Catena 113 (2014) 276–280

Contents lists available at ScienceDirect

Catena

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

Taxonomic distance between South African diagnostic horizons and the World Reference Base diagnostics

C.W. van Huyssteen [a,*], E. Michéli [b], M. Fuchs [b], I. Waltner [b]

[a] Department of Soil, Crop and Climate Sciences, University of the Free State, P.O. Box 339, Bloemfontein 9300, South Africa
[b] Department of Soil Science and Agricultural Chemistry, Szent István University, 2103 Gödöllő, Páter Károly u. 1, Hungary

ARTICLE INFO

Article history:
Received 2 March 2012
Received in revised form 26 July 2013
Accepted 12 August 2013

Keywords:
Classification
Diagnostics
Euclidean distance
Morpho-genetic
Property-based

ABSTRACT

A myriad of soil classifications exist internationally. These usually cater for unique national variations and conditions. These different classification systems, however, hinder international communication. This paper attempted to relate the South African Soil Taxonomic (SAT) soil classification system with the World Reference Base for Soil Resources (WRB) through taxonomic distance classification. A probability matrix of the presence of selected identifiers of the diagnostic elements (properties, horizons, materials) of the South African classification system and the WRB was constructed to determine the taxonomic relationships between them. Euclidean distance calculation on these data enabled numeric expression of the taxonomic similarities and dissimilarities between the South African and WRB diagnostics. Results proved encouraging and some recommendations can be made. For example, a >20% OC family for the organic O, as well as stagnic and gleyic families for the G horizon is proposed. It is further proposed that the WRB consider recognition of red apedal B, yellow-brown apedal B, and lithocutanic B horizons. Since the compared units are the basic building blocks of the two systems, the results presented here can be useful in the relation of soil classification in the South African Soil Taxonomy to the WRB.

1. Introduction

Humans classify the objects in their environment to create order, to reduce variability, to increase understanding, and to inventorise (Krasilnikov et al., 2009). A myriad of soil classification systems were developed, probably due to the relative young nature of the science, the unique local variation of soils, and lack of means of easy information exchange. A further consideration might be that scientists in each region consider different soil properties with varied interest. This situation creates obvious challenges for comparing soils and international communication.

The United States Department of Agriculture Soil Taxonomy (Soil Survey Staff, 2010) and the World Reference Base for Soil Resources (WRB; IUSS Working Group WRB, 2006) are used for international communication. The WRB has, however, been adopted as preferred soil correlation system by the International Union of Soil Sciences and the European Union (Jones et al., 2005).

The WRB is based on diagnostic horizons, properties, and materials, each with strict differentiating quantitative criteria and definitions. The WRB has two tiers: 32 reference soil groups (RSGs), determined by a key and qualifiers (that are accommodated as prefixes or suffixes to the RSG). Although it is stated that the differentiating criteria "should be measureable and observable in the field" the reality is quite different and in many cases almost complete chemical and physical analysis of the soil profile (horizons) is required to perform a classification. This is especially true where the local soil properties are not known.

Soil Classification — A Taxonomic System for South Africa (SAT) is used exclusively in South Africa (Soil Classification Working Group, 1991). The system follows a morpho-genetic approach, similar to that proposed by Kubiëna (1953). Very few chemical or physical analyses are therefore required to classify the soil. This feature makes the classification system uniquely suited to the low-technology (resource poor) environment experienced in Africa. The South African Taxonomy defines five topsoil and 25 subsoil diagnostic horizons, combinations of which give rise to 73 soil forms. The soil forms are subdivided into soil families, based on 19 sets of distinguishing properties. The final classification should also include the soil depth and topsoil texture.

The challenge is, however, to relate the South African Taxonomy with the WRB. Efforts in this regard are severely hampered by the differences in approaches (principles) between the two systems (morpho-genetic vs. property-based).

Taxonomic distance calculations, first promoted by Adanson (1763), are based on measures of similarity often applied in the phenetic view of numerical taxonomy, where the relative similarities or dissimilarities are measured based on different attributes without *a priory* weighting (Dunn and Everitt, 1982; Jardine and Sibson, 1971; Sneath, 1962; Sokal and Sneath, 1963).

Taxonomic distance calculation in soil science was first proposed by Hole and Hironaka (1960). Numerical soil classification has been applied

* Corresponding author. Tel.: +27 51 401 9247; fax: +27 51 401 2212.
E-mail addresses: vanhuyssteencw@ufs.ac.za (C.W. van Huyssteen), Micheli.Erika@mkk.szie.hu (E. Michéli), Fuchs.Marta@mkk.szie.hu (M. Fuchs), Waltner.Istvan@mkk.szie.hu (I. Waltner).

0341-8162/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.catena.2013.08.010

Table 1
Physical, chemical, and morphological properties of soils selected from the WRB diagnostic criteria that were used as dominant identifiers and coded against SAT and WRB diagnostics.

