COTTON YIELD FORECASTING FOR THE SOUTHEASTERN UNITED STATES

By

TAPAN BHARATKUMAR PATHAK

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To, my loving wife, my parents, and my family

ACKNOWLEDGMENTS

My PhD dissertation was one of the most important life changing journeys of my life. I have seen good times as well as hard times throughout my PhD student life. I enjoyed the good times and tried to learn and improve from the experiences of hard time. PhD degree was like a dream for me and this dream would not have been possible without constant high quality guidance and support from my major advisor Dr. James W. Jones. He is one of the best advisors and I was fortunate to have him as my PhD supervisor. He always encouraged me to explore innovative ideas but at the same time made sure that those ideas are executed in a systematic scientific manner. I have seen so much professional improvement in myself in last 5 years, which was solely due to his valuable guidance. There is so much to say about Dr. Jones and it is hard to narrate on a single page, but I must say that I will have my deepest respect for him throughout my career and will always try to follow his advice for my own professional development.

I would also like to express my deepest respect and gratitude towards my PhD co-advisor Dr. Clyde W. Fraisse. He was always helpful in providing me with new research ideas. His enthusiasm over implementing them always motivated me. I have a lot to learn from him and this dissertation would not have been possible without his support. I would like to thank my committee members, Dr. David Wright, Dr. James Heaney, and Dr. Gregory Kiker for their valuable guidance and recommendations. I am also thankful to Dr. Kenneth Boote for allowing me use his lab instruments for my research and also for his valuable advice. I would also like to extend my acknowledgement to Dr. Hoogenboom and Dr. Jasmeet Judge for providing me with experimental data to run my analyses. I am greatly thankful to all my friends at McNair

Bostick Simulation Laboratory (MBSL) for keeping cheerful yet non-disturbing work environment. I really enjoyed working with them. I enjoyed the company of “super friend” Dr. McNair Bostick, I will always miss him.

I would like to thank all my family members for their constant support and motivation throughout my PhD. I wish I could list out and thank all of them individually, but I am afraid that it will take up many more pages. I would like thank my mom and dad for their blessings, inspiration, and unconditional care. No words are enough to show my gratitude towards them. Finally, I have an opportunity to thank the most beautiful, loving, caring woman of my life, Rucha for her countless sacrifices and constant support throughout my PhD. She has seen all the ups and downs with me and I could not have done my PhD without her company. At the end, I would like to acknowledge our little lucky charm, who is soon to arrive in this world.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

COTTON YIELD FORECASTING FOR THE SOUTHEASTERN UNITED STATES
By

Tapan Bharatkumar Pathak

August 2010

Chair: James W. Jones
Cochair: Clyde W. Fraisse
Major: Agricultural and Biological Engineering

Cotton is the most important fiber crops in the United States, accounting for approximately 20% of the total production in the world and more than $25 billion in products and services annually. The Southeastern United States holds a major share of total cotton production of the country. While evidence clearly shows an increase in cotton planted over the time, climate variability is a major concern that could adversely affect its production in the southeastern United States. An effective way to reduce agricultural vulnerability to climate variability is through the implementation of adaptation strategies such as crop yield forecasts to mitigate negative consequences or take advantage of favorable conditions. The use of climate forecasts and climate indices to forecast cotton yield using the CROPGRO-Cotton model and principal component regression model, respectively, were assessed in this study. Using the crop model, in-season updating of the cotton yield forecast with real weather data along with the climatology significantly improved the accuracy of the forecast. With principal component regression models, cotton yield forecasts with significant skills were obtained with a lead time of approximately two months before cotton planting. Results indicated good potential for forecasting cotton yield for the southeastern United States.

CHAPTER 1
INTRODUCTION

Cotton is the single most important fiber crop in the world, accounting for more than 35% of the total fiber production. In terms of the world rankings, the United States is ranked second in cotton production, accounting for 19.9% of total world production (USDA ERS, 2009). The United States cotton industry is one of the major economic drivers of the country accounting for more than $25 billion in products and services annually. The Southeastern United States holds a major share of total cotton production in the United States; approximately, one fourth of total produced comes from this region. For instance, Georgia is ranked second to Texas in total cotton produced in the United States. In recent years, there has been an increased need for fiber supply, which has triggered increased cotton production in this region. Increases in acreage planted to cotton in Georgia and Alabama were approximately 26% and 41% in the last decade, respectively (NASS, 2007).

Although cotton is considered as a drought tolerant crop, climate variability may adversely impact cotton production. Especially, cotton produced under rain fed conditions could be severely affected by a variable climate. El Niño Southern Oscillation (ENSO) is a dominant phenomenon of climate variability in this region and other locations worldwide. The ENSO phenomenon is governed by the shift in sea surface temperature (SST) in the Pacific, which affects inter annual climate variability across most parts of the world including the southeastern United States (Hansen et al., 1998; Jones et al., 2003; Ropelewski and Halpert 1986, Kiladis and Diaz 1989; Mo and Schemm 2008; Mennis 2001). Hansen et al. (1998) showed that ENSO phase significantly influenced six major crops, including cotton, in the southeastern United

States. Their results showed that cotton area harvested in Alabama, Georgia, Florida, and South Carolina were significantly influenced by ENSO.

Jones et al. (2003) stated that the main reason that the climate variability is often so devastating to agriculture is that we do not know what to expect in the next growing season. Effective application of climate forecast and climate indices may provide an opportunity to tailor agricultural decisions for higher economic returns to growers. There have been many studies that evaluated the potential benefits of using climate forecasts on decision making processes in agriculture as a way to adapt to climate variability (Hansen et al., 2005; Podesta et al., 2002; Jones et al., 2003; Hansen et al., 1998).

An effective way to reduce agricultural vulnerability to climate variability is through an effective use of climate forecasts. One potential adaptation tool is yield forecasting based on climate information. Crop yield forecasts could be used by farmers to mitigate negative consequences of unfavorable climate, or benefit from anticipated favorable climate conditions (Baigorra et al., 2010). If growers know the expected cotton yield for the coming season, they may be able to decide on alternative management strategies to reduce the production risks (Jones et al., 2000; Hansen, 2005; Vedwan et al., 2005; Jagtap et al., 2002). Crane et al. (2010) conducted a research study to explore the potentials and constraints for farmers' application of climate forecasts. Their research documented farmers' perspectives of adjusting their decisions using climate forecasts to help them adapt to climate variability. Some of the documented long term decisions were crop type, buying appropriate crop insurance, and areas to plant. Pre-season crop yield forecasts would help growers in making the above mentioned decisions. For example, growers could purchase higher crop insurance coverage, plant different crop,

plant less etc. in order to compensate for an adverse effect of climate variability on their cotton yields.

Crop models have shown potential for use in forecasting crop yield if climate data are provided to the model is in terms of a forecast. The main advantage of a crop model is that it can produce a range of possible climate forecasts using uncertain climate forecasts. Before the start of crop growing season, weather is entirely uncertain, this translates into uncertainties in crop yield. If the model is updated during a season with observed weather data, some of the weather uncertainties are eliminated. Although weather uncertainties at the later stages of crop growth still impact the final crop yield, it may be possible to improve the accuracy of crop yield prediction by in-season updating the model with real weather data.

A model to simulate cotton growth and development has recently been developed for the Decision Support System for Agrotechnology Transfer (DSSAT) called CROPGRO-Cotton model (Jones et al., 1998; Hoogenboom et al., 2004). Since the CROPGRO-Cotton model is relatively new, it is important to calibrate and evaluate the model for t field conditions before it can be used for forecasting cotton yield. Hence, determining and understanding how sensitive the simulations of certain model processes are with respect to model parameters and estimating them is useful for model improvements.

While growth and development of crops are known to be influenced by weather during the growing season, it is a common practice to predict crop yield based on weather variables (Sakamoto, 1979; Idso et al., 1979; Walker, 1989; Alexandrove and Hoogenboom, 2001). However, crop yield predictions based on observed weather

cannot be made available before the planting season (Kumar, 2000). Attempts to obtain long-term forecasts using alternatives to weather variables such as climate indices that exhibit teleconnections with weather are limited. Large scale teleconnection indices greatly influence the climate and agriculture in the southeastern United States (Stenseth et al., 2003; Enfield, 1996; Bell and Jenowiak, 1994; Martinez et al., 2009). Using those large scale climate indices as an early indicator to cotton yield could provide valuable information to the growers in the Southeastern United States.

The overall research question addressed in this dissertation is “Do the use of climate forecasts and climate indices by crop model and empirical models provide potential in forecasting cotton yield for the southeastern United States?” Specific objectives include:

Objective 1: To conduct a global sensitivity analysis of the CROPGRO-Cotton model

Objective 2: To estimate model parameters and conduct an uncertainty analysis of the CROPGRO-Cotton model

Objective 3: To evaluate the use of in-season updates of cotton yield forecasts using the climate forecasts

Objective 4: To evaluate the use of climate indices for cotton yield forecasting in the Southeastern United States

CHAPTER 2
USE OF GLOBAL SENSITIVITY ANALYSIS FOR CROPGRO-COTTON MODEL DEVELOPMENT

Introduction

The need for information in agriculture is increasing due to market and economic pressures combined with the need for better management of our natural resources. Crop models, widely used as research and teaching tools, are now becoming important tools for agricultural decision makers, as the need for information in agriculture increases. Crop models range in complexity from simple ones with a few state variables to complex ones having large numbers of model parameters and state variables. The Decision Support System for Agrotechnology Transfer (DSSAT) (Jones et al., 1998; Hoogenboom et al. 2004) contains complex dynamic models that simulate crop growth and yield as a function of soil and weather conditions and crop management regimes. DSSAT can also be used to help researchers, extension agents, growers, and other decision-makers to analyze complex alternate decisions (Tsuji et al., 1998).

Cotton (*Gossypium hirsutum* L.) is the single most important textile fiber in the world, accounting for over 40 percent of total world fiber production. While some 80 countries from around the globe produce cotton, the United States, China, and India together provide over half the world's cotton. The United States, while ranking second to China in production, is the leading exporter, accounting for over one-third of global trade in raw cotton (MacDonald, 2000). Due to the importance of cotton in the world in general, and in the southeastern USA in particular, a model to simulate cotton growth and development is currently being developed by the DSSAT crop modeling group. The Cropping System Model (CSM, Jones et al., 2003) contains models of 21 crops based on the CROPGRO, CERES and other models. The new cotton model has been

developed using the CSM-CROPGRO crop template that allows its integration with other modules of the cropping system (Messina et al., 2004). The CROPGRO development team has used this approach in creating models for different species, including brachiaria grass (Giraldo et al., 1998), tomato (Scholberg et al., 1997), and velvet bean (Hartkamp et al., 2002, Boote et al., 2002). CROPGRO was originally developed as a process-oriented model for grain legumes, based on the SOYGRO, PNUTGRO, and BEANGRO models that consider crop carbon, water, and nitrogen balances (Boote et al., 1998). Its ability to represent different crops is attained through input files that define species traits and cultivar attributes (Boote et al., 2002). Outputs from CROPGRO models depend on a large number of model parameters associated with the species traits and cultivar attributes.

Determining and understanding how sensitive the simulations of certain model processes are with respect to model parameters is useful for guiding model developers. The effects of particular model parameter on a given output can be determined by measuring the relative influence of the model parameter on model output. Sensitivity analysis is useful for identifying the most and the least important model parameters to the given model output so that it can contribute to the simplification of a model (Saltelli et al., 2000). There are a number of methods and techniques available for performing sensitivity analysis (Saltelli et al., 2004), local and global sensitivity analyses are the most commonly used methods.

Ruget et al. (2002) performed a local sensitivity analysis on the STICS crop simulation model (Brisson et al., 1998) to determine how sensitive the simulation of processes in each module was to the model parameters. Leaf area index was sensitive

to all the model parameters of the leaf area index module (crop density, rate of LAI growth, and density effect on tillering), the cumulated root length was sensitivity to two of the model parameters of the root module (rooting depth for half water absorption, and rate of root deepening), whereas mineralization was most sensitive to humification depth. Xie et al. (2003) conducted local sensitivity analysis of the ALMANAC model (Kiniry et al., 1992) to input variables such as solar radiation, rainfall, soil depth, soil plant available water, and runoff curve number and the impact on grain yield of sorghum and maize. They found that runoff curve number change had the greatest impact on simulated yield.

Global sensitivity analysis differs from local methods by accounting for the variance of the model output associated with model parameters over their entire range of uncertainty. Homma and Saltelli (1996) explored methods of global sensitivity analysis of nonlinear models to calculate the fractional contribution of model parameters to the variance of model predictions. Makowski et al. (2004) used global sensitivity analysis to determine the contribution of generic model parameters to the variance of crop model predictions. A sensitivity analysis was performed for three output variables of the AZODYN wheat model (Jeuffroy and Recous, 1999) that included grain yield, grain protein content, and the nitrogen nutrition index. Out of thirteen different model parameters, five were found to have the most influence on grain yield and grain protein content. The only model parameter that affected the nitrogen nutrition index was the ratio of leaf area index to critical nitrogen concentration. This study concluded that model parameters with the least influence on important simulated processes may not need to be accurately estimated.

A sensitivity analysis allows modelers to rank model parameters in order of their influence on model output. Based on the rankings it can be used to identify model parameters that need a high accuracy in their estimates. Sensitivity analysis can also be used to check whether the behavior of the model output is as expected with respect to change in the input.

For the CSM crop models, it is practically impossible to measure or estimate all the model parameters with a high level of accuracy. The CSM-CROPGRO-Cotton model under development was initially parameterized using data from the literature (Messina et al., 2004) and evaluated for different environmental conditions using those parameters. Thus, there was uncertainty in the values of model parameters and how they may affect the output.

This study was conducted for two purposes. The first objective was to determine whether the global sensitivity analysis method would provide information on model performance that differs from the simpler local sensitivity method. This type of sensitivity analysis had not been used in the past with the CSM model. The central hypothesis in this case was that global sensitivity analysis will lead to improved understanding of the importance of model parameters since it accounts for the variance of model output associated with the variance of model parameters over the range on uncertainty in each parameter. The second objective was to determine how sensitive the prototype cotton model predictions are to an important subset of its crop growth and development parameters. Although the prototype model was based on an existing crop model it was not clear how these model parameters would affect the most critical outputs, such as yield and season length. We also did not know how these effects would differ between

rainfed and irrigated conditions. We hypothesized that the sensitivity of yield and season length to changes in model parameters does not vary with weather and irrigation. Results from this study were needed to guide further model parameter estimation efforts for improving the cotton model.

Materials and Methods

Overview of CSM-CROPGRO Cotton Model

The cotton model is based on the modular code of the CSM-CROPGRO model (Jones et al., 2003). This model simulates crop growth and development independent of location, season, and crop management system. Its flexible physiological framework provides a convenient template to implement a cotton model that can be immediately integrated with other crop models (Messina et al., 2004). CSM-CROPGRO is composed of several modules that make up a land unit in a cropping system. The primary modules are crop, soil, weather, soil-plant-atmosphere, and management. The soil module integrates information from four sub modules: soil water, soil temperature, soil carbon, and nitrogen dynamics. The soil is represented by a one-dimensional profile, consisting of a number of vertical soil layers. The main function of the weather module is to read or generate daily weather data required by the model, including minimum and maximum air temperatures, solar radiation, and precipitation. The soil-plant-atmosphere module computes daily soil evaporation and plant transpiration while the management module determines when field operations are performed by calling sub modules related to planting, harvesting, inorganic fertilization, irrigation, and application of crop residues or organic materials. The Crop module can predict the growth and development of a number of different crops, each crop has its own model parameter files. These modules describe the time changes that occur in a land unit due to management and weather.

The CSM-CROPGRO model has three sets of parameters that account for differences in development, growth, and yield between species, ecotypes, and cultivars (Boote et al., 2003). Cultivar parameters are specific to a particular variety, Ecotype parameters are for a group of cultivars, and Species parameters are common to all cultivars. Mainly the model cultivar parameters (Table 2-1) are vital to consider for sensitivity analysis.

Site and Experiment Description

Sensitivity analyses were conducted for two cropping seasons, 2003 for which we had observed data collected in an experiment conducted at the C.M. Stripling Irrigation Research Park (SIRP), Camilla, GA (31°11N, 84°12W), and 2000 which was a dry year to compare the results from irrigated and rainfed conditions. Daily weather data consisting of maximum and minimum temperature, solar radiation, precipitation, and wind speed were obtained from a local weather station at SIRP. Maximum temperatures varied from 24 to 33ºC; the minimum temperatures varied from 10 to 22ºC; and the average temperatures varied from 18 to 27ºC. The extremes for minimum temperatures occurred at the end of the growing season. Long term average precipitation (1939 to 2003) for June and July were 131.1 mm and 150.9 mm, respectively. During the 2000 cropping season the total precipitation for June and July were 63.2 mm and 102.4 mm, respectively. In 2003, June and July precipitation totaled 139.9 mm and 203.6 mm, respectively. Unlike 2003, no field experiment was conducted during 2000. The main reason of performing the sensitivity analysis for this year was to include a dry cropping season. Comparison of dry and wet years would allow us to evaluate the hypothesis that the importance of model parameters do not vary with irrigated and rainfed conditions.

A 34 ha field was planted with a late maturing cotton variety, DP 555, using a conventional tillage system. The field was sown during the first week of May with a plant population of approximately 110,000 plants per hectare. The soil type at the study site was classified as an Orangeburg loamy sand (Fine-loamy, siliceous, thermic Typic Paleudults) (Source - Mitchell County SCS Map, Soil Conservations Service). The experiment had two treatments; one was rainfed and the other was irrigated. All other inputs were the same for both treatments.

Sensitivity Analysis

The principle of sensitivity analysis is firstly to generate output variability associated with the variability of input, and secondly to assign the simulated output variability to the model parameters that affect it the most (Ruget et al., 2002). The most crucial step in sensitivity analysis is the selection of the model parameters and their uncertainty ranges. Including a large number of model parameters for global sensitivity analysis would result in an unrealistically high number of simulation runs and impractical computational load (Thorsen et al., 2001).

Local sensitivity analysis was performed on the entire set of cultivar model parameters listed in Table 2-1 and also on one of the species parameters, light extinction coefficient (KCAN). Messina et al. (2004) recommended that KCAN should be considered as a cultivar or ecotype parameter because initial estimates of KCAN from literature data showed variations between and within seasons, with planting dates, and between cultivars (Rosenthal and Gerik, 1991; Milroy et al., 2001; Milroy and Bange, 2003; Bange and Milroy, 2004). The main reason for including KCAN in the sensitivity analysis was to help answer the question whether KCAN was sufficiently important parameter to warrant its inclusion as an ecotype parameter. Results from this local

sensitivity analysis were used to select model parameters for the global sensitivity analysis.

Local sensitivity

Local sensitivity is often used to provide a normalized measure for comparing sensitivity of a model to several parameters. In order to measure relative sensitivity of an output relative to a particular model parameter, only that parameter is changed in the vicinity of a base value; all other parameters are fixed to their base values. Local sensitivity was calculated for model responses using the base and + and - 5% changes in the base value. Sensitivity indices were obtained by computing the change in the output relative to changes in parameters. Relative sensitivities for dry matter yield and season length were defined as follows:

$$\sigma_r(Y/\theta) = (\partial Y/Y)/(\partial\theta/\theta) \quad (2\text{-}1)$$

$$\sigma_r(M/\theta) = (\partial M/M)/(\partial\theta/\theta) \quad (2\text{-}2)$$

Where Y is simulated dry matter yield and M is simulated length of the growing season obtained for each level of an individual model parameter (θ) while keeping all other model parameters at their base values. $(\partial Y/Y)$ and $(\partial M/M)$ represent fraction changes in simulated outputs for dry matter and season length relative to the fraction changes in inputs $(\partial\theta/\theta)$, respectively. $\sigma_r(Y/\theta)$ and $\sigma_r(M/\theta)$ represent local sensitivities for the dry matter and season length, respectively.

Global sensitivity

Measuring model sensitivity for each model parameter θ separately with all other model parameters fixed at their single base values prevents the detection and

quantification of interactions. A key aspect of most global sensitivity methods is the ability to take these interactions into account.

Factorial design is a method of global sensitivity analysis that allows for simultaneous evaluation of the influence of many model parameters. It follows the classical theory of experimental design where nature is replaced by the simulated crop model (Box and Draper, 1987). In this study, selection of model parameters for global sensitivity analysis was based on results from the local sensitivity analysis. A simplification of the deterministic model can be used to represent the two output state variables, dry matter yield (Y) and season length (M), as a function of model parameters:

$$Y = f(\theta) \text{ and } M = f(\theta) \tag{2-3}$$

Where θ = model parameters

Complete factorial design uses all possible combinations of chosen factors and levels. For example, eight model parameters and three levels would create 3^8 combinations. For such a factorial experiment, the analysis can be expressed by decomposing the function Y= *f* (θ) and M = *f*(θ) into main effects and interactions:

$$Y_{abcdefgh} = \mu + \alpha_a + \beta_b + \ldots + \eta_{abcdefgh} \tag{2-4}$$

$$M_{abcdefgh} = \mu + \alpha_a + \beta_b + \ldots + \eta_{abcdefgh} \tag{2-5}$$

$Y_{abcdefgh}$ = *f* (a,b,c,d,e,f,g,h) and $M_{abcdefgh}$ = *f* (a,b,c,d,e,f,g,h) denote the model responses of dry matter yield and growing season length, respectively, when θ_1= a, θ_2 = b, θ_3 = c, θ_4 = d, θ_5 = e, θ_6 = f, θ_7 = g and θ_8 = h; μ is the overall mean of the model responses; α_a, β_b, …, ε_h represent the main effects of model parameters θ_1, θ_2…θ_8

when θ_1= a, θ_2 = b, …θ_8 = h; η_{ab} is the interaction between *a* and *b*, η_{gh} is the interaction between g and h, and so on.

The overall response variability can be separated into factorial terms as follows:

$$\sum_{abcdefgh}(Y_{abcdefgh}-\mu)^2 = m\sum_a \alpha_a^2 + m\sum_b \beta_b^2 + \sum_{ab}\eta_{ab}^2 + ... + \quad (2\text{-}6)$$

Where $\sum_{abcdefgh}(Y_{abcdefgh}-\mu)^2$, or total sum of squares (SS_T), represents the total variability in the model responses, $m\sum_a \alpha_a^2$ is the sum of squares (SS_1) associated with the main effect of *m* levels of model parameters θ_1, and so on. The NCSS (Hintze, 2004) statistical software was used for calculating the sum of squares of the main effects for the complete factorial design.

For the sensitivity analysis of a deterministic model, the main interest lies in comparing the contributions of the factorial terms to the total variability. The main effect sensitivity ($S_{i=1 \text{ to } 8}$) indicates the relative importance of individual model parameter uncertainty and can be calculated by dividing the corresponding main effect sum of squares by the total sum of squares (Equation 7).

$$S_i = \frac{SS_i}{SS_T} \quad (2\text{-}7)$$

Interaction sensitivity indices are measures of the interactive influences of the model parameters on the output variance and were calculated by dividing the interaction sum of squares of the model parameters by the total sum of squares.

Global sensitivity indices indicate the overall impact of model parameters on the output variance when model parameters vary over their entire range of uncertainty.

Global sensitivity indices for selected model parameters were calculated by following equation (Equation 8).

$$S_{global(i)} = \frac{SS_i + SS_{i,\sim i}}{SS_T} \tag{2-8}$$

Where, $S_{global(i)}$ indicates global sensitivity index, SS_i is the main effect sum of squares, $SS_{i,\sim i}$ is interaction sum of squares, and SS_T is total sum of squares, for parameters i = 1 to 8.

Global sensitivity analysis apportions the output variability to the variability in model parameters covering their entire range space, and hence it was important to decide the range of selected model parameters. The ranges of model parameters should be chosen such that they represent the expected extreme values of those parameters. The ranges of some of the model parameters SLAVR, KCAN, and XFRT were obtained from the literature (Bange and Milroy, 2000; Milroy et al., 2001; Milroy and Bange, 2003; Reddy et al., 1993; Reddy et al., 1992; Reddy et al., 1991; Messina et al., 2004).

Published data by Wright and Sprenkel (2006) on the ranges of different growth stages were used to determine the ranges of the model parameters that deal with crop growth duration. Information on certain crop growth stages was not available in the published literature so in order to make maximum use of available dataset it was assumed that the ratio of parameters on crop growth stages for the cotton model were same as the ratio of those parameters for the soybean model. Data for soybean parameters were obtained from Boote et al. (2003). In the case of SFDUR and PODUR, no information on ranges was available. Instead of using an arbitrary range, the

percentage variance from the mean value for soybean parameters was applied to the base cotton model parameters.

Results and Discussion

Local Sensitivity Analysis

Model parameters that were selected for global sensitivity analysis based on the initial local sensitivity analysis are shown in Tables 2-2 and 2-3. Sensitivity indices of the other parameters were either zero or very small. For both treatments and years, KCAN was the parameter that most affected the dry matter yield. For the 2000 rainfed condition, magnitudes and orders of local sensitivity indices were different from 2000 irrigated condition and 2003 irrigated and rainfed conditions (Table 2-2, Figure 2-1). As an example, dry matter yield was more sensitive to XFRT than EM-FL for 2000 rainfed condition whereas it was vice versa for all other cases. One reason for such differences could be because the model is non-linear in its responses and local sensitivity analysis does not consider the range of uncertainty.

Only three model parameters, EM-FL, FL-SD and SD-PM, affected the response of season length. The time between first seed and maturity (SD-PM) was the model parameter that influenced season length the most (Table 2-3).

Global Sensitivity Analysis

Based on the local sensitivity analysis results of all the cultivar model parameters and one species model parameter, eight parameters were selected for global sensitivity analysis (Table 2-4). Table 2-5 and 2-6 shows the calculated global sensitivity indices for dry matter yield and season length taking into account all main effects and interactions related to corresponding model parameter. The first ten interaction terms in order of their magnitudes of sensitivity indices are listed in Table 2-5.

Unlike local sensitivity indices, global sensitivity indices were consistent in terms of order of sensitivity across rainfed and irrigated conditions and years. For a given management condition and year, SLAVR had the highest sensitivity index followed by KCAN. Two model parameters, PODUR and SFDUR, showed the least influence on dry matter yield across both treatments and years. For lower specific leaf areas, thicker and smaller leaves would reduce light capture and net photosynthesis. On the other hand for higher specific leaf areas, leaves are thinner and larger resulting in increased light capture and hence increased net photosynthesis. Thus changing SLAVR would indirectly affect canopy net photosynthesis which eventually affects growth and yield. KCAN was the second most important parameter for dry matter yield. The light extinction coefficient, KCAN, is used to compute light interception depending on leaf area index. The highest interaction effect was obtained between KCAN and SLAVR. The model parameter XFRT was the third most important model parameter for crop yield. Parameter values of SLAVR and KCAN mainly control leaf expansion and light capture and hence control daily assimilates. XFRT regulates the partitioning of daily assimilates that goes to seed. XFRT being the third most important parameter, its interaction with SLAVR and KCAN being the second and third most important interactions was logical.

For 2000, irrigated and rainfed conditions provided similar sensitivity indices (Figure 2-2). There were small differences in the values due to irrigation and year; however the order of importance of model parameters was consistent. Year 2003 was a wet year and hence even with the rainfed conditions, the indices were consistent in

terms of values and order. Overall, global sensitivity indices did not show variations in order of importance with different treatments.

Season length was sensitive to only three model parameters, EM-FL, FL-SD, and SD-PM. Season length is mostly determined by parameters that control the duration of different crop growth stages. In our study three duration model parameters were selected that covered the whole crop season length. SD-PM was the most important model parameter, partly due to greater uncertainty in this parameter, affecting season length, followed by FL-SD, and EM-FL respectively (Table 2-6).

Comparison of Local and Global Sensitivity Analysis Results

For comparing local and global sensitivity analysis results, only rankings were taken into consideration. The reason for that was because local sensitivity index is the ratio of percentage change in the output response to the percentage change in the model parameter, whereas the global sensitivity index is the measure of percentage contribution of an individual model parameter to overall output variance.

Based on local sensitivity results, KCAN was the most important model parameter for cotton dry matter yield followed by SLAVR, which was opposite to global sensitivity results. The main reason for these differences was because local sensitivity analysis focuses on local impact of the model parameter on the model response where model parameters varied in small intervals around the base value of the model parameter. For nonlinear models, finding the most important model parameter with such sparse domain coverage may be misleading. On the other hand, global sensitivity analysis takes into account the main effects and interactions between parameters over their entire uncertainty ranges.

For this study, irrigated and rainfed conditions for dry and wet years were used to compare local sensitivity and global sensitivity analyses results. The model parameter rankings did not change with years and treatments for global sensitivity analysis, whereas they varied among the years and treatments for local sensitivity analysis. Both methods provided similar results for two variables (PODUR and SFDUR) showing that model sensitivity over the range of parameter uncertainty was small. Such information can be useful to focus additional research and possibly to simplify the model by reducing the number of model parameters that one has to estimate.

