2023

Building No-Tillage Maize Nitrogen Recommendation Algorithm that Considers Improvements in Soil Health

Dwarika Bhattarai
South Dakota State University, dwarikabh@gmail.com

BUILDING NO-TILLAGE MAIZE NITROGEN RECOMMENDATION ALGORITHM THAT CONSIDERS IMPROVEMENTS IN SOIL HEALTH

BY

DWARIKA BHATTARAI

A dissertation submitted in partial fulfillment of the requirements for the

Doctor of Philosophy

Major in Plant Science

South Dakota State University

2023

DISSERTATION ACCEPTANCE PAGE
Dwarika Bhattarai

This dissertation is approved as a creditable and independent investigation by a candidate for the Doctor of Philosophy degree and is acceptable for meeting the dissertation requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

David E. Clay
Advisor Date

David Wright
Department Head Date

Nicole Lounsbery, PhD
Director, Graduate School Date

This dissertation is dedicated to you, my dear grandfather, whose unwavering love, guidance, and inspiration have been a constant presence in my life. Though you are no longer with us in this world, your memory lives on in the lessons you taught me and the values you instilled in me.

I am forever grateful for the time we had together and the memories we shared. Your legacy will continue to motivate and inspire me throughout my life.

Rest in peace, my dear grandfather. Your memory will always be cherished and your impact on my life will never be forgotten.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to all those who have supported me throughout my academic journey and the completion of this dissertation.

First and foremost, I would like to thank my advisor Dr. David E. Clay, for his guidance, mentorship, and unwavering support throughout this process. Dr. Clay's expertise and insights have been invaluable, and I am grateful for the opportunity to work with him. I would also like to express my appreciation to my committee members, Drs. Sharon A. Clay, Gary Hatfield, Jason Clark, Shin-Yi Marzano, and Warren Rusche, for their valuable feedback and contributions to this dissertation.

I am deeply indebted to my parents, Tenth and Lila Devi Bhattarai, my wife, Pratiksha K C, and my son, Davion Lok Bhattarai, for their love, encouragement, and unwavering support. Their belief in me and my abilities has been a driving force throughout this process.

I would like to acknowledge the support and encouragement of my friends and colleagues, Dr. Shaina Westhoff, Dr. Janet Miller, Dr. Stephanie Bruggeman, Dr. Deepak Raj Joshi, Graig Reicks, Ben Groebner, Andrew Ahlersmeyer, Shailesh Pandit, Skye Brugler, Andrew Engel, Mehmet Budak, and Prashant K C, who have provided me with invaluable advice and support throughout my academic journey. I would like to thank all the farmers who took part in this study.

Lastly, my sincere thanks to South Dakota Nutrient Research and Education Council for funding this research. Without their generous support, this research would not have been possible.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

BUILDING NO-TILLAGE MAIZE NITROGEN RECOMMENDATION ALGORITHM THAT CONSIDERS IMPROVEMENTS IN SOIL HEALTH

DWARIKA BHATTARAI

2023

In long-term no-till fields, farmers have reported that less N is required to optimize maize (*Zea mays* L.) yields in long term no-tillage fields than conventional tillage fields. These reductions may be attributed to improved soil health resulting from increasing soil organic matter, higher soil microbial activities, and improved water and nutrient use efficiency. However, the impact of soil health measurements on fertilizer-N requirement has not been determined. The objective of this dissertation was to compare different regional N recommendation models to measured values and develop a maize fertilizer-N recommendation model, using machine learning approaches, that includes adjustments based on soil health measurements. The research was conducted for three years at 16-dryland sites that were under no-tillage practice for at least 6-years. The effect of six N rates (0, 45, 90, 135, 180, and 224 kg N ha^{-1}) on maize grain yield was evaluated. Soil samples for nitrate-N (NO_3-N), ammonium-N (NH_4-N), pH, EC, and phospholipid fatty acid (PLFA) were collected from various depths before planting and after harvest. Climate variations influenced the maize yield across experimental sites. Comparison of error rates and bias showed that at lower cost/value ratios the current South Dakota and North Dakota N models had lower error rates and biases than models used in Nebraska, Iowa, and Minnesota. Further, using soil health measurements the support vector machine (SVM) algorithm outperformed several other machine learning algorithms for forecasting the soil yield potential. The top five predictor variables were total N, total C, growing degree days (GDD), soil microbial biomass, and bacterial biomass. The overall findings from this study suggested that soil organic C, total N, inorganic N, soil microbial biomass in addition to the climate variables, rainfall, and temperature, can be used to predict the soil yield potential.

STATEMENT OF PROBLEM

Agricultural practices focusing on maximizing crop yield has the potential to negatively impact the soil and the environment. We believe that these risks can be reduced by adopting climate-smart practices. This dissertation is focused on the development of climate smart practices for no-tillage soils located in a semiarid frigid soil. Common climate smart practices include reducing the tillage intensity, planting cover crop, and adoption a 4R nutrient fertilizer approach. The 4R nutrient approach includes applying the right fertilizer, at the right, using the right product, at the right time, at the right place. This project is focused on the right rate.

The overuse of N fertilizers can cause soil acidification, nutrient imbalances, soil erosion, and water pollution. The over application of N fertilizer contributes to greenhouse gas (GHG) emissions, further exacerbating climate change (Thies et al., 2020; Bhattarai et al., 2021; Sainju et al., 2019). The impact of climate change is not observed evenly across the globe but is often focused on people who are least able to manage the problem. Therefore, the challenge is to develop an N recommendation system that considers soil health and ecosystem services while maintaining crop productivity at lower costs.

No-tillage systems mainly rely on organic matter decomposition and nutrient cycling to maintain soil fertility, making it important to consider the current soil nutrient status and crop requirements before applying any fertilizer. Therefore, N fertilizer recommendation in no-till systems differs from the conventional farming systems and can be a more challenging task. No-tillage can be separated into two-time frames, transition, and long-term systems. During the transition period soil organic matter increases and it can often require more N than conventional system. However, different results may be observed in long-term systems and require less N fertilizer than conventionally tilled systems. This N rate reduction may be attributed to several factors including: 1) reduced N loss from leaching and volatilization, 2) improved soil organic C and fertility thereby enhancing nutrient cycling, 3) reduced soil compaction due to better root growth, and 4) the retention of crop residues on the soil surface that helps to conserve soil moisture, reduce soil erosion, and improve fertilizer efficiency (St. Luce et al., 2021; Blanco-Canqui & Ruis, 2018). Globally, agronomists have struggled with optimizing yields and minimizing N rates. One approach for achieving this goal is to consider management induced changes in soil health. Classical soil health measurements include measures of soils physical, chemical, and biological properties. For example, water infiltration, biological activity, and soil organic matter. To date, integrated soil health information that combines physical, chemical, and biological properties has not been integrated into state-based N recommendation guideline in central US (Clark, 2019; Franzen, 2018; Kaiser et al., 2022). Preliminary research data from North Dakota and producer interviews suggests that as a soil transitions from a cultivated to a no-tillage system produces changes in soil health that can result in the amount of N needed to economically optimize yields. However, a scientific justification for this N rate reduction has not been identified. This research focuses on two main objectives, 1) to validate if the regional N recommendation models are appropriate for long-term no-till corn N recommendations, and 2) to use machine learning approaches to predict the soil yield potential considering various soil health measurements as the predictor variables.

CHAPTER 1: ARE THE REGIONAL N RECOMMENDATION MODELS APPROPRIATE FOR LONG-TERM NO-TILL MAIZE?

ABSTRACT

The current maize (*Zea mays*, L.) nitrogen (N) recommendation model in South Dakota (SD) is based on the yield goal approach, which is easy to understand but requires modifications considering climatic variations, and fertilizer and maize grain price to adjust in the cropping systems of semi-arid regions of SD. Research is required to determine the best regional N recommendation model that fits in the frigid soils with sub-humid to semi-arid moisture regime of SD. The objective of this study was to compare maize N fertilizer algorithms on their ability to predict N recommendation in long-term no-tillage fields located in frigid semi-arid environment. The research was conducted for three years at 16-dryland sites that were under no-tillage practice for at least 6-years. The effect of six N rates (0, 45, 90, 135, 180, and 224 kg N ha^{-1}) on maize grain yield was evaluated and the economic optimum N rates were calculated at N-to-maize price ratios of 4.11, 5.48, 6.85, and 8.23. Maize yield was influenced by climatic variations at different locations. Comparison of error rates and bias showed that at 4.11 price ratio the current SD N model had lower error rate and bias than models used in Eastern North Dakota (ND), Nebraska, Iowa, and Minnesota. At the price ratio of 4.11, EONR and the current SD N model recommended the same amount of N, 174 kg N ha^{-1}. However, price-based adjustments to the current SD N recommendation model were recommended.

INTRODUCTION

Developing a universal maize (*Zea mays* L.) N model has challenged soil scientists and agronomists for many years (Morris et al., 2018; Rodriguez et al., 2019; Tremblay et al., 2012). One of the early attempts at managing N fertilizers was the use of the yield goal model (Equation 1). However, this model, while easy to understand, had several basic flaws that included not accounting for changes in soil health, tillage, soil organic C, available water, climatic conditions, differential mineralization, and price of fertilizer or selling price of the product (Kim et al., 2008). This basic model is,

$$\text{N Recommendation} = \text{constant} \times \text{Yield Goal} - \text{N credits} \qquad (1)$$

In equation (1), the constant ranges from 18 to 26.8 kg N Mg^{-1} maize grain, and N credits include soil NO_3-N, N from previous legume crops, and N in the irrigation water. Agronomists have proposed numerous modifications of this model since the 1970s (Morris et al., 2018). The most common modification is to consider the potential of the soil to mineralize N. Most widely adopted N models do not consider the N mineralization potential as a credit for many reasons including: 1) the models are too complicated, 2) chemical extraction procedures do not provide consistent results across sites and years and 3) the methods are not easily integrated into commercial soil testing laboratories.

The second component of equation 1 that has been widely discussed is the yield goal. Research shows that there is often a poor correlation between the yield goal and the economic optimum N rate (Morris et al., 2018; Andraski & Bundy, 2002; Fox & Piekielek, 1987; Lory & Scharf, 2003). These poor correlations are attributed to not considering the economic constraints and the model's inability to consider soil and climatic variation. In addition, equation 1 does not correlate with the N provided by soil, fertilizer and maize costs, changes in fertilizer use efficiency, and it suggests that simply increasing the N rate will increase the yield (Sawyer et al., 2006). Due to these and other limitations of the yield goal-based approach, many soil scientists and agronomists are replacing or modifying the approach. For example, Iowa replaced the yield goal approach with the maximum return to N (MRTN) approach, whereas Nebraska integrated organic matter into the algorithm. Numerous other states have integrated the cost of fertilizer and value of product into the calculations. Given the wide variability in models used to predict N recommendations, research is needed to determine what type of model is best suited for a region that contains frigid soils and moisture that varies from subhumid to semiarid. Therefore, the objective of this study was to compare maize N fertilizer algorithms on their ability to predict N recommendation in long-term no-tillage fields located in frigid semi-arid environment.

MATERIAL AND METHODS

Experimental details

On farm experiments were conducted in 2019, 2020, and 2021. The study sites were in long-term no-till farmer's fields (>6 years no-till). Over three years, the experiment was conducted in 16 dryland sites located at six different counties of South Dakota (Figure 1-1). All locations are characterized as a hot summer humid continental climate (Köppen climate: Dfa). Out of 16 sites, one site was planted to maize following fallow, two sites were seeded following livestock grazing, two sites were seeded following a winter cover crop, four sites were seeded to maize following soybean (*Glycine max* L.), and seven sites were seeded to maize following winter wheat (*Triticum* *aestivum* L). Agronomic details including the planting dates, harvest dates, maize varieties, previous crops, and plant population of each site are shown in Table 1-1.

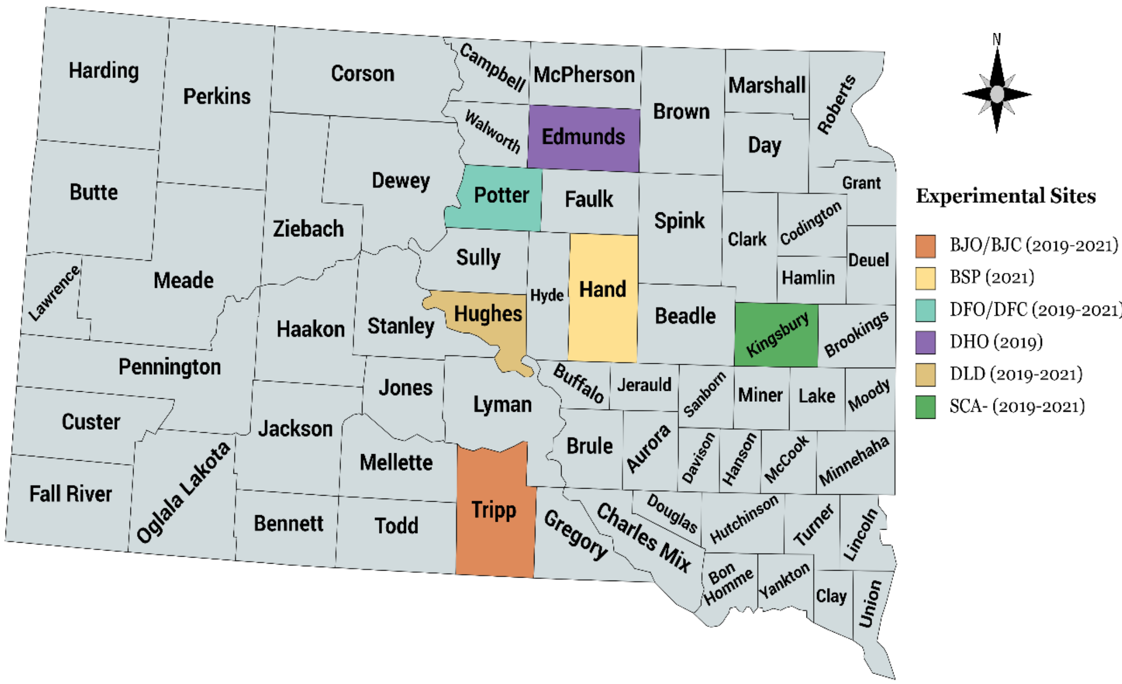


Figure 1- 1: The map of South Dakota highlighted with the counties where field experiments were conducted from 2019-2021. Each color in the legend represents the experimental field names with the different years of experiment.
Source: https://www.mapchart.net/usa-counties.html

The experimental design was a randomized complete block design (RCBD) with four N rates (28, 84, 140, and 196 kg N ha^{-1}) in 2019 and six N rates (0, 45, 90, 134, 179, and 224 kg N ha^{-1}) in 2020 and 2021. The experiments contained four blocks and urea (46-0-0) was the fertilizer source. Nitrogen treatments were manually broadcast applied between the V2 and V4 maize growth stages. Each plot had dimensions of 15.24 × 4.6 m (50 ft × 15 ft). Based on initial soil test values phosphorus and K fertilizers were applied (Clark, 2019; Supplementary Table 1).

