Data-driven analysis of soil quality indicators using limited data

Mansonia Pulido Moncada (1,2), Donald Gabriels (2), and Wim Cornelis (2)

(1) Instituto de Edafología, Facultad de Agronomía, Universidad Central de Venezuela, Maracay, Venezuela
(mansoniapulido@gmail.com), (2) Dept. of Soil Management, Ghent University, Ghent, Belgium (Wim.Cornelis@UGent.be)

The difficult question of which variables to include as a minimum data set of soil quality (SQ) indicators may be simplified by statistical methods, which allow working with databases including categorical and numerical variables commonly used for assessing SQ. The aims of this study were: i) to identify soil structural related parameters that may associate to SQ at different geographic areas and ii) to test the potential power of using decision trees in setting up a framework for SQ assessment, and in determining structural soil properties, visually evaluated, that could be included in the estimation of soil physical properties such as saturated hydraulic conductivity (Ks). SQ was evaluated by visual soil assessment (VSA) in the field and a limited number of physical and chemical soil properties (bulk density (BD), air capacity, plant available water capacity (PAWC), saturated hydraulic conductivity (Ks), water stable aggregates (WSA), particle size distribution, soil organic carbon (SOC) and cation exchange capacity (CEC)) determined in the laboratory. Using categorical and numerical data of those physical, chemical and morphological properties of soils in both tropical and temperate areas, classification trees and model trees were grown. Parameters related to SQ differed between geographic areas. Ks was the strongest variable determining the SQ in 'tropical' soils, but WSA, SOC and PAWC were also key variables in determining differences in SQ. For 'temperate' soils PAWC was the only variable selected by the tree building algorithm. SOC, clay, and CEC were the discriminating variables of the model constructed from the combined data set. Statistically significant relationships between measured and visual parameters are promising in demonstrating the SQ description required for merging morphological, physical and chemical properties for minimum data set of SQ indicators. Thresholds of different predicting variables could be better established when SQ frameworks involve VSA. We also proved that prediction of KS with a model tree was more accurate when morphological parameters were included as predictor variables, and that in that case the model tree showed a simpler structure compared to the tree built only from chemical and physical soil properties. In conclusion, decision trees are encouraging in the selection of SQ indicators. Moreover, including morphological properties in the prediction of key soil properties such as Ks seems promising. VSA could render morphological response variables for predicting other soil properties and developing SQ frameworks (agricultural interest) more capable of representing structural dynamic.