Heterogeneity, Climate Change, and Crop Yield Distributions: Solvency Implications for Publicly-Subsidized Crop Insurance Programs

Daniel Schuurman and Alan Ker

University of Guelph and Michigan State University

Abstract

Climate change continues to fuel concern about the future cost of publicly-subsidized crop insurance programs in developed nations. These changes in climate are expected to alter the upper and lower tails of crop yield distributions differently. This may best be captured by modeling the climate-yield relationship heterogeneously across different parts of the yield distribution. To this end, we consider a mixture model with the parameters expressed as nonparametric functions (to capture any non-linearities) of weather variables estimated by machine learning methods (neural net). By doing so, we are able to identify possibly heterogeneous effects of climate change on each component, the mixing probabilities, and thus all moments of the yield distribution. We find changing climate alters, quite significantly, the entire shape of the yield distribution. The overall probability of the lower tail tends to increase as temperatures rise, to the point where some yield distributions become positively skewed. Across a range of climate change scenarios, premium rates for fixed guarantees are expected to rise 20-66% relative to no climate change by 2040. However, if we allow the yield guarantees to also fall because of additional losses from climate change, premium rates (albeit not comparable given yield guarantees are different) increase notably smaller (6-14%) suggesting less solvency issues than first thought.

1 Introduction

Publicly-subsidized crop insurance programs indemnify producers for naturally occurring yield losses. These programs are the prevailing domestic agricultural policy in many developed nations (Bielza et al., 2007; Mahul and Stutley, 2010). They transfer large amounts of government monies, in the form of premium subsidies, to the agricultural production sector. With respect to the United States, producers receive roughly a 60% premium subsidy (Rosch, 2021; Ker et al., 2017) on most insurance products. Furthermore, large amounts of public monies are transferred to the private insurance industry to operate the program via an administrative and operating expense reimbursement as well as the asymmetric sharing of underwriting gains and losses. The U.S. program covers a wide range of crops including major grain crops (i.e. corn, wheat), tree crops (i.e. almonds, oranges), livestock, and specialty crops. In 2022, liabilities for all programs totalled $64.7 billion, premiums totalled $6.8 billion, and, subsidies totalled $4.1 billion. Over the past decade the costs of these programs has continued to rise, surpassing $80 billion. Given the size of public monies used to operate crop insurance programs, understanding how changing climate will effect yields and thus future public outlays is important.

The entire shape, not just the conditional mean, of the yield distribution is expected to change as climate alters the inputs to crop production. Using the law of minimum technology, Hennessy (2009) demonstrated how changes to inputs such as water supply could alter the skewness of a yield distribution. Predicting cotton yields under rising temperatures, Tack et al. (2012) found rising temperatures not only asymmetrically alter cotton yield distributions but do so differently both across and within states. Not surprisingly, modeling the relationship between weather and yield has recently been found to improve premium rate estimation (Ramsey, 2020; Liu and Ramsey, 2022; Tack et al., 2018, 2015). Thus, to predict the effect of climate change on insurance costs, we require a methodology that captures the possibly heterogeneous and non-linear effects of weather across the yield distribution. To this end, we consider mixture distributions with parameters estimated as nonparametric functions of weather variables via a neural network (machine learning methods) to model yields. While this represents a methodological contribution in general and specifically to the modeling of crop yields, this approach also provides a more realistic prediction of insurance losses

under climate change for any type of insurance (i.e. property, casualty, etc.). Furthermore, in a flavor similar to Ramsey (2020) and Liu and Ramsey (2022), we use cross-validation methods and Bayesian model averaging to reduce uncertainty in our predicted densities from the neural network.

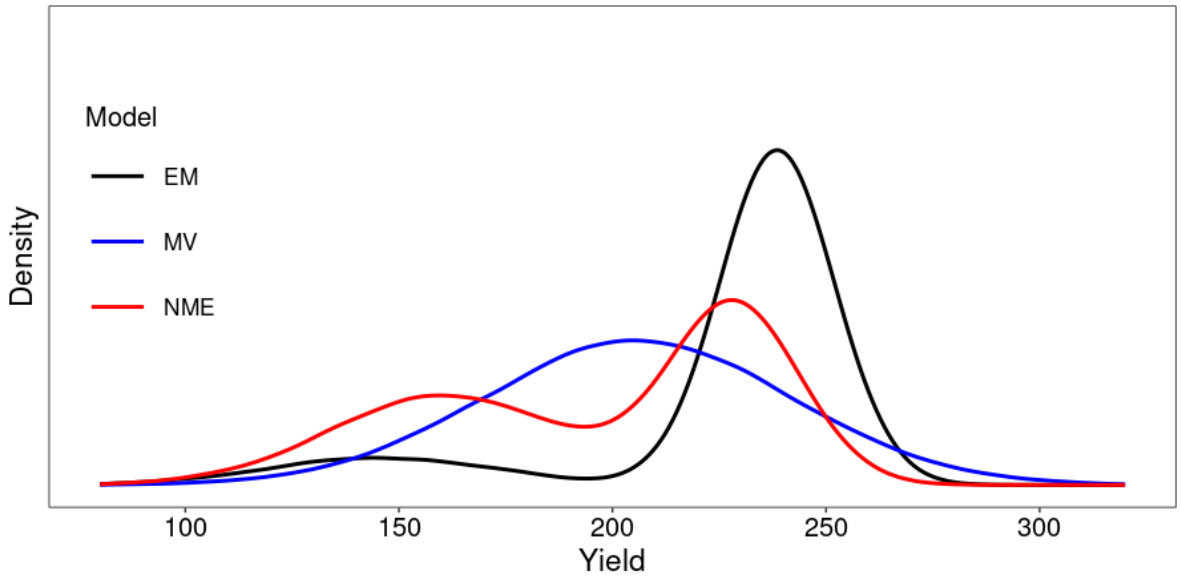
To give a sense of the economic importance of the methodological contribution that allows the climate-yield relationship to vary across the yield distribution when predicting future crop losses and associated public costs, we illustrate in Figure 1 predicted yield densities for corn in Wapello County, Iowa, using three different methodologies.[1] The first method, "Mean-Variance" (MV), is a normal distribution but with its mean and variance estimated as nonparametric functions of weather variables using a neural network. The second method, denoted *EM*, is a mixture distribution estimated using an Expectations-Maximization (EM) algorithm **without** conditioning on weather variables. Finally, the third method and method proposed in this manuscript, denoted *NME*, is also a mixture distribution but estimates the mixing probabilities (and by default the mixture parameters) as nonparametric functions of weather variables using a neural network nested within the EM algorithm.

The distinction between the *EM* and *NME* distributions illustrates the effect of climate. We see a significant shifting of probability to the lower tail caused by the yield-weather relationship in the *NME* method that is not possible with the *EM* method. This illustrates the significance of including weather in the parameter estimation process. The distinction between the *MV* and *NME* distributions illustrates the importance of accounting for the effect of climate change beyond the conditional mean of crop yield. As noted, the *NME* methodology adds significant probability to the lower tail without adding probability to the upper tail; this cannot be accommodated with the *MV* distribution.[2] Note, we do not impose any restrictions on the *NME* method that would cause this shifting of probability from the upper to the lower tail; this effect is completely driven by the yield-weather relationship as depicted in the historical data.

[1] Throughout the manuscript we use Wapello County as an illustrative example. Wapello was chosen for its representativeness; parameter estimates for Wapello are within a standard deviation of the means for all counties in Iowa.

[2] This effect will be most pronounced using a normal distribution as is done in Figure 1.

Figure 1: Predicted Conditional Densities of Wapello County in 2050 under RCP 4.5[3]



While differences in the predicted yield distributions are significant, the consequences for premium rates and thus costs of climate change via publicly subsidized crop insurance programs are also notable. The guarantee at the 90% coverage level under no climate change using the NME model is 189.9 bushels/acre for Wapello County with a corresponding premium rate of 4.2% in 2050 (see Figure 1). Under RCP 4.5, the premium rate for this same guarantee using the NME model is 16.7% representing an increase of 398%. If we allow the guarantee to shift downward with climate change, the guarantee for the NME would fall from 189.9 to 167.5 bushels/acre at the 90% coverage level.[4] The corresponding premium rate is 5.4% (a 29% increase in premium rates). All three rates are relevant in understanding the effect of climate change on premium rates. The first rate difference (4.2% to 16.7%) illustrates the effect of climate change relative to no climate change (two different yield distributions but identical guarantees) and thus represents the change in losses caused by a changing climate. Because the guarantees are equivalent the rates are directly comparable. The second rate difference (4.2% to 5.4%) allows the guarantee (expected yield) to shift downward with climate change as well. In this case, the guarantee for the 90% coverage

[3]Representative concentration pathway (RCP) 4.5 is a moderate global carbon emission scenario with emissions peaking in 2040 before declining.

[4]The yield guarantee for the insurance contract is the coverage level times expected yields. As losses increase, the expected yield necessarily decreases and thus the yield guarantee also decreases.

level fell 13% from 189.9 to 167.5 bushels/acre. The 5.4% premium rate is not comparable to the initial rate of 4.2% because of lower guarantees under climate change. Nonetheless, the premium rate the farmer would face in 2050 is 5.4% suggesting less of a solvency issue than first thought.

