Crop Failures from Temperature and Precipitation Shocks: Implications for U.S. Crop Insurance

Molly Sears* Jeffrey M. Perloff† Wolfram Schlenker‡ Ximing Wu§

November 10, 2025

Keywords: crop yield, risk, insurance, temperature, precipitation, conditional distribution

Abstract

The effects of global warming on crop yield risk are critically important to U.S. agriculture, particularly to crop insurance programs. We introduce a nonparametric model, using a copula density approach, to construct flexible conditional yield distributions given temperature and precipitation. This approach facilitates probabilistic predictions of quantities such as the probability of crop disasters and large crop insurance payouts in response to temperature and precipitation shocks. We use our model to estimate the probability of a yield shortfall. By combining our estimated conditional distribution with projected climate data, we simulate the probability of catastrophic yields in response to global warming. Our approach has two advantages over the traditional approaches. First, our nonparametric, copula approach allows us to estimate

*Michigan State University, Department of Agricultural, Food, and Resource Economics
†University of California, Berkeley, Department of Agricultural and Resource Economics
‡Harvard University, Kennedy School of Government, NBER and CEPR
§Texas A&M University, Department of Agricultural Economics

complex, flexible interaction effects of temperature and precipitation. Second, because we know the full distribution, we can coherently examine the effects on not only mean yields as in regression analyses, but also the effects on the probability of disastrous outcomes, variance, skewness, and other risk measures.

1 Introduction

Many articles show that temperature and precipitation, as well as their links to drought and humidity, have a complex, nonlinear relationship with crop yield and quality (Peng et al., 2004; Schlenker and Roberts, 2009; Welch et al., 2010; Fezzi and Bateman, 2015; Tack et al., 2015; Zhang et al., 2015; Kawasaki and Uchida, 2016; Eck et al., 2020; Li et al., 2021; dos Santos et al., 2022; Perry et al., 2020; Boyer et al., 2023; Sumner et al., 2025). For example, the APSIM crop model in Lobell et al. (2013) offers physiological explanations for why extreme heat has a stronger impact on crop yields than precipitation. Both temperature and precipitation affect the water balance of a crop, but extreme heat affects it through multiple pathways, amplifying its impact. Precipitation primarily supports crop growth by replenishing soil moisture. In contrast, extreme heat negatively affects crop productivity in two ways: it accelerates soil moisture loss through both evaporation and plant transpiration, and it increases plants' water requirements to maintain the same level of carbon uptake. These dual effects of extreme heat, on both water supply and plant demand, make yields more sensitive to temperature spikes than to changes in precipitation, which primarily influences water supply alone. Accurately capturing the complex interplay between temperature and precipitation is therefore essential for understanding crop responses.

It is generally possible to adapt to extreme heat. However, farmers achieve a lower sensitivity to extreme heat at the cost of a lower average yield. Schlenker et al. (2013) found that for warming by 2°C, the beneficial effects of a lower sensitivity to extreme heat is roughly offset by the loss in average yield. This result aligns with our expectations, as the envelope theorem suggests that for sufficiently small changes, the first-order effect of temperature on yields accurately approximates the total effect once the system has adapted. Altogether, there are three major components to track in this relationship: (i) how crop yields respond to gradual, long term changes in temperature and precipitation, (ii) how crop yields respond to extreme changes in temperature and precipitation, and (iii) the distribution of crop yield responses to changes in these climatic variables. While the first two components

are relatively well-studied, the third requires more attention.

Recent work has increasingly turned to copula-based models to examine the joint effects of climate variables on crop yield and price risk. Several studies have used copulas to characterize the dependence between climate extremes, such as drought and heat, and agricultural yields. For example, Alidoost et al. (2019) employ copulas to explore the joint distribution of climate variables, crop yields, and prices, highlighting the importance of multivariate approaches to understand agricultural risk. Similarly, Li et al. (2021) and Ribeiro et al. (2019) apply copula frameworks to evaluate the effects of drought on crop yields in China and Southern Europe, respectively. These studies underscore how copulas allow for flexible modeling of nonlinear dependencies. Gaupp et al. (2017) extend this to the spatial dimension, showing how copulas can capture spatial dependence in simultaneous crop failures across major wheat-producing regions, while Leng and Hall (2019) use a variety of copula families to assess drought impacts on global staple crops. In addition to modeling average relationships, copulas are particularly useful for capturing tail dependence, the tendency of extreme values in one variable to occur alongside extreme values in another. Du et al. (2018) emphasize this in their work on yield resilience, and Hochrainer-Stigler et al. (2019) show how copulas can simulate extreme drought risk under climate change.

An extensive literature examines how to incorporate knowledge about yield-weather relationships into designing superior crop insurance programs. For example, see Annan et al. (2014), Tack et al. (2018), Maestro et al. (2016), Belasco et al. (2020), Bucheli et al. (2022), and Regmi et al. (2023). Copulas have also been widely used to inform agricultural insurance. Goodwin and Hungerford (2015) assess copula-based approaches to modeling systemic risk for crop insurance and reinsurance purposes, while Ghosh et al. (2011) model the joint distribution of price and yield for revenue insurance pricing. Finally, Bokusheva (2018) examines whether weather-yield dependence is stable over time using Bayesian copulas.

Our work contributes to a growing literature that uses copula-based models to understand joint dependencies between climate variables and agricultural yields. While several studies

have estimated unconditional or joint distributions of yields and weather (e.g., Alidoost et al. 2019; Ribeiro et al. 2019; Du et al. 2018), our approach is distinct in two ways. First, we focus on estimating the full conditional distribution of crop yields given temperature, precipitation, and other agronomic and environmental variables. This allows us to simulate yield outcomes under specific climate scenarios and investigate the likelihood of catastrophic crop failures under future conditions. Where conventional regression methods estimate conditional means, and quantile regressions estimate target specific quantiles, our copula-based framework enables a more flexible and complete characterization of the distribution, from central tendencies to extreme tail risks.

Second, and most importantly, we employ nonparametric copula estimation, which avoids the limitations of traditional parametric copulas used in much of the existing literature. Parametric copulas, such as those used by Ribeiro et al. (2019) and Du et al. (2018) impose functional forms that potentially constrain the ability to capture complex, nonlinear dependence structures. In contrast, our nonparametric approach provides a data-driven representation of the joint distribution, allowing for greater flexibility in modeling the interactions between weather extremes and yield responses. This flexibility is crucial for uncovering hidden or asymmetric relationships, particularly in the tails of the distribution where risk is most concentrated.

Our approach yields internally consistent estimates for a range of risk measures, including mean, variance, skewness, tail probabilities, and yield thresholds. These quantities are derived from a single joint distribution, ensuring coherence across moments. In contrast, estimating separate regressions for each moment can lead to logically inconsistent outcomes. For example, Chebyshev's inequality places a mathematical bound on tail probabilities based on the variance, but this relationship can break down if variance and tail risk are modeled separately. By producing consistent risk metrics, our model offers a reliable foundation for risk management and policy design.

These insights have practical implications for the design of climate-resilient crop insur-

ance programs, including Area Yield Protection and weather-indexed insurance. Our results complement earlier work on improving insurance design under climate risk (e.g., Annan et al. 2014; Tack et al. 2018; Maestro et al. 2016) by offering a forward-looking, nonparametric framework for evaluating how warming and weather volatility may influence both average losses and extreme outcomes.

