

Predicting weathering indices in soils using FTIR spectra and random forest models

Maryam Ghebleh Goydaragh ^{a,e,*}, Ruhollah Taghizadeh-Mehrjardi ^{b,c}, Ahmad Golchin ^d, Ali Asghar Jafarzadeh ^a, Marcos Lado ^e

^a Department of Soil Science, Faculty of Agriculture, University of Tabriz, Iran

^b Institute of Geography, Eberhard Karls University Tübingen, Tübingen, Germany

^c Faculty of Agriculture and Natural Resources, Ardakan University, Ardakan, Iran

^d Department of Soil Science, Faculty of Agriculture, University of Zanjan, Iran

^e Centro de Investigaciones Científicas Avanzadas, Faculty of Sciences, University of A Coruna. A Zapateira s/n, 15071 A Coruna, Spain

ARTICLE INFO

Keywords:

Aridisols
FTIR spectroscopy
Weathering indices
Random forest
Soil formation
Spatial variability

ABSTRACT

Weathering indices based on the relative proportions of different chemical elements are a useful tool to investigate the degree of weathering of soils. This characterization is missing in West Azerbaijan, northern Iran, and thus the main goals of this work were to assess the suitability of different indices to determine soil weathering, and to predict weathering indices using Fourier Transform Mid-infrared (FTIR) spectroscopy and Random Forest (RF) models. Soil major elemental oxides were determined in 12 profiles by X-ray Fluorescence (XRF), and the degree of chemical weathering was characterized using various indices: Chemical Index of Alteration (CIA), Chemical Index of Weathering (CIW), Mineralogical Index of Alteration (MIA), Weathering Index of Parker (WIP), $\text{SiO}_2/\text{R}_2\text{O}_3$, Vogt Index (V) and Chemical Proxy of Alteration (CPA). For each index, a RF model based on the FTIR spectra of the samples was optimized and calibrated. The soils showed a weak to moderate degree of weathering, and CIA, CIW, MIA and V were the most appropriate for this characterization. The constructed models showed R^2 values of 0.79, 0.75 and 0.71 for MIA, CIA, and CIW, respectively. The important spectral bands for prediction were those related to the presence of smectitic clays, which indicates that weathering is related to the alteration of primary minerals and the neoformation of smectites. The predicted spatial distribution of the weathering indices showed that Typic Haplalcids and Typic Haploxerepts are at an earlier stage of weathering compared to Haplargids and Haplocambids in the area. The combination of FTIR spectroscopy and RF models is a rapid, efficient, and cost-effective technique to predict weathering indices in large datasets when XRF data are limited.

1. Introduction

Soil types and their characteristics are the result of the interplay between processes and soil forming factors (Jenny, 1941; Osat et al., 2016). Following the model of soil formation proposed by Jenny (1941), the pedogenesis depends on several parameters such as the nature of the parent material, climate, time, organisms and topography. In this context, weathering is explained as transformation of parent materials by physicochemical and biological processes (Oliva et al., 2003). Among all the processes that can lead to weathering, the most significant ones are those attributed to the chemical transformation of existing minerals,

which involve the depletion and removal of some primary minerals and the formation of secondary ones and amorphous oxy-hydroxides (Jeleńska et al., 2008). Assessing the weathering intensity, and how soil properties have been modified by different chemical and physical processes, provides information about what is happening and is going to happen with our present soils.

The degree of weathering of soils and the conditions of soil development are commonly characterized using chemical weathering indices. These have been designed to quantify the changes in soil properties and rock materials during their evolution (Parker, 1970; Price and Velbel, 2003). In general, their variation with depth is continuous, gradual or

* Corresponding author.

E-mail addresses: Maryam.ghebleh@tabrizu.ac.ir (M. Ghebleh Goydaragh), ruhollah.taghizadeh-mehrjardi@mnf.uni-tuebingen.de (R. Taghizadeh-Mehrjardi), Marcos.lado@udc.es (M. Lado).

steady, for homogeneous parent materials (Price and Velbel, 2003) and therefore they can be applied to compare the degree of evolution of various soil horizons. In this case, there are numerous chemical weathering indices adjusted to different rock types, climate regimes, tectonic and topographic locations, vegetation, soil development and human influence (Oliva et al., 2003; Osat et al., 2016). Among them, some of the most commonly applied are the Chemical Index of Alteration (CIA - Nesbitt and Young, 1982; Price and Velbel, 2003; Shao et al., 2012), the Chemical Index of Weathering (CIW - Harnois, 1988; Baumann et al., 2014), the Weathering Index of Parker (WIP - Parker, 1970; Sorokina and Gysev, 2018), the Plagioclase Index of Alteration (PIA - Baumann et al., 2014), the Vogt Index (V - Vogt, 1927), and the Mineralogical Index of Alteration (MIA - Voicu and Bardoux, 2002). Some of these indices such as the CIA and CIW indices are appropriate to evaluate the incipient stage of soil weathering rather than well-weathered soils (Zhang et al., 2007). They can be used to evaluate the effect of climate on weathering intensity (Nadlonek and Bojakowska, 2018), and their values are related to climate and are sensitive to differences in rainfall (Fiantis et al., 2010). In addition, All of these mentioned indices are defined as ratios of immobile to mobile elements present in soils, including Ca, Na, K, Al, Fe, K, P, Mn, Mg or Ti contents, which are commonly measured as oxides using X-ray fluorescence (XRF) spectrometry (Parker, 1970; Osat et al., 2016).

Numerous studies have proved the suitability of these chemical weathering indices to characterize soil evolution under different climates and parent materials, such as weakly weathered soils developed on sedimentary rocks from southern Iran (Abbaslou et al., 2013), or soils developed on volcanic ash in Indonesia (Fiantis et al. (2010)). For example, Duzgoren-Aydin et al. (2002) used thirty different indices including CIW, CIA, PIA, PIW, and the ratio $\text{SiO}_2/\text{R}_2\text{O}_3$ (where $\text{R}_2\text{O}_3 = \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3 + \text{TiO}_2$) to evaluate the suitability of soils for engineering applications, and concluded that the ratios of some mobile oxides to one or more immobile ones are important parameters to describe weathering of soils developed on pyroclastic rocks. Similarly, Price and Velbel (2003) used the CIA, CIW, PIA, and V indices for characterizing the weathering of soil profiles in North Carolina, and found that these indices are sensitive to geochemical alterations such as hydrothermal variations along faults or alterations caused by the water table. Dengiz et al. (2013) calculated some weathering indices for a Regosol, a Cambisol and a Vertisol sampled in Turkey under the same parent materials (limestone marls), and concluded that indices such as CIA, PIA and WIP were strongly influenced by topographic position.

As mentioned before, weathering indices are calculated from the soil elemental composition measured using XRF spectroscopy. This technique is costly and time-consuming, and thus the characterization of the spatial variability of weathering conditions is limited by the amount of samples that can be analyzed. An interesting alternative to this method is the use of other spectroscopic techniques such as Near Infrared (NIR) or Fourier Transform Mid-infrared (FTIR) spectroscopies, which allow spectral measurements of high numbers of samples with minimum cost, time and sample processing.

In this context, FTIR spectra seems very suitable to predict weathering indices due to the distinctive signatures of soil components in the Mid-infrared (MIR) region (400 to 4000 cm^{-1}), where different molecular functional groups can be easily recognizable through spectral libraries (Reeves, 2010). In fact, FTIR has been successfully used to explain the evolution of different soil horizons (Chapman et al., 2001). For example, Ben-Dor and Banin (1994) measured soil reflectance in the 400–1100 nm region to estimate SiO_2 , Fe_2O_3 , CaCO_3 , Al_2O_3 contents, and mass loss on ignition using a simple linear regression. Also, FTIR and vis-NIR (350 to 2500 nm) have been used for the quantitative estimation of clay minerals (Saikia and Parthasarathy, 2010), since FTIR can be used as a complementary method to X-ray diffraction (XRD) to identify clay minerals (Madejová, 2003).

Recently, the combination of spectral data and multivariate statistical approaches has achieved generality to quantify soil properties

(Viscarra Rossel et al., 2006). Different statistical models have been used for this purpose, such as Multiple Linear Regression, Partial Least Squares Regression (PLSR), Artificial Neural Networks (ANN), or Random Forest (RF). The RF model is an extension of tree-based classification and regression algorithms, and it has been found to be the best model for classification and quantification of soil properties (Viscarra Rossel and Behrens, 2010). Hobley et al. (2017) tested a combination of FTIR-ATR data and RF model to determine charcoal and litter in the soil, and found that RF was the best model to predict both charcoal and litter contents of the samples. Rial et al. (2016) applied RF and FTIR-ATR spectra to predict SOC in NW Spain, with a good performance of the model. Despite these examples, its use in combination with spectroscopic data is not extensive, and to our knowledge, it has not been used to relate spectroscopic data and soil weathering indices.

During the last decade, several studies have been conducted to analyze the type of clays and soil properties in Northern Iran (Moradi et al., 2012; Farid Giglo et al., 2014; Asghari Saraskanroud et al., 2017; Goydaragh et al., 2019; Goydaragh et al., 2021). Whereas, the study of their major elemental oxides has been evaluated (Goydaragh et al., 2019), the characterization of their weathering degree using chemical indices has not been analyzed so far. Moreover, although there are some examples relating FTIR spectra and WI through PLSR, the potential use of RF to estimate the variability of soil weathering indices from a limited number of samples has not been investigated. Thus, the main goal of this paper was to evaluate the intensity of weathering and alteration of soils in the Miandoab county, West Azerbaijan, Iran, using different weathering indices and to examine the suitability of FTIR spectroscopy and RF model for predicting different weathering indices as a rapid and inexpensive technique.