Season length response was sensitive to only three model parameters, EM-FL, FL-SD, and SD-PM. Out of these three SD-PM was the most important model parameter followed by FL-SD, and EM-FL, respectively. These rankings were the same for both sensitivity analysis methods. The reason for season length response being sensitive to only three model parameters was because those model parameters were the crop growth duration model parameters selected for this analysis.

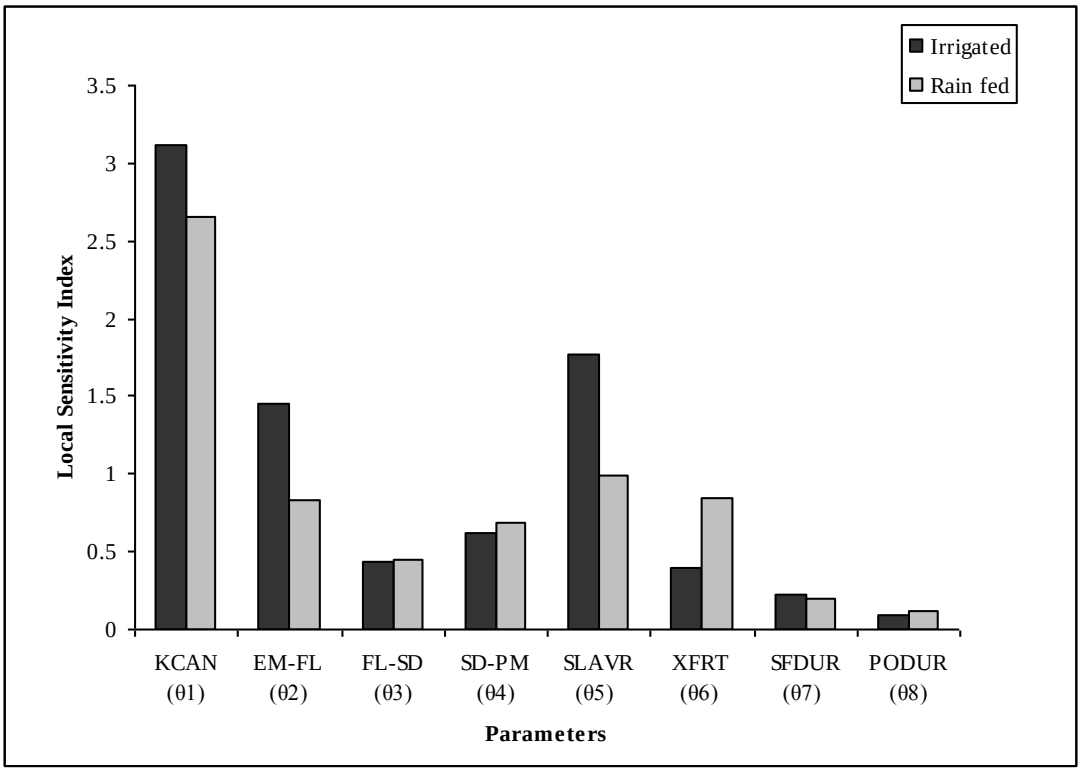
Conclusions

This study evaluated how sensitive the cotton model predictions were to a selected set of model parameters. Local and global sensitivity analyses were used to determine dry matter yield and season length sensitivity to model parameters under irrigated and rainfed conditions for two cropping seasons. Results demonstrated that, in accordance with our first hypothesis, global sensitivity analysis improved our understanding of how sensitive the prototype cotton model was to the selected set of parameters over the ranges of parameter uncertainties. In addition to accounting for the variance of model output associated with the variance of model parameters over the entire range on uncertainty, it had an advantage of considering the interactions among

model parameters. The most influencing model parameter on dry matter yield was the specific leaf area (SLAVR). Local sensitivity analysis indicated that the extinction coefficient (KCAN) was the most influencing model parameter.

The global sensitivity analysis results also demonstrated that, consistent with our second hypothesis, sensitivity of dry matter yield and season length to the selected set of model parameters did not vary between irrigated and rainfed conditions or with years. However, that was not true for local sensitivity analysis. Results from this study indicated that more research is needed to reduce the range of uncertainties of both KCAN and SLAVR. Experiments have shown variations in KCAN values for different cultivars and hence it was suggested that it should be included in ecotype set of model parameters rather than in species file. That suggestion was supported from the results of this study. The parameters selected for this study were associated with crop growth and development. Additional studies are needed to assess model sensitivity to soil water and nitrogen parameters which were held constant for this study.

Global sensitivity analysis can be a valuable tool for application with large, highly non-linear models, such as the DSSAT-CSM models. However, the use of a complete factorial design and analysis of variance can result in a large number of simulation runs when there are many model parameters. The choice of model parameters to be evaluated should be considered with care, taking into account available resources.



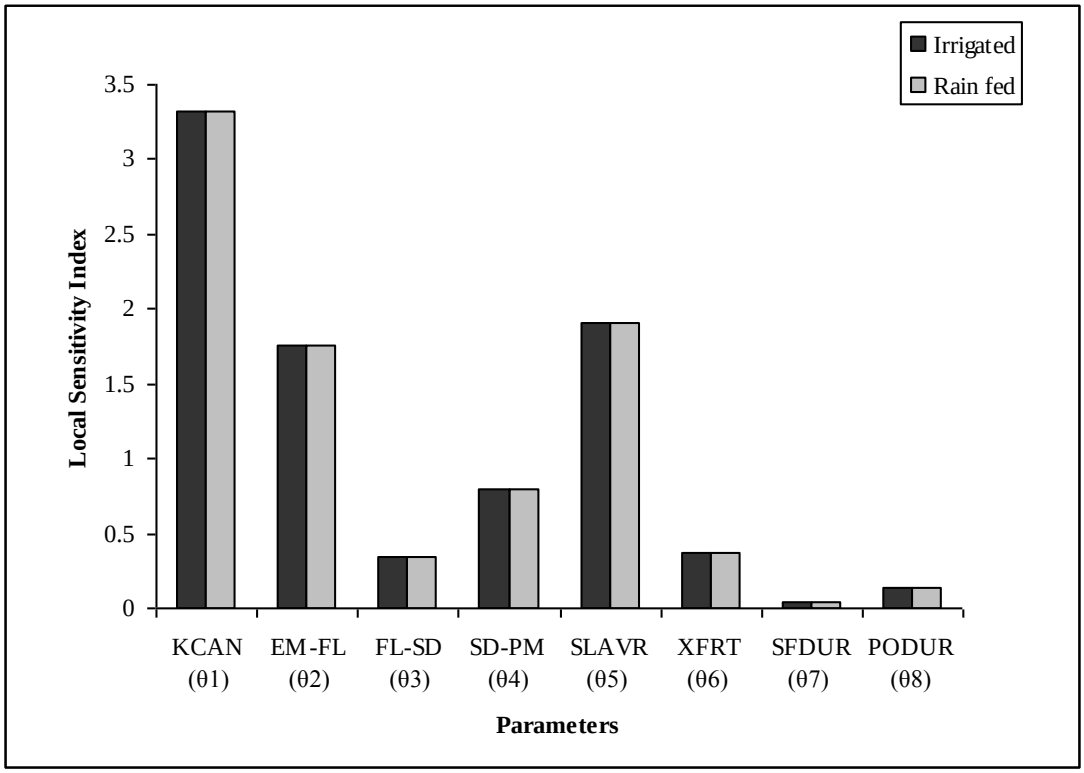


Figure 2-1. Local Sensitivity indices for seed cotton yield for irrigated and rainfed conditions for 2000 (top) and 2003 (bottom).

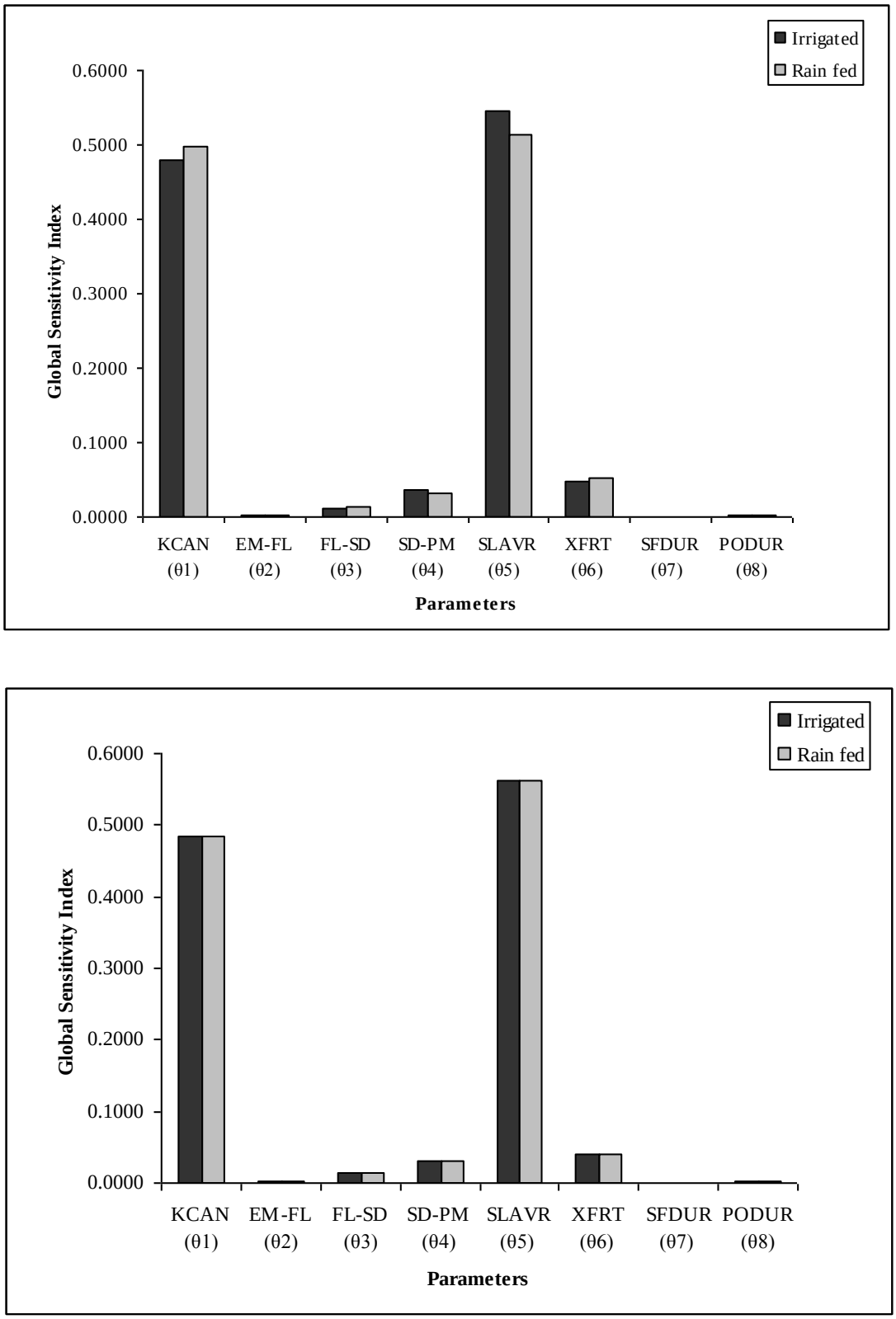


Figure 2-2. Global Sensitivity indices for seed cotton yield for irrigated and rainfed conditions for the year 2000 (top) and 2003 (bottom).

Table 2-1. List of cultivar parameters in CSM-CROPGRO–Cotton Model

Numbers	Parameters	Definitions
1	CSDL	Critical Short Day Length below which reproductive development progresses with no day length effect (for short day plants) (hour)
2	PPSEN	Slope of the relative response of development to photoperiod with time (positive for short day plants) (1/hour)
3	EM-FL	Time between plant emergence and flower appearance (R1) (photothermal days)
4	FL-SH	Time between first flower and first boll (R3) (photothermal days)
5	FL-SD	Time between first flower and first seed (R5) (photothermal days)
6	SD-PM	Time between first seed (R5) and physiological maturity (R7) (photothermal days)
7	FL-LF	Time between first flower (R1) and end of leaf expansion (photothermal days)
8	LFMAX	Maximum leaf photosynthesis rate at 30 C, 350 ppm CO2, and high light (mg CO2/m2-s)
9	SLAVR	Specific leaf area of cultivar under standard growth conditions (cm2/g)
10	SIZLF	Maximum size of full leaf (three leaflets) (cm2)
11	XFRT	Maximum fraction of daily growth that is partitioned to seed + shell
12	WTPSD	Maximum weight per seed (g)
13	SFDUR	Seed filling duration for boll cohort at standard growth conditions (photothermal days)
14	SDPDV	Average seed per boll under standard growing conditions (#/boll)
15	PODUR	Time required for cultivar to reach final boll load under optimal conditions (photothermal days)
16	KCAN*	Canopy light extinction coefficient (* species parameter)

Table 2-2. Local sensitivity indices with respect to selected model parameters for dry matter yield

Parameter(s)	Base	Year 2000		Year 2003	
Θ	Value	Irrigated	Rainfed	Irrigated	Rainfed
KCAN (θ_1)	0.72	3.11	2.65	3.32	3.32
EM-FL (θ_2)	27.87	1.45	0.83	1.76	1.76
FL-SD (θ_3)	11.65	0.43	0.44	0.34	0.34
SD-PM (θ_4)	27.68	0.62	0.69	0.79	0.79
SLAVR (θ_5)	170.00	1.76	0.99	1.91	1.91
XFRT (θ_6)	0.72	0.40	0.84	0.37	0.37
SFDUR (θ_7)	35.00	0.21	0.19	0.03	0.03
PODUR (θ_8)	8.00	0.09	0.12	0.13	0.13

Table 2-3. Local sensitivity indices with respect to selected model parameters for season length

Parameter(s)	Base	Year 2000		Year 2003	
Θ	Value	Irrigated	Rainfed	Irrigated	Rainfed
KCAN (θ_1)	0.72	0.00	0.00	0.00	0.00
EM-FL (θ_2)	27.87	0.31	0.32	0.32	0.32
FL-SD (θ_3)	11.65	0.10	0.10	0.10	0.10
SD-PM (θ_4)	27.68	0.42	0.43	0.43	0.43
SLAVR (θ_5)	170.00	0.00	0.00	0.00	0.00
XFRT (θ_6)	0.72	0.00	0.00	0.00	0.00
SFDUR (θ_7)	35.00	0.00	0.00	0.00	0.00
PODUR (θ_8)	8.00	0.00	0.00	0.00	0.00

Table 2-4. Model parameters and their range of uncertainty selected for the global sensitivity analysis

Parameter(s) Θ	Uncertainty Range Minimum	Base	Maximum	Reference For Uncertainty Range
KCAN (θ_1)	0.50	0.72	0.95	Rosenthal and Gerik, 1991;Milroy et al., 2001; Milroy and Bange, 2003; Bange and Milroy, 2004
EM-FL (θ_2)	27.72	27.87	28.02	Wright and Sprenkel, 2006
FL-SD (θ_3)	9.03	11.65	14.27	Wright and Sprenkel, 2006; Boote et al., 2003
SD-PM (θ_4)	21.46	27.68	33.91	Wright and Sprenkel, 2006; Boote et al., 2003
SLAVR (θ_5)	90.00	170.00	250.00	Reddy et al. 1993; Reddy et al., 1992; Reddy et al., 1991
XFRT (θ_6)	0.50	0.72	0.95	Reddy et al. 1993; Reddy et al., 1992; Reddy et al., 1991
SFDUR (θ_7)	31.12	35.00	38.88	Boote et al., 2003
PODUR (θ_8)	5.82	8.00	10.18	Boote et al., 2003

Table 2-5. Global sensitivity indices for dry matter yield including main effects and interactions

Parameter(s)	2000 Irrigated Sensitivity Indices		2000 Rainfed Sensitivity Indices		2003 Irrigated Sensitivity Indices		2003 Rainfed Sensitivity Indices	
θ	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices
KCAN (θ_1)	0.3848	0.4802	0.4109	0.4976	0.3761	0.4832	0.3762	0.4832
EM-FL (θ_2)	0.0002	0.0022	0.0002	0.0022	0.0002	0.0023	0.0002	0.0023
FL-SD (θ_3)	0.0057	0.0122	0.0065	0.0131	0.0058	0.0132	0.0058	0.0132
SD-PM (θ_4)	0.0203	0.0359	0.0196	0.0325	0.0165	0.0311	0.0165	0.0310
SLAVR (θ_5)	0.4491	0.5451	0.4262	0.5132	0.4549	0.5631	0.4550	0.5631
XFRT (θ_6)	0.0224	0.0479	0.0291	0.0532	0.0182	0.0400	0.0182	0.0400
SFDUR (θ_7)	0.0002	0.0009	0.0001	0.0010	0.0000	0.0011	0.0000	0.0011
PODUR (θ_8)	0.0004	0.0020	0.0007	0.0027	0.0010	0.0029	0.0010	0.0029
KCAN x SLAVR	0.0716	n/a	0.0651	n/a	0.0855	n/a	0.0855	n/a
KCAN x XFRT	0.0107	n/a	0.0100	n/a	0.0089	n/a	0.0089	n/a
SLAVR x XFRT	0.0096	n/a	0.0102	n/a	0.0082	n/a	0.0082	n/a
SD-PM x SLAVR	0.0065	n/a	0.0045	n/a	0.0059	n/a	0.0059	n/a
KCAN x SD-PM	0.0053	n/a	0.0046	n/a	0.0047	n/a	0.0047	n/a
KCAN x SD-PM x XFRT	0.0038	n/a	0.0029	n/a	0.0034	n/a	0.0034	n/a
FL-SD x SLAVR	0.0018	n/a	0.0017	n/a	0.0020	n/a	0.0020	n/a
KCAN x FL-SD	0.0015	n/a	0.0016	n/a	0.0016	n/a	0.0016	n/a
KCAN x SD-PM x SLAVR	0.0010	n/a	0.0006	n/a	0.0011	n/a	0.0010	n/a
EM-FL x FL-SD x SD-PM	0.0006	n/a	0.0006	n/a	0.0006	n/a	0.0006	n/a

Table 2-6. Global sensitivity indices for season length including main effects and interactions

Parameter(s)	2000 Irrigated Sensitivity Indices		2000 Rainfed Sensitivity Indices		2003 Irrigated Sensitivity Indices		2003 Rainfed Sensitivity Indices	
θ	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices	Main Effects & Interactions	Global Indices
EM-FL (θ_2)	0.0040	0.0339	0.0043	0.0344	0.0040	0.0343	0.0040	0.0343
FL-SD (θ_3)	0.1756	0.2058	0.1826	0.2127	0.1990	0.2307	0.1990	0.2307
SD-PM (θ_4)	0.7829	0.8142	0.7732	0.8038	0.7589	0.7896	0.7589	0.7896
EM-FL x FL-SD x SD-PM	0.0163	n/a	0.0154	n/a	0.0165	n/a	0.0165	n/a
FL-SD x SD-PM	0.0076	n/a	0.0072	n/a	0.0077	n/a	0.0077	n/a
EM-FL x SD-PM	0.0073	n/a	0.0071	n/a	0.0064	n/a	0.0064	n/a
EM-FL x FL-SD	0.0063	n/a	0.0067	n/a	0.0074	n/a	0.0074	n/a

Table 2-7. Model parameter rankings based on local and global sensitivity indices for dry matter yield

Parameter(s)	Year 2000		Year 2003	
Θ	Irrigated	Rainfed	Irrigated	Rainfed
Local Sensitivity Analysis Rankings				
KCAN (θ1)	1	1	1	1
EM-FL (θ2)	3	4	3	3
FL-SD (θ3)	5	6	6	6
SD-PM (θ4)	4	5	4	4
SLAVR (θ5)	2	2	2	2
XFRT (θ6)	6	3	5	5
SFDUR (θ7)	7	7	8	8
PODUR (θ8)	8	8	7	7
Global Sensitivity Analysis Rankings				
KCAN (θ1)	2	2	2	2
EM-FL (θ2)	6	6	6	6
FL-SD (θ3)	5	5	5	5
SD-PM (θ4)	4	4	4	4
SLAVR (θ5)	1	1	1	1
XFRT (θ6)	3	3	3	3
SFDUR (θ7)	8	8	8	8
PODUR (θ8)	7	7	7	7

Table 2-8. Model parameter rankings based on local and global sensitivity indices for season length

Parameter(s)	Year 2000		Year 2003	
Θ	Irrigated	Rainfed	Irrigated	Rainfed
Local Sensitivity Analysis Rankings				
EM-FL (θ2)	3	3	3	3
FL-SD (θ3)	2	2	2	2
SD-PM (θ4)	1	1	1	1
Global Sensitivity Analysis Rankings				
EM-FL (θ2)	3	3	3	3
FL-SD (θ3)	2	2	2	2
SD-PM (θ4)	1	1	1	1

CHAPTER 3
UNCERTAINTY ANALYSIS AND PARAMETER ESTIMATION OF CROPGRO – COTTON MODEL

Introduction

Applications of crop simulation models have become an important part of the agricultural research process. Because decision making processes may use results obtained from simulation models, consideration of model uncertainties in decision making processes has become increasingly important. Uncertainties in models can be categorized into three major sources: model parameters, model input data, and model structure. Crop models, such as CROPGRO-Cotton are complex and have many parameters. With limited measurement availability, estimates of model parameters are uncertain. Model structure is uncertain since it is typically a simplified representation of the system being studied. Lastly, model input data, such as initial conditions, are also imperfect to some extent and hence contribute towards output uncertainty (Makowski et al., 2006; Tolson and Shoemaker, 2008).

Different sources of uncertainties are important, however, the scope of this research was limited to addressing parameter uncertainty. Model parameters have been a significant source of uncertainty in model prediction in previous studies. Brazier et al. (2000) showed that parameters such as hydraulic conductivity have been recorded with a large variance for a single soil (Nielsen et al., 1973, Warrick and Nielsen, 1980) yet they are often input to the model as a single value. If model output is sensitive to hydraulic conductivity, the major portion of model output uncertainty could come from parameter uncertainty. Wang et al. (2005) performed parameter estimation and uncertainty analysis on crop yield and soil organic carbon simulated with the EPIC model where only parameter uncertainty was considered. In that study, a total of nine

corn yield and soil organic carbon related parameters were used for uncertainty analysis. Although only parameter uncertainty was considered, the observed corn yield and soil organic carbon fell within the 95% confidence limit of the predictions.

There have been significant advances in methodologies for uncertainty assessment. Blasone et al. (2008) referenced some of the widely used uncertainty analysis methods that include Classical Bayesian (Vrugt et al., 2003; Thiemann et al, 2001), Pseudo-Bayesian (Beven and Binely, 1992), data assimilation (Moradkhani et al, 2005), and multi model averaging methods (Georgekakos et al., 2004; Ajami et al., 2007). These methods differ in their underlying assumptions, complexity, and the way different sources of error are treated. Montanari, (2007) suggested that the selection of an uncertainty analysis method is subjective and should take into account issues such as model complexity, type of observed dataset available, and reliability of uncertainty assessment methods. A generalized likelihood uncertainty estimation (GLUE) technique introduced by Beven and Binley, (1992) is one of the most widely used and accepted uncertainty analysis techniques in environmental simulation modeling and it has also been used in crop modeling (Wang et al., 2005; He et al., 2009; Makowski et al., 2006). The main reasons for its popularity are its simple yet robust theory derived from Bayesian inference and its flexibility in implementation. Stedinger et al. (2008) counted a total of more than 500 citations of GLUE applications in various simulation modeling studies.

The GLUE methodology was developed out of the Hornberger-Spear-Young (HSY) method of sensitivity analysis (Whitehead and Young, 1979; Hornberger and Spear, 1981; Young, 1983). This method works on a phenomenon called equifinality.

The equifinality thesis suggests that there may be more than one parameter set that is acceptable for simulation and should be considered in assessing uncertainty in predictions (Beven, 2006). In the GLUE methodology, model parameter sets are weighted based on their agreement with related observations using subjective likelihood measures. Beven and Binley (1992) acknowledged that the choice of likelihood function used within the GLUE framework is subjective and the choice may greatly influence the resulting parameters and their uncertainties. These weights or probabilities are subsequently used to derive predictive uncertainty in output variables.

The CROPGRO-Cotton model is a part of Decision Support System for Agrotechnology Transfer (DSSAT) software package that includes models of 21 crops (Jones et al., 2003). It is a part of the CROPGRO family of crop models with many of the same level of details. Pathak et al. (2007) performed a global sensitivity analysis on the CROPGRO-Cotton model cultivar parameters and found that eight out of fifteen cultivar parameters were important to crop yield and physiological maturity outputs of the model. An accurate estimate of uncertainties associated with those important model parameters is needed for model applications and improvement. It may be difficult to find a single optimal fit of model parameters to observed datasets in complex simulation models due to the fact that there might be more than one parameter set that gives equally good results. This equifinality nature of this simulation model needs to be addressed. Parameter estimates of the cotton model have only been tested a few times under field experiments (Messina et al., 2004; Zamora et al., 2009). There have been no studies published in referred articles to address uncertainties associated with the CROPGRO-Cotton model.

Our main research question was, “What are the estimated values and uncertainties in genotype parameters and how do those translate into uncertainties in model outputs?” The objective of this research was to estimate the important model parameters and associated uncertainties using the GLUE technique.

Materials and Methods

Description of Field Sites and Measured Datasets

Data collected on four experiments at three sites were used for this study: 1) and 2) University of Florida - North Florida Research and Education Center (NFREC), located in Quincy, Florida, 3) University of Florida Plant Science Research and Education Unit, Citra, Florida, and 4) a cooperator’s farm located in Mitchell County, Georgia. Characteristics of the study sites including soil, weather, and management information are shown in Table 3-1. All field plots were planted with the full season cotton cultivar DeltaPine-555, the most widely grown cultivar in the Southeast USA. All four experimental sites were irrigated and fertilized during the cropping season. Timing and amounts of applied water and fertilizers were recorded.

Above ground biomass (leaf, stem, and boll) and leaf area index (LAI) were measured at each location at an interval of approximately two to three weeks with three to four replications. There were a total of 81 observations considered for this study consisting of 16 LAI, 24 leaf weight, 17 boll weight, and 24 stem weight measurements across all four experiments. In the vegetative stage sampling period, plants within one meter of row were cut at the soil surface and separated into leaf, stem, and bolls. Samples were then oven-dried at about 70°C for 48 hours and weighed to obtain dry biomass. LAI measurements were obtained using the LAI-2000 instrument (Li-Cor Inc., Lincoln, NE) for experiments 1 and 2. A leaf area meter (model LI-3100, Li-Cor Inc) was

used to measure LAI for experiments 3 and 4. Information on phenology (anthesis dates and maturity dates) were only collected for experiments 1, 3, and 4 (Table 3-1). The fields were visited approximately every two weeks in order to determine these dates.

The CROPGRO-Cotton Model

The CROPGRO-Cotton model is a member of the CROPGRO group of models in DSSAT that has been tested for several crops including soybean (Boote et al., 1998) and peanut (Boote et al., 1998; Gilbert et al., 2002). It simulates the effects of weather, soil, and, management on crop growth and development (Boote et al., 1998; Jones et al., 1998; Jones et al., 2003). The CROPGRO-Cotton model, using the same features and level of details as other the CROPGRO crop models, was developed recently. Only a few studies have been reported on this cotton model evaluation and applications (Zamora et al., 2009, Messina et al., 2004, Guerra et al., 2005).

Soil inputs consist of one dimensional soil physical properties such as lower, upper, and saturated water holding capacities, bulk density, and PH (Jones et al., 2003; Jones et al., 1998). The soil module integrates information from soil temperature, soil water, soil carbon, and nitrogen dynamics sub modules (Jones et al., 2003) to simulate growth and yield. Weather data consisted of daily values of minimum and maximum air temperatures, solar radiation, and precipitation. Management inputs consisted of information on amount of irrigation, fertilization, planting dates, plant population etc. The model was provided with information on soil, weather, and management specific to each experiment for model simulations (Table 3-1).

The CROPGRO model has three sets of parameters that account for differences in development, growth, and yield between species, ecotypes, and cultivars (Boote et al., 2003). In this study, cotton cultivar parameters were estimated along with the

uncertainties associated with them. A detailed description of cultivar parameters and uncertainty analysis procedures are given in the following sections.

Uncertainty Analysis and Parameter Estimation Procedures

The GLUE procedure was used in this study where 55,000 parameter sets were sampled from their prior distributions using Monte-Carlo simulations, and model outputs were obtained for each of those parameter sets. The prior distribution represents the original information about the uncertainties of parameters based on previous studies. The primary reason for running such a large number of simulations in GLUE is to obtain an adequate number of acceptable parameters for estimating the posterior distribution of parameters. Each of the model outputs was assigned likelihoods based on their agreement with related field observations. Using the likelihood values, posterior distributions were estimated using the Bayesian approach. The GLUE procedure was first performed on phenology parameters (EM-FL and SD-PM). Once the estimated parameters were obtained for these phenology parameters, the GLUE procedure was performed on the remaining cultivar parameters. This parameter estimation and uncertainty analysis procedure was based on recommendations in Boote et al. (1998) who suggested to first estimate phenology parameters before estimating other parameters. Descriptions of input parameters, uncertainty ranges, prior distributions, likelihood functions, and posterior distributions used in this research are given in the following sections.

