Food Security

Predicting food shortages

Assignments

Keep working on your projects!

Project work day will be April 27

Discussion question: What is the most challenging aspect of your project right now?

Due Friday 5pm

Climate change in the news

Climate change in the news

https://grist.org

Warming temperatures trigger earliest spring on record in parts of eastern US

Unseasonably early blooms can wreak havoc on allergies, disease vectors, and agriculture.

Aleksandr Kondratov/Getty Images

Parts of central Texas and the Louisiana coast, southern Arkansas, southern Ohio, the D.C. area, *New York City, and the New Jersey coastline all clocked their earliest spring on record,* said Theresa Crimmins, director of the National Phenology Network

Phenologists — people who study biological life cycles — use two metrics to delineate the change in seasons: **First bloom, when plants begin to flower, and first leaf-out, when leaves unfurl**. This year, first bloom and first leaf-out started creeping up the East Coast *between three and four weeks ahead of schedule*. That's not entirely unusual; natural variation in seasons results in an early spring every few years. But, in some places, spring arrived extremely early — *earlier than any time in the past four decades.*

Earlier springs are associated with a host of problems for human health. Recent research shows that the lengthening growing season has led to an *allergy season that is 21 percent more intense and 20 days longer,* on average, in North America. Shortened winters allow insects that carry disease, such as ticks and mosquitos, to get active earlier and *spread pathogens to other animals and humans*.

And early spring is a *nightmare for farmers* across the country who are already struggling to adapt to rapidly shifting environmental conditions. Mississippi's blueberry crop was imperiled a couple of weeks ago when a hard frost descended on the state after a spate of abnormally warm days caused blueberry bushes to bloom early. One farmer in the state estimated that t*he frost wrecked 80 percent of his crop.*

How climate change has already impacted food production

In 2022...

- In Texas, the cotton harvest was hit hard by drought.
- Hurricane Ian blew oranges off the trees in Florida.
- Rice farmers in California left half their fields empty for lack of water
- Unseasonable freezes and drought left almond trees without harvests
- Cattle ranchers sent more cows to slaughter because drought-stunted pastures can't support calving, raising prices for years to come

https://www.usatoday.com/story/news/nation/2022/12/07/climate-change-effects-hit-farmers-us-rice-citrus-almond-crops/82584

Food security and climate change

Predicted impacts on food security

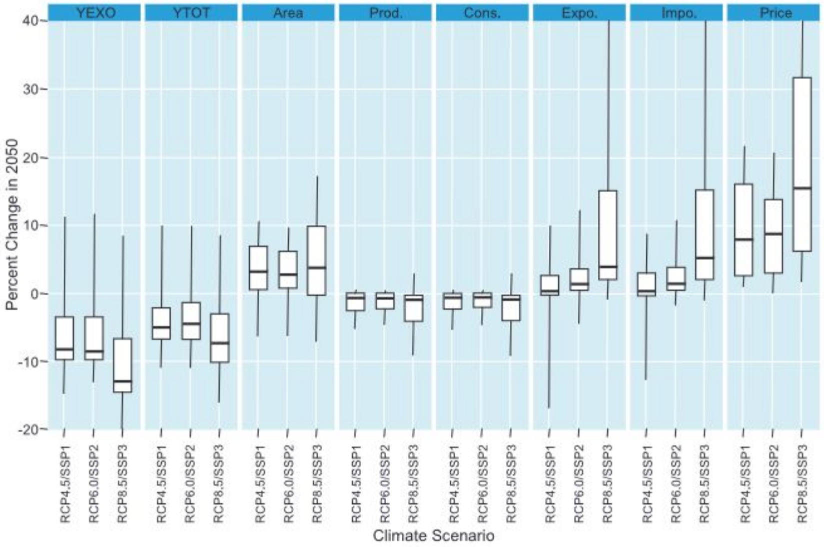


Figure ES-4. Climate-change effects on agricultural commodities in 2050 under different SSPs and RCPs. *The more pessimistic "high concentration/low international cooperation" scenario (RCP8.5/SSP3) shows much larger and more variable climate-change effects for the five commodities (coarse grains, rice, wheat, oilseeds and sugar), than the "medium concentration/middle of the road" (RCP6.0/SSP2) and "low concentration/sustainable development" (RCP4.5/SSP1) scenarios. All are compared to baseline of SSPs with no climate change. Results are from three GCMs and five economic models, aggregated across thirteen regions (n = 75). YEXO = yield effect of climate change without technical or economic adaptation, YTOT = realized yields after adaptation, AREA = agricultural area in production, PROD = total production, CONS = consumption, Expo = exports, IMPO = imports, PRICE = prices.*

