SUSTAINABILITY
CONSORTIUM

FOOD, BEVERAGE & AGRICULTURE

Regional Corn Modeling at a High Resolution Scale: A Yield Based Approach and Blue vs Green Water Assessment

Center for Agricultural and Rural Sustainability
University of Arkansas Division of Agriculture
Arkansas Cooperative Extension Service

White Paper #105 | February 2011

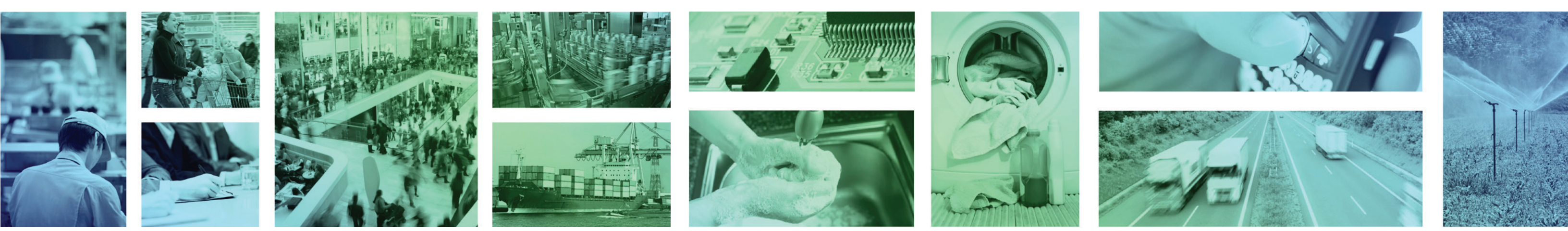

Table of Contents

Table of Contents i

Executive Summary 1

1. Introduction 2

2. Phase 1 2

  2.1. Regional Corn Yield Modeling 3

    2.1.1. Methods 3

    2.1.2. Results 5

    2.1.3. Discussion 8

  2.2. Blue vs. Green Water Estimation 8

    2.2.1. Water Use Efficiency 8

    2.2.2. Estimating Corn Evapotranspiration 10

3. Phase 2 11

  3.1. Background 11

  3.2. Methods 12

    3.2.1. Study Area 12

    3.2.2. Weather Inputs 13

    3.2.3. Soil Inputs 13

    3.2.4. Management Inputs 13

    3.2.5. Historical Yields 13

    3.2.6. Genetic Coefficient Estimation 14

    3.2.7. Validation 14

  3.3. Sensitivity Analysis 15

  3.4. Discussion 15

4. Conclusion 15

Literature Cited 17

Figures

Figure 1. USDA's ERS Region 1. Located between 36.042°N to 46.625°N and 99.292°W to 82.125°W. 2
Figure 2. ISRIC-WISE soil distribution through ERS Region 1. WISE soil profiles have to added to FAO-Unesco Soil Map of the World at 5min resolution. 4
Figure 3. 2000 SAGE corn yield data. The data covers the globe at a 5 min resolution. 4
Figure 4. Optimized distribution of cultivars through ERS Region 1. 6
Figure 5. Predicted vs. observed graph for yield calibration data. 6
Figure 6. Map of the percent error within the ERS Heartland Region. 7
Figure 7. Predicted vs. observed graph for yield validation data. 7
Figure 8. Percent errors for validation run. 8
Figure 9. Yield based evapotranspiration of corn for the globe. 10
Figure 10. Yield-derived global blue vs green water map. 11
Figure 11. Simulated yield versus genetic coefficient P1 for the year 1999 (Top Left), simulated yield versus genetic coefficient P5 for the year 2001 (Top Right), simulated yield versus genetic coefficient G2 for the year 1999 (Bottom Right), and simulated yield versus genetic coefficient G5 for the year 2001 (Bottom Left). 16

Tables

Table 1. Nine pre-selected cultivars and their subsequent physiological genetic information used in the calibration process 5
Table 2. Water use efficiency (WUE) values published for various management regimes for maize production. 9
Table 3. Assigned WUE values per corn yield. 10
Table 4. Management parameter used for the CERES-Maize simulations. 13
Table 5. CERES-Maize genetic coefficient definitions. 14

Executive Summary

Agricultural water use currently accounts for as much as 70 percent of total water use. Couple this with an expected 50 percent increase in global water resource demand in the next 40 years and a potential shift in rainfall patterns associated with climate change; one can begin to see the major challenges ahead for the current generations. Predicting the future agricultural water demands patterns is essential to ensure a sustainable world for a growing human population. We developed a crop water demand simulation process incorporating the CERES-Maize model in the Decisions Support System for Agrotechnology Transfer (DSSAT) Cropping System Model (CSM) program, version 4.0, the MATrix LABoratory program (MATLAB), and regional comprehensive datasets including the NASA Agroclimatology Archive, the International Soil Reference and Information Centre (ISRIC) World Inventory of Soil Emission Potentials (WISE) soils database, the Center for Sustainability and the Global Environment (SAGE) Harvest Area and Yields of 175 crops and Crop Calendar database, and the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS).

Phase 1 of the project involved a preliminary a calibration of the CERES-Maize model to the largest corn production regions in the U.S., USDA Economic Research Service (ERS) Region 1, also known as the Heartland Region. The calibration procedure exhibited relative success at modeling at the regional scale, obtaining a R^2 value of 0.8. However, the results were not replicated in the validation step, producing a R^2 value of 0.07

In addition, global evapotranspiration was estimated for the entire globe. The process began by obtain water use efficiencies (WUE) from the current scientific literature. The WUE values were then applied to global corn yields to calculate the evapotranspiration of corn. Finally, based on the equation $Q = P - ET - RO$, blue and green water was calculated for the globe. The areas where ET was greater than the difference between precipitation and runoff were designated blue water areas, while the areas where ET was less than the difference were designated as green water areas.

Given the poor predictive ability of the model produced during Phase 1, the calibration strategy was revised for the regional crop modeling procedure. Maize production ERS region 1 was re-evaluated during Phase 2 and a temporal dimension was added to the process. Phase 2 focused on derived a grid specific definition of maize genetic coefficients to act as predictor of maize yields. A preliminary sensitivity analysis was conducted to determine the optimal starting ranges for calibration process. Coding for the calibration procedure is currently underway and new results are expected soon.

1. Introduction

Global water resource demand is expected to increase by as much as 50 percent in the next 40 years as the human population reaches nine billion. Agricultural water use currently accounts for as much as 70 percent of total water use. Increasing demand for food, feed, fiber and fuel will increase demand for irrigation water and put greater pressure on an already stressed resource. In addition, global climate change will shift rainfall patterns, compounding water stress in certain regions. Predicting future cropping patterns is essential to ensure a sustainable world for a growing human population.

The goal of this project was to develop a crop water demand simulation process incorporating the CERES-Maize model in the Decisions Support System for Agrotechnology Transfer (DSSAT) Cropping System Model (CSM) program, version 4.0 (Jones & Kiniry, 1986; Jones, et al., 2003; Hoogenboom, et al., 2004). The CERES-Maize model, part of DSSAT, was chosen for the study as it represents one of the most tested and widely used crop simulation models available. The model simulates six corn phenological stages and the resulting biomass production, and grain yield using site specific input data. Final yield is determined by soil, weather, and management characteristics such as fertilizer application, irrigation, planting and harvesting date, plant poplulation, and others. The goal of this project was to simulate global corn yield based on a range of input parameters at the highest geospatial resolution supported by the data. A customized user-interface in the MATrix LABoratory program (MATLAB) was created to allow for the input of global datasets within DSSAT.

2. Phase 1

The first phase of the project involved a preliminary a calibration of the CERES-Maize model to the largest corn production regions in the U.S., the United States Department of Agriculture's (USDA) Economic Research Service (ERS) Region 1, also known as the Heartland Region (Figure 1). The Heartland Region represents not only the most production cropping regions in the U.S., but is one of the major producers of corn in the world. Cropping practices within the region are well established and documented.

