Lecture 9:
Soil Moisture Data - From Coarse to Fine Grained Resolution

Instructor: Danny Rorabaugh

and Michela Taufer

THE UNIVERSITY OF TENNESSEE
KNOXVILLE

BIG ORANGE. BIG IDEAS.®

Today's Outline

- Learn about problems on Soil Moisture Data
  - Limits of satellite date
  - Workflow for data refinement

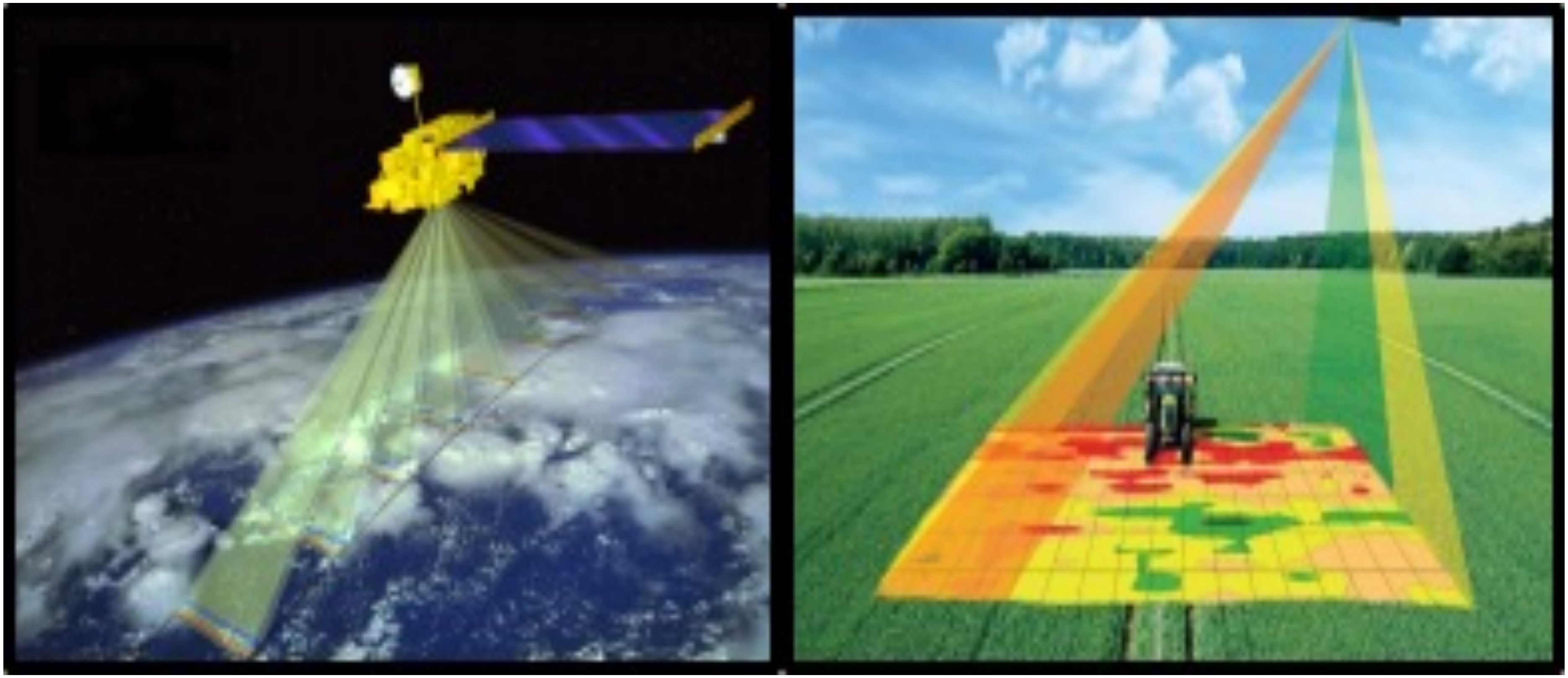
  From: D. Rorabaugh and co-authors: "SOMOSPIE: A modular SOil MOisture SPatial Inference Engine based on data driven decisions. 2016 (Paper in preparation)

- Assignment 10 (last of the semester)
- Project
  - Start writing the 2-page paper
  - Revision of slides

BIG ORANGE BIG IDEAS

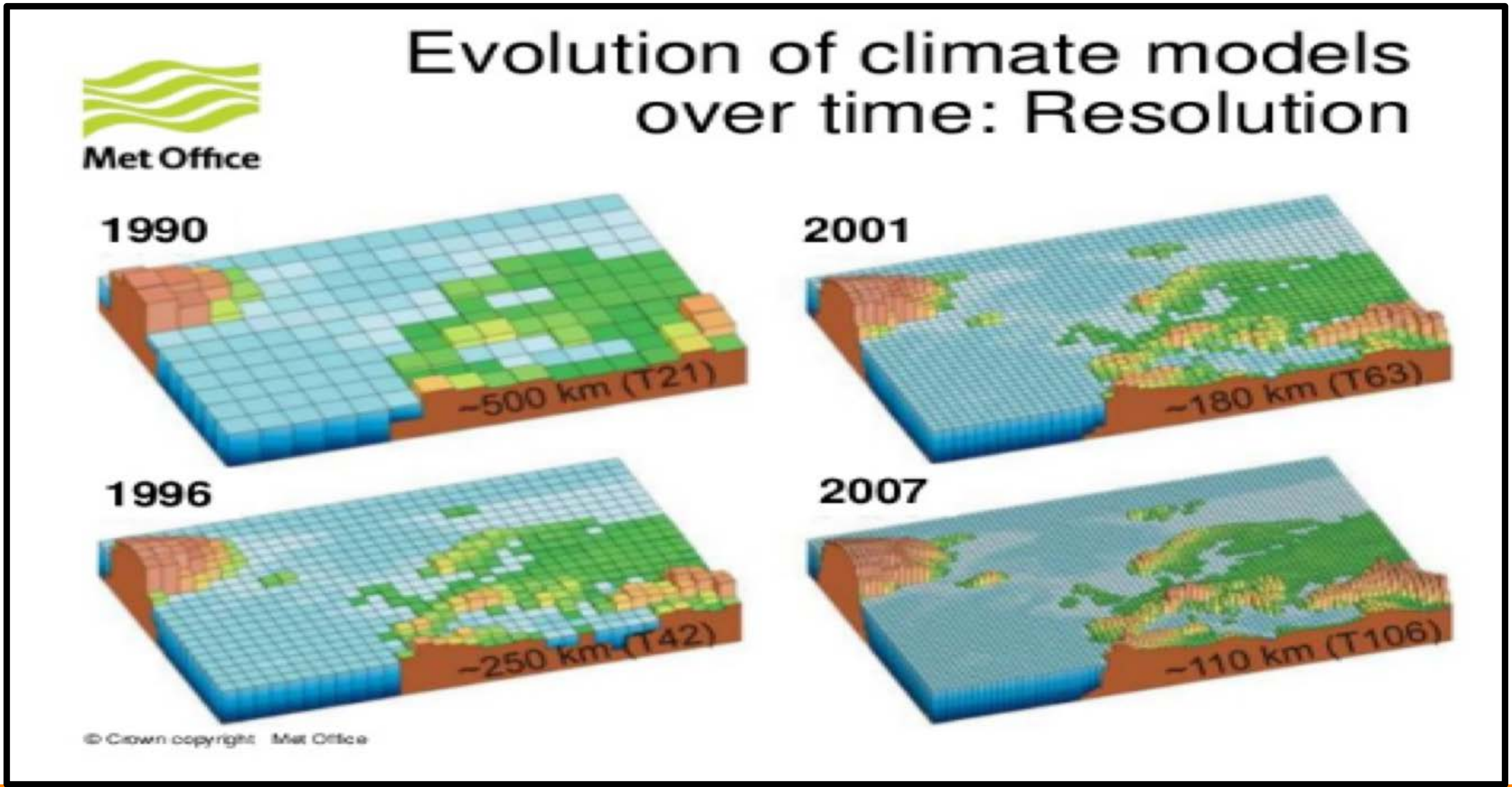
Collecting Soil Moisture Data

- Satellites collect raster data across the surface of the Earth (see [10] [13] in Rorabaugh and co-authors papers

