Integration of remote sensing and crop growth modeling for nitrogen management decision support in corn

Kelly Robert Thorp
Iowa State University
Integration of remote sensing and crop growth modeling for nitrogen management decision support in corn

by

Kelly Robert Thorp

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Agricultural Engineering

Program of Study Committee:
William D. Batchelor, Co-major Professor
Brian L. Steward, Co-major Professor
Amy L. Kaleita
Matthew J. Helmers
Petrutza C. Caragea

Iowa State University
Ames, Iowa

2006

Copyright © Kelly Robert Thorp, 2006. All rights reserved.
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Kelly Robert Thorp

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

Co-major Professor

Signature was redacted for privacy.

For the Major Program
For my parents and my brother
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>vii</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
</tbody>
</table>

## CHAPTER 1. GENERAL INTRODUCTION .............................................. 1

1.1. Introduction ......................................................... 1

1.2. Dissertation Overview ............................................ 6

1.3. References .......................................................... 8

## CHAPTER 2. USING AERIAL HYPERSONTRAL REMOTE SENSING IMAGERY TO ESTIMATE CORN PLANT STAND DENSIY  .......... 12

2.1. Abstract ................................................................. 12

2.2. Introduction .......................................................... 12

2.3. Materials and Methods ............................................. 15

   2.3.1. Data Collection ................................................ 15

   2.3.2. Ground-Reference Data Processing ............................. 20

   2.3.3. Hyperspectral Data Processing ................................ 21

   2.3.4. Statistical Analysis .......................................... 21

2.4. Results and Discussion ............................................. 23

2.5. Conclusions ............................................................ 30

2.6. Acknowledgements .................................................... 31

2.7. References ............................................................. 32

## CHAPTER 3. STATISTICAL PROCEDURES FOR VALIDATING CERES-MAIZE SIMULATIONS OF SPATIAL CORN YIELD VARIABILITY ...... 36

3.1. Abstract ................................................................. 36

3.2. Introduction ............................................................ 36

3.3. Materials and Methods .............................................. 40
CHAPTER 4. METHODOLOGY TO LINK PRODUCTION AND ENVIRONMENTAL RISKS OF PRECISION NITROGEN MANAGEMENT STRATEGIES IN CORN

4.1. Abstract

4.2. Introduction

4.3. Methods

4.3.1. Data Preparation

4.3.2. Model Calibration

4.3.3. Nitrogen Prescription Analysis

4.3.4. Cumulative Probability Distributions

4.3.5. Nitrogen Management Decisions

4.4. Results

4.4.1. Calibration Results

4.4.2. Grid Cell #4 Example

4.4.3. Statistics Over 100 Grid Cells Per N Rate

4.5. Discussion

4.5.1. Environmental Cost of Optimizing Economic Return

4.5.2. Opportunity Cost of Environmental Protection

4.6. Conclusions

4.7. References
CHAPTER 5. GENERAL CONCLUSIONS ......................................................... 96
5.1. Conclusions .................................................................................. 96
5.2. Recommendations ......................................................................... 98

ACKNOWLEDGEMENTS ......................................................................... 102
BIOGRAPHICAL SKETCH ................................................................. 105
VITA .................................................................................................. 106
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Framework for making management decision in precision agriculture</td>
<td>6</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Study area for remote sensing of corn plant stand density</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Ground-based corn plant population sensing system</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Measured versus predicted plant stand density for the top case</td>
<td>25</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Histograms of corn plant stand density in each plot</td>
<td>26</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Rank plot of top $R^2$'s for each case</td>
<td>27</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Histograms of times wavelengths were used as a predictor variable</td>
<td>28</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Dominant soil types for grid cells at the study site</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Spatial variation of RMSEP across the study site</td>
<td>50</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Spatial variation of mean effective tile drainage rate across the study site</td>
<td>51</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Spatial variation of mean saturated hydraulic conductivity across the study site</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Ninety-five percent confidence ellipse for fitted parameters in grid cell #13</td>
<td>54</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Spatial variation of ninety-five percent confidence ellipse area</td>
<td>55</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Dominant soil types for grid cells at the study site</td>
<td>66</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Spatial variation of RMSE for the model calibration</td>
<td>75</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Simulated versus measured corn yield for the model calibration</td>
<td>76</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>Grid cell #4 histogram for yield over 37 years of historical weather</td>
<td>79</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>Cumulative beta probability of yield in grid cell #4</td>
<td>80</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>Grid cell #4 histogram for unused nitrogen over 37 years of historical weather</td>
<td>81</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>Cumulative exponential probability of unused nitrogen in grid cell #4</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>Plot of marginal net return for grid cell #4</td>
<td>85</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Nitrogen prescription map for optimizing marginal net return</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>Nitrogen prescription map for achieving the environmental objective</td>
<td>89</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>Opportunity cost for managing nitrogen to achieve the environmental objective</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>Opportunity cost in grid cell #4 for a range of unused nitrogen restrictions</td>
<td>92</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 2.1. Summary of important characteristics for the three plots ........................................... 17
Table 2.2. Corn growth stage on each remote sensing image collection date ......................... 18
Table 2.3. The highest $R^2$ obtained for multiple linear regressions in each case ............... 24
Table 2.4. Summary of cases used for spectral evaluation ......................................................... 27
Table 3.1. Comparison of RMSE and validation errors against total rainfall ....................... 48
Table 3.2. Simulation results aggregated by dominant soil type ............................................... 56
Table 4.1. Field-level measured versus predicted yield relationships ....................................... 77
Table 4.2. Summary statistics of distributional parameters over 100 grid cells ..................... 83
Table 4.3. Field-level statistics for the economically optimum nitrogen prescription ............ 87
Table 4.4. Field-level statistics for the environmental nitrogen prescription ......................... 90
ABSTRACT

This dissertation describes efforts to move toward a completely integrated remote sensing and crop growth modeling tool for developing precision nitrogen management recommendations for corn. Aerial hyperspectral remote sensing imagery collected throughout the 2004 growing season was used to estimate corn plant stand density, and a machine vision system was used to map corn population on the ground. Multiple linear regression analysis was used to assess the ability of all combinations of three reflectance bands to estimate corn plant population at resolutions of 2 m, 6 m, and 10 m. Coefficients of multiple determination of up to 0.82 were achieved in this endeavor. Although some limitations apply, remote sensing can be used as a tool to provide corn plant population inputs for crop growth simulations. A cross validation technique and bivariate confidence ellipses were used to evaluate CERES Maize simulations of spatial corn yield variability across an Iowa cornfield. Results indicated that the model performed most poorly when using the wettest or driest growing seasons to validate the model, because the model parameters fitted under the conditions of moderate growing seasons were less flexible for simulating yield in growing seasons with more extreme weather. Results also indicated that topography affects the model performance spatially. CERES-Maize was also used to simulate yield and unused nitrogen remaining in the soil at harvest for a sequence of historical weather data. Simulations were run for 13 spring-applied nitrogen rates over a cornfield divided into 100 0.2 ha grid cells. A methodology based on cumulative probability distributions was then developed to use model output for assessing the link between yield and nitrogen left behind for various nitrogen rates in each grid cell. This methodology can be used to develop precision nitrogen management strategies that address both the economic and environmental concerns of nitrogen management practices. Although the three projects in this dissertation furthered the development of remote sensing, crop growth modeling, and decision support technologies, more work is required to obtain a completely integrated tool for development of precision nitrogen management strategies in midwestern cornfields.
CHAPTER 1. GENERAL INTRODUCTION

1.1 Introduction

Precision agriculture began in the mid 1980s when new technologies were first used to vary fertilizer rates and blends across agricultural fields (Robert, 2002). At this time, a key breakthrough was the realization that managing agricultural fields according to their soil and crop growth spatial variability, instead of uniformly, could potentially offer greater benefit in terms of profitability and environmental stewardship. Interest in precision agriculture grew throughout the 1990s as microprocessors, geographic information systems (GIS), global positioning systems (GPS), automatic control systems, and sensors were rapidly developed (Zhang et al., 2002). Together these technological developments offered the ability to detect spatial variability across agricultural fields, store the information, analyze the data to obtain site-specific management plans, and administer agricultural inputs site-specifically across fields. However, since the turn of the century, a number of issues have slowed progress toward the end goal, which is widespread adoption of precision agriculture on commercial farms. Robert (2002) divides these issues into three categories: socio-economical, agronomical, and technological. The socio-economical challenges of precision agriculture include the costs and skill required for adoption of the technology. Many producers find the cost of adoption to be too high, and many do not have and are unwilling to obtain the skills necessary to utilize the new technologies on their farms. Agronomical barriers include lack of information, misuse or misinterpretation of information, inadequate sampling and scouting techniques, lack of qualified agroconsultants, and lack of a solid agronomic basis for site-specific recommendations (Pierce and Nowak, 1999). Technological challenges exist in the area of application equipment, sensors for detection of variability in crop growth and soil conditions, global positioning systems and services, software for acquisition, processing, interpretation, and storing of crop and soil information, and the logistics of utilizing remote sensing imagery. These issues must be addressed before precision agriculture will be adopted by a majority of crop producers.

As a product of the Agricultural and Biosystems Engineering Department at Iowa State University, this dissertation addresses several of the technological challenges associated
with precision agriculture. Specifically, the dissertation describes efforts to unite remote sensing and crop simulation modeling within the context of precision agriculture, a direction supported by Moran et al. (1997). The end goal of this work was to move towards a complete crop sensing, crop growth modeling, and decision support system for development of nitrogen prescription maps that address both the economic and environmental concerns of nitrogen fertilization practices in midwestern cornfields.

Some researchers have proposed the use of remote sensing technology for direct assessment of crop nitrogen needs. Blackmer et al. (1996) digitized aerial color photographs over plots of irrigated corn that were fertilized with varying nitrogen rates, and they measured the final yield in each plot. They found high correlations between yield and raw digital numbers in the red, green, and blue regions, and they concluded that digital numbers could be used to detect yield reductions due to nitrogen stress. Sripada et al. (2005) performed a similar experiment in corn with various nitrogen rates applied both at planting and at the pretassel stage, and grain yields were shown to respond positively to applications of nitrogen at both times. The green difference vegetation index was then used to compute economically optimal nitrogen rates at the pretassel stage relative to the performance of high nitrogen reference strips. Results of Scharf and Lory (2002) also support the idea that a reference strip at a high nitrogen rate is needed to predict sidedress nitrogen rates for corn from digitized aerial photographs. Remote sensing has also been used to estimate wheat tiller density at growth stage 25, and this crop growth parameter has been used as a decision aid for early nitrogen applications in wheat (Flowers et al., 2001; Flowers et al., 2003). These studies have shown how spectral reflectance measurements can be used to identify nitrogen deficient areas of crops; however, these techniques are useful only if an area of crops known to be non-deficient in nitrogen is available for relative comparison. In these cases, levels of nitrogen stress are estimated merely through the spatial variability of reflectance information across the field, and absolutely no knowledge of the pathways and means by which nitrogen flows through the agricultural system is used to understand the cause of plant nitrogen stress. In addition, these researchers claim that remote sensing can be used to reduce losses of excess nitrogen to the environment, but their methodology includes no procedure for understanding how nitrogen moves through the soil matrix and out of the
agriculture system and also how this movement varies spatially and temporally. Because nitrogen movement depends spatially on soil properties and temporally on weather patterns, the use of point-in-time spectral measurements is a severely limited method for addressing both the economic and environmental concerns of nitrogen management. In addition to the limitations of this methodology, remote sensing technology itself has continued to suffer from a host of limitations. Some of these include data availability, image calibration and atmospheric correction, cloud cover and shadowing influences (Moran et al., 1997), and soil background effects (Huete et al, 1985; Scharf and Lory, 2002). Remote sensing offers the ability for rapid estimation of the spatial variability of crop growth parameters across a field, but its limitations preclude it from being the sole technology used for solving a problem as complicated as nitrogen management in agricultural systems.

The CERES-Maize process-based crop growth model (Jones and Kiniry, 1986) has also been used to study nitrogen management for corn. This model utilizes carbon, nitrogen, and water balance principles to simulate, in homogenous units, the daily processes that occur during plant growth and development. The model has been shown to adequately simulate corn growth, development, and yield on plot-level, field-level, and regional scales for many locations around the world (Hodges et al., 1987; Carberry et al., 1989; Liu et al., 1989; Jagtap et al., 1993; Pang et al., 1998; Garrison et al., 1999; Paz et al., 1999; Fraisse et al., 2001). Inputs required for model execution include management practices (plant genetics, plant population, row spacing, planting and harvest dates, and fertilizer application amounts and dates), environmental factors (soil type, drained upper limit, lower limit, saturated hydraulic conductivity, root weighting factor, and effective tile drain spacing), and weather conditions (daily minimum and maximum temperature, solar radiation, and precipitation). Since CERES-Maize utilizes nitrogen balances for crop growth analysis, it can be conveniently extended to calculate surface and subsurface losses of nitrate-nitrogen. For example, the model has undergone several modifications such that nitrate nitrogen in run-off (Gabrielle et al., 1995), tile flow (Garrison et al., 1999), and leaching (Gabrielle et al., 1996) can be simulated as part of the crop production process. To explore the use of crop models within a precision agriculture context, Paz et al. (1999) used CERES-Maize to develop nitrogen fertilizer prescriptions for a 16 ha cornfield divided into 224 management zones. A key input
for the CERES-Maize crop model is corn population. However, because manual collection of population data in each of the 224 zones would be quite labor intensive, this model input parameter was assumed to be constant across the entire 16 ha field area. The authors cited this as a major limitation in their work. Similarly, Batchelor et al. (2002) cite that the requirement of needing site-specific model input parameters is the major limitation of using crop growth models for applications in precision agriculture.

One favorable solution to these problems involves the coupling of remote sensing technology and crop growth modeling, because together these two precision management tools have complementary functionality (Moran et al., 1997; Inoue, 2003). To explain, whereas crop models are excellent for crop growth analysis in the temporal domain, a tool such as remote sensing is better suited for crop growth analysis in the spatial domain. Conversely, whereas model input requirements have limited the use of crop models for spatial analyses, several practical problems, including cloud cover and flight availability, have reduced the potential of remote sensing as a temporal analysis tool. However, with the integration of these tools, the problems associated with one are offset by the benefits of the other. Initial work toward uniting these technologies began in the mid 1980’s when Wiegand et al. (1986a) demonstrated how vegetation indices computed from remote sensing images could be used to estimate crop growth variables of LAI, biomass, and yield. Since crop models also use these variables in simulations of crop growth, Wiegand et al. (1986b) reported efforts to utilize remote sensing-based estimates of crop growth parameters either as selected model inputs or for updating the progress of simulations throughout the season. Seidl et al. (2004) computed a normalized vegetation index from remote sensing imagery and used the index to update the leaf weight state variable in the CROPGRO-Soybean model. Their technique showed potential for improving simulations of soybean yield as long as remote sensing imagery was collected at the proper time. Another method for coupling process-based crop growth models with remote sensing information involves reinitialization or recalibration of the crop model using remote sensing-based estimates of a model state variable (Moulin et al., 1998). The reinitialization strategy involves iterative readjustment of a model initial condition in order to minimize the error between a simulated state variable and remote sensing-based estimates of that state variable later in the growing season. Maas
(1988) developed this technique to adjust the initial condition of green leaf area index (GLAI) in a grain sorghum model such that the error between simulated GLAI and remotely sensed GLAI on three dates in the growing season was less than 0.1. The technique is called recalibration when, instead of an initial condition, a model parameter is adjusted. If a crop growth model is able to output the several geometrical and spectral variables required for a radiation transfer model, then the crop model can also be recalibrated by directly comparing simulated and measured reflectance (Moulin et al., 1998). Several techniques have been developed for uniting crop growth models and remote sensing technology.

When used within a precision agriculture context, the goal of uniting remote sensing technology and crop growth models is to generate a robust tool for obtaining site-specific management plans for crops. Jones and Barnes (2000) present a useful figure for understanding how such a management tool should work (Figure 1.1). After remote sensing data calibration, image interpretation provides estimates for crop growth parameters that can be subsequently used as inputs or for updating or recalibrating crop growth simulations. These crop growth simulations serve to model the functionality of the agricultural system such that various management scenarios can be tested on a computer before actually implementing a scenario in the field. Finally, the decision support system serves to automate crop model simulations, compare management alternatives, and present the user with the best-case scenario given the management goals.

The overall goal of the work presented in this dissertation was to develop crop sensing and crop growth modeling technologies within the precision agriculture framework shown in Figure 1.1. Specifically, remote sensing images were used to estimate corn plant population within management zones across an Iowa cornfield. This was done with the idea that the remote sensing-based population estimates could be used as an input or for recalibration of spatial crop growth simulations across the field. Also, the CERES-Maize crop growth model was used within the framework of the Apollo decision support system to 1) calibrate and validate the model for spatial simulations of corn yield and 2) analyze the performance of nitrogen prescriptions across a cornfield over 37 years of historical weather information. Simulation output for yield and nitrogen left behind over the 37 growing seasons was then used to develop a methodology for assessing the production and
environmental risks of precision nitrogen management strategies across the cornfield. Using this methodology, it is possible to develop nitrogen prescription maps that achieve a proper balance between the economic and environmental concerns of nitrogen management in this cornfield.

![Diagram](image)

Figure 1.1. A proposed framework for making management decisions within a precision agriculture context (Jones and Barnes, 2000).

1.2 Dissertation Overview

This dissertation is presented as a compilation of three articles that are currently in various stages of the review process for refereed publication. Chapter 2 contains a paper entitled "Using aerial hyperspectral remote sensing imagery to estimate corn plant stand density," which describes the use of reflectance information from remote sensing imagery to estimate corn plant population across an Iowa cornfield. On this project, I was responsible for coordinating the collection of remote sensing imagery, collecting ground truth information, processing and interpreting all remote sensing and ground reference information, and preparing the final manuscript. Colleagues of mine who provided assistance on the project include Dr. Brian L. Steward, Dr. Amy L. Kaleita, and Dr. William D. Batchelor. Dr. Steward and Dr. Kaleita are associate and assistant professors,
respectively, within the Agricultural and Biosystems Engineering Department at Iowa State University. Dr. Steward specializes in the use of machine vision technology on agricultural vehicles for mapping of crop characteristics, particularly corn population. The corn population sensing system developed by researchers in his lab was essential for collecting the rigorous ground reference dataset of corn plant population for this work. Dr. Kaleita specializes in remote sensing technology and provided assistance in the processing and interpretation of remote sensing imagery for this project. At the onset of this project, Dr. Batchelor served as a full professor in the Agricultural and Biosystems Engineering Department at Iowa State University, but he left midway through the completion of this dissertation and became the department head for the Agricultural and Biological Engineering Department at Mississippi State University. Dr. Batchelor specializes in the use of crop growth models for applications in precision agriculture and provided insight on how plant population measurements from remote sensing imagery might be useful for crop growth simulations within a precision agriculture context. The paper presented in Chapter 2 will be submitted for publication in Transactions of the ASAE after this dissertation has been successfully defended.

Chapter 3 contains a paper entitled “Statistical procedures for validating CERES-Maize simulations of spatial corn yield variability,” which describes the use of a cross validation technique and bivariate confidence intervals on fitted model parameters to evaluate the performance of the CERES-Maize crop growth model in simulating corn yield spatially across the field. On this project, I was responsible for running the required model simulations, performing the statistical analysis on the simulation results, and preparing the manuscript for publication. Colleagues who assisted in this work were Dr. William D. Batchelor and Dr. Joel O. Paz. Dr. Paz was a research associate working under Dr. Batchelor during his time as a professor in the Department of Agricultural and Biosystems Engineering at Iowa State University. On this project, Dr. Batchelor and Dr. Paz were responsible for the development of the Apollo decision support system, which was used to automate the crop growth model simulations across the field. They also provided assistance in teaching me how to perform crop model simulations and how to properly interpret the results. The paper
presented in Chapter 3 was submitted for publication in Transactions of the ASAE on September 27, 2005, and it is currently under review.

