

**USING DIGITAL AGRICULTURE METHODOLOGIES TO GENERATE
SPATIAL AND TEMPORAL PREDICTIONS OF N
CONSERVATION, MANAGEMENT AND MAIZE YIELD**

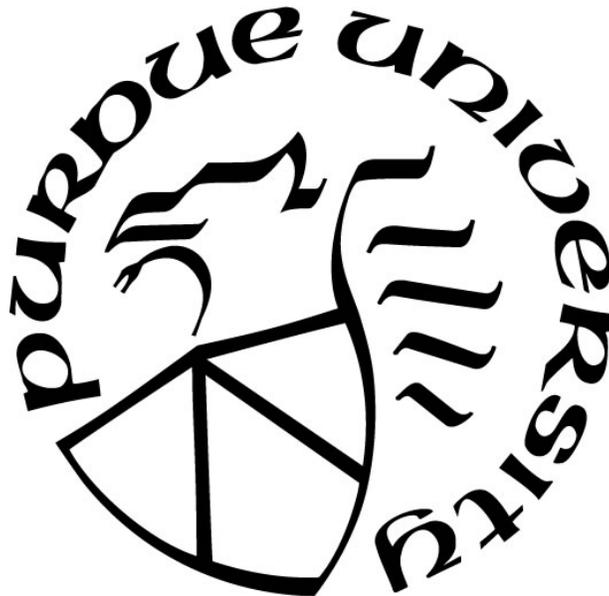
by
Min Xu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Agronomy

West Lafayette, Indiana

December 2018

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Shalamar D. Armstrong, Chair

Department of Agronomy

Dr. Tony J. Vyn

Department of Agronomy

Dr. Haishun Yang

Department of Agronomy and Horticulture, University of Nebraska - Lincoln

Dr. Tonglin Zhang

Department of Statistics

Approved by:

Dr. Ronald F. Turco

Head of the Graduate Program

To Yiheng

ACKNOWLEDGMENTS

The completion of this dissertation would not have been possible without the support and guidance of many individuals.

First, I would like to thank my major advisor, Dr. Shalamar Armstrong, who has been a great mentor and gave me support and guidance than I could have ever hoped for. I would also like to thank Dr. Brad Joern, who was my academic advisor from August 2011 to May 2017, for his encouragement, persistent help, and trust along the way.

I would like to extend my sincere gratitude to my committee members: Drs. Tony Vyn, Haishun Yang, and Tonglin Zhang for many generous and helpful suggestions.

Much love and appreciation go to my amazing friends at Purdue. Their intelligence, friendship and work ethic have been a great inspiration to me.

Last but not least, I would like to thank my family, whose love, support and sacrifice helped lead me to that light at the end of the tunnel.

TABLE OF CONTENTS

LIST OF TABLES	8
LIST OF FIGURES	9
ABSTRACT	11
CHAPTER 1. INTRODUCTION	13
1.1 Introduction	13
1.2 References	16
CHAPTER 2. THE FEASIBILITY OF SATELLITE REMOTE SENSING AND SPATIAL INTERPOLATION TO ESTIMATE COVER CROP BIOMASS AND NITROGEN UPTAKE IN A SMALL WATERSHED	20
2.1 Abstract.....	20
2.2 Introduction	21
2.3 Materials and Methods	23
2.3.1 Study area.....	23
2.3.2 Ground cover crop biomass and N uptake sampling.....	24
2.3.3 Satellite remote sensing indices	24
2.3.4 Spatial interpolation	25
2.4 Results and Discussion.....	26
2.4.1 Linear correlation between ground samples and different satellite remote sensing vegetation indices	26
2.4.2 The performance of cover crop estimation using soil adjusted vegetation index ...	28
2.4.3 Spatial interpolation using ordinary kriging and inverse distance weighting	29
2.5 Summary and Conclusions	30
2.6 Acknowledgements	31
2.7 References	31
CHAPTER 3. A NEW MODELING APPROACH: BIG DATA AND MULTIVARIATE SPATIAL STATISTICS TO GENERATE IN-SEASON FIELD SPECIFIC MAIZE YIELD PREDICTIONS 46	
3.1 Abstract.....	46
3.2 Introduction	46

3.3	Materials and Methods	49
3.3.1	Field N management and maize yield	49
3.3.2	Weather and crop data	50
3.3.3	Field and soil attributes	51
3.3.4	Landsat remote sensing vegetation index.....	51
3.3.5	Yield correlations with weather and remote sensing indices	52
3.3.6	Spatial autocorrelation analysis.....	52
3.3.7	Multivariate spatial autoregressive model.....	53
3.3.8	Model goodness of fit.....	53
3.4	Results and Discussion	54
3.4.1	Overview of maize yield	54
3.4.2	Yield correlations with weather and remote sensing indices	55
3.4.3	Yield spatial structure and performance of the MSAR model	57
3.4.4	Prediction accuracy of the MSAR model.....	58
3.4.4.1	The MSAR model performance in a drought year	58
3.4.4.2	The MSAR model performance in a typical year	59
3.4.4.3	The MSAR model performance on new fields.....	59
3.5	Conclusions	60
3.6	Acknowledgements	61
3.7	References	61
CHAPTER 4. INCORPORATE MULTIVARIATE SPATIAL STATISTICS TO A PROCESS-BASED NITROGEN TRANSFORMATION MODEL FOR FIELD-SCALE MAIZE GRAIN YIELD PREDICTION.....		78
4.1	Abstract.....	78
4.2	Introduction	79
4.3	Materials and Methods	81
4.3.1	Historical maize yield data	81
4.3.2	The process-based N Model in MapWindows GIS + MMP Tools.....	82
4.3.3	Weather and crop data	83
4.3.4	Field and soil attributes	84
4.3.5	Maize yield residual analysis	84

4.3.6	Statistical measures for model performance	85
4.4	Results and Discussion	86
4.4.1	Performance of the N Model	86
4.4.2	Analysis of maize yield residual of the N Model	87
4.4.3	Performance of the MSAR model adjusted yield.....	88
4.5	Conclusions	90
4.6	Acknowledgements	91
4.7	References	92
CHAPTER 5. CONCLUSIONS		107
5.1	Conclusions	107
5.2	References	110
APPENDIX A. SUPPLEMENTARY FIGURES		111
APPENDIX B. SUPPLEMENTARY TABLES.....		114
VITA.....		126
PUBLICATION.....		130

LIST OF TABLES

Table 2.1 Definition of remote sensing VI's. Surface reflectance of bands are designated in the formula as B (blue), G (green), R (red), and NIR (near-infrared).	37
Table 2.2 Leave-one-out cross validation of linear models between cover crop biomass and remote sensing indices and spatial interpolation.....	37
Table 2.3 Leave-one-out cross validation average out-of-sample R ² values of linear models between cover crop N uptake and remote sensing indices.	38
Table 2.4 Means and standard deviations of cover crop biomass estimation in the study watershed.	39
Table 2.5 Means and standard deviations of cover crop N uptake estimation in the study watershed.	40
Table 3.1 Dominant soil types and soil properties of the study fields.....	66
Table 3.2 Cropping history and soil types of the fields in this study.....	67
Table 3.3 Planting date of each field.	67
Table 3.4 General weather conditions during maize growing season. Values are the means across fields for any particular year.	68
Table 3.5 Relationship between maize grain yield and seasonal rainfall and temperature.	69
Table 3.6 Maize grain yield, Moran's I, yield estimates from multivariate spatial autoregressive model, and model goodness of fit of the calibration dataset.....	70
Table 3.7 Multivariate spatial autoregressive model prediction.....	71
Table 4.1 Advantages and disadvantages of the two modeling approaches.	97
Table 4.2 Maize N uptake function in the N Model in the MapWindows GIS + MMP Tools. ...	98
Table 4.3 Input parameters of each field for the N Model.....	98
Table 4.4 Maize grain yield, MSAR adjusted yield estimates, and model goodness of fit of the calibration dataset.	99
Table 4.5 Model goodness of fit comparison of the N Model and MSAR adjusted grain yield in a drought year (2012) and a typical year (2016).....	100

LIST OF FIGURES

Figure 2.1 Cover crop treatments in the study watershed.....	41
Figure 2.2 Linear relationship between remote sensing indices and cover crop biomass.	42
Figure 2.3 Linear relationship between remote sensing indices and cover crop N uptake.....	43
Figure 2.4 Estimated cover crop biomass of the study watershed.....	44
Figure 2.5 Estimated cover crop N uptake of the study watershed.	45
Figure 3.1 Spatial neighbors based on queen’s contiguity.	72
Figure 3.2 Biplot derived from the first two principal components. EV, early vegetative stage; LV, late vegetative stage; CP, critical period; GF, grain filling period; Pcp, precipitation, mm; Rad, solar radiation, MJ m ⁻² ; VPD, vapor pressure deficit, kPa; KDD, killing degree days, d.°C; GCVI, green chlorophyll vegetation index. Each point represents the mean within a soil type and site-year.	73
Figure 3.3 The relationship between maize grain yield and (A) green chlorophyll vegetation index (GCVI); (B) grain filling period cumulative killing degree days (GF KDD) (d.°C); (C) critical period cumulative killing degree days (CP KDD) (d.°C); and (D) critical period mean vapor pressure deficit (CP VPD) (kPa); using calibration dataset from 2010 to 2016, and each point represents the mean within a soil type and site-year.....	74
Figure 3.4 The negative linear relationship between Moran’s I statistic of grain yield and grain yield. Each point represents a site-year in the calibration dataset (n=35).	75
Figure 3.5 Comparison of maize grain yield predicted using MSAR model and actual grain yield of the calibration dataset (n=35). The 1:1 line (solid red line) and ±15% deviation (dashed black line) are shown. Fitted linear regression model is shown (solid black line).	76
Figure 3.6 Comparison of maize grain yield predicted using MSAR model and actual grain yield of the validation dataset (n=82). The 1:1 line (solid red line) and ±15% deviation (dashed black line) are shown. (A) Fitted linear regression model using all site-years is shown (solid black line). (B) Fitted linear regression model excluding 2012 site-years is shown (solid black line).	77
Figure 4.1 General weather conditions during maize growing season during 2010 – 2016. (A) Early Vegetative Stage; (B) Late Vegetative Stage; (C) Critical Period; and (D) Grain-fill Stage.	

Each point represents the average temperature and precipitation across all fields within a year. The red dash lines indicate long-term weather averages assuming a planting date of May 1st.

..... 101

Figure 4.2 Comparison of the N Model predicted and actual maize grain yields at three levels: (A) Soil type within a field; (B) Individual fields; and (C) All fields within a year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown. Data from the calibration dataset with 35 site-years. 102

Figure 4.3 Relationship between yield residual at the soil type level and EV precipitation. Each point represents the mean of a soil type within a site-year. Data from the calibration dataset with 35 site-years. 103

Figure 4.4 Relationship between Moran's I statistic of yield residual and (A) Actual yield; and (B) EV Precipitation. Each point represents the mean of a site-year. Data from the calibration dataset with 35 site-years. 104

Figure 4.5 Comparison of the MSAR model adjusted maize grain yield and actual maize grain yields at three levels: (A) Soil type within a field; (B) Individual fields; and (C) All fields within a year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown. Data from the calibration dataset with 35 site-years. 105

Figure 4.6 Comparison of (A) the N Model yield prediction and (B) the MSAR adjusted yield prediction using the validation dataset (n=82). Each point represents the mean of a site-year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown. 106

ABSTRACT

Author: Xu, Min. PhD

Institution: Purdue University

Degree Received: December 2018

Title: Using Digital Agriculture Methodologies to Generate Spatial and Temporal Predictions of N Conservation, Management and Maize Yield.

Committee Chair: Shalamar Armstrong

The demand for customized farm management prescription is increasing in order to maximize crop yield and minimize environmental risks under a changing climate. One great challenge to modeling crop growth and production is spatial and temporal variability. The goal of this dissertation research is to use publicly available Landsat imagery, ground samples and historical yield data to establish methodologies to spatially quantify cover crop growth and in-season maize yield. First, an investigation was conducted into the feasibility of using satellite remote sensing and spatial interpolation with minimal ground samples to rapidly estimate season-specific cover crop biomass and N uptake in the small watershed of Lake Bloomington in Illinois. Results from this study demonstrated that remote sensing indices could capture the spatial pattern of cover crop growth as affected by various cover crop and cash crop management systems. Soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI) and triangular vegetation index (TVI) were strongly correlated with cover crop biomass and N uptake for low and moderate biomass and N uptake ranges (0-3000 kg ha⁻¹ and 0-100 kg N ha⁻¹). The SAVI estimated cover crop biomass and N uptake were +/- 15% of observed value. Compared to commonly used spatial interpolation methods such as ordinary kriging (OK) and inverse distance weighting (IDW), using the SAVI method showed higher prediction R² values than that of OK and IDW. An additional advantage for these remote sensing vegetation indices, especially in the context of diverse agronomic management practices, is their much lower labor requirements compared to the high density ground samples needed for a spatial interpolation analysis.

In the second study, a new approach using the multivariate spatial autoregressive (MSAR) model was developed at 10-m grid resolution to forecast maize yield using historical grain yield data collected at farmers' fields in Central Indiana, publicly available Landsat imagery, top 30 cm soil organic matter and elevation, while accounting for yield spatial autocorrelation. Relative mean

error (RME) and relative mean absolute error (RMAE) were used to quantify the model prediction accuracy at the field level and 10-m grid level, respectively. The MSAR model performed reasonably well (absolute RME < 15%) for field overall yield predictions in 32 out of 35 site-years on the calibration dataset with an average absolute RME of 6.6%. The average RMAE of the MSAR model predictions was 13.1%. It was found that the MSAR model could result in large estimation error under an extreme stressed environment such as the 2012 drought, especially when grain yield under these stressed conditions was not included in the model calibration step. In the validation dataset (n=82), the MSAR model showed good prediction accuracy overall ($\pm 15\%$ of actual yield in 56 site-years) in new fields when extreme stress was not present. The novel approach developed in this study demonstrated its ability to use elevation and soil information to interpret satellite observations accurately in a fine spatial scale.

Then we incorporated the MSAR approach into a process-based N transformation model to predict field-scale maize yield in Indiana. Our results showed that the linear agreement of predicted yield (using the N Model in the Mapwindow GIS + MMP Tools) to actual yield improved as the spatial aggregation scale became broader. The proposed MSAR model used early vegetative precipitation, top 30 cm soil organic matter and elevation to adjust the N Model yield prediction in 10-m grids. The MSAR adjusted yield predictions resulted in more cases (77%) that fell within 15% of actual yield compared to the N Model alone using the calibration dataset (n=35). However, if the 2012 data was not included in the MSAR parameter training step, the MSAR adjusted yield predictions for 2012 were not improved from the N Model prediction (average RME of 24.1%). When extrapolating the MSAR parameters developed from 7 fields to a dataset containing 82 site-years on 30 different fields in the same region, the improvement from the MSAR adjustment was not significant. The lack of improvement from the MSAR adjustment could be because the relationship used in the MSAR model was location specific. Additionally, the uncertainty of precipitation data could also affect the relationship.

Through the sequence of these studies, the potential utility of big data routinely collected at farmers' fields and publicly available satellite imagery has been greatly improved for field-specific management tools and on-farm decision-making.

CHAPTER 1. INTRODUCTION

1.1 Introduction

Increasing population, soil erosion at an alarming rate, and significant changes in weather patterns present challenges to farmers. According to the United Nation (UN), the world population is projected to reach 9.8 billion in 2050 and the upward trend in population size is expected to continue (UN, 2017). The UN Food and Agriculture Organization (FAO) estimated that farmers will have to produce 50% more food by 2050 to meet the needs of the world's population (FAO, 2017). However, fertile soil was being lost at a rate of 24 billion tons every year around the world due to intensive agriculture. In the United States, 1.73 billion tons of topsoil loss via erosion was reported (USDA, 2007). Moreover, studies have indicated that extreme events such as heat waves (especially frequencies of hot nights) and large storms are likely to become more frequent or more intense with climate change (Walthall et al., 2012); at the same time, the occurrence of drought has also been on the rise (IPCC, 2007). Climate models suggest that rainfall may become less predictable (Walthall et al., 2012). Although there will be gains in some crops in some regions of the world, the overall impacts of climate change on agriculture are expected to be negative, threatening global food security (Nelson et al., 2009; Zhao et al., 2017).

The United States is the largest maize producer in the world, contributing approximately 40% of global maize production annually (Smith et al., 2014). Over the past few decades, US maize yield trend has been linear with a mean 1.5% annual increase (Lobell and Azzari, 2017). Plant breeding and improved agronomic practices have contributed to increased agricultural productivity during recent past decades. Smith et al. (2014) reported that for single-cross hybrids during 1930 to 2011, genetic gains achieved via improved phenotyping, marker assisted breeding, genetic engineering, and seed treatments contributed 59% to 79% of the maize yield gain. Agronomic management practices also play an important role in closing maize yield gap by reducing soil erosion, improving water and nutrient use efficiency, and control of weeds, insects and disease. These practices include better nitrogen (N) application timing, source, placement, and rate decisions (Randall and Vetsch, 2005; Randall and Sawyer, 2008; Mueller et al., 2012); better soil drainage (Cardwell, 1982; Kladivko et al., 2005; Nash et al., 2015); early planting dates (Kucharik, 2008); increased plant density (Grassini et al., 2011); crop rotation (Porter et al., 1997; Sindelar et

al., 2015); conservation tillage (Zhang and Blevins, 1996; Licht and Al-Kaisi, 2005); cover cropping (Kramberger et al., 2009; Lacey and Armstrong, 2015); and other practices. To feed the world in 2050, gains from both plant breeding and agronomic management practices are still needed to sustainably increase agricultural production with an improved environmental footprint by making the plants more resilient in the face of extreme weather, weed competition, and other pest attacks in future agriculture.

Maize production is heavily influenced by the environment. While the average maize yield has been increasing in the past, the yield variability increased as well. Lobell and Azzari (2017), using Landsat satellite images, detected that maize yield heterogeneity in US Corn Belt was rising, both between and within fields. They found average yield differences between the best and worst soils more than doubled from 2000 to 2015. They rationalized this trend could be the result of increased plant density, which disproportionately raised yields on productive soils. The results indicated that yield gains in this region were increasingly derived from the more productive soils. However, variation in soil N contributions via organic matter mineralization among soils and within soils from year to year have made fertilizer N recommendations difficult (Joern and Sawyer, 2006). Moreover, maize yield response to N fertilizer was found to vary depending on topography, soil type, groundwater levels, farm management and ever-changing environmental conditions. As a result, the selection of a site-specific optimum N rate was difficult to predict based on the large temporal and spatial variability of the N supply and crop N demand (Setiyono et al., 2011; Puntel et al., 2016). Therefore, data-driven nutrient management decisions based on field-specific information is essential.

Digital Agriculture is the integration of digitalized data sources (yield maps, fertilizer application, weather information, real-time satellite imagery, unmanned aviation systems, soil mapping, field sensors) with advanced crop and environment models, to provide a better understanding of nutrient management, which helps farmers make informed and actionable on-farm decisions to optimize resource management, reduce production loss and improve food security in a changing climate. So Digital Agriculture technologies have the potential to increase N use efficiency by matching the N requirements within field zones (Mamo et al, 2003; Mulla, 2013; Puntel et al., 2016). Many crop models have been developed to quantify maize yield response to applied N fertilizer, crop rotation, and the impact of climate change at various scales (Jones et al., 2003; Yang et al., 2004; Setinoyo et al., 2011; Holzworth et al., 2014). Statistical

models and algorithms were also used to predict maize yield and the potential effects of climate change (Lobell et al., 2011; Lobell et al., 2015). Unfortunately, these approaches have not fully resolved the needed improvements from N management since N losses from maize-based systems are still high with negative environment impacts (Puntel et al., 2016). The EPA cites agricultural runoff as the leading cause of pollution of lakes and rivers, and the hypoxic zone in the Gulf of Mexico in 2017 was determined to be the largest since 1985 (NOAA, 2017). Planting cover crops was identified as the most effective management practice that reduce nitrate-N loss (INLRS, 2015). In practice, however, cover crop establishment, biomass production, and nutrient uptake could be affected by greater variability in weather, soil type, landscape position, soil drainage and water table, and nutrient management practices.

The goal of this dissertation is to establish digital methodologies to spatially quantify cover crop growth and maize grain yield. We believe that digital methodologies that spatially analyze the big data routinely collected at farmers' fields such as yields, fertilizer inputs, and publicly available satellite images, can provide the farmer with powerful field-specific tools to rapidly assess productivity and manage conservation to minimize environmental risks in maize-based agricultural systems. The results from this dissertation can be implemented to predict field-specific maize grain yield to agronomic management and to advance field adaptive N management with cover crop growth. This local, data-driven approach could offer insights to enhance the ability in decision making and implementation.

Therefore, the specific research objectives are:

- 1) Develop algorithms that use Landsat satellite images and minimal ground samples to predict season-specific cover crop biomass and N uptake on a small watershed in the US Corn Belt; and compare the estimation accuracy with common spatial interpolation methods.
- 2) Develop a multivariate spatial autoregressive model using Landsat satellite imagery, historical yield data, soil survey and digital elevation to predict field-scale maize yield in Indiana.
- 3) Incorporate multivariate spatial statistics into a process-based N transformation model to predict maize yield on a field scale in Indiana.

1.2 References

- Cardwell, V.B. 1982. Fifty years of Minnesota corn production: Sources of yield increase. *Agronomy Journal*. 74: 984-990.
- FAO. 2017. The future of food and agriculture - Trends and challenges. Rome.
- Grassini, P., J. Thorburn, C. Burr, K.G. Cassman. 2011. High-yield irrigated maize in the Western U.S. Corn Belt: I. On-farm yield, yield potential, and impact of agronomic practices. *Field Crops Research*. 120: 142-150.
- Holzworth, D.P., N.I. Huth, P.G. deVoil, E.J. Zurcher, N.I. Herrmann, G. McLean, K. Chenu, E.J. van Oosterom, V. Snow, C. Murphy, A.D. Moore, H. Brown, J.P.M. Whish, S. Verrall, J. Fainges, L.W. Bell, A.S. Peake, P.L. Poulton, Z. Hochman, P.J. Thorburn, D.S. Gaydon, N.P. Dalgliesh, D. Rodriguez, H. Cox, S. Chapman, A. Doherty, E. Teixeira, J. Sharp, R. Cichota, I. Vogeler, F.Y. Li, E. Wang, G.L. Hammer, M.J. Robertson, J.P. Dimes, A.M. Whitbread, J. Hunt, H. van Rees, T. McClelland, P.S. Carberry, J.N.G. Hargreaves, N. MacLeod, C. McDonald, J. Harsdorf, S. Wedgwood, and B.A. Keating. 2014. APSIM – evolution towards a new generation of agricultural systems simulation. *Environment Modelling & Software*. 62: 327-350.
- ILNLRs. 2015. Illinois nutrient loss reduction strategy.
- IPCC. 2007. *Climate Change 2007: Impacts, adaptation and vulnerability*. Parry, M.L., O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson (Eds) Contribution of working group II to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.
- Joern, B., and J. Sawyer. 2006. Nitrogen and corn use. In: Sawyer, J., E. Nafziger, G. Randall, L. Bundy, G. Rehm, and B. Joern. *Concepts and rationale for regional nitrogen rate guidelines for corn*. Iowa State University Extension Publication, PM2015.
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, and L.A. Hunt. 2003. The DSSAT cropping system model. *European Journal of Agronomy*. 18: 235-265.
- Kladivko, E.J., G.L. Willoughby, and J.B. Santini. 2005. Corn growth and yield response to subsurface drain spacing on Clermont silt loam soil. *Agronomy Journal*. 97: 1419-1428.

- Kramberger, B., A. Gselman, M. Janzekovic, M. Kaligalic, and B. Bracko. 2009. Effects of cover crops on soil mineral nitrogen and on the yield and nitrogen content of maize. *European Journal of Agronomy*. 31: 103-109.
- Kucharik, C.J. 2008. Contribution of planting date trends to increased maize yields in the central United States. *Agronomy Journal*. 100: 328-336.
- Lacey, C., and S. Armstrong. 2015. The efficacy of winter cover crops to stabilize soil inorganic nitrogen after fall-applied anhydrous ammonia. *Journal of Environmental Quality*. 44: 442-448.
- Licht, M.A., and M. Al-Kaisi. 2005. Corn response, nitrogen uptake, and water use in strip-tillage compared with no-tillage and chisel plow. *Agronomy Journal*. 97: 705-710.
- Lobell, D.B., W. Schlenker, and J. Costa-Roberts. 2011. Climate trends and global crop production since 1980. *Science*. 333(6042): 616-620.
- Lobell, D.B., D. Thau, C. Seifert, E. Engle, and B. Little. 2015. A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*. 164: 324-333.
- Lobell, D.B., and G. Azzari. 2017. Satellite detection of rising maize yield heterogeneity in the U.S. Midwest. *Environmental Research Letters*. 12: 014014.
- Mamo, M., G.L., Malzer, D.J. Mulla, D.R. Huggins, and J. Strock. 2003. Spatial and temporal variation in economically optimum nitrogen rate for corn. *Agronomy Journal*. 95: 958-964.
- Mueller, N.D., J.S. Gerber, M. Johnston, D.K. Ray, N. Ramankutty, and J.A. Foley. 2012. Closing yield gaps through nutrient and water management. *Nature*. 490: 254-257.
- Mulla, D.J. 2013. Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. *Biosystems Engineering*. 114(4): 358-371.
- Nash, P., K. Nelson, and P. Motavalli. 2015. Reducing nitrogen loss with managed drainage and polymer-coated urea. *Journal of Environmental Quality*. 44: 256-264.
- Nelson, G.C., M.W. Rosegrant, J. Koo, R. Robertson, T. Sulser, T. Zhu, C. Ringler, S. Msangi, A. Palazzo, M. Batka, M. Magalhaes, R. Valmonte-Santos, M. Ewing, and D. Lee. 2009. *Climate change: impact on agriculture and costs of adaptation*. International Food Policy Research Institute. Washington, D.C.

- NOAA, 2017. Gulf of Mexico ‘dead zone’ is the largest ever measured: June outlook foretold New-Jersey-sized area of low oxygen. Available from <http://www.noaa.gov/media-release/gulf-of-mexico-dead-zone-is-largest-ever-measured>.
- Porter, P.M., J.G. Lauer, W.E. Lueschen, J.H. Ford, T.R. Hoverstad, E.S. Oplinger, and R.K. Crookston. 1997. *Agronomy Journal*. 89: 441-448.
- Puntel, L.A., J.E. Sawyer, D.W. Barker, R. Dietzel, H. Poffenbarger, M.J. Castellano, K.J. Moore, P. Thorburn, and S.V. Archontoulis. 2016. Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*. 7: 1-18.
- Randall, G.W., and J.A. Vetsch. 2005. Corn production on a subsurface-drained mollisol as affected by fall versus spring application of nitrogen and nitrapyrin. *Agron. J.* 97: 472-478.
- Randall, G.W., and J.E. Sawyer. 2008. Nitrogen application timing, forms and additives. In: Final report: Gulf hypoxia and local water quality concerns workshop, upper Mississippi river sub-basin hypoxia nutrient committee, American Society of Agricultural and Biological Engineers. St. Joseph, Michigan: 73-85.
- Setiyono, T.D., H. Yang, D.T. Walters, A. Dobermann, R.B. Ferguson, D.F. Roberts, D.J. Lyon, D.E. Clay, and K.G. Cassman. 2011. Maize-N: A decision tool for nitrogen management in maize. *Agronomy Journal*. 103(4):1-8.
- Sindelar, A.J., M.R. Schmer, V.L. Jin, B.J. Wienhold, and G.E. Varvel. 2015. Long-term corn and soybean response to crop rotation and tillage. *Agronomy Journal*. 107: 2241-2252.
- Smith, S., M. Cooper, J. Gogerty, C. Loffler, D. Borcharding, and K. Wright. 2014. Maize. In: Smith, S. B. Diers, J. Specht, and B. Carver (Eds) Yield gains in major U.S. field crops. CSSA Special Publication 33. Madison, WI, USA.
- UN. 2017. World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100. <https://www.un.org/development/desa/en/news/population/world-population-prospects-2017.html>
- USDA. 2007. National resources inventory – Soil Erosion on cropland.
- Walthall, C.L., J. Hatfield, P. Backlund, L. Lengnick, E. Marshall, M. Walsh, S. Adkins, M. Aillery, E.A. Ainsworth, C. Ammann, C.J. Anderson, I. Bartomeus, L.H. Baumgard, F.

- Booker, B. Bradley, D.M. Blumenthal, J. Bunce, K. Burkey, S.M. Dabney, J.A. Delgado, J. Dukes, A. Funk, K. Garrett, M. Glenn, D.A. Grantz, D. Goodrich, S. Hu, R.C. Izaurralde, R.A.C. Jones, S-H. Kim, A.D.B. Leaky, K. Lewers, T.L. Mader, A. McClung, J. Morgan, D.J. Muth, M. Nearing, D.M. Oosterhuis, D. Ort, C. Parmesan, W.T. Pettigrew, W. Polley, R. Rader, C. Rice, M. Rivington, E. Rosskopf, W.A. Salas, L.E. Sollenberger, R. Srygley, C. Stöckle, E.S. Takle, D. Timlin, J.W. White, R. Winfree, L. Wright-Morton, and L.H. Ziska. 2012. Climate change and agriculture in the United States: Effects and adaptation. USDA Technical Bulletin 1935. Washington, D.C.
- Yang, H.S., A. Dobermann, J.L. Lindquist, D.T. Walters, T.J. Arkebauer, and K.G. Cassman. 2004. Hybrid-maize – a maize simulation model that combines two crop modeling approaches. *Field Crops Research*. 87: 131-154.
- Zhang, Z., and R.L. Blevins. 1996. Corn yield response to cover crops and N rates under long-term conventional and no-tillage management. *Journal of Sustainable Agriculture*. 8(1): 61-72.
- Zhao, C., B. Liu, S. Piao, X. Wang, D.B. Lobell, Y. Huang, M. Huang, Y. Yao, S. Bassu, P. Ciais, J. Durand, J. Elliott, F. Ewert, I.A. Janssens, T. Li, E. Lin, Q. Liu, P. Martre, C. Muller, S. Peng, J. Penuelas, A.C. Ruane, D. Wallach, T. Wang, D. Wu, Z. Liu, Y. Zhu, Z. Zhu, and S. Asseng. 2017. Temperature increase reduces global yields of major crops in four independent estimates. *PNAS*. 114(35): 9326-9331.

CHAPTER 2. THE FEASIBILITY OF SATELLITE REMOTE SENSING AND SPATIAL INTERPOLATION TO ESTIMATE COVER CROP BIOMASS AND NITROGEN UPTAKE IN A SMALL WATERSHED

Citation: Xu, M., C.G. Lacey, and S.D. Armstrong. 2018. The feasibility of satellite remote sensing and spatial interpolation to estimate cover crop biomass and nitrogen uptake in a small watershed. *Journal of Soil and Water Conservation*. 73(6): 682-692. doi: 10.2489/jswc.73.6.682

2.1 Abstract

The adoption of winter cover crops has been identified as one of the most effective best management practices to reduce non-point Nitrogen (N) loss via subsurface drainage in a watershed in Midwestern Corn Belt. Understanding the variation of cover crop growth and N cycling is vital for watershed modeling efforts that simulate cover crop adoption. However, there is a dearth of watershed cover crop studies that describe the variation in cover crop growth and N cycling and compare the ability of spatial analytical methodologies to predict cover crop biomass and N uptake within diverse agronomic management practices and heterogeneous soil landscapes. Therefore, the objective of this study is to compare the feasibility of satellite remote sensing and spatial interpolation methods to predict cover crop biomass and N uptake in a small watershed (100 – 10,000 ha). This study was undertaken during 2015 - 2017 in the Lake Bloomington watershed (374 ha) in McLean County, Illinois. Within the small watershed, daikon radish/oats (R/O), annual ryegrass/daikon radish (A/R), cereal rye/daikon radish (C/R), were adopted on 78% of row crop land area for both years. Strong linear relationships were observed between soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI), and triangular vegetation index (TVI) and cover crop biomass, (average R^2 of 0.77, 0.78, 0.76, respectively) and N uptake (average R^2 of 0.68 for all 3 vegetation indices). Cover crop biomass and N uptake estimated by the SAVI method were 86% and 85%, 93% and 94%, 107% and 98% of the ground observed value during 2016 spring, 2016 fall and 2017 spring, respectively. Moreover, spatial pattern and different cover crop management fields in the study watershed were also captured by the SAVI method. Ordinary kriging (OK) and inverse distance weighting (IDW) showed similar mean cover crop biomass and N uptake values for fields with cover crops; however, both spatial interpolation methods showed lower prediction R^2 values than that of the SAVI method. The results of this study

suggest that the combination of spatially accurate satellite imagery and limited ground sampling could be used for repeated small watershed assessment of cover crop growth. Furthermore, this can be used to understand cumulative cover crop impacts on soil and water quality in response to conservation practices and weather within a drainage basin.

Key words: Cover crop biomass—cover crop nitrogen uptake—remote sensing—vegetation indices—spatial interpolation—watershed conservation

2.2 Introduction

Estimations of the Gulf of Mexico hypoxic zone in 2017 revealed that it is at its largest since 1985 dead zone mapping (22730 km²), four times of the Gulf Hypoxia Task Force target of 5000 km² (NOAA 2017). Studies have estimated that agricultural land in Mississippi River Basin is responsible for 71% of nitrogen (N) load to the hypoxia zone, among which Illinois, Indiana, Iowa and Missouri account for 48 % of the N load (EPA 2014). Cover crops after summer cash crops have been identified as the most effective in-field management practice for improved soil and water quality in row crop agriculture. Cover crops are vital to reducing N loading to subsurface drainage by scavenging post-harvest soil residual inorganic N, as well as N mineralized from soil organic matter during the fallow season (Dinnes et al. 2002; Tonitto et al. 2006; Constantin et al. 2010; Kaspar et al. 2012). Furthermore, the sequestered N can be recycled to the soil after winter kill or spring termination of the cover crops, with mineralization rate depending on cover crop species (Lacey and Armstrong 2014). Cover crops also exhibit beneficial effects on reduced sediment loss (Yeo et al. 2014), soil aggregate stability (Roberson et al. 1991), and water retention (Quemada and Cabrera 2002). Therefore, the adoption and management of cover crops is vital to reduce non-point N loss via subsurface drainage. However, cover crop growth is affected by weather, landscape positions and the cultural practices of farmers within the watershed. Thus, there is a need to better understand cover crop growth and N cycling variation.

Cover crops are expected to reduce 30% of total N load to the Gulf of Mexico hypoxia zone in multiple Midwest state nutrient loss reduction strategies (Christianson et al. 2016; MONLRS 2014; ILNLRs 2015; IANRS 2016). Diverse physical, environmental, and farm management factors affect the variability of cover crop growth, thus simple arithmetic mean of sparse ground sampling is not adequate. Currently, estimation of cover crop biomass production has been largely based on

extrapolation results from plot experiments to match implementation acreages (Hively et al. 2009). Thus, there is a need for more studies that describe the variation in cover crop growth and N cycling and that compare abilities of analytical methods to predict cover crop biomass and N uptake within diverse agronomic management practices and heterogeneous soil landscapes. Cover crop establishment, biomass production, and nutrient uptake could be affected by greater variability in soil type, landscape position, catchment area, soil drainage, and nutrient management relative to establishment in controlled experimental studies of plots. Thus, it is important to evaluate the accuracy and utility of different spatial analytic approaches to predict and estimate the variability in cover crop growth and N uptake.

