The response of maize, sorghum, and soybean yield to growing-phase climate revealed with machine learning

A L Hoffman, A R Kemanian and C E Forest

1 Department of Meteorology and Atmospheric Science, The Pennsylvania State University, PA, United States of America
2 Department of Plant Science, The Pennsylvania State University, PA, United States of America
3 Earth and Environmental Systems Institute, The Pennsylvania State University, PA, United States of America

E-mail: kxa15@psu.edu

Keywords: random forest, vapor pressure deficit, climate, temperature stress, agriculture, technological adaptation, crop yield

Abstract

Accurate representation of crop responses to climate is critically important to understand impacts of climate change and variability in food systems. We use Random Forest (RF), a diagnostic machine learning tool, to explore the dependence of yield on climate and technology for maize, sorghum and soybean in the US plains. We analyze the period from 1980 to 2016 and use a panel of county yields and climate variables for the crop-specific developmental phases: establishment, critical window (yield potential definition) and grain filling. The RF models accounted for between 71% to 86% of the yield variance. Technology, evaluated through the time variable, accounted for approximately 20% of the yield variance and indicates that yields have steadily increased. Responses to climate confirm prior findings revealing threshold-like responses to high temperature (yield decrease sharply when maximum temperature exceed 29 °C and 30 °C for maize and soybean), and reveal a higher temperature tolerance for sorghum, whose yield decreases gradually as maximum temperature exceeds 32.5 °C. We found that sorghum and soybean responded positively to increases in cool minimum temperatures. Maize yield exhibited a unique and negative response to low atmospheric humidity during the critical phase that encompasses flowering, as well as a strong sensitivity to extreme temperature exposure. Using maize as a benchmark, we estimate that if warming continues unabated through the first half of the 21st century, the best climatic conditions for rainfed maize and soybean production may shift from Iowa and Illinois to Minnesota and the Dakotas with possible modulation by soil productivity.

1. Introduction

Regional climate strongly influences the geographical distribution and yield of grain crops, and therefore, the food system. Our understanding of climate variable effects on grain yield has advanced recently through analyses that use historical climate and yield datasets (Schlenker and Roberts 2009, Lobell et al. 2014). These analyses take advantage of the non-controlled experiments conducted by producers as they grow crops in different locations and years for which records are aggregated at scales from fields to regions or even countries. When paired with the climate during the growing season and by using appropriate statistical techniques, these datasets can provide robust responses between crop yields and climate.

Schlenker and Roberts (2009) provided a novel way of analyzing the crop yield response to temperature using a county-level panel of yield and climate variable for maize, cotton, and soybean east of the 100th meridian in the United States. Among the predictors were the number of days exposed to discrete temperature ranges in 1 °C increments. Their analysis revealed that maize and soybean yield have a broad optimum temperature range that falls off abruptly when temperature increases above 29 °C and 30 °C, respectively, with no evidence of technological adaptation to temperature stress. This analysis highlighted the power of combining climate and...
yield records, despite the potential contamination of the climate signal by diseases, pests, soils and other factors.

Machine learning models, and specifically Random Forest (RF), are well suited for crop yield modeling and have been used in bioinformatics (Cutler et al. 2007), ecology (Rehfildt et al. 2012), and agronomy (Saha et al. 2017). Recently, Hoffman et al. (2017) applied RF as a statistical crop model, isolating technology from climate signals in Sub-Saharan countries. The immediate implication is that if one were to use a data-rich region over a large climate domain as used in Schlenker and Roberts (2009), it would be possible to uncover previously unknown relationships between climate variables and crop yield. Furthermore, representing the crop cycle with fidelity, rather than using fixed calendar days from planting to maturity, can help extract the distinct responses across crops' phenological stages.

The impact of climate variables on crop yield can be summarized as follows. Precipitation relates to crop yield by affecting the water balance: when soil water limits the ability of crops to match the atmospheric demand for evaporation, stomata close reducing water loss but also photosynthesis. The yield response is curvilinear and mediated by soil water storage; excess precipitation can reduce yield (Kremer et al. 2008, Sadras et al. 2011, Anderson et al. 2015). Temperature operates through several mechanisms. First, temperature controls the development rate (Sofield et al. 1977); in general, the higher the temperature the faster the development rate and the lower the solar radiation or water capture. Second, high temperature causes heat stress affecting both grain number and grain size (Broekelhurst 1977, Quattar et al. 1987, Miralles et al. 1998, Commuri and Jones 2001). Third, high temperature in C3 plants favor photorespiration and reduce net CO2 assimilation (Zelitch 1982). Finally, extremely low or high temperatures damage both vegetative and reproductive plant structures. Collectively, these processes express in large data panels as threshold-like responses where crop yield decreases sharply above a certain temperature (Schlenker and Roberts 2009, Lobell et al. 2014). Other factors like atmospheric dryness quantified through the vapor pressure deficit (VPD) relate to maize yield (Lobell et al. 2014), beyond the expected reduction in water use efficiency as VPD increase (Tanner 1981, Kemanian et al. 2005), hinting at yet unaccounted for processes affecting yield.

In this paper, we apply RF as a diagnostic crop model using high resolution data in the central region of the United States. Our specific objectives are to: (1) quantify the magnitude of association for precipitation (PRCP), maximum and minimum temperature (TMAX and TMIN, respectively), extreme degree days (EDD), solar radiation (SRAD), and vapor pressure deficit (VPD) in determining the yield of the summer crops (maize, sorghum, and soybean) and (2) estimate phase-specific responses to these variables for each crop. Specifically, we analyze the relationships between climate and grain yield of maize, sorghum, and soybean from 1980 to 2016 in 18 states in the central region of the United States (figure 1) where the majority of these crops are produced.

## 2. Data and methods

### 2.1. Yield and climate data

We used county-level yield data for each state from the United States Department of Agriculture National Agriculture Statistics Service (USDA NASS 2012) (A.2.1). We used this data in conjunction with weather data at 4 km resolution from 1980 to 2016 from the MetData high-resolution gridded surface meteorological product (Abatzoglou 2011). We considered six fundamental climate variables as potential predictors (table 1) and included year as a metric to capture technology improvement with time (please see Section A.1 for more information on variable selection). For each climate variable in table 1, we computed four predictors based on phenological phases defined as establishment, critical window, and grain filling, and averages for the entire cycle from planting to maturity.

