WEATHER- AND PROCESS-BASED MODELS FOR THE ESTIMATION OF MAIZE AND SOYBEAN GROWTH, DEVELOPMENT, AND YIELD

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRAUDATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

VIJAYA RAJ JOSHI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Dr. AXEL GARCIA y GARCIA AND Dr. JEFFREY A. COULTER

JANUARY 2020

Acknowledgements

Table of Contents

iii

List of Tables

List of Figures

Chapter 2

Chapter 4

CHAPTER 1. Introduction

1.1 Background

Field experiments in agricultural studies carried out at multiple sites and over several growing seasons are instrumental in improving crop management efforts, which help to reduce risks from abiotic and biotic stressors in crop production, enhance resources use efficiency, increase yield, and minimize negative environmental effects. However, results from such experiments can be specific to certain locations and can have narrower applicability. Additionally, results from field experiments can vary depending on spatial and temporal variability in crop management practices, weather conditions, and soil properties. In such context, crop models offer opportunities to overcome the shortcomings of field experiments conducted over limited periods and locations by simulating crop growth, development, and yield at various scenarios of weather and soil conditions (Boote et al., 1996; Sinclair and Seligman, 1996).

Crop models ranging in complexity and data requirements have been used in scientific research and as decision support tools (Jones, 1993; Jame and Cutforth, 1996; Setiyono et al., 2011). The use of crop models in the agricultural sector can be broadly categorized into three areas: (i) for synthesis of research knowledge, (ii) for management and decision making in cropping system, and (iii) for analysis of agricultural policies (Boote et al., 1996). Crop models, in general, are developed using two distinct approaches (Mathieu and Aires, 2016; Roberts et al., 2017). One approach uses statistical models whereas another approach uses process-based mathematical models of soil and plant processes. Statistical models have been extensively used to understand the association between crop growth, yield, and pest infestation with several weather variables and soil properties. Weather data-based crop yield estimation model

is one of the most frequently used statistical based crop models (Thompson, 1969; Mathieu and Aires, 2016). In this model, weather data, such as, average air temperature, rainfall or agroclimatic indices are associated with crop yield to develop regression equations for yield estimations. Process-based crop models are the mathematical representation of dynamic cropping systems. Such model include the interaction effects among soil, weather, crop management, and crop genotype through a series of equations characterizing basic physical, biochemical and physiological processes in cropping systems and simulate crop growth, development, and yield at hourly and daily time steps (Jones et al., 2003; Keating et al., 2003).

This dissertation is an integrated effort of field research and simulation studies focused on the application of weather-based and process-based models for crop growth, development and yield estimations. In this dissertation, Chapter 2 is about a study that evaluated the use and relative importance of readily available weather data to develop weather data-based yield estimation models for Maize (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.] in the US central Corn Belt. Total rainfall, average air temperature, and the difference between maximum and minimum air temperature at weekly, biweekly, and monthly time-scales from May to August were used to train multiple linear regression, general additive, and support vector machine models to estimate county-level maize and soybean grain yields for Iowa, Illinois, Indiana, and Minnesota. Chapter 3 is about a study that carried out calibration and evaluation of the simulation accuracy of a process-based crop model, CERES-Maize available within Decision Support System for Agrotechnology Transfer. Evaluations were done on the model ability to simulate maize phenology, final grain yield, and temporal change in aboveground biomass, leaf area index, and aboveground shoot nitrogen content at

variable nitrogen regimes at the growing conditions of two southern Minnesota sites located at Waseca and Lamberton. Chapter 4 is about a study conducted at Lamberton that compared site-specific maize grain yield estimation accuracy of a stand-alone process-based crop model, CERES-Maize, with a data-integration approach where satellite multispectral data were used to spatially optimize soil parameters in the CERES-Maize model.

CHAPTER 2. Weather-based yield estimation models for maize and soybean in the US central Corn Belt

Synopsis

Weather conditions regulate growth and yield of crops, especially in rainfed agricultural systems. This study evaluated the use and relative importance of readily available weather data to develop yield estimation models for maize and soybean in the US central Corn Belt. Total rainfall (Rain), average air temperature (Tavg), and the difference between maximum and minimum air temperature (Tdiff) at weekly, biweekly, and monthly time-scales from May to August were used to estimate county-level maize and soybean grain yields for Iowa, Illinois, Indiana, and Minnesota. Step-wise multiple linear regression (MLR), general additive (GAM), and support vector machine (SVM) models were trained with Rain, Tavg, and with/without Tdiff. For the total study area and at individual state level, SVM outperformed other models at all temporal levels for both maize and soybean. For maize, Tavg and Tdiff during July and August, and Rain during June and July were relatively more important whereas for soybean, Tavg in June and, Tdiff and Rain during August were more important. The SVM model with weekly Rain and Tavg estimated overall maize yield with root mean square error (RMSE) of 591 kg ha^{-1} and soybean yield with RMSE of 205 kg ha^{-1}. Inclusion of Tdiff in model considerably improved yield estimation for both crops; however, the magnitude of improvement varied with model, location, and temporal level of weather data. This study shows the applicability of weather data for reliable yield estimation of maize and soybean as a decision support tool in the US central Corn Belt.

2.1 Introduction

Maize (*Zea mays* L.) and soybean [*Glycine max* (L.) Merr.] production in the US

central Corn Belt is predominantly rainfed and highly dependent on in-season weather

conditions (Franzluebbers et al., 2011; Green et al., 2018). In several areas of this

region, weather variability has been reported to account for more than 75% of maize

yield variation (Ray et al., 2015). Across the US, 40 and 37% yield variability in maize

and soybean was explained by in-season weather conditions, respectively, with an even

greater percentage in the Corn Belt (Leng et al., 2016). Precipitation and air temperature

during the growing season, particularly at late vegetative and early reproductive stages

of development, are reported to cause significant yield deviations from average in both

maize and soybean (Teasdale and Cavigelli, 2017). Given the strong dependency of

rainfed maize and soybean yield on weather conditions, readily available meteorological

data, such as average air temperature (Tavg) and rainfall (Rain), can be used to develop

yield estimation models as decision support tools. Yield estimation is one of the most

important factors for decision-making on crop insurance, crop management, storage

needs, and marketing (Horie et al., 1992; Peng et al., 2018). Thus, the development of a

yield estimation model based on easily accessible weather data can aid farmers in

making informed decisions (Chen et al., 2019) and policy-makers in drafting policies on

trade and food security (Basso et al., 2013; van der Velde et al., 2019).

Several attempts have been made at developing weather-based crop yield

estimation models. For example, Kaul et al. (2005) used field-specific weekly rainfall and

United States Department of Agriculture (USDA) soil rating values to estimate maize and

soybean yield at multiple locations in Maryland, US. Using Tavg and Rain of June, July,

and August with mid-May planting progress, Westcott and Jewison (2013) developed

yield models for maize and soybean in the US. The effects of delayed planting from unfavorable weather conditions at the beginning of the growing season were accounted in their model. However, mid-May planting progress, as measured by the percent of maize plantings by mid-May, can vary within a county and its effect can be difficult to estimate. In addition, the June Rain in their model was used as the deviation amount from normal and only in the years when the June Rain was lower than 10% tail of its distribution. Data on such weather variables are not always easily available and requires calculation before it can be used in the yield estimation model. Mathieu and Aires (2018, 2016) compared more than 50 agro-climatic indices and different weather-impact models for maize yield estimation in the US. They found that Standardized Precipitation Evapotranspiration Index (SPEI) and Tavg in July were the best predictors. Derivation of agro-climatic indices such as SPEI, however, requires data that includes several other parameters, for e.g. potential evapotranspiration, which may not be readily available. Apart from yield estimation, weather-based yield models have also been used to assess crop yield sensitivity to weather variation (Hu et al., 2003; Cai et al., 2014; Mourtzinis et al., 2015) and to predict yield in future or altered climatic conditions (Lobell and Burke, 2010; Verón et al., 2015; D'Agostino and Schlenker, 2016).

Weather-based modeling is one of several methods for crop yield estimation. Broadly, this method can be considered a statistical approach for yield estimation, in which empirical regression equations are developed between yield and other predictors such as several weather parameters collected across multiple site-years (Thompson, 1969; Mathieu and Aires, 2016). In weather-based models, weather data such as Tavg, Rain, or agroclimatic indices (e.g. Palmer drought severity index, SPEI, and growing degree days) are used as predictors in regression equations. In recent decades with the

advancement of remote sensing tools, several spectral indices such as normalized difference vegetation index (Kriegler et al., 1969; Rouse et al., 1974) and enhanced vegetation index (Huete et al., 2002) are also being used to develop crop yield estimation models (Thomasson et al., 2004; Prasad et al., 2006; Franch et al., 2019). Other approaches for crop yield estimation include field observation/sampling and crop simulation modeling. The field observation/sampling is based on careful observation or destructive measurements of representative crop samples. For example, the USDA estimates in-season crop yield based on grower-reported surveys and field-measurement surveys. These surveys are designed to collect data on cultivated area, crop density, row spacing, and yield components (USDA-NASS, 2012). In crop simulation modeling, computer-based models are run with soil, weather, cultivar, and crop management information in order to estimate crop yield (Morell et al., 2016; Lobell and Asseng, 2017; Bhattarai et al., 2017). Crop models ranging in complexity and data requirements have been used in scientific research and as decision support tools (Jones, 1993; Jame and Cutforth, 1996; Setiyono et al., 2011).

Weather-based yield estimation models have several advantages over field observation/sampling and crop simulation modeling approaches. Field observation/sampling often entails trained personnel to collect representative samples from the field for meticulous measurements (USDA-NASS, 2012). Therefore, field observation/sampling approach can be time consuming and expensive. While crop modeling can be advantageous for estimating crop yield from the sub-field to national level (Fraisse et al., 2001; van Wart et al., 2013; Morell et al., 2016), it requires several soil and crop parameters for calibration and evaluation (Verón et al., 2015; Mathieu and Aires, 2016). In contrast, weather-based models rely only on weather parameters or

7

comparatively fewer parameters. Therefore, these models can be advantageous in terms of cost and scalability in both temporal and spatial domains. In previous studies, weather-based models have been used in yield estimation ranging from short-term, such as current year to long-term, such as at the mid or end of the century under future climate change scenarios (Schlenker and Roberts, 2009; Chen et al., 2019). Similarly, regarding spatial scale these models have been applied at the field, regional, and global level (Lobell et al., 2006; Ray et al., 2015; Powell and Reinhard, 2016). Despite these comparative advantages, weather-based models have some limitations. One of these arises from the nature of space-time data, which are often spatially and temporally correlated (Verón et al., 2015; Mathieu and Aires, 2016). In addition, collinearity and interaction among the weather variables can make such data difficult to model (Shi et al., 2013). Moreover, the spatial and temporal level of data used to develop a model can strongly influence the accuracy of yield estimation (Lobell and Burke, 2010; Mathieu and Aires, 2016). Even with these limitations, weather-based models or statistical models in general, are still widely used by the scientific community (Lobell and Burke, 2010; Shi et al., 2013; Mathieu and Aires, 2016).

For model development, multiple linear regression have been widely used, but are gradually being replaced by other more advanced approaches such as mixed effects models (Mathieu and Aires, 2016) and general additive models (Chen et al., 2019). Recent studies have also explored the potential of machine learning approaches, such as random forest model (Jeong et al., 2016), support vector machine regression (Oguntunde et al., 2018), and neural network (Crane-Droesch, 2018). As the availability of weather data has increased with improvement in temporal frequency and spatial resolution of data, more advanced statistical models provide opportunity to develop more

accurate crop yield estimation models. Most studies have used Rain, Tavg, solar radiation, and vapor pressure on monthly or seasonal levels as predictors (Schlenker and Roberts, 2009; Sibley et al., 2014; Lobell et al., 2014; Verón et al., 2015; Tack et al., 2015; Mathieu and Aires, 2016). A limited number of studies have used weather data at daily, weekly, and biweekly time-scales and with more advanced statistical models for crop yield estimation (Kaul et al., 2005; Tack et al., 2015). While the difference between daily maximum and minimum air temperature (Tdiff) has been shown to significantly affect crop yield, its use in yield estimation models is less frequent (Hu et al., 2003; Lobell, 2007; Verón et al., 2015; Tack et al., 2015). With relatively greater increment in daily minimum air temperature as compared to maximum air temperature, studies have showed a decrease in Tdiff on a global scale in the last 50 year (Vose et al., 2005; Lewis et al., 2013). However, the role of Tdiff in maize and soybean yield estimation model development is poorly understood. Understanding the implication of Tdiff on maize and soybean production is crucial for designing resilient cropping systems in the US Corn Belt and global food security, in general. The overarching goal of this study is to assess yield estimation models for maize and soybean grown in the US central Corn Belt using freely and accessible weather data. The specific objectives are to: (1) compare the accuracy of multiple linear regression (MLR), general additive model (GAM), and support vector machine (SVM) regression model to estimate maize and soybean yields, and (2) evaluate if inclusion of Tdiff in model improve model performance.

2.2 Materials and methods

2.2.1 Study period and study area

Historical weather data and grain yield of maize and soybean for a period of 28 years (1990 – 2017) were included for this study. The study period was selected until 2017 since the daily weather data at county level were available until 2017 at the time of data analysis. The study was limited up to 1990 only as it allowed to consider uniform effects from the genetic, agronomic, and technological advancements in crop yield over the study period. The study area focused on the central Corn Belt region of the US and included the major rainfed maize and soybean agricultural districts of Iowa (IA), Illinois (IL), Indiana (IN), and Minnesota (MN). These four states accounted for almost 50 and 44% of total US maize and soybean grain production in 2018 (USDA-NASS, 2019). Only those agricultural districts with non-irrigated maize and soybean production were included. The number of counties included from IA, IL, IN, and MN were 68, 39, 33, and 29, respectively (Figure 2.1). Altogether, the study involved a 28−year period and 169 counties from 16 agricultural districts in four states comprising a total data set of 4732 site-years.

2.2.2 Data collection

Daily data on Tavg and Rain were obtained at the county level from the fine-scaled weather data set for contiguous US developed by Schlenker and Roberts (2009) based on the PRISM Climate Group (2018) weather data. From the difference between daily maximum and minimum air temperature, Tdiff was calculated. The data were prepared at weekly, biweekly and monthly time-scales.

County level annual grain yield data of maize and soybean for the study period were obtained from USDA-NASS (USDA-NASS, 2018) using "nass_data" function from "nassR" package (Dinterman and Eyer, 2018) in R (R Core Team, 2018).

2.2.3 Model development

Yield estimation models were developed for each crop at the total study area level and at individual state level. For model development, in-season (May−August) weather data of Tavg and Rain with/without Tdiff were used as weather variables. Three different statistical models were assessed, which included MLR, GAM, and SVM. Each model was developed with weather data at weekly, biweekly, and monthly time scale.

First, time-series yield data were linearly detrended for each of the 169 counties to remove non-weather effects such as advancement in technology, improved agronomic practices, and cultivar changes over years. The detrending was carried out for maize and soybean yields using the "detrend" function from "pracma" package (Borchers, 2018) in R (R Core Team, 2018). Linear detrending is a common approach for trend removal in crop yield-climate relationship studies, especially for the removal of yield increment in a shorter time-period. Other frequently used detrending approaches are first differencing and non-linear regressions (Lobell and Field, 2007; Lu et al., 2017). After trend removal, MLR, GAM, and SVM yield estimation models were developed.

In MLR, the mean response variable, *y* was modeled as a linear function of *n* explanatory variables x_1, x_2, x_3, x_n, and can be written as:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \cdots b_n x_n + \epsilon \qquad (1)$$

where *b* refers to a regression coefficient and ϵ is the residual. The regression coefficient in MLR is estimated based on the ordinary least square regression that minimizes the sum of squared errors. For this study, MLR was modeled as:

$$y_{i,j} = b_0 + \sum (\beta. X)_{i,j} + \epsilon_{i,j} \tag{2}$$

where $y_{i,j}$ denotes estimated yield of the ith county in the jth year, b_0 refers to the intercept or mean yield, β denotes a vector of regression coefficients for *X* vector of weather variables, namely Tavg, Rain and Tdiff from May to August.

In GAM, the linear relationship as described in MLR was modeled as a smooth function in order to capture any non-linearities in the model (Hastie and Tibshirani, 1990), and can be written as :

$$y = b_0 + f_1 x_1 + f_2 x_2 + f_3 x_3 + \cdots f_n x_n + \epsilon \tag{3}$$

where *f* refers to smooth functions on explanatory variable *x*, which is estimated using scatterplot smoother (Hastie and Tibshirani, 1990). For this study, GAM was modeled as:

$$y_{i,j} = b_0 + \sum (F. X)_{i,j} + \epsilon_{i,j} \tag{4}$$

where *F* denotes a vector of smoother functions for *X* vector of weather variables as mentioned above.

In SVM, a regression function is setup from a training dataset with a margin of tolerance defined by ϵ as the ϵ-insensitive zone. The SVM ignores any error from data points within this margin of tolerance. Non-negative slack variables outside this margin are used to measure the deviation of training samples. The goal is to find an optimal

hyperplane by choosing the function that minimizes the deviation from the insensitivity

parameter (Cortes and Vapnik, 1995; Vapnik, 1999). In a given dataset of (x_i, y_i) pairs,

where i refers to 1,..,n observations, a linear function in SVM can be expressed as:

$$y = f(x) = \omega.x + b \tag{5}$$

where $\boldsymbol{\omega}$.x denotes the product between weighing vector $\boldsymbol{\omega}$ and input vector x, and b

refers to bias term. The objective function in regression analysis using SVM is to

minimize $\boldsymbol{\omega}$. Errors less than ϵ (tolerance margin) are not considered for the objective

function and only the non-negative slack variables across the tolerance margin or ϵ-

insensitive zone are considered (Vapnik, 1999). In non-linear regression, the data are

first linearized through incorporation of kernel functions to apply linear functions. Some

of the commonly used kernels include polynomial, radial basis function and sigmoid. In

this study, radial basis function kernel was used for SVM.

All data analyses were performed in R v. 3.5.1 (R Core Team, 2018) using the

caret package (Kuhn, 2008). The datasets were randomly allocated for training (75% of

total data) and testing (25% of total data) for model development and evaluation,

respectively. For MLR and GAM, data were transformed using the BoxCox

transformation (Box and Cox, 1964), whereas for SVM the data were centered and

scaled before processing. Each model was trained using five-fold cross validation with

five repetitions. For the MLR model, the 'lmStepAIC' method was used for stepwise

selection of variables and selection of the model with the lowest AIC value. For the GAM

model, the 'gam' method from the 'mgcv' package (Wood, 2018) was used. For the SVM

model, the 'svmRadial' method from the 'kernlab' package (Karatzoglou et al., 2018) was

used. The MLR did not have tuning parameters. For GAM, the smoothing parameter

estimation method was set to 'REML' and the selection type to 'TRUE', which penalizes

for new terms in the model and allows removal of terms (Wood, 2017, 2018). For SVM,

the sigma and cost parameters were tuned using the 'tuneLength' command, which was

set to test 15 default values for each parameter.

2.2.4 Importance of weather variables for yield estimation

The relative importance of weather variables in estimating maize and soybean

yields was determined using the 'varImp' function in the caret package (Kuhn, 2008) in R

(R Core Team, 2018). The relative importance was determined by using a locally

weighted least square regression (loess) smoother between crop yield and each weather

variable. The R^2 value from each model was compared with the intercept-only null

model, which was used to determine the relative importance of weather variables. Only

the weather variables grouped at monthly scale were used for this purpose.

2.2.5 Model evaluation

An independent dataset (not used for model development) was used for model

evaluation. Model fitness was evaluated using the root mean square error (RMSE),

normalized RMSE (*n*RMSE), correlation coefficient (*r*), and Willmott's index of

agreement (*d*) (Willmott, 1981).

The RMSE between estimated and actual yield was calculated using the

equation below:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (E_i - A_i)^2}{n}} \qquad (6)$$

where *n* refers to number of observations, *E* and *A* denotes estimated and actual yield. Then, *n*RMSE was calculated as $RMSE/mean(Actual\ yield)$. Lower RMSE and *n*RMSE values indicate better model performance.

The *r* ranges from −1 to +1, with −1 indicating perfect negative linear correlation and +1 indicating perfect positive linear correlation between estimated and actual yield. It was calculated using the equation below:

$$r = \frac{\sum_{i=1}^{n}(A_i - \overline{A})\,(E_i - \overline{E})}{\sqrt{\sum_{i=1}^{n}(A_i - \overline{A})^2}\,\sqrt{\sum_{i=1}^{n}(E_i - \overline{E})^2}} \tag{7}$$

The *d* varies between 0 and 1, with 1 indicating a perfect match and 0 indicating no match between estimated and actual yield and was calculated using the equation below:

$$d = 1 - \frac{\sum_{i=1}^{n}(A_i - E_i)^2}{\sum_{i=1}^{n}(|E_i - \overline{A}| + |A_i - \overline{A}|)^2} \tag{8}$$

2.3 Results

2.3.1 Average maize and soybean yield and yield deviation

Over the total study area, the average yield for maize and soybean was 11,950 and 3,680 kg ha^{-1}, respectively (Table 2.1). The coefficient of variation (CV) was 12% for maize and 11.3% for soybean. For maize, the highest average yield was from MN, followed by IA, IL, and IN. The CV ranged from 9.1 to 13.3% across states. For soybean, the highest average yield was in IL, followed by IA. In IN and MN, the soybean yield was similar at around 3,580 kg ha^{-1}. The CV for soybean yield was 10.4 – 11.5% in all states.

Considerable maize and soybean yield loss occurred in 1993, to a greater extent in IA and MN, due to flooding (Phillips, 1994). Lower than average yield in soybean was observed in 2003, and can be attributed to dry conditions during August and September (Brumm and Hurburgh, 2003). In all states, except MN, the highest negative maize and soybean yield deviation from average occurred in 2012, a historic drought year that caused a significant reduction in US crop yields (USDA-NASS, 2013) (Fig. 2.2, 2.3). In 1994, 2016, and 2017, comparatively yield in all states was above average ranging from 293 to 1635 kg ha^{-1} in maize and from 49 to 508 kg ha^{-1} in soybean (Fig. 2.2, 2.3).

2.3.2 Model development

The SVM model predicted the maize and soybean yield with the lowest RMSE and therefore, performed the best followed by GAM and MLR at all temporal levels of weather data (Fig. 2.4). Inclusion of Tdiff in the model improved yield estimation for both crops; however, the magnitude of improvement varied with model and temporal level of weather data. With the SVM model for maize, inclusion of Tdiff at weekly level did not improve the yield estimation. Similarly, adding Tdiff in the MLR model did not improve yield estimation at monthly level. However, estimation improved significantly ($P < 0.05$) in all other cases. For soybean, inclusion of Tdiff in the SVM model did not improve the estimation until monthly level. Adding Tdiff to the GAM and MLR models significantly ($P < 0.05$) improved yield estimation at all levels.

Within SVM models for maize, no significant difference in yield estimation occurred between weekly and biweekly models; yet, the RMSE from both was significantly lower compared to the monthly model (Fig. 2.3). Similar results were obtained for soybean yield estimations. This reveals the greater efficiency and superior

16

performance of the SVM model compared to the GAM and MLR models. At the state level, the SVM model produced the lowest RMSEs as shown in the Figs. S1−S8 (Electronic Supplementary Material).

2.3.3 Importance of weather variables for yield estimation

In estimating maize yield across the study area, rainfall during June and July and temperature during July and August were more important than other monthly weather variables (Fig. 2.5). The Tavg and Tdiff during July and August were both important. Rainfall during May and August, which coincide with early vegetative and late reproductive stages in maize, respectively were comparatively less important in estimating maize yield.

In estimating soybean yield, Tavg during June, when plants are in early reproductive stages, was the most important weather variable; rainfall during August was the second-most important variable (Fig. 2.5). As in maize, Tdiff during July and August was important for soybean yield estimation.

2.3.4 Model evaluation

The best performing models based on the lowest RMSE (Fig. 2.4), were used to estimate maize and soybean yield using an independent test dataset that was not used previously for model development (Fig. 2.6). As in model development, the SVM model outperformed the other models. The SVM model with weekly weather data of only Tavg and Rain as predictors estimated maize yield with the lowest RMSE of 591 kg ha^{-1} (4.9% nRMSE). The r and d-index were also the highest for the SVM model. After the SVM model, the lowest RMSE was obtained from the GAM model (792 kg ha^{-1}), followed by

the MLR with RMSE of 1065 kg ha^{-1} (Fig. 2.6). The GAM and MLR models used weekly weather data of Tavg, Rain and Tdiff.

As in maize, the SVM model outperformed the other models in soybean yield estimation. The SVM model with weekly Tavg and Rain estimated soybean yield with the lowest RMSE of 205 kg ha^{-1} (5.5% nRMSE). Like in maize, the r and d-index were also highest for the SVM model. Similarly, the next lowest RMSE was obtained from the GAM model (274 kg ha^{-1}) and the MLR model produced the highest RMSE of 320 kg ha^{-1} (Fig. 2.7).

2.3.5 Yield estimation at state level

Models were also evaluated for both crops at individual state level. Since the SVM models gave the lowest RMSEs during model development (Figs. S1−S8, Electronic Supplementary Material), only those models were used for evaluation at the state level.

Overall, maize yield was estimated with RMSE lower than 800 kg ha^{-1} in all states (Fig. 2.8). The lowest RMSE of 472 kg ha^{-1} (3.8% nRMSE) was obtained for IA, followed by MN with 505 kg ha^{-1} (4.2% nRMSE), IL with 660 kg ha^{-1} (5.4% nRMSE), and IN with 733 kg ha^{-1} (6.8% nRMSE). The d-index values were above 0.9 in all states. Except for IN, the values for r were equal or greater than 0.9 in other states.

In all states, soybean yield was estimated with RMSE lower than 250 kg ha^{-1} (Fig. 2.9). The lowest RMSE for soybean yield of 178 kg ha^{-1} (4.8% nRMSE) was obtained for IA, followed by IL with 197 kg ha^{-1} (5.1% nRMSE), MN with 199 kg ha^{-1} (5.7% nRMSE), and IN with 219 kg ha^{-1} (6.1% nRMSE). As for maize, the d-index values for soybean

were also above 0.9 in all states. The values of r were 0.9, 0.86, 0.84, and 0.89 for IA, IL, IN and, MN respectively.

