Geoderma 329 (2018) 54–60

Contents lists available at ScienceDirect

Geoderma

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

Allocating soil profile descriptions to a novel comprehensive soil classification system

Farzin Shahbazi[a,*,1], Jingyi Huang[b], Alex B. McBratney[b], Philip Hughes[b]

[a] Soil Science Department, Faculty of Agriculture, University of Tabriz, Iran
[b] Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia

ARTICLE INFO

Handling editor: M. Vepraskas

Keywords:
Soil taxonomy
Principal component analysis
Diagnostic horizons
Diagnostic properties
Allocation
Nearest-neighbour distance

ABSTRACT

Previous work has been put into the creation of a comprehensive soil classification system (CSCS) using a harmonised dataset of 23 soil properties at 18 depth intervals. The classification consists of selected soil taxa from the US Soil Taxonomy, World Reference Base for Soil Resources, the Australian Soil Classification, and the New Zealand Soil Classification. In this paper, the CSCS was tested for allocation using data for from 44 soil profiles collected in Iran. A distance-based algorithm was used to allocate and name the soil profiles according to the CSCS. It was found that 36 Iranian soil profiles are close to the existing taxa of the CSCS in the taxonomic space. Three Iranian profiles with distances between 25 and 30 taxonomic units to the closest CSCS taxa were added to the CSCS and assigned with new systematic names. Allocating the remaining 5 Iranian taxa would require regenerating the nomenclature system. The CSCS has shown advantages for allocating soil profiles from other regions of the world other than the USA, Australia and New Zealand. It also enables cross-referencing with existing soil classification systems. In the future, the CSCS can be further improved by adding taxa from other global or regional soil classification systems.

1. Introduction

The last few decades have seen an escalation in the importance and scope of soil science. The cataloguing and taxonomical identification of soil has become an increasing priority for a large number of nations. This is in part due to the recognition that soil security is as pertinent a priority as issues such as water security (McBratney et al., 2014). Unfortunately, communication between these nations has not been managed the same way the science has, creating a series of soil classification systems that are sometime niche, sometimes with similar terms that mean different things (for example, the Salic horizon of the USDA Soil Taxonomy and World Reference Base of Soil Resources), sometimes duplicated. Identifying truly unique soil types in a different system is therefore a challenge, as even though there have been great strides towards globalizing the study of soil, non-standard descriptions and methodologies (e.g. duplex soils in the Australian Soil Classification) still percolate through smaller and regional soil classification systems. This can be exacerbated by geopolitical issues which obfuscate communication and ergo understanding between government agencies of differing nations.

1.1. Novel classification development

Methods have been recently developed to look at these systems from a numerical perspective. Based on theories developed from the 'fifties to the turn of the century (Hole and Hironaka, 1960; Bidwell and Hole, 1964) viewing soil profile descriptions (SPD's), primarily from the Australian Soil Classification or the US based Soil Taxonomy (ST), in terms of properties and depths (Hughes et al., 2017a). Spline functions have been used to transform these soil descriptions which are typically recorded by horizon with associated horizon depths into standardized depth increments (Bishop et al., 1999). Of all the properties collected and splined in this fashion, 23 specific soil properties have been selected because of pedological importance and availability. These properties stacked according to depth into a single vector 414 integers long (excluding the pedon or taxonomic id) have been converted via scaled principal component analysis into a specific taxonomic space. The formula established from this process can be used to project other similarly collected and processed data into the same environment for taxonomic comparison. With SPD's or taxonomic descriptions converted in this way, other numeric techniques can be used for comparison (Hughes et al., 2018a). Of these techniques, convex hull analysis and

* Corresponding author at: Soil Science Department, Faculty of Agriculture, University of Tabriz, 5166616471, Iran.
E-mail addresses: shahbazi@tabrizu.ac.ir, farzin.shahbazi@sydney.edu.au (F. Shahbazi).
[1] Alternative address: Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia.

https://doi.org/10.1016/j.geoderma.2018.05.017
Received 27 February 2018; Received in revised form 3 May 2018; Accepted 13 May 2018
0016-7061/

numeric distances can provide insight into how well a soil description from one taxonomic system complements a profile or description from another. Taxonomic distances can also be used to determine if there is any duplication between systems in the data and remove unnecessary reference taxa. Furthermore, other mathematical techniques such as dendrogram analysis (Sokal and Rohlf, 1962) can demonstrate how well all taxa from all systems processed in the aforementioned manner, slot together within a global collection of soil profile or taxonomic descriptions (Hughes et al., 2017b). With all the given soils standardized and organized, a comprehensive logical nomenclature is possible, using the highest of the calculated dendrogram as a guide to the allocation of groups then applying a constant and a vowel to create a word, not unlike the other nomenclature systems such as organic chemistry's IUPAC (Hughes et al., 2018b). The culmination of all these methodologies is the production of a database comprised of soil profile descriptions which are sourced from a variety of locales and taxonomic systems, which are standardized in properties and depth increments, which also have their own systematic name and these in turn can be used as a reference for unknown SPD's or taxa from differing systems, a Comprehensive Soil Classification System (CSCS).

