Science Advances
AAAS

Supplementary Materials for

An unprecedented fall drought drives Dust Bowl–like losses associated with La Niña events in US wheat production

Lina Zhang *et al.*

Corresponding author: Xiaomao Lin, xlin@ksu.edu

Sci. Adv. **10**, eado6864 (2024)
DOI: 10.1126/sciadv.ado6864

This PDF file includes:

Supplementary Text

Conceptual framework

Crop production (P) is defined as the product of yield (Y) and harvested area (HA). HA can be further expressed as the product of planted area (PA) and harvestable fraction ($\frac{HA}{PA}$). A proportion of PA may be abandoned ($fab = \frac{PA-HA}{PA}$). Final county-level production can be calculated as:

$$P = \frac{HA}{PA} \times PA \times Y = (1 - f_{ab}) \times PA \times Y \quad [S1]$$

Given the marginal relationship between pre-season fall precipitation and planted area (fig. S12), we assume that weather conditions during both wheat pre-season and in-season do not impact planted areas. It seems evident that climate conditions have little bearing on farmers' planting decisions in Kansas. Instead, planting decisions are more likely guided by non-environmental influences (i.e., long-term agricultural strategies, socio-economic factors, and policy incentives). Thus, in this study, both pre-season and in-season climate conditions affect production variability presumably through crop abandonment and yield variations. To estimate the impacts of climate-induced changes in abandonment and yields on total production across the Dust Bowl decade (1931-1940), the latest decade (2013-2022), and year 2023, we decomposed actual production [S1] into components: expected base production (P_{base}), changes in production due to abandonment (ΔP_{fab}),changes in production due to reductions in per-acre-harvested yield (ΔP_Y), and changes in production due to technology (ΔP_{tech}).

We firstly substituted yields (Y) with the sum of expected base yields (Y_{base}) and yield anomalies (ΔY) in equation [S1], and introduced the terms of abandonment in normal years ($\overline{f_{ab}}$), resulting in the following:

$$P = \left(1 - f_{ab} + \overline{f_{ab}} - \overline{f_{ab}}\right) \times PA \times (Y_{base} + \Delta Y + [Y_{lowess} - Y_{base}]) \quad [S2]$$

where Y_{lowess} is expected yields (see Methods). The difference between Y_{lowess} and Y_{base} represents the change of yield driven by technology.

We next broke down equation [S2] into four components as,

$$P = \left(1 - \overline{f_{ab}}\right) \times PA \times Y_{base} + (\overline{f_{ab}} - f_{ab}) \times PA \times (Y_{base} + \Delta Y) + \left(1 - \overline{f_{ab}}\right) \times PA \times \Delta Y + \left(1 - f_{ab}\right) \times PA \times (Y_{lowess} - Y_{base}) \quad [S3]$$

where $1 - \overline{f_{ab}}$ represents harvested area ratio in normal years; and $\overline{f_{ab}} - f_{ab}$ defines changes in abandonment for years of interest relative to the expected abandonment in normal years. We then extracted each of four components from [S3]:

$$P_{base} = \left(1 - \overline{f_{ab}}\right) \times PA \times Y_{base} \quad [S4]$$

$$\Delta P_{fab} = (\overline{f_{ab}} - f_{ab}) \times PA \times (Y_{base} + \Delta Y) \quad [S5]$$

$$\Delta P_Y = \left(1 - \overline{f_{ab}}\right) \times PA \times \Delta Y \quad [S6]$$

$$\Delta P_{tech} = (1 - f_{ab}) \times PA \times (Y_{lowess} - Y_{base}) \quad [S7]$$

We here focused on climate-driven production change for the years of interest; thus, the f_{ab} and ΔY were simulated in our RF models (Eqs. [1] and [2] in Main text). Note that the average abandonment ($\overline{f_{ab}}$) and expected base yields (Y_{base}) in normal years were defined as the 30-year averages, calculated by using climate data from the period 1981-2010. We defined PA as the 30-year average over 1981-2010 when we calculated production changes

due to abandonment and yield reductions. This allowed us to compare production losses from climate-driven changes in both abandonment and yields among years. The base abandonment ($\overline{f_{ab}}$) was calculated as,

$$\overline{f_{ab}} = \frac{1}{30}\sum_{y=1981}^{2010} fab_y \quad [S8]$$

The expected base yield (Y_{base}) was calculated from climate-driven ΔY, simulated using the RF model in the main text, and average expected yield ($\tilde{Y}$) (e.g., 30-yr expectation value) derived from the loess regression,

$$Y_{base} = \frac{1}{30}\sum_{y=1981}^{2010} \left(\Delta Y_y + \tilde{Y}\right) \quad [S9]$$