Satellite remote sensing provides a tool for rapid estimation of cover crop biomass production. Vegetation indices (VI) based on surface reflectance in the visible and near-infrared are widely used to estimate canopy biomass, leaf area index (LAI), leaf chlorophyll concentration, and crop N status (Haboudane 2004; Gitelson 2004; Hansen and Schjoerring 2003; Prabhakara et al. 2015). Hively et al. (2009) first adopted NDVI derived from SPOT 5 satellite image to estimate cover crop biomass for fields with $> 210 \text{ kg ha}^{-1}$ of vegetation. They also estimated cover crop N uptake by multiplication with species-specific N concentration. Yet, Prabhakara et al. (2015) compared multiple indices and found that percent ground vegetation cover and the range of vegetation biomass greatly affected the performance of VI's. Interference from non-vegetation signal (such as soil and crop residue reflectance) and variable sensitivity of VI's at different biomass range also affected the result. However, there is a dearth of cover crop studies that describe the variation in cover crop growth and N cycling in small watersheds and compare the ability of spatial analytical methods to predict cover crop biomass and N uptake within diverse agronomic management practices and heterogeneous soil landscapes. There is a need to directly evaluate the performance of multiple satellite VI's in small watersheds to increase the accuracy of the prediction of cover crop biomass and N uptake, and compare that to other spatial interpolation methods.

OK and IDW are commonly used interpolation methods to characterize soil and crop spatial variability and interpolate between sampled points at both field scale (Franzen and Peck 1995; Kravchenko 2003; Sajid et al. 2013; Santos et al. 2015) and regional scale (Malone et al. 2014). The estimate at an unsampled location is based on a linear weighted combination of measurements at surrounding locations. The weights of OK are estimated from the semivariance function. With a proper semivariogram model, kriging provides the best linear unbiased estimate (Li and Heap

2014). The weights of IDW method are defined as the inverse function of the distance from the point of interest to the sampled points (Li and Heap 2008). The assumption of IDW method is that points close to each other are more similar in values than points that are far apart. The performance of spatial interpolation methods was found to be affected by data density, spatial distribution, temporal variation and sample size (Li and Heap 2011, 2014).

Estimation accuracy as well as required sample size are key considerations when studying cover crop growth in small watersheds. This study focused on the estimation of cover crop growth over a sampling domain of 100 – 10,000 ha. The objective of this study was to compare the feasibility of satellite remote sensing vegetative indices and of spatial interpolation methods for rapid assessment of cover crop biomass and N uptake with minimal ground truth sample requirements.

2.3 Materials and Methods

2.3.1 Study area

This study was undertaken in the Lake Bloomington watershed in McLean County, Illinois. The fields in this study were under corn-soybean rotation with both crops represented each year. In 2016, corn was planted in 24% of the agricultural land (90.7 ha), and soybeans were planted in 76% of the agricultural land (283.3 ha). N fertilizers were either fall applied (anhydrous ammonia) or spring sidedressed (urea ammonium nitrate) for corn production, with N rates ranging from 160 kg N/ha to 323 kg N/ha. Cover crops were aerial seeded into the standing cash crop each fall (8/28/2015 to 9/9/2015, 9/10/2016 to 9/11/2016). The cover crop species used in this study were oats (*Avena sativa L.*), daikon radish (*Raphanus sativus L.*), annual ryegrass (*Lolium multiflorum*) and cereal rye (*Secale Cereal L.*). The cover crop treatments were daikon radish/oats (R/O) or annual ryegrass/daikon radish (A/R) interseeded within soybean and cereal rye/daikon radish (C/R) interseeded within corn. A no cover crop control (No CC) was also included (Figure 2.1). Seeding rates were: R/O, 40.3 kg/ha oats and 4.5 kg/ha daikon radish; C/R, 77.3 kg/ha cereal rye and 6.7 kg/ha daikon radish; A/R, 28.6 kg/ha annual ryegrass and 5.0 kg/ha daikon radish. Oats and daikon radish plants were terminated by the low temperature in December, and cereal rye and annual rye were chemically terminated using a non-selective herbicide (Glyphosate and 26, 2, 4-Dichlorophenoxyacetic acid) at least 2 weeks before the anticipated planting of the cash crop. The fields in the study watershed were tile drained.

2.3.2 Ground cover crop biomass and N uptake sampling

Aboveground cover crop biomass was sampled in the fall before the first frost (12/12/2015, 11/18/2016) and in the spring before cover crop termination (4/6/2016, 4/11/2017). For every 6 ha, 1m² quadrant was randomly sampled with the sampling location geo-referenced using the iGIS app (Geometryit, 2017) on iPad Pro (Apple, Inc., 2016) devices. No biomass was collected from no cover crop control fields due to minimal weed presence. For 2015 fall, 2016 spring, 2016 fall, and 2017 spring, 25, 31, 31, and 38 aboveground cover crop biomass samples were collected, respectively. Plant samples were oven dried at 60 °C and weighed for dry biomass, and ground to pass through a 1mm sieve. Total percent N was determined with a Flash 2000 NC using a dry combustion method. Cover crop N uptake was calculated by multiplying the percent N by the dried biomass weight.

2.3.3 Satellite remote sensing indices

Landsat 8 Operational Land Imager (OLI) images (30m spatial resolution) covering the study watershed (path: 23, row:32) were downloaded within 3 days from ground sampling dates (imagery dates: 12/15/2015, 4/5/2016, 11/15/2016, 4/8/2017). Since various VI's showed different sensitivity in different plant biomass ranges (Prabhakara et al. 2015), a selection of 9 remote sensing VI's has been used to study cover crop biomass production (Table 2.1). For each GPS-referenced sampling point, B-spline interpolation method was used to extract surface reflectance value of the visible and near-infrared bands using SAGA GIS. Then the 9 VI's were computed using formula in Table 2.1. Linear regression was performed to study the correlation between 9 VI's and cover crop biomass and N uptake. The model goodness of fit R² of the linear regression was used to evaluate the validity of the linear relationship between cover crop biomass and VI's of the 4 sampling seasons in the study watershed.

Leave-one-out cross validation was performed using caret package in R to estimate out-of-sample R² and root mean square error (RMSE) as estimates of extrapolation accuracy.

$$\text{out - of - sample } R^2 = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}$$

where P_i and O_i are predicted and observed value at point i , respectively.

SAVI was found to have slightly higher prediction accuracy for low to medium aboveground biomass compared to EVI and TVI (Shen et al., 2008). In addition, our analysis of EVI, TVI, and SAVI suggested that these indices performed very similarly when estimating both cover crop biomass and N uptake. Thus to reduce redundancy, we selected SAVI to simply demonstrate the comparison of remote sensing VI with the spatial interpolation methods. Continuous cover crop biomass and N uptake map in the study watershed were estimated from SAVI values based on the linear relationship with SAVI. Fall 2015 sampling data was excluded from the comparative analysis due to dense cloud cover (99.18% land cloud cover) that resulted in poor correlations between VI's and cover crop biomass ($R^2 < 0.10$). Therefore, biomass and N uptake estimation maps were only generated for 2016 spring, 2016 fall and 2017 spring season.

2.3.4 Spatial interpolation

OK and IDW were performed only on cover crop adopted fields in this study watershed. Both OK and IDW estimate the value of variable Z at an unknown location x_0 , $Z^*(x_0)$, based on the data from the nearest known locations, $Z(x_i)$ as

$$Z^*(x_0) = \sum_{i=1}^n w_i Z(x_i)$$

where w_i are the weights assigned to each $Z(x_i)$, and n is the number of nearest local neighbors used for interpolation.

In kriging, the w_i values are calculated based on the spatial structure of data distribution, which is represented by sample variogram (Goovaerts 1998). In this study, an omni-directional exponential variogram was used. The variogram parameters (sill, range, and nugget) were estimated using maximum likelihood. Variogram estimation and OK were performed using the geoR package in R. Leave-one-out cross validation was performed to assess the accuracy of kriging estimates.

For IDW,

$$w_i = \frac{(1/d_i^p)}{\sum_{i=1}^n 1/d_i^p}$$

where d_i is the distance between the unknown location and a known location, and p is the power parameter. The performance of IDW is affected by the power function and the number of neighbors (Mueller et al. 2001; Kravchenko 2003). The power function p controls the significance of sampled locations on the unknown locations, and the number of neighbor locations defines how many nearest observations influence the estimation of unknown locations. In this study, IDW was performed using the `gstat` package in R with the power parameter ranging from 1 to 3. The optimum power parameter and number of neighbors were chosen based on prediction R^2 using leave-one-out cross validation.

2.4 Results and Discussion

2.4.1 Linear correlation between ground samples and different satellite remote sensing vegetation indices

Among the 3 sampling seasons, the cover crop biomass was found to be linearly correlated to 9 different VI's in this study (Figure 2.2). NDVI, the most commonly used vegetation index, showed good correlation with cover crop biomass with linear R^2 values of 0.79, 0.67 and 0.74 for 2016 spring, 2016 fall and 2017 spring season, respectively. This was comparable to Hively et al. (2009), who found NDVI could explain 73% of cover crop biomass variation. NDVI saturation was not experienced in our study where biomass ranged from 0 to 3000 kg ha⁻¹, however, SAVI appeared to be superior to NDVI as the inclusion of soil adjusting factor ($L=0.5$) could lessen spectral responses due to background soil reflectance, especially when crop reflectance is minimal. In addition to SAVI, TVI and EVI showed greater correlation with cover crop biomass relative to NDVI. When considering the average of all 3 cover cropping seasons, SAVI, TVI and EVI showed similar linear goodness of fit of with cover crop biomass. More specifically, we observed that SAVI resulted in the best linear predictor in 2016 spring ($R^2 = 0.80$). EVI performed the best in 2016 fall ($R^2 = 0.74$), and TVI was the best in 2017 spring ($R^2 = 0.82$). Prabhakara et al. (2015) found that TVI was most accurate in estimating high ranges of cover crop fall biomass ($R^2 = 0.86$) in their Maryland field study due to the high detectability of TVI on high biomass. Different from

NDVI, TVI does not reach an upper limit, thereby reducing the saturation effect and increasing correlation. They also reported similar linear goodness of fit R^2 values between fall biomass and other indices studied in this study.

Previous researchers have estimated cover crop N uptake by multiplying a constant cover crop N concentration by measured or estimated cover crop biomass (Hively et al. 2009) or using system models based on N supply and demand (Yeo et al. 2014; Lee et al. 2016). However, cover crops N concentration and its variability are affected by N fertilization, weather conditions, soil characteristics, landscape position, and sampling time, all of which are uncontrolled on areas larger than plots (Dean and Weil 2009; Hively et al. 2009). The literature has demonstrated that VI's allow for direct, non-destructive measurements that have been found to strongly correlate with leaf chlorophyll content which is a direct indicator of leaf N status (Tucker and Sellers 1986; Filella et al. 1995; Yoder and Pettigrew-Crosby 1995; Moran et al. 2000; Lemaire et al. 2008; Schlemmer et al. 2013). Thus in this study, we elected to adopt VI's to estimate cover crop N uptake with the intent to capture the variability in plant response to diverse management practices implemented by producers over whole fields. In general, we found that the linear correlation between VI's and cover crop N uptake was weaker compared to biomass (Figure 2.3). SAVI, TVI and EVI resulted in a better positive linear correlation with cover crop N uptake relative to the other indices evaluated. More specifically, in the cover cropping seasons of 2016 spring, 2016 fall and 2017 spring, the greatest explanation of cover crop N uptake variation were explained by SAVI (78%), EVI (57%), and TVI (73%), respectively. Reasons for less sensitivity of VI's to cover crop N uptake relative to above ground biomass could be related to background soil reflectance, crop canopy structure, and crop N partitioning that impact cover crop N concentration, chlorophyll and spectra responses.

Overall, we observed that the relationships between VI's and cover crop biomass and N uptake were less evident in the fall compared to spring. This trend could be attributed to less groundcover from slow growth, leaf yellowing and fall frost damage of cover crop above ground biomass. Prabhakara et al. (2015) pointed out that remote sensing VI's are most sensitive to healthy green vegetation and do not detect yellowed and browned leaves. The extent of leaves with yellowing or frost damage in the early winter could also reduce the correlation between remote sensing VI's and cover crop biomass.

2.4.2 The performance of cover crop estimation using soil adjusted vegetation index

The SAVI, TVI, and EVI showed similar prediction accuracy in cover crop biomass and N uptake (Table 2.2 and Table 2.3). SAVI was used to estimate cover crop production since SAVI was equal or superior to the other VI's while only requiring red and NIR spectra, which is commonly used and can easily be derived using various crop canopy sensors.

For cover crop above ground biomass, the SAVI out-of-sample R^2 and RMSE were 0.77 and 437 kg ha⁻¹, 0.68 and 231 kg ha⁻¹, and 0.79 and 270 kg ha⁻¹ in 2016 spring, 2016 fall and 2017 spring, respectively (Table 2.2). The RMSE were 33.6%, 26.2%, and 45.5% of the 3 seasonal biomass means. Greater percentage of RMSE in the spring was associated with greater range of observed biomass as affected by cover crop species (Table 2.4). For N uptake, the SAVI out-of-sample R^2 and RMSE were 0.74 and 15.3 kg ha⁻¹, 0.49 and 10.4 kg ha⁻¹, and 0.69 and 11.7 kg ha⁻¹ in 2016 spring, 2016 fall and 2017 spring, respectively (Table 2.3). The RMSE were 34.2%, 34.2%, and 55.7% of the 3 seasonal N uptake means. The higher percentages of N uptake RMSE to the observed mean were related to less robust linear relationship between SAVI and N uptake.

The SAVI method showed slight overestimation at low range of cover crop biomass and N uptake and underestimation at the high range across different cover crop treatments during different sampling seasons (Table 2.4 and Table 2.5). For the CR/R treatment, where growing cover crops were present in all three seasons, the means of cover crop biomass estimated at 30-m resolution using the SAVI method was 86%, 106%, and 88% of the ground sample means for 2016 spring, 2016 fall and 2017 spring, respectively. Cover crop N uptake estimated by the SAVI method in the CR/R treatment accounted for 85%, 101% and 90% of the N uptake measured directly by ground sampling in each of the 3 seasons. The SAVI method estimated spring biomass and N uptake means of the R/O treatment were similar with no cover crops, since radish and oats were terminated by the cold temperature over the winter. For cover crops in AR/R treatment, SAVI estimated mean biomass and N uptake were within 10% of observed value, except for biomass in 2017 spring. Standard deviation of estimated biomass and N uptake using SAVI method was also found comparable to those of ground samples across cover crop treatments across 3 sampling seasons (Table 2.4 and Table 2.5).

2.4.3 Spatial interpolation using ordinary kriging and inverse distance weighting

There was no consistent trend for overestimation or underestimation of cover crop biomass and N uptake using OK or IDW method. For OK, the prediction R^2 using leave-one-out cross validation varied from 0.35 to 0.56 for cover crop biomass and from 0.36 to 0.52 for cover crop N uptake during the 3 sampling seasons (Table 2.2 and Table 2.3). IDW showed greater out-of-sample R^2 (0.41 – 0.64 for biomass and 0.40 – 0.62 for N uptake) when compared to OK. Both spatial interpolation methods showed lower R^2 and higher RMSE values compared to the SAVI method. For all cover crop treatments, the estimated mean of cover crop biomass and N uptake in this study watershed using both OK and IDW methods were within 20% of ground observation when there were cover crops standing (Table 2.4 and Table 2.5).

Both OK and IDW estimates at an unknown location were influenced by local neighbors and their weights. Therefore, the accuracy of spatial point interpolation models was dependent on the density and spatial arrangement of data points (Lam 2013; MacEachren and Davidson 1987). For kriging, the quality of the semivariogram model, which captures the major spatial features (sill, range, and nugget effect) of a variable, is vital for estimation accuracy (McBratney and Webster 1986; Goovaerts 1998; Kravchenko 2003). Kravchenko (2003) reported kriging was less precise when a reliable sample variogram could not be obtained from the data. Hughes and Lettenmaier (1981) found kriging offer no clear advantage over conventional least squares when sample size is less than 50. Burrough and McDonnell (1998) suggested that at least 50-100 sample points to achieve a stable variogram for kriging. Webster and Oliver (2001) concluded that variogram derived from sample size less than 50 are often erratic with no evident spatial structure. Thus, with data clustering within different cover crop treatments and cash crop management zone across the study watershed, OK did not capture the spatial pattern and variability of cover crop growth (Figure 2.4 and Figure 2.5). When extrapolating cover crop production in small watersheds, ground sample density as well as biomass clustering due to different agronomic management limited the performance of OK. For this study, according to the literature for a kriging sample size, 50 samples per management zone in the small watershed would be required, which equated to 450 samples and a sampling rate of 1.2 samples per hectare. This would translate to excessive labor, time, analysis and operation cost compared to the sample size used to estimate cover crop growth and N uptake variability at the same scale using the SAVI method. The SAVI method used a total sample

size of 31 and 38 for 2016 and 2017, respectively, which is equivalent to 0.08 samples ha⁻¹ and 0.10 samples ha⁻¹.

For IDW, the number of neighbors and the power parameter have large influence on estimation accuracy (Li and Heap 2008). Researchers have reported that IDW worked well with regularly spaced data, however, it was unable to account for sample spatial clustering (Li and Heap 2008, 2014). In small watersheds, sample clustering is inevitable due to variation in landscape position, cash cropping systems and associated nutrient and tillage management, and cover crop selection from field to field. The clustering effects and variation make it difficult for IDW to capture the spatial pattern of cover crop (Figure 2.4 and Figure 2.5), especially for points on the field borders.

2.5 Summary and Conclusions

Spatial and temporal heterogeneity are important characteristics that influence both cover crop growth and N uptake in small watersheds. In our study we found that depending on the heterogeneity of the cropping system management, remote sensing VI's and spatial interpolation methodologies could be used as tools to rapidly estimate and predict cover crop biomass and N uptake in agricultural cropping systems in small watersheds. One limitation of our study was that the feasibility of using satellite remote sensing VI's is dependent on the available images during the time window of interest that are not obscured by heavy cloud cover. The cover cropping season of fall 2015 was omitted from the analysis due to heavy cloud cover, which restricted the number of cover crop growing periods that we analyzed from 4 to 3. Another limitation was that spatial interpolation methods are sensitive to unequal variance. Possible sources that influence unequal variance in both cover crop biomass and N uptake include multiple uncontrolled variables such as weather, cropping system management, cover crop selection and growth. For rapid assessment of cover crop biomass and N uptake, we performed the spatial interpolation using a non-clustering sampling scheme, instead of increasing sample density in cropping system zones to reduce unequal variance.

Despite these limitations, our study found that SAVI, EVI and TVI were strongly correlated with both cover crop biomass and N uptake in a small watershed for low and moderate biomass and N uptake ranges (0-3000 kg ha⁻¹ and 0-100 kg N ha⁻¹). Moreover, they could be used as successful predictors of cover crop biomass production and N uptake. Cover crop biomass and N

uptake estimated using SAVI were +/- 15% of observed value, and spatial interpolation methods predictions were within 20% of measured biomass and N uptake on cover crop adopted fields. However, the SAVI method showed higher prediction R^2 values than that of OK and IDW.

Spatial variation of the cover crops growth is important for studying nutrient transport fate and nutrient management. Remote sensing indices could capture the spatial pattern as affected by various cover crop and cash crop management systems, which are common in small watersheds. In contrast, spatial interpolation is useful where cover crop and nutrient management practices are relatively uniform. Where diverse agronomic management practices exist, it could be laborious to collect high density ground samples for a spatial interpolation analysis, compared to the remote sensing VI's. For future studies, images from unmanned aerial vehicles that allow for more flexible data collection could be used to advance our understanding of cover crop biomass and N uptake over time across various landscapes.

2.6 Acknowledgements

The authors would like to thank the Illinois Nutrient Research and Education Council for funding. The author would like to show gratitude to Michael Ruffatti (support scientist, Illinois State University, Normal, Illinois), Richard Roth (PhD student at Department of Agronomy, Purdue University, West Lafayette, Indiana), Travis Deppe (Nutrient Loss Reduction Manager, Illinois Corn Growers Association, Bloomington, Illinois), Ben Bruening (hydrologist, Minnesota Department of Agriculture, Rochester, Minnesota), and Houston Miller (MS student at Department of Agronomy, Purdue University, West Lafayette, Indiana) for help with sample collection and preparation.

2.7 References

- Apple, Inc. (2016) Apple iPad Pro 35GB model with Cellular + WiFi options. Cupertino, California. <https://www.apple.com/ipad-pro/>
- Burrough, P. A., and R. McDonnell. 1998. Principles of geographical information systems. Oxford: Oxford University Press.

- Christianson, R., C. Wong, and M. McDonald. 2016. Upper Mississippi Nutrient (Loss) Reduction Strategies - Illinois, Iowa, and Minnesota (White Paper). Center for Watershed Protection and Walton Family Foundation.
- Constantin, J., B. Mary, F. Laurent, G. Aubrion, A. Fontaine, P. Kerveillant, and N. Beaudoin. 2010. Effects of catch crops, no till and reduced nitrogen fertilization on nitrogen leaching and balance in three long-term experiments. *Agriculture, Ecosystems & Environment* 135 (4):268-278.
- Dean, J. E., and R. R. Weil. 2009. Brassica cover crops for nitrogen retention in the Mid-Atlantic coastal plain. *Journal of Environmental Quality* 38 (2):520-528.
- Dinnes, D. L., D. L. Karlen, D. B. Jaynes, T. C. Kaspar, J. L. Hatfield, T. S. Colvin, and C. A. Cambardella. 2002. Nitrogen management strategies to reduce nitrate leaching in tile-drained Midwestern soils. *Agronomy Journal* 94:153-171.
- EPA. 2014. Nutrient pollution: EPA needs to work with states to develop strategies for monitoring the impact of state activities on the Gulf of Mexico hypoxic zone. Report No. 14-P-0348.1-24.
- Filella, I., L. Serrano, J. Serra, and J. Peñuelas. 1995. Evaluating wheat nitrogen status with canopy reflectance indices and discriminant analysis. *Crop Science* 35:1400-1405.
- Franzen, D. W., and T. R. Peck. 1995. Field soil sampling density for variable rate fertilization. *Journal of Production Agriculture* 8:568-574.
- Geometryit. 2017. iGIS (Version 8.3) [Mobile application software].
<https://www.geometryit.com/igis/>
- Gitelson, A. A. 2004. Wide Dynamic Range Vegetation Index for remote quantification of biophysical characteristics of vegetation. *Journal of Plant Physiology* 161 (2):165-173.
- Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of soils* 27:315-334.

- Haboudane, D. 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sensing of Environment* 90 (3):337-352.
- Hansen, P. M., and J. K. Schjoerring. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment* 86 (4):542-553.
- Hively, W. D., M. Lang, G. W. McCarty, J. Keppler, A. Sadeghi, and L. L. McConnell. 2009. Using satellite remote sensing to estimate winter cover crop nutrient uptake efficiency. *Journal of Soil and Water Conservation* 64 (5):303-313.
- Hughes, J. P., and D. P. Lettenmaier. 1981. Data requirement for kriging: Estimation and network design. *Water Resources Research* 17 (6):1641-1650.
- IANRS (Iowa Nutrient Reduction Strategy). 2016. Iowa nutrient reduction strategy. A science and technology-based framework to assess and reduce nutrients to Iowa waters and the Gulf of Mexico. Iowa Department of Agriculture and Land Stewardship. Ames, IA: Iowa Department of Natural Resources, and Iowa State University College of Agriculture and Life Sciences.
- ILNLRs (Illinois Nutrient Loss Reduction Strategy). 2015. Illinois nutrient loss reduction strategy. Springfield, IL: Illinois Environmental Protection Agency.
<http://www.epa.illinois.gov/topics/water-quality/watershed-management/excess-nutrients/nutrient-loss-reduction-strategy/index>
- Kaspar, T. C., D. B. Jaynes, T. B. Parkin, T. B. Moorman, and J. W. Singer. 2012. Effectiveness of oat and rye cover crops in reducing nitrate losses in drainage water. *Agricultural Water Management* 110:25-33.
- Kravchenko, A. N. 2003. Influence of spatial structure on accuracy of interpolation methods. *Soil Science Society of America Journal* 67:1564-1571.
- Lacey, C. G., and S. D. Armstrong. 2014. In field measurements of nitrogen mineralization following fall applications of N and the termination of winter cover crops. *Air, Soil and Water Research*:53.

- Lam, N. S. 2013. Spatial Interpolation Methods: A Review. *The American Cartographer* 10 (2):129-150.
- Lee, S., I. Y. Yeo, A. M. Sadeghi, G. W. McCarty, W. D. Hively, and M. W. Lang. 2016. Impacts of watershed characteristics and crop rotations on winter cover crop nitrate-nitrogen uptake capacity within agricultural watersheds in the Chesapeake Bay region. *PLoS One* 11 (6):e0157637.
- Lemaire, G., C. Francois, K. Soudani, D. Berveiller, J. Pontailier, N. Breda, H. Genet, H. Davi, and E. Dufrene. 2008. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sensing of Environment* 112 (10):3846-3864.
- Li, J., and A. D. Heap. 2008. A review of spatial interpolation methods for environmental scientists No. Record 2008/23. Geoscience Australia, Canberra.
- Li, J., and A. D. Heap. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics* 6 (3-4):228-241.
- Li, J., and A. D. Heap. 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling & Software* 53:173-189.
- MacEachren, A. M., and J. V. Davidson. 1987. Sampling and isometric mapping of continuous geographic surfaces. *The American Cartographer* 14 (4):299-320.
- Malone, R. W., D. B. Jaynes, T. C. Kaspar, K. R. Thorp, E. Kladivko, L. Ma, D. E. James, J. Singer, X. K. Morin, and T. Searchinger. 2014. Cover crops in the upper midwestern United States: Simulated effect on nitrate leaching with artificial drainage. *Journal of Soil and Water Conservation* 69 (4):292-305.
- McBratney, A. B., and R. Webster. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science* 37:617-639.
- MONLRS (Missouri Nutrient Loss Reduction Strategy). 2014. Missouri nutrient loss reduction strategy. Jefferson City, MO: Missouri Department of Natural Resources.
<https://dnr.mo.gov/env/wpp/mnrsc/docs/nlrs-strategy-2014.pdf>

- Moran, J. A., A. K. Mitchell, G. Goodmanson, and K. A. Stockburger. 2000. Differentiation among effects of nitrogen fertilization treatments on conifer seedlings by foliar reflectance: a comparison of methods. *Tree Physiology* 20:1113-1120.
- Mueller, T. G., F. J. Pierce, O. Schabenberger, and D. D. Warncke. 2001. Map quality for site-specific fertility management. *Soil Science Society of America Journal* 65:1547-2558.
- NOAA. 2017. Gulf of Mexico 'dead zone' is the largest ever measured: June outlook foretold New Jersey-sized area of low oxygen. Available from <http://www.noaa.gov/media-release/gulf-of-mexico-dead-zone-is-largest-ever-measured>.
- Prabhakara, K., W. D. Hively, and G. W. McCarty. 2015. Evaluating the relationship between biomass, percent groundcover and remote sensing indices across six winter cover crop fields in Maryland, United States. *International Journal of Applied Earth Observation and Geoinformation* 39:88-102.
- Quemada, M., and M. L. Cabrera. 2002. Characteristic moisture curves and maximum water content of two crop residues. *Plant and Soil* 238:295-299.
- Roberson, E. B., S. Sarig, and M. K. Firestone. 1991. Cover crop management of polysaccharide-mediated aggregation in orchard soil. *Soil Science Society of America Journal* 55:734-739.
- Sajid, A., R. Rudra, and G. Parkin. 2013. Systematic Evaluation of Kriging and Inverse Distance Weighting Methods for Spatial Analysis of Soil Bulk Density. *Canadian Biosystems Engineering* 55 (1):1.1-1.13.
- Santos, D., E. G. Souza, L. H. P. NÓBrega, C. L. Bazzi, and F. N. Queiroz. 2015. Physical properties of soils and soybean yields after planting cover crops. *Engenharia Agrícola* 35 (2):280-292.
- Schlemmer, M., A. Gitelson, J. Schepers, R. Ferguson, Y. Peng, J. Shanahan, and D. Rundquist. 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation* 25:47-54.
- Shen, M., Y. Tang, J. Klein, P. Zhang, S. Gu, A. Shimono, and J. Chen. 2008. Estimation of aboveground biomass using in situ hyperspectral measurements in five major grassland ecosystems on the Tibetan Plateau. *Journal of Plant Ecology* 1(4): 247:257.

- Tonitto, C., M. B. David, and L. E. Drinkwater. 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.
- Tucker, C. J., and P. J. Sellers. 1986. Satellite remote sensing of primary production. *International Journal of Remote Sensing* 7 (11):1395-1416.
- Webster, R., and M. Oliver. 2001. *Geostatistics for Environmental Scientists*. Chichester: John Wiley & Sons, Ltd.
- Yeo, I. Y., S. Lee, A. M. Sadeghi, P. C. Beeson, W. D. Hively, G. W. McCarty, and M. W. Lang. 2014. Assessing winter cover crop nutrient uptake efficiency using a water quality simulation model. *Hydrology and Earth System Sciences* 18 (12):5239-5253.
- Yoder, B. J., and R. E. Pettigrew-Crosby. 1995. Predicting nitrogen and chlorophyll content and concentrations from reflectance spectra (400 - 2500 nm) at leaf and canopy scales. *Remote Sensing of Environment* 53:199-211.

Table 2.1 Definition of remote sensing VI's. Surface reflectance of bands are designated in the formula as B (blue), G (green), R (red), and NIR (near-infrared).

Index	Name	Formula
NDVI	Normalized difference vegetation index	$(\text{NIR} - \text{R})/(\text{NIR} + \text{R})$
GNDVI	Green normalized difference vegetation index	$(\text{NIR} - \text{G})/(\text{NIR} + \text{G})$
SR	Simple ratio	NIR/R
SAVI	Soil adjusted vegetation index ($L = 0.5$)	$[(\text{NIR} - \text{R})/(\text{NIR} + \text{R} + L)](1 + L)$
G - R	Green minus red	$\text{G} - \text{R}$
EVI	Enhanced vegetation index	$2.5(\text{NIR} - \text{R})/(\text{NIR} + 6\text{R} - 7.5\text{B} + 1)$
TVI	Triangular vegetation index	$0.5[120(\text{NIR} - \text{G}) - 200(\text{R} - \text{G})]$
NGRDI	Normalized green red difference index	$(\text{G} - \text{R})/(\text{G} + \text{R})$
VARI	Visible atmospherically resistant index	$(\text{G} - \text{R})/(\text{G} + \text{R} - \text{B})$

Source: Prabhakara et al. (2015).

Table 2.2 Leave-one-out cross validation of linear models between cover crop biomass and remote sensing indices and spatial interpolation.

Assessment Method	2016 Spring [†]	2016 Fall	2017 Spring
Remote Sensing Indices			
SAVI	0.77 (437) [‡]	0.68 (231)	0.79 (270)
TVI	0.76 (446)	0.62 (253)	0.80 (261)
EVI	0.76 (442)	0.71(218)	0.78 (277)
Spatial Interpolation Method			
OK	0.56 (607)	0.35 (329)	0.41 (449)
IDW	0.64 (548)	0.46 (299)	0.41 (451)

[†] Sample mean and sample size: 2016 Spring, mean = 1302 kg ha⁻¹, n = 31; 2016 Fall, mean = 882 kg ha⁻¹, n = 31; 2017 Spring, mean = 593 kg ha⁻¹, n = 38.

[‡] Number outside parenthesis is average out-of-sample R² value, number inside parenthesis is RMSE (kg ha⁻¹).

Table 2.3 Leave-one-out cross validation average out-of-sample R^2 values of linear models between cover crop N uptake and remote sensing indices.

Assessment Method	2016 Spring [†]	2016 Fall	2017 Spring
Remote Sensing Indices			
SAVI	0.74 (15.3) [‡]	0.49 (10.4)	0.69 (11.7)
TVI	0.75 (15.2)	0.47 (10.6)	0.70 (11.4)
EVI	0.72 (15.8)	0.51 (10.2)	0.68 (11.7)
Spatial Interpolation Method			
OK	0.52 (20.8)	0.36 (11.6)	0.37 (16.5)
IDW	0.62 (18.7)	0.54 (9.8)	0.40 (16.1)

[†] Sample mean and sample size: 2016 Spring, mean = 44.8 kg ha⁻¹, n = 31; 2016 Fall, mean = 30.4 kg ha⁻¹, n = 31; 2017 Spring, mean = 21.0 kg ha⁻¹, n = 38.

[‡] Number outside parenthesis is average out-of-sample R^2 value, number inside parenthesis is RMSE (kg ha⁻¹).

Table 2.4 Means and standard deviations of cover crop biomass estimation in the study watershed.

Season		Cover Crop Species [†]			
		A/R	C/R	R/O	No CC
----- Mean of Cover Crop Biomass (kg ha ⁻¹) -----					
2016 Spring	Ground Sampling	-	1681.7	0	-
	SAVI method	-	1442.3	239.3	225.3
	Kriging method	-	1618.4	-	-
	IDW method	-	1674.5	-	-
2016 Fall	Ground Sampling	700.8	668.4	1187.1	-
	SAVI method	677.3	706.6	926.3	35.5
	Kriging method	763.9	796.3	1077.8	-
	IDW method	748.4	785.6	1113.8	-
2017 Spring	Ground Sampling	389.7	1275.0	0	-
	SAVI method	494.2	1121.9	34.2	62.7
	Kriging method	419.9	1090.3	-	-
	IDW method	416.7	1066.7	-	-
----- Standard Deviation of Cover Crop Biomass (kg ha ⁻¹) -----					
2016 Spring	Ground Sampling	-	672.1	0	-
	SAVI method	-	650.8	164.4	262.3
	Kriging method	-	491.4	-	-
	IDW method	-	528.2	-	-
2016 Fall	Ground Sampling	163.6	629.6	244.4	-
	SAVI method	233.6	485.8	281.9	135.0
	Kriging method	114.2	416.1	158.6	-
	IDW method	124.1	472.6	168.3	-
2017 Spring	Ground Sampling	157.5	346.9	0	-
	SAVI method	226.5	283.7	107.1	146.6
	Kriging method	149.7	193.1	-	-
	IDW method	196.5	182.0	-	-

[†] Cover Crop Planting Area: AR/R (136.4 ha, 2016 Fall and 2017 Spring); CR/R (239.3 ha, 2016 Spring; 51.6 ha, 2016 Fall and 2017 Spring); R/O (51.6 ha, 2016 Spring; 103.0 ha, 2016 Fall and 2017 Spring); No CC (83.0 ha, all seasons)

Table 2.5 Means and standard deviations of cover crop N uptake estimation in the study watershed.

