Geoderma 235–236 (2014) 271–278

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Data-driven analysis of soil quality indicators using limited data

Mansonia Pulido Moncada [a,b,*], Donald Gabriels [b], Wim M. Cornelis [b]

[a] Instituto de Edafología, Facultad de Agronomía, Universidad Central de Venezuela, Av. Universidad vía El Limón, Maracay 2101, Aragua, Venezuela
[b] Department of Soil Management, UNESCO Chair on Eremology, Ghent University, Coupure links 653, B-9000 Ghent, Belgium

ARTICLE INFO

Article history:
Received 13 February 2014
Received in revised form 3 July 2014
Accepted 13 July 2014
Available online 3 August 2014

Keywords:
Decision trees
Soil structure
Visual soil assessment
Saturated hydraulic conductivity
Soil properties

ABSTRACT

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 (K_s). 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 (K_s), 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. K_s 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 K_S 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 K_s 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.

1. Introduction

Soil quality (SQ) is defined as 'the capacity of the soil to function within ecosystem and land-use boundaries to sustain biological productivity, maintain environmental quality, and promote plant and animal health' (Doran et al., 1996). The capacity of soil to function can be reflected by measured soil physical, chemical and biological properties, also known as soil quality indicators (SQIs) (Shukla et al., 2006).

Soil organic carbon (SOC) has been found to be the most important governing factor for monitoring SQ changes (Shukla et al., 2006). However, it is unlikely that a sole ideal indicator can be used for assessing SQ in any soil condition because of the multitude of properties involved and the dynamic condition of soils. Therefore, 'SQIs based decision tools that effectively combine a variety of information for multi-objective decision-making are needed' (Karlen and Stott, 1994).

Overall SQIs are intended to make complex information more accessible to decision makers. However, their applicability can be restricted not only to different soil types but also to multiple regions and management systems because of the site-specific nature of some SQIs (Andrews et al., 2003). Therefore, SQIs selected for evaluating soil functions must be truly representative of the complexity of the soil.

SQ is strictly related to soil structure and much of the environmental damage in intensive arable lands originates from soil structure degradation (Pagliai et al., 2004). Hence, soil structure as the most complex property of the soil is a key factor in the functioning of soil (Mooney et al., 2006).

The concept of the minimum data set of SQIs that reflects sustainable management goals and specific soil structure conditions is widely accepted, but has relied primarily on expert opinion to select minimum data set components (Doran and Parkin, 1994; Karlen et al., 1997;

* Corresponding author at: Department of Soil Management, UNESCO Chair on Eremology, Ghent University, Coupure links 653, B-9000 Ghent, Belgium.

E-mail addresses: mansonia.pulido@UGent.be, mansoniapulido@gmail.com (M. Pulido Moncada), donald.gabriels@UGent.be (D. Gabriels), wim.cornelis@UGent.be (W.M. Cornelis).

http://dx.doi.org/10.1016/j.geoderma.2014.07.014
0016-7061/

Larson and Pierce, 1991). The difficult question of which variables to include in an index of SQ may be simplified by statistical methods (Andrews et al., 2002). Soil structure is usually described in situ using classes or categories rather than continuous variables. Such soil structure classes cannot be used directly in classical statistical regressions for estimating soil properties from others (Pachepsky and Rawls, 2003), but techniques for developing tree-based models or decision trees enable us to work with databases including categorical and numerical variables (Clark and Pregibon, 1992). These are exploratory techniques based on uncovering structure in data, and partition the samples to find both the best predictors and best grouping of samples. Decision trees derive knowledge rules from the data that subsequently can be used to estimate the impact of proposed measures.

Decision trees are familiar to pedologists because the main output is similar to most soil classification schemes. These techniques have been successfully used to explore databases containing categorical and numerical variables in some branches of soil science (McKenzie and Jacquier, 1997). For instance, in agro-ecology, decision trees have been used to evaluate how population dynamics of soil organisms is affected by changes of different biological and physicochemical environmental attributes and agricultural practices (Debeljak et al., 2007).

In soil physics, their use has been mainly restricted to predicting soil hydraulic properties. For instance, Pachepsky and Rawls (2003) found that qualitative morphological observations of soil could be translated into quantitative soil hydraulic parameters, using a classification tree (tree-based model). The authors also demonstrate from decision trees the usefulness of the grade of structure as a predictor of water retention, which indicates a potential for observed aggregate-size distribution to be used in pedotransfer functions (PTFs).