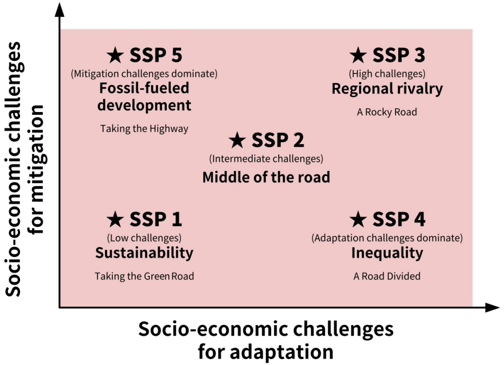
USDA

https://www.usda.gov/sites/default/files/documents/FullAssessment.pdf

Predicted impacts on food security

Figure ES- 5. Relative risks to food availability for different SSPs. *The risks to food availability would be lowest under the economic conditions described in in SSP 1 and SSP 5 for a given scenario of climate change, with poorer nations being at higher risk across all food production, distribution and trade categories for all SSPs. Shading represents higher or lower risks for each SSP from climate change. Risks reflect the informed judgment of the authors of this report based on the available literature.*

Not just an issue of agriculture



Foods at particular risk

Coffee - rising temperatures are estimated to reduce the suitable coffee-growing land by 50% by 2050 and support the growth of coffee rust fungus

Chocolate - yield will decline as soon as 2030. Growing regions will shift to South America

Shellfish - at risk from hotter and more acidic oceans

Corn - yield for corn grain will decrease within the U.S. Corn Belt by 20 to 40% from 1991-2000 levels by 2046-2055 with large economic impact

Livestock - impacted by more heat stress, disease, and water availability

Rice - increase and severity of hot weather can cause yields to decrease by 40% within this century

Honey - Shifts in the flowering plant cycles can cause nutritional stress on bee populations.

https://www.tastingtable.com/1016241/how-climate-change-is-impacting-these-foods/

Impacts on health

Impact of climate change on food production could cause over 500,000 extra deaths in 2050

03 March 2016

https://www.oxfordmartin.ox.ac.uk/news/201601-climate-food-production/

- Lead to average per-person reductions in food availability of 3.2% (99 kcal per day), in fruit and vegetable intake of 4.0% (14.9g per day), and red meat consumption of 0.7% (0.5g per day).
- Almost three-quarters of all climate-related deaths expected to occur in China (248,000) and India (136,000).

Paper Deep Dive

Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data

Jiaxuan You, Xiaocheng Li, Melvin Low, David Lobell, Stefano Ermon
Department of Computer Science, Stanford University
{jiaxuan, mwlow, ermon}@cs.stanford.edu
Department of Management Science and Engineering, Stanford University
chengli1@stanford.edu
Department of Earth System Science, Stanford University
dlobell@stanford.edu

Motivation

It is estimated that 795 million people still live without an adequate food supply (FAO 2015), and that by 2050 there will be *two billion more people* to feed (Dodds and Bartram, 2016). Ending hunger and improving food security are primary goals in the 2030 Agenda for Sustainable Development of the United Nations (United Nations 2015).

A central challenge of addressing food security issues is yield estimation, namely being able to accurately **predict crop yields well before harvest**. Agricultural monitoring, ***especially in developing countries, can improve food production and support humanitarian efforts*** in light of climate change and droughts (Dodds and Bartram 2016)

Background

Existing approaches rely on ***survey data and other variables related to crop growth*** (such as weather and soil properties) to model crop yield. These approaches are ***very successful in the United States***, where data are plentiful and of relatively high quality. Comprehensive surveys of weather parameters such as the Daymet (Thornton et al. 2014) and land cover types such as the Cropland Data Layer (Boryan et al. 2011) are publicly available and greatly facilitate the crop yield prediction task.

However, information about weather, soil properties, and precise land cover data are ***typically not available in developing countries***, where reliable yield predictions are most needed

Goal: be able to predict crop yields in countries without extensive agricultural data/monitoring

Brainstorm

What kind of data would you want to have to be able to approach this problem?

What kind of methods would you apply?

How would you measure success?

What difficulties might you face?

