Geoderma 326 (2018) 111–122

Contents lists available at ScienceDirect

Geoderma

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

The mapping of soil taxonomic units via fuzzy clustering – A case study from the Outer Carpathians, Czechia

Matěj Horáček[a,*], Pavel Samec[b,d], Jozef Minár[c]

[a] Department of Physical Geography and Geoecology, Faculty of Science, University of Ostrava, Chittussiho 10, 710 00 Ostrava, Czechia
[b] Department of Geoinformatics, Faculty of Science, Palacký University in Olomouc, 17. listopadu 50, 771 46 Olomouc, Czechia
[c] Department of Physical Geography and Geoecology, Faculty of Natural Sciences, Comenius University in Bratislava, Mlynská dolina, 842 15 Bratislava, Slovakia
[d] Department of Forest Ecology, Faculty of Forestry and Wood Technology, Mendel University, Zemědělská 3, Brno CZ-613 00, Czechia

ARTICLE INFO

Editor: A.B. McBratney

Keywords:
Digital soil mapping
Fuzzy c-means clustering
Soil taxonomy
Defuzzification

ABSTRACT

The paper presents a new method for the digital mapping of taxonomic soil units via fuzzy taxonomy and fuzzy clustering. In principle, this respects the continuous character of the soil but can result in traditional crisp soil maps. A part of the Moravskoslezské Beskydy Mts. (Czech Republic) was the testing area. Fuzzified taxonomic soil information from 106 soil pits with 75 geomorphometric parameters (potential environmental covariates of soil units) derived from a 10 m LIDAR Digital Elevation Model (DEM) was used for the input data. Another 60 soil pits were used for the validation of the results. Generalized linear models (GLM) of the relationship between fuzzified taxonomic soil information (membership of soil pits to particular soil types) and all geomorphometric parameters were used for the selection of 8 geomorphometric parameters as environmental covariates of soil taxonomic units (Catchment Slope, Elevation, Local Upslope Curvature, Gradient Difference, Melton Ruggedness Number, Slope Height, Modified Catchment Area, Multiresolution Index of Ridge Top Flatness). The fuzzy c-means clustering of selected covariates led to the delimitation of soil-landscape units. The taxonomic meanings of the soil-landscape units (membership of the particular soil type to the particular type of soil-landscape unit) were determined on the basis of membership of soil pits to the soil types, and membership of the pixels with soil pits to the given type of soil-landscape unit. Every pixel obtained an individual membership to each soil type in this way. The defuzzification process (the assignment of a given pixel to only one resultant soil subtype) considered the first and second largest membership of pixels to particular soil types. To express the scientific reliability of the results, methods for measuring uncertainty and two modified confusion indexes are calculated. This approach shows 26% full agreement, 64% partial agreement and 10% disagreement between the modelled and observed point data. The result significantly exceeds the accuracy of conventional soil maps in the tested area, as well as in some other previously investigated regions of the Western Carpathians.

1. Introduction

Soil taxonomy maps are the most common and oldest types of soil maps. Traditional soil taxonomy maps are ‚double crisp‘– crisp in terms of both taxonomy and geographical space (Burrough et al., 1997). Although soil has long been seen as a continuous natural system (McBratney and De Gruijter, 1992; Odeh et al., 1992), in the past specific methods for mapping soil as a continuum were missing. From the 1990s many methods were developed which use fuzzy logic and allow mapping of soils that respects their natural character, bringing more detailed information which was lost in the traditional approach.

Soil taxonomies are constructed as systems of conventional nominal units, although the relevant diagnostic characteristics generally vary continuously. Only some transitional soil units reflect this continuous character in principle, although these are still nominal units. By assigning a certain soil pit to a nominal unit, information is lost as to how representative of this unit the soil pit is, and whether the pit is related also to another soil unit. This problem can be solved using fuzzy expression of the membership of soil pits to nominal taxonomic units (e.g. Burrough, 1989; McBratney and De Gruijter, 1992; De Gruijter et al., 1997; Carré and Girard, 2002; Lagacherie, 2005; Balkovič et al., 2007; Balkovič et al., 2013).

