

# GC45C-09: Machine learning emulators and empirical models combining climate and global crop models for seasonal agricultural production



**Alex Ruane<sup>1</sup>, Andres Castellano<sup>1</sup>, and Jonas Jägermeyr<sup>2,1</sup>**

*<sup>1</sup>NASA Goddard Institute for Space Studies, New York City*

*<sup>2</sup>Columbia University Climate School, Center for Climate Systems Research*

**AGU Fall Meeting – Chicago – December 15th, 2022**



Goddard Institute for Space Studies  
New York, N.Y.



CIDs identify the specific climatic conditions associated with biophysical responses:  
 Ruane et al., 2022: The Climatic Impact-Driver Framework – Earth’s Future

ipcc INTERGOVERNMENTAL PANEL ON climate change	Climatic Impact-driver																																
	Heat and Cold		Wet and Dry					Wind			Snow and Ice				Coastal		Open Ocean			Other													
Asset	Mean air temperature	Extreme heat	Cold spell	Frost	Mean precipitation	River flood	Heavy precipitation and pluvial flood	Landslide	Aridity	Hydrological drought	Agricultural and ecological drought	Fire weather	Mean wind speed	Severe wind storm	Tropical cyclone	Sand and dust storm	Snow, glacier and ice sheet	Permafrost	Lake, river and sea ice	Heavy snowfall and ice storm	Hail	Snow avalanche	Relative sea level	Coastal flood	Coastal erosion	Mean ocean temperature	Marine heatwave	Ocean acidity	Ocean salinity	Dissolved oxygen	Air pollution weather	Atmospheric CO <sub>2</sub> at surface	Radiation at surface
Crop systems																																	

None/low confidence    Low/moderate    High

Impacts and risk relevance

Assessment of all impact sectors and climate changes in each region:  
 Ranasinghe et al., 2021: IPCC AR6 WGI Chapter 12

## Synthesis outputs from AgMIP Global Gridded Crop Model Intercomparison using weather indices

### GGCMI Phase 2 (Franke et al., 2020a,b):

- 12 models harmonized for planting date, fertilizers, cultivars
- Sensitivity tests across CO<sub>2</sub>, Temperature, Water and Nitrogen (+/- 1°C tests with high N)
- Maize, wheat, rice and soy; global 0.5° x 0.5° resolution; 30 years

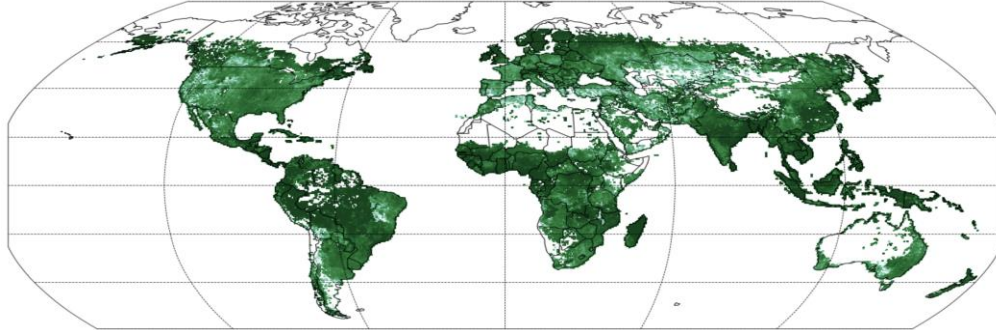
### 40 seasonal weather indices:

- Mean temperature and precipitation
- High heat days (30°C and 35°C) and cold day thresholds (5°C, 0°C)
- Number of rainy days, consecutive dry days
- 5 growing season periods (full GS, planting, pre-anthesis, anthesis, post-anthesis)

