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.

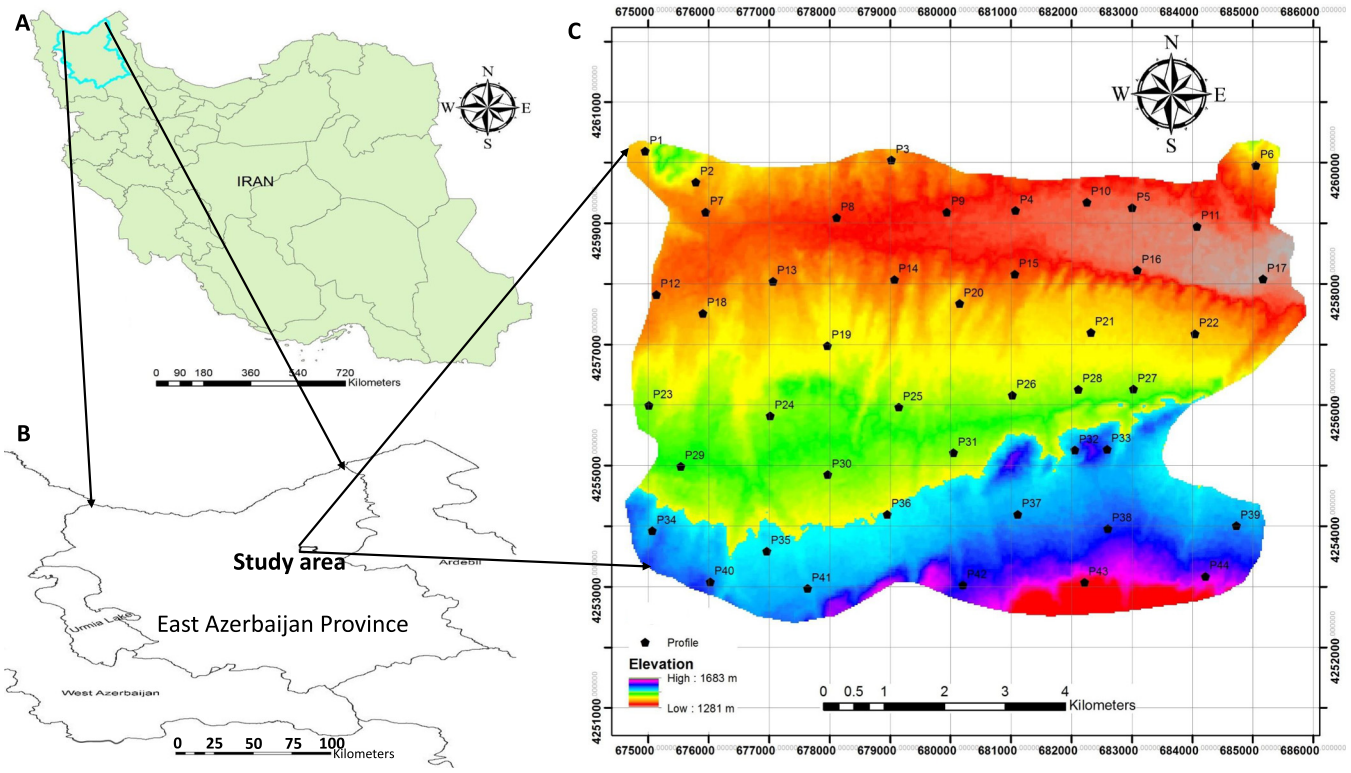

Fig. 1. Location of East Azerbaijan Province (A), study area (B) and DEM accompanying with spatial distribution of benchmark profiles (C).

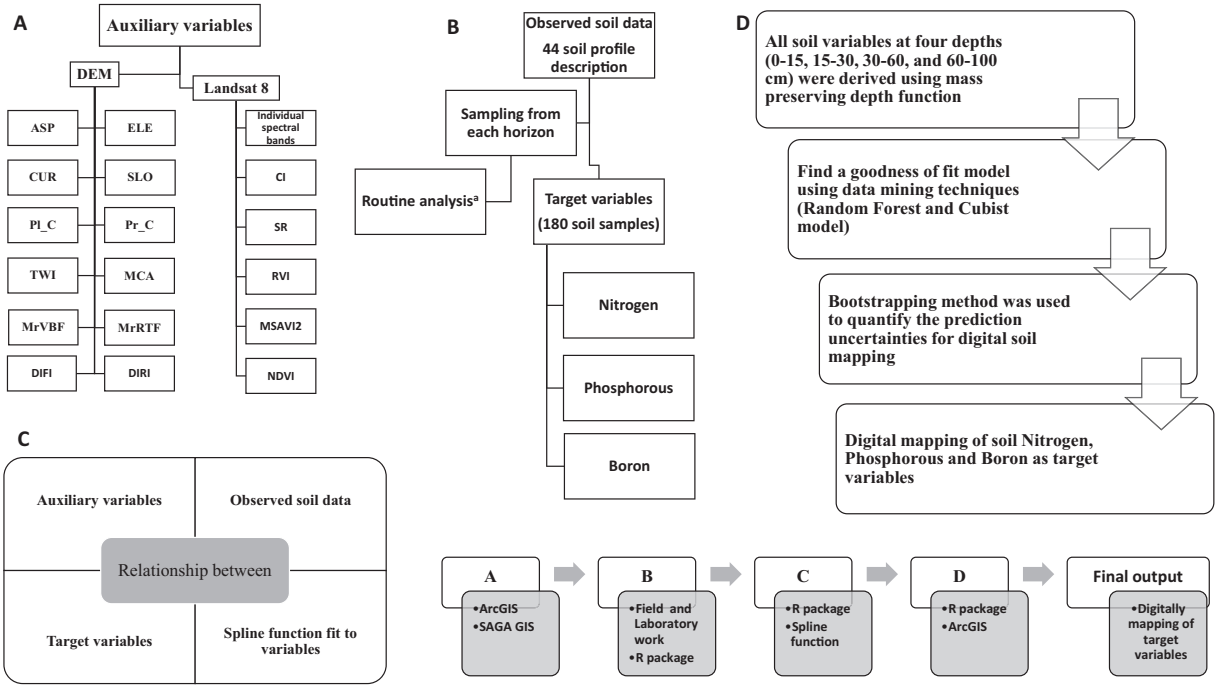

Fig. 2. Simplified algorithm of research processes.

[a]: Routine analysis was not used in this work (i.e. texture, bulk density, acidity, electrical conductivity, organic carbon, carbonate calcium equivalent, cation exchange capacity, sodium absorption ratio, and exchangeable sodium percent); ASP: aspect; ELE: elevation; SLO: slope; CUR: curvature; Pl_C: plan curvature; Pr_Curvature: profile curvature; MCA: modified catchment area; TWI: topographic wetness index; MrRTF: multiresolution index of the ridge top flatness; MrVBF: multiresolution index of valley bottom flatness; DIFI: diffuse insolation; DIRI: direct insolation; Individual bands (LSB2: Landsat band 2; LSB3: Landsat band 3; LSB4: Landsat band 4; LSB5: Landsat band 5; LSB6: Landsat band 6; LSB7: Landsat band 7); CI: clay index; SR: salinity ration; RVI: ratio vegetation index; MSAVI2: modified soil adjusted vegetation index; NDVI: normalized difference vegetation index.

Table 1
Summary of applied auxiliary data variables considered in this study.

Auxiliary data	Description	Definition
Individual bands	Blue, green, red, NIR, SWIR1, SWIR2	B2-B7
Vegetation and soil indices	NDVI	(NIR - red) / (NIR + Red)
	RVI	(SWIR1) / (NIR)
	MSAVI2	$(2*\text{NIR} + 1-\sqrt{(2*NIR + 1)^2 - 8*(NIR - Red)})$ *0.5
Geology indices	CI	(SWIR1/SWIR2)
	SR	(Red - NIR) / (Red + NIR)

NIR: near infrared; SWIR1: shortwave infrared 1; SWIR2: shortwave infrared 2; NDVI: normalized difference vegetation index; RVI: ratio vegetation index; MSAVI2: modified soil-adjusted vegetation index; CI: clay index; SR: salinity ratio.