Season		Cover Crop Species [†]			
		A/R	C/R	R/O	No CC
----- Mean of Cover Crop N Uptake (kg ha ⁻¹) -----					
2016 Spring	Ground Sampling	-	57.9	0	-
	SAVI method	-	49.4	9.9	9.4
	Kriging method	-	56.2	-	-
	IDW method	-	58.3	-	-
2016 Fall	Ground Sampling	22.7	24.5	41.5	-
	SAVI method	23.9	24.7	31.7	1.4
	Kriging method	24.4	29.1	38.3	-
	IDW method	24.3	28.8	39.1	-
2017 Spring	Ground Sampling	16.7	42.8	0	-
	SAVI method	17.7	38.7	1.5	2.8
	Kriging method	17.1	33.7	-	-
	IDW method	17.1	35.1	-	-
----- Standard Deviation of Cover Crop N Uptake (kg ha ⁻¹) -----					
2016 Spring	Ground Sampling	-	21.0	0	-
	SAVI method	-	21.4	5.4	8.7
	Kriging method	-	15.7	-	-
	IDW method	-	17.6	-	-
2016 Fall	Ground Sampling	5.3	23.1	7.4	-
	SAVI method	7.3	15.4	8.8	4.9
	Kriging method	3.9	14.1	4.3	-
	IDW method	4.1	17.6	5.7	-
2017 Spring	Ground Sampling	6.0	17.4	0	-
	SAVI method	7.6	9.5	3.9	5.2
	Kriging method	5.4	8.1	-	-
	IDW method	7.2	9.5	-	-

[†] Cover Crop Planting Area: AR/R (136.4 ha, 2016 Fall and 2017 Spring); CR/R (239.3 ha, 2016 Spring; 51.6 ha, 2016 Fall and 2017 Spring); R/O (51.6 ha, 2016 Spring; 103.0 ha, 2016 Fall and 2017 Spring); No CC (83.0 ha, all seasons)

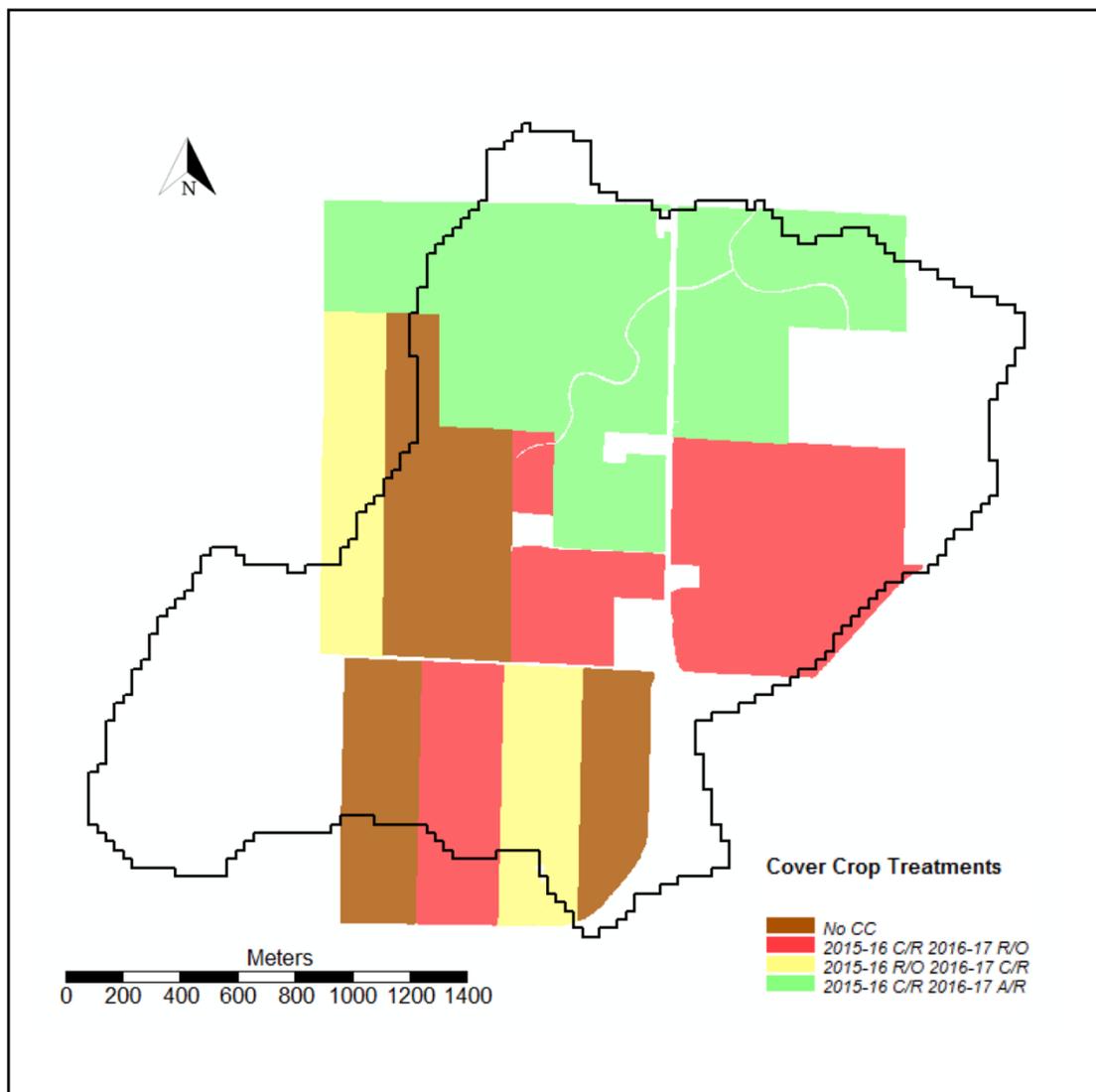


Figure 2.1 Cover crop treatments in the study watershed.

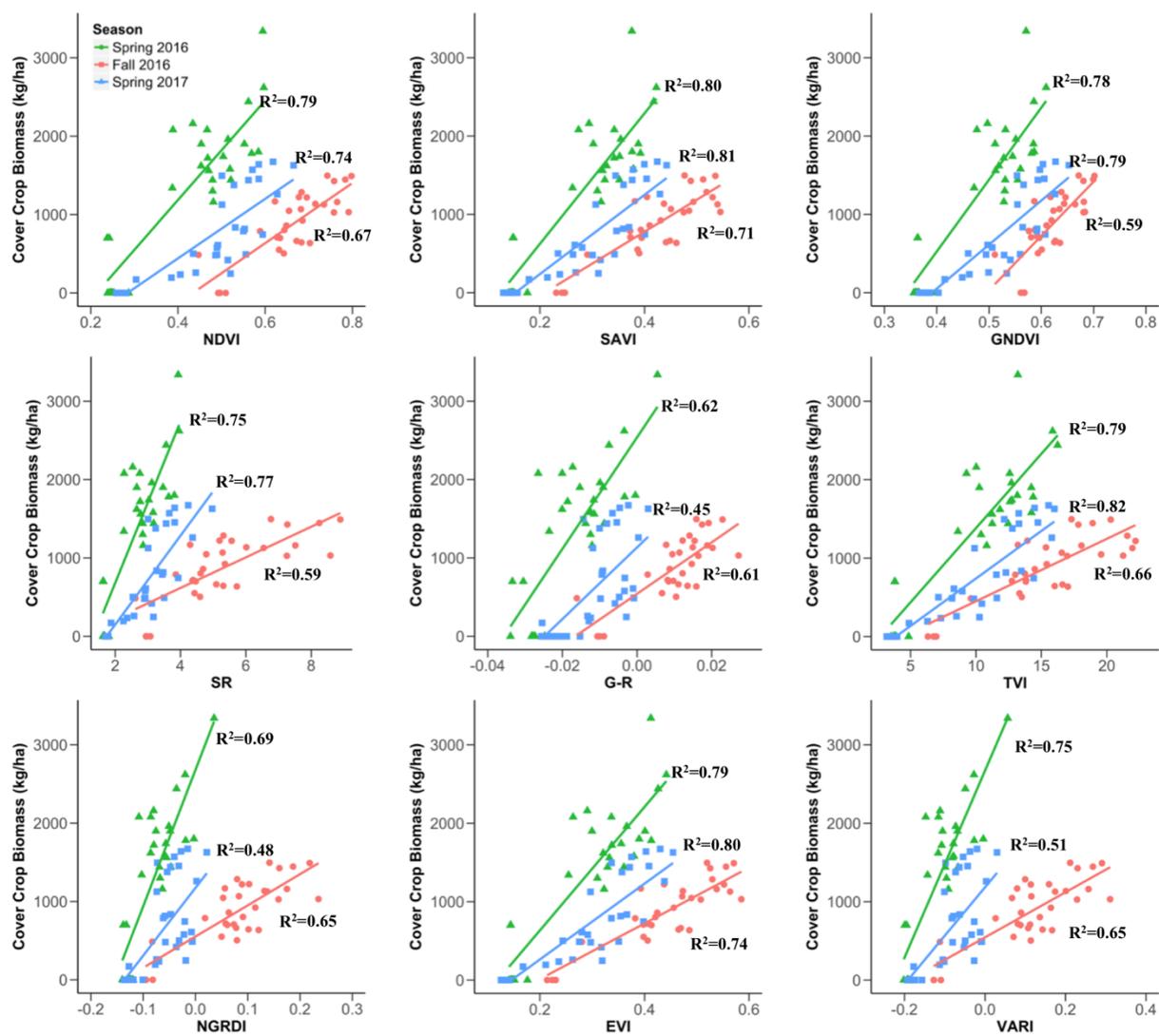


Figure 2.2 Linear relationship between remote sensing indices and cover crop biomass.

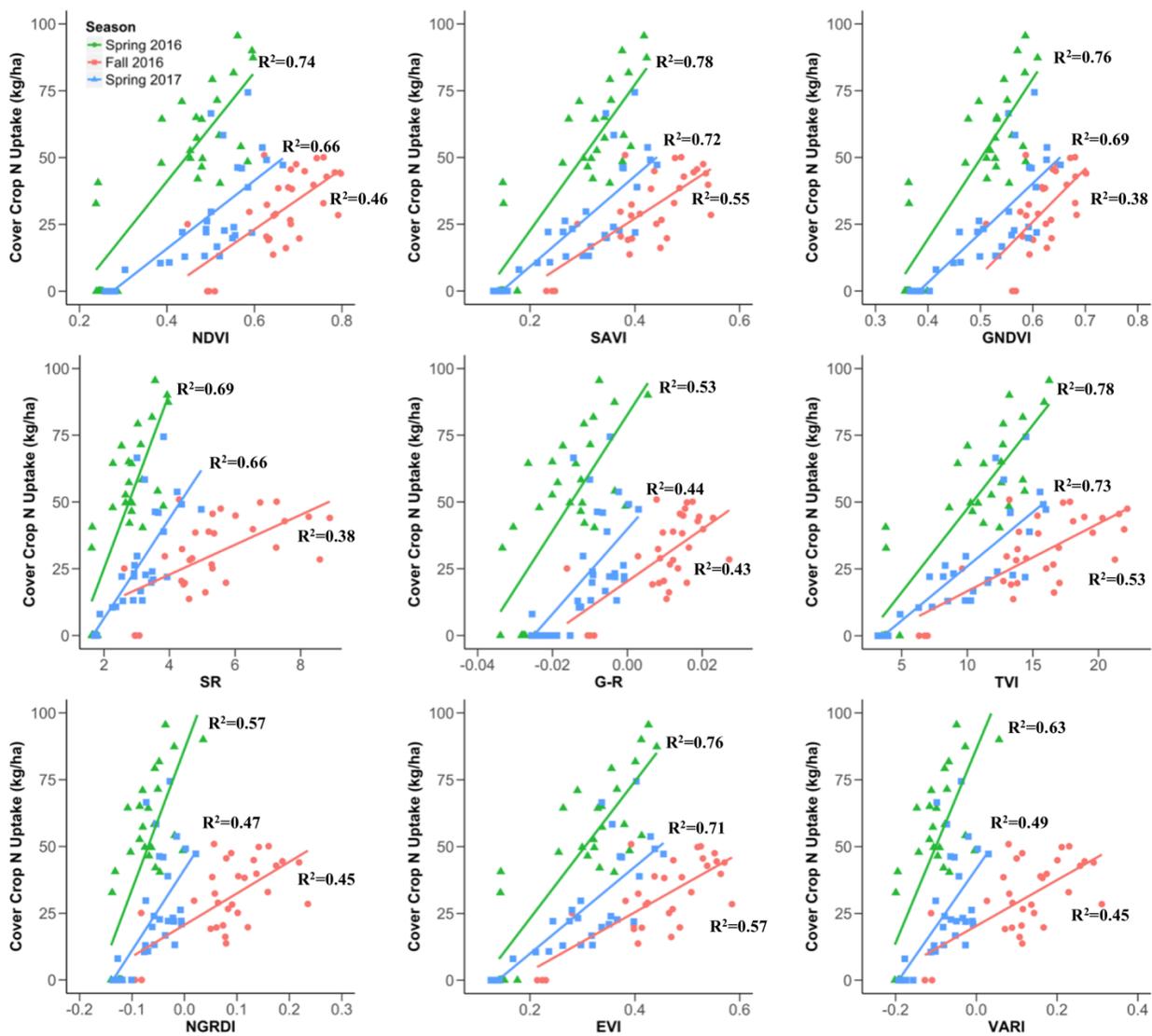


Figure 2.3 Linear relationship between remote sensing indices and cover crop N uptake.

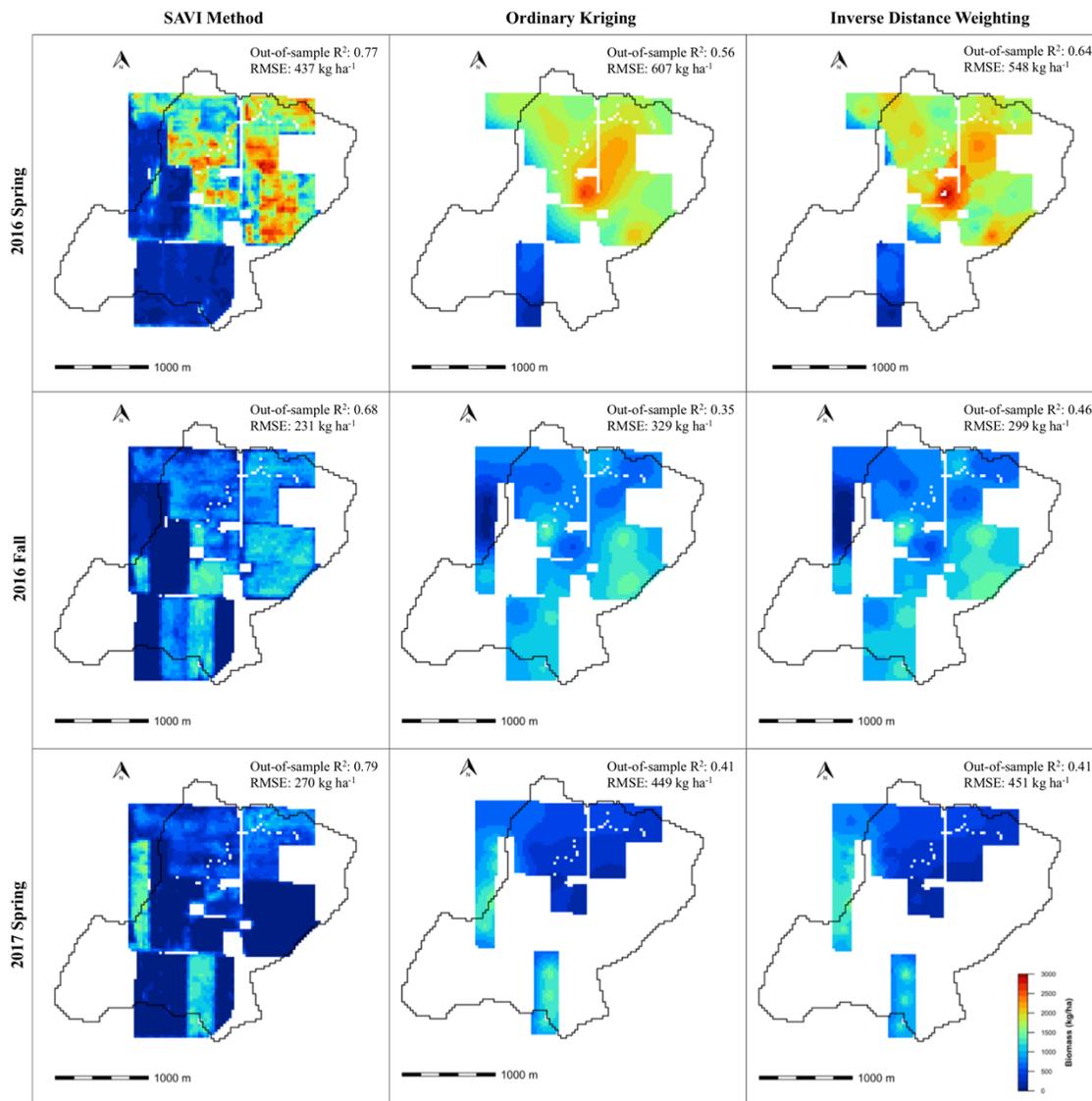


Figure 2.4 Estimated cover crop biomass of the study watershed.

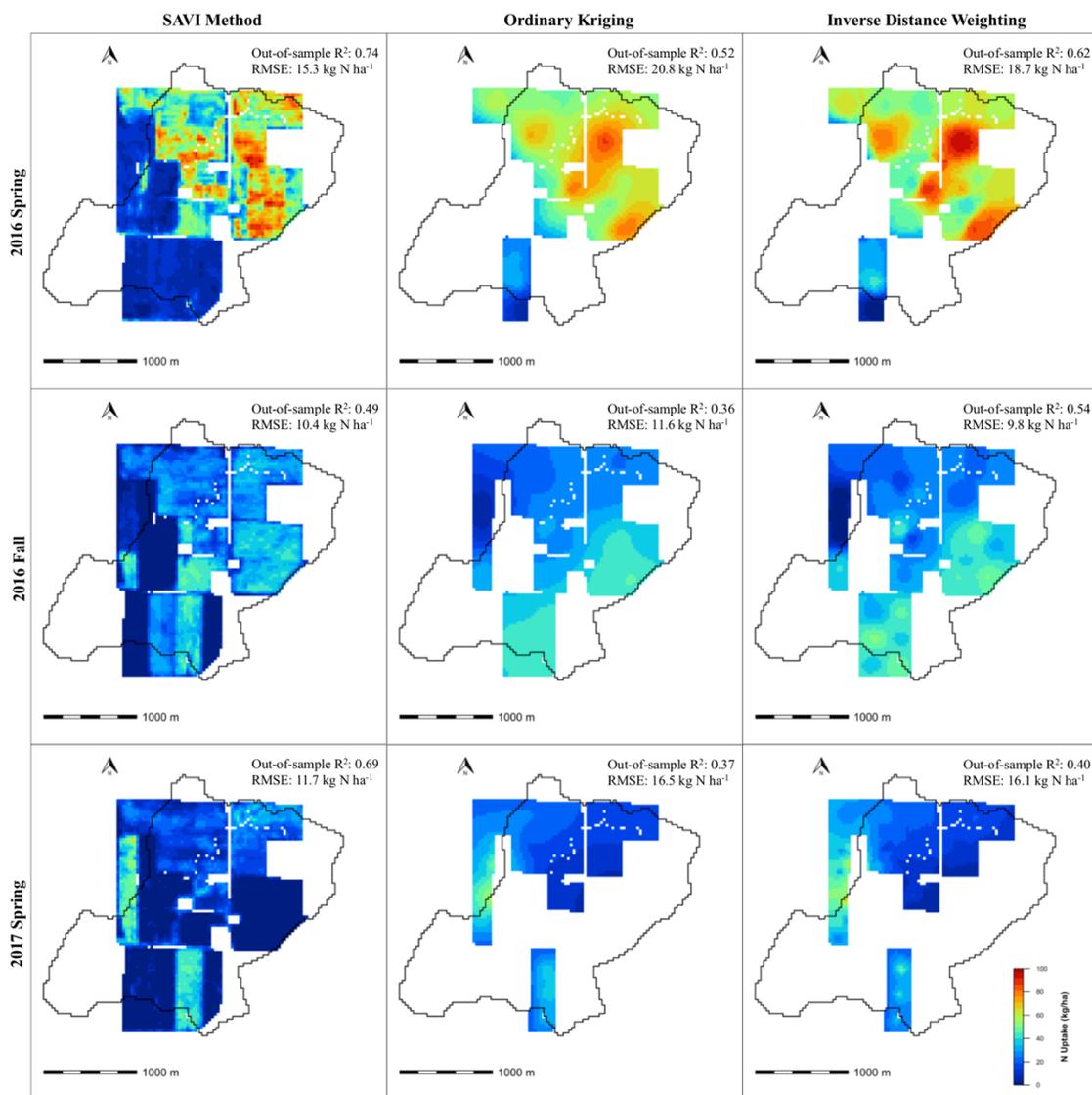


Figure 2.5 Estimated cover crop N uptake of the study watershed.

CHAPTER 3. A NEW MODELING APPROACH: BIG DATA AND MULTIVARIATE SPATIAL STATISTICS TO GENERATE IN-SEASON FIELD SPECIFIC MAIZE YIELD PREDICTIONS

3.1 Abstract

Accurate in-season yield prediction at the field scale is essential to the development of algorithms for improved N rate guidelines for farmers. One great challenge of modeling crop yield at the field scale is spatial and temporal yield variability. In this study, a new approach using the multivariate spatial autoregressive (MSAR) model was developed at 10-m grid resolution to forecast maize yield using timely remote sensing maps, site-specific top 30 cm soil organic matter (SOM) and elevation, while accounting for yield spatial autocorrelation. The calibration dataset contained 35 maize site-years, including 7 fields in Central Indiana during 2010-2016. Principal component analysis (PCA) revealed Landsat Green chlorophyll vegetation index (GCVI) was the variable that most closely associated with grain yield ($r = 0.70$). Thus, GCVI was used as the year-specific variable in the MSAR model. The predictability of the MSAR model was tested on a separate dataset of 82 maize site-years. Relative mean error (RME) and relative mean absolute error (RMAE) were used to quantify the model prediction accuracy at the field level and 10-m grid level, respectively. The MSAR model performed reasonably well for overall field maize productivity in 32 out of 35 site-years of the calibration dataset (absolute RME < 15%) with an average absolute RME of 6.6%. The average RMAE of the MSAR model was 13.1%. In the validation dataset, the MSAR model showed good prediction accuracy overall ($\pm 15\%$ of observed yield in 56 site-years) in new fields when extreme stress was not present. The strength of the MSAR method lies in its ability to use elevation and soil information to interpret satellite observations accurately in a fine spatial scale. The novel approach developed in this study can be used with sensor-based farm management to guide in-season N application and empower farmers to derive value from the big data sets that are routinely collected.

3.2 Introduction

With the fast development of precision agricultural machinery and sensor technology, large volumes of data are routinely collected by farmers. These data are under-utilized due to difficulties

in data processing and interpretation. Better methods for using historical, field-specific data will promote their potential values for use in smart farming (Wolfert et al., 2017). Rapid and accurate in-season crop yield prediction can be used to inform policy, marketing, and insurance decisions at various levels (Basso et al., 2013).

It has long been recognized that maize yield is determined by a combination of soil-plant-atmosphere processes that are highly affected by prevailing climatic conditions, existing status of soil water and N reserves, and biotic stresses such as weeds, insects and disease pressure. Additionally, large temporal and spatial variability is present in crop nutrient demand, soil nutrient supply, economic optimum N rate (EONR), and maize yield due to intra-season weather variability and interactions with soil and landscape properties (Basso et al., 2013). Maize yield variability is expected to further increase under climate change (Urban et al., 2012). Currently, there is no reliable method to account for such variability and, therefore, the ability to predict maize yield and N rates at the field scale is limited (Morris et al., 2018). While county and regional level models are useful to inform policy and market decisions, field-scale models that can predict crop yield response to agronomic management practices are needed to inform data-driven crop management for farmers and their advisors. In this study, field scale is defined as individual fields as managed by farmers.

The U.S. Department of Agriculture (USDA) yield forecasting and estimating program provides monthly maize yield estimates at the county level based on field measurements and surveys beginning August 1st to November 1st (USDA-NASS, 2012). However, this approach takes significant resources such as labor and time. Another approach for maize yield prediction is the use of statistical models. Statistical models are among the earliest models that were used in yield predictions at regional scales (Thompson, 1969). They can provide insights of crop yield trend under climate change (Lobell et al., 2013). Statistical models can be used to efficiently predict regional-scale yield with less complex computing (Basso et al., 2013). Yet, results of statistical models cannot be extrapolated to other areas and time due to variation in soils, landscape, and weather. Moreover, the spatial resolution and measurement error of commonly used agrometeorological input variables (e.g. temperature and rainfall) of statistical models have limited their applicability at the field scale.

In recent years, computer crop simulation models have been developed to forecast maize yield and evaluate agronomic consequences of crop management practices with the growing availability of data. Rather than regional yield estimation, crop models can be used at various scales including individual fields. Crop models simulate dynamic soil and crop processes based on soil, crop management, and weather data and have demonstrated reasonable prediction of maize yield from county level to regional level (Morell et al., 2016). However, due to the uncertainty and similarity of input parameters at the field scale, there is a great risk associated with prediction capability of maize yield at the field and sub-field scale (Cheng et al., 2016). Furthermore, particular skills and extensive data requirements for model calibration of cultivar coefficients and other input parameters, and lack of representative weather and soil data have constrained the use of crop models to predict maize yield response to N at finer scales (Puntel et al., 2016). Additionally, crop models usually assume homogeneity at the field scale, which can lead to considerable uncertainty when predicting nutrient requirement and maize yield. Due to the spatial and temporal heterogeneity of the dynamic interactions among management, soil and other environmental factors, the universal parameters settings for crop models to accurately predict maize yield may be unsatisfactory (Morris et al., 2018).

Advances in remote sensing data acquisition and processing offer promise for predicting agricultural production remotely across a range of spatial scales. Using remote sensing data to forecast crop yields at the field scale is particularly useful for understanding how crop growth responds to management and environmental factors (Lobell, 2013). Crop leaves absorb red and blue light by chlorophyll, but reflect green and near-infrared (NIR) radiation. Vegetation Indices (VIs) are mathematical combinations or ratios of spectral bands which are used to describe relationships between dynamic crop biophysical status and remote sensing observations (Benedetti and Rossini, 1993; Doraiswamy et al., 2003; Kogan et al., 2012; Bolton and Friedl, 2013). Blackmer et al. (1995) showed that maize canopy reflectance changed with N rates, and grain yield was correlated with the reflected light. Research efforts on various VIs have provided potential applications to site-specific forecasting of crop yield. Nevertheless, most of them rely on calibrated relationships between VIs and yields that are specific to locations and years, with ground measurements needed for each time (Baez-Gonzalez et al., 2002). To solve this issue, an approach called scalable satellite-based crop yield mapper (SCYM) was developed. It uses plant biomass generated by crop models to calibrate VIs, and then translates VIs to crop yield (Lobell et al., 2015;

Jin et al., 2017). However, the accuracy of SCYM depends on the accuracy of the crop model being used.

Precision farming involves crop management of high-yielding and low-yielding zones within the field. Accurate in-season yield prediction is the first step in the development of algorithms for in-season N management. There is a great need to develop a method that can rapidly and precisely capture spatial variability of grain yield. A common challenge of the above-mentioned crop yield forecast methods is the lack of modeling of spatial yield heterogeneity at the field scale. To address this problem, Colonna et al. (2004) used the spatial autoregressive (SAR) model to deal with spatially correlated residuals in regression analysis of soybean yield in a 20-acre field in East Central Illinois from 1999 to 2001. They showed that the SAR model gave consistent unbiased estimates of soybean yields. However, one limitation of the SAR model is that it does not accommodate variables that change with time.

In this study, a new approach using the MSAR model was developed to forecast field-scale maize yield using historical yield monitor data, timely remote sensing maps, and soil and landscape properties. The algorithms aim to account for the spatial heterogeneity feature to predict maize yield of individual fields. The objectives were to (1) identify factors that relate to maize yield at the field scale using historical yield data from individual fields, and (2) to forecast maize yield at the field scale by combining in-season satellite remote sensing VI and soil and landscape properties using a MSAR model. The results of our model can be used with sensor-based farm management to guide in-season N application and empower farmers to derive useful information from data they routinely collect.

3.3 Materials and Methods

3.3.1 Field N management and maize yield

For this study, estimates of model parameters were derived using a calibration dataset and the performance of the model was tested on a validation dataset. The calibration data used in this study are from 2010 to 2016 on 7 fields in south Cass and north Carrol county, Indiana (N 40°37'26", W 86°25'20"), USA. The main soil information is outlined in Table 3.1. Although crop rotation varied, either maize or soybean was planted with each field having 5 years of maize between 2010 and 2016 (See Table 3.2 for details). Maize planting time varied from mid-April to mid-May over

the studied time period (Table 3.3). The validation data included 30 fields in the same region with a total of 82 maize site-years from 2010 to 2016 (Table B.1). For all fields in the calibration and validation datasets, N fertilizer was applied in the form of urea ammonium nitrate at the rate of 218 kg N ha⁻¹ for maize following soybean and 252 kg N ha⁻¹ for maize following maize. Fertilizers were applied at the time of planting unless otherwise noted (Table 3.3). Additionally, all sites received 50 kg N ha⁻¹ as starter fertilizer for a total N rate of 268 kg N ha⁻¹ for maize following soybean, and 302 kg N ha⁻¹ for maize following maize. The recommended agronomic optimum N rate (AONR) for this region (northcentral Indiana) using field trials with sidedress N application was 237 kg N ha⁻¹ for maize following soybean. Furthermore, paired trials of crop rotations from 2007 to 2010 showed that the average AONR for maize following maize was 49 kg N ha⁻¹ greater than for maize following soybean (Camberato and Nielsen, 2017).

Maize yield monitor data were gridded to 10m x 10m by averaging harvest points within each grid and were reported at 15.5% moisture. Yield values more than 3 standard deviations away from the field mean yield were removed and then replaced by the mean of its nearest queen neighbors (Figure 3.1).

3.3.2 Weather and crop data

Daily weather variables (max/min air temperature, precipitation, solar radiation, and relative humidity) were obtained from the National Weather Station-Cooperative Observer Network (NWS-COOP) and all fields in this study shared the common nearest weather station (Logansport, IN). Maize maturity was assumed to be 1500 growing degree days (GDD, base 10 °C) as Relative Maturity records of planted maize hybrids were not available. Four growth stages were distinguished: early vegetative (EV, 0-483 GDD), late vegetative (LV, 484 -631 GDD), critical period (CP, 632-922 GDD), and effective grain filling (GF, 923-1361 GDD). Daily nighttime temperature was calculated as the mean of daily average temperature and daily minimum temperature. Mean values of daily temperatures (maximum, minimum, average, and nighttime) and vapor pressure deficit (VPD) of each of the four growth stages were computed separately. VPD is the difference between the air water vapor pressure and the saturated water vapor pressure, and depends on air temperature and humidity. High VPD increases plant water demand for sustaining diffusion of CO₂ into leaves (Lobell et al., 2013). VPD has been widely used as a drought measure of the atmosphere drying force on plants.

Cumulative sum of precipitation, solar radiation, and killing degree days (KDD) were calculated for each of the four growth stages. For KDD, 29 °C was used as the threshold to quantify temperatures that may reduce maize yields since studies have found that maize yields started to drop when maximum temperatures exceeded 29 °C, especially during silking and early grain filling stages (Schlenker and Roberts, 2009; Butler and Huybers, 2015). Daily KDD is calculated by,

$$KDD_d = \begin{cases} T_{max} - 29, & \text{if } T_{max} > 29, \\ 0, & \text{if } T_{max} \leq 29. \end{cases} \quad (1)$$

KDD is considered as a good measure of cumulative evaporative demand during crop growth (Urban et al., 2015). High KDD is expected to reduce mean yields while increasing variability by increased desiccation, accelerated crop development or direct damages to plant tissue or enzymes (Butler and Huybers, 2013).

3.3.3 Field and soil attributes

Digital elevation model (DEM) and soil properties in 10m x 10m grids were obtained for the fields from the University of Missouri Web-based clipper application (<http://clipper.missouri.edu>). Soil organic matter data was extracted from the USDA-NRCS SSURGO database and gridded to 10-m spatial resolution. Top 30 cm soil organic matter (SOM) amount was calculated using soil bulk density and organic matter content of each soil horizon (Table B.2). Top 30 cm was selected because it is the common sampling depth for soil N determination.

3.3.4 Landsat remote sensing vegetation index

Remote sensing imagery of Landsat 7 ETM+ (2010 – 2012, revisit time: 18 days) and Landsat 8 OLI (2013 – 2016, revisit time: 16 days) (Path: 21/22, Row: 32) were obtained during the maize growing season (May – October). Green chlorophyll vegetation index (GCVI, Gitelson et al., 2003) was computed to assess maize plant growth and yield development.

$$GCVI = NIR/Green - 1, \quad (2)$$

Studies have found that GCVI was positively related to maize leaf area index (LAI) ranging from 0 to 6.5, canopy N content, and survey-based yield (Gitelson et al., 2003; Nguy-Robertson et al., 2012; Schlemmer et al., 2013; Burke and Lobell, 2017). The advantage of GCVI over other common indices is that GCVI does not saturate at high LAI values and remains sensitive to

variation in LAI for denser canopies commonly found in the Midwest (Lobell et al., 2015). For each season, Landsat images were used when the median GCVI values for each field are the maximum during the growing season in this study, which commonly occurred during maize R1 to R2 development stage.

3.3.5 Yield correlations with weather and remote sensing indices

Maize yield was averaged by soil type (defined by a map unit of SSURGO) on each field for each maize year as the soil polygon was the smallest spatial unit. Principal component analysis (Zhang and Yang, 2018) was conducted in R using the stat package (R Core Team, 2017) to identify variables that were associated with grain yield. Variables used in the PCA included cumulative precipitation, cumulative solar radiation, cumulative KDD, and average VPD for each of the four growth stages identified above and GCVI. A biplot with points representing the scores of observations and vector representing the loadings of variables on the first two principal components was created to represent the data structure. The points on the biplot are helpful to highlight groups of homogeneous observations when considering all variables at the same time. The variable vectors on the biplot can be used to interpret correlation between variables. The cosine of the angle between any two variable vectors approximates their correlation. The lengths of the variable vectors are approximately proportional to the standard deviations of the variables, and they indicate how well the variables are represented by the graph. Moreover, the larger the projection of a point on a variable vector, the more this observation deviates from the average in the variable (Kroonenberg, 2008). Scatterplots were also used to explore nonlinear relationships between temperature variables and maize yields. Pearson correlation coefficients (r) were computed where appropriate to quantify relationships between variables indicated as meaningful by the PCA. For nonlinear relationships, a regression model was used and R^2 values were reported.