2. Materials and methods

2.1. Study area

The study area is situated in Miandoab county, West Azerbaijan Province, Iran (between longitudes $46^{\circ} 0' 0'' - 46^{\circ} 15' 0''\text{E}$ and latitudes of $36^{\circ} 52' 30'' - 37^{\circ} 2' 30''\text{N}$), and covers approximately 5000 ha (Fig. 1). Elevations ranged from 1283 to 1325 m above sea level. This area is situated on a high-elevation old alluvial fan and terrace, on the Zarrine Rud river, and it was formed during the Cenozoic (Roozitalab et al., 2018). There is some evidence of erosion and redistribution of the eroded material in the area. The climate is arid and semi-arid, being aridic and mesic the main soil moisture and temperature regimes. The average annual air temperature is 12.8 °C, and the average annual rainfall is 272.3 mm. The land use type is mainly agricultural, dominated by nuts and fruits orchards and cereal crops (Fig. 2). According to the Soil Taxonomy classification system (Soil Survey Staff, 2014), the soils in this area belong to two orders (Aridisols and Inceptisols), four suborders (Calcids, Cambids, Argids and Xerepts), four great groups (Haplocalcids, Haplocambids, Haplargids and Haploxerepts), and four subgroups (Typic Haplocalcids, Typic Haplocambids, Typic Haplargids and Typic Haploxerepts). Representative profiles of the four subgroups with their respective horizons are presented in Fig. 3, where their topographic location in a typical catena of the region is also shown.

2.2. Soil sampling and analysis

Soil samples were collected from 12 profiles distributed along the study area (Fig. 1), based on the previous knowledge about the spatial distribution of the main soil types. After recognition and characterization of soil genetic horizons in each profile up to a depth of 2 m, samples consisting of a homogenous mixture of all material in each horizon were collected. In total, the number of collected samples was 45.

Samples were transported to the laboratory, air-dried and sieved through a 2-mm mesh to retrieve the fine-earth fraction for analyses. The main soil properties were determined by standard methods. The hy-

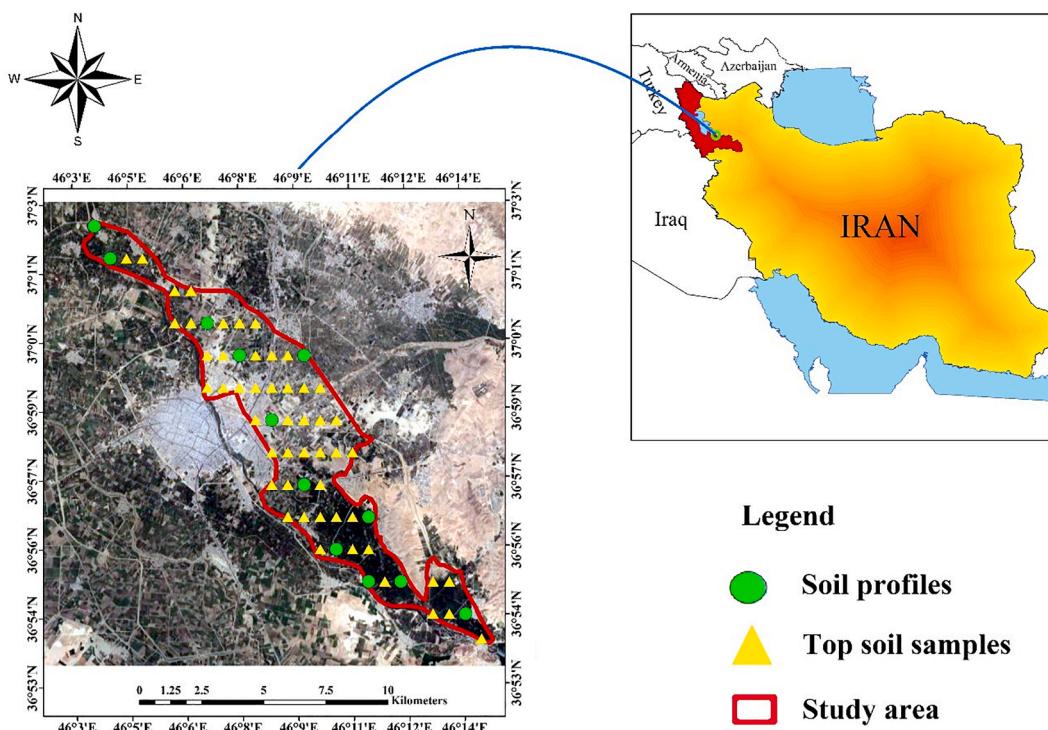


Fig. 1. Aerial picture of the study area and sampling scheme showing the location of the sampled profiles and top soil points.

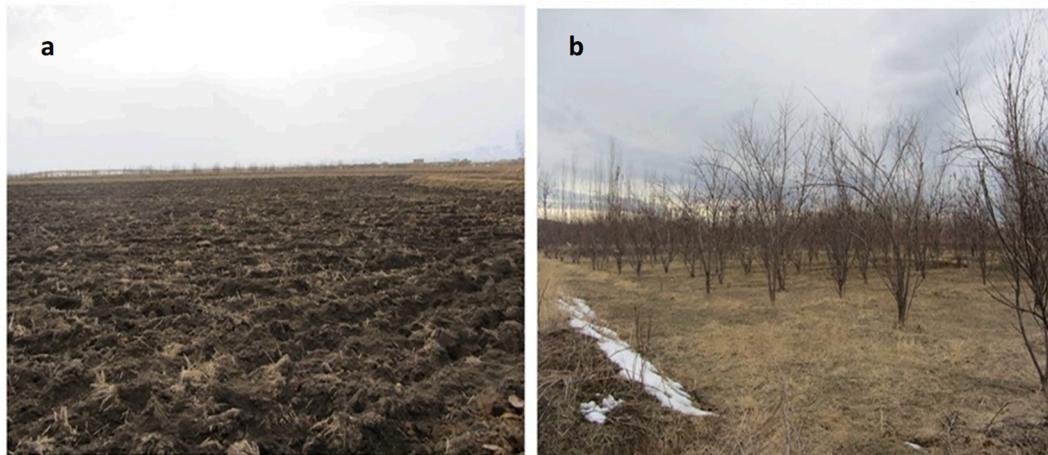


Fig. 2. View of the common crops of the area. a) fallow soil in a field dedicated to cereal crops; b) almond orchard.

drometer method was used to measure particle size distribution (Gee and Bauder, 1986); organic carbon (OC) was analyzed in a Flash EA 1112 elemental analyzer (ThermoQuest, USA) after carbonates removal by the addition of HCl; Calcium Carbonate Equivalent (CCE), and Cation Exchange Capacity (CEC) were measured by a titration method (Nelson and Sommers, 1996), and Bower method (Bower et al., 1952), respectively. Soil pH, electrical conductivity (EC_e), and soluble cations (Na^+ , Ca^{2+} and Mg^{2+}) were analyzed in saturated paste extract (Nelson, 1982). Based on the [Soil Survey Staff \(2014\)](#), the Sodium Adsorption Ratio (SAR) was determined as:

$$\text{SAR} = \text{Na}^+ / \sqrt{\text{Ca}^{2+} + \text{Mg}^{2+}} \quad (1)$$

2.3. Elemental composition and weathering indices

Major elemental oxides were measured by XRF in a S4 Pioneer XRF

spectrometer (Bruker, USA), which were used to calculate several weathering indices. These indices were selected based on previous studies evaluating soil weathering (Price and Velbel, 2003; Bahlburg and Dobrzinski, 2011), and are described in [Table 1](#). All weathering indices were calculated using molar percentages of the oxides (the ratio of the weight percent of oxide to the molar mass of the oxide) from the total elemental composition.

The weathering index of Parker (Parker, 1970) indicates the amount of alkaline cations (Na, K, Mg, and Ca) released by mineral hydrolysis and subsequently leached.

Vogt's Index is widely used to determine which clays have been weathered in the soil (Roaldset, 1972). The CIA and CIW are essentially interpreted as a measure of the extent of conversion of feldspars to clays such as kaolinite (Nesbitt and Young, 1984). The MIA explains the degree of mineralogical weathering (i.e., the ratio of transformation of a primary mineral into its secondary by products - Voicu and Bardoux,

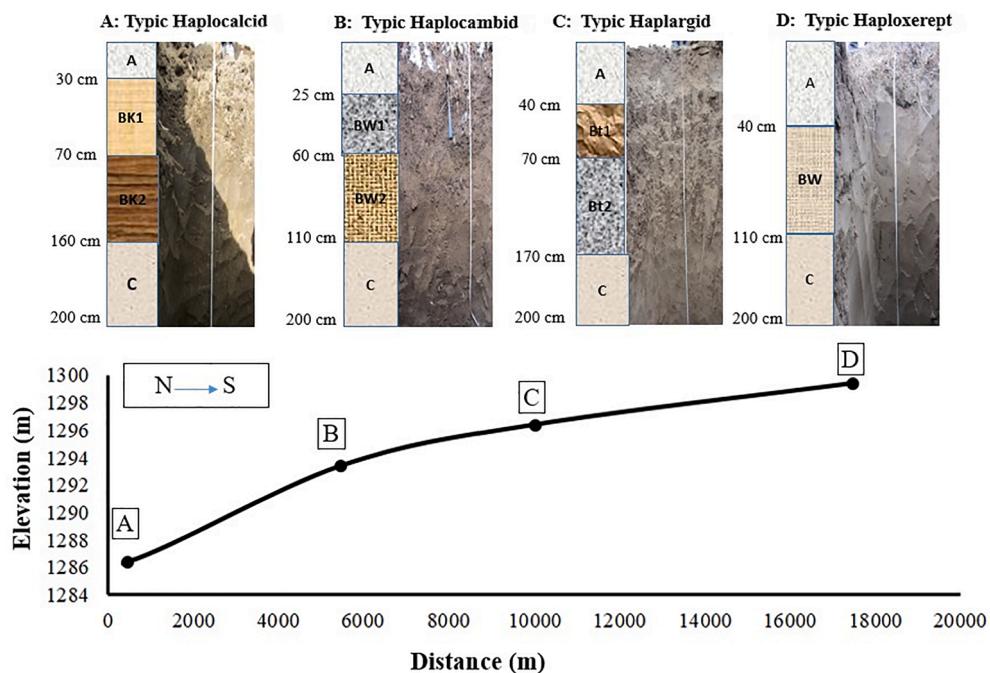


Fig. 3. The profiles of different subgroups and their topographic location along a transect in the study area. A: mineral surface horizon; Bk1 and Bk2: accumulation of secondary carbonates; BW, Bw1 and Bw2: structural B horizon; Bt1 and Bt2: illuvial accumulation of clays; C: unconsolidated parent material. A: Typic Haplocalcids; B: Typic Haplocambids; C: Typic Haplargids; D: Typic Haploxerupts.