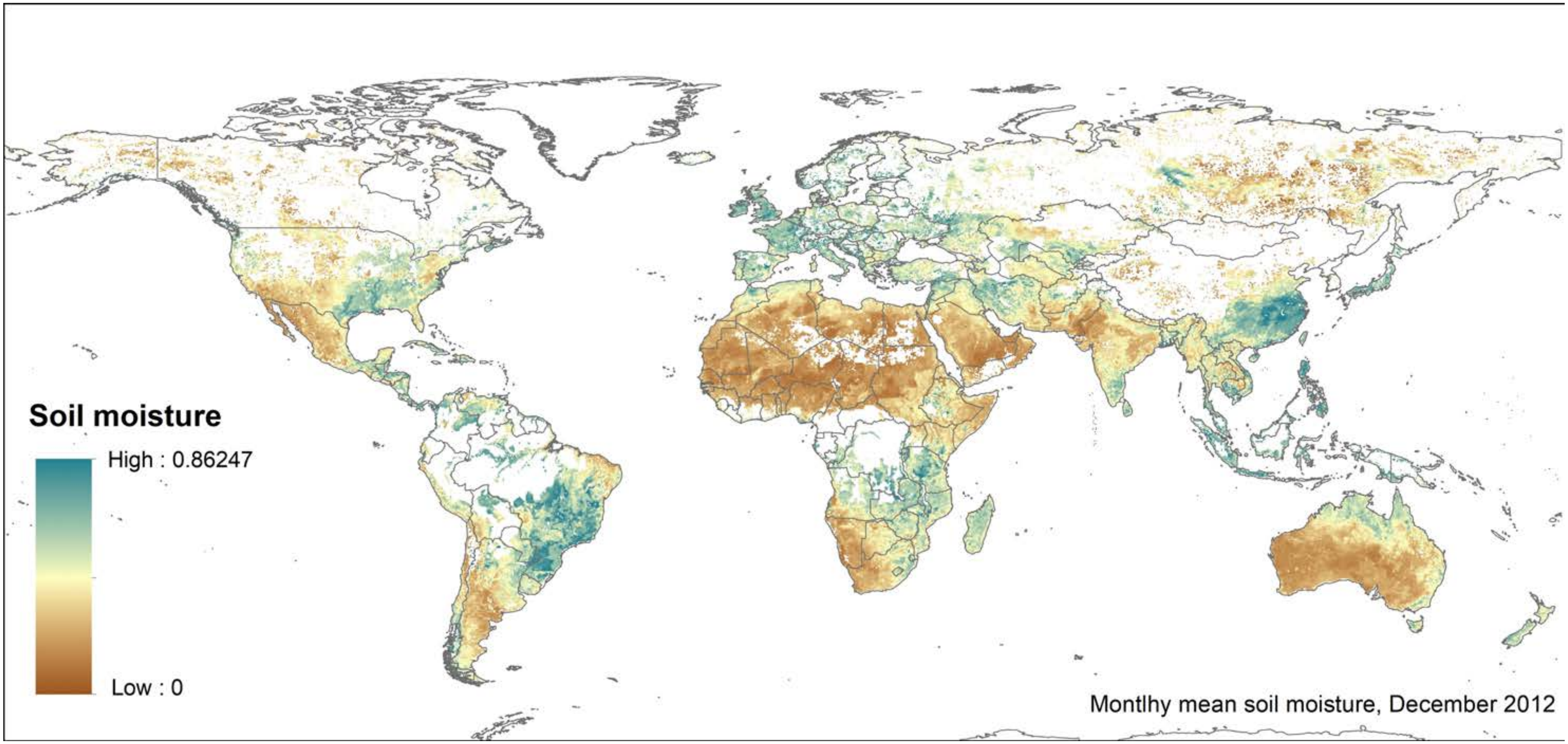


Collecting Soil Moisture Data

- Soil moisture can be used for precision farming, climate prediction, and wild fire propagation modeling



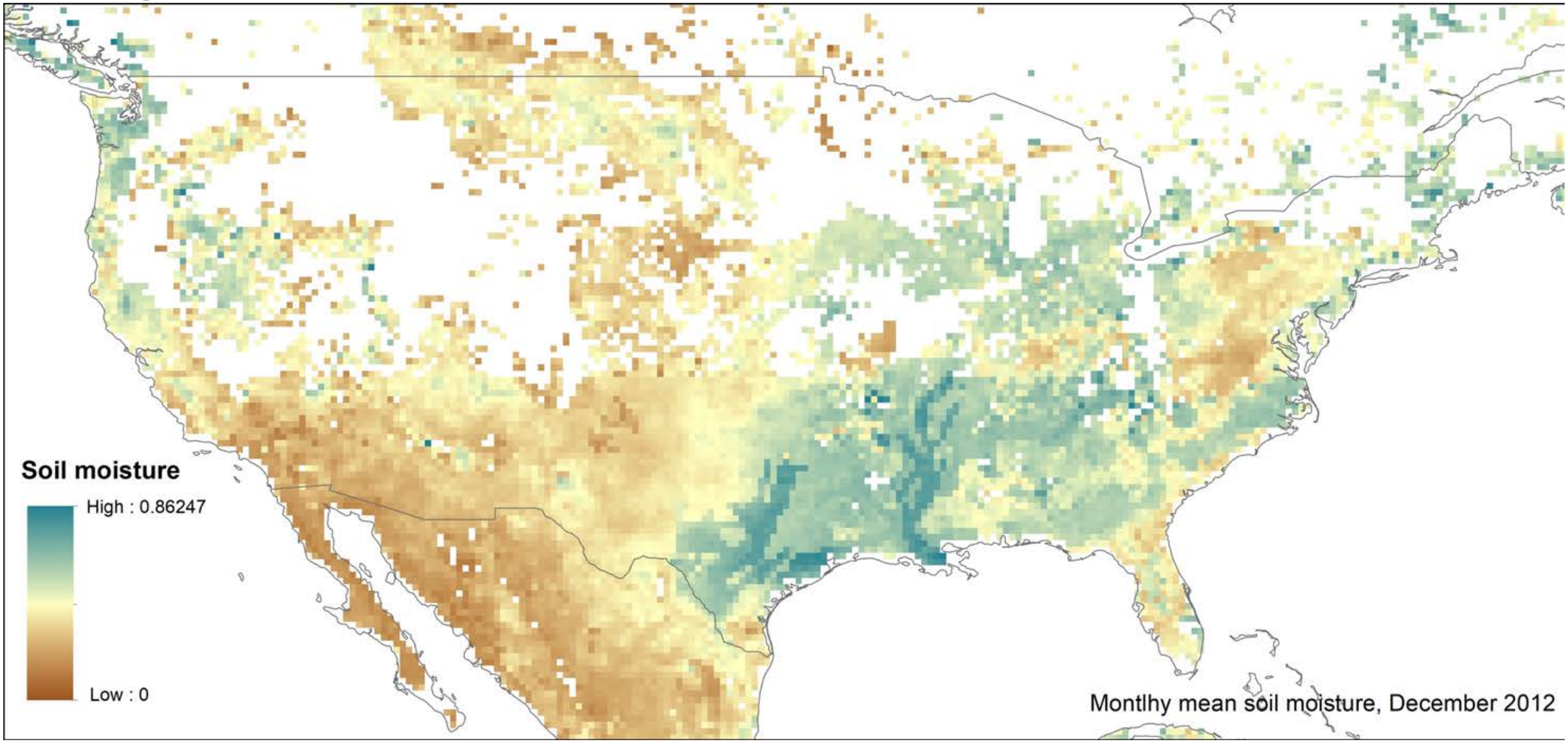
Gaps in Soil Moisture



Monthly soil moisture (m3/m3) averages for December 2000 with gaps where data cannot be collected accurately because of dense vegetation, snow cover, and extremely dry surfaces. Averaged from daily data from ESA-CCI soil moisture database (http://www.esa-soilmoisture-cci.org/).

BIG ORANGE BIG IDEAS

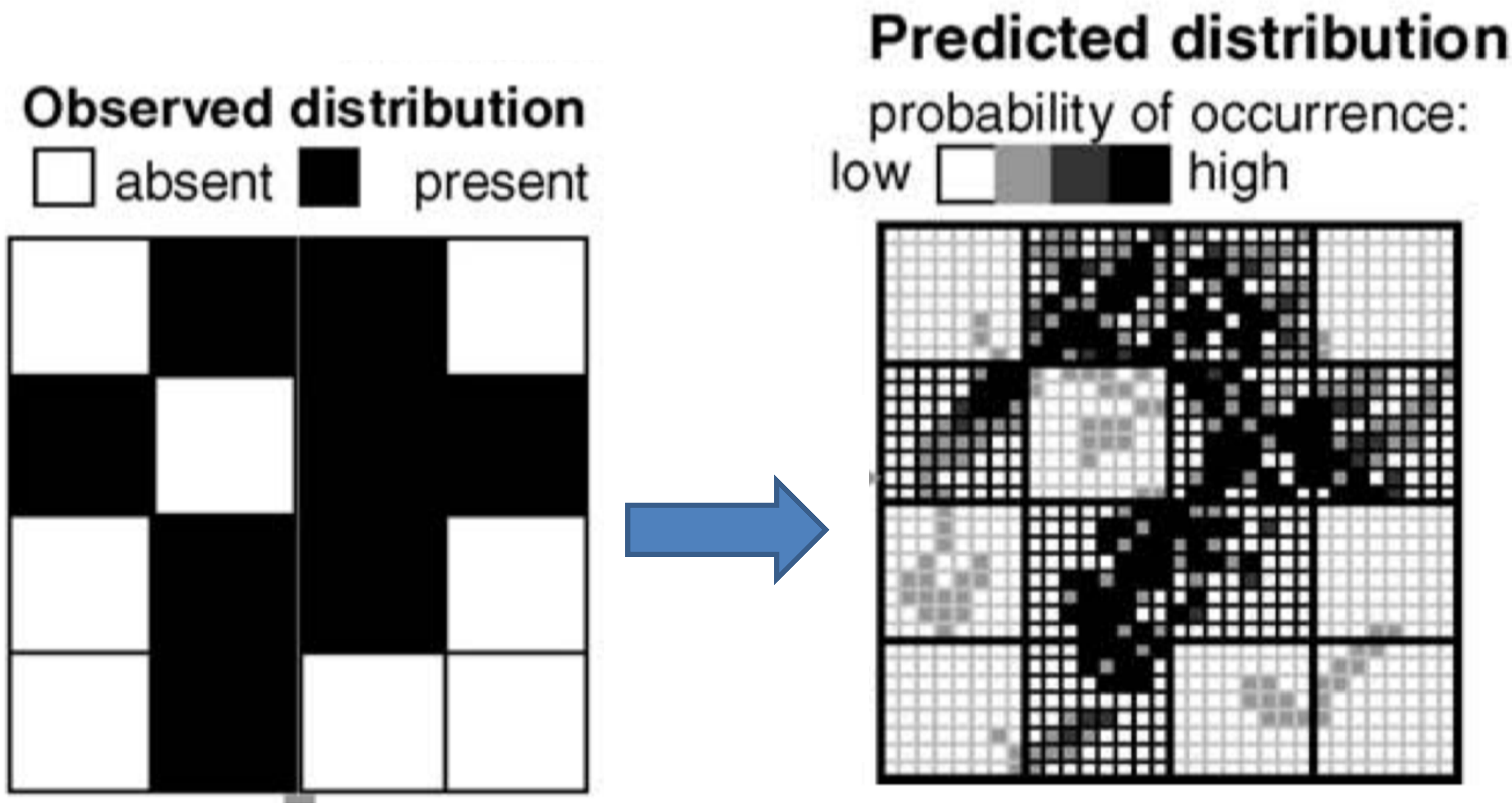
Gaps in Soil Moisture



Monthly soil moisture (m3/m3) averages for December 2000 with gaps where data cannot be collected accurately because of dense vegetation, snow cover, and extremely dry surfaces. Averaged from daily data from ESA-CCI soil moisture database (http://www.esa-soilmoisture-cci.org/).