2.4 Discussion and conclusions

Crop yield estimation is an important factor in decision making process for farm management. In this study, weather-based models for maize and soybean yield estimation were developed using easily available weather data of air temperature and rainfall. With statistical modeling, such weather data hold tremendous opportunities in estimating in-season crop yields in a quick and economical way.

The SVM model clearly outperformed the GAM and MLR models. Using only Tavg and Rain data at the weekly level, the SVM model estimated county average maize yield with less than 5% nRMSE and soybean yield with less than 6% nRMSE (Fig. 2.6, 2.7). Comparatively better yield prediction from the SVM than MLR has been reported in other studies as well. Chen et al. (2016) found that the SVM model was more accurate at predicting rice yield from weather variables than the MLR model. Similarly for maize, better model predictions with the SVM model compared to the MLR were reported by Karimi et al. (2008). Superior performance of the SVM model can be attributed to its ability to model nonlinear functions and high dimensional data (Vapnik, 1999). Crop yield data in relation to weather parameters are nonlinear in nature (Schlenker and Roberts, 2006, 2009; Lobell et al., 2011; Tack et al., 2015). Therefore, unsurprisingly, the MLR model which only modeled the linear relationship between weather variables and yield, performed the worst in this study (Fig. 2.6, 2.7). The GAM model, despite being able to model nonlinear relationships, did not give better estimations compared to the SVM model, but were better than those with the MLR model.

This study demonstrated the importance of Tdiff in maize and soybean yield estimation. Although the lowest RMSEs were obtained from models with Tdiff, the differences from models with and without Tdiff were not always significant (Fig. 2.4). Improvement in yield estimates with Tdiff were dependent on the temporal level of weather data, statistical model, and crop. Regardless of the temporal level of weather data, inclusion of Tdiff always significantly improved yield estimation with the GAM model (Fig. 2.4). Inclusion of Tdiff in the SVM model at the weekly level for maize and the weekly and biweekly levels for soybean did not improve yield estimates. However, including Tdiff in the model at the monthly level produced significantly lower RMSE for both crops. Yield estimates for maize and soybean obtained from monthly weather data with the SVM model were more accurate (lower RMSEs) than those from weekly and biweekly weather data from the GAM and MLR models.

The variable importance analysis showed that Tdiff played a key role in model development (Fig. 2.5). For maize, Tavg and Tdiff during July and August were important weather variables. Similarly, Tdiff during August and July were important for soybean. This highlights the important implication of Tdiff on maize and soybean yield. Lesser Tdiff owing to higher night temperature has been reported to reduce maize yield (Peters et al., 1971; Chang, 1981; Cantarero et al., 1999) and soybean yield (Peters et al., 1971). Higher night temperature, especially during the reproductive phase, can reduce crop yield by increasing respiration and the duration of the reproductive phase (Chang, 1981; Cantarero et al., 1999). Increased respiration reduces stored photo-assimilates in crops, thereby reducing their translocation to grains. Higher values of Tdiff have also been shown to reduce crop yields. Lobell (2007) found a nonlinear response of Tdiff to cereal grain yield and a negative response of maize yield to increased Tdiff.

This finding can have an important implication on crop production and climate change studies. Most studies on climate change have focused primarily on the effects of increased air temperature, but difference between maximum and minimum air temperature have received less attention. Results from this study as well as similar results from past studies (Chang, 1981; Cantarero et al., 1999; Lobell, 2007) suggest that Tdiff is important to consider while assessing the impact of climate change on maize and soybean production. As the Tdiff has been decreasing on a global scale (Vose et al., 2005), it is critical to assess its impacts on future maize and soybean production in the US Corn Belt. Decrease in Tdiff due to warmer nights can be advantageous in some regions with too low air temperature as in northern MN but can be damaging to crops in regions with high daily air temperature near optimal. Understanding such implications of weather conditions on crop production will help to design cropping systems that are resilient to future weather situations.

This study demonstrates the usefulness of weather data for reliable estimation of maize and soybean yield in the US central Corn Belt. This study also highlights the comparative significance of weather variables in estimating maize and soybean yield. Improvements in weather-based crop yield estimation are highlighted with respect to different statistical models and weather variables. The comparative analyses of model performances indicated that the SVM model with radial kernel is better as compared to multiple linear regression and general additive model in estimating maize and soybean yields. The inclusion of the difference in daily maximum and minimum air temperature as an explanatory variable improved the performance of all three models assessed. However, such improvement showed to be dependent on the temporal level of weather data, statistical model, and crop type. Due to inherent micro-environments in the

agricultural landscape, future research should investigate the performance of such

weather-based models in explaining spatial crop yield variability occurring within field.

Table 2.1. Total study area and state average detrended yield and coefficient of variation (CV) of maize and soybean during the study period (1990−2017).

State	Maize		Soybean	
	Average yield (kg ha^{-1})	CV (%)	Average yield (kg ha^{-1})	CV (%)
Iowa	12,260	9.8	3,700	11.1
Illinois	12,160	12.2	3,830	10.4
Indiana	10,740	13.3	3,580	10.9
Minnesota	12,300	9.1	3,580	11.5
Total area	11,950	12.0	3,680	11.3
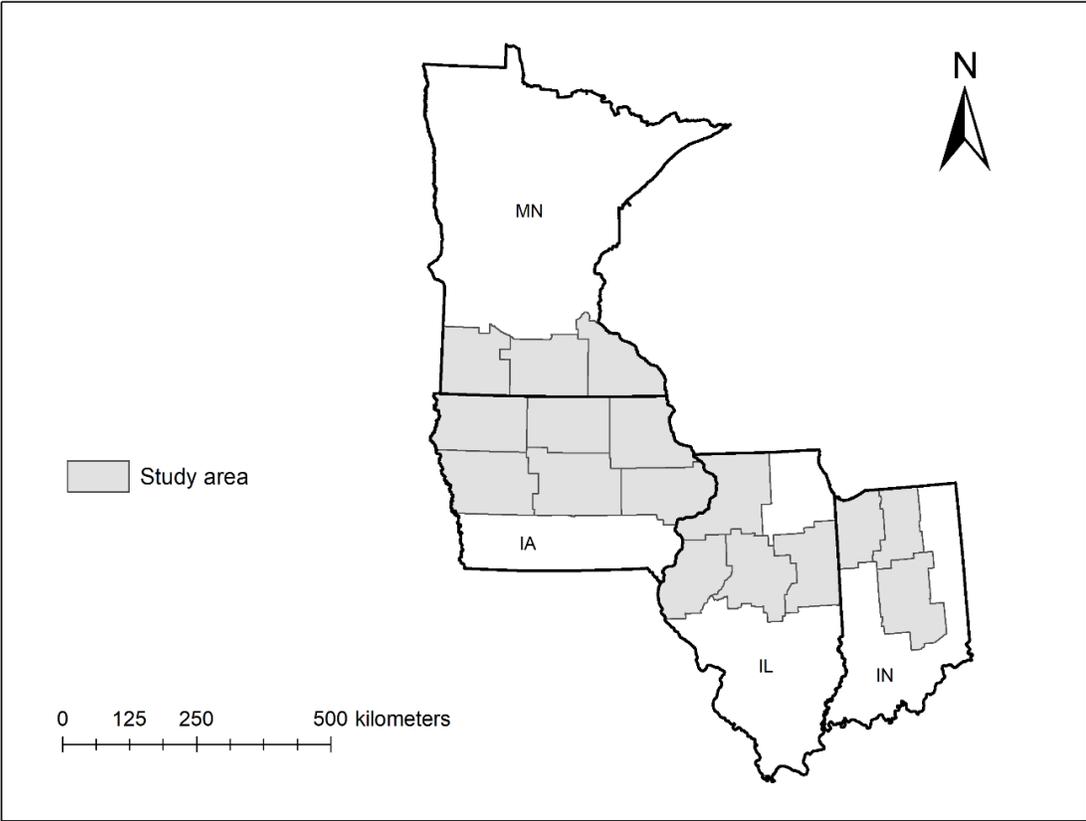
Figure 2.1. Map of the US states and the agricultural districts included in this study. In the map, MN, IA, IL, and IN refers to Minnesota, Iowa, Illinois, and Indiana, respectively.
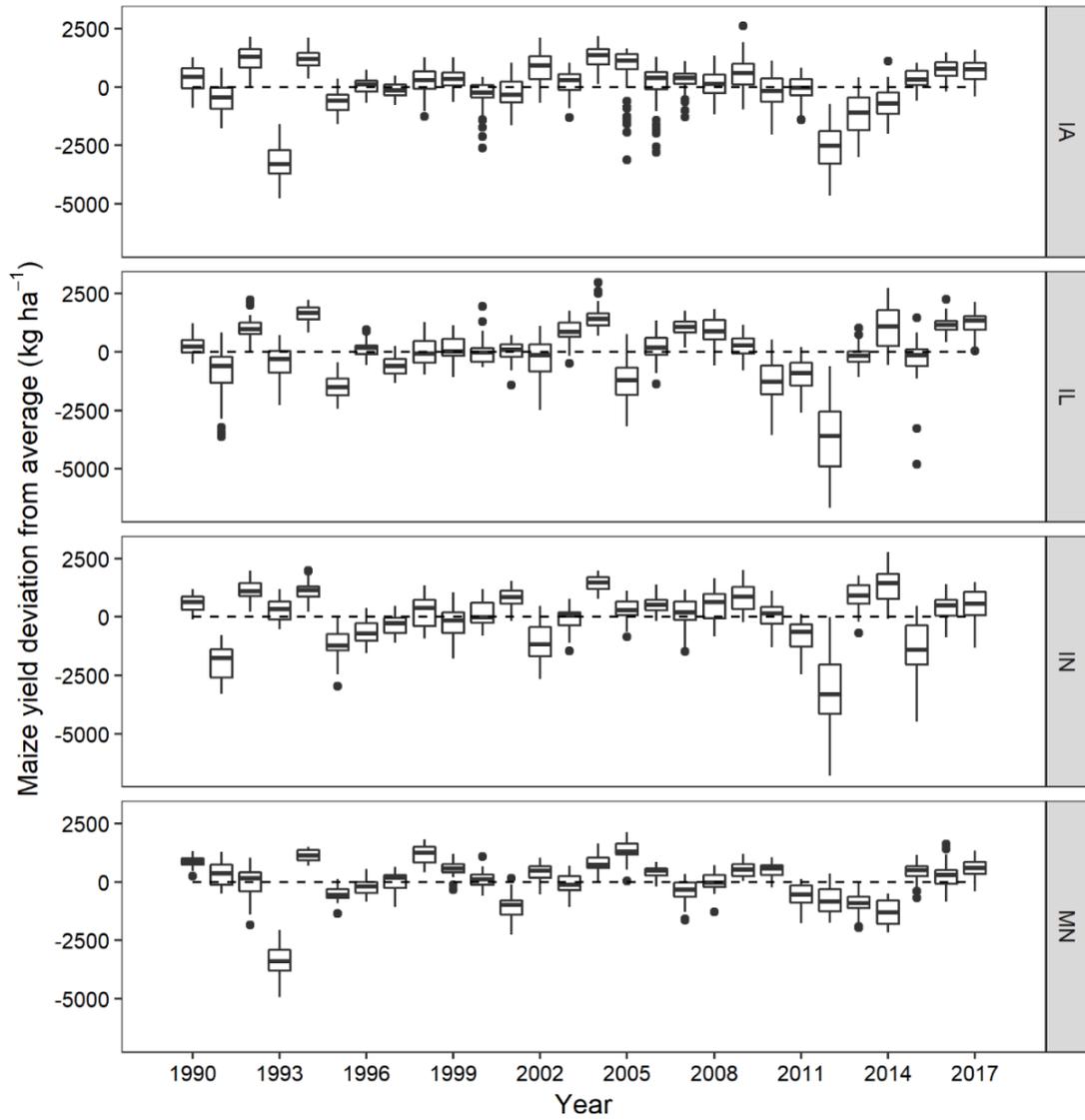
Figure 2.2. Maize yield deviation from average in Iowa (IA), Illinois (IL), Indiana (IN), and Minnesota (MN) from 1990−2017. The dotted black line represents the zero-reference line.
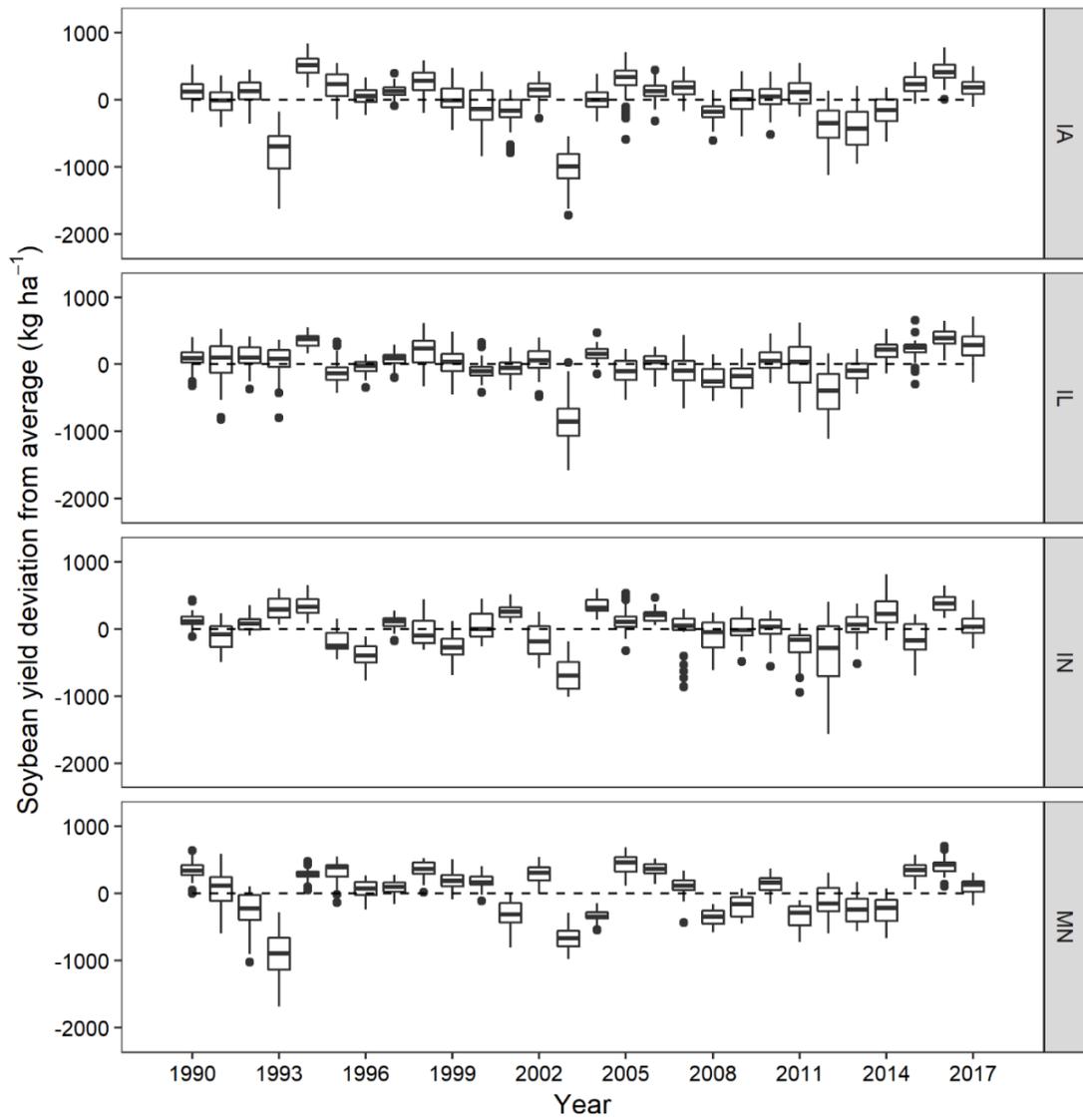
Figure 2.3. Soybean yield deviation from average in Iowa (IA), Illinois (IL), Indiana (IN), and Minnesota (MN) from 1990−2017. The dotted black line represents the zero-reference line.
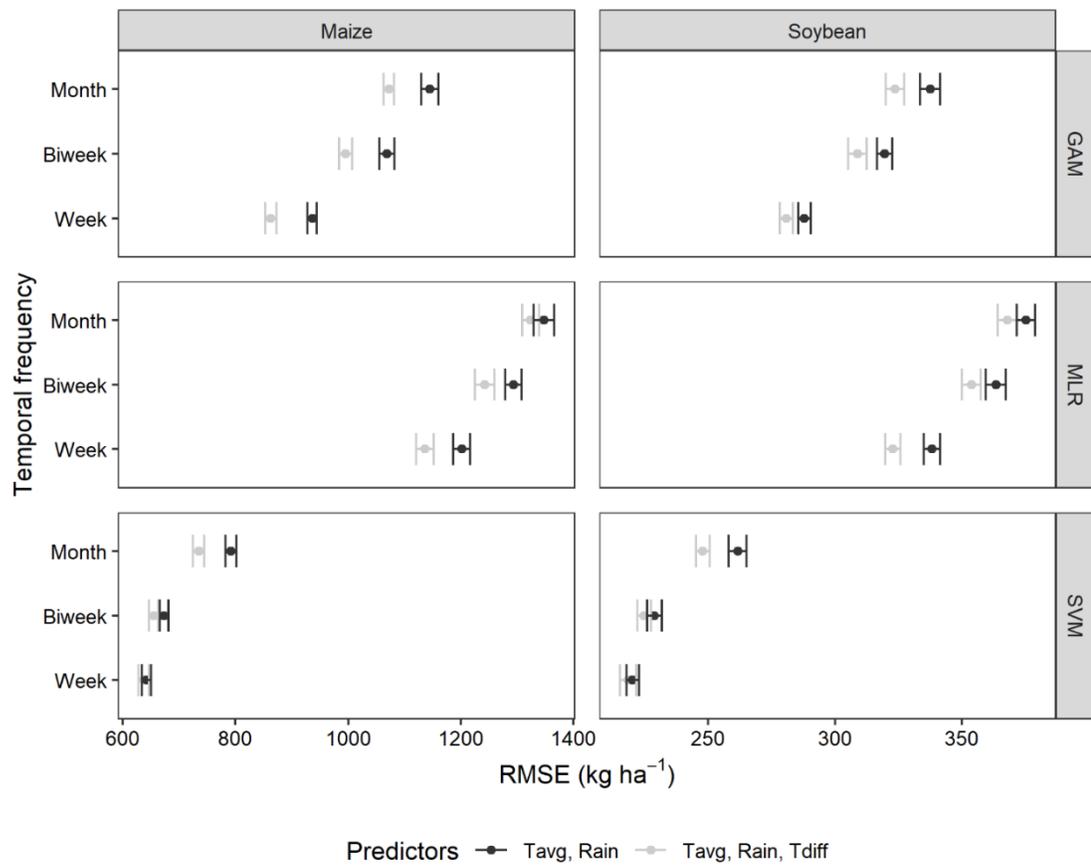
Figure 2.4. Root mean square error (RMSE) of maize and soybean yield estimation during model development for the generalized additive model (GAM), multiple linear regression (MLR), and support vector machine (SVM) models trained with weather predictors at weekly, biweekly and monthly time-scales.
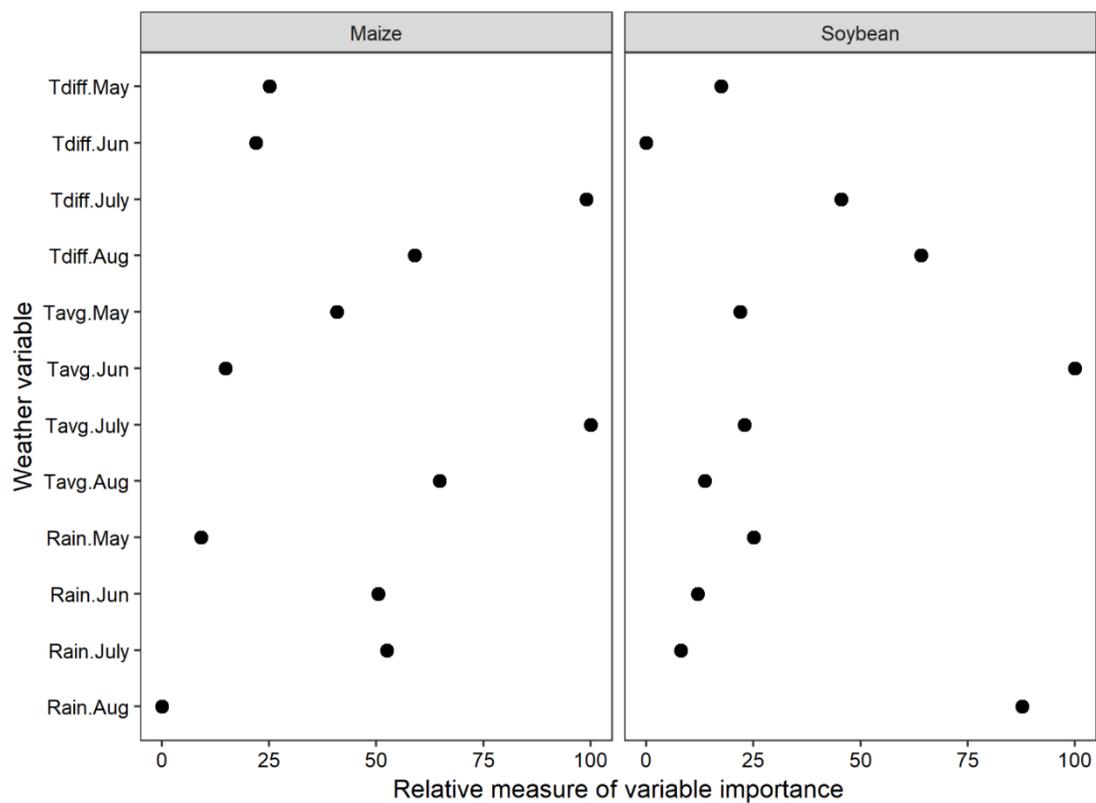
Figure 2.5. Relative measure of monthly weather variable importance scores in estimating maize and soybean yield across Iowa, Illinois, Indiana, and Minnesota. (Importance scores were normalized and scaled between 0 and 100).
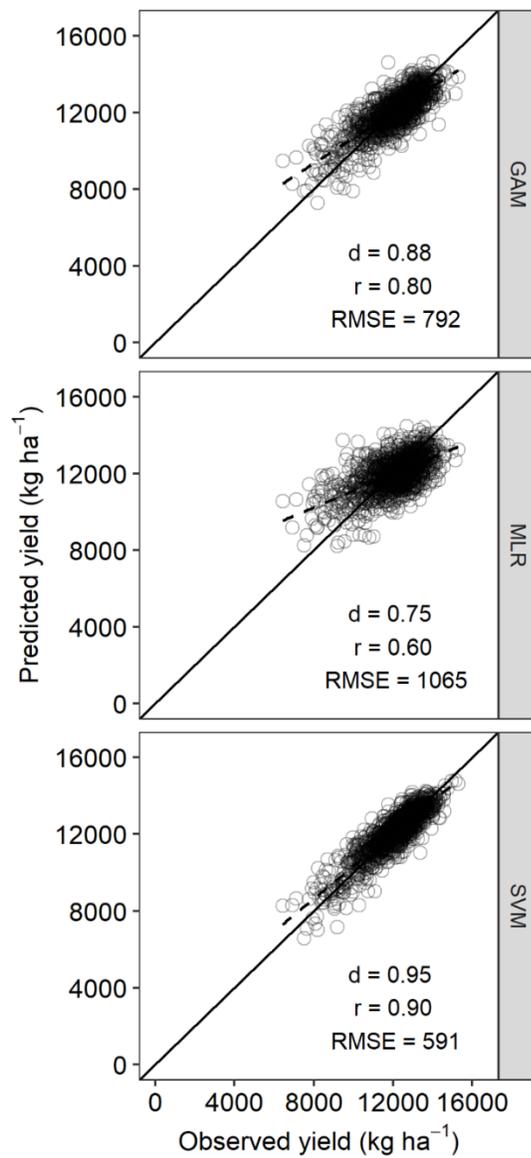
Figure 2.6. Scatterplots of observed versus predicted maize yield from the generalized additive model (GAM), multiple linear regression (MLR), and support vector machine (SVM) models. The diagonal black line is the 1:1 line. The dashed black line represents the linear regression between observed and predicted yields.
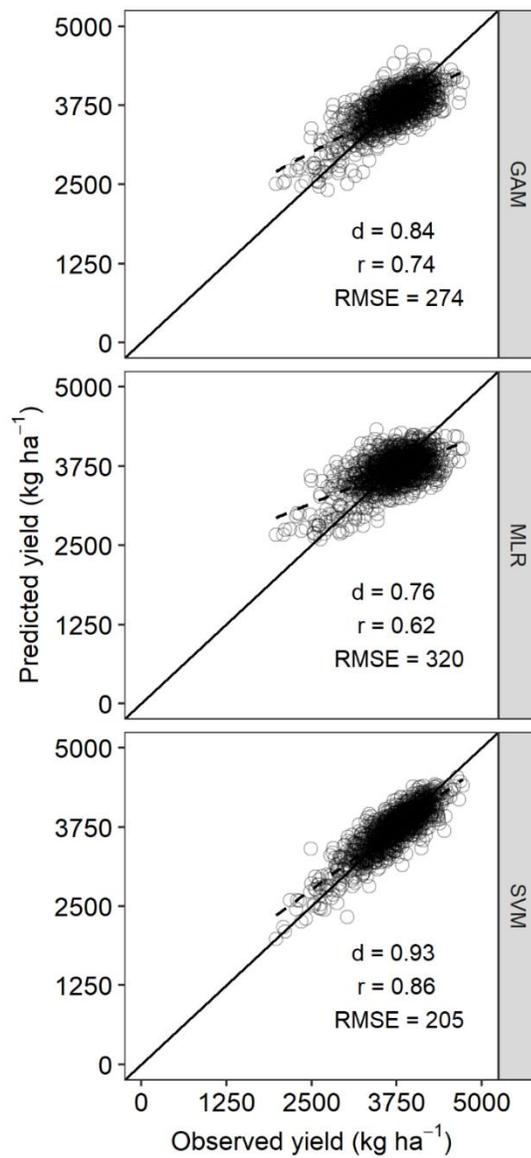
Figure 2.7. Scatterplots of observed versus predicted soybean yield from the generalized additive model (GAM), multiple linear regression (MLR), and support vector machine (SVM) models. The diagonal black line is the 1:1 line. The dashed black line represents the linear regression between observed and predicted yields.
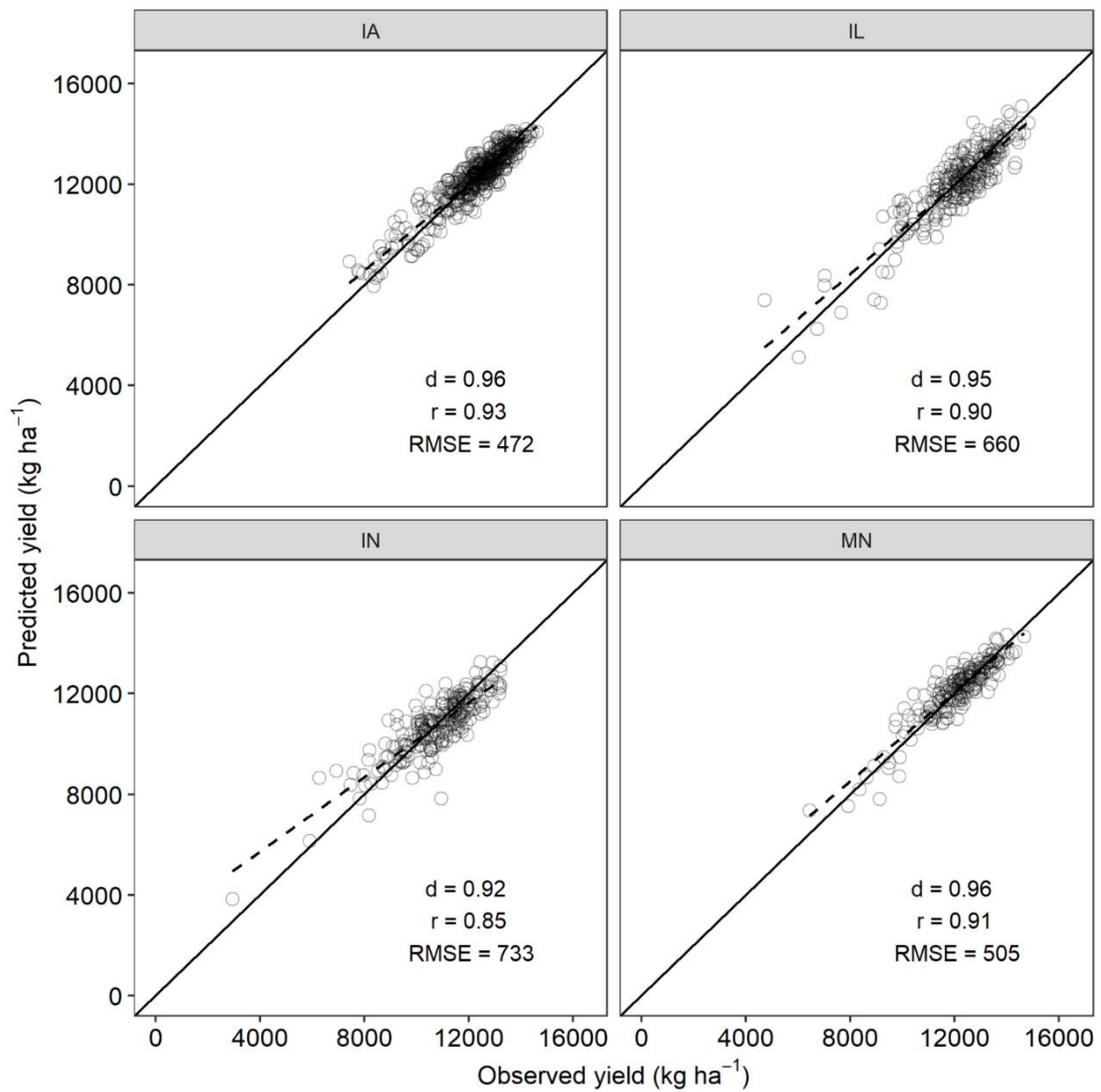
Figure 2.8. Scatterplots of observed versus predicted yield of maize using support vector machine model in Iowa (IA), Illinois (IL), Indiana (IN), and Minnesota (MN). The diagonal black line shows the 1:1 line. The dashed black line represents the linear regression between observed and predicted yields.
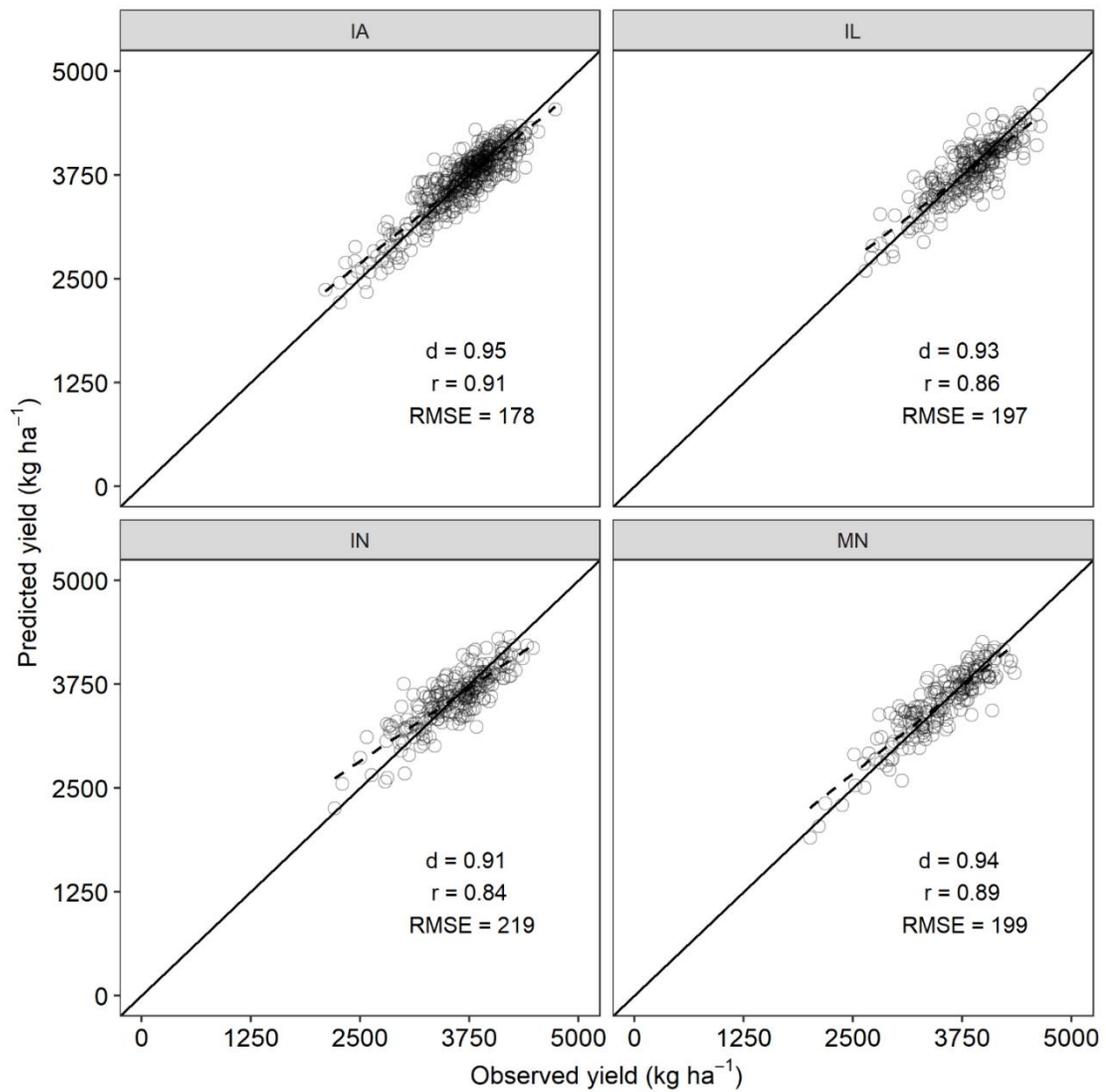
Figure 2.9. Scatterplots of observed versus predicted yield of soybean using support vector machine model in Iowa (IA), Illinois (IL), Indiana (IN), and Minnesota (MN). The diagonal black line shows the 1:1 line. The dashed black line represents the linear regression between observed and predicted yields.

