



Using Neural Network Classifier Approach for Statistically Forecasting Extreme Corn Yield Losses in Eastern United States

J. Mathieu, F. Aires

► To cite this version:

J. Mathieu, F. Aires. Using Neural Network Classifier Approach for Statistically Forecasting Extreme Corn Yield Losses in Eastern United States. *Earth and Space Science*, 2018, 5 (10), pp.622-639. 10.1029/2017EA000343 . hal-03880928

HAL Id: hal-03880928

<https://hal.science/hal-03880928>

Submitted on 3 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



Earth and Space Science

RESEARCH ARTICLE

10.1029/2017EA000343

Key Points:

- A neural network classifier forecasts extreme corn yield loss probability using four simple and monthly weather predictors
- Among the Eastern United States regions, 71% of yield loss extremes are well detected, with less than 15% of nonextremes misclassified as extremes
- The neural network provides a yield loss severity index that can help anticipating, in July, strong production losses

Supporting Information:

- Supporting Information S1

Correspondence to:

J. A. Mathieu,
jordane.mathieu@obspm.fr

Citation:

Mathieu, J. A., & Aires, F. (2018). Using neural network classifier approach for statistically forecasting extreme corn yield losses in Eastern United States. *Earth and Space Science*, 5, 622–639. <https://doi.org/10.1029/2017EA000343>

Received 27 FEB 2017

Accepted 23 MAY 2018

Accepted article online 30 MAY 2018

Published online 27 OCT 2018

Using Neural Network Classifier Approach for Statistically Forecasting Extreme Corn Yield Losses in Eastern United States

J. A. Mathieu¹  and F. Aires^{1,2} 

¹Laboratoire d'Etudes du Rayonnement et de la Matière en Astrophysique et Atmosphère Observatoire de Paris, Paris, France, ²Estellus, Paris, France

Abstract This paper presents a statistical method for forecasting extreme corn yield losses caused by weather extremes. A neural network classifier approach is tested over the Eastern United States (time series of 35 years) to detect extreme yield losses for corn from weather-related information. We first developed a methodology to rank a series of climate-based predictors according to the accuracy with which they classify extreme from nonextreme yield losses. The classification methodology is adapted in order to be trained with a limited number of extreme cases. Using four weather predictors—the average temperature in July and August, and the SPEI (Standardized Precipitation-Evapotranspiration Index) in June and July—71% of the extreme cases are well classified by this statistical model. Furthermore, the neural network output represents a good yield severity index and can provide an early quantitative warning for extreme yield anomalies.

Plain Language Summary The impact of weather on agriculture is very important and requires building models that describe the links and effects between weather and crop growth. These models are often efficient but they have great difficulty predicting very low crop yields. These situations refer to extreme yield losses and require other methodologies. Indeed, existing models are not efficient. Besides, very few data are available for these extreme events, which makes models difficult to calibrate. Moreover, the choice of good weather predictors is critical for the quality of the forecast. Our methodology identifies the weather indicators that best discriminate between extreme and nonextreme yield losses. We compare the ability of a variety of weather indicators to represent extreme yields: these indicators encompass single climate variables, water supply indices, but also agroclimatic indices calculated for a month, a year, or a growing season. The goal of the statistical modeling is not to estimate a yield but instead to define a yield severity index that would indicate extreme yield losses. It is extremely difficult to predict precisely in July or August the extreme losses. That is why we are searching for a probability of extreme cases rather than a strict prevision of the yield.

1. Introduction

“CHICAGO (August 27, 2013): Extreme weather forced the Federal Crop Insurance Program to pay out a record-breaking \$17.3 billion in crop losses last year. Payments made to farmers to cover losses from drought, heat and hot wind alone accounted for 80 percent of all farm losses, with many Upper Midwest and Great Plains states hit hardest. According to the Natural Resources Defense Council much of these losses could have been prevented using water-smart strategies.” (Natural Resources Defense Council, 2013). With a possible increasing frequency of abnormal climate conditions in the coming decades, it is more and more important for crop yield forecasting systems to accurately predict the impact of weather anomalies. Farmers require accurate estimates for crop insurance purposes, forward marketing and delivery planning, planning harvest or storage requirements, and cash-flow budgeting. Several actors (from growers, to planners, distribution centers, insurances, policy makers, and commodity traders) are severely impacted by abnormal low yields, so it is crucial to anticipate these events (Schlenker & Roberts, 2009).

Low yields can also be related to disease outbreak (Oerke, 2006) or political turmoil (Macours & Swinnen, 2000). Cerda et al. (2017) quantified yield losses of coffee due to pests and diseases and identified the most important predictors of coffee yields and yield losses. Disease impact or political turmoil are present in the yield anomaly record, but we focus here on the weather data that cannot explain such anomalies.

©2018. The Authors.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Thus, a weather information-based model will only be able to predict a percentage of the yield anomalies variance. Note, however, that weather impacts can also induce pest and disease outbreaks through indirect effects (Prakash et al., 2015).

The climate-based predictors used for yield forecasts are often chosen as simple weather variables such as temperature and precipitation. They are then applied on mechanistic or statistical models. For instance, Kotlowski (2007) used a nonparametric regression with a single climate variable (temperature or precipitation) to show that summer months' (June–August) weather variables are the most important factors explaining U.S. corn yield production. A long list of agroclimatic indices have been analyzed instead of, or in addition to, simple weather predictors, most often at the regional scale: monthly mean temperature, solar radiation, cumulative precipitation, and evapotranspiration (Mathieu & Aires, 2018; Peltonen-Sainio et al., 2010). In Gouache et al. (2015), each model is a linear combination of five to seven climate variables, at the departmental level. Landau et al. (2000) have built a parsimonious model in which every parameter reflected a known climate effect on the UK crop–environment system to allow mechanistic interpretation. Bannayan and Sanjani (2011) identified relationships of low- and high-wheat yield years to maximum and minimum air temperatures with a simple correlation analysis at the regional level. A multivariate analysis in Gobin (2010) showed significant statistical relationships between yield and vapor pressure deficit, temperature, growing season length, water logging, and drought; these predictors were included in a regional crop model. Despite all these publications on predicting yields in general, few of them focused on extreme agricultural events (important yield losses, for instance).

Many papers focus on the impact of extreme climate events on national food balances (M. Brown & Funk, 2008); heat waves are more frequent in recent years with important consequences on crop yields (Boyer et al., 2013; Fink et al., 2004). However, extreme weather events do not always cause extreme crop yield: it depends on occurrence, crop sensitivity, and farmers adaptation (e.g., use of irrigation, adjustment of sowing and harvest dates, and selection of adapted crops, Olesen et al., 2011; van der Velde et al., 2010). Thus, many studies have considered events in agriculture (Anaman, 2002; Lal et al., 2012; Triantafyllou & Sarris, 2011) but by focusing on extreme weather, not extreme low yields. For example, Powell and Reinhard (2016) showed, with econometric techniques, that “the number of days with extreme high temperatures [...] has significantly increased since the early 1900s.” Motha (2011) made an overview of the flooding, droughts and hurricanes that impacted food production. Fink et al. (2004) (in Europe) and Boyer et al. (2013) (in the United States) analyzed extreme weather events, but not extreme yields. Others studies aim to improve the predictive ability of biophysical crop and grassland simulation models under extreme weather conditions (mainly high/low temperatures and water deficit/excess, Bellocchi et al., 2014). Sivakumar et al. (2005) commented on the forecasting of extreme events occurrence, but they do not provide an input model able to estimate extreme low yield probabilities.

Few studies actually focus specifically on extreme yields by using agroclimatic indices (they generally intend to predict the interannual variability instead). Ciaia et al. (2005) analyzed extreme drought and heat to show that productivity reduction in eastern and western Europe can be explained by rainfall deficit and extreme summer heat, respectively. Waha et al. (2013) analyzed how much of the changes in historic agricultural production in major producing regions can be attributed to the occurrence of extreme weather events. McCool et al. (1986) and Debaje (2017) determined crop yield losses from air pollutants (e.g., ozone), and Vergara-Diaz et al. (2015) evaluated the accuracy of vegetation indices in predicting grain yield and assessing disease severity. Finally, using weeds density or weed relative leaf area, Edalat et al. (2011) evaluated the goodness of fit of different yield loss models during the 2008 and 2009 growing seasons.

Several studies that have attempted to predict yields from weather predictors have highlighted the difficulty of accurately predicting the extremely low yields (Mathieu & Aires, 2016). In agricultural studies, one of the first difficulties is the small amount of data and this difficulty is exacerbated when looking at extreme data. Extra care must then be taken in the methodology to ensure obtaining robust extreme models. This requires addressing several difficulties such as the following: the very limited amount of data available to properly calibrate a statistical model, the need to make some ad hoc methodological choices with potential important consequences, or the limited information content of the weather predictors, or the need to define reliable diagnostics for realistically measuring the quality of the models.

Recently, Ben-Ari et al. (2016) provided a very interesting and well-documented study on the corn and wheat yield over France and Spain, comparing the ability of both (agro)climatic predictors and model simulations

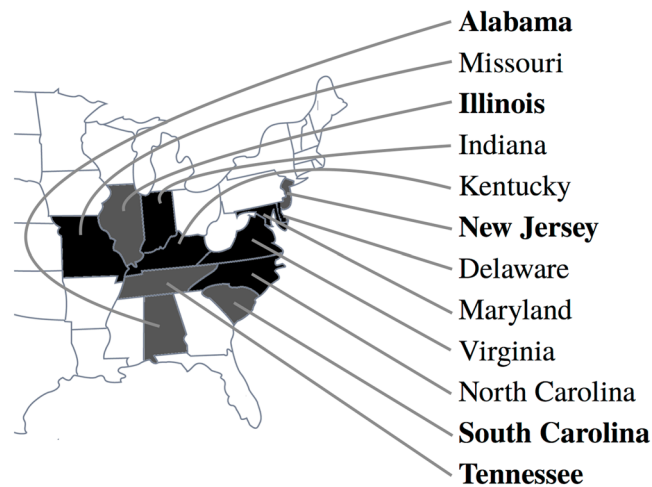


Figure 1. Spatial distribution of the 12 considered states. Bold names indicate the states that will be emphasized in the following.

