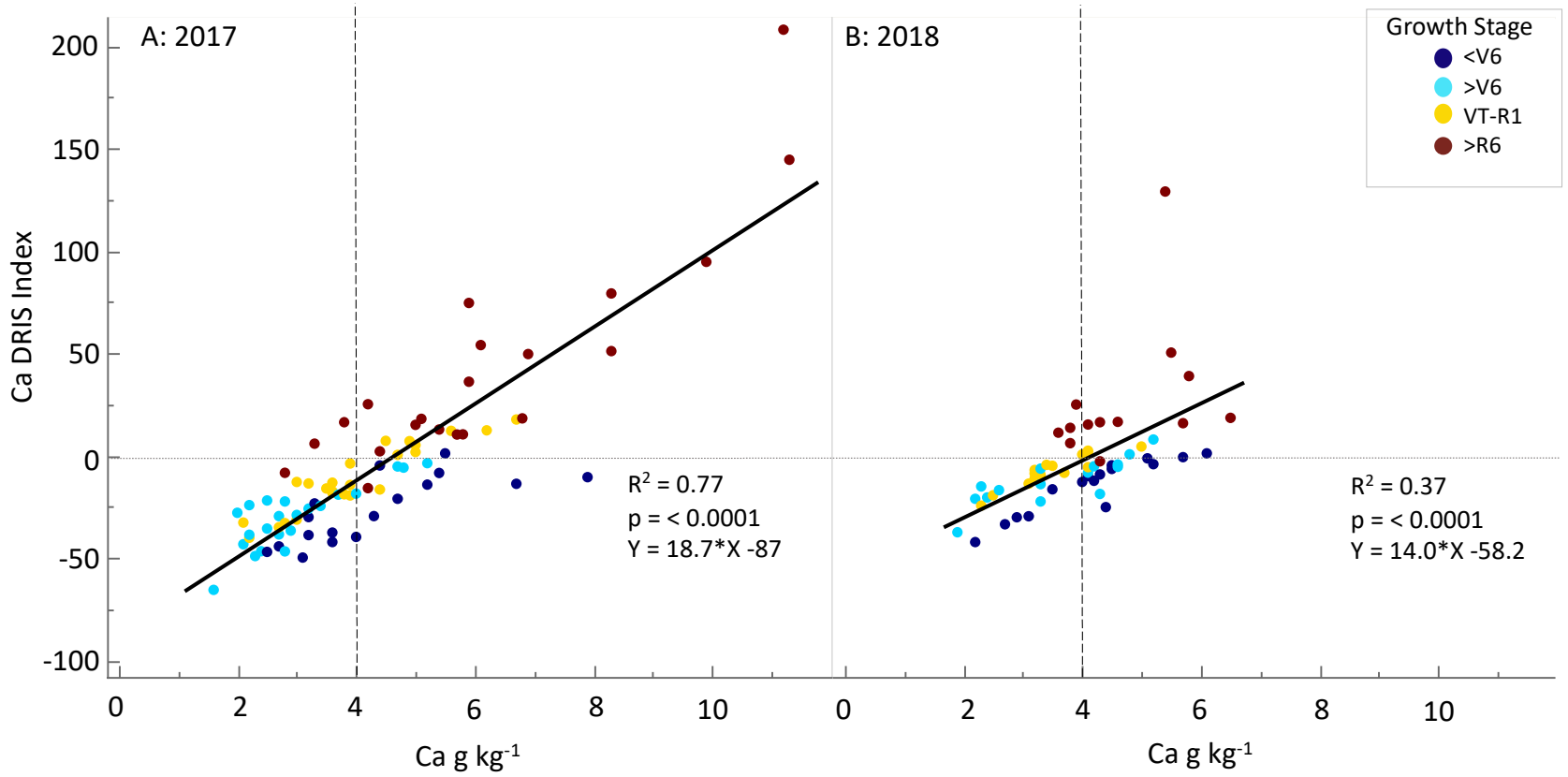

Figure 2-6. Relationship between Ca DRIS Index using macronutrients and plant tissue Ca concentration for samples taken at four growth stages from fields of NC Corn Yield Contest study participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

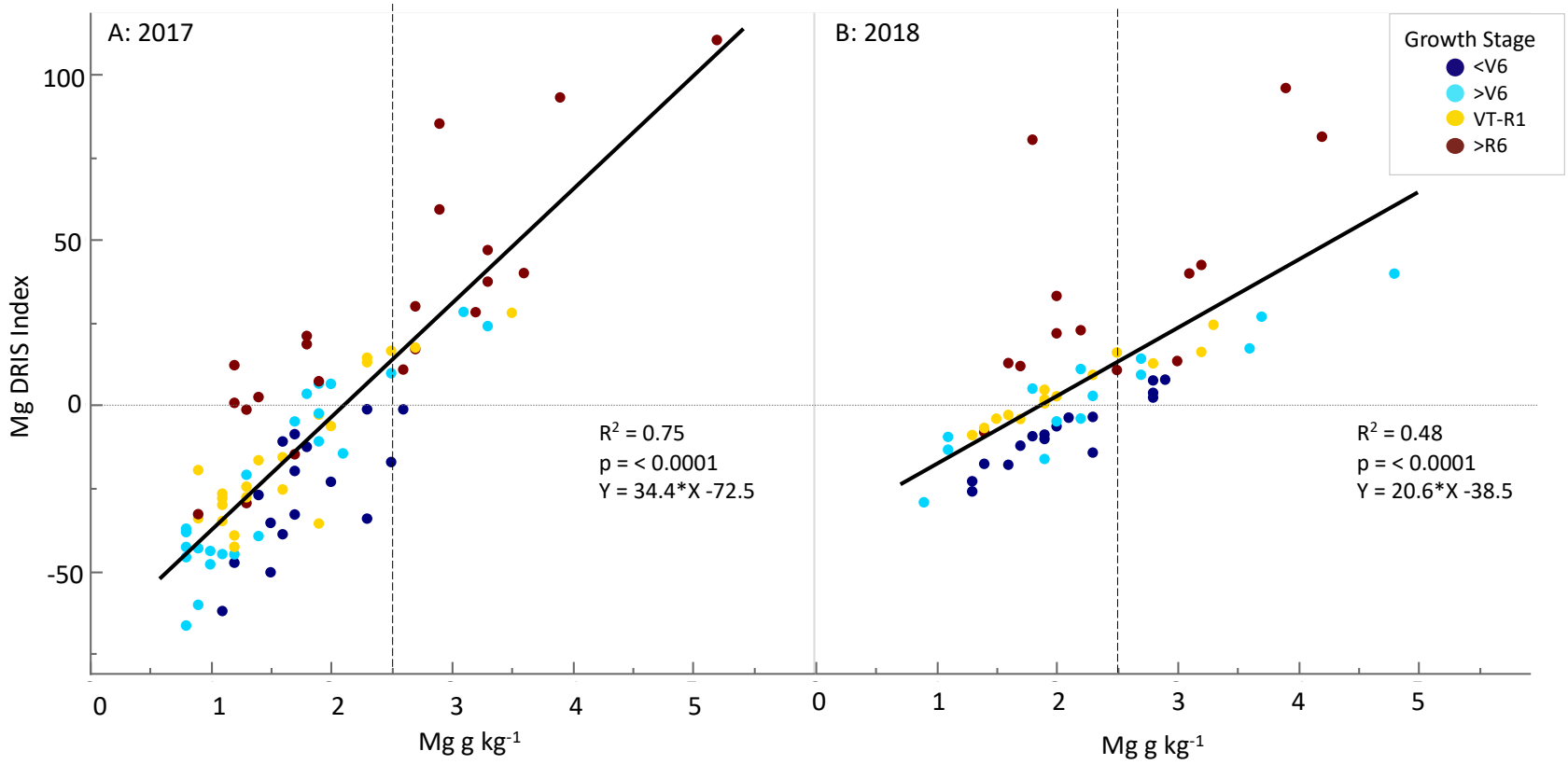

Figure 2-7. Relationship between Mg DRIS Index using macronutrients and plant tissue Mg concentration for samples taken at four growth stages from fields of NC Corn Yield Contest study participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

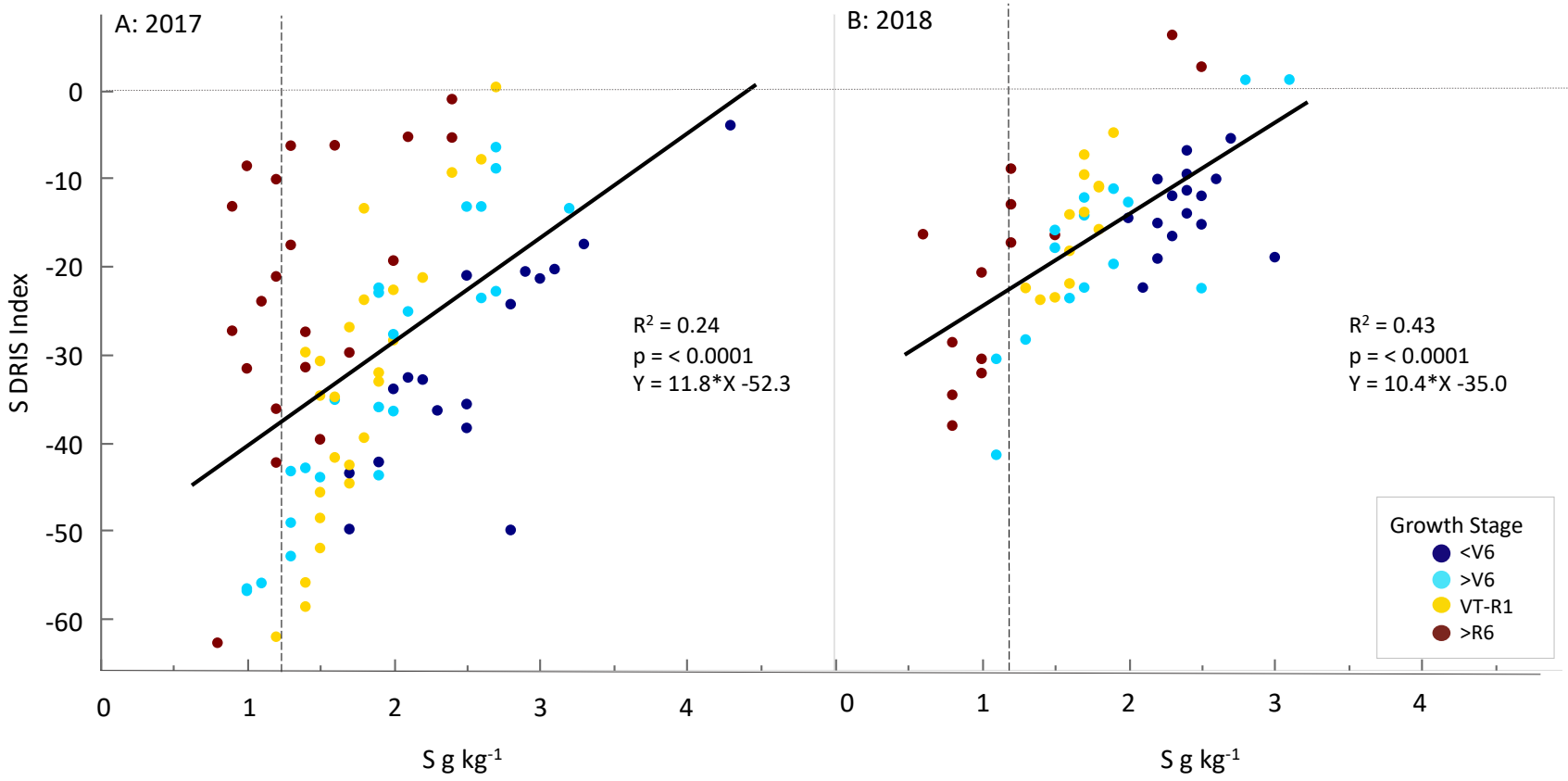

Figure 2-8. Relationship between S DRIS Index using macronutrients and plant tissue S concentration for samples taken at four growth stages from fields of NC Corn Yield Contest study participants in 2017 (A) and 2018 (B). The dashed vertical line indicates the critical value for the tissue nutrient concentration at the VT-R1 growth stage (Campbell and Plank, 2000). The dashed horizontal line at DRIS = 0 indicates nutrient balance.

The nutrient balance index (NBI) is the sum of the absolute value of all DRIS indices corresponding to a tissue sample. The mean NBI, averaged over growth stages for each field, was generally higher in 2017 than in 2018 (Figures 2-9A and B); the NBI means were 181 and 103 in 2017 and 2018, respectively. In 2017, there was a negative relationship between the NBI and yield; however, in 2018, the trend was positive but not statistically significant (p = .07) (Figures 2-3A and B). Walworth and Sumner (1987) showed an inverse relationship between yield and NBI based on results from numerous studies.

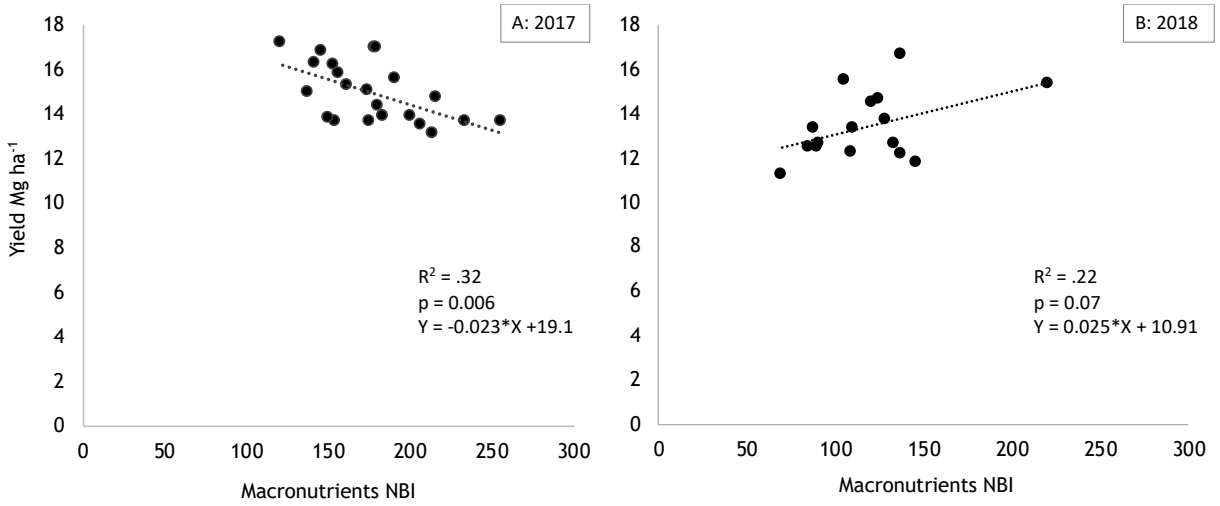

Figure 2-9. Relationship of yield and the macronutrient nutrient balance index (NBI) for fields in 2017 (A) and 2018 (B).

LASSO Models Performance in Predicting Yield

The final LASSO model outcome varied depending on the combination of explanatory variables included and the method of dataset partitioning: complete, by year, or by growth stage. Modeling that included management practices could be performed only in 2017, as too many fields in 2018 were missing management practice data.

Based on R^2 values, the poorest models were obtained when the full dataset was utilized to run the LASSO algorithm regardless of the number of explanatory variables (Table 2-7). Model performance, with one exception, tended to improve when the dataset was divided by year, while improvement in model performance when divided by growth stage depended on explanatory variable combinations. In nearly all cases, calculations that included two-way interactions resulted in better fit models than without, with exceptions for the tissue macro- and micronutrient variable combination (TMM) in 2017, the tissue macronutrients variable combination (TM) and TMM at > R6, and the soil macronutrient variable combination (SM) at < V6 and VT-R1. The best-fit model that included only nutrient variables was that containing soil micro and macronutrients (SMM) with two-way interactions for the 2018 dataset (R^2 = .89). However, the best overall fit model was produced when management practices were combined with tissue and soil macro and micronutrients (TSMM+MP) and using only 2017 data (R^2 = .94) (Table 2-7). The addition of management practices (MP), which could only be performed for 2017 data, improved model performance, and MP alone had stronger models than any combination containing only nutrients for all but the SMM with two-way interactions (Table 2-7). The number of explanatory variables retained in the models was between one and twenty-six (Table 2-8). The interpretation becomes more difficult as the number of variables increases, which increases the complexity of the model. Therefore, finding the ideal model, one that has the highest accuracy with the least number of variables, becomes a balance between model fit and interpretability.

Table 2-7. Coefficient of determination (R2) and root mean square error (RMSE) from least absolute selection and shrinkage operator (LASSO) regression models for grain yield as a function of different combinations of explanatory variables for: the full dataset; by year (2017 and 2018); and by growth stage (< V6, > V7, VT-R1, and > R6). Abbreviations for variable combinations: TM = tissue macronutrients; TMM = tissue macro- and micronutrients; SM = soil macronutrients; SMM = soil macro- and micronutrients; TSM = tissue and soil macronutrients; TSMM = tissue and soil macro- and micronutrients; All soil = soil macro- and micronutrients and pH, CEC, base saturation, exchangeable acidity, humic matter %, and weight/volume ratio; and MP = management practices (planting date, planting density, row width, use of starter fertilizer, tillage practice, irrigation use, and fertilizer application rate for N, P, K, and S).

	Year						Growth stage							
	2017 & 2018		2017		2018		<V6		>V7		VT-R1		>R6	
Variable combinations	R^2	RMSE[a]	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
TM	.12	1.42	.17	1.20	.37	1.08	.31	1.36	.12	1.29	.12	1.46	.48	1.06
+ Interactions	.26	1.30	.50	0.93	.38	1.06	.43	1.23	.53	0.94	.34	1.26	.38	1.12
TMM	.15	1.40	.26	1.14	.38	1.07	.39	1.27	.07	1.33	.12	1.45	.48	1.07
+ Interactions	.28	1.29	.15	1.22	.45	1.01	.67	.94	.39	1.08	.26	1.34	.28	1.26
SM	.10	1.44	.21	1.17	.47	0.98	.26	1.41	.13	1.28	.30	1.29	.09	1.40
+ Interactions	.33	1.24	.50	0.93	.55	0.91	.16	1.50	.35	1.11	.17	1.41	.21	1.30
SMM	.16	1.39	.27	0.12	.47	0.98	.24	1.42	.14	1.28	.41	1.19	.10	1.40
+ Interactions	.61	0.95	.86	0.48	.89	0.45	—	—	—	—	—	—	—	—
TSM	.23	1.33	.35	1.06	.53	0.93	.43	1.23	.23	0.21	.48	1.12	.50	1.04
+ Interactions	.28	1.29	.49	0.94	.55	0.91	—	—	—	—	—	—	—	—
TSMM	.33	1.24	.49	0.93	.49	0.96	.16	1.49	.37	1.09	.23	1.36	.59	0.94
All Soil	.25	1.31	.50	0.93	.65	0.80	.16	1.50	.21	1.23	.54	1.05	.50	1.04
MP	.50	0.97	.79	0.61	—	—	—	—	—	—	—	—	—	—
TM + MP	.56	0.92	.82	0.55	—	—	—	—	—	—	—	—	—	—
TMM+MP	.64	0.83	.89	0.43	—	—	—	—	—	—	—	—	—	—
TSM+MP	.70	0.76	.86	0.48	—	—	—	—	—	—	—	—	—	—
SM+MP	.59	0.89	.81	0.57	—	—	—	—	—	—	—	—	—	—
TSMM+MP	.83	0.58	.94	0.32	—	—	—	—	—	—	—	—	—	—

[a]The RMSE was calculated from the difference between actual and predicted yield with units of Mg ha^{-1}. Yield range was 11.3 to 17.1 Mg ha^{-1}.