As a soil pit can relate to several nominal units, a mapped pixel in geographical space can relate to several spatial (soil landscape) units. It is necessary to group pixels which have similar environmental condition and thus similar characters of pedogenesis. One useful method is

---

[*] Corresponding author.
E-mail address: matej.horacek@osu.cz (M. Horáček).

https://doi.org/10.1016/j.geoderma.2018.04.012
Received 12 October 2017; Received in revised form 9 April 2018; Accepted 10 April 2018

fuzzy c-means clustering of environmental covariates (e.g. Lee et al., 1988; Frazier and Cheng, 1989; De Bruin and Stein, 1998; Lark, 1999; Deng and Wilson, 2006; Zhu et al., 2010; Yang et al., 2011; Sun et al., 2012; Wang et al., 2012; Yang et al., 2016).

Generally used environmental covariates have various degrees of scientific reliability, spatial variability and coupling with soil properties. Geomorphometric parameters are the most frequently used type of environmental covariate. Digital soil mapping without the use of geomorphometric parameters is rare: McBratney et al. (2003) summarise 67 studies, 56 of which used geomorphometric parameters. There are 14 studies which model taxonomic units, only two of which did not use geomorphometric parameters. The increasing accuracy of digital elevation models (DEMs) and the quantity of relevant geomorphometric parameters provide an ideal set of potential environmental covariates, especially for mountainous and hilly areas with homogeneous land cover and lithology, given the limited availability of other relevant covariates.

There are several ways for making fuzzy maps of nominal units (Hengl et al., 2004), which contain more information than crisp maps. But these maps are frequently less readable for end users. So the best way is to model using fuzzy logic, followed by defuzzification to produce maps of nominal units that are more informative than those produced without fuzzy logic. Our aim here is to show how this can be achieved, and to assess the possible improvement over traditional mapping.

The proposed method was developed for the digital soil mapping of a forested monotonous flysch region of the Outer Carpathians, as an alternative to conventional soil mapping. Mapping in mountainous areas generally suffers from a much lower density of available point soil data, than in agricultural areas (because of worse accessibility and lesser economic importance). On the other hand the diversity of taxonomic soil units is higher because the range of environmental conditions is much bigger in mountainous areas. As a result, traditional soil taxonomy maps can be highly erroneous. Use of environmental covariates is therefore an ideal solution. To allow for natural fuzziness, we therefore used a combination of definition of types of soil-landscape units by fuzzy clustering of geomorphometric parameters, and determination of their taxonomic (fuzzy) meaning by using fuzzified taxonomic information from the soil pits. The naturally continuous character of the soil (in both taxonomic and geographical dimensions) is thus fully respected during modelling. The final defuzzification of the information obtained permits production of a crisp soil taxonomy map, which is required by most users, yet is of significantly higher quality than a traditional crisp map.

The international taxonomy system of the World Reference Base (WRB) is not intended to substitute for national taxonomy systems (IUSS Working Group WRB, 2015). Soil mappers have to choose whether they want to model in the WRB, for example, Balkovič et al. (2007) or Hengl et al. (2007), or in a national classification system, like Kempen et al. (2009) or Zhu et al. (2010). The goal outlined above is the reason for the choice of the national taxonomy system in this study. If a national soil database is used for modelling, the choice of a national classification system would generally be more suitable by avoiding the elimination of some information during the conversion to WRB.

2. Materials and methods

2.1. Study area

The highest part of the Outer Western Carpathians in the Moravian-Silesian Beskids of Czechia (Moravskoslezské Beskydy Mts., Radhošťský hřbet ridge, 104.5 km^2, centred at 49°29′27.345″N, 18°18′3.051″E, Fig. 1), was chosen as the study area. Flysch, composed of alternating sandstone and claystone layers is the dominant parent soil material. Its mineralogical composition is 20–30% quartz, about 10% feldspar, 33–45% mica and 15–25% clay minerals. The content of calcium is about 0.5% (Menčík et al., 1983). The area has been disrupted by numerous landslides (Břežný and Pánek, 2017).