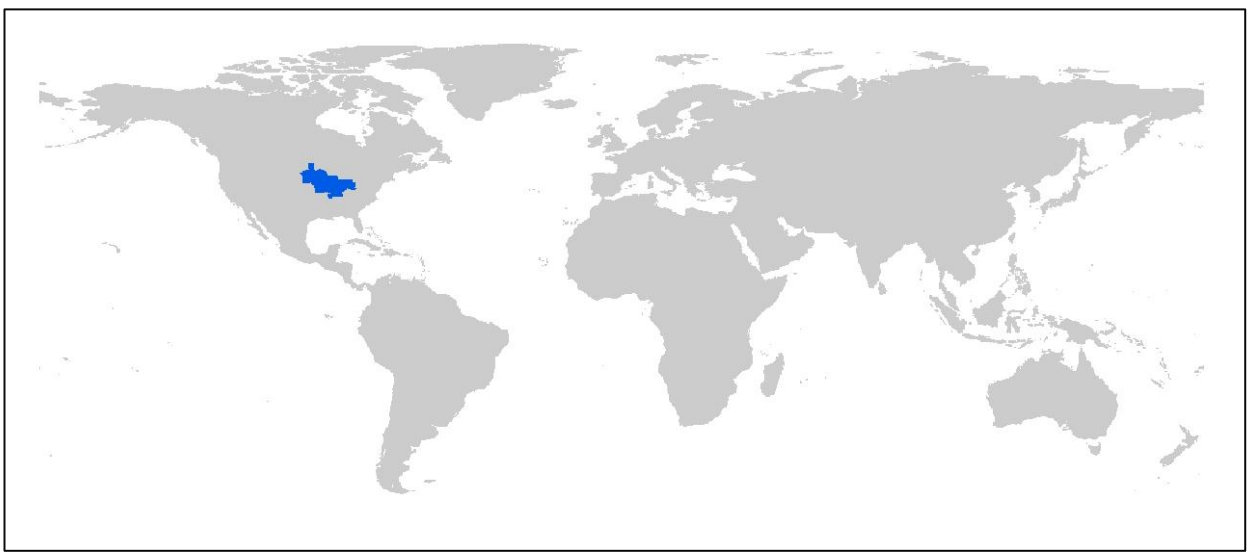

Figure 1. USDA's ERS Region 1. Located between 36.042°N to 46.625°N and 99.292°W to 82.125°W.

The objectives of this Phase 1 were (i) to develop a robust methodology for calibrating pre-established cultivar definitions in DSSAT for the CERES-Maize crop model from publically available global datasets, (ii) to evaluate the performance of the model at the regional scale under non-limiting N conditions, and (iii) to predict irrigation demands for the region in question, and (iv) estimate blue vs. green distribution on a global scale.

2.1. Regional Corn Yield Modeling

2.1.1. Methods

2.1.1.1. Study Area

Phase 1 covered maize production in the USDA Economic Research Service (ERS) U.S. Farm Resource Region 1 (36.042°N to 46.625°N and 99.292°W to 82.125°W) also known as the Heartland Region 2000 and 2005. The ERS regions depict geographic specialization in production of U.S. farm commodities and are derived from four major sources including the older farm production region classifications, a cluster analysis of U.S. farm characteristics (Sommer & Hines, 1991), the USDA Land Resource Region, and the National Agricultural Statistics Service (NASS) Crop Reporting Districts. The region is comprised of roughly 550 counties in nine states. The Heartland region was chosen as it represents not only one of the largest areas of maize production regions in the U.S., but the entire world.

2.1.1.2. Weather Inputs

Weather inputs were obtained from the NASA Agroclimatology Archive, one component of NASA's POWER (Prediction of Worldwide Energy Resource) project. POWER was created to allow access to data derived from NASA's Surface Meteorological and Solar Energy (SSE) project for those interested in the design of renewable energy systems. The Agroclimatology archive was developed with agricultural Decision Supports Systems (DSS) in mind and provides easy download of historical data for specific site locations. The parameters contained in this dataset are based upon solar radiation derived from satellite observations and meteorological data from the Goddard Earth Observing System assimilation model. The archive is globally comprehensive at 1° resolution dating back to July 1983 to near present time. Parameters selected from the this archive include daily estimates of insolation on a horizontal surface (MJ/m^2), daily mean, maximum, and minimum temperatures at 2m above ground surface (°C), and precipitation (mm).

2.1.1.3. Soil Inputs

Soil inputs were obtained from the International Soil Reference and Information Centre (ISRIC) World Inventory of Soil Emission Potentials (WISE) soils database. The WISE database is a globally comprehensive dataset at one of the highest resolutions available at a 5 min resolution. The data were created using the FAO-Unesco Soil Map of the World (DSMW) and soil parameter estimates derived from ISRIC's global WISE soil profile database (Batjes, 2006). The WISE database contains 1125 globally distributed soil profiles that are georeferenced and classified according to the FAO system. Soil profiles were assigned to the FAO classification within the ERS region.

2.1.1.4. Management Inputs

Corn yield values were obtained from the Center for Sustainability and the Global Environment (SAGE) Harvest Area and Yields of 175 crops database (Monfreda, Ramankutty, & Foley, 2008). The dataset is represents a high resolution, 5 min, and is globally comprehensive (Figure 3). Planting and harvesting dates were obtained from SAGE's Crop Calendar Dataset was used to determine planting and harvesting dates. The dataset was complied by digitizing and georeferencing existing observations of cropping systems in a gridded database (Sacks, Deryng, Foley, & Ramankutty, 2010).

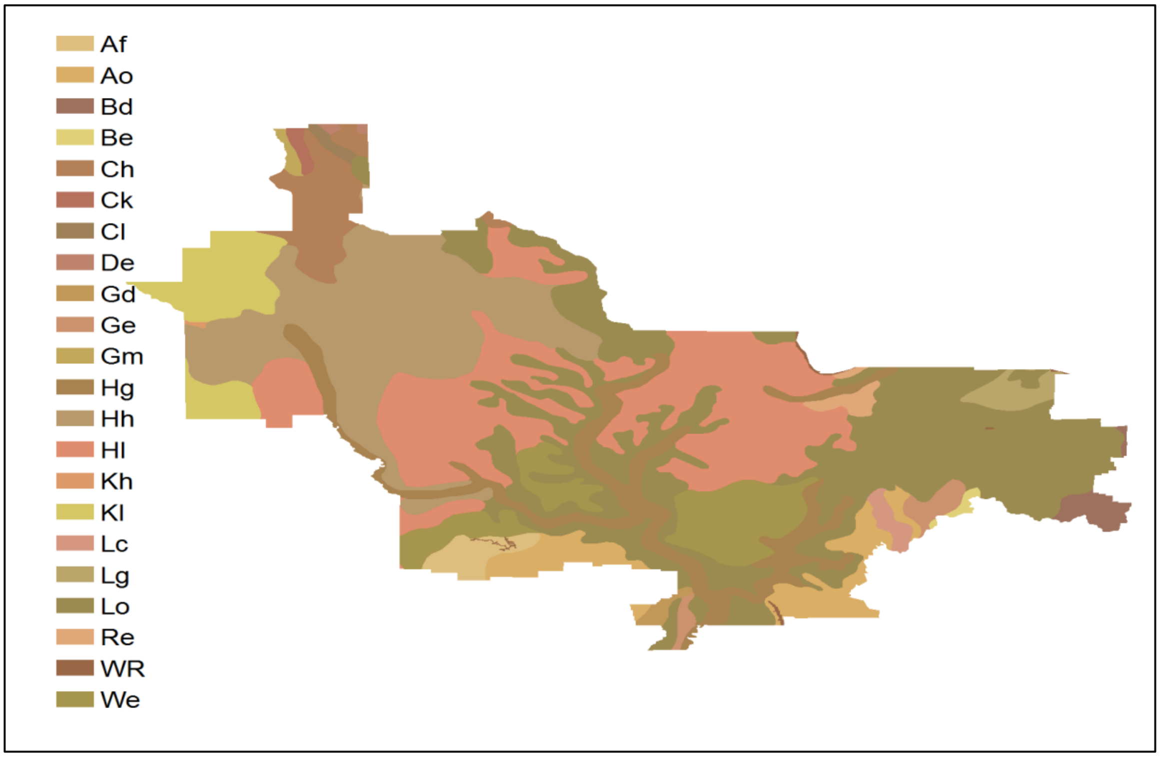

Figure 2. ISRIC-WISE soil distribution through ERS Region 1. WISE soil profiles have to added to FAO-Unesco Soil Map of the World at 5min resolution.

