Principal Component Analysis of Crop Yield Response to Climate Change

Ruohong Cai, John C. Bergstrom, Jeffrey D. Mullen, Michael E. Wetzstein, and W. Don Shurley

Principal Component Analysis of Crop Yield Response to Climate Change

Ruohong Cai

John C. Bergstrom

Jeffrey D. Mullen

Michael E. Wetzstein

W. Don Shurley

Graduate Research Assistant, Professor, Associate Professor, Professor, and Professor,

Department of Agricultural and Applied Economics, University of Georgia, Athens, GA 30602

_____

Dept. of Agricultural & Applied Economics

College of Agricultural & Environmental Sciences

University of Georgia

_____

Principal Component Analysis of Crop Yield Response to Climate Change

Ruohong Cai

John C. Bergstrom

Jeffrey D. Mullen

Michael E. Wetzstein

W. Don Shurley

Department of Agricultural and Applied Economics
University of Georgia
Athens, GA 30602-7509
tsai@uga.edu

The objective of this study is to compare the effects of climate change on crop yields across different regions. A Principal Component Regression (PCR) model is developed to estimate the historical relationships between weather and crop yields for corn, soybeans, cotton, and peanuts for several northern and southern U.S. states. Climate change projection data from three climate models are applied to the estimated PCR model to forecast crop yield response. Instead of directly using weather variables as predictor variables, the PCR model uses weather indices transformed from original weather variables by the Principal Component Analysis (PCA) approach. A climate change impact index (CCII) is developed to compare climate change effects across different regions. The key contribution of our study is in identifying a different climate change effects in crop yields in different U.S. states. Specifically, our results indicate that future warmer weather will have a negative impact for southern U.S. counties, while it has insignificant impact for northern U.S. counties in the next four decades.

Key words: Principal component regression, Crop yield response, Climate change.

Introduction

Contemporary, state-of-the art general circulation models (GCM) including the Australian CSIRO 3.5, Canadian CGCM 3.1 and Japanese MIROC 3.2 models all predict that average temperature will keep rising and precipitation will have a mild change for most states in the continental United States for the rest of the century assuming greenhouse gas emissions follow the IPCC SRA1B scenario[1]. Climate change has direct impacts on agriculture. For example, variations in climatic conditions such as a late spring, a rainy planting season, or a hot and dry growing season could all affect crop yields. Although agricultural technologies have been greatly improved in recent decades, many researchers believe that these possible variations in both temperature and precipitation still have adverse impact on crop yields in many regions (Tao et al. 2006; Schelenker and Roberts 2006; Lobell, Cahill, and Field 2007; Almaraz et al. 2008). While most researchers agree that warmer climate can be harmful to crops in the US, limited studies directly compare the effects of climate change on crop yields across different regions. Different regions have different agricultural technologies and environmental conditions such as soil properties; therefore, climate change is expected to have different effects on different regions. For the same reason, it is hard to directly compare assessment results generated in different regions. Besides, different studies differ in their scope, climate variable selection, scope and methodology. In this study, a PCA approach has been developed to compare the effects of climate change across different regions.

---

[1] A future world of very rapid economic growth, low population growth and rapid introduction of new and more efficient technology. Major underlying themes are economic and cultural convergence and capacity building, with a substantial reduction in regional differences in per capita income. In this world, people pursue personal wealth rather than environmental quality (IPCC 2007).

Generally, researchers use two major methodologies to study the relationship between weather and crop yields. One is a crop growth model, and the other is regression analysis. Both methodologies have their advantages and disadvantages.

Crop growth modeling is a computer-based simulation approach used by many agronomists. It is a mathematical representation of the integration of the disciplines of biology, physics, and chemistry (Jones 1993; Hoogenboom 2000; Jones et al. 2003). Based on the agronomic knowledge of the crop growing process, this model incorporates weather information including temperature, precipitation, solar radiation, and humidity with other factors including fertilizer applications and soil properties to simulate their impact on that process through a set of mathematical equations. The expected crop yields and their variance are generated by the model simulation. The Decision Support System for Agrotechnology Transfer (DSSAT) is a very popular software package using this method (Hoogenboom 2000).

However, crop growth modeling is highly complex and requires extensive information such as weather, soil and management options in order to simulate the crop growing process. Such information is usually incomplete and sometimes unavailable (Walker 1989). Because of this disadvantage of crop simulation modeling, regression analysis is an alternative method used for predicting yields in many yield forecasting studies by economists (Horie, Yajima, and Nakagawa 1992; Kandiannan et al. 2002; Tannura, Irwin, and Good 2008).

The regression analysis approach applies statistical methods to link historical crop yields to weather variables. Compared to the crop growth modeling approach, data limitations are less of a concern in the regression analysis approach. Furthermore, Tannura, Irwin and Good (2008) found that regression models have high explanatory power.

The first step for constructing a weather-crop yield regression model is choosing the appropriate weather factors affecting crop growth. In the literature, temperature and precipitation have been demonstrated to have significant impacts on crop yields. Agronomically, high temperature would affect soil moisture levels which could decrease crop yields if irrigation water supply is not sufficient (Mitchell et al. 1990). On the other hand, precipitation maintains necessary soil moisture for crop growth. Different temperatures could change growing season lengths inducing variations in crop yields. For example, high temperature tends to shorten many crop growing seasons. A short growing season exposes crops to less solar radiation needed for photosynthesis. In the long run, temperature and precipitation change could alter cropping patterns in many regions (Lotsch et al. 2007). Unstable temperature and precipitation will cause unpredictable variations in crop yields. Overall, temperature and precipitation are the two most important weather factors affecting crop yields. Therefore, we used monthly temperature and precipitation during the growing season as weather variables for our weather-yield model.

