EVALUATION OF SOIL SPATIO-TEMPORAL VARIABILITY AND COTTON YIELD USING TRADITIONAL SOIL ANALYSIS, REMOTE SENSING, AND MACHINE LEARNING TECHNIQUES

Franciele M. Carneiro
Getulio de F. Seben Junior
Francielle Morelli Ferreira
Leticia B. Santos
Murilo de S. Martins
Fagner Augusto Rontani
Luciano S. Shiratsuchi
LSU AgCenter
Baton Rouge, LA
Paulo Eduardo Teodoro
Unesp, Ilha Solteira, SP, Brazil
Carlos Antonio Silva Junior
Unemat, Sinop, MT, Brazil

Abstract

The soil type, fertility, and physical characteristics affect tremendously cotton yield. Even though there are many techniques used to classify soil through soil type provided from official websites, such as Web Soil Survey (WSS), it is still labor-intensive and costly to map soil characteristics using traditional techniques. One great tool to help in this issue is the use of Remote Sensing (RS), using sensors that can be very helpful to detect the variability of soil classification and complement traditional techniques. The main objective of this research is to evaluate cotton yield and the spatial-temporal variability of soil using traditional (soil fertility and physical properties and data from WSS), proximal soil sensor and remote sensing (passive sensors) data analyzed by machine learning approaches. This work was developed in a commercial field on a farm in Newellton, LA. The cotton was seeded on May 2nd, 2021. The important dataset to be extracted from as planted maps is the RTK GPS elevation (high precision dataset). The experimental design was Latin Square to impose N rates and regular grid sampling of 2.5 acre to collect for soil fertility. The traditional techniques used soil texture and WSS. Soil sensor used was: GSSI Profiler EMP400 (soil electromagnetic induction sensor) at 5, 10, 15 kHz frequencies; and the RS data were collected from satellite images from Planet Scope 2 and Sentinel 2A (passive sensor). Statistical analysis for soil variability was used to make a visual comparison among traditional and sensor-based maps. For training and test results Pearson correlation (Heat map) was used between sensors and cotton yield data, machine learning (Random Forest - RF). The metric parameters were determination coefficient (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RSME). According to RF feature importance, the great value for R^2 was 0.84, using soil and remote sensing data inputs. While for RS bands from satellite imageries the greater correlation between RS bands and cotton yield was NIR and Red.