Our paper proceeds as follows: Section 2 details the modeling approach, Section 3 describes the data, Section 4 shows the model output and compares its features to a standard linear model, Section 5 presents results of the future simulations and implications for crop insurance policy, and Section 6 concludes.

2 Copula Density Approach

Estimating multivariate or conditional distributions can be difficult. Researchers have developed various strategies to address this challenge. A common approach is to assume a parametric form, such as the multivariate Gaussian distribution. While reasonable, such models impose strong structural assumptions—including symmetry, elliptical contours, and tail independence—which may not reflect the true nature of agricultural data. Alternatively, nonparametric methods, such as kernel density estimation, offer greater flexibility but suffer from the curse of dimensionality, especially when modeling multiple interacting climate variables (Scott, 2015).

Because of these challenges, much of the literature focuses on estimating specific features of the conditional distribution rather than the full distribution itself. A common example is the use of regression models to estimate the conditional mean and conditional variance, as in the Just and Pope (1979) production function framework. Others have employed quantile regression to estimate conditional quantiles of yield (Koenker and Bassett Jr., 1978). While it is theoretically possible to recover a full distribution from a complete set of quantiles, it is difficult in practice: estimated quantiles often violate monotonicity and cannot be guaranteed

to form a valid distribution function (He, 1997; Chernozhukov et al., 2010).

We use a conditional density approach, which does not require separate models for the mean, median, percentiles, or higher moments. Instead, it estimates them in a unified model. We calculate these moments based on a common conditional distribution, so they are consistent with each other. In contrast, when one uses separate regression analyses, the results may conflict. The copula approach provides a flexible way of constructing or estimating multivariate densities. The separation of the marginal and copula densities makes the estimation problem easier, allowing the use of a "divide and conquer" approach, which mitigates the curse of dimensionality. Second, copula densities are often easier to estimate than joint densities because this approach removes the variations associated with marginal densities.

Let Y denote a univariate random variable (e.g., log yield), and let $X = (X_1, X_2, \ldots, X_d)$ be a d-dimensional vector of environmental and agronomic variables (e.g., temperature, precipitation, and soil texture). Let $F(y, x)$ denote the joint cumulative distribution function (CDF) of (Y, X), and $f(y, x)$ its associated joint density. The marginal distributions are $F_Y(y)$ for yield and $F_j(x_j)$ for each component X_j, with corresponding marginal densities $f_Y(y)$ and $f_j(x_j)$ for $j = 1, \ldots, d$.

By Sklar's Theorem (Sklar, 1959), the joint distribution can be expressed as a copula composition:

$$F(y, x) = C\left(F_Y(y), F_1(x_1), \ldots, F_d(x_d)\right), \tag{1}$$

where C is a copula function that captures the dependence structure between the variables. Differentiating both sides with respect to y and x yields the joint density:

$$f(y, x) = f_Y(y) f_1(x_1) \ldots f_d(x_d)\, c\left(F_Y(y), F_1(x_1), \ldots, F_d(x_d)\right), \tag{2}$$

where $c(\cdot)$ is the copula density, which is a density function on the unit hypercube $[0, 1]^{d+1}$ with uniform margins.

The copula formulation allows the joint density to be decomposed into marginal densities and a copula density, enabling a flexible, modular approach to multivariate density estimation. Let $f_X(x_1, \ldots, x_d)$ be the joint density of X. Then, Sklar's Theorem implies:

$$f_X(x_1, \ldots, x_d) = f_1(x_1) \ldots f_d(x_d)\, c_X\left(F_1(x_1), \ldots, F_d(x_d)\right), \tag{3}$$

where c_X is the copula density associated with X. Using Equation (1), the conditional density of Y given $X = x$ is:

$$f(y \mid x) = \frac{f(y, x)}{f_X(x)} = f_Y(y)\, \frac{c\left(F_Y(y), F_1(x_1), \ldots, F_d(x_d)\right)}{c_X\left(F_1(x_1), \ldots, F_d(x_d)\right)}. \tag{4}$$

Because $c_X(F_1(x_1), \ldots, F_d(x_d))$ is constant with respect to y, we can write:

$$f(y \mid x) = a_0\, f_Y(y)\, c\left(F_Y(y), F_1(x_1), \ldots, F_d(x_d)\right), \tag{5}$$

where

$$a_0 = \left[c_X\left(F_1(x_1), \ldots, F_d(x_d)\right)\right]^{-1} \tag{6}$$

is a normalization constant ensuring that $f(y \mid x)$ integrates to 1.

In contrast, conventional kernel-based methods often estimate the conditional density via the ratio:

$$\hat{f}(y \mid x) = \frac{\hat{f}(y, x)}{\hat{f}_X(x)}, \tag{7}$$

where $\hat{f}(y, x)$ and $\hat{f}_X(x)$ are nonparametric estimators of the joint and marginal densities. However, as Zellner (1978) and others have shown, the presence of a random variable in the denominator can render this ratio estimator unstable. This can lead to moments that do not exist, or a distribution that may be bimodal or even explosive.

The copula-based estimator avoids this problem by estimating the conditional density from the marginal and copula components. In our application, we estimate both the marginal

density $f_Y(y)$ and the copula density $c(F_Y(y), F_1(x_1), \ldots, F_d(x_d))$ using the multivariate Exponential Series Estimator (ESE), a method particularly well-suited for flexible copula density estimation (Wu, 2010; Gao et al., 2015; Chang and Wu, 2015).

3 Data

We illustrate the usefulness of the copula density approach by estimating the conditional distributions of U.S. corn and soybean yields, given temperature and precipitation. Our analysis focuses on the primarily unirrigated agricultural regions east of the 100th meridian. To do this, we combine publicly available county-level yield data with finely resolved historical and projected weather data.

We evaluate model performance by comparing root mean square out-of-sample prediction errors (RMSE) from both linear and nonlinear approaches. For the linear model, we calculate RMSE as the square root of the average squared difference between predicted and actual yields. For the nonlinear model, we use a similar procedure, but the predicted yield is taken from the "nearest" temperature–precipitation combination in a discretized output grid. Conditional mean yields are computed on a temperature–precipitation grid at 49 quantiles, ranging from the 2nd to the 98th percentile in 2% increments. We use these quantiles to construct the temperature–precipitation grid because it balances coverage of the data with statistical stability. The most extreme values in the tails are rare, which makes conditional mean estimates at those points noisy and unreliable in a copula framework. Trimming to the 2–98% range avoids sparse tail regions and boundary effects in density estimation, while still covering almost the full observed climate space.

County-level yield data for corn and soybeans are obtained from the National Agricultural Statistics Service (NASS) for the years 1950–2016. NASS also provides state-level weekly data on the share of planted and harvested acreage. We use these data to define each state's growing season: it begins at the start of the first week when 50% of the crop is planted and

ends at the conclusion of the first week when 50% is harvested. To avoid endogeneity due to adaptive planting or harvesting behavior in response to weather, we fix the growing season to its average start and end dates over the 1950–2016 period.

