A REGRESSION METAMODEL TO REPLACE SWAT IN CROP YIELD PREDICTION FOR BIG CREEK WATERSHED

By

Andrew Johnson

B.A., Augustana College, 2010

A Thesis
Submitted in Partial Fulfillment of the Requirements for the
Master of Science

Department of Geography and Environmental Resources
In the Graduate School
Southern Illinois University Carbondale
August 2013

THESIS APPROVAL

A REGRESSION METAMODEL TO REPLACE SWAT IN CROP YIELD PREDICTION FOR BIG CREEK WATERSHED

By

ANDREW JOHNSON

A Thesis Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of

Geography and Environmental Resources

Approved by:

Dr. Christopher Lant, Chair

Dr. Silvia Secchi

Dr. Justin Schoof

Graduate School
Southern Illinois University Carbondale
06/25/2013

AN ABSTRACT OF THE THESIS OF

Andrew Johnson, for the Master of Science degree in Geography and Environmental Resources, presented on 25 June, 2013, at Southern Illinois University Carbondale.

TITLE: A REGRESSION METAMODEL TO REPLACE SWAT IN CROP YIELD PREDICTION FOR BIG CREEK WATERSHED

MAJOR PROFESSOR: Dr. Christopher Lant

ABSTRACT

Climate change will have a significant impact on the productivity of agricultural lands and ecosystem services in the coming decades. Variability in temperature and precipitation will alter many central U.S. watersheds. Simulation models such as the Soil and Water Assessment Tool (SWAT) offer the ability to model changes in watersheds by varying inputs. Unfortunately, SWAT requires a large number of input parameters and computation time to process the output data. Regression metamodels offer an alternative that seeks to replace the simulation model with a regression equation. This research created a linear regression metamodel to approximate SWAT in crop yield prediction. Results show that regression models can account for 45-84 percent of variance in yields for corn, soybean, alfalfa, switchgrass, and cotton in Big Creek Watershed. The coefficient of variation for each of these models ranged from 13 to 41 percent. These metamodels were able to reduce simulation time from hours to minutes. The tradeoff for utilizing metamodels is computation time versus accuracy. The results of this research indicate that the considerable reduction in computation time coupled with a moderate degree of accuracy in predicting crop yields leads to the use of

metamodels over SWAT. Regression coefficients for each metamodel can reveal how various weather and farm management techniques impact crop yields. These metamodels will be utilized by the Agent Based Model to determine how farmers will respond to future economic policies and crop prices based on a series of climate scenarios.

ACKNOWLEDGEMENTS

I would like to sincerely thank my committee: Dr. Christopher Lant, Dr. Justin Schoof, and Dr. Silvia Secchi for their support and guidance throughout this work. I would also like to thank Dr. Christopher Lant for giving me the opportunity to work on the National Science Foundation Coupled Natural Human Systems grant. I also thank other members of the project for their support and guidance. Specifically, Girmay Misgna, Dr. Blanca Perez Lapena, Dr. Steven Kraft, Dr. Seth Soman, Awoke Teshagar, Mukesh Bhattarai, and Tim Stoebner.

TABLE OF CONTENTS

Contents

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Climate change will have a significant impact on the productivity of agricultural lands and ecosystem services in the coming decades (Baker & Allen, 1993; Meehl et al., 2007). Possible global warming scenarios for the future range from 0.2°C to 0.5°C increase per decade (Meehl et al., 2007). This will lead to variability in temperature and precipitation, which will potentially alter many Central U.S. watersheds. Changes associated with the timing and intensity of weather events will lead to variation in crop yields (Porter and Semenov, 2005). Crop yields are predicted to increase under the lowest assumed increase in temperature of 2°C, but yields will decrease if temperatures increase more than 3°C. (Parry et al., 2007; Schlenker and Roberts, 2009; Lobell et al., 2012). The sensitivity of corn, soybean, alfalfa, switchgrass and cotton crops to climate change will need to be measured in order to understand the severity of impacts associated with climate change.

One approach to study the climate dynamics of a watershed is through the use of hydrologic models. These models are beneficial in that they quantify and simplify the varying processes occurring within the watershed and allow the user to predict how various factors affect the makeup of the watershed. Simulation models are parameter and computationally intensive, which often necessitates the use of metamodels over simulation models (Wu & Babcock, 1994; Galelli et al., 2010). The goal of this study is to create a metamodel for multiple crops at the hydrologic response unit (HRU) scale. These metamodels will use output data from a deterministic watershed model, Soil and Water Assessment Tool (SWAT), and input data from various scripts and equations to simplify and reduce the amount of time needed to assess land use and climatic scenarios for Big Creek Watershed. Output data from the metamodel will be used to construct an agent-based model to determine how farmers will respond to future economic policies and crop prices based on a series of climate change scenarios. This research is a part of the Coupled Natural-Human Systems (CNH) grant funded by the National Science Foundation, which has one of the goals to model Central U.S. watersheds to determine impacts on the landscape associated with climate change. The research performed in this study will serve as part of the foundation for modeling on the HRU scale. Having metamodels for each watershed in the selected Central U.S. watersheds will greatly decrease computation time and assist in determining how watersheds will be affected as a result of climate change.

1.1 Study Area

This research created a series of crop-specific metamodels for Big Creek Watershed, located in Southern Illinois with a diverse range of land cover and land uses. Big Creek watershed has an area of 133 km^2 and is part of the larger Cache River Watershed (Figure 1.1). Mean annual precipitation is 1220 mm and mean yearly temperature is 14.2° Celsius (C) with the coldest month being January at 1°C and hottest July with 26.2°C (NOAA 2013). The predominant farmed crops are corn and soybeans, but the dominant land use types within the watershed are pastureland and forestland. Big Creek lies near the Shawnee National Forest, and, has land that is varied topographically, ranging from a low elevation of 50 meters to a high elevation of 192 meters, and contains many areas with highly erodible soil. Up to 70 percent of sediment loads in the Cache River can be traced back to Big Creek, and, therefore, this watershed has a significant impact on the Cache, which is an ecologically sensitive river that is part of the Cypress Creek National Wildlife Refuge (Guetersloh, 2001). This refuge contains an ecologically diverse swampland unique in Illinois, and sedimentation can have a significant impact on water quality and wildlife habitat within the swampland.

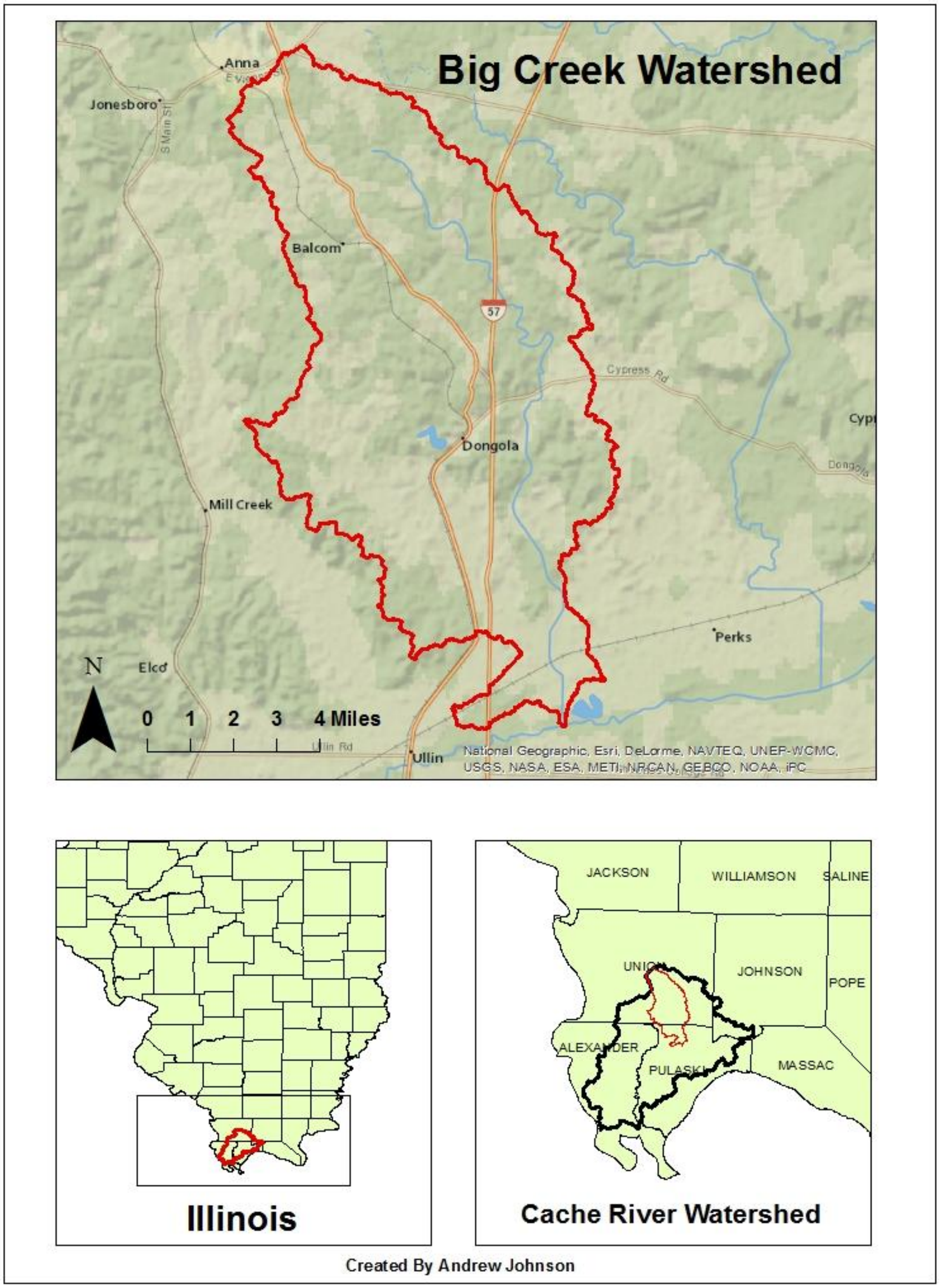


Figure 1.1 Big Creek Watershed

1.2 Research Questions

The goal of this research is to circumvent the Soil and Water Assessment Tool in crop yield prediction in Big Creek Watershed. There are three specific research questions that this thesis seeks to investigate.

1) What factors can significantly explain crop yields in Big Creek Watershed?
2) How accurately can a statistical equation such as regression reproduce the results of a sophisticated process-based watershed model such as SWAT?
3) How will crop yields in Big Creek Watershed respond to climate change scenarios?

The main need to replace SWAT is linked to computational and parameterization efficiency. In order to run all the climate scenarios in a timely manner, the metamodel is needed to quickly derive the output of crop yield without having to run the simulation through SWAT. SWAT also requires a large number of parameters to perform its calculations. Efficiency can be increased by having a small number of statistically significant parameters in a regression-based metamodel. The key is to measure the sensitivity of each crop to changes in climate and measure those changes with the variables in the regression model. These are some of the key reasons why metamodels are used for this research.

1.3 Justification

This research is part of a Coupled Natural-Human Systems (CNH) grant funded by the National Science Foundation. There is limited time to process the large amount of data associated with this project so metamodels are needed. This research represents a critical part of the project by decreasing computation time by replacing SWAT with a metamodel. Figure 1.2 represents a conceptual framework for the entire CNH project and this research will replace SWAT calculations in the watershed environment box. Once this part of the project is complete, then the output from the metamodel will be used to construct an Agent- Based Model to determine how farmers will react to changes in government policies and crop prices, which will vary in the coming decades because of global climate change.

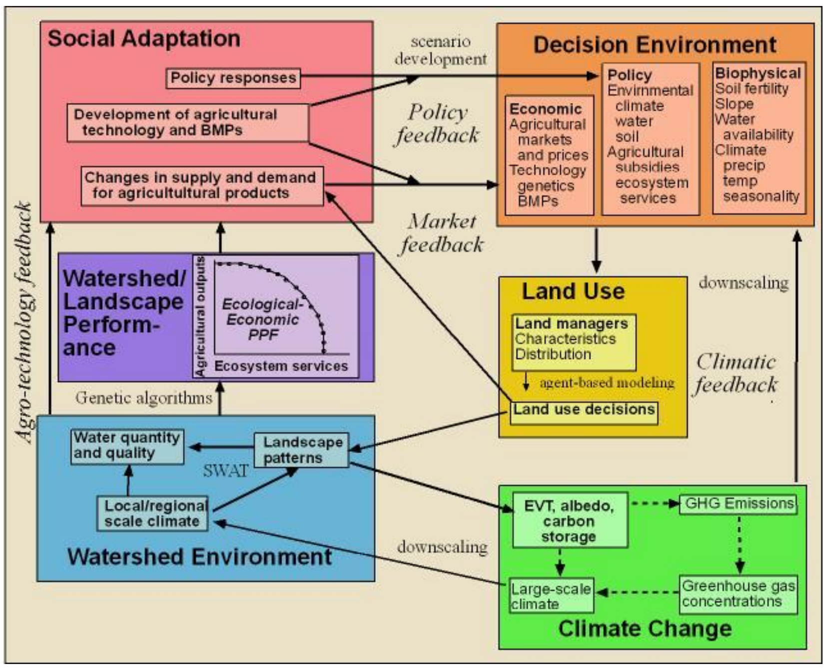


Figure 1.2 CNH Project Conceptualization (Lant. et al. 2009)

Two main sections will make up the literature review for this research. The first section will discuss the main factors that affect crop yields in Big Creek Watershed. Determining possible variables for the metamodel will be needed through a combination of a priori information found in the literature and variable selection during model creation. A thoughtful review of the literature will help lead to a model that maximizes explained variance of crop yield, while at the same time remaining parsimonious. The second section will detail the capabilities of the simulation model SWAT and how it has come to be widely used. Then the discussion will turn to metamodels and detail why and how they are used for various fields of research.

CHAPTER 2

LITERATURE REVIEW

2.1 Agricultural Landscape

2.1.1 Agricultural Management

There are a variety of factors that are vital for both crop growth and maximizing crop yields. Factors such as weather, soil properties and farm management techniques all play a role. In order to achieve high crop yields there are certain farm management practices that can be implemented. Farm management represents tillage choices, fertilizer applications, and crop rotation schemes. Altering any one of these management options can affect the yields for a given crop.

Nitrogen fertilizer represents one of the major factors contributing to crop yields for certain crops such as corn and cotton that are not nitrogen-fixers (Reidell et al., 2009). Fertilizer amounts can vary depending on the type of crop and if the crop is grown in continuous rotation or rotated with other crops. Corn grown in continuous rotation requires more nitrogen over corn-soybean rotations due to the lack of nitrogen that is in the soil for the crop to use. Varying the amount of nitrogen fertilizer also plays a key role in crop yields with high rates of fertilizer leading to significantly higher yields for corn crops (Reidell et al., 2009; Coulter et al., 2011). Farmers play a key role in the management choices they make such as rotation possibilities and fertilizer amounts. The choices of these farmers can ultimately impact crop yields and the health of streams and lakes in the watershed. Many Midwestern watersheds that are under intense monoculture farming techniques can significantly affect the water quality of the streams and lakes due to the amount of nitrogen and phosphorous that runoff from the farm field (Guetersloh, 2001). In the Midwest, corn and soybean are dominant crops. Around 43 percent of acres planted with corn are treated with 25 percent more fertilizer than the crop actually needs, though this has decreased from 59 percent in 1996 (USDA, 2013). Not only does the amount of fertilizer applied to a farm field make a difference, but the tillage regime also plays a part in not only the crop yield but also the transfer of the nitrogen into streams and lakes.

Tillage practices can have a significant effect on crop yields. Three main types of tillage practices that farmers employ are no-till, conservation tillage, and conventional tillage. In order for a farm plot to be classified as conservation tillage, there needs to be at least thirty percent crop residue left on the field following harvest. No-till indicates that one hundred percent of crop residue was left on the field following harvest, and conventional tillage indicates that less than fifteen percent residue was left on the field (Horowitz et al. 2010). Conventional tillage has the greatest impact on the landscape because leaving the soil tilled can lead to excessive rates of erosion and sediment loss during intense storm events. This can lead to a loss of key soil nutrients and impair the streams, rivers, and lakes with sediments along with nitrogen and phosphorous fertilizers (Horowitz et al. 2010). These benefits along with subsidies from the federal government has led farmers to increasingly use conservation tillage and no-till techniques, which can significantly reduce erosion and act as a carbon sink (West & Marland, 2002; Yadev et al., 2007). Edwards et al., (1988) studied the effect of tillage and crop rotation on corn, soybean, and wheat crops. Their research indicated that soybean yields were higher when no-till and conservation practices were employed as opposed to conventional tillage. Conservation and no-till practices helped maintain soil moisture and soil organic matter, which have a net positive impact on yields. Around 35 percent of cropland planted in the U.S. associated with eight major crops (corn, soybean, oat, wheat, barley, rice, sorghum and cotton) employed no till practices (USDA, 2013). Around 45 percent of planted soybean cropland employed no-till while only 25 percent did for corn cropland (USDA, 2013,). Not only does tillage have an impact on crop yields but use of rotation also impacts crop yield.

Crop rotation is another management technique that can impact agricultural yields. In the Midwest, corn and soybeans are rotated to cycle nitrogen and break pest cycles. Employing crop rotation benefits the environment because it improves soil properties such as bulk density, organic matter, and available water capacity (Grover et al., 2009). Employing crop rotation techniques has led to a boost in yields for various crops. Edwards et al. (1988) discovered that a corn-soybean rotation with conservation tillage resulted in 12 percent higher yields for corn. Grover et al. (2009), studied long term crop yield variability on four crops---corn, soybeans, alfalfa and oats. They found that crops that were rotated produced yields 10-12 percent higher than crops that were grown continuously without rotation. These studies have demonstrated the boost in yield with crop rotations and farmers have responded by adopting these practices. Around 84 percent of corn and 94 percent of soybeans planted in the U.S. is rotated with other crops (USDA, 2013). While the majority of farmers employ crop rotation schemes, there is still a small percentage that maintains continuous cropping practices. Continuous cropping practices impact yields by what is known as yield drag or penalty. This refers to the effect of the yield of a crop depending on what was planted the year prior or two years prior. Nitrogen availability along with weather and crop residue accumulation are the main factors that encourage yield penalties (Gentry et al. 2013). Gentry et al. (2013) studied the effects of crop rotation on corn yield and found continuous corn crops to have lower yields over a soybean-corn rotation in all years except one. An accurate representation of farming practices in Big Creek necessitates incorporating all possible rotations that farmers utilize.

2.1.2 Soil and Topography

Soil physical, biological, and chemical properties are key factors that affect crop yield. Physical properties such as bulk density, soil organic carbon, and soil texture directly impact how well a crop can grow. Without sufficient soil organic carbon, the plant will not be able to extract the necessary nutrients to achieve maximum growth (Kravchencko and Donald, 2000; Yadev, 2007). In addition, soil texture impacts what type of crops can be planted. For instance, soils high in clay content are poor soils for crop growth because these soils have low infiltration rates and often lead to oversaturation. On the other hand, soils high in silt and loam are very good for plant growth. In poorer soils, perennial crops such as switchgrass and alfalfa are better alternatives than corn or soybeans. The texture of the soil can also impact how well water is held by the soil and how well it can move into and out of the soil profile. One measure to determine how much water is held in the soil is with available water capacity (AWC). AWC is calculated by taking the difference between field capacity (fully saturated) and wilting point -- the deficit of water in the soil that will lead to the plant beginning to wilt (Neitsch et al., 2011). Available water capacity determines how well a plant will grow because the crop needs sufficient water to reach maturation and the crop cannot be exposed to prolonged water stress, otherwise, crop yield will be significantly impacted. The moisture content of the soil plays a critical role in the water stress of the plant. If there is a water deficit due to drought and the soil is dry then the plant will begin to experience water stress due to evapotranspiration exceeding precipitation. Hydraulic conductivity measures how well water can move through the soil; this is based on the distribution of macropores and micropores within the soil structure. Soils with higher hydraulic conductivity can move water through the soil column faster --aiding crop yields -- as opposed to waterlogged soils, which are not suitable for crops. Soil temperature is another factor that affects crop yields and the planting date of the crop. If the soil temperature is not warm enough, the seedlings will not sprout and crop development will be hindered. This is the case for corn when temperatures are below 12.5°C and where there is also high soil saturation, which can decrease yields considerably (Dwyer et al., 2000).

Topography is another factor that can affect crop yields. Topography has a direct impact on soil physical, biological and chemical properties. It can influence the distribution of soil particles and it can affect the water holding capacity of the soil (Kravchenko and Bullock, 2000). Numerous studies have been performed that analyze the impact of topography on crop yields. For instance, topographic variables such as slope and elevation have been significantly correlated with crop yields. Steeper slopes lead to less organic matter and more erosion, which leads to poorer crop yields. Generally, slopes steeper than 5 percent are not suitable for crop production (Kravchenko and Bullock, 2000; Jiang and Thelen, 2004).

2.1.3 Weather and Climate

Weather and climate are a major factor affecting crop yield variability across the Central U.S (Grover et al., 2009). A plant's ability to grow and mature depends on how much water and temperature stress it experiences during the growth cycle. As precipitation falls on the watershed, plants and the soil intercept the water and it is evapo-transpired back into the atmosphere. Precipitation that makes it to the ground is available for plant use, but evapotranspiration represents the majority of water loss in a watershed. About sixty percent of the precipitation that falls on a given parcel of land is evapo-transpired back into the atmosphere (Neitsch et al., 2011). Direct measurements of evapotranspiration can be done with evaporation pans or lysimeters, but when time and resources are not available there are empirical equations that can be used to calculate it. Thornthwaite (1948) is credited with the first calculation of evapotranspiration and he used the term potential evapotranspiration to describe the amount of water that a plant would transpire if it was not at a loss of water and was at a certain height above ground. He also studied the effect of water deficit on plants as the difference between precipitation and potential evapotranspiration, which varies depending on the soil moisture (Thornthwaite, 1948). Over time though, Thornthwaite's evapotranspiration equation began to be replaced with newer methods that incorporated additional parameters.

Many other methods were introduced to calculate evapotranspiration and were either based on temperature methods or solar radiation methods. The Penman-Monteith method was the most widely used method and takes into account variables such as air temperature, solar radiation, relative humidity, and wind speed. It is widely used because it has been tested and measured against empirical data and performed well among all the different reference evapotranspiration (RET) methods (Jensen et al., 1990). More recently, the Penman-Monteith method has been measured against not so well known methods. Yoder et al. (2005), measured the Penman-Monteith method against seven other RET methods in the Cumberland Plateau of the Southeastern U.S. He compared both radiation based RET methods and temperature based methods. His results indicated that the Penman-Monteith method was the best method with a coefficient of determination of 0.91 against measured lysimeter readings in the field. The next best method was the original Penman method with a 0.91 r^2 and the Turc method with a 0.90 r^2 (Yoder et al., 2005). For scientists wishing to employ simpler methods who lack the necessary data, the Turc method is an attractive alternative. The Turc method only requires daily solar radiation along with average daily temperature. Accurately measuring PET is crucial for predicting crop yields in simulation models and metamodels.

Another weather variable that directly impacts crop yields is temperature. In order for crops to begin growing, there needs to be a temperature threshold that must be reached. This is known as a crop base temperature and varies from crop to crop and determines when a crop will begin to grow. Once the air temperature and soil temperature remain above the base temperature, the crop will grow. Not only do low temperatures affect crop growth, but extreme high temperatures can also stunt crop growth. If the temperature exceeds the high temperature threshold then the crop will begin to experience heat stress and this decreases crop growth and yield depending on how many days the crop experiences stress (Nietsch et al., 2011). This is especially the case when crops are in their early phenological growth cycle.

While day-to-day weather patterns affect the crop there are also long term climatic trends that impact the growth and yield of a crop. As climate change is estimated to impact average precipitation and temperature levels across the planet, there is increasing research into how climate change will impact individual watersheds. IPCC assessments indicate that the timing and intensity of precipitation and temperature events will become more volatile with increasing CO_2 in the atmosphere (Meehl et. al, 2007). This will play out not only on a global scale but at a regional and local scale. Climatic variability will be enhanced in the Midwest due to the wide range of precipitation and temperature regimes that exist in different locations across the Midwest. Each region in the US has a dominant crop or series of crops that grows best with regard to precipitation, temperature and soil conditions. For example, the Midwest is dominated by corn and soybean agriculture. The appropriate rates of evapotranspiration, precipitation, temperature, and soil type make this region ideal for growing these two crops. Another example is the Great Plains region. The soils there are less ideal and the precipitation is drastically less, so as a result, this region grows a lot of wheat. Simulation modeling can help determine how severe the impacts associated with climate change will be in watersheds all across the Midwest.