2.3. Soil data base

2.3.1. Sampling and analysis

An efficient sampling method is needed to cover the whole range of the aforementioned environmental variables. To achieve samples that capture the variability of each auxiliary variable, Latin hypercube sampling method (LHS) conditioned by these environmental data offers a useful solution (Minasny and McBratney, 2006). Accordingly, forty-four soil profiles were excavated across the study area (Table 2), following a conditioned Latin hypercube sample of the available covariate data. Site access and budget constraints limited the number of samples. The soil profiles were described and classified based on their genetic horizons. In total, 180 horizons were observed, and samples were collected from each of the horizon for laboratory analysis. The dominant soils in the study area are Inceptisols (Soil Survey Staff, 2014). Typic Calcixerepts were the main sub-group. The shallowest soil profile was about 50 cm, while the deepest was > 190 cm. Although a comprehensive suite of analyses (physical and chemical) were conducted on the collected, not all were used in this research. In this study, we only considered soil N, P (associated with fertility) and B (correlated with EC and SAR and represents the toxicity). Total soil N (Bremner and Mulvaney, 1982), P (Olsen and Sommers, 1982), and B (Bingham, 1982) were analyzed using the Kjeldahl, Olsen and Carmine methods respectively.

2.3.2. Fitting soil mass-preserving splines with R

The biggest limitation to advancing DSM in both the vertical and lateral dimensions is the disparity in observation depths from one profile to the next within the soil data. A relatively straightforward approach is first to harmonise the soil profile information by fitting a depth function to each profile, then integrating the depth function to output information at a set of standardized depths for all profiles. This is followed by spatial modelling individually for each standard depth. In this study, all collected soil profile data were harmonised using mass preserving spline depth functions (Bishop et al., 1999) at four standard depths (H1: 0–15, H2: 15–30, H3: 30–60, and H4: 60–100 cm). Specifically, we followed the R coded routine for fitting mass preserving splines as demonstrated in Malone et al. (2017).

With the splines fitted to the observed soil data and integrated to outputs set at the prescribed depth intervals, our next step was to develop the spatial models. This involved the spatial intersection of the soil data with a stacked library of environmental covariates. This step requires two data sets:

- Spatial point vectors of data representing N, P, and B observations for each of the standard depth intervals at each sample sit location.
- A stack of rasters all at the same extent and resolution. In this study these consisted of the 23 environmental covariates that were compiled and described in Section 2.2.2.

The method of intersection of the point observations and the covariates was the nearest neighbour sampling, following the R coded routine in Malone et al. (2017) using the auxiliary R packages: “raster” (Hijmans, 2016), “sp” (Bivand et al., 2013) and “rgdal” (Bivand et al., 2016).

2.4. Data mining techniques

This study compares the applicability of Cubist and Random Forest models in the study area for each target variable. The two models are described in greater detail below.

2.4.1. Random Forest

Tree-based methods are atypical statistical models – they do not utilise distributions, likelihoods or design matrices; metrics typically associated with modelling. Regression Trees are tree-based models that have been widely used in DSM (Taghizadeh-Mehrjardi et al., 2016; Ivezić et al., 2012; Giasson et al., 2011; Rudiyanto et al., 2018). Random Forest may also be used for both regression and classification purposes (Dharumarajan et al., 2017). Random Forest operates via a resampling approach or boosting, where for regression, the prediction is the average of the individual tree outputs, whereas in classification, the trees vote by majority on the correct classification mode (Grimm et al., 2008). Fitting a Random Forest model usually requires the user to select arbitrary values or optimise a number of tuning parameters. The main parameters that can be adjusted include the number of trees to build “*ntree*”, and the number of variables (covariates) that are randomly sampled as candidates at each decision tree split “*mtry*”. In this study, these two parameters were optimised by iterating “*mtry*” values from 1 to the total number of covariates (23) and “*ntree*” values from 100 to 10,000 in increments of 100 (Hengl et al., 2015). Our objective for the iteration procedure was the maximisation of prediction accuracy based on the out-of-bag data from each Random Forest model. Due to our small data set, we used all the available data in order to optimise the *ntree* and *mtry* variables. To facilitate Random Forest model fitting in R, we used the Random Forest R package (Liaw and Wiener, 2002).

2.4.2. Cubist model

The Cubist model has seen much usage in soil science studies (Ma et al., 2017; Rudiyanto et al., 2018). It has an ability to “*mine*” non-linear relationships in data but does not have the issues of finite predictions that occur for regression trees. The Cubist model first partitions the data into internal subsets in which their characteristics are similar with respect to the target variable and the covariates. In this way the Cubist model is similar to a regression tree model, except that at the terminal nodes, a multivariate linear model is fitted based on covariate that were significant for the given data partition. The Cubist algorithm of Quinlan (1992), has three parameters that can potentially be optimised. These include: the maximum number of rules or partitions in the data that can be explored; the number of committee of boosted iterations of the algorithm (similar to the *ntree* parameter in the Random Forest model), and the extent of allowable extrapolation. Some self-governance in the model fitting internalised within the function ensures the partitions in the data are not made to the specifications of the user, if those specifications are not warranted, for example, the creation of unnecessary number of rule sets. From experience, this feature helps to avoid overfitting, which is often difficult to control when comparing to the Random Forest model. To facilitate Cubist model fitting in R we used the Cubist R package (Kuhn et al., 2016).

2.5. Model goodness of fit

Some of the common goodness of fit diagnostics includes the root mean square error (RMSE), bias, coefficient of determination (R^2)