CHAPTER 3. Simulation of maize growth, development, and yield at variable nitrogen regimes using the CERES-Maize model

Synopsis

Determination of maize response to nitrogen (N) remains challenging due to differences in requirement among genotypes and their interaction with weather conditions and soil texture. Field studies across multiple site-years are instrumental in understanding maize response to N, but this approach has economic constraints and time limitations. Process-based crop simulation models can help to overcome the shortcomings of short-term empirical research, but it requires careful calibration and evaluations. We assessed the accuracy of the process-based CERES-Maize model for conditions in the US upper Midwest. A study on four N fertilizer rates for maize production was conducted during the growing seasons of 2016 and 2017 in southwest (Lamberton) and southern (Waseca) Minnesota. The model accurately simulated the dates of anthesis and maturity at both locations with a normalized root mean square error (nRMSE) of 1%. At Lamberton, simulation accuracy showed a wide range of variability. Final grain yield in both years was simulated within 16% nRMSE, but aboveground biomass was simulated with nRMSE as high as 30% and aboveground N content and LAI were simulated with nRMSEs as high as 38%. At Waseca, however, aboveground biomass over the growing season and final grain yield in both years were simulated with a 15% nRMSE, and aboveground N and LAI at both years were simulated with 21% nRMSE. Overall, the accuracy of the model was better with optimal growing conditions compared to no N fertilization. This study showed that CERES-Maize is a promising tool for understanding maize response to N for conditions in the US upper

Midwest, but model applications for LAI and aboveground N content determination warrant further improvements.

3.1 Introduction

Nitrogen (N) is crucial for maize (*Zea mays* L.) growth, development, and yield (Sinclair and Horie, 1989). Managing soil N supply for optimum maize growth is a challenge due to N losses, variation in soil properties and weather conditions, and differences in N requirement among genotypes (Mamo et al., 2003; Scharf et al., 2005; Dhital and Raun, 2016). Since N is typically the most yield-limiting nutrient for maize production, there is a tendency for growers to over-apply N fertilizer relative to crop requirements for maximum grain yield (Sela et al., 2016). This has economic, health, and environmental consequences such as contamination of ground and surface waters and greenhouse gas emissions (Andraski et al., 2000; Shcherbak et al., 2014; Venterea et al., 2016). The upper Midwest is one of the most intensive agricultural areas in the United States, where maize is cultivated using a high level of N fertilizer. It has been estimated that more than 50% of the N and phosphorus load into the hypoxic zone of the northern Gulf of Mexico comes from maize and soybean [*Glycine max* L. (Merr.)] production (Alexander et al., 2008). As maize and soybean dominates the US Midwest cropping systems, watersheds in the region have the greatest nutrient loads as compared to other regions (NAWQA/USGS, 2014). As such, N fertilization is causing nitrate pollution in ground and surface waters, and threating water quality, human health, and aquatic ecosystems downstream (NAWQA/USGS, 2014; MEQB, 2015). Therefore, improvement in N fertilizer use efficiency for N loss reduction is extremely important.

Efficient use of N fertilizers in maize production relies on understanding of maize response to N. However, determination of optimum N fertilizer requirements remain for maize remains challenging due to the interaction effects of crop N response with weather conditions and soil texture (Tremblay et al., 2012). This is even more difficult to determine in rainfed systems, such as in the majority of the US Corn Belt, due to high variability in the frequency, intensity, duration, amount, and timing of rainfall and in daily air temperature. Several previous multi-location and multi-year studies have reported maize yield response to fertilizer N across a range of weather conditions and are useful for N management in maize cultivation (Tremblay et al., 2012; Cambouris et al., 2016; Alotaibi et al., 2018). Based on the premise of multi-location and multi-year field experiments, a data-driven recommendation method of maximum return to N has been implemented in much of the US Corn Belt for optimum N fertilizer rate determination (Nafziger et al., 2005). This is based on a quadratic-plateau response of maize grain yield to N fertilizer rate and the prices of N fertilizer and maize. This generalized method is adaptable with changing economic conditions; however, does not address within-field variability in soils and/or inter-and intra-seasonal weather variation affecting N use in maize (Melkonian et al., 2008). Although, field studies across multiple site-years are critical for understanding the maize response to N, field experiments to incorporate the effects of all possible weather conditions and soil types for every available maize genotype are not feasible, considering economic and time limitations. Decision support tools, such as process-based crop simulation models, offer opportunities to overcome the shortcomings of field experiments conducted over limited time periods and locations by simulating crop growth, development, and yield for various weather and soil conditions.

Crop simulation models are powerful tools to predict crop growth, development, and yield as a function of cultivar, soil, weather, and agronomic management. These models incorporate the chemical, bio-physical, and physiological processes within a cropping system (Jones et al., 2003; Keating et al., 2003). Therefore, crop models have been frequently used as decision support tools to understand crop response to stresses, predict yield, and evaluate fertilizer and irrigation management options. As maize is an important crop globally, several maize models exist worldwide and are in use for the prediction of maize growth and yield parameters. Some of the widely used maize models include CERES-Maize (Jones and Kiniry, 1986), APSIM-Maize (Keating et al., 2003), HYBRID-Maize (Yang et al., 2004), and STICS (Brisson et al., 1998). The CERES-Maize (Jones and Kiniry, 1986) model is one of over 42 crop models in the Decision Support System for Agrotechnology Transfer (DSSAT) software package (Hoogenboom et al., 2019). The CERES-Maize model has been evaluated and used around the world, including the US Midwest. For example, O'Neal et al. (2002) and Andresen et al. (2001) used the CERES-Maize model to study the effects of weather variability on maize yield in the US upper Midwest. More recently, Elliott et al. (2018) used the model to investigate the effects of drought on US historical maize yield. Frequent instances of the CERES-Maize application on N management studies are also present. For example, Ewing and Runck (2015), Thorp et al. (2007), Miao et al. (2006), and Pang et al. (1998) used the CERES-Maize model to assess N management practices in the US upper Midwest. Recently, Banger et al. (2018) and Jeong and Bhattarai, (2018) used the CERES-Maize model to simulate the effects of several fall- and spring-applied N management strategies on N loss and maize production in Illinois. In most of the previous attempts, either a generic genotype already presented in the model was used without calibration, (Andresen et al., 2001; Miao et al., 2006; Ewing and Runck, 2015;

Elliott et al., 2018) or calibrated only for yield (O'Neal et al., 2002) or yield and phenology (Banger et al., 2018). Studies on the CERES-Maize model that involve calibration for local growing conditions and evaluation for phenology, yield, and temporal change in crop parameters are limited for US upper Midwest. Moreover, only few studies (Pang et al., 1998; Banger et al., 2018) have investigated model performance to simulate crop parameters in response to N fertilizer rates in different soil types and weather conditions of the US upper Midwest. Assessment of in-season crop parameters such as temporal change in aboveground biomass, leaf area index (LAI), and aboveground shoot N content can be used to monitor within-season crop growth status and facilitate dynamic and location-specific crop management efforts. Therefore, the objective of this study was to calibrate and evaluate the CERES-Maize model to simulate maize phenology, yield, and temporal change in aboveground biomass, LAI, and aboveground shoot N content at variable N regimes under different growing conditions in the US upper Midwest.

3.2 Materials and methods

3.2.1 Study sites and years

Field experiments were conducted during the growing seasons of 2016 and 2017 at the University of Minnesota Southwest and Southern Research and Outreach Centers, located at Lamberton (44°14′19″ N and 95°18′50″ W) and Waseca (44°04′41″ N and 93°31′28″ W), MN, USA, respectively. These experiments were part of the Minnesota Long-Term Agricultural Research Network sites and that were in a maize-soybean rotation since 2013 and were managed with conventional tillage. Both study sites have an elevation of about 350 m ASL and have a hot-summer humid continental climate. The dominant soil types were Normania loam (Fine-loamy, mixed, superactive,

mesic Aquic Hapludoll) at Lamberton and Webster clay loam (Fine-loamy, mixed, superactive, mesic Typic Endoaquolls) at Waseca (Soil Survey Staff, 2018).

3.2.2 Experiment setup and data collection

Field experiments at both sites were configured in a randomized complete block design with four replications. Treatments include four N fertilizer rates of 0 (as control), 80, 100, and 120% of the University of Minnesota guideline rate for maize production (Kaiser et al., 2016) at each location. In both years at each site the 0, 80, 100, and 120% N rate treatments received a total of 0, 112, 140, and 168 kg N ha^{-1}, respectively. Both years, 5 d before planting, all plots except those of non-N fertilized control treatment received 73 kg N ha^{-1} in the form of urea [$CO(NH_2)_2$] and 17 kg S ha^{-1} as calcium sulfate dihydrate ($CaSO_4.2H_2O$) that was uniformly broadcast on the soil surface through Gandy spreader (Gandy Co., Owatonna, MN). At both sites, the remaining amount of N for each treatment was side-dressed in the form of urea-ammonium nitrate at the five-leaf collar maize phenological stage using a disc injector. Plots were 5.1 m × 6.1 m (8 maize rows wide) in size. At both sites, a maize hybrid 'DEKALB 53-56RIB' was planted in a row spacing of 0.76 m for a target population of 86,500 plants ha^{-1}. In 2016, planting was done on 25 April at both sites, whereas in 2017, planting was done on 29 and 24 April in Lamberton and Waseca, respectively. Field preparation was carried out using a conventional field cultivator. Soil profile data on soil texture were obtained from web soil survey produced by the United States Department of Agriculture Natural Resources Conservation Service (Soil Survey Staff, 2018). Soil textural class in the profile was mainly loam in Lamberton and clay loam in Waseca (Table 3.1). At both sites, soil samples were collected one week before planting in 2017 from each plot from the 0−30- and 30−60-cm depths. Each soil sample was composed of two subsamples collected

within each plot. Soil samples were air-dried, sieved to pass through a 2-mm mesh, and analyzed for ammonium-N, nitrate-N, organic carbon, cation exchange capacity, and pH (Table 3.2).

At both sites, plant data collection included phenology, LAI, aboveground biomass, and aboveground shoot N content every 7 to 14 d from one month after planting to two weeks after silking. Out of eight maize rows in a plot, two rows on both edges were considered border. The middle two rows were used for LAI readings and final grain yield. For LAI readings, AccuPAR model LP-80 (METER Group Inc., Pullman, WA, USA) was used, which estimates LAI non-destructively based on photosynthetic active radiation readings above and below the crop canopy. Three LAI readings were taken at the center of each plot and averaged. Destructive sampling for aboveground biomass and aboveground shoot N content was performed in the remaining 6th row. After emergence, plant stand counts from a 3-m length were taken from the rows used for aboveground biomass sampling. Stand counts in rows used for final grain harvest were taken at maize physiological maturity. At each aboveground biomass sampling, three plants from each plot were cut near soil surface and chopped into pieces to facilitate drying. All biomass samples were air-dried in a forced-air oven at 60°C until constant mass for dry aboveground biomass determination. The dried biomass samples from each plot were ground and analyzed for aboveground shoot N concentration. Final grain yield from the center two rows of each plot were determined with a plot combine and adjusted to 155 g kg^{-1} moisture.

Daily maximum and minimum air temperature (°C), solar radiation (MJ m^{-2}), and rainfall (mm) were obtained from the weather stations located within 200 m of the experiments.

3.2.3 CERES-Maize model calibration and evaluation

The cropping system model CERES-Maize (Jones and Kiniry, 1986), within DSSAT v. 4.7.5 (Jones et al., 2003; Hoogenboom et al., 2019), was used for this study. The DSSAT incorporates the dynamics of soil-plant-atmosphere interactions to simulate crop growth, development, and yield. Crop models in the DSSAT requires a minimum dataset of daily weather, crop management information, soil profile data, and cultivar information. Weather data includes daily maximum and minimum air temperature (°C), solar radiation (MJ m^{-2}), and precipitation (mm), which were obtained from the weather stations. The WeatherMan tool in the DSSAT was used to prepare weather file for the CERES-Maize model. Crop management information on planting date, planting rate, plant spacing, and fertilization strategy used in the study were entered through the XBuild tool in the DSSAT. Soil profile data on soil texture (Table 3.1) obtained from the Web Soil Survey (Soil Survey Staff, 2018). Total N (%), ammonium-N (ppm), nitrate-N (ppm), organic carbon (%), cation exchange capacity, and soil pH obtained from soil analyses (Table 3.2) were entered through the SBuilt tool in the DSSAT. The SBuilt was also used to estimate the missing data on bulk density, drained upper and lower limit, saturated water content, and saturated hydraulic conductivity based on the soil texture through pedo-transfer functions.

Calibration of CERES-Maize requires the estimation of six genotype-specific coefficients: thermal time from seedling emergence to end of juvenile phase (P1), extent of delay in development for each hour with daylength above 12 hours (P2), thermal time from silking to physiological maturity (P5), phyllochron interval between successive leaf tip appearances (PHINT), maximum possible number of kernels per plant (G2), and kernel growth rate during grain-filling stage under optimum conditions (G3). The first four

coefficients (P1, P2, P5, and PHINT) regulate maize development, whereas the last two coefficients (G1 and G2) regulate maize grain yield.

For model calibration, cultivar information on days to anthesis and physiological maturity, LAI at anthesis, aboveground biomass at anthesis and physiological maturity, and final grain yield from the 120% N treatment at Waseca during the 2017 growing season was used. This treatment gave the highest yield with the least variability, and therefore assumed to represent stress-free conditions. Generalized Likelihood Uncertainty Estimation (GLUE), a Bayesian parameter estimation procedure within DSSAT, was used for calibration. At the first step, development-related parameters were calibrated using anthesis and physiological maturity dates. Subsequently, growth and yield-related parameters were calibrated using measured aboveground biomass, LAI, and yield. At each step, GLUE was set to run for a total of 10,000 iterations to estimate the genotype coefficients. After model calibration, the model was evaluated against the data collected from the remaining treatments of 2017 and all treatments of 2016 in Waseca, and all treatments of 2016 and 2017 in Lamberton. The model was evaluated for its ability to simulate phenology, temporal change in aboveground biomass, LAI, and aboveground shoot N content, and final grain yield at different N levels in different soils and weather conditions. Root mean square error (RMSE) and normalized RMSE (nRMSE) were used as the evaluation metrics. The RMSE between simulated and measured values was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (S_i - M_i)^2}{n}} \tag{1}$$

where *n* refers to number of observations, *S* and *M* denote simulated and measured values. Then, nRMSE was calculated as RMSE / (mean of measured values). Lower RMSE and nRMSE values indicate better model performance.

3.3 Results and discussion

3.3.1 Weather conditions during the growing season

The historical (1981-2010) average cumulative rainfall during the growing season (May-October) was 492 mm for Lamberton (SWROC/UMN, 2018) and 613 mm for Waseca (SROC/UMN, 2018). Total cumulative rainfall during the growing season at both locations and in both years were above the historical averages. During the 2016 and 2017 growing seasons, the cumulative rainfall at Lamberton was 723 and 651 mm and at Waseca was 1195 and 657 mm, respectively. At both locations more rainfall occurred during July, August, and September of 2016 than the historical averages (Fig. 3.1). At Waseca, total monthly rainfall in 2016 was more than two times greater than the historical average in July and August, and more than four times greater than the historical average in September (Fig 3.1).

The historical average monthly air temperatures for both locations during the growing season were nearly identical (SROC/UMN, 2018; SWROC/UMN, 2018). As compared to the historical average air temperature of the growing season, 2016 was warmer at both locations. Waseca was warmer by 1.4°C, whereas Lamberton was warmer by 0.7°C. In 2016, monthly average air temperature was higher in all months in Waseca and in all months except May and July in Lamberton (Fig 3.1). In 2017, monthly average air temperatures was higher in all months except May and August at both locations.

3.3.2 Model calibration

The calibration procedure gave cultivar coefficients values, which had the maximum probability to minimize the difference between measured and simulated data (Table 3.3). The calibrated coefficients accurately simulated anthesis and physiological maturity dates with 0.4% nRMSE (Table 3.4). During calibration, aboveground biomass was simulated with a nRMSE of 7% at anthesis and 9% at physiological maturity. The maximum LAI was simulated with a nRMSE of 12% and final grain yield was simulated with a nRMSE of 15% (Table 3.4). After satisfactory simulation results in calibration, the calibrated model coefficients were evaluated to simulate phenology, aboveground biomass, LAI, aboveground shoot N content, and yield at different N rates in the growing conditions of Waseca and Lamberton.

3.3.3 Model evaluation

Upon evaluation against observed data across all N treatments in Waseca in 2016, anthesis and maturity date were simulated accurately within 1% nRMSE (Table 3.5). Across all N treatments, aboveground biomass over the 2016 growing season in Waseca was simulated with a nRMSE of about 15%. Aboveground shoot N and LAI over the 2016 growing season were simulated with a nRMSE of about 20%. In rest of the N treatments in Waseca in 2017 (except 120% N which was used for calibration), anthesis and maturity dates were accurately simulated. Aboveground biomass over the 2017 growing season was simulated with a nRMSE of 15% and aboveground shoot N content was simulated with a nRMSE of 18%. Leaf area index over the 2017 growing season across all N treatments was simulated with an nRMSE of 21% (Table 3.5).

Upon evaluation against observed data across all N treatments in Lamberton in 2016, anthesis date and maturity dates were accurately simulated within one day of the observed result (Table 3.5). While LAI over the growing season across all N treatments was simulated with a nRMSE less than 19%, aboveground biomass and aboveground shoot N content were simulated with higher nRMSEs of 30 and 35%, respectively. In 2017 as well, anthesis and maturity dates in all treatments were simulated accurately within one day of the observed results. Aboveground biomass over the growing season in 2017 across all N treatments was simulated with an overall nRMSE of 25% whereas aboveground shoot N content was simulated with a nRMSE of 14%. A rather large difference in simulation accuracy occurred in LAI with a nRMSE of 38% (Table 3.5).

