



Comparison of some machine learning methods for assessing the percentage of soil organic carbon

Emanuele Barca (1), Giovanna Vessia (2), and Annamaria Castrignanò (3)

(1) Water Research Institute, IRSA –Bari, Italy (emanuele.barca@ba.irsra.cnr.it), (2) University, Campus Universitario - Palazzo Ex Rettorato, Department of Engineering and Geology (INGEO), Chieti Scalo (CH), Italy (g.vessia@unich.it), (3) CNR – ISAFOM , Cosenza, Italy (annamaria.castrignano@entecra.it)

The knowledge about soil organic carbon content plays a primary role in the precision agriculture framework. Nowadays, there is a large availability of powerful methods able to assess the value of such key variable, each of them capable to predict the response under different conditions of size or degree of nuisance of the training matrix. Therefore, having the chance of using different prediction methods can be an added value for users. In the present contribution, two different methods, namely support vector machine (SVM) and random forest (RF), have been used to assess the percentage of soil organic carbon. The estimation has been carried out using as covariates 216 different spectra. By means of the PCA a feature selection has been carried out and the number of covariates has been reduced to five. The training dataset has a size of 90 elements and the test set of 45 in the ratio 2/3 and 1/3 of the total dataset. The error analysis showed that the RF provides better results than SVM method, proving that RF behaves better than SVM with a relatively small training dataset. The dataset comes from the Bonis catchment in Calabria (Southern Italy) and has been collected in the frame of the Alforlab project.