Geoderma 287 (2017) 66–70

Contents lists available at ScienceDirect

Geoderma

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

Estimating soil bulk density with information metrics of soil texture

Miguel Ángel Martín [a,*], Miguel Reyes [b], F. Javier Taguas [a]

[a] Dpt. of Applied Mathematics, E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, Avda. Complutense s/n, 28040, Madrid, Spain
[b] Dpt. of Applied Mathematics, E.T.S.I. Informáticos, Universidad Politécnica de Madrid, Campus de Montegancedo s/n, 28660 Boadilla del Monte, Madrid, Spain

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form 29 August 2016
Accepted 4 September 2016
Available online 15 September 2016

Keywords:
Bulk density
Soil texture heterogeneity
Soil particle size distribution
Information entropy

ABSTRACT

Using the Shannon Information Entropy (IE) as a soil structure metric is proposed to analyze the effect of the particle size distribution (PSD) heterogeneity on soil bulk density values. A data base including 6239 soil samples from Florida is used. For each soil the IE is computed using mass proportions of the seven texture fractions that the data base provides. The range of IE values is divided into subintervals of equal length to study how differences in the soil texture metric are reflected in differences in soil bulk density values. The total range of IE values divided into equal subintervals, each subinterval corresponds to the group of soils with IE in this subinterval, average bulk density for soils in each subinterval is found, and the average information entropy value in any of the subintervals is plotted versus the average soil bulk density values. Coefficients of determination of the linear regressions 'average IE vs. average bulk density' were 0.99 and 0.98 for 10 and 15 subintervals respectively. Predictions based in that linear relationship give the mean predicted error (MPE) equal to 0.0015 g/cm^3 over the total number of soils, and the normal distribution of prediction errors with the standard deviation of to 0.16 g/cm^3. These results strongly support the hypothesis that Information Entropy can serve as an indicator of the typical bulk density for a soil with a given PSD. Values of IE were also computed for all samples in the database using only three texture fractions: clay, silt and sand content. Simple linear regression using the IE value as the input variable was implemented to predict bulk density value. Additionally, several published bulk density pedotransfer functions (PTFs), including organic carbon (OC) content and texture inputs, were applied to the same data base. Results show the root-mean-squared error of predictions close to 0.16 g/cm^3 when the IE is used as the sole input. Estimation of bulk density using Information Entropy as predictor became worse for soils in horizon A than in the horizon E, respectively, possibly due to the influence of the different organic carbon content in those horizons.
Overall, the Information Entropy metric of soil texture provides a useful input for estimating bulk density, which also might be used together with other inputs as depth or OC content.

1. Introduction

Bulk density (ρ_b) is an important soil physical property needed for estimating soil water characteristic and used as the input parameter for water and nutrient transport models (Boucneau et al., 1998). Accurate evaluation of bulk density is needed to get precise estimates of soil organic carbon (OC) stocks by means of weight-to-volume conversions (Howard et al., 1995).

Different factors as depth, organic matter content or compaction have influence on the bulk density values. Overall, differences in bulk density values among soils are primarily attributed to differences in particle size distribution (PSD) (Manrique and Jones, 1991). Large variations are observed in how PSD data are included in pedotransfer functions (PTFs) used to estimate bulk density values (Rawls, 1983; Manrique and Jones, 1991; Tomasella and Hodnett, 1998; Bernoux et al., 1998; Kaur et al., 2002; De Vos et al., 2005; Benites et al., 2007; Nanko et al., 2014). There are great differences in weights given to textural components in multiple regressions derived for different data sets. In spite of large number of existing bulk density PTFs, the search for additional properties as inputs is an important direction for improving PTF accuracy and reliability (Wösten et al., 2001). As it has been recently noticed, developing physics-based PTFs should be explored (Pachepsky and Romano, 2015).

The quantitative analysis of the influence of the PSD on bulk density values appapently should benefit from introduction of a metric which can effectively summarize the effect of variation in the PSD on variations in bulk density. As suggested by Tranter et al. (2007), bulk density can conceptually be represented as the sum

$$\rho_b = \rho_m + \Delta\rho \qquad (1)$$

where the component ρ_m is seen as the typical bulk density of mineral matter for a soil with a given PSD and average structure features and

* Corresponding author.
E-mail address: miguelangel.martin@upm.es (M.Á. Martín).

http://dx.doi.org/10.1016/j.geoderma.2016.09.008
0016-7061/

$\Delta\rho$ the variation in the bulk density attributed to other structural factors such as organic matter.

The objectives of this study are: (i) to propose the metric of soil texture based on the information theory, (ii) to study how differences in the soil texture metric correspond to differences in the typical soil bulk density values of soils and to analyze the basic structure of the correspondence, (iii) to evaluate the accuracy of predictions of soil bulk density using the information metric as a single input, and (iv) to compare predictions made by using the information metric as single input with those made using traditional PTFs which include OC content along with textural inputs as predictors.

2. PSD heterogeneity metrics

The soil texture information metrics proposed here -the Shannon Information Entropy-, is related to PSD heterogeneity. Indeed heterogeneity should influence bulk density values. Briefly, in a sandy soil solely composed by sand particles only pores greater than certain size might be expected to be present. On the contrary, in the case of a more heterogeneous texture with additional clay fraction, some of the parts of such pore space should be occupied by smaller particles, with no significant increase in volume. Similar arguments might be held true across the whole range of size particles: the higher the degree of heterogeneity is the higher chance of denser packing would be expected. One metric of heterogeneity is the Shannon Entropy (Shannon, 1948). A related entropy-like parameter, the Balance Entropy Index (BEI), has already been used as a predictor of soil water retention estimation (Martín et al., 2005). Results of this work suggested that the index had a potential to reflect probable packing of soil particles in soils with heterogeneous PSD's. This is consistent with the fact that the BEI provides additional information at smaller scales when a scaling in the PSD takes place (Martín and Taguas, 1998; Martín et al., 2001).