3.3.6 Spatial autocorrelation analysis

Maize yield in the fields often exhibits certain spatial patterns. In this study, existence of spatial clusters of grain yield was tested by the global Moran's I statistic with queen neighbor weight matrix. Moran's I statistic is commonly used as a measure of overall spatial autocorrelation (Fu et al., 2014). Moran's I is calculated as

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (i \neq j), \quad (3)$$

where n is the number of observations in the field of interest, x_i and x_j are the observations at locations of i and j , \bar{x} is the mean of observation, and w_{ij} is the spatial weight between locations of i and j . The theoretical range of Moran's I is between -1 and 1 (Zhang and Lin, 2016). Large positive Moran's I value indicates positive spatial autocorrelation, where the target value is similar to its neighbors. Large negative Moran's I value means negative spatial autocorrelation, where the target value is different with its neighbors (e.g., a high value in a low value neighborhood). A Moran's I value close to 0 implies spatial randomness (Tu and Xia, 2008).

3.3.7 Multivariate spatial autoregressive model

The MSAR model was developed to predict maize yield at a 10-m grid resolution using site-specific and year-specific variables while accounting for yield spatial autocorrelation. The MSAR model is defined as

$$y_{i,j} = \rho \mathbf{W} y_{ij} + \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{X}_j \boldsymbol{\beta}_j + \epsilon, \quad (4)$$

where $y_{i,j}$ is maize yield at site i in year j , ρ is the autocorrelation parameter ($-1 < \rho < 1$), \mathbf{W} is the queen neighbor weight matrix, \mathbf{X}_i is site-specific terrain values (elevation and 0-30 cm SOM), \mathbf{X}_j is year-specific GCVI values, and ϵ is the error term with $N(0, \sigma^2)$ distribution. The parameters of this model include ρ , $\boldsymbol{\beta}_i$, $\boldsymbol{\beta}_j$ and σ^2 . Queen neighbor weight matrix was used as it is the commonly used spatial weight for a regular grid of points in agricultural studies (Colonna et al., 2004). GCVI was the only variable that varies over time to reduce model overfitting issues. The interpretation of the MSAR model is that the response of a variable depends not only on the explanatory variables, but also on its neighboring units. The maximum likelihood estimator of ρ was derived using the Newton-Raphson algorithm. Parameters of the MSAR model were estimated by the maximum likelihood method conditioning on ρ .

3.3.8 Model goodness of fit

Since there is no single method to best assess the goodness of fit, both graphical and numerical methods were used to highlight different features of model performance (Archontoulis and Miguez, 2015). Graphical comparisons were used to visually assess the goodness of fit. Numerical

statistical indices, such as mean error (ME) and mean absolute error (MAE), were used in model evaluation. Relative error term, such as RME and RMAE, were used to compare model goodness of fit among different fields or years. In this study, absolute RME less than 15% was considered as good prediction overall, and RMAE less than 15% was considered as good prediction at the 10-m pixel level. The distance between RMAE and the absolute value of RME indicates the direction of model yield predictions. For instance, a closer distance suggested the model yield prediction was mainly underestimation (positive RME) or overestimation (negative RME), while a greater distance suggested the error originated from both underestimated and overestimated sites in the field.

$$ME = \bar{y} - \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$RME = \frac{ME}{\bar{y}} \times 100\% \quad (7)$$

$$RMAE = \frac{MAE}{\bar{y}} \times 100\% \quad (8)$$

where n is the number of data points, y_i and \hat{y}_i are the observed and predicted values, respectively, and \bar{y} is the mean of observed value.

3.4 Results and Discussion

3.4.1 Overview of maize yield

The study period (2010 – 2016) exhibited diverse weather conditions during the maize growing seasons (Table 3.4). As a result, maize yield varied annually by soil types within a field and among fields (average maize yield by soil type within a site-year ranged from 7.1 to 15.5 Mg ha⁻¹). To characterize the growing season environment and their impact on yield, principal component analysis serves as a useful tool to visualize the correlation between grain yield and environmental variables. The PCA resulted in data points being clustered into 4 groups, with observations within 2010, 2011, and 2012 clustered by year, and observations from other years clustered into one group.

The 2010 data cluster was characterized by a large amount of rainfall during EV and GF stages, the 2011 data clustered by high GF precipitation and warmer CP temperatures, and the 2012 clustered by warmer CP temperature and low precipitation during the growing season, especially in the LV stage (Table 3.4). The results suggest that the variation in weather patterns during the growing season was a major driver for grain yield variation, despite soil type differences among fields (Figure 3.2). Therefore, large uncertainty may exist in in-season maize yield prediction from crop models when using historical average weather record due to the large annual weather variations. In the combined cluster by year 2013 through 2016, greater variation along the grain yield vector direction was found in years 2013 and 2015, which corresponded to their larger standard deviation of grain yield between fields (2013: 1.9 Mg ha⁻¹; 2014: 0.2 Mg ha⁻¹; 2015: 1.4 Mg ha⁻¹; 2016: 0.9 Mg ha⁻¹). This could be due to highly variable soil N dynamics caused by different N application timing, different soil N mineralization and N losses through leaching and denitrification as affected by warm temperature due to late planting and high rainfall during the first half of growing season.

3.4.2 Yield correlations with weather and remote sensing indices

Our PCA suggested that GCVI was the variable most closely associated with grain yield (Figure 3.3, $r = 0.70$), as remotely sensed GCVI can provide a measure of crop canopy state as well as both spatial and temporal information. This high correlation is attributed to the fact that GCVI captured differences in nutrient stress that was correlated with yield (Burke and Lobell, 2017). Another variable that was highly correlated with grain yield was GF KDD (Figure 3.3, $r = -0.62$). For every 1 KDD accumulation during the effective grain filling stage, a 0.3% grain yield decrease was found in our study. Butler and Huybers (2015) found that with the high temperature sensitivity during the grain filling stage, KDD accumulation could increase respiration losses, accelerate the growth phase, and lower final kernel weight. In comparison, the relationship between growing season total KDD and grain yield was less evident ($r = -0.48$). A higher maize yield reduction rate (0.7%) was reported for each KDD unit during the growing season (Auffhammer and Schlenker, 2014). A quadratic relationship was found between grain yield and GF average temperature (Table 3.5), and the optimum GF average temperature was around 23 °C.

During the critical period of kernel setting, maize grain yield was negatively associated with KDD and VPD (Figure 3.3, $r = -0.46$ and -0.49 , respectively). It is also important to note that KDD

and VPD were positively correlated to each other during the CP ($r = 0.77$). Lobell et al. (2013) found that low yield was associated with KDD through the link of KDD to increased VPD, not via a confounding effect with rainfall amount. In our study, the yield reduction due to high CP VPD was apparent only at the highest VPD levels (Figure 3.3). These high CP VPD levels were the result of low precipitation and warm temperature during the 2012 growing season. Similarly, Lobell et al. (2014) found that rainfed maize yield was negatively correlated with July (i.e. generally coinciding with the CP) VPD in the Midwest. The correlation between GF KDD and grain yield was stronger than that of CP KDD and grain yield. The relatively stronger correlation between GF KDD and grain yield could be attributable to the greater negative impact of heat stress on the length of effective grain filling (shortened up to 8 days) and final kernel weight during GF than in the CP itself (Ederira et al., 2014). However, CP KDD and GF KDD also showed a high linear correlation ($r = 0.92$). Since the data used in this study lack explicit experimental design, it wasn't possible to verify causal relationships (although sensitive factors associated with productivity were identified).

Grain yield was weakly associated with precipitation during each of the four growth stages despite the great range of rainfall variation in maize growth stages across years (Table 3.4 and Table 3.5). A quadratic relationship was found between maize yield and seasonal total precipitation ($R^2 = 0.40$) and the optimum seasonal precipitation was found to be 479 mm. This optimal value is similar to the 462 mm reported by Meng et al. (2016) for modern long-maturity maize varieties in the Chinese Maize Belt. The relatively weak response of grain yield to precipitation compared to KDD and VPD in this study could be related to the greater response of plant water deficit to warming than precipitation reduction found by Lobell et al. (2013). Precipitation measurements are subject to error. This error may be partially responsible for the weak relationship between grain yield and precipitation found in both statistical and process-based crop models (Schlenker and Roberts, 2009; Lobell et al., 2013). The prevalence of this weak relationship in the models based in the Midwest suggests that grain yield's weak response to rainfall is also a basic feature of the normally high maize productivity in the region resulting from adequate rainfall in most years except for 2012 (Table 3.4).

3.4.3 Yield spatial structure and performance of the MSAR model

Moran's I coefficient indicated significant spatial autocorrelation structure of maize grain yield (Table 3.6), suggesting the existence of spatial clustering, where high (or low) grain yield values appeared in geographic groups in the field. Moran's I value found in the calibration dataset of 35 site-years ranged from 0.15 to 0.85, varying considerably across fields and across growing seasons. Grain yield was negatively associated with Moran's I (Figure 3.4, $r = -0.78$), suggesting high yield spatial heterogeneity under unfavorable growth environments and low yield spatial heterogeneity under favorable growth conditions. Years that are considered very favorable for crop growth (2014 and 2016) were associated with smaller Moran's I value, while years with either inadequate precipitation (2012) or too much rainfall (2010 and 2015) overall showed larger Moran's I value. This wide range of Moran's I values indicated the spatially varying grain yield response to uniform applied N fertilizer was due to different soil N supply and water stress levels result from interactions of different soil type, landscape positions and weather characteristics. For example, the difference in soil N supply between a high and a low OM soil is likely to be greater in a growing season with low N loss potential compared to that when N loss is high. Wong and Asseng (2006) found that on a sub-field scale, when N was adequate for crop growth, wheat yield spatial variability increased with seasonal rainfall as sites with higher plant-available water conserved more water in wet seasons to give higher yield response than sites with low plant-available water. Therefore, models ignoring the observed spatial patterns of grain yield and assuming complete spatial randomness (uniform yield patterns) could lead to biased estimates for grain yield on different soils in a field.

In this study, the MSAR model used for maize yield prediction accounted for the spatial autocorrelation parameter ρ , soil top 30 cm SOM information, and landscape variable elevation. Landsat GCVI was also included in the MSAR model since GCVI was the most correlated variable across growing seasons. Moreover, the spatial resolution of GCVI (30m) was useful to quantify crop status at the field scale during the growing season. To reduce the chance of model overfitting, GCVI was used without interactions with weather variables in the MSAR model, even though multiple weather variables were identified above to have association with maize grain yield. Additionally, the confounding effects between GCVI and weather variables could lead to model multicollinearity and prediction errors. For instance, GCVI was found to be negatively associated

with GF KDD ($r = -0.77$). Therefore, in this study, GCVI alone was used to represent the result of crop growth status response to growing season weather conditions.

For overall field-level yields, the MSAR model performed reasonably well in 32 site-years (absolute RME < 15%, Figure 3.5), and the MSAR model was significantly superior to the corresponding linear regression model for 24 site-years among the 35 site-years. (Table 3.6). The average absolute RME of the MSAR model across the 35 site-years was 6.6%. For 10-m pixel yield predictions, the MSAR model performance was considered good in 22 site-years and the MSAR model resulted in lower average RMAE in 21 site-years compared to the linear regression model, with the average RMAE of the MSAR model being 13.1%. The spatial autocorrelation parameter ρ did not vary greatly from field to field (ranging from 0.77 to 0.91) for 7 of the studied fields, even though Moran's I showed a wide range both across fields and years. The relative prediction error we found in this study was smaller compared to the results found by Morell et al. (2016) at the county scale using crop models. They also found that model accuracy improved when upscaling from field scale to larger spatial domains. These results suggested that MSAR model was appropriate to be used for grain yield prediction across years.

3.4.4 Prediction accuracy of the MSAR model

To test the prediction accuracy, we tested the ability of the MSAR model to forecast grain yield (a) in a drought year (2012), (b) in a typical year (2016), and (c) at different fields across years. For (a) and (b), the dataset except the forecast year was used to calibrate MSAR model coefficients for the 7 fields used in the calibration, and then the coefficients were used to forecast the maize grain yield for the particular year. For (c), the average of the MSAR coefficients of the 7 calibration fields (35 site-years) was used to forecast maize grain yield in 2010 – 2016 for 82 site-years (10 site-years with maize as the previous crop, 72 site-years with soybean as the previous crop) on 30 different validation fields (15 fields under maize-soybean rotation: ms; the other 15 fields had multiple maize – soybean rotation: mms) in the same region. The exact number of maize years in mms rotation varied from 2 to 3.

3.4.4.1 The MSAR model performance in a drought year

For the drought year of 2012, the MSAR model showed satisfactory prediction of overall field yield except on field 6. However, only 2 out of 5 fields were within 15% RMAE at 10-m pixel

level (Table 3.7). As expected, the predicted maize yield was generally greater than actual yield. This could be explained by the lack of drought stressed grain yields in the training dataset.

3.4.4.2 The MSAR model performance in a typical year

The 2016 growing season was considered a typical year of central Indiana (Table 3.4) with favorable weather conditions for maize growth. The predicted maize yield in 2016 using MSAR model showed a strong agreement with actual yield for all 7 fields (Table 3.7). The predicted maize yield was mostly lower than actual yield, likely due to favorable growth condition after the GCVI acquisition date (typically late July to early August) that promoted kernel growth.

3.4.4.3 The MSAR model performance on new fields

The MSAR model was used to predict maize grain yield in 82 site-years from fields that are different from the data used to derive model coefficients. These fields were located in the same region as the fields in the calibration dataset, and the growing season characteristics were very similar to the weather conditions discussed above. As the N rate application followed a similar routine among all of these fields (268 kg N ha⁻¹ for maize following soybean, 302 kg N ha⁻¹ for maize after maize), N rates were not considered as a variable in this study. Among all 82 site-years, MSAR model grain yield predictions were $\pm 15\%$ of actual yield in 56 site-years (Figure 3.6). At the 10-m pixel level, RMAE value was found to be less than 15% in 45 site-years. The inferior grain yield prediction for site-years in 2012 changed the slope (i.e. deviated further from 1) between observed yield and predicted yield. This could be attributed to the extremely low rainfall during the growing season that may have altered the relationship of spatial yield response to GCVI, top 30 cm SOM, and field elevations. The influence of crop rotation could also affect the relationship of maize yield with weather, SOM and elevation during different growing conditions (Riedell et al., 2009). In our study, the effect of crop rotation on maize yield was not consistent. Yield advantages of maize-soybean rotation over multiple maize – soybean rotations were 7.3%, -20.2%, 2.1%, 6.9%, 6.8% and -6.4%, respectively, from 2011 to 2016. Moreover, fields in this study received N rates at planting (except for the sidedress N timing in 2013) that were higher than the state recommendation (Camberato and Nielsen, 2017). The results of the MSAR model might have been different for lower N rates and different N timing, and new parameter coefficients should be further considered.

In summary, the MSAR model also showed good prediction accuracy overall in new fields when extreme stress was not present at fine spatial scales. The strength of the MSAR method lies in its ability to use elevation and soil information to interpret satellite observations accurately in a fine spatial scale, and this method can be applied to new fields without the need for historical calibration.

3.5 Conclusions

Timely and accurate assessment of in-season maize grain yield at the field scale is needed for on-farm management decision support tools. Historical yield data analysis using spatial statistics and satellite imagery shows a great potential to forecast maize yield and identify major limitations in crop systems at the field scale. This study made use of historical grain yield data collected at farmers' fields and publicly available Landsat data to identify factors that affect maize yield, and further used them to predict maize yield at the field scale by combining in-season GCVI and soil and landscape properties using a newly developed MSAR model.

- PCA suggested that GCVI was the variable that most closely associated with grain yield ($r = 0.70$), which is attributable to its high correlation with plant LAI and the ability to capture differences in nutrient stress that was correlated with maize grain yield. GF KDD was also found to be highly correlated with grain yield ($r = -0.62$). We found that a 0.3% yield decrease was associated with each 1 KDD accumulation during the effective grain filling stage.
- The MSAR model performed reasonably well in 32 out of 35 site-years of the calibration dataset (absolute RME < 15%) with an average absolute RME of 6.6%. The average RMAE of the MSAR model was 13.1%. When the MSAR model was applied to the validation dataset containing 82 site-years, grain yield predictions were $\pm 15\%$ of observed yield in 56 site-years.
- The MSAR model resulted in large yield prediction errors under extreme stressed environmental conditions, such as 2012 growing season. To improve the model performance for a drought year, the inclusion of more years of similar stress into a separate calibration dataset is needed to better capture the yield responses under these conditions.

This novel approach can be further coupled with crop models and field sensors to monitor crop stress and yield in an even finer spatial context.

3.6 Acknowledgements

The authors would like to thank the Mylet Farms, IN for providing the historical on-farm yield monitor data for this research project.

3.7 References

- Auffhammer, M., and W. Schlenker. 2014. Empirical studies on agricultural impacts and adaptation. *Energy Economics*. 46: 555-561.
- Archontoulis, S.V., and F.E. Miguez. 2015. Nonlinear regression models and applications in agricultural research. *Agronomy Journal*. 107(2): 786-798.
- Baez-Gonzalez, A.D., P. Chen, M. Tiscareno-Lopez, and R. Srinivasan. 2002. Using satellite and field data with crop growth modeling to monitor and estimate corn yield in Mexico. *Crop Science*. 42: 1943-1949.
- Basso, B., D. Cammarano, and E. Carfagna. 2013. Review of crop yield forecasting methods and early warning systems. Report presented to first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy.
- Benedetti, R., and P. Rossini. 1993. On the use of NDVI profiles as a tool for agricultural statistics: the case study of wheat yield estimate and forecast in Emilia Romagna. *Remote Sensing of Environment*. 45: 311-326.
- Blackmer, T.M., J.S. Schepers, and G.E. Meyer. 1995. Remote sensing to detect nitrogen deficiency in corn. In: Robert, P.C., R.H. Rust, and W.E. Larson (Eds.) *Proceedings of site-specific management for agricultural systems: Second International Conference*. ASA-CSSA-SSSA. Minneapolis, MN, USA. 505-511.
- Bolton, D.K., and M.A. Friedl. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*. 173: 74-84.
- Burke, M., and D.B. Lobell. 2017. Satellite-based assessment of yield variation and its determinants in smallholder African systems. *PNAS*. 114: 2189-2194.

- Butler, E.E., and P. Huybers. 2013. Adaptation of US maize to temperature variations. *Nature Climate Change*. 3: 68-72.
- Butler, E.E., and P. Huybers. 2015. Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*. 10: 1-8.
- Camberato, J.J., and R.L. Nielsen. 2017. Nitrogen management guidelines for corn in Indiana. *Applied Crop Research Update*. 1-10.
- Cheng, Z., J. Meng, and Y. Wang. 2016. Improving spring maize yield estimation at field scale by assimilating time-series HJ-1 CCD data into the WOFOST model using a new method with fast algorithms. *Remote Sensing*. 8: 303.
- Colonna, I., M. Ruffo, G. Bollero, and D. Bullock. 2004. A comparison of geostatistical and spatial autoregressive approaches for dealing with spatially correlated residuals in regression analysis for precision agriculture applications. *Annual Conference on Applied Statistics in Agriculture*. Kansas State University. 310-326.
- Doraiswamy, P.C., S. Moulin, P.W. Cook, and A. Stern. 2003. Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote Sensing*. 69: 665-674.
- Ederira, J.I.R., L.I. Mayer, and M.E. Otegui. 2014. Heat stress in temperate and tropical maize hybrids: Kernel growth, water relations and assimilate availability for grain filling. *Field Crops Research*. 166: 162-172.
- Fu, W.J, P.K. Jiang, G.M. Zhou, and K.L. Zhao. 2014. Using Moran's I and GIS to study the spatial pattern of forest litter carbon density in a subtropical region of southeastern China. *Biogeoscience*. 11: 2401-2409.
- Gitelson, A.A., A. Vina, T.J. Arkebauer, D.C. Rundquist, G. Keydan, and B. Leavitt. 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical Research Letters*. 30: 1248.
- Jin, Z., G. Azzari, and D.B. Lobell. 2017. Improving the accuracy of satellite-based high-resolution yield estimation: A test of multiple scalable approaches. *Agricultural and Forest Meteorology*. 247: 202-220.

- Kogan, F., L. Salazar, and L. Roytman. 2012. Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. *International Journal of Remote Sensing*. 33: 2798-2814.
- Kroonenberg, P.M. 2008. *Applied multiway data analysis*. Wiley Series in Probability and Statistics.
- Lobell, D.B. 2013. The use of satellite data for crop yield gap analysis. *Field Crops Research*. 143: 56-64.
- Lobell, D.B., G.L. Hammer, G. McLean, C. Messina, M.J. Roberts, and W. Schlenker. 2013. The critical role of extreme heat for maize production in the United States. *Nature Climate Change*. 3(5): 497-501.
- Lobell, D.B., M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, R.M. Rejesus, and G.L. Hammer. 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science*. 344: 516-519.
- Lobell, D.B., D. Thau, C. Seifert, E. Engle, and B. Little. 2015. A scalable satellite-based crop yield mapper. *Remote Sensing of Environment*. 164: 324-333.
- Meng, Q., X. Chen, D.B. Lobel, Z. Cui, Y. Zhang, H. Yang, and F. Zhang. 2016. Growing sensitivity of maize to water scarcity under climate change. *Scientific Reports*, 6: 19605.
- Morell, F.J., H.S. Yang, K.G. Cassman, J. Van Wart, R.W. Elmore, M. Licht, J.A. Coulter, I.A. Ciampitti, C.M. Pittelkow, S.M. Brouder, P. Thomison, J. Lauer, C. Graham, R. Massey, and P. Grassini. 2016. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? *Field Crops Research*. 192: 1-12.
- Morris, T.F., T.S. Murrell, D.B. Beegle, J.J. Camberato, R.B. Ferguson, J. Grove, Q. Ketterings, P.M. Kyveryga, C.A.M. Laboski, J.M. McGrath, J.J. Meisinger, J. Melkonian, B.N. Moebius-Clune, E.D. Nafziger, D. Osmond, J.E. Sawyer, P.C. Scharf, W. Smith, J.T. Spargo, H.M. van Es, and H. Yang. 2018. Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agronomy Journal*. 110 (1): 1-37.
- Nguy-Robertson, A., A. Gitelson, Y. Peng, A. Vina, T. Arkebauer, and D. Rundquist. 2012. Green leaf area index estimation in maize and soybean: Combining vegetation indices to achieve maximal sensitivity. *Agronomy Journal*. 104: 1336-1347.

- Puntel, L.A., J.E. Sawyer, D.W. Barker, R. Dietzel, H. Poffenbarger, M.J. Castellano, K.J. Moore, P. Thorburn, and S.V. Archontoulis. 2016. Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*. 7: 1-18.
- R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Riedell, W.E., J.L. Pikul, A.A. Jaradat, T.E. Schumacher. 2009. Crop rotation and nitrogen input effects on soil fertility, maize mineral nutrition, yield, and seed composition. *Agronomy Journal*. 101:870-879.
- Schlemmer, M., A. Gitelson, J. Schepers, R. Ferguson, Y. Peng, J. Shanahan, and D. Rundquist. 2013. Remote estimation of nitrogen and chlorophyll contents in maize at leaf and canopy levels. *International Journal of Applied Earth Observation and Geoinformation*. 25: 47-54.
- Schlenker, W., and M.J. Roberts. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *PNAS*. 106: 15594-15598.
- Thompson, L.M. 1969. Weather and technology in the production of corn in the U.S. *Corn Belt Agronomy Journal*. 61: 453-456.
- Tu, J. and Xia, Z.G. 2008. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluations. *Science of the Total Environment*. 407(1): 358-378.
- Urban, D.W., M.J. Roberts, W. Schlenker, and D.B. Lobell. 2012. Projected temperature changes indicate significant increase in interannual variability of U.S. maize yields. *Climatic Change*. 112(2): 525-533.
- Urban, D.W., J. Sheffield, and D.B. Lobell. 2015. The impacts of future climate and carbon dioxide changes on the average and variability of US maize yields under two emission scenarios. *Environmental Research Letters*. 10: 1-9.
- USDA-National Agricultural Statistical Service (NASS). 2012. The yield forecasting and estimating program of NASS, by the Statistical Methods Branch, Statistics Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C. NASS Staff Report No. SMB 12-01.

- Wolfert, S., L. Ge, C. Verdouw, and M. Bogaardt. 2017. Big data in smart farming – a review. *Agricultural Systems*. 153: 69-80.
- Wong, M.T.F., and S. Asseng. 2006. Determining the causes of spatial and temporal variability of wheat yields at sub-field scale using a new method of upscaling a crop model. *Plant and Soil*. 283: 203-215.
- Zhang, T., and G. Lin. 2016. On Moran's I coefficient under heterogeneity. *Computational Statistics and Data Analysis*. 95: 83-94.
- Zhang, T., and B. Yang. 2018. Dimension reduction for big data. *Statistics and Its Interface*, 11: 295-306.

Table 3.1 Dominant soil types and soil properties of the study fields.

Soil Name	Soil mukey [†]	0-30 cm SOM (%)	Top soil texture	Classification
Cyclone, 0-2% slopes	165160	4.5	Silty Clay Loam	Mesic Typic Argiaquolls
Kendall-Fincastle, 0-1% slopes	165177	1.8	Silt Loam	Mesic Aeric Endoaqualfs
Patton, 0-2% slopes	162446	4.7	Silt Loam	Mesic Typic Endoaquolls
Starks, 0-1% slopes	165240	1.8	Silt Loam	Mesic Aeric Endoaqualfs
Rockfield-Williamstown, 1-6% slopes	165231	1.3	Silt Loam	Mesic Oxyaquic Hapludalfs
Starks, 0-3% slopes	162458	1.8	Silt Loam	Mesic Aeric Endoaqualfs
Patton, loamy substratum, 0-2% slopes	165221	5.0	Silty Clay Loam	Mesic Typic Endoaquolls
Camden, 2-6% slopes	165144	1.3	Silt Loam	Mesic Typic Hapludalfs
Fincastle, 0-2% slopes	165424	2.4	Silt Loam	Mesic Aeric Epiaqualfs
Camden, 0-1% slopes	165143	1.3	Silt Loam	Mesic Typic Hapludalfs

[†] SSURGO soil map key.

Table 3.2 Cropping history and soil types of the fields in this study.

	Area (ha)	2010	2011	2012	2013	2014	2015	2016	Soil mukey [†]
Field 1	27.84	Maize	Maize	Soybeans	Maize	Maize	Soybeans	Maize	165177 (51%), 165160 (39%),
Field 2	39.61	Maize	Soybeans	Maize	Maize	Maize	Soybeans	Maize	^{165221 (60%)} 162446 (89%), 162458 (7%),
Field 3	31.11	Maize	Soybeans	Maize	Maize	Maize	Soybeans	Maize	^{162452 (40%)} 165177 (51%), 165160 (49%)
Field 4	51.82	Maize	Maize	Soybeans	Maize	Maize	Soybeans	Maize	165177 (62%), 165160 (36%),
Field 5	36.03	Maize	Soybeans	Maize	Maize	Soybeans	Maize	Maize	^{1655221 (70%)} 165160 (59%), 165177 (37%),
Field 6	16.04	Soybeans	Maize	Maize	Maize	Soybeans	Maize	Maize	^{165220 (40%)} 165160 (77%), 165177 (23%)
Field 7	12.03	Soybeans	Maize	Maize	Maize	Soybeans	Maize	Maize	165160 (81%), 165177 (19%)

[†] SSURGO soil map key.

Table 3.3 Planting date of each field.

	Planting Date						
	2010	2011	2012	2013	2014	2015	2016
Field 1	4/12	5/14	-	5/13 (6/19) [†]	4/24	-	4/18
Field 2	4/14	-	4/5	5/16 (6/12)	5/3	-	4/27
Field 3	4/18	-	4/3	5/14 (6/8)	5/5	-	4/20
Field 4	4/17	5/19	-	5/14 (6/8)	5/5	-	4/20
Field 5	4/13	-	4/6	5/16 (6/11)	-	5/3	4/26
Field 6	-	5/10	4/13	5/15 (6/11)	-	5/1 (6/3)	4/26
Field 7	-	5/9	4/17	5/15 (6/12)	-	4/30	4/27

[†] Date in parentheses is N sidedress date.

Table 3.4 General weather conditions during maize growing season. Values are the means across fields for any particular year.

Period	Variable	Unit	2010	2011	2012	2013	2014	2015	2016	Long-term [†]
EV	MaxT	°C	23.1	26.1	23.4	26.4	24.9	24.6	23.8	24.1
	AveT	°C	17.1	20.5	16.2	20.5	18.7	19.0	17.9	18.4
	MinT	°C	11.0	14.9	9.0	14.6	12.5	13.3	11.9	12.6
	Precipitation	mm	363	230	81	171	180	306	181	215
LV	MaxT	°C	29.5	29.7	31.3	26.6	28.7	25.4	27.4	29.0
	AveT	°C	24.0	23.6	23.7	22.3	23.3	20.8	21.9	23.1
	MinT	°C	18.6	17.4	16.1	18.0	18.0	16.1	16.4	17.3
	Precipitation	mm	110	17	2	39	25	117	65	43
CP	MaxT	°C	30.1	33.0	33.7	28.5	26.8	27.3	28.4	28.6
	AveT	°C	23.7	26.9	26.3	22.4	20.6	22.1	23.5	22.9
	MinT	°C	17.3	20.8	19.0	16.2	14.4	17.0	18.7	17.1
	Precipitation	mm	57	24	17	102	50	124	97	86
GF	MaxT	°C	30.5	29.7	31.1	28.5	28.2	27.7	28.4	28.0
	AveT	°C	25.1	23.3	24.3	22.1	22.7	21.5	23.1	22.2
	MinT	°C	19.7	16.9	17.6	15.8	17.2	15.3	17.9	16.3
	Precipitation	mm	184	143	131	29	92	65	119	123
Total	MaxT	°C	26.7	28.8	27.6	27.5	26.6	26.1	26.0	26.5
	AveT	°C	20.8	22.8	20.5	21.6	20.6	20.5	20.6	20.7
	MinT	°C	15.0	16.8	13.4	15.6	14.7	14.9	15.1	14.9
	Precipitation	mm	715	415	231	342	348	613	462	467

[†] Assume average planting date of May 1st.

Table 3.5 Relationship between maize grain yield and seasonal rainfall and temperature.

Period	Variable	Relationship	Pearson Correlation (r)	Model R ²
EV	Precipitation	Quadratic		0.36
LV	Precipitation	Quadratic		0.24
CP	Precipitation	Linear	0.30	
GF	Precipitation	Linear	-0.23	
Total	Precipitation	Quadratic		0.40
GF	Temperature	Quadratic		0.31

Table 3.6 Maize grain yield, Moran's I, yield estimates from multivariate spatial autoregressive model, and model goodness of fit of the calibration dataset.

Field	Year	Mean Y_a	Moran's I of Y_a	ρ	Mean Y_{hat}	ME	MAE	RME [†]	RMAE	
unit		Mg ha ⁻¹			Mg ha ⁻¹	Mg ha ⁻¹	Mg ha ⁻¹	%	%	
1	2010	9.73	0.69	0.86	11.78	-2.05	2.42	-21.1	24.9	
	2011	12.77	0.41		11.42	1.35	2.14	10.6	*	16.8
	2013	11.63	0.65		12.57	-0.93	2.01	-8.0		17.3
	2014	14.26	0.37		13.98	0.28	0.94	2.0	*	6.6
	2016	14.79	0.62		13.87	0.91	1.65	6.2	*	11.2
2	2010	7.66	0.85	0.91	8.46	-0.80	3.16	-10.5	*	41.3
	2012	10.89	0.73		10.41	0.48	1.11	4.4		10.2
	2013	8.79	0.74		10.31	-1.51	2.09	-17.2	*	23.8
	2014	14.05	0.50		13.63	0.42	0.90	3.0	*	6.4
	2016	14.61	0.38		13.32	1.29	1.46	8.8	*	10.0
3	2010	12.75	0.40	0.77	13.25	-0.50	1.58	-4.0	*	12.4
	2012	11.68	0.64		12.57	-0.89	1.62	-7.6	*	13.8
	2013	14.25	0.43		13.62	0.64	1.30	4.5		9.1
	2014	13.71	0.35		13.84	-0.13	1.43	-1.0	*	10.4
	2016	15.00	0.24		14.00	1.00	1.43	6.7		9.5
4	2010	11.52	0.68	0.86	12.51	-0.99	2.38	-8.6	*	20.6
	2011	11.61	0.55		12.76	-0.24	1.84	-2.1	*	15.9
	2013	13.42	0.46		13.38	0.05	0.98	0.3	*	7.3
	2014	13.85	0.40		13.71	0.14	0.88	1.0	*	6.4
	2016	15.01	0.34		13.91	1.10	1.34	7.3	*	8.9
5	2010	10.75	0.76	0.87	10.80	-0.05	1.93	-0.5	*	18.0
	2012	10.64	0.69		9.92	0.72	1.37	6.8		12.9
	2013	12.99	0.59		13.08	-0.08	1.31	-0.6		10.1
	2015	10.41	0.81		11.38	-0.96	2.19	-9.3	*	21.0
	2016	14.48	0.38		14.30	0.18	0.63	1.2	*	4.3
6	2011	14.58	0.15	0.83	12.76	1.82	2.23	12.5		15.3
	2012	10.75	0.41		12.74	-2.00	2.01	-18.5	*	18.7
	2013	13.79	0.28		13.12	0.67	0.87	4.8		6.3
	2015	13.19	0.62		13.12	0.08	0.94	0.6		7.2
	2016	12.75	0.37		13.28	-0.53	0.66	-4.1	*	5.2
7	2011	14.32	0.43	0.85	12.58	1.74	2.20	12.2		15.4
	2012	11.90	0.50		12.71	-0.81	1.13	-6.8	*	9.5
	2013	13.47	0.31		12.86	0.61	1.16	4.5	*	8.6
	2015	11.33	0.59		12.76	-1.44	2.01	-12.7	*	17.8
	2016	13.04	0.26		13.00	0.04	0.58	0.3	*	4.5

† “*” indicates the relative error is smaller compared to the linear model without spatial autocorrelation structure.

Table 3.7 Multivariate spatial autoregressive model prediction.