The objective of our manuscript is to predict changes in public crop insurance costs under various climate change scenarios. We use Iowa corn as this crop-region combination is most prevalent in the literature. We find that restrictions imposed on the yield-weather relationship by methodological choice can affect the results. A second contribution of our manuscript is methodological in nature; embedding a neural network inside the expectations-maximization algorithm notably increases flexibility in modeling the yield-weather relationship. Moreover, this latter contribution can be applied to any other type of insurance such as property, casualty, etc.

The remainder of the manuscript proceeds as follows. The next section discusses the relevant literature. The following two sections outline our data (including climate change scenarios) and estimation methods, respectively. The fifth section presents our estimation results, forecasts the yield distributions, and recovers future premium rates. The final section provides a summary of our findings and discusses the implications for the efficacy of publicly subsidized insurance programs.

2 Relevant Literature

Schlenker and Roberts (2006) and Deschênes and Greenstone (2007) introduced methods that empirically consider the effects of climate change on expected yield and profit respectively. Following these seminal works, others introduced new methods to control potential non-linear effects of weather variables (McIntosh and Schlenker (2006), Schlenker and Roberts (2009), and Roberts et al. (2013)) and model the effects of climate change on expected crop yields (D'Agostino and Schlenker, 2016; Schlenker and Roberts, 2009, 2006; Lobell et al., 2013; Keane and Neal, 2020a; Challinor et al., 2014; Gammans et al., 2017). However, in an insurance setting, identifying the effect of climate change on lower tail probabilities, not just the expected yield, is necessary.

Several authors considered multiple moments to model the climate-yield relationship over

the entire yield distribution. Tack et al. (2012) modeled the first three moments of cotton yields as functions of weather variables. They find high temperatures negatively affect all moments and moderate temperatures positively affect the first two moments. Malikov et al. (2020) used conditional quantile regression methods to examine the climate-yield relationship for corn and soybean. They find greater sensitivity to rising temperatures at lower quantiles. Ramsey (2020) introduced weather into the rate-making process with spatial quantile models but did not offer projections under future warming scenarios. Liu and Ramsey (2022) incorporate weather information into crop insurance estimation using a Bayesian approach but do not explicitly examine the climatic drivers of yield distributions or offer projections of insurance costs under climate change scenarios.

Others have considered the effect of insurance programs on farmer adaptation to climate change as insurance offsets some of the additional production risk caused by climate change. Annan and Schlenker (2015) found slower adaptation to rising temperatures for insured versus non-insured corn acreage. Chemeris et al. (2022) found a negative relationship between participation in crop insurance programs and adaptation to climate change, also using U.S. corn yields. Using partial moments of variance, Connor and Katchova (2020) found a relationship between greater crop insurance participation and increased downside corn and soybean risk.

Perhaps most similar to this manuscript, Tack et al. (2018) predicted rising temperatures to lower the mean and raise the variance of county corn yields. Under these changes, they projected premium rates at the 90% coverage level to rise by 39% across the Corn Belt under a 1° C warming. Their approach directly models the mixture parameters as functions of climate and indirectly — via potential clustering – the mixing weights. Conversely, our approach models the mixing weights directly as functions of climate and by default indirectly affects the estimated mixture parameters. We do this by drawing on features of mixture density networks (Bishop, 1994) and kernel mixture networks (Ambrogioni et al., 2017). Specifically, we use a neural network within the EM algorithm to capture expected non-linearities in the relationship between mixing probabilities and weather. While the EM algorithm is commonly used to estimate mixture models, using neural networks to capture non-linearities between latent variables (climate) and component probabilities has not been

previously considered.

Neural networks and other forms of machine learning are increasingly being used in crop yield prediction (Van Klompenburg et al., 2020). The semi-parametric neural network used by Crane-Droesch (2018) and the deep neural network used by Keane and Neal (2020b) are trained with mean-squared error to predict the conditional mean of corn yields. They both find improvement in out-of-sample prediction accuracy of expected yields compared to conventional methods. Despite the capacity of neural networks to approximate complex functions with limited functional restriction, conditional density estimation using neural networks has received limited attention.

A substantial portion of climate-yield models using econometric (Schlenker and Roberts, 2006, 2009; Lobell et al., 2013; Burke and Emerick, 2016; Hsiang, 2016; Roberts et al., 2013; Ramsey, 2020; Keane and Neal, 2020a; Tack et al., 2012, 2018) and machine learning approaches (Crane-Droesch, 2018; Keane and Neal, 2020b; Khaki et al., 2020) use panel data of county-level observations. By controlling for differences with fixed effects, there is an assumption that the climate-yield relationship only differs across counties in the conditional mean. Conversely, we estimate *separate* models for each individual county thus allowing for both distributional and climate-yield relationship differences *across* counties. However, this does not come without trade-offs, modeling each county individually while allowing heterogeneity does suffer from fewer observations to estimate the climate-yield relationship. These approaches are complementary as one estimates the average effect while the other estimates individual effects.

3 Data

We consider county-level corn yields for Iowa using data obtained from National Agricultural Statistics Service for 1951-2019. Ideally, more disaggregated data would be used but unfortunately, it does not exist over sufficient years required to undertake the empirical analyses. Iowa was chosen because it is the largest corn-producing state at 58.4 million tonnes (2020) and accounts for over 16% of US production (USDA, 2021). Also, corn yields in Iowa are particularly sensitive to rising temperatures and the greater evaporative demand expected under climate change (Schlenker and Roberts, 2006, 2009; Ortiz-Bobea et al., 2019; Phillips

et al., 1996; Ort and Long, 2014; Ting et al., 2021).

Historical daily weather data obtained for Iowa from 1950-2019 was constructed in the same manner as Schlenker and Roberts (2009).[5] The county-level weather data uses a balanced set of weather stations that limits variation due to station coverage (Auffhammer et al., 2013). Consistent with much of the previous work on climate change and yields, we use Growing Degree Days (GDD), Extreme Heat Degree Days (HDD), Vapor-Pressure Deficit (VPD), and precipitation as our weather variables over the April-September growing season (Tack et al., 2012, 2018; Ortiz-Bobea et al., 2019; Tolhurst and Ker, 2015; Schlenker and Roberts, 2009).

Growing degree days represent the accumulated, typically beneficial, exposure to heat over the growing season bounded between 10 and 29°C and high degree days represent damaging heat exposure above 29°C. Temperature exposure is estimated based on interpolations of daily minimum and maximum temperatures using a double sine curve following methods outlined by Ortiz-Bobea (2021). Vapor pressure deficit is based on diurnal temperatures and humidity; relating water and heat stress to plant physiology (Roberts et al., 2013; Lobell et al., 2013). Following Roberts et al. (2013), we define two variables, one for the entire growing season and one for July and August (VPD_{ja}). Linear and quadratic measures of cumulative precipitation in millimeters are included over the growing season and for July and August.

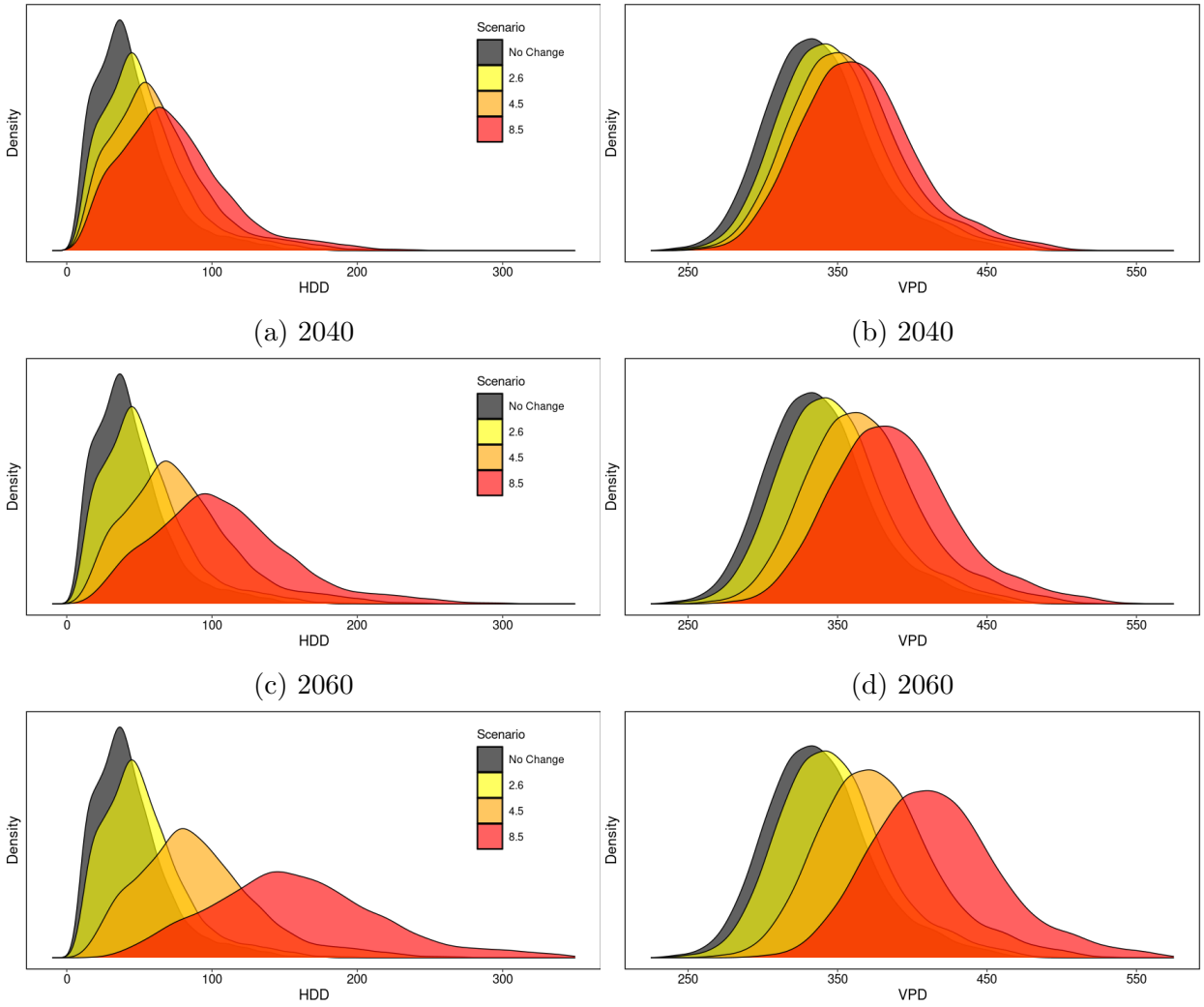
Climate change scenarios for 2020-2080 used for yield predictions in Section 5 are based on annual uniform shifts of daily maximum and minimum temperatures (see Lobell and Burke (2010); Tack et al. (2012, 2018); Roberts et al. (2017)). All temperature changes were based on average predictions under representative concentration pathways (RCP) 2.6, 4.5, and 8.5 for the Midwest by the National Oceanic and Atmospheric Administration (Vose and Wehner, 2017; Wuebbles et al., 2017). Relative to a 1976-2005 baseline, mean temperatures are predicted to rise 1.4°C under RCP 2.6 in the early century (2021-2050) before plateauing in the mid- to late-century. For mid-century (2036-2065) and late-century (2071-2100) temperatures are predicted to rise 2.34 and 3.09°C for RCP 4.5 and 2.94 and 5.27°C for RCP 8.5.