We then examined and integrated county-level production for each component to state-level production,

$$\Delta P_{fab,y} = \sum_{c=1}^{n} \Delta P_{fab,c,y} \quad [S10]$$

$$\Delta P_{Y,y} = \sum_{c=1}^{n} \Delta P_{Y,c,y} \quad [S11]$$

$$\widehat{P_{base}} = \sum_{c=1}^{n} P_{base,c} \quad [S12]$$

where *y* refers to specific year of interest (e.g., 2023), *c* represents specific county, n equals to 105 (number of counties in Kansas, US). Lastly, we expressed production changes relative to climate-driven averaged production over 1981-2010 ($\widehat{P_{base}}$) as,

$$\%P'_y = \frac{\Delta P_{fab,y} + \Delta P_{Y,y}}{\widehat{P_{base}}} \times 100 \quad [S13]$$

where $\%P'$ is the relative production change (%) (shown in Fig. 2 in Main text).

The most critical climate driver of abandonment or yield loss

While RF models have built-in methods for assessing the relative importance of predictor variables, such as feature importance, these methods typically assess variable importance at a global scale. No insight is given as to what is driving the prediction for a given observation point. Thus, in this study, we integrated a game theory (Shapley additive explanations, SHAP) into the RF models we created (equations [1] and [2] in Main text) to quantify the most critical climate drivers in wheat abandonment and yield changes. The SHAP values provide local model explanations and the contribution of each predictor variable for each data point via estimating the expected marginal contribution of a predictor across all possible combinations of predictors. Specifically, the magnitude of the SHAP value represents the strength of contribution to the prediction. A larger absolute magnitude implies a stronger influence, while a smaller absolute magnitude indicates a weaker influence. We based our selection of predictors on the magnitudes for both directions (positive and negative). The more positive and more negative ones were selected as predictors. Moreover, a positive SHAP value means that it contributes to the increase in predicted value, and vice versa. Thus, we focused on the climate variables with more positive SHAP values for *fab* and with more negative SHAP values for ΔY.

The process to identify the climate drivers for yield and abandonment extremes includes five steps (fig. S6): 1) we defined an extreme abandonment event as the abandonment above the 90th percentile, and an extreme yield losses event as the yield variation below the 10th percentile across all years (1926-2023) for each county in Kansas; 2) for each observed crop yield extreme event, we used SHAP to break down the predicted value derived from the RF model into the contribution of individual climate variables; 3) climate variables with the top three highest SHAP values and the top three lowest negative SHAP values were identified as the primary influencers for the abandonment and yield loss events, respectively; 4) based on the climate variables we used and the prevalent climate causes, we focused on four climate extreme events (low precipitation, high precipitation, cold temperatures, and warm temperatures) during four specific growth periods based on 10th or 90th percentiles of historical probability distributions. Note, a cold event occurs when Tn falls below its 10th percentile, whereas a warm event occurs when Tx exceeds its 90th percentile; 5) we determined the most critical climate driver for each abandonment or yield loss extreme. For example, for extreme abandonment, if the climate variable with the highest SHAP value (step 3) corresponded to any of four types of climate extremes above (step 4), it was deemed as the most pivotal climate extreme event; if no correspondence was found, we proceeded to look for the second highest variable, and subsequently to the third if necessary. Finally, all extremes in abandonment (or yield) associated with the identical primary climate drivers were grouped together, resulting in 16 distinct groups. These groups were then used to calculate the percentage of area driven by the same primary driver relative to the total area experiencing extreme abandonment (or yield losses).