After choosing weather factors, the next step in developing a weather-crop yield regression model is to select specific predictor variables. A typical growing season is seven or eight months or even longer for different crops in different regions. This means there could be about 14 to 16 predictor variables for both temperature and precipitation in the weather-crop yield regression model. One disadvantage of having too many predictor variables is a low degree of freedom which causes unstable estimation results. Although this issue could be solved by increasing the sample size, weather and yield data are not always available for early years.

In the literature, many methodologies have been utilized to reduce the number of weather variables in regression models. Instead of using calendar months, some researchers divide the growing season by the stages of crop growth, therefore decreasing the number of weather

variables (Dixon et al. 1994; Kafumann and Snell 1997). The active growing season for corn in the State of Georgia is from April to October; therefore we have seven monthly temperature variables. If we divide the growing season by the four general growth stages for corn, we could reduce the number of temperature variables to four. Agronomically, this method makes more sense than using calendar months. However, it is hard to specify the exact boundary between two crop growth stages. Also, the timing of crop growth stages varies from year to year, and from region to region. Alternatively, researchers intensively utilized statistical variable selection techniques to reduce the number of calendar months or weeks as predictor variables.

Researchers have used t-statistics, R-squared, and other statistical techniques to select significant weather variables. Statistical methods exclusively lean on data while ignoring agronomic implications of different growing season months; therefore, statistical methods alone could make agronomic mistakes due to data limitations. For example, utilization of traditional statistical variable selection techniques could drop important months and keep unimportant months by chance. In addition, the weather variables are correlated, a direct variable selection approach will lead to unstable estimated coefficient. Therefore, using statistical variable selection techniques alone to select appropriate weather variables is unreliable when developing weather-crop yield regression models.

It is commonly agreed that the effects of weather conditions on crop yields are not simple linear relationships (Deschenes and Greenstone 2007; Schlenker and Roberts 2009). For example, the response of crop yields to a one degree Celsius increase in temperature depends on the baseline temperature. Therefore, some previous studies include quadratic terms for weather variables in order to account for the nonlinearity of weather variables effects on crop yields.

However, the issue of too many weather variables becomes even more severe when quadratic terms for weather variables are introduced into the regression models.

Agronomically speaking, crop growth is a cumulative process. Weather conditions in any growing months would affect final realized crop yields. For this reason, it is preferable to keep all growing season months in the regression model. However, as mentioned earlier, keeping too many predictor variables in the regression model could lead to low degrees of freedom. To balance the choice between keeping all the growing months in the model while reducing the number of variables, Principal Component Regression (PCR) is applied in this study. Additionally, using PCR model also helps eliminate possible severe multicollinearity issue in multiple regression models (Dixon et al. 1994). When severer multicollinearity exists among predictor variables, the estimated coefficients are sensitive to the selection of predictor variables.

Instead of taking original weather variables as predictor variables, the PCR model uses Principal Components (PCs) generated by using Principal Component Analysis (PCA) technique. PCA is a variable compression technique. It transforms a large number of interrelated variables to a new set of uncorrelated PCs which are linear combinations of the original variables (Jolliffe 2002). Therefore, each principal component contains information on all weather variables. By using PCA, none of growing season months will be completely omitted regardless of which variable selection technique is applied. Many previous studies have used PCR models (Pandzic and Tminic 1992; Yu, Chu, and Schroeder 1997; Hansen, Jorgensen, and Thomsen 2002; Martinez, Baigorria, and Jones 2009).

PCA generates the same number of weather indices as the original weather variables (including both original and quadratic terms) and orders them by the magnitude of variances. In order to reduce the number of predictor variables, the rule of thumb for most previous related

studies is to take only the first several PCs with eigenvalue greater than or equal to one as predictor variables. The eigenvalue represents the variance displayed by each PCs axis. Some previous researchers even used only the first principal component in a PCR model (Martinez et al., 2009). It should be noted that it is not appropriate to use only the PCs with large variances in the PCR model (Jolliffe 1982; Jolliffe 2002). Hadi and Ling (1998) showed that it is possible for the PCs with the smallest variance to be the only predictor variable correlated to response variable. Weather variables with larger variance are not necessary more important than weather variables with smaller variance for crop growth. By using only the first several principal components with larger variances, it is possible to omit important predictor variables with small variances. Therefore, instead of simply taking the first several principal components as predictor variables, we use statistical variable selection technique to select an appropriate subset of PCs.

In previous studies, application of PCA in climatic data modeling has concentrated on investigating geographic patterns of temperature and precipitation. The constraints and interdependency of spatiotemporal climate data can be identified by the use of PCA (Preisendorfer 1988). Since different units and magnitudes could dominate the grouping of principal components, temperature and precipitation variables need to be standardized before PCA is employed (Jolliffe 2002).

The study conducted by Kantanantha, Serban and Griffin (2010) is one of few to use PCR to study the relationships between crop yields and weather indicators. However, in their study, temperature and precipitation variables were processed under PCA separately, which leaves multicollinearity issue between precipitation and temperature unsolved. Also, they only use original terms for temperature and precipitation variables. In this study, we include quadratic terms for temperature and precipitation to account for the nonlinear relationships between crop

9

yields and weather. Besides the mean value of weather indicators, the effects of weather variability on crop yields are also likely to decrease crop yields (Porter et al. 2005), such as increased extreme weather events. For this reason, we include the difference of mean daily maximum and minimum temperatures of the month to account for the effects of extreme temperature events on crop yields.

