Geophysical Research Letters

Supporting Information for

20th-century regional climate change in the central United States attributed to agricultural intensification

Ross E. Alter[1], Hunter C. Douglas[1], Jonathan M. Winter[2], Elfatih A. B. Eltahir[1]

[1]Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA.

[2]Department of Geography, Dartmouth College, Hanover, New Hampshire, 03755, USA.

Contents of this file

Additional Supporting Information (Files uploaded separately)

Caption for Movie S1

Introduction

The following methodology, figures, and tables provide important details about this study, which determined that agricultural intensification has been a greater forcing of observed regional climate change during the summer in the central United States than greenhouse gas (GHG) emissions over the last century.

The methodology section is split into six sections: The Experimental Design section describes the MIT Regional Climate Model (MRCM) – the regional climate model used in this study – and outlines the procedures used to set up, validate, and run simulations with MRCM. The University of Delaware Observational Data section describes the gridded temperature and precipitation datasets that were used to validate the output from MRCM and also describes the procedure by which the data were analyzed in support of this study. The ISD Humidity Data section describes the observational humidity data that were used for trend analysis over the central United States. The NASS Crop Data section describes the source of the crop data used for this study and how the data were analyzed in support of our observational and modeling experiments. The Statistical Analysis section explains each statistical test that was used to support the quantitative results of this study. Finally, there is a statement on Code Availability at the end of Text S1.

The supporting figures and tables provide supplemental views of the primary data relationships described in the main text and figures.

Text S1. Methodology

Experimental Design

In this study, the MIT Regional Climate Model (MRCM) is used to examine the climatological effects of agricultural intensification within the central United States. MRCM maintains much of the structure of the Regional Climate Model version 3 (RegCM3) (Pal et al., 2007) but includes several improvements, including coupling to the IBIS land surface scheme (Winter et al., 2009), a new bare soil albedo assignment method (Marcella, 2012), new convective cloud and convective rainfall autoconversion schemes (Gianotti and Eltahir, 2014a, 2014b), and modified boundary layer height and boundary layer cloud schemes (Gianotti, 2012). A new irrigation module has also been implemented within the IBIS land surface scheme (Marcella, 2012). Details on the workings of the irrigation module can be found in previous studies that implemented the module over West Africa (Marcella and Eltahir, 2014).

Previous work has confirmed that MRCM can successfully simulate regional climate in the central United States during the summer, especially precipitation (Winter and Eltahir, 2012). Given this validation, we adopted the same version of MRCM in designing experiments to approximate agricultural intensification in the central United States. More detailed model description and general performance of MRCM can be found in previous studies of MRCM simulations over West Africa (Im and Eltahir, 2014; Im, Gianotti, & Eltahir, 2014; Im, Marcella, & Eltahir, 2014; Marcella and Eltahir, 2014) and the central United States (Winter et al., 2009; Winter and Eltahir, 2012).

The original model domain for the MRCM control simulation (CONT) covers most of the eastern two-thirds of the contiguous United States plus part of southern Canada. It is centered at 40.5°N and 91.5°W with 30 km horizontal grid increments (122 grid points zonally and 80 grid points meridionally) and 18 vertical sigma levels from the surface to the 50 hPa level. The size of the domain and location of the boundaries were chosen to minimize interference from the elevated topography of the Rocky Mountains yet also include the bulk of regional weather and climate patterns that affect the central United States. The figures in the paper are shown with 9 model grid cells removed from the boundary in each direction so as to focus on the interior of the model domain that was least affected by boundary conditions yet encompass all of the major cropland areas of the central United States.

Figure S6 presents the model domain and land-use distribution for the MRCM simulations. Both the CONT and experimental (EXP) simulations account for (rain-fed) cropland and irrigated cropland across the model domain. Gridded irrigation data from the year 2005 were assimilated into MRCM from the Historical Irrigation Dataset (Siebert et al., 2015), and a grid cell was designated as "irrigated cropland" if irrigation covered at least 25% of the total grid cell area. At grid cells that are designated as "irrigated cropland", irrigation is triggered when soil moisture in the root zone (top 1 m of the soil) reaches 75% of relative field capacity, i.e., field capacity divided by the porosity of a soil layer. At this point, water is added until the root zone reaches soil saturation, which is assumed to be relative field capacity. This irrigation procedure is repeated as necessary for both CONT and EXP simulations during each model year from July to September. In comparison to irrigation schemes that continuously saturate the root zone, this procedure allows for a more demand-based irrigation system that maintains more realistic levels of soil moisture over irrigated cropland.

Cropland in MRCM is represented as a generic C4 crop rather than individual crop types, based on the parameters and implementation of Agro-IBIS in Kucharik and Brye (2003), and crop phenology was specified to begin in July of each model year. Adding C3 crops to MRCM is an important avenue for future work as it would further refine the results of this study.

