

Article

Assessing the Effectiveness of the Actuaries Climate Index for Estimating the Impact of Extreme Weather on Crop Yield and Insurance Applications

Qimeng Pan ¹, Lysa Porth ² and Hong Li ^{2,*}¹ KPMG Advisory (China) Limited, Beijing 100738, China; patricia.pan@kpmg.com² Gordon S. Lang School of Business and Economics, University of Guelph, Guelph, ON N1G 2W1, Canada; lporth@uoguelph.ca

* Correspondence: lihong@uoguelph.ca

Abstract: This paper investigates the effectiveness of the Actuaries Climate Index (ACI), a climate index jointly launched by multiple actuarial societies in North America in 2016, on predicting crop yields and (re)insurance ratemaking. The ACI is created using a variety of climate variables reflecting extreme weather conditions in 12 subregions in the US and Canada. Using data from eight Midwestern states in the US, we find that the ACI has significant predictive power for crop yields. Moreover, allowing the constituting variables of the ACI to have data-driven rather than pre-determined weights could further improve the predictive accuracy. Furthermore, we create the county-level ACI index using high-resolution climate data and investigate its predictive power on county-level corn yields, which are more relevant to insurance practices. We find that although the self-constructed ACI index leads to a slightly worse fit due to noisier county-specific yield data, the predictive results are still reasonable. Our findings suggest that the ACI index is promising for crop yield forecasting and (re)insurance ratemaking, and its effectiveness could be further improved by allowing for the data-driven weights of the constituting variables and could be created at higher resolution levels.

Keywords: actuarial climate index; crop yield forecasting; agricultural insurance; high-resolution data; probit model



Citation: Pan, Q.; Porth, L.; Li, H. Assessing the Effectiveness of the Actuaries Climate Index for Estimating the Impact of Extreme Weather on Crop Yield and Insurance Applications. *Sustainability* **2022**, *14*, 6916. <https://doi.org/10.3390/su14116916>

Academic Editor: Luca Salvati

Received: 29 April 2022

Accepted: 31 May 2022

Published: 6 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In recent decades, global climate change has become one of the most complicated and serious challenges facing human society. Extreme weather risks impose non-negligible threats to many aspects, including resource shortage, food consumption, and human health ([1]). The Intergovernmental Panel on Climate Changes (IPCC) reports that the average observed global mean surface temperature during the decade 2006–2015 was 0.87 °C higher than that over the last half of the 19th century. The National Oceanic and Atmospheric Administration (NOAA) reports that seven of the eight warmest years on record have happened since 2001, and all of the ten warmest years have taken place since 1995. This means that even though the temperature may not have increased significantly, extreme climate events are becoming more frequent.

Agriculture is the foundation of human development and social stability. The yields of many crops, and in many locations, are strongly related to weather variability, and severe weather conditions pose threats to agricultural production ([2]). The increase in extreme climate events exacerbates the volatility of agricultural production and severe agricultural loss. For example, Ref. [3] posit that rising temperatures tend to lower maize and wheat yields in some countries. The reduction has been so large that it offsets the increase in crop yields brought by technological advances, such as biotechnology, farming equipment and practices, etc., and [4] found that extremely high temperatures can reduce corn yields

by approximately 7% in the US. Moreover, Ref. [2] indicate that extreme weather events may have more negative impacts on crops planted in tropical areas and higher latitude regions. There is significant research that focuses on the impact of extreme weather risks on crop yield using specific climate variables or weather index variables, such as the Palmer Drought Severity Index (PDSI), Cooling Degree Day (CDD), temperature, precipitation, etc. Although the majority of the existing studies focus on the changes in the average yield over time, it is the frequency of severe losses that matters to insurers in agriculture ([5]). In particular, for insurance purposes, it is not just frequency, but also severity that impacts agricultural loss.

In 2016, the American Academy of Actuaries (AAA), the Casualty Actuarial Society (CAS), the Canadian Institute of Actuaries (CIA), and the Society of Actuaries (SOA) launched the Actuaries Climate Index (ACI) [6]. The ACI is a climate index that measures the observations of extreme weather and sea levels within the continent of the United States and Canada. The ACI is intended to help actuaries, insurance companies, governments, and the general public better understand the potential effects of climate trends and extreme weather events. Actuaries not only focus on how to measure the risks facing people in their everyday lives, but also assist in mitigating, managing, and predicting the future risks facing human society. As climatologists, environmentalists, and agricultural scientists build models to assess the potential climate changes and their impacts on the environment and agriculture, actuaries establish models to analyze the effects of uncertain climate events on the financial losses of various entities. Therefore, the ACI may provide actuaries with reliable information about extreme weather conditions, which are important for estimating and modeling climate-related insurance and financial risks.

To examine the effectiveness and feasibility of the ACI, this paper adopts the ACI, as well as the constituting variables of the ACI, to build statistical models for estimating the impacts of extreme weather events on crop yields. As an illustration, our analysis focuses on the Midwest region in the US, i.e., the major crop production region in the country. In addition, this research develops a methodology for assessing and predicting the agricultural loss for crop insurance and reinsurance applications using a combination of linear regression and probit regression models. We aim to better understand the key components of the ACI that are most important for predicting crop yields. The ACI consists of multiple climate variables used in previous studies to predict crop yields (see more details in Sections 2 and 3). It could be treated as a standardized index that integrates climate information in multiple aspects (including average weather and extreme weather events) and is thus a potential standardized predictor that could be adopted by researchers and insurance practitioners to predict crop yields and price crop yield (re)insurance products. Since the ACI was recently launched on 30 November 2016, there is very little research examining the importance and feasibility of the index for forecasting and managing agriculture risks. Therefore, to the best of our knowledge, this paper provides the first discussion on the strengths and weaknesses of the ACI for agriculture (re)insurance applications.

In this paper, we apply five regression models based on different geographical scales and predictors. For each model, we consider both the linear regression model and probit regression. The probit model is used to identify the event of insurance claims. Specifically, the binary response variable is 1 when the (detrended) corn yield is below the 25th quantile of the empirical distribution of yields over the sample, and thus indicates an insurance claim. When an insurance claim occurs, the linear regression model will be used to predict the corn yield in that year, based on which the severity of insurance loss can be calculated.

The first three models are designed using corn yields from eight Midwestern states in the United States from 1961 to 2018 as the dependent variable and weather variables derived from the ACI as the independent variables [7,8]. Specifically, Model 1 considers the monthly ACI as the independent variable; Model 2 derives several individual components from the monthly ACI as independent variables, including temperature above the 90th percentile and below the 10th percentile, a maximum monthly 5-day rainfall period, annual

consecutive dry days, and wind speed above the 90th percentile; Model 3 considers similar individual ACI components (excluding wind speed) but based on an annual, rather than monthly, temporal scale to investigate the impact of averaging out the monthly weather variations on corn yield prediction. Finally, a high-level spatial resolution climate dataset including monthly county-level climate information is employed to develop a high-level resolution ACI. Models 4 and 5 derive similar weather variables from high-level resolution data and apply them to predict county-level crop yields. Model 4 focuses on Iowa, while Model 5 covers the whole Midwest region. Models 4 and 5 serve as a comparison to Models 1 to 3 when more high-resolution yield data (county-level) are used in the prediction. From an insurance point of view, county-level data are more relevant as real-world insurance products rarely cover corn yields of a whole state, let alone the whole Midwest region.

The results suggest that the ACI has reasonable predictive powers on corn yields. Specifically, the linear regression Model 1 and probit Model 1 both include July and September ACI as two significant variables and produce an R^2 of 0.8917 and a pseudo R^2 of 0.1617, respectively. Further, the linear regression Model 2 presents an R^2 of 0.9535, which is the largest R^2 among the five models. The probit Model 2 produces a pseudo R^2 of 0.3451. The R^2 and pseudo R^2 of the linear regression Model 3 and probit Model 3 are reduced to 0.7504 and 0.0739, respectively, indicating that monthly variations are important in yield prediction. When high-resolution data are used, the goodness of fit was slightly worse compared to the Midwest region regression. One possible reason might be the data become noisier at the county level. Specifically, Model 4 generates an R^2 of 0.7982 for the linear regression and a pseudo R^2 of 0.2864 for the probit regression. Finally, the R^2 of linear regression Model 5 is 0.7126, while the pseudo R^2 of probit Model 5 is 0.2341.

Among the five models, Model 2 leads to the best goodness of fit. However, the county-level regressions, which are more valuable for insurance practices, also lead to reasonable goodness of fit. Therefore, our results indicate that the ACI index can be promising in modeling and predicting crop yields. Moreover, it would be beneficial to the insurance industry if the weights of the constituting variables of the ACI could be determined in a data-driven way, rather than being predetermined. Finally, the insurance sector could benefit if the ACI could be published at more high-resolution levels, such as the county level.

The rest of the paper is organized as follows. Section 2 provides a literature review. Section 3 introduces the construction of the ACI and its applications on (re)insurance. Section 4 introduces the data. Section 5 outlines the statistical models. Section 6 discusses the empirical results. Finally, Section 7 concludes the paper and Section 8 outlines future research directions. Other materials are delegated to the Appendix A.

2. Literature Review

2.1. Statistical Models for Crop Yields Prediction

Regression models are commonly adopted to assess the extreme climate risk probability distribution and the impacts of extreme weather events on agricultural production ([9]). Production function models usually use cross-section or time-series data to evaluate the impacts of temperature, precipitation, labor, and fertilization on crop yields. Regression analysis is employed to quantify the marginal effects of these factors and predict crop yields in different climates. Compared to agronomic process-based models and economic models, statistical models are easy to implement and have lower data requirements ([10]). For example, the most basic statistical model only needs historical yields and weather data. Furthermore, statistical models are more transparent in dealing with uncertainties compared to many other models. In particular, a statistical model is unable to reflect the way that weather affects crop yields if it has relatively low goodness of fit and many of the predictors are insignificant ([11]).

Time horizon is shown to be an important factor in crop yield forecasting. Ref. [12] recommend using a reweighting method on weather data over a long period, such as over 100 years, and employ a Generalized Linear Model (GLM) linked with a fractional logit

regression model to build a weather index to estimate crop yield loss more reliably. Further, ref. [13] shows that time horizon is an essential factor in crop insurance ratemaking and develops a conditional Weibull model to determine the effects of sample period span and weather variation on yield risk and pricing estimation.

In terms of regression methods, many researchers adopt three main types of statistical approaches, time series, cross-section, and panel data methods, to predict crop yields and related insurance losses ([10,11,14]). Ref. [11] discuss the advantages and disadvantages of these methods and compare them to process-based simulation methods. Ref. [14] use panel regression to assess the impacts of temperature and precipitation on crop yields in the Pampas since 1971 and find that climate has relatively small but significant negative effects. Ref. [5] employ the Superposed Epoch Analysis (SEA) in a time-series regression to assess the impact of extreme weather events on crop production and observe that both drought and extreme heat lead to national production reduction globally.