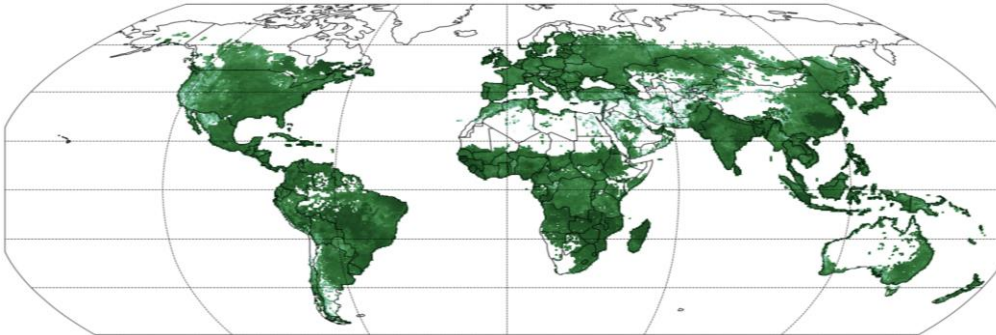
### Extreme Gradient Boosting (XGBoost) Machine Learning model

- 3 features targeting yield outputs independently at each grid cell<sub>3</sub>

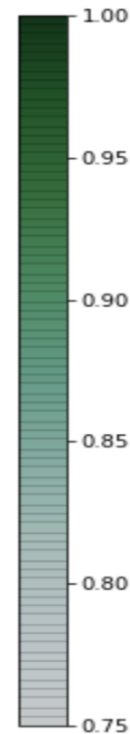
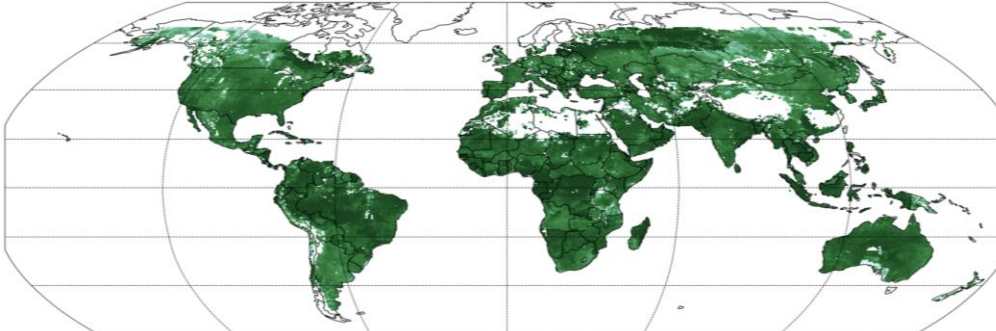
pDSSAT



