Corteva – Climate Change Impact on Crop Yield

PURDUE UNIVERSITY | The Data Mine

Alex, Alexander, Brady, Josh, Mihika, Sahasra, Sandilya, Temi

PROJECT BACKGROUND

This project explores the relationships between **crop yield** and **climate change**. We set out to provide meaningful results that will assist Corteva in ensuring progress in agricultural science, enriching the lives of those who consume and produce. Our project follows three workstreams, summarized by the following research questions:

- In the future, what areas of the world will have climates that match current Midwestern climates?
- Will crop yield anomalies escalate in the future?
- Can we quantify the likelihood of yield stability in the future?

METHODS & TOOLS

- **Computed heat maps** of temperatures and precipitations in the Midwest to verify data integrity against existing maps using public access data from the ISIMIP climate data repository.
- **Modelling of the heat maps** used GFDL-ESM4 model, which is a coupled chemistry-carbon-climate Earth System, along with ssp370 scenario which represents an environment with moderately high greenhouse emissions.
- **Calculated regions** that had the highest corn yield from USDA corn database, combining that with climate classification information (**Fig. 1**).
- **Computed crash years classifications** based on historical data, finding the climate classifications that were correlated to poor or anomalous yields, in this case, 1988 (**Fig 2**)
- **Exploratory analysis** started with Indiana, focusing on historical climate and yield data to identify correlations between climatic factors and crop performance.
- Using R, Python, and API keys, we **obtained public access data** from the USDA and NASA.
- **Exploratory analysis of national yield data** using ArcGIS and R.

$$Yield\ Residual_t = \beta_0 + \beta_1 * Temp\ Residual_t + \beta_2 * Prcp\ Residual_t + \varepsilon_t$$

We detrend yield, sum of seasonal precipitation, and average seasonal temperature by county and use the residuals from those trends as our variables of interest. Each of these variables used in our main model can be seen in **Fig. 4**. One year of interest is **2012**, where the Midwest saw low rainfall and high temperatures, resulting in a clear large drop in US yields as revealed by one of the **lowest yield residuals** on the graph. This model which aims to capture variation in detrended yield using detrended climatic variables is derived from the methods of Kukal and Irmak (see references).

RESULTS

We find high statistical significance (**99.99%**) in the linear relationships between detrended yield and the climatic variables for both corn and soybeans. With relatively low R-squared values (**10.16%**), we know additional linear predictors or a model that captures nonlinear effects would explain more variation in yield.

The residuals of our main model can provide value about what **years** and **locations** were able to be **predicted** with higher accuracy. For example, 2003 was the largest drop in yields relative to trend, but 1988 was the crash year that was predicted with higher accuracy. In other words, our model did a **better job** using climate data in predicting the crash in 1988 compared to 2003 or 2012.