Our historical weather dataset extends the Schlenker and Roberts (2009) dataset to 2016. The data are based on a consistent set of approximately 2,000 weather stations with daily observations of minimum and maximum temperature. We interpolated missing values using an inverse-distance weighted average of percentile ranks across nearby stations. Specifically, for each weather station and variable, we first estimate its cumulative distribution function (CDF). If a station has a missing value on a given day, we compute the inverse-distance weighted average percentile from the other stations' values that day and use the corresponding percentile of the missing station's CDF to impute the value. For instance, if the interpolated percentile is 80%, we substitute the 80th percentile of the missing station's CDF as the estimated value. Cross-validation exercises, where observed values are omitted and then interpolated using this method, show that it performs well for temperature, which tends to be spatially smooth. Importantly, maintaining a fixed set of weather stations ensures that our estimates are not affected by changes in station coverage over time, which is critical in a panel setting where we rely on deviations from long-term averages.

We assign each 2.5×2.5-mile PRISM grid to its 10 nearest weather stations. We then regress monthly PRISM values on the corresponding station data to obtain station-specific weights for each grid cell. Using these weights, we construct daily minimum and maximum temperatures for each grid cell. We then fit a sinusoidal curve between each day's minimum and maximum to approximate within-day temperature distributions. To generate county-level measures, we average across grid cells within a county using satellite-derived land-use weights, which are fixed over time.

Prior work has established that the yield–weather relationship is highly nonlinear, especially with respect to temperature. Thus, constructing weather variables on a fine spatial grid before aggregating to the county level is essential. For example, Tack et al. (2015) demon-

strated that models that average temperatures before applying nonlinear transformations perform worse than those that apply transformation prior to averaging. For precipitation, which is more localized and harder to interpolate at the daily level, we instead use monthly totals from the PRISM Climate Group.

For future projections, we use data from the NASA NEX-GDDP dataset[1], which provides daily, bias-corrected minimum and maximum temperatures and precipitation from 21 CMIP5 climate models at a 0.25° spatial resolution. The dataset spans a historical period from 1950–2005 and includes projections for 2006–2099 under RCP 4.5 and RCP 8.5 scenarios. We apply the same processing steps to the projected data as with the historical data: we fit sinusoidal temperature curves for each day, compute growing degree days, and aggregate to the county level based on the agricultural area that lies within each grid cell whose centroid falls inside the county boundary.

4 Copula Density Analysis

Previous research identified extreme heat as the most reliable predictor of corn and soybean yields, particularly in dryland regions where irrigation cannot offset heat-related damage (Schlenker and Roberts, 2009). Our model incorporates climate effects using a flexible specification that allows for rich interactions among variables. However, the curse of dimensionality limits our ability to model numerous interactions simultaneously. Consequently, we concentrate on the interaction between extreme heat and precipitation, motivated by prior findings that water availability may buffer the harmful effects of high temperatures (Ortiz-Bobea et al., 2019).

Throughout the analysis, our temperature variable refers specifically to extreme heat degree days, or the number of degree days above 29°C (84.2°F) for corn and 30°C (86°F) for soybeans, measured during the middle third of the growing season, when crops are most vulnerable to heat stress.

[1] `https://www.nccs.nasa.gov/services/data-collections/land-based-products/nex-gddp`

We use a two-step estimation strategy. In the first step, we regress the logarithm of county-level yield, temperature, and precipitation on county fixed effects, county-specific quadratic time trends, and year fixed effects. This removes confounding variation due to spatial and temporal heterogeneity. We experimented with including additional variables, such as moderate growing degree days (days between 10°C and 29/30°C), but found they had minimal impact. In the second step, we use the residuals from the first regression as inputs to estimate the conditional density of yield, allowing for flexible interactions between extreme heat and total seasonal precipitation.

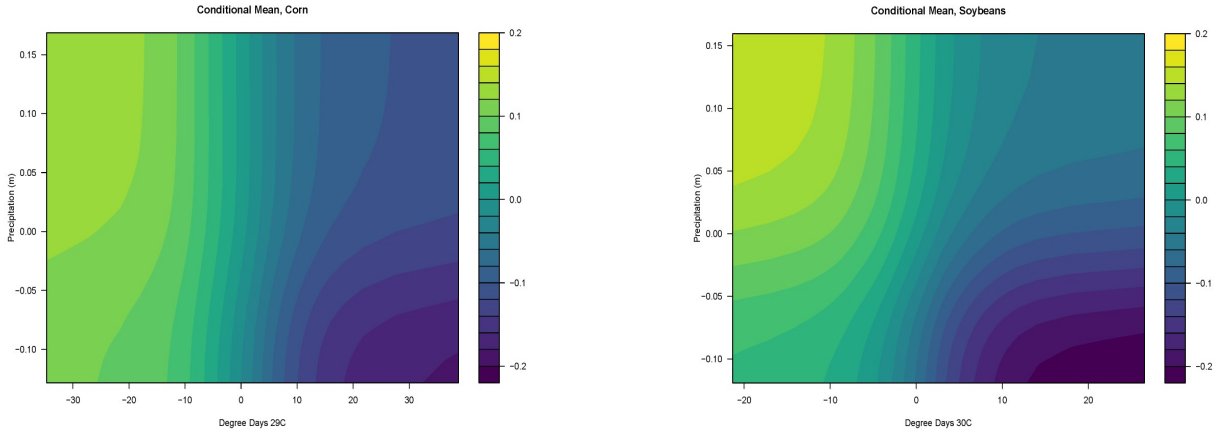
4.1 Contour Maps

The figures below present contour plots of the conditional mean, standard deviation, coefficient of variation, and probability of a yield shortfall for corn and soybeans during the middle third of the growing season. In a standard linear regression, these contours would appear straight and parallel, reflecting additively separable effects. In contrast, our nonlinear copula-based approach produces curved, non-parallel contours, illustrating the advantages of modeling flexibility.

4.1.1 Conditional Mean

Figure 1 displays contour maps of the conditional mean log yield for corn and soybeans as functions of temperature and precipitation, restricted to counties east of the 100th meridian where irrigation is uncommon. Estimates for "hot" and "cool" counties, classified by historical average temperatures, are shown in Figure 2. Because the first-stage regression removes fixed effects, both temperature and precipitation are measured relative to county-level historical averages. For example, a precipitation value of -0.1 denotes 10 cm less rainfall than the historical average for that county during the growing season.

The contour maps reveal that the lowest average yields for both crops occur under hot and dry conditions. For corn, yields peak under relatively cool temperatures regardless of rainfall.



(a) Corn yield

(b) Soybean yield

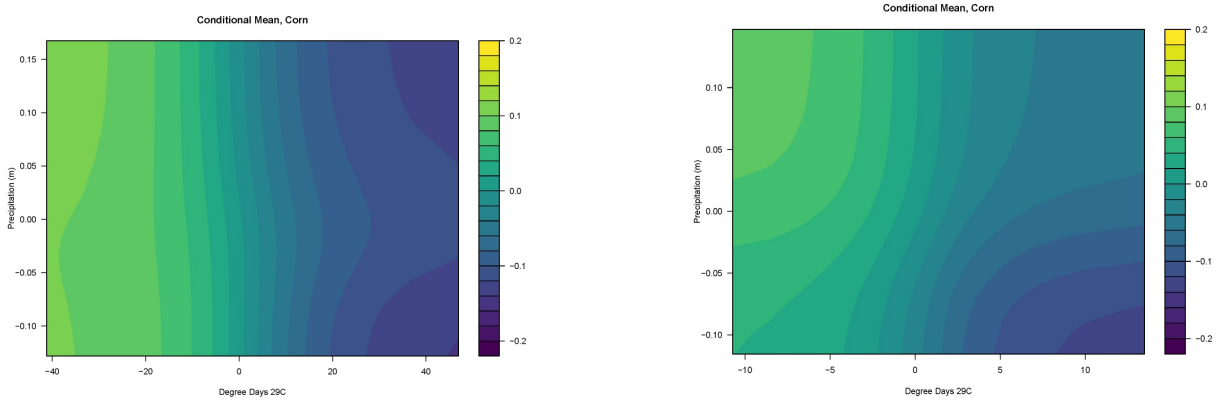
Figure 1: Conditional mean for corn yield (a) and soybean yield (b).