LPJML



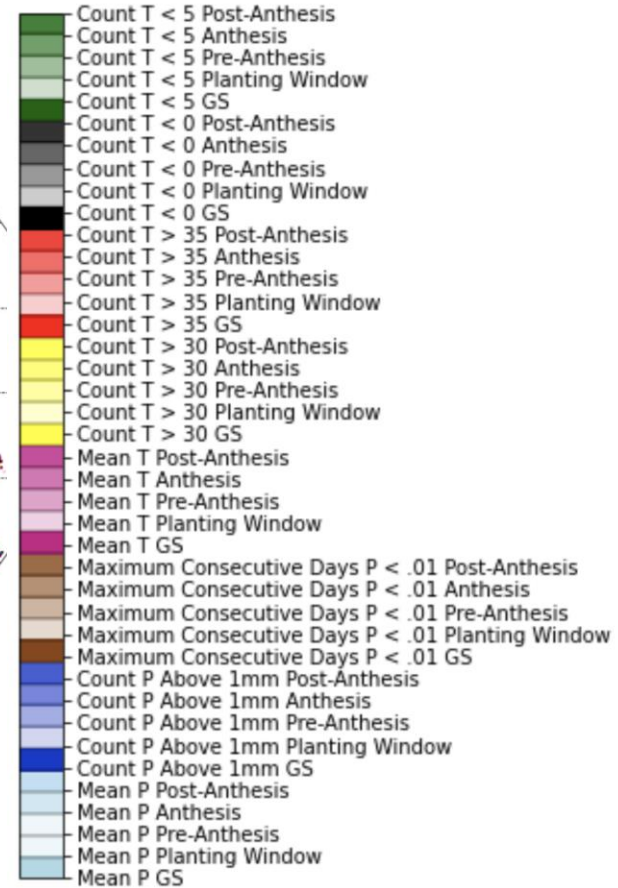
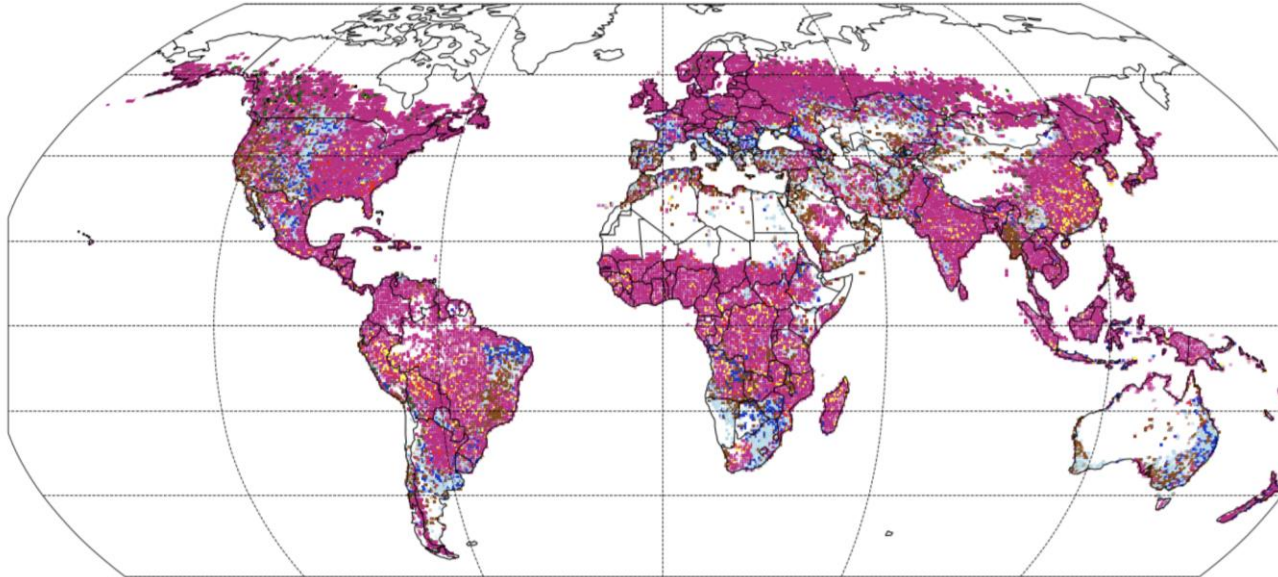
EPIC TAMU