The LASSO calculations that included interactions often resulted in more complex models and dividing the dataset by year tended to result in more complex models than when divided by the growth stage (Table 2-8). The model with the highest performance from the full 2017 and 2018 datasets with two-way interactions was SMM with two-way interactions (Table 2-7), but the model included 26 variables making it too complex to interpret (Table 2-8).

The performance of LASSO was based on how well the model prediction of yield fit the actual yield data and was reported as the non-adjusted R^2. The size of a model is another factor to consider in determination of the ideal or best-fit model. Generally, increasing the number of explanatory variables increases model R^2. In our case, there was a moderate positive correlation (R^2 = .51) between the number of model parameters and model performance using the complete dataset. Dividing the dataset by year strengthened this correlation for both 2017 (R^2 = .66) and 2018 (R^2 = .61). Dividing the dataset by growth stage strengthen the correlations further. For <V6, > V7, VT-R1, and > R6, the R^2 were .80, .91, .81, and .93, respectively) (Table 2-8). The combination and type of explanatory variables influenced the performance of the model more than the number of variables within the model for the complete dataset, and similarly for the dataset divided by year. For the growth stage datasets, the size of the model was more important, however, few models had R^2 > .50.

Table 2-8. The number of retained variables within the least absolute selection and shrinkage operator (LASSO) regression models for grain yield as a function of different combinations of explanatory variables for: the full dataset; by year (2017 and 2018); and by growth stage (< V6, > V7, VT-R1, and > R6). Abbreviations for variable combinations: TM = tissue macronutrients; TMM = tissue macro- and micronutrients; SM = soil macronutrients; SMM = soil micro- and macronutrients; TSM = tissue and soil macronutrients; TSMM = tissue and soil micro- and macronutrients; All soil = soil micro- and macronutrients and pH, CEC, base saturation, exchangeable acidity, humic matter %, and weight/volume ratio; and MP = management practices (planting date, planting density, row width, use of starter fertilizer, tillage practice, irrigation use, and fertilizer application rate for N, P, K, and S).

Variable combinations	Year			Growth Stage			
	2017 & 2018	2017	2018	<V6	>V7	VT-R1	>R6
	Number of retained variables						
TM	4	3	3	3	3	3	6
+ Interactions	13	12	4	5	10	6	4
TMM	6	5	3	4	2	3	4
+ Interactions	19	7	16	10	5	4	3
SM	4	4	5	4	2	3	2
+ Interactions	10	5	8	1	6	3	3
SMM	6	5	5	4	3	5	2
+ Interactions	26	26	22	—	—	—	—
TSM	7	8	6	5	4	7	5
+ Interactions	11	17	17	—	—	—	—
TSMM (n=21)	14	10	5	2	7	3	8
All Soil	7	6	12	4	4	7	6
MP	9	10	—	—	—	—	—
TM + MP	12	14	—	—	—	—	—
TMM+MP	12	17	—	—	—	—	—
TSM+MP (n=22)	17	18	—	—	—	—	—
SM+MP	12	13	—	—	—	—	—
TSMM+MP	21	20	—	—	—	—	—
	Correlation of the number of variables retained in a model versus model R^2						
R^2	.51	.66	.61	.80	.91	.81	.93

Among the tissue macronutrients, N and Mg had the highest model retention rate in the complete dataset (58 and 52%, respectively) (Figure 2-10A). When divided by year, the tissue macronutrients with the highest model retention rate were N and S in 2017 and Ca in 2018. Only Mg was retained at approximately the same rate in both years. Of the tissue micronutrients, Mn and B were retained more often than other micronutrients in the complete dataset (Figure 2-

10B). Soil P, K, Ca, and Mg were the soil nutrients with the highest retention rate (>55%) in the complete dataset; whereas in the by-year dataset, soil nutrients with retention rates > 55% were Ca, Mg, S, Mn, and Zn in 2017, and only K and Mn in 2018 (Figure 2-10C).

Relative to the complete and by-year datasets, the retention rate for most tissue macro- and micronutrients decreased when the dataset was divided by growth stage (Figure 2-11). The earliest growth stage (< V6) models retained N, P, S, Cu, and B more often than the three other growth stages (> V7, VT-R1, and > R6) (Figure 2-11A and B). Tissue Mg, on the other hand, was retained more often in the >V7 growth stage models (Figure 2-11A).

Similar to the tissue nutrients, the soil nutrients were retained in models at a lower frequency when the dataset was divided by the growth stage than when the whole dataset was used or divided by year. The soil nutrients retained most often were K, Ca, Mg, and S, but the model retention rate changed depending on the growth stage. At the < V6 growth stage, Mg and S were retained more than other nutrients. At the > V7 stage, Ca and Mg were retained at a higher rate, and by the tasseling stage (VT-R1), K, Ca, Mg, and Zn were retained more often.

The division of the dataset into growth stages included both years, as it was not possible to divide each growth stage by year due to the size of initial dataset. Taking into account the large difference between years, it is likely that the growth-stage specific models would differ by year as well. The similarity among retained variables within full dataset and the dataset divided by growth stage indicated that tissue N, P, Mg, S, and B likely had a strong influence on yield.

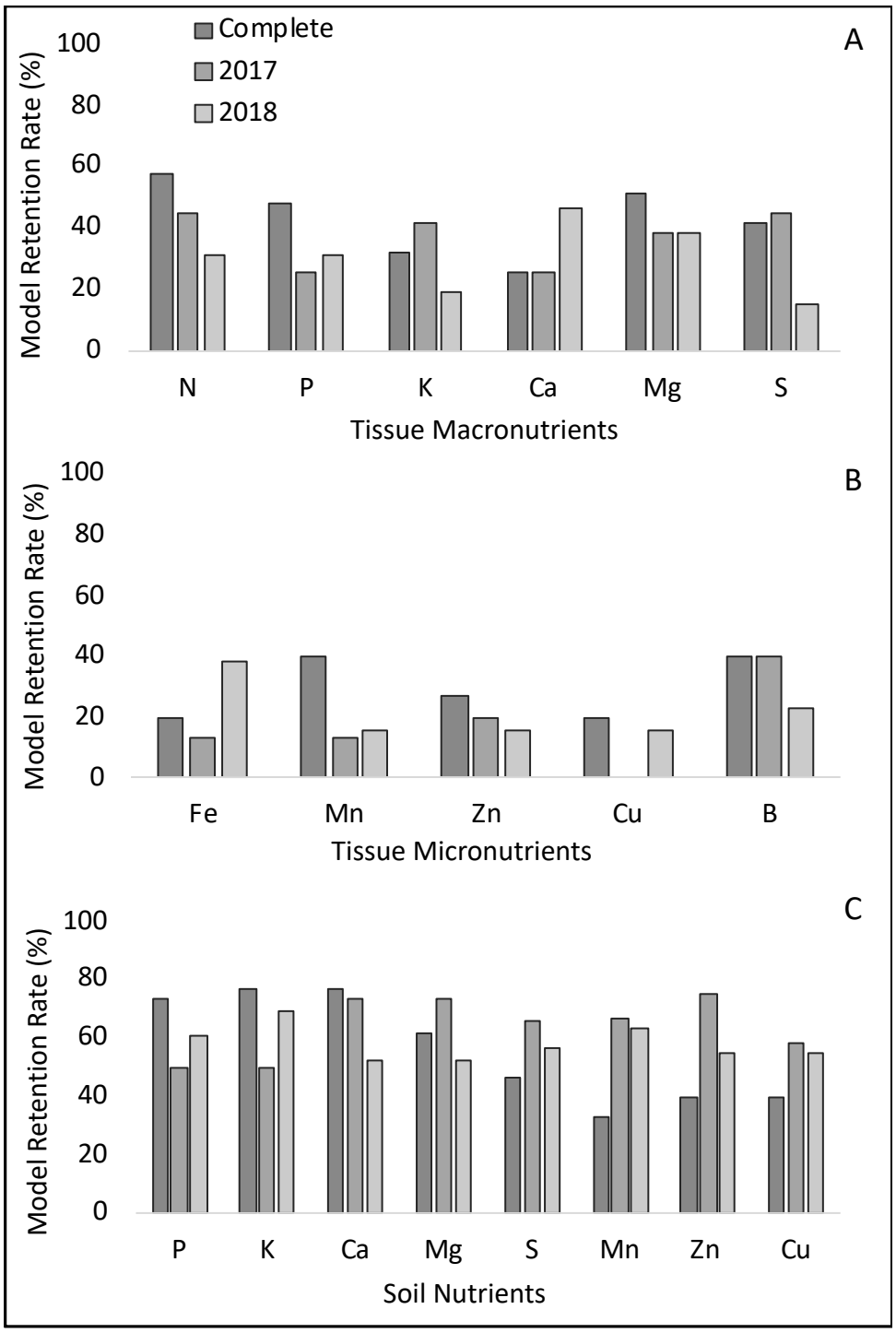

Figure 2-10. Model retention rates (%) of explanatory variables among the complete (2017 and 2018), 2017, and 2018 datasets for tissue macronutrients (A), tissue micronutrients (B), and soil nutrients (C). Data from tissue and soil samples taken in 2017 and 2018 from fields of NC Corn Yield Contest study participants.

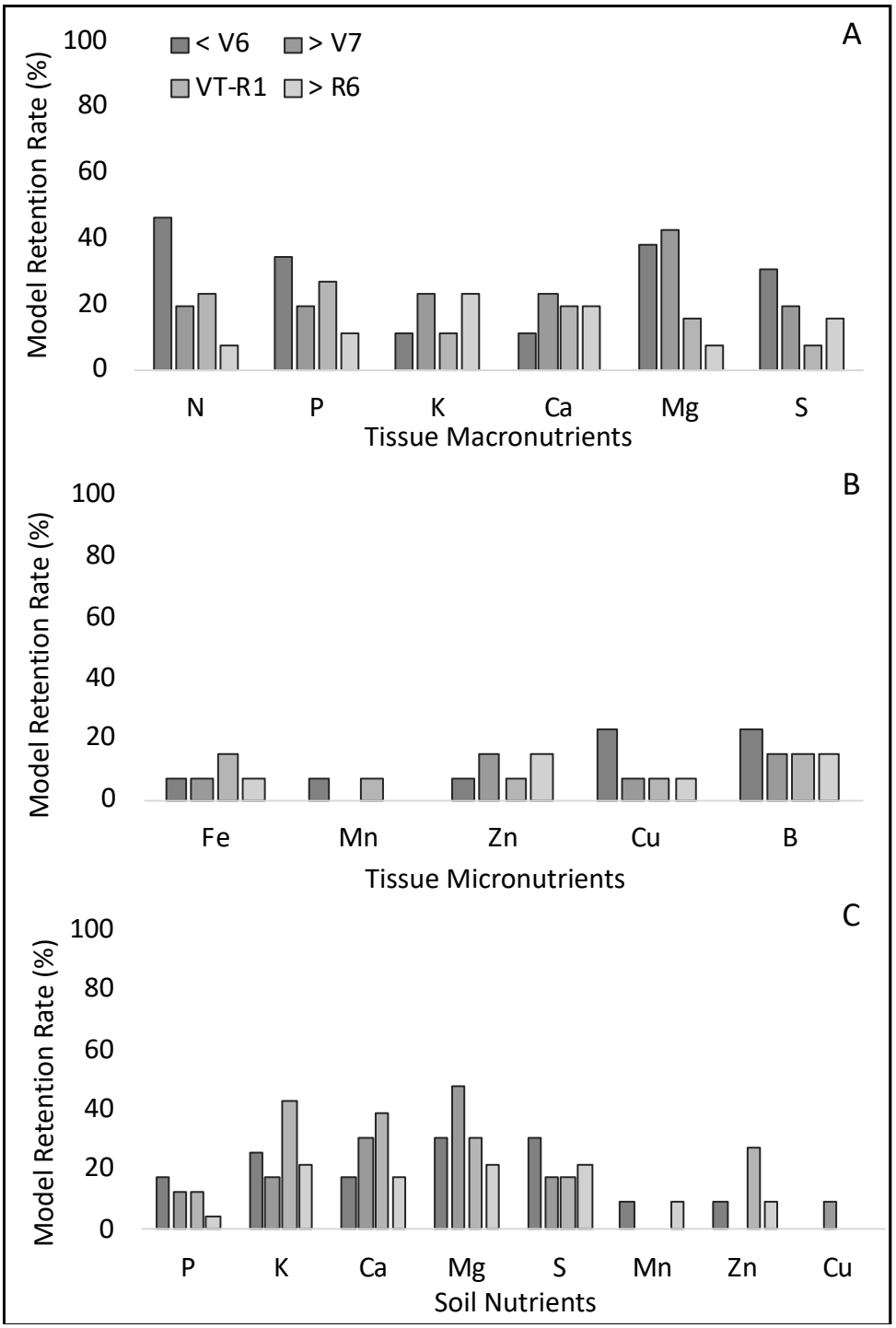

Figure 2-11. Model retention rates of explanatory variables among the datasets divided by four growth stages (< V6, > V7, VT-R1, and > R6) for tissue macronutrients (A), tissue micronutrients (B), and soil nutrients (C). Data from tissue and soil samples taken in 2017 and 2018 from fields of NC Corn Yield Contest study participants at four growth stages.

Relative to the complete dataset, the LASSO approach resulted in stronger model performance when data were divided by either year or growth stage as well as when management practices were included. The differences in model strength between 2017 and 2018 indicated that yield was likely influenced by environmental conditions that differed greatly between years. When the LASSO computation was run by growth stage, the stronger models were those that included the two-way interactions. Model performance at the early vegetative and maturity growth stages (< V6 and R6) was often better than at the middle two growth stages (>V7 and VT-R1). Our dataset was too small to validate our models to elucidate predictive performance. However, the trends in the LASSO analysis regarding which explanatory variables were retained in the models most often were consistent with trends using the DRIS system relative to the nutrients furthest away from zero. Unlike DRIS, the LASSO regression allowed us to include soil characteristics and management practices to determine their potential influence on yield. LASSO can be useful as a tool to determine the variables with the strongest influence and point out possible factors limiting yields.

Comparing the LASSO outcomes with DRIS, tissue Mg and S were nutrients indicated in both models as factors likely affecting yield. However, unlike DRIS, the LASSO models indicated that tissue N, P, and B also influenced yield. The measure of tissue nutrients is important, but it only indicates after the fact what the plant accumulated. It is equally important to determine: nutrient availability, soil parameters that influence nutrient availability, and management practices that contribute to yield. Therefore, LASSO provides a robust tool to investigate multiple variables simultaneously and determine which are most important. LASSO does not lead to a

diagnosis of deficiency or quantify a prescriptive solution. However, it does elucidate what variables likely had the largest influence on yield, leading the way to finding a solution.

Machine Learning Performance

The dataset scenario influenced the performance among all algorithms for predicting the yield classes: low (< 13.5 Mg ha^{-1}); medium (13.5 to 15 Mg ha^{-1}); and high (> 15 Mg ha^{-1}). The complete tissue and soil dataset collected in 2017 and 2018 (T&S) resulted in a lower correct-classification rate compared to any of the 2017 datasets scenarios (Table 2-9). Figures 2-12, 2-13, 2-14, and 2-15 show the change in percent correct classification by each algorithm as a function of the number of model variables; each figure represents a different dataset scenario. When not taking into account the number of explanatory variables within the model, the highest percent correct classifications for the T&S scenario were 87% and 84% for the six-fold and percent split validation techniques, respectively. However, these models contained ≥ 22 explanatory variable (Figure 2-12 A and B). The highest classification rates for the 2017 dataset scenario with tissue and soil parameters (2017 T&S) were 92% and 89% for the K-fold and percent split validation techniques, respectively (Figure 2-13 A and B); these models contained 23 and nine explanatory variables, respectively. On the other hand, the 2017 data with the tissue, soil, and management practices (2017 T&S + management) had 100% correct classification for both cross-validation techniques; these models contained between eight and 26 explanatory variables from the K-fold (Figure 2-14A); and between five and 26 explanatory variables for the percentage

Table 2-9. Machine learning performance statistics for best-fit models from support vector machine, artificial neural network, and decision tree algorithms for determining yield class: low, < 13.5 Mg ha-1; medium, 13.5 to 15 Mg ha-1; and high, >15 Mg ha-1, using two cross-validation methods: K-fold with six folds (KF), and percentage split (PS) (66%); and four combinations of explanatory variables: complete tissue and soil data collected in 2017 and 2018 ("T&S"); 2017 tissue and soil data ("2017 T&S"); 2017 T&S plus management practices; and 2017 management practices alone. Data collected in 2017 and 2018 from fields of NC Corn Yield Contest study participants.

	Support vector machine		Artificial neural network		Decision tree	
	Validation method					
Variable combination	KF	PS	KF	PS	KF	PS
	Percent correct classification					
2017 & 2018 tissue & soil (T&S)	66	50	79	75	79	67
2017 T&S	79	71	88	80	82	84
2017 T&S + management	85	71	98	100	98	94
2017 Management only	76	69	92	93	93	93
	Kappa statistic					
2017 & 2018 tissue & soil (T&S)	.48	.29	.68	.63	.68	.50
2017 T&S	.68	.57	.83	.70	.72	.75
2017 T&S + management	.77	.57	.97	1.00	.97	.91
2017 Management only	.64	.51	.88	.90	.89	.90
	AUROC (Area Under the Receiver Operating Characteristics)					
2017 & 2018 & soil (T&S)	.754	.673	.916	.893	.849	.766
2017 T&S	.870	.828	.962	.949	.881	.895
2017 T&S + management	.915	.887	.994	1.00	.987	.956
2017 Management only	.860	.786	.994	1.00	.989	.992

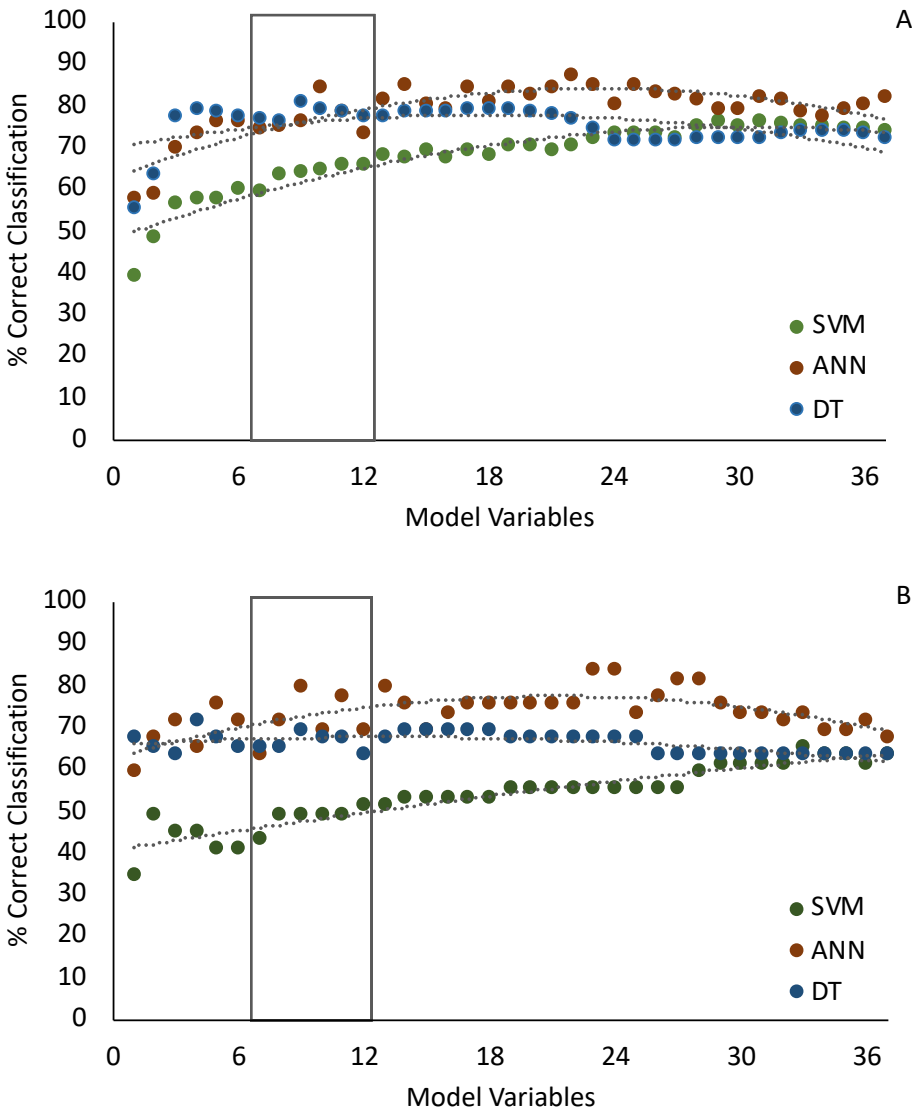

Figure 2-12. Algorithm performance for the complete tissue and soil dataset collected in 2017 & 2018 as the percent correctly classified into low, medium, and high yield classes by three machine learning algorithms: support vector machine (SVM); artificial neural network (ANN); and decision tree (DT): as a function of the number of explanatory variables. Data are shown by the machine-learning cross- validation method: K-fold with 6 folds (A) and 66% percent split (B). The box indicates the best-fit models for the scenario, which included between seven and twelve explanatory variables.

