A Copula-based Approach to Simulate Climate Impacts on Yield: Some Preliminary Findings

D Sheng[1,2], DM Lambert[1]*, C Hellwinckel[1]

1. University of Tennessee, Department of Agricultural & Resource Economics, Knoxville, Tennessee

2. Graduate Research Assistant

*Corresponding author: dlamber1@utk.edu

Selected paper presented at the Southern Agricultural Economics Association Meeting,

San Antonio, Texas, February 6-9, 2016

Abstract

One approach to modeling the *ex-ante* effects of extreme weather events on production entails statistical downscaling of global climate model ensembles to regional or sub-regional levels. However, statistical downscaling of global climate ensembles is highly technical and computationally expensive in terms of time, data storage, and human resource costs. Our approach considered the use of copulas to generate multivariate empirical distributions of crop yields and meteorological information. The fundamental idea is that retrospective data on drought, floods, or other extreme weather events are statistically associated with crop performance. We posit the historical record of crop production and meteorological events embody all of the necessary information needed to understand what happens to agricultural production when extreme weather events prevail. These trends are evident in publicly available records. By mining the correlation structure between yields and growing conditions embedded in the historical record, we can simulate sequentially poor weather years and the corresponding effects on agricultural production.

Key words: simulation, copula, crops, preliminary findings

JEL Classification: Q15

Introduction

The U.S. agricultural sector is expected to experience declining yield growth trends for major crops under various climate projections into the 21st century (USDA-ERS, 2015). Many factors are expected to be closely related to various in crop production, especially the prospect of more frequent extreme weather events and prolonged drought. An increase in the severity, duration and area affected by drought has been observed to occur over larger geographic areas since the 1970s (Li et al, 2009). From 1980 to 2014, billion-dollar disaster droughts have occurred 22 times, costing on average 9.4 billion USD per event (NOAA, 2015).

Our research questions are motivated by the trends and associations depicted in Figures 1, 2, and 3. The National Ocean and Atmospheric Administration (NOAA) complied monthly historical temperature and precipitation records and interpolated to the county level (NOAA, 2015). NOAA used these data to calculate various drought indictors, including the Palmer Modified Drought Index (PMDI). The PMDI is a standardized variable NOAA uses to classify drought conditions into categories such as *moderate*, *extreme*, or *severe*. According to Figure 1, from1970 to 2011, there are three or four years when extreme drought impacted more than 30% the counties in lower 48 states (Figure 1). Outstanding years include 1977, 1988, and 2011. The drought of 1988 appears to be flanked by periods where 15 to 20% of the counties were impacted by moderate or extreme drought. From 2000 to 2003 there are consecutive periods where counties experienced severe drought. A similar sequence appears again, 2006 – 2008. However, the spatial distribution of impacted counties suggests somewhat of a different story. While the relative frequency of drought appears to be clustered (as seen in the bar chart of Figure 1), the spatial-temporal appears less structured, although periods 2001 – 2003 exhibit some degree of regional persistence. Figure 2 presents the same data from a different, with corn yield deviates (left *y*-axis) and PMDI (right *y*-axis) against time (*x*-axis). In some years and in certain regions, variations in corn yield and the PMDI are coupled. In other years, the co-movement is less evident.

[Figure 1]

[Figure 2]

How did crop production respond to these events? Figure 2 compares corn and soybean production during the "drought years" of 1977, 1988, 2000, and 2007. The data are ASD-level, de-trended yield deviates reported by NASS. The lower (higher) the yield deviation, the worse (better) the yield was from a historical average. Holding many other factors constant, we would expect to see some correspondence between the PMDI and yield deviates. Regions with low PMDIs are impacted by drought. In the absence of irrigation or drought resistant corn or soybean varieties, low PMDI's (extreme or severe drought) should correspond with negative yield deviations (green to dark blue shading) (Figure 2). Close inspection of Figure 2 suggests there may be some corresponded, but it is not immediately evident.

[Figure 3]

We are left with the following question: can the impact of drought on crop production be simulated using the hypothesized correlation between drought and crop production which one would naturally expect to find in the historical record? Our preliminary research attempts to answer this question by developing a framework for simulating how extreme weather events –

such as prolonged drought – could impact crop yields if drought events become more frequent and persist for longer periods. Changes in production should also impact price, and eventually crop planting decisions. In theory, higher national (state) prices (driven by production shortfalls in a particular region), should encourage producers in regions unencumbered by drought (but possibly with access to groundwater irrigation) to intensify production of the crop during shortfalls. Our approach taps into the historical climate and crop yield data to generate bivariate copula for replicating the empirical distributions of crop yields. The basic idea is that, by resampling from the lower tails of yield distributions, we can simulate the impacts of drought on production under a variety of temporal and geographic scenarios.