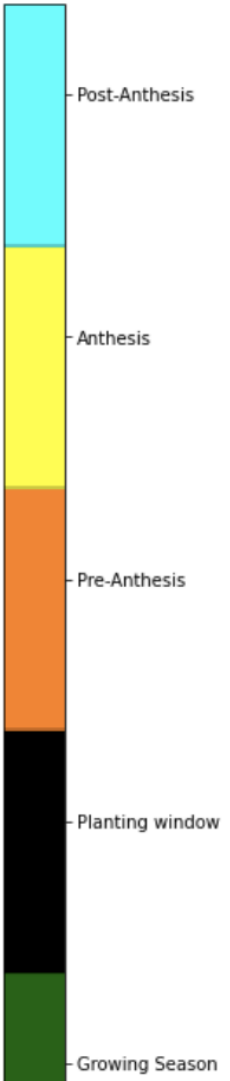
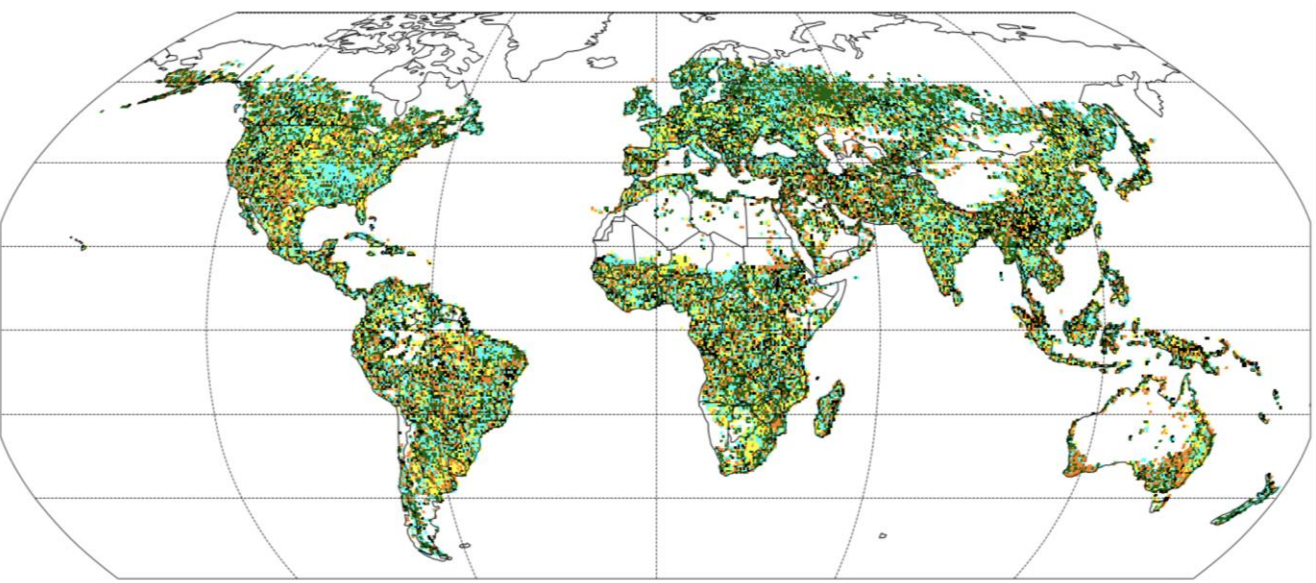
$R^2$  vs. underlying GGCM  
Rainfed Maize



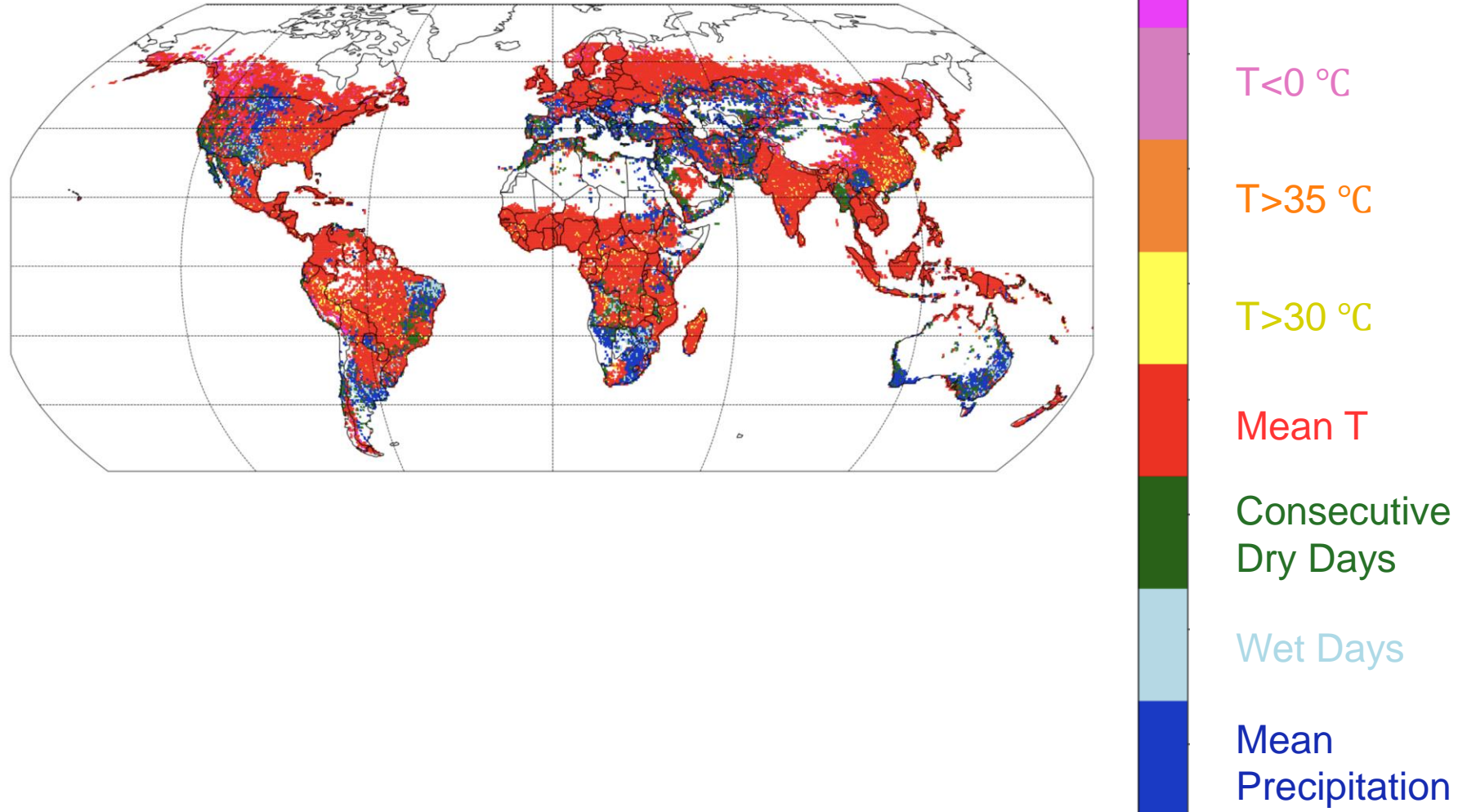
Most Important Rainfed Maize Feature (pDSSAT)



