Geoderma 337 (2019) 1196–1207

Contents lists available at ScienceDirect

Geoderma

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

Retrospective monitoring of the spatial variability of crystalline iron in soils of the east shore of Urmia Lake, Iran using remotely sensed data and digital maps

Farzin Shahbazi[a,*], Alex McBratney[b], Brendan Malone[b,c], Shahin Oustan[a], Budiman Minasny[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
[c] CSIRO Agriculture and Food, Bruce E. Butler Laboratory, PO Box 1700, Canberra, ACT 2601, Australia

ARTICLE INFO

Handling Editor: Michael Vepraskas

Keywords:
Data mining
Digital soil mapping
Modeling
Pedogenesis
Soil ripening

ABSTRACT

The soil ripening process can be defined as change in the soil/sediment matrix from exposure to air in previously inundated areas. Early pedogenesis crystalline iron oxide is one of the most important diagnostic criteria to monitor the aforementioned process. In this study, the east shore of the shrinking Urmia Lake located in the north west of Iran was selected for investigation. This article aims to express the importance of advanced digital soil mapping to link between a difficult-to-measure soil attribute (i.e. crystalline iron oxides) and remotely sensed data as easy-to-measure variables. Additionally, the spatial distribution of clay and soil total iron density was also assessed. In this article, both Landsat 8 and Landsat 7 imagery were acquired for July 2017 and 1999 respectively. Eighteen remotely sensed data (environmental covariates) were employed for analysis, and a bootstrapping method was used to analyze the associated uncertainty of the created maps. Three data mining techniques namely; multiple linear regression, decision tree modeling and Cubist modeling were candidate models for mapping the target variables across the study area. Regarding optimality, a multiple linear regression model was fitted to predict clay content, while a Cubist model was fitted to predict both soil total iron density (SFeD) and crystalline iron oxide (Fe_{crys}). From the summary output, spectral bands in the visible region did not predict SFeD as strongly as that found for Fe_{crys}. As we expected, Fe_{crys} content increases with increasing the distance from the shore. This means that the ripened soils have been established far from the water body.

1. Introduction

During the 1990s, Urmia Lake in northwestern Iran was once the largest salt-water lake in the Middle East. But now its water level is dropping, especially in the eastern regions. The highest and lowest water levels were recorded in 1995 and 2015 respectively. The decline in water level has been measured as eight meters (ULRP, 2015). Nasiri et al. (2015) reported that Urmia Lake has shrunk by up to 56% during the last 15 years. Similar massive decline of water resources has occurred in other salt lakes around the world such as the Aral Sea (Wish-Wilson, 2002), and Great Salt Lake (Mohammed and Tarboton, 2012).

The receding of water from Urmia Lake is greater on the eastern and southern shores where there is relatively flat or gently sloping terrain (Heydari, 2013). Dam construction, excessive use of groundwater, construction of causeways, and unsustainable agricultural development have been the most important agents for the water decline (Ghalibaf and Mousavi, 2014). Moreover, the occurrence and frequency of floods and droughts have also had a substantial impact on the amount of runoff entering Urmia Lake (Razmara et al., 2013). Ultimately, the lake salinity has sharply increased from 160 g/L in high water years, to > 340 g/L (supersaturated salt water) in recent years, resulting in the appearance of salt crystals upon some surfaces (Hesami and Amini, 2016).

Soil shrinkage in wetlands is accompanied by some modifications in its physical and chemical characteristics via redox processes (Pezeshki and Delaune, 2012). The receding of water on the east shore of Urmia Lake naturally exposes to the atmosphere sediments that were once inundated with water. This process initiates the early diagenesis of iron oxides (Haese et al., 1997). Soil iron oxides differ in solubility, reducibility and extractability (Kabata-Pendias, 1993) which tremendously influence the release of precipitated trace metals (Sipos et al., 2014). This research provides a pathway to study the effect of pedogenic processes and type of parent material on iron oxide phases as well as soil chronosequence on the east shore of Urmia Lake in the

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

https://doi.org/10.1016/j.geoderma.2018.11.024
Received 12 March 2018; Received in revised form 10 November 2018; Accepted 12 November 2018

future.

To recognize and monitor soil properties a lot of soil survey needs to be carried out (Ma et al., 2017). Digital soil mapping (DSM) is a key for quantifying and assessing the variation of soil properties in un-sampled and scarcely sampled areas (McBratney et al., 2003). Minasny and McBratney (2016) have reported that soil attributes can be predicted by soil forming factors (s), climate (c), organism (o), relief (r), parent materials (p), age (a) and spatial position (n), with spatially correlated errors (e). The knowledge on the spatial prediction of soil attributes can make a major contribution to the digital soil assessment which was introduced by Carré et al. (2007). Such assessments may focus on different issues such as land contamination and soil biological activity (Khamseh et al., 2017; Shao et al., 2016; Shahbazi et al., 2013).

The incorporation of remote sensing (RS) data (Kim et al., 2012), as well as digital elevation model (DEM) data and derivatives thereof, (Mulder et al., 2011; Boettinger, 2010) have been used with success in DSM studies. Notwithstanding the importance of DEM and derivative data in soil spatial modeling, it was not employed in our research because there were no sharp differences in relief, and climate for that matter, across the study area. Derivatives of RS spectral data have been widely used as environmental covariates in modeling soil-landscape relationships (McBratney et al., 2003). The use of Landsat spectral data has been particularly useful in arid and semi-arid areas (Boettinger et al., 2008). In addition of individual spectral bands, principal component analysis of bands may be important to predict some soil properties (Taghizadeh-Mehrjardi et al., 2016; Malone et al., 2009).

Numerous prediction methods have been developed and introduced to correlate ancillary variables and soil properties through the DSM framework proposed by McBratney et al. (2003). For example, linear regression (LR) modeling has previously been used to predict the distribution of soil organic carbon and compared with Random Forests in Africa (Hengl et al., 2015). In Finland, decision tree modeling (DT) was used for mapping geomorphic surfaces (Luoto and Hjort, 2005). DT models and other non-linear model types such as Cubist models (Ma et al., 2017; Quinlan, 1992) are generally preferred in DSM studies because they generally attain higher accuracies when evaluated against actual observations, in comparison to the linear model counterparts.

Uncertainty analysis is also necessary for establishing how reliable a digital soil map might be (Brown and Heuvelink, 2005). Sun et al. (2012) used available Fe as a soil quality index and estimated the uncertainty of their maps according to the standard deviation of the soil quality index and the derived soil quality information with a confidence level above 0.9. Ma et al. (2017) performed the efficiency of empirical uncertainty quantification through fuzzy clustering and cross validation method. Liddicoat et al. (2015) performed bootstrapping for quantification of uncertainties across very large mapping extents.

The hypothesis of this study is that the mineralogical and chemical changes in iron forms will be a consequence of air penetration into the previously waterlogged materials. Early pedogenesis crystalline iron is an important assessment property to identify the ripened soils. For this purpose, the map of Fe_{crys} was created not only for the present time (i.e. 2017) but also for the past which in this case was 1999. The spatial distribution of clay and soil total iron density (SFeD) was also evaluated using advanced data mining techniques across the study area based on co-migration and/or co-association of iron and clay (Stonehouse and Arnaud, 1971). This study can be considered as a starting point for future investigations in the east shore of Urmia Lake.

2. Materials and methods

2.1. Study area

This study was performed on the east shore of Urmia Lake located in the east Azerbaijan province, Iran (Fig. 1). The average annual precipitation is 341 mm. The minimum and maximum temperature of the region ranges between 0 and −23 °C in winter and up to 39 °C in summer (IRIMO, 2012). The mean annual water level in Urmia Lake is currently more than four meters below the critical level (1274 m above sea level) needed to sustain its own ecosystem. In October 2015, the water level reached the lowest level which caused the southern parts of the lake becoming totally dried (ULRP, 2015).

The study area represents three separate sites which cover approximately 223 km^2, 275 km^2 and 119 km^2 for sites 1–3, respectively. They are delimited by longitude 45°22′18″–46°01′17″ E and latitude 37°14′58″–38°07′12″ N and consist predominantly of swamp and marsh. Three strata were determined at each site by the bathymetry of Urmia Lake: i) 1271–1273 m; ii) 1273–1275 m; and iii) 1275–1278 m above sea level (Fig. 1-D). Iron oxides have accumulated in the surface of the study area particularly in the second stratum of each site (Fig. 1-C). To study changes in soils not affected by inundation, control sites were selected from the external part of stratum 3 at each site where they were not inundated in either 1999 or 2017 (Fig. 1-A and B). For simplicity, the flowchart of the employed procedures is illustrated in Fig. 2.

2.2. Soil sampling and analysis

2.2.1. Routine analysis

A total of 157 topsoil samples (0–10 cm) were collected from the study area according to the stratified random sampling method (McKenzie et al., 2008) during June 2017. Samples (both disturbed and undisturbed) were collected with a grid of 400 ha from the entire of the study area. This method allowed us to shift sampling points where the land surfaces had been covered by salt crusts. With this method, 60, 71 and 26 samples were collected from sites 1–3, respectively. The samples were then transported to the laboratory for preparation and analysis. In the laboratory, disturbed samples were air-dried, then sieved through a 2 mm sieve. The undisturbed samples (clods for clay soils, cylinders for sandy soils) were used for measuring bulk density (BD) in the laboratory (Blake and Hartge, 1986).

Some soil physical and chemical properties were also determined. Particle size analysis (soil texture) was determined using the hydrometer method (Gee and Or, 2002) with clay defined as particles < 0.002 mm, silt (0.002–0.05 mm), and sand (0.05–2 mm). Soil electrical conductivity (EC) was measured using a 1:5 (soil/water ratio) water extract (Rhoades, 1996). Organic carbon (OC) was measured by wet oxidation with chromic acid and back titration with ferrous ammonium sulphate according to the Nelson and Sommers (1996). Carbonate calcium equivalent (CCE) was measured by neutralization-titration method (Allison and Moodie, 1965).

