EGU21-7828, updated on 28 May 2022
https://doi.org/10.5194/egusphere-egu21-7828
EGU General Assembly 2021

Digital Mapping of Soil Particle Size Distribution in an Alluvial Plain Using the Random Forest Algorithm

Fuat Kaya and Levent Başayiğit
Isparta University of Applied Sciences, Faculty of Agriculture, Department of Soil Science and Plant Nutrition, Isparta, Turkey

Soil maps are an important source of data in monitoring natural resources and land use planning. However, in many countries, soil maps were prepared at a reconnaissance level. This detail is not enough for land use planning. Soil texture is one of the most important soil physical properties that affect water holding capacity, nutrient availability, and crop growth. The spatial distribution of soil texture at a high resolution is essential for crop planning and management. Digital soil mapping is the method of spatial data generation with the advantages of current technologies. It supplies fast, accurate, and reproducible results.

In this study, a soil texture map with 30 m spatial resolution was produced for an alluvial plain covering an area of approximately 10,000 ha. In the study, 11 Topographic Environmental Variables obtained from NASA's ASTER Global Digital Elevation model were used. Another input parameters were clay, silt, and sand values determined for 91 soil samples obtained through field studies.

R Core Environment (3.6.1) and related packages were used for environmental variable extraction, modeling, and spatial mapping. For model building, 70 % of data was used and the rest of the data was used for validation. Random Forest Algorithm offers interpretability for pedological information extraction by determining the importance of environmental variables in digital soil mapping. Random Forest Algorithm is preferred because of working in small data sets, harmoniously. The most important topographic environmental variables for clay were elevation, aspect, and slope. For sand, it was the elevation, aspect, and topographic wetness index. And for silt, it was the elevation, slope length, and planform curvature. Root Mean Square Error (RMSE), was used as a model performance measure. In the train data, R^2 values for clay, sand and silt were 0.84, 0.75, 0.85 and RMSE values were 5.23 %, 3.03 %, 5.48 % respectively. In the test data, R^2 and RMSE values were 0.26, 0.11, 0.10 and 11.8 %, 6.74 %, 13.71 % respectively.

There are high differences between RMSE values of training and test data sets. This event may be caused by the small sample size and to be discussed subject in different studies. High resolution (30 m) data of clay, silt, and sand contents can be useful for hydrological studies and for the preparation of land use plans. Digital soil maps can guide policymakers in creating site-specific land management plans. As well as it can be used for monitoring soil fertility and providing ecosystem services. This study revealed important results regarding the use of digital soil mapping in practice with its analytical and statistical accuracy.