Framing the problem

We consider the problem of **predicting the average yield of a type of crop** for a region of interest based on a sequence of remotely sensed images taken before the harvest. Specifically, we are interested in the average yield per unit area in a given geographical region, e.g., a county or district. **As input, we are given a sequence of multispectral images**

Data

Focus on US soybean crops (commonly used)

MODIS satellite data: surface reflectance, land surface temperature, and land cover type (for masking) at 500m resolution

Images collected 30 times a year, from the 49th day to the 281th day at 8-days intervals.

Output data: yearly average soybean yields at the county-level measured in bushels per acre, publicly available on the USDA website. Data from 11 states from 2003 to 2015

Data Problem!

“Given the scarcity of labeled training data (less than 10,000 data points), directly training a deep model end-to-end is not feasible.”

Data Problem!

“Given the scarcity of labeled training data (less than 10,000 data points), directly training a deep model end-to-end is not feasible.”

“Pre-training on popular benchmarks from computer vision like Imagenet is also not appropriate, because remotely sensed images are multi-spectral and taken from a bird’s eye viewpoint.”

Data Problem!

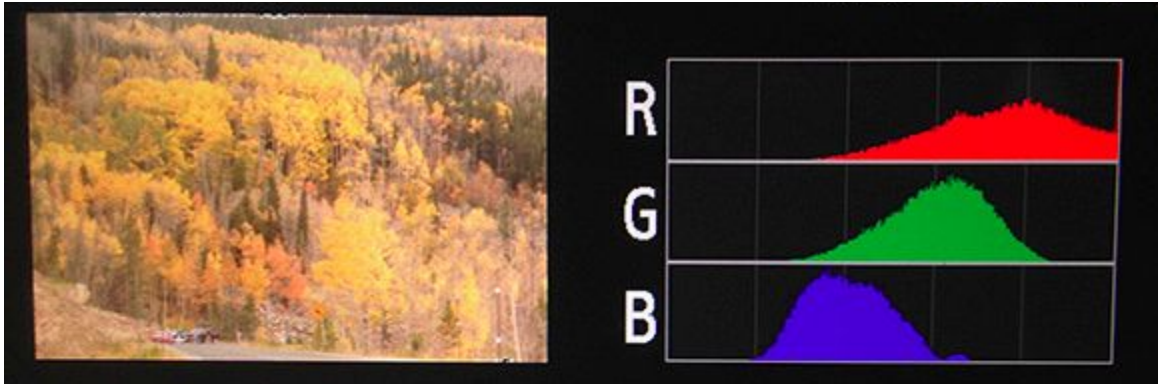
“Given the scarcity of labeled training data (less than 10,000), directly training a deep model end-to-end is not feasible.”

“We therefore designed a dimensionality reduction technique under the assumption of ***permutation invariance***. Our approach is based on the following intuition: we don’t expect the average yield to depend (much) on the position of the image pixels, since they merely indicate the locations of the cropland.”

Data Problem!

“Given the scarcity of labeled training data (less than 10,000), directly training a deep model end-to-end is not feasible.”

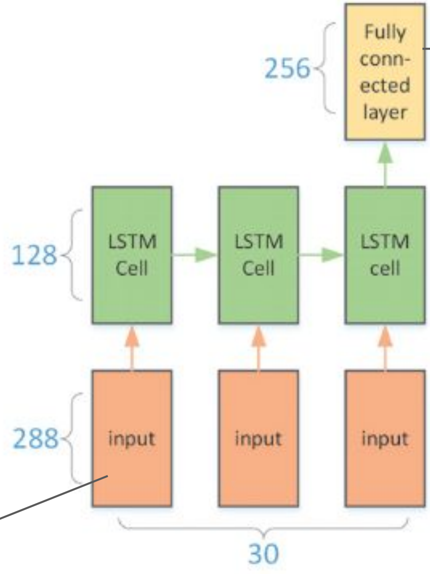
“Assuming permutation invariance holds, only the number of different pixel types in an image (pixel counts) are informative. In other words, there is no loss of information in mapping the high-dimensional image into a **histogram** of pixel counts”