2.2.2. Soil total iron density calculation

Due to the high variability of soils, analysing spatial patterns of soil properties were recently conducted as a stock condition (Ma et al., 2017; Tang et al., 2016). Soil total iron was measured using digestion of samples with aqua regia (McGrath and Cunliffe, 1985). Soil total iron density (SFeD) for a given soil depth (0–10 cm) was calculated by (Eq. (1)).

$$SFeD = Fe_t * BD * h/100 \quad (1)$$

where $SFeD$ is expressed in kg/m^2, Fe_t is the soil total Fe (g/kg), BD is the soil bulk density (g/cm^3) and h is the thickness of soil horizon (cm).

2.2.3. Determination of soil crystalline iron

Iron was determined in both acid ammonium oxalate extracts (Fe_o) and in dithionate-citrate-bicarbonate extracts (Fe_d) of all samples taken from the study area (Loeppert and Inskeep, 1996). Iron concentration in the extracts was measured by atomic absorption spectrometry (Shimatzu, AA-6300). Oxalate extraction dissolves much of the iron from the amorphous materials but very little from the crystalline iron oxides, whereas dithionate extraction dissolves a large proportion of the crystalline iron oxides as well as much of the amorphous materials

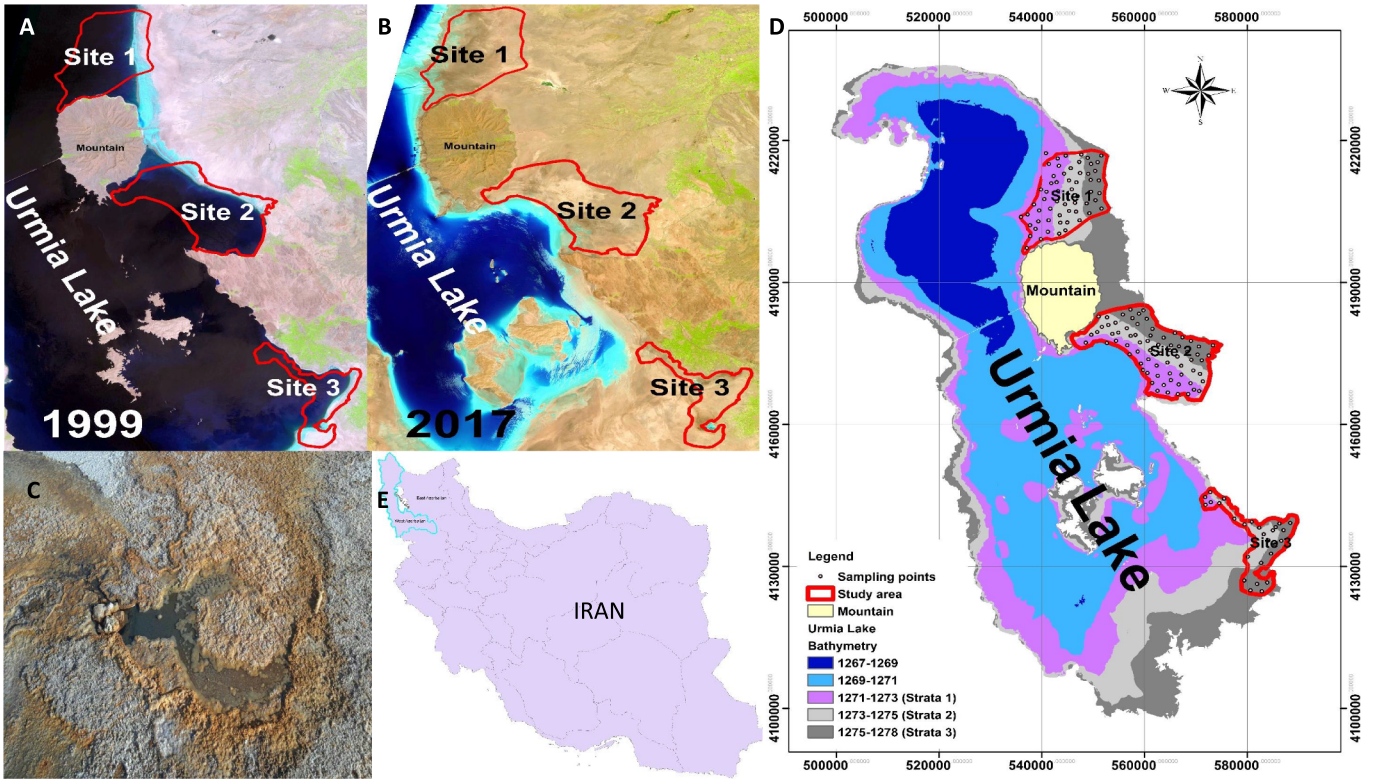

Fig. 1. Location of the study area.
A: Represents the study area in 1999; B: the study area at present scenario (2017); C: a photograph of accumulated iron for the present scenario; D: three strata within three sites on the east shore of Urmia Lake as well as represents the sampling points; and E: location of the study area in country of Iran.

(McKeague and Day, 1966). The Fe_{crys} was therefore calculated by a subtraction of Fe_o from Fe_d.

2.2.4. Statistical analysis

For fitting most statistical models, it is presumed that the data set are normally distributed (Shapiro and Wilk, 1965). Summary statistics were computed prior to modeling to check for normality of the data. A formal way for normality testing includes the Anderson-Darling test statistic. For the data to be normally distributed, the *p*-value calculated by Anderson-Darling should be > 0.05. If not, a way to make the data to be more normal is to transform it (e.g. via log- or square root transformation).

2.3. Environmental covariates

The next task was to gather a suite of environmental covariates thought to be useful (and available) for predicting soil properties within a DSM framework for our study region. Due to harmony in climate and topography, terrain derivatives and climatic data were not used for our research. Instead, from our own prior investigations, we envisaged that soil variation could be captured by investigation of the soil spectral properties. Therefore, we collected a number of remotely sensed imagery scenes from the Landsat 8 ETM^+ as well as Landsat 7 satellite platform acquired in July 2017 and 1999, respectively. We used both the reflectance of the individual bands, together with indices derived from combinations of the individual bands (Table 1). All covariates were transformed to a spatial resolution of 30 m using ArcGIS 10.2 (ESRI, 2011).

For some background, various methods have been studied for representing and mapping soil variability using multispectral remote sensing data. Common approaches include using the individual bands, band ratios, principal components of each of the individual bands, and multiband classification of the bands (Chavez and Kwarteng, 1989). In this study, Landsat imagery derived data were categorized as three orders of different indices: i) vegetation, soil and water; ii) landscape; and iii) geology. Bartholomeus et al. (2007) revealed that vegetation effects on Fe prediction. The presence of vegetation with Normalized Difference Vegetation Index (NDVI) shows a strong sensitivity on the spatial distribution of iron. Soil-Adjusted Vegetation Index (SAVI) also represents better information concerning vegetation and soil characteristics according to Gilabert et al. (2002). Additional to these, Visible Atmospherically Resistant Index (VARI) is the only vegetation index that is significant when mapping with RGB as well as it measures the reflectance of vegetation versus soil (Gitelson et al., 2002). Normalized Difference Moisture Index (NDMI) is sensitive to moisture levels, has been used to monitor droughts (Skakun et al., 2003). Based on this NDMI could potentially be useful for identifying soil formation processes such as those associated with iron redox processes.

With respect to landscape indices, we calculated Normalized Burn Ratio (NBR2) as a new metric for quantifying burn severity (Parks et al., 2014). The low value indicates bare ground and recently burned areas. This index was used in this article because there is no or little vegetation in the salty areas.

In terms of soil indices, Ducart et al. (2016) used Clay index (CI) in Brazil and reported that it discriminates clays in areas related to mafic regolith. Clay index (Boettinger et al., 2008), salinity ratio (SR) (Metternicht and Zinck, 2003), ferrous mineral (FMI) and iron oxide (IOI) indices (Drury, 2016) were four calculated derivative Landsat imagery indices that were used in our research to represent the soil geochemical variability across the study area. Four examples of ancillary data derived from Landsat spectral data across the study area are shown in Fig. 3.

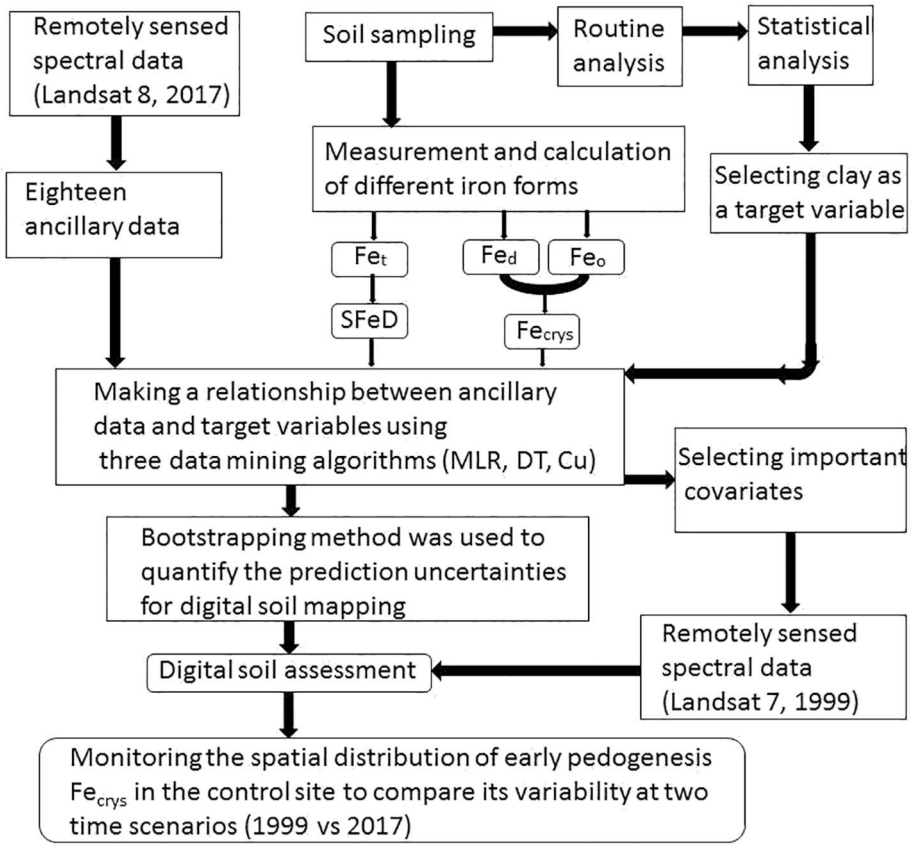

Fig. 2. Simplified flowchart of the research across in this study.
Fe_t: Total iron (g/kg); SFeD: soil total iron density (kg/m^2); Fe_d: dithionate-extractable iron (g/kg); Fe_o: oxalate-extractable iron (g/kg); Fe_{crys}: crystalline iron (g/kg); MLR: multiple linear regression; DT: decision trees; Cu: Cubist model.