From Coarse to Fine Grained Data

- Satellites with radar sensors ranging between 25 and 50 km



From: https://pdfs.semanticscholar.org/92a6/b6b8cec29640d7ae0284824a78e18c127435.pdf

BIG ORANGE BIG IDEAS

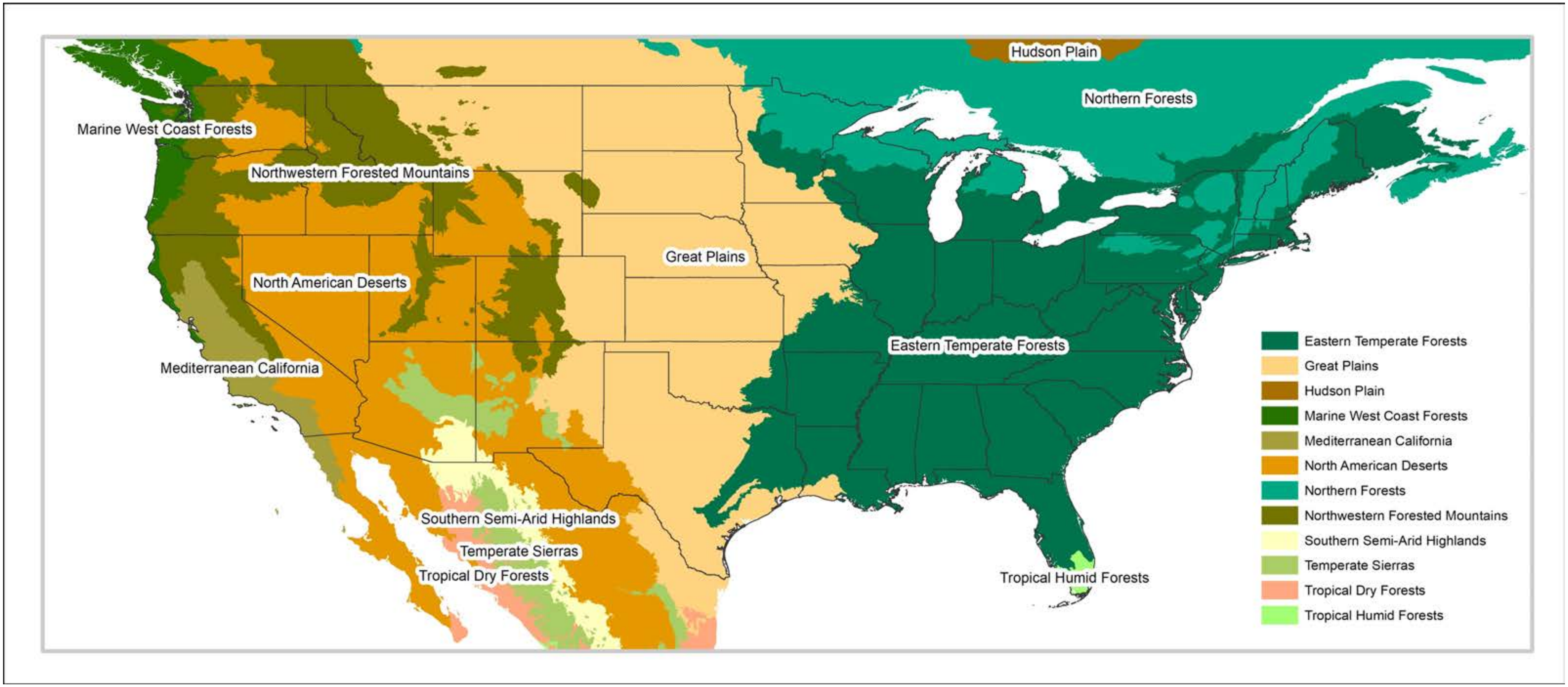
Dealing with Heterogenous Data

Dataset	Spatial resolut.	Temporal resolution	Variable / Description	Source
ESA-CCI	0.25 degrees	Daily, 1978-2013	soil moisture (m^3/m^3)	European Space Agency
Digital surface model (DSM)	≈30 meters	Static ('Current')	Land surface characteristics	The Japan Aerospace Exploration Agency
CEC	n/a	Static ('Current)	Ecoregion boundaries	Commission for Environmental Cooperation

- Satellite date is from European Space Agency
- Land surface characteristic (topography data) are local features
- Ecoregion boundary allows the simplification of predictions
    - When working inside an ecoregion we can ignore the climate impact and heavily rely on topographical data

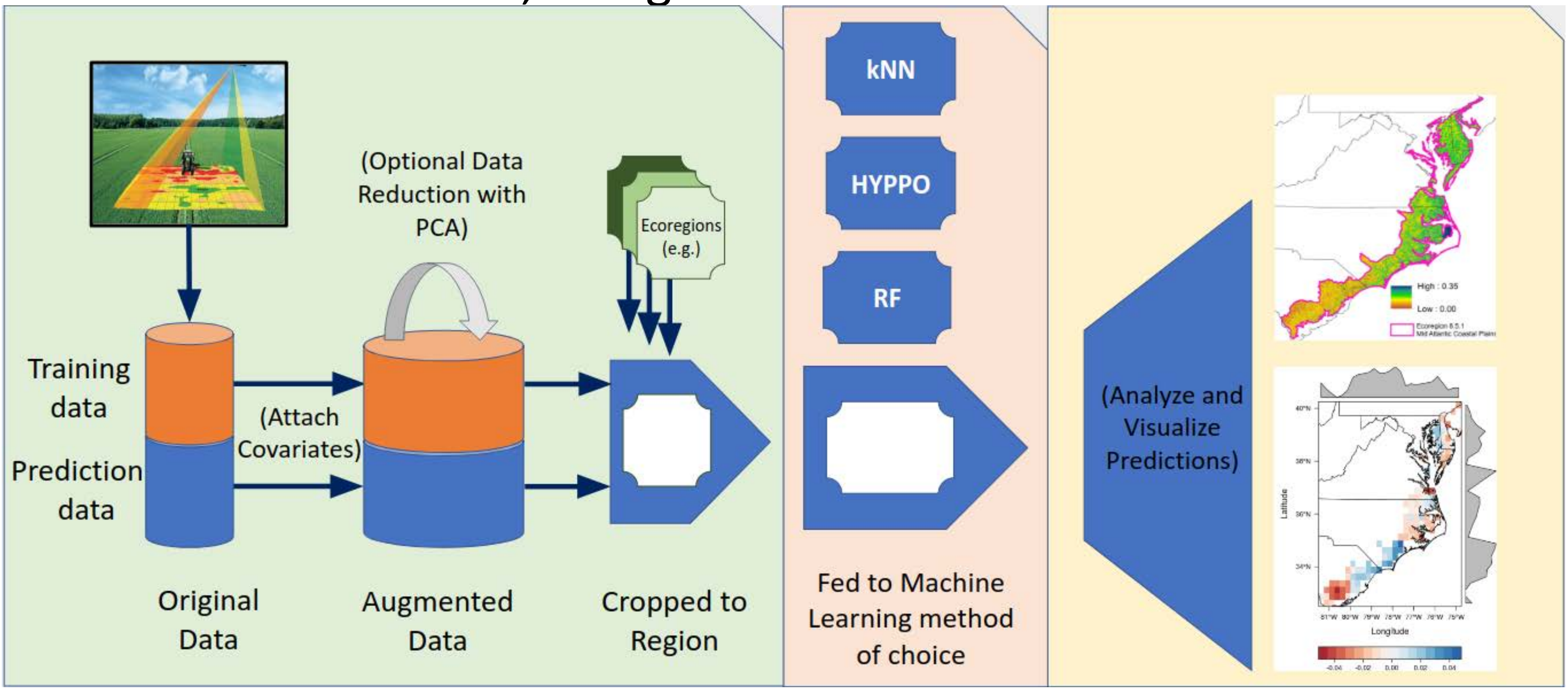
Integrate Climate Regions in Workflow

- Ecoregions across the conterminous United States play a key role to simplify the prediction process



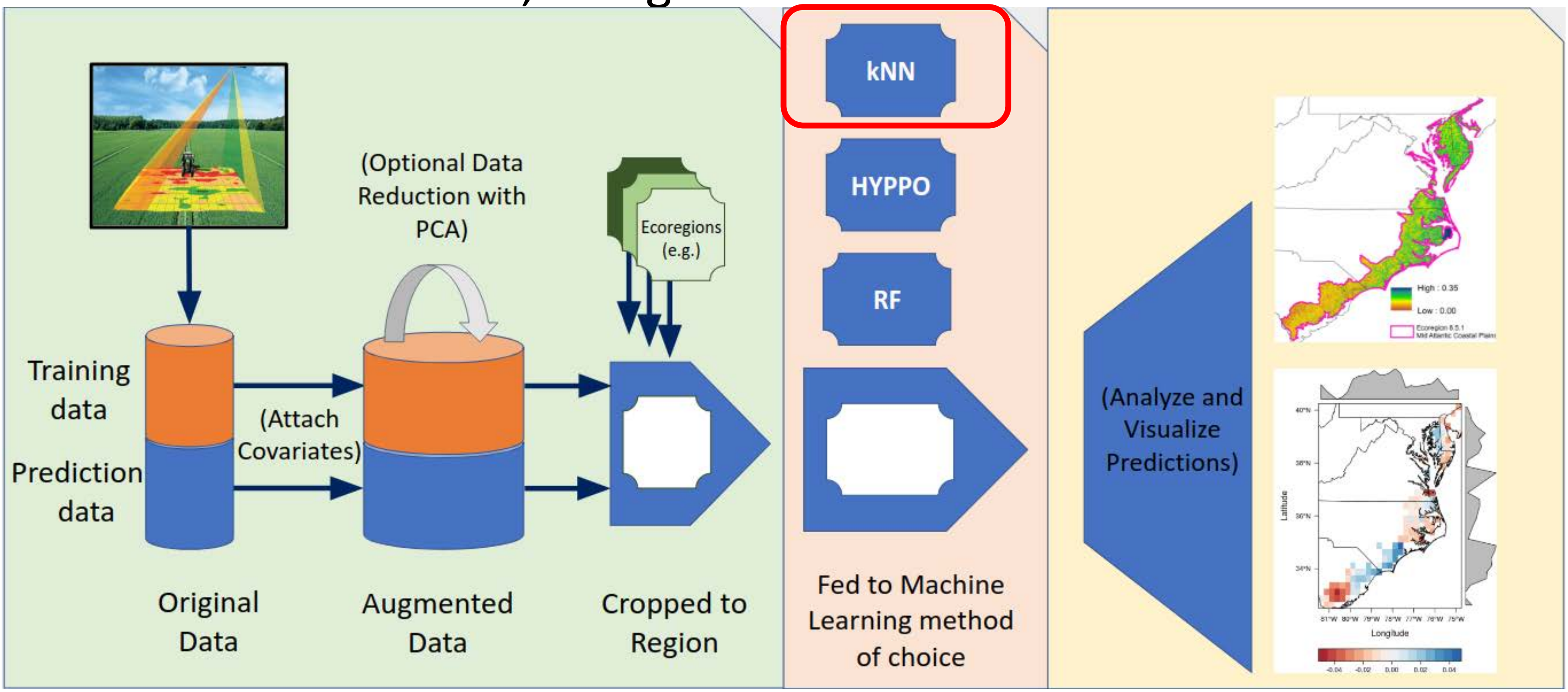
A Workflow to Build Fine Grained Data

- Build a workflow that transforms raw, heterogenous, grained data into accurate, fine grained data