The key factor in long-term climatic trends that impact crop yields will be the timing and intensity of precipitation and temperature events. Current CO_2 levels of 400 parts per million will continue to increase through the end of the century. Higher CO_2 levels alone lead to higher crop yields due to increased photosynthesis and decreased transpiration, but this will be offset by increased temperatures, which will decrease yields (Baker and Allen, 1993). Enhanced CO_2 levels in the future will not only affect the mean temperatures, they will also affect variability (Porter & Semenov, 2005). Extreme events will become more extensive and unpredictable than past climates. Crop yields will be affected depending on the timing of the event. Different climatic events can alter the growth of the crop depending on the growth stage that the crop is in. Slight increases in temperature are problematic for crops such as wheat during the flowering phenological stage (Wheeler, 2012). Vital plant processes such as photosynthesis and evapotranspiration will also be impacted. The optimum temperature range for photosynthesis is between 20°C and 25°C for agricultural crops such as corn, soybeans, and wheat (Qaderi & Reed, 2009). Any deviation from this range will affect yield potential and the nutrients in the crops. Increases in minimum temperatures will also allow crops to be planted earlier, thus, leading to the possibilities of corn and soybean crops expanding north in the central US towards the Canadian border (Butler and Huybers, 2013).

Measuring the sensitivity of these climatic changes in crops represents a major area of interest. The delicate interplay of temperature, precipitation, and CO_2 can affect different crops in different ways across many different regions. Butler and Huybers (2013) measured the sensitivity of corn to high temperatures in the central US. They measured the sensitivity of corn to a 2°C increase in temperature. Results indicated that yield losses will vary depending on the region and the adaptations that the farmer implements. Counties located in warmer parts of the country are less sensitive to yield losses due to cultivar adaptation. Furthermore, yield losses from increased temperatures can be decreased by a factor of two with adaptation (Butler & Huybers, 2013). This of course relates to the hypothesized 2°C increase, but these results can vary depending on how extreme future climates will be.

A key way to measure and analyze impacts on agricultural land and ecosystem services are through the use of software models. Mapping of a watershed can be done through Geographic Information Systems software. The use of high powered computers and advanced programs can lead to better prediction and understanding of climate change impacts on agricultural land in the future under varying climatic scenarios.

2.2 Modeling

2.2.1 Simulation Models

Many different types of software can model watersheds but one type of software that is widely used is the Soil and Water Assessment Tool (SWAT). SWAT was developed by Dr. Jeff Arnold of the USDA for the Agricultural Research Service in the early 1990's. It is used by water resource managers, scientists and governmental officials to model a variety of watershed characteristics depending on the input variables (Couckuyt et al., 2009). SWAT is a physically-based continuous time model. Instead of modeling based on probability theory, it models the physical processes occurring in the watershed and uses a priori theory and mathematical equations to model these processes. The major components in SWAT are climate, hydrology, erosion, nutrients, plant growth, and management practices (Neitsch et al., 2011). The SWAT model is highly flexible and can be used to determine long-term impacts associated with land use change, and climate change or it can model short-term impacts such as crop rotation impacts on non-point source pollution in streams and lakes. It can also model at different scales from the HRU scale to sub-basin to watershed scale. The watershed is divided into sub-basins and HRU's, which are not spatially explicit in SWAT. The HRU's represent similar soil, topographic and land use properties. Since they are not spatially explicit they represent percentages of the sub-basins. The ability to be used at different scales and model different processes makes this model a powerful tool.

SWAT is a product of numerous legacy models, specifically Chemicals, Runoff, and Erosion from Agricultural Management (CREAMS), Groundwater Loading Effects on Agricultural Management Systems (GLEAMS), and Environmental Impact Policy Climate (EPIC). CREAMS modeled how daily rainfall affected the hydrology of the watershed. GLEAMS modeled the pesticide impacts on the watershed. EPIC modeled the crop growth in the watershed and how the crops responded to various inputs and soil erosion. These three models led to the creation of the Simulation of Water Resources in Rural Basins (SWRRB) model and ultimately SWAT (Gassman et al., 2007). Because of its outgrowth from these models it is used by non-governmental and governmental organizations such as the USDA and EPA. It is also used around the world to address watershed questions and help in watershed management and this wide rate of use has led to its success as a watershed simulation model. The SWAT model also has wide support by a community of developers and users. Because SWAT is open source and free, anybody has the ability to look at the source code and alter it to their specific needs.

There are a few reasons why SWAT is so widely used by many different people. One primary reason is that it has the ability to predict how certain variables such as land use and management affects water quality, sediment transport, and agricultural yields. It also is physically-based, which means the model utilizes empirical equations to describe the physical processes occurring in the watershed (Grayson et al., 1992). This is more advantageous to use over stochastic models when modeling complex physical processes across the watershed. SWAT was also created with the idea of utilizing readily available data such as stream gauge data from the USGS or weather data from NOAA. This has lead to a diverse array of applications.

Numerous peer-reviewed studies document the benefits and uses of SWAT. Applications of SWAT range from climate change modeling (Stonefelt et al., 2000; Fontaine et al 2001; Jha et al., 2006), to hydrologic predictions (Harmel et al., 2006; Cao et al., 2006), and farm pollutant effects (Saleh et al., 2000; Stewart 2006). In addition, there are numerous studies that relate to the robustness of the SWAT model in its ability to accurately model stream flow characteristics along with sedimentation and nutrients (Arnold et al., 1999; Rosenthal and Hoffman, 1999; Santhi et al., 2006).

The majority of published literature relating to SWAT has to do with the use of farm management techniques and how these practices affect the watershed from erosion to water quality in streams and lakes. The advantage of SWAT deals with its ability to measure these changes at the sub-basin and HRU scale. Tillage and fertilizer applications are managed by SWAT for each crop that is grown. These practices are implemented at certain points within the growing season based on the fraction potential heat unit index that SWAT uses when implementing these practices (Neitsch et al., 2011). The user can control what type of tillage is used and how much fertilizer is to be applied. Conservation practices can also be implemented to determine how techniques such as terracing and riparian buffers will affect sedimentation and nutrient losses into streams within the watershed.

The robustness of the model has also been tested by comparing different calibration and validation techniques (Bekele & Nicklow, 2007; Zhang et al., 2009; Moriasi et al., 2012). Bekele and Nicklow (2007), looked at the automatic calibration of SWAT using a genetic algorithm. They calibrated the daily stream flow and sediment concentration in the Big Creek Watershed. Their findings indicated that automatic calibration of SWAT using a genetic algorithm does a good job of depicting the parameters, but it tended to overestimate the parameters, which could be due to a lack of data given the small scale of the Big Creek watershed. SWAT has been tested for many different applications and its wide use as a watershed simulation model lends itself as a capable and robust model.

2.2.2 Metamodels

2.2.2.1 What is a metamodel

A metamodel is essentially a model of a model (Broad et al., 2010). Metamodels are created because there is often a need to approximate real world systems and simulation models with a mathematic equation. The metamodel can perform at higher speeds than the simulation model and most of the time can approximate the real world system nearly as accurately. The metamodel mirrors the input and output processes that take place in the simulation model. Often the simulation model requires large amounts of inputs and the metamodel is more attractive because it reduces the amount of inputs necessary and aims to achieve a similar result with the output. By operating in this manner, it makes it more efficient to create a metamodel than to run the simulation model. This is especially the case if the time and resources for a project are limited. Metamodels have many different forms and they are applied in a wide variety of fields from engineering to agriculture.

2.2.2.2 Metamodel Form

The modeling process begins by deciding what metamodel form will best fit the data at hand. The researcher must determine the appropriate metamodel class and form. Metamodel classes such as neural networks, kriging, polynomial and linear regression are frequently used in the literature and provide robust techniques for creating the metamodel. Multiple linear regression is the simplest form to use (Equation 2.1) and is a common metamodel technique. Polynomial metamodels apply a higher order polynomial fit to the data depending on the order chosen by the researcher (Equation 2.2). Depending on the polynomial order of the independent variable there will be K coefficients for each independent variable. Kriging models utilize a general linear regression form to measure the correlation in the residuals of the observations with the regression model (Equation 2.3). Jin et al. (2007), studied the various metamodelling techniques to discover which one was the most efficient and robust, and in terms of model construction and efficiency the polynomial regression was suggested by the authors as the first metamodel class to use. Pineros Garcet et al. (2005), assessed how a kriging metamodel performed against a neural network model in nitrate leaching prediction, and found that the kriging model performed better in terms of RMSE and model efficiency than the neural network metamodel. If one metamodel class does not perform adequately, the researcher can choose another class to reach the desired results.

Equation 2.1 Multiple Linear Regression Equation

Equation 2.2 Polynomial Regression Equation

Equation 2.3 Kriging Equation

There are several approaches to selecting independent variables in the process of fitting regression models. One approach is known as all-possible-subsets regression. In this method the researcher examines all the potential combinations of independent variables and measures their goodness-of-fit with each other. This function can be utilized in the R software and the all-subsets can be based on R^2, adjusted R^2 or Mallows C. Variables that have the highest goodness-of-fit should be included in the equation, whereas, variables with low goodness-of-fit should be omitted from the final equation due to the lack of explanatory or predictive power, unless they have theoretical justification.

Another approach in variable selection is stepwise regression, which is an alternative to all subsets regression. This procedure examines the contribution of each variable and if the variable has a significant contribution to the equation then it is added, whereas, if the variable does not have a significant contribution then it is omitted. Stepwise is usually based on Akaike Information Criterion (AIC). The variable with the greatest contribution is added first along with subsequent variables, and this is known as forward addition. Another approach is backward elimination, which adds all the variables first and then systematically eliminates them using AIC. Both of these methods are not exclusive but are often used in tandem in the beginning stages of model identification.

Models that are intended for predictive purposes have two primary goals. The first is that the model should be developed with the best possible predictors. A model that is not efficient or robust will not have good predictive power. The second goal is that the model should be a parsimonious model -- it should limit the explanatory variables and have as few variables as possible while trying to maximize variance explained (Burt et al., 2009). This research intends to use these models for predictive purposes and will seek a balance between being parsimonious and including the necessary variables in crop yield prediction.

2.2.2.3 Metamodel Applications

Numerous studies utilize metamodels as a means to replace a simulation model. Metamodels are primarily used in the fields of engineering and computer science, but there are numerous studies where metamodels are applied to agricultural issues. For agricultural applications, metamodels have been employed for measuring irrigation needs, along with nonpoint source pollution (Bouzaher et al., 1992; Wu & Babcock, 1996; Galelli et al., 2010). Galelli et al. (2010), built a metamodel to measure water demand for the Muzza-Bassa irrigation district. They utilized crop parameters along with soil and meteorological parameters as inputs to the metamodel. For metamodel class they chose a state-dependent parameter model and achieved a high R^2 of around 80 percent for the calibration and validation datasets (Galelli et al., 2010). Wu & Babcock (1996), modeled nitrate water pollution in the Central US using a regression metamodel to approximate the EPIC simulation model. Input parameters utilized for the model included management systems, soil properties, and weather parameters. They found that most of the coefficients were significant at the 0.1 alpha level and that the model was able to predict 75 percent of nitrogen runoff and 73 percent of nitrogen leaching. Bouzaher et al. (1992) also looked at modeling non-point source pollution using a nonlinear regression metamodel. Their metamodel approximated the PRZM and STREAM simulation models. Most of the explanatory variables for the metamodel represented soil properties such as organic matter and bulk density. The metamodel form they used was a non-linear regression metamodel that was able to explain more than 80 percent of the variance in non-point source pollution amounts (Bouzaher et al., 1992). These studies demonstrate that metamodels are able to explain their chosen phenomena with a moderate-high degree of accuracy.

CHAPTER 3

METHODOLOGY

3.1 Parameter Selection

The parameters that were used in each of the eight watersheds were: weather, tillage, soil, topography, fertilizer, and crop rotation. Each of these parameters was justified based on theory or empirical evidence as found in the literature. Variables were calculated at different temporal scales from yearly variables to seasonal and monthly variables. These were varied in order to capture certain weather impacts on crop phenological stage during the course of the growing season. The methodology followed the workflow in Figure 3.1.

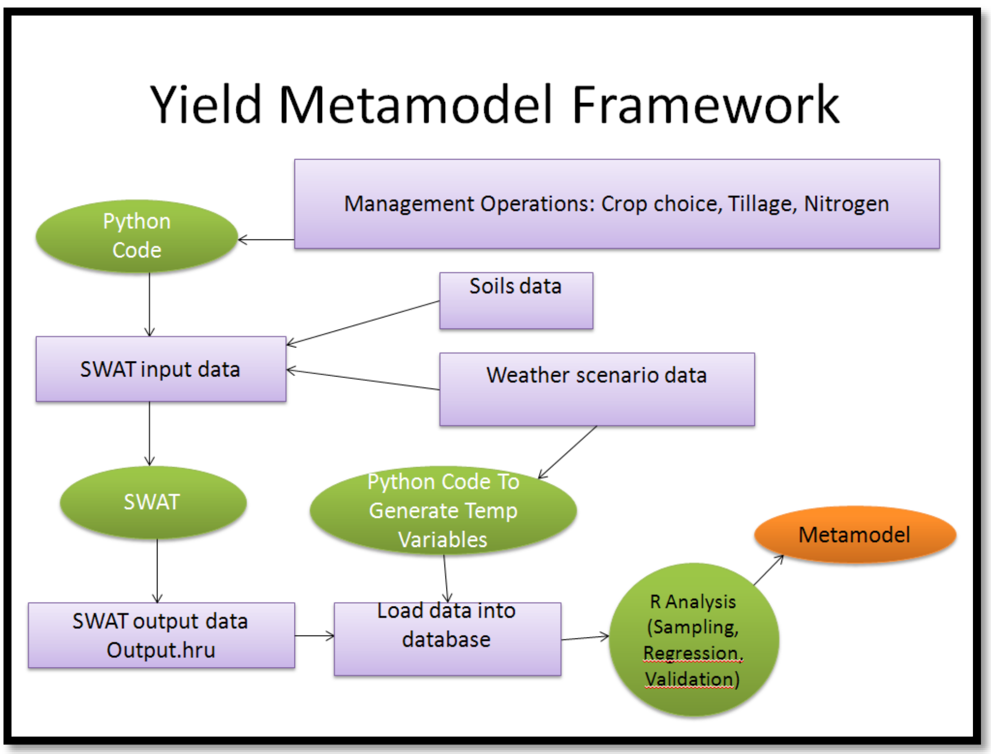


Figure 3.1 Metamodel Methodology Framework

3.1.1 Weather

Weather variables represent the dominant influence on crop yields. Daily mean temperature and precipitation were used from both historical records from 1974-2010 and future climate data for the Anna, IL weather station. The future climate data was downscaled to Big Creek Watershed by Dr. Justin Schoof. Daily weather data were provided for precipitation, solar radiation, wind speed, relative humidity, and minimum and maximum temperature for the time period 2006-2095.

In order to approximate the crop growth model in SWAT, the metamodel utilizes mean daily temperature data or maximum daily temperature to derive values for temperature stress. Daily temperature values were input into the python script to calculate temperature stress as an output. These data were then brought into excel and aggregated using pivot tables to apply the final value for the specific year and applied to the whole watershed. Monthly sums and seasonal sums are used as potential explanatory variables in the model. Temperature stress is a function of three components. Daily average air temperature, the optimal temperature of the crop, and the base temperature of the crop are all utilized to derive temperature stress. Under optimal conditions, the temperature stress for a given crop is zero and approaches one as temperature stress increases (Neitsch et al., 2011). The crop receives a temperature stress of 1 if the daily average air temperature is below the base temperature and above the high temperature threshold. For daily temperature between the base temperature and high temperature threshold, the temperature stress value ranges from 0-1 for the day. The data needed for optimal temperature and base temperature for each plant was extracted from the SWAT plant database. Table 3.1 shows the base, optimal, and high temperatures for each crop.

Table 3.1 Crop Temperature Thresholds

Crop Temperature Thresholds Celsius					
	Corn	Soy	Cotton	Alfalfa	Switch grass
Base	8	10	15	4	12
Optimum	25	25	30	20	25
High	42	40	45	36	38

Equation 3.1 details the formulas that were used to calculate temperature stress. This is the same formula that SWAT uses to calculate temperature stress. Due to the lack of sensitivity of the temperature stress in the hot months of the growing season, there were two monthly temperature stress variables created for July and August that were based on daily maximum temperature instead of daily average temperature. April and May temperature stress were calculated using daily average temperature to capture the effects of the cool temperatures on the growth of the crop in the early phenological stage.

tstress = 1 when

tstress = 1-exp ———— when

tstress = 1- exp ———— when

tstress = 1 when

Equation 3.1 Temperature Stress Equation Utilized by SWAT (Neitsch et al. 2011)

An alternative temperature variable was used in addition to temperature stress. The growing degree day (GDD) method was used to calculate GDD's for each month of the growing season from April through September along with total annual GDD's. Each day where the temperature rises above the base temperature of the crop, the GDD value was the difference between the daily average temperature and base temperature. Each positive daily GDD was summed over the month to get the monthly value (McMaster and Smika, 1997). Once temperature reaches the high temperature threshold for the crop, GDD was not added. Table 3.1 shows the base, optimal and high temperature thresholds for each crop that were used in the GDD calculation.

Moisture variables were represented by total annual precipitation, monthly precipitation during the summer months and the water surplus (WS) method. WS was calculated by taking the difference between precipitation and the potential evapotranspiration method developed by Thornthwaite (Thornthwaite, 1948). Each day of the growing season WS was calculated and then summed up over the entire season to derive a monthly and yearly WS. WS was calculated for the May-August time period due to the importance of precipitation on the crop during its maximum growth stage. If the water surplus for a crop is too low then there will be a negative impact on the yields of that specific crop, but if precipitation exceeds potential evapotranspiration then there will be a water surplus in the watershed and the crop will have higher yields. Potential evapotranspiration was derived according to the Turc method (Yoder et al., 2005). The Turc method (Equation 3.2) incorporates daily average air temperature and solar radiation as detailed in the literature review section. This formula was implemented in excel to calculate that daily reference evapotranspiration.

Equation 3.2 Turc Reference Evapotranspiration Formula

Where:

A_t = 1.0 for RH_{mean} = 50%

T_{avg} = daily average temperature in °C

R_s = solar radiation (MJ m^2/day)

λ = 2.49

3.1.2 Tillage

Tillage was a categorical variable in the model that takes the value of no-till, conservation tillage, or conventional tillage. Dummy codes generated by R are used to indicate the level of tillage in the model. The baseline value for the tillage dummy variable was set to conventional tillage. Tillage varies with each crop rotation that is input into SWAT in order to achieve the full range of possible input values.

3.1.3 Fertilizer

Fertilizer represents the amount of nitrogen fertilizer applied to the crop. Data for fertilizer are available at the national level from the USDA Economic Research Service. Data are also utilized at the county level from the Conservation Technology Information Center (CTIC) databases. Fertilizer amounts vary depending upon the crop being modeled. Corn was modeled with a nitrogen fertilizer range of 112-222 kg/ha, cotton with a range of 33-96 kg/ha (Roberts et al.

1999), switchgrass from 0-134 kg/ha (Schmer et al, 2008). For the winter wheat rotation in the soybean model, 75 kg/ha of fertilizer was applied to winter wheat. Soybeans and alfalfa generally do not require nitrogen fertilizer so the amount was set to zero for both crops.

3.1.4 Soil

Soil properties utilized for the metamodel were determined by reviewing the literature to determine which soil properties correlate highest with crop yields. The soil properties also had to be utilized by the SWAT input file. For the metamodel to be consistent, soil organic carbon, available water holding capacity, and hydraulic conductivity are three variables that were chosen based on their correlation with high crop yields. Bulk density was considered, but it has a high correlation with soil organic carbon so this variable was excluded from the model. These data were obtained by the Natural Resources Conservation Service (NRCS) SSURGO soils database, which contains extensive information on soil physical, biological, and chemical properties for every county in the United States. Soil property data are extracted from the soils table stored in the SWAT access database. Soil data are provided for each horizon or as a total for the entire soil column.

3.1.5 Crop Rotation

Crop rotation represents the crop choices that farmers have over a given time frame. A three-year crop rotation scheme was employed for Big Creek Watershed. Examples of three-year crop rotations would be corn-soybean-corn, corn-corn-soybean or alfalfa-alfalfa-alfalfa. Corn was rotated with cotton and soybeans. Soybean was rotated with corn and cotton. Cotton was rotated with corn and soybeans. Switchgrass and alfalfa, both perennials, were planted continuously. Table 3.2 details the crop rotations used in the models. The first column represents the crop that was planted in the first year and the second and third columns represent the crops in the second and third years of the three-year crop rotation. Crops and rotations were determined based on the cropland data layer from the National Agricultural Statistics Service (NASS) and potential future crops were added to the rotation based on hypothesized climate conditions in the future. For example, cotton or switchgrass could be grown in Big Creek watershed under a warmer climate and, as a result, cotton was added to the rotation possibilities for Big Creek. Crop rotations were stored in a lookup table, which the python script read from to generate the desired permutation that was simulated through SWAT. Crop rotation dummy variables were generated by the R statistical software. Baseline dummy rotation was corn-corn-corn for the corn model, corn-corn-soybean for the soybean model, and corn-corn-cotton for the cotton model.

Table 3.2 Crop Rotations Used in the Models

First Year (C-2)	Second Year (C-1)	Third Year (C)
Corn	Corn	Corn
Corn	Soybean	Corn
Soybean	Soybean	Corn
Corn	Corn	Soybean
Soybean	Corn	Soybean

Table 3.2 (Continued)

Soybean	Soybean	Soybean
Cotton	Corn	Soybean
Cotton	Soybean	Soybean
Corn	Soybean/WW	Soybean
Corn	Cotton	Corn
Soybean	Cotton	Corn
Cotton	Corn	Corn
Cotton	Soybean	Cotton
Soybean	Cotton	Cotton
Corn	Corn	Cotton
Corn	Soybean	Cotton
Alfalfa	Alfalfa	Alfalfa
Switchgrass	Switchgrass	Switchgrass

3.2 SWAT

3.2.1 SWAT Data

Three initial data sources were needed to begin the SWAT procedure. These data sources were topographic, land use, and soils data. Topographic data were derived from a 30-meter digital elevation model (DEM). This DEM was obtained from the USGS National Elevation Dataset. Land use data came from the USGS National Land Cover dataset. Soils data came from the NRCS SSURGO database, which is based on a 1:24,000 scale. All three of these datasets were projected into the World Geodetic System Datum UTM Zone 15N and overlaid to begin the watershed delineation process.

3.2.2 SWAT Setup

The next step in setting up SWAT is to load in the necessary layers to create the HRU's. First, SWAT creates sub-basins within the watershed based on the digital elevation model (DEM) that the user inputs. SWAT then delineates those sub-basins into smaller HRU's with a unique combination of land use, soils, and slope (Figure 3.2).

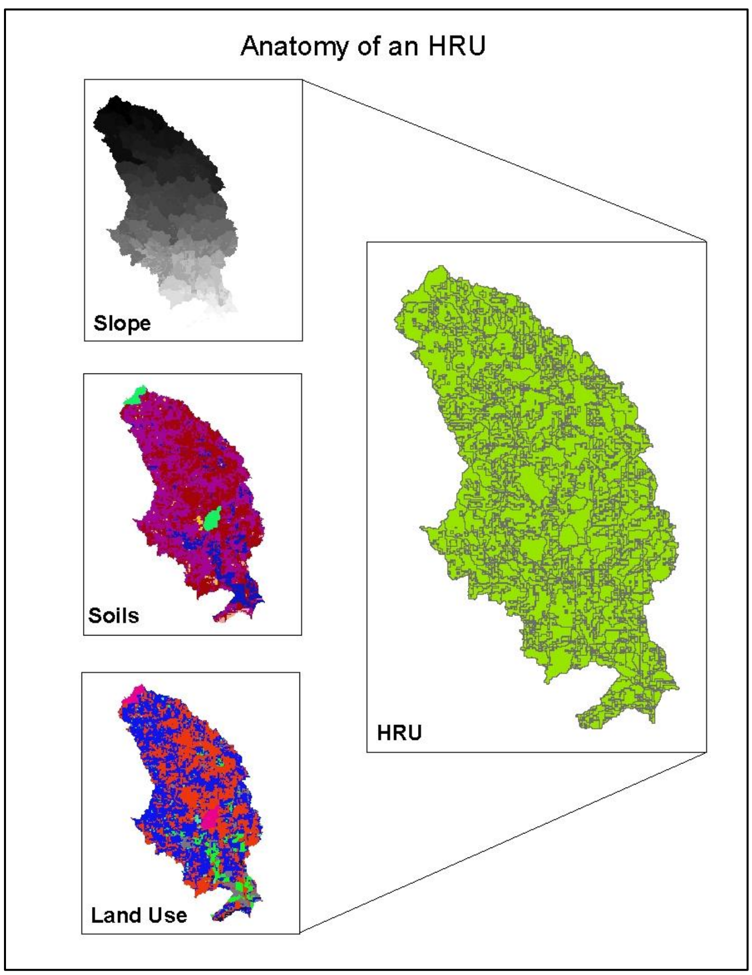


Figure 3.2 Anatomy of Big Creek Hydrologic Response Units

HRU's are not spatially explicit by default in the SWAT model, but for this project the HRU's were made to be spatially explicit. In order for SWAT to accurately model watershed conditions, the model has to be calibrated to existing conditions within the watershed using three steps. The first step is to use part of the observation data that the user obtains. The second step is run the model over a range of values for the unknown parameters in order to achieve best-fit results. The third step is to apply the calibrated model to the remaining observation data that was held out in the first step (Neitsch et al., 2011).