In contrast, soybean yields are highest when rainfall is abundant and temperatures are low, indicating a stronger interaction between temperature and moisture availability. These results highlight the potential bias from imposing linear or additively separable functional forms, given the clear crop-specific and nonlinear weather–yield relationships.

For corn, we further divide the sample into two groups based on the historical average growing-season temperature of each county: the hotter third of observations ("hot" counties) and the colder third ("cool" counties). This approach captures persistent regional climate differences rather than short-term weather variability. Figure 2 presents conditional mean contour plots for each group. Hot counties exhibit pronounced sensitivity to high temperatures regardless of precipitation, indicating that moisture cannot fully offset heat stress. In contrast, cool counties are less sensitive to relatively high temperatures, and abundant precipitation can mitigate much of the yield loss caused by warming.

4.1.2 Conditional Standard Deviation

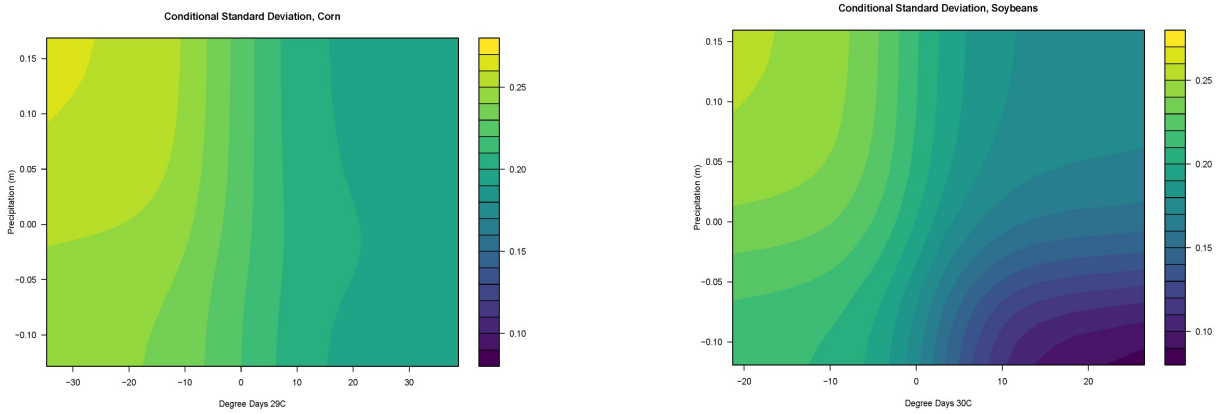
Figure 3 shows the conditional standard deviation of yield, an absolute measure of production risk. For both crops, the standard deviation peaks under cooler, wetter conditions, while extreme heat—though damaging to yields—is associated with lower variability, particularly



(a) Hot counties (b) Cold counties

Figure 2: Conditional mean for corn yield in hot counties (a) and cold counties (b).

in corn. This suggests that extreme heat reduces uncertainty by consistently producing poor outcomes. Precipitation plays a limited role in variability except under cooler conditions.



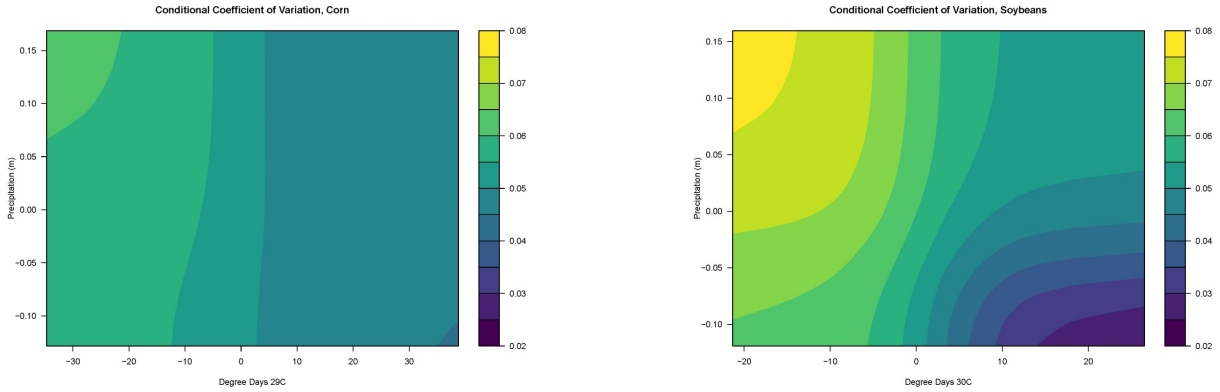
(a) Corn yield (b) Soybean yield

Figure 3: Conditional standard deviation for corn yield (a) and soybean yield (b).

Soybeans display a more complex pattern. High temperatures combined with low precipitation yield both low mean yields and low variability, whereas cool, wet conditions produce the highest variability. Across the full range of conditions, corn exhibits a higher maximum standard deviation than soybeans, indicating greater absolute production risk.

4.1.3 Coefficient of Variation

Figure 4 presents the coefficient of variation (CV), or the standard deviation normalized by the mean. This captures relative rather than absolute risk. Patterns broadly mirror those of the standard deviation but differ in scale. In corn, relative risk at high temperatures is largely unaffected by precipitation. In soybeans, both extremes of temperature and precipitation can raise relative risk, and the magnitude of these fluctuations can be larger (or smaller) than those seen in corn, reflecting different crop responses to weather stress.



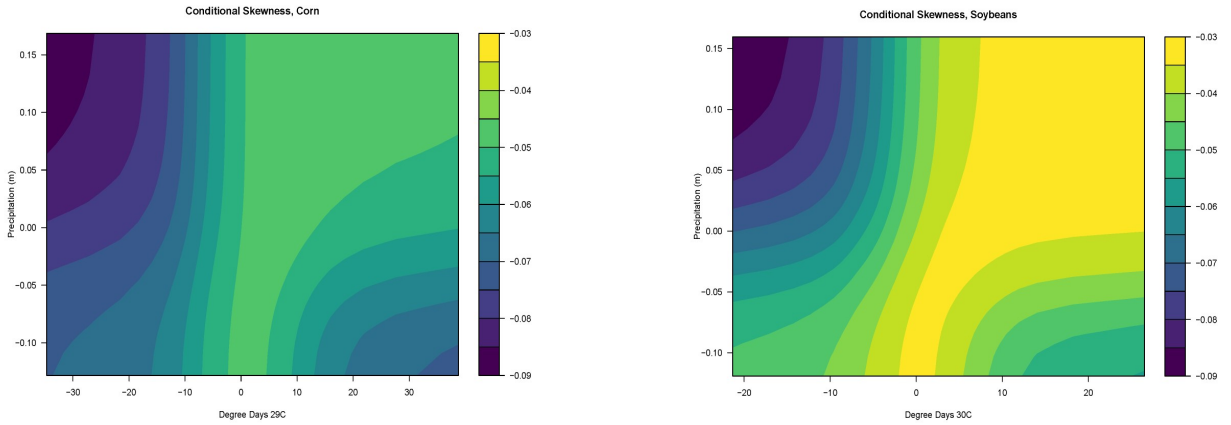
(a) Corn yield (b) Soybean yield

Figure 4: Conditional coefficient of variation for corn yield (a) and soybean yield (b).