Baseline soil samples from four blocks were collected from four depths (0-5, 5-15, 15-30, and 30-60 cm) before planting maize and before the treatment application at each site. Plant residues on the soil surface were carefully removed and the samples were randomly collected from 15-20 random spots within each block using a standard soil probe with 1.9 cm inner diameter. Soil samples, for each depth per block, were mixed thoroughly, air-dried to a constant weight, and sieved through 2-mm mesh before soil analyses. Soil inorganic N, nitrate-N (NO_3-N) and ammonium-N (NH_4-N), were extracted using 1M KCl (1:10 soil to KCl ratio) (Kim et al., 2008), quantified by cadmium reduction method (Clark et al., 2019) analyzed using Astoria Analyzer (Astoria-Pacific).

Maize was seeded with no-tillage planters and most sites had 75 cm row spacing. However, there were several sites that deviated from this convention. For example, the BJC site in 2020 was planted with a 150 cm row spacing and a cover crop mixture was seeded at the maize V3 growth stage. The cover crop mixture at this site were oats (*Avena sativa*, L., 13 kg ha^{-1}), flax (*Linum usitatissimum*, L., 2.2 kg ha^{-1}), mung bean (*Vigna radiata*, L., 4.5 kg ha^{-1}), guar (*Cyamopsis tetragonoloba*, L. 2.2 kg ha^{-1}) and red clover (*Trifolium pratense* var. *sativum* (Schreb.), 1.1 kg ha^{-1}). In addition, a narrow row spacing (50 cm) was used at the Hughes County sites.

Maize ears and stover were hand-harvested from 9.29 m^2 area marked from the center of each plot. Based on these values plant populations, harvest indexes, and corn yields at 15.5% moisture were determined.

Table 1- 1. Agronomic information including planting and harvest date, corn varieties and their maturity days, previous crops, and plant population of different experiment sites across the experiment years, 2019-2021.

Field name	**Counties**	**Planting Date**	**Harvest Date**	**Maize Maturity Days**	**Previous crop(s)**	**Plant population** (per ha)
BJO19	Tripp	5/15/2019	10/19/2019	106	Wheat	60500
DFO19	Potter	5/14/2019	10/08/2019	101	Wheat	62000
DHO19	Edmunds	5/26/2019	10/14/2019	88	Soybean	65000
DLD19	Hughes	5/15/2019	10/15/2019	105	Wheat	52000
SCA19	Kingsbury	5/16/2019	10/18/2019	97	Soybean	79000
BJC20	Tripp	4/29/2020	10/01/2020	99	Wheat	60500
BJO20	Tripp	4/29/2020	10/01/2020	100	Wheat	60500
DFO20	Potter	5/11/2020	10/06/2020	96	Wheat	62000
DLD20	Hughes	4/30/2020	9/26/2020	99	Oats + Barley	52000
SCA20	Kingsbury	4/27/2020	10/08/2020	97	Soybean	79000
BJO21	Tripp	5/3/2021	10/18/2021	99	Wheat-fallow-livestock	59000
BSP21	Hand	5/4/2021	10/5/2021	102	Fallow	69000
DFC21	Potter	5/4/2021	10/18/2021	101	Cover crops mix	65500
DFO21	Potter	5/5/2021	10/12/2021	102	Wheat-fallow-livestock	64000
DLD21	Hughes	5/6/2021	10/7/2021	100	Wheat	86500
SCA21	Kingsbury	4/30/2021	9/28/2021	105	Soybean	79000

Nitrogen fertilizer response and EONR calculations

The delta yield values were calculated to assess the improvement in N recommendation tool using the equation,

$$N\ resposnse_{EONR} = Y_{EONR} - Y_{0N} \quad (2)$$

Where Y_{EONR} is the maize yield at the EONR and Y_{0N} is the maize yield when N was not applied (Kim et al., 2013; Lory & Scharf, 2003). The yield at EONR was calculated at the fertilizer-to-maize price ratio of 4.11. The correlation of delta yield was calculated with the EONR and the maize yield.

Economic optimum N rates were calculated using quadratic and quadratic-plateau models using PROC GLM and PROC NLIN procedures, respectively, in SAS Studio (v3.8, Enterprise Edition, SAS Institute Inc., Cary, NC, USA). The best model with the lower root mean squared error (RMSE) was selected for further calculations. Economic optimum N rate is the point where the last increment of fertilizer provides a yield increase that can pays for the additional amount of fertilizer applied (CNRC, 2022). Both models were developed by plotting maize yield (kg ha^{-1}) against applied N rates (kg N ha^{-1}), excluding the 0 N rate to exclude the possible bias. The EONRs were calculated using the first derivative of the selected model and fertilizer cost-to-maize price ratios (4.11, 5.48, 6.85, and 8.23) using a maize price of US\$262 Mg^{-1} grain at 15.5% moisture and N fertilizer (Urea) costs of US\$450 ton^{-1}, US\$600 ton^{-1}, US\$750 ton^{-1}, and US\$900 ton^{-1}, respectively. The fertilizer-to-maize price ratios considered were 0.075, 0.10, 0.125, and 0.15 (Kaiser et al., 2022). The EONR values were calculated using the equation,

$$EONR = \frac{\frac{\$/(kg\ N\ fertilier)}{\$/(corn\ grain)} - b}{2c} \quad (3)$$

Where EONR is economic optimum N rate (kg N ha^{-1}), \$/ (kg of N) is cost of N (46% of Urea fertilizer cost), \$/ (kg maize grain) is selling price of maize, b is the linear coefficient, c is the quadratic coefficient from the models. The economic optimum yield, also written as the yield at EONR, was calculated by substituting the EONR in the quadratic and quadratic-plateau models and solving the equation (Bhattarai et al., 2021).

Nitrogen recommendation algorithms

Across experimental sites, the EONRs for no-till maize were calculated using existing algorithms that are used in South Dakota, North Dakota (ND), Western Minnesota, Iowa, and Nebraska. The South Dakota N recommendation (NR; kg N ha^{-1}) model is,

$$NR = k \times YG - STN - LC \tag{4}$$

In equation (4), YG is the yield goal (Mg ha^{-1}), k is 21.4 kg N Mg^{-1} grain for the historic model, STN is the amount of NO_3-N (kg N ha^{-1}) at 0-to-60 cm soil depth, and LC is the previous legume crop credit (44 kg N ha^{-1}) (Clark, 2019). An updated N recommendation model for South Dakota has reduced the value of k from 21.4 to 17.86 kg N Mg^{-1} grain (Clark, 2023). The yield goal was 11.29 Mg ha^{-1}, which was determined as the average yield across experimental sites at the maximum N fertilizer rate. North Dakota N recommendation model has updated the N recommendation model to the MRTN approach from yield goal approach (Franzen, 2018). Nitrogen recommendation model in ND is classified as East ND and West ND approaches for long-term no-till systems. The ND N recommendation model includes soil test N at 0-to-60 cm soil depth and organic matter content (https://www.ndsu.edu/pubweb/soils/corn/).

The Western Minnesota recommendation model is,

$$NR = (MRTN\ for\ maize\ /\ maize) - 0.60 \times STN - irrigation\ N \qquad (5)$$

where the MRTN for maize/maize is the fertilizer-N requirement based on the N cost/crop price ratio, and STN is the amount of NO_3-N (kg N ha^{-1}) contained in the 0-60 cm soil depth (Kaiser et al., 2022), and irrigation N is the N applied by the irrigated water. Based on the Minnesota model, the MRTN for the 4.11, 5.48, 6.85, and 8.23 fertilizer-to-maize price ratios were 213, 196, 185, and 174 kg N ha^{-1}.

The Iowa N recommendation algorithm used the MRTN approach., that does not consider preseason soil NO_3-N. Based on Sawyer et al. (2006), the recommendations for maize-maize rotations were 188, 175, 165, and 155 kg N ha^{-1} and the recommendations for maize-soybean rotations were 135, 125, 118, and 110 kg N ha^{-1} for the 4.11, 5.48, 6.85, and 8.23 fertilizer-to-maize price ratios, respectively. The Nebraska N recommendation algorithm was,

$$NR = [39 + 21.4 \times YG - \frac{2.505(YG \times OM)}{10} - 9(NO_3 - N) - LC - irrigation\ N]f_A \times f_R \ (6)$$

Where, YG was the yield goal (Mg ha^{-1}), OM is the organic matter content (up to 30 g OM kg^{-1}), NO_3-N was the average amount of nitrate-N contained in the surface 120 cm (3 mg kg^{-1} soil was considered for the 60 to 120 cm NO_3-N content), LC was the soybean credit (50 kg N ha^{-1}), f_A was the correction factor for application time, and f_R was the correction factor for the maize/N price ratio (Shapiro et al., 2019).

Statistical Analysis

For each model, the root mean square error (RMSE) and bias were calculated using equations 7 and 8, respectively.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)^2}{n}} \tag{7}$$

$$Bias = \frac{\sum_{i=1}^{n}(\hat{x}_i - x_i)}{n} \tag{8}$$

In equations 7 and 8, n is the number of comparisons, i is which individual sample, x_i means the measured EONR values, and $\hat{x}_i$ is the mean of the measured N recommendations. A positive bias value indicates that the model overestimated the N recommendation while a negative bias indicated that the model underestimated the N requirements.

RESULTS AND DISCUSSION

Rainfall and Temperatures

In frigid semiarid environments, it is likely that yields will be reduced by cool and drought conditions during some stage on growth. Generally, corn yield is most sensitive to drought during reproductive stages is most resistant to cool conduction conditions early in the growing season and least able to manage drought conductions during reproductive growth stages and least sensitive as the plant approaches the R6 growth stage. To address seasonal differences, the climate data was separated into prior to tasseling and the entire growing season.

Rainfall and growing degree days (GDD, based on 10 ºC) varied by year and site (Table 1-2). In South Dakota, generally requires 780 growing degree days by silking and 1390 days by maturity for corn with a maturity rating of 100 days. In all three years, GDD were less than 1390 in 2019 (1280) and greater than 1390 in 2019, (1633) and 2020 (1491). It is likely that low temperatures reduced yields in 2019.

The seasonal cumulative precipitation at corn at black layer was higher in 2019 than 2020 and 2021. However, slightly different results were observed at the tasseling stage when precipitation was less in 2021 than 2019 and 2020. It is likely that the limited precipitation prior to tasseling in 2021, reduced yields. One approach that has been used to increase yields in water stressed systems is called the skip row approach. In this approach, the seeding rate within a field remains the same but the row width increases. This approach has been used to increase the amount of stored soil moisture during the early growth stages.

Table 1- 2: Pre-planting water depth at 0-to-60 cm soil depth (cm), cumulative rainfall (cm) from planting to tasseling and harvesting stages, and cumulative GDDs from planting to harvesting across different experimental sites, 2019-2021. The GDDs were calculated using the base temperature of 10 ºC.

Years	Sites	Counties	Water depth (cm)	Cumulative Rainfall (cm)		Cumulative GDD
				Tassel	Harvest	Harvest
2019	BJO19	Tripp	19.43	45	69.1	1441
	DFO19	Potter	20.32	28	48.2	1101
	DHO19	Edmunds	15.20	18.5	42.1	1128
	DLD19	Hughes	21.41	12.3	35.2	1389
	SCA19	Kingsbury	14.29	33.8	66.1	1235
2020	BJC20	Tripp	26.20	25.7	39.1	1584
	BJO20	Tripp	24.10	25.7	39.1	1584
	DFO20	Potter	19.63	14.9	23.1	1266
	DLD20	Hughes	17.68	17.7	28	1557
	SCA20	Kingsbury	23.20	22.5	42	1398
2021	BJO21	Tripp	24.60	8.4	30.8	1812
	BSP21	Hand	20.95	9.1	26.5	1601
	DFC21	Potter	18.97	8.4	27.4	1497
	DFO21	Potter	20.56	8.4	27.4	1492
	DLD21	Hughes	20.38	4.2	23.2	1785
	SCA21	Kingsbury	21.31	7.8	34.6	1460

Maize yields across years

Water availability is one of the key factors influencing maize yields (Kim et al., 2013; Kim et al., 2008). Water impacts the plant demand for N and the ability of the soil to mineralize N and to transport the inorganic N to the growing plant. Previous studies have shown that soil water directly impacts microbial activity and because nitrate moves to the root in the transpiration stream, increasing transpiration increases nitrate uptake and nitrogen use efficiency (Clay et al., 2006; Clay et al., 1990). Mineralization of organic matter, driven by soil microbes and environmental factors, has potential to impact the N requirements (Cotrufo et al., 2019).

In dryland no-tillage dryland systems, surface residue reduces evaporation and increases precipitation use efficiency. This increased water can result in improved N use efficiency. Other approaches to improve water use efficiency is to reduce the seeding rate or use a skip row planting technique.

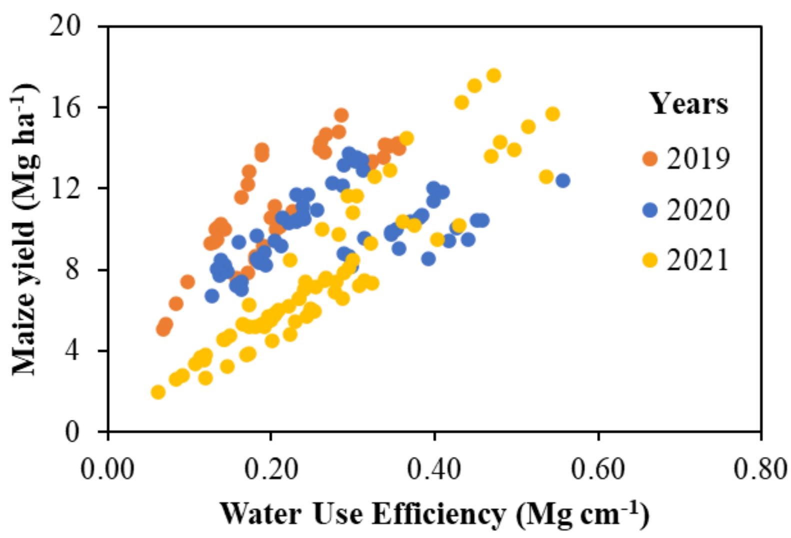


Figure 1- 2: Scatter plot showing the relationship between maize yield (Mg ha^{-1}) and water use efficiency (Mg cm^{-1}) in the experimental sites across 2019, 2020, and 2021.