Despite the effort done to include morphological properties of soil structure as potential predictors of the soil hydraulic properties (Lilly et al., 2008; Pachepsky and Rawls, 2003; Vereecken et al., 2010), thus far no unified approach exists on how to best include structural properties in PTFs. According to Vereecken et al. (2010) soil structure predictors in particular can suffer from the absence of a uniform protocol or definition, or may depend on the experience of the observer. However, the visual examination and evaluation of soil structure methods (Ball et al., 2007; Mueller et al., 2009; Shepherd, 2009) could be considered for collecting dependable morphological data for predicting other soil properties.

We hypothesized that the use of such decision tree approaches that relate morphological, physical and chemical soil properties to soil structure, hence SQ, enables the possibility of developing SQ frameworks more capable of representing structural dynamics in specific environments. The objective of this study was to identify soil structural related parameters that may be linked to SQ at different geographic areas and to test the potential power of using decision trees in setting up a framework for SQ assessment, using a limited number of categorical and numerical variables from both 'tropical' and 'temperate' soils.

2. Materials and methods

2.1. Study area and data collection

Ten soils were selected, with six located in a tropical environment (V_1–V_6; central-northern part of Venezuela) and four in a temperate one (B_1–B_4; Flanders Region of Belgium) (Table 1). In the tropical area the data set was collected from a SQ evaluation study in different agricultural areas, where a large part of the country cereal and vegetable production takes place (Pulido Moncada et al., 2014a). The temperate data set includes samples taken from different soil types, which represent common soils in the Flanders Region. The fields selected differ in factors that affect SQ such as soil type, soil management and vegetation type (Table 1), which provide a wide range of SQ. At the sampling moment, the water content of the soils was near field capacity and crops were in flowering.

Three transects were randomly laid out along the soils. Disturbed and undisturbed soil samples were taken at two spots in each transect (36 and 24 observation points in the tropical and temperate areas, respectively). At each spot, disturbed samples and soil blocks for visual examination were taken from the upper layer to 20 cm depth, whereas core samples to 10 cm depth.

2.2. Physical, chemical and morphological soil properties

In this study, properties most frequently evaluated when assessing SQ were selected as measured properties.

At each spot, three 100 cm^3 core samples were taken and used to determine the following soil properties: i) dry bulk density (BD) based on the core method; ii) air capacity (AC, $\theta_{h=0\ cm}$–$\theta_{h=-10\ kPa}$) and plant

Table 1
General description of the 'tropical' (V_1–V_6) and 'temperate' (B_1–B_4) soils.

Soil	Textural class	USDA class (Soil survey staff, 2010)	Geographic coordinates	Drainage status[a]	Soil use and management[b]
V_1	Sandy clay loam	Typic kandiustult	10° 22′ N 67° 12′ W	Well drained	Fruit cropping, no-till
V_2	Clay loam	Fluventic haplustoll	10° 15′ N 67° 37′ W	Well drained	Permanent pasture, no-till, no trampling
V_3	Loam	Typic endoaqualf	10° 21′ N 68° 39′ W	Imperfectly drained	Maize mono-cropping, conventional tillage
V_4	Loam	Aquic haplustoll	8° 46′ N 67° 45′ W	Moderately well drained	Grazing, no-till, permanent cattle
V_5	Silt loam	Typic rhodustalf	9° 0′ N 67° 41′ W	Moderately well drained	Cereal crops with fallow periods, conventional tillage
V_6	Silty clay	Aquic haplustalf	9° 02′ N 67° 41′ W	Moderately well drained	Grazing with natural vegetation, trampling
B_1	Sandy loam	Inceptisol	50° 59′ N 3° 31′ E	Well drained	Cereal mono-cropping, conventional tillage
B_2	Silt loam	Alfisol	50° 46′ N 3° 35′ E	Moderately well drained	Cereal mono-cropping, conventional tillage
B_3	Silt loam	Alfisol	50° 47′ N 3° 25′ E	Moderately well drained	Rotation of corn and winter wheat, conventional tillage
B_4	Loam	–	50° 47′ N 2° 49′ E	Well drained	Rotation of cereal and grass, reduce tillage, no trampling

Conventional tillage in these areas in Venezuela can be described as multiple passes of the harrow and plough during each cultivation period as well as a yearly or a two year subsoiling, whereas in Flanders conventional tillage comprised primary tillage with cultivator + mouldboard plough, followed by secondary tillage with harrow and seed drill.

[a] The soil drainage class indicates the possibility to evacuate excess of water from a soil based on the soil unit's classification name. The FAO soil drainage classes are: not applicable; excessively drained; soils extremely drained; well drained; moderately well drained; imperfectly drained; poorly drained; very poorly drained; water bodies.

[b] Current and over the last 10 years.