Parameter selection and prior distributions

A complex simulation model such as the CROPGRO-Cotton model is generally heavily parameterized. Thorsen et al. (2001) suggested that the inclusion of a large number of parameter sets would result in an unrealistic number of simulation runs that

would be too large to compute. A systematic way of selecting parameters for uncertainty analysis is to perform sensitivity analysis on parameters to get information on which parameters are important. Pathak et al. (2007) performed a global sensitivity analysis of the CROPGRO-Cotton model to cultivar and species parameters and suggested a list of parameters (EM-FL, FL-SD, SD-PM, SLAVAR, KCAN, SFRT, SFDUR, and PODUR) to which important model outputs such as cotton yield and physiological maturity are sensitive.

The CROPGRO-Cotton model was initially parameterized using data from the literature (Messina et al., 2004; Zamora et al., 2009). Very little knowledge was available on real ranges and distributions of these parameters. For this study, a uniform prior distribution was assumed for each of the selected parameters as it has been the most reported sampling distribution in similar studies reported in the literature (Beven, 2001; Stedinger et al., 2008). Because of a lack of information, it was also assumed that the parameters are independent. The only information needed to sample parameters from uniform distribution was their minimum and maximum values. The ranges of model parameters should be chosen such that they represent expected extreme values of those parameters. The ranges of model parameters (SLAVAR, KCAN, and XFRT) were obtained from the literature (Bange and Milroy, 2000; Milroy et al., 2001; Milroy and Bange, 2003; Reddy et al., 1992; Messina et al., 2004). Information on ranges of certain parameters (LFMAX, EM-FL, FL-SD, and SD-PM) were not available, so in order to make maximum use of available dataset, the percentage variance from the mean values for the CROPGRO-Soybean (Boote et al., 2003) parameters were applied to the

base cotton model parameters. Table 3-2 shows selected parameters and their uncertainty ranges used in the GLUE analysis.

Likelihood function

The likelihood function is a measure of how well a model outputs obtained from a set of parameters fit to field observations. The calculation of the likelihood function is an important part of the GLUE procedure. There have been several likelihood functions used in previous applications of GLUE (Beven and Freer, 2001). The choice of likelihood function should be such that it can reduce the uncertainties in parameters and simulated outputs should provide close agreements with observed data. He et al. (2009) compared four likelihood functions for parameter estimation of the CERES-Maize model using the GLUE method. He found that the Gaussian likelihood function (Makowski et al., 2006) was the best choice because the results from this function resulted in the least uncertainties in parameters and improved model predictions more than the other functions that were evaluated. Based on his findings, the Gaussian likelihood function was used in this study, as is shown below.

$$L(\theta_i / O) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\Pi\sigma^2}} \exp\left(\frac{(\overline{O} - P_i(\theta))^2}{2\sigma^2} \right) \tag{3-1}$$

Where, $L(\theta_i / O)$ is likelihood of parameters, given the observations (O). $\overline{O}$ is a mean of replications of observation, $P_i(\theta)$ is ith model response, and N is the number of observations. With complex simulation models such as the CROPGRO-Cotton model, it is difficult to know the model error variance that is needed for the Gaussian function. Alternatively, variances in measurements were assumed to be equal to the model error variance in previous studies (Van Oijen et al., 2005; He et al., 2009). Another alternative

that Wang et al. (2005) used was to assume that the error variance was equal to the error variance between observed and simulated values using the very best parameter set. For this study, error variance σ^2 was replaced with the measurement variance σ_o^2 for LAI, leaf weight, stem weight, and boll weight. Each data point in the time-series of LAI, leaf weight, stem weight, and boll weights contained measurement variances specific to that particular data point. The measurement variance was unknown for anthesis date and maturity date; hence the error variance in the above equation for anthesis and maturity date was replaced with the minimum error variance between measured and simulated results (Wang et al., (2005). The likelihoods were calculated for each of the data points separately and were integrated together by taking the product.

It can be seen from equation 3-1 that the likelihood values of N number of observations were combined by taking a product to get a global likelihood. The main benefit of using product of likelihoods is that as the number of observations increases, the global likelihood response becomes steeper. He et al. (2009) compared three likelihood combination methods including the product and found that the Gaussian likelihood function updated with product function provided the best combination for CERES-Maize model.

Posterior distribution

The posterior distribution was derived from the behavioral simulations, that is, the number of parameter sets not ruled out of the analysis by near zero likelihood. The threshold to derive behavioral and non behavioral parameter sets is subjective in the GLUE methodology (Beven and Binley, 1992). The common practice is to select a cutoff

for likelihood weight to identify behavioral parameters for use in determining posterior distributions. For this study, a likelihood threshold of 0.0001 was selected. Parameter sets having likelihood values above that cutoff limit were considered behavioral parameters to construct the posterior distribution. Once parameter sets are distinguished between behavioral and non behavioral by using likelihood values, likelihood weights were calculated by normalizing behavioral likelihood values as shown in equation 3-2.

$$L_w(\theta_i) = \frac{L(\theta_i / O)}{\sum_{i=1}^{N} L(\theta_i / O)} \tag{3-2}$$

$Lw(\theta_i)$ is the normalized likelihood weight where the sum of $Lw(\theta i)$ is equal to 1.0. The $L(\theta_i / O)$ is the global likelihood value obtained by taking the product of all the likelihood values. The expected values and variances for each parameter were calculated using equations 3-3 and 3-4.

$$\hat{\mu} = \sum_{i=1}^{N} L_w(\theta_i).\theta_i \tag{3-3}$$

$$\hat{\sigma} = \sum_{i=1}^{N} L_w(\theta_i).(\theta_i - \hat{\mu})^2 \tag{3-4}$$

Uncertainty bounds for model predictions

Uncertainty bounds represent the uncertainties in model predictions that were associated with uncertainties in model parameters. A total 5000 sets of input parameters were sampled from their posterior distributions using Monte Carlo sampling and the CROPGRO-Cotton model was run to obtain outputs. The uncertainties in model outputs were estimated from their empirical cumulative distributions. An appropriate

quantiles of cumulative distribution were selected to form uncertainty bounds around the model outputs. In this analysis 95% confidence interval (uncertainty bound) was estimated from the values of 2.5% and 97.5% quantiles of the cumulative distribution of model outputs. The primary reasons for obtaining 5000 simulations runs was to obtain good estimates or prediction distributions and of prediction quantiles.

Statistical methods for model testing

The simulated values of LAI, leaf weight, stem weight, and boll weight were analyzed using the following statistical measures:

$$MeanDeviation = \frac{1}{n}\sum_{i=1}^{n}(Si - Oi) \tag{3-5}$$

$$RootMeanSquaredError(RMSE) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Si - Oi)^{2}} \tag{3-6}$$

$$d = 1 - \left[\frac{\sum_{i=1}^{n}(S_i - O_i)^2}{\sum_{i=1}^{n}\left(\left|Si - \overline{O}\right| + \left|O_i - \overline{O}\right|\right)^2}\right] \tag{3-7}$$

Where, n is the number of observations, S_i and O_i are the ith simulated value with mean parameters and observed values, respectively. Mean deviation indicates bias in the simulations, positive value of mean deviation corresponds to over prediction by the model and vice versa. The RMSE corresponds to the magnitude of the mean difference between predicted and measured values. The d-index (Willmott, 1982) corresponds to the agreement between model simulations and observations, 0 representing no agreement and 1 representing perfect agreement.

Results and Discussion

Simulations Using Unmodified Parameters

Simulations using unmodified cultivar parameters showed good agreement with observed LAI, leaf weight, stem weight, and boll weight for all four experiments. The unmodified parameters currently available in DSSAT were initially estimated using the experimental dataset (Messina et al., 2004). Although the overall model agreement with the measurements was fairly good, it was observed that the model over-predicted LAI and above ground biomass when simulated using original model parameters. Figure 3-1 shows agreement between measured and simulated values with original model parameters for LAI, leaf weight, stem weight, and boll weight for all four experiments. Except for a few data points, the model consistently over predicted the values as the majority of the data points fell below the 1:1 line.

Tables 3-3 summarizes the average observations, simulations, and mean differences for LAI, leaf weight, stem weight, and boll weight for all four experiments. The positive values of mean differences indicated that the model was over-predicting and vice versa. For instance, the mean differences between simulated and measured average boll weights were 1454 kg ha^{-1} and 1536 kg ha^{-1} for experiments 1 and 3, respectively (Table 3-3). In general average observed stem weight was over predicted by the model for all four experimental sites, average observed average LAI and leaf weight were over predicted in three out of four experiments, and average observed boll weight was over predicted in two out of four experiments.

Figure 3-2 shows the time series of model simulations and observed values for LAI, leaf, stem, and boll weight at different growth stages using original model parameters. During the early stages of crop growth the model was either under

predicting the above ground biomass and LAI or had good agreement with the measured values. However, during the reproductive stage of crop growth, the model over predicted them through maturity. For example, if we observe the LAI, leaf weight, and stem weight time series plots for experiments 2, 3, and 4, the model under predicting the first two or three measurements but then over predicted throughout the remaining growth cycle.

In general, time series plots and statistics show that the model simulations tended to overestimate observed LAI, leaf weight, stem weight, and boll weight for all four experiments. Hence estimating parameters regulating the crop growth should improve the overall model predictions.

Parameter Estimates and Uncertainties

Comparison with prior distribution

A total of eight cultivar parameters of the CROPGRO-Cotton model associated with plant growth and development were estimated along with their uncertainties. Table 3-4 shows means, standard deviations, and coefficient of variations (CV) for prior and posterior distributions of parameters. The prior means of model parameters shown in Table 3-4 were based on the 55,000 randomly sampled parameter sets from their corresponding uniform prior distributions.

The posterior distributions for all eight parameters were narrowed to smaller ranges compared to their prior distributions (Fig. 3-3, Table 3-4). These narrower ranges are, of course, dependent on the data from the four treatments used in this analysis. For example, the SLAVAR had prior uncertainty range of 90-250 $cm^2\ g^{-1}$ which was subsequently reduced to 171-175 $cm^2\ g^{-1}$ in its posterior distribution and the SLAVAR coefficient of variation (CV) was changed from 27% to only 0.8%, respectively.

The KCAN parameter had a prior uncertainty range of 0.5-1, which was reduced to 0.61-0.67 and the CV was reduced from 17% to 3%. Those two parameters were the most important parameters to simulated cotton yield according to the sensitivity analysis results obtained by Pathak et al. (2007). Uncertainty in those parameters would significantly affect the uncertainty in model outputs, hence predicting and reducing the amount of uncertainty in these parameters improved the model performance. The highest CV value among all eight parameters in the Pathak et al (2007) study was for SLAVAR which lowered down to 0.8% in the posterior distribution compared to 27% in the prior distribution. One of the main reasons for narrower uncertainties in posterior distributions was the integration of the large number of observations from the four different sites into the GLUE to estimate uncertainties. In the Gaussian likelihood function (eq. 3-1), as the number of observations N increases the parameter space also may tend to decrease and thus result in a narrower range. That was the primary reason there were only 81 behavioral parameter sets that were concentrated towards the narrower uncertainty range in this study.

Comparison with DSSAT default values

An interesting thing to notice was that the expected values of all the parameters were not so different from their DSSAT default values except for KCAN (Table 3-4). The estimated value for KCAN was 0.64 which was lower than its DSSAT base value of 0.8. The KCAN is an important parameter as it is responsible for light capture by the canopy. Hence changing the value of KCAN would have a direct impact on daily photosynthesis, which subsequently impacts LAI and above ground biomass. As expected, the estimated value for KCAN was lower than the DSSAT default value because it would control the model over-prediction of LAI and above ground biomass.

Model predictions based on estimated parameters

The comparison between Figure 3-1 and 3-4 as well as Table 3-3 and 3-5 showed that the estimated parameters improved the overall agreement of measured LAI and above ground biomass with simulated results. Other than KCAN, no estimated parameters were very different from their corresponding DSSAT base values. As shown in the initial model results with unmodified parameters, the model over predicted the observations in general. Since estimated KCAN was lower than the corresponding DSSAT base values, the over estimation by the model was controlled significantly and model predictions were improved.

Table 3-6 shows the comparison of RMSE and d-statistics between measured and simulated LAI, leaf weight, stem weight, and boll weight obtained from DSSAT parameter base values and maximum likelihood estimated parameter values, respectively. For instance, RMSE of boll weight for default values were reduced from 1636 kg/ha to 535 kg/ha and 1842 kg/ha to 819 kg/ha for experiments 1 and 3, respectively. This represents a reduction of RMSE of approximately 55% and 67%, respectively. In this study, the d-statistics were improved for most of the variables across all the experiments. The highest d-statistic value of 0.99 was obtained for stem weight for experiments 2 and 3, which represents almost perfect agreement of measured with simulated stem weights. The lowest d-statistic of 0.45 was obtained for boll weight in experiment 4. The reason for a low agreement could be due to measurement errors because the agreement of the model with other variables (LAI, leaf weight, and stem weight) of experiment 4 was in good agreements with measured values. The RMSE and d-statistics showed a close agreement of the measured and simulated LAI and biomass components overall. The average RMSE of simulated LAI,

stem weight, and boll weight from all four experiments were reduced with the GLUE estimated parameters compared to the DSSAT default parameters. The average RMSE for leaf weight was increased; however the increase was marginal (Table 3-6). It is also important to note that the estimated parameters generally improved the model by reducing model error.

Model output uncertainties

Output uncertainties for LAI, leaf weight, boll weight, and stem weight were estimated from 95% confidence interval obtained from 2.5% and 97.5% quantiles of their cumulative distributions from 5000 simulations generated from posterior distributions of model parameters. The output uncertainties in terms of standard deviations (STDEV) and coefficient of variability (CV) were compared in Table 3-7 along with the averages and ranges. The CV for model outputs obtained with prior parameter distribution ranged between 29% and 56% which were subsequently reduced to 4-13% with the simulated outputs obtained from the posterior distribution of parameters.

Figure 3-5 shows the simulated 95% confidence limits around the average simulated values in addition to the measured values of LAI, leaf weight, stem weight, and boll weight for all four sites. Results show that the 80% of the means of data from experiment 1, 90% from experiment 2, 87% from experiment 3, and 53% of data from experiment 4 were covered by the uncertainty bounds. Overall, approximately 79% of the data were within the uncertainty bounds. Reasons for the remaining 21% of data that were not within the uncertainty limits could be measurement errors also shown in Fig 3-5 or the subjective selection of cut-off of likelihood values.

It is important to note that the output uncertainties shown in Table 3-6 only consider uncertainties associated with eight model parameters of cotton growth and

development. The parameters and their uncertainties obtained in this analysis considered four treatments. Results could be different if other datasets were used. This study did not take into account the other sources of uncertainties. The uncertainties in model structure and model input data were not within the scope of this research.

Conclusion

The mean values of parameters after estimation improved model predictions relative to the original parameters available in DSSAT for delta pine 555 cotton cultivar. The prior uncertainties in the most important parameters (Pathak et al., 2007), SLAVAR and KCAN, were reduced from 27% and 13% to 0.8% and 3%, respectively. The general agreement of the model with measurements using the GLUE estimated parameters was good with the d-statistics ranged between 0.99 for experiments 2 and 3 and 0.45 for experiment 4. This study also demonstrated an efficient prediction of uncertainties in model parameters and outputs using the widely accepted GLUE technique. Overall, approximately 79% of the data were within the uncertainty bounds that were determined from the predicted confidence interval from the uncertainty analysis.

The limitation of this study was that the uncertainties in model structure, and model inputs were not examined. In further research, these uncertainties should be investigated. Also, parameter selection was concentrated only on cultivar parameters, and one species parameter. Uncertainty analysis should be performed on other important parameters such as soil parameters.

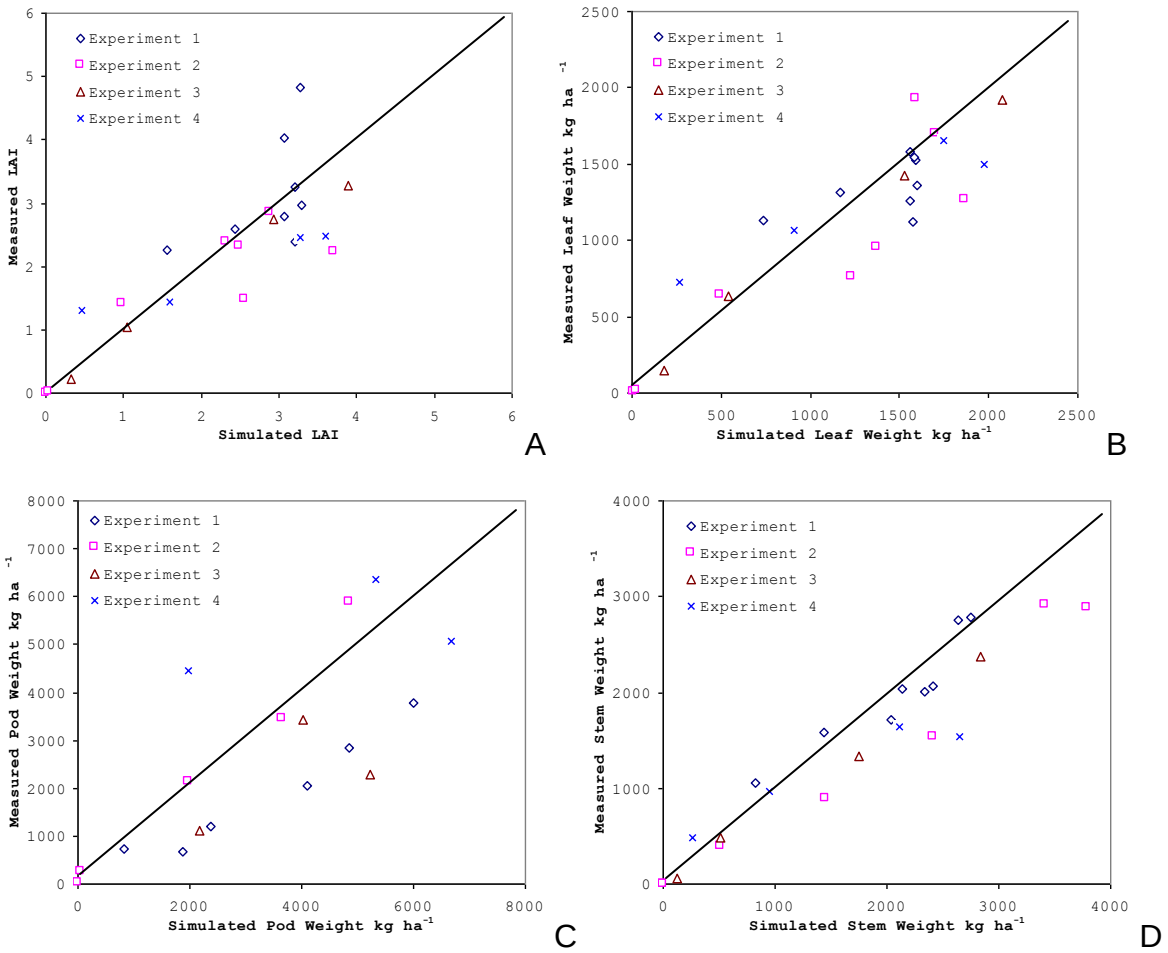


Figure 3-1. Scatter plot of simulated vs. measured values for A) leaf area index, B) leaf weight, C) boll weight, D) stem weight for all four experiments. The simulated results were obtained using unmodified cultivar parameters.

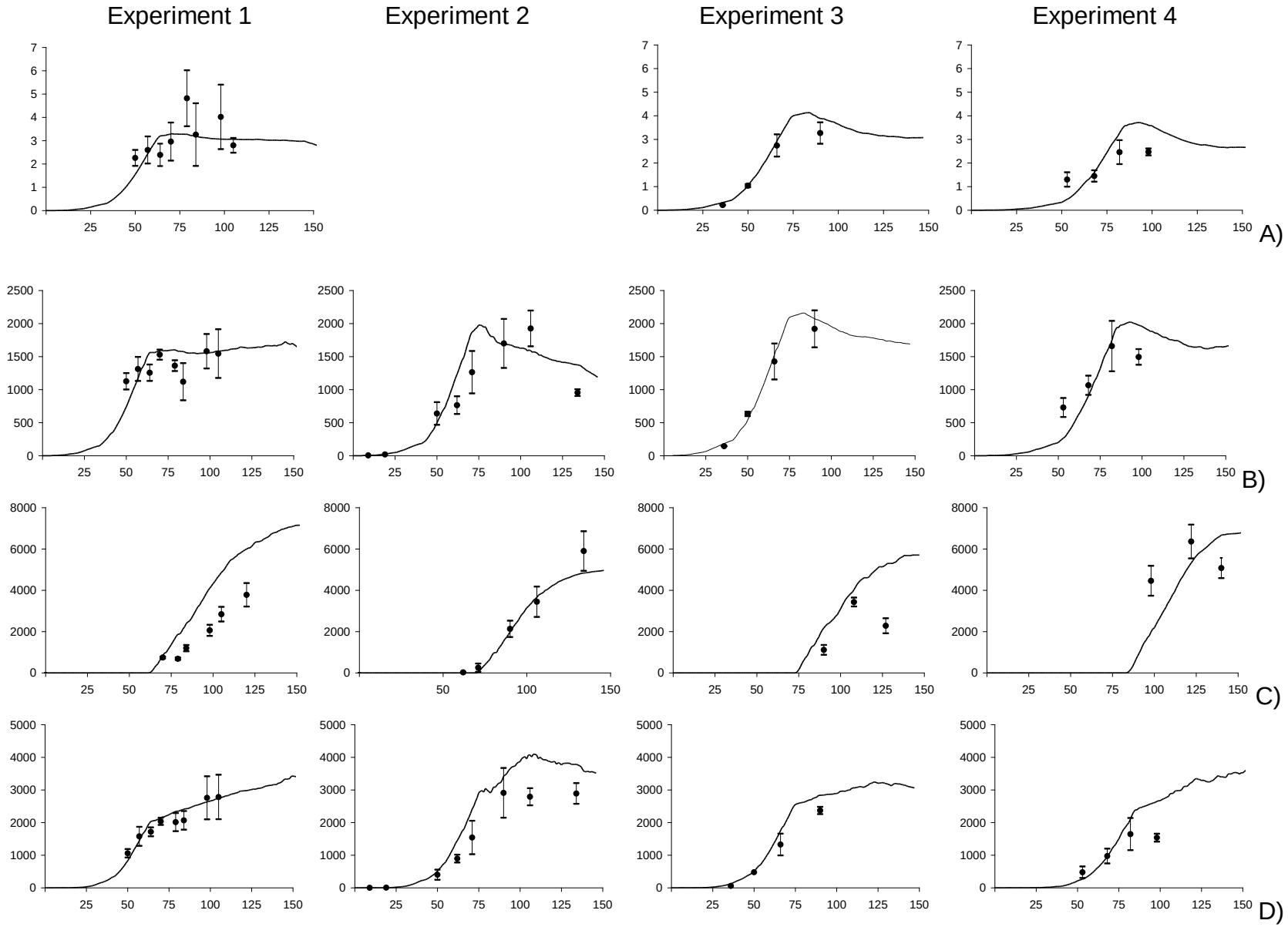


Figure 3-2. Simulated and observed values for A) leaf area index, B) leaf weight (kg/ha), C) boll weight (kg/ha), and D) stem weight (kg/ha) for experiments 1-4 using unmodified cultivar parameter values.

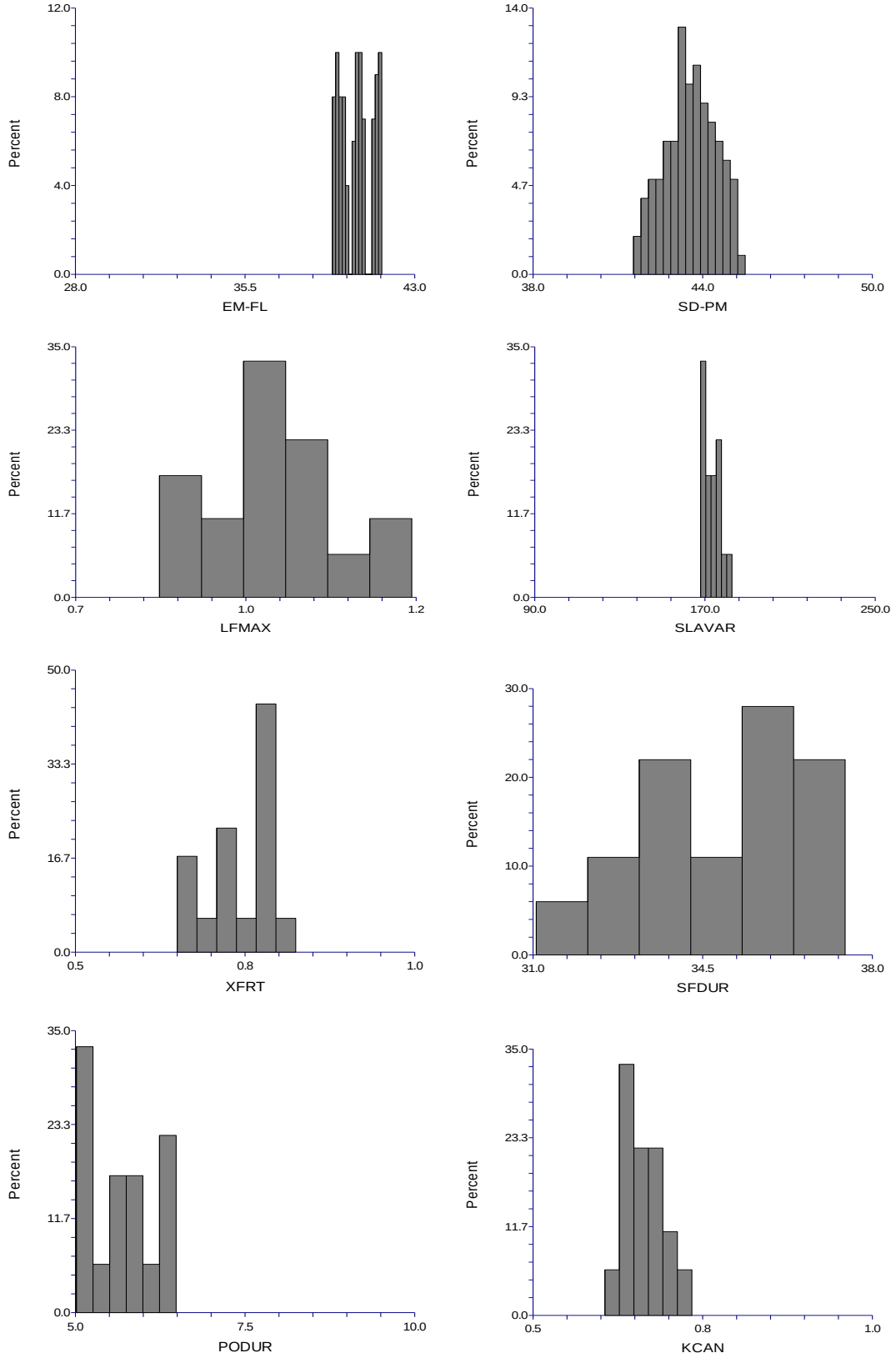


Figure 3-3. Posterior probability distributions of model parameters

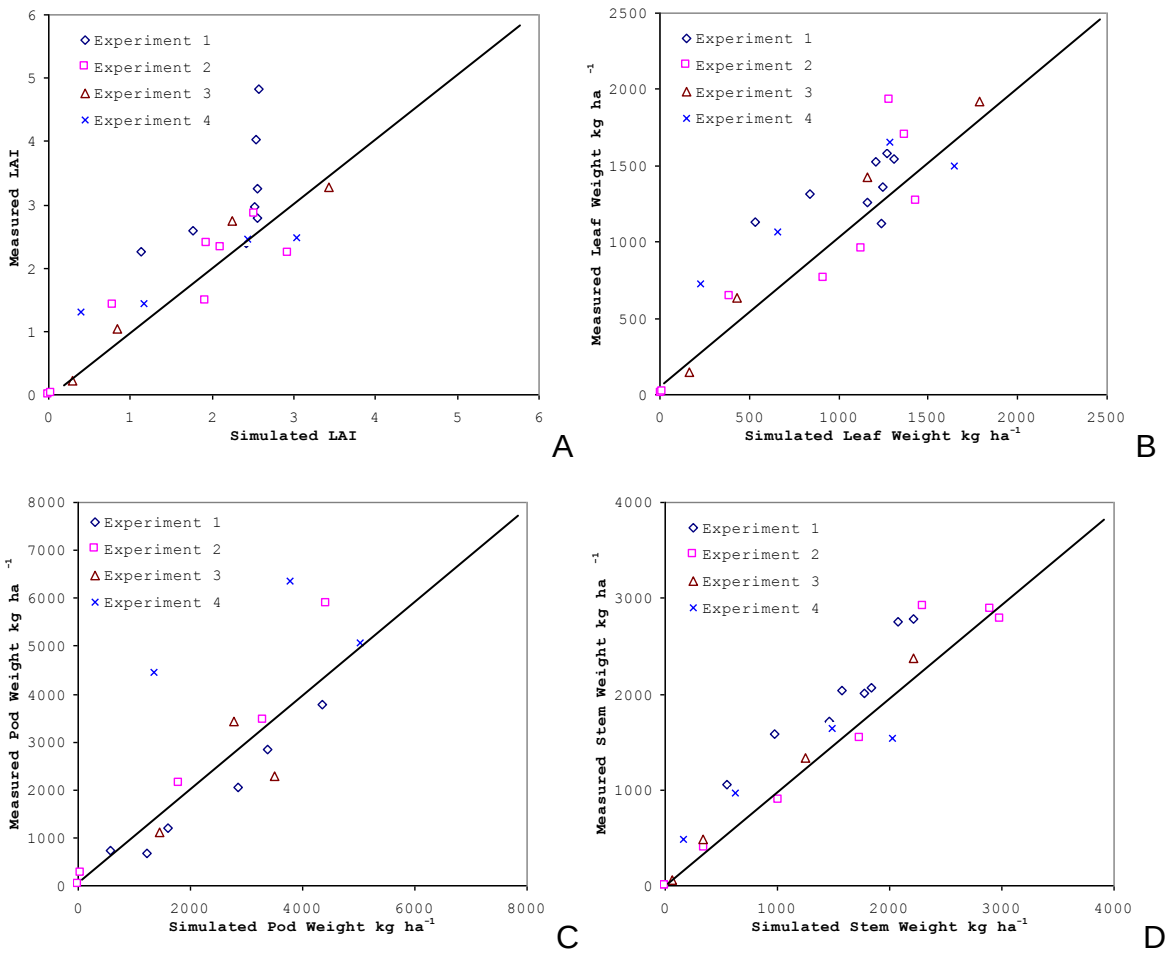


Figure 3-4. Scatter plot of simulated vs. measured values for A) leaf area index, B) leaf weight, C) boll weight, D) stem weight for all four experiments. The simulated results were obtained using estimated cultivar parameters.

