2019-20 Modeling the Future Challenge Project Report

An Analysis on the Impact of Climate Change on Corn in Minnesota

Parth Asawa, Brandon Guo, Richard Liu, Eric Zhang

Executive Summary

Corn is a fundamental part of the US agricultural industry, being the crop of the greatest gross tonnage produced and highest monetary value yielded each year.

In this report, we focused our analysis of climate change's impact on corn yield and insurance claims in Southern Minnesota through an analysis of how temperature and precipitation are expected to change over the next 30 years, and how these changes are projected to affect the relationship between the agriculture and the insurance industries with regards to crop loss and indemnities.

Data regarding regional climate trends and historical indemnities is used in order to predict changes to the agricultural industry, and how insurance industries can best react to mitigate their losses. Sources such as the USDA (United States Department of Agriculture) RMA (Risk Management Agency), NOAA (National Oceanic Atmospheric Administration), and the USDA NASS (National Agricultural Statistics Service) are useful in developing a thorough understanding of how regional climate changes, and how indemnities and crop losses are affected by those changes.

Our methodology rests on the observation that both temperature and precipitation are stochastic processes, i.e. they cannot ever be fully determined in the future, and thus warrant stochastic modeling and simulation. Thus, temperature and precipitation are modeled individually as random processes viz. Markov process and random walk, respectively. The two results are combined by means of a synthesis model to predict the volume of indemnities. This synthesis model incorporates feature engineering and is trained by optimization on past data. This pipeline (forecasts of temperature, precipitation, and the synthesis model) is performed 1000 times, and combined with a Monte Carlo average to deduce the forecasts for indemnity loss.

The results from our modeling demonstrated a clear increase in the volatility of temperature and precipitation, with anomalous values occurring in short spikes rather than extended periods of time. After training and generating the combined model of these factors against crop loss, we were able to quantify the expected total payout of insurance companies over time using the upper bound of a 99.7% confidence interval to approximately $29.5 billion in total over the next 30 years, which in comparison to current yearly loss, led to our recommendations focusing on increased premiums, diversification of the agricultural market with crops that are more resilient to climate change and optimized growing patterns, and an overall improvement to the infrastructure that supports irrigation and crop-growth.

Introduction

Background Information

Agricultural Industry Background

Corn is one of the crops that is globally most susceptible to climate change due to its vulnerability to extreme heat and susceptibility to precipitation damage. Estimates suggest that only two degrees of warming globally would contribute to a $20-40\%$ decrease in corn production globally with four degrees leading to a $40-60\%$ decrease [1]. Corn alone consists of over $4 billion of the Minnesota Agricultural Industry, and just under 50% of all crop production of the state in a single year [2].

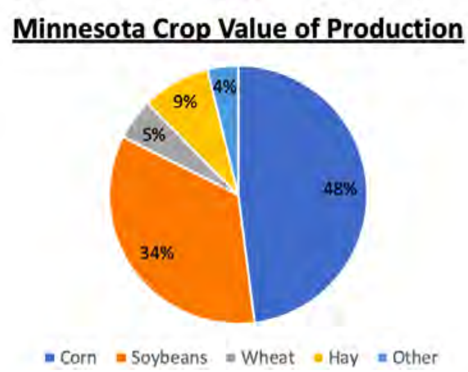

Figure 1: Percentage breakdown of crop production in 2018, with corn being the most widespread form of agriculture.

Overall corn production in the US is concentrated in a small region, mostly in Minnesota, Iowa, Illinois, and Nebraska [3]. Counties in Minnesota generate more than 20 million bushels of corn annually, so naturally it makes up a large part of their economy. Southern Minnesota specifically is cut through by the Minnesota River, which has a large fertile floodplain to grow corn, as well as provide water for irrigation.

There are two main types of insurance in Minnesota: crop yield insurance, which covers losses due to natural disasters, and crop revenue insurance, which covers losses due to fluctuating prices. For our analysis we are mostly concerned with crop yield insurance, since climate change affects natural disasters such as floods and hail, however we do have recommendations for both. There are three main types of coverage: multi-peril crop insurance, which covers all types of losses; crop hail, which covers fire, hail, and storage losses; and replant, which covers the cost of replanting crops after a failure [4]. Multi-peril Crop Insurance is a part of the Federal Crop Insurance Corporation, while crop hail insurance is generally provided by a private insurance provider [5]. Corn is a especially susceptible to hail damage, so it is important for farmers to purchase crop hail insurance in regions with higher amounts of hail [6] [7].

Climate Forecast Background

We choose to focus our analysis of climate factors on precipitation and temperature. Among the commonly accepted Top 7 Factors in Crop Production [8], virtually every factor (soil, nutrients, damage, etc) that affects crop production is affected by the amount and volatility of precipitation and temperature .

According to the United Nations Framework on Climate Change (2010): "To prevent dangerous interference with the climate system, the scientific view is that the increase in global temperature should be below 2 degrees Celsius." Global warming has resulted in an increasing global mean temperature, however, the mean temperature in Minnesota is increasing faster than the global mean. Forecasts suggest that even if the global mean temperature only raises by 2 degrees Celsius, Minnesota's mean temperature will see a rise of 3.3 degrees Celsius [9].

Furthermore, Minnesota's precipitation is also expected to change significantly. Multi-day events of heavy rainfall which can be damaging to crop production have seen a 37% increase in recent years in the region, suggesting an acceleration of weather events that significantly impact crop yield. It's important to note that while warming during the winter does indeed increase the length of the growing seasons, it also directly correlates to an increase in the number of pests and soil erosion, both of which negatively impact crop yield during the harvesting season. Furthermore, when this warming occurs during the summer, it directly corresponds to a lower crop yield.

Additionally associated with global warming is the increase in volatility of weather events and the increased presence of extreme events which cause crop damage and soil loss. Our analysis robustly incorporates increasing volatility values to best model this trend, so that we make the most accurate insurance and policy recommendations as evidenced by the literature on global warming in relation to agriculture.

Finally, warmer winters have led to reductions in the snowpack which has sped up the process of water release to agriculture, in addition to precipitation in the winter falling as rain instead of snow, increasing wildfire risk and the severity of droughts that occur in the summer [10].

Modeling

Data Methodology

We draw data from three main sources in our modeling: The National Oceanic and Atmospheric Administration's Climate Data Online archive of historical weather, The National Agricultural Statistics Service of the USDA's Quick Stats Data, and the US Department of Agriculture Risk Management Agency's Cause of Loss files.

Data from these sources allows us to specifically gain data regarding climate change over time, crop production over time, and how indemnities have changed over time, which we can then analyze in different pairwise models to make appropriate policy and insurance recommendations while considering the multifaceted ways climate change can affect agricultural indemnities.