Chapter 4 contains a paper entitled "Methodology to link production and environmental risks of precision nitrogen management strategies in corn," which describes the use of CERES-Maize simulation output over a 37-year weather sequence to develop a methodology for linking yield and nitrogen left behind for precision management of nitrogen in an Iowa cornfield. On this project, I was responsible for running the required model simulations, performing the statistical analysis on the simulation results, developing the methodology to link production and environmental risk, and preparing the final manuscript for publication. Colleagues who assisted in this work include Dr. William D. Batchelor, Dr. Joel O. Paz, Dr. Brian L. Steward, and Dr. Petrutza C. Caragea. Again, Dr. Batchelor and Dr. Paz were responsible for development of the Apollo decision support system, which was used to automate the crop model simulations for this project. They also assisted me in understanding how to implement the crop model for applications in precision agriculture. In addition, Dr. Batchelor presented the original idea for the methodology that I developed in this work. Dr. Steward was included as an author for his continued contributions in moving toward our overall goal of uniting the sensing and modeling technologies. Finally, Dr. Caragea is an assistant professor in the Department of Statistics at Iowa State University, and she provided assistance with the statistical analysis of crop model simulation output in this work. The paper presented in Chapter 4 was submitted for publication in Agricultural Systems on January 17, 2005. After the peer review, a modified draft was submitted on September 7, 2005, and the paper was accepted for publication on September 23, 2005. The paper is currently in press and will appear in the journal later this year. Chapter 5 details the conclusions and future recommendations for this area of work, and it is followed by an acknowledgements section, a brief biography of myself, and my current curriculum vitae.

1.3 References


CHAPTER 2. USING AERIAL HYPERSPECTRAL REMOTE SENSING IMAGERY TO ESTIMATE CORN PLANT STAND DENSITY

A paper to be submitted to *Transactions of the ASAE*

Kelly R. Thorp, Brian L. Steward, Amy L. Kaleita, William D. Batchelor

2.1 Abstract

Aerial hyperspectral remote sensing imagery was collected on three dates over three plots of corn. The imagery had a spatial resolution of 1 m and a spectral resolution of 3 nm between 471 nm and 828 nm. A machine vision corn plant population sensing system was also used to map every row of corn within the three plots, and a complete inventory of corn plants was generated as a rich ground reference dataset for remote sensing image analysis. A multiple linear regression analysis was performed to predict corn plant stand density using reflectance in combinations of three wavebands, and $R^2$'s of up to 0.82 were found. Estimates of corn plant stand density were best when using imagery collected at the later vegetative growth stage. Quantization effects due to row width complicated corn plant stand density estimates at 2 m spatial resolution, and better estimations were typically seen at resolutions of 6 m and 10 m. For the best-case scenarios, the first predictor variable in the regression model typically fell in the blue reflectance region (473 to 492 nm). The second predictor variable was typically in the longer green and shorter red wavelengths (584 to 635 nm), and reflectance for the third predictor variable was typically at the red edge (729 nm) or in the near-infrared region. Because results for the second and third predictor variables tended to straddle between important regions of typical vegetative reflectance spectra, it is expected that multiple linear regressions using a greater number of bands would improve the distinction between important spectral ranges for estimating corn plant stand density.

2.2 Introduction

Corn plant population, or plant stand density, is an important crop growth parameter that influences corn (*Zea mays* L.) yield. Duncan (1958) and Duncan (1984) determined that the weight of grain produced by individual corn plants decreases as the plant population
increases, because at higher stand densities neighboring corn plants must compete more fiercely for resources. On the other hand, once corn plant population decreases beyond the level at which population pressure limits yield, the average yield per plant cannot continue to increase, because plant genetics limit the weight of grain that a single plant can produce. Thus, for a given set of environmental conditions, there exists an optimum corn plant stand density at which corn yield will be maximized. Furthermore, due to the development and usage of corn hybrids that yield more at higher plant densities, recommended optimum planting densities have increased since the 1960s (Duvick and Cassman, 1999).

Spatial variability in corn plant population arises as a result of planter performance issues (Nielsen, 1995), emergence delays or failure (Nielsen, 1991), and early-season plant death due to stress. When these problems occur, the distribution of corn plants within the crop row, or the plant spacing, also becomes spatially variable across the field. The effect of interplant spacing variability on corn yield is unclear. Several studies have shown that corn yield decreased on the order of 159 kg ha\(^{-1}\) (3 bu acre\(^{-1}\)) for each 2.54 cm (1 in) increase in the standard deviation of plant spacing (Krall et al., 1977; Nielsen, 1991). Nafziger (1996) found that corn plants growing on either side of a “skip” compensated for only 47% of the missing plant’s grain at 44,479 plants ha\(^{-1}\) (18,000 plants acre\(^{-1}\)) and 19% of the missing plant’s yield at 74,131 plants ha\(^{-1}\) (30,000 plants acre\(^{-1}\)), thus reducing overall crop yield. Although the yield of each plant in a “double” was 10% to 17% less than uniformly spaced plants, the net effect of doubles was to increase yield at all populations. Because both skips and doubles increased plant spacing variability but had opposite effects on yield, the researchers concluded that the skips and doubles affected yield mainly through changes in plant stand density. Vanderlip et al. (1988) found that plant spacing variability accounted for 5% to 23% of grain yield variability, and Liu et al. (2004) found no significant relationship between these two variables. Barbieri et al. (2000) planted corn in 0.35 m rows, one half the conventional width, and increased plant spacing to maintain a plant population consistent with conventional methods. They found the narrow rows to increase grain yield by 27% to 46% under the condition of low nitrogen availability. Other experiments have shown that the use of narrow rows at higher than recommended plant populations can also significantly increase grain yield (Hunter et al., 1970; Porter et al., 1997; Widdicombe and Thelen, 2002).
Although the effect of plant spacing variability on grain yield is unclear, results clearly show that row width modifications have potential to increase yield. Therefore, it can be concluded that, in addition to the total population, the distribution of plants over an area is also important for optimizing yield.

Since corn plant population has been known to have significant effects on grain yield, this crop growth parameter has been a topic of precision agriculture research. In terms of management, variable-rate seeding has been marketed to producers as a means to optimize yield spatially across the field. However, Bullock et al. (1998) cautions that this practice may not be economically beneficial for producers until more extensive information on the spatial relationship between plant population and crop yield is obtained for their fields. Other researchers have developed sensing technology for corn plant population and plant spacing variability measurement. Birrell and Sudduth (1995) mapped corn population at harvest with a mechanical sensor mounted on a combine corn header. Plant populations measured by the sensor were within 5% of plant population measured manually by hand. Plattner and Hummel (1996) developed an optical sensor to map corn population at harvest, and the sensor was able to estimate average plant spacing with an error of 6.2%. Using a machine vision approach, Shrestha and Steward (2003) developed a sensing system for measurement of corn plant population and plant spacing in early growth stage corn. The sensing system utilized a video camera and a global positioning system (GPS) receiver to collect and locate image frames along corn rows, and video processing algorithms were developed for sequencing consecutive image frames, segmenting corn plants from soil background, and determining the geographic position of each corn plant in the row. The system plant counts and manual plant counts were correlated with an $r^2$ of 0.90. Further developments in this work include a chain code methodology for delineating plant boundaries in sequenced video frames (Shrestha and Steward, 2005) and a statistical approach for improving the robustness of video processing algorithms over a wider range of field conditions (Shrestha et al., 2004a).

In addition to ground-based systems, aerial and satellite imaging systems have been regularly used to monitor the status of crop growth, and researchers have related spectral reflectance information obtained from these systems to crop growth parameters such as
emergence date (Wanjura et al., 2003), percent canopy cover (Maas, 1998; Thorp et al., 2004), biomass development (Thenkabail et al., 2000), leaf area index (Bouman, 1992), and yield (GopalaPillai and Tian, 1999). However, there were no studies found in literature where remote sensing was used to detect spatial variability in corn plant stand density.

Due to the influence of corn plant population on crop yields, corn growth models typically require plant population inputs for crop growth simulations. Consequently, the use of corn growth models for applications in precision agriculture requires knowledge of the spatial variability of plant population across the cornfield (Batchelor et al., 2002). Paz et al. (1999) used the CERES-Maize crop model to develop nitrogen fertilizer prescriptions for a 16 ha cornfield divided into 224 management zones. Corn population was assumed to be constant across the entire 16 ha field area, because manual collection of population data in each of the 224 zones would be quite labor intensive. The authors cited this as a major limitation in their work. One favorable solution to this problem involves the coupling of precision management tools that have complementary functionality (Moran et al., 1997). To explain, whereas crop models are excellent for crop growth analysis in the temporal domain, a tool such as remote sensing is better suited for crop growth analysis in the spatial domain. Conversely, whereas model input requirements have limited the use of crop models for spatial analyses, several practical problems, including cloud cover and flight availability, have reduced the potential of remote sensing as a temporal analysis tool. However, with the integration of these tools, the problems associated with one are offset by the benefits of the other. With this concept in mind, our main objective was to explore the use of aerial hyperspectral remote sensing technology as a means to estimate variability in corn plant stand density. Secondary objectives were to identify the most useful spectral ranges and to determine the spatial and temporal limitations of using remote sensing for this purpose.

2.3 Materials and Methods

2.3.1 Data Collection

Data collection occurred over three sections of a cornfield at Iowa State University’s Agronomy and Agricultural Engineering Research Center west of Ames, Iowa, USA
(93.77879°W, 42.00988°N). The three data collection regions are aptly named Plot North (PN), Plot South (PS), and Plot West (PW) and are arranged as shown in Figure 2.1. Each plot was approximately 1 ha in land area (Table 2.1). On June 4, 2004, corn was planted in 76.2 cm (30 in.) rows in PN and PS. The planting of PN and PS was coordinated as part of another research project investigating the effects of planter speed, planter row unit design, and compaction on corn population and yield. These various treatments in addition to manual thinning introduced spatial variability in corn plant population over a relatively small area. A conventional planting methodology was used to sow PW on June 13, 2004 (Table 2.1).

Figure 2.1. Data collection occurred over three sections of a cornfield in Iowa. Plots were arranged as shown on this 1 m spatial resolution image collected July 25, 2004.
Table 2.1. Summary of important characteristics for the three data collection regions.

<table>
<thead>
<tr>
<th></th>
<th>Plot North</th>
<th>Plot South</th>
<th>Plot West</th>
</tr>
</thead>
<tbody>
<tr>
<td>Land Area</td>
<td>0.9 ha</td>
<td>1.1 ha</td>
<td>1.0 ha</td>
</tr>
<tr>
<td>Tillage</td>
<td>Conventional</td>
<td>No-Till</td>
<td>Conventional</td>
</tr>
<tr>
<td>Planting Date</td>
<td>June 4, 2004</td>
<td>June 4, 2004</td>
<td>June 13, 2004</td>
</tr>
<tr>
<td>Planned Population Variability</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Ground Reference Data Collection</td>
<td>June 23, 2004</td>
<td>June 23, 2004</td>
<td>June 30, 2004</td>
</tr>
<tr>
<td>Remote Sensing Date 1</td>
<td>June 22, 2004</td>
<td>June 22, 2004</td>
<td>June 22, 2004</td>
</tr>
<tr>
<td>Remote Sensing Date 3</td>
<td>September 3, 2004</td>
<td>September 3, 2004</td>
<td>September 3, 2004</td>
</tr>
<tr>
<td>Number of cells at 2 m</td>
<td>2,576</td>
<td>2,599</td>
<td>2,159</td>
</tr>
<tr>
<td>Number of cells at 6 m</td>
<td>259</td>
<td>259</td>
<td>210</td>
</tr>
<tr>
<td>Number of cells at 10 m</td>
<td>88</td>
<td>88</td>
<td>75</td>
</tr>
</tbody>
</table>

Aerial hyperspectral remote sensing imagery was collected over the study area using the hyperspectral focal plane scanner and data acquisition system developed by scientists at the Institute for Technology Development at Stennis Space Center in Mississippi (Mao, 2000). A Cessna single engine aircraft was used as the platform for remote sensing data collection, and the sensor was fastened in a gyro stabilized mount to minimize the effect of airplane roll, pitch, and yaw on data quality. The scanner collected data between 471 nm and 828 nm at a 3 nm bandwidth for a total of 120 bands of spectral information. The spatial resolution of the imagery was 1 m. Remote sensing data was collected over the entire study area on three dates in the summer of 2004: June 22, July 25, and September 3. These dates corresponded to corn growth stages V5, V15, and R4 for PN and PS and V3, V12, and R2 for PW (Table 2.2). Prior to remote sensing data collection, calibration tarps showing eight grayscale levels from white to black were laid out in an area near the study site. A spectroradiometer (1500, GER Corporation, Millbrook, NY, USA) was used to measure the spectral reflectance from each panel between 286 nm and 1102 nm at a bandwidth of approximately 1.5 m. Pilots then captured hyperspectral remote sensing imagery over both the study area and over the calibration tarps.
Ground reference data was collected using the machine vision-based corn plant population sensing system developed by Shrestha and Steward (2003). System components were mounted on a 4x4 Kawasaki all-terrain vehicle (ATV) for data collection in the field (Figure 2.2). A digital camcorder (DCR-TRV900, Sony Corporation, New York, NY, USA) was used for video acquisition of crop rows, and a special mount was designed to hold the camera at the front center of the vehicle. Special features of the camera mount included a metal frame skirted with translucent white cloth for diffusion of sunlight in the camera's field of view and a connection mechanism that isolated the camera from vehicle vibrations. The camera was mounted at a height of 0.53 m above the ground, and this provided a 0.4 m by 0.3 m field of view. Video of crop rows was recorded onto miniDV tapes. A global positioning system (GPS) receiver (GG24-RTK, Thales Navigation, Santa Clara, CA, USA) was used to obtain the geographic coordinates of the ATV in the field. The antennae for the GPS receiver was mounted above the storage box at the rear center of the vehicle, and the distance between the video camera and the antennae along the longitudinal axis of the vehicle was 1.73 m. A GPS encoder/decoder (VMS 200, Red Hen Systems, Inc., Fort Collins, CO, USA) was used to convert GPS strings to an audio signal that could be recorded on the soundtrack of the miniDV tapes. A corrugated aluminum storage box was mounted on the rear of the ATV for storage of the global positioning equipment.
On the days of ground reference data collection at the study site, a second GPS receiver (GG24-RTK, Thales Navigation, Santa Clara, CA, USA) was placed at the location of a benchmark on the research farm. This receiver was used as a base station to improve the accuracy of position measurements at the rover receiver on the ATV. The two GPS receivers communicated with each other via a radio link (RFM-96W, Pacific Crest Corporation, Santa Clara, CA, USA). The video camera was then set to collect video in progressive scan mode with a shutter speed of 1/1000 s. Due to the movement of the ATV, these settings were essential to insure that high quality video was collected. After fully zooming out the camera and allowing it to automatically focus on the scene, the camera’s auto focus function was turned off. If left on, the auto focus was found to continually overcompensate as it attempted to adjust the camera’s focus during data collection, and this ultimately caused blurriness in the video. Prior to video collection, the camera’s white balance was also adjusted to insure a more natural video color. All other camera controls were used at their default settings. Ground reference data was collected in PN and PS on June 23, 2004 and PW on June 30, 2004 (Table 2.1). The system was used to collect information over every crop row contained within the area of the three plots. Video frames collected were 480 by 720 pixels in size with
24-bit color resolution, and GPS information was recorded on the soundtrack at a frequency of 5 Hz. The ATV was operated at an average speed of 1 m s\(^{-1}\).

2.3.2 Ground-Reference Data Processing

Following the initial development of their machine vision system, Shrestha and Steward (2003) packaged their algorithms for system operation and video processing into a C++ application named ESCOPE. Characteristics of the software include two operation modes, including "real-time mode" for automatic collection of corn plant population and spacing information in the field and "laboratory mode" for analysis of pre-recorded videotapes in the laboratory. The ESCOPE software also provides three options for image segmentation, including a new algorithm that significantly reduces the processing time required for this task (Shrestha et al., 2004b). In addition, a manual plant count adjustment algorithm and graphic user interface was developed such that a user could visually inspect and make corrections to the automatic plant counting algorithm results on the computer screen. Due to the difficulties in automatically delineating plants at higher growth stages, such corrections were most needed when attempting to count larger plants. To generate a ground-reference dataset of corn plant population for this work, the ESCOPE software first was used in laboratory mode to segment the video frames that were recorded during the data collection effort. To save time, the fast image segmentation algorithm (Shrestha et al., 2004b) was used as a first choice. However, when poor field conditions or video quality warranted a more robust algorithm, the slower algorithm presented in Shrestha and Steward (2003) was used. Plant identification and counting was then performed on all the sequenced images of crop rows using an image segmentation algorithm (Shrestha and Steward, 2003) combined with a chain code approach (Shrestha and Steward, 2005). However, because crop rows were mainly recorded at higher growth stages in this work, manual adjustments were made in a majority of the sequenced images to insure the accuracy of plant locations. After these adjustments, ESCOPE produced a text file containing the geographic coordinates of all marked plants in the sequenced images. Because video was recorded and analyzed on every crop row, the ground-based system was used to generate a complete inventory of all corn plant locations within our study area. The generation of this dataset was quite costly in terms
of manual effort; however, it enabled a unique investigation into the use of remote sensing imagery as an alternative way to estimate plant population spatially across cornfields.

2.3.3 **Hyperspectral Data Processing**

The hyperspectral imagery was prepared for analysis using both spatial and spectral preprocessing. First, since raw image spatial distortions can be produced by changes in aircraft attitude during the scanner-based image collection process, a correction procedure, developed by Yao et al. (2001), was implemented to remove as much spatial distortion in the raw hyperspectral imagery as possible. Next, the images were georeferenced to the Universal Transverse Mercator (UTM) coordinate system using a field boundary map that was obtained with a meter-level accuracy backpack GPS unit (Pathfinder Pro XRS, Trimble Navigation Limited, Sunnyvale, CA, USA). For spectral correction, a minimum noise fraction (MNF) transformation (Green et al., 1988) was used to remove sensor noise in the raw reflectance data. Then, by matching the digital numbers of the calibration tarps in each image to the reflectance measurements taken of the tarps on the ground, the imagery was calibrated to percent reflectance with an empirical line calibration procedure (Smith and Milton, 1999). These pre-processing steps were performed separately for each of the three remote sensing image collection dates using the Environmental for Visualizing Images (ENVI) software (Version 4.2, Research Systems, Inc., Boulder, CO, USA).

2.3.4 **Statistical Analysis**

In preparation for statistical analysis, ground reference data and spectral reflectance data were aggregated at three separate spatial resolutions. Since all the remote sensing images were originally collected at 1 m spatial resolution, the reflectance measurements in each waveband were averaged over square blocks of 4, 36, and 100 raster units to decrease the spatial resolution of the imagery to 2 m, 6 m, and 10 m, respectively. The total number of grid cells for PN, PS, and PW resulting from this aggregation process are given in Table 2.1. Using ArcGIS (Version 9, ESRI, Redlands, CA, USA), the raster grids for images on each date and at each spatial resolution were then used to clip the ground reference corn population measurements over each of the three plots. The total number of plants within
each raster grid cell area was then determined, and the plant counts were normalized by the
grid cell area to generate raster maps of corn plant stand density. Since corn population was
measured on the ground only once during the season, we assumed that the plant stand was
well established at the time of the ground-based measurements and that the corn population
did not change significantly throughout the remainder of the growing season.

Multiple linear regression analysis (Neter et al., 1996) was used to relate reflectance
measurements to plant density across the plots. A Visual Basic program was written to
compute the slopes and coefficient of multiple determination, $R^2$, for linear regressions of all
1-, 2-, and 3-band combinations of reflectance and plant stand density. The highest values
for $R^2$ were achieved when using three reflectance bands to predict corn plant stand density.
Thus, the linear statistical model of greatest value in this work can be written as

$$ Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i, $$

where $Y_i$ is the response variable, corn plant density, in the $i$th grid cell, and $X_{i1}$, $X_{i2}$, and $X_{i3}$
are the average reflectance in three wavebands over the area of the $i$th grid cell. The
parameters of the model are $\beta_0$, $\beta_1$, $\beta_2$, and $\beta_3$. The error term is $\varepsilon_i$. Since a total of 120
bands were available, the total number of 2-band combinations was $120 \times 119 / (2 \times 1) = 7,140$
and the total number of 3-band combinations was $120 \times 119 \times 118 / (3 \times 2 \times 1) = 280,840$. The
longest computing time for a single plot was approximately 2 hours for the linear regressions
on 3-band combinations at the 2 m spatial resolution. After computation was complete, the
band combinations tested for each case of plot, spatial resolution, and image collection date
were placed in order according to the $R^2$ value. An evaluation was then conducted to
determine the spatial resolution, spectral wavelengths, and temporal considerations necessary
for estimating corn plant stand density using remote sensing imagery. For spatial and
temporal evaluations, conclusions were drawn based on the performance of the top
regression model for each case. To insure that the cases with poorer results did not influence
the succeeding spectral evaluation, investigations into the important spectral bands for
estimating corn plant stand density were performed only on the cases that provided a top $R^2$
of greater than 0.70. A rank plot of $R^2$ values showed a natural break at this location. For
these top performing cases, the results for the top 100 3-band regression models were
separated from the rest, and a count was made of the number of times that a particular
waveband was used as the first, second, or third predictor variable in these models. Histograms of these counts then provided information on the waveband range and center for the three spectral regions most highly correlated with corn plant stand density when using multiple linear regression on three wavebands.