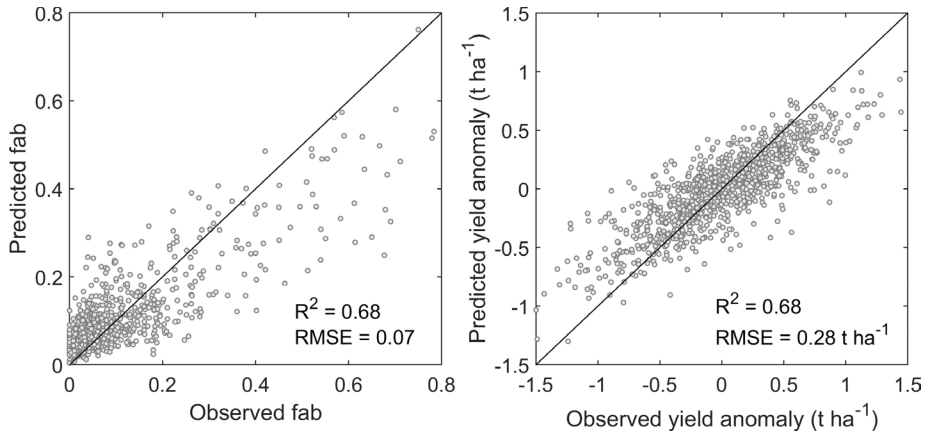

Fig. S1.
Predicted *vs.* observed abandonment rates (fab) and yield anomaly (t ha^{-1}). Black lines represent the 1:1 line. We showed the coefficient of determination (R^2) and root mean square error (RMSE).

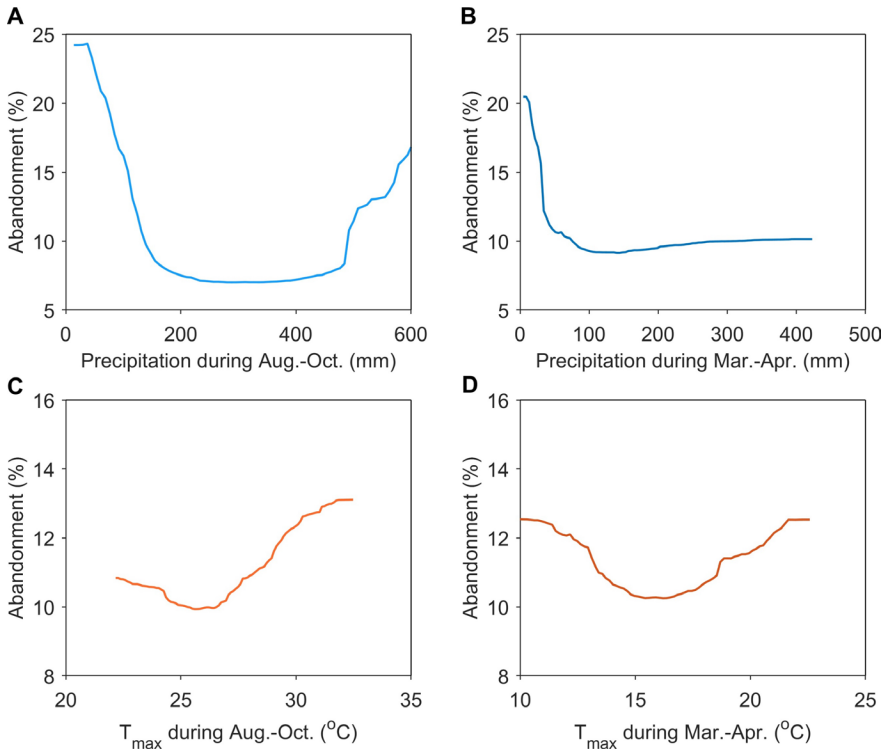

Fig. S2.
Partial response of crop abandonment to precipitation in Fall (A) and Spring (B), as well as to maximum temperature in Fall (C) and Spring (D) taken from our random forest (RF) model.