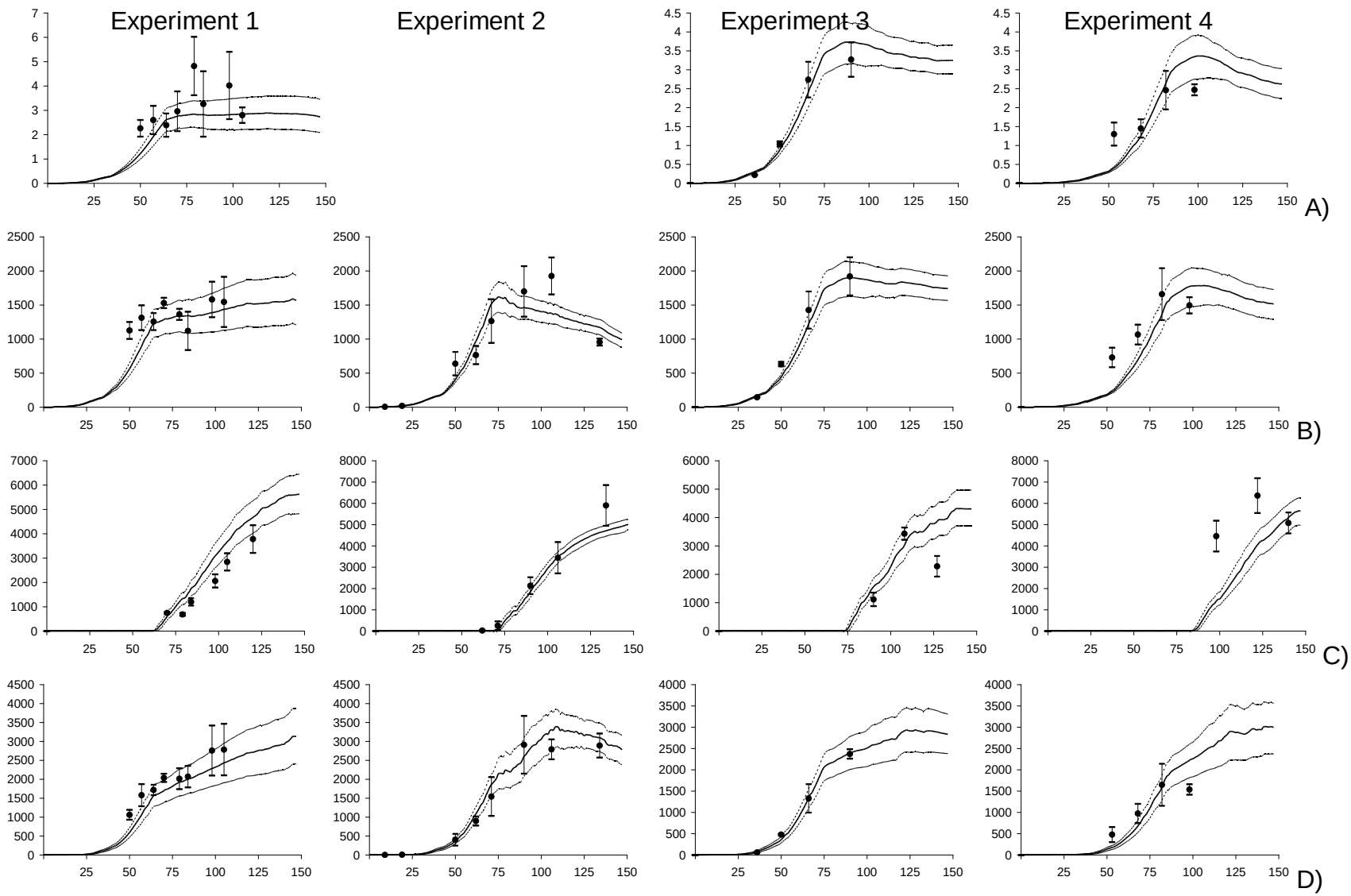


Figure 3-5. Simulated and observed values for A) leaf area index, B) leaf weight, C) boll weight, and D) stem weight for four experiments using estimated cultivar parameter values. Vertical bars indicate the standard deviations of observed values. Dotted lines represent 2.5% and 97.5% confidence interval of 5000 simulations.

Table 3-1. Information about the experimental sites, planting date, type of soils, and weather characteristics

Experiment	1	2	3	4
Sites	Citra, FL	Camilla, GA	Quincy, FL	Quincy, FL
Latitude	29°24' N	31°11' N	30°36' N	30°36' N
Longitude	82°17' W	84°12'W	84°33' W	84°33' W
Elevation (m)	20	54	63	63
Planting date	19-Jun-06	12-Apr-04	06-Jun-06	05-May-06
Soil type	Troup Sand	Troup sand	Dothan sandy loam	Dothan sandy loam
Mean seasonal temperature (^{o}C)	27	25	22	22
Mean seasonal precipitation (mm)	2.4	3.7	3.5	3.5
Reference of weather data	www.fawn.com	www.georgiaweather.net	www.fawn.com	www.fawn.com

Table 3-2. The CROPGRO-Cotton parameters and uncertainty ranges used for GLUE prior distributions.

Parameter	Definition	Uniform Distribution		
		Minimum	Base	Maximum
EM-FL	Duration between emergence and flowering	28.00	40.00	43.00
SD-PM	Duration between first seed to physiological maturity	38.00	45.00	50.00
LFMAX	Maximum leaf photosynthesis rate	0.70	1.10	1.20
SLAVR	Specific leaf area of cultivar under normal growth condition	90.00	170.00	250.00
XFRT	Maximum fraction of daily growth that is partitioned to seed+shell	0.50	0.80	0.95
SFDUR	Seed filling duration for boll cohort at standard growth condition	31.00	35.00	38.00
PODUR	Time required for cultivar to reach final boll load under optimal condition	5.00	5.00	10.00
KCAN	Light Extinction Coefficient	0.50	0.80	0.95

Table 3-3. The CROPGRO-Cotton average model predictions using DSSAT default parameters in comparison with corresponding measured data

Variable Name	Experiment	Mean Observed	Mean Simulated	Mean Diff
First flower	1	57	51	-6
Maturity	1	142	146	4
LAI	1	3.14	2.88	-0.25
Leaf wt kg/ha	1	1354	1422	68
Stem wt kg/ha	1	2002	2073	71
Boll wt kg/ha	1	1886	3339	1454
Leaf wt kg/ha	2	909	1036	127
Stem wt kg/ha	2	1430	1955	525
Boll wt kg/ha	2	2352	2103	-249
First flower	3	65	61	-4
LAI	3	1.82	2.04	0.23
Leaf wt kg/ha	3	1031	1081	50
Stem wt kg/ha	3	1059	1305	246
Boll wt kg/ha	3	2276	3812	1536
First flower	4	64	72	8
LAI	4	1.92	2.23	0.31
Leaf wt kg/ha	4	1238	1224	-13
Stem wt kg/ha	4	1159	1495	336
Boll wt kg/ha	4	5300	4657	-644

Table 3-4. Parameter uncertainties and fundamental statistics of prior and posterior distributions

Parameter	DSSAT default	Prior			Posterior		
		Mean	STDEV	CV	Mean	STDEV	CV
EM-FL	40	35.44	4.34	12.30%	40	0.61	1.50%
SD-PM	45	44.07	3.53	8.00%	44	2.06	4.70%
LFMAX	1.10	0.95	0.14	15.20%	1.05	0.11	5.40%
SLAVAR	170	170.00	45.91	27.10%	173	7.5	0.80%
XFRT	0.80	0.73	0.13	17.90%	0.77	0.02	2.90%
SFDUR	35	35.00	2.01	5.80%	36	0.83	2.40%
PODUR	5	7.50	1.44	19.10%	5.2	0.4	1.80%
KCAN	0.80	0.73	0.13	17.80%	0.64	0.05	3.20%

Table 3-5. The CROPGRO-Cotton average model predictions using estimated parameters in comparison with corresponding measured data

Variable Name	Experiment	Mean Observed	Mean Simulated	Mean Diff
First flower	1	57	52	-5
Maturity	1	142	142	0
LAI	1	3.14	2.26	-0.88
Leaf wt kg/ha	1	1354	1101	-253
Stem wt kg/ha	1	2002	1559	-444
Boll wt kg/ha	1	1886	2330	444
Leaf wt kg/ha	2	909	820	-90
Stem wt kg/ha	2	1430	1412	-18
Boll wt kg/ha	2	2352	1916	-436
First flower	3	65	62	-3
LAI	3	1.82	1.70	-0.12
Leaf wt kg/ha	3	1031	884	-146
Stem wt kg/ha	3	1059	964	-95
Boll wt kg/ha	3	2276	2572	296
First flower	4	64	72	-8
LAI	4	1.92	1.76	-0.16
Leaf wt kg/ha	4	1238	954	-284
Stem wt kg/ha	4	1159	1077	-82
Boll wt kg/ha	4	5300	3379	-1921

Table 3-6. Comparison of RMSE, and d-statistics of simulated LAI, and biomass components for four sites based on DSSAT default model parameters and GLUE estimated parameters

Experiments	Variables	DSSAT default parameters		GLUE estimated parameters	
		RMSE	d-statistics	RMSE	d-statistics
1		0.76	0.63	1.11	0.57
3		0.33	0.98	0.28	0.98
4	LAI	0.81	0.80	0.55	0.87
1		258.27	0.67	330.24	0.53
2	Leaf	333.84	0.93	287.47	0.94
3	Weight	105.69	0.99	180.87	0.98
4	Kg/ha	346.36	0.88	381.01	0.83
1		1630.34	0.75	535.15	0.95
2	Boll	497.71	0.98	681.73	0.97
3	Weight	1842.07	0.45	819.64	0.75
4	Kg/ha	1803.42	0.53	2340.07	0.45
1		229.93	0.95	474.14	0.82
2	Stem	684.69	0.94	240.62	0.99
3	Weight	316.40	0.97	116.08	0.99
4	Kg/ha	611.40	0.81	349.32	0.91

Table 3-7 Comparison of output uncertainties in model outputs of LAI and above ground biomass components for prior and posterior distribution

Experiments	Variables	Prior				Posterior			
		STDEV (kg/ha)		CV (%)		STDEV (kg/ha)		CV (%)	
		Mean	Range	Mean	Range	Mean	Range	Mean	Range
1		1.38	1.03-1.49	56	56-58	0.26	0.12-0.34	10.7	10.1-12.0
3		0.8	0.1-1.6	47	40-52	0.14	0.01-0.28	6.97	4.29-8.75
4	LAI	0.92	0.14-1.55	51	41-56	0.18	0.02-0.31	8.79	5.64-10.0
1		497	360-568	44	43-45	119.13	53.76-163.2	9.98	9.33-11.35
2		246	0.6-481	29	10-38	62.55	2.13-114.86	7.05	5.23-11.10
3	Leaf Weight kg/ha	302	45-594	35	32-39	67.32	6.18-136.85	6.34	3.74-8.07
4		366	64-611	39	31-43	89.26	11.57-154.08	8.09	5.04-9.36
1		613	18-1268	45	40-61	217.26	89.47-357.21	9.71	7.64-14.92
2		562	14-917	34	20-60	87.69	0-138.13	4.58	2.74-7.01
3	Boll Weight kg/ha	892	504-1226	36	36-37	236.75	141.68-302.59	8.57	7.83-9.07
4		1153	485-1653	38	36-42	263.43	152.24-334.50	8.19	6.27-10.82
1		760	478-986	46	45-50	183.94	73.77-279.83	10.98	10.25-12.45
2		604	2-962	43	29-62	165.83	0.39-292.98	10.14	6.78-14.19
3	Stem Weight kg/ha	411	57-871	53	40-75	101.05	15.78-211.07	13.02	8.80-20.88
4		501	104-896	52	44-63	125.44	27.42-220.36	12.13	9.91-15.53

CHAPTER 4
IN-SEASON UPDATES OF COTTON YIELD FORECASTS USING CROPGRO-COTTON MODEL

Introduction

There has been significant progress in developing models to simulate growth and development of agricultural crops (Ritchie, 1994). These crop simulation models have shown potential for use in forecasting crop yield using weather forecasts (De Wit and Diepen, 2007; Hansen et al., 2005; Larow et al., 2005) as well as using historical weather records (Lembke and Jones, 1972, Wright et al., 1984). Simulation models are usually deterministic; hence if inputs to the model and model parameters are accurately measured, yield forecasts can be made even in abnormal climate situations, if those can be predicted (Bannayan and Crout, 1999).

To facilitate simulation model forecasts of crop yield, daily weather data need to be provided to the model in terms of a forecast. It is a common practice to generate weather ensembles stochastically (Lawless and Semenove 2005; Bannayan and Crout 1999; Ahmed et al., 1976; Grondona et al., 2000; Podesta et al., 2002) to represent uncertainty of weather conditions that are provided as an input to the model. However, the main limitations of weather generators are that they do not simulate extreme values well and they assume that the observed relationships between weather variables will remain the same in the future (Jones et al., 2009). As an alternative to using generated weather series, historical weather records could be used to represent the possible weather for the next season. There has been handful of studies made where historical weather data were used to simulate crop yield, which was represented as frequency distribution of expected yields (Lembke and Jones, 1972, Wight et al., 1984).

Studies have shown that instead of using a static range of weather to forecast yields, updating the model in-season using observed weather improves accuracy of model forecasts (Bannayan and Crout 1999; Lawless and Semenov 2006; Semenov and Porter 1995). Before the start of crop growing season, weather is entirely uncertain which translates into uncertainties in crop yield forecasts. If the model is updated in-season with observed weather data, some of the weather uncertainties are eliminated. Since weather uncertainties early in the season are eliminated via use of observed data, it is possible to improve accuracy of crop yield prediction by in-season updating the model with real weather data. For example, Bannayan and Crout, (1999) clearly showed that the in-season updating the SUCROS model improved the forecasting accuracy of winter wheat yield.

Crop yield in southeastern United States is affected significantly by El Niño Southern Oscillation (ENSO) (Jones et al., 2000; Hansen et al., 1998). In this region, El Niño events are characterized by lower winter temperature and higher rainfall and La Niña events have the opposite effects. Since ENSO influences agricultural yield in the southeastern United States, it may be possible to forecast crop yield tailored to different ENSO phases using in-season updates of real weather data. Royce et al. (2009) compared the predictive potential of three different ENSO classifications on crop yields for the Southeastern United States, including cotton. The three ENSO classifications they used were based on Japan Meteorological Agency (JMA) index, Oceanic Niño Index (ONI), and Multivariate ENSO index (MEI). According to that study, March-May MEI ENSO index and January-April ONI ENSO index showed the strongest association with cotton yield in the southeastern United States. The assumption was that the ENSO

conditions occurring immediately prior to or during the spring-summer crop season climate influence in the southeastern United States that affects crop yields which would not be captured by categorical annual JMA ENSO index. They found that monthly ENSO indices were better predictors of crop yield than the JMA. That research was based on the historical county yield data which does not distinguish between irrigated and rain fed practices. Simulation models can be effective for more robust analysis since they can quantify the effects under rainfed conditions alone.

The CROPGRO-Cotton model (Messina et al., 2004; Pathak et al., 2007) is a complex simulation model that has been recently added to Decision Support System for Agrotechnology Transfer (DSSAT) (Jones et al., 1998; Jones et al., 2003) group of models. It has been calibrated using field conditions in the southeastern United States but has not been utilized to obtain in-season updates of cotton yield forecasts. Specific research questions addressed in this study were:

1). Do in-season updates of cotton yield forecasts improve accuracy over the forecast obtained before season?

2). Which ENSO index (ONI, MEI, or JMA) provides the best cotton yield forecasting accuracy?

3). Do in-season updates on cotton yield forecasts tailored to ENSO have better potential in forecasting cotton yield than the cotton yield forecasts obtained using climatology alone?

The objectives of this study were 1) to evaluate in-season updates of cotton yield forecasts, 2) to evaluate the use of different ENSO indices in forecasting cotton yield and 3) to compare ENSO-based forecasts with those based on climatology.

Material and Methods

Outline of the Forecasting Method

The CROPGRO-Cotton model requires daily weather data to simulate growth and development. In order to utilize the model to forecast cotton yield, weather data inputs must be provided in terms of a forecast. As a first step, historical weather data up to the point when the cotton yield forecast is made were replaced with the real weather data for that season. The process of updating the model with real weather data was repeated two times during the season (July 1, and August 1) during 1951-2005. Cotton yield forecasts obtained before season, and in-season using past weather data as a forecast (climatology) were compared against their simulated cotton yields using real weather data to evaluate an accuracy of the forecasts, to obtain cotton forecasts tailored to ENSO phases as a next step, historical weather data for the current ENSO phase were assumed to be a forecast of future weather. Cotton yield forecasts tailored to ENSO were compared with the cotton yield forecasts obtained using climatology alone. This study was focused on Quincy, Florida, as a case study of the Southeastern United States.

Model Description and Input Data

Cotton was simulated using the CROPGRO-Cotton model under rain fed conditions with no nitrogen stress. The planting date for all the simulations was kept on May 1 to represent a typical planting date in the southeastern United States (Pettigrew, 2002). The model has been calibrated for Quincy, Florida location (Chapter 2) for the Delta Pine 555 cultivar. This cultivar parameter set was used to simulate cotton yield for this location. Site specific details including soil type and cultivar are described in detail in Chapter 2.

Comparison between Before-Season and In-Season Cotton Yield Forecasts Based on Climatology

Inputs of historical weather data for 1951-2005 were provided to the model to simulate cotton yield using May 1 as a planting date for all years. The mean and standard deviations of simulated cotton yields represent “before-season” cotton yield forecast and associated uncertainties obtained using climatology as a forecast.

In the next step, historical weather data up to July 1 were replaced with observed weather data up to that point for all the years. For example, in order to forecast cotton yield for the year 1951, all the historical weather records from 1952-2005 were replaced with 1951 observed weather up to July 1. Those updated weather data were then provided to the model to simulate updated cotton yields on July 1. The mean and standard deviations of those simulated cotton yields represents “in-season” updated cotton yield forecasts and associated uncertainties using climatology as a forecast. Another in-season update of cotton yield forecasts was obtained on August 1 for all the years.

The residual errors of “before-season” and “in-season” updated cotton yield forecasts were obtained by comparing with their true simulated values obtained using observed weather data for each of the years. In order to evaluate if the forecasting accuracy of in-season updated cotton yield was improved significantly, an F-test was performed to see if the standard deviation of residual error (σ_{ε}) between observed and forecasted cotton yield obtained before season was significantly different from that obtained in-season with updated weather data. The F-test was also performed to evaluate if the differences in average standard deviation of simulated cotton yield forecast across all the years reflecting the uncertainties in weather ($\overline{\sigma}_{s}$) were

statistically significant for before season and in-season. These two F-tests provided the basis to answer the first research question.

Comparison of Cotton Yield Forecasts Based on ENSO Indices

Royce et al, (2009) compared predictive potentials of ONI, MEI, and JMA based ENSO classification for three crops of the southeastern United States including cotton. Based on the findings, January-April ONI based monthly ENSO classification and March-May MEI based monthly ENSO classification showed the strongest association with the historic cotton yield. In this study, ONI and MEI based ENSO classification was carried out using the approach of Royce et al, (2009). For example, under MEI classification, historical years having El- Niño phase for March-May months were used to create MEI El- Niño realizations as input to crop model. Similarly, under ONI classification, historical years that had El- Niño phase for January-April months were used to create ONI El- Niño realization as input to crop model. Similarly, La-Niña and neutral phase realizations as input to the models were created. Since JMA is a yearly classification for ENSO, historical years of data that fall under El- Niño, La-Niña, and neutral phase were used to create realizations as input to the crop model, respectively.

Before-season and in-season updates of cotton yield forecasts were then obtained under El-Niño phase, for instance, by providing the model with weather data specific to that particular ENSO phase based on one of three ENSO indices being compared. In order to evaluate which ENSO index showed the highest forecasting potential, the F-test was performed to see if σ_{ε} obtained by three ENSO indices were significantly different from each other. Similarly, $\overline{\sigma}_{s}$ of forecasted cotton yields obtained based on three ENSO indices were also evaluated by the F-test to investigate if the differences

were statistically significant. These statistical evaluations were carried out to answer the second research question.

Comparison between Climatology Based and ENSO Tailored Cotton Yield Forecast

The particular ENSO index that showed the lowest mean standard deviations of model simulations reflecting weather uncertainties ($\overline{\sigma}_s$) and the lowest standard deviation of the residual error of observed and mean simulated cotton yields (σ_ε) were compare with climatology based cotton yield forecasts. Statistical comparisons between climatology based and ENSO based cotton yield forecasts were carried out with F-test in the same manner as described in the above two sections to answer third research question.

Results and Discussion

Comparison between Before-Season and In-Season Cotton Yield Forecasts based on Climatology

Cotton yield forecasts by the CROPGRO-Cotton model that were obtained before the season for 1951-2005 using climatology are shown in Table 4-1. In-season updates of cotton yield forecasts obtained on July 1 and August 1 over the period of 1951-2005 are shown in Table 4-2, and Table 4-3, respectively. Overall, the standard deviations of simulated cotton yield forecasts reflecting weather uncertainties (σ_s) with in-season updates on July 1, and August 1 were reduced in 80% and 90% of the years, respectively. Similarly, the reductions in the residual errors (ε) with in-season updated cotton yield forecasts obtained on July 1 and August 1 were observed 65% and 56% of the time compared to before season cotton yield forecasts. An example of how uncertainties in cotton yield forecasts are reduced with in-season updates with observed

weather is shown in Figure 4-1. The average cotton yield forecast and standard deviation around the mean were higher for forecasts obtained before the season. When, the cotton yield forecasts were updated with observed weather data, average cotton yield forecast approached the observed cotton yields.

Statistical comparison of mean standard deviations of simulated cotton yield forecasts across all the years reflecting weather uncertainties ($\bar{\sigma}_s$) and the standard deviation of residual errors between observed and mean simulated forecast (σ_ε) are shown in Table 4-4. Although the results show reductions of σ_ε and $\bar{\sigma}_s$ on July 1 updated forecasts compared to the before-season cotton yield forecasts, the differences were not statistically significant. However, σ_ε and $\bar{\sigma}_s$ were reduced considerably for August 1 updates. These reductions were highly significant ($p<0.01$). The reduction in σ_ε and $\bar{\sigma}_s$ of August 1 forecast updates relative to before-season cotton yield forecasts were approximately 35% and 32%, respectively.

These results agree with the statement by Wright et al. (1984) that the standard deviation decreased over time with updated forecasts. The main reason was due to the reductions in uncertainties in the weather data over time as observed weather replaced forecast weather.

Comparison of Cotton Yield Forecasts based on ENSO Indices

El-Niño phase

Measures of model deviations for cotton yield using El-Niño forecast based on JMA, MEI, and ONI are shown in Tables 4-5, 4-6, and 4-7. In general, $\bar{\sigma}_s$ and σ_ε for cotton yield forecasts based on MEI were lowest among three ENSO indices based forecasts. Table 4-14 shows statistical comparison of cotton yield forecasts based on all

three ENSO indices for three phases including El-Niño phase. Results show that the MEI was best among the three ENSO indices compared in forecasting the cotton yield before season and in-season for El-Niño forecasts. Although σ_ε of cotton yield forecast based on MEI was 30% lower than JMA and 21% lower than ONI, the differences were not significant. Unfortunately, between 1951 and 2005 there were only 9-12 years classified as El-Niño under those three ENSO indices. Better statistical testing could have been made if the sample size was larger.

La-Niña phase

Cotton yield forecasts obtained using La-Niña forecasts based on JMA, MEI, and ONI are presented in Tables 4-8, 4-9, and 4-10. In general, MEI performed better than other two ENSO indices based cotton yield forecasts. The σ_ε for cotton yield forecasts obtained using MEI based La-Niña forecasts were lowest for before season cotton yield forecast and July 1 in-season updated cotton yield forecasts (Table 4-14) whereas, ONI based La-Niña forecasts were lowest for August 1 in-season updated cotton yield forecasts. But, σ_ε and $\bar{\sigma}_s$ for cotton yield forecasts for all three ENSO indices were not statistically significant under the La-Niña phase.

Neutral phase

Interestingly, cotton yield forecasts obtained using Neutral forecasts based on JMA index showed improved forecasting accuracy compared to the yield forecasts based on other two ENSO indices. (Tables 4-11, 4-12, and 4-13). The σ_ε of cotton yield forecasts based on JMA was significantly different from ONI for the forecasts obtained before the season (Table 4-14). Other than that σ_ε and $\bar{\sigma}_s$ of cotton yield forecasts based on the three ENSO indices did not show statistically significant differences.

Comparison between Climatology-Based and ENSO-Tailored Cotton Yield Forecasts

Although cotton yield forecasting accuracy measures (σ_ε and $\bar{\sigma}_s$) were not significantly different, the ENSO indices that showed the lowest σ_ε and $\bar{\sigma}_s$ for a given phase were compared with climatology based cotton yield forecast. Based on the best results discussed above, MEI was used in the comparison for El-Niño and La-Niña phases, and JMA was used for the Neutral phase.

Statistical comparisons between climatology based and ENSO tailored cotton yield forecasts are shown in Table 4-15. Interestingly, ENSO tailored cotton yield forecasts for El-Niño and La-Niña did not show improvement over climatology-based forecasts. On the contrary, climatology-based cotton yield forecasts showed lower σ_ε compared to ENSO based cotton yield forecasts. The standard deviations were lower under the ENSO tailored cotton yield forecasts, but those results were not significantly different.

Interestingly, the standard deviation of cotton yield forecasts obtained before the season and on July 1 under the Neutral phase based on JMA index showed highly statistically significant reductions compared to the climatology based cotton yield forecasts. In general, ENSO tailored cotton yield forecasts for Neutral phase shows better predictability compared to climatology based cotton yield forecasts. But, for El-Niño and La-Niña phases the cotton yield forecasts did not show statistically significant differences from climatology based forecasts.

Conclusions

Accuracy of cotton yield forecasts was improved with in-season updating the CROPGRO-Cotton model predictions using observed weather data. The in-season updates of cotton yield forecasts were statistically significant for August 1 forecasts but

not statistically significant for July 1 forecasts. On August 1 updates, approximately 90% of the cotton yield forecasts showed reductions in their standard deviations, and 56% of the cotton yield forecasts showed reductions in residual errors compared to before-season cotton yield forecasts.

In general, ENSO indices did not show statistically significant differences in σ_{ε} and $\bar{\sigma}_s$ for El-Niño and La-Niña phases. For the Neutral phase, the JMA index based cotton yield forecasts were better. The $\bar{\sigma}_s$ for JMA based cotton yield forecasts was significantly lower than ONI based cotton yield forecasts.

