Reminders

- PRA2 due 2/8
- Course project progress report 1 due 2/27
- Come to OH for course project discussion!

Artificial Intelligence Methods for Social Good

Lecture 7

Case Study: Estimate Crop Yield from Remote Sensing Data

17-537 (9-unit) and 17-737 (12-unit)

Instructor: Fei Fang

feifang@cmu.edu

Outline

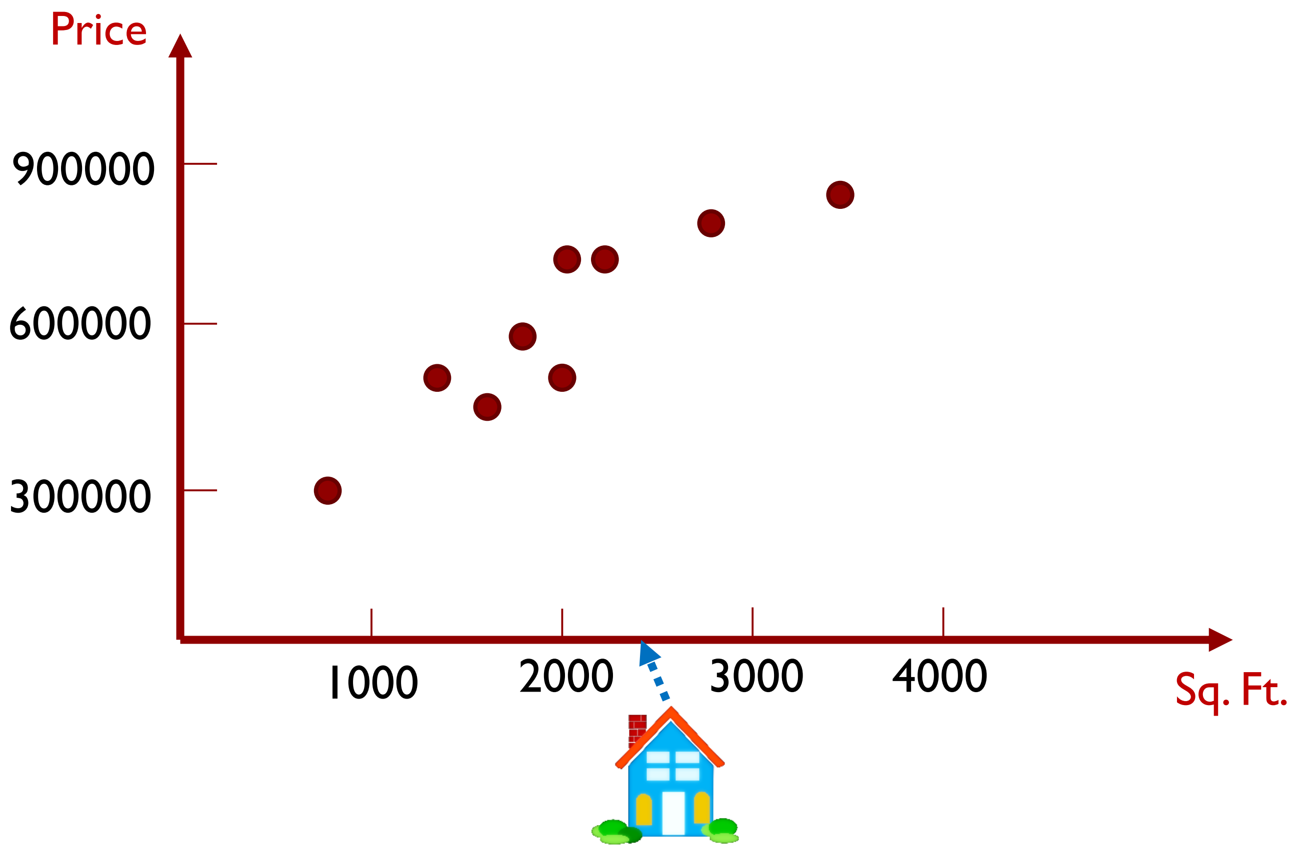
- Gaussian Process Regression
- Estimate Crop Yield
- Discussion

Learning Objectives

- Describe the following concepts
  - K-NN Regression, Gaussian Process, GP Regression
- For the crop yield estimation problem, briefly describe
  - Significance/Motivation
  - Task being tackled, i.e., what is being predicted/estimated
  - Data usage, i.e., what data is used and how it is processed
  - Domain-specific considerations
  - Machine learning method used
  - Evaluation process and criteria

Recall: Regression Example

- Predict house price



Non-parametric regression

- If we don't want to represent the relationship between x and y using an explicit function with to-be-learned parameters, can we still make predictions?

- Simplest approach: Nearest neighbor regression

Nearest Neighbor Regression

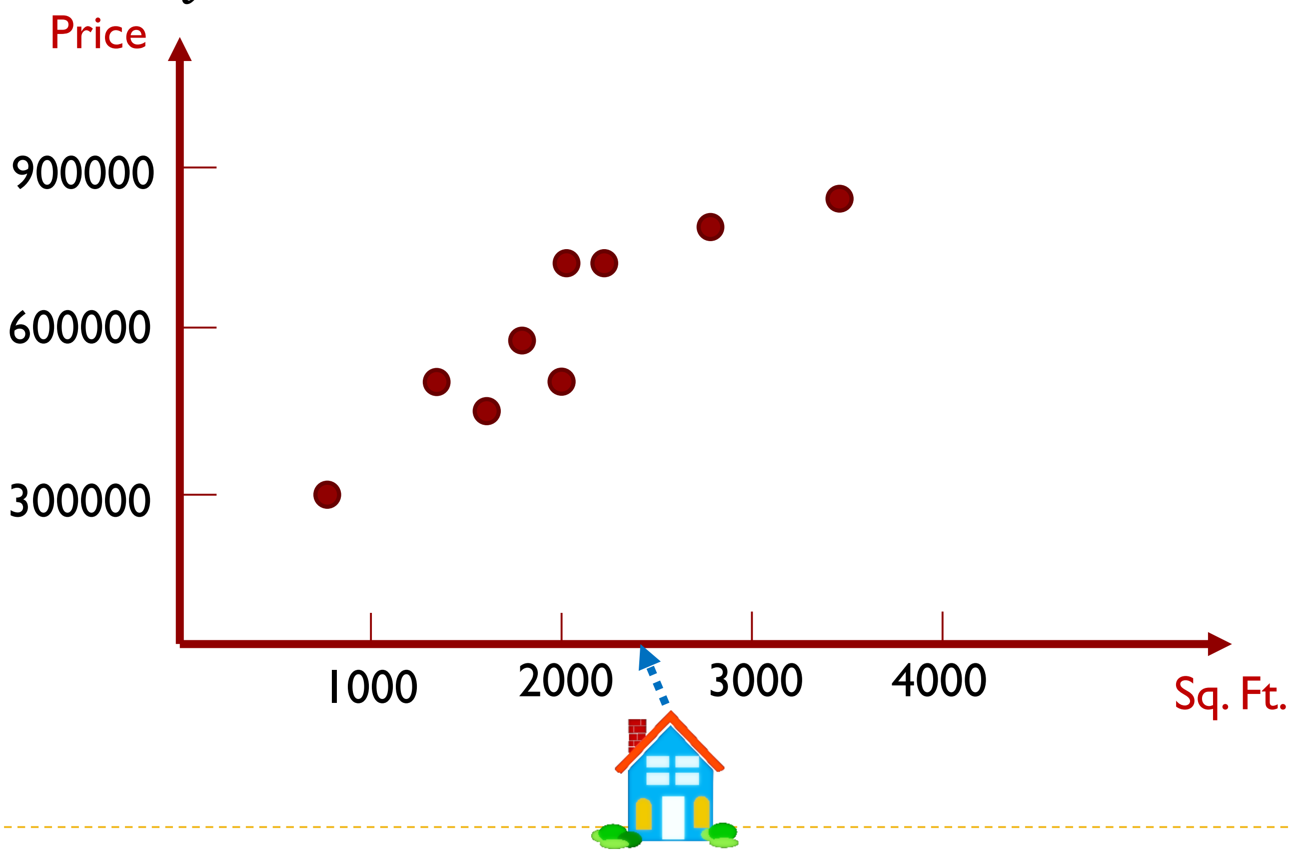
- 1-NN regression
  - Predicted value = value of "closest" point in training data
  - Distance metric: Manhattan:
    - $\text{Dist}(x_i, x_q) = \sum_k |x_i[k] - x_q[k]|$
    - (Scaled) Euclidean distance:

$$\text{Dist}(x_i, x_q) = \sqrt{\sum_k a_k \left(x_i[k] - x_q[k]\right)^2}$$

  - Limitation: poor performance in areas with little data; sensitive to noise

1-NN Regression

- Input: Training data $\{(x_i, y_i)\}$, query point x_q
- $i_{NN} = \underset{i}{\operatorname{argmin}}\, Dist(x_i, y_i)$, Output: $y_{i_{NN}}$



Nearest Neighbor Regression

- K-NN regression
  - Predicted value = average value of k "closest" points
  - Robust to noise
  - Limitation: poor performance in areas with little data or boundary; discontinuous predictions
  - Choose k: cross validation

Nearest Neighbor Regression

- K-NN regression
  - Predicted value = average value of k "closest" points
  - Robust to noise
  - Limitation: poor performance in areas with little data or boundary; discontinuous predictions
  - Choose k: cross validation

Find $(x_{NN_1}, x_{NN_2}, \ldots x_{NN_k})$ such that $\forall x_i \notin \{x_{NN_1}, \ldots x_{NN_k}\}$

distance $(x_i, x_q) \geq$ distance (x_{NN_k}, x_q)

Predict $\hat{y}_q = \frac{1}{k} \sum_{j=1}^{k} y_{NN_j}$

Poll 1

- Given the following training data, what is the price of Alice's house with house size = 1300 sqft through k-NN with $k = 3$
  - A: 220
  - B: 235
  - C: 213.3
  - D: 256.7
  - E: Neither of the above
  - F: I don't know