(WOFOST) to predict extreme yield losses. Similarly to our study, Ben-Ari et al. (2016) also quantify in length the robustness of their results. In particular, the sensitivity to several methodological choices (such as the used thresholds or the definition of extremes). In their study, a sensitivity analysis was performed independently for each predictor which is a limitation they propose to explore in their perspectives. In fact, Kebede et al. (2012) showed that the combined effect of heat and severe moisture deficit stress has a much greater impact for nonirrigated corn than their independent effect. Similarly, Schlenker and Roberts (2009) explored interactions between temperature and precipitation outcomes and showed that greater precipitation partially mitigates damages from extreme high temperatures.

Our paper is an extension of the Ben-Ari et al. (2016) and Mathieu and Aires (2016) studies. It uses a nonlinear statistical model that intends to detect extreme yield losses of corn over the Eastern United States from weather-related information only. We compare and measure in a systematic way the ability of a variety of weather predictors to detect the extreme corn yields. These predictors encompass single climate variables, water supply indices, and also agroclimatic indices calculated for a month, a year, or a growing season to assess which type of information is the more related to extreme yields. Our aim is to provide a methodology that (1) identifies the weather predictors that best discriminate between extreme and nonextreme yield losses (considering as extreme, an event whose probability of occurring is less than 5%), (2) analyzes in agricultural terms, this ranking of the agroclimatic predictors and their combination, and (3) with a simple but accurate classification model, uses the best combination of predictors to estimate the probability of having an extreme yield loss.

Section 2 introduces the databases used in this study. The methodological aspects are summarized in section 3, and the main results are presented in section 4. Finally, section 5 presents our conclusions and perspectives for this work.

2. Database

2.1. Yield Data

Maize is one of the major crops in the United States, and we focus here on this cereal. The corn yield data (in bushel per acre) were collected by the U.S. Department of Agriculture (USDA) at the county level (more than 700 counties). A long historical record is available from at least 1979 to 2013, but the yield time series are not complete in all counties. To analyze extreme yield events, we focus on the 12 more weather-sensitive states (see Figure 1) as identified in Mathieu and Aires (2018).

The U.S. Drought Monitor provides a national database to track the severity of droughts in the United States. For long-term drought characterization, the Palmer Drought Severity Index (PDSI) (Palmer, 1965) is the most widely used drought index. A 50-year high-resolution global data set of PDSI can be found at Sheffield et al. (2006, 2012). Extreme yield losses often correspond to a drought event (Figure S1 in the supporting information).

2.2. Weather Data

Monthly and daily minimum and maximum temperature, as well as monthly precipitation data were collected for the period 1979–2013, from the ERA-Interim reanalysis of the European Center for Medium-Range Weather Forecasts (Uppala et al., 2005). De Wit et al. (2010) demonstrate for Europe that the ERA-Interim data set is highly suitable for regional crop yield forecasting. The U.S. territorial organization and its subdivision into counties are used to project the ERA-Interim data, from its original 75 km × 75 km grid scale into a county level scale.

2.3. Agroclimatic Indices

Agroclimatic variables are derived from monthly and daily temperature and rainfall data in order to generate meteorological variables that are believed to impact the corn yield. Generally, the agroclimatic variables intend to represent the thermal or the water impact on a crop. Thermal indices include cumulative measures (such as Growing Degree Days, Corn Heat Units, Cold Degree Days), threshold measures (e.g., number of days for which temperature is above 30 °C), or data-driven measures (such as first fall frost date, length of the growing season, and date of the growing season start/end). Water indices include cumulated precipitation on different periods of time (growing season or summer months), or soil moisture (volumetric soil water layer 2, from ERA-Interim). Formulas used to calculate these indices are available, for instance, in Lepage et al. (2012) and are summarized in Table S2. Standardized Precipitation-Evapotranspiration Index (SPEI) comes from Begueria et al. (2014), which calculated the potential evapotranspiration with the FAO-56 Penman-Monteith estimation.

In order to more easily compare the various agroclimatic variables in terms of normalized anomalies, we decide in the following, to resscale them as relative anomalies from −1 to 1. In the following, *predictor* will refer to weather relative anomaly.

3. Methodology for Extreme Events Detection

This section details the methodology used to define the extreme yield loss (section 3.1) and the statistical tools used to predict them (sections 3.2 to 3.5).

3.1. Extreme Yield Loss Definition

It is first necessary to distinguish the influence of climate from the other factors—such as fertilization, soil type, or mechanization—on maize yield. Nonweather factors are assumed to correspond to long-term variations on yield time series $y(t)$, whereas weather factors correspond to interannual variations along the long-term trend. As proposed in Mathieu and Aires (2016), a nonlinear mixed-effect model (ME) with a sigmoid function is used to represent the annual trend $\bar{y}(t)$. The detrended and reduced yield a is given by the following:

for each year t ,

$$a(t) = \frac{y(t) - \bar{y}(t)}{\bar{y}(t)} \quad (1)$$

Yield anomaly $a(t)$ represents the corn yield variations as a percentage around the trend. For instance, $a(t) = -0.5$ means that the yield was 50% lower than the expected one from the interannual trend of year t . Although equation (1) represents the relative anomaly, rather the actual anomaly [which is the numerator in equation (1)], it will be referred hereafter as simply *anomaly*.

Figure 2 illustrates the frequency distribution of a values. This histogram takes into account all the 12 states and the 35 years (1979–2013). Most values (82%) range from −0.3 to 0.3. Forty-six percent of the anomalies are negative. Positive outliers (larger than 0.5) are extremely infrequent because it is rare to obtain +50% yield in one particular year. Here, we only focus on extreme yield losses (extreme negative anomalies). In order to define what is an extreme yield loss for our application, we have to choose a good threshold T beyond which all anomalies will correspond to an extreme yield loss. A good threshold on yield anomaly for low extremes should not be too high (to take only extreme cases) and not too low (to get enough extreme observations). We consider as extreme an event whose probability is less than 5% (this is an ad hoc choice, that is discussed in the supporting information, Bommier, 2014). In terms of anomaly threshold, it corresponds to the value $a < -0.45$. In the supporting information S1, results obtained with this threshold are compared with those obtained with other thresholds to assess the robustness of our conclusions.

3.2. Working With Unbalanced Data Set

By definition, the proportion of extremes (5%) is significantly lower than nonextreme cases. Our approach uses a classification model to discriminate between extreme and nonextreme cases. When calibrated on such unbalanced data set, the classification would emphasize nonextreme cases. In order to obtain good performances on the extreme cases, the classification methodology needs to be adapted. Different resampling techniques have been proposed in the literature (More, 2016). We choose here a simple oversampling method that consists in duplicating the elements of a (minority) class a certain number of times until the two classes have similar number of samples.

3.3. Training and Validation

In statistical learning, the data set is divided into a training set and a validation set. The training set is used to estimate the model parameters value and the validation set is used to assess the strength and utility of the predictive relationship after estimating the model parameters. This section emphasizes two factors to be considered when building the training and validation data sets: space correlations and extreme distribution.

Climate data are spatially correlated. Therefore, training and validation sets are rarely independent if samples are randomly chosen in the overall weather data set. As shown by Le Rest (2013), the cross-validation method in presence of spatial autocorrelation needs to be adapted, otherwise it will lead to nonreliable results and false conclusions. Indeed, one of the hypotheses in cross validation is that the samples are independent from each other. In geosciences, due to spatial correlations (i.e., two close pixels) or temporal correlations (same pixel but two consecutive days), samples are often not totally independent from each other. And this can make the cross validation not reliable.

Suppose that the data of a county c of a year t belongs to the training set. The aim of training is to estimate at best the county c . Let us suppose that other data of the same year belong to the validation set: the counties close to the county c are very similar to c in terms of weather and yield (the behavior regularity can be seen on the annual weather or yield maps), and will therefore be very well estimated by the model. This efficient sampling is therefore not reliable to validate the model (i.e., to test its ability to generalize and to provide a good estimate even when data are new from the training set).

In order to use the cross-validation method, the training set must be independent from the validation set and no close counties of the same year can be distributed in the learning and validation sets. If we do not do that, the results will certainly be excellent, but not realistic. To solve this issue, our analysis uses independent years for the training and the validation data sets (as weather from one year is not correlated to the next one (Danforth & Yorke, 2006; Slingo & Palmer, 2011)).

The fact that the two considered classes have a similar size (thanks to the oversampling method) does not provide a greater diversity of extreme cases; both the training and the validation sets should include a sufficient number of extreme years to obtain satisfactory results, that is, the best possible estimate of extreme cases, while limiting the errors of generalization (in the validation set). Among the 35 available years, four of them (1980, 1983, 1988, and 2012) account for a very high number of extreme anomalies in the counties. These years correspond to a severe year of drought. Six years (1991, 1999, 2002, 2003, 2005, and 2011) account for a small number of extreme anomalies. Only two of them correspond to a year of drought. A cross-validation analysis shows that the best results are obtained when the training set includes two or three of the four *large extreme years* and at least three of the six *small-extreme years*. The other extreme years are then used in the validation set.

Thus, to determine the training and validation sets, we use a cross-validation method, adapted to both (1) the noncorrelation of two successive years (cf. comment above), and (2) to the distribution of years comprising a large number of extremes in the training set.