### 2.2. Methods to compute growing season phases

#### 2.2.1. Planting date

For each year, we estimated the planting date for each county based on county-level temperatures. Other work has used latitude and elevation to estimate planting dates (Yang et al. 2017), but using

| Variable | Units | Definition |
|----------|-------|------------|
| TMAX | °C | daily maximum temperature |
| TMIN | °C | daily minimum temperature |
| SRAD | W/m² | surface downwelling shortwave flux |
| PRCP | mm | accumulated precipitation |
| VPD | hPa | vapor pressure deficit |
| EDD | °C d | extreme degree days (calculated via cap method) |
temperatures allowed us to simulate farmer adaptation to cold or warm years. We estimate that planting occurs once the 21-day moving average rises to a crop-specific threshold temperature (table 2). Planting temperatures for maize, sorghum, and soybean were 10, 15, and 12 °C, respectively. While there is some latitudinal variation in these planting thresholds, a single threshold for each crop is chosen for simplicity and we note that our results still lie within the range of observed planting dates and that of other work (Bondeau et al. 2007, Waha et al. 2012, Yang et al. 2017, National Agricultural Statistics Service 2012). We assume planting occurs on a single day, though it realistically takes approximately 3 weeks to plant 80-90% of the target area.

Due to sorghum’s and soybean’s low tolerance to cold temperatures, planting happened once the temperature planting threshold is reached after the last spring frost occurred. If threshold temperatures for planting are never met, the crop is not planted.

2.2.2. Crop developmental phases
We considered the entire growing season, but also partitioned the growing season for each crop into three non-overlapping phases: establishment, critical window, and grain filling. This approach should provide a more consistent analytical basis than calendar-based approaches, except that it does not account for planting and harvest time windows or differences in phenology among genotypes, both of which are arguments in favor of a wider calendar window.

The establishment phase is meant to capture conditions during germination, emergence, and early vegetative stages. The critical window aims to capture the reproductive phases that determine yield potential, and the grain filling phase captures conditions during grain growth. There is unavoidable overlap between consecutive phases (Figures A1-A3). These phases represent the phenology well for maize and sorghum, but for soybean the staggered flowering and pod filling phases make the discretization more ambiguous (please see Section A.3.1 for more information on growth phases). For soybean, we consider the establishment phase to span from planting to the start of flowering (R1), the critical window to contain the second through fourth reproductive stages (R2-R4), and grain filling to capture the final stages of reproduction (R5-R8) (Setiyono et al. 2007, Berglund and McWilliams 2015, Fehr and Caviness 1977).
Table 2. Phase-specific thermal times, base temperature, and planting thresholds for maize, sorghum, and soybean. Asterisk denotes that a frost restriction was used. The unit °C d represents a degree day.

| Phase abbrev. | Maize | Sorghum | Soybean |
|---------------|-------|---------|---------|
| Establishment (°C d) | ES | 600 | 600 | 500 |
| Critical window (°C d) | CW | 600 | 400 | 500 |
| Grain filling (°C d) | GF | 850 | 600 | 1000 |
| Growing season (°C d) | GS | 2050 | 1600 | 2000 |
| \( T_{\text{base}} \) (°C) | — | 8 | 8 | 5 |
| \( T_{\text{plant}} \) (°C) | — | 10 | 15* | 12* |

Figure 2. Fractional explained variance for each crop model. The values included in this plot were derived using the raw, unscaled permutation accuracy metric (Strobl et al 2007) and the explained variance metric that emerges from the RF model.

The planting date algorithm was designed to allow for year to year adaptation, and the growing season algorithm was designed to allow for early termination of the grain filling phase under extreme conditions. For all crops, the length of these phases is determined by thermal time and crop-specific base temperature (table 2). Using thermal time to compute the growth phases results in a shorter growing season under warm conditions. In the event that the plant experiences extremely cold late-season conditions we force-ended grain filling prematurely. We consider harvest for all crops to occur once grain filling is completed and did not account for post-maturity yield losses.

2.3. Modeling

Random Forest (RF) is a non-parametric regression technique based on a special ensemble of classification and regression trees (CART) (Breiman et al 1983, Breiman 2001). RF models are constructed by using an ensemble of CARTs wherein each tree is grown using a random subset of the training data and random predictor selection at each node of each tree. The ensemble reduces variance in the final model, while the random predictor selection reduces correlation between trees and keeps bias low. Because each tree is trained on a subset of the data, the remaining data (about one-third of the sample) is left out for testing and computation of internal metrics.

We estimate how county level crop yields in each year depend on a set of fundamental climate variables (table 1). The list contains temperature information in the daily maximum and minimum temperatures (TMAX and TMIN, respectively), and extreme degree days (EDD) calculated via cap method (A.2.2.2). Information regarding moisture and its availability are included as accumulated precipitation (PRCP) and vapor pressure deficit (VPD) (A.2.2.3). We also include surface downwelling shortwave flux (SRAD).

Each crop was modeled independently with each model consisting of six fundamental variables and the year as predictors. We also include all four phases: establishment, critical window, grain filling, and the growing season average. RF can be used to estimate the sensitivity of yields to individual climate predictors and estimate their functional responses. Though many of the predictors in this analysis are closely related, the relation can be explored and made explicit.

We considered the following states: New Mexico, Colorado, Texas, Oklahoma, Kansas, Nebraska, South Dakota, North Dakota, Wisconsin, Minnesota, Iowa,
Michigan, Ohio, Illinois, Indiana, Missouri, Arkansas, and Louisiana for a total of 1631 counties.

3. Results

3.1. Model performance and general responses

Overall, the RF crop models captured 86%, 71%, and 81% of the variance in maize, sorghum, and soybean yields, respectively (figure 2). The variable importance metrics highlight that yield dependence on year is significant and approximately linear (figure A7), explaining yield increases of 90 kg/ha (1.3%) per year for maize, 29.3 kg/ha (0.79%) for sorghum, and 29.4 kg/ha (1.1%) for soybean.

In general, we found that yields were largely driven by environmental conditions during the grain filling period. However, accumulated precipitation and extreme degree days over the entire growing season had a significant effect on yields, particularly for maize (figure 3). A most intriguing response was found for VPD: though VPD during the grain filling period (VPD-GF) drove sorghum and soybean yields, maize yields were driven by VPD during the critical window (VPD-CW). Yield sensitivity to SRAD was low when compared with the other climate variables (figure 3).

3.2. Maize

Maize yields were strongly influenced by VPD during the critical window (figure 5(d)), decreasing by 127 kg/ha (1.9%) per 1 hPa increase in VPD.

More than any other crop in this study, extreme degree days accumulated throughout the growing season (EDD-GS) related to maize yield (figure 6(a)). In general, if EDD-GS exceeded 55 °C d, yields dropped below average. Optimal TMIN-GF and TMAX-GF ranges for maize growth emerged from figures 4(a) and (d). A wide plateau of suitable TMIN-GF ranged from 8°C – 19°C and for TMAX-GF from 21°C – 29°C (table 3). Once TMAX-GF exceeded 29 °C, yields rapidly decreased by 82 kg/ha (1.2%) for every additional °C. Increases in TMIN consistently increased maize yields, so long as TMIN-GF did not exceed 16 °C (figure 4(d)).