4.1.4 Conditional Skewness and Studentized Skewness

Higher moments offer further insight into yield distribution shapes. Figure 5 reports conditional skewness estimates for both crops. Skewness is predominantly negative, indicating a heavier lower tail (more frequent low-yield outcomes), and this negative skew intensifies under weather extremes, particularly under high temperatures with low precipitation and low temperatures with high precipitation.

To evaluate how weather affects crop insurance programs, we estimate the probability of extreme yield shortfalls. Our model predicts the likelihood that yields fall below critical thresholds. In U.S. crop insurance, an example of where this estimation is particularly



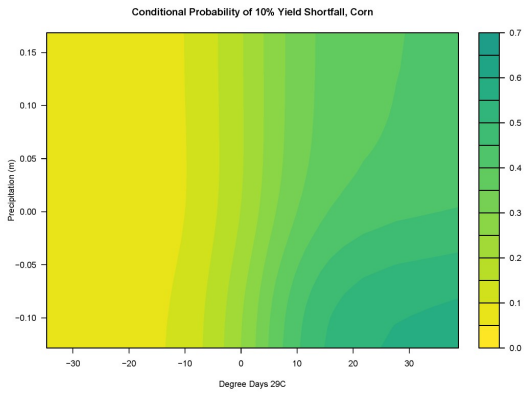
(a) Corn yield

(b) Soybean yield

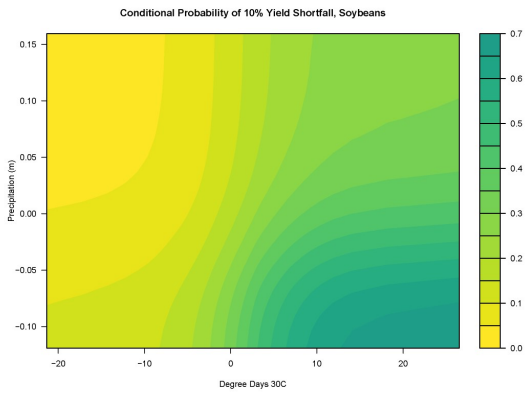
Figure 5: Conditional skewness for corn yield (a) and soybean yield (b).

relevant is Area Yield Protection (AYP), which compensates producers when county-level yields fall below a selected coverage level, typically ranging from 70% to 90% of the county's historical average yield. While farm-level Revenue Protection and Yield Protection policies now dominate the market, county yields (and yield shortfalls) remain central to many crop insurance products. They inform premium rating, validate farm-level yields for revenue and yield protection policies, and trigger supplemental area-based coverage like SCO and ECO. Our estimates of extreme county yield shortfall probabilities are therefore relevant well beyond AYP.

We estimate the probability that county-level yields drop at least 10% below the historical average (i.e., to 90% or less of the benchmark). Figure 6 shows that while the spatial patterns of shortfall probability are similar for corn and soybeans, the probability is generally higher for soybeans, underscoring subtle but important differences in vulnerability between the crops.



(a) Corn yield



(b) Soybean yield

Figure 6: Conditional probability of a 10% yield shortfall in corn yields (a) and soybean yields (b).

5 Dependence Model Fit

5.1 Comparisons to Linear Model

We evaluate in-sample and out-of-sample predictive accuracy of our nonlinear copula model relative to a standard linear regression of yield on temperature and precipitation. Using 14 randomly selected years between 1950 and 2016, we predict conditional mean county yields and compute root mean squared error (RMSE) for both models.

Our nonlinear model uses a grid of temperature and precipitation quantiles (2% to 98% in 2% steps) to estimate conditional means by numerical integration, while the linear model directly predicts conditional means. The RMSE results, summarized in Table 1, show the nonlinear model performs slightly better in-sample for corn (RMSE 0.0415 vs. 0.0428) and equivalently for soybeans (0.0285 for both). The linear model outperforms the nonlinear in out-of-sample corn predictions (0.0471 vs. 0.0557), but differences are small, indicating comparable predictive capability. We observe no significant sensitivity to the choice of temperature–precipitation grid for the linear model, reinforcing the robustness of these findings.

Table 1: RMSE for in-sample and out-of-sample predictions of conditional mean county yields

	In-Sample		Out-of-Sample	
Crop	Linear	Nonlinear	Linear	Nonlinear
Corn	0.0428	0.0415	0.0471	0.0557
Soybeans	0.0285	0.0285	0.0288	0.0376

5.2 Yield Shortfall Prediction

We further evaluate how well the nonlinear copula model predicts yield shortfalls, defined as yields 10% below historical means. Using data from 1950–2002 for estimation, we predict shortfalls for 2003–2016 and compare predictions to actual shortfall occurrences. There were 4,782 actual shortfall events and 16,469 non-shortfall observations in this period.

As shown in Table 2, the nonlinear model identifies 2,060 of the shortfalls but also predicts many false positives (3,015). The linear model identifies fewer shortfalls (1,568) but also has fewer false positives (1,829). This indicates the nonlinear model is somewhat more sensitive to shortfalls, though at the cost of specificity.

Table 2: Prediction of yield shortfalls, 2003–2016

	Linear Model		Nonlinear Model	
	Shortfall	No Shortfall	Shortfall	No Shortfall
Predicted Shortfall	1,568	1,829	2,060	3,015
Predicted No Shortfall	3,214	14,640	2,722	13,454

5.3 Probability of Large Losses

To examine how the nonparametric copula performs under particularly poor yield events, we focus on counties with yield losses exceeding 20% below historical averages during 2003–2016. For each yield shortfall event, we compute the model's predicted probability of a shortfall and average these probabilities over all such events. We repeat this for periods without shortfalls. The model predicts, on average, a 21.56% chance of shortfall during

actual shortfall years/counties, compared to only 13.42% during non-shortfall years/counties. This demonstrates that the nonparametric copula effectively quantifies the increased risk of extreme adverse outcomes, a feature unavailable in mean-based linear models.

In sum, our findings support the presence of tail dependence in crop yield–weather relationships. Relative to linear models, the nonlinear copula approach delivers comparable predictive accuracy for mean yields and adds unique capacity for predicting rare, extreme events. To evaluate the robustness of our copula framework, we are extending the analysis to compare the nonparametric copula with benchmark Gaussian and Student's t copulas using out-of-sample validation, with an 80/20 train-test split. Preliminary results suggest that the t copula performs best in capturing tail dependence, while the nonparametric copula offers flexibility that improves upon the Gaussian in many cases. Full model comparisons are ongoing.