A Workflow to Build Fine Grained Data

- Build a workflow that transforms raw, heterogenous, grained data into accurate, fine grained data

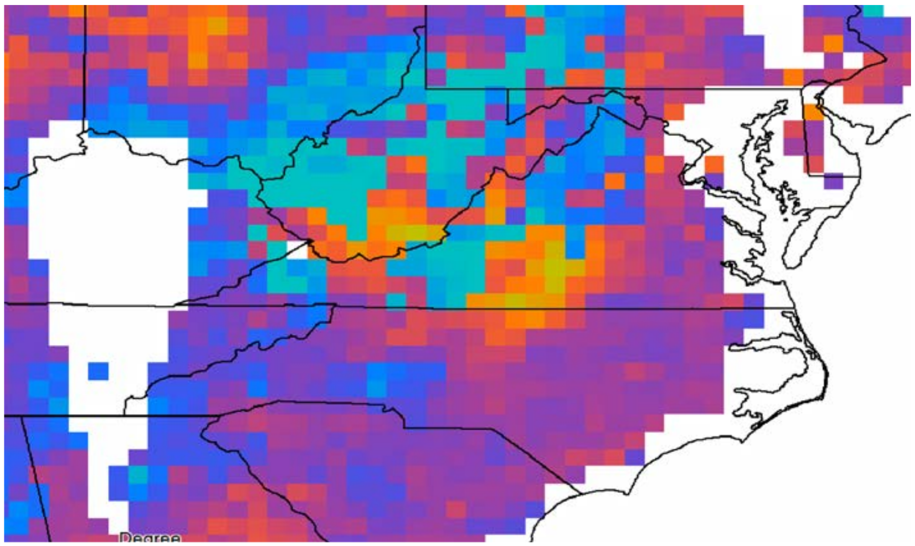


Data

- Soil moisture data are collected by ESA satellites as part of its Climate Change Initiative
- Topographical parameters are derived from a Digital Elevation Model in SAGA GIS
    - Surrogate of overland flow of water (e.g., terrain slope)
    - Surrogate of potential incoming solar radiation (e.g., south or north slopes)

Training and Prediction Data

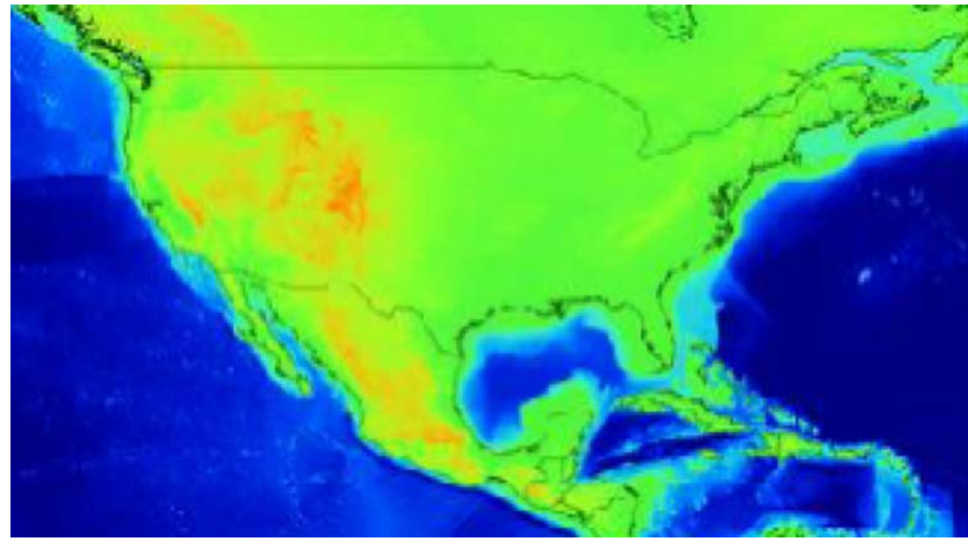
Training Data:



27 km x 27km pixels
Each pixel is a vector:

- latitude and longitude of the centroid
- average soil moisture ratio in pixel
- 15 topographic parameters in centroid (optimal)

Prediction Data:



1 km x 1 km pixels
Each pixel is a vector:

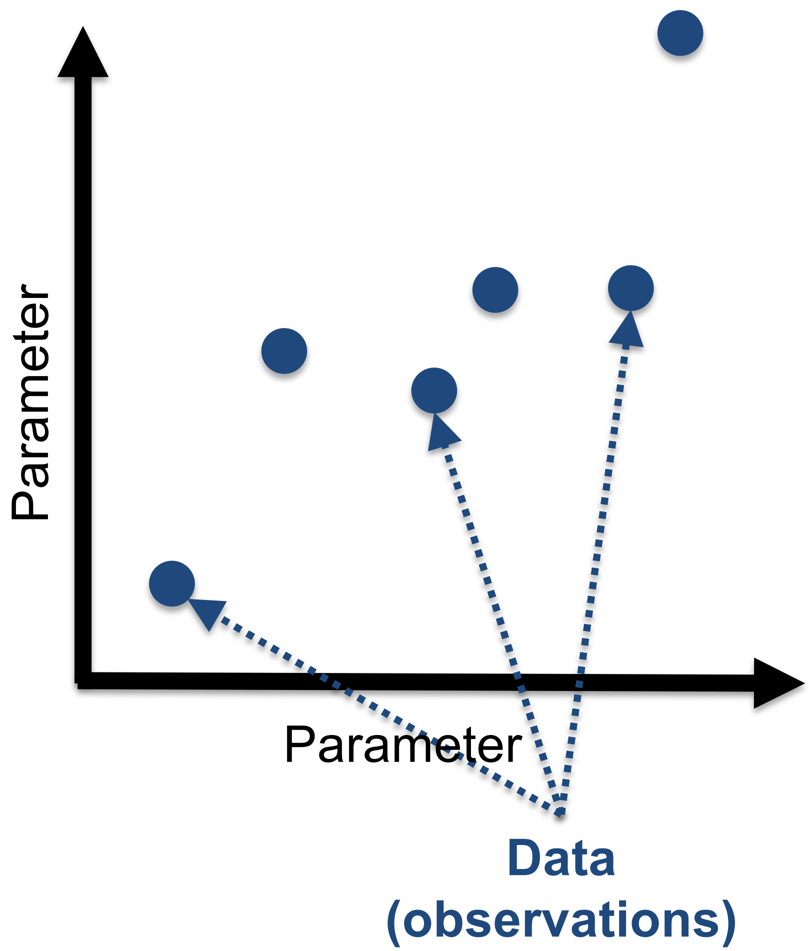
- latitude and longitude of the centroid
- 15 topographic parameters in centroid (optimal)

Modeling Methods

- ***K Nearest Neighbors:*** KNN assigns each point in the testing set a soil moisture that is the weighted average of the soil moisture values of its neighbors
- ***Surrogate-based model:*** SBMs are built by fitting a polynomial surface to sampled data points by using e.g., least squares regression
- ***Hybrid Piecewise Polynomial*****:** HYPPO combines kNN with a Surrogate Based Model to build a global polynomial model of a surface. HYPPO uses a polynomial approximation in each neighborhood of k nearest points to predict soil moisture
- Other methods possible ….

BIG ORANGE BIG IDEAS

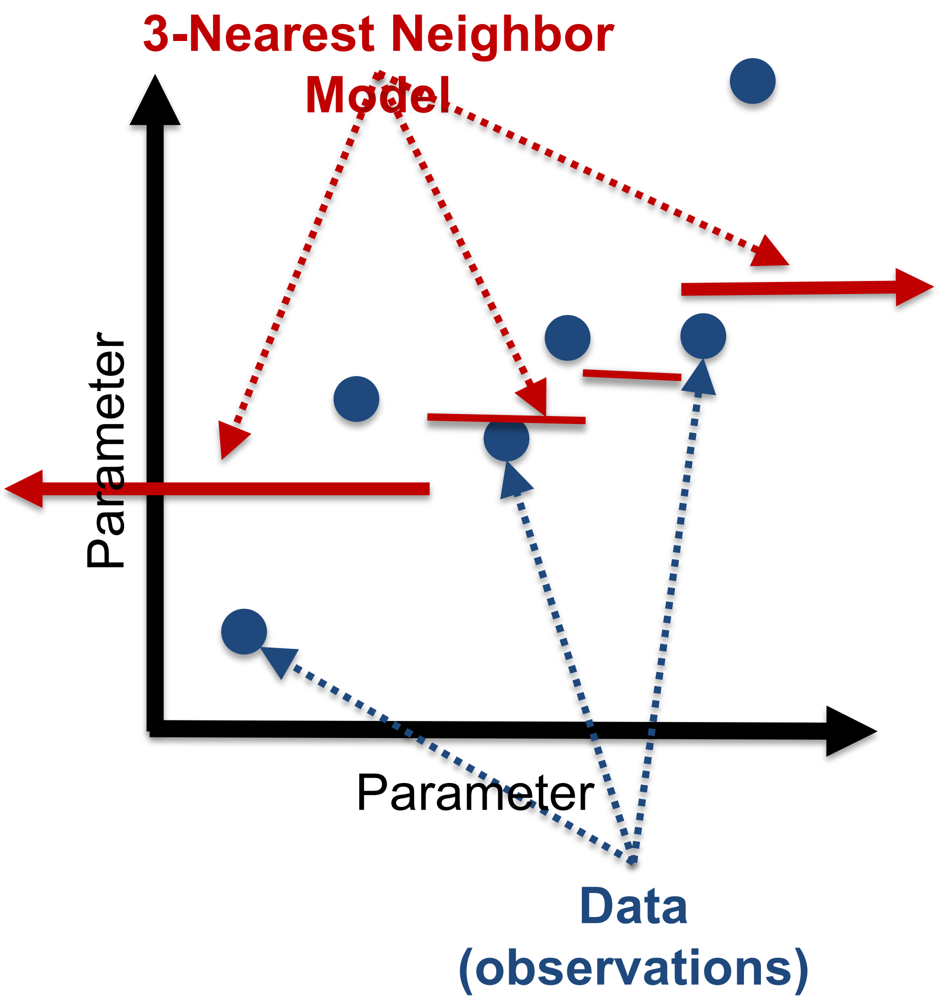
K Nearest Neighbors Model



- Data → small amount of data to create simple, **local models**
- Model → **average** of k nearest neighbors
- Best for → surfaces that are locally **flat**