Fig. 1

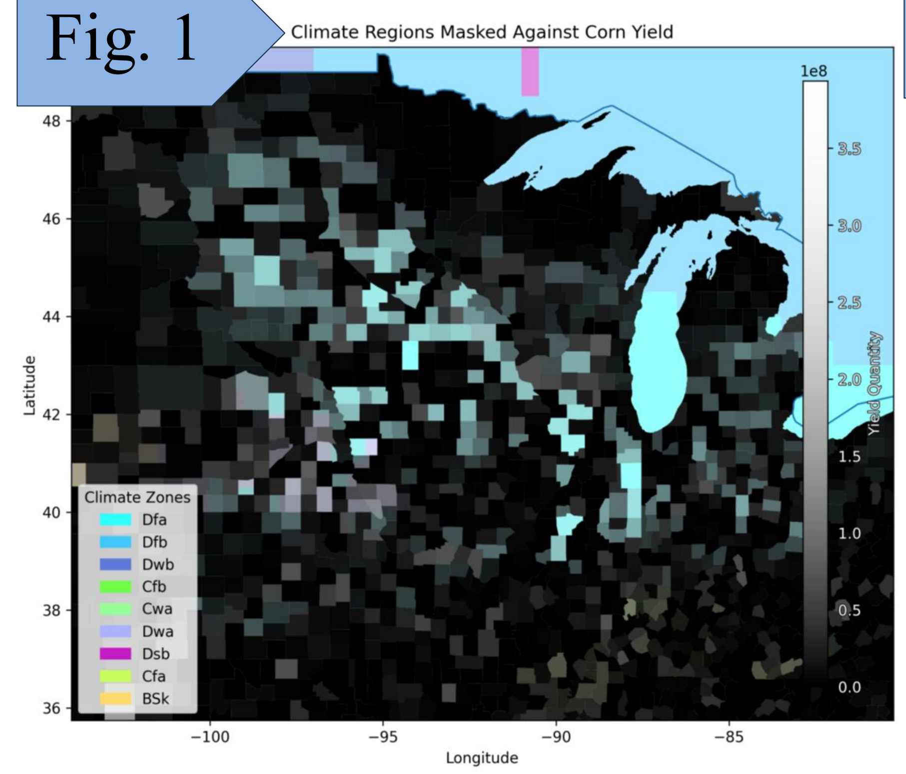


Fig. 2

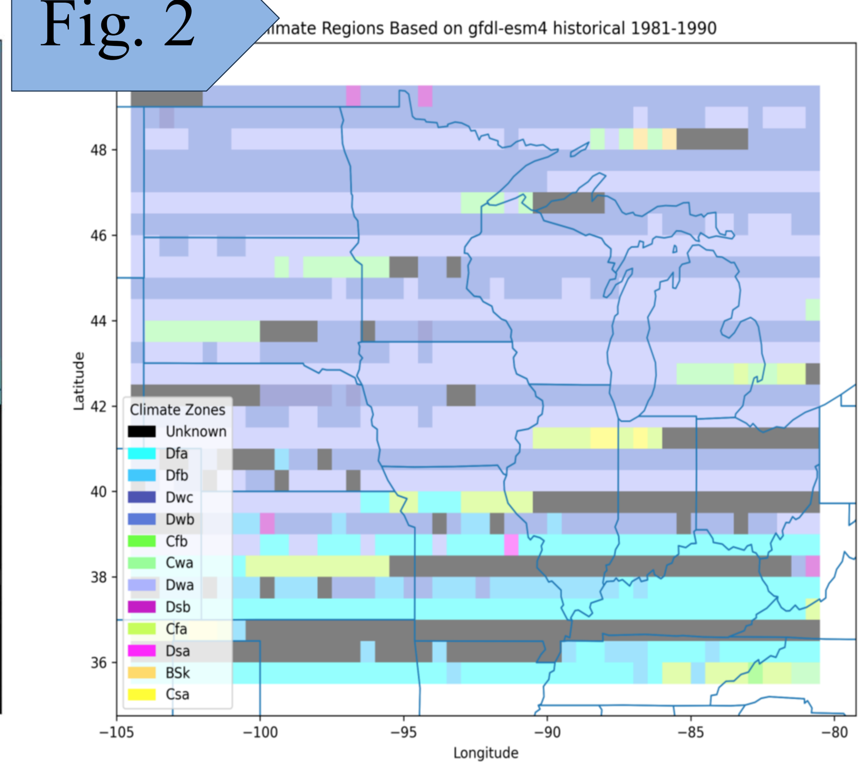