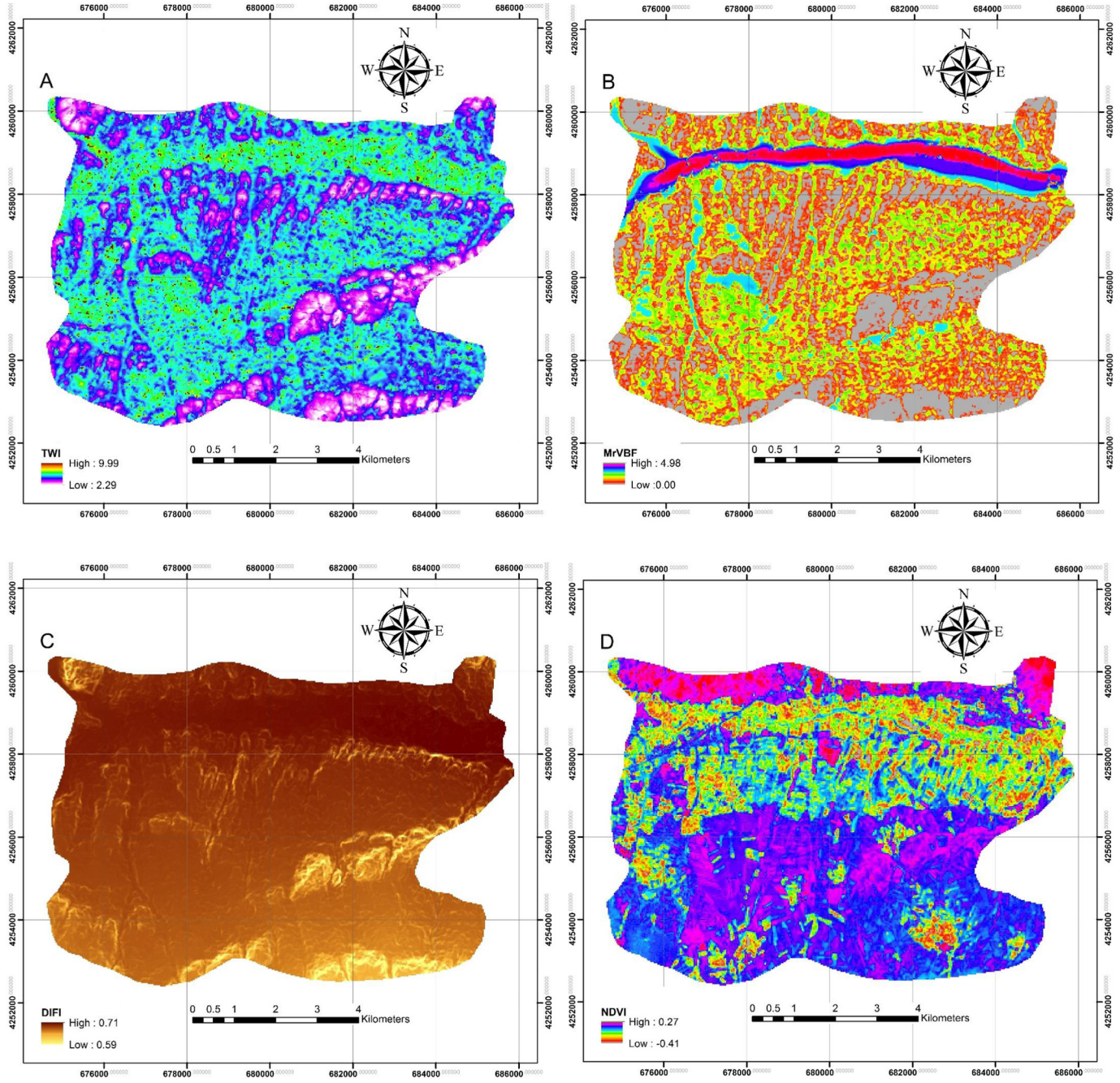

Fig. 3. Created some environmental auxiliary variables using DEM- and Landsat spectral-data across the study area.
TWI: Topographic wetness index (A); MrVBF: Multi-resolution valley bottom flatness index (B); DIFI: Diffusion insolation index (C); NDVI: Normalized difference vegetation index (D).

value, and concordance correlation coefficient (CCC). The RMSE is popular Eq. (1) and bias is also called the mean error of prediction Eq. (2). The R^2 is a measure of goodness of fit between the observations and their corresponding predictions and measures the precision of the relationships Eq. (3). CCC, or more formally—Lin's concordance correlation coefficient, on the other hand is a single statistic that both evaluate the accuracy and precision of the relationship. It is often referred to as the goodness of fit along a 45° line Eq. (4).

$$RMSE = \sqrt[2]{\frac{\sum_{i=1}^{n} (obs_i - pred_i)^2}{n}} \quad (1)$$

$$bias = \frac{\sum_{i=1}^{n} (obs_i - pred_i)}{n} \quad (2)$$

$$r = \frac{\sum_{i=1}^{n} (obs_i - \overline{obs})(pred_i - \overline{pred})}{\sqrt[2]{\sum_{i=1}^{n} (obs_i - \overline{obs})^2} \sqrt[2]{\sum_{i=1}^{n} (pred_i - \overline{pred})^2}} \quad (3)$$

where, *obs* is the observed soil property, *pred* is the predicted soil property from a given model, and *n* is the number of observations. $\overline{obs}$ and $\overline{obs}$ are absolute value of observed and predicted soil properties respectively.

Table 2
Summary of the data of the described soil profiles across the study area.

Soil profile	Longitude (east)	Latitude (north)	Depth (cm)	Number of horizons	Soil sub-group
1	47° 00′ 20″	38° 28′ 22″	75	2	Typic Xerorthents
2	47° 01′ 54″	38° 28′ 05″	69	2	Typic Xerorthents
3	47° 03′ 07″	38° 28′ 15″	50	2	Typic Xerorthents
4	47° 04′ 32″	38° 27′ 46″	55	2	Typic Xerorthents
5	47° 05′ 51″	38° 27′ 46″	199	5	Typic Calcixerepts
6	47° 07′ 17″	38° 27′ 07″	150	5	Vitrandic Calcixerepts
7	47° 01′ 01″	38° 29′ 49″	190	5	Vertic Haploxeralfs
8	47° 02′ 29″	38° 27′ 44″	185	5	Vertic Haploxeralfs
9	47° 03′ 44″	38° 27′ 46″	190	6	Vertic Haploxeralfs
10	47° 05′ 20″	38° 27′ 50″	135	4	Aquic Haploxerepts
11	47° 06′ 35″	38° 27′ 35″	125	3	Vertic Haploxerepts
12	47° 00′ 25″	38° 27′ 05″	120	4	Typic Calcixerepts
13	47° 01′ 45″	38° 27′ 11″	89	3	Typic Haploxerepts
14	47° 03′ 08″	38° 27′ 11″	145	4	Typic Calcixerepts
15	47° 04′ 30″	38° 27′ 12″	85	3	Typic Calcixerepts
16	47° 05′ 54″	38° 27′ 13″	170	5	Typic Calcixerepts
17	47° 07′ 19″	38° 27′ 06″	119	5	Fluventic Haploxerepts
18	47° 00′ 57″	38° 26′ 55″	135	4	Typic Haploxerepts
19	47° 02′ 21″	38° 26′ 36″	180	3	Typic Calcixerepts
20	47° 03′ 52″	38° 26′ 57″	190	5	Typic Calcixerepts
21	47° 05′ 21″	38° 26′ 40″	150	4	Vertic Calcixerepts
22	47° 06′ 32″	38° 26′ 38″	130	4	Vertic Calcixerepts
23	47° 00′ 18″	38° 26′ 06″	155	6	Typic Calcixerepts
24	47° 01′ 41″	38° 25′ 59″	148	5	Typic Calcixerepts
25	47° 03′ 09″	38° 26′ 02″	180	3	Typic Calcixerepts
26	47° 04′ 26″	38° 26′ 07″	99	4	Typic Haploxerepts
27	47° 05′ 49″	38° 26′ 09″	150	4	Calcic Haploxerepts
28	47° 05′ 12″	38° 26′ 10″	145	4	Calcic Haploxerepts
29	47° 00′ 39″	38° 25′ 33″	160	5	Typic Haploxerepts
30	47° 02′ 19″	38° 25′ 27″	140	3	Typic Calcixerepts
31	47° 03′ 46″	38° 25′ 37″	148	4	Vertic Haploxerepts
32	47° 05′ 08″	38° 25′ 36″	160	4	Calcic Haploxerepts
33	47° 05′ 30″	38° 25′ 37″	160	4	Typic Calcixerepts
34	47° 00′ 19″	38° 24′ 59″	190	5	Typic Calcixerepts
35	47° 01′ 37″	38° 24′ 47″	170	5	Calcic Haploxerepts
36	47° 02′ 59″	38° 25′ 05″	180	5	Typic Calcixerepts
37	47° 04′ 28″	38° 25′ 03″	140	5	Typic Calcixerepts
38	47° 05′ 30″	38° 24′ 55″	135	4	Typic Calcixerepts
39	47° 06′ 57″	38° 24′ 55″	140	4	Typic Calcixerepts
40	47° 00′ 58″	38° 24′ 31″	160	4	Typic Calcixerepts
41	47° 02′ 04″	38° 24′ 26″	170	5	Calcic Haploxerepts
42	47° 03′ 50″	38° 24′ 25″	100	4	Calcic Haploxerepts
43	47° 05′ 13″	38° 24′ 27″	120	4	Typic Calcixerepts
44	47° 06′ 35″	38° 24′ 28″	100	4	Typic Calcixerepts

$$\rho_c = \frac{2\rho\sigma_{pred}\sigma_{obs}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} - \mu_{obs})^2} \quad (4)$$

where, μ_{pred} and μ_{obs} are the means of the predicted and observed values respectively. ${\sigma_{pred}}^2$ and ${\sigma_{obs}}^2$ are the corresponding variances. ρ_c is the correlation coefficient between the predictions and observations. These criteria were used to select the optimal fitted model to achieve digital lateral and vertical maps in four depths and entire of the study area.