The area is allocated to two groups of the Köppen climate classification: Dfb (Warm summer continental or hemiboreal climates) and Dfc (Subarctic or boreal climates): the latter is only for the top part of the ridge. The mean annual temperature is 4–5 °C, while annual precipitation is 1000–1200 mm (Tolasz et al., 2007).

Production forest covers 94.2% of the area and its composition is strongly affected by man. Potentially, over 60% of the area is covered by four units of potential geobiocoenosis in the sense of forest typology (Průša, 2001): *Abieto - Fagetum mesotrophicum* (fresh, nutrient-medium Fir – Beech: 19.12%), *Abieto – Fagetum fastigiosum – lapidosum mesotrophicum* (slope-stony, nutrient medium Fir – Beech: 16.61%), *Abieto – Fagetum eutrophicum* (nutrient-rich Fir – Beech: 15.45%) and *Piceeto – Fagetum mesotrophicum* (fresh, nutrient-medium Spruce – Beech: 10.07%). The rest of the territory is classified with another 48 units of potential geobiocoenosis (ÚHÚL, 2016).

2.2. Taxonomy system and existing soil map

In this study, the "Taxonomy Classification System of the Soils of the Czech Republic" (Taxonomický klasifikační systém půd České republiky - TKSP) (Němeček et al., 2011) is used. The highest hierarchical level of the TKSP is Reference Classes, which correspond to Reference Soil Groups in the WRB. A lower (basic) taxonomic level is Soil Type. It is unassignable to the WRB. The next lowest level is Soil Subtype, which in practice is the most used level of the TKSP. Soil subtypes correspond to principal qualifiers in the WRB (Krasilnikov et al., 2009). However, in the TKSP the specification of the soil subtype is put after the name of the soil type, as in biological binomial nomenclature - genus and species. The list of soil subtypes used in this paper and their correlation with the WRB units are in Table 1. Both principal qualifiers in the WRB and the soil subtype in TKSP often express a transition to another soil unit (soil group or soil type). E.g. the transition of fluvisol to cambisol (soil types) made up of the following soil subtypes: fluvisol modal (haplic fluvisol) – fluvisol cambic (does not exist in the WRB) – cambisol fluvic (fluvic cambisol) – cambisol modal (haplic cambisol). We can model such soil subtypes as a result of the combination of soil types. Naturally, transitions do not exist between all soil types. On the other hand, some soil subtypes can pass to more than one other soil type. For example, cambisol dystric, which in the TKSP has transitions to podzosol and cryptopodzosol. Furthermore, some subtypes are not transitions to other soil types, for example umbric subtypes, as there is no umbrisol in the TKSP. TKSP soil subtypes defined as transitions between soil types are listed in Table 2.

The most current and detailed (1: 50,000) soil map of the study area was made by the Nature Conservation Agency of the Czech Republic (Agentura ochrany přírody a krajiny České republiky - AOPK) (AOPK, 2007, Fig. 2). This shows that the dominant soil subtype is cambisol modal (KAm; haplic cambisol in the WRB), which occurs in 57% of the area. Cambisol rankric (KAs; skeletic cambisol) covers 20.1%, cambisol dystric (KAd; dystric cambisol) 15.1% and cryptopodzosol modal (KPm; entic podzosol) 3.8% of the area (AOPK, 2007).

2.3. Input data

The modelling was based on the data from 74 soil pits from the soil databases of the Forest Management Institute (Ústav pro hospodářskou úpravu lesa – ÚHÚL) (ÚHÚL, 2013) and the data from 32 soil pits from our own field survey. For validation, the data of another 60 soil pits from the soil databases of ÚHÚL and our own field survey were used (Fig. 3). For the collection of soil data the conventional field survey described by Samec et al. (2014) was used. The taxonomic information of soil pits was fuzzified by determining the membership of soil pits to particular soil types. The determination of fuzzy taxonomic information by a pedologist is principally the same as the assignment of a nominal