Finally, crop yields are seasonal and are affected by a variety of weather variables. Variable selection is necessary to reduce the complexity of regression models and determine which variables contribute most to crop yield prediction. Three regularization techniques based on the Ordinary Least Squares (OLS) regression, including ridge regression, lasso regression, and elastic net, can be applied to obtain the optimum regression models. All of the approaches aim to shrink the regression coefficients: ridge regression shrinks the parameters close to zero, while lasso regression shrinks them exactly to zero. Therefore, ridge regression may prevent multicollinearity and reduce the model complexity, while lasso regression can select the most significant variables and reduce the model complexity. In other words, ridge regression works well when most parameters affect the result, while lasso works well when only a few parameters are related to the response. Elastic net combines the penalty approaches of both the lasso and ridge regression.

2.2. Machine Learning-Based Methods and Simulation Approaches for Crop Yields Prediction

The process of variable selection is also a machine learning (ML)-based method. ML methods are widely employed to assess and forecast the impacts of climate signals on crop yields. Ref. [15] use the random forest approach to explore how climate and other variables affect crop yields and the relationship between crop yields and scale-compatible climate data from 1962 to 2014 in sub-Saharan Africa. This model captures the negative response of crop yields to increasing maximum temperature and low precipitation. Ref. [16] adopt the lasso regression model with three ML approaches that include support-vector machines, random forest, and neural networks to build various empirical models for forecasting wheat yields from 2000 to 2014 in Australia. In general, they find that precipitation and wet day frequency have a positive correlation with crop yields, while maximum temperature and potential evapotranspiration have a negative correlation.

The empirical approach usually assumes that climate does not have a trend, so the estimation and forecasting process may disregard the sensitivity of crop yields to technology and climate trends. Ref. [17] demonstrate that statistical approaches tend to overestimate yield losses from warming, especially in dry years, while the Crop Environment Resource Synthesis Maize (CERES-Maize) model (a simulation method) captures the interaction effects of warming and moisture. The simulation model simulates dynamic weather factors to establish a mathematical model to predict crop yields under specific conditions. At present, representative crop simulation models include CERES, used in the US; Agricultural Production Systems Simulator (APSIM), employed in Australia; and World Food Studies (WOFOST), developed in the Netherlands ([18]).

2.3. Extreme Weather Events for Crop Yields Prediction

Except for typical temperature and precipitation variables ([3,14,15,19,20]), the PDSI is another weather variable commonly considered in many pieces of research ([1,10,21]). PDSI subsumes the effects of both precipitation and temperature and provides a locally relative scale ranging from very wet to very dry conditions. Ref. [1] use positive PDSI

values to represent wet spells and negative PDSI values to represent drought conditions to assess whether longer time series weather data can provide a more accurate weather index for capturing the impacts of weather events on crop losses. Ref. [21] develops a summer average PDSI index (June, July, and August) and employs it to monitor significant weather events affecting crop yield growth during the critical growing season. While other weather variables (such as temperature, precipitation, various degrees of integration or disaggregation levels, etc.) may have been used, the analysis does not suggest that using other PDSI index types, months or month combinations, or direct temperature and precipitation measures have a qualitative impact on the outcomes.

The normalized difference vegetation index (NDVI) and crop moisture index (CMI) are also used to represent drought conditions. NDVI is a satellite-derived remote sensing measure that can estimate the health of vegetation during any given period by measuring greenness, which is thought to be closely correlated to crop yields. NDVI is easy to measure on a regular basis and is not subject to manipulation by agricultural producers or insurers. Most research focuses on the applicability of NDVI in index-based crop insurance, not indemnity-based crop insurance. Ref. [22] investigate whether NDVI can be a reliable indicator in developing index-based agricultural insurance products and conclude there is a wide variation in relationships between crop yields and NDVI depending on locations and crop types. Ref. [23] show that while NDVI, as a factor of catastrophic drought index insurance in Zimbabwe, has a relatively high correlation with corn and cotton yield loss, the relationship between NDVI and crop yield varies across crops and districts. The CMI indicates general conditions and not local variations caused by isolated rain. It is the sum of the evapotranspiration anomaly (which is generally negative or slightly positive) and excess moisture (either zero or positive) and is developed from some of the moisture-accounting procedures used in computing the PDSI. The CMI gives the short-term or current status of purely agriculture drought or moisture surplus and can change rapidly from week to week.

Cooling degree days (CDD), growing degree days (GDD), and diurnal temperature range (DTR) are also widely considered to assess the impacts of temperature or extreme heat ([1,10,14,22,24–26]). CDD refers to the degrees that a day's average temperature is above 65 degrees Fahrenheit, and GDD represents the mean daily temperature above a certain temperature during the growing season. The DTR means the range of temperature between the high of the day and the cool of the night. Ref. [24] developed a weather index using whole-season CDD (from May to September) and June–July total CDD. The reason for using the June–July periods is that crop growth is frequently adversely affected by heat units. Ref. [22] use the GDD standard of 80 degrees Fahrenheit to represent the instances of extreme heat. Ref. [24] calculate the GDD regarding a base of 46.4 degrees Fahrenheit and a ceiling of 89.6 degrees Fahrenheit to measure the impacts of temperature on agricultural productivity. Ref. [14] calculate the monthly means of DTR and then average them across the growing season. They then estimate the impacts of precipitation, temperature, and DTR on crop yields on a county basis through time-series regression models and conclude that, compared to temperature, DTR has had relatively small negative impacts on corn, wheat, and soy in the Pampas since 1971. Ref. [26] demonstrate that a greater DTR indicates a higher warm temperature and a lower cool temperature in a day, both of which have negative effects on crop yields.

2.4. Other Variables Affect Crop Yields

Technology is one of the main reasons for growing yields for many crop types in many countries. Technology contributes positively to long-term crop yields by making the crops more resistant to extreme weather events, such as drought and strong wind ([27]). Ref. [15] use time and country as fixed effect variables to represent technology changes and conclude that technological advances especially explained the increasing yields in sub-Saharan Africa from 1962 to 2014. Ref. [28] demonstrate that under similar drought conditions in 1988 and 2012 in 12 states in the Corn Belt, the yield losses, as well as loss–cost ratios and loss ratios that reflect crop insurance risk, were significantly lower in the second

period. In 2012, 91% of corn planted in Iowa was a genetically engineered variety that is more resistant to heat and moisture stresses, but no genetically engineered variety was available in 1988.

In addition, other factors, including soil quality, the management skill of farmers, government policy, and solar radiation are considerable consequences of crop insurance ([10,13,21,29,30]). Ref. [10] uses soil productivity as an independent variable to control farm heterogeneity through time and shows that better soil results in higher yields under the same weather and technology conditions. Ref. [13] document that failing to account for the spatial dependence of each dataset would lead to the erroneous estimation of some coefficients. The government also plays a vital role in responding to unusual and unexpected natural disasters ([29]). In particular, if the government encourages farmers to plant a specific crop with agricultural subsidies, the yields of this crop in that period were found to be higher than normal. Finally, Ref. [30] find that solar radiation is correlated to the increase in corn yields in the US Corn Belt from 1984 to 2013.

2.5. Extreme Weather Events and Insurance Applications

Extreme weather incidents result in huge losses, including those from crop insurance, for insurance and reinsurance firms. Many large payments caused by extreme events depend on the risk loading included in the premium since the risk loading takes into account the likelihood and severity of extreme events ([31]). Measuring the risk of extreme events provides more information to actuaries for climate insurance modeling. Ref. [32] use simulated Cooling Degree Day to build up a weather call option and measure the tails of the payments of the call option using conditional tail expectation (CTE) and Value-at-Risk (VaR). They then utilize a number of simulation-based methods to forecast the distributions of future daily temperature and insurance payments in California. The statistical ensemble models are essential to simulate and predict future indemnities or costs for insurance products so that actuaries could develop a better and more reasonable premium rate.

3. Background of the Actuarial Climate Index

The ACI has six components, including warm temperatures (T90: based on 90th percentile of both daily maximum and minimum temperature), cool temperatures (T10: based on 10th percentile of both daily maximum and minimum temperature), precipitation (P), drought (D), wind power (WP), and sea-level changes (S). In practice, the ACI employs the maximum number of consecutive dry days in a year (CDD) and the maximum consecutive 5-day precipitation amount in a month (Rx5day) to represent drought and precipitation, respectively. All components are presented as standardized anomalies and summarized on a monthly and seasonal basis. The standardized anomalies are determined by taking the difference between the component value in the current period and its mean value over the 1961–1990 reference period and dividing it by the standard deviation of the reference period. The ACI is the average of the standardized anomalies of six components (T10 is subtracted in the index) and thus focuses on extreme weather conditions.

The ACI is closely related to insurance rate-making and risk management. Insurance companies develop the premium for a given individual based on the information of the (homogeneous) groups or regions expected to have the same costs as the individual. Actuaries generally group individuals with the same anticipated costs together as one block and then derive a universal product premium for them. Risk grouping is an important step to utilize statistical methods more effectively and efficiently (On Risk Classification, 2011 [33]). Statistical models are typically used to predict the potential costs of providing coverage and other necessary additional expenses for these homogeneous blocks of individuals. The ACI data may provide reliable information on the frequency and severity of extreme weather events and could serve as useful variables in these statistical models.

Reinsurance is a type of insurance for insurance companies. Extreme climate events represent a possible failure for catastrophe reinsurance companies. By far, North America has the largest proportion of the reinsurance premiums capital ([34]). In the past

30 to 40 years, the number of high-cost weather events in the US has increased by four times. In 2011 alone, the US experienced fourteen extreme climate-related events, each resulting in more than a \$1 billion loss (Climate Science Watch). The occurrence of unusual extreme climate events and their potentially large losses are of concern to reinsurance companies ([35]). Based on measurements from meteorological and coastal tide stations in the United States and Canada, the ACI is generated from observations of climate changes in temperature, sea level, precipitation, and wind speed. It is hypothesized that the ACI may improve the accuracy of estimating the frequency of rare events and contribute to building risk models that could be used by reinsurance companies in future business decisions.

The Actuaries Climate Risk Index (ACRI) is under development to represent the relationships between economic losses and climate variables using regression analysis. The intent is to develop several ACRI that can be used for a wide range of fields related to climate risk ([36]). However, the ACRI has not been developed for agriculture applications as of the writing of this paper. In fact, the results of this research could provide some insights into an ACRI applicable to assessing the climate risks in crop yields.

4. Data

This section introduces the crop data, the actuarial climate index, the data used to reconstruct the ACI, and the high-resolution climate data.

4.1. Crop Yields Data

One of the 12 subregions identified by the Actuaries Climate Index is the Midwest (MID) US, which contains eight states: Illinois, Indiana, Iowa, Michigan, Minnesota, Missouri, Ohio, and Wisconsin. The Midwest region is the region of focus for the empirical analysis in this research. A set of state-level data for yields (bushel/acre) and areas planted (acres) for corn covering a 58-year (from 1961 to 2018) historical period in the eight states in the Midwest region is gathered from the United States Department of Agriculture (USDA)'s National Agricultural Statistical Service (NASS). We also calculate the Midwest-level data by taking the weighted average of the eight state-level yield series. The corn yields are matched to the Midwest monthly ACI dataset for each year from 1961 to 2018. The trend and distribution of the crop yields data are shown in Figures A1–A5 in the Appendix A.