32 bins times 9 spectral bands, at 30 time points

Model

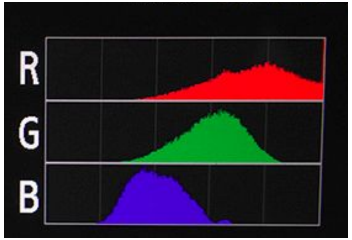
Model



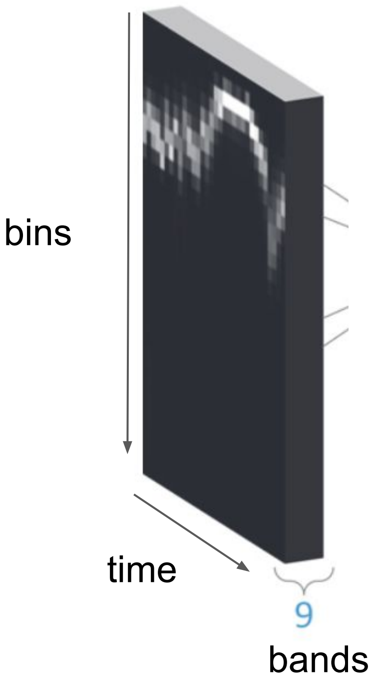
(b) The LSTM structure

predict crop yield

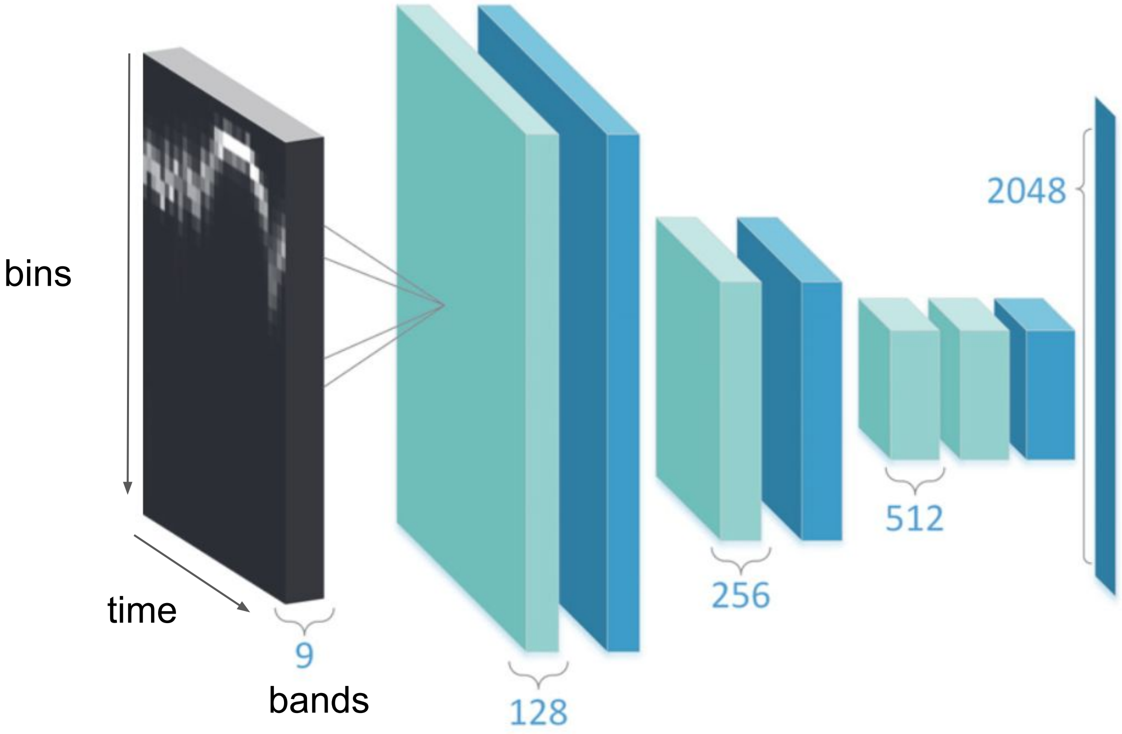
Each input is a histogram of the satellite image for one time point



Model



Model



(c) The CNN structure (no pooling)

Baseline Models

NDVI at each time point was given as input to:

Ridge regression

Decision trees

DNN with 3 hidden layers and 256 neurons each.

The hyperparameters in these models are optimized in cross-validation

All are evaluated with Root Mean Squared Error

Results

Results

	Baselines			Deep models	
Year	Ridge	Tree	DNN	LSTM	CNN
2011	9.00	7.98	9.97	5.83	5.76
2012	6.95	7.40	7.58	6.22	5.91
2013	7.31	8.13	9.20	6.39	**5.50**
2014	8.46	7.50	7.66	6.42	5.27
2015	8.10	7.64	7.19	6.47	6.40
Avg	7.96	7.73	8.32	6.27	5.77

Table 1: The RMSE of county-level model performance.

