Geoderma 219–220 (2014) 106–116

Contents lists available at ScienceDirect

Geoderma

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

A new approach to soil classification mapping based on the spatial distribution of soil properties

Alessandra Fagioli da Silva [a,*], Maria João Pereira [b], João Daniel Carneiro [b], Célia Regina Lopes Zimback [a], Paulo Milton Barbosa Landim [c], Amilcar Soares [b]

[a] UNESP — Univ Estadual Paulista, Faculdade de Ciências Agronômicas, Departamento de Recursos Naturais, Rua José Barbosa de Barros, Caixa-Postal: 237, 1780 CEP: 18610-307 Botucatu, SP, Brazil
[b] CERENA, Instituto Superior Técnico, Universidade Técnica de Lisboa, IST, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
[c] UNESP — Univ Estadual Paulista, Instituto de Geociências e Ciências Exatas, Departamento de Geologia Aplicada, Av. 24-A, 1515 Bairro Bela Vista, CEP 13506-900 Rio Claro, SP, Brazil

ARTICLE INFO

Article history:
Received 27 February 2013
Received in revised form 18 November 2013
Accepted 10 December 2013
Available online 25 January 2014

Keywords:
Diagnostic attributes
Geostatistical simulation
Multidimensional scaling
Soil mapping

ABSTRACT

This study presents a new approach to classifying types of soil based on the probability classes of the relevant set of attributes. Two key ideas are addressed in this study: (i) the use of stochastic simulations to generate a local cumulative distribution function or extreme classes of each attribute and (ii) the use of a multidimensional scaling (MDS) technique to visualize and quantify the relative importance of each attribute in the classification process. After the simulated realizations, the weighted "distances" attributes extreme values (probability classes) of each grid node are calculated and the MDS algorithm is applied for the spatial representation of the grid nodes in a new Cartesian reference frame based on the "distances" of the probability classes of attributes. This allows the classification of soil types based on the clusters in the MDS space, after expert validation. In the second step, a sensitivity analysis of the attributes is performed with MDS: each attribute is made "neutral" one at a time, by assuming the median rather than the extreme values in each grid node before the distance evaluation, and the consequent impact on the shape and centroid displacement of the clusters (soil types) in the MDS reference frame is calculated. Hence, the spatial uncertainty of the soil type/classes and the influence of various properties are evaluated in the MDS reference frame. This method is applied to soils in a region of Brazilian in which the previous classification of soil types has been a crucial tool for precision agriculture management. Using the MDS algorithm, the selected attributes (horizon, textural gradient, colors, saturation, sand content, and clay content) were represented in a two-dimensional plot and grouped into eight clusters distinguished from each other by their characteristics. A sensitivity analysis shows that the horizon and saturation attributes had the greatest influence on determination of the clusters, i.e., the soil types.

1. Introduction

Soil mapping is a crucial physical environment tool for rational land planning and environmental management that balances economic and social development with the conservation and protection of natural resources. However, there are several limitations on the ability to collect data about land and/or its attributes, such as the high cost of surveys, the large areas to be mapped, and the difficulty in accessing those areas. Usually, one must consider limitations related to problems with the accuracy and reliability of information obtained primarily from qualitative interpretations.

The patterns of soil in the landscape are the result of the effects of soil-forming factors, namely, climate, organisms, parent material, topography, and time. These factors, particularly the climate and parent material, produce broad patterns of soil in geographic space. The complex combination of these factors causes repetitive patterns of soil in the landscape, which form the basis for soil classification, identification, and mapping. In traditional soil classification, both the conceptual and real classes are qualitative and have no clear-cut boundaries. Pedologists have, in the past, attempted to avoid qualitative classification by applying numerical classification to complex soil data. Soil classification, in both geographic and taxonomic space, is a simple representation of complex and occasionally repetitive patterns of soil in the landscape. Class identification, in contrast, seeks to match classes that are conceptually described (usually in a classification system) with reality, represented by data obtained by soil morphological descriptions and measurements (Odeh and McBratney, 2005).

The main purpose of any classification is data reduction, whereby a complex system represented by some set of data is made more explicit. Almost all soil surveys are accompanied by some forms of grouping, such as the so-called "natural" system of classification or the technically interpretative form. These classifications are composed of mutually

* Corresponding author at: Rua Wantuil Goulart Barbosa, 361, CEP 86360-000 Bandeirantes, PR, Brazil, Tel.: +55 43 99673558.
E-mail addresses: alefagioli@hotmail.com (A.F. da Silva), maria.pereira@ist.utl.pt (M.J. Pereira), joao.carneiro@ist.utl.pt (J.D. Carneiro), czimback@gmail.com (C.R.L. Zimback), plandim@rc.unesp.br (P.M.B. Landim), asoares@ist.utl.pt (A. Soares).

0016-7061/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.geoderma.2013.12.011

exclusive classes to conform to the discontinuous soil variation embedded in the traditional soil surveys. However, soil variation is more continuous than discrete. This problem has been the motivation for paradigm shifts in soil classification; hence the applications of pedometrics. Pedometrics is the use of quantitative methods for the study of soil distribution and genesis under uncertainty in soil models that describe deterministic or stochastic variation, vagueness, and lack of knowledge of soil properties and processes. The pioneering efforts in pedometrics are the computer-based numerical classification research of Hole and Hironaka (1960), Moore and Russell (1967), and de Gruijter (1977).

Many quantitative methods of soil classification have been proposed as applications of numerically hard fuzzy classification systems and logistic regression and were designed specifically for situations with dichotomous (nominal or ordinal) dependent variables. Classical linear regression, which is typically used for continuous dependent variables, and classification trees search for combinations of values of independent variables that best predict the value of the dependent variable (Odeh and McBratney, 2005).