4.2. The Actuaries Climate Index Data

The standardized ACI data over the same period were obtained from the Actuaries Climate Index official website (<https://actuariesclimateindex.org/data/>, accessed on 31 May 2022). The data include: (1) the monthly and annual ACI indices and (2) the five individual components of the ACI (there are no sea-level change records for the Midwest region). The individual components are available at a monthly frequency for the whole Midwest region, but only at an annual frequency for the individual states.

We selected the variables that are included in the statistical models based on the growing season of corn (from April to September). The combined ACI variables selected in this paper are ACI_4, ACI_5, ACI_6, ACI_7, ACI_8, and ACI_9. In addition, the individual components of the ACI are June to July CDD (CDD_6 and CDD_7), May to August Rx5Days (Rx5Day_5, Rx5Day_6, Rx5Day_7, and Rx5Day_8), April, May and September T10 (T10_4, T10_5, and T10_9), June to August T90 (T90_6, T90_7, and T90_8), and June WP90 (WP90_6). Summer is the growing season for corn, so most of the variables utilized in this research are summer variables. The reason for adopting the June wind speed is that tornados usually happen in June in the Midwest. Frozen nights often take place in the spring and autumn, so low-temperature variables in April, May, and September were selected.

4.3. The High-Level Resolution Climate Data

The ACI data is composed of grid-level data, where each grid is 2.5 degrees longitude (about 278.3 km) by 2.5 degrees latitude (about 277.5 km). Based on this, each grid is developed with state and region-level data. The grid level scale is relatively large since

station-level data is commonly used in climate-related research, and the concern is that large-scale data might eliminate extreme local observations and lead to a misestimation of results. In order to investigate the impact of using more high-resolution data on crop yield prediction, we employ a high-level spatial resolution global climate dataset called TerraClimate in this research ([37]). This dataset is based on an approximate 4 km by 4 km scale and contains monthly climate data, including historical weather data for global land surfaces, such as precipitation, PDSI, temperature, vapor pressure, etc. Variables similar to those used to construct the ACI index are selected, including PDSI, precipitation, minimum and maximum temperature, and soil moisture.

5. Statistical Models

This paper applies the linear regression and probit regression models as the two main statistical methods to estimate corn yields in the Midwest region of the US. Linear regression is widely used in estimation and prediction, measuring correlation and model fit.

In insurance pricing, the total expected claim amount is typically composed of the severity and frequency of the claims. A simple decomposition could be displayed by a continuous variable and a binary variable (such as the response variables discussed in Section 5.2). The binary variable could illustrate whether a claim happened, and the continuous variable could indicate the amount of a claim. GLM is a widely used method to estimate the expectation of the binary variable ([38,39]). In particular, logit and probit models are typical models utilized in insurance pricing and are commonly used to estimate and predict the possibility of claims ([40]). When the number of failures (the frequency of claims; also, the number of claims) is far less than the occurrence of success (the number of zeros in the response variable), and the independent variables include continuous variables, the probit model is shown to be preferred over the logit model ([41]).

Five crop yield models were designed based on both linear and probit regression models, using different scale-level corn yields as the dependent variable and weather variables derived from the ACI and TerraClimate data as the independent variables. The details and results for each model will be discussed in Section 6.

5.1. Linear Regression Model

Linear regression is the first model applied, and is expressed as follows:

$$Y_{ik} = \beta_{0k} + \beta_j X_{jik} + \gamma t + \epsilon_{ik}, \quad (1)$$

where Y_{ik} represents the weighted average corn yields of the whole Midwest (Models 1 and 2) for year i ($i = 1961, \dots, 2018$) or corn yields of state k in the Midwest (Model 3) for year i ($i = 1961, \dots, 2016$), X_{jik} is the group of j explanatory variables described in Section 3 for year i and state k , and β_j is the estimated coefficient of each explanatory variable. In addition, β_{0k} is the intercept for state k , t is the time effect, and ϵ_{ik} represents the error terms following a normal distribution with $(0, \sigma^2)$.

Before estimating the models, we perform a KPSS test on the average corn yields of the Midwest region and the state-level corn yields. The results are shown in Table A7 in the Appendix A. We see that all corn yield series are non-stationary. Therefore, the time effect is included in the model specification (1) to capture this trend. For the probit model, the response variables (corn yields) are detrended and are therefore stationary. Hence, no time variable is included in the regressions.

5.2. Generalized Linear Model—Probit Regression Model

The second model is probit regression, which estimates corn yield losses with the selected variables. Since the corn yields have an increasing trend overall, detrending is necessary to prepare the data and ensure a more accurate and representative result. After detrending, we transform the corn yields into binary variables in the following two steps: Step 1: Compute the 25th percentile of the detrended corn yields;

Step 2: Set the yields lower than the 25th percentile as “1”, which means there is a loss, and the part above the 25th percentile data as “0”, which means there is no loss.

Increased temperatures, drought conditions, and extreme weather disasters (EWDs) could reduce crop yields. For instance, corn yields decrease by about 10% with each degree Celsius increase in the US ([42]), and EWDs would produce an average of 19.9% reduction in national cereal production in North America ([5]). However, choosing the 25th percentile of detrended corn yields as the threshold of losses is a consideration for sample effectiveness and efficiency. A 10th percentile threshold would make the number of losses too small to generate a convincing result, particularly for Model 1, which includes only 58 observations. In addition, probit regression focuses on demonstrating that different resolution databases might produce inconsistent results for estimating crop yield losses.

The linear regression model is written as follows:

$$Y_{ik} = \beta_{0k} + \beta_j X_{jik} + u \quad (2)$$

The probit regression model is:

$$P(Y_{ik} = 1 | X_{jik}) = \Phi(\cdot) \quad (3)$$

where Y_{ik} represents the weighted average corn yields of the whole Midwest (Models 1 and 2) for year I ($i = 1961, \dots, 2018$) or corn yields of state k in the Midwest (Model 3) for the year i ($i = 1961, \dots, 2016$), X_{jik} is the group of j explanatory variables described in Section 3 for year i and state k , and β_j is the estimated coefficient of each explanatory variable. Moreover, β_{0k} is the intercept for state k , and $\Phi(\cdot)$ is the cumulative standard normal distribution function.

6. Empirical Analysis

In this section, we applied the five model specifications for crop yield prediction. Model 1 and Model 2 use the region-level annual weighted average corn yields as dependent variables and monthly combined ACI and individual components of the ACI as independent variables, respectively. Model 3 employs state-level annual corn yields and annual individual components of the ACI as dependent and independent variables. Model 4 utilizes the county-level annual corn yields in Iowa as dependent variables and monthly standardized high-level spatial resolution weather variables as independent variables. Model 5 is the same as Model 4, but the research area expands to the whole Midwest region.

6.1. Midwest Analysis with the Combined Actuaries Climate Index

In this sub-section, the effectiveness of the individual combined ACI variables is assessed for estimating the corn yields in the whole Midwest region. The linear regression results are shown in Table 1 column 2, which has an R^2 value of 0.8917 and two significant variables, July and September ACI. The coefficient of the July ACI illustrates that extreme weather conditions may have negative effects on corn yields. The coefficient of the September ACI is positive. One possible reason is that high daytime warm temperatures might offset the adverse effects of cold nights in September. The coefficients of ACI_4, ACI_5, ACI_6, and ACI_8 are not significant. Column 3 shows the linear regression results without the insignificant variables, dropping variables starting from the one with the highest p -value, step by step. All remaining variables are significant, and the R^2 value is as high as 0.8901. The fit of the linear regression model implies that changes in ACI variables explain about 89% of the changes in corn yields across the Midwest. Overall, the model is significant, as shown by the significance of the F-test statistic at a 0.1% significance level.

Column 4 is the results of probit regression and contains two significant variables as well. The pseudo R^2 value of 0.1617 is calculated by the formula:

$$1 - \frac{\loglikelihood (full\ model)}{\loglikelihood (null\ model)} \quad (4)$$

which is also known as McFadden R^2 . The lower pseudo R^2 might imply that the combined ACI might have fewer insights when modeling corn yield losses since a pseudo R^2 value between 0.2 and 0.4 could represent a very good fit.

Table 1. Midwest analysis with combined ACI variables.

Variables	Linear Regression	Linear Regression	Probit Regression
Intercept	66.8829 ***	67.15695 ***	−0.76085
ACI_4	3.5330		−0.44542
ACI_5	−1.5629		0.09833
ACI_6	−3.2987		0.80559
ACI_7	−6.9927 *	−6.24302 †	0.76382 †
ACI_8	−5.6327	−6.16164 †	0.55087
ACI_9	8.2069 *	7.80479 *	−1.11165 *
time	1.8737 ***	1.86356 ***	
N	58	58	58
R^2 (Pseudo R^2)	0.8917	0.8901	0.1617
AIC			69.582
F-test statistic	58.82 ***	84.22 ***	

Note: The outcome variable is the annual weighted average corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. The second column is the result of the linear regression model. Column 3 is the result of the linear regression with fewer variables. Column 4 is the result of the probit regression model. “†”, “*”, and “***” indicate significance at 10%, 5%, and 0.1% levels, respectively.

Midwest Analysis with Individual Component Variables of the Actuaries Climate Index

The results of the linear regression model with individual components of the ACI are shown in Table 2. The R^2 value is as high as 0.9548 and the F-test statistic is high enough to imply that all explanatory variables together significantly explain corn yields. The R^2 values are comparable to those reported in existing studies, such as [43], indicating reasonable goodness of fit. Furthermore, the R^2 values using the individual components are higher than using the original ACI variables. This indicates that relaxing the predetermined weights of the variables in the ACI construction could improve the predictive accuracy of corn yields.

Four individual variables are statistically significant, and one of them (T10_9) is significant at a 10% level. Using the Farrar–Glauder test to diagnose multicollinearity among variables, the results show that CDD_6 and CDD_7 are highly collinear variables. Therefore, the next step is to remove CDD_6 and CDD_7 from the linear regression. The results shown in Table 2 column 3 are similar to the first model. Drop variables with the highest p -value step by step to get the simplest model with variables that are all significant. The final simplest model results in an R^2 of 0.9535 and five significant variables, shown in Table 2, column 4.

The backward stepwise selection method is employed as a comparison to check the accuracy of the manual stepwise selection results. The results shown in Table A4 column 2 have the same estimated coefficients and significance levels as the manual selection results. Variable changes during each dropping step of the manual selection method could be observed, such as the estimated coefficients, level of significance, and the value of R^2 and pseudo R^2 . This paper continues to use the manual selection method in the following analysis but provides the results from backward stepwise selection for comparison and reference in the Appendix A.