Technology change has an important role in long-run crop yield changes since it improves the crop yields over time. In previous studies, researchers usually add an additional trend predictor into the weather-crop yield regression model to represent technology change. Possible candidates for this trend predictor include GDP and a linear or nonlinear time trend (Buller 1972; Choi and Helmberger 1993; McCarl, Villavicencio and Wu 2008). Nonlinear time trends were also used since yield improvement is not necessary linear which could happen when improved crop varieties were used.

In this study, the primary purpose of the regression analysis is to investigate the relationship between crop yields and weather. By adding GDP or a time trend to represent technology change, the significance of the model will be mostly explained by this technology trend removal instead of weather variables. Although it is statistical appropriate to add GDP or a time trend to the regression model, it is better to use de-trended yield to investigate the connections between weather variables and crop yields. Another issue related to using GDP or a time trend variable is spatial difference. Assuming a linear technology trend, some places could benefit from larger technology advances. By using a technology trend predictor, spatial differences in technology will determine the main trend when forecasting changes in crop yields.

Methodology

In this study, a weather-crop yield PCR model is developed to study the response of crop yields to weather changes. Before presenting the PCR model, it is necessary to show how PCs are generated. The following is a generalized expression of the crop yield regression model. It is a statistical function that demonstrates the historical relationship between weather variables and crop yields. After holding all other inputs such as fertilizers and insect infestations constant, equation (1) estimates the connections between crop yields and weather conditions:

(1) $\qquad y = X\beta + \epsilon$

where X is a matrix of p random variables of weather indicators with dimension $n \times p$, y is a vector of crop yields with n observations, β is a vector of p regression coefficients and ϵ is a vector of error terms. The weather vectors within X matrix include monthly temperature, square term of monthly temperature, monthly precipitation, square term of monthly precipitation, and the monthly difference of mean daily maximum and minimum temperatures.

As previously mentioned, the PCR model uses PCs of weather indicators as predictor variables. The first step of the PCR model is generating PCs of the original variables. PCA derives the same number of PCs as the original variables. These PCs are ranked by the magnitude of their variances. The first PC of weather indicators with the largest variance could be represented as a linear combination of original variables:

(2) $\qquad \underbrace{X\boldsymbol{\alpha_1}}_{(n\times 1)} = x_1\alpha_{11} + x_2\alpha_{12} + \cdots + x_p\alpha_{1p} = \sum_{j=1}^{p} \alpha_{1j}x_j,$

where $\boldsymbol{\alpha_1}$ is an eigenvector of covariance matrix $\boldsymbol{\Sigma}$ of $\boldsymbol{X}$ corresponding to its 1st largest eigenvalue λ_1. The second and following PCs could be generated in a similar way. A list of PCs of weather indicators is:

(3)
$$\underset{(n \times p)}{\underline{Z}} = \underset{(n \times p)}{\underline{X}} \, \underset{(p \times p)}{\underline{A}} = [X\alpha_1, X\alpha_2, \ldots, X\alpha_p],$$

where all the PCs are mutually independent.

Since A is orthogonal, we can rewrite equation (1) as:

(4)
$$\underset{(n \times 1)}{\underline{y}} = X\beta + \epsilon = XAA'\beta + \epsilon = \underset{(n \times p)}{\underline{Z}} \, \underset{(p \times 1)}{\underline{\gamma}} + \epsilon$$

Equation (4) is a general expression of the PCR model. The PCs matrix Z replaces the original weather variables matrix X. Z has the exact same dimension $n \times p$ as X. Since a key advantage of the PCR model compared to a standard regression model is reducing the number of predictor variables, it is preferable to select a subset of principal components to use as predictor variables in the PCR model. Besides, if all the PCs are used as the predictor variables, the resulting model will be equivalent to the standard regression model (Jolliffe 2002). As mentioned previously, many studies only selected the first several principal components claiming that these components explain most of the variance in the data. This selection rule for PCs is not appropriate since even if a particular principal component explains only a small variance in the weather variables, this does not necessarily mean it has weak power in explaining the variance in crop yields (Jolliffe 1982). However, there is no general advice existing about the PCs selection rules. Out of the several variable selection techniques discussed by Jolliffe (2002), we chose to use p-value criterion where the coefficients with a p-value less than 10% be kept in the model.

A PCR model for which *p-k* components have been eliminated is:

(5)
$$\underset{(n \times 1)}{\underline{y}} = \underset{(n \times k)}{\underline{Z_k}} \, \underset{(k \times 1)}{\underline{\gamma_k}} + \epsilon_k,$$

where γ_k is a vector of k PCR coefficients, Z_k is an $(n \times k)$ matrix whose columns are the

corresponding subset of columns of $\boldsymbol{Z}$, and ϵ_k is the appropriate error term. These k PCs that remained in the model of equation (5) are not necessarily the first k PCs.

Compared to many other research areas, climate change study is special in that future observations (climate change projection data) are usually already available before conducting the research. This also requires a different approach in PCR forecasting. To predict future crop yields with estimated PCR model, a direct approach is to generate a new set of PCs by directly transforming future weather variables. However, a new set of PCs would have completely different eigenvectors from those transformed from historical weather variables. Therefore, this approach is not appropriate. An alternative approach (Approach 1) is to use eigenvectors associated with historical weather variables to construct PCs for future weather variables. However, using eigenvectors based on historical data to generate PCs for future data is inappropriate, since previous eigenvectors are uniquely selected for historical data.

To solve above issue of utilizing a PCR model for forecasting purposes, we propose to transform weather variables into PCs using both historical climate data and future climate change data simultaneously (Approach 2). One set of future climate change data is specifically related to a particular estimated PCR model. Thus, although we only have one set of historical weather data, there will be three different PCR models for forecasting based on three climate change projection data sets. Approach 2 avoids applying historical eigenvectors to future data sets. Meanwhile, a concern about its prediction performance is raised since estimated model is a function of future observations, which is different from traditional econometric approaches. To test the prediction performance for Approach 1 and Approach 2, a Monte Carlo experiment is designed to compare their mean squared errors (Monte Carlo experiment procedure and results are showed in the Appendix section). After running the experiment for 1000 times, it is observed

13

that the differences between the averages mean squared error for both approaches are insignificant. Therefore, Approach 2 is chosen for its advantage mentioned earlier. To the best of our knowledge, this approach has never been conducted in previous literature.

