Predicting Soil Moisture

PURDUE UNIVERSITY | The Data Mine | JOHN DEERE

Abigail Jones, Soumya Kakarlapudi, Hansheng Liu, Luhanhai Ying, Sparsh Mathur, Erik Ohst, Atharva Thakur, Praneeth Kukunuru

INTRODUCTION

ABOUT:
John Deere is a global leader in the delivery of agricultural, turf, construction, and forestry equipment.

PROBLEM:
Soil moisture impacts many aspects of agriculture. If the soil is too wet, growers must wait until the ground dries to perform tasks such as tilling their land, planting crops, or applying fertilizer. Soil moisture can also affect important plant level outcomes. The issue is it is nearly impossible for growers to get a holistic view of their land's soil moisture throughout a growing season to help them make decisions.

GOAL:
Compile/organize various types of data that are known to have an impact on soil moisture content (i.e., topography, weather, soil type, etc.). Visualize this data effectively to showcase relationships among these attributes. Finally, leverage these variables of interest to create a soil moisture index model that estimates soil moisture for a given location.

COLLECTING DATA

SOIL DATA:

- For soil data, we are using the USDA's SSURGO database accessed via their API. This allows us to access the data with SQL queries
- We are focusing on attributes we believe are relevant for soil moisture
  - Mainly looked at soil composition
  - Future goals include topography (e.g. slope)
- SSURGO contains sand, silt, and clay percentages that we can use for modeling.
- The soil triangle can be used to classify the soil texture depending on the percentages of sand, silt, and clay

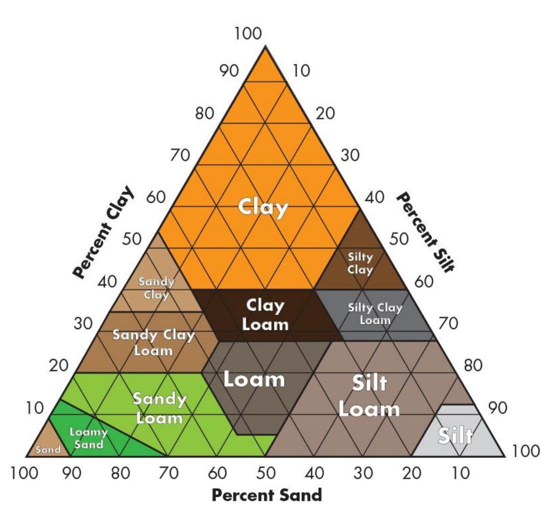


Image source: oercommons

WEATHER DATA:

- Researched accurate and credible data sources for Midwest weather
  - Identified MeteoStat as best option for Python integration with our specific requirements
- Worked with Dr. Bruce Erickson to understand how the weather data affects soil types
- Developed methods and functions to pull weather data accurately and efficiently
- Weather attributes that were used in model include average, minimum, and maximum temperature, precipitation, snow, wind direction and speed, and air pressure

MERGING SOIL AND WEATHER DATA:

- A challenge: weather changes constantly while the soil data doesn't
- Merging soil characteristics with weather data using location – the soil data does not shift as much over time, so we are not focusing on time for merging at present

MODELING

Target:
- Feature extraction: Identify the most important features in the weather variables
- Prediction model: Build the model that effectively predict the soil data

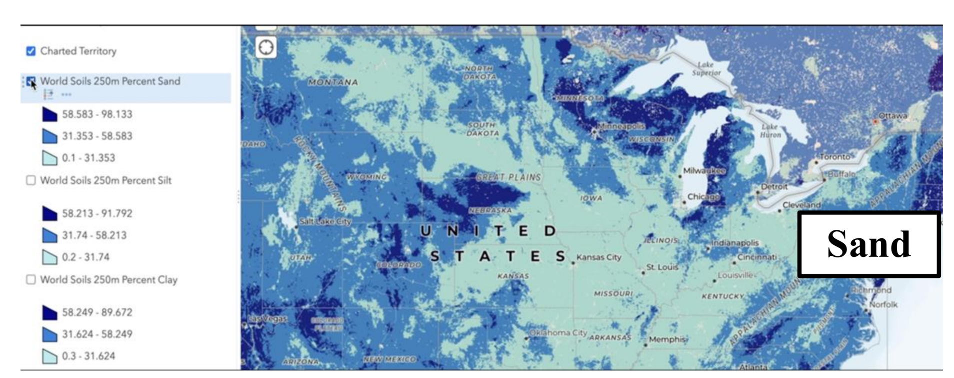
Descriptive statistics:
- Summarize the central tendency, dispersion and shape of the dataset's distribution
- Correlation matrix: feature importance list (temperature > precipitation > other weather variables)

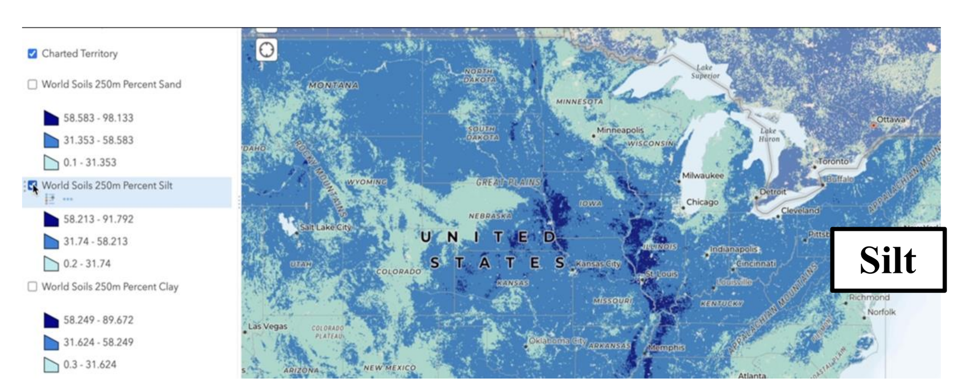
PCA:
- 3 principal components can explain 80% information of 7 weather variables
- Soil type ~ PC1 (Soil Temperature Factor mainly related to temperature) + PC2 (Soil Wind Risk Assessment related to precipitation and wind speed) + PC3 (Soil Precipitation-Wind Direction Pattern related to precipitation and wind direction)

