

SUB-FIELD CROP YIELD AND YIELD STABILITY ANALYSIS WITH MULTI-SOURCE
DATASETS AND ADVANCED CROP SIMULATION MODEL

By

Guanyuan Shuai

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Geological Sciences – Doctor of Philosophy

2022

ABSTRACT

SUB-FIELD CROP YIELD AND YIELD STABILITY ANALYSIS WITH MULTI-SOURCE DATASETS AND ADVANCED CROP SIMULATION MODEL

By

Guanyuan Shuai

The combination of precision agriculture technology and big data allows farmers to improve the sustainability in crop production by improving profit and minimizing environmental loss. However, knowledge about managing sub-field yield variations over the years (yield stability) is less understood, so the actual adoption rate of precision technology is low, especially in small farms. The overarching goal of this dissertation was to evaluate the spatial and temporal variability of maize and soybean yields related to landscape characteristics and management in the U.S. Midwest (Chapter 1).

Chapter 2 aims to develop a yield stability map using satellite images for ten maize/soybean main production states in the Corn Belt. This study uses Landsat images over eight years and common land unit polygons as the primary dataset to classify each field into stable high, stable low, and unstable areas. I further quantified nitrogen (N) balance and loss and associated monetary and environmental loss for each yield stability class. Across a wide range of crop growth environments, this large-scale analysis has shown reliable and consistent subfield yield patterns that could improve fertilizer application.

Chapter 3 investigates the potential of these low-yielding areas in supporting biofuel crop (switchgrass) production to replace current crops and subsequent improvement in soil and climate change. I also proposed planting a cover crop, rye, to evaluate the effect on the soil nitrate level. The switchgrass and cover crop production were simulated by the Systems Approach to Land Use Sustainability (SALUS) model, which also modeled soil organic matter dynamics under the

proposed and current (continuous maize or maize/soybean rotation) scenarios. Benefits such as reduced soil nitrogen, improved soil structure, and higher heterogeneity of landscape composition demonstrate the potential of yield stability maps to support sustainable agricultural production.

In Chapter 4, I combined the yield stability maps with remote sensing images to improve the maize yield prediction at subfield scale. In this study, I first made a further division in the unstable area by including the topographic features and modeled cumulative crop water stress. The modeled water stress was combined with Landsat Analysis Ready Dataset (ARD) to predict maize yield using a random forest algorithm. Results showed that incorporating the modeled water stress with vegetation index derived from Landsat images produced the highest accuracy in estimated maize yield compared to satellite images alone. This work emphasized the yield stability map's ability to capture subfield spatial variations of soil and plant available water stress information.

Chapter 5 compares the accuracy in yield level estimation between yield stability maps and high-resolution Planet images (3-m, bi-weekly) and analyzes the temporal changes of spatial variation of crop conditions during the growing season. The yield stability maps could predict similar maize yield levels in stable zones compared to satellite images. The temporal analysis in the spatial variation indicated greater changes in the field at the beginning or end of the season, while spatial patterns of crop growth dynamics remain stable in the mid-season. Early- or late-season changes did not always impact the final yield. This analysis provided an in-depth analysis of the in-season image so farmers could better monitor crops in real-time.

Conclusions from these research projects and recommendations on interpretation yield and satellite images for sustainable production are presented in Chapter 6.

ACKNOWLEDGEMENTS

It has been a long journey for me to complete the PhD program especially under Covid over last two years. I would like to thank every individual who helped me obtain my Ph.D. degree at Michigan State University.

I want to express my best and most sincere gratitude to my Ph.D. advisor and chair of my dissertation committee, Dr. Bruno Basso. I am grateful that Bruno offered me the opportunity to pursue higher education six years ago. I am very fortunate that I have been involved in world-class projects since I was here. Under the supervision of Bruno, I gained so much knowledge from many disciplines which was totally unfamiliar to me before I came to MSU. I also learned from him how top scientists come up with real scientific questions and propose solutions to real-world applications. I appreciate Bruno's financial support for many years as a graduate student. Bruno also gave me a lot of care when I was depressed midway through this period, and his enthusiastic and positive personality helped me get through this difficulty. I want to thank Bruno's consistent patience with me whenever I have any research or life problems. I cannot successfully get my degree without Bruno's comprehensive guidance.

I would also want to thank my other committee members. Dr. David Hyndman is always very nice and offered help on my research planning and cared about my personal health. Dr. G. Philip Robertson has provided massive support on my projects. He proposed many innovative ideas to strengthen my work and helped the scientific writing. I would like to thank him for giving me a second chance when I made some stupid mistake during my comprehensive exam. His critical thinking really impacts me. Dr. Jeffrey Andresen helped me understand the macro- and micro-

climate knowledge and guided me in linking this knowledge with my project. They broadened my recognition of the agricultural system from their professions.

I would like to thank the members in Bruno's lab, including Brian Baer, Lydia Rill, Ruben Ulbrich, and Richard Price, who are lovely lab mates from life to research. I enjoyed my time with them in the lab.

I would like to thank my parents, like Bruno, who always had my back. You are the harbor. Whenever I have happiness or difficulties, I can share them with you. Your unconditional love and support make me believe that I can finish anything if I want.

The last and most important person I want to thank is my girlfriend, Jiaxin Zhang. Abroad life was never easy for me alone, and you made it easier and happier. Although our personalities differ a lot from each other, I still believe that we will have a bright future.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
KEY TO ABBREVIATIONS.....	xiii
CHAPTER 1: INTRODUCTION TO THE DISSERTATION	1
1.1 Rationale and background.....	1
1.2 Objectives and Structure of the Dissertation.....	6
1.3 Yield stability analysis	8
1.4 System Approach for Land Use Sustainability (SALUS) process-based model	9
CHAPTER 2: YIELD STABILITY ANALYSIS REVEALS SOURCES OF LARGE- SCALE NITROGEN LOSS FROM THE US MIDWEST.....	11
2.1 Abstract	11
2.2 Introduction	11
2.3 Dataset and Method.....	14
2.3.1 NDVI Stability Classes	17
2.3.2 Yield and N Uptake Estimates	18
2.3.3 Nitrogen Balance Calculations	19
2.3.4 Monetary value and Environmental losses	20
2.3.5 Yield Stability Validation	21
2.4 Results	21
2.5 Discussion	28
2.6 Conclusion.....	30
2.7 Acknowledgements	31
CHAPTER 3: POTENTIAL OF USING YIELD STABILITY MAP TO SUPPORT SUSTAINABILITY OF AGRICULTURAL LANDSCAPE.....	32
3.1 Abstract	32
3.2 Introduction	33
3.3 Study area and Dataset	37
3.4 Method	39
3.4.1 Spatial analysis of yield stability maps.....	39
3.4.2 Cover crop/switchgrass simulations	39
3.4.3 Life cycle assessment for maize and switchgrass	42
3.4.4 Ecosystem biocontrol index and avian richness under proposed management practices	43
3.5 Results	43
3.5.1 Proportions of stability class along the distance from field boundary.....	43
3.5.2 SALUS simulations results for switchgrass, cover crop, and maize/soybeans.....	47
3.5.3 Greenhouse warming intensity under switchgrass and maize production systems	58
3.5.4 Ecological benefits from changing maize to switchgrass	60

3.6	Discussion	61
3.7	Conclusion.....	64
CHAPTER 4: SUBFIELD MAIZE YIELD PREDICTION IMPROVES WHEN IN-SEASON CROP WATER DEFICIT IS INCLUDED IN REMOTE SENSING IMAGERY-BASED MODELS		
4.1	Abstract	65
4.2	Introduction	66
4.3	Study area and dataset	70
4.4	Methods.....	73
4.4.1	Landsat 7 and 8 images mosaic	73
4.4.2	Cloud removal for mosaic VI	74
4.4.3	Yield stability zones and Cumulative Drought Index (CDI) calculations	74
4.3.3.1	Stability zone calculation.....	74
4.3.3.2	Subdivision of unstable zones using topography features	75
4.3.3.3	Cumulative drought index calculations	75
4.3.3.4	Yield predictions and validation	77
4.5	Results	78
4.5.1	Accuracy of the new sub-field scale yield prediction model.....	78
4.5.2	SALUS model testing	84
4.5.3	Long-term pattern of CDI	89
4.6	Discussion	92
4.7	Conclusion.....	96
4.8	Acknowledgment	96
CHAPTER 5: ASSESSING IN-SEASON IMPACTS OF SPATIAL AND TEMPORAL CHANGES OF MAIZE GROWTH ON YIELD AT FIELD SCALE: A COMPARISON BETWEEN IMAGERY AND YIELD STABILITY MAPS.....		
5.1	Abstract	97
5.2	Introduction	98
5.3	Study area and data	101
5.4	Methods.....	104
5.3.1	Yield level prediction using yield stability maps and remote sensing.....	104
5.3.2	Optimal image selection for yield estimation using clustering analysis.....	105
5.3.3	Short duration crop growth dynamics impacts on yield	106
5.5	Results	107
5.4.1	Accuracy of yield level prediction using yield stability maps and in-season remote sensing	107
5.4.2	Variance in optimal imagery dates and impacts on yield prediction	109
5.4.3	Rate of change and related environmental factors	110
5.6	Discussion	114
5.7	Conclusion.....	115
CHAPTER 6: CONCLUSIONS AND RECOMMENDATION		
APPENDICES		
APPENDIX A: CHAPTER 2 Supplementary Information		

APPENDIX B: CHAPTER 3 Supplementary Information	128
REFERENCES	135

LIST OF TABLES

Table 1 State-level yield stability trends for 2010-2017.....	25
Table 2 N removed by harvest, N fertilizer surplus, and apparent N use efficiency (NUE) within yield stability classes.	26
Table 3 Surplus fertilizer N loss from stable and unstable low yield areas and its monetary value, embedded energy, and associated CO ₂ -equivalent emissions.....	27
Table 4 LS area at different patch size.....	47
Table 5 Simulated state-level average maize yield (Mg ha ⁻¹ y ⁻¹) in the LS areas under four rotations: continuous maize (MZ), maize soybean rotation (MZ-SB), continuous maize with rye (MZ-RY), and maize soybean rotation with rye (MZ-SB-RY) and two tillage methods: conventional (CT) and no-till (NT). Simulations were from 2009 to 2019.....	50
Table 6 Simulated state-level average soybean yield (Mg ha ⁻¹ y ⁻¹) under MZ-SB and MZ-SB-RY in the LS areas. Simulated state-level average rye biomass (Mg ha ⁻¹ y ⁻¹) under MZ-RY and MZ-SB-RY.....	51
Table 7 Simulated state-level average soil organic carbon change (Mg ha ⁻¹ y ⁻¹) under conventional tillage.....	54
Table 8 Simulated state-level average soil organic carbon change (Mg ha ⁻¹ y ⁻¹) under no-till. ...	55
Table 9 Simulated soil nitrate level (kg ha ⁻¹ y ⁻¹) under conventional tillage.....	56
Table 10 Simulated soil nitrate level (kg ha ⁻¹ y ⁻¹) under no-till.....	57
Table 11 State-level average GWI (Mg CO ₂ eq ha ⁻¹ y ⁻¹) for MZ, MZ-SB and SW in the LS area.	59
Table 12 State-level NLCD statistical information and grass compositions and changes under maize and switchgrass scenarios.....	61
Table 13 Three most important features for yield prediction with all GCVIs and CDI (GCVI DOY in parentheses). Unit of IncMSE is percentage (%).	82
Table 14 CDI during entire grain filling period for Field27	92
Table 15 Average state-level N fertilizer rates and Coefficient of Variation (CV, %) as reported by ARMS (NASS 2016) and University-recommended rates based on statewide Maximum Return to Nitrogen (MRTN) databases (Sawyer et al., 2006). All values are kg N ha ⁻¹ y ⁻¹	121

Table 16 Average N rate, N fertilizer surplus and percent contribution from each yield stability classes.....	122
Table 17 Cultivar information for RY16 species.....	128
Table 18 Cultivar information for SW_UL species.....	129
Table 19 Values used in the estimation of net emissions associated with nitrogen fertilizer additions. Values of RS, N_AG, N_BG, R for maize, soybean, and switchgrass were from Tables 11.a in the IPCC, 2019.	130
Table 20 State-level stability area, field edge area, LS area, and LS edge area. All values are in ha expect the last column as percentage.	131
Table 21 Proportion of stability class within each buffer area. Stability area value is in ha and HS, LS, UN columns are the percentages (%).	132

LIST OF FIGURES

Figure 1 Crop yield stability maps for (A) ten U.S. Midwest states and subregions of (B) 10,000 km ² , (C) 196 km ² , and (D) 118 ha. Colors represent yield stability areas for 0.09 ha portions of fields planted to maize or maize-soybean for at least three years during 2010-2017 (~30 million ha in total).	24
Figure 2 Fractions of stability class within each buffer area. FC indicates field center. Buffer area was determined as the area between two neighboring buffers defined in the method.	45
Figure 3 Fraction of LS area with each buffer area to all LS in the field.	46
Figure 4 Mean yield of switchgrass over 11 years in the LS areas.	52
Figure 5 SOC change over 11 years of switchgrass in the LS areas.	53
Figure 6 Annual soil nitrate levels after crop (MZ or MZ/SB) or rye harvest in the MZ-RY (left) and MZ-SB-RY (right) scenarios.	58
Figure 7 Reclassified NLCD dataset (a), proportions of grassland under maize (b) and switchgrass scenarios (c), and associated grassland changes (d).	60
Figure 8 Accuracy assessment for maize yield prediction models in Iowa (IA), Illinois (IL), Indiana (IN), and two tiles (022006 and 023007) in Michigan (MI) and in three testing years: good, average, and poor years. “VI” refers to multi-date gap-filled Landsat ARD images as model input, “VI & ST” refers to combination of multi-date gap-filled Landsat ARD GCVI and ST images, while “VI & CDI” refers to input including multi-date images and CDI together. Points on the lines indicate the prediction dates when new Landsat image were acquired since the maize grain filling period.	81
Figure 9 Scatterplot of predicted and actual yield from four situations: a) composited GCVI using pixel samples, b) composited GCVI and CDI using pixel samples, c) composited GCVI using field samples, and d) composited GCVI and CDI using field samples.	83
Figure 10 Subfield-level CDI (A), yield maps (B), yield predictions using VI alone (C) and using VI and CDI (D) in 2012. For actual and predicted yield maps in each field, we used the same color ramp.	84
Figure 11 Scatter plots of SALUS simulated yield vs yield monitor data at zone level for the selected 16 fields. Yield stability classes include high stable (HS), medium stable (MS), low stable (LS), unstable-depression (UNDEP), unstable-hilltop (UNHIL), and unstable-other (UNOTH). The 1:1 line is added in each subplot.	87
Figure 12 Scatter plots of yield monitor data vs CDI during the whole grain filling period at zone level for the same fields in Figure 4. Regression lines (blue lines) and grey area (the 95% confidence level interval) are added to show the response of maize yield to CDI.	88

Figure 13 Box plots showing distribution of whole grain filling period CDI by stability zone in each state.....	89
Figure 14 Cumulative distribution of CDI during whole grain filling period from 1990 to 2019 in four fields.....	91
Figure 15 Selected study sites, 2018-2020.	103
Figure 16 Accuracies on the maize yield level prediction by linear regression (LR), multi-linear regression (MLR), random forest (RF), and yield stability map (YS) for each state in the stable high (HS) and stable low (LS) zones.	108
Figure 17 Accuracy assessment for maize yield level prediction models dates when image is added in acquisition time order. This assessment included all the fields in the study area.	109
Figure 18 Weekly difference between the field level optimal image and that determined by the whole dataset (a) and between the field and subfield level optimal image (b).	110
Figure 19 Percentage of ‘Change’ and ‘Non change’ classes during each successive images for each state.	111
Figure 20 Percentage of HL class in the high yielding area (a) and LH class in the low yielding area (b) by the stability zone for each state.	112
Figure 21 The relationships between environmental factors and the rate of change in June (a) and September(b). Gray areas are the standard errors.	113
Figure 22 Estimated yield using NDVI-stability class approach vs yield monitor data.	123
Figure 23 Estimated yield vs. yield monitor data for corn (orange) and soybeans (blue) for each state.	124
Figure 24 The acreage percentages of stable high yield, stable low yield, and unstable yield areas of Iowa calculated from yield monitor combine harvesters (left-hand bars) or NDVI data (right-hand bars).	125
Figure 25 Yield stability class areas derived from combine-based yield monitor data (left-hand panels) and LANDSAT NDVI data (right-hand panels) for two representative Iowa fields (top and bottom rows). The LANDSAT NDVI spatial resolution is 30 m; yield monitor resolution is 2 m.	126
Figure 26 (A) County-level N loss in 2012-13 growing season and (B) mean 2013 nitrate concentration at the 100 Midwest Stream Quality Assessment sites from Van Metre et al. (2016). The red circles denote areas with concurrent higher N loss and nitrate concentrations. The blue rectangles are areas with lower N loss and nitrate concentration.....	127
Figure 27 Validation of simulated rye biomass vs. biomass from literature. Red circle includes some outliers.....	134

KEY TO ABBREVIATIONS

AI	Artificial Intelligence
ARD	Analysis Ready Dataset
ARMS	Agricultural Resource Management Survey
C	Carbon
CDI	Cumulative drought index
CDL	Cropland Data Layer
CLU	Common Land Unit
CSM	Crop Simulation Model
DA	Digital Agriculture
DEM	Digital Elevation Model
DOY	Day of the Year
ERS	Economic Research Service
ET	Evapotranspiration
GCVI	Green Chlorophyll Vegetation Index
GHG	Greenhouse Gas
HS	Stable High Yield
LAI	Leaf Area Index
LS	Stable Low Yield
MAPE	Mean Absolute Percentage Error
MRTN	Maximum Return to Nitrogen
N	Nitrogen

NASS	National Agricultural Statistics Service
NDVI	Normalized Difference Vegetation Index
NIR	Near Infrared
NLCD	National Land Cover Datasets
NUE	Nitrogen Use Efficiency
PA	Precision Agriculture
RF	Random Forest
RS	Remote Sensing
SALUS	Systems Approach for Land Use Sustainability
SOC	Soil Organic Carbon
SOM	Soil Organic Matter
SSM	Site-Specific Management
SSURGO	Soil Survey Geographic Database
ST	Surface Temperature
TPI	Topographic Position Index
UN	Unstable yield
USDA	United States Department of Agriculture
VI	Vegetation Index
VRT	Variable Rate Technology

CHAPTER 1: INTRODUCTION TO THE DISSERTATION

1.1 Rationale and background

Agricultural systems are incredibly complex because they involve multiple interactions, and the basic science behind even simple questions such as how much fertilizer to apply is still uncertain. Nevertheless, farmers need to increase fertilizer use efficiency so that food can be produced while minimizing the environmental problems attributed to over-fertilization. To that end, this project aims to advance the science involved in making agricultural systems more sustainable and efficient concerning economics, energy, and the environment.

Over the last two decades, the application of sensors and geospatial technologies in agriculture has made it possible to assess the spatial-temporal variability of crop yield. These technologies have allowed farmers to visualize their data and become more aware of the inefficiency of their management strategies, particularly the application of nitrogen (N) fertilizer. As variability within fields is the norm rather than the exception, the typical practice of uniform N fertilizer application causes some parts of the field to be under-fertilized and others over-fertilized, leading to N losses to the environment and to reduced profit (in low-yielding areas where a crop does not completely utilize the applied N).

Since the 1990s, Precision agriculture (PA) was introduced to apply variable inputs according to spatial variability of crop growth. Although farms equipped with PA were reported to have a higher profit of \$66 per acre than nonadopters, this difference mainly comes from larger farms (Schimmelpfennig, 2016). The cost for variable rate technology (VRT) machines is better amortized on large farms compared to small farms, like other mechanical technologies. VRT adoption was more common on farms over 1,700 acres than on small farms, with the highest adoption rate of 40% found on farms over 3,800 acres. Basso & Antle (2020) further indicated that

PA did not bring widespread economic and environmental benefits, with one possible reason being the lack of effective policies to incentivize the adoption of these technologies. Another large barrier is the challenge of interpreting field data, such as variability within a field and its relationship to weather and soil, which impedes the effective use of PA technology and development of effective decisions for adaptive decision making.

The challenges and trade-offs farmers face when making decisions could be eased by novel emerging integrated digital technologies. Digital agriculture (DA) inherits principles from PA and integrates modern machines, big-data collection, advanced computer tools, and information and communication technologies (ICTs), to make on- and off-farm decisions (Ozdogan et al., 2017). Technologies such as guidance and tracking systems using Global Positioning System (GPS), remote sensing images, and crop simulation models have been used since the 1990s, while more recent cutting-edge technology, such as artificial intelligence (AI), and the Internet of Things (IoT), bring DA to the new level. The core idea of DA is that big-data analysis and data-driven solutions can bridge the gaps between agronomic recommendations and actual practices.

With accurate geographic information, the GPS guidance system on tractors and combine harvesters is the basis for high-precision operations, i.e., seedling, fertilizer application and yield harvest, thus increasing time-cost efficiency with less over- and under-applications (Schimmelpfennig, 2016). Farmers equipped with GPS guidance can increase the efficiency of their field management operations by avoiding errors in the applications of inputs; eliminating areas where inputs were applied twice (overlap) and areas where the input was not applied. The first use of yield monitors was also about three decades ago. Combined with a guidance system, harvester-mounted yield monitors simultaneously record yield data and GPS coordinates to generate yield maps which capture within-field yield variability (Murphy et al., 1995). Yield maps

over multiple years are the most valuable dataset for farmers to characterize field performance and understand how their management practices affect the output. The VRT is also based on a guidance system to apply fertilizer and other chemicals following the prescribed variable rates in a field (Clark & McGuckin, 1996). All these PA technologies are the fundamental tools to collect extensive data about yield, soil, and fertilizer use.

Remote sensing is another major component of big data, especially high spatial and temporal resolution satellite images (i.e., Sentinel, Planet), and drones are becoming available to farmers (Jung et al., 2021; Tang et al., 2002). Many facilities provide platforms for free and open access to satellite images (e.g., Copernicus Open Access Hub (European Space Agency), Earth Explorer (United States Geological Survey)). Drones now offer centimeter resolution imaging of crop plants and are less dependent on weather conditions as they fly below the cloud canopy. Advanced drones are even able to spray fertilizer and insecticides. The agronomic analysis of remote sensing images ranges from real-time crop biophysical variable assessment, phenotyping, prescriptive analytics about fertilizer and irrigation, to yield forecasting. Satellite spectral information can be used to calculate vegetation indices (VIs) using various band combinations (Bannari et al., 1995). VIs derived from near-infrared and red-edge bands are relevant to vegetation conditions that positively correlate with final crop yields, such as leaf area index (LAI) and biomass (Basso et al., 2004). In addition to spectral information, the three-dimensional point cloud collected by lidar sensors could further improve accuracy in crop biomass and height estimation. Based on a comprehensive review from Basso & Liu (2019), seasonal NDVI images are the most commonly used data for building linear regression models to forecast crop yield. The expensive hyperspectral sensors could further detect the crop nitrogen status based on the relationship between foliar N concentration and leaf reflectance measurements. Thermal images could evaluate

the evapotranspiration in the field, which is used to schedule irrigation if crops are under water-deficit status. These recent advances in remote sensing information allows farmers to monitor their crops at high spatial, temporal, and spectral resolutions.

Converting complex and large-scale data into decisions requires associated expertise. Modeling can identify the optimal management since it is difficult to conduct field experiments over space and time to find the answer (Basso & Ritchie, 2015). There are many well-developed models, such as DSSAT, APSIM, and Systems Approach for Land Use Sustainability (SALUS), which have proven useful in agricultural research worldwide (Rauff & Bello, 2015). Process-based crop models can simulate crop growth and development across various soil, weather, crop variety, and management information. Moreover, the physical, chemical, and physiological limits on crop growth and yield can be obtained from model outputs to inform management strategies. Soil N and carbon dynamics are also modeled to provide detail about soil greenhouse gas emissions and to find C sequestration implementations. Soil water calculations are performed daily to indicate the drought impacts and the potential for nitrogen loss. Crop models can also assess the impact of climate change on crop production at local and global scales (Angulo et al., 2013; Rosenzweig et al., 2014; Tao et al., 2018). For example, global wheat production was predicted to decrease by 6% for each °C of temperature increase (Asseng et al., 2015). Overall, crop models explain the interactions between Genotype, Environment, and Management ($G \times E \times M$) and provide alternative management options for long-term sustainable production (Jones et al., 2003).

As the state-of-the-art method, cloud computing and AI are a part of our daily lives, from SIRI to self-driving vehicles, and are now involved in agricultural activities. The primary task of AI is to analyze the collected data using deep learning algorithms and other machine learning methods by considering multiple factors, such as weather variability, environmental conditions, or

market conditions. Convolutional neural networks (CNN) and transfer learning models have shown great potential in yield prediction, price forecasting, crop health monitoring, and pest management (Fountas et al., 2020). The advanced computation of real-time data on cloud platforms relieves farmers from time-consuming and complex data analysis so they can focus on making comprehensive decisions. Advanced robotics are designed and trained for agricultural tasks, like weed control and crop harvest, to reduce labor costs. While the agricultural labor force has been in shortage for many years due to decreased rural population, AI and agri-robots are becoming a possible option for many large-scale agricultural businesses that cannot find enough employees (Bampasidou & Salassi, 2019).

DA is still developing at a rapid rate, driven by many technological advances, however, maturity of DA technology alone is unlikely to attain general adoption in farms. Weltzien (2016) mentioned that DA offers only modest returns since collected data only lies with the computers and not agricultural products. The complexity of the agriculture system and inherent variability in climate, soil, and topography, and associated impacts on crop growth and yield are still not well understood (Basso & Antle, 2020). There is a lack of advanced computation tools which could unlock the full potential of spatial-temporal datasets and convert them into manageable actions for farmers. These data pipelines require the thoughtful coordination of agronomy and climatology experts, farmers, data scientists, and PA companies. The scalability is also an important issue since practical, sustainable solutions need to be developed for both large-scale industrial and smaller-scale systems (Nery et al., 2018). Meanwhile, government and business support and education of farmers are essential to promote digital agriculture and achieve a sustainable future.

1.2 Objectives and Structure of the Dissertation

With the projected increasing population in the world, we will be challenged to guarantee food security without converting new land to cultivation. Unfortunately, the pressure to increase food production often leads to nutrient loss through run-off and leaching, resulting in negative environmental impacts. Farmers already recognize that any fertilizer not taken up by crops can be lost from their fields, thereby lowering profits. However, it's not an easy task to limit over-fertilization while maintaining or even increasing food production. To maximize profits and minimize risk, N fertilizer would optimally be applied at a rate where the farmer's risk-adjusted expected payoff is maximized. That does not necessarily mean applications should be at the rate that produce the highest yields. For over a decade, it has been hoped that future gains in N-fertilizer efficiency (and associated environmental protection) could be achieved through precision management of food production, however, that promise has yet to be realized. This research aims to achieve that goal.

While PA and "Big-Data" are related, they are not the same thing. PA describes recently developed technologies and principles used to assess and manage spatial (within a field) and temporal (within a season and between years) variability associated with all aspects of agricultural production. Big-Data refers to collecting, analyzing, and synthesizing large datasets that may (or may not) originate from PA equipment. Big-Data techniques offer the capability to collect and analyze data at a magnitude that was impossible in the past because of technological or analytical constraints (Sonka, 2016). In addition, farmers have easy access to remote sensing through aerial imagery, Unmanned Aerial Vehicles (UAVs), soil sensing technologies, and other on-the-go vehicle-mounted sensors that can be linked to hand-held mobile devices or computers and store data "in the cloud". Approximately 70% of US tractors have GPS with auto steering technologies

and 40% of all maize farms used yield monitors (Schimmelpfennig, 2016). Such monitors demonstrate that significant yield differences commonly exist within a field (Basso et al., 2001, 2016; Batchelor et al., 2002), and that these patterns of variability within a field can also differ from year to year (S. M. Albarenque et al., 2016; Basso et al., 2007). Remote sensing imagery confirms large differences in canopy development related to yield variability (Basso et al., 2001, 2016). The magnitude of variability can be used to evaluate whether to implement a spatially variable management plan. Farmers and researchers are inundated with crop yield and imagery data, but such data are of little value until they are translated into actionable information that improves economic and environmental efficiency. Proper assessments of yield variability would integrate several years of data with different crops because a given limiting factor can exert different spatial and temporal influences on yield (Basso et al., 2012).

The overarching goal of this research is to develop and enable solutions to improve the environmental and economic returns of U.S. agricultural systems given the knowledge of spatial and temporal variability of maize and soybean yields related to landscape characteristics and management. We will combine spatial and temporal yield variability with advanced analytical tools to investigate the hypothesis that site-specific management strategies based on these new data can reduce the cost of crop production, limit off-site environmental impacts, and maintain or increase yields.

The objectives of this research are to:

1. Develop yield stability maps from remote sensing images.
2. Analyze stability maps, spatial and temporal variability of N uptake, and nitrate leaching in maize and soybean crops across U.S. Corn Belt.
3. Map low and stable zones across the US Midwest.

4. Evaluate alternative management scenarios to be implemented on stable low zones
5. Use the SALUS crop model to model bioenergy grown and cover crops productivity in stable low zones.
6. Improve the prediction of subfield maize yields by linking remote sensing images with a simulated crop drought index.
7. Couple remote sensing and yield stability maps to understand in-season dynamics of crop growth and its relationship to final grain yield.

To fulfill these objectives, we aim to develop a novel geospatial system approach that links remote sensing images and other geospatial data layers with crop information to improve our understanding of the U.S. Corn Belt crop production systems.

1.3 Yield stability analysis

Among multiple categorical datasets, historical yield maps are often considered as a higher priority in management zone (MZ) delineation. The rationale for using plant as sensors, and yield as a critical layer for MZ delineation evolves from the fact that if variables have shown spatial pattern different than the yield, managing these factors would lead to erroneous applications of input, as the crop has not responded to that variability. Yield stability concept has been around for few years (Blackmore, 2000; Basso, 2007) but the use of remote sensing as layer for yield stability analysis is a key point of this dissertation (Basso et al., 2019). The identification of yield stability is carried out by considering the level of yield obtained within the field and the degree of stability over the years using spatial yield monitor data (Blackmore, 2000). Specifically, the spatial variability of yield is calculated as the relative percentage difference of crop yield from the field's average yield at each point mapped, according to Equation 1:

$$\overline{y_i} = \frac{1}{n} \sum_{k=1}^n \left(\frac{y_{ik} - \overline{y_k}}{\overline{y_k}} \times 100 \right) \dots \dots \dots \text{Equation (1)}$$

where $\overline{y_i}$ is the average percentage difference at pixel i , $\overline{y_k}$ is the average yield at year k , y_{i_k} is the yield at location i at year k , and n is the number of years of input yield maps.

The final yield stability zones are created by overlaying multiple years of the relative percentage difference of yield at pixel level. Fields are divided into high and low yielding areas using 0 as the threshold to classify the average relative percentage difference of yield. The temporal variance of yield, expressed as the degree of stability, is calculated as Equation 2:

$$\overline{\sigma_i^2} = \frac{1}{n} \sum_{k=1}^n (y_{i_k} - \overline{y_{i_n}})^2 \dots \dots \dots \text{Equation (2)}$$

where $\overline{\sigma_i^2}$ is the temporal variance value at location i , y_{i_k} is the yield at location i at year k , $\overline{y_{i_n}}$ is the average yield over the n years. In this dissertation, we used the absolute value of $\overline{\sigma_i^2}$ less than or equal to 15 as threshold to determine whether a pixel is classified as stable or unstable.

Finally, an individual field is classified into three categories: 1) stable high area, 2) stable low area, and 3) unstable area.

1.4 System Approach for Land Use Sustainability (SALUS) process-based model

The System Approach for Land Use Sustainability (SALUS) model simulates crop growth and development in this dissertation. SALUS simulates not only crop yields over multiple years but also soil and water dynamics under various management strategies. Typical management, such as rotations, irrigation and fertilizer applications, tillage methods, etc., can be set up in the model. The main components of SALUS include crop growth models, soil organic matter modules, and soil water balance and temperature modules at daily steps. Crop stage is determined by accumulated thermal time, and crop growth is affected by the intercepted light using solar radiation data and simulated Leaf Area Index (LAI), and nutrient deficiency. Daily biomass accumulation is a function of photosynthetically active radiation, radiation use efficiency, and LAI. The water

and nutrient stress factors are calculated every day, and their impacts on LAI and biomass reductions in the presence of stresses are estimated. The SOM module derives from CENTURY with some modifications, which simulates SOM dynamics for three soil organic carbon pools (active, slow, and passive). More details can be found in Basso & Ritchie (2015).

CHAPTER 2: YIELD STABILITY ANALYSIS REVEALS SOURCES OF LARGE-SCALE NITROGEN LOSS FROM THE US MIDWEST

A version of this chapter appeared in the journal Scientific Reports (doi: 10.1038/s41598-019-42271-1)

2.1 Abstract

Loss of reactive nitrogen (N) from agricultural fields in the U.S. Midwest is a principal cause of the persistent hypoxic zone in the Gulf of Mexico. We used eight years of high-resolution satellite imagery, field boundaries, crop data layers, and yield stability classes to estimate the proportion of N fertilizer removed in harvest (NUE) versus left as surplus N in 8 million maize (*Zea mays*) fields at subfield resolutions of 30×30 m (0.09 ha) across 30 million ha of 10 Midwest states. On average, 26% of subfields in the region could be classified as stable low yield, 28% as unstable (low yield some years, high others), and 46% as stable high yield. NUE varied from 48% in stable low yield areas to 88% in stable high yield areas. We estimate regional average N losses of 1.12 (0.64-1.67) Tg N y⁻¹ from stable and unstable low yield areas, corresponding to USD 485 (267-702) million dollars of fertilizer value, 79 (45-113) TJ of energy, and greenhouse gas emissions of 6.8 (3.4-10.1) MMT CO₂ equivalents. Matching N fertilizer rates to crop yield stability classes could reduce regional reactive N losses substantially with no impact on crop yields, thereby enhancing the sustainability of maize-based cropping systems.

2.2 Introduction

Reactive nitrogen loss to the environment is one of the most widespread and recalcitrant environmental problems in major crop-producing regions of the world today. It is especially problematic in those areas where N fertilizer is used extensively, such as the United States, China, and Europe (Lassaletta et al., 2014; Mueller et al., 2017; Robertson & Vitousek, 2009; Vitousek

et al., 2009), and leads to coastal (Howarth et al., 2011) and surface water eutrophication (Van Metre et al., 2016), groundwater contamination (Nolan et al., 1997), elevated rates of N deposition from gaseous emissions of NH_3 and NO_x (X. Liu et al., 2013), atmospheric greenhouse gas loading (Shcherbak et al., 2014), and stratospheric ozone depletion (Ravishankara et al., 2009). The ~110 Tg of N fertilizer applied annually to crops (FAO, 2017), often in excess of plant requirements (Conant et al., 2013; Lassaletta et al., 2014), is almost twice that entering the biosphere during pre-industrial times (Vitousek et al., 2013). Future reactive N losses from excessive N fertilizer use will be further exacerbated by rising demands for food and other agricultural products as global population and affluence increase (Tilman et al., 2011). Solutions to the imbalance between crop N requirements and the regional amount of N fertilizer applied have been elusive, in part because of the difficulty of linking large-scale effects to small-scale practices (Stuart et al., 2015).

The pressure on farmers to increase crop yields for greater economic return often leads to excessive N fertilizer application, despite its economic and environmental cost (Cui et al., 2018; Pannell, 2017; Robertson & Vitousek, 2009; X. Zhang et al., 2015). In most of the world, fertilizer is applied uniformly at the beginning of a cropping season in anticipation of high yields and efficient use, even though farmers recognize that yields and N use will be neither uniform nor necessarily efficient in any given year, and that fertilizer not taken up by crops will be lost from fields, thereby lowering profits and harming the environment.

Ideally, N application rates should vary across fields to match the well-known variability of crop growth conditions, which are largely a function of soil, weather, and position in the landscape (Basso et al., 2009; Kravchenko et al., 2005). However, applying fertilizer at variable rates across fields (precision agriculture) is challenging because precision agriculture requires a detailed understanding of subfield variability and the relationship of this variability to weather and

crop growth. For more than a decade, gains in N fertilizer use efficiency (and associated gains in environmental protection) were hoped to be achieved through precision N management, but that promise has yet to be realized.

Current use of variable rate technologies is low even in technologically advanced countries like the U.S. due to the complexity of converting geospatial information on soil and plant status into appropriate crop management information, and low economic returns based on current application practices (Babcock & Pautsch, 1998; Y. Liu et al., 2006; Schimmelpfennig, 2016). In 2012 yield mapping was used on about half of U.S. maize and soybean farms, but variable rate application technology, whether for seeds, pesticides, lime, or fertilizers, on only 16-26% in aggregate (Schimmelpfennig, 2016).

Here we provide a novel remote sensing approach to derive fine-scale yield stability classes that respond differentially to N fertilizer based on historical yield performance. Using non-commercial widely available remote sensing imagery, we quantify the spatial and temporal variability of maize and soybean yields based on in-season growth patterns and provide for subfield areas a partial N balance for maize years, calculated as N fertilizer addition less plant harvest N removal. This value, a form of N use efficiency (the fraction of N input harvested as product), defined as N recovery efficiency or nitrogen partial factor productivity (X.-P. Chen et al., 2011; Conant et al., 2013; Cui et al., 2018; Ladha et al., 2005; Lassaletta et al., 2014; W. Zhang et al., 2013; X. Zhang et al., 2015), provides a conservative estimate of the amount of N fertilizer not used by the crop and thereby unnecessarily lost from the system, subsequently providing an estimate of the monetary and environmental savings that could be realized by more precise application of N fertilizers using available geospatial technologies at the subfield scale.

2.3 Dataset and Method

Cropland data layers (CDL). CDLs are annual raster-format land-use maps created by the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS). In 2006, CDLs had a spatial resolution of 56m and land-use categories were based on the Landsat 5 TM, Landsat 7 Enhanced Thematic Mapper (ETM+), the Indian Remote Sensing RESOURCESAT-1 (IRS-P6), and Advanced Wide Field Sensors (AWiFS). Since 2008, the CDLs utilized Landsat TM/ETM+ and AWiFS imagery for production of a 30m product covering the continental US. For our study period CDLs were processed using the Albers Equal-Area Conic Projection with the North American Datum 1983 (NAD83); we re-projected from Albers to the dominant Universal Transverse Mercator (UTM) zone with a spheroid and datum of World Geodetic System 1984 (WGS84). We used CDL data to extract fields in the study area grew either maize or soybeans.

Satellite Data. We acquired Google Earth Engine Landsat 5, 7, and 8 images (30 m resolution) between 2010 and 2017 that were consistent with CDL data. The projection of Landsat imagery is dominant UTM with a datum of WGS 84. In 2010 and 2011, Landsat 5 data were preferred due to the SLC failure in the Landsat 7 images. In 2012, only Landsat 7 was used, and the gaps were filled by applying a medium-kernel to the Landsat 7 SLC-off images. For 2013-2017, Landsat 8 data was used as the main source, supplemented by Landsat 7 images. For each growing season, Landsat images during the last two weeks of July were preferred to represent variation in crop growth as there is a high correlation between NDVI and crop yield during this period (Maestrini & Basso, 2018b). While other satellite imagery is also available, National Aeronautics and Space Administration (NASA) Landsat provides the longest continuous quality-assured record at the appropriate scale for subfield analysis.

Areas with cloud cover were replaced with clear pixels from images collected during adjacent periods. Two thresholds were applied to identify cloudy pixels: one threshold of 0.2 was applied to the near infrared (nir) band and another threshold of 1 to the simple cloud-likelihood score image, an internal index in Google Earth Engine derived using a combination of brightness, temperature, and normalized difference snow index (NDSI). After replacing cloudy areas, one composite Landsat image covering ten states was created for each year. The NDVI image was then calculated using red and nir bands of this composite image (Rouse Jr et al., 1973).

Common Land Unit (CLU). The CLU data layer is a standardized Geographic Information System (GIS) layer that characterizes the smallest unit of land with a permanent contiguous boundary and common land cover and management. The CLU was established by the USDA Farm Service Agency (FSA) to map the nation's farm fields, rangeland, and pastureland at a confidence level of 90% with a tolerance of 3m from ground features visible in the imagery (<https://datagateway.nrcs.usda.gov/>). This layer is used to implement farm service programs such as crop monitoring and disaster assistance. We used the CLU polygons as the spatial boundary for within-field yield variability analysis. The spatial reference of the CLU layer is UTM dominant zone with a datum of NAD83. We converted the original NAD83 datum into WGS84 to keep consistent with other data sources.