Overall at both locations and in both growing seasons, phenology was simulated accurately within nRMSE of 1%. The model successfully captured the effects of cooler-than-average air temperatures in August of 2017 at both locations, which delayed physiological maturity (Fig. 3.1). Accurate simulations of phenological events using CERES-Maize have been reported in several previous studies (Persson et al., 2009; Yang et al., 2009; Anothai et al., 2013), confirming the results of this study. Although aboveground biomass was simulated with a nRMSE of around 15% in Waseca in both years, in Lamberton was simulated with higher nRMSEs of 30% and 25% in 2016 and 2017, respectively (Table 3.5). As with aboveground biomass, simulation of aboveground shoot N content also showed larger variation. In Waseca, aboveground shoot N content was simulated with a nRMSE of about 20% in both years, whereas in Lamberton, nRMSE for aboveground shoot N content was 35% in 2016 and 14% in 2017. Simulations of LAI also showed high variability in accuracy. In Waseca, LAI was simulated with a nRMSE of about 20% in both years, whereas in Lamberton the nRMSE

44

ranged from 18% in 2016 to 38% in 2017 (Table 3.5). Overall, simulation results in Waseca in both years were more consistent and accurate as compared to results in Lamberton.

3.3.3.1 Aboveground biomass

Upon evaluation against measured data from individual treatment, aboveground biomass in Waseca over the 2016 growing season was simulated with a nRMSE less than 12% for all treatments except the non-N-fertilized control, which had a nRMSE of 30% (Fig. 3.2). Aboveground biomass simulation in control treatment in 2017 also had a higher nRMSE of 55%. Aboveground biomass over the growing season with the 80 and 100% N treatments was simulated with nRMSEs less than 14% (Fig. 3.3).

As in Waseca, simulated aboveground biomass over the growing season in the control treatment in Lamberton also had higher nRMSEs as compared to the other treatments. The nRMSE for simulated aboveground biomass in the control treatment in Lamberton was about 60% for both years (Fig. 3.4, 3.5). However, simulation results from higher N rates had greater accuracies. In 2016, aboveground biomass over the growing season in Lamberton was simulated with nRMSEs of 33, 24, and 13% in 80, 100, and 120% N treatments, respectively (Fig. 3.4). In 2017, the corresponding nRMSEs were below 21%. The model usually overestimated aboveground biomass and especially for the control treatment (Fig. 3.5).

3.3.3.2 Leaf area index

The difference between simulated and measured LAI was highly variable. In Waseca in 2016 the model underestimated the maximum LAI and simulated LAI with a

nRMSE of 31, 12, 15 and 21% in the control, 80, 100, and 120% N treatments, respectively (Fig. 3.6). In 2017, the model overestimated LAI at early vegetative stages for all N treatments and simulated LAI with a nRMSE of 39, 27 and 13% in the control, 80, and 100% N treatments, respectively (Fig. 3.7).

As in Waseca, simulated LAI over the growing season in the control treatment in Lamberton also had higher nRMSE as compared to the other treatments. The nRMSE for simulated LAI in the control was 38% in 2016 and was 44% in 2017 (Fig. 3.8, 3.9). The model simulated LAI more accurately with nRMSEs of 11 and 9% in the 100 and 120% N treatments in 2016 (Fig. 3.8). However, in 2017 the discrepancy between measured and simulated LAI was much higher. For the 80, 100, and 120% N treatments, the nRMSEs for simulated LAI ranged from 33 to 37% (Fig. 3.9). The model frequently overestimated LAI at early vegetative stages and underestimated the maximum LAI for all N treatments in 2017 (Fig. 3.9).

Despite simulating a similar number of total leaves at maturity as compared to actual number of leaves (measured, 18 vs simulated, 18-20), growth simulation in the model was rapid and the model added new leaves quickly at earlier growth stages. This rapid addition of new leaves in the model simulation not only caused overestimation of aboveground biomass and LAI at earlier growth stages but also caused premature peak of LAI (Fig 3.6-3.9). This was more apparent in the simulation of N-stressed treatments. This is parallel with the observation of Liu et al. (2014), who reported that CERES-Maize failed to adjust for a slower growth rate in nutrient-limited conditions. This result is also in agreement with previous finding of Basso et al. (2016), who reviewed the performance of CERES models and found that simulation errors of CERES models are usually higher under marginal growing conditions, including nutrient deficit situations.

Overall, simulation results in Waseca in both years were more consistent and closer to measured results than in Lamberton. High variability of simulated results in Lamberton was partly due to lower simulated soil water content, which induced water stress from 70 days after planting in both years with greater extent in 2017 season. This water stress reduced leaf expansion, photosynthesis, and simulated early senescence that caused underestimation of aboveground biomass and LAI after the late vegetative stage, particularly in treatments with higher N rates.

3.3.3.3 Aboveground shoot N content

Simulation of aboveground shoot N content also showed a wide range of variability. In Waseca in 2016, contrary to other crop variables, the model simulated aboveground shoot N content accurately with a nRMSE of only 7% in the control treatment (Fig. 3.10). However, the model usually overestimated the aboveground shoot N contents in other N treatments and simulated the aboveground shoot N content with a nRMSE of 26, 24 and 18% in the 80, 100, and 120% N treatments, respectively (Fig. 3.10). In 2017, the model simulated aboveground shoot N content with a nRMSE of 15% in the control treatment, which was still lower as compared to 17 and 20% nRMSE for the 80 and 100% N treatments, respectively (Fig. 3.11).

As in Waseca, the model frequently overestimated the aboveground shoot N content in Lamberton in 2016. The discrepancy between measured and simulated aboveground shoot N content was even higher as compared to Waseca, with nRMSEs ranging from 39 to 27% across all N treatments (Fig. 3.12). In 2017, the differences between measured and simulated aboveground shoot N content was relatively lower, with nRMSEs of only 6% for control and about 15% for the other treatments (Fig. 3.13).

47

Overall, aboveground shoot N simulation was frequently overestimated in 2016 at both locations, whereas underestimated in 2017 at both locations. This difference in model estimates might be attributed to the difference in sampling time. At both locations, samples for aboveground shoot N content were mostly from later growth stages in 2016 (2 samples before and 3 samples after 75 days after planting) as compared to earlier growth stages in 2017 (3 samples before and 1 sample after 75 days after planting). Thus, these results show that the model tends to underestimate aboveground shoot N content at earlier vegetative stages, whereas overestimate it at later vegetative to early reproductive stages.

Higher inconsistencies in shoot N simulation as compared to measured results have also been reported in past studies (Liu et al., 2012; Basso et al., 2016; Fang et al., 2017). Simulation of aboveground shoot N content in CERES-Maize (Jones and Kiniry, 1986) depends on crop N demand, which is calculated based on critical N concentration. The CERES-Maize uses a growth stage based approach (Jones and Kiniry, 1986) to calculate critical N concentration, which depends on two coefficients: maximum level for critical tissue N concentration, CTCNP1 (1.52 as default value) and a coefficient with which N concentration changes with growth stage, CTCNP2 (0.16 as default value). Some studies have reported improved simulations of aboveground shoot N content and N uptake by optimizing these coefficients for local growing conditions (Liu et al., 2012; Fang et al., 2017). An alternative approach to calculate critical N concentration is based on the aboveground biomass approach as described by Plénet and Lemaire (2000). This alternative approach (Plénet and Lemaire, 2000) of critical N determination in maize depends on the N dilution curve and relies on the allometric relationship between aboveground biomass and shoot N content. Recent studies have reported higher

accuracies in aboveground shoot N concentration and N uptake simulations in maize by altering the algorithm to calculate critical N concentration from the growth stage based approach to aboveground biomass based approach (Fang et al., 2017; Yakoub et al., 2017; Ratjen et al., 2018).

3.3.3.4 Grain yield

Overall, final grain yield at both locations was simulated satisfactorily, with an overall nRMSE of less than 16% in both years (Fig. 3.14). In Waseca, yield was slightly overestimated with nearly equal nRMSEs of 15% in both years. In contrast, yield in Lamberton was slightly underestimated with an nRMSE of 16% in 2016 and 13% in 2017. This underestimation in yield can be partly attributed to the water stress simulated in the model. The range of nRMSE in yield simulations in this study has been frequently reported in scientific publications and is considered as a satisfactory agreement between simulated and measured yields. For example, Jing et al. (2017) used the CERES-Maize model to simulate maize yield in eastern Canada and reported an overall nRMSE of 20%. In an another study using the CERES-Maize, Li et al. (2015) in northwestern China obtained nRMSEs ranging from 13 to 19%. A study by Persson et al. (2009) reported nRMSEs from 1 to 28% while evaluating the CERES-Maize in the southeastern United States.

3.4 Conclusions

This study assessed the performance of CERES-Maize model to simulate maize growth, development, grain yield, and temporal change in leaf area index, and aboveground shoot N content at different N regimes during 2016 and 2017 at two locations in southern Minnesota. The model accurately simulated anthesis and maturity

date at both locations, within an nRMSE of 1%. In Waseca, aboveground biomass over the growing season and final grain yield at both years were simulated within 15% nRMSE. Aboveground shoot N and LAI over the growing season at both years were simulated within 21% nRMSE. In Lamberton, however, simulation accuracy showed a wide range of variability due to simulated water stress. Although, final grain yield in both years at Lamberton was simulated within 16% nRMSE, aboveground biomass over the growing season was simulated with higher nRMSEs of up to 30%. Leaf area index and aboveground shoot N content over the growing season were simulated with even higher nRMSEs of up to 38%. Overall, CERES-Maize model estimates were more reliable for phenology, aboveground biomass, and yield whereas less reliable for LAI and aboveground shoot N content. Within each crop parameter, the model tended to give more accurate estimates in simulations with optimal growing conditions without N stress. This study shows CERES-Maize as a promising tool for understanding maize response to N, especially while determining maize phenology, aboveground biomass, and yield in the growing conditions of southern Minnesota. However, model applications for LAI and aboveground shoot N content determination warrant further model calibration, evaluation and improvements.

Table 3.1. Soil texture and soil textural classes used in the study.

Horizon depth (cm)	Sand (%)	Silt (%)	Clay (%)	Soil textural class
Webster clay loam soil, Waseca, MN				
0-25	25	42	33	Clay loam
25-50	25	42	33	Clay loam
50-107	35	37	28	Clay loam
107-200	44	35	21	Loam
Normania loam soil, Lamberton, MN				
0-23	35	39	26	Loam
23-43	38	37	25	Loam
43-71	40	36	24	Loam
71-99	41	36	23	Loam
99-200	38	37	25	Loam

Table 3.2. Soil pH, organic matter content, cation exchange capacity, nitrate-nitrogen, and ammonium-nitrogen in the 0-30 and 30-60 cm depths before planting in 0, 80, 100, and 120% nitrogen treatment plots at Waseca and Lamberton, MN in 2017.

Treatment	Soil pH	Organic matter (%)	Cation exchange capacity	Nitrate-nitrogen (ppm)	Ammonium-nitrogen (ppm)
			Waseca (0-30 cm)		
0	6.6 (0.6)	4.5 (0.8)	28 (4.5)	3.7 (0.8)	16.3 (14.2)
80	6.4 (0.4)	4.9 (0.6)	28 (3.2)	4.3 (2.2)	10.8 (6.9)
100	6.6 (0.3)	4.5 (0.6)	26.5 (2.9)	3.3 (0.5)	4.3 (1.9)
120	6.7 (0.3)	4.5 (0.9)	26.0 (4.6)	3.0 (0.8)	7.7 (9.3)
			Waseca (30-60 cm)		
0	6.9 (0.5)	2.6 (0.5)	26.4 (1.4)	3.5 (1.0)	11.3 (8.9)
80	7.2 (0.5)	1.9 (0.3)	23.0 (1.8)	2.8 (1.8)	7.1 (6.3)
100	7.1 (0.6)	2.2 (0.4)	24.5 (1.7)	3.1 (0.7)	9.1 (6.1)
120	7.0 (0.2)	2.3 (0.3)	23.9 (1.5)	2.8 (1.0)	8.9 (12.0)
			Lamberton (0-30 cm)		
0	5.8 (0.3)	4.0 (0.7)	19.0 (2.2)	4.0 (0.3)	5.9 (0.6)
80	5.6 (0.2)	4.0 (0.7)	19.8 (1.3)	4.1 (0.9)	5.7 (1.1)
100	5.7 (0.4)	3.9 (0.4)	19.8 (1.1)	4.2 (0.5)	6.2 (1.0)
120	5.9 (0.2)	3.9 (0.7)	19.0 (0.9)	4.6 (1.3)	6.3 (1.6)
			Lamberton (30-60 cm)		
0	7.2 (0.7)	2.1 (0.3)	21.1 (3.5)	3.1 (0.6)	3.9 (0.6)
80	7.2 (0.8)	2.3 (0.6)	23.4 (5.6)	3.6 (0.7)	4.1 (1.3)
100	7.2 (0.5)	2.2 (0.5)	20.7 (3.9)	3.7 (1.4)	3.5 (0.5)
120	7.2 (0.8)	2.3 (0.5)	22.1 (3.9)	3.9 (0.8)	4.1 (0.7)

Table 3.3. Calibrated values of cultivar coefficients for CERES-Maize obtained using generalized likelihood uncertainty estimation procedure in DSSAT v. 4.7.5.

Maize cultivar coefficient	Calibrated values
Thermal time from seedling emergence to the end of the juvenile phase (P1)	216.8
Extent to which development is delayed for each hour that daylength is above 12.5 hours (P2)	0.13
Thermal time for silking to physiological maturity (P5)	763.3
Phyllochron interval between successive leaf tip appearances (PHINT)	44
Maximum possible number of kernels per plant (G2)	700.9
Kernel growth rate during linear grain filling stage under optimum conditions (G3)	10.46

Table 3.4. Measured and simulated values of different crop variables with their respective root mean square error (RMSE) and normalized RMSE during model calibration.

Variable	Measured	Simulated	RMSE	nRMSE (%)
Phenology (Julian day)				
Anthesis	201 (0)	201 (0)	0	0
Maturity	266 (0)	267 (0)	1	0.4
Aboveground biomass (kg ha^{-1})				
Anthesis	10407 (634)	10845 (1.2)	678	6.5
Maturity	24305 (1674)	26173 (243)	2124	8.7
LAI (m^2 m^{-2})	5.5 (0.46)	4.96 (0)	0.66	11.9
Yield (kg ha^{-1})	12968 (328)	14858 (439)	1905	14.7

Table 3.5. Root mean square error (RMSE) and normalized RMSE (%, in parenthesis) between measured and simulated data averaged over the growing season across all nitrogen treatments at Waseca and Lamberton, MN in 2016 and 2017.

Study site/Year	Phenology (day)		Aboveground biomass (kg ha^{-1})	Aboveground shoot N (%)	LAI (m^2 m^{-2})
	Anthesis	Maturity			
Waseca					
2016	0	3 (1.2)	900 (14.8)	0.4 (20.4)	0.6 (19.8)
2017[‡]	0	0	1172 (15.3)	0.5 (18.2)	0.7 (20.6)
Lamberton					
2016	1 (1)	0	1769 (30.1)	0.6 (34.6)	0.4 (18.1)
2017	1(1)	1(1)	1065 (25.1)	0.4 (13.6)	0.9 (37.9)