House size (sqft)	Sale price ($)
1200	220
1000	170
800	150
1500	250
1800	300

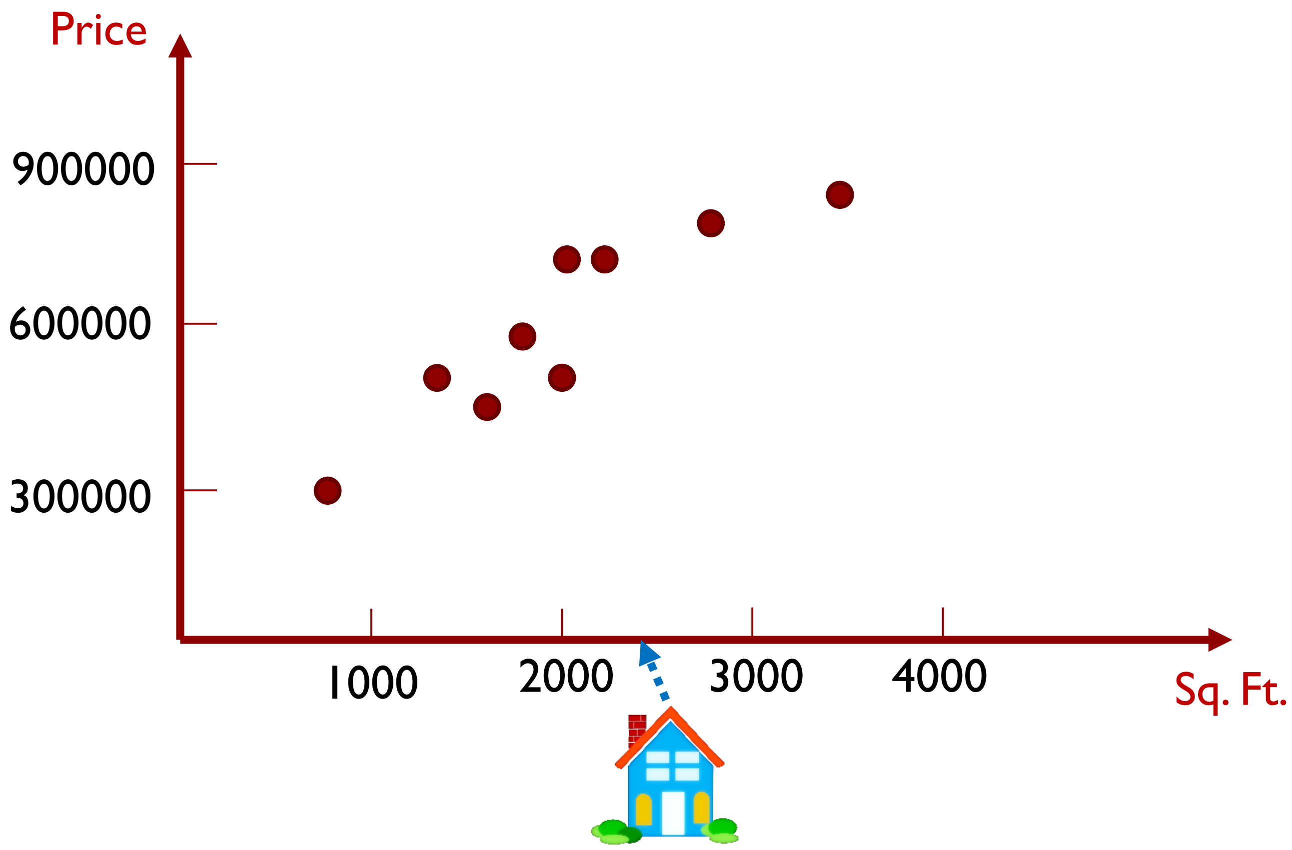
Kernel Regression

- Weighted K-NN regression
  - Predicted value = weighted average value of k "closest" point in training data
  - Smaller distance → Higher weight
    - E.g., weight=$\frac{1}{Dist(x_i,x_q)}$

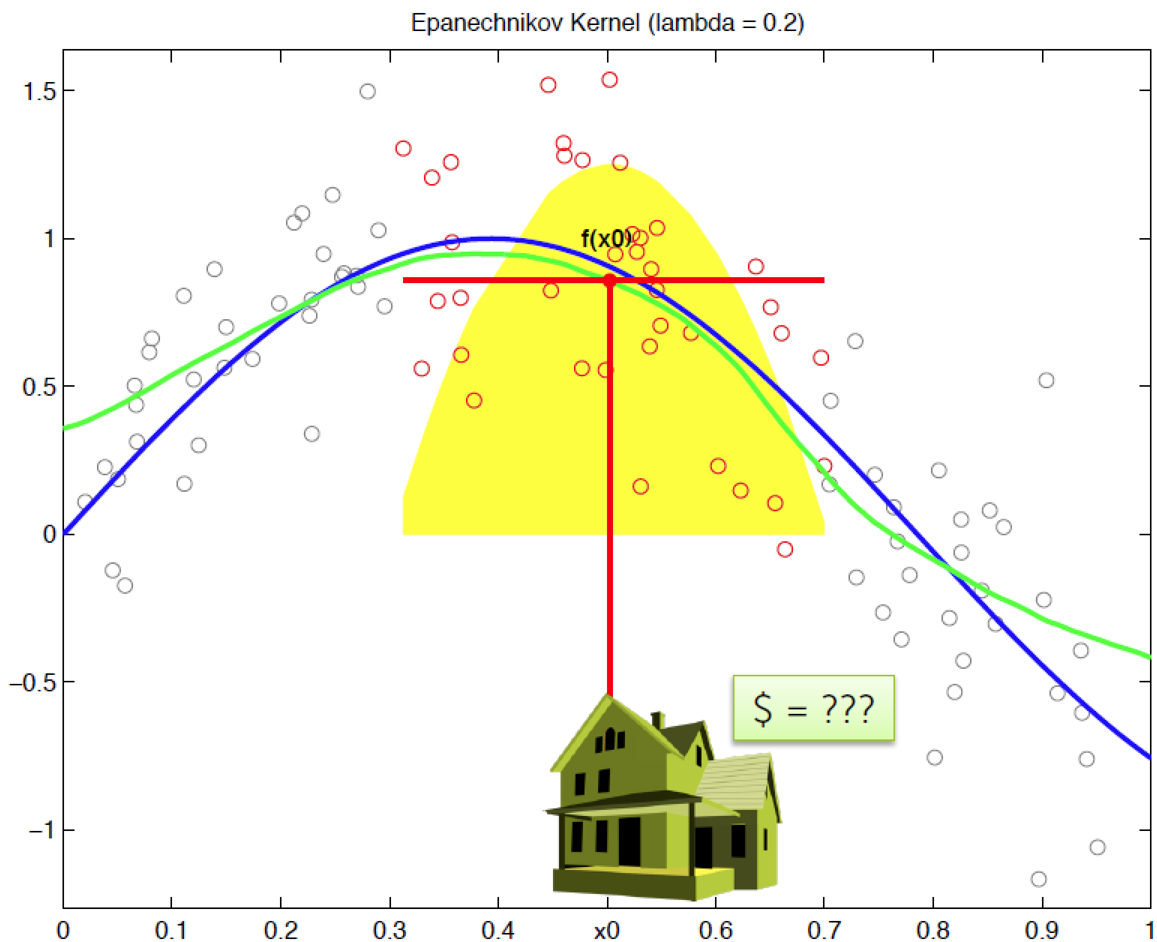
- Kernel regression
  - Predicted value = weighted average value of all points in training data
  - Weight is a function of distance: kernel
  - Example kernel: Gaussian, Triangle, Uniform
    - Gaussian: $kernel_\lambda(|x_i - x_q|) = e^{-\frac{|x_i - x_q|^2}{\lambda}}$

Kernel Regression

- Predict house price



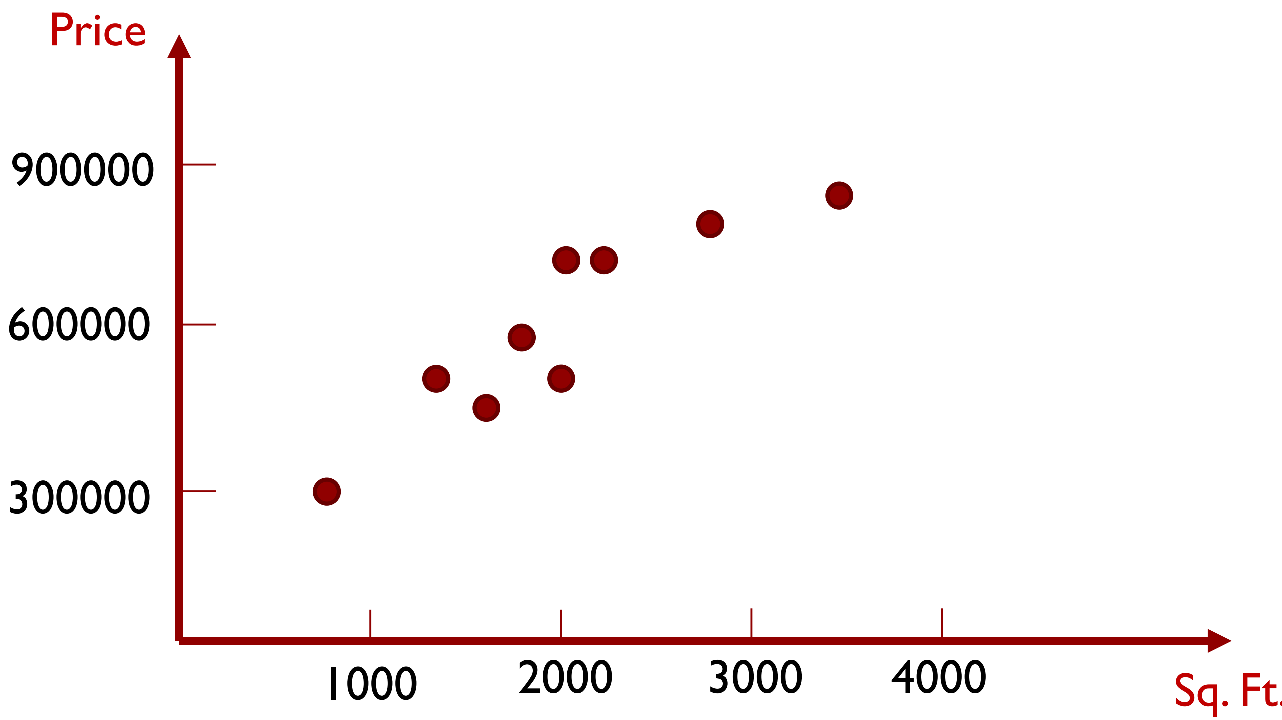
Kernel Regression



https://www.coursera.org/learn/ml-regression

Kernel Regression

- Kernel regression
  - Choose bandwidth parameter λ: cross validation
    - Small λ: Overfitting to nearest neighbors; Large λ: Smooth out