To simulate agricultural intensification, we sought to develop a straightforward method that could accurately capture the increase in crop productivity over the past century. Depending on the choice of start year, crop production in the central United States (especially for corn/maize) increased by a factor of ~5-7 over the last 50-100 years, based on Figure 2b (derived from NASS [2016]) and Mueller et al. (2016, Fig. S3). To ensure that we achieved an actual productivity increase in the simulations that was consistent with these observed trends, we multiplied the net photosynthetic rate directly by a factor of 6.

Net primary productivity (NPP) was expected to rise in tandem with photosynthesis, and we identified simulated increases in NPP by a factor of 4-6 over most of the cropped areas due to the enhanced photosynthesis, which coincides with the observed increase in crop yield over the last 50-100 years in the central United States (Fig. 2b). However, the

mean NPP in our simulations with enhanced photosynthesis was about 2000 g C m^{-2} yr^{-1}, which is significantly higher than the estimated mean NPP for corn of 971 g C m^{-2} yr^{-1} for 2007/2008 in Li et al. (2014). Simulated NPP values are based on the default value from IBIS for a generic C4 crop and were not calibrated against observations.

In the absence of observed, long-term changes in regional photosynthesis and evapotranspiration, the observed changes in temperature can provide another means for anchoring our prescribed agricultural forcing to reality. Indeed, the link between agricultural intensification and temperature has already been rigorously established through recent observational analyses (Mueller et al., 2016, 2017). Therefore, to test the feasibility of our photosynthetic approximation from a climate perspective, we compared the observational and simulated results for surface air temperature. We found that adding the results of the MRCM and CMIP5 temperature simulations [Figs. 3 and 4] approximately equals the observed change in surface air temperature over the region of significant rainfall change (ROSC, see University of Delaware Observational Data subsection for definition) in the central United States (-0.35°C) [Figs. 1 and 2]. Thus, combining simulations with both natural and anthropogenic forcings, including our photosynthetic approximation, seem to correctly reproduce the magnitude of observed temperature change in the central United States.

Each of the simulations spans 30 years continuously from January 1982 to December 2011. The initial and boundary conditions used by MRCM are specified according to the ERA-Interim reanalysis (Uppala et al., 2008) – the third generation ECMWF reanalysis project – which has horizontal grid increments of 1.5° x 1.5° at 6-hour intervals. Although additional spin-up time is not added, this should not introduce a significant error because initial soil moisture conditions are specified from long-term offline simulations that bring soil moisture in equilibrium with the climate of the region. Sea surface temperature (SST) is prescribed by the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation (OI) SST dataset (Reynolds et al., 2002) with horizontal grid increments of 1° x 1° and weekly temporal resolution for the entire simulation period. Each simulation utilized time-evolving historical greenhouse gas concentrations from 1982-2011 based on the SRES A2 scenario, as specified by RegCM3 (Pal et al., 2007). To minimize the dependence of the model simulations on initial conditions, each of our five experiments had a different starting date (i.e., 1982.1.1, 1982.1.2, 1982.1.3, etc.) to generate time-lagged ensemble members. The climatological response to agricultural intensification derived from the EXP simulations is based on the arithmetic mean of the five ensemble members.

Validation of the CONT simulation was performed with the University of Delaware gridded observational dataset (Willmott and Matsuura, 2015), which is described below.

University of Delaware Observational Data

Observational analyses were carried out using the Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (V3.01) dataset from the University of Delaware (UDel) (Willmott and Matsuura, 2015). This quality-controlled, gridded

dataset is built from land-based observations at weather stations. It provides monthly values for mean air temperature and total precipitation from 1900 to 2010 at a resolution of 0.5° latitude by 0.5° longitude for land areas across the entire globe.

To determine the time periods for comparison in the observational rainfall analysis, the UDel data within the model domain were first split into 24 subregions of 5° latitude by 5° longitude. Within each subregion, the precipitation data at each grid cell were then split into two 30-year periods (separated by at least 5 years), and the first and last years of each time period were varied. The period of most significant change was defined as the time period between the two 30-year periods for which the greatest number of grid cells in each subregion exhibited statistically significant change ($p=0.05$). This was found to be the years 1950-1969 over most of the subregions in the central United States [not shown]. To add robustness to the analysis, the two periods of comparison were extended to be 40 years long, i.e., 1910-1949 and 1970-2009. Data before 1910 were excluded from the analysis because the UDel dataset draws heavily from the Global Historical Climatology Network (Peterson and Vose, 1997), where station coverage drops off rapidly before 1910: Within the model domain, there were 795 active stations in the year 1910 but only 671 in 1900 (a 16% decrease).

The period 1950-1969 was also a time of substantial development in agricultural production within the Corn Belt of the central United States: Corn production in the region nearly doubled over this time period [Fig. 2]. The time periods straddling this initial period of agricultural development, 1910-1949 and 1970-2009, are referred to as pre-DEV and full-DEV, respectively.

Based on the findings of the above analysis, the region of significant change (ROSC) was defined as a rectangular area within the central United States that encompasses the majority (77.8%) of grid cells exhibiting statistically significant change in precipitation between the pre-DEV and full-DEV time periods, according to the Kolmogorov-Smirnov test (see Statistical Analysis subsection for details). The ROSC extends from 39 to 47°N and from 100 to 82°E. The size and position of the ROSC were chosen to capture as much of the observed rainfall increases as possible while remaining small enough to not overly dilute the impact of the areas with the largest increases in observed rainfall.

