Geoderma 264 (2016) 340–349

Contents lists available at ScienceDirect

Geoderma

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

Testing the pedometric evaluation of taxonomic units on soil taxonomy — A step in advancing towards a universal soil classification system

Erika Michéli [a], Vince Láng [a,*], Phillip R. Owens [b], Alex McBratney [c], Jon Hempel [d]

[a] Department of Soil Science and Agricultural Chemistry, Szent István University, Gödöllő, Hungary
[b] Department of Agronomy, Purdue University, West Lafayette, IN, USA
[c] Department of Environmental Sciences, The University of Sydney, Australia
[d] USDA NRCS National Soil Survey Center, Lincoln, NE, USA

ARTICLE INFO

Article history:
Received 16 April 2015
Received in revised form 14 September 2015
Accepted 16 September 2015
Available online 28 September 2015

Keywords:
Universal soil classification
Centroids
Soil taxonomy
Taxonomic distance

ABSTRACT

Most existing soil classification systems were developed to understand and provide information on soils, their natural properties and potential use for certain purposes. The conceptual developments of the systems took place before the recent boom of observation technologies, data storage and data processing achievements that can support to determine or predict soil differences. Until the recent past soils under agricultural or forestry use received more attention than other soils, such as anthropogenically modified or urban soils, or soils of the cold regions. The broader view of soil functions and the understanding of global environmental processes require a better understanding and description of all soils. Precisely recorded and harmonized data is needed to serve the new era of modern agricultural practices, other land uses as well as different scientific applications. The soil science community is challenged to apply the accumulated knowledge on soil formation, soil differences and functions, as well as, new tools of robust data processing to evaluate current systems and define objective relationships for better future classification systems. The evaluation of existing soil classification systems may help the understanding of taxonomic relationships of differentiated soil groups and improve our methods of classifying soils. This paper is summarizing the approaches and methods of evaluation that was applied for the great group and higher levels of Soil Taxonomy. Simple statistical and pedometric methods were applied on centroids and, calculated on the basis of properties commonly used to define the classification units. The centroids provide an objective tool to evaluate the concepts of taxa and the taxonomic relationships between them. Examples of conceptual evaluations and detailed discussions of the taxonomic distance calculations between the great groups within their orders and members of other orders are provided. The presented methods and relationships were found very useful for the evaluation purpose. The extension of the methods for other systems, other data bases and the combination of those is in progress. The initial results suggest that the objective, pedometric approaches can support the development of an envisioned Universal Soil Classification System.

1. Introduction

In the past century, most developed countries elaborated their own classification systems, based on the genetic principles of Dokuchaev (1883). International communication and data exchange, collections of data and compilation of regional and world soil maps have been based on the different editions of Soil Taxonomy (Soil Survey Staff, 1960, 2010), the legend of the FAO/UNESCO Soil Map of the World (FAO, 1974) and the World reference Base of Soil Resources (FAO/ISRIC/ISSS, 1998; IUSS Working Group, 2006).

The largest available harmonized data sets are also based on these systems. The ISRIC-WISE Harmonized Global Soil Profile Database (Batjes, 2008, 2009) is maintained by ISRIC and includes 10,253 profiles. Most data were collected in national surveys and correlated or converted to the FAO legend and the WRB and partially to Soil Taxonomy. Since the national data collection and the analytical standards are very diverse the manually harmonized data base includes some inconsistencies and subjectivity (Lang et al., 2013a). The United States Pedon Database maintained by the United States Department of Agriculture, Natural Resources Conservation Service (USDA-NRCS) and includes several thousands of soil profiles and growing continuously.

The observation, sample collection, characterization and laboratory measurements have been based on uniform USDA standards minimizing the subjectivity of Pedon data. The high number of profiles and the consistency of the database make it the best set for evaluation of soil classification units. The objective of the presented research was to utilize the available data to numerically evaluate the classification units of Soil Taxonomy. The work included the evaluation of concepts, simple statistical and pedometric methods applied on calculated

* Corresponding author.
E-mail address: langvince@gmail.com (V. Láng).

http://dx.doi.org/10.1016/j.geoderma.2015.09.008
0016-7061/

centroids for the great group and higher level taxa. The great group level and equivalent levels in other systems were agreed by the members of the Universal Soil Classification System IUSS Working Group as level where major soil differences should be evaluated (Hempel et al., 2013).

The idea of using calculated taxonomic distances to express the level of similarity and dissimilarity between different soil taxonomic units was first applied in the 1960's (Hole and Hironaka, 1960; Bidwell and Hole, 1964a, 1964b; Sarkar et al., 1966) but only with local data and limited scope. The rediscovery of taxonomic distance calculations has been initiated in the 20th century by Minasny and McBratney (2007) who incorporated taxonomic distances into spatial prediction and digital mapping of soil classes. Minasny et al. (2009) derived taxonomic distances for the WRB Reference Soil Groups (RSGs) based on the presence and absence of key properties.