3.4. Classification Criteria

As shown in Mathieu and Aires (2016), a regression analysis can be used to estimate yield but the forecast of extreme yield losses is very difficult because there are less data to fit the statistical model. We decide instead here to forecast the probability of being an extreme yield loss year, which is less ambitious.

In this paper, we focus on the detection of extreme corn yield losses and a specific model needs to be developed. The goal is to estimate the probability that a year will be extreme or not. We denote by $\mathbb{P}(\text{extreme} \mid \text{weather})$ the conditional probability of facing an extreme yield loss given some weather

Table 1
Confusion Matrix of a Classification Problem With Two Classes, 0 (Negative) and 1 (Positive)

		Predicted			
		1	0		
Observed	1	TP	FN	$TPR = \frac{\sum \text{True Positive}}{\sum \text{Positive}}$	$FNR = \frac{\sum \text{False Negative}}{\sum \text{Positive}}$
	0	FP	TN	$FPR = \frac{\sum \text{False Positive}}{\sum \text{Negative}}$	$TNR = \frac{\sum \text{True Negative}}{\sum \text{Negative}}$

Notes. Each column represents the instances in a predicted class, while each row represents the instances of the real classes. Each measure defined here can be used to quantify the performance of a classifier. The acronyms are the following: TP = true positive; TN = true negative; FP = false positive; FN = false negative; TPR = true positive rate; TNR = true negative rate; FPR = false positive rate; and FNR = false negative rate. For more details, see text (section 3.4).

conditions. In this section, we introduce some criteria that can be used to assess the performance of a weather predictor and help in comparing the several classifiers.

A corn yield anomaly lower than a given threshold T (to be defined) will be considered as extreme (Figure S2). In order to work with class and not anomaly values a , extreme yield anomalies were assigned the value 1, and nonextreme the value 0. Table 1 defines all common measures used to quantify the quality of a classifier when two classes are considered. True positive (TP) represents the number of positive (i.e., extreme) well classified as positive. False negative (FN) refers to the number of positive incorrectly classified as negative. False positive (FP) represents the number of negative (i.e., nonextreme) classified as positive (so, misclassified). Finally, true negative (TN) refers to the number of negative well classified as negative.

TP rate (TPR), FN rate (FNR), FP rate (FPR), and TN rate (TNR) refer to a percentage. TPR represents the proportion of positive well classified, and FNR represents the proportion of positive incorrectly classified. Similarly, FPR is the ratio between the number of negative events wrongly categorized as positive, and TNR measures the proportion of negatives that are correctly identified as such (Fawcett, 2006). The goal of a good classification is to get a high TPR and TNR, while limiting FNR and FPR.

3.5. Classification Model

A statistical model is first trained to predict the state of a yield (extreme loss or not) given several weather variables. The output of this model is a yield losses severity index that behaves like the probability for the county to face an extreme yield loss. The yield losses severity index is modeled using a neural network (NN) with one hidden layer of 10 neurons and one output neuron. All transfer functions are tan-sigmoid. The network is trained using weather variable as inputs and binary outputs (extreme or not) and doing so conducts the network to predict a real value between 0 and 1 that estimates the probability to be extreme or not (Bishop, 1996). This NN is trained using the Levenberg-Marquardt back-propagation algorithm over a balanced sample of 50,000 observations at the county level (with 50% of the cases being of an extreme yield loss). The NN returns a real value between 0 and 1, but we actually need a binary value to decide if it is an extreme (1) or a nonextreme (0) year. It is therefore necessary to define a *cutoff point* C on the NN output that will discriminate between 0 and 1 states. For instance, if $C = 0.5$, all NN outputs lower than 0.5 will be classified as *nonextreme* and all outputs higher than 0.5 will be classified as *extreme*. The values of the quality measures described in Table 1 will depend on this chosen cutoff point. Figure 2 synthesizes the several steps and data sets used in this study.

3.6. Receiver Operating Characteristic Curve

The receiver operating characteristic (ROC) curve is a classical classifiers evaluation tool (C. Brown & Davis, 2006; Fawcett, 2004). In a ROC curve, TPR and FPR are plotted, based on the cutoff point choice. Each point on the ROC curve is associated to a particular cutoff point (decision threshold) that discriminates between the two groups. A perfect discrimination (no overlap in the distributions of the two groups) would have a ROC curve that passes through the upper left corner (Zweig & Campbell, 1993). Accuracy is measured by the area under the ROC curve. An area of 1 represents a perfect classification; an area of 0.5 represents a useless classification. This area measures the discrimination quality, that is, the ability to correctly classify extreme from nonextreme cases.

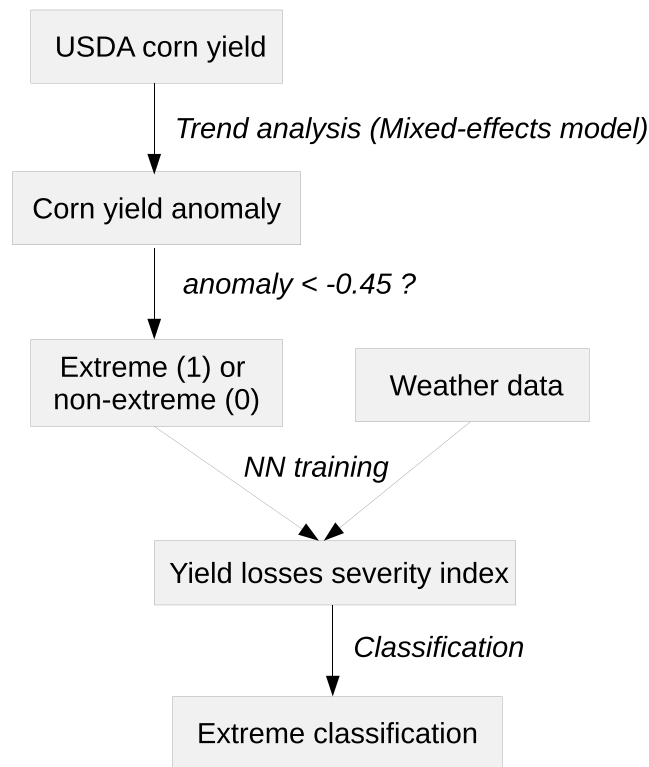


Figure 2. Scheme of the data sets and methods used in the study.

Figure 3 (right), represents our application, and in a more explicit way, the TPR and FPR rates when the cutoff point value is varying (x axis). Usually, the chosen cutoff point C is the one that minimizes the sum of false negative and false positive (FN + FP), that is, the number of misclassified events, but other quality measures can also be used to choose it. Indeed, in addition to the minimization of FN + FP, we also want avoiding a too high FPR. Thus, we can define a maximum value of FPR that should not be exceeded. For instance, we can choose to have $FPR < 15\%$. So, two conditions may be checked to choose the best cutoff point C (see Figure 3, right):

$$\begin{aligned} &\text{Minimization of } FN + FP \\ &\text{Under the constraint } FPR < 15\% \end{aligned}$$

In Figure 3 (right), the cutoff value minimizing FN + FP is highlighted by a black vertical line, while the $FPR < 15\%$ threshold is denoted by the green vertical line. For this classifier, two different results are obtained when using these two cutoff values: the TPR is certainly better (84% against 75%) when we only minimize FN + FP (intersection between the blue curve and the black vertical line), but the FPR is larger (21% against 15%) —intersection between the orange curve and the vertical black line. This shows that, instead of using simple quality measures, it is preferable to combine them with additional constraints. Impact of these constraints on the results is tested in section 4.3.

3.7. Conditional Probabilities

The link between weather and extreme yield can be analyzed in two symmetric ways. First, given extreme/nonextreme information, what can be said about the weather conditions? This first point of view is linked to the conditional probability $\mathbb{P}(\text{weather} \mid \text{extreme})$ of observing certain climatic events for an extreme low yield year. Second, given the weather information, what can be said about the extreme/nonextreme probabilities? This second point of view is linked to the conditional probability $\mathbb{P}(\text{extreme} \mid \text{weather})$ of observing extremely low yields at the end of the year, given a weather state for this year. The conditional probabilities $\mathbb{P}(\text{weather} \mid \text{extreme})$ and $\mathbb{P}(\text{extreme} \mid \text{weather})$ are different and they do not inform us about the same thing. The comparison of these two conditional probabilities is given in section 4.1.

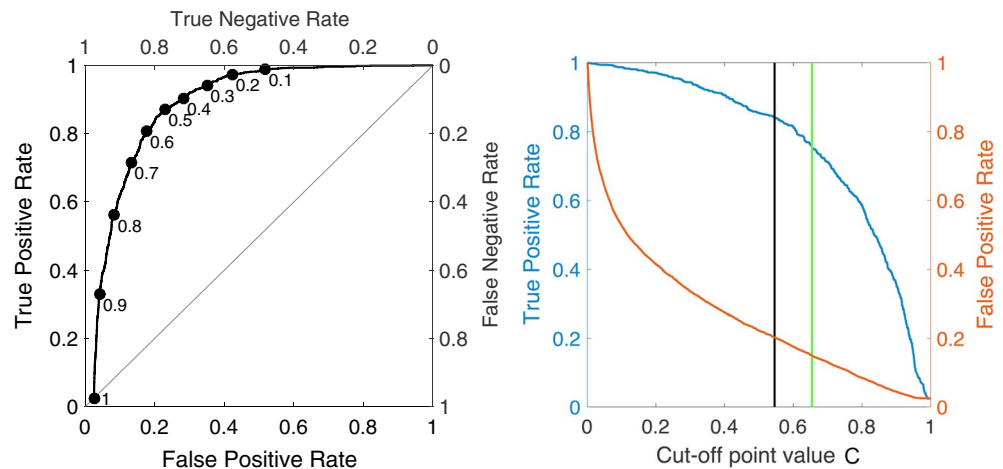


Figure 3. Choice of the cutoff point C . (left) A regular ROC curve in our application with the four different quality measures. Black points and numbers represent different values of the cut-off point. (right) The variations of TPR (left axis) or FPR (right axis) as a function of the cutoff point value (x axis). The cutoff value that minimizes the sum of false negative and false positive ($FN + FP$; i.e., the number of misclassified events) is indicated by the black vertical line, whereas the cutoff value that minimizes $FN + FP$ under the constraint $FPR < 15\%$ is indicated by the green vertical line.