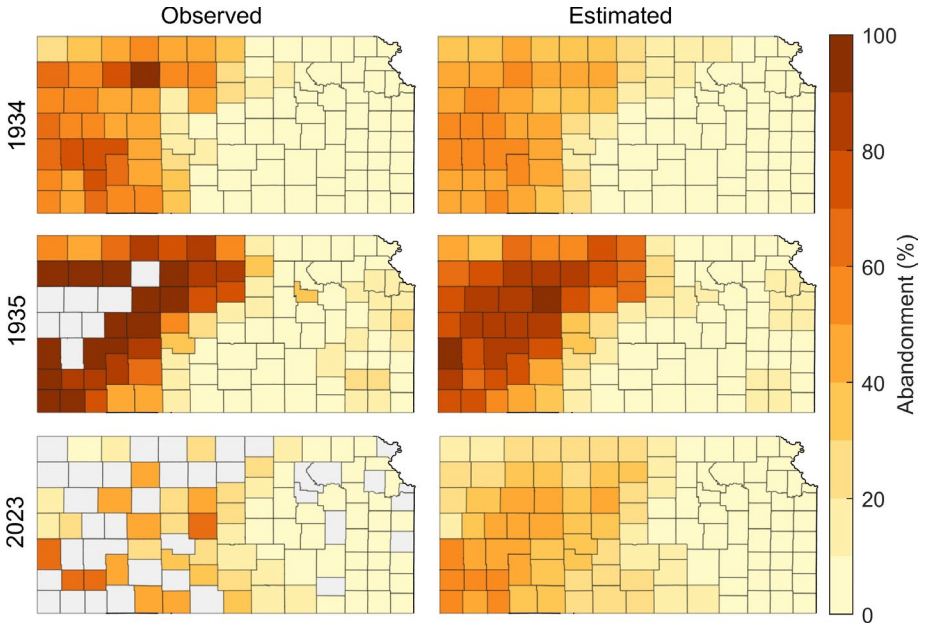

Fig. S3.
Robustness check of the estimation. Spatial distributions of the observed and estimated abandonment in typical years (1934, 1935, and 2023). The gray regions for the observed abandonment in 1935 and 2023 panels represent areas where county-level data on areas were unavailable.

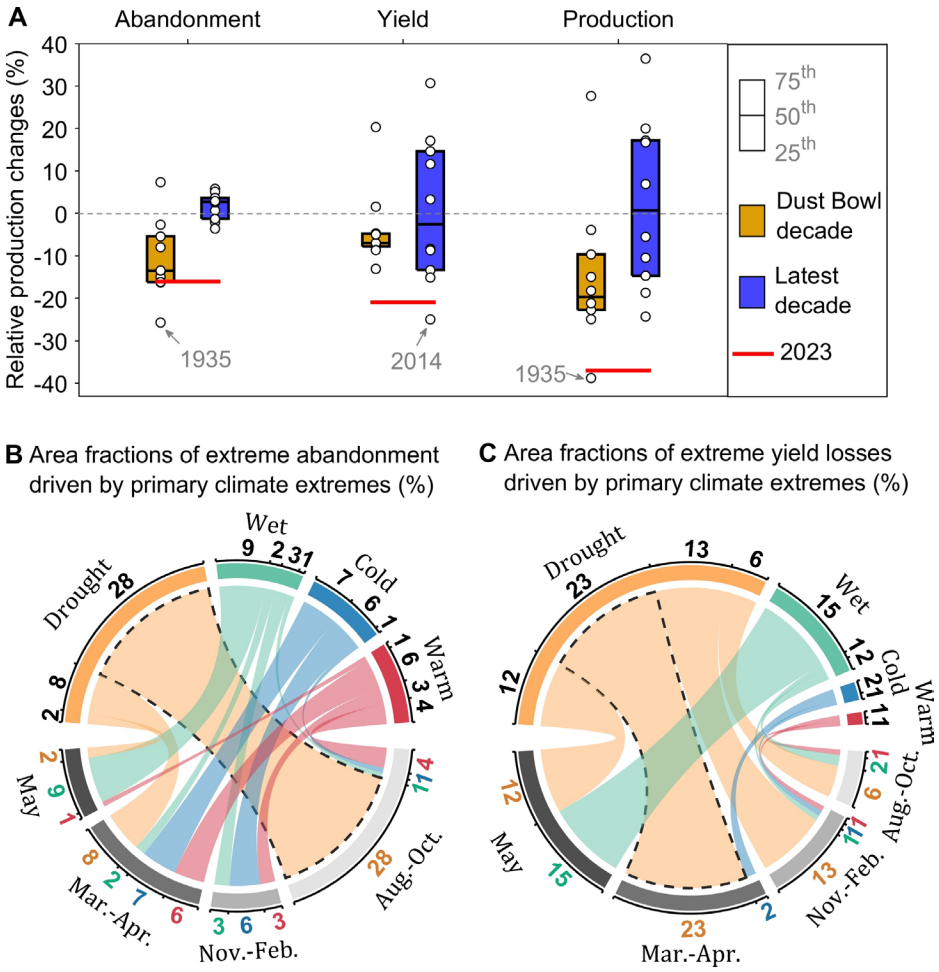

Fig. S4.
The decomposition of wheat production and primary climate drivers of extreme abandonment and yield loss. Similar to Fig. 2B and Fig. 3, E and F, but the accumulated precipitation was replaced with the Palmer Z Index in the models.