Table 1

Description of the weathering indices calculated in this study. All oxides contents should be expressed in molar units.

Weathering Indices	Equation	Reference
Vogt Index (V)	$(Al_2O_3 + K_2O)/(MgO + CaO^* + Na_2O)$	(Vogt, 1927)
Weathering Index of Parker (WIP)	$[(Na_2O/0.35) + (MgO/0.9) + (K_2O/0.25) + (CaO^*/0.7)] \times 100$	(Parker, 1970)
Chemical Index of Alteration (CIA)	$[(Al_2O_3)/(Al_2O_3 + CaO^* + Na_2O + K_2O)] \times 100$	(Nesbitt and Young, 1982)
Chemical Index of Weathering (CIW)	$[(Al_2O_3)/(Al_2O_3 + CaO^* + Na_2O)] \times 100$	(Harnois, 1988)
SiO_2/R_2O_3	$(SiO_2)/(Al_2O_3 + TiO_2 + Fe_2O_3)$	(Singh et al., 1998)
Chemical Proxy of Alteration (CPA)	$[(Al_2O_3)/(Al_2O_3 + Na_2O)] \times 100$	(Cullers, 2000)
Mineralogical Index of Alteration (MIA)	$2 \times (CIA-50)$	(Voicu and Bardoux, 2002)

* Calcium oxide content should be limited to the one present in silicate minerals.

2002). For the calculation of some indices involving CaO content, like V, WIP, CIA and CIW, the amount of this oxide was limited to the one present in silicate minerals. Hence, according to the McLennan's method, the amount of CaO was subtracted from phosphate (P_2O_5). If the remaining number was lower than the amount of Na_2O , the CaO was applied as CaO, and in the opposite case, the CaO^* was assumed to be equivalent to Na_2O (McLennan, 1993).

2.4. FTIR spectroscopic analysis

For FTIR analysis, dried soil samples were ground in a mortar, and 1–2 mg of each soil sample was mixed with 200 mg of Potassium Bromide (KBr), ground again to homogenize the sample, and then pressed using a hydraulic press to prepare a pellet. Afterwards, a Bruker Vector 22 spectrophotometer was used to determine the absorbance of the pellets in the mid-infrared region (400 to 4000 cm^{-1} with a resolution of 2 cm^{-1}). Spectra were preprocessed for baseline and scatter correction using Extended Multiplicative Signal Correction (EMSC) (Martens and Stark, 1991), which are standard preprocessing steps in FTIR analysis. EMSC is particularly useful to remove unwanted additive and multiplicative effects in spectral data (Skogholz et al., 2019).

2.5. Spectral model development for weathering indices

A chemometric model (Mohanty et al., 2016) was applied to detect relations between the weathering indices and FTIR spectra, and to recognize the most relevant spectral bands for this purpose. The process consisted on fitting a RF model with weathering indices as dependent variables and spectral data as predictors. RF was chosen because its advantages compared to other machine-learning methods: it is non-parametric, it can be used in datasets with more predictors than observations, and it does not overfit (Díaz-Uriarte and De Andres, 2006; Wiesmeier et al., 2011). Moreover, it is able to manage small-n-large-p problems, and it is relatively strong to noise and outliers (Guio Blanco et al., 2018).

A RF model is an ensemble of decision trees that can be used for both regression and classification purposes (Dharumarajan et al., 2017). Based on the improvement of the residual sum of squares (RSS), the initial dataset is recursively divided into increasingly homogeneous subsets according to the target variable in each tree. All predictors are examined at each division to decide which value of the predictor and which predictor best subdivides the dataset. After the forest is generated, each tree in the forest predicts the target variable, and the predictions are averaged to give a single final prediction. In the RF models, the observations set aside from model training, referred as “out of bag”

(OOB), are used to predict the accuracy of model (Breiman, 2001).

A set of hyperparameters can be optimized in order to enhance RF model's performance. These include the number of trees (*num.trees*), the size of the predictor subset (FTIR spectral bands *mtry*), the maximum depth of the tree (*max.depth*), and the minimum number of samples in terminal nodes (*min.nod.size*). There are different methods of optimizing hyperparameters such as random search and grid search. In our study, we used particle swarm optimization (PSO) that has been successfully used for optimization problem in many studies (Pham et al., 2018; Pham et al., 2020). Briefly, it consists of a group of particles (i.e., random potential solutions); each particle in each group moves randomly in a search space and is influenced by surrounding situation during movement (Chen et al., 2017). Based on its current situation and velocity, the status of each particle is updated for each iteration in order to identify optimal solutions (Hajihassani et al., 2018). The final optimal solutions (i.e., hyperparameters) are selected as the one maximizing the accuracy of predictions based on out-of bag (OOB) data. The search for the optimal *num.trees* was done from 50 to 3000, at steps of 100 trees; *mtry* was tested from 5 to 1000, at steps of 50; *max.depth*, from 3 to 30 at unit steps; and *min.node.size*, from 2 to 10, also at unit steps. Table 2 shows the values for these hyperparameters in the models constructed for each of the weathering indices that showed the best performance.

2.6. Model validation

To calibrate and validate the model, data was split into training and testing sets. A proportion of 70% of the soil samples were randomly selected and used for constructing the model, while the remaining 30% was used for validation.

The performance of the model was evaluated comparing predicted values and observed data in the test set. Several parameters were calculated, including coefficient of determination (R^2 - Eq. (2)), Root Mean Squared Error (RMSE - Eq. (3)), Ratio of Performance to Deviation (RPD - Eq. (4)), Mean Absolute Error (MAE - Eq. (5)) and Ratio of Performance to Interquartile distance (RPIQ - Eq. (6)).

$$R^2 = 1 - SS_{error}/SS_{total} \quad (2)$$

$$RMSE = \sqrt{1/n \sum_{i=1}^n (O_i - P_i)^2} \quad (3)$$

$$RPD = SD/RMSE \quad (4)$$

$$MAE = 1/n \sum_{i=1}^n |P_i - O_i| \quad (5)$$

$$RPIQ = (Q3 - Q1)/RMSE \quad (6)$$

where O_i and P_i are the observed and predicted weathering indices

Table 2

Hyperparameters of the Random Forest model providing the best performance for each weathering index.

	Num.Trees	mtry	Max.Depth	Min.Nod.Size
CIA	700	650	4	5
CIW	600	600	5	7
WIP	500	450	5	6
CPA	500	750	6	5
MIA	700	650	4	5
SiO ₂ /R ₂ O ₃	800	550	5	4
V	400	500	7	4

Num.Trees: the number of trees; mtry: the size of the predictor; Max.Depth: the maximum depth of the tree; Min.Nod.Size: the minimum number of samples in terminal nodes; CIA: Chemical Index of Alteration; CIW: Chemical Index of Weathering; WIP: Weathering Index of Parker; CPA: Chemical Proxy of Alteration; MIA: Mineralogical Index of Alteration; V: Vogt Index.

values at the i th point; n is the number of observation in the model; and SD is the standard deviation of the measured values (Zeraatpisheh et al., 2019). The RPD values were calculated as the ratio of the standard deviation of the measured values to RMSE: values below 1.4 indicate a poor performance, values between 1.4 and 2.0 are considered as a moderate performance, and values above 2.0 suggest a good estimation (Qi et al., 2018). The RPIQ was computed using the first (Q1) and third (Q3) quartiles of the measured values and the RMSE: values below 1.4 indicate a very poor model, those between 1.4 and 1.7 characterize a fair model, values between 1.7 and 2.0 are typical of good models, those between 2.0 and 2.5 denote a very good model, and values above 2.5 appear in excellent models (Nawar and Mouazen, 2017). In summary, the best estimation showed the highest R^2 , RPD and RPIQ and lowest RMSE and MAE.

RF provides also the relative importance of each spectral band in the prediction of the weathering indices. RF models estimate the importance of a variable (in this case, spectral bands) by computing changes in node purity and the estimation error when OOB data for that variable are permuted whereas all other variables remain unchanged (Rasaei and Bogaert, 2019). With this procedure, it is possible: 1) to identify the variables that are relevant for the estimation of the dependent variable; 2) to define a small number of variables that produce a good estimation of the output variable (Genuer et al., 2010).