Literature Review

There are several approaches to simulate the impacts of drought or other adverse weather conditions on yield. The first line of research uses stochastic weather generators to simulate the effects of extreme weather events on crop yields (Torriani et al., 2007; Xiong et al., 2009; Kapphan, Carlanca, 2012; Wang, Wang and Liu, 2011). An example of a plant growth model using weather generators is the Environmental Policy Integrated Model (EPIC) system. When sub-regional information on plant requirements and management practices is available, EPIC can be calibrated and then downscaled to different spatial levels to reflect regional variation in production (USDA-ERS, 2015).

The second general approach combines regression analysis and theoretical distributions to generate yield forecasts for the mean of crop yield in response to different climate or soil projections. (McCarl, Villavicencio, and Wu, 2008; Woodard and Garcia, 2008; Goodwin, 2001). Nelson and Preckel (1989) used the conditional beta distribution to fit the crop yield for its range and skewness. Zanini et al (2001) found that Weibull distribution best summarized yield distribution of corn and soybeans of Illinois Farms. Tack (2013) analyzed county-level corn yield data of U.S. Corn Belt and found evidence in favor of generalized maximum-entropy distribution to fit crop yields. Sherrick et al. (2004) compared the performance of alternative parametric distributions for modeling yields on Illinois farms. Ramirez, Misra and Field (2003) suggested a family of parametric distributions that allowed for non-normal yields. Other researches apply non-parametric approach to model yield distributions. For example, Goodwin and Ker (1998) and Ker and Goodwin (2000) suggested a nonparametric kernel density method to model crop yields. Ker and Coble (2003) introduced a semi-parametric estimator that combines the desirable properties of both parametric and non-parametric estimators. Compared to previous studies which focus on the mean of yield distribution, Harri et al (2011) and Tack, Harri and Coble (2012) focused on higher moments of crop yield.

The above approaches clearly suggest the importance of modeling yields and climate drivers as joint distributions. Copulas provide another approach to model joint distributions. Sklar (1959) introduced the concept of copula. Nelsen (2006) summarized the details of construction and properties of copulas. In the applied agricultural economics literature, copula-based approaches have seen more frequent use in the investigation of systematic production and price, crop insurance, and meteorological-induced disasters (Xu et al., 2010; Okhrin, Odening, and Xu, 2012; Feng and Hayes, 2014; Annan, Tack, Harri and Coble, 2014; Goodwin and Hungerford, 2014).

Methods

We use copulas to generate multivariate empirical distributions of drought severity (with NOAA's PMDI) and crop yields. This analysis examines yield deviates and the PMDI index at the Agriculture Statistic District (ASD) level for mainly corn and soybeans. The drought index we use is the Palmer Z-index. Future research will use county-level data and the PMDI to generate copula and estimate copula other major crops and their association with the drought index.

The copulas are calibrated using historical crop production and weather data. We investigate the possibility of simulating the impacts of extreme weather events on production by focusing on the lower percentiles of the simulated yield distributions to replicate poor growing conditions. The key idea is that retrospective data on drought, floods, or other extreme weather events are statistically associated with crop performance (Mishra and Cherkauer, 2010). In other words, the historical record of crop production and meteorological events – defined by key variables like precipitation and temperature – embody all of the necessary statistical information needed to simulate what happens to agricultural production when extreme weather events prevail. Put simply, when plants do not have adequate water, production decreases. These trends are evident in publicly available records for rain-fed crops. By mining the correlation structure between yields and growing conditions embedded in the historical record, we hope to simulate at any given time step, sequentially poor weather years and the corresponding effects on agricultural production.

Copula models are reliable and flexible method to re-construct and amplify the multivariate correlation structure. Let H be an n-dimensional distribution function with the marginals $F_1, F_2, \ldots, F_n$. Then there exists an n-copula C,

(1) $$H(x_1, x_2, \dots, x_n) = C\big(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\big), \ \ x_1, x_2, \dots, x_n \in R.$$

Let $F_1^{(-1)}, F_2^{(-1)}, \dots, \ F_n^{(-1)}$ be quasi-inverses of $F_1, F_2, \dots, F_n$, respectively. That is,

(2) $$C(u_1, u_2, \dots, u_n) = H(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \dots, \ F_n^{(-1)}(u_n)), \ u_1, u_2, \dots, u_n \in [0,1] .$$