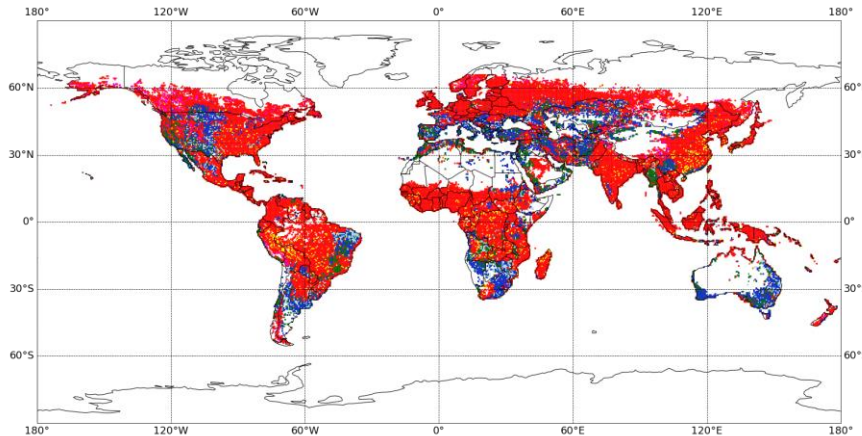
### Phenology of Most Important Rainfed Maize Feature (pDSSAT)