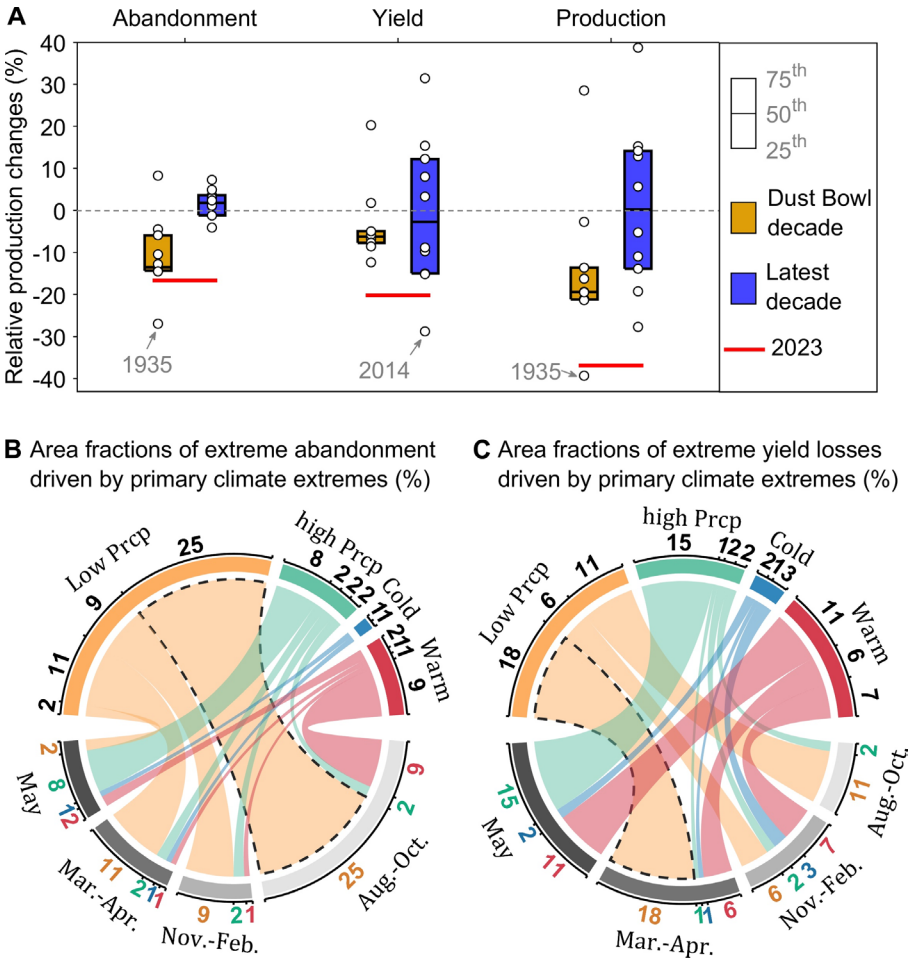

Fig. S5.
The decomposition of wheat production and primary climate drivers of extreme abandonment and yield loss. Similar to Fig. 2B and Fig. 3, E and F, but the maximum and minimum temperature were replaced by the fraction of warm days and cold days in the models, respectively.

