AGU 2025 | GC13F-0713 | Mon Dec 15 • New Orleans

Beyond Climate: ML-Powered Forecasting of Crop Yields Using Soil & Satellite Data

Arit Prince
American High School • Stanford YIP • Bright STARS

Arya Prince
UC Berkeley (Data Science) • Science and Beyond

github.com/ScienceAndBeyond/CropCast

Image: AI-generated

MOTIVATION & DATA

*Can integrating **climate**, **Vegetation Indices (NDVI & EVI)**, and **soil properties** significantly improve crop yield predictions over climate-only models?*

Practical Goal

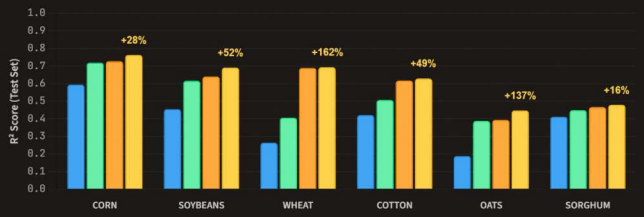
Help farmers make informed planting decisions by predicting yields across different crops, with future integration of commodity prices for profit optimization.

734	11	15	20K
COUNTIES	STATES	YEARS	RECORDS

Crops Analyzed (Records)

Corn	6,680	Soybeans	5,679
Cotton	1,724	Oats	1,647
Sorghum	1,212	Wheat	1,014

MODEL PERFORMANCE: R^2 COMPARISON (TEMPORAL HOLDOUT 2022–2024)



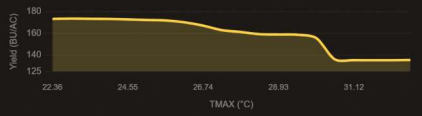
+74%
Average Improvement over Climate-Only

$R^2 = 0.76$
Best Model (Corn, All Features)

Climate Only | + Vegetation | + Soil | All Features

KEY DISCOVERIES

Quantifying Climate Risk: Corn Yield vs TMAX



⚠ **Heat Stress Threshold:** Corn yield plummets 38 BU/AC when TMAX exceeds 29°C—identifying actionable thresholds for adaptation planning.

Vegetation Extremes Beat Averages

Using **min/max** of NDVI & EVI (capturing stress periods + peak health) outperforms seasonal means. Sorghum: EVI max alone = 33% of prediction power!

Top Predictor by Crop

Corn:	TMAX	20.7%	Wheat:	pH	28.8%
Sorghum:	EVI	33.2%	Cotton:	SRAD	21.9%
Soy:	EVI	22.7%	Oats:	TMIN	16.0%

DATA PIPELINE & ML WORKFLOW

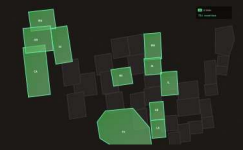
NOAA gridMET	MODIS	SoilGrids	USDA NASS
Climate	MOD13A3 Vegetation	Soil	Yields
→ GEE monthly aggregation	→ Google Earth Engine	→ WCS parallel downloads	→ QuickStats API (6 crops)
→ County reduceRegions	→ County reduceRegions	→ Raster mask by county	→ Filter ≥1000 samples
→ Apr–Sep growing season	→ Annual min/max NDVI & EVI	→ 0–30cm depth average	→ County-year BU/AC values
6 features	4 features	4 features	Target: YIELD

Merge — Inner join FIPS+year → 20,353 records → **Preprocess** — Drop missing, 14 features → **Temporal Split** — Train: '10–'21, Test: '22–'24 → **Random Forest** — 200 trees/crop, max_features=0.5

Acknowledgments

- Stanford Young Investigators Program (YIP) for research inspiration and guidance
- Google Earth Engine for computational resources
- Data providers: USDA NASS, NOAA gridMET, NASA MODIS, ISRIC SoilGrids
- My family for their unwavering support and encouragement throughout this project

STUDY AREA



Top States (records):

MN	3,644
TX	3,643
IA	3,262
IL	2,857
NE	2,500

Code & Contact

github.com/ScienceAndBeyond

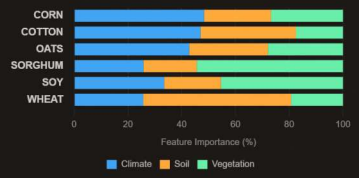
aritprince@gmail.com

CROP PROFILES

CORN	SOYBEANS	WHEAT
$R^2 = 0.76$	$R^2 = 0.69$	$R^2 = 0.69$
+28% vs climate	+52% vs climate	+162% vs climate

COTTON	OATS	SORGHUM
$R^2 = 0.63$	$R^2 = 0.45$	$R^2 = 0.48$
+49% vs climate	+137% vs climate	+16% vs climate

Feature Category Importance by Crop



KEY TAKEAWAYS

1. **Multi-source data essential**—climate alone explains 19–60% of variance; combined features reach 45–76%.
2. **Crop-specific insights**—Wheat benefits most from soil pH; Sorghum from vegetation indices.
3. **Extreme values matter**—Min/max of NDVI/EVI outperform averages for yield prediction.
4. **Actionable thresholds**—PDP reveals critical TMAX limit (~29°C) for corn yield collapse.

⚠ Limitations

- County-level aggregation masks within-county variability
- No irrigation or management practice data
- Static soil properties (no annual dynamics)
- Generic Apr–Sep growing season applied to all crops
- Temporal validation limited to 3 years (2022–2024)

Future Work

- Incorporate irrigation data from USGS
- Crop-specific NDVI masking (USDA CDL)
- Deep learning models (LSTM, Transformers)
- Single-crop deep-dive study (corn across Midwest)
- Farmer decision tool with commodity prices