Geoderma 315 (2018) 36–48

Contents lists available at ScienceDirect

Geoderma

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

The extrapolation of soil great groups using multinomial logistic regression at regional scale in arid regions of Iran

Farideh Abbaszadeh Afshar[a], Shamsollah Ayoubi[b], Azam Jafari[c,*]

[a] Department of Soil Science, College of Agriculture, University of Jiroft, Kerman, Iran
[b] Department of Soil Science, College of Agriculture, Isfahan University of Technology, Isfahan 84156-83111, Iran
[c] Department of Soil Science, College of Agriculture, Shahid Bahonar University of Kerman, Kerman 76169-14111, Iran

A R T I C L E I N F O

Handling Editor: A.B. McBratney

Keywords:
Digital soil mapping
Extrapolation
Calibrated-in
Multinomial logistic regression

A B S T R A C T

Soil information is essential for sustainable management of ecosystems. In many parts of Iran, soil information is either not available or difficult to obtain. Therefore, when no detailed map or soil information is accessible in a region of interest, one way to obtain information is to extrapolate information from other parts having soil information. This study was conducted to determine whether - machine-learning extrapolation method extracted from the reference region, i.e. Zarand can estimate soil classes in the interest region, i.e. Bam and reduce the costs of soil mapping. To identify similarities between reference and interest regions, homology of soil forming factors was determined using Gower's similarity index. Then, the multinomial logistic regression was extracted from the reference region and applied into the interest region to estimate soil classes. Moreover, soil classes were predicted in the interest region using direct soil observations; finally, the accuracy of the soil maps obtained from both methods was assessed. Based on Gower index, the study regions, namely Bam and Zarand, were to some extent similar in term of soil forming factors. The results showed that although the soil map derived from the extrapolation process indicated appropriate spatial coverage of soil classes in the interest area, the resultant predictive map of calibrated-in process was slightly more accurate, i.e. higher κ, lower Brier scores. Acceptable levels of predictive accuracy (60%) were achieved using extrapolated model while costs simultaneously significantly lowered. This study put forward the view that the extrapolation method was quite useful in predicting soil classes within areas where soil mapping by calibrated-in method might be too costly or time consuming or where soil observations may not be sufficient. Nevertheless, this research encouraged us to use extrapolated method to fill the gaps in the present soil map of Iran and to apply it as the base map to increase and improve the efficiency of digital soil mapping.

1. Introduction

"Soil mapping is a crucial physical environment tool for rational land planning and environmental management" which relates soil/soil classes to topographic position in certain landforms, geological units, vegetation communities, and/or land uses (Cook et al., 1996; McBratney et al., 2003; Scull et al., 2003). The integration process between soil forming factors and pedology can be qualitative or computer-based. The qualitative integration process is defined as traditional soil survey (TSM) and the computer-based integration process is known as digital soil mapping (DSM) (Brungard et al., 2015; Grinand et al., 2008; Jafari et al., 2012; McBratney et al., 2003; Scull et al., 2003, 2005; Taghizadeh-Mehrjardi et al., 2014, 2015).

As traditional soil mapping methods are costly and time-consuming, many scientists have developed and used mathematical methods to estimate soil properties (Schloeder et al., 2001; Thomas et al., 2000; Yemefack et al., 2005). McBratney et al. (2003) introduced and developed the SCORPAN model; several researchers have used geostatistical techniques (Emadi and Baghernejad, 2014; Mousavifard et al., 2013; Salehi et al., 2013; Wälder et al., 2008), expert systems (Smith et al., 2012; Van Zijl et al., 2012; Zhu et al., 2001), unsupervised classification (Boruvka et al., 2008; Triantifilis et al., 2012) or machine-learning methods (Behrens and Scholten, 2006; Bui and Moran, 2003; Kim et al., 2012; Lemercier et al., 2012; Stum et al., 2010). The most important purpose of digital soil mapping is to make a soil map for regions with no soil information using environmental low-cost variables. To achieve this objective, the proposed approach involves construction of a model in one place having data and extending it to areas

* Corresponding author at: Department of Soil Science, College of Agriculture, Shahid Bahonar University of Kerman, Afzalipour Campus, Pajoohesh Sq., Kerman Postal code: 7616914111, Iran.
E-mail addresses: f.abbaszadeh@ujiroft.ac.ir (F. Abbaszadeh Afshar), ayoubi@cc.iut.ac.ir (S. Ayoubi), a.jafari@uk.ac.ir (A. Jafari).

https://doi.org/10.1016/j.geoderma.2017.11.030
Received 14 March 2017; Received in revised form 5 November 2017; Accepted 23 November 2017

lacking information.

It is important that soil classes be surveyed based on the spatial distribution of the soil-forming factors or in other words, environmental variables, in both methods (TSM and DSM). Therefore, it can be implied that the inference model for prediction of soil properties/soil classes can be similar if similar geomorphology, hydrology and pedology processes occur or if the spatial distribution of environmental variables is similar in two areas (Caten et al., 2011; Grinand et al., 2008; Lagacherie et al., 2001). In the other words, if a model is developed for an area, this model can be generalized to the other similar areas. Therefore, extrapolation process can be a useful tool in reducing costs and time.