Model for each year is trained on all preceding years

Model Problem!

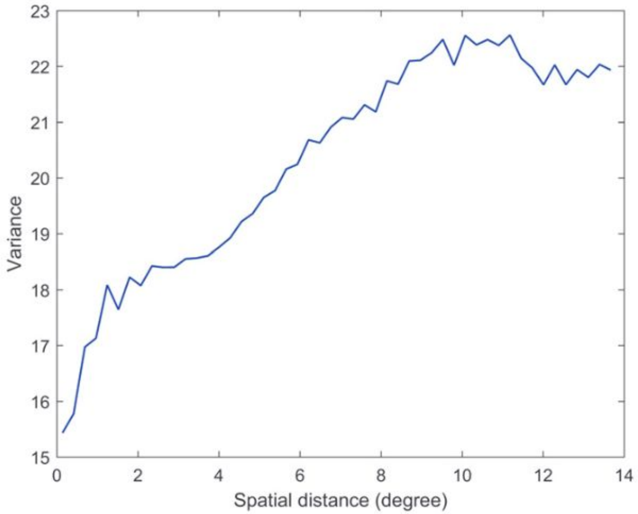
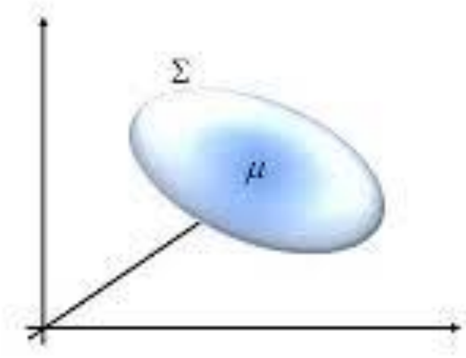


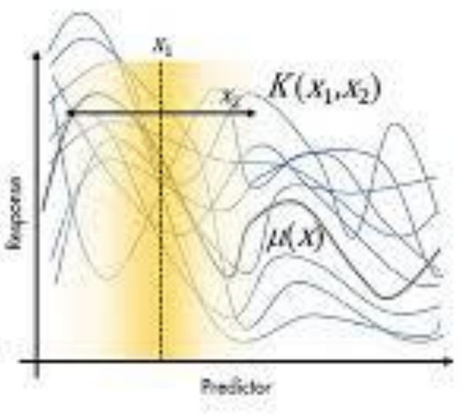
Figure 2: A variogram on the absolute prediction error of the proposed CNN model.

There are spatial-temporal correlations that these models don’t capture.

Gaussian Processes

A way to approximate a function (relationship) using probability distributions





Model

Add a correction to the output of the model based on a Gaussian Process

Output of full model

$$y(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}$$

Gaussian process

Output of deep neural network

$$f(\boldsymbol{x}) \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}')),$$

Model

Add a correction to the output of the model based on a Gaussian Process

Output of full model

$$y(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}$$

Gaussian process

Output of deep neural network

$$f(\boldsymbol{x}) \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}')),$$

Correlation of outputs based on location and year

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left[-\frac{\|\boldsymbol{g}_{\text{loc}} - \boldsymbol{g}'_{\text{loc}}\|_2^2}{2r_{\text{loc}}^2} - \frac{\|g_{\text{year}} - g'_{\text{year}}\|_2^2}{2r_{\text{year}}^2} \right]$$

Model

Add a correction to the output of the model based on a Gaussian Process

Output of full model

$$y(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{h}(\boldsymbol{x})^T \boldsymbol{\beta}$$

Gaussian process

Output of deep neural network

$$f(\boldsymbol{x}) \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}')),$$

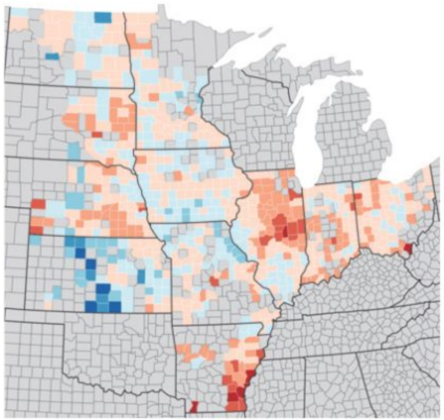
Correlation of outputs based on location and year