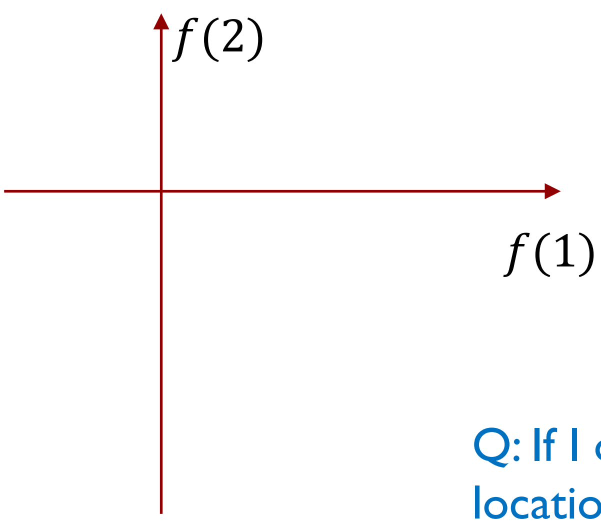
Gaussian Process

- GP is a stochastic process: a collection of random variables indexed by context $\boldsymbol{x}$: $\{f(\boldsymbol{x})\}_{x \in \mathcal{X}}$
  - $\boldsymbol{x}$ often represents context in time or space
  - E.g., temperature in different locations ($\mathcal{X}$ is the set of locations, $\boldsymbol{x}$ is a location, $f(\boldsymbol{x})$ is temperature at location $\boldsymbol{x}$)
- $\mathcal{X}$ can be an infinite set
- Every finite collection of those random variables has a multivariate normal distribution

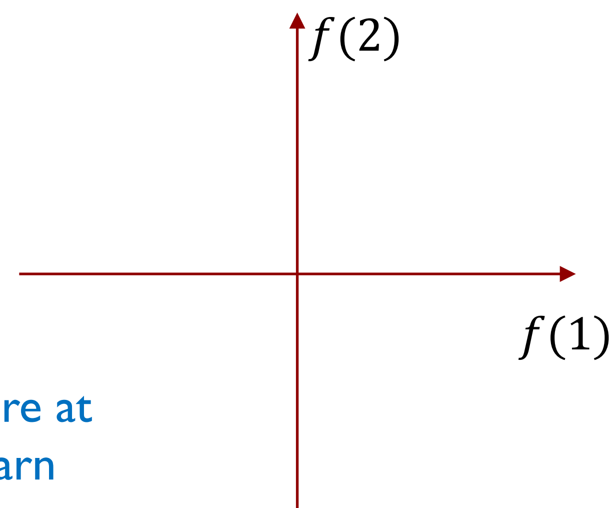
Example

- (Normalized) Temperature at two locations:

$$\begin{bmatrix} f(1) \\ f(2) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$



$$\begin{bmatrix} f(1) \\ f(2) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.6 \\ 0.6 & 2 \end{bmatrix}\right)$$

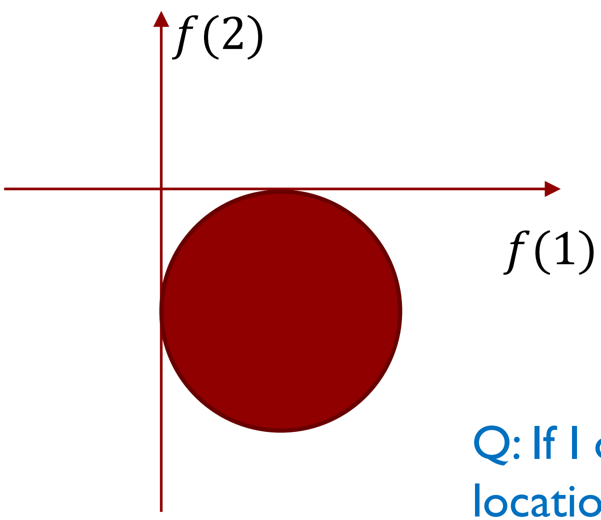


Q: If I observed temperature at location 1 is 0.5, can we learn anything about the temperature at location 2?

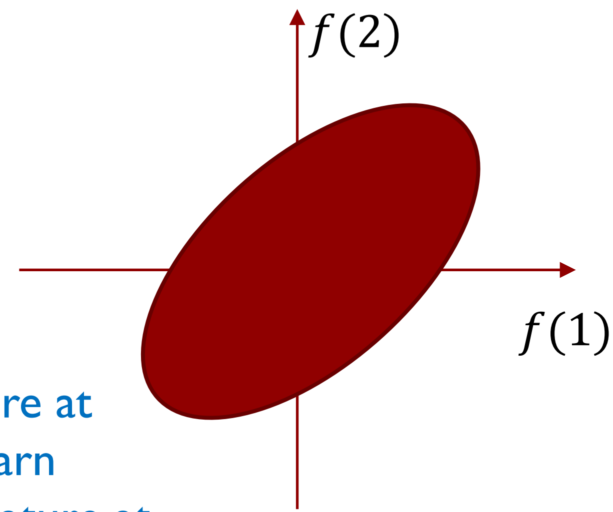
Example

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

$$\begin{bmatrix} f(1) \\ f(2) \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix})$$



$$\begin{bmatrix} f(1) \\ f(2) \end{bmatrix} \sim \mathcal{N}(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.6 \\ 0.6 & 2 \end{bmatrix})$$

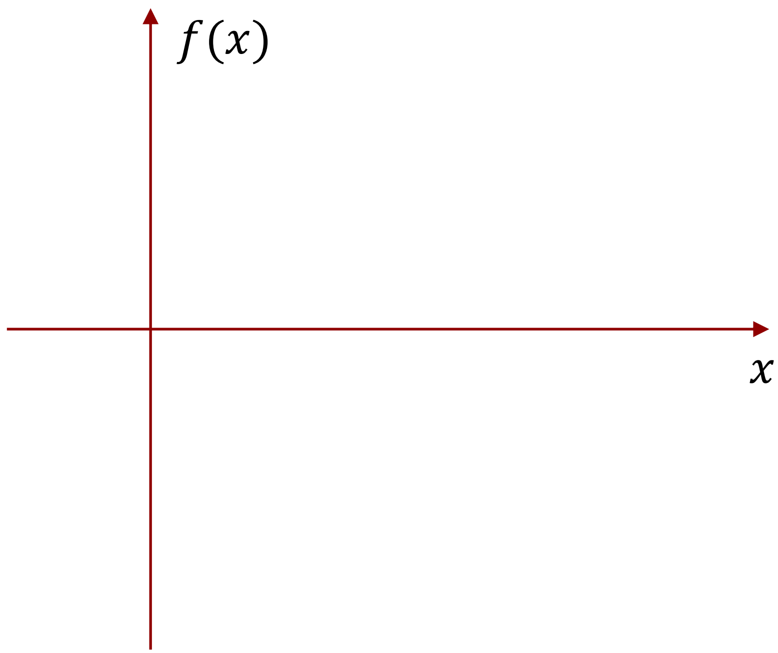


Q: If I observed temperature at location 1 is 0.5, can we learn anything about the temperature at location 2?

Example

- (Normalized) Temperature at three time points:

$$\begin{bmatrix} f(1) \\ f(2) \\ f(3.2) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 & 0.2 \\ 0.7 & 1 & 0.6 \\ 0.2 & 0.6 & 1 \end{bmatrix}\right)$$



There are infinite time points and the temperature at all these time points are correlated, how to represent the correlation?

Gaussian Process

- A Gaussian process can be defined as
$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$$
- $m(\boldsymbol{x})$ is the expectation of random variable $f(\boldsymbol{x})$
- Kernel function $k(\boldsymbol{x}, \boldsymbol{x}')$ defines covariance $\mathrm{cov}(f(\boldsymbol{x}), f(\boldsymbol{x}'))$
  - E.g., Radial basis function (RBF) kernel

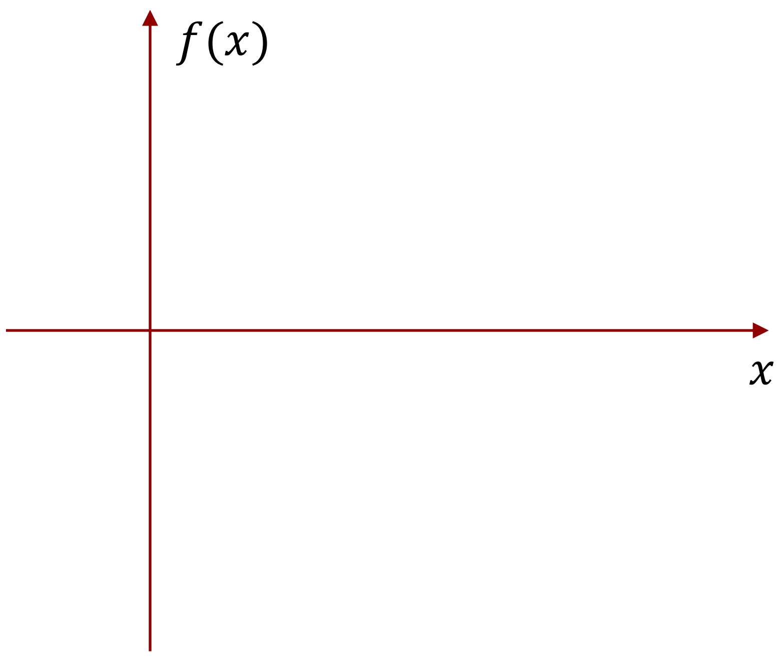
$$k_{RBF}(\boldsymbol{x}, \boldsymbol{x}') = \sigma^2 \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2r^2})$$

- Marginal and conditional prob. distributions are also Gaussian

Example

- (Normalized) Temperature at all time points:

$$m(\boldsymbol{x}) = 0, k_{RBF}(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_2^2}{2})$$



GP is a distribution over functions!