The results of the probit regression model with individual component ACI variables are shown in Table 3. We use the same variable selection approaches as for the linear regressions to obtain the simplest model, and the simplest model is shown in column 4. Only three significant variables are seen in the simplest probit model, and two of them, T10_5 and T10_9 (May and September minimum temperature), are significant at a 10% level. This result is not as good as the results from the linear regression model. However, the pseudo R^2 value of the probit model indicates a good fitness since it is close

to 0.4. Furthermore, even though the pseudo R^2 is gradually decreasing, the AIC value and p -value of the Chi-square test are both declining as well. It might demonstrate that the simplest model, as a whole, has a better fit compared to the original probit model (ANOVA chi-square test is based on the current model and the null model). The backward stepwise selection results are the same as the simplest model and are shown in Table A4, column 3.

Table 2. Midwest analysis with individual component variables, linear regression model.

Variables	Linear Regression	Linear Regression without CDD	Simplest Linear Regression
Intercept	67.11794 ***	67.17533 ***	67.32980 ***
CDD_6	2.14337		
CDD_7	−1.86621		
Rx5Day_5	0.41072	0.40784	
Rx5Day_6	−0.46413	−0.47983	
Rx5Day_7	1.98230	1.98341	1.98864 †
Rx5Day_8	2.86024 *	2.83325 *	2.79095 *
T10_4	−0.45676	−0.44421	
T10_5	0.11886	0.01343	
T10_9	−2.54756 †	−2.61742 †	−2.85146 *
T90_6	0.72151	0.66719	
T90_7	−5.77264 ***	−5.78398 ***	−5.45714 ***
T90_8	−6.13024 ***	−6.09450 ***	−5.76785 ***
WP90_6	−0.62446	−0.57842	
Time	1.84387 ***	1.83969 ***	1.83177 ***
N	58	58	58
R ²	0.9548	0.9547	0.9535
AIC	425.5799	421.6728	411.2117
F-test statistic	64.82 ***	79.01 ***	174.2 ***

Note: The outcome variable is the annual weighted average corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. The second column is the result of the original linear regression model. Column 3 is the result of the linear regression model without CDD variables. Column 4 is the result of the linear regression model that removed all insignificant variables. “†”, “*”, and “***” indicate significance at 10%, 5%, and 0.1% levels, respectively.

Table 3. Midwest analysis with individual component variables, probit regression model.

Variables	Probit Regression	Probit Regression without CDD	Simplest Probit Regression
Intercept	−0.83975 **	−0.84186 **	−0.7595 ***
CDD_6	−0.52693		
CDD_7	0.66512		
Rx5Day_5	0.12407	0.10113	
Rx5Day_6	0.09689	0.08025	
Rx5Day_7	−0.22456	−0.21207	
Rx5Day_8	−0.08674	−0.11425	
T10_4	0.21453	0.18391	
T10_5	0.35710	0.35282	0.4222 †
T10_9	0.42315	0.39679	0.4200 †
T90_6	0.15687	0.18821	
T90_7	0.75301 *	0.76816 *	0.7353 **
T90_8	0.48152	0.48877	0.4278
WP90_6	0.24637	0.22502	
N	58	58	58
Pseudo R ²	0.4015	0.3973	0.3451
AIC	67.686	63.963	53.426
ANOVA Chi-square test (p-value)	0.01402 *	0.005767 **	0.0001338 ***

Note: The outcome variable is a binary variable transformed from the annual weighted average corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. Column 2 is the result of the original probit regression model. Column 3 is the result of the probit regression model without CDD variables. Column 4 is the result of the probit regression model that removed almost all insignificant variables. “†”, “*”, “**”, and “***” indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

6.2. Midwest Analysis with Annual State-Level Data

The results in the previous section show that some variables are not significant in predicting corn yields. One potential reason is that a relatively large spatial scale dataset is used and thus some extreme records may be averaged in some places. To further investigate this issue, the following analysis employs the Midwest state-level raw data of the ACI.

The ACI's state-level data is a yearly dataset rather than monthly data and only contains four independent variables, Rx5days, Tn10, Tx90, and CDD, from 1961 to 2016. Similarly, we use linear regression and probit regression models to estimate the impacts of extreme weather on the Midwest state-level corn yields. State-level corn yields are also detrended and transformed to follow a binomial distribution to prepare for the probit regression. Among the four variables, Tn10 and Tx90 represent the percentage of days when the daily minimum temperature was lower than the 10th percentile of the base period (1961–1990) and the maximum temperature was higher than the 90th percentile of the base period (1961–1990), respectively. These definitions are slightly different from the terms of T10 and T90.

Before running the regression, we standardize the state-level raw data. Specifically, 30 years (from 1961 to 1990) are selected as the reference period, and then the two steps below are followed (using CDD as an example):

Step 1: Calculate the mean and standard deviation of CDD in the reference period 1961 to 1990, written as μ_f and σ_f ;

Step 2: Calculate the standardized CDD for year i in each state k within the Midwest from 1961 to 2016, using the formula $\frac{CDD_{(i,k)} - \mu_f}{\sigma_f}$.

The results in Table 4, columns 2 and 3, show that all explanatory variables together might be meaningful evidence since the F-test statistics are significant. Two out of the four variables are significant at least a 1% level of significance. The R^2 value of the linear regression model is 0.7504, and the pseudo R^2 value of the probit regression model is 0.0739. After adding a state effect variable to the linear regression model, the R^2 value grows to 0.7607, and the significance of the state variable might imply that locations are highly correlated to crop yields. However, all of these R^2 s are still smaller than the results in Section 6.1. This result might be due to the fact that yearly data is an averaged evidence of the entire year and ignores the monthly or daily extrema. In addition, the state-level might still be a large spatial scale and may eliminate local extrema.

Table 4. Midwest analysis with annual state-level individual component variables.

Variables	Linear Regression	Linear Regression	Probit Regression
Intercept	67.87581 ***	76.29208 ***	−0.62538 ***
Rx5Day	0.10005	−0.63556	0.06065
Tn10	−2.32393 *	−2.51860 *	0.09919 †
Tx90	−5.73952 ***	−5.80187 ***	0.29302 ***
CDD	1.47515	0.77219	−0.06731
Time	1.67055 ***	1.66590 ***	
State		−0.29164 ***	
N	448	448	448
R^2 (Pseudo R^2)	0.7504	0.7607	0.0739
AIC	3809.78	3792.93	492.89
F-test statistic	265.8 ***	233.7 ***	

Note: The outcome variable in columns 2 and 3 is state-level corn yields and in column 4 is a binary variable transformed from the state-level corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. Column 2 is the result of the linear regression model. Column 3 is the result of the linear regression model added with a state effect variable. Column 4 is the result of the probit regression model. "†", "*", and "***" indicate significance at 10%, 5%, and 0.1% levels, respectively.

6.3. Iowa Analysis with Standardized High-Level Resolution Climate Data

In this subsection, we use the TerraClimate dataset, which is generated based on approximate 4 km by 4 km grid-level data and contains monthly county-level historical weather data to predict crop yields. Especially, June to July PDSI (pdsi_6 and pdsi_7), May to August precipitation (pr_5, pr_6, pr_7, and pr_8), June to July soil moisture (soil_6 and soil_7), June to August maximum temperature (tmmx_6, tmmx_7, and tmmx_8), and April, May, and September minimum temperature (tmmn_4, tmmn_5, and tmmn_9) data from 1961 to 2018 in Iowa are selected based on the corn-growing season to estimate the corresponding county-level yearly corn yields.

We then replicate the ACI development approach by selecting a 30-year period from 1961 to 1990 as the reference period and standardizing Iowa's high-level resolution climate data in the following steps (using PDSI as an example):

Step 1: Calculate the mean and standard deviation of PDSI for each month m in the reference period 1961 to 1990, written as $\mu_f(m)$ and $\sigma_f(m)$;

Step 2: Calculate the standardized PDSI for month m of year i in each county k within Iowa from 1961 to 2018, using the formula $\frac{pdsi_{(m, i, k)} - \mu_f(m)}{\sigma_f(m)}$.

Same as the estimation uses ACI data, the following linear regression is applied:

$$Y_{ik} = \beta_{0k} + \beta_j X_{jik} + \gamma t + \epsilon_{ik} \quad (5)$$

where Y_{ik} represents the corn yields of county k in Iowa for year i ($i = 1961, \dots, 2018$), X_{jik} is the group of j explanatory variables described above for year i and county k , and β_j is the estimated coefficient of each explanatory variable. β_{0k} is the intercept of county k , t is the time effect, and ϵ_{ik} represents the error terms following a normal distribution with $(0, \sigma^2)$.

The results are shown in Table 5, column 2. The value of R^2 is 0.7982, indicating that this model with standardized high-level resolution climate data is appropriate. Further, except for two insignificant variables, July PDSI and June precipitation, all other variables are statistically significant. In addition, the F-test statistic is high enough to imply that all explanatory variables together significantly explain corn yields. Moreover, we remove pr_6 and pdsi_7 in the next two steps to get better-fit models as a result (shown in Table 5, columns 3 and 4), the significance level of the estimated coefficients is the same as the second column, and the R^2 value of 0.7982 does not change. Compared to the results in Section 6.1, this simplest linear regression model contains more effective coefficients. The backward stepwise selection results in the same as the simplest model and is shown in Table A6, column 2.

Table 5. Iowa analysis with standardized high-level resolution climate data using the linear regression model.

Variables	Linear Regression	Linear Regression without pr_6	Simplest Linear Regression
Intercept	65.01353 ***	65.01403 ***	65.01863 ***
pdsi_6	3.07422 *	3.06977 *	2.68789 ***
pdsi_7	−0.46299	−0.45688	
pr_5	−2.78124 ***	−2.77376 ***	−2.77632 ***
pr_6	−0.02197		
pr_7	4.14047 ***	4.13523 ***	4.03047 ***
pr_8	−2.83129 ***	−2.83299 ***	−2.83344 ***
tmmn_4	−0.61862 *	−0.62150 *	−0.61710 *
tmmn_5	−0.54592 †	−0.56722 †	−0.56257 †
tmmn_9	5.08470 ***	5.08345 ***	5.08391 ***
tmmx_6	0.60512 †	0.60769 †	0.60951 †
tmmx_7	−5.01798 ***	−5.01431 ***	−5.00016 ***

Table 5. Cont.

Variables	Linear Regression	Linear Regression without pr_6	Simplest Linear Regression
tmmx_8	−9.61604 ***	−9.61887 ***	−9.62043 ***
soil_6	3.12003 ***	3.09629 ***	3.04412 ***
soil_7	−8.29489 ***	−8.28874 ***	−8.29728 ***
Time	1.89845 ***	1.89844 ***	1.89820 ***
N	5742	5742	5742
R ²	0.7982	0.7982	0.7982
AIC	49,197.69	49,195.69	49,193.75
F-test statistic	1506 ***	1614 ***	1738 ***

Note: The outcome variable is the annual county-level corn yields of the farms in Iowa reported to the USDA. Descriptions of other variables are in Table A1. Column 2 is the result of the linear regression model. Column 3 is the result of the linear regression model without June precipitation. Column 4 is the result of the same linear regression model without another insignificant variable, July PDSI. “†”, “*”, and “***” indicate significance at 10%, 5%, and 0.1% levels, respectively.