2.4. Soil attribute modeling

2.4.1. Multiple linear regression

Multiple linear regression (MLR) is a common model for soil spatial predictions. Demattê et al. (2007) used this model for evaluating the relationship between analytical and spectral parameters of six important classes of tropical Brazilian soils. They reported the importance of spectral data obtained by Interface Region Imaging Spectrograph (IRIS) for the prediction of clay content at both local and global scales. Forkuor et al. (2017) recently used MLR for mapping of soil properties in south-western Burkina Faso. Jones (1973) had previously reported the efficiency of MLR to predict the soil clay content in West Africa. In this study, we first put all covariates (18 ancillary data) in the model and then performed a stepwise regression to determine statistically significant variables.

2.4.2. Decision trees

Decision Trees (DT) has been widely used for DSM (Taghizadeh-Mehrjardi et al., 2014). In this model, several independent variables correlated with direct or indirect relationships to a target variable are defined with a tree structure, which is generated by partitioning the data recursively into a number of groups (Breiman et al., 1984). Nodes,

Table 1
Ancillary data variables considered in this study.

Ancillary data	Description	Definition[b]
Individual bands	Blue, Green, Red, NIR, SWIR1, SWIR2	B2–B7
PCA of bands	Principal component analysis of six individual bands	PCA (B2–B7)
False colour composite	Combination of B4, B5 and B6	B456
	Combination of B4, B6 and B7	B467
Vegetation, soil and water indices	Normalized Difference Vegetation Index (NDVI)	(NIR − Red) / (NIR + Red)
	Soil-Adjusted Vegetation Index (SAVI)	((NIR − Red) / (NIR + Red + L[a])) × (1 + L)
	Visible Atmospherically Resistant Index (VARI)	(Green − Red) / (Green + Red − Blue)
	Normalized Difference Moisture Index (NDMI)	(NIR − SWIR1) / (NIR + SWIR1)
Landscape index	Normalized Burn Ratio 2 (NBR2)	(SWIR1 − SWIR2) / (SWIR1 + SWIR2)
Geology indices	Clay Index (CI)	(SWIR1 / SWIR2)
	Salinity Ratio (SR)	(Red − NIR) / (Red + NIR)
	Ferrous Minerals Index (FMI)	SWIR1 / NIR
	Iron Oxide Index (IOI)	Red / Blue

NIR: near infrared; SWIR1: shortwave infrared 1; SWIR2: shortwave infrared 2.
[a] The L value varies depending on the amount of green vegetative cover as well as in areas with moderate green vegetative cover, L = 0.5.
[b] Assign between Landsat 8 and 7 different band numbers.

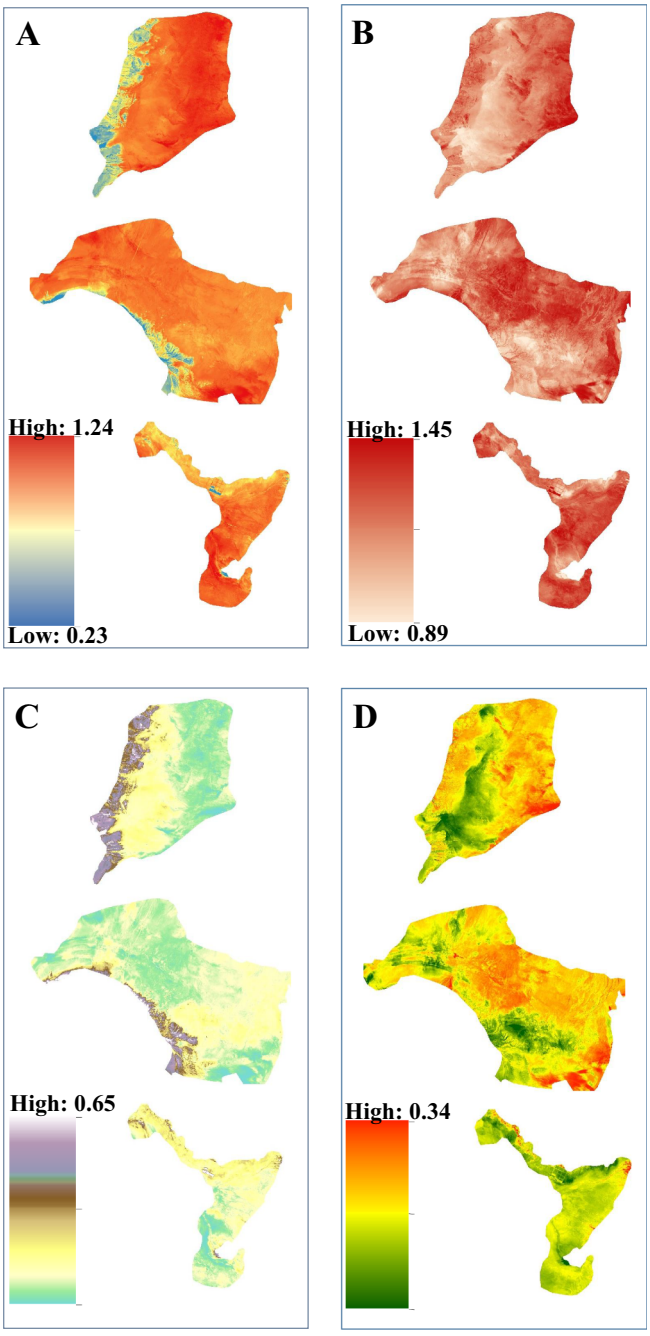

Fig. 3. Four examples of applied Landsat spectral data derived to model across the study area.
A: Ferrous minerals index (FMI); B: iron oxide index (IOI); C: normalized difference moisture index (NDMI); D: normalized difference vegetation index (NDVI).

branches and leaves are the components of the constructed decision tree. Each branch of the tree ends in a terminal node which it is uniquely defined by a set of rules. A criticism often raised about DTs though is that the terminal node itself is a single value, which could culminate in the creation of a class-like map of numerical data in situations where there are few terminal nodes. To facilitate DT model fitting in R we used "*rpart*" package (Therneau et al., 2017).

2.4.3. Cubist model

An advanced form of the DT model is the Cubist model (Quinlan, 1992). The efficiency of this model in DSM has been reported earlier (e.g. Malone et al., 2017; Kidd et al., 2015). Cubist models are presented as a series of rules, each starting with a conditional 'if' statement that subsets the data based on criteria of the variables used in the model. It provides information about the conditions for each rule, the regression model for each rule, and information about the diagnostics of the model fit, plus the frequency of which the covariates were used as conditions and/or within a model. In this study, we arbitrarily set calibration parameters of the Cubist model to: 5 rules, 10% of data extrapolations and 5 committees, which suggested that 5 boosting iterations were supplied to predict and calculate the contribution of the covariates. To facilitate Cubist model fitting in R we used the "*cubist*" package (Kuhn et al., 2016) to characterize the relationship between target variables and covariates.

2.4.4. Model validation

Each fitted model was assessed using four criteria: i) the root mean square error (RMSE); ii) coefficient of determination (R^2); iii) bias; and iv) Lin's concordance correlation coefficient (Concordance) (Lin, 1989). R^2 measures the precision of the relationship between the observations and their corresponding predictions. Bias, also called the mean error of prediction, indicates if the model under or over predicts. Concordance evaluates both the accuracy and precision of the relationship, often referred to as the goodness of fit along a 1:1 line (Eqs. (2)–(5)).

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

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

$$bias = \frac{\sum_{i=1}^{n}(obs_i - pred_i)}{n} \quad (4)$$

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.

$$\rho_c = \frac{2\rho\sigma_{pred}\sigma_{obs}}{\sigma_{pred}^2 + \sigma_{obs}^2 + (\mu_{pred} - \mu_{obs})^2} \quad (5)$$

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 be used for mapping clay, SFeD and Fe_{crys} across the extent of the study area.

2.5. Maps and their associated uncertainties

Minasny and McBratney (2002) reported that uncertainty analysis is essential for pedotransfer functions. There are some approaches for quantifying the empirical prediction uncertainties: i) universal kriging prediction variance; ii) Bootstrapping; iii) Empirical uncertainty quantification through data partitioning and cross validation; and iv) Empirical uncertainty quantification through fuzzy clustering and cross validation (Malone et al., 2017). In this study, we used a bootstrapping approach to calculate the mean of the prediction realisations from each bootstrap sample. For running the bootstrap, we selected arbitrarily 200 iterations. For each iteration, 100% of the sample data size was 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 an out-of-bag sample set. For each iteration, out-of-bag goodness of fit diagnostics measures (as described above) was evaluated. The variances of the predictions were then estimated after stacking all created maps in the modeling process. Prediction intervals (PI) which

refer to the interval between the 90% upper and lower prediction limits, were calculated. Through this, we generated variance information at every node of our prediction grid of 30-m resolution, which was used to calculate PI. This range was used as criteria to assess the magnitude of uncertainty across the study area.

2.6. Digital soil assessment

Very simply, digital soil assessment (DSA) can be likened to the quantitative modeling of difficult-to-measure soil attributes. McBratney et al. (2012) defined that DSA is the translation of DSM outputs into decision making aids that are framed by the particular, contextual human-value system which addresses the question/s at hand. In this study, for the present scenario, we first created the spatial maps of clay and SFeD using the best fitted models across the study area according to our speculation that there is co-migration or co-association of clay and iron (Stonehouse and Arnaud, 1971). The digital map of Fe_{crys} was also created. Based on these inputs we then assessed the impact of water level decline on the soils on the eastern shore of Urmia Lake.

The next issue concerning to DSA involved assessing the changes in soil not impacted by water inundation with the study area after two decades. Such areas are located at the external part of the third strata in each site i.e. areas furthest from the lake shoreline, where waterlogging was not evident in either 1999 or 2017 (see Fig. 1-A and B together). In total, 77.4 km^2 was selected from sites 1–3 (areas about 33, 27.6 and 16.8 km^2, respectively). For mapping the spatial distribution of Fe_{crys} across the selected control sites for 1999, parameters from the fitted 2017 model were applied to the associated RS covariates that were acquired for 1999. For this, we assigned between Landsat 8 and Landsat 7 different bund numbers. While there may be assumptions required around the stationarity of the model parameters through time with this approach, it provides an efficient means to utilise time series RS data to assess likely changes in soil through time in situations where sampling could not be achieved during both time periods. Such an approach has been applied by Waring et al. (2014) for assessing changes in soil carbon over time with a region of Australia. The changes of Fe_{crys} as well as the occurrence of the soil ripening process for both time scenarios were monitored using the created digital maps of control sites. Moreover, comparing the mean values of predicted Fe_{crys} maps will lead to understanding the link between soil chemistry, pedogenesis and soil surveying.

3. Results and discussions

3.1. Observed soil data

Descriptive statistics of measured routine analysis and iron dependant properties are summarized in Table 2. Routine soil analysis revealed that the collected soils were high in electrical conductivity (30 dS/m, on average), varied from coarse to fine texture, were high in organic carbon content (2.37 g/100 g, on average), and medium in total iron content (1.62%, on average). According to the results, the highest variation between maximum and minimum values was recorded for Fe_{crys} (114-fold) followed by Fe_t (90-fold) and SFeD (85-fold), respectively. This variation was also high for sand, silt and clay. Notwithstanding sand followed by Fe_{crys} had the highest coefficient of variation (CV) between soil properties which indicated their strong variation in the study area, but all variables could be categorized as moderate variability (CV < 10% = low variability, 10% < CV < 90% = moderate variability, and CV > 90% = extreme variability) (Fang et al., 2012). There was a positive skewness with coefficients varying between 0.03 and 1.97 for all properties. The observations of Anderson-Darling *p*-value revealed that all variables except clay data deviated from normal. Subsequently, the data of SFeD and Fe_{crys} were normalized by square-root transformation.

3.2. Selection of parsimonious model

Table 3 shows the performances of MLR, DT and Cubist models on calibration and validation data set to predict clay, SFeD and Fe_{crys} across the study area. SFeD and Fe_{crys} are the sqrt-transformed results of the original data set. The results revealed that the DT model was not suitable for our data and environmental covariates in our study area for all target variables.

In terms of clay, all predictions were generally quite good with calibration using DT (R^2 = 0.48) followed by MLR (R^2 = 0.31) and Cubist model (R^2 = 0.26). The performance of R^2 in validation dataset decreased in MLR (R^2 = 0.12) followed by Cubist (R^2 = 0.11) and DT (R^2 = 0.08). On the other hand, the highest concordance value in validation data set for clay prediction was observed in implementation of MLR. Generally, MLR was selected as the best fitted model to predict the spatial distribution of clay across that study area.

Also, the obtained results showed that the performance of R^2 with calibration in terms of SFeD in DT (0.54), followed by Cubist (0.29) and MLR (0.26), while those values decreased in validation. We obtained the highest difference of R^2 between calibration and validation data set in DT (72%), followed by MLR (49%) and Cubist model (35%). The Cubist model represents the highest value of concordance in validation data set too. A similar trend was also observed in terms of Fe_{crys}. Based on these results, the Cubist model was selected in this research for digitally mapping of both SFeD and Fe_{crys} across the study area.

3.3. The importance of covariates in models

3.3.1. Multiple linear regression model

A stepwise MLR was used to model the relationships between clay

Table 2
Summary of descriptive statistics of measured soil properties in this study (n = 157).

	Min	Max	Mean	SD	CV	Skewness	Kurtosis	AD	p-Value
EC (dS/m)	0.61	122	30.06	15.86	52.77	1.97	8.23	3.21	4×10^{-8}
OC (g/100 g)	0.83	5.22	2.37	1.41	59.21	0.63	−1.03	8.36	2×10^{-16}
CCE (g/100 g)	11.21	49.81	32.81	12.17	37.08	0.12	−1.36	5.71	4×10^{-14}
Sand (%)	0.05	89.78	35.36	26.34	74.49	0.38	−1.15	3.51	8×10^{-9}
Silt (%)	0.01	61.94	26.76	17.36	65.38	0.03	−1.18	2.09	2×10^{-5}
Clay (%)	0.01	79.54	38.09	14.41	37.82	0.23	−0.19	0.36	0.45
Fe_t (g/kg)	0.61	54.94	16.17	8.97	55.49	0.62	1.51	0.77	0.04
SFeD (kg/m^2)	0.11	9.39	2.33	1.32	56.69	1.09	4.34	0.89	0.03
Fe_d (g/kg)	0.19	12.09	2.98	2.04	68.43	1.18	2.38	2.63	2×10^{-3}
Fe_o (g/kg)	0.02	1.91	0.41	0.24	59.23	1.62	8.53	1.11	0.01
Fe_{crys} (g/kg)	0.09	10.24	2.58	1.93	74.93	1.24	1.92	3.96	7×10^{-9}

EC: Electrical conductivity; OC: organic carbon; CCE: carbonate calcium equivalent; Fe_t: total iron; SFeD: soil total iron density; Fe_d: dithionate-extractable iron; Fe_o: oxalate-extractable iron; Fe_{crys}: crystalline iron; SD: standard deviation; CV: coefficient of variation; AD: Anderson-Darling amount; p-value: calculated by Anderson-Darling for normality testing.

Table 3
The statistical criteria for evaluating the performance of predictions in calibration and validation dataset across the study area.

	R^2			Concordance			RMSE[a]			Bias		
	MLR	DT	Cu	MLR	DT	Cu	MLR	DT	Cu	MLR	DT	Cu
Calibration dataset (in the bag)												
Clay	0.31	0.48	0.26	0.48	0.65	0.39	0.09	0.09	0.09	−0.001	0.001	−0.76
SFeD	0.51	0.54	0.45	0.67	0.71	0.59	0.001	0.003	0.002	0.001	0.001	0.012
Fe_{crys}	0.44	0.57	0.39	0.61	0.72	0.54	0.003	0.01	0.003	0.001	0.001	−0.05
Validation dataset (out of bag)												
Clay	0.12	0.08	0.11	0.27	0.25	0.23	0.09	0.09	0.09	−0.22	−0.02	−0.65
SFeD	0.26	0.15	0.29	0.46	0.36	0.47	0.001	0.003	0.002	0.01	−0.01	0.012
Fe_{crys}	0.27	0.21	0.29	0.47	0.42	0.45	0.003	0.01	0.003	0.01	0.006	−0.05

[a] Normalized RMSE (root mean square error); SFeD: soil iron density (sqrt (kg/m^2)); Fe_{crys}: crystalline iron (sqrt (g/kg)); MLR: multiple linear regression; DT: decision trees; Cu: Cubist model.

and remotely sensed data (see Table 1) as quantitative predictors. The results revealed that only nine covariates including B2, B5, B6, B7, CI, FMI, NDMI, PCA and SAVI were significantly predictors of clay across the study area. Band 2 and NDMI ($P < 0.00$) were the most important covariates, followed by B6, B7, CI and PCA ($P < 0.001$) while B5, FMI and SAVI ($P < 0.01$) were the last ones.

3.3.2. Cubist model

The output generated from fitting the Cubist model revealed the existence of five conditional rulesets, each with an associated model, to predict SFeD and Fe_{crys} (Tables 4 and 5). The results showed that the first ruleset has the least amount of standard error in predictions.

Table 6 summarizes the contribution of all environmental covariates in the Cubist model as the best selected one for predicting SFeD and Fe_{crys} across the study area. In terms of SFeD, nine covariates played an important role. IOI and VARI appeared the most frequently used variables, where they were included in 60% of ruleset models. CI on the other hand was less frequently used with inclusion in 12% of ruleset models. Usage of the other variables were between these two extremes i.e. 12% and 60%. The results indicated that visible bands (B2, B3, B4), composite (B456, B467) and also NIR band (B5) were not identified as predictors in terms of SFeD in this research. While B2 (corresponding to the visible blue band) was identified as the common model predictor (included in 60% of ruleset models) in terms of Fe_{crys} across the study area. This outcome could be interpreted as meaning that Fe_{crys} is more sensitive to the visible bands compared to the SFeD.

3.4. Digital maps and soil assessment

The spatial distribution of clay, SFeD and Fe_{crys} and also their associated PI as an indicator of quantitative uncertainty analysis were then assessed for the three sites using selected models in this research. The maps of SFeD and Fe_{crys} were back-transformed to their original scale for mapping and interpretation. The impact of water level decline on target variables was expressed by dividing each site into existent

Table 4
Constructed rulesets of the fitted Cubist model to predict SFeD (sqrt (kg/m^2)) across the study area.

Ruleset	Model	Se
1	1/1: if CI > 1.12 then: SFeD = 17.22–22.83 CI + 47.3 NBR2 − 0.01 B7 + 7.52 IOI + 9.8e−5 B6 + 5.3 VARI	0.29
	1/2: If CI < 1.12 then: SFeD = −6.07–0.01 B7 + 9.79 IOI + 0.01 B6–14.4 NBR2 + 9.4 VARI	0.28
2	SFeD = 0.87 + 3.5e−5 B7	0.35
3	SFeD = −11.46 − 19.23 NDMI + 31.44 IOI − 12.63 FMI + 33 VARI − 0.01 PCA	0.36
4	SFeD = 0.71 + 3.9e−5 B6	0.36
5	SFeD = −10.46 − 20.33 NBR + 31.39 IOI − 13.39 FMI + 33 VARI − 0.01 PCA	0.38

Se: Standard error; B6: shortwave infrared 1; B7: shortwave infrared 2; CI: clay index; NBR2: normalized burn ratio 2; IOI: iron oxide index; VARI: visible atmospherically resistant index; NDMI: normalized difference moisture index; FMI: ferrous minerals index; PCA: principal component analysis of bands.

Table 5
Constructed rulesets of the fitted Cubist model to predict Fe_{crys} (sqrt (g/kg)) across the study area.

Ruleset	Model	Se
1	Fe_{crys} = 3.51–0.01 B2 + 0.01 B3	0.37
2	Fe_{crys} = 12.71–15.19 NDMI - 10.13 FMI	0.41
3	Fe_{crys} = 3.64–0.01 B2 + 0.01 B4	0.39
4	4/1: if NDMI > 0.004 then: Fe_{crys} = 1.51–2.04 NDMI	0.46
	4/2: if NDMI < 0.004 then: Fe_{crys} = 7.01–62.8 SAVI	0.61
5	Fe_{crys} = 5.82 + 0.01 PCA − 0.01 B7–0.01 B6–13.63 NDMI − 0.01 B2	0.47

Se: Standard error; B2: blue; B3: green; B4: red; B6: shortwave infrared 1; NDMI: normalized difference moisture index; FMI: ferrous minerals index; SAVI: soil-adjusted vegetation index; PCA: principal component analysis of bands.

Table 6
Contributions of the environmental covariates predictors in the Cubist model.

Covariates	SFeD (sqrt (kg/m^2))		Fe_{crys} (sqrt (g/kg))	
	Conds	Model	Conds	Model
B2				60%
B3				20%
B4				20%
B5				
B6		40%		20%
B7		40%		20%
PCA		40%		20%
B456				
B467				
NDVI				
SAVI				4%
VARI		60%		
NDMI		40%	20%	56%
NBR2		20%		
CI	20%	12%		
SR				
FMI		40%		20%
IOI		60%		

Conds: conditions; B2: blue; B3: green; B4: red; B5: near infrared; B6: shortwave infrared 1; B7: shortwave infrared 2; PCA: principal component analysis of six individual bands; B456: combination of B4, B5 and B6; B467: combination of B4, B6 and B7; NDVI: normalized difference vegetation index; SAVI: soil-adjusted vegetation index; VARI: visible atmospherically resistant index; NDMI: normalized difference moisture index; NBR2: normalized burn ratio 2; CI: clay index; SR: salinity ratio; FMI: ferrous minerals index; IOI: iron oxide index.

Table 7
Calculated mean values of target variables and interval ranges across the study area using the best fitted models.

	Site 1			Site 2			Site 3		
	Stratum 1	Stratum 2	Stratum 3	Stratum 1	Stratum 2	Stratum 3	Stratum 1	Stratum 2	Stratum 3
Area (km^2)	88.1	76.9	58.3	94	90	91.2	7.3	44.2	67.8
Clay (%)	30.76	32.41	39.91	38.37	48.21	48.92	39.56	40.28	41.52
PI (clay)	17.82	9.56	9.03	12.92	9.82	9.61	14.19	13.24	12.01
SFeD	1.16	1.83	2.39	2.21	2.58	2.83	2.55	2.83	2.91
PI (SFeD)	0.44	0.22	0.19	0.31	0.24	0.23	0.28	0.31	0.33
Fe_{crys}	0.93	1.72	2.21	2.12	2.88	3.13	2.21	2.73	2.77
PI (Fe_{crys})	0.44	0.33	0.25	0.37	0.36	0.33	0.32	0.32	0.32

PI (clay): mean prediction interval range for prediction of clay in percent; SFeD: mean prediction of soil total iron density (kg/m^2); PI (SFeD): mean prediction interval range for prediction of soil total iron density (kg/m^2); Fe_{crys}: mean prediction of crystalline iron (g/kg); and PI (Fe_{crys}): mean prediction interval range for prediction of crystalline iron (g/kg).

stratum. Maps were created for all sites in this research, but we show only those from site 1 in this manuscript. The maps created for sites 2 and 3 are given in the Supplementary material. Table 7 represents overall findings partitioned by zoning of each stratum within each site. Detailed explanations in terms of target variables are fully addressed in the next sections.

3.4.1. Clay

Fig. 4 shows the spatial distribution of clay that resulted from using the stepwise MLR model. Clay varies largely across the study area from 0% to 100%. Raizada et al. (2003) have reported that the soils of the waterlogged area have higher concentrations of clay-loam. The results showed that on average the second site has the highest clay content (43.5%), followed by the third site (40.45%) and the first one (34.36%). In terms of the first site, the highest clay on average was observed in the stratum 3 (39.91%), followed by stratum 2 (32.41%) and then stratum 1 (30.76%) respectively. The prediction interval which was calculated by the difference between upper and lower 90% limit of prediction bounds to define the level of confidence (Xiong et al., 2015) showed that in the direction from the first-to-third strata PI on average

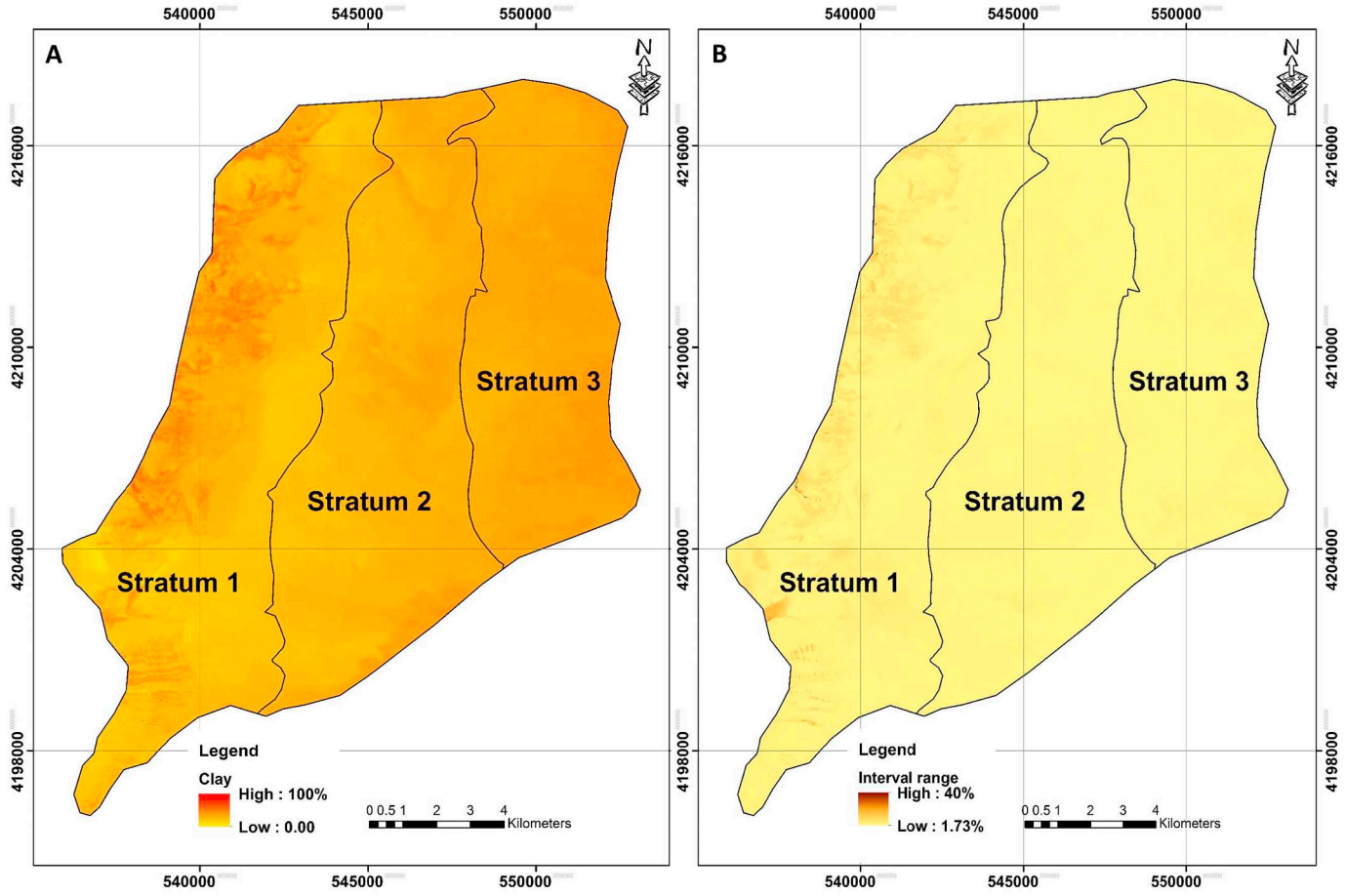

Fig. 4. The spatial distribution of clay content (%) across the first site within the study area using stepwise MLR model.
A: Mean prediction (%); B: prediction of interval range (%).

decreases. The similar trend was also found for the other two sites in terms of both mean clay content and mean PI (see Table 7). It means that the accuracy and precision of prediction of clay content was boosted with increments in distance from the shore. It is speculated that the variation observed may be related to soil formation processes such as soil ripening. Also, the nature of clay particles, which may differ over the strata, determines soil properties and their behaviours and functions largely governs some geomorphic processes such as weathering, erosion and deposition (Viscarra Rossel, 2011).

3.4.2. Soil total iron density

According to the summary outputs (see Table 7) using Cubist model, the highest content of SFeD belongs to the third site (2.76 kg/m^2, on average), followed by the second site (2.54 kg/m^2, on average) and then the first one (1.79 kg/m^2, on average). Detailed observation via separation of each stratum within the sites also revealed that the increment in distance from the shore to the external part of the study area in terms of sites 1 and 2 has positive correlation with SFeD and negative correlation with PI. In terms of site 3, we could not find the same trend. A possible reason is that there is very little geographic separation between the strata within site 3, especially on the northern part. Although, the highest and lowest PI were calculated for the first (0.35 kg/m^2, on average) and the third strata (0.25 kg/m^2, on average). Generally, the distribution pattern of SFeD is approximately similar to the distribution of clay across that study area. It means that spatial variability of clay corresponds to the distribution of SFeD, as we expected. Moskovchenko et al. (2017) reported that metal concentrations are positively correlated with the contents of clay fractions. Fig. 5 shows the digital map of predicted SFeD and its associated quantified uncertainty analysis in the first site.

3.4.3. Crystalline iron

Fig. 6 illustrates the spatial distribution of Fe_{crys} for the first site of the study area using Cubist model. As reported by Table 7, the highest value of Fe_{crys} relevant to the site 2 (2.71 g/kg), followed by site 3 (2.57 g/kg) and then site 1 (1.62 g/kg). Generally, it shows that the soils on the second site are the most ripened ones due to high in Fe_{crys}. The results also revealed that stratum 3, within all sites, has the highest Fe_{crys} as we expected to observe because these soils have been exposed to the atmosphere the longest and consequently, there has been sufficient time to form crystalline iron. Iron oxide formation can also be established from the youngest to the oldest soils (Moody and Graham, 1995).

Quantification of prediction uncertainties for the created digital maps using bootstrapping method showed that the extracted outputs from stratum 3 and 1 has high and low in accuracy and precision, respectively. This may be due to the presence of salt crystals in scattered and sometimes mounded deposits on the land surfaces at the first stratum or due to the occurrences of sediments close to the lake which are difficult to capture within the modeling process with sufficient data in terms of sampling and the quality of predictive variables used.

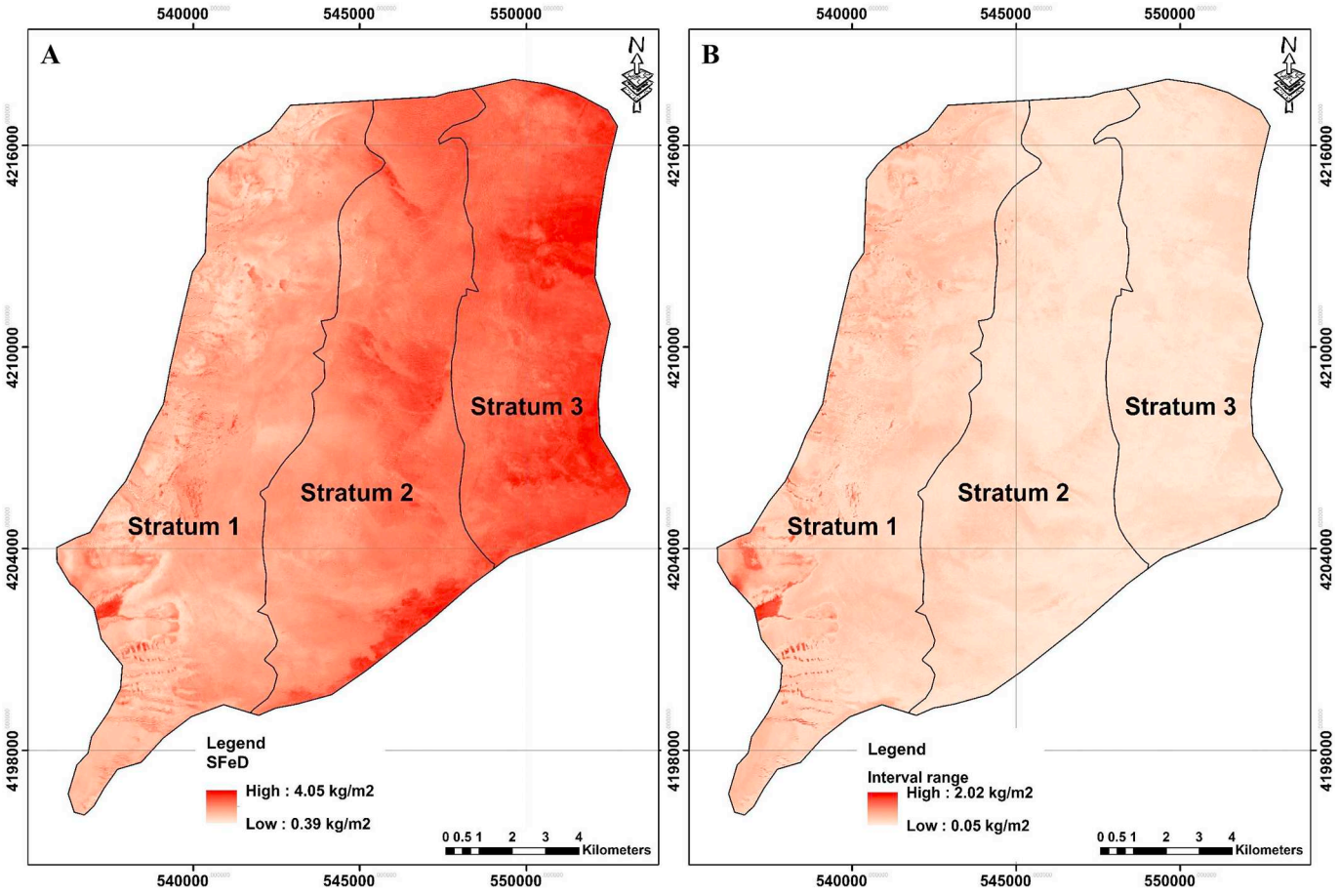

Fig. 5. The spatial distribution of SFeD (kg/m^2) across the first site within the study area using Cubist model.
A: Mean prediction (kg/m^2); B: prediction of interval range (kg/m^2).

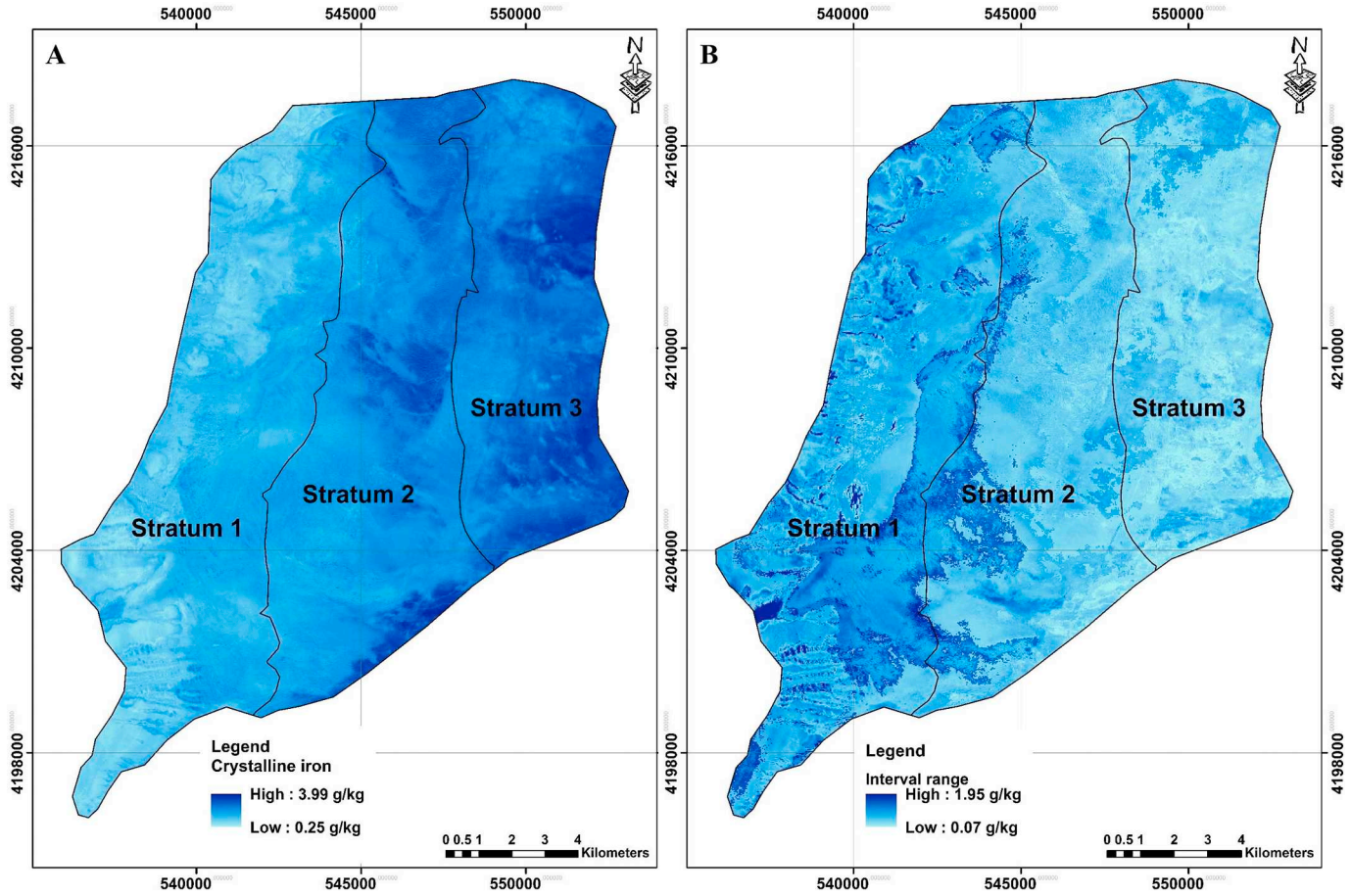

Fig. 6. The spatial distribution of Fe_{crys} (g/kg) across the first site within the study area using Cubist model.
A: Mean prediction (g/kg); B: prediction of interval range (g/kg).

3.5. Monitoring the changes of crystalline iron

Fig. 7 represents the changes of Fe_{crys} for a control site which was extracted from the first site via created maps and also their associated PI to compare two scenarios. The results showed that mean value of Fe_{crys} decreased over the two decades contrary to our expectation, if we consider only soil formation without paying attention to soil leaching. This value was recorded to be 2.75 g/kg for 2017, while it was recorded to be 3.81 g/kg for 1999. It means that despite the soil formation process of ripening that act upon the control sites, Fe_{crys} has decreased. The possible reason for this is could be that the area is located on the basin of Urmia Lake as well as has been established on the direction of Ajichay River, which iron may be leached. Iron depletion was mostly observed in the northern part of the first site. Coating of iron oxides by other constituents such as carbonates can also justify the event (Tack and Verloo, 1997). Moreover, crystallization of amorphous iron oxides over the time reduces their surface area for light reflection. Meanwhile, the results of the prediction mean interval range was shown that created maps of 1999 have lower confidences than created ones for 2017 as we expected. It occurred because a vast selected control sites in 1999 was nearby to the Urmia Lake at that date. This observation confirms our findings from 2017 as well.

3.6. Future works

This research provided a pathway to start further works in the future such as micromorphological observations and vis-NIR spectroscopy to find more details on different forms of iron oxides. According to Czech et al. (2012) who have reported that the highest iron content were determined in the area with highest total content of zinc and lead as well as because of the relationships between iron and trace metal fraction in soil (Sipos et al., 2014), it was also advised to study the condition of trace metal association with soil iron phases.

4. Conclusions

This study sorts to understand pedogenesis in the east shore of Urmia Lake, Iran. The results presented above demonstrate that the spatial distribution of Fe_{crys} as a difficult-to-measure property either for the current or the previous periods using environmental covariates were successfully mapped.

The results revealed that the DT model was not suitable for all target variables. MLR was identified as the best model to predict the spatial distribution of clay. While, the Cubist model was selected as the best one in terms of SFeD and Fe_{crys} across the study area.

Remote sensing imagery, particularly those encompassing the visible region played an important role in the prediction of Fe_{crys}, while they were not identified to model the distribution of SFeD very well. The obtained results were also revealed that the Visible Atmospherically Resistant Index (VARI) was important to predict SFeD, while it was not used in terms of Fe_{crys} across the study area.

The results showed that Fe_{crys} increased with increment in distance from the shore at both timestamps of 1999 and 2017 which corresponds to the ripening process. Comparing both time scenarios showed that Fe_{crys} had decreased during the past two decades, to reveal the possibility of leaching due to iron mobility.

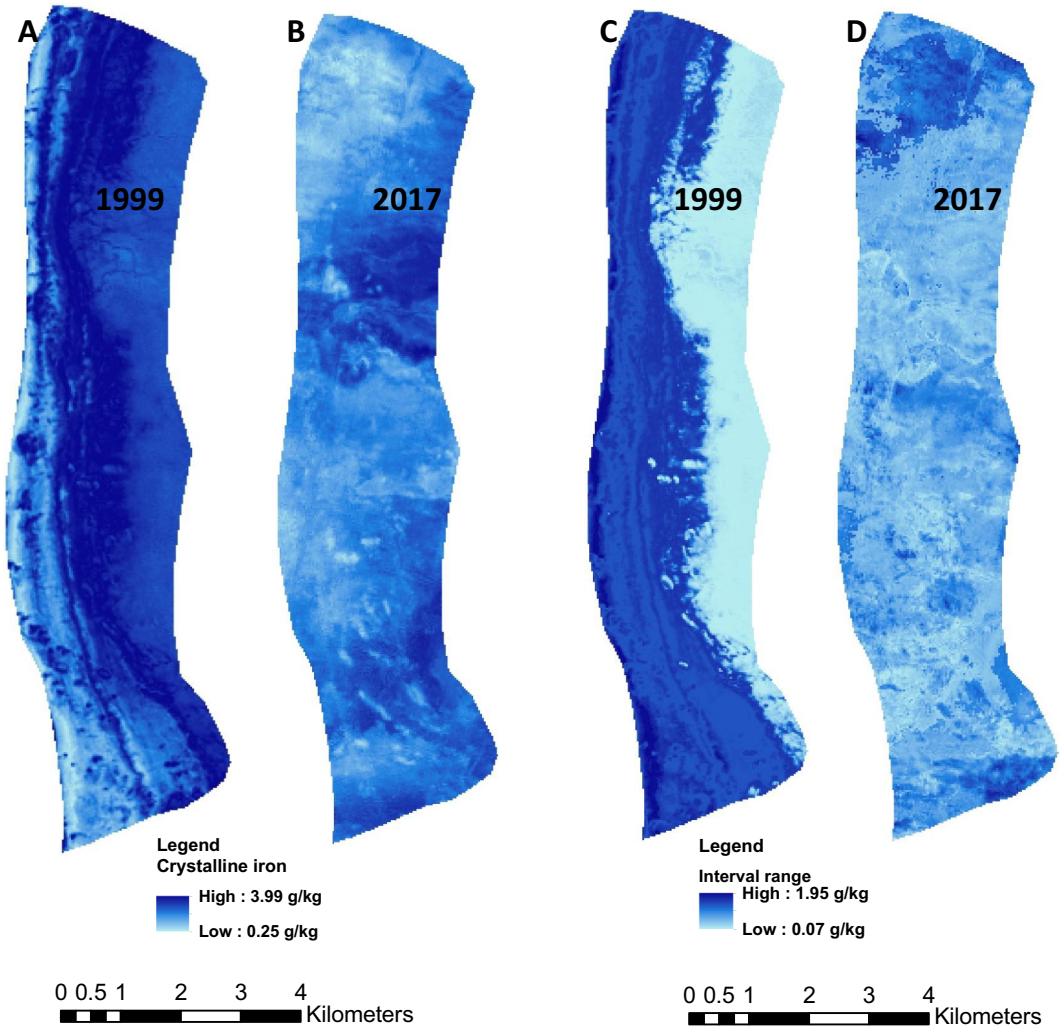

Fig. 7. Comparing predicted mean value of Fe_{crys} and its associated interval range for two scenarios of 1999 and 2017.
A and B: Mean prediction (g/kg) for 1999 and 2017 respectively; C and D: prediction of interval range (g/kg) for 1999 and 2017 respectively.

Acknowledgements

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 while based at the University of Sydney, Australia.

Appendix A. Supplementary data

Supplementary data have been associated with this article at https://doi.org/10.1016/j.geoderma.2018.11.024. These data include the Google Map of the sampling points and created digitally maps of clay, SFeD and Fe_{crys} for both sites 2 and 3.

References

Allison, L.E., Moodie, C.D., 1965. Carbonate. In: Black, C.A. (Ed.), Methods of Soil Analysis. Part 2, 2nd ed. Agron. Monogr., vol. 9. ASA, CSSA, and SSSA, Madison, WI, pp. 1379–1400.
Bartholomeus, H., Epema, G., Schaepman, M., 2007. Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. Int. J. Appl. Earth Obs. Geoinf. 9, 194–203.
Blake, G.R., Hartge, K.H., 1986. Bulk density. In: Klute, A. (Ed.), Methods of Soil Analysis, Part 1: Physical and Mineralogical Methods, 2nd ed. Agron. Monogr., vol. 9. ASA, CSSA, and SSSA, Madison, WI, pp. 363–375.
Boettinger, J.L., 2010. Environmental covariates for digital soil mapping in the western USA. In: Boettinger, J.L., Howell, D.W., Moore, A.C., Hartemink, A.E., Kienast-Brown, S. (Eds.), Digital Soil Mapping. Progress in Soil Science. vol. 2. Springer, Dordrecht, pp. 17–27.
Boettinger, J.L., Ramsey, R.D., Bodily, J.M., Cole, N.J., Kienast-Brown, S., Nield, S.J., Saunders, A.M., Stum, A.K., 2008. Landsat spectral data for digital soil mapping. In: Harteming, A.E., McBratney, A.B., Mendonca-Santos, M. (Eds.), Digital Soil Mapping with Limited Data. Springer, Dordrecht, pp. 193–203.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Chapman & Hall, New York.
Brown, J.D., Heuvelink, G.B.M., 2005. Assessing uncertainty propagation through physically based models of soil water flow solute transport. Encycl. Hydrol. Sci. 6, 79.
Carré, F., McBratney, A.B., Mayr, T., Montanarella, L., 2007. Digital soil assessments: beyond DSM. Geoderma 142, 69–79.
Chavez, P.S., Kwarteng, A.Y., 1989. Extracting spectral contrast in Landsat thematic mapper image using selective principal component analysis. Photogramm. Eng. Remote. Sens. 55, 339–348.
Czech, T., Gambuœ, F., Wieczorek, J., 2012. Spatial distribution of various forms of iron in heavy metal polluted soils. Ecol. Chem. Eng. A 19, 1239–1246.
Demattê, J.A.M., Nanni, M.R., Formaggio, A.R., Epiphanio, J.C.N., 2007. Spectral reflectance for the mineralogical evaluation of Brazilian low clay activity soils. Int. J. Remote Sens. 28, 4537–4559.
Drury, S.A., 2016. Image Interpretation in Geology, second ed. Nelson Thornes, London.
Ducart, D.F., Silva, A.M., Toledo, C.L.B., Assis, L.M., 2016. Mapping iron oxides with Landsat-8/OLI and EO-1/Hyperion imagery from the Serra Norte iron deposits in the Carajás Mineral Province, Brazil. Braz. J. Geol. 46, 331–349.
ESRI, 2011. ArcGIS Desktop: Release 10.2. Environmental Systems Research Institute, Redlands, CA.
Fang, X., Xue, Z.J., Li, B.C., An, S.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.
Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression models. PLoS One 12

(1). https://doi.org/10.1371/journal.pone.0170478.
Gee, G.W., Or, D., 2002. Particle-size analysis. In: Dane, J.H., Topp, G.C. (Eds.), Methods of Soil Analysis. Part 4. Physical Methods. Agron. Monogr., vol. 9. ASA, CSSA, and SSSA, Madison, WI, pp. 255–293.
Ghalibaf, M.B., Mousavi, Z., 2014. Development and environment in Urmia Lake of Iran. Eur. J. Sustain. Dev. 3, 219–226.
Gilabert, M.A., Gonzalez-Piqueras, J., Garcia-Haro, F.J., Melia, J., 2002. A generalized soil-adjusted vegetation index. Remote Sens. Environ. 82, 303–310.
Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87.
Haese, R.R., Wallmann, K., Dahmke, A., Kretzmann, U., Muller, P.J., Schulz, H.D., 1997. Iron species determination to investigate early diagenetic reactivity in marine sediments. Geochim. Cosmochim. Acta 61, 63–72.
Hengl, T., Heuvelink, G.B.M., Kempen, B., Leenaars, J.G.B., 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 (6). https://doi.org/10.1371/journal.pone.0125814.
Hesami, A., Amini, A., 2016. Changes in irrigated land and agricultural water use in the Lake Urmia basin. Lake Reservoir Manage. 32, 288–296.
Heydari, N., 2013. Physical Chemistry of Urmia Lake. Department of Environment of Iran, Urmia Lake Restoration Staff.
IRIMO, 2012. Islamic Republic of Iran Meteorological Organization.
Jones, M.J., 1973. The organic matter content of the savanna soils of West Africa. Eur. J. Soil Sci. 24, 42–53.
Kabata-Pendias, A., 1993. Behavioural properties of trace metals in soils. Appl. Geochem. 8, 3–9.
Khamseh, A., Shahabzi, F., Oustan, Sh., Najafi, N., Davatgar, N., 2017. Impact of tailings dam failure on spatial features of copper contamination (Mazraeh mine area, Iran). Arab. J. Geosci. 10 (244). https://doi.org/10.1007/s12517-017-3040-y.
Kidd, D., Webb, M., Malone, B., Minasny, B., McBratney, A., 2015. Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. Soil Res. 53 (8), 932–955.
Kim, J., Grunwald, S., Osborne, T.Z., Robbins, R., Yamataki, H., Rivero, R.G., 2012. Spatial resolution effects of remote sensing images on digital soil models in aquatic ecosystems. In: Minasny, B., Malone, B.P., McBratney, A. (Eds.), Digital Soil Assessments and Beyond. Taylor & Francis Group, London, pp. 121–125.
Kuhn, M., Weston, S., Keefer, C., Coulter, N., 2016. C code for Cubist. Cubist: Rule- and Instance-based Regression Modeling. R Package Version 0.0.19. https://CRAN.R-project.org/package=Cubist.
Liddicoat, C., Maschmedt, D., Clifford, D., Searle, R., Herrmann, T., Macdonald, L.M., Baldock, J., 2015. Predictive mapping of soil organic carbon stocks in South Australia's agricultural zone. Soil Res. 53, 956–973.
Lin, L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.
Loeppert, R.L., Inskeep, W.P., 1996. Iron. In: Sparks, D.L. (Ed.), Methods of Soil Analysis. Part 3, 3rd ed. Agron. Monogr., vol. 9. ASA, CSSA, and SSSA, Madison, WI, pp. 639–664.
Luoto, M., Hjort, J., 2005. Evaluation of current statistical approaches for predictive geomorphological mapping. Geomorphology 67, 299–315.
Ma, Y., Minasny, B., Wu, C., 2017. Mapping key soil properties to support agricultural production in Eastern China. Geoderma Reg. 10, 144–153.
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, pp. 271.
McBratney, A.B., Mendoça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52.
McBratney, A.B., Minasny, B., Wheeler, I., Malone, B.P., Linden, D.V.D., 2012. Frameworks for digital soil assessment. In: Minasny, B., Malone, B.P., McBratney, A. (Eds.), Digital Soil Assessments and Beyond. Taylor & Francis Group, London, pp. 9–15.
McGrath, S.P., Cunliffe, C.H., 1985. A simplified method for the extraction of the metals Fe, Zn, Cu, Ni, Cd, Pb, Cr, Co and Mn from soils and sewage sludges. J. Sci. Food Agric. 36, 794–798.
McKeague, J.A., Day, J.H., 1966. Dithionate- and oxalate-extractable Fe and Al as aids in differentiating various classes of soils. Can. J. Soil Sci. 46, 13–22.
McKenzie, N.J., Webster, R., Ryan, P.J., 2008. Sampling using statistical methods. In: McKenzie, N.J., Grundy, M.J., Webster, R., Ringrose-Voase, A.J. (Eds.), Guidelines for Surveying Soil and Land Resources, Second edition. CSIRO, Australia, pp. 319–326.
Metternicht, G.I., Zinck, J.A., 2003. Remote sensing of soil salinity: potentials and constraints. Remote Sens. Environ. 85, 1–20.
Minasny, B., McBratney, A.B., 2002. Uncertainty analysis for pedotransfer functions. Eur. J. Soil Sci. 53, 417–429.
Minasny, B., McBratney, A.B., 2016. Digital soil mapping: a brief history and some lessons. Geoderma 264, 301–311.
Mohammed, I.N., Tarboton, D.G., 2012. An examination of the sensitivity of the Great Salt Lake to changes in inputs. Water Resour. Res. 48, W11511. https://doi.org/10.1029/2012WR011908.
Moody, L.E., Graham, R.C., 1995. Geomorphic and pedogenic evolution in coastal sediments, central California. Geoderma 67, 181–201.
Moskovchenko, D.V., Kurchatova, A.N., Fefilov, N.N., Yurtaev, A.A., 2017. Concentrations of trace elements and iron in the Arctic soils of Belyi Island (the Kara Sea, Russia): patterns of variation across landscapes. Environ. Monit. Assess. 189 (210). https://doi.org/10.1007/s10661-017-5928-0.
Mulder, V.L., Bruin, S., Schaepmana, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — a review. Geoderma 162, 1–19.
Nasiri, A., Saghafian, B., Pourhoseini, S.P., 2015. Analysis of relation between Lake Urmia's area with meteorological and hydrological droughts using Landsat satellite images. J. Sci. Res. Dev. 2 (6), 306–315.
Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. In: Sparks, D.L. (Ed.), Methods of Soil Analysis. Chemical Methods. Part 3. ASA, CSSA, and SSSA, Madison, WI, pp. 961–1010.
Parks, S.A., Dillon, G.K., Miller, C., 2014. A new metric for quantifying burn severity: the relativized burn ratio. Remote Sens. 6, 1827–1844.
Pezeshki, S.R., Delaune, R.D., 2012. Soil oxidation-reduction in wetlands and its impact on plant functioning. Biology 1, 196–221.
Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of AI92, 5th Australian Conference on Artificial Intelligence. World Scientific, Singapore, pp. 343–348.
Raizada, S., Chadha, N.K., Hasan, J., Maheshwari, U.K., 2003. Physical soil characteristics of waterlogged and saline lands of five districts of Haryana, India. J. Indian Fish. Ass. 30, 1–8.
Razmara, P., Massah Bavani, A.R., Motiee, H., Torabi, S., Lotfi, S., 2013. Investigating uncertainty of climate change effect on entering runoff to Urmia Lake Iran. Hydrol. Earth Syst. Sci. Discuss. https://doi.org/10.5194/hessd-10-2183-2013.
Rhoades, J.D., 1996. Salinity: electrical conductivity and total dissolved solids. In: Sparks, D.L. (Ed.), Methods of Soil Analysis. Chemical Methods. Part 3. ASA, CSSA, and SSSA, Madison, WI, pp. 417–435.
Shahbazi, F., Aliasgharzad, N., Ebrahimzad, S.A., Najafi, N., 2013. Geostatistical analysis for predicting soil biological maps under different scenarios of land use. Eur. J. Soil Biol. 55, 20–27.
Shao, D., Zhan, Y., Zhou, W., Zhu, L., 2016. Current status and temporal trend of heavy metals in farmland soil of the Yangtze River Delta region: field survey and meta-analysis. Environ. Pollut. 219, 329–336.
Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.
Sipos, P., Choi, C., Nemeth, T., Szalai, Z., Poka, T., 2014. Relationship between iron and trace metal fractionation in soils. Chem. Speciat. Bioavailab. 26, 21–30.
Skakun, R.S., Wulder, M.A., Franklin, S.E., 2003. Sensitivity of the thematic mapper enhanced wetness difference index to detect mountain pine beetle red-attack damage. Remote Sens. Environ. 86, 433–443.
Stonehouse, H.B., Arnaud, R.J., 1971. Distribution of iron, clay and extractable iron and aluminium in some Saskatchewan soils. Can. J. Soil Sci. 51, 283–292.
Sun, X.L., Zhao, Y.G., Zhao, M.S., Zhang, G.L., 2012. Spatial agricultural soil quality evaluation based on digital soil maps and uncertainty analysis. In: Minasny, B., Malone, B.P., McBratney, A. (Eds.), Digital Soil Assessments and Beyond. Taylor & Francis Group, London, pp. 37–42.
Tack, F.M.G., Verloo, M.G., 1997. Single extractions versus sequential extraction for the estimation of heavy metal fractions in reduced and oxidised dredged sediments. Chem. Speciat. Bioavailab. 11, 43–50.
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. https://doi.org/10.3390/f7110267.
Therneau, T., Atkinson, B., Ripley, B., 2017. rpart: Recursive Partitioning and Regression Trees. R Package Version 4.1-11. https://CRAN.R-project.org/package=rpart.
ULRP, 2015. Report of Lake Urmia Conditions. Urmia Lake Restoration Program. University of Sharif, Iran.
Viscarra Rossel, R.A., 2011. Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. J. Geophys. Res. 116, F04023. https://doi.org/10.1029/2011JF001977.
Waring, C., Stockmann, U., Malone, B.P., Whelan, B., McBratney, A., 2014. Is percent 'projected natural vegetation soil carbon' a useful indicator of soil condition? In: Hartemink, A.E., McSweeney, K. (Eds.), Soil Carbon. Progress in Soil Science. Springer, pp. 219–227.
Wish-Wilson, P., 2002. The Aral Sea environmental health crisis. J. Rural Remote Environ. Health 1 (2), 29–34.
Xiong, X., Grunwald, S., Myers, B., Kim, J., Harris, W.G., Bliznyuk, N., 2015. Assessing uncertainty in soil organic carbon modeling across a highly heterogeneous landscape. Geoderma 251-252, 105–116.