[‡]Only includes 0, 80, and 100% N treatments
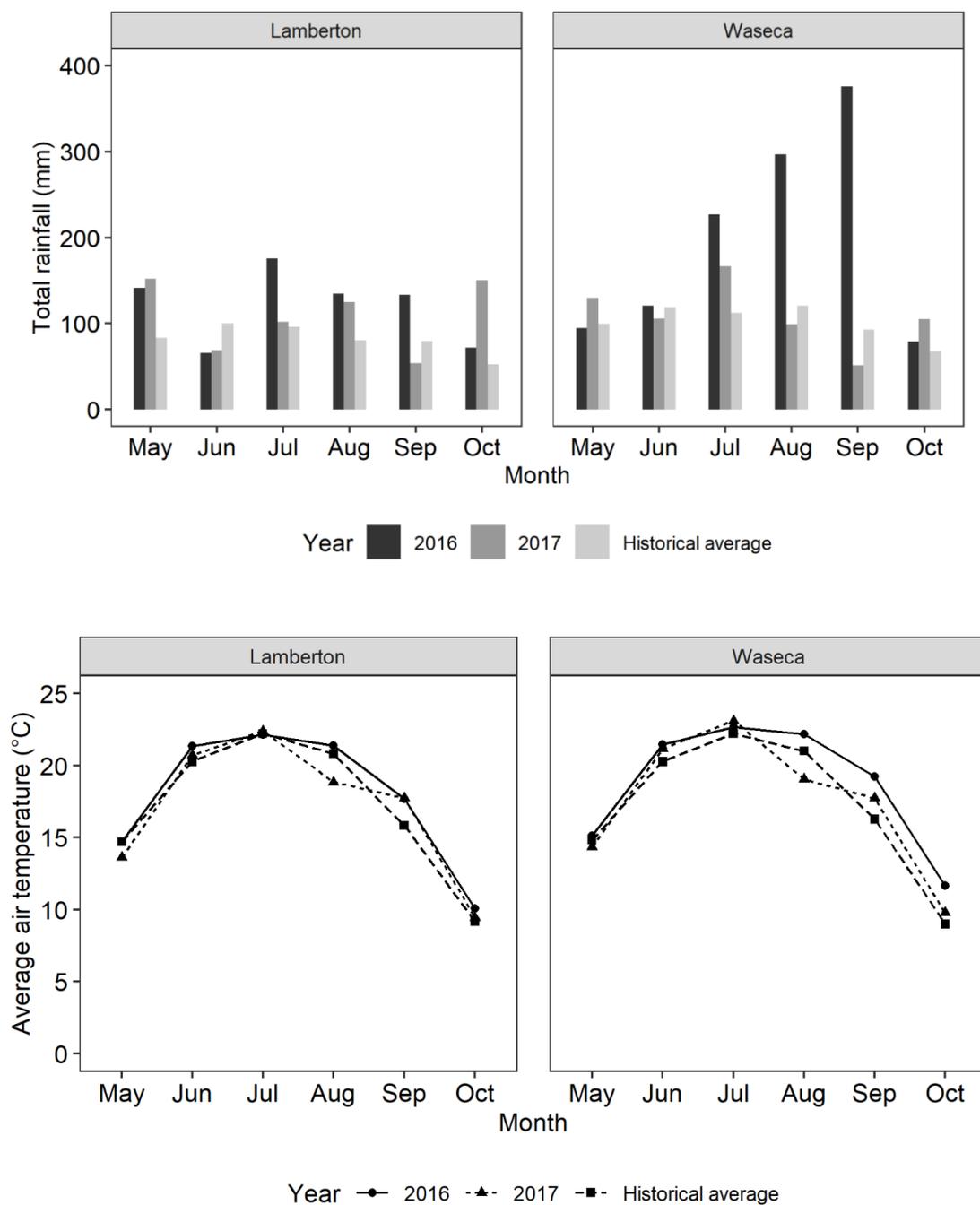
Figure 3.1. Comparisons of monthly total rainfall (top) and average air temperature (bottom) with historical averages (1981-2010) in Lamberton and Waseca, MN during the 2016 and 2017 growing seasons.
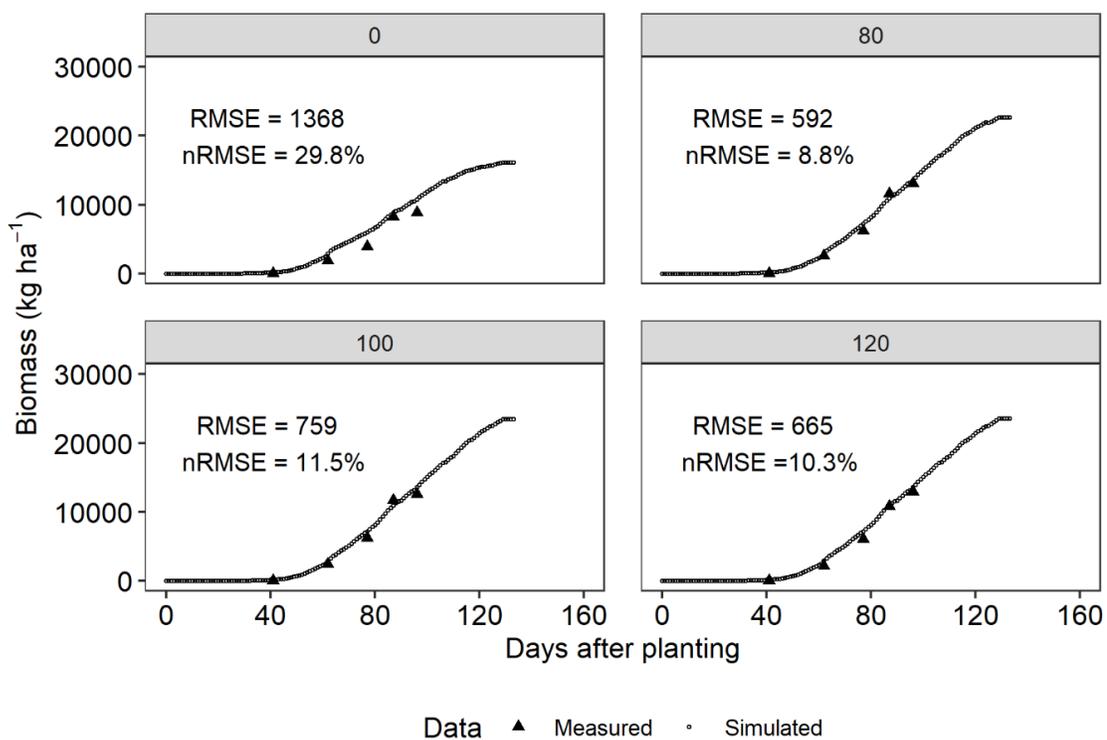
Figure 3.2. Measured and simulated aboveground biomass at different days after planting at 0, 80, 100, and 120% N rates in Waseca, MN during 2016.
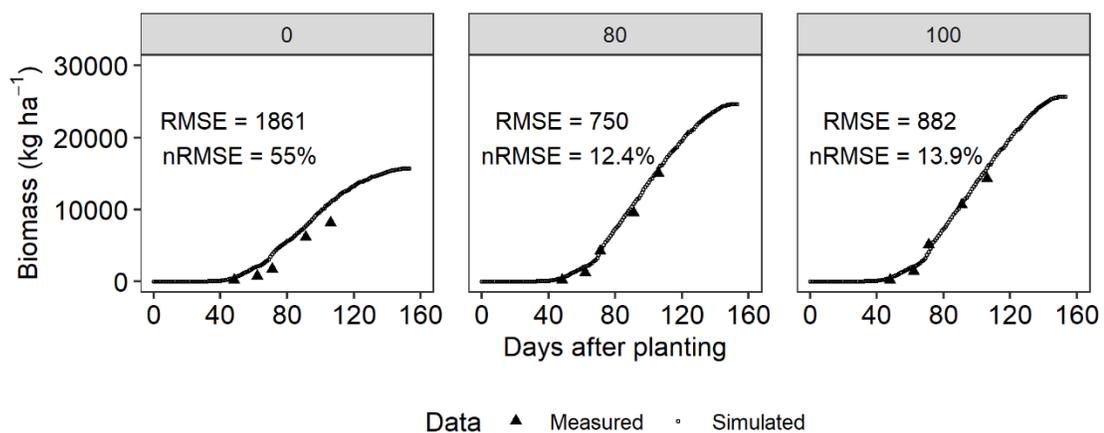
Figure 3.3. Measured and simulated aboveground biomass at different days after planting at 0, 80, and 100% N rates in Waseca, MN during 2017.
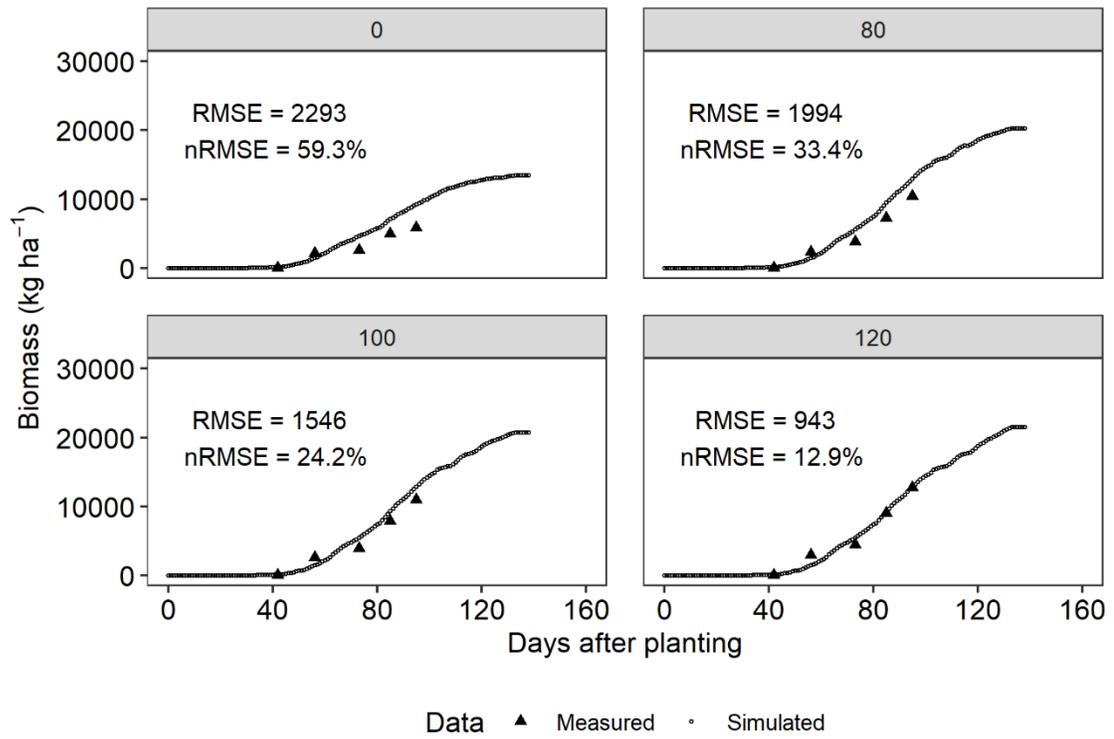
Figure 3.4. Measured and simulated aboveground biomass at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2016.
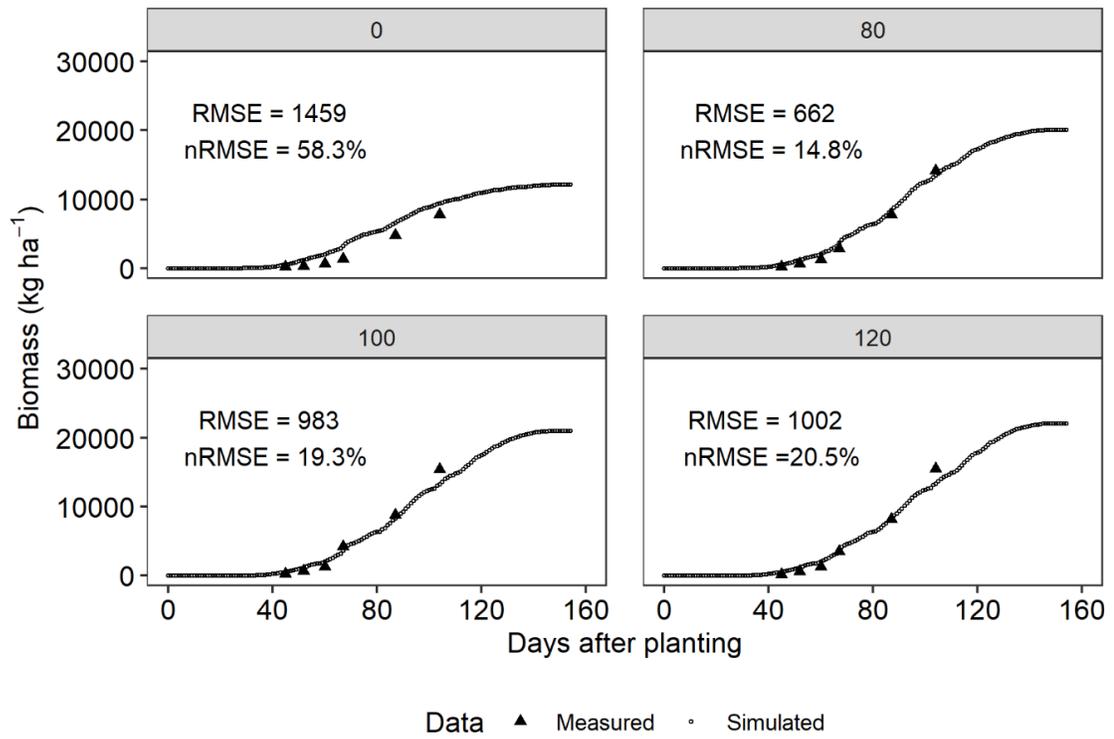
Figure 3.5. Measured and simulated aboveground biomass at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2017.
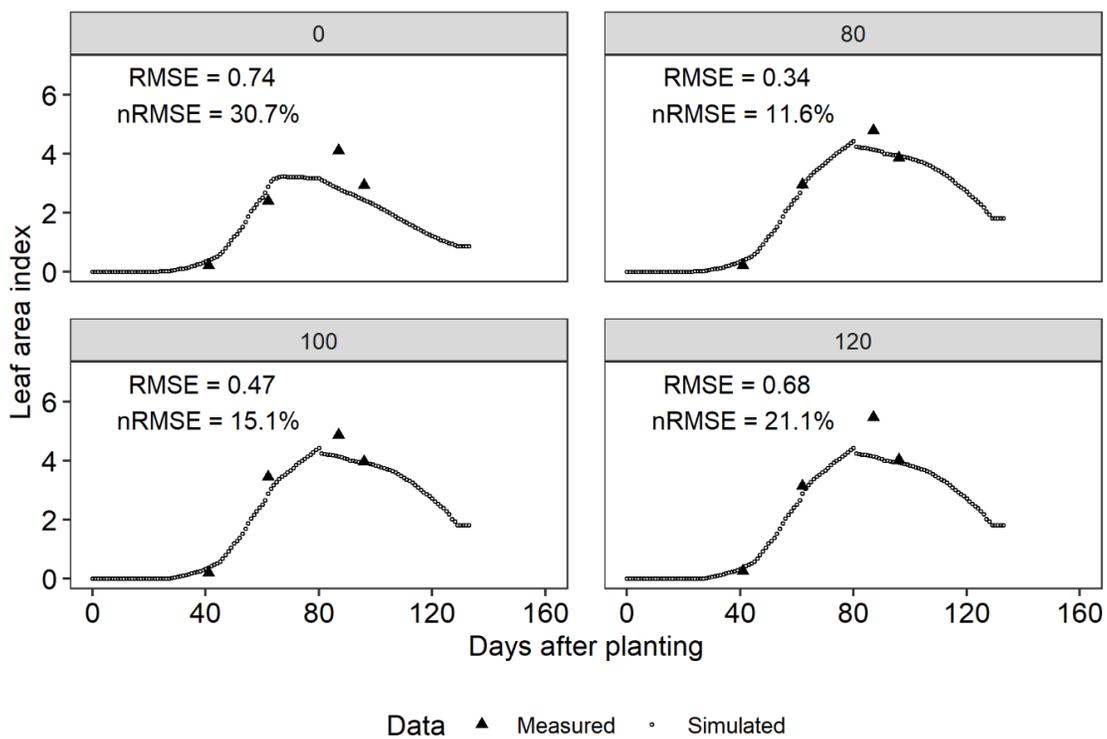
Figure 3.6. Measured and simulated leaf area index at different days after planting at 0, 80, 100, and 120% N rates in Waseca, MN during 2016.
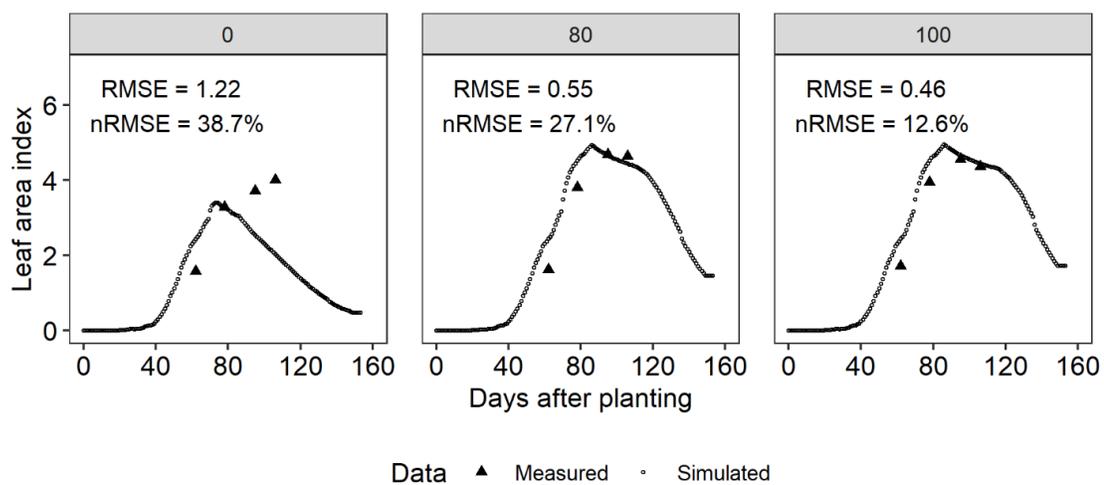
Figure 3.7. Measured and simulated leaf area index at different days after planting at 0, 80, and 100% N rates in Waseca, MN during 2017.
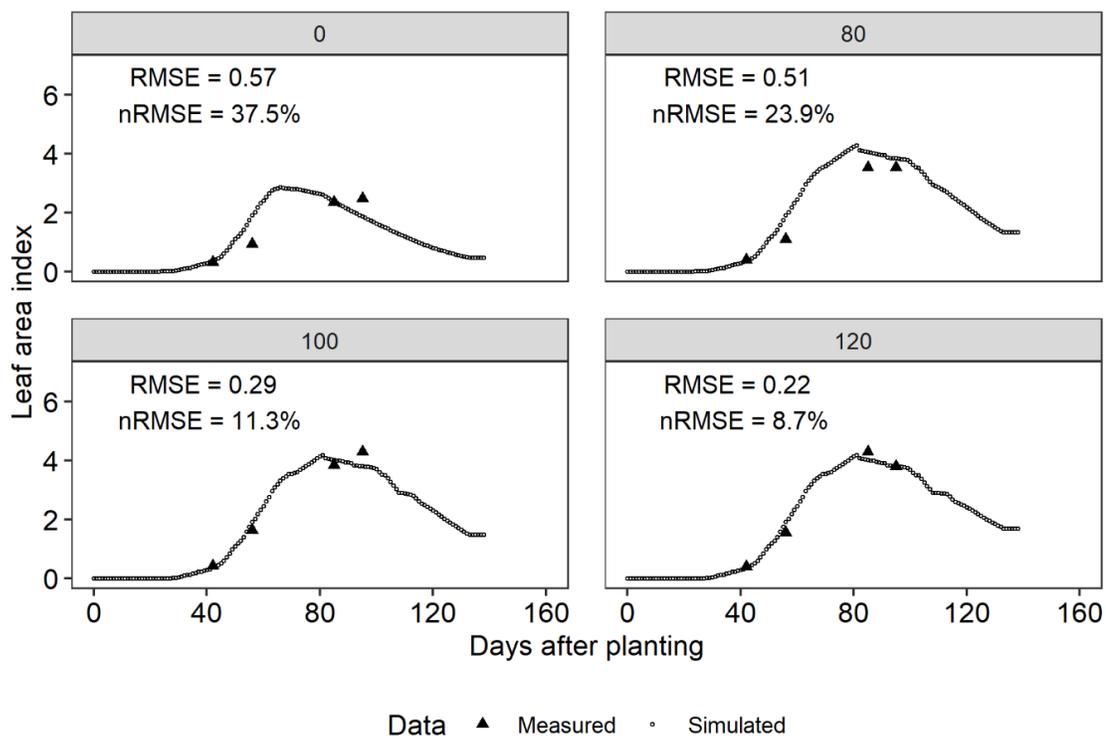
Figure 3.8. Measured and simulated leaf area index at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2016.
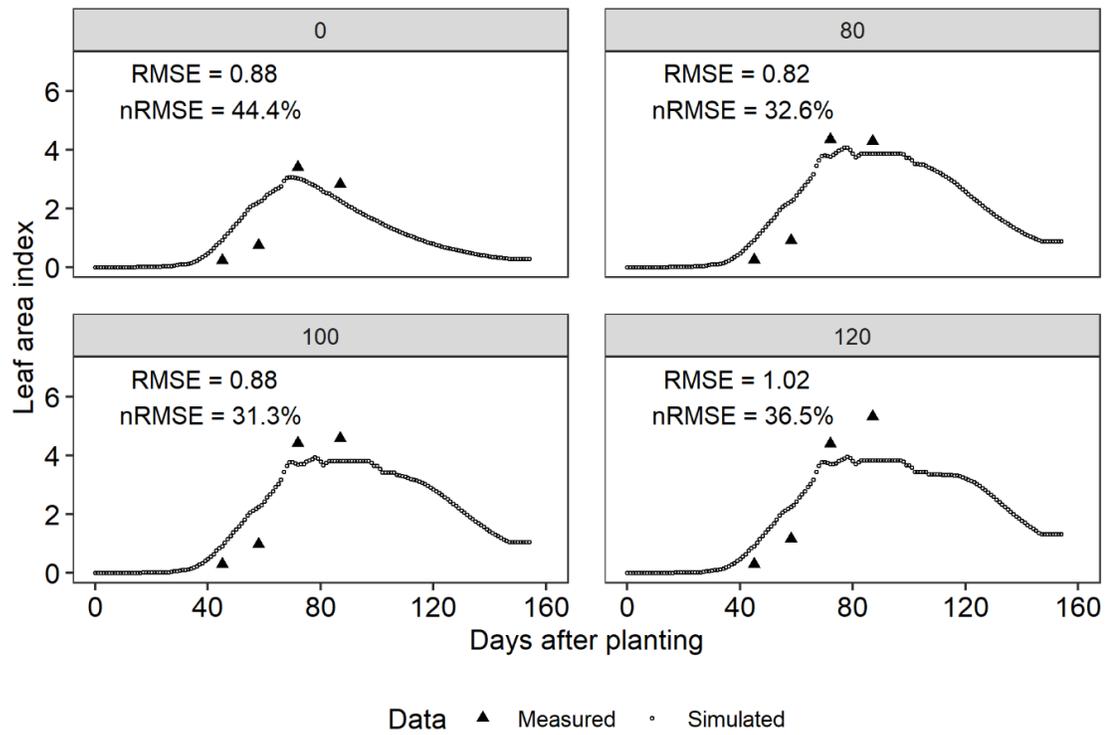
Figure 3.9. Measured and simulated leaf area index at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2017.
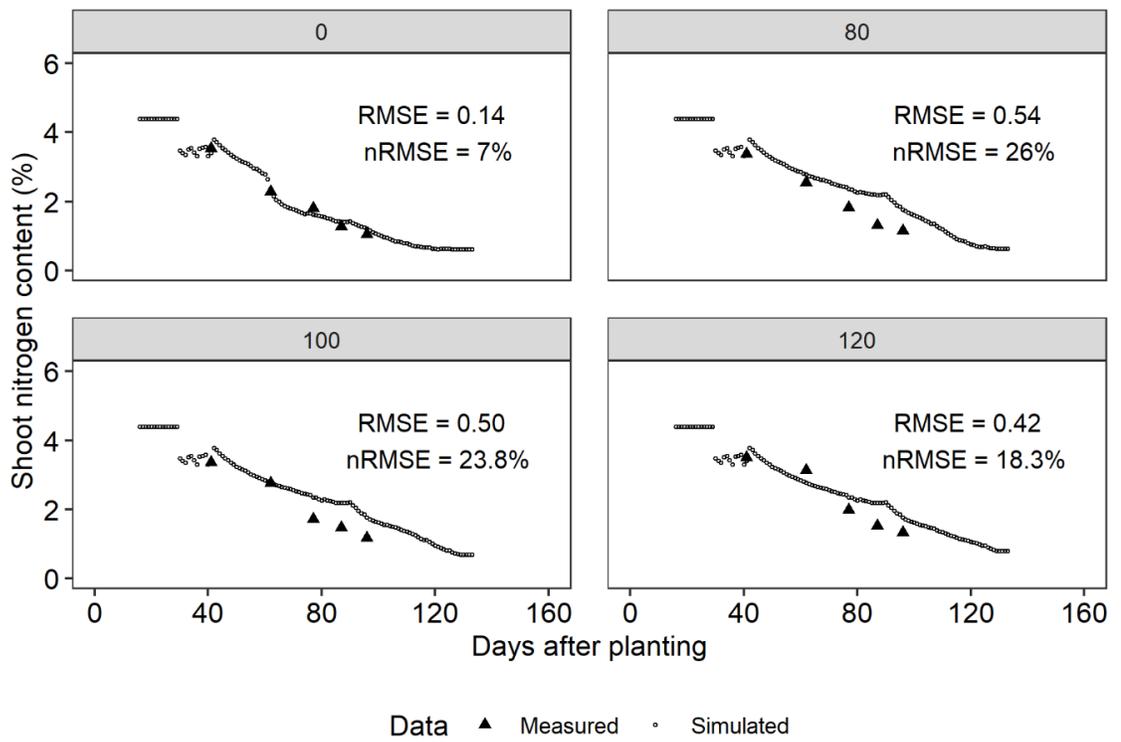
Figure 3.10. Measured and simulated aboveground shoot nitrogen content (%) at different days after planting at 0, 80, 100, and 120% N rates in Waseca, MN during 2016.
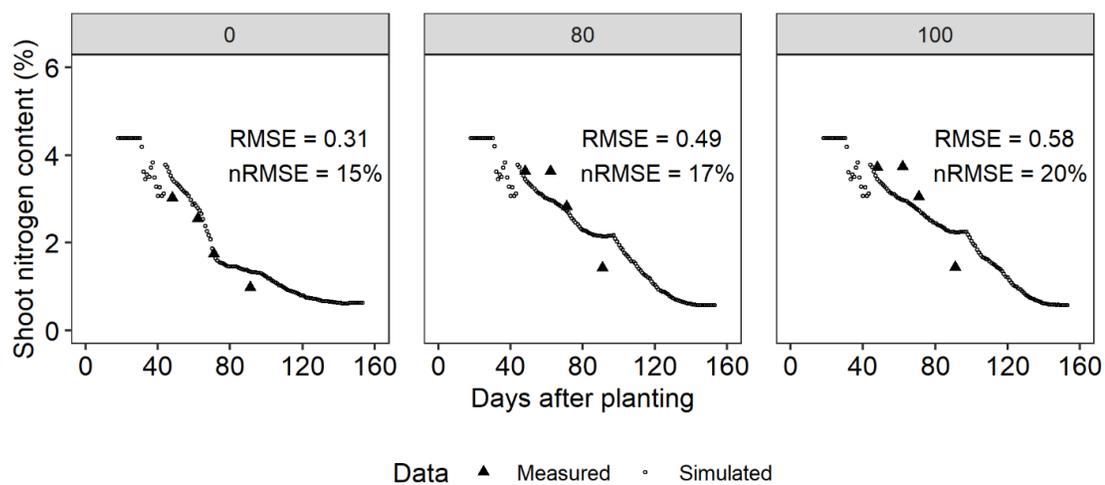
Figure 3.11. Measured and simulated aboveground shoot nitrogen content (%) at different days after planting at 0, 80, and 100% N rates in Waseca, MN during 2017.
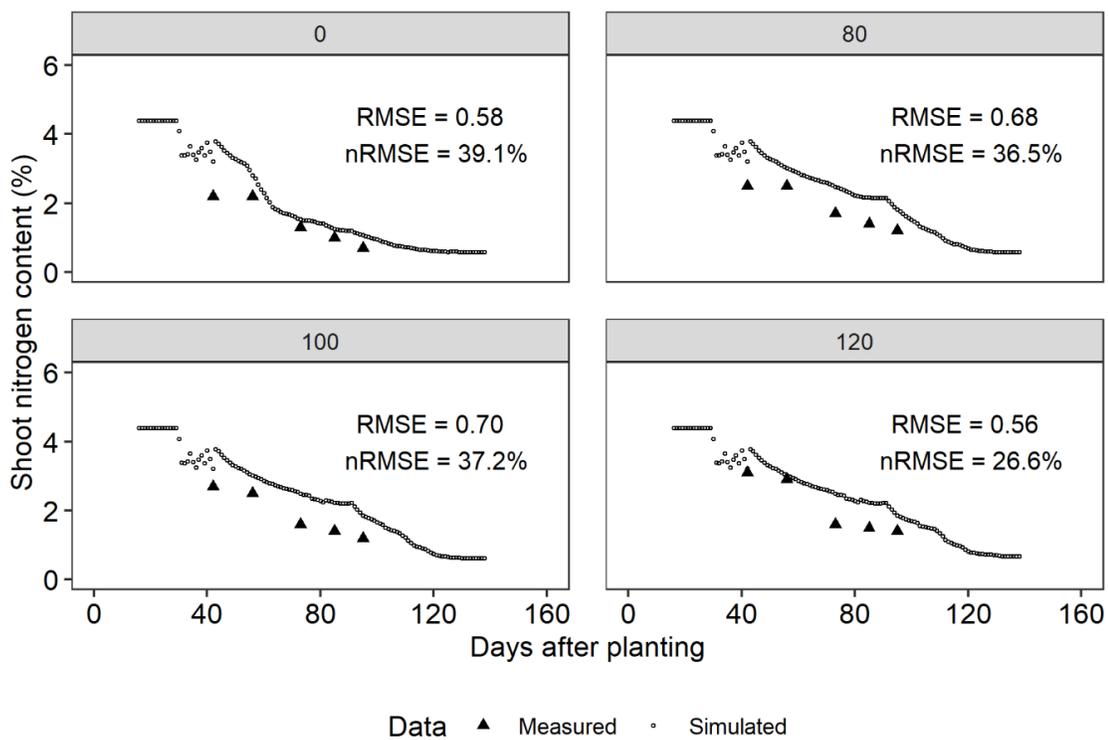
Figure 3.12. Measured and simulated aboveground shoot nitrogen content (%) at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2016.
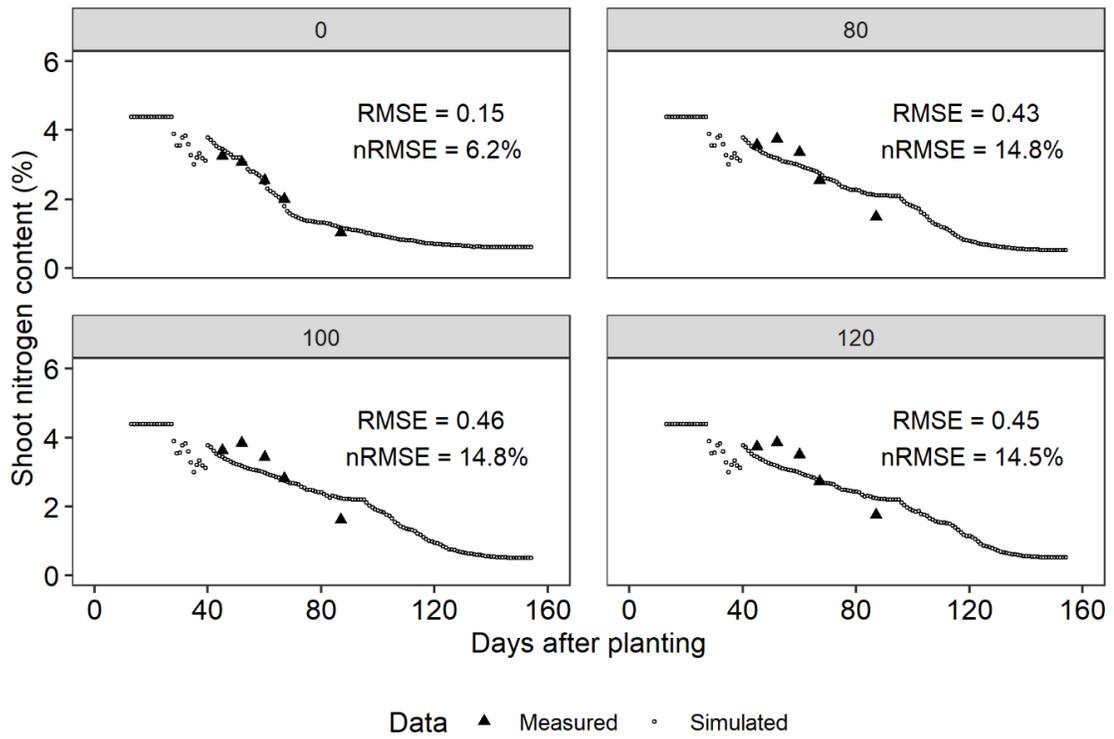
Figure 3.13. Measured and simulated aboveground shoot nitrogen content (%) at different days after planting at 0, 80, 100, and 120% N rates in Lamberton, MN during 2017.
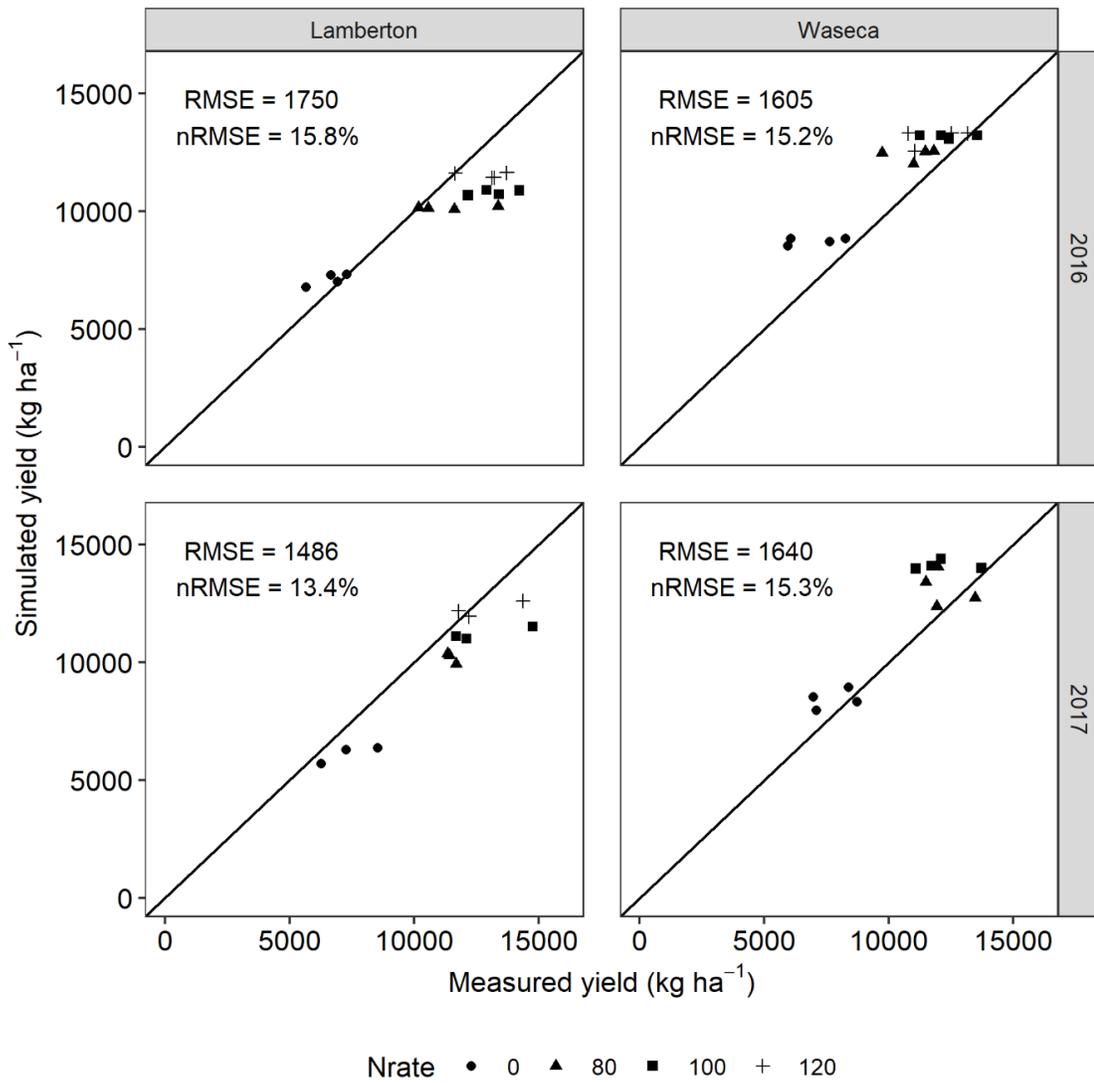
Figure 3.14. Measured and simulated maize grain yield at 0, 80, 100, and 120% N rates in Lamberton and Waseca, MN during 2016 and 2017.