Next, define *h* to be the density function of distribution *H*. Then, the corresponding density function of equation (1) can be written as:

(3) $$h(x_1, x_2, \dots, x_n) = c\big(F_1(x_1), F_2(x_2), \dots, F_n(x_n)\big) \prod_{i=1}^{n} f_i(x_i).$$

Finally, let *c* denote the density function of the copula *C*. Then the corresponding density function of equation (2) can be summarized as:

(4) $$c(u_1, u_2, \dots, u_n) = \frac{h\Big(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2), \dots, F_n^{(-1)}(u_n)\Big)}{\prod_{i=1}^{n} f_i(x_i)}.$$

Feng and Hayes (2014) compare the technical differences between copulas. The copula family generally consists of parametric and nonparametric copulas. Typical parametric copulas are Gaussian, *t*, Clayton, and Gumble copulas. The Gaussian copula assumes linear correlation and zero dependence in the tails of the distribution. However, linear correlations do not contain all relevant information on the dependence and joint distributions with the same correlation coefficient, resulting in unexpected behavior, particularly in the tails (SAS 13.2 User’s Guide, 2014). The Gaussian copula does permit tail dependence (co-movement of tails between the distributions of random variables), whereas the parameterized *t*, Clayton, and Gumble copulas do. A *t*-copula allows for non-zero tail dependence, but imposes symmetry in the dependence distribution in both tails of the distribution (Ahmed and Goodwin, 2015). The Gumble copula implies a stronger upper tail dependence, while the reverse is true for the Clayton copula. Given the variety of copula, model selection is an issue. In our preliminary analysis we use a *t*-copula, which allows for strong dependence in alternative tails.

Let $\Theta=\{(v,\Sigma): v\in(1,\infty), \Sigma\in R^{m\times m}\}$ and let t_v be a univariate *t* distribution with *v* degrees of freedom. The Student's *t* copula can be written as

$$C_\Theta(u_1, u_2, \dots, u_m) = t_{v,\Sigma}(t_v^{-1}(u_1), t_v^{-1}(u_2), \dots, t_v^{-1}(u_m)), \tag{5}$$

where $t_{v,\Sigma}$ is the multivariate Student's *t* distribution with a correlation matrix Σ with v degrees of freedom.

We calibrate the *t*-copula using Kendall's τ rank correlation coefficient, a nonparametric measure of association, against the empirical distributions of the de-trended, ASD-level crop yield deviates and the PMDI. The degrees of freedom parameter, *v*, is estimated with maximum likelihood. Once the yield and PMDI data were fit with the copula, yield and PMDI data points were simulated for n = 1,000 replicates, the observed and simulated empirical distributions compared.

Preliminary Results

Kendall's τ (a measure of association between crop yield deviates and the drought index) suggests there are regional similarities with respect to the strength of association between the drought index and variation in production. The panels in Figure 4 summarize the association (Kendall's τ) between the weather index and nine major crops. The strength of association varies regionally for each crop. In general, ASD highlighted in blue are regions where the expected relationship between rain-fed crops and the drought index holds. Production regions highlighted in brown or yellow indicate regions where the association between variation in yield and the drought index are weak or negative. In these cases, crops were likely irrigated (for example, corn and hay in the western US).

[Figure 4]

The top-left panel of Figure 5 plots the relationship between the historical corn yield deviates and the Z index (n = 40 observations of corn yield and the Palmer Z index). The observed data is transformed to the unit square using a kernel estimator. A *t*-copula is then used to fit the multivariate distribution using Kendall's τ (a non-parametric measure of association) as the correlation measure. One-thousand (1,000) pseudo-observations are generated with the fitted copula (bottom-left panel, Figure 5). The empirical distribution of the pseudo-observation population (red line) nearly approximates the observed distribution of the historical relationship (blue line) between drought and crop production (in this example, corn).

[Figure 5]

The region-specific simulated empirical distributions could provide flexibility in terms of simulating extreme weather conditions. For example, to force a drought scenario over consecutive years, yields corresponding with the lower tails of the empirical distribution (e.g., “Poor” in Figure 5) could be pulled during a Monte Carlo simulation. Alternatively, “Excellent” weather conditions (or even use of irrigation) could be simulated by repeatedly sampling the upper tail of the empirical distribution.