2.6. Quantifying uncertainty analysis

Bootstrapping was used in this paper as it is a popular non-parametric approach for quantifying prediction uncertainties (Efron and Tibshirani, 1993). It is commonly used in digital soil mapping (Malone et al., 2018). Viscarra Rossel et al. (2015) performed bootstrapping for quantification of uncertainties across the very large mapping extent of the Australian continent at 100 m resolution. We found in this study too, that bootstrapping is amenable for smaller data sets as well. The bootstrap routine works for first selecting the number of iterations or bags to fit a model. In this study we selected arbitrarily 250 iterations. For each iteration 100% of the sample data size is selected at random with replacement. This procedure always leaves a small proportion of the data out of the model fitting procedure which is akin to the out-of-bag data from the Random Forest model. For each iteration, out-of-bag goodness of fit diagnostics measures (as described above) were evaluated. In order to map each target variable and for each depth, each bootstrap model was applied to the stack of covariates, to create a map. This procedure was performed separately for N, P, and B for the four depths, using the best selected modelling type.

2.6.1. Creation of maps with associated uncertainty estimates

After selecting the appropriate model to estimate the variability of N, P, and B, the main priority of this paper was to create thematic maps of the three target variables using some environmental covariates in selected depths over entire the study area with a distinct coordinate reference system. Using the outputs from the bootstrap modelling, the mean of the predictions from each bootstrap sample was derived. The variance of the predictions was then estimated from the bootstrap prediction maps using Eq. (5) in R environment.

$$Var(X) = \frac{1}{1-n}\sum_{i=1}^{n}(xi - \mu)^2 \quad (5)$$

where, the symbol μ is the mean bootstrap prediction, and x_i is the *ith* bootstrap map. From the estimate of the variance we were then able to derive the upper prediction limit, lower prediction limit, and prediction limit range of each map. The procedure for advancing this workflow is

Table 3
Statistical summary of soil N, P, and B content at four standardized depths.

Layer (cm)	Min	Max	Average	Std. Dev.	CV	Q25*	Q50*	Q75*
N (H1)	0.001	0.23	0.09	0.06	60.35	0.06	0.08	0.12
N (H2)	0.007	0.17	0.07	0.03	48.81	0.04	0.07	0.09
N (H3)	0.001	0.14	0.05	0.03	65.36	0.02	0.05	0.07
N (H4)	0.001	0.09	0.04	0.03	67.99	0.02	0.04	0.06
P (H1)	0.001	49.85	10.13	11.31	111	3.92	6.61	11.16
P (H2)	0.49	35.21	6.43	7.17	112	2.71	3.99	6.31
P (H3)	0.52	30.31	3.32	5.38	162	1.09	1.89	3.31
P (H4)	0.11	29.21	2.49	4.42	177	0.76	1.39	2.68
B (H1)	0.002	3.03	0.81	0.67	83.33	0.27	0.70	1.12
B (H2)	0.04	2.71	0.76	0.58	75.41	0.31	0.65	1.03
B (H3)	0.001	2.89	0.82	0.67	81.85	0.44	0.63	1.04
B (H4)	0.008	3.61	0.83	0.74	89.71	0.36	0.71	1.05

H1: standardized depth 0–15 cm; H2: standardized depth 15–30 cm; H3: standardized depth 30–60 cm; H4: standardized depth 60–100 cm; N: soil nitrogen (g/100 g); P: soil phosphorous (mg/kg); B: Soil boron (mg/kg); Min: minimum; Max: maximum; Std. Dev: standard deviation; CV: coefficient of variation; Q25, Q50, and Q75 refers to the 25% quartile, median, and 75% quartile.

described in (Malone et al., 2017).

3. Results and discussion

3.1. Descriptive statistics of observed soil data

Descriptive statistics of N, P, and B are summarized in Table 3. The observations revealed that the mean and median are not equivalent, indicating the distribution these data deviates from normality. Subsequently, the data were normalized by square root transformation. The transformation was performed after the fitting of the mass preserving depth splines. According to the analysis, the average N and P levels ranged from 0.04 to 0.09 g/100 g and 2.49 to 11.31 mg/kg, respectively while B returned similar values from the top to bottom (0.76 to 0.83 mg/kg) in the study area. Nitrogen and phosphorous exhibited a decreasing trend with depth as expected but for B this trend was not present. The coefficient of variation in terms of P was high for the four standard depths. Although this coefficient was improved by about 55% through square root transformation, normalized P appeared to have the highest variability across the study area (CV equal to 60%) in the 60–100 cm depth. Overall, all target soil properties could be categorized as having moderate variability, a CV of 10% indicates a low variability and 10%–90% indicates a moderate variability, and CV > 90% indicates extreme variability (Fang et al., 2012).

3.2. Environmental characteristics

Some important environmental covariates of the study area were previously mapped in Fig. 3. Summary statistics of these rasters revealed that TWI varies from 2.29 to 9.99 across the entire area. 83% of the study area had moderate TWI values (5–7). In 12% of the area, TWI was > 7%. MrVBF was calculated from zero to five. It was then classified into three categories (class 1: < 1.21; class2: 1.21–3.14; and class 3: 3.14–4.98) for further visualization. The class 1 in an area extension of approximately 55% followed by class 2 (37%) and class 3 (8%) were categorized. Diffuse insolation also varied from 0.59–0.71. The three created class ranges (class1: < 0.68; class2: 0.68–0.7; class 3: 0.7–0.71) showed that 60% of study area was located in class 3, 36% in class 2, and 4% in class 1. Finally, as NDVI varied between −0.41 and 0.27 in the study area, it was classified into two classes of low and moderate (class1: < 0.1 corresponding to barren areas of rock, sand, or snow; class 2: 0.1–0.3 representing shrub and grassland). The highest values (0.6 to 0.8) were associated with temperate and tropical rainforests (Jones, 2014) which was much higher than any value that appeared on the study area. Almost 90% of the study area was classified in class 1 (barren areas of rock). The rest of the environmental covariates were analyzed similar to the rules previously explained.

3.3. Data mining techniques

3.3.1. Random forest

The results showed that the R^2 and concordance for the calibration models were high compared to the out-of-bag samples when using Random Forest. While Taghizadeh-Mehrjardi et al. (2016) had described the efficiency of the Random Forest Model for predicting soil organic carbon in Iran, and Hengl et al. (2015) used Random Forests with apparent success for mapping soil properties of Africa at 250 m resolution, we could not avoid issues of overfitting when using this model. We presume this is because of the small data set used in this study.

3.3.2. Cubist model

Table 4 shows the performance of the Cubist model on calibration data (in the bag) and on validation data (out of bag). These analyses are based on using the square root transformed values. For easy interpretation, the normalized square root of N, P, and B were nominated as NSN, NSP, and NSB respectively.

Generally, all predictions of NSN were quite good with Cubist calibration especially in H1:NSN (R^2 = 0.46) followed by H2:NSN (R^2 = 0.37), H3:NSN (R^2 = 0.19), H4:NSN (R^2 = 0.18) but the

Table 4
Comparison of the performance of the prediction of target values using Cubist calibration and Cubist validation.