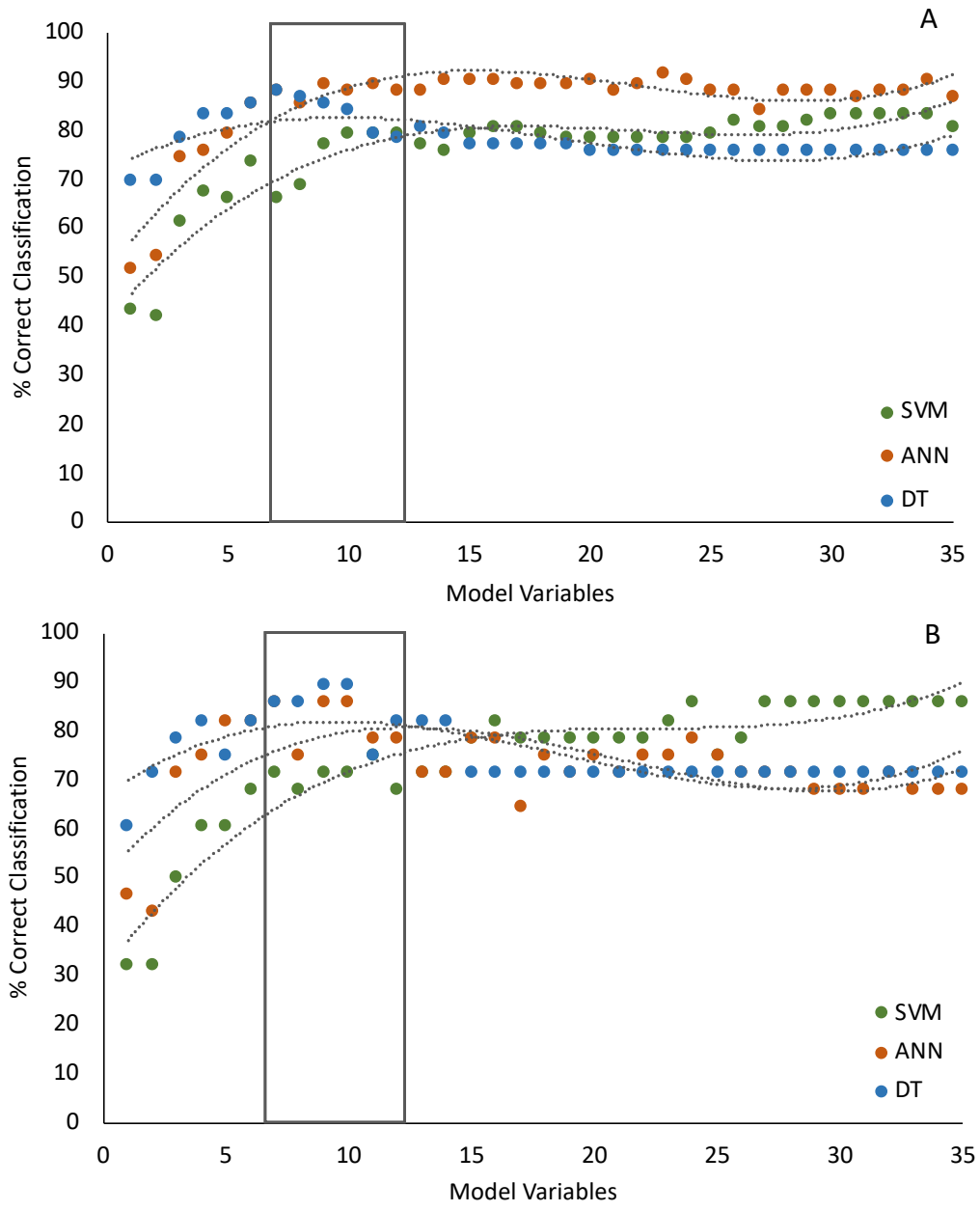

Figure 2-13. Algorithm performance for the dataset scenario containing 2017 tissue and soil data as the percent correctly classified into low, medium, and high yield classes by three machine learning algorithms: support vector machines (SVM); artificial neural networks (ANN); and decision trees (DT): as a function of the number of explanatory variables. Data are shown by the machine-learning cross-validation method: K-fold with 6 folds (A) and 66% percent split (B). The box indicates the best-fit models for the scenario, which included between seven and twelve explanatory variables.

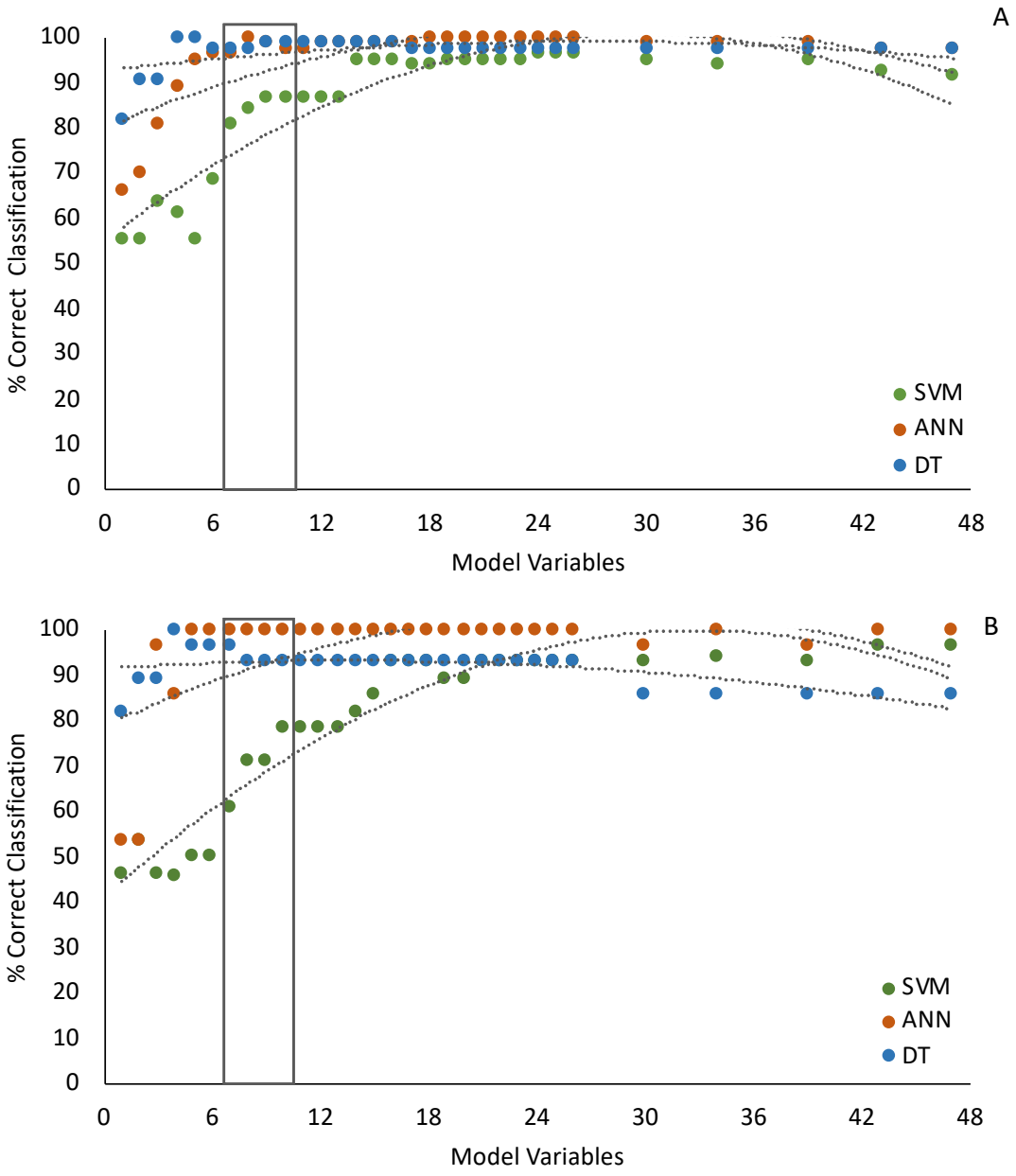

Figure 2-14. Algorithm performance for the dataset scenario containing the 2017 tissue and soil data plus management practices as the percent correctly classified into low, medium and high yield classes by three machine learning algorithms: SVM (support vector machine); ANN (artificial neural network); and DT (decision tree): as a function of the number of explanatory variables. Data are shown by the machine-learning cross-validation method: K-fold with 6 folds (A) and 66% percent split (B). The box indicates the best-fit models for the scenario, which included between seven and ten explanatory variables.

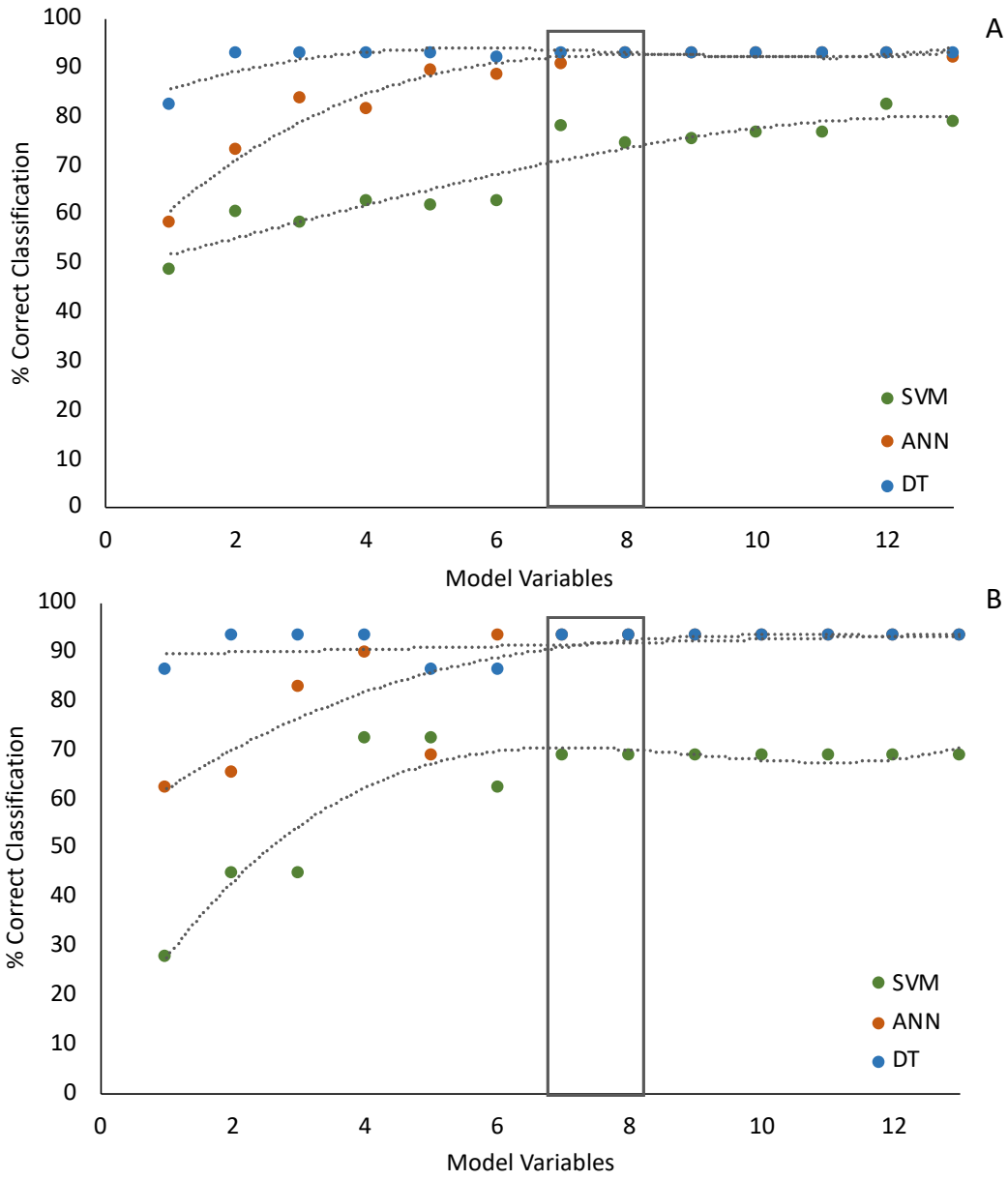

Figure 2-15. Algorithm performance for the dataset scenario with 2017 only management practices as the percent correctly classified into low, medium and high yield classes by three machine learning algorithms: SVM (support vector machine); ANN (artificial neural network); and DT (decision tree); as a function of the number of explanatory variables. Data are shown by the machine-learning cross-validation method: K-fold with 6 folds (A) and 66% percent split (B). The box indicates the best-fit models for the scenario, which included between seven and eight explanatory variables.

split validation techniques (Figure 2-14B). In at least one algorithm, the 2017 scenario with management practices only (Figure 2-15) performed better than the scenarios without management practices (Figures 2-12 and 2-13) but not as well as the 2017 scenario that contained all measured variables (Figure 2-14 A and B).

Classification performance differed by algorithm type and the model cross-validation method (Table 2-9). The ANN always had a higher rate of correct classifications compared to SVM, and was equal to or better than DT 75% of the time (Table 2-9). The DT had high classification rates with the least number of explanatory variables 88% of the time (Figures 2-12 to 2-15). The SVM had the lowest performance of all the algorithms, regardless of the scenario.

The kappa statistic and AUROC score are a measure of precision in algorithm classification. The kappa statistic with ANN was good to optimal at .63 to 1, and the AUROC score was very good to optimal at .89 to 1 (Table 2-9). The kappa statistic with SVM was fair to good (.29 to .77) while the AUROC score was moderate to very good (.673 to .915). With one exception (kappa statistic = .50), DT had a kappa statistic that was moderate to very good (.68 to .97) and an AUROC score that was good to very good (.77 to .99).

Our goal was to determine what factors most likely influenced yield by reducing the size of the model to improve interpretability. The SVM algorithm does not perform well when the dataset contains a high amount of noise in the data, which could be irrelevant data or data that is not easily interpreted (Nettleton, Orriols-Puig, & Fornells, 2010). This may be the reason that the SVM classification performance increased as the number of explanatory variables increased (Figures 2-12 to 2-15). The SVM algorithm uses support vectors to differentiate between classes, and more model variables might have compensated for the higher noise/error likely due to

environmental factors that also influenced yield. However, as our main goal was to seek a smaller model, this type of dataset appeared better suited to ANN or DT model algorithms, as these algorithms reduced model complexity by decreasing the number of variables and improving the ability for interpretation. A comparison of the validation methods showed that the K-fold validation yielded higher correct-classification rates compared to the 66% split method in most cases. This is to be expected from a relatively small dataset, as the K-fold method shuffles though all of the data K-1 times, whereas the percent split method holds out a random sample percentage to use in model development just once. The poor algorithm performance of the T&S scenario was likely due to differences in environmental factors between the two years. Environmental conditions can greatly influence plant growth, but were not quantified in our study; therefore, using a dataset from a single year eliminated the environmental differences that occurred between years and improved the model performance.

When the data were randomized per the negative control method, each of the three algorithms were unable to predict yield class for any of the four scenarios (Table 2-10). The mean correct classification of the randomized dataset by SVM, ANN, and DT ranged from 28.0 to 34.3%, which was similar to what could be expected from random chance when classifying the respective 3 (33%) classes. Kappa statistics ranged from -0.07 to -0.02 and AUROC ranged from 0.458 to 0.481. A negative Kappa statistic and an AUROC of < 0.5 indicated that the performance of the negative control was worse than what would be predicted by random assignment. Therefore, the performance of the three algorithms for all scenarios was better than random assignment, indicating that true learning occurred.

Table 2-10. Negative control outcomes of the machine learning algorithms: support vector machine (SVM), artificial neural network (ANN), and decision tree (DT), generated from 10 randomized datasets with the class value randomized within each. The performance statistics are reported as the percent of correct classification, Kappa statistic, and area under the receiver operator curve (AUROC) and shown by the cross-validation methods of K-fold cross-validation (KF) (six folds) and percentage split (PS) (66%).

	SVM		ANN		DT	
Dataset Scenario	KF	PS	KF	PS	KF	PS
			% Correct Classification			
Complete Tissue and Soil	37	37	34	33	33	35
2017 Tissue & Soil (T&S)	34	28	31	30	33	31
2017 T&S + Management	37	33	34	30	35	30
2017 Only Management	33	33	30	30	32	32
			Kappa Statistic			
Complete Tissue and Soil	.04	.05	-.01	.02	-.01	.01
2017 Tissue & Soil (T&S)	-.03	-.07	-.04	-.04	-.02	-.02
2017 T&S + Management	-.01	.04	-.02	-.05	.01	-.06
2017 Only Management	-.05	-.12	-.07	-.09	-.06	-.03
			AUROC			
Complete Tissue and Soil	.491	.503	.493	.483	.492	.502
2017 Tissue & Soil (T&S)	.481	.458	.463	.479	.480	.468
2017 T&S + Management	.503	.497	.480	.482	.503	.455
2017 Only Management	.486	.495	.440	.460	.424	.477

Like with LASSO, choosing the ideal model is a balance between a high percent classification rate and a low number of explanatory variables. Considering the four dataset scenarios, the number of explanatory variables retained within the best-fit models ranged from seven to thirteen depending on the algorithm and cross-validation technique. The two scenarios with management practices (Figures 2-14 and 2-15) had better performance with a smaller number of variables than the scenarios containing only tissue and soil variables (Figures 2-12 and 2-13). The only variable retained in all three dataset scenarios that included tissue and soil data was percent humic matter (Table 2-11). Of the tissue parameters, the 2017 and 2018 T&S scenario retained tissue Mg and Mn, and the 2017 T&S scenario retained tissue Mg and Na.