2.4 Results and Discussion

For all cases of plot, spatial resolution, and image collection date, the greatest values for $R^2$ were found when using linear combinations of three reflectance bands (Table 2.3). The relationship between reflectance spectra and plant stand density at 2 m spatial resolution generally provided relatively low $R^2$'s ranging from 0.02 to 0.42. The reason for this result can be explained in terms of the crop row width. In this study, crop rows were planted at a width of 0.76 m (30 in). Thus, depending on the location of the 2 m spatial resolution raster grid relative to the crop rows, some raster cells would contain three crop rows while adjacent cells would contain only two crop rows. If a raster cell contained three crop rows, the plant count for the cell would be significantly higher than for the cells containing only two crop rows. Given the low $R^2$'s at 2 m spatial resolution, it is evident that this crop row quantization effect was unable to be detected within the remote sensing imagery. These results make sense, since light interaction within a plant canopy is not restricted to the bounds of a raster grid whereas plant population can be discretely measured within that grid. At 6 m and 10 m spatial resolution, the presence of a greater number of rows within the raster grid cells reduced the effect of row quantization on plant counts within the raster grid, and $R^2$'s were always higher for these lower spatial resolution cases.

When comparing the $R^2$'s across the three image collection dates, results were highest for the July 25 image for PN and PS. For both of these plots, an $R^2$ of 0.79 was achieved when relating reflectance spectra from July 25 to corn plant stand density at the 6 m spatial resolution. Also, the highest $R^2$ in the entire study, 0.82, was achieved when relating reflectance spectra from July 25 to corn plant stand density in PS at the 10 m spatial resolution (Figure 2.3). For PW, the September 3 image gave the best results; however, the highest $R^2$ for this plot was only 0.66 at the 10 m spatial resolution. On the June 23 image collection date, corn plants were still in their early vegetative growth stages (Table 2.2), and
canopy closure had not yet occurred. At early growth stages, the effects of soil background on reflectance spectra have been known to hamper the analysis of remote sensing imagery for vegetation (Thorp et al., 2004). Similarly in this study, results indicated that the use of remote sensing imagery to estimate the vegetative growth parameter of plant stand density was less reliable at earlier growth stages before the canopy had closed. One would also expect to see a decline in the ability to use reflectance spectra to estimate corn plant stand density at the end of the growing season as the plants reach physiological maturity and lose vegetative vigor. Our results for PN and PS confirm this, since the $R^2$s for the September 3 image were lower than that for the July 25 image. Plants in PN and PS were at the V15 and R4 growth stages on July 25 and September 3, respectively (Table 2.2). Plant population in PW was most correlated to reflectance on September 3 when plants were at the R2 growth stage. Thus, reflectance spectra in remote sensing images was best used to estimate the vegetative growth parameter of plant stand density when plants were at the upper vegetative or lower reproductive growth stages.

Another interesting result was found when comparing the $R^2$s for PN and PS on each image collection date across all spatial resolutions. For the June 23 image collection date, the $R^2$s at each spatial resolution were always higher for PN. Prior to planting, PN was tilled using a conventional tillage method while PS was managed using no-till practices (Table

### Table 2.3

The highest $R^2$ obtained in a multiple linear regression of a 3-band reflectance combination and plant stand density for each case of plot, spatial resolution, and image collection date.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Date</th>
<th>$R^2$ at each Spatial Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2 m</td>
</tr>
<tr>
<td>North</td>
<td>6/22/04</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>7/25/04</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>9/3/04</td>
<td>0.32</td>
</tr>
<tr>
<td>South</td>
<td>6/22/04</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>7/25/04</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>9/3/04</td>
<td>0.40</td>
</tr>
<tr>
<td>West</td>
<td>6/22/04</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>7/25/04</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>9/3/04</td>
<td>0.15</td>
</tr>
</tbody>
</table>
2.1). As a result, it is expected that the higher proportion of residue covering the surface of PS increased the soil brightness in that plot. Since increasing soil brightness has been shown to cause reduced correlations of reflectance spectra to vegetative growth parameters at early growth stages (Thorp et al., 2004), the residue cover in PS probably increased the difficulty in detecting corn plant stand density variability on the June 23 date relative to PN. However, on the July 25 and September 3 dates, multiple linear regression gave higher $R^2$s for PS than PN across all spatial resolutions. Also, the $R^2$s for PW were lower than that of PN and PS for all cases of image date and spatial resolution. It is expected that the total corn population variability across each plot determined the relative performance of estimating population from reflectance in the plot. The standard deviations for corn plant density, aggregated at the 6 m spatial resolution, were 0.91, 1.20, and 0.53 plants m$^{-2}$ for PN, PS, and PW, respectively. In addition, a histogram of the data shows that plant density ranged from 2 to 9 plants m$^{-2}$ for
PN and PS, but it only ranged from 6 to 9 plants m\(^{-2}\) for PW (Figure 2.4). Thus, the \(R^2\)’s for relating reflectance spectra to corn plant stand density were higher as the total variability of corn plant stand density across the plot increased. This result shows that the artificial introduction of plant stand variability in PN and PS increased the potential for using remote sensing images to detect that variability, simply because there was more variability to detect.

![Histograms of stand density in each plot aggregated at 6 m spatial resolution.](image)

Figure 2.4. Histograms of stand density in each plot aggregated at 6 m spatial resolution.

Investigations into the important spectral bands for estimating corn plant stand density were performed only on the cases that provided a top \(R^2\) of greater than 0.70. A rank plot of the top \(R^2\)’s showed a natural break at this location (Figure 2.5). For PN, these cases included July 25 at the 6 m and 10 m spatial resolutions (Table 2.3). Similarly for PS, the cases meeting this criterion included both the July 25 and September 3 dates at both the 6 m and 10 m spatial resolutions. For PW, no cases had a top \(R^2\) greater than 0.70. Selecting the wavebands combinations used in the top 100 regression models for these six highest performing cases meant that the total number of combinations used in the spectral analysis was 600. The range of \(R^2\)’s for the top 100 regression models for each case is given in Table
2.4. Creating histograms of the number of times that the reflectance in a particular waveband was used as a predictor variable in Equation 2.1 provided insight on the wavelengths of greatest interest when using reflectance data to estimate corn plant population in a 3-band multiple linear regression procedure (Figure 2.6).

Figure 2.5. A rank plot of top $R^2$s for each case demonstrates the natural break that occurs near $R^2 = 0.70$.

Table 2.4. The $R^2$ range for the top 100 regression models from the top cases used in the spectral evaluation.

<table>
<thead>
<tr>
<th>Plot</th>
<th>Date</th>
<th>S.R. (m)</th>
<th>$R^2$ Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>7/25/04</td>
<td>6</td>
<td>0.7949 to 0.7876</td>
</tr>
<tr>
<td>North</td>
<td>7/25/04</td>
<td>10</td>
<td>0.7505 to 0.7262</td>
</tr>
<tr>
<td>South</td>
<td>7/25/04</td>
<td>6</td>
<td>0.7892 to 0.7853</td>
</tr>
<tr>
<td>South</td>
<td>7/25/04</td>
<td>10</td>
<td>0.8191 to 0.8009</td>
</tr>
<tr>
<td>South</td>
<td>9/3/04</td>
<td>6</td>
<td>0.7831 to 0.7771</td>
</tr>
<tr>
<td>South</td>
<td>9/3/04</td>
<td>10</td>
<td>0.7351 to 0.7265</td>
</tr>
</tbody>
</table>
For the first predictor variable in the regression model, $X_1$, the mean of the wavelengths of interest was 482.2 nm with a standard deviation of 9.4 nm. Thus, 68% of the time, the reflectance values most useful for the first predictor variable in the regression model fell within the range of 473 nm and 492 nm, which corresponds to blue light in the electromagnetic spectrum. According to the results of Thenkabail et al. (2000), this range of wavelengths represents the minimum crop-to-soil reflectance ratio within the blue and green portions of the spectrum. Due to the high absorption of blue light by chlorophyll and the relatively high reflectance of blue light from soil, reflectance variability within this range of wavelengths relates to variability in vegetative growth. Reflectance information within this range of wavelengths was also found useful for computing narrow-band vegetation indices and relating index values to percent canopy cover in soybeans (*Glycine max* (L.) Merr.) (Thorp et al., 2004). It is also interesting to note that this range of wavelengths approached...
the spectral limits of the remote sensing system, and the reported wavelength range may have been less narrow in the case that spectral data was available at wavelengths shorter than 471 nm.

For the second predictor variable in the regression model, \( X_2 \), the mean of the wavelengths of interest was 609.5 nm with a standard deviation of 25.8 nm. Thus, 68% of the time, the reflectance values most useful for the second predictor variable in the regression model fell within the range of 584 nm and 635 nm. This range of wavelengths straddles the upper green and lower red portions of the electromagnetic spectrum. Thenkabail et al. (2000) found this portion of the spectrum to be important for agricultural crop studies, because the first derivative of reflectance spectra for crops reaches a minimum within this range. Also, similar to the blue region of the spectrum, the red portion of the spectrum is useful for detecting variability in vegetative growth due to the high absorbance of red light by chlorophyll in leaves. Thorp et al. (2004) also found that reflectance data within the range of this distribution for predictor variable, \( X_2 \), was important for development of narrow-band vegetation indices.

For the third predictor variable in the regression model, \( X_3 \), the mean of the wavelengths of interest was 749.0 nm with a standard deviation of 34.6 nm. The wavelength range within one standard deviation of the mean was not calculated due to the severe non-normality of this histogram. The most striking feature for the \( X_3 \) histogram is the spike at the wavelength of 729 nm. This wavelength likely corresponds to the location of the red edge (Horler et al., 1983), which is the transition point between absorption of visible red and reflection of near-infrared in plant leaves. At this point, the change in slope of reflectance spectra per unit change in wavelength reaches a maximum. Other wavelengths of interest occur mainly within the near-infrared portion of spectrum, and the unique response of vegetation to incident light in these wavelengths has been demonstrated in countless remote sensing investigations. Due to the internal cellular structure of plant leaves, very little near-infrared radiation is absorbed by a crop canopy, and up to 50% of incident near-infrared light can be reflected back toward the sensor (Knipling, 1970). The usefulness of reflectance spectra in the near-infrared region and at the red edge for detection of vegetation is well known and was an expected result.
2.5 Conclusions

Reflectance information collected over corn canopies was shown to have a strong relationship with corn plant stand density at mid-season. This work fills a research gap in the arena of corn population sensing, which to date has only been developed for counting plants during the early stages of corn growth (Shrestha and Steward, 2003) and while harvesting (Birrell and Sudduth, 1995). Effective use of remote sensing imagery for estimating population was shown to depend heavily on timing. Therefore, for efforts to be fruitful, plans for data reconnaissance must be well executed to acquire imagery when corn plants are reaching the later vegetative growth stages. If image collection dates are too early, results may be hampered by the strong influence of soil background on reflectance spectra. If remote sensing images are collected too late in the growing season, the onset of reproductive development and senescence prevents the use of reflectance spectra for estimating plant population.

The characteristics of the ground reference dataset in this work demonstrate the usefulness of ground-based crop sensing systems for testing the effectiveness of remote sensing technology. Since the entire area of each plot was mapped for corn plant geographic locations, no assumptions were made regarding corn population in unmeasured locations and there was no extrapolation of population measurements to larger areas based on strategic sampling. This was possible due to the existence of the ground-based corn population sensing system developed by Shrestha and Steward (2003). Future research in agricultural remote sensing will benefit from the development of ground-based sensing systems that can relatively quickly generate maps of important crop growth and soil parameters across the field. By first acquiring a detailed map of these parameters on the ground, a truer assessment of the limitations of remote sensing can be obtained as camera systems are incorporated on aerial and satellite platforms farther away from the scene. Then, it is possible to determine whether remote sensing offers any advantages over ground-based data collection and whether remote sensing images can be used to accurately estimate the true variability of crop parameters on the ground. For example, this study showed that remote sensing offers an advantage over ground-based data collection at mid-season; however, corn plant stand
density could not be estimated at higher spatial resolutions due the effects of row quantization within an image’s raster grid.

Histograms resulting from an evaluation of important spectral wavebands indicated that the range of wavelengths useful for the second and third predictor variable tended to span more than one type of electromagnetic radiation. For example, in the histogram for the third predictor variable, results were good when selecting the third waveband from either the red edge location or the near-infrared region; however, these locations in vegetative spectra are quite distinct from each other. Such results indicate the potential for improving estimates of corn plant population if multiple linear regressions were performed using combinations of a greater number of bands. Unfortunately, computing technology currently makes testing higher numbers of band combinations more impractical, because a great amount of time is required to make the necessary calculations. Exploration of alternative analysis techniques, such as genetic algorithms, may help overcome this current limitation in processing of hyperspectral data.

Use of remote sensing imagery to detect population variability in production cornfields will depend on the level of variability that exists over the area of interest. In this study, remote sensing was effective at estimating corn plant stand density for plots having artificial variability incorporated as part of the experimental design. However, results for the plot that was planted using the conventional methodology were less favorable. Since most producers aim for uniform plant populations in their cornfields, further investigations are needed into the level of population variability that can exist in production cornfields before remote sensing images are used to estimate this variability. Ground-based sensing systems such as those developed by Shrestha and Steward (2003) and Birrell and Sudduth (1995) offer much potential for completing this objective.

### 2.6 Acknowledgements

This journal paper was supported by the Hatch Act and State of Iowa funds. The authors also express sincere thanks to the Institute of Technology Development in Urbana, IL and NASA for collection of aerial hyperspectral remote sensing imagery and to Pioneer Hi-Bred International, Inc. in Johnston, IA for use of their data collection equipment.
2.7 References


CHAPTER 3. STATISTICAL PROCEDURES FOR VALIDATING CERES-MAIZE SIMULATIONS OF SPATIAL CORN YIELD VARIABILITY

A paper submitted to Transactions of the ASAE

Kelly R. Thorp, William D. Batchelor, Joel O. Paz

3.1 Abstract

Validation of crop models is often neglected due to limitations in available measured data. In this work, a cross validation approach was used to validate the CERES-Maize crop growth model in spite of limited measured data. Simulations were run for an Iowa cornfield divided into 100 grid cells. Five growing seasons of measured information were available for calibration of two parameters, tile drainage rate and saturated hydraulic conductivity, in each grid cell. Cross validation requires that the model be calibrated five times in each grid cell by alternatively leaving out one season of measured information. The model is then validated using the fitted parameters to simulate the growing season left out of the calibration. To evaluate model performance in each grid cell, the root mean squared error of prediction (RMSEP) was computed as the average error between measured and simulated yield for the five validation runs. Results indicated that the model performed most poorly when using the wettest or driest growing seasons to validate the model, with validation errors up to 1400 kg ha\(^{-1}\). Model parameters fitted under moderate weather conditions were less flexible for simulating yield in growing seasons with more extreme weather conditions. Spatial variability in model performance across grid cells indicated that topography may influence the ability of the model to simulate yield, because the model does not account for surface and sub-surface run-on between neighboring grid cells, and the dynamics of this process would be more complex for a sloped topography.

3.2 Introduction

In the past decade, process-oriented crop growth models have been implemented to understand the core questions in precision agriculture: how do the parameters that describe agricultural systems vary over space and time and how can this information be used to
improve management of crop production inputs. Fundamentally, crop models have been used to identify the factors that limit crop yield and cause observable spatial yield variability (Batchelor et al., 2002). Specific yield-limiting factors that have been studied include water stress, nitrogen stress, soybean cyst nematode, and weeds (Paz et al., 1998; Paz et al., 1999, Paz et al., 2001b, Paz et al., 2002). Models have also been coupled with other precision agriculture technologies to better understand yield variability. Remote sensing images have been particularly useful for delineating zones of crop growth spatial variability within a field (GopalaPillai and Tian, 1999), and field data collection of crop model input parameters have been planned according to the results of remote sensing image analyses (Basso et al., 2001). Once the yield limiting factors are understood, crop modeling techniques can be extended to develop variable-rate management strategies that optimize producer economic returns. Methodologies for development of prescriptions that optimize economic return have been developed for plant population and variety selection in soybeans (Paz et al., 2001a; Paz et al., 2003) and for nitrogen in corn (Paz et al., 1999). In the special case of nutrient management, crop model simulations are also useful for understanding the role of nutrients in both crop production and environmental quality. Long-term prescriptions for reducing the environmental impacts of corn production have been developed using crop models to simulate yield and nitrogen left behind for many seasons of historical weather (Thorpe et al., 2005b). Other applications of crop growth models in precision agriculture include yield forecasting (Hodges et al., 1987; Liu et al., 1989), yield gap analysis (Paz et al., 2004), and simulation of crop response to climate change or genetic modification (Boote et al., 1996).

For all crop modeling applications, the underlying assumption is that the model can accurately simulate the processes occurring within the agricultural system of interest. However, no model can simulate these processes perfectly. To improve the accuracy of simulation results, model calibration techniques are employed to fine-tune model input parameters over an expected range for site-specific conditions. To calibrate a crop model, one or several model input parameters are adjusted in an iterative fashion to solve for the parameter values that minimize error between simulated model output and observed quantities. For example, Liu et al. (1989) adjusted phenological coefficients for maize until the simulated dates for silking and maturity closely matched observed dates. Also, Jones and
Carberry (1994) adjusted potential kernel number and potential kernel growth rate to minimize error between measured and simulated corn yield. Similar techniques have been employed to calibrate crop models for agricultural systems having highly-restrictive claypan soil layers (Fraisse et al., 2001) and/or tile drains (Garrison et al., 1999). A variety of techniques have been explored for implementing calibration procedures to estimate model input parameters. In general, most techniques involve an optimization algorithm that solves for the parameter set that maximizes or minimizes an objective function (Jones and Carberry, 1994; Calmon et al., 1999; Paz et al., 1999; Irmak et al., 2001). Typically, the root mean square error (RMSE), or related error statistics, between simulated output and measured values is the objective function to be minimized in the calibration of crop models (Kobayashi and Salam, 2000; Gauch et al., 2003).

Model validation is an assessment of the ability of a calibrated model to perform adequately for calibration-independent datasets. On the matter of model validation, two conflicting beliefs exist among researchers. Those who are skeptical of model performance believe that validation is an important step for demonstrating that a calibrated model can provide acceptable simulations for datasets not used in the calibration. To validate a model, measured observations are often partitioned into two groups: one group for model fitting during the calibration phase and the other for model testing during the validation phase. By setting aside a portion of the measured data for model validation, it is possible to assess whether the model can be confidently used to simulate crop responses under conditions other than those explicitly defined in the calibration dataset. Other researchers feel that removing information from the measured data available for model calibration reduces the ability of the model to simulate the true processes occurring within the agricultural system. Also, given that the sample size of measured datasets is often limited by the time and/or money available to collect the data, modelers are often reluctant to withhold information during the calibration phase. However, if all available measured data are used for model fitting, any effort to validate the model is fruitless, because repeated use of any portion of the calibration dataset in the validation phase results in statistically biased estimates of model performance. As a result of this dilemma, the model validation step has often been approached less rigorously in
comparison to model calibration, because measured datasets are usually not comprehensive enough to allow for both steps to occur (Jones and Carberry, 1994).

Leave-one-out (LOO) cross validation (Efron and Gong, 1983; Efron and Tibshirani, 1998) is a statistical procedure that can be used to validate crop models in the instance of limited measured data. With the LOO cross validation technique, observations in the measured dataset are iteratively and exhaustively used for both model calibration and model testing, resulting in an estimate of model predictive performance that is more reliable than estimates from the two-group partition method and less biased than estimates derived from calibration-dependent datasets (Jones and Carberry, 1994). Given a measured dataset having $n$ total observations, LOO cross validation requires the model to be calibrated and independently validated $n$ times. For the $j$th measured observation, model calibration procedures are performed using the $n - 1$ other measured observations, leaving out the $j$th observation each time. After fitting the model with the $j$th observation left out, the fitted model is used to simulate the $j$th observation, and the error between observed and simulated values is calculated as a measure of model performance. This process is repeated until all $n$ observations have been left out and used for model validation one time. The LOO cross validation estimate of model simulation accuracy is then calculated as the root mean square error of prediction (RMSEP) between observed and simulated model output for $n$ independent model validation procedures. Cross validation techniques have been successfully used for validation of crop growth models in the work of Jones and Carberry (1994) and Irmak et al. (2000); however, these techniques have not been used within a precision agriculture framework.

