Comparison of climate change impact on rainfed maize yield in Kansas using statistical and process-based models

by

Meenakshi Rawat

M.S., Central University of Punjab, 2017

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Carl and Melinda Helwig Department of Biological and Agricultural Engineering
Carl R. Ice College of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2023

Approved by:

Major Professor
Vaishali Sharda

Copyright

Abstract

Changing climate and the projected increase in variability and frequency of extreme events make an accurate prediction of crop yield critically important for addressing emerging challenges in food security. Precise and timely prediction of crop yield can provide valuable information to agronomists, producers, and decision-makers. Even without considering climate change, several factors including environment, management, and genetics and their complex interactions make the prediction of crop yield challenging. In this study a statistical-based Multiple Linear Regression (MLR) model was developed to predict rainfed maize yield in Kansas and compared with yield predictions of the DSSAT process-based model to assess the impact of synthetic climate change scenarios of 1 and 2 °C temperature rise. Historic weather, soils, and crop management data were collected and converted to model-compatible formats to simulate and compare maize yield using both models. It was found that DSSAT had a large Root Mean Square Error (RMSE) compared to the MLR model whereas the correlation coefficients (r) were 0.93 and 0.70 for MLR and DSSAT, respectively. These results indicated that predicted yields from the MLR model had a stronger association with the observed yields than the simulated yields from DSSAT. Analysis of climate change impact showed that the reduction in rainfed maize yield predicted by DSSAT was 8.7% and 18.3% for the synthetic scenarios of 1 and 2 °C temperature rise respectively. Reduction in rainfed maize yield predicted by the MLR model was nearly 6% in both scenarios. Due to the extreme heat effect, predicted impacts under uniform climate change scenarios were considerably more severe for the process-based model than for the statistical-based model.

Keywords: Climate impacts, DSSAT, Model inter-comparison, Multiple Linear Regression (MLR) model

Table of Contents

List of Figures

List of Equations

Acknowledgments

I would like to express my deepest gratitude to my major professor, Dr. Vaishali Sharda, for her expertise, dedication, and unwavering commitment to excellence have shaped this thesis in profound ways. During the past two years, Dr. Sharda provided me with knowledge and expertise in hydrology, DSSAT modeling, climate change, oral and poster presentation, and paper writing, and also offered me the chance to present my research on different platforms, which added precious experiences that are valuable for my career. Dr. Sharda's guidance and mentorship have not only broadened my knowledge but I am also truly grateful for the countless hours she spent on reviewing my work, providing constructive feedback, and pushing me to strive for the highest standards.

I would like to extend my sincere appreciation to my committee members, Dr. Xiaomao Lin and Dr. Kraig Roozeboom, for their invaluable contributions and support throughout my master's program. Dr. Lin's expert guidance on statistical model development for corn yield prediction has been instrumental in advancing my research. Dr. Roozeboom's valuable suggestions on management practices for maize at county scale were truly helpful. His extensive experience and insights have provided me with a broader perspective on agricultural practices and their impact on crop yield. I am grateful for the rigorous discussions and the critical insights my committee members provided, which have helped shape my research in meaningful ways.

I am grateful to the faculty members and staff of Carl and Melinda Helwig, Department of Biological and Agricultural Engineering, Kansas State University. I would like to thank Rintu, who started the journey with me and helped me in my ups and downs from starting and I am forever grateful for your friendship. A special thanks to my lab mates at K-State, Kelechi, Ikenna, Heather, and Manavjot for their company, invaluable support, and encouraging words.

At last, I want to thank my parents and my sister for giving me unconditional love, trust, continuous support, and encouragement throughout my life. I am proud of having the best parents in the world. I truly appreciate my friend, Lokendra, your unwavering belief in me and your constant encouragement have provided me with the strength to overcome challenges and stay focused. I am grateful for the countless discussions, and brainstorming sessions, which have enriched my research journey in immeasurable ways.

Chapter 1 - Introduction

Globally, average temperature and precipitation have increased at the average rate of 0.18 °C (NOAA, 2023) since 1981 and 1.016 mm (NOAA, 2022) per decade since 1901, respectively. For the United States, this average rate of increase in temperature and precipitation has been 0.27 °C (Menne et al., 2018) since 1981 and 5.08 mm (NOAA, 2022) per decade since 1901 respectively. The state of Kansas forms a part of the central Great Plains and lies in the mid-western part of the United States and faces an east to west precipitation gradient with plentiful rain in the east and insufficient precipitation in the west (Frankson et al., 2022). During 1895-2015, the average temperature in Kansas has increased by 0.059 ± 0.03 °C per decade with annual mean precipitation for the western third, the central third, and the eastern third of Kansas being 531 mm, 660 mm, and 945 mm, respectively (Lin et al., 2017). These uneven changes in temperature and precipitation have been shown to impact crop productivity (Maitah et al., 2021).

Wheat, maize, and rice are the world's leading crops (Kennett et al., 2020), with maize being the second most-grown crop in the world after wheat (FAO-ESS, 2021). Maize is the major crop grown in the USA and globally, and the USA contributes 50% of global maize production (FAO, 2021; Erenstein et al. 2022). It is a global staple food, a major source of the human diet as well as livestock feed, and has varied industrial and energy uses (Erenstein, 2010). Maize production in Kansas ranks 6th in the nation and is the 2nd major crop grown in Kansas after winter wheat, with the Kansas maize industry playing a significant role in the US agricultural economy (USDA-NASS, 2022).

Changing climate adversely affects crop production (Arora, 2019), which is a challenge in the face of a growing global population. The changing climate results in frequent warming events, and irregular precipitation (Malhi et al., 2021), which in turn impact crop production and induce biotic (Shahzad et al., 2021) and abiotic stresses (Paraschivu et al., 2019). The impact of climate on crop yield varies by crop type and geographical location (Challinor et al. 2014; Su et al., 2021), and precise and timely prediction of crop yield can provide valuable information to agronomists, producers, and decision-makers (Cooper et al. 2020; Dai and Li 2013; Filippi et al., 2017; Hipólito et al., 2018). It has been shown that increasing temperatures reduce maize yields in the US Midwest with higher temperatures being associated with short grain filling periods that faster reproductive development (Abendroth et al., 2021). Timely precipitation could mitigate rising temperature variations though significant yield losses have been reported due to fluctuations in precipitation (Malhi et al., 2021). It has been found that a 1 °C rise in maximum daily temperature (T_{max}) reduced maize and rice yields in the southeastern USA by 34% and 8.3% respectively (Sharma et al., 2022), and results from a statistical model have indicated an 8.3% maize yield reduction globally with a 1 °C rise in temperature (Lobell & Field, 2007). In addition, 43% to 44% reduction of maize yields have been found to occur under a slow-warming scenario (B1 scenario that assumes resource efficient technology), and 79% to 74% yields reduction found in fast-warming scenario (A1FI that assumes continued use of fossil fuels, which results in the largest increase in CO_2-concentrations and temperatures (IPCC, 2000)) in the US (Schlenker & Roberts, 2009). Changing climate and the projected increase in variability and frequency of extreme events make an accurate prediction of crop yield critically important for addressing emerging challenges in food security (Ansarifar et al., 2021).

To assess the climate change impact on crop yield, two distinct approaches have been used worldwide. The first approach uses process-based models that require detailed input data for soils, weather, and crop management practices to simulate in-season crop growth and end-of-season crop yield. The second approach links crop yield data to weather variables to make predictions using statistical methods. These two approaches have their strengths and weaknesses (Roberts et al., 2017).

Process-based crop models are non-linear models, e.g., APSIM (Keating et al., 2003), DSSAT (Jones et al., 2003), and RZWQM (Ma et el., 2012), that can simulate growth, development, and yield of a crop. These models operate at a daily time-step and require inputs of weather data including minimum temperature, maximum temperature, solar radiation, precipitation, wind speed, and humidity; soil properties which include physical, chemical and engineering properties; and crop management information contain planting date, plant population, and cultivar. Process-based crop models use mechanisms that are based on established physical processes that link environmental conditions like weather and soils to crop growth and production, which is a key strength of these models (Roberts et al., 2017). They are typically developed for field-scale decision making (Brisson et al., 2003; Jones et al., 2003; Keating et al., 2003; van Ittersum et al., 2003), but currently are used to evaluate climate change scenarios (Southworth et al., 2000a; Challinor et al., 2014; Lobell & Asseng, 2017; Roberts et al., 2017). These models were originally not designed to simulate future conditions and may be missing some key processes related to extreme climate conditions (van Oort et al., 2011; White et al., 2011). Overall, process-based model are data intensive and complex, and their prediction accuracy may not be high (Akhavizadegan et al., 2021; Bassu et al., 2014).

Statistical models are used to determine the relationship between predictors and response variables and have strengths and weaknesses that are different than process-based models. These models have become increasingly common for crop yield prediction because of the better availability of weather and crop datasets. Crop data can be collected from field experiments, farmer surveys, official government statistics, and other resources (Lobell & Asseng, 2017). The most common concern in the use of these models, however, is that it is difficult to distinguish between the effects of highly correlated weather variables, such as temperature and precipitation (Lobell & Ortiz-Monasterio, 2007; Sheehy et al., 2006).

Though several studies have considered the impacts of future climate change on crop production (Southworth et al., 2000; Edmonds & Rosenberg, 2005; Fischer et al., 2005; Parry et al., 2005; Lobell et al., 2006; Schlenker & Roberts, 2009; Challinor et al., 2014; Arora, 2019; Malhi et al., 2021; Sharma et al., 2022), and have studied comparison and combination of process-based, regression, and machine-learning models (Irmak et al., 2005; Roberts et al., 2017; Leng & Hall, 2020), these models have not been compared for their ability to predict yield and impact of climate change at the county scale. Keeping this in mind, this study was undertaken to develop a new statistically-based Multiple Linear Regression (MLR) model for predicting rainfed maize yield in Kansas; to compare the predicted yield from the MLR model with that of a process-based model (DSSAT); and to assess the impact of climate change on maize yield for synthetic climate change scenarios of 1 and 2 °C temperature rise.

Chapter 2 - Review of Literature

2.1. Problem Statement

Research in the area of crop yield prediction has grown richer worldwide, especially with the growing global population and climate change. The United Nations (UN) has projected that the population will reach around 9.8 billion by 2050. To fulfill the demand for food, feed, fiber, and bioenergy of the growing population, crop production needs to be increased. Simultaneously, climate change results in increasing variability and extremes, and this may have significant negative impacts on crop production. Studies have shown that rainfed crops are more affected by climate change than irrigated crops. Therefore, changing climate and the projected increase in variability and frequency of extreme events make an accurate prediction of crop yield critically important for addressing challenges in food security. Several factors, including environment, management, and genetics and their complex interactions, make the prediction of crop yield challenging. Precise and timely prediction of crop yield can provide valuable information to agronomists, producers, and decision-makers. In this section, worldwide literature pertaining to crop yield prediction by using models assessing climate change's impact on it is reviewed and studied.

2.2. Literature

A recent research study (Kamath et al., 2021) was conducted to forecast the yield of major crops in India. They used the classification algorithm Random Forest (Hastie et al., 2009) on historical data from the years 1997 to 2012 and compared the accuracy of the Random Forest model with 3 other classifiers namely K-Nearest Neighbors (KNN) (Mucherino et al., 2009), XGBoost (Chen & Guestrin, 2016), and Logistic regression (Sammut & Webb, 2010). The authors found that the Random Forest classifier produced highly accurate results, a 98%

accuracy, with a relatively lower error as compared to other three models tested because it creates decision trees for different sets of the training data and then takes their average to create a single accurate decision tree. Another study of use of machine learning-based framework to forecast maize yields in three US Corn Belt states (Illinois, Indiana, and Iowa) (Shahhosseini et al., 2020) under different environmental conditions and management at agricultural district and state-level scales found that optimized weighted ensemble model is the most precise model with RMSE of 1138 kg/ha. The authors concluded that the proposed ensemble model performed better than the nine individual models used in the study. Along with machine-learning algorithms, statistical models have been widely used worldwide for estimating direct relationship between maize yields and climate.

Lobell & Field, (2007) developed statistical models to assess the impact of climate trends on global yield (wheat, rice, maize, soybean, barley, sorghum). They defined growing season months for each crop and time-series data for yield and climate (temperature and precipitation) variables were analyzed to evaluate the relationship between yield and climate variables. They found that climate predictors explained 47% year-to-year maize yield variability and caused 8.3% yield reduction with 1 °C rise in temperature globally. Similarly, another study (Tao et al., 2008) focused on 22 provinces in China and aimed to develop MLR models for estimating crop yields of rice, maize, soybean, and wheat. The study revealed that maize yield reduced from 2.4% to 20.7% for every 1 °C increase in the maximum temperature (T_X) during the growing season and a reduction in maize yield ranging from 1.6% to 27.9% was observed for every 1 °C increase in the minimum temperature (T_N) during the growing season. These findings emphasize the sensitivity of maize yield to temperature changes and highlight the potential impact of rising temperatures on crop production. Another study (Hatfield et al., 2011) found that simulated maize yields in the central Corn Belt in the U.S. reduced 5 to 8% per 2 °C temperature increase, which lead to the prediction that a temperature rise of 0.8 °C could decrease maize yield by 2-3%. Likewise, a study using three different regression models using weather variables (growing degree day, extreme degree day, vapor pressure deficit, precipitation, and their interactions) for maize yield prediction in Illinois (Roberts et al., 2013a) found maize yields reduction in the range of 10 to 60% for the years 2040-2069 relative to 1981-2010 baseline scenario from one of the best performing models. Likewise, a study by Feng et al., (2021) also proposed a statistical approach to estimate the risk of crop yield reduction under dry, hot, and compound dry-hot condition (precipitation deficit and high-temperature extremes) due to concurrent changes in climate-yield relationships. Standardized Precipitation Index (SPI) (McKee et al., 1993) was used to characterize dry conditions and the Standardized Temperature Index (STI) (Feng et al., 2019) for hot conditions. They found that risk of maize yield reduction increased in the top ten maize-producing countries of the world under compound dry-hot conditions. Similarly, a new predictive model was developed for the Corn Belt Region by Ansarifar et al., (2021), and RMSE obtained for linear regression for corn in the years 2015, 2016, and 2017 were 1390, 1330, and 1190 kg ha^{-1} whereas, for the interaction regression model, RMSE were 1020, 810, and 900 kg ha^{-1} respectively showing that the interaction model outperformed the linear model with respect to prediction accuracy.