Fuchs et al. (2011) studied the taxonomic relationship of Hungarian soil types based on their dominant soil forming processes to provide numerical support to the improvement of the criteria of the taxonomic units. Soil taxonomic distance calculations were also applied to study the correlation possibilities of different national soil classification systems to the WRB. The calculated distances supported the expert knowledge based correlations with objective measures of taxonomic relatedness of compared classification units of different systems. (Lang et al., 2013b; van Huyssteen et al., 2014).

2. Materials

2.1. Soil Taxonomy

Soil Taxonomy is a result of experience and contributions of many soil scientists around the world over many decades. The early works and the publication of the 7th Approximation (1960) were coordinated by Dr. Guy Smith. From the very beginning the aim of the efforts was to develop a system that can accommodate all soils of the world (Eswaran, 1999). Starting with the first edition, Soil Taxonomy was published in 1975 and was based on accumulated knowledge from development of other classification systems. Based on new experiences and continuous development, 12 updated editions of the Keys to Soil Taxonomy have been published by the National Cooperative Soil Survey. The keys provide the bifurcated key with specific rules for determining the proper taxonomic classification. This system requires identifying diagnostic horizons, properties and materials to determine the soil group. On the first level of the Soil Taxonomy 12 soil orders are defined by the key. Each soil order is further divided into suborders followed by further divisions for great groups, subgroups, families and continues to more than 24,000 individual series. In this research the definitions and criteria of the 11th edition of the Keys to Soil Taxonomy were used (Survey Staff, 2010).

2.2. Databases

Legacy data for the centroid calculations were derived from the United States Pedon Database maintained by the United States Department of Agriculture, Natural Resources Conservation Service (National Cooperative Soil Survey, 2012) and included 38,321 number of pedons for evaluation. The WISE 3.1 database maintained by ISRIC (Batjes, 2008, 2009). The ISRIC-WISE v3.1 dataset is a compilation of soil profile data, collected from 149 countries worldwide and includes 10,253 pedons, and 3569 have Soil Taxonomy related information, thus available for evaluation. Many pedons in each database were excluded due to the lack of complete data.

2.3. Consolidation and harmonization of databases

The described datasets were integrated into a new data structure to accommodate all above mentioned and future datasets in one database. The database was quality checked and filtered to provide sufficient data for the further calculations. Pedons with the following features were excluded from this study:

1. Pedons with non-continuous sampling to the bottom of the pedon or to a maximum depth of 1.5 m
2. Pedons with no records on the US Soil Taxonomy Great Group and higher taxonomic levels
3. Pedons with laboratory methods not following the standards discussed in Soil Survey Laboratory Information Manual, Report No. 45, Version 2.0 (Burt, 2011)
4. Great Groups with less than 5 pedons available for centroid calculations were excluded from taxonomic distance studies

This resulted a total number of 41,890 included in the further calculations. The geographical distribution of the profiles, with the distribution of profiles by orders is presented on Figs. 1 and 2, respectively. As most of the pedons are derived from the United States Pedon Database, and only pedons with Soil Taxonomy related classification were included a bias can be observed geographically. This bias will be addressed with the inclusion of other international and national databases and taxonomic units of other classification systems.

2.4. Calculation of centroids

The calculations were limited to soil chemical and physical properties which are widely available in most databases, or can be derived through pedotransfer functions and a limited number of soil morphological properties, due to low availability of many morphologic properties. The properties with the units of measure are presented in Table 1.

Profiles in the database were not sampled uniformly for laboratory analysis. Sampling was either according to genetic horizons or certain defined layers. To make comparison possible, each profile was converted into 5 cm layers to a depth of 50 cm, 10 cm layers between 50 and 110 cm and to 20 cm layers from 110 to a depth of 150 cm. Values were calculated on a weighted mean bases with the R software (R Development Core Team, 2012) aqp package.

Based on soil classification records, mean values were calculated for each Soil Taxonomy Great Group for each layer and each property. Higher level classification unit centroids were calculated from the lower level centroids to avoid bias by lower level units with high membership. Soil Taxonomy Great Group centroids with less than 5 contributing members to the centroid were excluded from the distance calculations, but were used as contributors to the higher taxonomic level centroids.

2.5. Calculation of distances

On the basis of the matrix, the taxonomic distances between Soil Taxonomy Orders and Great Groups were calculated via R software (R Development Core Team, 2012) HDMD package using Mahalanobis distance metrics. Mahalanobis distance was chosen over Euclidean distance as it takes the covariance into account. Different soil properties often correlate, thus the distance calculation method needs to be selected accordingly. Mahalanobis distance calculations requires both larger datasets and higher computational times over Euclidean distance calculations. The equation below was used for the taxonomic distance calculations:

$$dij = ((xi - xj)t\, S - 1(xi - xj))1/2 \quad (1)$$

where: dij is the element of distance matrix D with size (c × c), where c is the number of groups (in this case the number of soil classification units taken into consideration), S represents the covariance matrix. The value of dij represents the taxonomic distance between soil group i and group j, and x refers to a vector of indicators of the soil identifiers.