Table 2-11. From the machine learning results, the ranking of yield model explanatory variables retained in the best-fit models using four combinations of explanatory variables (Table 8): complete tissue and soil data collected in 2017 and 2018 ("T&S"); 2017 tissue and soil data ("2017 T&S"); 2017 T&S plus management practices; and 2017 management practices alone. Ranking determined using a single-attribute evaluator feature selection function. The number of variables for the best-fit models ranged from seven to thirteen. Bolded explanatory variables appeared in more than one scenario. HM % = humic matter percent, CEC= cation exchange capacity, Ac= exchangeable acidity, BS% = base saturation, NBI Macro = nutrient balance index calculated from DRIS macronutrients, NBI All = nutrient balance index calculated from DRIS macro- and micronutrients, Growth stage = <V6, >V7, VT-R1, and >R6, year = 2017 and 2018.

Rank in model	Explanatory variable combination			
	2017 and 2018 tissue and soil (T&S)	2017 T&S	2017 T&S and management practices	2017 management practices
	Explanatory variable			
1	**Soil Mn**	**HM %**	**Fertilizer K**	**Fertilizer K**
2	**Soil Cu**	**CEC**	**Fertilizer N**	**Plant Density**
3	Year	**Soil Mg**	**Fertilizer P**	**Fertilizer N**
4	**Tissue Mg**	**Ac**	**Plant Density**	**Fertilizer S**
5	Tissue Mn	Soil Ca	**HM %**	Planting Date
6	Fe:Mn	Soil P	**Fertilizer S**	**Fertilizer P**
7	NBI all	**Soil Mn**	**Soil Mg**	**Tillage**
8	**HM %**	**Soil Cu**	**Tillage**	Manure
9	Soil Zn	**Tissue Mg**	**Ac**	**—**
10	Soil Class	pH	**CEC**	**—**
11	Growth Stage	Tissue Na	—	—
12	NBI Macro	Soil K	—	—
13	BS%	Soil Na	—	—

In contrast, the 2017 scenario that included the management practices did not retain any tissue parameters. Additionally, the DRIS NBI for macronutrients and for macro- and micronutrients together were retained in the complete scenario, indicating that these measurements were important variables for predicting yield. Soil Mn and Cu had the greatest model influence among the retained variables in the complete dataset scenario, and though both were also retained in the 2017 T&S scenario, they were ranked lower. Soil Mg was the only soil nutrient retained in the 2017 T&S scenario that was also retained in the 2017 scenario with T&S + management practices (Table 2-11). Cation exchange capacity and Ac were also retained in both 2017 scenarios. The variables with the highest model influence when the scenario included

management practices were the application rates of fertilizer N, P, and K, along with the planting density. The addition of the management practices increased the correct classification rate; however, management practices alone had a low classification rate.

Unfortunately, due to small sample numbers we were not able to analyze the 2018 data with management practices, therefore were only able to make inferences on the results of 2017. The results of the ML analysis indicated that management practices had a greater influence on yields than did the soil parameters we measured. Few individual tissue nutrients were among the variables most often retained in the best-fit models. However, variables derived from multiple tissue parameters, for example the NBI for macronutrients and NBI for micro-and macronutrient, did appear as influential variables in the model from the T&S scenario. By themselves, individual nutrient tissue concentrations can help guide in-season nutrient management, but they are more of a barometer of soil and environmental conditions affecting the plant and not as indicative of overall yield performance.

Conclusion

The DRIS method uses only tissue nutrient pairs for analysis of factors influencing yield. Thus, DRIS lacks the ability to compensate for environmental and genetic influences that can strongly affect nutrient uptake. On the other hand, LASSO and ML methods allow for a greater variety of explanatory variables such as soil parameters and management practices to be incorporated into yield models.

Regardless of the analysis method, the tissue nutrient indicated most often as influencing yield was Mg. Tissue S was an influential variable in the DRIS and LASSO methods but was not retained as an explanatory variable in any of the ML best-fit models. Our study indicated that the

addition of soil parameters to LASSO or ML analyses greatly improved model performance. The addition of management practices greatly improved the performance of the LASSO and ML models, indicating that relative to the soil and tissue parameters tested, management practices had the greatest influence on yield models in our study.

CHAPTER 3

Detecting Phenotypic Differences in Corn via Machine Learning of Diurnal Hyperspectral Imagery from an Off-Nadir View Mast-Mounted Camera

Abstract

Traditional field experimentation methods incorporating soil and crop science are often complex and labor intensive. Recently, incorporation of remote sensing hyperspectral cameras have facilitated the non-invasive collection of high-resolution spectral data providing greater insight into chemical and biological processes. Over the past decade, use of hyperspectral imaging (HSI) has increased rapidly in precision agriculture mainly due to the demand for more sustainable and efficient agricultural practices. Implementing a pole-mounted off-nadir view camera for image capture in varying directions throughout a field is an efficient method for diurnal and seasonal data collection. Such a system can be close to the canopy, operate semi-autonomously throughout the day over an entire growing season, and provide higher spatial and temporal resolution than imagery acquired from satellites or aerial platforms. However, the changing relationship between solar irradiance and off-nadir view angles throughout the day can reduce the image signal-to-noise ratio, which may influence the information that can be gained from an image. In addition, compiling and interpreting HSI data can be time consuming and difficult due to the large file size. Supervised machine learning is one method of analyzing the "big data" generated by HSI technology. Our objective was to study the application of hyperspectral imaging technology for detecting diurnal variation of canopy reflectance in a trial of two hybrid corn (*Zea mays* L.) varieties under two irrigation regimes (irrigated and non-irrigated). We implemented three machine learning algorithms: a support vector machine, an

artificial neural network, and a decision trees to analyze the diurnal changes in canopy reflectance to determine if image data could reliably classify the experimental treatments. We used three dataset scenarios: the complete dataset; the dataset divided by field (northeast and southwest fields); and by-hour datasets divided by the time of image capture at 11 and 40 minutes past the hour. The ML algorithms were able to distinguish between hybrids but not irrigation regimes, although the yield differences were between irrigation regimes and not hybrids. Classification performance was highest among images of the NE field which received forward scattering light, but the performance varied by the time of image capture. The highest correct classification for the NE field dataset was 89 and 91% from ANN with two cross-validation methods, but this increased to >80% for all algorithms when the data where divided by time of sensing for all but the 10:40AM, 2:40 PM, and 4:40 PM data. The explanatory variables with the most model influence were not uniform among datasets and changed depending on the time of image capture and field location. The NE field and hourly dataset scenarios showed that wavebands between approximately 440 to 500 nm were the variables with the highest influence on the ML algorithm models. When the dataset was divided by hour, the influence of the individual hyperspectral wavebands on the ML models in the NE and SE fields were similarly ranked from about 418 to 640 nm, opposing from about 640 to 700 nm, and again similar from about 700 to 798 nm. Our results suggest that when using an off-nadir hyperspectral camera it would be advisable to avoid view angles that would encounter back scattering light situations and thus provide greater reliability in using hyperspectral imaging to characterize crop canopies.

Introduction

It is projected that the global population will rise to nearly 10 billion by 2050 (United Nations Dept. of Economic Affairs, 2019). In tandem with the greater food demand from population rise is higher demand for animal protein and renewable energy sources like biofuel. These increased demands will place a substantial burden on worldwide agriculture production (FAO, 2012). Field corn (*Zea mays* L.) is an important commodity crop used for animal feed, biofuel, and for food and industrial products. Over the past three decades, corn production has increased significantly due to advances in breeding programs focused on hybrids with higher yields and increased pest and disease tolerance (Bender et al., 2013; Luetchens & Lorenz, 2018; Vyn & Tollenaar, 1998). However, continual increases in production are pressured by climate factors such as increases in annual temperature, especially during the cropping season, creating conditions for heat and water stress that may potentially reduce yields. Increased heat stress throughout the growing season has the potential to reduce corn yields at a time when increases in production from agriculture are most critical (Lobell et al., 2013; Welikhe et al., 2016).

To overcome the potential shortfall caused by environmental stressors, field research is needed to better understand the effect of diurnal changes on corn physiology. The resource and labor requirements to collect and quantify all the data needed for thorough crop and soil science field experiments are high and often times need to be pared down to stay within budget. In addition, field experiments present confounding factors such as weather that cannot be controlled and can make statistical analysis difficult. One of most efficient ways to understand these complex systems is to collect multiple types of data and use data science techniques to analyze and look for strong correlations within the data.

Recently, one of the keys components to understanding physiological responses of crops to environmental stress has been the use reflectance data gathered through the use of remote sensing images. Remote sensing cameras have allowed researchers to gather information about crop phenotypic and physiological differences using spectral reflectance over multiple wavelengths and determining relationships between the reflectance signature and the status of the crop. Remote sensing allows for non-destructive imaging of large areas of land in a short amount of time. Quantifying electromagnetic radiation from the crop and soil can be used to determine yield, nutrient and water status, crop stress, and other biological functions (Mulla, 2013). The spectral information can be analyzed using indices developed to determine various abiotic and biotic conditions within a crop. The data is often used to calculate predictive spectral indices that can be utilized in precision agriculture to determine the spatial variability of nutrient needs within a field and apply appropriate amounts only where required. These indices are calculated using specific combinations of spectral bands. They include various Normalized Difference Vegetation Indices (NDVI), the Optimized Soil Adjusted Vegetation Index (OSAVI), and the Photochemical Reflectance Index (PRI) (Bannari, Morin, Bonn, & Huete, 1995). These can help to determine crop status and then be used to extrapolate a potential yield (Gamon, Serrano, & Surfus, 1997; David J. Mulla, 2013).

The adoption of remote sensing technology has allowed for an increased amount of data collection, but traditional multiband remote sensing in the visible and near-infrared spectral range is limited by the number of wavebands (typically three to eight) that can be detected and analyzed. Hyperspectral sensors increase the amount of data collected by dividing this spectral range into hundreds of wavebands. Hyperspectral imaging systems quantify in a two-dimensional

raster array the light radiating from the objects being sensed. Typically, the light passes through an aperture and is projected onto a spectrograph focal plane. The incoming light is then dispersed by diffraction gratings onto a two-dimensional sensor array which records the information as an image slice. Each slice represents a single relatively narrow waveband, and these slices are compiled to create a hypercube. The hypercube is a 3-D representation of the 2D image, with the third dimension comprising the wavebands sensed. The result is a wavelength-intensity map of the area being sensed. The hyperspectral data can then be analyzed in relation to ground-truth data collected directly from the field. Hyperspectral imaging models have been shown to outperform multiband vegetative indices when predicting crop performance over time (Aguate et al, 2017). However, hyperspectral imaging can generate large amounts of data that can be difficult to interpret.

Hyperspectral remote sensing technology has been used for precision agriculture since the early-2000s (Erives & Fitzgerald, 2005; Yang, Prasher, Whalen, & Goel, 2002), and over the last decade the technology has advanced from sensors in airplanes and satellites to include sensors that can be mounted on unmanned aerial vehicles (UAV's), on tractor implements, and in a fixed position within a field (Mulla, 2013). Initially, satellite and airborne systems were relatively limited by data access and ground resolution. The UAV's can provide high-spatial-resolution imagery but can be restrictive due to the additional knowledge base and special license needed for operation. Field-based sensing systems permit crop monitoring over a long-term research study. This allows for image capture throughout a day, over many days, up to the entire growing season. In addition, a field-deployed hyperspectral camera can provide very high ground resolution especially compared to satellite imagery. This technology is not without its drawbacks,

and an off-nadir view camera, which is often used in these systems, introduces complexity to the image that can increase the error/noise within the data and increase the difficulty of gaining useable information from the image. Dynamic factors that may need to be considered include the effect of canopy temperature on reflectance, plant architecture, the temporal resolution, weather, and time of day when data is collected (Coops, Hermosilla, Hilker, & Black, 2017; Cure, Flagler, & Heagle, 1989; Roujean, Leroy, & Deschamps, 1992). All of these can affect reflectance, and consideration of how light and environmental conditions interact with differing plant physiology when utilizing this data can reduce misinterpretation.

One method of analyzing the "big data" generated by technology such as hyperspectral imaging is machine learning (ML). Machine learning has gained favor as a method to 'learn' from large datasets through the use of mathematical algorithms. In the last decade, ML has been implemented in the field of agriculture (Liakos et al., 2018). The ML programs are either supervised or unsupervised learning types, and have different learning models such as regression, clustering, and classification. One typical task for ML is to classify variables into categories using a portion of the dataset, known as a training set, to learn from and build a model. The model is then validated against another portion of the dataset, known as a validation set.

Our goal was to use ML on site-specific, hyperspectral imaging data to evaluate whether a ML algorithm could classify experimental treatments in a field study and determine the explanatory variables most influential in the model. These explanatory variables could then be compared to physiological changes within the corn crop over the growing season. We hypothesized that ML can correctly classify treatments using hyperspectral imaging data. We have the following objectives: 1) determine if a hyperspectral imaging system can detect diurnal

changes in field-grown corn consistently over the growing season; 2) determine if differences in hyperspectral leaf reflectance can distinguish two hybrids grown under two irrigation regimes; 3) determine if phenotypic differences within the canopy can be accurately determined using machine learning techniques.

Materials and Methods

Experimental Design

The hyperspectral image data was collected from a field study at the Cunningham Research Station in Kinston, NC (35º17'46"N 77º34'13"W) in the 2016 growing season. The field study consisted of two corn genotypes, N74R and N78S (Syngenta Biotechnology Inc.), grown under three different irrigation regimes. However, due to the limited field of view (FOV) of the hyperspectral image camera, scans were limited to the full irrigated and non-irrigated treatments. The experimental design was unconventional and created to favor the hyperspectral camera FOV and the associated sampling. A randomized complete block design was not possible due to the controllable dimensions of the irrigation system. The experimental design for the field consisted of four blocks: two fully irrigated blocks, and two non-irrigated blocks (Figure 3-1). Each treatment block was divided in half laterally and each half contained two plots each with a randomized hybrid. This resulted in each irrigation block containing two replicates of two hybrids randomly assigned.

Hyperspectral Camera Design

The hyperspectral imaging camera was designed by Dr. Michael Kudenov's lab in the College of Electrical and Computer Engineering at NC State University. The camera design comprised a paraxial objective lens, a slit oriented parallel to the *x*-axis placed directly in front of

the paraxial lens, a collimator made of an aspheric achromat, an N-SF11 high-dispersion equilateral prism, a CaF2 low-dispersion equilateral prism, a flat-fold mirror, a paraxial reimaging lens, a focal plane array, and a monochrome Sony ICX414 imaging sensor (Allied Vision Technologies Manta G-003) (Kudenov, Lowenstern, Craven, & LaCasse, 2017). The camera image sensor was 6.5 x 4.9 mm and comprised 656 x 1100 spatial pixels that were 100 μm^2 in size with 492 wavelength samples that had a relative focal plane array from 400 to 900 nm.

The camera was mounted on a rotating platform equipped with a motor to facilitate image capture from each side of the field and placed atop a 15.2 m (50 ft) mast located in the center of the field. The imagery was collected from ~6:10 AM to 8:40 PM at alternating 30-minute intervals: 30 min for one side of the field and 30 min for the other; thus each side of the field was imaged hourly. Images taken at 10 min past the hour were of the portion of the field SW of the camera, and images taken at 40 min past the hour were of the field half to the NE.

The field was planted on 18 Apr. 2016. Each hybrid treatment plot consisted of 12-rows and was 9.12 m wide and 18.24 m long. Within each plot were two adjacent sections of 12.16 m length flanked on each end with a 3.04 m long border. One section was designated for phenotyping and one for destructive sampling. Four plots made up each irrigation treatment block, and two hybrid genotypes were randomly assigned to the plots within each irrigation block (Figure 3-1). Between each irrigation treatment was a 9.12-by-18.24-m buffer. Irrigation was provided by a programmable linear movement overhead sprinkler system that had actuating bi-directional sprinkler head nozzles that were "on" and "off" orthogonal and traveled parallel to the irrigation block.

Figure 3-1. Diagram of experimental design for the field study located at Cunningham Research Station, Kingston, NC containing two irrigated and two non-irrigated plot, and two corn hybrids N78S and N74R.

Four reflectance tiles were positioned within the field at the four cardinal points, approximately 27.5 m from the centrally located pole mount for the camera (Figure 3-2). The reflectance tiles were constructed of 0.61 m square steel panels. Using flat sheen latex outdoor paint from Valspar (Valspar Corp., Minneapolis, MN), one half of each tile was painted to achieve 40% reflectance and the other half, 20% reflectance. These reflectance tiles, which will hereafter be referred to as reference tiles, served as reference points for data normalization during image

pre-processing, both to adjust for fluctuations in solar irradiance and for georeferencing the images. The primer used on the tiles corroded over time, and there was notable degradation by the end of the season. To account for this, a correction was made for each wavelength by generating a linear degradation model based on tile reflectance at the beginning and end of the season. The model was used to estimate the actual reflectance of the tiles for each day of the study.

Preprocessing of the raw image data was performed using MATLAB computer software (MATLAB version R2020a.1, The Mathworks, Inc. Natick, MA). First, spectral calibration of the raw data was performed using coefficients from laboratory measurements which allowed for the focal plane array to be mapped to the wavelength. Second, the datacubes were spatially corrected for rotation and orientation against the coordinates for the centroids of the reference tiles in each image. Next, each datacube was corrected for telluric H_2O and O_2 absorption lines, and low signal to noise regions of the final spectra were removed. Finally, reflectance spectra were calculated as a function of the reference tiles. Images captured prior to 8:10 AM and after 4:40 PM were not usable due to many light-saturated 2-nm bands because of the low sun angle. This resulted in 18 usable images per day, nine for each side of the field, and 54 for the 2016 season. After interpolation, the final spectra spanned 418 to 798 nm.

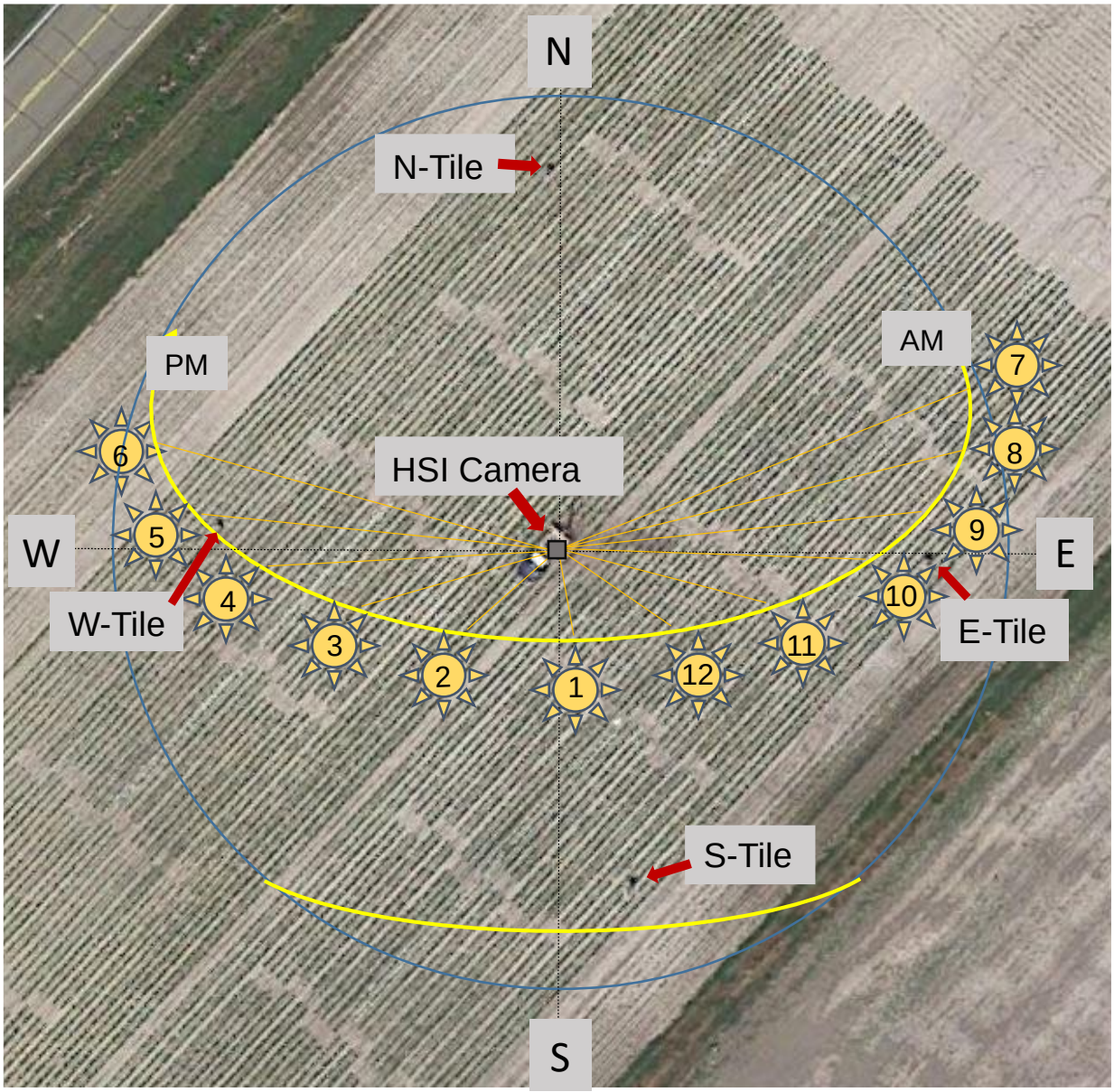

Figure 3-2. Diagram of Cunningham Research Station study field overlaid with labels for the four reflectance tiles, the hyperspectral imaging camera (HSI), and the path of the sun relative to the field.