3.2.3 SWAT Calibration & Validation

Big Creek had a total of 1644 HRU's after calibration (Figure 3.3). Big Creek watershed was calibrated according to data in the years 1999, 2000, and 2001 and validated on the years 2002 and 2003. After SWAT was calibrated and validated it was ready to run simulations with different climate scenarios. Historical climate data and future scenarios were run through SWAT (Table 3.4-3.5). Since SWAT operates at a daily time step, daily data were required to run a three-year crop rotation simulation for solar radiation, maximum and minimum temperature, precipitation, relative humidity, and wind speed. These data were formatted in excel files and then output to the necessary SWAT files. SWAT simulations were able to output accurate yield values compared against historical averages. Utilizing historical yield data for Union County from NASS, the mean corn and soybean yields of the SWAT runs were compared against historical averages from 1974-2010. The historical average for corn was 6.71 t/ha, while the SWAT simulation was 7.9 t/ha. For soybeans, the historical average was 2 t/ha, and the mean yield for the SWAT runs was 2.5 t/ha. These mean yields overestimate the historical averages, but they also incorporate potential future climate scenarios not just historical climate.

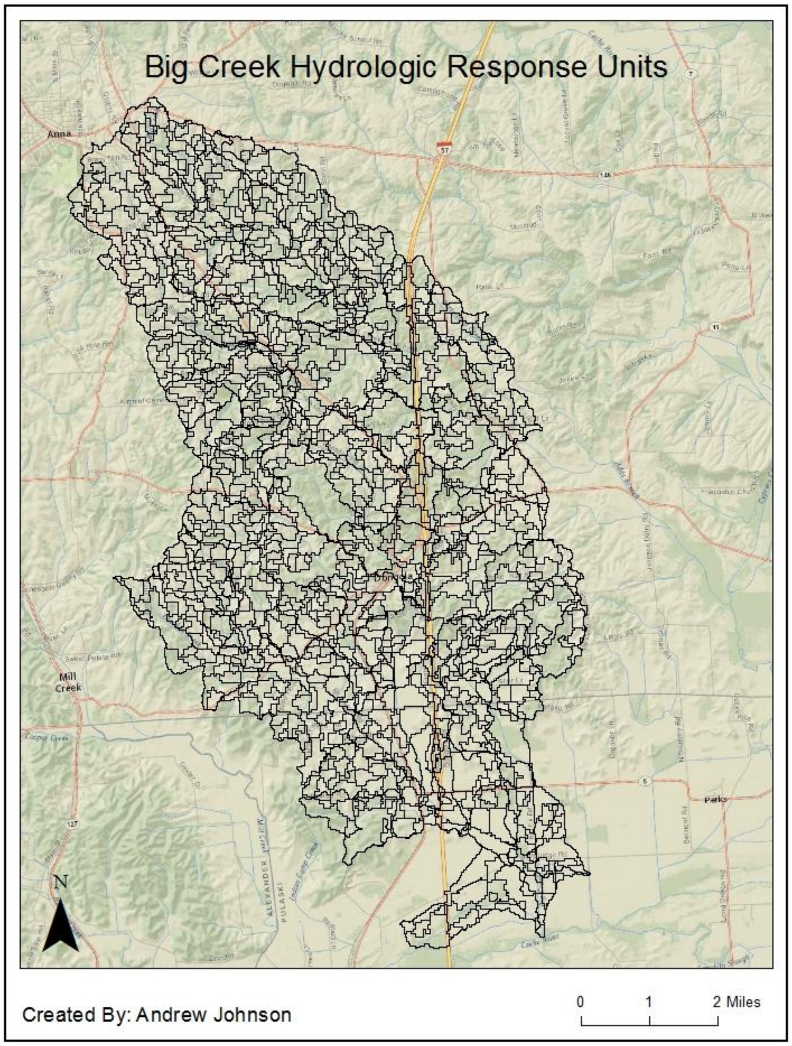


Figure 3.3 Big Creek Hydrologic Response Units

3.3 Python

The python programming language was instrumental in the replacement of SWAT and ultimately in the creation of the metamodel itself. Numerous python scripts were created for pre-processing, post-processing, SWAT automation, and calculating the independent variables. Table 3.3 shows the python scripts that were used along with the dedicated task that each script performed. How python was used will be detailed according to the tasks that the script performed.

Table 3.3 Python Scripts Used in this Research

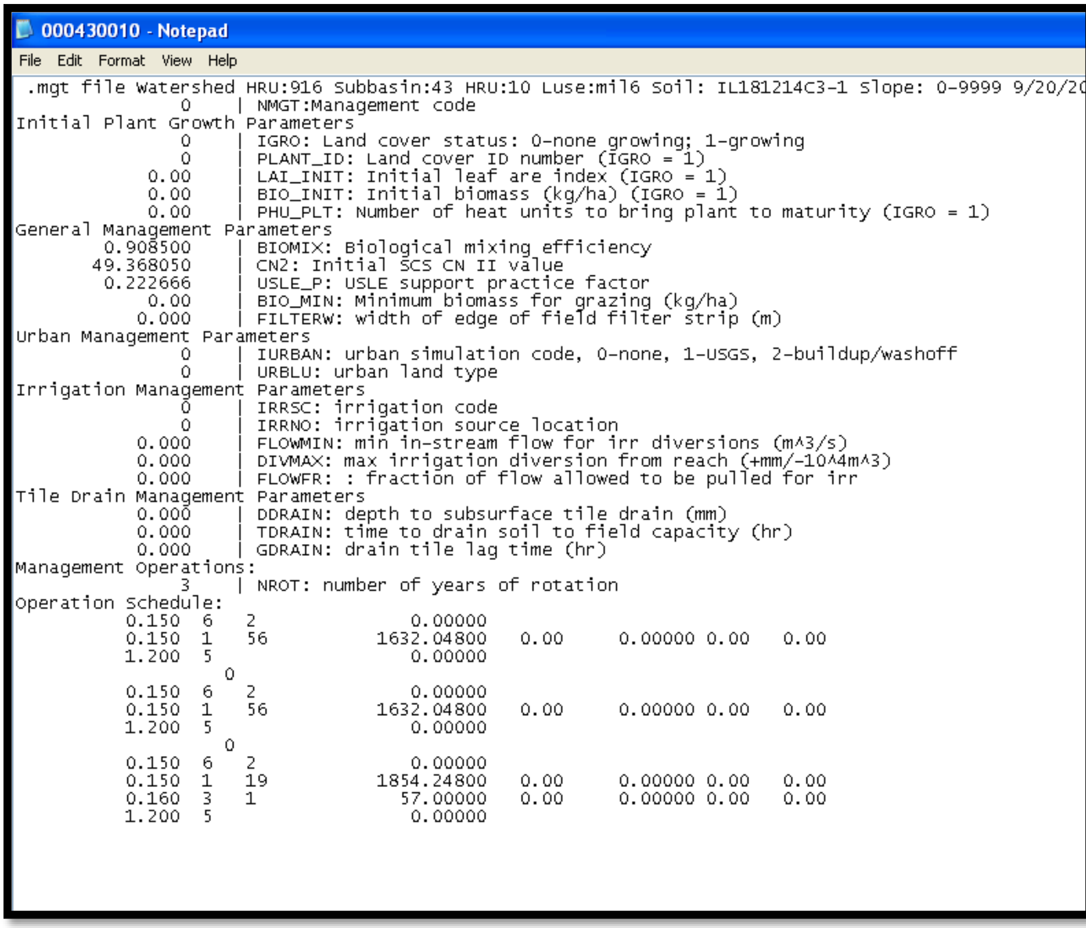
Python Scripts	
Name	**Task**
TempSplit.py	preprocessing
CombinationsV10.py	SWAT automation
Extract.py	post processing
CRN_tstress.py	Variable Creation
CRN_GDD.py	Variable Creation
Soy_tstress.py	Variable Creation
Soy_GDD.py	Variable Creation
CTN_tstress.py	Variable Creation
CTN_GDD.py	Variable Creation
Alf_tstress.py	Variable Creation
Alf_GDD.py	Variable Creation
SWG_tstress.py	Variable Creation
SWG_GDD.py	Variable Creation

3.3.1. Pre-processing Script

Pre-processing involves taking the raw weather data in a text file format and bringing it into Microsoft Excel to format it for SWAT use. SWAT requires that each daily observation in the weather file have a year value, unique ID value, and value associated with the specific weather variable. All three of these values must be concatenated and then output into the appropriate SWAT text file. The raw weather data must be bounded by zeroes and this is where the python script was implemented. The script ensured quality control for the data that were fed into SWAT.

3.3.2. SWAT Automation Script

The automation script was a rotation generator script to generate all the permutations possible in the population dataset. These permutations represented the possible rotations for each crop along with the tillage application. The number of possible permutations varied according to the crop. For example, soybeans and corn incorporated cotton as a possible rotation while alfalfa and switchgrass did not because they are perennial crops and are implemented in 5-year cycles. The possible rotations were based on available empirical data regarding farmer practices and potential rotations that might exist under a future climate scenario. The second script replaced all the rotations in each of the mgt files (Figure 3.4) within SWAT. The mgt file is where the management information for each HRU is stored for SWAT to process when the simulation is run. Figure 3.4 represents a sample mgt file where the python script replaces the bottom half of the file where the operation schedule is listed. Those numbers changed depending on the unique permutation that is called up by the script. The third script was used to call up SWAT to run all of the rotations for a given scenario.



000430010 - Notepad

File Edit Format View Help

```
 .mgt file Watershed HRU:916 Subbasin:43 HRU:10 Luse:mil6 Soil: IL181214C3-1 Slope: 0-9999 9/20/20
               0    | NMGT:Management code
Initial Plant Growth Parameters
               0    | IGRO: Land cover status: 0-none growing; 1-growing
               0    | PLANT_ID: Land cover ID number (IGRO = 1)
            0.00    | LAI_INIT: Initial leaf are index (IGRO = 1)
            0.00    | BIO_INIT: Initial biomass (kg/ha) (IGRO = 1)
            0.00    | PHU_PLT: Number of heat units to bring plant to maturity (IGRO = 1)
General Management Parameters
        0.908500    | BIOMIX: Biological mixing efficiency
       49.368050    | CN2: Initial SCS CN II value
        0.222666    | USLE_P: USLE support practice factor
            0.00    | BIO_MIN: Minimum biomass for grazing (kg/ha)
           0.000    | FILTERW: width of edge of field filter strip (m)
Urban Management Parameters
               0    | IURBAN: urban simulation code, 0-none, 1-USGS, 2-buildup/washoff
               0    | URBLU: urban land type
Irrigation Management Parameters
               0    | IRRSC: irrigation code
               0    | IRRNO: irrigation source location
           0.000    | FLOWMIN: min in-stream flow for irr diversions (m^3/s)
           0.000    | DIVMAX: max irrigation diversion from reach (+mm/-10^4m^3)
           0.000    | FLOWFR: : fraction of flow allowed to be pulled for irr
Tile Drain Management Parameters
           0.000    | DDRAIN: depth to subsurface tile drain (mm)
           0.000    | TDRAIN: time to drain soil to field capacity (hr)
           0.000    | GDRAIN: drain tile lag time (hr)
Management Operations:
               3    | NROT: number of years of rotation
Operation Schedule:
          0.150  6   2          0.00000
          0.150  1   56      1632.04800   0.00      0.00000 0.00    0.00
          1.200  5              0.00000
                   0
          0.150  6   2          0.00000
          0.150  1   56      1632.04800   0.00      0.00000 0.00    0.00
          1.200  5              0.00000
                   0
          0.150  6   2          0.00000
          0.150  1   19      1854.24800   0.00      0.00000 0.00    0.00
          0.160  3   1         57.00000   0.00      0.00000 0.00    0.00
          1.200  5              0.00000
```

Figure 3.4 Sample MGT File Representing One HRU

3.3.3. Independent Variable Construction

The last script was used to extract the data from the input and output files. Python was also used to calculate temperature stress and growing degree days. Because these variables need to be calculated based on certain conditions, such as average temperature, python was necessary to implement conditional statements. Python is a fundamental part of the methodology and these scripts helped reduce the time necessary to extract the data and run all the scenarios.

3.4 Metamodels

Once the data were sampled using a simple random sample technique, a regression-based metamodel was used to mimic SWAT. Methodology for metamodel creation was based on previous research. The metamodel class linear regression was used to estimate the parameters. Linear regression was chosen because of the interpretability of the coefficients and the ability to incorporate categorical variables into the model, which other metamodel forms cannot (e.g. Kriging and ANN). The data were input into a regression equation to derive the necessary output. The output dependent variable represents crop yield derived from SWAT output HRU files. The explanatory variables were derived from SWAT files and python scripts. The process of creating these metamodels was performed utilizing the open source statistics software R. These equations were applied to each HRU to determine various crop yields under various climate scenarios in any given HRU within Big Creek Watershed.

3.4.1 Metamodel Calibration

In order to estimate the metamodel coefficients, the simulation model SWAT must be run over the entire range of input values used by the metamodel. SWAT runs were chosen based on their ability to maximize the variance of the climate variables and management practices. With each permutation, SWAT generated a yearly crop yield value that was placed into a database along with the explanatory variables associated with the crop yield. Once the population dataset was created, a simple random sample was taken from the dataset in order to create the metamodel. The sample size used was 20 percent. The statistical software R was used in order to perform an all-subsets regression procedure on the sampled dataset. All-subsets regression performs an exhaustive forward and backward search through all of the variables and outputs the best predictors for each number of variables (Burt et al., 2009). For example, the user can select to output the best three predictors so R would output the top three variable combinations for each number of predictors based on some criteria such as R^2 or AIC. The best regression model was chosen based on the coefficient of determination, and root mean square error. The signs of the coefficients were analyzed along with any potential problems arising from multicollinearity between the variables. The calibration model was then measured against a validation model.

3.4.2 Metamodel Validation

The metamodel was validated by taking another sample from the population dataset in order to compare the results from the first sample with that of the second sample. The models were compared based on their R^2 and RMSE. Corn and soybean yields produced by the metamodel were compared against historical average corn and soybean yields in Union County. Once the metamodel was validated, it was used for predicting crop yields in Big Creek Watershed under various climate scenarios.

3.4.3 Climate Scenarios

Climate scenarios were employed in order to capture variability associated with temperature and precipitation. Dr Justin Schoof provided downscaled climate data that corresponded with a low, medium, and high scenario for CO_2 levels in the atmosphere. He utilized regional climate projections RCP 26, RCP 45, and RCP 85 for each of the four climate models: IPSLCM, CNRM, MRI, and MPI. For 48 years in historical, MRI, and IPSLCM weather scenarios, SWAT was run and the output data were collected and combined into one large population dataset. In order to capture the possible realizations with future climate, some of the data were adjusted to account for a wide range of weather. These data were varied beyond what the models predicted. Annual precipitation was varied from a low of 200 mm to a high of 2000 mm (Table 3.4). Temperature was varied from a low annual average temperature of around 12°C to a high annual average temperature near 18°C (Table 3.5).

The year column in Table 3.4 represents the final year used to calibrate the metamodel from the historical and downscaled data. Years run with historical weather include: 1974-1976, 1978-1980, 1980-1982, 1986-1988, 1995-1997, 1998-2000, and 2004-2006. Years that were run under the RCP85 carbon scenario of the IPSLCM model were: 2016-2018, 2044-2046, 2059-2061, and 2082-2084. Years that were run through the RCP26 carbon scenario for the IPSLCM model were 2050-2052. Years that were run through the RCP85 carbon scenario of the MRIC model were 2044-2046, and 2072-2074. The years with altered precipitation values include 1988B, 2052A, and 2052B. The years where mean annual temperature was altered were 2018 and 2084. A total of 48 years were run through SWAT. Multiply that by the number of permutations for each crop and the dataset was populated with hundreds of thousands of observations that represent HRU's in Big Creek Watershed.

Table 3.4 Sixteen Years Run Through SWAT with the Year Representing the Final Year in the Three Year Rotation Sequence

Year	Total Precip (mm)	Total PET (mm)
1976	895.8	443.763
1980	805.5	450.987
1982	1407	454.170
1988A	988.6	467.627
1988B	743.3	467.627
1997	1145	446.596
2000	1243	439.352
2006	1540	444.888
2046	1776	395.589
2074	2030	399.615
2093	1615	404.653
2052A	217.3	401.715
2052B	308.3	401.715
2061	888.4	399.140
2084	1362.1	418.469
2018	1050.3	363.384

Table 3.5 Average Air Temperature °C For Each Month in the Growing Season

	Temperature Variables	(All averages of daily avg air temp Celsius)					
Year	April	May	June	July	Aug	Sept	Annual
1976	15.583	17.377	23.212	26.211	23.740	20.355	13.804
1980	14.008	19.647	25.053	28.816	28.529	23.563	14.622
1982	12.467	22.355	23.035	26.881	24.768	20.730	14.534
1988A	15.047	20.337	24.918	26.979	27.589	21.952	14.716
1988B	15.047	20.337	24.918	26.979	27.589	21.952	14.716
1997	11.205	16.735	22.713	25.592	23.792	20.323	13.333
2000	13.088	20.747	23.697	24.685	25.723	20.405	13.948
2006	17.048	18.874	24.068	25.629	25.679	18.590	14.705
2046	18.325	21.816	25.565	27.166	24.795	20.220	15.025
2074	20.755	23.897	25.618	26.840	23.692	20.167	15.265
2093	19.538	22.894	28.338	28.110	24.861	16.630	15.856
2052A	17.368	25.274	26.598	29.568	26.947	19.318	15.751

Table 3.5 (Continued)

2052B	17.368	25.274	26.598	29.568	26.947	19.318	15.751
2061	19.77333333	25.07903226	25.4	25.89032	26.9371	20.83833	16.15356
2084	22.8805	27.15935484	30.096	31.075	26.74419	21.703	17.750
2018	13.105	16.688	19.957	23.048	20.040	17.208	12.144

CHAPTER 4

RESULTS

Table 4.1 details the five crop models with their associated explanatory variables and coefficients for each variable. R^2, RMSE, and the coefficient of variation are reported for each model at the end of the table. The variables included in the model are: nitrogen, slope length, available water capacity, May water surplus, August water surplus, April growing degree days, June growing degree days, and annual potential evapotranspiration. Dummy variables in the model to account for management techniques are: first year rotation as cotton (C-2), first year rotation as soy (C-2), second year rotation as cotton (C-1), second year rotation as soy (C-1), second year rotation as winter wheat (C-1), conservation tillage and no-till. Each of the regression models were validated by taking a new sample and comparing the results against the calibration model.

Table 4.1 Regression Models For Each Crop

Variables	Corn	Soy	Alfalfa	SWG	Cotton
Intercept	-2.396	-7.601	-1.92092	-43.48	-7.634
Nitrogen (kg)	0.0318	0	0	0.08068	0.0181
Slope Length (m)	-0.0787	-0.01459	-0.03543	-0.06845	-0.008
Available Water Capacity (mm)	0.00731	0.001159	0.002248	0.01119	0.002
May Water Surplus (mm)	0.00942	0.006381	-0.00296	-0.01681	0.006
August Water Surplus (mm)	0.00905	0.004245	0.0309	0.0413	0.001
April Growing Degree Days (Days)	0.01615	0.00581	0.006883	0.004744	0.004
June Growing Degree Days (Days)	-0.0351	-0.00064	-0.0277	-0.01309	-0.005
PET (mm)	0.0401	0.02758	0.053334	0.115	0.021
First year rotation cotton (C-2)	0.2499	0.01352			0.062
First year rotation soy (C-2)	0.1097	0.007692			0.168

Table 4.1 (Continued)

Second year rotation cotton (C-1)	0.08394	-0.1278			-0.175
Second year rotation soy (C-1)	0.3518	0.01228			0.065
Second year rotation winter wheat		-0.5487			
Conservation till	0.1351	-0.0045	-0.01208	0.4525	-0.006
No-till	0.1211	-0.01051	0.007657	0.1277	-0.16
Calibration R^2	0.7059	0.6233	0.8381	0.7672	0.4745
Calibration RMSE	1.679	0.7563	0.9077	2.363	0.844
Validation R^2	0.7074	0.6263	0.8405	0.7593	0.4735
Validation RMSE	1.669	0.7534	0.8953	2.4	0.841
Mean Yield (t/ha)	7.976	2.556	6.635	8.804	2.037
Coefficient of Variation	0.21050	0.298	0.136804	0.26840	0.4143

Table 4.2 Standardized Coefficients For Each Explanatory Variable

Standardized Coefficients					
	Corn	Soy	Alfalfa	SWG	Cotton
Nitrogen	0.341	0	0	0.608	0.24
AWC	0.182	0.073	0.077	0.172	0.154
SLOPE	-0.226	-0.105	-0.145	-0.135	-0.046
April GDD	0.429	0.451	0.247	0.076	0.233
June GDD	-0.747	-0.557	-0.841	-0.186	-0.5
MAYWS	0.156	0.269	-0.064	-0.195	0.246
AUGWS	0.159	0.17	0.462	0.442	0.048
PET	0.379	0.792	0.688	0.686	0.431

Results for each model are reported with univariate, bivariate, and multivariate statistics.

4.1 Soybean Model

Univariate Statistics

Table 4.3 details the summary statistics for each of the explanatory variables along with the dependent variable.

Table 4.3 Summary Statistics for Soybean Model Explanatory Variables

Variable	Min	Mean	Median	Max
Precipitation (mm)	217.3	1063.8	988.6	2029.7
Nitrogen (Kg/ha)	0	0	0	0
Slope Length (meters)	0.1907	11.4117	8.5028	66.7935
AWC (millimeters)	7.428	291.121	243.2	443.1
AprilTS (Days)	3.468	13.244	14.864	25.658
MayTS (Days)	0.2791	4.1927	2.0181	14.8085
JulyTS (Days)	0.0416	15.3963	16.445	21.230
AugTS (Days)	0.3284	15.225	18.988	20.528
AprilGDD (Days)	57.0	194.6	169.8	397.7
MayGDD (Days)	142.7	348.2	333.1	531.9
JuneGDD (Days)	223.9	449.2	602.9	602.9
JulyGDD (Days)	315.2	519.7	523.3	653.3
AugGDD (Days)	233.6	470.5	486.1	574.4
GDD (Days)	1164	2589	2586	3465
PET (mm)	338.2	417.6	402.8	467.6
AprWS (mm)	-31.449	33.584	25.557	143.429
MayWS (mm)	-36.31	50.81	41.74	141.25
JuneWS (mm)	-58.2	25.96	33.18	114.29
JulyWS (mm)	-34.844	22.304	24.576	110.61
AugWS (mm)	-38.35	24.49	23.08	126.95
WS (mm)	-185.05	668.3	686.9	1380.41
YLD (t/ha)	0.00	2.556	2.685	7.275

Bivariate Statistics

Figures 4.1-4.7 display the bivariate plots of soybean yields for each explanatory variable.