Comparison between climatology and ENSO based cotton yield forecast results showed that the ENSO tailored cotton yield forecasts did not show significant improvement for El-Niño and La-Niña phases over climatology based cotton yield forecasts. But, under the Neutral phase, the standard deviations of ENSO tailored cotton yield forecasts were significantly lower than climatology based cotton yield forecasts. That shows that the in-season updates of cotton yield forecasts in neutral phases have good potential over using climatology based cotton yield forecast.

This study was conducted with historical weather records over 1951-2005. The limitation of this study was that there were only 9 to 12 years under El-Niño and La-Niña categories. For better statistical comparisons, more years of weather data are needed to increase the number of years in each ENSO phase. Also, only one location was used as a case study in the southeastern United States. More locations should be analyzed.

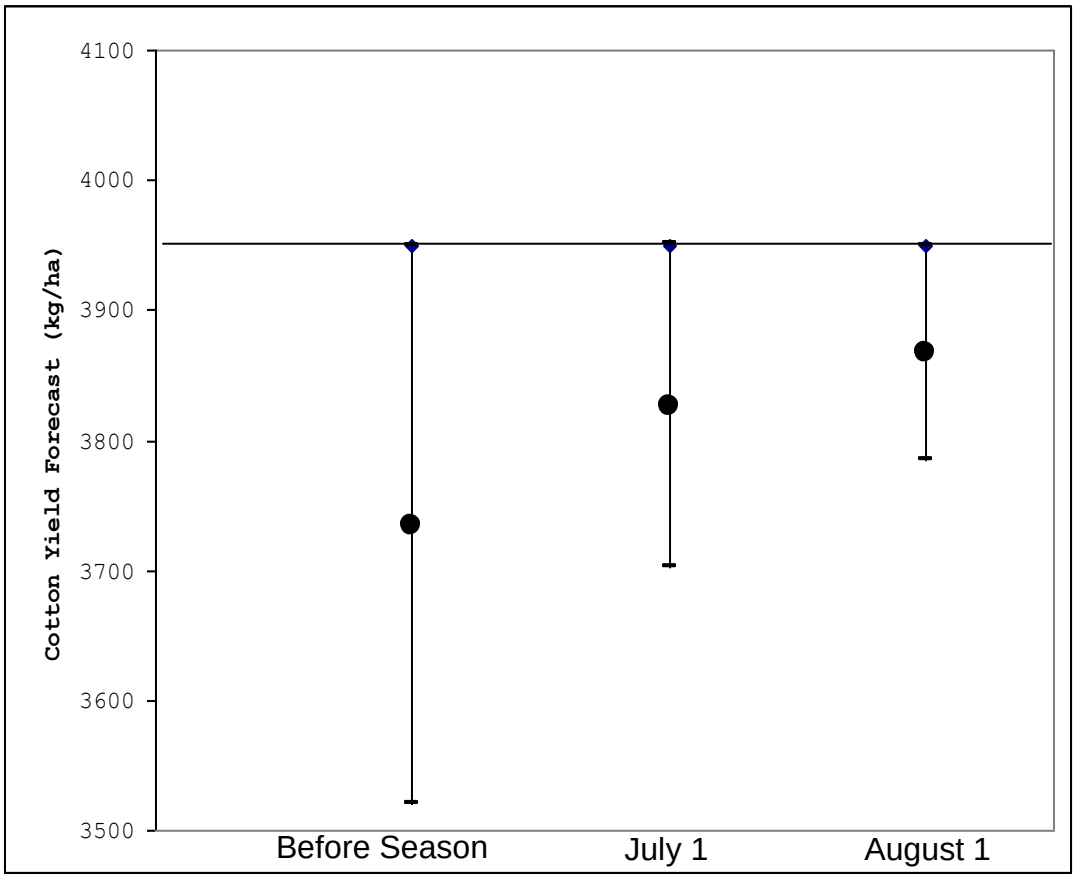


Figure 4-1. Distribution of forecasted cotton yields for the 1980 cotton season at Quincy, Florida simulated using 1951-2005 historical weather data. The solid horizontal line represents “observed” cotton yield for 1980 obtained using model simulation using observed weather.

Table 4-1. Measures of model deviations for seed cotton yield (Quincy, FL) before season using no forecast except climatology

Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	Residual Error (ε) (Kg/ha)	Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	Residual Error (ε) kg/ha
1951	3370	3387	415	-17	1978	3947	3689	455	258
1952	2981	3393	416	-412	1979	3730	3710	459	20
1953	3448	3385	415	63	1980	3950	3735	468	215
1954	3047	3393	412	-346	1981	4286	3743	461	543
1955	3632	3382	414	250	1982	4295	3758	463	537
1956	3529	3382	419	147	1983	2494	3817	436	-1323
1957	4118	3372	402	746	1984	4123	3810	478	313
1958	3311	3393	416	-82	1985	4142	3830	476	312
1959	3671	3398	416	273	1986	4165	3846	478	319
1960	3694	3414	423	280	1987	3987	3875	483	112
1961	3420	3430	421	-10	1988	3585	3912	492	-327
1962	3623	3439	422	184	1989	3876	3930	490	-54
1963	3645	3448	424	197	1990	3463	3953	488	-490
1964	3624	3456	431	168	1991	3882	3964	495	-82
1965	3992	3457	420	535	1992	4096	3969	500	127
1966	3340	3491	428	-151	1993	4055	3978	497	77
1967	3497	3499	431	-2	1994	4382	3995	498	387
1968	3524	3515	440	9	1995	3807	4033	504	-226
1969	3576	3536	435	40	1996	4138	4044	511	94
1970	3484	3554	437	-70	1997	3699	4069	506	-370
1971	3109	3570	434	-461	1998	3599	4108	507	-509
1972	2361	3602	416	-1241	1999	4173	4114	514	59
1973	3708	3610	445	98	2001	4531	4142	516	389
1974	3014	3631	437	-617	2002	4158	4172	522	-14
1975	3990	3626	444	364	2003	5079	4187	511	892
1976	3946	3643	456	303	2004	4342	4208	529	134
1977	2132	3702	398	-1570	2005	4079	4256	529	-177

Table 4-2. Measures of model deviations for seed cotton yield (Quincy, FL) in-season (July 1) using no forecast except climatology

Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε kg/ha
1951	3370	3540	368	-170	1978	3947	3665	405	282
1952	2981	3316	424	-335	1979	3730	3684	467	46
1953	3448	3427	413	21	1980	3950	3826	413	124
1954	3047	3474	361	-427	1981	4286	3848	420	438
1955	3632	3456	385	176	1982	4295	3811	427	484
1956	3529	3474	347	55	1983	2494	3846	348	-1352
1957	4118	3532	341	586	1984	4123	3851	409	272
1958	3311	3396	346	-85	1985	4142	3943	419	199
1959	3671	3350	372	321	1986	4165	4043	430	122
1960	3694	3382	400	312	1987	3987	3904	405	83
1961	3420	3393	404	27	1988	3585	3755	553	-170
1962	3623	3635	356	-12	1989	3876	3988	438	-112
1963	3645	3461	329	184	1990	3463	3915	516	-452
1964	3624	3522	399	102	1991	3882	3949	377	-67
1965	3992	3470	394	522	1992	4096	4008	426	88
1966	3340	3327	395	13	1993	4055	3944	499	111
1967	3497	3541	361	-44	1994	4382	4077	441	305
1968	3524	3521	412	3	1995	3807	3934	569	-127
1969	3576	3301	498	275	1996	4138	4135	430	3
1970	3484	3614	399	-130	1997	3699	4096	448	-397
1971	3109	3650	397	-541	1998	3599	3766	644	-167
1972	2361	3535	345	-1174	1999	4173	4097	415	76
1973	3708	3632	394	76	2001	4531	4217	458	314
1974	3014	3472	507	-458	2002	4158	4189	474	-31
1975	3990	3743	412	247	2003	5079	4321	430	758
1976	3946	3603	390	343	2004	4342	4324	438	18
1977	2132	3630	437	-1498	2005	4079	4057	565	22

Table 4-3. Measures of model deviations for seed cotton yield (Quincy, FL) in-season (August 1) using no forecast except climatology

Year	Observed	Expected	σ_s	ε	Year	Observed	Expected	σ_s	ε
	kg/ha	kg/ha	kg/ha	(Kg/ha)		kg/ha	kg/ha	kg/ha	kg/ha
1951	3370	3653	287	-283	1978	3947	3735	238	212
1952	2981	2537	368	444	1979	3730	3626	311	104
1953	3448	3433	286	15	1980	3950	3867	218	83
1954	3047	3431	258	-384	1981	4286	3852	423	434
1955	3632	3525	218	107	1982	4295	4055	265	240
1956	3529	3516	246	13	1983	2494	3149	473	-655
1957	4118	3673	224	445	1984	4123	3578	286	545
1958	3311	3512	242	-201	1985	4142	4240	280	-98
1959	3671	3491	251	180	1986	4165	4236	279	-71
1960	3694	3454	290	240	1987	3987	3934	410	53
1961	3420	3585	289	-165	1988	3585	4013	388	-428
1962	3623	3540	310	83	1989	3876	3968	372	-92
1963	3645	3535	218	110	1990	3463	3251	538	212
1964	3624	3442	233	182	1991	3882	3800	254	82
1965	3992	3664	395	328	1992	4096	3852	474	244
1966	3340	3318	256	22	1993	4055	4179	357	-124
1967	3497	3593	234	-96	1994	4382	4245	406	137
1968	3524	3512	404	12	1995	3807	4362	271	-555
1969	3576	3441	237	135	1996	4138	3881	522	257
1970	3484	3735	275	-251	1997	3699	4047	313	-348
1971	3109	3577	260	-468	1998	3599	3536	194	63
1972	2361	3230	488	-869	1999	4173	4043	280	130
1973	3708	3456	353	252	2001	4531	4313	357	218
1974	3014	3301	239	-287	2002	4158	4304	314	-146
1975	3990	3894	250	96	2003	5079	4667	291	412
1976	3946	3541	457	405	2004	4342	4456	423	-114
1977	2132	2620	434	-488	2005	4079	4174	389	-95

Table 4-4. Statistical comparison of cotton yield forecasts obtained before-season with in-season updated cotton yield forecasts. *** represents significant difference in σ_ε and SD for in-season updated cotton yield forecast with the σ_ε and SD for before season cotton yield forecasts at 0.01 probability level. The "ns" represents non-significant correlations.

	Cotton Yield Forecast based only on climatology		
	Before Season	01-Jul	01-Aug
σ_ε	457	420^{ns}	299***
$\overline{\sigma}_s$	458	423	320***

Table 4-5. Measures of model deviations for seed cotton yield (Quincy, FL) using El Niño forecast based on JMA index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1952	2981	3197	603	-216	3103	659	-122	2466	319	515
1958	3311	3179	605	132	3262	526	49	3429	311	-118
1964	3624	3213	614	411	3274	565	350	3294	308	330
1966	3340	3274	623	66	3080	588	260	3199	319	141
1970	3484	3323	632	161	3393	568	91	3634	363	-150
1973	3708	3366	638	342	3435	572	273	3420	357	288
1977	2132	3571	492	-1439	3469	611	-1337	2628	413	-496
1983	2494	3661	582	-1167	3729	376	-1235	3148	427	-654
1987	3987	3607	689	380	3735	596	252	3880	428	107
1988	3585	3676	718	-91	3501	824	84	3932	476	-347
1992	4096	3700	717	396	3789	568	307	3794	520	302
2003	5079	3841	664	1238	4065	514	1014	4478	384	601

Table 4-6. Measures of model deviations for seed cotton yield (Quincy, FL) using El Niño forecast based on MEI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1953	3448	3257	429	191	3259	512	189	3343	239	105
1958	3311	3282	435	29	3229	417	82	3416	149	-105
1969	3576	3410	453	166	3084	637	492	3415	262	162
1980	3950	3591	473	360	3628	501	323	3889	162	61
1983	2494	3815	129	-1321	3786	172	-1292	2862	283	-368
1987	3987	3716	489	271	3650	444	337	3686	248	301
1992	4096	3817	510	279	3873	520	223	3607	334	490
1993	4055	3819	506	236	3755	625	300	4051	244	5
2005	4079	4118	551	-39	3836	726	243	4005	259	74

Table 4-7. Measures of model deviations for seed cotton yield (Quincy, FL) using El Niño forecast based on ONI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1958	3311	3337	515	-26	3346	521	-35	3460	286	-149
1966	3340	3439	531	-99	3179	565	161	3254	248	86
1969	3576	3459	539	118	3157	754	420	3432	212	144
1973	3708	3528	549	180	3534	540	174	3318	448	390
1983	2494	3881	309	-1387	3800	275	-1306	2988	569	-494
1987	3987	3784	591	203	3801	589	186	3775	516	213
1992	4096	3886	636	210	3832	562	264	3698	590	398
1995	3807	3969	619	-162	3777	893	30	4350	228	-543
2003	5079	4006	524	1073	4084	514	995	4566	232	513

Table 4-8. Measures of model deviations for seed cotton yield (Quincy, FL) using La Niña forecast based on JMA index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1955	3632	3384	537	248	3490	468	142	3513	211	119
1956	3529	3357	535	172	3431	474	98	3538	244	-9
1957	4118	3340	488	778	3507	448	611	3715	248	403
1965	3992	3442	530	550	3488	448	504	3768	409	224
1968	3524	3499	563	25	3598	471	-74	3615	429	-91
1971	3109	3627	549	-518	3682	505	-573	3603	132	-494
1972	2361	3678	414	-1317	3603	341	-1242	3415	281	-1054
1974	3014	3704	538	-690	3610	437	-596	3350	133	-336
1975	3990	3629	569	361	3826	492	164	3970	246	20
1976	3946	3616	582	330	3549	508	397	3641	430	305
1989	3876	3973	632	-97	4030	553	-154	4043	333	-167
1999	4173	4140	659	33	4070	559	103	4077	304	96

Table 4-9. Measures of model deviations for seed cotton yield (Quincy, FL) using La Niña forecast based on MEI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1954	3047	3409	308	-362	3500	283	-453	3405	223	-358
1955	3632	3343	318	289	3425	294	207	3468	251	164
1956	3529	3316	337	213	3422	284	107	3498	253	31
1962	3623	3406	331	217	3639	272	-16	3536	335	87
1964	3624	3382	346	242	3530	326	94	3397	245	227
1968	3524	3455	365	69	3499	308	25	3578	431	-54
1971	3109	3593	313	-484	3650	310	-541	3550	187	-441
1974	3014	3670	276	-656	3524	352	-510	3314	212	-300
1976	3946	3561	369	385	3498	308	448	3609	434	337
1999	4173	4079	417	94	4065	337	108	4045	327	128

Table 4-10. Measures of model deviations for seed cotton yield (Quincy, FL) using La Niña forecast based on ONI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1955	3632	3384	357	248	3500	278	132	3446	235	186
1956	3529	3352	383	177	3449	307	80	3522	261	7
1968	3524	3499	414	25	3609	312	-85	3686	227	-162
1971	3109	3662	328	-553	3748	273	-639	3550	132	-441
1974	3014	3748	269	-734	3661	299	-647	3331	150	-317
1975	3990	3623	365	367	3845	276	145	3912	280	78
1976	3946	3603	423	343	3564	312	382	3682	200	264
1999	4173	4137	448	36	4093	354	80	4072	276	101

Table 4-11. Measures of model deviations for seed cotton yield (Quincy, FL) using Neutral forecast based on JMA index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed	Expected	σ_s	ε	Expected	σ_s	ε	Expected	σ_s	ε
1951	3358	3484	224	-126	3613	216	-255	3658	287	-300
1953	3449	3481	225	-32	3477	316	-28	3419	300	30
1954	3038	3495	208	-457	3563	207	-525	3503	212	-465
1959	3671	3490	225	181	3392	236	279	3484	239	187
1960	3697	3524	235	174	3465	262	232	3474	293	223
1961	3419	3528	228	-109	3494	258	-75	3641	274	-222
1962	3623	3535	230	88	3691	226	-68	3510	311	113
1963	3644	3544	231	100	3528	195	116	3567	182	77
1967	3494	3600	235	-106	3625	226	-131	3632	206	-138
1969	3575	3638	238	-63	3343	373	233	3450	229	125
1978	3947	3788	254	159	3724	248	224	3791	193	156
1979	3771	3813	258	-42	3776	317	-5	3681	275	90
1980	3949	3851	267	98	3919	274	30	3913	187	36
1981	4301	3838	248	463	3928	276	373	3834	460	467
1982	4295	3854	252	441	3848	279	447	4072	189	223
1984	4126	3926	272	200	3940	262	186	3599	296	527
1985	4143	3931	270	212	3990	258	153	4261	260	-118
1986	4165	3948	271	217	4130	290	35	4278	243	-113
1990	3472	4070	259	-598	3995	367	-523	3173	595	299
1991	3892	4074	282	-182	4012	231	-120	3801	251	91
1993	4056	4085	286	-29	4016	353	40	4209	355	-153
1994	4382	4098	282	284	4159	296	223	4272	443	110
1995	3812	4148	284	-336	4004	435	-192	4409	208	-597
1996	4160	4170	294	-10	4199	258	-39	3799	555	361
1997	3710	4187	280	-477	4201	301	-491	4138	265	-428
2001	4530	4248	299	282	4303	301	227	4332	347	198
2002	4261	4283	306	-22	4262	323	-1	4309	301	-48
2004	4347	4331	307	16	4394	261	-47	4426	433	-79
2005	4019	4374	305	-355	4137	415	-118	4162	404	-143

Table 4-12. Measures of model deviations for seed cotton yield (Quincy, FL) using Neutral forecast based on MEI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1951	3370	3504	280	-134	3649	249	-279	3699	285	-329
1952	2981	3532	270	-551	3458	312	-477	2582	426	399
1959	3671	3504	281	167	3445	258	226	3541	267	130
1960	3694	3529	306	165	3408	232	286	3475	280	219
1961	3420	3546	286	-126	3458	238	-38	3561	282	-141
1966	3340	3616	285	-276	3416	287	-76	3353	264	-13
1970	3484	3674	297	-190	3715	265	-231	3803	251	-319
1973	3708	3727	306	-19	3738	239	-30	3538	365	170
1978	3947	3802	313	145	3779	259	168	3738	229	209
1979	3730	3832	317	-102	3782	305	-52	3703	269	27
1982	4295	3867	308	428	3937	326	358	4065	280	230
1984	4123	3933	348	190	3914	262	209	3566	276	557
1986	4165	3966	326	199	4149	296	16	4265	275	-100
1988	3585	4067	343	-482	3908	454	-323	4036	368	-451
1995	3807	4177	335	-370	4146	474	-339	4394	268	-587
2001	4531	4266	353	265	4336	291	195	4378	342	153
2003	5079	4292	307	787	4403	278	676	4693	310	386
2004	4342	4356	378	-14	4440	320	-98	4549	460	-207

Table 4-13. Measures of model deviations for seed cotton yield (Quincy, FL) using Neutral forecast based on ONI index

		Before-Season			In-Season (July 1)			In-Season (August 1)		
Year	Observed kg/ha	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)	Expected kg/ha	σ_s kg/ha	ε (Kg/ha)
1952	2981	3488	358	-507	3390	337	-409	2520	398	461
1953	3448	3462	372	-14	3487	338	-39	3476	295	-28
1954	3047	3477	362	-430	3561	321	-514	3529	227	-482
1957	4118	3437	348	681	3580	291	538	3701	216	417
1959	3671	3471	372	200	3398	312	273	3516	234	155
1960	3694	3498	375	196	3446	367	248	3490	264	204
1961	3420	3509	377	-89	3475	365	-55	3634	268	-214
1962	3623	3516	379	107	3681	312	-58	3579	317	44
1963	3645	3525	380	120	3510	291	135	3588	190	57
1964	3624	3543	382	81	3602	364	22	3522	188	102
1965	3992	3529	372	463	3529	334	463	3681	387	311
1967	3497	3581	385	-84	3621	336	-124	3650	211	-153
1970	3484	3638	390	-154	3705	361	-221	3809	226	-325
1972	2361	3716	302	-1355	3634	265	-1273	3252	530	-891
1978	3947	3769	409	178	3736	352	211	3804	191	143
1979	3730	3795	413	-65	3776	407	-46	3704	276	26
1980	3950	3826	421	124	3920	388	30	3923	175	27
1981	4286	3820	409	466	3942	388	344	3887	442	399
1982	4295	3836	412	459	3854	364	441	4119	193	176
1984	4123	3899	429	224	3930	380	193	3636	249	487
1986	4165	3932	429	233	4128	407	37	4305	233	-140
1990	3463	4055	427	-592	4007	404	-544	3240	585	223
1991	3882	4059	442	-177	4008	329	-126	3848	257	34
1993	4055	4070	446	-15	4038	418	17	4258	335	-203
1994	4382	4082	445	300	4160	411	222	4302	437	80
1997	3699	4171	448	-472	4191	427	-492	4146	259	-447
2002	4158	4267	470	-109	4270	422	-112	4359	296	-201
2004	4342	4306	479	36	4405	391	-63	4487	418	-145

Table 4-14. Statistical comparison of cotton yield forecasts tailored to ENSO forecasts by three indices. “a” represents that the σ_ε for JMA-Neutral is significantly lower than σ_ε of ONI-Neutral. All other values were not statistically different from each other.

Type of climate forecast	Cotton Yield Forecast based on ENSO forecast					
	Before Season		01-Jul		01-Aug	
	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)
El-Niño						
JMA	716	631	660	580	398	385
MEI	517	441	533	506	241	242
ONI	636	535	606	579	383	370
La-Niña						
JMA	582	550	528	475	399	283
MEI	363	338	334	307	264	289
ONI	412	373	376	301	248	220
Neutral						
JMA	262[a]	260	246	285	264	303
MEI	330	313	289	297	312	305
ONI	403	401	373	360	309	296

Table 4-15. Statistical comparison of cotton yield forecasts using only climatology forecast with ENSO tailored cotton yield forecasts using MEI. "***" represents that the standard deviations of expected cotton yield forecasts obtained from JMA-Neutral is significantly lower than the standard deviations of expected cotton yield forecasts obtained from climatology.

Type of climate forecast	Cotton Yield Forecast Comparison					
	Before Season		01-Jul		01-Aug	
	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)	σ_ε kg/ha	$\bar{\sigma}_s$ (kg/ha)
Climatology	471	464	487	434	261	342
MEI-El Niño	517	441	533	506	241	242
Climatology	324	437	299	397	274	291
MEI-La Niña	363	338	334	307	264	289
Climatology	257	467	239	430	244	314
JMA-Neutral	262	260***	246	285***	264	303

CHAPTER 5
COTTON YIELD FORECASTING FOR THE SOUTHEASTERN USA USING CLIMATE INDICES

Introduction

Cotton is the most important fiber crop in the United States, accounting for approximately 20% of the total production in the world and more than $25 billion in products and services annually (USDA ERS, 2009). Cotton production in the southeastern United States averages about 22% of the total upland cotton production in the United States (USDA ERS, 2009). Georgia and Alabama hold the major share of total cotton produced in the southeastern United States. Recently, there has been a significant increase in cotton planted those two states. While comparing average acreage planted, there was an increase of about 26% in Georgia 41% in Alabama during the last decade (NASS, 2007).

While evidence clearly shows an increase in cotton planted over time, climate variability is a major concern that could adversely affect its production in the southeastern United States. An effective way to reduce agricultural vulnerability to climate variability is through the implementation of adaptation strategies. For example crop yield forecast could be used by farmers to mitigate negative consequences of unfavorable climate forecast, or benefit from anticipated favorable climate conditions (Baigorra et al., 2007). If growers know the expected cotton yield for the coming season, they may be able to decide on alternative management strategies to reduce the cotton production risk (Jones et al., 2000; Hansen, 2005; Vedwan et al., 2005; Jagtap et al., 2002). For example, growers could purchase appropriate crop insurance ahead of time in order to compensate for an adverse effect of climate variability on their cotton yields.

Given that the times at which different crop yield forecasts may be made differ, the techniques to obtain them also vary. Seasonal cotton yield forecasts are useful if they are available as early as February, before the growing season starts, so that growers can make use of them to decide on purchasing seeds, fertilizers, and insurance policies in advance (personal communication with David Wright and Clyde Fraisse). For instance the deadline for growers to make crop insurance-related decisions in this region is March 15. So, if they would like to make insurance decisions based on expected yields for the coming season, then the crop yield forecasts should be available before that date.

While the growth and development of crops are known to be influenced by weather during the growing season, it is a common practice to predict crop yield based on weather variables (Sakamoto, 1979; Idso et al., 1979; Walker, 1989; Alexandrove and Hoogenboom, 2001). However, crop yield predictions based on observed weather cannot be made available before the planting season (Kumar, 2000). Attempts to obtain long-term forecasts using alternatives to weather variables, such as climate indices that exhibit teleconnections with weather, are limited.

The El Niño-Southern Oscillation (ENSO) phenomenon is one of the most significant drivers of climate and agricultural variability in the southeastern United States (Hansen et al., 1998; Ropelewski and Halpert 1986, Kiladis and Diaz 1989; Mo and Schemm 2008; Mennis 2001). Although, the ENSO effects in southeastern United State are stronger during the winter, their effects are not very strong during the summer months (Baigorria et al., 2007). Baigorria et al. (2010) also showed that cotton was not strongly affected by ENSO itself, however, ENSO in conjunction with other oceanic

(Pacific and Atlantic) and atmospheric patterns may be useful for forecasting cotton yield. In a parallel study, Martinez et al. (2009) used sea surface temperature (SST) in the tropical North Atlantic (TNA), the Pacific-North American (PNA) index and Bermuda High Index (BHI) along with the ENSO index to predict corn yields in the southeastern United States. The findings of Martinez et al. (2009) showed good potential for using climate indices to forecast corn yield in this region.

Large scale teleconnection indices greatly influence the climate and agriculture in the southeastern United States (Stenseth et al., 2003; Enfield, 1996; Barnston et al., 1991; Bell and Jenowiak, 1994; Martinez et al., 2009). The research question addressed was; are there teleconnections between large scale climate indices and cotton yield that would provide the basis for forecasting cotton yield in southeastern United States? The objectives of this study were to evaluate the relationships between large scale climate indices and cotton yield and to evaluate the skill of cotton yield forecasts.

Materials and Methods

Historic Cotton Yield Data

County level yield data for cotton were obtained from National Agricultural Statistical Services (NASS, 2008) for 57 years from 1951- 2007 for a total of 64 cotton producing counties in Georgia and Alabama. Cotton producing counties in Florida were not considered in this study because there was only one county, Santa Rosa that had a county reported cotton yield available for 57 years. The NASS cotton yields did not distinguish between irrigated and rain-fed cotton production.

The time series of historic cotton yield over the time period between 1951 and 2007 showed a gradual upward trend. The factors contributing towards this upward

trend likely include effects of fertilizers, pesticides, improved cultivars, and enhanced management practices, but not necessarily climate (Wigley and Qipu, 1983). In order to evaluate the correlations of historic cotton yield with climate indices, the effects of non climatic influences on historic yield needed to be removed. For this study, technological improvements in cotton yield over time was assumed to be linear, hence county cotton yields were detrented by removing the linear trend. The percentage deviation of yield from the trend line (% residual) was computed for each year (Eq. 4-1).

$$Y_{residuals} = \left(\frac{Y_{observed}}{Y_{trend}} - 1 \right) * 100 \qquad (5\text{-}1)$$

Climate Data

Monthly average temperature and monthly cumulative precipitation for 64 weather stations in Georgia and Alabama were obtained from National Climatic Data Center (NCDC) for 1966-2007. Climate data for May-September months were used in the analysis instead of full year of data since those months typically coincides with the cotton growing season in the southeastern United States. The main reason for limiting the climate data to 1966-2007 instead of 1951-2007 was the fact that there were many counties, especially in Alabama, that had missing climate data between 1951 and 1965.

Atmospheric and Oceanic Climate Indices

Stenseth et al. (2003) stated that any single climate index may possibly explain only a relatively small part of the local climate variability. They suggested that one should use climate indices that pick up most of the relevant climate-weather variations for the specific system under study. In this study, a total of seven climate indices were used: Oceanic Niño Index (ONI), Tropical North Atlantic (TNA) SST index, Atlantic Meridional Mode (AMM) index, North Oscillation Index (NOI), North Pacific (NP) pattern,

Quasi-Biennial Oscillation (QBO) index, and Tropical North Hemisphere (TNH) index. Descriptions of each of those indices as well as their effects on climate of southeastern United States as described in the literature are as follows.

Oceanic niño index (ONI)

The ONI anomalies are the running means of SST anomalies in the NIÑO 3.4 region (5°N-5°S, 120°-170°W) and were obtained from the NOAA climate prediction center. The ONI has become the National Oceanic and Atmospheric Administration (NOAA) standard for categorizing the ENSO events in the tropical pacific. However, continuous monthly values of ONI were selected in this study and not the ENSO phases.