NASS statistics. County-level maize and soybean yields were obtained by USDA NASS (https://www.nass.usda.gov/Quick_Stats/), and farmer-reported N application rates to maize were provided by the Agricultural Resource Management Survey (ARMS).

N fertilizer rate. Nitrogen fertilizer application rates for individual fields are not directly represented in USDA or other databases. We thus estimated application rates by two independent measures. First are rates reported by farmers in the 2016 USDA ARMS of maize farmers. This

survey, conducted on a 5-year cycle for each major commodity, is based on in-person interviews with farmers (1209 farmers in the 10 states covered in our analysis), and we use reported average rates, with variation among rates within states to provide a measure of uncertainty, for all fields within a county.

The second measure was constructed from university-based recommendations. For six states in the region, university extension recommendations are based on the Maximum Return to Nitrogen (MRTN) calculator (<http://cnrc.agron.iastate.edu/About.aspx>) (Stuart et al., 2014). The MRTN calculator determines an average economically optimal maize N fertilizer rate for fields within IL, IA, MI, MN, OH, and WI based on thousands of research trials conducted for maize following maize and maize following soybean rotations. Because calculator values are generally used to define a baseline value for any given field, with actual fertilizer recommendations rates usually greater than MRTN calculated values, the calculator provides a second conservative estimate of fertilizer use in the region.

We used the MRTN calculator to estimate optimal fertilizer rates using a common 1:10 price ratio of maize and N fertilizer (USD 4/bu and USD 0.42/lb, respectively, equivalent to USD 157/MT and USD 210/MT, respectively) (Sawyer et al., 2006). For states where MRTN data were not available, we used ARMS rates plus 10.3%, the average difference between MRTN and ARMS rates for all states with both data available. We held the amount of applied N constant from 2010 to 2017.

Crop residues are assumed to be recycled internally and to not accumulate or increase SOM, thus they are not considered here.

2.3.1 NDVI Stability Classes

Crop yield stability classes have long been used to create management zones within the field based on inter-annual yield variation (Basso et al., 2007; Blackmore, 2000; Lark, 1998; Maestrini & Basso, 2018a, 2018b). Typically, data are collected from georeferenced yield monitor systems mounted on harvesters, with yield stabilities resolved to a few m².

We estimated yield stabilities in the absence of high-resolution yield monitor data by analyzing the year-to-year variability of satellite-derived NDVI for 0.09 ha subfield areas within individual fields. We used eight years of available imagery and CDL data to determine how a given 30×30 m subfield pixel changed from year to year relative to the mean NDVI for the entire field. This analysis was performed to determine areas that had, over the studied years, 1) consistently higher NDVI compared to the mean NDVI of the field (high & stable yields, SH), 2) a consistently lower NDVI (low & stable yields, SL), or 3) an inconsistent (lower some years, higher others) NDVI (unstable yields, U).

For the i th year, the average NDVI value of the j th CLU polygon was calculated and all pixels within this polygon were classified into two types (higher than average of the field and lower than average of the field for a given year):

$$\overline{NDVI}_{i,j} = \frac{1}{n} \sum_{k=1}^n NDVI_{i,j,k} \dots\dots\dots \text{Equation (3)}$$

$$rNDVI_{i,j,k} = \frac{NDVI_{i,j,k} - \overline{NDVI}_{i,j}}{\overline{NDVI}_{i,j}} \dots\dots\dots \text{Equation (4)}$$

$$snNDVI_{i,j,k} = \begin{cases} \text{High NDVI} & \text{if } NDVI_{i,j,k} \geq \overline{NDVI}_{i,j} \\ \text{Low NDVI} & \text{if } NDVI_{i,j,k} < \overline{NDVI}_{i,j} \end{cases} \dots\dots\dots \text{Equation (5)}$$

where n and k indicate the total number of pixels and k th pixel in the j th polygon. $rNDVI$ is relative NDVI, and $snNDVI$ is the spatially normalized NDVI for a given year.

We then determined the temporal variation of $snNDVI$ expressed as degree of stability, or $tnNDVI$ (temporally normalized NDVI). We calculated year-by-year variation for each maize and soybean pixel and then classified each pixel as stable high, stable low, or unstable NDVI, using the following equations:

$$\overline{snNDVI}_{j,k} = \frac{\sum_{i=1}^8 snNDVI_{i,j,k}}{8} \dots\dots\dots \text{Equation (6)}$$

$$tnNDVI_{j,k} = \sqrt{\frac{1}{8} \sum_{i=1}^8 (snNDVI_{i,j,k} - \overline{snNDVI}_{j,k})^2} \dots\dots\dots \text{Equation (7)}$$

$$Stability_{j,k} = \begin{cases} SH & \text{if } NDVI_{i,j,k} \geq \overline{NDVI}_{i,j} \text{ and } tnNDVI_{j,k} < 0.15 \\ SL & \text{if } NDVI_{i,j,k} \geq \overline{NDVI}_{i,j} \text{ and } tnNDVI_{j,k} < 0.15 \\ U & \text{if } tnNDVI_{j,k} \geq 0.15 \end{cases} \dots\dots\dots \text{Equation (8)}$$

Stable high NDVI pixels were identified where the pixelwise NDVI over the eight years was always greater than the average of the field with a $tnNDVI$ less than 0.15. Stable low NDVI pixels were identified where the mean NDVI for each pixel for the eight years was always lower than the average of the field with a $tnNDVI$ less than 0.15. Unstable areas were identified where the $tnNDVI$ was greater than 0.15.

2.3.2 Yield and N Uptake Estimates

We estimated subfield yields by deconvoluting observed USDA NASS county-level yields for any given year into high and low yield areas using Equations 3-5 based on their respective

NDVI values to maintain the proportionality of differences among NDVI using the following equations:

$$Yield_{i,j} = Area_{high,i,j} \times Yield_{high,i,j} + Area_{low,i,j} \times Yield_{low,i,j} \dots\dots\dots \text{Equation (9)}$$

$$\frac{Yield_{high,i,j}}{Yield_{low,i,j}} = \frac{\overline{NDVI}_{high,i,j}}{\overline{NDVI}_{low,i,j}} \dots\dots\dots \text{Equation (10)}$$

Where $Area_{high,j}$ and $Area_{low,j}$ indicated the acreage of high and low NDVI areas, respectively.

$\overline{NDVI}_{high,j}$ and $\overline{NDVI}_{low,j}$ are mean NDVI values for area. We directly assigned the county-level yield to $Yield_{i,j}$ based on the assumption that summed crop yields for all fields within a county are equal to the county-level yield.

Grain N uptake (NUP) was determined by multiplying yield by grain N percent (1.2%) (Boone et al., 1984; Ciampitti & Vyn, 2012). Residues are assumed to be retained on the soil and decomposed. NUE, defined as kg grain kg⁻¹ N applied assuming no change in total soil N (Ladha et al., 2005), was derived by dividing grain N uptake (NUP) by N applied (N_{app}).

$$NUE = NUP / N_{app} \dots\dots\dots \text{Equation (11)}$$

2.3.3 Nitrogen Balance Calculations

Estimated N loss for a given subfield area was calculated as the difference between N fertilizer applied and grain N output for a given year (Conant et al., 2013; Cui et al., 2018; Lassaletta et al., 2016). Grain N output was calculated as yield × N content; modern grain varieties have a remarkably consistent grain N content of 1.2% (Boone et al., 1984; Ciampitti & Vyn, 2012).

This difference assumes stable internal stores of N (as soil organic matter including crop residue), since any increase in internal storage might otherwise be mis-attributed to N loss. Soil organic matter is known to be stable or declining across the US Midwest except where permanent no-till or cover crops are present (Baranski et al., 2018; Horowitz et al., 2010), so we expect no regional changes in soil organic matter that would significantly reduce our loss estimates. Our calculation also assumes there are no significant changes in other N inputs across the region. Nitrogen inputs additional to fertilizer in these systems include atmospheric N deposition and N fixed by a prior soybean crop. We do not include either in our analysis because both are negligible ($<10 \text{ kg N ha}^{-1} \text{ yr}^{-1}$) in comparison to the amount of N fertilizer used, and their inclusion would, in any case, increase rather than decrease our estimates of N loss. Thus their exclusion makes our environmental loss term even more conservative.

We do not differentiate between hydrologic and gaseous losses of N except to estimate the amount of nitrous oxide (N_2O) emissions (both direct and indirect) from added fertilizer, calculated using IPCC emission factors (Hergoualc'h et al., 2019).

2.3.4 Monetary value and Environmental losses

Direct monetary value and environmental losses were calculated as:

$$\text{USD Loss (USD / ha)} = 0.42 \text{USD / kg} \times \text{NLoss} \dots\dots\dots \text{Equation (12)}$$

$$\text{Energy Loss (MJ / ha)} = 70 \text{MJ / kg} \times \text{NLoss} \dots\dots\dots \text{Equation (13)}$$

$$\text{CO}_2 \text{ equivalent (kg / ha)} = \text{NLoss} \times 6.04 \dots\dots\dots \text{Equation (14)}$$

The monetary value of lost fertilizer N is based on the 2017 fertilizer value of USD 210 USD/MT (0.42 USD per kg N). The coefficients used in Equations 13 and 14 were obtained from Hergoualc'h et al. (2019) and Hood & Kidder (1992).

2.3.5 *Yield Stability Validation*

We verified yield estimates (at 30m resolution) for maize and soybeans against data from combine harvester monitors (at 2m resolution) for 508 maize fields across the region (Figure 22-Figure 25). Most fields were in a maize and soybean rotation for at least 3 years. Yield data were collected from combined harvesters equipped with yield monitors and recorded as point data at a 2m interval for each row. Stability maps were generated from yield monitor data using the same procedure as for NDVI images. Comparisons of our calculated yield stability classes vs. high-resolution yield monitor data for the 508 fields are given in Figure 22-Figure 24.

2.4 Results

Figure 1 shows subfield yield stability classes for each field in the region planted with maize or soybeans for at least three years of the eight-year study period, including fields planted continuously to maize, maize-soybeans, or (infrequently) more complex rotations. Croplands that did not meet this three-year requirement appear as white pixels in figures. Examination of county and section-level subregions (Table 1, Figure 1) shows remarkably high yield variation for a region with yields often considered uniformly stable and efficient. On average, 46% of the cropland analyzed had stable high yields, 26% stable low yields, and 28% unstable yields (Table 1).

There were remarkably few differences among states with respect to the proportion of areas in each class. Stable high yield areas constitute 38–52% of total maize and soybean cropland across these states, with the highest proportions in Wisconsin, Iowa, Minnesota, and the Dakotas (Table 1)—all located in the northwest portion of the region. Likewise, the proportion of land in the low stability class was between 19 and 31% among all states, with unstable areas constituting the remaining 16–38% of total maize and soybean cropland, with the highest proportions in Michigan, Indiana, and Ohio—the easternmost states in the region.

We calculated fertilizer application rates based on the USDA ARMS survey and university-based recommendations. N fertilizer rates from ARMS (Table 15), not including manure inputs, averaged $160 (\pm 24 \text{ SD}) \text{ kg N ha}^{-1} \text{ yr}^{-1}$ across the 10 states, ranging from 117 to $197 \text{ kg N ha}^{-1} \text{ yr}^{-1}$. These values are self-reported by farmers and lower than university recommendations, and thus provide a conservative, likely minimum fertilizer rate for maize.

University-based N fertilizer recommendations are based on the maximum return to nitrogen (MRTN) database for most states in the region (Sawyer et al., 2006). MRTN rates are ~10% higher than those reported in the ARMS survey and range from 129 to $209 \text{ kg N ha}^{-1} \text{ yr}^{-1}$ (Table 2), for an average rate of $177 (\pm 27 \text{ SD}) \text{ kg N ha}^{-1} \text{ yr}^{-1}$. Farmers in general and U.S. farmers specifically more often use N fertilizer recommendations from fertilizer and seed dealers than from university extension, and when used, MRTN rates are commonly used as starting points for N rate decisions (Pannell, 2017; Stuart et al., 2014). We thus set our maximum range to 20% greater than local MRTN values to obtain a plausible bracketing.

Using our remotely sensed based crop yields and estimated N fertilizer rates, we calculated average annual N uptake, NUE, and surplus N loss for each stability class (Table 2). Estimated reactive N losses (surplus N) from stable high yield areas ranged from none (indicating the partial use of another source of N such as manure or residual N fixed by soybeans in maize-soybean rotations) to $142 \text{ kg N ha}^{-1} \text{ yr}^{-1}$. Annual N losses from stable high yield areas averaged only 51 kg N ha^{-1} . In contrast, estimated N losses from stable low yield areas averaged 83 kg N ha^{-1} , and unstable yield areas had intermediate N losses of 63 kg N ha^{-1} . NUE varied correspondingly (Table 2), from an average of 76% for high yield areas to 57% for low yield areas. In unstable subfield areas NUE averaged 70%.

We estimate that over-fertilization of stable low yield areas every year and of unstable areas in low yielding years costs farmers in the region ~USD 485 (267-702) million per year in unused N fertilizer lost to the environment (Table 3). Losses would be even higher in years with very unfavorable growing conditions, such as 2012 when most unstable areas had low yields. At a small watershed scale, county-level surface water nitrate concentrations from USGS appear well correlated with N loss patterns predicted by crop stability classes (Figure 26) (Van Metre et al., 2016).

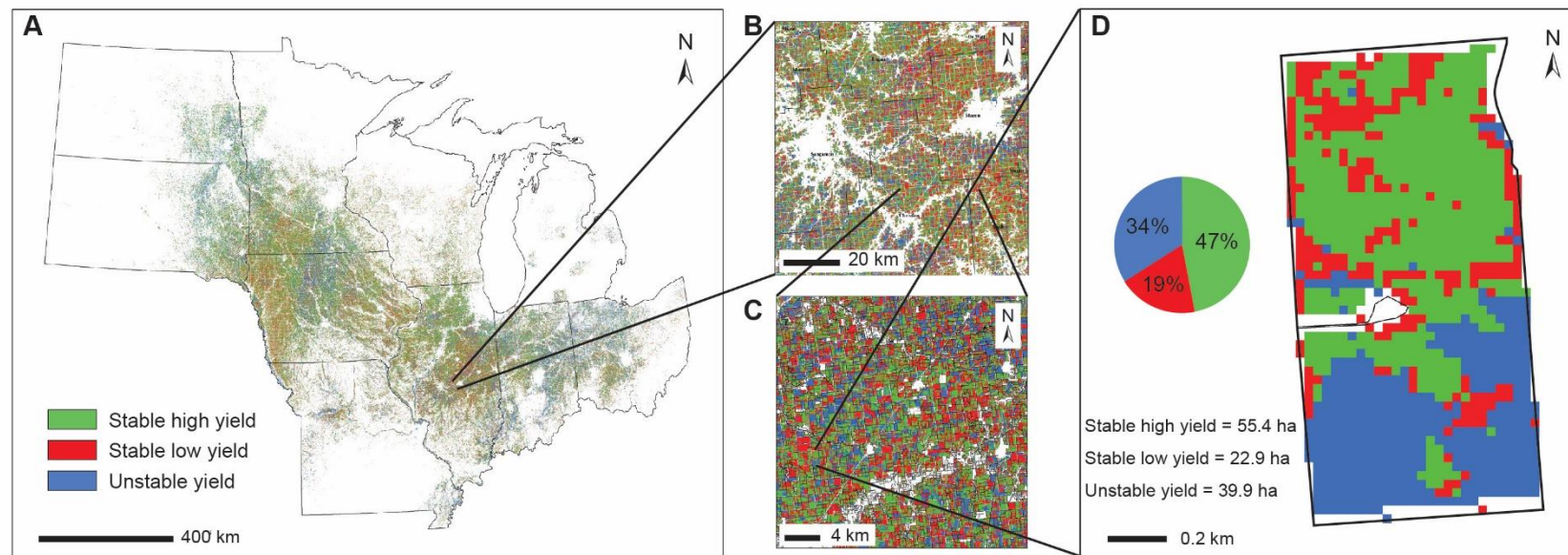


Figure 1 Crop yield stability maps for (A) ten U.S. Midwest states and subregions of (B) 10,000 km², (C) 196 km², and (D) 118 ha. Colors represent yield stability areas for 0.09 ha portions of fields planted to maize or maize-soybean for at least three years during 2010-2017 (~30 million ha in total).

Table 1 State-level yield stability trends for 2010-2017.

State	Area (ha)	Percentage of area (%)*			Unstable area yield class (%)	
		Stable high yield	Stable low yield	Unstable yield	High yield	Low yield
Illinois	6,516,484	47 (± 7)	30 (± 5)	23 (± 8)	64	36
Indiana	3,153,424	41 (± 7)	25 (± 4)	34 (± 10)	64	36
Iowa	7,497,549	51 (± 8)	31 (± 8)	19 (± 14)	69	31
Michigan	121,673	38 (± 11)	24 (± 10)	38 (± 17)	64	36
Minnesota	3,894,599	51 (± 8)	23 (± 6)	26 (± 11)	67	33
Missouri	1,414,243	41 (± 10)	29 (± 8)	30 (± 15)	61	39
North Dakota	704,829	50 (± 13)	19 (± 12)	31 (± 14)	67	33
Ohio	1,830,759	42 (± 8)	27 (± 7)	31 (± 13)	62	38
South Dakota	2,064,051	51 (± 14)	22 (± 11)	28 (± 19)	68	32
Wisconsin	867,204	52 (± 5)	31 (± 3)	16 (± 5)	68	32
Average		46	26	28	65	35

*Numbers after $\pm$ are the standard deviation values calculated from county-level stability statistics

Table 2 N removed by harvest, N fertilizer surplus, and apparent N use efficiency (NUE) within yield stability classes.

State	Fertilizer N Rate	Harvested N			Surplus N			NUE		
		Stable High	Stable Low	Unstable	Stable High	Stable Low	Unstable	Stable High	Stable Low	Unstable
IL	179 – 229	145	108	131	34 – 84	71 – 121	48 – 98	63 – 81	47 – 60	57 – 73
IN	168 – 251	135	99	122	33 – 116	69 – 152	46 – 129	54 – 80	39 – 59	49 – 73
IA	152 – 208	148	115	138	4 – 60	37 – 93	14 – 70	71 – 97	55 – 76	66 – 91
MI	143 – 198	131	95	118	12 – 67	48 – 103	25 – 80	66 – 92	48 – 66	60 – 83
MN	154 – 212	145	115	135	9 – 67	39 – 97	19 – 77	68 – 94	54 – 75	64 – 88
MO	189 – 260	118	85	105	71 – 142	104 – 175	84 – 155	45 – 62	33 – 45	40 – 56
ND	137 – 190	111	83	102	26 – 79	54 – 107	35 – 88	58 – 81	44 – 61	54 – 74
OH	155 – 234	136	100	122	19 – 98	55 – 134	33 – 112	58 – 88	43 – 65	52 – 79
SD	134 – 182	118	89	109	16 – 64	45 – 93	25 – 73	65 – 88	49 – 66	60 – 81
WI	111 – 155	132	100	122	0 – 23	11 – 55	0 – 33	85 – 119	65 – 90	79 – 110
Total Average	152 – 212	132	99	120	22 – 80	53 – 113	33 – 92	63 – 88	48 – 66	58 – 81

Values are for maize in stable high yield, stable low yield, and unstable yield areas by US state. All values are kg N ha⁻¹ y⁻¹ except NUE is kg grain N kg⁻¹ N fertilizer.

Table 3 Surplus fertilizer N loss from stable and unstable low yield areas and its monetary value, embedded energy, and associated CO₂-equivalent emissions.

State	Surplus N Loss (Gg y ⁻¹)	Monetary Value (million USD y ⁻¹)	Embedded Energy (10 ⁶ GJ y ⁻¹)	CO _{2eq} Emissions (Mt y ⁻¹)
IL	192 – 380	80.6 – 159.5	13.4 – 26.6	1.2 – 2.3
IN	88 – 241	36.9 – 101.1	6.1 – 16.9	0.5 – 1.5
IA	87 – 304	36.6 – 127.6	6.1 – 21.3	0.5 – 1.8
MI	6 – 17	2.3 – 7.0	0.4 – 1.2	0.0 – 0.1
MN	57 – 204	23.9 – 85.8	4.0 – 14.3	0.3 – 1.2
MO	72 – 132	30.3 – 55.6	5.0 – 9.3	0.4 – 0.8
ND	35 – 85	14.7 – 35.7	2.6 – 5.9	0.2 – 0.5
OH	40 – 133	16.7 – 55.8	2.8 – 9.3	0.2 – 0.8
SD	51 – 134	21.3 – 56.1	3.5 – 9.3	0.3 – 0.8
WI	8 – 43	3.6 – 18.1	0.6 – 3.0	0.1 – 0.3
Total Average	1155 (636 – 1673)	485 (267 – 702)	78.8 (45 – 113)	6.8 (3.5 – 10.1)

Embedded energy refers to the energy cost of producing surplus N. CO₂-equivalents are greenhouse gas emissions during fertilizer manufacture plus nitrous oxide emissions from applied fertilizer. The N fertilizer price used in this study is the 2017 price of 210 USD/MT.

2.5 Discussion

Stable high yield subfield areas made up 46% of maize-soybean cropland in the 10 Midwest states analyzed here and appear to utilize fertilizer N much more efficiently than low yield areas, which appeared incapable of supporting crop growth at a level sufficient to utilize the majority of the N applied. As a result, stable low yielding areas appeared to contribute most of the reactive N lost to the environment (~44%), with another 31% lost from unstable yield areas in years with low yields (Table 1, Table 2, Table 16). The cost of this lost N is substantial, both to farmers in terms of the monetary value of unused fertilizer N and to society in terms of reactive N that pollutes aquifers, inland and coastal surface waters, and adds to the atmosphere's greenhouse gas burden.

All states had a similar proportion of stable high yield subfield areas (38-52%) and stable low yield areas (19-31%), though there appeared to be a greater percentage of unstable areas in eastern states. This greater percentage may be due to a greater proportion of fields in eastern states with shallower soils and rolling terrains, which create greater dependence on rainfall amounts and distributions compared to prairie states like Iowa and Minnesota, which in general have deeper soils and greater soil water availability. Unstable areas were high yielding 65% of the time, on average (Table 1).

How practical might geospatial N fertilizer recommendations based on crop yield stability classes be in reality? Non-commercial satellite imagery sufficient to generate NDVI-based yield stability maps is available for most if not all fields in the U.S., and most grain farmers also have access to harvest combine monitors that can produce annual yield maps for individual fields with even greater precision than satellite imagery. Once stable low yield areas are identified, N fertilizer rates could be adjusted downward to match consistently low crop yields in those areas and thereby

reduce both economic costs and environmental harm (Basso et al., 2016; Basso, Ritchie, et al., 2011). Variable rate technology for N fertilizer application is commercially available if not, before now, profitable (Schimmelpfennig, 2016). In unstable yield areas, a modeling-based strategy that, at the time of fertilization, incorporates recent and projected weather could better match N fertilizer rates to that year's crop needs (Basso, Ritchie, et al., 2011).

Our estimated cost savings for avoided fertilizer use covers only the monetary value of lost fertilizer, and thus represents only one facet of a full economic analysis that must also include, at the farm scale, the costs of equipment, labor, and information handling, and at the societal scale, the costs of environmental degradation and mitigation. The absolute economic savings, while likely significant, are thus uncertain.

However, our analysis applies only to N fertilization. Performing this analysis for other inputs such as phosphorus, lime, seeds, herbicides, and labor may show that stable low yielding areas are not only environmentally vulnerable but are also economically unprofitable (Basso et al., 2016; Basso, Sartori, et al., 2011). In such cases it may be that these areas are better planted to conservation strips (Schulte et al., 2017) or to perennial biofuel crops (Brandes et al., 2018; Davis et al., 2012; Robertson et al., 2017) .

There are several other sources of uncertainty in our analysis. Perhaps the greatest is our estimate of farmer N fertilizer use. Because there are no verifiable sources of reliable N application rates in the U.S., we used a range bracketed by the ARMS survey of self-reported rates for minimum likely values and university recommended rates plus 20% for likely maximum values. We believe this provides a reasonable, conservative estimate of likely minimum and maximum fertilizer rates used by farmers. Estimates of NUE (Table 2), however, suggest that even these fertilizer estimates are too conservative insofar as our NUE estimates are higher than those reported

for field studies (Ladha et al., 2005). If so, then we might expect NUE for low yielding areas to be <50%, with correspondingly higher N losses. The relative importance of different yield stability classes, however, should remain unchanged.

Another source of uncertainty is our assumption that farmers apply N fertilizer at similar rates to all fields cropped similarly. Direct evidence for this is lacking, but we infer that this is mostly the case from very low coefficients of variation for farmer-reported N fertilizer rates in the ARMS survey (Table 15).

And finally, uncertainties in yield stability validation, while low (Figure 22), stem from mismatched resolutions between yield monitor data (2m) and our satellite-based approach (30m), and by the fact that much more spatial variation is captured using high-resolution yield monitor data from harvesters at their finer resolution. The time interval between NDVI imagery and yield monitor data can introduce additional error.

Multiple approaches are needed to address the widespread problem of excess agricultural N in the environment (Lassaletta et al., 2016; Robertson & Vitousek, 2009). Within and edge of field remediation measures as well as 4R (right source, right rate, right time, and right place) approaches to improving fertilizer use efficiency offer promise for reducing the environmental burden of agricultural N use (Christianson et al., 2012; IPNI, 2012; Jaynes & Isenhardt, 2014; Roley et al., 2016; Schulte et al., 2017; Tonitto et al., 2006; Zhou et al., 2010). Variable rate N fertilizer application based on subfield yield stability as described here provides an additional, nonexclusive means for meeting environmental goals thus far difficult to achieve.

2.6 Conclusion

Evaluating the effectiveness of fertilizer application at subfield scale is important for farmers to adjust their strategies to increase economic return and cause less environmental damage.

This study utilized public satellite images to perform yield stability analysis in the U.S Corn Belt and assessed nitrogen use efficiency for each stability class. Results showed that around half of cropland produced stable high yield, with equal proportion of stable low and unstable yielding area. More than one third of nitrogen was possibly not absorbed by plants in the stable low area, implying that attention should be paid to considering historical yield in determining application practices.

2.7 Acknowledgements

We thank E. Anderson of NASS for access to county-specific fertilizer rate variability information in the ARMS survey, and J.A. Andresen, J.W. Jones, T.G. Reeves, J.T. Ritchie, and anonymous reviewers for many insightful comments on earlier versions of the manuscript. Financial support was provided by USDA/NIFA (award 2015-68007-23133), the U.S. National Science Foundation's Dynamics of Coupled Natural and Human Systems Program (award 1313677), the NSF Long-term Ecological Research Program (award 1637653), the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (awards DE-SC0018409 and DE-FC02-07ER64494), and Michigan State University AgBioResearch.

CHAPTER 3: POTENTIAL OF USING YIELD STABILITY MAP TO SUPPORT SUSTAINABILITY OF AGRICULTURAL LANDSCAPE

3.1 Abstract

Despite that previous research has revealed the benefits of switchgrass and cover crops, current adoption rates are still low across the U.S. This study examined the potential of yield stability map to offer opportunities to convert areas with stable low yield (LS) into switchgrass. In addition to 10 states in Chapter 1, we added Arkansas, Kansas, Nebraska, and Pennsylvania in this study. We first estimated the proportion of LS along the field boundary and found that about 40% of LS areas were located from 0 – 30m from field edges and over 90% of LS areas were connected in patches of at least 3 pixels, which was set as the minimum area for switchgrass production. Then we ran the SALUS model to simulate switchgrass growth and development using a calibrated cultivar parameterized in previous research, and also rye as a cover crop in the area without switchgrass production after crop harvest. The simulated rye biomass has been calibrated with 50 values reported from public literature and attained a correlation coefficient of 0.65. On average, the estimated maize, soybean, switchgrass, and rye yields were 8.6, 2.3, 5.9, and 0.8 Mg/ha in the LS areas. Soil organic carbon content increased $0.1 \text{ Mg ha}^{-1} \text{ y}^{-1}$ after 10 years of switchgrass and $0.3 \text{ Mg ha}^{-1} \text{ y}^{-1}$ for continuous maize under no-till, while 0.2 and $0.4 \text{ Mg ha}^{-1} \text{ y}^{-1}$ SOC loss happened under continuous maize and maize/soybean rotation scenarios with conventional tillage. Little SOC change was observed when modeling rye after crop harvest, but the soil nitrate level decreased by 20% annually after rye harvest in the LS areas. The life-cycle assessment indicated that switchgrass could offset $3.6 \text{ Mg CO}_2\text{eq ha}^{-1} \text{ y}^{-1}$ in the atmosphere, while the maize and soybean rotation resulted in positive GHG emissions. The last benefit explored in this paper is the biocontrol index after the change from maize or soybeans to switchgrass. For the fields in the study

area, the proportion of grassland increased from 16% to 20%, with over 6% increase in Illinois, Iowa, and North Dakota. This led to an increase of 0.025 in the biocontrol index. Overall, this study presented a new solution to manage fields with yield stability to improve the soil, climate, and ecosystem.

3.2 Introduction

The Fourth U.S. National Climate Assessment (NCA4) reported that the annual temperature in the United States has increased by 0.7 °C over the last few decades with increased frequency and intensity of drought and extreme precipitation events, which pose detrimental effects on crop production (Reidmiller et al., 2019). According to previous literature, The warming climate is driven by concentrations of atmospheric greenhouse gases (GHG), including carbon dioxide (CO₂), methane (CH₄), and nitrous oxide (N₂O), which will contribute to further increase in temperature and variability in climate (Easterling & Fahey, 2018). GHG emissions from agricultural sector account for around 9.3% of total U.S. emissions in 2018 (618 MMT CO₂ eq), with around 80% emissions from agricultural soil management through fertilizer applications (Hockstad & Hanel, 2018). Croplands expanded by over 1 million acres per year from 2008-16, and this land use change produced additional GHG emissions (T. J. Lark et al., 2020). Nevertheless, it is possible to remove carbon from the atmosphere into soils, the largest terrestrial C pool, and mitigate climate change with a suite of management practices. In 2015, the French Ministry of Agriculture proposed the “4 per 1000 initiative: soils for food security and climate” at the Paris climate conference and suggested that a 0.4% increase in soil organic carbon (SOC) in actively managed agricultural fields could provide a net CO₂ removal of ~1 Gt C y⁻¹ globally, almost equivalent to a large emitter such as the European Union (Chabbi et al., 2017; Soussana et

al., 2019). This initiative targets improving soil quality, especially for degraded soils, and enhancing future food security.

As a part of soil organic matter (SOM), SOC is one of the most important indicators of soil health as it is involved in biological, physical, and chemical processes. Soils with higher SOM content has higher potential crop production than low SOM levels (Oldfield et al., 2019). There are many specific benefits: 1) reduced fertilizer input requirements due to nutrient release from SOM; 2) better water holding capacity in the soil could alleviate some drought effects (Hudson, 1994); and 3) reduced erosion with stable soil aggregates (Lal, 2005). Minimal disturbance of soil and the addition of organic materials are two primary practices for building carbon in the soil. No-till or reduced tillage methods leave crop residues on the ground to prevent soil losses from the intense precipitation, especially during the fallow period (Montgomery, 2007). Rotation with crops that return residues to the soil, cover crops, organic fertilizers, and perennial biomass crops, such as switchgrass, are all excellent choices for improving the SOC (Halvorson et al., 2002; Liebig et al., 2008). Similar to conservation tillage, cover crops are primarily planted to provide continuous ground cover and biomass into soil after the crop harvest (Langdale et al., 1991). Cover crops, such as cereal rye, clovers, and sorghum, have a positive effect on soil structure and living organisms in the soil. Wulanningtyas et al. (2021) observed significant improvements in SOC and some soil chemical properties in surface soil under 18 years of no-till or cover crop (rye/hairy vetch) management. Switchgrass is a native perennial C4 grass with a deep root system in North America, showing great productivity even in degraded or marginal lands. Switchgrass offers greater potential in reducing CO₂ emissions than forest and grasslands (Lewandowski et al., 2003). A recent study found that switchgrass could accumulate 0.2 – 1.3 Mg C ha⁻¹ year⁻¹ with low vulnerability to predicted climate change impacts (4% – 6% decrease on average) (Martinez -

Feria & Basso, 2020). Meanwhile, switchgrass is also one of the principal biofuel crops and could produce similar amounts of ethanol as maize with much fewer inputs. In addition to increasing SOM, adding cover crops or switchgrass in the cropland also increases the biodiversity in the landscape thus supporting multiple ecosystem services, such as pest suppression, pollination, and conservation of bird species. These ecosystem services are also essential to advance the sustainability of agricultural systems.

Extensive research on the benefits of cover crops and switchgrass has been done for about two decades, however, the scale of adoption of both crops is still small. A cover crop summary provided by Economic Research Service (ERS) in 2021 indicated that cover crop adoption only increased from 3.4 percent in 2012 to 5.1 percent in 2017, with the most increase in Maryland and the Eastern US (Wallander et al., 2021). According to the USDA census of agriculture, the harvested switchgrass covered fewer than 1000 acres, while the harvested hay covered 57 million in 2017 (Hanson et al., 2020). The primary adoption limitation may be the high economic cost and extra labor cost despite government financial support such as the Environmental Quality Incentives Program (EQIP) and the Conservation Stewardship Program (CSP). Farmers also lack experience with managing these crops. Additionally, Runck et al. (2020) estimated that around 4% of the current production area (1.4 million hectares) is needed to supply cover crop seeds for the US maize production area, with rye and hairy vetch requiring as much as 4.5% and 12% respectively. Similarly, Gelfand et al. (2013) indicated that millions of hectares of existing cropland and marginal land could be converted into switchgrass production to satisfy the mandates for low-C biofuels (Calvin et al., 2021; Martinez-Feria et al., 2022). Switchgrass production could even result in a net increase in emissions when the carbon released from land use conversion outweighs the

reductions in bioenergy system emissions. Both economic and environmental costs on such a scale could significantly impact current food production systems.

Regarding the current low adoption scale, Brandes et al. (2018) offered the opportunity of integrating perennials into agronomically and economically underperforming parts of maize/soybean fields. In this study, underperforming parts of maize/soybean fields were converted into perennials without expanding the additional land use into cropland. This idea promoted the re-design of agricultural landscapes for sustainability. However, there are two major difficulties in implementing this method in the actual world: 1) the underperforming area is only determined by soil properties, which actually only account for a part of yield variation; and 2) the spatial variation of crop yield also changes by year, and this temporal variability makes it challenging to plant perennials which require areas that consistently underperform along years. Instead, yield stability maps can help re-design subareas within fields that are persistently unprofitable or environmentally unsustainable, the low and stable (LS) area (Basso & Antle, 2020).

The objectives of this project are to: 1) analyze the spatial distribution of LS area within the field using yield stability maps developed in our previous research; and 2) propose alternative crops in the LS area which produce low economic values and high environmental burden. Specifically, we proposed to use switchgrass in the LS. In the meantime, cover crop, rye, was introduced on the rest of the stability zones during the winter. We ran the SALUS model to simulate switchgrass and cover crop biomass, as well as differences in SOC, soil nitrate level. We also conduct a life cycle assessment to estimate the net emissions in bio-fuel production. Further, we quantified the biological and ecological benefits by changing the existing LS into switchgrass. These alternative scenarios provide a conservative opportunity for bringing environmental benefits that could be realized with yield stability map.

3.3 Study area and Dataset

This research extended the study area from Chapter 1 to include Arkansas, Kansas, Nebraska, and Pennsylvania. These states also produce a large amount of maize, especially for irrigated fields in eastern Kansas and Nebraska. In 2021, the study area produced 11,580,425 thousand bushel of maize, which accounted for 77% of total maize production in the U.S (Sassenrath et al., 2022). Nebraska has 8.3 million irrigated acres, which account for 14.9% of the total irrigated acreage in the U.S. Irrigated maize production occurs on over one-half of all irrigated acres in Kansas, with the highest acreage peak occurring in 2013 at almost 2.0 million acres (Rogers & Lamm, 2008). The climates in Kansas and Nebraska are typical Midwestern climate with hot summers and cold winters. The maize planted area is located in the wetter region while western region is drier. Arkansas has a humid subtropical climate with a significant amount of precipitation every month, and Pennsylvania is characterized by a humid continental climate with annual rainfall of 40 inches.

Yield stability mapping. To evaluate the within-field spatial distribution of crop yield stability classes, we used the satellite-based maize/soybean yield stability maps created in Chapter 1. Stability maps for Arkansas, Kansas, Nebraska, and Pennsylvania were produced in this study following the same algorithm. The unstable zone was further classified into hilltop, depression, and slope (constant slope/backslope or flat) using the 1/3 arc-second (approximately 10 meters) DEM datasets downloaded from the USGS National Elevation Dataset (Arundel et al., 2015). This classification is based on the observation that crop yield within unstable field zones is primarily affected by the interactions between weather and landscape position (Martinez-Feria & Basso, 2020). Specifically, DEMs were used to calculate slope and topographic position index (TPI) at pixel scale under a neighborhood of 9×9 pixels, and then classification was conducted using

positive/negative local standard deviation of TPI values as the threshold. The common land unit (CLU) polygons were used as field boundaries.

Environment dataset for SALUS modeling. The daily gridMET weather dataset at high-spatial resolution (~4-km, 1/24th degree) surface meteorological data was used from 1989 to 2019 (Abatzoglou, 2013). This dataset has been validated through an extensive network of weather stations across the western United States. Gridded Soil Survey Geographic (gSSURGO) soil data were used for all the states at a spatial resolution of 10 meters (USDA, 2011). Unique soil types in the SSURGO rasters is defined by map unit keys (MUKEYs), which are unique numerical keys used to join the tabular tables with detailed soil properties at multiple layers. These datasets were used to run the SALUS model to simulate switchgrass and rye biomass.

SALUS model validation dataset. To validate the accuracy of the rye biomass simulation, we have collected 55 rye biomass results from these literature (De Bruin et al., 2005; Kaspar et al., 2007, 2012; Krueger et al., 2011; Martinez-Feria et al., 2016; Moore et al., 2014; Pantoja et al., 2016). These results were obtained through field experiments conducted in many states and under several soils and cropping systems over more than ten years.

Land cover dataset. We downloaded national land cover datasets (NLCD) from the U.S. Geological Survey for 2011, 2013, and 2016 so we can evaluate the potential ecosystem changes from current maize/soybean planting areas to proposed rye and switchgrass crops. The NLCD are land cover maps at a 30-m resolution that have been used for thousands of applications that require information on land use or land cover change over time. These thematic maps categorize land cover into several classes, such as water, urban, agriculture, forest, etc.

3.4 Method

3.4.1 *Spatial analysis of yield stability maps*

We created multiple non-overlapping buffers at specified distances (30, 60, 90, and 120 meters, corresponding to 1, 2, 3, and 4 Landsat pixels) inside each field, which were used to calculate the area proportions for each stability class within different buffers. We reported the state-level statistics on stability area, field edge area, LS area, and LS area at field edge (0 – 30m from field edge).

We also assessed the spatial connectivity of LS areas. The LS pixels were segmented into patches with a unique ID using ENVI software if they are connected in an 8-neighbor configuration. Then we calculated the area of LS patches with at least specific sizes (1, 3, 5, 7, and 9), where size 1 means all LS pixels were included.

3.4.2 *Cover crop/switchgrass simulations*

The SALUS model was used to simulate switchgrass, rye, maize, and soybean biomass and associated changes in SOC and soil nitrate under the current and proposed cropping system.

For all the simulations, we ran the SALUS model in the LS areas since it is vulnerable to nitrogen and monetary loss, shown in Chapter 1. The area-based majority soil type and weather information from the nearest station was used as input. We made further adjustments for the soil properties and rainfall based on the landscape position and historical yield. Specifically, we reduced the soil depth to 30 cm for LS.

Firstly, we simulated maize/soybean and continuous maize cropping systems as the baseline scenario. The model was run for a 11-year from 2009 to 2019. Soil was initialized in 2009 with 30 kg N/ha of inorganic N in the soil. Planting dates for maize and soybean were estimated at county scale using the day of year when 50% was planted according to NASS. This estimation

was done under the following steps: 1) Because progress reports were released in 1-3 week intervals at the state-level, we interpolated daily values linearly between reported dates, and every day after the first report day had a percent planted value; 2) every year we computed the centroids for maize production areas within each state, using county centroids weighted by the total land area planted under maize in each county; 3) we interpolated between state-year centroids by fitting a generalized additive model to the date of planting using Equation 15; and 4) we predicted county-level planting dates using the latitude and longitude based on the county centroids.

$$DP = s(lat, long) + ti(year \times lat) + ti(year \times long) \dots\dots\dots \text{Equation (15)}$$

Where DP is the date of planting at 50% level, $s(lat, long)$ is the two-dimensional smoothing function with Latitude and Longitude as covariates and $ti(year \times lat)$ and $ti(year \times long)$ are tensor functions of the random effect of year and Latitude and Longitude, respectively.