MATLAB was used to extract data from datacubes captured between 8:11 AM and 4:40 PM from 20 to 27 July 2016, with exception of 22 July. Two of these days, 25 and 27 July, had significant cloud cover during a portion of the day; thus, these days did not include all timepoints in the extracted data. In total, 90 images were sampled. Twenty pixels were randomly selected within each hybrid plot along the centers of the four middle rows. The non-plant structures, such

as flags and soil, were avoided. An effort was made to choose sun-leaves, to avoid adding additional noise to the data from shade leaves, which reflect differently. The mean and standard deviation of each 20-pixel set for 192 two-nm bands from 418 to 798 nm of the visible and near-infrared spectrum were retained and compiled for further analysis.

Analyses

The compiled data samples were analyzed using multiple techniques. The reflectance spectra for the eight plots sampled were plotted to ascertain if the data was consistent with those of typical vegetation spectral reflectance signatures (Figure 3-3). The final step was to apply supervised machine learning algorithms to perform the task of spectral pattern recognition and determine what variables were most important in identifying a specified class. There were three class categories used: two irrigation treatment classes (yes or no); two hybrid classes (A=N78S and B=N74R); and four hybrid by irrigation treatment classes (N78S irrigated, N78S non-irrigated, N74R irrigated, and N74R non-irrigated).

Machine Learning Techniques

Supervised machine learning (ML) techniques were used to determine if there were any 2-nm spectral bands or other explanatory variables that were important in differentiating (classifying) the treatments. The ML was performed using Waikato for Knowledge Analysis (WEKA) Version 3.8.2 software (Department of Computer Science, University of Waikato, New Zealand). Three different ML model algorithms were used in variable selection for the predictive model: a type of support vector machine (SVM) known as sequential minimal optimization; an artificial neural network (ANN) algorithm called multilayer perceptron; and a decision tree (DT) algorithm known as J48 (Witten et al., 2011).

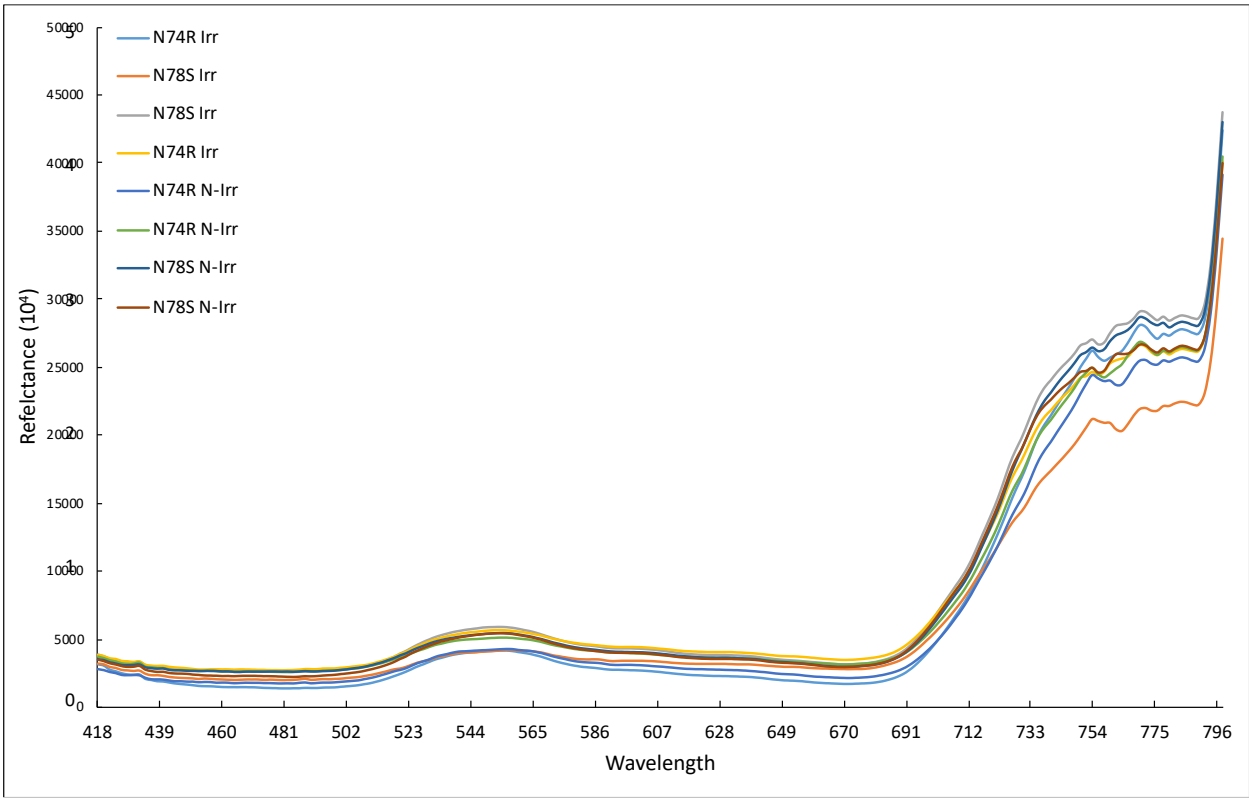

Figure 3-3. Mean reflectance signature (n = 20) for eight treatments plots in irrigated and non-irrigated blocks containing two hybrids (N78S and N74R) with two replications. Hyperspectral imagery was captured with a pole-mounted, off-nadir-view camera located in the center of a corn field study at Cunningham Research Station, Kinston, NC.

The algorithms were subject to cross-validation to evaluate machine learning performance (Smith & Frank, 2016). Two cross-validation approaches, percentage split ("PS") and stratified hold-out ("K-fold"), were used to improve orthogonal (independent) confirmation and overall performance accuracy. Orthogonal confirmation is used in machine learning to improve confidence in the outcome by using differing validation techniques to classify similar attributes. The PS cross-validation technique works by dividing the training and validation data by a set percentage. For each ML run, the split percentage was set at 66% of randomly selected data to be used as the training set and 34% as the test set. The K-fold cross-validation technique works by randomly dividing the data into a set number *k* of equal-sized groups or folds, where *k* − 1

folds of randomly ordered data were used as the training set for each iteration and one-fold was used as a validation set. This process is iterated *k* times. The K-fold was set at *k* = 8 folds for the complete dataset and for the northeast, and southwest datasets. Five folds were used for the dataset divided by time. The number of folds was chosen based on the number of sampled plots (8) for the former and the number of dates (5) within a dataset for the latter. Feature selection was used to reduce redundancy and irrelevant variables within the model, to help guide model reduction, and to determine ranking of variables during algorithm runs.

There were 202 possible explanatory variables considered for model inclusion: the 192 2-nm bands of the visible and near-infrared spectrum; NDVI and PRI; date and time of image acquisition; and at that time: sun azimuth, sun zenith, plot azimuth, plot to camera zenith, sun to camera zenith, and plot to sun azimuth (Figure 3-4). The dataset was partitioned into multiple dataset scenarios: the complete scenario comprised image data from both the NE and SW sides of the field; the NE and SW field scenarios comprised image data from the corresponding field half; and the hourly scenario comprised image data from a specific timepoint, one for each hour at 11 and 40 minutes past the hour, captured on dates from 20 to 27 July, 2016, excluding 22 July.

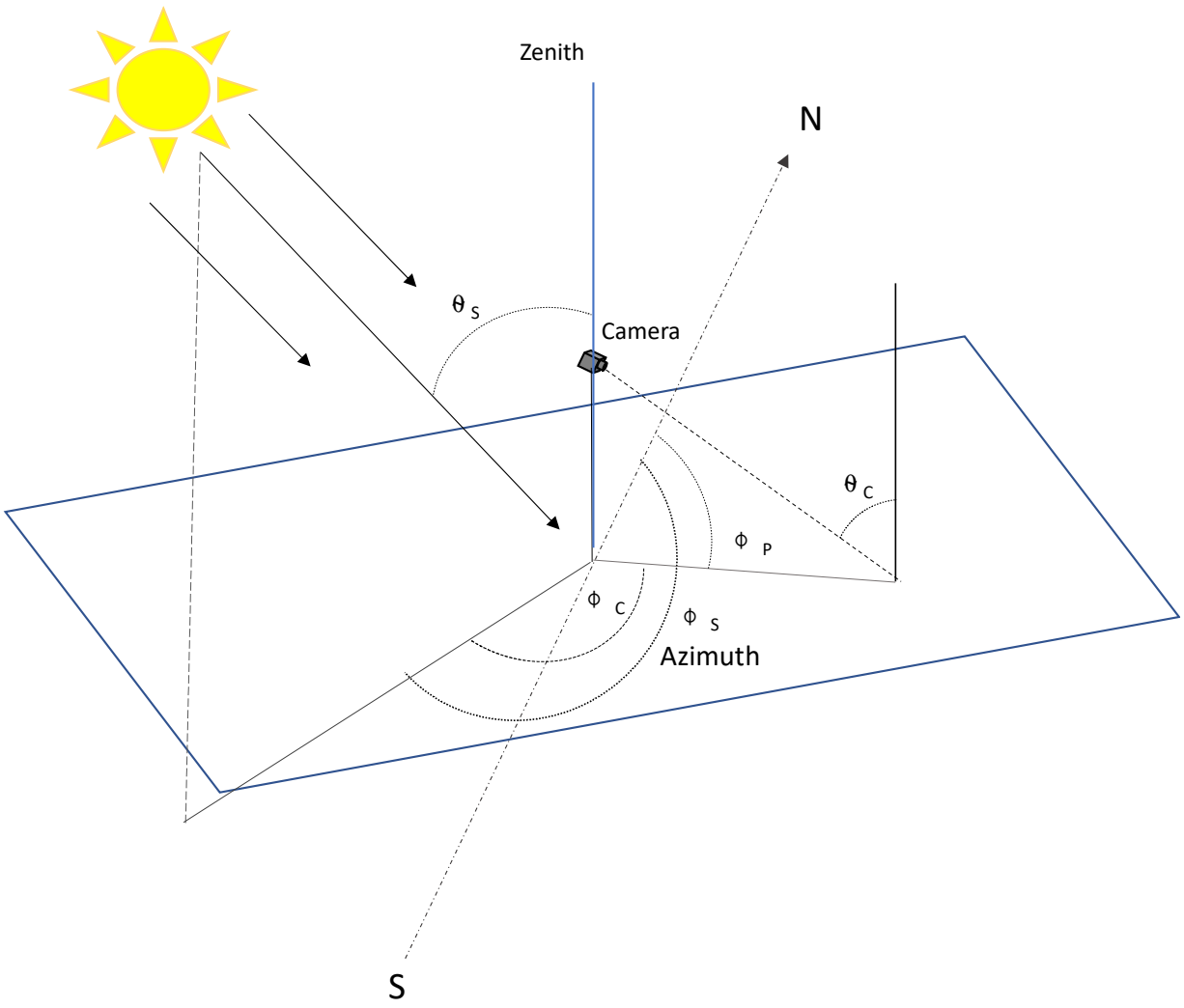

Figure 3-4. Diagram of field set-up for pole-mounted off-nadir view camera and angles determined at the time of image capture for: sun zenith (θ_S); sun azimuth (ϕ_S); plot azimuth (ϕ_P); camera to sun azimuth (ϕ_C); camera view zenith (θ_C).

The predictive models were then evaluated using three parameters: the percent of correctly classified observations; the Kappa statistic; and the weighted average of the Area Under the Receiver Operating Characteristic Curve (AUROC) for the classes. The percent of correctly classified observations is based on the ability of the algorithms to distinguish the designated classes. The Kappa statistic is a measure of model reliability used to assess whether the algorithm results, or agreement of the model to the data, did not occur just by chance (Vieira et al., 2010).

The Kappa statistic, which is related to the true positive rate, ranges from zero to one, where zero is agreement by chance and one is perfect agreement. The true positive rate is used in ML to refer to the number of observations that were actually identified correctly (Wang & Zheng, 2013). Therefore, a Kappa statistic of > 0.51 means greater than 50% chance of the correct classification between classes. The AUROC is related to the true negative rate and is a performance metric for how well the model distinguishes between classes. The values range from zero to one, and the higher the AUROC, the better the model is able to distinguish between classes.

Results and Discussion

The corn yield for the sensed plots differed by irrigation treatment, but no difference between hybrids was detected (Table 3-1). The performance of the algorithms in classifying irrigation and hybrids differed among the three dataset scenarios. Regardless the dataset scenario, all three algorithms classified irrigation correctly <50% of the time, which is considered no better than chance (data not shown). Similarly, the four treatment combinations (hybrids X irrigation) were classified correctly <50% of the time. Due to the experimental layout, including as explanatory variables the plot angles relative to the sun or the camera improved model performance only in classifying irrigation. This would be expected, as the irrigation regimes were located opposite from each other and had very distinct locations within the image. This would be the equivalent of providing the GPS coordinates for the location of a treatment and thus did not provide useful information in discerning irrigation regimes spectrally rather than geographically.

Table 3-1 . Mean corn grain yield (Mg ha-1) within hybrid and irrigation for the hyperspectral imaging study at Cunningham Research Station, Kinston, NC.

Treatment	df	Yield
Hybrid		Mg ha^{-1}
N74R (*n* = 8)		12.1a[a]
N74S (*n* = 8)		12.6a
Irrigation		
Yes (*n* = 8)		13.6a
No (*n* = 8)		11.1b
	ANOVA	
Hybrid	1	ns[†]
Irrigation	1	***

[a]Means within a treatment followed by the same letter are not statistically significant according to LSD (.05)
*Significant at the .05 probability level.
***Significant at the .001 probability level.
[†]ns, nonsignificant.

Water stress detection indices often incorporate wavelengths higher in the NIR than in our study, for example the water reflectance trough in the region of 950 to 970 nm (J. Penuelas, Pinol, Ogaya, & Filella, 1997). The maximum wavelength in our study was only 798 nm, which may help to explain the ML algorithms' poor performance in classifying irrigation. Therefore, only hybrid classification is presented hereafter.

Using the complete dataset scenario and without regard to the time of sensing, the best correct classification rates ranged from 72 to 77%, and the models with the best correct classification rates had from 103 to 193 explanatory variables (Table 3-2). The corresponding kappa statistics from these models ranged from .44 to .53, and the AUROC ranged from .721 to .851. These models were better than chance but only classified a specific hybrid correctly three out of four times, which would not prove useful if accuracy is the goal. Additionally, the models contained a high number of explanatory variables making interpretation difficult. It is likely that

the strong difference in illumination and viewing angles between the NE and SW fields contributed to the low classification because these variables were unaccounted for in the model development. Depending on the direction of the field relative to the camera, illumination could have been forward scattering in the case of the NE field or backscattering in the case of the SW field (Figure 2). Soares Galvão et al. (2008) found that backscattering directions resulted in better discrimination between soybean varieties over forward scattering directions in off-nadir image data. Dividing our dataset by field would control for this influential factor.

When the data were divided by field, the correct classification rates for SVM and ANN in the NE field were greater than their performance with the complete dataset; the classification performance of DT was little changed (Table 3-2). In contrast and relative to the complete dataset, the correct classification rate in the SW field was slightly lower for SVM and ANN and somewhat lower with DT (Table 3-2). The model sizes for SVM and ANN in the NE field were little changed from the complete dataset and ranged from 175 to 195 variables; whereas the model size with the DT varied greatly depending on the cross-validation technique (Table 3-2). In the SW field, DT had the models with the fewest explanatory variable for both validation techniques (Table 3-2). The kappa statistics for the models with the best correct classification ranged from .48 to .82 in the NE field and from .18 to 45 in the SW field (Table 3-2). The AUROC ranged from .732 to .973 in the NE field and from .558 to .816 in the SW field (Table 3-2).

Table 3-2. Among the models tested without regard to the time of sensing, the highest achieved correct hybrid classification rate (%), the number of variables, the kappa statistic, and the area under the receiver operator curve (AUROC) in the associated model for three dataset scenarios using three machine learning model algorithms: Support Vector Machine (SVM), Artificial Neural Network (ANN) and a Decision Tree (DT) with two cross-validation techniques: K-fold (eight-folds)(KF) and percent split (66%) (PS).

	Algorithm					
	SVM		ANN		DT	
	Cross-validation method					
Dataset Scenario	KF	PS	KF	PS	KF	PS
	% Correct classification (number of model variables)					
Complete Dataset	72 (183)	74 (173)	76 (193)	77 (103)	72 (193)	75 (113)
Northeast Field	84 (195)	85 (175)	89 (175)	91 (185)	76 (135)	74 (1)
Southwest Field	67 (195)	66 (104)	69 (94)	73 (84)	59 (1)	58 (1)
	Kappa statistic					
Complete Dataset	.44	.47	.52	.53	.44	.48
Northeast Field	.68	.69	.77	.82	.51	.48
Southwest Field	.34	.31	.38	.45	.18	.19
	AUROC					
Complete Dataset	.721	.737	.851	.789	.731	.756
Northeast Field	.840	.845	.956	.973	.732	.743
Southwest Field	.672	.658	.773	.816	.558	.593

Controlling for the influence of back scattering and forward scattering light in the dataset improved classification performance for SVM and ANN in the forward scattering direction, but reduced it in the back scattering direction. However, even with the improved accuracy in differentiating the hybrids, the models contained numerous variables. The large model size also indicated that there were not specific parts of the spectrum that were most important for differentiating between hybrids. There was a high degree of variability in spectral variables between hours and among days, which could have contributed to decreased classification performance.