ENSO exhibits strong correlations with temperature and precipitation in the United States (Ropelewski and Halpert, 1986) and southeastern United States in particular (Mote, 1986, Hansen et al., 1998, O'Brian et al., 1996). Since studies show clear evidence of impact of ENSO on southeastern United States climate, ONI index was chosen to represent ENSO index in this analysis.

Tropical north Atlantic (TNA) index

The TNA index (Enfield et al., 1999) is an anomaly of the average of the monthly SST from 5.5^{o}N-23.5^{o}N, 15°-57.5^{o}W. The TNA data were obtained from the Physical Science Division of the Earth System Research Laboratory (ESRL) at National Oceanic and Atmospheric Administration (NOAA).

Association of TNA with the climatic conditions of southeastern United States has been documented in the literature. For instance, Wang et al. (2008) demonstrated that the variability in the summer precipitation for the southeastern United States is strongly associated with Atlantic SST. Enfield, (1996) investigated a teleconnection between

Atlantic SST and precipitation in the southeastern United States. Results from the study by Martinez et al. (2009) clearly indicated an association of Atlantic SSTs with temperature and precipitation in the southeastern United States.

Atlantic meridional mode (AMM) index

The AMM SST index is a gathering of cross equatorial meridional SST anomalies in the tropical Atlantic index (Chiang and Vimont, 2004; Takeshi et al., 2010). The AMM index is correlated with TNA index and plays an important role in inter-annual and decadal climate variability and is closely linked with hurricane activities in the southeastern United States (Xie et al., 2005; Vimont and Kossin 2007; Kossin and Vimont 2007).

North oscillation index (NOI)

The NOI index represents differences between the sea level pressure anomalies at the north pacific height (NPH) in the northeast Pacific and near Darwin, Australia (Schwing et al., 2002). Monthly NOI index values were obtained from the Pacific Fisheries Environmental Laboratory (PFEL) of NOAA.

The NOI is closely related to El Niño and La Niña events where positive values of NOI are reflective of La Niña conditions and negative values of NOI are closely linked with El Niño conditions (Lee and Sydeman, 2009). Since, NOI is linked closely with ENSO events; it could impact climatic conditions of southeastern United States.

North pacific (NP) pattern

Another sea level pressure based index used in this study was NP index, that represents area weighted SLP over the region 30°N-65°N, 160°E-140°W (Trenberth and Hurrell, 1994). The positive phase of the NP pattern is associated with enhanced

cyclonic circulations of pacific jet stream over the southeastern United States which affects climatic conditions in this region (Bell and Jenowiak, 1994).

Tropical north hemisphere (TNH) index

The TNH pattern, first classified by Mo and Livezey (1986) reflects large-scale changes in both the location and eastward extent of the pacific jet stream and thus this pattern significantly modulates the flow of marine air into the United States and to the southeastern United States (Washington et al., 2000). Barnston et al. (1991) found that the negative phase of TNH pattern was often observed when the pacific warm condition (El Niño) is present, which would eventually affect the climate conditions in the southeastern United States.

Quasi-biennial oscillation (QBO) index

The QBO represents the oscillation of zonal winds in the stratosphere (at 30 mb) over the equator in the Pacific that blow eastward or westward in a cycle that averages about 28 months. The effect of QBO in southeastern United States is evident due to its close association with the ENSO (Thompson et al., 2001). For instance, during an El Niño years with an easterly QBO the temperature tend to be below normal across the southeastern United States (Barnston et al., 1991).

Correlation Analysis

The correlation analysis of climate indices with precipitation and temperature and with cotton yield residuals was carried out using Pearson correlation method. The statistical significance of the correlations was evaluated at $p < 0.10$.

In the following sections correlations of climate indices with temperature, precipitation, and cotton yield are described. The correlation analysis of cotton yield with precipitation and temperature was not performed in this study, because in a parallel

study Baigorria et al. (2007) correlated cotton yield with surface temperature and rainfall for 57 cotton producing counties in Georgia and Alabama. The results showed that surface temperature and rainfall in this region were significantly correlated with cotton yield for more than 50% of the counties processed. Results also showed that the correlations were highly significant, especially for July rainfall for 84% of the counties. Since the clear evidence of significant correlation of cotton yield with surface temperature and rainfall was shown by Baigorria et al. 2007, it was not re analyzed in this study.

Correlations of Climate Indices with Temperature and Precipitation

It was unclear whether the monthly climate indices prior to the cotton growing season were significantly correlated with monthly temperature and precipitation during the cotton growing season for all cotton growing counties in Georgia and Alabama selected in this study. Climate indices for January and February were correlated with May-September temperature and precipitation for 1966-2007.The significant correlations of climate indices with temperature and precipitation during the cotton growing season could provide a basis for using climate indices to forecast cotton yield.

Correlations of Climate Indices with Cotton Yield

Climate indices for January and February were correlated with county cotton yield for 1966-2007. Correlation analyses were carried out for 64 cotton producing counties of Georgia and Alabama. Since the correlations of cotton yield with temperature and precipitation during the growing season is known (Baigorria et al., 2007), identifying statistically significant correlations of climate indices with climate variables and cotton yield would help us understand the behavior of climate- cotton yield interactions and the forecasting of cotton yield.

Principal Component Regression

All climate indices were summarized using a principal component analysis (PCA). The PCA is a data transformation technique that transforms original highly correlated variables into a new set of independent variables in a way that the first principal component (PC) describes the highest variance followed by the second PC and so on (Massy, 1965).

The motivation for using principal components of climate indices instead of their original values was because of the significant correlations among the climate indices (Barnston et al., 1991; Barreiro et al., 2005). If highly correlated climate indices were used in the regression model, then the assumption of mutually independent explanatory variables is violated. Instead, the principal components of climate indices transform them into mutually independent variables that can be effectively used in multiple linear regression models. Another leading advantage of using PCA is that it is an efficient data-reduction technique. For instance, if the maximum variances of two climate indices are summarized by principal component 1, then instead of using two indices one can use just one PC to utilize the information.

The principal component regression (PCR) model uses principal components as explanatory variables. The dependent variable for this model was detrended historic cotton yield and independent variables were principal components of climate indices. The general form of model is shown below:

$$Y_{residuals} = \beta_o + \beta_1 X_1 + ... + \beta_p X_p + \varepsilon \qquad (5\text{-}2)$$

Where, $Y_{residuals}$ is cotton yield residual for a given county, and $X_1 \ldots Xp$ are the principal components of climate indices retained in the PCR model based on backward

stepwise regressions. The residual error ε from multiple regression analysis was tested for normality. This is important because of the assumption in regression analysis that the residual errors are normally distributed.

Leave One Out Cross Validation

The yield forecasts from different models for each county were evaluated using cross validation. This statistical validation approach can be used to validate the model when data are limited. With the cross validation approach, observed data are iteratively and exhaustively used for model testing, resulting in more reliable evaluation than getting estimates from the two-group partition method and less biased than estimates derived from calibration-dependent dataset (Jones and Carberry, 1994). In this method, n-1 data were used to estimate the parameters for the regression model and the one left out data point was used for model evaluation. By an iterative process, all the data points were used for validation. The skill of the forecast was evaluated using the statistically significant correlations and mean squared error (MSE) between observed and forecasted cotton yield.

Categorical Yield Forecast – Contingency Table

Yield forecasts obtained from the principal component regression models were divided into two categories; above average and below average cotton yield. A negative cotton yield residuals falls in the below average yield category and vice versa. A 2x2 contingency table (Table 5-1) was used to evaluate the skill of the forecast. Categorical cotton yield forecasts were evaluated for statistical significance using the Pearson's chi-square test of association (Plackett, 1983). This method tests an association between observed and forecasted categorical cotton yield. Statistical significance were evaluated at $p < 0.10$.

Since it is not possible to forecast absolute cotton yields accurately, knowing whether to expect an above average or below average yield for the coming season is also useful information to growers. The percentages of correct yield forecasts, the probability of detecting (POD) above average yield, and the probability of detecting below average yields were calculated using the following equations and Table 5-1.

$$PercentCorrect = 100 * [(A + E) / I] \quad (5\text{-}3)$$

$$POD_{AboveAverage} = E / F \quad (5\text{-}4)$$

$$POD_{BelowAverage} = A / C \quad (5\text{-}5)$$

The $PercentCorrect$ in (3-3) represents the ratio of total number of correct yield forecasts to the total number of forecasts. The $POD_{AboveAverage}$ (3-4) shows the ratio of total number of correct above average forecast to the total number of observed above average cotton yield residuals. The $POD_{BelowAverage}$ (3-5) shows the ratio of total number of correct below average yield forecasts to the total number of observed below average cotton yield residuals.

Results and Discussion

Historic Cotton Yield Data

Time-series of historic cotton yield showed a gradual upward trend over the time period of 1951-2007. The gradual upward trend was expected due to technological improvements over time. It was assumed that the technological trend over time was linear; the historic cotton yield data fit well the linear trend line. Once, the trend was removed, variability in cotton yield residuals was observed around the trend line closer to zero. This variability in cotton yield residuals could be due to the effect of climate variability, henceforth were analyzed for their correlations with climate indices.

Correlation Analysis

Correlations of climate indices with temperature and rainfall

Overall, the correlations of climate indices with temperature showed statistically significant results at $p < 0.10$. Figure 5-1 shows the most prominent correlations of all seven climate indices with temperature. It was interesting to note that all seven climate indices exhibited maximum correlations with temperature during the month of July. Climate indices such as NOI, NP, and TNH exhibited statistically significant correlations with more counties than the other indices. Except for one county in Alabama, all seven counties together showed statistically significant correlations with temperature in all cotton producing counties considered in this study.

As can be seen from Figure 5-1 the NOI, NP, and TNH indices were negatively correlated with July temperature. This means that when these climate indices are in a negative phase during January and February, July temperature is expected to be above average for the southeastern United States. On the other hand, AMM, QBO, TNA, and ONI were positively correlated with July temperature. AMM, TNA, and ONI are Atlantic and Pacific SSTs and hence, warming of the SST during January and February was responsible for higher July temperatures.

The correlations of all seven climate indices with rainfall are shown in Figure 5-2. The NOI and NP exhibited a negative correlations with July rainfall; opposite in sign to their correlations with July temperature. TNH did not exhibit strong correlations with July rainfall; however, its correlation was strong for September rainfall in a majority of the counties in Georgia and several counties in Alabama. It was interesting to note that the TNA showed negative correlations with June rainfall and positive correlations with

September rainfall. These opposite correlations with June and September could eventually affect the cotton yield in the same manner because higher rainfall during the early growing season is favorable to cotton but the same during the maturity would adversely affect cotton yield.

Correlation of climate indices with cotton yield

Figure 5-3 shows correlations of climate indices with cotton yields. The NOI index was positively correlated with cotton yield. NOI was positively correlated with July precipitation and negatively correlated with July temperature; hence a positive correlation with cotton yield was expected for NOI. The NP pattern had a strong negative correlation with July temperature, but the correlations with cotton yield was positive for all the counties. This means that during the negative phase of NP, it is likely to have higher July temperature and lower cotton yield. The month of July typically coincides with the flowering stage of cotton growth in southeastern United States. Increased temperature during July stimulates the photosynthesis and leaf expansion and crop water requirements. This could increase water stress in non-irrigated cotton plants and also reduce the allocation of daily assimilates to the fruit. As a result, final cotton yield could be either positively or negatively affected by July temperature.

TNA and AMM showed negative correlations with cotton yield. As mentioned in the previous section, TNA exhibited negative correlation with June and positive correlation with September rainfall. Both would adversely affect cotton yield and hence be consistent with the negative correlation of TNA with cotton yield. TNH was also negatively correlated with cotton yield while it had a positive correlation with September rainfall. Rainfall during the late maturity stage of cotton has limited or no use to the plant, on the contrary; it could create an adverse effect. Frequent precipitation during

September and October months is the main cause of one of the most common cotton disease called hard lock of cotton, which would significantly reduce cotton yield.

It was interesting to note that NOI was significantly correlated with rainfall for more counties than it was correlated with cotton yields. The possible explanation could be an effect of irrigation. If irrigation is a typical practice in those counties, then the effect of rainfall on cotton yield would be negligible and hence yield for those counties may lack significant correlations with NOI.

As can be seen in Figure 5-3, a single month climate index alone was not sufficient to correlate significantly with cotton yield for all counties. All seven indices together were significantly correlated with cotton yield and climate for all the counties. In general, the correlations of climate indices with cotton yield and climate during the growing season showed potential use of forecasting cotton yield for the southeastern United States.

Principal Component Regression

Table 5-3 shows significant PCs retained in the backward stepwise regression procedure and Table 5-4 shows the loadings of climate indices in the principal components. Principal component 1 (PC1) was retained in 44% of the regression models for counties in Georgia. The loadings show that the TNA had the highest loading in PC1, followed by ONI and NOI indices. In contrast, PC1 was only significant in 22% of the counties in Alabama. Conversely, Principal Component 3 (PC3) was significant in 43% of the models for counties in Alabama but only in 20% of the models for counties in Georgia. It can be seen from Table 5-4 that the PC3 had the highest loadings from the QBO index. This was expected because correlations of QBO with cotton yield were more prominent in counties of Alabama than in Georgia (Figure 5-3). Interestingly,

Principal Component 11 (PC 11) was significant for approximately 50% of the counties for all the counties in Georgia and Alabama. Based on the loadings in PC 11 it can be seen that the TNA and AMM had maximum loadings compared to other indices.

Significant PCs retained by backward stepwise regression model were not in top to down order (Table 5-3). In other words, PC1 was not necessarily the most significant PC for all the counties. This was due to the fact that the first PC explains the maximum variance among the explanatory variables but not necessarily of the dependent variable in principal component regression (Sutter et al., 1992). For example, loadings of QBO index were higher in PC3 compared to PC1. Since QBO correlation with cotton yield was more prominent in Alabama; PC3 was significant more times than PC1.

Leave One Out Cross Validation

The accuracy of cotton yield forecasts by principal component regression models were evaluated with the leave one out cross validation approach. 77% of the counties in Georgia and 70% of the counties in Alabama showed significant correlations between observed and cross validated cotton yield residuals (Figure 5-4, Table 5-2). The highest significant correlation of 0.50 was obtained for Shelby County in Alabama whereas the lowest significant correlation of 0.22 was obtained for Colquitt County in Georgia and Lee County in Alabama. A total of 8 out of 34 counties in Georgia and 9 out of 30 counties in Alabama did not show significant correlations between observed and forecasted cotton yields. The possible reasons for this could be that the proportion of irrigated cotton yields for those counties might be higher than rainfed cotton. The irrigated cotton could diminish the direct impact of rainfall on the cotton yield and hence result in non-significant correlations. For example, all seven climate indices showed statistically significant correlations with climate during the cotton growing season in

Calhoun County in Georgia but did not show a significant correlation with cotton yield and subsequently no predictability. It is possible that the climate indices show significant correlation with climate during the cotton growing season but not with cotton yield, if the reported cotton yield came from irrigated practices.

Baigorria et al. (2008) stated that the highest yield counties in the southeastern United States showed weakest predictability. The speculation was that the climate based yield predictability is weaker in counties with greater proportion of cultivated areas under irrigation. The results obtained in this study are in agreement with that speculation because some of the highest cotton yielding counties, such as Dooly, Colquitt, Mitchell, and Crisp showed little or no predictability. Unfortunately, the NASS data do not distinguish between irrigated and rainfed cotton yield, but this could be the possible explanation of weaker predictabilities for those counties.

Figure 5-5 shows the time series comparisons between historic cotton yield residuals and cross validated cotton yield for the highest and the lowest correlated counties in Georgia and Alabama. It was evident that the principal components regression models of climate indices were able to capture year-to-year variability in cotton yield fairly well. The mean squared errors (MSE) between the historic cotton residuals and cross validated cotton yields were within the range of 0.03-0.10. It can be seen from Figure 5-6 that the average errors across all the counties was normally distributed with mean being very close to 0. Overall, the cotton yield residuals forecasted with principal components of climate indices showed good predictability of cotton yields.

Categorical Yield Forecast – Contingency Table

Results obtained from the contingency table are shown in Table 5-5 and Figures 5-7 and 5-8. A total 67% of the counties in Georgia and 73% of counties in Alabama showed statistically significant association between categorical observed and cross validated cotton yield forecast.

A total of 94% of the counties showed that the categorical cotton yield forecast obtained from the cross validated cotton yield at the lead time of approximately two to three months before planting (In February) was correct more than 50% of the time. The highest percent correct score was obtained for Bulloch County in Georgia that had correct categorical forecasts 73% of the time based on cross validated results. Figure 5-7 shows the spatial distributions of percent correct categorical cotton yield forecasts for counties in Georgia and Alabama. It was interesting to note that except for Thomas County in Georgia, the categorical cotton yield forecast was correct more than 50% of the time even though eight counties that were not significantly correlated with cross validated cotton yields. This shows that even if the model was not able to significantly capture the variability of the observed cotton yield residuals, a majority of forecasts were in the right category (above average or below average) more than 50% of the time.

As we know, the probability of detecting above average or below average yield is 50% purely by chance. With the categorical forecast of cotton yield obtained from the principal components of climate indices, the probability of detecting above average or below average yields was better than chance for 88% and 94% of the counties, respectively. In applications of these results, growers might be more interested in

knowing if they would expect a better than average yield or vice versa for their decision making processes.

Conclusions

The association of seven climate indices with climate during the cotton growing season and with county average historic cotton yields was evaluated using Pearson's correlations. January and February monthly climate indices, which were months prior to the cotton growing season, exhibited statistically significant correlations with climate during the cotton growing season as well as with cotton yields. July temperature showed the strongest correlations with all seven climate indices whereas the strongest correlation of climate indices with rainfall varied between June, July, August, and September.

The accuracy of cotton yield forecasts based on principal component regressions were evaluated with the leave one out cross validation approach. With a lead time of approximately 2 months before the typical planting period on the southeastern United States, about 77% of the counties in Georgia and 70% of the counties in Alabama showed statistically significant correlations between observed and forecasted cotton yields. The MSE between observed and cross validated cotton yield residual forecasts were in the range of 3-11%.

Categorical cotton yield forecasts obtained from the cross validated results were evaluated using the skill measures of percent correct forecasts, probability of detecting above average cotton yield, and probability of detecting below average yields. 94% of the counties showed the categorical cotton yield forecast obtained at a lead time of approximately two months before planting (In February) was correct for more than 50% of the time. The probabilities of detecting above and below average yields were better

than chance in 88% and 94% of Alabama and Georgia counties, respectively. The principal component regression of climate indices showed potential to become a useful tool to forecast cotton yield with a lead time of approximately two months before the typical cotton planting time of May first week in the southeastern United States.

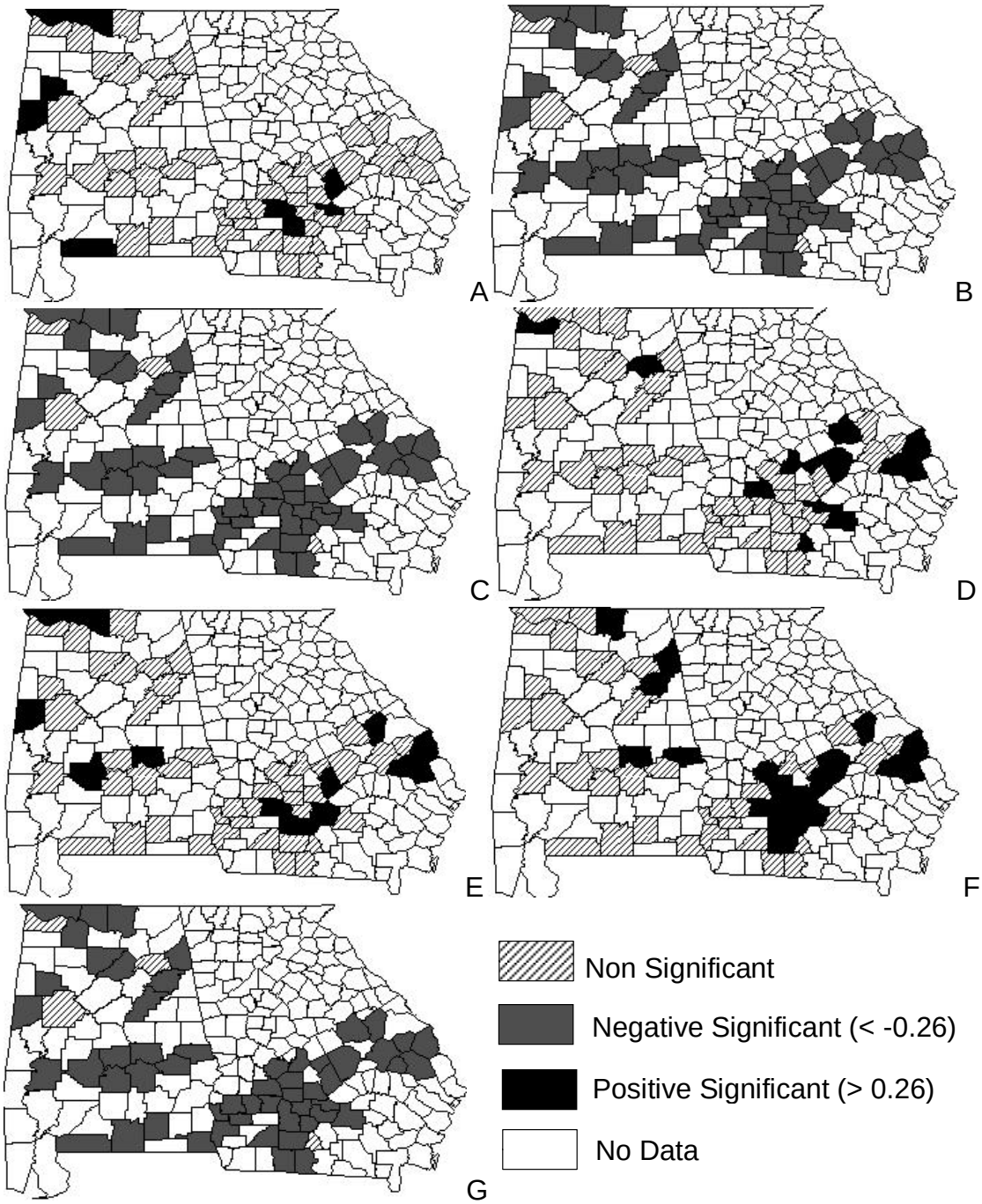


Figure 5-1. Correlations of climate indices with temperature during the cotton growing season. A) AMM-July temperature B) NOI-July temperature C) NP-July temperature D) QBO-July temperature E) TNA-July temperature F) ONI-July temperature G) TNH-July temperature. Correlations greater or less than 0.26 and -0.26 were significant at $p < 0.10$.

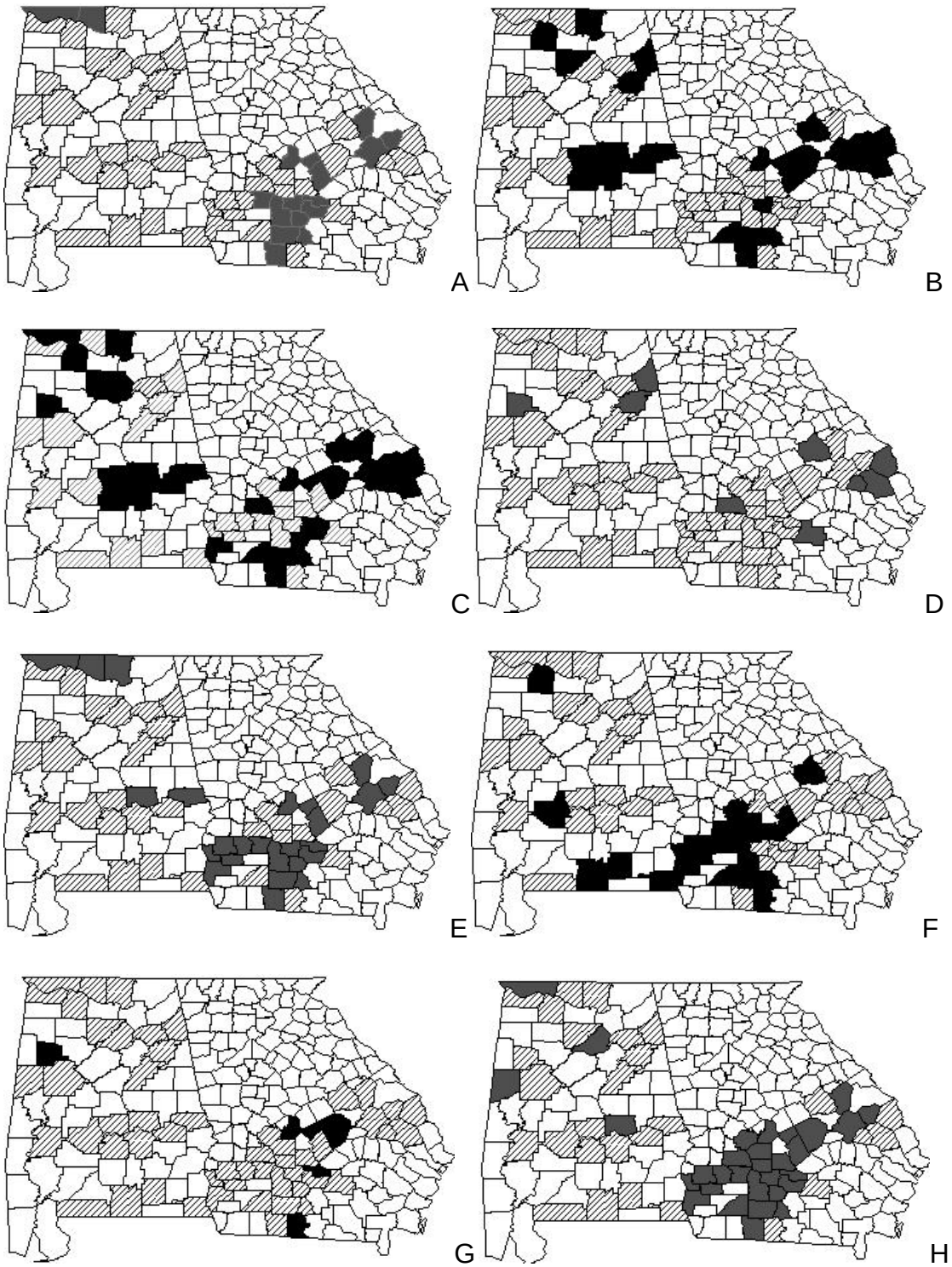


Figure 5-2 Correlations of climate indices with rainfall during the cotton growing season. A) AMM-June rainfall B) NOI-July rainfall C) NP-July rainfall D) QBO-July rainfall E) TNA-June rainfall F) TNA-September rainfall G) ONI-August rainfall H) TNH-September rainfall. Legends are same as Figure 3-1. Correlations greater or less than 0.26 and -0.26 were significant at $p < 0.10$.

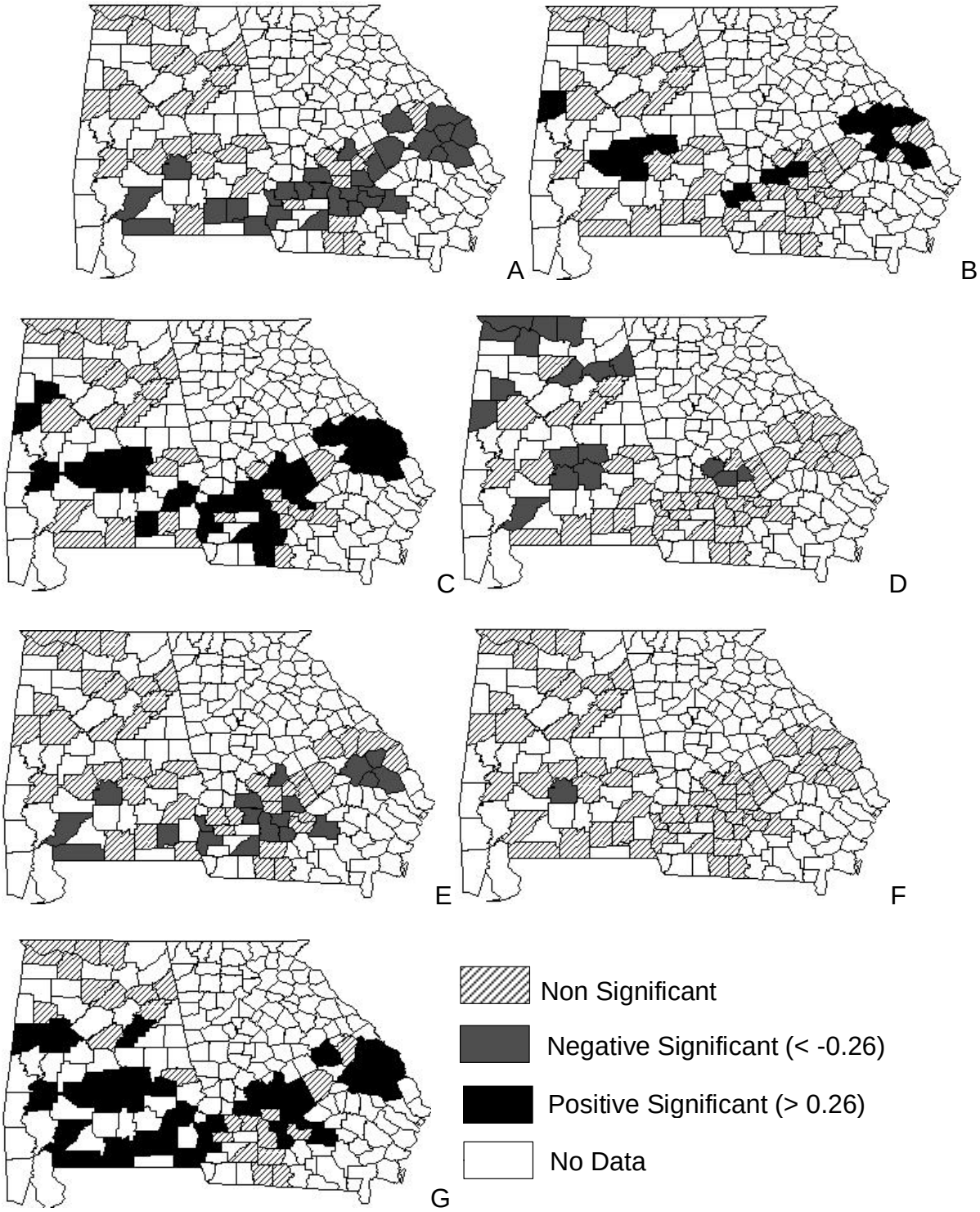


Figure 5-3. Correlations climate indices with cotton yield. A) AMM B) NOI C) NP D) QBO E) TNA F) ONI G) TNH. Correlations greater or less than 0.26 and -0.26 were significant at $p < 0.10$.