Maize yields were highest when approximately 630 mm of precipitation accumulated over the growing season (figure 5(a)). At low levels of precipitation, increasing precipitation by 100 mm steadily and linearly increased yields by 632 kg/ha (4.8%). We consider optimal growing conditions for maize to exist when PRCP-GS exceeds 400 mm.

3.3. Sorghum

The model for sorghum yields had the lowest explanatory power of all three crops with an R² of 0.71 (figure 2). Following time, the most important yield driver was VPD-GF (figure 4(e)). On average, grain yields decreased by 44 kg/ha (1.2%) for every 1 hPa increase in VPD-GF. At very high VPD-GF (greater than 15 hPa), sensitivity to increasing VPD-GF is reduced and yields only drop by 3.8 kg/ha. This reduced sensitivity during warm and dry conditions is also evident in the response to EDD-GS (figure 6(b)).

In general, yields decrease by about 19 kg/ha for every extra 10 °C d EDD-GS, but once 80 °C d are accumulated, that response drops to less than 1 kg/ha. We note that flattening at the edges of these partial dependence plots with RF should be interpreted as data sparsity (A.3.2).
Figure 4. Partial dependence plot for phase-specific maximum daily temperature (TMAX) and minimum daily temperature (TMIN) for each growth phase, including the growing season. Colors represent the growth phase: establishment (blue), critical window (green), grain filling (red), growing season (black). Responses from the 2.5–97.5th percentiles are plotted.

Table 3. Optimal grain filling temperature ranges (derived from partial dependence information plotted in figure 4).

|       | TMAX (°C) |       | TMIN (°C) |
|-------|-----------|-------|-----------|
|       | min       | max   | min       | max   |
| Maize | 20.7      | 29.8  | 9.2       | 19.3  |
| Sorghum | 25.3      | 32.5  | 14.4      | 19.3  |
| Soybean | 21.3      | 30.4  | 12.4      | 19.1  |

Optimal TMIN-GF ranges from 14.5 °C to 19 °C, while optimal TMAX-GF ranges from 25 °C to 30 °C (figures 4(b) and (e)). Yields were likely to be above average as long as TMIN-GF ranged from 12.5 °C to 20 °C and TMAX-GF remained under 32.5 °C. Sorghum yields were highest when TMAX-GF was near 29 °C, and declined rapidly above 32.5 °C. Yet, sorghum was sensitive to cooler minimum temperatures, and yields benefit from increased TMIN throughout the growing season, as long as TMIN-GF does not exceed 19 °C (figure 4(e)).

Sorghum yields were highest at approximately 420 mm of PRCP-GS. Below this threshold, sorghum yields steadily increased by about 180 kg/ha per 100 mm accumulated. Above this threshold, yields decreased by about 13 kg/ha with each additional 50 mm. While PRCP-GF is not as important as PRCP-GS, we found that precipitation accumulated during this phase consistently increases yields (figure 5(b)).

3.4. Soybean

Soybean yield was strongly affected by conditions during grain filling simply because that phase occupied most of the growing cycle (figure 3). Yields respond most strongly to TMAX-GF with yields slowly increasing until 30 °C and then falling off sharply at 98 kg/ha (4.2%) with each additional degree (figure 4(c)). These results indicated an optimal range of both TMIN-GF and TMAX-GF between 9-19 °C and 23-29 °C, respectively.

Similar to the other crops, soybean yields are also affected by VPD-GF (figure 5(f)). In general, yields decrease by 31 kg/ha (1.3%) per 1 hPa increase in VPD-GF.

Unlike sorghum and maize, soybean yields responded to EDD-GF and PRCP-GF instead of the growing season totals (figures 5(c) and 6(c)). Yields increase by approximately 71 kg/ha (3%) per 50 mm until 200 mm have accumulated during this phase; once 200 mm accumulate, yield improvement stagnates. Yield response curves to precipitation accumulated over the growing season indicates that yields optimize at 494 mm. Similar to observations for sorghum, too much precipitation during the establishment phase (PRCP-ES > 120 mm) lowers yields.
4. Discussion

While prior work focused on whole season analyses, single crops, or limited climate variables, we found that precipitation, temperature and dryness of the atmosphere affect the yield of all crops, but with distinct and consequential patterns when considering each developmental phase. When variable importance is assembled by growing phase regardless of the climate variable, the grain filling phase and total growing season tend to be more important, except for maize for which the critical window is equally important (figure A4).

In general, increasing EDD and VPD lead to reductions in yield. Maize and soybean are grown in comparable areas and both were sensitive to EDD, but maize was more sensitive to VPD (mostly during flowering), and soybean was more sensitive to both TMAX and TMIN (figure 4, 5, 4). Sorghum is grown in a drier environment and was more sensitive to VPD (mostly during grain filling) but not very responsive to EDD which may reflect its known tolerance to heat.

Unlike Lyon et al (2003) and Lobell et al (2014), we found no interactive effect between VPD and time on yield (not shown). Recent reports seem to agree that VPD, although positively correlated to higher temperature, may affect yield independently of other factors (Lobell et al 2014, Hoffman et al 2017). No prior report separated its effect on different crops and growing phases.

A comparable maize yield response to VPD was found by Lobell et al (2014) for the period 61-90 days after planting (roughly equivalent to the critical window), which our analysis suggest is a robust response across a wider geography (figure 5(d)). Maize is open pollinated and drought stress causes pollination failures as well as embryo abortion, but water stress effects are still exclusively related in the literature and
modeling to the ratio of the water demand and supply (e.g. Messina et al (2019)). Our analysis strongly suggests that there is a direct effect of dryness of the air, and we speculate that it relates to silk elongation rate (Westgate and Boyer (1985)), silk receptivity, or pollen germination. In addition, the negative effect in yield expresses at relatively low VPD and ceases at 12 hPa (maximum daytime VPD of 20 hPa). Neither of these effects are seen in the autogamous sorghum and soybean (figures 5(e) and (f)).

Temperatures during grain filling emerge as an important driver of yields in the central US (Lobell et al 2014, Prasad et al 2015), likely due to a medley of grain abortion, decreased grain size potential, and decreased grain growth. Prasad et al (2015) found that average temperatures ranging from 25°C to 37°C quadratically decreased individual grain weight when imposed at the start of grain filling as a result of decreased floret fertility. We might be overestimating temperature effect during grain filling and underestimating it on the critical phase because we assume planting to occur on a single day of the year, but the overall response with a broad optimum that falls in the extremes stands.