The harvest date was set to October 31st for maize and soybeans. We used conventional tillage (CT) and no-till (NT) in the simulation to test the effect of tillage method on the SOC content. Total nitrogen applications for maize were the same in Chapter 1, and we used the application rate in Missouri for Arkansas, and that in Ohio for Pennsylvania. Nitrogen application was split: 1/ 3 applied at planting and 2/3 with a side-dress at 45 days after planting. Fertilizer depths were 10.16 cm (4 in) at planting and 5.08 cm (2 in) at side-dress. Population was set as 8 seeds/m² for maize, and 38 seeds/m² for soybeans. Row spacing was 30 inches for maize, and 20 inches for soybeans, and planting depths were 2 inches for maize and 1 inch for soybeans. We applied automatic sprinkler irrigation in Nebraska and Kansas with a 30 cm irrigation management depth. the threshold for automatic irrigation was set to 60% of the max available water and stopped at the 90% value.

The second scenario considered the adoption of cover crops. As a common cover crop, rye was applied over winter after the maize or soybean harvest. The RY16 species was used here (Table 17). We conducted experiments under both CT and NT. Rye was planted the day after maize harvest, and we set removal date on May 1st. The maize and soybean planting dates were adjusted to be the day after rye was harvested if needed. Population was set as 200 seeds/m², row spacing was 15 cm, and planting depth was 3 cm. We applied 30 kg N/ha at planting since we found more than 10 percent of simulations produced zero biomass when no fertilizer was applied. The rye simulation model was evaluated against the agronomic yields collected from the literature.

The last scenario we proposed was to plant switchgrass. One possible reason is the low quality of soils to support crop growth, and plants do not respond to the applied fertilizers. Instead, switchgrass needs much less nitrogen compared to maize and is able to produce high biomass even in poor soils. This makes switchgrass an ideal alternative crop in the LS area. For this simulation, we used SW_UL cultivar (Table 18), which has been calibrated for the Midwest (Martinez-Feria et al., 2022). Tillage was done to 4 inches on May 15th, 2009, and switchgrass was planted on June 1st, 2009, instead of May 1st to avoid issues with temperatures below 10 degrees C, the base temperature for the selected switchgrass cultivar to grow. Population was set as 600 seeds/m² with a row spacing of 1 cm and planting depth of 1 cm. We added 56 kg N/ha each year in the SALUS model to as proxy of nitrogen fixation from switchgrass. We harvested yearly on November 1st with a 65% harvesting efficiency ratio (ratio of agronomic yield to peak above ground biomass), which typically ranges from 60% to 80%.

The outputs of the SALUS model include biomass, grain yield, soil organic carbon, and soil nitrate level.

3.4.3 Life cycle assessment for maize and switchgrass

We further assessed the gap in the life-cycle emissions between continuous maize/maize soybean rotation and switchgrass production in the LS areas. According to Martinez-Feria et al. (2022), this was calculated as the difference in life-cycle greenhouse warming intensity (GWI) in both scenarios. The equation for GWI calculation is listed below:

$$GWI = GHG_{Fuel} + GHG_{seed} + GHG_{AgCh} + GHG_{FertSyn} + GHG_{FertApp} + GHG_{N_2O} + GHG_{\Delta SOC} + GHG_{FeedStock} \dots\dots\dots \text{Equation (16)}$$

Where GHG_{Fuel} , GHG_{seed} , GHG_{AgCh} , $GHG_{FertSyn}$, $GHG_{FertApp}$, and GHG_{N_2O} indicated GHG emitted by fossil fuels used for field operations, seed production, agrochemicals other than nitrogen, nitrogen synthesis (4.5 kg CO₂ for a kg of N) (Robertson et al., 2000), nitrogen application operations (26 kg CO₂eq ha⁻¹) (Gelfand et al., 2013), and the biogenic direct and indirect N₂O emissions. GHG related to fuel, seedling, and agrochemicals were not considered here since I only considered the difference caused by nitrogen applications. The N₂O emissions were estimated through the tier 1 method in the IPCC Guidelines for National Greenhouse Gas Inventories (Table 19) (Hergoualc'h et al., 2019). The ΔSOC indicated the change in soil C sequestration associated with fertilizer management, and GHG was calculated by multiplying this change by 44/12 to convert to CO₂eq. Feedstock indicates the potential emissions when biofuel is produced from switchgrass biomass and maize grain. Biofuel production was calculated as the product of the biomass yield and net GHG offset, which was estimated as -0.737 Mg CO₂eq per Mg of biomass. Here we used maize grain yield for biofuel production.

To compare emissions under two scenarios, we used the reported nitrogen rates in Chapter 1 for maize to calculate $GHG_{FertSyn}$ and $GHG_{FertApp}$. The ΔSOC was the annual difference averaged

between successive growing season, and biofuel production was estimated through biomass/yield values from all crops.

3.4.4 *Ecosystem biocontrol index and avian richness under proposed management practices*

Converting LS to grasslands is going to change the composition of the landscape. The current grassland distribution is composited from national land cover datasets (NLCD) in 2011, 2013, and 2016, where pixels are at least twice labeled as grassland in these datasets. This criterion removes grassland that exists within a short period or misclassification in the NLCD dataset. We calculated the grassland composition under maize and switchgrass scenarios using a moving window analysis in ArcGIS. The landscape of a pixel is defined as the area of a radius of 1.5 km around this pixel, which has been shown to influence arthropods in previous studies (Meehan et al., 2012). Then, a biocontrol index (BCI) was calculated before and after planting switchgrass using the relationship developed by Meehan et al. (2012) in Equation 17. Biocontrol refers to suppression of invertebrate crop pests by predatory arthropods in agricultural landscapes.

$$BCI = 0.44 + 0.62 \times \text{proportion of grass in landscape} \dots\dots\dots \text{Equation (17)}$$

Forest was not included in this equation since the effect of forest on BCI was not statistically significant (Meehan et al., 2012).

3.5 Results

3.5.1 *Proportions of stability class along the distance from field boundary*

Of the 35.6 million ha of maize and soybean fields, there was 18.7 million ha of HS, 10.1 million ha of LS, and 6.8 million ha of UN (Table 20Table 21). This result was similar to Chapter 1 after we extended the study area to 14 states. The highest proportion of HS was 66% in Nebraska, with the lowest percentage of UN at 3%. Similarly, HS in Kansas occupied 59% of the field area, and UN was about 12%. On the other hand, Arkansas only had 32% of field acreage as HS, while

half of the field area had unstable yields. State-level variation in the LS percentage was much lower, with the highest percentage of LS (35%) observed in Pennsylvania and the lowest at 13% in North Dakota.

In the area 0 – 30 m from field boundary (field edge), the percentages of HS and LS were 42% and 39%, respectively (Figure 2). As the distance from field boundaries increases, the percentage of HS increases while the percentage of LS decreases. About 59% of field center areas were identified as HS and 22% as LS. The percentage of UN was consistent (19% or 20%) irrespective of location in the field. When we only focus on LS locations (Figure 3), about 30 to 40 percent of LS pixels were in the field edge, which decreased to around 20 percent in the field center. Around half of LS is found in the field edge in Wisconsin, while only 5% of field centers produce stable low yields. One exception is North Dakota where LS was mostly located in the field center.

The connectivity analysis indicated that 93% of LS was contiguous with at least 3 pixels, while the remaining 7% was isolated pixels (Table 4). This percentage dropped at a low rate when the patch size threshold increased, and over 80% of LS areas were spatially connected with at least 9 pixels (2 acres). On average, Iowa and Nebraska had greater proportions of large LS patches than other states.

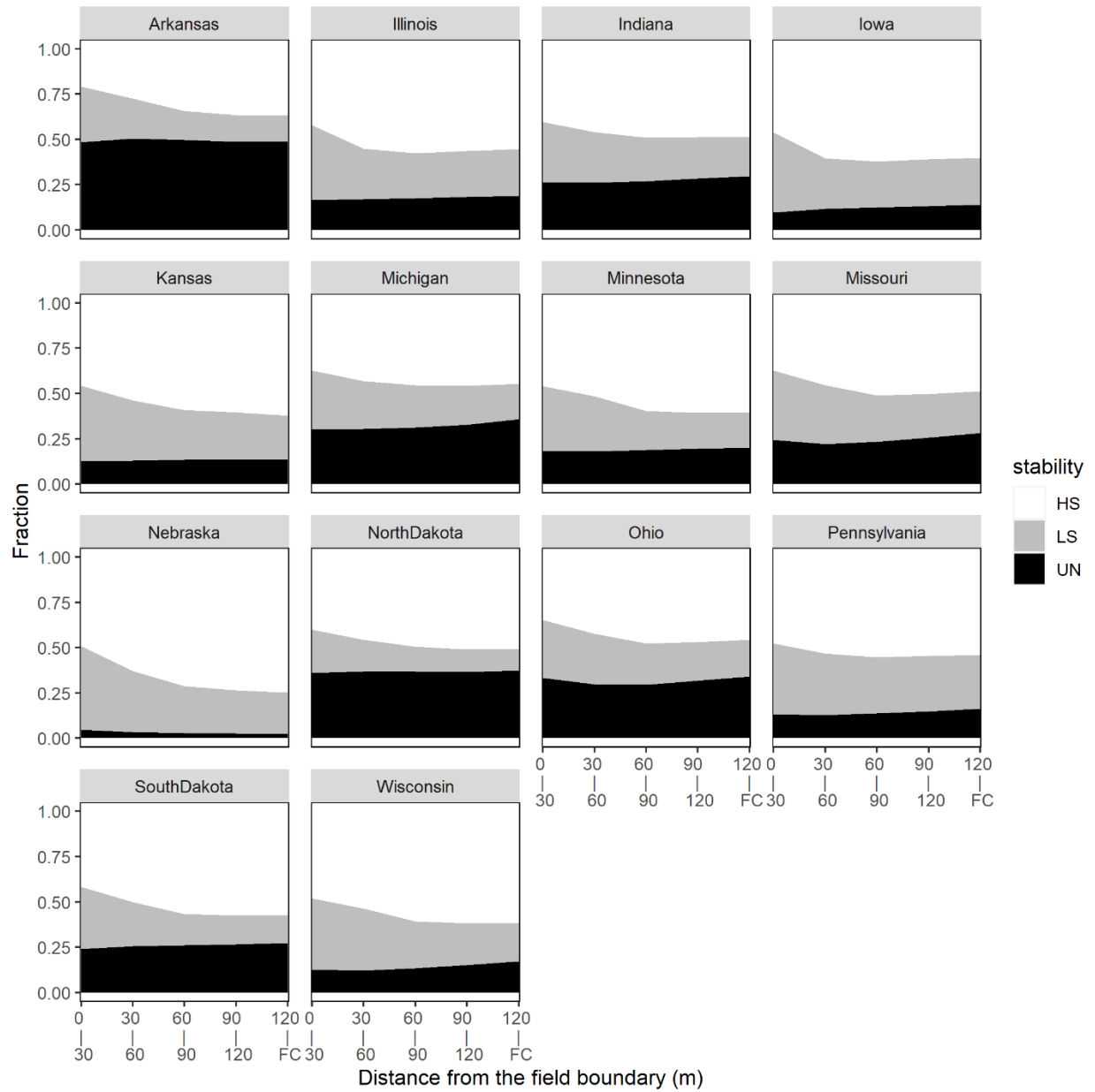


Figure 2 Fractions of stability class within each buffer area. FC indicates field center. Buffer area was determined as the area between two neighboring buffers defined in the method.

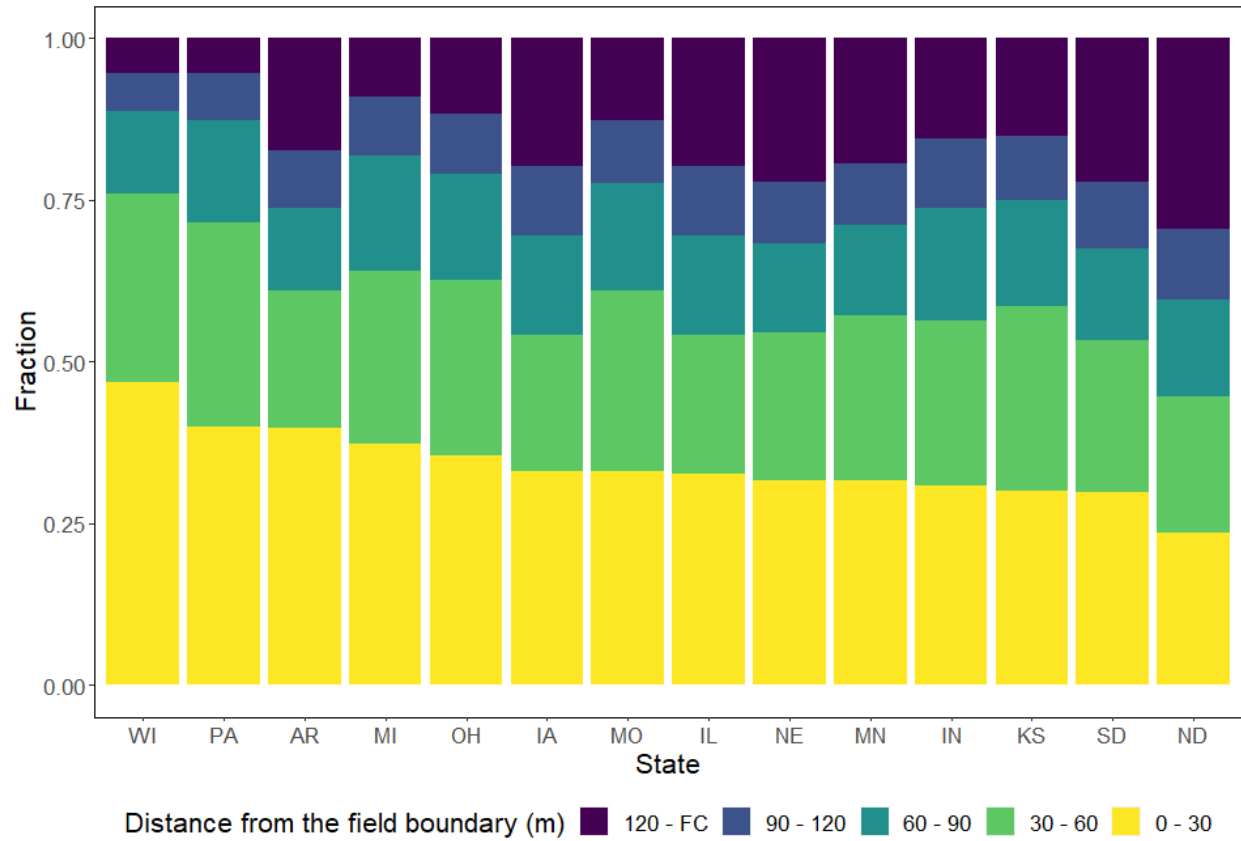


Figure 3 Fraction of LS area with each buffer area to all LS in the field.

Table 4 LS area at different patch size.

State	LS area (ha) with patches that are larger than each size				
	1 pixel	3 pixels	5 pixels	7 pixels	9 pixels
AR	24,609	19,705	17,249	15,684	14,492
IL	2,043,111	1,935,956	1,865,622	1,812,912	1,770,405
IN	796,340	729,133	686,421	654,763	628,945
IA	2,330,849	2,218,900	2,144,690	2,087,593	2,040,746
KS	253,078	232,597	218,478	207,365	198,166
MI	506,268	444,545	406,948	379,915	358,031
MN	993,808	927,382	882,972	848,933	820,832
MO	409,367	375,003	353,185	336,189	321,999
NE	1,396,031	1,336,283	1,292,433	1,257,039	1,227,007
ND	94,624	78,143	69,021	62,918	58,305
OH	456,333	412,475	385,337	365,196	349,023
PA	67,856	55,689	48,809	43,819	39,812
SD	414,243	267,696	251,416	238,802	228,502
WI	268,363	242,262	224,182	209,689	197,407
Sum	10,054,880	9,275,769	8,846,763	8,520,817	8,253,672

3.5.2 SALUS simulations results for switchgrass, cover crop, and maize/soybeans

The simulated state-level average maize yield across different managements was listed in Table 5. Continuous maize with CT produced 8.6 Mg grain ha⁻¹ y⁻¹ for all states, and NT produce very similar maize yields. The highest yield (9.6 Mg/ha) in the LS was observed in irrigated fields in Nebraska, with the lowest yield (5.6 Mg/ha) in Arkansas. Under conventional tillage, the continuous maize with rye (MZ-RY) scenario produced 0.8 Mg/ha less maize grain than the continuous maize (MZ) scenario, partly due to later planting dates was as the day after rye harvest (May 1st). But this difference was not found in the NT situation, which suggested that the NT system could improve the maize yield when planting date is delayed. Adding soybeans in the

rotation again caused positive (0.3 Mg/ha for CT and 0.2 Mg/ha for NT), yet not significant, effects on the maize yield. Similarly, soybean yield was consistent across all the scenarios, which produced an average of 2.3 Mg/ha (Table 6).

We obtained a R^2 of 0.65 and a root mean square error of 620 kg/ha by comparing the simulated and reported rye biomass in the literature, after removing outliers in the red circle from Figure 27. This suggests that our model could provide accurate rye biomass results. Across the states and over 10 years of simulation, mean rye biomass was estimated as ~700 kg/ha and ~900 with CT and NT, respectively, with little difference between MZ and the maize soybean rotation (MZ-SB) scenarios (Table 6). Rye yields in NT system was higher than CT, possibly due to higher soil moisture reserved in the NT. We also found higher rye yield in the southern states compared to the northern states in the study region, which was consistent with previous literature that higher temperature has a positive temperature on the rye growth (Pantoja et al., 2016).

In this study, we used the SOC changes in the soil profile from 0 to 40 cm, which represented over 90% of SOC change through all the soil profiles in the simulation results. SOC decreased at a rate of $0.19 \text{ Mg ha}^{-1} \text{ y}^{-1}$ under MZ with CT, while NT resulted in a positive carbon sequestration in the soil ($0.3 \text{ Mg ha}^{-1} \text{ y}^{-1}$) (Table 7-Table 8). Adding soybeans in the continuous maize further decreased $0.18 \text{ Mg ha}^{-1} \text{ y}^{-1}$ C in the soil. Using rye as a cover crop added $0.21 \text{ Mg ha}^{-1} \text{ y}^{-1}$ of carbon to the soil in the CT situation, but this effect was smaller than the NT as a conservation method. Similar to soybeans, the small increase in SOC from rye is due to the low simulated rye biomass. Under NT, SOC in the MZ-RY was less than that in the MZ alone because of lower maize yield in the MZ-RY as we mentioned earlier. With an average yield of $5.9 \text{ Mg ha}^{-1} \text{ y}^{-1}$, switchgrass contribute $0.06 \text{ Mg ha}^{-1} \text{ y}^{-1}$ increase in SOC (Figure 4-Figure 5).

We examined the soil nitrate level after maize/soybean harvest and rye harvest in Table 9-Table 10. The amount of nitrate in the soil was 85.3 kg/ha with CT, and 34.3 kg/ha with NT under the continuous maize scenario. The low soil N in the NT on the harvest day was because applied nitrogen has leached into the environment in the growing season, due to higher soil water content. The maize/soybean rotation produced about half the amount of soil nitrate because we did not apply fertilizer when soybeans were planted. Planting rye after crop harvest could reduce the soil nitrate level to 70.2 and 27.2 kg/ha with both tillage methods, which accounted for ~20% of soil nitrate without rye. The effect of rye on reducing environmental nitrogen load has been widely observed in previous literatures (Kaspar et al., 2007, 2012; Williams et al., 2018). Annual soil nitrate levels indicated that the drought in 2012 increased soil nitrate when maize crops absorbed little of the applied nitrogen, while this high level of soil nitrate was not found for soybean fields (Figure 6).

Table 5 Simulated state-level average maize yield ($\text{Mg ha}^{-1} \text{ y}^{-1}$) in the LS areas under four rotations: continuous maize (MZ), maize soybean rotation (MZ-SB), continuous maize with rye (MZ-RY), and maize soybean rotation with rye (MZ-SB-RY) and two tillage methods: conventional (CT) and no-till (NT). Simulations were from 2009 to 2019.

State	MZ		MZ-RY		MZ-SB		MZ-SB-RY	
	CT	NT	CT	NT	CT	NT	CT	NT
AR	5.5 (± 0.1)	5.6 (± 0.1)	5.4 (± 0.2)	5.6 (± 0.2)	5.7 (± 0.1)	5.8 (± 0.1)	5.7 (± 0.2)	5.8 (± 0.2)
IL	8.7 (± 0.6)	8.8 (± 0.7)	7.8 (± 0.8)	8.8 (± 0.6)	9.1 (± 1.4)	9.2 (± 1.5)	8.3 (± 0.8)	9.2 (± 0.7)
IN	8.6 (± 0.8)	8.7 (± 0.8)	7.9 (± 0.8)	8.7 (± 0.8)	9.0 (± 1.4)	9.0 (± 1.4)	8.6 (± 0.8)	9.0 (± 0.8)
IA	8.7 (± 0.8)	8.9 (± 0.8)	7.6 (± 1.0)	8.8 (± 0.8)	9.0 (± 1.7)	9.1 (± 1.7)	8.6 (± 1.0)	9.1 (± 0.9)
KS	7.1 (± 0.9)	7.1 (± 0.8)	6.9 (± 0.6)	7.1 (± 0.8)	7.3 (± 1.2)	7.3 (± 1.2)	7.2 (± 0.7)	7.4 (± 0.9)
MI	7.6 (± 0.7)	7.6 (± 0.7)	7.4 (± 0.7)	7.6 (± 0.7)	7.8 (± 0.4)	7.8 (± 0.4)	7.7 (± 0.7)	7.8 (± 0.7)
MN	8.8 (± 0.7)	8.9 (± 0.7)	8.3 (± 0.7)	9.0 (± 0.7)	9.0 (± 1.7)	9.1 (± 1.7)	8.9 (± 0.8)	9.1 (± 0.8)
MO	7.0 (± 0.8)	7.1 (± 0.8)	6.5 (± 0.6)	7.1 (± 0.8)	7.2 (± 0.9)	7.2 (± 0.9)	7.0 (± 0.7)	7.3 (± 0.8)
NE	9.6 (± 1.2)	9.6 (± 1.2)	8.4 (± 1.2)	9.5 (± 1.2)	9.8 (± 1.8)	9.8 (± 1.8)	9.0 (± 1.2)	9.7 (± 1.2)
ND	6.7 (± 0.7)	6.8 (± 0.7)	6.7 (± 0.8)	6.9 (± 0.8)	6.6 (± 1.2)	6.7 (± 1.2)	6.5 (± 0.7)	6.7 (± 0.7)
OH	8.4 (± 0.8)	8.4 (± 0.8)	7.8 (± 0.7)	8.5 (± 0.8)	8.8 (± 1.6)	8.8 (± 1.6)	8.4 (± 0.7)	8.7 (± 0.9)
PA	7.5 (± 0.4)	7.5 (± 0.4)	7.5 (± 0.5)	7.6 (± 0.5)	7.3 (± 0.6)	7.3 (± 0.6)	7.3 (± 0.7)	7.3 (± 0.7)
SD	8.0 (± 0.9)	8.1 (± 0.9)	7.5 (± 0.8)	8.3 (± 0.8)	8.4 (± 1.5)	8.5 (± 1.5)	8.3 (± 1.0)	8.5 (± 1.0)
WI	8.0 (± 0.4)	8.0 (± 0.4)	7.6 (± 0.5)	8.0 (± 0.4)	8.1 (± 1.5)	8.2 (± 1.5)	8.0 (± 0.6)	8.2 (± 0.7)
Average	8.6 (± 1.0)	8.7 (± 1.1)	7.8 (± 1.0)	8.7 (± 1.0)	8.9 (± 1.6)	8.9 (± 1.6)	8.4 (± 1.1)	8.9 (± 1.1)

*Numbers after $\pm$ are the standard deviation values calculated from state-level statistics

Table 6 Simulated state-level average soybean yield ($\text{Mg ha}^{-1} \text{ y}^{-1}$) under MZ-SB and MZ-SB-RY in the LS areas. Simulated state-level average rye biomass ($\text{Mg ha}^{-1} \text{ y}^{-1}$) under MZ-RY and MZ-SB-RY.

State	SB				RY			
	MZ-SB		MZ-SB-RY		MZ-RY		MZ-SB-RY	
	CT	NT	CT	NT	CT	NT	CT	NT
AR	1.2 (± 0.1)	1.3 (± 0.1)	1.2 (± 0.1)	1.3 (± 0.1)	2.0 (± 1.1)	3.6 (± 0.7)	2.1 (± 0.5)	2.6 (± 0.4)
IL	2.1 (± 0.2)	2.2 (± 0.2)	2.1 (± 0.2)	2.2 (± 0.2)	1.2 (± 0.8)	1.4 (± 0.8)	1.1 (± 0.4)	1.4 (± 0.5)
IN	2.2 (± 0.2)	2.3 (± 0.2)	2.2 (± 0.2)	2.3 (± 0.2)	0.9 (± 0.7)	1.2 (± 0.7)	1.0 (± 0.3)	1.3 (± 0.5)
IA	2.4 (± 0.2)	2.4 (± 0.2)	2.3 (± 0.2)	2.4 (± 0.2)	0.4 (± 0.3)	0.4 (± 0.3)	0.4 (± 0.2)	0.5 (± 0.3)
KS	2.3 (± 0.1)	2.1 (± 0.1)	2.2 (± 0.1)	2.1 (± 0.1)	1.7 (± 0.8)	2.4 (± 0.7)	1.4 (± 0.4)	2.0 (± 0.4)
MI	2.1 (± 0.2)	2.2 (± 0.2)	2.1 (± 0.2)	2.2 (± 0.2)	0.4 (± 0.1)	0.5 (± 0.2)	0.5 (± 0.2)	0.5 (± 0.2)
MN	2.5 (± 0.2)	2.5 (± 0.2)	2.5 (± 0.2)	2.5 (± 0.2)	0.1 (± 0.1)	0.1 (± 0.1)	0.1 (± 0.1)	0.1 (± 0.1)
MO	1.9 (± 0.2)	1.9 (± 0.2)	1.9 (± 0.2)	1.9 (± 0.2)	1.7 (± 0.9)	2.1 (± 0.8)	1.3 (± 0.4)	1.8 (± 0.5)
NE	2.5 (± 0.2)	2.4 (± 0.2)	2.5 (± 0.2)	2.4 (± 0.2)	0.7 (± 0.3)	1.0 (± 0.4)	0.8 (± 0.2)	1.1 (± 0.3)
ND	2.2 (± 0.2)	2.3 (± 0.2)	2.2 (± 0.2)	2.3 (± 0.2)	0.01 (± 0.01)	0.01 (± 0.01)	0.01 (± 0.01)	0.01 (± 0.02)
OH	2.2 (± 0.2)	2.3 (± 0.2)	2.2 (± 0.2)	2.3 (± 0.2)	0.9 (± 0.5)	1.2 (± 0.5)	1.0 (± 0.3)	1.2 (± 0.3)
PA	2.5 (± 0.1)	2.6 (± 0.1)	2.5 (± 0.1)	2.6 (± 0.1)	0.9 (± 0.4)	0.9 (± 0.4)	0.7 (± 0.3)	0.8 (± 0.3)
SD	2.1 (± 0.3)	2.2 (± 0.3)	2.1 (± 0.3)	2.2 (± 0.3)	0.2 (± 0.1)	0.2 (± 0.1)	0.2 (± 0.1)	0.3 (± 0.1)
WI	2.5 (± 0.2)	2.5 (± 0.2)	2.4 (± 0.2)	2.5 (± 0.2)	0.2 (± 0.1)	0.2 (± 0.1)	0.2 (± 0.1)	0.2 (± 0.1)
Average	2.3 (± 0.3)	2.3 (± 0.3)	2.3 (± 0.3)	2.3 (± 0.3)	0.7 (± 0.7)	0.9 (± 0.8)	0.7 (± 0.5)	0.9 (± 0.7)

*Numbers after $\pm$ are the standard deviation values calculated from state-level statistics

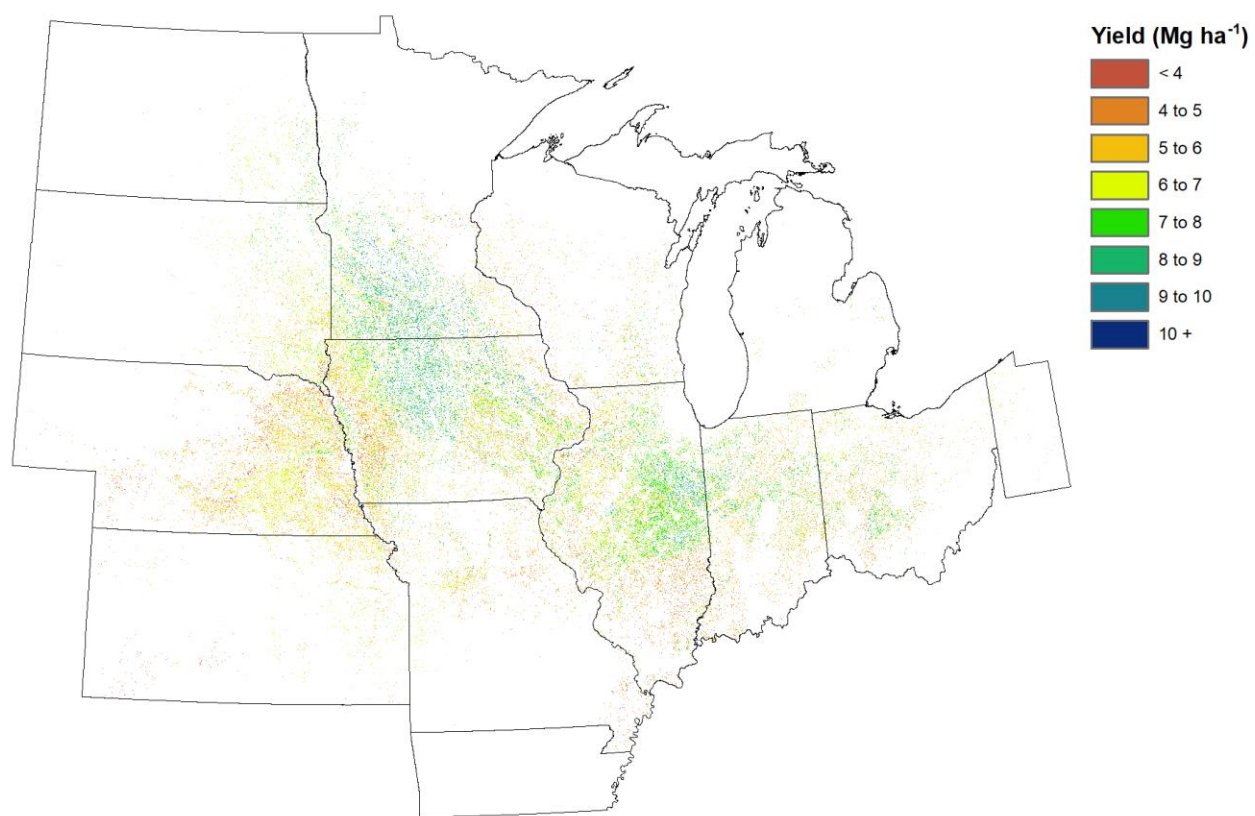


Figure 4 Mean yield of switchgrass over 11 years in the LS areas.

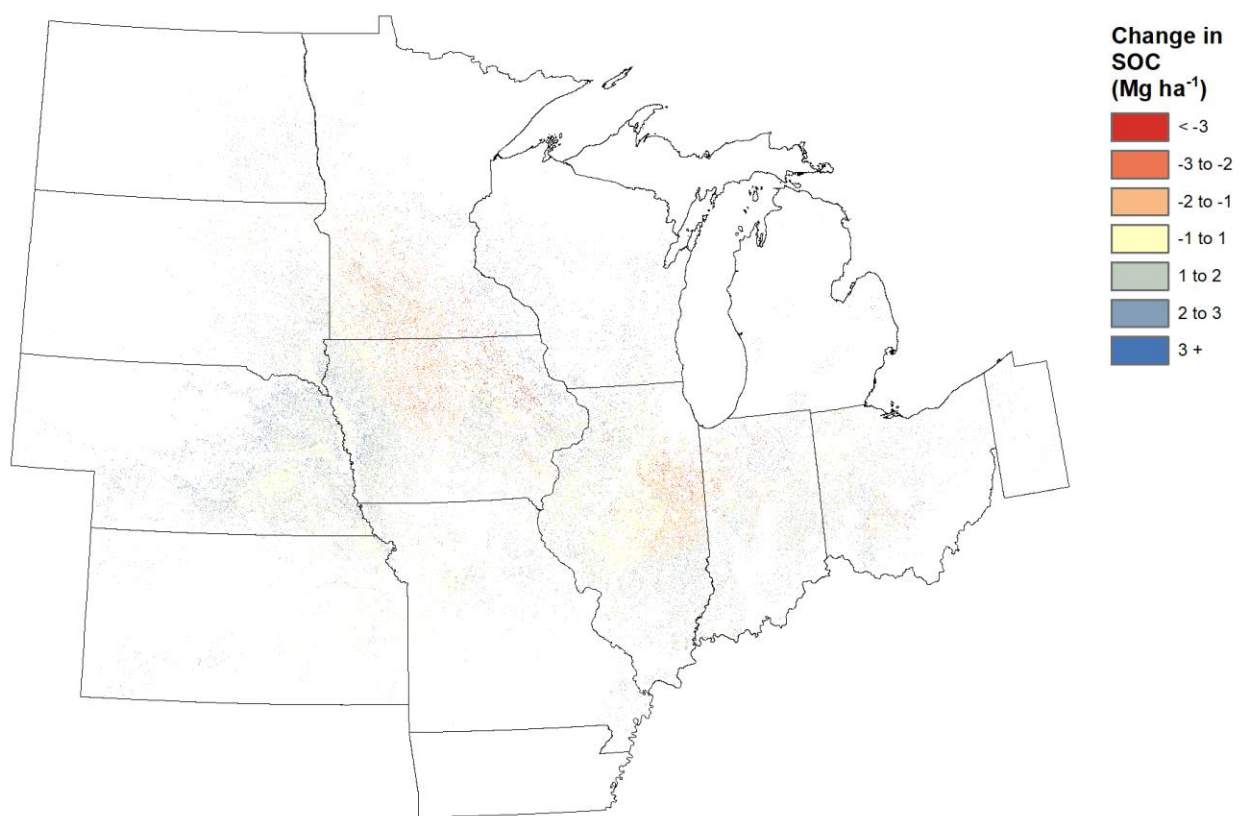


Figure 5 SOC change over 11 years of switchgrass in the LS areas.

Table 7 Simulated state-level average soil organic carbon change ($\text{Mg ha}^{-1} \text{ y}^{-1}$) under conventional tillage.

State	MZ	MZ-RY	MZ-SB	MZ-SB-RY
AR	0.01 (± 0.23)	0.01 (± 0.14)	0.01 (± 0.24)	0.09 (± 0.13)
IL	-0.4 (± 0.61)	-0.05 (± 0.27)	-0.58 (± 0.66)	0.11 (± 0.25)
IN	-0.06 (± 0.57)	0.08 (± 0.26)	-0.15 (± 0.59)	0.27 (± 0.24)
IA	-0.19 (± 0.75)	-0.01 (± 0.34)	-0.46 (± 0.72)	0.1 (± 0.40)
KS	-0.13 (± 0.44)	0.1 (± 0.21)	-0.21 (± 0.39)	0.25 (± 0.20)
MI	-0.16 (± 0.65)	0.15 (± 0.32)	-0.23 (± 0.63)	0.21 (± 0.32)
MN	-0.93 (± 0.97)	-0.23 (± 0.46)	-1.03 (± 0.92)	-0.26 (± 0.54)
MO	-0.11 (± 0.54)	0.05 (± 0.23)	-0.22 (± 0.53)	0.23 (± 0.19)
NE	0.41 (± 0.39)	0.19 (± 0.21)	0.17 (± 0.42)	0.4 (± 0.21)
ND	-0.76 (± 0.50)	-0.22 (± 0.30)	-0.83 (± 0.47)	-0.29 (± 0.30)
OH	-0.06 (± 0.59)	0.1 (± 0.28)	-0.16 (± 0.61)	0.32 (± 0.26)
PA	0.05 (± 0.43)	0.18 (± 0.25)	0.01 (± 0.40)	0.41 (± 0.20)
SD	-0.14 (± 0.73)	-0.02 (± 0.34)	-0.39 (± 0.62)	0.11 (± 0.38)
WI	0.21 (± 0.53)	0.23 (± 0.32)	0.05 (± 0.55)	0.34 (± 0.32)
Average	-0.19 (± 0.71)	0.02 (± 0.32)	-0.37 (± 0.70)	0.15 (± 0.36)

*Numbers after $\pm$ are the standard deviation values calculated from state-level statistics

Table 8 Simulated state-level average soil organic carbon change (Mg ha⁻¹ y⁻¹) under no-till.

State	MZ	MZ-RY	MZ-SB	MZ-SB-RY	Switchgrass
AR	0.19 (±0.12)	0.12 (±0.13)	0.2 (±0.13)	0.17 (±0.14)	0.11 (±0.06)
IL	0.21 (±0.23)	0.1 (±0.22)	0.16 (±0.25)	0.14 (±0.23)	-0.03 (±0.12)
IN	0.32 (±0.22)	0.19 (±0.20)	0.29 (±0.23)	0.25 (±0.21)	0.03 (±0.12)
IA	0.3 (±0.24)	0.21 (±0.22)	0.24 (±0.25)	0.26 (±0.24)	0.18 (±0.13)
KS	0.29 (±0.19)	0.18 (±0.18)	0.25 (±0.18)	0.27 (±0.20)	0.04 (±0.08)
MI	0.35 (±0.20)	0.27 (±0.20)	0.29 (±0.20)	0.31 (±0.20)	0.09 (±0.11)
MN	0.14 (±0.28)	0.11 (±0.27)	0.11 (±0.30)	0.13 (±0.29)	-0.07 (±0.17)
MO	0.32 (±0.19)	0.2 (±0.18)	0.27 (±0.20)	0.26 (±0.21)	0.05 (±0.10)
NE	0.44 (±0.16)	0.33 (±0.16)	0.43 (±0.16)	0.4 (±0.17)	0.1 (±0.08)
ND	0.2 (±0.15)	0.21 (±0.16)	0.12 (±0.15)	0.14 (±0.15)	0.08 (±0.07)
OH	0.34 (±0.22)	0.21 (±0.21)	0.31 (±0.23)	0.27 (±0.23)	0.03 (±0.12)
PA	0.45 (±0.13)	0.36 (±0.14)	0.41 (±0.14)	0.4 (±0.17)	0.11 (±0.07)
SD	0.31 (±0.19)	0.24 (±0.19)	0.26 (±0.19)	0.28 (±0.19)	0.08 (±0.09)
WI	0.42 (±0.20)	0.38 (±0.20)	0.41 (±0.20)	0.43 (±0.20)	0.12 (±0.10)
Average	0.3 (±0.23)	0.2 (±0.22)	0.25 (±0.24)	0.25 (±0.23)	0.06 (±0.13)

*Numbers after ± are the standard deviation values calculated from state-level statistics

Table 9 Simulated soil nitrate level (kg ha⁻¹ y⁻¹) under conventional tillage.

State	N rate (kg/ha)	MZ	MZ-RY	MZ-SB	MZ-SB-RY
AR	224	101.6 (±142.3)	22.2 (±73.8)	88.4 (±99.9)	57.7 (±67.9)
IL	204	51.2 (±85.5)	26.5 (±50.8)	39.2 (±63.4)	16.2 (±30.5)
IN	209	42.6 (±73.4)	17.2 (±39.5)	35.2 (±60.4)	14 (±25.5)
IA	180	101.0 (±115)	89.7 (±97.9)	42.1 (±51)	45.2 (±43.8)
KS	180	85.1 (±118.4)	65.8 (±108.4)	47.4 (±81)	25.6 (±53)
MI	171	42.8 (±71.7)	30.3 (±52.1)	29.5 (±40.6)	23.5 (±22.5)
MN	183	138.0 (±127.2)	142.1 (±120.0)	56.3 (±56.8)	76.3 (±52.6)
MO	224	75.7 (±92.6)	45.2 (±66.1)	52 (±78.3)	21.6 (±43.9)
NE	224	121.1 (±147.4)	108.6 (±131.3)	45.8 (±77.9)	34 (±57)
ND	164	144.2 (±133.1)	155.1 (±118.3)	63.6 (±50.6)	87.6 (±43.4)
OH	195	45.1 (±77.1)	17.9 (±38.3)	31.5 (±59.3)	12.5 (±27.5)
PA	195	102.5 (±67.9)	46.5 (±32.1)	46.4 (±60.7)	18.8 (±24)
SD	158	122.9 (±127.5)	131.1 (±123.7)	54.7 (±60.6)	61 (±51.6)
WI	155	38.3 (±61.9)	33.9 (±45.9)	22.3 (±35)	26.3 (±23.3)
Average		85.3 (±108.1)	70.2 (±93.8)	42.5 (±61.0)	34.5 (±44.7)

*Numbers after ± are the standard deviation values calculated from state-level statistics

Table 10 Simulated soil nitrate level (kg ha⁻¹ y⁻¹) under no-till.