Physical properties	Chemical properties	Morphological properties
Consolidated material	Organic material saturated <30 consecutive days	≥5% Red/black nodules
<20% Cracks	Organic material saturated ≥30 consecutive days	>15% Nodules
>10 cm Horizontal crack spacing	Organic carbon accumulation	>40% Nodules
Vertical cracks >10 cm apart	<0.5% Organic carbon	Indurated nodules or plates
<50% (v/v) Rock structure	>0.6% Organic carbon	>5% Oximorphic colours
<80% Gravel	<5.9 pH	>90% Reductimorphic colours or
Irreversible hardening	<50% BS	≥15% Red mottles or
>50% Indurated	>50% BS	Dark/red colour
Indurated by $CaCO_3$	Effervescence with HCl	Darker than overlying
<10% Dipersible clay	Strong effervescence with HCl	Grey dry colour
<10% Weatherable minerals	>0.50% AlO + FeO	Grey moist colour
≥4.5 MPa penetration resistance	>15% Exchangeable sodium percentage	Munsell value and chroma ≤3
≥4 MPa penetration resistance	>2% Calcium carbonate equivalent	Pale pedfaces and bright interiors
≥50% Slaking	Calcium carbonate equivalent ≥15%	Evidence of clay illuviation (clay bridging/films)
Clay increase	Cation exchange capacity <16 $cmol_c\ kg^{-1}$ clay	≥5% (v/v) Secondary carbonates
Doubling of clay content over 7.5 cm	Si accumulation	>25% Stratification
≥8% Clay		≥10% (v/v) Si nodules/fragments
≥30% Clay		Under albic
Very fine sand, loamy very fine sand or finer texture		Wedge-shaped peds
Loamy sand or finer texture		Slickensides
		Massive, blocky, columnar or prismatic structure
		≥10 cm Peds

in multiple studies since, including some with distance metrics. Most of these studies, however, focused on smaller areas and/or datasets and therefore had only limited applicability or scope (Bidwell and Hole, 1964a,b; McBratney et al., 2000; Sarkar et al., 1966). National and international application of the method has not been published for almost another decade (McBratney et al., 2009). Carré and Jacobson (2009) incorporated distance metrics into their OSACA model application and used it to allocate soils to existing classifications or to derive centroids and create new classifications based on clustering.

Minasny et al. (2009) were first to apply taxonomic distance metrics at an international level for the WRB Reference Soil Groups. Their concept-based approach focused on the dominant identifiers, or diagnostic criteria, of the system instead of deriving centroids.

With the centroid-based approach the taxonomic distance metrics are generally based on actual numerical data, where a centroid is calculated for the group in question and for each selected attribute ("dominant identifier"). However, distance calculations can also be based on the presence or absence of certain features, derived from scientifically sound concepts (Minasny et al., 2009). These concept- and centroid-based approaches were enhanced for the semi-quantitative analysis and correlation of different classification systems (Láng et al., 2010, 2013).

Probably the simplest and most common method to calculate taxonomic relationships is the Euclidean distance (Dunn and Everitt, 1982; Webster, 1977). The taxonomic distance (d_{ij}) is based on Pythagoras' theorem, so for points x_i and x_j with two variables it can be expressed as:

$$d_{ij} = \sqrt[2]{\left(x_{i1} - x_{j1}\right)^2 + \left(x_{i2} - x_{j2}\right)^2} \quad (1)$$

Extending the same principles to multiple dimensions (representing multiple variables) gives:

$$d_{ij} = \sqrt[2]{\sum_{k=1}^{p} \left(x_{ik} - x_{jk}\right)^2} \quad (2)$$

where p is the number of dimensions observed.

In application to soil classification, the taxon being considered forms the centre and the distance of the other taxa from the centre are calculated based on the selected soil properties (diagnostic criteria in this instance). Interpretation can therefore not be made between the different taxa, but only between the two (i.e. the central and one additional) taxa under consideration. For example, two taxa may have the same distance from the centre, but may be equally far from each other, because the direction of the departure is not specified.

The purpose of this paper was to explore the relationship between diagnostic horizons in the South Africa Soil Taxonomy and the diagnostic horizons, materials, and properties in the WRB.

2. Methods

This taxonomic distance calculation focussed only on relating 33 possibly related WRB diagnostic horizons, properties, and materials with the SAT diagnostic horizons. The WRB qualifiers and SAT family criteria were therefore excluded. In addition the fulvic and melanic horizons, andic and vitric properties, and tephric material were excluded, because recent pyroclastic deposits are not known in South Africa. All the anthropogenic diagnostics in the WRB and SAT were excluded. Firstly because the hortic, irriagric, plaggic, terric or anthraquic horizons are not known to exist in South Africa and secondly because the diagnostics for the man-made soil deposit are not scientifically well developed. The physical, chemical, and morphological properties selected as dominant identifiers are presented in Table 1.

The taxonomic distance calculation was done on the concept based approach (Láng et al., 2010; Minasny et al., 2009), by constructing a coded matrix table. The matrix table expressed the probability that a property (selected as dominant identifier) must be present (code: 1), cannot be present (code: 0), or is likely to be present (code: 0.5) in the compared diagnostic elements. The codes are the probability values (0, 0.5, 1) of the presence of the selected dominant identifier (physical, chemical, and morphological criteria) of the selected diagnostics. The assignment of codes was based on the soil properties and diagnostic criteria in the compared systems (IUSS Working Group WRB, 2006; Le Roux et al., 1999; Soil Classification Working Group, 1991), personal experience, and expert judgement (the coded matrix and calculated taxonomic distances are available as online supplementary material.)

Taxonomic distances were calculated, based on the matrix table, as the Euclidean distance between the different taxa by using the R software package (Baier and Neuwirth, 2007).

Mathematically a taxonomic distance value of 0.0 indicates an exact similarity. Assessment of the data has shown that values less than 1.0 indicate large similarities, while values approaching 2.0 and above indicate large differences.