Modeled yield responses of maize and soybean yields to TMAX-GF compare favorably to those found in other work wherein yields increase slowly with increases until a threshold is reached, after which yields are relatively insensitive to precipitation but decrease slightly if too much precipitation is accumulated (Sadras et al 2011, Hsiang et al 2013, Anderson et al 2015, Hoffmann et al 2017, Cooper et al 2020). Since these are county-level yields, it likely reflects minor losses in well-drained soils and sharp losses in wetter soils.

As expected, grain yields are strongly affected by PRCP-GS and exhibit non-linear responses to accumulated precipitation (figures 5(a)–(c)). The modeled yield response curves are similar to those in other works wherein yields show significant linear increases until a threshold is reached, after which yields are relatively insensitive to more precipitation. We found that yields respond to growing season precipitation exhibited clear threshold behaviors (figures 5(a)–(c)); once an adequate amount of precipitation accumulated, yields optimized. In the case of maize, there are other important predictor variables like VPD-CW and EDD-GS, but they do not exhibit clear thresholds like temperature and precipitation do. To illustrate the point, we plot the optimal growing conditions for maize based on the well-defined optimal ranges for temperature during grain filling and threshold for precipitation accumulated over the growing season. We use the temperature ranges defined in table 3 and consider the threshold for precipitation at 400 mm. We plot these conditions for 2012 and 2016 (figure 7).

Crop yields in 2012 were historically low as a result of drought and a very warm growing season, while the year 2016 was average. Optimal conditions for maize growth in 2016 occurred in the Corn Belt, but it is clear why 2012 had historically low yields: very few counties met the conditions for optimal growth. The band of optimal TMAX-GF temperature shifted north into a narrow band north of Iowa, while low precipitation in most of the central US severely limited yield. It seems that current conditions, although variable, overlay the best climate for maize production with some of the most productive soils in the nation (Iowa, Illinois), but this may change.

To understand how the optimal growing region for maize might change in the future, we generated an optimal growing region figure for 2064 in a business-as-usual climate change scenario using the same ranges and thresholds identified in this analysis (figure 8.) We use statistically downscaled
Figure 7. Optimal temperature and precipitation ranges are plotted for maize in 2016 (top) and 2012 (bottom). Counties which lie within the optimum TMAX and TMIN ranges are highlighted in red and blue, respectively. Counties which experienced more than 400 mm of precipitation are colored grey. Counties that experienced optimal growing conditions are therefore darkest (dark purple).

data from future experiments run by the Geophysical Fluid Dynamics Laboratory model (GFDL) for the Fifth Coupled Model Intercomparison Project, or CMIP5 (Abatzoglou and Brown 2012). While there are uncertainties associated with projected climate data for a single year (and scenario), figure 8 should be considered a snapshot of what could be expected. We find that mid-century growing conditions in a business-as-usual scenario may shift optimal growing conditions north and west resulting in better conditions for growth in Minnesota and the Dakotas. We note the remarkable similarity between the optimal growing conditions in 2012 and those in 2064. The exact direction and extent of these shifting optimal conditions depend on accurate estimates of precipitation which are more uncertain than temperature. It is also important to note that any shifts in optimal growing conditions will
be modulated by soil productivity, as they are today (Williams et al 2016). Nonetheless, figure 8 indicates that the growing conditions of 2012 are likely to occur more often in the future in the RCP8.5 warming scenario.

While environmental conditions drive crop yields, this work also highlights time (year) as an important variable in the analysis, despite limitations in the NASS database that need to be taken into account (A.3.4). Year cannot be directly compared to individual climate variables because we have partitioned each of the climate variables into phases. More equitable comparisons might be to compare year against the aggregated importance of a climate variable for all phases, or phase for all climate variables (figure A4). Nonetheless, time is still an important predictor variable in this analysis as it is representative of technological advances, and likely a minor to moderate response to CO₂. As a standalone variable, time explained 22% of the yield variation for maize and soybean, but only 10% for sorghum. As yield potentials are reached more often (Grassini et al 2011) the environment becomes increasingly dominant. Sorghum seems to have room for further improvement and could become a resilient crop option if the marketing opportunities develop, because increasing temperatures and precipitation variability may make it a preferable choice due to its tolerance to high temperature and water stress.

5. Conclusions

Using local climate data, we trained random forest models for maize, soybean, and sorghum and were able to highlight novel, phase-specific yield responses to climate variables, and confirm prior analyses over a wide geographic range. In order to use phase specific climate information, we developed a reproducible and scalable method to estimate planting date and identify phenological phases for each crop on a county-level. Maize exhibited a unique strong response of yield to atmospheric humidity during the critical phase encompassing from before to after flowering, as well as a strong sensitivity to exposure to extreme temperatures. Soybean has a strong response to both maximum and minimum temperatures. All crops had threshold-like responses to high temperature, as reported earlier by Schlenker and Roberts (2009), although we document a comparatively greater tolerance to high temperature for sorghum (32.5°C vs 29°C–30°C for maize and soybean). Using maize as an example, warm and dry years like 2012 shrink the area of optimum growing conditions and shift it northward toward higher latitudes and westward toward higher elevations (figure 7). Climate conditions similar to those in 2012 are representative of what could be expected in the period of the 2050s-2070s in a business-as-usual warming scenario (figure 8). If warming continues unabated, during the decade of the 2060’s, we expect the best climatic conditions for rainfed maize and soybean production.
may shift from Iowa and Illinois to Minnesota and the Dakotas, a shift that will be modulated by soil productivity.

Acknowledgments

This material is based upon work that is supported by the Network for Sustainable Climate Risk Management (SCRiM) under NSF cooperative agreement GEO-1240507, the USDA National Institute of Food and Agriculture under Grant #2014-68002-21768, and Hatch Appropriations under Project #PEN04710 and Accession #1020049. We would like to thank the two anonymous reviewers for their insights and suggestions that led to substantial improvements in the manuscript. We declare no conflicts of interest for this research.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A. Supplementary information

A.1. Modeling choices and variable selection

A.1.1. Random Forest

Predictive performance of RF only increases when unique information is added by predictors, so correlated variables do not affect the performance of the model. However, the relative importance of correlated predictors is split among correlated variables.

To determine which variable and growth phase had the largest effect on yields, we utilized the permutation accuracy importance measure (Strobl et al 2007). This variable importance measure is computed by measuring the change in model error as a result of permuting a single predictor; if the variable is important, the permutation error is large.