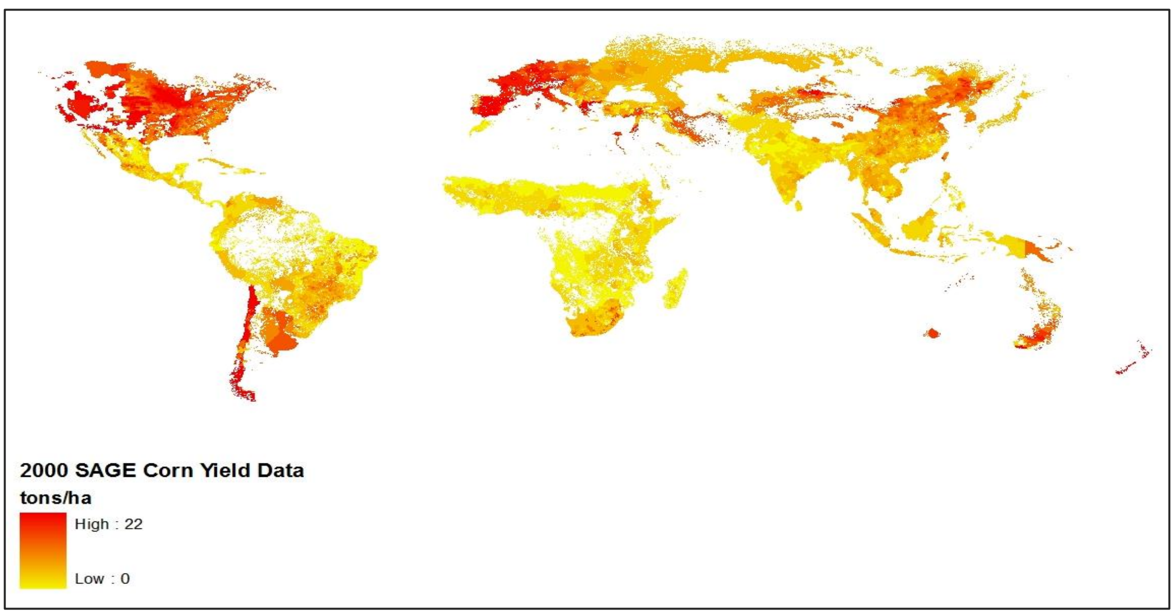

Figure 3. 2000 SAGE corn yield data. The data covers the globe at a 5 min resolution.

2.1.1.5. Data Processing and ERS Region 1 Calibration

A customized Matlab user-interface was developed to facilitate the regional modeling process. The MATLAB interface was designed to format large geospatial data files into input files for use in DSSAT.

The DSSAT-CERES Model was executed from within the MATLAB interface, and the outputs were exported into MATLAB for post-processing.

The calibration process involved selecting nine cultivar varieties that are common to the region a fitting the best cultivar to each pixel in the ERS Region 1 using an optimization function. To begin, the NASA POWER data were sub-sampled to a 5 min resolution to allow for homogeneity in spatial scale during the calibration process. The region contained 11,357 grids in total. Once all input data were set to a common spatial scale, one of the nine pre-selected cultivars (Table 1) were selected using am optimization function. The function of the optimization function was to select the cultivar that produced a yield that was closest to the actual yield for each grid cell.

Table 1. Nine pre-selected cultivars and their subsequent physiological genetic information used in the calibration process

Cultivar	P1	P2	P5	G2	G3	PHINT
H. OBREGON	360	0.8	685	908	10	38.9
H6	310	0.3	685	908	10	38.9
PIO 3147	255	0.76	685	908	10	38.9
A632xVA26	240	0.3	685	908	10	38.9
B14xC103	180	0.5	685	908	10	38.9
DeKalb XL45	150	0.4	685	908	10	38.9
F16xF19	165	0.3	685	908	10	38.9
F478xW705A	140	0.3	685	908	10	38.9
CP170	120	0.3	685	908	10	38.9

2.1.1.6. Model Validation

Using the best fit cultivars selected during the calibration step, validation of the model involved applying the geospatial cultivar definitions to the same input data for 2005, with the exception of SAGE yield data. Considering the SAGE yield dataset only contained information for 2000, the USDA's National Agricultural Statistics Service (NASS) county level were used for the validation process. The coefficient of determination for validation and the Nash-Sutcliffe Coefficient were used to evaluate the predictive ability of the model.

2.1.2. Results

2.1.2.1. Calibration

The distribution of the selected cultivars and a percent error plot of the predicted yield compared to the observed yield can be found in Figures 4 and 5. Figure 6 shows the absolute value of the model error for each pixel. The R2 value for the results of the calibration run was 0.7169. The Nash-Sutcliffe coefficient (E) was 0.67, suggesting that calibration was reasonable.

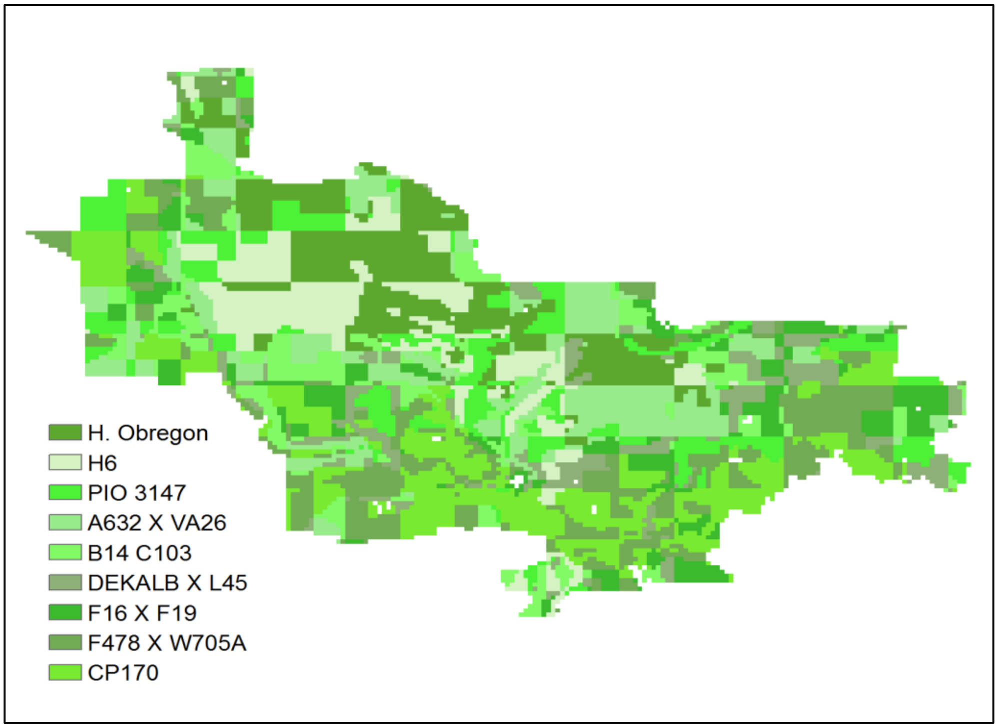

Figure 4. Optimized distribution of cultivars through ERS Region 1.