Year	Field	Mean Y_a	Mean $Y_{\hat{a}}$	ME	MAE	RME	RMAE
		----- Mg ha ⁻¹ -----				----- % -----	
2012	2	10.89	10.27	0.62	1.22	5.71	11.2
2012	3	11.68	13.29	-1.61	1.89	-13.8	16.2
2012	5	10.64	9.59	1.04	1.61	9.8	15.1
2012	6	10.75	14.01	-3.27	3.27	-30.4	30.4
2012	7	11.90	13.00	-1.10	1.29	-9.3	10.8
2016	1	14.79	13.42	1.37	1.77	9.3	12.0
2016	2	14.61	12.75	1.86	1.97	12.7	13.5
2016	3	15.00	13.62	1.38	1.71	9.2	11.4
2016	4	15.01	13.40	1.61	1.78	10.7	11.8
2016	5	14.48	14.14	0.33	0.68	2.3	4.7
2016	6	12.75	13.64	-0.89	0.97	-7.0	7.6
2016	7	13.04	12.92	0.12	0.64	0.9	4.9

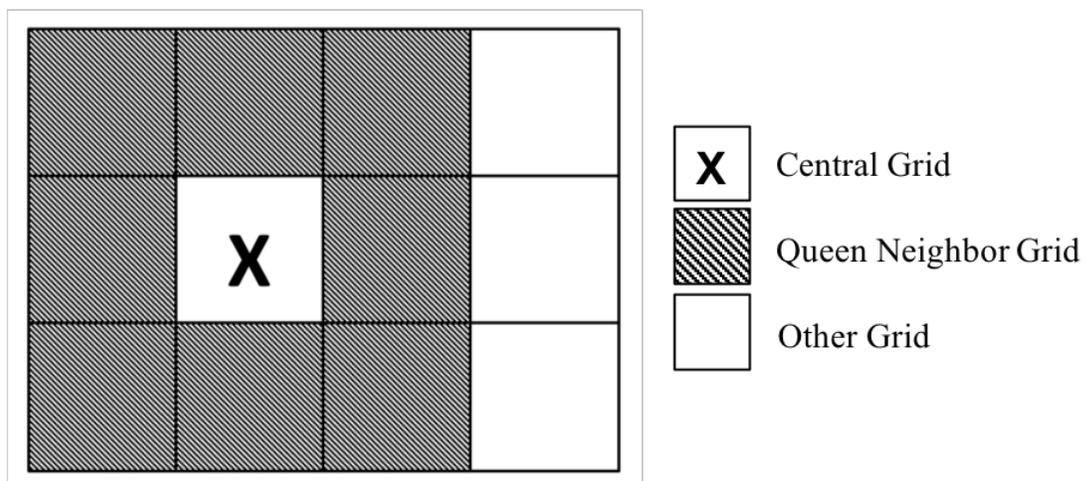


Figure 3.1 Spatial neighbors based on queen's contiguity.

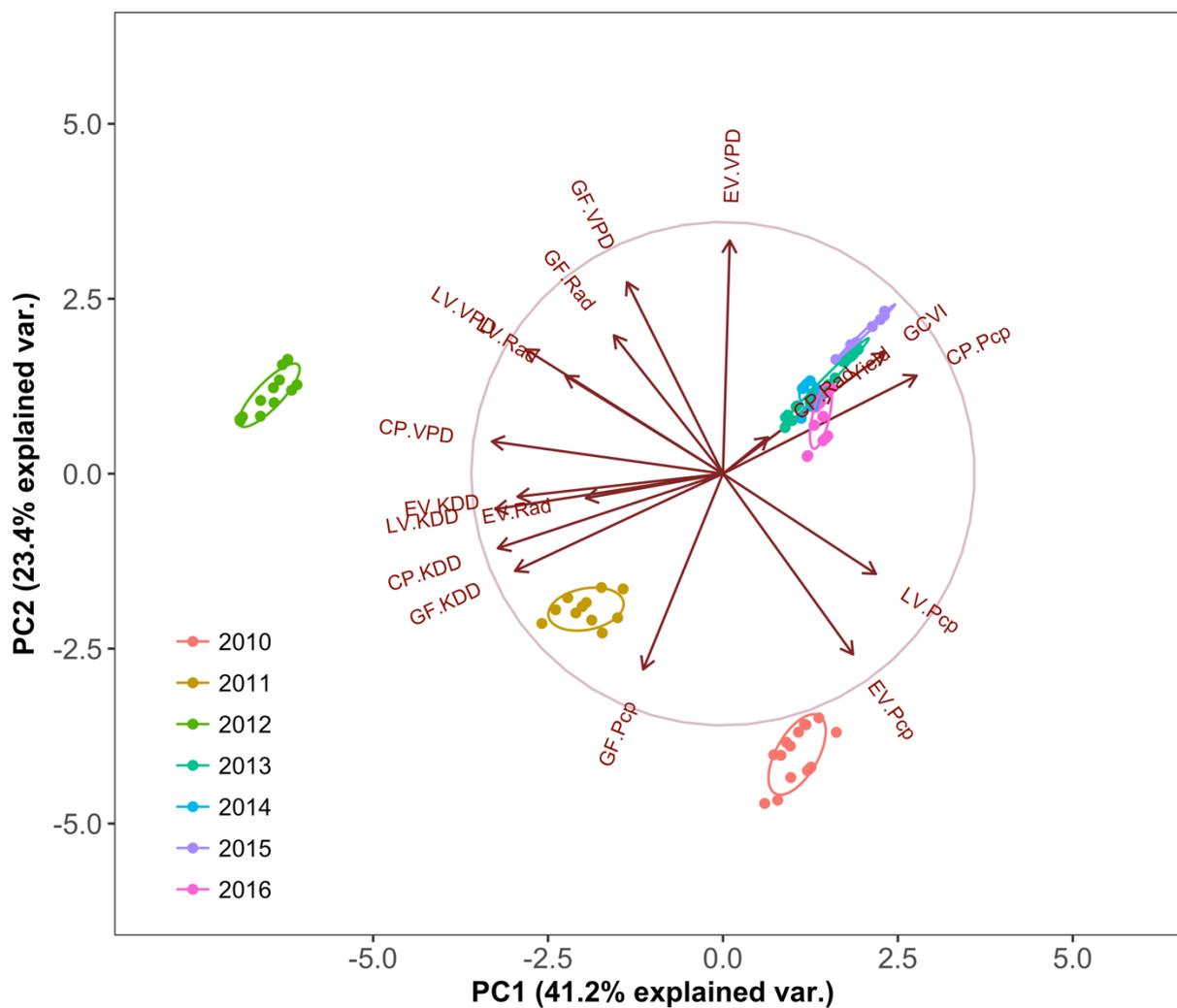


Figure 3.2 Biplot derived from the first two principal components. EV, early vegetative stage; LV, late vegetative stage; CP, critical period; GF, grain filling period; Pcp, precipitation, mm; Rad, solar radiation, MJ m^{-2} ; VPD, vapor pressure deficit, kPa; KDD, killing degree days, d°C ; GCVI, green chlorophyll vegetation index. Each point represents the mean within a soil type and site-year.

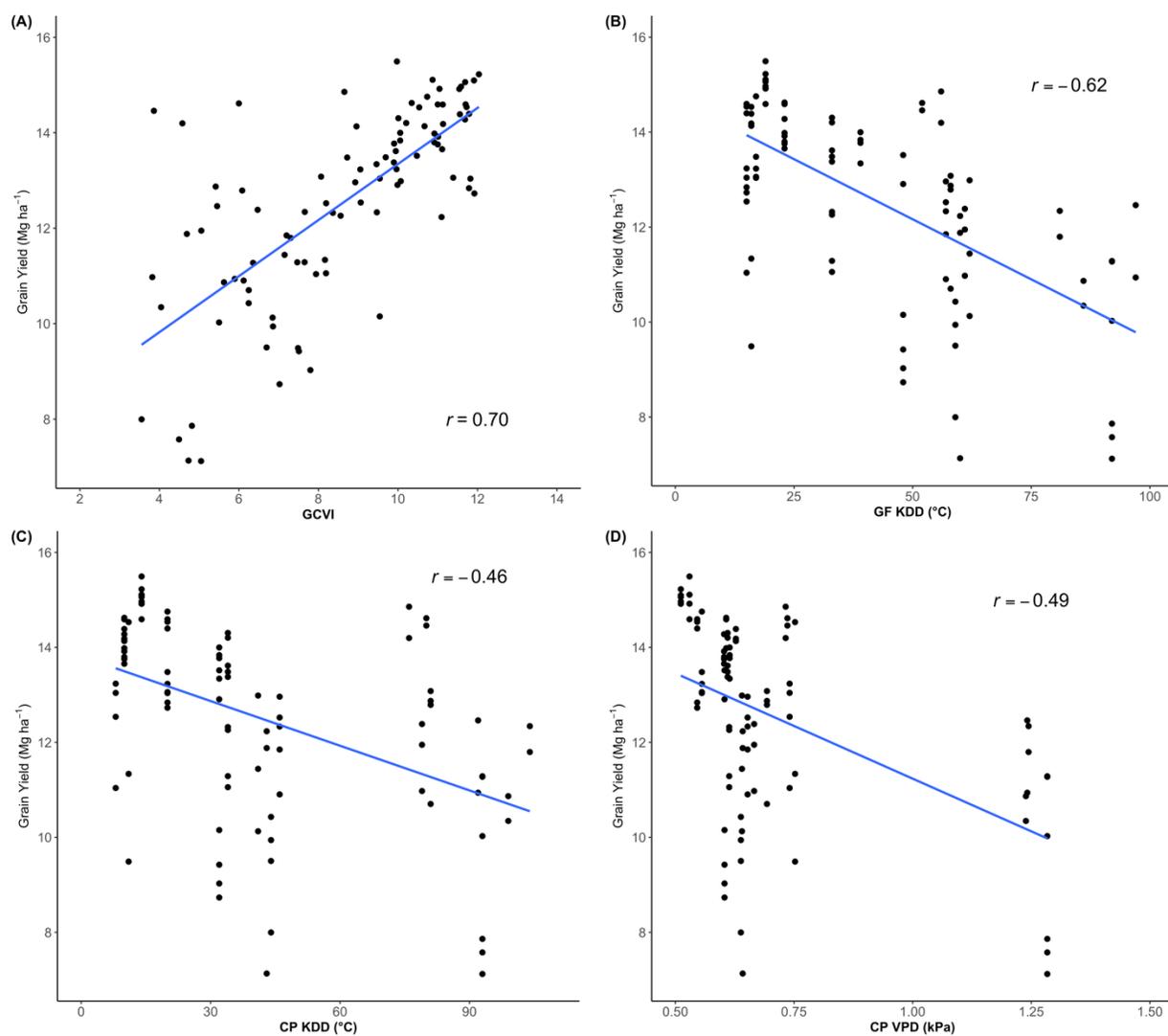


Figure 3.3 The relationship between maize grain yield and (A) green chlorophyll vegetation index (GCVI); (B) grain filling period cumulative killing degree days (GF KDD) (d.°C); (C) critical period cumulative killing degree days (CP KDD) (d.°C); and (D) critical period mean vapor pressure deficit (CP VPD) (kPa); using calibration dataset from 2010 to 2016, and each point represents the mean within a soil type and site-year.

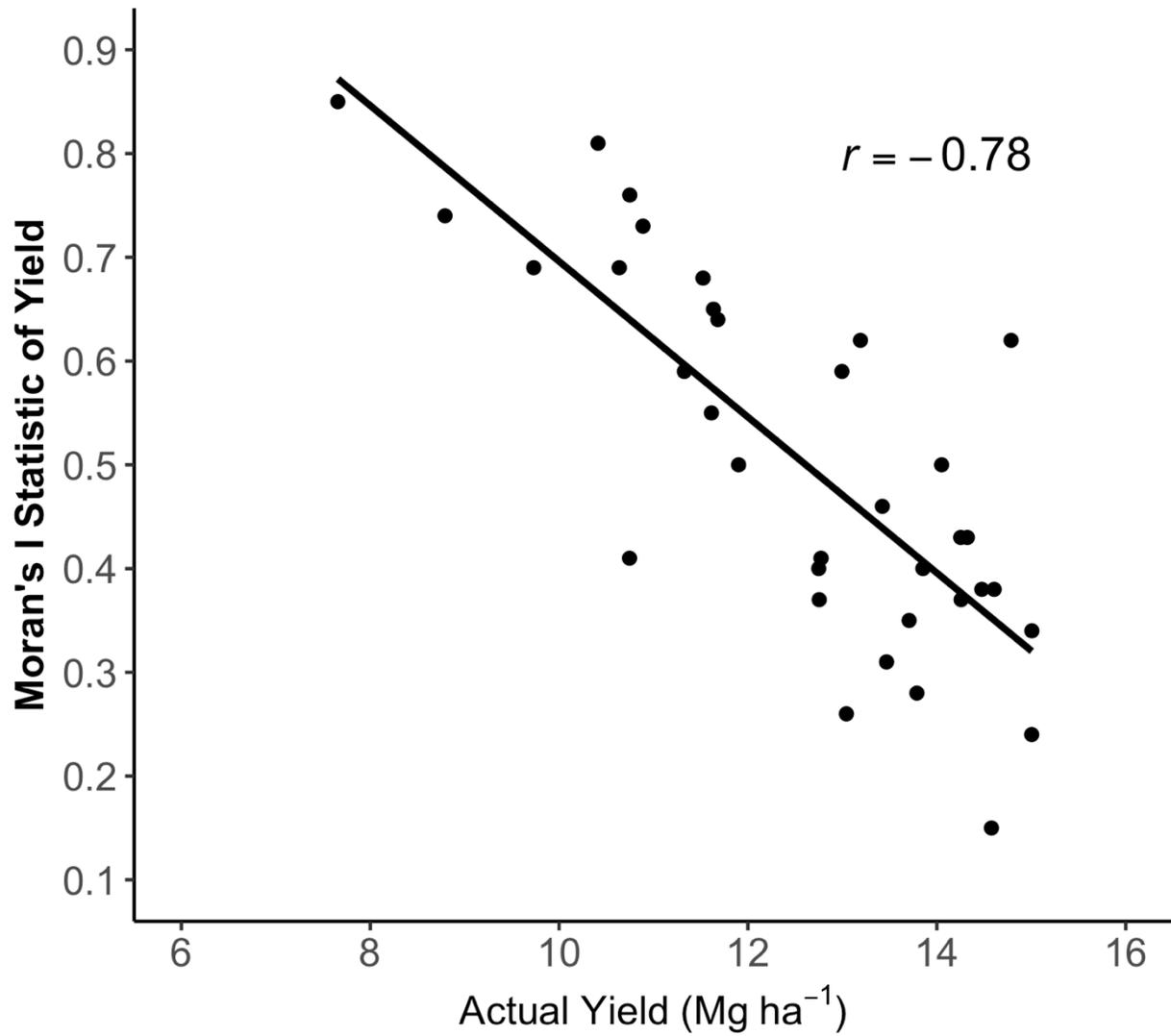


Figure 3.4 The negative linear relationship between Moran's I statistic of grain yield and grain yield. Each point represents a site-year in the calibration dataset ($n=35$).

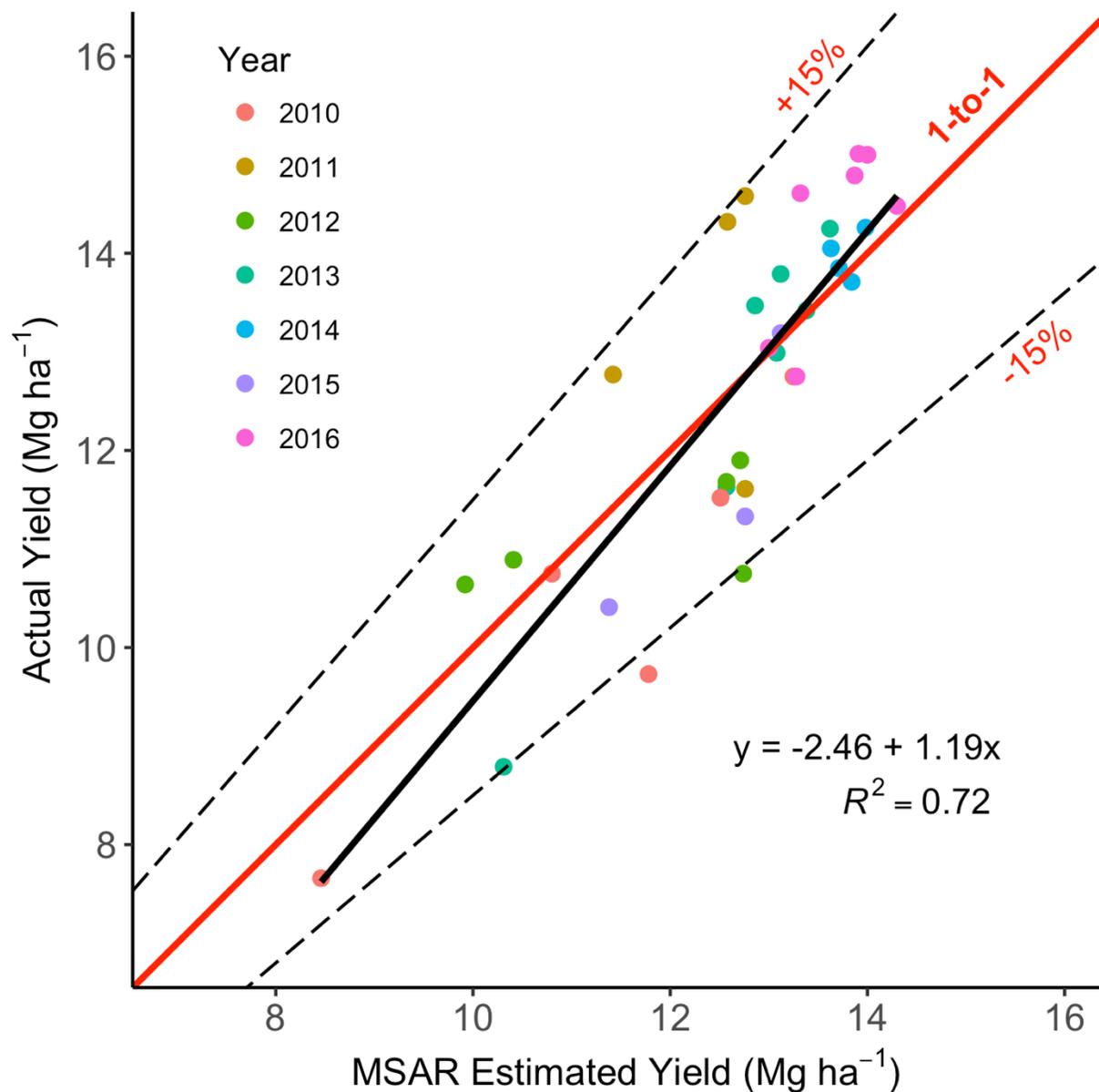


Figure 3.5 Comparison of maize grain yield predicted using MSAR model and actual grain yield of the calibration dataset ($n=35$). The 1:1 line (solid red line) and $\pm 15\%$ deviation (dashed black line) are shown. Fitted linear regression model is shown (solid black line).

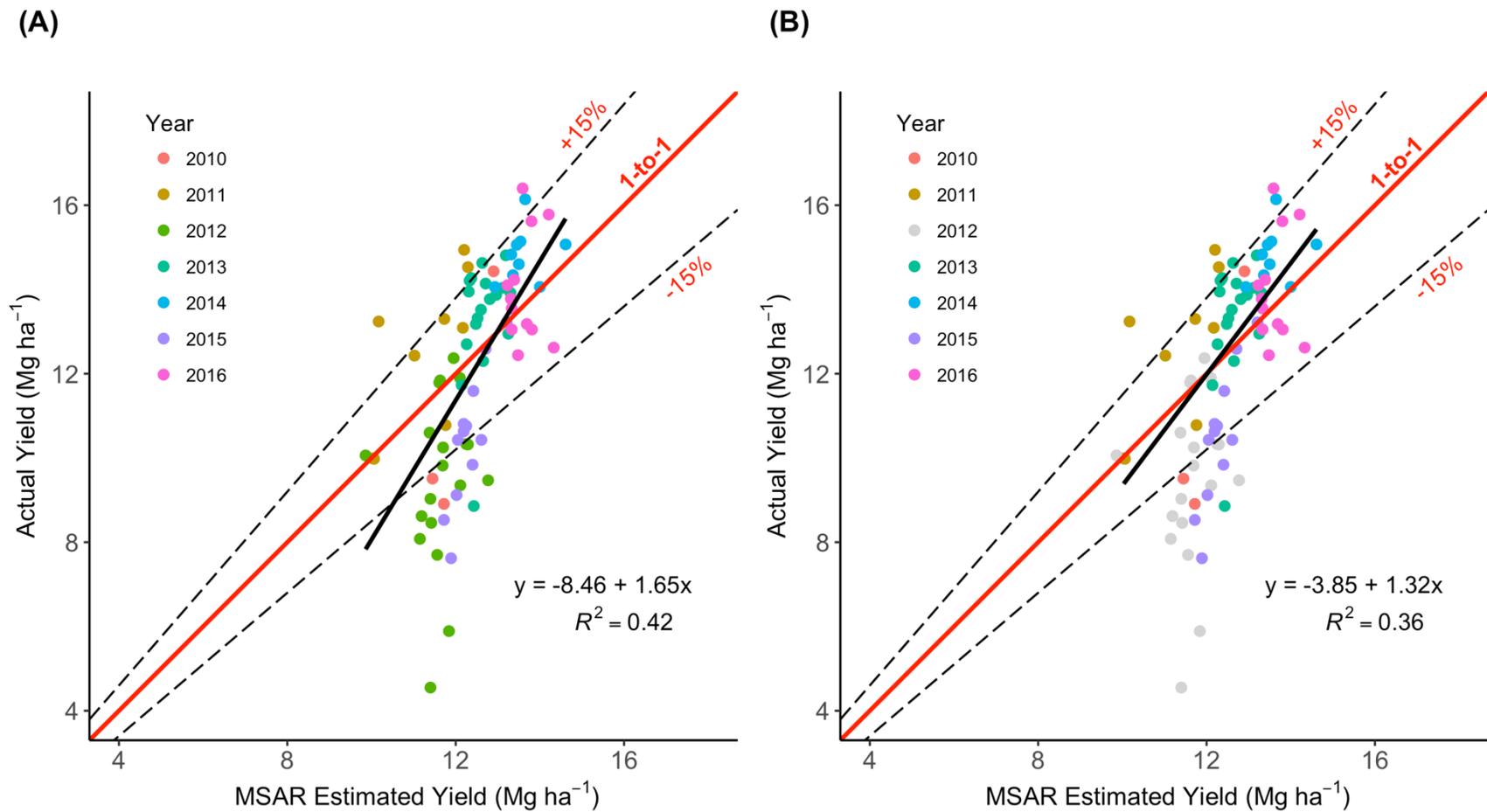


Figure 3.6 Comparison of maize grain yield predicted using MSAR model and actual grain yield of the validation dataset ($n=82$). The 1:1 line (solid red line) and $\pm 15\%$ deviation (dashed black line) are shown. (A) Fitted linear regression model using all site-years is shown (solid black line). (B) Fitted linear regression model excluding 2012 site-years is shown (solid black line).

CHAPTER 4. INCORPORATE MULTIVARIATE SPATIAL STATISTICS TO A PROCESS-BASED NITROGEN TRANSFORMATION MODEL FOR FIELD-SCALE MAIZE GRAIN YIELD PREDICTION

4.1 Abstract

Agricultural system models are being widely used for crop yield forecasts and site-specific management decisions. Many efforts have been devoted to improving field-scale maize yield prediction for on-farm decision tools. In this study, we evaluated the performance of the N Model package in the Mapwindow GIS + MMP Tools (referred to as the “N Model”) in field-scale maize grain yield prediction, and further identified factors that affected yield prediction accuracy. A multivariate spatial autoregressive (MSAR) model was then incorporated to the N Model to adjust maize yield prediction based on historical maize yield collected from farmers’ fields from 2010 to 2016 in central Indiana, USA. The N Model predicted higher grain yield than actual yield in most cases, and the linear agreement of predicted and actual yield improved as the spatial aggregation scale became broader using a dataset containing 35 site-years on 7 fields. However, at the soil type level within a field, only 56% of the yield predictions by the N Model fell within 15% of the actual yield. Our analysis revealed that the difference between actual and N-Model-predicted yield was linearly associated with early vegetative stage (EV) precipitation ($r = 0.33$). Moreover, the residual of N Model predicted yield showed significant spatial heterogeneity in all site-years. The proposed MSAR model used EV precipitation, top 30 cm soil organic matter and field elevation while accounting for spatial autocorrelation in 10-m grids to adjust yield predictions. The MSAR adjusted yield predictions resulted in more cases (77%) that fell within 15% of actual yield compared to the N Model alone when using the whole dataset. However, if the 2012 data was not included in the training process, the MSAR adjusted yield predictions did not improve in the drought year of 2012 (average RME of 24.1%). When extrapolating the MSAR parameters developed from the 7 fields above to a dataset containing 82 site-years on 30 different fields in the same region, the improvement from the MSAR adjustment was not significant. The lack of improvement from the MSAR adjustment could be that the relationship used in the MSAR model was location specific. Additionally, the uncertainty of precipitation data could also affect the

relationship. The MSAR approach proposed here could be used to fine tune the N Model for in-season grain yield prediction and N management practices at the field scale.

4.2 Introduction

Agricultural system models have become increasingly common for crop yield forecasting in recent years with the growing availability of data from the farm. These models have been used by producers to make site-specific management decisions on nutrient management practices (Fountas et al., 2006; Basso et al., 2016). Researchers have been using models to understand nitrogen (N) dynamics (Cannavo et al., 2008) and to answer questions that cannot be addressed with field research such as climate change impacts (Lobell and Asseng, 2017). The results can be further used by policymakers to inform development assessments. Two main model development approaches have been used, statistical models and process-based simulation models. Both modeling approaches have advantages and disadvantages (See Table 4.1 for details), and both are dependent on the accuracy of input data. Errors in weather measurement, precipitation in particular (Lobell, 2013), could introduce bias into predictions made with both types of models. The selection of which approach to use also depends on the modeling purpose and scale. However, it is worth noting that all statistical models have predictors and form of functions informed by process understanding, and all process-based models have empirical testing and calibration (Lobell and Asseng, 2017).

Statistical models use historical weather observations, satellite imagery, and crop growth and yield data to develop relatively simple regression equations that can be used to forecast crop production and economy. Data has been taken from field measurements, farmer surveys, government statistics, or a combination of the above (Lobell and Asseng, 2017). The performance of statistical models differed by climate variables and spatial scales. Statistical models were found to be able to reproduce many of the key features of process-based model response to changing temperature and precipitation, and the performance of statistical models improved as the spatial scale of analysis became broader (Lobell and Burke, 2010). In the literature, statistical models have been applied at relatively low spatial resolution, often at the scale of counties, regions or countries rather than field scale (Lobell and Burke, 2010). Additionally, Auffhammer and Schlenker (2014) reported that spatial averaging farm-level data over a county or temporal

averaging over the growing season can hide important interactions and nonlinearities. Moreover, averages derived from large regions will show less variation from year to year relative to averages over smaller areas (Lobell and Burke, 2010).

Process-based crop models, on the other hand, have been widely used at various scales, from individual fields to large regions. Process-based crop models attempt to simulate plant physiological response or soil nutrient availability to weather and management factors over time using mathematical descriptions of physiological, chemical, and physical processes on homogeneous land in order to predict how a plant will grow under specific environmental conditions (Batchelor et al., 2002). Processes often include photosynthesis, respiration, growth and partitioning, development of reproductive structures, transpiration, soil chemical transformations, water and nutrient uptake (White and Hoogenboom, 2009). Most process-based models calculate at daily time steps during crop growing season and require initial conditions such as soil nutrients and water status, planting date and density. Irrigations, fertilizer applications, crop rotation, and other factors may also be considered. Following the first crop model by de Wit (1965), various process-based models, such as CERES-Maize (Jones and Kiniry, 1986), WOFOST (van Diepen et al., 1989), DSSAT (Jones et al., 2003), APSIM (Keating et al., 2003), Hybrid-Maize (Yang et al., 2004), Maize-N (Setiyono et al., 2011) and AgMaize (Tollenaar et al., 2018), have been developed to understand crop growth, development and production under different environment conditions at various scales. Moreover, the predictability of process-based models on maize yield and N management has been assessed at multistate levels (Archontoulis et al., 2014; Thompson et al., 2015; Morell et al., 2016; Puntel et al., 2018), and these models have been further used to study the impact of climate change (Lobell and Asseng, 2017).

One problem when using process-based models at the field scale is that these models usually assume homogeneity of environmental conditions within individual fields. It is commonly observed that maize yield response to applied N varies considerably across and within fields, and from year to year due to soil, landscape, weather and crop management factors that affect N supply - demand relationships (Morris et al., 2018). Even within the same region and year, field-to-field variation in yield and yield gap can be remarkably high (Farmaha et al., 2016). As anticipated, maize yields were found to be significantly correlated across space due to the systemic nature of weather, nutrient stress, disease, and pest damage (Ozaki et al., 2008). Therefore, there is considerable uncertainty in results produced by the use of crop models.

Spatial statistics is helpful in analyzing the yield spatial and temporal variability at the field scale in the agricultural system. Historical yield data provides us with prior knowledge on the variation that exists within the field over time, which can be used to study the spatial and temporal variability for improved model accuracy. Spatial autoregressive model was found to be able to adjust for spatial autocorrelation inherent in maize yield (Zhang et al., 2010). In Chapter 3, a multivariate spatial autoregressive (MSAR) model that adjusted for spatial autocorrelation and used the season-specific remote sensing vegetation index predicted grain yield accurately when extreme stress was not present at the field scale.

Reliable and comprehensive datasets are needed for an efficient model evaluation. Ideally the data have to cover several aspects of the soil-plant-atmosphere continuum, but in reality, such data are rare (Archontoulis et al., 2014). The yield data that farmers routinely collected provides an opportunity to evaluate model performance at the commercial field level across a range of growing season conditions.

The objectives of this study were to: (1) evaluate the maize grain yield prediction accuracy of the N Model in the Mapwindow GIS + MMP Tools in a spatial context at the field scale; (2) identify factors affecting maize yield prediction at the field scale; and (3) incorporate a multivariate spatial autoregressive (MSAR) model into the N Model to adjust maize yield prediction based on historical maize yield collected from farmers' fields from 2010 to 2016 in central Indiana, USA.

4.3 Materials and Methods

4.3.1 Historical maize yield data

For this study, estimates of the MSAR model parameters were derived using a calibration dataset and the performance of the MSAR model adjustment was tested on a validation dataset. The calibration data included in this study are from 2010 to 2016 on 7 fields with a total of 35 maize site-years in south Cass and north Carrol county, Indiana (N 40°37'26", W 86°25'20"), USA. The validation data included 30 fields in the same region with a total of 82 maize site-years (Table B.1). Details on agronomic practices were described in Chapter 3. Total N application was uniform in each of the site-years; 268 kg N ha⁻¹ and 302 kg N ha⁻¹ was applied as urea ammonium nitrate (UAN, 28-0-0) for maize following soybean and maize following maize, respectively. Maize yield monitor data were gridded to 10m x 10m by averaging harvest points within each grid and were

reported at 15.5% moisture. Yield values more than 3 standard deviations away from the field mean yield were removed and then replaced by the mean of its nearest queen neighbors.

4.3.2 The process-based N Model in MapWindows GIS + MMP Tools

The N Model in the MapWindows GIS + MMP Tools by Joern and Hess at Purdue University was used to derive maize yield prediction at a 10m x 10m grid. From this point on, this N Model will be called the “N Model” for simplicity. The N Model was selected due to its ability to predict maize N uptake in a fine spatial context. Furthermore, model calibration is not required for the N Model. The N Model is based on the daily dynamic balance of plant N uptake and soil N transformation processes. Processes considered in the N Model include soil and manure N mineralization, nitrification, denitrification, ammonia volatilization, nitrate leaching and crop N uptake, driven by temperature, soil moisture, soil texture, soil organic matter (Zhao, 2013). The impact of crop rotation was considered in the soil N mineralization process. In the N Model, N mineralization potential for the top 30 cm soils under maize after soybean was estimated to supply 22.4 kg N ha⁻¹ per percent of organic matter (OM), and soils under maize after maize supplies 11.2 kg N ha⁻¹ per percent of OM. Soil physical and chemical properties were obtained from the USDA-NRCS web soil survey. Daily precipitation, maximum and minimum air temperature were obtained from the U.S. National Weather Service Cooperative Observer Program (NWS COOP).

The N Model inputs include emergence date, maximum rooting depth, previous crop, field boundary, fertilizer and manure N source, rate, placement and timing. Maize N uptake estimation is based on a set of N uptake rates segmented by maize thermal development stages (Table 4.2), and maize N uptake is then translated to maize yield by assuming one kilogram of plant N uptake would yield 56 kilograms of maize grain. The comprehensive review by Ciampitti and Vyn (2012) reported that the average N internal efficiency (NIE = grain yield/plant N uptake) for modern maize hybrids was 56 kg kg⁻¹. However, a wide range of NIE (0.1 – 123.4 kg kg⁻¹) was found under different growth environments. At a given maize yield, 12 Mg ha⁻¹ for example, plant total N uptake was found to vary between 150 kg ha⁻¹ to 350 kg ha⁻¹ in modern maize hybrids. In fact, the variability in NIE depends on many factors (hybrid, N management, plant density, other crop management, weather and soil properties) and the interactions between these factors, which could affect the predictability of the N Model.

A default yield potential parameter of 12.6 Mg ha⁻¹ (200 bu/A) is used to refrain extreme N uptake values. In this study, yield potentials for each field were set as 110% of the highest historical yield on each field, and maximum rooting depth was set to be 122 cm. Detailed input parameters used for the N Model were presented in Table 4.3.

4.3.3 Weather and crop data

Daily weather variables (max/min air temperature, precipitation, solar radiation, and relative humidity) were obtained from the National Weather Station-Cooperative Observer Network (NWS-COOP) and all fields in this study shared the common nearest weather station (Logansport, IN). The same weather source was used by the N Model. Soil moisture was estimated using Irris Scheduler software (Purdue University). Maize maturity was assumed to be 1500 growing degree days (GDD, base 10 °C) as Relative Maturity records of planted maize hybrids were not available. Four growth stages were distinguished: early vegetative (EV, 0-483 GDD), late vegetative (LV, 484-631 GDD), critical period (CP, 632-922 GDD), and effective grain filling (GF, 923-1361 GDD). Daily nighttime temperature was calculated as the mean of daily average temperature and daily minimum temperature. Mean values of daily average temperature, nighttime temperature, vapor pressure deficit (VPD), and soil moisture of each of the four growth stages were computed separately. VPD has been widely used as a drought measure of the atmosphere's drying force on plants.

Cumulative sum of precipitation, solar radiation, and killing degree days (KDD) were calculated for the four growth stages. In this study, 29 °C based KDD was used to quantify temperatures that may reduce maize yields since studies have found that maize yields started to drop when maximum temperatures exceeded 29 °C, especially during silking and early grain filling stages (Schlenker and Roberts, 2009; Butler and Huybers, 2015). Daily KDD is computed by,

$$KDD_d = \begin{cases} T_{max} - 29, & \text{if } T_{max} > 29, \\ 0, & \text{if } T_{max} \leq 29. \end{cases} \quad (1)$$

Weather conditions during the growing seasons compared to long-term average conditions are presented in Figure 4.1.

4.3.4 Field and soil attributes

Digital elevation model (DEM) and soil properties in 10m x 10m grids were obtained for the fields from the University of Missouri Web-based clipper application (<http://clipper.missouri.edu>). Soil organic matter data was extracted from the USDA-NRCS SSURGO database and gridded to 10-m spatial resolution. Top 30 cm soil organic matter (SOM) amount was calculated using soil bulk density and organic matter content of each soil horizon (Table B.2). Top 30 cm was selected because it is the common sampling depth for soil N determination.