A.1.2. Fundamental vs composite variables

For this research, we originally defined a set of composite and fundamental climate variables that are relevant for crop yields prediction (table A1). The work presented in this manuscript only uses the fundamental variables referenced in table 1 because we found that the inclusion of the composite variables did not improve explanatory power of the models.

The fundamental variables can be used to derive more complex climate variables (e.g. maximum temperature and vapor pressure are used to compute minimum relative humidity). A powerful feature of RF is its ability to compute and rank variables by their importance to predicting the response variable. These complex, or composite variables are a better representation of the environmental conditions, so naturally they arise as more important to a regression. However, whether these variables actually provide new information through variable interactions is unknown. An objective of this work was to test if composite variables are necessary to understand the response of crops to climate, or if they simply repeat information.

We concluded that composite variables are not necessary. In an experiment with all four phases, all 15 climate variables from table 1, as well as year and two variables that measured cold temperature stress during grain filling, the explanatory power of the model marginally increased (table A2).

A.1.3. Latitude

To determine whether our results depend on location, we computed average latitude within a county and included it as a predictor. We found no changes in any of the temperature thresholds. However, when latitude was added as a predictor, the steepness of the slope of maize yields once TMAX-GF > 29°C decreased, which is a very similar response identified in Schlenker and Roberts (2009). We did not include latitude as a predictor in the final model because it did not improve the explanatory power of the fundamental variable model. We note that it did not significantly affect the shape of the variable response functions, and it was tightly correlated with temperature.

A.1.4. Solar radiation

Until now, we have not assessed the impact of SRAD on yields because the other variables are more important (figure 3). Based on theory, we know that solar radiation is an important fundamental variable, but the results of the model are not showing this. For this reason, we tested two separate permutations of the SRAD variable: (1) solar radiation adjusted for optimal temperature conditions (SRADTMP) and (2) solar radiation adjusted for optimal temperature and atmospheric moisture conditions (SRADVPD). However, this was not a directly comparable test because we effectively created a composite variable in which more than one environmental variable was represented. For each of these composite solar radiation variables, the importance increased.

These results do not validate those of Tollenaar et al (2017) that suggested that solar brightening has caused maize yields to increase from 1983-2013. We suspect that this effect has been absorbed into the time (year) variable because solar brightening has been consistently increasing since 1983 (figure 1 in Tollenaar et al (2017)). Despite possible issues with the variability in the SRAD variable, we observe an increase in SRAD-GF from 1980–2016 in line with Tollenaar et al (2017), but we only observe half (650 MJ m⁻²) of what they report (not shown).
Table A1. Potential climate predictors used in this study. Fundamental variables used in the final models are denoted with an asterisk. Variables not found in MetData, but computed from MetData are denoted with a "■".

| Fundamental variables               | Composite variables                          |
|-------------------------------------|----------------------------------------------|
| TMAX\(°C\) daily maximum temperature | RMIN (%) daily minimum relative humidity     |
| TMIN\(°C\) daily maximum temperature | RMAX (%) daily minimum relative humidity     |
| EDD\(°C \text{d}\) extreme degree days (calculated via cap method) | SPH (kg of water vapor/kg of air) daily mean specific humidity |
| PRCP\(\text{mm}\) accumulated precipitation | PDSI (unitless) Palmer drought severity index |
| VPD\(\text{hPa}\) vapor pressure deficit | FDIFF (unitless) fraction of diffuse radiation (Bristow et al 1985) |
| SRAD\((\text{W m}^{-2})\) surface downwelling shortwave flux | SRADTMP\(\text{x}\) adjusted solar radiation for optimal temperature conditions (Olson et al 2012) |
| WINDSPEED\((\text{m s}^{-1})\) daily mean wind speed | SRADVPPD\(\text{x}\) adjusted solar radiation for optimal temperature and vapor conditions (Olson et al 2012) |
| TAVG\(\text{°C}\) daily average temperature |                                             |

Table A2. Comparison of explanatory power (R²) between crop models that only include fundamental variables and those that include all variables from table 1. Models with all predictors also include two variables measuring cold temperature strain during grain filling, while both models include year as a predictor.

|                      | sorghum | maize | soybean |
|----------------------|---------|-------|---------|
| Fundamental variable model | 0.713   | 0.857 | 0.811   |
| All variable model    | 0.738   | 0.869 | 0.826   |

A.2. Data preparation

A.2.1. Yield data—filtering

Models for maize, sorghum, and soybean have approximately 30 000, 18 000, and 27 000 yield records, respectively.

Yield data from the USDA is available from census and survey data, but because the census does not report county-level yields, we used survey data in this work. Survey data is partitioned into various sectors and further into groups. The yield data used in this study come from the field crop group in the crop sector. We extracted the irrigated, rainfed, and unspecified yield data for sorghum, maize, and soybean. We attempted to preserve as much yield data as possible while limiting contamination from irrigated yields. If rainfed yields exist for a given county and year, that yield is included. However, rainfed yields are inconsistently reported and often not reported at all, so we filled in missing data in this record with unspecified yields. Since most irrigated area is west of the 100th meridian, instead of eliminating these records altogether (as done in Schlenker and Roberts (2009)), we filtered out potentially irrigated yields using census data from the 2012 Census of Agriculture (USDA NASS 2012). In general, irrigation has increased with time, so using irrigation data from this decade will likely guarantee our estimates are conservative. For each reporting county, we compute the area irrigated as a function of area planted. If that fraction is greater than 25%, we eliminate all yields from those counties. In the case that irrigation is reported but unpublished (denoted by ‘D’ in the record), we assume that the county is irrigated. If no irrigation is reported, we assume the county is primarily rainfed. Therefore, if the fraction irrigated is less than 25%, we filled in the unspecified yields when available. While low yields are permitted, we only consider the records from a county that contains at least one instance of non-zero and non-missing data. All yields were converted from bushels per acre to kg/ha assuming 13%, 15.5%, and 13% moisture for sorghum, maize (grain), and soybean, respectively.

Yield data on the county-level reduces the need to aggregate climate data, which can potentially obscure subtle yield responses in statistical crop models and bias predictions in process-based crop models (Pierce and Running 1995, Baron et al 2005, Tack et al 2015, Hoffmann et al 2016, Hoffman et al 2017).

A.2.2. Climate data

This work utilized several predictor variables not included in MetData (see bold variables in table 1). For each of these the appropriate MetData variables were used to compute the new variable at each individual grid point for each year and each day. These values were subsequently aggregated to the county-level and used to compute the phase-specific variables used in the analysis. Though variables like temperature show small daily variations across a county, we elected to compute these variables at each individual grid points (instead of using the county average) for consistency with the MetData.
Figure A1. Segment plot depicting average growing season for maize. If a state has no data, then maize is either not planted there, or is planted in counties with harvest area irrigation over 25%. Blue represents the establishment phase, green represents the critical window, red represents the grain filling phase, while black represents a two-week drying period (not included in this study).