4. Results

Section 4.1 identifies how weather predictors are distributed for extreme years, and how extreme years are distributed according to weather predictors. Section 4.2 presents a ranking analysis that compares climatic predictors according to their discriminating performance (i.e., between extreme and nonextreme yield loss cases). Section 4.3 shows results of the classification analyses.

4.1. Joint Distribution of Weather Predictors and Extreme Years

Figure 4 represents the distribution of four important weather predictors T_{july} (mean temperature in July), T_{august} (same for August), $\text{SPEI}_{\text{june}}$ (mean Precipitation-Evapotranspiration Index in June), and $\text{SPEI}_{\text{july}}$ (same for July). These predictors have been selected by Mathieu and Aires (2018) as those which best estimate maize yields. Besides, the choice of T and SPEI for the months of June, July, and August, is supported by the Kotlowski (2007) analysis, as mentioned in section 1. In these plots, data have been separated into those that are related to an extreme yield loss and those that are not. Weather variables with the smallest intersection of the two distributions will be best predictors for the extreme detection. In Figure 4, the vertical black lines delineate the optimal value (henceforth called ρ) of the weather predictor that discriminates better extreme and nonextreme events. For T_{july} , the optimal value is $\rho = 0.05$; for $\text{SPEI}_{\text{july}}$ -0.35 ; for $\text{SPEI}_{\text{june}}$ -0.15 ; and for T_{august} -0.05 . This threshold is successively calculated by maximizing the TPR when we classify as extreme the data with $T_{\text{july}} > \rho$ (or $T_{\text{august}} > \rho$, or $\text{SPEI}_{\text{june}} < \rho$, or $\text{SPEI}_{\text{july}} < \rho$). These results depend on the threshold taken to define the extreme anomalies (see section 3.1).

Figure 5 represents the scatter plots of four weather predictors ($\text{SPEI}_{\text{june}}$, $\text{SPEI}_{\text{july}}$, T_{july} , and T_{august}) that have been selected, thanks to the histogram analysis of Figure 4. Red dots are for extreme cases and green dots for nonextreme cases. The value written on this four subfigures corresponds to the thresholds identified in Figure 4. The probability $\mathbb{P}(w | e)$ is the proportion of extreme (red) dots conditional to a weather state (e.g., weather conditions w under or lower than a certain threshold). For instance, on Q1, we can see that when we observe a particular weather state w , $T_{\text{july}} > 0.05$ and $\text{SPEI}_{\text{july}} < -0.35$, 64% of the events are extreme cases. In other words, $\mathbb{P}(T_{\text{july}} > 0.05 \text{ and } \text{SPEI}_{\text{july}} < -0.35 | \text{extreme}) = 0.64$. This means that this particular weather state is particularly prone to extreme corn yield losses. The probability $\mathbb{P}(e | w)$ refers to the proportion of extreme data (red) when knowing that they are on a specific subsquare. For instance, on the top left figure, we can see that 20% of the data with $T_{\text{july}} > 0.05$ and $\text{SPEI}_{\text{july}} < -0.35$ are related to extreme yield losses.

How are weather predictors distributed for extreme years? Scatter-plots of each predictor against the others have been plotted. In Figure 5, four of them are illustrated. As mentioned in section 2.3, we consider weather anomalies. Scatter plots show that extreme data are not randomly distributed. During years with extreme yields, we note that $T_{\text{july}} \gg 0$, $\text{SPEI}_{\text{july}} \ll 0$, and $T - \text{SPEI} \gg 0$ in summer months; that is, the gap between

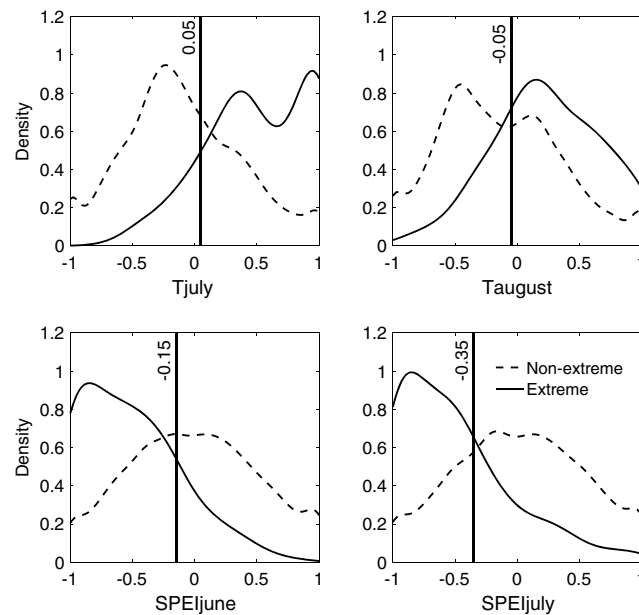


Figure 4. Histogram of four important weather predictor (Tjuly, Taugust, SPEIjune, and SPEIjuly) anomalies, after separating data into extreme (solid line) and nonextreme (dashed line) yield loss events. The vertical black line delineates the best (p) value that discriminates between extreme and nonextreme case for the four specific weather variables.

temperature and SPEI is highly pronounced (Figure 5). However, the converse is not true: in a lot of cases, these three items are checked although they do not refer to an extreme yield.

How are extreme years distributed according to weather predictors? Figure 5 reports on the climatic conditions observed for extreme yield loss: the probability of observing Tjuly > 0.05 and SPEIjuly < -0.35 is equal to 64%. Conversely, the likelihood that the yield loss will be extreme if Tjuly > 0.05 and SPEIjuly < -0.35 can be very different. Figure 5 shows that if Tjuly > 0.05 and SPEIjuly < -0.35 , 20% of the cases with this configuration will give an extreme yield loss (QI, at the bottom right).

4.2. Ranking Variables

In this paragraph we identify the weather predictors that best explain extreme yield losses in order to discriminate extreme and nonextreme cases. According to Figure 4, weather predictors that reveal more information about extreme yield losses are Tjuly, SPEIjuly, and SPEIjune. Not all meteorological or agrometeorological variables have an ability to predict corn yield fluctuations. Some of them give information on the final yield because they represent a biological phenomenon and they intervene at a critical time in the development of the plant. Therefore, as shown in Kotlowski (2007) and Mathieu and Aires (2018), information varies strongly from a weather predictor to another: there is a small number of strong predictors for corn yield and many predictors with hardly any influence on yields.

As mentioned in section 3.4, we denote by 1 the extreme yield loss anomalies and by 0 the others. Our procedure to select the best predictors uses a simple NN. A forward induction process is adopted: 50 different learning sets are chosen; for each one, the predictor that obtains the higher TPR in generalization is selected. We select as first predictor the weather indicator that is most frequently selected among the 50 trials (i.e., SPEIjuly). Considering SPEIjuly as first input of a NN classifier, the second input is similarly chosen among all the predictors as the one that, combined with SPEIjuly, defines a two-input NN classifier that gives the greatest TPR in generalization. This process is repeated until a desired number of inputs is selected. The second and third selected inputs are Tjuly and SPEIjune.

We tested a total of 59 weather variables or agroclimatic indices (Table S1). The selection process allows us to classify them according to their ability to discriminate extreme/nonextreme yields. Once this ranking is done, as it is not desirable to keep 59 inputs, a cross validation is used to know how many we should keep: the first four inputs (Tjuly, SPEIjuly, SPEIjune, and Taugust) are sufficient since classification does not improve by adding a fifth variable.

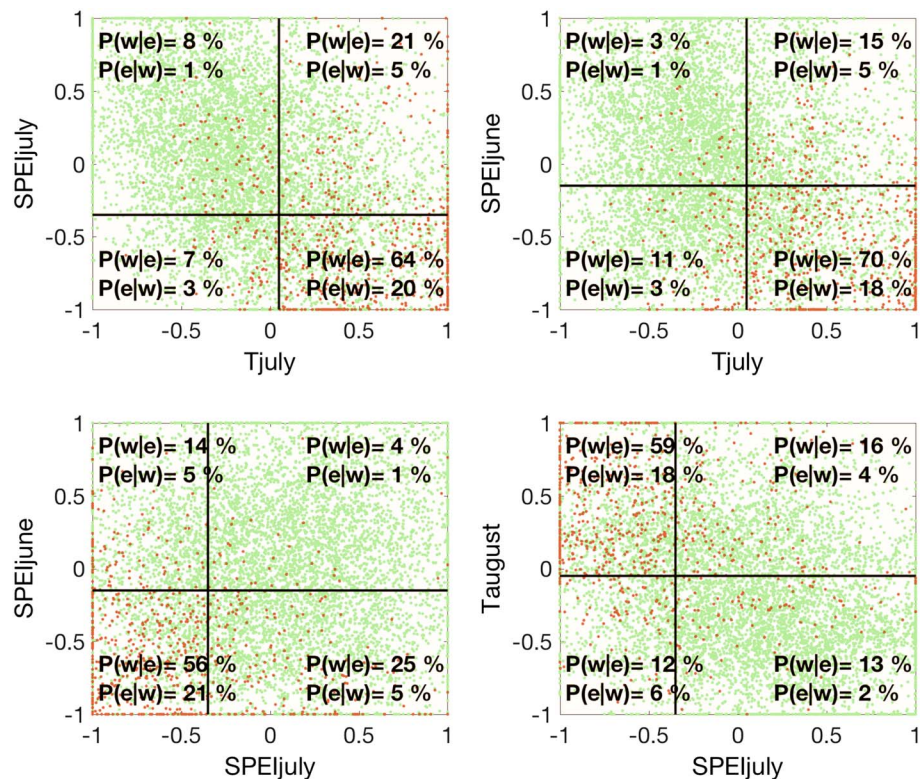


Figure 5. Scatter plots of four weather predictors: SPEIjune, SPEIjuly, Tjuly, and Taugust. Red dots are for extreme cases and green for nonextreme. Horizontal and vertical delimitations have been defined in Figure 4. As noted in section 3.1, 5% of yield anomalies are extreme anomalies.