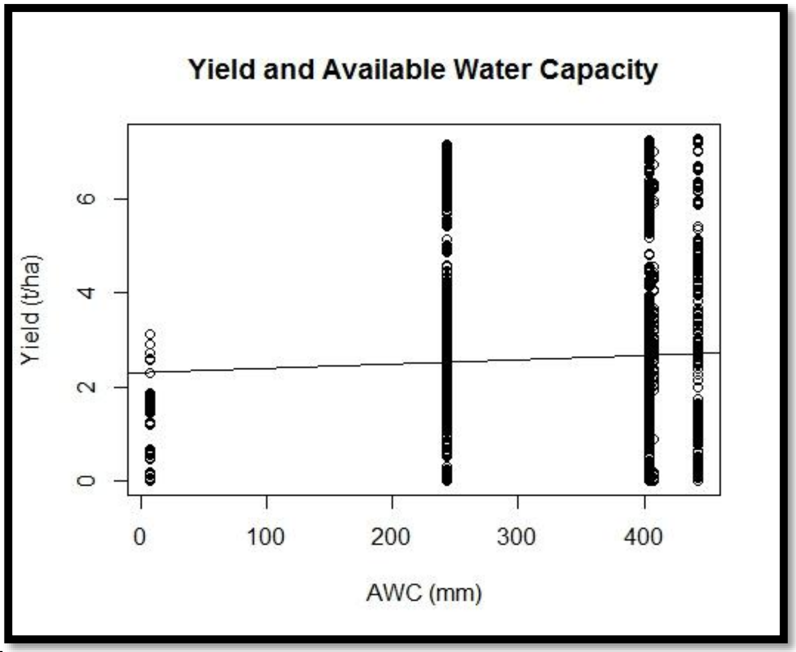


Figure 4.1 Bivariate Plot of Soybean Yield and Available Water Capacity

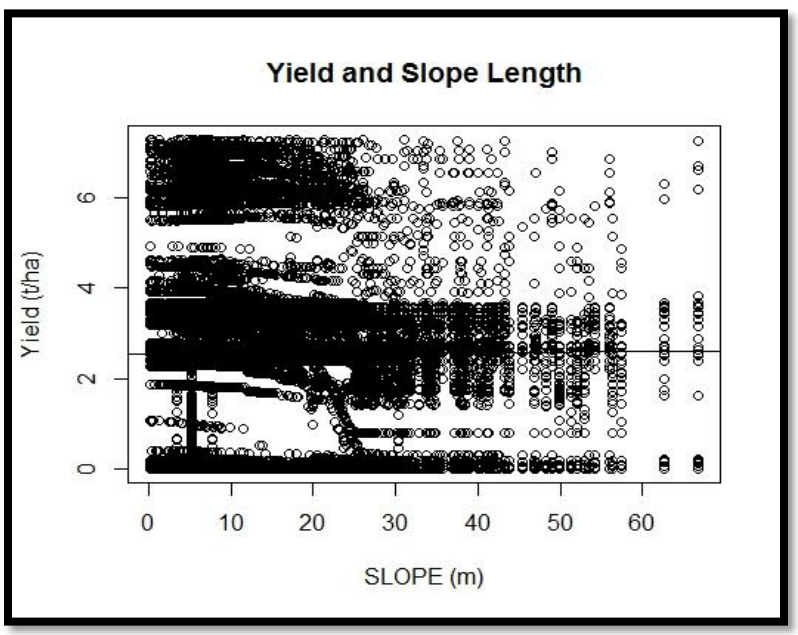


Figure 4.2 Bivariate Plot of Soybean Yield and Slope Length

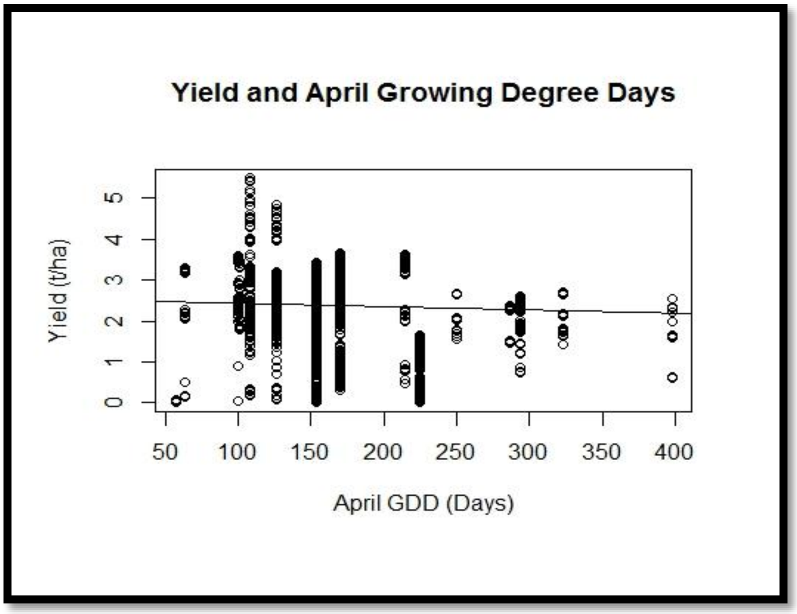


Figure 4.3 Bivariate Plot of Soybean Yield and April Growing Degree Days

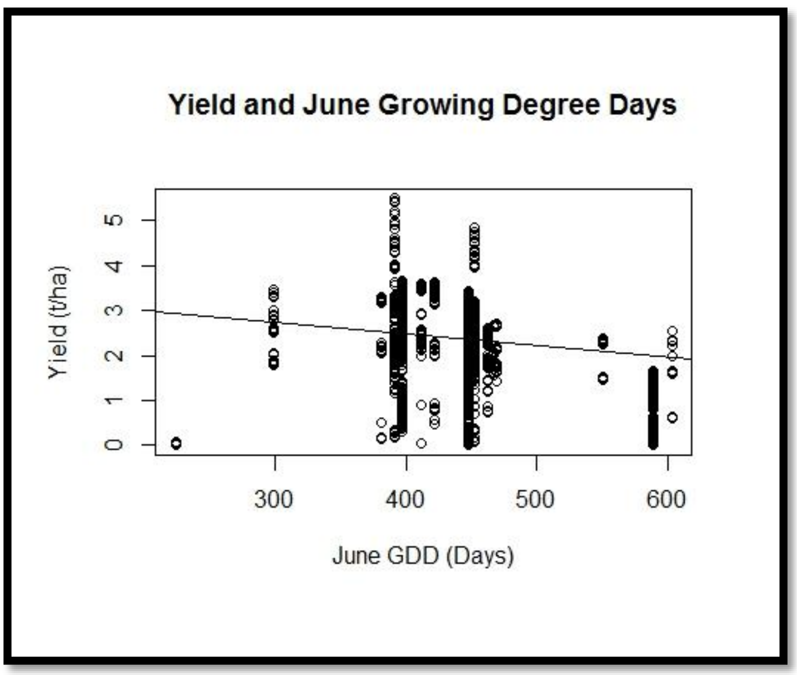


Figure 4.4 Bivariate Plot of Soybean Yield and June Growing Degree Days

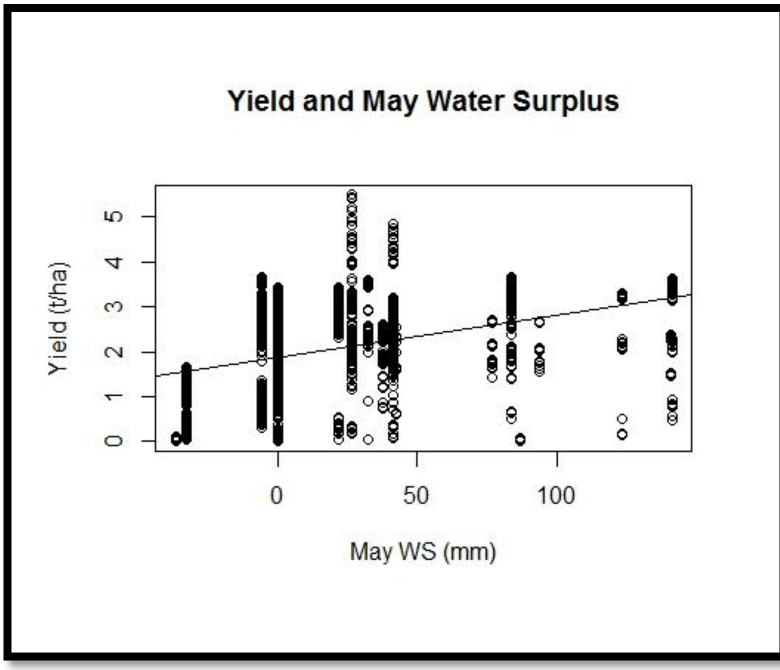


Figure 4.5 Bivariate Plot of Soybean Yield and May Water Surplus

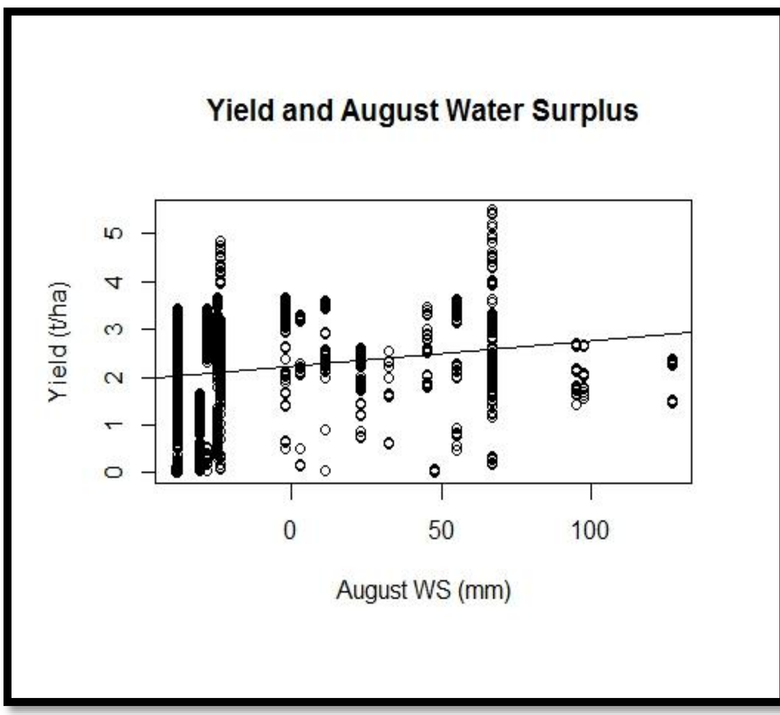


Figure 4.6 Bivariate Plot of Soybean Yield and August Water Surplus

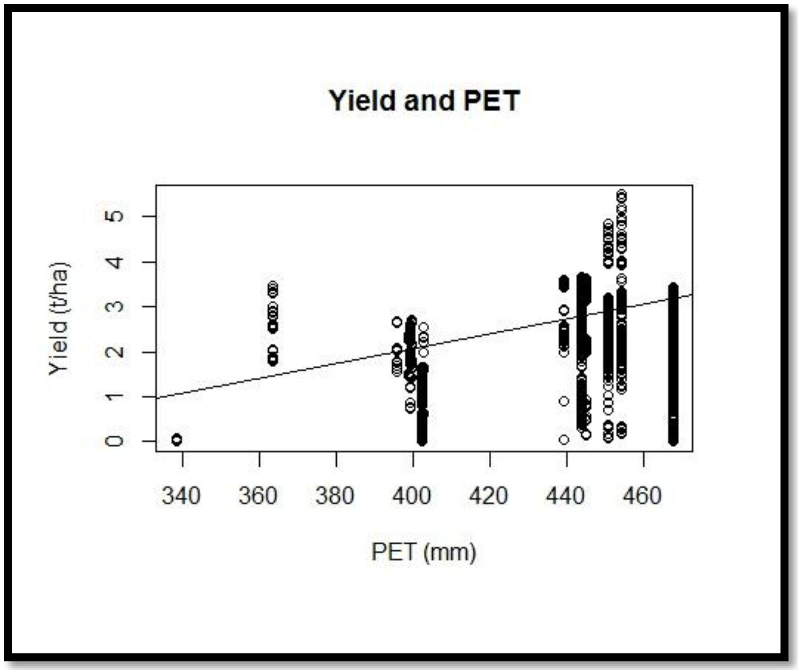


Figure 4.7 Bivariate Plot of Soybean Yield and Potential Evapotranspiration

Multivariate Statistics

The soybean regression model was able to explain 62 percent of the variance in yield. The mean yield for the model was 2.3 t/ha. The RMSE of the model was 0.75t/ha. The explanatory variables and their corresponding coefficients are listed in Table 4.4. All variables were significant at the 0.05 alpha level except for first year rotation soybean, second year rotation soybean, conservation tillage and no-till.

Table 4.4 Regression Model for Soybean

Soybean Model				
Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-7.601	0.03434	-225.64	<2e-16
SLOPE	-0.0138	0.0002805	-52.014	<2e-16
AWC	0.001136	0.00003131	37.008	<2e-16

Table 4.4 (Continued)

MayWS	0.006439	0.00007806	81.748	<2e-16
AugWS	0.004071	0.00008436	50.317	<2e-16
AprGDD	0.00581	0.00005133	113.2	<2e-16
JuneGDD	-0.007	0.0000495	-140.24	<2e-16
PET	0.02758	0.0000803	343.496	<2e-16
FYRCT	0.0239	0.008477	1.595	0.008
FYRS	0.002897	0.005775	1.332	0.60
SYRCT	-0.1326	0.009564	-13.364	<2e-16
SYRS	0.0028	0.005774	2.126	0.61
SYRWW	-0.5487	0.01223	278.986	<2e-16
TillCSV	-0.00207	0.006054	-0.739	0.73
TillNT	-0.00039	0.006044	-1.738	0.94

4.2 Corn Model

Univariate Statistics

Table 4.5 details the summary statistics for each of the explanatory variables along with the dependent variable.

Table 4.5 Summary Statistics for each Variable

Variable	Min	Mean	Median	Max
Precipitation (mm)	217.3	1063.8	988.6	2029.7
Nitrogen (Kg/ha)	100	159.7	156.0	223
Slope Length (meters)	0.1907	11.4117	8.5028	66.7935
AWC (millimeters)	7.428	291.121	243.2	443.1
AprilTS (Days)	2.188	8.764	9.722	15.488
MayTS (Days)	0.2002	1.9296	1.2245	6.0975
JulyTS (Days)	0.1582	4.2992	3.5535	16.6153
AugTS (Days)	0.9094	3.792	2.6041	8.6627
AprilGDD (Days)	153.2	260.5	227.5	426.1
MayGDD (Days)	269.3	410.6	395.1	593.9
JuneGDD (Days)	358.7	514.9	511.6	610.1
JulyGDD (Days)	466.5	579.6	588.4	645.3

Table 4.2 (Continued)

AugGDD (Days)	373.2	540.8	549.4	636.4
GDD (Days)	2098	3101	3096	3815
PET (mm)	363.4	424.1	418.5	467.6
AprWS (mm)	-28.02	31.439	21.642	143.429
MayWS (mm)	-41.22	47.78	40.27	141.25
JuneWS (mm)	-58.2	21.02	30.36	114.29
JulyWS (mm)	-35.844	22.304	4.576	110.61
AugWS (mm)	-38.35	19.83	11.10	126.95
WS (mm)	-195.05	617.3	521.0	1380.41
YLD (t/ha)	0.142	7.976	8.442	15.984

Bivariate Statistics

Figures 4.8-4.15 display the explanatory variables plotted against crop yield along with the regression line.