Training data: $(\boldsymbol{x}_i, y_i), i = 1..N$
Can you learn anything about the temperature at a new time point $\boldsymbol{x}^*$?

Yes! Conditional prob. dist. is Gaussian!

Can compute the posterior mean and variance of $f(\boldsymbol{x}^*)$

Gaussian Process Regression

- Given training data $\{(\boldsymbol{x}_i, y_i \sim f(\boldsymbol{x}_i)), i = 1..N\}$, kernel function k, test data $\boldsymbol{x}^*$, predict mean and variant of $f(\boldsymbol{x}^*)$ conditioned on the value of the training data

input: X (inputs), $\mathbf{y}$ (targets), k (covariance function), σ_n^2 (noise level), $\mathbf{x}_*$ (test input)

$L := \text{cholesky}(K + \sigma_n^2 I)$

$\boldsymbol{\alpha} := L^\top \backslash (L \backslash \mathbf{y})$ } predictive mean eq. (2.25)

$\bar{f}_* := \mathbf{k}_*^\top \boldsymbol{\alpha}$

$\mathbf{v} := L \backslash \mathbf{k}_*$ } predictive variance eq. (2.26)

$\mathbb{V}[f_*] := k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{v}^\top \mathbf{v}$

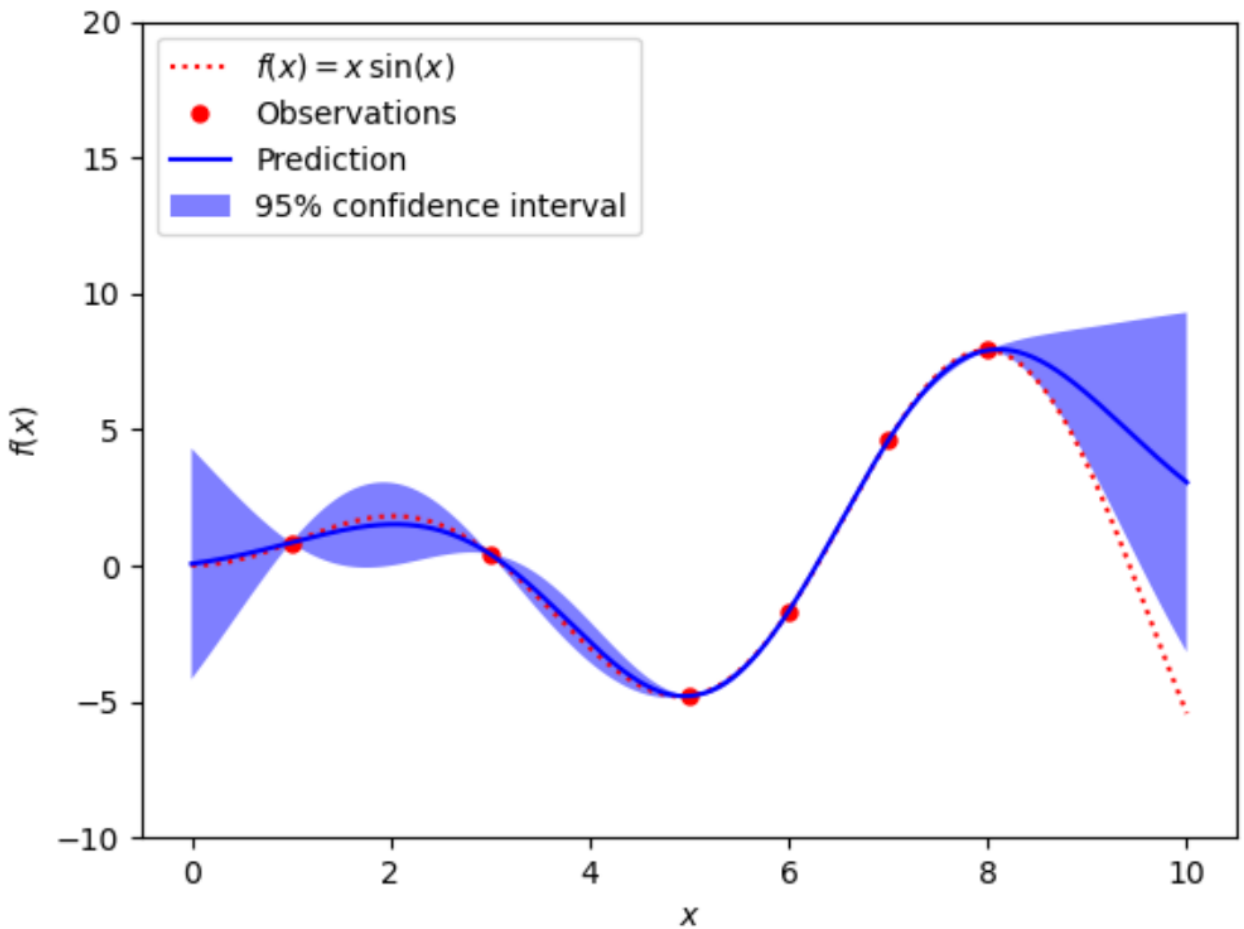
$\log p(\mathbf{y}|X) := -\frac{1}{2}\mathbf{y}^\top \boldsymbol{\alpha} - \sum_i \log L_{ii} - \frac{n}{2} \log 2\pi$ eq. (2.30)

return: $\bar{f}_*$ (mean), $\mathbb{V}[f_*]$ (variance), $\log p(\mathbf{y}|X)$ (log marginal likelihood)

GPR in scikit-learn implements this algorithm (Alg. 2.1 in reference [4])

GP Regression in Practice

- Existing code packages, e.g., scikit-learn



Use GaussianProcessRegressor in sklearn.gaussian_process

https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_noisy_targets.html

GP Regression in Practice

```python
import numpy as np
from matplotlib import pyplot as plt

from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import RBF, ConstantKernel as C

np.random.seed(1)


def f(x):
    """The function to predict."""
    return x * np.sin(x)
```

```python
X = np.atleast_2d([1., 3., 5., 6., 7., 8.]).T

# Observations
y = f(X).ravel()

# Mesh the input space for evaluations of the real function, the prediction and
# its MSE
x = np.atleast_2d(np.linspace(0, 10, 1000)).T

# Instantiate a Gaussian Process model
kernel = C(1.0, (1e-3, 1e3)) * RBF(10, (1e-2, 1e2))
gp = GaussianProcessRegressor(kernel=kernel, n_restarts_optimizer=9)

# Fit to data using Maximum Likelihood Estimation of the parameters
gp.fit(X, y)

# Make the prediction on the meshed x-axis (ask for MSE as well)
y_pred, sigma = gp.predict(x, return_std=True)
```


https://scikit-learn.org/stable/auto_examples/gaussian_process/plot_gpr_noisy_targets.html

Outline

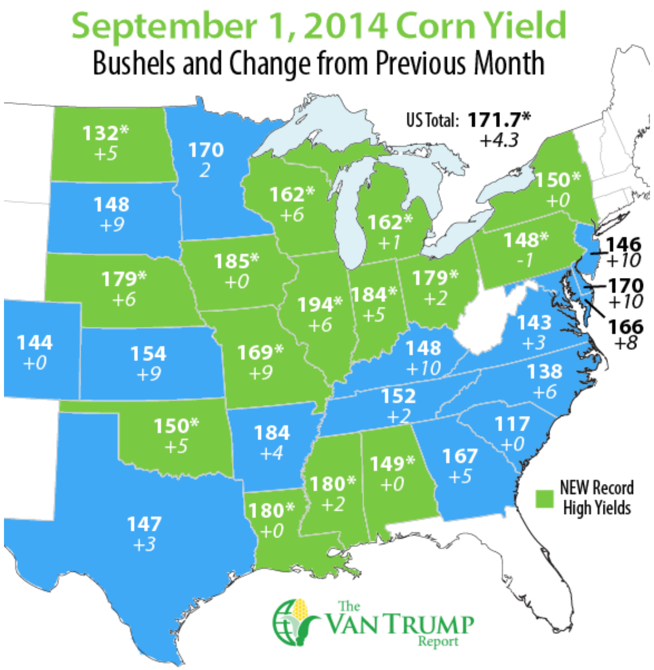
- Gaussian Process Regression
- Estimate Crop Yield
- Discussion