For model comparison, a study (Liu et al., 2022) in the Corn Belt compared multiple machine learning (ML) models (RF, XGBoost, and Long Short Term Memory with attention ($LSTM_{att}$)) and a process-based crop model (DSSAT) for predicting rainfed maize yield using meteorological variables and soil properties. They found that the RMSE was about three times in DSSAT compared to $LSTM_{att}$, and $LSTM_{att}$ explained 73% of the spatiotemporal variance of the observed maize yield, while DSSAT explained 16%. They concluded that the $LSTM_{att}$ outperforms other ML models, as well as DSSAT and suggested that $LSTM_{att}$ has great potential to predict yield in changing climate. Another study (Joshi et al., 2021) showed that weather conditions regulate growth and yield of crops, especially in rainfed cropping systems. They developed yield estimation models (step-wise Multiple Linear Regression (MLR), General Additive (GAM), and Support Vector Machine (SVM)) for maize and soybean in the US Corn Belt. They found that SVM outperformed all other models, and RMSE for maize and soybean were 591 and 205 kg ha^{-1} respectively. Similarly, Roberts et al., (2017) compared and combined the process-based and statistical models for maize yield prediction in the Corn Belt region. They found that process-based models predicted actual outcomes slightly better than the statistical model, but the combined model performed significantly better than either model. To assess the climate change impact, they found that predicted impacts are considerably more severe under the statistical model, combined model, and full model as compared to the process model. Another study (Southworth et al., 2000) showed the effect of climate variability on maize yields for three hybrid maturities (long, medium, and short) in the 10 representative agricultural regions in the midwestern Great Lakes states of the US. They used DSSAT CERES-Maize model to estimate the impact of climate variability on maize yield and found that increased variability of the future climate scenarios increased the variability of the year-to-year crop yields. They found that the yields of long-season maize decreased in the southern and central locations for all scenarios ranging from 0 – 45% loss. Medium-season maize yields reduced in the range of 15.1 to 40%, whereas short-season maize showed increase in yields ranging from 5.1 to 35%. These literature sources have utilized various approaches, including machine learning, process-based, and statistical modeling, for predicting crop yields. However, among these approaches, statistical models have emerged as a reliable method for evaluating the relationship between climate and yield on a global scale.

2.3. Major Takeaways

Conducting this review of literature, we have noticed that there is a wide range of models that have been utilized for predicting maize yield globally. Among these models, statistical models have been extensively developed to examine the relationship between yield (as the response variable) and climate (as the predictor variables) due to availability of climate as well as crop yield data, and a relatively lower number of model inputs. Some other studies have concentrated on comparing different models. However, there is limited literature available that specifically focuses on the DSSAT model, which is a process-based model, and its comparison with statistical models.

Given this research gap, our study aims to address three key objectives. Firstly, we aim to develop a new statistical-based Multiple Linear Regression (MLR) model for climate-yield relationship at Kansas county-scale. Secondly, to compare the performance of the MLR with the DSSAT model, which is widely used in agricultural research and to evaluate their strengths, weaknesses, and suitability in the context of climate change scenarios. Lastly, compare the predicted maize yield from both models to assess the impact of climate change on maize yield for synthetic climate change scenarios of 1 and 2 °C temperature rise .

Chapter 3 - Methodology

3.1. Study Region

The study region (Figure 3.1) selected is the state of Kansas located in climatic zones B and C which comprise of dry and moist subtropical mid-latitude climates (Beck et al., 2020). Kansas is a part of the midwestern United States and extends from 36°59’36.11” N to 40°00’10.07” N, and from 94°35’18.20” W to 102°03’04.43” W. Selection of Kansas counties as shown in Figure 3.1 were based on data availability and maize grown under rainfed condition which is discussed in further detail in the sections that follow.

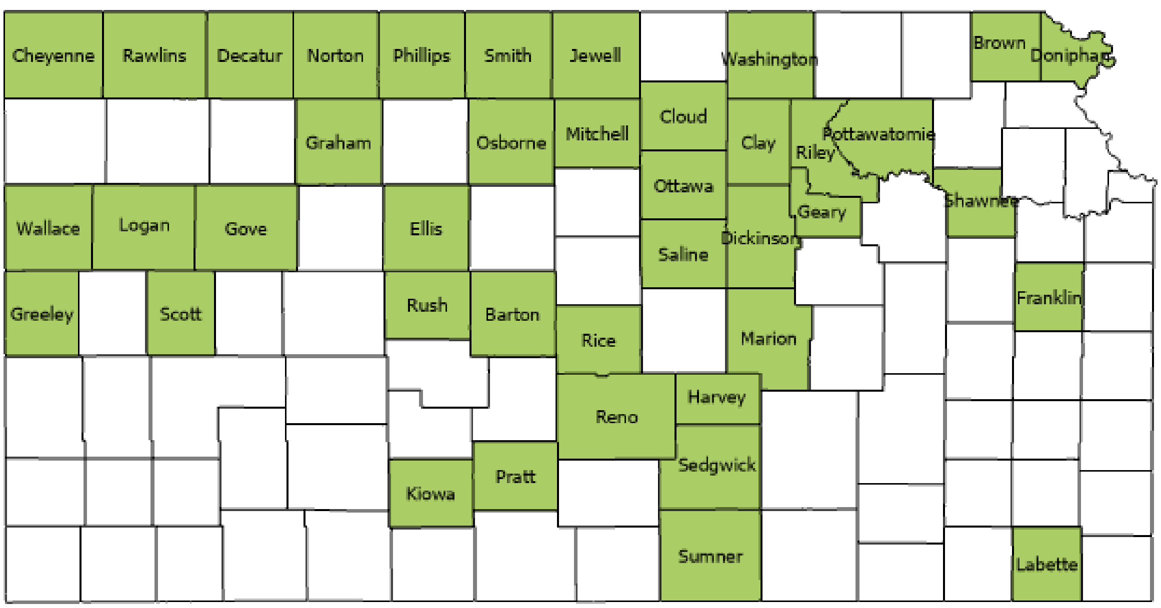


Figure 3.1. Map of study area

For this study, we used maize because it is one of the major crops grown in USA, and Kansas is one of the largest producers of maize in the country. Currently, Kansas ranks the 6th in maize production in the nation (USDA-NASS, 2022). Maize is the 2nd most grown crop in Kansas after winter wheat (FAO-ESS, 2021) with 66.4% of maize growing areas being rainfed and 33.6% under irrigated conditions in the last 10 years (USDA-NASS, 2021). Kansas has varying temperature and precipitation county-to-county with an east to west decreasing precipitation with average annual precipitation ranging from 1172 mm to 435 mm and a north to south increasing average annual temperature trend ranging between 10.7 °C to 14.3 °C (40 year average, Figures. 3.2 (a) and (b) (Rupp et al., 2022)).

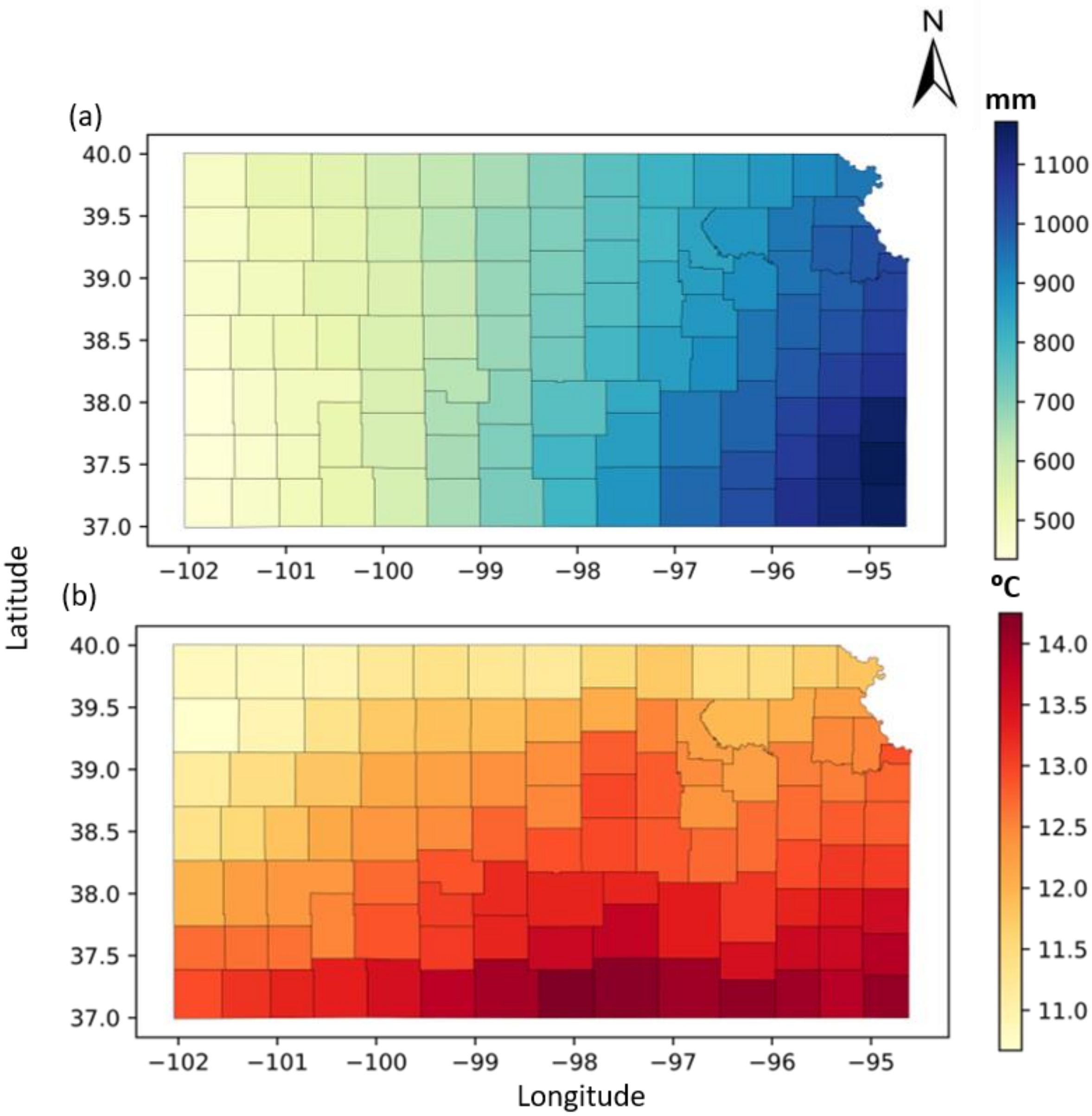


Figure 3.2. (a) Annual precipitation; (b) Annual mean temperature across Kansas from 1981 to 2020

3.2. Data and Data Sources

3.2.1. Climate Data

Daily historic temperature and precipitation data used for the study were obtained from the Parameter-elevation Relationships on Independent Slopes Model (PRISM) (Rupp et al., 2022) and include daily maximum temperature (T_X), daily minimum temperature (T_N), and precipitation (PRECIP) at a 4 km grid. Because it is standard practice to use at least 30 years of historic data to capture climate variations (Feenstra et al., 1998), we used data from 1981 to 2020 in our study. The gridded raster data for these climate variables were processed and aggregated at county level for each growing season. Growing season for maize was selected based on information available in literature based on Corn Belt region of the USA, and from Kansas Crop Planting Guide and K-state maize crop performance tests (Shroyer et al., 1996; Corn Crop Performance Tests, 2022). For this study, we carefully considered the selection of growing seasons for Kansas in our Multiple Linear Regression (MLR) model. To ensure comparability of results, we selected April to September, as the growing season for the MLR model, as this period was consistent across all counties in Kansas. This approach enabled us to examine the relationship between climate factors and yield outcomes across different counties both temporally and spatialy. In contrast, the growing seasons varied from county to county in the DSSAT model. This variation was due to the significant influence of factors such as planting date and crop management information, in addition to soil and weather conditions, on the simulation of crop growth and development. By accounting for these crucial management variables, the DSSAT model simulations could provide a comprehensive understanding of how different agricultural practices impact crop performance. By considering the specific growing seasons and accounting for the diverse range of factors involved, our study aimed to capture the complexities and nuances of the relationship between climate and rainfed maize yield. This approach allowed us to draw meaningful insights and make informed comparisons between the MLR and DSSAT models, advancing our understanding of the factors influencing crop productivity in Kansas.

Solar radiation data, which is a required weather input for the process-based crop model (in addition to T_X, T_N, and PRECIP), was obtained from NASA POWER (POWER, Data Access Viewer, 2022), which is available at 1° resolution and was interpolated to county scale.

3.2.2. Rainfed Maize Yields Data

The historical rainfed maize yield data for Kansas counties was obtained from the USDA-NASS Quick Stats database (USDA-NASS, 2017) and K-State Research and Extension (KSRE) reports (Corn Crop Performance Tests, 2022). For this study, we used annual county-level maize grain yields from 1981 to 2020. USDA-NASS database has data for irrigated, rainfed, and average crop yields for all crops at county, state, and national level. We selected counties that had maize grown under rainfed conditions and had historic yield data available for at least 20 years for the period of study (1981-2020). This resulted in selection of 40 counties out of 105 counties in Kansas. Maize yield data was reported in bu ac^{-1} which was converted to kg ha^{-1} by multiplying with a conversion factor of 62.8 (Li et al., 2019).

3.3. Multiple Linear Regression (MLR) Model

MLR models were developed to evaluate the relationship between time series for historic rainfed maize yield and climate variables. For development of MLR, we assumed that residuals are normally distributed and there was no autocorrelation between predictors (James et al., 2021). The predictor variables used for MLR model were the growing degree days (GDD), extreme degree days (EDD), total precipitation for the month of May (PRECIP-May), total precipitation for the month of June (PRECIP-June), total precipitation for the month of July (PRECIP-July), total precipitation for the month of August (PRECIP-August), average temperature for the month of May (TEMP-May), average temperature for the month of June (TEMP-June), average temperature for the month of July (TEMP-July) parameters, and their interactions. Figure 3.3 represents the procedure used for developing the MLR.