CHAPTER 4. Improving site-specific maize yield estimation by integrating satellite multispectral data into a crop model

Synopsis

Integrating remote sensing data into crop models offers opportunities for improved crop yield estimation. To compare site-specific yield estimation accuracy of a stand-alone crop model with a data-integration approach, a study was conducted in 2016-2017 with nitrogen (N) fertilized and unfertilized treatments across a heterogeneous 7-ha maize field. For each treatment, yield data were grouped into five classes resulting into 109 spatial zones. In each zone, the CERES-Maize model was run using the GeoSim plugin within QGIS. In the data integration approach, maize biomass estimated using satellite imagery at the five (V5) and ten (V10) leaf-collar stages were used to optimize the total soil nitrogen concentration (SLNI) and soil fertility factor (SLPF) in CERES-Maize. Without integration, maize yield was simulated with root mean square error (RMSE) of 1264 kg ha^{-1}. Optimization of SLNI improved yield simulations at both V5 and V10. However, better simulations were obtained from optimization at V10 (RMSE 1026 kg ha^{-1}) as compared to V5 (RMSE 1158 kg ha^{-1}). Optimization of SLPF together with SLNI did not further improve the yield simulations. This study shows that integrating remote sensing data into a crop model can improve site-specific maize yield estimations as compared to stand-alone crop modeling approach.

4.1 Introduction

Early estimation of crop yield is important to farmers, government agencies, and policy makers to improve crop production efficiency and detect potential biotic and abiotic risks that affect crop yield (Basso et al., 2013; Chen et al., 2019). Mounting

pressures to address environmental problems resulting from crop production (Alexander et al., 2008; Mclellan et al., 2015) and increasing competition for greater economic efficiency (Koch et al., 2004) have directed research efforts for site-specific crop yield estimation. Spatially-explicit estimation of crop yield not only helps to explain the spatial variability of crop growth within a field but also to optimize crop management efforts and reduce risks (Raun et al., 2002). In recent decades, technological advancement in satellite-based global positioning system (GPS), improved sensor capabilities, computational tools, and geographic information systems (GIS) have greatly enabled digital data-driven approaches for site-specific crop yield estimation (Earl et al., 2000; Stafford, 2000; Thessler et al., 2011; Kamilaris et al., 2017). Frequently used approaches include yield maps (Birrell et al., 1996; Blackmore, 2000), remote sensing images (Shanahan et al., 2001; Peralta et al., 2016; Ban et al., 2017), and process-based crop simulation models (Batchelor et al., 2002; Thorp et al., 2015).

Sensors mounted on combine harvesters calculate the mass of grain per unit of area harvested, which together with GPS receivers provide grain yield measurements at geo-referenced points to produce yield maps that are effective in visualizing spatial variability of crop yield (Grisso et al., 2009). Historical yield maps help to locate high and low yielding regions within a field and are useful in estimating site-specific yield (Blackmore, 2000; Maestrini and Basso, 2018). However, yield maps alone are less revealing as to the cause of yield variation (Grisso et al., 2009). Furthermore, historical yield maps can vary widely between years due to differences in growing season weather making future in-season yield estimates challenging (Ray et al., 2015; Leng et al., 2016). Remote sensing images, in this regard, are advantageous to monitor in-season crop

growth patterns in response to the effects of weather, pests, disease, and other management issues.

Remote sensing devices, such as multispectral and hyperspectral sensors, when mounted on aerial or satellite platforms can cover large areas and can give rapid assessment of within-field variability of crop growth. Spectral indices, such as the normalized difference vegetation index (NDVI) (Kriegler et al., 1969; Rouse et al., 1974) obtained from remote sensing images correlate with crop growth status, which help to estimate site-specific crop yield (Shanahan et al., 2001; Peralta et al., 2016; Guan et al., 2017; Maestrini and Basso, 2018). Additionally, time-series spectral data from remote sensing devices are also used to develop models to characterize temporal change in crop growth, such as leaf area index, which have been found promising for crop yield estimation (Ban et al., 2017). Despite distinguishing crop growth and yield variability and being advantageous at both temporal and spatial scales, remote sensing images alone, similar to yield maps, have limitations in determining yield-limiting factors, such as nutrients, water, or pests.

Process-based crop models incorporate soil, weather, cultivar, and crop management information to simulate crop growth and yield (Jones et al., 2003; Keating et al., 2003). Therefore, crop models have an advantage over yield maps and remote sensing images by helping to identify the sources of yield variation in a way that can be used to optimize crop management efforts. The major limitation of crop models comes from their point-based nature. Since crop models typically only incorporate input data from one point in space, they have reduced capability at larger spatial scales. For example, heterogeneity in soil properties across a field requires extensive parameterization for larger spatial application of crop models (Thorp et al., 2015). Due to

differences in field topography and soil properties, events like runoff are inevitable and can change soil moisture conditions, nitrogen (N) levels, and ultimately yield. The point-based nature of crop models do not account for such spatial processes of runoff from adjoining regions (Batchelor et al., 2002). High resolution soil sampling, in terms of space and time, can detect spatial processes but can be impractical. Additional constraints for crop model applications are the requirement of multiple site-years of field-collected data on a particular cultivar for calibration and evaluation purposes. In practical situations, multiple sites and years of field data are not always available for detailed calibration and evaluation. In addition, new cultivars are released frequently, making any previous calibration efforts obsolete. In such data-scarce situations, one option is to integrate soil and crop parameters obtained from remote sensing images into crop models. As satellite or aerial images have an advantage of rapid, site-specific estimation of soil and crop bio-physical parameters across larger spatial scales, they offer great opportunities to address the limitations of crop models by supporting spatial modeling applications and facilitating spatial calibration of models (Delécolle et al., 1992; Batchelor et al., 2002; Thorp et al., 2015).

Numerous attempts have been made at integrating remote sensing data into point-based crop models for site-specific crop yield estimation (Dente et al., 2008; Fang et al., 2008; Thorp et al., 2015; Ban et al., 2019). Primary schemes of data integration have been to replace, update/re-initialize, or optimize the model inputs that are missing or difficult to measure across the field (Delécolle et al., 1992; Guérif and Duke, 2000; Launay and Guerif, 2005; Dorigo et al., 2007). For example, Dente et al. (2008) integrated leaf area index (LAI) estimated from remote sensing images into the CERES-Wheat model to map wheat (*Triticum aestivum* L.) yield variability in Southern Italy at

catchment scale to optimize sowing date, wilting point, and field capacity parameters, which improved yield estimates as compared to a no-integration approach. Ines et al. (2013) integrated Moderate Resolution Imaging Spectroradiometer (MODIS) LAI and remotely estimated soil moisture into the CERES-Maize model and obtained improved maize (*Zea mays* L.) yield simulations compared to stand-alone crop model outputs. More recently, Ban et al. (2019) also integrated MODIS derived seasonal LAI as well as water stress factors to the CERES-Maize model for improved accuracy of maize yield estimations. Similar improvements with the use of data integration have been reported by Guo et al. (2018), Launay and Guerif (2005), and Li et al. (2015). Although data integration has been shown to be an effective method to enhance site specific yield estimation, most of the previous conclusions have been drawn from state or regional level studies (Dente et al., 2008; Fang et al., 2008; Ines et al., 2013; Ban et al., 2019) or plot-scale experiment (Li et al., 2015). Limited studies have been carried out to address within field variability (Guo et al., 2018; Gaso et al., 2019). Studies at the field scale are important to examine the efficacy of data-integration for site-specific crop production optimization and for precision management of agricultural inputs. Besides, previous studies on data integration have frequently used LAI as the state variable to connect remote sensing and crop modeling (Dente et al., 2008; Fang et al., 2008; Ines et al., 2013; Li et al., 2015; Gaso et al., 2019). Surprisingly, crop parameters other than LAI, such as biomass and canopy N content, have been less explored. Additionally, time series measurements of crop parameter have been used in past studies with little attention given to early growth stage measurements. Options for crop management are restricted at later growth stages (late vegetative to reproductive stages). Studies on data-integration at early stages of crop development will aid in decision making, for example in precision application of fertilizers and irrigation. Despite several studies,

integration of remote sensing data into crop models is not yet a common practice and

several details of the integration processes still need to be worked out. Some of the

crucial information that are lacking include the identification of process to correspond

remote sensing measurements to state variables of crop models, improvement of data

integration method, and implementation of the integration process to several aspects of

crop production. The objectives of this study were to (i) develop an approach for

improved maize yield estimation by integrating multispectral data of maize canopy at

early vegetative stages into a crop growth model and (ii) compare site-specific yield

estimation accuracy of a stand-alone crop model with a data-integration approach where

soil parameters in the crop model were spatially optimized from satellite images derived

biomass at five (V5) and ten leaf-collar (V10) stages.

4.2 Materials and methods

4.2.1 Study site and year

The field study was conducted during the growing seasons of 2016 and 2017 at

the University of Minnesota Southwest Research and Outreach Center located near

Lamberton, MN, USA (44°14′19″ N and 95°18′50″ W) at an elevation of about 350 m

ASL. The dominant soil type is Normania clay loam (Fine-loamy, mixed, superactive,

mesic Aquic Hapludoll) (Soil Survey Staff, 2018). The region has a hot-summer humid

continental climate.

4.2.2 Field experiment and data collection

The experiment was carried out across a heterogeneous 7-ha field in strips with

and without N fertilizer; the fertilized strips received sidedress applications of urea

ammonium nitrate at the six leaf-collar stage (V6) at the rate of 118 and 135 kg N ha^{-1} in

2016 and 2017, respectively. Four replications of the two treatments were randomized in a complete block design. The maize hybrid 'Pioneer P0297' was planted with a row spacing of 0.76 m for a target population of 88,900 plants ha^{-1}. Each strip contained a minimum of 18 rows and was at least 230 m long (Fig. 4.1).

In 2016, soil samples were collected before planting at 34 geo-referenced points from the 0−30-, 30−60-, 60−90-, and 90−120-cm depths for analysis of soil texture. Each soil sample was composed from two sub-samples collected around each point. Soil samples from the 0−30- and 30−60-cm depths were analyzed for total N, ammonium-N, nitrate-N, organic carbon, cation exchange capacity, and pH. In 2017, soil samples were collected before planting from an additional 8 sampling points within the N fertilized treatments at 0−30- and 30−60-cm depths for total N, ammonium-N, and nitrate-N. In both years, soil samples were air-dried and sieved to pass a 2 mm mesh before lab analysis.

In both years, plant stand counts across 3 m lengths in three different rows were determined around each geo-referenced sampling point. In 2016, five plant samples were collected within a 2-m radius of geo-referenced points at the eight leaf-collar (V8), tasseling (VT), and physiological maturity stages and air dried in forced-air oven at 60°C until constant mass for dry biomass. As in 2016, plant samples were collected in 2017 at three dates near five leaf-collar (V5), ten leaf-collar (V10), and VT growth stages for dry biomass. In both years, maize grain yield was measured using a field-scale combine (John Deere S600, Deere and Co., Moline, IL, USA) equipped with a yield monitoring device (John Deere StarFire^{TM} 3000 GPS receiver and GreenStar^{TM} 2620 display, Deere and Co., Moline, IL, USA). Yield data points in both years were processed using

GreenStar™ Apex 3.7.9 farm management software following the John Deere S-series combine calibration guidelines (Deere and Company, 2012).

Daily weather data of maximum and minimum air temperature (°C), solar radiation (MJ m^{-2}), and rainfall (mm) were obtained from the weather station located approximately 800 m from the field.

4.2.3 Satellite imagery and image processing

RapidEye Ortho - Level 3A satellite images were used for this study. The images are orthorectified tile products with geometric and terrain corrections in Universal Transverse Mercator map projection (Planet, 2016). The RapidEye satellite used a multi-spectral imager with five channels sensitive between 440 and 850 nm and providing images with a spatial resolution of 5 m (Table 4.1).

Satellite images acquired on 26 June (Julian day 177) and 17 July (Julian day 195) in 2017 were used for this study, which corresponded to the V5 and V10 stages of maize development, respectively. Digital numbers (DN) in images were first converted to top of atmosphere radiance (RAD) values by multiplying the DN with radiometric scale factor given in meta-data file of the images. The RAD values (Wm^{-2} μm^{-1} sr^{-1}) were then converted to reflectance (REF) values (dimensionless) using the formula (NASA, 2003):

$$REF\ (i) = RAD(i)\ \times\ \frac{\pi \times SunDist^2}{EAI(i) \times \cos(SolarZenith)} \tag{1}$$

where i refers to corresponding band, SunDist denotes the Earth-Sun distance at the image acquisition date in Astronomical units, EAI refers to Exo-Atmospheric Irradiance and SolarZenith denotes solar zenith angle in degrees obtained by subtracting sun elevation from 90°. SunDist values for each imagery acquisition date were obtained from

Chander et al. (2009). The EAI values for each band were obtained from RapidEye (Planet, 2016). The sun elevation values for each imagery acquisition date were obtained from the meta-data file of the images. The REF values of spectral bands were then used to calculate vegetative indices. Previous studies have shown strong correlation of maize growth status with NDVI and red-edge NDVI (RENDVI), (Freeman et al., 2007; Martin et al., 2007; Kross et al., 2015). Therefore, only these two indices were considered for this study. The NDVI (Kriegler et al., 1969; Rouse et al., 1974) and RENDVI (Gitelson and Merzlyak, 1994; Mutanga and Skidmore, 2004) for each date were calculated as follows:

$$NDVI = ((NIR - R)) / ((NIR + R)) \qquad (2)$$

$$RENDVI = ((NIR - RE)) / ((NIR + RE)) \qquad (3)$$

where NIR, R, and RE denotes reflectance values of near-infrared, red and red-edge spectral bands, respectively. Regression between both indices and maize biomass measured at V5 and V10 were conducted. The best performing index based on higher coefficient of determination (r^2) was used for biomass estimations across the field. All image processing steps and calculations were carried out using the ModelBuilder and Raster Calculator tools in ArcGIS Desktop 10.5.1 (ESRI Inc., 2017).

4.2.4 CERES-Maize model

The Cropping System Model (CSM)-CERES-Maize (Jones and Kiniry, 1986), one of 42 crop simulation models within the Decision Support System for Agrotechnology Transfer (DSSAT) v. 4.7.5 (Jones et al., 2003; Hoogenboom et al., 2019) was used for this study. The DSSAT incorporates the dynamics of soil-plant-atmosphere interactions

and simulates maize growth, development, and yield as a function of genotype, weather, soil, and crop management information (Jones et al., 2003).

4.2.4.1 Model inputs

The minimum dataset to run DSSAT crop models includes daily weather data, crop management information, soil profile data, and cultivar information. The minimum weather inputs are daily maximum and minimum air temperature (°C), solar radiation (MJ m^{-2}), and precipitation (mm). Weather data were obtained from the automated weather station located at the research site. The WeatherMan tool in DSSAT was used to prepare weather files for the CSM-CERES-Maize model. Crop management information including sowing date, seeding rate, tillage, and fertilization strategy used in the study were entered through XBuild, a DSSAT tool for describing experiments. Similarly, SBuild, a tool for creating and modifying soil profile data as required for model simulations, was used to enter soil texture, total N%, ammonium-N (ppm), nitrate-N (ppm), organic carbon%, cation exchange capacity, and soil pH obtained from soil analyses. SBuild was also used to estimate the missing data on bulk density, saturated water content, field capacity, wilting point, and saturated hydraulic conductivity based on the soil texture through pedo-transfer functions. Cultivar information on days to anthesis and physiological maturity, biomass at anthesis, and yield was obtained from the fertilized treatment of the 2016 growing season.

4.2.4.2 Geospatial data management

The measured and estimated soil properties at each soil sampling point and depth in 2016 were associated with their respective geographic coordinates in a point shapefile. Soil properties at each sampling depth were interpolated into soil layers using ordinary kriging. Altogether, 44 soil layers were created, which included measured and

estimated soil properties at different soil profile depths. Similarly, yield data points from each fertilized and unfertilized plot from the 2017 growing season were interpolated into yield layers. The interpolated yield layers were resampled to 5 m to match the spatial resolution of satellite imagery. Then, resampled yield layers were grouped into five classes using an isocluster unsupervised classification algorithm, resulting in 109 zones (51 in fertilized and 58 in unfertilized treatment). These zones were later used as base layer polygons for crop simulation. All kriging, resampling and classification were carried out in ArcGIS Desktop 10.5.1 (ESRI Inc., 2017). Average values of yield and each soil property for each zone were calculated using zonal statistics within the raster analysis toolbox in QGIS 3.6.2 (QGIS Development Team, 2019). Average values of estimated biomass from remote sensing imagery for each zone were also obtained using the same procedure. Zonal statistics tool appended the averaged data to the respective zone in the attribute table. Soil files and crop management files required to run CSM-CERES-Maize in each zone were created as required for the DSSAT file format using template files in Geospatial Simulation version 1.3 (GeoSim) (Thorp and Bronson, 2013) in Quantum GIS 3.6.2 software (QGIS Development Team, 2019). The simulation control tool in GeoSim was then used to run simulations in each zone, which transferred all the required geospatial data located in the attribute table from the respective zone to the model.

4.2.4.3 Model calibration

Calibration of CSM-CERES-Maize requires the estimation of six genotype-specific coefficients: P1 (thermal time from seedling emergence to end of juvenile phase), P2 (extent of delay in development for each hour with daylength above 12 hours), P5 (thermal time for silking to physiological maturity), PHINT (phyllochron

interval between successive leaf tip appearances), G2 (maximum possible number of kernels per plant), G3 (kernel growth rate during grain-filling stage under optimum conditions). The first four coefficients (P1, P2, P5, and PHINT) regulate the development of maize, whereas the last two coefficients (G1 and G2) regulate yield of maize. For model calibration, phenology observations, biomass and yield data collected from the N fertilized treatments in 2016 were used. Generalized Likelihood Uncertainty Estimation (GLUE), a Bayesian parameter estimation procedure within DSSAT was used for calibration. Initially, phenology-related parameters were calibrated using anthesis and physiological maturity dates. Later, yield specific parameters were estimated using measured biomass and yield. At each step, GLUE was set to run for a total of 10,000 iterations to estimate the genotype coefficients.

4.2.4.4 Spatial optimization and data integration

Agricultural fields are inherently heterogenous due to spatially variable soil characteristics. Yet, because of their point-based nature, crop models such as in DSSAT assume homogenous field areas. Therefore, at a field-scale crop simulation, site-specific optimization of soil parameters is required to account for spatial variability in soil properties. As of now, however, DSSAT does not currently have a tool for spatial optimization of soil parameters. In this study, biomass estimated from satellite imagery was used to account for the effect of the spatial variability of soil on maize growth. For this, CSM-CERES-Maize was run using the simulation optimizer tool in GeoSim which performed spatial optimization. Geosim employs a simulated annealing algorithm and facilitates the optimization process by adjusting the user-selected parameters to improve the agreement between measured and simulated data (Thorp and Bronson, 2013). In this study, for site-specific soil parameter optimization, maize biomass estimated from

remote sensing images at V5 and V10 was used (Fig. 4.2). Soil parameters chosen for optimization were total soil N concentration (SLNI) in top 30-cm of soil and soil fertility factor (SPLF). As the research involved N-fertilized and unfertilized treatments and N is highly mobile nutrient, spatial processes such as leaching and runoff could have changed the soil N level. To account for spatial variability in soil N level, SLNI was selected. The SLPF was chosen to address soil fertility issues other than N that were not accounted for in the model. Based on soil analyses, initial values of SLNI were set at 0.2 and 0.1% for N fertilized and unfertilized treatments, respectively. The SLPF was set to 1 for both treatments. For optimization, SLNI was allowed to vary from 0.01 to 0.4%, whereas SLPF was allowed to vary from 0.7 to 1. Both parameters were optimized independently as well as simultaneously to reduce the error between estimated biomass from satellite imagery and biomass simulated by the model. The optimization involved initialization of SLNI and SLPF with the optimized values.

4.2.4.5 Model evaluation

The model was evaluated for phenology, biomass, and grain yield data collected during the 2017 growing season. Two separate evaluations were conducted, without and with site-specific optimization of soil properties. For evaluation, root mean square error (RMSE) and normalized RMSE (nRMSE) were used. The RMSE between simulated and measured values was calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Si-Mi)^2}{n}} \tag{4}$$

where n refers to number of observations, S and M denote simulated and measured values, respectively. Then, nRMSE was calculated as RMSE / (mean of measured values). Lower RMSE and nRMSE values indicate better model performance.

4.3 Results and discussions

4.3.1 Weather conditions during the growing season

Growing seasons during both study years received more rainfall than the historical average (Table 4.2). Total cumulative rainfall was 723 and 651 mm during the 2016 and 2017 growing seasons, respectively. In both years, May was wetter whereas June was drier than the historical average (Table 4.2).

Considering the historical average of the growing season, 2016 was warmer by 0.75°C whereas 2017 was cooler by 0.02°C. During both years, June, September, and October were warmer whereas July was cooler (Table 4.2). August in 2017 was also cooler by 1.7°C than the historical average. Monthly average air temperature during the 2016 growing season was often higher than in 2017 (Fig. 4.3).

4.3.2 Relation between vegetative indices and maize biomass

At the V5 stage, measured maize biomass ranged from 112 to 519 kg ha^{-1}, whereas at the V10 stage, maize biomass ranged from 1343 to 5079 kg ha^{-1}. The r^2 between biomass and RENDVI at the V5 and V10 stages was 0.44 and 0.58, respectively, whereas r^2 with NDVI was 0.55 and 0.63, respectively (Fig. 4.4).

At both stages, NDVI had greater r^2 than RENDVI. For both indices, greater r^2 was achieved at V10 than at V5. The NDVI values increased with the growth stages; from 0.24 to 0.35 at V5 and from 0.49 to 0.68 at V10. At both stages, lower and higher NDVI readings were associated with lower and higher biomass, respectively (Fig. 4.4).

The predictive ability of NDVI to explain maize biomass variability has been well documented (Freeman et al., 2007; Martin et al., 2007; Kross et al., 2015). As observed

in this study, increase in NDVI as maize grows and develops has also been reported by Shaver et al. (2010) and Martin et al. (2007). However, as maize growth progresses, NDVI tends to saturate after canopy closure. In this situation, NDVI can have lower explanatory ability to distinguish biomass variability (Sharma et al., 2015). As the red edge region of the electromagnetic spectrum can penetrate deeper into the crop canopy due to lower chlorophyll absorption, red edge-based indices such as RENDVI have been shown to better explain growth variability at later growth stages (Li et al., 2014; Sharma et al., 2015).

The regression equations for the relationship between NDVI and biomass (Fig. 4.4) were used to estimate maize biomass for all pixels in the RapidEye imagery. The spatial variability of estimated biomass ranged from 105 to 714 kg ha^{-1} at V5 (Fig. 4.5) and from 1209 to 5775 kg ha^{-1} at V10 (Fig. 4.6). The ranges of estimated biomass did not vary greatly with the measured biomass. The use of vegetative indices calculated from remote sensing data to estimate crop growth parameters is frequent in literature. For example, Kross et al. (2015) used RapidEye imagery to calculate several vegetative indices, including NDVI and RENDVI, to estimate LAI and biomass of maize and soybean (*Glycine max* L. Merr.). Parallel to the finding of this study, Kross et al. (2015) also found NDVI better in estimating biomass as compared to other indices. Gitelson et al. (2003) proposed new indices based on near infrared, red, and green bands of the spectrum for improved estimation of LAI and leaf biomass in maize. In addition to these vegetation indices, physical models involving physical laws and the inversion of canopy reflectance spectra are also used (Dorigo et al., 2007). For example, Thorp et al. (2012) used the inversion of the PROSAIL radiative transfer model to estimate wheat growth parameters that affected LAI and biomass in the CERES-Wheat model. Unlike statistical

models, physical models require estimation of several variables of soil, leaf, and canopy to simulate the reflectance at leaf and canopy level (Dorigo et al., 2007).

4.3.3 Model calibration genetic coefficients

The calibrated coefficients obtained (Table 4.3) gave satisfactory simulation of phenology. Days to anthesis was accurately predicted, within one day of the observed results and days to physiological maturity was also accurately predicted. During calibration, biomass was simulated with a RMSE of 2312 kg ha^{-1} at anthesis and RMSE of 1678 at physiological maturity; yield was simulated with a RMSE of 2175 kg ha^{-1}.

4.3.4 Model evaluation with and without spatial optimization

Upon evaluation with 2017 data, the calibrated model simulated anthesis and physiological maturity accurately. Physiological maturity in 2017 was delayed and occurred in late September (as compared to the usual time in early September) due to cooler than average air temperatures in July and August (Table 4.2), and the model correctly captured this delay in physiological maturity.

Without spatial optimization, the model overpredicted maize biomass at V5 with RMSE of 264 kg ha^{-1} (nRMSE 101%), which was more than twice the average measured biomass. The results were similar for the N fertilized and unfertilized treatments (Table 4.4). Site-specific optimization of SLNI based on V5-estimated biomass slightly reduced the simulated biomass RMSE to 245 kg ha^{-1} (nRMSE 94%). Simultaneous adjustment of SLNI and SLPF did not further improved the simulation (Table 4.4). Improvement in biomass simulations after SLNI optimization was greater in the N fertilized treatment which reduced the RMSE from 275 to 236 kg ha^{-1} (nRMSE from 102 to 88%).

Poor simulations of maize biomass at V5 could have been improved from better calibration with more in-season biomass samplings. Also, maize biomass data from multiple sites or years could have given more accurate cultivar coefficients during calibration to simulate biomass at early growth stages. Only slight improvement in simulation after optimization of SLNI and SLPF also refers to the calibration of genotypic specific coefficients or any crop parameters rather than soil parameters. Another possible reason could have been due to earlier emergence of maize in the model than in the actual field conditions, which led to greater biomass simulations at V5. Previous studies have used remote sensing imagery to calibrate genotype specific coefficients and crop management parameters. Thorp et al. (2015) for example, used LAI estimates from NDVI readings to spatially optimize SLAVR of DSSAT/CROPGRO-Cotton model which regulates potential specific leaf area in cotton (*Gossypium hirsutum* L.). Similarly, Li et al. (2015) used remotely estimated LAI and canopy nitrogen content in DSSAT/CERES-Wheat model to optimize several crop genotype specific parameters including PHINT and crop management parameters such as plant population. However, our study did not consider optimization of any genotype specific parameters and only optimized soil-related parameters to account for spatial variability in soil properties.