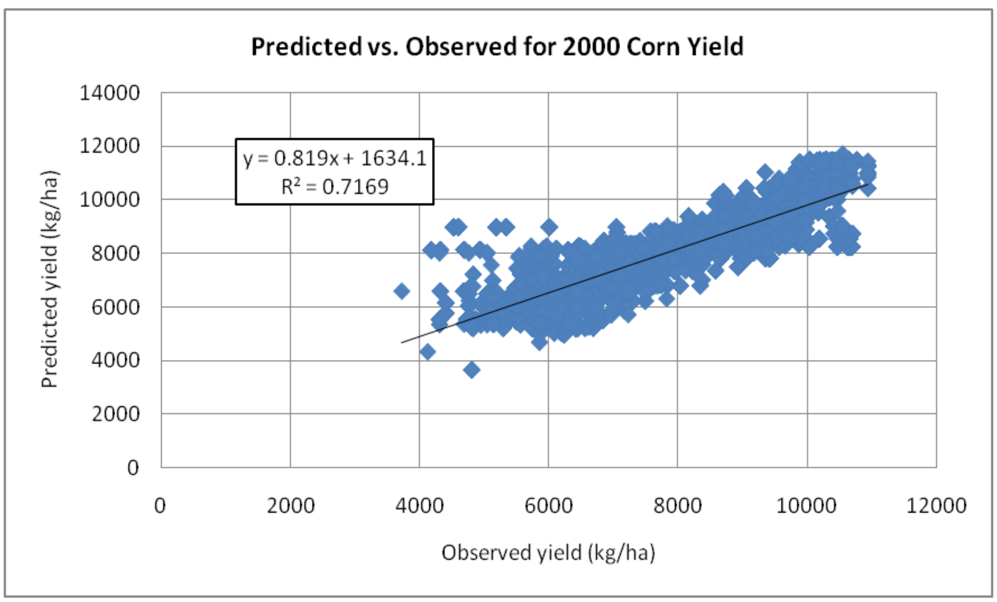

Figure 5. Predicted vs. observed graph for yield calibration data.

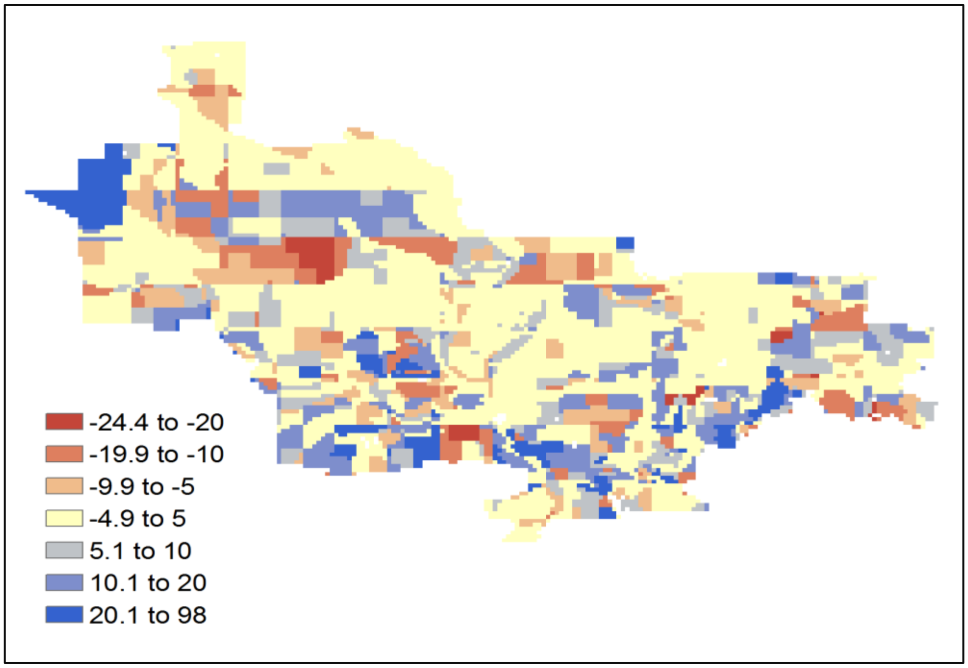

Figure 6. Map of the percent error within the ERS Heartland Region.

2.1.2.2. Validation

Validation results for ERS Heartland Region indicated additional challenges for calibration. The predicted vs. observed yield from 2005 (Figure 7) show the wide range of predictions. The geospatial distribution of simulation error (percent) suggests a clustering of over and under-estimates (Figure 8). The coefficient of determination for validation was 0.0763, and Nash-Sutcliffe Coefficient was -0.33.

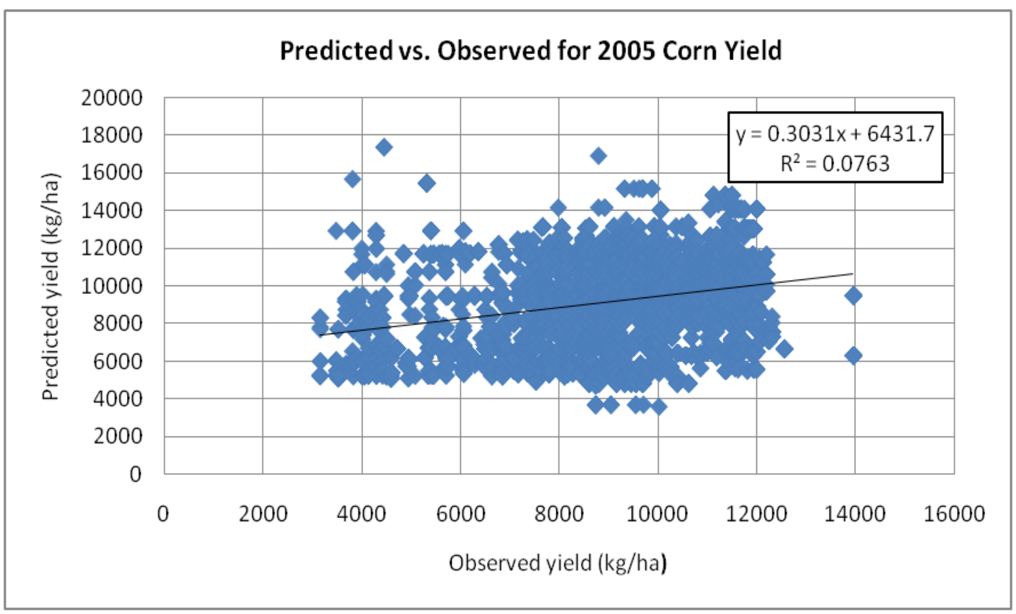

Figure 7. Predicted vs. observed graph for yield validation data.

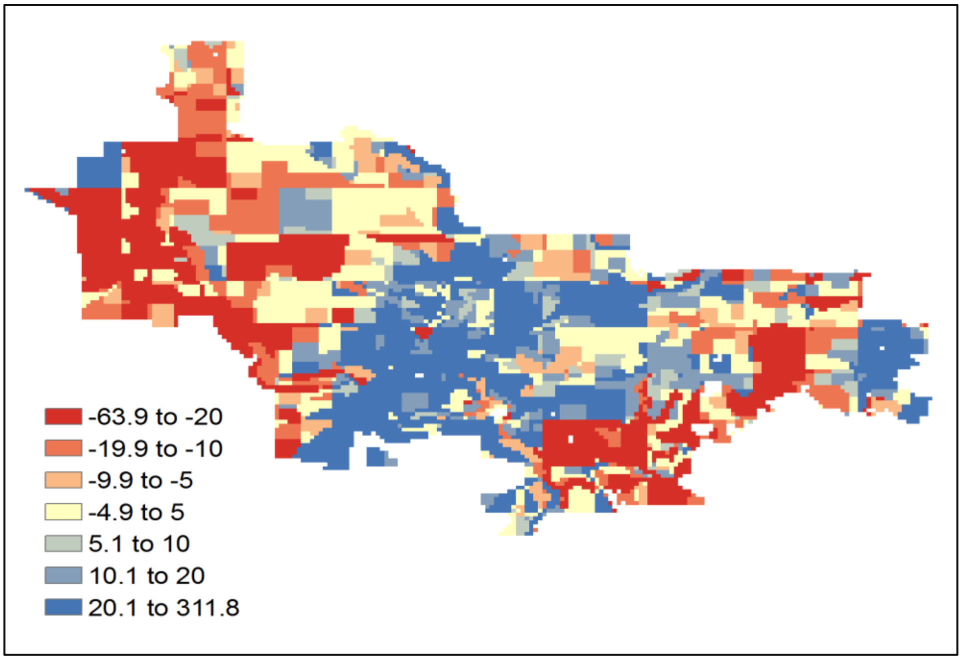

Figure 8. Percent errors for validation run.