A.2.2.1. TAVG
Daily average temperature was computed using:

\[
TAVG = \frac{(TMAX + TMIN)}{2} \quad (A1)
\]

Many crop models use a skewed value so that it is more representative of daytime conditions, which are arguably more important to plants.

A.2.2.2. EDD
The ‘cap method’ (Gilmore and Rogers, 1958) is used to compute extreme degree days.

\[
EDD = \sum_{t=1}^{N} DD_{30,t} \ast normFactor, \quad (A2)
\]
Figure A2. Same as figure A1 but for sorghum.

\[ DD_{30,t} = \begin{cases} 0 & \text{TAVG}_t < 30^\circ C \\ TAVG_t - 30 & \text{TAVG}_t \geq 30^\circ C \end{cases} \quad \text{(A3)} \]

where

\[ \text{normFactor}_t = \frac{DD_{30,t}}{TMAX_t - TMIN_t} \quad \text{(A4)} \]

Where \( t \) represents each day, while \( N \) represents the number of days in each growth phase. We normalize EDD by the diurnal temperature because EDD is typically computed with hourly data (as in Lobell et al (2013)).

A.2.2.3. VPD

Vapor pressure deficit was computed using the maximum and minimum values for temperature and relative humidity. The standard means were computed for both temperature and relative humidity. Average temperature was used to estimate the saturated vapor pressure using Teten’s equation while average relative humidity was used to estimate vapor pressure using...
the approximation that $RH = e/e_s$. We use the following variation of Teten’s equation:

$$e_s (hPa) = 6.112 \times \exp \left( \frac{17.269 \times T(\degree C)}{T(\degree C) + 237.3} \right) \quad (A5)$$

There are many ways to compute VPD (Abtew and Melesse 2013). Lobell et al (2014) used the difference in saturation vapor pressure computed using TMAX and TMIN. As with average temperature, we could have also used average temperature skewed towards the daytime to compute the saturation vapor pressure.

A.2.2.4. Derived radiation variables: FDIFF, SRADTMP, SRADVPD

These variables were computed because plants do not utilize solar radiation uniformly under all conditions. Under extreme conditions radiation is utilized less efficiently. We use the equations outlined in Bristow et al (1985), Olson et al (2012), and Olson et al (2012) for FDIFF, SRADTMP, and SRADVPD, respectively.
A.2.2.5. Flags and frost events

Though not discussed much in the manuscript, when the 10-day moving average temperature drops below the base temperature for the grain filling phase, the early grain filling termination flag is raised. We also computed the number of late-season frosts during grain filling using the minimum temperature to compute how many times the minimum temperature dropped below 0 °C.

A.3. Potential sources of error

A.3.1. Growth phases

A novel aspect of this work was the estimation of annual planting dates so that we can approximate the yield response to climate during individual growth phases, but it is possible that we introduced additional uncertainty to our model as a result. The individual phases for each state are included in Figures A1–A3. While planting dates are largely dependent on environmental conditions, farmers also use historical knowledge to inform their decision making which is not included in this work (Waha et al. 2012). Similar to the methodology of Waha et al. (2012), it could be worthwhile to estimate the planting date in the warmest states (i.e. Texas and Louisiana) using seasonal precipitation as the limiting factor rather than temperature. Another potential source of this uncertainty is using a single planting threshold temperature for each county.

Based on the results and the physiology of soybean, we suspect that there may be several minor limitations interpreting soybean growth phases. Our results suggest that we are still capturing the establishment, or early-season, signal with the modeled critical window. In other words, the vegetative (establishment) phase for soybean is likely not long enough in our model. Therefore, the phase which we intended to be the critical window is actually contained within the grain filling phase, and the grain filling phase effectively represents all of soybean’s reproductive phases (R1-R8). This is one reason why grain filling might arise as the most informative phase for each variable.
A.3.2. Partial dependence sensitivity to noise
As mentioned before, a prominent feature in many partial dependence plots is the decreased sensitivity of yields to extreme values on either end of the range. This ‘flattening’ is a feature of random forest, and as noise increases, the more conservative predictions will become. At the bounds of our data for each predictor variable, there are fewer data points, and thus more noise. It is also possible that there is a lack of response (i.e. other variables are dominating the response). Though we often plot the innermost 95th percentile of data, the effect is still evident in many cases. These features should not be considered robust.

A.3.3. Scale
We do not parse out field-level crop information or mask out areas of a county that do not grow a given crop. We elected to estimate the county averages in this manner because the available data for crop areas are calibrated to represent the year 2000 (Sacks et al 2010). Though temporally varying crop areas can potentially be estimated using satellite data (Schlenker and Roberts 2009), the use of satellite
products introduces additional uncertainty because these algorithms are based on single-cropped systems.

We assume there are minimal differences from state to state. However, soil composition, the use of applications (fertilizer, fungicide, insecticide, etc), and farm-subsidies can change with state. When including state as a predictor, we did not find any significant differences in the shape or thresholds of the yield responses to individual climate variables (only vertical compression).

A.3.4. Yield quality
Yield data from the USDA NASS is commonly used for crop modeling studies, but few studies acknowledge the limitations of these data, and we address several of these here. We note concerns with inconsistencies in irrigated data and county-level yields. First, we note spatial and temporal inconsistencies in the reporting of irrigated, rainfed, and unspecified yields. Not every county reports yield data each year and therefore each year does not necessarily represent the same spatial coverage. For instance, since the year 2008, only five states have reported maize yields that discriminate irrigated and rainfed yields (personal comm. Chris Hawthorne). Additionally, if an insufficient fraction of survey responses is collected in a given county for rainfed or irrigated yields of any crop, then the responses are aggregated for a larger area called ‘other (combined) counties’ (Johnson 2014, Bock and Kirkendall 2017). These counties were omitted from our analysis because the NASS does not record which counties were aggregated in the data.

We used unspecified yields in this study, but masked out counties that irrigate over 25% of the harvested area for a given crop. This was because there were no consistent patterns in the data that would have allowed us to extrapolate what fraction of the

---

**Figure A7.** Partial dependence plot for year. All responses are plotted and year 1 and 37 correspond to 1980 and 2016, respectively.

**Figure A8.** Similar to figure 8 but for a moderate warming scenario (RCP 4.5) (Abatzoglou and Brown 2012).
unspecified yields were rainfed and which were irrigated. For this reason, studies that use rainfed yields from the USDA NASS have not used it to study large regions (Lobell et al 2013).