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp\left[-\frac{\|\boldsymbol{g}_{\text{loc}} - \boldsymbol{g}'_{\text{loc}}\|_2^2}{2r_{\text{loc}}^2} - \frac{\|g_{\text{year}} - g'_{\text{year}}\|_2^2}{2r_{\text{year}}^2} \right]$$

Parameters of GP are selected via hyperparameter optimization

Model

Add a correction to the output of the model based on a Gaussian Process



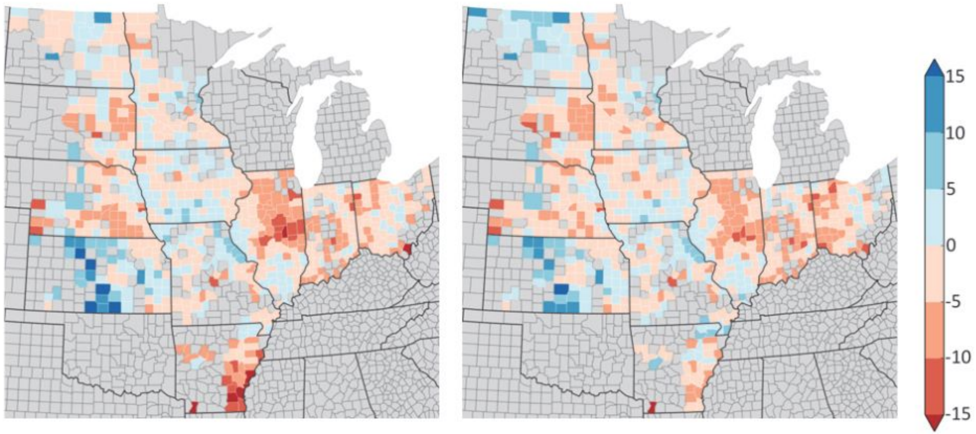
(a) CNN

“Intuitively, we believe the errors are
due to properties that are not observable in remote
sensing images (e.g., due to soil). The GP part
learns these patterns from past training data and
effectively corrects for them”

Errors in CNN model

Results

Add a correction to the output of the model based on a Gaussian Process



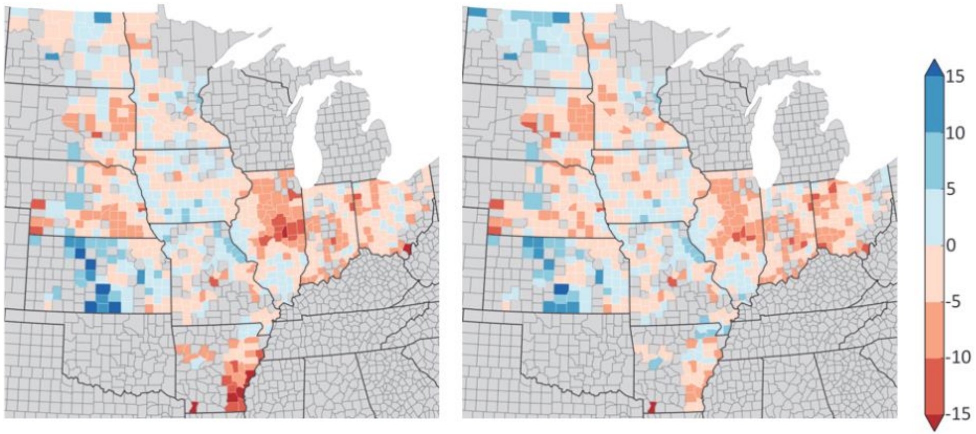
(a) CNN (b) CNN+GP

Figure 3: County-level error maps before and after adding the GP. The color represents the prediction error in bushel per acre.

Deep models			
LSTM	LSTM + GP	CNN	CNN + GP
5.83	5.77	5.76	**5.7**
6.22	6.23	5.91	**5.68**
6.39	5.96	**5.50**	5.83
6.42	5.70	5.27	**4.89**
6.47	**5.49**	6.40	5.67
6.27	5.83	5.77	**5.55**

Results

Add a correction to the output of the model based on a Gaussian Process



(a) CNN (b) CNN+GP

Figure 3: County-level error maps before and after adding the GP. The color represents the prediction error in bushel per acre.