2.7. Implementation of the model to predict weathering indices for unavailable XRF data: An illustrative example

The developed models were used to characterize the spatial variability of the weathering indices in topsoils of the study area. For this purpose, a second sampling scheme was developed. A total of 50 topsoil (0–0.25 m depth) samples were collected in a regular grid (sampling points separated by 700 × 1000 m) across the entire area. These samples were transported to the laboratory, air-dried and sieved to obtain the fine earth fraction. FTIR spectra from these samples were obtained following the steps described in 2.4. These spectra were used as input data in the RF models to calculate the values of the weathering indices for which the models showed a good performance.

2.8. Data analysis

All data processing was implemented in the R-Studio environment (R Core Team, 2016), including baseline removal with the package (Liland and Mevik, 2015), extended multiplicative signal correction with the EMSC package (Liland, 2016), construction of RF models using the ranger package (Wright and Ziegler, 2015), training the RF model using the caret package (Kuhn et al., 2017), and optimization the RF model using the pso package (Bendtsen and Bendtsen, 2011).

3. Results and discussion

3.1. Soil types and properties

Table 3 shows the main morphological and physicochemical characteristics of the representative four subgroups in the whole area.

The most distinctive feature of the analyzed profiles was the presence of cambic, argillic and calcic B horizons (in 68%, 16% and 16% of the soil profiles, respectively), which are usually found in arid and semi-arid regions (Buringh, 1981). All B horizons showed angular blocky structure, while in the A and C horizons the structure varied from granular to angular block, and from single granular to massive, respectively. CCE content was high, and an increase with depth was observed in all profiles. The morphology of the accumulations of carbonates in subsurface horizons were in filaments (stage I), prominent nodules (stage II) or plugged horizons (stage III) (Gile et al., 1966). Secondary carbonate nodules appeared in Typic Haplocalcids, and led to the formation of a calcic horizon. An argillic horizon with clay coatings and Fe mottles

Table 3

Morphological, chemical and physical properties of representative profiles corresponding to the four soil subgroups found in the study area. (from Goydaragh et al., 2019).

Horizon	Depth (cm)	Color (dry)	Color (moist)	Texture	Structure	Special feature	EC_e dSm $^{-1}$	pH	CEC cmol kg $^{-1}$	SAR (%)	CCE	OC	Clay	Silt	Sand
Typic Haplocalcids															
A	0–30	10YR5/2	10YR3/3	SIL	2mgr		4.9	7.9	12.6	4.5	11.5	1.8	18.9	55.8	25.3
Bk1	30–70	10YR5/2	10YR2/2	L	2mabk	+	9.5	7.8	13.3	10.3	13.3	1.2	14.0	46.9	38.9
Bk2	70–160	10YR6/4	10YR4/4	L	3mabk	+	12.2	8.5	8.3	11.9	14.0	0.1	7.0	44.9	48.0
C	160–200	10YR5/2	10YR3/3	S	sg		5.0	7.9	3.1	8.6	14.1	0.8	4.0	2.69	93.2
Typic Haplocambids															
A	0–25	10YR4/3	10YR3/4	C	3cabk		1.4	8.0	29.4	4.8	7.0	1.0	40.0	18.5	41.4
Bw1	25–60	10YR6/3	10YR4/3	CL	2cabk		1.4	8.0	16.4	8.0	11.5	0.8	38.1	27.8	34.0
Bw2	60–110	10YR6/4	10YR4/6	L	2mabk		3.0	8.5	13.1	12.1	17.0	0.4	21.0	36.9	42.0
C	110–200	10YR6/4	10YR4/4	L	ma		1.7	9.0	11.9	9.2	13.6	0.4	12.8	44.0	43.0
Typic Haplargids															
A	0–40	10YR6/3	10YR3/4	CL	1vfgr		3.8	7.9	19.7	10.6	12.8	0.9	30.2	49.8	19.9
Bt1	40–70	10YR5/3	10YR4/3	SiC	2fabk	+++	3.8	8.3	17.1	12.0	13.4	0.5	42.0	42.2	15.7
Bt2	70–130	10YR5/2	10YR3/4	C	2fabk	+++	8.7	8.4	18.7	32.3	13.1	0.5	59.9	23.1	16.9
C	130–200	10YR5/3	10YR4/4	L	ma		7.4	8.5	16.7	37.5	13.2	0.3	24.1	27.4	47.4
Typic Haploxerepts															
A	0–40	10YR5/3	10YR3/4	SL	1fgr		0.7	7.9	10.6	1.0	11.3	0.4	10.0	26.9	63.0
Bw	40–110	10YR5/4	10YR3/4	L	1cabk		0.6	7.9	9.1	0.5	13.4	0.3	11.1	44.8	44.0
C	110–200	10YR6/4	10YR3/6	SL	sg		0.6	8.0	8.2	0.4	14.3	0.2	9.0	31.8	59.1

+: CaCO₃ nodules, +++: Clay coatings, SIL: Silty Loam, L: Loam, S: Sandy, C: Clay, CL: Clay Loam, SiC: Silty Clay, SL: Sandy Loam. 1: Weak, 2: Moderate, 3: Strong. vf: Very Fine, f: Fine, m: Medium, c: Coarse, g: Granular, abk: Angular Blocky, ma: Massive, EC: Electrical Conductivity, CEC: Cation Exchange Capacity, SAR: Sodium Adsorption Ratio, CCE: Calcium Carbonate Equivalent, OC: Organic Carbon.

appeared in Typic Haplargids. Typic Haplocambids and Typic Haploxerepts showed cambic horizons.

The high EC_e and SAR values of the soils indicated that little leaching happened in this area. The pH values ranged from 7.9 to 9.0 (moderately basic to slightly alkaline) as a consequence of carbonate and salts accumulation. According to pH, EC_e, and SAR values, the Typic Haploxerepts and Typic Haplocambids were categorized as normal soils, the Typic Haplocalcids as saline, and the Typic Haplargids as saline-sodic. The OC values were low in all profiles, and the clay contents generally reduced with depth. The higher silt content in the surface layers in comparison to the subsurface layers in the Typic Haplocalcids and Typic Haplargids may be associated with eolian additions, which have a fundamental role in the pedogenesis of many arid and semi-arid regions (Khresat et al., 2004). CEC values were lower than 25 cmol_c kg $^{-1}$, as a consequence of the small quantities of organic matter.

3.2. Weathering indices

Descriptive statistics of weathering indices for sampled profiles are summarized in Table 4. The median and the mean values of all indices were similar, which is indicative of a normal distribution of the data. In addition, the coefficients of variation (CV) were low for CIA, CIW, WIP, CPA, SiO₂/R₂O₃ and moderate for MIA and V (Wilding, 1985), showing a fairly homogenous weathering stage of the soils in this area.

Thus, all twelve profiles were quite similar in terms of weathering despite the fact that they belong to different soil groups. The CIA and CIW indices had the same pattern (Fig. 4A and B), and resembled closely

each other in all profiles. Bahlburg and Dobrzinski, (2011) and Nesbitt and Young, (1982) reported that unweathered materials like basalts or fresh granites have values of CIA between 30 and 45%, and between 45 and 55%, respectively, while byproducts of weathering reach values close to 100%. In the present study, the range of CIA values was between 31.10 and 50.04%, which is similar to CIA and CIW values between 35 and 55% found by Abbaslou et al. (2013) in northern and eastern Iran. Therefore, these soils can be considered at incipient to intermediate stages of development.

WIP describes the total amount of alkaline soil metals (Na, K, Ca, and Mg) that are mobile during weathering (Bouchez et al., 2012). Values close to 0 indicate a high degree of soil weathering, whereas values between 60 and 100 show a low degree of weathering (Bahlburg and Dobrzinski, 2011). In these profiles, the values of WIP were all in a narrow range between 65.09 and 82.19 (Fig. 4C), and therefore assumed to be young (Fiantis et al., 2010).

The CPA is considered by some authors (e.g., Buggle et al., 2011) as the most suitable geochemical proxy or index for silicate weathering. Like CIA and CIW, CPA varies between 1 and 100, with higher values indicating a more intensive weathering. Fig. 4D shows that CPA values were relatively high (between 75.24 and 89.17) irrespective of the profile type. The low Na contents of these soils (1.1 to 2.9% with an average of 1.8%) is the reason for these high values of the index, misleading towards an intense weathering (Buggle et al., 2011).

The MIA index evaluates the degree of mineralogical weathering with values in the where a value of 100% indicates the complete transformation of primary minerals into their respective weathered

Table 4
Descriptive statistics of the weathering indices calculated from the soil profiles (n = 45).

Weathering Indices	Minimum	Maximum	Mean	SD	Median	CV (%)
CIA	31.10	50.04	41.25	5.06	41.55	12.26
CIW	32.81	55.31	44.85	6.19	45.13	13.81
WIP	64.72	90.97	76.95	5.95	76.70	7.73
CPA	75.24	89.17	84.21	3.69	85.54	4.39
MIA	12.21	50.08	32.50	10.54	33.11	32.45
SiO ₂ /R ₂ O ₃	3.70	6.77	4.35	0.61	4.27	14.12
V	0.43	0.92	0.68	0.12	0.67	18.74

SD: Standard Deviation; CV: coefficient of variation; CIA: Chemical Index of Alteration; CIW: Chemical Index of Weathering; WIP: Weathering Index of Parker; CPA: Chemical Proxy of Alteration; MIA: Mineralogical Index of Alteration; V: Vogt Index.