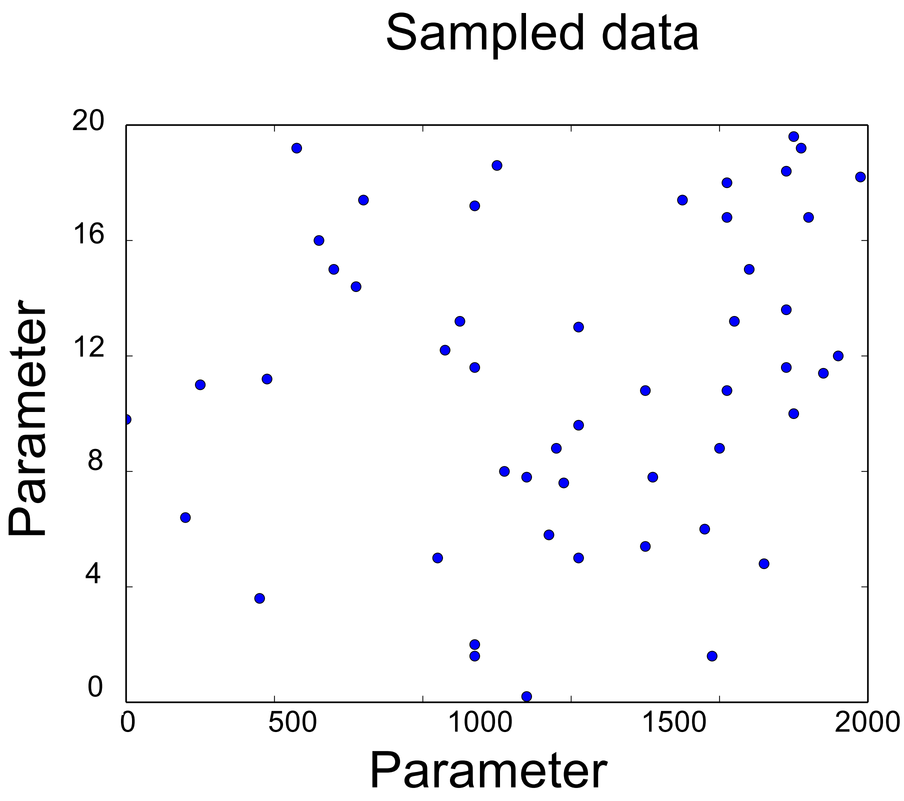
BIG ORANGE BIG IDEAS

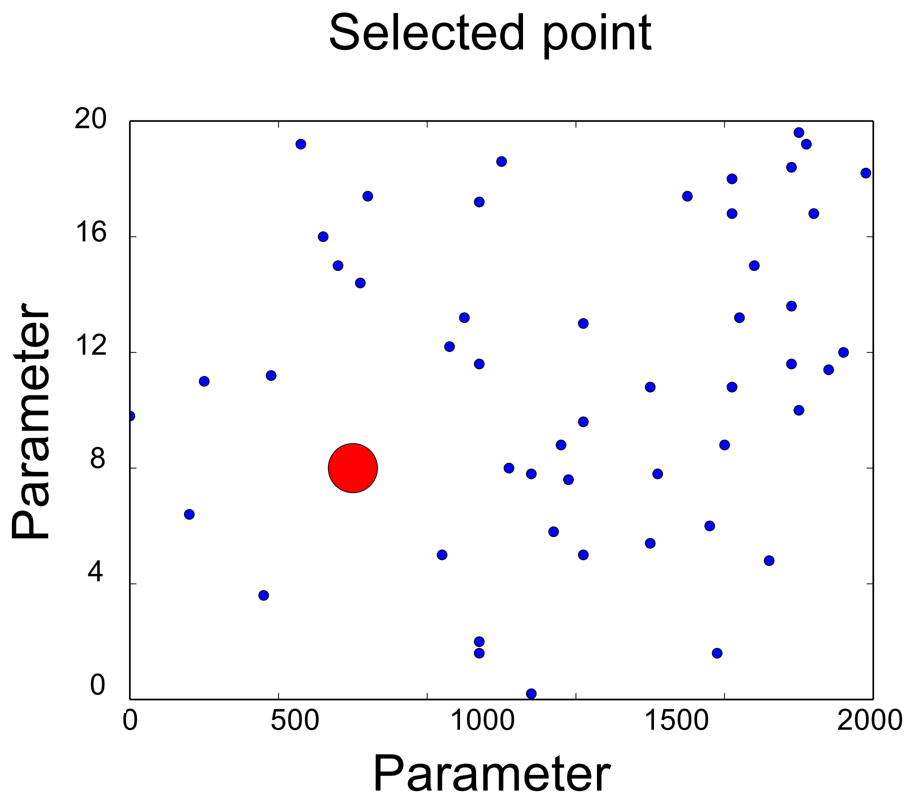
K Nearest Neighbors Model



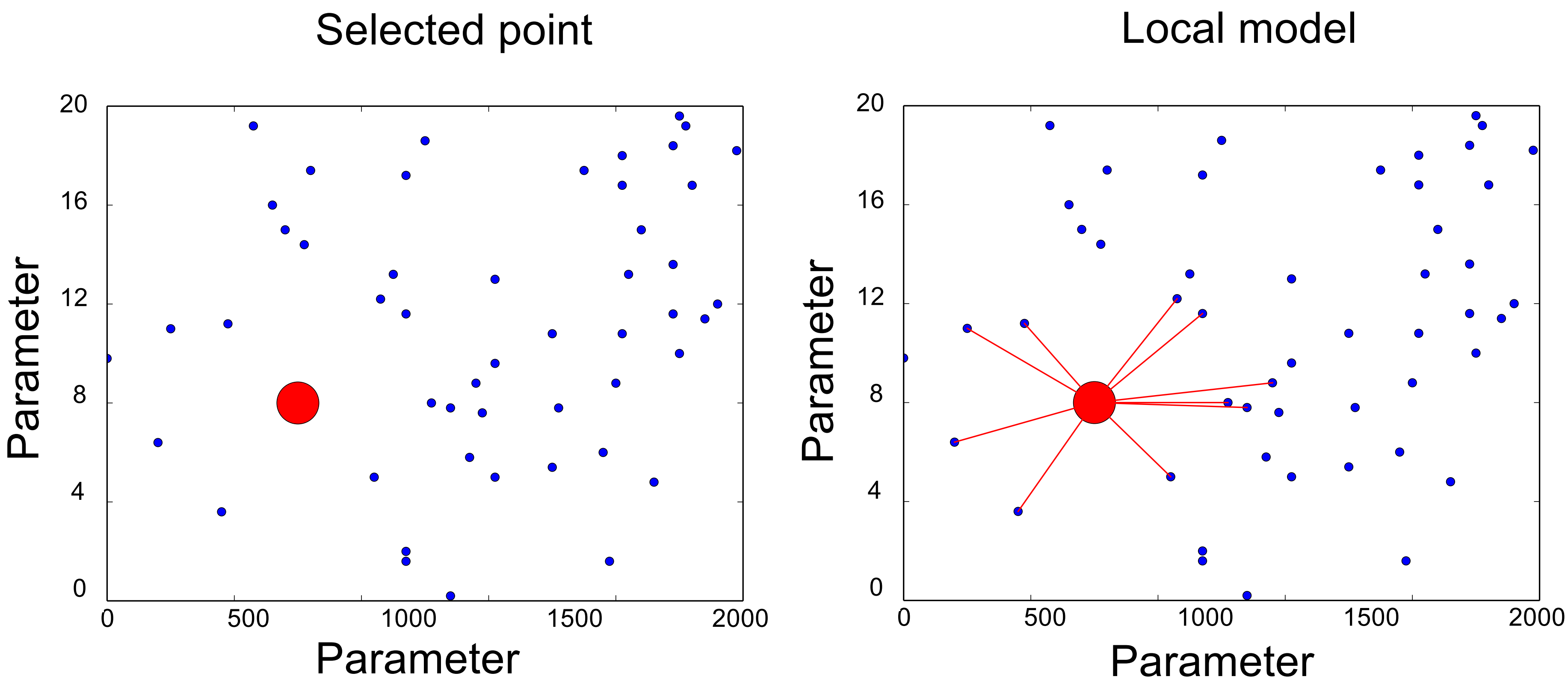
- Data → small amount of data to create simple, **local models**
- Model → **average** of k nearest neighbors
- Best for → surfaces that are locally flat