2.1.3. Discussion

The calibration procedure exhibited relative success at modeling at the regional scale. The line of best fit for the system had a slope of 0.8 and a y-intercept value of 1,634, suggesting the model was over-predicting yields at the lower end of the observed yield range and under-predicting toward the higher end. The over-prediction of low yield pixel could be a result of low agricultural land use in these pixels. The underlying assumption of regional modeling using a grid-based approach is that each grid represents a homogenous set of inputs. Thus, yield is assumed to be uniform over the pixel. If a particular pixel exhibits a high percentage of an alternative land use, other than agricultural, overall yields will be lowered for the pixel. A lower yielding cultivar would then be selected under the current process. One explanation for the over-prediction is the nine pre-selected cultivars represent varieties that are too high yielding. More testing is needed to characterize the source of this error.

While a high R^2 values were obtained from the calibration run, these results were not replicated in the validation step. The validation run produced a R^2 value of 0.07 and a Nash-Sutcliffe Coefficient of -0.33. In other words, the models residual variance was larger than the data variance. The low predictive ability of the model can likely be attributed to a lack of temporal resolution during the calibration process. Calibrating maize cultivars to one year of yield data proved insufficient to act a predictive model.

2.2. Blue vs. Green Water Estimation

In addition to the development of the CERES-Maize model, the integration of yield-derived evapotranspiration data with precipitation and basin-level runoff data as a method to classify blue and green water areas was explored.

2.2.1. Water Use Efficiency

Vegetative water use refers the quantity of water used for metabolic growth and production and can be described as the amount of water needed to meet the water loss through evapotranspiration.

Water use efficiency (WUE) can be defined as the ratio of crop yield (Y) to water loss through evapotranspiration (ET) (Al-Kaisi & Yin, 2003; Tijani, et al., 2008; Howell, et al., 1998):

$$WUE = \frac{Y}{ET} \qquad [1]$$

Published WUE values ($kg\ ha^{-1}\ mm^{-1}$) were obtained and assigned to four different yield intensities: low, medium, high, and very high. Al-Kaisi and Yin (2003) established WUE values for three different plant population densities of 57,000 (plants ha^{-1}), 69,000 (plants ha^{-1}), and 81,000 (plants ha^{-1}) for the years 1998-2000. Howell et al. (1998) obtained WUE values for both short season and full-season cultivars in Bushland, Texas. The short season cultivar (Pioneer 3737) had a WUE of 15.2 $kg\ ha^{-1}\ mm^{-1}$ for a grain yield of 1.132 $kg\ m^{-2}$ and the full-season cultivar (Pioneer 3245) had a WUE of 15.7 $kg\ ha^{-1}\ mm^{-1}$ for a grain yield of 1.43 $kg\ m^{-2}$. In 2007, Tijani et al. studied maize water use efficiency for multiple fallow treatments in Nigeria and found that native fallow under fertilizer application resulted in the largest WUE (16 $kg\ ha^{-1}\ mm^{-1}$) while native fallow without fertilization had the smallest (10 $kg\ ha^{-1}\ mm^{-1}$). Similarly, Grassini et al. (2009) analyzed maize productivity for both fully irrigated and rainfed conditions in the Western U.S. Corn-Belt and obtained a WUE value of 37 $kg\ ha^{-1}\ mm^{-1}$, the representing our largest yield intensity WUE. A summary of the combined published WUE values as well as the values assigned to each of the four yield intensity categories are shown in Table 2 below.

Table 2. Water use efficiency (WUE) values published for various management regimes for maize production.

WUE ($kg\ ha^{-1}\ mm^{-1}$)	**Descriptor**	**Source**	
10.97			
13.89	57,000 plants ha^{-1}		
12.93			
12.18			
15	69,000 plants ha^{-1}	Al-Kaisi & Yin (2003)	
14.61			
13.57			
14.76	81,000 plants ha^{-1}		
14.36			
16.33	Fertilized Native furrow	Tijani et al. (2008)	
9.98	Non-fertilized Native furrow		
37	Irrigated	Grassini et al. (2009)	
15.2	Short Season (Pioneer 3737)	Howell et al. (1988)	
15.7	Full-Season (Pioneer 3245)		
Low=10 $kg\ ha^{-1}\ mm^{-1}$	Medium=15 $kg\ ha^{-1}\ mm^{-1}$	High=25 $kg\ ha^{-1}\ mm^{-1}$	Very High=35 $kg\ ha^{-1}\ mm^{-1}$

2.2.2. Estimating Corn Evapotranspiration

Using 5 arc-minute actual global corn yield data obtained from SAGE (Ramankutty, Olejniczak, & Foley, 2000) we hypothesized that estimated evapotranspiration (mm) of corn could be obtained by taking the ratio of corn yield (CY) (kg ha^{-1}) to WUE (kg ha^{-1} mm^{-1}).

$$ET_C = \frac{CY}{WUE} \quad [2]$$

The globe was gridded into 5 arc-minute pixels and using ArcGis the corn yield was divided by an assigned WUE value which resulted in the estimated ET_C for each area. For pixels with yields less than 5 T ha^{-1}, 5 to 10 T ha^{-1}, 10 to 15 T ha^{-1}, and greater than 15 T ha^{-1}, WUE respective values of 10, 15, 25, and 35 kg ha^{-1} mm^{-1} will be used (Table 3).

Table 3. Assigned WUE values per corn yield.

Corn Yield Range (T ha^{-1})	**WUE value assigned (kg ha^{-1} mm^{-1})**
< 5	10
5 to 10	15
10 to 15	25
> 15	35

The global estimated ET_C is shown in Figure 9 below. Approximate evapotranspiration for corn over the entire growing period is between 500 and 800 mm (Brouwer & Heibloem, 1986). The estimated ET_c for corn using the WUE method is between 0 and 667 mm with the densest corn production areas corresponding to 450 to 667 mm which seems like a good estimate based on typical corn ET values.

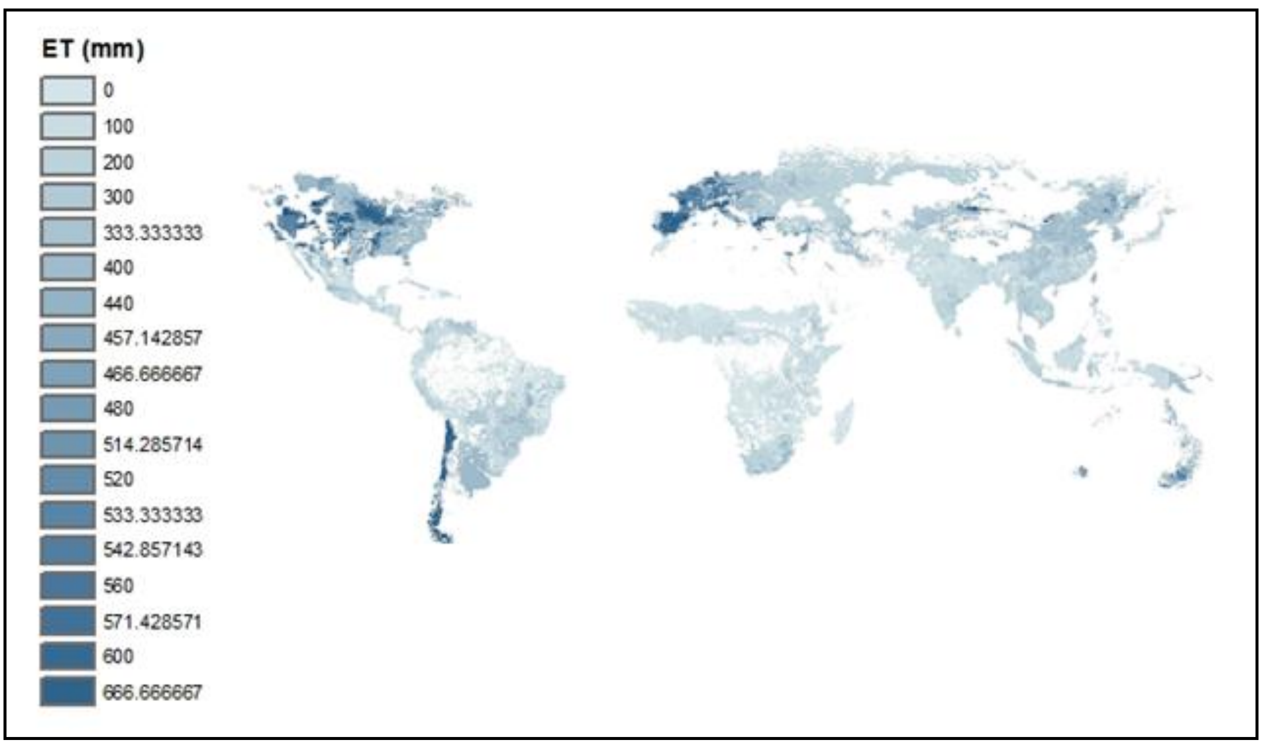

Figure 9. Yield based evapotranspiration of corn for the globe.