[5]The March 2020 version of the "Daily Weather Data for Contiguous United States (1950-2019)" dataset was obtained from http://www.columbia.edu/ ws2162/links.html

Table 1: Mean values of temperature-dependent weather variables under alternative climate scenarios

	2020	2040	2060	2080	2020	2040	2060	2080
	No Change				*RCP 2.6*			
GDD	1,714.7	1,714.7	1,714.7	1,714.7	1,714.7	1,784.9	1,784.9	1,784.9
HDD	38.5	38.5	38.5	38.5	38.5	47.4	47.4	47.4
VPD_{gs}	338.5	338.5	338.5	338.5	338.5	347.6	347.6	347.6
VPD_{ja}	135.7	135.7	135.7	135.7	135.7	139.3	139.3	139.3
	RCP 4.5				*RCP 8.5*			
GDD	1,741.8	1,851.4	1,938.5	1,999.0	1,776.1	1,912.9	2,070.7	2,250.7
HDD	42.0	57.4	72.5	84.6	46.7	68.1	101.2	152.9
VPD_{gs}	342.0	356.6	368.6	377.3	346.6	365.1	388.0	416.7
VPD_{ja}	137.1	142.8	147.5	150.8	138.9	146.1	155.0	166.2

We shift daily historical county-level temperatures for observations from 1976-2005 based on the slope of temperature changes between reference, early-, mid-, and late-century periods used by the NOAA (see Supplemental Appendix section 1.1 for additional detail). Temperature adjustments change both degree day and vapor pressure deficit variables (see Table 1). The distribution of weather events under climate change shifts upwards and exhibits greater inter-annual variation (see Figure 2).



(a) 2040 (b) 2040

(c) 2060 (d) 2060

(e) 2080 (f) 2080

Figure 2: Extreme heat degree days (left) and vapor pressure deficit (right) differences across all counties at 2040, 2060, and 2080 across four scenarios (No climate change and RCPs 2.6, 4.5 and 8.5)

4 Methodology

As previously noted, we use a Normal mixture to model yields similar to Ker (1996), Goodwin et al. (2000), Woodard and Sherrick (2011), Tolhurst and Ker (2015), and Ng and Ker (2021). Mixtures are an attractive option because of their flexibility; they can approximate any continuous distribution within a desired error. Our main purpose is to allow weather to alter

the probability of each mixture and affect all underlying parameters and moments of the estimated yield distribution. This section summarizes the expectation maximization with embedded trends introduced in Tolhurst and Ker (2015) before presenting the neural mixture estimation that introduces weather variables into the estimation. Finally, the competing mean-variance methodology is outlined.

Specifically, the three models considered are: (i) Mean-Variance (MV) model is $y \sim N(\mu, \sigma^2)$ where the parameters are estimated as nonparametric functions of weather variables via a neural network; (ii) Expectation-Maximization (EM) model is $y \sim \sum_{k=1}^{K} \pi_k N(\alpha_k + \beta_k * t, \sigma_k^2)$ where the parameters are NOT estimated as functions of weather variables; and (iii) Neural Mixture Estimation (NME) model is $y \sim \sum_{k=1}^{K} \pi_k(\omega_t) N(\alpha_k + \beta_k * t, \sigma_k^2)$ where the parameters are estimated as nonparametric functions of weather variables via a neural network. While the marginal yield distributions of the three models are nested, the models are not because the underlying parameter functions are non-nested. For the interested reader, we provide additional detail about the models in the Supplemental Appendix. The proposed methodology is the NME model which incorporates weather into the estimated parameters. The EM model is used as a comparison to the NME model to illustrate the effect of including weather into the estimation of the parameters. The MV model is used as a comparison to the NME model to illustrate the effect of capturing weather heterogeneously across the yield distribution.

4.1 Expectation Maximization (EM)

In the context of normal mixtures with embedded trends, the EM estimates the probability of each component (π_k); the intercept (α_k) and slope (β_k) of component k's trend; and the standard deviation (σ_k) of component k.

$$y_t = \sum_{k=1}^{K} \pi_k N(\alpha_k + \beta_k t, \sigma_k^2) \tag{1}$$

This is accomplished by iterating between "Expectation" and "Maximization" steps to maximize the likelihood function. Given initialized values for each of the four parameters for each component, the "Expectation" step calculates ρ_{kt}, the posterior probability of yield

observation at time t belonging to component k.

$$\rho_{kt} = \frac{\pi_k N(y_t; \alpha_k, \beta_k, \sigma_k)}{\sum_{l=1}^{K} \pi_k N(y_t; \alpha_k, \beta_k, \sigma_k)} \tag{2}$$

By construction, $\sum_k \rho_{kt} = 1$ for each observation t. The adapted "Maximization" step updates the parameters using maximum likelihood estimation and weighted least squares. Component probability is the average of respective posterior probabilities.

$$\pi_k = \frac{\sum_t \rho_{kt}}{T} \tag{3}$$

Instead of the weighted means approach used in the EM, weighted least squares are used to estimate the component-specific trends. The posterior probabilities of component k are the diagonal of the weight matrix (W) and X represents a T x 2 matrix of a constant and time trend.

$$[\alpha_k, \beta_k]' = (X'WX)^{-1}(X'Wy) \tag{4}$$

The homoskedastic variance is calculated as the weighted sum of the squared difference between the observation and fitted values from WLS.

$$\sigma_k^2 = \frac{\sum_t \rho_{kt}(y_t - \hat{y}_{kt})^2}{\sum_t \rho_{kt}} \tag{5}$$

The algorithm returns the updated parameters to the "Expectation" step until convergence of the log-likelihood function $\ell(y_t; \pi_k, \alpha_k, \beta_k, \sigma_k)$ is reached.[6]

4.2 Neural Mixture Estimation

To introduce weather into EM algorithm, we use a neural network within the "Expectation" step of the algorithm to estimate posterior probabilities each component as nonparametric functions of weather.[7] Our approach draws on features of mixture density networks (MDN) (Bishop, 1994) and kernel mixture networks (KMN) (Ambrogioni et al., 2017). In MDNs, all

[6] If the reader is interested in more detail concerning EM we direct them to the initial work by Hartley (1958) and Dempster et al. (1977) and, for greater detail, McLachlan and Krishnan (2007).

[7] See Figure 2.A in the Appendix for the architecture of a single layer feed-forward neural network used in this context.

parameters (μ_k, σ_k, π_k) of the k-component Gaussian mixture model are directly controlled by the neural network. In KMNs, the network only learns component mixture weights (π_k) with predetermined means μ_k and trainable scale parameters σ_k.