In water limited systems, the stomates close to prevent water loss which in turn reduce CO_2 fixation during photosynthesis. Across the study sites, there was a strong relationship between maize yields and water use efficiency (Figure 1-2). Generally increasing water use efficiency improved maize yields. The optimum maize yield across the dryland sites in 2019, 2020, and 2021 were 11.20, 10.04, and 7.78 Mg ha^{-1}, respectively. Lower yields in 2021 were attributed to water stress during grain filling (Table 1-2 and Figure 1-2).

Fertilizer-N responsive and non-responsive sites

Out of 16 dryland sites, 12 were responsive to the fertilizer-N whereas the other four sites were non-responsive to N. Sites that were not responsive to N could be attributed to several factors including drought, higher N mineralization rates, high seeding rates, and high residual N concentrations (Table 1-1 and Table 1-2). The responsive sites generally had higher yield than the non-responsive sites which increased with increasing fertilizer-N (Figure 1-3). Maize yield from the responsive sites revealed that there was no yield difference ($p=0.066$) over years (data not shown); however, the yields were lower ($p<0.001$) in the low rainfall year of 2021(5.54 Mg ha^{-1}) than the higher rainfall year of 2020 (8.11 Mg ha^{-1}) in non-responsive sites.

Although the yields were comparable to the sites that were responsive to N, four sites were non-responsive to fertilizer-N (Table 1-3). In 2019, there was enough moisture to support maize development (Table 1-1) and mobilize the applied N, which resulted in response to N fertilizer. One of the sites in Tripp country was not responsive to N fertilizer in 2020. The lack or response was attributed to the farmer planting an in-season cover crop. In the Tripp and Potter County sites in 2021 limited rainfall most likely limited growth and most likely resulted in the sites being non-responsive to N (Table 1-1). Maize yield at the Hughes county sites was responsive despite the low rainfall, which can be related to closer row spacing and high plants density compared to other sites (Table 1-1).

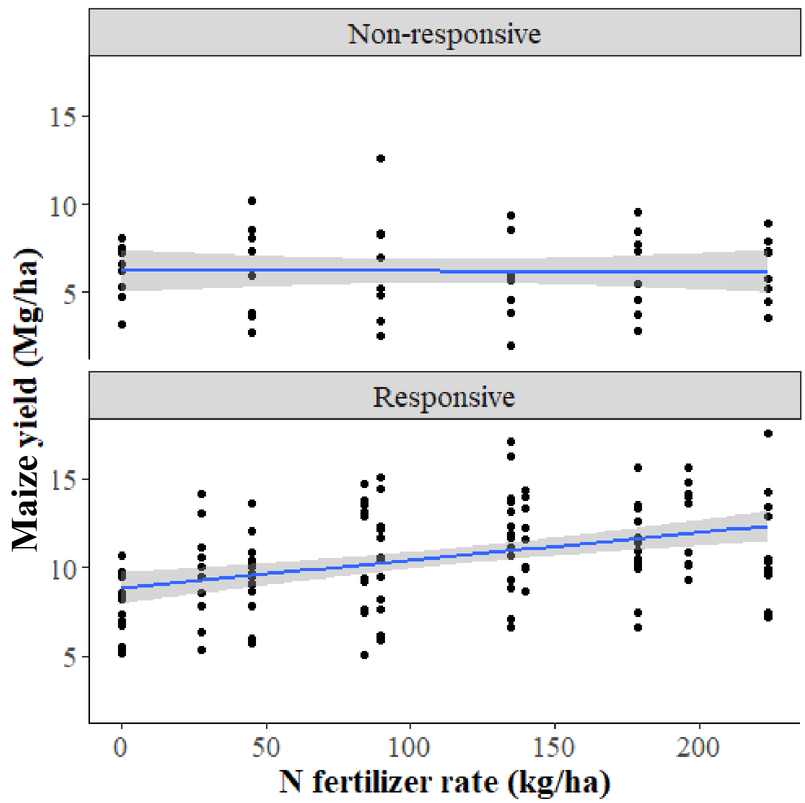


Figure 1- 3: Maize yield (Mg ha^{-1}) in response to fertilizer-N (kg ha^{-1}) between responsive and non-responsive sites averaged across three years. Each point in the plot represents the yield for each site.

Delta yield correlated to the EONR

The correlation among the maize yield at EONR, EONR, and delta yield were influenced by the fertilizer/maize price ratio (Figure 1-4). The relationship between the EONR and the maize yield at EONR were non-significant for all price ratios; however, the EONR showed strong positive correlation with delta yield. These findings were similar to Lory & Scharf (2003) and can help farmers identify if they need to apply more fertilizer to produce more maize economically. The EONR for each price ratio was highly correlated to delta yield at lower price ratios (4.11 and 5.48) as compared to the higher price ratios (6.85 and 8.23). Traditionally, the yield goal approach has been used in the corn producing areas of central United States. However, research suggests that modifications are needed (Derby et al., 2005; Kim et al., 2013; Lory & Scharf, 2003).

Table 1- 3: Maize yield (Mg ha^{-1}) at EONR (price ratio 4.11) with N responsiveness across experimental sites, 2019-2021.

Years	Site	Counties	Optimum Yield (Mg ha^{-1})	N Response
2019	BJO19	Tripp	11.68	Responsive
	DFO19	Potter	14.42	Responsive
	DHO19	Edmunds	10.03	Responsive
	DLD19	Hughes	13.48	Responsive
	SCA19	Kingsbury	13.9	Responsive
2020	BJC20	Tripp	7.27	Non-responsive
	BJO20	Tripp	10.69	Responsive
	DFO20	Potter	10.81	Responsive
	DLD20	Hughes	10.65	Responsive
	SCA20	Kingsbury	13.38	Responsive
2021	BJO21	Tripp	3.79	Non-responsive
	BSP21	Hand	6.75	Responsive
	DFC21	Potter	6.8	Non-responsive
	DFO21	Potter	7.45	Non-responsive
	DLD21	Hughes	5.81	Responsive
	SCA21	Kingsbury	14.15	Responsive

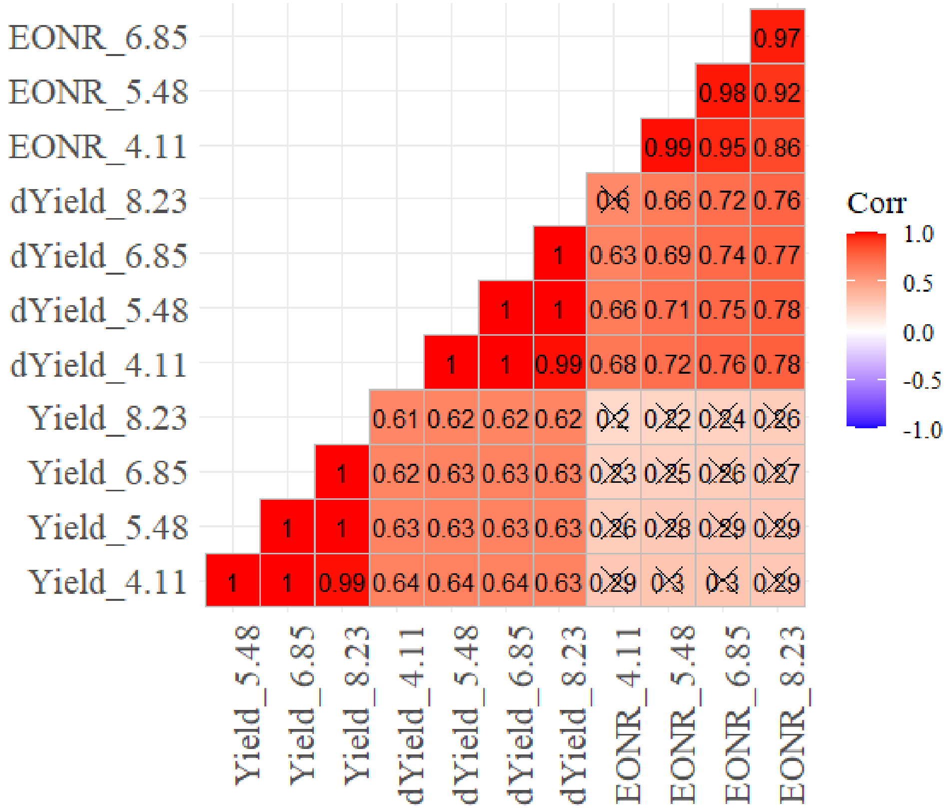


Figure 1- 4. Pearson correlation coefficients between maize yield, delta yield, and EONR at different fertilizer-N to maize cost ratios. The different cost ratios were 4.11, 5.48, 6.85, and 8.23. Correlation coefficients overlapped by "×" sign represents non-significant relationship at p= 0.05. In the figure, dYield' means delta yield and 'EONR' mean economic optimum N rate.

Nitrogen recommendation models across states

The RMSE and bias values were impacted by the fertilizer to maize price ratio as well as the recommendation model (Table 1-4). Historically, different states use different approaches to calculate the recommended fertilizer N, which is mostly targeted for tilled systems. We viewed a negative bias as not acceptable because it could be viewed as reducing yields. Most of the bias values (N recommendation – EONR) across four states were greater than zero meaning that the recommendations were higher than actual requirements.

Table 1- 4: Comparison of N recommendation models from South Dakota (SD), North Dakota (ND), western Minnesota (MN), Iowa (IA), and Nebraska (NE) using root mean square errors and bias. Data from all the fertilizer-N responsive dryland sites were included in this analysis.

Fertilizer/ maize price ratio	Root mean square errors (RMSE)						
	Historic SD Model	West MN Model	IA Model	NE Model	Updated SD Model	East ND Model	West ND Model
	kg N ha^{-1}						
4.11	102 (40)‡	95 (16)	97 (17)	97 (18)	94 (0)	107 (57)	92 (-19)
5.48	101 (51)	87 (13)	90 (15)	89 (17)	88 (11)	96 (46)	83 (-7)
6.85	101 (62)	79 (13)	82 (15)	80 (9)	82 (21)	88 (49)	73 (3)
8.23	103 (72)	72 (14)	77 (15)	73 (7)	80 (31)	77 (39)	67 (-6)

‡ Bias values in parentheses.

Table 1- 5: Economic optimum N rates (EONR) and the state N recommendations as influenced by the fertilizer/maize price ratio and the state recommendation models from South Dakota (SD), North Dakota (ND), western Minnesota (MN), Iowa (IA), and Nebraska (NE). Numbers inside the parentheses represent the confidence interval of N recommendations across the sites at α=0.05.

Price ratio	Nitrogen recommendations							
	Actual EONR	Historic SD Model	West MN Model	IA Model	NE Model	Updated SD Model	East ND Model	West ND Model
	kg N ha^{-1}							
4.11	174 (±51)	214 (±12)	190 (±12)	191 (±16)	192 (±15)	174 (±12)	230 (±3)	155 (±3)
5.48	162 (±47)	214 (±12)	175 (±11)	177 (±15)	179 (±13)	174 (±12)	211 (±3)	155 (±3)
6.85	152 (±42)	214 (±12)	166 (±9)	167 (±15)	161 (±12)	174 (±12)	201 (±3)	155 (±3)
8.23	142 (±38)	214 (±12)	156 (±8)	157 (±13)	149 (±11)	174 (±12)	181 (±3)	136 (±3)

For the historical South Dakota model, the RMSE values were similar across the price ratios with the lowest bias was for the 4.11 price ratio. In comparison to the other state recommendation models, the historical South Dakota model had the greatest RMSE and bias values because the model did not consider the fertilizer-N to the seed price ratio. This has resulted in the same N recommendations across all the price ratios for South Dakota (Table 1-5). Nitrogen recommendation calculated using the historical South Dakota model was higher than the other models.

Except for the lowest price ratio, the Western Minnesota model had lower RMSE values than the Eastern ND, Iowa and Nebraska models. The bias was marginally lower than the historic SD model but higher than the updated SD and Western ND models (Table 1-4). This model uses N recommendation based on the MRTN values subtracted from the soil test N from 60 cm soil depth which might be the reason for lower error rates. At the two lower fertilizer-to-corn price ratios, the Western Minnesota recommendation model had lower N recommendations at the 5.48 price ratios, as compared to the historic South Dakota, Iowa, and Nebraska models (Table 1-5).

The Iowa N recommendation model was similar to the Western Minnesota model as compared to the historic South Dakota model. The error rates and bias values were lower for the Iowa model in comparison with the historic South Dakota model (Table 1-4). This model uses the MRTN approach too; however, unlike the Western Minnesota model, it does not consider the soil test N level before planting. The Iowa N recommendation was similar to that with the Western Minnesota model at all the price ratios.

The Nebraska N recommendation model that considers soil organic matters and yield goal; is more complex than the other three state models. The RMSE values of the Nebraska model, at all price ratios, were comparable to those of the Western Minnesota and Iowa models; however, the bias value at the highest price ratio, 8.23, was the lowest among all the state recommendation models (Table 1-4). At this higher price ratio, the N recommendation using the Nebraska model was closer to the EONR. As this model considers the organic matter content, it can provide the N recommendation more precisely. Long-term no-till systems may have large amounts of organic matter accumulated on the soil surface because the surface residue is not harvested, and the soil is minimally disturbed. The lowest bias value at higher price ratio could be due to the amount of organic matter considered in the recommendation model. Expectedly, the N recommendation at 8.23 fertilizer-to-corn price ratio was 149 kg N ha^{-1}, respectively (Table 1-5).