Keeping in mind the constraint $FPR < 15\%$, the TPR rises from 50% (with only the SPEIjuly input) to 63% with the two first selected inputs (SPEIjuly and Tjuly). Adding SPEIjune, the TPR reaches 70%. The TPR levels off at 73% with the fourth input (Taugust).

4.3. Detection Model for the Classification of Extreme Yield Losses

NNs are common models for classification task. In this study, we use the *patternnet* function of the Matlab software to build a NN classifier with 10 neurons in its hidden layer.

As explained in section 3.4, the quality measure used to define the cutoff point enables a trade-off between the usual minimization of $FN + PF$ and a subjective constraint on FPR (e.g., $FPR < 15\%$). Table 2 gives the TPR obtained with a constraint on FPR varying between 5% and 40%: in each case, in order to quantify the impact of the training and the validation set on the results, 200 NNs are run with a different training and validation set.

Considering the constraint $FPR < 15\%$, our classifier can detect 71% of extreme yield losses. In other words, $FNR = 100 - 71\% = 29\%$ of extreme years are misclassified. When we make the constraint on FPR less restrictive, the proportion of extreme years well detected (TPR) rises but more and more years are considered as extreme, whereas they are not (FPR rises too). Standard deviation of the results are quite stable when the constraint on FPR is varying. It enables checking the results robustness and the generalization quality of the classifier. Figure 6 quantifies the sensitivity of TPR and FPR to the training and validation sets. The NN classifier is run with 200 different sets of training and validation. Average value of TPR and FPR are illustrated and a confidence range is given with the second and the eighth decile. These results are satisfactory: sensitivity of TPR to the training set is not negligible but considered natural for this study. The TPR varies more with the training set than the FPR: the mean standard deviation for TPR is 0.08 and 0.05 for FPR. The largest standard deviation for TPR is 0.15 (obtained for a cutoff point of 0.86), and the largest standard deviation for FPR is 0.13 (obtained for a cutoff point of 0.06) (see Figure 6).

The values of the different quality measures (TPR, FPR ...) change only slightly when focusing on the original database (and not on the over-sampled data set).

Table 2
Average TPR in the Validation Set When the Constraint on FPR is Varying Between 5% and 40%

Constraint on FPR	TPR	sd_{TPR}	$Cutoff_{opt}$	$sd_{Cutoff_{opt}}$
5%	44%	0.04	0.81	0.09
10%	60%	0.09	0.72	0.07
15%	71%	0.08	0.62	0.09
20%	78%	0.07	0.55	0.11
25%	83%	0.08	0.51	0.12
30%	85%	0.07	0.52	0.13
35%	86%	0.08	0.52	0.13
40%	87%	0.08	0.53	0.14

Notes. In each case, 200 NNs are run with a different training and validation set. Average cutoff point value and standard deviation are also given. Optimal cutoff point minimizes FN + FP under the constraint on FPR.

Figure 7 illustrates the results for three years: one referring to a nonextreme drought (2013), another referring to a moderate drought (2000), and a third referring to an extreme drought (1983). Yield anomaly observations are compared to the NN output (in $[0, 1]$) and to the NN output after applying the cutoff point C (in $\{0, 1\}$). The optimal cutoff point for this classifier is $C = 0.60$. The NN output can be seen as a yield losses severity index, showing which regions are the most impacted by the weather conditions and the most subject to yield losses. The closer to 1 the NN output is, the higher the probability of having an extreme yield loss is. Conversely, the closer to 0 the NN output is, the lower the probability of having an extreme yield loss is. Before applying the cutoff point, the map of the NN outputs contains a lot of information. The spatial consistency of the NN output is good. The spatial correlation between yield anomalies and NN output for the two extreme years is quite high (-0.51 for 1983, and -0.59 for 2000). Counties whose NN output is very close to the optimal cutoff point are the ones whose classification “0/1,” is the most ambiguous. The classifier quality depends on the variability of the optimal cutoff point C between the spatial regions and between years. Clearly, for the two years with a drought (1983 and 2000), the cutoff point estimated for the whole data set is too small and results would be better (FPR would be lower, i.e., less nonextreme counties would be predicted as extreme) if the cutoff point was higher (around 0.70). Some years displayed a higher FPR than 15%, even after a fixed constraint of $FPR < 15$ was used for the whole data set. It should be noted that, for this application, it is more important to assign the extreme label to a yield (even if it is wrong) than to miss a real extreme case.

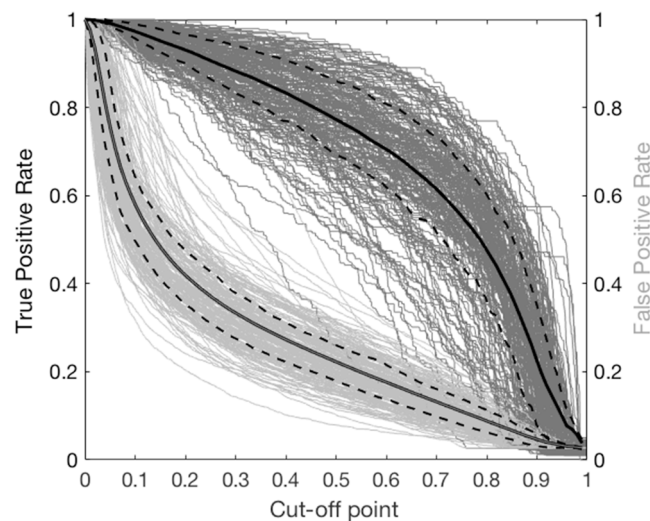


Figure 6. TPR and FPR curves variability in the validation set for 200 different training and validation sets. The black bold curves represent the average value of FPR (or TPR) against the cutoff point value. Black dotted lines represent the second and the eighth decile of the two colored curves.

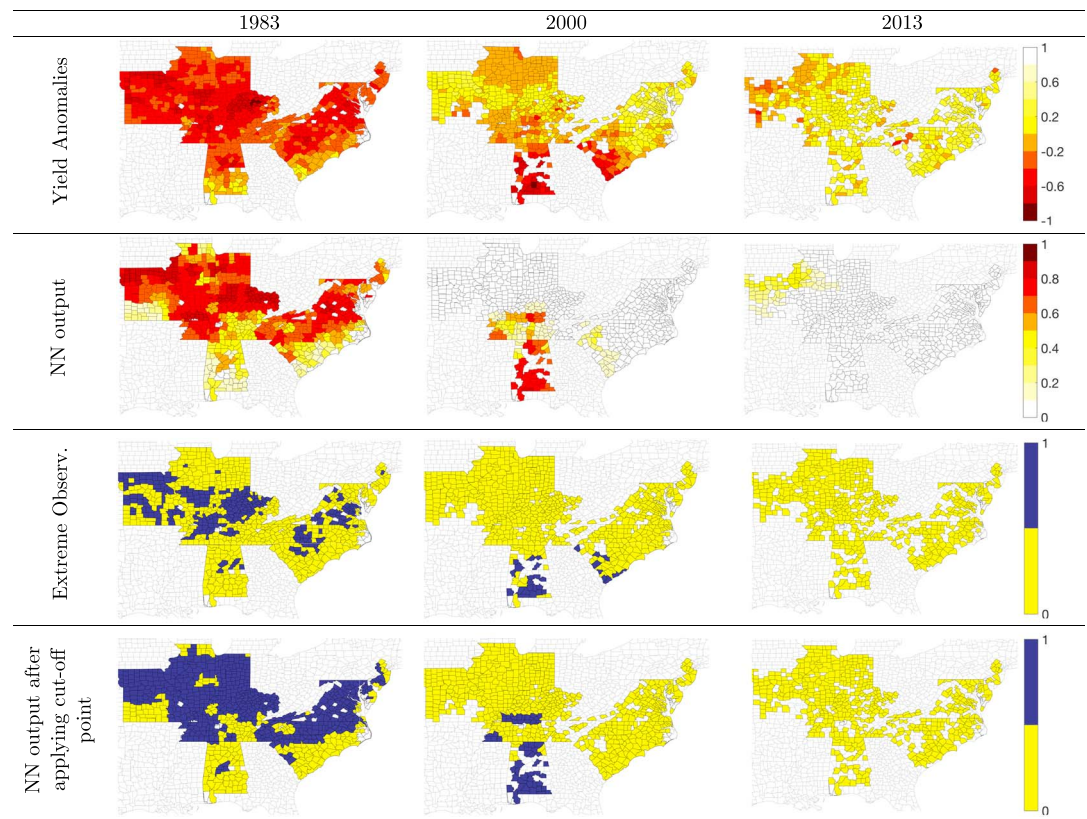


Figure 7. From top to bottom, eastern U.S. maps of corn yield anomalies for years 1983, 2000, and 2013; neural network (NN) output for the same years; observation for the same years; classification decision for this NN.