Recent work at Iowa State University has focused on the use of crop growth models to study precision agriculture questions on the sub-field-level scale (Batchelor et al., 2002). To utilize crop models in this way, researchers have developed optimization techniques to calibrate crop model parameters uniquely for many grid cells or management zones across the field area. These optimization techniques solve for the parameter set in each grid cell that minimizes the RMSE between measured and simulated yield over multiple growing seasons. The greatest limitation in this endeavor has been that many growing seasons are required to generate measured yield datasets of adequate sample size. Assuming that a yield monitor is
available and functioning properly on the day of harvest, only one new yield value can be
generated in each grid cell per year for continuous corn. Similarly, for a corn/soybean crop
rotation, typical of agriculture in the midwestern United States, the sample size of a measured
corn yield dataset in a grid cell can only be increased by five every decade. Because of the
sample size limitations of measured yield datasets in this case, the objective of this work was
to explore LOO cross validation as a procedure for evaluating the ability of the CERES-
Maize crop model to simulate corn yield across an Iowa cornfield divided into 100 equally-
sized grid cells. A second objective was to examine the magnitude of and to identify the
causes of spatial variability in parameter estimates and model performance across the field.

3.3 Materials and Methods

3.3.1 Crop Growth Model

The CERES-Maize crop growth model (Jones and Kiniry, 1986) is a computer
program that utilizes carbon, nitrogen, and water balance principles to simulate the processes
that occur during the growth and development of corn plants within an agricultural system.
The model calculates the growth and development of corn plants within a homogeneous area
on a daily time step, and the final crop yield is computed on the date of harvest. Inputs
required for model execution include management practices (plant genetics, plant population,
row spacing, planting and harvest dates, and fertilizer application amounts and dates),
environmental factors (soil type, drained upper limit, lower limit, and saturated hydraulic
conductivity), and weather conditions (daily minimum and maximum temperature, solar
radiation, and precipitation). CERES-Maize has been widely used to simulate the collective
effect of plant genetics, management practices, weather, and soil conditions on the growth,
development, and yield of corn plants. The model has been shown to perform adequately on
plot-level, field-level, and regional scales for a wide variety of corn hybrids, climatic
conditions, and soil types around the world (Hodges et al., 1987; Carberry et al., 1989; Liu et
al., 1989; Jagtap et al., 1993; Pang et al., 1998; Garrison et al., 1999; Paz et al., 1999; Fraisse
et al., 2001).
3.3.2 Apollo

Given the usefulness of crop model simulations for precision agriculture applications, researchers at Iowa State University have recently developed a new decision support software called Apollo (Batchelor et al., 2004). Designed to automate the use of crop models for solving various problems in precision agriculture, Apollo offers a Windows-based environment for calibration and validation of the DSSAT family of crop models on the sub-field-level scale. To calibrate the model, the simulated annealing algorithm (Corana et al., 1987; Goffe et al., 1994) has been implemented to solve for the parameter set that minimizes the RMSE between measured and simulated yield. Up to ten model input parameters can be optimized for several hundred equally-size grid cells or variably-sized management zones. The software utilizes the model calibration results in each zone to generate nitrogen and plant population prescriptions, to forecast yields, and to explain the causes of yield spatial variability across agricultural fields.

3.3.3 Data Preparation

The study area was a 20.25 ha section of a production cornfield near Perry, IA, USA (41.93080° N, 94.07254° W). This area was divided into 100 grid cells, each 45 m by 45 m in size. A digitized soil survey indicated that five primary soil types were present in the study area: Canisteo silty clay loam, Clarion loam, Nicollet loam, Harps loam, and Okoboji silty clay loam. Estimates of the physical properties for these soils were obtained from two sources. Ratliff et al. (1983) provided the drained upper limit (DUL) (cm$^3$ cm$^{-3}$) and lower limit for various soil textures. Values for the saturated hydraulic conductivity ($K_{SAT}$) (cm d$^{-1}$), bulk density (BD) (g cm$^{-3}$), and soil pH at various soil depths were obtained from the county soil survey (USDA-SCS, 1981). Saturated moisture content (SAT) (cm$^3$ cm$^{-3}$) was calculated from BD using

$$\text{SAT} = 0.92 \times \left(1 - \frac{\text{BD}}{2.65}\right).$$

(3.1)

Each of the 100 grid cells was assigned the soil properties for the soil type that covered the largest area within the grid cell (Figure 3.1). A Visual Basic for Applications (VBA) script was created within the ArcGIS 9 software to create the grid layout, clip the digital soil survey
by grid cell, determine the soil type covering the largest area, and write the soil parameters to a soil file for crop model runs (Thorp et al., 2005a).

Figure 3.1. Soil types for the 20.25 ha study area divided into 100 grid cells

Yield and weather data were collected directly at the site during five growing seasons. This information was used to develop the yield and weather files necessary for crop model runs. Corn yield was measured in each of the 100 grid cells using a yield monitor on a grain combine at the conclusion of the 1994, 1996, 1998, 2000, and 2002 growing seasons.
Measured yield data was used to optimize crop model input parameters by minimizing the RMSE between observed and simulated yield during the model calibration phase. The data was also used to calculate RMSEP between observed and simulated yield using LOO cross validation. The VBA script in ArcGIS 9 was extended to clip the yield data by grid cell, calculate the average yield for each grid cell, and write the yield files to a disk. Weather data, including solar radiation, maximum and minimum daily temperature, and precipitation amount, was collected daily using a weather station directly at the site. This information was used to develop weather files for model simulations of the five seasons of corn production.

Soil water content and initial nutrient levels were not available for this site. Appropriate values were assumed and assigned uniformly to each grid cell across the study area. Initial soil water content was set to 0.3 cm$^3$ cm$^{-3}$, a value just below the DUL for the soils in the field. Initial nutrient levels were set arbitrarily to 0.1 g elemental N, P, and K per Mg soil. For the purpose of this study, it was assumed that the soil profile contained only a negligible amount of nutrients at the beginning of the season, and that spring-applied fertilizer applications served to raise the nutrient concentrations to levels that would support plant growth. Plant population information was collected during the 1996 growing season only, and the 1996 values for plant population in each grid cell were used to approximate plant population in the grid cells for the other four years. Model inputs for management practices, including planting date, harvest date, and fertilizer application rates and dates, were set according to the producer’s actual practice in each of the five growing seasons.

3.3.4 Cross Validation

Using the Apollo decision support software to automate crop model runs spatially across the 100 grid cells, CERES-Maize simulations of corn yield spatial variability were evaluated by independently applying LOO cross validation techniques to each grid cell. Since five seasons of measured corn yield were available, the Apollo calibration module was used to optimize crop model input parameters in each grid cell five different times, leaving one season of measured data out of the calibration each time. For each of the five groups of four growing seasons, optimum parameters for each grid cell were determined by minimizing
the RMSE between measured and simulated yield. In the context of this study, the RMSE for each group of four growing seasons can be defined as

\[
RMSE_i = \left( \frac{1}{n} \sum_{j=1}^{n} (Y_{m,i,j} - Y_{s,i,j})^2 \right)^{0.5},
\]

where \( Y_{m,i,j} \) is the measured yield and \( Y_{s,i,j} \) is the simulated yield in the \( i \)th grid cell for the \( j \)th of \( n \) seasons of yield data. Two model parameters were optimized with this calibration procedure, including the \( K_{SAT} \) of the deep impermeable layer (cm day\(^{-1}\)) and the effective tile drainage rate (day\(^{-1}\)). These parameters govern the movement of water through the soil profile and can be used to mimic the effect of water stress on corn plants. Water stress is typically the greatest factor influencing yield loss and spatial variability in the rain-fed agricultural systems of Iowa. Although a first estimate of the \( K_{SAT} \) value was obtained from the county soil survey, the range of parameter values for most soil types is very wide. The calibration procedure served to fine-tune this parameter to more accurately represent the water table dynamics at the specific location of each grid cell. If a grid cell was properly drained, the calibration procedure generated a large value for the \( K_{SAT} \) parameter. In this case, excess water is more quickly lost out the bottom of the profile and water tables are kept low or never form, which allow roots to grow deep in the soil profile. The calibration procedure would give small values for the \( K_{SAT} \) parameter if a grid cell were poorly drained. This causes water to move more slowly through the bottom soil layer, water tables are kept high, and roots grow to more shallow depths within the soil profile. The effective tile drainage rate controls the speed at which water is lost through a tile when the water table is above the tile line. Because the calibration procedure was run five times leaving one season of yield data out each time, five unique optimized parameter sets were generated during the calibration phase of the LOO cross validation procedure. Apollo stored each of these optimum parameter sets in a database for future use.

To complete the LOO cross validation procedure, the Apollo validation module was used to simulate corn yield in each grid cell for the growing season that was left out of each of the five calibration trails. The results of this analysis were used to calculate, in each grid cell, the LOO cross validation estimate of model prediction error or the RMSEP for the model. In the context of this study, RMSEP can be defined as
\[ \text{RMSEP}_i = \left( \frac{1}{n} \sum_{j=1}^{n} (Y_{m_{i,j}} - Y_{s_{i,j}})^2 \right)^{0.5}, \tag{3.3} \]

where \( Y_{m_{i,j}} \) is the measured yield value for the \( i \)th grid cell in the \( j \)th of \( n \) growing seasons, and \( Y_{s_{i,j}} \) is the simulated yield value in the \( i \)th grid cell obtained using the model that was calibrated by leaving out the data for the \( j \)th growing season. To consider the model performance for an individual validation year, the validation error was calculated as the absolute value of the difference between \( Y_{m_{i,j}} \) and \( Y_{s_{i,j}} \).

### 3.3.5 Parameter Variation

The LOO cross validation procedure generated five unique sets of the two model input parameters for each grid cell. Each of these parameter sets minimizes the error between measured and simulated yield in a grid cell for the four growing seasons used in the calibration. Theoretically, these five optimized parameter sets should be identical within each grid cell, because the optimizer would arrive at the same result regardless of the growing seasons used in the calibration if model simulations were perfect. However, since this is not the case in reality, there will exist some variability in the five estimates for the two model parameters in each grid cell. This variability in parameter estimates represents another measure of model performance. If the model performed well at simulating corn yield in a grid cell, the optimized parameter estimates should not be significantly changed by alternatively leaving one season of data out of the calibration, and the parameter variation should be low. On the other hand, if the model performed poorly at simulating corn yield in a grid cell, the optimized parameter estimates might change significantly when the data from one growing season is left out of the calibration, and the variation in parameter estimates should be higher.

A variety of techniques were used to describe the variation in parameter estimates in this work. The first approach involved simple univariate statistics for individual parameters. The sample means and standard errors were computed independently for each of the two parameters in each grid cell. These statistics were then used to compute \( 100(1-\alpha)\% \)
univariate confidence intervals around the sample mean for parameters in each grid cell based on Student’s t-distribution. Confidence intervals were calculated according to

\[ \bar{x}_{i,k} \pm t_{\alpha/2,n-1} \left( s_{i,k} / \sqrt{n} \right), \]  

where \( \bar{x}_{i,k} \) and \( s_{i,k} \) are the sample mean and sample standard deviation for the \( k \)th parameter in the \( i \)th grid cell, \( n \) is the sample size, and \( t_{\alpha/2,n-1} \) is the value of Student’s t-distribution at \( \alpha/2 \) on \( n-1 \) degrees of freedom. Since optimized parameters existed as a set of two, multivariate techniques were also used to describe the parameter variation and to compute 100(1-\( \alpha \))% bivariate confidence ellipses around the sample means for parameter estimates in each grid cell. The procedure for developing 100(1-\( \alpha \))% confidence regions around multivariate parameter sets is described in Johnson and Wichern (2002). The 100(1-\( \alpha \))% confidence region around the sample mean of the two-dimensional parameter set, as applied to the \( i \)th grid cell in this two-parameter study, is the ellipse determined by all \((\mu_1, \mu_2)\) such that

\[ n \begin{bmatrix} \bar{x}_{i,1} - \mu_1 \\ \bar{x}_{i,2} - \mu_2 \end{bmatrix}^T S_i^{-1} \begin{bmatrix} \bar{x}_{i,1} - \mu_1 \\ \bar{x}_{i,2} - \mu_2 \end{bmatrix} \leq \frac{p(n-1)}{(n-p)} F_{p,n-p}(\alpha), \]  

where \( p \) is the number of parameters, \((\bar{x}_{i,1}, \bar{x}_{i,2})\) is the set of sample means for the two parameters, \( F_{p,n-p}(\alpha) \) is the value of the F-distribution at \( \alpha \) on \( p \) and \( n-p \) degrees of freedom, and \( S_i^{-1} \) is the inverse of the 2 x 2 variance-covariance matrix for the parameter estimates in the \( i \)th grid cell. Explicitly, \( S_i \) is defined as

\[ S_i = \frac{1}{n-1} \begin{bmatrix} \sum_{j=1}^{n} (x_{i,j} - \bar{x}_{i,1})^2 & \sum_{j=1}^{n} (x_{i,j} - \bar{x}_{i,1})(x_{i,j} - \bar{x}_{i,2}) \\ \sum_{j=1}^{n} (x_{i,j} - \bar{x}_{i,1})(x_{i,j} - \bar{x}_{i,2}) & \sum_{j=1}^{n} (x_{i,j} - \bar{x}_{i,2})^2 \end{bmatrix}, \]  

where \((x_{i,1,j}, x_{i,2,j})\) is the set of two parameter estimates obtained with the \( j \)th growing season left out of the calibration. In order to solve Equation 3.5, the matrices must be expanded, and the quadratic equation is used to write \((\bar{x}_{i,1} - \mu_1)\) in terms of \((\bar{x}_{i,2} - \mu_2)\). Then, since \( \bar{x}_{i,1} \) and \( \bar{x}_{i,2} \) are known, the boundary of the confidence ellipse can be determined by alternatively
adjusting the value of $\mu_2$ around $\bar{x}_{i,2}$ and solving the quadratic equation for $\mu_1$. Several measures of multivariate parameter variation were also obtained from $S_i$ for each grid cell. As defined in Johnson and Wichern (2002), the generalized variance was computed as the determinant of $S_i$, and the total variance was computed as the trace of $S_i$. In addition, the area ($A$) of the $100(1-\alpha)%$ bivariate confidence region for the $i$th grid cell was computed as

$$A_i = \pi \cdot \chi^2_{p,\alpha} |S_i|^{0.5},$$

(3.7)

where $\chi^2_{p,\alpha}$ is the value of the chi-square distribution at $\alpha$ on $p$ degrees of freedom. All confidence intervals and confidence ellipses were computed using a $\alpha$-level of 0.05.

### 3.4 Results

Results of all of the calibration runs indicated that the optimization routine had performed remarkably well at adjusting $K_{SAT}$ of the deep impermeable layer and the effective tile drainage rate to simulate spatial yield variability across the field. The field level average RMSEs across all 100 grid cells for each calibration did not exceed 500 kg ha$^{-1}$ with standard deviations no greater than 300 kg ha$^{-1}$ (Table 3.1), indicating that corn yield variability was simulated with less than 10% error in most grid cells during the calibration phase. Comparing the field average RMSE for each model calibration against the total rainfall for the growing season that was left out of the calibration demonstrates that the model had the greatest difficulty simulating growing seasons with extreme weather, such as drought or flood (Table 3.1). Of the five calibrations, the lowest field level average RMSE was achieved when the driest of all the growing seasons, year 2000, was left out of the calibration. In this growing season, only 345.8 mm of rain fell on the study site, nearly half of the yearly average for this area. The field level average RMSE for this calibration was 268.9 kg ha$^{-1}$. The second lowest RMSE occurred when the wettest of the five growing seasons, 1998, was left out of the calibration. In 1998, 815.8 mm of rain fell on the study site, and the field level average RMSE was 389.0 kg ha$^{-1}$ when data from this growing season was left out of the calibration. For the three remaining calibrations in which both the wettest and driest years were left in, the field level average RMSEs were greater and ranged from 397.0 kg ha$^{-1}$ to 461.9 kg ha$^{-1}$. Since the model performed better when growing seasons with
extreme rainfall conditions were left out, it is expected that the model had difficulty simulating the effect of water stress on corn yield for those growing seasons.

In comparing the field level average calibration RMSE to the field level average validation error for the season left out of the calibration, greater errors were seen as expected (Table 3.1). In addition, the standard deviations in field level average validation error were larger than the standard deviation in field level average calibration RMSE. This occurred because the model data used for validation is independent of the data used to optimize the parameters during model calibration. The tendency for the model to perform more poorly for

<table>
<thead>
<tr>
<th>Years Used for Calibration</th>
<th>Validation Year</th>
<th>Total Rainfall for the Validation Year (mm)</th>
<th>Calibration RMSE over 100 grid cells (kg ha⁻¹)</th>
<th>Validation error over 100 grid cells (kg ha⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Average Std. Dev.</td>
<td>Average Std. Dev.</td>
</tr>
<tr>
<td>1994, 1996, 1998, 2002</td>
<td>2000</td>
<td>345.8</td>
<td>268.9 146.7</td>
<td>1388.03 738.1</td>
</tr>
<tr>
<td>1994, 1996, 2000, 2002</td>
<td>1998</td>
<td>815.8</td>
<td>389.0 242.5</td>
<td>1130.1 962.0</td>
</tr>
<tr>
<td>1994, 1998, 2000, 2002</td>
<td>1996</td>
<td>756.9</td>
<td>449.6 269.2</td>
<td>675.7 929.3</td>
</tr>
<tr>
<td>1996, 1998, 2000, 2002</td>
<td>1994</td>
<td>716.5</td>
<td>397.0 229.8</td>
<td>1039.7 883.2</td>
</tr>
</tbody>
</table>

the growing seasons with extreme weather conditions was also apparent in the validation results. The greatest error occurred when the model was validated with the data from 2000, the drought year. In this case, the model was calibrated using four growing seasons that all had adequate rainfall, and the optimum parameters generated by this calibration were less flexible for simulating the more extreme water stress conditions that occurred during the drought in 2000. As a result, the model was less able to accurately simulate corn yield, and greater validation error was seen for this year. The second greatest validation error occurred for 1998, the wettest year in the dataset. In this case, the model was calibrated using growing
seasons having less rainfall, and the optimum parameters for these years were less flexible for simulating the effect of greater levels of precipitation in 1998. Since the highest validation errors occurred for the wettest and driest seasons in the dataset, the lowest validation errors occurred for the three remaining growing seasons (1994, 1996, and 2002) that all had relatively moderate rainfall. Lower validation errors were obtained for these three growing seasons, because their respective calibration datasets each contained both the wettest and the driest year. Thus, model calibration was performed over a wide range of weather conditions, and the fitted parameters values were flexible enough to more accurately simulate crop yield in growing seasons having moderate weather.

Since the LOO cross validation procedure was performed independently for each grid cell, a map of RMSEP shows the spatial variation of the model’s ability to simulate corn yield across this agricultural field (Figure 3.2). Because grid cells with similar RMSEP tend to cluster together, there is evidence that spatial patterns in model performance exist across the field. Overlaying a topographic contour on the map suggests that model performance may be related to topography. For instance, one cluster of large RMSEP occurs at an easting of 411,001 m and a northing of 4,642,450 m. As indicated by the topography contour, this marks the location of a distinct area of converging flow that cuts through the southern half of the study area. Greater RMSEP could result from reductions in plant population as a result of washout during heavy rains in the early season. Another cluster of large RMSEP occurs at an easting of 411,070 m and a northing of 4,642,350 m. Here, the contour lines are closer together, indicating that this portion of the field is more sloped than other areas. At this location, higher RMSEP could result from the inability of the model to account for surface run-on and subsurface water flow between neighboring grid cells, the dynamics of which would be more significant on a sloped topography.