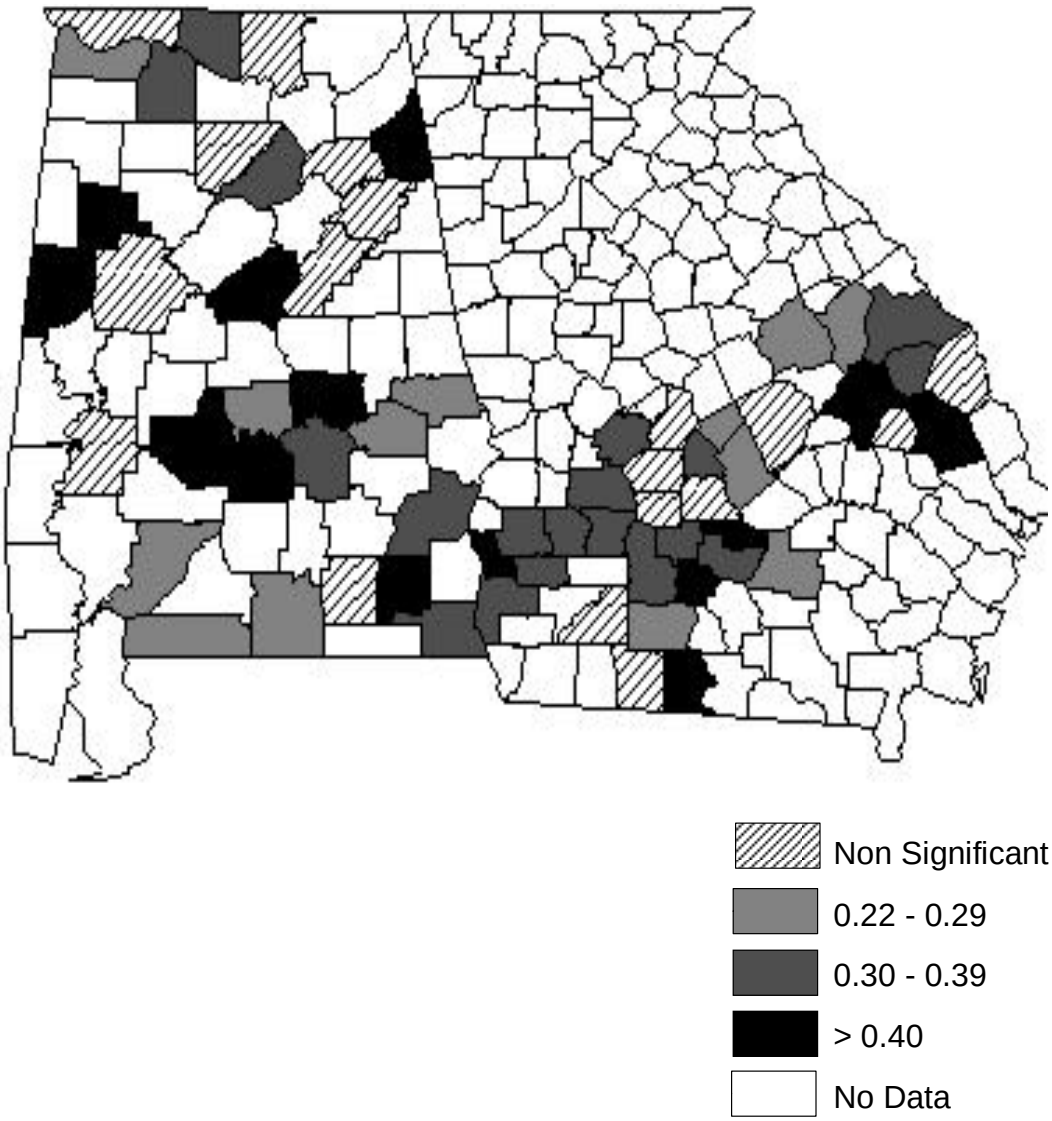


Figure 5-4. Correlations of historic cotton residuals with cross validated cotton yield residuals using the principal component regression of January and February climate indices. Correlations greater than 0.22 are significant at $p < 0.10$ (N = 57).

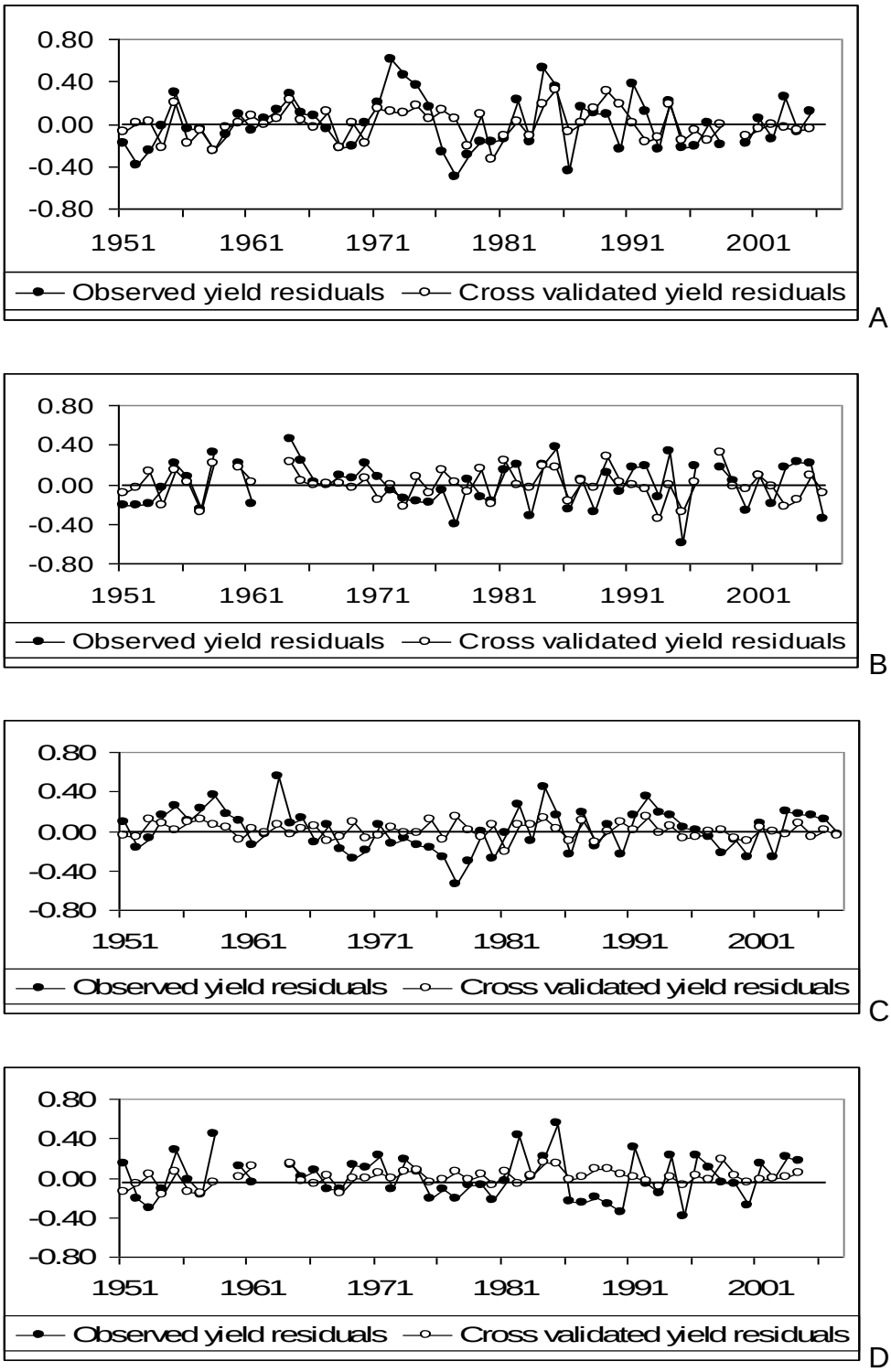


Figure 5-5. Time series comparison between observed and cross validated cotton yield residuals for four counties of Alabama and Georgia that showed the maximum (A and B) and minimum correlations (C and D). A) Bulloch, B) Shelby, AL C) Colquitt, GA, and D) Lee, AL.

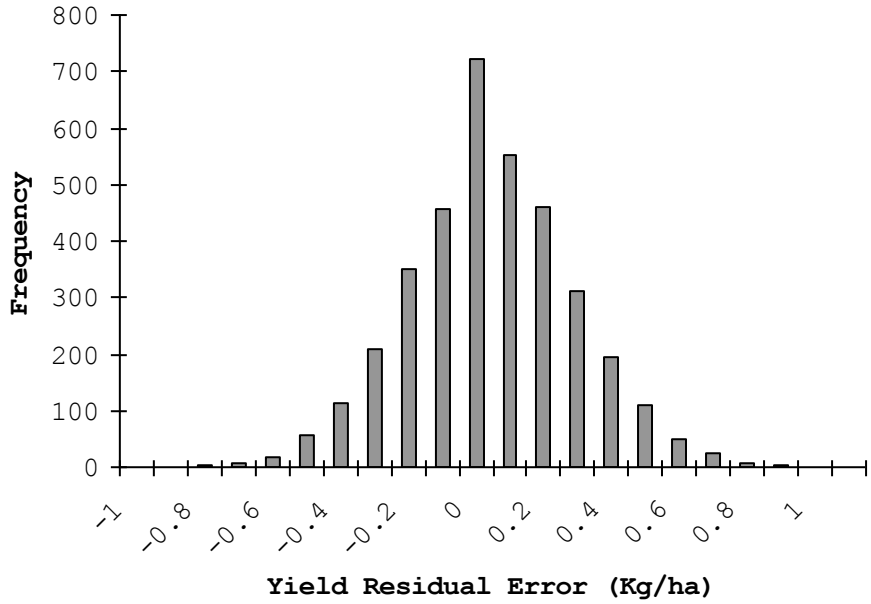


Figure 5-6. Histogram of residual errors across the entire cross validated county cotton yields. The errors were normally distributed with $\mu = -0.0001$ and $\sigma = 0.24$

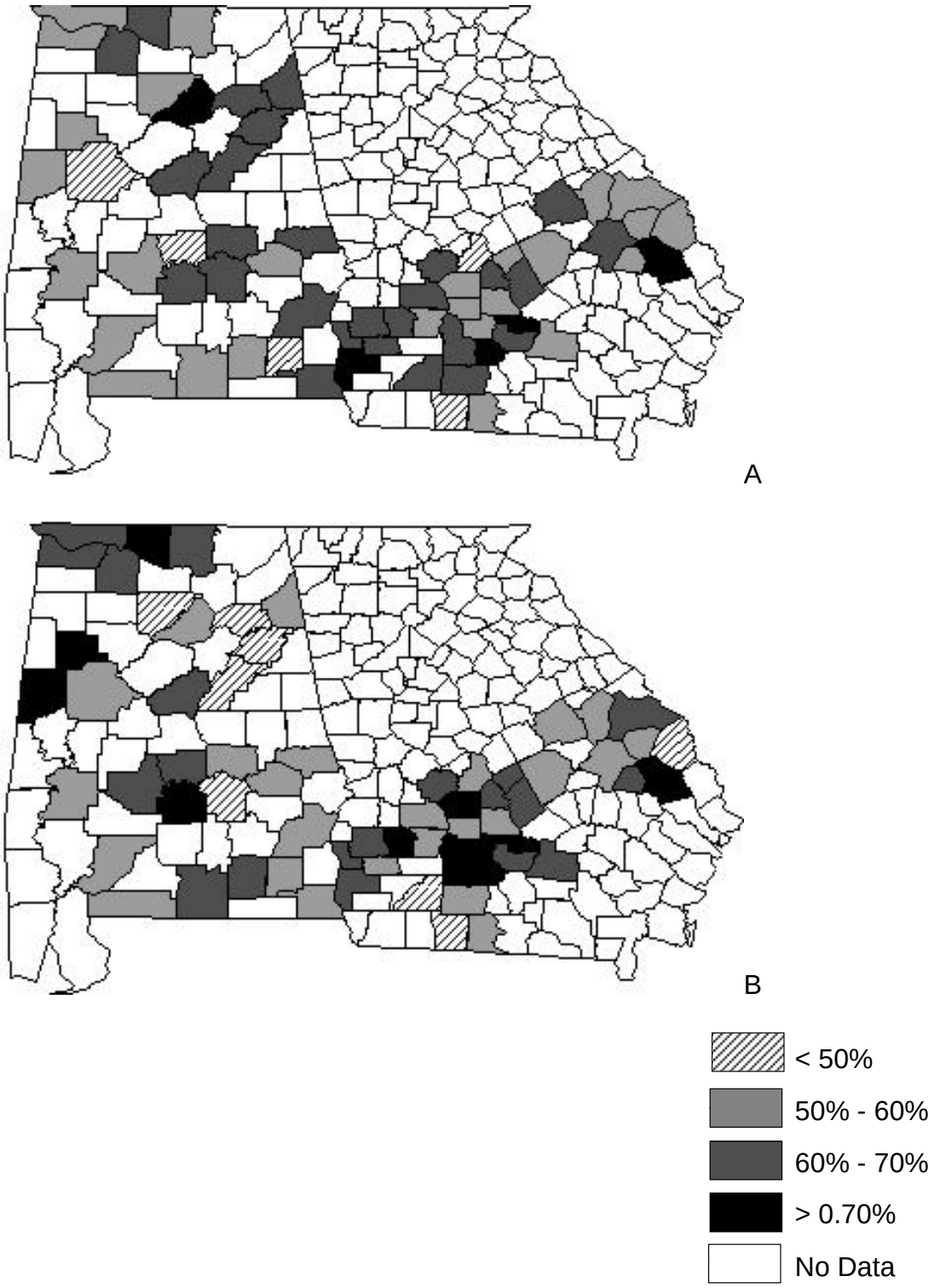


Figure 5-7. Probability of detecting county level cotton yields using cross validated cotton forecasts for two categories. A) Probability of detecting above average cotton yield. B) Probability of detecting below average cotton yield.

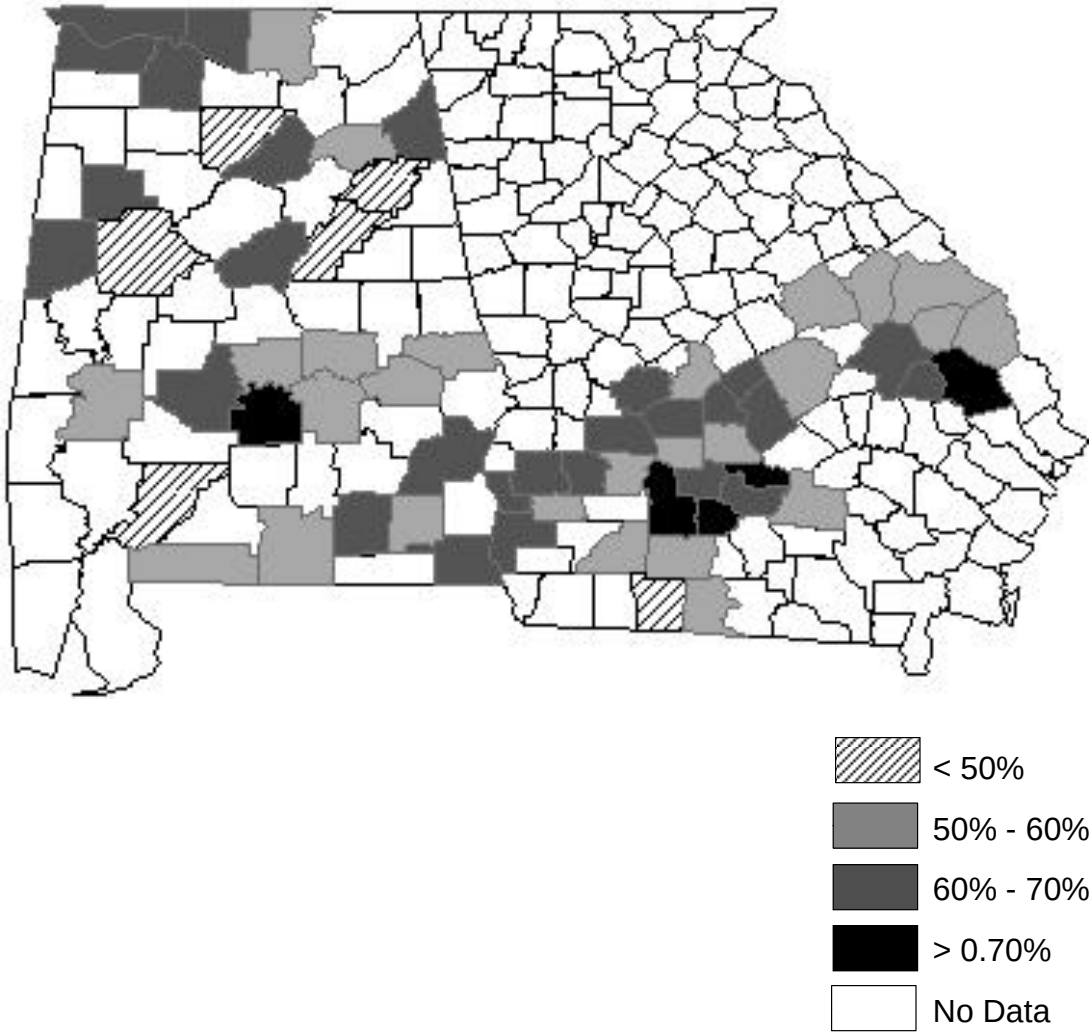


Figure 5-8. Percent correct cross validated cotton yield forecasts based on principal components regression model of climate indices and contingency table.

Table 5-1. A 2x2 contingency table for categorical cotton yields.

Observed Yield	Forecasted Yield Below	Above	Total
Below	A	B	C
Above	D	E	F
Total	G	H	I

Table 5-2. Pearson's correlations and MSE for cross validated cotton yields for the counties in Georgia and Alabama. The ***, **, and * represents significance at 0.01, 0.05, and 0.1 probability levels. The "ns" represents non-significant correlations.

Georgia			Alabama		
Counties	Correlations	MSE	Counties	Correlations	MSE
BEN HILL	0.42***	0.07	AUTAUGA	0.47***	0.07
BLECKLEY	0.23*	0.07	BARBOUR	0.30**	0.08
BROOKS	0.42***	0.04	BLOUNT	0.38***	0.03
BULLOCH	0.46***	0.06	CALHOUN	ns	0.09
BURKE	0.30**	0.06	CHEROKEE	0.44***	0.06
CALHOUN	0.33**	0.03	COFFEE	ns	0.11
CANDLER	ns	0.08	COLBERT	0.25*	0.06
CLAY	0.48***	0.04	COVINGTON	0.29**	0.07
COFFEE	0.28**	0.10	CULLMAN	ns	0.07
COLQUITT	0.22*	0.04	DALE	0.45***	0.09
CRISP	ns	0.05	DALLAS	0.46***	0.06
DODGE	0.28**	0.06	ELMORE	0.42***	0.10
DOOLY	ns	0.09	ESCAMBIA	0.25*	0.07
EARLY	0.37***	0.05	ETOWAH	ns	0.06
EMANUEL	0.42***	0.07	FAYETTE	0.46***	0.10
HOUSTON	ns	0.06	HOUSTON	0.35**	0.09
IRWIN	0.35***	0.05	LAUDERDALE	ns	0.08
JEFFERSON	0.29**	0.08	LAWRENCE	0.30**	0.08
JENKINS	0.31**	0.08	LEE	0.23*	0.06
LAURENS	ns	0.08	LIMESTONE	0.37***	0.08
LEE	0.33**	0.07	LOWNDES	0.49***	0.07
MACON	0.37***	0.06	MACON	0.28**	0.05
MITCHELL	ns	0.06	MADISON	ns	0.09
PULASKI	0.35***	0.06	MARENGO	ns	0.09
RANDOLPH	0.37***	0.05	MONROE	0.29**	0.05
SCREVEN	ns	0.08	MONTGOMERY	0.30**	0.07
SUMTER	0.34***	0.07	PICKENS	0.44***	0.05
TERRELL	0.31**	0.04	SHELBY	0.50***	0.05
THOMAS	ns	0.05	TALLADEGA	ns	0.04
TIFT	0.46***	0.04	TUSCALOOSA	ns	0.06
TURNER	0.40***	0.06			
WASHINGTON	0.29**	0.07			
WILCOX	ns	0.06			
WORTH	0.30**	0.05			

Table 5-3. Significant principal components (PCs) of principal component regression models for cotton producing counties of Georgia and Alabama

Georgia		Alabama	
Counties	Significant PCs	Counties	Significant PCs
BEN HILL	3,5,12,13,14	AUTAUGA	2,4,5,9,11
BLECKLEY	11,12	BARBOUR	8,11
BROOKS	5,8,13	BLOUNT	12
BULLOCH	1,9,13,14	CALHOUN	1
BURKE	1,6,11	CHEROKEE	2,3,5,11,12
CALHOUN	5,8,11	COFFEE	3,7
CANDLER	1,9	COLBERT	3,6,11
CLAY	1,5,6,11	COVINGTON	3,5
COFFEE	7,9,13	CULLMAN	2
COLQUITT	3,14	DALE	1,5,7,8,11,14
CRISP	3,6	DALLAS	4,5,9,11
DODGE	5,11,12	ELMORE	1,2,5,9,11
DOOLY	11	ESCAMBIA	1
EARLY	1,3,8,14	ETOWAH	2,3
EMANUEL	1,5,9,12,13	FAYETTE	3,8,9,11
HOUSTON	1,11	HOUSTON	3,4,5,7,11
IRWIN	9,12	LAUDERDALE	3
JEFFERSON	1,7,11,13	LAWRENCE	3,6,11
JENKINS	1,2,13	LEE	5,11
LAURENS	12	LIMESTONE	3
LEE	6,9,13	LOWNDES	1,2,3,4,9,11
MACON	1,3,11,14	MACON	5,11,13
MITCHELL	1	MADISON	3
PULASKI	3,6,11	MARENGO	6
RANDOLPH	1,5,6,11	MONROE	6
SCREVEN	12	MONTGOMERY	2,7,8
SUMTER	1,6,11,13	PICKENS	1,3,6,9,11
TERRELL	5,11,12	SHELBY	2,4,5,9,10,11
THOMAS	1	TALLADEGA	1
TIFT	2,8,11,12,14	TUSCALOOSA	6
TURNER	1,6,8,11,12,14		
WASHINGTON	6,11		
WILCOX	5,6,14		
WORTH	3,5,12,14		

Table 5-4. Loadings of principal components (PCs) of climate indices

Indices	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14
ONI_Jan	0.73	0.44	0.30	0.13	-0.20	-0.32	-0.03	0.09	0.03	-0.01	0.07	0.02	0.00	0.07
ONI_Feb	0.74	0.45	0.33	0.10	-0.17	-0.32	0.00	0.05	0.01	-0.02	0.07	-0.01	0.01	-0.07
AMM_Jan	0.30	-0.89	-0.15	0.04	-0.02	0.18	0.12	-0.07	0.02	-0.11	0.15	0.03	0.07	0.00
AMM_Feb	0.56	-0.77	-0.08	0.03	-0.02	0.10	0.08	0.01	-0.06	0.23	0.09	-0.02	-0.06	0.00
NOI_Jan	-0.72	-0.40	0.04	0.28	-0.22	-0.13	-0.06	0.23	-0.33	-0.05	0.01	0.01	0.00	0.00
NOI_Feb	-0.56	-0.45	-0.16	-0.40	-0.36	-0.15	-0.03	0.31	0.23	0.02	0.01	-0.01	0.00	0.00
NP_Jan	-0.46	-0.25	0.00	0.54	0.55	-0.25	0.15	0.14	0.17	0.02	0.01	0.00	0.00	0.00
NP_Feb	-0.55	-0.37	0.15	-0.28	-0.04	-0.52	0.33	-0.29	-0.06	0.02	0.00	0.00	0.00	0.00
QBO_Jan	0.15	0.28	-0.91	0.11	-0.07	-0.15	-0.01	-0.04	-0.04	-0.01	0.02	-0.14	0.02	0.01
QBO_Feb	0.16	0.23	-0.92	0.09	-0.11	-0.15	0.04	-0.02	0.01	0.02	-0.02	0.14	-0.02	-0.01
TNA_Jan	0.61	-0.72	0.00	0.14	-0.15	-0.03	0.07	-0.04	0.07	-0.23	-0.09	-0.02	-0.06	0.00
TNA_Feb	0.80	-0.52	0.05	0.12	-0.10	-0.08	0.08	0.07	-0.02	0.13	-0.16	0.00	0.06	0.00
TNH_Jan	-0.26	0.54	0.08	0.21	-0.35	0.30	0.61	0.05	0.03	0.01	0.00	-0.01	0.00	0.00
TNH_Feb	-0.58	-0.17	0.14	0.48	-0.46	0.02	-0.30	-0.24	0.14	0.07	0.00	0.00	0.00	0.00

Table 5-5. Skills of categorical cross validated cotton yield forecasts for counties of Georgia and Alabama.

Counties	Georgia % correct	POD Above	POD Below	Counties	Alabama % correct	POD Above	POD Below
BEN HILL***	71.93	0.72	0.71	AUTAUGA*	55.77	0.48	0.64
BLECKLEY**	61.40	0.57	0.66	BARBOUR**	60.78	0.67	0.57
BROOKS	56.14	0.56	0.56	BLOUNT***	64.58	0.71	0.58
BULLOCH***	73.21	0.70	0.76	CALHOUN	38.46	0.64	0.08
BURKE*	59.65	0.57	0.62	CHEROKEE**	61.11	0.65	0.57
CALHOUN	57.89	0.62	0.55	COFFEE**	61.11	0.58	0.64
CANDLER**	61.40	0.57	0.66	COLBERT**	61.11	0.58	0.63
CLAY***	68.52	0.69	0.68	COVINGTON**	59.62	0.54	0.64
COFFEE	56.36	0.53	0.62	CULLMAN	47.92	0.56	0.39
COLQUITT*	59.65	0.62	0.57	DALE	53.19	0.48	0.58
CRISP*	57.89	0.57	0.59	DALLAS***	64.15	0.58	0.69
DODGE**	63.16	0.62	0.64	ELMORE**	55.56	0.61	0.52
DOOLY**	64.91	0.57	0.72	ESCAMBIA***	59.26	0.59	0.59
EARLY***	68.42	0.71	0.66	ETOWAH*	51.85	0.62	0.43
EMANUEL***	62.50	0.67	0.59	FAYETTE**	64.58	0.57	0.70
HOUSTON	52.73	0.48	0.57	HOUSTON**	61.54	0.69	0.54
IRWIN***	68.42	0.68	0.69	LAUDERDALE**	61.11	0.56	0.66
JEFFERSON	56.14	0.57	0.56	LAWRENCE**	64.81	0.60	0.69
JENKINS	51.85	0.52	0.52	LEE**	58.82	0.61	0.57
LAURENS	54.39	0.59	0.50	LIMESTONE***	66.67	0.63	0.70
LEE	58.18	0.59	0.58	LOWNDES***	70.83	0.68	0.74
MACON***	66.07	0.68	0.64	MACON	53.85	0.58	0.50
MITCHELL	54.39	0.65	0.39	MADISON*	59.26	0.54	0.65
PULASKI**	63.16	0.62	0.64	MARENGO*	55.32	0.59	0.52
RANDOLPH**	63.64	0.62	0.65	MONROE	50.00	0.50	0.50
SCREVEN	52.63	0.59	0.46	MONTGOMERY	54.35	0.63	0.48
SUMTER**	61.40	0.64	0.59	PICKENS**	63.83	0.58	0.71
TERRELL***	66.67	0.62	0.71	SHELBY**	63.46	0.61	0.67
THOMAS	28.07	0.23	0.32	TALLADEGA	48.98	0.64	0.33
TIFT***	71.93	0.75	0.70	TUSCALOOSA***	50.00	0.44	0.56
TURNER***	66.67	0.59	0.75				
WASHINGTON**	59.26	0.62	0.57				
WILCOX	54.39	0.52	0.57				
WORTH***	71.93	0.69	0.75				

CHAPTER 6
CONCLUSIONS AND FUTURE WORK

The main focus of this dissertation was to use climate forecasts and climate indices to forecast cotton yield for the southeastern United States using the CROPGRO-Cotton model and an empirical model. Overall, the use of climate information provided significant skill in forecasting cotton yield for the southeastern United States.