Conclusions

This study used a *t*-copula to fit historical drought and crop yield data with the eventual goal of simulating the impacts of drought on crop production, commodity prices, and land use decisions. The method shows potential with respect to alternative methods relying on downscaling algorithms that push global climate predictions to disaggregated spatial units. The proof-of-concept method appears to replicate well the empirical distributions of drought and crop yield variation when measures of association between these variables are significant. In cases where the historical record indicates negative associations between drought and crop production, the results are less clear. One obvious reason is that producers will irrigate when drought conditions prevail. Another reason could be that the impact of drought on productivity will vary, depending on what stage of growth the plant is in when drought occurs.

The preliminary findings suggest the following research avenues. Future steps suggest the following extensions:

Future extensions of this work are stated as follows:

1. Disaggregate the ASD analysis to counties.
2. Focus on production instead of yield deviates.
3. Hold other factors constant, such as irrigation, topography, soil productivity, and other unobservable effects.

4. Investigate the use of other copula allowing for different tail dependencies.
5. The NOAA data is recorded at monthly intervals. Determine which month is most strongly (positively) correlated with production indices at a spatially disaggregated level.

References

Ahmed, M. and Goodwin, B. (2015). “Copula-Based Modeling of Dependence Structure among International Food Grain Markets.” Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting. San Francisco, CA, July 26-28.2015

Annan, F., Tack, J., Harri, Ardian. And Coble, K. (2014). “Spatial Pattern of yield Distributions: Implications for Crop Insurance.” *American Journal of Agricultural Economics* 96(1): 253–268.

Feng, X. and D.J. Hayes. (2014). Is Government Involvement Really Necessary: Implications for Systemic Risk and Crop Reinsurance Contracts. Economics Presentations, Posters and Proceedings. Paper 4. http://lib.dr.iastate.edu/econ_las_conf/4.

Goodwin, B.K. and Hungerford, A. (2014). “Copula-Based Models of Systemic Risk in U.S. Agriculture: Implications for Crop Insurance and Reinsurance Contracts.” *American Journal of Agricultural Economics* 97 (3): 879–896.

Goodwin, B.K.and Ker, A.P. (1998)."Nonparametric Estimation of Crop Yield Distributions: Im plications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80 (1): 139–53.

Goodwin, B. K. (2001). “Problems with Market Insurance in Agriculture.” *American Journal of Agricultural Economics* 83:643–649.

Harri, A., Coble, K., Ker, A.P. and Goodwin. B.J. (2011). “Relaxing heteroscedasticity assumptions in area-yield crop insurance rating.” *American Journal of Agricultural Economics* 93 (3):707 – 717.

Kapphan, I., Carlanca, P. and Holzkaemper, A. (2012). “Climate Change, Weather Insurance Design and Hedging Effectiveness.” *The Geneva Papers* 37:286–317.

Ker, A.P., and Coble, K. (2003). "Modeling Conditional Yield Densities." *American Journal of Agricultural Economics*, 85(2): 291–304.

Ker, A.P. and Goodwin, B.K. (2000)."Nonparametric Estimation of Crop Insurance Rates Revisi ted." *American Journal of Agricultural Economics*, 82 (2):463–78.

Li, Y.P., Ye, W., Wang, M. and Yan, X.D. 2009. “Climate Change and Drought: A Risk Assessment of Crop-Yield Impacts”. *Climate Research* 39:31–46.

Marshall, E., Aillery, M., Malcolm, S. and Williams R. “Climate Change, Water Scarcity, and Adaptation in the U.S. Fieldcrop Sector.” Economic Research Report 201. Economic Research Service. Accessed November, 2015. www.ers.usda.gov/publications/err-economic-research-report/err201.

McCarl, B.A., Villavicencio, X. and Wu. X. (2008). "Climate Change and Future Analysis: Is Stationarity Dying?" *American Journal of Agricultural Economics* 90(5):1241–1247.

Mishra, V., and K.A. Cherkauer. (2010). Retrospective Droughts in the Crop Growing Season: Implications to Corn and Soybean Yield in the Midwestern United States. *Agricultural and Forest Meteorology* 150:1030–1045.

Nelsen, R.B. *An Introduction to Copulas*. New York: Springer, 2006.

Nelson, C.H. and Preckel, P.V. (1989). "The Conditional Beta Distribution as a Stochastic Production Function." *Journal of Agricultural and Applied Economics* 71(2):370–378.

NOAA (National Ocean and Atmospheric Administration). (2015). Data Access Portal, http://www.ncdc.noaa.gov/data-access/quick-links#indices.

Okhrin, O., Odening, M. and Xu, W. (2012). "Systemic Weather Risk and Crop Insurance: The Case of China." *Journal of Risk and Insurance* 80 (2): 351–372.

