The replication files are organized in various subfolders:
dataSTATA: has all the STATA files needed for the analysis

- weather_corn, weather_cotton, weather_soybeans average the fine-scaled weather data over the growing area of each crop, where the area is obtained from the cropland data layer (average of the years 2010-2019). The data are provided by month, so various growing-season definition can be chosen (e.g., sum March-August).
- yieldData: has the annual county-level yield data from the National Agricultural Statistics Service.

codeSTATA: has the replication code

- 1_analysis.do calls several subroutines to conduct the analysis
- 2_figures.do calls subroutines to plot the figures

resultSTATA: has the saved results after running 1_analysis.do that are used in the figures

Note: the data set has been updated to include the latest advances, so the results differ slightly (but all major results go through)

- It includes a new interpolation procedure that includes more weather station than what was used in the PNAS article. (Missing values are cross-interpolated using a station's own cumulative density function). It is the same data set that is published online: http://www.columbia.edu/~ws2162/links.html
- The data is extended to the year 2020 to include the more recent years 2006-2020.
- The original data average weather using a one-time satellite scan of the entire cropland area in a county and hence the weather was the same for all crops. The new version uses the crop-specific area, i.e., weather in a county differs by crop as it places different weights on the PRISM grids. The weights are the average cropland area for the last decade (2010-2019) as reported in the cropland data layer.
- The results cluster the error by state to account for spatial correlation. The original paper used an adjustment of the Conley routine that looks for spatial correlation within a year. The routine is provided for reference.