In the following text, we present an overview of each of our data sources, including what data points and measurements are being explored in our analysis.

USDA RMA Cause of Loss (Primary Data Set) [11]

- **Scope and Parameters of data:** Yearly indemnity data for corn for the selected counties in Minnesota aggregated from the monthly indemnity data (historic insurance claims) from 2001 to 2019 from the USDA Risk Management Agency's "Cause of Loss Files".
- **Purpose of data:** The indemnity data provides us direct information on how the insurance industry has been impacted over time, which we then use to train and create a model against changes in the climate to forecast future losses to insurance companies.
- **Motivation:** The usage of this primary data set is important to our analysis as the insurance claims that have been filed by farmers provide us direct information of how the insurance portion of the agricultural industry has been impacted over time, which is the only way we could correlate our insurance recommendations to the simultaneously occurring changes in the climate.

USDA National Agricultural Statistics Production Data (Supporting Data Set) [12]

- **Scope and Parameters of data:** Total amount of corn production for the selected counties by year.
- **Purpose of data:** The production data is used in exploratory analysis to study the key parameters that have large impact on crop yield and thus may be correlated

to loss. The values themselves are not used in our analysis, but yielded important insights for our modeling and parameter selection.

- **Motivation:** While indemnity data can allow us to see the overall losses due to climate factors, production data allows us to investigate total impact of climate change on corn. Production data is a key indicator of the overall health of the corn industry.

NOAA Climate Data Archive Daily Summaries (Supporting Data Set) [13]

- **Scope and Parameters of data:** Daily extreme and maximum precipitation/temperature of the corn-growing counties between 1950 and 2018 sourced from "Climate Data Online" archives of the NOAA specific to the region of Minnesota.
- **Purpose of data:** Daily precipitation and temperature data is extracted as time series data and converted to PDSI values (as detailed in the methodology), then used for the forecasting for each of the parameters.
- **Motivation:** The usage of the NOAA Climate Archives provide us with official information of climate, but also allows us to analyze specific factors that interest us such as maximum daily temperatures and precipitation or the occurrence of multi-day events, which were necessary to our definition of the two important climate factors we chose to focus on as described in the Background Information section.

NOAA Climate Data Archive Yearly Summaries (Supporting Data Set) [13]

- **Scope and Parameters of data:** Yearly extreme and maximum precipitation/temperature of the corn-producing counties between 1950 and 2018 sourced from "Climate Data Online" archives of the NOAA specific to the region of Southern Minnesota.
- **Purpose of data:** In conjunction with the USDA datasets on loss and production, yearly data (converted to PDSI values detailed below) is used to justify the use of our temperature and precipitation parameters as a predictor of output.
- **Motivation:** Similar to the previous data set motivation, the usage of this supporting data set was motivated by the need for specific data as we defined climate change in our recommendations to insurance companies.

What additional data would have helped our model?

The changes in foreign supply and demand (or even domestic outside of Minnesota) caused by climate change could exacerbate the results we see in our models below. This is important to note because while our region of interest is solely Southern Minnesota, the data from this wider scale would give a better picture of changes that we might not be able to identify with models limited by information consisting of production, climate, and indemnity data solely from Southern Minnesota.

Mathematics Methodology

Overview

Our overall goal in our modeling is to predict future values of crop loss using forecasts of precipitation and temperature. Thus, we must first independently arrive at those two forecasts in order to get a coherent understanding of crop loss.

Here we present the workflow of our modeling methodology, with key steps labeled numerically:

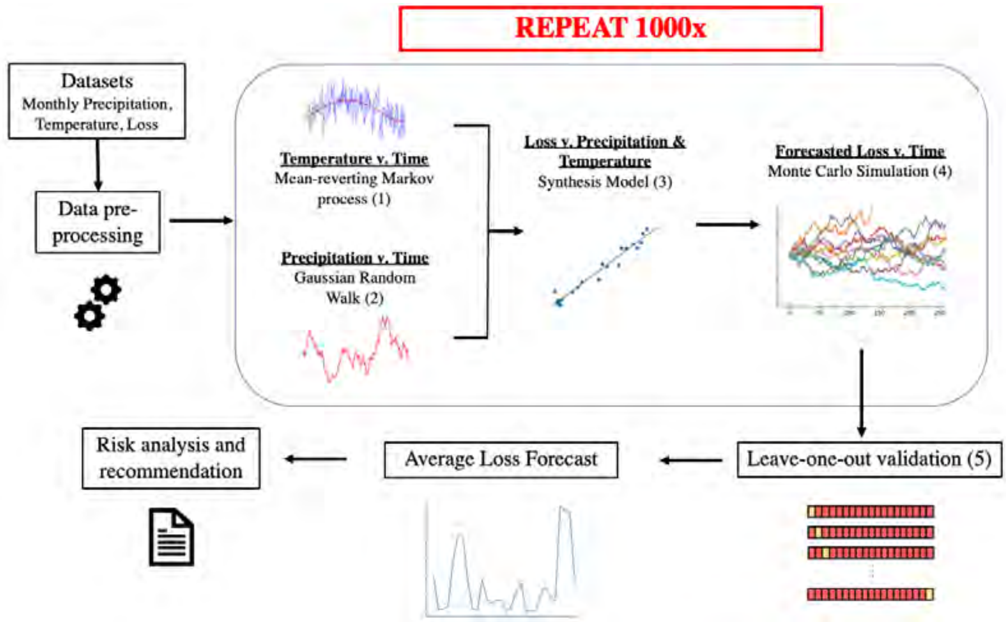

Figure 2: Analysis workflow to forecast the next thirty years of crop loss.

To summarize, after classic pre-processing of our data files we begin the iterative modeling process. To forecast temperature, we use a mean-reverting process around a sinusoidal fit, as shown in (1). The precipitation is forecasted using a random walk methodology (2). These two independently stochastic models are combined via a synthesis model, which incorporates a regression along with feature engineering of our two previous forecasts (3). This model allows us to make one simulation of future indemnities based on our simulated parameters i.e. temperature precipitation (4). This methodology, steps (2), (3), (4) are repeated 1000 times, and the algorithm is subjected to leave-one-out validation (5) to prevent overfitting. The resulting forecast is determined by averaging across our 1000 simulations.

Our methodology can be partitioned into three major steps:

1. Temperature and Sinusoidal simulations (Markov process, random walk)
2. Synthesis model (regression)

3. Indemnity forecast of simulations (Monte Carlo)

Each section of the analysis will begin by introducing the methodology of that particular analysis, followed by stating a few key assumptions. Then, the theory is presented, followed by a justification of why the specific model is appropriate for the task at hand.