State	N rate (kg/ha)	MZ	MZ-RY	MZ-SB	MZ-SB-RY
AR	224	46.7 (±16.2)	2.6 (±11.1)	81.3 (±69.4)	22.2 (±37)
IL	204	34.8 (±32.5)	18.2 (±19.7)	32.7 (±45.6)	14.6 (±18.2)
IN	209	25.6 (±31.2)	12.2 (±15.8)	31.9 (±49.2)	13.2 (±18.5)
IA	180	36.8 (±35.5)	35.1 (±33.2)	26.5 (±35.9)	28 (±22.7)
KS	180	45.2 (±37.8)	27.5 (±34.5)	39 (±56.5)	16.1 (±28.2)
MI	171	23.6 (±34.9)	21.9 (±26.2)	20.7 (±29.8)	19.1 (±12.2)
MN	183	39.5 (±38.3)	44.9 (±37.0)	26.5 (±23.7)	45 (±24.7)
MO	224	51.6 (±37.9)	25.1 (±25.9)	39.6 (±52.6)	14.4 (±23.3)
NE	224	27.9 (±37.8)	24.7 (±38.5)	28.9 (±47.8)	18.9 (±26.2)
ND	164	35.3 (±37.3)	46.3 (±29.9)	28 (±20)	50.8 (±23.1)
OH	195	28.3 (±37)	12.5 (±18.4)	25.4 (±47.2)	10.6 (±18.2)
PA	195	71.6 (±40)	32.9 (±20.3)	31.8 (±49)	15.2 (±17.2)
SD	158	44.0 (±44.5)	52.4 (±46.3)	23.7 (±30.5)	32.4 (±20.6)
WI	155	22.0 (±25.5)	22.4 (±21.2)	13.5 (±23.6)	20 (±11.9)
Average		34.3 (±36.0)	27.2 (±31.1)	28.8 (±41.3)	27.2 (±23.5)

*Numbers after ± are the standard deviation values calculated from state-level statistics

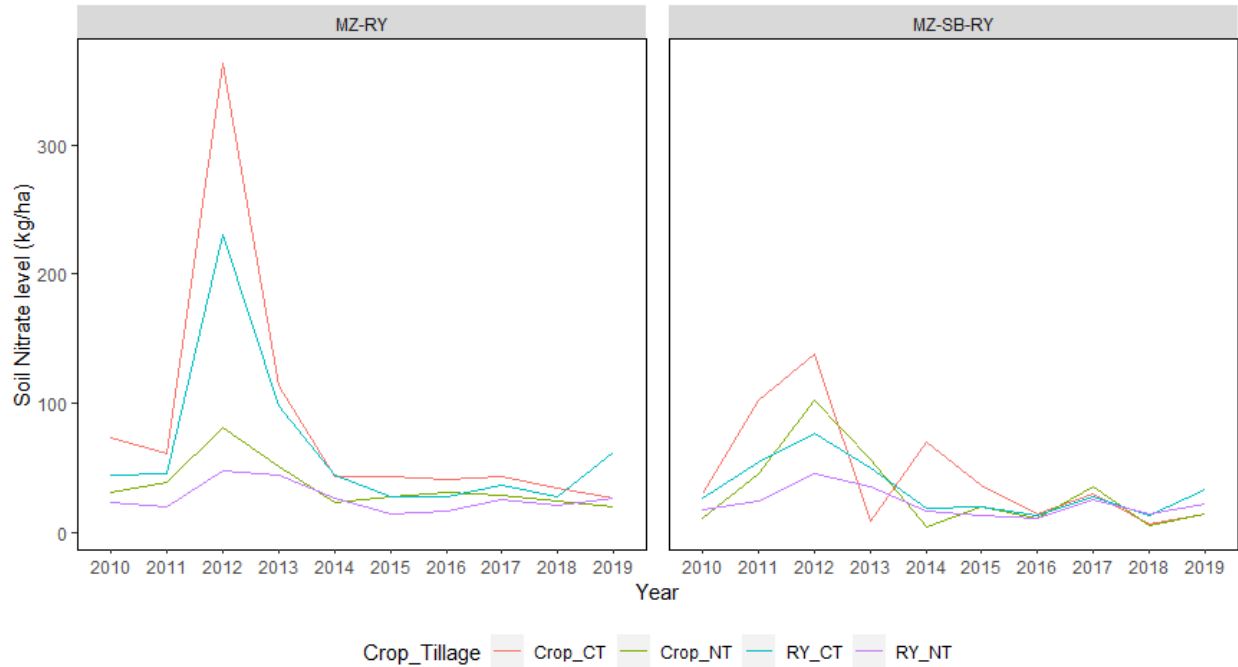


Figure 6 Annual soil nitrate levels after crop (MZ or MZ/SB) or rye harvest in the MZ-RY (left) and MZ-SB-RY (right) scenarios.

3.5.3 Greenhouse warming intensity under switchgrass and maize production systems

We further estimated emitted GWI in terms of CO₂ equivalence from crop production in MZ, MZ-SB, and SW (Table 11). The application of fertilizer was a large component of the GHG emissions in the agricultural activities (5.1 Mg CO₂eq ha⁻¹ y⁻¹). Projected across the LS area, state-mean GWI values were expected to range from -0.2 to 4.9 Mg CO₂eq ha⁻¹ y⁻¹, averaging at 2.4 Mg CO₂eq ha⁻¹ y⁻¹. With enhanced SOC content in the NT system, the mean GWI substantially decreased to 0.4 Mg CO₂eq ha⁻¹ y⁻¹. Although there was no GHG emission from fertilizer synthesis and application for SB, the lack of biofuel production resulted in GWI values of 5.9 and 3.2 Mg CO₂eq ha⁻¹ y⁻¹ for CT and NT respectively. With no fertilizer input, GWI was -3.6 Mg CO₂eq ha⁻¹ y⁻¹ in the switchgrass system.

Table 11 State-level average GWI (Mg CO₂eq ha⁻¹ y⁻¹) for MZ, MZ-SB and SW in the LS area.

State	N rate for MZ (kg/ha)	MZ		MZ-SB		SW
		CT	NT	CT	NT	
AR	224	4.7 (±0.9)	3.0 (±0.2)	6.1 (±0.9)	5.4 (±0.4)	-3.0 (±0)
IL	204	3.2 (±2.2)	1.0 (±1)	6.2 (±2.6)	3.4 (±1.2)	-3.5 (±0.2)
IN	209	2.2 (±2.1)	0.8 (±0.9)	4.9 (±2.4)	3.3 (±1.2)	-3.5 (±0.2)
IA	180	1.6 (±2.8)	-0.2 (±1)	5.2 (±2.9)	2.6 (±1.4)	-3.7 (±0.2)
KS	180	2.3 (±1.9)	0.9 (±1)	4.6 (±1.8)	2.9 (±1.1)	-3.4 (±0.2)
MI	171	1.8 (±2.4)	-0.1 (±0.9)	3.9 (±2.3)	2.0 (±0.9)	-3.6 (±0.1)
MN	183	4.2 (±3.6)	0.3 (±1.1)	7.2 (±3.5)	2.9 (±1.4)	-3.8 (±0.2)
MO	224	3.9 (±1.9)	2.1 (±0.7)	6.4 (±1.9)	4.4 (±0.9)	-3.5 (±0.2)
NE	224	0.2 (±1.4)	0.2 (±1)	3.9 (±2.1)	2.9 (±1.2)	-3.6 (±0.2)
ND	164	4.4 (±1.8)	0.8 (±0.7)	6.6 (±1.9)	3.1 (±1)	-3.9 (±0.1)
OH	195	1.8 (±2.2)	0.3 (±1)	4.7 (±2.5)	2.8 (±1.3)	-3.6 (±0.1)
PA	195	1.9 (±1.7)	0.4 (±0.6)	3.9 (±1.6)	2.4 (±0.7)	-3.7 (±0.1)
SD	158	1.3 (±2.9)	-0.6 (±0.9)	4.2 (±2.5)	1.7 (±1.3)	-3.9 (±0.2)
WI	155	-0.3 (±2)	-1.1 (±0.7)	3.2 (±2.4)	1.6 (±1.3)	-3.7 (±0.2)
Average		2.1 (±2.7)	0.4 (±1.2)	5.3 (±2.8)	2.9 (±1.4)	-3.6 (±0.2)

*Numbers after ± are the standard deviation values calculated from state-level statistics

3.5.4 Ecological benefits from changing maize to switchgrass

Based on the latest NLCD image, 86.7, 47.8, and 133.8 million ha of cropland, forest, and grassland were in the fourteen maize/soybean production states (Figure 7, Table 12). However, considering the definition of landscape (radius of 1.5 km) in the study, the proportion of grassland around maize and soybean fields was only 16% under the current scenario, with the highest proportion (36.1%) observed in Kansas and the lowest in Arkansas (2.3%). Grassland proportion in the center states (Illinois and Indiana) was 13%, even lower than the average value. Planting switchgrass increased the average proportion of grassland to 20%, and Illinois, Iowa, and North Dakota had an increased rate above 6%. This higher percentage of grassland in the landscape led to a positive effect on the biocontrol index (an increase of 0.025 on average).

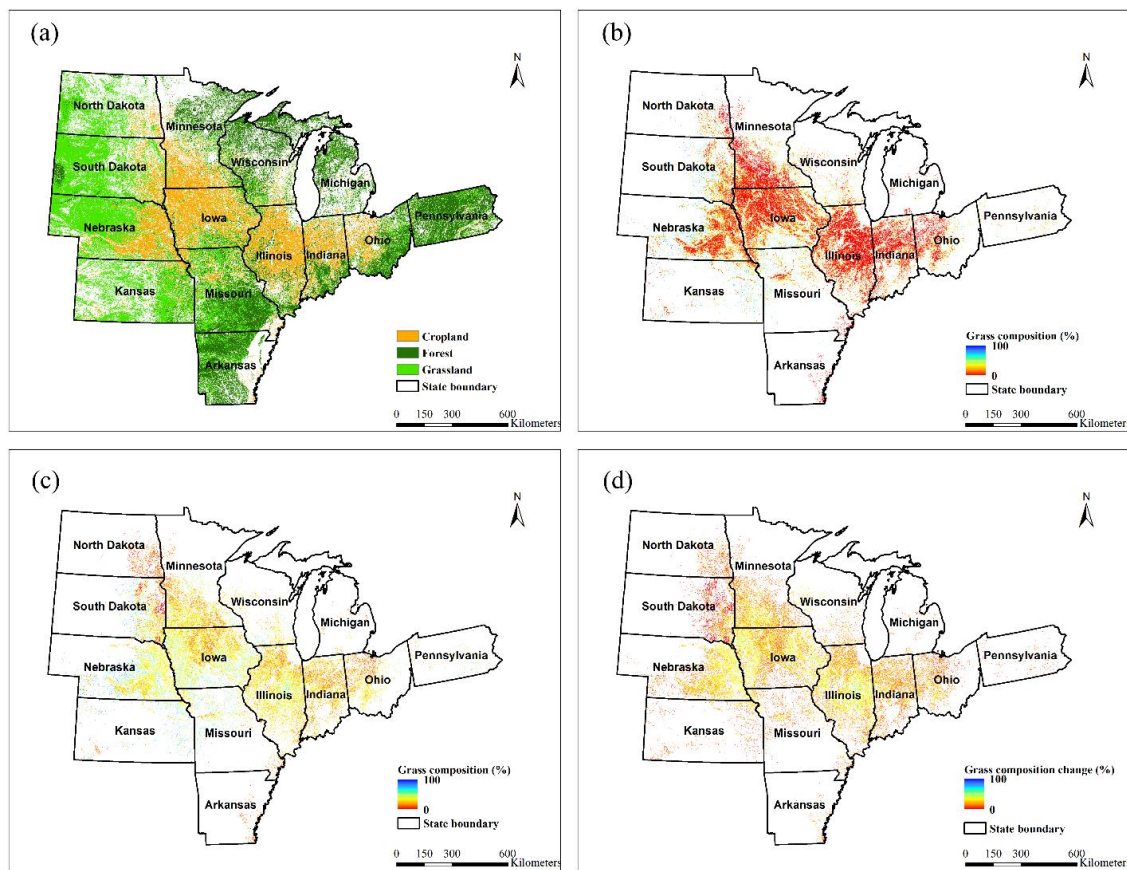


Figure 7 Reclassified NLCD dataset (a), proportions of grassland under maize (b) and switchgrass scenarios (c), and associated grassland changes (d).

Table 12 State-level NLCD statistical information and grass compositions and changes under maize and switchgrass scenarios.

State	Crop (ha)	Forest (ha)	Grass (ha)	Proportion of grassland (%)			BCI change
				Maize	Switchgrass	change	
AR	139474	5566078	2227120	2.3	4.9	2.6	0.02
IL	6503184	2166356	1070000	5.4	12	6.6	0.04
IN	3140567	2141514	775962	6.7	10.4	3.7	0.02
IA	7492762	1100346	1983989	10.1	16.9	6.8	0.04
KS	869252	948759	9257258	36.3	39.5	3.2	0.02
MI	2053029	2947478	735689	9.6	12.9	3.3	0.02
MN	3886614	3937433	1664092	8	12.5	4.5	0.03
MO	1414235	6644014	5464797	25.3	28.9	3.6	0.02
NE	4470153	372658	10558412	17.6	18.8	1.2	0.01
ND	703671	302621	6327033	18.8	25	6.2	0.04
OH	1820514	3417143	1482176	10.3	14	3.7	0.02
PA	197357	7069522	1628224	28.8	30.1	1.3	0.01
SD	2082771	628303	11073021	24.7	27.5	2.8	0.02
WI	867306	4784137	1337768	20.1	25.1	5	0.03

3.6 Discussion

The 4 per 1000 initial proposed the increase of carbon stock in the soil to mitigate climate change. We proposed to improve soil organic carbon content by shifting maize or soybean production into switchgrass in the LS area defined from yield stability analysis. About 40 percent of LS were located in the field edge area, consistent with the phenomenon that small and sparse plants are always found in this area. Yield reduction in field edge could result from poor soil due to pressure from heavy mechanical machinery, which is not able to support crop growth. Crop yield loss in the field edge could be above 30 percent if next to a mature woodland, which competes with crops for sunlight and nutrients (Pierce & Milhollin, 2020). Microclimate or herbicide shift

could also cause this edge effect on maize yields. Along with the edge effect, around 90% of LS are contiguous in a patch with at least 3 pixels ($3 \times 30 \text{ m} \times 30 \text{ m}$, 270 m^2), and over 80% were in patches of at least 9 pixels ($9 \times 30 \text{ m} \times 30 \text{ m}$, 810 m^2). This spatially continuous property of LS, together with distribution mainly on the field edge, creates an easily operated environment for farmers to plant switchgrass since isolated LS areas would be difficult to manage.

Our simulation indicated that NT and switchgrass are two effective ways to increase SOC in soils. Compared to the continuous maize system with conventional tillage, NT and switchgrass were estimated to increase 0.3 and $0.06 \text{ Mg ha}^{-1} \text{ y}^{-1}$ SOC (Table 7-Table 8). The change in the SOC from switchgrass was a result of lower biomass (5.9 Mg/ha) compared to maize (8.6 Mg/ha). This neutral effect on SOC was reasonable in the LS, which is SOC richer than marginal land in previous studies. Putting additional nitrogen could increase the switchgrass yield and related soil carbon sequestration (Martinez-Feria et al., 2022). Soybeans has negative effects on the SOC sequestration because of not only little biomass or residue for SOC buildup but also nitrogen-rich soybean residue in the soil often leads to vigorous growth of decomposer bacteria and fungi microbes (Hall & Russell, 2019). Cereal rye also added carbon in the soil, which was positively related to rye yield. Thus, rye can be used for SOC sequestration in southern states of this study area, such as Arkansas and Nebraska. In general, NT is still the best way to sequester carbon in this study, and long-term use of switchgrass could achieve similar effects. As a cover crop, rye could reduce the soil nitrate level by 20%, which contributes to better water quality in the environment (Table 9-Table 10). These results were consistent with previous studies (Kaspar et al., 2007, 2012; Krueger et al., 2011; Williams et al., 2018). Meanwhile, farmers need more education and money to upscale the cover crop adoption. Thus, government support is necessary

to promote switchgrass and rye production in order to increase the sustainability of the agricultural system.

In face of a changing climate, putting crop land into biofuel production can offset a large proportion of CO₂eq emitted from fertilizer synthesis, application, and associated nitrogen emissions (Table 11) (Martinez-Feria et al., 2022). Additionally, crop rotation can protect soil structure, sometimes improved crop yield, and lower concentrations of specific pests and disease (Berzsenyi et al., 2000; Halvorson et al., 2002; Karlen et al., 2013; Peters et al., 2003; Rusch et al., 2013). Production of soybeans, that cannot be used to produce biofuel, resulted in an increased GHG emissions, although nitrogen application was not required by these crops. It is important to balance biofuel crop production and other recommended management practices. Although we did not observe large SOC increase, planting switchgrass for biofuel production has a positive effect (-3.6 Mg CO₂eq ha⁻¹ y⁻¹) on the reduction of GHG in the atmosphere (Table 11).

Monocultures in the current cropping system greatly reduce biodiversity (Price, 2008). The average proportion of grassland providing biocontrol functions was only about 16% for maize and soybean fields (Table 12). The situation was even worse in some center states of the Corn Belt, such as Illinois and Indiana. The intercropping system is a common practice to enrich biodiversity, which also works for switchgrass and maize/soybeans (Afrin et al., 2017; Maitra & Ray, 2019; Snapp et al., 2010). The grassland proportion increased to 20% under this intercropping system and positively affected the biocontrol index. The negative relationship between biocontrol index and insecticide use found in Meehan et al. (2012) and other studies suggest switchgrass could also reduce insect damage and insecticide use. Biological control is effective because the abundance of natural enemies of crop pests is higher in a more diverse landscape with perennial-dominated plants (Bianchi et al., 2006; Risch et al., 1983).

3.7 Conclusion

Subfield areas that have been consistently underperforming over the years offer great opportunities to shift from current crops to biofuel crops. This paper examined several benefits of this shift, including improvements in the soil organic carbon, greenhouse gas emissions, and ecological diversity in the landscape. Results indicated that no-till was the best strategy in the LS to improve SOC and switchgrass could reduce the GHG in the atmosphere compared to the maize or maize/soybean system. Adding rye as a cover crop in the rotation could lower the amount of nitrate in the soil to further protect the environment. These results highlight the potential of using yield stability maps to support the long-term sustainability of agricultural systems.

CHAPTER 4: SUBFIELD MAIZE YIELD PREDICTION IMPROVES WHEN IN-SEASON CROP WATER DEFICIT IS INCLUDED IN REMOTE SENSING IMAGERY-BASED MODELS

A version of this chapter appeared in the journal *Remote Sensing of Environment* (doi: 10.1016/j.rse.2022.112938)

4.1 Abstract

In-season prediction of crop yield is a topic of research studied by several scientists using different methods. Seasonal forecasts provide critical insights to different stakeholders who use the information for strategic and tactical decisions. In this study, we propose a novel scalable method to forecast in season subfield crop yield through a machine learning model based on remotely sensed imagery and data from a process-based crop model on a cumulative crop drought index (CDI) designed to capture the impact of the in-season crop water deficit on yields. To evaluate the performance of our proposed model, we used 352 growers' fields of different sizes across the states of Michigan, Indiana, Iowa, and Illinois, with 2,520 respective yield maps generated by combine harvesters equipped with precise high-resolution yield monitor sensors over multiple years (2006 - 2019). We obtained high resolution digital elevation model, climate, and soil data to execute the SALUS model and calculate the CDI for each field used in the study. We used Landsat Analysis Ready Dataset (ARD) products generated by USGS as the image source to calculate the green chlorophyll vegetation index (GCVI). We found that the inclusion of the CDI in remote sensing-based random forest models substantially improved in-season subfield maize yield prediction. The addition of the CDI in the yield prediction model showed that the greatest improvements in predictions were observed in the driest year (2012) in our case study. The proposed approach also showed that the subfield spatial variations of maize yield are better

captured with the inclusion of CDI for most fields. The earliest prediction in the growing season with GCVI and CDI together outperformed the latest prediction with GCVI alone, highlighting the potential of CDI for predicting spatial variability of maize yields around grain filling period, which is on average close to two months before typical crop harvest in the US Midwest.

4.2 Introduction

Crop yield is a result of the interactions and dynamic feedbacks between Genotype, Environment, and Management (G x E x M). These interactions are highly variable within a field, between fields and across years (Maestrini & Basso, 2021; Messina et al., 2009). Accurate in-season crop yield prediction is valuable for farmers, governments, and commodity traders to make strategic decisions (Basso & Liu, 2019). On a small scale, seasonal subfield-scale yield prediction allows farmers to make financial decisions and analyze the effects of management practices on crop yield, resource use efficiency, and environmental outcomes (Hunt et al., 2019). Public crop yield forecasts in the U.S. provided by the United States Department of Agriculture (USDA) are available only at the county level, thus offer limited value for farmers in identifying yield-limiting factors within their own fields (Xiao et al., 2017). Hunt et al. (2019) pointed out that only a few studies have focused on within-field yield estimation, mainly due to the limited availability of high-resolution yield data. With the advancement of precision agriculture technologies, producer yield maps offer opportunities to validate the performance of subfield yield prediction models (Basso & Liu, 2019; Maestrini & Basso, 2018b; Martinez-Feria & Basso, 2020) and to develop a robust crop yield prediction system at subfield scale to help farmers increase production and assess the agricultural sustainability of management practices on a long-term basis.

Remote sensing (RS) observations with repeated visits and large-scale coverage are used for agricultural monitoring, with an example being the NASA Landsat archive. The NASA Landsat

satellites have been acquiring images for about five decades. These images are freely accessible for the scientific community, at a 30m resolution and 16-day cycle (Wulder et al., 2016). Spectral information from Landsat data can be used to calculate vegetation indices (VIs) using various band combinations, such as normalized difference vegetation index (NDVI) or enhanced vegetation index (EVI) (P.-Y. Chen et al., 2006). VIs derived from near-infrared and red-edge bands are relevant to detect vegetation conditions that are positively correlated with final crop yield, such as leaf area index (LAI) and crop biomass (Dong et al., 2019; Waldner et al., 2019). Based on a comprehensive review from Basso & Liu (2019), seasonal NDVI images are the most commonly used data for building linear regression models of crop yield. NDVI is linearly correlated with leaf area index (LAI) till LAI reaches the value of 3, after which the relationship flattens out due to light saturation with no changes in NDVI value corresponding to increasing LAI due to leaf extensions and plant biomass accumulation. Hunt et al. (2019) assessed the capacity of Sentinel-2 data to estimate within-field wheat yield variability. The data provided accurate estimates of within-field yield variability with an RMSE of 0.66 t/ha. Given the current diversity of satellite data, spectral bands other than optical bands, such as thermal and microwave bands, can now be integrated into models to improve yield estimation. Guan et al. (2017) indicated that Ku-band backscatter, thermal-based ALEXI evapotranspiration (ET), and X-band vegetation optical depth (VOD) provide extra information on crop stresses beyond EVI. Basso & Liu (2019) reported that additional agrometeorological information, including growing degree days, soil characteristics and landscape parameters are also often combined with remote sensing data to predict yield (Balaghi et al., 2008; Betbeder et al., 2016; Saeed et al., 2017). One main drawback of these approaches is that the statistical models are locally parameterized and may generate poor results when conditions in the forecasting year or region are not represented in the trained model (Kang & Özdoğan, 2019).

On the other hand, process-based crop simulation models (CSMs) can predict crop yield at any time or location using localized soil, climate and management practices. Crop state variables such as phenological development, shoot and root biomass accumulation, soil water and nitrogen dynamics are updated at daily steps. CSMs have the capacity to capture the feedback and synergies occurring at the soil-plant-atmosphere continuum. This allows for a more systematic understanding of the processes leading to yield formation and environmental outcomes to the point of using the simulated results to improve management strategies before the season (strategic approach) or in-season adaptation (tactical approach) (Jones et al., 2017).

Integration of CSM and RS data using data assimilation approaches have been applied to predict yield at regional and national scales since the 80's (X. Jin et al., 2018; Maas, 1988). De Wit & Van Diepen (2007) proposed the use of an ensemble Kalman filter for yield prediction in Spain, France, Italy and Germany from 1992 to 2000. Their results fit crop yield data well, with R^2 values of 0.66 for winter wheat and 0.56 for maize. Huang et al. (2015) assimilated RS LAI data into the WOFOST model using a similar framework, which contributed to improved accuracy of predicted winter wheat yield from R^2 of 0.23 to 0.48. The data assimilation approach assumes that the LAI generated by remote sensing is more accurate than the one generated by the crop model. If that was the case, forcing the adjustment of LAI without changing parameters in the model to reproduce the new LAI would not guarantee that the model would continue to deviate from the real values observed in the field. Despite successful applications of data assimilation, poor results are often found when there is a lack of correlation between LAI and yield or when there are significant errors in phenology simulations (Curnel et al., 2011; Nearing et al., 2012). Lobell et al. (2015) regressed a series of simulated green chlorophyll vegetation indices (GCVI) with crop yield under various management settings of crop models, which were then applied to the

GCVI from actual RS data for yield prediction. Zhao et al. (2020) directly added the crop water stress index (SI) calculated from a biophysical model as an additional feature to VI in wheat yield prediction. This method achieved high accuracy when multiple years with diverse climate conditions were included in the analysis. The calculated SI was based only on climate and crop data, so fields sharing the same crop and climate information were projected to suffer the same level of stress. This is unlikely in real, heterogeneous fields. Crop water stress levels are known to differ between and within fields based on interactions of landscape position and weather (Martinez-Feria & Basso, 2020). RS has often been used in monitoring water stress in crop fields, however, some effects of drought on crops may not be reflected by spectral information. For example, while drought during the grain filling period would result in fewer kernels and low kernel weight, these differences may not be pronounced enough to cause detectable spectral variations (Thelen, 2007). Also, remote sensing may not be able to detect drought occurred at the mid-season because of signal saturation when crop canopy closes, especially in maize plants. Since crop yields are more vulnerable to cumulative drought effects, another disadvantage of using remote sensing alone to detect drought is the lack of high-frequency continuous observations due to cloud coverage, especially in the U.S. Great Lake region (Kamara et al., 2003). Compared to RS, crop models that simulate the daily crop water balance can more accurately evaluate the impact of cumulative crop drought stress at any growth stage.

In this study, we posed the following research question: Can the integration of water stress from a crop model with remote sensing images improve maize yield prediction? To address this question, we coupled historical crop yield stability maps (Basso et al., 2019; Maestrini & Basso, 2018a, 2018b) with the cumulative drought index (CDI) calculated from the SALUS crop model. The CDI is defined as the actual crop water use over the potential evapotranspiration. This index

varies within a field based on the different interactions occurring between soil, weather, genotype management and landscape characteristics.

The objective of this research was to improve subfield yield predictions by developing an hybrid model based on the integration of modelled crop water stress and remotely sensed vegetation indices. Here we combined VIs derived from Landsat images (RS) with simulated CDI to predict field-level maize yield for 352 fields across the US Midwest. Results were validated using crop yield data from combine harvesters equipped with yield-monitoring sensors.

4.3 Study area and dataset

The study was conducted on 352 fields, of which 186 fields were located in Michigan, 76, fields in Indiana, 64 in Iowa, and 26 fields in Illinois. Crops in these fields are growing under a continental climate with warm and humid summers. The average annual precipitation is 787 mm, with 60% of rain falling during the growing season. The average length of the growing season for summer crops in this region is 130 to 150 days. The size of selected fields ranged from 1.3 ha to 155 ha. Maize was the most frequently planted crop in these fields, with soybeans and wheat also in rotation. Here, we only focused on maize. After examining USDA crop progress reports. ([https://www.nass.usda.gov/Charts and Maps/Crop Progress & Condition/index.php](https://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/index.php)), the growing seasons of 2012 (poor year), 2016 (average year), and 2018 (good year) were selected for Michigan to include different climatic conditions (USDA, 2016). Similarly, we used 2014 (good), 2015 (poor), and 2018 (average) for Indiana and 2012 (poor), 2016 (good) and 2017 (average) for Iowa. For Illinois, only 2013 (poor) and 2016 (good) growing seasons were used as no average year was found in the report. Of 352 fields, 130 to 150 fields were planted with maize in 2012, 2016, and 2018, and 20 to 45 fields in other years.

Yield monitor dataset. For these fields, we obtained a total of 2,520 high quality yield maps collected directly by farmers from multiple years up to 2019, from combine harvesters equipped with a yield monitor sensor. The raw yield- monitor data is a point shapefile at a 2-m interval. We processed the raw yield data into yield maps based on the steps from Maestrini & Basso (2018b). In brief, the boundary of each field was determined by the envelope of all yield points in ArcGIS. From each crop dataset, the median yield was used to define a lower ($0.1 \times \text{median}$) and higher ($3 \times \text{median}$) bounds, which were applied to remove yields outside this range. Then a spherical kriging model with a fixed radius of 20 m and a minimum requirement of 12 points was applied to convert yield shapefile into raster file at the same resolution. This raster file was smoothed using mean value within 10×10 pixels and then clipped within the boundary of the field to get the final yield map.

Digital Elevation Model (DEM). We obtained the 1/3 and 1/9 arc-second (approximately 10 and 3 meters, respectively) DEM datasets from the USGS National Elevation Dataset (Arundel et al., 2015).

Historical daily weather dataset. We used gridded weather data from gridMET which is a dataset of daily high-spatial resolution ($\sim 4\text{-km}$, $1/24^{\text{th}}$ degree) surface meteorological data covering the contiguous US from 1979-yesterday (Abatzoglou, 2013). gridMET blends spatial attributes of gridded climate data from PRISM with desirable temporal attributes (and additional variables) from regional reanalysis (NLDAS-2) using climatically aided interpolation. The resulting product is a spatially and temporally complete, high-resolution ($1/24^{\text{th}}$ degree $\sim 4\text{-km}$) gridded dataset of surface meteorological variables.”

Soil dataset. Gridded Soil Survey Geographic (gSSURGO) data were downloaded for all the fields with a spatial resolution of 10 meters. Soil information for the SALUS model includes

maximum ponding height (mm), soil class description, and for each soil layer, depth to the base of layer (cm), bulk density (Mg/m^3), clay/silt/coarse fragment content (%), saturated soil water content (m^3/m^3), drained upper and lower limits (m^3/m^3), saturated hydraulic conductivity (cm/h), organic C (%), total N (%), pH, and soil hospitality factor (for root growth).

Landsat Analysis Ready Dataset (ARD). Landsat ARD products are generated by USGS. They meet the highest scientific standards and can be used directly to monitor landscapes without any further processing (Dwyer et al., 2018). The ARD products are available for Landsat 4 to 8 over the continental United States. These products are processed to create a common tiling scheme, which is modified from the Web-Enabled Landsat Data (WELD) system. Each tile contains $5,000 \times 5,000$ 30-m pixels acquired in a day. These ARD products have shown great potential in terrestrial monitoring.

Fields in Michigan are located in tiles of path 6/row 22 (022006) and path 7/row 23 (023007). For both tiles, any individual Landsat scene is insufficient to fill the entire tile, leaving some black areas in the images. To make tiles more spatially complete, Landsat 8 surface reflectance (SR) images were the main images used after 2012. Landsat 7 images acquired one day before or after the Landsat 8 acquisition, if they existed, were also downloaded as complementary material to fill the areas not covered by Landsat 8. For example, we downloaded a Landsat 7 image acquired on June 17th if a Landsat 8 image from June 18th had been obtained. For 2012, only Landsat 7 ARD were collected for analysis. Image acquisition dates were set as June 1st to September 30th, which covered most stages of the growing season according to local planting dates. Tiles for Illinois, Indiana, and Iowa are 020010, 022008, and 018007. In addition, we also downloaded ARD surface temperature (ST) images at same acquisition dates as SR images. These images were derived by applying the single channel algorithm on Landsat 8 Band

10, and Landsat 7 Band 6. Since the ARD product is designed for direct use, the only preprocessing step was a cloud and shadow mask by thresholding on the quality assessment (QA) band.

4.4 Methods

4.4.1 Landsat 7 and 8 images mosaic

For mosaicking Landsat 7 and 8 images, the spectral differences for maize due to adjacent acquisition dates were assumed to be negligible as the impact of surface changes between two images would be sufficiently small within such a short period (Roy et al., 2016). However, differences between sensors must be accounted for before mosaicking. First, we built linear relationships between SR from Landsat 7 and 8 using the overlapping clear pixels. Cloud/shadow and saturated pixels were excluded. An additional filter proposed by Ju et al. (2012) was applied to the overlapped pixel set to remove land cover and surface condition changes that may have occurred in the one-day period between the two sensors (Equation 18). Corresponding Landsat 7 and 8 SR values captured one-day apart were only discarded when the SR difference in the blue band was greater than 100% of their average. Afterward, Landsat 7 SR was converted to pseudo-Landsat-8 SR using the linear relationships, filled cloud/shadow in Landsat 8, and black area in the tile. This mosaicking step was not conducted on ST images because thermal emittance are highly variable in space and time. Thus, Landsat 8 ST was used if the clear area on this image was higher than Landsat 7, and vice versa. This resulted in lower percentage of clear area in ST images than SR images. The green chlorophyll vegetation index (GCVI) was calculated after image mosaicking using Equation 19. We used this index because it is less influenced by LAI saturation of dense canopies when compared to other VIs (Gitelson et al., 2003).

$$\frac{\left| \rho_{blue}^{SR,L8} - \rho_{blue}^{SR,L7} \right|}{0.5 \times \left| \rho_{blue}^{SR,L8} + \rho_{blue}^{SR,L7} \right|} > 1 \dots \dots \dots \text{Equation (18)}$$

$$GCVI = \rho_{NIR} / \rho_{GRN} - 1 \dots \dots \dots \text{Equation (19)}$$

where ρ_{NIR} and ρ_{GRN} are SR of near-infrared and green bands. This index is nearly linear to LAI values based on an empirical relationship derived from (Nguy-Robertson et al., 2012).

4.4.2 Cloud removal for mosaic VI

Cloud and shadow contamination often hinder crop monitoring with optical images in these states. We applied the gap-filling algorithm developed by Yan & Roy (2018) to fill gaps in mosaicked Landsat GCVI and ST data. The main steps of the gap-filling algorithm include: 1) image segmentation using spatial-temporal information from the multi-temporal GCVI image, which was divided into two groups (≥ 3 pixels and < 3 pixels) for the subsequent process, 2) clustering segments, and 3) for segments with missing pixels (S_{gap}) in a specified target image, extracting an alternative segment (S_{alt}) with similar time-series spectral information and non-missing pixel values in the target image and replacing GCVI values of missing pixels in S_{gap} with their equivalent alternative pixel in S_{alt} . The similarity measurement in the above steps was based on a refined spectral angle mapper similarity metric. More details about this algorithm can be found in Yan & Roy (2018).

4.4.3 Yield stability zones and Cumulative Drought Index (CDI) calculations

4.3.3.1 Stability zone calculation

Following the principles of precision agriculture and site-specific management, farmers divide crop fields into areas of uniform conditions either based on soil properties or yield, known as management zones (MZ), as a baseline for many farming decisions (Nawar et al., 2017). Following an established protocol described in Chapter 1, we created yield stability maps as MZs in fields that have at least 3-years of crop yield data (Blackmore, 2000).

4.3.3.2 Subdivision of unstable zones using topography features

Martinez-Feria & Basso (2020) showed strong evidence using thousands of yield maps that unstable zones clearly respond to the spatial interactions between soil, weather, and topography, rather than just responding to soil variability. To better capture these interactions, we performed an additional division of the unstable areas based on the landscape position of each unstable pixel. Specifically, we calculated topographic position index (TPI) by comparing the elevation of each cell to the mean elevation of a specified vicinity around that cell (Weiss, 2001). We used 50 meters as the threshold to define the nearby cells. Unlike the six classes in the data's general topography description (hilltop, depression, upper/middle/flat/lower slopes), the unstable zone was classified into three classes: UN-hilltop (UNHIL) ($TPI \geq 1SD$), UN-depression (UNDEP) ($TPI < -1SD$) and UN-other (UNOTH) including all slope classes, where SD is the standard deviation of TPI values for each field. We used this classification because most unstable pixels are found in the depression, slopes or hilltop positions in these fields (Martinez-Feria & Basso, 2020).

4.3.3.3 Cumulative drought index calculations

The validated SALUS process-based model was used to calculate the Cumulative drought index (CDI) values along with crop phenology at a daily time step. We ran SALUS simulations by stability zone to obtain variations in subfield-scale water stress and CDI values. Since more than one soil type may co-exist within a zone, only the soil type that occupies the largest area in each zone was used to represent that zone. In this spatial version of SALUS, soil properties are parameterized automatically based on the stability maps characteristics. For instance, compared to the low-yielding areas—which are characterized by compacted soil, low hydraulic conductivity, and shallow soil profile—soils in the high-yielding areas are deeper and better represent higher soil water holding capacity, as shown by high and stable yield independent of rainfall (Maestrini

& Basso, 2018a). Variables such as the soil hospitality factor (SHF), time to ponding, ponding capacity, saturated hydraulic conductivity, and runoff and rainfall water routing coefficients are modified depending on topographic position. In wet seasons, the final yield for UN-depression is impacted by the SALUS anoxia routine. The decrease in yield is congruent with observations made in previous work (Martinez-Feria & Basso, 2020). Plant population is also reduced to match observed plant heterogeneity measured in these fields. This modification is made based on the evidence that sparse vegetation density is often associated with poor soils (Drury et al., 1999).

Except for the above modifications, model input parameters were the same for all stability zones. The model was run with a 30-year maize-soybean rotation from 1990 to 2019, but the individual fields' crop rotation history was included in the 30-year rotation. Maize cultivar duration (maturity group) and initial soil inorganic N were provided by farmers for some of the fields included in the analysis, otherwise it was estimated by the model or domain knowledge. Weather data was derived from the gridMet grid point nearest to the center of the stability zone. Planting dates were taken from NASS crop progress data at state-level and harvest dates were set to October 15th for maize and soybeans, and July 15th for wheat. Row spacings were 30 inches for maize, 20 inches for soybeans, and 7.5 inches for wheat, while planting depths were 2 inches for maize and 1 inch for soybeans and wheat. No-tillage was implemented in most fields, unless growers provided tillage details. Nitrogen application rates were derived from the USDA NASS Arms dataset and from Chapter 1. The nitrogen rates used for maize were 204 kg/ha for Illinois, 209.5 kg/ha for Indiana, 180 kg/ha for Iowa, and 170.5 kg/ha for Michigan (USDA - NASS, 2017). Nitrogen application was split: 1/3 applied at planting, and 2/3 with a side-dress at 45 days after planting. Fertilizer depth was set as 2 inches below the soil surface. The outputs of the SALUS

model include daily LAI, biomass, grain yield, plant extractable soil water (PESW) contents at multiple soil layers, drought and nitrogen factors, and phenology dates.

To assess the effect of drought stress on crop yield, we calculated the cumulative drought index during the grain filling period (the end of vegetative stage) when crop yield is most sensitive to water stress.

$$CDI = \sum_{i=1}^m ET_a^i / ET_p^i \dots\dots\dots \text{Equation (20)}$$

Where ET_a and ET_p indicate the actual and potential crop evapotranspiration, which are the output from SALUS model. The potential evapotranspiration is calculated using modified Penman-monteith equation in SALUS. The ratio of ET_a and ET_p is the drought index (DI) for a single day and CDI is the cumulative DI over m days and the superscript i is the i th day since the beginning of grain filling period to the prediction date. A value of 1 for DI indicates no water stress on that day, meaning that the evaporative demand is fully satisfied; values less than 1 indicate different extents of water stress under limited soil water and precipitation conditions (e.g. 0.8 means that the soil could only supply 80% of the atmospheric evaporative demand, translating in a plant growth occurring at 80% capacity).