The mean model parameter values for effective tile drainage rate also exhibited spatial patterns that were related to topography (Figure 3.3). When the effective tile drainage rate is high, water is lost more quickly out the tile when the water table is above the specified tile drainage depth. When the effective tile drainage is low, drainage due to tile drains occurs more slowly when water is above the specified depth of tile drains. During growing seasons of adequate rainfall, it is expected that the water table will be high relative to the soil surface,
and this is especially true for the areas of the field at lower elevations. At the location of the converging flow area in the southern portion of the field (easting of 411,001 m and a northing of 4,642,450 m), there exists a strip of grid cells with high mean values for effective tile drainage rate. Since many of the growing seasons in this study had average and above average rainfall, the grid cells located in the area of converging flow required higher values
for effective tile drainage rate in order to reduce soil water content such that the error between measured and simulated yield could be minimized. Grid cells tending to have lower mean values for effective tile drainage rate tended to occur in the southeastern corner and in the east-central portion of the study area. The location of these grid cells generally corresponded to areas of the field at higher elevations, and thus tile drainage either has little effect or does not physically exist within these grid cells.

Figure 3.3. Spatial variation of mean effective tile drainage rate across 100 grid cells with topographic contour lines overlaid
Spatial relationships between topography and the mean model parameter values for $K_{\text{SAT}}$ of the deep impermeable layer in each grid cell were also apparent in the results (Figure 3.4). A high value for $K_{\text{SAT}}$ in a grid cell indicates that the model required rapid movement of water out the bottom of the soil profile in order to minimize the error between measured and simulated yield. Lower values for $K_{\text{SAT}}$ indicated that the model required slow movement of water out the bottom of the profile. The combination of high rainfall and low
$K_{SAT}$ allows for the formation of a simulated water table that can serve to limit root growth in the early season and reduced yield at harvest. In dry years, there may not be enough water available to form a water table. As expected, the grid cell located in the center of the area of converging flow has a low value for the $K_{SAT}$ parameter. Since this grid cell is located in an area of relatively low elevation, there is an increased chance that water table effects could reduce corn yield. Thus, the model uses a low $K_{SAT}$ value in this grid cell to raise the water table such that roots can grow only to shallow depths in the soil profile and corn yield is subsequently reduced. In the southeastern corner and in the center portions of the field, there are collections of grid cells having high mean values for the $K_{SAT}$ parameter. As before, the locations of these grid cells correspond to areas of the field with higher elevation. As a result, water table effects do not seriously hamper corn yield in these areas of the field, and the high values for $K_{SAT}$ prevent yield-reducing water tables from forming in these grid cells.

In addition to RMSEP, another measure of model performance in each grid cell is the variation in parameter values obtained using LOO cross validation. Since two parameters were calibrated in this work, the parameter variation can be characterized as the area of the ellipse formed by the $100(1-\alpha)$% confidence region around the mean parameter values in each grid cell. The ellipse for Grid cell #13 provides an example (Figure 3.5). Mapping the confidence ellipse area for each grid cell with topographic contours overlaid gives another indication of the relationship between model performance and topography across the field (Figure 3.6). As with RMSEP in Figure 3.2, a cluster of grid cells with high variability in optimized parameter values exists at an easting of 411,070 m and a northing of 4,642,350 m. Here, the contour lines are bunched more closely together indicating more rapid elevation changes that could be linked to the failure of the model to account for surface run-on and subsurface water flow between neighboring grid cells. Unfortunately, it is difficult to pick out many other patterns that are similar between model performance errors measure by RMSEP (Figure 3.2) and ninety-five percent confidence ellipse area (Figure 3.6).

To further understand the spatial patterns of the model's performance, the median RMSEP and the median ellipse area was found for grid cells having the same dominant soil type. In addition, the mean and standard deviation were computed for effective tile drainage rate parameters and $K_{SAT}$ parameters in grid cells having the same soil type. Because the
water dynamic properties within a grid cell should not change from year to year, a larger variability, or standard deviation, in optimized parameter values represents greater difficulty on the part of the model to consistently find the same value for a parameter in spite of using different growing seasons in the calibration. Results of this analysis were consistent with the topologic trends for each soil type. The median RMSEP and the median ellipse area for grid cells dominated by Harps loam and Clarion loam were greater than that of the other soil types, indicating that the model had more difficulty explaining yield variability in grid cells having these soil types (Table 3.2). In addition, the standard deviations for soil hydraulic conductivity were greater for these soil types, indicating the model was less consistent in optimizing this parameter when using different datasets in the calibration. Clarion loam is a
gently sloping, well-drained soil on convex upland knolls (USDA-SCS, 1981), meaning the soil type is generally present on the sideslopes of the swell and swale topography typical of the central Iowa countryside. Therefore, a possible explanation for the greater median

Figure 3.6. Spatial variation of ninety-five percent confidence ellipse area across 100 grid cells with topographic contour lines overlaid

RMSEP, greater median ellipse area, and greater variability in optimized values for soil hydraulic conductivity in Clarion grid cells is that the model does not adequately account for
surface run-on and sub-surface water flow between neighboring grid cells, the dynamics of which would be more significant for the sloped Clarion soil type. Harps loam is a poorly drained soil that surrounds closed depressions on upland flats (USDA-SCS, 1981). This soil type may also exhibit a slightly sloped topography, although not as pronounced as the Clarion loam. The median RMSEP for grid cells dominated by the Okoboji silty clay loam were significantly lower than that of the other soil types. Okoboji silty clay loam is a very poorly drained soil occurring in concave depressions on uplands (USDA-SCS, 1981). Although ponded water can collect on these soils during times of heavy rainfall, low RMSEP values for the Okoboji dominated grid cells indicate that the model performed well in mimicking the water flow dynamics in these poorly drained areas.

Table 3.2. Median RMSEP, mean and standard deviation of effective tile drainage rate (ETDR), mean and standard deviation of saturated hydraulic conductivity (SHC), and median ellipse area for each soil type

<table>
<thead>
<tr>
<th>Soil Type</th>
<th>Median RMSEP (kg ha⁻¹)</th>
<th>Mean ETDR (day⁻¹)</th>
<th>Standard Deviation ETDR (day⁻¹)</th>
<th>Mean SHC (cm day⁻¹)</th>
<th>Standard Deviation SHC (cm day⁻¹)</th>
<th>Median Ellipse Area (cm day⁻²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clarion Loam (138B)</td>
<td>1156.5</td>
<td>0.1049</td>
<td>0.0527</td>
<td>1.0377</td>
<td>0.7755</td>
<td>0.3560</td>
</tr>
<tr>
<td>Canisteo Silty Clay Loam (507)</td>
<td>1084.9</td>
<td>0.1160</td>
<td>0.0682</td>
<td>0.3415</td>
<td>0.2951</td>
<td>0.0691</td>
</tr>
<tr>
<td>Nicollet Loam (55)</td>
<td>1098.0</td>
<td>0.0909</td>
<td>0.0523</td>
<td>0.4309</td>
<td>0.2206</td>
<td>0.0707</td>
</tr>
<tr>
<td>Okoboji Silty Clay Loam (6)</td>
<td>942.9</td>
<td>0.1149</td>
<td>0.0569</td>
<td>0.5513</td>
<td>0.5048</td>
<td>0.0692</td>
</tr>
<tr>
<td>Harps Loam (95)</td>
<td>1175.9</td>
<td>0.1217</td>
<td>0.0559</td>
<td>0.6747</td>
<td>0.5582</td>
<td>0.2619</td>
</tr>
</tbody>
</table>

3.5 Conclusions

The calibration and validation modules of the Apollo software proved to be valuable for rapidly generating the simulation output necessary for implementing the LOO cross validation procedure. Results demonstrated the importance of having growing seasons with extreme weather conditions contained in the calibration dataset, because the ability of the model to simulate an independent dataset is improved when the calibration dataset spans a wide range of weather conditions for a study site. Likewise, model performance will decline
when attempting to simulate growing seasons with weather conditions much different from that contained in the calibration dataset. Given this result, a major issue for applying crop models to solve problems in agriculture is obtaining measured information for model calibration during the extreme growing seasons. Once a measured dataset adequately represents the majority of weather conditions that might be seen at a particular study site, a researcher can be more confident in using the model to simulate independent growing seasons in either the past or future. However, until a well-rounded measured dataset for a study site is obtained, there is a greater chance that the model will perform poorly when attempting to simulate a more extreme growing season. This could be especially troubling if, when using the model to predict future events, a more extreme year was encountered. Often, it may take years or decades of work at a study site to obtain such a dataset. As the size of the measured dataset for a particular study site increases, an interesting area of research will be to determine the level of influence that the dataset for a particular growing season contributes to the overall optimization. In this way, datasets that contribute nothing can be removed to reduce bulkiness in measured datasets and to reduce the time required to calibrate the crop model. In regards to the spatial variability of model performance across this study site, it can be concluded that topography is a possible cause for this. Further work is needed to determine the proportion of model error due to topography and, if necessary, to explore methods for appropriately simulating the effect of topography on the water balance within the model.

3.6 References


USDA-SCS, 1981. Soil Survey of Boone County, Iowa. United States Department of Agriculture – Soil Conservation Service (USDA-SCS), Iowa Agriculture and Home Economics Experiment Station Cooperative Extension Service, Iowa State University, and State of Iowa Department of Soil Conservation, Ames, IA, USA.
CHAPTER 4. METHODOLOGY TO LINK PRODUCTION AND ENVIRONMENTAL RISKS OF PRECISION NITROGEN MANAGEMENT STRATEGIES IN CORN

A paper accepted by *Agricultural Systems*

Kelly R. Thorp, William D. Batchelor, Joel O. Paz, Brian L. Steward, Petrutza C. Caragea

4.1 Abstract

A new decision support system called Apollo, which runs the CERES-Maize crop growth model, was used to study the corn (*Zea mays* L.) yield response and the nitrogen (N) dynamics of a cornfield in central Iowa, USA. The model was calibrated to minimize error between simulated and measured yield over five growing seasons. Model simulations were then completed for 13 spring-applied N rates in each of 100 grid cells with varying soil properties. For each N rate and grid cell, simulations were repeated for 37 years of historical weather information collected near the study site. Model runs provided the crop yield and unused N in the soil at harvest for all combinations of N rate, grid cell, and weather year. Using these simulated datasets, a methodology involving cumulative probability distributions was developed such that the yield and unused N resulting from each N rate applied in each grid cell could be directly linked according to their probability of occurrence over the 37 simulated growing seasons. These cumulative probability distributions were used to evaluate the economic and environmental risks of two alternate precision N management strategies for the study area. In the first strategy, N rates were selected to maximize the producer’s marginal net return in each grid cell. The environmental cost of this management strategy, in terms of N left behind, was determined to be 56.2 kg ha\(^{-1}\) on average over all grid cells. In the second strategy, N rates were selected to insure that the amount of N left in the soil at harvest would not exceed 40 kg ha\(^{-1}\) in 80% of growing seasons. The producer’s opportunity cost for reducing N rates to achieve this environmental objective was calculated to be $48.12 ha\(^{-1}\) on average over all grid cells. The overall goal of this work was to develop a methodology for directly contrasting the production and environmental concerns of N management in agricultural systems. In this way, N management plans can be designed to achieve a proper balance between production and environmental goals.
4.2 Introduction

With the increased use of yield monitors on grain combines in the past decade (Searcy et al., 1989), crop yield has repeatedly been shown to exhibit substantial spatial variation across individual fields (Jaynes and Colvin, 1997). In addition, airborne and satellite remote sensing imagery has shown similar variation in crop growth and development throughout the growing season (GopalaPillai and Tian, 1999). Such datasets have demonstrated the potential for variable-rate applications of nitrogen (N) fertilizer, based on the site-specific crop need. Applying N fertilizer site-specifically also makes sense from an environmental perspective. Bakhsh et al. (2000) identified several properties of the agricultural landscape that altered its susceptibility to movement and loss of nitrate-nitrogen (NO$_3$-N). These include soil type, topography, soil moisture, tile drainage, and tillage practices. Thus, a robust method for determining appropriate N application rates must consider the spatial variability of the agricultural landscape's NO$_3$-N loss risk, as well as the spatial variability of the crop's N need. In this way, N prescriptions can be tailored to address both production and environmental concerns.

In addition to the spatial aspects of N management, there is also a complex temporal problem that arises due to the unpredictable nature of weather patterns. Precipitation events drive the movement of NO$_3$-N through the agricultural system, and rainfall is necessary for the crop to uptake NO$_3$-N from the soil. However, problems arise when precipitation events, NO$_3$-N availability, and crop need do not coexist in time (Dinnes et al., 2002). For example, in the midwestern United States, N is most commonly applied in the fall or spring, prior to planting corn. Nitrogen applied at these times has the greatest potential for loss to the environment, because snow melt and heavy rains in the spring season can move NO$_3$-N out of the agricultural system prior to crop uptake. A similar problem exists during seasons of drought. For this case, suppose NO$_3$-N is made available through side-dress applications of N fertilizer at mid-season. Although NO$_3$-N is now available during the time of peak N demand, the lack of water prevents the crop from removing all the NO$_3$-N from the soil. The excess NO$_3$-N is then available for loss during precipitation events that occur after harvest, when the crop no longer needs it. Unfortunately, when making N management decisions, knowledge of future weather patterns and precipitation events is limited to the accuracy of
seasonal forecasts. However, large sets of historical weather data now exist for many portions of the world, and these datasets can be used as an indicator of probable future weather patterns for an area. In this way, historical weather data becomes a useful set of information for the development of N management strategies that are conscious of the influence of weather patterns on NO₃-N movement through the agricultural system.

Over the past decade, researchers have focused on a wide variety of methods for developing site-specific N prescriptions. An arsenal of sensing techniques has been employed for identifying N deficient areas of crops, including airborne and satellite remote sensing (Blackmer et al., 1996; Flowers et al., 2003), multispectral camera systems on ground vehicles (Noh et al., 2003), and chlorophyll meter readings of individual corn leaves (Schepers et al., 1992). Although these sensing techniques have successfully identified N deficiencies, they do not effectively account for the spatially varying properties of the agricultural landscape or weather patterns that affect NO₃-N losses from the agricultural system. Other researchers have attempted to develop yield response functions by regressing crop yield against soil nutrient measurements, such as late-spring NO₃-N concentration (Katsvairo et al., 2003) and soil organic matter (Schmidt et al., 2002). High \( r^2 \) values for the relationship between crop yield and soil nutrient levels have not been consistently obtained with this approach, because the temporal aspects of N movement through the agricultural system cannot be adequately characterized by a single equation. As a result, soil nutrient concentrations based on point-in-time measurements have not been helpful for developing variable-rate N recommendations. The greatest limitation in these approaches is that none of them can adequately account for the fact that N movement depends heavily on the temporal pattern of weather encountered during the growing season.

The CERES-Maize crop growth model (Jones and Kiniry, 1986) is another tool that has been used to study precision N management for corn (Zea mays L.) (Paz et al., 1999; Batchelor et al., 2002). This model utilizes carbon, N, and water balance principles to simulate, in homogenous units, the daily processes that occur during plant growth and development. The final corn yield for the simulated growing season is then calculated on the harvest date. The model has been shown to adequately simulate corn growth, development, and yield on plot-level, field-level, and regional scales for many locations around the world.
Inputs required for model execution include management practices (plant genetics, plant population, row spacing, planting and harvest dates, and fertilizer application amounts and dates), environmental factors (soil type, drained upper limit, lower limit, saturated hydraulic conductivity, root weighting factor, and effective tile drain spacing), and weather conditions (daily minimum and maximum temperature, solar radiation, and precipitation). Since CERES-Maize utilizes N balances for crop growth analysis, it can be conveniently extended to calculate surface and subsurface NO$_3$-N losses. For example, the model has undergone several modifications such that NO$_3$-N in run-off (Gabrielle et al., 1995), tile flow (Garrison et al., 1999), and leaching (Gabrielle et al., 1996) can be simulated as part of the crop production process. Since CERES-Maize can collectively account for many of the spatial and temporal factors that affect crop yield and N movement through the agricultural system, it serves as a very useful and appropriate tool for developing N management strategies that address both the economic and the environmental concerns of corn production.

A new decision support system called Apollo runs CERES-Maize and other DSSAT crop models for management zones within a field (Batchelor et al., 2004). Apollo is an interface that can be used to calibrate and validate model parameters and execute model runs to achieve a variety of precision farming objectives, such as prescription analysis and yield gap analysis. In this work, the Apollo system was used to calibrate CERES-Maize and run N prescriptions for an Iowa cornfield divided into 100 grid cells.

The overall objective was to use the results of the prescription simulations to develop a methodology for estimating the economic and environmental trade-offs of N management strategies for this cornfield. The existence of a trade-off between the production and environmental concerns of N management is an important concept, because of the dual opposing role that N plays in crop production and environmental quality. Whereas N fertilizer is beneficial for maximizing crop production, unused N fertilizer that is lost from the agricultural system poses a threat to environmental quality, wildlife welfare, and human health. Therefore, N management strategies of the future must aim to find the appropriate balance between these opposing concerns. The first step in this endeavor is to develop a methodology for predicting how a particular N management strategy will affect corn yield and unused N remaining in the soil at harvest. With such a methodology, N management
strategies can be developed and implemented with a direct understanding of the cost to the producer and the cost to the environment. In addition, the methodology could aid in the development of environmental legislation and producer compensation programs that aim to reduce the environmental risk of agricultural N management.

4.3 Methods

4.3.1 Data Preparation

The study area included a 20.25 ha section of a production cornfield near Perry, IA, USA (41.93080° N, 94.07254° W). This area was divided into 100 grid cells, each 45 m by 45 m in size. A digitized soil survey indicated that five primary soil types were present in the study area: Canisteo silty clay loam, Clarion loam, Nicollet loam, Harps loam, and Okoboji silty clay loam. Estimates of the physical properties for these soils were obtained from two sources. Ratliff et al. (1983) provided the drained upper limit (DUL) (cm$^3$ cm$^{-3}$) and lower limit for various soil textures. In addition, values for the saturated hydraulic conductivity ($K_{SAT}$) (cm d$^{-1}$), bulk density (BD) (g cm$^{-3}$), and soil pH at various soil depths were obtained from the county soil survey (USDA-SCS, 1981). Saturated moisture content (SAT) (cm$^3$ cm$^{-3}$) was calculated from BD using

$$SAT = 0.92 \left(1 - \frac{BD}{2.65}\right).$$

Each of the 100 grid cells was assigned the soil properties for the soil type that covered the largest area within the grid cell (Figure 4.1). A Visual Basic for Applications (VBA) script was created within the ArcGIS 8.2 software to create the grid layout, clip the digital soil survey by grid cell, determine the soil type covering the largest area, and write the soil parameters to a soil file for crop model runs (Thorp et al., 2005a). This soil file was used for both the model calibration and the N rate prescription simulations.

Five seasons of measured corn yield were available for crop model calibration. Measured yield datasets were obtained using a yield monitor on a grain combine during the 1994, 1996, 1998, 2000, and 2002 growing seasons. The VBA script in ArcGIS was extended to clip the yield data by grid cell, calculate the average yield for each grid cell, and
write the yield files to a disk. The yield files were used to compare measured and simulated yield during the model calibration phase.

Figure 4.1. Soil types for the 20.25 ha study area divided into 100 grid cells

Weather files were created based on the availability of 37 years of historical weather data collected near Perry, IA. These historical weather datasets allowed for the simulation of N rate performance over the weather conditions of the past 37 growing seasons. In addition, weather information for 5 of the 37 growing seasons was used in the model calibration. For
years 1966 to 1995, weather data was collected at the Perry grain elevator, 10 km from the study site. This data was obtained from a historical weather database maintained by the Department of Agronomy at Iowa State University (http://mesonet.agron.iastate.edu). For years following 1995, weather data was collected using a weather station directly at the site.

Soil water content, initial nutrient levels, and plant population were not available for this site. Appropriate values were assumed and assigned uniformly to each grid cell across the study area. In addition, since individual growing seasons were simulated independently, initial conditions were specified uniformly for each growing season and carry over of soil water and nutrients between growing seasons was ignored. Initial soil water content for each simulation was set to 0.3 cm$^3$ cm$^{-3}$, a value just below the DUL for the soils in the field. Initial N levels were set arbitrarily to 0.1 g elemental N per Mg soil. For the purpose of this study, it was assumed that the soil profile contained only a negligible amount of N at the beginning of the season. In practice, a producer would subtract pre-season soil nutrient levels from the N fertilization rate recommendations generated with the simulation methodology developed in this work. Finally, plant population was set to 7.4 plants m$^{-2}$ based on the average of population measurements collected during the 1996 growing season. These approximations for soil water content, initial nutrient levels, and plant population were used for both the model calibration phase and for the N rate prescription simulation phase.