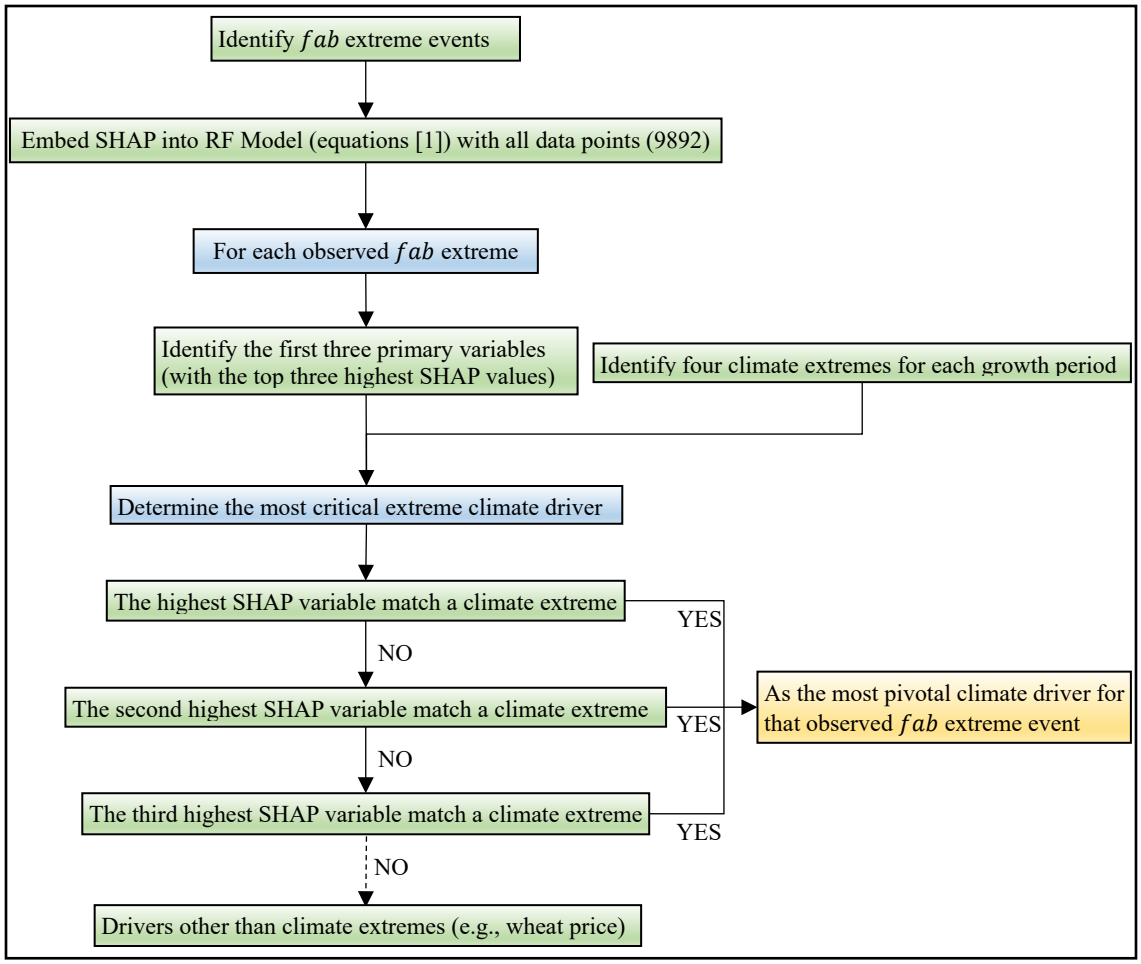

Fig. S6.
Workflow of identifying the critical climate driver of each extreme abandonment. This workflow was also carried out for the extreme yield loss (ΔY; equation [2] in the main text), but using the lowest SHAP variable. This process was repeated for each observed extreme abandonment or yield loss occurrence. After the workflow, all extremes in abandonment or yield losses associated with the same primary climate driver were used to calculate the area fraction. Additionally, the area fraction for extremely low wheat prices (<10th) for *fab* was also calculated. The detailed descriptions can be found in Supplementary Text.

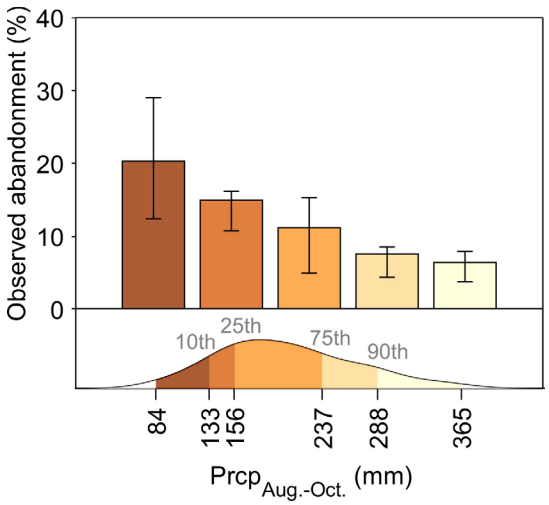

Fig. S7.
Effects of fall precipitation on wheat abandonment in Kansas. The state-level observed abandonment (%) binned over fall precipitation (Prcp) from 1926-2023. The error bars indicate the 25th and 75th percentiles across all years within each precipitation bin. The lower panel displays the state-level precipitation distribution color-coded over 1926-2023 and precipitation percentiles are shown for defining the bins used.