The results of the probit regression model using standardized high-level resolution climate data in Iowa are shown in Table 6. Column 3 is the probit regression model removing the most insignificant variable, June soil moisture. We then apply the same approach in Section 6.1 to drop other insignificant variables, June precipitation, June maximum temperature, April minimum temperature, and July PDSI, step by step, to get the simplest probit model and display it in column 4. All remaining variables are statistically significant at a 0.1% level of significance, except for the May minimum temperature. The pseudo R² values of the original probit model and simplest probit model, 0.2867 and 0.2864, are very close and could represent that both models are a good fit.

Table 6. Iowa analysis with standardized high-level resolution climate data using the probit regression model.

Variables	Probit Regression	Probit Regression without soil_6	Simplest Probit Regression
Intercept	−0.837520 ***	−0.83781 ***	−0.83610 ***
pdsi_6	−0.192399	−0.18535	−0.25164 ***
pdsi_7	−0.060811	−0.07146	
pr_5	0.265546 ***	0.26434 ***	0.26529 ***
pr_6	−0.009730	−0.01235	
pr_7	−0.214757 ***	−0.20945 ***	−0.21372 ***
pr_8	0.213755 ***	0.21389 ***	0.20966 ***
tmmn_4	0.031181	0.03109	
tmmn_5	0.043589 †	0.04375 †	0.05174 *
tmmn_9	−0.239507 ***	−0.23952 ***	−0.23903 ***
tmmx_6	−0.020039	−0.01969	
tmmx_7	0.362900 ***	0.36229 ***	0.37875 ***
tmmx_8	0.602808 ***	0.60295 ***	0.59262 ***
soil_6	−0.009438		
soil_7	0.395052 ***	0.39028 ***	0.38068 ***
N	5742	5742	5742
Pseudo R ²	0.2867	0.2867	0.2864
AIC	4785.19	4783.22	4777.72
ANOVA Chi-square test (p-value)		0.8769	0.7723

Note: The outcome variable is a binary variable transformed from the annual county-level corn yields of the farms in Iowa reported to the USDA. Descriptions of other variables are in Table A1. Column 2 is the result of the original probit regression model. Column 3 is the result of the probit regression model without June soil moisture. Column 4 results from the probit regression model without all other insignificant variables, July PDSI, June precipitation, and April minimum and June maximum temperature. “†”, “*”, and “***” indicate significance at 10%, 5%, and 0.1% levels, respectively.

Furthermore, from column 2 to column 4, the AIC value has a tiny decrease, and the p -value of the ANOVA Chi-square test is decreased but still insignificant. This demonstrates that the final simplest probit model could slightly improve the goodness of fit compared to the base model (ANOVA chi-square test is based on the current model and the original probit regression model). The backward stepwise selection results in the same as the simplest model and is shown in Table A6, column 3.

6.4. Midwest Analysis with Standardized High-Level Resolution Climate Data

We now expand the investigated area to the whole Midwest region and apply the ACI development approach to the Midwest high-resolution climate data. Similarly, we use the same self-selected variables in the linear regression model. Table 7 column 2 shows that all variables are statistically significant at a 0.1% level except for June soil moisture. The R^2 value of 0.7126 is acceptable, although it is smaller than the R^2 values of the models using ACI data. It might imply that changes in high-level resolution weather variables could explain about 71.26% of the corn yield changes across the Midwest. The F-test statistic is significant and demonstrates that all these variables together fit and explain corn yields well.

Table 7. Midwest analysis with standardized high-level resolution climate data, the linear regression model.

Variables	Linear Regression	Linear Regression (Lasso Selection)	Simplest Linear Regression (Lasso Selection)
Intercept	62.41806 ***	63.696265 ***	63.593053 ***
pdsi_5		−7.358564 ***	−7.356485 ***
pdsi_6	4.87800 ***	14.477384 ***	14.471175 ***
pdsi_7	−3.88711 ***	−6.178563 ***	−6.178364 ***
pr_5	1.26769 ***	0.676166 ***	0.663937 ***
pr_6	5.45665 ***	2.114477 ***	2.120822 ***
pr_7	9.70496 ***	8.912569 ***	8.913359 ***
pr_8	0.49006 ***	−0.733426 ***	−0.730487 ***
pr_9		1.170014 ***	1.169836 ***
tmmn_4	1.80134 ***	−0.418122 *	−0.425323 *
tmmn_5	4.53892 ***	−0.097777	
tmmn_6		4.952973 ***	4.916094 ***
tmmn_7		6.389840 ***	6.375701 ***
tmmn_8		7.222588 ***	7.210296 ***
tmmn_9	2.99060 ***	−4.724990 ***	−4.729296 ***
tmmx_5		4.479994 ***	4.394181 ***
tmmx_6	7.97478 ***	4.156479 ***	4.189060 ***
tmmx_7	−7.27719 ***	−11.281069 ***	−11.271298 ***
tmmx_8	−9.84868 ***	−17.532366 ***	−17.524793 ***
tmmx_9		7.555767 ***	7.564631 ***
soil_6	0.52865	0.909000 **	0.910250 **
soil_7	−8.86023 ***	−8.340331 ***	−8.339541 ***
Time	1.67237 ***	1.566900 ***	1.566997 ***
N	39,567	39,567	39,567
R ²	0.7126	0.7297	0.7297
AIC	352,752.4	350,338.1	350,336.1
F-test statistic	6539 ***	4854 ***	5085 ***

Note: The outcome variable is the annual county-level corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. Column 2 is the result of the linear regression models with variables selected based on experience. Columns 3 and 4 result from the linear regression models with variables determined by both the Lasso approach and experience, and column 4 removes one insignificant variable, May minimum temperature. “*”, “**”, and “***” indicate significance at 5%, 1%, and 0.1% levels, respectively.

The Least Absolute Shrinkage and Selection Operator (Lasso) regression approach is introduced to select variables and explore whether some variables could be included or dropped and increase the effectiveness of the high-level resolution climate data for estimating corn yields. Excluding the variables that are not in the growing season (broadly, April to September), variables selected to be supplementary variables in the linear regression model are May PDSI (pdsi_5), September precipitation (pr_9), June to August minimum temperature (tmmn_6, tmmn_7, and tmmn_8), and May and September maximum temperature (tmmx_5 and tmmx_9). The results are in Table 7, columns 3 and 4. All variables are

statistically significant, except the May minimum temperature. After dropping it from the regression, all variables are significant. The R^2 value remains as 0.7297 and is slightly larger than 0.7126, indicating that all these variables together are fitted better than the regression with only subjectively selected variables. The F-test statistics and AIC values show the same conclusion. The backward stepwise selection method is applied to the linear regression model with the variables selected based on both the Lasso approach and experience. The results are the same as the simplest linear model and are shown in Table A7, column 2.

The results of the probit regression model with standardized high-level resolution climate data are shown in Table 8. Column 3 is the result of the simplest probit model without the insignificant variables, July PDSI and August precipitation. All remaining variables are statistically significant at a 0.1% level of significance, except for the June PDSI. The pseudo R^2 of the simplest model, 0.2341, is the same as the pseudo R^2 of the original probit regression and could represent a good fitness.

Table 8. Midwest analysis with standardized high-level resolution climate data, the probit regression model.

Variables	Probit Regression	Simplest Probit Regression	Probit Regression (Lasso Selection)	Simplest Probit Regression (Lasso Selection)
Intercept	−0.757158 ***	−0.756608 ***	−0.724359 ***	−0.726714 ***
pdsi_5			0.625632 ***	0.626712 ***
pdsi_6	−0.074283 *	−0.019064 †	−0.900663 ***	−0.906566 ***
pdsi_7	0.064830		0.267352 ***	0.269595 ***
pr_5	−0.107747 ***	−0.105575 ***	−0.064896 ***	−0.065205 ***
pr_6	−0.283505 ***	−0.280150 ***	−0.104547 ***	−0.111829 ***
pr_7	−0.455160 ***	−0.439158 ***	−0.445186 ***	−0.443344 ***
pr_8	0.001954		0.040559 ***	0.041241 ***
pr_9			−0.014300	
tmmn_4	−0.136050 ***	−0.137153 ***	−0.060383 ***	−0.063977 ***
tmmn_5	−0.152961 ***	−0.153493 ***	−0.051977 †	−0.066655 *
tmmn_6			−0.043355	
tmmn_7			−0.319458 ***	−0.328385 ***
tmmn_8			−0.226305 ***	−0.226689 ***
tmmn_9	−0.156244 ***	−0.155197 ***	0.051380 *	0.034804 †
tmmx_5			−0.086972 **	−0.075420 **
tmmx_6	−0.349582 ***	−0.352383 ***	−0.372841 ***	−0.411686 ***
tmmx_7	0.324255 ***	0.321780 ***	0.544415 ***	0.555092 ***
tmmx_8	0.480141 ***	0.479800 ***	0.711735 ***	0.708800 ***
tmmx_9			−0.151694 ***	−0.134287 ***
soil_6	0.067226 ***	0.073061 ***	0.064598 **	0.066103 **
soil_7	0.432060 ***	0.430947 ***	0.410868 ***	0.410685 ***
N	39,567	39,567	39,567	39,567
Pseudo R^2	0.2341	0.2341	0.2537	0.2536
AIC	35,294	35,293	34,407	34,406
ANOVA				
Chi-square test (p -value)		0.2529	$<2 \times 10^{-16}$ ***	$<2.2 \times 10^{-16}$ ***

Note: The outcome variable is a binary variable transformed from the annual county-level corn yields of the farms in the Midwest reported to the USDA. Descriptions of other variables are in Table A1. Columns 2 and 3 are the results of the probit regression models with the variables selected based on experience, and column 3 removes two insignificant variables, July PDSI and August precipitation. Columns 4 and 5 result from the probit regression models with variables selected based on both the Lasso approach and experience, and column 5 removes two insignificant variables, September precipitation and June minimum temperature. “†”, “*”, “**”, and “***” indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

Columns 4 and 5 are the results of probit regressions using the variables selected through the Lasso approach. September precipitation and June minimum temperature are removed due to insignificance, and the simplest model is shown in column 5. With more variables in the probit regression, the pseudo R^2 of these two models is moderately

improved to 0.2536. In addition, the AIC value and p -value of the ANOVA Chi-square test of the probit regression model with more variables (Lasso selection) are smaller than the values of the probit model only with self-selected variables and demonstrate that the probit model with more variables could better explain the effectiveness of the high-level resolution ACI for estimating corn yield losses (the ANOVA Chi-square test is based on the current model and the original probit regression). The backward stepwise selection method is applied to the original probit model and the probit model with the variables selected based on both the Lasso approach and experience. The results are the same as the simplest models and are shown in Table A7, columns 3 and 4.