Management practice model inputs were changed between the model calibration phase and the N rate prescription simulation phase. To calibrate the model, the producer’s actual planting date, actual harvest date, and actual fertilizer application rates and dates were used for each of the 5 growing seasons available for calibration. For the N rate prescription simulations, the planting and harvest dates were assumed to be uniform across all 37 growing seasons. In this case, the dates of planting and harvest were set to April 25 and October 12, respectively, based on the average of the 5 years of known management practice dates for this producer and study site. Also, the model was set to apply all N on April 15 in each of the 37 seasons included in the prescription simulations. Values for N rate were left blank in the model input file, such that the Apollo decision support system could alternatively input various N rates to test during the N prescription simulations.
4.3.2 Model Calibration

Paz et al. (1999) developed a technique to calibrate the CERES-Maize crop growth model for tile-drained soils in the Midwestern United States. The technique implements the simulated annealing algorithm to adjust model input parameters to minimize the error between measured and simulated yield within an area of interest. In this work, this technique was implemented within the Apollo calibration module to calibrate two CERES-Maize model parameters: $K_{SAT}$ of the deep impermeable layer and effective tile drainage rate. Although a first estimate of the $K_{SAT}$ value was obtained from the county soil survey, the range of parameter values for most soil types in the survey is very wide. The calibration procedure served to fine-tune this parameter to more accurately represent the water table dynamics of each grid cell. If a grid cell was properly drained, the calibration procedure generated a large value for the $K_{SAT}$ parameter. In this case, excess water is more quickly lost out the bottom of the profile and water tables are kept low or never form, which allow roots to grow deep in the soil profile. The calibration procedure would give small values for the $K_{SAT}$ parameter if a grid cell was poorly drained. This causes water to move more slowly through the bottom soil layer, water tables are kept high, and roots grow to more shallow depths within the soil profile. The effective tile drainage rate controls the speed at which water is lost through tile lines when the water table is above the tile. For each of the 100 grid cells at the study site, the technique of Paz et al. (1999) was used to solve for the optimum set of the two model parameters that minimized the root mean square error (RSME) between simulated and measured yield for the five available seasons of measured yield data. Parameters were calibrated uniquely for each grid cell to account for spatial variability within the field. The objective function to be minimized during model calibration with the simulated annealing algorithm can be written as

$$\text{RMSE}_i = \left( \frac{1}{n} \sum_{j=1}^{n} (Y_{m_{i,j}} - Y_{S_{i,j}})^2 \right)^{0.5},$$

where $Y_{m_{i,j}}$ is the measured yield and $Y_{S_{i,j}}$ is the simulated yield in the $i$th grid cell for the $j$th of $n$ seasons of yield data. The model calibration procedure in Apollo provided the final minimized RMSE between measured and simulated yield for each grid cell, which represents the error associated with optimizing the two soil parameters within the grid cells over the five
calibration growing seasons. These RMSE values also serve as a performance indicator for the calibration procedure, where a RMSE of less than 1000 kg ha\(^{-1}\) is roughly less than 10% error. After completing a satisfactory calibration with acceptable RSME values, the calibrated model parameters for each grid cell were used in an N prescription analysis within the Apollo software.

Model validation is important for providing evidence that a calibrated model is performing sensibly for calibration-independent datasets. Such model testing procedures were especially important for this work because model calibration was carried out using data from only five growing seasons, and the calibrated model was then used to simulate crop yield and unused N for an extended set of historical weather over 37 growing seasons. Thus, the later simulations will only perform as well as the calibration has successfully captured the key drivers of the observed spatial variability. The details concerning model validation at this study site have been explored and presented in previous work (Thorp et al., 2005b).

4.3.3 Nitrogen Prescription Analysis

Prescription analyses in Apollo use three nested loops to simulate crop yield and N pooling for a set of N rates, management zones, and historical weather years. First, Apollo loops through a series of user-defined N rates, running CERES-Maize each time to assess the yield response and N pools for each N rate. To facilitate simulation of fall, spring, and side-dress applications, the user can also specify the fertilizer application date. A second nested loop repeats the process for each user-defined management zone or grid cell, and the third loop repeats the entire process for all the available years of historical weather data for the field location. Thus, the Apollo prescription module calculates information useful for studying yield response and N pools, as if precision N management strategies had been used during the weather patterns of previous growing seasons. In order to develop N management strategies for the future, we simulate and analyze how N rates would have performed in the past.

The Apollo prescription module was run for the study site to simulate the crop yield response and the amount of N in four pools, including NO\(_3\)-N in the soil at harvest, NH\(_4\) in the soil at harvest, total NO\(_3\)-N leached, and total NO\(_3\)-N lost out the tile. The four N pools
were summed to generate a value for total unused N at the end of the growing season. Model simulations were run for 13 N application rates over 37 years of historical weather data near the study site (1966 to 2002). Simulated N rates ranged from 80 to 320 kg ha\(^{-1}\) at increments of 20 kg ha\(^{-1}\). While running the prescription analysis, Apollo generated a text file containing the simulation results for all combinations of N rates, grid cells, and weather years. Thus for this work, the prescription output file contained 48,100 entries (13 rates * 37 years * 100 grid cells).

4.3.4 Cumulative Probability Distributions

Prescription analyses in Apollo have the potential to generate a very large amount of simulated data, depending on the number of grid cells, N rates, and weather years used in the simulation. To condense this dataset for interpreting the effect of historical weather patterns on N mobility in the agricultural system, a methodology involving cumulative probability distributions, which give the probability that a variable takes a value lesser than or equal to a specified quantity, was developed. Two families of cumulative probability curves were calculated for each grid cell: one for yield and the other for unused N left in the soil at harvest. The first family provides the cumulative probability of yield for each N rate over the number of weather years, 37 in this case. Each curve in this family represents the probability of obtaining crop yield by applying the associated rate of N fertilizer consistently over a 37-year period. The second family gives the cumulative probability of unused N for each N rate over the number of weather years. Each curve in this family represents the probably, or risk, of leaving unused N in the soil when applying the associated N rate consistently over a 37-year period. Using these two families of cumulative probability distributions together, the economic and environmental costs of applying various N rates can be compared, and N rates can be selected to accomplish objectives associated with both crop production and environmental protection. Because the two families of curves are unique for each grid cell, the complete set of curves for all grid cells can be used to develop variable-rate N management plans that achieve the production or environmental objectives for the entire field. It is convenient to fit probability distributions to the data generated from historical weather simulations, because the dataset will ultimately be used to develop N fertilizer
recommendations for future growing seasons in which weather conditions are unknown. By fitting probability distributions to the simulated data for yield and unused N for past weather years, the effect of unknown weather patterns on future yield and unused N can be characterized in terms of chance or probability. Thus, N recommendations can be designed with a level of certainty that a given yield threshold will be achieved or that an unused N threshold will not be exceeded.

To proceed with this analysis, a Visual Basic application was written to manipulate the data within the Apollo prescription file and to fit appropriate probability distributions. Automation of this process within Visual Basic was important because of the large number of datasets to which distributions were fitted. Distributions were fit for both yield and unused N for each N rate in each grid cell (2 variables * 13 rates * 100 grid cells = 2600 distributional fits). Initial work focused on fitting normal probability distributions to both the yield and unused N datasets. However, further investigations using histograms and the Shapiro-Wilk test (Shapiro and Wilk, 1965; Royston, 1995) revealed that the datasets were oftentimes severely non-normal. In an effort to remedy this problem, several alternative distributions were explored. Based on the results of chi-squared goodness-of-fit tests and visualization of distributions fitted to histograms, the beta probability distribution was chosen for use with the simulated yield data and the exponential probability distribution was selected for use with the simulated data for unused N.

The flexibility of the beta distribution proved to be helpful for fitting the heavily skewed simulated yield datasets in this research. The general formula for the probability density function of the beta distribution is

\[ f(x) = \frac{(x-a)^{p-1}(b-x)^{q-1}}{B(p,q)(b-a)^{p+q-1}} , \quad a \leq x \leq b , \quad p > 0 , \quad q > 0 , \quad (4.3) \]

where \( p \) and \( q \) are shape parameters, \( a \) and \( b \) are the respective upper and lower bounds of the distribution, and \( B(p,q) \) is the beta function (Johnson et al., 1994). The beta function is

\[ B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1} dt . \quad (4.4) \]

To fit a beta distribution, all four parameters, \( p \), \( q \), \( a \), and \( b \), must be estimated. In this work, the lower bound, \( a \), was assumed to be 0 for all yield datasets, since yield cannot be negative.
The three remaining parameters were estimated using a combination of maximum likelihood and method of moments estimation in an iterative procedure, all implemented within the Visual Basic algorithm. Initially, the upper limit, $b$, was set to the largest value in each yield dataset. Next, the method of moments estimators (Johnson et al., 1994) for both $p$ and $q$ were calculated. All four parameters were then used to compute the value of the likelihood function for the beta distribution (Gnanadesikan et al., 1967). This process was iterated several hundred times while incrementing the value of $b$ by one in successive iterations. The values of $p$, $q$, $a$, and $b$ that gave the maximum value for the likelihood function were assumed to be the parameters that provided the best fit of Equation 4.3 to a given yield dataset. A unique set of distributional parameters was calculated for each of the 2600 histograms of simulated yield data. These parameters were then used to calculate the cumulative beta probability distribution for each yield dataset. The cumulative beta probability distribution, also known as the incomplete beta function ratio, can be expressed as

$$I_x(p,q) = \frac{\int_0^x t^{p-1}(1-t)^{q-1} \, dt}{B(p,q)}, \quad 0 \leq x \leq 1, \ p > 0, \ q > 0,$$

(Johnson et al., 1994). This equation restricts the lower and upper bound to 0 and 1, respectively. Therefore, the parameter values, $a$ and $b$, were used to scale each yield data value down to within this required range. To solve the beta function (Equation 4.4) and the incomplete beta function ratio (Equation 4.5) for each yield dataset, numerical approximations, translated from the C programming language, were incorporated into the Visual Basic algorithm (Press et al., 1992).

Histograms of the simulated datasets for unused N in the soil at harvest suggested that an exponential distribution would provide an appropriate fit. The probability density function of the exponential distribution can be written as

$$f(x) = \frac{1}{\beta} e^{-(x-\mu)/\beta}, \quad x \geq \mu, \ \beta > 0,$$

(4.6)

where $\mu$ is a location parameter and $\beta$ is a scale parameter. The location parameter for all exponential distributions was set such that the lower asymptote was equal to the smallest
value for unused N in the dataset. As a result, parameter estimation was much simpler for
the exponential distribution compared to the beta distribution, because estimation of only one
parameter, $\beta$, was required, and the maximum likelihood estimator of $\beta$ is simply the sample
mean. The cumulative exponential distribution function can be expressed as

$$F(x) = 1 - e^{-(x-\mu)/\beta}, \quad x \geq \mu, \quad \beta > 0,$$

(Johnson et al., 1994). A short segment of code was written to calculate the cumulative
exponential distributions for each of the 2600 sets of simulated unused N datasets and added
to the Visual Basic application.

4.3.5 Nitrogen Management Decisions

The families of cumulative beta probability distributions for yield and the families of
cumulative exponential distributions for unused N at harvest were used to assess the
economic and environmental consequences associated with N management decisions. A
comparison of two N management strategies was carried out to demonstrate how the
distributions could be used. The objective of the first N management strategy was to
maximize the profitability of the management practice for the producer. Paz et al. (1999)
presented a simple equation to calculate marginal net return for N fertilizer management in
corn:

$$\text{Marginal Net Return} = Y \cdot P_c - N \cdot P_n,$$

(4.8)

where $Y$ is the corn yield (kg ha$^{-1}$), $P_c$ is the price of corn ($\$ \text{ kg}^{-1}$), $N$ is the N application rate
(kg ha$^{-1}$), and $P_n$ is the price of N fertilizer ($\$ \text{ kg}^{-1}$). For this study, $P_c$ was set to $0.086 \text{ kg}^{-1}$
and $P_n$ was set to $0.46 \text{ kg}^{-1}$, which are current market values for corn and N fertilizer in
Iowa. Managing N to optimize long-term marginal net return assures that producers can
achieve the maximum possible profit from their corn crop. However, this practice has been
shown to have significant environmental implications (Burkart and James, 1999; Goolsby et
al., 2001), because unused NO$_3$-N is highly susceptible to loss from the agricultural system.
Therefore, the objective of the second N management strategy was to reduce the level of
applied N to achieve an environmental objective: leave less than 40 kg ha$^{-1}$ of N in soil at
harvest with a probability of 80%. Because the two families of cumulative probability curves
link yield to unused N in the soil at harvest, they allow for the quantification of any
production losses that a producer might incur for managing N at levels below the production optimal. In this way, the producer’s opportunity cost for applying reduced N rates can be calculated by extending Equation 4.8:

\[
\text{Opportunity Cost} = (Y_{\text{max}} - Y_{\text{red}}) \cdot P_c - (N_{\text{max}} - N_{\text{red}}) \cdot P_N
\]

(4.9)

where \( Y_{\text{max}} \) and \( N_{\text{max}} \) are the yield achieved and N rate used when maximizing net return and \( Y_{\text{red}} \) and \( N_{\text{red}} \) are the reduced yield and reduced N rate for managing N to achieve the environmental objective. For the purpose of introducing this methodology, the analysis was described in detail using grid cell #4 as an example. By then repeating the analysis for all 100 grid cells, two variable-rate N prescription maps were generated for the study area: one for maximizing the producer’s net return and the other for accomplishing the environmental objective.

4.4 Results

4.4.1 Calibration Results

Cell-level differences between measured and simulated yield as indicated by RMSE ranged from 50 kg ha\(^{-1}\) to 1075 kg ha\(^{-1}\) with an average RMSE of 490 kg ha\(^{-1}\) across all grid cells (Figure 4.2). The average RMSE for grid cells dominated by the Canisteo, Clarion, Nicollet, Harps, and Okoboji soil types were 406 kg ha\(^{-1}\), 649 kg ha\(^{-1}\), 404 kg ha\(^{-1}\), 472 kg ha\(^{-1}\), and 318 kg ha\(^{-1}\), respectively. Errors for grid cells dominated by Clarion loam were greater on average than that of the other soil types, indicating that the model had more difficulty explaining yield variability in grid cells having this soil type. Clarion loam is a gently sloping, well-drained soil on convex upland knolls (USDA-SCS, 1981), meaning the soil type is generally present on the sideslopes of the swell and swale topography typical of the central Iowa countryside. Therefore, a possible explanation for the greater average RMSE in Clarion grid cells is that the model does not adequately account for surface run-on and sub-surface water flow between neighboring grid cells, the dynamics of which would be more significant for a sloped topography. Another interesting note is that the aggregation of high RMSE grid cells across the center of the field corresponds to the location of the north sideslope of a well-defined gully that cuts through the study area. These results also suggest
that model calibration performance is spatially linked to topography. Errors for grid cells dominated by the Okoboji silty clay loam were significantly lower than that of the other soil types. Okoboji silty clay loam is a very poorly drained soil occurring in concave depressions on uplands (USDA-SCS, 1981). Low RMSE values for the Okoboji dominated grid cells indicate that the model performed well in mimicking the water flow dynamics in these poorly drained areas.

![Image of RMSE values for each grid cell](image-url)

Figure 4.2. Root mean square error for the two-parameter calibration in each grid cell
The map of cell-level calibration errors (Figure 4.2) is useful for assessing the spatial distribution of RMSE across the study area; however, the RMSE associated with an individual growing season in the calibration dataset is lost during the averaging process of the RMSE calculation (Equation 4.2). In order to visualize the error associated with individual growing seasons, a one-to-one plot of simulated versus measured yield was constructed (Figure 4.3). This plot illustrates the relationship between measured and simulated yield for each of the five calibration years and for each of the 100 grid cells, where the vertical distance between an individual data point and the one-to-one line represents the difference between measured and simulated yield. Cell-level yield for the 1994, 1998, and 2002 growing seasons tended to cluster at 10,600 kg ha\(^{-1}\); however, deviations from the one-to-one line are more apparent for the 1998 and 1994 seasons than for the 2002 growing season. Cell-level yield during the 1996 and 2000 growing seasons tended to cluster around
9,250 kg ha\(^{-1}\) and 7,500 kg ha\(^{-1}\), respectively. These lower yielding growing seasons improved the ability of the model to explain year-to-year variation by increasing the yield range in the calibration dataset. Overall, the model was able to explain much of the yield variability \((r^2 = 0.89)\) when considering all 100 grid cells and all five growing seasons used for calibration (Table 4.1). The strength of linear regression was reduced when considering the fit for any individual year, because the cell-level yields for single seasons tended to cluster together in relatively narrow ranges (Figure 4.3). However, RMSE calculations for any given production year showed that deviations between field-level measured and simulated yields were not greater than 600 kg ha\(^{-1}\) (Table 4.1). Since past research has supported RMSE and related statistics over correlation-regression analysis for testing crop model accuracy (Kobayashi and Salam, 2000), it is expected that the lower \(r^2\) values for individual seasons can be ignored in favor of the more descriptive RMSE calculations.

Table 4.1. Field-level measured versus predicted yield relationships

<table>
<thead>
<tr>
<th>Production Year</th>
<th>Measured yield (kg ha(^{-1}))</th>
<th>Predicted yield (kg ha(^{-1}))</th>
<th>RMSE (kg ha(^{-1}))</th>
<th>(r^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>10788</td>
<td>11224</td>
<td>520</td>
<td>0.30</td>
</tr>
<tr>
<td>1996</td>
<td>9137</td>
<td>9039</td>
<td>217</td>
<td>0.63</td>
</tr>
<tr>
<td>1998</td>
<td>10218</td>
<td>10773</td>
<td>598</td>
<td>0.47</td>
</tr>
<tr>
<td>2000</td>
<td>7484</td>
<td>7617</td>
<td>286</td>
<td>0.26</td>
</tr>
<tr>
<td>2002</td>
<td>10625</td>
<td>10786</td>
<td>323</td>
<td>0.31</td>
</tr>
<tr>
<td>All years</td>
<td>9650</td>
<td>9888</td>
<td>490</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Values for the two optimized parameters, resulting from applying the calibration procedure within each grid cell, were within the expected ranges. The average value for effective tile drainage rate was 0.107 day\(^{-1}\) with a standard deviation of 0.059 day\(^{-1}\) across 100 grid cells. The minimum value for this parameter was 0.011 day\(^{-1}\) in grid cell #29, and the maximum value was 0.246 day\(^{-1}\) in grid cell #26. These values fell within the expected range of 0.01 day\(^{-1}\) and 0.25 day\(^{-1}\) for effective tile drainage rate. The average value for saturated hydraulic conductivity of the deep layer was 0.790 cm day\(^{-1}\) with a standard
deviation of 0.646 cm day\(^{-1}\) across 100 grid cells. The minimum value for this parameter was 0.094 cm day\(^{-1}\) in grid cell #83, and the maximum value was 1.976 cm day\(^{-1}\) in grid cell #54. These values fell within the expected range of 0.001 cm day\(^{-1}\) and 2 cm day\(^{-1}\) for saturated hydraulic conductivity.

### 4.4.2 Grid Cell #4 Example

Two families of cumulative probability distributions were produced for each grid cell. Since it is not feasible to show the resulting 200 graphs, the results of grid cell #4 were selected randomly and presented as an example. The cumulative beta probability distributions of yield for grid cell #4 demonstrate how corn yield and N rate can be related for making N management decisions. For each of the 13 N rates, model simulations provided 37 values for yield in grid cell #4, representing the seasonal corn yield achieved with an N rate given the weather conditions of the past 37 growing seasons. An example histogram for the simulated yield values in grid cell #4 at an N rate of 220 kg ha\(^{-1}\) demonstrates how the yield datasets can be well characterized using a beta distribution (Figure 4.4). Distributions of simulated yield for other N rates and grid cells typically resembled this histogram in which the lack of an upper tail skewed the distribution heavily to the left. One interpretation of this occurrence is that the weather conditions for most seasons allowed corn yields to approach the potential upper limit for yield in this field. In other seasons when weather conditions were less favorable, simulated yields were significantly lower than the yield potential which created the tapering effect on the left side of the distribution. For this histogram, the four beta parameters, \(p\), \(q\), \(a\), and \(b\), estimated for the fit were 3.60, 0.98, 0, and 11633, respectively. By fitting unique beta distributions to each of the 13 yield histograms for N rates in grid cell #4 and calculating the cumulative beta probability distribution using the resulting beta parameters, a family of curves that explain the chance of corn yield response to N rates given the weather conditions of the past 37 years was generated (Figure 4.5). At the time N management decisions are made, these curves are used with the understanding that corn yield response in grid cell #4 will be heavily influenced by the unknown pattern of weather encountered between the N application and harvest. However, based on the patterns of weather seen in previous years, it is possible to
describe future corn yield in terms of chance or probability, such that N management decisions can be made less blindly. Cumulative beta probability distributions of yield in grid cell #4 generally showed an increasing trend with N rate at equal probability levels. An exception occurred in the range of the 0% and 20% probability thresholds of yield between 4,000 kg ha\(^{-1}\) and 8,000 kg ha\(^{-1}\). A likely explanation is that small levels of N stress aided root development early in the season, which better prepared the crop to combat water stress and increase yield in these relatively low yielding years. This area of the curve is also perhaps less important, because producers will likely be more interested in yield probabilities much greater than 10%. The beta distribution also has difficulty fitting the high yield end of the histogram in Figure 4.4, and the resulting probability curves in Figure 4.5 are awkwardly shaped with no roll-off effect in the upper portion of the distribution. These results provide

---

Figure 4.4. Histogram of the grid cell #4 simulated yield values over 37 years of historical weather at an N rate of 220 kg ha\(^{-1}\) and a fitted beta distribution
evidence that further analysis may be more accurate if focused away from the tails of the distribution. Finally, the yield distributions for grid cell #4 did not change significantly for N rates above 280 kg ha\(^{-1}\), indicating that yield was not affected by increasing N rates above this level.