We scaled all of the CMIP5 global climate model output to a middle-of-the-road horizontal resolution for the CMIP5 models (ACCESS models, 1.25° x 1.875°), which corresponds to a 6 x 10 grid in the ROSC (Fig. 3). The results from the MRCM simulations and UDel observations were also scaled to this resolution for the histograms in Fig. 3d and 3e, creating 60 points for each dataset. Scaling upward to the coarsest CMIP5 model resolution (3.71° x 3.75°, BCC-CSM1.1 model) would have resulted in only about 15 grid cells within the ROSC, which would have contributed too few samples for a meaningful histogram. Furthermore, regional aspects of the simulations and the observations would have been lost at this coarsest resolution. Finally, bias correction was not conducted for this study.

ISD Humidity Data

A set of station data, the Integrated Surface Database (ISD) (*NCEI*, 2016b), was used to obtain historical observations of humidity. Stations in this database record parameters such as air temperature, dew point temperature, and air pressure. Of the stations collocated with areas exhibiting significant increases in precipitation in the UDel dataset, only those that had data before 1950 and over 80% data coverage were selected. Twelve stations were identified that met these criteria, and none of these stations included data from earlier than 1938.

Specific humidity was calculated from dew point temperature and air pressure using the following formulas (derived from Rogers and Yau [1989]):

$$SH = a\frac{P_w}{P}$$

$$P_w = 6.1094e^{\left(17.625T_d/(243.04+T_d)\right)}$$

Where:
SH = specific humidity
a = ratio of molecular weight of water to that of dry air = 0.622
P_w = vapor pressure of water at a given dew point temperature
P = atmospheric pressure
T_d = dew point temperature

Because very little data was available for the pre-DEV period, using the Kolmogorov-Smirnov test (see Statistical Analysis subsection) for determining statistically significant change was not considered appropriate. Instead, a simple linear least squares regression was applied to the specific humidity data, and the magnitude of the trend was noted.

NASS Crop Data

The U.S. Department of Agriculture's National Agricultural Statistics Service (NASS) maintains rich archives of the nation's historical crop production data, which were used to obtain records of crop production, yield, and harvested acreage for key summer crops (NASS, 2016). Data were extracted for corn and soybeans, as these represent the vast majority of summertime agricultural production in the Corn Belt of the central United States (NASS, 2016). Mean annual crop production for corn and soybeans was calculated for the pre-DEV and full-DEV climate periods for individual counties within the model domain, and these were used to calculate the differences in crop productivity between the two periods [Fig. 1, Fig. S1].

A time series of production, harvested acreage, and yield of corn in the Corn Belt was also produced using state-level data from NASS (NASS, 2016) [Fig. 2]. The Corn Belt, as defined by the National Centers for Environmental Information (NCEI, 2016c),

includes: Colorado, Illinois, Indiana, Iowa, Michigan, Minnesota, Nebraska, Ohio, South Dakota, and Wisconsin. Pennsylvania, while technically a part of the Corn Belt, was excluded from this analysis as it is not located within the central United States.

Statistical Analysis

In calculations using the UDel dataset, statistical significance was determined using a two-sample Kolmogorov-Smirnov (K-S) test. The K-S test is used to determine the likelihood that two samples of data arise from the same distribution function. It is based on the maximum difference between the empirical cumulative distributions of the two samples and identifies shifts in both the means of the data and changes in their variance (Smirnov, 1939; Sheskin, 2007). The K-S test is also nonparametric, i.e., it makes no assumptions about the distributions of the data. Because of these factors, it is an appropriate test for comparing distinct samples of unknown distribution, such as weather data (Sheskin, 2007). The K-S test was chosen over the Student's t-test because meteorological data are not necessarily normally distributed, and the t-test assumes normality. In this study, statistical significance was defined at the 95% level, i.e., $p \leq 0.05$.

For the simulations, we used the chi-square test (χ^2) to determine the consistency of the rainfall increases. The chi-square test analyzes categorical data (e.g., positive vs. negative) to determine whether the distribution of categories is different enough from the expected outcome to reject the null hypothesis. In our case, the null hypothesis is that the distribution of positive and negative rainfall changes in the agricultural intensification simulation is similar to the expectation of 50% positive and 50% negative changes. The dark green (brown) grid cells in the simulation results [Fig. 3] show where agricultural intensification causes rainfall increases (decreases) during at least 62% of the 150 model years, which yields a p-value of 0.036 when using the chi-square test. Therefore, in the grid cells that pass this threshold, the consistency of the rainfall increases due to agricultural intensification is statistically significant at the 5% level.