Due to nonlinearity in the correlations among many soil variables, and indeed in the correlations of many soil variables with ancillary variables, robust methods such as generalized linear models (GLMs), generalized additive models (GAMs), and regression trees (RTs) have been developed and applied (McBratney et al., 2000). Moral et al. (2010) used distribution maps of physical soil variables to generate potential management zones by means of principal component analysis and a fuzzy cluster algorithm. However, most of these methods share identical limitations: the classification is based on experimental sample values, i.e., there is no spatial inference in the classification.

Therefore, pedometricians have begun applying geostatistics (Odeh and McBratney, 2005), to quantify the spatial continuity patterns of classes identified at sample locations using variograms and then using kriging methods for the spatial inference of the classes. A different approach involves the classification being calculated after the spatial inference (kriging) of the attributes at each grid node of the study area. Both approaches rely on the smoothing effect of kriging of the attributes, which may have an impact on posterior classification, and both approaches lack an assessment of the spatial uncertainty of attributes or classes (Goovaerts, 1997). In this study, we sought to address these limitations using stochastic simulation methods, because it allows the assessment of spatial uncertainty (Goovaerts, 1997; Soares, 2001), to calculate extreme values (local probabilities) of attributes in each grid node of a study area and to evaluate the relative weight of each attribute in the classification using a multidimensional scaling (MDS) technique (Cox and Cox, 2001), because it reduces the dimensionality of the data and does not require that the data have correlation each other.

Thus, we propose a new hybrid model for mapping the taxonomic classification of soils that combines geostatistical simulation methods and multivariate analysis, namely, indicator sequential simulation (SIS) (Goovaerts, 1997), direct sequential simulation (DSS) (Soares, 2001), and MDS (Cox and Cox, 2001). The SIS and DSS methods provide estimates of the local probability distributions of a set of quantitative and qualitative soil attributes, taking into account the spatial autocorrelation of these attributes, for each grid node. Then, the weighted "distances" between grid nodes are calculated based on classes of properties, and an MDS technique is applied to highlight clusters of samples in the new MDS reference frame. After expert validation, the clusters are converted into soil types and mapped. Lastly, a sensitivity analysis of the influence of each attribute to the final clusters (soil types) is performed with the MDS: each attribute is excluded, one at a time, by assigning its median value to each grid node, and the resulting displacement of each cluster in the new MDS reference frame is calculated and assumed to be proportional to the relative importance of the attribute in the classification.

This new method was applied to a case study site in the countryside of the State of São Paulo, Brazil.

2. Study area

2.1. Area description

The study site, which has an area of approximately 1200 ha, is the Edgardia experimental farm owned by the Faculty of Agricultural Sciences of São Paulo State University, which is located in the municipality of Botucatu in San Paulo State in southeast Brazil, between −22° 47′ 30″S 48° 26′ 15″W and 22° 50′ S 48° 22′ 30″ W, at an altitude between 475 and 725 m.

The region has a mesothermal humid without a dry season. The climate is classified as Cfa according to the Köppen system, with an average temperature in the coldest month below 18 °C and an average temperature in the warmest month above 22 °C. The mean annual rainfall is 1530 mm. Remnants of natural vegetation, characterized by massive formations of isolated plants of seasonal forest, savanna, and grassland, can be observed.

2.2. Soil properties

The characterization of soil in Brazil is based on the methodology defined by Embrapa (2006). Soil types are defined in terms of categorical (horizon and color) and continuous attributes (texture gradient, degree of saturation, sand content, and clay content):

- Horizon: the diagnostic horizon identified in the subsurface. Soils with a B horizon are coded as 1, and soils without a B horizon are coded as 2. The first corresponds to deeper soil than the second, that is more developed.
- Color: the soil color, according to the Munsell chart (Munsell, 1998), in one of four classes – brown, gray, yellowish red, and red – coded as colors 1, 2, 3, and 4, respectively.
- Textural gradient: the ratio of the average clay content of the B horizon to that of the A horizon. When this ratio is greater than 1.7, it is assumed that textural horizon B is present in the soil, this means that the clay content of the B horizons (subsurface horizon) is larger than the horizon (surface horizon), or soils show clay migration processes underground.
- Degree of saturation (V): an attribute obtained by calculating the ratio of the base that is extractable (Eq. (1)). A base saturation greater than or equal to 50% indicates that the soil is eutrophic, and a base saturation less than 50% indicates that the soil is dystrophic.

$$V = \frac{S}{T} \times 100 \quad (1)$$

where ($S = Ca^{2+} + Mg^{2+} + K^{+} + Na^{+}$) and ($T = Ca^{2+} + Mg^{2+} + K^{+} + H^{+} + Al^{3+}$); *S*: base sum; and *T*: cation exchange capacity.

The sand and clay contents are determined from a granulometric (particle size) analysis performed according to the Empresa Brasileira de Pesquisa Agropecuária (1979) methodology. For the purpose of subdividing the soil according to texture (Table 1 and Fig. 2), the following three categories are used:

a) Clayey: containing more than 35% clay.
b) Medium texture: containing between 15 and 35% clay.
c) Sandy: containing less than 15% clay and more than 70% sand.

Soil data were collected by Carvalho et al. (1991) on a sampling grid (Fig. 1) of 95 samples that were collected per horizon. The purpose of this field survey was to build a semi-detailed soil map.

The general characteristics of the soil samples are summarized in Table 1.