6.5. Rolling Window Predictive Analysis

Prediction is critical in insurance pricing, as mentioned earlier, with which actuaries could develop and renew the premium of insurance products. To evaluate the effectiveness of the ACI and high-level resolution ACI for predicting crop yields, this paper employs rolling window regression and the Diebold and Mariano test to compare the predictive accuracy and forecast ability of the four linear regression models described in Section 5. Selecting a 30-year window width and running rolling window regression, we then compare each regression with a corresponding simple random walk model. The results are shown in Table 9, and p -values could reveal the predictive accuracy of the two methods in each model. Based on $\alpha = 0.05$, the p -values suggest that the two methods in Model 1 and Model 2 have the same forecast accuracy, while the linear regressions of Model 3 and Model 5 are more accurate than the simple random walk model, which might illustrate that the higher-level resolution ACI is better at predicting corn yields than the ACI.

Table 9. Predictive ability of different spatial scale ACIs.

	Model 1	Model 2	Model 3	Model 5
p -value	0.2616	0.0748	1.332×10^{-8}	$<2.2 \times 10^{-16}$

Model 1: Linear regression with the monthly combined ACI in the Midwest.

Model 2: Linear regression with monthly individual component variables of the ACI in the Midwest.

Model 3: Linear regression with annual state-level individual component variables of the ACI in the Midwest.

Model 5: Linear regression with monthly high-level (county-level) resolution ACI in the Midwest.

7. Conclusions and Discussion

This paper examines the effectiveness of the Actuaries Climate Index (ACI), which is a climate index published recently by multiple actuarial associations in North America, in corn yield prediction and yield (re)insurance ratemaking. The ACI is intended to help actuaries, insurance companies, governments, and the general public understand the potential effects of climate trends and extreme weather events better. We constructed five linear regression models and five probit regression models to examine the predictive power of the ACI and related climate variables on different geographical and temporal scales. The probit regressions are used to identify insurance claims, which occur when the corn yield falls below a certain threshold calculated from the historical yields. When an insurance claim occurs, the linear regression models will then be used to predict the corn yield in that year, based on which the insurance severity can be calculated.

The empirical results show that the ACI provides valuable information in predicting crop yields and yield losses, as the goodness of fit from both the linear and the probit regression models are satisfying. However, we also see that when the weights of the climate variables which constitute the ACI could be determined using a data-driven method rather than being predetermined, the predictive accuracy could be further improved.

Moreover, in order to investigate the predictive power of the ACI at higher resolution levels, we construct the ACI index at the county level using a dataset (TerraClimate dataset) consisting of 4 km by 4 km resolution climate data. We find that, although the predictive accuracy at the county level is slightly worse than that at the Midwest region level, possibly due to noisier data at the county level, the predictive accuracy is also reasonable. The county-level data are more relevant to insurance practices because real-life insurance products typically do not cover corn yields of as large a geographical location as a state or the whole Midwest region (which consists of eight states). Hence, the ACI could further benefit the insurance industry if it could be developed to be a higher-resolution index.

8. Future Research

There are several directions for future research. First, since the aim of this paper is to evaluate the effectiveness of the Actuaries Climate Index, only the ACIs and the related climate variables are included in the regression models. It would be interesting to include additional variables that capture the potential spatial–temporal correlations of corn yields in the regression models in future research. Second, more advanced statistical or machine learning models, such as distributional forest and neural networks, could be used to develop accurate predictive models of corn yields. Third, utilizing the established predictive models, it would be interesting to explore the possibility of constructing index-based corn yield insurance products, the payments of which are not determined by the actual loss of the insured farms but by the realization of an external index which cannot be affected by the behavior of the farmers (the insured).

Author Contributions: Data curation, Q.P.; formal analysis, Q.P., L.P. and H.L.; investigation, Q.P., L.P. and H.L.; methodology, Q.P., L.P. and H.L.; project administration, L.P. and H.L.; resources, L.P.; software, Q.P.; supervision, L.P. and H.L.; validation, L.P.; writing—original draft, Q.P.; writing—review and editing, Q.P., L.P. and H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC), [RGPIN-2020-05387] and [DGEER-2020-00347].

Data Availability Statement: Data are not publicly available, though the data may be made available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Abbreviations and descriptions of variables included in the dataset.

Variable Abbreviation	Variable Description
CDD	Maximum number of consecutive dry days in a year with precipitation less than one millimeter
Rx5Day	Maximum rainfall per month in five consecutive days
T10	Change in frequency of cooler temperatures below the 10th percentile
Tn10	Percentage of days when the daily minimum temperature is less than the 10th percentile of the reference period
T90	Change in frequency of warmer temperatures above the 90th percentile
Tx90	Percentage of days when the daily maximum temperature is greater than the 90th percentile of the reference period
WP90	Frequency of wind speed above the 90th percentile
psdi	Palmer Drought Severity Index
pr	Precipitation
tmmn	Average monthly minimum temperature
tmmx	Average monthly maximum temperature
soil	Soil moisture
time	Estimation year

Table A2. Descriptive summary of Actuaries Climate Index.

Midwest ACI Data				
N = 58				
Variables	Min.	Max.	Mean	Std. dev.
CDD_6	−1.5200	1.6200	−0.2228	0.826686
CDD_7	−1.5100	1.5800	−0.2228	0.837129
Rx5Day_5	−2.3600	3.1700	0.2216	1.075954
Rx5Day_6	−2.8600	2.7900	0.3293	1.086263
Rx5Day_7	−1.9500	2.4500	0.1081	1.013655
Rx5Day_8	−1.8600	2.6100	0.0603	1.026407
T10_4	−1.5800	5.3400	0.0966	1.257771
T10_5	−1.5000	2.0600	−0.0107	0.931239
T10_9	−1.7300	2.9100	−0.1960	0.947877
T90_6	−1.4900	3.4200	0.0419	0.869365
T90_7	−1.4200	3.3100	−0.0452	1.072289
T90_8	−1.1400	3.2300	−0.0090	0.916306
WP90_6	−2.0000	2.4700	−0.0519	0.959081
Midwest State-Level ACI Data				
N = 448				
Variables	Min.	Max.	Mean	Std. dev.
Rx5Days	−3.0069	3.3768	0.1416	7.055175
Tn10	−3.4417	2.6049	−0.6697	2.834646
Tx90	−2.6841	3.5244	−0.0742	3.126361
CDD	−1.7955	3.2369	−0.0929	5.162458

Table A3. Descriptive summary of standardized TerraClimate high-resolution data.

Variables	Iowa				Midwest			
	Min.	Max.	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.
pdsi_5	−2.2089	2.4257	0.2369	0.9235	−2.8968	2.7909	0.2197	0.9634
pdsi_6	−2.1255	2.8809	0.2810	0.9640	−3.0956	3.2119	0.2450	0.9805
pdsi_7	−2.0579	3.1031	0.2855	1.0049	−3.3342	3.4489	0.2567	1.0043
pr_5	−2.1374	4.1245	0.2233	1.0853	−2.0385	4.9146	0.1300	1.0406
pr_6	−1.9785	4.4164	0.2075	1.1000	−2.0814	5.6135	0.1683	1.0733
pr_7	−2.1185	4.7036	−0.0063	1.0682	−2.1026	5.3021	0.0378	1.0226
pr_8	−1.6730	3.9577	0.0393	1.0099	−2.0023	4.9549	−0.0004	0.9564
pr_9	−1.6081	3.4128	−0.1191	0.8677	−1.7247	5.2737	−0.0442	0.9430
tmmn_4	−4.2684	3.0962	0.0633	1.0847	−3.7637	3.3897	0.0744	1.0185
tmmn_5	−2.3074	3.3706	0.1385	0.9498	−3.5180	3.3233	0.1413	0.9930
tmmn_6	−3.2213	3.9836	0.3211	1.0449	−5.1971	3.4176	0.2040	0.9881
tmmn_7	−3.2254	3.5331	0.0854	1.0876	−4.7108	3.1205	0.1064	1.0149
tmmn_8	−2.9774	3.8303	0.1778	1.0520	−3.6873	3.3509	0.1649	1.0146
tmmn_9	−3.0840	3.0610	0.1807	1.0536	−3.3779	3.0657	0.0988	0.9565
tmmx_5	−2.3430	2.9852	−0.0271	0.9485	−3.9462	2.9478	0.0162	0.9473
tmmx_6	−3.6915	3.0492	−0.0573	0.9724	−5.0084	2.8003	0.0058	0.9499
tmmx_7	−3.8262	3.5587	−0.1971	1.1176	−5.0404	4.2726	−0.0735	1.0334
tmmx_8	−2.6017	3.7864	−0.0768	0.9583	−3.6573	3.6903	0.0115	0.9895
tmmx_9	−3.9046	2.7694	0.1893	1.0727	−4.0176	2.6574	0.0828	0.9721
soil_6	−1.4541	3.6462	0.3125	1.0964	−1.5698	4.4239	0.1609	1.0362
soil_7	−1.3344	4.8123	0.2354	1.1375	−1.3969	6.0008	0.1532	1.0864

Table A4. KPSS test for stationary of corn yields.

Region	KPSS Test for Stationarity p-Value
Midwest	Below 0.01
IL	Below 0.01

Table A4. Cont.

Region	KPSS Test for Stationarity <i>p</i> -Value
IN	Below 0.01
IA	Below 0.01
MI	Below 0.01
MO	Below 0.01
OH	Below 0.01
WI	Below 0.01

Note: the KPSS test is applied to the yearly corn yield in the whole Midwest region and each state within the Midwest. The null hypothesis is that the corn yield is stationary, and the alternative hypothesis is that the corn yield has a unit root. The *p*-values of all regions are smaller than the printed *p*-values, indicating that the corn yield is non-stationary. *p*-values shown in the table above are output from R.

Table A5. Backward stepwise selection for Model 2.

Variables	Simplest Linear Regression	Simplest Probit Regression
Intercept	67.32980 ***	−0.7595 ***
CDD_6		
CDD_7		
Rx5Day_5		
Rx5Day_6		
Rx5Day_7	1.98864 †	
Rx5Day_8	2.79095 *	
T10_4		
T10_5		0.4222 †
T10_9	−2.85146 *	0.4200 †
T90_6		
T90_7	−5.45714 ***	0.7353 **
T90_8	−5.76785 ***	0.4278
WP90_6		
Time	1.83177 ***	
N	58	58
R ² (Pseudo R ²)	0.9535	0.3451
AIC	411.2117	53.426
F-test statistic/ANOVA		
Chi-square test (<i>p</i> -value)	174.2 ***	0.0001338 ***

Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide information about the changes in estimated coefficients, significance levels, R² (Pseudo R²) values, and F-test statistics during each step. “†”, “*”, “**”, and “***” indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

Table A6. Backward stepwise selection for Model 4.