4.3.5 Maize yield residual analysis

Maize yield residual, defined as actual yield minus the N Model predicted yield ($y = y_a - y_N$), was used for this analysis. Maize yield residual was averaged by soil types for each site year. For each of the four development stages identified above, scatterplots were used to explore the relationship between season-specific weather variables and maize yield residuals (Figure A.1). Season-specific variables included average temperature, cumulative precipitation, average soil moisture, cumulative excess water, average VPD, cumulative solar radiation, cumulative KDD, and average nighttime temperature.

Overall spatial autocorrelation of yield residual was tested using the global Moran's I statistic with queen neighbor weight matrix. Moran's I value is calculated as

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (i \neq j), \quad (2)$$

where n is the number of observations in the field of interest, x_i and x_j are the observations at locations of i and j , \bar{x} is the mean of observations, and w_{ij} is the spatial weight between locations of i and j . The theoretical range of Moran's I is between -1 and 1 (Zhang and Lin, 2016). Large positive Moran's I value indicates positive spatial autocorrelation, where the target value is similar to its neighbors. Large negative Moran's I value means negative spatial autocorrelation, where the target value is different with its neighbors (e.g., a high value in a low value neighborhood). A Moran's I value close to 0 implies spatial randomness (Tu and Xia, 2008).

Maize yield residuals, similar to actual maize yields, showed high spatial autocorrelation, where the yield residual of one site is related to the values of its neighboring sites. Multivariate spatial autoregressive model (MSAR) could account for spatial autocorrelation of maize yield

residual and this model was used to estimate maize yield residual using site-specific and year-specific variables. The MSAR is defined as

$$y_{i,j} = \rho \mathbf{W} y_{ij} + \mathbf{X}_i \boldsymbol{\beta}_i + \mathbf{X}_j \boldsymbol{\beta}_j + \epsilon, \quad (3)$$

where $y_{i,j}$ is maize yield residual at site i in year j , ρ is the autocorrelation parameter ($-1 < \rho < 1$), \mathbf{W} is the queen neighbor weight matrix, \mathbf{X}_i is site-specific terrain values (elevation and 0-30 cm SOM), \mathbf{X}_j is early vegetative stage precipitation (year-specific), and ϵ is the error term with $N(0, \sigma^2)$ distribution. The parameters of this model include ρ , $\boldsymbol{\beta}_i$, $\boldsymbol{\beta}_j$ and σ^2 . Queen neighbor weight matrix was used as it is the commonly used spatial weight for a regular grid of points in agricultural studies (Colonna et al., 2004). The maximum likelihood estimator of ρ was derived using the Newton-Raphson algorithm. Parameters of the MSAR model were estimated by the maximum likelihood method conditioning on ρ . The grain yield prediction at a given site i in year j adjusted by the MSAR model was calculated by:

$$y_{hat(i,j)} = \widehat{y}_{i,j} + y_{N(i,j)} \quad (4)$$

where $y_{hat(i,j)}$ is the MSAR adjusted grain yield prediction, $\widehat{y}_{i,j}$ is MSAR estimated yield residual, and $y_{N(i,j)}$ is grain yield prediction by the N Model at site i in year j .

4.3.6 Statistical measures for model performance

Graphical comparison was used to visually assess the goodness of fit. Mean error (ME) and mean absolute error (MAE) were used to provide absolute error for estimating maize yield residual. Relative error terms for maize yield, such as RME and RMAE, were used to compare model goodness of fit among different fields or years using the N Model and the MSAR model.

In this study, absolute RME less than 15% was considered as good prediction overall, and RMAE less than 15% was considered as good prediction at the 10-m pixel level. The distance between RMAE and the absolute value of RME indicates the direction of model yield predictions. For instance, a closer distance suggested the model yield prediction was mainly an underestimation (positive RME) or overestimation (negative RME), while a greater distance suggested the error originated from both underestimated and overestimated sites in each field.

$$ME = \bar{y} - \frac{1}{n} \sum_{i=1}^n \widehat{y}_i \quad (5)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (6)$$

$$RME = \frac{ME}{\bar{y}_a} \times 100\% \quad (7)$$

$$RMAE = \frac{MAE}{\bar{y}_a} \times 100\% \quad (8)$$

where n is the number of data points, y_i and \hat{y}_i are the actual and predicted maize yield residual, respectively, p is the number of model parameters, and \bar{y}_a is the mean of actual yield.

4.4 Results and Discussion

4.4.1 Performance of the N Model

A spatial evaluation of the performance of the N Model was conducted to test whether the N Model can be used as a decision tool for various Midwestern production systems. Predicted maize grain yield using the N Model and the actual grain yield were compared at three spatial aggregation levels and the linear agreement of predicted and actual yield improved as the aggregation scale became broader (Figure 4.2). At the soil type level within a field, there was no obvious trend in the relationship between actual and predicted yield. Around 56% of the yield predictions by the N Model fell within 15% of the actual yield. The predicted yield was greater than actual yield for 78% of all cases. Individual cases of overestimations exceeding 15% of actual yield included those occurring on low OM soils during a dry growing season (2012), high OM soils under maize after maize during seasons with normal (2016) or higher than normal precipitation (2010), and in a late planting season (2013). Sources of yield overestimation could be from unaccounted biotic and abiotic plant stresses, model omission of N immobilization processes for maize after maize, and uneven emergence from late planting. Greater model predicted grain yield than actual yield was common in farmers' fields where it is not feasible to ensure timely nutrient supply and perfect control of biotic stresses (Morell et al., 2016; Cassman et al., 2003). The N Model predicted grain yield was below actual yield in 22% of the soil type-year cases. Model underestimation more than 15% of the actual yield corresponded to soils with low OM during seasons with higher than normal precipitation such as in 2010 and 2015. This is likely due to an overestimation of N leaching loss under high precipitation levels.

At the individual field scale, field-average grain yields predicted by the N Model were greater than actual yields for 94% of all the site-year cases. The N Model yield prediction was considered reasonable for 74% of the site-year cases. Large yield overestimations corresponded to fields 1 and 2 during 2010 and 2013 growing season. This could be related to rainfall spatial variation in specific site-years, especially for summer thunderstorms, when weather data collected at a station located 10 km away could not depict what happened on a farm.

When site-year cases were aggregated by growing season, the N Model predicted yield and actual yield showed a distinct linear pattern ($r = 0.91$), with predicted yield slightly greater than actual yield in all years. The average ratio between predicted and actual yield was 1.1. The N Model predicted yield was within 15% of actual yield in all years except for 2013. Additionally, the range of predicted and actual grain yield became narrower with increasing spatial aggregation. This trend was also found by Morell et al. (2016), suggesting that both low and high portions of predicted yields were averaged out when moving from soil type within a field to across fields within a year. In summary, even though the N Model was built at the spatial level of soil type within a single field, its prediction accuracy increased with spatial aggregation, especially when several farms were considered for maize production prediction. Further improvement of field and sub-field scale predictions is needed.

4.4.2 Analysis of maize yield residual of the N Model

At the soil type level within a site-year, maize yield residuals ranged widely from -7.2 Mg ha^{-1} to 6.8 Mg ha^{-1} , with a mean of -1.2 Mg ha^{-1} in the calibration dataset. The yield residual did not show any significant relationship with any of the weather and stress variables considered in this study with the exception of EV precipitation (Figure A.1). Maize yield residual displayed a trend that moved toward zero with increasing EV precipitation ($r = 0.33$, Figure 4.3). This overestimation of maize yield could be due to the high plant N uptake rate during V6 to V12 stage assumed by the N Model ($0.42 \text{ kg N ha}^{-1}$ per GDD), together with more readily available N in the soil from applied N fertilizers when excess spring precipitation was not present (Table 4.2). In comparison, physiological studies of modern maize hybrids in response to various plant density and N stress factors found that high levels of N uptake rate didn't start until around V10 stage (Ciampitti et al., 2013; Bender et al., 2013). According to Bender et al. (2013), maize N uptake followed a sigmoidal (S-shaped) pattern, with N uptake rate of modern maize hybrids peaking

during V12 to R1 growth stage (around 556-833 GDD) at 0.46 kg N ha⁻¹ per GDD. They reported a lower N uptake rate during V6 to V12 stage at 0.24 kg N ha⁻¹ per GDD. Additionally, soil N loss was found to have an exponential increase with spring precipitation, leading to a higher economic optimum N rate (EONR) with above normal spring precipitation for both maize after maize and maize after soybean crop sequences using a 16-year field experiment dataset in central Iowa (Puntel et al., 2016). Therefore, further adjustment of the module controlling plant N uptake rate in the N Model might be helpful for improved maize yield prediction accuracy at the soil type spatial level.

At the individual field scale, maize yield residuals showed significant spatial autocorrelation as suggested by Moran's I coefficient. Moran's I of the yield residual ranged from 0.28 to 0.88 using the calibration dataset, and varied among site-years. Moreover, a significant negative linear relationship was found between Moran's I of yield residual and actual grain yield ($r = -0.70$, Figure 4.4 A). The Moran's I increased as actual grain yield decreased. This result indicated high spatial heterogeneity of yield residual during unfavorable growth conditions for maize. Therefore, the N Model showed higher spatial uncertainty during harsh environments. Moreover, maize yield residual Moran's I showed a quadratic relationship with EV precipitation (Figure 4.4 B); Moran's I values increased when EV precipitation was greater than 166 mm, compared to the long-term average EV precipitation of 215 mm. The EV stage of the growing season was found to be most susceptible to excess moisture conditions. In Indiana, the consequences of excess precipitation can be worse if adequate or systematic subsurface drainage is lacking. This could be attributed to less carbohydrate storage, poorly developed roots, and higher risk of plant submergence (Mukhtar et al., 1990; Zaidi et al., 2004; Ren et al., 2014). Therefore, excess moisture stress in the EV stage may affect the response of plant growth on different landscape positions and soil types within a field, which can lead to increased grain yield heterogeneity.

4.4.3 Performance of the MSAR model adjusted yield

The performance of MSAR model adjustment was evaluated at three steps: 1) evaluate the ability to reproduce grain yield using the whole calibration dataset; 2) evaluate the ability to forecast grain yield in a drought year (2012) and a typical year (2016); 3) evaluate the forecast accuracy when extrapolating to different fields across a range of growing seasons using the validation dataset. For (2), the entire dataset except the forecast year was used to calibrate MSAR

model coefficients for the 7 calibration fields, and then the coefficients were used to forecast the maize grain yield for the particular year. For (3), the average of the MSAR coefficients of the 7 calibration fields (35 site-years) was used to forecast maize grain yield in 2010 – 2016 for 30 different validation fields (82 site-years) under various crop rotation practices in the same region.

The MSAR adjusted yield predictions showed better accuracy across three spatial aggregation levels compared to the N Model using the calibration dataset (Figure 4.2 and Figure 4.5). The MSAR adjusted yield predictions resulted in more cases that fell within 15% of actual yield compared to the N Model alone (73 vs. 53, 33 vs. 26, and 7 vs. 6 at the soil type level, site-year level, and year level, respectively). And the MSAR adjusted yield predictions were closer to the 1:1 line and had lower average RME. This result indicated that by taking account of the EV precipitation, the spatial autocorrelation of the yield residual, and the static soil and elevation properties, the MSAR model was capable to improve the prediction accuracy of the N Model.

At the field mean level, the MSAR adjusted yield predictions was superior to the N Model in 25 site-years out of the 35 site-years in the dataset, and MSAR model showed better accuracy compared to a linear model without the spatial autocorrelation specification in 17 site-years (Table 4.4). The spatial autocorrelation parameter showed a narrow range across fields (0.79 – 0.94). The spatial variation improvement was not significant for the remaining 18 site-years possibly because the EV precipitation alone cannot account for all observed spatial variation. Additionally, the measurement error associated with precipitation from nearby weather stations might also affect the relationship. Another source of error for using weather variables for empirical model adjustment was that weather variables other than the selected variable (e.g. temperature, solar radiation, and relative humidity) may alter the empirical estimates through a classical omitted variable problem (Auffhammer and Schlenker, 2014). Moreover, the presence of these other phenomena and their correlations with precipitation may be location specific.

When historical grain yields of each specific field were used to train MSAR model parameters, the MSAR adjusted yield predictions did not improve in the drought year of 2012 (average RME of 24.1%) due to the lack of drought stressed environment in MSAR model training process (Table 4.5). In comparison, the N Model itself showed a satisfying agreement with actual yield in 2012 on the studied fields (Absolute RME < 15%). Conversely, in a typical year such as 2016, the MSAR adjustment exhibited improvement in grain yield prediction compared to the N Model on

6 out of 7 fields, at both field mean level and the 10-m grid level. The N Model overestimated grain yield at all fields, and the mean absolute RME was 13.1% in 2016. The MSAR adjusted yield showed an average absolute RME of 8.1%. This finding suggested that the MSAR adjustment was useful for improving field-scale grain yield predictions of the N Model during a growing season without extreme stresses.

The ability to use the MSAR model to improve maize grain yield prediction accuracy of the N Model was further assessed using a validation dataset containing 82 site-years on 30 different fields in the same region. It was found that the improvement from the MSAR adjustment was not significant (Figure 4.6). The MSAR adjustment did not increase the number of cases where grain yield prediction was within 15% of actual yield, or the average overall accuracy (mean absolute RME was 17.4% and 17.5% for MSAR adjusted and N Model yield predictions, respectively). The N Model showed more overestimation cases with the average ratio between predicted and actual yield of 1.1, where the ratio for the MSAR adjusted yield prediction averaged 1.0. The lack of improvement from the MSAR adjustment could be because the relationships between yield residual and EV precipitation, soil and elevation was location specific. As discussed above, the uncertainty of the precipitation data and other omitted variables could also affect the relationship.

4.5 Conclusions

Robust agricultural process-based models under various environment conditions are needed for site-specific management guidelines. In this study, field-scale maize grain yield prediction results from the N Model in the Mapwindow GIS + MMP Tools showed that the N Model yield prediction was considered reasonable for 74% of the site-year cases (n=35). This finding provides strong evidence that the soil N processes considered in the model are capable of reflecting plant-soil dynamics that are relevant to final yield determination. Our results showed that the N Model predicted higher grain yield than actual yield in most cases, and the linear agreement of predicted and actual yield improved as the spatial aggregation scale became broader. Yield overestimation could be from unaccounted biotic and abiotic plant stresses, the lack of N immobilization process inclusion for maize after maize cultivation, and uneven emergence associated with instances of late planting.

Historical yield and management data are valuable in refining the N Model predictions. Our analysis revealed that yield residuals were linearly associated with early vegetative stage (EV) precipitation ($r = 0.33$). This overestimation could be the result of the high plant N uptake rate during V6 to V12 stage assumed by the N Model ($0.42 \text{ kg N ha}^{-1} \text{ per GDD}$), and more N was considered readily available from applied N fertilizers when excess spring precipitation was not present. We found higher spatial heterogeneity of yield residuals when maize yields of the fields were low ($< 10 \text{ Mg ha}^{-1}$), indicating the N Model had greater spatial uncertainty during harsh environments.

The proposed MSAR model can be used as a tool to adjust field-scale maize yield prediction. The MSAR adjusted yield predictions resulted in more cases that fell within 15% of actual yield compared to the N Model alone when using the whole dataset ($n=35$). However, if the 2012 data was not included in the training process, the MSAR adjusted yield predictions did not improve in the drought year of 2012 (average RME of 24.1%). This finding suggested that the MSAR adjustment was useful for improving field-scale grain yield predictions of the N Model during a growing season without extreme stresses. When extrapolating the MSAR model parameters to a dataset containing 82 site-years on 30 different fields in the same region, the improvement from the MSAR adjustment was not significant. The lack of improvement from the MSAR adjustment could be that the relationship used in the MSAR model was too location specific. The uncertainty of the precipitation data and other omitted variables could also affect these relationships. The approach proposed here could be used to fine tune the N Model for in-season grain yield prediction and N management practices at the field scale.

4.6 Acknowledgements

The authors would like to thank Dr. Brad C. Joern and Phil Hess in their help to evaluate the N Model in MapWindows GIS + MMP Tools. The authors would like to extend gratitude to the Mylet Farms for providing the historical on-farm data for this research project.

4.7 References

- Archontoulis, S.V., F.E. Miguez, and K.J. Moore. 2014. Evaluating APSIM maize, soil water, soil nitrogen, manure, and soil temperature modules in the Midwestern United States. *Agronomy Journal*. 106: 1025-1040.
- Auffhammer, M., and W. Schlenker. 2014. Empirical studies on agricultural impacts and adaptation. *Energy Economics*. 46: 555-561.
- Basso, B., B. Dumont, D. Cammarano, A. Pezzuolo, F. Marinello, and L. Sartori. 2016. Environmental and economic benefits of variable rate nitrogen fertilization in a nitrate vulnerable zone. *Science of the Total Environment*. 545-546: 227-235.
- Batchelor, W.D., B. Basso, and J.O. Paz. 2002. Examples of strategies to analyze spatial and temporal yield variability using crop models. *European Journal of Agronomy*. 18: 141-158.
- Bender, R.R., J.W. Haegerle, M.L. Ruffo, and F.E. Below. 2013. Nutrient uptake, partitioning and remobilization of modern transgenic insect-protected maize hybrids. *Agronomy Journal*. 105: 161-170.
- Boote, K.J., J.W. Jones, J.W. White, S. Asseng, J.I. Lizaso. 2013. Putting mechanisms into crop production models. *Plant, Cell and Environment*. 36: 1658-1672.
- Butler, E.E., and P. Huybers. 2015. Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*. 10: 1-8.
- Cannavo, P., S. Recous, V. Parnaudeau, and R. Reau. 2008. Modeling N dynamics to assess environmental impacts of cropped soils. *Advances in Agronomy*. 97: 131-174.
- Cassman, K.G., A. Dobermann, D.T. Walters, and H. Yang. 2003. Meeting cereal demand while protecting natural resources and improving environmental quality. *Annual Review of Environment and Resources*. 28: 315-358.
- Ciampitti, I.A., and T.J. Vyn. 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.

- Ciampitti, I.A., S.T. Murrell, J.J. Camberato, M. Tuinstra, Y. Xia, P. Friedemann, and T.J. Vyn. 2013. Physiological dynamics of maize nitrogen uptake and partitioning in response to plant density and N stress factors: I. Vegetative phase. *Crop Science*. 53: 2105-2119.
- Colonna, I., M. Ruffo, G. Bollero, and D. Bullock. 2004. A comparison of geostatistical and spatial autoregressive approaches for dealing with spatially correlated residuals in regression analysis for precision agriculture applications. *Annual Conference on Applied Statistics in Agriculture*. Kansas State University. 310-326.
- de Wit, C.T. 1965. Photosynthesis of leaf canopies. *Agricultural Research Report No. 663*. Centre for Agricultural Publications and Documentation. Wageningen.
- Farmaha, B.S., D.B. Lobell, K.E. Boone, K.G. Cassman, H.S. Yang, and P. Grassini. 2016. Contribution of persistent factors to yield gaps in high-yield irrigated maize. *Field Crops Research*. 186: 124-132.
- Fountas, S., D. Wulfsohn, B.S. Blackmore, H.L. Jacobsen, and S.M. Pedersen. 2006. A model of decision-making and information flows for information-intensive agriculture. *Agricultural Systems*. 87(2): 192-210.
- Jones, C.A., and J.R. Kiniry. 1986. *CERES-Maize: A simulation model of maize growth and development*. Texas A&M University Press, Temple, TX.
- Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie. 2003. The DSSAT cropping system model. *European Journal of Agronomy*. 18: 235-265.
- Keating, B.A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, and C.J. Smith. 2003. An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*. 18: 267-288.
- Lobell, D.B. 2009. Crop responses to climate: time-series models. In Lobell, D. and Burke, M. (Eds.) *Climate change and food security: Adapting agriculture to a warmer world*. Springer.
- Lobell, D.B., and M.B. Burke. 2010. On the use of statistical models to predict crop yield responses to climate change. *Agricultural and Forest Meteorology*. 150: 1443-1452.

- Lobell, D.B. 2013. Errors in climate datasets and their effects on statistical crop models. *Agricultural and Forest Meteorology*. 170: 58-66.
- Lobell, D.B., and S. Asseng. 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*. 12: 1-12.
- Morell, F.J., H.S. Yang, K.G. Cassman, J. Van Wart, R.W. Elmore, M. Licht, J.A. Coulter, I.A. Ciampitti, C.M. Pittelkow, S.M. Brouder, P. Thomison, J. Lauer, C. Graham, R. Massey, and P. Grassini. 2016. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? *Field Crops Research*. 192: 1-12.
- Morris, T.F., T.S. Murrell, D.B. Beegle, J.J. Camberato, R.B. Ferguson, J. Grove, Q. Ketterings, P.M. Kyveryga, C.A.M. Laboski, J.M. McGrath, J.J. Meisinger, J. Melkonian, B.N. Moebius-Clune, E.D. Nafziger, D. Osmond, J.E. Sawyer, P.C. Scharf, W. Smith, J.T. Spargo, H.M. van Es, and H. Yang. 2018. Strengths and limitations of nitrogen rate recommendations for corn and opportunities for improvement. *Agronomy Journal*. 110 (1): 1-37.
- Mukhtar, S., J.L. Baker, and R.S. Kanwar. 1990. Corn growth as affected by excess soil water. *Transactions of the American Society of Agricultural Engineers*. 33: 437-442.
- Ozaki, V. A., S.K. Ghosh, B.K. Goodwin, and R. Shirota. 2008. Spatio-temporal modeling of agricultural yield data with an application to pricing crop insurance contracts. *American Journal of Agricultural Economics*. 90: 951-961.
- Puntel, L.A., J.E. Sawyer, D.W. Barker, R. Dietzel, H. Poffenbarger, M.J. Castellano, K.J. Moore, P. Thorburn, and S.V. Archontoulis. 2016. Modeling long-term corn yield response to nitrogen rate and crop rotation. *Frontiers in Plant Science*. 7: 1-18.
- Puntel, L.A., J.E. Sawyer, D.W. Barker, P.J. Thorburn, M.J. Castellano, K.J. Moore, A. VanLoocke, E.A. Heaton, and S.V. Archontoulis. 2018. A systems modeling approach to forecast corn economic optimum nitrogen rate. *Frontiers in Plant Science*. 9: 1-15.
- Ren, B., J. Zhang, X. Li, X. Fan, S. Dong, P. Liu, and B. Zhao. 2014. Effects of waterlogging on the yield and growth of summer maize under field conditions. *Canadian Journal of Plant Science*. 94: 23-31.
- Schlenker, W., and M.J. Roberts. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *PNAS*. 106: 15594-15598.

- Setiyono, T.D., H. Yang, D.T. Walters, A. Dobermann, R.B. Ferguson, D.F. Roberts, D.J. Lyon, D.E. Clay, and K.G. Cassman. 2011. Maize-N: A decision tool for nitrogen management in maize. *Agronomy Journal*. 103: 1276-1283.
- Thompson, L.J., R.B. Ferguson, N. Kitchen, D.W. Franzen, M. Mamo, H. Yang, and J.S. Schepers. 2015. Model and sensor-based recommendation approaches for in-season nitrogen management in corn. *Agronomy Journal*. 107: 2020-2030.
- Tollenaar, M., K. Dzotsi, S. Kumudini, K. Boote, K. Chen, J. Hatfield, J.W. Jones, J.I. Lizaso, R.L. Nielsen, P. Thomison, D.J. Timlin, O. Valentinuz, T.J. Vyn, and H. Yang. 2018. Modeling the effects of genotypic and environmental variation on maize phenology: The phenology subroutine of AgMaize crop model. In J.L. Hatfield, M.V.K. Sivakumar, and J.H. Prueger (eds) *Agroclimatology: Linking agriculture to climate*, Agronomy Monograph 60. 1-10.
- Tu, J. and Xia, Z.G. 2008. Examining spatially varying relationships between land use and water quality using geographically weighted regression I: Model design and evaluations. *Science of the Total Environment*. 407(1): 358-378.
- van Diepen, C.A., J. Wolf, H. van Keulen, and C. Rappoldt. 1989. WOFOST: a simulation model of crop production. *Soil Use and Management*. 5(1): 16-24.
- van Oort, P.A.J., T. Zhang, M.E. de Vries, A.B. Heinemann, and H. Meinke. 2011. Correlation between temperature and phenology prediction error in rice (*Oryza sativa* L.). *Agricultural and Forest Meteorology*. 2011. 151: 1545-1555.
- White, J.W., and G. Hoogenboom. 2009. Crop response to climate: ecophysiological models. In Lobell, D. and Burke, M. (Eds.) *Climate change and food security: Adapting agriculture to a warmer world*. Springer.
- White, J.W., G. Hoogenboom, B.A. Kimball, and G.W. Wall. 2011. Methodologies for simulating impacts of climate change on crop production. *Field Crop Research*. 124: 357-368.
- Yang, H.S., A. Dobermann, J.L. Lindquist, D.T. Walters, and T.J. Arkebauer. 2004. Hybrid-maize – a maize simulation model that combines two crop modeling approaches. *Field Crops Research*. 87: 131-154.

- Zaidi, P.H., S. Rafique, P.K. Rai, N.N. Singh, and G. Srinivasan. 2004. Tolerance to excess moisture in maize (*Zea mays* L.): susceptible crop stages and identification of tolerant genotypes. *Field Crops Research*. 90: 189-202.
- Zhang, L. L. Lei, and D. Yan. 2010. Comparison of two regression models for predicting crop yield. *Proceedings of the IEEE International Symposium on Geoscience and Remote Sensing (IGARSS)*, July 25-30, 2010, IEEE, New York, USA.
- Zhang, T., and G. Lin. 2016. On Moran's I coefficient under heterogeneity. *Computational Statistics and Data Analysis*. 95: 83-94.

Table 4.1 Advantages and disadvantages of the two modeling approaches.

	Process-based models	Statistical models
Advantages	<ul style="list-style-type: none"> • Integrate effects of genetics, environment and crop management • Can be used for problem-based research • Can be used to study the impact of climate change, resource management on the modeled processes, as well as crop production 	<ul style="list-style-type: none"> • Limited reliance on field calibration data (Lobell and Burke, 2010) • Transparent assessment of model uncertainties • Potential to capture effects of poorly understood processes, such as pest dynamics
Disadvantages	<ul style="list-style-type: none"> • Performance depends heavily on the data used to develop and evaluate them • May be missing key processes related to extreme climate conditions (White et al., 2011; van Oort et al., 2011) • Can be difficult to calibrate due to a large number of uncertain parameters (Lobell and Burke, 2010) • Often parameter uncertainty is ignored (Lobell and Burke, 2010) • Improvements are needed to relate to genetics, soil fertility, water stress and pest damage (Boote et al., 2013) 	<ul style="list-style-type: none"> • Difficulty in extrapolating beyond historical extremes (Lobell, 2009) • Collinearity between predictor variables • Assumptions of stationarity • Low signal-to-noise ratios • Model overfitting problem when historical records are limited

Table 4.2 Maize N uptake function in the N Model in the MapWindows GIS + MMP Tools.

GDD (°C)	N Uptake Rate (kg ha ⁻¹ °C ⁻¹)
0-278	0.0605
278-556	0.4234
556-833	0.2822
833-1111	0.2016
1111-1389	0.1008
1389-1528	0.1008

Table 4.3 Input parameters of each field for the N Model.

	Emergence Date							Yield
	2010	2011	2012	2013	2014	2015	2016	Potential [†]
Field 1	4/24	5/24	-	5/19	5/9	-	4/27	16.3 (260)
Field 2	4/26	-	4/21	5/22	5/13	-	5/12	16.1 (257)
Field 3	5/1	-	4/18	5/20	5/13	-	5/4	16.6 (264)
Field 4	5/1	5/26	-	5/20	5/13	-	5/4	16.6 (264)
Field 5	4/25	-	4/21	5/22	-	5/9	5/11	15.9 (254)
Field 6	-	5/19	4/27	5/21	-	5/8	5/11	16.1 (257)
Field 7	-	5/15	5/3	5/21	-	5/8	5/12	15.7 (251)

[†] Numbers outside parentheses are in Mg ha⁻¹, numbers within parentheses are in bu A⁻¹.

Table 4.4 Maize grain yield, MSAR adjusted yield estimates, and model goodness of fit of the calibration dataset.

Field	Year	Mean Y_a	ρ	Mean Y_{hat}	ME	MAE	RME [†]	RMAE		
unit		Mg ha ⁻¹		Mg ha ⁻¹	Mg ha ⁻¹	Mg ha ⁻¹	%	%		
1	2010	9.73	0.87	10.44	-0.72	2.33	-7.4	A B	24.0	A B
	2011	12.77		11.02	1.75	2.39	13.7		18.7	
	2013	11.63		13.81	-2.17	3.18	-18.7	A	27.3	A
	2014	14.26		13.86	0.39	1.59	2.8	A B	11.2	
	2016	14.79		14.42	0.36	1.67	2.5	A B	11.3	A B
2	2010	7.66	0.94	7.23	0.43	2.76	5.6	A B	36.1	A B
	2012	10.89		9.81	1.07	1.97	9.9	B	18.1	B
	2013	8.79		12.72	-3.93	3.98	-44.7	A	45.3	A
	2014	14.05		12.41	1.64	1.85	11.7	B	13.2	B
	2016	14.61		13.93	0.68	1.17	4.6	A B	8.0	A B
3	2010	12.75	0.79	12.15	0.59	1.90	4.7	A	14.9	A
	2012	11.68		10.47	1.21	2.26	10.4		19.3	
	2013	14.25		14.93	-0.68	1.34	-4.8		9.4	A
	2014	13.71		14.48	-0.78	1.54	-5.7	A	11.2	A
	2016	15.00		15.22	-0.22	0.97	-1.4	A B	6.5	A B
4	2010	11.52	0.84	11.32	0.21	2.27	1.8	A	19.7	A
	2011	11.61		12.76	-1.14	2.01	-9.9	A B	17.3	A B
	2013	13.42		13.87	-0.45	1.78	-3.3	A	13.2	A
	2014	13.85		13.46	0.40	1.58	2.9	A	11.4	A
	2016	15.01		13.97	1.03	1.36	6.9	B	9.0	B
5	2010	10.75	0.91	10.81	-0.06	2.05	-0.6	A B	19.0	A B
	2012	10.64		9.76	0.88	2.69	8.3		25.3	
	2013	12.99		14.35	-1.35	2.35	-10.4	A	18.1	A
	2015	10.41		10.10	0.32	3.09	3.0	A	29.7	A
	2016	14.48		14.58	-0.10	1.16	-0.7	A B	8.0	A B
6	2011	14.58	0.80	14.09	0.49	1.41	3.4	A B	9.7	A B
	2012	10.75		9.79	0.96	1.65	8.9	A B	15.4	
	2013	13.79		13.86	-0.08	0.97	-0.5	A B	7.1	A
	2015	13.19		12.71	0.48	1.13	3.6	B	8.6	A B
	2016	12.75		14.57	-1.82	1.84	-14.3	A	14.4	A
7	2011	14.32	0.83	13.95	0.37	1.54	2.6	A B	10.7	A B
	2012	11.90		10.59	1.31	1.95	11.0		16.4	
	2013	13.47		14.00	-0.54	1.25	-4.0	A	9.2	A
	2015	11.33		10.47	0.86	2.18	7.6		19.3	A
	2016	13.04		14.92	-1.88	1.91	-14.4	A	14.7	A

† “A” indicates the relative error is smaller compared to the N Model, and “B” indicates the relative error is smaller compared to the linear model without spatial autocorrelation structure.

Table 4.5 Model goodness of fit comparison of the N Model and MSAR adjusted grain yield in a drought year (2012) and a typical year (2016).

Year	Field	The N Model				MSAR Adjusted		
		Mean Y_a	Mean Y_N	RME	RMAE	Mean Y_{hat}	RME	RMAE
		Mg ha ⁻¹	Mg ha ⁻¹	%	%	Mg ha ⁻¹	%	%
2012	2	10.89	11.62	-6.7	11.4	8.69	20.2	27.9
2012	3	11.68	11.91	-2.0	12.5	9.31	20.3	25.4
2012	5	10.64	11.47	-7.8	11.4	8.63	18.9	32.5
2012	6	10.75	12.20	-13.5	14.0	7.90	26.4	28.2
2012	7	11.90	12.34	-3.7	8.8	7.80	34.5	34.8
2016	1	14.79	15.95	-7.9	13.5	14.29	3.3	11.8
2016	2	14.61	16.62	-13.8	14.4	13.77	5.7	8.8
2016	3	15.00	15.74	-4.9	9.8	15.29	-1.9	6.8
2016	4	15.01	15.45	-3.0	8.1	13.62	9.3	10.8
2016	5	14.48	15.89	-9.8	10.9	14.59	-0.8	8.6
2016	6	12.75	16.20	-27.0	27.0	15.07	-18.1	18.2
2016	7	13.04	16.37	-25.6	25.6	15.41	-18.2	18.3

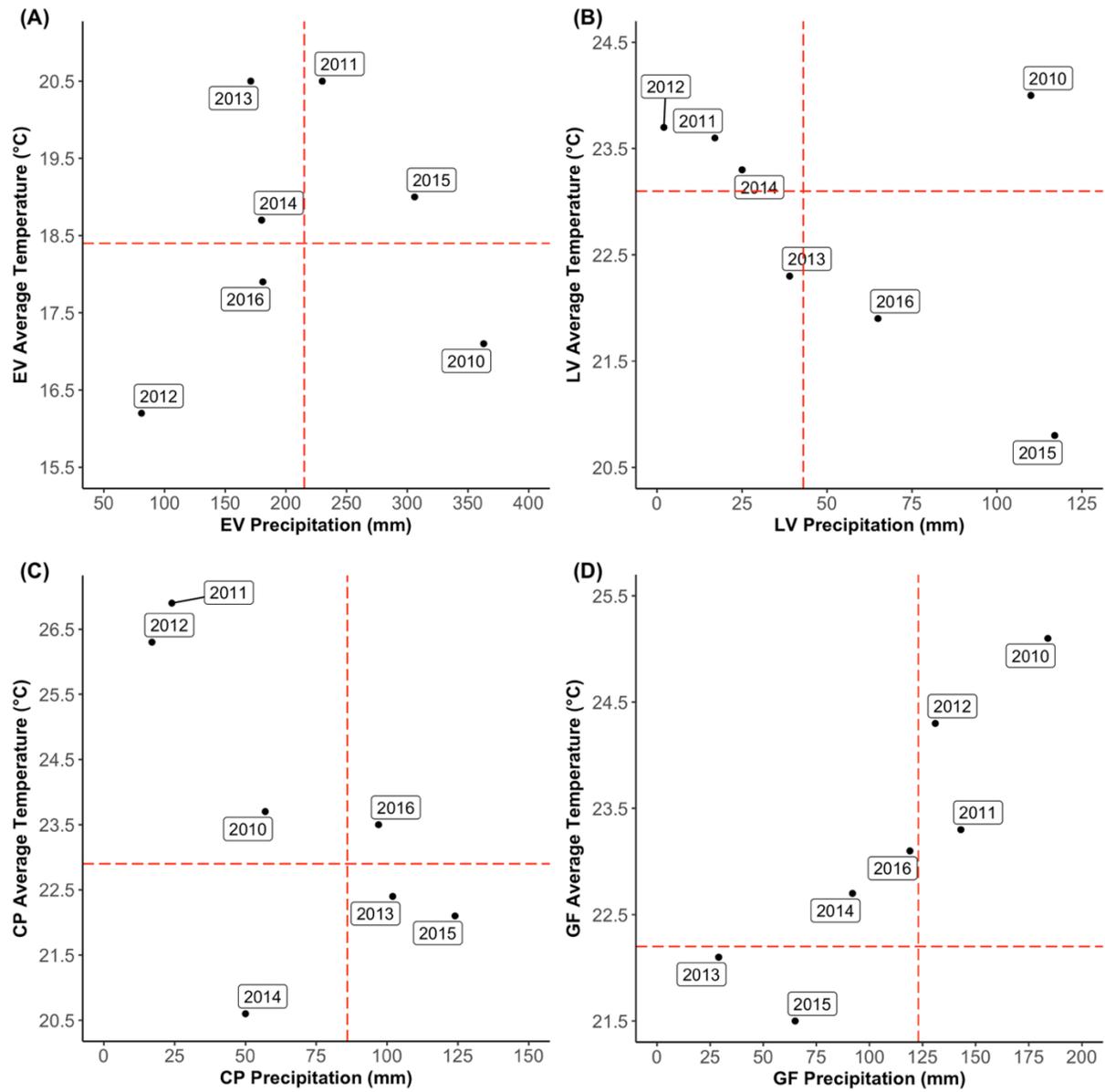


Figure 4.1 General weather conditions during maize growing season during 2010 – 2016. (A) Early Vegetative Stage; (B) Late Vegetative Stage; (C) Critical Period; and (D) Grain-fill Stage. Each point represents the average temperature and precipitation across all fields within a year. The red dash lines indicate long-term weather averages assuming a planting date of May 1st.