Forecasting Temperature

We will forecast temperature by beginning with a simple sinusoidal fit that estimates seasonal and temporal behavior of the data. We begin by using literature and the intuitions from our exploratory analysis to form a general four-parameter model for temperature. We then learn individual parameters of the model via gradient descent to minimize a least-squares loss function.

To simplify modeling, we make some assumptions about the data. We assume that the NOAA temperature data takes a consistent measurement, and this measurement is a good "summary" of temperature that ends up influencing our loss output. In this section, we propose that the temperature in a given month is weakly correlated to the temperature in all past months. The justification here is that even though hot months may precede hot months, and cold months cold, this is largely due to seasonal variation (i.e. two hot months will mostly likely occur during summer) that matters more than the monthly variation. We also postulate that our data will exhibit a trend of increasing average temperature over time, largely due to climate change. We will assume that the trend we identify will extend into the future in the same manner. This seems reasonable because experts forecast that climate change will only grow at an increasing rate (i.e. faster than linear), which means our model is making a conservative forecast and likely not excessively extrapolating [14].

Based on our exploratory analysis, we observe that temperature obeys a sinusoidal trend with respect to time, in particular the seasonal variations that are cyclic every year. We propose the model below that has been applied in literature regarding temperature forecasting [15, 16].

$$T_m(t) = a + bt + c\sin\left(\frac{2\pi t}{12} + d\right) \tag{1}$$

where t is in months. We can then learn the parameters a, b, c, d by gradient descent, minimizing our cost function with respect to our true temperatures $T(t)$ and our estimation for the mean temperature T_m:

$$J(t_1, t_2, \ldots, t_n, T_m) = \frac{1}{n}\sum_{t_1,\ldots,t_n} (T(t) - T_m(t))^2 \tag{2}$$

Gradient descent is a mathematical technique where a loss function (Equation 2), which quantifies the difference between our predictions and the data, is minimized by taking directional derivatives until a local minimum is reached. We use this method to optimize our parameters a, b, c, d so that the equation above has the smallest value, which is mathematical evidence that it is an optimized "fit" for the data at hand.

It is worth also mentioning here the bt term attempts to model global warming trends of increasing average temperatures, and this is well-cited in temperature models used by [17]. So the bt term is indeed important to estimating the means of temperature.

We now begin our forecasting. We propose a mean-reverting Ornstein-Uhlenbeck processing that is well cited in weather derivative pricing as a stochastic model for temperature [18]. Concretely, we propose that the sequences of temperatures $\{T(n)\}$ be modeled by the stochastic differential equation, where at some time t:

$$dT = dT_m + \alpha(T_m - T) + \sigma dW \tag{3}$$

where W is Brownian motion [18]. But because we are dealing with discrete time series (as opposed to particle motion) we work with the discrete form of Equation (3), namely that

$$T_t = T_t^m + \alpha(T_{t-1}^m - T_{t-1}) + \sigma_t \epsilon_t \tag{4}$$

which is just the discrete form of Equation (3), where ϵ_t is a random normal. This allows to get one stochastic simulation of $T(t)$. It remains for us to propose values of σ_t and α, the daily volatility and the mean-reverting speed.

To deduce σ_t, we note that since our temperature value is roughly periodic (a key assumption for our earlier sine model), we can make a discrete-time Fourier transform (DFFT) to our daily temperature time series. Then Benth et al. derives the following expansion of σ^t based on some parameters $c_0, c_1, \ldots$:

$$\sigma^2(t) = c_0 + \sum_{i=1}^{I} c_i \sin\left(\frac{2\pi i t}{365}\right) + \sum_{j=1}^{J} c_j \cos\left(\frac{2\pi j t}{365}\right) \tag{5}$$

where the t indices are discrete time point $t = 0, 1, 2, \ldots 364$ and the parameters $c_0, c_1, \ldots$ are optimized by the same gradient descent method outlined in Equation (2) [16].

Concretely, we claim that the volatility $\sigma(t)$ is dependent on the day of the year, so we look at the 365 historical variances in our data, one for each day. A nonlinear regression is performed then to estimate c_0, which gives us discrete volatility.

The mean-reverting parameter α for this process is also estimated by Wang et al., who propose the following closed form for α

$$\alpha = -\log\left(\frac{\sum_{i=1}^{n}((T_{i-1} - T_{i-1}^m)/\sigma_{i-1}^2)(T_i - T_i^m)}{\sum_{i=1}^{n}((T_{i-1} - T_{i-1}^m)/\sigma_{i-1}^2)(T_{i-1} - T_{i-1}^m)}\right) \tag{6}$$

which results in considering the weather time series as a ideal martingale [16]. Although the closed form itself is messy, conceptually it is a weighted average of how more the temperature T_t deviates from the mean T_m than does T_{t+1}, for all t in the range. We compute α using a simple iterative approach over our time intervals to estimate $\alpha = 0.3043$.

Thus the resulting value $T(t)$ can now be estimated for each increment of t. We elucidate the implementation of this stochastic process with the following pseudo-code, which is implemented on Python and TensorFlow backend:

Algorithm 1 Mean-reverting stochastic to simulate temperature given mean temperature $T_m(t)$, mean-reverting speed α, volatility σ, and last observed temperature T, at some time n in the future.

$c \leftarrow 0$ {counter variable}
$T \leftarrow T$ {last-seen temperature value}
while $c < n$ **do**
  $\epsilon \leftarrow \mathcal{N}(0, 1)$ {random normal sample}
  $T \leftarrow T + \alpha(T - T_m(t)) + \sigma\epsilon$ {execution of Equation 4}
  $c \leftarrow c + 1$ {increment counter}
end while
return T

Forecasting Precipitation

To measure precipitation, we will use the Palmer Drought Severity Index (PDSI) monthly data to forecast the future. This is an index that spans from -10 (dry) to 10 (wet) to measure the relative amount of precipitation during a certain time period.

Because we believe that there is a weak relationship between *months* of precipitation (i.e. a wet month has and subsequent wet month, especially given seasonal variations), we propose the use of a random walk to best model precipitation. Literature regarding precipitation, often citing El Niño, suggest that precipitation will be more volatile and variant in the future as the climate itself becomes more unpredictable.

For this reason, we use the Brownian motion limit of random walk, a process whereby each new precipitation value differs from the previous precipitation value by a difference which is normally distributed. We choose this model primarily because its forecasts mathematically shows increasing volatility over time [19]. The rationale here for the random walk is that, unlike temperature, precipitation does not seem to exhibit any sort of cumulative or seasonal trend, but it does seem to be more volatile over time (which is agreed upon by the literature) [9, 20]. Therefore it seems reasonable to construct random walks starting from our past data and aggregate our results into a model for the next thirty years. This seems like the most optimal way to forecast something that, mathematically speaking, we have no reliable way to accurately estimate, even with complex models like neural networks. We show the specific methodology further down in this section.