4.3.3.4 Yield predictions and validation

The machine-learning algorithm Random Forest (RF) has been applied in yield prediction in several studies (Liaw & Wiener, 2002). RF is an ensemble learning method constructed from multiple decision trees. In each tree, a random subset of input features is extracted for making decisions. The estimated yield value is the mean fitted responses from all trees. Here, RF regression was conducted using the R “randomForest” package (Liaw & Wiener, 2002). The parameters for RF regression were the square root of the number of input features—which was used to split the

data at each node—and the number of trees was 500. To test if date of prediction impacted model accuracy, the first prediction date was set as the day of the year (DOY) when all the fields were already five days into the grain filling phase, and predictions were made again each time during the growing season a new RS image became available. For each prediction, we used GCVI alone, GCVI plus ST, and then combined GCVI and CDI up to the prediction date to measure the effect of CDI on accuracy. Twenty percent of pixels were randomly selected as training and eighty percent of the data were used to validate the model. Since the number of images and image acquisition dates are not consistent across location (Landsat ARD tile) and year due to cloud cover, we fit a separate model for each tile-year combination, totaling up to 14 models. Multiple fields are contained in a single tile, so the same model was used for each field under the same tile. The performance of the model built from each prediction was validated using the coefficient of determination (R^2) and mean absolute percentage error (MAPE) using Equation 21. We compared different years' improvements with CDI in the remotely sensed yield prediction using a t-test, with the null hypothesis that improvements are the same each year.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Yield_{actual} - Yield_{predicted}}{Yield_{actual}} \right| \dots \dots \dots \text{Equation (21)}$$

where n is the number of predicted pixels.

4.5 Results

4.5.1 Accuracy of the new sub-field scale yield prediction model

For each tile-year, we validated the performance of the sub-field yield prediction models with and without CDI using yield monitor maps in Figure 8. There were not VI and ST results for the first prediction dates of good and poor years in Indiana because both ST images did not meet

the gap filling requirements due to high cloud coverage. This situation also happened for the poor year in MI (022006). Overall, adding CDI to RS data improved the R^2 and MAPE between predicted and actual yield, compared to predictions based on RS data alone, with the biggest improvement in the poor year of MI. In the poor year, the first predictions with VI alone produced R^2 values of 0.27 for MI (022006) and 0.49 for MI (023007), which were improved to 0.52 and 0.62, respectively, after adding CDI to the model. Correspondingly, the MAPE decreased from 10.72% to 8.49% (022006) and 19.87% to 16.87% (023007). As the prediction date approached the end of the season, the differences in R^2 between models with two input variables decreased to 0.15 and 0.1 for the last prediction. The two-sample t test results showed that improvements were significantly lower in the average and good years with t values of 4.05 and 4.89, $p < 0.01$, which rejected the null hypothesis. In most cases, the VI and ST results (blue lines) lies within red and green lines, suggesting that adding ST images in the model could improve prediction accuracy, to a lesser extent than models using CDI. We still found large improvements in IL (good year) by including CDI in the yield prediction, and moderate improvements in IN (average and poor year), while few improvements were observed in other tile-years. In addition, we calculated feature importance in the VI and CDI regression model using the importance function from the RF package, with the three most important features listed in Table 13. The last prediction models at the end of season were chosen so that images acquired at all crop stages can be examined. IncMSE is the increase of the mean squared error (estimated with out-of-bag-CV) when one variable is randomly permuted in the training stage. Relative to GCVI at any date, CDI was always the most important feature with the highest IncMSE values occurring in MI. High importance values for GCVI images acquired early in the season (DOY 150 to 170) were observed, while sometimes mid- or late-season images also showed high importance due to closely explained variations in

crop yields. For other states, CDI was also the most important feature in good years and the second most important feature in average and poor years, except the average year in IA. In summary, CDI provided unique information on water stress that improved crop yield predictive power.

We also tested the scalability of CDI in maize yield prediction using the “two-window” approach (Deines et al., 2021). To composite multi-temporal images into two images; the first GCVI image comes from the maximum value observed in an “early season” window between day-of-year (DOY) 165–200 (June 14– July 19) and the second for a “late season” window between DOY 201–240 (July 20 – August 28). After this composition step, images from all Landsat tiles and years were combined together to run a RF model using 20 percent of pixels, randomly selected, as a training set and the remaining as a validation set (Figure 9). This resulted in an accuracy of R^2 as 0.68 and MAPE of 13.99%. When we added the CDI accumulated over all grain filling season to match this image composition method in the model, the R^2 increased to 0.76 and MAPE decreased to 11.62%. Another comparison was conducted by selecting training samples at field level instead of pixel level, that is, twenty percent of fields were randomly picked for training. Results indicated that using fields as training samples caused lower accuracy, however, CDI was able to improve the prediction.

We plotted subfield variations of CDI at the end of the season, and actual and predicted yield maps in 2012 in Figure 10. The yield variations between the left and right sections of field 27 were not revealed by CDI, so adding CDI to VI did not improve the yield prediction. On the other hand, the differences between high and low yields were captured in CDI images in the other fields, especially the northern and southeastern parts in field 70. These captured variations were also found in the predicted maps with VI and CDI together, which confirmed the higher accuracy of yield prediction.

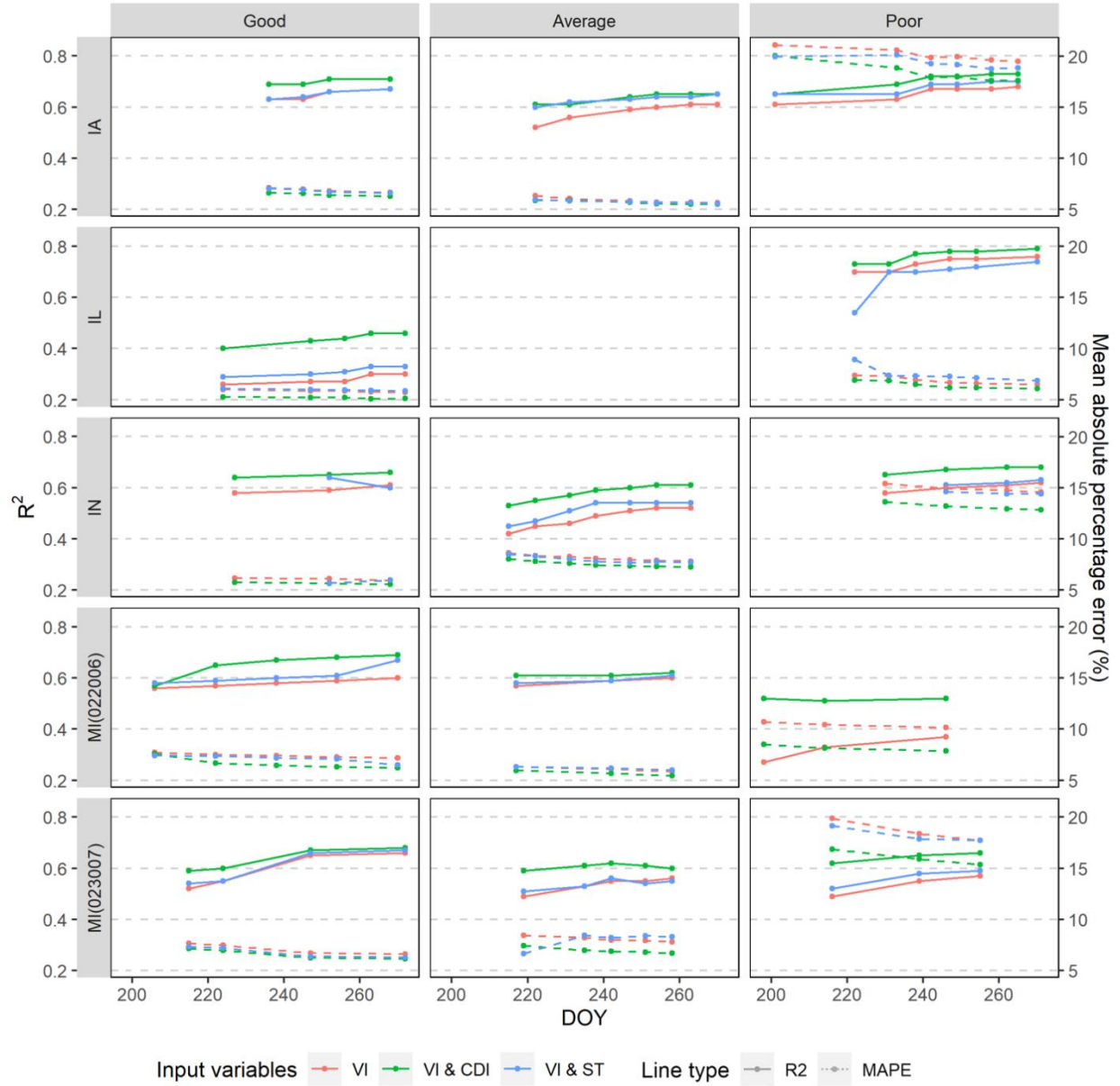


Figure 8 Accuracy assessment for maize yield prediction models in Iowa (IA), Illinois (IL), Indiana (IN), and two tiles (022006 and 023007) in Michigan (MI) and in three testing years: good, average, and poor years. “VI” refers to multi-date gap-filled Landsat ARD images as model input, “VI & ST” refers to combination of multi-date gap-filled Landsat ARD GCVI and ST images, while “VI & CDI” refers to input including multi-date images and CDI together. Points on the lines indicate the prediction dates when new Landsat image were acquired since the maize grain filling period.

Table 13 Three most important features for yield prediction with all GCVIs and CDI (GCVI DOY in parentheses). Unit of IncMSE is percentage (%).

Tile	Year					
	Good		Average		Poor	
	Variable	IncMSE	Variable	IncMSE	Variable	IncMSE
IA	CDI	68.1	GCVI (231)	57.5	GCVI (242)	71.1
	GCVI (252)	62.8	GCVI (247)	39.1	CDI	56.2
	GCVI (165)	55.2	GCVI (254)	36.4	GCVI (258)	55.7
IL	CDI	24.3			GCVI (222)	27.3
	GCVI (224)	18.7			CDI	25.2
	GCVI (263)	16.1			GCVI (238)	22.9
IN	CDI	57.6	GCVI (238)	32.7	GCVI (214)	69
	GCVI (220)	54.7	CDI	31.3	CDI	63.9
	GCVI (188)	51.1	GCVI (231)	28.6	GCVI (246)	47.3
MI (022006)	CDI	64.1	CDI	65.2	CDI	62.8
	GCVI (214)	42.7	GCVI (217)	49.3	GCVI (158)	46.6
	GCVI (166)	38.8	GCVI (169)	46	GCVI (190)	43.6
MI (023007)	CDI	57.8	CDI	49	CDI	87
	GCVI (159)	56.6	GCVI (258)	34.7	GCVI (247)	57.8
	GCVI (255)	45.1	GCVI (155)	34.7	GCVI (188)	53

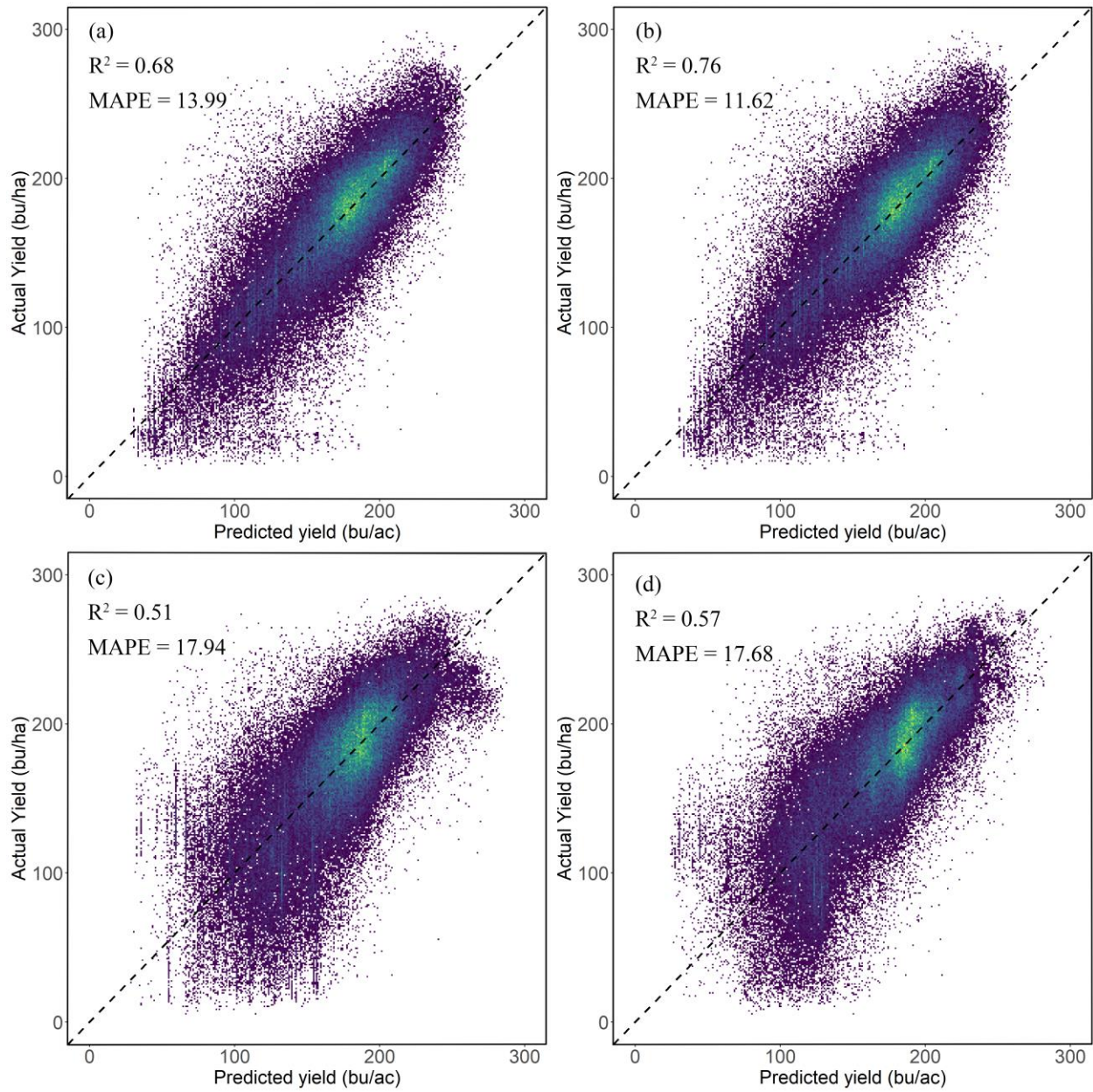


Figure 9 Scatterplot of predicted and actual yield from four situations: a) composited GCVI using pixel samples, b) composited GCVI and CDI using pixel samples, c) composited GCVI using field samples, and d) composited GCVI and CDI using field samples.

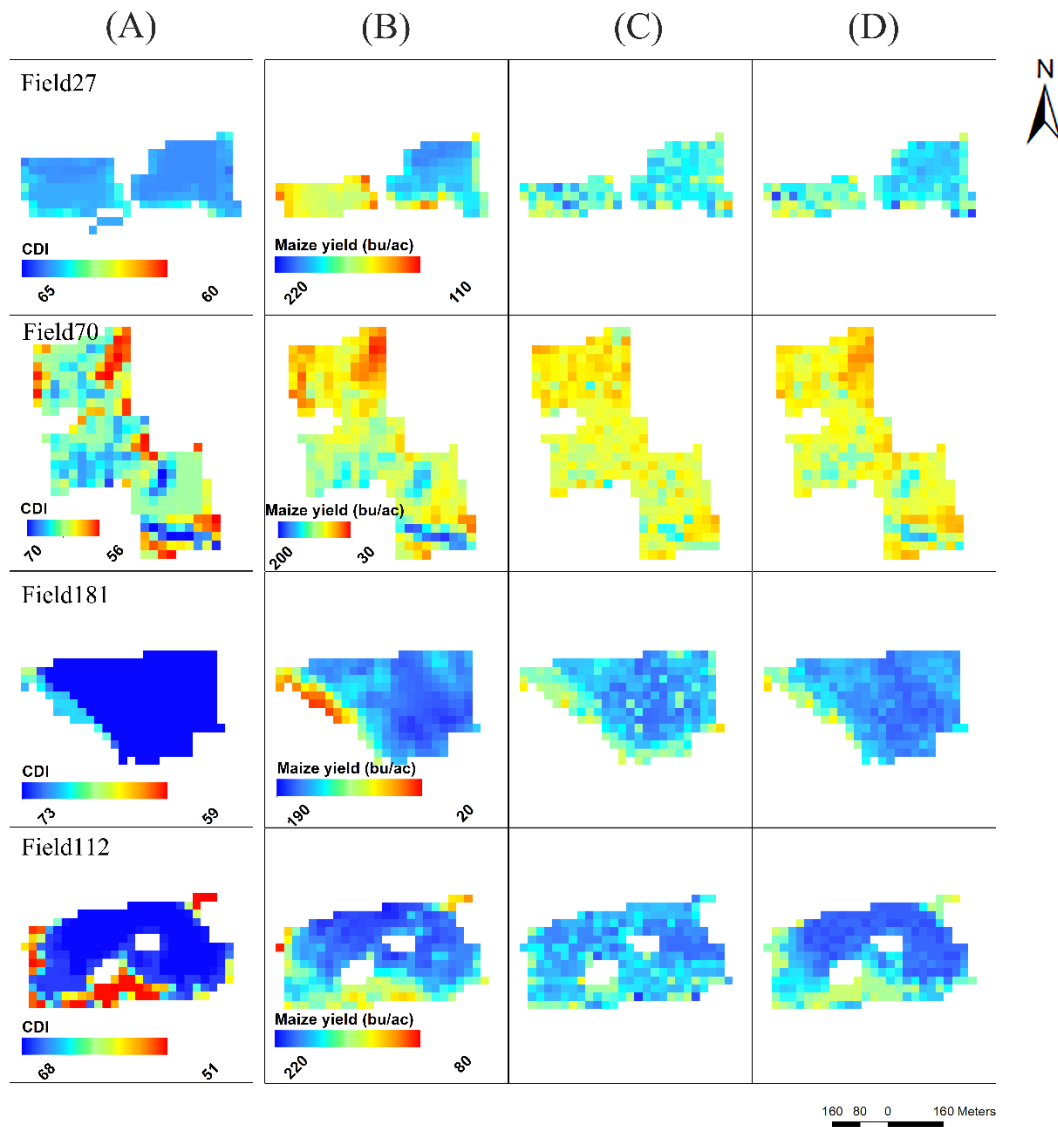


Figure 10 Subfield-level CDI (A), yield maps (B), yield predictions using VI alone (C) and using VI and CDI (D) in 2012. For actual and predicted yield maps in each field, we used the same color ramp.

4.5.2 SALUS model testing

We conducted detailed interpretation of SALUS simulated yield and CDI to help understand the reason why CDI could improve yield prediction. To test the accuracy of SALUS in simulating maize yield, we compared maize yield simulated by SALUS outputs with yield monitor

datasets in randomly selected fields (16) from four states (Figure 11). The zone-level actual yield is the average yield within each zone. There are one-year or two-year data shown in some fields since not all fields were planted with maize in the three testing years. SALUS performed well in estimating maize yield in fields 160, 167, 328 and 335, while HS and MS yield was overestimated in fields 261, 268 and 296. Yield underestimation was found in fields 39 and 122. Across all the field-years, SALUS overestimated 0.25 ton/ha in Illinois, 3.92 ton/ha in Indiana, 2.47 ton/ha in Iowa and underestimated 3.24 ton/ha in Michigan. Despite the fact that SALUS was run completely uncalibrated (no model reparameterization based on observed yield) the crop yield was simulated with moderate to low errors. For the stable zones, HS zones always produced the highest yields followed by MS and LS, which was consistent across the field-year. Yield differences for unstable zones varied by field. For example, yield in UNDEP was as high as MS in Field 122, while it can be as low as LS in Field 160. Despite these field-specific characteristics, the zone-level variation of actual yield were reflected in the SALUS simulation, which demonstrated the strength of using stability maps to spatially execute the SALUS model.

The comparison of maize yield with the addition of the simulated CDI during the whole grain filling period is shown in Figure 12, for the same fields shown in Figure 11. Compared to other years, yield and CDI were lower in the poor year, however, HS and MS could remain high CDI value due to greater water availability in the soil profile (such as Field 122). Blue regression lines in the subplot showed that CDI positively reflected crop yield variability for most fields, that is, higher yields were associated with higher CDI (lower water stress). We also observed negative correlation in fields 148 and 328 and little correlation in fields 160 and 167 where limiting factors other than water stress influence crop yield. These results provided evidence that CDI can be used as a proxy to indicate within-field crop yield variations. We mapped the range of CDI by stability

zone from all the fields to show the within-field CDI variations at the state level in Figure 13. We found that depressional areas (UNDEP) tended to have the highest CDI (indicative of low water stress), followed by HS, MS, and the remaining zones. The difference in CDI between zones was greater in the poor year than in the average and good years. This provides evidence that the spatial variation in drought stress within fields is exacerbated when precipitation is scarce. Highest CDI was observed in good years, followed by average and poor years for IL, IA, and MI, while the average year had higher CDI than the good year in IN, even though not significant. IL had the highest and lowest CDI over all the state-year combinations, while the difference in CDI among years was the smallest in MI. This indicates that the within-field water stress variation differs not only at field scale (Figure 11), but also at state scale since more climate and environments are present across a state.

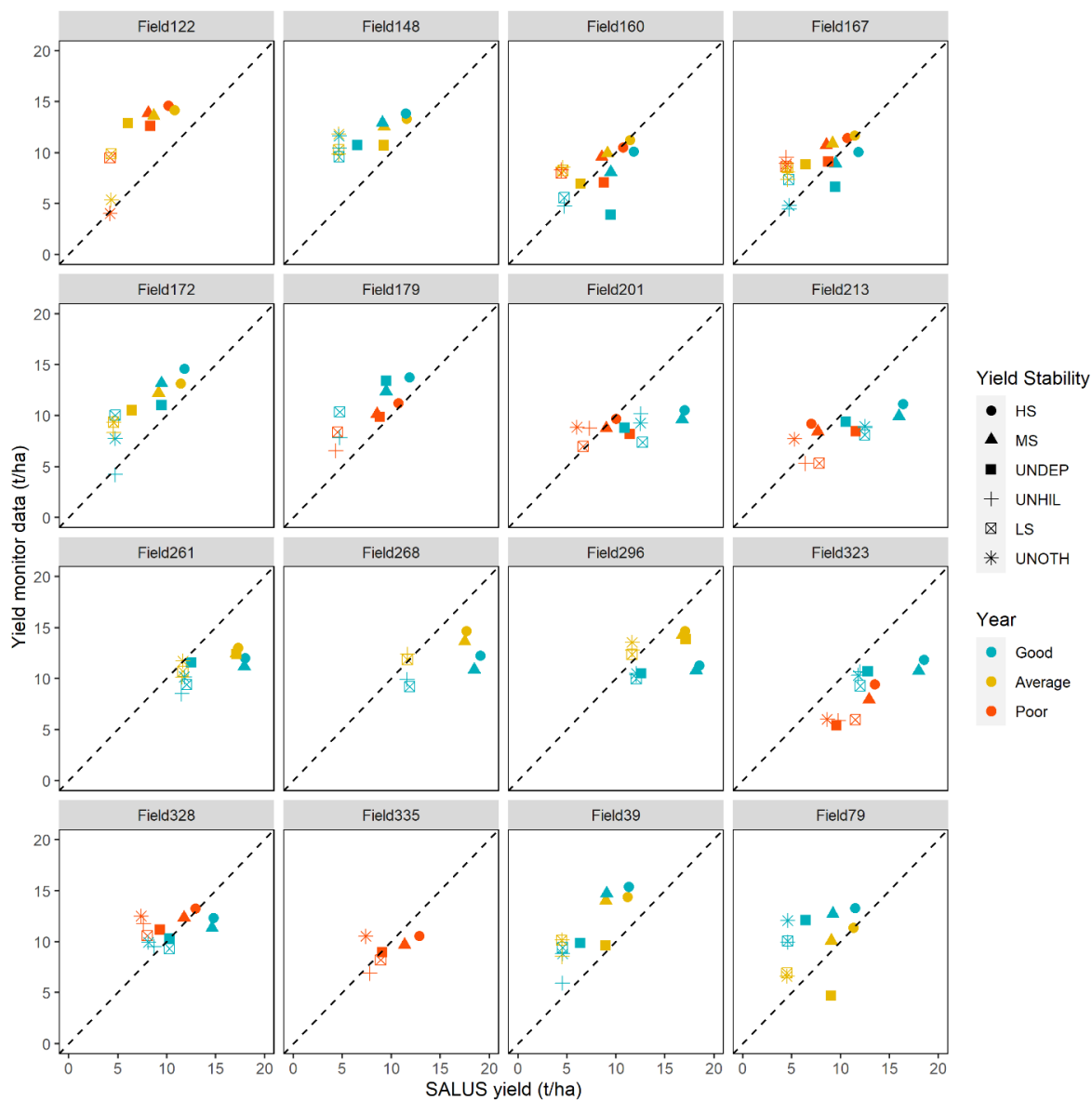


Figure 11 Scatter plots of SALUS simulated yield vs yield monitor data at zone level for the selected 16 fields. Yield stability classes include high stable (HS), medium stable (MS), low stable (LS), unstable-depression (UNDEP), unstable-hilltop (UNHIL), and unstable-other (UNOTH). The 1:1 line is added in each subplot.

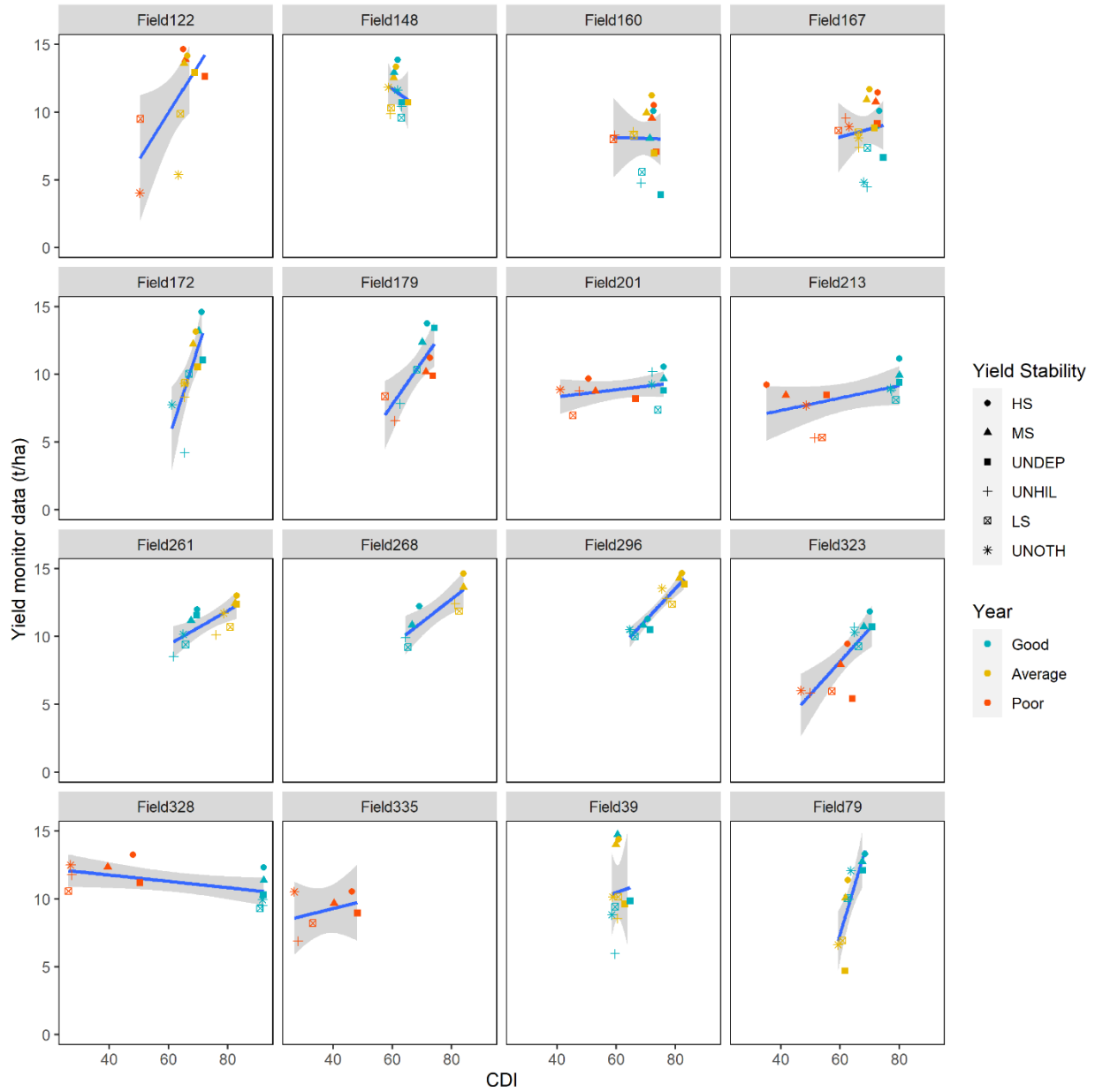


Figure 12 Scatter plots of yield monitor data vs CDI during the whole grain filling period at zone level for the same fields in Figure 4. Regression lines (blue lines) and grey area (the 95% confidence level interval) are added to show the response of maize yield to CDI.

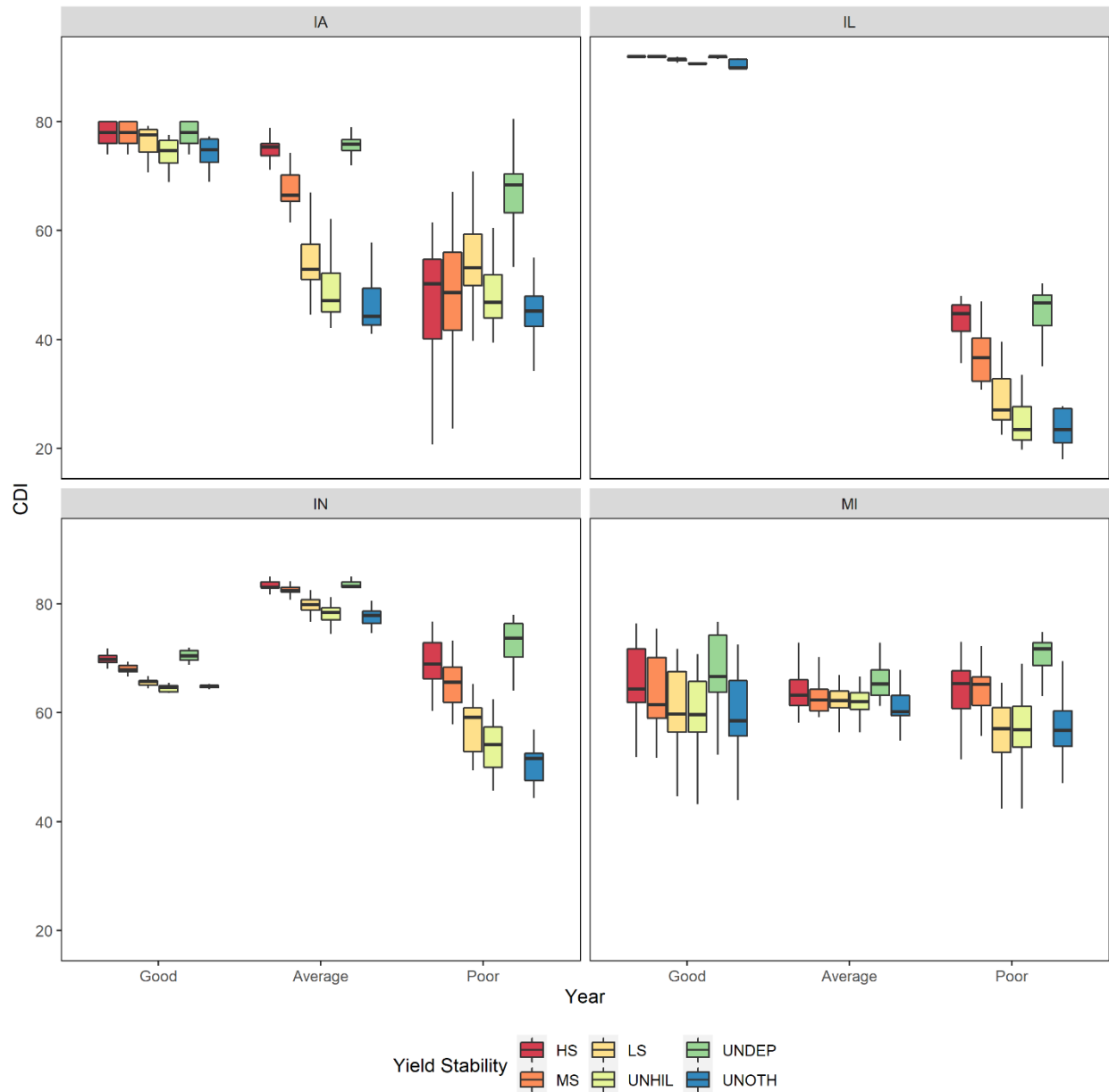


Figure 13 Box plots showing distribution of whole grain filling period CDI by stability zone in each state.

4.5.3 Long-term pattern of CDI

Although we mainly focused on demonstrating the potential of CDI to improve in-season yield predictions, we also wanted to show that the 30-year output of the CDI from SALUS could help classify locations in the probability of occurrence of CDI outcomes based on historical weather. We examined CDI over 30 years in each individual field and provide preliminary

interpretation on its values to be used as historical benchmarking means in absence of in the season CDI value. The cumulative probability function of CDI of the entire maize grain filling period for four fields is shown in Figure 14 and the CDI values for field 27 is reported in

Table 14. The curve in field 339 shows a bimodal distribution of CDI, with most values ranging from 15 to 45 and 65 to 80. CDI was evenly distributed in other fields, with CDI between 40 and 80. Similar to Figure 5, CDI in UNDEP was always the largest, followed by HS, MS, and UN zones, which suggested that the spatial variations of CDI were consistent among years. Based on this information, one could calculate the similarity in climate condition between the current season and historical years and pick the most likely historical year to estimate CDI values. This could be done by building statistical models to predict CDI with weather, soil, and other factors. For example, if the climate situation in 2021 is similar to that in 2000 for field 27, it is possible that CDI values in 2021 are around 75 for HS, MS, and UNDEP, and around 65 for the other stability zones. Users could apply these values without running crop models to better predict maize yield until they get the actual CDI in 2021. Together with CDI-yield relationships in Figure 12, one could further estimate crop yield within a reasonable range.

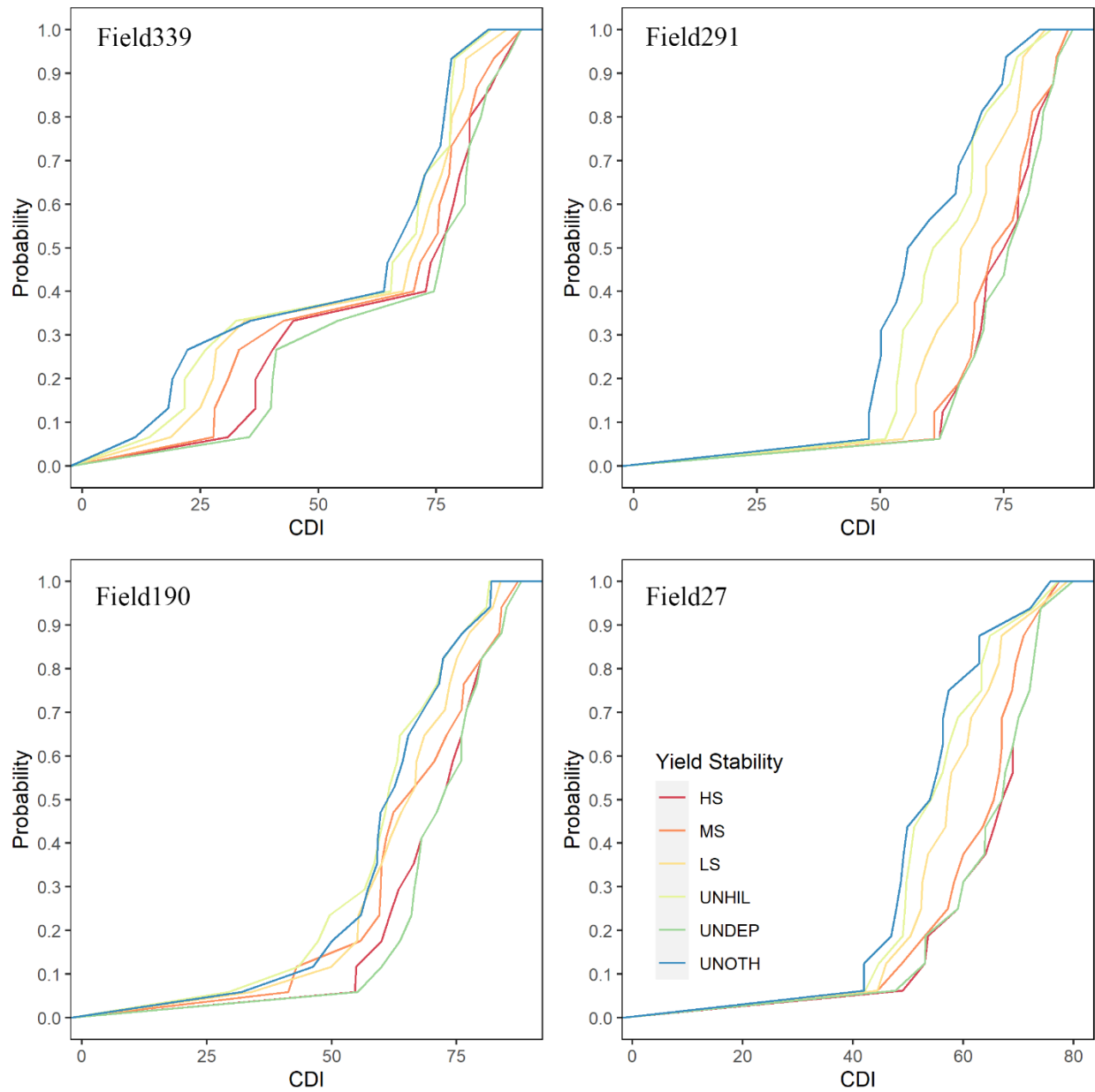


Figure 14 Cumulative distribution of CDI during whole grain filling period from 1990 to 2019 in four fields.

Table 14 CDI during entire grain filling period for Field27

Year	Stability					
	LS	MS	HS	UNDEP	UNOTH	UNHIL
1990	54.6	61.1	65.0	65.0	53.7	54.6
1992	54.6	60.0	60.0	60.0	54.0	54.6
1994	63.1	78.0	78.0	78.0	62.6	63.1
1996	38.6	35.1	41.3	39.7	33.7	37.7
1998	64.1	72.9	76.0	76.0	61.9	63.8
2000	66.6	74.4	75.0	75.0	64.9	66.5
2002	58.7	55.4	67.0	64.5	57.6	59.0
2004	69.8	77.0	77.0	77.0	63.9	69.8
2006	63.0	68.3	72.0	72.0	60.2	62.8
2008	28.9	36.8	43.3	43.0	28.6	28.9
2009	54.2	60.4	63.4	63.2	52.5	54.1
2011	64.5	66.5	67.8	67.6	64.0	64.5
2012	57.4	64.3	67.8	70.5	56.4	57.9
2013	68.5	75.5	77.0	77.0	66.1	68.2
2015	61.9	70.5	73.8	73.7	54.8	61.3
2016	62.0	60.0	66.0	63.7	59.6	61.1

4.6 Discussion

In this study, we showed that the inclusion of the CDI in remote sensing-based machine learning models improves in-season maize yield prediction. The greatest improvements in predictions were observed in the dry year (2012) (Figure 8, Figure 10). This is consistent with recent research which suggest that combining RS and crop model could enhance field-scale crop water stress detection and subsequent yield estimation (Hunt et al., 2019; Sayago et al., 2017; Zhao et al., 2020). However, stress information within the field is seldom described, and with the method proposed here, we can better appreciate the impact of spatial variability of water stress as a function of soil type and depth differences, position in the landscape, and management which will affect plant stands at subfield scale. the earliest prediction in the growing season with VI and CDI together outperformed the latest prediction with VI alone, highlighting the potential of CDI for predicting spatial variability of maize yield in late July, two months before crop harvest. The

degree of improvement decreased as more images were acquired because drought effects were eventually captured by the images. This mid-season prediction capability allows farmers to make decisions about sale prices and grain storage early in the season. Stresses other than water deficit may dominate in other study years that have sufficient precipitation, especially the wet year (2018), thus CDI did not enhance predictive capacity in that year as much as in the dry year. This study points out that an index similar to CDI but designed to capture N stressed could also lead to improvement in wet years. Yet, our analysis underlined the superior information provided by CDI beyond any single GCVI image in most years. This was also revealed in Figure 8 which demonstrates that yield prediction improvement from CDI was greater than the prediction based on surface temperature images. This is due to the fact that CDI takes temperature, precipitation, soil texture and topography effect into account, rather than just canopy temperature. It is worth noting that images acquired in early and mid-June also contributed large portions of the yield predictions. One possible reason is that early-season images captured the spatial variation in crop emergence, which explain the variations in the final yield to some extent W. Liu et al. (2004). In the scalable approach, models using CDI and GCVI again explained 76% of yield variation with a MAPE of 11.62%, better than models using GCVI as input. Sampling by field resulted in lower prediction accuracy than from the pixel sampling method, which is reasonable considering the inter-field variations. Nevertheless, the improvement from adding CDI is consistent. This supports the use of CDI for subfield level yield prediction over state and country scale.