As indicated by the coefficient of variation (CV,%), the variability of individual wavelengths was relatively high for the NE field compared to the SW field irrespective of hybrid or time of day (Figure 3-5). This is likely why the ML performance was generally higher for the SE field data. One explanation is that the bidirectional reflectance from the canopy was affected differently by the back scattering and forward scattering light. Sandmeier and Itten (1999) demonstrated that the view angle in the backscattering direction showed a more pronounced effect on reflectance in the visible part of the spectrum than in the NIR. At specific sensing times in our study, image hotspots resulting from the bidirectional reflectance function could have contributed to lower classification performance when the sun was lower in the sky or when the camera was sensing the antisolar point, which is the point directly opposite the direction of the sun form the observation point. Hotspots are phenomenon that occur in remote sensing when the angle of incidence and the view angle align in the back scattering direction causing bright spots to appear in the image (Jupp & Strahler, 1991).

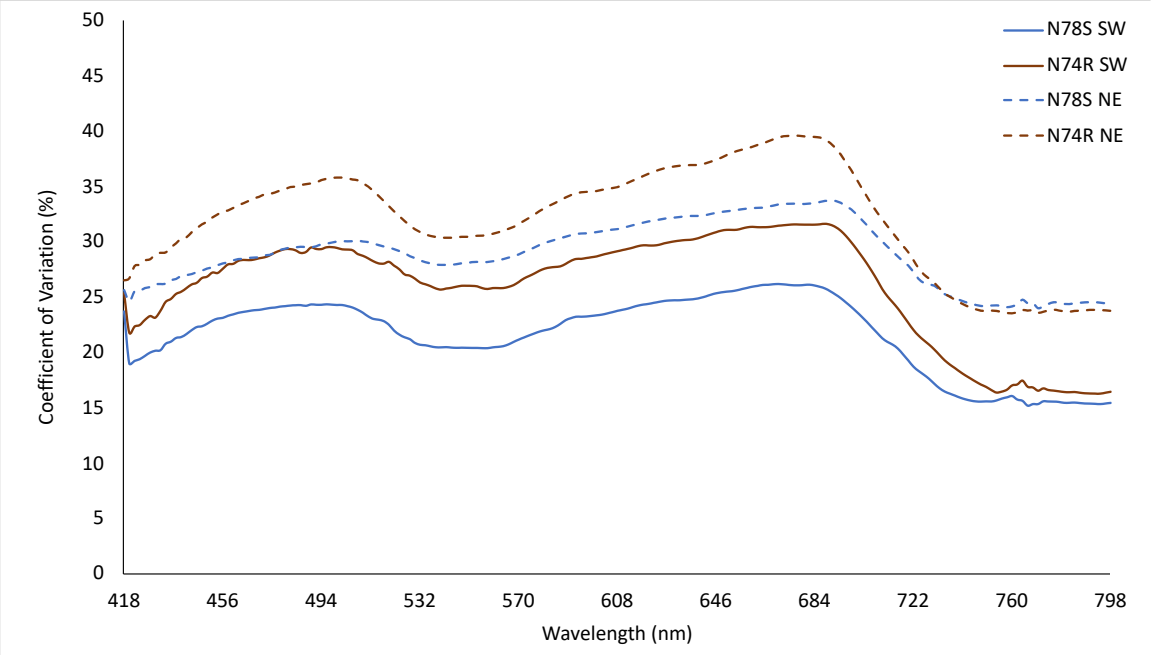

Figure 3-5. The CV (%) as a function of wavelength, field location (NE or SW), and corn hybrid (N78S and N74R) for imagery acquired from 8:11 to 4:40 PM on 20 to 27 July, 2016, with a pole-mounted off-nadir view hyperspectral camera.

The highest hybrid classification rate regardless of model algorithm was when the data were analyzed by time of sensing (Tables 3-3 and 3-4). The correct classification rates varied greatly throughout the day and ranged from 43 to 100% depending on the field location and time of sensing. In general, the correct classification rate was higher overall for the times sensed from the NE field than those from the SW field (Figure 3-6A and B, respectively). In the SW field (Table 3-3), the correct classification rate using DT was >80% only at 8:11 AM and 4:11 PM; whereas for ANN, the correct classification rate was >80% for at least one cross-validation technique at all times except 10:11, 12:11, 1:11, and 2:11 PM; while none of the times had >80% correct classification rate for SVM (Table 3-3). In contrast, the NE field data resulted in >80% correct classification from all algorithms regardless of cross-validation technique at all times except 2:40 and 4:40 PM. However, only the 2:40 PM data resulted in correct classification rates <80% with all three algorithms (Table 3-3).

Table 3-3. The highest achieved correct hybrid classification rates, the corresponding number of explanatory variables (in parentheses) in the associated models, the kappa statistic, and the area under the receiver operator curve (AUROC) from hourly dataset scenarios for the SW field using three machine learning algorithms: Support Vector Machine (SVM), Artificial Neural Network (ANN), and a Decision Tree (DT), all with two cross-validation techniques: K-fold (five folds) (KF) and percentage split (66%) (PS). Bolded numbers indicate correct classification rates > 80%

	Algorithms					
	SVM		ANN		DT	
Hour	KF	PS	KF	PS	KF	PS
	Correct Classification (%) & Model size					
8:11 AM	77 (92)	75 (62)	**92** (12)	**88** (12)	**85** (1)	**81** (1)
9:11 AM	65 (4)	71 (4)	**90** (36)	79 (16)	70 (96)	71 (1)
10:11 AM	63 (2)	43 (24)	60 (2)	50 (1)	55 8)	55 (8)
11:11 AM	73 (2)	64 (2)	**93** (20)	**93** (3)	78 (1)	57 (1)
12:11 PM	55 (192)	43 (192)	60 (192)	57 (192)	60 (192)	50 (7)
1:11 PM	43 (3)	43 (1)	68 (7)	72 (2)	53 (5)	42 (1)
2:11 PM	70 (10)	57 (1)	73 (2)	64 (1)	68 (1)	71 (44)
3:11 PM	73 (15)	69 (1)	**85** (192)	**88** (55)	71 (1)	69 (1)
4:11 PM	73 (5)	75 (10)	**83** (21)	69 (1)	**83** (2)	**88** (15)
	Kappa statistic					
8:11 AM	.54	.52	.83	.75	.71	.64
9:11 AM	.30	.44	.70	.59	.40	.46
10:11 AM	.25	.02	.20	.17	.10	.47
11:11 AM	.45	.31	.85	.85	.55	.22
12:11 PM	.10	-.12	.20	.16	.2	-.04
1:11 PM	0	0	.35	.44	.05	0
2:11 PM	.40	.19	.45	.34	.35	.44
3:11 PM	.46	.41	.71	.74	.42	.41
4:11 PM	.46	.52	.67	.74	.67	.75
	AUROC					
8:11 AM	.771	.778	.920	.968	.832	.833
9:11 AM	.650	.729	.898	.813	.686	.750
10:11 AM	.625	.511	.565	.778	.516	.778
11:11 AM	.725	.667	.948	.896	.725	.625
12:11 PM	.550	.438	.658	.563	.584	.479
1:11 PM	.500	0500	.633	.646	.533	.500
2:11 PM	.700	.604	.618	.917	.723	.729
3:11 PM	.729	.722	.901	.889	.635	.722
4:11 PM	.729	.778	.826	.934	.799	.873

Table 3-4. The highest achieved correct hybrid classification rates, the corresponding number of explanatory variables (in parentheses) in the associated models, the kappa statistic, and the area under the receiver operator curve (AUROC) from hourly dataset scenarios for the NE field using three machine learning algorithms: Support Vector Machine (SVM), Artificial Neural Network (ANN), and a Decision Tree (DT), all with two cross-validation techniques: K-fold (five folds) (KF) and percentage split (66%) (PS). Bolded numbers indicate correct classification rates > 80%.

	Algorithms					
	SVM		ANN		DT	
Hour	KF	PS	KF	PS	KF	PS
8:40 AM	**88** (2)	**88** (1)	**88** (1)	**94** (1)	**88** (1)	**94** (2)
9:40 AM	**90** (1)	**88** (4)	**90** (1)	**88 (**1)	**88** (1)	**88** (1)
10:40 AM	**83** (192)	79 (192)	**90** (192)	71 (192)	58 (1)	57 (44)
11:40 AM	**93** (182)	**100** (1)	**90** (162)	**100** (46)	**88** (1)	**100** (1)
12:40 PM	**95** (4)	**100** (1)	**95** (1)	**93** (2)	**95** (1)	**100** (1)
1:40 PM	**93** (192)	**93** (172)	**88** (172)	**86** (2)	**88** (1)	**86** (1)
2:40PM	73 (1)	63 (131)	77 (2)	69 (4)	77 (2)	75 (1)
3:40 PM	**90** (82)	**81** (1)	**85** (42)	**81** (1)	**83** (1)	**81** (1)
4:40 PM	67 (2)	**88** (1)	**88** (92)	**81** (2)	73 (1)	63 (7)
	Kappa statistic					
8:40 AM	.75	.75	.75	.87	.75	.87
9:40 AM	.79	.73	.92	.73	.75	.73
10:40 AM	.73	.70	.85	.59	.32	.34
11:40 AM	.85	1	.80	1	.75	1
12:40 PM	.90	1	.90	.86	.85	1
1:40 PM	.85	.86	.65	.71	.70	.71
2:40PM	.46	.20	.54	.31	.54	.43
3:40 PM	.71	.64	.71	.64	.67	.64
4:40 PM	.33	.71	.75	.56	.46	.20
	AUROC					
8:40 AM	.875	.900	.842	.967	.843	950
9:40 AM	.896	.896	.990	.883	.835	.867
10:40 AM	.925	.885	.980	.905	.653	.688
11:40 AM	.925	1	.905	1	.798	1
12:40 PM	.950	1	.953	1	1	1
1:40 PM	.925	.929	.873	.959	.832	.857
2:40PM	.729	.600	.710	.667	.707	700
3:40 PM	.854	.850	.837	.800	.779	.850
4:40 PM	.667	.833	.901	.700	.684	.567

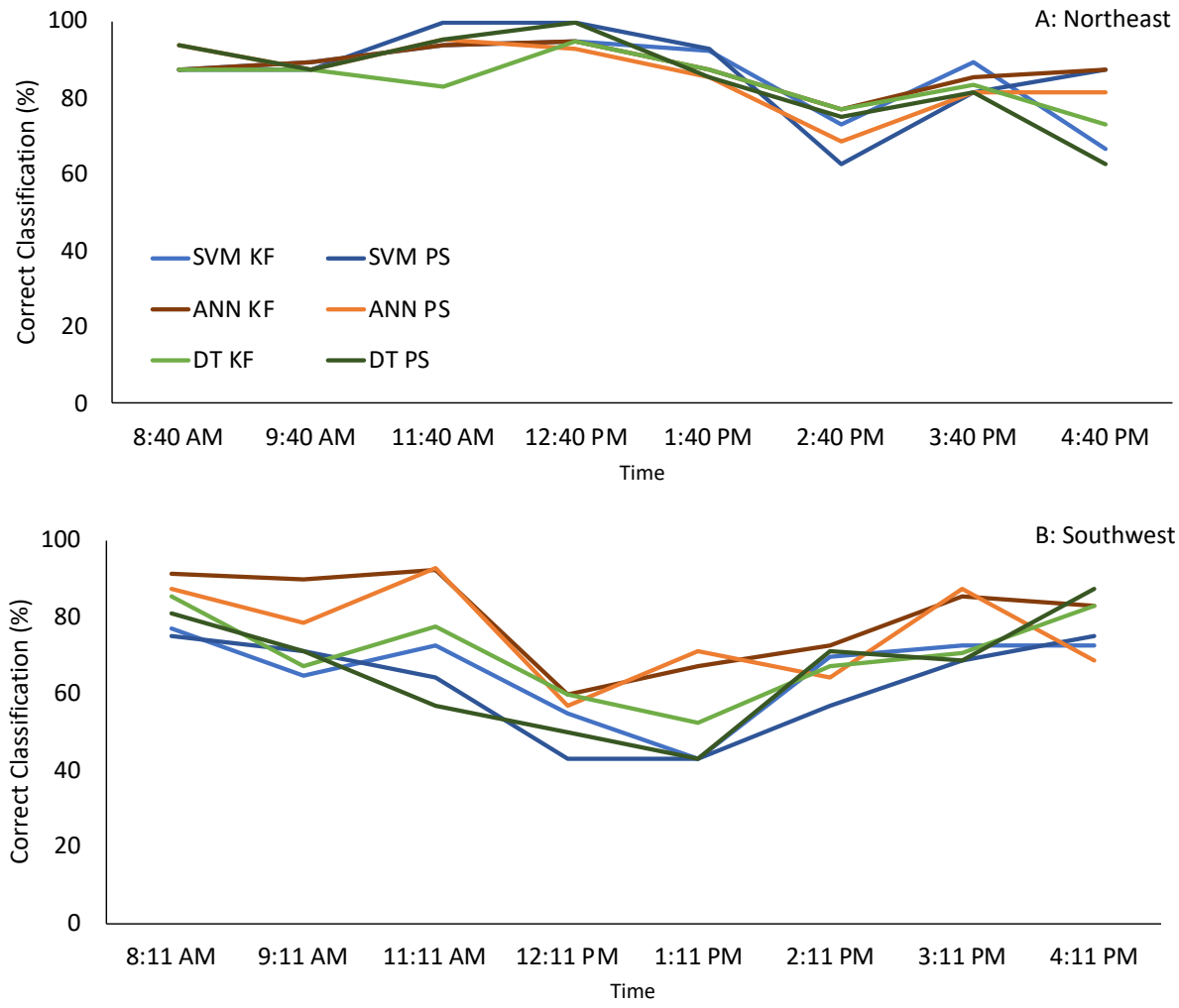

Figure 3-6. Highest hybrid classification rate achieved from by-hour dataset scenarios from three machine learning model algorithms: Support Vector Machine (SVM),), Artificial Neural Network (ANN), and a Decision Tree (DT), all with two cross-validation techniques: K-fold (five folds) (KF) and percentage split (66%) (PS), using hyperspectral image data collected hourly from 20 to 27 July 2016 in fields located in the SW (A) and NE (B) direction from the hyperspectral camera.

The goal of a best-fit model is one with a high classification rate and a small number of variables to minimize the complexity of interpretation and reduce overfitting of the model. Therefore, a model with very high number of explanatory variables, as was the case with all models from the complete dataset scenario and most from the NE and SW field scenarios, is far too large to be useful. The sizes of the models with the best classification rates for the data

divided by time of sensing varied drastically, from including all explanatory variables used in the scenario to only one (Tables 3-3 and 3-4). For the SW field data, there was only one time of day, 8:11 AM, when more than one algorithm had a classification rate >80% and a model containing < 20 explanatory variables (Table 3-3). Conversely, for the NE field data, 63% of the models had a correct classification rate >80% and < 20 explanatory variables; about two-thirds of those models had only one explanatory variable (Table 3-4).

The ranking of the influence of the spectral bands as explanatory variables when using the single-attribute selection operator varied greatly between NE and SW field dataset scenarios, and among the hourly datasets (Figure 3-7). The lower its rank value, the more influence a variable had on classification performance. For the NE field and hourly dataset scenarios, the wavelengths between ≈ 440 to 500 nm were the explanatory variables with the highest influence on the ML algorithm models (Figure 3-7A). However, this was not the case in the SW field and hourly dataset scenarios (Figure 3-7B). In fact, the NE and SE field datasets scenarios showed opposing trends in variable ranking from approximately 440 to 668 nm. However, when the dataset was divided by hour, the NE and SE field rankings were similar from about 418 to 640 nm, opposing from about 640 to 700 nm, and again similar from about 700 to 798 nm. Additionally, the spectral bands between ~450 and 460 had the lowest variability in rank among the by-hour dataset scenarios, indicating that these bands were less influenced by diurnal changes in irradiance angle.

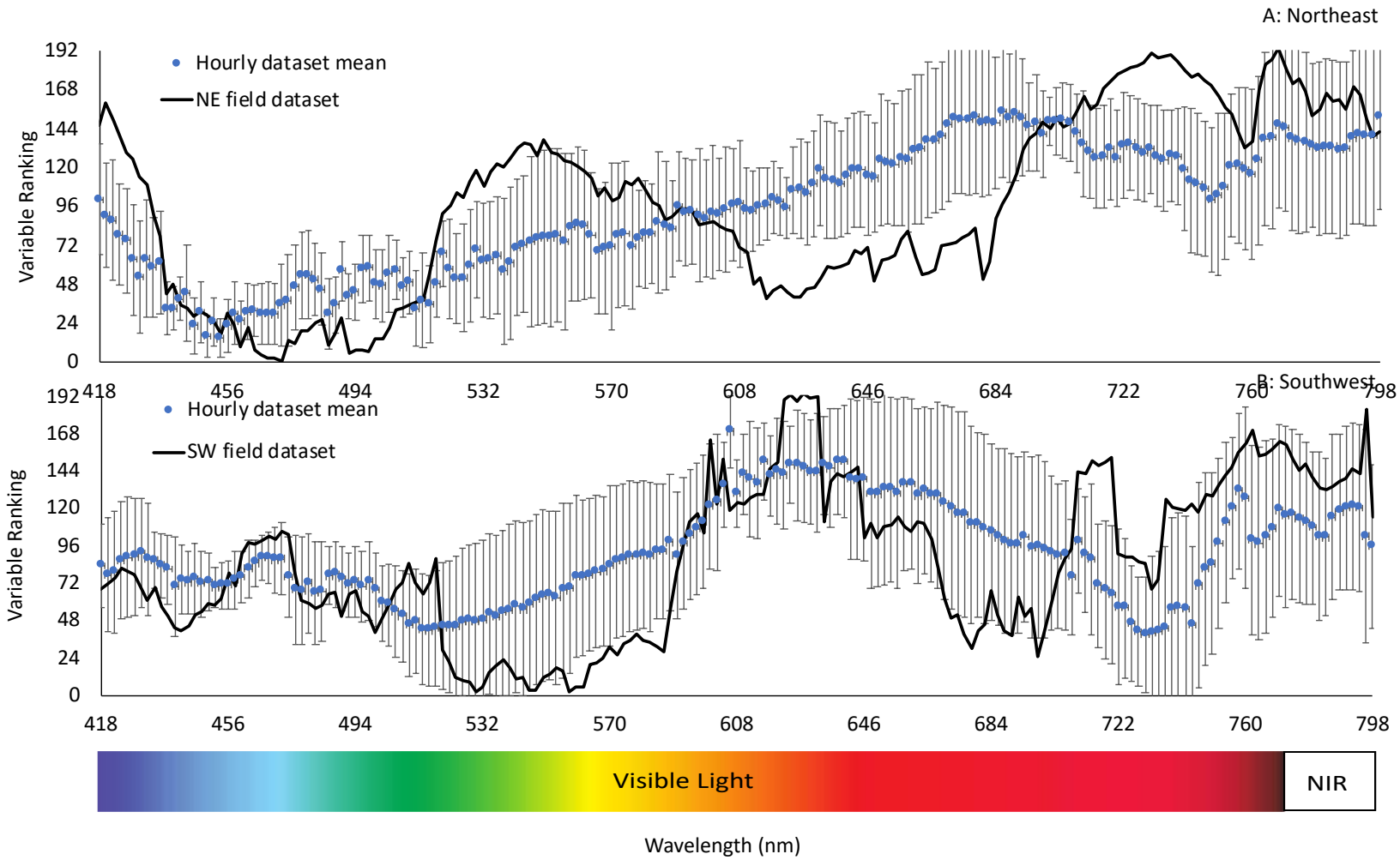

Figure 3-7. The ranking of the influence of the wavelengths as explanatory variables in classifying hybrids for NE (A) and SE (B) field scenarios (black line); and the mean rank of influence among the hourly dataset scenarios (blue dots) for 40 minutes after the hour (A) and 11 minutes after the hour (B). The lower the rank value the more influence the variable was to classification performance. Error bars are ± one standard deviation.