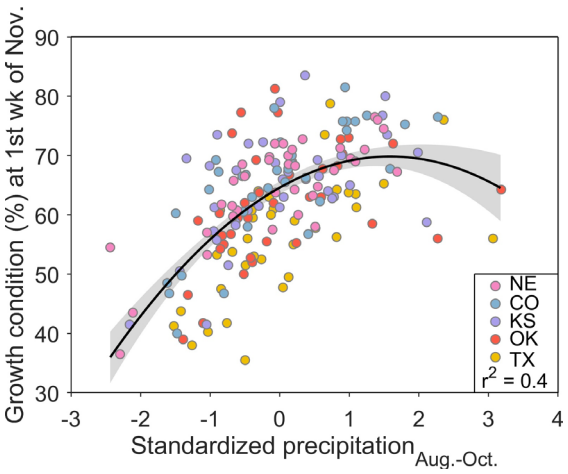

Fig. S8.
Relationship between wheat growth index conditions and precipitation in early November (first week) from 1987 to 2023. The quadratic relationship between winter wheat growth conditions and precipitation during August-October across the five wheat states (Nebraska, NE; Colorado, CO; Kansas, KS; Oklahoma, OK; and Texas, TX). State-level growth conditions were derived from USDA-NASS.

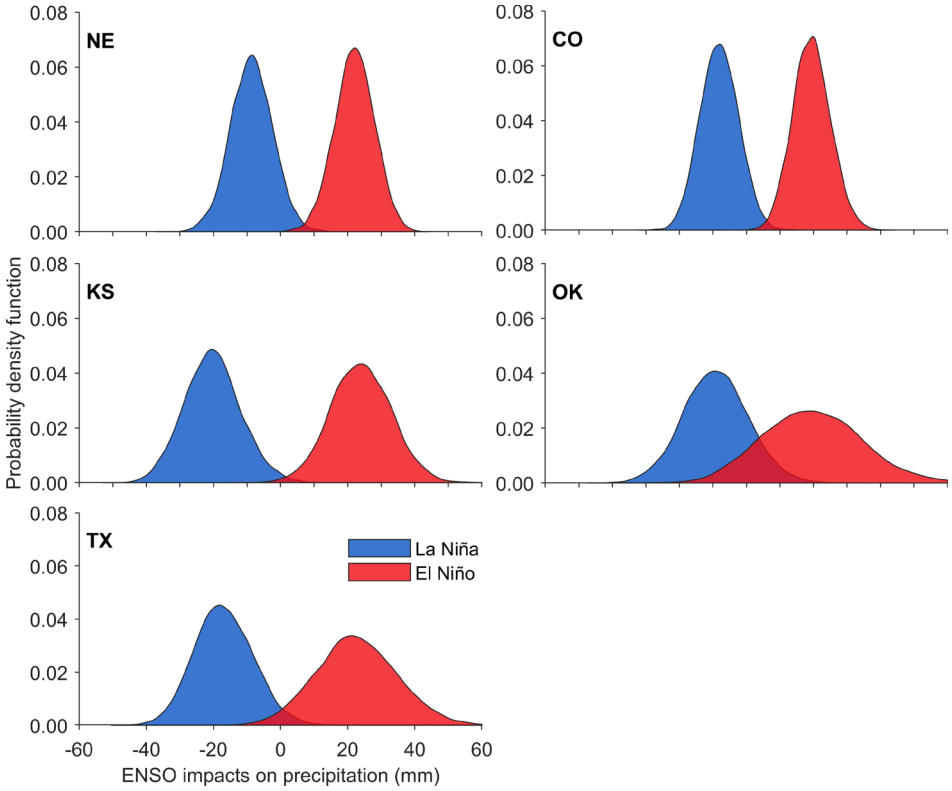

Fig. S9.
Impacts of ENSO phases on linear detrended precipitation anomalies. Distributions of area-weighted average fall precipitation anomalies during El Niño (red) and La Niña (blue) phases in Nebraska (NE), Colorado (CO), Kansas (KS), Oklahoma (OK), and Texas (TX) based on average harvested area during 1981-2010. Negative and positive values indicate ENSO phases with decreased and increased precipitation, respectively. Distribution of precipitation anomalies in both El Niño and La Niño phases were plotted based on bootstrapping approach (n = 10,000) (see Main Methods).

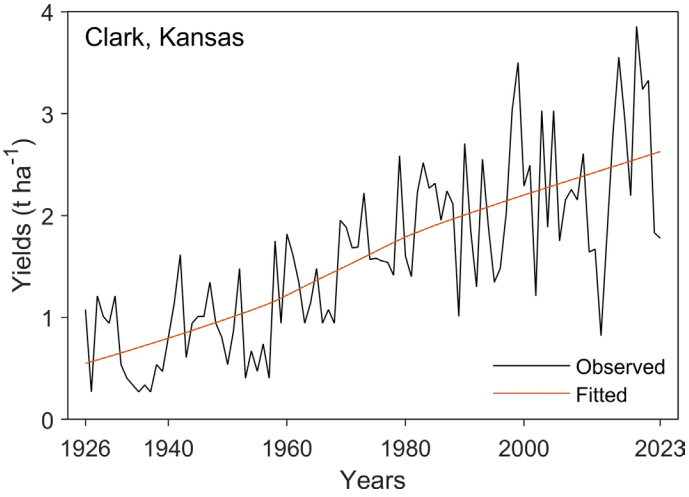

Fig. S10.
An example of yield trend used in this study through locally weighted smoothing regression with a span parameter of 0.75.