In the NME algorithm, the neural network directly controls the mixing weights of the components but the cost function is the negative log-likelihood.[8] The trend and standard deviation parameters are found using maximum likelihood estimation as previously outlined in the "Maximization" step of the EM. This procedure reduces issues of over-fitting and mode collapse that arise with simultaneous training of all distribution parameters in MDN (Cui et al., 2018; Makansi et al., 2019; Rothfuss et al., 2019) and is less restrictive in parameter selection than the KMN.[9]

Estimation of the technological trend, the average change in yield attributed to time, outside of the neural network is advantageous because: (i) it allows for standard estimates of intercepts and slopes of technology within the distribution unrecoverable within neural networks; and (ii) only weather information is used to estimate component probabilities ensuring the network is not using time to over-fit and capture non-technological processes.[10]

In general, the input layer for a network consists of a set of J variables, $X_t = (x_{1t}, x_{2t}, ..., x_{Jt})$ for each year t. The hidden layer contains H neurons, where each is a linear combination of the weighted sum of the inputs. For example, the first neuron in the hidden layer, h_1, is a the linear combination defined as: $w_{1,0} + w_{1,1}x_{1,t} + w_{1,2}x_{2,t} + ... + w_{1,J}x_{J,t}$.[11] All combinations undergo non-linear activation with the commonly-used leaky rectified linear unit function (LReLU) to scale neuronal contributions (Maas et al., 2013).[12] Using a soft-max function, the network outputs the posterior probability ρ of each observation at time t belonging to component k. A soft-max function regularizes the network's output for each observation between 0 and 1 satisfying the necessary conditions $\sum_{k=1}^{K} \rho_{kt} = 1$ and $\rho_{kt} \geq 0 \; \forall \; k$ for posterior

[8] Our choice of $K = 2$ is based on Tolhurst and Ker (2015) who found $K = 2$ using Akaike and Bayesian information criterion for the same Iowa corn data.

[9] The structure of KMNs is more restrictive than MDN but less prone to overfitting (Rothfuss et al., 2019).

[10] We model the mixing probabilities without a trend component as Tolhurst and Ker (2015) who find no evidence of a trend. We test this assumption and similar to Tolhurst and Ker (2015), we can not reject the null of no trend in any of the 99 counties (0.05 level of significance). Multiple testing correction made using Holm-Bonferronni.

[11] Another common notation is $G = \mathbf{w}^T\mathbf{x} + b$, where $\mathbf{x}$ is a vector of all inputs and $\mathbf{b}$ is bias.

[12] In comparison to the ReLU, the leaky ReLU does not allow for nodes in the hidden layer to shut off.

probabilities of normal mixtures.

Each component's intercept, slope, and variance are estimated in the maximization step. We restrict the estimated time trends in the upper and lower components to not cross.[13] Through iteration between the neural network (expectation) and maximization steps, the log-likelihood is maximized. Instead of directly updating the posterior probabilities, as done in the typical EM, the posterior probabilities are determined by the neural network. The weights are derived using the Root Mean Squared Propagation gradient descent method (RMSprop) on the negative of the log-likelihood until convergence.[14]

Machine learning algorithms often require several high-level decisions on their structure, optimization methods, and parameters. Failure to control for over-fitting and properly generalize the model results in poor predictive performance.[15] In the case of Gaussian mixture models, regularization and averaging methods are necessary to prevent mode collapse, account for convergence to local minima, and improve out-of-sample performance (Makansi et al., 2019).[16] To improve out-of-sample performance and rate of convergence, a ridge regularization (L_2 penalty) parameter (λ) is selected for each county.[17] This parameter adjusts the loss function by subtracting the sum of squared neural network weights from the negative log-likelihood function: $\ell(y_t; \pi_k, \alpha_k, \beta_k, \sigma_k) - \lambda \sum w^2$. The optimal regularization parameter results in a trade-off between reduced over-fitting and regularization added bias.[18]

Convergence to local minima is one of the main limitations of this approach. Using neural networks, Crane-Droesch (2018) demonstrated large improvements in out-of-sample error for expected corn yield predictions by "bagging"; averaging over a large number of bootstrapped samples (Breiman, 1996). Due to the large computational cost of our proposed model and the increased frequency of mode collapse in bootstrapped samples, leave-one-out (LOO)

[13] In these cases the intercept from the lower component was set to be equal to the intercept of the upper component. This occurred in 3 of the 99 cases.

[14] Detail concerning the RMSprop method is in Appendix section 1.3.

[15] A model's ability to generalize refers to its accuracy in out-of-sample predictions using data not in the training set.

[16] Virtually all neural networks require high-level decisions for generalization. In particular, likelihoods from mixture models are known to be relatively flat across the dimensions of the parameter space.

[17] The L_2 parameter is selected with grid-search for each county using cross-validation with a rolling forecasting origin. The selected parameter minimizes the average negative log-likelihood on the forecast sample.

[18] Generally, regularization is the class of methods needed to modify maximum likelihood to give reasonable predictions and estimates using new data.

cross-validation methods were used to create the model space.

To average over our models we implement Bayesian Model Averaging (BMA) which improves predictive performance over using any single model (Madigan and Raftery, 1994). In our setting, consider a set of J estimated LOO densities for a given county ($\hat{f}_1, ..., \hat{f}_J$). The county's set of models is assigned weights based on their performance in a hold-out sample. The posterior distribution of our quantity of interest, in this case, future observations Δ, under model M_j is $pr(\Delta|M_j, D)$. We assume a uniform prior over the set of models and as such the posterior probability of model M_j is $pr(M_j|D)$ and the posterior distribution of Δ given data D is the average of each model's posterior distribution weighted by each model's probability.

$$pr(\Delta|D) = \sum_{j=1}^{J} pr(\Delta|M_j, D)pr(M_j|D) \tag{6}$$

The posterior probability of model M_j is given as

$$pr(M_j|D) = \frac{pr(D|M_j)pr(M_j)}{\sum_{j=1}^{J} pr(D|M_j)pr(M_j)} \tag{7}$$

and

$$P(D|M_j) = \int pr(D|\theta_j, M_j)pr(\theta_j|M_j)d\theta_j \tag{8}$$

where $pr(M_j)$ is the prior probability, θ_j is our vector of model parameters $(,\pi_k, \alpha_k, \beta_k, \sigma_k)$, and $pr(\theta|M_j)$ is our prior for θ_j. Using the Laplace method, the integral from equation (8) can be approximated using the Bayesian Information Criterion yielding a posterior probability of $pr(M_j|D) = exp{-\frac{1}{2}BIC_j}$ (Dasgupta and Raftery, 1998).

The resulting conditional density estimate $\hat{f}$ for a county becomes

$$\hat{f} = \sum_{j=1}^{J} \omega_j \hat{f}_j \tag{9}$$

where

$$\omega_j = \frac{exp(-\frac{1}{2}BIC_j)}{\sum_{j=1}^{J} exp(-\frac{1}{2}BIC_j)} \tag{10}$$

4.3 Mean-Variance

The mean-variance model is created to offer a comparison to the NME by limiting the model to have only a single normal component. In this way, the weather-yield relationship is restricted to by symmetric across the distribution. In our construction of the MV, we do not impose linear time trends on the component or a fixed component variance as done in the NME. Rather, we allow a neural network to directly determine the mean μ and standard deviation σ using information on weather and time. This flexibility in the first two moments ensures that it is only more restrictive in higher moments.

The MV model contains two separate neural networks, one for each parameter of the predicted normal distribution. Each network is simultaneously updated to maximize the negative log-likelihood. Identical to the Bayesian model averaging approach used for the *NME*, leave-one-out cross-validation over each year of observations is used to create the model space.

5 Empirical Application

5.1 Density Estimation Results

We estimate model parameters for each of the 99 Iowa counties separately using data from 1950-2006 and test predictive performance on a hold-out sample from 2007-2019.[19] A summary of the parameter estimates for the three methodologies is located in Table 2. The differences between *NME* and *EM* directly highlight the effect of including weather in the estimation process. The differences between the *NME* and *MV* highlight the difference in distributional assumptions.

The *EM* methodology illustrates more variation in the lower component estimates (comparing the minimum to mean parameter estimates). This is expected as the lower component is more affected by weather and has significantly fewer observations to estimate the parameters versus the upper component (Tolhurst and Ker, 2015). Including weather in the estimation process leads to a less strict assignment of component probability and thus we

[19]This split roughly achieves an 80/20 split of training/test data

would expect less extreme parameter estimates. The MV model's average slope of the sample falls between the upper and lower trends in the NME. However, it has a smaller variance and slightly higher intercept than the NME.

Table 2: Summary Statistics for Estimated In-Sample Parameters for *NME*, *EM*, and *MV* Methodologies - Iowa Corn (1950-2006)

Model	Parameter	Min	Median	Mean	Max	Std. Dev.
NME	π_l	0.04	0.24	0.28	0.58	0.13
	α_l	25.32	47.37	45.71	63.59	8.60
	α_u	32.69	50.12	49.67	64.02	6.35
	β_l	0.64	1.52	1.53	2.28	0.37
	β_u	1.68	2.03	2.03	2.38	0.16
	σ_l	5.32	20.52	20.04	30.39	4.72
	σ_u	8.52	12.72	12.82	17.91	2.25
EM	π_l	0.03	0.21	0.27	0.76	0.18
	α_l	2.03	46.35	43.74	59.57	10.31
	α_u	30.65	49.65	48.84	63.89	6.06
	β_l	0.09	1.34	1.34	2.18	0.50
	β_u	1.76	2.15	2.16	2.65	0.20
	σ_l	0.05	17.21	16.73	28.58	5.66
	σ_u	2.73	9.54	9.39	14.06	2.57
MV	α^a	39.06	58.65	57.99	74.24	5.67
	β^a	1.04	1.80	1.74	2.14	0.26
	σ	16.63	17.69	18.19	20.76	2.10

[a] These parameters are estimated from the MV in-sample predictions for the sake of comparison to EM and NME. They are not outputs of the MV model or should be interpreted as estimates of a technological trend.