Precipitation data was summed for each pixel for the appropriate growing season. Monthly basin-level runoff data, provided by the University of New Hampshire's Global Runoff Data Centre (GRDC), was summed by growing season. Based on the equation Q = P – ET – RO, where P = precipitation and RO = runoff, the total runoff was subtracted from the total precipitation. The areas where ET was greater than the difference between precipitation and runoff were designated blue water areas, while the areas where ET was less than the difference were designated as green water areas. Figure 8 shows the map of the blue and green water.

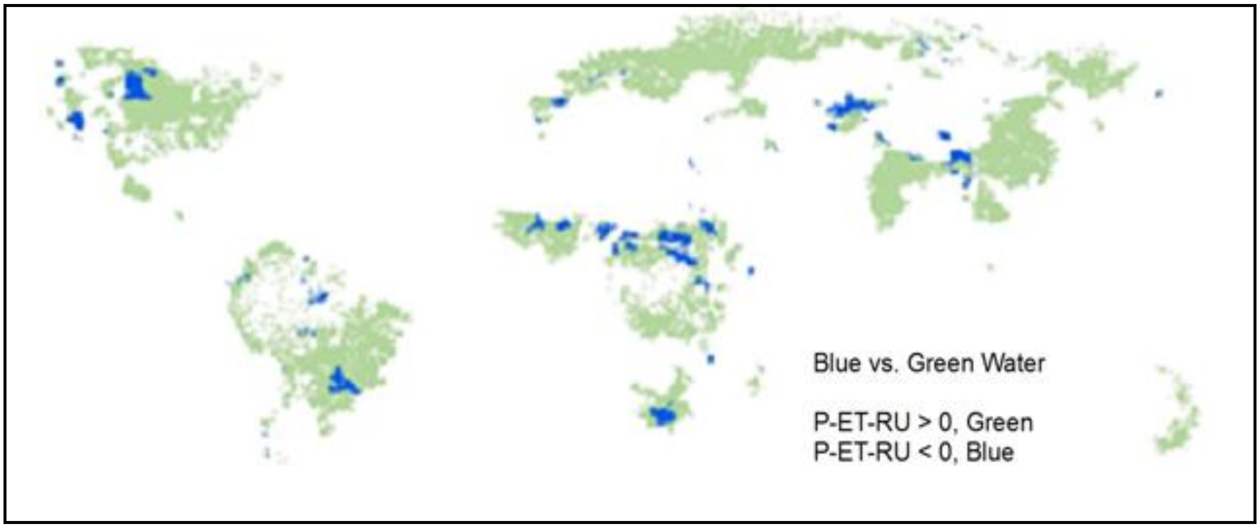

Figure 10. Yield-derived global blue vs green water map.

3. Phase 2

Given the poor predictive ability of the model produced during Phase 1, the calibration strategy was revised for the regional crop modeling procedure. Maize production ERS region 1 was re-evaluated during Phase 2 and a temporal dimension was added to the process. The CERES-Maize model was also given another look to help devise a new calibration strategy.

The objectives of this study were (i) to develop a robust methodology for calibrating cultivar coefficients for the CERES-Maize crop model from the most basic of production information, i.e. yield, to a high resolution gridded dataset, (ii) to evaluate the performance of the model at the regional scale under non-limiting N conditions, and (iii) to predict irrigation demands for the region in question. This study is the initial phase of a project to calibrate and evaluate the CERES-Maize crop model for use in determining the future impacts of increased water use efficiency under climate change scenarios.

3.1. Background

The CERES-Maize crop model is a dynamic simulation model that operates on a daily time step to predict crop growth at the plot-scale. However, policy and agricultural recommendations are rarely implemented at the plot-scale. In response to this demand, crop models, such as CERES-Maize, have been adapted to national and regional scales to aid policy makers with agricultural decisions concerning at variety of issues. Areas of interest include maize production in response to climate change (Southworth, et al., 2000; Jones & Thornton, 2003; Xiong, Matthews, Holman, Lin, & Xu, 2007), global food security (Parry, Rosenzweig, Iglesias, Fischer, & Livermore, 1999), and within season of corn yield

(Quiring & Legates 2008). While the CERES-Maize model has been extensively tested and validated at the regional scale, several challenges remain when crop modeling scales larger than the farm.

Difficulties arise for any model is applied scale for which it was not designed. Models such as the CERES-Maize are designed to simulate crop growth, soil-plant-atmosphere interactions, and nutrient dynamics at a point in space over time. Each point, in this case representing a corn stalk, is then distributed over the designated plot area based on management inputs such as planting density and row spacing. After the growth cycle, outputs such as yield, irrigation**,** and evapotranspiration are considered uniform across the plot. This approach works well at the plot-scale as the assumption of homogeneity of input parameters typically holds true across the plot. Unfortunately, the same cannot be said about input parameters at the regional scale. Thus, the introduction of errors becomes an inherent consequence of the regional modeling process, forcing the modeler to dance with making generalizations necessary for the regional modeling and minimizing error.

With the advancements of GIS-based systems over the past decade, it has become possible to improve input parameter definition at the regional scale through the use of geospatially explicit data bases. The CERES-Maize model depends heavily on defining soil-plant-atmosphere interactions. As a result, soil and climatic inputs must be defined at a high resolution to minimize error. One such database used to characterize region specific soils is the ISRIC-WISE dataset (Batjes, 2006). Coupled with a gridded datset of past and present weather condition, such as the NASA Prediction of Worldwide Energy Resource) POWER Agroclimitololgy archive (NASA POWER Team, 2010), it becomes possible to simulate and evalute the accuracy of simulated yields. While, soils and climate can be applied to the regional with a high level of certaintiy, information concerning management practices are less obtainable.

Comperhensive regional datasets contiaining information describing region specific management applications are pracitically non-existant. Information on planting and harvesting dates can found from such sources ans USDA National Agriculutral Statistics Service or the SAGE Crop Calandar Dataset (Sacks, Deryng, Foley, & Ramankutty, 2010). However, defining cultivar(s) for a specific region is muc h more complex. Typical approaches addressing the issue of maize vareties include either assigning a regional common and pervioulsy defined vareity to represent the mixture cultivars in the region (Ines, Gupta, & Loof, 2002; Guo, Lin, Mo, & Yang, 2010) or deriving new cultivar definitions from crop vareity trials and applying these values to the region (Quiring & Legates, 2008; Yang, Wilkerson, Buol, Bowman, & Heiniger, 2010). In either case, one or a few cutlivar defintions are used to represent a large area which potentially contains many sets of cultivars, introducing error.

The inherent error associated with using one or a few varieties to represent an entire region offers the potential for improvement. When modeling on a high resolution grid, each grid begins to exemplify a field unit. It is to this author's knowledge that no attempts have been made at calibrating the CERES-Maize model on grid by grid basis.