6 Simulation of Climate Change Impacts on Crop Yields

To assess the potential impact of global warming on crop yields, we use our estimated model to simulate the effects of simultaneous shocks to temperature and precipitation. These shocks are derived from scientific climate projections, focusing on two representative concentration pathways (RCPs): RCP 4.5, representing a moderate, low-emission scenario, and RCP 8.5, representing a high-emission scenario. The resulting simulations provide valuable inputs for analyzing agricultural risk and informing crop insurance program design under changing climate conditions.

6.1 Historical Baseline: Probability of Yield Shortfalls (1980–2005)

We first evaluate the baseline probability of a yield shortfall by county during the historical period of 1980–2005. For each county and year, we calculate the probability that the

average county yield falls at least 10% below its historical mean. These probabilities are then averaged across all years to produce a spatial map of average shortfall risk, as depicted in Figure 7. Results show that southern counties generally exhibit higher probabilities of yield shortfalls during this period. However, several counties in Wisconsin and Michigan also faced relatively elevated risks. It is important to note that these probabilities represent county-level average yields and a relatively severe threshold (10% below the historical mean). Individual farms may experience higher shortfall probabilities due to local variability.

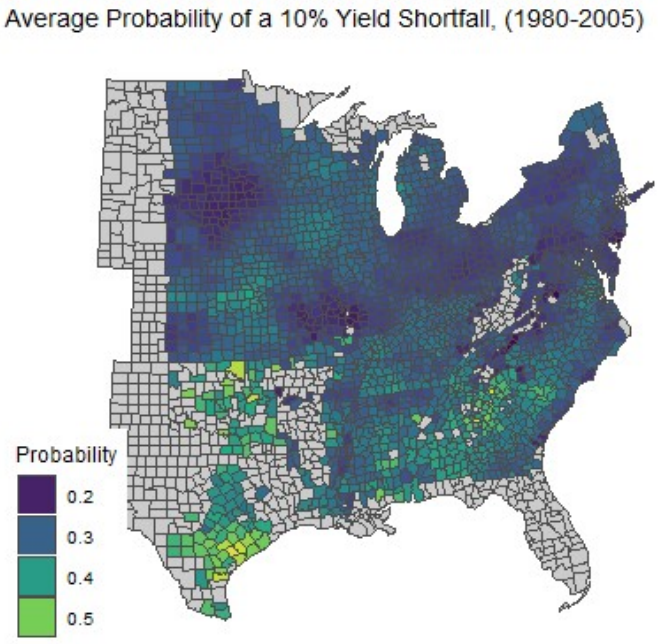
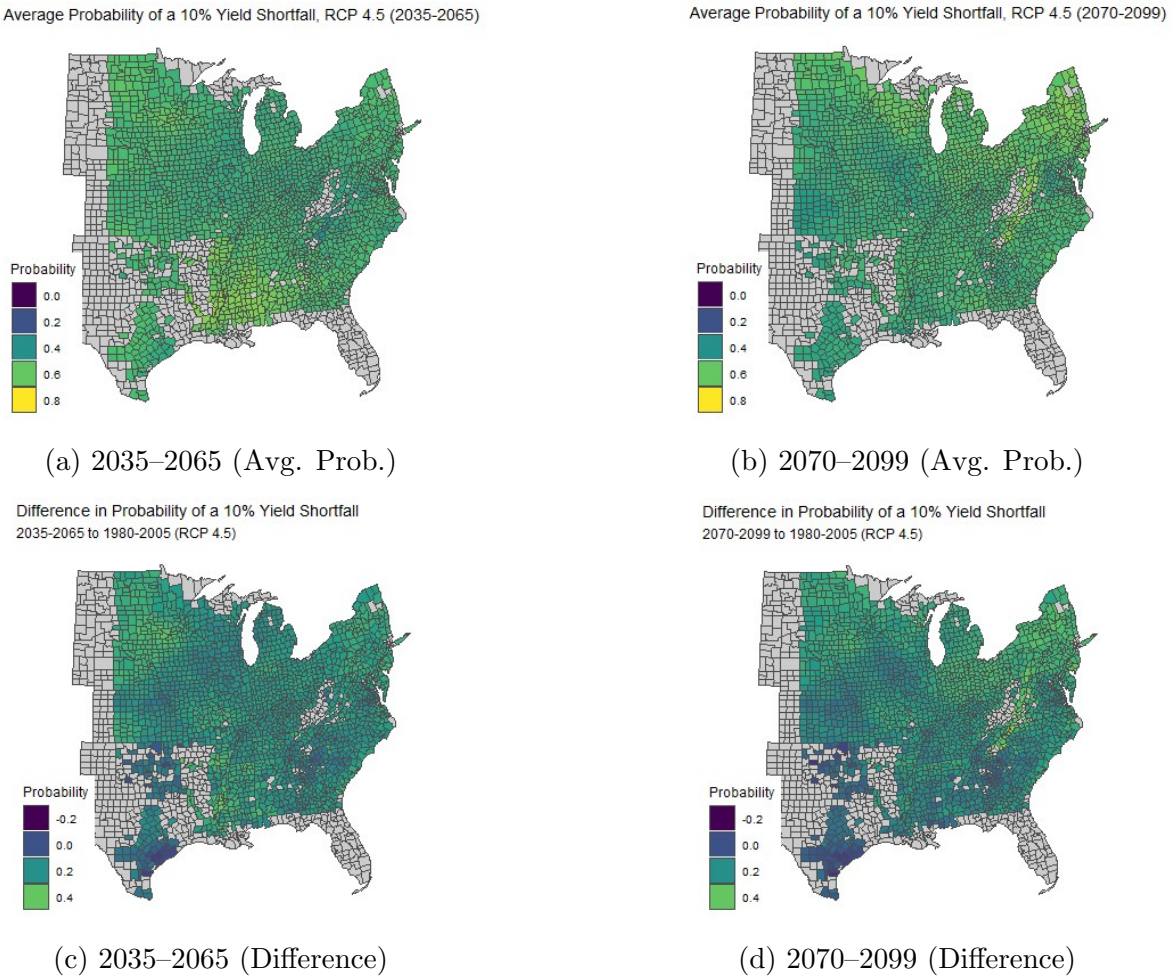


Figure 7: Average probability of a 10% yield shortfall in corn, 1980–2005.

6.2 Future Projections under RCP 4.5 (2035–2065 and 2070–2099)

Next, we project yield shortfall probabilities under the moderate emissions scenario (RCP 4.5) for two future periods: mid-century (2035–2065) and late-century (2070–2099) in Figure 8. These simulated shortfalls are measured relative to the same historical county means to enable consistent comparison. Our results indicate a general increase in the probability of yield shortfalls across the region compared to the historical baseline. Notably, northern counties are projected to experience larger increases in shortfall risk by the end of the century,

suggesting that climate change may exacerbate adverse yield outcomes even in areas currently less vulnerable.



(a) 2035–2065 (Avg. Prob.)

(b) 2070–2099 (Avg. Prob.)