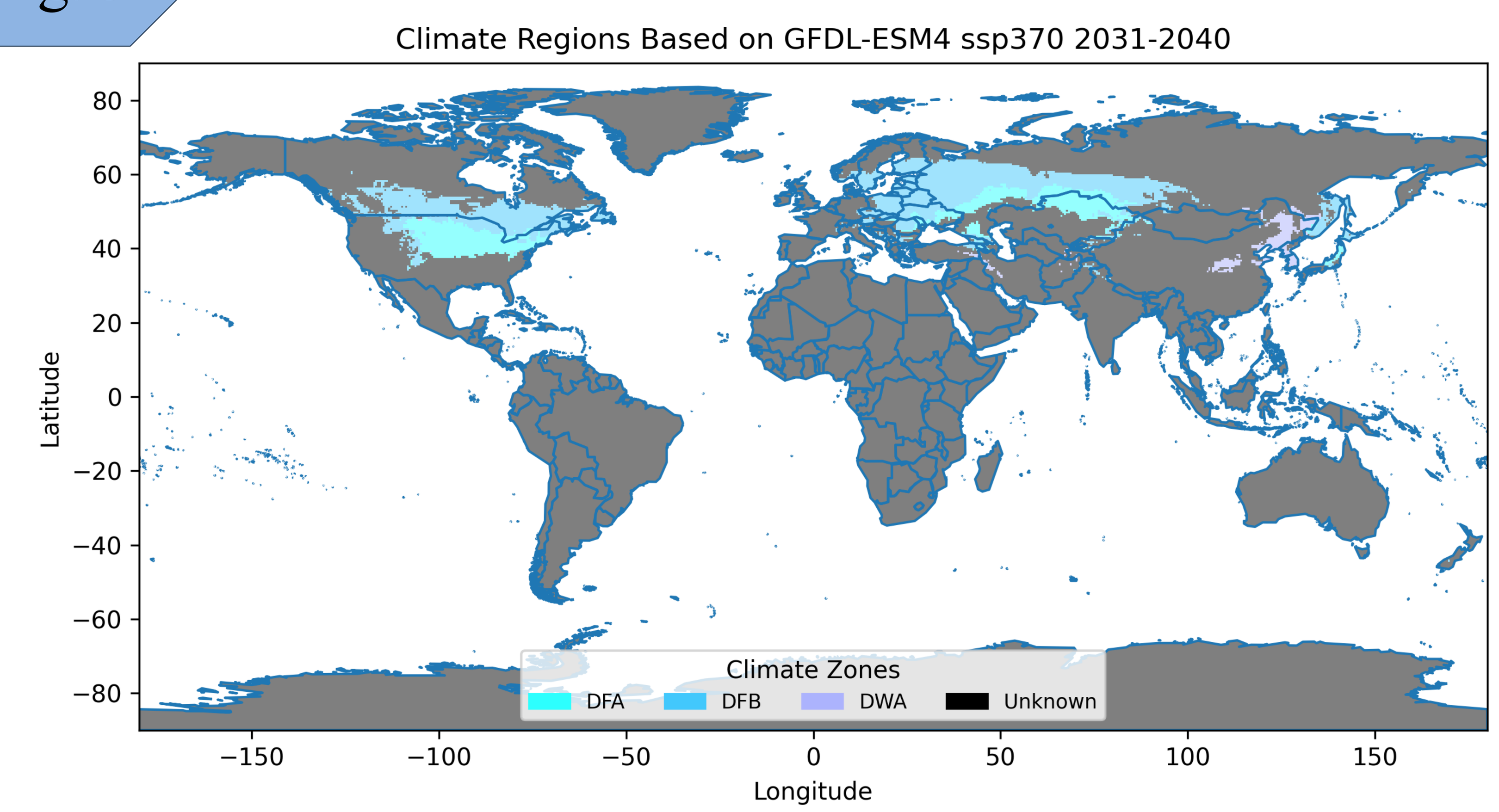
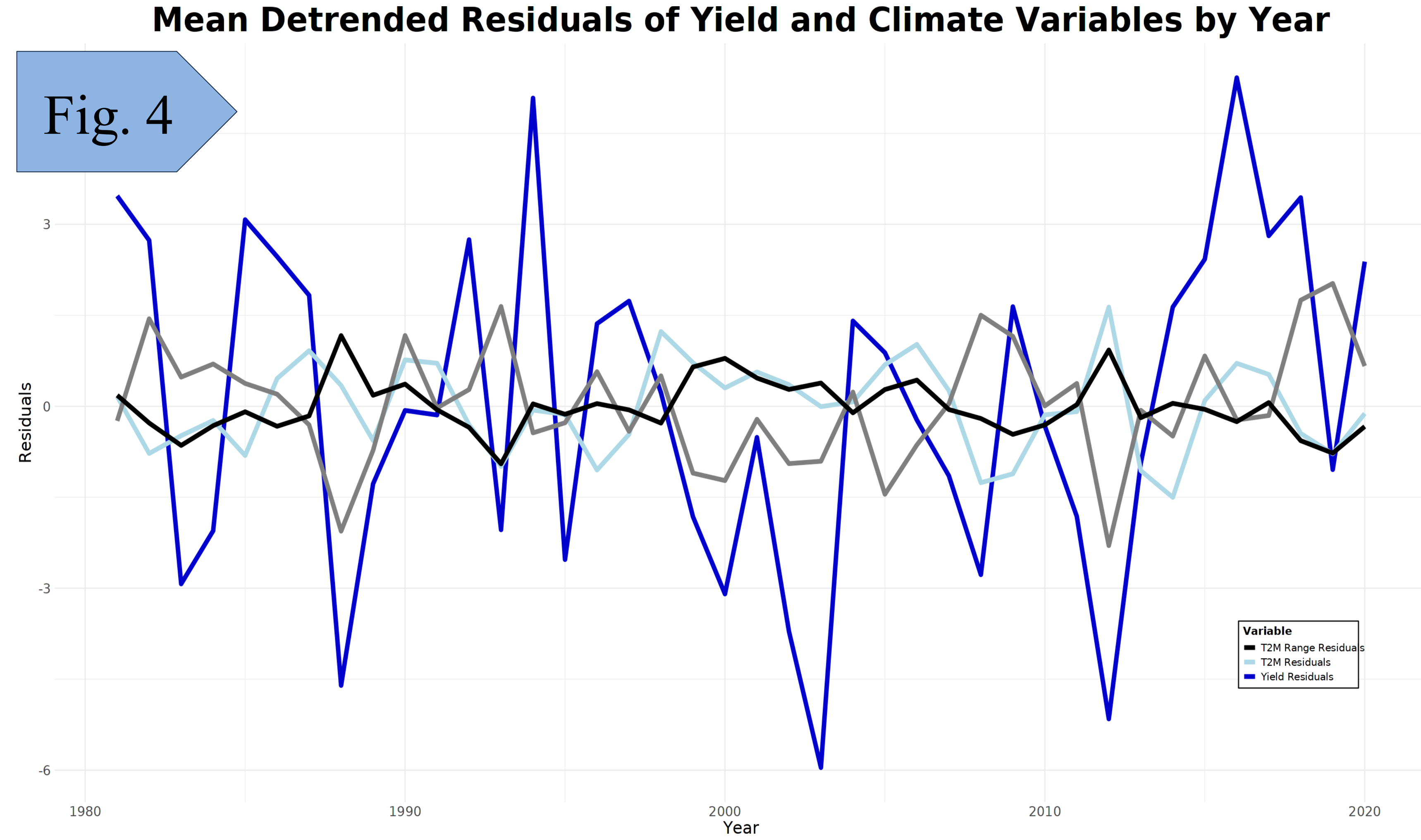