Considering nRMSE results, the biomass simulation results were comparatively better at the V10 than at V5. Without spatial optimization, biomass at the V10 was simulated with RMSE of 1255 kg ha^{-1} (nRMSE 36%) (Table 4.4). Biomass in the N-fertilized treatment (nRMSE 19%) was simulated better than in the unfertilized treatment (nRMSE 46%). The RMSE of N-fertilized simulated biomass was 654 kg ha^{-1} whereas the unfertilized treatment was 1608 kg ha^{-1}. Biomass in the unfertilized treatment was underpredicted because of a greater N stress simulation by the model.

86

Following site-specific optimization of SLNI using V10-estimated biomass, the overall RMSE for simulated biomass was reduced by 12.8% from 1255 to 1094 kg ha^{-1}. Simultaneous optimization of SLNI and SLPF reduced the RMSE by 22.7% to 969 kg ha^{-1}. The optimization improved the biomass simulation in both N fertilized and unfertilized treatments. Optimization of only SLNI reduced the RMSE from 654 to 617 kg ha^{-1} in the fertilized treatment. The RMSE was reduced even lower to 507 kg ha^{-1} when both SLNI and SLPF were optimized simultaneously (Table 4.4). The trend was similar in the unfertilized treatment. Optimizing SLNI alone reduced the RMSE from 1608 to 1383 kg ha^{-1} in the unfertilized treatment. Simultaneous optimization of both SLNI and SLPF further reduced the RMSE to 1240 kg ha^{-1}. Thus, optimization of SLNI resulted in 5.6 and 13.9% reduction in RMSE in the N fertilized and unfertilized treatments, respectively, whereas simultaneous optimization of both SLNI and SLPF resulted in 22.8 and 22.4% reduction in RMSE in N fertilized and unfertilized treatments, respectively. Following optimization, greater reduction in RMSE in unfertilized treatments could be from the improvement in the representation of soil N level which eventually decreased the N stress simulation in the model.

The average measured maize grain yield ranged from 10547 to 11009 kg ha^{-1} among N-fertilized strips and 4393 to 5984 kg ha^{-1} among unfertilized strips (Table 4.5). Without spatial optimization, overall maize yield was simulated with RMSE of 1264 kg ha^{-1} (nRMSE 15.7%) (Fig. 4.7). As with biomass, yield in the N-fertilized treatment was also simulated better than the unfertilized treatment. The RMSE of simulated yield in the N-fertilized treatment was 1132 kg ha^{-1} (nRMSE 10%), whereas in the unfertilized treatment it was 1370 kg ha^{-1} (nRMSE 25%). The coefficients of variation of the average simulated yield without optimization were less than that of the measured yield in both N-

fertilized and unfertilized strips (Table 4.5), indicating that soil properties were more homogenous in the model than in the actual field conditions.

Spatial optimization of SLNI from data integration at the V5 and V10 stages improved the overall yield simulations (Fig. 4.7). However, slightly better yield simulations were obtained from optimization of SLNI at V10 (nRMSE 12.8%) as compared to V5 (nRMSE 14.4%). Only optimizing SLNI using V5-estimated biomass reduced the overall RMSE by 8.4% from 1264 to 1158 kg ha^{-1}. Simultaneous optimization of SLNI and SLPF slightly increased the RMSE from 1264 to 1271 kg ha^{-1}. Following optimization of SLNI at V5, the improvement in yield simulations were from fertilized strips, which lowered the over-predicted yield. Yield simulations in unfertilized strips, however, remained unchanged (Table 4.5). In addition to decreasing the yield in all strips, simultaneous optimization of SLNI and SLPF at V5 increased the coefficients of variation of average simulated yields. This increment in coefficients of variation indicates the creation of heterogeneous soil conditions in the model in terms of soil N concentration and soil fertility level.

Spatial optimization of SLNI only using V10-estimated biomass reduced the RMSE by 18.8% and gave the lowest RMSE of 1026 kg ha^{-1}. Simultaneous optimization of SLNI and SLPF using V10- estimated biomass produced a RMSE of 1058 kg ha^{-1}, which was slightly more than from the optimization of SLNI alone. With biomass at both V5 and V10, simultaneous optimization of SLNI with SLPF did not give better yield simulation as compared to single optimization of SLNI. After spatial optimization of SLNI, improvement in yield simulation occurred in both the N-fertilized and unfertilized treatments. Improvement in the N-fertilized treatment occurred from optimization at both V5 and V10 whereas improvement in the unfertilized treatment occurred only from

optimization at the V10 stage. In addition to improvement in yield simulations, optimization of SLNI at V10 improved the coefficients of variation of the average simulated yield of all unfertilized strips. This improvement in coefficients of variation showed the enhancement in representing spatial heterogeneity in soil N concentration in the model.

After data integration, the overall improvement in yield simulation was from better representation of soil N levels in the field. Estimation of biomass at the V10 stage from satellite imagery depicted the spatial variability in maize growth. As the study site has an elevation ranging from 337 to 345 m with a downward slope towards the east (Fig. 4.8), it is likely that zones at lower areas may have received run-off water together with N and higher levels of organic matter from upland N-fertilized treatment areas. As reported by Batchelor et al. (2002) and Fraisse et al. (2001), enhancement in crop models for accurate site-specific yield estimation requires consideration of spatial events such as runoff from uplands. Although this study did not calculate any two- or three- dimensional movement of water or nutrients between the spatial zones, this study has shown an option for point-based simulation to account for the effects of these spatial events at field scale. Accounting for such spatial events in the crop model not only improves the estimation of site-specific yield, but also enhances the decision-making process to address spatial and temporal variability of the soil-crop relationship within the field. Understanding of this variability within the field is essential for making make informed decisions on crop management for greater economic efficiency and addressing environmental problems. As the variability in the soil-crop relationship within the field eventually results in differences in crop yield, early estimation of spatially-explicit crop yield at the field scale can aid in precision application of agricultural inputs.

4.4 Conclusions

This study evaluated the accuracy of site-specific maize yield estimation of the stand-alone CERES-Maize model with a data integration approach in which satellite multispectral images were used to optimize soil parameters in CERES-Maize. Spatial optimization of SLNI using estimated biomass at both V5 and V10 stages improved the overall yield simulations. More accurate yield estimations, however, were obtained from optimization at V10 as compared to V5 stage. Subsequent optimization of SLNI and SLPF did not further improve the yield simulations. This study shows that integrating remote sensing data into a crop model better depicts the whole soil-crop relationship within a field as compared to the crop model alone. This integrated approach is promising for site-specific maize yield estimation and for planning spatially variable management of agronomic inputs like fertilizers and irrigation aiming at improving resources use efficiency in crop production. Future studies on the use of higher-resolution images for site-specific optimization would further enhance the application of crop models in precision agriculture. Further research on crop model uncertainties due to soil inputs at the field scale and from different spatial interpolation methods used to estimate soil properties would help in improving soil data collection and crop model applications at large scale.

Table 4.1. Spectral bands and their respective range of wavelengths for RapidEye imagery.

Spectral Bands	Wavelength (nm)
Blue	440-510
Green	520-590
Red	630-685
Red Edge	690-730
Near Infrared	760-850

Table 4.2. Comparison of 2016 and 2017 monthly average air temperature (Tavg; °C) and total rainfall (Rain; mm) with historical averages at Lamberton, MN, USA.

Month	Historical average (1961-2014)		Deviation from historical average[‡]			
			2016		2017	
	Tavg	Rain	Tavg	Rain	Tavg	Rain
May	14.4	88	+0.3	+53	−0.8	+64
June	20.2	104	+1.2	−38	+0.5	−36
July	22.5	91	−0.4	+85	−0.1	+11
Aug	20.5	80	+0.9	+55	−1.7	+45
September	16.1	78	+1.6	+55	+1.6	-24
October	9.2	52	+0.9	+20	+0.3	+98

‡ "+" denotes above and "−" denotes below historical average.

Table 4.3. Calibrated values of cultivar coefficients for CERES-Maize obtained using generalized likelihood uncertainty estimation procedure in DSSAT v. 4.7.5.

Maize cultivar coefficient	Calibrated values
Thermal time from seedling emergence to the end of the juvenile phase (P1)	159.7
Extent to which development is delayed for each hour that daylength is above 12.5 hours (P2)	1.409
Thermal time for silking to physiological maturity (P5)	669.4
Phyllochron interval between successive leaf tip appearances (PHINT)	38.9
Maximum possible number of kernels per plant (G2)	875.3
Kernel growth rate during linear grain filling stage under optimum conditions (G3)	8.7

Table 4.4. Root mean square error (RMSE; kg ha^{-1}) and normalized RMSE (nRMSE; %) between measured and simulated maize biomass in nitrogen (N) fertilized and unfertilized treatments at five (V5) and ten leaf-collar (V10) stages with and without spatial optimization.

Optimization[‡]	Average	N Fertilized	N unfertilized
	RMSE (nRMSE)		
	Biomass at V5		
None	264 (101)	275 (102)	253 (99)
SLNI	245 (94)	236 (88)	253 (99)
SLNI and SLPF	245 (94)	236 (88)	252 (99)
	Biomass at V10		
None	1255 (36)	654 (19)	1608 (46)
SLNI	1094 (31)	617 (18)	1383 (39)
SLNI and SLPF	969 (28)	507 (14)	1240 (36)

‡None refers to without any spatial optimization. SLNI refers to optimization of total soil nitrogen concentration (SLNI) only. SLNI and SLPF refer to simultaneous optimization of SLNI and soil fertility factor (SLPF).

Table 4.5. Average measured and simulated maize yield (kg ha^{-1}) with coefficient of variation (%; in parenthesis) of nitrogen (N)-fertilized and unfertilized strips with and without spatial optimization.

Yield[‡]	F1[¥]	F2	F3	F4	UF1[¥]	UF2	UF3	UF4
Measured	10861	11009	10858	10547	5984	5903	4393	5719
	(5.0)	(3.1)	(3.4)	(5.6)	(7.6)	(20.7)	(30.0)	(15.2)
No optimization	11966	11945	11864	11799	4652	4705	4627	4735
	(2.3)	(1.5)	(1.3)	(2.2)	(3.6)	(2.3)	(3.9)	(2.0)
V5 SLNI	11568	11577	11530	11534	4652	4705	4627	4735
	(1.1)	(1.0)	(1.0)	(1.3)	(3.5)	(2.4)	(3.9)	(2.0)
V5 SLNI and SLPF	10452	10681	11350	11060	4626	4678	4625	4686
	(1.2)	(1.1)	(5.5)	(8.0)	(3.7)	(2.4)	(3.9)	(2.1)
V10 SLNI	11603	11837	11702	11727	5091	5119	4860	5102
	(1.8)	(2.3)	(1.9)	(2.4)	(4.4)	(3.7)	(8.4)	(4.9)
V10 SLNI and SLPF	9750	11378	10852	10741	5069	5076	4819	5074
	(10.8)	(9.3)	(11.1)	(12.6)	(4.4)	(4.5)	(9.0)	(5.7)

‡No optimization refers to without any spatial optimization. V5 and V10 refers to optimization done using estimated biomass at five and ten leaf-collar stages, respectively. SLNI refers to optimization of total soil nitrogen concentration (SLNI) only. SLNI and SLPF refers to simultaneous optimization of both SLNI and soil fertility factor (SLPF); ¥F1 to F4 represents N-fertilized strips, whereas UF1 to UF4 represents N-unfertilized strips.

Figure 4.1. Map of the study site, which delineates the study area and shows soil and plant sampling points for the 2017 growing season.
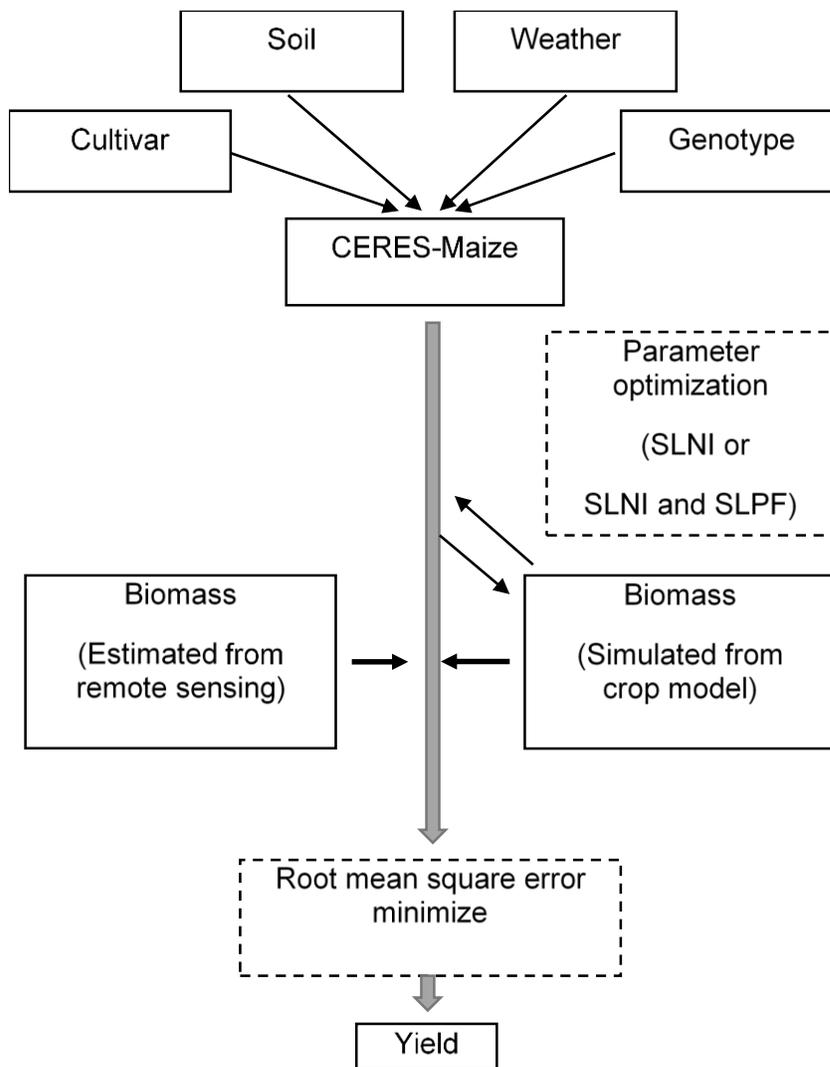
Figure 4.2. Flow chart of using biomass data estimated from satellite imagery into the CERES-Maize model for spatial optimization of total soil nitrogen concentration (SLNI) or SLNI and soil fertility factor (SLPF). (Boxes in solid lines indicate inputs and outputs and boxes in dashed lines indicate optimization process).
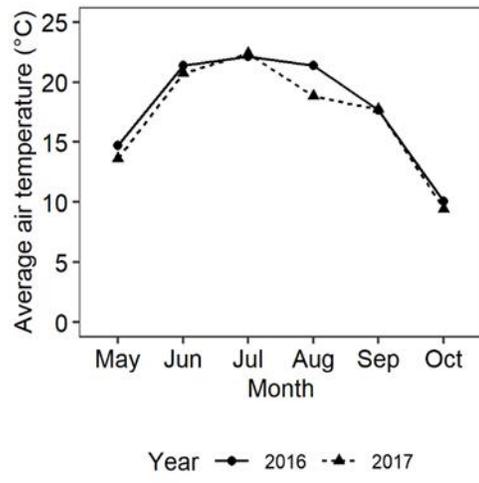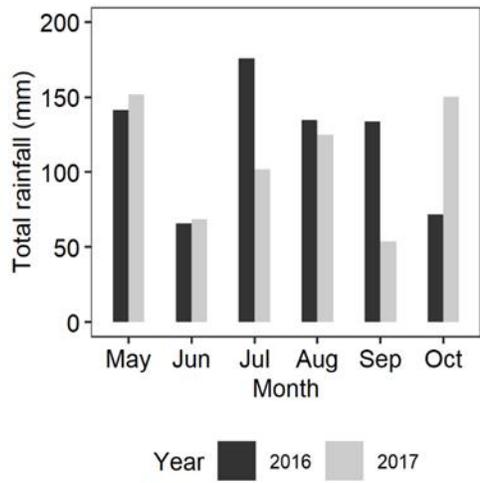
Figure 4.3. Total monthly rainfall and monthly average air temperature during 2016 and 2017 growing seasons at the study site.
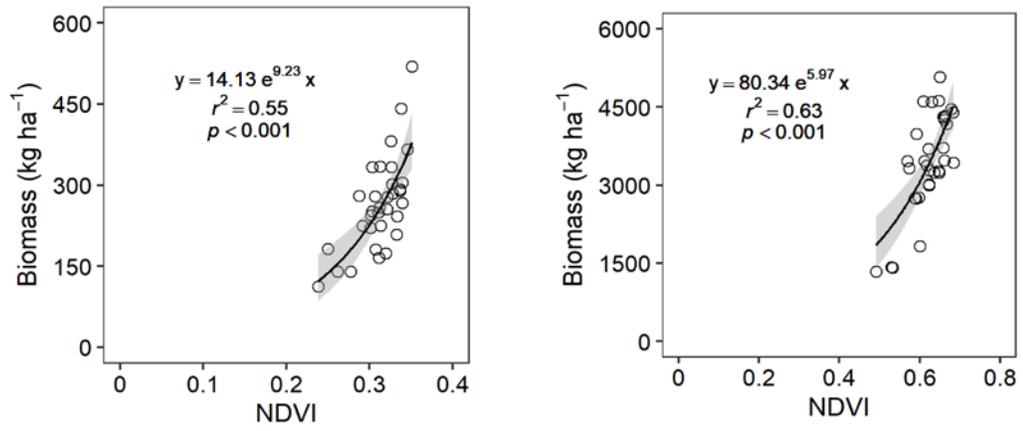
Figure 4.4. Relationship between normalized difference vegetation index (NDVI) and maize biomass at five (left) and ten (right) leaf-collar growth stages.

Figure 4.5. Spatial variability in maize biomass at the five leaf-collar stage estimated using NDVI values from RapidEye satellite imagery.

Figure 4.6. Spatial variability in maize biomass at ten-leaf collar stage estimated using NDVI values from RapidEye satellite image.
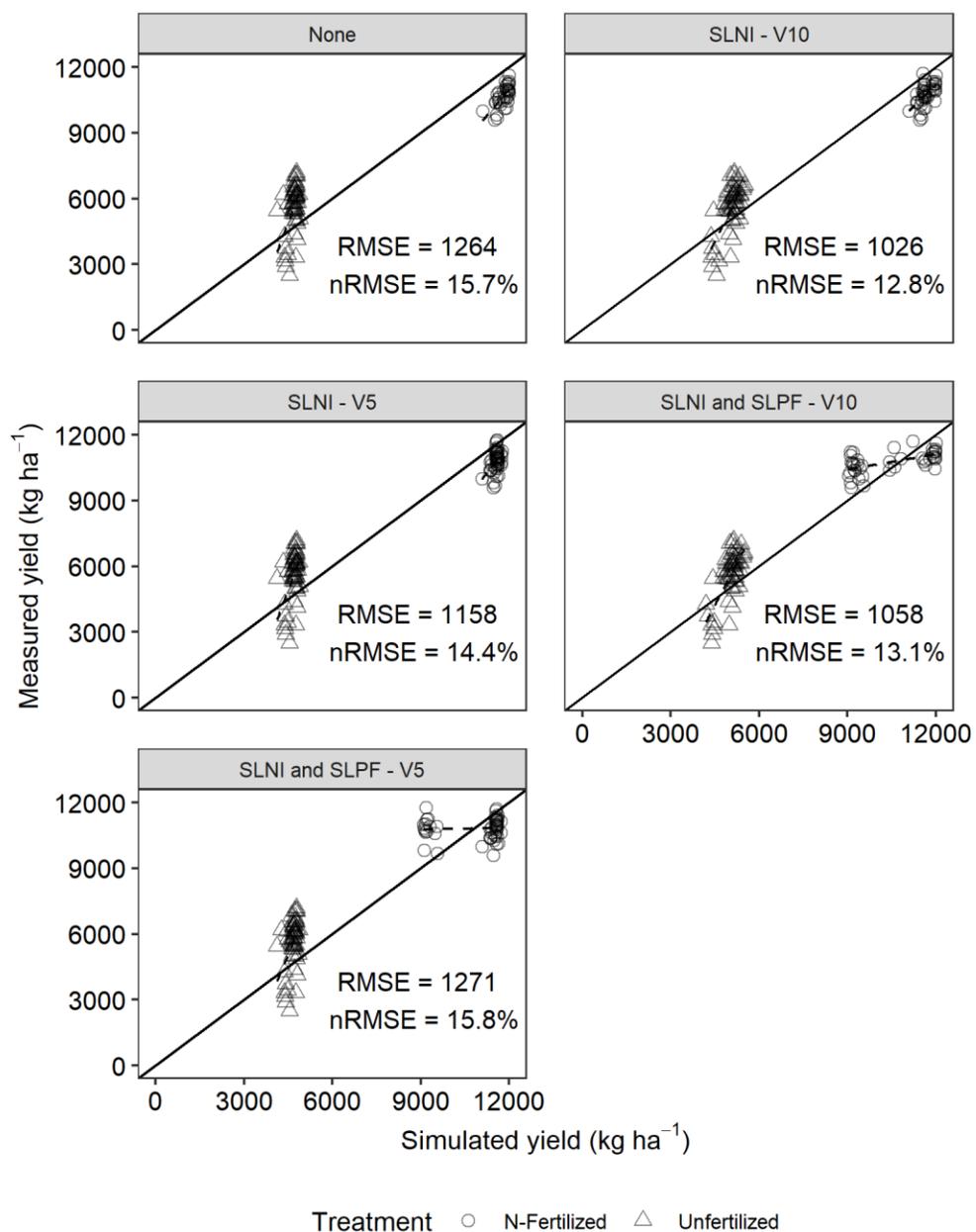
Figure 4.7. Scatterplots of simulated and measured maize yield. The diagonal black line is the 1:1 line. The dashed black line represents the linear regression between simulated and measured yields.
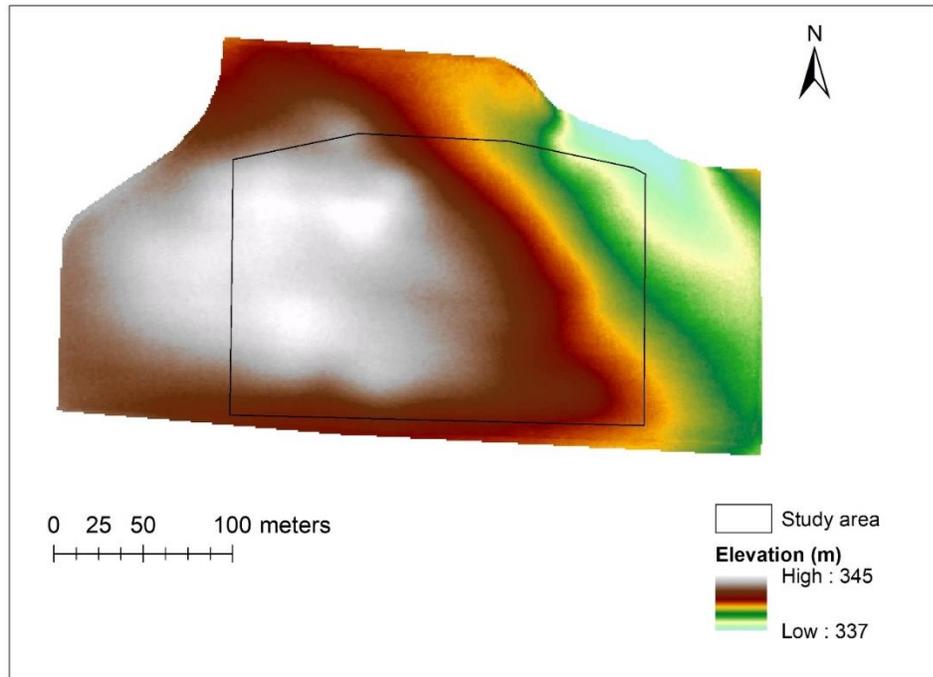
Figure 4.8. Digital elevation map of the study site. (Source: Minnesota Department of Natural Resources and Minnesota Geospatial Information Office, MnTOPO, 2018).

References

Alotaibi, K.D., A.N. Cambouris, M. St. Luce, N. Ziadi, and N. Tremblay. 2018. Economic optimum nitrogen fertilizer rate and residual soil nitrate as influenced by soil texture in corn production. Agron. J. 110(6): 2233–2242. doi: 10.2134/agronj2017.10.0583.

Andraski, T.W., L.G. Bundy, and K.R. Brye. 2000. Crop management and corn nitrogen rate effects on nitrate leaching. J. Environ. Qual. 29(4): 1095–1103. doi: 10.2134/jeq2000.00472425002900040009x.

Andresen, J.A., G. Alagarswamy, C.A. Rotz, J.T. Ritchie, and A.W. LeBaron. 2001. Weather impacts on maize, soybean, and alfalfa production in the Great Lakes region, 1895–1996. Agron. J. 93(5): 1059–1070. doi: 10.2134/agronj2001.9351059x.

Anothai, J., C.M.T. Soler, A. Green, T.J. Trout, and G. Hoogenboom. 2013. Evaluation of two evapotranspiration approaches simulated with the CSM-CERES-Maize model under different irrigation strategies and the impact on maize growth, development and soil moisture content for semi-arid conditions. Agric. For. Meteorol. 176: 64–76. doi: 10.1016/j.agrformet.2013.03.001.

Banger, K., E.D. Nafziger, J. Wang, U. Muhammad, and C.M. Pittelkow. 2018. Simulating nitrogen management impacts on maize production in the U.S. Midwest (P.C. Struik, editor). PLoS One 13(10): e0201825. doi: 10.1371/journal.pone.0201825.

Basso, B., D. Cammarano, and E. Carfagna. 2013. Review of crop yield forecasting methods and early warning systems. In: Report presented to first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural

statistics. FAO, U.N. Rome, Italy, 18 - 19 July.

