ESTIMATING SOIL MOISTURE USING MULTISPECTRAL SATELLITE IMAGERY

Patel, S., Thompson, R., Jordan, D., Grant, M.
Data Science Masters Capstone, University of Washington

INTRODUCTION

BACKGROUND

Soil moisture is an important characteristic in agriculture as it has been shown to correlate strongly with plant health and crop yields. However, accurate soil moisture readings are expensive and impractical for capturing high-resolution variability in soil moisture at scale. Remote sensing (e.g. satellite imagery) offers a potential low-cost solution [1].

OBJECTIVE

Determine whether machine learning models can be used to accurately estimate soil moisture with high-resolution multispectral satellite imagery, physical characteristics and environmental factors.

DATA PREPARATION

We gathered data from four fields in Eastern Washington and Western Idaho during the 2012-2014 growing seasons.

IN-SITU MEASUREMENTS

Field measurements provided by our project sponsors consisted of:

- **Crop/soil properties**
  seasonal; 12 sites/field
- **Soil moisture**
  daily; 12 sites/field; 5 depths
- **Weather**
  hourly; field-level

Figure 1: Location of the fields used in this study

MULTISPECTRAL SATELLITE IMAGERY

From Planet Labs' RapidEye multispectral satellite imagery we calculated a vegetative index called Normalized Difference Red-Edge (NDRE) which is based on two wavelengths (Equation 1).

$$NDRE = \frac{NIR - RE}{NIR + RE}$$

Equation 1: Normalized Difference Red Edge (NDRE) equation. NIR = near-infrared (725-760 nm); RE = red-edge (700-725 nm)

NDRE has been shown to respond strongly to chlorophyll in plants [2]; thus, it serves as a good indicator of soil moisture by capturing the health and vigor of surface vegetation.

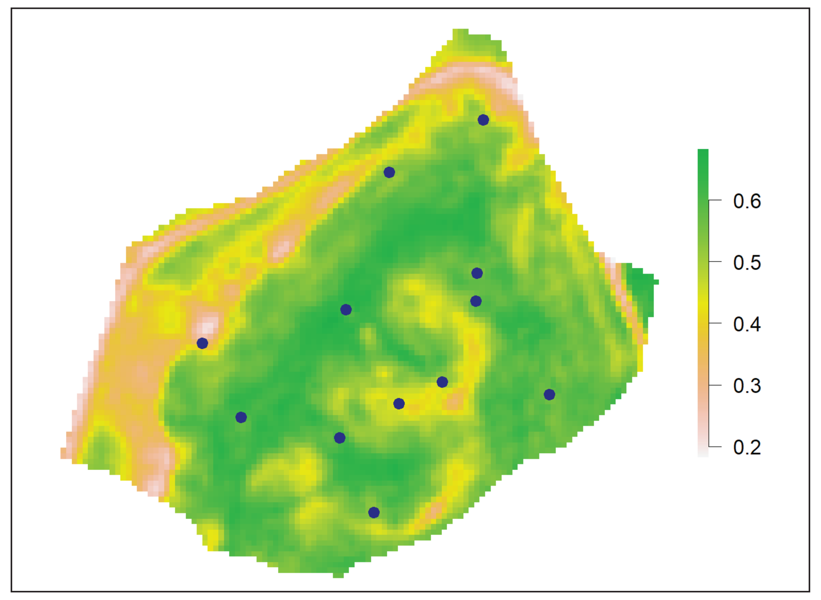

Figure 2: Example NDRE image with twelve monitoring locations. Greener colors represent wetter locations; browner colors represent drier locations. Blue circles represent the monitoring locations.

DATA CLEANING & MERGING

- **NDRE images:** we cropped the images to the farm shapefiles, then averaged the values from the 9 pixels surrounding each of the twelve in-situ sensors
- **In-situ soil moisture:** data was collected at five depths; however, due to excessive missing data, we averaged the values at the top three depths
- **Data merge:** we merged all data on field, date, and sensor location

METHODS & RESULTS

SATELLITE-BASED MODELING

We trained several machine learning (ML) models to estimate soil moisture using NDRE imagery, soil properties and weather; models included:

- **Ridge Regression**
- **XGBoost** (gradient boosted trees)
- **Neural Network** (3-layer network)

We compared the models against a physical Soil Moisture Routing (SMR) model (Figure 3).