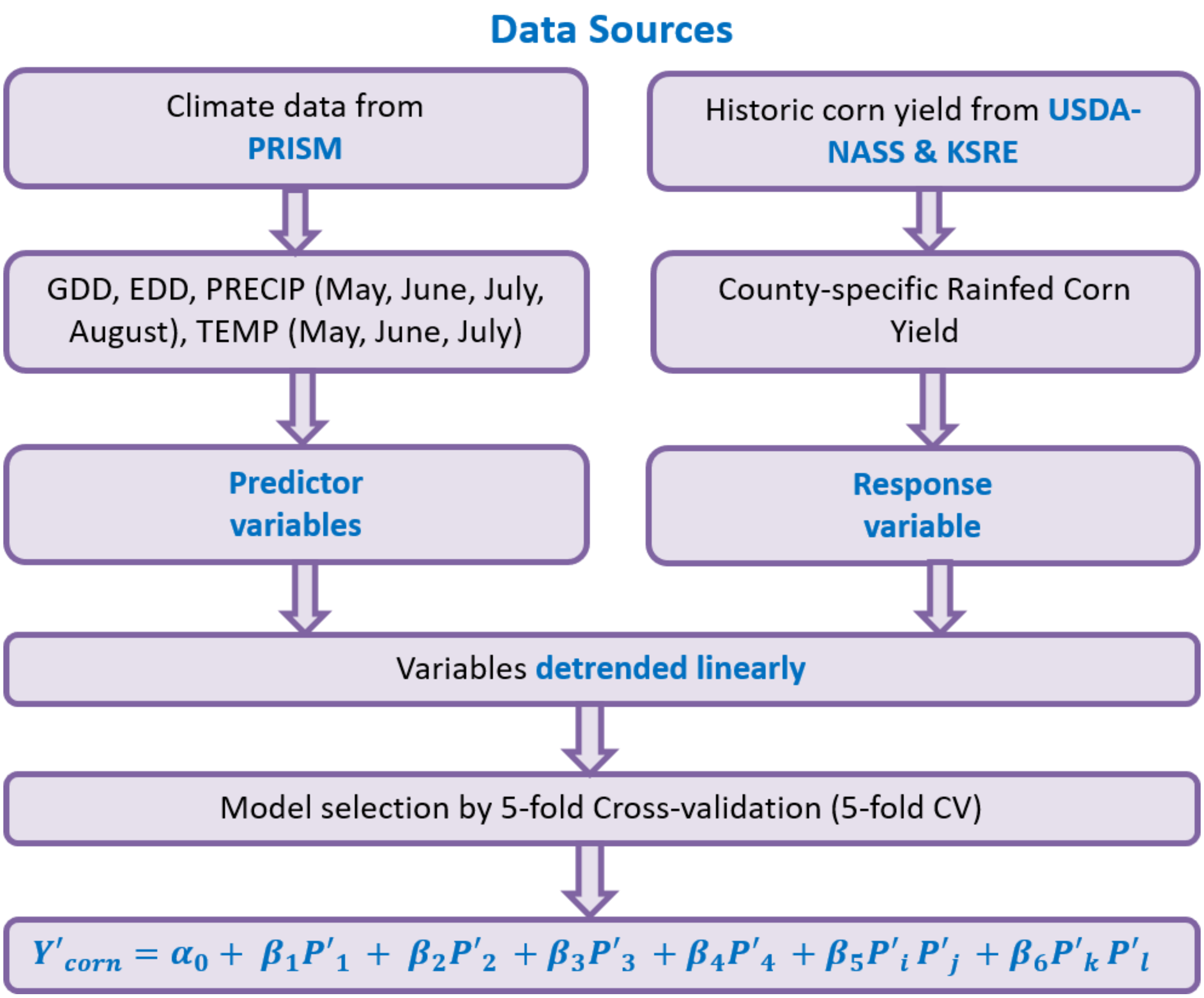


Figure 3.3. Flowchart showing the procedure for development of county-specific Multiple Linear Regression (MLR) model

GDD are used to estimate the thermal time required for a specific cultivar/crop to develop or accumulated exposure to heat over the growing season.

$$\boldsymbol{GDD_d = \frac{T_{X,d} + T_{N,d}}{2} - T_{baseline}}$$

3.1. Equation for calculation of daily GDD for maize

$$\boldsymbol{T_{X,d} = \begin{cases} T_{X,d} & if\ T_{low} < T_{X,d} < T_{high}, \\ T_{baseline} & if\ T_{X,d} \le T_{baseline}, \\ T_{high} & if\ T_{X,d} \ge T_{high} \end{cases}}$$

3.2. Maximum, minimum and baseline temperature ranges for maize

where GDD_d (Equation 3.1) is daily heat unit defined on each day, d, T_X maximum temperature, T_N minimum temperature, and $T_{baseline}$ baseline temperature (typically set to 9 °C), using the representation (Equation 3.2) (Abendroth et al., 2010; Butler & Huybers, 2015).

EDD accounts for growing degree days above the threshold, with a high of (T_{high}) 29 °C for maize. It generally has a strong negative relation with maize yields (Roberts et al., 2013a), and is defined as:

$$\boldsymbol{EDD_d = \begin{cases} T_{X,d} - T_{high} & if\ T_{X,d} > T_{high}, \\ 0 & if\ T_{X,d} \le T_{high}. \end{cases}}$$

3.3. Equation for calculation of daily EDD for maize

where EDD_d (Equation 3.3) is daily heat unit above threshold T_{high} (29 °C) defined on each day, d.

After getting the daily GDD and EDD from the equations 3.1 and 3.3 respectively, we calculated the accumulated GDD and EDD for the growing season (April to September) of maize. From the time series data obtained from USDA-NASS, we observed that generally maize yields increase with time due to technological advancement, improvement in seed variety, and changes in management practices etc. (Troy et al., 2015).

To isolate the variability of rainfed maize yield due only to climate variables, the time-series data was detrended linearly to remove the trend of technological advancement on yields and interannual variability within climate variables. Linear detrending is the most common detrending process and consists of removing a straight line trend component from a time series (Wu et al., 2007). Detrending also minimizes the influence of slowly changing factors such as soil and crop management (Lobell & Field, 2007). Therefore, all climate variables (predictors) and rainfed maize yields (response variable) were detrended linearly. Then, county-specific MLR models were developed to evaluate the relationship between detrended rainfed maize yield and detrended climate variables (9 variables as described above) from 1981-2020. We selected 6 predictor variables from the original 9 predictors that were tested for MLR model development by using 5-fold cross-validation (5-fold CV) (Roberts et al., 2017). 5-fold CV divides the data into five roughly equal parts, then trains the model on four parts and tests the model on the fifth. In addition, it gives the model's performance of the test part each time. This procedure was replicated for all five parts of the data, then $R^2$and RMSE were obtained for all tested parts, then averaged at county-scale. Six variables out of the 9 variables resulting in high R^2 and low RMSE for each county were chosen (4 independent and 2 interactions) to develop a model in the form given by equation 3.4:

$$\boldsymbol{Y'_{maize} = \alpha_0 + \beta_1 P'_1 + \beta_2 P'_2 + \beta_3 P'_3 + \beta_4 P'_4 + \beta_5 P'_i P'_j + \beta_6 P'_k P'_l}$$

3.4. MLR model equation

where Y'_{maize} is detrended predicted maize yield, α_0 intercept, $\beta_1 - \beta_6$ estimated coefficients of 4 independent variables and 2 interaction variables, $P'_1 - P'_4$ detrended independent predictor variables, $P'_i P'_j - P'_k P'_l$ detrended interaction predictor variables.

To obtain 40 county-specific MLR models, the model was run for selection of variables by 5-fold CV. The performance of these models' were assessed by coefficient of determination

(R^2) and root mean square error (RMSE) (as described by equations 3.7 and 3.5, respectively in section 3.5 below) were then plotted spatially across Kansas and are discussed in chapter 4. From the county-specific MLR model, we obtained different independent variables for each county. Box plots were used to visualize the range and variability of county-level independent variable coefficients.

3.4. Process-Based Model, DSSAT

The process-based crop simulation model Decision Support System for Agrotechnology Transfer (DSSAT) version 4.8 (Jones et al., 2003) was used to simulate maize yield for years 1981-2020 for this study. DSSAT version 4.8 comprises models for 42 crops, and the model can simulate crop growth, development, and yield along with several management strategies (Sharda et al., 2017). DSSAT's Crop Estimation through Resources and Environmental Synthesis (CERES) Maize (Ritchie, 1998) calculates maize growth along with daily nutrient and water balance by utilizing physical principles of soil water, nutrient, crop growth, and end of the season crop yield and yield components. The CERES Maize model simulates six different phenological stages for a maize plant with each stage controlled by genetic traits of the cultivar and their interaction with the environment (Sharda et al., 2021). For this study, we used maize cultivar PIO 3489, which has been calibrated in north-east Kansas. DSSAT requires daily weather data, soil information at a different depths of the soil profile, detailed crop management information, and cultivar as input parameters to simulate crop growth and development at a specific location (Hoogenboom et al., 2021).

Daily climate variables required to run DSSAT were obtained from PRISM (Rupp et al., 2022) and NASA POWER (POWER, Data Access Viewer, 2022) and converted to DSSAT format. The soil data was retrieved from the Natural resource Conservation Service (NRCS)

Gridded Soil Survey Geographic (gSSURGO) database (Gridded Soil Survey Geographic (GSSURGO) Database, 2021) available at 30-meter resolution. These data were aggregated to a 4 km grid scale and ArcGIS 10.8 Desktop (ESRI, ArcGIS Desktop 10.8, 2021) toolbox called Soil Data Development Toolbox was used to extract the selected soil properties for Kansas counties. DSSAT requires several physical, chemical and engineering properties of soil, which resulted in extraction of fourteen soil properties from the gSSURGO database. These soil properties are available at six different soil depth, providing valuable information about the soil characteristics. These properties include dry bulk density at 1/3 bar pressure, clay, silt, and sand percentages, soil pH, soil saturated hydraulic conductivity, soil organic matter, available water content at 1/3rd bar and 15 bar pressure, and cation exchange capacity. It is important to note that some properties, such as slope percentage, farmland classification, runoff and drainage class are not depth-dependent and are applicable across the entire soil profile. Crop management data required for running DSSAT were also obtained from USDA-NASS (USDA-NASS, 2021) and KSRE (Corn Crop Performance Tests, 2022). This included planting dates and plant population for each county based on discussions with K-state agronomists and the Kansas maize planting guide. Figure 3.4 shows the input requirement of the DSSAT for simulation of yield as discussed above.

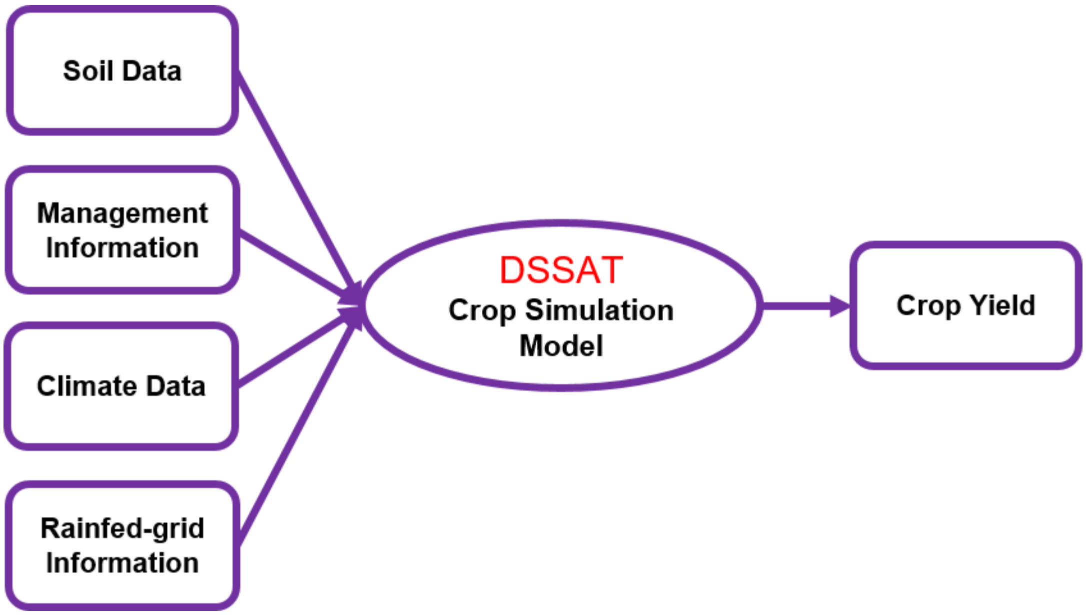


Figure 3.4. Flowchart represents the inputs required by Decision Support System for Agrotechnology Transfer (DSSAT).

After obtaining climate, rainfed maize yields, crop management, and soil properties datasets, area under rainfed maize was extracted using information from Global Map of Irrigation Areas (GMIA) (GMIA, 2005) and CropScape (*CropScape - NASS CDL Program*, 2022). The data from CropScape were available at a 30 m resolution whereas GMIA data for agriculture Area Equipped for Irrigation (AEI) was at a 10 km resolution. GMIA data was downscaled to 4 km with rainfed grid defined as having AEI less than 5%. The data was processed in ArcGIS Pro 3 (ArcGIS Pro 3, 2022) to get rainfed maize yield gridded area for Kansas and was then merged with pre-proceesed gSSURGO soil data. The gridded soil data was converted into DSSAT compatibale format, and rainfed maize yields were simulated at a 4 km grid using DSSAT model. The model was run for 1114 rainfed grid points, with 702 dominant soil across 40 counties as shown in the Figure 3.5. Then the gridded simulated yields were aggregated at county-scale, and averaged over dominant soil types for 40 years. The DSSAT simulated yields were also detrended linearly as discussed above in MLR for comparison of predicted yields from the models.

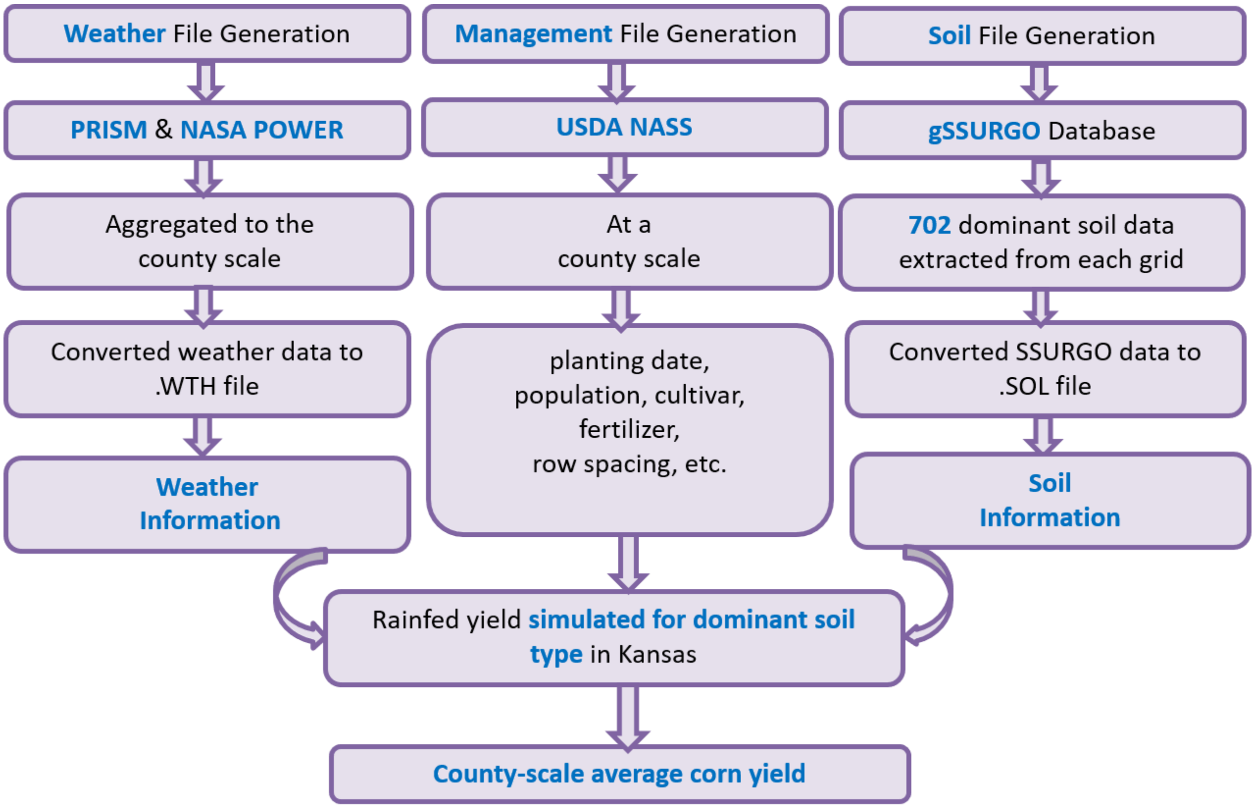


Figure 3.5. Flowchart showing the weather, soil data and management information required for simulating county-specific yields from DSSAT model.