Basso, B., L. Liu, and J.T. Ritchie. 2016. A Comprehensive Review of the CERES-Wheat, -Maize and -Rice Models' Performances. Advances in Agronomy. Academic Press Inc. p. 27–132

Bhattarai, M.D., S. Secchi, and J. Schoof. 2017. Projecting corn and soybeans yields under climate change in a Corn Belt watershed. Agric. Syst. 152: 90–99. doi: 10.1016/J.AGSY.2016.12.013.

Boote, K.J., J.W. Jones, and N.B. Pickering. 1996. Potential uses and limitations of crop models. Agron. J. 88(1): 704–716.

Borchers, H.W. 2018. Package "pracma". The Comprehensive R Archive Network (CRAN). https://cran.r-project.org/web/packages/pracma/pracma.pdf.

Box, G.E.P., and D.R. Cox. 1964. An analysis of transformations. J. R. Stat. Soc. Ser. B 26(2): 211–243. doi: 10.1111/j.2517-6161.1964.tb00553.x.

Brisson, N., B. Mary, D. Ripoche, M.H. Jeuffroy, F. Ruget, et al. 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. I. Theory and parameterization applied to wheat and corn. Agronomie 18(5–6): 311–346. doi: 10.1051/agro:19980501.

Brumm, T.J., and C.R.J. Hurburgh. 2003. Quality of the 2003 soybean crop in the United States. Agricultural and biosystems engineering technical reports and white papers-19, Iowa State University.

Cai, R., D. Yu, and M. Oppenheimer. 2014. Estimating the spatially varying responses of

corn yields to weather variations using geographically weighted panel regression. J. Agric. Resour. Econ. 39(2): 230–252.

Cambouris, A.N., N. Ziadi, I. Perron, K.D. Alotaibi, M. St. Luce, et al. 2016. Corn yield components response to nitrogen fertilizer as a function of soil texture (X. Yang, editor). Can. J. Soil Sci. 96(4): 386–399. doi: 10.1139/cjss-2015-0134.

Cantarero, M.G., A.G. Cirilo, and F.H. Andrade. 1999. Night temperature at silking affects set in maize. Crop Sci. 39(3): 703–710. doi: 10.2135/cropsci1999.0011183X003900020017x.

Chang, J. 1981. Corn yield in relation to photoperiod, night temperature, and solar radiation. Agric. Meteorol. 24: 253–262. doi: 10.1016/0002-1571(81)90049-2.

Chen, K., R.A. O'Leary, and F.H. Evans. 2019. A simple and parsimonious generalised additive model for predicting wheat yield in a decision support tool. Agric. Syst. 173(2019): 140–150. doi: 10.1016/J.AGSY.2019.02.009.

Chen, H., W. Wu, and H.-B. Liu. 2016. Assessing the relative importance of climate variables to rice yield variation using support vector machines. Theor. Appl. Climatol. 126(1–2): 105–111. doi: 10.1007/s00704-015-1559-y.

Cortes, C., and V. Vapnik. 1995. Support-vector networks. Mach. Learn. 20(3): 273–297. doi: 10.1007/BF00994018.

Crane-Droesch, A. 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. Environ. Res. Lett. 13(11): 114003. doi: 10.1088/1748-9326/aae159.

D'Agostino, A.L., and W. Schlenker. 2016. Recent weather fluctuations and agricultural yields: implications for climate change. Agric. Econ. 47(S1): 159–171. doi: 10.1111/agec.12315.

Dhital, S., and W.R. Raun. 2016. Variability in Optimum Nitrogen Rates for Maize. Agron. J. 108(6): 2165. doi: 10.2134/agronj2016.03.0139.

Dinterman, R., and J. Eyer. 2018. nassR: An alternative for downloading various USDA data from Quick Stats through R. https://rdrr.io/github/rdinter/nassR/.

Elliott, J., M. Glotter, A.C. Ruane, K.J. Boote, J.L. Hatfield, et al. 2018. Characterizing agricultural impacts of recent large-scale US droughts and changing technology and management. Agric. Syst. 159: 275–281. doi: 10.1016/J.AGSY.2017.07.012.

Ewing, P.M., and B.C. Runck. 2015. Optimizing nitrogen rates in the midwestern United States for maximum ecosystem value. Ecol. Soc. 20(1): 18–28. doi: 10.5751/ES-06767-200118.

Fang, Q.X., L. Ma, T.J. Trout, L.H. Comas, K.C. DeJonge, et al. 2017. Modeling N concentration and uptake for maize hybrids under growth stage-based deficit irrigations. Trans. ASABE 60(6): 2067–2081. doi: https://doi.org/10.13031/trans.12405.

Fraisse, C.W., K.A. Sudduth, and N.R. Kitchen. 2001. Calibration of the CERES-Maize model for simulating site-specific crop development and yield on claypan soils. Appl. Eng. Agric. 17(4): 547–556. doi: 10.13031/2013.6453.

Franch, B., E.F. Vermote, S. Skakun, J.C. Roger, I. Becker-Reshef, et al. 2019. Remote sensing based yield monitoring: Application to winter wheat in United States and

Ukraine. Int. J. Appl. Earth Obs. Geoinf. 76: 112–127. doi:
10.1016/J.JAG.2018.11.012.

Franzluebbers, A., J. Steiner, D. Karlen, T. Griffin, J. Singer, et al. 2011. Rainfed
Farming Systems in the USA. Rainfed Farming Systems. Springer Netherlands,
Dordrecht. p. 511–560

Green, T.R., H. Kipka, O. David, and G.S. McMaster. 2018. Where is the USA Corn Belt,
and how is it changing? Sci. Total Environ. 618: 1613–1618. doi:
10.1016/J.SCITOTENV.2017.09.325.

Hastie, T.J., and R.J. Tibshirani. 1990. Generalized additive models. Monographs on
statistics and applied Probability. 1st ed. Chapman and Hall/CRC.

Hoogenboom, G., C.H. Porter, V. Shelia, K.J. Boote, U. Singh, et al. 2019. Decision
Support System for Agrotechnology Transfer (DSSAT) Version 4.7.5. DSSAT
Foundation, Gainesville, Florida, USA.

Horie, T., M. Yajima, and H. Nakagawa. 1992. Yield forecasting. Agric. Syst. 40(1–3):
211–236. doi: 10.1016/0308-521X(92)90022-G.

Hu, Q., G. Buyanovsky, Q. Hu, and G. Buyanovsky. 2003. Climate effects on corn yield
in Missouri. J. Appl. Meteorol. 42(11): 1626–1635. doi: 10.1175/1520-
0450(2003)042<1626:CEOCYI>2.0.CO;2.

Huete, A., K. Didan, T. Miura, E.. Rodriguez, X. Gao, et al. 2002. Overview of the
radiometric and biophysical performance of the MODIS vegetation indices. Remote
Sens. Environ. 83(1–2): 195–213. doi: 10.1016/S0034-4257(02)00096-2.

Jame, Y.W., and H.W. Cutforth. 1996. Crop growth models for decision support systems. Can. J. Plant Sci. 76(1): 9–19. doi: 10.4141/cjps96-003.

Jeong, H., and R. Bhattarai. 2018. Exploring the effects of nitrogen fertilization management alternatives on nitrate loss and crop yields in tile-drained fields in Illinois. J. Environ. Manage. 213: 341–352. doi: 10.1016/J.JENVMAN.2018.02.062.

Jeong, J.H., J.P. Resop, N.D. Mueller, D.H. Fleisher, K. Yun, et al. 2016. Random forests for global and regional crop yield predictions (J.L. Gonzalez-Andujar, editor). PLoS One 11(6): e0156571. doi: 10.1371/journal.pone.0156571.

Jing, Q., J. Shang, T. Huffman, B. Qian, E. Pattey, et al. 2017. Using the CSM-CERES-Maize model to assess the gap between actual and potential yields of grain maize. J. Agric. Sci. 155(2): 239–260. doi: 10.1017/S0021859616000290.

Jones, J.W. 1993. Decision support systems for agricultural development. Springer, Dordrecht. p. 459–471

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, et al. 2003. The DSSAT cropping system model. Eur. J. Agron. 18(3–4): 235–265. doi: 10.1016/S1161-0301(02)00107-7.

Jones, C.A., and J.R. Kiniry, editors. 1986. CERES-Maize: A simulation model of maize growth and development. Texas A&M University Press, College Station, Texas, U.S.

Kaiser, D., F. Fernandez, and J.A. Coulter. 2016. Fertilizing corn in Minnesota. https://extension.umn.edu/crop-specific-needs/fertilizing-corn-minnesota (accessed 10 October 2018).

Karatzoglou, A., A. Smola, and K. Hornik. 2018. kernlab (v 0.9-27)-An S4 package for kernel methods in R. https://cran.r-project.org/web/packages/kernlab/vignettes/kernlab.pdf.

Karimi, Y., S.O. Prasher, A. Madani, S. Kim, S.O. Madani, et al. 2008. Application of support vector machine technology for the estimation of crop biophysical parameters using aerial hyperspectral observations. Can. Biosyst. Eng. 50(13): 1–8. http://www.csbe-scgab.ca/docs/journal/50/c0715.pdf (accessed 6 April 2019).

Kaul, M., R.L. Hill, and C. Walthall. 2005. Artificial neural networks for corn and soybean yield prediction. Agric. Syst. 85: 1–18.

Keating, B.., P.. Carberry, G.. Hammer, M.. Probert, M.. Robertson, et al. 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18(3–4): 267–288. doi: 10.1016/S1161-0301(02)00108-9.

Kriegler, F.J., W.A. Malia, R.F. Nalepka, and W. Richardson. 1969. Preprocessing transformations and their effects on multispectral recognition. Proceedings of the sixth international symposium on remote sensing of environment, Ann Arbor, MI, USA, 13–16 October. p. 97–131

Kuhn, M. 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28(5): 1–26.

Leng, G., X. Zhang, M. Huang, G.R. Asrar, and L.R. Leung. 2016. The role of climate covariability on crop yields in the conterminous United States. Sci. Rep. 6(1): 1–11. doi: 10.1038/srep33160.

Lewis, S.C., D.J. Karoly, S.C. Lewis, and D.J. Karoly. 2013. Evaluation of historical

diurnal temperature range trends in CMIP5 models. J. Clim. 26(22): 9077–9089. doi: 10.1175/JCLI-D-13-00032.1.

Li, Z.T., J.Y. Yang, C.F. Drury, and G. Hoogenboom. 2015. Evaluation of the DSSAT-CSM for simulating yield and soil organic C and N of a long-term maize and wheat rotation experiment in the Loess Plateau of Northwestern China. Agric. Syst. 135: 90–104. doi: 10.1016/j.agsy.2014.12.006.

Liu, S., J.Y. Yang, C.F. Drury, H.L. Liu, and W.D. Reynolds. 2014. Simulating maize (Zea mays L.) growth and yield, soil nitrogen concentration, and soil water content for a long-term cropping experiment in Ontario, Canada. Can. J. Soil Sci. 94(3): 435–452. doi: 10.4141/CJSS2013-096.

Liu, H. long, J. yi Yang, P. He, Y. lu Bai, J. yun Jin, et al. 2012. Optimizing Parameters of CSM-CERES-Maize Model to Improve Simulation Performance of Maize Growth and Nitrogen Uptake in Northeast China. J. Integr. Agric. 11(11): 1898–1913. doi: 10.1016/S2095-3119(12)60196-8.

Lobell, D.B. 2007. Changes in diurnal temperature range and national cereal yields. Agric. For. Meteorol. 145(3–4): 229–238. doi: 10.1016/J.AGRFORMET.2007.05.002.

Lobell, D.B., and S. Asseng. 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environ. Res. Lett. 12(1): 015001. doi: 10.1088/1748-9326/aa518a.

Lobell, D.B., M. Bänziger, C. Magorokosho, and B. Vivek. 2011. Nonlinear heat effects on African maize as evidenced by historical yield trials. Nat. Clim. Chang. 1(1): 42–

45. doi: 10.1038/nclimate1043.

Lobell, D.B., and M.B. Burke. 2010. On the use of statistical models to predict crop yield responses to climate change. Agric. For. Meteorol. 150(11): 1443–1452. doi: 10.1016/J.AGRFORMET.2010.07.008.

Lobell, D.B., K.N. Cahill, and C.B. Field. 2006. Weather-based yield forecasts developed for 12 California crops. Calif. Agric. 60(4): 211–215. doi: 10.3733/ca.v060n04p211.

Lobell, D.B., and C.B. Field. 2007. Global scale climate–crop yield relationships and the impacts of recent warming. Environ. Res. Lett. 2(1): 014002. doi: 10.1088/1748-9326/2/1/014002.

Lobell, D.B., M.J. Roberts, W. Schlenker, N. Braun, B.B. Little, et al. 2014. Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. Science 344(6183): 516–519. doi: 10.1126/science.1251423.

Lu, J., G.J. Carbone, and P. Gao. 2017. Detrending crop yield data for spatial visualization of drought impacts in the United States, 1895–2014. Agric. For. Meteorol. 237–238: 196–208. doi: 10.1016/J.AGRFORMET.2017.02.001.

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. Agron. J. 95(4): 958–964. doi: 10.2134/AGRONJ2003.9580.

Mathieu, J.A., and F. Aires. 2016. Statistical weather-impact models: an application of neural networks and mixed effects for corn production over the United States. J. Appl. Meteorol. Climatol. 55(11): 2509–2527. doi: 10.1175/JAMC-D-16-0055.1.

Mathieu, J.A., and F. Aires. 2018. Assessment of the agro-climatic indices to improve crop yield forecasting. Agric. For. Meteorol. 253–254: 15–30. doi: 10.1016/J.AGRFORMET.2018.01.031.

MEQB. 2015. Beyond the status quo: 2015 EQB water policy report. Minnesota Environmental Quality Board, MN.

Miao, Y., D.J. Mulla, W.D. Batchelor, J.O. Paz, P.C. Robert, et al. 2006. Evaluating management zone optimal nitrogen rates with a crop growth model. Agron. J. 98(3): 545–553. doi: 10.2134/agronj2005.0153.

Morell, F.J., H.S. Yang, K.G. Cassman, J. Van Wart, R.W. Elmore, et al. 2016. Can crop simulation models be used to predict local to regional maize yields and total production in the U.S. Corn Belt? F. Crop. Res. 192: 1–12. doi: 10.1016/J.FCR.2016.04.004.

Mourtzinis, S., J.E. Specht, L.E. Lindsey, W.J. Wiebold, J. Ross, et al. 2015. Climate-induced reduction in US-wide soybean yields underpinned by region- and in-season-specific responses. Nat. Plants 1(2): 1–4. doi: 10.1038/nplants.2014.26.

NAWQA/USGS. 2014. Nutrient delivery to the Gulf of Mexico. Natl. Water Qual. Assess. Program, United States Geol. Surv. https://water.usgs.gov/nawqa/ (accessed 26 August 2018).

O'Neal, M.R., J.R. Frankenberger, and D.R. Ess. 2002. Use of CERES-Maize to study effect of spatial precipitation variability on yield. Agric. Syst. 73(2): 205–225. doi: 10.1016/S0308-521X(01)00095-6.

Oguntunde, P.G., G. Lischeid, and O. Dietrich. 2018. Relationship between rice yield

and climate variables in southwest Nigeria using multiple linear regression and support vector machine analysis. Int. J. Biometeorol. 62(3): 459–469. doi: 10.1007/s00484-017-1454-6.

Pang, X.P., S.C. Gupta, J.F. Moncrief, C.J. Rosen, and H.H. Cheng. 1998. Evaluation of nitrate leaching potential in Minnesota glacial outwash soils using the CERES-Maize model. J. Environ. Qual. 27(1): 75–85. doi: 10.2134/jeq1998.00472425002700010012x.

Peng, B., K. Guan, M. Pan, and Y. Li. 2018. Benefits of seasonal climate prediction and satellite data for forecasting U.S. maize yield. Geophys. Res. Lett. 45(18): 9662–9671. doi: 10.1029/2018GL079291.

Persson, T., A. Garcia y Garcia, J. Paz, J. Jones, and G. Hoogenboom. 2009. Maize ethanol feedstock production and net energy value as affected by climate variability and crop management practices. Agric. Syst. 100(1–3): 11–21. doi: 10.1016/j.agsy.2008.11.004.

Peters, D.B., J.W. Pendleton, R.H. Hageman, and C.M. Brown. 1971. Effect of night air temperature on grain yield of corn, wheat, and soybeans. Agron. J. 63(5): 809–809. doi: 10.2134/agronj1971.00021962006300050046x.

Phillips, S. 1994. The Soil Conservation Service responds to the 1993 Midwest floods. Historical notes number 4, Enonomic and Social Sciences Division, United States Department of Agriculture.

Plénet, D., and G. Lemaire. 2000. Relationships between dynamics of nitrogen uptake and dry matter accumulation in maize crops. Determination of critical N

concentration. Plant Soil 216(1–2): 65–82. doi: 10.1023/A:1004783431055.

Powell, J.P., and S. Reinhard. 2016. Measuring the effects of extreme weather events on yields. Weather Clim. Extrem. 12: 69–79. doi: 10.1016/J.WACE.2016.02.003.

Prasad, A.K., L. Chai, R.P. Singh, and M. Kafatos. 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8(1): 26–33. doi: 10.1016/J.JAG.2005.06.002.

PRISM Climate Group. 2018. PRISM gridded climate data, Parameter-elevation Regression on Independent Slopes Model (PRISM) Climate Group, Oregon State University. Prism. http://www.prism.oregonstate.edu/ (accessed 12 January 2018).

Ratjen, A.M., G. Lemaire, H. Kage, D. Plénet, and E. Justes. 2018. Key variables for simulating leaf area and N status: Biomass based relations versus phenology driven approaches. Eur. J. Agron. 100: 110–117. doi: 10.1016/j.eja.2018.04.008.

Ray, D.K., J.S. Gerber, G.K. MacDonald, and P.C. West. 2015. Climate variation explains a third of global crop yield variability. Nat. Commun. 6(5989): 1–9. doi: 10.1038/ncomms6989.

Roberts, M.J., N.O. Braun, T.R. Sinclair, D.B. Lobell, and W. Schlenker. 2017. Comparing and combining process-based crop models and statistical models with some implications for climate change. Environ. Res. Lett. 12(9). doi: 10.1088/1748-9326/aa7f33.

Rouse, J.W.J., R.H. Hass, J.A. Schell, and D.W. Deering. 1974. Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations. NASA SP-351.

Washington D.C. p. 309–317

Scharf, P.C., N.R. Kitchen, K.A. Sudduth, J.G. Davis, V.C. Hubbard, et al. 2005. Field-scale variability in optimal nitrogen fertilizer rate for corn. Agron. J. 97(2): 452–461. doi: 10.2134/agronj2005.0452.

Schlenker, W., and M.J. Roberts. 2006. Nonlinear effects of weather on corn yields. Rev. Agric. Econ. 28(3): 391–398. doi: 10.1111/j.1467-9353.2006.00304.x.

Schlenker, W., and M.J. Roberts. 2009. Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. Proc. Natl. Acad. Sci. U. S. A. 106(37): 15594–15598. doi: 10.1073/pnas.0906865106.

Sela, S., H.M. van Es, B.N. Moebius-Clune, R. Marjerison, J. Melkonian, et al. 2016. Adapt-N outperforms grower-selected nitrogen rates in northeast and midwestern United States strip trials. Agron. J. 108(4): 1726–1734. doi: 10.2134/agronj2015.0606.

Setiyono, T.D., H. Yang, D.T. Walters, A. Dobermann, R.B. Ferguson, et al. 2011. Maize-N: a decision tool for nitrogen management in maize. Agron. J. 103(4): 1276–1283. doi: 10.2134/AGRONJ2011.0053.

Shcherbak, I., N. Millar, and G.P. Robertson. 2014. Global metaanalysis of the nonlinear response of soil nitrous oxide (N2O) emissions to fertilizer nitrogen. Proc. Natl. Acad. Sci. U. S. A. 111(25): 199–204. doi: 10.1073/pnas.1322434111.

Shi, W., F. Tao, and Z. Zhang. 2013. A review on statistical models for identifying climate contributions to crop yields. J. Geogr. Sci. 23(3): 567–576. doi: 10.1007/s11442-013-1029-3.

Sibley, A.M., P. Grassini, N.E. Thomas, K.G. Cassman, and D.B. Lobell*. 2014. Testing Remote Sensing Approaches for Assessing Yield Variability among Maize Fields. Agron. J. 106(1): 24. doi: 10.2134/agronj2013.0314.

Sinclair, T.R., and T. Horie. 1989. Leaf nitrogen, photosynthesis, and crop radiation use efficiency: a review. Crop Sci. 29(1): 90–98. doi: 10.2135/cropsci1989.0011183X002900010023x.

Sinclair, T.R., and N.G. Seligman. 1996. Crop modeling: From infancy to maturity. Agron. J. 88(5): 698–704. doi: 10.2134/agronj1996.00021962008800050004x.

Soil Survey Staff. 2018. Web Soil Survey. Nat. Resour. Conserv. Serv. United States Dep. Agric. Web Soil Surv. https://websoilsurvey.sc.egov.usda.gov/ (accessed 21 August 2018).

SROC/UMN. 2018. Historic weather reports. South. Res. outreach center, Univ. Minnesota, Waseca, MN. https://sroc.cfans.umn.edu/weather-sroc/historic-reports (accessed 28 October 2018).

SWROC/UMN. 2018. Historic weather reports. Southwest Res. outreach center, Univ. Minnesota, Lamberton, MN. https://swroc.cfans.umn.edu/weather/historic-reports (accessed 27 October 2018).

Tack, J., A. Barkley, and L.L. Nalley. 2015. Effect of warming temperatures on US wheat yields. Proc. Natl. Acad. Sci. U. S. A. 112(22): 6931–6936. doi: 10.1073/pnas.1415181112.

Teasdale, J.R., and M.A. Cavigelli. 2017. Meteorological fluctuations define long-term crop yield patterns in conventional and organic production systems. Sci. Rep.

7(688): 1–10. doi: 10.1038/s41598-017-00775-8.

Thomasson, J.A., J.R. Wooten, S. Gogineni, R. Sui, and B.M. Kolla. 2004. Remote sensing and weather information in cotton yield prediction. Ecosystems' dynamics, agricultural remote sensing and modeling, and site-specific agriculture. International Society for Optics and Photonics, San Diego, California, United States. p. 127–136

Thompson, L.M. 1969. Weather and technology in the production of corn in the U. S. Corn Belt. Agron. J. 61(3): 453–456. doi: 10.2134/agronj1969.00021962006100030037x.

Thorp, K.R., R.W. Malone, and D.B. Jaynes. 2007. Simulating long-term effects of nitrogen fertilizer application rates on corn yield and nitrogen dynamics. Trans. ASABE 50(4): 1287–1303. https://pubag.nal.usda.gov/pubag/downloadPDF.xhtml?id=12862&content=PDF.

Tremblay, N., Y.M. Bouroubi, C. Bélec, R.W. Mullen, N.R. Kitchen, et al. 2012. Corn response to nitrogen is influenced by soil texture and weather. Agron. J. 104(6): 1658–1671. doi: 10.2134/agronj2012.0184.

USDA-NASS. 2012. The yield forecasting program of NASS. SMB staff report number SMB 12-01, Statistical Methods Branch, National Agricultural Statistics Service, United States Department of Agriculture.

USDA-NASS. 2013. Crop production 2012 annual summary. National Agriculture and Statistics Service, United States Department of Agriculture.

USDA-NASS. 2018. Statistics by State. National Agricultural Statistics Service, United States Department of Agriculture.

Vapnik, V.N. 1999. An overview of statistical learning theory. IEEE Trans. neural

    networks 10(5): 988–999.

    http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.332.356&rep=rep1&type=

    pdf (accessed 3 April 2019).

van der Velde, M., I. Biavetti, M. El-Aydam, S. Niemeyer, F. Santini, et al. 2019. Use and

    relevance of European Union crop monitoring and yield forecasts. Agric. Syst.

    168(1): 224–230. doi: 10.1016/J.AGSY.2018.05.001.

Venterea, R.T., J.A. Coulter, and M.S. Dolan. 2016. Evaluation of Intensive "4R"

    Strategies for Decreasing Nitrous Oxide Emissions and Nitrogen Surplus in Rainfed

    Corn. J. Environ. Qual. 45(4): 1186. doi: 10.2134/jeq2016.01.0024.

Verón, S.R., D. de Abelleyra, and D.B. Lobell. 2015. Impacts of precipitation and

    temperature on crop yields in the Pampas. Clim. Change 130(2): 235–245. doi:

    10.1007/s10584-015-1350-1.

Vose, R.S., D.R. Easterling, and B. Gleason. 2005. Maximum and minimum temperature

    trends for the globe: An update through 2004. Geophys. Res. Lett. 32(23): L23822.

    doi: 10.1029/2005GL024379.

van Wart, J., K.C. Kersebaum, S. Peng, and M. Milner. 2013. Estimating crop yield

    potential at regional to national scales. F. Crop. Res. 143: 34–43. doi:

    10.1016/J.FCR.2012.11.018.

Westcott, P.C., and M. Jewison. 2013. Weather effects on expected corn and soybean

    yields. FDS-13g-01. A report from the Economic Research Service, United States

    Department of Agriculture.

Wood, S.N. 2017. Generalized additive models : an introduction with R. 2nd ed. Chapman and Hall/CRC.

Wood, S.N. 2018. Mixed GAM computation vehicle with automatic smoothness estimation. Package "mgcv" v. 1.8-26. https://cran.r-project.org/web/packages/mgcv/mgcv.pdf.

Yakoub, A., J. Lloveras, A. Biau, J.L. Lindquist, and J.I. Lizaso. 2017. Testing and improving the maize models in DSSAT: Development, growth, yield, and N uptake. F. Crop. Res. 212: 95–106. doi: 10.1016/j.fcr.2017.07.002.

Yang, H.S., A. Dobermann, J.L. Lindquist, D.T. Walters, T.J. Arkebauer, et al. 2004. Hybrid-maize - A maize simulation model that combines two crop modeling approaches. F. Crop. Res. 87(2–3): 131–154. doi: 10.1016/j.fcr.2003.10.003.

Yang, Z., G.G. Wilkerson, G.S. Buol, D.T. Bowman, and R.W. Heiniger. 2009. Estimating genetic coefficients for the CSM-CERES-Maize model in North Carolina environments. Agron. J. 101(5): 1276–1285. doi: 10.2134/agronj2008.0234x.