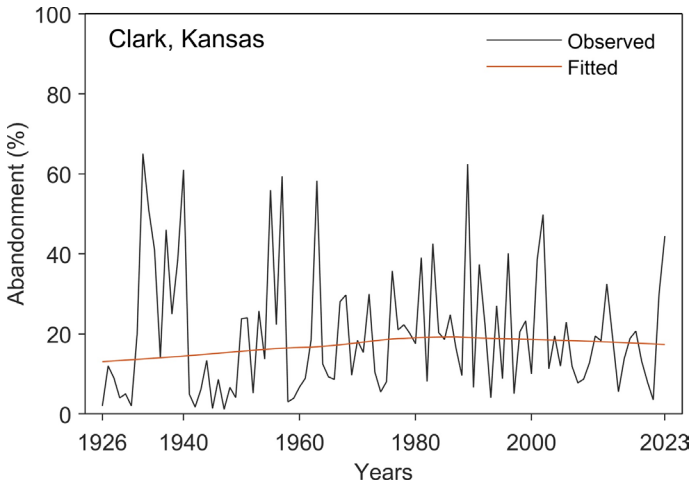

Fig. S11.
An example of crop abandonment variations (Clark County in Kansas) with a locally weighted smoothing regression (a span of 0.75) used for detrending process.

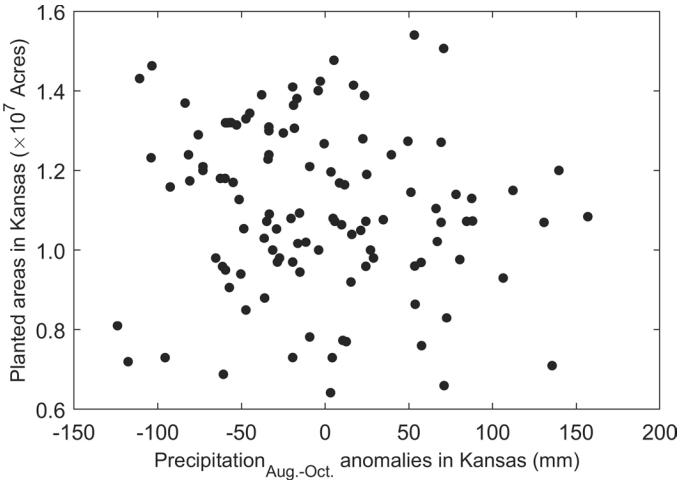

Fig. S12.
State-level planted areas of winter wheat *vs.* pre-season precipitation anomalies (with a base period of 1981-2010) across 1909-2023 in Kansas. Planted area data were derived from the United States Department of Agriculture-National Agricultural Statistics Service (USDA-NASS).

Table S1.
Modeling performance.

Models	Yield changes		Abandonment fraction	
	R^2	RMSE (t ha^{-1})	R^2	RMSE (-)
Model 1 (Prcp, Tx, and Tn)	0.68	0.28	0.68	0.07
Model 2 (Palmer Z-Index, Tx, and Tn)	0.70	0.27	0.65	0.08
Model 3 (Prcp, f_{wd}, and f_{cd})	0.70	0.27	0.69	0.07

Notes: Model 1 is the model we used in Main text with predictors including the accumulated precipitation (Prcp), mean maximum temperature (Tx), and mean minimum temperature (Tn) during four specific growth periods (fall, winter, spring, and early summer). In Model 2, the Prcp was replaced with the Palmer Z-Index. We also substituted the Tx and Tn in Model 1 by the fraction of warm days (f_{wd}) and cold days (f_{cd}) for Model 3.

Table S2.
Available years for planted and harvested areas data from USDA-NASS.

States	County-level	State-level
Nebraska	1956 - 2023	1909 - 2023
Colorado	1929 - 2023	1909 - 2023
Kansas	1926 - 2023	1909 - 2023
Oklahoma	1926 - 2023	1909 - 2023
Texas	1968 - 2023	1909 - 2023