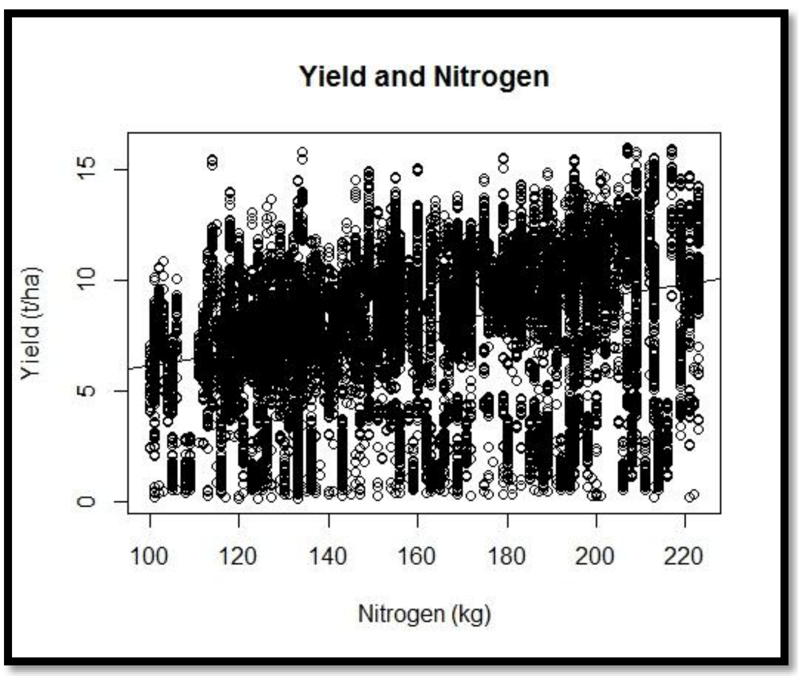


Figure 4.8 Bivariate Plot of Corn Yield and Nitrogen

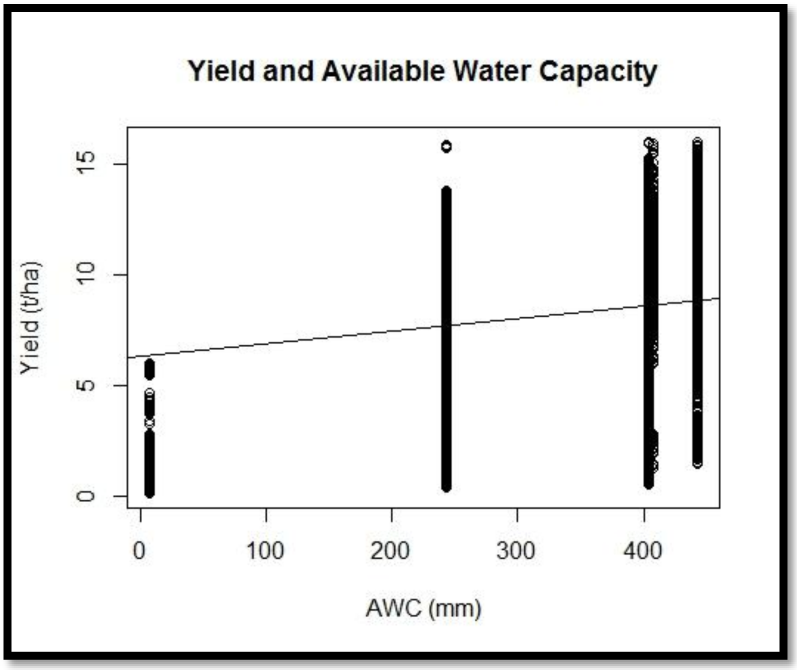


Figure 4.9 Bivariate Plot of Corn Yield and Available Water Capacity

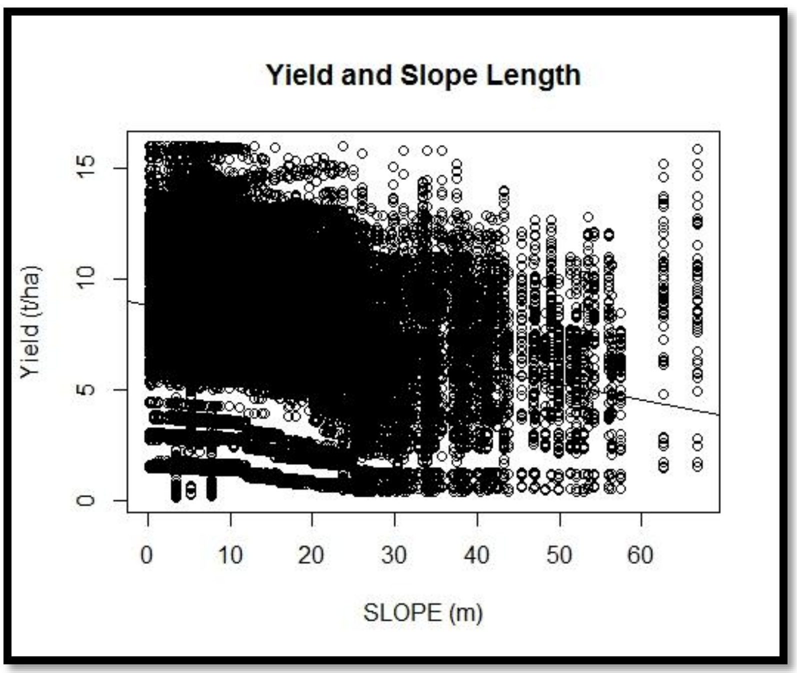


Figure 4.10 Bivariate Plot of Corn Yield and Slope Length

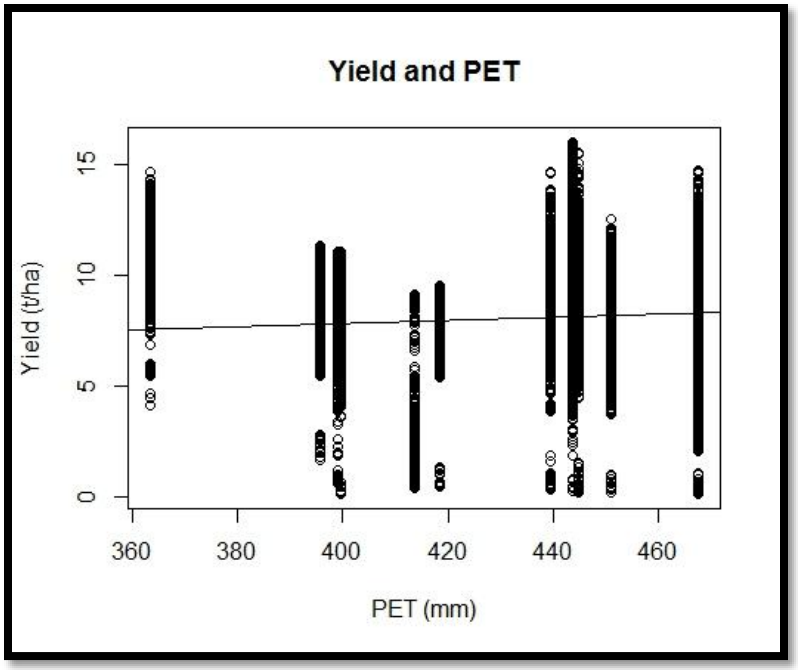


Figure 4.11 Bivariate Plot of Corn Yield and Potential Evapotranspiration

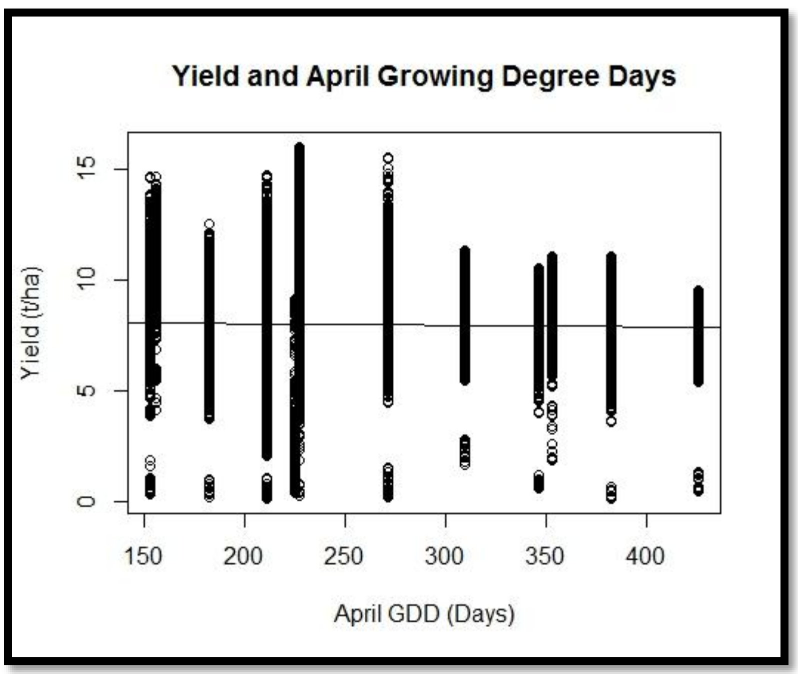


Figure 4.12 Bivariate Plot of Corn Yield and April Growing Degree Days

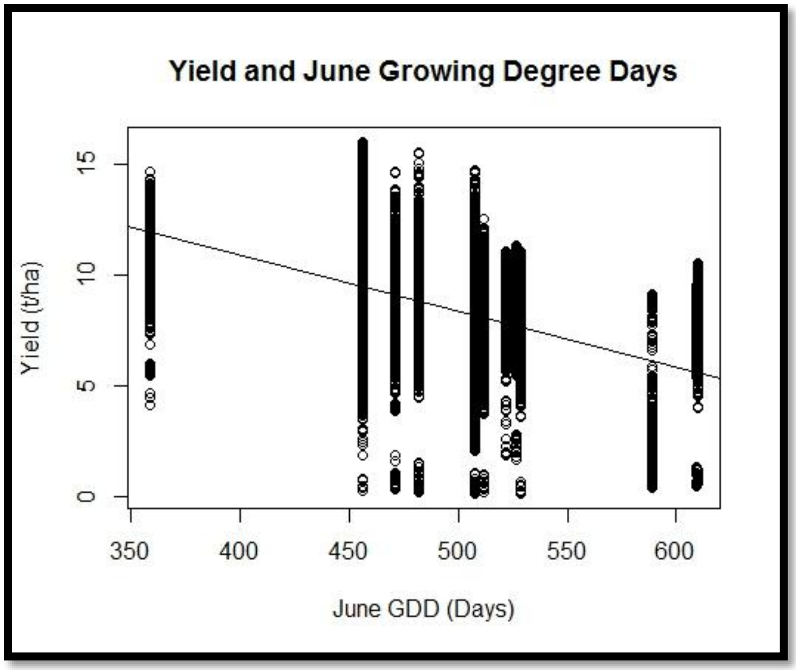


Figure 4.13 Bivariate Plot of Yield and June Growing Degree Days

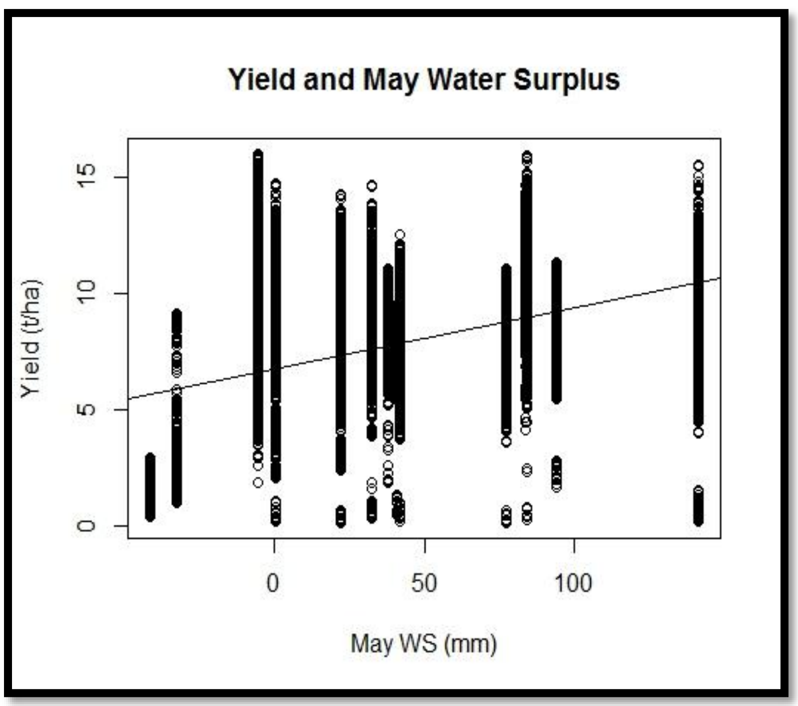


Figure 4.14 Bivariate Plot of Yield and May Water Surplus

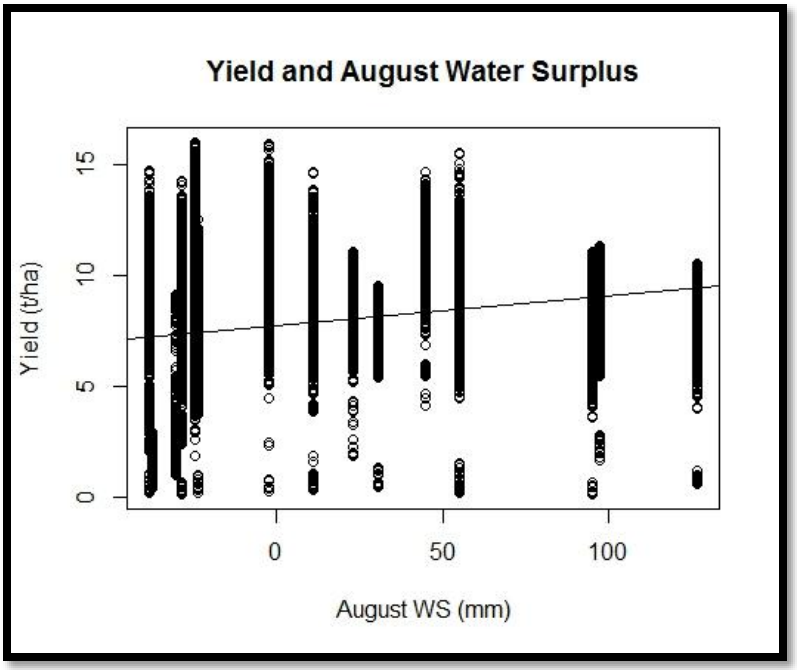


Figure 4.15 Bivariate Plot of Corn Yield and August Water Surplus

Multivariate Statistics

The corn regression model was able to explain 70 percent of the variance in yield. The mean yield for the model was 7.976 tons per hectare. The root mean square error of the model was 1.679 tons per hectare. The explanatory variables and their corresponding coefficients are listed in Table 4.6. All variables were significant at the 0.05 alpha level.

Table 4.6 Corn Regression Mode

Coefficients	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-2.396	0.1243	-19.282	< 2e-16
N	0.03185	0.000166	191.911	< 2e-16
SLOPE	-0.07871	0.000598	-131.643	< 2e-16
AWC	0.007305	6.9E-05	105.89	< 2e-16
AprGDD	0.01615	0.000108	149.456	< 2e-16

Table 4.6 (Continued)

JuneGDD	-0.03512	0.000124	-282.624	< 2e-16
MayWS	0.009416	0.000255	36.918	< 2e-16
AugWS	0.009047	0.000331	27.352	< 2e-16
PET	0.0401	0.000298	134.438	< 2e-16
FYRCT	0.2499	0.01817	13.751	< 2e-16
FYRS	0.1097	0.01152	9.528	< 2e-16
SYRCT	0.08394	0.01429	5.873	<4.3E-09
SYRS	0.3518	0.0141	24.946	< 2e-16
TillCSV	0.1351	0.01302	10.375	< 2e-16
TillNT	0.1211	0.01303	9.292	< 2e-16

4.3 Switchgrass Model

Univariate Statistics

Table 4.7 details the summary statistics for each of the explanatory variables along with the dependent variable.

Table 4.7 Switchgrass Explanatory Variable Summary Statistics

Variable	Min	Mean	Median	Max
Precipitation (mm)	217.3	1063.8	988.6	2029.7
Nitrogen (Kg/ha)	1.00	66.89	68.00	125.00
Slope Length (meters)	0.1907	11.412	8.503	66.794
AWC (millimeters)	7.428	291.121	243.2	443.1
AprilTS (Days)	5.116	15.442	16.996	27.145
MayTS (Days)	0.3594	5.0416	3.7537	14.7176
JulyTS (Days)	0.3218	11.6303	11.6164	29.900
AugTS (Days)	0.087	8.630	8.274	22.403
AprilGDD (Days)	32.40	144.29	111.30	347.74
MayGDD (Days)	145.3	279.5	271.1	469.9
JuneGDD (Days)	238.7	394.5	391.6	542.9
JulyGDD (Days)	342.5	459.0	464.4	591.3
AugGDD (Days)	249.2	414.2	425.4	512.4
GDD (Days)	1220	2129	2120	2999
PET (mm)	363.4	424.1	418.5	467.6
AprWS (mm)	-31.449	33.439	22.642	143.429
MayWS (mm)	-41.22	50.78	40.27	141.25

Table 4.7 (Continued)

JuneWS (mm)	-58.2	27.02	42.36	114.29
JulyWS (mm)	-35.844	24.304	9.576	110.61
AugWS (mm)	-38.35	18.83	2.10	126.95
WS (mm)	-185.05	621.3	686.0	1380.41
YLD (t/ha)	0.00	8.804	9.654	23.205

Bivariate Statistics

Figures 4.16-4.23 display the plots of the explanatory variables against switchgrass yield with the regression line.

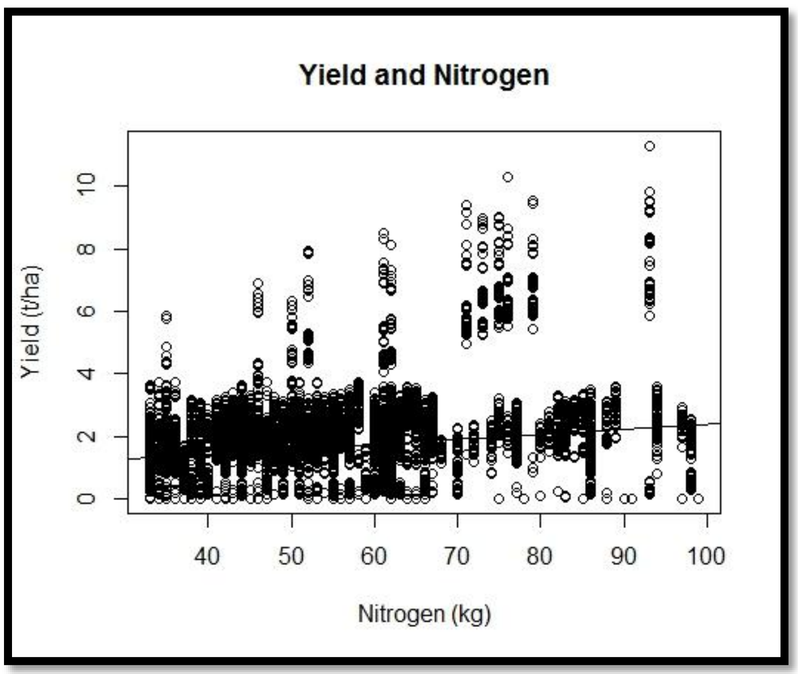


Figure 4.16 Bivariate Plot of Switchgrass Yield and Nitrogen

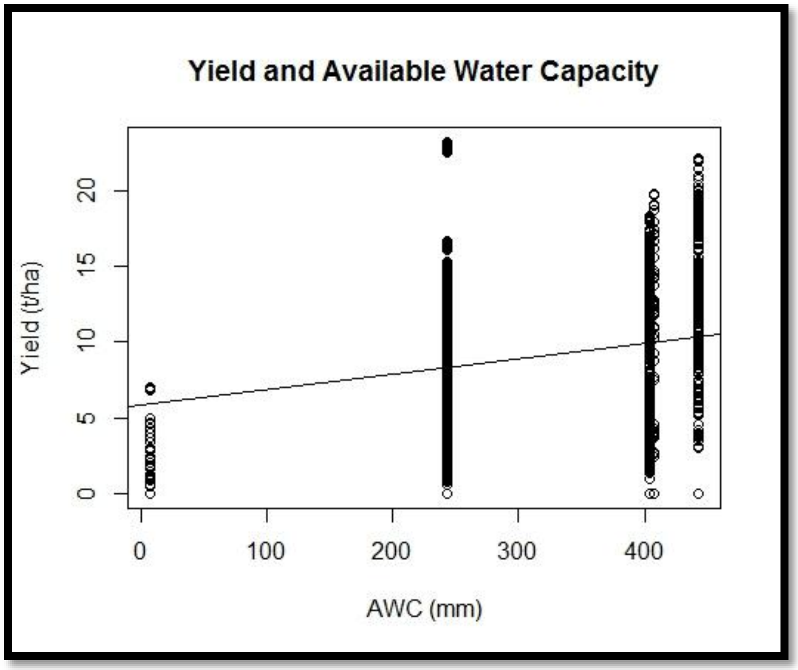


Figure 4.17 Bivariate Plot of Switchgrass Yield and Available Water Capacity

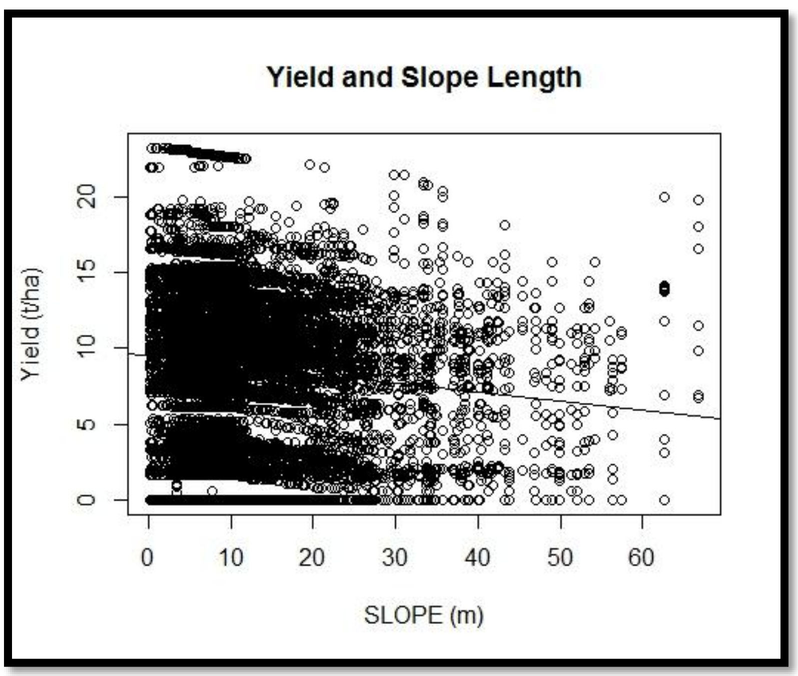


Figure 4.18 Bivariate Plot of Switchgrass Yield and Slope Length

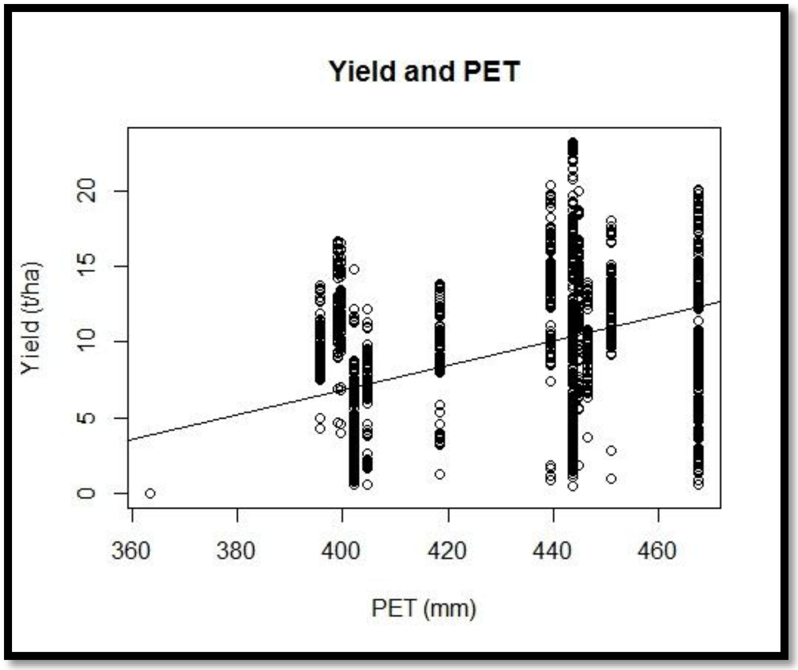


Figure 4.19 Bivariate Plot of Switchgrass Yield and Potential Evapotranspiration

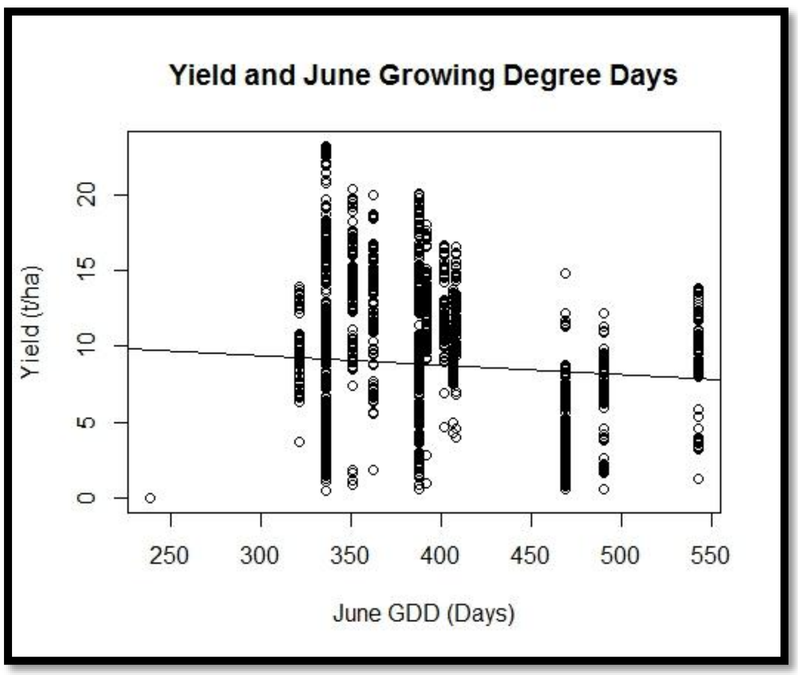


Figure 4.20 Bivariate Plot of Switchgrass Yield and June Growing Degree Days

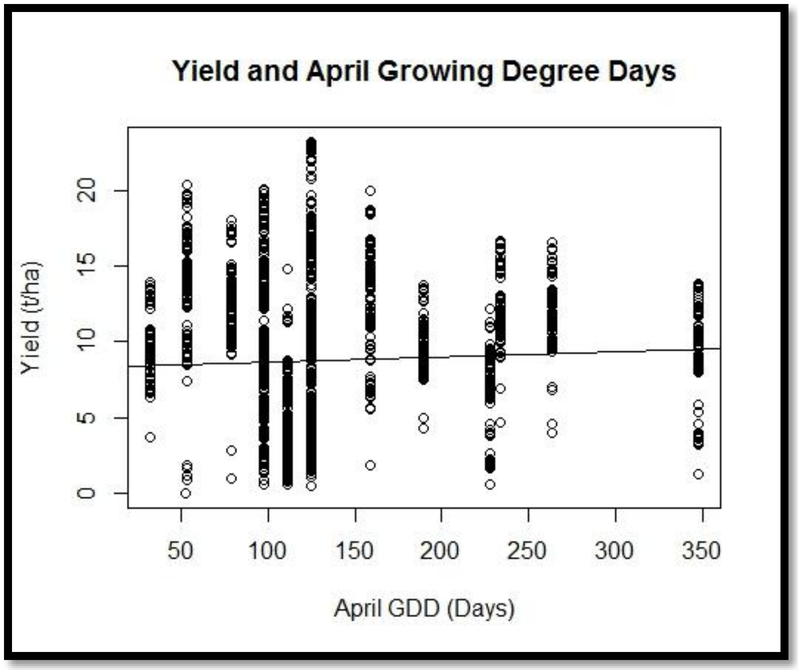


Figure 4.21 Bivariate Plot of Switchgrass Yield and April Growing Degree Days

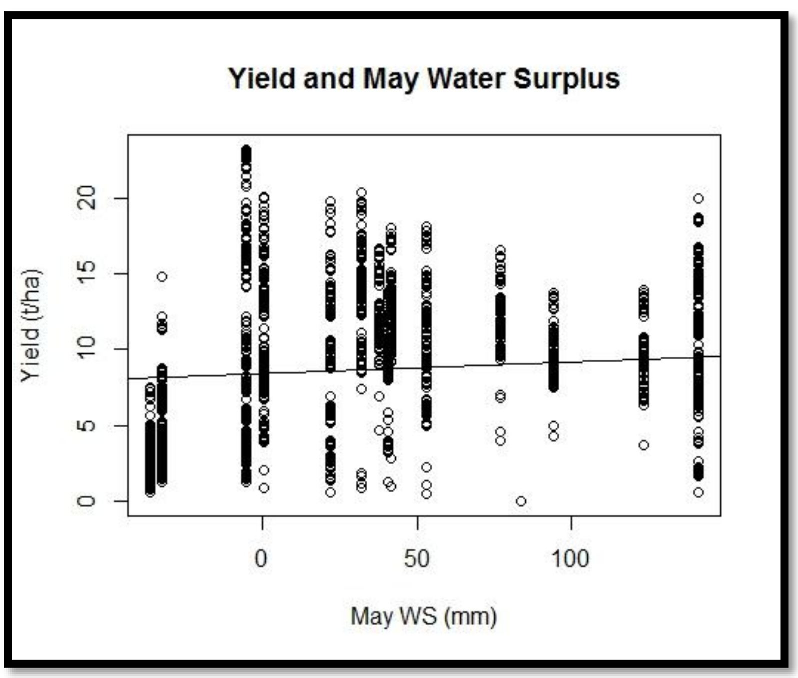


Figure 4.22 Bivariate Plot of Switchgrass Yield and May Water Surplus

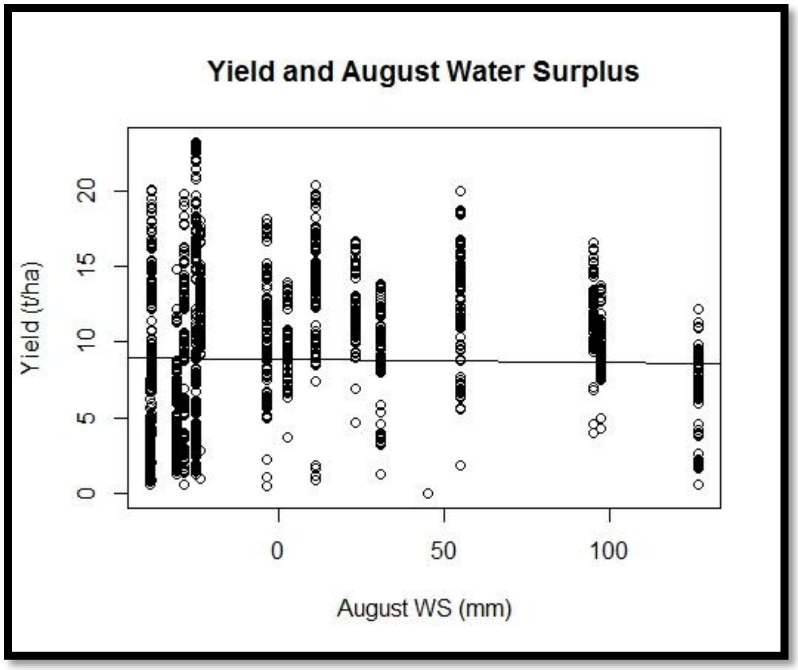


Figure 4.23 Bivariate Plot of Switchgrass Yield and August Water Surplus

Multivariate Statistics

The switchgrass regression model was able to explain 77 percent of the variance in yield. The mean yield for the model was 8.804 tons per hectares. The root mean square error of the model was 2.363 tons per hectare. The explanatory variables and their corresponding coefficients are listed in table 4.8. All the variables are significant at the 0.05 alpha level.

Table 4.8 Switchgrass Regression Model

Switchgrass Model				
Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-43.48	0.4133	-105.197	< 2e-16
N	0.08068	0.000581	138.871	< 2e-16
SLOPE	-0.06845	0.002251	-30.405	< 2e-16
AWC	0.01119	0.000259	43.273	< 2e-16

Table 4.8 (Continued)

AprGDD	0.004744	0.000437	10.857	< 2e-16
JuneGDD	-0.01309	0.000442	-29.601	< 2e-16
MayWS	-0.01681	0.000859	-19.572	< 2e-16
AugWS	0.0413	0.001129	36.588	< 2e-16
PET	0.115	0.001003	114.698	< 2e-16

4.4 Alfalfa Model

Univariate Statistics

Table 4.9 displays the summary statistics for each of the explanatory variables along with the dependent variable.

Table 4.9 Alfalfa Explanatory Variable Summary Statistics

Variable	Min	Mean	Median	Max
Precipitation (mm)	217.3	1063.8	988.6	2029.7
Nitrogen (Kg/ha)	0	0	0	0
Slope Length (meters)	0.1907	11.4117	8.5028	66.7935
AWC (millimeters)	7.428	291.121	243.2	443.1
AprilTS (Days)	0.1026	2.917	1.809	9.259
MayTS (Days)	0.054	0.400	0.230	0.986
JulyTS (Days)	0	0	0	0
AugTS (Days)	0	0	0	0
AprilGDD (Days)	216.2	371.1	347.5	568.4
MayGDD (Days)	393.3	525.7	506.4	717.9
JuneGDD (Days)	478.7	633.4	627.5	782.9
JulyGDD (Days)	590.5	706.0	708.0	839.3
AugGDD (Days)	497.2	661.6	672.0	760.4
GDD (Days)	3167	4211	3167	5137
PET (mm)	363.4	424.1	418.5	467.6
AprWS (mm)	-31.449	33.439	21.642	143.429
MayWS (mm)	-36.22	51.78	40.27	141.25
JuneWS (mm)	-58.2	27.02	42.36	114.29
JulyWS (mm)	-35.844	25.304	38.576	110.61
AugWS (mm)	-38.35	18.83	11.10	126.95

Table 4.9 (Continued)

WS (mm)	-185.05	623.3	686.0	1380.41
YLD (t/ha)	0.481	6.635	7.00	9.02

Bivariate Statistics

Figures 4.24-4.30 display the explanatory variables plotted against the alfalfa yield along with the regression line.