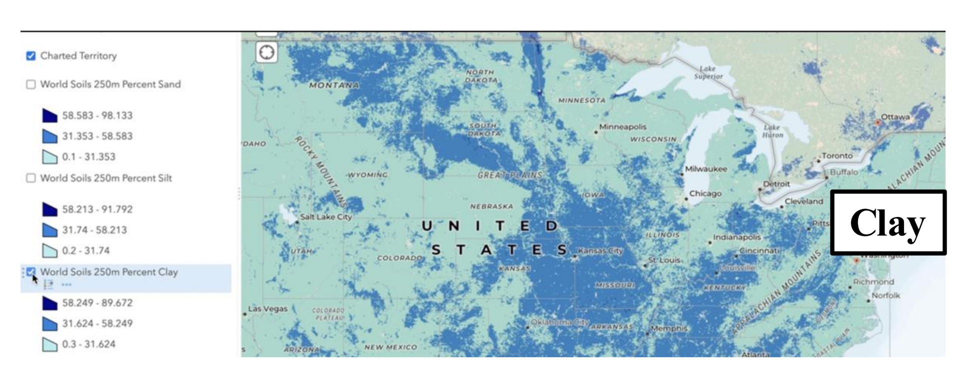
Time-series:
- Constructing time series data for exploring the relationship between date, sensor, and variables.
- XGBoost: The prediction accuracy of R^2 of the obtained model on the test set reaches 0.9, which indicates that the model can effectively capture the time correlation of data and has good prediction ability for future data

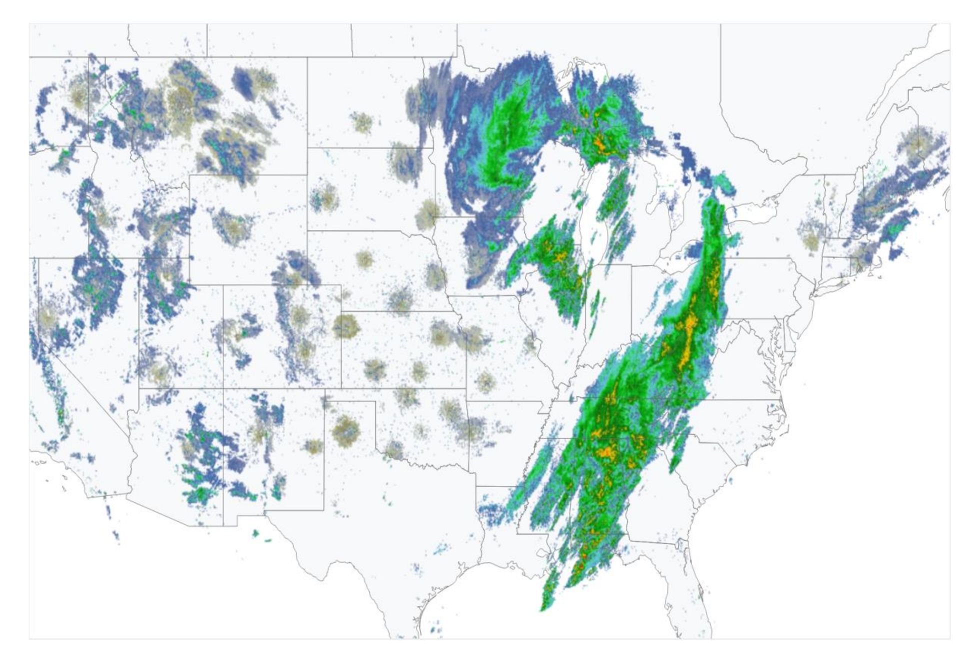
VISUALIZING THE DATA

We utilized visualization resources such as Python libraries, ArcGIS, and Tableau
- Below are :
  - Images of different layers of soil types: sand, silt, and clay to show which areas have more soil type content than others
  - Areas that received more precipitation than others









Precipitation: (Which Areas Receive Most Precipitation)

CONCLUSION

Over the course of the project, our team has:

- Created visualizations that effectively show how soil and weather attributes interact over time and space
- Created a database of soil and weather attributes
- Found which attributes were significant in creating a model
- Achieved a model that can predict future soil moisture data based on the soil type and weather conditions

FUTURE GOALS

These are some future goals we have in mind as we handoff the project to John Deere:
- Test the model extensively to determine the true accuracy of the model
- Expand upon the modeling work to increase the accuracy of the model
- Work on integrating the backend logic of the modeling and the visualization into a user-friendly front-end program
- Train model on larger database so it would be more accurate for more locations

ACKNOWLEDGEMENTS

A special thank you to all the people that helped make this project successful!

- Our Mentors
  - Dylan Roth
  - Hannah Price
- The Data Mine team
- Dr. Bruce Erickson
- Speakers from John Deere
  - Andy Greenlee
  - Stephen Schwartz
  - Adam Seabert
  - Erik Sorensen

DEPENDENCIES

- ArcGIS
- Tableau
- USDA SSURGO
- MeteoStat
- XGBoost

ArcGIS | tableau | XGBoost | meteostat

The Data Mine Corporate Partners Symposium 2024