For the observations, we also utilized a recently developed statistic – the consistency of relative change index (CRCI) (Alter et al., 2015) – that can be used to determine the consistency of the sign of relative change for any temporal variable. For this study, the CRCI is used to determine how frequently rainfall exhibits relative increases during a "full agricultural development" period compared to a "pre-agricultural development" period [not shown]. In general, the CRCI is calculated by counting the number of times that rainfall values in the later period (for example, full-DEV) exceed the mean of the observed rainfall in the earlier period (for example, pre-DEV). In this respect, the mean rainfall of the earlier period (for example, pre-DEV) is used as a threshold for determining how often the later period contains positive or negative rainfall anomalies. The CRCI is then sorted by percentile rank for each grid cell value in relation to the corresponding values associated with all grid cells in the study area. In comparing the pre-DEV and full-DEV time periods using the UDel precipitation dataset, grid cells with CRCI greater than or equal to the 90th percentile are clustered in the areas with the largest

rainfall increases in the central United States [not shown]. This indicates consistent increases in rainfall compared to the vast majority of other grid cells in the domain.

Code Availability

All computer code used to conduct the analyses in this paper is available upon request.

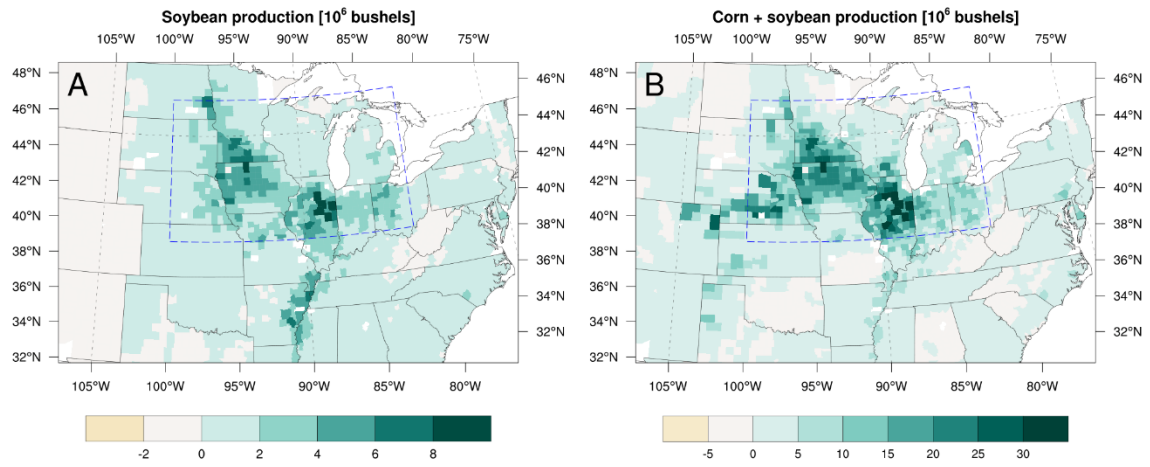

Figure S1. Historical changes in (a) soybean production and (b) the sum of corn and soybean production from the period before agricultural development (pre-DEV, 1910-1949) to the period of full agricultural development (full-DEV, 1970-2009). The blue dotted line encloses the region where a large proportion of grid cells has experienced statistically significant increases in observed rainfall (region of significant change – ROSC) according to the Kolmogorov-Smirnov test (p=0.05).

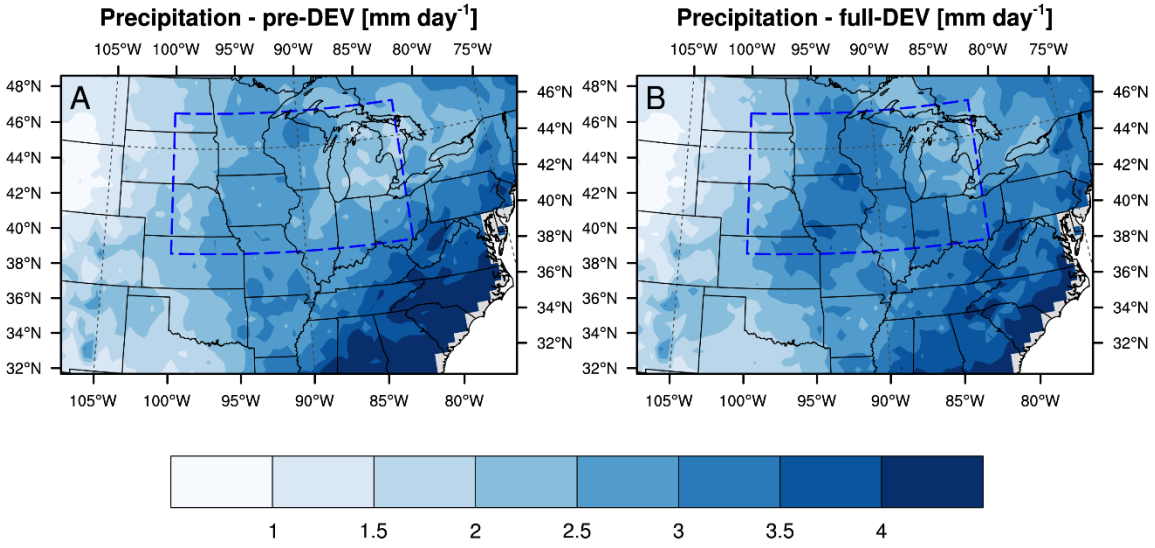

Figure S2. Mean observed July-August precipitation over the central United States for (a) the period before agricultural development (pre-DEV, 1910-1949) and (b) the period of full agricultural development (full-DEV, 1970-2009). The blue box over the region of significant change (ROSC) is the same as in Fig. S1.