Before we present the methodology, we propose a few assumptions and conditions for random walk: we assume there are no prominent seasonal effects of precipitation on a monthly basis, so the random walk is well justified. This claim can be verified when the data is presented in the Results section, where historic precipitation does not seem to follow any such trends. We also assume that PDSI is a measure of purely climate-related precipitation and water measurements, and not affected by human impacts on water levels. This means the precipitation we are measuring is purely "natural" and thus its changes can be directly attributed to climate change.

With those assumptions in mind, we are ready to present the random walk model: Given monthly reads of precipitation $P_1, P_2, ...P_n$, we estimate the next monthly precipitation amount P_{n+1} as

$$P_{n+1} = P_n + k \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, \sigma) \tag{7}$$

where σ the historical volatility across all of our data that builds the normal value from which we randomly sample k i.i.d. from a normal distribution from each step of the random walk. We compute σ to be 0.9568 by taking the standard deviations of all of the monthly differences in precipitation, which remains constant throughout. Each new precipitation value, therefore, randomly deviates from the past one by an amount that is, over many trials, normally distributed.

Combining models

We now wish to combine the precipitation and temperature time forecasts and prediction for future loss using a synthesis model. Thus in this model, we present the methodology assuming we have already arrived at our temperature forecasts and one iteration of the random walk forecast for precipitation. Our goal here is to synthesize these models to predict a yearly loss amount, which will then be aggregated over 1000 iterations to achieve an average loss estimate.

Whereas our previous forecasts were performed on a monthly basis, we choose to perform loss predictions on a yearly basis. The justification is that predicting loss on a monthly basis requires a level of granularity that simply cannot be represented with our random walk and sinusoidal regression model. Furthermore, if it is of low value to the actuarial industry to present loss on a monthly basis, especially if these values would naturally be more inconsistent. We forecast instead yearly loss predictions, which we believe will be more instrumental and suitable with our methodology.

To achieve this, we propose a modified polynomial regression model. We find our loss values per year from the USDA dataset, and these amounts will be used for our model to learn its parameters.

For our assumptions, we assume that the precipitation and temperature are independent. This allows us to deploy our model without worrying about confounding variables. While there may indeed be some correlation between the two, because our models came up with precipitation and temperature independently, an independence assumption seems reasonable.

Moreover, we assume that crop growth and production follows a roughly quadratic relationship with climate variables from an optimal value. This is well-justified by Figure 3 (presented below), a graph presented by the US Global Climate Change Research Program, which shows the conditions for optimal plant growth [17]. This clearly shows a quadratic relationship with respect to some optimal value.

This particular graph and its mathematical implications will be implemented in our model, as well. Finally, we assume that the cumulative effect of different climate variables is additive, which follows largely from the independence assumption, and implies our resulting model is quite appropriate for combining these two parameters.

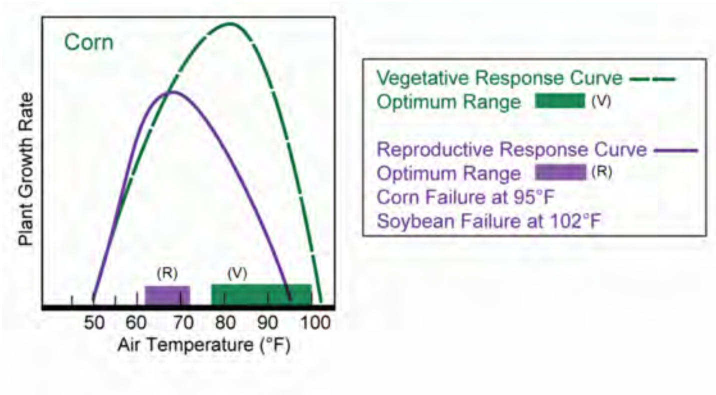

Figure 3: Studies from the US Global Climate Change Research Program suggest that corn's growth is roughly quadratic, with optimal rate achieved at some ideal temperature. These observations are incorporated in our synthesis model.

We are now ready to present the methodology. The underlying premise of this regression is that we believe extreme climate behavior, whether it be high (temperature or precipitation) or low, have similar effects on loss (generally some negative one). This is not much an assumption but rather a hypothesis of the data that the regression will explore.

The data presented in the above Figure also suggest that there exist some optimal values of temperature and precipitation T_{ideal} and p_{ideal} where loss is minimized. We do not know precisely where these values lie, but thankfully we can learn them by running gradient descent on past data.

Given yearly precipitation reads $p = p_1, p_2, \ldots p_n$, temperature reads $T = T_1, T_2, \ldots T_n$, and crop losses $y_1, y_2, \ldots y_n$, we seek to fit the following function:

$$\hat{y} = \beta_0 * (T - T_{ideal})^2 + \beta_1 * (p - p_{ideal})^2$$

where our parameters $\beta_0, \beta_1, T_{ideal}, p_{ideal}$ are learned by gradient descent to optimize least squares error (similar to our temperature model).

The rationale for this model is that values of T and p that deviate from what the the data suggests is "ideal" contribute in some way to loss, whether or not they are larger or smaller. This is where the squared difference comes in. The amount with which they contribute to loss is unknown, though, so parameters β_0 and β_1 are introduced so that the model can learn the relative weight of these deviations (i.e. whether or not a significantly non-ideal precipitation is more impactful than a non-ideal temperature).

Monte Carlo Simulations

Each of the predictions made by the Monte Carlo simulation are now averaged together to generate the final results. Concretely, given loss predictions $L_1, L_2, \ldots, L_{1000}$ we compute the average simulation L as the simple average:

$$L = \frac{1}{1000} \sum_{i=1}^{1000} L_i \tag{8}$$

Model Results

Temperature Results

Based on our gradient descent model, we arrive at the following time-based modeling for temperature, where t is in months.

$$\hat{T}(t) = 44.69 + 0.0019t + 33.69 \sin(1.728t + 0.704) \tag{9}$$

This equation is the mean temperature over time, and will serve as the baseline for our mean-reverting model. The 0.0019 term indicates the average amount of degrees the temperature has risen in per month based on the model. So above the sinusoidal trend for seasonality, which is expected, we also see a non-negligible increase in temperature over time. We can check our model on our historical data after our optimization process to ensure it is an accurate representation of the data before we extrapolate for future time points.