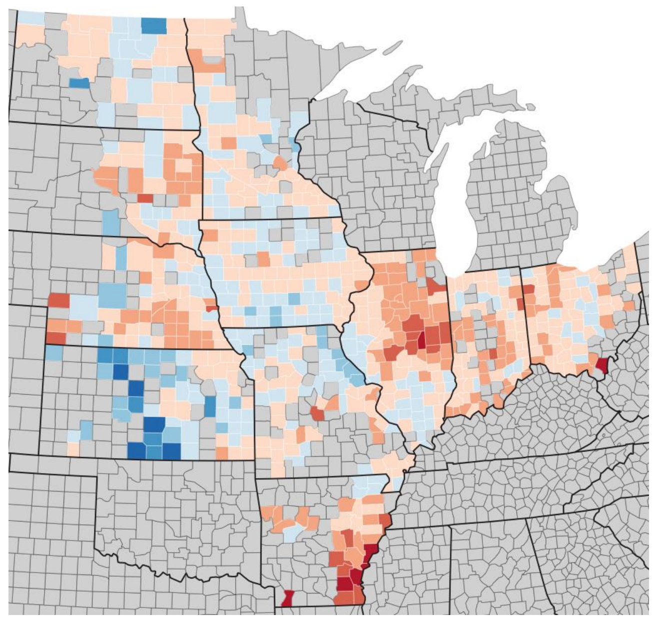
Agricultural Monitoring and Forecasting

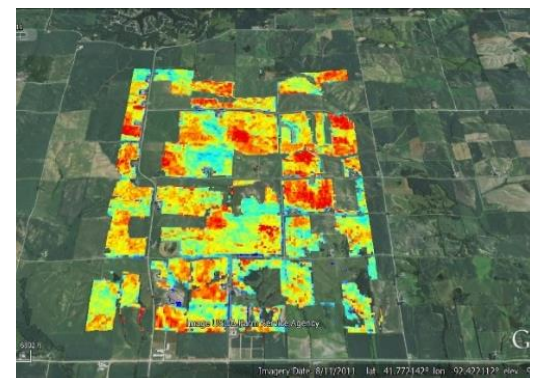
- Agricultural monitoring and forecasting:
  - Forecast production and demand
  - Real-time monitoring of food prices
  - Climate change effects

- Increase productivity with information technologies

Crop Yield Prediction







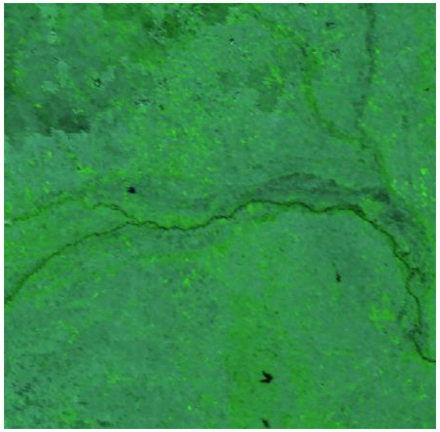
State level

County level

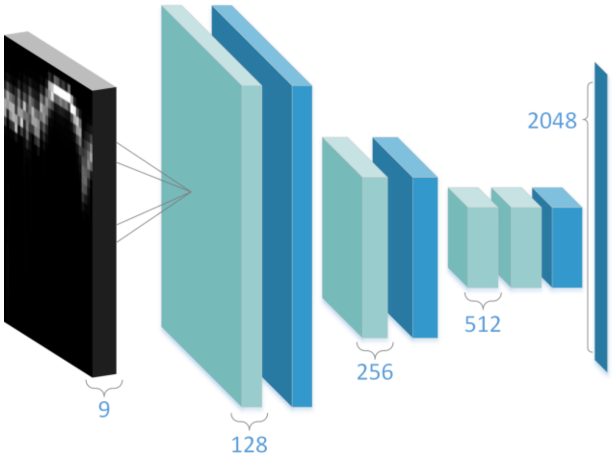
Field level

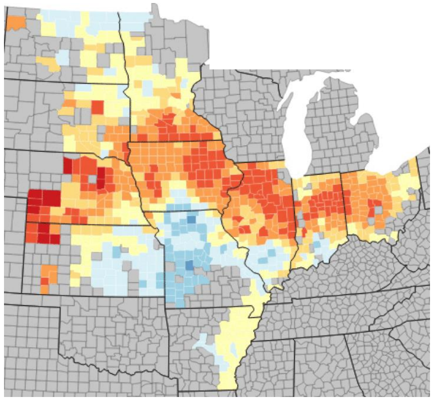
Crop Yield Prediction

Input:
Remote Sensing Data



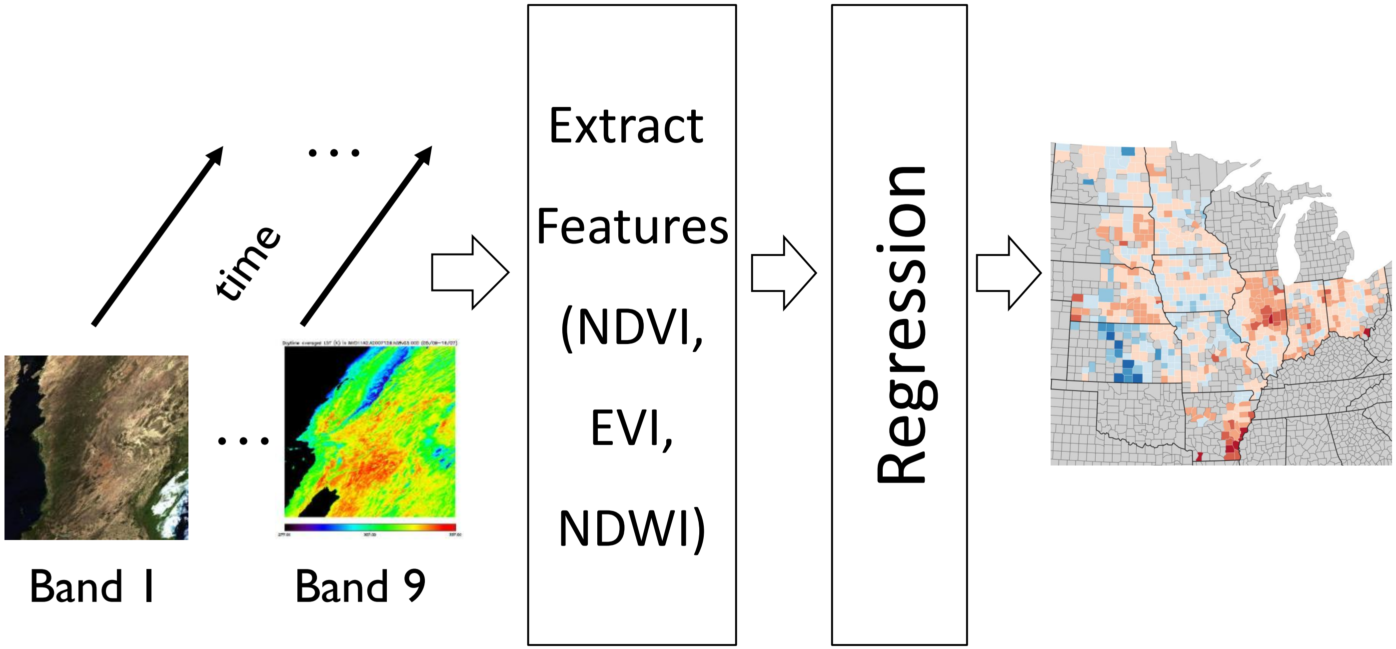
Output:
Crop yield





Deep Gaussian Process

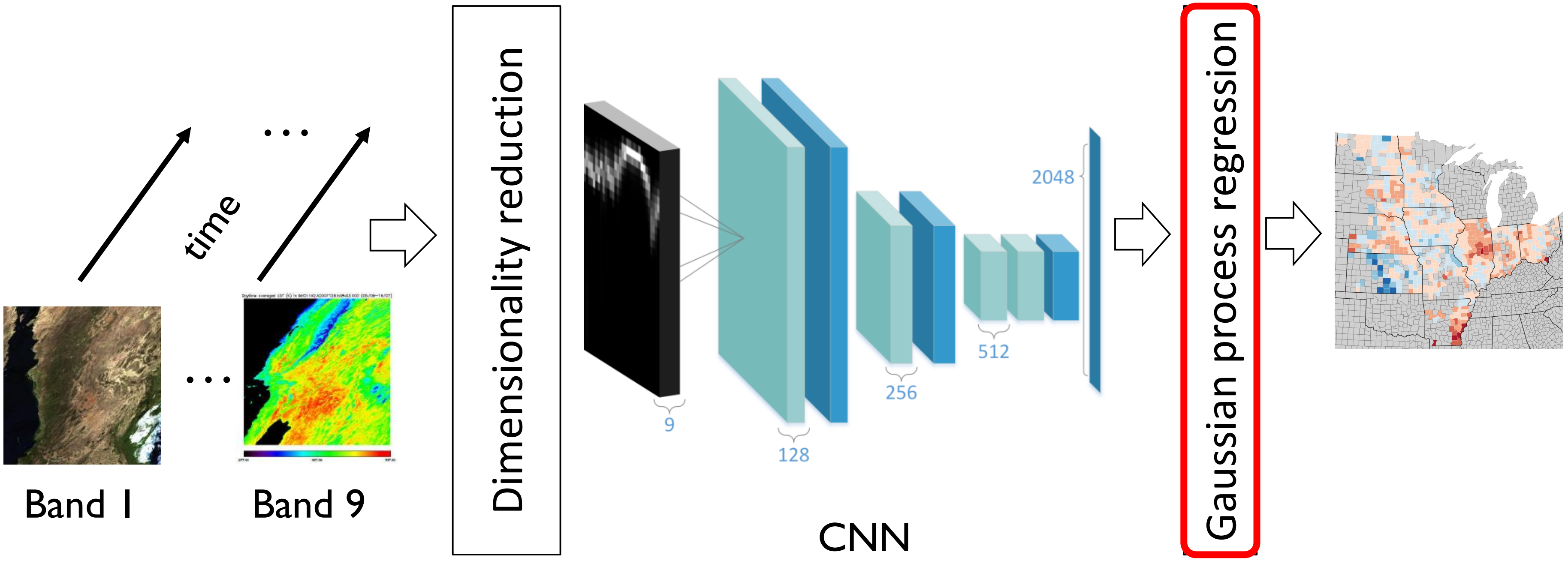
Existing Approaches



Hand crafted features

Spatial structure not modeled

New Approach - Deep Gaussian Process



Deep Gaussian Process for Crop Yield Prediction

- $\boldsymbol{x} = \left(\boldsymbol{I}^{(1)}, \dots, \boldsymbol{I}^{(T)}, \boldsymbol{g}_{loc}, g_{year}\right)$ denote an original data point represented by image sequences, geographic locations that have crop yield data, year
- $y(\boldsymbol{x}) = f(\boldsymbol{x}) + h(\boldsymbol{x})^T \boldsymbol{\beta}$

  When is it purely a prediction task?