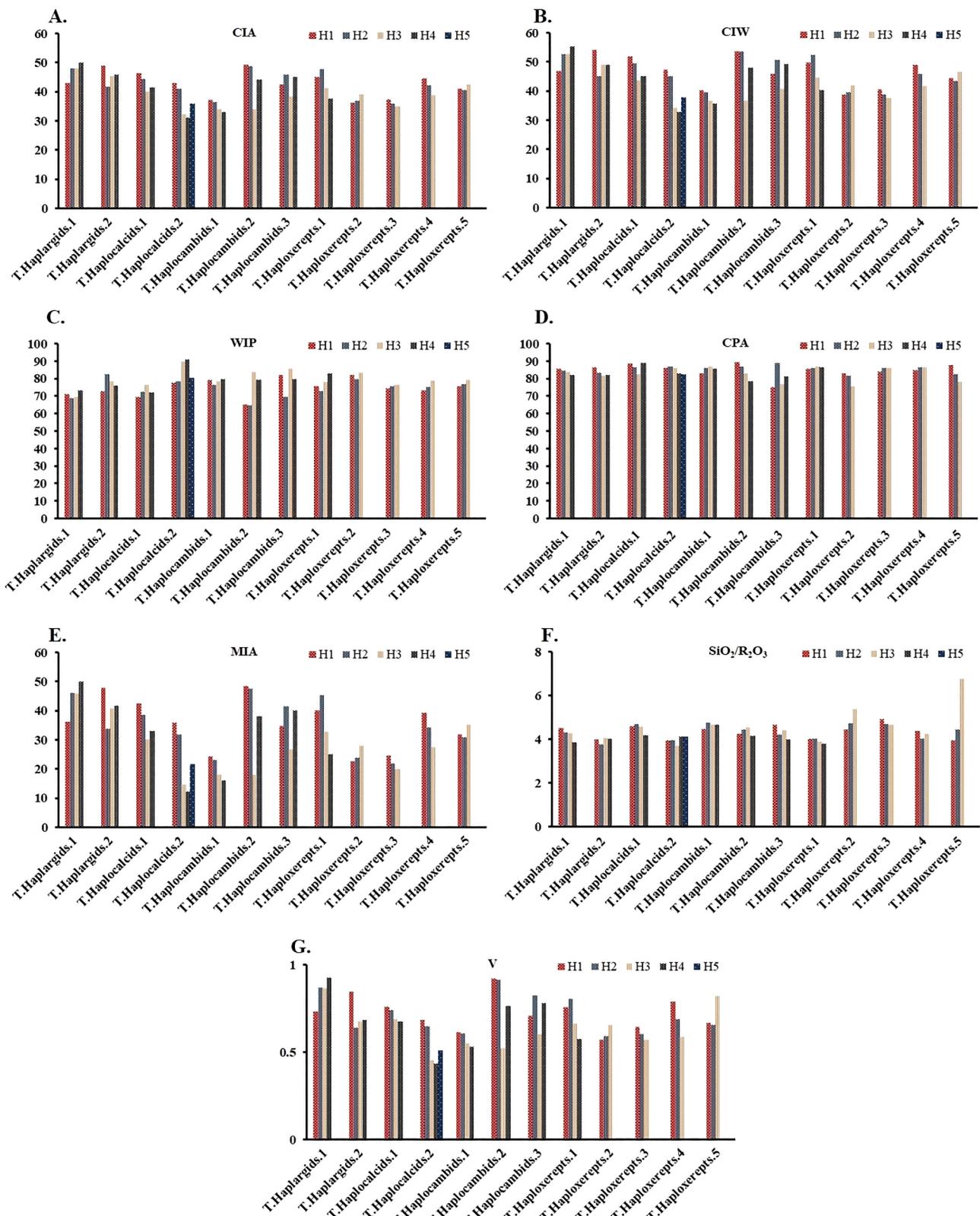


Fig. 4. The values of different weathering indices in the twelve reference profiles. CIA: Chemical Index of Alteration; CIW: Chemical Index of Weathering; WIP: Weathering Index of Parker; CPA: Chemical Proxy of Alteration; MIA: Mineralogical Index of Alteration; $\text{SiO}_2/\text{R}_2\text{O}_3$: $\text{SiO}_2/\text{R}_2\text{O}_3$; V: Vogt Index; H1, H2, H3, H4 and H5: The different horizons.

products (Voicu and Bardoux, 2002). In this study, MIA (Fig. 4E) showed a range of values between 12.21 and 50.08%, with an average of 32.50%, which indicates an incipient to weak weathering and the presence of high Fe contents (Mohanty et al., 2016). The same can be concluded from the values of the ratio $\text{SiO}_2/\text{R}_2\text{O}_3$ (where $\text{R}_2\text{O}_3 = \text{Al}_2\text{O}_3 + \text{Fe}_2\text{O}_3 + \text{TiO}_2$), which varied over a narrow range between 3.70 and 6.77, illustrating that these soils were fairly homogeneous (Singh et al., 1998) and weakly to moderately weathered (Fig. 4F). Similarly, Vogt values varied from 0.43 to 0.92 with an average of 0.68 (Fig. 4G), where values of lower than 1 indicate that soils are not highly weathered, while a Vogt value of infinite indicates that soils are close to complete weathering (Vogt, 1927).

In summary, all weathering indices except CPA seemed to be the adequate for the evaluation of weathering stages in this area, and all of them indicated the presence of young soils. The high pH of the soils due to the presence of carbonates, and the low amount of rainfall in the region, limited silicate weathering, feldspar transformations and the formation of clay minerals (Brady and Weil, 2013).

3.3. FTIR spectra

Fig. 5 shows the FTIR spectra of the 45 samples after baseline and EMSC corrections. FTIR spectra are good predictors of soil composition, since both inorganic and organic soil components have a distinctive signature in the MIR region of the spectra (Mohanty et al., 2016; Rial et al., 2016; Zannah et al., 2016). In this case, due to the low organic matter content of the soils (Table 3), the spectra are dominated by the presence of inorganic functional groups. Peaks at 3627, 3692 cm^{-1} are associated with OH vibrations in 1:1 clays, although that at 3627 cm^{-1} is also related to the presence of smectites (Madejová, 2003), which are very common clay minerals in arid and semi-arid regions. The band with a peak at 1030 cm^{-1} results from Si–O–Si stretch vibrations (Madejová, 2003; Goydaragh et al., 2019). Peaks at 800 and 780 cm^{-1} are related to symmetric stretching vibrations of Si–O bonds, typical of quartz and silica (Müller et al., 2014); and the peaks around 550–445 cm^{-1} correspond to Al–O–Si and Si–O–Si bending vibrations (Goydaragh et al., 2021). The peak at 1440 cm^{-1} is assigned to carbonates, and those at 712, 875 cm^{-1} indicate the presence of calcite instead of dolomite (Müller et al., 2014). Despite the dominance of inorganic soil components in the spectral signature, some peaks denote the presence of organic functional groups, especially those at 2920 and 2850 cm^{-1} that correspond to aliphatic C–H vibrations (Janik et al., 2007). Due to the clear signature of the inorganic soil components, the use of spectral data has been successfully applied for the prediction of different soil properties, and even soil weathering indices (Vaculíková and Plevová, 2005; Mohanty et al., 2016; Saikia et al., 2016; Xu et al., 2019).

3.4. Modeling weathering indices using FTIR data

Transmission MIR spectra of the soil samples were used as predictors of the weathering indices using RF models. Table 5 indicates the prediction accuracy of the RF models on calibration and validation data for all the indices. From the results obtained with the validation set, it can be observed that not all the models had the same performance: some indices, including MIA ($R^2 = 0.79$, RMSE = 5.06, MAE = 4.25, RPD = 2.29 and PRIQ = 3.16), CIA ($R^2 = 0.75$, RMSE = 2.22, MAE = 1.95, RPD = 2.10 and PRIQ = 3.21) and CIW ($R^2 = 0.71$, RMSE = 3.02, MAE = 2.74, RPD = 1.94 and PRIQ = 2.45) were predicted reasonably well. Other indices like V, were estimated with a moderate success ($R^2 = 0.57$, RMSE = 0.08, MAE = 0.06, RPD = 1.59 and PRIQ = 2.05). Finally, $\text{SiO}_2/\text{R}_2\text{O}_3$ ($R^2 = 0.41$, RMSE = 0.58, MAE = 0.33, RPD = 1.36 and PRIQ = 0.89), CPA ($R^2 = 0.44$, RMSE = 2.57, MAE = 2.14, RPD = 1.40 and PRIQ = 1.56), and WIP ($R^2 = 0.29$, RMSE = 5.07, MAE = 3.78, RPD = 1.24 and PRIQ = 1.09) were poorly predicted. The different performance of the models can also be observed in Fig. 6, which shows the predicted vs. observed values in the validation dataset. MIA, CIA and CIW performed well, although the model slightly underpredicted high values of MIA, while the points appeared distributed along the 1:1 line in the case of CIA and CIW.