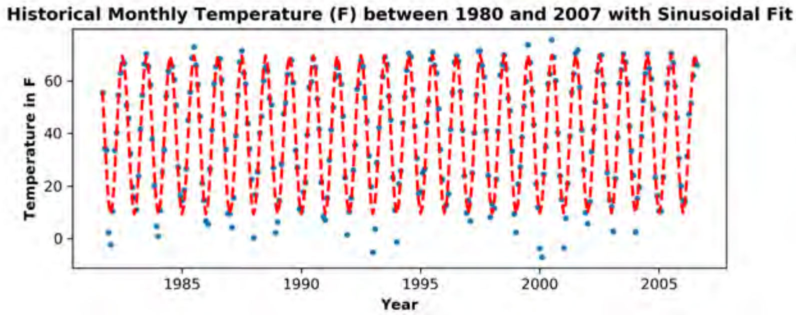

Figure 4: Sinusoidal fit on historical temperature monthly values between 1980 and 2007 overlaid on true data.

Now, based on the algorithm described in the methods section, we are able to arrive at a simulated forecast for the temperature. We showcase a portion of one simulated forecast, showing the mean temperature, represented by a sine curve, as well as our forecast. It is clear that the forecast is mean-reverting with respect to the sinusoidal curve of best fit.

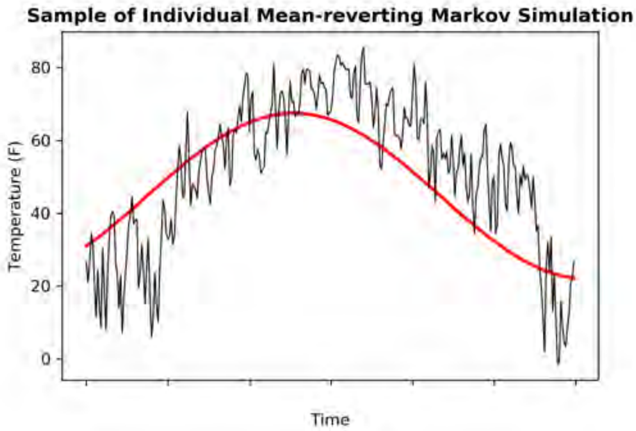

Figure 5: Mean-reverting simulation, in black, for a period of time on training data. Overlaid over sinusoidal fit.

This is one individual simulation that will be eventually combined by Monte Carlo methods, and we can see that the simulation is indeed mean-reverting and seems to better capture the volatility than a simple sinusoidal fit. We choose not to overlay the simulation over past data here, because we do not wish to use a single random sample to generalize conclusions about accuracy. We will perform validation later in the methodology, when we have combined individual simulations.

Precipitation Results

We present our aggregated random walks for the PDSI values until 2050:

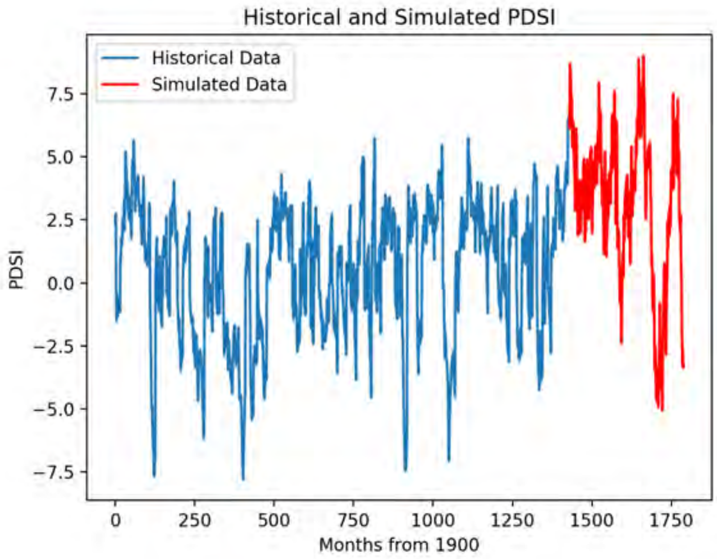

Figure 6: A PDSI simulation using a random walk.

Figure 6 (above) shows that forecasted PDSI demonstrates a larger volatility than the historical data. Additionally, extreme PDSI values are projected by the model, which surpass historic extremes, particularly for positive PDSI. This could indicate the possibility for higher amounts of extreme months, which poses interesting directions for analysis. The average results indicate a higher average precipitation but continues to show dry periods. The complete implications of such projected behavior will be discussed in the Recommendations section.

Combined Results

Using our synthesis model, converted our monthly forecasts into yearly forecasts, then combined the two yearly forecasts (temperature and precipitation) to learn a model for loss, based on past data to learn our parameters. The historical data of loss versus our prediction is plotted below. The prediction themselves are not far off, but most importantly the model is able to capture movement of loss over time quite accurately.

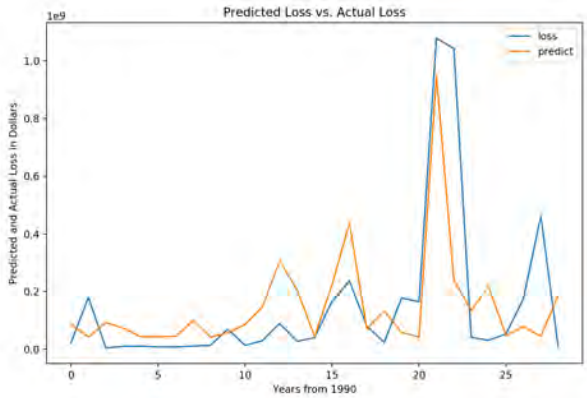

Figure 7: Historical values of loss versus predicted values of loss based on precipitation and temperature parameters. Predicted forecast shows large parallels with historical ground truth loss values.

The purposes of the previous few results were to justify the accuracy and appropriateness of the individual pieces of our model. We are now able to combine the models to forecast future loss, which we show here.

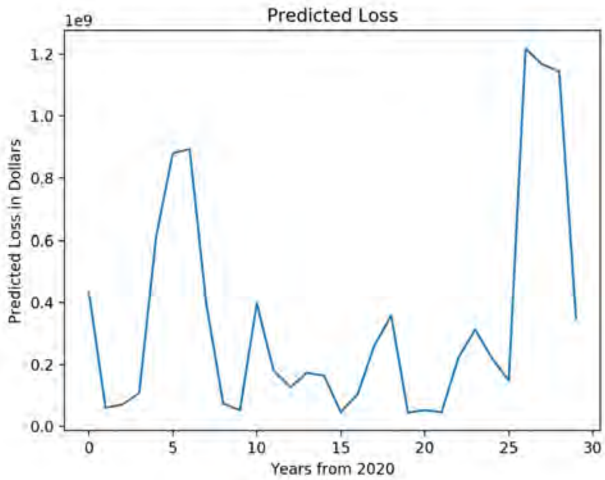

Figure 8: An example of forecasted loss values based on the model established in Figure 7 (computed from one specific random walk). Large spikes in loss suggest an increasingly high volatility of loss values in the future, but not necessarily a consistent increase in yearly loss values.