Variables	Simplest Linear Regression	Simplest Probit Regression
Intercept	65.01863 ***	−0.83610 ***
pdsi_6	2.68789 ***	−0.25164 ***
pdsi_7		
pr_5	−2.77632 ***	0.26529 ***
pr_6		
pr_7	4.03047 ***	−0.21372 ***
pr_8	−2.83344 ***	0.20966 ***
tmmn_4	−0.61710 *	
tmmn_5	−0.56257 †	0.05174 *
tmmn_9	5.08391 ***	−0.23903 ***
tmmx_6	0.60951 †	
tmmx_7	−5.00016 ***	0.37875 ***

Table A6. Cont.

Variables	Simplest Linear Regression	Simplest Probit Regression
tmmx_8	−9.62043 ***	0.59262 ***
soil_6	3.04412 ***	
soil_7	−8.29728 ***	0.38068 ***
Time	1.89820 ***	
N	5742	5742
R ² (Pseudo R ²)	0.7982	0.2864
AIC	49,193.75	4777.72
F-test statistic/ANOVA	1738 ***	0.7723
Chi-square test (p-value)		

Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide information about the changes in estimated coefficients, significance levels, R² (Pseudo R²) values, and F-test statistics during each step. “†”, “*”, “***”, and “****” indicate significance at 10%, 5%, and 0.1% levels, respectively.

Table A7. Backward stepwise selection for Model 5.

Variables	Simplest Linear Regression (Lasso Selection)	Simplest Probit Regression	Simplest Probit Regression (Lasso Selection)
Intercept	63.593053 ***	−0.756608 ***	−0.726714 ***
pdsi_5	−7.356485 ***		0.626712 ***
pdsi_6	14.471175 ***	−0.019064 †	−0.906566 ***
pdsi_7	−6.178364 ***		0.269595 ***
pr_5	0.663937 ***	−0.105575 ***	−0.065205 ***
pr_6	2.120822 ***	−0.280150 ***	−0.111829 ***
pr_7	8.913359 ***	−0.439158 ***	−0.443344 ***
pr_8	−0.730487 ***		0.041241 ***
pr_9	1.169836 ***		
tmmn_4	−0.425323 *	−0.137153 ***	−0.063977 ***
tmmn_5		−0.153493 ***	−0.066655 *
tmmn_6	4.916094 ***		
tmmn_7	6.375701 ***		−0.328385 ***
tmmn_8	7.210296 ***		−0.226689 ***
tmmn_9	−4.729296 ***	−0.155197 ***	0.034804 †
tmmx_5	4.394181 ***		−0.075420 **
tmmx_6	4.189060 ***	−0.352383 ***	−0.411686 ***
tmmx_7	−11.271298 ***	0.321780 ***	0.555092 ***
tmmx_8	−17.524793 ***	0.479800 ***	0.708800 ***
tmmx_9	7.564631 ***		−0.134287 ***
soil_6	0.910250 **	0.073061 ***	0.066103 **
soil_7	−8.339541 ***	0.430947 ***	0.410685 ***
Time	1.566997 ***		
N	39,567	39,567	39,567
R ² (Pseudo R ²)	0.7297	0.2341	0.2536
AIC	350,336.1	35,293	34,406
F-test statistic /ANOVA	5085 ***	0.2529	<2.2 × 10 ^{−16} ***
Chi-square test (p-value)			

Note: Results from backward stepwise selection show the same significant variables as the results from the manual selection approach. While the results are the same, the process of the manual selection method would provide information about the changes in estimated coefficients, significance levels, R² (Pseudo R²) values, and F-test statistics during each step. “†”, “*”, “***”, and “****” indicate significance at 10%, 5%, 1%, and 0.1% levels, respectively.

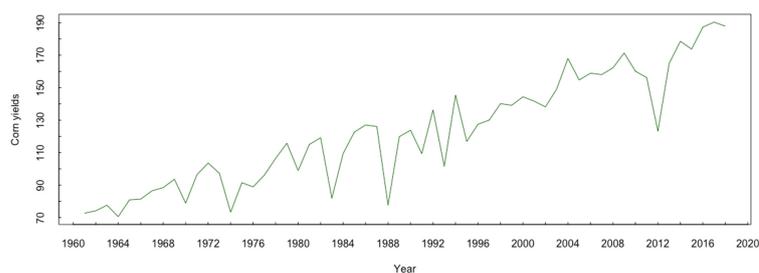


Figure A1. Annual weighted average corn yields in the Midwest (N = 58). Note: The 1958–2018 annual weighted average corn yields in the Midwest show an upward trend.

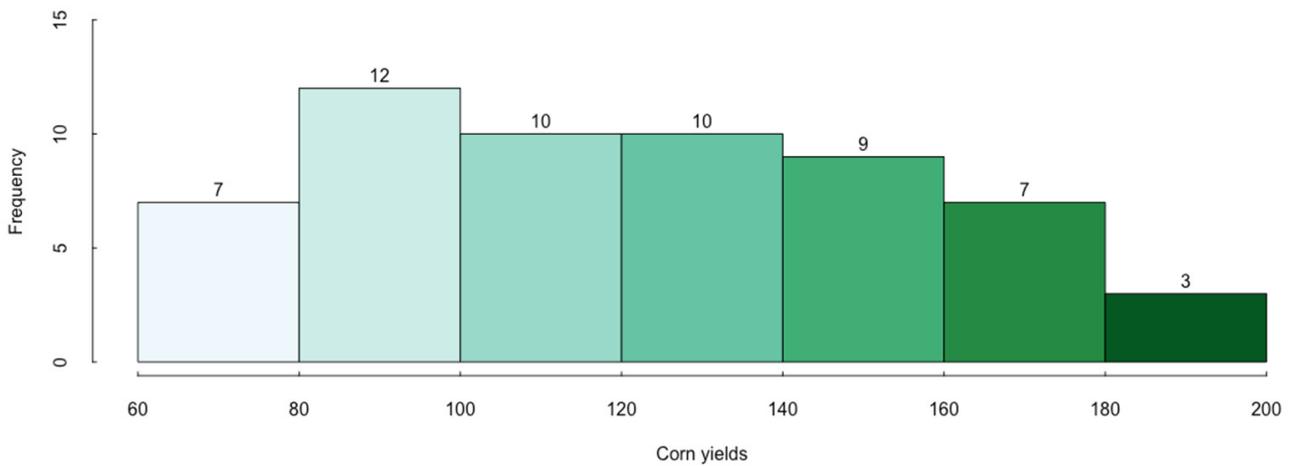


Figure A2. Distribution of annual weighted average corn yields in the Midwest (N = 58). Note: The 1958–2018 annual weighted average corn yields in the Midwest show a roughly right-skewed distribution.

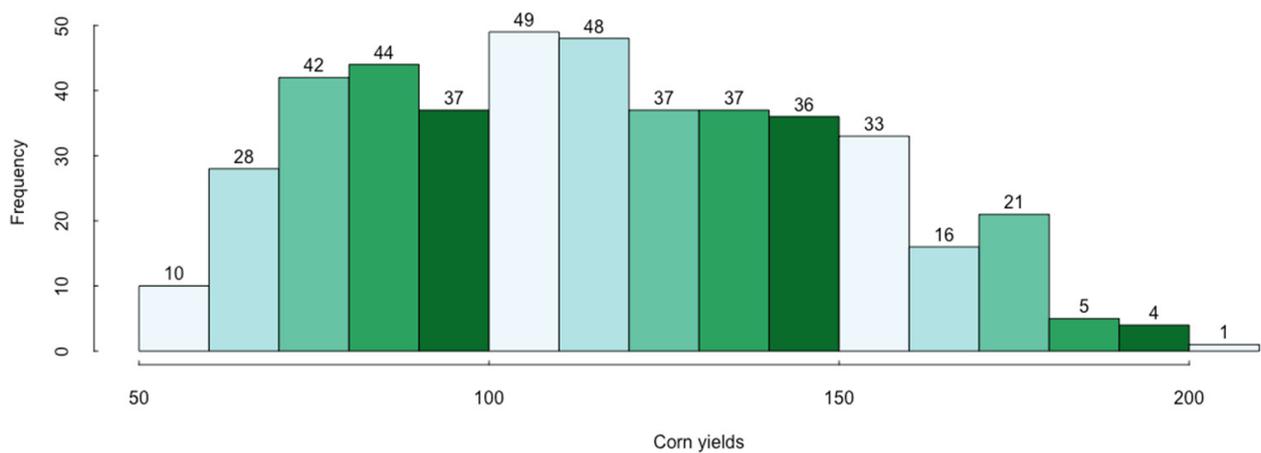


Figure A3. Distribution of annual state-level corn yields in the Midwest (N = 448). Note: The annual corn yields from 1958 to 2018 of all states in the Midwest show an approximately right-skewed distribution.

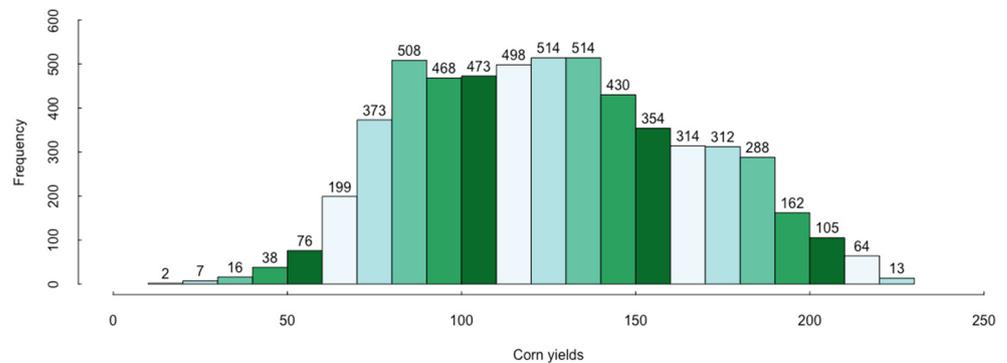


Figure A4. Distribution of annual county-level corn yields in Iowa (N = 5742). Note: The annual corn yields from 1958 to 2018 of all counties in Iowa show a roughly normal distribution.

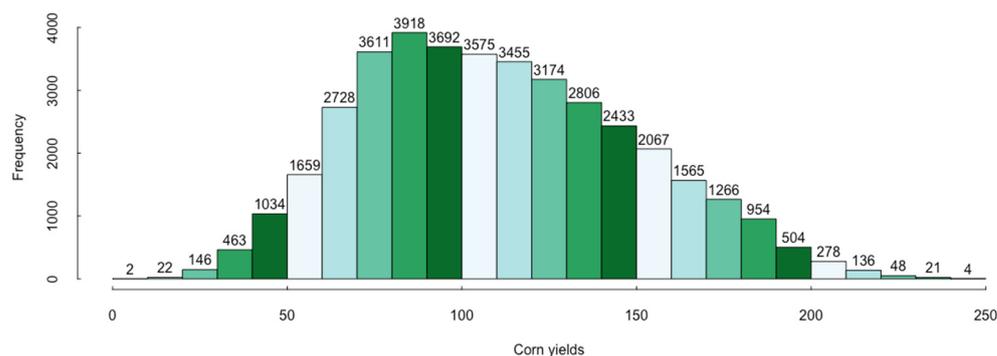


Figure A5. Distribution of annual county-level corn yields in the Midwest (N = 39,567). Note: The annual corn yields from 1958 to 2018 of all counties in the Midwest show a right-skewed distribution.