Considering the importance of soil properties and management practices, the historic South Dakota and North Dakota N recommendation algorithms have been updated (Franzen, 2018). The updated South Dakota model reduced the amount of grain N requirements from 21.4 to 17.86 kg N Mg^{-1} grain, which is approximately 20% reduction in the grain fertilizer-N requirement. Compared to the historic model, the updated South Dakota model reduced the N recommendation from 214 kg N ha^{-1} to 174 kg N ha^{-1} (Table 1-5). The error rates and bias values at all price ratios were also reduced (Table 1-4).

Similarly, both Eastern and Western ND N recommendation model are based on the MRTN approach. The Eastern ND model, which is more relevant to our experimental sites, had higher RMSE and biases for the 4.11 fertilizer to grain value ratio than the updated SD model. These findings indicate that this model would overestimate the N recommendation in SD no-tillage fields, and therefore would not be considered as a climate smart practice. The western ND model had slightly lower RMSE values than the updated SD model. However, this model also had negative biases. Negative bias suggests that the model underestimated the N requirement (Table 1-4).

Usually, areas with higher productivity require less fertilizer-N as those areas are believed to be high in organic matter and soil moisture content (Franzen, 2018). As mentioned earlier, the Nebraska model considered soil organic matter in the equation because the mineralization of organic matter is correlated to higher yield (Shapiro et al., 2019). Kim et al. (2013) conducted a similar analysis of tilled system on experiments conducted between 2002 and 2006. This analysis showed that the biases were dependent on the fertilizer price to corn price ratio and generally increased with increasing ratio. Differences between Kim et al. (2013) and those reported in this study could be attributed to increasing corn yields, reducing reliance on fertilizer, and increased mineralization. Our research results suggested that the ND model might work the best for the long-term no-till farmers; however, the amount of fertilizer that needs to be adjusted in the equation may vary based on soil types and the environment (Franzen, 2018).

Although different N recommendation models from different states were based on various approaches, most of them did not consider soil organic matter, soil microbial activities, soil moisture content, or other management factors. Soil moisture content plays important roles in the mineralization of organic matter and transportation of plant available-N. Approximately 60% water-filled pore space is required to gain the maximum mineralization driven by soil microbes (Linn & Doran, 1984; Wang et al., 2004). Soil microbial activities and N mineralization is highly influenced by the temperature (Andersson & Nilsson, 2001; Bell et al., 2008). In addition, soil management factors such as soil tillage (Lal, 1993), and cover crops (Blanco-Canqui et al., 2015) are some of the most important factors influencing the soil N availability.

Based on our current study, the EONR calculated at the lowest price ratio matched with the N recommended by the updated SD model. The higher positive bias at higher price ratios using the updated SD model suggested considering the price ratios in the model. Table 1-6 shows the maize N recommendations for long-term no-till maize for South Dakota based on our current study.

Table 1- 6: Maize N recommendations (kg N ha^{-1}) for long-term no-till maize for South Dakota based on the current study, considering N cost and maize grain price.

$/kg N	**$/kg maize**								
	0.1	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**	**0.45**	**0.5**
	kg N ha^{-1}								
1	128	154	167	175	180	184	187	189	191
1.5	88	128	147	159	167	173	177	180	183
2	49	101	128	143	154	161	167	171	175
2.5	9	75	108	128	141	150	157	163	167
3	0	49	88	112	128	139	147	154	159
3.5	0	23	68	96	114	128	137	145	151
4	0	0	49	80	101	116	128	136	143
4.5	0	0	29	65	88	105	118	128	135
5	0	0	9	49	75	94	108	119	128

CONCLUSIONS

Our results suggested that, considering the historic (10-year average) and the lowest price ratio (4.11), the updated South Dakota N recommendation model provided a lower EONR estimate and was supported by the lowest bias and RMSE. The bias and error rates were high at higher price ratios for the updated SD model. Farmers need to consider N cost and maize grain price to calculate the recommended maize N rate. Table 1-5 suggests maize N recommendation based on our study in long-term no-till sites.

Considering the fertilizer price ratio of the past two years (price ratio 8.23), the Nebraska model estimated the lowest EONR. This indicates that to offset the raised fertilizer cost, N recommendation model needs to consider soil organic matter as well as the price ratio as an input to the model. Further, improving soil health, reduced erosion, and increased water retention in long-term no-tillage systems can supply additional N to the crops. This can account for the mineralizable N available to the crops during the growing season. Additional work needs to be conducted to integrate the fertilizer to corn value ratio into the recommendation.

LITERATURE CITED

Andersson, S., & Nilsson, S. I. (2001). Influence of pH and temperature on microbial activity, substrate availability of soil-solution bacteria and leaching of dissolved organic carbon in a mor humus. *Soil Biology and Biochemistry*, *33*(9), 1181-1191.

Andraski, T. W., & Bundy, L. G. (2002). Using the presidedress soil nitrate test and organic nitrogen crediting to improve corn nitrogen recommendations. *Agronomy Journal*, *94*(6), 1411-1418.

Bell, C., McIntyre, N., Cox, S., Tissue, D., & Zak, J. (2008). Soil microbial responses to temporal variations of moisture and temperature in a Chihuahuan Desert grassland. *Microbial ecology*, *56*(1), 153-167.

Bhattarai, D., Kumar, S., & Nleya, T. (2021). Nitrogen and sulfur fertilizers effects on growth and yield of Brassica carinata in South Dakota. *Agronomy Journal*, *113*(2), 1945-1960.

Bhattarai, D., Abagandura, G. O., Nleya, T., & Kumar, S. (2021). Responses of soil surface greenhouse gas emissions to nitrogen and sulfur fertilizer rates to Brassica carinata grown as a bio-jet fuel. *GCB Bioenergy*, 13(4), 627-639.

Blanco-Canqui, H., Shaver, T., & Lindquist, J. (2015). Cover Crops and Ecosystem Services: Insights from Studies in Temperate Soils. Agronomy & Horticulture. In: *Faculty Publications*.

Blanco-Canqui, H., & Ruis, S. J. (2018). No-tillage and soil physical environment. *Geoderma*, 326, 164-200.

Clark, J. D., Veum, K. S., Fernández, F. G., Camberato, J. J., Carter, P. R., Ferguson, R. B., ... & Shanahan, J. F. (2019). United States Midwest soil and weather conditions influence anaerobic potentially mineralizable nitrogen. *Soil Science Society of America Journal*, 83(4), 1137-1147.

Clark, J. D. (2019). Fertilizer recommendations guide. Original Gerwig, J. & Gelderman, R. (2005). https://extension.sdstate.edu/sites/default/files/2019-03/P-00039_0.pdf

Clark, J. D. (2023). Fertilizer recommendations guide. Original Gerwig, J. & Gelderman, R. (2005). https://extension.sdstate.edu/sites/default/files/2023-03/P-00039-2023.pdf

Clay, D., Kim, K. I., Chang, J., Clay, S., & Dalsted, K. (2006). Characterizing water and nitrogen stress in corn using remote sensing. *Agronomy Journal*, *98*(3), 579-587.

Clay, D., Malzer, G., & Anderson, J. (1990). Ammonia volatilization from urea as influenced by soil temperature, soil water content, and nitrification and hydrolysis inhibitors. *Soil Science Society of America Journal*, *54*(1), 263-266.

CNRC. (2022). Corn nitrogen rate calculator. http://cnrc.agron.iastate.edu/About.aspx

Cotrufo, M. F., Ranalli, M. G., Haddix, M. L., Six, J., & Lugato, E. (2019). Soil carbon storage informed by particulate and mineral-associated organic matter. *Nature Geoscience*, *12*(12), 989-994.

Derby, N. E., Steele, D. D., Terpstra, J., Knighton, R. E., & Casey, F. X. (2005). Interactions of nitrogen, weather, soil, and irrigation on corn yield. *Agronomy Journal*, *97*(5), 1342-1351.

Fox, R. H., & Piekielek, W. P. (1987). Yield response to N fertilizer and N fertilizer use efficiency in no-tillage and plow-tillage corn. *Communications in Soil Science and Plant Analysis*, *18*(5), 495-513.

Franzen, D. W. (2018). North Dakota Fertilizer Recommendation Tables and Equations. *North Dakota State University Extension Service*. https://www.ndsu.edu/fileadmin/soils.del/pdfs/North_Dakota_Fertilizer_Recommendation_Tables_and_Equations_SF882.pdf

Kaiser, D. E., Fernandez, F., Wilson, M., Coulter, J. A., & Piotrowski, K. (2022). Fertilizing corn in Minnesota. *Regents of the University of Minnesota.* Available online: https://extension. umn. edu/crop-specific-needs/fertilizing-corn-minnesota (accessed on 8 April 2022). https://drive.google.com/file/d/1f3-BrF4bD5EMu4pq4ofCQXUbgnzxeTwy/view

Kim, K. I., Clay, D., Clay, S., Carlson, G. C., & Trooien, T. (2013). Testing corn (Zea mays L.) preseason regional nitrogen recommendation models in South Dakota. *Agronomy Journal*, *105*(6), 1619-1625.

Kim, K. I., Clay, D. E., Carlson, C., Clay, S., & Trooien, T. (2008). Do synergistic relationships between nitrogen and water influence the ability of corn to use nitrogen derived from fertilizer and soil? *Agronomy Journal*, *100*(3), 551-556.

Lal, R. (1993). Tillage effects on soil degradation, soil resilience, soil quality, and sustainability. *Soil and tillage Research*, *27*(1-4), 1-8.

Linn, D. M., & Doran, J. W. (1984). Effect of water-filled pore space on carbon dioxide and nitrous oxide production in tilled and nontilled soils. *Soil Science Society of America Journal*, *48*(6), 1267-1272.

Lory, J., & Scharf, P. (2003). Yield goal versus delta yield for predicting fertilizer nitrogen need in corn. *Agronomy Journal*, *95*(4), 994-999.

Morris, T. F., Murrell, T. S., Beegle, D. B., Camberato, J. J., Ferguson, R. B., Grove, J., Ketterings, Q., Kyveryga, P. M., Laboski, C. A., & McGrath, J. M. (2018). Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agronomy Journal*, *110*(1), 1-37.

Rodriguez, D. G. P., Bullock, D. S., & Boerngen, M. A. (2019). The origins, implications, and consequences of yield-based nitrogen fertilizer management. *Agronomy Journal*, *111*(2), 725-735.

Sainju, U. M., Ghimire, R., & Pradhan, G. P. (2019). Nitrogen fertilization I: Impact on crop, soil, and environment. *Nitrogen Fixat*, 9, 1-9.

Sawyer, J., Nafziger, E., Randall, G., Bundy, L., Rehm, G., & Joern, B. (2006). Concepts and rationale for regional nitrogen rate guidelines for corn. *Iowa State University-University Extension, Ames, Iowa*, *28*.

Shapiro, C. A., Ferguson, R., Wortmann, C. S., Maharjan, B., & Krienke, B. T. (2019). Nutrient management suggestions for corn. *University of Nebraska-Lincoln Extension*. https://extensionpublications.unl.edu/assets/pdf/ec117.pdf

St. Luce, M., Ziadi, N., Chantigny, M. H., & Braun, J. (2021). Long-term effects of tillage and nitrogen fertilization on soil C and N fractions in a corn–soybean rotation. *Canadian Journal of Soil Science*, 102(2), 277-292.

Thies, S., Joshi, D. R., Bruggeman, S. A., Clay, S. A., Mishra, U., Morile-Miller, J., & Clay, D. E. (2020). Fertilizer timing affects nitrous oxide, carbon dioxide, and ammonia emissions from soil. *Soil Science Society of America Journal*, 84(1), 115-130.

Tremblay, N., Bouroubi, Y. M., Bélec, C., Mullen, R. W., Kitchen, N. R., Thomason, W. E., Ebelhar, S., Mengel, D. B., Raun, W. R., & Francis, D. D. (2012). Corn response to nitrogen is influenced by soil texture and weather. *Agronomy Journal*, *104*(6), 1658-1671.

Wang, W., Smith, C., & Chen, D. (2004). Predicting soil nitrogen mineralization dynamics with a modified double exponential model. *Soil Science Society of America Journal*, *68*(4), 1256-1265.

CHAPTER 2: SOIL HEALTH MEASUREMENTS IMPACT MAIZE YIELD IN NO-TILL SYSTEMS- PREDICTIONS USING MACHINE LEARNING APPROACHES.

ABSTRACT

Farmers have reported maize (*Zea mays* L.) requires less N in long-term no-tillage than conventional management. This apparent reduction may be attributed to improved soil health resulting from increasing soil organic matter, higher soil microbial activities, and improved water and nutrient use efficiency. However, the impact of soil health measurements on fertilizer-N requirement has not been determined. This project aims to assess the feasibility of machine learning algorithms that use soil health measurements to forecast yield predictions. The research was conducted for three years at 16-dryland sites that were under no-tillage practice for at least 6-years. The effect of six N rates (0, 45, 90, 135, 180, and 224 kg N ha^{-1}) on maize grain yield was evaluated. Predictor variables for soil yield potential consisted of climatic data, and soil health measurements including nitrate-N (NO_3-N), ammonium-N (NH_4-N), pH, EC, soil respiration, and soil microbial biomass that were collected from various depths before planting. Positive correlations of maize yield were seen with water infiltration, soil respiration, and soil microbial biomass. Support vector machine (SVM) algorithm outperformed several other machine learning algorithms for forecasting the soil yield potential. The top five predictor variables were total N, total C, growing degree days (GDD), soil microbial biomass, and bacterial biomass. The overall findings from this study suggested that soil health measurements in addition to the climate variables, rainfall, and temperature, can be used to predict the soil yield potential.

INTRODUCTION

Soil scientists have struggled for over 50 years to create techniques that can predict biological activity and microbial mineralization. This research tested the ability of chemical extraction and biological incubations as tools for forecasting the soil N supplying power. Despite improving our understanding of N cycling, most N recommendation models do not consider biological activity. For example, the maximum return to N approach recommends a fixed N application rate does not consider the ability of soil to mineralize N. The lack of integrating soil biological activity into many N recommendation is not consequence for the lack of trying, but it is the result of that we do not understand how to integrate this information into the recommendation (Franzluebbers, 2018; Yost et al., 2018).