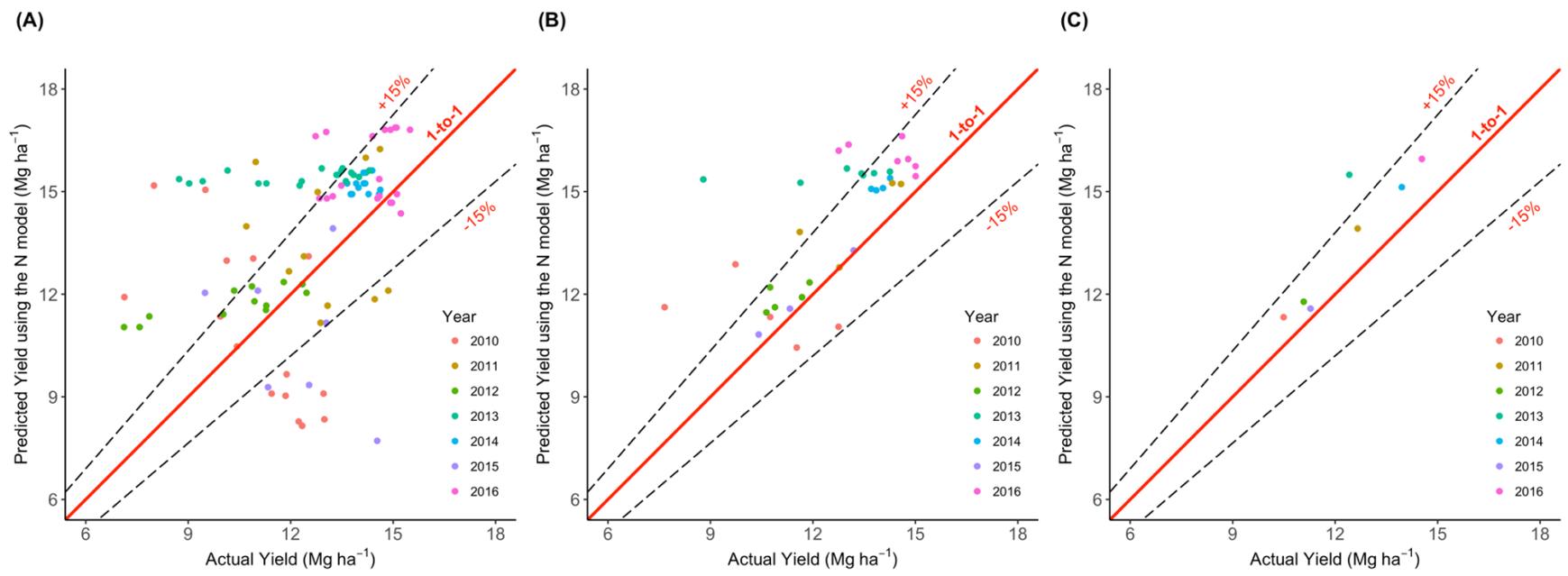


Figure 4.2 Comparison of the N Model predicted and actual maize grain yields at three levels: (A) Soil type within a field; (B) Individual fields; and (C) All fields within a year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown. Data from the calibration dataset with 35 site-years.

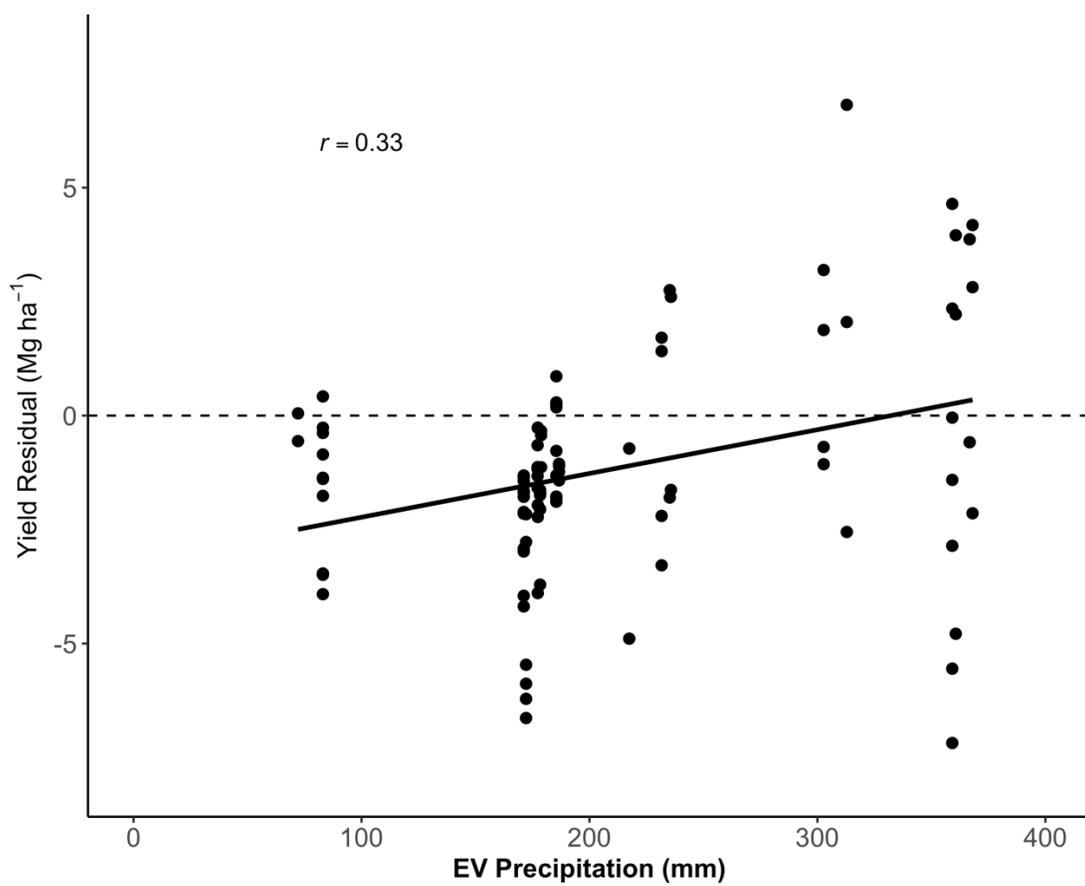


Figure 4.3 Relationship between yield residual at the soil type level and EV precipitation. Each point represents the mean of a soil type within a site-year. Data from the calibration dataset with 35 site-years.

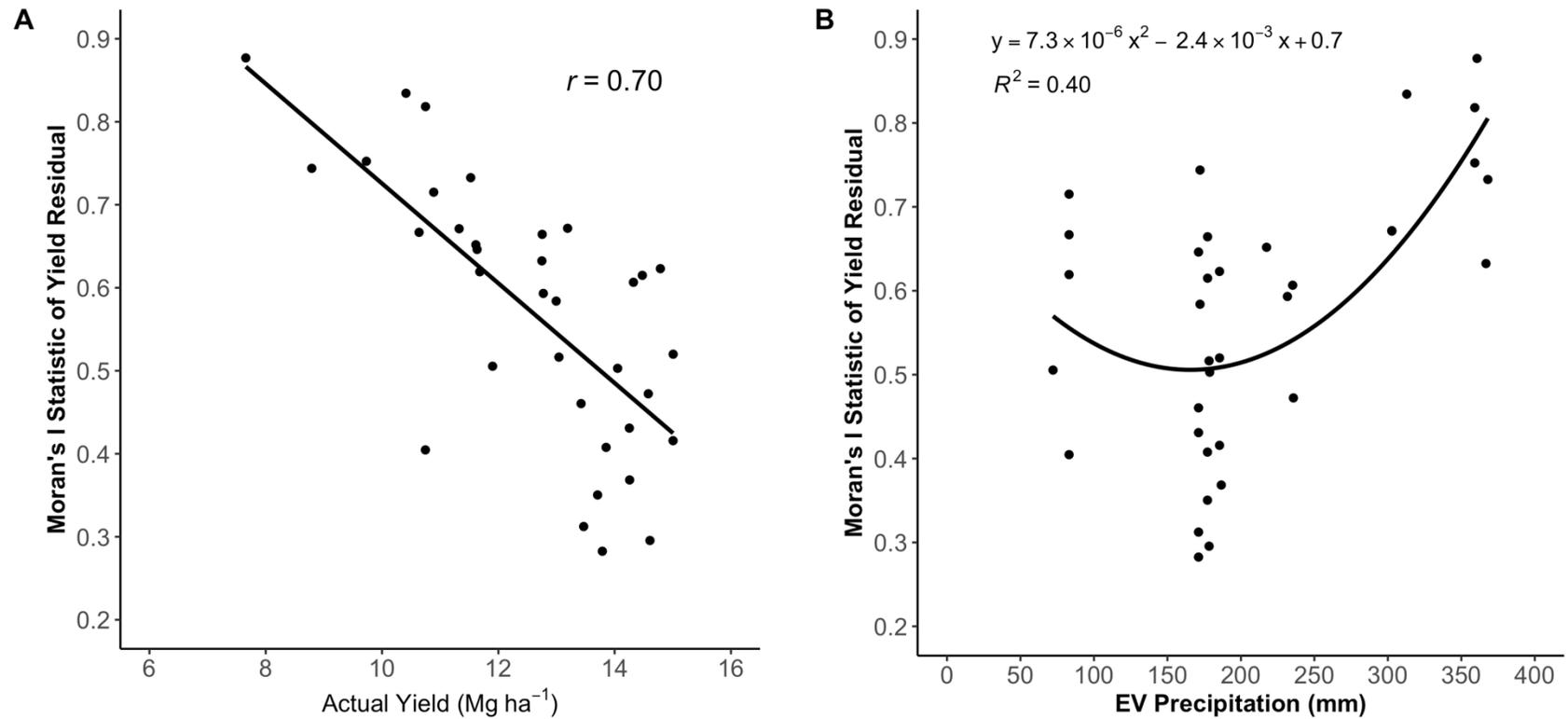


Figure 4.4 Relationship between Moran's I statistic of yield residual and (A) Actual yield; and (B) EV Precipitation. Each point represents the mean of a site-year. Data from the calibration dataset with 35 site-years.

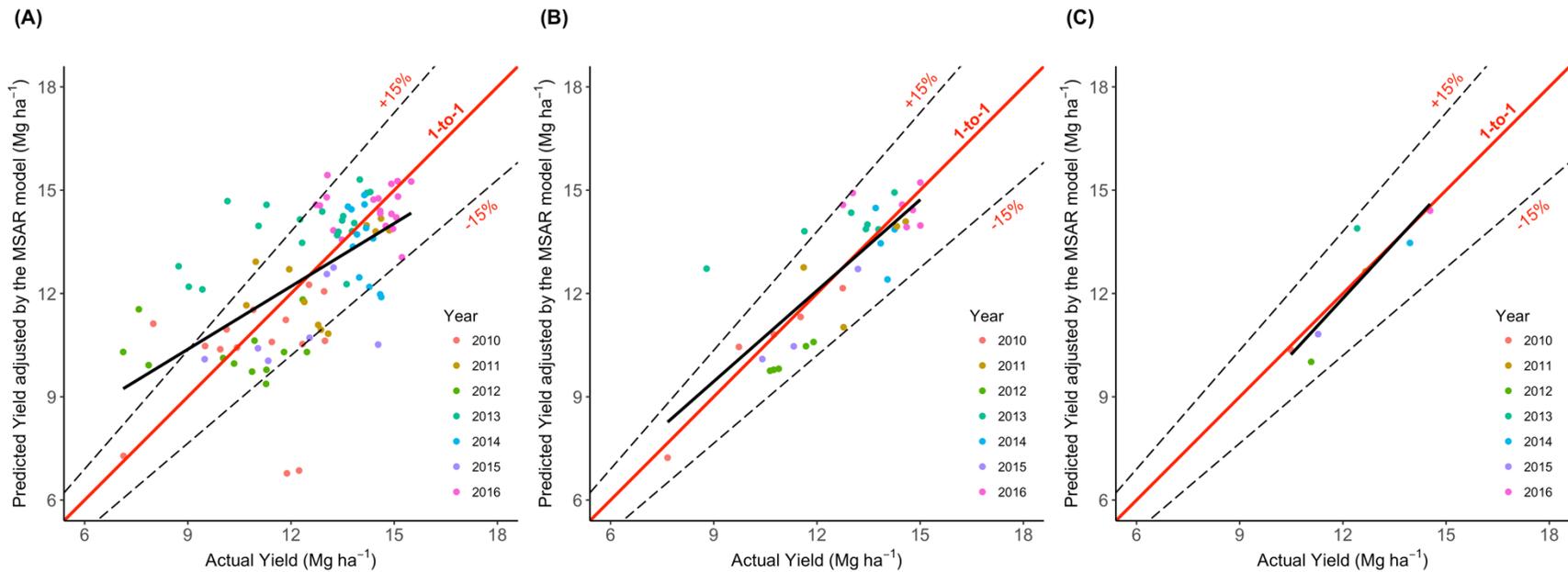


Figure 4.5 Comparison of the MSAR model adjusted maize grain yield and actual maize grain yields at three levels: (A) Soil type within a field; (B) Individual fields; and (C) All fields within a year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown. Data from the calibration dataset with 35 site-years.

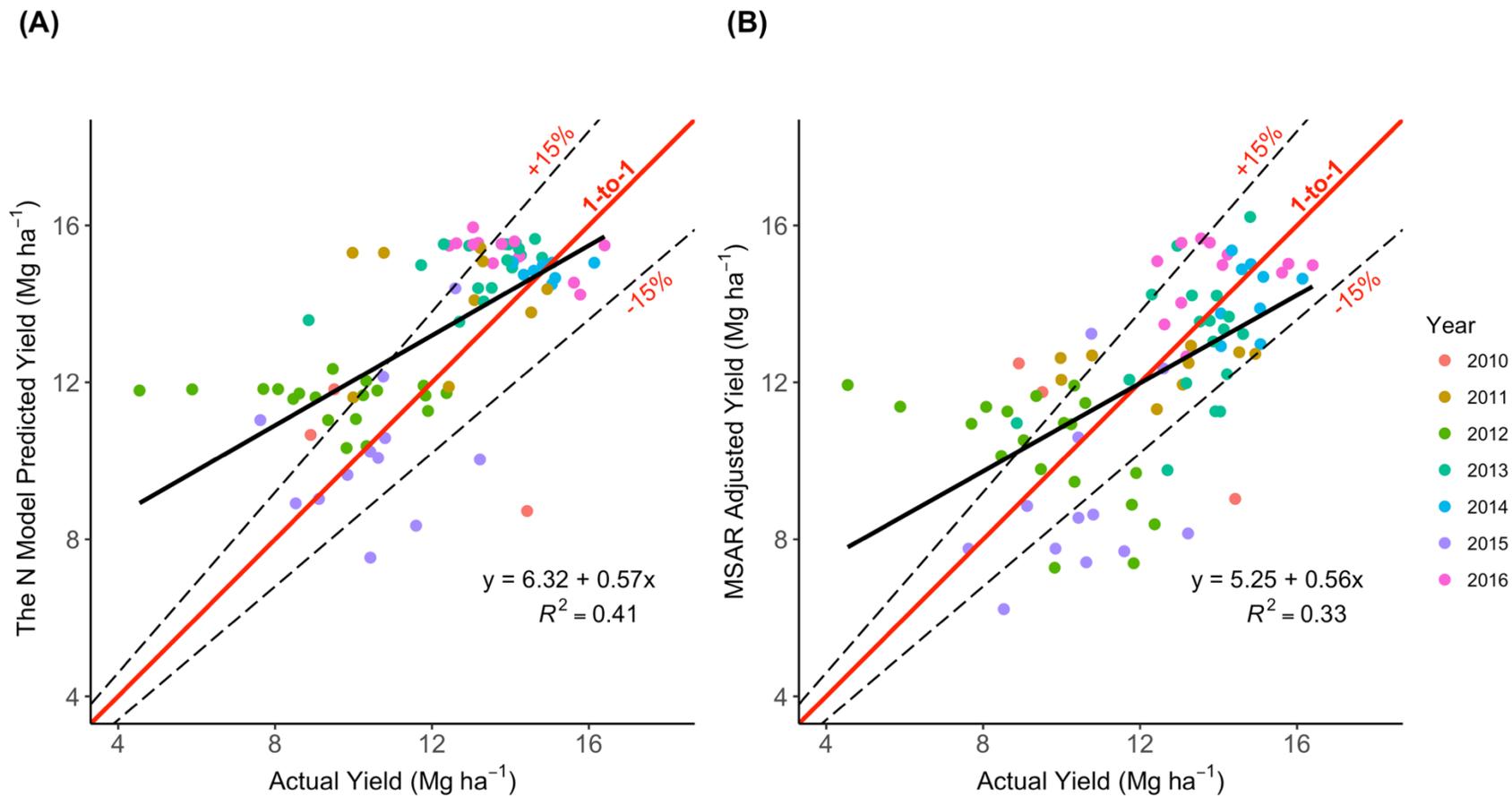


Figure 4.6 Comparison of (A) the N Model yield prediction and (B) the MSAR adjusted yield prediction using the validation dataset ($n=82$). Each point represents the mean of a site-year. The 1:1 line (solid red line) and $\pm 15\%$ deviation (dotted black lines) are shown.

CHAPTER 5. CONCLUSIONS

5.1 Conclusions

The demand for more prescription-based and customized farm management is increasing. Nitrogen (N) management is one of the key elements in the Midwest maize-based agricultural systems. Researchers have been devoted to developing N rate guidelines and N conservation practices for farmers to maximize maize grain yield and maintain environmental sustainability. However, economic optimum N rate was found to vary by soil type and region due to interactions among soil properties and weather conditions (Tremblay et al., 2012). On the other hand, the adoption of winter cover crops has been identified as one of the most effective management practices in reducing non-point N loss via subsurface drainage. Still, cover crop growth is affected by weather, landscape positions and the agronomic practices. Therefore, digital technologies and analytics that can spatially assess the cover crop growth and maize yield under a changing climate are necessary.

In recent years, orbital satellite, unmanned aerial vehicle, soil, water and plant sensors and farm machinery generate large volumes of data containing spatially-precise information that can be analyzed to help make informed decisions and manage risks. The goal of this dissertation research was to use publicly available Landsat data, limited ground-truth samples and historical yield data to establish methodologies to spatially quantify cover crop growth and in-season maize grain yield. The goal was addressed through the following objectives:

- 1) Develop algorithms that use Landsat satellite images and minimal ground samples to predict season-specific cover crop biomass and N uptake on a small watershed in the US Corn Belt; and compare the estimation accuracy with common spatial interpolation methods.
- 2) Develop a multivariate spatial autoregressive model using Landsat satellite images, historical yield data, soil survey and digital elevation to predict field-scale maize yield in Indiana.
- 3) Incorporate multivariate spatial statistics into a process-based N transformation model to predict maize yield on a field scale in Indiana.

Results from Objective 1 confirmed the combination of spatially accurate satellite imagery and limited ground sampling could be used for repeated small watershed assessment of cover crop growth. We found that soil adjusted vegetation index (SAVI), enhanced vegetation index (EVI)

and triangular vegetation index (TVI) were strongly correlated with cover crop biomass and N uptake for low and moderate biomass (0-3000 kg ha⁻¹) and N uptake ranges (0-100 kg N ha⁻¹) in the small watershed of Lake Bloomington in east-central Illinois. These vegetation indices could be used as successful predictors of cover crop biomass production and N uptake. Compared to commonly used spatial interpolation methods such as ordinary kriging (OK) and inverse distance weighting (IDW), using the SAVI method showed higher prediction R² values than that of OK and IDW. Additionally, it would be labor- and resource-intensive to collect high density ground samples for a spatial interpolation analysis when diverse agronomic management practices exist, compared to the remote sensing vegetation indices. In our study, the SAVI estimated cover crop biomass and N uptake were +/- 15% of observed value. This finding demonstrated that remote sensing indices could capture the spatial pattern as affected by various cover crop and cash crop management systems, which are common in small watersheds.

In Objective 2, a new approach using the multivariate spatial autoregressive (MSAR) model was developed at 10-m grid resolution to forecast maize yield using historical grain yield data collected from farmers' fields, publicly available remote sensing maps, site-specific top 30 cm soil organic matter (SOM) and elevation, while accounting for yield spatial autocorrelation. Principal component analysis (PCA) suggested that Landsat green chlorophyll vegetation index (GCVI) during the R1 to R2 development stage in maize was the variable that most closely associated with grain yield ($r = 0.70$), which is attributable to its high correlation with plant LAI and its ability to capture differences in nutrient stress tolerance that are correlated with maize grain yield. Therefore, GCVI was used as the year-specific variable in the MSAR model. In the study, the MSAR model performed reasonably well for overall field maize productivity in 32 out of 35 site-years of the calibration dataset (absolute RME < 15%) with an average absolute RME of 6.6%. The average RMAE of the MSAR model was 13.1%. Because the empirical relationship used in MSAR model was based on historical crop yield, remote sensing of GCVI, and soil and landscape properties, a constant N conversion efficiency and constant harvest index were assumed. Therefore, the MSAR model could result in large prediction errors under extreme stressed environmental conditions such as the 2012 growing season, especially when grain yields under these stressed conditions were not included into the model calibration. Weather conditions for crop growth after the GCVI acquisition date (around R2) may also affect the performance of the MSAR model. In the validation dataset (n=82), the MSAR model showed good prediction accuracy overall ($\pm 15\%$ of observed yield in

56 site-years) in new fields when extreme stress was not present. The novel approach developed in this study demonstrated its ability to use elevation and soil information to interpret satellite observations accurately in a fine spatial scale.

Results from Objective 3 indicated historical yield and management data are valuable in refining the N Model predictions of yield. We evaluated the performance of the process-based N Model developed by Joern and Hess at Purdue University in the Mapwindow GIS + MMP Tools in field-scale maize grain yield prediction, and further identified factors that affected yield prediction accuracy. The N Model was considered reasonable (i.e. RME < 15%) for 74% of the site-year cases (n=35). This finding provides strong evidence that the soil N processes considered in the model are capable of capturing the plant-soil dynamics. Our results showed that the N Model predicted higher grain yield than actual yield in most cases, and the linear agreement of predicted and actual yield improved as the spatial aggregation scale became broader. We found that the yield residuals were linearly associated with early vegetative stage (EV) precipitation ($r = 0.33$). This overestimation could be the result of the high plant N uptake rate during V6 to V12 stage assumed by the N Model ($0.42 \text{ kg N ha}^{-1}$ per GDD), and more N readily available from applied N fertilizers when excess spring precipitation was not present. Additionally, the residual of N Model predicted yield showed significant spatial heterogeneity in all site-years.

A MSAR model was then incorporated into the N Model to adjust maize yield prediction. The proposed MSAR model used EV precipitation, top 30 cm soil organic matter and elevation while accounting for spatial autocorrelation in 10-m grids. The MSAR adjusted yield predictions resulted in more cases (77%) that fell within 15% of actual yield compared to the N Model alone when using the whole dataset. If the 2012 data was not included in the training process, the MSAR adjusted yield predictions did not improve the yield predictions in the 2012 drought year (average RME of 24.1%). When extrapolating the MSAR parameters developed from 7 fields to a dataset containing 82 site-years on 30 different fields in the same region, the improvement from the MSAR adjustment was not significant. The lack of improvement from the MSAR adjustment could be because the relationship used in the MSAR model was location specific. Additionally, the uncertainty of precipitation data could also affect the relationship. The unaccounted impacts from crop rotation and tillage management on variability in maize plant N internal efficiency and grain yield could also affect the transferability of the MSAR model parameters. It is also important to note that the data set used in this study are from fields where N fertilizers were applied (mostly at

planting) at a total N rate that is above the state recommendation. For data sets containing variable N rates and timings, N rate and timing should also be considered as parameters when establishing the MSAR model. The approach proposed here could be used to fine tune the N Model for in-season grain yield prediction at the field scale using historical data.

Through the sequence of research studies, the utility of big data routinely collected at farmers' fields and publicly available satellite data has been improved for field-specific maize yield prediction or cover crop growth estimation. We developed a novel approach to predict in-season field-specific maize grain yield, the result of which can be further implemented to agronomic management adjustment. For farmers who adopted cover crop to reduce nutrient loss to the environment, we developed a method that can be used to rapidly assess cover crop biomass production and N uptake. These results can be used to advance field adaptive N management. For future studies, unmanned aerial vehicles, as well as plant and soil sensors that allow for more flexible and comprehensive data collection, could be used to advance our understanding of cover crop growth and maize yield response to management over time across various landscapes.

5.2 References

Tremblay, N., Y.M. Bouroubi, C. Belec, R.W. Mullen, N.R. Kitchen, W.E. Thomason, S. Ebelhar, D.B. Mengel, W.R. Raun, D.D. Francis, E.D. Vories, and I. Ortiz-Monasterio. 2012. Corn response to nitrogen is influenced by soil texture and weather. *Agronomy Journal*. 104: 1658-1671.

APPENDIX A. SUPPLEMENTARY FIGURES

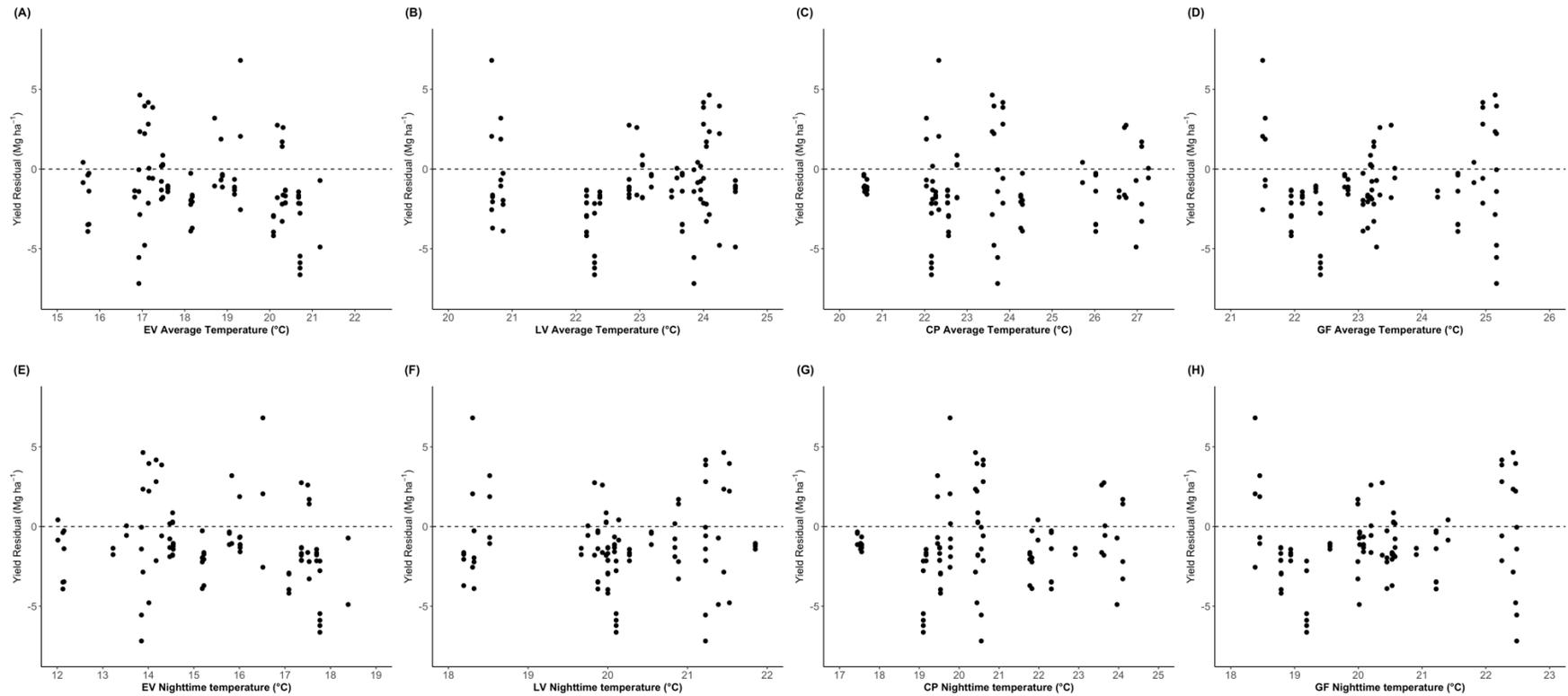


Figure A.1 Relationship between yield residual at the soil type level and (A-D) Average mean temperature during EV, LV, CP, and GF, respectively; (E-H) Average nighttime temperature during EV, LV, CP, and GF, respectively; (I-L) Cumulative solar radiation during EV, LV, CP, and GF, respectively; (M-P) Cumulative KDD during EV, LV, CP, and GF, respectively; (Q-T) Cumulative precipitation during EV, LV, CP, and GF, respectively; (U-X) Average soil moisture during EV, LV, CP, and GF, respectively; (Y-AB) Average VPD during EV, LV, CP, and GF, respectively. Each point represents the mean of a soil type within a site-year. Data from the calibration dataset with 35 site-years.

Figure A.1 continued

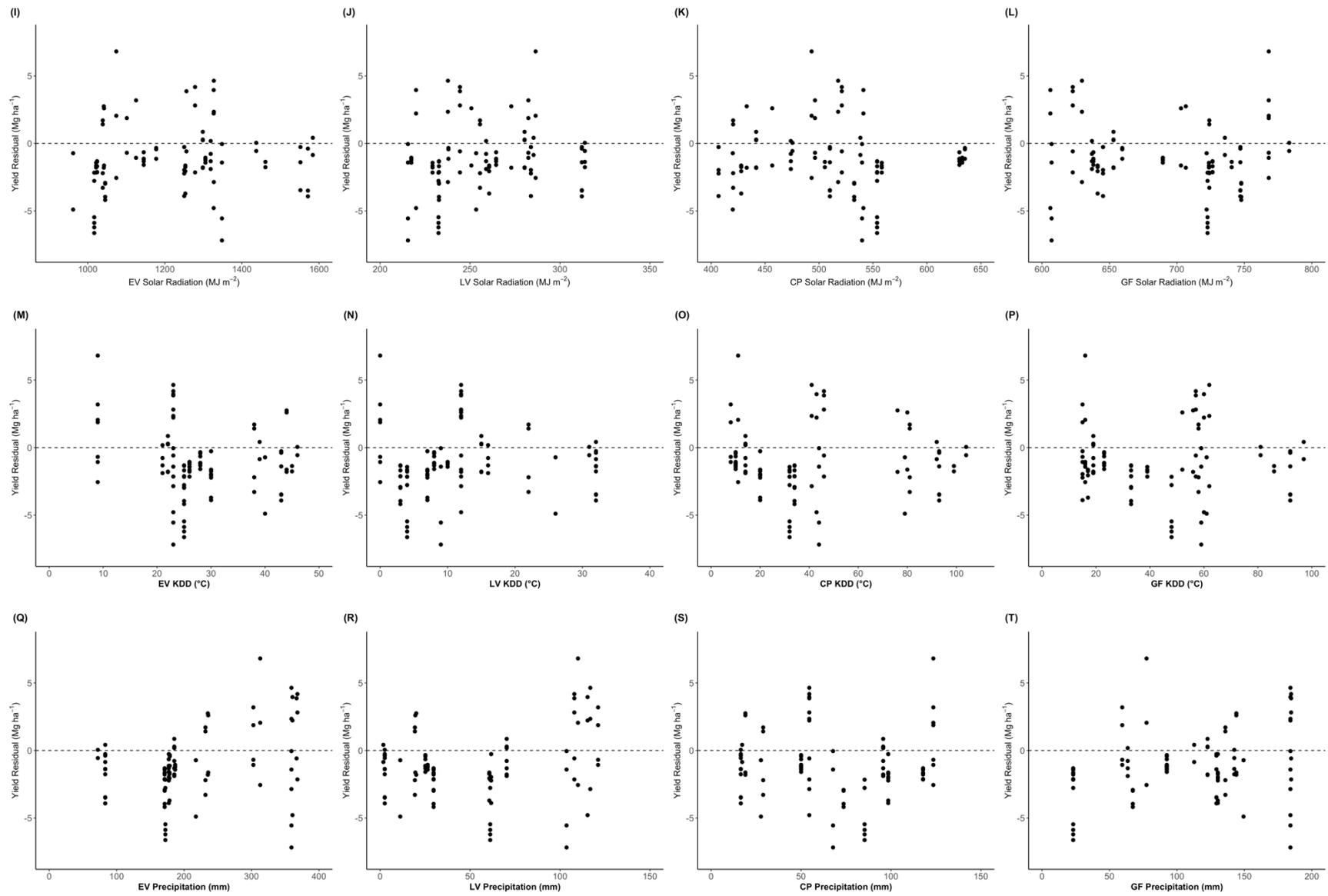
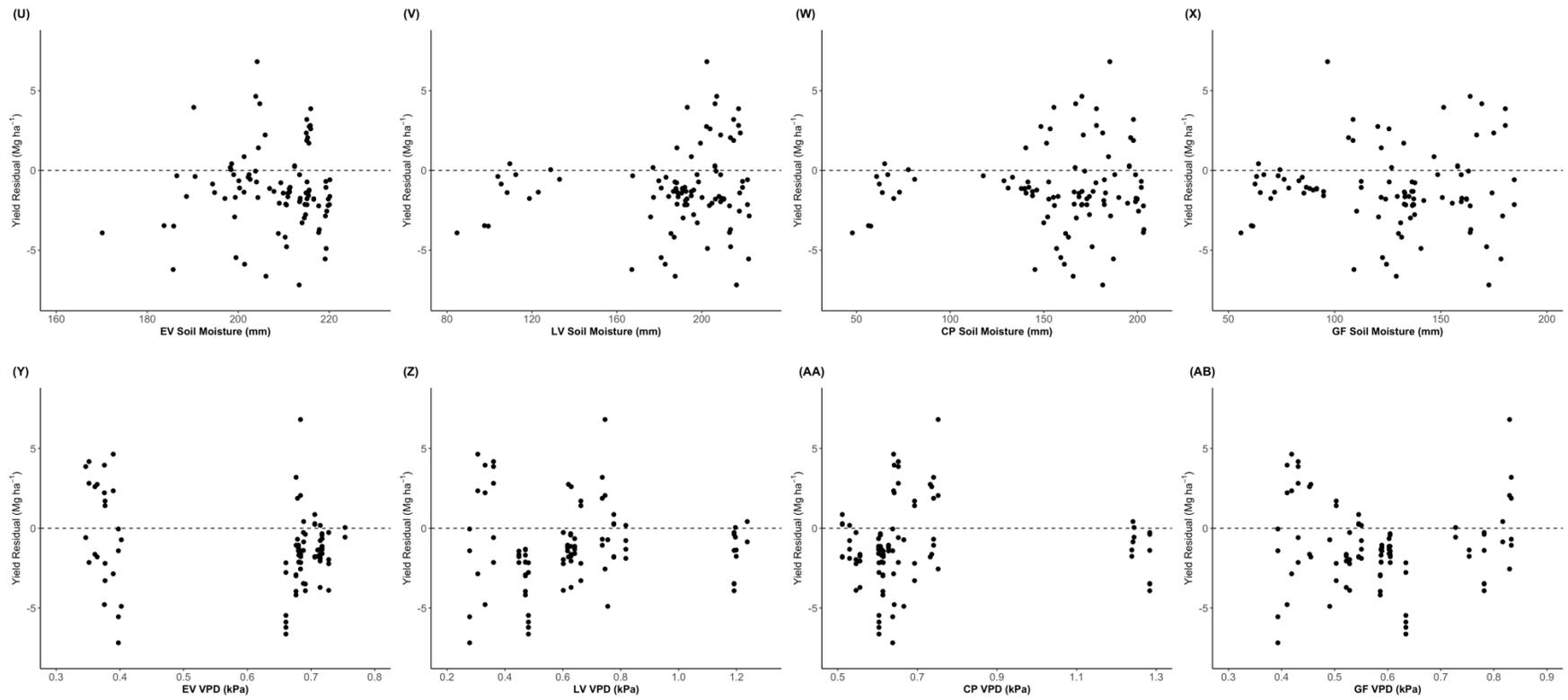


Figure A.1 continued



APPENDIX B. SUPPLEMENTARY TABLES

Table B.1 Planting date, crop rotation, previous crop, soil types, and grain yield of the 82 site-years in the validation dataset.