Fig. 3



Mean Detrended Residuals of Yield and Climate Variables by Year

Fig. 4



CONCLUSIONS

- Based on current corn growth in the United States, it is possible to **predict** which **climate classifications** can have **ideal corn yield** based on the Koppen-Geiger climate classifications.
- Key areas identified for future corn growth include the Midwest, Central Europe, and Northeastern China (**Fig. 3**).
- The DWA climate zone was very apparent in the Midwest during the 1988 crash year (**Fig. 2**), suggesting that Northeast China (also DWA, **Fig. 3**) may be more vulnerable to yield instability.
- Based on the R-squared values from the main yield-climate models, we conclude that a **greater amount** of **variance** in soybean yields can be explained by climate compared to corn yields.

FUTURE GOALS

- Focused analysis of yield stability on a specific timeframe or geographic region, especially in the **2000-2020** range where yearly yield residuals seem to closely match yearly summed precipitation.
- Improving model accuracy **using non-linear predictors** of yield residuals and allowing for more linear predictors in the model such as soil characteristics or the use of irrigation pivots.
- Creating a **testing framework** for the predicted climate classifications using historical climate data.
- In a new direction, data collected from USDA to enable a **prediction** of **corn yield itself**, leveraging models trained on the climate data.

REFERENCES & SOURCES

- **USDA** ERS API (corn and soybean yield data, 1940-2023)
- **US Census Bureau** for county latitude and longitude data
- **NASA Power** Data Services for climate data (1981 – 2022)
- **ISIMIP** Climate Database for temperature and precipitation

Kukal & Irmak (2018), Climate Driven Crop Yield and Yield Variability and Climate Change Impact on the U.S. Great Plains Agricultural Production, *Scientific Reports* (8), Article Number: 3450

ACKNOWLEDGEMENTS

Thank you to our T.A. Smarika and the Data Mine staff that helped our project run smoothly. And a huge thank you to our Corteva mentors - Renee, Oscar, and Jose!