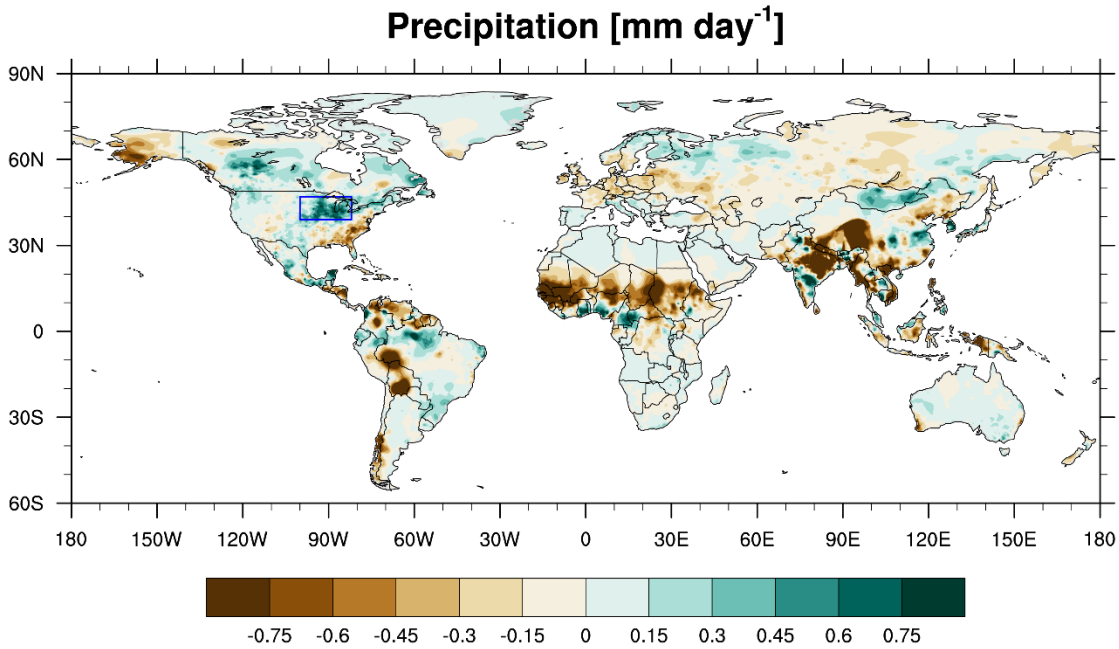

Figure S3. Absolute change in July-August precipitation from pre-DEV (1910-1949) to full-DEV (1970-2009). The blue box over the region of significant change (ROSC) in the central United States is the same as in Fig. S1.

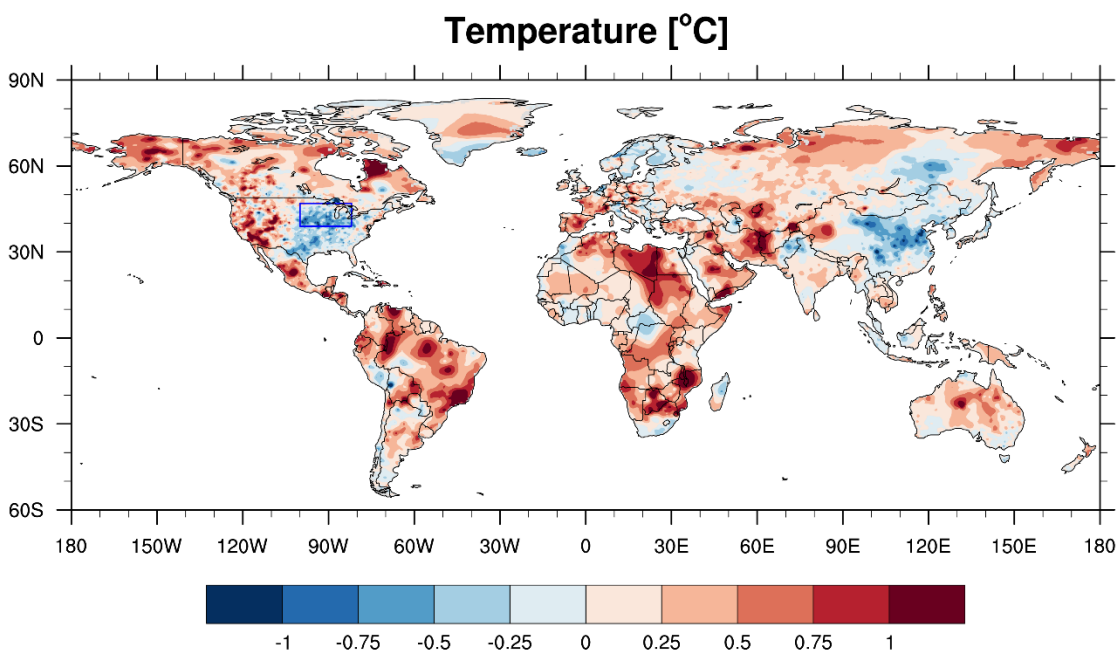

Figure S4. Absolute change in July-August surface air temperature from pre-DEV (1910-1949) to full-DEV (1970-2009). The blue box over the region of significant change (ROSC) in the central United States is the same as in Fig. S1.