If we focus on the *NME* parameter estimates, several points are worth noting: (i) lower and upper component intercepts (α_l, α_u) of two-component models are close to each other as expected given these represent yields from the early 1950s where variance was relatively low; (ii) the probability in the lower component (π) is smaller than the upper component in 88 of 99 counties for the *NME* method; and (iii) the technology slope parameter is greater in the upper versus lower tail and thus illustrates a yield distribution with increasing variance over time. Consistent with this last result, Ramsey (2020) found technological change to

be greater at the 75^{th} percentile than the 25^{th} in 44 of 70 Illinois counties. Interestingly, we find diverging trends to be reduced upon weather inclusion. Incorporating climate in the *NME* method reduces *upper* component trends by 0.13 bushels per year while raising *lower* component trends by 0.19 bushels per year bringing slope parameters noticeably closer together for the *NME* versus the *EM* methodology. Again, this is because including weather in the estimation process leads to less strict assignment of component probability and thus we would expect less extreme parameter estimates, exactly what we see.

Table 3 compares the in-sample and hold-out performances between the methodologies. It is expected that the *EM* methodology performs best in-sample given it maximizes the likelihood based only on observed yields whereas the *MV* and *NME* maximize a penalized likelihood and incorporates weather in the estimation process.[20]

Table 3: Mean negative log-likelihood (NLL) estimates from in-sample and validation on the hold-out sample for alternative density estimation models

Model	Measure	NLL
NME	In-sample	-4.148
	Hold-out	-4.323
EM	In-sample	-4.073
	Hold-out	-4.837
MV	In-sample	-4.159
	Hold-out	-4.648

There are three important conclusions regarding model performance in the hold-out sample. First, the results show the inclusion of weather information improves density predictions because the *MV* and *NME* methods make noticeable improvements over the *EM*. This is consistent with previous work estimating insurance premiums (Tack et al., 2012, 2018; Ramsey, 2020; Liu and Ramsey, 2022).[21]

Second, given the inclusion of weather information, more flexible assumptions on the

[20]The reported negative log-likelihood in Table 3 is not penalized.

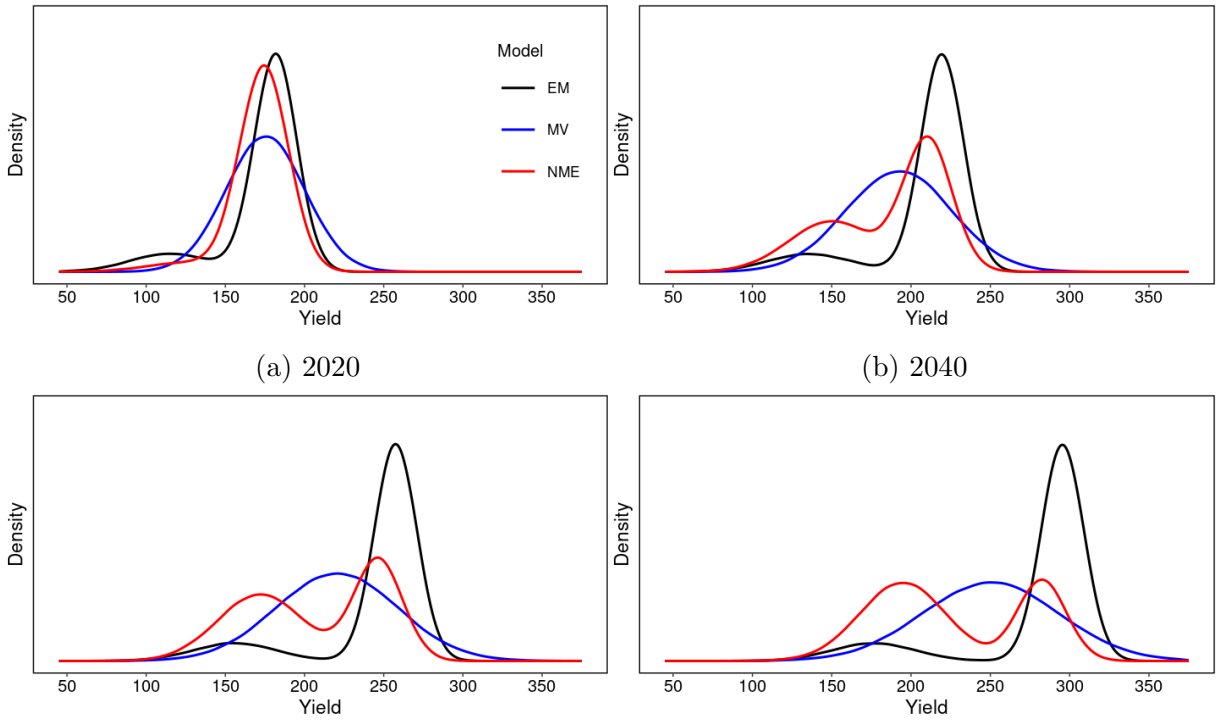
[21]In addition, we evaluate the importance of each weather variable and combinations of weather variables on NME performance (see Figure A.4a in the Supplement)

shape of the yield distribution improve predictive performance. The reduction in performance from the in-sample to hold-out sample is 11.8% in the *MV* approach compared to a 4.2% decline in the *NME* approach. Importantly, the *NME* methodology predicts better than the *MV* methodology in 89 of the 99 counties. Third, the complexity of how the weather-yield relationship is modeled across the distribution appears to be less important than distributional flexibility (at least in our case). We demonstrate this by replacing the neural network in our *NME* with a linear model. Despite hold-out performance declining from -4.323 to -4.451, it remains superior to predictions from the MV model (-4.648) which uses neural networks. The use of the neural network is guided by previous literature such as Crane-Droesch (2018) and Keane and Neal (2020b) who find machine learning methods tend to outperform commonly used econometric methods for crop mean yield prediction problems. More importantly, we had no prior knowledge as to possible functional forms for the correct mapping from climate to mixing probabilities. Of course, for other crop-region combinations, the relative performance of the neural network versus linear model could be vastly different.

5.2 Climate Effects on Yield Distributions

In this subsection, we examine the distributional consequences of climate change by predicting future conditional densities using historical weather patterns adjusted by RCP temperature projections. Current literature examining yield distributions under climate change finds changes to multiple moments of crop yield distributions (Tack et al., 2012, 2015, 2018; Malikov et al., 2020). Looking specifically at corn, the two-component neural mixture allows for covariates to non-linearly and heterogeneously effect distribution components.

Figure 3 illustrates the *NME*'s capacity to predict changes to lower component probability, diverging components, and rising variance in comparison to the *EM* and *MV* models for the representative county. Allowing weather to affect all moments of the predicted yield distribution has significant economic implications for the overall shape and thus premium rates and future government outlays. Not surprisingly but nonetheless interestingly and economically relevant for predicting future insurance costs, the predicted conditional densities become increasingly different over time across the three methods.



(a) 2020 (b) 2040

(c) 2060 (d) 2080

Figure 3: Estimated Conditional Densities of Wapello County using three methodologies for 2020, 2040, 2060, and 2080 under RCP 4.5

Table 4 summarizes the moments of the predicted yield distributions for the three methods in 20-year intervals (see Figures A.5-A.8 in Supplemental Appendix for all scenarios). The greatest changes are projected to occur to the probability of the lower tail. In comparison to a *MV* model, the *NME* predicts yield distributions with notably lower mean, lower variance, lower negative skewness, and greater excess kurtosis. Note, the MV distribution can only represent changes in the kurtosis of the data-generating process (tailedness) by adjusting the variance. The lower mean and negative skewness caused by incorporating weather heterogeneously in different parts of the yield distribution also cannot be accommodated in our *MV* model. In comparison to the *EM* model, the *NME* predicts yield distributions with notably lower mean and variance, less negative skewness, and less kurtosis.[22] These differences are caused by a significant increase in lower component probability caused by climate change that is captured within the *NME* methodology but not the *EM* methodology. The

[22]Under a normal distribution skewness is 0 and kurtosis is 3. Note, under a mixture of normals skewness may be present

comparison of the *NME* and *EM* illustrates the effect of changing climate.