The models developed in the present study perform slightly worse than those obtained by Mohanty et al. (2016), which used PLSR for predicting weathering indices based on spectral data too. In their case, the study comprised a larger number of samples and a wider range of values for the weathering indices than in the present study. It can be argued that the performance of the models could be negatively affected by the limited number of samples in this study. However, during the last decade, the number of observations included in machine learning models to predict soil properties have significantly differed among studies, and good performances have been obtained using a number of samples similar to the one in the present study. For instance, Liu et al., (2020) used 44 soil samples to successfully estimate soil heavy metals concentration using PLSR. Also, Massawe et al., (2018) used only 33 soil profiles to predict soil taxa by RF in an area covering 11,600 km^2 . Du et al., (2009) used 56 top soil samples to predict nitrogen, phosphorous, potassium and organic matter content using PLSR. In these studies, the performance of the machine learning models was good despite the low number of samples. On the opposite side, Wang et al., (2018) and Da Silva Chagas et al., (2016) have used more than 300 samples in their predictions and reported R^2 of 0.45 and 0.56, respectively. Thus, the models' prediction capacity and performance in different study areas seems to be mostly affected by other variables like topography or the diversity of parent materials. In our case, it is most probably that the poorer performance of the models compared to those of Mohanty et al. (2016) is caused by the incipient weathering of the soils in the present study, which resulted in narrow range of values for the weathering

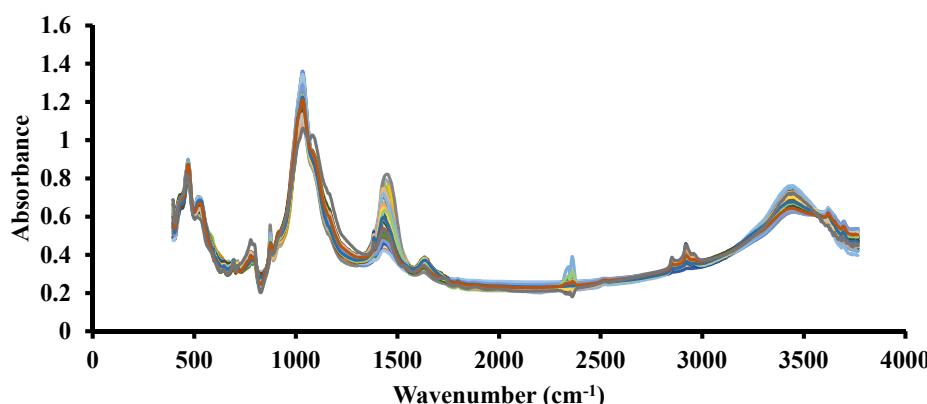


Fig. 5. FTIR spectra of the analyzed soil samples ($n = 45$).

Table 5

Performance of the Random Forest models build to predict weathering indices based on FTIR spectra.

	Calibration					Validation				
	R ²	RMSE	MAE	RPD	RPIQ	R ²	RMSE	MAE	RPD	RPIQ
MIA	0.85	3.76	3.31	2.59	4.19	0.79	5.06	4.25	2.29	3.16
CIA	0.82	2.2	1.88	2.39	3.70	0.75	2.22	1.95	2.10	3.21
CIW	0.78	2.78	2.39	2.16	3.44	0.71	3.02	2.74	1.94	2.45
V	0.85	0.05	0.04	2.06	3.70	0.57	0.08	0.06	1.59	2.05
SiO ₂ /R ₂ O ₃	0.45	0.27	0.21	1.37	2.20	0.41	0.58	0.33	1.36	0.89
CPA	0.42	2.54	2.07	1.34	1.50	0.44	2.57	2.14	1.40	1.56
WIP	0.57	3.33	2.69	1.55	2.03	0.29	5.07	3.78	1.24	1.09

R²: Coefficient of determination; RMSE: Root Mean Square Error; RPD: Ratio of Performance to Deviation; MAE: Mean Absolute Error; RPIQ: Ratio of Performance to Interquartile distance; MIA: Mineralogical Index of Alteration; CIA: Chemical Index of Alteration; CIW: Chemical Index of Weathering; V: Vogt Index; CPA: Chemical Proxy of Alteration; WIP: Weathering Index of Parker.

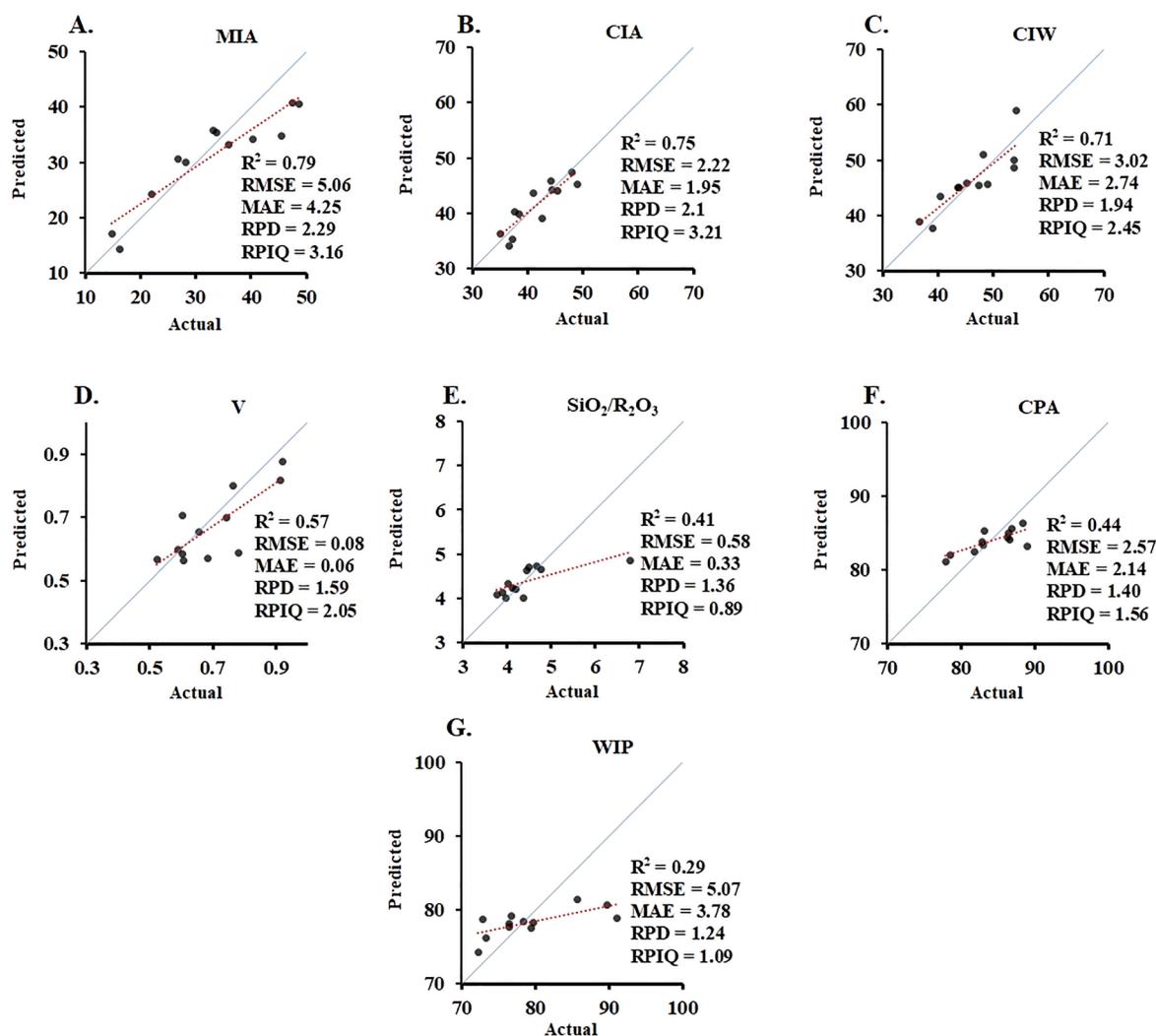


Fig. 6. Scatter plots of the measured against predicted weathering indices (%) based on the RF models. CIA: Chemical Index of Alteration; CIW: Chemical Index of Weathering; WIP: Weathering Index of Parker; CPA: Chemical Proxy of Alteration; MIA: Mineralogical Index of Alteration; V: Vogt Index;

indices. Nevertheless, the results show that even with a limited range of values for the indices, some of them can be predicted reasonably well using FTIR spectra. Similarly, Goydaragh et al. (2019) and Zhang and Hartemink (2018) found acceptable levels of accuracy for predicting different major elemental oxides and weathering indices, respectively, using RF models.

In addition to the capability to predict the weathering indices using spectral data, the RF models identified the most important spectral bands for the estimation of each weathering index. These bands are

presented in Fig. 7 for those indices that showed a good to moderate performance: CIA, MIA, CIW and V. The results showed that the majority of the important bands to estimate CIA, MIA, CIW and V indices were similar: those located at 1300–1400 cm⁻¹, which can be assigned to montmorillonite and nontronite (McConnell and Hossner, 1989), and result from the presence of smectites in these soils; bands at 800–870 cm⁻¹, related to AlMgOH groups resulting from isomorphic substitutions of Al by Mg in the octahedral layer of smectites (Madejová, 2003); and bands at 1000–1100 cm⁻¹, assigned to O—Si—O stretching