An alternative approach is to consider another component of most N models, yield. There are numerous studies that have created yield forecasts using some combination of historical measurement, in-season measurement, and modeling (Basso & Liu, 2019; Rosenzweig et al., 2013). However, because historical measurements may not be useful as farmers transition to climate smart practices and because many agronomists are often not willing to invest the time to learn how to use process-based models, alternative approaches are needed (Drummond et al., 2003; Geisseler & Wilson, 2020; Ngwira et al., 2014; Puntel et al., 2019). Machine learning algorithms based on soil health measurements can help fill this gap (Dhaliwal et al., 2022; Joshi et al., 2022).

A review on crop yield prediction by Van Klompenburg et al. (2020) found that temperature, rainfall, and soil type were the most widely used variables to predict crop yield. Although including soil biotic parameters such as soil microbial biomass, soil respiration, and other related variables to directly predict crop yield is not common, studies have shown that soil bacterial communities can improve soil quality and can predict soil properties like soil pH, bulk densities, and nutrient concentrations (Hermans et al., 2020). Previous research showed that soil health provide a good assessment of the soil and that the implementation of soil health practices (cover crops, reduced tillage, and rotations) may have a neutral to negative impact on yield (Miner et al., 2020). However, this analysis may not consider the impact of soil health improvement on the ability of a soils resilience to adverse climatic conditions, such as drought. For example, Clay et al. (2014) showed that increasing soil organic matter improved the soil productivity in the 2012 drought and that the adoption or cover crops and reduced tillage increased SOM (Clay et al., 2012; Joshi et al., 2023). Therefore, given that: 1) machine learning techniques can quickly process a large amount of data, 2) N recommendations are a function of the soil supplying power and yield potential, and 3) biological information is rarely considered in N recommendations, our goal is to assess the feasibility of machine learning algorithms that uses soil health measurements to forecast yield predictions.

MATERIALS AND METHODS

Experimental details

No-till maize on-farm experiments were conducted in 2019, 2020, and 2021. The study sites were in long-term no-till fields (>6 years no-till). Over three years, the experiment was conducted in 16 dryland sites in South Dakota (SD) (Figure 2-1). All the locations are characterized as hot summer humid continental climate (Köppen climate: Dfa). Agronomic details of each site are shown in Figure 2-1. Average temperature and cumulative precipitation information can be found in Supplementary Figure 1.

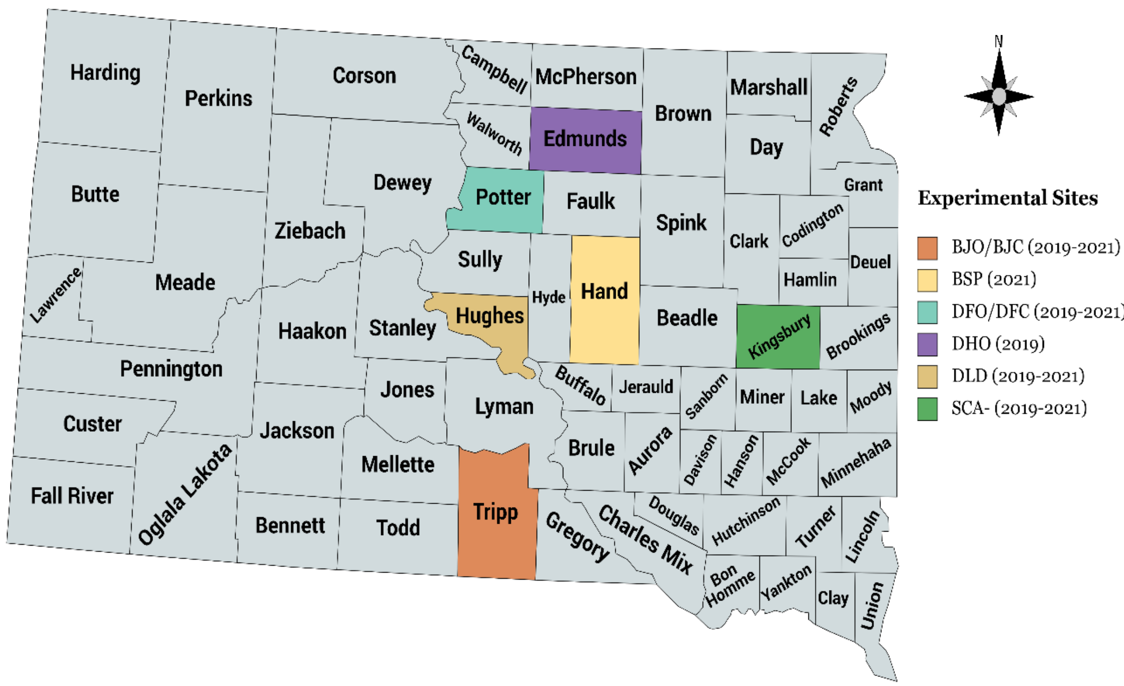


Figure 2- 1: The map of South Dakota highlighted with the counties where field experiments were conducted from 2019-2021. Each color in the legend represents the experimental field names with the different years of experiment.
Source: https://www.mapchart.net/usa-counties.html

The experimental design was a randomized complete block design (RCBD) with four N rates (28, 84, 140, and 196 kg N ha^{-1}) in 2019 and six N rates (0, 45, 90, 134, 179, and 224 kg N ha^{-1}) in 2020 and 2021 with four replications. Urea (46-0-0) fertilizer was applied as the source of N. Nitrogen treatments were manually broadcast applied between the V2 and V4 maize growth stages. Each plot had dimensions of 15.24 × 4.6 m (50 ft × 15 ft).

Maize was planted by the farmers and most of the sites had 75 cm row spacing. One site in 2020 (BJC 2020), was planted with cover crops at 150 cm row spacing and cover crops were planted at maize V3 growth stage. The cover crop species planted were oats (*Avena sativa*, L., 13 kg ha^{-1}), flax (*Linum usitatissimum*, L., 2.2 kg ha^{-1}), mung bean

(*Vigna radiata*, L., 4.5 kg ha^{-1}), guar (*Cyamopsis tetragonoloba*, L. 2.2 kg ha^{-1}) and red clover (*Trifolium pratense* var. *sativum* (Schreb.), 1.1 kg ha^{-1}). All the sites in Hughes County were planted at 50 cm row spacing. Phosphorus and K fertilizers were applied based on the soil test results (Clark, 2019). Soil characteristics information is shown in Supplementary Table 1. Maize ears were hand-harvested from 9.29 m^2 area marked from the center of each plot and calculated the final corn yield at 15.5% moisture.

Soil sampling and measurements

Baseline soil samples from four blocks were collected from four depths (0-5, 5-15, 15-30, and 30-60 cm) before planting maize and before the treatment application at each site. Plant residues on the soil surface were carefully removed and the samples were randomly collected from 15-20 random spots within each block using a standard soil probe with a 1.9 cm inner diameter. Soil samples, for each depth per block, were mixed thoroughly, air-dried to a constant weight, and sieved through a 2-mm mesh screen before soil analyses. Soil inorganic N, nitrate-N (NO_3-N) and ammonium-N (NH_4-N), were extracted using 1M KCl (1:10 soil to KCl ratio) (Kim et al., 2008), quantified by cadmium reduction method (Clark et al., 2019) analyzed using Astoria Analyzer (Astoria-Pacific). Soil pH and electrical conductivity were determined using pH and EC meter (Mettler Toledo). Total N, total C including d^{13}C and d^{15}N isotopes were analyzed

Table 2- 1: Agronomic information including planting and harvest date, corn varieties and their maturity days, previous crops, and plant population of different experiment sites across the experiment years, 2019-2021.

Field name	**Counties**	**Planting Date**	**Harvest Date**	**Maize Maturity Days**	**Previous crop(s)**	**Plant population** (per ha)
BJO19	Tripp	5/15/2019	10/19/2019	106	Wheat	60500
DFO19	Potter	5/14/2019	10/08/2019	101	Wheat	62000
DHO19	Edmunds	5/26/2019	10/14/2019	88	Soybean	65000
DLD19	Hughes	5/15/2019	10/15/2019	105	Wheat	52000
SCA19	Kingsbury	5/16/2019	10/18/2019	97	Soybean	79000
BJC20	Tripp	4/29/2020	10/01/2020	99	Wheat	60500
BJO20	Tripp	4/29/2020	10/01/2020	100	Wheat	60500
DFO20	Potter	5/11/2020	10/06/2020	96	Wheat	62000
DLD20	Hughes	4/30/2020	9/26/2020	99	Oats + Barley	52000
SCA20	Kingsbury	4/27/2020	10/08/2020	97	Soybean	79000
BJO21	Tripp	5/3/2021	10/18/2021	99	Wheat-fallow-livestock	59000
BSP21	Hand	5/4/2021	10/5/2021	102	Fallow	69000
DFC21	Potter	5/4/2021	10/18/2021	101	Cover crops mix	65500
DFO21	Potter	5/5/2021	10/12/2021	102	Wheat-fallow-livestock	64000
DLD21	Hughes	5/6/2021	10/7/2021	100	Wheat	86500
SCA21	Kingsbury	4/30/2021	9/28/2021	105	Soybean	79000

from air dried soil using mass spectrometry (Clay et al., 2015). The C3 plants have $d^{13}C$ values between -18 to -23 and C4 plants have values from -12 to -14.

Bulk densities were determined by drying the soil at 105ºC for 48 hours. Soil respiration was measured using Solvita burst test (Haney et al., 2008) and expressed in the form of CO_2-C. Water infiltration rate was measured using in-situ steady state double ring infiltration method (Bodhinayake et al., 2004).

Soil samples from 0- to 5-cm soil depth were collected and analyzed for soil microbial community structure using Veum et al. (2019). Phospholipid fatty acids were extracted following a modified protocol described by Buyer & Sasser, (2012), Fiedler et al., (2021), and Joshi et al. (2022) using a 19:0 phosphatidylcholine internal standard for PLFA and a check sample to confirm the final values. Extracts were analyzed using a Shimadzu GC-2010 Plus gas chromatograph (Shimadzu Corporation, Japan) using a flame ionization detector. The gas chromatograph was calibrated using a calibration standard provided by MIDI Sherlock (No. 1208, MIDI, Inc., Newark, DE) using PLFAD2 method.

The extracted fatty acids were characterized into different microbial groupings using the MICSOILV2 method from MIDI Sherlock Software system (MIDI, Inc., Newark, DE). The Sherlock PLFA Analysis Software determines abundance and type of microbial community by assigning fatty acids into different functional groups associated with each community type (Veum et al., 2019).

Table 2- 2: Predictor variables used for machine learning algorithms from pre-plant soil samples.

Variables	Description	Unit	Data source
N rate	Fertilizer-N rates applied	kg N ha^{-1}	Data collection
NO_3-N	Soil nitrate-N, 0-15 cm	kg N ha^{-1}	Data collection
NH_4-N	Soil ammonium-N, 0-15 cm	kg N ha^{-1}	Data collection
TIN	NO_3-N + NH_4-N, 0-15 cm	kg N ha^{-1}	Data collection
$pH_{1:1}$	Soil pH, 0-15 cm		Date collection
$EC_{1:1}$	Soil electrical conductivity, 0-15 cm	µS cm^{-1}	Data collection
Clay	Clay percent, 0-15 cm	%	Web Soil Survey
Bacteria	Soil bacterial biomass, 0-5 cm	µg C g^{-1} soil	Data collection
Fungi	Soil fungal biomass, 0-5 cm	µg C g^{-1} soil	Data collection
AMF	Soil arbuscular mycorrhizal fungi biomass, 0-5 cm	µg C g^{-1} soil	Data collection
Biomass	Soil microbial biomass, 0-5 cm	µg C g^{-1} soil	Data collection
OM	Soil organic matter, 0-15 cm	%	Data collection
CO_2-C	Solvita soil respiration	mg kg^{-1}	Data collection
Infiltration	Infiltration rate	mm hr^{-1}	Data collection
PPT	Cumulative precipitation from planting to tasseling growth stage	cm	Mesonet, SDSU
GDD	Cumulative growing degree days at maize harvest, base 10ºC		Mesonet, SDSU
TN	Organic + inorganic N, 0-5 cm	%	Data collection
$d^{15}N$	^{15}N content, 0-5 cm	‰	Data collection
TC	Total carbon content, 0-5 cm	%	Data collection
$d^{13}C$	^{13}C content, 0-5 cm	‰	Data collection
Soil_water	Soil water depth, 0-15 cm	cm	Data collection

Data generation for machine learning algorithms

Maize yield, from each individual plot, from the dryland experimental sites across three years were used as the response variable in the dataset. The dataset included 21 predictor variables (soil health measurements) and yield as a response variable (Table 2-2). The dataset was divided into training and testing datasets. The randomly selected training portion contained 70% of the whole dataset, whereas the testing components contained 30%.

Machine learning algorithms

Maize yield for each individual plot was predicted using five different algorithms that included linear regression, two regularizations of linear regression- ridge and LASSO regressions, random forest, and support vector machine (SVM). These models are commonly used to predict crop yields (Dhaliwal et al., 2022; Joshi et al., 2022; Ransom et al., 2019). All the model development and validation process was performed using R software (R Core Team, 2022). A “set.seed” function with a value of “123” was used to make the predictions reproducible.

Linear relationship between the response and predictor variables can be determined using linear regression (Montgomery et al., 2021). These models make several assumptions including linear relationships, constant variance, and little to no multicollinearity between the predictor variables. In linear regression, the chance of getting biased prediction with high error is common due to the simplicity of the model. Ridge and LASSO (least absolute shrinkage and selection operator) regressions are the regularizations of linear estimates (James et al., 2013) that reduce model complexity and prevent the model from obtaining high variances. The regularization is implemented by adding a penalty equivalent to square of the magnitude of the coefficients in ridge regression, whereas in LASSO regression, magnitude is considered instead of square of magnitude. They result in a reduced magnitude of the coefficients while the number of features remain constant in the training dataset.

Random forest is a robust machine learning approach that is based on the recursive partitioning principle and can be used for regression and classification purposes (Breiman, 2001; Ransom et al., 2019). This algorithm does not require specific information about the relationship between the response and predictor variables. Strengths of the RF approach is that the accuracy and robustness of the model generally increases with the number of trees in the forest. However, a weakness of the RF approach is that it can over smooth predictions when the training datasets are relatively small (Koparan et al., 2022).

Support Vector Machine (SVM) is a machine learning algorithm that separates the data into different classes using a line or a hyperplane. This approach was selected because it can solve non-linear prediction problems (Ahmad et al., 2014). The line or hyperplanes represent decision boundaries and are often used to classify continuous outputs (Brereton & Lloyd, 2010). However, due to the computational time requirement, the method may not be well suited for large data sets. Additional information on this approach is available in Cortes & Vapnik (1995).