The underlying reason for yield prediction improvement in our study is that stability maps, with additional topographic division for the unstable zones, are able to capture the water stress effects on yield (Figure 12). The spatial comparison of CDI among yield stability zones indicated that water availability in depressional areas is higher, and in agreement with conclusions from our

previous work (Martinez-Feria & Basso, 2020). Deep soils and redistribution of precipitation in fields allows UNDEP zones to retain more water. However, higher soil water availability does not always result in higher yield. UNDEP zones are also reported to yield 30% less yield due to the excess water in wet years (Li et al., 2019). On the other hand, in the low soil water area, modern maize cultivars may be planted to tolerate a short period of drought without much effect on final yield (Golbashy et al., 2010). For example, although the late-June precipitation in 2016 was as low as in 2012, the moderate rainfall events in early July of 2016 partially relieved plants from the drought situation. Thus, farmers would need to make tactical decisions in the growing season, and the integration of CDI and remote sensing can help them better manage their crops based on the weather in a specific year, such as scheduling irrigation or adjusting N fertilizer application.

Another important phenomenon that can be observed from Figure 12 is that fields' yield-water stress response differ, underscoring the importance that fields should be modeled individually even when under similar weather. This is because other field-dependent information - soil, topography, management – vary from field to field, which also plays a role in determining water stress. The state-scale comparisons in Figure 13 further supported this point. We also noted that the spatial pattern of CDI within the field is consistent over 30 years (Figure 14), implying that stability maps are reliable irrespective of interannual weather variability. Moreover, the long-term historical information (

Table 14) offered the chance estimate CDI values in upcoming growing seasons, which will be tested in our future work.

The primary dataset, yield maps, are increasingly available in large commercial farms or from companies, as indicated in Deines et al. (2021). Other environmental factors, such as topography and soil, are also public now. Moreover, our method is not limited to the use of Landsat

imagery as input. As UAV technology and commercial airborne high-resolution imagery are available at high frequency, farmers can access images with no cloud contamination regularly during the season. Finally, the ability of crop models to describe processes of crop growth across various conditions facilitates the extended application in other regions and other crops.

This approach is not free of uncertainties. Although yield losses due to water stress were well described in Figure 11, we observed that the crop yield simulated by uncalibrated SALUS model can be improved. Model errors are mainly due to unrepresentative inputs into the SALUS model as results of the scales of the model input (weather at 4km scale rather than on site; soil depth and level of fertility are crudely estimated with heuristics, as well as inability to capture the true number of plants per m² needed by the model vs the real heterogenous plant stand conditions and the true tillage practices occurring in a field. On this last topic, tillage and cover crops detection is improving using machine learning and remote sensing images (Deines et al., 2021; Zulauf & Brown, 2019). Proximal sensing (like a digital soil map of clay, organic C, pH) could help improve the accuracy of soil input in the crop model. Another uncertainty is the effect of irrigation on water stress and crop yield. Although most fields in Michigan are rainfed, some farmers reported the use of irrigation equipment in some of their fields. Thus, the assumption of no irrigation in our model would underestimate CDI and yield for the irrigated fields, which may explain why crop yields in some fields had a minor response to CDI (Figure 12). These yield underestimations or overestimations can be improved with more detailed planting date, cultivar information, and nitrogen application rates at finer scale. Improvements on satellite image quality are also required. Although the gap filling algorithm could address cloud and shadow contamination to a great extent, the performance was still poor for areas with cloud cover.

Overall, the improved yield prediction and deep understanding of water stress showed the opportunities of the proposed method in improved prediction of maize yield at subfield scale.

4.7 Conclusion

This study presented a new method to integrate crop water stress from process-based crop models with remote sensing images for maize yield prediction for over three hundred fields. The SALUS crop model was run based on yield stability zones and topographic features, and the simulated water stress was in a good relationship with crop yield. Results indicated that this method was able to provide accurate yield mapping two months before crop harvest. This study explored the possibility of mapping subfield crop information, which is critical progress toward the target of precision agriculture.

4.8 Acknowledgment

This research was supported in part by the US Department of Agriculture National Institute of Food and Agriculture award numbers: 2019-67012-29595, 2018-67003-27406, 2015-68007-23133 and by the Great Lakes Bioenergy Research Center, US Department of Energy, Office of Science, Office of Biological and Environmental Research under award numbers DE-SC0018409 and DE-FC02-07ER64494. The authors wish to thank Lydia Rill and Ruben Ulbrich for their valuable support on data processing.

CHAPTER 5: ASSESSING IN-SEASON IMPACTS OF SPATIAL AND TEMPORAL CHANGES OF MAIZE GROWTH ON YIELD AT FIELD SCALE: A COMPARISON BETWEEN IMAGERY AND YIELD STABILITY MAPS

5.1 Abstract

While historical yields and remote sensing observations serve as two of the most important sources in predicting crop yields, there is relatively little research to compare the accuracy derived from both type of information. Utilizing biweekly Planet images from June through September (2018 - 2020) and observed yield maps over 115 fields located in Indiana, Iowa, Michigan, and Minnesota, we found that yield stability maps were able to provide accurate yield level estimations in stable yield areas (HS and LS). This accuracy was similar to a model trained on images acquired from June through August. Results imply that yield stability map can be used as guidance to schedule fertilizer applications in the stable yield zones before the growing season without any in-season information. We also conducted an experiment to determine if the best correlated images vary by field and if yield predictions based on each best image would improve the accuracy. Overall results suggested that the images did improve accuracy. On the other hand, a significant improvement was found when predictions were carried out for each field. These results suggest that the difference among fields is the biggest limitation accurate large-scale yield predictions. One strength of remote sensing predicting approaches is the ability to monitor crops within season in real-time. By examining the changes between two successive acquired images, we found the greatest observed spatial and temporal variations of crop growth at the beginnings and near the ending of growing seasons, while relatively pretty stable during the mid-season. However, these large temporal changes did not always impact final yields since crop growth is a dynamic process. We further analyzed the impacts of environmental

factors on crop growth during these two periods, and results showed that rainfall and soil water availability impacted the rates of crop vegetative growth and senescence. This study is an initial investigation to assess temporal changes in the spatial variation of crop conditions within a growing season and suggests a need for more comprehensive knowledge about crop monitoring methodology and procedures.

5.2 Introduction

Precision management at the subfield scale through remote sensing and big data technology approaches hold promise for sustaining high crop yields and avoiding continuing negative environmental impacts of current intensive agricultural production systems (J. Liu et al., 2022; Veldhuizen et al., 2020; Weiss, 2001). The concurrent availability of high-resolution remote sensing imagery, advanced computing, and newly available high quality, ground truth yield data opens the door to reliable assessment of subfield scale crop growth and performance (Deines et al., 2021; Hatfield et al., 2020). The immediate and primary science questions and objectives are focused on agronomic challenges of optimum nitrogen applications to maximize maize yields (Hunt et al., 2019; Ji et al., 2021; Z. Jin et al., 2019; Kayad et al., 2019) while mitigating nitrogen leaching and greenhouse gas emissions (Basso et al., 2019; J. Liu et al., 2022). These approaches and tools also enable large, systemic analysis of food security, landscape biodiversity, and land use diversification informed by patterned plant growing capacity of the landscape (Hernández-Ochoa et al., 2022).

The heterogenous response of the landscape to agricultural management can be separated into homogeneous yield stability zones. Characterization of common yield zones with similar soil properties, topographic features, and historical yield has been conducted for decades (Nawar et al., 2017; Plant, 2001). Identification of common zones improves profitability, nitrogen use

efficiency and conserves other resources in cereal, fruit, and nut crop production systems (Basso et al., 2019; Basso, Ritchie, et al., 2011; Koch et al., 2004; Martinez-Feria & Basso, 2020; Panda et al., 2010). In previous Chapters, we defined fields into three zone: high stable yields (HS), low stable yields (LS), and unstable yields based on observed yield monitor data. Identifying high and low yield zones allows farmers to adjust management strategies to match the yield potential of stable zones, thereby avoiding nitrogen loss and reduced profitability (Basso, Ritchie, et al., 2011). In contrast, unstable zones require in season tactical management as yields in unstable zone are driven by hillslope position and in-season weather factors (too much or too little water) (Basso et al., 2019; Martinez-Feria & Basso, 2020). In-season remote sensing imagery is needed to estimate yields for unstable zones in real time.

Remote sensing images provide real-time spatial information about the crop conditions at the subfield scale needed for tactical in season management decisions. The post-hoc crop yield estimation of maize is well established in the literature and validated at the landscape scale and on small holder fields (Deines et al., 2021; Z. Jin et al., 2019; Padilla et al., 2018). The Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and Ratio Vegetation Index (RVI) are commonly used indices to represent crop biophysical parameters, such as leaf area index and chlorophyll content (Clevers & Gitelson, 2013; Hassan et al., 2019; Xue & Su, 2017). These indices are prone to saturation, limiting their usefulness in the high leaf area index setting of most maize fields. In contrast, the Green Chlorophyll Vegetation Index (GCVI) does not saturate to the same degree (Gitelson et al., 2003). Deines et al. (2021) used GCVI to compare corn yields relative to onboard yield monitor data. They found that yield estimates with GCVI improved with a two-image window and harmonic best fit metric which captured additional information about the grain filling stage beyond the observed maximum.

Empirical yield estimation models directly from images are not generalizable, especially at subfield scale, though some patterns are evident. Johnson (2014) compared the correlations between many satellite products, such as NDVI, surface temperature and rainfall, with corn and soybean yields in the United States, and found that the strongest correlations came from NDVI images in early August or late July. Lobell et al. (2015) compiled multi-temporal Landsat images into two images at early-season (DOY 161–200, or mid-June to mid-July) and peak-season (DOY 201–240, or mid-July to late-August), and proposed a scalable method for maize yield mapping over the US. Burke & Lobell (2017) replicated this method to detect yield variation in small fields in Kenya with yield estimation accuracy similar to survey-based measures. In Chapter 4, we combined the cumulative drought index derived from crop simulation model output using yield stability maps as zones with remote sensing images to capture the subfield variation of maize yield. While the body of literature on in-season sub-field scale yield estimates is large, translation into potential management practices is still difficult because management practices also depend on the fields' characteristics, such as soil and landscape situations.

Use of yield stability maps and remote sensing vegetation indices to observe in season dynamics is needed to better understand the mechanistic drivers of yield heterogeneity and to improve yield and production efficiency while reducing nutrient pollution and agricultural greenhouse gas emissions. Despite great efforts, current skill and accuracy in estimating subfield crop yield still remains low with only about half of the variation explained (Asseng et al., 2015; Azzari et al., 2017; Burke & Lobell, 2017). Crop monitoring analysis frequently assumes uniform crop development so that nitrogen prescriptions can be based on a single image acquired at a critical date (e.g. V6 for maize). Heterogeneous crop emergence due to variations in soil quality, soil moisture, and other relating factors, and delay in emergence can lead to subsequent

yield decreases (S. Albarenque et al., n.d.). Yield stability zones may also be misleading in optimizing inputs for crop production when the estimated variability of crop biophysical parameters are not accurate due to non-uniform emergence (S. M. Albarenque et al., 2016). These challenges raise two central objective questions: 1) Do remote sensing images provide benefit for in-season tactical management beyond yield stability maps? and 2) Does in-season remote sensing of crop growth dynamics help explain end of season yields? In this paper, we work to assess the ability of remote sensing and yield stability methods to estimate yield zones and clarify the role of in-season heterogeneity in crop growth on yield. Our specific objectives are: 1) Assess the utility of remote sensing for the in-season estimation of yield zones compared to historic yield stability zones. 2) Determine the optimal within-season timing and variance of remote sensing image selection for estimation of yield. 3) Test the impacts and mechanisms of remote sensing vegetative indexes on final yields and yield stability zones.

5.3 Study area and data

We investigated maize yield heterogeneity across four Midwest states, sampling 34 fields from Michigan (MI), 29 fields from Indiana (IN), 35 fields from Iowa (IA), and 17 fields from Minnesota (MN). These fields were selected based on the availability of remote sensing images. All fields were under maize-soybean or maize-soybean-wheat rotations and were analyzed for the maize rotation. These states are all in the Corn Belt region with continental climates typified by cold winters and hot, humid summers (Modified Köppen classification types Df). Average rainfall in these states during the summer months typically range from 7.6 to 12.7 cm per month, which, together with stored soil moisture in the winter and spring, are usually sufficient to meet crop requirements in most years. The maize crop is normally planted from the late April to mid-May when soil conditions are favorable and harvested in October.

PlanetScope imagery: Planet provides high-resolution, real-time images for agricultural monitoring in RGB and near infrared at global coverage. We utilized multi-temporal clear and cloudless PlanetScope time series images from 2018 to 2020 for IN, MI, and MN, while images for IA were obtained only from 2018 to 2019 due to the lack of yield data in 2020. Planet scope images are gap-filled, radiometrically calibrated, atmospheric and orthorectification corrected, and harmonized to sentinel 2, prior to acquisition (Gao & Zhang, 2021). Tile numbers are 19E-190N (Indiana), 22E-197N and 22E-198N (Iowa), 28E-195N (Michigan), and 20E-201N and 20E-202N (Minnesota), with each tile containing 8000×8000 pixels at 3-meter resolution (Figure 15). For each year, we acquired images for the first and 15th day of May-September with an additional image from September 30th to cover most of the maize growing season. We resampled Planet images into 10-meter resolution to match the standard resolution of our environmental data set.

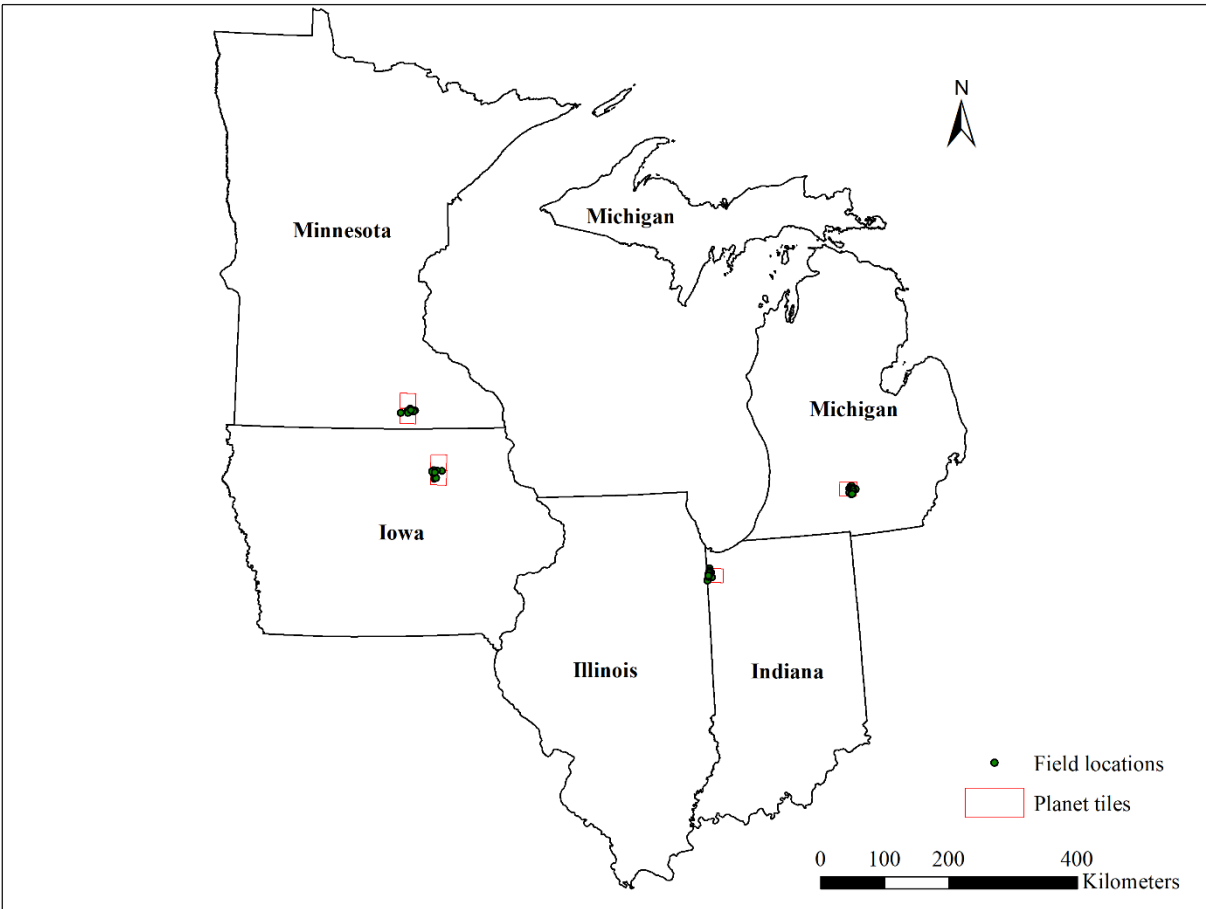


Figure 15 Selected study sites, 2018-2020.

Agronomic and environmental datasets: We built an environmental dataset that included elevation, weather, and soil spatial data. For elevation data, we acquired a 10-meter digital elevation model (DEM) from the USGS National Elevation Dataset (Arundel et al., 2015) and calculated the topographic position index (TPI) for the classification of hillslope position (Weiss, 2001). We characterized the daily average max and min temperature and rainfall for each field with gridMET at ~ 4-km resolution (Abatzoglou, 2013). We characterized soil properties from the Gridded Soil Survey Geographic (gSSURGO) soil dataset at 10-meter resolution with the soil characteristics; soil organic matter (SOM) content, bulk density (BD), soil depth, and soil water availability.

We use yield stability maps to evaluate the yield performance at subfield scale. We obtained 1375 high resolution yield shapefiles based on observed yield monitor data for the study period directly from farmers and processed them into 2-m resolution yield rasters. Specifically, each field was split into high and stable (HS), low and stable (LS), and unstable (UN) areas based on the multi-year average and standard deviation of historical yield for any specific location. The details of the yield stability zone calculation are presented in Chapter 1. The stability maps were also resampled to a standard 10m raster.

5.4 Methods

5.3.1 Yield level prediction using yield stability maps and remote sensing

We compared the accuracy of in-season crop yield zones estimated with remote sensing models with yield stability zones based on historical observations. For each year assessed, we evaluated the potential applicability of historical yield stability in the estimation of current year yield zones. We classified the yield map of each year and field into two discrete high or low yield classes by comparing the pixel's yield to the field average yield. Then we calculated the percentages of both classes in HS and LS zones respectively. The historical UN zone was not included given no assumption of high or low yield in this zone in any specific year.

We used a vegetative index with three modeling approaches to leverage remote sensing imagery to estimate current year yield stability zones. For all models, the estimated yields were translated into high or low yield zones for comparison to the historic yield stability zones approach. In the first modeling approach, we used linear regression (LR) of yield by CGVI using the single optimal image from the season, where the single optimal image was defined as the image with maximum correlation with yield. For the second modeling approach we used all the images in each growing season, progressively and sequentially as inputs to a series of

multiple linear regression (MLR) models of yield. In the last modeling approach, we applied a series of random forest (RF) regression models to estimate yield using the same, sequential image inputs as the second MLR model-based approach. An RF is an ensemble learning method constructed from multiple decision trees. In each tree, a random subset of input features is extracted for making decisions. The estimated yield value is the mean fitted responses from all trees. RF regression was conducted using the R “randomForest” software package (Liaw & Wiener, 2002). We used the square root of the number of input variables to split the data at each node and the number of trees was 500. For all three models, 20 percent of pixels were randomly selected as training data. Then, the trained models were used to estimate yield and yield zones for each field.

5.3.2 *Optimal image selection for yield estimation using clustering analysis*

We tested the hypothesis that the optimal image date differs between the field and the subfield scale. At the field level, we computed the correlation between GCVI and crop yield to obtain each field’s optimal image. To assess the sub field scale, we performed a cluster analysis on the GCVI curve for each pixel in one growing season with the time-series Kmeans (TK-means) method using the tslearn python package (Tavenard et al., 2020). This time-series clustering was carried out once for each field-year combination. Similar to the K-means algorithm, the TK-means evaluates the weighted distance between m time series

$Z = \{z_1, z_2, z_3, \dots, z_m\}$ and the centroids of k clusters $C = \{c_1, c_2, c_3, \dots, c_k\}$, where weights are assigned to time series values according to the importance of a time span in the clustering process. Details can be found in Huang et al. (2016). We used soft dynamic time warping as the distance metric suggested by Janati et al. (2020) to avoid the susceptibility of Euclidean distance to time shifts in the GCVI curves expected in this data set. We used 8 as the cluster number and

did not optimize the segmentation as the objective of clustering was to evaluate if there was variation in yield estimation accuracy between clusters. Similar to field level analysis, we obtained the optimal image for each cluster in each field. To evaluate whether the patch (field or subfield) level optimal date or the field itself influenced the yield prediction, we conducted yield estimates for LR models by each optimal image, the field, and the cluster in each field.

5.3.3 *Short duration crop growth dynamics impacts on yield*

We evaluated the change in yield zone classification between image collection periods to assess what information might be gained from early season remotelysensed growth dynamics for subsequent in-season management. Each GCVI image was classified as high (H) or low (L) based on the field's average GCVI value. Then the change between successive classified GCVI images was determined by associating labels from both dates, producing four temporal classes (HH, HL, LH, and LL). For example, a pixel was labelled as high-high (HH) on July 15th if it was classified as high on July 1st and July 15th images. HL/LH indicated areas with changes, and HH/LL indicated areas with no change. We computed the state average percentage of change vs. non change over time and across historical yield stability zones to test if short term dynamics in crop growth related to crop yield outcomes.

Finally, we explored the physical mechanisms of short-term change in GCVI between periods using a linear mixed effect model. We used the absolute difference in the GCVI values between successive images as the rate of change (ROC). Using a continuous variable for ROC allowed us to evaluate linear trends with environmental predictors not available with the categorical temporal classes. To test the hypothesis that crop growth differs among soil, topography, and weather factors, we fit the following model:

$$ROC = SW + Rain + SD + BD + \alpha_{LP} + \mu_y + \mu_{y(F)} \dots \dots \dots \text{Equation (22)}$$

Where SW, SD, and BD indicate soil water availability ((upper limit – lower limit)×soil depth), soil depth, and bulk density, Rain is the cumulative precipitation amount for the period of ROC calculation, α_{LP} is a categorical factor showing the pixel's location in the landscape (hilltop, depression, and other). μ_y is the random effect of the year, and $\mu_{y(F)}$ is the random effect of the field within year.

5.5 Results

In this section, we first present the correlations between historical yield stability zones and in-season remote sensing models with end of season yield zones and their potential to inform subsequent in-season tactical management decisions. Second, we identify the variance in the optimal image date. Third, we investigate the environmental variables driving the short term dynamics through the observed rates of changes of the GCVI.

5.4.1 Accuracy of yield level prediction using yield stability maps and in-season remote sensing

For each state, yield monitor data validated the performance of the yield level prediction by remote sensing models and historic yield stability zones (Figure 16). The accuracy was calculated as fraction of pixels that were correctly classified. Across all four states historical yield stability zones outperformed the LR and MLR model approaches but was outperformed by the RF. Under optimum conditions, using post-hoc optimal images for LR and the full image set for MLR and RF, the LR, MLR, RF and YS correctly estimated 58%, 66%, 86%, and 76% of observed yield levels respectively in the HS and LS zones. The YS estimate performance was equivalent to the RF on images from June to August and accuracy was consistent across the four states in any given year.

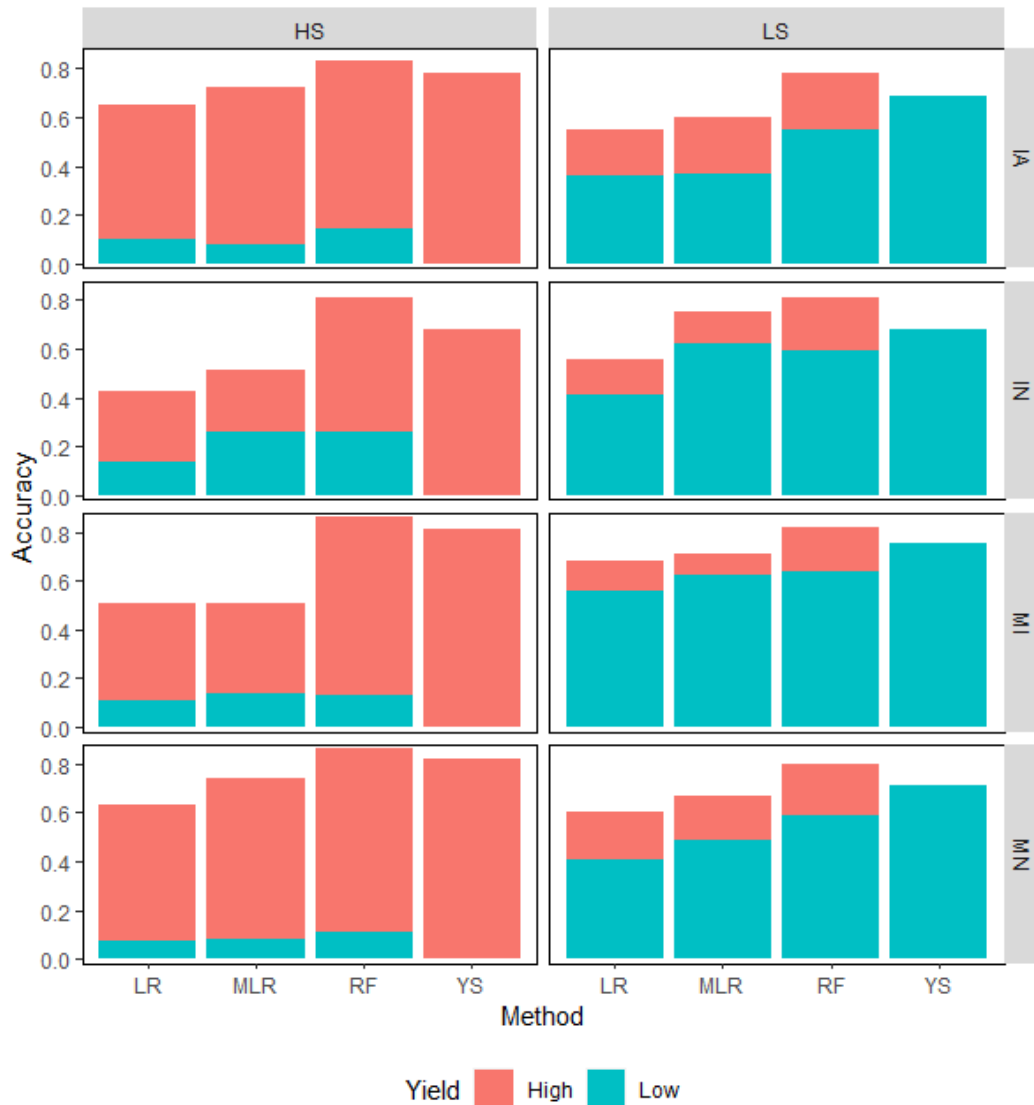


Figure 16 Accuracies on the maize yield level prediction by linear regression (LR), multi-linear regression (MLR), random forest (RF), and yield stability map (YS) for each state in the stable high (HS) and stable low (LS) zones.

Iteratively including images across a season improved both the MLR and RF models prediction accuracy (Figure 17). The MLR model prediction accuracy increased dramatically, 6%, with the addition of the Aug 1st image, but only increased moderately, 1% to 3%, with the addition of other images. The accuracy of the RF model increased rapidly through June, then gradually increased with every additional image until September.

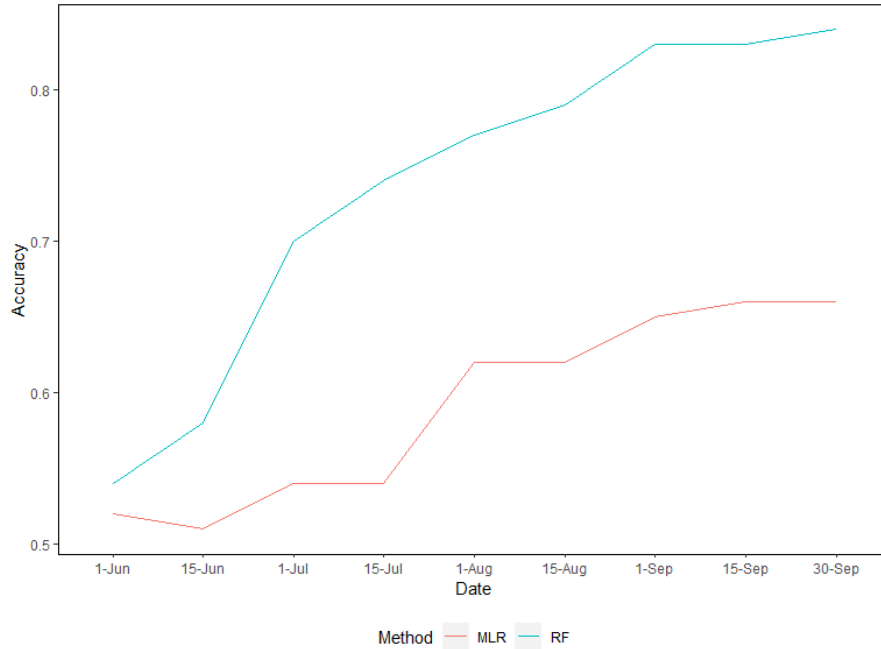


Figure 17 Accuracy assessment for maize yield level prediction models dates when image is added in acquisition time order. This assessment included all the fields in the study area.

5.4.2 Variance in optimal imagery dates and impacts on yield prediction

The optimal date of images acquired for LR varied both between fields and within fields. At the between fields, optimal image acquisition date range was large, spanning Mid-July to early-September and within each field there was an additional one month range from the field scale optimal image. The standard deviation of the optimal date for field from the average of the total data set was 3.8 weeks while the standard deviation of subfield optimal images to the field average was 2.4 weeks (Figure 18). The largest proportion of fields (29%) had an optimal image from 8/1, two weeks before the optimal date of the full dataset. Nevertheless, optimal image for 79% of fields were between 7/15 and 8/15, and on 9/1 for 13% of the fields. At the subfield scale, 35% of subfield areas had the same optimal image date as the field.

Regressions done by each optimal image produced similar accuracy of 60% compared to using one optimal image for all the fields. However, regression modelled by field and cluster

produced a much higher accuracy at 77% and 80% respectively. This indicated that changing the optimal image does not impact the final predicted yield and clustering fields into several parts did not lead to significant improvement, although understanding differences among fields appears to be a key to improve this objective.

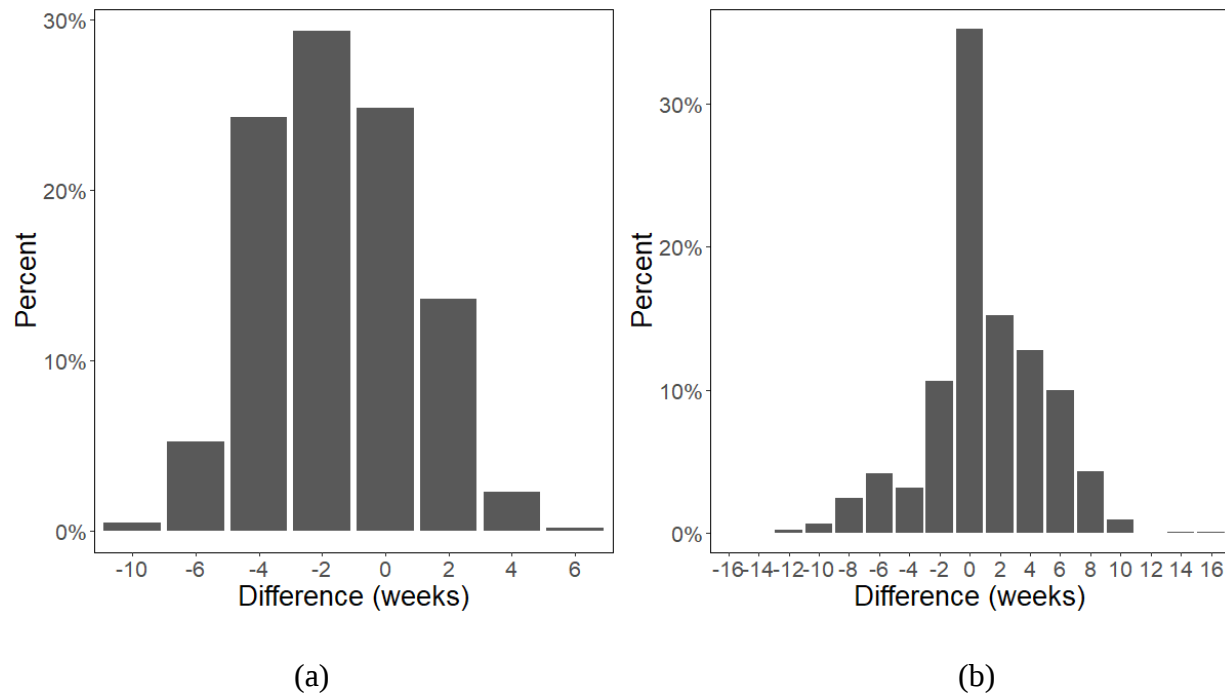


Figure 18 Weekly difference between the field level optimal image and that determined by the whole dataset (a) and between the field and subfield level optimal image (b).

5.4.3 Rate of change and related environmental factors

Most of the field areas were classified as ‘Non Change’ (HH or LL) with little change during a single period. Changes in the spatial pattern of GCVI were more common at the beginnings and ends of the season. The first month of crop growth was associated with the largest percentage of changing areas, with 29% from 6/1 to 6/15 and 31% from 6/15 to 7/1 (Figure 19). Fields became more stable in the mid-season (July and August) and the percentage

of area labelled as HL or LH decreased to 19.6%. In September, this percentage remained low in most periods but increased to around 25% for 9/15 to 9/30 in MI and 9/1 to 9/15 in MN.

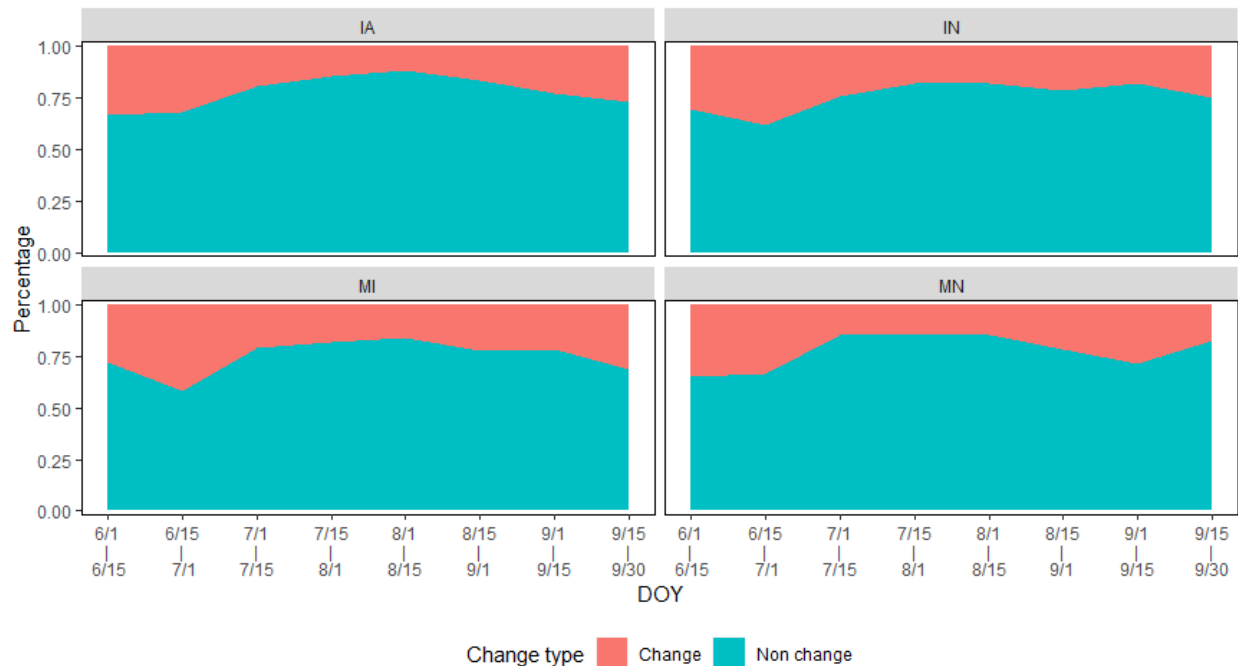
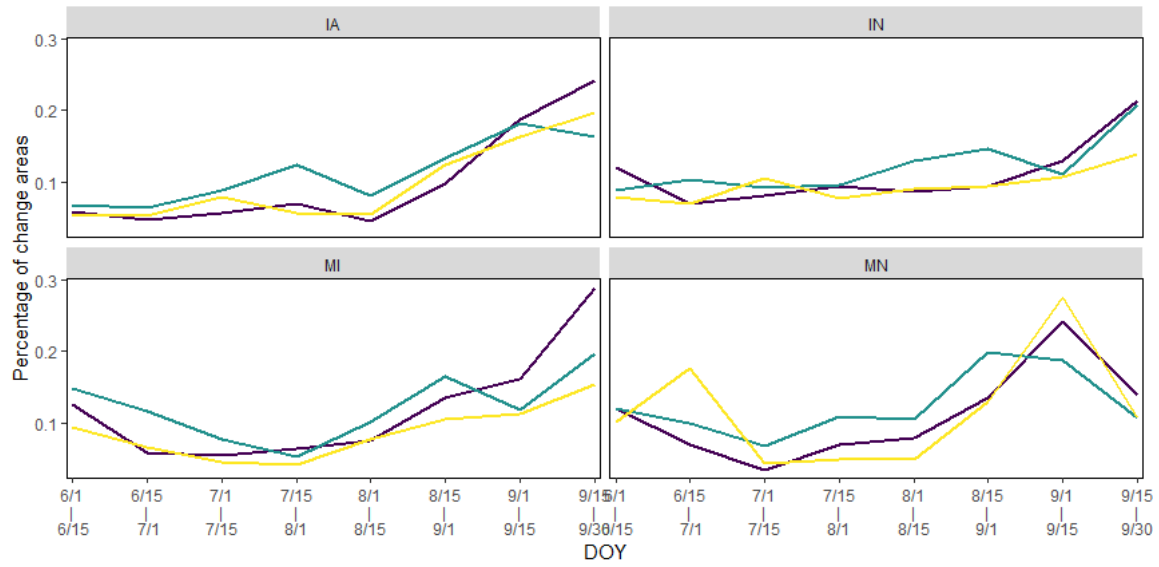
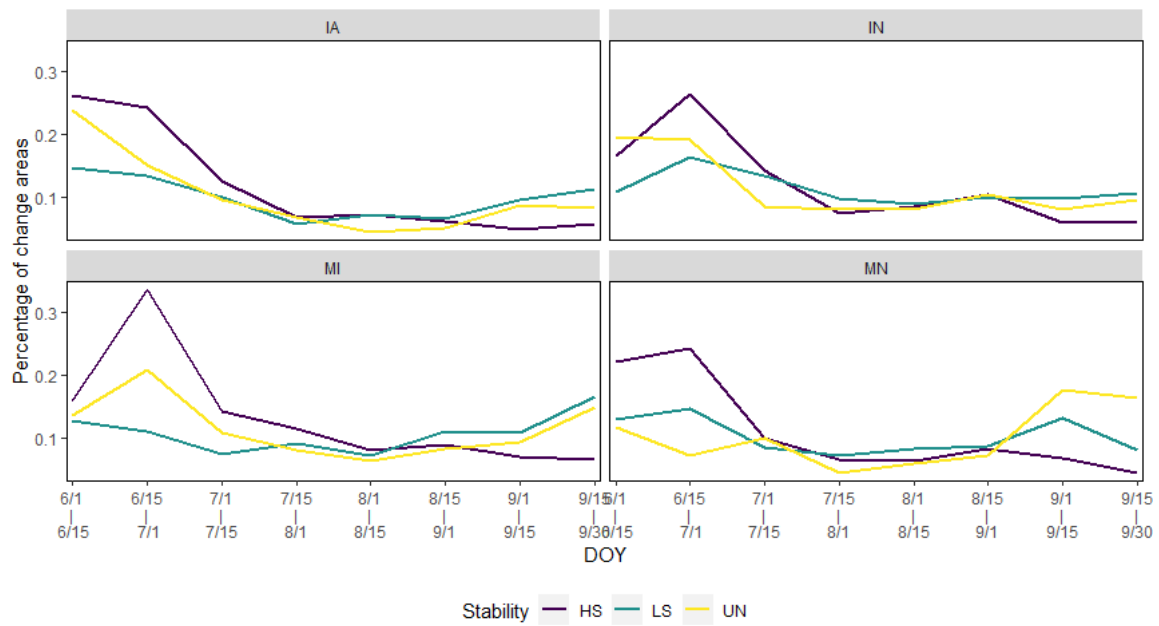


Figure 19 Percentage of ‘Change’ and ‘Non change’ classes during each successive images for each state.