In short, the NN used as a classifier predicts well years with a low or high number of extremes. The spatial localization of the prediction is good but could be improved: counties close to extreme counties are often predicted as extreme, whereas they are not. This is normal because they have the same climate, but this same climate does not have the same effects. There is probably a lack of site-specific information. The same meteorological conditions will not always lead to extreme yield losses in two different federal states (different agricultural policies, different financial resources, ...): the cutoff point could be optimized for each federal state.

The behaviors of time series of the yield anomalies and NN outputs are very similar: their temporal correlations are represented in Figure 8. Note that in the first five figures, the opposite of the NN output is represented and not the NN output itself. The opposite variations are thus much better represented: these time series are clearly anticorrelated. The map of the United States in Figure 8 gathers the temporal correlation between yield anomalies and the NN outputs for all counties. In most of the considered regions (60%), the correlation is less than -0.6 , illustrating the quality of the NN output as a predictor of the yield variability, even if it was designed for classification (extreme/nonextreme) and not yield prediction.

The NN procedure may go through further posttraining analysis in order to estimate the model response to input variations. For example, the NN behavior can be looked on a regular grid basis, using two predominant input variables, such as SPEI_{July} and T_{July} (Figure 9): positive anomalies of July temperature associated with negative anomalies of July SPEI leads to extreme classification. Similarly, positive anomalies of SPEI in July associated with high positive anomalies of July temperature leads to the classification extreme. In Figure 9 (left), the cutoff point value $C = 0.60$ is underlined by a black curve. All cases below this curve are classified by the NN as nonextreme, and all cases above this curve are classified as extreme. Looking at the left/upper part of the figure, the separating line between extreme cases and nonextreme is quite linear (even if a linear regression or a binary decision tree would not give similar results for these cases). The upper right corner is not very reliable as very few samples are available for this domain (see Figure 9, right). It can be seen in

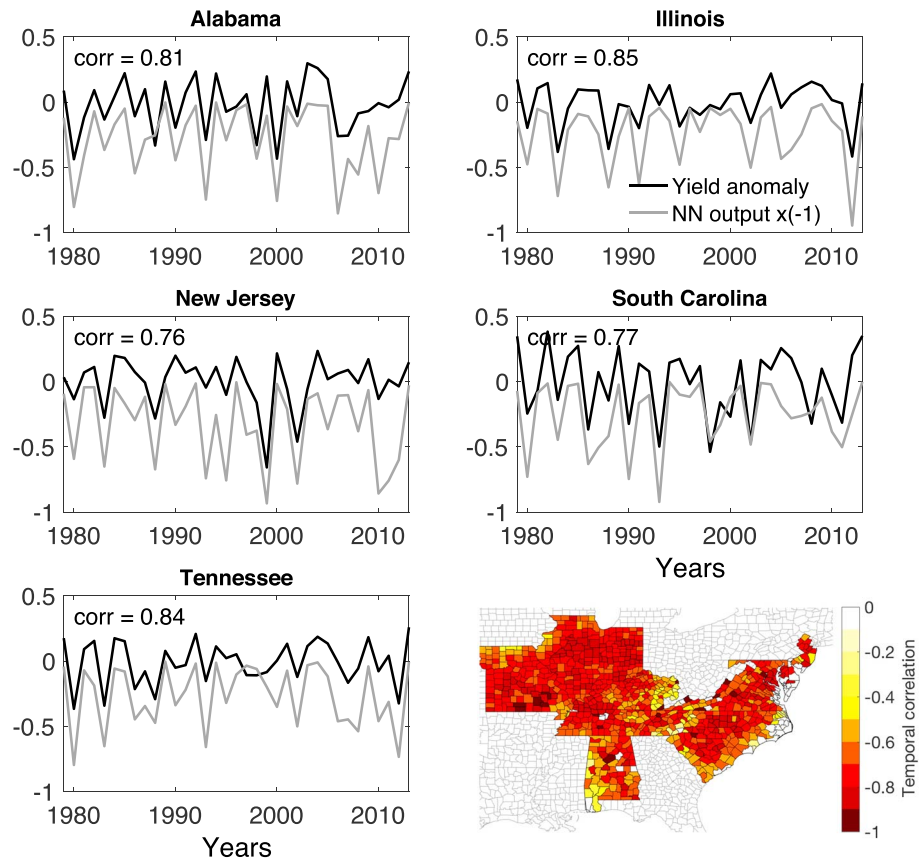


Figure 8. Time series of yield anomalies and neural network (NN) output $\times (-1)$ (the opposite of the NN output) for five counties in several distant states (chosen in Figure 1). (bottom right) Temporal correlation between the time series of corn yield anomaly and the time series of the NN output, at the county scale. The time series of the NN output and the yield anomaly are negatively correlated, because they are varying in opposite ways.

Figure 9 (right), that more samples with high T_{july} and low $\text{SPEI}_{\text{july}}$ would be extremely valuable to improve the calibration of our extreme cases classification.

The spatial feature of Figure 9 (left) shows for a same value of $\text{SPEI}_{\text{july}}$ (horizontal axis) that the probability of extremes is not the same for all values of T_{july} (vertical axis); and conversely, for a specific value of T_{july} , the probability of extremes might not be the same for all values of $\text{SPEI}_{\text{july}}$. This means that there are interactions between these two predictors to estimate the probability of extremes, and the NN is able to catch this behavior. The NN separators are clearly multivariate.

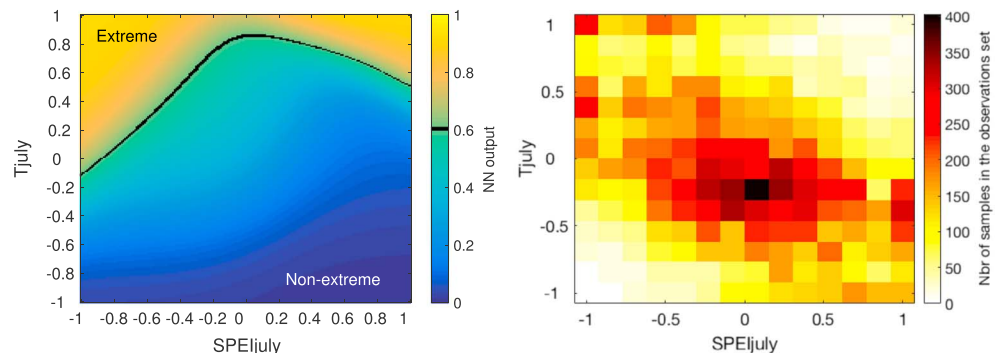


Figure 9. Neural network (NN) classifier (on the two most important inputs, T_{july} and $\text{SPEI}_{\text{july}}$) behavior on a regular grid of these two inputs. The optimal cutoff point C to forecast extreme or nonextreme situations is 0.60 and is represented by a black discriminant line. The right figure represents the number of $\{T_{\text{july}}, \text{SPEI}_{\text{july}}\}$ samples in the observation set.

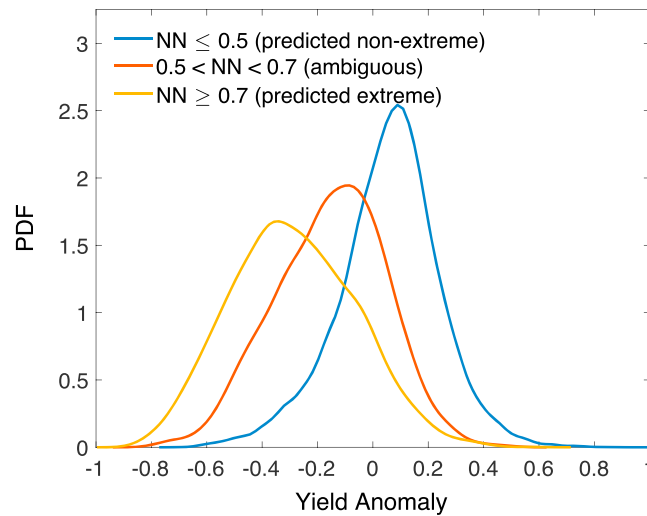


Figure 10. Probability distribution function (PDF) of the yield anomalies for samples that are classified as (1) *extreme* with high probability (yellow curve), (2) as *nonextreme* with high probability (blue curve), and (3) as *ambiguous* (red curve). NN = neural network.

According to this posttraining analysis, cases with $T_{\text{july}} \ll 0$ are always classified as nonextreme (for all SPEI_{july} values), which could show the limits of the classification. It can be surprising to see that if $T_{\text{july}} \ll 0$, corn reacts in the same ways if $\text{SPEI}_{\text{july}} \ll 0$ or $\text{SPEI}_{\text{july}} \gg 0$. However, Cutforth et al. (1986) showed that for corn root growth during emergence, sensitivity to water content decreases with decreasing temperature. When temperatures are much lower than average, the root growth rate evolves very little, whatever the relative available water.

Figure 10 shows the NN response for all yield anomalies of the database. The yellow curve illustrates the distribution of yield anomalies that are considered as extreme with high degree of certainty ($\text{NN}_{\text{output}} \geq 0.7$). Blue curve illustrates the distribution of yield anomalies that are considered as nonextreme with high degree of certainty ($\text{NN}_{\text{output}} \leq 0.5$). Finally, the red curve illustrates the distribution of yield anomalies that are considered as extreme or as nonextreme with little certainty ($0.5 < \text{NN}_{\text{output}} < 0.7$: the NN output is close to the cutoff point $C = 0.60$).

Yellow and blue histograms are not intersecting too much, which means that the NN classifier is doing a good job in separating extreme from nonextreme cases. The average value of yellow distribution (when $\text{NN}_{\text{output}} \geq 0.7$) is close to -0.4 , whereas the average value of blue distribution (when $\text{NN}_{\text{output}} \leq 0.5$) is close to 0.1 . It is satisfactory to obtain an average value of the red distribution (when $0.5 < \text{NN}_{\text{output}} < 0.7$), close to -0.1 : the three averages of these distributions are well separated. In that sense, the NN output performs well as a yield losses severity index.