(c) 2035–2065 (Difference)

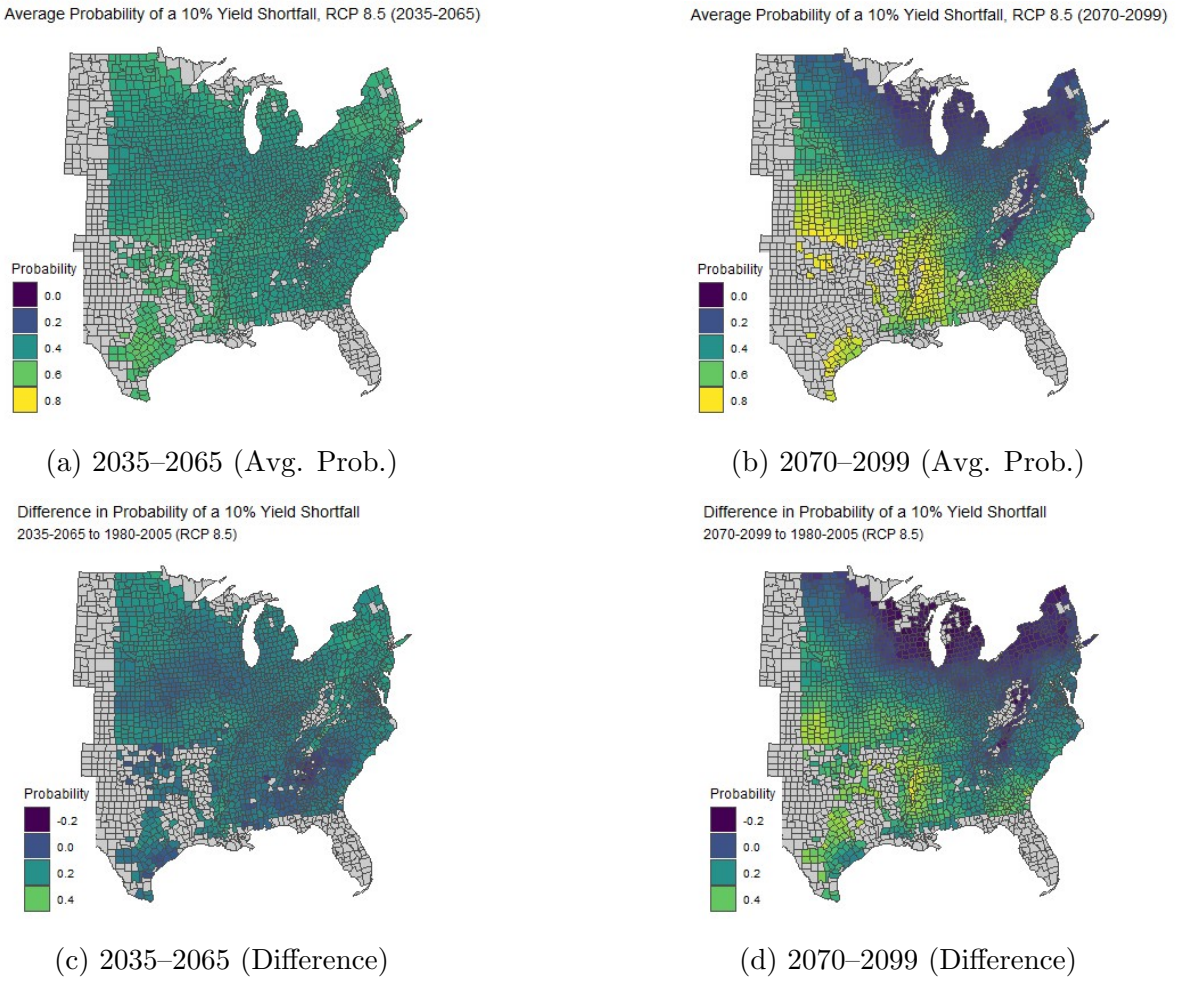
(d) 2070–2099 (Difference)

Figure 8: (a,b) Average probability of a 10% yield shortfall in corn under RCP 4.5 for two future periods, and (c,d) difference in probability relative to 1980–2005.

6.3 Future Projections under RCP 8.5 (2035–2065 and 2070–2099)

Figure 9 illustrates the analogous projections under the high-emission scenario (RCP 8.5). By mid-century, we observe a widespread increase in yield shortfall risk across nearly all counties, reflecting the more severe climate shocks anticipated under this pathway. By the late-century period, the spatial pattern shifts: southern counties become the hardest hit, with substantially elevated probabilities of yield shortfalls, while northern counties appear

to experience some improvement relative to the historical baseline. This divergence aligns with findings in the broader climate and agricultural economics literature, which suggests potential northward shifts in suitable growing conditions under extreme warming scenarios.



(a) 2035–2065 (Avg. Prob.)

(b) 2070–2099 (Avg. Prob.)

(c) 2035–2065 (Difference)

(d) 2070–2099 (Difference)

Figure 9: (a,b) Average probability of a 10% yield shortfall in corn under RCP 8.5 for two future periods, and (c,d) difference in probability relative to 1980–2005.

Our simulation results underscore the heightened risks that climate change poses to agricultural productivity, with important spatial heterogeneity depending on emissions trajectories and time horizons. These insights highlight the need for regionally tailored risk management strategies and insurance products capable of accommodating changing yield risk profiles in a warming world.

7 Conclusion

Our findings demonstrate that a nonparametric copula model effectively captures the complex, nonlinear relationships between temperature, precipitation, and crop yields across different crops and regions. This framework allows us to move beyond modeling only conditional means or variances and instead characterize the entire conditional distribution of yields, including its higher moments and tail behavior. Using the estimated model, we simulate the potential effects of climate change on yield shortfall risks under two representative concentration pathways (RCP 4.5 and RCP 8.5). These simulations indicate that the probability of yield shortfalls is likely to rise across most of the region, with particularly large increases projected in northern counties under the moderate emissions scenario and in southern counties under the high-emissions scenario by late century. The spatial heterogeneity in these projections underscores the importance of modeling nonlinear dependencies between temperature and precipitation, as well as the value of distributional information for assessing agricultural climate risks.

While this approach performs comparably to traditional linear models in predicting conditional mean yields, its real strength lies in its ability to estimate the probability of large yield losses, which is outside the scope of a linear model. Preliminary validation using an 80/20 train–test split against other copulas suggests that tail dependence is an important feature of the yield–weather relationship: the Student's t copula generally achieves the highest log-likelihood, particularly during colder periods and earlier decades, while the nonparametric copula improves upon the Gaussian but does not consistently surpass the t. Ongoing work extends this comparison across additional crops and model specifications to better understand the trade-offs between flexibility and statistical efficiency.

This modeling capability has important practical implications for policymakers and stakeholders in agricultural risk management. For instance, crop insurance programs could leverage farm- or field-specific weather data integrated with our model to predict the likelihood that a grower experiences a yield shortfall relative to county averages. Such probabilistic

assessments can help insurers more accurately price premiums, forecast indemnities, and anticipate evolving risk patterns under changing climate conditions. Similarly, the ability to characterize yield distribution tails provides growers and policymakers with richer information on downside risks and the potential for extreme losses, which is increasingly critical as climate variability intensifies.

Looking forward, a promising extension of this research would be to apply our modeling framework to field-level yield and weather data. This would enable direct comparison of model-predicted yield shortfalls with actual insurance indemnity payouts, offering a robust test of predictive accuracy at finer spatial scales. Moreover, exploring the model's applicability to other crop insurance products, such as Prevent Plant, could yield valuable insights. If early-season weather indicators reliably forecast yield shortfall probabilities at the end of the growing season, growers could use this information to make more informed planting and management decisions under uncertainty.

Overall, our nonparametric modeling approach offers a flexible and powerful tool for quantifying agricultural climate risks, supporting better-informed decision-making in a world of growing environmental variability and economic uncertainty.