The forecast shows two large spikes in loss which seem larger than what has historically been true, along with smaller valleys with relatively low loss. Thus the model does not seem to suggest that loss necessarily will increase monotonically in the future but rather there is a generally larger volatility in the amount of loss. The implications of this forecast as well as the recommendations we conclude are discussed in depth in the Recommendations section.

Analysis

Risk Analysis

The risk of crop failures and reduced crop yields will likely increase in the coming decades due to the effects of climate change. This risk will ultimately be the burden of insurance companies, who will have to raise premiums. The price of increased premiums will be passed on to the farmer, who in turn will pass it on to the consumer. Therefore, it is inevitable that our food prices increase in the coming decades due to climate change [21].

Risks to the Agricultural Industry

Agriculture is a big part of the Minnesotan economy. In fact, employment in industries related to agriculture accounts for 15% of total jobs. Specifically, in rural Minnesota, agriculture accounts for 24% of jobs and even in metro areas, agricultural accounts for 13% of jobs [22]. A decrease in crop production in the years leading to 2050 will lead to unemployment for many farmers, as they will be unable to keep up with production elsewhere in the country, where the effects of climate change are not so pronounced. If crop production declines, the economy will face a directly reduction of agricultural jobs, which is a crucial component of the Minnesota economy and could lead to an increase in unemployment rates in the state.

Overall corn production will decrease as temperature increases according to our model (Figure 4). However, our model does not fully capture the extremely cold months, represented by the blue dots which lie under the red dotted line. This means that not only will the average temperature increase in the future, but cold weather months will also become more extreme. Therefore, it is imperative that farmers not only be prepared for extreme hot weather, but also for extreme cold weather. Extreme cold weather, such as an early (or late) season frost, could mean disaster for a corn harvest if not dealt with correctly.

Identification of Ancillary Risks

Since corn accounts for 48% of Minnesota agricultural production, a reduction to corn production could mean widespread unemployment that ripples across the local area, even to other industries up and down the supply chain. Suppliers of farm equipment could see reduced sales, especially for equipment used on plants highly susceptible to climate change such as corn. More troubling, though, heightened food prices due to a restricted and irregular supply of corn could mean reduced living conditions for many families, who will have to spend a higher percentage of their income on food. The dual risks of increased unemployment and higher food costs could mean many families will move out of the Minnesota region in the near future.

In addition, the highly variable water supply due to anomalous precipitation from the effects of climate change means a greater frequency and magnitude of droughts. Excessive precipitation poses a significant flooding and hail risk for farmers, as both of these events severely damage corn crops. This risk will ultimately be passed on to the insurance companies who insure farmers through crop-hail insurance. Significant damage will also be made to infrastructure in the event of a flood if a community is not sufficiently prepared.

Drought poses a more sizable risk, as it would lead to water scarcity which not only impacts the agricultural industry but also the municipal water supply [10]. Crops cannot be grown without a steady supply of water, and people cannot survive without reliable drinking water. Water shortages will also have a negative impact on the local wildlife, which may lead to more intrusion of wildlife into urban areas, causing disruptions with vehicles.

Furthermore, the aridity caused by the variations in temperature significantly increase the risk of wildfires which can cause heavy damage to infrastructures such as those that California has dealt with in recent years if not properly prepared for. Spikes in heat can also leads to spikes in heat-related illness, exasperate the spread of infectious disease, and increase the amount of requests for emergency food assistance [10].

All of these are reasonable risks that must be considered in the face of highly volatile precipitation and rising temperatures.

Recommendations

Insurance Recommendations

In our recommendations to insurance companies, we will focus on the results of our models regarding expected losses of indemnities over the next 30 years until 2050. We begin by considering the graph of relative frequencies of predicted losses from our random walks, shown below:

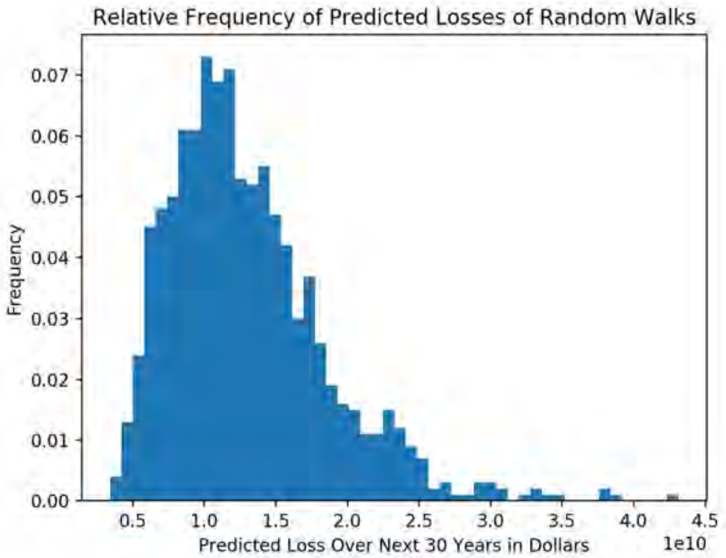

Figure 9: Histogram of predicted loss, in tens of billions USD, by random walks shows a right skew to larger values of loss. This suggests that increased climate change and volatility runs the risk of large, unprecedented losses in the next thirty years.

One of the first interesting aspects of our result is the skewness to the right of the figure. Specifically, this skewness to the right indicates that there is a chance of extremely large losses for farmers that would need to be covered by insurance companies. This risk can be quantified with the following two sample statistics from our model of indemnity based on average losses from a simulated random walk:

Mean Total Indemnity Loss over the next 30 Years	\$13.1 billion
Standard Deviation of Total Indemnity Loss over the next 30 Years	\$5.47 billion

Obviously, insurance companies need to be careful, so we consider the upper bound of a 95% confidence interval of losses over the next 30 years. It's important to acknowledge here that the data and simulated losses are indeed quite variant, but insurance companies are most likely interested in protecting themselves from loss, which is why we choose to consider the upper bound of such an interval. This upper bound comes out to approximately \$23.8 billion, which is thus a reasonable expectation of an upper bound to indemnities that insurance companies should expect to pay out in total over the next 30 years.

With the effects of climate change becoming more and more dire each year, insurance providers will need to re-evaluate their risks and potentially raise premiums in order to keep up with the higher probability of disaster. Some types of insurance, like crop-hail insurance, will be especially susceptible to the increase in frequency of abnormal weather events brought on by an increase in overall temperature. Atmospheric phenomenon are all driven by heat, so an increase in it will mean more frequent events such as hail, droughts, and thunderstorms, which all have the potential to wreak havoc on corn crops.