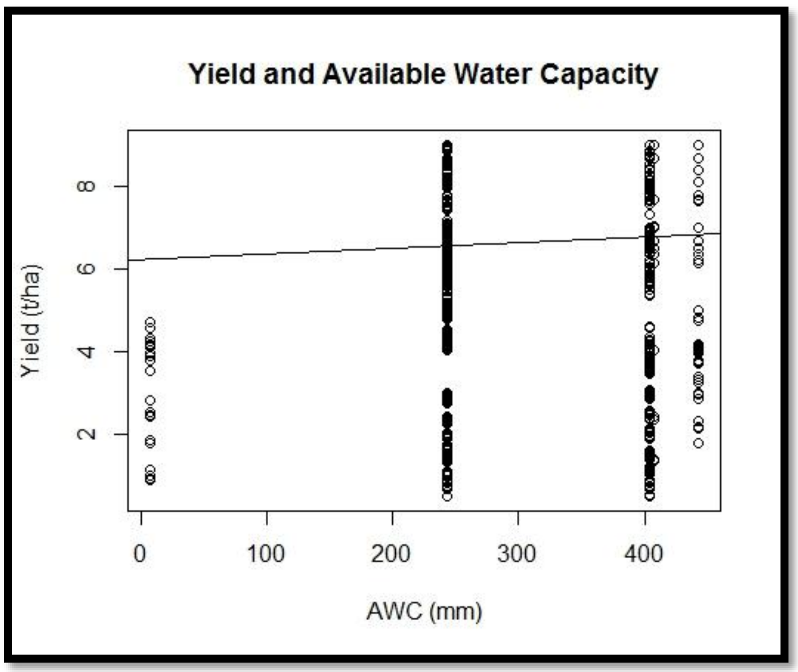


Figure 4.24 Bivariate Plot of Alfalfa Yield and Available Water Capacity

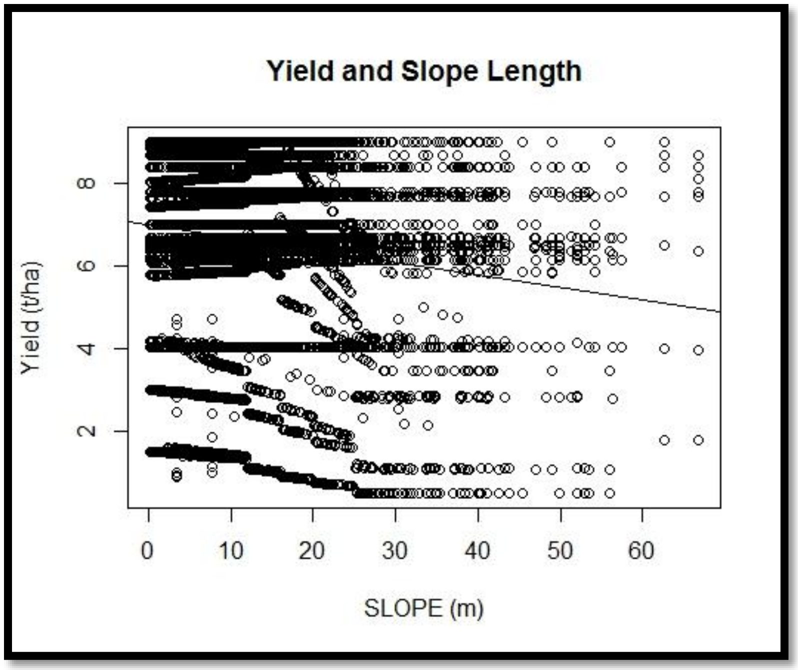


Figure 4.25 Bivariate Plot of Alfalfa Yield and Slope Length

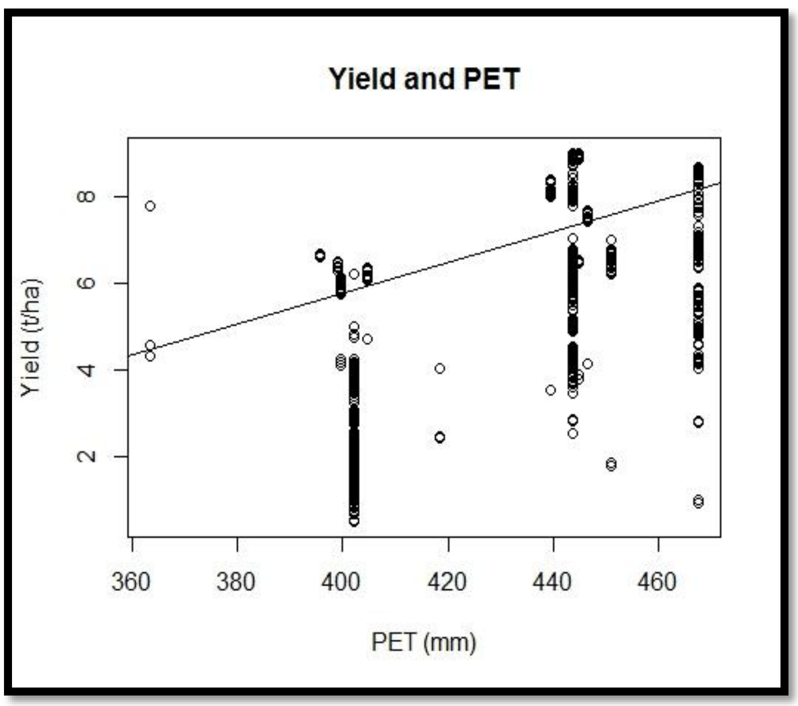


Figure 4.26 Bivariate Plot of Alfalfa Yield and Potential Evapotranspiration

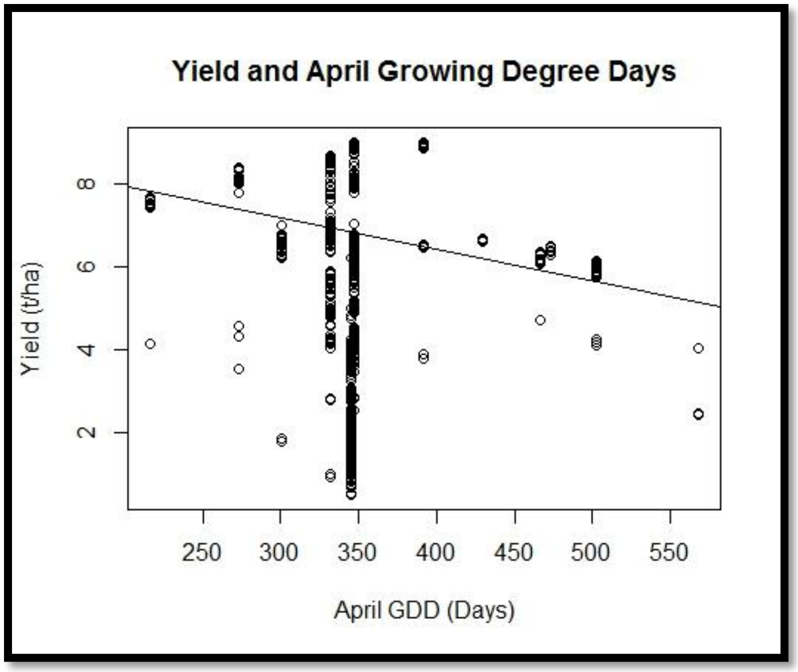


Figure 4.27 Bivariate Plot of Alfalfa Yield and April Growing Degree Days

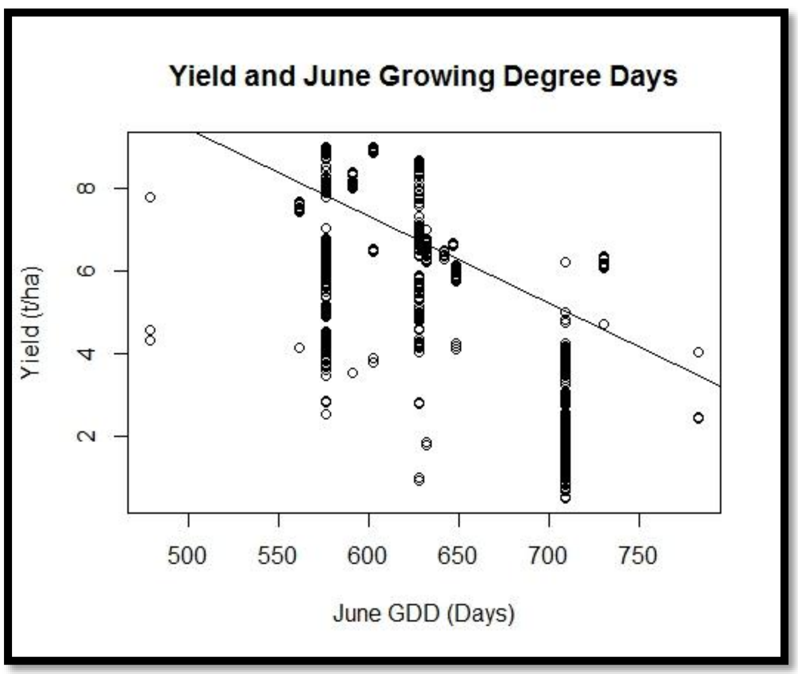


Figure 4.28 Bivariate Plot of Alfalfa Yield and June Growing Degree Days

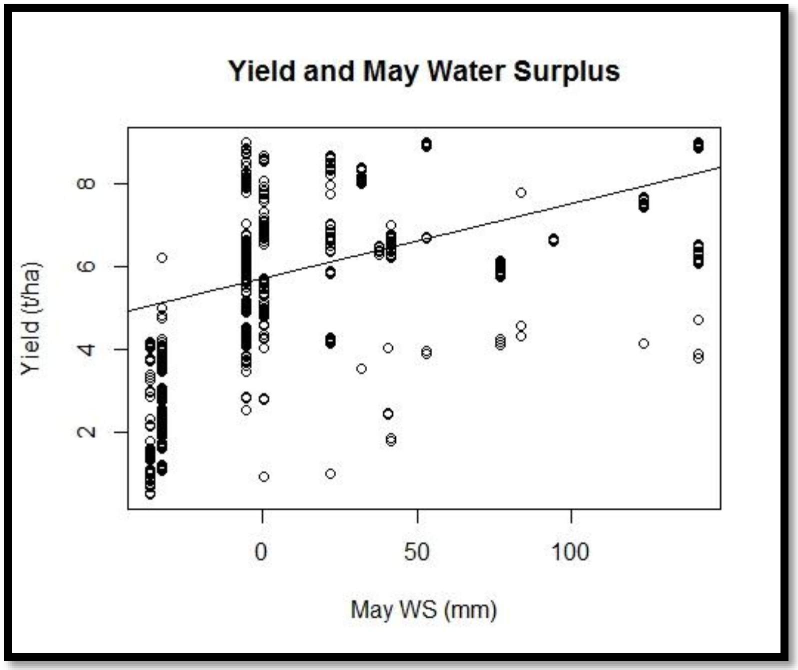


Figure 4.29 Bivariate Plot of Alfalfa Yield and May Water Surplus

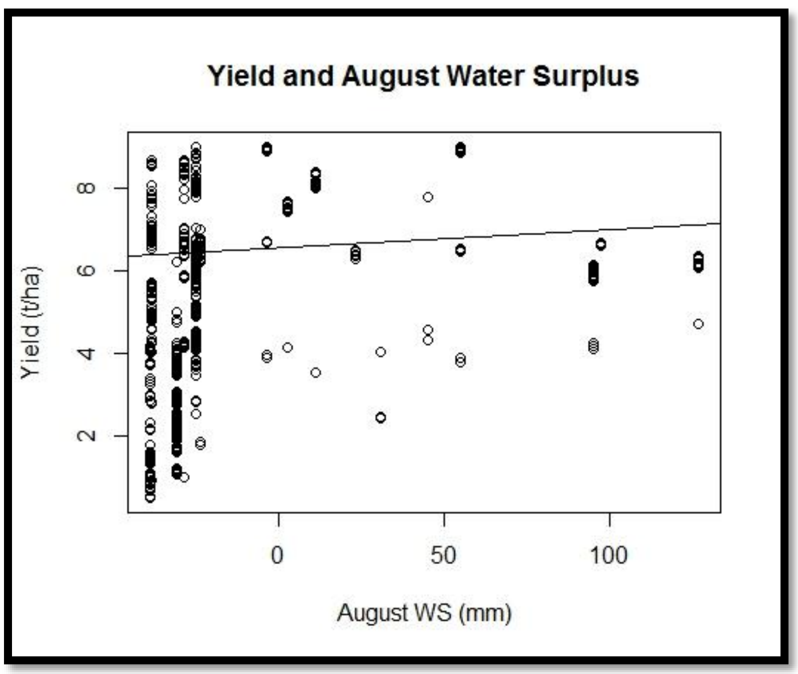


Figure 4.30 Bivariate Plot of Alfalfa Yield and August Water Surplus

Multivariate Statistics

The alfalfa regression model was able to explain 84 percent of the variance in yield. The mean yield for the model was 6.635 tons per hectares. The RMSE of the model was 0.9077 t/ha. The explanatory variables and their corresponding coefficients are listed in Table 4.10. All the variables are significant at the 0.05 alpha level except for the tillage dummy variables.

Table 4.10 Alfalfa RegressionModel

Alfalfa Model				
Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-1.9209214	0.1629349	-11.79	<2e-16
SLOPE	-0.0354903	0.0008595	-41.292	<2e-16
AWC	0.0022667	0.0001003	22.609	<2e-16
AprGDD	0.0062184	0.0001597	38.943	<2e-16
JuneGDD	-0.02663	0.0001649	-161.523	<2e-16
MayWS	-0.0029632	0.0003372	-8.787	<2e-16
AugWS	0.0208292	0.0004598	45.3	<2e-16
PET	0.053334	0.0003849	138.572	<2e-16

4.5 Cotton Model

Univariate Statistics

Table 4.11 displays the summary statistics for the explanatory variables along with the dependent variable in the cotton model.

Table 4.11 Cotton Explanatory Variable Summary Statistics

Variable	Min	Mean	Median	Max
Precipitation (mm)	217.3	1063.8	988.6	2029.7
Nitrogen (Kg/ha)	33.0	55.1	53.0	99.0
Slope Length (meters)	0.1907	11.4117	8.5028	66.7935
AWC (millimeters)	7.428	291.121	243.2	443.1
AprilTS (Days)	8.335	19.971	21.866	29.447
MayTS (Days)	0.2791	9.134	6.918	23.6031

Table 4.11 (Continued)

JulyTS (Days)	0	0	0	0
AugTS (Days)	0	0	0	0
AprilGDD (Days)	8.1	104.2	75.1	262.6
MayGDD (Days)	73.3	238.2	228.0	473.7
JuneGDD (Days)	231.4	361.3	318.6	589.0
JulyGDD (Days)	300.2	424.5	393.4	610.4
AugGDD (Days)	269.4	380.2	332.4	563.1
GDD (Days)	1159	1878	1600	3322
PET (mm)	363.4	424.1	418.5	467.6
AprWS (mm)	-31.449	48.439	52.642	140.429
MayWS (mm)	-36.22	69.78	79.27	141.25
JuneWS (mm)	-58.2	46.02	50.36	114.29
JulyWS (mm)	-35.844	38.304	38.576	110.61
AugWS (mm)	-38.35	30.83	30.10	126.95
WS (mm)	-185.05	804.3	935.3	1380.41
YLD (t/ha)	0.00	1.658	1.735	11.307

Bivariate Statistics

Figures 4.31-4.38 display the plots of the explanatory variables against the cotton yield along with the regression line.