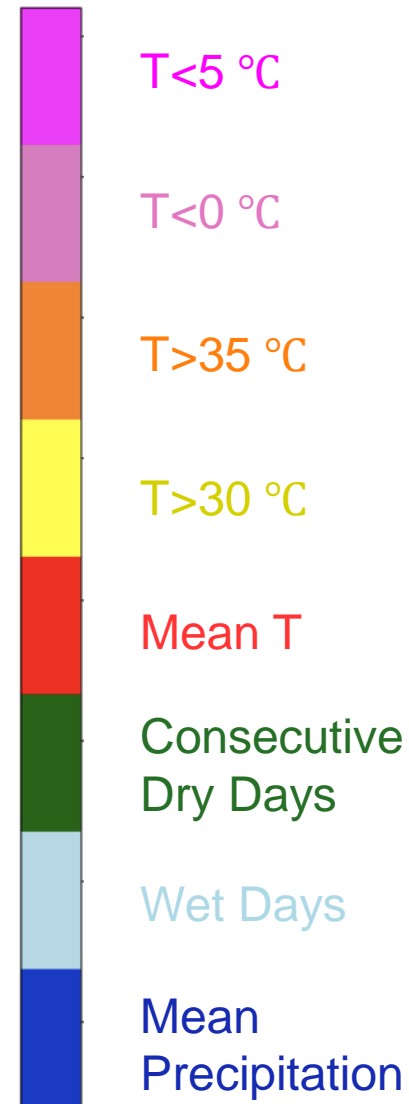
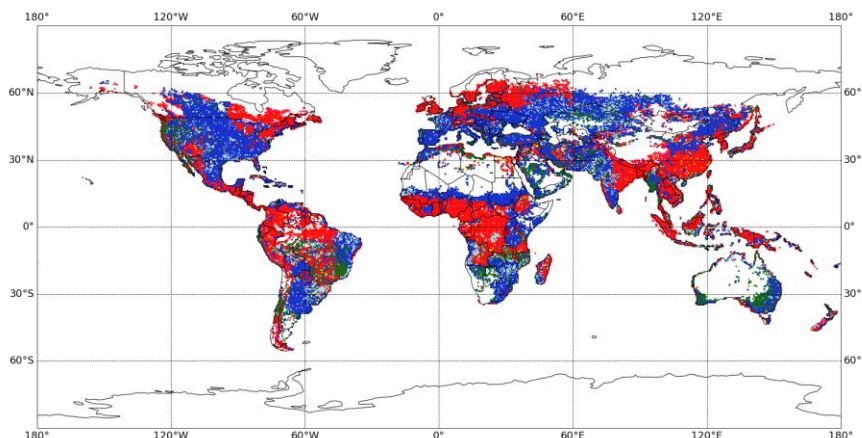
Most Important Rainfed Maize Variable (pDSSAT)