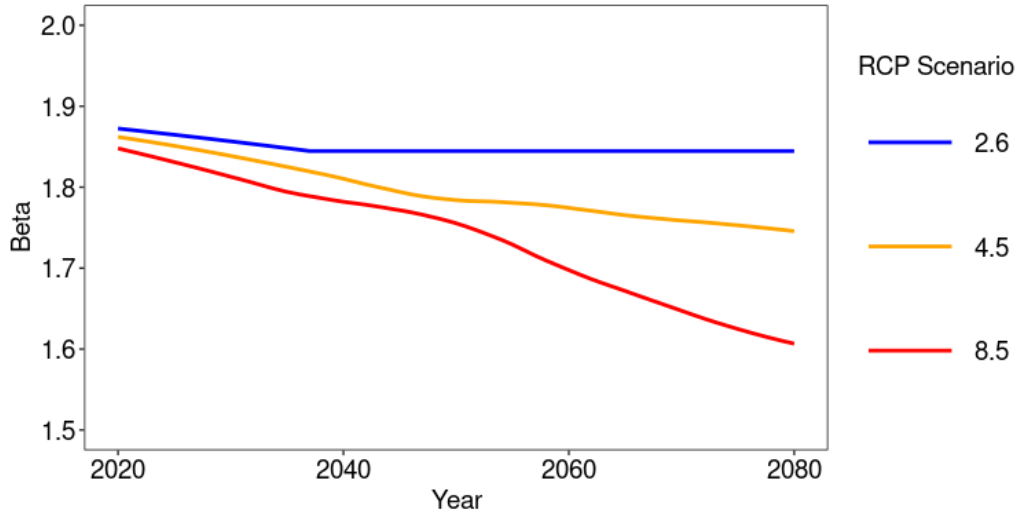
Table 4: Out-of-sample comparisons of production-weighted crop yield moments between models of data-generating processes under RCP 4.5

Model	Moment	2020	2040	2060	2080
NME	Mean	180.64	212.23	244.33	276.89
	St. Dev.	23.43	27.36	30.83	33.78
	Skewness	-0.80	-0.62	-0.45	-0.31
	Kurtosis	3.99	3.76	3.91	4.19
EM	Mean	188.36	224.00	263.62	303.22
	St. Dev.	26.38	31.18	36.68	42.35
	Skewness	-1.37	-1.39	-1.40	-1.40
	Kurtosis	4.90	4.94	4.95	4.98
MV	Mean	183.22	215.83	249.49	284.63
	St. Dev.	30.36	35.52	41.33	46.81

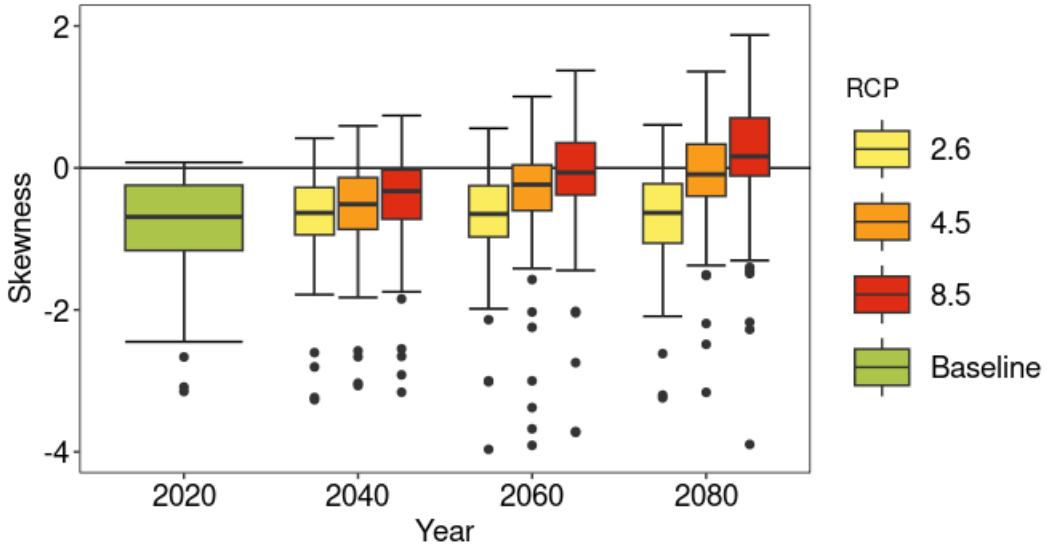
In general, we find the yield distributions predicted by both the *MV* and *NME* models continue to shift upwards from technological improvements. Uniquely in the NME, probability moves significantly between the two components driven solely by changes in the underlying distribution of climate. Between components, the probability typically shifts from the *upper* to the *lower*, increasing probability below the mean. These temperature-induced changes in component probability imply the average rate of technological improvement (changes in mean yields) declines as probability shifts to the lower (slower) component (Figure 4a) with the greatest decline corresponding to the greatest warming scenario. For the same reason, distributions tend to become more positively skewed under warming with greater effects for more severe warming scenarios (Figure 4b).[23,24] While the decreasing magnitude of negative skewness under greater warming may seem counter-intuitive, greater warming causes notably increased probability in the lower component not simply increased probability in the lower tail of the lower component which would necessarily cause greater negative skewness.

[23] "More positive skewness" means negatively skewed distributions become less negatively skewed and positively skewed distributions become more positively skewed.

[24] Positive changes in skewness occur in 84 counties under RCP 2.6 and 86 under RCP 8.5.



(a) The median technological trends of Iowa counties under 3 climate scenarios.



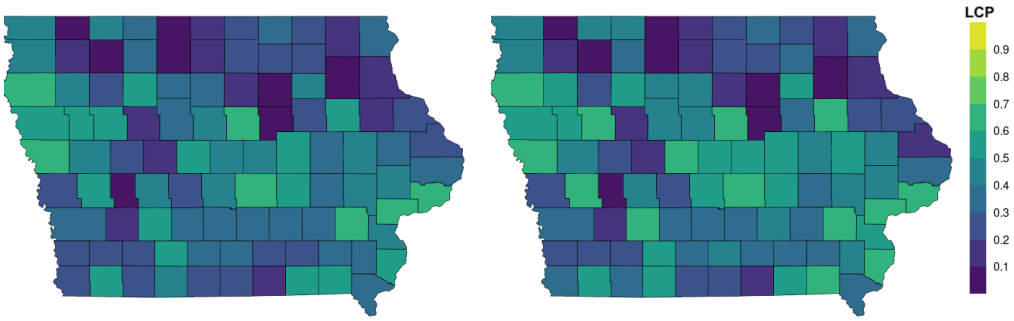
(b) Skewness Changes.

Figure 4: First and Third Moment Effects

The significant geographical heterogeneity in measures of lower probability suggests modeling the yield-weather relationship independently at the county level is informative and interesting. We tested the null of $\pi_i = \pi_j$ for the 4851 pairs of counties (100*99/2-99). We strongly reject the null of $\pi_i = \pi_j \; \forall \; i \neq j$. Using Holm-Bonferroni (Holm, 1979) to correct for multiple testing, we reject $\pi_i = \pi_j$ for 4488 of the 4851 unique pairs. While the vast majority of counties are predicted to experience positive changes, much larger increases are

expected in Southern and Eastern Iowa. These spatial differences are likely due to differences in historical climate conditions and factors such as soil type and the ability to retain moisture under greater evaporative pressure (Ortiz-Bobea et al., 2019; Kane et al., 2021). Additionally, we evaluate the relative influence of each of the variables on lower component probability and find all variables to be of similar importance. Quite interestingly, in 1977, 1980, 1983 and 1988, when Iowa corn yields were poor, the darker counties (with smaller lower component probability) in Figure 5 typically experienced smaller losses. In 1977, the 7 counties with the smallest lower components had average yields 17 bushels above the remaining 92 counties. In 1980, 1983, and 1988, average yields were 14, 19, and 20 bushels higher among the "darker" counties. When we predict on our hold-out sample of 2012, a particularly dry year, our model consistently predicts small lower components for the darker counties. On average, the darkest 7 counties yielded 151 bu/ac in 2012 while the remaining 92 yielded 129 bu/ac.

Our findings regarding changes to the moments of future corn yield distributions are consistent with Malikov et al. (2020) and Tack et al. (2018) who similarly find greater variance and negative impacts of climate change on the mean (relative to no climate change). Our adaptation of the EM algorithm complements this work and highlights how climate change shifts probability from the upper tail to the lower tail implying changes to the probability of yield outcomes and a decreasing rate of technological change over time. The change to the "fatness" of tails and the significant increase in low probability events is particularly important when considering the economic costs of climate change (Nordhaus, 2011; Weitzman, 2011).



(a) RCP 4.5 in 2020

(b) RCP 8.5 in 2020

(c) RCP 4.5 in 2040

(d) RCP 8.5 in 2040

(e) RCP 4.5 in 2060

(f) RCP 8.5 in 2060

(g) RCP 4.5 in 2080

(h) RCP 8.5 in 2080

Figure 5: Predicted county lower component probability (LCP) under RCP 4.5 and 8.5

5.3 Financial Implications of Rising Temperatures

This section examines the impact of future climate scenarios (RCP 2.6, 4.5, and 8.5) on expected loss and crop insurance subsidies. The probability of crop loss is given by the area under the yield density below the level of guaranteed coverage. For a contract guaranteeing τ percent of expected yield (Y^e), where the expected yield is calculated as the mean of the predicted distribution, the expected loss (EL) is given by:

$$EL = P(y < \tau Y^e)(\tau Y^e - E(y|y < \tau Y^e)). \tag{11}$$

The actuarially fair premium rate is the ratio of expected loss (EL) to total liability (τY^e). The probability and expected value of loss are recovered from the predicted yield density.