The statistical model described here appears to give quantitative information on yield. The value of the NN output informs us about two things. On the one hand, it informs us about the precision and assurance of the NN as to its decision: outputs much lower than 0.5 are clearly nonextreme, and those much stronger than 0.7 are clearly extreme. On the other hand, the output of the NN informs us about the probability to face an extreme yield loss: if the output of the NN is close to zero, with high probability, the final yield will be correct or better than the others years. Figure 10 shows that almost none of the extreme yield losses are misclassified by the NN. The model behaves like a yield loss severity index.

5. Discussion and Conclusion

This study introduces a NN classifier to forecast extreme yield loss probability in agriculture and, in particular, in the production of corn over the United States. Without agronomic assumptions, in a data-driven approach, our statistical classification model achieves good accuracy. The NN classifier identifies well the years with a lot, a moderate, or a small number of extreme yield losses and provides an estimate of the extreme probability that can be used as a good yield loss severity index. The model is general and can be used as it is, for any state in the Eastern United States. The weather-to-yield relationship is complex, and incomplete (other missing factors

impact the yield, too). By construction, the NNs are well suited to take into account the interactions among the inputs. This is an important feature; the NN is able to find and exploit the simultaneous combination of high heat and low moisture that is devastating for the crop yields. Furthermore, an analysis is given in the supporting information in order to determine whether, and to what extent, the inherent choices taken in our methodology impact our results. This allows assessing the robustness of our methodology. It was shown that the parameters used in the model (such as the thresholds T used to define the extremes or the constraint on the FPR), do not have a significant impact on the results: according to this sensitivity analysis, we are fairly confident about the result's robustness.

Our conclusions are consistent with agronomic observations on the critical stages of corn plant lifetime (Dardy & Lauer, 2006; Ritchie et al., 1993). The NN behavior makes sense in terms of crop growth since July is the flowering period of corn. A high temperature in July affects the photosynthesis that becomes less effective. Corn yield may also be reduced due to high air temperatures (35 °C and higher) during pollination. Corn is highly sensitive to moisture in July and August, because it is the period of grain formation. During moisture stress, especially at low relative humidity, high temperatures can desiccate silks and damage or kill pollen. However, pollination will not be affected by high temperatures if there is enough moisture in the soil because pollen shed usually occurs during morning hours. The moisture stress during this stage may cause up to 50% yield reduction (Wiatrak, 2010).

The proposed approach differs from standard crop yield modeling approaches especially for extreme yield modeling, from the following points. It does not require specific a priori expert knowledge, and our conclusions are drawn from data. We focus on observed maize yield on the most weather sensitive states (Mathieu & Aires, 2018) covering three contrasting climate zones (i.e., hot humid, mixed-humid and cold, Baechler et al., 2010), from 1979 to 2013. We first analyze weather predictors independently and then in a multivariable classification model. Model prediction performances are evaluated by a cross-validation process. Our methodology is very general and can easily be applied to other crops or other countries.

It is interesting to list the differences with the Ben-Ari et al. (2016) study. Our main contributions are as follows: (1) we developed a multivariate and hierarchical predictor selection where Ben-Ari performs a sensitivity analysis independently for each predictor; (2) we use a multivariate and nonlinear classifier (i.e., NN) instead of a univariate threshold technique, that is an important point, because some predictors interact to act on the corn yield (e.g., heat and water); (3) our study is a new application over the whole Eastern United States; and (4) we performed a comprehensive spatial analysis of the weather sensitivity. Some of these innovations were actually mentioned in the *Room for improvement* section in Ben-Ari et al. (2016).

It can be surprising that the main predictors for the detection of negative extreme yields are simple predictors such as monthly-mean temperature and SPEI. This was mentioned too in Ben-Ari et al. (2016) where it was noted "It is puzzling that simple indicators [...] perform as well as (and sometimes even outperform) composite indicators and crop model simulation for most of the crop-country combinations". Kotlowski (2007) shows that the most important weather attributes for corn are those related to summer months. In particular, his analysis confirms that the crop growth strongly depend on the amount of water in July. The lack of months of early growth can be explained by the fact that the April/May predictors provide less information on the future development of the plant, and are less precise as to the possible evolution and capacity of the plant to recover from difficult climatic conditions. Kotlowski (2007) shows too, that aggregation of monthly weather conditions into seasons decreased the information gain. Using higher temporal resolution (daily values for instance) would probably improve the results. However, Ben-Ari et al. (2016) seem to show that there is a gain of information when considering data on static periods of plants. Combining variables does not seem to be relevant either, and the NN is more likely to do it alone. If there is too much freedom in the inputs of the model but not enough data to calibrate the model, then it is detrimental for the quality of the estimator. It is necessary to find a complexity (model and inputs information) that is compatible with the number of data available to calibrate all this. It is therefore a form of regularization. If we had more data in the historical record, which could describe a wide range of weather situations, with really variable yield outcomes, we could use more complex things in our model. But this is not the case.

What is surprising too, as mentioned by Ben-Ari et al. (2016), is that even the agronomic models are not better or even less than those simple models. Either this means that these models are not sufficiently complete (variables which influence the yields are not taken into account), or it suggests that there is also a problem of model calibration with historical data.

The goal of our statistical modeling is not to estimate a yield but instead to define a yield severity index that would indicate extreme yield losses. Besides, it is extremely difficult to predict precisely in July or August the extreme losses. The causes of extreme yield losses are numerous and can occur just before harvest. Besides, the small proportion of extreme cases in the database is not sufficient to accurately capture yield values for various weather patterns. Providing a probability of extreme yield loss represents less information than the estimation of the actual yield loss. However, this probability estimation requires less input information than what would be required for the yield estimation.

The classification of extreme yield loss is satisfactory but not perfect. This is to be expected: the forecasting of extreme cases is difficult, in particular, because of the rarity of samples available to calibrate the statistical models. An important range of moderate yield loss anomalies (between normal and extreme years) can be classified as ambiguous: the NN output is close to the cutoff point value C .

Our model can be used in several ways. The information provided by the classifier is valuable for many people: farmers could avoid a one time shortfall at the end of the year, by anticipating the signature of an insurance contract. Insurance companies could hedge their risks against large and nonexpected claims for compensation. However, insurance offers are still limited, as in Europe (Bielza Diaz-Caneja et al., 2009). Nongovernmental organizations could anticipate a possible lack of food on a specific region. Food manufacturers could avoid running out of raw material stock, anticipating their supply from another region (Lizumi et al., 2013). Food shops could adapt the food supply too, by searching for another manufacturer. Since the NN forecasts well the years with potentially high extremes yield losses from simple weather information, it can be used in climate simulations to investigate the impact of climate change in agriculture. The model provides also a yield loss severity index that can be used for obtaining early warnings.

There are numerous ways to improve the accuracy of our extreme classifier, and we mention here two of them. (1) The number of nonextreme yields that are classified as extreme is higher than expected: the constraint on the false positive rate ($FPR < 15\%$) at the U.S. scale is not totally satisfactory. A partial solution to this issue would be to specialize the classifier to the local conditions, by adapting, for instance, the cutoff point value C for each state (assuming that enough extreme cases are available in the historical record to calibrate a model for each state or district). (2) The classification could also be improved by adding other predictors. However, it is difficult to find good predictors impacting crop yield, available for more than 30 years. For instance, some studies use vegetation indices (Panda et al., 2010; Tan et al., 2013; Xijie, 2013) such as the Normalized Difference Vegetation Index or the Fraction of Photo-synthetically Active Radiation. Potentially, we could also use non weather-related information, such as soil properties or topography.

Acknowledgments

This work was supported by the PSL Research University. The authors are grateful to the National Agricultural Statistics Service of the United States Department of Agriculture (NASS) for the quality of their survey and studies used in this work. We thank the European Center for Medium-Range Weather Forecasts (ECMWF) for providing meteorological data. SPEI database used in this study can be downloaded at Begueria et al. (2014). Finally, we deeply thank two reviewers and the Editor who handled this manuscript and for their useful comments.

References

- Anaman, K. (2002). Assessing the economic and social impacts of extreme events on agriculture and the use of meteorological information to reduce adverse impacts, *Agrometeorology related to extreme events*. Geneva: World Meteorological Organization.
- Baechler, M., Williamson, J., Gilbride, T., Cole, P., Hefty, M., & Love, P. (2010). Guide to determining climate regions by county, building America best practices series (Vol. 7.1, Tech. Rep.). US Department of Energy. Retrieved from http://apps1.eere.energy.gov/buildings/publications/pdfs/building_america/ba_climateguide_7_1.pdf
- Bannayan, M., & Sanjani, S. (2011). Weather conditions associated with irrigated crops in an arid and semi arid environment. *Agricultural and Forest Meteorology*, 151(12), 1589–1598.
- Begueria, S., Vicente-Serrano, S., Reig, F., & Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, 34, 3001–3023. Retrieved from <http://sac.csic.es/spei/database.html>
- Bellocchi, G., Villalobos, F., Donatelli, M., Christensen, O., Rojas, O., Confalonieri, R., et al. (2014). Extending existing models to capture vegetation response to extreme weather events: The MODEXTREME project. In D. P. Ames (Ed.), *International environmental modelling and software society, 7th International Congress on Environmental Modelling and Software*. San Diego, CA.
- Ben-Ari, T., Adrian, J., Klein, T., Calanca, P., Van der Velde, M., & Makowski, D. (2016). Identifying indicators for extreme wheat and maize yield losses. *Agricultural and Forest Meteorology*, 220, 130–140.
- Bielza Diaz-Caneja, M., Conte, C., Catenaro, R., Gallego Pinilla, F., Dittmann, C., & Stroblmaier, J. (2009). Agricultural insurance schemes, executive summary.
- Bishop, C. M. (1996). *Neural networks for pattern recognition*. Oxford: Clarendon Press.
- Bommier, E. (2014). Peaks-over-threshold modelling of environmental data. Retrieved from <https://www.google.fr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKewiFwaSPvpDWAhWccRQKHQoBCA4QFgg7MAI&url=https>
- Boyer, J., Byrne, P., Cassman, K., Cooper, M., Delmer, D., Greene, T., et al. (2013). The U.S. drought of 2012 in perspective: A call to action. *Global Food Security*, 2(3), 139–143.
- Brown, C., & Davis, H. (2006). Receiver operating characteristic curves and related decision measures: A tutorial. *Chemometrics and Intelligent Laboratory Systems*, 80, 24–38.
- Brown, M., & Funk, C. (2008). Food security under climate change. *Science*, 319(5863), 580–581.
- Cerda, R., Avelino, J., Gary, C., Tixier, P., Lechevallier, E., & Allinne, C. (2017). Primary and secondary yield losses caused by pests and diseases: Assessment and modeling in coffee. *PLoS ONE*, 12(1), e0169133.

- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., & Allard, V. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437, 529–533.
- Cutforth, H., Shaykewich, C., & Cho, C. (1986). Effect of soil water and temperature on corn (zea mays L.) root growth during emergence. *Canadian Journal of Soil Science*, 66(1), 51–58.
- Danforth, C., & Yorke, J. (2006). Making forecasts for chaotic physical processes. *Physical Review Letters*, 96(14), 144102.
- Dardy, H., & Lauer, J. (2006). Plant physiology: Critical stages in the life of a corn plant. Retrieved from corn.agronomy.wisc.edu/Management/pdfs/CriticalStages.pdf
- De Wit, A., Baruth, B., Boogaard, H., van Diepen, C., van Kraalingen, D., Micale, F., et al. (2010). Using ERA-Interim for regional crop yield forecasting in Europe. *Climate Research*, 44, 41–53.
- Debaje, S. (2017). Estimated crop yield losses due to surface ozone exposure and economic damage in India. *Environmental Science and Pollution Research International*, 21(12), 7329–7338.
- Edalat, M., Ghadiri, H., Hamzehzarghani, H., & Kazemeini, S. (2011). Prediction of corn yield loss due to different redroot pigweed density and irrigation level using empirical models. *Australian Journal of Crop Science*, 5, 187–196.
- Fawcett, T. (2004). ROC graphs: Notes and practical considerations for researchers. *Pattern Recognition Letters*, 27, 882–891.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874.
- Fink, A., Brücher, T., Krüger, A., Leckebusch, G., Pinto, J., & Ulbrich, U. (2004). The 2003 European summer heatwaves and drought—Synoptic diagnosis and impacts. *Weather*, 59, 209–216.
- Gobin, A. (2010). Modelling climate impacts on crop yields in Belgium. *Climate Research*, 44, 55–68.
- Gouache, D., Bouchon, A., Jouanneau, E., & Le Bris, X. (2015). Agrometeorological analysis and prediction of wheat yield at the departmental level in France. *Agricultural and Forest Meteorology*, 209, 1–10.
- Kebede, H., Fisher, D., & Young, L. (2012). Determination of moisture deficit and heat stress tolerance in corn using physiological measurements and a low-cost microcontroller-based monitoring system. *Journal of Agronomy and Crop Science*, 198, 118–129.
- Kotlowski, W. (2007). Qualitative models of climate variations impact on crop yields (Tech. Rep.) Laxenburg, Austria: IIASA Interim Report - IIASA. IR-07-034.
- Lal, L., Delgado, J., Gulliford, J., Nielsen, D., Rice, C., & Van Pelt, R. (2012). Adapting agriculture to drought and extreme events. *Journal of Soil and Water Conservation*, 67(6), 162A–166A.
- Landau, S., Mitchell, R., Barnett, V., Colls, J., Craigon, J., & Payne, R. (2000). A parsimonious, multiple-regression model of wheat yield response to environment. *Agricultural and Forest Meteorology*, 101, 151–166.
- Le Rest, K. (2013). Méthodes statistiques pour la modélisation des facteurs influençant la distribution et l'abondance de populations: Application aux rapaces diurnes nichant en France, Biologie de l'environnement, des populations, écologie, Université de Poitiers, Poitiers.
- Lepage, M., Bourgeois, G., & Bélanger, G. (2012). Canada: Indices agrométéorologiques pour l'aide à la décision dans un contexte de climat variable et en évolution., Centre de référence en agriculture et agroalimentaire du Québec (CRAAQ), Québec, QC.
- Lizumi, T., Sakuma, H., Yokozawa, M., Luo, J., Challinor, A., Brown, M., et al. (2013). Prediction of seasonal climate-induced variations in global food production. *Nature Climate Change*, 3, 904–908.
- Macours, K., & Swinnen, J. (2000). Causes of output decline in economic transition: The case of central and eastern European agriculture. *The Journal of Agricultural Science*, 28, 172–206.
- Mathieu, J., & Aires, F. (2016). Statistical weather impact models: an application of neural network and mixed-effects for corn production over the United States. *Journal of applied Meteorology and Climatology*, 55(11), 2509–2527.
- Mathieu, J., & Aires, F. (2018). Impact of agro-climatic indices to improve crop yield forecasting. *Agricultural and Forest Meteorology*, 15, 253–254.
- McCool, P., Musselman, R., Teso, R., & Oshima, R. (1986). Determining crop yield losses from air pollutants. *California Agriculture*, 40, 9–10.
- More, A. (2016). Survey of resampling techniques for improving classification performance in unbalanced datasets. *ArXiv e-prints*.
- Motha, R. (2011). Chapter 30: The impact of extreme weather events on agriculture in the United States.
- Natural Resources Defense Council (2013). Soil matters: How the federal crop insurance program should be reformed (Tech. Rep.). Natural Resources Defense Council. Retrieved from <https://www.nrdc.org/media/2013/130827>
- Oerke, E. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144, 31–43.
- Olesen, J., Trnka, M., Kersebaum, K., Skjelvåg, A., Seguin, B., Peltonen-Sainio, P., et al. (2011). Impacts and adaptation of European crop production systems to climate change. *European Journal of Agronomy*, 34, 96–112.
- Palmer, W. (1965). Meteorological drought (Research Paper No. 45). Washington, DC: US Weather Bureau.
- Panda, S., Ames, D., & Panigrahi, S. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. *Remote Sensing*, 2(10), 673–696.
- Peltonen-Sainio, P., Jauhainen, L., Trnka, M., Olesen, J. E., Calanca, P., Eckersten, H., et al. (2010). Coincidence of variation in yield and climate in Europe. *Agriculture, Ecosystems and Environment*, 139(4), 483–489.
- Powell, J., & Reinhard, S. (2016). Measuring the effects of extreme weather events on yields. *Weather and Climate Extremes*, 12, 69–79. <https://doi.org/10.1016/j.wace.2016.02.003>
- Prakash, A., Rao, J., Mukherjee, A. K., Jeyaveeran, B., Pokhare, S., Adak, T., et al. (2015). Climate change impact on crop pests.
- Ritchie, S., Hanway, J., & Benson, G. (1993). How a corn plant develops, Iowa State University of Science and Technology Cooperative Extension Service, Ames. Retrieved from www.publications.iowa.gov/18027
- Schlenker, W., & Roberts, M. (2009). Nonlinear temperature effects indicate severe damages to US crop yields under climate change. *Proceedings of the National Academy of Sciences of the United States of America*, 106(37), 15,594–15,598.
- Sheffield, J., Goteti, G., & Wood, E. (2006). Development of a 50-yr high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19, 3088–3111. Retrieved from <http://hydrology.princeton.edu/data.pdsi.php>
- Sheffield, J., Wood, E., & Roderick, M. (2012). Little change in global drought over the past 60 years. *Nature*, 491, 435–438.
- Sivakumar, M., Motha, R., & Das, H. (2005). *Natural disasters and extreme events in agriculture: Impacts and mitigation*. Berlin Heidelberg: Springer-Verlag.
- Slingo, J., & Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Science*, 369(1956), 4751–4767.
- Tan, C., Samanta, A., Jin, X., Tong, L., Ma, C., Guo, W., et al. (2013). Using hyperspectral vegetation indices to estimate the fraction of photosynthetically active radiation absorbed by corn canopies. *International Journal of Remote Sensing*, 34(24), 8789–8802.
- Triantafyllou, A., & Sarris, A. (2011). Extreme volatility in agricultural commodity markets and implications for food security.
- Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. DaCosta, Fiorino, M., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society*, 131, 2961–3012.

- van der Velde, M., Wriedt, G., & Bouraoui, F. (2010). Estimating irrigation use and effects on maize yield during the 2003 heatwave in France. *Agriculture, Ecosystems and Environment*, 135(1-2), 90–97.
- Vergara-Diaz, O., Kefauver, S., Elazab, A., Nieto-Taladriz, M., & Araus, J. (2015). Grain yield losses in yellow-rusted durum wheat estimated using digital and conventional parameters under field conditions. *The Crop Journal*, 3(3), 200–210.
- Waha, K., Coumou, D., & Müller, C. (2013). Extreme weather events and agriculture.
- Wiatrak, P. (2010). Environmental conditions affecting corn growth, Clemson University. Retrieved from http://www.clemson.edu/extension/rowcrops/corn/guide/environmental_conditions.html
- Xijie, L. (2013). Remote sensing, normalized difference vegetation index (NDVI), and crop yield forecasting, (PhD thesis). University of Illinois at Urbana-Champaign.
- Zweig, M., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39, 561–77.