In order to achieve the overall research question presented in Chapter 1, the dissertation research was organized into four specific objectives under four main chapters; global sensitivity analysis of CROPGRO-Cotton model (Chapter 2), parameter estimation and uncertainty analysis (Chapter 3), In-season updates of cotton yield forecasts using the CROPGRO-Cotton model (Chapter 4), and cotton yield forecasting for the southeastern United States using climate indices (Chapter 5).

The global sensitivity analysis results improved our understanding of how sensitive the CROPGRO-Cotton model is to the selected parameters over the range of parameter uncertainties. The specific leaf area (SLAVR) followed by extinction coefficient (KCAN), and fraction of daily assimilates allocated to seed (XFRT) were important model parameters that influenced the simulated cotton yield. The duration between emergence and flowering (EM-FL), and first seed to physiological maturity (SD-PM) parameters were most important parameters for physiological maturity. Results also showed that global sensitivity analysis was a better method than local sensitivity analysis due to the fact that local sensitivity analysis did not take into account the interactions among the parameters.

Results from global sensitivity analysis were utilized to estimate parameters and perform an uncertainty analysis in Chapter 3. The parameter estimates obtained by the

generalized likelihood uncertainty estimation technique represented an improvement in the parameters previously available in DSSAT for DP-555 cotton cultivar. The output uncertainty confidence intervals at 95% limit covered approximately 80% of the measurements. This study also demonstrated an efficient prediction of uncertainties in model parameters and outputs using the widely accepted GLUE technique. There was good overall agreement of the CROPGRO-Cotton model with the field measurements using the estimated parameters.

The parameter estimation and uncertainties of the CROPGRO-Cotton model in chapter 3 provided the basis in using it for cotton yield forecasting. In-season updating the CROPGRO-cotton model with observed weather data along with the climatology improved the accuracy of the cotton yield forecasts over time. The reduction in the residual errors and standard deviations were statistically significant with in-season updates. Approximately 90% of the cotton yield forecasts showed reduction in standard deviations and 56% of the cotton yield forecasts showed reduction in residual errors among the 55 years tested. In general, three ENSO indices among them selves and the comparison between climatology based and ENSO tailored cotton yield forecasts did not show statistically significant differences in the standard deviations and residual errors of forecasted cotton yield.

As an alternative of using the crop model, cotton yield forecasts were also evaluated using historical county yield and climate data in an empirical model in this dissertation. Chapter 5 was mainly focused on forecasting county cotton yield using climate indices. The empirical principal component regression models of climate indices provided significant skills in forecasting cotton yield for the southeastern United States.

In general, with a lead time of approximately 2 months before the typical planting period in the southeastern United States in May, about 77% of the counties in Georgia and 70% of the counties in Alabama showed statistically significant correlations between observed and forecast cotton yields. The MSE between observed and cross validated cotton yield forecasts were in the range of 0.03-0.11. In addition to that, the skills of categorical cotton yield forecasts were evaluated with a contingency table. 94% of the counties showed the categorical cotton yield forecast obtained at a lead time of approximately two months before planting (In February) was correct more than 50% of the time.

In general, results showed potential for significant skill in using climate forecasts to forecast cotton yield for the southeastern United States. Improvements to the current forecasts can be made as and when climate forecasts are improved. However, in order to use these forecasts in decision making, users must integrate their perceptions of forecast uncertainty in the context of their goals, constraints, and risk tolerance as they manage their agricultural production systems (Jones et al., 2003).

LIST OF REFERENCES

Ahmed, J., C.H.M. Van Bavel, and E.A. Hiller. 1976. Optimizing of crop irrigation strategy under a stochastic weather regime: a simulation study. Water Resource Res. 12(6): 1241-1247.

Ajami, N.K., Q. Duan, and S. Sorooshian. 2007. An integrated hydrologic Bayesian multi-model combination framework: confronting input, parameter and model structural uncertainty. *Water Resource Res*. 43:W01403. doi:10.1029/2005WR004745.

Alexandrov, V.A., and G. Hoogenboom. 2001. Climate variation and crop production in Georgia, USA, during the twentieth century. *Clim Res.* 17:33-43.

Baigorria, G.A., J.W. Hansen, N. Ward, J.W. Jones, and J.J. O'Brien. 2008. Assessing predictability of cotton yields in the southeastern USA based on atmospheric circulation and surface temperatures. *J of Applied Meteorology and Climatology.* 47: 76-91.

Baigorria, G.A., J.W. Jones, D.W. Shin, A. Mishra, and J.J. O'Brien. 2007. Assessing uncertainties in crop model simulations using daily bias-corrected regional circulation model outputs. *Clim Res.* 34:211-222.

Baigorria, G.A., M. Chelliah, K.C. Mo, C.C. Romero, J.W. Jones, J.J. O'Brian, and R.W. Higgins. 2010. Agron. J. 102: 187-196.

Bange, M. P., and S. P. Milroy. 2000. Timing of crop maturity in cotton impact of dry matter production and partitioning. *Field Crops Res*. 68: 143-155.

Bange, M. P., and S. P. Milroy. 2004. Growth and dry matter partitioning of diverse cotton genotypes. *Field Crops Res.* 87: 73-87.

Bannayan, M., and N.M.J. Crout. 1999. A stochastic modelling approach for real-time forecasting of winter wheat yield. *Field crop res.* 62: 85-95.

Barnston, A.G., R.E. Livezey, and M.S. Halpert. 1991. Modulation of Southern Oscillation-North Hemisphere mid-winter climate relationships by the QBO. J. of Climate. 4: 203-217.

Barreiro, M., P. Chang, L. Ji, R. Saravanan, and A. Giannini. 2005. Dynamical elements of predicting boreal spring tropical Atlantic sea surface temperatures. *Dyn. Atmos. Oceans*. 39: 61-85.

Bell, G. D. and J. E. Janowiak. 1994. Atmospheric circulation during the Midwest floods of 1993. *Bull. Amer. Met. Soc*. 76: 681-696.

Beven, K.J. 2001. How far can we go in distributed hydrological modelling*?. Hydrology and Earth System Sciences*. 5(1): 1-12.

Beven K. 2006. A manifesto for the equifinality thesis. *J Hydrol*. 320(1):18–36.

Beven, K., and A. Binley. 1992. The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process*. 6: 279-298.

Beven, K.J., and J. Freer. 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems. *J. Hydrology*. 249: 11–29.

Blasone, R. S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski. 2008. Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling. *Adv. Water Resour*. 31: 630–648.

Boote, K. J., J. W. Jones, and G. Hoogenboom. 1998. Simulation of crop growth: CROPGRO Model. *In*: Peart R.M., Curry R.B., eds. Agricultural systems modeling and simulation. New York: Marcel Dekker, 651-692.

Boote, K. J., J. W. Jones, G. Hoogenboom, and N. B. Pickering. 1998. The CROPGRO model for grain legumes. *In:* Tsuji, G. Y., Hoogenboom, G., and Thornton, P. K., eds. Understanding Options for Agricultural Production. Dordrecht, the Netherlands: Kluwer Academic, 99-128.

Boote K.J., J.W. Jones, W.D. Batchelor, E.D. Nafziger, and O. Myers. 2003. Genetic coefficients in the CROPGRO-Soybean model: Links to field performance and genomics. *Agron. J.* 95: 32-51.

Boote, K.J., M.I.Mínguez, and F. Sau. 2002. Adapting the CROPGRO legume model to simulate growth of faba bean. *Agron. J*. 94:743-756.

Box, G.E.P, and N.R. Draper. 1987. *Empirical model building and response surfaces*. John Wiley & Sons, New York.

Brazier, R.E., K.J. Beven, J. Freer, and J. S.Rowan. 2000. Equifinality and uncertainty in physically based soil erosion models: application of the GLUE methodology to WEPP—the water erosion prediction project—for sites in the UK and USA. *Earth Surface Processes and Landforms.* 25: 825–845.

Brisson N., B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Dürr, G. Richard, N. Beaudoin, S. Recous, X. Tayot, D. Plenet, P. Cellier, J.M. Machet, J.M. Meynard, and R. Delécolle. 1998. Stics: a generic model for the simulation of crops and their water and nitrogen balances. I. theory and parameterization applied to wheat and corn. *Agronomie*. 18:311-346

Chiang, J. C. H., and D. J. Vimont. 2004. Analogous Pacific and Atlantic meridional modes of tropical atmosphere-ocean variability. *J. Clim.* 17: 4143–4158.

Crane, T.A., Roncoli, C., Paz, J., Breuer, N., Broad, K., Ingram, K.T., and Hoogenboom, G. 2010. Forecast Skill and Farmers' Skills: Seasonal Climate Forecasts and Agricultural Risk Management in the Southeastern United States. *Weather, Climate, and Society.* 2: 44-59

De Wit, A.J., and C.A. van Diepen. 2007. Crop model data assimilation with the emsemble Kalman filter for improving regional crop yield forecasts. *Agri. and forest metero.* 146: 38-56.

Enfield, D.B. 1996. Relationships of inter-American rainfall to tropicalAtlantic and Pacific SST variability. *Geophysical Research Letters.* 23(23): 3305–3308.

Enfield, D.B., D.A. Mayer, and L. Cid Serrano. 1999. How ubiquitous is the dipole relationship in tropical Atlantic sea surface temperatures? *Journal of Geophysical Research.* 104(C4): 7841–7848.

Georgekakos, K.P., D.J. Seo, H. Gupta, J. Schaake, and M.B. Butts. 2004. Characterizing stream flow simulation uncertainty through multimodel ensembles. *J Hydrol.* 298 (1-4): 222-41.

Gilbert, R.A., K.J. Boote, and J.M. Bennett. 2002. On-farm testing of the PNUTGRO crop growth model in Florida. *Peanut Sci.* 29:58-65.

Giraldo, L. M., L. J. Lizcano, A. J. Gijsman, B. Rivera, and L. H. Franco. 1998. Adaptation of the DSSAT model for simulation of *Brachiara decumbens* production. *Pasturas Tropicales.* 20: 2-12.

Grondona, M.O., Podesta´ , G.P., Bidegain, M., Marino, M., Hordij, H., 2000. A stochastic precipitationgenerator conditioned on ENSO phase: A case study in southeastern South America. *Journal of Climate.* 13: 2973–2986.

Guerra, L.C., A.Garcia y Garcia, G. Hoogenboom, C.W. Bednarz, and J.W. Jones. 2005. Evaluation of a new model to simulate growth and development of cotton. 2005. Annual Meeting Abstracts (CD-ROM). American Society of Agronomy, Salt Lake City, Utah.

Hansen, J.W. 2005. Integrating seasonal climate prediction and agricultural models for insights into agricultural practice*. Philos. Trans. Roy. Soy. London*, 360B: 2037-2047.

Hansen, J.W., A.W. Hodges, J.W. Jones. 1998. ENSO influences on agriculture in southeastern United States. *Journal of Climate*. 11(3): 404-411.

Hartkamp, A.D., G. Hoogenboom, R. Gilbert, T. Benson, S.A. Tarawali, A. Gijsman, W. Bowen, and J.W. White. 2002. Adaptation of the CROPGRO growth model to velvet bean as a green manure cover crop: II. Model evaluation and testing. *Field Crops Research.* 78(1):27-40.

He, J., M.D. Dukes, J.W. Jones, W.D. Graham, and J. Judge. 2009. Applying GLUE for estimating CERES-Maize genetic and soil parameters for sweet corn production. *Trans. of ASABE.* 52(6): 1907-1921.

Hintze, J. (2004). NCSS and PASS. Number Cruncher Statistical Systems. Kaysville, Utah. WWW.NCSS.COM

Homma, T., and A. Saltelli. 1996. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*. 52: 1-17.

Hoogenboom, G., J.W. Jones, P.W. Wilkens, C.H. Porter, W.D. Batchelor, L.A. Hunt, K.J. Boote, U. Singh, O. Uryasev, W.T. Bowen, A.J. Gijsman, A. du Toit, J.W. White, and G.Y. Tsuji. 2004. Decision Support System for Agrotechnology Transfer Version 4.0 [CD-ROM]. University of Hawaii, Honolulu, HI.

Hornberger, G.M., and R. C. Spear. 1981. An approach to the preliminary analysis of environmental systems. *J. Environmental Management*. 12: 7-18.

Idso, S.B., J.L. Hatfield, R.D. Jackson, and R.J. Reginato. 1979. Grain yield prediction: Extending the stress-degree-day approach to ac- 1975–1995. *Remote Sens. Environ*. 8:267–272

Jagtap, S.S., J.W. Jones, P. Hildebrand, D. Letson, J.J. O'Brien, G. Podesta, D. Zierden, and F. Zazueta. 2002. Responding to stakeholder's demands for climate information: From research to application in Florida. *Agric. Syst.* 74: 415-430.

Jeuffroy, M. H., and S. Recous. 1999. Azodyn: a simple model for simulating the date of nitrogen deficiency for decision support in nitrogen fertilization. *Eur J of Agron.* 10: 129-144.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. The DSSAT cropping system model. *Eur J Agron*. 18: 235-265.

Jones, J.W., J.W. Hansen, F.S. Royce, and C.D. Messina. 2000. Potential benefits of climate forecasting in agriculture. *Agri, ecosys., and env.* 82:169-184.

Jones, J.W, G. Y. Tsuji, G. Hoogenboom, L. A. Hunt, P. K. Thornton, P. W. Wilkens, D. T. Imamura, W. T. Bowen, U. Sing. 1998**.** Decision support system for agrotechnology transfer; DSSAT v3. In: G.Y. Tsuji, G. Hoogenboom and P.K. Thornton. *Understanding options for agricultural production*. Dordrecht: Kluwer Academic, 157–177.

Jones, P. D., C. G. Kilsby, C. Harpham, V. Glenis, and A. Burton. 2009. UK ClimateProjections science report: Projections of future daily climate for the UK from the Weather Generator. University of Newcastle, UK.

Jones P.N., P.S. Carberry. 1994. A technique to develop and validate simulation models. *Agric. Syst*. 46:427-442.

Kiladis, G. N., and H. F. Diaz. 1989. Global climatic anomalies associated with extremes in the Southern Oscillation. *J. Climate.* 2: 1069-1090.

Kiniry, J.R., J.R. Williams, P.W. Gassman, and P. Debaeke. 1992. A general, process-oriented model for two competing plant species. *Trans. ASAE.* 35: 801–810.

Kossin, J. P., and D. J. Vimont. 2007. A more general framework for understanding Atlantic hurricane variability and trends. Bull*. Amer. Meteor. Soc*. 88: 1767-1781.

Kumar, V. 2000. Forecasting wheat yield using time series analysis: A case study for the Canadian Prairies. *Agronomy Journal*. 92:1047-1053

Larow, T., J. Sardonia, and D. Shin. 2005. Real time seasonal forecasts at FSU and its application to crop models. American geophysical union abstract No. A33B-06.

Lawless, C., M.A. Semenov. 2005. Assessing lead-time for predicting wheat growth using a crop simulation model. *Agri. and forest meteorology.* 135: 302-313.

Lee, D. E. and W. J. Sydeman. 2009. North Pacific climate change mediates offspring sex ratio in northern elephant seals. *Journal of Mammalogy.* 90: 1-8.

Lembke, W.D., and B.A. Jones. 1972. Selecting a method for irrigation scheduling using a simulation model. Trans. ASABE. 15(2): 284-286.

MacDonald, S. 2000. The new agricultural trade negotiations: Background and issues for the U.S. cotton sector. Cotton and Wool Situation Outlook, CWS 2000: 17-23.

Makowski, D., C. Naud, H. Monod, M. Jeuffroy, and A. Babotin. 2004. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. In Proc. Of 4th SAMO Conf. (Sensitivity Analysis of Modeling Output), 8-11. Santa Fe, USA.

Makowski, D., J. Hillier, D. Wallach, and B. Andrieu. 2006. Parameter estimation for crop models. In *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Application*, 101-103. D. Wallach, D. Makowski, and J. W. Jones, eds. Amsterdam, The Netherlands: Elsevier.

Martinez, C.J, G.A. Baigorria, J.W. Jones. 2009. Use of climate indices to predict corn yields in southeast USA. International Journal of Climatology 29(11): 1680-1691. doi:10.1002/joc.1817.

Massy, W.F. 1965. Principal component regression in exploratory statistical research. *J. of American Statistical Association.* 60(309): 234-256.

Mennis, J. 2001. Exploring relationships between ENSO and vegetation vigour in the south-east USA using AVHRR data. *International Journal of Remote Sensing*. 22(16): 3077-3092.

Messina, C.D., J.T. Ritchie, J.W. Jones, K.J. Boote, G. Hoogenboom, and V.K. Whalen. 2004. A simulation model of cotton growth and development for CSM. Annual Meetings Abstracts (CD-ROM). American Society of Agronomy, Madison, WI.

Milroy, S.P., and M.P. Bange. 2003. Nitrogen and light response of cotton photosynthesis and implications for crop growth. *Crop Sci*. 43: 904-913.

Milroy, S.P., M.P. Bange, and V.O. Sadras. 2001. Profiles of leaf nitrogen and light in reproductive canopies of cotton (*Gossypium hirsutun*). *Ann. Bot.* 87: 325-333.

Mo, K. C., and J. E. Schemm.2008. Relationships between ENSO and drought over the southeastern United States. *Geophys. Res. Lett.*, 35: L15701, doi:10.1029/2008GL034656.

Mo, K. C., and R. E. Livezey. 1986. Tropical–extratropical geopotential height teleconnections during the Northern Hemisphere winter. *Mon. Wea. Rev.* 114: 2488–2515.

Montanari, A. 2007. What do we mean by uncertainty? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrol. Processes*. 21:841-845. doi:10.1002/hyp.6623.

Moradkhani, H., K. L. Hsu, H. Gupta, and S. Sorooshian. 2005. Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resour. Res.* 41:W05012. doi:10.1029/2004WR003604.

Mote, T.L. 1986. Influence Of ENSO on maximum, minimum, and mean temperatures in the southeast United-States. *Physical Geography.* 17(6) 497-512.

National Agricultural Statistics-United States Department of Agriculture (NASS/USDA). 2008. *Acreage*. Released June 30, 2008. Available at: http://www.nass.usda.gov/Publications/.

Nielsen, D.R., J.W. Biggar, and K.T. Erh. 1973. Spatial variability of field measured soil water properties. *Hilgardia* 42: 215-259.

O'Brien, J.J.O., T. S, Richards, A.C. Davis. 1996. The Effect Of El Niño On U.S. Landfalling Hurricanes. *Bulletin Of The American Meteorological Society*. 77(4) 773-774.

Pathak, T.B., C.W. Fraisse, J.W. Jones, C.D. Messina and G. Hoogenboom. 2007. Use of global sensitivity analysis for CROPGRO cotton mode development. *Transactions of the ASABE.* 50 (6): 2295–2302.

Pettigrew, W.T. 2002. Improved yield potential with an early planting cotton production system. *Agron. J.* 94: 997-1003.

Plackett, R.L. 1983. Karl person and the Chi-squared test. International Statestical Review. 51(1): 59-72.

Podesta´, G., D. Letson, C. Messina, F. Royce, A. R. Ferreyra, J. Jones, J. Hansen, I. Llovet, M. Grondona, and J. J. O'Brian. 2002. Use of ENSO-related climate information in agricultural decision making in Argentina. a pilot experience. *Agricultural System.* 74: 371-392.

Reddy, V. R., D. N. Baker, and H. F. Hodges. 1991. Temperature effects on cotton canopy growth, photosynthesis, and respiration. Agron. J. 83: 699-704.

Reddy, K. R., H. F. Hodges, and J. M. McKinion. 1993. Temperature effects on Pima cotton leaf growth. Agron. J. 91:851-858.

Reddy, K. R., V. R., Reddy, and H. F. Hodges. 1992. Temperature effects on early season cotton growth and development. Agron. J. 84: 229-237.

Ritchie, J.T. 1994. Classification of crop models. In Uhlir, P.F., C.J. Carter eds. 3-14. Crop modeling and related environmental data.

Ropelewski, C.F., M.S. Halpert. 1986. North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Monthly Weather Review*. 114(12): 2352-2362.

Rosenthal, W.D., and T.J. Gerick. Radiation use efficiency among cotton cultivars. 1991. Agron. J. 83: 655-658.

Royce, F., C. Fraisse, and G. Baigorria. 2009. Best ENSO classification index for summer crop yields in the southeastern USA. SECC planning meeting. Gainesville, FL.

Ruget, F., N. Brisson, R. Delecolle, and R. Faivre. 2002. Sensitivity analysis of crop simulation model, STICS, in order to choose the main parameters to be estimated. *Agronomie* 22:133-158.

Saltelli, A., S. Tarantola, and F. Campolongo. 2000. Sensitivity analysis as an ingredient of modeling. *Statistical Science*. 15: 377-395.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto. 2004. *Sensitivity Analysis in practice. A guide to assessing scientific models*. New York, N.Y.: John Wiley and Sons.

Sakamoto, C.M. 1979. The Z-index as a variable for crop yield estimalyzing a longer yield series. *Agric. Meteorol*. 19:305–313.

Scholberg, J. M. S., K. J. Boote, J. W. Jones, and B. L. McNeal. 1997. Adaptation of the CROPGRO model to simulate the growth of field-grown tomato. In: Kropff, M. J. et al. (Ed.) Systems Approaches for Sustainable Agricultural Development: Applications of Systems Approaches at the Field Level. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 133-151.

Schwing, F.B., T. Murphree, P.M. Green. 2002. The Northern Oscillation Index (NOI): a new climate index for the northeast Pacific. *Progress in Oceanography*. 53: 115-139.

Semenov, M.A. and J.R. Porter. 1995. Climatic variability and the modeling of crop yields. *Journal of Agricultural and Forest Meteorology*. 73:265-283.

Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder. 2008. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res. 44*: W00B06. doi:10.1029/2008WR006822.

Stenseth, N.C., G. Ottersen, J.W. Hurrell, A. Mysterud, M. Lima, K.S. Chan, N.G. Yoccoz, and B. Adlandsvik. 2003. Studying climate effect on ecology through the use of climate indices: the North Atlantic Oscillation, El Niño Southern Oscillaiton and beyond. Proc*. Royal Soc. Lond.* DOI 10.1098/rspb.2003.2415.

Sutter, J.M., J.H. Kalivas, P.M. Lang. 1992. Which principal components to utilize for principal component regression. J. of Chemometrics. 6(4): 217-225.

Takeshi, D., T. Tozuka, T. Yamagata. 2010. The Atlantic meridional model and its coupled variability with the Guinea Dome. J. Clim. 23: 455-475.

Thiemann, M., M. Trosset, H. Gupta, S. Sorooshian. 2001. Bayesian recursive parameter estimation for hydrological models. *Water Resour Res*. 37(10): 2521–35.

Thompson, D. W. J., and J. M. Wallace. 2001. Regional climate impacts of the Northern Hemisphere annular mode. *Science.* 293: 85-89.

Thorsen, M., J. C. Refsgaard, S. Hansen, E. Pebesma, J. B. Jensen, and S. Kleeschulte. 2001. Assessment of uncertainty in simulation of nitrate leaching to aquifers at catchment scale. *J. Hydrology*. 242: 210-227.

Tolson, B.A. and C.A. Shoemaker. 2008. Efficient prediction uncertainty approximation in the calibration of environmental simulation models. *Water Resources Research.* 44: W04411.

Trenberth, K. E., and J. W. Hurrell. 1994. Decadal atmosphere-ocean variations in the Pacific. *Clim. Dyn*. 9: 303-319.

Tsuji, G. Y., G. Hoogenboom, and P. K. Thornton. 1998. Understanding Options for Agricultural Production. Systems approach for sustainable agricultural development, Kluwer Academic Publisher, Dordrechtp, the Netherlands. 400 pp

USDA. 2009. Cotton and wool year book. Economic Research Service. Available at: http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1282

Van Oijen, M., Rougier, J., Smith, R. 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. *Tree Physiology.* 25: 915-927.

Vedwan, N., K. Broad, D. Letson, K.T. Ingram, G. Podestá, N.E. Breuer, J.W. Jones, and J.J. O'Brien. 2005. Assessment of climate information dissemination efforts by the Florida Climate Consortium. *Publication of the Southeast Climate Consortium Technical Report series*. SECC: 05-001.

Vimont, D. J., and J. P. Kossin. 2007. The Atlantic meridional mode and hurricane activity. *Geophys. Res. Lett.* 34: L07709, DOI:10.1029/2007GL029683.

Vrugt, J.A., H.V. Gupta, W. Bouten, S. Sorooshian. 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour Res.* 39(8):1201. doi:10.1029/2002WR001642.

Walker, G.K. 1989. Model for operational forecasting of western Canada wheat yield. *Agric. For. Meteorol*. 44:339–351.

Wang, X., X. He, J.R. Williams, R. C. Isaurralde, and J. D. Atwood. 2005. Sensitivity and uncertainty analyses of crop yields and soil organic carbon simulated with EPIC. *Trans ASABE* 48(3):1041-1054.

Wang, C., S. K. Lee, and D. B. Enfield. 2008. Climate response to anomalously large and small Atlantic warm pools during the summer. *J. Clim. 21*(11): 2437–2450.

Warrick, A. W. and D. R. Nielson. 1980. *Applications of Soil Physics*, ch.13, 319–344. New York: Academic Press.

Washington, R., A. Hodson, E. Isaksson, and O. Macdonald. 2000. Northern hemisphere teleconnection indices and the mass balance of Svalbard glaciers. Int. J. Climatol. 20: 473-487.

Whitehead, P.G., and P.C. Young. 1979. Water quality in river systems: Monte-Carlo analysis. *Water Resources Research.* 15: 451-459.

Wigley, T.M.L., and T. Qipu. 1983. Crop-climate modeling using spatial patterns of yield and climate. Part 1: Background and an example from Australia. *J. of Climate and Applied Meteorology*. 22(11): 1831-1841.

Wilmott, C.J. 1982. Some comments on the evaluation of model performance. Bulletin of American meteorological society. 63(11): 1309-1313.

Wright, J.R., C.L., Hanson, D. Whitmer. 1984. Using weather records with a forage production model to forecast range forage production. *J. range manage*. 37(1): 3-6.

Wright, D. L., and R. K. Sprenkel. 2006. Cotton growth and development. Available at http://edis.ifas.ufl.edu/AG235.

Xie, S.P., H. Xu, W.S. Kessler, and M. Nonaka. 2005. Air-sea interaction over the eastern Pacific warm pool: Gap winds, thermocline dome, and atmospheric convection. *J. Climate.* 18: 5-25.

Xie, Y., J. R. Kiniry, and J. R. Williams. 2003. The Almanac Model's Sensitivity to Input Variables. *Agricultural Systems*. 76:1-16.

Young, P.C. 1983. The validity and credibility of models for badly-defined systems. *In:* Beck, M.B., and van Straten, G., eds. Uncertainty and Forecasting of Water Quality. Berlin: Springer-Verlag, 69-98.

Zamora, D.S., Jose, S., Jones, J.W., Cropper, W.P. 2009. Modeling cotton production response to shading in a pecan alleycropping system using CROPGRO. *Agroforest Syst.* 76: 423-435.

BIOGRAPHICAL SKETCH

Tapan Pathak was born in a city called Vadodara and grew up in a beautiful city and a Gujarat state capitol Gandhinagar, India. He went to St.Xavier's High School. He earned his Bachelors degree in Agricultural Engineering from Gujarat Agricultural University in 2000. In January, 2001 he joined Utah State University for MS program and earned his MS degree in Irrigation Engineering in 2004. After completion of MS degree program, he joined UF for the Ph.D. in 2005 and finished his Ph.D. in 2010. He is currently working as a faculty in the school of natural resources at the University of Nebraska, Lincoln.