Table 5 illustrates the mean effect on premium rates at 90% coverage of different climate scenarios and assumptions on future technological progress. As in the introduction, we present two sets of premium rates: the first without changing the guarantee so rates are directly comparable (left side of Table 5); and the second, allowing the guarantee to also fall with climate change representing expected future rates but not comparable to the no climate change scenario (right side of Table 5). Overall, we find three main results.

First, the effect of climate change on expected losses and premium rates is non-linear over time. Consider the left side of Table 5 where premium rates are predicted using the yield guarantee under the *no climate change* scenario. Relative to the no climate change scenario, premium rates rise increasingly across the severity of climate scenarios due to increased probability of loss. Consistent with work by Tack et al. (2018) and Perry et al. (2020), premium rates (estimated at yield guarantees from a no climate change scenario) are projected to rise 20-66% by 2040 with greater losses over time due to the increased probability of loss. By 2080, premium rates are projected to rise 21-256% across all warming and technological change scenarios. However, when the yield guarantee is also allowed to fall, the rise in rates driven by shifting in the component probability is offset by declines in the yield guarantee relative to no climate change (right side of Table 5). Re-iterating, these premium rates represent what the farmer would pay (and rates for which pubic subsidies would be based) but are not comparable to the no climate change premium rates as the yield guarantees are much different. In this case, premium rates rise by only 6-14% by 2040. By

comparing rates across warming scenarios it is clear the magnitude of temperature increases under climate change determines the degree to which the yield guarantee dominates the effect on the lower component.

	Premium Rates (%) No Climate Yield Guarantee				Premium Rates (%) Changing Yield Guarantee			
Tech. Trend	2020	2040	2060	2080	2020	2040	2060	2080
				No Change in Climate				
None	2.55	2.55	2.55	2.55	2.55	2.55	2.55	2.55
50%	2.54	2.48	2.43	2.40	2.54	2.48	2.43	2.40
Historical	2.53	2.43	2.38	2.36	2.53	2.43	2.38	2.36
				RCP 2.6				
None	2.58	3.05	3.05	3.05	2.56	2.73	2.73	2.73
50%	2.57	2.98	2.93	2.90	2.56	2.64	2.58	2.53
Historical	2.56	2.93	2.88	2.86	2.54	2.58	2.50	2.47
				RCP 4.5				
None	2.79	3.60	4.42	4.95	2.65	2.86	2.90	2.85
50%	2.79	3.52	4.29	4.79	2.65	2.74	2.61	2.39
Historical	2.77	3.47	4.25	4.78	2.63	2.65	2.44	2.15
				RCP 8.5				
None	3.09	4.17	5.45	6.55	2.75	2.91	2.79	2.60
50%	3.08	4.09	5.32	6.41	2.75	2.77	2.39	1.91
Historical	3.07	4.04	5.29	6.43	2.73	2.65	2.12	1.54

Table 5: Premium rates (90% coverage) for NME approach under alternative assumptions on technological trends and climate scenarios with results allowing the yield guarantee to change and holding it fixed at 2020 value

Second, premium rates under climate change do not depend solely on temperature changes, but on the rate of temperature change over time *relative* to the rate of technological change over time. While the effect of continued yield trends lowers premium rates, climate change can drive rates in either direction based on the magnitude of temperature changes. Consider the left side of Table 5 where the yield guarantee only changes over time from technology trends. In this case, premium rates are driven almost entirely by temperature rises over time. When the yield guarantee changes depending on the climate scenario, differences across technological trends become pronounced (right side of Table 5). This is because the estimated technological trends are divergent, with a greater distance between

components at higher assumed trends. Therefore, any probability shifting from the upper to the lower component leads to a greater change in the yield guarantee relative to slower assumed technological trends under the same climate scenario.

Third, across all climate scenarios and realistic assumptions of technological trends (50%-100%), total costs (premium rate x expected yield) to insure corn acreage at 90% coverage will rise 8-14% by 2040 when considering the total effect of climate change on the probability of loss and yield guarantee. Consistent with previous literature we maintain that future financial costs of the federally-subsidized crop insurance program due to climate change will continue to rise, albeit more modestly. However, the inclusion of technological trends and heterogeneity of climate effects within the distribution points to different underlying mechanisms. Our findings suggest the effect of climate on the probability of yield loss will initially dominate the downward pressure on yield guarantees in the short term (2020-2040). Under continued rises in temperature beyond 2040, we predict the reverse to be true.[25]

Historically, climate change contributed to a large rise in insurance costs (Diffenbaugh et al., 2021). By allowing for weather to influence the shape of the yield distribution and component-specific rates of technological change, we find their relative magnitudes and rates of change point to smaller and eventually negative effects of climate change on insurance costs. Our results do not differ significantly from Perry et al. (2020) who predicts small rises (<5%) in loss-cost ratios in Iowa under a 1°C rise in temperature (see RCP 2.6 in 2040). Studying average county corn yields, Butler and Huybers (2013) finds decreasing sensitivity to climate change in warmer regions – supporting the decrease in climate effects over time.

Our results are conditional on historic trends in technological change and adaptation to rising temperatures. Past yield trends relied on improvements in seed genetics (Lusk et al., 2018), increases in fertilizer application, and higher planting densities (Perry et al., 2021). Continued technological advancement is crucial to offsetting the effects of climate change on mean crop yields (Ortiz-Bobea and Tack, 2018). In the future, new technologies (i.e. genetic improvements) or intensification of climate change may shift future production outside historical experiences (Dell et al., 2014).

[25]Note, our estimated premium rates across scenarios in 2020 (2.53 - 3.09%) are considerably lower than the Risk Management Agency's Area Yield Protection premium rates (4.71%). Lower predicted premium rates are consistent with Liu and Ker (2020) who found that several methodologies, all with smaller variance, outperform the RMA methodology.

6 Conclusion

The effect of changing climate conditions on crop yield distributions and publicly-subsidized crop insurance programs warrants careful attention. Building on the increasing use of weather information in crop yield density estimation (see Liu and Ramsey (2022); Ramsey (2020); Tack et al. (2012, 2015, 2018)), we introduce a new approach that utilizes neural networks directly in an EM algorithm within a mixture model thereby allowing for both asymmetry and non-linearity in the weather-yield relationship across the crop yield distribution. While this proposed methodology improves predictive performance relative to other models considered here, importantly, this approach allows climate change to affect the crop yield distribution heterogeneously and non-linearly across its support (if dictated by the historical yield-weather data). Moreover, this approach allows climate change to move probability from one location of the yield distribution to another location (i.e. upper to lower tail) altering the assumed rate of technological change.

Using uniform shifts in daily temperature to simulate three different RCP scenarios (2.6, 4.5, 8.5), we predict county crop yield distributions in Iowa for 2020-2080. Similar to Tack et al. (2012) and Malikov et al. (2020), we find rising temperatures impact multiple higher moments of the yield distribution. Our approach identifies the effects of climate change on the distributional shape of yields whereas most previous approaches focus only on the conditional mean. Consequently, we find rising temperature to increase probability in the historically smaller, lower component, and under larger temperature increases, more positive skewness, and multi-modality. While yield losses would rise significantly, in most cases doubling, if we allow the yield guarantee to drop with climate change as well, the resulting insurance costs from climate are much less dramatic.

Finally, methodology matters for Iowa corn and likely matters for most crop-region combinations. Our results are an order of magnitude different assuming a normal distribution and mapping weather into the parameters using a neural net. An approach that employs a flexible distribution and allows heterogeneous mapping of weather into different parts of the yield distribution appears to be preferred. The NME model accomplishes this through the use of mixtures and the mapping of climate to distributional parameters. Tack et al. (2018) does this by specifying the distribution nonparametrically and allowing the parameters of

the individual distributions to be functions of climate. Such sensitivity of results to distributional flexibility is not generally found but in the case of tail probabilities and premium rates combined with forecasting ahead decades, it is not surprising. It does not appear that the mapping of climate into the distributional parameters via nonparametric or nonlinear methods such as a neural net -- as opposed to linearly -- adds much.

Finally, methodology matters for Iowa corn and likely matters for most crop-region combinations. Our results are an order of magnitude different assuming a normal distribution and mapping weather into the parameters using a neural net. An approach that employs a flexible distribution and allows heterogeneous mapping of weather into different parts of the yield distribution appears to be preferred. The NME model accomplishes this through the use of mixtures and the mapping of climate to distributional parameters. Tack et al. (2018) does this by estimating the distribution nonparametrically and allowing the parameters of the individual distributions to be functions of climate. Such sensitivity of results to distributional flexibility is not generally found but in the case of tail probabilities and premium rates combined with forecasting ahead decades, it is not surprising. Finally, it does not appear that the nonparametric mapping of climate into the distributional parameters via a neural network adds much relative to a simple linear mapping.

Bibliography

Ambrogioni, L., Güçlü, U., van Gerven, M. A., and Maris, E. (2017). The kernel mixture network: A nonparametric method for conditional density estimation of continuous random variables. *arXiv preprint arXiv:1705.07111*.

Annan, F. and Schlenker, W. (2015). Federal crop insurance and the disincentive to adapt to extreme heat. *American Economic Review*, 105(5):262–66.