Deep models			
LSTM	LSTM + GP	CNN	CNN + GP
5.83	5.77	5.76	**5.7**
6.22	6.23	5.91	**5.68**
6.39	5.96	**5.50**	5.83
6.42	5.70	5.27	**4.89**
6.47	**5.49**	6.40	5.67
6.27	5.83	5.77	**5.55**

Uncertainty??

Results

Predicting earlier, based on fewer months

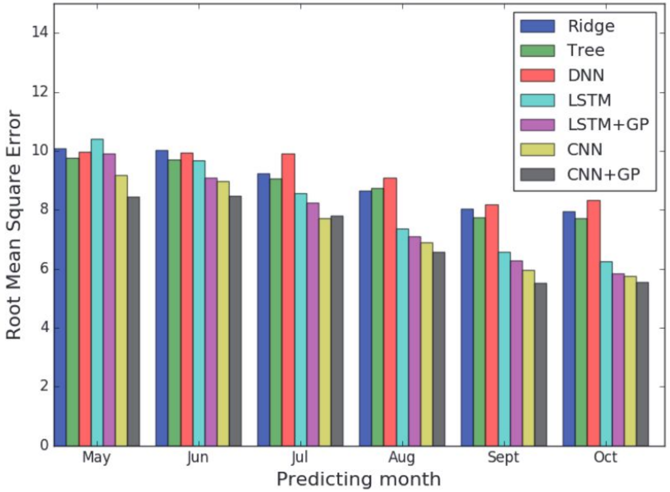


Figure 4: Model performance in each month measured in RMSE. The results are averaged from 2011 to 2015.

	July	August		September		October	
	Ours	USDA	Ours	USDA	Ours	USDA	Ours
MAPE	5.65	3.92	**3.37**	4.14	**3.41**	**2.48**	3.19

Table 2: The MAPE of US-level model performance, averaged from 2009 to 2015.

USDA uses surveys to predict yield

Results

Feature Importance

Results

Feature Importance: Training with permuted data

Traditionally, band 2 as a near infrared band, is viewed as a key factor in revealing crop growth but our model also focuses on band 7, a short-wave infrared band always ignored by traditional approaches.

The high dependence on land surface temperature, shown as band 8 and 9, is also confirmed by previous work.

Growth-related bands 2 and 7 are given higher relative importance in later months (when crop has grown), while temperature bands 8 and 9 are more significant in earlier months (when it has not yet).

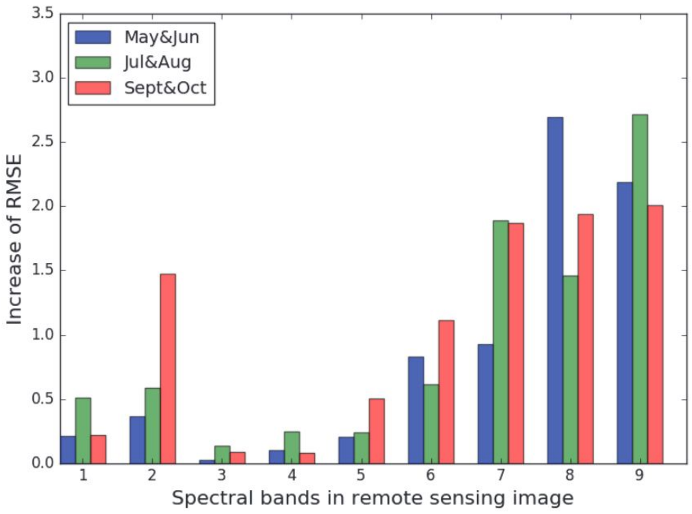
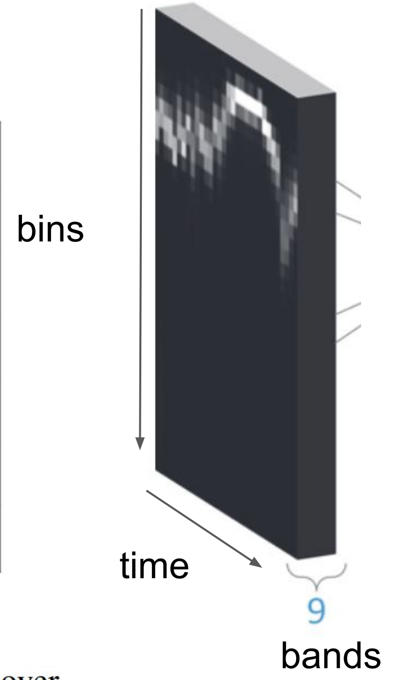


Figure 5: The increase of RMSE after permutation over bands. We evaluate predictions made in different months.