Individual periods did not provide explanatory value for end of the season yields. The majority of the transient HL and LH values stabilized by the end of the season (Figure 20). We did not observe significant differences among stability zones, except from 6/15 to 7/1 in MI and MN. In the high yielding area, the percentage of HL increased from less than 10% in the mid-season to 18.6% at the end of the growing season, but the decrease in GCVI did not correlate to reduced yields. In contrast, in the low yielding area, the largest percentage of the LH occurred during the beginning of the growing season. On average, 20% of the ultimately low yielding areas experienced good crop growth during the early period but still produced low yields. In general, trends in ROC dynamics were consistent across yield stability zones and any single period did not provide predictive power for end of season yields.



(a)



(b)

Figure 20 Percentage of HL class in the high yielding area (a) and LH class in the low yielding area (b) by the stability zone for each state.

Environmental factor models of ROC dynamics and the isolated partial effects help illuminate the mechanisms driving early and late season crop growth dynamics. Models for June

and September explained 94% and 86% of the variance in the ROC respectively, demonstrating that these variables can be used to understand crop growth variations in the field. In June, soils with higher water capacity and rainfall were observed to be associated with low increases in the GCVI, whereas these same environmental factors were also associated with higher GCVI in September (Figure 21). High bulk density soils (with low soil porosity and soil compaction) and shallow soils also resulted in low GCVI increases in June, but these impacts were weaker than those associated with water capacity and rainfall. The impacts of these two variables on ROC were also small in September when maize plants have accumulated most of their dry weight in the kernels.

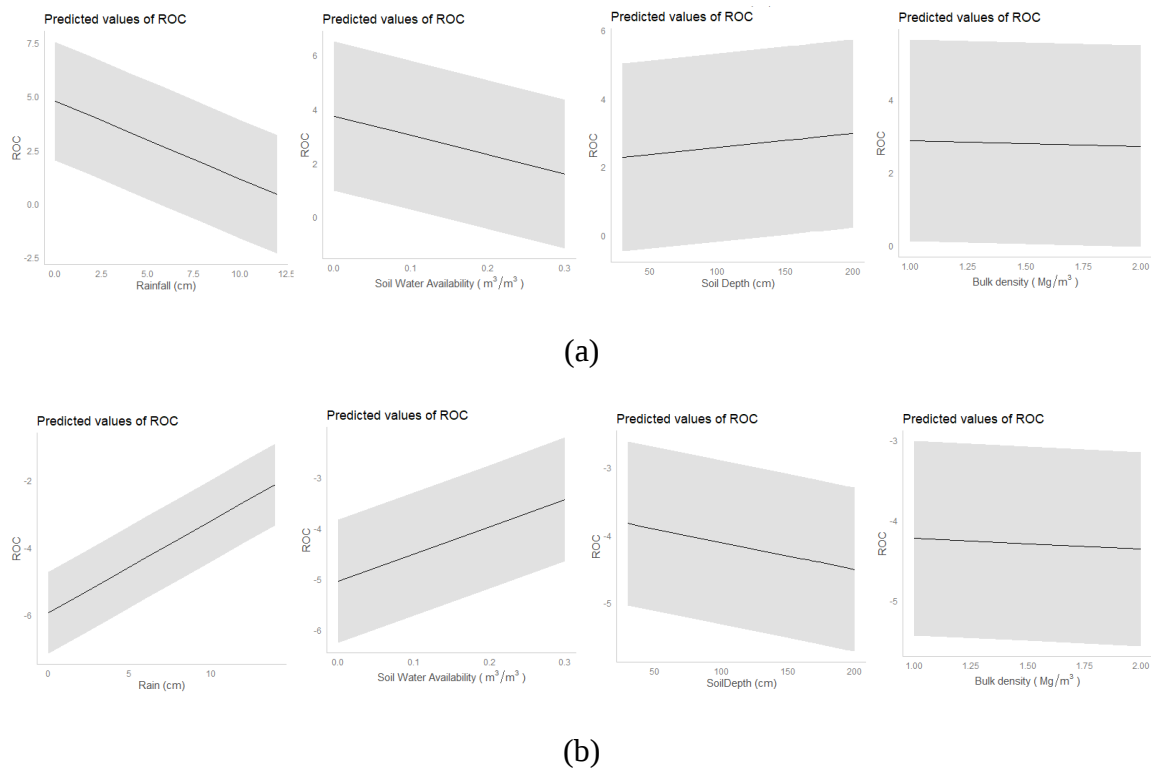


Figure 21 The relationships between environmental factors and the rate of change in June (a) and September(b). Gray areas are the standard errors.

5.6 Discussion

Overall, this study finds that integration of yield stability map and remote sensing is needed for in-season assessment of yield levels and tactical management. Our result concurs with Johnson (2014) that peak NDVI which occurs in the first week of August best correlates with end of season corn yield. The timing of peak NDVI (or GCVI) differs among and within fields due to various crop emergence date or growth conditions. The large range in optimal dates suggests historical best image date or sufficient image series should be established for each field (Figure 18). Even with optimal images selected by field, remote sensing is challenged to increase yield estimation accuracy above 50% at the sub field scale (Burke & Lobell, 2017). These images, when including availability, acquisition, and processing may be too late into the season to enable use in tactical management. Strategic management, particularly of the stable zones (i.e. variable rate Nitrogen application (Basso et al 2019), cultivar selection, land retirement, adaptive grazing (Nunes et al 2019), etc.) provides substantial information to inform spatially explicit baseline management. Uncertainty in the historic yield stability zones captured in unstable zones present the opportunity for powerful synergy with remote sensing integration, particularly when informed or augmented by physically-based model output (Jefferies et al 2020, Shuai & Basso, 2022). As a result, tactical management of unstable zones should be the primary focus of remote sensing integration.

Inertia of the historic yield zones, and the limitation of short duration crop growth dynamics, suggest heterogeneity in growth dynamics alone is an unreliable approach for yield estimation (Figure 20). Early changes in growth may be associated with either strong growth on soil with limited water or nutrient holding capacity that leads crops to ultimately struggle after early success. Alternatively, the images may simply be capturing non-crop, weedy plant growth.

The high yielding areas where there is a large percentage of HL break from the expectation of the green leaf area duration from anthesis to harvest positively correlating with final grain dry weight (Wolfe et al., 1988). The ultimate, mechanistic drivers of final yield at subfield scale are primarily related to water stress (Martinez-Feria & Basso, 2020).

The limitation of in-season remote sensing models of Maize yield zones and the value of historic yield stability zones open several avenues of future research. To improve management, automated integration of yield stability zones, remote sensing models and physically based models to estimate yield and nitrogen use efficiency at the subfield scale is needed. The data sets needed for this application, if tasked with learning over time, will provide the opportunity to isolate variables and mechanisms that explain yield variance between and within fields. The challenge of adapting these tools, here in development for relatively simple cropping systems, to diverse and highly integrated systems is needed to reduce environmental impact while sustaining exceptional yields (Hatfield et al., 2020). Learning from simple systems across a continuum of soil and climatic conditions may enable generalizable strategic and tactical approaches to precision management across the agroecological system.

5.7 Conclusion

This study evaluated in-season and real time methods to understand spatial crop yield zone variance. We found that: 1) Historic yield stability maps provide accurate estimation of corn yield in stable zones, which could be used to support fertilizer management. 2) Use of field and sub-field cluster specific models improved remote sensing estimations. 3) the remote sensing plant growth dynamics of any given month, while more variable at the beginning and the end of the season, did not have clear impacts on final yield or yield zones. Integrated methods including

historical yield zones, in-season remote sensing, and physical modeling are needed to estimate in-season yield zones to inform tactical management.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATION

The purpose of this dissertation is to advance the understanding of subfield yield variability over time and space using yield monitor datasets, process-based crop simulation models, satellite images, and other geographic information (Chapter 2 - 5). The following chapters investigate the potential use of yield stability maps in estimating nitrogen balance at subfield scale and related economic and environmental losses, providing land for biofuel crop production without additional land use expansion, estimating subfield maize yield combined with remote sensing, and providing prior knowledge about yield level without in-season information for maize/soybean fields in the U.S. Midwest.

Chapter 1 stated the background of many aspects of digital agriculture and that gap between the theoretical and actual use of precision technology and the large datasets is due to the lack of appropriate data interpretation. It clarified the importance of yield stability analysis and site-specific management strategies based on analysis. A description of the yield stability calculation process and a simple introduction of the SALUS model was also included and used in the following chapters.

Chapter 2 utilized the advantage of remote sensing images (extensive spatial coverage and long-term availability) to perform yield stability analysis for over 30 million ha of fields in the U.S. Corn Belt. The free access to satellite images on the Google Earth Engine allows government, researchers, and even farmers to conduct this analysis. They could also use the USDA reported nitrogen application rates, or the rates that are applied by farmers, to determine potential nitrogen losses due to inappropriate fertilizer application strategies and estimate the subsequent environmental consequences given these nitrogen losses.

Chapter 3 presented a new solution to manage the stable low-yielding subfield area with biofuel crops, switchgrass. Despite decades of research on the benefits of switchgrass, the adoption rate is still low. The LS area were common contiguous in fields, so it would be very convenient to upscale the biofuel production if farmers could plant it in their own fields with precision technology. Our results indicated that no-till was the best strategy to save carbon in the soil. With less nitrogen fertilization, switchgrass produces similar amounts of biofuel to offset GHG in the atmosphere. Planting rye cover crops could efficiently remove soil nitrate to protect water quality. The change in the land use composition in the landscape created additional habitat for wildlife, thus showing positive effects on the ecological aspects. This proposed management is proposed based on the strategic use of yield stability maps for long-term sustainable production at both the field and national level.

Chapters 4 and 5 mainly focus on the short-term tactical use of yield stability maps in maize growth monitoring and yield prediction. In Chapter 4, I developed a new methodology to integrate remote sensing and crop modeling to enhance maize yield prediction at the subfield scale. Water stress varies in fields by historical yield, weather, and landscape position. Thus, I used SALUS to simulate cumulative water stress using yield stability maps as the spatial unit. A good relationship between simulated water stress and crop yield and improved accuracy in estimated maize yield demonstrated the effectiveness of this method. This work provided a brand-new framework for combining crop models, remote sensing, and agronomic and environmental datasets for precise yield forecasts, which could give accurate information even two months before the crop harvest.

Chapter 5 assessed the ability of yield stability maps to estimate maize yield level in stable areas before the growing season. The accuracy of this estimation is similar to those derived from models trained on high-resolution PlanetScope images. Since historical yield is often used as the

proxy to determine the nitrogen rate, the possibility of inappropriate nitrogen application could be lower if farmers follow the information from yield stability maps. Another important topic I investigated in this chapter was to enhance crop monitoring by interpreting temporal changes on the spatial variation of crop conditions. I concluded that one or two images were not enough to fulfill this work since crop growth is a dynamic process, while temporal inspection of images was necessary to get in-depth knowledge of crop growth processes and the related final yield.

The digitalization of agricultural system is under rapid development and farmers are now, more than ever before, able to afford the latest technology, such as UAV, which increasingly points up the value of comprehensive knowledge about subfield variability both in seasonal crop conditions and final crop yield. Sensor observations should be integrated with process-based models as powerful tools to explore the underlying physical, chemical, and biological factors of crop status and offer reasonable management toward the requirements of sustainable production, as has been shown in this research project.

APPENDICES

APPENDIX A: CHAPTER 2 Supplementary Information

Table 15 Average state-level N fertilizer rates and Coefficient of Variation (CV, %) as reported by ARMS (NASS 2016) and University-recommended rates based on statewide Maximum Return to Nitrogen (MRTN) databases (Sawyer et al., 2006). All values are kg N ha⁻¹ y⁻¹.

State	ARMS $\pm$ CV (%)	MRTN
IL	183 $\pm$ 2%	191
IN	175 $\pm$ 4%	209
IA	158 $\pm$ 4%	173
MI	151 $\pm$ 5%	165
MN	160 $\pm$ 4%	177
MO	197 $\pm$ 4%	217
ND	143 $\pm$ 4%	158
OH	174 $\pm$ 11%	195
SD	138 $\pm$ 3%	152
WI	117 $\pm$ 5%	129

*https://www.nass.usda.gov/Quick_Stats

Table 16 Average N rate, N fertilizer surplus and percent contribution from each yield stability classes.

State	Average Fertilizer N Rate	Average Surplus N			Average Contribution to Surplus N		
		Stable High	Stable Low	Unstable	Stable High	Stable Low	Unstable
IL	204	59	96	73	26	42	32
IN	210	75	111	88	27	41	32
IA	180	32	65	42	23	47	30
MI	171	40	76	53	24	45	31
MN	183	38	68	48	25	44	31
MO	225	107	140	120	29	38	33
ND	164	53	81	62	27	41	32
OH	195	59	95	73	26	42	32
SD	158	40	69	49	25	44	31
WI	133	12	33	17	19	53	27
Total Average	182	52	83	63	25	44	31

Values are for maize in stable high yield, stable low yield, and unstable yield areas by US state. All values are kg N ha⁻¹ y⁻¹ except percentage (%) of average surplus N contributed from each area.

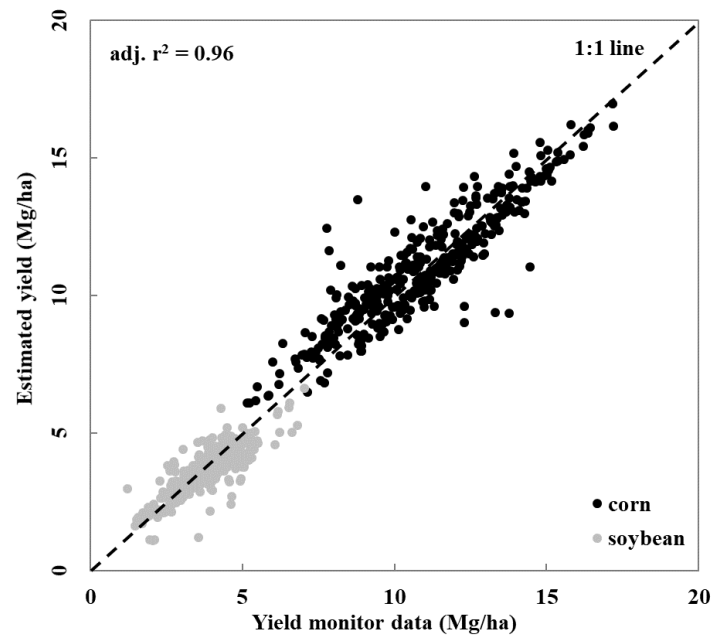


Figure 22 Estimated yield using NDVI-stability class approach vs yield monitor data.

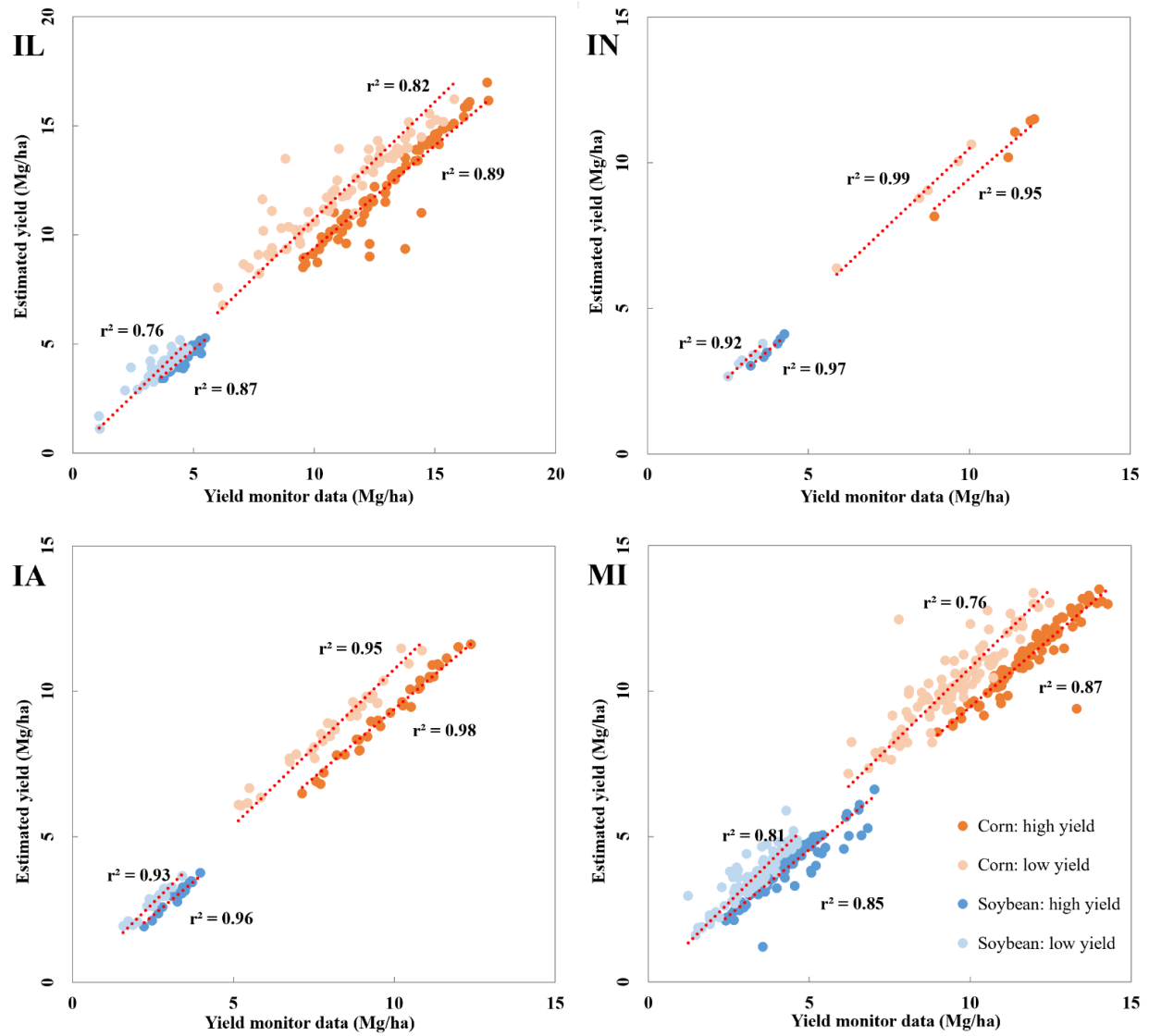


Figure 23 Estimated yield vs. yield monitor data for corn (orange) and soybeans (blue) for each state.

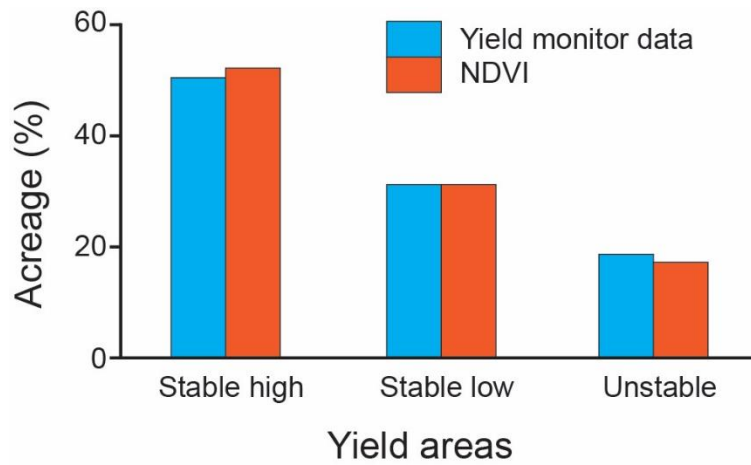


Figure 24 The acreage percentages of stable high yield, stable low yield, and unstable yield areas of Iowa calculated from yield monitor combine harvesters (left-hand bars) or NDVI data (right-hand bars).

Yield monitor-based

Landsat-based

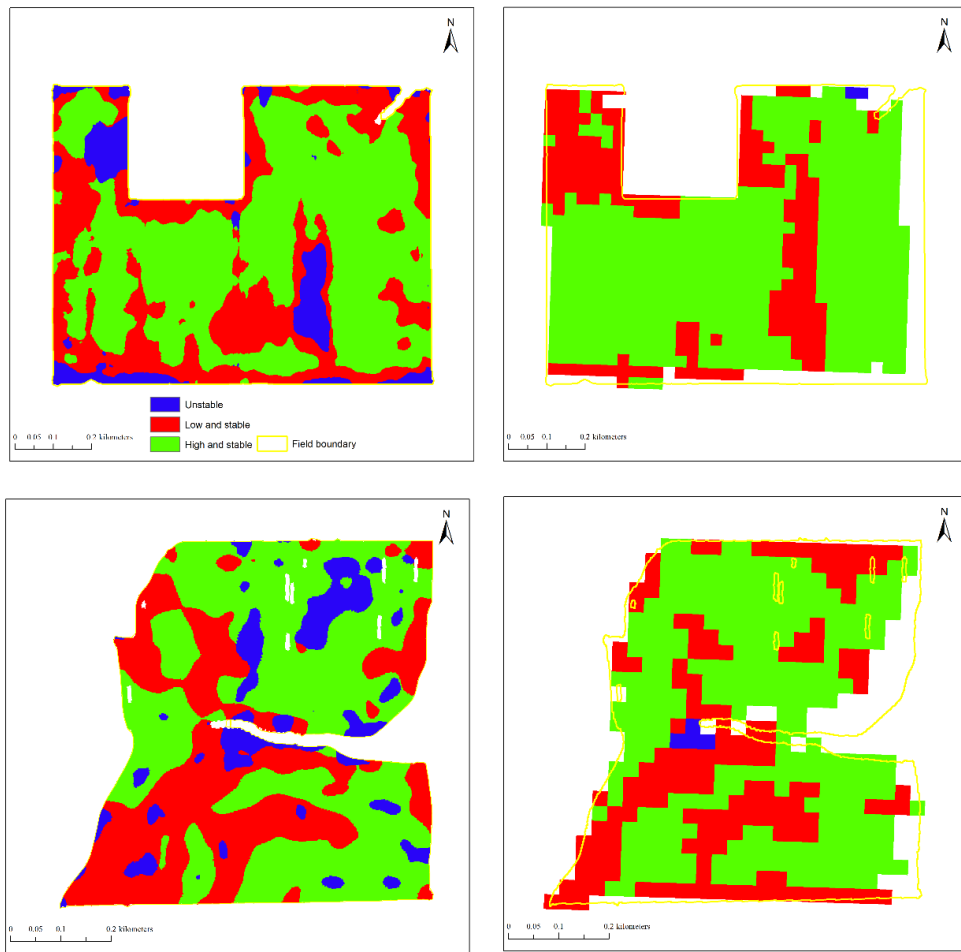


Figure 25 Yield stability class areas derived from combine-based yield monitor data (left-hand panels) and LANDSAT NDVI data (right-hand panels) for two representative Iowa fields (top and bottom rows). The LANDSAT NDVI spatial resolution is 30 m; yield monitor resolution is 2 m.

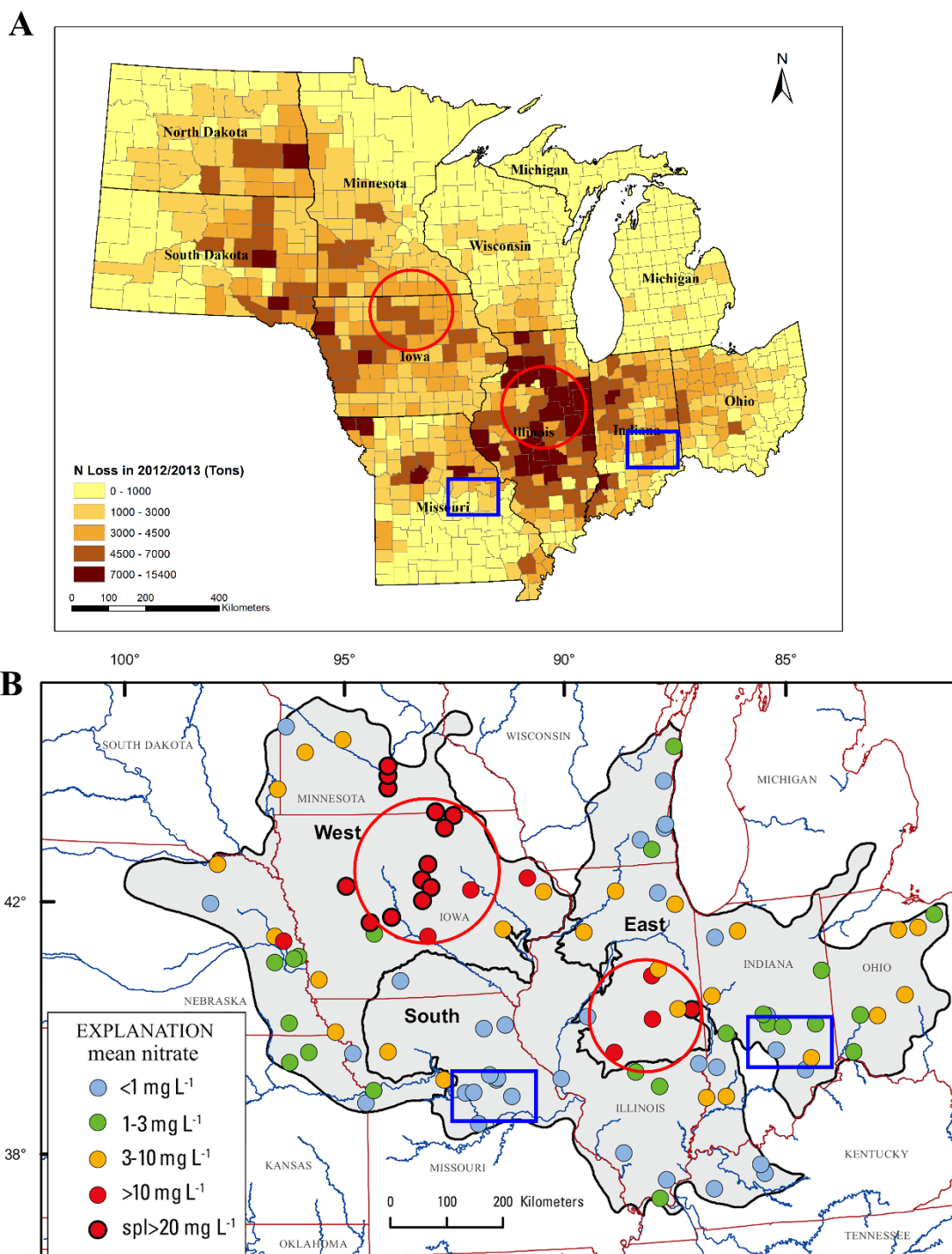


Figure 26 (A) County-level N loss in 2012-13 growing season and (B) mean 2013 nitrate concentration at the 100 Midwest Stream Quality Assessment sites from Van Metre et al. (2016). The red circles denote areas with concurrent higher N loss and nitrate concentrations. The blue rectangles are areas with lower N loss and nitrate concentration.

APPENDIX B: CHAPTER 3 Supplementary Information

Table 17 Cultivar information for RY16 species

Parameter	Value	Parameter	Value	Parameter	Value
CHeight	1	PlntN_Hf	0.0180	TTtoMatr	1600
EmgInter	15	PlntN_Mt	0.0140	TbaseDev	0
EmgSlope	6	PlntP_Em	0.0040	ToptDev	15
GrnN_Mt	0.0230	PlntP_Hf	0.0030	relLAI_P1	0.01
GrnP_Mt	0.0037	PlntP_Mt	0.0024	relLAI_P2	0.95
HrvIndex	0.42	RUEmax	1.6	relTT_P1	0.30
LAImax	3	SnParLAI	1	relTT_P2	0.5
Nfixing	N	SnParRUE	1	relTT_Sn	0.8
PhotoSynID	C3	Species_Name	Rye		
PlntN_Em	0.0226	TTtoGerm	20		

CHeight: Approx. Height of Crop (m); EmgInt: Intercept of emergence time calculation; EmgSlp: Slope of emergence time calculation; GrnN_Mt: Optimal N in grain at maturity (g/g); GrnP_Mt: Optimal Phosphorus in grain at maturity (g/g); HrvIndex: Harvest index (Mg/M); LAImax: maximum potential leaf area index (m²/m²); PhotoSynID: Photosynthesis Type; PlntN_Em, PlntN_Hf, PlntN_Mt: Optimal N in plant at emergence, halfway to maturity, maturity (g/g); PlntP indicates optimal phosphorus in the plant (g/g); RUEmax: Maximum potential radiation use efficiency (g/MJ); SnParLAI: Leaf area index decline rate parameter after senescence starts; SnParRUE: Radiation use efficiency decline rate parameter after senescence starts; TTtoGerm, TTtoMatr: Development time to germinate, mature (degree day); TbaseDev, ToptDev: Minimum and optimal temperature for plant development (degree C); relLAI_P1, relLAI_P2: Relative leaf area index at first and second point in the growing season ((deg. day)/(deg. day)); relTT_P1, relTT_P2: Relative development time at first and second point in the growing season ((deg. day)/(deg. day)); relTT_Sn: Fraction of growing season when leaf area starts declining ((deg. day)/(deg. day)).

Table 18 Cultivar information for SW_UL species

Parameter	Value	Parameter	Value	Parameter	Value
CHHeight	1.5	PlntN_Mt	0.0075	Species_Name	Switchgrass Upland
EmgInter	15	PlntP_Em	0.0140	TTtoGerm	20
EmgSlope	6	PlntP_Hf	0.001	TTtoMatr	1200
GrnN_Mt	0	PlntP_Mt	0.0007	TbaseDev	10
GrnP_Mt	0.0022	RUEmax	3.5	ToptDev	25
HrvIndex	0	RWPC1	0.6	relLAI_P1	0.1
LAImax	7	RWPC2	0.26	relLAI_P2	0.95
Nfixing	N	RootPartFac	1	relTT_P1	0.15
PhotoSynID	C4	RootSenFac	0.0024	relTT_P2	0.6
PlntN_Em	0.01	SnParLAI	1	relTT_Sn	0.7
PlntN_Hf	0.09	SnParRUE	1		

RWPC1, RWPC2: Parameters in linear partitioning of biomass; RootPartFac, RootSenFac: Root Partition and Senescence Factor, A simple multiplier to change the amount of daily biomass that goes to the roots.

Table 19 Values used in the estimation of net emissions associated with nitrogen fertilizer additions. Values of RS, N_AG, N_BG, R for maize, soybean, and switchgrass were from Tables 11.a in the IPCC, 2019.

Category	Name	Description	Unit	Value	Source
IPCC Tier 1 N ₂ O emissions	N ₂ O _{CO₂eq}	GWP of N ₂ O	kg CO ₂ eq / kg N ₂ O	256	IPCC, 2019
	EF1	Emissions factor	kg CO ₂ eq / kg N fertilizer	0.01	
	RS	ratio of below-ground root biomass to above-ground shoot biomass for crop	kg dm/kg dm		
	N_AG	N content of aboveground residue	kg N / kg dm		
	N_BG	N content of aboveground residue	kg N / kg dm		
	R	C:N of soil organic matter	kg C / kg N	11	
	Frac_GASF	fraction of synthetic fertilizer N that volatilizes as NH ₃ and NO _x	kg N / kg N fertilizer	0.11	
	Frac_LEACH	fraction of all N added to/mineralized in managed soils in regions where leaching/runoff occurs that is lost through leaching and runoff, kg N (kg of N additions)	kg N / kg N fertilizer	0.24	
	EF4	emission factor for volatilized N ₂ O emissions from atmospheric deposition of N on soils and water surfaces	kg CO ₂ eq / kg N fertilizer	0.01	
	EF5	emission factor for leached N ₂ O emissions from atmospheric deposition of N on soils and water surfaces	kg CO ₂ eq / kg N fertilizer	0.011	

Table 20 State-level stability area, field edge area, LS area, and LS edge area. All values are in ha except the last column as percentage.

State	Stability area	Field edge area	LS area	LS_{edge} area	Fraction of LS in field edge
AR	139,474	36,644	24,609	9,787	40
IL	6,503,184	1,615,547	2,043,111	666,347	33
IN	3,140,567	784,208	796,340	245,424	31
IA	7,492,762	1,806,249	2,330,849	771,332	33
KS	869,252	171,131	253,078	76,022	30
MI	2,053,029	602,943	506,268	189,038	37
MN	3,886,614	885,464	993,808	313,935	32
MO	1,414,235	359,689	409,367	135,224	33
NE	4,470,153	803,805	1,396,031	442,466	32
ND	703,671	102,007	94,624	22,187	23
OH	1,820,514	526,720	456,333	161,956	36
PA	197,357	67,367	67,856	27,100	40
SD	2,082,771	384,037	414,243	123,757	30
WI	867,306	328,300	268,363	125,737	47
Sum	35,640,889	8,474,111	10,054,880	3,310,302	33

Table 21 Proportion of stability class within each buffer area. Stability area value is in ha and HS, LS, UN columns are the percentages (%).

State	distance	Stability area	HS	LS	UN	State	distance	Stability area	HS	LS	UN
AR	0 - 30	36,644	18	27	55	NE	0 - 30	803,805	41	55	4
	30 - 60	28,614	32	18	50		30 - 60	867,941	60	37	3
	60 -90	22,734	37	14	48		60 -90	715,342	70	27	3
	90 -120	17,075	39	13	48		90 -120	569,310	74	24	2
	120 - FC	34,407	39	12	49		120 - FC	1,513,755	78	20	2
IL	0 - 30	1,615,547	42	41	17	ND	0 - 30	102,007	42	22	36
	30 - 60	1,484,472	52	30	18		30 - 60	122,389	48	16	36
	60 -90	1,178,262	55	27	18		60 -90	113,072	52	13	35
	90 -120	813,455	54	27	19		90 -120	91,168	54	11	35
	120 - FC	1,411,450	51	29	20		120 - FC	275,035	54	10	36
IN	0 - 30	784,208	41	31	28	OH	0 - 30	526,720	36	31	33
	30 - 60	765,300	45	27	28		30 - 60	465,909	42	27	31
	60 -90	607,928	48	23	29		60 -90	346,879	47	22	31
	90 -120	395,771	48	22	30		90 -120	211,085	47	20	33
	120 - FC	587,360	47	21	32		120 - FC	269,922	47	20	33
IA	0 - 30	1,806,249	46	43	11	PA	0 - 30	67,367	48	40	12
	30 - 60	1,640,431	57	30	13		30 - 60	64,180	55	34	11
	60 -90	1,345,728	59	27	14		60 -90	36,063	58	30	12
	90 -120	950,185	58	27	15		90 -120	17,194	58	29	13
	120 - FC	1,750,169	56	26	18		120 - FC	12,552	57	29	14

Table 19 (cont'd)

	0 - 30	171,131	42	44	14		0 - 30	384,037	43	32	25
	30 - 60	213,879	53	34	13		30 - 60	417,433	51	23	26
KS	60 -90	155,205	60	27	13	SD	60 -90	354,967	57	17	26
	90 -120	103,673	64	24	12		90 -120	276,962	58	15	27
	120 - FC	225,365	74	17	9		120 - FC	649,372	57	14	29
MI	0 - 30	602,943	39	31	30	WI	0 - 30	328,300	49	38	13
	30 - 60	545,120	44	25	31		30 - 60	241,437	55	32	13
	60 -90	423,522	46	22	32		60 -90	146,532	62	24	14
	90 -120	232,000	46	20	34		90 -120	73,993	64	21	15
	120 - FC	249,443	45	18	37		120 - FC	77,044	64	19	17
MN	0 - 30	885,464	45	35	20	Sum	0 - 30	8,474,111	42	39	19
	30 - 60	859,300	51	30	19		30 - 60	8,083,431	51	30	19
	60 -90	662,556	60	21	19		60 -90	6,386,427	56	25	19
	90 -120	488,974	61	19	20		90 -120	4,412,481	57	23	20
	120 - FC	990,320	61	19	20		120 - FC	8,284,439	59	22	19
MO	0 - 30	359,689	37	37	26						
	30 - 60	367,026	44	31	25						
	60 -90	277,637	49	25	26						
	90 -120	171,636	49	23	28						
	120 - FC	238,245	44	22	34						

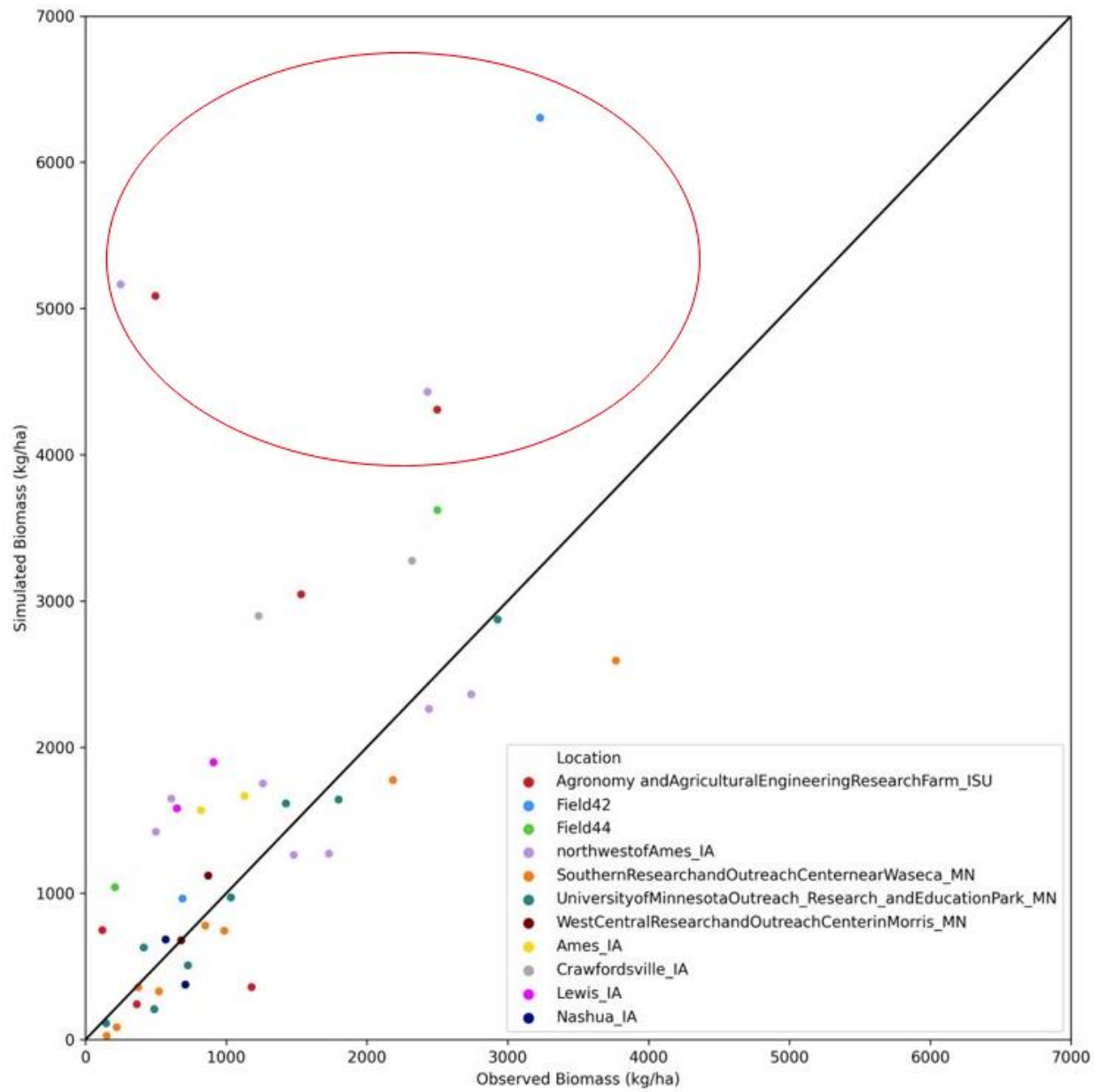


Figure 27 Validation of simulated rye biomass vs. biomass from literature. Red circle includes some outliers.