Another major issue in the USDA NASS data is that the number of counties for which survey-based yield estimates are available (irrigated, rainfed, or unspecified) has been steadily decreasing with time (Bock and Kirkendall 2017). By 1998, the halfway point in this analysis, 70%, 55%, and 54% of yields had already been reported for sorghum, maize, and soybean, respectively. A major reason for this discrepancy is because yield data from aggregated areas cannot be utilized, and aggregating yields has become more common.

Lastly, we considered all yields from a given county provided that there was one instance of non-zero, non-missing data.

A4. Extended discussion

A4.1. Fractional variable importance in fundamental variable models
This section contains figures generated from a rather unconventional method to structure the information contained in figure 3, but it is a valid method because we use the raw, unscaled permutation metric (Strobl et al 2007). Figure A4 is highlighting the relative importance of the climate in each phase for each crop, while figure A5 is highlighting the importance of each phase for each variable in the maize model. Summing the values for each phase in figure A5, one can create figure A4.

A4.2. Variable importance.
A4.2.1. VPD
Historically, VPD over the growing season in the region of interest has increased by 1.2 hPa, 1.4 hPa, and 1.3 hPa for maize, sorghum, and soybean, respectively from 1980-2016, changes that are much smaller than those reported for regions of China by Zhang et al (2017).

Increasing the frequency of TMAX above 29 °C, 32.5 °C, and 30 °C will reduce yields for maize, sorghum, and soybean, respectively. However, yield loss can be higher for maize if the increase in TMAX is accompanied by an increase in VPD.

A4.2.2. Year
Year was included as a predictor because we acknowledge that technology, including improved fertilizers, fungicides, and development of robust and high-yield varieties, have historically increased yields. The fact that year arose as an important predictor was unsurprising given the results in Grassini et al (2013) and Hoffman et al (2017). Interestingly, year did not arise as one of the most important predictors in Lobell et al (2014), but this might be explained by aggregating the importance of all the climate phases (similar to figure A4).

A4.3. Future projections
We also plotted optimal growing conditions for a moderate warming scenario, RCP4.5 (Figure A8) (Abatzoglou and Brown 2012).

ORCID iDs
A L Hoffman  https://orcid.org/0000-0002-9622-3721
A R Kemanian  https://orcid.org/0000-0002-7862-3527
C E Forest  https://orcid.org/0000-0002-2643-0186

References
Abatzoglou J T 2011 Development of gridded surface meteorological data for ecological applications and modelling Int. J. Climatol. 33 121–31
Abatzoglou J T and Brown T J 2012 A comparison of statistical downscaling methods suited for wildfire applications Int. J. Climatol. 32 772–80
Abtew W and Melesse A Vapor Pressure Calculation Methods Evaporation and evapotranspiration: Measurements and estimations (Dordrecht: Springer) pp 53–62
Anderson C J, Babcock B A, Peng Y, Gassman P W and Campbell T D 2013 Placing bounds on extreme temperature response of maize Environ. Res. Lett. 10 124001
Baron C, Sultan B, Balme M, Sarr B, Traore S, Lebel T, Janicot S and Dingkuhn M 2005 From GCM grid cell to agricultural plot: scale issues affecting modelling of climate impact Phil. Trans. R. Soc. B 360 2095–108
Berglund D and McWilliams D 2015 Soybean growth and management quick guide number A-1174 North Dakota State University Extension Service Fargo, ND (https://ag.ndsu.edu/pubs/plantsci/rowcrops/a1174.pdf)
Bock M E and Kirkendall N J eds 2017 Improving Crop Estimates by Integrating Multiple Data Sources (Washington, DC: Academies Press)
Bondeau A et al 2007 Modelling the role of agriculture for the 20th century global terrestrial carbon balance Glob. Change Biol. 13 679–706
Breiman L 2001 Random Forests Mach. Learn. 45 5–32
Breiman L, Friedman J, Olshen R, Stone C, Steinberg D and Colla P 1983 CART: Classification and Regression Trees (Belmont, CA: Wadsworth)
Bristow K L, Campbell G S and Saxton K E 1985 An equation for separating daily solar irradiation into direct and diffuse components Agric. Forest Meteorol. 35 123–31
Brocklehurst P A 1977 Factors controlling grain weight in wheat Nature 266, 348–9
Commuri P D and Jones R J 2001 High temperatures during endosperm cell division in maize Crop Science 41 1122–30
Cooper M, Tang T, Gho C, Hart T, Hammer G and Messina C 2020 Integrating genetic gain and gap analysis to predict improvements in crop productivity Crop Sci. 60 582–604
Cutler D R, Thomas C, Edwards J, Beard K H, Cutler A, Hess K T, Gibson J and Lawler J J 2007 Random forests for classification in ecology Ecology 88 2783–92
Fehr W R and Caviness C E 1977 ‘Stages of soybean development’ (http://lib.dr.iastate.edu/specialreports/87)
Gilmore E C and Rogers J S 1958 Heat units as a method of measuring maturity in corn Agron. J. 50 611–15
Grassini P, Eskridge K M and Cassman K G 2013 Distinguishing between yield advances and yield plateaus in historical crop production trends Nat. Commun. 4 2918
Grassini P, Thorburn J, Burr C and Cassman K G 2011 High-yield irrigated maize in the Western US Corn Belt: I. On-farm yield, yield potential, and impact of agronomic practices Field Crops Res. 120 142–50

References
Abatzoglou J T 2011 Development of gridded surface meteorological data for ecological applications and modelling Int. J. Climatol. 33 121–31
Abatzoglou J T and Brown T J 2012 A comparison of statistical downscaling methods suited for wildfire applications Int. J. Climatol. 32 772–80
Abtew W and Melesse A Vapor Pressure Calculation Methods Evaporation and evapotranspiration: Measurements and estimations (Dordrecht: Springer) pp 53–62
Anderson C J, Babcock B A, Peng Y, Gassman P W and Campbell T D 2013 Placing bounds on extreme temperature response of maize Environ. Res. Lett. 10 124001
Baron C, Sultan B, Balme M, Sarr B, Traore S, Lebel T, Janicot S and Dingkuhn M 2005 From GCM grid cell to agricultural plot: scale issues affecting modelling of climate impact Phil. Trans. R. Soc. B 360 2095–108
Berglund D and McWilliams D 2015 Soybean growth and management quick guide number A-1174 North Dakota State University Extension Service Fargo, ND (https://ag.ndsu.edu/pubs/plantsci/rowcrops/a1174.pdf)
Bock M E and Kirkendall N J eds 2017 Improving Crop Estimates by Integrating Multiple Data Sources (Washington, DC: Academies Press)
Bon...
Hoffman A L, Kemanian A R and Forest C E 2017 Analysis of climate signals in the crop yield record of sub-Saharan Africa Global Change Biology 24 143–57