Models' performance

All the machine learning algorithms were tuned using different hyperparameters. The training dataset (70% of the original dataset) was validated using 10-fold cross validation technique, which splits the data into 10 folds, estimates the error rates, and generates a model with the lowest error rate (Refaeilzadeh et al., 2009). The models' performance was evaluated by root mean squared error (RMSE), goodness of fit (R^2), and mean absolute error (MAE) using Equations 1-3.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_p)^2}{n}} \qquad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_p)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^{n}|(y_i - y_p)|}{n} \quad (3)$$

In equations above, y_i and y_p are measured and predicted maize yield values, respectively, $\bar{y}_i$ is the mean of all measured yield, and *n* is the number of samples. The best model was selected based on the lowest error rates and greater goodness of fit.

Software and graphics

Statistical analyses including machine learning algorithms were run and analyzed using R programming software (R Core Team, 2022). The machine learning algorithms were built using “caret” package (v 6.0-88) in RStudio. All the graphics were finalized using GraphPad Prism (v8.2, GraphPad software, LLC).

RESULTS AND DISCUSSION

Correlation across variables

Farmers are interested in learning and incorporating information about soil health into estimates about the soil’s yield potential. Current approaches for estimating the soil yield potential are based on averaging previous yields and do not consider how management changes affect soil health and the yield potential (Reitsma et al., 2016; Shapiro, 2008). To assess the importance of different chemical and biological measurements a correlation analysis (Figure 2-2) was conducted between the measured values (Table 2-2). Positive correlations between microbial biomass and Solvita soil respiration suggest that increasing microbial activity had a positive impact on maize yields. Others have reported similar responses (Franzluebbers et al., 2000; Haney et al., 2008). The positive correlation between precipitation and yield and water infiltration and yield suggests that water availability limited yields. The impacts of water availability on yield have been reported numerous times (Kim et al., 2008; Ransom et al., 2021).

Maize yield was negatively correlated with several measurements including soil $pH_{1:1}$ and $EC_{1:1}$. At the study sites, $pH_{1:1}$ ranged from 5.8 to 8. Over this range of values, increasing pH reduced yields. This would be expected and could be attributed to many factors including reduced nutrient availability. For $EC_{1:1}$, the range of values was from 0.13 to 2.3 dS/m. Based on these values, soils with an $EC_{1:1}$ of 2 or more would be considered as saline soils, and therefore yield reductions with an increasing $EC_{1:1}$ would be expected.

Many of the measured values were correlated to each other (Figure 2-2). For example, the water infiltration rate was highly correlated with clay content, fungi, and arbuscular mycorrhizal fungi (AMF) biomass, whereas total N and C were correlated with soil bacterial biomass, and NO_3-N, and CO2-C respiration was correlated to water infiltration, microbial biomass, mainly bacterial biomass. The positive correlation between water infiltration rate and fungal biomass indicates that in these long-term no-tillage fields building the fungal population had a positive impact on water infiltration.

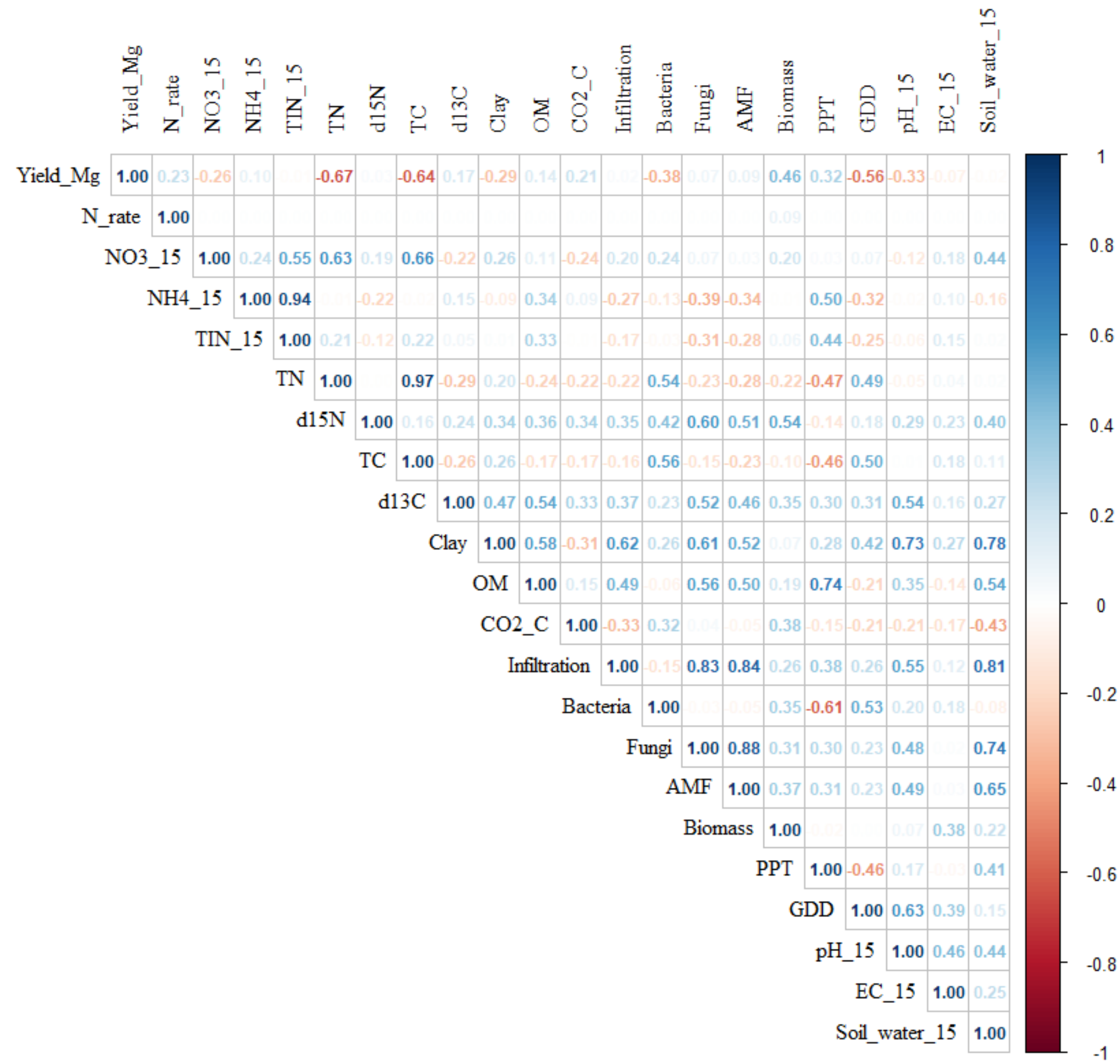


Figure 2-2: Correlation across response and predictor variables. Nitrate and ammonium N are the amounts in the surface 15 cm, TN and TC are the concentration of organic N and C in the soil, d13C is ^{13}C isotope in ‰, d15N is ^{15}N isotope in ‰, clay is g clay kg^{-1} soil, OM is organic matter in g kg^{-1} soil, CO2-C is Solvita soil respiration in mg C kg^{-1} soil, infiltration is the rate of water infiltration in cm h^{-1}, bacteria, fungi, AMF and biomass are in µg C g^{-1} soil as measured using the PLFA technique, PPT is cumulative precipitation at tasseling stage in cm, GDD is growing degree days, pH is soil pH at 1:1, EC is soil electrical conductivity at 1:1 solution in µS cm^{-1}, and soil water is soil water depth at surface 15 cm measured in cm.

Machine learning model performance

The performance of five different machine learning algorithms that predict maize yield on training, and testing datasets are summarized in Table 2-3. The table showed that the random forest training algorithms had the highest R^2 and lowest RMSE compared to the other regression models in the training dataset; however, the SVM algorithms had the highest R^2 and lowest RMSE for the testing dataset.

The performance of all three regression models were similar with similar R^2 values for each dataset. The R^2 value on the training dataset for random forest and SVM models were 0.96 and 0.94, respectively. However, in the testing dataset, the R^2 values and the error rates for the SVM were 0.93, 0.84 (RMSE), and 0.59 (MAE), respectively, whereas the random forest model had lower R^2 value and higher error rates as compared to the SVM (Table 2-3).

Table 2- 3: Machine learning models performance ability to predicate maize yields at each N rate. The training data was used to build the models and the testing data was independent data that was not included in creating the models. Comparison among linear regression, ridge regression, LASSO regression, random forest, and support vector machine (SVM) using root mean squared error (RMSE), mean absolute error (MAE), and goodness of fit (R^2) in the training, and testing datasets are shown.

Models	Training data			Testing data		
	RMSE	MAE	R^2	RMSE	MAE	R^2
	Mg ha^{-1}	Mg ha^{-1}		Mg ha^{-1}	Mg ha^{-1}	
Linear regression	1.12	0.80	0.89	1.23	0.99	0.85
Ridge regression	1.38	1.08	0.83	1.41	1.24	0.80
LASSO regression	1.12	0.80	0.89	1.22	0.98	0.85
Random Forest	0.61	0.40	0.96	1.17	0.79	0.86
SVM	0.80	0.48	0.94	0.84	0.59	0.93

Linear regression and its regularization models had larger training and testing error compared to the other models; this suggested low variance and highly biased predictions. Although the training data error was small using the random forest model, larger error rate in testing dataset suggested higher possibility of low bias and high variance (overfitting) conditions.

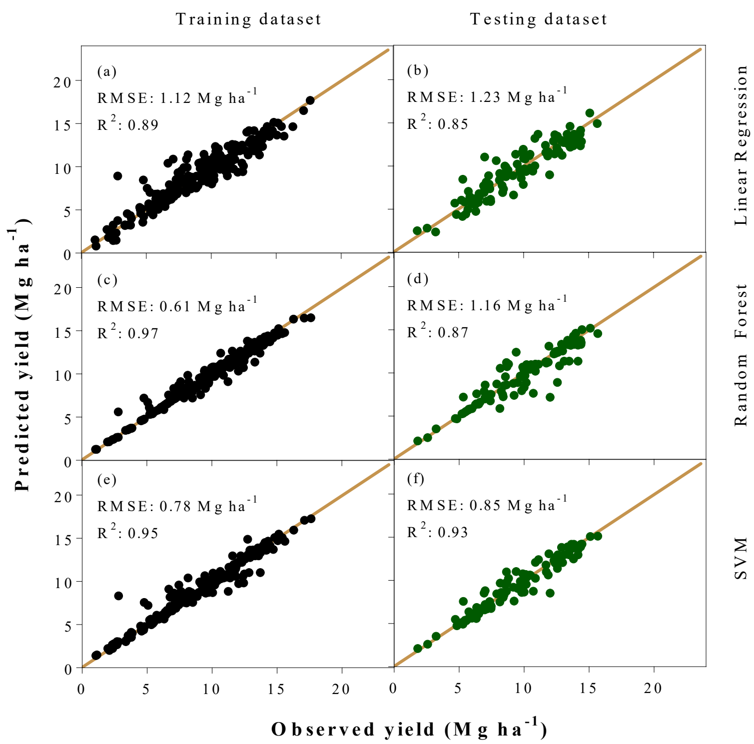


Figure 2- 3: Observed vs. predicted maize yield (Mg ha^{-1}) plots for Linear regression a) training dataset, b) testing dataset; Random forest c) training dataset, d) testing dataset, and support vector machine (SVM) e) training dataset, f) testing dataset. The brown solid lines indicate 1:1 relation between the observed and predicted yields.

Scatterplots of predicted maize yield by linear regression, random forest, and the SVM on training and testing datasets are in Figure 2-3. The brown line in the figure indicates that there was a 1:1 relationship between the observed and predicted maize yields. Deviation of the data points from the 1:1 line (Figures 2-3a, 2-3b) suggested that the maize yields were generally underestimated. The random forest model underestimated the predicted yield at higher values for the testing datasets (Figures 2-3c, 2-3d) and the error rates and goodness of fit for the training dataset were the best compared to any other models. The testing dataset for SVM showed the best fit across all three datasets and among all machine learning models (Figures 2-3e, 2-3f).

Predictor variable performance

The importance of predictor variables was determined using the best model fit, the SVM (Figure 2-4). Interestingly, total soil N was the most influencing factor to predict the maize yield followed by total C, growing degree days, soil microbial biomass, and bacterial biomass. Of these measurements, total C and N and microbial biomass would be considered as soil health measurements. It was interesting that total C was a better predictor than soil organic matter. There were many measurements that were only minimally important, several of these variables were Solvita soil respiration, N rate, EC, fungal biomass, and water infiltration. These findings were contrary to the correlation analysis discussed earlier. Different results between machine learning and classical statistics are attributed to machine learning approaches behaving as a black box.

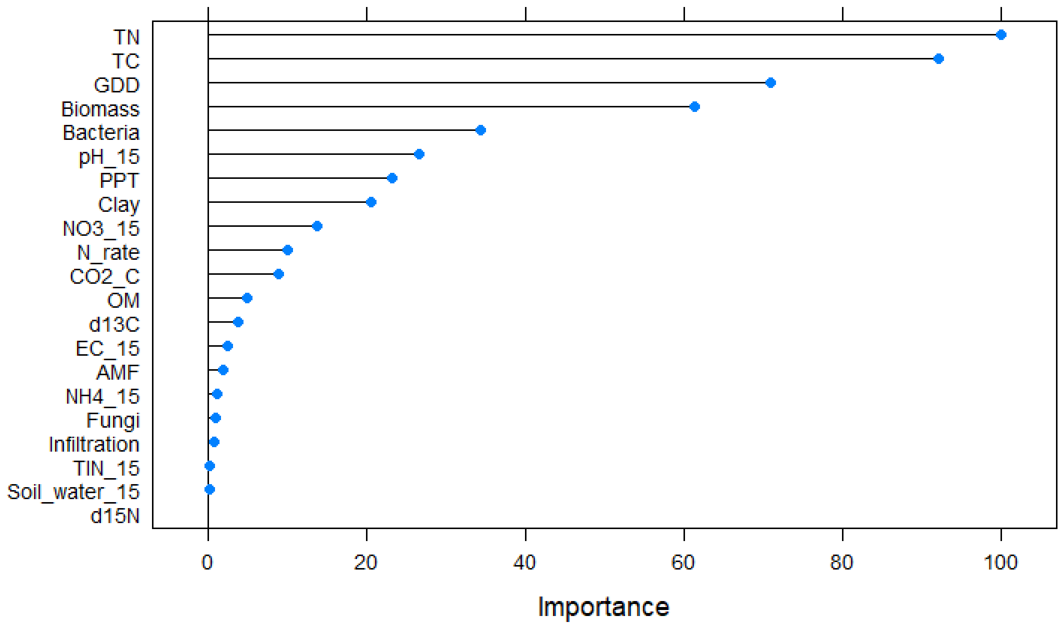


Figure 2- 4: Variable importance plot based on support vector machine algorithms. X-axis represents relative importance on a scale of 100 and Y-axis shows the different predictor variables. Nitrate and ammonium N are the amounts in the surface 15 cm, TN and TC are the concentration of organic N and C in the soil, d13C is ^{13}C isotope in ‰, d15N is ^{15}N isotope in ‰, clay is g clay kg^{-1} soil, OM is organic matter in g kg^{-1} soil, CO2-C is Solvita soil respiration in mg C kg^{-1} soil, infiltration is the rate of water infiltration in cm h^{-1}, bacteria, fungi, AMF and biomass are in µg C g^{-1} soil as measured using the PLFA technique, PPT is cumulative precipitation at tasseling stage in cm, GDD is growing degree days, pH is soil pH at 1:1, EC is soil electrical conductivity at 1:1 solution in µS cm^{-1}, and soil water is soil water depth at surface 15 cm measured in cm.