1.2. Introducing data from Iran

In order to truly test the efficiency and global nature of the CSCS, data from as wide a range of countries is required. Iran, being several thousand kilometres distant from the United States, New Zealand or Australia is quite possibly also taxonomically distant in its soil composition. It is instructive to be able to use an Iranian database to determine similarity and differences between soils as well as identify common management issues in an environment where effective dissemination of information is rare. The Iranian province of East Azerbaijan is an area in which agriculture can be roughly segregated into two major activities- tubers and cereal crops. The landscape consists of several mountain ranges which provide a host of conditions for differing soil types. As with many agricultural areas in the Middle East, farming has been practised for thousands of years (Shahbazi and De la Rosa, 2010). Several locations within the East Azerbaijan region have been sampled; a total of 44 profiles have been sampled and processed. Because of the limited soil profile data used to construct the CSCS, the incorporation of this data into a world system gives a rare opportunity to add region-specific soil profiles to the global soil database.

1.3. Aims

With a novel method for taxonomic comparison and the addition of some ‘exotic’ data from Iran, comes the opportunity of testing the new system with untried information. The purpose of the CSCS is to have a system that is simple, complementary, understandable, transferrable and reproducible. Many of these ideas can be tested by some simpler objectives. These ideally should be centred on the ability of the new system to allocate soil profiles automatically, to be able to identify ant extreme divergence within the new group and to seamlessly add any suitable profiles to the existing system.

The aims of this study therefore are:

- To allocate and name 44 soil profiles (considered as unknown profiles) collected in the north west of Iran.
- To identify any profiles that diverge substantially from known systems within the CSCS.
- To expand the CSCS by adding any Iranian profiles that are considered to be sufficiently divergent.

2. Materials and methods

2.1. Study area

The soil profiles used in this study were collected in the Ahar County, East Azerbaijan Province, Iran (Fig. 1). Briefly, the site was located between longitude 47°00′–47°07′43″ east and latitude 38°24′04″–38°28′33″ north. Based on the synoptic meteorological station report, the area has a warm summer continental climate with mean annual temperature of 10.8 °C, and mean annual precipitation of 299.4 mm. The parent materials were mainly limestone, old alluvium, and volcano-sedimentary rocks. A detailed soil survey of the area was reported by Shahbazi et al. (2009). According to the USDA ST (2006), the dominant soils are classified as Inceptisols, Entisols, and Alfisols (refer to Table 1). Typic Calcixerepts is the major subgroup, followed by Typic Xerorthents, and Calcic Haploxerepts. The main crops are wheat, alfalfa, sugar beet, potato and maize.

2.2. Comprehensive soil classification system (CSCS)

Creation of a CSCS can be summarized with a flowchart (Fig. 2). In brief, the CSCS was created by sequentially adding soil taxa of World Reference Base (WRB) for Soil Resources, Australian Soil Classification (ASC), and New Zealand Soil Classification (NZSC) that were at the equivalent classification level of the great groups of the ST to the ST Great Group allocations, and applying a thinning algorithm to remove the redundancy of the system (Hughes et al., 2018a).

It was noted that all the soil taxa in the CSCS included 23 soil properties measured at 18 depth intervals from the soil surface to 1.5 m (i.e. 414 soil variables in total). The 23 soil properties included coarse fractions, clay, silt and sand contents, soil colour expressed in red, green and blue, water content, ice evidence, bulk density, organic carbon content, carbonate content, pH, cation exchange capacity, exchangeable cations (i.e. calcium, sodium, magnesium, potassium), acid saturation, base saturation, exchangeable sodium percentage, electrical conductivity, and gypsum content. Scaling of the 414 variables was carried out:

$$\mathbf{A}_{i,j} = \frac{\mathbf{A}_{i,j} - \mathrm{mean}(\mathbf{A}_{i,j})}{\mathrm{SD}(\mathbf{A}_{i,j})} \quad (1)$$

where $\mathbf{A}_{i,\,j}$ was the j[th] soil variable for the i[th] taxon, mean($\mathbf{A}_{i,\,j}$) and SD ($\mathbf{A}_{i,\,j}$) were the average and standard deviation of the j[th] soil variable calculated using an initial soil database comprising of the centroids of the ST and ASC systems, and $\mathbf{A}_{i,\,j}$ was the rescaled j[th] soil variable for the i[th] taxon.

PCA was then applied to the scaled initial soil database to calculate a principal component (PC) space. All the calculations were conducted in R Software (R Core Team, 2017). The “*prcomp*” function of the “*stats*” Package was used to calculate the principal component space of the initial soil database and the “*predict*” function was used to project the different soil taxa onto the PC space.

The CSCS has 493 taxa from the ST (171 taxa), WRB (4), ASC (240), and NZSC (78). Hughes et al. (2018a) created a hierarchical nomenclature, but not allocation, system using the 493 taxa. The systematic names of the 493 taxa were shown in Fig. 3. The most important soil properties that have a main impact on the classification were acid saturation, base saturation, organic carbon content, and pH and were discussed in detailed in Hughes et al. (2017a).

The CSCS was then classified into three tiers, including 15 nomenclature groups at the first tier, 86 nomenclature subgroups at the second tier, and 493 nomenclature sub-subgroups at the third tier, respectively. The third tier is the operational one for allocation. A sequence of a sequence of 18 consonants (“B”, “D”, “F”, “G”, “H”, “J”, “K”, “L”, “M”, “N”, “P”, “R”, “S”, “T”, “V”, “W”, “Y”, and “Z”) was used to name the taxa at each tier alphabetically with “A” and “E” inserted between the