Available online at www.sciencedirect.com
SCIENCE DIRECT®
Ecological Modelling 182 (2005) 221–228
ECOLOGICAL MODELLING
www.elsevier.com/locate/ecolmodel

An entropy-based heterogeneity index for mass–size distributions in Earth science

Miguel Angel Martín[a,*], José-Manuel Rey[b], Fco. Javier Taguas[a]

[a] *Departamento de Matemática Aplicada, Escuela Técnica Superior de Ingenieros Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain*
[b] *Departamento de Análisis Económico, Universidad Complutense, Campus de Somosaguas, 28223 Madrid, Spain*

Abstract

A quantitative classification of soil texture is proposed based on an entropic index that can be easily computed from knowledge of the fractional contents of soil textural classes. It is first shown that the index formula supplies a number that agrees with the entropy dimension when the corresponding soil particle-size distribution (PSD) displays self-similar fractal features. In the absence of self-similarity, the index is further shown to retain information-theoretic content so that it becomes a meaningful diversity index in the general case. The index is defined by balancing Shannon's entropy in an appropriate way to deal with the high variability of the interval lengths used to report soil particle size classes. The performance of the proposed formula is illustrated for standard textural data reported as clay–silt–sand soil mass fractions. The index induces a classification of a continuum of textural classes that may distinguish soils within the same standard textural class, thus establishing a continuous characterisation of textures that is complementary to the usual classification, but requires no additional information. Finally, it is shown how the balanced-entropy index might also be used as a measure of body size diversity for living organisms.

Keywords: Soil texture; Biodiversity; Shannon's entropy; Information theory; Fractal geometry

1. Introduction

The classification of soil and sediment textures plays an important role in the Earth sciences. In particular, the statistical description of soil particle-size distributions (PSD) is of great importance in the study of soil physical properties.

The usual classification of textures defines textural classes grouping together soils with mass percentages of clay, silt and sand between certain prescribed limits. Different classifications of soil textures have been proposed (Folk, 1954; Shepard, 1954; Baver et al., 1972; Soil Conservation Service (SCS), 1975; Vanoni, 1984). These systems differ both in the particle-size limits chosen to separate the size groups and in the percentage limits established to define each textural class.

Since many different combinations of clay, silt and sand may correspond to the same textural class, soil samples of rather diverse composition appear indistinguishable under the grouping that these classes establish. Shirazi and Boersma (1984) proposed a classification based on the addition of new information to the conventional texture triangle used by the United States Department of Agriculture (USDA). By integrating on the textural triangle geometric means and standard deviations obtained from mechanical analysis of soil samples, they derive a new diagram which provides greater resolution in detecting

* Corresponding author. Tel.: +34-91-336-58-13; fax: +34-91-336-58-17.
E-mail address: mamartin@mat.etsia.upm.es (M.A. Martín).

0304-3800/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2004.04.002

classified soil samples within a textural region. However, because of its character of first approximation to PSD, a classification system of textures should keep a trade-off between simplicity and taxonomic power. A more efficient statistical description may thus be inadequate for classification purposes if it is achieved at the cost of obtaining extra non-trivial information.

The issue is whether the characterisation of soil textures can be refined and unified without requiring any further information than that employed by any of the standard classification systems, e.g. by the USDA or the International Society of Soil Sciences (ISSS), that is, data of soil mass percentages of primary particles only.

The goal of this note is to propose a uniparametric continuous characterisation of textures by means of an index built from Shannon's entropy (Shannon, 1948a,b) that can be computed from soil mass fractions of primary particles. The use of the index as a textural parameter arose from a fractal modelling for PSD (Martín and Taguas, 1998). Under the fractal model, the index is the so-called *entropy dimension* of the underlying fractal distribution which in turn yields rich information on the scaling behaviour of mass distribution with respect to particle sizes. However, if the fractal model is not assumed to describe the distribution of soil particle masses, the index can still be understood from information theory and can be shown to carry information about the heterogeneity of a PSD.

The ideas above also appear to be useful in ecology, namely for evaluating diversity of body size distribution in living organisms, which is a problem remarkably having common features with that of evaluating PSD textural heterogeneity. However, further biodata work is needed to illustrate the use of the index in this context.

The use of the index for practical classification of textures stems from (Martín and Taguas, 1998; Martin et al., 2001) and its role for measuring body size diversity was addressed in (Martín and Rey, 2002). The general ideas behind the theoretical framework are discussed in Sections 2 and 3. In Section 4, a practical study using clay, silt and sand percentages corresponding to 171 real soil data from Soil Conservation Service (SCS, 1975) is performed to show the ability of the proposed parameter to characterise soil textures. Section 5 comments on the possible use of the balanced-entropy index to evaluate body size diversity.

2. Models and parameters for texture that require no extra information

The challenge is to find relevant parameters, maybe through suitable models, to characterise PSD without requiring any more information than that supplied by usual textural data. Assume for the sequel that textural data are supplied by the fractions (P_1, P_2, P_3) of the mass of soil particles with characteristic sizes respectively within the intervals I_1, I_2, I_3, which are prescribed to report textures. The basic choice $I_1 = [0, 0.002]$ (mm), $I_2 = [0.002, 0.05]$ (mm) and $I_3 = [0.05, 2]$ (mm) used by the USDA classification will be considered in this paper. It may be noted that any other choice, varying either the number or the size of the intervals, may be considered within the scheme of the model and the accompanying parameter described below.

Under the point of view of the statistical description of PSD, infinitely many different distribution models may be conceived to fit given textural data (P_1, P_2, P_3), even under strong assumptions like, for instance, log-normality. Of course, each one of them would predict differently the distribution of mass inside the intervals I_1, I_2, I_3, when nothing is known from the given data. The selection of a best model to describe the real distribution would require extra data on particle sizes at a finer resolution than those reported by I_1, I_2, I_3.

However, for classification purposes, the problem has further subtle shades, since the little amount of textural information, as reported above, is all that one has to design a parameter that differentiates textures. Such a parameter should ideally capture some meaningful feature of the PSD, rather than describing the entire distribution.

The approach of the fractal model below links in fact both aspects. First, a distribution model, unequivocally determined from usual textural data without extra information, is constructed to fit the data and to replicate unknown data at smaller size scales. Second, the model provides a easily computable parameter which proves relevant for the characterisation of soil PSD's.

2.1. A fractal model for PSD

PSD may be thought of as a mass distribution in the interval $I = [0, 2]$ (mm) of particle sizes assigning to each interval of sizes $[a, b]$ the mass of soil