k Nearest Neighbors Model

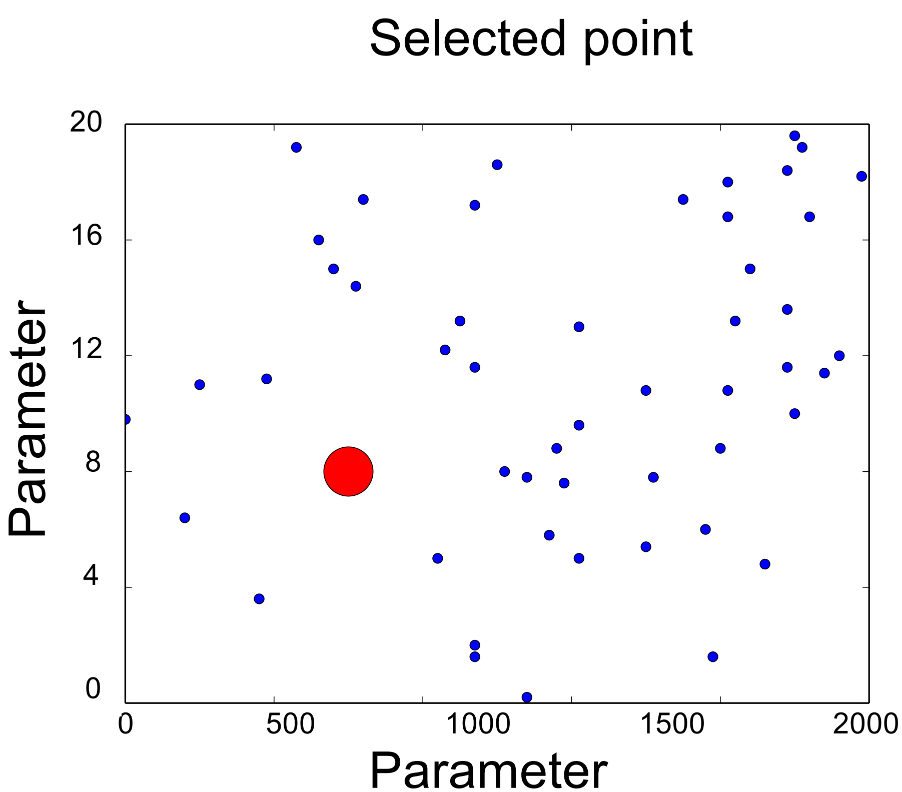


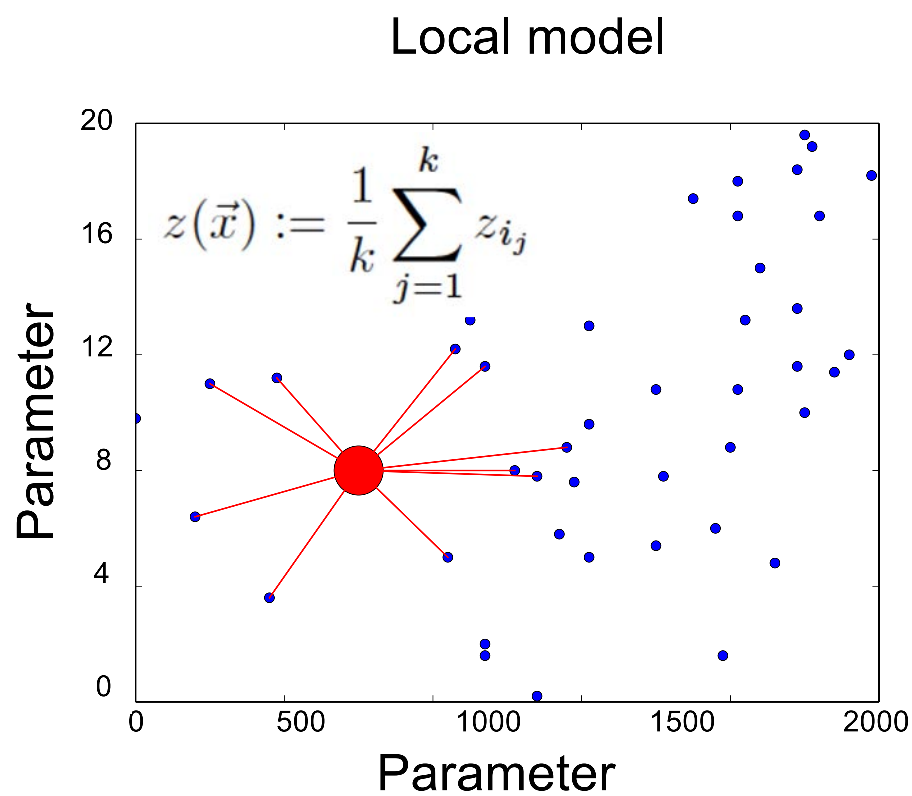


k Nearest Neighbors Model



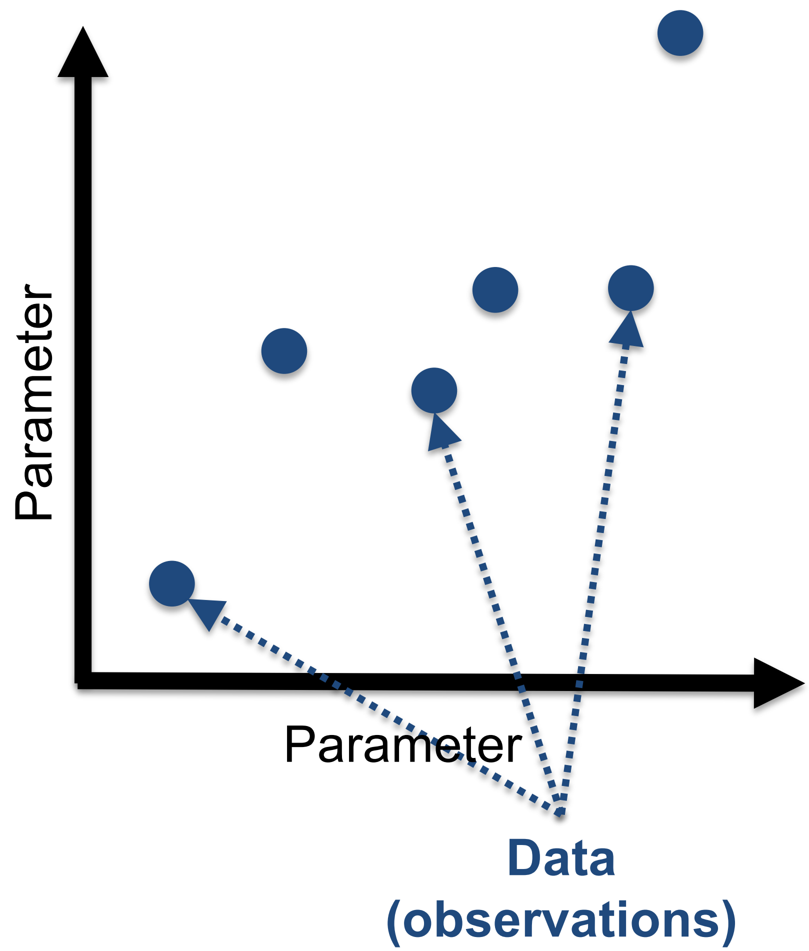
k Nearest Neighbors Model





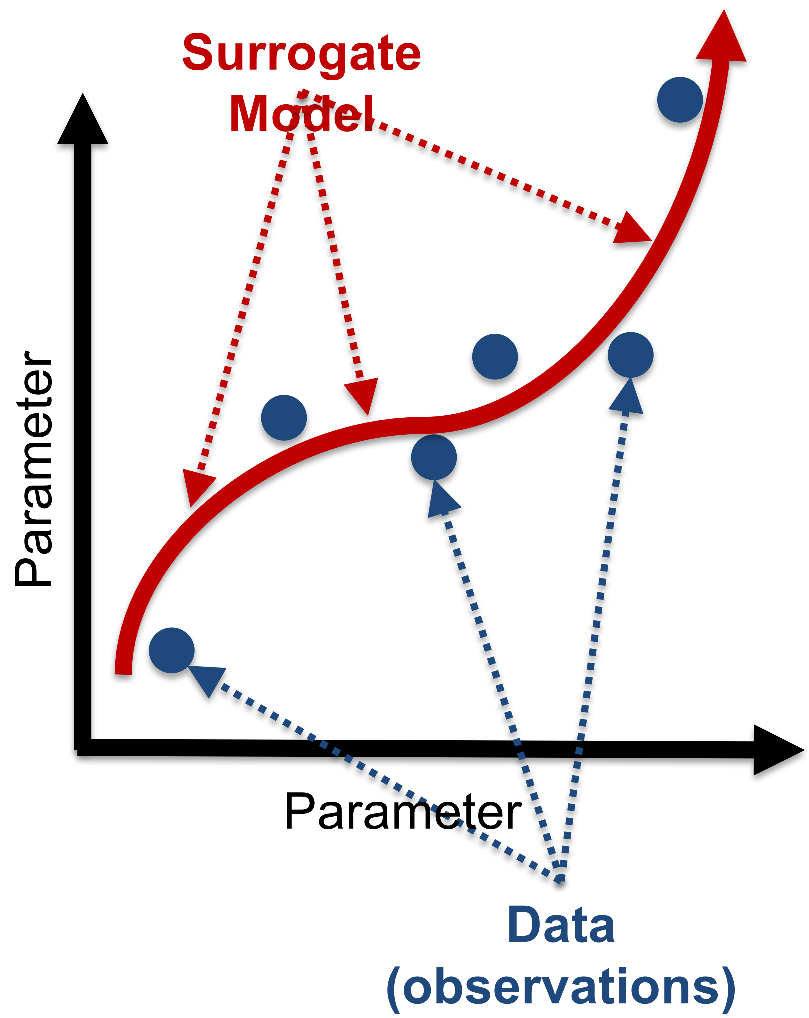
Polynomial with degree zero

Surrogate-Based Modeling



- Data → all sampled data to create a **single global model**
- Model → fit a **polynomial** to data (continuous and differentiable)
- Best for→ finding underlying **global trends** when they exist