Heterogeneity is an ambiguous term that needs to be quantified to be introduced as input parameter for bulk density prediction. This issue is addressed with concepts from the information theory introduced in the pioneering work of Shannon (1948).

Formally speaking, the PSD of granular media may be considered as a continuous mass particle-size distribution $\boldsymbol{\mu}$ supported on the interval of grain sizes. Limited information on PSD is usually determined over a set of size ranges that covers all the sizes present in the sample. Grains sorted according to the size thus appear distributed over size classes $\boldsymbol{J}_1, \boldsymbol{J}_2, \ldots, \boldsymbol{J_k}$ defined by those size ranges. Particle size analysis provides the mass proportions $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p_k}$, where $\boldsymbol{\mu}(\boldsymbol{J_i}) = \boldsymbol{p_i}$, $1 \le \boldsymbol{i} \le \boldsymbol{k}$, and $\sum_{\boldsymbol{i}=1}^{\boldsymbol{k}} \boldsymbol{p_i} = 1$. The Shannon Information Entropy (IE) (Shannon, 1948) of the PSD is defined by

$$H = -\sum_{i=1}^{k} p_i \log\ p_i \tag{2}$$

where log denotes the binary logarithm. Products $p_i \log p_i$ are set to zero if $p_i = 0$.

The number $H = H(p_1, p_2, \ldots, p_k)$ is expressed in information units (bits) and its extreme values are $\log 2^{2k}$, which corresponds to the most even (homogeneous) case — where all the intervals have the same cumulative mass — and 0, which corresponds to the most uneven (heterogeneous) case — where the whole mass is concentrated in a single interval. The Shannon Information Entropy $H = H(p_1, p_2, \ldots, p_k)$ is a widely accepted measure of heterogeneity in the mass distribution μ among partitions $P = (J_1, J_1, \ldots, J_1)$. It can be shown that any measure of heterogeneity with natural properties for such goal must be a multiple of $H(p_1, p_2, \ldots, p_k)$ (Khinchin, 1957). Entropy has long been used in the life sciences as a plausible metric of biodiversity (Margalef, 1958) in the sense of evenness or heterogeneity of the diversity of species in an ecosystem. A similar use of entropy to measure pedodiversity has been discussed in (Ibáñez et al., 1998).

3. Materials and methods

3.1. Data set

The data set used to evaluate the influence of soil texture heterogeneity on soil bulk density values is USKSAT. This data base is comprised from journal publications and technical reports containing coupled data on laboratory Ksat, USDA textural class, and bulk density obtained in the United States. Detailed information can be found in Pachepsky and Park (2015). Percentages of seven textural fractions (very coarse sand, coarse sand, medium sand, fine sand, very fine sand, silt and clay) are available. Soil samples not reporting both, texture and bulk density, and soil samples with inconsistent textural data (the sum of mass texture fractions not agreeing with the total mass) were excluded. Soil samples with the same textural data but bulk density values differing less than 0.1 g/cm^3 were combined and considered as a single soil. Finally, soil samples having same textural data and collected at the same depth, were grouped and the average value of their bulk density values was used. Based in these selection criteria, total of 6239 soils were considered for the study. Of those, 3956 samples are sand, 570 are loamy sand, 698 are sandy loam, 27 are loam, 27 are silty loam, 4 are silt, 666 are sandy clay loam, 26 are clay loam, 3 are silt clay loam, 118 are sandy clay and 144 are clay. In the comparisons with different PTFs, four groups of soils have been considered: (1) the total data base (6239 samples), (2) soils from horizon A (1088 samples), (3) soils from horizon E (1113 samples) and (4) sandy soils (1264 samples). Selected descriptive statistics of the basic soil properties are in Table 1.

3.2. Methods

Firstly, Information Entropy of soil texture was computed, following Eq. (2), for all samples (6239) using the seven available fraction contents. In order to analyze how differences in the Information Entropy explain differences in the typical soil bulk density value of related soils, the entire range of information entropy values was first divided into 10 subintervals of equal length. Variations in the corresponding mean bulk density values in the subintervals were less than 0.04 g/cm^3. The mean bulk density value of soil samples with information entropy within each subinterval was computed and the mean bulk density value was plotted against the mean information entropy value. Linear regression 'average IE vs. average bulk density' was computed. The same analysis was repeated dividing the range of information entropy values into 15 subintervals of equal length. Variation of mean bulk density within subintervals were less than 0.02 g/cm^3.

The linear regression was used to predict bulk density values of all the soils samples (6239) with the Information Entropy value as the sole input. The prediction quality was determined by computing complementary indices: Eq. (3), the mean predicted error (MPE); Eq. (4),

Table 1
Summary statistics of basic soil properties of USKSAT data base.

Dataset	Statistic	OC	Sand (%)	Silt (%)	Clay (%)
All soils	Min	0.01	0.20	0	0
(6239)	Max	15.61	99.90	94.50	93.40
	Mean	0.59	85.53	5.65	8.83
Horizon A	Min	0.08	2.40	0	0
(1088 soils)	Max	11.55	99.90	53.10	67.70
	Mean	1.60	89.92	6.10	3.98
Horizon E	Min	0.01	36.40	0.10	0.10
(1113 soils)	Max	3.61	99.70	42.60	21.00
	Mean	0.22	95.54	2.97	1.48