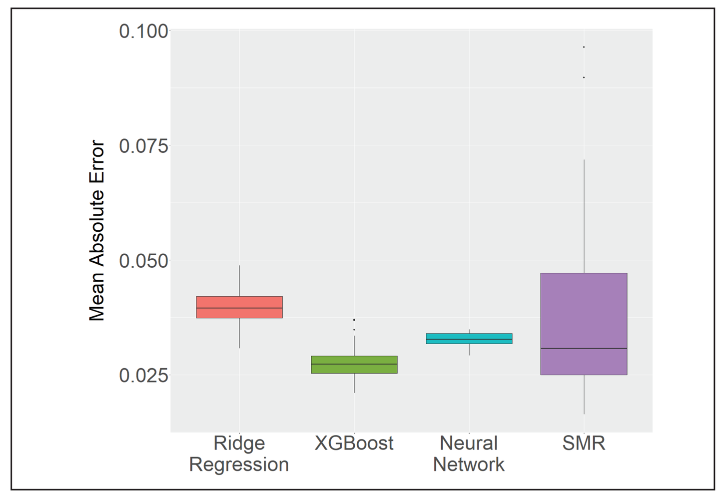

Figure 3: Mean Absolute Error for three machine learning models and the physical Soil Moisture Routing (SMR) model

XGBoost yielded the lowest MAE, at 0.027; this is an improvement over the results obtained from the physical model, which yielded a MAE of 0.035. Thus, we were able to obtain more accurate results with less input data.

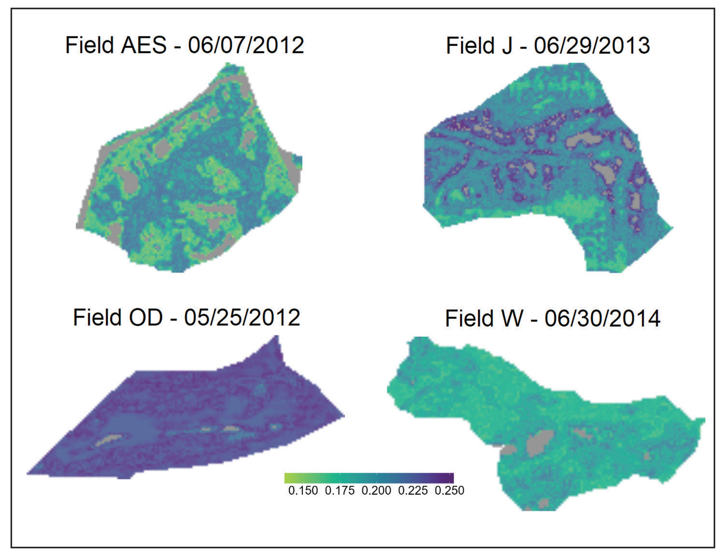

Figure 4: Estimated soil moisture (m^3/m^3) generated by the XGBoost model for each of the four fields

BARE SOIL PREDICTIONS

Because NDRE is only sensitive to vegetation, the XGBoost model is ineffective over bare soil.

To estimate soil moisture in the absence of vegetation, we trained a convolutional neural network (CNN) using the physical characteristics of the soil and weather data. We extended this model to yield two-week soil moisture forecasts. This gives farmers the ability to predict soil moisture weeks in advance even when no vegetation is present. The CNN model yielded a MAE of 0.043.

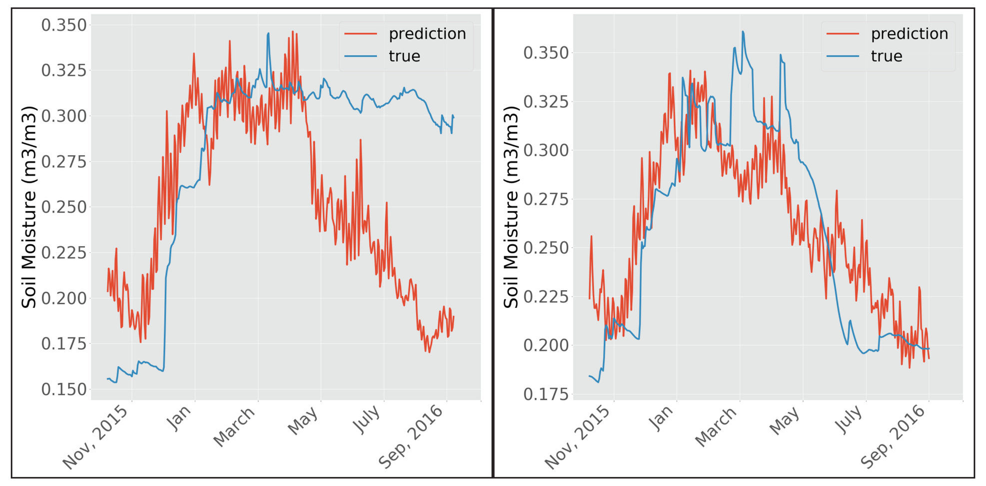

Figure 5: Observed soil moisture (red) compared with soil moisture predictions generated by the CNN model (blue) for two fields