From the beginning of soil genesis to evolution, soils are influenced by environmental factors. Therefore, similarities in environmental factors in different regions increase the possibility of extrapolation. The idea of extrapolating environmental variables has previously been explored mainly to identify areas with similar climates for crop production all over the world. Prescott (1938) coined the term "homoclime", referring to areas or regions in the world with similar climate. Also, Mallavan et al. (2010) introduced a methodology called "homosoil", which assumes homology of soil forming factors (lithology, climate, and topography) between a reference area and the region of interest. With this definition, they presented a methodology for quantitative extrapolation of soil information between the region of interest and other reference areas with good coverage of soil data as well as with similar soil-forming factors. They used Gower similarity index for finding similarity of soil forming factors. The significance of this work and other similar studies is the limited time, lack of legacy soil information and finite financial resources to collect new soil samples in many places around the world. On the other hands, insufficient sampling wastes time and money because it cannot provide the required level of accuracy and precision for successful management (Mueller et al., 2004).

Arid and semi-arid regions cover approximately 36% of Earth's surface, one of which is Iran. In most of these regions, legacy soil information either is difficult to obtain or no data exist. In such conditions, we have to extrapolate the extracted model from similar areas. Mallavan et al. (2010) and Salehi et al. (2013) explained that the rules and models extracted from the reference area could be applied in the interest region. Therefore, our objectives in this study were i) to compare two survey areas in terms of soil forming factors using similarity index, ii) to assess the extrapolation of digital soil mapping model derived from reference region, i.e. Zarand district, to estimate soil classes in the interest region, i.e. Bam region, and iii) to estimate and compare costs between extrapolation and interpolation processes.

2. Materials and methods

2.1. Description of interest region (Bam)

The study area, an area of 100,000 ha, located in the Bam region, southeast Iran, about 200 km from the city of Kerman, between 58° 4′ 17″ to 58° 28′ 8″ E longitudes and 28° 52′ 51″ to 29° 9′ 29″ N latitudes (Fig. 1). The area is surrounded by mountains, mostly limestone and volcanic, from north-west to south-east. The major landforms in the study area include old and young alluvial fans, bajada, pediment, clay flat (playa) and hills. Mean annual precipitation and temperature of the region are 59 mm and 23 °C, respectively. The soil moisture and temperature regimes of the study area are aridic and hyperthermic, respectively (Soil Survey Staff, 2010).

2.2. Description of reference area (Zarand region)

Zarand region, the area studied by Jafari et al. (2012, 2012), is located between 56° 16′ to 56° 36′ E and 30° 37′ to 30° 53′ N, about 70 km far from Kerman Province, southeast Iran (Fig. 1). It covers an area of 90,000 ha, surrounded by mountains (limestone, dolomite, and shale). Major parent materials principally include limestone, dolomite and shale. Main landforms in the area include alluvial fans, coalescing alluvial fans (bajada), playa, hills and sand dunes. The soil moisture and temperate regimes of the Zarand region are similar to those of Bam (Jafari et al., 2012, 2012). Pistachio is the most important crop in the study area, which is mostly being grown in the playa. The study area has an aridic soil moisture regime with mean annual precipitation, temperature, and potential evapotranspiration of 61 mm, 22 °C and 1750 mm, respectively.

2.3. Ancillary environmental variables derived from the reference and interest regions

The ancillary variables included terrain attributes, remote sensing indices and geomorphology map.

i- Relief attributes: These attributes were derived from a Digital Elevation Model (DEM) at a spatial resolution of 30 m downloaded from the website (METI and NASA, 2012) and were calculated using System for Automated Geoscientific Analyses (SAGA GIS) and ArcGIS (ESRI). The primary and secondary terrain attributes drawn from DEM included slope, curvature, plan curvature, profile curvature, aspect, stream power index (SPI), total insolation (TI), SAGA wetness index (SWI), topographic wetness index (TWI), LS factor, multi-resolution ridge top flatness (MRRTF) and multi-resolution valley bottom flatness (MRVBF) (Olaya, 2004).

ii- Remote sensing indices: One scene of the Landsat Enhanced Thematic Mapper (ETM) acquired in 2005 was downloaded (U.S. Geology Survey, 2005) and used to extract remote sensing indices including the Normalized Difference Vegetation Index (NDVI; Boettinger et al., 2008), Ratio Vegetation Index (RVI; Pearson and Miller, 1972), Perpendicular Vegetation Index (PVI; Richardson and Wiegand, 1977) and Clay Index (CI; Boettinger et al., 2008).

iii- Geomorphology Map: Air photo interpretation (API) was used to map geomorphological entities based on their formation processes, general structure(s) and morphometries. The geomorphological entities were defined through a nested geomorphic hierarchy defined by Toomanian et al. (2006). During stereoscopic delineation on landscapes in the study area, we employed our existing knowledge in soil-landscape relationships together with geology, topography and geomorphology. Then, interpreted air photos of the study area were imported into a GIS environment and geomorphic surfaces were mapped and inserted in GIS via on-screen digitization following an ortho-photo geo-referencing. 18 and 17 geomorphic surfaces were identified in Zarand and Bam regions, respectively (Figs. 2 and 3, Table 1). Sampling was done mostly on the basis of the geomorphology map in Bam and Zarand regions.

The stratified sampling scheme was adopted for the Bam and Zarand district using digital maps of geology, geomorphology and topography for stratification (Figs. 2 and 3). Sampling strata were defined so as to represent the differences in landform (geomorphology), topography (DEM) and lithology. Within each stratum, sampling locations were randomly chosen so that the sample size was proportional to the stratum area. This resulted in 126 profiles in both areas, which were then described, sampled, analyzed and classified using the USDA soil classification system (Soil Survey Staff, 2010).

All environmental variables and geomorphology (geomorphic surfaces) were projected to the same reference system (WGS 84 UTM 40 N) and also the same resolution. The values of the terrain attributes and levels of geomorphology map were then converted into a table for all the point locations. Soil great groups for all the locations were also added to this table which was then imported into R software (R Development Core Team, 2013) for predictive mapping.