Our model of simulated PDSI data to 2050 (Figure 6) supports the idea of volatility increasing in the years to come, with the extrapolated data demonstrating clear increases in maxes and minimums of the data, and in general deviations from the mean. We are also inclined to believe in the strength of this model, as if we consider the extrapolated data in the same color as historical data as shown below:

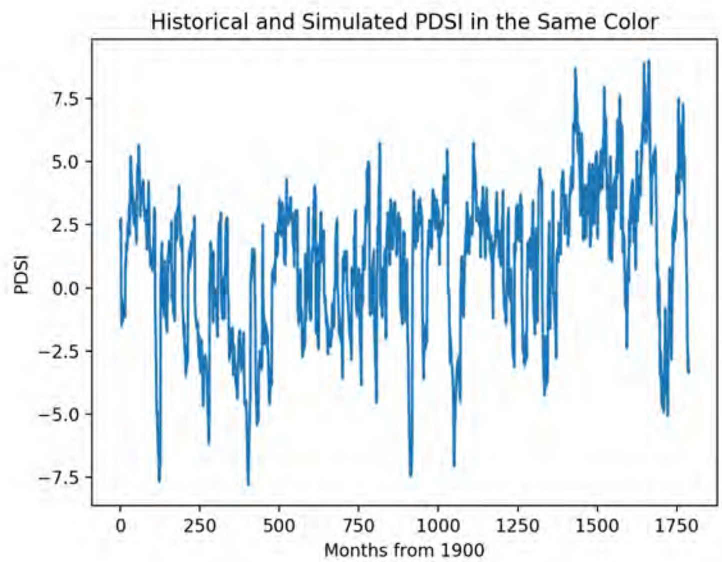

Figure 10: Combined PDSI time series (historical and simulated) suggests an increasing volatility of PDSI values in the future, especially regarding more wet PDSI values.

The graph is especially notable in that a new high for precipitation is achieved at least 6 different instances (with respect to the historical highs) — the severity of such severe spikes precipitation without proper readiness systems very likely leading to immediate crop damage. Insurance companies should not expect these events of severe precipitation to continue over longer periods of time, as even in our model, they only appear in spikes rather than large plateaus. Thus overall, it is highly likely that crop-hail insurers will have to raise premiums in order to cover the higher costs of the abnormal weather events. Consequently, crop-revenue insurance will be heavily impacted, as an increase in the number and severity of weather events directly correlates to a greater volatility in the supply and demand schemes of the crop market. The purpose of crop-revenue insurance is to protect agricultural workers against the volatility of said market, thus the increased fluctuations in crop yield due to the changing temperature and its direct correlation with the volatile nature of weather events suggest that crop-revenue insurance will need to be wider spread to be able to cater to these larger fluctuations, hence forcing insurance companies to either raise the premiums for an increased coverage, or decrease the amount of loss in market fluctuation that they are willing to cover for agricultural workers.

Since our temperature forecast (Figure 4) does not accurately model the extremely cold months, extra caution is needed to prepare for those cold months. For insurance providers, this means reevaluating risks of crop failures due to extreme cold events such as frosts. At the same time, the possibility of extreme hot weather events such need to be factored in as well. Overall, this would mean increasing premiums for crop-hail insurance providers. Indeed, it behooves insurance companies for using dynamic values for premiums that incorporate these discrepancies. Perhaps in areas where there are temperature levels (high or low) that cannot be accurately captured by a least-squares regression, insurance companies should adjust their premiums to ensure that they do not suffer unprecedented losses.

Public Policy Recommendations

In our public policy and governmental regulation recommendations we will focus on mitigating the factors that actually result in crop damage and hence indemnities for insurance companies. While the two factors with which we focused our definition of climate change on are temperature and precipitation, the issue of temperature is often times out of the scope of public policy to deal with, so instead we focus on the infrastructural solutions that can be proposed to combat the problems that arise with precipitation and water in general.

Aside from its direct impact in the form of rain on crop production, water supply also affects crop production in a region due to its effect on irrigation systems. Specifically, in regions where irrigation systems are developed to support the current level of crop production, irrigation systems that face a lack of water will not be able to successfully distribute water to crops, directly leading to crop damage and decreased yield [23]. One way to mitigate the effects of volatile precipitation and severe weather events, such as droughts, is to build a more resilient water supply. The state of Minnesota could build additional canals and wells in order to provide a sustainable supply water during times of drought. Investments could also be made into drip irrigation systems, which have been shown to be much more effective than traditional flood irrigation [24]. By investing in the water supply, as well as further conserving the use of water, the effects of droughts on agricultural production and urban life could be largely mitigated.

Floods must be considered as well, as they can wreak havoc on unprepared communities. Structural mitigation, such as the construction of dikes and levees, as well as non-structural mitigation, such as moving people away from flood-prone areas, are both needed to reduce the effect of floods [25] [26]. These techniques must be used selectively in high flood risk areas of the state to minimize flood damages as well as reduce costs.

Now aside from the direct mitigation of precipitation, public policy could potentially be beneficial if we consider the economic aspects of the corn and insurance markets. Specifically, one of the two major types of insurance as discussed in the background information is crop-revenue insurance, which is dictated not by how much the farmers produce, but rather how much revenue they can generate off of it.

Because of the increased volatility in the production of corn that is directly correlated with the increased volatility of PDSI indexes that we see in the extrapolated portion of the graph of PDSI's, corn revenue is susceptible to severe price fluctuation. The government should make active efforts to protect the economy by educating the relevant producers about the necessities and benefits of hedging prices in the commodity market.

For example: farmers grow corn but risk that the price of corn will have declined by the time they are ready to sell them. Farmers could hedge the risk by selling corn futures, which lock in a price for their corn early in the growing season. In general, we would recommend diversification of the market with other crops that may be more resilient to climate change, as farmers are incentivized to create optimized growing patterns for different crops that are more likely to survive in the more extreme scenarios of the current climate, than hedge their bets on crops with a smaller ideal growth condition interval and smaller variety of crops [20].

Further Recommendations

In the face of a warming climate, corn will be more and more difficult to grow. Instead of changing the climate to adapt to the corn, we could change the corn to adapt to the climate. Already 92% of corn grown in the United States is genetically modified [27], so it should not be too difficult to invest in the creation of a new strain of corn that is more resilient within a warming climate [28].

While this is an interesting and increasingly more novel approach to this issue, there are several ways communities themselves can still participate in the mitigation of climate change and harm to the economy. Communities that are on the frontlines of climate change in Minnesota (such as farmers) stand to suffer the most. Communities that depend heavily on resources that are projected to change with the climate ought to proactively take steps to identify sources of alternatives or sources that can provide stability to income, basic needs (drinking water, healthcare assistance, etc). Electric and gas companies should take steps to mitigate the risks of wildfires immediately, and increase monitoring of their systems to locate the existences of wildfires in a quicker amount of time.