Variable	calibration (in the bag)					Validation (out of bag)				
	R^2	CCC	MSE	RMSE	Bias	R^2	CCC	MSE	RMSE	Bias
H1:NSN	0.46	0.58	0.005	0.07	0.004	0.13	0.24	0.01	0.11	0.007
H2:NSN	0.37	0.49	0.003	0.05	0.007	0.11	0.15	0.006	0.08	0.004
H3:NSN	0.19	0.33	0.005	0.07	0.006	0.02	0.01	0.009	0.09	0.01
H4:NSN	0.18	0.27	0.004	0.06	0.004	0.01	0.01	0.007	0.08	0.006
H1:NSP	0.59	0.71	0.91	0.92	0.02	0.17	0.24	2.24	1.46	−0.07
H2:NSP	0.58	0.69	0.51	0.69	−0.02	0.11	0.16	1.51	1.21	0.04
H3:NSP	0.52	0.61	0.21	0.44	−0.03	0.01	0.05	1.55	1.19	−0.08
H4:NSP	0.41	0.51	0.24	0.46	−0.03	0.01	0.01	1.27	1.08	−0.11
H1:NSB	0.47	0.62	0.07	0.28	−0.004	0.22	0.38	0.13	0.36	−0.01
H2:NSB	0.47	0.61	0.05	0.23	0.02	0.21	0.37	0.09	0.31	0.007
H3:NSB	0.46	0.59	0.07	0.25	0.03	0.16	0.29	0.13	0.36	0.04
H4:NSB	0.46	0.59	0.08	0.28	0.03	0.11	0.21	0.17	0.41	0.03

H1: standardized depth 0–15 cm; H2: standardized depth 15–30 cm; H3: standardized depth 30–60 cm; H4: standardized depth 60–100 cm; NSN: normalized soil nitrogen (sqrt (g/100 g)); NSP: normalized soil phosphorous (sqrt (mg/kg)); NSB: normalized soil boron (sqrt ((mg/kg)); R2: correlation coefficient; CCC: concordance; MSE: mean square error; RMSE: root mean square error.

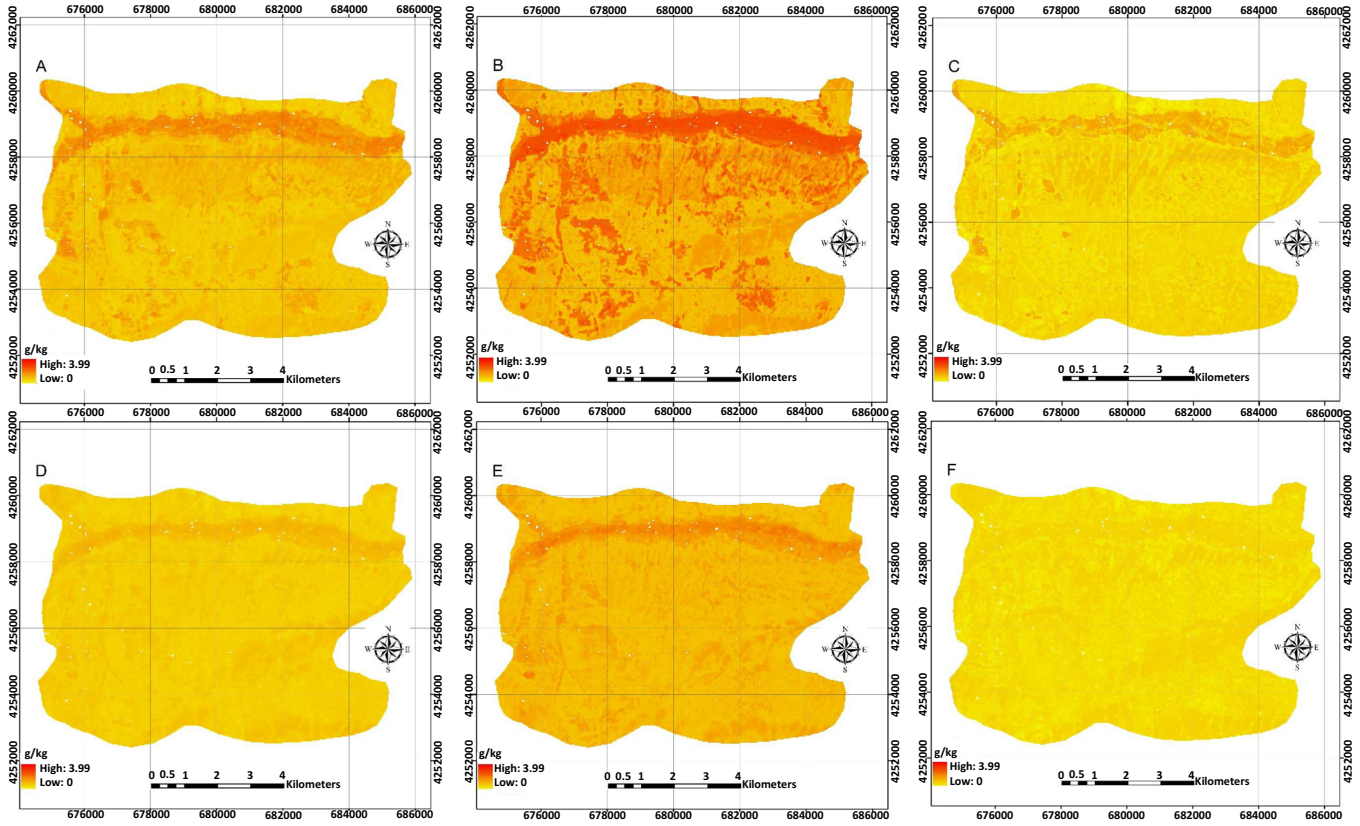

Fig. 4. Predicted maps of nitrogen content (g/kg) across the study area using Cubist model for the two standardized depths.
Mean, upper 90% limit prediction, and lower 90% prediction respectively for H1 (A–C); mean, upper 90% limit prediction, and lower 90% prediction respectively for H2 (D–F).

performance of R^2 against the validation did not compare well except for the first depth increment (R^2 = 0.11). A similar trend was also observed in the concordance coefficient. The trend of Cubist calibration for NSP was almost similar to NSN insomuch that the R^2 and concordance values decreased with depth. The distribution of P was assessed in H1 and H2 standardized depths. Cubist calibration for the four depths did not distinctly vary for B (R^2 = 0.46–0.47). Notwithstanding the decreasing performance of R^2, the concordance against validation was not good, but the Cubist model showed that it has more efficiency to explain the spatial variability of B in four standardized depths compared with N and P. This may be due to the prevailing agricultural practices which may be done in some parts of the study area - fruit trees have different requirements in C, N and water than cereal crops, and local knowledge may necessitate different management procedures. The challenge to estimate subsurface N and P using environmental covariates remained.

The useful feature of the Cubist model is that it does not unnecessarily overfit or partition the data (Malone et al., 2017). In general, the contribution of all environmental covariates in the Cubist model for the prediction of N, P and B at any acceptable depth included: aspect, elevation, TWI, MrRTF, MrVBF, RVI, and individual bands 3 and 7 were eight important covariates in the models among the 23 covariates. For H1:NSN it was found MrVBF was identified as a major driver in the model. Both TWI and MrVBF were important variables in the H2:NSN model. TWI was used more than MrVBF in the model (90% vs 17%). MrVBF also contributed in H1:NSP. RVI was used in H2:NSP model (74%). Bands 3 and 7 and also elevation contributed in the H1:NSB model. In terms of H2:NSB, elevation, MrVBF (68%), TWI (60%), and aspect (32%) played important roles in the model. MrRTF was utilized in the model of H3:NSB. For H4:NSB it was found that not only MrRTF but also aspect played important roles. Ma et al. (2017) have reported similar outputs for the role of environmental covariates in mapping of some surface soil properties in China. These terrain derived attributes are emerging as major predictors for elements in the subsurface horizons.

3.4. Digital soil mapping

According to the main aim of this paper and research flowchart (Fig. 2), digital soil maps of N, P and B were created across the study area. The presented values are the back-transformations of the modelled predictions.

3.4.1. Nitrogen

Fig. 4 shows spatial distribution of N (g/kg) for mean, upper 90% limit prediction and lower 90% limit prediction using Cubist model in two soil depth (H1and H2).