Based on the above discussion, we modify equation (5) to a PCR model that is applicable to forecasting purpose. Suppose we have three sets of climate change data $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2$, and $\boldsymbol{\Phi}_3$ each with dimension of $(m \times 1)$. Each of them are combined with historical weather data X to generate three sets of weather data $\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Psi}_3$ as follows:

(6a)
$$\underbrace{\boldsymbol{\Psi}_1}_{((n+m)\times p)} = \left[\underbrace{X'}_{(p\times n)} , \underbrace{\boldsymbol{\Phi}_1'}_{(p\times m)} \right]'$$

(6b)
$$\boldsymbol{\Psi}_2 = \left[X', \boldsymbol{\Phi}_2' \right]'$$

(6c)
$$\boldsymbol{\Psi}_3 = \left[X', \boldsymbol{\Phi}_3' \right]'$$

$\boldsymbol{\Psi}_1$, $\boldsymbol{\Psi}_2$ and $\boldsymbol{\Psi}_3$ were then transformed into PCs matrices Z_1, Z_2 and Z_3:

(7a)
$$\underbrace{Z_1}_{((n+m)\times p)} = \underbrace{\boldsymbol{\Psi}_1}_{((n+m)\times p)} \underbrace{A_1}_{(p\times p)} = \left[\boldsymbol{\Psi}_1\alpha_{11}, \boldsymbol{\Psi}_1\alpha_{12}, \dots, \boldsymbol{\Psi}_1\alpha_{1p} \right] = [\dot{Z}_1', \ddot{Z}_1']'$$

(7b)
$$Z_2 = \boldsymbol{\Psi}_2 A_2 = \left[\boldsymbol{\Psi}_2\alpha_{21}, \boldsymbol{\Psi}_2\alpha_{22}, \dots, \boldsymbol{\Psi}_2\alpha_{2p} \right] = [\dot{Z}_2', \ddot{Z}_2']'$$

(7c)
$$Z_3 = \boldsymbol{\Psi}_3 A_3 = \left[\boldsymbol{\Psi}_3\alpha_{31}, \boldsymbol{\Psi}_3\alpha_{32}, \dots, \boldsymbol{\Psi}_3\alpha_{3p} \right] = [\dot{Z}_3', \ddot{Z}_3']'$$

The PCR model will result in the following, where each of them have a unique PCs matrix $\dot{Z}_1$, $\dot{Z}_2$ and $\dot{Z}_3$, depending on the climate change data that will be used for yield forecasting. Although the PCs used in the PCR model are still based on historical weather data, they are different since future weather data are combined to generate PCs.

14

$$
\text{(8a)} \qquad \underbrace{y}_{(\boldsymbol{n}\times\boldsymbol{1})} = \underbrace{\dot{Z}_1}_{(\boldsymbol{n}\times\boldsymbol{p})} \underbrace{\boldsymbol{\gamma}_1}_{(\boldsymbol{p}\times\boldsymbol{1})} + \epsilon
$$

$$
\text{(8b)} \qquad y = \dot{Z}_2\boldsymbol{\gamma}_2 + \epsilon
$$

$$
\text{(8c)} \qquad y = \dot{Z}_3\boldsymbol{\gamma}_3 + \epsilon
$$

After estimation, these three climate models were then used to predict crop yields based on the estimated PCR models (see equations 8a, 8b, and 8c).

The predicted yields from the PCR model using future climate change data were used to generate a Climate Change Impact Index (CCII). Forty one predicted yields for certain crops in selected counties were compared between three climate change models. The number of years for which MIROC 3.2 (warmest climate scenario) generates lower crop yields as compared to the CSIRO 3.5 (coldest climate scenario) was recorded. The CCII is generated by dividing the number of these particular years by total years and the state-level CCII is generated by take the average of county-level CCII. The equation for a state-level CCII is as follows:

$$
\text{(9)} \qquad CCII_s = \frac{\sum_{i=1}^{c} \varphi}{c} * \frac{1}{41}
$$

where s denotes specific states, c denotes number of counties in specific states, φ denotes the number of years for which MIROC 3.2 (warmest) generates lower crop yields as compared to CSIRO 3.5 (coldest). Each county has different crop acreage; CCII for a county with higher crop acreage should have more weight than a county with lower crop acreage in equation (9). Therefore, weighted state-level CCII was also generated. The CCII was generated by using the above formula for several northern and southern U.S. states. A high value of CCII indicates that global warming is a serious influence on crop yields, while a low value of CCII indicates that global warming is less of a concern.

Data

Climate represents long-run weather patterns. Therefore, to study the historic effects of climate on agriculture, the longest possible period is preferred. Due to data availability, 1960 is the earliest year with both of the required crop yield and weather data available. Thus, our study observes the period 1960-2009.

Because they are the top producing states for corn, soybeans, cotton and peanuts, we selected the northern U.S. states of Minnesota, Nebraska, Indiana, Illinois, Iowa, and the southern U.S. states of Georgia , Alabama, Texas for our empirical application. County-level crop yields and weather data in these states were analyzed using the PCR model. County-level data were used since there are large variations in crop yields due to different local environmental conditions.