Year	Farm ID	Field Size	Planting Date	Yield [†]	SD	Rotation	Previous Crop	Soil mukey [§]
		ha		Mg ha ⁻¹	Mg ha ⁻¹			
2011	8	5.31	5/10	9.98	2.40	Other [‡]	Soybean	165222 (100%)
2013	8	9.46	5/06	13.95	1.52	Other	Soybean	165222 (56%), 165231 (44%)
2016	8	4.15	4/25	15.78	0.92	Other	Soybean	165231 (100%)
2013	9	8.62	5/12	14.14	1.66	Other	Soybean	165160 (53%), 165162 (47%)
2014	9	8.62	4/24	14.06	1.48	Other	Maize	165160 (53%), 165162 (47%)
2016	9	4.05	4/25	16.40	1.11	Other	Soybean	165162 (100%)
2010	10	2.23	4/14	14.43	1.16	Other	Soybean	162424 (88%), 162462 (12%)
2012	10	2.58	4/10	10.33	1.47	Other	Soybean	162424 (76%), 162458 (14%), 162462 (10%)
2013	10	9.91	5/07	12.70	1.81	Other	Maize	162446 (94%), 162458 (4%), 162462 (2%)
2015	10	11.63	5/09	8.53	3.78	Other	Soybean	162424 (17%), 162446 (80%), 162458 (3%)
2010	11	1.23	4/13	9.51	2.97	Other	Soybean	165162 (74%), 165222 (26%)
2012	11	3.57	4/02	10.06	2.27	Other	Soybean	165231 (66%), 165162 (25%), 165222 (9%)
2013	11	2.64	5/13	14.27	2.07	Other	Maize	165236 (88%), 165222 (12%)
2015	11	20.51	4/30	12.59	2.81	Other	Soybean	165160 (71%), 165231 (11%), 165236 (11%), 165162 (4%), 165222 (3%)
2012	12	5.60	4/06	10.25	1.93	Other	Soybean	165177 (62%), 165236 (34%), 165239 (4%)
2014	12	2.12	4/22	14.06	2.47	Other	Soybean	165236 (91%), 165239 (9%)
2015	12	5.60	5/04	10.43	3.55	Other	Maize	165177 (62%), 165236 (34%), 165239 (4%)
2012	13	7.05	4/06	8.46	2.21	Other	Soybean	165177 (52%), 165160 (28%), 165239 (18%), 165231 (2%)

Table B.1 continued

Year	Farm ID	Field Size	Planting Date	Yield [†]	SD	Rotation	Previous Crop	Soil mukey [§]
		ha		Mg ha ⁻¹	Mg ha ⁻¹			
2012	14	9.60	4/10	11.84	1.01	Other	Soybean	162423 (100%)
2013	14	9.60	5/15	13.92	1.61	Other	Maize	162423 (100%)
2015	14	9.60	5/07	7.62	4.15	Other	Soybean	162423 (100%)
2013	15	14.76	5/12	13.87	1.57	Other	Soybean	165160 (60%), 165162 (38%), 165231 (2%)
2010	16	1.71	4/10	8.91	2.44	Other	Soybean	165231 (100%)
2012	16	1.71	4/03	9.35	2.35	Other	Soybean	165231 (100%)
2013	16	1.71	5/13	14.81	0.85	Other	Maize	165231 (100%)
2015	16	2.32	5/02	10.76	3.56	Other	Soybean	165231 (74%), 165236 (26%)
2011	17	22.82	5/10	13.30	1.42	CS	Soybean	162446 (62%), 162458 (38%)
2013	17	22.82	5/07	14.63	1.33	CS	Soybean	162446 (62%), 162458 (38%)
2015	17	22.82	5/07	13.22	1.41	CS	Soybean	162446 (62%), 162458 (38%)
2012	18	22.30	4/11	11.79	1.86	CS	Soybean	162446 (66%), 162458 (33%), 162453 (1%)
2011	19	35.04	5/13	14.53	2.03	CS	Soybean	162424 (58%), 162446 (26%), 162455 (5%), 162453 (4%), 162454 (4%), 162462 (3%)
2013	19	35.04	5/07	12.30	0.98	CS	Soybean	162424 (58%), 162446 (26%), 162455 (5%), 162453 (4%), 162454 (4%), 162462 (3%)
2015	19	35.04	5/08	11.59	3.34	CS	Soybean	162424 (58%), 162446 (26%), 162455 (5%), 162453 (4%), 162454 (4%), 162462 (3%)
2011	20	12.17	5/10	13.24	2.70	CS	Soybean	162446 (82%), 162458 (10%), 162453 (8%)
2013	20	12.17	5/07	14.21	1.27	CS	Soybean	162446 (82%), 162458 (10%), 162453 (8%)
2015	20	12.17	5/07	10.63	2.60	CS	Soybean	162446 (82%), 162458 (10%), 162453 (8%)

Table B.1 continued

Year	Farm ID	Field Size	Planting Date	Yield [†]	SD	Rotation	Previous Crop	Soil mukey [§]
		ha		Mg ha ⁻¹	Mg ha ⁻¹			
2011	21	2.43	5/09	10.78	3.40	Other	Soybean	165222 (100%)
2012	21	2.43	4/13	12.37	0.65	Other	Maize	165222 (100%)
2013	21	2.43	5/15	11.73	1.64	Other	Maize	165222 (100%)
2016	21	2.43	4/26	13.18	0.42	Other	Soybean	165222 (100%)
2011	22	28.68	5/09	14.94	1.64	Other	Soybean	165160 (57%), 165177 (40%), 165231 (3%)
2013	22	28.68	5/15	13.77	1.36	Other	Soybean	165160 (57%), 165177 (40%), 165231 (3%)
2016	22	28.68	4/27	13.05	0.49	Other	Soybean	165160 (57%), 165177 (40%), 165231 (3%)
2012	23	9.60	4/10	10.32	1.39	CS	Soybean	165177 (98%), 165231 (2%)
2016	23	9.60	4/25	15.62	0.84	CS	Soybean	165177 (98%), 165231 (2%)
2012	24	14.02	4/12	9.47	2.37	CS	Soybean	165160 (48%), 165163 (42%), 165231 (10%)
2014	24	14.02	4/26	15.07	1.41	CS	Soybean	165160 (48%), 165163 (42%), 165231 (10%)
2016	24	14.02	4/19	12.62	1.37	CS	Soybean	165160 (48%), 165163 (42%), 165231 (10%)
2011	25	14.07	5/14	9.99	2.67	CS	Soybean	165231 (44%), 165240 (38%), 165221 (14%), 165143 (4%)
2013	25	14.07	5/13	13.32	1.71	CS	Soybean	165231 (44%), 165240 (38%), 165221 (14%), 165143 (4%)
2015	25	14.07	5/19	10.43	3.55	CS	Soybean	165231 (44%), 165240 (38%), 165221 (14%), 165143 (4%)
2012	26	6.00	4/12	5.89	2.33	CS	Soybean	165231 (63%), 165163 (25%), 165160 (12%)
2014	26	6.00	4/26	14.60	2.21	CS	Soybean	165231 (63%), 165163 (25%), 165160 (12%)
2016	26	6.00	4/19	14.23	1.90	CS	Soybean	165231 (63%), 165163 (25%), 165160 (12%)

Table B.1 continued

Year	Farm ID	Field Size ha	Planting Date	Yield [†] Mg ha ⁻¹	SD Mg ha ⁻¹	Rotation	Previous Crop	Soil mukey [§]
2012	27	13.46	4/12	7.70	3.41	CS	Soybean	165144 (33%), 165143 (25%), 165221 (20%), 165240 (18%), 165155 (4%)
2014	27	13.46	4/26	16.14	1.86	CS	Soybean	165144 (33%), 165143 (25%), 165221 (20%), 165240 (18%), 165155 (4%)
2016	27	13.46	4/19	12.44	1.52	CS	Soybean	165144 (33%), 165143 (25%), 165221 (20%), 165240 (18%), 165155 (4%)
2012	28	10.38	4/12	4.55	2.02	CS	Soybean	165163 (34%), 165231 (34%), 165230 (32%)
2014	28	10.38	4/26	14.34	2.27	CS	Soybean	165163 (34%), 165231 (34%), 165230 (32%)
2016	28	10.38	4/19	13.55	1.84	CS	Soybean	165163 (34%), 165231 (34%), 165230 (32%)
2012	29	32.19	4/11	8.08	3.47	CS	Soybean	165144 (29%), 165143 (27%), 165221 (14%), 165247 (14%), 165240 (9%), 165249 (7%)
2014	29	32.19	4/26	15.14	1.45	CS	Soybean	165144 (29%), 165143 (27%), 165221 (14%), 165247 (14%), 165240 (9%), 165249 (7%)
2016	29	32.19	4/18	13.05	1.76	CS	Soybean	165144 (29%), 165143 (27%), 165221 (14%), 165247 (14%), 165240 (9%), 165249 (7%)

Table B.1 continued

Year	Farm ID	Field Size ha	Planting Date	Yield [†] Mg ha ⁻¹	SD Mg ha ⁻¹	Rotation	Previous Crop	Soil mukey [§]
2012	30	30.58	4/12	8.62	3.39	CS	Soybean	165144 (37%), 165240 (20%), 165221 (19%), 165193 (13%), 165247 (6%), 165210 (4%), 165167 (1%)
2014	30	30.58	4/26	14.83	1.79	CS	Soybean	165144 (37%), 165240 (20%), 165221 (19%), 165193 (13%), 165247 (6%), 165210 (4%), 165167 (1%)
2016	30	30.58	4/18	13.78	2.16	CS	Soybean	165144 (37%), 165240 (20%), 165221 (19%), 165193 (13%), 165247 (6%), 165210 (4%), 165167 (1%)
2012	31	22.22	4/12	9.03	3.75	CS	Soybean	165234 (26%), 165189 (20%), 165240 (19%), 165247 (18%), 165221 (10%), 165144 (5%), 165210 (2%)
2014	31	22.22	4/27	15.06	1.37	CS	Soybean	165234 (26%), 165189 (20%), 165240 (19%), 165247 (18%), 165221 (10%), 165144 (5%), 165210 (2%)
2016	31	22.22	4/19	14.10	2.40	CS	Soybean	165234 (26%), 165189 (20%), 165240 (19%), 165247 (18%), 165221 (10%), 165144 (5%), 165210 (2%)

Table B.1 continued

Year	Farm ID	Field Size ha	Planting Date	Yield [†] Mg ha ⁻¹	SD Mg ha ⁻¹	Rotation	Previous Crop	Soil mukey [§]
2011	32	33.54	5/14	12.43	2.59	CS	Soybean	165240 (38%), 165221 (32%), 165143 (12%), 165247 (11%), 165234 (3%), 165210 (2%), 165230 (2%)
2013	32	33.54	5/13	13.52	1.86	CS	Soybean	165240 (38%), 165221 (32%), 165143 (12%), 165247 (11%), 165234 (3%), 165210 (2%), 165230 (2%)
2015	32	33.54	5/09	9.12	3.09	CS	Soybean	165240 (38%), 165221 (32%), 165143 (12%), 165247 (11%), 165234 (3%), 165210 (2%), 165230 (2%)
2012	33	14.40	4/05	11.90	1.85	Other	Soybean	165177 (64%), 165160 (36%)
2013	33	5.23	5/14	14.04	1.18	Other	Maize	165160 (100%)
2013	34	22.31	5/14	12.95	2.04	Other	Soybean	165177 (90%), 165231 (5%), 165222 (5%)
2012	35	15.93	4/06	9.82	2.16	Other	Soybean	165160 (75%), 165177 (10%), 165230 (8%), 165239 (5%), 165231 (2%)
2013	35	15.93	5/16	8.86	1.34	Other	Maize	165160 (75%), 165177 (10%), 165230 (8%), 165239 (5%), 165231 (2%)
2015	35	15.93	5/03	10.81	2.35	Other	Soybean	165160 (75%), 165177 (10%), 165230 (8%), 165239 (5%), 165231 (2%)
2012	36	16.56	4/09	10.60	1.63	Other	Soybean	162458 (100%)
2011	37	12.73	5/13	13.09	2.14	CS	Soybean	162446 (61%), 162458 (39%)
2013	37	12.73	5/07	13.18	1.66	CS	Soybean	162446 (61%), 162458 (39%)
2015	37	12.73	5/07	9.84	3.01	CS	Soybean	162446 (61%), 162458 (39%)

[†] average grain yield at 15.5% moisture;

[‡] multiple maize – soybean rotation. The number of continuous maize years varies from 2 to 3;

[§] SSURGO soil map key.

Table B.2 Soil types and soil properties of the fields in maize yield prediction dataset (data from SSURGO soil database).

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
Po	162446	Patton	28	SIL	0.19	0.23	1.31	5	6.7	27.7	4.74
			79	SICL	0.19	0.2	1.36	1.25	7	27	
			96	SICL	0.19	0.2	1.36	1.25	7.6	24.6	
			152	FSL	0.14	0.13	1.54	0.75	7.9	14.2	
RtB	162453	Rush	23	SIL	0.16	0.11	1.45	1.5	6.2	10.5	1.32
			69	SICL	0.19	0.18	1.5	0.75	5.5	15	
			137	L	0.14	0.14	1.5	0.75	5.5	15	
			145	L	0.12	0.1	1.5	0.75	6.2	14	
			152	LCOS	0.04	0.02	1.7	0.25	7.9	3	
SrA	162458	Starks	25	SIL	0.16	0.12	1.46	2	5.7	13.2	1.79
			97	SICL	0.19	0.19	1.47	0.75	5.8	24.4	
			142	L	0.15	0.14	1.52	0.75	6.5	16.4	
			200	SIL	0.14	0.06	1.5	0.25	7.4	7.7	
KgA	165177	Kendall	25	SIL	0.16	0.14	1.45	2	6.2	14.5	1.79
			124	SICL	0.19	0.2	1.5	0.75	5.9	16.5	
			152	L	0.13	0.12	1.7	0.75	7.9	10.5	
Sn	165236	Sloan	30	SIL	0.16	0.18	1.45	3	7	20	3.00
			145	CL	0.19	0.19	1.5	2	7.3	19	
			178	L	0.14	0.12	1.5	0.75	7.5	12.5	
Pk	165222	Pella	38	SICL	0.19	0.22	1.35	4	7	26	4.00
			81	SICL	0.19	0.2	1.5	1.5	7.2	19	
			94	SICL	0.19	0.18	1.6	0.75	7.9	13.5	
			152	SIL	0.15	0.14	1.7	0.75	7.9	13.5	

Table B.2 continued

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
Cz	165160	Cyclone	25	SICL	0.19	0.22	1.3	4.5	6.3	26.9	4.50
			36	SIL	0.16	0.22	1.3	4.5	6.3	26.9	
			51	SICL	0.19	0.16	1.38	1.25	6.9	21.2	
			124	SICL	0.19	0.18	1.4	1.25	7.3	24.1	
			152	L	0.15	0.15	1.44	1.25	7.3	19.7	
			200	L	0.14	0.09	1.46	0.75	8.1	12.7	
FbB	165163	Fincastle	23	SIL	0.16	0.12	1.45	2.5	6.2	13	2.08
			107	SICL	0.19	0.18	1.5	0.75	5.5	16.5	
			122	CL	0.18	0.19	1.6	0.75	6.5	16	
			152	L	0.14	0.11	1.8	0.75	7.9	10	
RrB2	165231	Rockfield	23	SIL	0.16	0.16	1.45	1.5	5.9	12.5	1.32
			81	SICL	0.19	0.2	1.5	0.75	5.3	16.5	
			122	SICL	0.16	0.18	1.6	0.75	6.2	14.5	
			145	L	0.14	0.15	1.6	0.75	7.2	12.9	
			165	L	0.14	0.1	1.8	0.25	8.2	9	
CaA	165143	Camden	23	SIL	0.16	0.13	1.45	1.5	6.2	14	1.32
			94	SICL	0.19	0.18	1.5	0.75	6.2	16.5	
			165	CL	0.17	0.19	1.6	0.75	6.2	13.5	
			203	L	0.14	0.09	1.65	0.75	7.9	8	
OdB2	165210	Ockley	20	SIL	0.16	0.1	1.45	1.5	6.1	9	1.42
			38	SIL	0.16	0.11	1.45	1.25	6.1	10	
			46	SIL	0.15	0.16	1.5	0.75	5.8	14	
			94	CL	0.17	0.17	1.5	0.75	5.6	14	
			124	SCL	0.13	0.11	1.55	0.75	7	11	
			200	COS	0.03	0.01	1.85	0.25	8	2	

Table B.2 continued

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
Pg	165221	Patton	18	SICL	0.19	0.23	1.31	5	6.7	27.7	5.00
			30	SICL	0.19	0.22	1.23	5	6.7	27.7	
			102	SICL	0.19	0.2	1.36	1.25	7	27	
			152	FSL	0.14	0.14	1.53	0.75	7.9	14.2	
RoA	165230	Rockfield	23	SIL	0.16	0.16	1.45	1.5	5.9	12.5	1.32
			81	SICL	0.19	0.2	1.5	0.75	5.3	16.5	
			122	SCL	0.16	0.18	1.6	0.75	6.2	14.5	
			145	L	0.14	0.15	1.6	0.75	7.2	12.5	
			165	L	0.14	0.1	1.8	0.25	7.9	9	
RwA	165234	Rush	23	SIL	0.16	0.11	1.45	2	6.2	12	1.70
			74	SICL	0.19	0.18	1.5	0.75	5.8	17.5	
			89	CL	0.18	0.2	1.6	0.75	5.8	15.5	
			124	SCL	0.13	0.12	1.6	0.75	6.2	12.5	
			142	SL	0.07	0.05	1.7	0.25	7.9	4	
			165	COS	0.03	0.01	1.7	0.25	7.9	3	
StA	165240	Starks	25	SIL	0.16	0.11	1.45	2	6.2	15.5	1.79
			89	SICL	0.19	0.2	1.5	0.75	5.8	18	
			117	L	0.14	0.13	1.6	0.75	6.5	14	
			152	SIL	0.15	0.07	1.7	0.25	7.2	7.5	
WoA	165247	Waynetown	23	SIL	0.16	0.11	1.45	1.5	6.2	11.5	1.32
			81	SICL	0.19	0.19	1.5	0.75	6.1	17.5	
			94	L	0.15	0.15	1.6	0.75	6.1	13	
			137	SCL	0.13	0.12	1.6	0.75	7.2	14	
			152	LCOS	0.04	0.04	1.8	0.25	8.2	2.5	

Table B.2 continued

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
FcA	162424	Fincastle	25	SIL	0.16	0.12	1.45	2.5	5.9	15	2.42
			33	SIL	0.16	0.13	1.45	2	5.9	15	
			69	SICL	0.19	0.18	1.5	0.75	5.8	20	
			127	CL	0.19	0.19	1.6	0.25	6.9	20	
			150	L	0.15	0.15	1.8	0.25	7.4	10	
			200	L	0.13	0.11	1.8	0.25	8	10	
Ge	162425	Gessie	13	SIL	0.16	0.13	1.45	1.5	7.9	14.5	1.07
			91	SIL	0.16	0.13	1.5	0.75	7.9	12	
			152	S	0.05	0.02	1.7	0.25	7.9	2	
RuB	162454	Russell	25	SIL	0.16	0.1	1.45	1.75	6.2	12.5	1.58
			91	SICL	0.19	0.18	1.5	0.75	5.3	19	
			152	CL	0.18	0.19	1.6	0.75	6.2	17	
			203	L	0.14	0.11	1.9	0.25	7.9	9	
RuC	162455	Russell	25	SIL	0.16	0.1	1.45	1.75	6.2	12.5	1.58
			91	SICL	0.19	0.18	1.5	0.75	5.3	19	
			152	CL	0.18	0.19	1.6	0.75	6.2	17	
			203	L	0.14	0.11	1.9	0.25	7.9	9	
SeA	162462	Xenia	25	SIL	0.16	0.11	1.45	2	6.5	12.4	1.79
			76	SICL	0.19	0.2	1.5	0.75	5.3		
			127	CL	0.18	0.19	1.6	0.75	6.5	22.2	
			147	L	0.14	0.12	1.6	0.25	7.9	15.7	
			200	L	0.14	0.11	1.9	0.25	7.9	12.1	
Ba	165139	Beaucoup	28	SICL	0.19	0.22	1.35	2.5	6.7	24.5	2.43
			124	SICL	0.19	0.21	1.5	1.5	7	18.5	
			165	SIL	0.15	0.17	1.6	0.75	7.9	12.5	
Ss	165239	Sloan	38	SIL	0.16	0.18	1.45	3	6.7	19	3.00
			74	L	0.15	0.17	1.5	2	6.7	16.5	
			122	L	0.15	0.15	1.5	0.75	7.2	14	

Table B.2 continued

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
Cy	162423	Cyclone	36	SIL	0.16	0.14	1.34	4.5	6.3	14.4	4.50
			51	SIL	0.16	0.16	1.4	1.25	6.9	21.2	
			124	SICL	0.19	0.19	1.41	1.25	7.3	24.1	
			152	L	0.15	0.15	1.44	1.25	7.3	19.7	
			200	L	0.14	0.09	1.44	0.75	8.7	12.7	
FaA	165162	Fincastle	25	SIL	0.16	0.12	1.45	2.5	5.9	12.5	2.21
			33	SIL	0.16	0.13	1.45	0.75	5.9	14.2	
			69	SICL	0.19	0.18	1.5	0.75	5.8	22.9	
			127	CL	0.18	0.19	1.6	0.25	6.9	21.3	
			150	L	0.14	0.15	1.8	0.25	7.4	15.7	
			200	L	0.13	0.11	1.8	0.25	8	12.1	
Pb	165218	Palms	89	MUCK	0.25	0.33	0.2	75	6.5	185	75.00
			203	L	0.14	0.15	1.6	1.5	7.6	8.5	
CaB2	165144	Camden	23	SIL	0.16	0.12	1.3	1.5	6.2	14	1.31
			94	SICL	0.19	0.18	1.5	0.75	6.2	16.5	
			165	CL	0.17	0.19	1.6	0.75	6.2	13.5	
			178	L	0.14	0.09	1.7	0.75	7.9	8	
Cr	165155	Cohoctah	25	FSL	0.14	0.15	1.45	4.5	7.9	16	4.05
			81	FSL	0.14	0.12	1.6	2	7.9	10	
			112	FSL	0.14	0.11	1.6	1.25	7.9	8.5	
			152	FSL	0.14	0.08	1.6	0.75	7.9	5	
FtC3	165167	Fox	15	CL	0.14	0.17	1.65	0.75	6.2	17	0.75
			66	CL	0.15	0.15	1.65	0.75	6.2	15.5	
			152	COS	0.03	0.03	1.7	0.25	7.9	1	
Ma	165189	Mahalasville	30	SICL	0.19	0.25	1.5	4.5	6.7	23	4.50
			114	SICL	0.19	0.21	1.5	1.25	6.7	18.5	
			127	L	0.14	0.12	1.6	0.75	7.2	12.5	
			165	COS	0.04	0.02	1.7	0.25	7.9	2.5	

Table B.2 continued

Soil musym [†]	Soil mukey [‡]	Soil Series Name	Horizon Depth	Texture	AWC [§]	Wilt Pt	Bulk Density	OM	pH	CEC	0-30 cm OM
			cm		cm cm ⁻¹	cm cm ⁻¹	g cm ⁻³	%		cmol ⁺ kg ⁻¹	%
MfC3	165193	Martinsville	18	CL	0.18	0.19	1.55	1.25	6.1	14.5	1.05
			102	SCL	0.16	0.18	1.6	0.75	5.8	15	
			155	SCL	0.16	0.15	1.6	0.75	6.5	12.5	
			178	L	0.14	0.1	1.8	0.25	7.9	7	
Wr	165249	Westland	25	L	0.15	0.17	1.45	3.5	6.7	19	3.09
			112	CL	0.17	0.19	1.6	1.25	6.7	17	
			130	SCL	0.14	0.14	1.7	0.75	7.2	11	
			152	COS	0.03	0.02	1.8	0.75	7.9	4.5	

[†] SSURGO soil map unit symbol;

[‡] SSURGO soil map key;

[§] Available water capacity.

VITA

Min Xu

Education

Ph.D. Agronomy, Purdue University, West Lafayette, IN

December 2018

GPA 4.0/4.0

M.S. Applied Statistics, Purdue University, West Lafayette, IN

May 2016

GPA 3.91/4.0

M.S. Agronomy, Purdue University, West Lafayette, IN

December 2014

GPA 4.0/4.0

B.S. Resource and Environmental Science, China Agricultural University, Beijing, China

June 2011

GPA 3.79/4.0

Research Experience

Graduate Research Assistant, Purdue University, Advisor: Dr. Shalamar D. Armstrong

May 2017 – Dec 2018

Use satellite remote sensing and multivariate spatial statistics to predict corn yield at the field scale

August 2016 – Dec 2018

- Developed a multivariate spatial autoregressive model using Landsat satellite data, historical yield data, soil survey and digital elevation to predict field scale corn yield in Central Indiana.
- Evaluated the performance of the process-based N Model in the Mapwindow GIS + MMP Tools in predicting corn yield at the field scale.
- Incorporated a multivariate spatial autoregressive model to the N Model to adjust field-scale corn yield prediction.

Estimate cover crop biomass and nitrogen uptake at the watershed scale using Landsat remote sensing and spatial interpolation

May 2017 - May 2018

- Assessed cover crop biomass production and nitrogen uptake of the small watershed of Lake Bloomington in Illinois using Landsat satellite data and limited ground samples.
- Compared the performance of satellite remote sensing to common spatial interpolation methods (ordinary kriging and inverse distance weighting) in estimating cover crop growth on such scale.

Graduate Research Assistant, Purdue University, Advisor: Dr. Brad C. Joern*August 2011 - April 2017**The use of high-carbon soil amendments to improve crop productivity and soil health in reclaimed mined soils*

January 2015 - November 2016

- Conducted a greenhouse study to study high carbon amendments (biochar and composted mulch) impacts on nutrient cycling and wheat growth of a newly reclaimed coal-mined soil under greenhouse conditions.
- Analyzed soil and plant samples from the pot experiment for aggregate stability, microbial activity, soil and plant macro- and micro-nutrients concentration, and plant biomass production.
- Studied the dynamics of nutrient cycling and wheat production under fertilized and unfertilized management scenarios.

Liquid swine manure application timing and InstinctTM impacts on net soil nitrogen mineralization, corn nitrogen uptake and yield

August 2011 - September 2014

- Conducted field-scale swine manure application experiment from 2011 to 2013 at four locations in Indiana.
- Conducted soil incubation to estimate soil nitrogen mineralization potentials for soils collected at four fields with eleven different manure and nitrification inhibitor treatments.
- Analyzed weekly nitrous oxide emissions from the soil surface after manure applications through in-situ sample collection and gas chromatograph.
- Studied soil nitrogen mineralization, corn plant growth and grain yield response of manure application timing and the use of nitrification inhibitor.

Undergraduate Research Assistant, China Agricultural University, Advisors: Dr. Zhong Liu and Dr. Baoguo Li

March 2010 - June 2011

Establishment and modeling analysis of county level soil system

- Data processing of soil surveys and construction of soil system in Lishu County, Jilin Province, China.
- Achieved visualization geography network data model combining with factors affecting soil system establishment (climate, hydrology, soil, traffic, administrative division), remote sensing data through the GIS platform provided by Mapguide Open Source 2.0.

Publication

Xu, M., C.G. Lacey, and S.D. Armstrong. 2018. The feasibility of satellite remote sensing and spatial interpolation to estimate cover crop biomass and nitrogen uptake in a small watershed. *Journal of Soil and Water Conservation*. 73(6): 682-692. doi: 10.2489/jswc.73.6.682

Conference Abstracts

Xu, M., H. Yang, T.J. Vyn, T. Zhang, and S.D. Armstrong. 2018. A new modeling approach: Big data and multivariate spatial statistics to generate in-season field-specific maize yield predictions. ASA Meetings Abstract #112398. Oral Presentation. Agronomic Production Systems Division. Annual Meetings Nov. 4-7, Baltimore, MD.

Xu, M., C.G. Lacey, and S.D. Armstrong. 2018. The feasibility of satellite remote sensing and spatial interpolation to estimate cover crop biomass and nitrogen uptake in a small watershed. ASA Meetings Abstract #112424. Poster (1st place). Agronomic Production Systems Division. Annual Meetings Nov. 4-7, Baltimore, MD.

Martin, J.M., **M. Xu**, and Brad C. Joern. 2016. High-carbon amendment impacts on reclaimed mined land nutrient cycling and wheat growth under greenhouse conditions. SSSA Meetings Abstract #100763. Poster. Soil and Water Management and Conservation Division. Annual Meetings Nov. 6-9, Phoenix, AZ.

<https://scisoc.confex.com/crops/2016am/webprogram/Paper100763.html>

Xu, M., B.C. Joern, and J.J. Camberato. 2014. Liquid swine manure application timing and nitrification inhibitor impacts on soil N transformations and corn growth. SSSA Meetings Abstract #165-4. Poster. Soil Fertility and Plant Nutrition Division. Annual Meetings Nov. 2-5, Long Beach, CA.

<https://scisoc.confex.com/crops/2014am/webprogram/Paper86736.html>

Xu, M., B.C. Joern, and J.J. Camberato. 2012. Effects of swine manure application time and Instinct on corn growth and mineralizable nitrogen on two Indiana soils. SSSA Meetings Abstract #393-21. Poster. Soil Fertility and Plant Nutrition Division. Annual Meetings Oct. 21-24, Cincinnati, OH.

<https://scisoc.confex.com/crops/2012am/webprogram/Paper71029.html>

Teaching Experience

Soil Fertility (AGRY 365) Teaching Assistant, Purdue University

Spring Semester 2013 and 2014

- Prepared and led two weekly labs (18 students in each lab) covering common soil and plant analysis such as estimating cation exchange capacity, organic matter, and nutrient concentration.
- Created a lab preparation booklet to help future TAs organize the labs.
- Worked one-on-one with students to improve their understanding of basic chemistry calculations and soil chemistry.

Soil Physics (AGRY 465/560) Teaching Assistant, Purdue University

Fall Semester 2013

- Organized and taught three weekly labs (5-8 students in each lab) focusing on measurement of soil physical properties, such as soil temperature dynamics, bulk density, particle density, root penetration, soil water retention, and aggregate stability.
- Mentored senior undergraduate students on job searching and graduate school application; gave suggestions to new graduate students on critical thinking, time management.

Soil Science (AGRY 255/270) Teaching Assistant, Purdue University
Fall Semester 2012

- Organized and taught two weekly discussion sessions (14 students in each session) focusing on fundamental soil formation, physical and chemical properties, soil fertility, and soil classification.
- Assisted students to build their knowledge and skills in soil science, to reinforce newly learned concepts to be used in the real world.

Certifications

Machine Learning: Regression (Nov 2018, License MRCRRPWRR6T7)

Machine Learning Foundations: A Case Study Approach (Jul 2018, License KMXHA5HJ7G9Z)

SAS Certified Base Programmer for SAS 9 (Dec 2015, License BP055216v9)

Professional Certificate for Analysis of Soil, Plant and Environment (Mar 2011)

National Computer Rank Examination – C Language (Oct 2009)

Awards and Scholarships

Purdue Agronomy – George D. Scarseth Scholarship (Oct 2018)

Purdue Agronomy – George D. Scarseth Scholarship (Sep 2014)

China Agricultural University - Beijing Outstanding Undergraduate (Jun 2011)

China Agricultural University – Outstanding student scholarship (2010)

Samsung – Scholarship for Agricultural Talents (2009)

China Agricultural University – Scholarship for Excellent students (2008)

PUBLICATION

Xu, M., C.G. Lacey, and S.D. Armstrong. 2018. The feasibility of satellite remote sensing and spatial interpolation to estimate cover crop biomass and nitrogen uptake in a small watershed. *Journal of Soil and Water Conservation*. 73(6): 682-692. doi: 10.2489/jswc.73.6.682