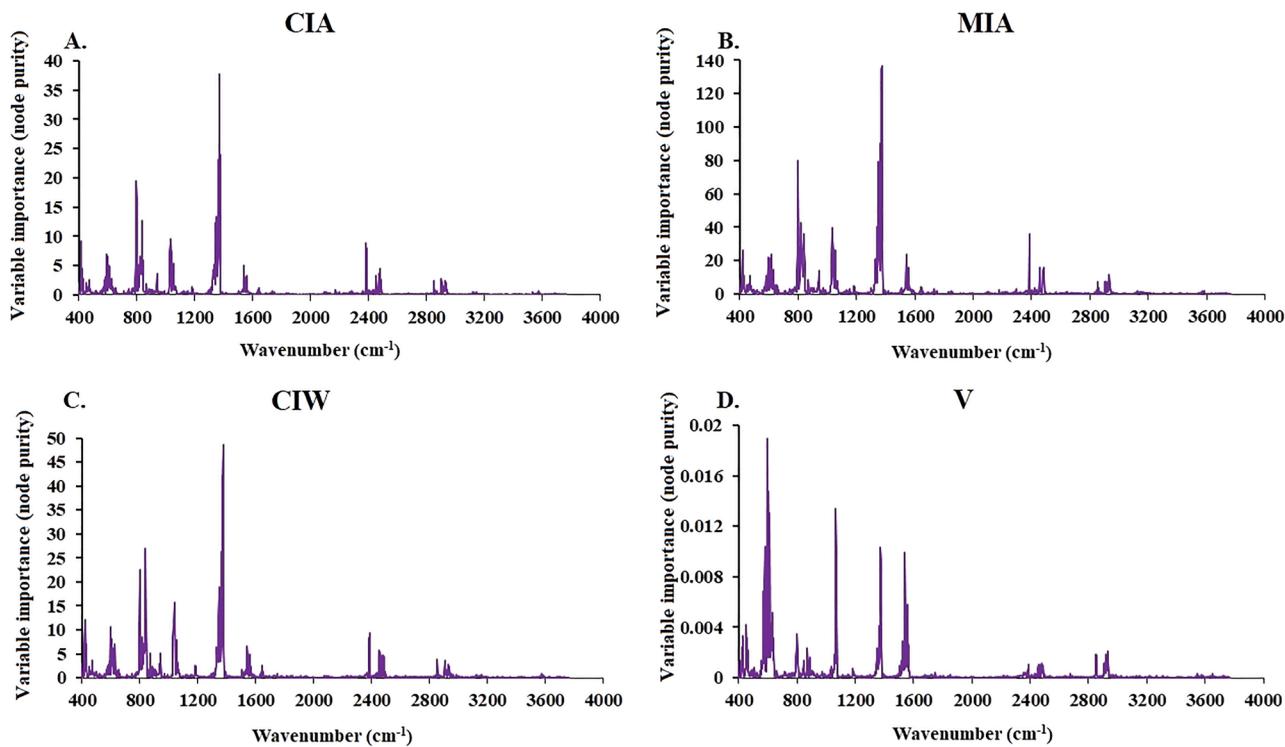


Fig. 7. Relative importance of each FTIR band for prediction of weathering indices. CIA: Chemical Index of Alteration; MIA: Mineralogical Index of Alteration; CIW: Chemical Index of Weathering; V: Vogt Index.

of silicates. Less important bands include those at 2400–2500 cm^{-1} , related to the presence of carbonates; those in the range 400–500 cm^{-1} , denoting the presence of clay minerals such as illite, kaolinite and smectite (Saikia and Parthasarathy, 2010); and bands at 580–650 cm^{-1} , associated with Si—O—Si bending vibrations and the amounts of octahedral Al, Mg and Fe in certain clays (Hayashi and Oinuma, 1965). In the case of the V index, the picture was slightly different, being the most important bands those at 550–600 cm^{-1} (Si—O bending bands), followed by those at 1066 cm^{-1} (O—Si—O stretching), 1373 cm^{-1} (referred above as belonging to montmorillonite and nontronite) and 1540 cm^{-1} . This last band can be related to Si—O stretching (Du and Zhou, 2011) and aromatic C=C and C=O of organic matter (Calderón et al., 2011).

It must be noted that, despite the significant presence of carbonates in these soils (Table 3) and their distinctive FTIR spectral signature with peaks at 1440, 875 and 712 cm^{-1} (Fig. 5), these do not appear as important bands in the prediction of CIA, MIA, CIW and V. This is consistent with the formulation of the indices, which require the removal of carbonates to define the amount of CaO used in their calculation (Table 1), and strengthens the specificity of the models that appear to be directly related to the relative amount of smectites in the inorganic fraction. In arid and semiarid soils, smectites are common clays, and they have a neoformation origin (Wilson, 1999). In light of all these facts, it seems logical to assume that in our soils, CIA, MIA, CIW and V are reflecting mainly the weathering of K-bearing minerals and the relative increase of smectites. In addition to the IR absorption bands of the clay minerals, bands of minor importance corresponding to organic materials could indicate the formation of organo-mineral complexes, more likely to occur with increasing amount of smectites.

In summary, the combination of FTIR spectra and RF models allowed not only to predict weathering indices based on spectral signatures, but also to identify the important bands that were related to the weathering processes that the indices try to describe.

3.5. Application of spectral models to predict weathering indices for unavailable XRF data

The developed RF models were used to estimate weathering indices for 50 top soil samples distributed regularly across the study area, for which no XRF data was available (Fig. 1). The predicted values of CIA, MIA, CIW and V are presented in Fig. 8 superimposed on the map showing the mapping units of the four soil subgroups found in the area. Despite the fact that all of the measured and predicted values showed a narrow range of values resulting from a weak to moderate stage of development of the soils, and that weathering was not intense enough to produce significant variations (Nesbitt and Young, 1982), some interesting trends can be extracted from the spatial pattern of the indices. All four weathering indices showed a similar spatial distribution, with the highest values concentrated in the central part of the study area, while the northernmost and southernmost parts concentrate low values of the indexes. This seems to be related to the different types of soils: Typic Haplocalcids and Typic Haploxerepts showed low values for all indices, pointing to an earlier stage of weathering in these soil units compared to Haplargids and Haplocambids. Considering the important bands for prediction, it seems that the weathering of these soils is related to the alteration of Ca and Na containing minerals and the subsequent leaching of these cations, resulting in soils enriched in Al-containing secondary minerals. In particular, weathering could be related to the neoformation of smectites and an increase in clay content, which is higher in Haplargids and Haplocambids compared to Haplocalcids and Haploxerepts (Table 3).

4. Conclusions

The main goal of this study was to analyze the weathering conditions of the most representative soils of West Azerbaijan, Northern Iran, using chemical indices, and to demonstrate the suitability of FTIR spectra and RF to predict these weathering indices and to identify the major soil components related to soil weathering. The results showed that the soils

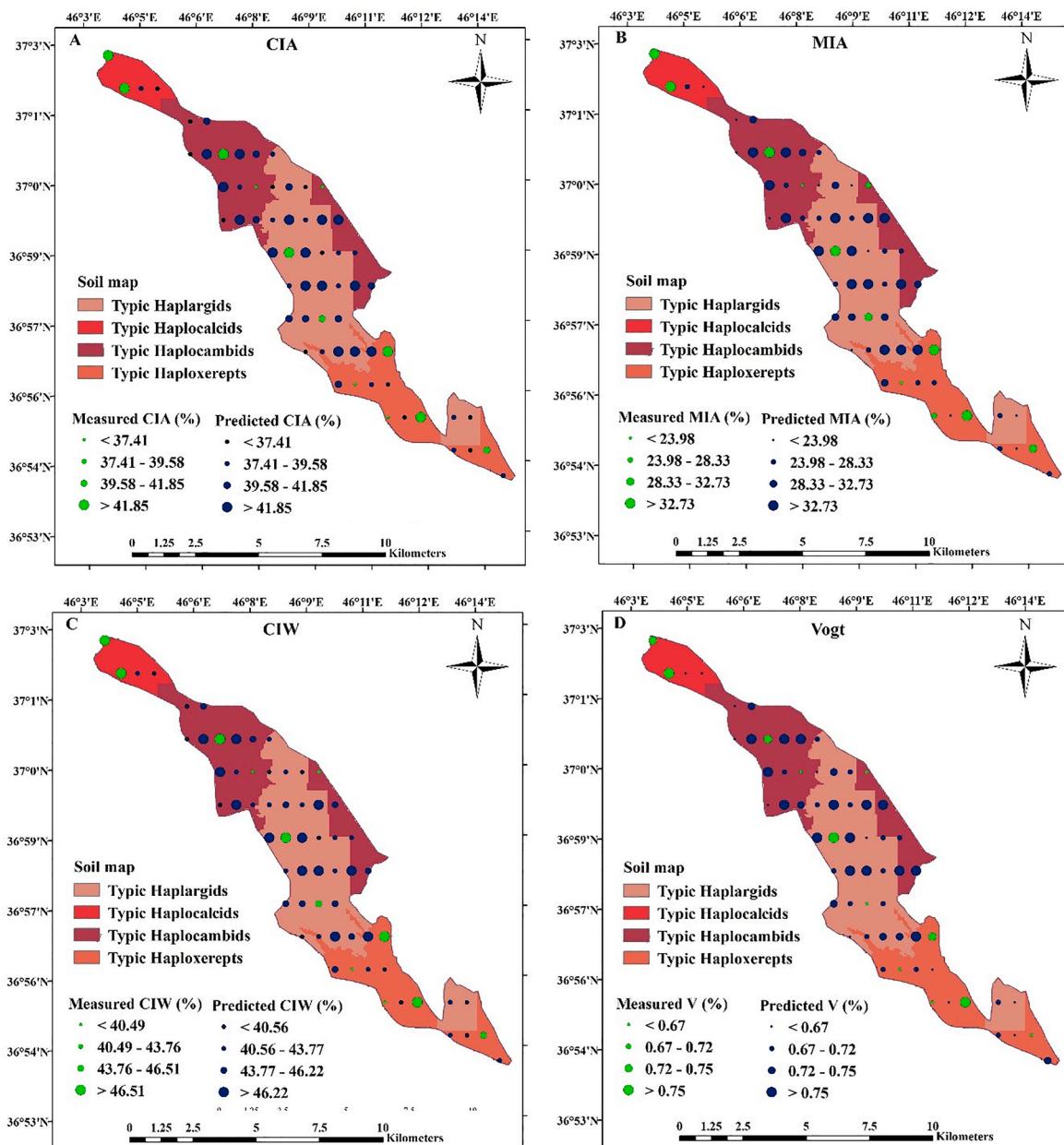


Fig. 8. Predicted values of CIA, MIA, CIW and V in the topsoil of 50 locations across the study area.

in the study area are at an early stage of development and fairly homogeneous in terms of weathering. From all the indices, CIA, CIW, MIA and V were the most suitable to evaluate the weathering intensity, and the combination of FTIR spectra and RF models allowed their prediction with high accuracy. In addition to the predictions, the RF models provided valuable information on the most important weathering processes these indices were describing, which are related to the alteration of K-bearing minerals and the relative increase of smectites in these soils.