The historical weather data which includes monthly average temperature, monthly average temperature difference and monthly total precipitation were retrieved from the National Climate Data Center (NCDC). We selected monthly temperature and precipitation as weather variables because of the availability of these variables in the climate data for our time period of analysis. Data was retrieved only for months during which crops are grown. The following table summarizes planting months used in the PCR model associated with specific crops and states.

Monthly average temperatures used in this study were the difference between monthly mean maximum temperature and monthly mean minimum temperature. It is also preferable to include some predictors to account for variations in precipitation. Although daily precipitation data is available to calculate precipitation variation within a month, it is not available in the NCDC data. Thus, for the purpose of forecasting yield response under climate change, we did not account for precipitation variation.

To predict future climate change with alternative greenhouse gas scenarios, many climate change models have been developed by atmospheric scientists. Different climate change models provide different climate change projections based on different approaches and underlying scenarios. In this study, three climate change projections were developed by the United States Department of Agriculture (USDA) Forest Service as part of the 2010 Renewable Resources Planning (RPA) Act assessment of natural resource demand and supply in the U.S. The projections were derived from global climate models: CGCM 3.1, CSIRO 3.5 and MIROC 3.2 which are based on the SRA1B scenario from the Special Report on Emission Scenarios (SRES) of IPCC (Coulson *et al*. 2010; IPCC 2007; Nakicenovic *et al*. 2000). Climate change projections from these three climate models provide monthly projection of temperature and precipitation up to the year 2100. In general, as the time horizon increases, crop yield response forecasting will become more unreliable. Therefore, we only use the climate change data up to the year 2050. The following weather indicator projections were used for crop yield forecasting: total annual precipitation, monthly mean maximum temperature and monthly mean minimum temperature.

Annual corn, soybeans, cotton and peanuts yield data was collected from USDA's National Agricultural Statistical Service (NASS) for the past 50 years for Illinois, Minnesota, Nebraska, Iowa, Indiana, Georgia, Texas and Alabama at county-level. Yield data were de-trended to a 2009 technology level.

A large amount of data including historical weather and crop yield data and climate change projection data are needed for crop yield response estimation. However, most of original data are not in the correct form; therefore, SAS was used to reconstruct the original data in a manner compatible with estimating the relationships between climate change and crop yields.

Results and Discussion

Yield prediction is the main concern in this paper; therefore, we don't discuss the loadings of weather variables in the specific PCs and estimated coefficients in the PCR model which are of less importance for prediction. We used the estimated coefficients from our estimated PCR model to predict future crop yields up to the year 2050, incorporating predicted changes in temperature and precipitation based on the US Forest Service climate change predictions. The corn yield response to climate change projected by three climate change models: CSIRO 3.5 (coldest scenario), CGCM 3.1 (middle scenario), MIROC 3.2 (warmest scenario) under the SRA1B scenario were forecasted and compared for both counties.

For Mitchell County, Georgia (see figure 1), 31 out of 41 forecast years showed lower predicted corn yields under MIROC32 (warmest scenario) compared to predicted corn yields under CSIRO35 (coldest scenario). Consistent with previous studies, this result indicates that warming temperatures under future climate change scenarios will tend to reduce corn yields. For Hancock County, Illinois (see figure 2), 18 out of 41 forecast years showed lower corn yields under MIROC32 (warmest scenario) compared to predicted corn yields under CSIRO35 (coldest scenario). This result for Hancock County is different from Mitchell County which indicates the existence of different effects of climate change on different regions. Farms in the northern U.S. may actually benefit from warming temperatures compared to farms in the southern U.S. where temperatures are comparatively higher.

To provide stronger evidence for the above hypothesis, county level studies were conducted for all available counties in the eight specified states. Similar analyses were replicated for soybeans, cotton and peanuts. Generated state-level CCIIs are presented in Table 1. We observe that CCII weighted by county acreage has close value to the unweighted CCII. It is

observed that northern states have a lower CCII value compared to southern states in terms of corn and soybeans. For corn, CCII for northern states range from 0.506 to 0.546, while CCII for southern states range from 0.560 to 0.692. Especially, CCII for Georgia and Alabama are both close to 0.70 which is significant higher than CCII for northern states. For soybeans, CCII for southern states range from 0.488 to 0.561, while CCII for southern states range from 0.598 to 0.706. Again, CCII for Georgia and Alabama are both close to 0.70, much higher than that of northern states. The results indicate that corn and soybeans' yields generally have a mild decrease due to predicted global climate change in the northern U.S. states studied, and a relatively more pronounced negative effect in the southern U.S. states studies where warm temperatures and periodic drought already pose significant constraints to crop production. This result is consistent with the expectation that a probable impact of global climate change, should it occur as predicted, would be to shift some cropping patterns from the southern U.S. to the northern U.S.. Cotton and Peanuts yield data are not available in the Corn Belt states, therefore their CCII are only compared between Georgia and Texas. Cotton CCII value is higher in Georgia compared to Texas, while Peanuts CCII value is higher in Texas compared to Georgia.

Conclusions

In this research, we conducted an econometric analysis of weather factors influencing crop yields using county level data from major producing states for corn, soybeans, cotton and peanuts. Specifically, a PCR model was developed with weather indices for monthly temperature and precipitation and their quadratic terms. We used an estimated PCR model to forecast the future crop yields in response to weather change projections based on three climate change models: CSIRO 3.5 (coldest), CGCM 3.1 (middle) and MIROC 3.2 (warmest). The southern U.S.

counties generally displayed lower predicted corn and soybeans yields associated with warming temperature climate change projections, while the coldest climate change projections tend to result in higher predicted corn and soybeans yields. This indicates that global warming could have a negative impact on southern counties. In the northern U.S. counties studied, the warmest climate change projections resulted in slightly higher predicted corn and soybeans yields compared to predicted corn and soybeans yields under the coldest climate change projections. This demonstrates that global warming trends may benefit corn and soybeans production in the northern U.S., while negatively impacting corn and soybeans production in the southern U.S.. The results suggest that farmers in the south could cope with climate change by switching to crops with better warm tolerance.