Surrogate-Based Modeling



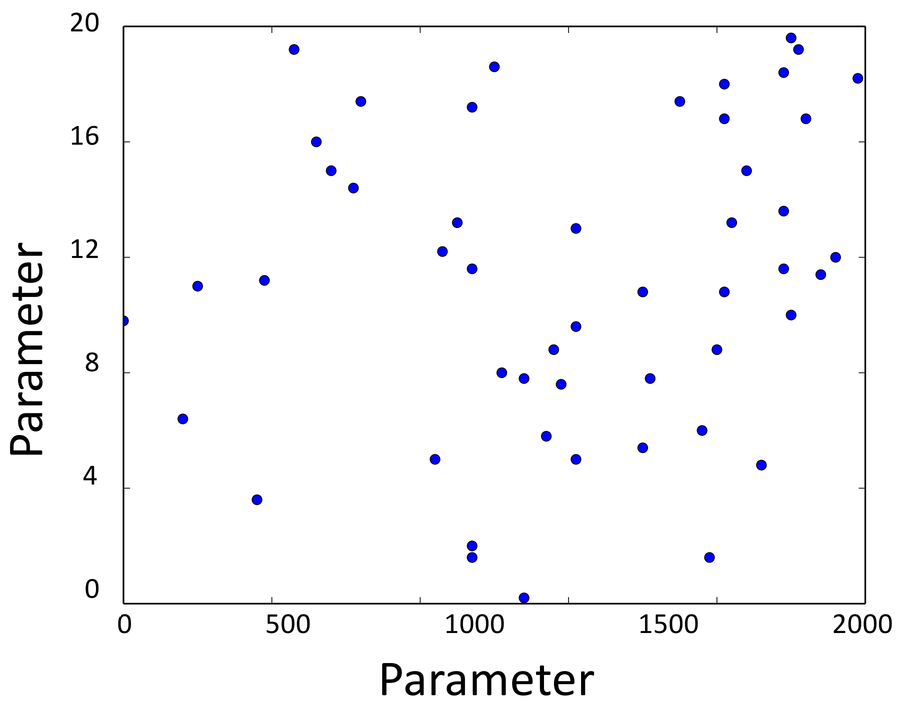
- Data → all sampled data to create a **single global model**
- Model → fit a **polynomial** to data (continuous and differentiable)
- Best for→ finding underlying **global trends** when they exist

T. Johnston et. al., Performance Tuning of MapReduce Jobs Using Surrogate-Based Modeling, ICCS, 2015

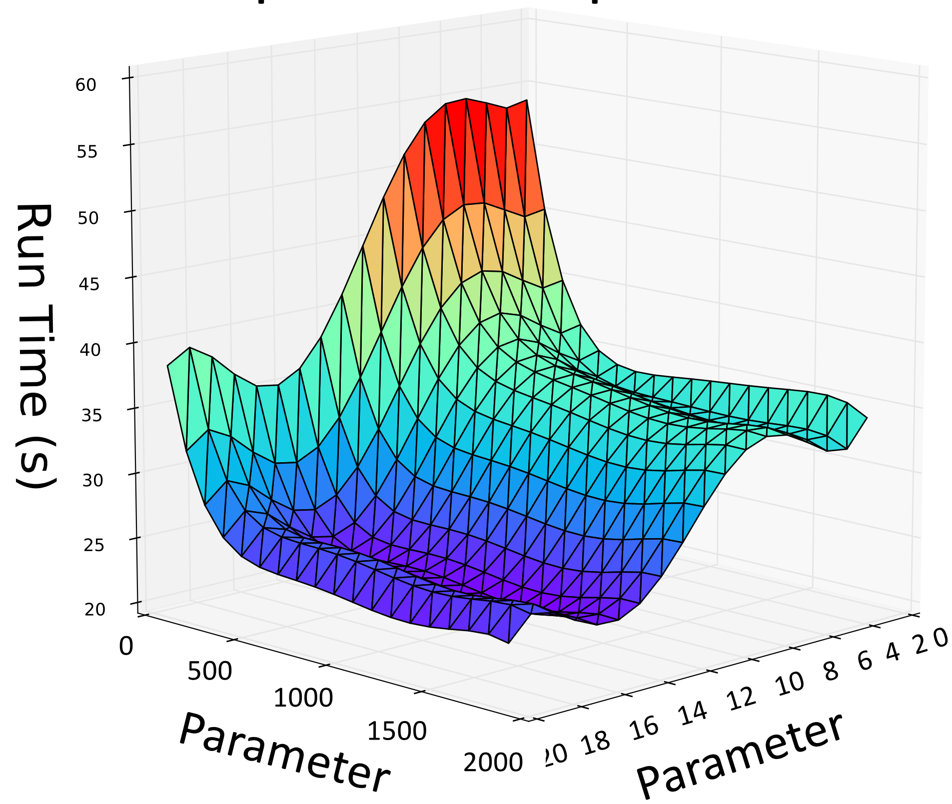
BIG ORANGE BIG IDEAS

Surrogate-Based Modeling

- A single polynomial is globally fit to the parameter space



Sampled data



Polynomial model with degree *d*

BIG ORANGE BIG IDEAS

Surrogate-Based Modeling
+ k Nearest Neighbors
HYPPO

k Nearest Neighbors

- Use **local** data (k points)
- Compute the average (many simple local models)

Surrogate-Based Modeling

- Use **all** sampled data
- Construct one **polynomial** (*single complex global model*)

HYPPO

- Use **local** data
- Construct many **polynomials** (*many local models*)

Travis Johnston, Connor Zanin, Michela Taufer: HYPPO: A Hybrid, Piecewise Polynomial Modeling Technique for Non-Smooth Surfaces. SBAC-PAD 2016: 26-33

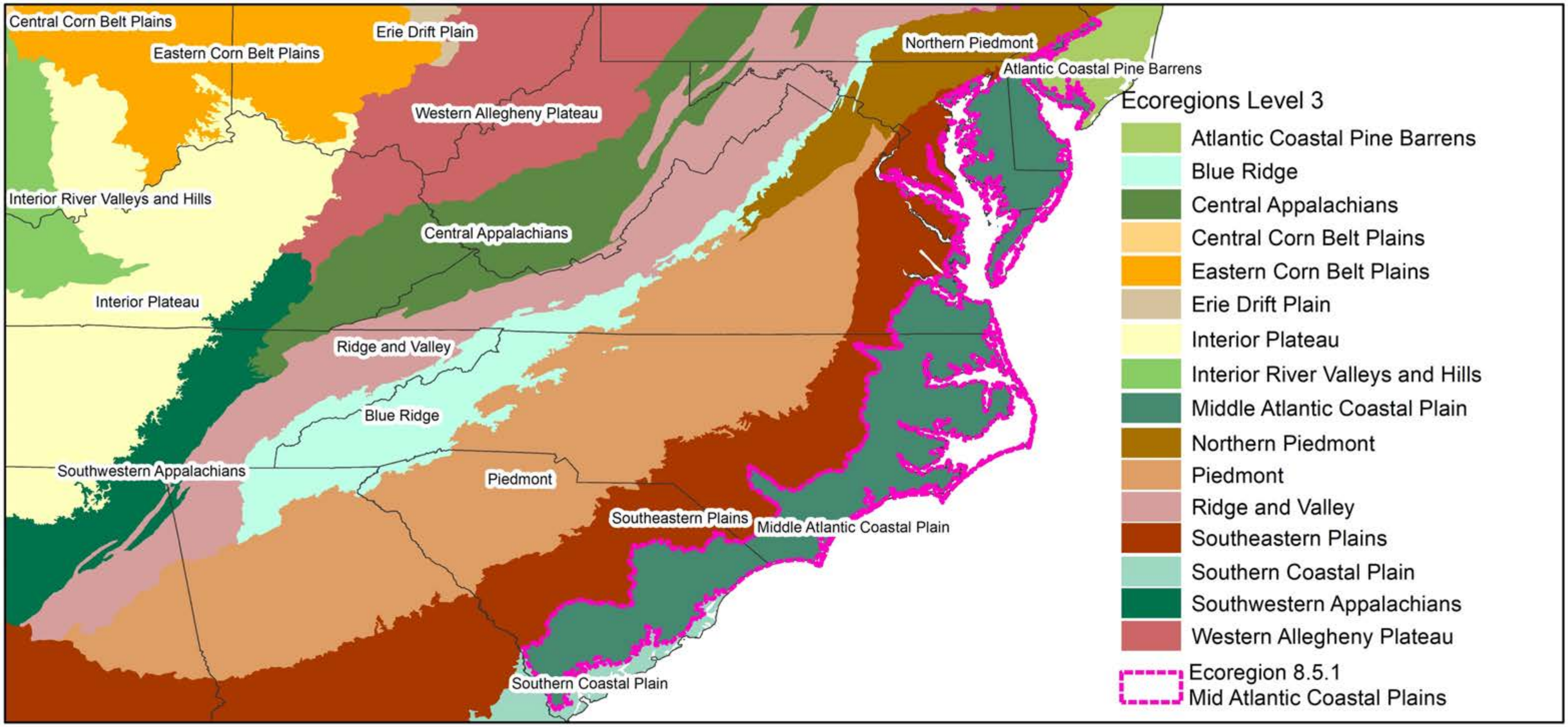
Modeling Methods

- ***K Nearest Neighbors:*** kNN assigns each point in the testing set a soil moisture that is the weighted average of the soil moisture values of its neighbors
- ***Surrogate-based model:*** SBMs are built by fitting a polynomial surface to sampled data points by using e.g., least squares regression
- ***Hybrid Piecewise Polynomial*:** HYPPO combines kNN with a Surrogate Based Model to build a global polynomial model of a surface. HYPPO uses a polynomial approximation in each neighborhood of k nearest points to predict soil moisture