CHALLENGES

ACCURACY VS. SCALABILITY

Throughout this process we had to think about balancing the accuracy of our model with its ability to estimate soil moisture at other locations.

- Several of our predictors came from in-situ measurements (e.g. sand, silt, clay); this data is not easily acquired, so other locations would depend on less precise estimates from regional maps.
- All training data was localized, from the Palouse region; it is unclear how well the models would scale to locations in other parts of the world.

LIMITATIONS

IN-SITU DATA

- Limited to 12 locations per field
- Missing soil moisture data at numerous times/locations/depths

SATELLITE DATA

- Limited images available (~one/week)
- Cloud cover obscured many images
- NDRE ineffective for bare soil conditions

The limitations from each source were amplified when merged; this resulted in a narrow set of training data.

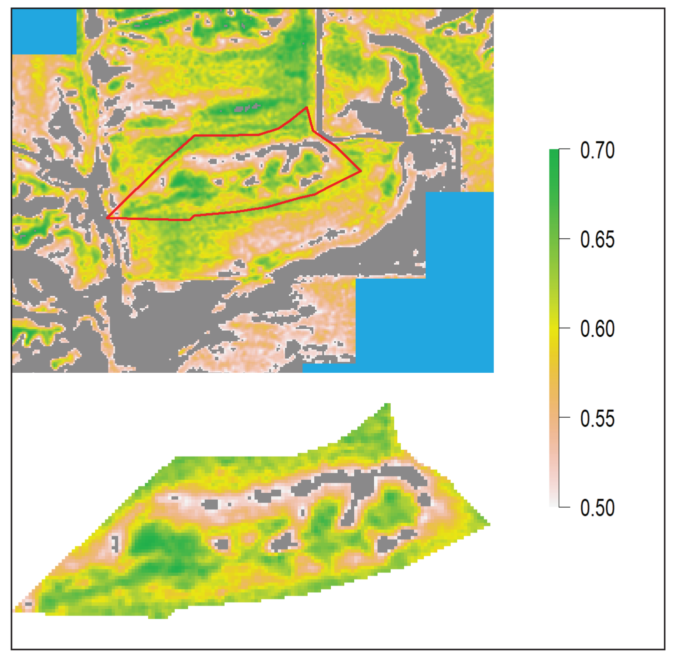

Figure 6: NDRE images with missing data due to cloud cover (blue) and lack of vegetation (gray). The red line in the top image outlines the field; the bottom image shows the data after being subset to the field bounds.

EXTENSIONS & FUTURE WORK

PHYSICAL MODEL AND INTERPOLATION

A physical (hydrological) model could be utilized to estimate soil moisture beyond the 12 sensor locations. While less precise, a kriging or splining method could be used to spatially interpolate soil moisture across the fields where no vegetation or soil moisture readings are present.

BEYOND THE PALOUSE

It is unclear how our models would perform for fields beyond the Palouse. To develop a more robust model, future work would benefit from new training data encompassing a wider range of physical characteristics.

ACKNOWLEDGMENTS

A thanks to our sponsors: **Matt Yourek**, PhD student, Department of Crop and Soil Sciences, Washington State University; **Dr. David Brown**, Professor, Department of Crop and Soil Sciences, Washington State University; **Dr. Erin Brooks**, Professor, Department of Soil and Water Systems, University of Idaho; and our Data Science Capstone advisor: **Dr. Megan Hazen**, University of Washington

REFERENCES

[1] Yourek MA (2016) An Investigation of Crop Senescence Patterns Observed in Palouse Region Fields Using Satellite Remote Sensing and Hydrologic Modeling. University of Idaho.
[2] De Benedetto D et al. (2013) An approach for delineating homogeneous zones by using multi-sensor data. Geoderma 199:117-127

GitHub