Overall, this research contributes to the literature in a number of ways. First, it is one of the first applications of PCA to estimate the relationships between weather and crop yields. We improve upon previous PCR models by adding quadratic terms for weather variables and temperature variations. Although these terms have been considered in traditional regression models, they have never been applied in the PCR models. We also argue that previous related studies made mistakes by generating separate PCs for temperature and precipitation. To the best of our knowledge, this research is the first to note that it is inappropriate to apply different future data sets (for example, climate change data under different scenarios) to the same estimated PCR model. Future data is a determinant in estimating the PCR model (e.g., future data affects the PCR model by influencing how PCs are standardized and transformed).We also contribute to the literature by demonstrating different effects of climate change in northern and southern U.S. regions, while most previous climate change impact studies focused only in one region.

Crop yield response modeling is complex due to the growth process of crops; therefore, it is hard to implement a comprehensive model that considers all the influential factors. The results reported in this study are subject to several limitations. First, we assumed that there is no CO_2 fertilization effect for crop growth. Although numerous previous studies have demonstrated improvements in crop yields with CO_2 fertilization, most of these studies are based on crop simulation models. Some actual field research indicates a much smaller increase in crop yields under a higher CO_2 concentration environment (Long et al., 2006). In order to focus our study on weather and crop yield connections, we decided to exclude increased CO_2 effects not just to simplify the model, but also because of the inability to specify these effects.

Agronomically, the timing of precipitation is relatively more important than the amount of precipitation. Due to the availability of climate change data, we did not include the distribution of precipitation in the model. Although extreme temperature events were considered by including maximum and minimum temperature, extreme precipitation events such as drought or flood were not considered in this study. In future climate change studies where adequate climate change data is available, we recommend considering such extreme precipitation events.

Another important issue is the availability of irrigation water. Adequate irrigation water could mitigate the negative effects of drought and heat. Therefore, adverse weather could have smaller effects on regions with high irrigation water availability. This study is limited in excluding the effects of irrigation.

References

Almaraz, J., F. Mabood, X. Zhou, E. Gregorich, and D. Smith. 2008. "Climate Change, Weather
Variability and Corn Yield at a Higher Latitude Locale: Southwestern Quebec." *Climatic
Change* 88: 187 – 197.

Buller, O. 1972. "Influence of Research and Policy on Crop Yields in Kansas." *Transactions of
the Kansas Academy of Science* 75:20-28.

Choi, J.S., and P.G. Helmberger. 1993. "How Sensitive are Crop Yields to Price Changes and
farm Programs?" *Journal of Agricultural and Applied Economics* 25: 237-244.

Coulson, D.P., L.A. Joyce, D.T. Price, D.W. McKenney, R. Siltanen, P. Papadopol, and K.
Lawrence. Climate Scenarios for the Conterminous United States at the County Spatial
Scale Using SRES Scenarios A1B and A2 and Prism Climatology. Fort Collins, CO: U.S.
Department of Agriculture, Forest Service, Rocky Mountain Research Station, August,
2010.

Deschenes, O., and M. Greenstone. 2007. "The Economic Impacts of Climate Change: Evidence
from Agricultural Output and Random Fluctuations in Weather." *American Economic
Review* 97: 354-385.

Dixon, B. L., S. E. Hollinger, P. Garcia, and V. Tirupattur. 1994. "Estimating Corn Yield
Response Models to Predict Impacts of Climate Change." *Journal of Agricultural and
Resources Economics* 19:58-68.

Donner, S.D., and C.J. Kucharik. 2008. "Corn-Cased Ethanol Production Compromises Goal of
Reducing Nitrogen Export by the Mississippi River." *Proceedings of the National
Academy of Sciences* 105: 4513–4518.

Hadi, A.S., and R.F. Ling. 1998. "Some Cautionary Notes on the Use of Principal Components Regression." *The American Statistician* 52: 15-19.

Hansen, P.M., J.R. Jorgensen, and A. Thomsen. 2002. "Predicting Grain Yield and Protein Content in Winter Wheat and Spring Barley Using Repeated Canopy Reflectance Measurements and Partial Least Squares Regression." *Journal of Agricultural Science* 139:307–318.

Hoogenboom, G. 2000. "Contribution of Agrometeorology to the Simulation of Crop Production and its Applications." *Agricultural and Forest Meteorology* 103:137–157.

Horie, T., M. Yajima, and H. Nakagawa. 1992."Yield forecasting" *Agricultural Systems* 40: 211-236.

IPCC, Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. IPCC, Geneva, Switzerland, 2007, 104 pp.

Jolliffe, I.T. 1982. "A Note on the Use of Principal Components in Regression." *Applied Statistics* 31: 300-303.

Jolliffe, I.T. 2002. *Principal Component Analysis*, 2nd. ed. New York: Springer.

Jones, J.W., G. Hoogenboom, C.H. Porter, K.J. Boote, W.D. Batchelor, L.A. Hunt, P.W. Wilkens, U. Singh, A.J. Gijsman, and J.T. Ritchie. 2003. "The DSSAT Cropping System Model." *European Journal of Agronomy* 18:235-265.

Kafumann, R.K., and S. E. Snell. 1997. "A Biophysical Model of Corn Yield: Integrating Climatic and Social Determinants." *American Journal of Agricultural Economics* 79:178-190.