The PRI and NDVI were included as explanatory variables in the complete and the NE and SW field dataset scenarios. Both PRI and NDVI differed by time of image capture (Figure 3-8 and 3-9), but only NDVI differed between hybrids; there was no hybrid x time interaction for either PRI or NDVI (Table 3-5). That NDVI differed by hybrid indicates that it alone could discriminate between them. The PRI was eliminated from the models on the first run in all three scenarios, as it was ranked as having zero merit in the algorithms (not shown). Though there was no detectable difference in PRI between hybrids regardless of field location or time of day, there were subtle differences in PRI by time in the SW field (Figure 3-8), where PRI peaked around mid-day (Figure 3-8B). Hilker et al. (2008) found that 60% of the changes in PRI from a tree canopy could be explained by directional reflectance effects rather than physiological changes. Due to the relatively small changes in reflectance signals when using PRI on a landscape scale, it has been suggested that careful attention be paid to the effects from canopy structure (Josep Penuelas & Filella, 1998).

When first conceived, PRI was formulated with data collected at the leaf scale using reflectance at 531 and 570 nm (Gamon et al., 1997). It is possible that the PRI formulation should be adjusted to incorporate different wavelengths when working at the canopy scale, as was the case with our study. For a rice (Oryza sativa) canopy, Inoue et al. (2008) found that using 550 rather than 570 nm as the reference wavelength resulted in a stronger correlation with radiation use efficiency.

In contrast to PRI, NDVI was the highest ranked variable in the complete and NE field dataset scenarios but was eliminated from the SW dataset scenario on the first run (not shown). There are different formulations of NDVI, each using slightly different regions of the visible (red)

and near-infrared parts of the spectrum, and each with strengths and weakness depending on their application. Choosing the optimum formulation is somewhat dependent on the spectral resolution of the sensor, for example two broad bands over a wide range of wavelengths versus two narrow hyperspectral bands within the same range, and on the application of the index.

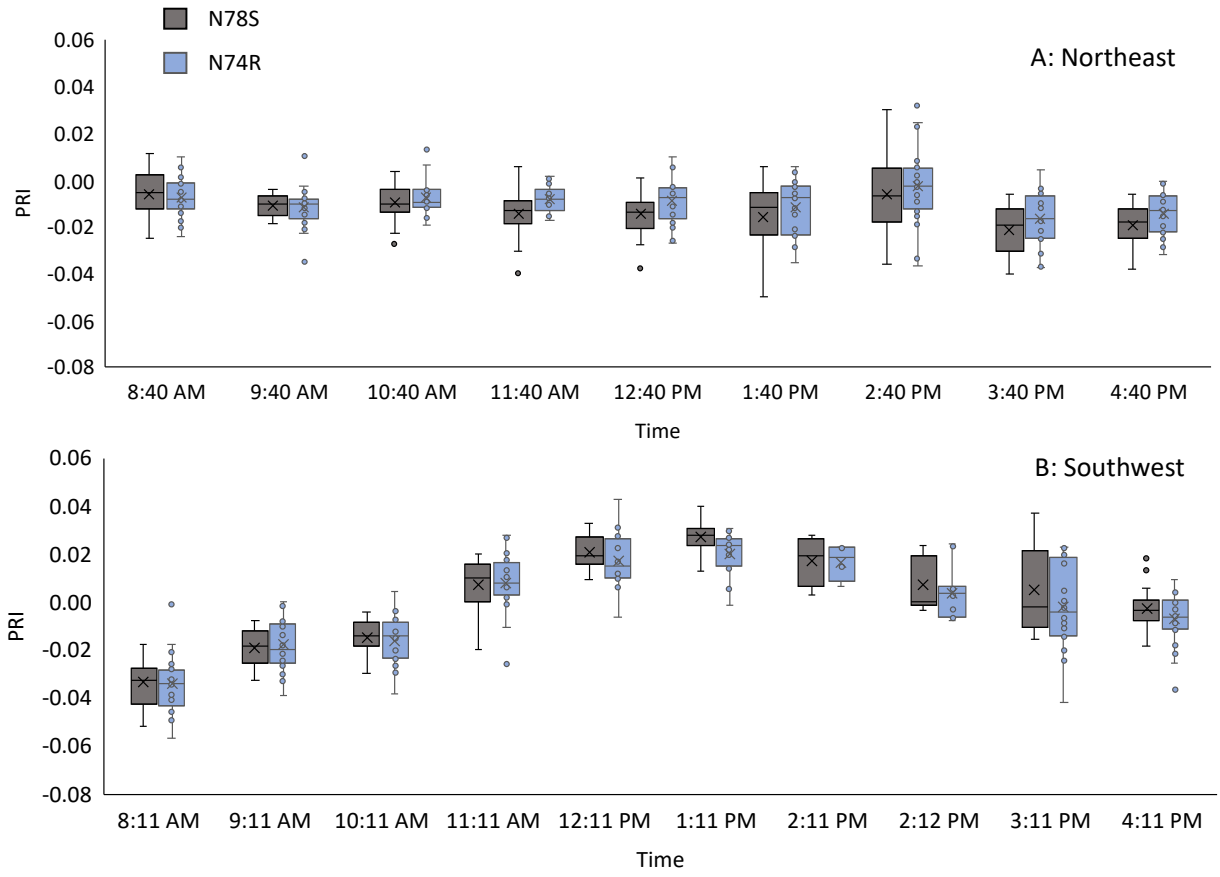

Figure 3-8. The photochemical reflectance index (PRI) calculated from hourly hyperspectral imagery acquired from 20 to 27 July 2016 of fields facing NE and SW planted with two corn hybrids (A=N78S and B=N74R) with two irrigation regimes (irrigated and non-irrigated). The box and whisker plots show the extreme upper and lower values excluding outliers and a box representing the inner 50% of the values. The "X" in the box is the mean and the mid-line the median; the upper and lower parts of the box are 25% of the values above and below the median

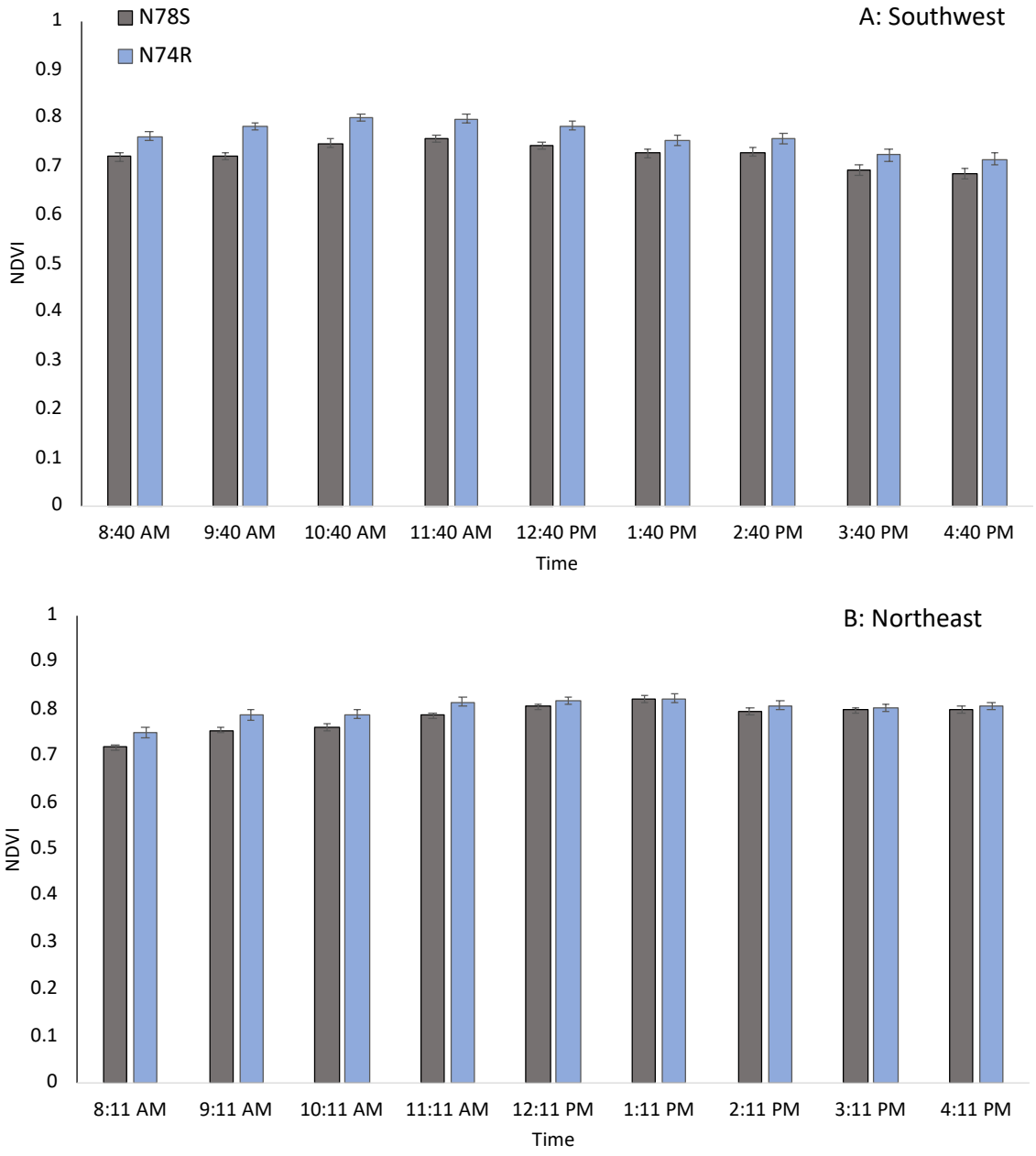

Figure 3-9. The normalized difference vegetation index (NDVI) calculated from hourly hyperspectral imagery acquired between 20 and 27 July 2016 of fields in the SW (A) and NE (B) directions planted with two corn hybrids (N78S and N74R) with two irrigation regimes.

We used NDVI wavebands of R650 (red) and R750 (NIR) based on a previous study using these red and NIR bands (Metternicht, 2003). It is possible that different wavebands or a wider band of the spectrum would have improved classification performance. A study by Schlerf et al. (2007) which used broader bands than 2 nm in the red and NIR parts of the spectrum showed

that NDVI was more highly correlated with tree stand attributes in the off-nadir view position, especially in the forward scattering light direction, but that this was influenced by the tree species. The irradiance reflected from a vegetative canopy is bidirectionally reflected and is expressed as a bidirectional reflectance distribution function (BRDF) but changing the view geometry relative to the plant would likely influence this function due to differences in forward and back scattering light (Figure 3-10).

Incorporation of the BRDF into our analysis was beyond the scope of our study, but would have likely improved the overall model performance, especially when using the complete dataset scenario. However, we did attempt to account for variability from view geometry changes by Including the angles for the azimuth and zenith pertaining to the sun, camera, and plot within the image. However, this did not improve the classification performance when differentiating between hybrids. The use of a radiation transfer function to account for the changes in view angle from the off-nadir view camera would likely have improved the performance outcome (Jacquemoud, Bacour, Poilve, & Frangi, 2000). In addition to changes in the viewing geometry, canopy architecture and leaf structure can also influence reflectance.

Table 3-5 . Mean photochemical reflectance index (PRI) and normalized difference vegetation index (NDVI) calculated from hyperspectral image data sensed between 20 and 27 July, 2016 with a pole-mounted off-nadir view camera for two corn hybrids captured from a northeast and southwest field at 11 and 40 minutes past the hour between 8 AM and 4 PM.

	Vegetative index			
	PRI		NDVI	
	Field			
Hybrid	Northeast	Southwest	Northeast	Southwest
N78S	-0.01	0.01	0.72	0.78
N74R	-0.01	-0.01	0.76	0.80
	ANOVA			
Hybrid	ns[†]	ns	**	**
Time	***	***	***	***
Hybrid x Time	ns	ns	ns	ns

**Significant at the 0.01 probability level. ***Significant at the 0.001 probability level.

[†]ns, nonsignificant.

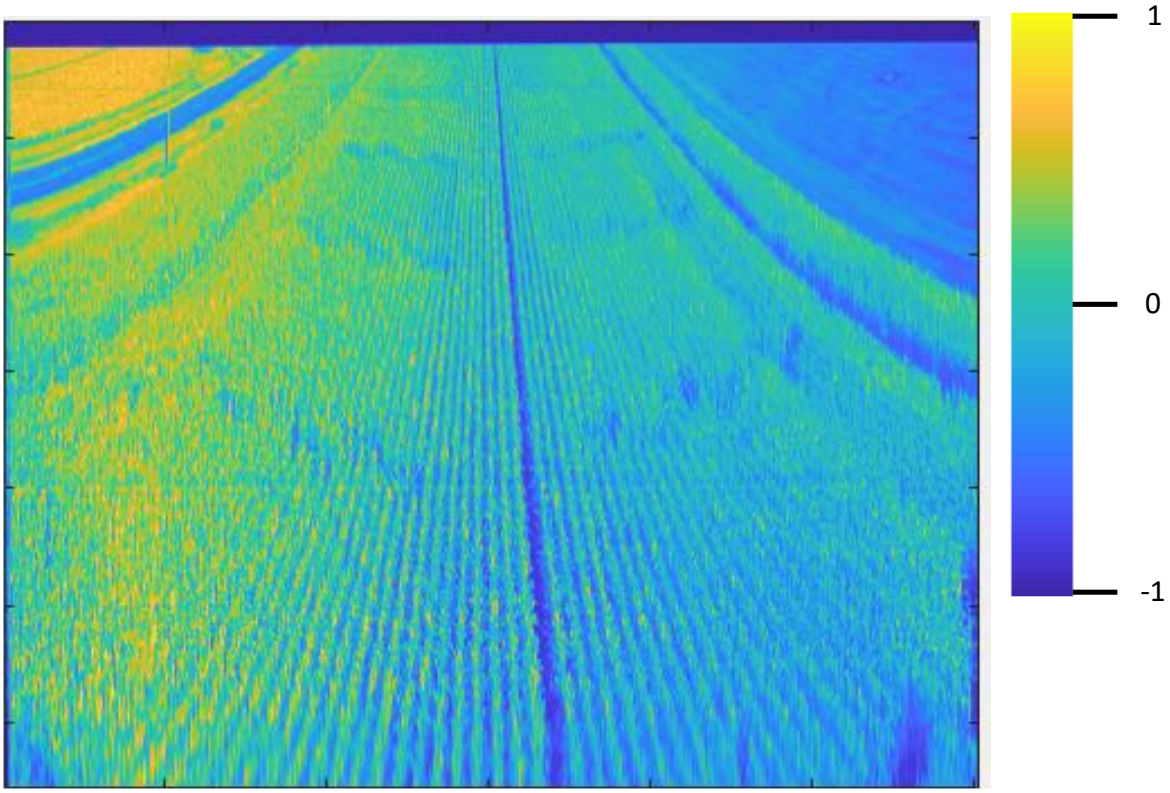

Figure 3-10. Hyperspectral image from MATLAB output showing NDVI displayed in a Parula colormap. Image was captured at 8:11 AM on 23 July 2016 in the back-scattering direction showing visible sign of the influence of view angle on NDVI as evidenced by the brighter plots located on the left side of the image.

The canopy structure differed between the hybrids in our study. The N78S hybrid had a more upright leaf architecture compared to the N74R hybrid and this difference could be observed in the images. Confounding the directional reflectance variability from the biophysical properties of the crop is the angular variability introduced from the off-nadir view sensor. Within the canopy, the resulting reflectance is dependent on leaf direction as well as light distribution but is further confounded by the view direction (Xie et al. 2017). An off-nadir view camera introduces the possibility of capturing more shade leaves, which differ in physiology to sun leaves and would likely absorb and reflect light differently. While we aimed to select leaves that were located near the top of the canopy, it is possible that leaves located in the lower canopy could have been represented in the initial pixel sampling in the pre-processing step. Lower leaves are not exposed to the same light regime as the sunlit leaves and have a different physiology. In addition to physiological difference between sun and shade leave, Xie et al., (2017) demonstrated that new corn leaves had drastically different reflectance than mature leaves.

Conclusion

We were unable to detect differences between two irrigation regimes using hourly hyperspectral imaging data sensed from an off-nadir view camera. We were able to correctly classify between two hybrids when machine learning was applied to the hyperspectral imaging data; however, the classification performance was higher in the NE field with forward scattering light than the SW field with back scattering light. Classification improved when analyzed by time of sensing, and overall, the correct classification rate was higher for the times sensed from the NE field than those from the SW field. Algorithm performance differed more in the SW field,

where ANN had greater classification performance than SVM and DT. In contrast, the NE field data resulted in >80% correct classification from all algorithms.

The rank of influence to the ML algorithm models varied greatly for spectral band explanatory variables. For the NE field and hourly dataset scenarios, the spectral bands between ≈ 440 to 500 nm were the explanatory variables with the highest influence, but this was not the case for the SW field data. The NE and SE field datasets scenarios showed opposing trends in variable ranking from approximately 440 to 668 nm. Interestingly, the spectral bands between ≈ 450 and 460 had the lowest variability in rank among the by-hour dataset scenarios, indicating that these bands were less influenced by diurnal changes in irradiance angle. NDVI differed between hybrids and was the variable with the highest rank of influence in the complete and NE fields datasets, while it was ranked as one of the lowest in the SE dataset scenario.

The use of a field-based hyperspectral imaging camera offers higher ground resolution than an air-based system and can collect crop data throughout the day and season. However, the use of an off-nadir view camera introduces complexity into the image processing and interpretation. The effects of changing solar irradiance angles, view angle, differing canopy architecture, and influence of sun and shade leaves influence the reflectance detected by the sensor, but not uniformly. Machine learning algorithms were able to best differentiate between hybrids in images with forward scattering light. Perhaps when using an off-nadir hyperspectral camera, it would be advisable to avoid view angles that would encounter back scattering light situations and thus provide greater reliability in using hyperspectral imaging to characterize crop canopies.

References

Austin, R., Osmond, D., & Shelton, S. (2019). Optimum nitrogen rates for maize and wheat in North Carolina. *Agronomy Journal*, *64*(9), 10.

Baker, W. H., Bell, P. F., Campbell, C. R., Cox, F. R., Donohue, S. J., Gascho, G. J., … Unruh, L. (2000). Reference sufficiency ranges for plant analysis in the southern region of the United States. *Southern Cooperative Series Bulletin*, *394*, 3–4, 11.

Baldock, J. O., & Schulte, E. E. (1996). Plant analysis with standardized scores combines DRIS and sufficiency range approaches for corn. *Agronomy Journal*, *88*(3), 448.

Bannari, A., Morin, D., Bonn, F., & Huete, A. R. (1995). A review of vegetation indices. *Remote Sensing Reviews*, *13*(1–2), 95–120.

Baumhardt, R., Stewart, B., & Sainju, U. (2015). North American Soil Degradation: Processes, Practices, and Mitigating Strategies. *Sustainability*, *7*(3), 2936–2960.

Bender, R. R., Haegele, J. W., Ruffo, M. L., & Below, F. E. (2013). Nutrient uptake, partitioning, and remobilization in modern, transgenic insect-protected maize hybrids. *Agronomy Journal*, *105*(1), 161–170.

Binkley, D., & Vitousek, P. (1989). Soil nutrient availability. In R. W. Pearcy, J. R. Ehleringer, H. A. Mooney, & P. W. Rundel (Eds.), *Plant Physiology Ecology* (pp. 75–78).

Blevins, D. G. (1985). Role of potassium in protein metabolism in plants. In Robert D. Munson (Ed.), *Potassium In Agriculture* (pp. 413–424).

Bruetsch, T. F., & Estes, G. O. (1976). Genotype variation in nutrient uptake efficiency in corn. *Agronomy Journal*, *68*(3), 521.