3.5. Model Evaluation

We developed time series plots for 40 counties in Kansas, as depicted in Figure 4.4 and Figure 4.5. Figure 4.4 represents the time series plot for each county, categorized into agricultural district regions. To gain a comprehensive understanding, we calculated the average predicted and observed yields for the entire Kansas state and time series were developed in Figure 4.5. To evaluate the performance of the MLR and DSSAT models, predicted yield value (y_{pred}) from both models were compared with the observed value (y_{obs}) using statistical metrics (Plevris et al., 2022). These include Root Mean Square Error (RMSE), Pearson correlation coefficient (r), and coefficient of determination (R^2) given by equations 3.5, 3.6, and 3.7 respectively.

RMSE measures the difference in magnitude between predicted and observed values and is given by:

$$\textbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{obs,i} - y_{pred,i}\right)^2}$$

3.5. Equation for calculation of Root Mean Square Error **(**RMSE)

where $y_{obs,i}$ and $y_{pred,i}$ are the observed and predicted yield for the ith data record, and $\bar{y}_{obs}$ and $\bar{y}_{pred}$ are their corresponding averages.

The Pearson correlation coefficient, r measures the degree of linear association between predicted and observed values and ranges between -1 to +1.

$$r = \frac{\sum_{i=0}^{n}\left(y_{obs,i} - \bar{y}_{obs}\right)\left(y_{pred,i} - \bar{y}_{pred}\right)}{\sqrt{\sum_{i=0}^{n}\left(y_{obs,i} - \bar{y}_{obs}\right)^2}\sqrt{\sum_{i=0}^{n}\left(y_{pred,i} - \bar{y}_{pred}\right)^2}}$$

3.6. Equation for calculation of Pearson correlation coefficient (r)

The coefficient of determination R^2 measures how much variance in the observed yield can be explained by predictors in a linear regression fit and is given by:

$$R^2 = 1 - \frac{\sum_i [y_{obs,i} - f(y_{pred,i})]^2}{\sum_i (y_{obs,i} - \bar{y}_{obs})^2}$$

3.7. Equation for calculation of coefficient of determination (R^2)

3.6. Synthetic Climate Change Scenarios

To assess the impact of temperature rise on maize yield as simulated by the MLR and the process-based CERES Maize model, we developed a synthetic climate dataset to mimic the changes in climate that we may expect in the future. For this, we chose two scenarios of 1 and 2 °C rise in daily temperature (Roberts et al., 2017). It is important to note that these scenarios are not tied to any future global circulation models and are hypothetical (Roberts et al., 2017). After making these adjustments to the historical temperature dataset, both MLR and DSSAT were run to simulate new maize yields under climate change scenarios. After obtaining the original and new predicted maize yields from historical and new hypothetical climate data respectively, we measured the differences between the predicted yield county-wise for both scenarios of 1 and 2 °C rise. Then, changes in yields were plotted spatially, county-to-county from both models and scenarios. To evaluate the overall yield changes across Kansas, bar plots were developed for both models and scenarios.

Chapter 4 - Results and Discussion

4.1. Multiple Linear Regression (MLR) Model

The R^2 values for the MLR models for the 40 counties of Kansas ranged from 0.32 to 0.87 (Figure 4.1) which means that at least 32% of the variability in year-to-year maize yield change can be explained by climate variables with maximum and minimum variability observed in Rice and Graham counties, respectively. The variability in R^2 (Figure 4.1) showed a similar pattern to the variability of PRECIP-July across Kansas. This shows that much of the model's explanatory power is derived from PRECIP-July. The lower R^2 observed in few counties could be attributed to the high year-to-year fluctuation in rainfed maize yield. The wide range of R^2 could be likely attributed to missing and non-uniform observed yield data, variations in length of growing seasons within counties, climate responses at county-scale, changes in economics, and other conditions that influence crop management. Consideration of these factors would likely improve the performance of the model. However, the MLR model demonstrated the ability to explain 30% to 90% of the overall variability in rainfed maize yield. These county-scale results are relatively lower compared to global and state-scale studies. Lobell & Field, (2007) found 47% of year-to-year yield variability in maize yield at a global scale was explained by climate variables and similarly Roberts et al., (2013) reported an R^2 of 0.72 for a regression model between maize yields and climate variables for Illinois.

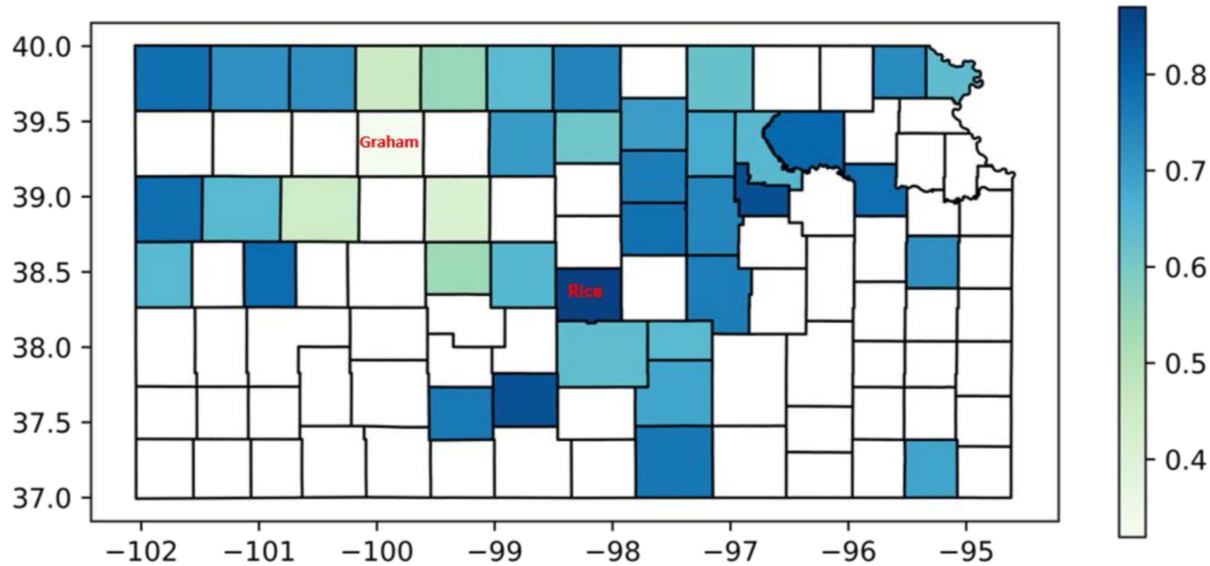


Figure 4.1. Spatial distribution of R^2 obtained from MLR model for selected rainfed maize producing counties of Kansas.

The RMSE for maize yield ranged from 483 kg ha^{-1} to 1388 kg ha^{-1} for 40 counties of Kansas as shown in Figure 4.2. The minimum and maximum values of RMSE were observed for Pratt (483 kg ha^{-1}) and Ellis county (1388 kg ha^{-1}) respectively. The likely cause of this is due to the non-uniform spatial distribution of the rainfed maize yield data from 1981 to 2020. We observed that range of NASS rainfed maize yields from 1981 to 2020 across Kansas: 200-1200 kg ha^{-1} in the north-east region, 100-700 kg ha^{-1} in the north-west region, and 500-800 kg ha^{-1} in the central region. The variability in these yields can be attributed to temperature and precipitation trends. Also, we found that in certain years, rainfed yields were considerably very less in western region and central region due to drought and heat, particularly in the year 1983, 2000, 2011, and 2012. The models' RMSE results are similar to studies that have considered soil parameters in addition to climate variables in their statistical models to predict maize yield (Ansarifar et al., 2021) and with studies conducted using a suite of statistical and machine learning models for maize (Shahhosseini et al., 2021) and soybean (Joshi et al., 2021).

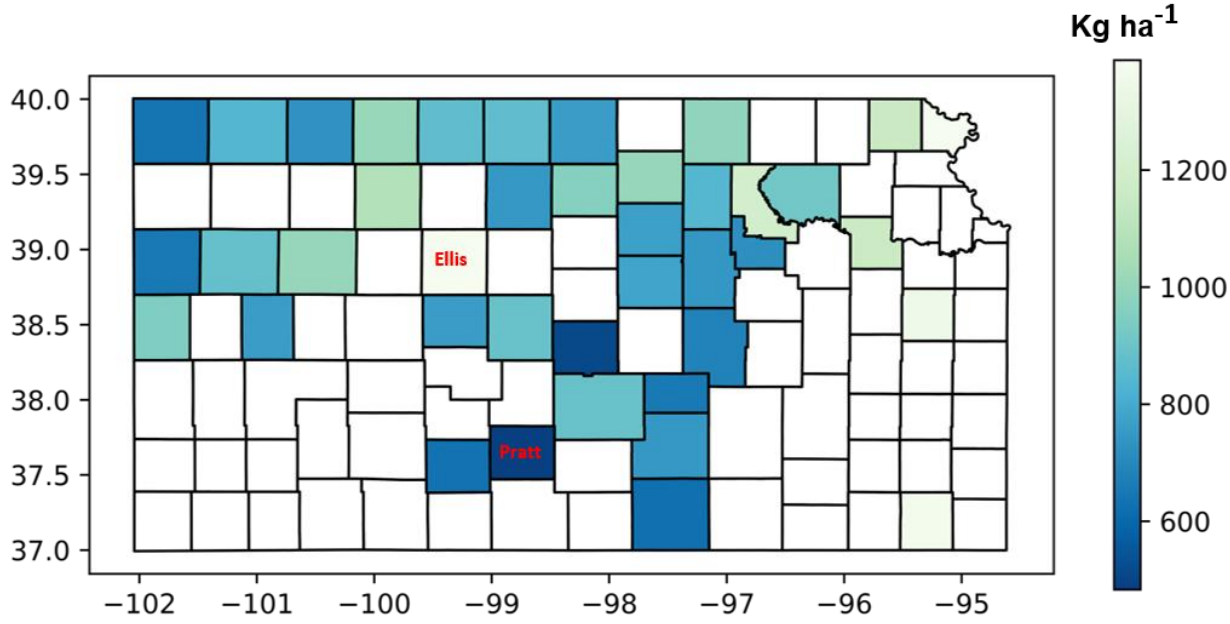


Figure 4.2. Spatial distribution of RMSE obtained from MLR model for selected rainfed maize producing counties of Kansas.

Figure 4.3 shows the range of coefficients of the predictor variables for the county-level MLR models with counts representing number of counties having those variables. Overall, in MLR models, 60% of the counties had GDD and 70% had EDD followed by 55% counties with PRECIP-June, and 52% with PRECIP-July as predictor variables. It shows that GDD and EDD are the most common variables for maize yield prediction along with PRECIP-June and PRECIP-July. It can also be observed that GDD and total PRECIP-July have a positive relationship whereas EDD has a strongly negative relationship with maize yield (Figure 4.3). Roberts et al., (2013) also found that GDD and EDD showed positive and negative relationship with the maize yield respectively. We found that range of temperature coefficients are wider compared to other variables as shown in figure 4.3.

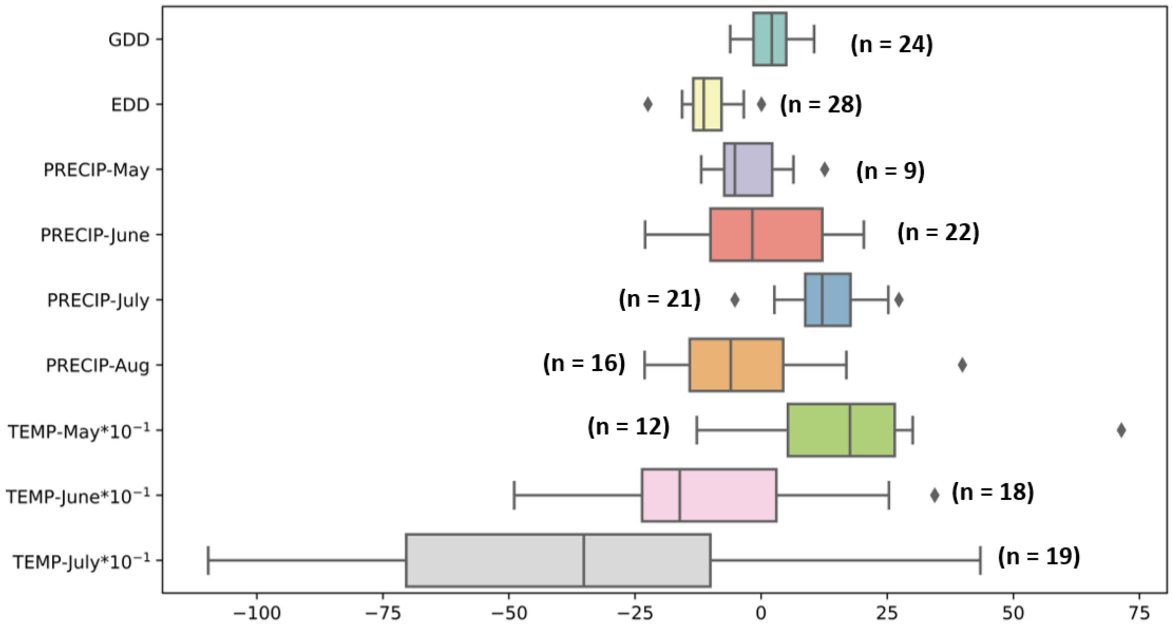


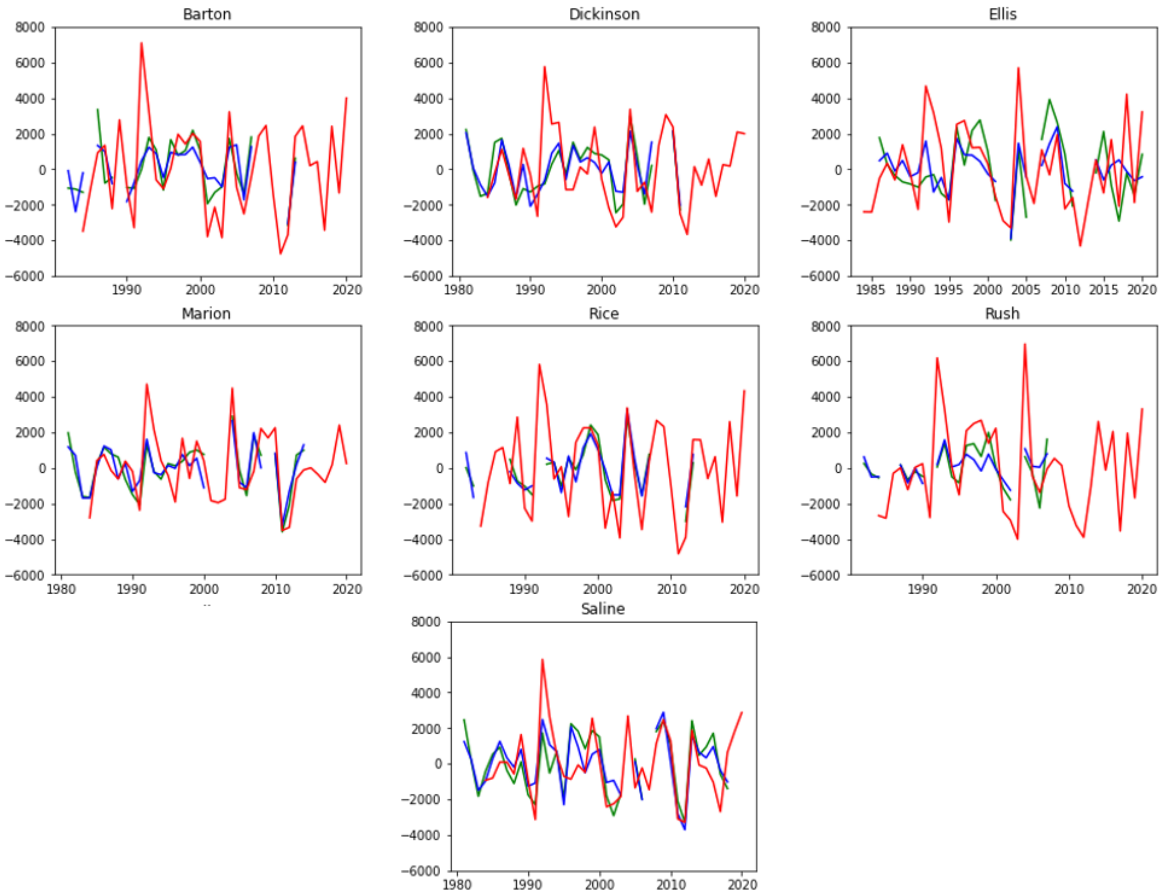
Figure 4.3. Range of coefficient for variables of county-level MLR models with counts for selected rainfed maize producing counties of Kansas.