Kandiannan, K., K.K. Chandaragiri, N. Sankaran, T.N. Balasubramanian, and C. Kailasam. 2002. "Crop-weather model for turmeric yield forecasting for Coimbatore District, Tamil Nadu, India." *Agricultural and Forest Meteorology* 112:133-137.

Kantanantha, N., N. Serban and P. Griffin. 2010. "Yield and Price Forecasting for Stochastic Crop Decision Planning." *Journal of Agricultural, Biological, and Environmental Statistics* 15:362-380.

Lobell, D.B., K.N. Cahill, and C.B. Field. 2007. "Historical Effects of Temperature and Precipitation on California Crop Yields." *Climatic Change* **81**: 187-203.

Lotsch, A. 2007. "Sensitivity of cropping patterns in Africa to Transient Climate Change." Working Paper, World Bank, Washington DC.

Martinez, C.J., G.A. Baigorria, and J.W. Jones. 2009. "Use of Climate Indices to Predict Corn Yields in Southeast USA." *International Journal of Climatology* 29:1680–1691.

McCarl, B.A., X. Villavicencio, and X. Wu. 2008. "Climate Change and Future Analysis: Is Stationary Dying?" *American Journal of Agricultural Economics* 90: 1241-1247.

Mitchell, J.F.B., S. Manabe, V. Meleshko, and T. Tokioka. 1990. "Equilibrium climate change – and its implications for the future." The IPCC Scientific Assessment. Cambridge University Press, New York, NY.

Nakicenovic, N. et al. Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge, UK: Cambridge University Press, 2000. 599 p.

Pandzic, K., and D. Tminic. 1992. "Principal Component Analysis of River Basin Discharge and Precipitation Anomaly Fields Associated with the Global Circulation." *Journal of Hydrology* 132: 343-360.

Porter, J. R., M. A. Semenov. 2005. "Crop Responses to Climatic Variation." *Philosophical Transactions: Biological Sciences* 360:2021–2035.

Preisendorfer, R.W. 1988. *Principal Component Analysis in Meteorology and Oceanography.* London: Elsevier.

Schlenker, W., and M.J. Roberts. 2009. "Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change." *Proceedings of the National Academy of Sciences* 106:15594–15598.

Tannura, M. A., S. H. Irwin, and D. L. Good. 2008. "Weather, Technology, and Corn and Soybean Yields in the U.S. Corn Belt. " Dept. Agr. Econ. Marketing and Outlook Research Report 2008-01,University of Illinois.

Tao, F., M. Yokozawa, Y. Xu, Y. Hayashi, and Z. Zhang. 2006. "Climate Changes and Trends in Phenology and Yields of Field Crops in China, 1981-2000." *Agricultural and Forest Meteorology* 138: 82-92.

Walker, G.K. 1989. "Model for Operational Forecasting of Western Canada Wheat Yield." *Agricultural and Forest Meteorology* 44:339-351.

Yu, Z.P., P.S. Chu, and T. Schroeder. 1997. "Predictive Skills of Seasonal to Annual Rainfall Variations in the U.S. Affiliated Pacific Islands: Canonical Correlation Analysis and Multivariate Principal Component Regression Approaches." *Journal of Climate* 10:2586–2599.
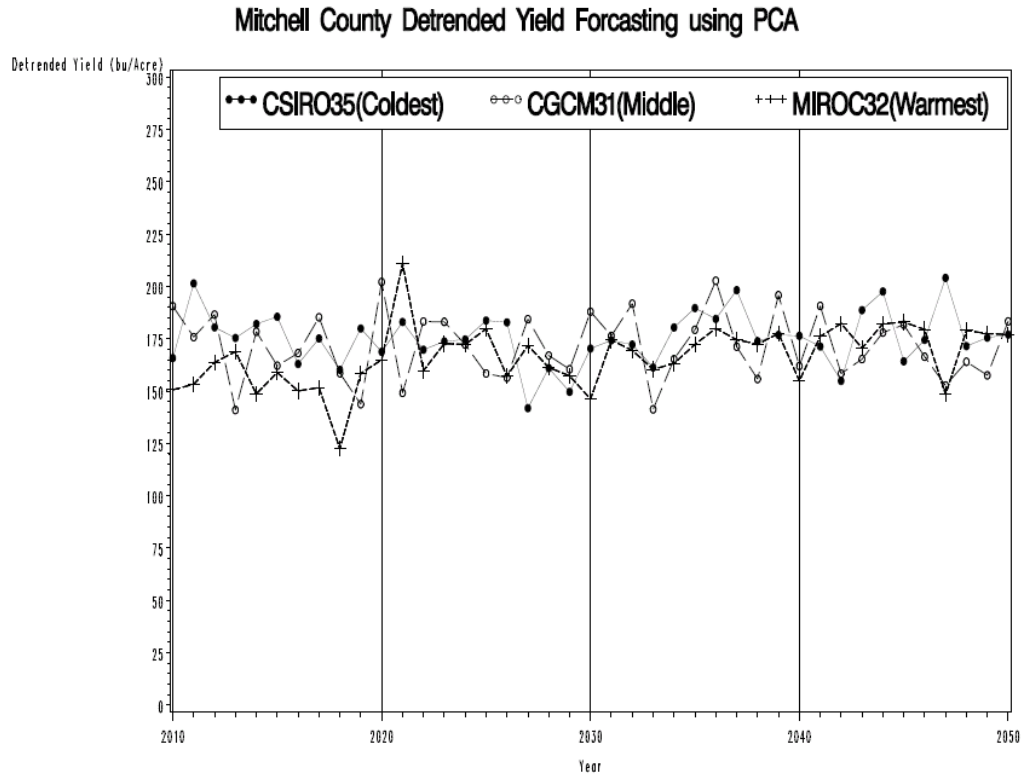
Figure 1. **Forecasted corn yields in Mitchell County, Georgia from 2010 to 2050**