Several efforts have been made to model crop yield in response to fertilizer (Dhakal & Lange, 2021); however, the soil health measurements were always overlooked. Earlier researchers have shown that addition of soil surface residue can improve soil physical, chemical, and biological properties (Clay et al., 2015; Turmel et al., 2015). Implementation of machine learning algorithms in yield modeling have provided more flexibility to the researchers in terms of variable selection and model performance testing (Lischeid et al., 2022; Shahhosseini et al., 2021). Our current results have shown that soil microbial biomass including fungal biomass increased with better water infiltration; maize yield was strongly correlated with these predictors in addition to precipitation and inorganic N.

CONCLUSIONS

Our results suggest that increasing soil microbial biomass and higher cumulative rainfall at the tasseling stage increase the soil yield potential in long-term no-till systems. Soil surface crop residues, mainly from C_3 plants, and minimal soil disturbance for long periods might have developed a high amount of organic matter on the soil surface, which were related with higher water infiltration rate, and higher soil microbial activities (higher respiration). These findings suggested three major findings. 1) Promoting soil microbial biomass maximizes the yield potential greater than the fertilizer-N, 2) Optimal precipitation is required at early reproductive stage to maximize the production, 3) Higher clay content and higher organic matter improve fungal population in the soil resulting in better water infiltration capacity of the soil.

Deep analysis of the predictor variables with the selection of the most important variables might be necessary to provide a robust model for farmer's understanding and use. All the dryland experimental sites were included in the current analysis; however, sub-setting N responsive and non-responsive sites and considering only the important soil health indicators in the model might make it more efficient. For example, maize yield in N responsive sites were poorly correlated with soil nitrate and ammonium N whereas the correlation was stronger in non-responsive sites (Supplementary figures 2 and 3).

LITERATURE CITED

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33, 102-109.

Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. *Advances in Agronomy*, 154, 201-255.

Bodhinayake, W., Si, B. C., & Noborio, K. (2004). Determination of hydraulic properties in sloping landscapes from tension and double-ring infiltrometers. *Vadose Zone Journal*, 3(3), 964-970.

Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.

Brereton, R. G., & Lloyd, G. R. (2010). Support vector machines for classification and regression. *Analyst*, *135*(2), 230-267.

Buyer, J. S., & Sasser, M. (2012). High throughput phospholipid fatty acid analysis of soils. *Applied Soil Ecology*, 61, 127-130.

Clark, J. D., Veum, K. S., Fernández, F. G., Camberato, J. J., Carter, P. R., Ferguson, R. B., ... & Shanahan, J. F. (2019). United States Midwest soil and weather conditions influence anaerobic potentially mineralizable nitrogen. *Soil Science Society of America Journal*, 83(4), 1137-1147.

Clay, D. E., Chang, J., Clay, S. A., Stone, J., Gelderman, R. H., Carlson, G. C., Reitsma, K., Jones, M., Janssen, L., & Schumacher, T. (2012). Corn yields and no-tillage affects carbon sequestration and carbon footprints. *Agronomy Journal*, 104(3), 763-770.

Clay, D. E., Clay, S. A., Reitsma, K. D., Dunn, B. H., Smart, A. J., Carlson, G. G., Horvath, D., & Stone, J. J. (2014). Does the conversion of grasslands to row crop production in semi-arid areas threaten global food supplies? *Global Food Security*, 3(1), 22-30.

Clay, D. E., Reicks, G., Carlson, C. G., Moriles-Miller, J., Stone, J. J., & Clay, S. A. (2015). Tillage and corn residue harvesting impact surface and subsurface carbon sequestration. *Journal of Environmental Quality*, 44(3), 803-809.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20, 273-297.

Dhakal, C., & Lange, K. (2021). Crop yield response functions in nutrient application: A review. *Agronomy Journal*, 113(6), 5222-5234.

Dhaliwal, J. K., Panday, D., Saha, D., Lee, J., Jagadamma, S., Schaeffer, S., & Mengistu, A. (2022). Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Computers and Electronics in Agriculture*, 199, 107107.

Drummond, S. T., Sudduth, K. A., Joshi, A., Birrell, S. J., & Kitchen, N. R. (2003). Statistical and neural methods for site–specific yield prediction. *Transactions of the ASAE*, 46(1), 5.

Fiedler, D. J., Clay, D. E., Joshi, D. R., Engel, A., Marzano, S. Y., Jakubowski, D., ... & Clay, S. A. (2021). CO2 and N2O emissions and microbial community structure from fields that include salt-affected soils. *Journal of Environmental Quality,* 50(3), 567-579.

Franzluebbers, A. J., Haney, R. L., Honeycutt, C. W., Schomberg, H. H., & Hons, F. M. (2000). Flush of carbon dioxide following rewetting of dried soil relates to active organic pools. *Soil Science Society of America Journal*, 64(2), 613-623.

Franzluebbers, A. J. (2018). Soil-test biological activity with the flush of CO2: III. Corn yield responses to applied nitrogen. *Soil Science Society of America Journal*, 82(3), 708-721.

Geisseler, D., & Wilson, R. (2020). Nitrogen in potato rotations with cover crops: Field trial and simulations using DSSAT. *Agronomy Journal*, 112(3), 2275-2287.

Haney, R., Brinton, W., & Evans, E. (2008). Soil CO2 respiration: Comparison of chemical titration, CO2 IRGA analysis and the Solvita gel system. *Renewable Agriculture and Food Systems*, 23(2), 171-176.

Hermans, S. M., Buckley, H. L., Case, B. S., Curran-Cournane, F., Taylor, M., & Lear, G. (2020). Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome*, 8, 1-13.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. *Springer*. 112

Joshi, D., Sieverding, H., Xu, H., Kwon, H., Wang, M., Clay, S. A., Johnson, J., Westhoff, S., & Clay, D. (2023). A global meta-analysis of cover crop response on soil carbon storage within a corn production system. *Agronomy Journal*, In press.

Joshi, D. R., Clay, D. E., Clay, S. A., Moriles-Miller, J., Daigh, A. L., Reicks, G., & Westhoff, S. (2022). Quantification and Machine learning based N2O-N and CO2-C emissions predictions from a decomposing rye cover crop. *Agronomy Journal*, *Machine learning in agriculture*.

Kim, K. I., Clay, D. E., Carlson, C., Clay, S., & Trooien, T. (2008). Do synergistic relationships between nitrogen and water influence the ability of corn to use nitrogen derived from fertilizer and soil? *Agronomy Journal*, *100*(3), 551-556.

Koparan, M. H., Rekabdarkolaee, H. M., Sood, K., Westhoff, S. M., Reese, C. L., & Malo, D. D. (2022). Estimating soil organic carbon levels in cultivated soils from satellite image using parametric and data-driven methods. *International Journal of Remote Sensing*, *43*(9), 3429-3449.

Lischeid, G., Webber, H., Sommer, M., Nendel, C., & Ewert, F. (2022). Machine learning in crop yield modelling: A powerful tool, but no surrogate for science. *Agricultural and Forest Meteorology*, 312, 108698.

Miner, G. L., Delgado, J. A., Ippolito, J. A., & Stewart, C. E. (2020). Soil health management practices and crop productivity. *Agricultural & Environmental Letters*, 5(1), e20023.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). Introduction to linear regression analysis. *John Wiley & Sons*.

Ngwira, A., Aune, J. B., & Thierfelder, C. (2014). DSSAT modelling of conservation agriculture maize response to climate change in Malawi. *Soil and Tillage Research*, 143, 85-94.

Puntel, L. A., Pagani, A., & Archontoulis, S. V. (2019). Development of a nitrogen recommendation tool for corn considering static and dynamic variables. *European Journal of Agronomy*, 105, 189-199.

R Core Team. (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing. In https://www.R-project.org/

Ransom, C. J., Kitchen, N. R., Camberato, J. J., Carter, P. R., Ferguson, R. B., Fernández, F. G., Franzen, D. W., Laboski, C. A., Myers, D. B., & Nafziger, E. D. (2019). Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. *Computers and Electronics in Agriculture*, 164, 104872.

Ransom, C. J., Kitchen, N. R., Sawyer, J. E., Camberato, J. J., Carter, P. R., Ferguson, R. B., ... & Shanahan, J. F. (2021). Improving publicly available corn nitrogen rate recommendation tools with soil and weather measurements. *Agronomy Journal*, 113(2), 2068-2090.

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. Encyclopedia of database systems. *Springer*, 5, 532-538. https://doi.org/10.1007/978-0-387-39940-9_565

Reitsma, K., Clay, D. E., Carlson, G., Bly, A., & Reicks, G. (2016). Estimating Yield Goals and Nitrogen, Phosphorus, Potassium, Iron, and Zinc Recommendations. In D. E. Clay, G. Carlson, S. A. Clay, & E. Byamukama (Eds.), *iGrow Corn*: Best Management Practices. South Dakota State University Extension. https://extension.sdstate.edu/sites/default/files/2019-09/S-0003-23-Corn.pdf

Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., & Janssen, S. (2013). The agricultural model intercomparison and improvement project (AgMIP): protocols and pilot studies. *Agricultural and Forest Meteorology*, 170, 166-182.

Shahhosseini, M., Hu, G., Huber, I., & Archontoulis, S. V. (2021). Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific reports*, 11(1), 1-15.

Shapiro, C. (2008). Setting Realistic Yield Goals. *Cropwatch*. https://cropwatch.unl.edu/setting-realistic-yield-goals

Turmel, M.-S., Speratti, A., Baudron, F., Verhulst, N., & Govaerts, B. (2015). Crop residue management and soil health: A systems analysis. *Agricultural Systems*, 134, 6-16.

Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177, 105709.

Veum, K. S., Lorenz, T., & Kremer, R. J. (2019). Phospholipid fatty acid profiles of soils under variable handling and storage conditions. *Agronomy Journal*, 111(3), 1090-1096.

Yost, M. A., Veum, K. S., Kitchen, N. R., Sawyer, J. E., Camberato, J. J., Carter, P., Ferguson, R., Fernández, F., Franzen, D., & Laboski, C. (2018). Evaluation of the Haney Soil Health Tool for corn nitrogen recommendations across eight Midwest states. *Journal of Soil and Water Conservation*, 73(5), 587-592.

SUMMARY

In conclusion, a no-tillage maize N recommendation model considering soil health measurements can be valuable tool for farmers to improve crop yield and promote sustainable agricultural practices. The historic N recommendation model for SD does not consider fertilizer to grain price ratio as well as soil health measurements. Our results from this three-years study at 16 sites suggested that considering the price ratios and soil health measurements can greatly improve the N recommendation and the maize yield potential. Integrating this information into the recommendation should be considered as a climate smart practice.

In long-term no-tillage systems, our first chapter from this study suggested that either reducing the constant (amount of N per yield goal) from 21.4 kg to 18 N Mg^{-1} grain on the yield goal based historic SD model can best predict the optimum N required by maize at lower price ratio. Considering of N and maize grain price ratio is suggested in the N recommendation model for SD. In addition, the Nebraska model with organic matter content as one of the input values had the lowest biasness suggesting that price ratios in addition to the soil health information can improve the N recommendation models.

Consideration of soil health measurements improved the predictions of maize yield potential using machine learning approach. Maize yield was correlated with different soil health measurements including soil NH_4-N, soil microbial biomass, organic matter, soil respiration, and precipitation. Similarly, water infiltration capacity was improved with greater fungal population, clay content, and organic matter content. Support vector machine approach, which was implemented to predict the important

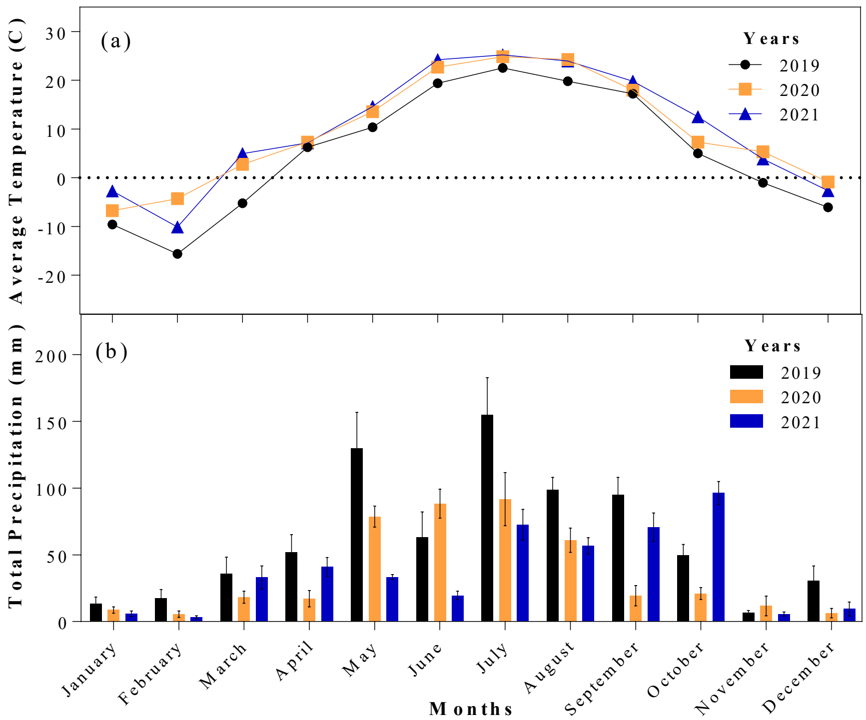
predictor variables for maize yield, suggested that total N, total C, growing degree days, soil microbial biomass, and bacterial biomass were the most important predictor variables of yield.