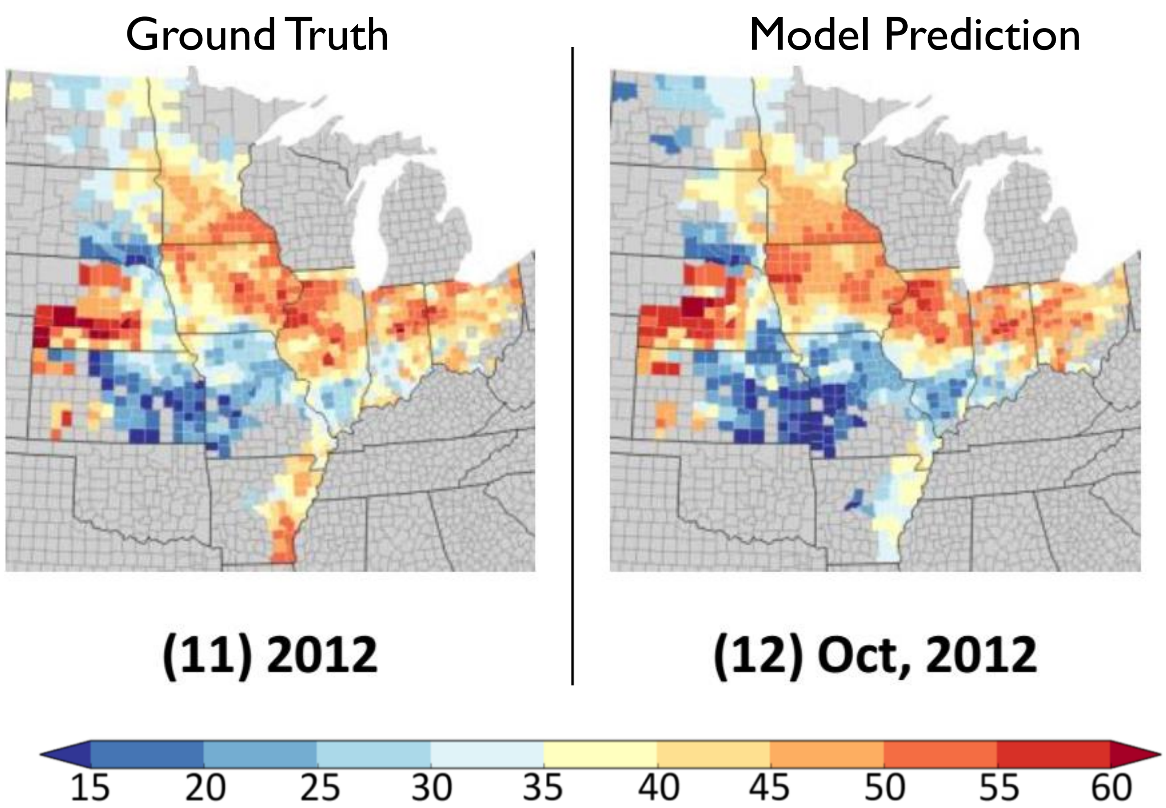
  - $f(\boldsymbol{x}) \sim \mathcal{GP}\left(\boldsymbol{0}, k(\boldsymbol{x}, \boldsymbol{x}')\right)$

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

  - $h(\boldsymbol{x})$ is a fixed set of basis functions
  - $\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{b}, \boldsymbol{B})$ is an independent random variable
  - $h(\boldsymbol{x})$ and b correspond to the final layer in the deep model (CNN), and $\boldsymbol{B} = \sigma_b I$

Hyperparameters to be tuned

Result


Ground Truth
Model Prediction
(11) 2012
(12) Oct, 2012
15
20
25
30
35
40
45
50
55
60

Comparison

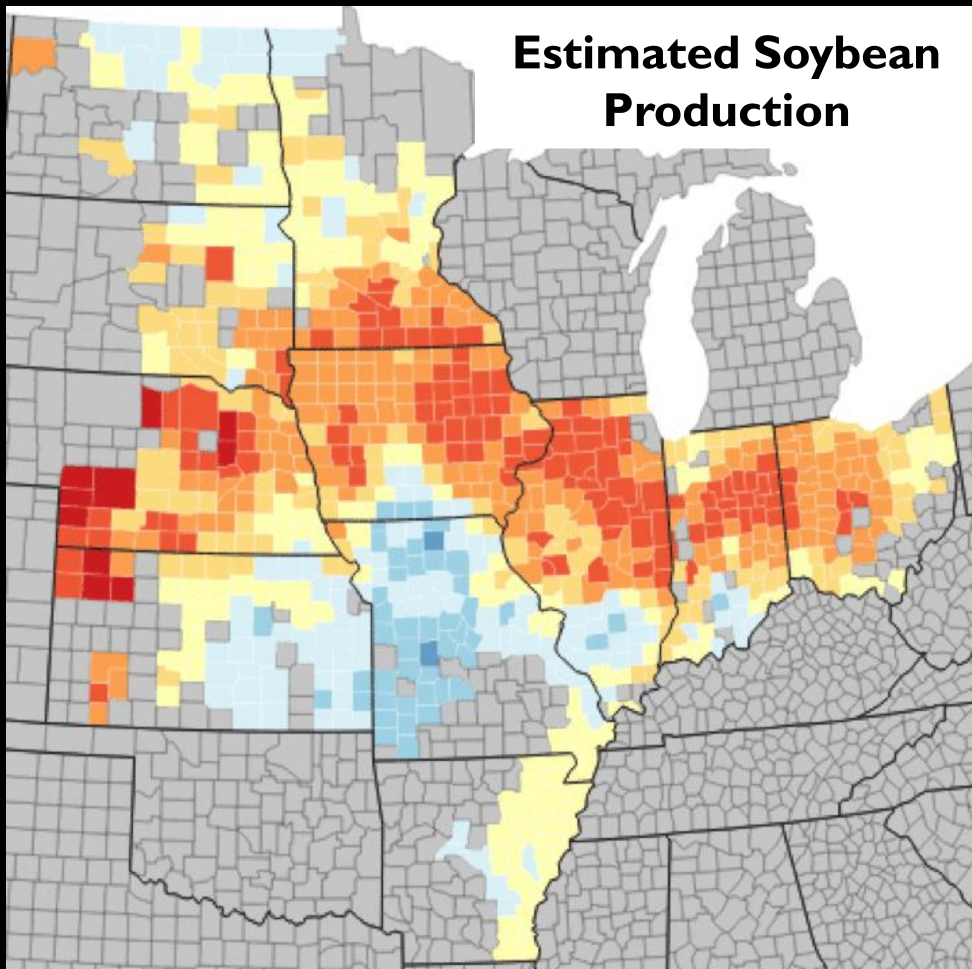
- The Mean Absolute Percentage Error (MAPE) of US-level model performance, averaged from 2009 to 2015

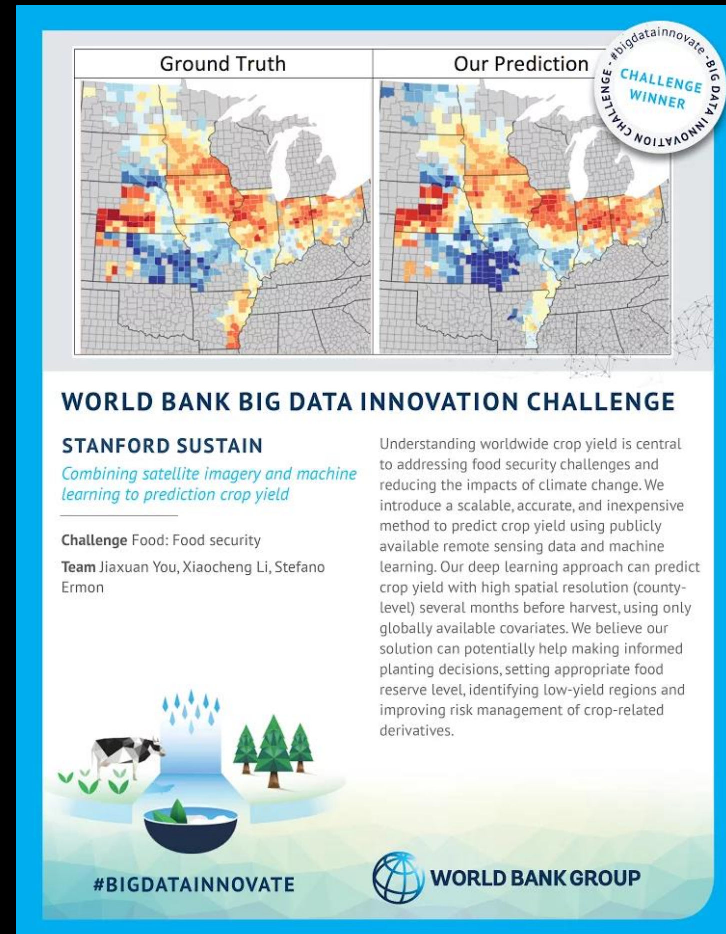
	July	August		September		October	
	Ours	USDA	Ours	USDA	Ours	USDA	Ours
MAPE	5.65	3.92	**3.37**	4.14	**3.41**	**2.48**	3.19

  - MAPE: a measure of prediction accuracy of a forecasting method

$$\text{MAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

A_t: Actual value
F_t: Predicted value


Estimated Soybean Production


Ground Truth
Our Prediction
#bigdatainnovate · BIG DATA INNOVATION CHALLENGE
CHALLENGE WINNER
WORLD BANK BIG DATA INNOVATION CHALLENGE
STANFORD SUSTAIN
Combining satellite imagery and machine learning to prediction crop yield
Challenge Food: Food security
Team Jiaxuan You, Xiaocheng Li, Stefano Ermon
Understanding worldwide crop yield is central to addressing food security challenges and reducing the impacts of climate change. We introduce a scalable, accurate, and inexpensive method to predict crop yield using publicly available remote sensing data and machine learning. Our deep learning approach can predict crop yield with high spatial resolution (county-level) several months before harvest, using only globally available covariates. We believe our solution can potentially help making informed planting decisions, setting appropriate food reserve level, identifying low-yield regions and improving risk management of crop-related derivatives.
#BIGDATAINNOVATE
WORLD BANK GROUP

Outline

- Gaussian Process Regression
- Estimate Crop Yield
- Discussion

Discussion

- In addition to predicting/estimating poverty level and crop yield, what can remote sensing data be used for?