PRECIP-June and PRECIP-July are important parameters because they play a major role during reproductive stage of maize along with TEMP-July that results in reduction of maize yield because high temperature shorten the reproductive period and hastens grain filling (Abendroth et al., 2021). During flowering, high day time temperature can result in pollen sterility, and reduced seed count, whereas high night time temperatures result in faster growing degree days (GDDs) accumulation that lead to early maize maturation (Herrero & Johnson, 1980; Cross et al., 2003; Echer et al., 2014). The distribution of these variables across Kansas depend upon location of county, weather condition, and management practices. Our analysis revealed that GDD emerged as a crucial predictor in the north-eastern and central regions of Kansas, while EDD and PRECIP-July did not exhibit any specific pattern across state. Understanding the dynamics of these variables is important for making informed decisions about crop growth and development, particularly in regions where farmers rely heavily on precipitation for agricultural production. By having accurate knowledge of local climate conditions, implementing appropriate management practices, and understanding the specific requirements of crop growth in these areas, we can strive towards achieving improved yields that directly benefit local farmers. This insight can pave the way for more sustainable and efficient agricultural practices, enhancing the overall resilience of the farming communities in Kansas.

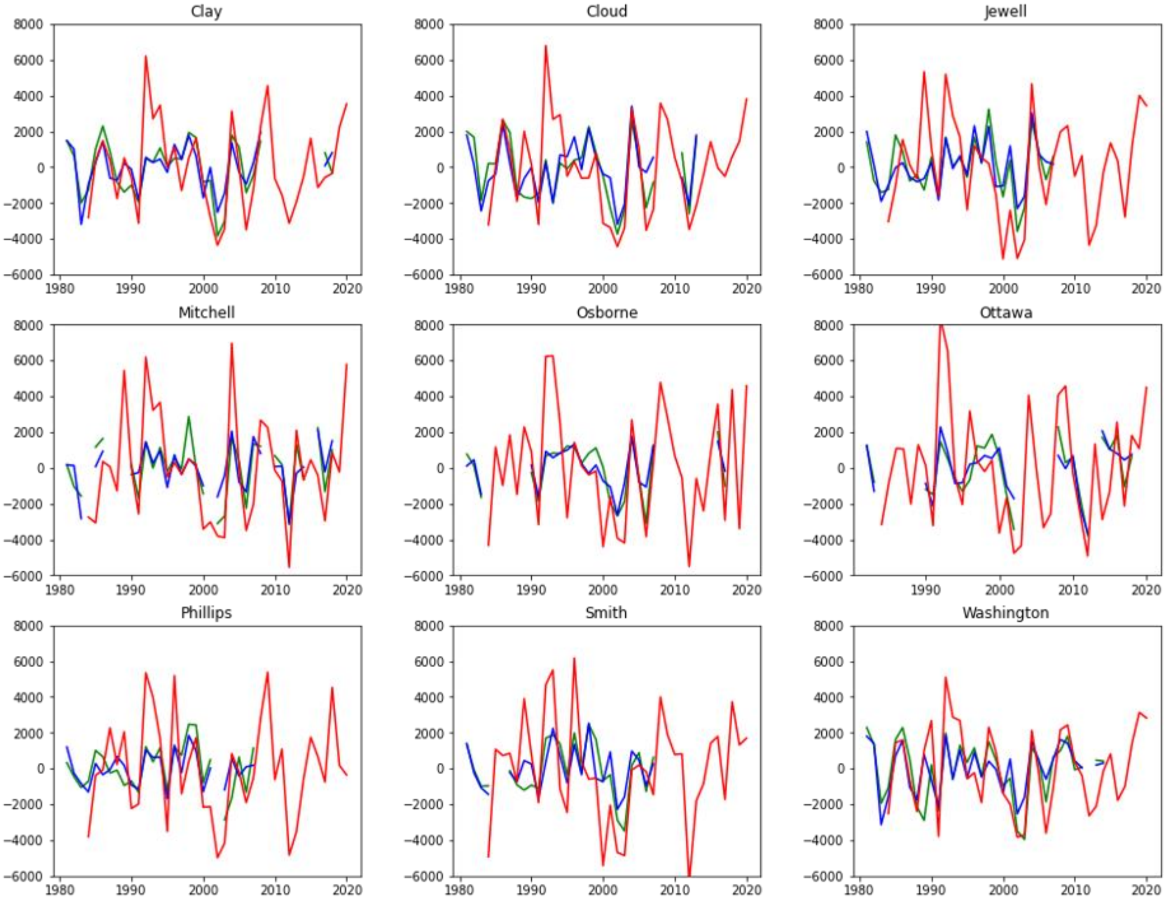
4.2. Comparison of MLR and Process-Based Models

However, evaluating the performance of each individual county (Figure 4.4) can be challenging. Therefore, we assessed the overall performance of both models across the entire state (Figure 4.5). The results indicate that the MLR model achieved an r value of 0.93 and an RMSE of 443 kg ha^{-1}, while the DSSAT model achieved an r value of 0.70 and an RMSE of 1408 kg ha^{-1}. These metrics provide insights into the performance of the models and their ability to predict maize yields in Kansas. High r value by MLR model across Kansas suggests that it is able to capture a significant portion of the variability in the observed NASS yields, implying a good fit to the data and lower RMSE indicates good level of accuracy in predicting maize yields in Kansas. On the other hand, the DSSAT shows weaker linear relationship between predicted and observed maize yields.

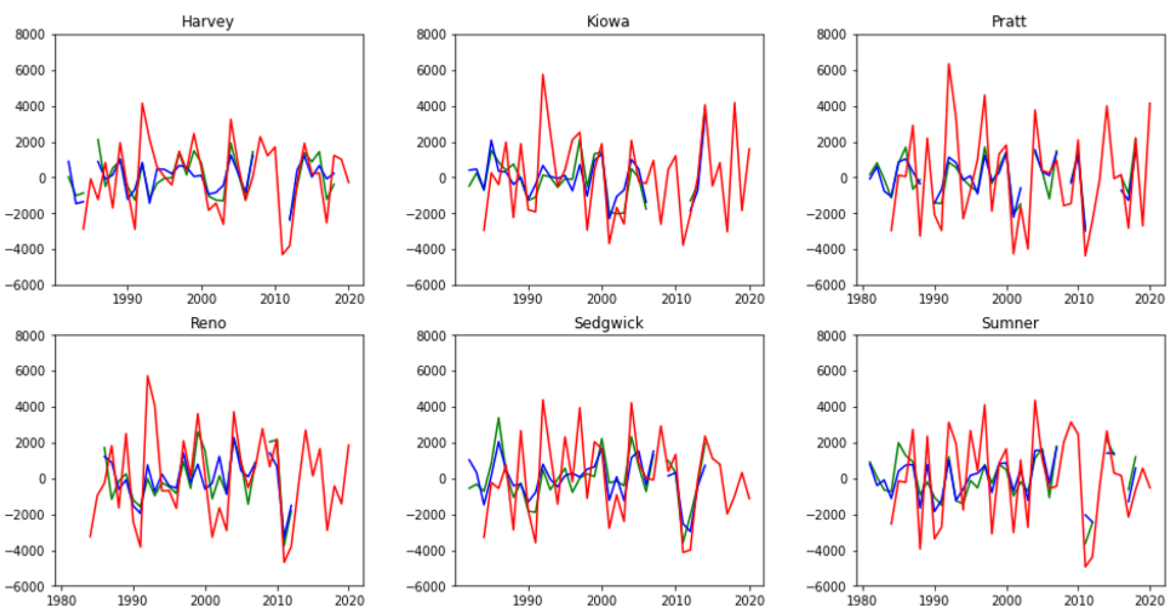
a)



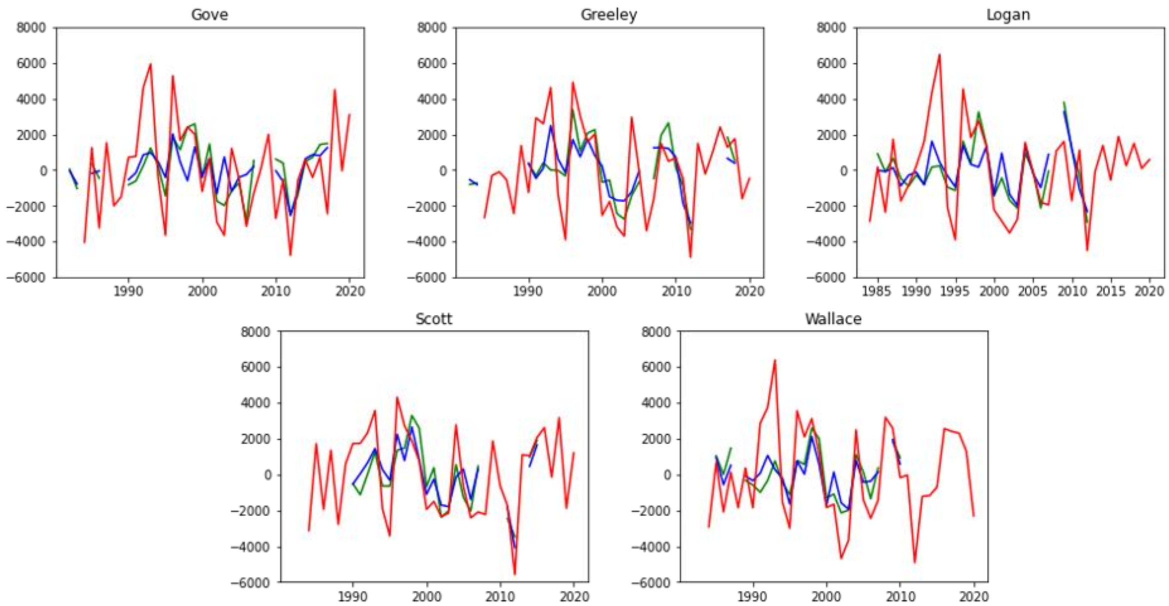
b)


Clay
Cloud
Jewell
Mitchell
Osborne
Ottawa
Phillips
Smith
Washington
8000
6000
4000
2000
0
−2000
−4000
−6000
1980
1990
2000
2010
2020

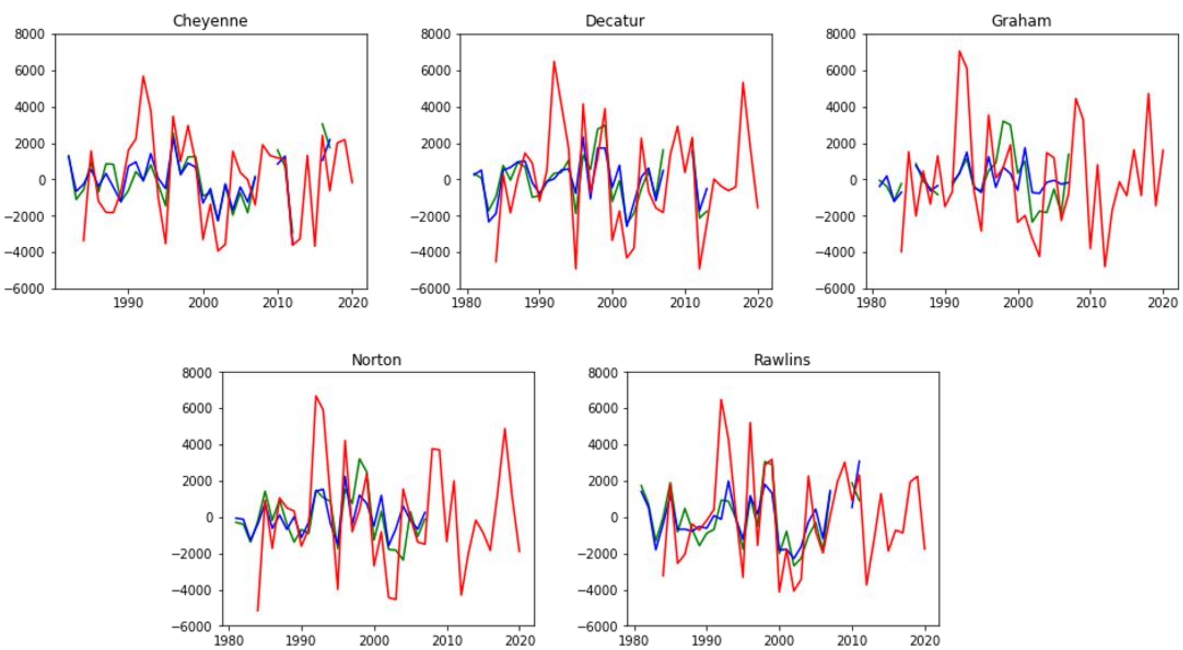
c)


Harvey
Kiowa
Pratt
Reno
Sedgwick
Sumner
8000
6000
4000
2000
0
−2000
−4000
−6000
1980
1990
2000
2010
2020

d)


Gove
Greeley
Logan
Scott
Wallace

e)