3.2. Methods

3.2.1. Study Area

This study covered maize production in the USDA Economic Research Service (ERS) U.S. Farm Resource Region 1 (36.042°N to 46.625°N and 99.292°W to 82.125°W) also known as the Heartland Region from 1997-2007. The ERS regions depict geographic specialization in production of U.S. farm commodities and are derived from four major sources including the older farm production region classifications, a cluster analysis of U.S. farm characteristics (Sommer & Hines, 1991), the USDA Land Resource Region, and the National Agricultural Statistics Service (NASS) Crop Reporting Districts. The region is comprised of roughly 550 counties in nine states. The Heartland region was chosen as it represents not only one of the largest areas of maize production regions in the U.S., but the entire world.

3.2.2. Weather Inputs

Weather inputs were obtained from the NASA Agroclimatology Archive, one component of NASA's POWER (Prediction of Worldwide Energy Resource) project. POWER was created to allow access to data derived from NASA's Surface Meteorological and Solar Energy (SSE) project for those interested in the design of renewable energy systems. The Agroclimatology archive was developed with agricultural Decision Supports Systems (DSS) in mind and provides easy download of historical data for specific site locations. The parameters contained in this dataset are based upon solar radiation derived from satellite observations and meteorological data from the Goddard Earth Observing System assimilation model. The archive is globally comprehensive at 1° resolution dating back to July 1983 to near present time. Parameters selected from the this archive include daily estimates of insolation on a horizontal surface (MJ/m^2), daily mean, maximum, and minimum temperatures at 2m above ground surface (°C), and precipitation (mm).

3.2.3. Soil Inputs

Soil inputs were obtained from the International Soil Reference and Information Centre (ISRIC) World Inventory of Soil Emission Potentials (WISE) soils database. The WISE database is a globally comprehensive dataset at one of the highest resolutions available at a 5' X 5' resolution. The data were created using the FAO-UNESCO Soil Map of the World (DSMW) and soil parameter estimates derived from ISRIC's global WISE soil profile database (Batjes, 2006). The WISE database contains 1125 globally distributed soil profiles that are georeferenced and classified according to the FAO system. Soil profiles were assigned to the FAO classification within the ERS region.

3.2.4. Management Inputs

The Center for Sustainability and the Global Environment's (SAGE) Crop Calendar Dataset was used to determine planting and harvesting dates. The dataset was complied by digitizing and georeferencing existing observations of cropping systems in a gridded database (Sacks, Deryng, Foley, & Ramankutty, 2010). In order to facilitate the ease of the modeling process, a region average for both mean planting and harvesting dates was used. Inputs such as seeding density, seeding depth, and row spacing were set to regional common values (Table 4).

Table 4. Management parameter used for the CERES-Maize simulations.

Model Parameter	**Value**
Row spacing (cm)	85
Seeding density (plants/cm)	7.2
Seeding depth (cm)	4
Planting date	May 12
Harvesting date	October 22

3.2.5. Historical Yields

Historical county-average maize yields were obtained between the years 1997-2007 for the counties of ERS region 1 from USDA's National Agricultural Statistics Service (NASS) dataset. In their current state, the data is not capable of being used for a grid base regional analysis and as such, need to be gridded. To achieve this goal, each county yield value was georeferenced to its corresponding county polygon in ArcGIS. Yields were then calculated for each pixel on a 5' X 5' grid using an area weighting approach. The process involved using equation [3];

$$Y_{i,n} = \sum_{i=1}^{n} Y_{c,n} \left(\frac{A_{i,c}}{A_c}\right) \qquad [3]$$

where $Y_{i,n}$ is equal to the cell grid yield in year n, $Y_{c,n}$ is the reported NASS county yield that overlaps grid cell i, A_c is the area of the county, and $A_{i,c}$ is the fraction of the grid cell i that falls within county polygon. The newly created grid base yield polygons were then converted to yearly raster for the modeling process.

3.2.6. Genetic Coefficient Estimation

CERES-Maize uses six genetic coefficients to simulate a maize plant's response to both photoperiod and temperature (Table 5). This study focused on deriving a genetic coefficient profile for each pixel to describe all potential cultivars with the pixel area. To accomplish this task, each coefficient was optimized in a step-like fashion, being with P1 and ending with G3. The Root Mean Squared Error (RSME) was used to determine the coefficient associated with the least amount of error in yield estimation [4];

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_s - Y_o)^2} \qquad [4]$$

where n is equal to the number of years with a non-zero yield value for the grid cell, Y_s is the simulated yield and Y_o is the area weighted average NASS yield.

Table 5. CERES-Maize genetic coefficient definitions.

Symbol	Description	Lower limit	Upper limit	Interval size
P1	Thermal time from seedling emergence to the end of the juvenile phase (expressed in degree days above a base temperature of 8°C) during which the plant is not responsive to changes in photoperiod.	150	300	10
P2	Extent to which development (expressed as days) is delayed for each hour increase in photoperiod above the longest photoperiod at which development proceeds at a maximum rate (which is considered to be 12.5 hours).	0.45	0.45	0
P5	Thermal time from silking to physiological maturity (expressed in degree days above a base temperature of 8°C).	600	1000	10
G2	Maximum possible number of kernels per plant.	450	1000	50
G3	Kernel filling rate during the linear grain filling stage and under optimum conditions (mg/day).	5	12	0.5
PHINT	Phylochron interval; the interval in thermal time (degree days) between successive leaf tip appearances.	38.9	38.9	0

3.2.7. Validation

In order evaluate the performance of the model resulting model, the derived genetic coefficients were used to simulate yields for the year 2007, using the RSME as the evaluation criteria. The RSME was applied to each grid cell.

3.3. Sensitivity Analysis

In order to reduce the potential number of simulations, a sensitivity analysis was conducted for each of the genetic coefficients with the exception of P2 and PHINT, which were held constant, to determine the optimal range for each coefficient for the calibration stage. Yang, Wilkerson, Buol, Bowman, & Heiniger (2010) found that P2 was constant with respect to yield below 0.39 and above 0.50, concluding a median value of 0.45 was sufficient to achieve yields in the middle of the variability range. PHINT was also held constant in compliance with most other predetermined cultivars in DSSAT 4.0. Using a broad set of ranges for each coefficient (see Table 2), the sensitivity analysis involved incremental changing one coefficient, while holding all other constant according to the mediums season variety, and running the CERES-Maize model. The resulting yield was plotted against the coefficient value for each coefficient. The analysis was run for 100 randomly selected pixels in ERS region 1.

Figures 11-14 show the results for each of the coefficient for a selected. Overall, all of the four coefficients show a universal positive trend within the ranges analyzed. The results suggest no drastic sensitivity. These could be a result of not extending the coefficient range enough during the analysis; however, more testing is needed.

3.4. Discussion

Coding for the calibration of the ERS region 1 is currently underway. More testing of the sensitivity of the inputs is needed to better define the genetic coefficients for the optimization process. The resulting product should provide a sufficient definition of maize cropping in the Heartland Region. Once a calibration procedure has been successfully validated, the process can be applied to future climate change scenarios.

4. Conclusion

The goal of this project was to evaluate the potential to integrate DSSAT into a global data framework for integrated crop production analysis. The results clearly indicate this is possible. The first objective of this phase was to develop a robust methodology for calibrating pre-established cultivar definitions in DSSAT for the CERES-Maize crop model from publically available global datasets. That objective has been met. The second objective was to evaluate the performance of the model at the regional scale under non-limiting N conditions. That objective was also met. The final objective was to predict irrigation demands for the region in question, and estimate blue vs. green distribution on a global scale. This objective was completed with high effectiveness at the global scale and less than satisfactory resolution at the regional scale. Regional data are more explicit so the biases of the model are more apparent at smaller scales. The remaining challenge is to develop a geospatial water management data input process. This work is ongoing under another program.

A research team called HarvestChoice (http://harvestchoice.org/about/at_a_glance), supported by the Bill & Melinda Gates foundation, is non-profit organization that focuses on generating knowledge bases for strategic investment in improve the quality of life in poverty stricken sub-Saharan Africa and South Asia. This group has been engaged in similar studies that are trying to predict and optimize future cropping systems. The next stage for this project is to link the model calibration and validation tools to Harvest Choice to provide a global set of tools for analyzing the impacts of environmental and economic pressures on global crop production.