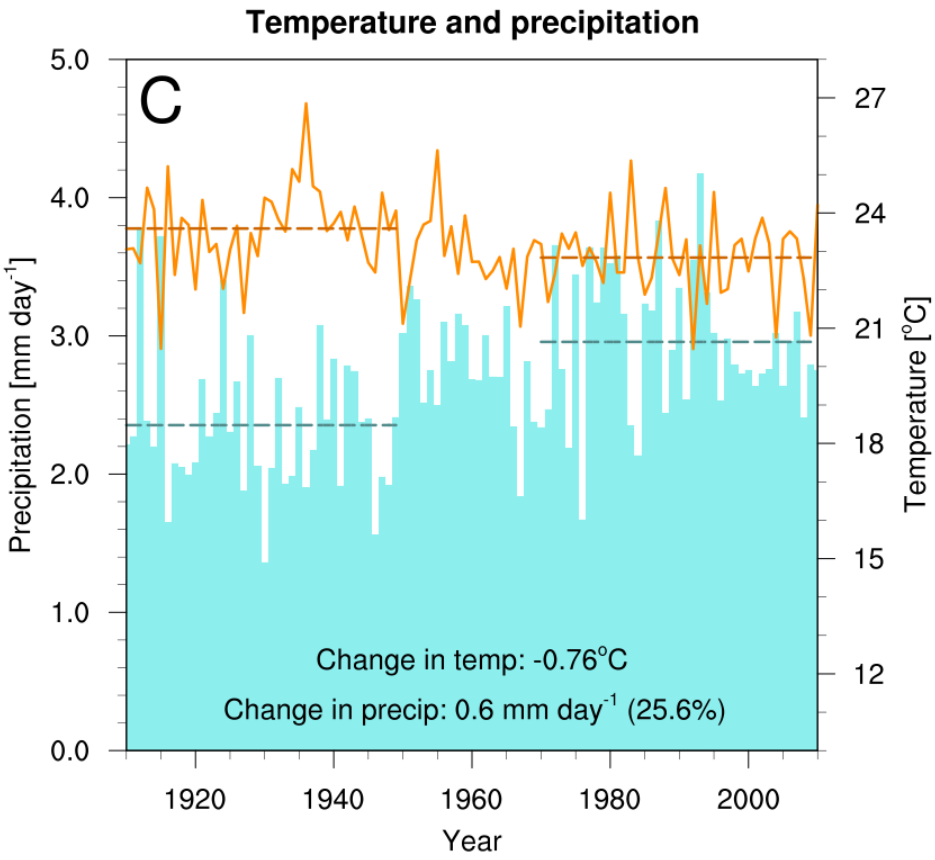

Figure S5. Same as Fig. 2c, except only for points in the central United States that have experienced statistically significant increases (p = 0.05) in precipitation from pre-DEV (1910-1949) to full-DEV (1970-2009).

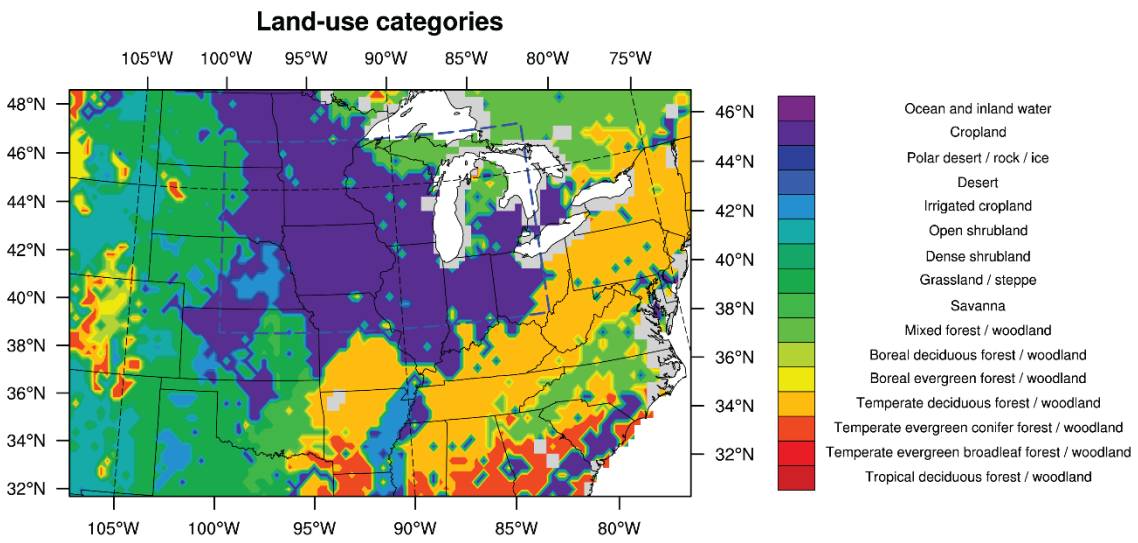

Figure S6. Dominant land-use categories within the model domain of MRCM, the regional climate model used in this study. The blue box over the region of significant change (ROSC) is the same as in Fig. S1.