Cheyenne
Decatur
Graham
Norton
Rawlins

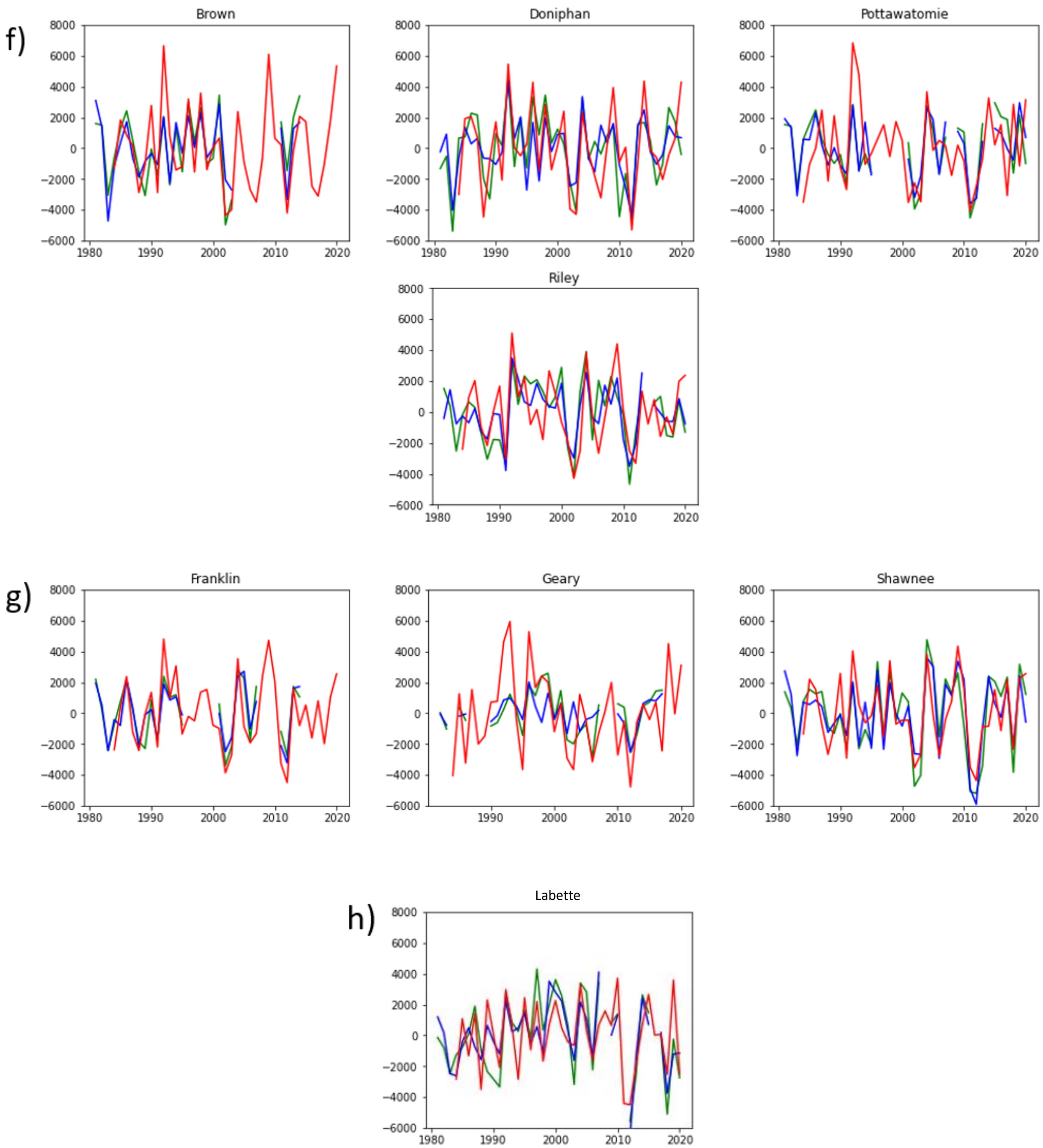


Figure 4.4. Time series of observed (green) and predicted yield from models (simulated yield from DSSAT in red, predicted yield from MLR in blue) where y-axis represents detrended predicted yield (a) Central, (b) North Central, (c) South Central, (d) West Central, (e) North Western, (f) North Eastern, (g) Eastern Central, (h) South Eastern regions in Kansas.

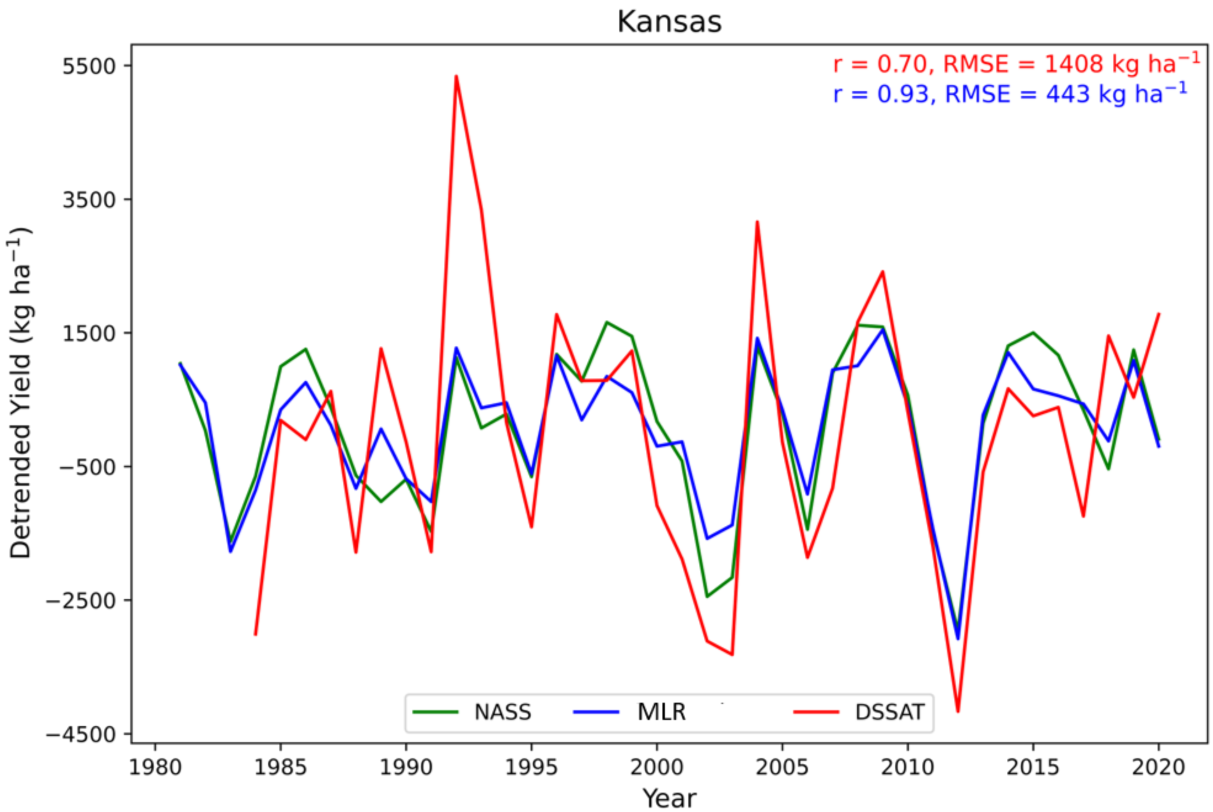


Figure 4.5. Time series plot of observed (NASS) and predicted yields from MLR and DSSAT models.

Due to unavailability of solar radiation data from NASA POWER for the period of 1981 to 1983, the time series were compared from 1984 to 2020. From time-series plot and statistical metrics analysis (Figure 4.5), we found that MLR model yields had a stronger association ($r = 0.93$) with the observed yield than the process-based DSSAT model ($r = 0.70$) and that the MLR model had a lower RMSE than DSSAT. A study by Liu et al., (2022) also compared rainfed maize yield from multiple machine learning (ML) and process-based DSSAT model for US Corn Belt and found that RMSE was about three times higher in DSSAT compared to the best performing ML model, and DSSAT explained 16% spatiotemporal variance of the observed maize yield. However, caution should be exercised when interpreting the results of comparisons because the MLR model was trained on observed yield data, while limited calibration was done in DSSAT (Leng & Hall, 2020). Also, DSSAT yields were simulated for the year 1984 to 2020 whereas for the MLR model, the yield was predicted based on observed USDA-NASS that has missing maize yield data and is not uniform. Nevertheless, it is essential to highlight the promise of the MLR for yield prediction compared to process-based models because it requires considerably fewer inputs and shows time-series patterns similar to observed yield.

4.3. Climate Change Impacts on Predicted Yields

The comparison of predicted maize yield from both models for two synthetic climate change scenarios of 1 and 2 °C rise in temperature (scenario I and II, respectively) showed that predicted maize yield reduction is greater in DSSAT in comparison to the MLR model (Figures 4.6 and 4.7). Using the MLR model, few counties showed an increase in predicted maize yield, but with DSSAT all counties showed reductions in both scenarios. In Scenario I, the predicted maize yield from MLR showed an increase within the range of 20 to 469 kg ha^{-1}, whereas a decrease in yield in the range of 19 to 1118 kg ha^{-1} was observed. In Scenario II, the range of increase in predicted yield varied between 8 to 455 kg ha^{-1}, and the decrease ranged from 19 to 1136 kg ha^{-1}. DSSAT showed reduction in the predicted rainfed maize yield in both scenarios. In Scenario I, reductions in predicted maize yield ranged between 280 and 980 kg ha^{-1} whereas, in Scenario II, reductions ranged from 590 to 1810 kg ha^{-1}. The difference between yield reductions under both scenarios as predicted by the MLR and DSSAT model could have several causes. In DSSAT, planting date and plant population vary county-to-county whereas in the MLR model, planting date was fixed for all selected counties.

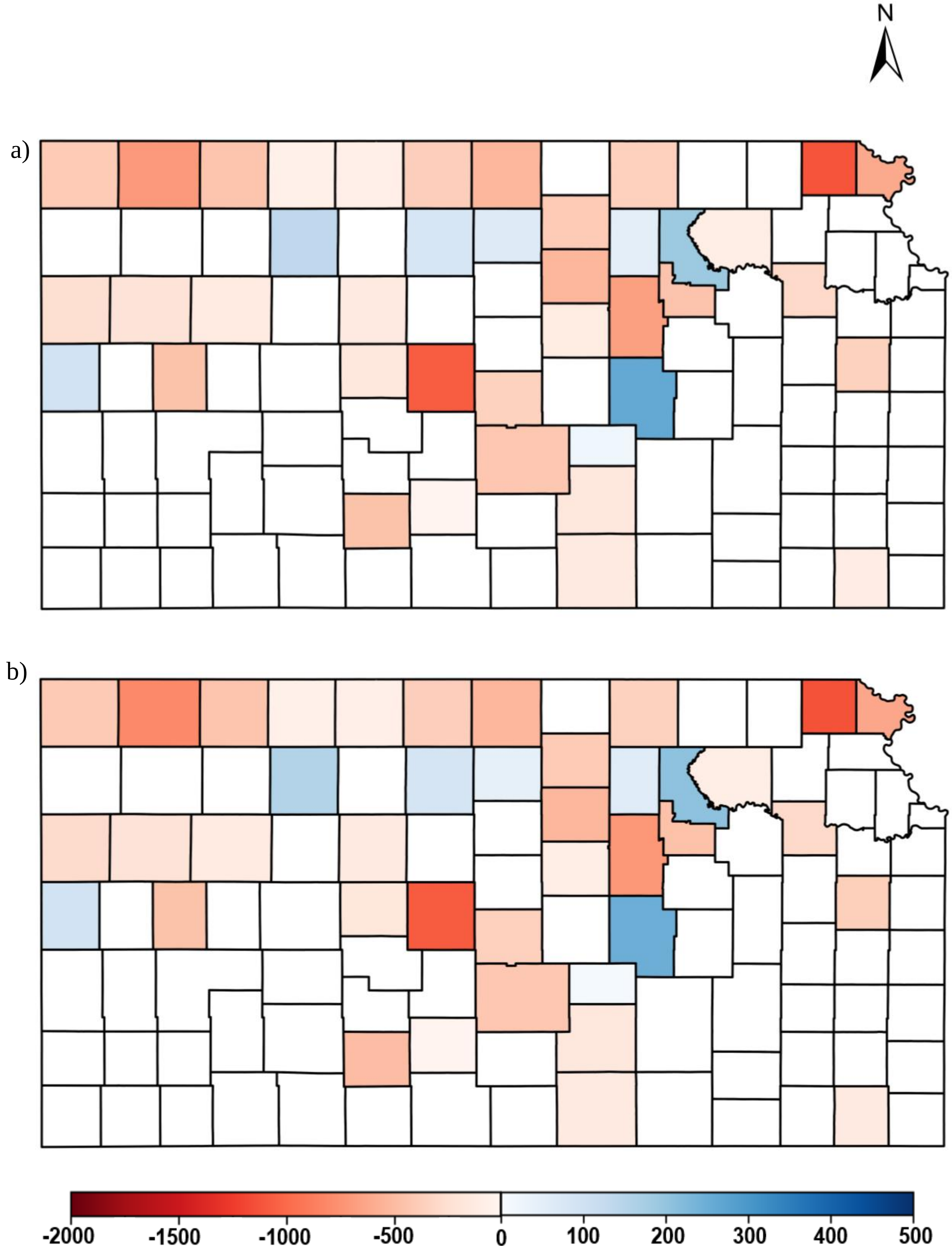


Figure 4.6. Predicted yield changes (kg ha^{-1}) for each county of Kansas from the MLR model, for (a) Scenario I as +1 and (b) Scenarios II as +2 °C rise in temperature

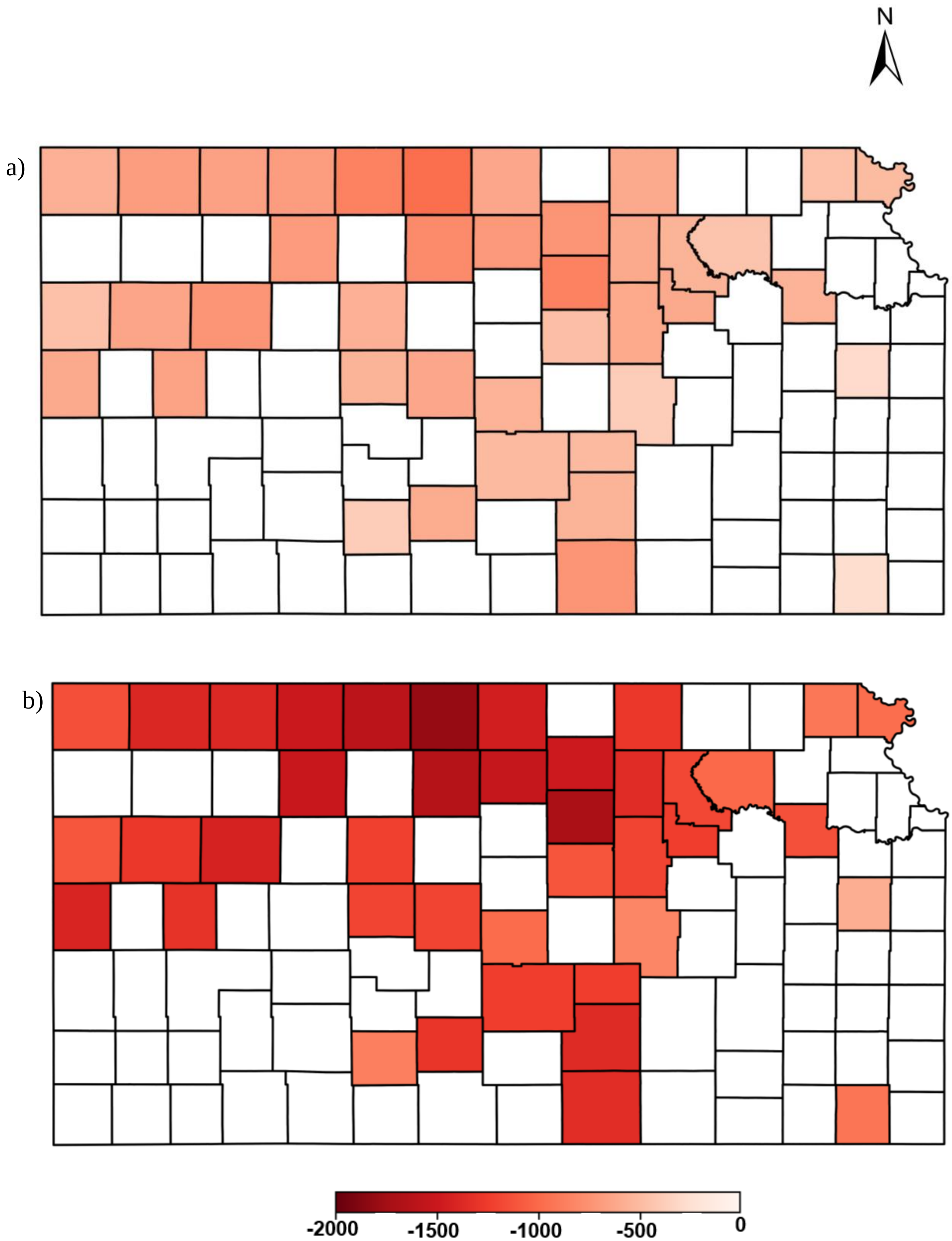


Figure 4.7. Predicted yield changes (kg ha^{-1}) for each county of Kansas from the DSSAT model, for (a) Scenario I as +1 and (b) Scenario II as +2 °C rise in temperature