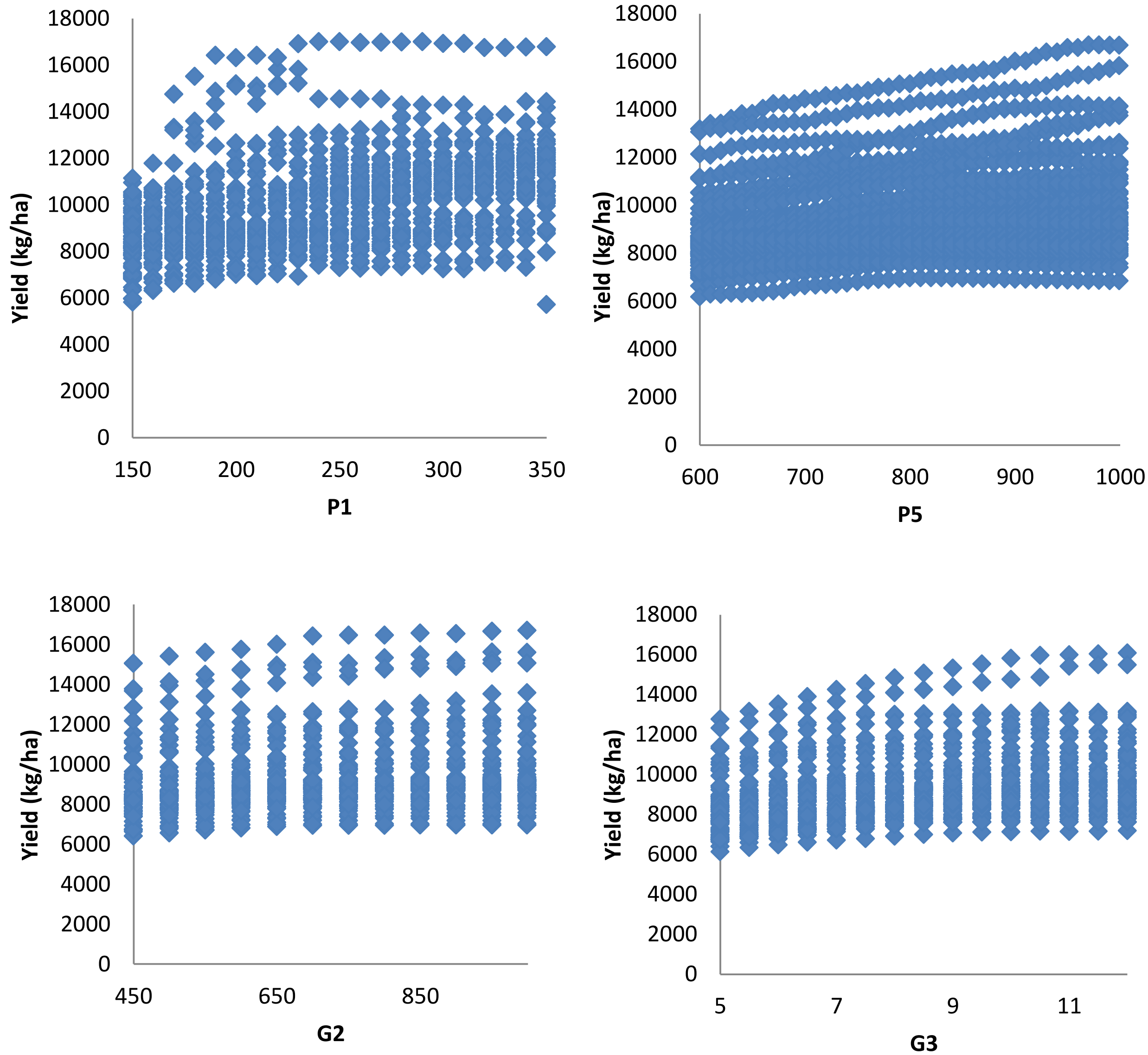

Figure 11. Simulated yield versus genetic coefficient P1 for the year 1999 (Top Left), simulated yield versus genetic coefficient P5 for the year 2001 (Top Right), simulated yield versus genetic coefficient G2 for the year 1999 (Bottom Right), and simulated yield versus genetic coefficient G5 for the year 2001 (Bottom Left).

Literature Cited

Al-Kaisi, M., & Yin, X. (2003). Effects of nitrogen rate, irrigation rate, and plant population on corn yield and water use efficiency. *Journal of Agronomy, 95*, 1475-1482.

Batjes, N. H. (2006). *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (version 1.0).* Wageningen: ISRIC - World Soil Information.

Brouwer, C., & Heibloem, M. (1986). *Irrigation water management: irrigation water needs.* (Vol. No. 3). Rome: FAO.

Grassini, P., Yang, H., & Cassman, K. G. (2009). Limits to maize productivity in the Western Corn-Belt: A simulation analysis for fully irrigated and rainfed conditions. *Agricultural and Forest Meteorology, 149*(8), 1254-1265.

Guo, R., Lin, Z., Mo, X., & Yang, C. (2010). Responses of crop yield and water use efficiency to climte change in the North China Plain. *Agricultural Water Management, 97*, 1185-1194.

Hoogenboom, G., Jones, J. W., Porter, C. H., Wilkens, P. W., Boote, K. J., Batchelor, W. D., et al. (2004). Decision Support System for Agrotechnology Transfer Version 4.0 [CD-ROM]. University of Hawii, Honolulu.

Howell, T., Tolk, J., Schneider, A., & Evett, S. (1998). Evapotranspirationl, yield, and water use efficiency of corn hybrids differing in maturity. *Journal of Agronomy, 90*, 3-9.

Ines, A. V., Gupta, A. D., & Loof, R. (2002). Application of GIS and crop growth models in estimating water productivity. *Agricutural Water Management, 54*, 205-225.

Jones, C. A., & Kiniry, J. R. (1986). *CERES-Maize, a simulation model of maize growth and development.* College Station: Texas A&M University Press.

Jones, P. G., & Thornton, P. K. (2003). The potential impacts of climate change on maize production in Africa and Latin America in 2055. *Global Environmental Change, 13*, 51-59.

NASA POWER Team. (2010). *Sufarce meteorology and solar energy (SSE) release 6.0 methodology VERSION 2.4.1.* NASA POWER Agroclimatology, Retrived from http://power.larc.nasa.gov/documents/SSE_Methodology.pdf.

Parry, M., Rosenzweig, C., Iglesias, A., Fischer, G., & Livermore, M. (1999). Climate change and world food security: a new assessment. *Global Environmental Change, 9*, S51-S67.

Quiring, S. M., & Legates, D. R. (2008). Application of CERES-Maize for within-season prediction of rainfed corn yields in Delaware, USA. *Agricutlural and Forest Meteorology, 148*, 964-975.

Ramankutty, N., Olejniczak, N. J., & Foley, J. A. (2000). Harvested Area and Yields of 175 crops (M3-Crops Data). Madison, Wisconsin, USA: Department of Geography, McGill University.

Sacks, W. J., Deryng, D., Foley, J. A., & Ramankutty, N. (2010). Crop planting dates: an analysis of global patterns. *Global Ecology and Biogeography, 19*, 607-620.

Southworth, J., Randolph, J. C., Habeck, M., Doering, O. C., Pfeifer, R. A., Rao, D. G., et al. (2000). Consequences of future climate change and changing climate variability on maize yields in the midwestern Unitied States. *Agriculture, Ecosystems and Environment, 82*, 139-158.

Tijani, F., Oyedele, D., & Aina, P. (2008). Soil moisture and water-use efficiency of maize planted in succession to different fallow treatment. *International Agrophysics, 22*, 81-87.

Xiong, W., Matthews, R., Holman, I., Lin, E., & Xu, Y. (2007). Modelling China's potential maize production at regional scale under climate change. *Climatic Change, 85*, 433-451.

Yang, Z., Wilkerson, G. G., Buol, G. S., Bowman, D. T., & Heiniger, R. W. (2010). Estimating genetic coefficents for the CSM-CERES-Maize model in North Carlonia Environments. *Agronomy Journal, 101*(5), 1276-1285.