Results

Feature Importance: Training with permuted data

Surveys show that soybean planting usually starts on day 110 and ends on day 190, while harvest usually starts on day 250.

The most useful data are collected during the growing season, peaking at days just before the harvest.

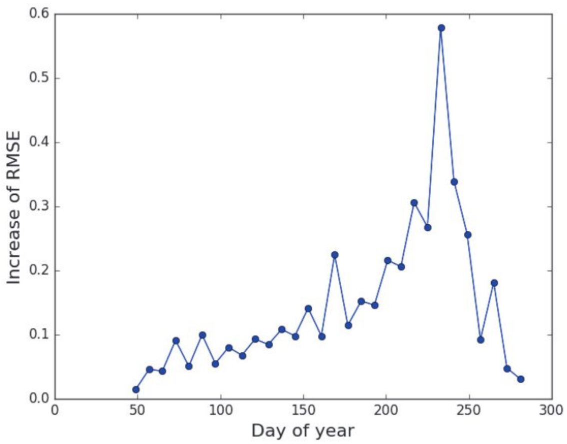
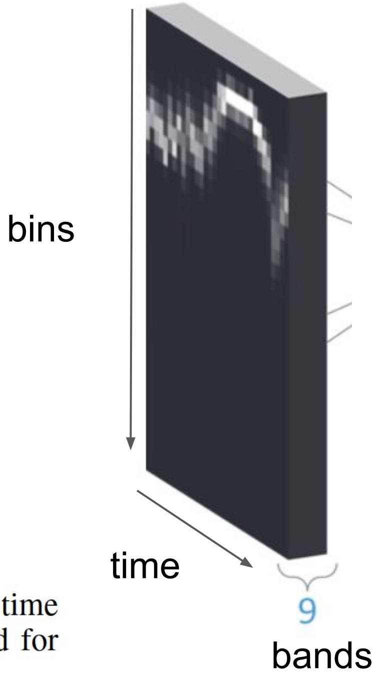


Figure 6: The increase of RMSE after permutation over time within a year. The model with complete data are used for evaluation.



Further Resources

Explainer:
https://www.worldbank.org/en/news/feature/2022/10/17/what-you-need-to-know-about-food-security-and-climate-change

All of the impacts of climate change on agriculture:
https://news.climate.columbia.edu/2018/07/25/climate-change-food-agriculture/

Visual explainer of Gaussian Processes:
https://distill.pub/2019/visual-exploration-gaussian-processes/

Summary

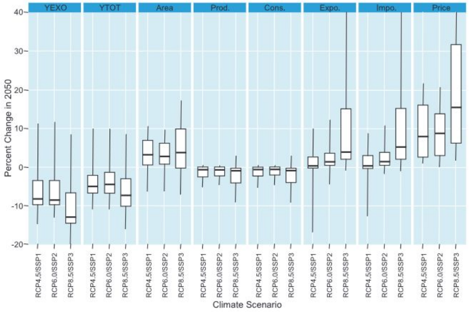
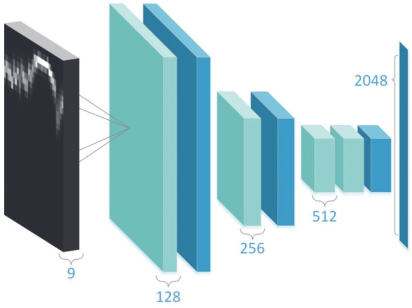


Figure ES-4. Climate-change effects on agricultural commodities in 2050 under different SSPs and RCPs. *The more pessimistic "high concentration/low international cooperation" scenario (RCP8.5/SSP3) shows much larger and more variable climate-change effects for the five commodities (coarse grains, rice, wheat, oilseeds and sugar), than the "medium concentration/middle of the road" (RCP6.0/SSP2) and "low concentration/sustainable development" (RCP4.5/SSP1) scenarios. All are compared to baseline of SSPs with no climate change. Results are from three GCMs and five economic models, aggregated across thirteen regions (n = 75). YEXO = yield effect of climate change without technical or economic adaptation, YTOT = realized yields after adaptation, AREA = agricultural area in production, PROD = total production, CONS = consumption, Expo = exports, IMPO = imports, PRICE = prices.*



(c) The CNN structure