Technique	Agriculture	Education	Environmental sustainability	Healthcare	Combating information manipulation	Public safety	Social care and urban planning	Transportation	Total
Cognitive modeling	0	0	0	0	0	1	0	0	1
Constraint satisfaction and optimization	2	5	31	48	20	26	9	59	173
Cognitive systems	1	2	2	7	2	3	1	5	20
Computer vision	3	8	12	20	6	12	7	19	79
Game playing and interactive entertainment	0	1	0	1	0	0	0	0	2
Game theory and economic paradigms	3	5	30	6	11	31	1	16	78
Human-AI collaboration	1	8	11	23	9	6	6	17	69
Human computation and crowd sourcing	1	5	6	20	45	12	11	15	98
Heuristic search and optimization	1	3	11	14	8	8	6	26	69
Knowledge representation and reasoning	0	0	0	5	3	2	0	1	11
Multiagent systems	2	7	47	19	16	22	8	31	122
Machine learning	12	27	65	174	53	65	36	92	460
Natural language processing	4	12	6	18	10	10	5	3	58
Planning, routing, and scheduling	9	4	48	43	14	28	31	84	210
Robotics	3	4	12	10	4	5	4	10	47
Reasoning under uncertainty	4	3	30	23	8	6	6	13	78
Total	40	78	225	344	155	177	90	253	1176

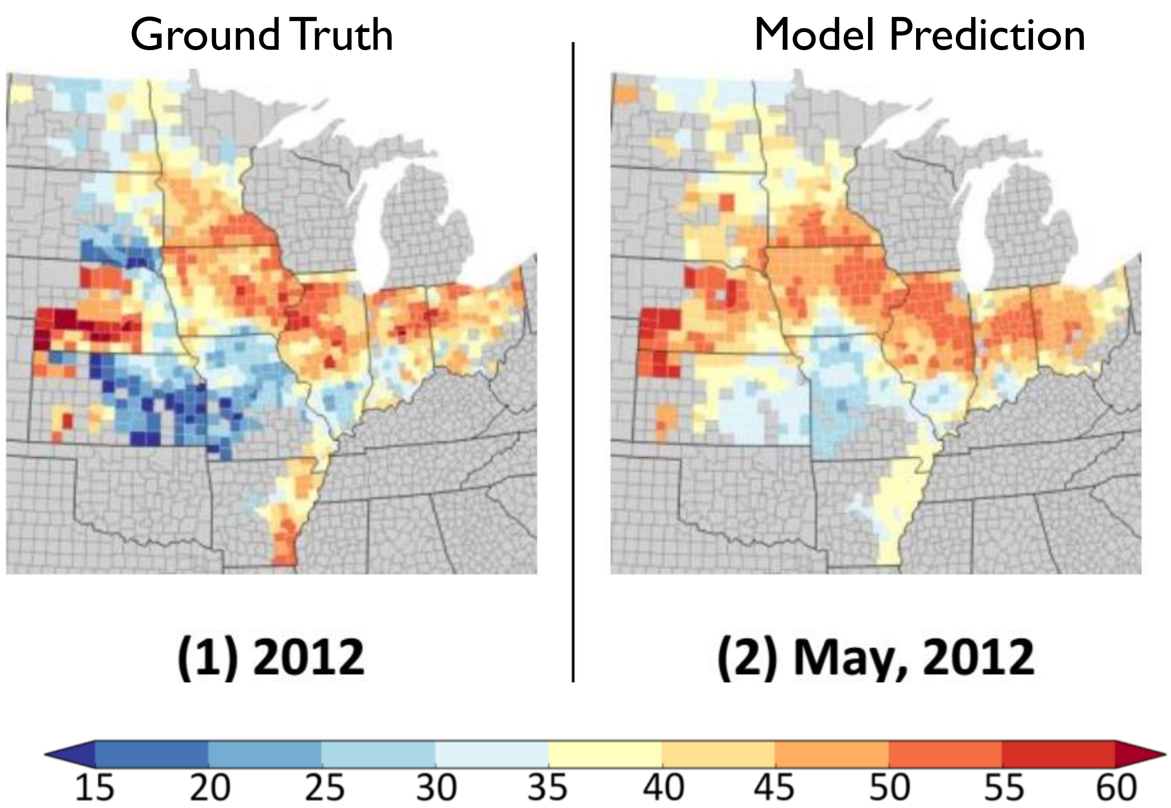
References

- *[1] Combining satellite imagery and machine learning to predict poverty*
- *[2] Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data*
- *[3] (video) Measuring progress towards sustainable development goals with machine learning*
- *[4] Gaussian Processes for Machine Learning, Chapter 2*