By considering these and other relevant factors, the model can provide farmers with more precise and targeted N recommendations, reducing the risk of over-application and associated environmental impacts. In addition, by promoting no-tillage practices, the model can help improve soil health and reduce erosion, leading to more sustainable agriculture practices and long-term benefits for both farmers and the environment.

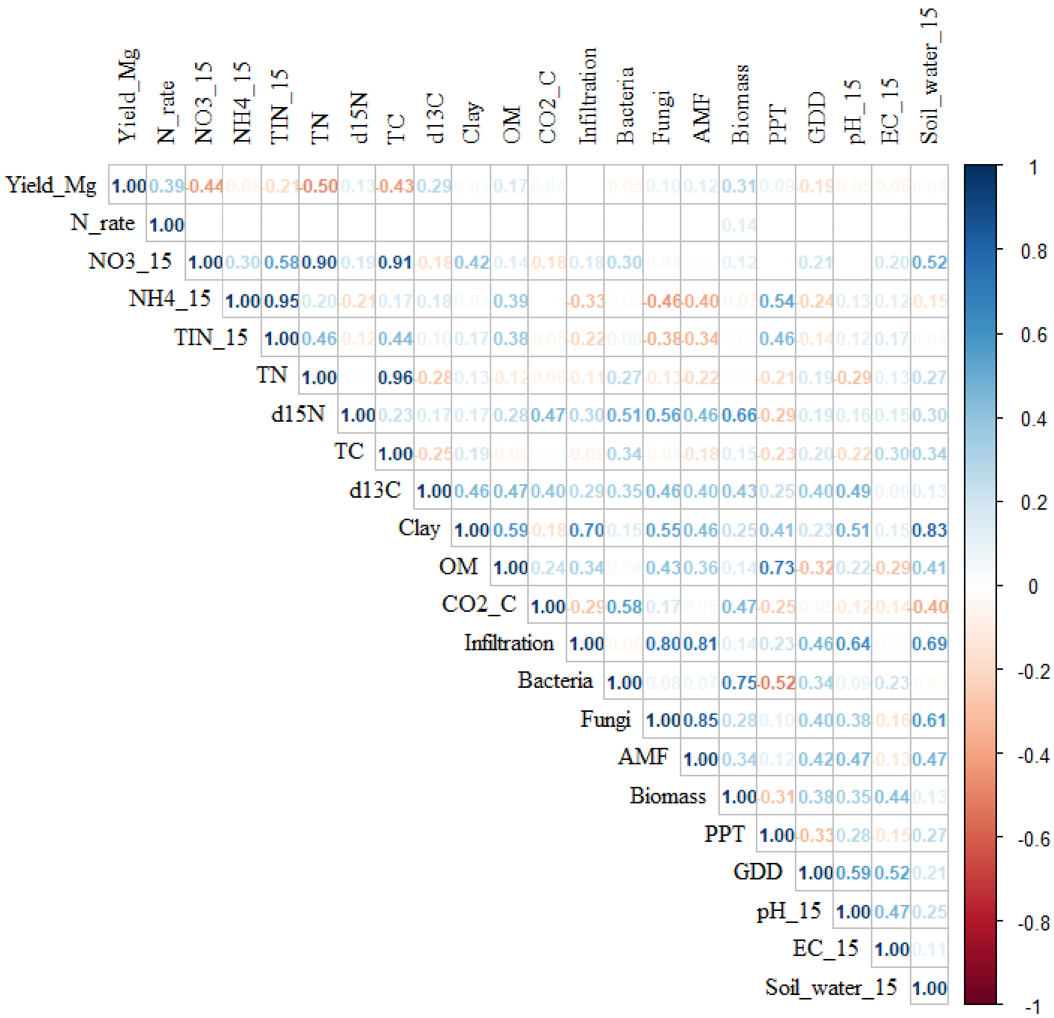
Overall, a no-tillage maize N recommendation model that considers soil health can be a valuable tool for optimizing crop production while also promoting sustainable agriculture practices.

Future research needs to consider the price factor in the N recommendation models in addition to the soil organic matter, climatic variations, and soil health measurements. Measurement of changes in N mineralization, and soil microbial community structure in-season might provide information regarding the N use efficiency and crop yield potential. Use of high-throughput advance technologies and machine learning can be an innovative way to obtain useful information.

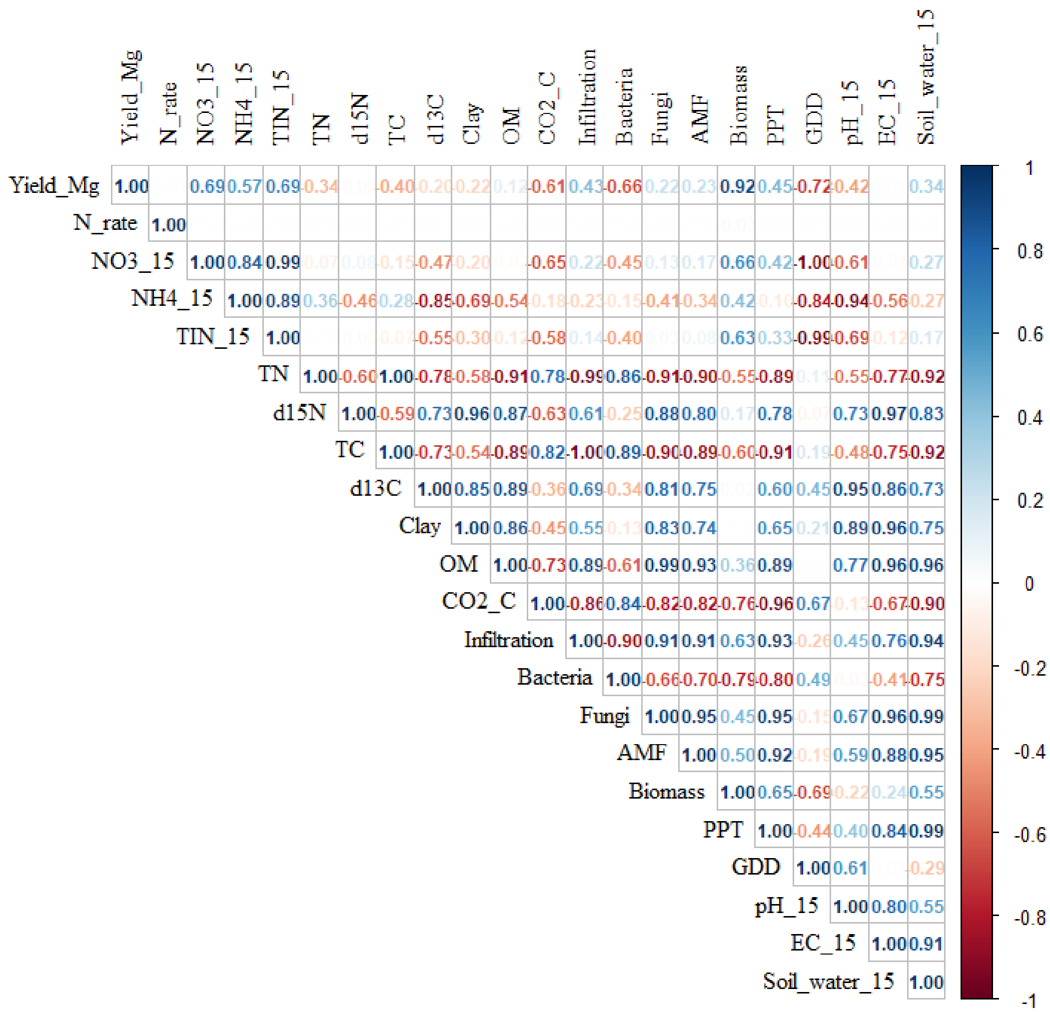
SUPPLEMENTARY FIGURES AND TABLES



Supplementary Figure 1: (a) Average air temperature and (b) cumulative precipitation throughout 2019, 2020, and 2021 averaged across all the experimental sites.



Supplementary Figure 2: Correlation across response and predictor variables in N responsive sites. Nitrate and ammonium N are the amounts in the surface 15 cm, TN and TC are the concentration of organic N and C in the soil, d13C is ^{13}C isotope in ‰, d15N is ^{15}N isotope in ‰, clay is g clay kg^{-1} soil, OM is organic matter in g kg^{-1} soil, CO2-C is Solvita soil respiration in mg C kg^{-1} soil, infiltration is the rate of water infiltration in cm h^{-1}, bacteria, fungi, AMF and biomass are in µg C g^{-1} soil as measured using the PLFA technique, PPT is cumulative precipitation at tasseling stage in cm, GDD is growing degree days, pH is soil pH at 1:1, EC is soil electrical conductivity at 1:1 solution in µS cm^{-1}, and soil water is soil water depth at surface 15 cm measured in cm.



Supplementary Figure 3: Correlation across response and predictor variables in N non-responsive sites. Nitrate and ammonium N are the amounts in the surface 15 cm, TN and TC are the concentration of organic N and C in the soil, d13C is ^{13}C isotope in ‰, d15N is ^{15}N isotope in ‰, clay is g clay kg^{-1} soil, OM is organic matter in g kg^{-1} soil, CO2-C is Solvita soil respiration in mg C kg^{-1} soil, infiltration is the rate of water infiltration in cm h^{-1}, bacteria, fungi, AMF and biomass are in µg C g^{-1} soil as measured using the PLFA technique, PPT is cumulative precipitation at tasseling stage in cm, GDD is growing degree days, pH is soil pH at 1:1, EC is soil electrical conductivity at 1:1 solution in µS cm^{-1}, and soil water is soil water depth at surface 15 cm measured in cm.

Supplementary Table 1. Soil series and soil test results of different experiment sites before planting, 2019-2021. The results are presented for organic matter (OM), soil pH (1:1), soil electrical conductivity ($EC_{1:1}$), soil nitrate-N (NO_3-N), Olsen-phosphorus, Potassium (K), and sum of cations at 0-15 and 15-60 cm depths.

SN	Field name	Soil series	Depth	OM	pH 1:1	EC 1:1	NO_3-N	Olsen-P	K	Sum of cations
				LOI			KCl	Ammonium acetate		
			cm	g kg^{-1}		dS m^{-1}	kg ha^{-1}	mg kg^{-1}	mg kg^{-1}	meq $100g^{-1}$
1	BJO19	Fine, smectitic, mesic Typic Haplusterts	0-15	46	8.0	0.21	6.44	4.38	380.00	42.83
			15-60	--	--	--	--	--	--	--
2	DFO19	Fine-silty, mixed, mesic Typic Argiustolls	0-15	40	6.5	0.09	13.16	11.74	537	20.9
			15-60	--	--	--	--	--	--	--
3	DHO19	Fine-loamy, mixed, superactive, frigid Typic Argiustolls	0-15	48	6.4	0.16	20.72	4.38	521.25	20.13
			15-60	--	--	--	--	--	--	--
4	DLD19	Coarse-silty over clayey, mixed, mesic Fluventic Haplustolls	0-15	33	7.0	0.32	4.48	20.20	609.00	19.00
			15-60	--	--	--	7.85	--	--	--
5	SCA19	Fine-silty, mixed, superactive, frigid Calcic Hapludolls	0-15	50	6.7	0.13	13.72	8.4	189.00	26.6
			15-60	--	--	--	--	--	--	--
6	BJC20	Fine, smectitic, mesic Typic Haplusterts	0-15	48	8.0	0.60	4.50	11.00	572.00	35.80
			15-60	--	--	--	--	--	--	--
7	BJO20	Fine, smectitic, mesic Typic Haplusterts	0-15	48	8.0	0.60	4.50	11.00	572.00	35.80
			15-60	--	--	--	--	--	--	--
8	DFO20	Fine-silty, mixed, superactive, mesic Typic Argiustolls	0-15	27	5.8	0.24	8.30	18.90 (Bray)	393.73	23.51
			15-60	17	7.3	0.46	28.70	5.50	164.98	27.84
9	DLD20	Coarse-silty over clayey, mixed, mesic Fluventic Haplustolls	0-15	34	6.9	0.30	11.20	19.50	500.00	17.10
			15-60	27	--	--	13.40	--	--	--
10	SCA20	Fine-silty, mixed, superactive, frigid Calcic Hapludolls	0-15	45	5.7	0.32	9.50	55.40 (Bray)	221.25	30.16
			15-60	23	7.2	0.59	28.70	6.00	161.83	32.05
11	BJO21	Fine, smectite, mesic Typic Haplusterts	0-15	38	8.45	0.39	7.84	112	368.43	36.5
			15-60	27	8.6	0.90	9.63	5.00	265.89	40
12	BSP21	Fine-loamy, mixed, superactive, mesic Typic Argiustolls	0-15	23	7.6	0.60	7.62	64.0	553.79	18.8
			15-60	19	7.4	2.30	5.6	59.0	342.15	21.6
13	DFC21	Fine-silty, mixed, superactive, mesic Typic, Argiustolls	0-15	27	7.2	0.38	12.32	10.0	1236.4	19.4
			15-60	17	7.9	0.49	7.62	6.00	1247.3	31.7
14	DFO21	Fine-silty, mixed, mesic Typic Argiustolls	0-15	29	6.8	0.25	12.99	8.00	650.22	14.8
			15-60	20	7.2	0.36	10.08	2.00	87.73	28.2
15	DLD21	Fine, smectite, mesic Vertic Argiustolls	0-15	32	7.9	0.24	9.52	4.00	319	28.4
			15-60	--	--	--	--	--	--	--
16	SCA21	Fine-silty, mixed, superactive, frigid Calcic Hapludolls	0-15	38	7.75	0.30	4.48	3.00	380.92	22.2
			15-60	18	7.15	0.41	4.48	2.00	553.72	30.3