Ramirez, C.A., Misra, S.K. and Field, J. (2003). "Crop-Yield Distributions Revisited." *American Journal of Agricultural Economics* 85:108–120.

SAS Institute Inc. (2014). "SAS/ETS® 13.2 User's Guide." Cary, NC: SAS Institute Inc.

Sherrick, B.J., Zanini, F., Schnitkey, G. and Irwin. S. (2004). "Crop Insurance Valuation under Alternative Yield Distributions." *American Journal of Agricultural Economics,* 86(2):406–419.

Sklar, A. (1959). "Fonctions de répartition *á* n dimensions et leurs marges." Publ. Inst. Statist. Univ. Paris (8): 229–231.

Tack, J. (2013). "A Nested Test for Common Yield Distributions with Application to U.S. Corn." *Journal of Agricultural and Resource Economics* 38 (1):64–77.

Tack, J., Harri, A and Coble, K. (2012). "More Than the Mean Effects: Modeling the Effect of Climate on the Higher Ordered Moment of Crop Yields" *American Journal of Agricultural Economics* 94(5): 1037–1054.

Torriani, D.S., Calanca, P., Schmid, S., Beniston, M. and Fuhrer. J. (2007). "Potential Effects of Changes in Mean Climate and Climate Variability on the Yield of Winter and Spring Crops in Switzerland." *Climate Research* 34(1):59–69.

Wang, J., Wang, E. and Liu, D.L. (2011). "Modelling the Impacts of Climate Change on Wheat Yield and Field Water Balance over the Murray-Darling Basin in Australia." *Theoretical and Applied Climatology* 104(3–4):285–300.

Woodard, J.D. and Garcia, P. (2008). "Basis risk and weather hedging effectiveness", *Agricultural Finance Review*, 68 (1): 99–117.

Xiong, W., Conway, D., Lin, E. and Holman, I. (2009). "Potential Impacts of Climate Change and Climate Variability on China's Rice Yield and Production." *Climate Research* 40(1):23–35.

Xu,W., Filler, G., Odening, M. and Okhrin, O. (2010). "On the Systemic Nature of Weather Risk." *Agricultural Finance Review* 70(2): 267–284.

Zanini, F.C., Sherrick, B.J., Schnitkey, G.D. and Irwin. H.S. (2001). "Crop Insurance Valuation Under Alternative Yield Distributions." Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO.