Hancock County Detrended Yield Forcasting using PCA
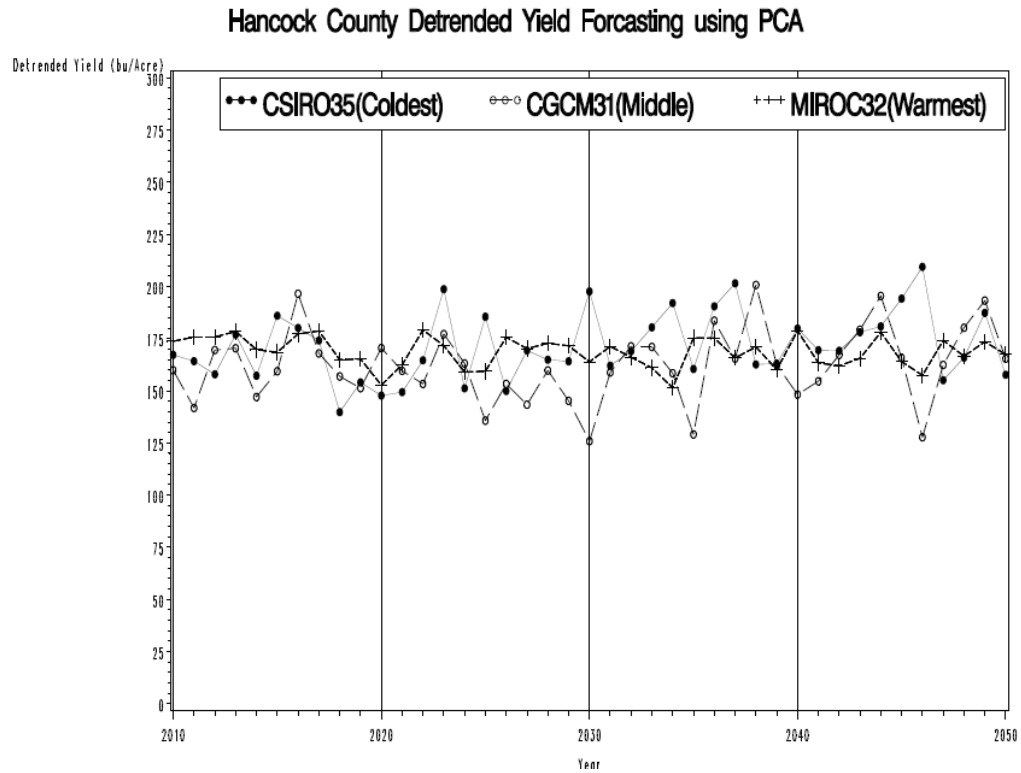
Figure 2.   **Forecasted corn yields in Hancock County, Illinois from 2010 to 2050**

Table 1.   **Climate Change Impact Index (by states, by crops)**

	Corn		Soybeans		Cotton		Peanuts	
	Original	Weighted	Original	Weighted	Original	Weighted	Original	Weighted
MN	0.511	0.528	0.4112	0.421				
IA	0.549	0.561	0.448	0.445				
NE	0.506	0.488	0.528	0.516				
IL	0.531	0.526	0.500	0.505				
IN	0.536	0.532	0.514	0.520				
TX	0.56	0.598	0.634	0.626	0.556	0.542	0.548	0.534
AL	0.691	0.706	0.545	0.590				
GA	0.692	0.687	0.594	0.630	0.634	0.633	0.497	0.489

Appendix

Monte Carlo Experiment:

1. Obtain data set (x_i, y_i), $i= 1, ..., 50$.
2. Bootstrap $\tilde{x}_i$ from data set in step 1, $i= 1, ..., 130$.
3. Generate $\tilde{y}_i = \tilde{x}_i \hat{\beta} + \tilde{e}_i$, where $\hat{\beta}$ is estimated from (x_i, y_i), $\tilde{e}_i$ is random normal error
4. Separate $(\tilde{x}_i, \tilde{y}_i)$ into three groups:

   $(x1_i, y1_i)$, i= 1,…, 50,

   $(x2_i, y2_i)$, i= 1,…, 40,

   $(x3_i, y3_i)$, i= 1,…, 40,
5. Generate Principal Components for $x1_i$ ($PC1_i$) based on PCA method
6. Estimate $\widehat{\beta 1}$ using $(PC1_i, y1_i)$
7. Generate Principal Components for $x2_i$ ($PC2_i$) based on mean, standard deviation, and eigenvectors from PC1
8. Forecast $\widehat{y2}_i = PC2_i \widehat{\beta 1}$
9. Generate Principal Components for $x3_i$ ($PC3_i$) based on mean, standard deviation, and eigenvectors from PC1
10. Forecast $\widehat{y3}_i = PC3_i \widehat{\beta 1}$
11. Generate Principal Components for $x1_i$ and $x2_i$ ($PC12_i$) based on PCA method
12. Estimate $\widehat{\beta 12}$ using $(PC12_i, y1_i)$
13. Forecast $\widehat{y12}_i = PC12_i \widehat{\beta 12}$
14. Generate Principal Components for $x1_i$ and $x3_i$ ($PC13_i$) based on PCA method
15. Estimate $\widehat{\beta 13}$ using $(PC13_i, y1_i)$
16. Forecast $\widehat{y13}_i = PC13_i \widehat{\beta 13}$
17. Compare Mean Squared Error (MSE) in step (8), (10), (12), and (16)

Monte Carlo simulation results
(average MSE based on 1,000 times of experiments)

	Mean Squared Error	
	Approach 1	Approach 2
Y2	378.77	379.81
Y3	384.56	383.29