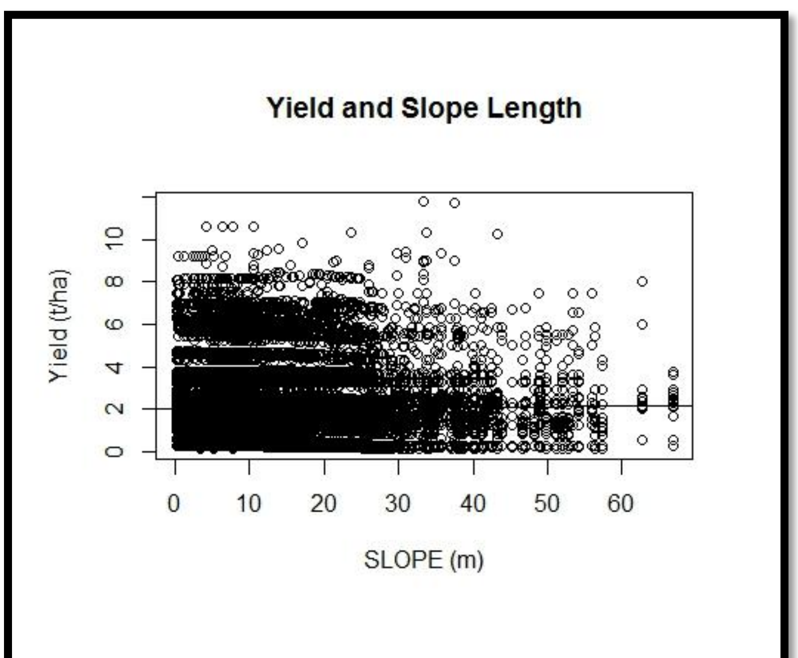


Figure 4.31 Bivariate Plot of Cotton Yield and Slope Length

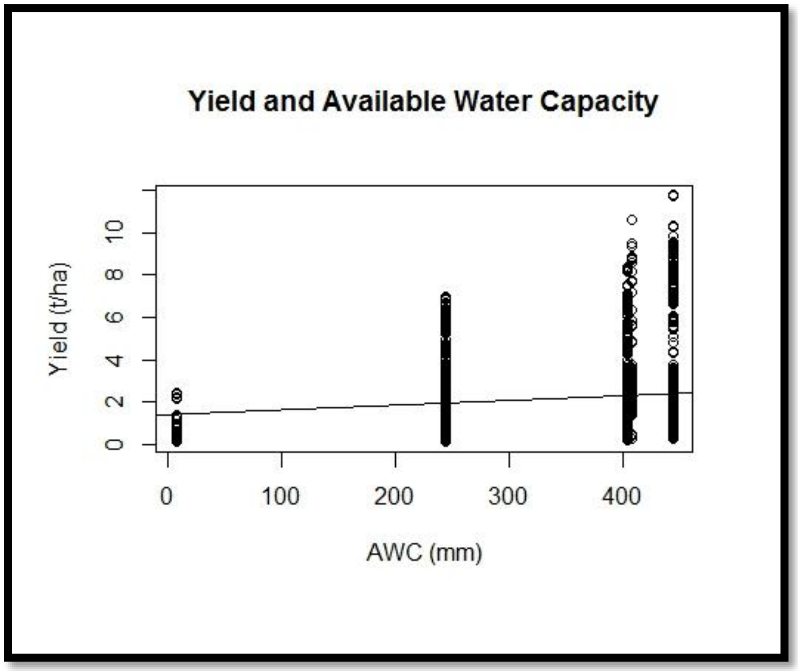


Figure 4.32 Bivariate Plot of Cotton Yield and Available Water Capacity

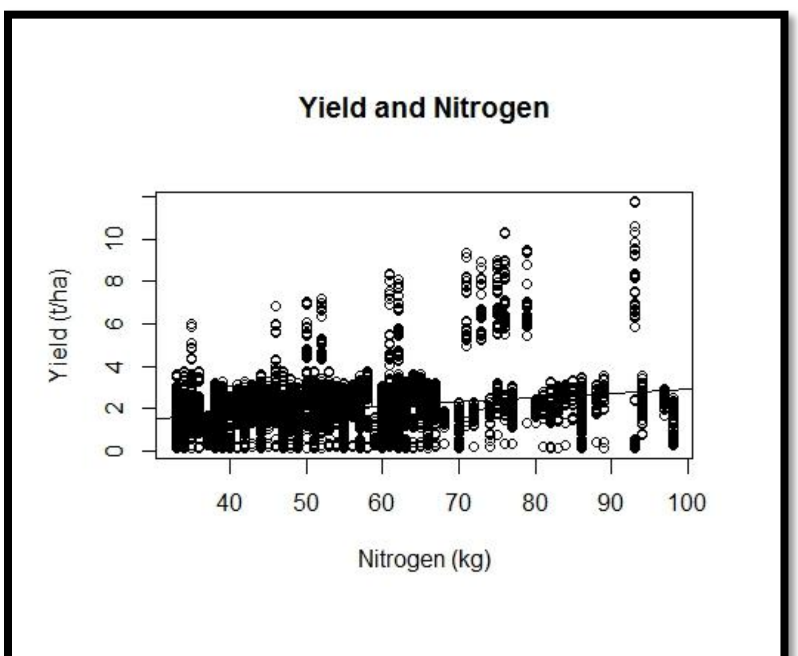


Figure 4.33 Bivariate Plot of Cotton Yield and Nitrogen

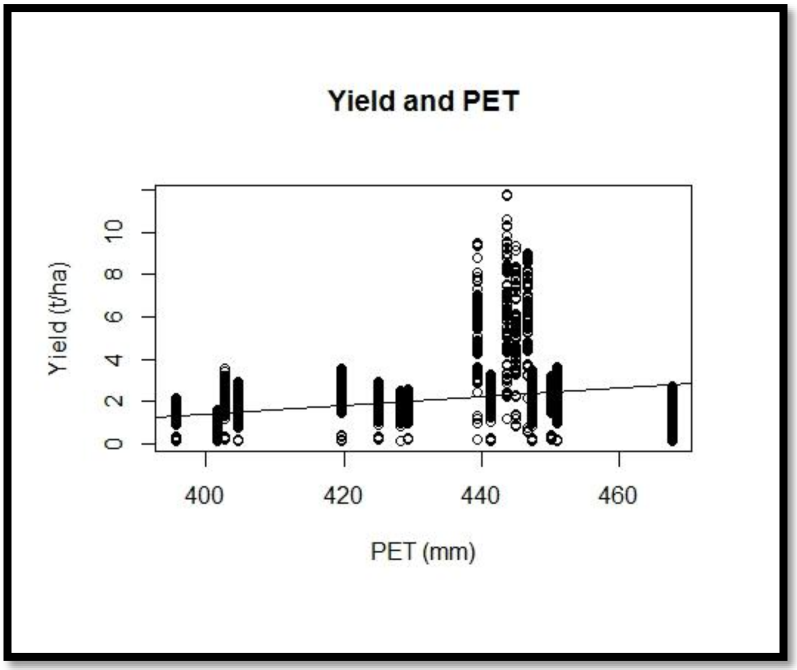


Figure 4.34 Bivariate Plot of Cotton Yield and Potential Evapotranspiration

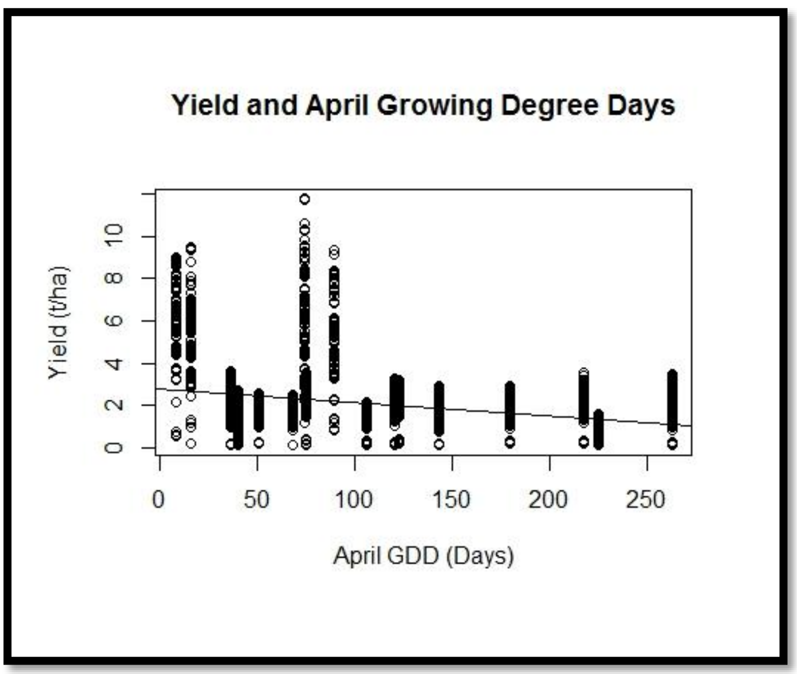


Figure 4.35 Bivariate Plot of Cotton Yield and April Growing Degree Days

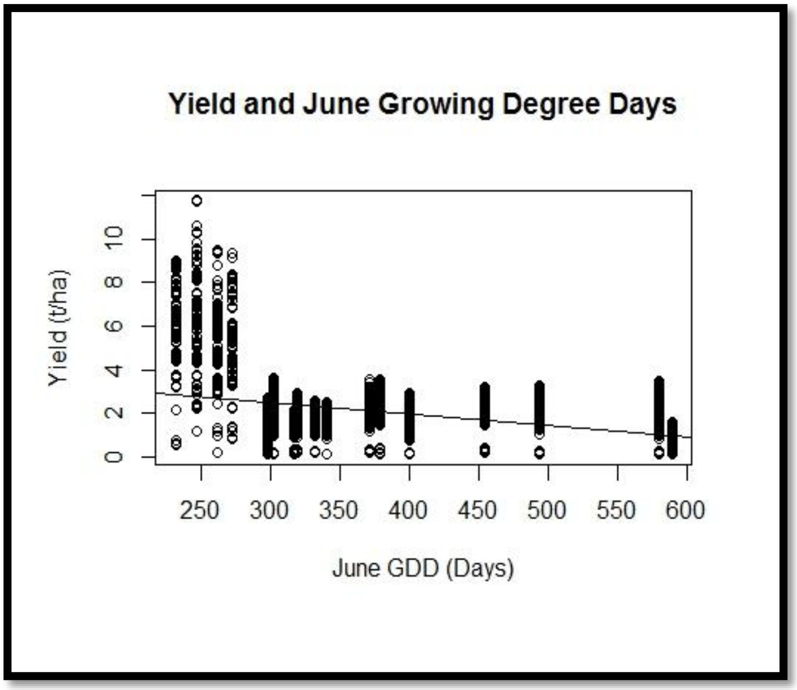


Figure 4.36 Bivariate Plot of Cotton Yield and June Growing Degree Days

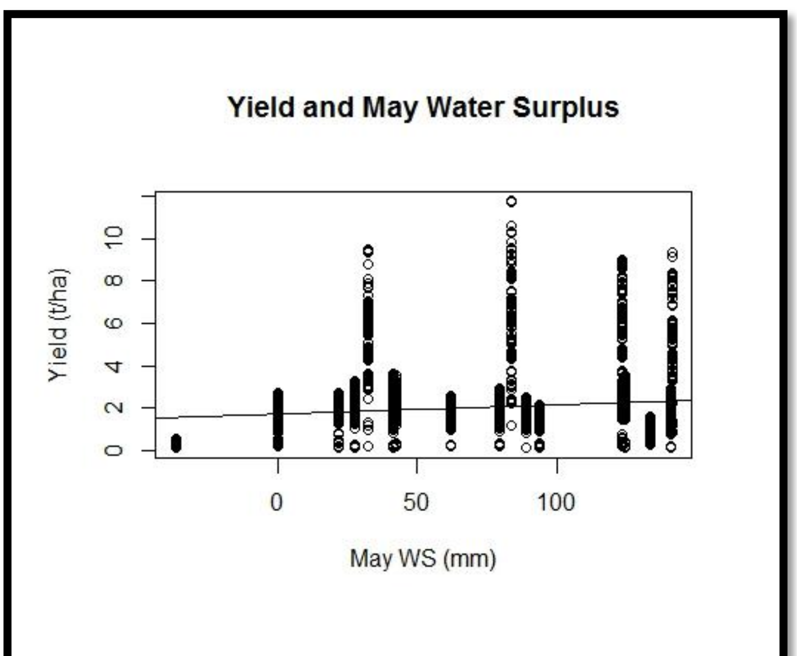


Figure 4.37 Bivariate Plot of Cotton Yield and May Water Surplus

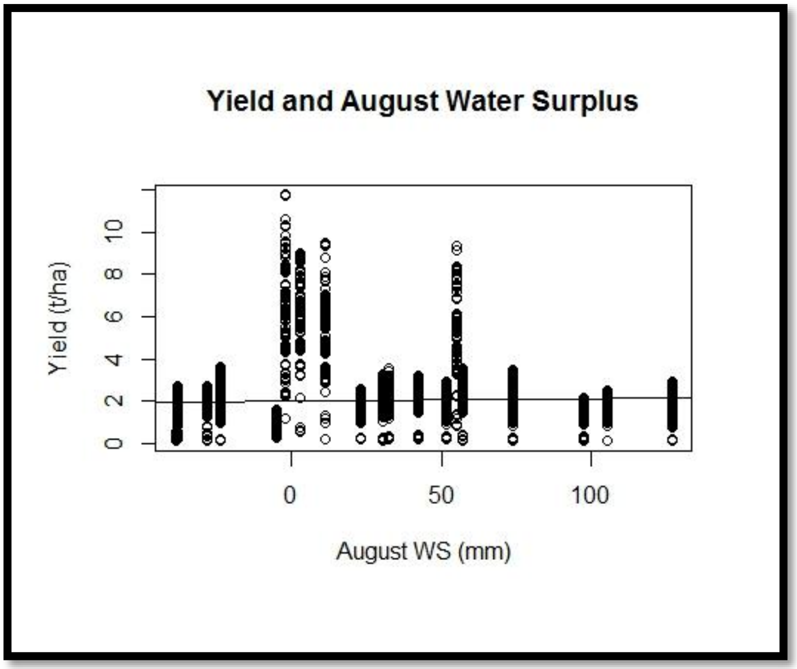


Figure 4.38 Bivariate Plot of Cotton Yield and August Water Surplus

Multivariate Statistics

The cotton regression model was able to explain 48 percent of the variance in yield. The mean yield for the model was 2.037 t/ha. The RMSE of the model was 0.8444 tons per hectare. The explanatory variables and their corresponding coefficients are listed in Table 4.12. All variables were significant at the 0.05 alpha level except for conservation tillage.

Table 4.12 Cotton Regression Model

Cotton Model				
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-7.634	0.081	-93.708	< 2e-16
SLOPE	-0.008	0.000	-17.118	< 2e-16

Table 4.12 (Continued)

N	0.018	0.000	91.1	< 2e-16
AWC	0.002	0.000	59.743	< 2e-16
AprGDD	0.004	0.000	46.826	< 2e-16
JuneGDD	-0.005	0.000	-109.944	< 2e-16
MayWS	0.006	0.000	71.542	< 2e-16
AugWS	0.001	0.000	12.997	< 2e-16
PET	0.021	0.000	125.51	< 2e-16
FYRCT	0.062	0.008	8.191	2.62E-16
FYRS	0.168	0.011	15.613	< 2e-16
SYRCT	-0.175	0.012	-15.166	< 2e-16
SYRS	0.065	0.008	8.317	< 2e-16
TillCSV	-0.006	0.007	-0.792	0.4283
TillNT	-0.016	0.007	-2.166	0.0303

CHAPTER 5

DISCUSSION

5.1 Discussion of Research Questions

1) What factors can significantly explain crop yields in Big Creek Watershed?

Thirty explanatory variables were originally created for the metamodel. This was reduced to fourteen explanatory variables that have been found to significantly explain crop yields for Big Creek Watershed. These variables were chosen utilizing an all-subsets regression function. The variables were further reduced by analyzing the multicollinearity of the variables and the variance inflation factor (VIF) of these variables. The VIF is an index that is used to measure the degree of multicollinearity between variables. In addition, variables were added and dropped depending on how much they increased R^2 and decreased RMSE. The final list of variables were: annual potential evapotranspiration, nitrogen fertilizer, slope length, available water capacity, April growing degree days, June growing degree days, May water surplus, August water surplus, rotation dummy variables, and tillage dummy variables. The dummy variables were not significant for every model, but they were still included in the final model due to the importance of farm management practices on crop yields and in the agent-based model to which these metamodels are to be applied. The variables were the same for each of the models in order to measure the effects of the variables on each of the crops. Each of the explanatory variables will be discussed to detail their relationship with yield.

Nitrogen Fertilizer

Nitrogen fertilizer was found to have a positive influence on crop yields across all of the models. This is validated both by the coefficients in the model and by theory. Increasing nitrogen fertilizer will lead to higher yields, but this is the case up to a threshold where crop yields will no longer respond to increases in nitrogen fertilizer. Due to that reason, this research narrowed the range of fertilizer application for each crop as discussed in the methods section. By narrowing the range, this variable had a linear relationship with yield. Figures 4.8, 4.16 and 4.33 display the relationship between nitrogen and crop yield. Nitrogen had the largest influence on switchgrass yields with a coefficient of 0.608 compared to 0.341 for corn and 0.24 for cotton (Table 4.2). Switchgrass nitrogen amount was varied with a larger range than corn or cotton and this could be an explanation as to why nitrogen had the largest influence on switchgrass yields. In these models yields responded strongly to increases in nitrogen fertilizer. This can be seen in corn yields where a nitrogen fertilizer of 68 kg/ha lead to yields of around 5 t/ha, whereas, increasing nitrogen fertilizer to 174 kg/ha lead to yields of around 10 t/ha. In those simulations the weather was the same and the only variable that changed between the simulations was the fertilizer amount. A potential explanation is that the SWAT model is sensitive to changes in nitrogen fertilizer amounts.

Slope Length

Slope length was found to have a negative influence on crop yields across all of the models. Crop yield and slope length had a negative relationship and this is confirmed in the coefficients and in theory. Figures 4.2, 4.10, 4.18, 4.25, and 4.31 detail the relationship between slope length and crop yield. For the majority of the crops, yields start to decrease with slope length greater than 30 meters. Slope had the largest influence on corn yields with a coefficient of -0.226 compared to -0.135 for switchgrass, -0.145 for alfalfa, -0.105 for soybeans, and -0.046 for cotton (Table 4.2). This would suggest that switchgrass, alfalfa, soybeans, and cotton are better suited to be grown on steeper slopes than corn.

Available Water Capacity

Available water capacity has a positive relationship with crop yield as supported by theory and the coefficients in the model. The larger values of AWC represent more water that the plant has access to. In Big Creek watershed, rates of available water capacity were not highly variable. Even with limited variability, AWC was a significant variable for each of the models. Figures 4.1, 4.9, 4.17, 4.24, and 4.32 detail the relationship of available water capacity with crop yield. Generally, as water capacity increases, the greater the yield; this is demonstrated by the regression line. AWC had the largest influence on corn yields with a coefficient of 0.182 compared to 0.172 for switchgrass, 0.154 for cotton, 0.077 for alfalfa, and 0.073 for soybeans (Table 4.2). Implications under future climate would indicate that maintaining sufficient soil moisture is key for corn and switchgrass and less crucial for soybean and alfalfa yields in Big Creek Watershed.

April Growing Degree Days

April GDD's had a positive influence on yields across most of the models. This is in line with theory where increasing number of GDD's will lead to better yields. This is especially the case in April, which is a crucial month for farmers to be able to get into the field and plant the crops. For many crops, April is generally the first month where farmers will begin to plant crops if soil and weather conditions allow. Future climate scenarios will lead to a greater number of GDD's, which will lead farmers to plant their crops sooner, thus, extending the growing season and increasing yields. Figures 4.3, 4.12, 4.21, 4.27, and 4.35 display the relationship between April GDD's and yield. The coefficient for alfalfa is positive in the model, whereas, the plot in Figure 4.4 displays a negative relationship. This indicates that there is multicollinearity that exists with June GDD's. Multicollinearity was measured by the VIF of the variable and commonly if the VIF is above 5 or 7.5 it should be dropped from the model. The VIF for the variables were not above 5 and in order to maintain comparability with each of the models the variable was kept in the model. While April GDD's did not have as strong of an influence in most of the models it was kept in the model to examine the effect of climate change on early season temperature variability. April GDD's had the biggest influence on soybean yields with a coefficient of 0.451 compared to 0.429 for corn, 0.247 for alfalfa, 0.233

for cotton and 0.076 for switchgrass (Table 4.2). These results suggest that a warmer April will benefit corn and soybean yields.

June Growing Degree Days

June GDD's had a negative relationship on yields across all of the models. This can be attributed to the hotter June having a negative influence on yields. Figures 4.4, 4.13, 4.20, 4.28, and 4.36 reveal this relationship clearly. Since the lowest value of GDD's for the month starts relatively high, the increasing number of GDD's leads to lower yields indicating the negative effect of higher temperatures on the crops for the month of June. Yields start to decrease around 500 GDD's for corn and soybean and around 350 GDD's for switchgrass. Alfalfa yields start to decrease around 600 GDD's and cotton yields start to decrease around 250 GDD's. Since many crops are still in the early phenological stage in this month, it is possible that the hotter climate will reduce yields in June. June growing degree days had a greater influence on alfalfa yields with a coefficient of -0.841 compared to -0.747 for corn, -0.557 for soybeans, -0.5 for cotton, and -0.186 for switchgrass (Table 4.2). The results suggest that a warmer June will negatively influence alfalfa and corn the most.

May Water Surplus

May water surplus had a positive influence on corn, soybean, and cotton yields as shown in Figures 4.5, 4.14, 4.22, 4.29, and 4.37. It had a negative relationship with alfalfa and switchgrass yields as shown in Figures 4.2 and 4.3. This suggests that a wetter May will influence corn, soybean, and cotton positively, but it will negatively influence alfalfa and switchgrass yields. Under future climate scenarios a wetter or drier May could affect crops differently depending on the severity of water deficit or surplus. May WS had a greater influence on soybean yields with a coefficient of 0.269 compared to 0.246 for cotton, -0.195 for switchgrass, 0.156 for corn, and -0.064 for alfalfa. A wetter May would be beneficial for corn and soybean crops and have a negative influence over switchgrass and alfalfa crops.

August Water Surplus

August water surplus had a positive influence on crop yields across all of the models. This can be attributed to a wetter August being beneficial for each of these crops. Often, soil water deficits reach their peak in August so this variable is in the model to better predict the effect of climate change on weather condition later in the growing season. Figures 4.6. 4.15, 4.23, 4.30, and 4.38 display the relationship between August water surplus and crop yield. August water surplus had the greatest influence on alfalfa yields with a coefficient of 0.462 compared to 0.442 for switchgrass, 0.17 for soybean, 0.159 for corn, and 0.048 for cotton (Table 4.2). This suggests that a wetter August will benefit alfalfa and switchgrass crops the most.

Annual Potential Evapotranspiration

Annual potential evapotranspiration had a positive influence on yields across every model. This is in line with theory because potential evapotranspiration is a proxy for warmth and the higher the PET the warmer the climate and the positive impact on crop yield. Figures 4.7, 4.11, 4.19, 4.25, and 4.34 display the relationship between PET and crop yield. Generally, PET values around 440 mm and higher lead to the highest yields for each of the crops. PET had the greatest influence on soybean yields with a coefficient of 0.792 compared to 0.688 for alfalfa, 0.696 for switchgrass, 0.431 for cotton, and 0.379 for corn yields (Table 4.2).

Dummy Variables First Year Rotation (C-2)

The first year rotation cotton dummy variable had a positive influence on crop yields across all of the metamodels. This suggests that if you plant cotton in the first year of the three year rotation then there will be a boost to the yield of the corn, soybean, and cotton crops two years later. Planting cotton two years prior would lead to an increase of 250 kg of corn compared to planting soybean or corn. This is similar to findings by Reddy et al. (2006), where corn yields increased 5-13 percent when rotated with cotton in the Mississippi River floodplain. The soybean rotation dummy variable had a positive influence on crop yields across all of the models. If the farmer planted soybean as the first crop in the three year rotation, there will be a positive influence on the yield two years later for corn, cotton, and soybeans. This is consistent with yield drag theory where switching from a continuously cropped practice to a rotated crop practice will be beneficial for crop yields the following year and two years in the future. Planting soybean two years prior would lead to an increase of 110 kg of corn compared to planting corn or cotton two years prior.

Dummy Variables Second Year Rotation (C-1)

The second year rotation cotton dummy variable had a positive influence on corn yields, and a negative influence on soybean and cotton yields. This suggests that if you plant cotton the year before it would be beneficial for corn yields but not for the other crops. Planting cotton the previous year would lead to an increase of 84 kg of corn compared to soybean or cotton being planted the previous year. The soybean dummy variable had a positive influence on yields across all of the models. This suggests that if the farmer planted soybean the year before there would be a boost in yields for all of the crops. This is expected since soybeans are a nitrogen fixer and can be beneficial for crops such as corn or cotton. The magnitude of yield boost varies from crop to crop. The biggest boost was associated with corn where soybeans would lead to 352 kg increase compared to corn or cotton being planted the previous year. The winter wheat coefficient came out to be significant and had a negative effect on soybean yields. Soybean yields are expected to decrease with double cropping due to the shortened season soybeans have when they are planted with winter wheat.

Tillage Dummy Variables

The dummy variable for conservation tillage was found to be significant in both the corn and cotton model. Conservation tillage was found to be insignificant for the soybean model. The coefficients for corn and cotton were both positive, which is what was expected. Conservation tillage should have a positive impact on yields and increase them compared to conventional tillage. The dummy variable for no-till was found to be significant in the corn model at a 0.05 alpha value. It was found to be insignificant in the cotton and soybean model. The coefficients were positive for the corn and switchgrass models, but negative for the cotton model. Even though these coefficients were not statistically significant they still have some practical significance associated with them. In the literature, tillage practices are beneficial for soil properties, but not as strongly for crop yields as are other farm management practices.

2) How accurately can a statistical equation such as regression reproduce the results of a sophisticated process based watershed model such as SWAT?

The results suggest that the metamodels can reproduce the results of SWAT in predicting crop yields with a moderate degree of accuracy for corn, soybeans, alfalfa, switchgrass, and cotton. The models were able to explain between 48-85 percent of the variance in crop yield across the different crops. The RMSE of the models ranged from 0.75 to 2.53. The coefficient of variation for each of these models ranged from 13 to 41 percent. The corn model was able to capture 70 percent of the variability in yields. The soybean model was able to capture 63 percent of the variability in yields. The switchgrass model was able to capture 77percent of the variability in yields, the alfalfa model was able to capture 84 percent of the variability in yields, and the cotton model was able to capture 48 percent. The cotton model captured the least amount of variability in cotton yields compared to the other models. Potential explanations could relate to the selected variables and to the climate conditions run through SWAT. For numerous SWAT runs, yields were considerable lower and were simulated near 0 t/ha. This could be due to the poor growing conditions for cotton under historical and current climate conditions. A diagnostic of weather conditions revealed a high amount of temperature stress days accumulated by cotton throughout the growing season. This indicates that there is considerable temperature stress on cotton yields associated with colder temperatures in Big Creek. There was also a lack of variability in the soil variables such as available water capacity. AWC less than 200 mm and slopes greater than 3 percent lead to poor cotton yields. The HRU's with those conditions had the lowest cotton yields. There is potential for cotton to be grown in Big Creek Watershed under future climate conditions, but topographic and soil limitations will lead to cotton being planted only in certain parts of Big Creek Watershed such as the lowland areas near the Cache River.

The results from the corn, switchgrass, and alfalfa model are in line with results from previous agricultural metamodel literature. Additionally, the metamodels are able to capture the sensitivity of crop yields in response to various weather and farm management scenarios. This is key to capture the sensitivity and magnitude of change in yields from scenario to scenario in order for the Agent Based Model to predict how climate will affect farmer's decisions. The metamodels were a computationally efficient alternative to running the simulation model SWAT. A small set of statistically significant variables is more advantageous than setting up and running SWAT in order to derive one output variable from the model. This was demonstrated by the runtime of the model. The SWAT model can take hours to process various scenarios depending on the computer used. At the minimum it took a half hour to run through the 14 corn permutations for one three-year rotation sequence. This was on a Core i5 laptop, whereas, running the simulations on older computers took a least an hour to run through one three-year rotation sequence. Multiply that by the various climate scenarios and it can take hours to a whole day to run through all the scenarios. This is in contrast to the metamodel, which can be run in seconds to minutes by feeding in the data into the regression equation. The results of this research indicate that metamodels offer a 99 percent reduction in computation time coupled with a moderate degree of accuracy. The reduction in time and parameters demonstrates that metamodels represent a valuable alternative to running large computationally intensive simulation models.

3) How will crop yields in Big Creek Watershed respond to various climate scenarios?

Crop yields in Big Creek Watershed will vary depending on the severity of climate change over the course of the century. Figure 5.1 details the effect on yield under various climate scenarios. Future climate data were derived from downscaled climate data associated with the IPSLCM model at a high (RCP85) carbon scenario. The figures demonstrate that crop yields will decrease under various climate scenarios. Under baseline climatic conditions, corn yields range from 2.73 to 7.51 t/ha. Climate A represents a possible climate scenario for mid-century and climate B represents a possible climate scenario in late century. For climate A, the data for precipitation and temperature were averaged over the decade in the 2050's utilizing data from the IPSCLM model. For climate B, data for precipitation and temperature were averaged over the decade in the 2080's from the IPSLCM model. For climate A, yields decrease drastically from a low of 0 to a high of 4.71 t/ha. In climate B, the yields rebound somewhat from 0.83 to 5.77 t/ha. This could be due to the effect of higher CO_2 levels in the atmosphere by late century, which leads to increased water use efficiency due to stomata conductance. Assessments by the IPCC indicate that the effects of CO_2 in mid-century might not be high enough to counteract the extreme temperature and precipitation events that will decrease yields. This is especially the case for C_3 crops such as soybeans and wheat, which respond greatest to CO_2 fertilization. Figures 5.2-5.4 display the average crop yield for soybeans, alfalfa, and switchgrass. The trend holds similar for these crops where the average yields decrease from historical to climate A and then remain the same or increase in climate B. Table 5.1 details average yields for each crop under additional climate variability scenarios. The scenarios were based on the climate data that were used to calibrate the metamodel. Data were defined from the upper and lower limits of the variables to determine how varying moisture and temperature will impact the crops. All crops see larger yields with a warmer April, and see decreases under a hotter June. Alfalfa and switchgrass have the highest yields under a wet August scenario. This demonstrates that the variability in the early and late part of the growing season will affect crops differently and possibly encourage farmers to plant certain crops over others depending upon future climate change.