In the MLR model, the growing season was set to April to September whereas in DSSAT this was dependent on time it took for the crop to mature from the date of planting, hence making it more sensitive to environmental factors. Overall, predicted maize yield reduction is higher in DSSAT because DSSAT requires daily weather information, wheareas MLR used the average value of weather for the growing period of maize. The use of daily weather information in DSSAT allows for a more detailed and dynamic representation of crop responses to changing weather conditions whereas MLR reliance on average weather values for the growing period simplifies the weather input and may not capture the full range of variability experience by the crop. As a result, daily weather information generally leads to higher sensitivity and responsiveness to weather variations, resulting in greater predicted maize yield reduction in DSSAT compared to MLR. These results are opposite to those found by Roberts et al. (2017) who assessed the climate change impact on maize yield and found that predicted impacts are considerably more severe when using the statistical model, combined model (combination of statistical and process models) and full model (add all interactions with EDD with combined model specification) as compared to the process-based model.

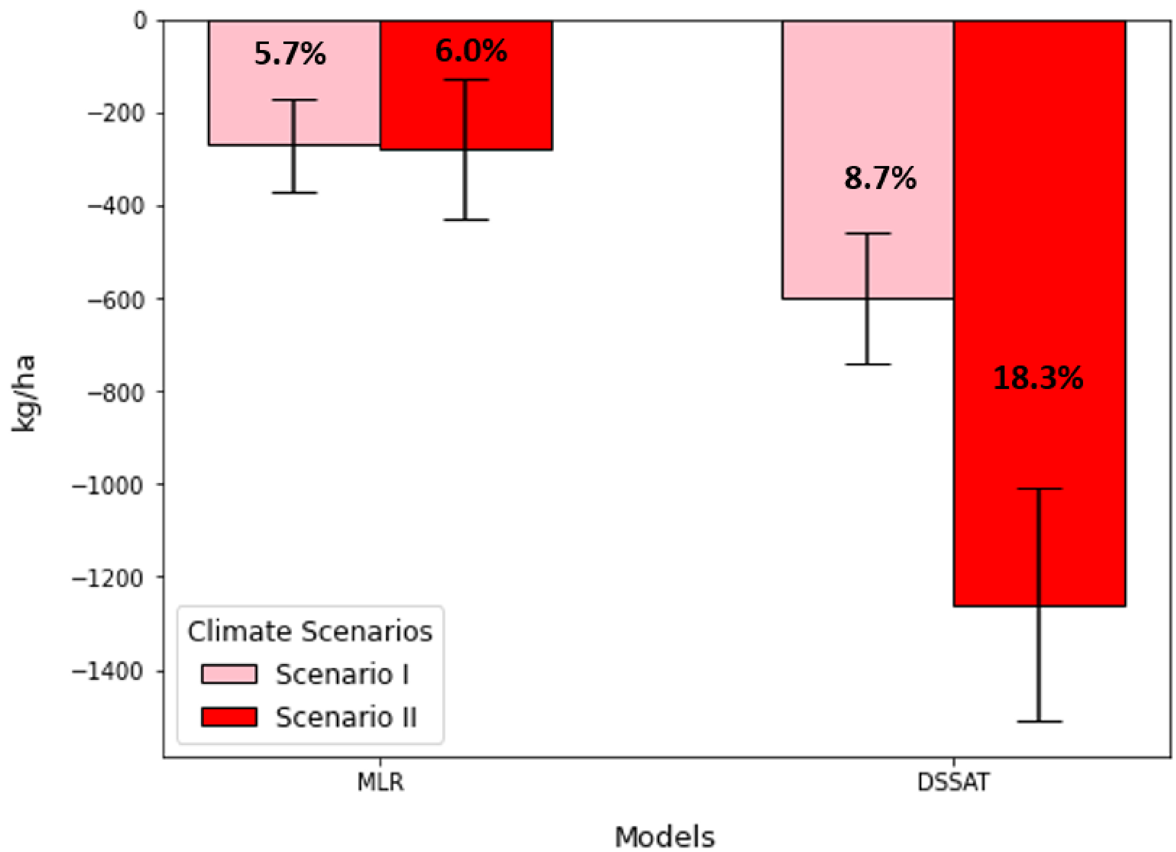


Figure 4.8. Average predicted yield change (kg ha^{-1}) for both models and each climate scenario along with the percentage change in the yield.

Figure 4.8 displays the overall impact of synthetic climate change scenarios across Kansas. It was found that reduction in predicted rainfed maize yield by MLR model was nearly 6% in both scenarios. These results are similar to those reported by a global maize yield prediction by regression model (Lobell & Field, 2007) who found that a 1 °C rise in temperature resulted in 8.3% yield reduction and those of a study conducted in central Corn Belt (Hatfield et al., 2011) that reported 5 to 8% reduction in simulated maize yields per 2 °C temperature increase. According to another study (Roberts et al., 2017), three different regression models developed for maize yield prediction in Illinois found reduction in yield to be in the range of 10 to 60% for the years 2040-2069 relative to the 1981-2010 baseline scenario. We found that reduction in predicted rainfed maize yield by DSSAT is 8.7% and 18.3% for Scenario I and

Scenario II, respectively. These results are similar to those reported by Roberts et al. (2017) with a process-based model showing a 0.112 kg ha^{-1} reduction in maize yield with a temperature increase of 2 °C. Overall, the results presented in figure 4.8 clearly indicate that irrespective of the type of model used to predict yield, predicted rainfed maize yield will reduce with an increase in temperature. Predicted climate change impacts are considerably more severe under the DSSAT process-based model compared to MLR model which could be explained by extreme heat effects (Leng & Hall, 2020; Liu et al., 2022), which are not captured by the MLR model because only DSSAT uses daily weather information.

Chapter 5 - Conclusions and Summary

In this study, we delved into the understanding how climate change affects maize yields in the agricultural areas of Kansas. Our research centered around comparing two types of models: a statistical Multiple Linear Regression (MLR) model and a processe based the Decision Support System for Agrotechnology Transfer (DSSAT) model. By doing so, we aimed to shed light on the most effective approach for predicting maize yields under changing climatic conditions.

We found that using the developed MLR model, at least 32% of predicted year-to-year rainfed maize yield variability was explained by climate variables in Kansas counties, which suggests that climate plays an important role in rainfed maize yield predictions. We also found that MLR outperformed DSSAT, as is evident from higher correlation coefficient (r=0.93), and lower root mean square error (RMSE=443 kg ha^{-1}) of MLR as compared to lower r (0.70) and higher RMSE (1408 kg ha^{-1}) observed with DSSAT. Notably, the DSSAT model's performance hinged largely upon crop management data such as planting date and plant population, and soil information, variables that were not a part of the MLR models.

This study demonstrated that rising temperature generally have a detrimental impact on maize yields in Kansas. As temperature rose, the predicted reductions in rainfed maize yields were more pronounced, particularly in the DSSAT model. This highlights the vulnerability of maize crops to the intensifying heat that is expected with climate change. However, it is important to recognize the limitations of these models. Both the MLR and the DSSAT models have their respective constraints when it comes to mechanistic understanding, computation demand and complexity. It is interesting that the two models project very different impacts from climate change. While we think these differences deserve further research, we do not think it is appropriate to conclude that process-based models generally imply greater damage from climate change than statistical-based models. It is worthwhile considering a wider set of models (Lobell & Asseng, 2017) as suggested by other studies.

References

Abendroth, L. J., Elmore, R. W., Boyer, M. J., & Marlay, S. K. (2010). Understanding corn development: A key for successful crop management. *Proceedings of the Integrated Crop Management Conference*, 10972934. https://doi.org/10.31274/icm-180809-33

Abendroth, L. J., Miguez, F. E., Castellano, M. J., Carter, P. R., Messina, C. D., Dixon, P. M., & Hatfield, J. L. (2021). Lengthening of maize maturity time is not a widespread climate change adaptation strategy in the US Midwest. *Global Change Biology*, *27*(11), 2426–2440. https://doi.org/10.1111/gcb.15565

Akhavizadegan, F., Ansarifar, J., Wang, L., Huber, I., & Archontoulis, S. V. (2021). A time-dependent parameter estimation framework for crop modeling. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-90835-x

Ansarifar, J., Wang, L., & Archontoulis, S. V. (2021). An interaction regression model for crop yield prediction. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-97221-7

ArcGIS Desktop | Desktop GIS Software Suite. (2020.). Retrieved March 12, 2023, from https://www.esri.com/en-us/arcgis/products/arcgis-desktop/overview

Arora, N. K. (2019). Impact of climate change on agriculture production and its sustainable solutions. *Environmental Sustainability*, *2*(2), 95–96. https://doi.org/10.1007/s42398-019-00078-w

Bassu, S., Brisson, N., Durand, J.-L., Boote, K., Lizaso, J., Jones, J. W., Rosenzweig, C., Ruane, A. C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Waha, K. (2014). How do various maize crop models vary in their responses to climate change factors? *Global Change Biology*, *20*(7), 2301–2320. https://doi.org/10.1111/gcb.12520

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2020). Publisher Correction: Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *7*(1), Article 1. https://doi.org/10.1038/s41597-020-00616-w

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y. M., Cellier, P., Debaeke, P., Gaudillère, J. P., Hénault, C., Maraux, F., Seguin, B., & Sinoquet, H. (2003). An overview of the crop model stics. *European Journal of Agronomy*, *18*(3–4), 309–332. https://doi.org/10.1016/S1161-0301(02)00110-7

Butler, E. E., & Huybers, P. (2015). Variations in the sensitivity of US maize yield to extreme temperatures by region and growth phase. *Environmental Research Letters*, *10*(3), 034009. https://doi.org/10.1088/1748-9326/10/3/034009

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, *4*(4), Article 4. https://doi.org/10.1038/nclimate2153

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

Climate at a Glance | National Centers for Environmental Information (NCEI). (2022.). Retrieved January 23, 2023, from https://www.ncei.noaa.gov/access/monitoring/climate-at-a-glance/

Climate Change: Global Temperature | NOAA Climate.gov. (2022). Retrieved January 24, 2023, from http://www.climate.gov/news-features/understanding-climate/climate-change-global-temperature

Cooper, M., Tang, T., Gho, C., Hart, T., Hammer, G., & Messina, C. (2020). Integrating genetic gain and gap analysis to predict improvements in crop productivity. *Crop Science*, *60*(2), 582–604. https://doi.org/10.1002/csc2.20109

Corn Crop Performance Tests. (2022). Retrieved March 12, 2023, from https://www.agronomy.k-state.edu/outreach-and-services/crop-performance-tests/corn/

CropScape—NASS CDL Program. (2021). Retrieved February 25, 2023, from https://nassgeodata.gmu.edu/CropScape/

Cross, R. H., Mckay, S. a. B., G. McHUGHEN, A., & Bonham-Smith, P. C. (2003). Heat-stress effects on reproduction and seed set in Linum usitatissimum L. (flax). *Plant, Cell & Environment*, *26*(7), 1013–1020. https://doi.org/10.1046/j.1365-3040.2003.01006.x

Dai, Z. Y., & Li, Y. P. (2013). A multistage irrigation water allocation model for agricultural land-use planning under uncertainty. *Agricultural Water Management*, *129*, 69–79. https://doi.org/10.1016/j.agwat.2013.07.013

DSSAT.net—Official Home of the DSSAT Cropping Systems Model. (2021). DSSAT.Net. Retrieved January 19, 2023, from https://dssat.net/

Echer, F. R., Oosterhuis, D. M., Loka, D. A., & Rosolem, C. A. (2014). High Night Temperatures During the Floral Bud Stage Increase the Abscission of Reproductive Structures in Cotton. *Journal of Agronomy and Crop Science*, *200*(3), 191–198. https://doi.org/10.1111/jac.12056

Edmonds, J. A., & Rosenberg, N. J. (2005). Climate Change Impacts for the Conterminous USA: An Integrated Assessment Summary. *Climatic Change*, *69*(1), 151–162. https://doi.org/10.1007/s10584-005-3613-8

Paraschivu, M., Cotuna, O., Paraschivu, M., & Olaru, A. (2019). Effects of Interaction Between Abiotic Stress and Pathogens in Cereals in the context of Climate Change: An Overview.

Annals of the University of Craiova—Agriculture, Montanology, Cadastre Series, 49(2), 413-424. https://anale.agro-craiova.ro/index.php/aamc/article/view/992

Erenstein, O. (2010). The Evolving Maize Sector in Asia: Challenges and Opportunities. *Journal of New Seeds*, *11*(1), 1–15. https://doi.org/10.1080/15228860903517770

Erenstein, O., Jaleta, M., Sonder, K., Mottaleb, K., & Prasanna, B. M. (2022). Global maize production, consumption and trade: Trends and R&D implications. *Food Security*, *14*(5), 1295–1319. https://doi.org/10.1007/s12571-022-01288-7

ESS: Crops Statistics—Concepts, Definitions and Classifications. (n.d.). Retrieved January 24, 2023, from https://www.fao.org/economic/the-statistics-division-ess/methodology/methodology-systems/crops-statistics-concepts-definitions-and-classifications/en/

Irmak, A., Jones, J. W., & Jagpat, S. S. (2005). Evaluation of the Cropgro-Soybean Model for Assessing Climate Impacts on Regional Soybean Yields. *Transcation of the ASAE. 48(6),* 2343-2353. https://doi.org/10.13031/2013.20073

FAOSTAT. (2022). Retrieved January 24, 2023, from https://www.fao.org/faostat/en/#home

Feenstra, J. F., Burton, I., Smith, J. B., & Tol, R. S. J. (1998). *Handbook on Methods for Climate Change Impact Assessment and Adaptation Strategies*.