Auffhammer, M., Hsiang, S. M., Schlenker, W., and Sobel, A. (2013). Using weather data and climate model output in economic analyses of climate change. *Review of Environmental Economics and Policy*.

Bielza, M., Stroblmair, J., Gallego, J., Conte, C. G., and Dittmann, C. (2007). Agricultural risk management in europe. Technical report.

Bishop, C. M. (1994). Mixture density networks.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Burke, M. and Emerick, K. (2016). Adaptation to climate change: Evidence from us agriculture. *American Economic Journal: Economic Policy*, 8(3):106–40.

Butler, E. E. and Huybers, P. (2013). Adaptation of us maize to temperature variations. *Nature Climate Change*, 3(1):68–72.

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S., Smith, D., and Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4(4):287–291.

Chemeris, A., Liu, Y., and Ker, A. P. (2022). Insurance subsidies, climate change, and innovation: Implications for crop yield resiliency. *Food Policy*, 108:102232.

Connor, L. and Katchova, A. L. (2020). Crop insurance participation rates and asymmetric effects on us corn and soybean yield risk. *Journal of Agricultural and Resource Economics*, 45(1):1–19.

Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003.

Cui, H., Radosavljevic, V., Chou, F., Lin, T., Nguyen, T., Huang, T., Schneider, J., and Djuric, N. (2018). Multimodal trajectory predictions for autonomous driving using deep convolutional networks. *CoRR*, abs/1809.10732.

D'Agostino, A. L. and Schlenker, W. (2016). Recent weather fluctuations and agricultural yields: implications for climate change. *Agricultural Economics*, 47(S1):159–171.

Dasgupta, A. and Raftery, A. E. (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American statistical Association*, 93(441):294–302.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? the new climate-economy literature. *Journal of Economic Literature*, 52(3):740–98.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22.

Deschênes, O. and Greenstone, M. (2007). The economic impacts of climate change: evidence from agricultural output and random fluctuations in weather. *American economic review*, 97(1):354–385.

Diffenbaugh, N. S., Davenport, F. V., and Burke, M. (2021). Historical warming has increased us crop insurance losses. *Environmental Research Letters*, 16(8):084025.

Gammans, M., Mérel, P., and Ortiz-Bobea, A. (2017). Negative impacts of climate change on cereal yields: statistical evidence from france. *Environmental Research Letters*, 12(5):054007.

Goodwin, B. K., Roberts, M. C., and Coble, K. H. (2000). Measurement of price risk in revenue insurance: implications of distributional assumptions. *Journal of Agricultural and Resource Economics*, pages 195–214.

Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2):174–194.

Hennessy, D. A. (2009). Crop yield skewness under law of the minimum technology. *American Journal of Agricultural Economics*, 91(1):197–208.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, pages 65–70.

Hsiang, S. (2016). Climate econometrics. *Annual Review of Resource Economics*, 8:43–75.

Kane, D. A., Bradford, M. A., Fuller, E., Oldfield, E. E., and Wood, S. A. (2021). Soil organic matter protects us maize yields and lowers crop insurance payouts under drought. *Environmental Research Letters*, 16(4):044018.

Keane, M. and Neal, T. (2020a). Climate change and us agriculture: Accounting for multidimensional slope heterogeneity in panel data. *Quantitative Economics*, 11(4):1391–1429.

Keane, M. and Neal, T. (2020b). Comparing deep neural network and econometric approaches to predicting the impact of climate change on agricultural yield. *The Econometrics Journal*, 23(3):S59–S80.

Ker, A. P. (1996). *Rating and yield predicting procedures for the group risk federal crop insurance program: A Nonparametric Approach*. North Carolina State University.

Ker, A. P., Barnett, B., Jacques, D., and Tolhurst, T. (2017). Canadian business risk management: Private firms, crown corporations, and public institutions. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie*, 65(4):591–612.

Khaki, S., Wang, L., and Archontoulis, S. V. (2020). A cnn-rnn framework for crop yield prediction. *Frontiers in Plant Science*, 10:1750.

Liu, Y. and Ker, A. P. (2020). Rating crop insurance contracts with nonparametric bayesian model averaging. *Journal of Agricultural and Resource Economics*, 45(2):244–264.

Liu, Y. and Ramsey, A. F. (2022). Incorporating historical weather information in crop insurance rating. *American Journal of Agricultural Economics*.

Lobell, D. B. and Burke, M. B. (2010). On the use of statistical models to predict crop yield responses to climate change. *Agricultural and forest meteorology*, 150(11):1443–1452.

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.

Lusk, J. L., Tack, J., Hendricks, N. P., et al. (2018). Heterogeneous yield impacts from adoption of genetically engineered corn and the importance of controlling for weather. *Agricultural Productivity and Producer Behavior*.

Maas, A. L., Hannun, A. Y., Ng, A. Y., et al. (2013). Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer.

Madigan, D. and Raftery, A. E. (1994). Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of the American Statistical Association*, 89(428):1535–1546.

Mahul, O. and Stutley, C. J. (2010). *Government support to agricultural insurance: challenges and options for developing countries*. World Bank Publications.

Makansi, O., Ilg, E., Cicek, O., and Brox, T. (2019). Overcoming limitations of mixture density networks: A sampling and fitting framework for multimodal future prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7144–7153.

Malikov, E., Miao, R., and Zhang, J. (2020). Distributional and temporal heterogeneity in the climate change effects on us agriculture. *Journal of Environmental Economics and Management*, 104:102386.

McIntosh, C. T. and Schlenker, W. (2006). Identifying non-linearities in fixed effects models. *UC-San Diego Working Paper*.

McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*. John Wiley & Sons.

Ng, H. and Ker, A. (2021). On the changing nature of canadian crop yield distributions. *Journal of Agricultural and Resource Economics*, 46(1):101–125.

Nordhaus, W. D. (2011). The economics of tail events with an application to climate change. *Review of Environmental Economics and Policy*.

Ort, D. R. and Long, S. P. (2014). Limits on yields in the corn belt. *Science*, 344(6183):484–485.

Ortiz-Bobea, A. (2021). The empirical analysis of climate change impacts and adaptation in agriculture. In *Handbook of agricultural economics*, volume 5, pages 3981–4073. Elsevier.

Ortiz-Bobea, A. and Tack, J. (2018). Is another genetic revolution needed to offset climate change impacts for us maize yields? *Environmental Research Letters*, 13(12):124009.

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.

Perry, E. D., Hennessy, D. A., and Moschini, G. (2021). Uncertainty and learning in a technologically dynamic industry: Seed density in us maize. *American Journal of Agricultural Economics*.

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(1):1–9.

Phillips, D. L., Lee, J. J., and Dodson, R. F. (1996). Sensitivity of the us corn belt to climate change and elevated co2: I. corn and soybean yields. *Agricultural systems*, 52(4):481–502.

Ramsey, A. F. (2020). Probability distributions of crop yields: a bayesian spatial quantile regression approach. *American Journal of Agricultural Economics*, 102(1):220–239.

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

Roberts, M. J., Schlenker, W., and Eyer, J. (2013). Agronomic weather measures in econometric models of crop yield with implications for climate change. *American Journal of Agricultural Economics*, 95(2):236–243.

Rosch, S. (2021). Federal crop insurance: A primer. Technical Report R46686, Congressional Research Service.

Rothfuss, J., Ferreira, F., Walther, S., and Ulrich, M. (2019). Conditional density estimation with neural networks: Best practices and benchmarks. *arXiv preprint arXiv:1903.00954*.

Schlenker, W. and Roberts, M. J. (2006). Nonlinear effects of weather on corn yields. *Review of agricultural economics*, 28(3):391–398.

Schlenker, W. and Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to us crop yields under climate change. *Proceedings of the National Academy of sciences*, 106(37):15594–15598.

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 us crop insurance program. *Agricultural economics*, 49(5):635–647.

Tack, J., Harri, A., and Coble, K. (2012). More than mean effects: Modeling the effect of climate on the higher order moments of crop yields. *American Journal of Agricultural Economics*, 94(5):1037–1054.

Ting, M., Seager, R., Li, C., Liu, H., and Henderson, N. (2021). Future summer drying in the midwestern united state in cmip5 models and the role of midlatitude storm tracks. *Journal of Climate*, 34:9043–9056.

Tolhurst, T. N. and Ker, A. P. (2015). On technological change in crop yields. *American Journal of Agricultural Economics*, 97(1):137–158.

Van Klompenburg, T., Kassahun, A., and Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177:105709.

Vose, R.S., E. D. K. K. L. A. and Wehner, M. (2017). *Climate Science Special Report: Fourth National Climate Assessment, Volume I*, chapter Temperature Changes in the United States, pages 185–206. U.S. Global Change Research Program, Washington, DC, USA,.

Weitzman, M. L. (2011). Fat-tailed uncertainty in the economics of catastrophic climate change. *Review of Environmental Economics and Policy*.

Woodard, J. D. and Sherrick, B. J. (2011). Estimation of mixture models using cross-validation optimization: implications for crop yield distribution modeling. *American Journal of Agricultural Economics*, 93(4):968–982.

Wuebbles, D. J., Fahey, D. W., and Hibbard, K. A. (2017). Climate science special report: fourth national climate assessment, volume i.