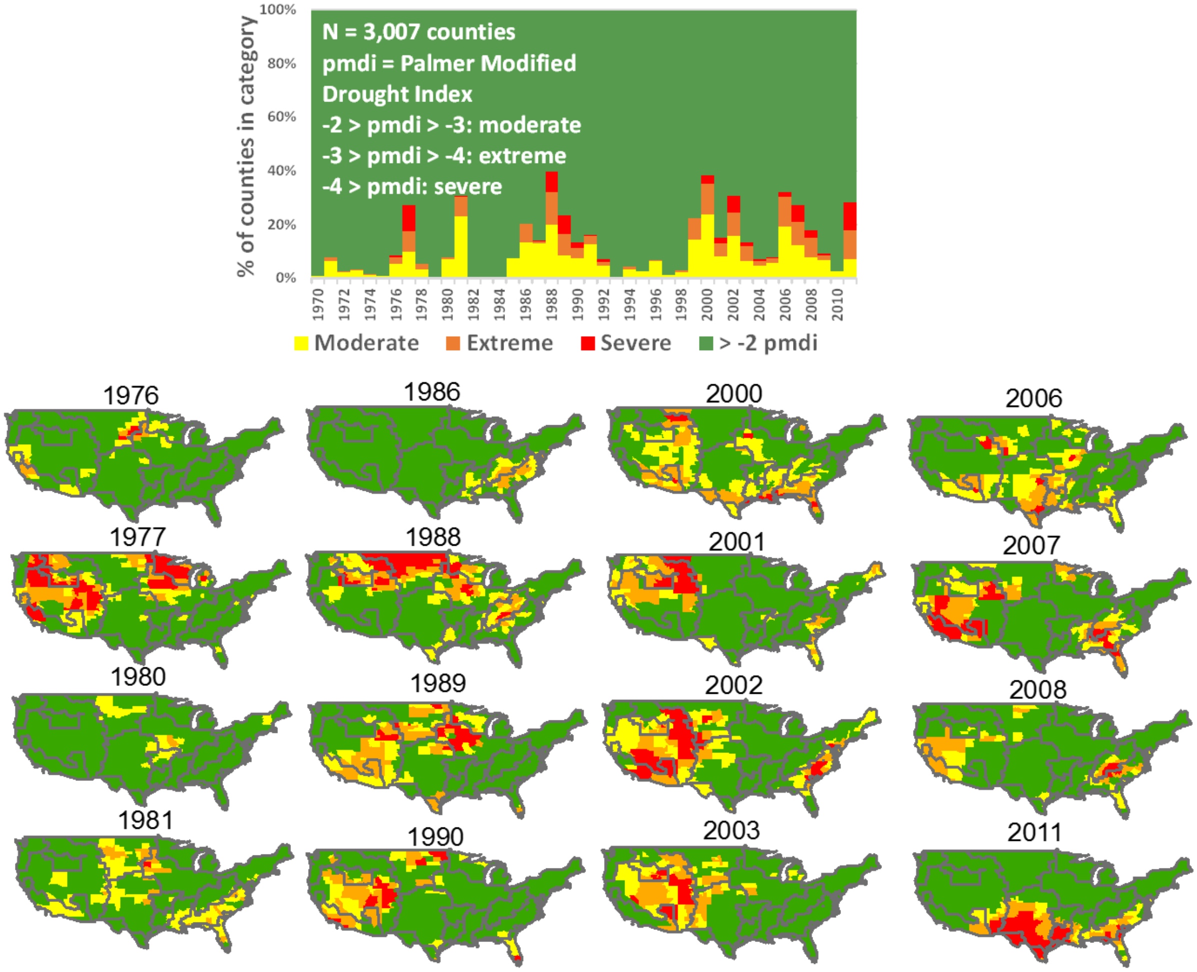

Figure 1. Historical distribution of droughts, 1970 – 2011

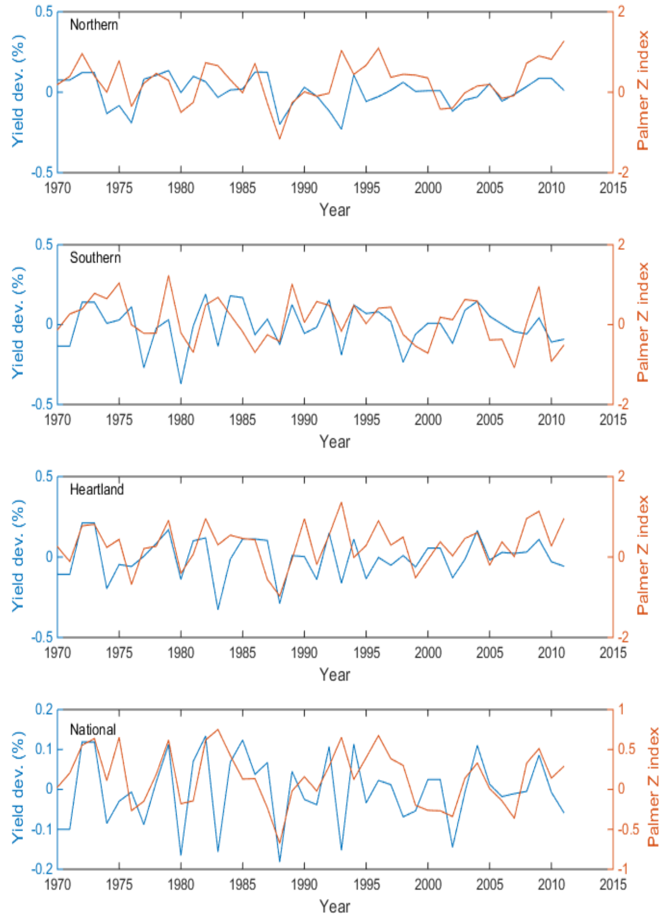

Figure 2. NOAA drought index and corn yield, 1970 – 2011 (Source: authors' analysis)