Cahill, S., Johnson, A., Osmond, D., & Hardy, D. (2008). Response of corn and cotton to starter phosphorus on soils testing very high in phosphorus. *Agronomy Journal*, *100*(3), 537–542.

Capehart, T. (2019). USDA ERS - Feedgrains sector at a glance. Retrieved June 3, 2019, from Corn and Other Feedgrains website: https://www.ers.usda.gov/topics/crops/corn-and-other-feedgrains/feedgrains-sector-at-a-glance/

Carter, E. K., Riha, S. J., Melkonian, J., & Steinschneider, S. (2018). Yield response to climate, management, and genotype: A large-scale observational analysis to identify climate-adaptive crop management practices in high-input maize systems. *Environmental Research Letters*, *13*(11).

Ciampitti, I. A., & Vyn, T. J. (2014). Understanding Global and Historical Nutrient Use Efficiencies for Closing Maize Yield Gaps. *Agron. J*, *106*, 2107–2117.

Coops, N. C., Hermosilla, T., Hilker, T., & Black, T. A. (2017). Linking stand architecture with canopy reflectance to estimate vertical patterns of light-use efficiency. *Remote Sensing of Environment*, *194*, 322–330.

Cure, W. W., Flagler, R. B., & Heagle, A. S. (1989). Correlations between canopy reflectance and leaf temperature in irrigated and droughted soybeans. *Remote Sensing of Environment*, *29*(3), 273–280.

Dibb, D. W., & Thompson, W. R. (1985). Interaction of potassium with other nutrients. In R. D. Munson (Ed.), *Postassium in Agriculture*.

dos Anjos Reis, R. (2002). DRIS norms universality in the corn crop. *Communications in Soil Science and Plant Analysis*, *33*(5–6), 711–735.

Duvick, D. N. (2005). Genetic progress in yield of United States maize (Zea mays L.). *Maydica*, *50*(3–4), 193–202.

Echarte, L., Nagore, L., Di Matteo, J., Cambareri, M., Robles, M., & Della Maggiora, A. (2013). Grain yield determination and resource use efficiency in maize hybrids released in different decades. *Agricultural Chemistry*, 19–36.

Elmore, R. W., Sawyer, J. E., & Boyer, M. J. (2019). Updating an old paradigm Corn growth, development, dry matter, and nutrient accumulation and partitioning. *Crops & Soils Magazine*, *March-Apri*, 34–37. Retrieved from www.certifiedcropadviser.

Elwali, A. M. O., Gascho, G. J., & Sumner, M. E. (1985). Dris norms for 11 nutrients in corn leaves. *Agronomy Journal*, *77*(3), 506.

Emmert-Streib, F., Dehmer, M., Emmert-Streib, F., & Dehmer, M. (2019). High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction*, *1*(1), 359–383.

Erives, H., & Fitzgerald, G. J. (2005). Automated registration of hyperspectral images for precision agriculture. *Computers and Electronics in Agriculture*, *47*(2), 103–119.

Fageria, V. D. (2001). Nutrient interactions in crop plants. *Journal of Plant Nutrition*, *24*(8), 1269–1290.

Forster, M., & Sober, E. (2011). AIC scores as evidence: a Bayesian interpretation. In *Philosophy of Statistics* (pp. 535–549). Elsevier.

Fox, R. H., & Piekielek, W. P. (1984). Soil magnesium level, corn (Zea mays L.) yield, and magnesium uptake. *Communications in Soil Science and Plant Analysis*, *15*(2), 109–123.

Fribourg, H. A., Bryan, W. E., Lessman, G. M., & Manning, D. M. (1976). Nutrient uptake by corn and grain sorghum silage as affected by soil Type, planting date, and moisture regime. *Agronomy Journal*, *68*(2), 260.

Gábor, A., & Banga, J. R. (2015). Robust and efficient parameter estimation in dynamic models of biological systems. *BMC Systems Biology*, *9*(1), 1–25.

Galvão, L. S., Roberts, D. A., Formaggio, A. R., Numata, I., & Breunig, M. (2009). View angle effects on the discrimination of soybean varieties and on the relationships between vegetation indices and yield using off-nadir Hyperion data. *Remote Sensing of Environment*, *113*, 846–856.

Gamon, J. A., Serrano, L., & Surfus, J. S. (1997). The photochemical reflectance index: an Optical indicator of photosynthietic radiation use efficiency across species, functional types, and nutrient levels. *Oecologia*, *112*(1997), 492–501.

Hallmark, W. B., Walworth, J. L., Sumner, M. E., deMooy, C. J., Pesek, J., & Shao, K. P. (1987). Separating limiting from non-limiting nutrients. *Journal of Plant Nutrition*, *10*(9–16), 1381–1390.

Hammer, G. L., Dong, Z., McLean, G., Doherty, A., Messina, C., Schussler, J., … Cooper, M. (2009). Can changes in canopy and/or root system architecture explain historical maize yield trends in the U.S. corn belt? *Crop Science*, *49*(1), 299–312.

Hanway, J. (1966). How a corn plant develops. *Iowa Agricultural and Home Economics Experiment Station Publications*, *48*(Sept), 1–18.

Hardy, D., & Tucker, M. (2013). *Understanding the soil test report*. Retrieved from http://ncagr.gov/agronomi/pdffiles/ustr.pdf

Hardy, D.H. (2014). Colorimetric Determination of Humic Matter with 0.2 N NaOH Extraction. In *Soil Test Methods From the Southeastern United States* (pp. 162–167). Retrieved from http://www.clemson.edu/agsrvlb/sera6/MethodsManualFinalSERA6.pdf#page=188

Hardy, David H, Tucker, M. R., & Stokes, C. E. (2014). *Crop Fertilization Based on North Carolina Soil Tests*. North Carolina Department of Agriculture and Consumer Services, Agronomic Division. Circular no. 1.

Hatfield, A. L. (1972). Soil test reporting a nutrient index system. *Communications in Soil Science and Plant Analysis*, *3*(5), 425–436.

Heiniger, R. W. (2014). *2014 North Carolina corn yield contest*. Retrieved from https://vernonjames.ces.ncsu.edu/wp-content/uploads/2015/05/Rules20141.pdf?fwd=no

Hilker, T., Coops, N. C., Hall, F. G., Black, T. A., Wulder, M. A., Nesic, Z., & Krishnan, P. (2008). Separating physiologically and directionally induced changes in PRI using BRDF models. *Remote Sensing of Environment*, *112*(2008), 2777–2788.

Hussain, H. A., Men, S., Hussain, S., Chen, Y., Ali, S., Zhang, S., … Wang, L. (2019). Interactive effects of drought and heat stresses on morpho-physiological attributes, yield, nutrient uptake and oxidative status in maize hybrids. *Scientific Reports*, *9*(1), 3890.

Inoue, Y., Peñuelas, J., Miyata, A., & Mano, M. (2008). Normalized difference spectral indices for estimating photosynthetic efficiency and capacity at a canopy scale derived from hyperspectral and CO2flux measurements in rice. *Remote Sensing of Environment,* 112 (1), 156-172.

Izsáki, Z. (2017). Effect of potassium supplies on the nutritional status of maize (Zea mays L.). *Communications in Soil Science and Plant Analysis*, *48*(19), 2347–2358.

Jacquemoud, S., Bacour, C., Poilve, H., & Frangi, J. P. (2000). Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, *74*(3), 471–481.

Jakobsen, S. T. (1993). Interaction between plant nutrients III. Antagonism between potassium, magnesium and calcium. *Acta Agriculturae Scandinavica B: Plant Soil Sciences*, *43*(1), 1–5.

Johnson, A. M., Osmond, D. L., & Hodges, S. C. (2005). Predicted impact and evaluation of North Carolina's phosphorus indexing tool. *Journal of Environmental Quality*, *34*(5), 1801–1810.

Jones, J. B., Eck, H. V., & Voss, R. (1990). Plant Analysis as an aid in fertilizing corn and grain sorghum. In R. L. Westerman (3rd ed.), Soil Testing and Plant Analysis (pp. 521-547) SSSA.

Jupp, D. L. B., & Strahler, A. H. (1991). A hotspot model for leaf canopies. *Remote Sensing of Environment*, *38*(3), 193–210.

Karlen, D. L., Flannery, R. L., & Sadler, E. J. (1988). Aerial accumulation and partitioning of nutrients by corn. *Agronomy Journal*, *80*(2), 232–242.

Khiari, L., Parent, L. E., & Tremblay, N. (2001). Selecting the high-yield subpopulation for diagnosing nutrient imbalance in crops. *Agronomy Journal*, *93*, 802–803.

Kudenov, M. W., Lowenstern, M. E., Craven, J. M., & LaCasse, C. F. (2017). Field deployable pushbroom hyperspectral imaging polarimeter. *Optical Engineering*, *56*(10), 1.

Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors (Switzerland)*, *18*(8), 1–29.

Lidon, F. C., Azinheira, H. G., & Barreiro, M. G. (2000). Aluminum toxicity in maize: Biomass production and nutrient uptake and translocation. *Journal of Plant Nutrition*, *23*(2), 151–160.

Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., & Schlenker, W. (2013). The critical role of extreme heat for maize production in the United States. *Nature Climate Change*, *3*(5), 497–501.

Long, N. V., Assefa, Y., Schwalbert, R., & Ciampitti, I. A. (2017). Maize yield and planting date relationship: A synthesis-analysis for U.S. high-yielding contest-winner and field research data. *Frontiers in Plant Science*, *8*, 2106.

Luetchens, J., & Lorenz, A. J. (2018). Changes in dynamic leaf traits in maize associated with year of hybrid release. *Crop Science*, *58*(2), 551–563.

Marschner, H., & Dell, B. (1994). Nutrient uptake in mycorrhizal symbiosis. *Plant and Soil*, *159*(1), 89–102.

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817.

McGinnis, M., Stokes, C., & Cleveland, B. (2014). Plant/Waste/Solution/Media Analysis Section. In *NCDA&CS Plant Tissue Guide* (Vol. 2012, pp. 1–25).

Mehlich, A. (1976). New buffer ph method for rapid estimation of exchangeable acidity and lime requirement of soils. *Communications in Soil Science and Plant Analysis*, *7*(7), 637–652.

Mehlich, A. (1984a). Mehlich 3 soil test extractant: A modification of Mehlich 2 extractant. *Communications in Soil Science and Plant Analysis*, *15*(12), 1409–1416.

Mehlich, A. (1984b). Photometric determination of humic matter In soils, A proposed method. *Communications in Soil Science and Plant Analysis*, *15*(12), 1417–1422.

Mehlich, A., Bowling, S. S., & Hatfield, A. L. (1976). Buffer pH acidity in relation to nature of soil acidity and expression of lime requirement. *Communications in Soil Science and Plant Analysis*, *7*(3), 253–263.

Melsted, S. W., Motto, H. L., & Peck, T. R. (1969). Critical plant nutrient composition values useful in interpreting plant analysis data. *Agronomy Journal*, *61*(1), 17–20.

Metternicht, G. (2003). Vegetation indices derived from high-resolution airborne videography for precision crop management. *International Journal of Remote Sensing*, *24*(14), 2855–2877.

Mulla, D.J., & Schepers, J. S. (1997). Key processes and properties for site-specific soil and crop management. In Robert P C (Ed.), *The site-specific management for agricultural systems* (pp. 1–18).

Mulla, David J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. *Biosystems Engineering*, *114*(4), 358–371.

Mundorf, T., Wortmann, C., Shapiro, C., & Paparozzi, E. (2015). Time of day effect on foliar nutrient concentrations in corn and soybean. *Journal of Plant Nutrition*, *38*(14), 2312–2325.

NCDA&CS. (2017). *North Carolina Agriculture Statistics.* (K. Krueger, Ed.). USDA NASS.

Nettleton, D. F., Orriols-Puig, A., & Fornells, A. (2010). A study of the effect of different types of noise on the precision of supervised learning techniques. *Artificial Intelligence Review*, *33*(4), 275–306.

Ohno, T., & Grunes, D. L. (1985). Potassium-magnesium interactions affecting nutrient uptake by wheat forage. *Soil Science Society of America Journal*, *49*, 685–690.

Parent, L. E., & Dafir, M. (1992). A theoretical concept of compositional nutrient diagnosis. In *J. Amer. Soc. Hort. Sci.* (Vol. 117).

Penuelas, J., Pinol, J., Ogaya, R., & Filella, I. (1997). Estimation of plant water concentration by the reflectance Water Index WI (R900/R970). *International Journal of Remote Sensing*, *18*(13), 2869–2875.

Penuelas, Josep, & Filella, I. (1998). Visible and near-infrared reflectance techniques for diagnosing plant physiological status. *Trends in Plant Science*, *3*(4), 151–156.

Rahmatullah, & Baker, D. E. (1981). Magnesium accumulation by corn (Zea mays L.) as a function of potassium-magnesium exchange in soils. *Soil Science Society of America Journal*, *45*(21), 899–903.

Rajkovich, S. R., Crozier, C. R., Smyth, T. J., Crouse, D., & Osmond, D. L. (2015). Updating North Carolina corn yields and nitrogen recommendations to match current production practices and new hybrids. *Crops, Forage & Turfgrass Management*, *1*(1), 1–8.

Rietra, R. P. J. J., Heinen, M., Dimkpa, C. O., & Bindraban, P. S. (2017). Effects of nutrient antagonism and synergism on yield and fertilizer use efficiency. *Communications in Soil Science and Plant Analysis*, *48*(16), 1895–1920.

Ritchie, S. W., Hanway, J. J., & Benson, G. O. (1986). *How a corn plant develops. Spec. Rep. 48. Rev*.

Roujean, J., Leroy, M., & Deschamps, P.-Y. (1992). A bidirectional reflectance model of the Earth's surface for the correction of remote sensing data. *JOURNAL OF GEOPHYSICAL RESEARCH*, *97*(D18), 455–475.

Sandmeier, S. R., & Itten, K. I. (1999). A field goniometer system (FIGOS) for acquisition of hyperspectral BRDF data. *IEEE Transactions on Geoscience and Remote Sensing*, *37*(2 II), 978–986.

Schlerf, M., Hill, J., Koetz, B., & Kneubuehler, M. (2007). Retrieving canopy structure from hyperspectral multi-angular satellite data. *Proceedings 5th EARSeL Workshop on Imaging Spectroscopy.*, *1*(April 23-25 2007), 1–7. Bruges, Belgium.

Smith, T. C., & Frank, E. (2016). Introducing machine learning concepts with WEKA. In E. Mathe & S. Davis (4th eds.), *Statistical Genomics. Methods in Molecular Biology* (Vol. 1418, pp. 353–378).

Soltanpour, P. N., Malakouti, M. J., & Ronaghi, A. (1995). Comparison of diagnosis and recommendation integrated system and nutrient sufficiency range for corn. *Soil Science Society of America Journal*, *59*(1), 133.

Stromberger, J. A., Tsai, C. Y., & Huber, D. M. (1994). Interactions of potassium with nitrogen and their influence on growth and yield potential in maize. *Journal of Plant Nutrition*, *17*(1), 19–37.

Sumner, M. E. (1977). Application of Beaufil's diagnostic indices to maize data published in the literature irrespective of age and conditions. In *Plant and Soil* (Vol. 46). Retrieved from

Terman, G. L., Allen, S. E., & Bradford, B. N. (1975). Nutrient dilution-antagonism effects in corn and snap beans in relation to rate and source of applied potassium. *Soil Science Society of America Journal*, *39*(4), 680–685.

Terman, G. L., Allen, S. E., & Giordano, P. M. (1973). Dry matter yield-N and S concentration relationships and ratios in young corn plants. *Agronomy Journal*, *65*(4), 633–636.

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Tolk, J. A., Howell, T. A., & Evett, S. R. (1999). Effect of mulch, irrigation, and soil type on water use and yield of maize. *Soil and Tillage Research*, *50*(2), 137–147.

Tollenaar, M., & Aguilera, A. (1992). Radiation use efficiency of an old and a new maize hybrid. *Agronomy Journal*, *84*(3), 536–541.

Tonitto, C., David, M. B., & Drinkwater, L. E. (2006). Replacing bare fallows with cover crops in fertilizer-intensive cropping systems: A meta-analysis of crop yield and N dynamics. *Agriculture, Ecosystems and Environment*, *112*(1), 58–72.

United Nations Dept. of Economic Affairs. (2019). How certain are the United Nations global population projections? In *Population Facts* (Vol. 2019).

USDA. (2020). Crop Values: 2019 summary. In *Crop Values 2019 Summary*. Retrieved from https://www.nass.usda.gov/Publications/Todays_Reports/reports/cpvl0220.pdf

van Roekel, R. J., & Coulter, J. A. (2011). Agronomic responses of corn to planting date and plant density. *Agronomy Journal*, *103*(5), 1414–1422.

Vieira, S. M., Kaymak, U., & Sousa, J. M. C. (2010). Cohen's kappa coefficient as a performance measure for feature selection. *2010 IEEE World Congress on Computational Intelligence, WCCI 2010*, 1–8.

Vyn, T. J., & Tollenaar, M. (1998). Changes in chemical and physical quality parameters of maize grain during three decades of yield improvement. *Field Crops Research*, *59*, 135–140.

Walworth, J. L., & Ceccotti, S. (1990). A re-examination of optimum foliar magnesium levels in corn. *Communications in Soil Science and Plant Analysis*, *21*(13–16), 1457–1473.

Walworth, J. L., & Sumner, M. E. (1987). The Diagnosis and Recommendation Integrated System (DRIS). In B. A. Stewart (Ed.), *Advances in Soil Science* (pp. 149–188). New York, NY: Springer New York.

Walworth, J. L., Woodward, J. J., & Sumner, M. E. (1988). Generation of corn tissue norms from small, high-yield data base. *Communications in Soil Science Plant Nutrition*, *19*(5), 563–577.

Wang, H., & Zheng, H. (2013). True positive rate. In W. Dubitzky, O. Wolkenhauer, K.-H. Cho, & H. Yokota (Eds.), *Encyclopedia of Systems Biology* (pp. 2302–2303).

Welikhe, P., Essamuah-Quansah, J., Boote, K., Asseng, S., El Afandi, G., Fall, S., … Ankumah, R. (2016). Impact of climate change on corn yields in Alabama. *Professional Agriculture Workers Journal*, *4*(1), 1–14.

Witten, I. H., Elbe, and F., Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. (3rd ed.). Morgan Kaufman

Woli, K. P., Sawyer, J. E., & Boyer, M. J. (2018). Corn Era Hybrid Macronutrients and Dry Matter in Plant Components. *Agronomy Journal*, *110*(5), 1648–1658.

Woli, K. P., Sawyer, J. E., Boyer, M. J., Abendroth, L. J., & Elmore, R. W. (2017). Corn era hybrid dry matter and macronutrient accumulation across development stages. *Agronomy Journal*, *109*(3), 751–761.

Woli, K. P., Sawyer, J. E., Boyer, M. J., Abendroth, L. J., & Elmore, R. W. (2018). Corn era hybrid macronutrient and dry matter accumulation in plant components. *Agronomy Journal*, *110*(5), 1648–1658.

Xie, D., Qin, W., Wang, P., Shuai, Y., Zhou, Y., & Zhu, Q. (2017). Influences of leaf-specular reflection on canopy BRF characteristics: A case study of real maize canopies with a 3-D scene BRDF model. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(2), 619–631.

Yang, C.-C., Prasher, S. O., Whalen, J., & Goel, P. K. (2002). PA—Precision agriculture: Use of hyperspectral imagery for identification of different fertilisation methods with decision-tree technology. *Biosystems Engineering*, *83*(3), 291–298.