Nitrogen varies largely across the study area which it was distinctly shown for H1 (Fig. 4-A) compared to H2 (Fig. 4-D). The north part of the study area has a high amount of N at both depths while for the south part of the study area it was high at only H1. This might be attributed to the fact that the north part has low slope and is close to the permanent river namely the Ahar Chay and land use in this area is variable. There are some seasonal rivers in the study area which lie from the south to the north aspect. As nitrogen has a high correlation with organic matter, comparing the results with other findings revealed that predicted maps using the Cubist model for soil organic carbon at both surface from (Zhao et al., 2013) and subsurface from (Taghizadeh-Mehrjardi et al., 2016) were accurate. There are about 20 villages in the study area and it was expected that nitrogen content would be high near the villages due to some cultivation and agricultural practices. The absence of nitrogen in the predictions indicates that the villagers do not

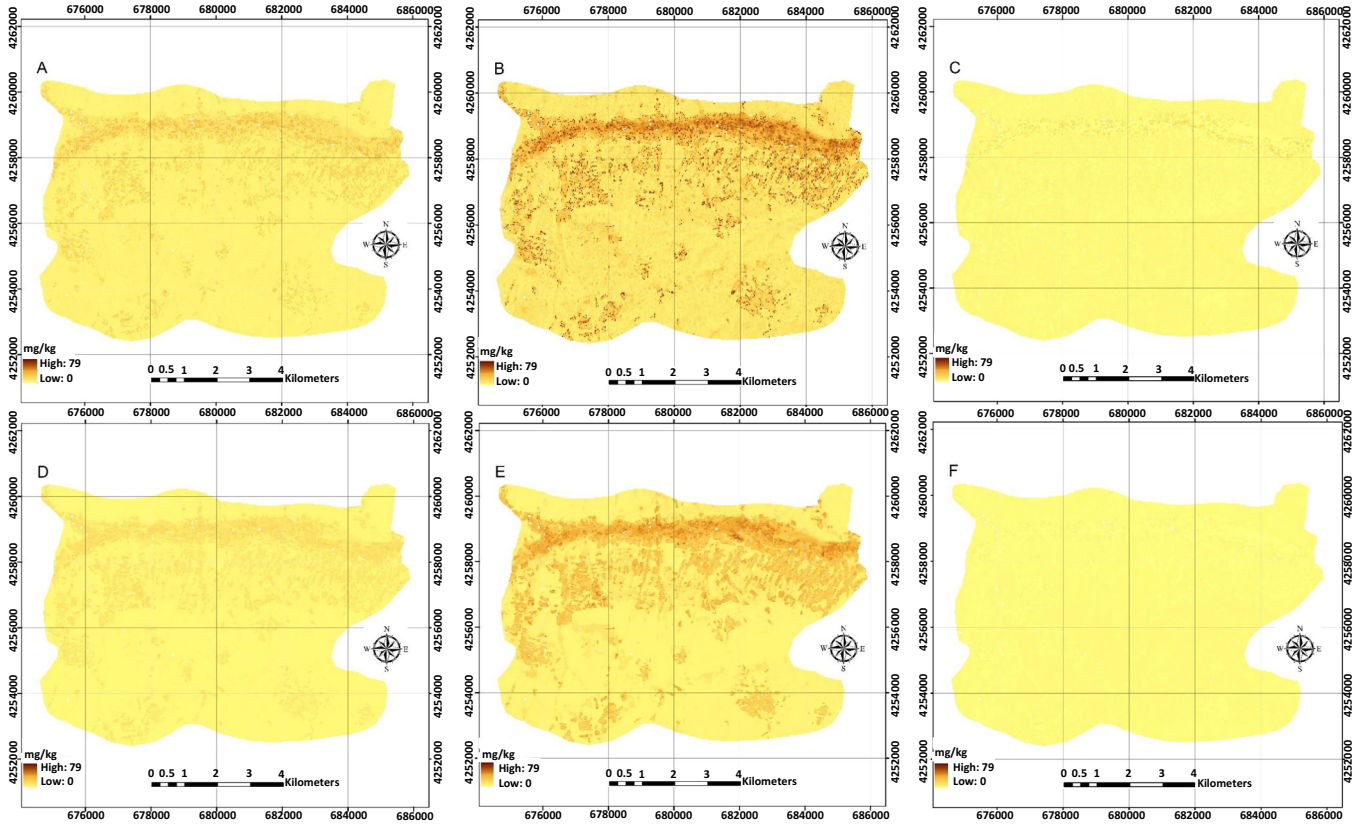

Fig. 5. Predicted maps of phosphorous content (mg/kg) across the study area using Cubist model for the two standardized depths.
Mean, upper 90% limit prediction, and lower 90% prediction respectively for H1 (A–C); mean, upper 90% limit prediction, and lower 90% prediction respectively for H2 (D–F).

Table 5
Attributes of boron prediction and uncertainty analysis of Cubist model entire the study area.

Depth	Mean*	90% upper prediction limit*	90% lower prediction limit*	Prediction interval range*
B (H1)	0.07–2.22	0.34–4.61	0.00–0.87	0.14–4.53
B (H2)	0.07–2.01	0.47–4.46	0.00–1.05	0.01–17.83
B (H3)	0.31–2.74	0.59–5.62	0.00–1.31	0.13–5.58
B (H4)	0.24–3.29	0.66–7.72	0.00–1.01	0.16–7.17

H1: standardized depth 0–15 cm; H2: standardized depth 15–30 cm; H3: standardized depth 30–60 cm; H4: standardized depth 60–100 cm; B: soil boron (mg/kg); *: (mg/kg).

tend to add nitrogen artificially. The prediction limit ranges for H1 and H2 were 0–3.74 g/kg and 0–1.41 g/kg respectively. Although the calculated range for two depths was small, the difference means that local management influences in the surface horizons and subsurface horizons are as we expect.

3.4.2. Phosphorous

Fig. 5 (A–F) also shows spatial distribution of P (g/kg) for mean, upper 90% limit prediction and lower 90% limit prediction using a Cubist model in two soil depth increments (H1and H2). For phosphorous, we could map for two depths using a similar method as with nitrogen prediction. The pattern of the Ahar Chay River was distinguished in the maps of phosphorous. The supplied environmental covariates in both models were extracted by DEM. There is no major difference between the phosphorous levels of H1 and H2. The prediction limit ranges were not mapped in this paper. The results demonstrated that the limitation range for H1 was 1.33–79.32 mg/kg while the limitation range was 0.38–54.35 mg/kg for H2. Therefore, the management practices might impact more on phosphorous content at the surface horizons. Tang et al. (2016) using ordinary kriging method reported that soil N and P stock have strong to moderate spatial dependence means that soil N stock was mainly controlled by intrinsic factors such as management practices compared to soil P. The distribution of P influences microbial activity on a regional scale (Cao et al., 2016). This can be used when predicting microbial ecology with DSM. In order to sample microbial units at large scales, taking into account such variables as phosphorous, accurate maps are a necessity.

3.4.3. Boron

The predicted maps for boron have got the most accurate maps among three studied variables at four standardized depths. The results of prediction are summarized in Table 5. Fig. 6 illustrates rescaled maps for visualization of B spatial distribution in four depths. The results showed that boron predictions have an increasing trend from the top of the study area to the bottom because it is mobile in the soil as well as subject to leaching (Price, 2006). Although, the capillary rise of soil moisture may cause a higher concentration of boron in H1 than in H2 (Nobel et al., 1997). The prediction interval values were 4.39 mg/kg, 17.82 mg/kg, 5.45 mg/kg, and 6.97 mg/kg for H1, H2, H3, H4 respectively. The prediction of B at these depths can produce an understanding of the most likely occurrences of toxicological problems.