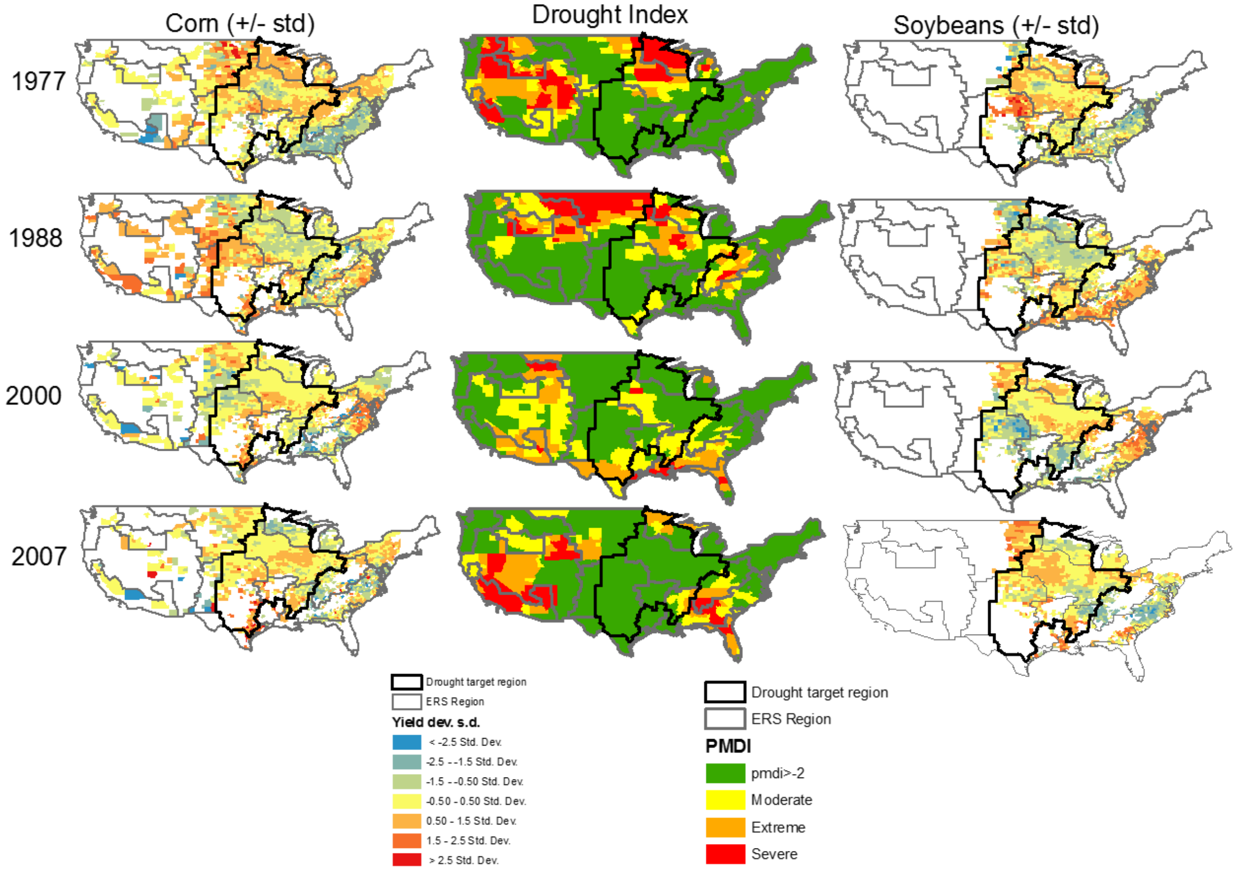

Figure 3. Yield deviates for corn and soybean and the PMDI index, 1977, 1988, 2000, 2007 (Source: authors' analysis)