Table 5.1 Crop Yields Under Various Climate Scenarios

Crop Yields (t/ha)				
Scenario	Corn	Soybean	Alfalfa	Switchgrass
Cold April	4.77	2.12	6.03	6.93
Hot June	2.96	1.95	3.32	5.54
Dry August	5.97	2.65	5.88	4.97
Dry May	5.65	2.39	7.17	8.67
Wet May	7.34	3.41	6.70	5.90
Hot April	9.06	4.04	8.05	8.13
Wet August	7.37	3.34	9.03	11.37
Baseline	7.97	2.55	6.63	8.8

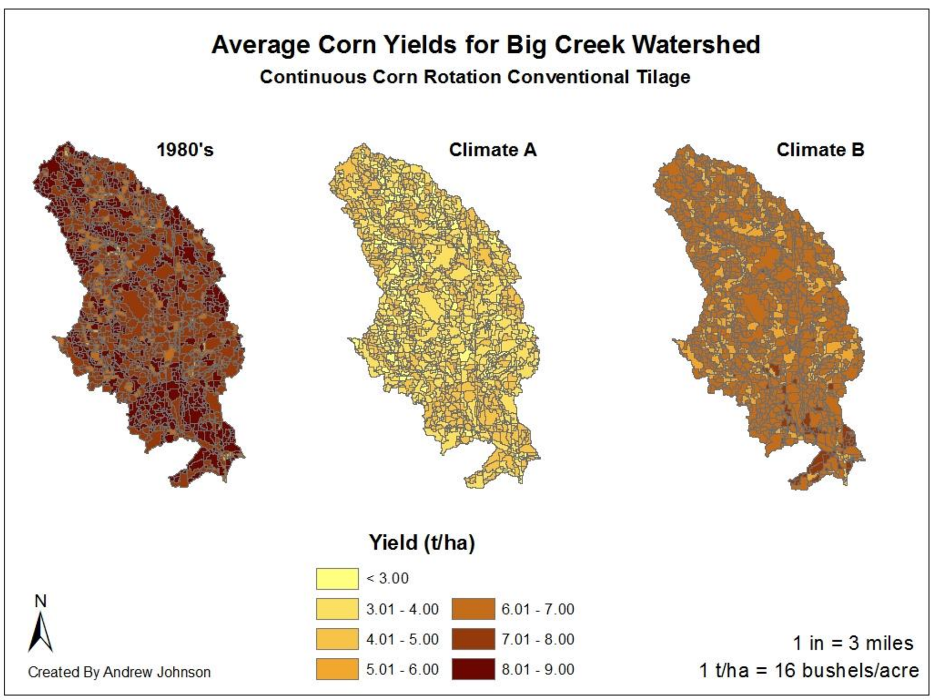


Figure 5.1 Corn Yields Under Various Climates

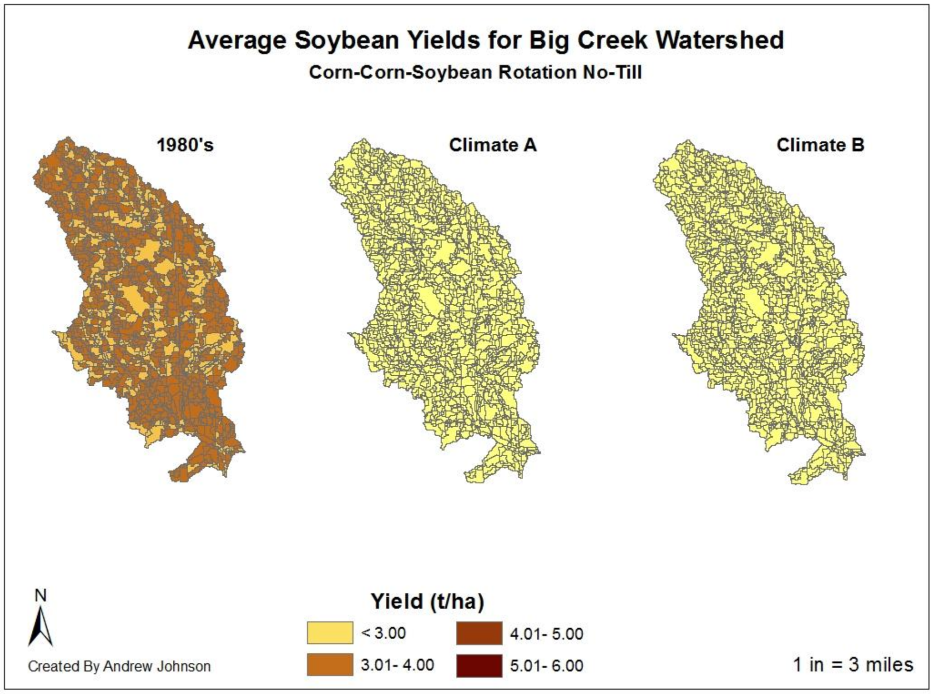


Figure 5.2 Soybean Yields Under Various Climates

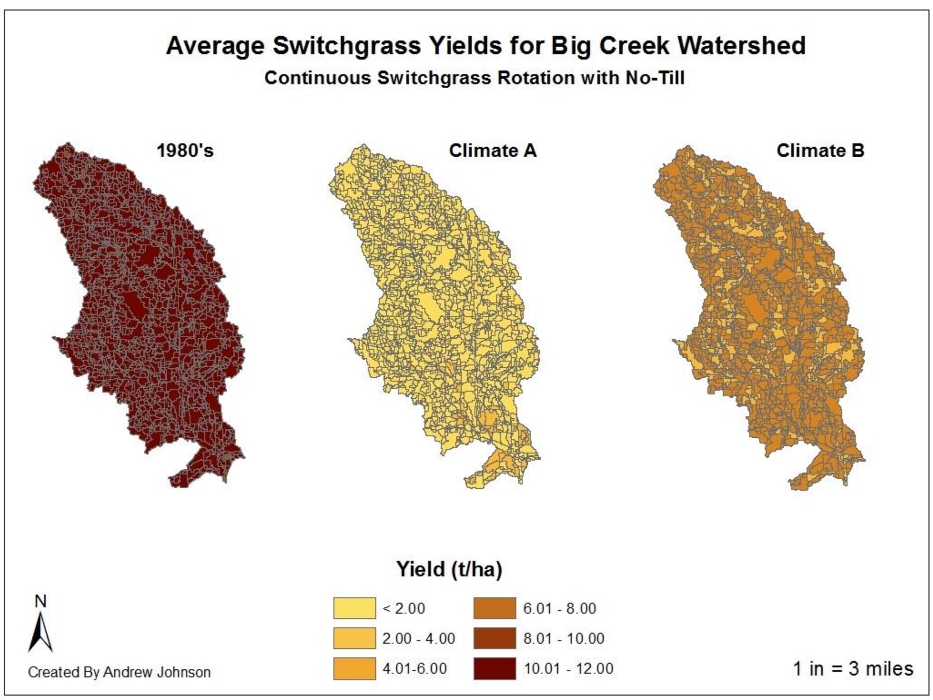


Figure 5.3 Switchgrass Yields Under Various Climates

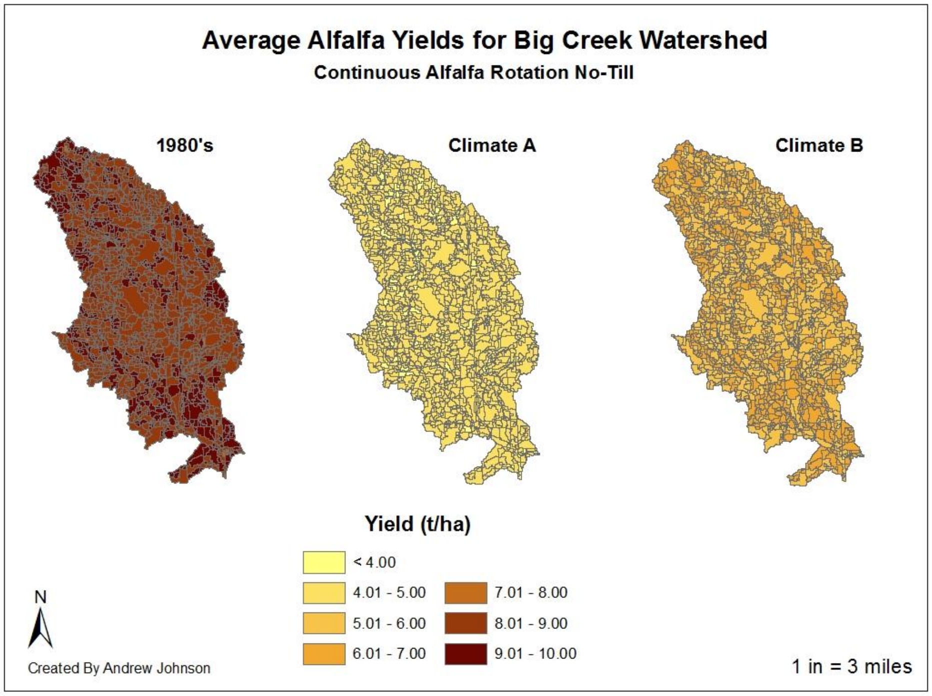


Figure 5.4 Alfalfa Yields Under Various Climates

Conclusion

Five metamodels were created for corn, soybean, switchgrass, alfalfa, and cotton crops in Big Creek watershed. This research has demonstrated that metamodels can be used to approximate SWAT. These metamodels can be used to predict historical crop yields and crop yields under future climate scenarios. These metamodels can also capture the variance associated with climate change impacts on temperature and precipitation. The accuracy of the metamodels in this research is consistent with results achieved by other agricultural metamodel studies. Additionally, metamodels are more advantageous to use than the simulation model by reducing computation time and utilizing fewer statistically significant parameters. Utilizing crop yield metamodels will lead to a more extensive analysis of climate change scenarios for Big Creek Watershed.

REFERENCES

Adams, Richard M., Cynthia Rosenzweig, Robert M. Peart, Joe T. Ritchie, Bruce A. McCarl, J. David Glyer, R. Bruce Curry, James W. Jones, Kenneth J. Boote, and L. Hartwell Allen. 1990. Global Climate Change and US Agriculture. *Nature* 345 (627): 219–224.

Arnold, J.G., and P.M. Allen. 1999. Automated methods for estimating baseflow and groundwater recharge from streamflow records. *Journal of the American Water Resources Association*. 35(2): 411-424

Baker, J. T., and L. H. Allen Jr. 1993. Contrasting Crop Species Responses to CO2 and Temperature: Rice, Soybean and Citrus. *Vegetation* 104 (1): 239–260.

Bekele, E.G., and J.W. Nicklow. 2005. Multiobjective management of ecosystem services by integrative watershed modeling and evolutionary algorithms. *Water Resources Research* 41(10): 1-10.

Bekele, E.G., and J.W. Nicklow. 2007. Multi-objective automatic calibration of SWAT using NSGA-II. *Journal of Hydrology* 341: 165-176.

Broad, D.R., H.R. Maier., and G.C. Dandy. 2010. Optimal Operation of Complex Water Distribution Systems Using Metamodels. *Journal of Water Resources Planning and Management* 136(4): 433-443.

Brown, R.A., N.J. Rosenberg, C.J. Hays, W.E. Easterling, and L.O. Mearns. 2000. Potential production and environmental effects of switch grass and traditional crops under current and greenhouse altered climate in the central United States: A simulation study. *Agriculture, Ecosystems and Environment* 78: 31-47.

Bouzaher, A., R. Cabe, A. Carriquiry, P. Gassman, P.G. Lakshminarayan, J.F. and Shogan. 1992. Metamodels and Nonpoint Pollution Policy in Agriculture. Working Paper 92-WP 97.

Burt, J.E., G.M. Barber, and D.L. Rigby. 2009. *Elementary Statistics for Geographers*. New York: The Guilford Press.

Butler, E.E., and P, Huybers. 2013. Adaptation of US Maize to Temperature Variations. *Nature* (3): 68-72.

Cao, W., W.B. Bowden, T. Davie, and A. Fenemor. 2006. Multi-variable and multi-site calibration and validation of SWAT in a large mountainous catchment with high spatial variability. Hydro. Proc. 20(5): 1057-1073.

Caviglia, O.P., V. O., Sadras, and F. H., Andrade. 2011. Yield and Quality of Wheat and Soybean in Sole- and Double-Cropping. *Agronomy Journal* 103 (4): 1081.

Christensen, J.H., B. Hewitson, A. Busuioc, A. Chen, X. Gao, I. Held, R. Jones, R.K. Kolli, W.-T. Kwon, R. Laprise, V. Magaña Rueda, L. Mearns, C.G. Menéndez, J. Räisänen, A. Rinke, A. Sarr and P. Whetton, 2007. *Regional Climate Projections In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Couckuyt, I. D. Gorissen, H. Rouhani, E. Laermans, and T. Dhaene. 2009. Evolutionary Regression Modeling with Active Learning: An Application to Rainfall Runoff Modeling. International Conference on Adaptive and Natural Computing Algorithms.548-558.

Coulter, J.A. C, Sheaffer. D, Wyse. M, Haar. P, Porter. S, Quiring. and D. Klossner. 2011. Agronomic Performance of Cropping Systems with Contrasting Crop Rotations and External Inputs. *Agronomy Journal* 103: 182-192.

Edwards, JH, DL Thurlow, and JT Eason. 1988. Influence of Tillage and Crop Rotation on Yields of Corn, Soybean, and Wheat. *Agronomy Journal.* 80: 76–80.

Ewert, F. 2012. Opportunities in Climate Change? *Nature Climate Change* 2:153-157.

Fontaine, T.A., J.F. Klassen, T.S. Cruickshank, and R.H. Hotchkiss. 2001. Hydrological response to climate change in the Black Hills of South Dakota, USA. *Hydrological Sciences Journal*. 46(1): 27-40

Galelli, S., C. Gandolfi, R. Soncini-sessa, and D. Agostani. 2010. Building a metamodel of an irrigation district distributed-parameter model. *Agricultural Water Management* 97: 187-200.

Gentry, Laura F., Matias L. Ruffo, and Fred E. Below. 2013. Identifying Factors Controlling the Continuous Corn Yield Penalty. *Agronomy Journal* 105(2): 295- 303.

Gassman, P. W., M. R. Reyes, C. H. Green, and J. G. Arnold. 2007. The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. *Transactions of the ASABE* 50(4): 1211-1250.

Gorissen, D., W. Hendrickx, K. Crombecq, and T. Dhaene. 2006. Integrated Grid Computing and Metamodeling. Proceedings of the Sixth IEEE International Symposium on Cluster Computing and the Grid.

Grayson, R.B., I.D. Moore, and T.A. McMahon. 1992. Physically Based Hydrologic Modeling 2. Is the Concept Realistic? Water Resources Research 26(10): 2659-2666.

Grover, K. H, Karsten. and G, Roth. 2009. Corn Grain Yields and Yield Stability in Four Long-Term Cropping Systems. *Agronomy Journal* 101: 940-946.

Guetersloh, M. 2001. Big Creek Watershed Restoration Plan: A Component of the Cache River Watershed Plan. 52. Springfield, IL. Illinois Department of Natural Resources.

Horowitz, J. R, Ebel., and K, Ueda. 2010. *"NoTill" Farming Is a Growing Practice*. USDA ERS. Economic Information Bulletin 70.

Jha, M., J.G. Arnold, P.W. Gassman, F. Giorgi, and R. Gu. 2006. Climate change sensitivity of water yield in the Upper Mississippi River Basin. *Journal of American Water Resource Association* 42(4): 997-1015.

Jiang, P.T., and others. 2004. Effect of Soil and Topographic Properties on Crop Yield in a North-Central Corn–soybean Cropping System. *Agronomy Journal* 96 (1): 252.

Jin, R., X. Du, and W. Chen. 2003. The Use of Metamodeling Techniques for Optimization Under Uncertainty. *Structural and Multidisciplinary Optimization* 25 (2): 99–116.

Harmel, R.D., R.J. Cooper, R.M. Slade, R.L. Haney, and J.G. Arnold. 2006. Cumulative uncertainty in measured stream flow and water quality data for small watersheds. Trans. ASABE 49(3): 689-701.

Krishna N. Reddy, Martin A. Locke, Clifford H. Koger, Robert M. Zablotowicz, and L. Jason Krutz (*2006*) Cotton and corn rotation under reduced tillage management: impacts on soil properties, weed control, yield, and net return. *Weed Science* 54(4): 768-774.

Kravchenko, A.N.B., and G. Donald. 2000. Correlation of Corn and Soybean Grain Yield with Topography and Soil Properties. *Agronomy Journal* 92 (1): 75.

Lant, C.L., S. Secchi, G. Misgna. 2009. Climate Change, Hydrology, and Landscapes of America's Heartland: A Multi-Scale Natural-Human System. National Science Foundation Grant Proposal.

Lobell, D. B., Adam Sibley, and J. Ivan Ortiz-Monasterio. 2012. Extreme Heat Effects on Wheat Senescence in India. *Nature Climate Change* 2 (3): 186–189.

Lobell, D.B., M. Banzinger, C. Magorokosho, and B. Vivek. 2011. Nonlinear Heat Effects on African Maize As Evidenced By Historical Yield Trials. *Nature Climate Change* 1:42-45.

McMaster, G.S. and D.E. Smika. 1997. Growing Degree-Days: One Equation, Two Interpretations. *Agriculture Forest Meteorology* 87: 291-300.

Meehl, G.A., T.F. Stocker, W.D. Collins, P. Friedlingstein, A.T. Gaye, J.M, A. Kitoh, R. Knutti, J.M. Murphy, A. Noda, S.C.B. Raper, I.G. Watterson, A.J. Weaver and Z.C. Zhao. 2007. Global Climate Projections. In: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (eds.)]. Cambridge University Press, UK.

Moriasi, D.N., B.N. Wilson, K.R. Douglas-Mankin, J.G. Arnold and P.H. Gowda. 2012. Hydrologic and water quality models: Use, calibration, and validation. *Transactions of the ASABE*. 55(4): 1241-1247.

Muleta, M., and J. Niklow. 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology* 206: 127-145.

Neitsch, S.L., J.G. Arnold, J.R. Kiniry, and J.R. Williams. 2011. Soil and Water Assessment Tool Theoretical Documentation Version 2009. Texas Water Resources Institute Technical Report No. 406.

NowData NOAA Online Weather Data. 2013. http://www.nws.noaa.gov/climate/xmacis.php?wfo=pah

Parry, M.L., O.F. Canziani, J.P. Palutikof and Co-authors 2007. Technical Summary Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, M.L. Parry, O.F. Canziani, J.P. Palutikof , P.J. van der Linden and C.E. Hanson, Eds., Cambridge University Press, Cambridge, UK, 23-78.

Peterson, T.A., and GE Varvel. 1989. Crop Yield as Affected by Rotation and Nitrogen Rate. I. Soybean. *Agronomy Journal* 81: 727-731.

Piñeros Garcet, J.D., A. Ordoñez, J. Roosen, and M. Vanclooster. 2006. Metamodelling: Theory, Concepts and Application to Nitrate Leaching Modelling. *Ecological Modelling* 193 (3-4): 629–644.

Porter, J. R., and M. A. Semenov. 2005. Crop Responses to Climatic Variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1463): 2021–2035.

Qaderi, M.M., and D.M., Reid. 2009. Crop Responses to Elevated Carbon Dioxide and Temperature. *Climate Change and Crops:* 1-9.

R version 2.15.2. 2012. The R Foundation for Statistical Computing

Riedell, W.E., J.L. Pikul Jr, A.A. Jaradat, and T.E. Schumacher. 2009. Crop Rotation and Nitrogen Input Effects on Soil Fertility, Maize Mineral Nutrition, Yield, and Seed Composition. *Agronomy Journal* 101(4): 870-879.

Rosenthal, W.D., and D.W. Hoffman. 1999. Hydrologic modeling/GIS as an aid in locating monitoring sites. *Transactions of the ASAE*. 42(6): 1591-1598.

Saleh, A., J.G. Arnold, P.W. Gassman, L.W. Hauck, W.D. Rosenthal, J.R. Williams, and A.M.S. McFarland. 2000. Application of SWAT for the upper north Bosque Watershed. Trans. ASAE 43(5): 1077-1087.

Santhi, C., R. Srinivasan, J.G. Arnold, and J.R. Williams. 2005. A modeling approach to evaluate the impacts of water quality management plans implemented in a watershed in Texas. *Environmental Modelling & Software.* 21: 1141-1157.

Stonefelt, M. D., Fontaine, T. A. and Hotchkiss, R. H. 2000. Impacts Of Climate Change On Water Yield In The Upper Wind River Basin. *JAWRA Journal of the American Water Resources Association* 36: 321–336

Thornthwaite, C.W. 1948. An Approach Toward a Rational Classification of Climate. *Geographical Review* 38(1): 55-94.

USDA Crop and Livestock Practices. 2013. http://www.ers.usda.gov/topics/farm-practices-management/crop-livestock-practices.aspx#.UZPq3kpvB-I

Walter, R.E. J, Pikul. A, Jaradat. and T, Schumacher. 2009. Crop Rotation and Nitrogen Input Effects on Soil Fertility, Maize Mineral Nutrition, Yield, and Seed Composition. *Agronomy Journal* 101: 870-879.

West, T.O., and G. Marland. 2002. A Synthesis of Carbon Sequestration, Carbon Emissions, and Net Carbon Flux in Agriculture: Comparing Tillage Practices in the United States. *Agriculture, Ecosystems & Environment* 9 (1): 217–232.

Wheeler, Tim. 2012. Agriculture: Wheat Crops Feel the Heat. *Nature Climate Change* 2 (3): 152–153.

Wilhelm, W. W., J. W. Doran, and James F. Power. 1986. Corn and Soybean Yield Response to Crop Residue Management Under No-tillage Production Systems. *Agronomy Journal* 78 (1): 184–189.

Wu, J., P.G. Lakshminarayan, and B.A. Babcock. 1996. Impacts of Agricultural Practices and Policies on Potential Nitrate Water Pollution in the Midwest and Northern Plains of the United States. Working Paper 96-WP 148.

Yadev, V., and G.P. Malanson, 2007. Progress in soil organic matter research: litter decomposition, modeling, monitoring and sequestration. *Progress in Physical Geography* 31(2): 131-154.

Yadev, V., and G. Malanson, 2008. Spatially explicit land use land cover and soil organic carbon transformations in Southern Illinois. *Agriculture, Ecosystems and Environment* 123: 280-292.

Yadev, V., G.P. Malanson, E. Bekele, and C. Lant. 2009. Modeling watershed-scale sequestration of soil organic carbon for carbon credit programs. *Journal of Applied Geograph*y 29:488-500.

Yoder, R.E., L.O., Odhiambo, and W.C., Wright. 2005. Evaluation of Methods For Estimating Daily Reference Crop Evapotranspiration At A Site In The Humid Southeast United States. *Applied Engineering in Agriculture* 21(2): 197-202.

Zhang, X., R. Srinivasan, and M. Van Liew. 2008. Multi-Site Calibration of the SWAT Model for Hydrologic Modeling, *Transactions of the ASABE*. 51(6): 2039-2049.

APPENDIX

APPENDIX A – PYTHON SCRIPTS

Growing Degree Day Script

```
#### GDD.py
####Created by Andrew Johnson
####Date: 3/20/13
'''
 This script calculates growing degree days for corn
'''

####Corn SPECIFIC PARAMETERS
top = 42.00
base = 8.00

myPath = "C:\\BigCreek\\TIN.txt"
inFile = open(myPath,'r')
myList = inFile.readlines()
outPath = "C:\\BigCreek\\TOUT.txt"
outFile = open(outPath,'w')

for day in myList:

   day.strip(",!?;:&*'=></#@)('\n'")

   day = float(day)

   if day <= base:
      GDD = 0
      outFile.write(str(GDD)+'\n')

   elif day >= base and day <= top:
      GDD = day-base
      outFile.write(str(GDD)+'\n')

   elif day >= top:

      GDD = 0
      outFile.write(str(GDD)+'\n')

   else:
      print "Error GDD did not calculate correctly" + str(day)
```

```
inFile.close()
outFile.close()
```

Temperature Stress Script

```
###CREATED BY: ANDREW JOHNSON
###CREATED ON: 3/22/13
### This script was created to calculated temperature stress given an
###input of daily average temperature


### CORN SPECIFIC PARAMETERS
optimal = 25.00
base = 8.00
### Corn starts to experience heat stress at 42 celsius when tavg > 2*Topt-Tbase

myPath = "C:\\BigCreek\\TIN.txt"
inFile = open(myPath,'r')
myList = inFile.readlines()
outPath = "C:\\BigCreek\\TOUT.txt"
outFile = open(outPath,'w')

for day in myList:
    day.strip(",!?;:&*'=></#@)('\n'")
    day = float(day)

    if day <= base:
        '[0] == "-":'
        tstress = 1
        outFile.write(str(tstress)+'\n')

    elif day >= 42.00:
        tstress = 1
        outFile.write(str(tstress)+'\n')

    elif day >= base and day <= optimal:
        tstress = 1-(2.7818281828**((-0.1054*((optimal-day)**2))/((day-base)**2)))
        outFile.write(str(tstress)+'\n')

    elif day >= optimal and day <= 42.00:
        tstress = 1-(2.7818281828**((-0.1054*((optimal-day)**2))/((50-day-
base)**2)))
```

```
        outFile.write(str(tstress)+'\n')

    else:
        print "Error temp stress did not calculate correctly"

inFile.close()
outFile.close()
```

Extract Output Data Script

```
from numpy import *
import os
import glob

workspace = "C:\\Documents and Settings\\siu850895153\\My Documents\\Dropbox\\SIUC\\SWAT\\outputfiles\\WW"
outfile = "C:\\Documents and Settings\\siu850895153\\My Documents\\Dropbox\\SIUC\\SWAT\\outputfiles\\WW\\AllOut.txt"

for file in glob.glob(os.path.join(workspace,'*.hru')):

    mylist = open(file,'r')

    fout = open(outfile, 'a+')

    mylines = mylist.readlines()

    for line in mylines[3297:4941]:
        mystr =  str(line)
        fout.write(mystr)



    mylist.close()
    fout.close()
```

Temperature Formatting Script

```
# This script replaces the zero in the second position with the negative sign

myPath = "C:\\BigCreek\\TIN.txt"
myFile = open(myPath,'r')
```

```
outPath = "C:\\BigCreek\\TOUT.txt"
outFile = open(outPath,'w')

myList = myFile.readlines()

for eachLine in myList:

    if eachLine[0] == "-":
        outFile.write("-"+eachLine[2:])
    else:
        outFile.write(eachLine)

myFile.close()
outFile.close()
```

VITA

Graduate School
Southern Illinois University

Andrew J. Johnson

andrewj87@gmail.com

Augustana College
Bachelor of Arts, Geography, May 2010

Special Honors and Awards:
    2010, Robin Hoffman Award For Service And Scholarship, Geography Department, Augustana College, Rock Island, Illinois

Thesis Title:
    A Regression Metamodel To Replace SWAT In Crop Yield Prediction For Big Creek Watershed

Major Professor: Dr. Christopher Lant