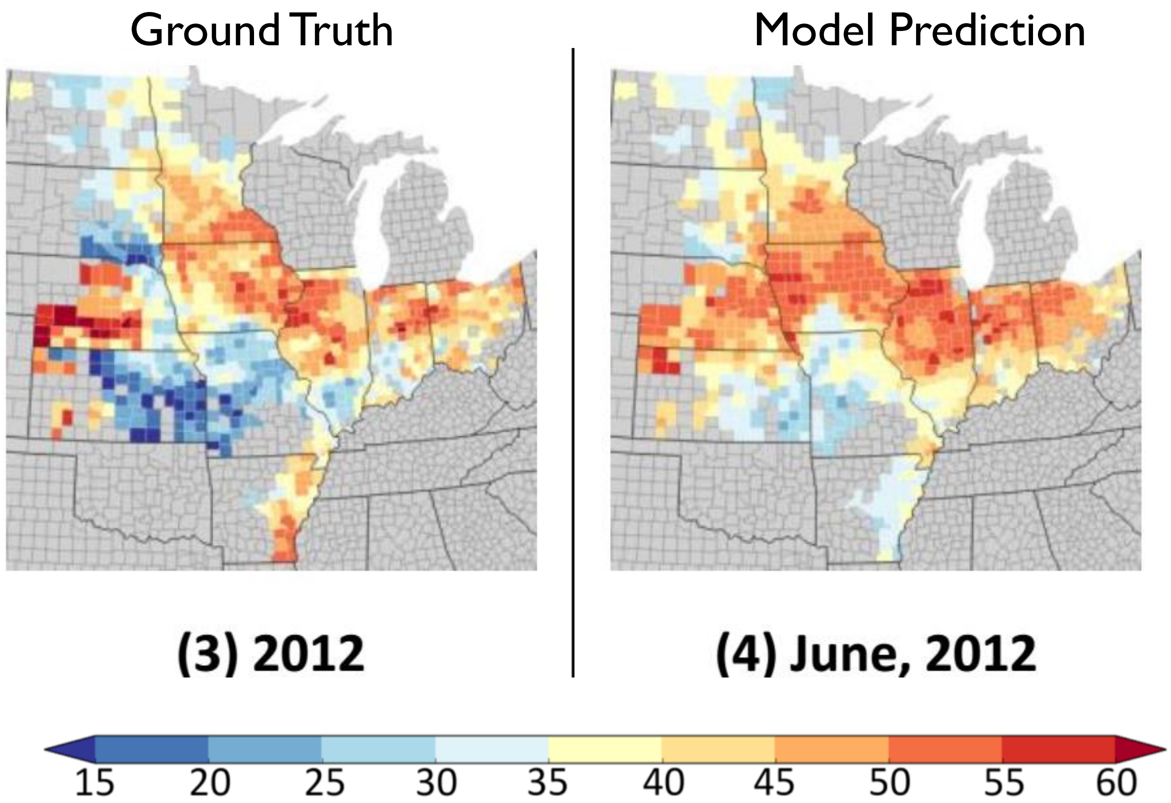
Acknowledgment

- The slides in this lecture are based on the slides provided by Stefano Ermon

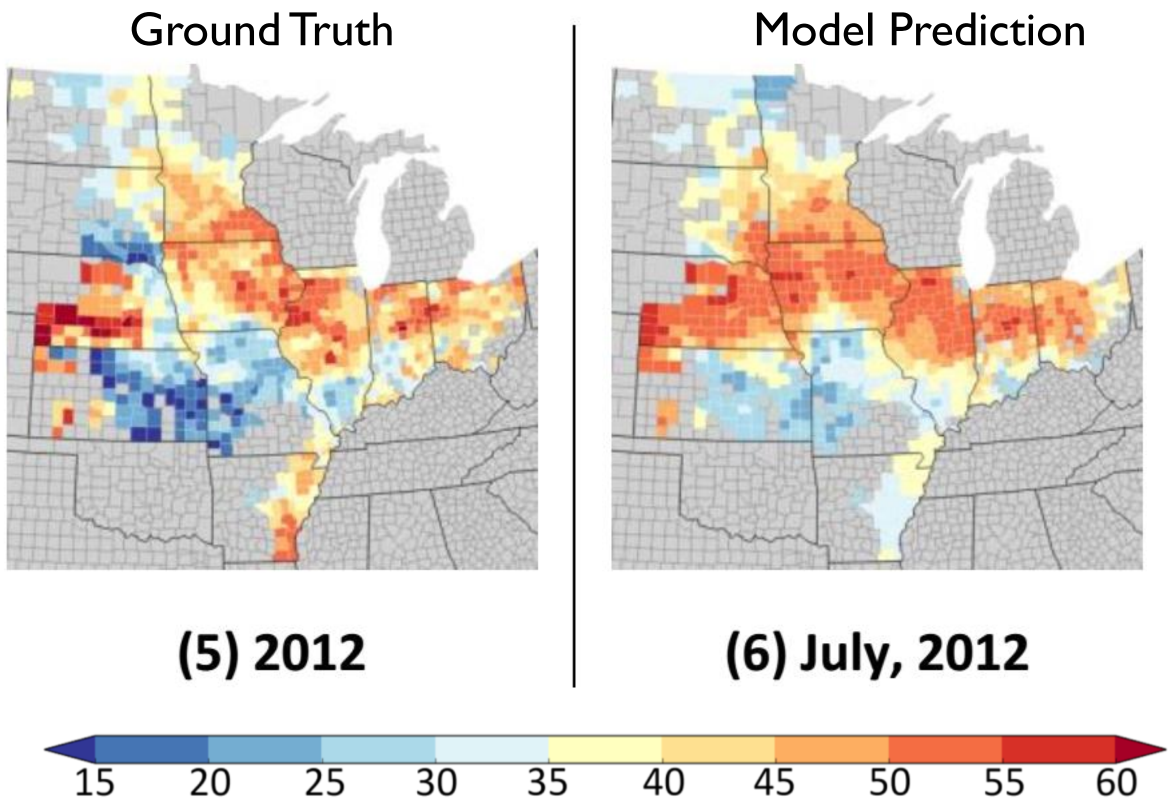
Result


Ground Truth
Model Prediction
(1) 2012
(2) May, 2012
15
20
25
30
35
40
45
50
55
60

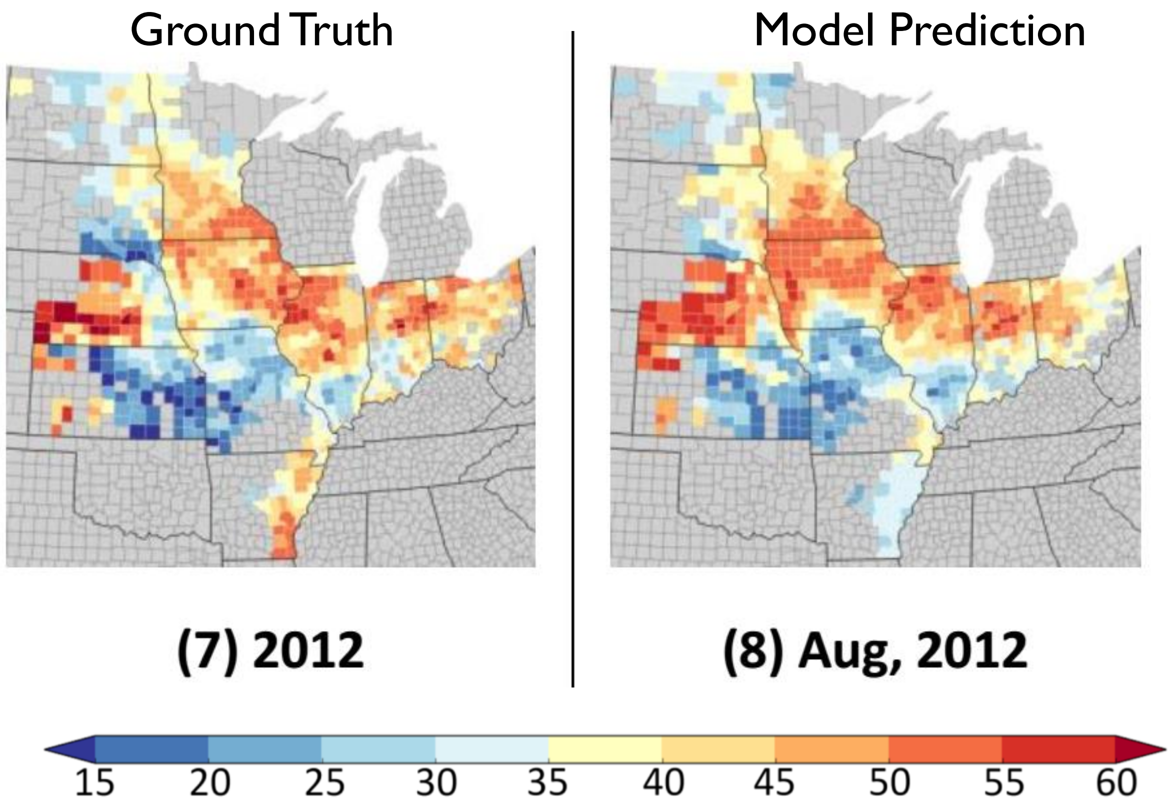
Result


Ground Truth
Model Prediction
(3) 2012
(4) June, 2012
15
20
25
30
35
40
45
50
55
60

Result

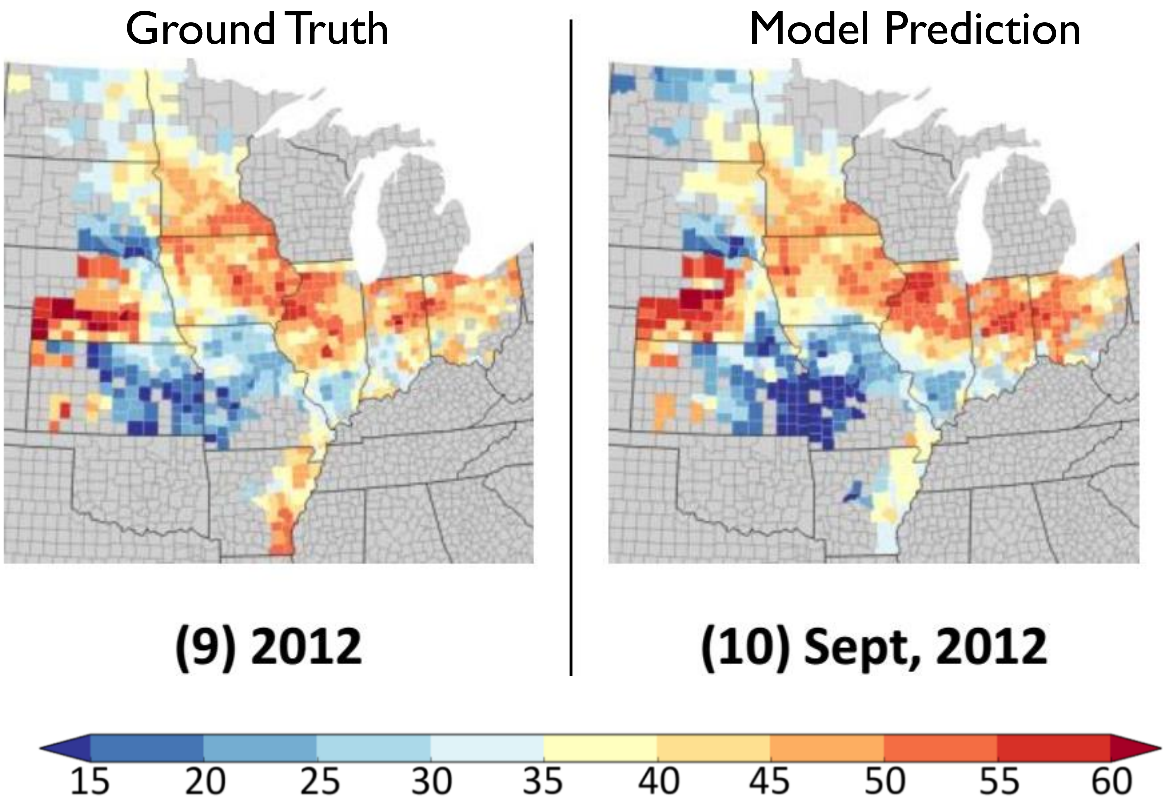

Ground Truth
Model Prediction
(5) 2012
(6) July, 2012
15
20
25
30
35
40
45
50
55
60

Result



Ground Truth
Model Prediction
(7) 2012
(8) Aug, 2012
15
20
25
30
35
40
45
50
55
60

Result

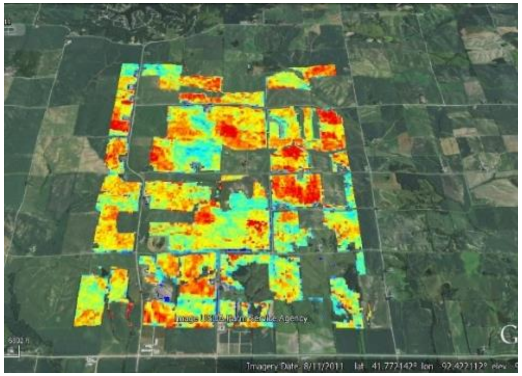

Ground Truth
Model Prediction
(9) 2012
(10) Sept, 2012
15
20
25
30
35
40
45
50
55
60

Possible Directions

Planet data

Prices? Poverty?

Global estimates



Field level predictions

Comparison

Year	Ours (Jul.)	USDA (Aug.)	Ours (Aug.)	USDA (Sept.)	Ours (Sept.)	USDA (Oct.)	Ours (Oct.)
2009	-4.26	-5.23	**-3.84**	-3.86	**1.12**	-3.64	**1.22**
2010	-7.02	**1.15**	-2.64	**2.76**	-5.93	**2.07**	-3.76
2011	7.22	**-1.43**	6.62	**-0.48**	7.14	**-1.19**	6.90
2012	11.3	-9.75	**1.04**	-11.75	**-1.63**	-5.50	**3.33**
2013	-1.47	-3.18	**3.17**	-6.36	**-2.36**	N/A	**-2.15**
2014	3.53	-4.42	**1.67**	**-0.54**	-4.42	-0.84	**-0.71**
2015	-4.77	**-2.29**	-4.62	**-5.15**	-8.38	**-1.67**	-4.26
Absolute Mean	5.65	3.92	**3.37**	4.14	**3.41**	**2.48**	3.19