Hoffmann H et al 2016 Impact of spatial soil and climate input data aggregation on regional yield simulation PLOS One 11 e015182

Hsiang S, Lobell D, Roberts M and Schlenker W 2013 Climate and crop yields in Australia, Brazil, China, Europe and the United States SSRN (http://doi.org/10.2139/ssrn.2977571)

Johnson D M 2014 An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States Remote Sensing of Environment 141 116–28

Kemanian A R, Stockle C O and Huggins D R 2005 Transpiration-use efficiency of barley Agri. Forest Meteorol. 130 1–11

Kremer C, C O Sockle, Kemanian A R and Howell T 2008 Response of crops to limited water: understanding and modeling water stress effects on plant growth processes American Society of Agronomy, Crop Science Society of America, Soil Science Society of America chapter A Canopy Transpiration and Photosynthesis Model for Evaluating Simple Crop Productivity Models (New York: Wiley) pp. 165–189

Lobell D B, Hammer G L, McLean G, Messina C, Roberts M J and Schlenker W 2013 The critical role of extreme heat for maize production in the United States Nat. Climate Change 3 497–501

Lobell D B, Roberts M J, Schlenker W, Braun N, Little B B, Rejeus R M and Hammer G L 2014 Greater Sensitivity to Drought Accompany Maize Yield Increase in the US Midwest Science 344 516–19

Lyon D J, Hammer G L, McLean G B and Blumenthal J M 2003 Simulation supplements field studies to determine no-till dryland corn population recommendations for semiarid Western Nebraska Agron. J. 95 884–91

Messina C D, Hammer G L, McLean G, Cooper M, van Oosterom E J, Tardieu F, Chapman S C, Doherty A and Gho C 2019 On the dynamic determinants of reproductive failure under drought in maize Silico Plants 1 dio003

Miralles D, Calderini D, Pomer K and D’Ambrogio A 1998 Dwarving genes and cell dimensions in different organs of wheat J. Exp. Botany 49 1119–27

National Agricultural Statistics Service 2012 ‘Field crops usual planting and harvesting dates’ (http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-2010.pdf)

Olson S N, Ritter K, Rooney W, Kemanian A, McCarl B A, Zhang Y, Hall S, Packer D and Mullet J 2012 High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops Biofuels Bioprod. Biorefining 6 640–55

Pierce L L and Running S W 1995 The effects of aggregating sub-grid land surface variation on large-scale estimates of net primary production Landscape Ecol. 10 239–53

Prasad P V V, Boote K J and Jr L H A 2006 Adverse high temperature effects on pollen viability, seed-set, seed yield and harvest index of grain-sorghum [Sorghum bicolor (L.) Moench] are more severe at elevated carbon dioxide due to higher tissue temperatures Agric. Forest Meteorol. 139 237–51

Prasad P V V, Djanaguiraman M, Perumal R and Ciampitti I A 2015 Impact of high temperature stress on floret fertility and individual grain weight of grain sorghum: sensitive stages an thresholds for temperature and duration Front. Plant Sci. 6 820

Prasad P V V, Pisipati S R, Mutava R N and Tiuinstra M R 2008 Sensitivity of grain sorghum to high temperature stress during reproductive development Crop Sci. 48 1911–17

Quattar S, Jones R and Crookston R 1987 Effect of water deficit during grain filling on the pattern of maize kernel growth and development Crop Sci. 27 726–30

Rehfeld G E, Crookston N L, Saenz-Romero C and Campbell E M 2012 North American vegetation model for land-use planning in a changing climate: A solution to large classification problems Ecol. Appl. 22 119–41

Sacks W, Deryng D, Foley J A and Ramankutty N 2010 Crop planting dates: an analysis of global patterns Glob. Ecol. Biogeogr. 19 607–20

Sadas V O, Grassini P and Steduto P 2011 Status of water use efficiency of main crops. SOLAW Background Thematic Report—TR07 Food and Agriculture Organization of the United Nations Rome 16–18 (http://fao.org/fileadmin/templates/solaw/files/thematic_reports/TR_07_web.pdf)

Saha D, Rau B M, Kaye J P, Montes F, Adler P R and Kemanian A R 2017 Landscape control of nitrous oxide emissions during the transition from conservation reserve program to perennial grasses for bioenergy Glob. Change Biol. Bioenergy 9 783–95

Schlenker W and Roberts M J 2009 Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change Proc. Natl Acad. Sci. 106 15594–8

Setiyono T D, Weiss A, Specht J E, Bastidas A M and Cassman K G 2007 Understanding and modeling the effect of temperature and daylength on soybean phenology under high-yield conditions Field Crops Res. 100 257–71

Sofield I, Evans L, Cook M and Wardlaw I 1977 Factors influencing the rate and duration of grain filling in wheat Aust. J. Plant Physiol. 4 785–97

Srobl C, Boulesteix A L, Zieleski A and Hotthorn T 2007 Bias in random forest variable importance measures: Illustrations, sources and a solution BMC Bioinform. 8 1

Tack J, Barkley A and Nalley L L 2015 Effect of warming temperatures on US wheat yields Proc. Natl Acad. Sci. 112 6931–6

Tanner C 1981 Transpiration efficiency of potato Agron. J. 73 59–64

Tollenaar M, Fridge H, Tjapi P, Stackhouse I P R W and Kumudini S 2017 The contribution of solar brightening to the us maize yield trend Nat. Clim. Change 7 275–80

USDA NASS 2012 2012 Census of Agriculture

Waha K, van Bussel L G J, Müller C and Bondeau A 2012 Climate-driven simulation of global crop sowing dates Glob. Ecol. Biogeogr. 21 247–59

Westgate M E and Boyer J S 1985 Osmotic adjustment and the inhibition of leaf, stem and silk growth at low water potentials in maize Plants 164 540–9

Williams A, Hunter M C, Kammerer M, Kane D A, Jordan N R, Mortensen D A, Smith R G, Snapp S and Davis A S 2016 Soil water holding capacity mitigates downside risk and volatility in us rainfed maize: time to invest in soil organic matter? Phil. One 11 e0160974

Yang X, Wilson J T and Wang J 2017 A spatially explicit crop planting initiation and progression model for the conterminous United States Ecol. J. Agron. 90 184–97

Zelitch I 1982 The close relationship between net photosynthesis and crop yield BioScience 32 796–802

Zhang S, Tao F and Zhang Z 2017 Spatial and temporal changes in vapor pressure deficit and their impacts on crop yields in China during 1980–2008 J. Meteorol. Res. 31 800–8