BIG ORANGE BIG IDEAS

Middle Atlantic Coastal Plains

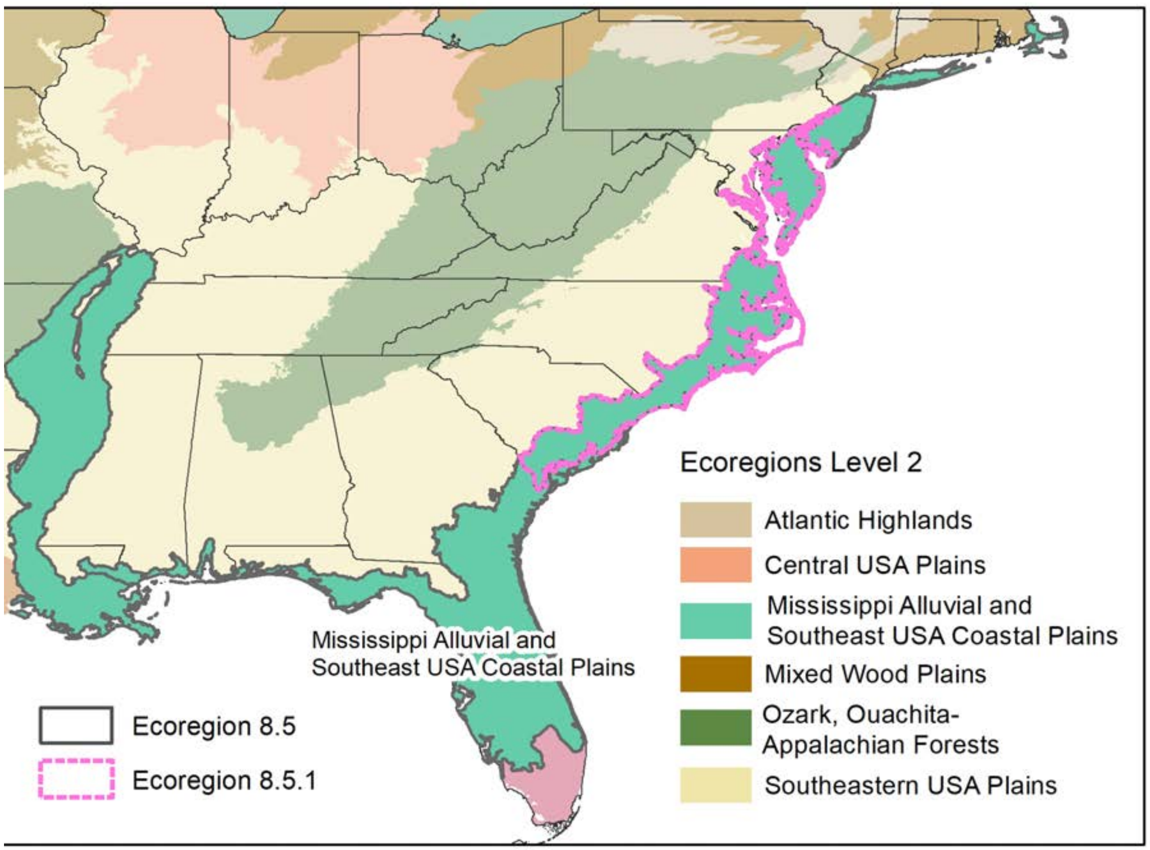
- Region with a broad range of moisture ratios



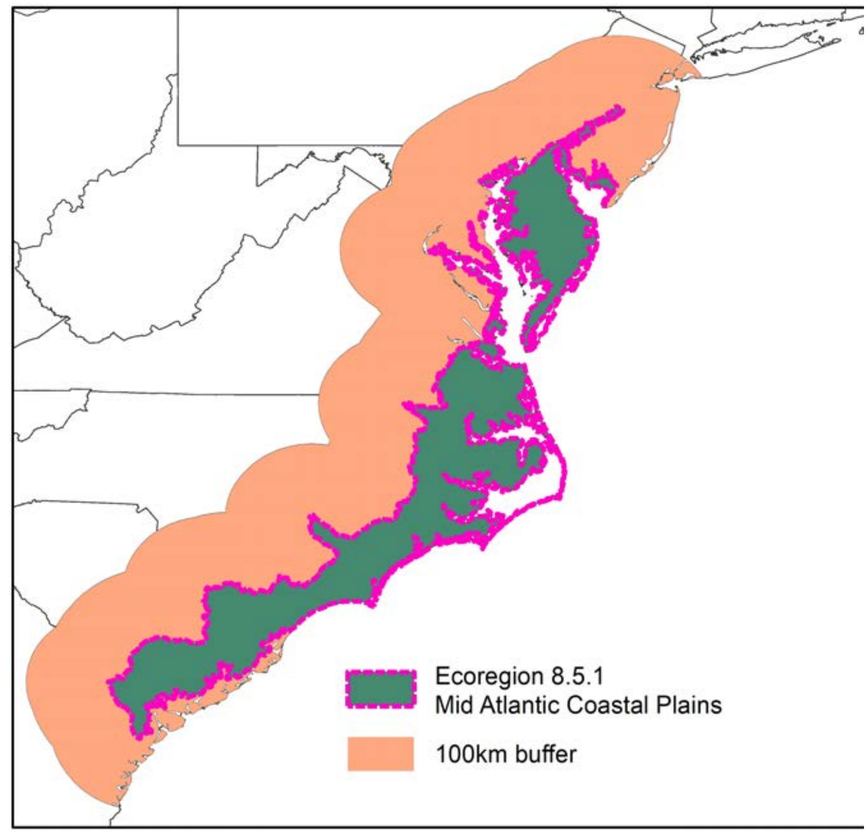
Middle Atlantic Coastal Plains

- Region with a broad range of moisture ratios (not in the assignment)

Level II ecoregion 8.5

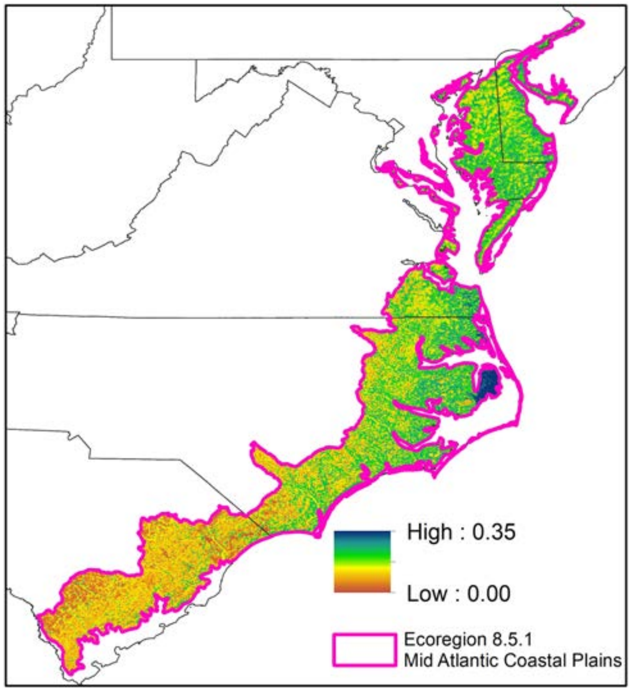


Level III ecoregion 8.5.1 + 100km buffer

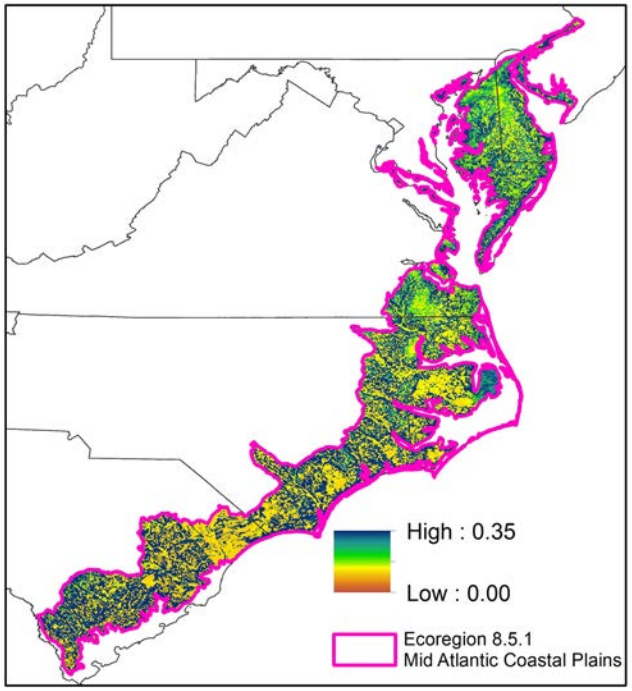


Predictions with different datasets

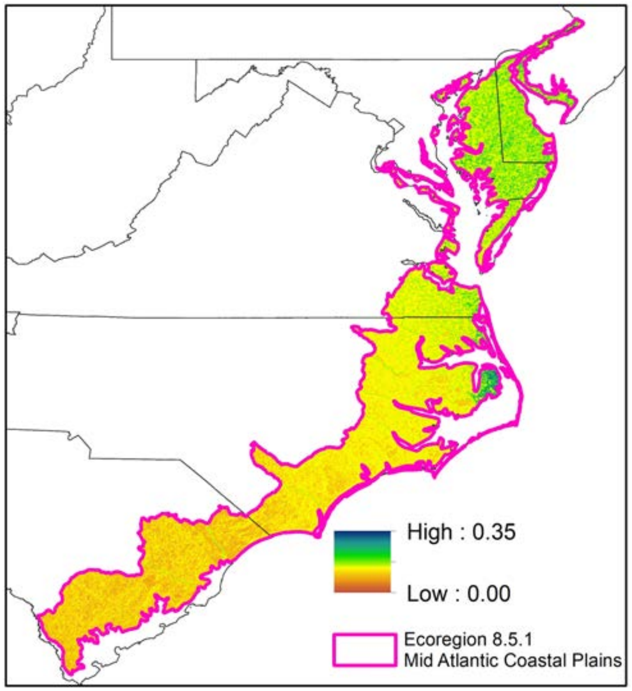
Level III ecoregion 8.5.1



Level II ecoregion 8.5



Level III ecoregion 8.5.1
+ 100km buffer



Assignment 10

Preprocessing data (different from dietary data):

- Problem 1: Remove all monthly columns from a given data frame except for a specified month
- Problem 2: Drop the rows that have an NA in that column
- Problem 3: Use KNN to predict data on the region of interest
- Problem 4: Create heatmaps for the training and prediction soil moisture values
  - Course-grained map: Original Soil Moisture Heatmap
  - Fine-grained map: Predicted Soil Moisture Heatmap

DUE DATE: Nov 19, 2018

BIG ORANGE BIG IDEAS

Project (I)

Motivation Describe the motivation of your work. To build the motivation, you can answer these questions:
What is the problem you are tackling?
How is the problem solved today?
Write a paragraph of 200 - 300 words
Contributions List between 2 and 4 contributions of your work.
Contributions are bullet points that define your solution. E.g.,
We build a system that ....
We validate the system accuracy by ....
We measure the performance of the system by ...
Write a section of 150 - 200 words

DUE DATE: Nov 19, 2018

Project (II)

Tests List the type of tests (measurements) you will perform.
E.g.,
What are your metrics of success?
Where do you run your tests?
What tests do you perform?
How many times do you run each test?
What do you measure?
Write a section of 250 - 350 words.
Slides Keep updating your slides. Submit a new version together with Assignment 10.

DUE DATE: Nov 19, 2018

- **November 12, 2018:** special office hours during lecture time
- Use the time to review assignments, ask questions on the project, debug parts of your code
- **NO TRADITIONAL LECTURE**

THE UNIVERSITY OF
TENNESSEE
KNOXVILLE

BIG ORANGE. BIG IDEAS.®

BIG ORANGE
BIG IDEAS