EECS498 Project Proposal: Investigating Algorithm Modeling Accuracy On Soil Carbon and Nitrogen Through Satellite Data

Stefanos A Frilingos - Senior EE, BSE
Marco Tulio Giachero Pajaro - Senior EE, BSE

1. Motivation

As the charge against climate change becomes more vital by the day, new technology to help this fight is essential. Taking in-person samples to monitor Soil Organic Carbon (SOC) and Soil Total Nitrogen (STN) is too expensive and time consuming. AI technologies that use remote sensing combined with Digital Elevation Model (DEM) data is a more efficient way to estimate soil properties such as SOC and STN, which are used as climate change indicators. Our research uses datasets produced by RADAR, image satellites such as Sentinel-1/2 (S1/S2), topographic data from a diverse region (to be chosen) to evaluate the most accurate Machine Learning algorithms for modeling SOC and STN.

2. Supporting Materials

After reading papers from the *Remote Sensing for Environment* publication, we decided to use the Sentinel datasets provided by the European Space Agency. From S2 images three spectral indices (**N**ormalized **D**ifference **V**egetation, **E**nhanced **V**egetation, **S**oil **A**djusted **T**otal **V**egetation) can be calculated as predictor variables, because of their strong correlation with SOC (Gholizadeh et al. 1997).

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}} \tag{1}$$

(Tucker, 1979)

$$\mathrm{EVI} = 2.5 \times \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + 6 \times \mathrm{RED} - 7.5 \times \mathrm{BLUE} + 1} \tag{2}$$

(Huete et al., 1997)

$$\mathrm{SATVI} = \frac{\mathrm{SWIR}_1 - \mathrm{RED}}{\mathrm{SWIR}_1 + \mathrm{RED} + 1} \times 2 - \frac{\mathrm{SWIR}_2}{2} \tag{3}$$

(Marsett et al., 2006)
BLUE, RED, NIR, SWIR1 and SWIR2 correspond to the wavelength bands B2, B4, B8, B11, and B12 of the S2 image, respectively. We also plan to use SARscape software to filter S1 radar data points with a Lee filter (Lee, 1986) and perform geocoding and radiometric calibration (Zhou et al., 2018a). Topographic measurements will be obtained from SRTM DEM or ASTER GDEM datasets. (Hu et al., 2017; Patel et al., 2016)

3. Plan

We aim to use the aforementioned predictor variables, after preprocessing our dataset, to model and predict SOC and STN in the study area. A 5-fold cross validation method will allow us to objectively test a mixture of the data for a holistic performance. To evaluate algorithm accuracy we will be calculating: determination coefficient, F1-score, root-mean squared error, ratio of performance to deviation, and mean absolute error. Inspired by previous research on modeling soil properties we will select 5 from the following ML algorithms: RF, CART, Bagged CART, XGBoost, BPNN, SVM, BRT, ANN. We will be using 70% of the data for training, 15% for validation and 15% for testing at the end, following the choice of the optimal algorithm. We are considering modeling the optimal parameter inputs for each algorithm by using the 'caret' library in R. If time allows, we would like to duplicate our investigation by performing the same exact modeling on a study area with different soil characteristics and at a different continent, and research algorithm performance change and factors for it.

References

Zhou, et al. “High-Resolution Digital Mapping of Soil Organic Carbon and Soil Total Nitrogen Using DEM Derivatives, Sentinel-1 and Sentinel-2 Data Based on Machine Learning Algorithms.” *Science of The Total Environment*, Elsevier, 13 Apr. 2020

Ma, Guolin, et al. “Digital Mapping of Soil Salinization Based on Sentinel-1 and Sentinel-2 Data Combined with Machine Learning Algorithms.” *Regional Sustainability*, Elsevier, 9 July 2021

Lee, J.-S., 1986. Speckle suppression and analysis for synthetic aperture radar images. Opt. Eng. 25 (5), 255636.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Re- mote Sens. Environ. 218, 89–103.

Huete, A.R., Liu, H.Q., Batchily, K., van Leeuwen, W., 1997. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. Remote Sens. Environ. 59 (3), 440–451.

Marsett, R.C., Qi, J., Heilman, P., Biedenbender, S.H., Carolyn Watson, M., Amer, S., Weltz, M., Goodrich, D., Marsett, R., 2006. Remote sensing for grassland management in the Arid Southwest. Rangel. Ecol. Manag. 59 (5), 530–540.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150.

Hu, Z., Peng, J., Hou, Y., Shan, J., 2017. Evaluation of recently released open global digital elevation models of Hubei, China. Remote Sens. 9 (3), 262.

Patel, A., Katiyar, S.K., Prasad, V., 2016. Performances evaluation of different open source DEM using Differential Global Positioning System (DGPS). Egypt. J. Remote Sens. Space Sci. 19 (1), 7–16.