Figure 4.5. Cumulative beta probability of yield in grid cell #4 for N rates of 80 kg ha\(^{-1}\) to 320 kg ha\(^{-1}\).

Similar to the cumulative beta probability distributions of yield, the cumulative exponential probability distributions of unused N left in the soil at harvest for grid cell #4 demonstrate how unused N and N rate can be related for making N management decisions. For each of 13 N rates, model simulations provided 37 values for unused N in grid cell #4, representing the seasonal post-harvest soil N content obtained with an N rate given the weather conditions of the past 37 growing seasons. An example histogram for the simulated unused N values in grid cell #4 at an N rate of 220 kg ha\(^{-1}\) suggests that the unused N datasets can be well characterized using an exponential distribution (Figure 4.6). Distributions of unused N for other N rates and other grid cells typically resembled this histogram with a
majority of growing seasons having relatively small amounts of N left in the soil, particularly for the lower N rates. In other seasons when weather conditions were not favorable to uptake

![Histogram of the grid cell #4 simulated unused N values over 37 years of historical weather at an N rate of 220 kg ha\(^{-1}\) and a fitted exponential distribution](image)

Figure 4.6. Histogram of the grid cell #4 simulated unused N values over 37 years of historical weather at an N rate of 220 kg ha\(^{-1}\) and a fitted exponential distribution

of N by plants, greater amounts of N were left in the soil and this created the tapering effect on the right side of the distribution. As N rates were increased, the right side of the distribution tended to taper off more slowly. For the histogram in Figure 4.6, a value of 13.73 was estimated for the exponential parameter, \(\beta\). The distribution was adjusted right by a factor of 16.58, the lowest unused N value in the 220 kg ha\(^{-1}\) dataset for grid cell #4. Similar to the curves for yield, the cumulative exponential probability distributions of unused N explain the *chance* that different N rates will leave N in the soil at harvest given the weather conditions of the past 37 growing seasons (Figure 4.7). This family of curves permits the addition of an environmental component to N management decisions, such that N applications can be designed to achieve a balance between production and environmental goals. Cumulative exponential probability distributions of unused N for grid cell #4 showed
an increasing trend with N rate at equal probability levels. An interesting feature of the unused N probability curves is that, even for the lowest N rate, the simulations always resulted a small amount of N, approximately 15 kg ha\(^{-1}\), left in the soil at harvest. This phenomenon in the simulation results may be attributed to mineralization of N at the end of the growing season after the crop was no longer taking up nutrients, or it may an artifact of the model.

![Figure 4.7. Cumulative exponential probability of unused N in grid cell #4 for N rates of 80 kg ha\(^{-1}\) to 320 kg ha\(^{-1}\)](image)

4.4.3 **Statistics Over 100 Grid Cells Per N Rate**

In the process of fitting a distribution and developing cumulative probability curves for yield and unused N, one data dimension, the number of weather years, is essentially removed from prescription simulation datasets. The data can then be described in terms of beta distribution parameters, exponential distribution parameters, and cumulative probability distributions over the two remaining data dimensions, number of N rates and number of grid cells. However, it is not feasible to show the histograms, distributional fits, and cumulative
probability curves for every simulated N rate in every grid cell within the study area. Instead, the mean and standard deviation was used to summarize the distribution of estimated beta and exponential parameters at each N level over all 100 grid cells (Table 4.2). As expected, the prescription simulations showed that an increase in N rate increased yield in grid cells on average, but it also increased the amount of unused N in the soil after harvest.

Table 4.2. Summary statistics (average and standard deviation) for the 37-year average yield, beta parameters, the 37-year average unused N, and exponential parameters for each N rate over 100 grid cells.

<table>
<thead>
<tr>
<th>N rate (kg ha(^{-1}))</th>
<th>Yield (kg ha(^{-1}))</th>
<th>Beta parameters</th>
<th>Unused N (kg ha(^{-1}))</th>
<th>Exponential param.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>p</td>
<td>q</td>
<td>(\beta)</td>
</tr>
<tr>
<td>Avg</td>
<td>SD</td>
<td>Avg</td>
<td>SD</td>
<td>Avg</td>
</tr>
<tr>
<td>80</td>
<td>5386</td>
<td>228</td>
<td>6985</td>
<td>133</td>
</tr>
<tr>
<td>100</td>
<td>6830</td>
<td>340</td>
<td>8068</td>
<td>163</td>
</tr>
<tr>
<td>120</td>
<td>7604</td>
<td>397</td>
<td>8831</td>
<td>162</td>
</tr>
<tr>
<td>140</td>
<td>8064</td>
<td>434</td>
<td>9460</td>
<td>188</td>
</tr>
<tr>
<td>160</td>
<td>8337</td>
<td>446</td>
<td>9926</td>
<td>114</td>
</tr>
<tr>
<td>180</td>
<td>8522</td>
<td>463</td>
<td>10553</td>
<td>220</td>
</tr>
<tr>
<td>200</td>
<td>8669</td>
<td>469</td>
<td>10959</td>
<td>171</td>
</tr>
<tr>
<td>220</td>
<td>8843</td>
<td>501</td>
<td>11680</td>
<td>237</td>
</tr>
<tr>
<td>240</td>
<td>8982</td>
<td>531</td>
<td>12407</td>
<td>349</td>
</tr>
<tr>
<td>260</td>
<td>9059</td>
<td>532</td>
<td>13151</td>
<td>420</td>
</tr>
<tr>
<td>280</td>
<td>9105</td>
<td>532</td>
<td>13727</td>
<td>555</td>
</tr>
<tr>
<td>300</td>
<td>9131</td>
<td>516</td>
<td>13891</td>
<td>590</td>
</tr>
<tr>
<td>320</td>
<td>9152</td>
<td>498</td>
<td>13904</td>
<td>581</td>
</tr>
</tbody>
</table>

Because yield increased with N rate, the upper limit beta parameter, \(b\), for the yield distributions also increased on average with N rate. Standard deviations for yield, unused N, and \(b\) all showed a general increasing trend with N rate. The beta shape parameter, \(p\), showed a decreasing trend with N rate on average across all grid cells. When the N rate increased above 180 kg ha\(^{-1}\), the \(p\) parameter was on average approximately 4.0 with a standard deviation of 0.7. The \(q\) parameter of the beta distribution decreased from 3.0 at 80 kg ha\(^{-1}\) to 0.91 at 160 kg ha\(^{-1}\) and then increased to 2.1 as N rate increased to 320 kg ha\(^{-1}\). For the exponential distributions of unused N, the average \(\beta\) parameter across all grid cells and its standard deviation both increased with increasing N rate. This was expected since a larger \(\beta\)
value makes the exponential distribution decay more slowly, representing an increased frequency of larger amounts of N left unused. For all but the highest two rates, the average location factor for the exponential distribution remained between 17 and 19 with a low standard deviation, indicating most N rates and most grid cells had at least one season with less than 20 kg ha\(^{-1}\) N left unused. The spatial distribution of summary statistics similar to those in Table 4.2 could also be studied by averaging over N rate instead of over grid cells; however, this makes less sense because only one N rate would be applied in a given season. An investigation into the spatial effect of grid cell location will be the subject of the next section, which illustrates the use of the cumulative probability curves to select N rates and develop N prescriptions that satisfy production and/or environmental requirements.

4.5 Discussion

4.5.1 Environmental Cost of Optimizing Economic Return

For a farming operation to be profitable, producers must use management practices that maximize marginal net return (Equation 4.8). Continuing with the analysis of grid cell #4, calculations were carried out to determine the relationship between N rates and marginal net return. Average marginal net return represents the average return for each N rate over all 37 growing seasons. Maximum and minimum marginal net return represents the greatest and least return achieved with each N rate in a single year. For grid cell #4, an N rate of 240 kg ha\(^{-1}\) maximized the average marginal net return over 37 growing seasons (Figure 4.8). If the producer applied this rate to grid cell #4 in each year, there would be a 50:50 chance that the marginal net return would be greater than or less than $689.40 ha\(^{-1}\). However, the uncertainty in marginal net return values for individual seasons is large for the 240 kg ha\(^{-1}\) N rate, because the range between minimum and maximum net return values widen significantly at the higher N rates. In one of the 37 years, the 240 kg ha\(^{-1}\) N rate may result in a net return of $953.76 ha\(^{-1}\) while in another it may result in a return of only $188.36 ha\(^{-1}\) (Figure 4.8). The cumulative beta probability distributions of yield now become useful for determining the expected yield from this management practice. By applying the 240 kg ha\(^{-1}\) N rate to grid cell #4, the producer could expect a 50:50 chance of yield greater than 9,681 kg
ha\(^{-1}\) in any given growing season. Similarly, the grower could expect an 80% chance of yield less than 11,333 kg ha\(^{-1}\) or a 20% chance of yield greater than 11,333 kg ha\(^{-1}\) (Figure 4.5).

![Graph showing the relationship between N rate and net return.](image)

Figure 4.8. The 240 kg ha\(^{-1}\) N rate maximizes average marginal net return in grid cell #4 over 37 growing seasons.

Given that the producer must apply 240 kg ha\(^{-1}\) of N in grid cell #4 to maximize average marginal net return over 37 growing seasons, the cumulative exponential probability distributions of unused N (Figure 4.7) are now useful for determining the environmental risk associated with this management practice. The curve for an N rate of 240 kg ha\(^{-1}\) in grid cell #4 shows that there is a 50:50 chance that the amount of unused N left in the soil will be greater than 30.6 kg ha\(^{-1}\). Similarly, there is an 80% probability that the amount of unused N left in the soil will be less than 49.4 kg ha\(^{-1}\) or a 20% probability that unused N will be greater than 49.4 kg ha\(^{-1}\). This represents the environmental risk associated with applying the 240 kg ha\(^{-1}\) N rate in grid cell #4, because N left unused in the soil will be highly susceptible to loss in the months between growing seasons. In terms of the quantity of N left unused in soil when optimizing producer economics, the grid cell #4 value of 49.4 kg ha\(^{-1}\) at the 80% probability level is relatively moderate, only slightly lower than the field average of 56.2 kg.
ha$^{-1}$. On the other hand, in grid cell #74, the N management practice that optimized marginal net return (260 kg ha$^{-1}$ N) had an 80% chance of leaving 120.9 kg ha$^{-1}$ or less of unused N in the soil at harvest. In other words, if N is managed to optimize economic return in this grid cell, nearly half of the applied N will remain in the soil after harvest in 20% of growing seasons.

By repeating this analysis for all 100 grid cells, an N prescription map was developed for optimizing marginal net return across the entire study area (Figure 4.9). Since the model

Figure 4.9. Nitrogen prescription for optimizing marginal net return over 37 growing seasons
runs for the prescription analysis were performed with the assumption of nearly zero initial N in the soil, measurements of the actual available N in the soil prior to a fertilizer application should be used as a credit for the N rates given in this prescription. The average N rate to optimize production across all 100 grid cells was 233 kg ha\(^{-1}\) with a standard deviation of 21 kg ha\(^{-1}\) (Table 4.3). Based on the weather of the past 37 growing seasons, this variable-rate N prescription would have an 80% chance of producing a field average crop yield of 11,009 kg ha\(^{-1}\) or less with a standard deviation of 589 kg ha\(^{-1}\) across grid cells. Also, the management practice would have an 80% chance of leaving a field level average of 56.2 kg ha\(^{-1}\) or less unused N in the soil at harvest with a standard deviation of 16.7 kg ha\(^{-1}\) across grid cells. Finally, this N prescription would have an 80% chance of achieving a field average marginal net return of less than $839.44 ha\(^{-1}\) with a standard deviation of $43.10 ha\(^{-1}\) over 100 grid cells.

Table 4.3. Field-level statistics over all 100 grid cells for the economically optimum prescription at the 80% probability threshold.

<table>
<thead>
<tr>
<th></th>
<th>N Rate (kg ha(^{-1}))</th>
<th>Yield (kg ha(^{-1}))</th>
<th>Unused N (kg ha(^{-1}))</th>
<th>Marginal Net Return ($ ha(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>233</td>
<td>11009</td>
<td>56.2</td>
<td>839.44</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>21</td>
<td>589</td>
<td>16.7</td>
<td>43.10</td>
</tr>
<tr>
<td>Minimum</td>
<td>160</td>
<td>8910</td>
<td>35.0</td>
<td>692.66</td>
</tr>
<tr>
<td>Maximum</td>
<td>260</td>
<td>11816</td>
<td>120.9</td>
<td>896.58</td>
</tr>
</tbody>
</table>

4.5.2 **Opportunity Cost of Environmental Protection**

Managing N to optimize marginal net return assures that producers can achieve the maximum possible profit from their corn crop. However, this practice can have significant environmental impacts in some areas of an agricultural field, because there is a greater chance that a large quantity of N will remain in the soil after harvest. To reduce the environmental impacts of corn production, producers must begin managing N to achieve a balance between environmental and production objectives. For example, assume that
lawmakers establish an environmental policy stating that unused N left in the soil after harvest must be less than 40 kg ha\(^{-1}\) 80% of the time. Cumulative exponential probability of unused N (Figure 4.7) can now be used to determine the rate of N that will meet this objective for grid cell #4 of the study area. Furthermore, we can use cumulative beta probability of yield (Figure 4.5) to determine the yield and the producer’s opportunity cost (Equation 4.9), which is the profit that the producer foregoes by managing N to meet the environmental restriction. With linear interpolation between the cumulative exponential probability curves for the 220 and 240 kg ha\(^{-1}\) N rates, an N rate of 222 kg ha\(^{-1}\) insures that the amount of unused N will be less than 40 kg ha\(^{-1}\) with a probability of 80% (Figure 4.7). Then using the cumulative beta probability of yield curves, an N rate of 222 kg ha\(^{-1}\) will give a crop yield of 11,006 kg ha\(^{-1}\) 80% of the time (Figure 4.5). The producer’s opportunity cost for reducing N rates can now be calculated using Equation 4.9. For grid cell #4, the producer’s opportunity cost for leaving less than 40 kg ha\(^{-1}\) of unused N in the soil at harvest with a probability of 80% is $20.01 ha\(^{-1}\).

By repeating this analysis for all 100 grid cells, an N prescription map was developed for meeting the environmental objective of leaving less than 40 kg ha\(^{-1}\) unused N in the soil at harvest with 80% probability across the entire study area (Figure 4.10). Again, since the model runs for the prescription analysis were performed with the assumption of nearly zero initial N in the soil, measurements of the actual available N in the soil prior to a fertilizer application should be used to credit the N rates given in this prescription. The average N rate to accomplish the environmental objective across all 100 grid cells was 194 kg ha\(^{-1}\) with a standard deviation of 41 kg ha\(^{-1}\) (Table 4.4). Based on the weather of the past 37 growing seasons, this variable-rate N prescription would have an 80% chance of producing a field average crop yield of 10,240 kg ha\(^{-1}\) or less with a standard deviation of 1146 kg ha\(^{-1}\) across grid cells. Also, this management practice would have an 80% chance of achieving a field average marginal net return of less than $791.32 ha\(^{-1}\) with a standard deviation of $79.81 ha\(^{-1}\) over 100 grid cells. The field level average opportunity cost incurred by the producer using this precision N management strategy would be $48.12 ha\(^{-1}\) with standard deviation of $53.13 ha\(^{-1}\) across the 100 grid cells. At this cost, the average amount of N left unused in the soil at harvest would be reduced by greater than 16.2 kg ha\(^{-1}\) in 20% of growing seasons.
Interestingly, in seven of the 100 grid cells the producer’s opportunity cost was negative, indicating that the prescription for optimizing net return already leaves less than 40 kg ha\(^{-1}\) of unused N in soil with 80% probability. For these grid cells, the N rate that forces the environmental objective is actually greater than the N rate that optimizes economic return. Obviously, the producer should use the economically optimal N rate for these grid cells. A map of the spatial distribution of opportunity cost demonstrates how an environmentally

![Map of the spatial distribution of opportunity cost demonstrating how an environmentally friendly N rate can be determined.](image)

Figure 4.10. Nitrogen prescription for not exceeding 40 kg ha\(^{-1}\) of unused N in the soil at harvest in 80% of growing seasons
conscience management scenario for N fertilizer is quite costly to the producer in some areas of the field, but in other grid cells it costs the producer nothing (Figure 4.11). As expected, the grid cells that require the lowest rates to meet the environmental objective (Figure 4.10) also cost the producer significantly to manage in this way (Figure 4.11).

Table 4.4. Field-level statistics over all 100 grid cells for the environmental objective of no more than 40 kg ha\(^{-1}\) of unused N in the soil at harvest in 80% of growing seasons

<table>
<thead>
<tr>
<th>N Rate (kg ha(^{-1}))</th>
<th>Yield (kg ha(^{-1}))</th>
<th>Unused N (kg ha(^{-1}))</th>
<th>Marginal Net Return ($ ha(^{-1}))</th>
<th>Opportunity Cost ($ ha(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>194</td>
<td>10240</td>
<td>40</td>
<td>791.32</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>41</td>
<td>1146</td>
<td>0</td>
<td>79.81</td>
</tr>
<tr>
<td>Minimum</td>
<td>90</td>
<td>6778</td>
<td>40</td>
<td>541.13</td>
</tr>
<tr>
<td>Maximum</td>
<td>238</td>
<td>11383</td>
<td>40</td>
<td>869.70</td>
</tr>
</tbody>
</table>

The environmental policy of leaving no more than 40 kg ha\(^{-1}\) of unused N in the soil after harvest with 80% probability was picked at random for demonstration purposes. However, if such a strategy were actually implemented to legislate requirements for agricultural N management, the future economic prosperity of the agricultural industry would depend heavily on a sound methodology for appropriately selecting the environmental restriction to be met. Inappropriate or unrealistic expectations could have severe effects on the economics of a farming operation. To aid in this policy making process, the probability curves can be manipulated to obtain the producer's opportunity cost for many possible environmental policy scenarios (Figure 4.12). To develop the plot, the producer's opportunity cost was determined at several different probability levels while incrementing the restriction for unused N by one over the range of 15 kg ha\(^{-1}\) to 65 kg ha\(^{-1}\). This type of plot would give a policymaker a tool for judging the degree of impact that a particular policy, in terms of the restriction for unused N left in the soil, would have on a producer. For the previous example, an environmental restriction of no more than 40 kg ha\(^{-1}\) of unused N in the soil with 80% probability was implemented. Figure 4.12 demonstrates that this scenario is relatively lenient for grid cell #4, since it costs the producer only $20.01 ha\(^{-1}\) in that grid cell.
However, if the environmental policy had been established to be no more than 20 kg ha\(^{-1}\) of unused N in the soil with 80% probability, the opportunity cost jumps to $360.95 ha\(^{-1}\) in grid cell #4. Such a policy would not be economically feasible for the corn grower. Effective policies for restricting the amount of unused N in the soil after harvest must strive to achieve a proper balance between environmental risk and producer economics.

![Figure 4.11. Producer's opportunity cost for leaving less than 40 kg ha\(^{-1}\) of unused N in the soil 80% of the time](image)
Figure 4.12. Producer’s opportunity cost in grid cell #4 for unused N restriction levels ranging from 15 kg ha\(^{-1}\) to 65 kg ha\(^{-1}\) at six different levels of probability.