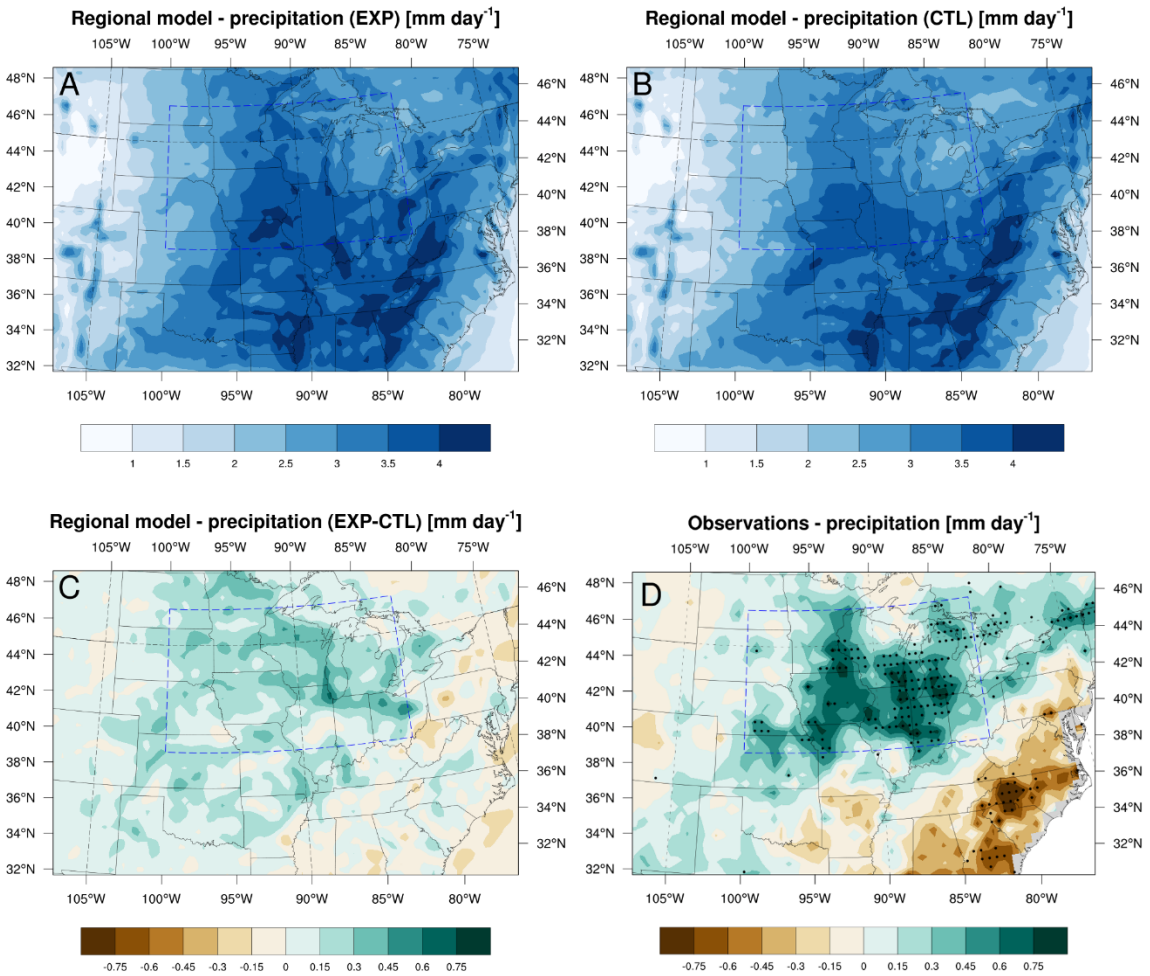

Figure S7. Mean July-August precipitation from the ensemble of regional climate model (MRCM) simulations with (a) agricultural intensification (EXP) and (b) no agricultural intensification (CTL). (c) Absolute change in precipitation when including agricultural intensification in MRCM, and (d) observed absolute change in mean July-August precipitation from the pre-DEV period (1910-1949) to the full-DEV period (1970-2009). The blue box over the region of significant change (ROSC) is the same as in Fig. S1.