Feng, S., Hao, Z., Zhang, X., & Hao, F. (2019). Probabilistic evaluation of the impact of compound dry-hot events on global maize yields. *The Science of the Total Environment*, *689*, 1228–1234. https://doi.org/10.1016/j.scitotenv.2019.06.373

Feng, S., Hao, Z., Zhang, X., & Hao, F. (2021). Changes in climate-crop yield relationships affect risks of crop yield reduction. *Agricultural and Forest Meteorology*, *304–305*, 108401. https://doi.org/10.1016/j.agrformet.2021.108401

Filippi, C., Mansini, R., & Stevanato, E. (2017). Mixed integer linear programming models for optimal crop selection. *Computers & Operations Research*, *81*, 26–39. https://doi.org/10.1016/j.cor.2016.12.004

Fischer, G., Shah, M., Tubiello, F. N., & van Velhuizen, H. (2005). Socio-economic and climate change impacts on agriculture: An integrated assessment, 1990-2080. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1463), 2067–2083. https://doi.org/10.1098/rstb.2005.1744

Global Map of Irrigation Areas (GMIA) | Tierras y Aguas | Organización de las Naciones Unidas para la Alimentación y la Agricultura | Land & Water | Food and Agriculture Organization of the United Nations. (n.d.). Retrieved March 12, 2023, from https://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/es/c/1029519/

Gridded Soil Survey Geographic (gSSURGO) Database. (2021). Natural Resources Conservation Service. Retrieved March 12, 2023, from http://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database

Hastie, T., Tibshirani, R., & Friedman, J. (2009). Random Forests. In T. Hastie, R. Tibshirani, & J. Friedman (Eds.), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (pp. 587–604). Springer. https://doi.org/10.1007/978-0-387-84858-7_15

Hatfield, J. L., Boote, K. J., Kimball, B. A., Ziska, L. H., Izaurralde, R. C., Ort, D., Thomson, A. M., & Wolfe, D. (2011). Climate Impacts on Agriculture: Implications for Crop Production. *Agronomy Journal*, *103*(2), 351–370. https://doi.org/10.2134/agronj2010.0303

Herrero, M. P., & Johnson, R. R. (1980). High Temperature Stress and Pollen Viability of Maize1. *Crop Science*, *20*(6), cropsci1980.0011183X002000060030x. https://doi.org/10.2135/cropsci1980.0011183X002000060030x

Hipólito, J., Boscolo, D., & Viana, B. F. (2018). Landscape and crop management strategies to conserve pollination services and increase yields in tropical coffee farms. *Agriculture, Ecosystems & Environment*, *256*, 218–225. https://doi.org/10.1016/j.agee.2017.09.038

Ho, T. K. (1995). Random decision forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, *1*, 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

Introduction to ArcGIS Pro—ArcGIS Pro | Documentation. (2021). Retrieved March 12, 2023, from https://pro.arcgis.com/en/pro-app/latest/get-started/get-started.htm

IPCC (Ed.). (2000). *Emissions scenarios: Summary for policymakers;a special report of IPCC Working Group III$Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). Linear Regression. In G. James, D. Witten, T. Hastie, & R. Tibshirani (Eds.), *An Introduction to Statistical Learning: With Applications in R* (pp. 59–128). Springer US. https://doi.org/10.1007/978-1-0716-1418-1_3

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., Wilkens, P. W., Singh, U., Gijsman, A. J., & Ritchie, J. T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, *18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7

Joshi, V. R., Kazula, M. J., Coulter, J. A., Naeve, S. L., & Garcia y Garcia, A. (2021). In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *International Journal of Biometeorology*, *65*(4), 489–502. https://doi.org/10.1007/s00484-020-02039-z

Kamath, P., Patil, P., S, S., Sushma, & S, S. (2021). Crop yield forecasting using data mining. *Global Transitions Proceedings*, *2*(2), 402–407. https://doi.org/10.1016/j.gltp.2021.08.008

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., Huth, N. I., Hargreaves, J. N. G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J. P., Silburn, M., Wang, E., Brown, S., Bristow, K. L., Asseng, S., … Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Kennett, D. J., Prufer, K. M., Culleton, B. J., George, R. J., Robinson, M., Trask, W. R., Buckley, G. M., Moes, E., Kate, E. J., Harper, T. K., O'Donnell, L., Ray, E. E., Hill, E. C., Alsgaard, A., Merriman, C., Meredith, C., Edgar, H. J. H., Awe, J. J., & Gutierrez, S. M. (2020). Early isotopic evidence for maize as a staple grain in the Americas. *Science Advances*, *6*(23), eaba3245. https://doi.org/10.1126/sciadv.aba3245

KS-rank22.pdf. (2022). Retrieved January 24, 2023, from https://www.nass.usda.gov/Statistics_by_State/Kansas/Publications/Economic_Releases/Rank/2022/KS-rank22.pdf

Kunkel, K. E. (2022). *State Climate Summaries for the United States 2022. NOAA Technical Report NESDIS 150*. NOAA NESDIS. https://statesummaries.ncics.org/chapter/ks

L. Ma, L. R. Ahuja, B. T. Nolan, R. W. Malone, T. J. Trout, & Z. Qi. (2012). Root Zone Water Quality Model (RZWQM2): Model Use, Calibration, and Validation. *Transactions of the ASABE*, *55*(4), 1425–1446. https://doi.org/10.13031/2013.42252

L818.pdf. (1996). Retrieved March 12, 2023, from https://bookstore.ksre.ksu.edu/pubs/l818.pdf

Leng, G., & Hall, J. W. (2020). Predicting spatial and temporal variability in crop yields: An inter-comparison of machine learning, regression and process-based models. *Environmental Research Letters*, *15*(4), 044027. https://doi.org/10.1088/1748-9326/ab7b24

Li, Y., Guan, K., Yu, A., Peng, B., Zhao, L., Li, B., & Peng, J. (2019). Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crops Research*, *234*, 55–65. https://doi.org/10.1016/j.fcr.2019.02.005

Lin, X., Harrington, J., Ciampitti, I., Gowda, P., Brown, D., & Kisekka, I. (2017). Kansas Trends and Changes in Temperature, Precipitation, Drought, and Frost-Free Days from the 1890s to 2015. *Journal of Contemporary Water Research & Education*, *162*(1), 18–30. https://doi.org/10.1111/j.1936-704X.2017.03257.x

Liu, Q., Yang, M., Mohammadi, K., Song, D., Bi, J., & Wang, G. (2022). Machine Learning Crop Yield Models Based on Meteorological Features and Comparison with a Process-Based Model. *Artificial Intelligence for the Earth Systems*, *1*(4). https://doi.org/10.1175/AIES-D-22-0002.1

Lobell, D. B., & Asseng, S. (2017). Comparing estimates of climate change impacts from process-based and statistical crop models. *Environmental Research Letters*, *12*(1), 015001. https://doi.org/10.1088/1748-9326/aa518a

Lobell, D. B., & Field, C. B. (2007). Global scale climate–crop yield relationships and the impacts of recent warming. *Environmental Research Letters*, *2*(1), 014002. https://doi.org/10.1088/1748-9326/2/1/014002

Lobell, D. B., Field, C. B., Cahill, K. N., & Bonfils, C. (2006). Impacts of future climate change on California perennial crop yields: Model projections with climate and crop uncertainties. *Agricultural and Forest Meteorology*, *141*(2), 208–218. https://doi.org/10.1016/j.agrformet.2006.10.006

Lobell, D. B., & Ortiz-Monasterio, J. I. (2007). Impacts of Day Versus Night Temperatures on Spring Wheat Yields:A Comparison of Empirical and CERES Model Predictions in Three Locations. *Agronomy Journal*, *99*(2), 469–477. https://doi.org/10.2134/agronj2006.0209

Maitah, M., Malec, K., & Maitah, K. (2021). Influence of precipitation and temperature on maize production in the Czech Republic from 2002 to 2019. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-89962-2

Malhi, G. S., Kaur, M., & Kaushik, P. (2021). Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. *Sustainability*, *13*(3), Article 3. https://doi.org/10.3390/su13031318

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The Relationship of Drought Frequency and Duration to Time Scales. *Eighth Conference on Applied Climatology,* https://www.droughtmanagement.info/literature/AMS_Relationship_Drought_Frequency_Duration_Time_Scales_1993.pdf

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. *Journal of Climate*, *31*(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1

Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). K-Nearest Neighbor Classification. In A. Mucherino, P. J. Papajorgji, & P. M. Pardalos (Eds.), *Data Mining in Agriculture* (pp. 83–106). Springer. https://doi.org/10.1007/978-0-387-88615-2_4

Parry, M., Rosenzweig, C., & Livermore, M. (2005). Climate change, global food supply and risk of hunger. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1463), 2125–2138. https://doi.org/10.1098/rstb.2005.1751

Plevris, V., Solorzano, G., Bakas, N., & Ben Seghier, M. (2022). *Investigation of performance metrics in regression analysis and machine learning-based prediction models*. https://doi.org/10.23967/eccomas.2022.155

POWER | Data Access Viewer. (2021). Retrieved March 12, 2023, from https://power.larc.nasa.gov/data-access-viewer/

PRISM Climate Group at Oregon State University. (2021). Retrieved January 16, 2023, from https://prism.oregonstate.edu/terms/

Ritchie, J. T. (1998). Soil water balance and plant water stress. In G. Y. Tsuji, G. Hoogenboom, & P. K. Thornton (Eds.), *Understanding Options for Agricultural Production* (pp. 41–54). Springer Netherlands. https://doi.org/10.1007/978-94-017-3624-4_3

Roberts, M. J., Braun, N. O., Sinclair, T. R., Lobell, D. B., & Schlenker, W. (2017). Comparing and combining process-based crop models and statistical models with some implications for climate change. *Environmental Research Letters*, *12*(9), 095010. https://doi.org/10.1088/1748-9326/aa7f33

Roberts, M., Schlenker, W., & Eyer, J. (2012). Agronomic Weather Measures in Econometric Models of Crop Yield with Implications for Climate Change. *American Journal of Agricultural Economics*, *95*, 236–243. https://doi.org/10.1093/ajae/aas047

Rupp, D. E., Daly, C., Doggett, M. K., Smith, J. I., & Steinberg, B. (2022). Mapping an Observation-Based Global Solar Irradiance Climatology across the Conterminous United States. *Journal of Applied Meteorology and Climatology*, *61*(7), 857–876. https://doi.org/10.1175/JAMC-D-21-0236.1

Sammut, C., & Webb, G. I. (Eds.). (2010). Logistic Regression. In *Encyclopedia of Machine Learning* (pp. 631–631). Springer US. https://doi.org/10.1007/978-0-387-30164-8_493

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Shahhosseini, M., Hu, G., & Archontoulis, S. V. (2020). Forecasting Corn Yield With Machine Learning Ensembles. *Frontiers in Plant Science*, *11*. https://www.frontiersin.org/articles/10.3389/fpls.2020.01120

Shahhosseini, M., Hu, G., Khaki, S., & Archontoulis, S. V. (2021). Corn Yield Prediction With Ensemble CNN-DNN. *Frontiers in Plant Science*, *12*, 709008. https://doi.org/10.3389/fpls.2021.709008

Shahzad, A., Ullah, S., Dar, A. A., Sardar, M. F., Mehmood, T., Tufail, M. A., Shakoor, A., & Haris, M. (2021). Nexus on climate change: Agriculture and possible solution to cope future climate change stresses. *Environmental Science and Pollution Research*, *28*(12), 14211–14232. https://doi.org/10.1007/s11356-021-12649-8

Sharda, V., Handyside, C., Chaves, B., McNider, R. T., & Hoogenboom, G. (2017). The Impact of Spatial Soil Variability on Simulation of Regional Maize Yield. *Transactions of the ASABE*, *60*(6), 2137–2148. https://doi.org/10.13031/trans.12374

Sharda, V., Mekonnen, M. M., Ray, C., & Gowda, P. H. (2021). Use of Multiple Environment Variety Trials Data to Simulate Maize Yields in the Ogallala Aquifer Region: A Two Model Approach. *JAWRA Journal of the American Water Resources Association*, *57*(2), 281–295. https://doi.org/10.1111/1752-1688.12873

Sharma, R. K., Kumar, S., Vatta, K., Bheemanahalli, R., Dhillon, J., & Reddy, K. N. (2022). Impact of recent climate change on corn, rice, and wheat in southeastern USA. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-21454-3

Sheehy, J. E., Mitchell, P. L., & Ferrer, A. B. (2006). Decline in rice grain yields with temperature: Models and correlations can give different estimates. *Field Crops Research*, *98*(2), 151–156. https://doi.org/10.1016/j.fcr.2006.01.001

Southworth, J., Randolph, J. C., Habeck, M., Doering, O. C., Pfeifer, R. A., Rao, D. G., & Johnston, J. J. (2000). Consequences of future climate change and changing climate variability on maize yields in the midwestern United States. *Agriculture, Ecosystems & Environment*, *82*(1), 139–158. https://doi.org/10.1016/S0167-8809(00)00223-1

Su, Y., Gabrielle, B., & Makowski, D. (2021). The impact of climate change on the productivity of conservation agriculture. *Nature Climate Change*, *11*(7), Article 7. https://doi.org/10.1038/s41558-021-01075-w

Tao, F., Yokozawa, M., Liu, J., & Zhang, Z. (2008). Climate–crop yield relationships at provincial scales in China and the impacts of recent climate trends. *Climate Research*, *38*, 83–94. https://doi.org/10.3354/cr00771

Troy, T. J., Kipgen, C., & Pal, I. (2015). The impact of climate extremes and irrigation on US crop yields. *Environmental Research Letters*, *10*(5), 054013. https://doi.org/10.1088/1748-9326/10/5/054013

USDA - National Agricultural Statistics Service—Research and Science—CropScape and Cropland Data Layers. (2021). Retrieved March 2, 2023, from https://www.nass.usda.gov/Research_and_Science/Cropland/sarsfaqs2.php#Section1_14.0

van Ittersum, M. K., Leffelaar, P. A., van Keulen, H., Kropff, M. J., Bastiaans, L., & Goudriaan, J. (2003). On approaches and applications of the Wageningen crop models. *European Journal of Agronomy*, *18*(3), 201–234. https://doi.org/10.1016/S1161-0301(02)00106-5

van Oort, P. A. J., Zhang, T., de Vries, M. E., Heinemann, A. B., & Meinke, H. (2011). Correlation between temperature and phenology prediction error in rice (Oryza sativa L.). *Agricultural and Forest Meteorology*, *151*(12), 1545–1555. https://doi.org/10.1016/j.agrformet.2011.06.012

White, J. W., Hoogenboom, G., Kimball, B. A., & Wall, G. W. (2011). Methodologies for simulating impacts of climate change on crop production. *Field Crops Research*, *124*(3), 357–368. https://doi.org/10.1016/j.fcr.2011.07.001

Wu, Z., Huang, N. E., Long, S. R., & Peng, C.K. (2007). On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, *104*(38), 14889–14894. https://doi.org/10.1073/pnas.0701020104