Finally, there are still politicians in Congress that deny the existence of climate change, ignoring the warnings and undisputed logic of many scientists. To accomplish effective change from the national level, communities should leverage the power of social media which has dramatically increased in influence in the past decade to counter disinformation about climate change and to spread awareness about it.

Appendix

Contributions

Authors are listed in alphabetical order in the title page and contributions section.

- Models were developed by BG and EZ.
- Technical computing was performed by BG and EZ.
- Background research was performed by PA and RL.
- Risk analysis and recommendations were performed by PA, RL, and EZ.
- Manuscript was assembled by PA and BG.

Acknowledgments

We would like to thank Professor Herschel Day of the University of Wisconsin Eau-Claire, our actuarial mentor, for all of his guidance in the formation of this paper. The work would not be as sophisticated as it is now without his instrumental feedback.

We would also like to thank our AP Statistics teacher Mr. Joe Kim of Monta Vista High School, our project mentor, who taught us many of the fundamentals used in this report.

Finally, we would like to thank The Actuarial Foundation for putting together the Modeling The Future Challenge that drove our research. Participating in the MTFC has allowed us to explore new topics in data science and develop a greater understanding of the nuances, impact, and importance of actuarial work. We strongly believe that this experience has significantly impacted all of our future career goals, and hope that both more students become aware of this competition, and that it continues to provide the same thrilling experiences to students in the many years to come!

Technical Computing

All of the modeling methodologies are performed using Python 3.x in a standard Python environment. All of the code scripts can be found in this Github link: https://github.com/bg49623/modeling_corn. The majority of computation was performed in Python base packages, but a few classical frameworks (Numpy 1.18, TensorFlow 2.x, Pandas, Matplotlib 3) were used for matrix algebra, gradient descent, data manipulation, and figure creation.

References

[1] J. Urton, "Climate Change Projected to Boost Insect Activity and Crop Loss."

[2] USDA-NASS, "2018 State Agriculture Overview." https://www.nass.usda.gov/Quick_Stats/Ag_Overview/stateOverview.php?state=Minnesota.

[3] USDA-NASS. https://www.nass.usda.gov/Charts_and_Maps/Crops_County/cr-pr.php, 2018.

[4] D. Larson, "Crop Insurance in Hutchinson and Glencoe, MN." https://www.insuranceservicesminnesota.com/products/farm-insurance/crop-insurance.

[5] S. Blair, "Crop-Hail Insurance: How Does it Work?." https://www.tciteam.com/crop-hail-insurance-how-does-it-work/.

[6] J. Kagan, "Crop-Hail Insurance." https://www.investopedia.com/terms/c/crophail-insurance.asp.

[7] M. Battaglia, C. Lee, W. Thomason, J. Fike, and A. Sadeghpour, "Hail Damage Impacts on Corn Productivity: A Review," *Crop Science*, vol. 59, pp. 1–14, 12 2018. 10.2135/cropsci2018.04.0285.

[8] UNL, "Top Seven Factors In Crop Production." https://extension.unl.edu/statewide/gage/Top%20Seven%20Factors%20in%20Crop%20Production.pdf.

[9] UMA, "Minnesota Climate Report." https://www.geo.umass.edu/climate/stateClimateReports/MN_ClimateReport_CSRC.pdf\.

[10] NCA, "Fourth National Climate Assessment - Chapter 24: Northwest." https://nca2018.globalchange.gov/chapter/24/.

[11] USDA-RMA, "Cause of loss historical data files." https://www.rma.usda.gov/Information-Tools/Summary-of-Business/Cause-of-Loss.

[12] NASS, "Quick Stats ." https://quickstats.nass.usda.gov/.

[13] NOAA, "Climate Data Online." https://www.ncdc.noaa.gov/cdo-web/.

[14] J. Cowles, B. Boldgiv, P. Liancourt, P. S. Petraitis, and B. B. Casper, "Effects of increased temperature on plant communities depend on landscape location and precipitation," *Ecology and Evolution*, vol. 8, no. 11, pp. 5267–5278, 2018.

[15] S. Zhu, E. K. Nyarko, and M. Hadzima-Nyarko, "Modelling daily water temperature from air temperature for the missouri river," *PeerJ*, vol. 6, p. e4894, June 2018.

[16] Z. Wang, P. Li, L. Li, C. Huang, and M. Liu, "Modeling and forecasting average temperature for weather derivative pricing," *Advances in Meteorology*, vol. 2015, pp. 1–8, 2015.

[17] "Global Climate Change Impacts in the United States 2009 Report Legacy site." https://nca2009.globalchange.gov/corn-and-soybean-temperature-response/index.html.

[18] Wang, Zhiliang, Li, Peng, Li, Huang, Chunyan, and Liu, "Modeling and forecasting average temperature for weather derivative pricing," Mar 2015.

[19] R. R. Marathe and S. M. Ryan, "On the validity of the geometric brownian motion assumption," *The Engineering Economist*, vol. 50, pp. 159–192, Apr. 2005.

[20] N. Diffenbaugh, T. Hertel, M. Scherer, and M. Verma, "Response of Corn Markets to Climate Volatility under Alternative Energy Futures." https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3519383/.

[21] A. Mazhirov, "Climate Change to Exacerbate Rising Food Prices." https://blogs.ei.columbia.edu/2011/03/22/climate-change-to-exacerbate-rising-food-prices/.

[22] "Minnesota Agriculture: The Foundation of Minnesota's Economy." https://www.leg.state.mn.us/docs/2008/other/080928.pdf, 2007.

[23] J. Elliott, "Constraints And Potentials Of Future Irrigation Water Availability On Agricultural Production Under Climate Change." https://www.pnas.org/content/pnas/111/9/3239.full.pdf.

[24] F. R. Lamm and T. P. Trooien, "Subsurface drip irrigation for corn production: a review of 10 years of research in kansas," *Irrigation Science*, vol. 22, p. 195–200, Jan 2003.

[25] "6. Flood Barriers in Defensive Strategies," *Security of Flood Defenses*, p. 61–67, Aug 2019.

[26] Q. Fan and M. Davlasheridze, "Flood risk, flood mitigation, and location choice: Evaluating the national flood insurance programs community rating system," *Risk Analysis*, vol. 36, p. 1125–1147, Sep 2015.

[27] B. Plumer, "How widespread are GM foods?," July 2015.

[28] R. Raman, "The impact of genetically modified (gm) crops in modern agriculture: A review," *GM Crops Food*, vol. 8, p. 195–208, Feb 2017.