4. Conclusions

In this study, the vertical and lateral variation of soil nitrogen, phosphorous, and boron were investigated using both the Random Forest and Cubist models. Our analysis revealed that the Random Forest model was not suitable for our data. We found the Cubist model to be the best model in this particular study. However, results showed that

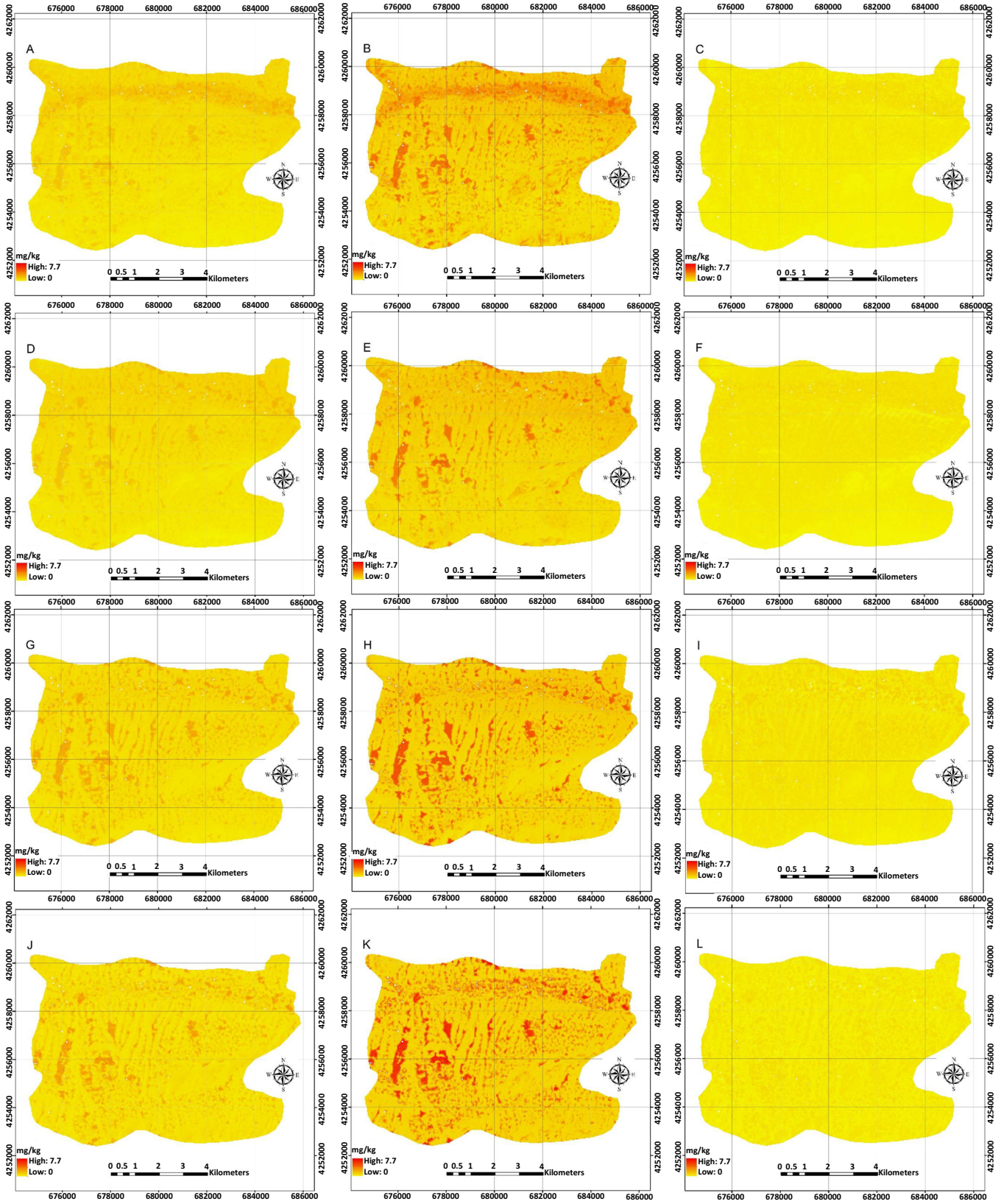

Fig. 6. Predicted maps of boron content (mg/kg) across the study area using Cubist model for the four standardized depths.
Mean, upper 90% limit prediction, and lower 90% prediction respectively for H1 (A–C); mean, upper 90% limit prediction, and lower 90% prediction respectively for H2 (D–F); mean, upper 90% limit prediction, and lower 90% prediction respectively for H3 (H–I); mean, upper 90% limit prediction, and lower 90% prediction respectively for H4 (J–L).

where the target variable was in the sub-surface, the models were not reliable.

The analysis revealed that terrain attributes play an important role for the estimation of subsurface boron content than the other two elements (nitrogen and phosphorous) in the study area. The boron maps produced in this study were the most accurate, when compared to those of soil nitrogen and phosphorous. Further visualization also revealed that the distribution of subsurface boron and the second horizons of nitrogen and phosphorous are close to the pattern of the seasonal and permanent rivers. This suggests that nitrogen and phosphorous may have been transported by illuviation. In this study, remote sensing data did not prove to be useful for the spatial prediction of either soil target variables.

This study revealed that DSM using important environmental covariates can have applications in planetary health and microbiology because those domains need accurate and precise soil samples (Cao et al., 2016). Soil nitrogen and phosphorous showed different spatial patterns across the whole study area indicating that fertilizers should be supplied with various ratios. The most important findings of this paper in terms of environmental hazards are that the inundated areas of the west part of the study area, according to boron concentrations, may be susceptible to contamination, which may require remediation.

Acknowledgment

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. This is a contribution to the work sabbatical program of Dr. Farzin Shahbazi in the University of Sydney. The first author would like to thank Prof. Alex McBratney for providing this opportunity.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.catena.2018.10.005. These data include the Google map of the most important areas described in this article.