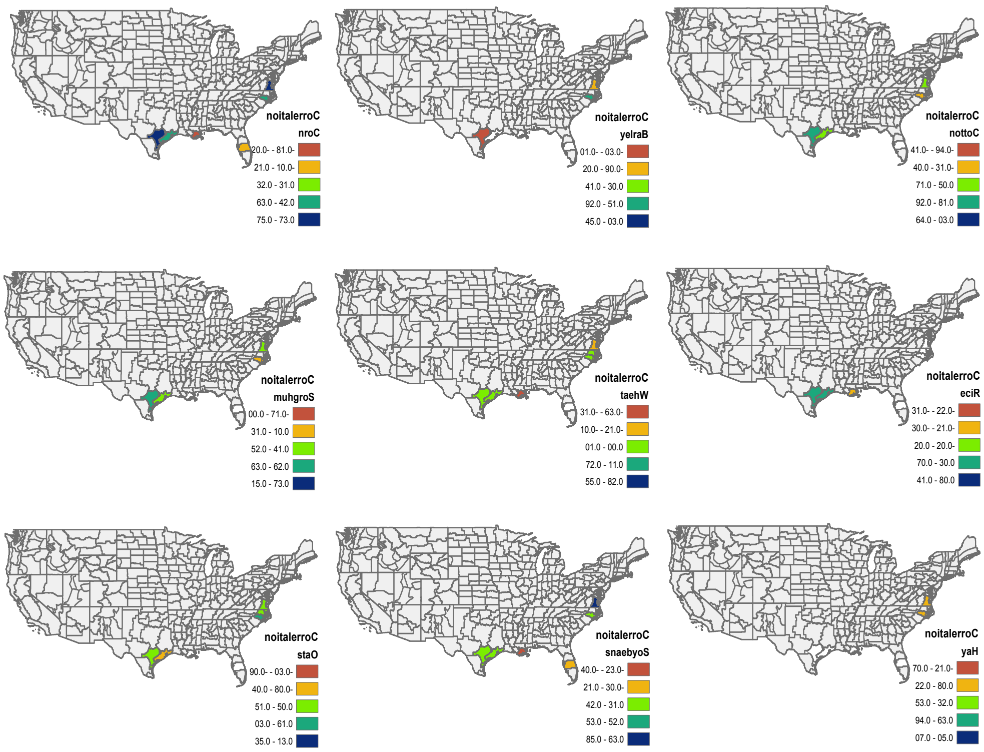

Figure 4. Kendall's tau correlations of the Palmer-Z index and corn production deviations (Source: authors' analysis)

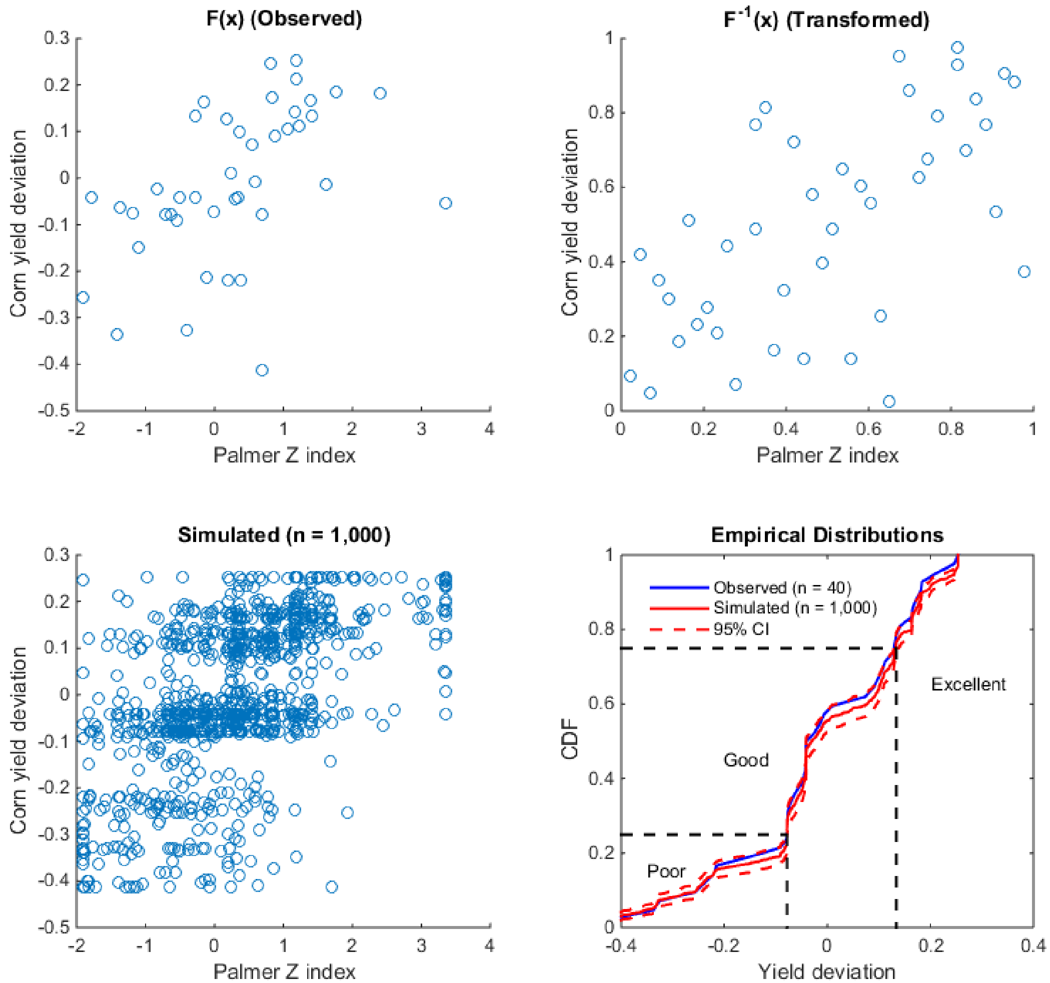

Figure 5. Generating crop yield and drought index copula with the corresponding simulated empirical distributions (Source: authors' analysis)