References

Alidoost, F., Su, Z., and Stein, A. (2019). Evaluating the effects of climate extremes on crop yield, production and price using multivariate distributions: A new copula application. *Weather and Climate Extremes*, 26:100227.

Annan, F., Tack, J., Harri, A., and Coble, K. (2014). Spatial pattern of yield distributions: Implications for crop insurance. *American Journal of Agricultural Economics*, 96(1):253–268.

Belasco, E. J., Cooper, J., and Smith, V. H. (2020). The development of a weather-based crop disaster program. *American Journal of Agricultural Economics*, 102(1):240–258.

Bokusheva, R. (2018). Using copulas for rating weather index insurance contracts. *Journal of Applied Statistics*, 45(13):2328–2356.

Boyer, C. N., Park, E., and Yun, S. D. (2023). Corn and soybean prevented planting acres response to weather. *Applied Economic Perspectives and Policy*, 45(2):970–983.

Bucheli, J., Dalhaus, T., and Finger, R. (2022). Temperature effects on crop yields in heat index insurance. *Food Policy*, 107:102214.

Chang, M. and Wu, X. (2015). Transformation-based nonparametric estimation of multivariate densities. *Journal of Multivariate Analysis*, 135:71–88.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

dos Santos, C. A. C. et al. (2022). Trends of extreme air temperature and precipitation and their impact on corn and soybean yields in nebraska, usa. *Theoretical and Applied Climatology*, 147:1379–1399.

Du, X., Hennessy, D. A., Feng, H., and Arora, G. (2018). Land resilience and tail dependence among crop yield distributions. *American Journal of Agricultural Economics*, 100(3):809–828.

Eck, M. A. et al. (2020). Influence of growing season temperature and precipitation anomalies on crop yield in the southeastern united states. *Agricultural and Forest Meteorology*, 291:108053.

Fezzi, C. and Bateman, I. (2015). The impact of climate change on agriculture: Nonlinear effects and aggregation bias in ricardian models of farmland values. *Journal of the Association of Environmental and Resource Economists*, 2(1):57–92.

Gao, Y., Zhang, Y., and Wu, X. (2015). Penalized exponential series estimation of copula densities with an application to intergenerational dependence of body mass index. *Empirical Economics*, 48:61–81.

Gaupp, F., Pflug, G., Hochrainer-Stigler, S., Hall, J., and Dadson, S. (2017). Dependency of crop production between global breadbaskets: a copula approach for the assessment of global and regional risk pools. *Risk Analysis*, 37(11):2212–2228.

Ghosh, S., Woodard, J. D., and Vedenov, D. V. (2011). Efficient estimation of copula mixture model: an application to the rating of crop revenue insurance. Technical report.

Goodwin, B. K. and Hungerford, A. (2015). Copula-based models of systemic risk in us agriculture: Implications for crop insurance and reinsurance contracts. *American Journal of Agricultural Economics*, 97(3):879–896.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, 51(2):186–192.

Hochrainer-Stigler, S., Balkovič, J., Silm, K., and Timonina-Farkas, A. (2019). Large scale extreme risk assessment using copulas: an application to drought events under climate change for austria. *Computational Management Science*, 16:651–669.

Just, R. E. and Pope, R. D. (1979). Production function estimation and related risk considerations. *American Journal of Agricultural Economics*, 61(2):276–284.

Kawasaki, K. and Uchida, S. (2016). Quality matters more than quantity: Asymmetric temperature effects on crop yield and quality grade. *American Journal of Agricultural Economics*, 98(4):1195–1209.

Koenker, R. and Bassett Jr., G. (1978). Regression quantiles. *Econometrica*, pages 33–50.

Leng, G. and Hall, J. (2019). Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. *Science of the Total Environment*, 654:811–819.

Li, Z., Zhang, Z., and Zhang, L. (2021). Improving regional wheat drought risk assessment for insurance application by integrating scenario-driven crop model, machine learning, and satellite data. *Agricultural Systems*, 19:1031–1041.

Lobell, D. B., Hammer, G. L., McLean, G., Messina, C., Roberts, M. J., and Schlenker, W. (2013). The critical role of extreme heat for maize production in the united states. *Nature Climate Change*, 3(5):497–501.

Maestro, T., Barnett, B. J., Coble, K. H., Garrido, A., and Bielza, M. (2016). Drought index insurance for the central valley project in california. *Applied Economic Perspectives and Policy*. Advanced Access.

Ortiz-Bobea, A., Wang, H., Carrillo, C. M., and Ault, T. R. (2019). Unpacking the climatic drivers of us agricultural yields. *Environmental Research Letters*, 14(6):064003.

Peng, S., Huang, J., Sheehy, J. E., Laza, R. C., Visperas, R. M., Zhong, X., Centeno, G. S., Khush, G. S., and Cassman, K. G. (2004). Rice yields decline with higher night temperature from global warming. *Proceedings of the National Academy of Sciences*, 101(27):9971–9975.

Perry, E. D., Yu, J., and Tack, J. (2020). Using insurance data to quantify the multidimensional impacts of warming temperatures on yield risk. *Nature Communications*, 11:4542.

Regmi, M., Tack, J., and Featherstone, A. M. (2023). Does crop insurance influence crop yield impacts of warming temperatures? a farm-level analysis. *Journal of the Agricultural and Applied Economics Association*, 2(4):808–822.

Ribeiro, A. F., Russo, A., Gouveia, C. M., and Páscoa, P. (2019). Copula-based agricultural drought risk of rainfed cropping systems. *Agricultural Water Management*, 223:105689.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to u.s. crop yields under climate change. *Proceedings of the National Academy of Sciences*, 106:15594–15598.

Schlenker, W., Roberts, M. J., and Lobell, D. B. (2013). Us maize adaptability. *Nature Climate Change*, 3(8):690–691.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8:229–231.

Sumner, E., Li, M., and Shr, Y. (2025). Is yield response enough? drought impacts on crop acreage throughout the production cycle. *American Journal of Agricultural Economics*.

Tack, J., Barkley, A., and Nalley, L. L. (2015). Effect of warming temperatures on us wheat yields. *Proceedings of the National Academy of Sciences*, 112(22):6931–6936.

Tack, J., Coble, K., and Barnett, B. (2018). Warming temperatures will likely induce higher premium rates and government outlays for the u.s. crop insurance program. *Agricultural Economics*, 49(5):635–647.

Welch, J. R., Vincent, J. R., Auffhammer, M., Moya, P. F., Dobermann, A., and Dawe, D. (2010). Rice yields in tropical/subtropical asia exhibit large but opposing sensitivities to minimum and maximum temperatures. *Proceedings of the National Academy of Sciences*, 107(33):14562–14567.

Wu, X. (2010). Exponential series estimator of multivariate densities. *Journal of Econometrics*, 156(2):354–366.

Zellner, A. (1978). Estimation of functions of population means and regression coefficients including structural coefficients: A minimum expected loss (melo) approach. *Journal of Econometrics*, 8(2):127–158.

Zhang, H.-L., Zhao, X., Yin, X.-G., Liu, S.-L., Xue, J.-F., Wang, M., Pu, C., Lal, R., and Chen, F. (2015). Challenges and adaptations of farming to climate change in the north china plain. *Climatic Change*, 129(1–2):213–224.