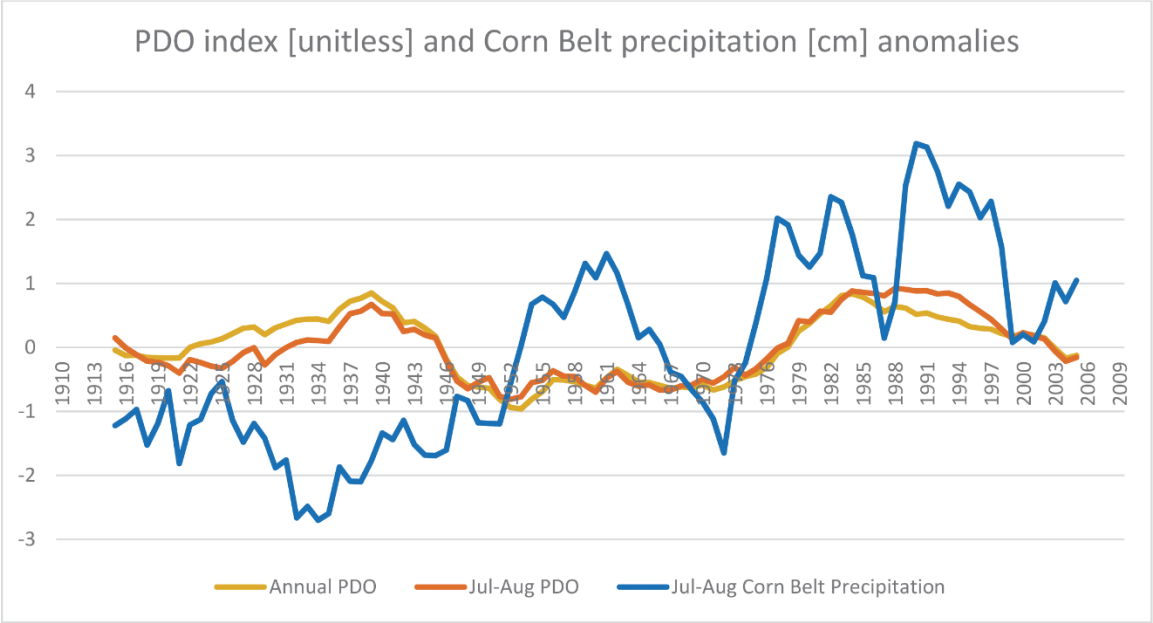

Figure S8. Time series of annual (light orange) and July-August (dark orange) anomalies of the Pacific Decadal Oscillation index (PDO) and July-August precipitation anomalies in the Corn Belt of the central United States (blue). The data are shown as 10-year moving averages.

Model Name	Modeling Center (or Group)	Institute ID	Ensembles used in this study
ACCESS1.0	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia	CSIRO-BOM	r1-r2
ACCESS1.3	“ ”	“ ”	r1-r3
BCC-CSM1.1	Beijing Climate Center, China Meteorological Administration	BCC	r1-r3
BCC-CSM1.1(m)	“ ”	“ ”	r1-r3
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	GCESS	r1
CMCC-CESM	Centro Euro-Mediterraneo per I Cambiamenti Climatici	CMCC	r1 (no surface specific humidity)
CMCC-CM	“ ”	“ ”	r1 (no surface specific humidity)
CMCC-CMS	“ ”	“ ”	r1 (no surface specific humidity)
CNRM-CM5	Centre National de Recherches Météorologiques / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique	CNRM-CERFACS	r1-r3
CNRM-CM5-2	“ ”	“ ”	r1
CSIRO-Mk3.6.0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence	CSIRO-QCCCE	r1-r3
GFDL-CM2.1	NOAA Geophysical Fluid Dynamics Laboratory	NOAA GFDL	r1-r3 (no surface specific humidity)
HadCM3	Met Office Hadley Centre	MOHC	r1-r3
NorESM1-M	Norwegian Climate Centre	NCC	r1-r3
NorESM1-ME	“ ”	“ ”	r1

Table S1. The global climate models from the Coupled Model Intercomparison Project 5 (CMIP5) consortium that were utilized in this study. The model simulations listed above did not account for historical land use changes. Some models did not have an explicit variable for specific humidity at the surface (variable name = “huss”); these models were excluded from the CMIP5 specific humidity analyses in this study. If a model provided more than three ensemble runs, we only utilized the first three runs. Note that the final year of most historical CMIP5 runs is 2005.

	Sign of observed change	Sign of predicted change	
		Agricultural intensification	Greenhouse gas (GHG) emissions
Rainfall	**Positive**[1,2]	**Positive**[3]	**Neutral**[3]
Surface temperature	**Negative**[1,2,3]	**Negative**[3,4]	**Positive**[3,4]
Surface humidity	**Positive**[2]	**Positive**[4]	**Positive**[4]

Table S2. Summary table of the signs of change for each climate variable analyzed in this study. The superscripts denote the figure numbers in the main text associated with each table value. Note that "GHG emissions" includes all non-agricultural forcings accounted for in the CMIP5 historical simulations (i.e., GHG emissions plus natural and other anthropogenic forcings).

Movie S1. Annual corn production in the central United States by county from 1910 to 2010.