References

Anonymous, 2012. Geological Map of Ahar 1:100000, Series Sheet 5466. Ministry of Mine and Metals, Geological Survey of Iran.
Assadpour, M., Heuss-Aßbichler, S., Jafari Bari, M., 2017. Boron contamination in the west of Lake Urmia, NW Iran, caused by hydrothermal activities. Procedia Earth Planet. Sci. 17, 554–557.
Bingham, F.T., 1982. Boron. In: Page (Ed.), Methods of Soil Analysis, 2nd ed. Chemical and Microbiological Properties, Agronomy Monograph 9.2 Part 2. pp. 431–447.
Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modeling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45.
Bivand, R.S., Pebesma, E., Gomez-Rubio, V., 2013. Applied Spatial Data Analysis with R, Second ed. Springer, NY.
Bivand, R.S., Keitt, T., Rowlingson, B., 2016. rgdal: bindings for the 'Geospatial' data abstraction library. In: R Package Version 1.2-15, . https://CRAN.R-project.org/package=rgdal.
Bremner, J.M., Mulvaney, C.S., 1982. Nitrogen-total. In: Page (Ed.), Methods of Soil Analysis, 2nd ed. Chemical and Microbiological Properties, Agronomy Monograph 9.2 Part 2. pp. 595–624.
Breunig, F.M., Galvão, L.S., Formaggio, A.R., 2008. Detection of sandy soil surfaces using ASTER-derived reflectance, emissivity and elevation data: potential for the identification of land degradation. Int. J. Remote Sens. 29, 1833–1840.
Cao, H., Chen, R., Wang, L., Jiang, L., Yang, F., Zheng, S., Wang, G., Lin, X., 2016. Soil pH, total phosphorus, climate and distance are the major factors influencing microbial activity at a regional spatial scale. Sci. Rep. 6 (25815), 1–10.
Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA), version 2.1.4. Geosci. Model Dev. 8, 1991–2007.
Dharumarajan, S., Hegde, R., Singh, S.K., 2017. Spatial prediction of major soil properties using random Forest techniques - a case study in semi-arid tropics of South India. Geoderma Reg. 10, 154–162.
Dobos, E., Carré, F., Hengl, T., Reuter, H.I., Tóth, G., 2006. Digital Soil Mapping as a Support to Production of Functional Maps. EUR 22123 EN, 68 p.. Office for Official Publications of the European Communities, Luxemburg.
Efron, B., Tibshirani, R., 1993. An Introduction to the Bootstrap. Chapman and Hall, London.
ESRI, 2011. ArcGIS Desktop: Release 10.2 Redlands. Environmental Systems Research Institute, CA.
Fang, X., Xue, Z., Li, B., An, S., 2012. Soil organic carbon distribution in relation to land use and its storage in a small watershed of the Loess Plateau, China. Catena 88, 6–13.
Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39, 1347–1359.
Giasson, E., Sarmento, E.C., Weber, E., Flores, C.A., Hasenack, H., 2011. Decision trees for digital soil mapping on subtropical basaltic steep lands. Sci. Agric. 68, 167–174.
Grimm, R., Behrens, T., Marker, M., Elsenbeer, H., 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island: digital soil mapping using random forests analysis. Geoderma 146, 102–113.
Hengl, T., Reuter, H.I., 2008. Geomorphometry: concepts, software, applications. In: Hengl, T., Reuter, H.I. (Eds.), Developments in Soil Science. vol. 33 Elsevier (772 p).
Hengl, T., Heuvelink, G.B., Kempen, B., Leenaars, J.G., Walsh, M.G., Shepherd, K.D., Sila, A., MacMillan, R.A., de Jesus, J.M., Tamene, L., Tondoh, J.E., 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. PLoS One 10, 1–26.
Hijmans, R.J., 2016. raster: geographic data analysis and modeling. In: R Package Version 2.5-8, . https://CRAN.R-project.org/package=raster.
Ivezić, V., Almås, A.R., Singh, B.R., 2012. Predicting the solubility of Cd, Cu, Pb and Zn in uncontaminated Croatian soils under different land uses by applying established regression models. Geoderma 170, 89–95.
Jenny, H., 1941. Factors of Soil Formation. McGraw-Hill, New York.
Jones, H.G., 2014. The Normalized Difference Vegetation Index. Oxford University Press (208 p).
Kuhn, M., Weston, S., Keefer, C., Coulter, N., 2016. C code for Cubist. Cubist: Rule- and instance-based regression modeling. In: R Package Version 0.0.19, . https://CRAN.R-project.org/package=Cubist.
Li, Y., Niu, S.L., Yu, G.R., 2016. Aggravated phosphorus limitation on biomass production under increasing nitrogen loading: a meta-analysis. Glob. Chang. Biol. 22, 934–943.
Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2 (3), 18–22.
Lopez-Granados, F., Jurado-Exposito, M., Atenciano, S., Garcia-Ferrer, A., de la Orden, M.S., Garcia-Torres, L., 2002. Spatial variability of agricultural soil parameters in southern Spain. Plant Soil 246, 94–105.
Ma, Y., Minasny, B., Wu, C., 2017. Mapping key soil properties to support agricultural production in Eastern China. Geoderma Reg. 10, 144–153.
Major, D.J., Baret, F., Guyot, G., 1990. A ratio vegetation index adjusted for soil brightness. Int. J. Remote Sens. 11, 727–740.
Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152.
Malone, B.P., Minasny, B., McBratney, A.B., 2017. Using R for Digital Soil Mapping. Springer (271p).
Malone, B.P., Odgers, N.P., Stockmann, U., Minasny, B., McBratney, A.B., 2018. Digital mapping of soil classes and continuous soil properties. In: McBratney, A.B., Minasny, B., Stockmann, U. (Eds.), Pedometrics. Springer International Publishing, Switzerland, pp. 373–413.
McBratney, A.B., Mendoça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McKenzie, N.J., Gessler, P.E., Ryan, P.J., O'Connell, D., 2000. The role of terrain analysis in soil mapping. In: Wilson, J., Gallant, J. (Eds.), Terrain Analysis: Principles and Applications. John Wiley and Sons, New York, pp. 245–265.
Minasny, B., McBratney, A.B., 2006. A conditioned Latin hypercube method for sampling in the presence of ancillary information. Comput. Geosci. 32, 1378–1388.
Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Digital mapping of soil carbon. Adv. Agron. 118, 1–47.
Mora-Vallejo, A., Claessens, L., Stoorvogel, J., Heuvelink, G.B.M., 2008. Small scale digital soil mapping in Southeastern Kenya. Catena 76, 44–53.
Nobel, R.O., Banuelos, G.S., Paull, J.G., 1997. Boron toxicity. Plant Soil 193, 181–198.
Olsen, S.R., Sommers, L.E., Page, et al., 1982. Phosphorous. In: Methods of Soil Analysis, 2nd ed. Chemical and Microbiological Properties, Agronomy Monograph 9.2 Part 2. pp. 403–430.
Orton, T.G., Pringle, M.J., Bishop, T.F.A., 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262, 174–186.
Pahlavan-Rad, M.R., Toomanian, N., Khormali, F., Brungard, C.W., Komaki, C.B., Bogaert, P., 2014. Updating soil survey maps using random forest and conditioned Latin hypercube sampling in the loess derived soils of northern Iran. Geoderma 232–234, 97–106.
Poggio, L., Gimona, A., 2014. National scale 3D modelling of soil organic carbon stocks with uncertainty propagation — an example from Scotland. Geoderma 232–234, 284–299.
Price, G., 2006. Australian soil fertility manual, Third edition. FIFA, CSIRO Publishing.
Qi, J., Chehbouni, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126.
Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of AI92, 5th Australian Conference on Artificial Intelligence. World Scientific, Singapore, pp. 343–348.
Rudiyanto, Minasny, B., Setiawan, B.I., Saptomo, S.K., McBratney, A.B., 2018. Open digital mapping as a cost-effective method for mapping peat thickness and assessing the carbon stock of tropical peatlands. Geoderma 313, 25–40.
Shahbazi, F., Jafarzadeh, A.A., 2010. Integrated assessment of rural lands for sustainable development using MicroLEIS DSS in West Azerbaijan, Iran. Geoderma 157, 175–184.
Shahbazi, F., Anaya-Romero, M., Braimoh, A.K., de la Rosa, D., 2014. Sustainable land use planning in west Asia using MicroLEIS decision support systems. In: Braimoh, A.K., Huang, H.Q. (Eds.), Vulnerability of Land Systems in Asia. John Wiley and Sons,

New York, pp. 179–194.
Soil Survey Staff, 2014. Keys to Soil Taxonomy, Twelfth edition. United State Department of Agriculture, Natural Resource Conservation Service.
Taghizadeh-Mehrjardi, R., Sarmadian, F., Minasny, B., Triantafilis, J., Omid, M., 2014. Digital mapping of soil classes using decision tree and auxiliary data in the Ardakan region, Iran. Arid Land Res. Manag. 28, 147–168.
Taghizadeh-Mehrjardi, R., Nabiollahi, K., Kerry, R., 2016. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. Geoderma 266, 98–110.
Tang, X., Xia, M., Guan, F., Fan, S., 2016. Spatial distribution of soil nitrogen, phosphorus and potassium stocks in Moso bamboo forests in subtropical China. Forests 7 (267), 1–12.
Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33, 309–319.
Tariq, M., Mott, C.J.B., 2007. The significance of boron in plant nutrition and environment-a review. J. Agron. 6, 1–10.
Taylor, G.R., Mah, A.H., Kruse, F.A., Kierein-Young, K.S., Hewson, R.D., Bennett, B.A., 1996. Characterization of saline soils using airborne radar imagery. Remote Sens. Environ. 57, 127–142.
Viscarra Rossel, R.A., Chen, C., Grundy, M.J., Searle, R., Clifford, D., Campbell, P.H., 2015. The Australian three-dimensional soil grid: Australia's contribution to the globalsoilmap project. Soil Res. 53, 845–864.
Wilson, J.P., Gallant, J.C., 2000. Secondary topographic attributes. In: Wilson, J.P., Gallant, J.C. (Eds.), Terrain Analysis: Principles and Applications. John Wiley & Sons, Inc., New York, pp. 87–131.
Zhao, M.S., Zhang, G.L., Zhao, Y.G., 2013. Variability of soil organic matter and its main factors in Jiangsu Province. Acta Ecol. Sin. 33, 5058–5066.