REFERENCES

REFERENCES

- Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology*, 33(1), 121–131.
- Afrin, S., Latif, A., Banu, N. M. A., Kabir, M. M. M., Haque, S. S., Ahmed, M. M. E., Tonu, N. N., & Ali, M. P. (2017). Intercropping empower reduces insect pests and increases biodiversity in agro-ecosystem. *Agricultural Sciences*, 8(10), 1120.
- Albarenque, S., Davidson, O., & Basso, B. (n.d.). Spatio-Temporal Variability of Maize Emergence Effects on Crop Yield and Growth. *ASA, CSSA and SSSA International Annual Meetings (2020)| VIRTUAL*.
- Albarenque, S. M., Basso, B., Caviglia, O. P., & Melchiori, R. J. M. (2016). Spatio-temporal nitrogen fertilizer response in maize: Field study and modeling approach. *Agronomy Journal*, 108(5), 2110–2122.
- Angulo, C., Rötter, R., Lock, R., Enders, A., Fronzek, S., & Ewert, F. (2013). Implication of crop model calibration strategies for assessing regional impacts of climate change in Europe. *Agricultural and Forest Meteorology*, 170, 32–46.
- Arundel, S. T., Phillips, L. A., Lowe, A. J., Bobinmyer, J., Mantey, K. S., Dunn, C. A., Constance, E. W., & Usery, E. L. (2015). Preparing The National Map for the 3D Elevation Program—products, process and research. *Cartography and Geographic Information Science*, 42(sup1), 40–53.
- Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., & White, J. W. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2), 143–147.
- Azzari, G., Jain, M., & Lobell, D. B. (2017). Towards fine resolution global maps of crop yields: Testing multiple methods and satellites in three countries. *Remote Sensing of Environment*, 202, 129–141.
- Babcock, B. A., & Pautsch, G. R. (1998). Moving from uniform to variable fertilizer rates on Iowa corn: Effects on rates and returns. *Journal of Agricultural and Resource Economics*, 385–400.
- Balaghi, R., Tychon, B., Eerens, H., & Jlibene, M. (2008). Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *International Journal of Applied Earth Observation and Geoinformation*, 10(4), 438–452.
- Bampasidou, M., & Salassi, M. E. (2019). Trends in US Farm Labor and H-2A Hired Labor. *Choices*, 34(1), 1–6.

- Bannari, A., Morin, D., Bonn, F., & Huete, A. (1995). A review of vegetation indices. *Remote Sensing Reviews*, 13(1–2), 95–120.
- Baranski, M., Caswell, H., Claassen, R., Cherry, C., Jaglo, K., Lataille, A., Pailler, S., Pape, D., Riddle, A., & Stilson, D. (2018). Agricultural conservation on working lands: Trends from 2004 to present. *Technical Bulletin*, 1950.
- Basso, B., & Antle, J. (2020). Digital agriculture to design sustainable agricultural systems. *Nature Sustainability*, 3(4), 254–256.
- Basso, B., Bertocco, M., Sartori, L., & Martin, E. C. (2007). Analyzing the effects of climate variability on spatial pattern of yield in a maize–wheat–soybean rotation. *European Journal of Agronomy*, 26(2), 82–91.
- Basso, B., Cammarano, D., Chen, D., Cafiero, G., Amato, M., Bitella, G., Rossi, R., & Basso, F. (2009). Landscape position and precipitation effects on spatial variability of wheat yield and grain protein in Southern Italy. *Journal of Agronomy and Crop Science*, 195(4), 301–312.
- Basso, B., Cammarano, D., & De Vita, P. (2004). Remotely sensed vegetation indices: Theory and applications for crop management. *Rivista Italiana Di Agrometeorologia*, 1(5), 36–53.
- Basso, B., Dumont, B., Cammarano, D., Pezzuolo, A., Marinello, F., & Sartori, L. (2016). Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Science of the Total Environment*, 545, 227–235.
- Basso, B., Fiorentino, C., Cammarano, D., Cafiero, G., & Dardanelli, J. (2012). Analysis of rainfall distribution on spatial and temporal patterns of wheat yield in Mediterranean environment. *European Journal of Agronomy*, 41, 52–65.
- Basso, B., & Liu, L. (2019). Seasonal crop yield forecast: Methods, applications, and accuracies. In *Advances in Agronomy* (Vol. 154, pp. 201–255). Elsevier.
- Basso, B., & Ritchie, J. T. (2015). Simulating crop growth and biogeochemical fluxes in response to land management using the SALUS model. *The Ecology of Agricultural Landscapes: Long-Term Research on the Path to Sustainability*, 252–274.
- Basso, B., Ritchie, J. T., Cammarano, D., & Sartori, L. (2011). A strategic and tactical management approach to select optimal N fertilizer rates for wheat in a spatially variable field. *European Journal of Agronomy*, 35(4), 215–222.
- Basso, B., Ritchie, J. T., Pierce, F. J., Braga, R. P., & Jones, J. W. (2001). Spatial validation of crop models for precision agriculture. *Agricultural Systems*, 68(2), 97–112.
- Basso, B., Sartori, L., Bertocco, M., Cammarano, D., Martin, E. C., & Grace, P. R. (2011). Economic and environmental evaluation of site-specific tillage in a maize crop in NE Italy. *European Journal of Agronomy*, 35(2), 83–92.
- Basso, B., Shuai, G., Zhang, J., & Robertson, G. P. (2019). Yield stability analysis reveals

- sources of large-scale nitrogen loss from the US Midwest. *Scientific Reports*, 9(1), 1–9.
- Batchelor, W. D., Basso, B., & Paz, J. O. (2002). Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy*, 18(1–2), 141–158.
- Berzsenyi, Z., Györfy, B., & Lap, D. (2000). Effect of crop rotation and fertilisation on maize and wheat yields and yield stability in a long-term experiment. *European Journal of Agronomy*, 13(2–3), 225–244.
- Betbeder, J., Fieuzal, R., & Baup, F. (2016). Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(6), 2540–2553.
- Bianchi, F. J. J. A., Booij, C. J. H., & Tscharntke, T. (2006). Sustainable pest regulation in agricultural landscapes: a review on landscape composition, biodiversity and natural pest control. *Proceedings of the Royal Society B: Biological Sciences*, 273(1595), 1715–1727.
- Blackmore, S. (2000). The interpretation of trends from multiple yield maps. *Computers and Electronics in Agriculture*, 26(1), 37–51.
- Boone, L. V., Vasilas, B. L., & Welch, L. F. (1984). The nitrogen content of corn grain as affected by hybrid, population, and location. *Communications in Soil Science and Plant Analysis*, 15(6), 639–650.
- Brandes, E., McNunn, G. S., Schulte, L. A., Muth, D. J., VanLoocke, A., & Heaton, E. A. (2018). Targeted subfield switchgrass integration could improve the farm economy, water quality, and bioenergy feedstock production. *GCB Bioenergy*, 10(3), 199–212.
- Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, 114(9), 2189–2194.
- Calvin, K., Cowie, A., Berndes, G., Arneth, A., Cherubini, F., Portugal-Pereira, J., Grassi, G., House, J., Johnson, F. X., & Popp, A. (2021). Bioenergy for climate change mitigation: Scale and sustainability. *GCB Bioenergy*, 13(9), 1346–1371.
- Chabbi, A., Lehmann, J., Ciais, P., Loescher, H. W., Cotrufo, M. F., Don, A., SanClements, M., Schipper, L., Six, J., & Smith, P. (2017). Aligning agriculture and climate policy. *Nature Climate Change*, 7(5), 307–309.
- Chen, P.-Y., Fedosejevs, G., Tiscareno-Lopez, M., & Arnold, J. G. (2006). Assessment of MODIS-EVI, MODIS-NDVI and VEGETATION-NDVI composite data using agricultural measurements: An example at corn fields in western Mexico. *Environmental Monitoring and Assessment*, 119(1–3), 69–82.
- Chen, X.-P., Cui, Z.-L., Vitousek, P. M., Cassman, K. G., Matson, P. A., Bai, J.-S., Meng, Q.-F.,

- Hou, P., Yue, S.-C., & Römheld, V. (2011). Integrated soil–crop system management for food security. *Proceedings of the National Academy of Sciences*, 108(16), 6399–6404.
- Christianson, L., Bhandari, A., Helmers, M., Kult, K., Sutphin, T., & Wolf, R. (2012). Performance evaluation of four field-scale agricultural drainage denitrification bioreactors in Iowa. *Transactions of the ASABE*, 55(6), 2163–2174.
- Ciampitti, I. A., & Vyn, T. J. (2012). Physiological perspectives of changes over time in maize yield dependency on nitrogen uptake and associated nitrogen efficiencies: A review. *Field Crops Research*, 133, 48–67.
- Clark, R. L., & McGuckin, R. L. (1996). Variable rate application technology: An overview. *Proceedings of the Third International Conference on Precision Agriculture*, 855–862.
- Clevers, J. G. P. W., & Gitelson, A. A. (2013). Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and-3. *International Journal of Applied Earth Observation and Geoinformation*, 23, 344–351.
- Conant, R. T., Berdanier, A. B., & Grace, P. R. (2013). Patterns and trends in nitrogen use and nitrogen recovery efficiency in world agriculture. *Global Biogeochemical Cycles*, 27(2), 558–566.
- Cui, Z., Zhang, H., Chen, X., Zhang, C., Ma, W., Huang, C., Zhang, W., Mi, G., Miao, Y., & Li, X. (2018). Pursuing sustainable productivity with millions of smallholder farmers. *Nature*, 555(7696), 363–366.
- Curnel, Y., de Wit, A. J. W., Duveiller, G., & Defourny, P. (2011). Potential performances of remotely sensed LAI assimilation in WOFOST model based on an OSS Experiment. *Agricultural and Forest Meteorology*, 151(12), 1843–1855.
- Davis, S. C., Parton, W. J., Grosso, S. J. Del, Keough, C., Marx, E., Adler, P. R., & DeLucia, E. H. (2012). Impact of second-generation biofuel agriculture on greenhouse-gas emissions in the corn-growing regions of the US. *Frontiers in Ecology and the Environment*, 10(2), 69–74.
- De Bruin, J. L., Porter, P. M., & Jordan, N. R. (2005). Use of a rye cover crop following corn in rotation with soybean in the upper Midwest. *Agronomy Journal*, 97(2), 587–598.
- De Wit, A. J. W. de, & Van Diepen, C. A. (2007). Crop model data assimilation with the Ensemble Kalman filter for improving regional crop yield forecasts. *Agricultural and Forest Meteorology*, 146(1–2), 38–56.
- Deines, J. M., Patel, R., Liang, S.-Z., Dado, W., & Lobell, D. B. (2021). A million kernels of truth: Insights into scalable satellite maize yield mapping and yield gap analysis from an extensive ground dataset in the US Corn Belt. *Remote Sensing of Environment*, 253, 112174.
- Dong, T., Liu, J., Shang, J., Qian, B., Ma, B., Kovacs, J. M., Walters, D., Jiao, X., Geng, X., &

- Shi, Y. (2019). Assessment of red-edge vegetation indices for crop leaf area index estimation. *Remote Sensing of Environment*, 222, 133–143.
- Drury, C. F., Tan, C., Welacky, T. W., Oloya, T. O., Hamill, A. S., & Weaver, S. E. (1999). Red clover and tillage influence on soil temperature, water content, and corn emergence. *Agronomy Journal*, 91(1), 101–108.
- Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: enabling analysis of the Landsat archive. *Remote Sensing*, 10(9), 1363.
- Easterling, D. R., & Fahey, D. W. (2018). Chapter 2: our changing climate. *Fourth National Climate Assessment. Volume II: Impacts, Risks, and Adaptation in the United States*.
- FAO. (2017). *World Fertilizer Trends and Outlook to 2020: Summary Report*. Food and Agriculture Organization of the United Nations Rome, Italy.
- Fountas, S., Espejo-García, B., Kasimati, A., Mylonas, N., & Darra, N. (2020). The future of digital agriculture: technologies and opportunities. *IT Professional*, 22(1), 24–28.
- Gao, F., & Zhang, X. (2021). Mapping crop phenology in near real-time using satellite remote sensing: Challenges and opportunities. *Journal of Remote Sensing*, 2021.
- Gelfand, I., Sahajpal, R., Zhang, X., Izaurrealde, R. C., Gross, K. L., & Robertson, G. P. (2013). Sustainable bioenergy production from marginal lands in the US Midwest. *Nature*, 493(7433), 514–517.
- Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*, 30(5).
- Golbashy, M., Ebrahimi, M., Khorasani, S. K., & Choukan, R. (2010). Evaluation of drought tolerance of some corn (*Zea mays* L.) hybrids in Iran. *African Journal of Agricultural Research*, 5(19), 2714–2719.
- Guan, K., Wu, J., Kimball, J. S., Anderson, M. C., Frolking, S., Li, B., Hain, C. R., & Lobell, D. B. (2017). The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. *Remote Sensing of Environment*, 199, 333–349.
- Hall, S. J., & Russell, A. E. (2019). Do corn-soybean rotations enhance decomposition of soil organic matter? *Plant and Soil*, 444(1), 427–442.
- Halvorson, A. D., Peterson, G. A., & Reule, C. A. (2002). Tillage system and crop rotation effects on dryland crop yields and soil carbon in the central Great Plains. *Agronomy Journal*, 94(6), 1429–1436.
- Hanson, E., Ritten, J., Miller, P., Nagler, A., & Gerace, S. (2020). *B-1360.1 Is growing switchgrass economically feasible? Economic considerations for switchgrass production in*

- the Upper Missouri River Basin. Laramie, WY: University of Wyoming Extension. Laramie, WY: University of Wyoming Extension. <http://www.wyoextension.org>*
- Hassan, M. A., Yang, M., Rasheed, A., Yang, G., Reynolds, M., Xia, X., Xiao, Y., & He, Z. (2019). A rapid monitoring of NDVI across the wheat growth cycle for grain yield prediction using a multi-spectral UAV platform. *Plant Science*, 282, 95–103.
- Hatfield, J. L., Cryder, M., & Basso, B. (2020). Remote sensing: advancing the science and the applications to transform agriculture. *IT Professional*, 22(3), 42–45.
- Hergoualc'h, K., Akiyama, H., Bernoux, M., Chirinda, N., Prado, A. del, Kasimir, Å., MacDonald, J. D., Ogle, S. M., Regina, K., & Weerden, T. J. van der. (2019). *N2O emissions from managed soils, and CO2 emissions from lime and urea application*.
- Hernández-Ochoa, I. M., Gaiser, T., Kersebaum, K.-C., Webber, H., Seidel, S. J., Grahmann, K., & Ewert, F. (2022). Model-based design of crop diversification through new field arrangements in spatially heterogeneous landscapes. A review. *Agronomy for Sustainable Development*, 42(4), 1–25.
- Hockstad, L., & Hanel, L. (2018). *Inventory of US greenhouse gas emissions and sinks. Environmental System Science Data Infrastructure for a Virtual Ecosystem*
- Hood, C. F., & Kidder, G. (1992). Fertilizers and energy. *Fact Sheet EES-58, November*.
- Horowitz, J., Ebel, R., & Ueda, K. (2010). *Economic Information Bulletin Number 70*.
- Howarth, R., Chan, F., Conley, D. J., Garnier, J., Doney, S. C., Marino, R., & Billen, G. (2011). Coupled biogeochemical cycles: eutrophication and hypoxia in temperate estuaries and coastal marine ecosystems. *Frontiers in Ecology and the Environment*, 9(1), 18–26.
- Huang, J., Tian, L., Liang, S., Ma, H., Becker-Reshef, I., Huang, Y., Su, W., Zhang, X., Zhu, D., & Wu, W. (2015). Improving winter wheat yield estimation by assimilation of the leaf area index from Landsat TM and MODIS data into the WOFOST model. *Agricultural and Forest Meteorology*, 204, 106–121.
- Huang, X., Ye, Y., Xiong, L., Lau, R. Y. K., Jiang, N., & Wang, S. (2016). Time series k-means: A new k-means type smooth subspace clustering for time series data. *Information Sciences*, 367, 1–13.
- Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation*, 49(2), 189–194.
- Hunt, M. L., Blackburn, G. A., Carrasco, L., Redhead, J. W., & Rowland, C. S. (2019). High resolution wheat yield mapping using Sentinel-2. *Remote Sensing of Environment*, 233, 111410.
- IPCC. (2019). *N2O Emissions From Managed Soils, and CO2 Emissions From Lime and Urea Application. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas*

Inventories.

- IPNI. (2012). *4R plant nutrition manual: A manual for improving the management of plant nutrition*. International Plant Nutrition Institute Norcross, Georgia, USA.
- Janati, H., Cuturi, M., & Gramfort, A. (2020). Spatio-temporal alignments: Optimal transport through space and time. *International Conference on Artificial Intelligence and Statistics*, 1695–1704.
- Jaynes, D. B., & Isenhardt, T. M. (2014). Reconnecting tile drainage to riparian buffer hydrology for enhanced nitrate removal. *Journal of Environmental Quality*, 43(2), 631–638.
- Ji, Z., Pan, Y., Zhu, X., Wang, J., & Li, Q. (2021). Prediction of crop yield using phenological information extracted from remote sensing vegetation index. *Sensors*, 21(4), 1406.
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92, 141–152.
- Jin, Z., Azzari, G., You, C., Di Tommaso, S., Aston, S., Burke, M., & Lobell, D. B. (2019). Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228, 115–128.
- Johnson, D. M. (2014). An assessment of pre-and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, 141, 116–128.
- Jones, J. W., Antle, J. M., Basso, B., Boote, K. J., Conant, R. T., Foster, I., Godfray, H. C. J., Herrero, M., Howitt, R. E., & Janssen, S. (2017). Brief history of agricultural systems modeling. *Agricultural Systems*, 155, 240–254.
- Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3–4), 235–265.
- Ju, J., Roy, D. P., Vermote, E., Masek, J., & Kovalsky, V. (2012). Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods. *Remote Sensing of Environment*, 122, 175–184.
- Jung, J., Maeda, M., Chang, A., Bhandari, M., Ashapure, A., & Landivar-Bowles, J. (2021). The potential of remote sensing and artificial intelligence as tools to improve the resilience of agriculture production systems. *Current Opinion in Biotechnology*, 70, 15–22.
- Kamara, A. Y., Menkir, A., Badu-Apraku, B., & Ibikunle, O. (2003). The influence of drought stress on growth, yield and yield components of selected maize genotypes. *The Journal of Agricultural Science*, 141(1), 43.
- Kang, Y., & Özdoğan, M. (2019). Field-level crop yield mapping with Landsat using a

- hierarchical data assimilation approach. *Remote Sensing of Environment*, 228, 144–163.
- Karlen, D. L., Cambardella, C. A., Kovar, J. L., & Colvin, T. S. (2013). Soil quality response to long-term tillage and crop rotation practices. *Soil and Tillage Research*, 133, 54–64.
- Kaspar, T. C., Jaynes, D. B., Parkin, T. B., & Moorman, T. B. (2007). Rye cover crop and gamagrass strip effects on NO₃ concentration and load in tile drainage. *Journal of Environmental Quality*, 36(5), 1503–1511.
- Kaspar, T. C., Jaynes, D. B., Parkin, T. B., Moorman, T. B., & Singer, J. W. (2012). Effectiveness of oat and rye cover crops in reducing nitrate losses in drainage water. *Agricultural Water Management*, 110, 25–33.
- Kayad, A., Sozzi, M., Gatto, S., Marinello, F., & Pirotti, F. (2019). Monitoring within-field variability of corn yield using Sentinel-2 and machine learning techniques. *Remote Sensing*, 11(23), 2873.
- Koch, B., Khosla, R., Frasier, W. M., Westfall, D. G., & Inman, D. (2004). Economic feasibility of variable-rate nitrogen application utilizing site-specific management zones. *Agronomy Journal*, 96(6), 1572–1580.
- Kravchenko, A. N., Robertson, G. P., Thelen, K. D., & Harwood, R. R. (2005). Management, topographical, and weather effects on spatial variability of crop grain yields. *Agronomy Journal*, 97(2), 514–523.
- Krueger, E. S., Ochsner, T. E., Porter, P. M., & Baker, J. M. (2011). Winter rye cover crop management influences on soil water, soil nitrate, and corn development. *Agronomy Journal*, 103(2), 316–323.
- Ladha, J. K., Pathak, H., Krupnik, T. J., Six, J., & van Kessel, C. (2005). Efficiency of fertilizer nitrogen in cereal production: retrospects and prospects. *Advances in Agronomy*, 87, 85–156.
- Lal, R. (2005). Soil erosion and carbon dynamics. In *Soil and Tillage Research* (Vol. 81, Issue 2, pp. 137–142). Elsevier.
- Langdale, G. W., Blevins, R. L., Karlen, D. L., McCool, D. K., Nearing, M. A., Skidmore, E. L., Thomas, A. W., Tyler, D. D., & Williams, J. R. (1991). Cover crop effects on soil erosion by wind and water. *Cover Crops for Clean Water*, 15–22.
- Lark, R. M. (1998). Forming spatially coherent regions by classification of multi-variate data: an example from the analysis of maps of crop yield. *International Journal of Geographical Information Science*, 12(1), 83–98.
- Lark, T. J., Spawn, S. A., Bougie, M., & Gibbs, H. K. (2020). Cropland expansion in the United States produces marginal yields at high costs to wildlife. *Nature Communications*, 11(1), 1–11.

- Lassaletta, L., Billen, G., Garnier, J., Bouwman, L., Velazquez, E., Mueller, N. D., & Gerber, J. S. (2016). Nitrogen use in the global food system: past trends and future trajectories of agronomic performance, pollution, trade, and dietary demand. *Environmental Research Letters*, 11(9), 095007.
- Lassaletta, L., Billen, G., Grizzetti, B., Anglade, J., & Garnier, J. (2014). 50 year trends in nitrogen use efficiency of world cropping systems: the relationship between yield and nitrogen input to cropland. *Environmental Research Letters*, 9(10), 105011.
- Lewandowski, I., Scurlock, J. M. O., Lindvall, E., & Christou, M. (2003). The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass and Bioenergy*, 25(4), 335–361.
- Li, Y., Guan, K., Schnitkey, G. D., DeLucia, E., & Peng, B. (2019). Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. *Global Change Biology*, 25(7), 2325–2337.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2(3), 18–22.
- Liebig, M. A., Schmer, M. R., Vogel, K. P., & Mitchell, R. B. (2008). Soil carbon storage by switchgrass grown for bioenergy. *Bioenergy Research*, 1(3), 215–222.
- Liu, J., Bowling, L., Kucharik, C., Jame, S., Baldos, U., Jarvis, L., Ramankutty, N., & Hertel, T. (2022). Multi-scale Analysis of Nitrogen Loss Mitigation in the US Corn Belt. *ArXiv Preprint ArXiv:2206.07596*.
- Liu, W., Tollenaar, M., Stewart, G., & Deen, W. (2004). Response of corn grain yield to spatial and temporal variability in emergence. *Crop Science*, 44(3), 847–854.
- Liu, X., Zhang, Y., Han, W., Tang, A., Shen, J., Cui, Z., Vitousek, P., Erisman, J. W., Goulding, K., & Christie, P. (2013). Enhanced nitrogen deposition over China. *Nature*, 494(7438), 459–462.
- Liu, Y., Swinton, S. M., & Miller, N. R. (2006). Is site-specific yield response consistent over time? Does it pay? *American Journal of Agricultural Economics*, 88(2), 471–483.
- Lobell, D. B., Thau, D., Seifert, C., Engle, E., & Little, B. (2015). A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*, 164, 324–333.
- Maas, S. J. (1988). Using satellite data to improve model estimates of crop yield. *Agronomy Journal*, 80(4), 655–662.
- Maestrini, B., & Basso, B. (2018a). Drivers of within-field spatial and temporal variability of crop yield across the US Midwest. *Scientific Reports*, 8(1), 1–9.
- Maestrini, B., & Basso, B. (2018b). Predicting spatial patterns of within-field crop yield variability. *Field Crops Research*, 219, 106–112.

- Maestrini, B., & Basso, B. (2021). Subfield crop yields and temporal stability in thousands of US Midwest fields. *Precision Agriculture*, 1–19.
- Maitra, S., & Ray, D. P. (2019). Enrichment of biodiversity, influence in microbial population dynamics of soil and nutrient utilization in cereal-legume intercropping systems: A Review. *Int. J. Biores. Sci*, 6(1), 11–19.
- Martinez-Feria, R. A., & Basso, B. (2020). Unstable crop yields reveal opportunities for site-specific adaptations to climate variability. *Scientific Reports*, 10(1), 1–10.
- Martinez-Feria, R. A., Basso, B., & Kim, S. (2022). Boosting climate change mitigation potential of perennial lignocellulosic crops grown on marginal lands. *Environmental Research Letters*.
- Martinez-Feria, R. A., Dietzel, R., Liebman, M., Helmers, M. J., & Archontoulis, S. V. (2016). Rye cover crop effects on maize: A system-level analysis. *Field Crops Research*, 196, 145–159.
- Martinez-Feria, R., & Basso, B. (2020). Predicting soil carbon changes in switchgrass grown on marginal lands under climate change and adaptation strategies. *GCB Bioenergy*, 12(9), 742–755.
- Meehan, T. D., Werling, B. P., Landis, D. A., & Gratton, C. (2012). Pest-suppression potential of midwestern landscapes under contrasting bioenergy scenarios. *PLoS One*, 7(7), e41728.
- Messina, C., Hammer, G., Dong, Z., Podlich, D., & Cooper, M. (2009). Modelling crop improvement in a G* E* M framework via gene-trait-phenotype relationships. *Crop Physiology: Interfacing with Genetic Improvement and Agronomy. The Netherlands: Elsevier*, 235–265.
- Montgomery, D. R. (2007). Soil erosion and agricultural sustainability. *Proceedings of the National Academy of Sciences*, 104(33), 13268–13272.
- Moore, E. B., Wiedenhoef, M. H., Kaspar, T. C., & Cambardella, C. A. (2014). Rye cover crop effects on soil quality in no-till corn silage–soybean cropping systems. *Soil Science Society of America Journal*, 78(3), 968–976.
- Mueller, N. D., Lassaletta, L., Runck, B. C., Billen, G., Garnier, J., & Gerber, J. S. (2017). Declining spatial efficiency of global cropland nitrogen allocation. *Global Biogeochemical Cycles*, 31(2), 245–257.
- Murphy, D. P., Schnug, E., & Haneklaus, S. (1995). Yield mapping-a guide to improved techniques and strategies. *Site-specific Management for Agricultural Systems*, 33–47.
- Nawar, S., Corstanje, R., Halcro, G., Mulla, D., & Mouazen, A. M. (2017). Delineation of soil management zones for variable-rate fertilization: A review. In *Advances in agronomy* (Vol. 143, pp. 175–245). Elsevier.

- Nearing, G. S., Crow, W. T., Thorp, K. R., Moran, M. S., Reichle, R. H., & Gupta, H. V. (2012). Assimilating remote sensing observations of leaf area index and soil moisture for wheat yield estimates: An observing system simulation experiment. *Water Resources Research*, 48(5).
- Nery, M., Santos, R., Santos, W., Lourenco, V., & Moreno, M. (2018). Facing digital agriculture challenges with knowledge engineering. *2018 First International Conference on Artificial Intelligence for Industries (AI4I)*, 118–119.
- Nguy-Robertson, A., Gitelson, A., Peng, Y., Viña, A., Arkebauer, T., & Rundquist, D. (2012). Green leaf area index estimation in maize and soybean: Combining vegetation indices to achieve maximal sensitivity. *Agronomy Journal*, 104(5), 1336–1347.
- Nolan, B. T., Ruddy, B. C., Hitt, K. J., & Helsel, D. R. (1997). Risk of Nitrate in groundwaters of the United States a national perspective. *Environmental Science & Technology*, 31(8), 2229–2236.
- Oldfield, E. E., Bradford, M. A., & Wood, S. A. (2019). Global meta-analysis of the relationship between soil organic matter and crop yields. *Soil*, 5(1), 15–32.
- Ozdogan, B., Gacar, A., & Aktas, H. (2017). Digital agriculture practices in the context of agriculture 4.0. *Journal of Economics Finance and Accounting*, 4(2), 186–193.
- Padilla, F. M., Gallardo, M., Peña-Fleitas, M. T., De Souza, R., & Thompson, R. B. (2018). Proximal optical sensors for nitrogen management of vegetable crops: A review. *Sensors*, 18(7), 2083.
- Panda, S. S., Hoogenboom, G., & Paz, J. O. (2010). Remote sensing and geospatial technological applications for site-specific management of fruit and nut crops: A review. *Remote Sensing*, 2(8), 1973–1997.
- Pannell, D. J. (2017). Economic perspectives on nitrogen in farming systems: Managing trade-offs between production, risk and the environment. *Soil Research*, 55(6), 473–478.
- Pantoja, J. L., Woli, K. P., Sawyer, J. E., & Barker, D. W. (2016). Winter rye cover crop biomass production, degradation, and nitrogen recycling. *Agronomy Journal*, 108(2), 841–853.
- Peters, R. D., Sturz, A. V., Carter, M. R., & Sanderson, J. B. (2003). Developing disease-suppressive soils through crop rotation and tillage management practices. *Soil and Tillage Research*, 72(2), 181–192.
- Pierce, R. A., & Milhollin, R. (2020). *Field borders for agronomic, economic and wildlife benefits*.
- Plant, R. E. (2001). Site-specific management: the application of information technology to crop production. *Computers and Electronics in Agriculture*, 30(1–3), 9–29.
- Price, A. E. (2008). *Corn Monoculture: No Friend of Biodiversity*.

- Rauff, K. O., & Bello, R. (2015). A review of crop growth simulation models as tools for agricultural meteorology. *Agricultural Sciences*, 6(09), 1098.
- Ravishankara, A. R., Daniel, J. S., & Portmann, R. W. (2009). Nitrous oxide (N₂O): the dominant ozone-depleting substance emitted in the 21st century. *Science*, 326(5949), 123–125.
- Reidmiller, D. R., Avery, C. W., Easterling, D. R., Kunkel, K. E., Lewis, K. L. M., Maycock, T. K., & Stewart, B. C. (2019). Fourth national climate assessment. *Volume II: Impacts, Risks, and Adaptation in the United States, Report-in-Brief*.
- Risch, S. J., Andow, D., & Altieri, M. A. (1983). Agroecosystem diversity and pest control: data, tentative conclusions, and new research directions. *Environmental Entomology*, 12(3), 625–629.
- Robertson, G. P., Hamilton, S. K., Barham, B. L., Dale, B. E., Izaurralde, R. C., Jackson, R. D., Landis, D. A., Swinton, S. M., Thelen, K. D., & Tiedje, J. M. (2017). Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes. *Science*, 356(6345), eaal2324.
- Robertson, G. P., Paul, E. A., & Harwood, R. R. (2000). Greenhouse gases in intensive agriculture: contributions of individual gases to the radiative forcing of the atmosphere. *Science*, 289(5486), 1922–1925.
- Robertson, G. P., & Vitousek, P. M. (2009). Nitrogen in agriculture: balancing the cost of an essential resource. *Annual Review of Environment and Resources*, 34(1), 97–125.
- Rogers, D. H., & Lamm, F. R. (2008). Kansas irrigation trends. *Proceedings of the 2012 Central Plains Irrigation Conference, Colby, Kansas, February 21-22*.
- Roley, S. S., Tank, J. L., Tyndall, J. C., & Witter, J. D. (2016). How cost-effective are cover crops, wetlands, and two-stage ditches for nitrogen removal in the Mississippi River Basin? *Water Resources and Economics*, 15, 43–56.
- Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A. C., Müller, C., Arneth, A., Boote, K. J., Folberth, C., Glotter, M., & Khabarov, N. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences*, 111(9), 3268–3273.
- Rouse Jr, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*.
- Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185, 57–70.
- Runck, B. C., Khoury, C. K., Ewing, P. M., & Kantar, M. (2020). The hidden land use cost of upscaling cover crops. *Communications Biology*, 3(1), 1–4.

- Rusch, A., Bommarco, R., Jonsson, M., Smith, H. G., & Ekbom, B. (2013). Flow and stability of natural pest control services depend on complexity and crop rotation at the landscape scale. *Journal of Applied Ecology*, 50(2), 345–354.
- Saeed, U., Dempewolf, J., Becker-Reshef, I., Khan, A., Ahmad, A., & Wajid, S. A. (2017). Forecasting wheat yield from weather data and MODIS NDVI using Random Forests for Punjab province, Pakistan. *International Journal of Remote Sensing*, 38(17), 4831–4854.
- Sassenrath, G., Mengarelli, L., Lingenfelser, J., Knapp, M., & Lin, X. (2022). Crop Production Summary-2021. *Kansas Agricultural Experiment Station Research Reports*, 8(3), 4.
- 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.
- Sayago, S., Ovando, G., & Bocco, M. (2017). Landsat images and crop model for evaluating water stress of rainfed soybean. *Remote Sensing of Environment*, 198, 30–39.
- Schimmelpfennig, D. (2016). *Farm profits and adoption of precision agriculture*.
- Schulte, L. A., Niemi, J., Helmers, M. J., Liebman, M., Arbuckle, J. G., James, D. E., Kolka, R. K., O’Neal, M. E., Tomer, M. D., & Tyndall, J. C. (2017). Prairie strips improve biodiversity and the delivery of multiple ecosystem services from corn–soybean croplands. *Proceedings of the National Academy of Sciences*, 114(42), 11247–11252.
- Shcherbak, I., Millar, N., & Robertson, G. P. (2014). Global metaanalysis of the nonlinear response of soil nitrous oxide (N₂O) emissions to fertilizer nitrogen. *Proceedings of the National Academy of Sciences*, 111(25), 9199–9204.
- Snapp, S. S., Blackie, M. J., Gilbert, R. A., Bezner-Kerr, R., & Kanyama-Phiri, G. Y. (2010). Biodiversity can support a greener revolution in Africa. *Proceedings of the National Academy of Sciences*, 107(48), 20840–20845.
- Sonka, S. (2016). Big data: fueling the next evolution of agricultural innovation. *Journal of Innovation Management*, 4(1), 114–136.
- Soussana, J.-F., Lutfalla, S., Ehrhardt, F., Rosenstock, T., Lamanna, C., Havlík, P., Richards, M., Chotte, J.-L., Torquebiau, E., & Ciais, P. (2019). Matching policy and science: Rationale for the ‘4 per 1000-soils for food security and climate’ initiative. *Soil and Tillage Research*, 188, 3–15.
- Stuart, D., Basso, B., Marquart-Pyatt, S., Reimer, A., Robertson, G. P., & Zhao, J. (2015). The need for a coupled human and natural systems understanding of agricultural nitrogen loss. *BioScience*, 65(6), 571–578.
- Stuart, D., Schewe, R. L., & McDermott, M. (2014). Reducing nitrogen fertilizer application as a climate change mitigation strategy: Understanding farmer decision-making and potential barriers to change in the US. *Land Use Policy*, 36, 210–218.

- Tang, S., Zhu, Q., Zhou, X., Liu, S., & Wu, M. (2002). A conception of digital agriculture. *IEEE International Geoscience and Remote Sensing Symposium*, 5, 3026–3028.
- Tao, F., Rötter, R. P., Palosuo, T., Gregorio Hernández Díaz-Ambrona, C., Mínguez, M. I., Semenov, M. A., Kersebaum, K. C., Nendel, C., Specka, X., & Hoffmann, H. (2018). Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments. *Global Change Biology*, 24(3), 1291–1307.
- Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., Payne, M., Yurchak, R., Rußwurm, M., & Kolar, K. (2020). Tslern, a machine learning toolkit for time series data. *J. Mach. Learn. Res.*, 21(118), 1–6.
- Thelen, K. (2007). Assessing drought stress effects on corn yield. *Field Crop Advisory Team Alert Newsletter. Michigan State Univ. Available Online at Http://Msue. Anr. Msu. Edu/News/Assessing*.
- Tilman, D., Balzer, C., Hill, J., & Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108(50), 20260–20264.
- Tonitto, C., David, M. B., & Drinkwater, L. E. (2006). Replacing bare fallows with cover crops in fertilizer-intensive cropping systems: A meta-analysis of crop yield and N dynamics. *Agriculture, Ecosystems & Environment*, 112(1), 58–72.
- USDA. (2011). Web soil survey. *Available Online at Websoilsurvey. Nrcs. Usda. Gov. Accessed, 18*.
- Van Metre, P. C., Frey, J. W., Musgrove, M., Nakagaki, N., Qi, S., Mahler, B. J., Wieczorek, M. E., & Button, D. T. (2016). High nitrate concentrations in some Midwest United States streams in 2013 after the 2012 drought. *Journal of Environmental Quality*, 45(5), 1696–1704.
- Veldhuizen, L. J. L., Giller, K. E., Oosterveer, P., Brouwer, I. D., Janssen, S., van Zanten, H. H. E., & Slingerland, M. A. (2020). The Missing Middle: Connected action on agriculture and nutrition across global, national and local levels to achieve Sustainable Development Goal 2. *Global Food Security*, 24, 100336.
- Vitousek, P. M., Menge, D. N. L., Reed, S. C., & Cleveland, C. C. (2013). Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1621), 20130119.
- Vitousek, P. M., Naylor, R., Crews, T., David, M. B., Drinkwater, L. E., Holland, E., Johnes, P. J., Katzenberger, J., Martinelli, L. A., & Matson, P. A. (2009). Nutrient imbalances in agricultural development. *Science*, 324(5934), 1519–1520.
- Waldner, F., Horan, H., Chen, Y., & Hochman, Z. (2019). High temporal resolution of leaf area data improves empirical estimation of grain yield. *Scientific Reports*, 9(1), 1–14.

- Wallander, S., Smith, D., Bowman, M., & Claassen, R. (2021). *Cover crop trends, programs, and practices in the United States*.
- Weiss, A. (2001). Topographic position and landforms analysis. *Poster Presentation, ESRI User Conference, San Diego, CA, 200*.
- Weltzien, C. (2016). Digital agriculture or why agriculture 4.0 still offers only modest returns. *Landtechnik, 71*(2), 66–68.
- Williams, A., Scott Wells, M., Dickey, D. A., Hu, S., Maul, J., Raskin, D. T., Chris Reberg-Horton, S., & Mirsky, S. B. (2018). Establishing the relationship of soil nitrogen immobilization to cereal rye residues in a mulched system. *Plant and Soil, 426*(1), 95–107.
- Wulanningtyas, H. S., Gong, Y., Li, P., Sakagami, N., Nishiwaki, J., & Komatsuzaki, M. (2021). A cover crop and no-tillage system for enhancing soil health by increasing soil organic matter in soybean cultivation. *Soil and Tillage Research, 205*, 104749.
- Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight, E. A., Shaw, J., Masek, J. G., & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment, 185*, 271–283.
- Xiao, J., Hart, C. E., & Lence, S. H. (2017). USDA Forecasts Of Crop Ending Stocks: How Well Have They Performed? *Applied Economic Perspectives and Policy, 39*(2), 220–241.
- Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors, 2017*.
- Yan, L., & Roy, D. P. (2018). Large-area gap filling of Landsat reflectance time series by spectral-angle-mapper based spatio-temporal similarity (SAMSTS). *Remote Sensing, 10*(4), 609.
- Zhang, W., Dou, Z., He, P., Ju, X.-T., Powlson, D., Chadwick, D., Norse, D., Lu, Y.-L., Zhang, Y., & Wu, L. (2013). New technologies reduce greenhouse gas emissions from nitrogenous fertilizer in China. *Proceedings of the National Academy of Sciences, 110*(21), 8375–8380.
- Zhang, X., Davidson, E. A., Mauzerall, D. L., Searchinger, T. D., Dumas, P., & Shen, Y. (2015). Managing nitrogen for sustainable development. *Nature, 528*(7580), 51–59.
- Zhao, Y., Potgieter, A. B., Zhang, M., Wu, B., & Hammer, G. L. (2020). Predicting wheat yield at the field scale by combining high-resolution Sentinel-2 satellite imagery and crop modelling. *Remote Sensing, 12*(6), 1024.
- Zhou, X., Helmers, M. J., Asbjornsen, H., Kolka, R., & Tomer, M. D. (2010). Perennial filter strips reduce nitrate levels in soil and shallow groundwater after grassland-to-cropland conversion. *Journal of Environmental Quality, 39*(6), 2006–2015.
- Zulauf, C., & Brown, B. (2019). Tillage Practices, 2017 US Census of Agriculture. *Farmdoc Daily, 9*(136).