GCLab

SOMOSPIE

Cyberinfrastructure for Soil Moisture Inference

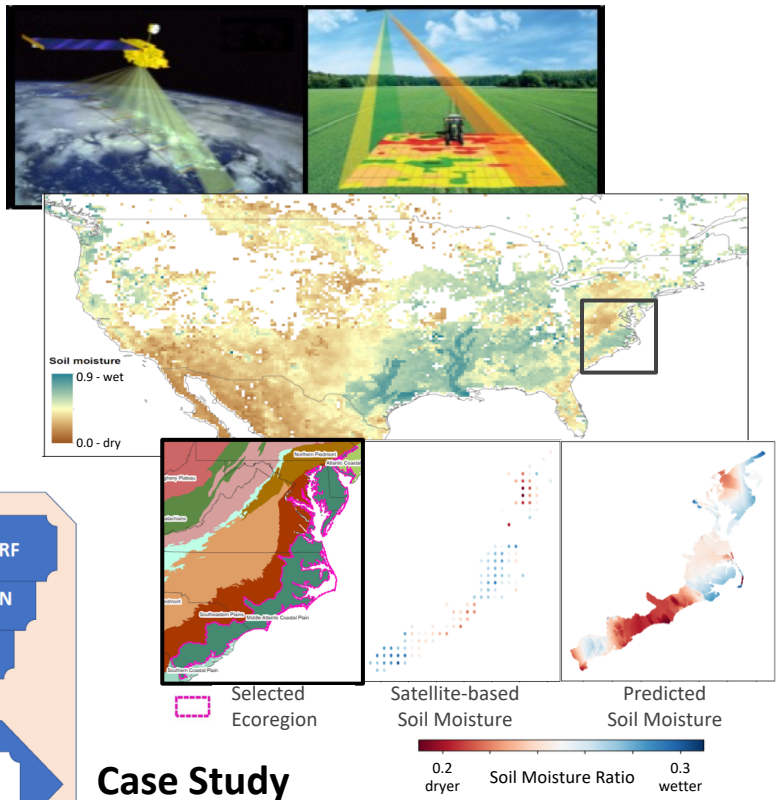
Danny Rorabaugh, Kae Suarez, Michela Taufer @ U. Tennessee, Knoxville
Mario Guevara, Ricardo Llamas, Rodrigo Vargas @ U. Delaware, Newark

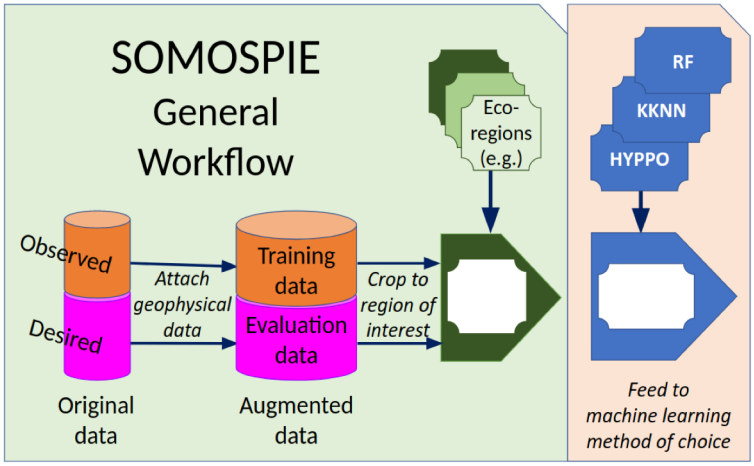
Project Overview

A UD+UTK collaboration of environmental and computer scientists is developing SOMOSPIE—a **SOil MOisture SPatial Inference Engine**—for generating gap-free soil moisture information at finer resolution than available from satellites, the primary source of soil moisture data over large areas. The engine consists of modular stages for processing spatial environmental data and generating predictions with machine learning techniques, especially where soil moisture data are too coarse or sparse for a given need (e.g., precision agriculture).

Methodology

The engine begins its work with soil moisture values attached to the coordinates of the coarse data (i.e., training data) as well as spatial coordinates at the desired resolution (i.e., evaluation data). These data are augmented with related geophysical data that are already available at the desired resolution, such as topographic parameters. Some additional preprocessing tools are available for this augmented data. The processed training data is then fed to machine learning method(s) of interest to generate a model on which the processed prediction data can be evaluated, producing a fine-grain, spatially complete soil moisture prediction.





Case Study

D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas, and M. Taufer, *SOMOSPIE: A modular SOil MOisture SPatial Inference Engine based on data-driven decisions*, IEEE eScience 2019, arXiv:1904.07754.

Validation Elements

Modeling Component

- Cross-validation

Data for Comparison

- "Ground truth" from ground sensors
- Random portion of satellite data set aside from testing data

Lessons Learned

Importance of Region Selection

- Selection of testing region can significantly influence prediction
- Specifically, climate influence can be controlled by region selection

Use of Topography as a Predictor

- Data redundancy can be reduced via PCA without greatly affecting prediction accuracy

THE UNIVERSITY OF TENNESSEE KNOXVILLE
XSEDE Jetstream
NSF
GCL