References

- Lee, J.; Nadolnyak, D. The Impacts of Climate Change on Agricultural Farm Profits in the U.S. In Proceedings of the Agricultural & Applied Economics Association's 2012 AAEA Annual Meeting, Washington, DC, USA, 12–14 August 2012; Volume 42, pp. 405–416.
- Wheeler, T.; Von Braun, J. Climate change impacts on global food security. *Science* **2013**, *341*, 508–513. [[CrossRef](#)] [[PubMed](#)]
- Lobell, D.B.; Schlenker, W.; Costa-Roberts, J. Climate trends and global crop production since 1980. *Science* **2011**, *333*, 616–620. [[CrossRef](#)] [[PubMed](#)]
- Schlenker, W.; Roberts, M.J. Nonlinear Temperature Effects Indicate Severe Damages to U.S. Crop Yields under Climate Change. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15594–15598. [[CrossRef](#)] [[PubMed](#)]
- Lesk, C.; Rowhani, P.; Ramankutty, N. Influence of extreme weather disasters on global crop production. *Nature* **2016**, *529*, 84–87. [[CrossRef](#)]
- ACI. Actuaries Climate Index: Development and Design. American Academy of Actuaries, Canadian Institute of Actuaries, Casualty Actuarial Society, Society of Actuaries. 2016. Available online: actuariesclimateindex.org (accessed on 23 June 2019).
- IPCC. 2018: Summary for Policymakers. In *Global Warming of 1.5 °C. An IPCC Special Report on the Impacts of Global Warming of 1.5 °C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways*; Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., et al., Eds.; World Meteorological Organization: Geneva, Switzerland, 2018; p. 32.
- de Jong, P.; Heller, G.Z. *Generalized Linear Models for Insurance Data*; Cambridge University Press: Cambridge, UK, 2008; pp. 1–196. [[CrossRef](#)]
- Song, J. Advances in research methods on the impacts of climate change on agricultural production. *Sci. Technol. Dev.* **2016**, *12*, 765–776. [[CrossRef](#)]
- Ward, P.S.; Florax, R.J.G.M.; Flores-Lagunes, A. Climate change and agricultural productivity in Sub-Saharan Africa: A spatial sample selection model. *Eur. Rev. Agric. Econ.* **2014**, *41*, 199–226. [[CrossRef](#)]
- Lobell, D.B.; Burke, M.B. On the use of statistical models to predict crop yield responses to climate change. *Agric. For. Meteorol.* **2010**, *150*, 1443–1452. [[CrossRef](#)]
- Coble, K.H.; Miller, M.F.; Rejesus, R.M.; Goodwin, B.K.; Knight, T.O. A Comprehensive Review of the RMA APH and COMBO Rating Methodology Final Report. *Prep. Sumaria Syst. Risk Manag. Agency* **2010**. Available online: <https://legacy.rma.usda.gov/pubs/2009/comprehensivereview.pdf> (accessed on 31 May 2022).
- Woodard, J.D. Impacts of Weather and Time Horizon Selection on Crop Insurance Ratemaking: A Conditional Distribution Approach. *N. Am. Actuar. J.* **2014**, *18*, 279–293. [[CrossRef](#)]
- Verón, S.R.; de Abelleira, D.; Lobell, D.B. Impacts of precipitation and temperature on crop yields in the Pampas. *Clim. Chang.* **2015**, *130*, 235–245. [[CrossRef](#)]
- Hoffman, A.L.; Kemanian, A.R.; Forest, C.E. Analysis of climate signals in the crop yield record of sub-Saharan Africa. *Glob. Chang. Biol.* **2018**, *24*, 143–157. [[CrossRef](#)] [[PubMed](#)]
- Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [[CrossRef](#)]
- Burke, M.; Lobell, D. Climate change and food security: Adapting agriculture to a warmer world. *Clim. Chang. Food Secur. Adapt. Agric. Warmer World* **2010**, *37*, 201. [[CrossRef](#)]
- Guo, J. Advances in impacts of climate change on agricultural production in China. *J. Appl. Meteorol. Sci.* **2015**, *26*, 1–11.
- Blanc, E. The Impact of Climate Change on Crop Yields in Sub-Saharan Africa. *Am. J. Clim. Chang.* **2012**, *1*, 1–13. [[CrossRef](#)]
- Sheehy, J.E.; Mitchell, P.L.; Ferrer, A.B. Decline in rice grain yields with temperature: Models and correlations can give different estimates. *Field Crops Res.* **2006**, *98*, 151–156. [[CrossRef](#)]

21. Rejesus, R.M.; Coble, K.H.; Miller, M.F.; Boyles, R.; Goodwin, B.K.; Knight, T.O. Accounting for weather probabilities in crop insurance rating. *J. Agric. Resour. Econ.* **2015**, *40*, 306–324. [[CrossRef](#)]
22. Turvey, C.G.; McLaurin, M.K. Applicability of the normalized difference vegetation index (NDVI) In index-based crop insurance design. *Weather Clim. Soc.* **2012**, *4*, 271–284. [[CrossRef](#)]
23. Makaudze, E.M.; Miranda, M.J. Catastrophic drought insurance based on the remotely sensed normalised difference vegetation index for smallholder farmers in Zimbabwe. *Agrekon* **2010**, *49*, 418–432. [[CrossRef](#)]
24. Deschênes, O.; Greenstone, M. The economic impacts of climate change: Evidence from agricultural output and random fluctuations in weather. *Am. Econ. Rev.* **2007**, *97*, 354–385. [[CrossRef](#)]
25. Huang, H.; Khanna, M. An Econometric Analysis of U.S. Crop Yield and Cropland Acreage: Implications for the Impact of Climate Change. In Proceedings of the Agricultural and Applied Economics Association (AAEA) Conferences 2010 Annual Meeting, Denver, CO, USA, 25–27 July 2010. [[CrossRef](#)]
26. Vogel, E.; Donat, M.G.; Alexander, L.V.; Meinshausen, M.; Ray, D.K.; Karoly, D.; Frieler, K. The effects of climate extremes on global agricultural yields. *Environ. Res. Lett.* **2019**, *14*, 054010. [[CrossRef](#)]
27. Porth, L.; Roznik, M.; Tan, K.; Zhu, W.; Porth, C.B. Predictive Analysis: The Effects of Technology and Weather on Crop Yield. Society of Actuaries. 2019. Available online: <https://www.soa.org/globalassets/assets/files/resources/research-report/2019/predictive-analysis-effects.pdf> (accessed on 31 May 2022).
28. Goodwin, B.K.; Piggott, N.E. Has Technology Increased Agricultural Yield Risk? Evidence from the Crop Insurance Biotech Endorsement. *Am. J. Agric. Econ.* **2020**, *102*, 1578–1597. [[CrossRef](#)]
29. Priest, G.L. The government, the market, and the problem of catastrophic loss. *J. Risk Uncertain.* **1996**, *12*, 219–237. [[CrossRef](#)]
30. Tollenaar, M.; Fridgen, J.; Tyagi, P.; Stackhouse, P.W.; Kumudini, S. The contribution of solar brightening to the US maize yield trend. *Nat. Clim. Chang.* **2017**, *7*, 275–278. [[CrossRef](#)] [[PubMed](#)]
31. Roebber, P.; Brazauskas, V.; Kravtsov, S. The Actuarial Utility of Weather and Climate Predictions. Submitted for Publication. 2017. Available online: <https://cpb-us-w2.wpmucdn.com/sites.uwm.edu/dist/a/122/files/2016/05/2017rbk-1y141el.pdf> (accessed on 31 May 2022).
32. Jin, Z.; Erhardt, R.J. Incorporating Climate Change Projections into Risk Measures of Index-Based Insurance. *N. Am. Actuar. J.* **2020**, *24*, 611–625. [[CrossRef](#)]
33. On Risk Classification. American Academy of Actuaries. 2011. Available online: https://www.actuary.org/sites/default/files/files/publications/RCWG_RiskMonograph_Nov2011.pdf (accessed on 31 May 2022).
34. Murnane, R.J. Climate research and reinsurance. *Bull. Am. Meteorol. Soc.* **2004**, *85*, 697–707. [[CrossRef](#)]
35. Della-Marta, P.M.; Liniger, M.A.; Appenzeller, C.; Bresch, D.N.; Köllner-Heck, P.; Muccione, V. Improved estimates of the European winter windstorm climate and the risk of reinsurance loss using climate model data. *J. Appl. Meteorol. Climatol.* **2010**, *49*, 2092–2120. [[CrossRef](#)]
36. Collins, D.; Gibson, R.; Kolk, S.; Lindman, C.; Mathewson, S.; Hall, D.; Guérard, Y. Actuaries Climate Risk Index Preliminary Findings. *Am. Acad. Actuar.* **2020**. Available online: <https://www.actuary.org/sites/default/files/2020-01/ACRI.pdf> (accessed on 31 May 2022).
37. Abatzoglou, J.T.; Dobrowski, S.Z.; Parks, S.A.; Hegewisch, K.C. TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Sci. Data* **2018**, *5*, 170191. [[CrossRef](#)]
38. Goldburd, M.; Khare, A.; Tevet, D.; Guller, D. *Generalized Linear Models for Insurance Rating*, 2nd ed.; Casualty Actuarial Society: Arlington, VA, USA, 2019.
39. Heras, A.; Moreno, I.; Vilar-Zanón, J. An application of two-stage quantile regression to insurance ratemaking. *Scand. Actuar. J.* **2018**, *2018*, 753–769. [[CrossRef](#)]
40. Frees, E.W. *Regression Modeling with Actuarial and Financial Applications*; Cambridge University Press: Cambridge, UK, 2010; pp. 1–565. [[CrossRef](#)]
41. Jin, Y.; Rejesus, R.M.; Little, B.B. Binary choice models for rare events data: A crop insurance fraud application. *Appl. Econ.* **2005**, *37*, 841–848. [[CrossRef](#)]
42. Zhao, C.; Liu, B.; Piao, S.; Wang, X.; Lobell, D.B.; Huang, Y.; Asseng, S. Temperature increase reduces global yields of major crops in four independent estimates. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9326–9331. [[CrossRef](#)] [[PubMed](#)]
43. Li, H.; Porth, L.; Tan, K.; Zhu, W. Improved index insurance design and yield estimation using a dynamic factor forecasting approach. *Insur. Math. Econ.* **2021**, *96*, 208–221. [[CrossRef](#)]