4.6 Conclusions

Precision management of N fertilizer has not become common practice in the midwestern United States, because the economic cost incurred by applying reduced N rates has not been adequately demonstrated. Such information has been difficult to generate because the dynamics of N movement in an agricultural system is highly complex and because it varies depending on spatial location and weather patterns. Crop growth models can serve as a useful tool to make sense of this complex dynamic system. Given the soil properties, management practices, and historical weather information for specific study areas, model simulations are able to demonstrate how various N management scenarios would have affected yield and unused N in the soil at harvest under the weather conditions of past growing seasons. By fitting cumulative probability distributions to the yield and unused N data, simulation results from past growing seasons can be used to look forward in time, and the uncertainty associated with the effect of unknown future weather on future yield and
future unused N left in the soil can be discussed in terms of probability. On the basis of chance, these probability distributions effectively unite yield and unused N left behind, the two most important variables for addressing the production and environmental concerns of N management in agricultural cropping systems.

4.7 References


USDA-SCS, 1981. Soil Survey of Boone County, Iowa. United States Department of Agriculture – Soil Conservation Service (USDA-SCS), Iowa Agriculture and Home Economics Experiment Station Cooperative Extension Service, Iowa State University, and State of Iowa Department of Soil Conservation, Ames, IA, USA.
CHAPTER 5. GENERAL CONCLUSIONS

5.1 Conclusions

This dissertation describes three projects that advance remote sensing, crop growth modeling, and decision support technology within the framework of precision agriculture, as modeled in Figure 1.1. Although development towards integration of these technologies was not completed fully, a significant contribution was made within each of the three areas in effort to move towards the end goal of a completely integrated remote sensing and crop growth modeling tool for making nitrogen management decisions in midwestern cornfields. Remote sensing technology was successfully used to estimate corn plant stand density across Iowa cornfields, although some limitations applied. Also, the CERES-Maize crop growth model was successfully validated for simulating spatial corn yield variability across Iowa cornfields. In the future, it is expected that corn plant population information obtained from remote sensing images could be used to further improve CERES-Maize simulations of spatial corn yield variability. The advancement in the area of decision support involved the development of the methodology for linking production and environmental risks of precision nitrogen management strategies based on the output of crop model simulations over an extended weather sequence. Together, these developments all contribute towards the completion of a fully integrated tool for developing site-specific nitrogen management plans that consider both the economic and environmental concerns of nitrogen management in midwestern cornfields.

Remote sensing technology was effective for estimation of corn plant stand density; however, the level of success depended on the spatial resolution of the analysis and the date of image collection. The best results were obtained when remote sensing imagery was collected while corn plants were at the later vegetative growth stage. At this time, soil background effects were minimized, yet the plants had not begun to senesce or to transfer nutrients from the leaves to the grain. For the analysis at the 2 m spatial resolution, results were relatively poor because of quantization effects due to the row width geometry. At lower spatial resolutions of 6 m and 10 m, the effect of row width quantization was reduced, because a greater number of rows were contained within the grid cells. Results were also
shown to be dependent on the total overall variability of corn plant stand density within the field. With greater spatial variability, corn plant population was estimated from reflectance with better results. The most useful spectral information in this study included reflectance in the blue region (473 to 492 nm), longer green and shorter red wavelengths (584 to 635 nm), and the red-edge (729 nm) and near-infrared region. When using reflectance information in these wavebands for multiple linear regression analysis, the best overall relationships between reflectance information and corn plant stand density were obtained when using imagery from the late vegetative corn growth stages at a spatial resolution of 6 m or 10 m in plots where the spatial variability in corn plant stand density was artificially increased.

A cross validation procedure and bivariate confidence ellipses were used to evaluate CERES-Maize simulations of spatial corn yield variability across an Iowa cornfield. Results indicated that, for the rain fed conditions of Iowa, the model performed most poorly when using the wettest or driest growing seasons to validate the model. This occurred because the model parameters fitted under moderate weather conditions were less flexible for simulating yield in growing seasons with more extreme weather conditions. The implication for this is that the best simulation results will be obtained for independent growing seasons when the weather in that season more closely matches the weather that occurred in the growing seasons used for model calibration. When the weather in the independent growing season is drastically different, there is less confidence that the model can accurately simulate the independent season. Because model simulations were performed spatially, it was possible to assess the cause of variability in model performance across the field. Results indicated that grid cells occurring at the sloped areas of the field typically had higher model performance errors. This likely happened because the model does not account for the effects of surface and sub-surface run-on between neighboring grid cells, and the dynamics of this process would be more influential in areas having a sloped topography.

The new decision support system called Apollo was effective for running the CERES-Maize model to simulate yield and nitrogen left behind for a sequence of historical weather data. Simulations were run for 13 spring-applied nitrogen rates over a cornfield divided into 100 0.2 ha grid cells. A methodology based on cumulative probability distributions was then developed to use model output for assessing the link between yield and nitrogen left behind
for various nitrogen rates in each grid cell. These cumulative probability distributions were used to evaluate the economic and environmental risks of two alternate precision nitrogen management strategies for the study area. In the first strategy, nitrogen rates were selected to maximize the producer's marginal net return in each grid cell. The environmental cost of this management strategy, in terms of nitrogen left behind, was determined to be 56.2 kg ha\(^{-1}\) on average over all grid cells. In the second strategy, nitrogen rates were selected to insure that the amount of nitrogen left in the soil at harvest would not exceed 40 kg ha\(^{-1}\) in 80% of growing seasons. The producer's opportunity cost for reducing nitrogen rates to achieve this environmental objective was calculated to be $48.12 ha\(^{-1}\) on average over all grid cells. On the basis of chance, the cumulative probability distributions developed from simulation output in this work effectively unite yield and unused nitrogen left behind, the two most important variables for addressing the production and environmental concerns of nitrogen management in agricultural cropping systems. Thus, using this methodology, precision nitrogen management decisions can be made while keeping both the economic and environmental objectives of the management practice in mind.

5.2 Recommendations

Estimation of corn plant stand density from remote sensing images may be improved with further investigations into the methods used to relate reflectance to population. In this work, the analysis was restricted to multiple linear regression on three wavebands due to limitations in computational speed. However, spectral analysis results suggested that there may be more than three wavelength ranges of importance, because the ranges of interest for three-band multiple linear regression often straddled between distinct areas of importance in vegetative reflectance curves. Multiple linear regression analyses using combinations of four to six wavebands may improve results by taking advantage of contrasts between more than three wavelength ranges of importance, although it will require significant computer time to perform this regression analysis. Investigations into the use of regression on principle components or partial least squares regression may be another pathway to improving the relationship between spectral reflectance and corn plant stand density. Also, explorations
using more efficient computational techniques, such as genetic algorithms, may be fruitful for this endeavor.

Another important aspect of the corn plant population sensing study was the contrasts made between ground-based crop sensing systems and remote sensing from airborne platforms. A ground-based corn plant population system was used to extensively map the location of every corn plant in each plot. This type of dataset is unique for remote sensing analysis, because none of the study area remained unsampled. Thus, it was possible to completely assess the ability of remote sensing to estimate corn plant stand density on a pixel-by-pixel basis. Future investigations in agricultural remote sensing will benefit from a similar approach in which a ground-based sensor is first used to extensively map the scene of interest. By acquiring a detailed map of soil or crop growth parameters on the ground, a truer assessment of the limitations of remote sensing technology can be obtained as camera systems are used on aerial and satellite platforms farther away from the scene. Then, it is possible to determine whether remote sensing offers any advantages over ground-based data collection and whether remote sensing images can be used to accurately estimate the true variability of crop parameters on the ground.

When using remote sensing technology to estimate corn plant stand density, an important factor is the level of population variability that exists in the scene of interest. Results presented in Chapter 2 clearly show that the ability of using remote sensing to detect spatial variability in corn plant population decreases as the level of total variability decreases. This may be a significant factor for using this technology in production cornfields, because producers currently tend to aim for uniform populations across their fields. More broad-scale investigations into the nature of spatial variability of corn populations would be appropriate to determine if remote sensing can be effective in the production setting. These investigations should also aim to understand what part of the spatial variability arises from agronomic factors and what part arises due to machine error. Ground-based corn population sensing systems such as the one used for this project would be useful for this type of investigation.

For the model validation project, an interesting area of research will evolve as the number of measured datasets for a particular study site continues to increase. Many growing
seasons will have moderate weather conditions, and occasionally a year with extreme weather conditions will be encountered. Because moderate growing seasons should be more common, it will be interesting to determine the level of influence that a particular growing season contributes to the overall model calibration. It is expected that the extreme growing seasons will have greater influence on the fitting of the model parameters, and it is possible that several of the moderate growing seasons will contribute very little. In this case, it may be beneficial to determine which growing seasons contributed nothing to the overall optimization and remove them from the calibration dataset for the sake of simplicity. Further model developments may be necessary to account for the limitations of using the model for grid-based or zone-based applications in precision agriculture. For this kind of work, the movement of water and nutrients laterally between neighboring grid cells is an issue. Further exploration is needed to determine the proportion of model error that can be attributed to topography and how the model could be modified for simulating the effect of topography on movement of water and nutrients.

Future work involving the methodology for linking the economic and environmental risks of precision nitrogen management strategies initially should focus on incorporating the code for calculating the probability curves into the Apollo decision support system. Currently, Apollo simply outputs the necessary simulation data and computation of the probability curves occurs in a separate program. In addition, there exists a need for testing this methodology in the field, although an appropriate test would take many years of field data collection. In such a test, a cornfield should be managed according to a precision nitrogen management recommendation generated with this methodology for several years. At the end of each growing season, the yield and nitrogen left behind should be measured in each management zone. If the model has correctly simulated the processes occurring in the field, fitting of probability distributions to the measured data should yield a result similar to that of the simulated data for the same nitrogen rate and grid cell.

As stated earlier, the work in this dissertation describes an effort to develop an integrated remote sensing and crop growth modeling tool for precision nitrogen management decision support in corn. Although advancements toward the model shown in Figure 1.1 were made, further work is necessary to fully develop this tool. As a next step toward
complete integration, an interesting project would be to assess the effect of remote sensing-based corn population inputs on the behavior of the crop growth model. A simple project would involve the collection of remote sensing imagery and other information necessary for crop model simulations over several seasons of corn production in a field. The objective would be to determine if remote sensing-based estimates of corn plant stand density improved CERES-Maize simulations of corn yield as opposed to assumed population values. Another interesting objective involves the way in which remote sensing imagery should be combined with the CERES-Maize model. The main question is to determine whether estimates of corn population from remote sensing images can be used in procedure to calibrate the model, or should the estimates simply serve as a model input. Once it is demonstrated how the two technologies should be integrated, a next step would be to develop a geographic information system (GIS) for more efficient and user-friendly union of remote sensing and crop growth modeling techniques. Ideally, the GIS should contain the corn population map derived from remote sensing images according to the predefined management zone boundaries for the field, and the GIS should also store all the other necessary parameters required for crop model simulations within the management zones. A decision support system such as Apollo could then be developed to search the GIS for appropriate spatial information required to complete model simulations. These are the first steps necessary to move toward complete integration of remote sensing and crop growth modeling for nitrogen management decision support in corn.
ACKNOWLEDGEMENTS

My time as a graduate student here in the Department of Agricultural and Biosystems Engineering at Iowa State University (ISU) has been made possible by contributions from Cargill Inc., Pioneer Hi-Bred International, Inc., the ISU Premium for Academic Excellence (PACE) award program, and the Harold and Katherine Guy Graduate Fellowship program. A majority of my tuition and my research assistantship stipend was covered by a gift from Cargill. Pioneer provided additional assistance with my stipend during the summer of 2004, in which I collected all of the data for the project on estimation of corn plant population from remote sensing imagery. The PACE award covered half of my tuition during my first year at ISU. I was also awarded the Harold and Katherine Guy Graduate Fellowship during my final year. This award provided $2,000 for personal expenses and $1,000 for research expenses and professional travel. Without these contributions to ISU, to the Agricultural Engineering Department, and to myself, my work here may not have been possible.

My learning experiences in the area of crop growth modeling were made possible with the help and guidance of Dr. Bill Batchelor, Dr. Joel Paz, and Dr. Jon Lizaso. These researchers were members of the Crop Modeling Laboratory when I started at ISU, and they were integral in helping me understand how crop growth models work and how to use them within the framework of precision agriculture. Sharing the laboratory with Dr. Paz and Dr. Lizaso was especially rewarding, because they were always available and open for helping me with whatever problem I was facing on a certain day. I also thank the Iowa corn grower, Mr. Ron Heck, who allowed our research group to collect data for use with our crop growth models in his production fields. A special thanks is also in order for Dr. Batchelor who served as my co-major professor and mentor for the crop modeling work completed in this dissertation. He was always able to provide encouragement and useful advice, and his passion for research in our field motivates his students to be successful. Also, during my first year at ISU, Dr. Batchelor took a sabbatical and spent a majority of his time creating the Apollo decision support system. Much of the work described in this dissertation would not have been possible without the availability of this software. I also have many fond memories of the discussions that Dr. Batchelor and I had while leading a group of undergraduates on a study abroad trip to Germany in August 2004. Dr. Batchelor is very knowledgeable about
the state of things in the agricultural engineering discipline and in the university system, and I found these conversations very informative and interesting. My enjoyment of these discussions was probably magnified by the fact that they typically took place in German beer gardens surrounded by beautiful German architecture and after a glass or two of Hefeweizen!

For the investigations into the use of remote sensing to estimate corn plant stand density, I thank all of the people who contributed to the design of the ground-based sensing system, the development and programming of algorithms for video processing, and the collection of data in the field. First, I would like to thank Dr. Brian Steward and Dr. Dev Shrestha. Dr. Shrestha preceded me as a Ph.D. student under the direction of Dr. Steward at ISU, and the main focus of their work was the development of the ground-based sensing system that was used for ground-reference data collection of corn plant population in this work. This system was necessary to generate the extensive ground-reference dataset in my work, and I appreciate the assistance of Dr. Shrestha and Dr. Steward in learning how to operate this system in the field. I am also grateful for the programming work of Mr. Michael Collins, Mr. Neal Buchmeyer, and others who preceded them. Their work greatly advanced the ESCOPE software environment for operating the sensing system, sequencing video frames, segmenting images, and counting corn plants. The ESCOPE software was also crucial for generating the ground-reference dataset for corn plant population in my work. I am also thankful for the folks at Pioneer who provided the all-terrain vehicle for ground-reference data collection in 2004 and who designed the hardware for retrofitting the sensing system on the vehicle. Their design made the ground-reference data collection effort run much more smoothly. A very special thanks is due to Ms. Sandra Wenke who spent many early mornings, many long and hot afternoons, and many late evenings out in the field collecting data with me during the summer of 2004. For the times when the circumstances of the data collection effort tested my temper, I appreciate her patience. Her enthusiasm and dedication to our project was very much appreciated and needed. For completion of this project, I am also indebted to my friends and former co-workers at the Institute for Technology Development (ITD) in Urbana, IL. All of the remote sensing imagery for this work was provided free of charge by ITD, and I am grateful for all the times they have let me borrow their hand-held radiometer and other field equipment. Particularly, I thank Mr. Tim
Gress and Mr. Ken Copenhaver of ITD. It was through my work experiences with their organization that I elected to go to graduate school and focus in the areas of remote sensing and precision agriculture in the first place. I appreciate their continued support of my research endeavors through their friendship and through their generous contributions of hyperspectral remote sensing imagery. Finally, I would like to give a special thanks to Dr. Brian Steward who served as my co-major professor and mentor for the corn population sensing work. Several different components had to come together in order for this corn population project to be completed, and Dr. Steward was very good at managing the different components to insure that the project was successful in the end. I appreciated his constant support and advice as I progressed on this project. Dr. Steward should also be credited with originally sparking my interest in ISU at the 2002 ASAE meeting in Chicago.

I would like to thank the members of my dissertation committee for their suggestions on my research direction and on the improvement of this draft. I also thank them for supporting my continued professional development. These people are Dr. Bill Batchelor, Dr. Brian Steward, Dr. Amy Kaleita, Dr. Matt Helmers, and Dr. Petrutza Caragea. I thank Dr. Batchelor and Dr. Steward for being excellent mentors during the process of obtaining this degree. I also thank Dr. Kaleita and Dr. Helmers for their contributions and support of my efforts with the Preparing Future Faculty (PFF) program at ISU. A special thanks is also due to Dr. Caragea for her guidance on the statistical problems encountered in this work.

Dr. Amy Kaleita also deserves special recognition for taking over the management of the Cargill research funds allocated for my work after Dr. Batchelor left ISU in December of 2004. I appreciate that I was allowed to remain in my office space and keep my desk and computer while I finished my research work and dissertation. I also appreciate her expertise and advice on the remote sensing part of the corn population sensing work.

I am also grateful for the friendships and professional connections that I have established with several other graduate students in the department. The most prominent of these people are Brian Gelder, Kendall DeJonge, Jeremiah Davis, and James Bosserd.

Finally, I thank my family and friends for their sustained and unconditional love and support throughout my life. I know you are all having a hard time believing it to be true. But after eleven years, yes, I am finally done with college.
Kelly Robert Thorp was born in Urbana, Illinois, USA, on September 26, 1976. He was raised on a farm near Gibson City, Illinois, where his parents grew corn and soybeans to earn their living and support their family. Both of his parents came from families who also earned their living through farming, and ancestors from both sides of his family have farmed in the central Illinois area since the mid-1800’s. From birth, Kelly was immersed in the activities of farm life in central Illinois.

After a brief stint at a private university in Indiana, Kelly enrolled at Parkland College in Champaign, Illinois, in January 1996. He completed his basic undergraduate engineering curriculum at Parkland, and graduated with an Associate of Science in Engineering Science in May 1998. Given his lifelong background in agriculture, he then decided to pursue a degree in agricultural engineering and enrolled at the University of Illinois at Urbana-Champaign (UIUC) in August 1998. While completing his undergraduate degree, he obtained part-time employment at the midwest office of the Institute for Technology Development (ITD)/Spectral Visions in Urbana, Illinois. There, he supplemented his education with practical experience in use of remote sensing for applications in precision agriculture. In May 2000, he completed a Bachelor of Science degree in Agricultural Engineering from UIUC.

Before completing his undergraduate degree, Kelly was offered a graduate fellowship from the Jonathan Baldwin Turner (JBT) Graduate Fellowship Program through the College of Agricultural, Consumer, and Environmental Sciences (ACES) at UIUC. He accepted this position and began working on a Master of Science in Agricultural Engineering in August 2000. As an M.S. student, Kelly’s research focused on the use of remote sensing imagery to develop weed maps for variable-rate applications of herbicide in soybeans. A significant component of the work involved testing the performance of remote sensing-based herbicide applications in the field, and reductions in herbicide application volume were on the order of 30% with no change in the level of weed control. In December 2002, Kelly completed the M.S. degree from UIUC, and three peer-reviewed publications were generated from his work. In June of 2003, Kelly left central Illinois and moved to central Iowa to begin working on a doctoral degree in agricultural engineering at Iowa State University.
NAME OF AUTHOR: Kelly Robert Thorp

DATE AND PLACE OF BIRTH: September 26, 1976, Urbana, IL

DEGREES AWARDED:
- B.S. in Agricultural Engineering, University of Illinois at Urbana-Champaign, 2000
- M.S. in Agricultural Engineering, University of Illinois at Urbana-Champaign, 2002

PROFESSIONAL EXPERIENCE:
- Graduate Research Assistant, Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA, 6/2003 - present
- Graduate Research Assistant, Department of Agricultural Engineering, University of Illinois at Urbana-Champaign, 8/2000 - 1/2003

HONORS AND AWARDS:
- Midwest Central ASABE Graduate Student of the Year, 2006
- Harold and Katharine Guy Graduate Fellowship, 2006
- ASAE IET Select Paper Award, 2005
- Premium for Academic Excellence (PACE) Award, 2003
- Jonathon Baldwin Turner Fellowship, 2000-2002
- Agricultural Engineering Top Transfer Student Award, 1999
- Jane Ryoti Mathematics Scholarship, 1998

PROFESSIONAL AFFILIATIONS:
- American Society of Agricultural and Biological Engineers (ASABE), 1998-present
- Alpha Epsilon Agricultural Engineering Honor Society, 1999-present
- Engineers for a Sustainable World, 2006-present
- SPIE: International Society for Optical Engineering, 2004
- Professional Engineer Intern in the State of Illinois, 2000

REFEREED PUBLICATIONS: