Catena 173 (2019) 71–82

Contents lists available at ScienceDirect

Catena

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

Evaluating the spatial and vertical distribution of agriculturally important nutrients — nitrogen, phosphorous and boron — in North West Iran

Farzin Shahbazi[a,*], Philip Hughes[b], Alex B. McBratney[b], Budiman Minasny[b], Brendan P. Malone[b]

[a] Soil Science Department, Faculty of Agriculture, University of Tabriz, Iran
[b] Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia

ARTICLE INFO

Keywords:
Covariates
Cubist model
Digital soil mapping
Random forest
Spatial distribution
Spline depth function

ABSTRACT

Soil legacy data is ubiquitous and usually contains routine soil analysis information. In Iran, like most places, legacy soil data constitutes genetic horizon soil information recorded from excavated soil profiles. Describing and sampling from each genetic horizon is assumed to be heterogeneous from site to site. Digital soil mapping (DSM) using observed data is valuable because it provides a means to exploit the available information together with leveraging commonly available information by way of environmental covariates. It creates a much more detailed view of soil at the landscape scale. The purpose of this paper is to model and map the spatial distribution of nitrogen, phosphorous and boron at four standardized depths: 0–15, 15–30, 30–60, 60–100 cm, in an area of 7300 ha in the north west of Iran, and compare different model types. To circumvent the issue of heterogeneous soil depth observations from site to site, mass-preserving soil depth function splines were used to harmonise the soil profile observed data to the aforementioned standard depths. This facilitated the spatial modelling of each of the target variables for each standard depth with the aim of creating digital soil maps. Twenty-three covariates were extracted from a publically available digital elevation model (DEM) as well as freely available Landsat 8 ETM$^+$ imagery. The DEM-derivative covariates used in this study were divided into three main categories: i) Morphometry; ii) hydrology; and iii) lighting visibility. Both Random Forest and Cubist were assessed as candidate models for predicting each target variable. The results showed that Cubist was the most accurate method. Terrain attributes play an important role in estimating N, P, and B, while optical images do not have significant role. The most important findings of this paper in terms of environmental hazards are that the inundated regions in the west part of the study area are susceptible to boron contamination, providing future guidance for remediation.

1. Introduction

Soil nitrogen (N) and phosphorus (P) are important macronutrients which can limit or co-limit plant growth (Li et al., 2016). Soil boron (B) is also important to plant development as a micronutrient (Tariq and Mott, 2007). Boron has also been linked to various toxicological issues too as shown in the work of Assadpour et al. (2017) in north-western Iran. Understanding the spatial variation of these nutrients will result in better management plans and assessment of potential environmental hazards.

Following the earlier work on soil forming factors (Jenny, 1941), digital soil mapping (DSM) is invaluable to understanding the spatial variation of soil properties as it provides an empirical framework for soil type or attribute mapping based on spatial data related and pseudo-related to the soil forming factors using numerical functions or models (McBratney et al., 2003). The State-of-the-Art of DSM is well understood in Iran. However, it has been focused only on a relatively small number of readily measured soil properties such as soil organic carbon and clay contents (Taghizadeh-Mehrjardi et al., 2016). As management parameters with strong spatial dependence (patchy distribution) will be more readily managed and an accurate site-specific fertilization schemes for precision farming more easily developed (Lopez-Granados et al., 2002), there is a requirement of thematic digital maps related to some macro- and micronutrients e.g. N, P. and B.

Digital soil mapping employs mathematical and statistical models which combine information from soil observations with information contained in environmental variables and remote sensing images to produce predictions of properties over a large scale at a defined resolution (Dobos et al., 2006). Numerous prediction methods have been utilized to find linear and non-linear relationships between soil organic

* Corresponding author at: Soil Science Department, Faculty of Agriculture, University of Tabriz, 5166616471, Iran.
E-mail address: shahbazi@tabrizu.ac.ir (F. Shahbazi).

https://doi.org/10.1016/j.catena.2018.10.005
Received 22 December 2017; Received in revised form 28 September 2018; Accepted 6 October 2018

carbon and ancillary data such as from a digital elevation model (DEM) and Landsat imagery (Hengl et al., 2015; Minasny et al., 2013; Malone et al., 2009; Mora-Vallejo et al., 2008).

Recent developments in DSM have highlighted the utility of methods to map the vertical and lateral variability of soils (Malone et al., 2017; Taghizadeh-Mehrjardi et al., 2016). The methodology is loosely termed pseudo 3-dimensional soil mapping. The digital soil information that is achieved from this 3-D soil mapping provides an ability like never before to properly represent soil within all environmental modelling and management endeavours. Whatever the terminology, despite some successes in DSM (Pahlavan-Rad et al., 2014; Taghizadeh-Mehrjardi et al., 2014) – albeit in relatively small mapping extents - the application of 3-D DSM methods has not sufficiently been examined in detail within Iran. As described in Malone et al. (2017), there are several methodologies that are potentially at hand. One method is to combine soil depth functions with spatial modelling of continuous soil attributes as exemplified in Malone et al. (2009). This is a two-step procedure and first involving the fitting of splines, followed by spatial modelling of the target variable for each standardized depth. More recently, one-step approaches such as that in Orton et al. (2016) and Poggio and Gimona (2014) have been proposed. While the one-step approaches are more mathematically concise and appealing to some for that matter, the two-step approach developed by Malone et al. (2009) has endured because of its flexible nature. For example, the values retrieved from a fitted spline to soil data at given standard depth are both soil attribute information and parameters of the spline. The spline fitted data at given depths represent just a different reality of the observed data, and can also be used to retrieve those actual observations when the predicted values are used as inputs in the spline model. Furthermore, the two-step approach does not limit the soil modeller to using linear-based spatial models, meaning that the whole gamut of data mining and machine learning approaches can be considered (Malone et al., 2018).

This research aims to investigate the spatial variation of N, P and B in a study area with north-western Iran using the combination of spline depth functions coupled with different data mining techniques for a comparative analysis. These models include the Random Forest and Cubist data mining algorithms. The created maps may help us to assess the occurred environmental hazard across the study area.

2. Material and methods

2.1. Study area

This study was focussed upon 7300 ha extent of land in East Azerbaijan Province, Iran (Fig. 1). There are about 20 villages as well as a permanent river namely the Ahar chay within the study area.

The study area is represented by different kinds of land uses (e.g. cereal crops and apple orchards) as well as different lithology (e.g. limestone, old alluvium and volcanic-sedimentary) (Anonymous, 2012). It lies between the latitudes of 38° 24′ 04″ and 38° 28′ 33″ North and the longitudes of 47° 00′ 00″ and 47° 07′ 43″ East. The climate is semiarid. Annual rainfall and temperatures on average are 295 mm and 11 °C, respectively. Average annual maximum and minimum temperatures are 16.3 °C and 5.3 °C which was reported for July and February, respectively. The humidity index is 0.45. The humidity index was calculated with CDBm$^+$, a software package within MicroLEIS DSS (Shahbazi and Jafarzadeh, 2010). The elevation varies from 1281 to 1683 m a.s.l. The main physiographical units in the study area are described as flat, alluvial plains, hillsides and mountains (Shahbazi et al., 2014).

2.2. Environmental covariates

Due to variation of elevation and parent material and even land uses, spatial distribution of N, P and B is likely to be estimated as some function of given environmental and land cover data. For this purpose, a DEM and Landsat imagery spectral data were used in this study.

All covariates used in this study were aligned to the same grid cell resolution and extent. Here, a 30 m grid was used and alignment of grids was performed using cubic spline resampling where needed. The coordinate reference system used in this study was WGS1984 UTM Zone 38.

2.2.1. DEM derived covariates

Derivatives of the DEM (described below) were estimated using various functions made available in both ArcGIS (ESRI, 2011) and SAGA GIS (Conrad et al., 2015). The flowchart of the procedures is presented in (Fig. 2).

Terrain analysis is an integral component of DSM (McKenzie et al., 2000). Using the available DEM, we generated a number of derivatives to which were classified under three broad categories. 1) Morphometry: with derivative including slope, aspect, and curvature (plan and profile). 2) Hydrology: which include the derivatives catchment area, multi resolution indices of valley bottom flatness (MrVBF) and ridge top flatness (MrRTF). 3) Lighting visibility: potential incoming solar radiation. DEM derivatives were classified. Slope, aspect and curvature are local morphometric terrain parameters. Plan and profile curvature are also horizontal and vertical components of curvature (Tarboton, 1997). Modified catchment area describes width and specific catchment area (Hengl and Reuter, 2008). Multi resolution indices of valley bottom flatness (MrVBF) and ridge top flatness (MrRTF) are two morphometric parameters that as the names suggest can identify areas of flatness at different scales in valley bottoms and ridge areas respectively (Gallant and Dowling, 2003). Specifically, MrVBF is a topographic index designed to identify areas of deposited material at a range of scales based on the observations that valley bottoms are low and flat relative to their surroundings and that large valley bottoms are flatter than smaller ones. Zero values indicate erosional terrain with values 1 and larger indicating progressively larger areas of deposition. With slight modification to the MrVBF algorithm, the same analysis can be performed for ridge top areas to estimate MrRTF. Potential incoming solar radiation is a topoclimatic variable that is used as a parameter for evaluating the positional aspect effect in a landscape. Derived from the DEM, this parameter is evaluated over a temporal range of dates, taking into account sun position, location and sunrise and sunset times (Wilson and Gallant, 2000).

2.2.2. Covariates derived by Landsat 8 ETM$^+$

Landsat 8 ETM$^+$ imagery acquired on July 10, 2013 was selected for further analysis in this project. This scene was selected due to minimal cloud coverage and maximum soil surface exposure. A brief description of auxiliary data derived by Landsat 8 imagery is summarized in Table 1.

Landsat 8 spectral bands 2 to 7 were selected as six individual bands with a collective wavelength range between 0.452 and 2.294 μm (blue, green, red, near infrared, shortwave infrared one and two). Clay index (Breunig et al., 2008) and Salinity Ratio (Taylor et al., 1996) were calculated to represent parent material and soil factors across the study area. Normalized difference vegetation index (NDVI) was also calculated. NDVI ranges between −1.0 and 1.0, and mostly represents the saturation of green for higher values and corresponds to actively growing vegetation. Any negative values are mainly generated from clouds, water and snow, while values near zero are mainly generated from rock and bare soil. RVI (ratio vegetation index), and MSAVI2 (modified soil adjusted vegetation index) were also calculated to represent the vegetation and soil situation at the study area (Qi et al., 1994; Major et al., 1990). The index of MSAVI2 minimises the effect of bare soil on the SAVI. Fig. 3 represents some calculated auxiliary rasters for the study area.