**pDSSAT Rainfed Maize**  
top predictor variable:

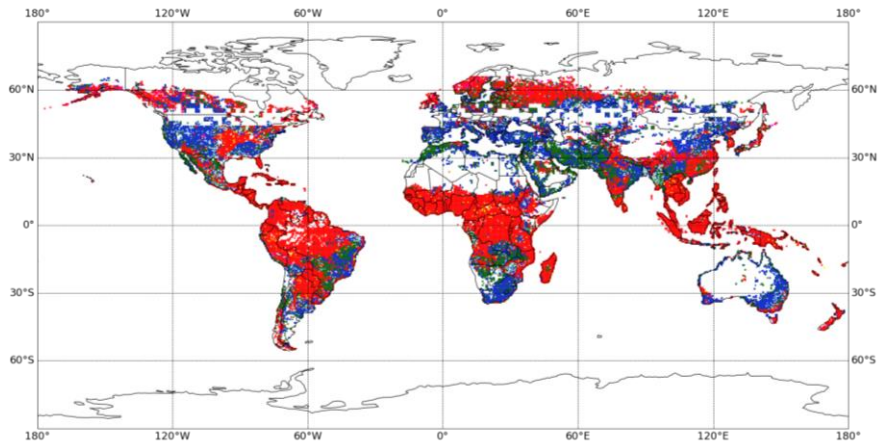


**LPJmL Rainfed Maize**  
top predictor variable:

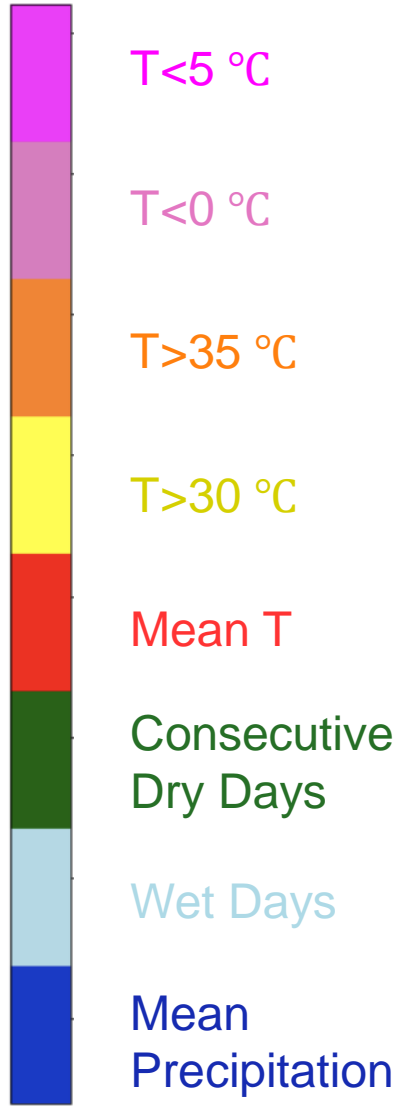
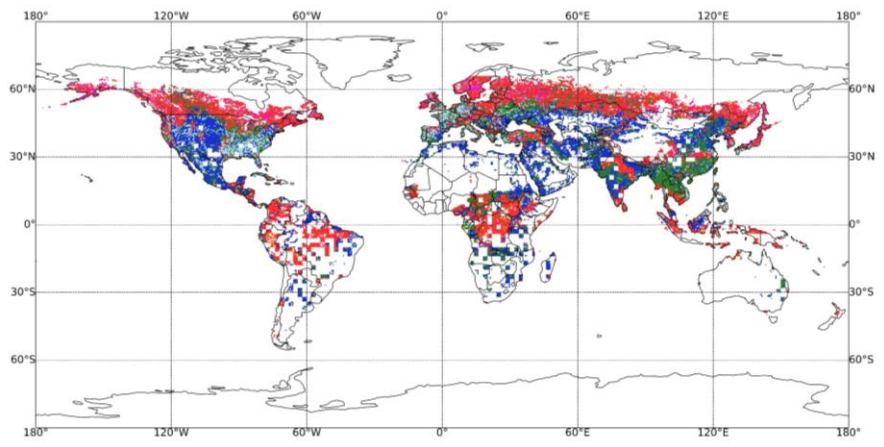




pDSSAT **Spring Wheat**  
top predictor variable:



pDSSAT **Winter Wheat**  
top predictor variable:



**ML emulators can capture bulk of historical crop model simulations with as few as 3 climate variables**

**Leading predictor features vary across region, system and model**

**Results point to common climate drivers for regional responses**

**Models respond most strongly to mean climate conditions**

**- Likely underestimate role of extreme events**

## Outputs from AgMIP Global Gridded Crop Model Intercomparison compared against reported production

### GGCMI Phase 3 (Jägermeyr et al., 2021) / ISIMIP Phase 3a:

- 10 harmonized GGCMI
- Maize, wheat, rice and soy; global 0.5° x 0.5° resolution
- 1980-2100 (SSP1-2.6, SSP3-7.0, SSP5-8.5)
- Focus here on 1980-2010 period

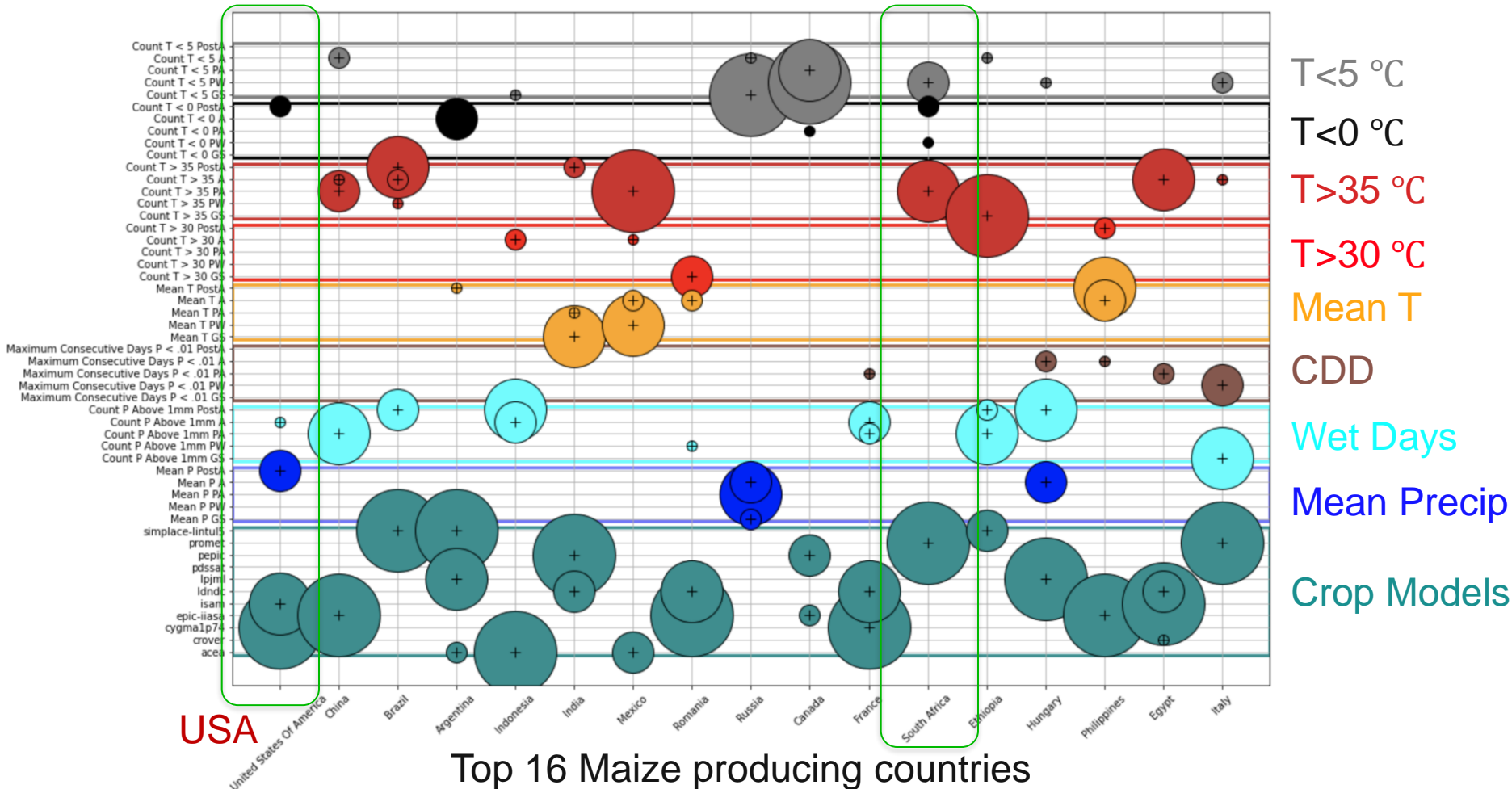
### FAOStat Country-Level Production

- Reports for top 20 producers for each crop species

### Extreme Gradient Boosting (XGBoost) Machine Learning model

- 5 features targeting yield outputs independently at each grid cell
- 5-fold cross-validation
- 40 climate variables (8 variables x 5 seasonal timings)
- 10 GGCMI aggregated to country level

## 5 feature model fit to FAO yields – size of circle represents prominence of features





## Models need enhanced cold and flood hazard responses

### Agricultural decision support

- Seasonal detection and attribution
- Resilience planning and disaster risk reduction
- Suitability analysis for management and breeding approaches
- Climate change adaptation and risk management

## Agricultural Model Intercomparison and Improvement Project (AgMIP) Machine Learning Team

- AgML launching soon!

### Machine Learning Applications

- Determination of model parameters
- Crop model emulators
- Model diagnostics and improvement prioritization
- Hybrid model development and applications
- Combine with remote sensing and additional predictors

Join us at AgMIP Town Hall  
Friday to learn more

**SAVE THE DATE!**

**9<sup>th</sup> AgMIP Global Workshop**

**June 26-30, 2023**

**Columbia University, New York, NY**

Learn more at [www.agmip.org](http://www.agmip.org)

A scenic landscape photograph of a valley. The foreground is dominated by lush green fields, likely cornfields, with some taller plants in the immediate foreground. In the middle ground, a small town or village is visible, featuring several buildings with red roofs and a prominent white steeple. The background consists of rolling green hills and mountains under a sky filled with large, white, fluffy clouds. The overall scene is bright and vibrant, suggesting a rural or agricultural setting.

**Thanks for your  
attention!**

[alexander.c.ruane@nasa.gov](mailto:alexander.c.ruane@nasa.gov)

[www.agmip.org](http://www.agmip.org)

# Model ensemble performance and climate information influence

## a) Maize production correlations with FAO