Overall, the results indicated the potential of FTIR spectra and RF for the successful prediction of weathering indices when XRF data are not available. FTIR combined with RF is an inexpensive, fast, and feasible technique, and therefore, this type of approach can be beneficial in terms of cost and time to predict weathering intensities in large datasets as long as XRF data is available for a representative set of samples.

Future work should focus on improving these chemometric models in several ways: 1) incorporating covariates like terrain attributes or remote sensing data to predict spatial distribution of weathering indices, and 2) testing different hybrid and deep-learning methods for predicting

weathering indices and evaluating the degree of weathering in this and other study areas.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors deeply thank to University of Tabriz (Tabriz, Iran) and University of A Coruna (A Coruna, Spain) for financial assistance in performing the research.

References

- Abbaslou, H., Abtahi, A., Baghernejad, M., 2013. Effect of weathering and mineralogy on the distribution of major and trace elements (Hormozgan province, Southern Iran). *Int. J. Forest, Soil and Erosion.* 3 (1), 15–25.
- Asghari Saraskanrou, S., Zeinali, B., Mohammadnejad, V., 2017. Analysis physical and chemical properties of soil and morphometric impacts on gully erosion. *Desert* 22 (2), 157–166. <https://doi.org/10.22059/JDESSERT.2017.64179>.
- Bahlburg, H., Dobrzinski, N., 2011. A review of the Chemical Index of Alteration (CIA) and its application to the study of Neoproterozoic glacial deposits and climate transitions. *Geol. Soc. London, Memoirs.* 36 (1), 81–92. <https://doi.org/10.1144/M36.6>.
- Baumann, F., Schmidt, K., Dorfer, C., He, J.S., Scholten, T., Kuhn, P., 2014. Pedogenesis, permafrost, substrate and topography: plot and landscape scale interrelations of weathering processes on the central-eastern Tibetan Plateau. *Geoderma* 226, 300–316. <https://doi.org/10.1016/j.geoderma.2014.02.019>.
- Ben-Dor, E., Banin, A., 1994. Visible and near-infrared (0.4–1.1 μm) analysis of arid and semiarid soils. *Remote Sens Environ.* 48 (3), 261–274. [https://doi.org/10.1016/0034-4257\(94\)90001-9](https://doi.org/10.1016/0034-4257(94)90001-9).
- Bendtsen, C., Bendtsen, M.C., 2011. Package ‘pso’.
- Bouchez, J., Gaillardet, J., Lupker, M., Louvat, P., France-Lanord, C., Maurice, L., Armijos, E., Moquet, J.-S., 2012. Floodplains of large rivers: Weathering reactors or simple silos? *Chem Geo.* 332, 166–184. <https://doi.org/10.1016/j.chemgeo.2012.09.032>.
- Bower, C.A., Reitemeier, R., Fireman, M., 1952. Exchangeable cation analysis of saline and alkali soils. *Soil Sci.* 73 (4), 251–262. <https://doi.org/10.1097/00010694-195204000-00001>.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Brady, N.C., Weil, R., 2013. Nature and properties of soils, the: Pearson new international edition. Pearson Higher Ed.
- Bugge, B., Glaser, B., Hambach, U., Gerasimenko, N., Marković, S., 2011. An evaluation of geochemical weathering indices in loess–paleosol studies. *Quat. Int.* 240 (1–2), 12–21. <https://doi.org/10.1016/j.quaint.2010.07.019>.
- Burrough, P., 1981. Introduction to the study of soils in tropical and subtropical regions. *Soil Sci.* 131 (1), 66. <https://doi.org/10.1097/00010694-198101000-00021>.
- Calderón, F.J., Reeves, J.B., Collins, H.P., Paul, E.A., 2011. Chemical differences in soil organic matter fractions determined by Diffuse-Reflectance Mid-Infrared spectroscopy. *Soil Sci. Soc. Am. J.* 75 (2), 568–579. <https://doi.org/10.2136/sssaj2009.0375>.
- Chapman, S., Campbell, C., Fraser, A., Puri, G., 2001. FTIR spectroscopy of peat in and bordering Scots pine woodland: relationship with chemical and biological properties. *Soil Biol. Biochem.* 33 (9), 1193–1200. [https://doi.org/10.1016/S0038-0717\(01\)00023-2](https://doi.org/10.1016/S0038-0717(01)00023-2).
- Chen, W., Panahi, M., Pourghasemi, H.R., 2017. Performance evaluation of GIS-based new ensemble data mining techniques of adaptive neuro-fuzzy inference system (ANFIS) with genetic algorithm (GA), differential evolution (DE), and particle swarm optimization (PSO) for landslide spatial modelling. *Catena* 157, 310–324. <https://doi.org/10.1016/j.catena.2017.05.034>.
- Collers, R.L., 2000. The geochemistry of shales, siltstones and sandstones of Pennsylvanian-Permian age, Colorado, USA: implications for provenance and metamorphic studies. *Lithos* 51 (3), 181–203. [https://doi.org/10.1016/S0024-4937\(99\)00063-8](https://doi.org/10.1016/S0024-4937(99)00063-8).
- Da Silva Chagas, C., de Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. *Catena* 139, 232–240. <https://doi.org/10.1016/j.catena.2016.01.001>.
- Dengiz, O., Sağlam, M., Özaytekin, H.H., Baskan, O., 2013. Weathering rates and some physico-chemical characteristics of soils developed on a calcic toposequences. *Carpath. J. Earth Env.* 8 (2), 13–24.
- 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.
- Díaz-Uriarte, R., De Andres, S.A., 2006. Gene selection and classification of microarray data using random forest. *J. BMC Bioinformat.* 7 (1), 3. <https://doi.org/10.1186/1471-2105-7-3>.
- Du, C., Zhou, J., 2011. Application of infrared photoacoustic spectroscopy in soil analysis. *Appl. Spectrosc. Rev.* 46 (5), 405–422. <https://doi.org/10.1080/05704928.2011.570837>.
- Du, C., Zhou, J., Wang, H., Chen, X., Zhu, A., Zhang, J., 2009. Determination of soil properties using Fourier transform mid-infrared photoacoustic spectroscopy. *Vib. Spectrosc.* 49 (1), 32–37. <https://doi.org/10.1016/j.vibspec.2008.04.009>.
- Duzgoren-Aydin, N., Aydin, A., Malpas, J., 2002. Re-assessment of chemical weathering indices: case study on pyroclastic rocks of Hong Kong. *Eng. Geol.* 63 (1–2), 99–119. [https://doi.org/10.1016/S0013-7952\(01\)00073-4](https://doi.org/10.1016/S0013-7952(01)00073-4).
- Farid Giglo, B., Arami, A., Akhzari, D., 2014. Assessing the Role of Some Soil Properties on Aggregate Stability Using Path Analysis (Case Study: Silty-Clay-Loam and Clay-Loam Soil from Gully Lands in North West of Iran). *Ecopersia* 2 (2), 513–523.
- Fiantis, D., Nelson, M., Shamshuddin, J., Goh, T., Van Ranst, E., 2010. Determination of the geochemical weathering indices and trace elements content of new volcanic ash deposits from Mt. Talang (West Sumatra) Indonesia. *Eurasian. J. Soil Sci.* 43 (13), 1477–1485. <https://doi.org/10.1134/S1064229310130077>.
- Gee, G.W., Bauder, J.W., 1986. Particle-size analysis.Pp.255-293. In: Klute, A. (ED.), Methods of soil analysis, Part 1: Physical and mineralogical methods. Second ed. Agronomy. 9, 383-411.
- Generu, R., Poggi, J.-M., Tuleau-Malot, C., 2010. Variable selection using random forests. *Pattern Recognit. Lett.* 31 (14), 2225–2236. <https://doi.org/10.1016/j.patrec.2010.03.014>.
- Gile, L.H., Peterson, F.F., Grossman, R.B., 1966. Morphological and genetic sequences of carbonate accumulation in desert soils. *Soil Sci.* 101 (5), 347–360. <https://doi.org/10.1097/00010694-196605000-00001>.
- Goydaragh, M.G., Taghizadeh-Mehrjardi, R., Jafarzadeh, A.A., Triantafilis, J., Lado, M., 2021. Using environmental variables and Fourier Transform Infrared Spectroscopy to predict soil organic carbon. *Catena* 202, 105280. <https://doi.org/10.1016/j.catena.2021.105280>.
- Goydaragh, M.G., Jafarzadeh, A.A., Shahbazi, F., Oustan, S., Taghizadeh-Mehrjardi, R., Lado, M., 2019. Estimation of elemental composition of agricultural soils from West Azerbaijan, Iran, using mid-infrared spectral models. *R. Bras. Eng. Agríc. Ambiental.* 23 (6), 460–466. <https://doi.org/10.1590/1807-1929/agriambi.v23n6p460-466>.
- Guio Blanco, C.M., Brito Gomez, V.M., Crespo, P., Lieb, M., 2018. spatial prediction of soil water retention in a Páramo landscape: Methodological insight into machine learning using random forest. *Geoderma* 316, 100–114. <https://doi.org/10.1016/j.geoderma.2017.12.002>.
- Hajihassani, M., Jaled Armaghani, D., Kalatehjari, R., 2018. Applications of Particle Swarm Optimization in Geotechnical Engineering: A Comprehensive Review. *Geotech. Geol. Eng.* 36 (2), 705–722. <https://doi.org/10.1007/s10706-017-0356-z>.
- Harnois, L., 1988. The CIW index: a new chemical index of weathering. *Sediment. Geol.* 55 (3–4), 319–322. [https://doi.org/10.1016/0037-0738\(88\)90137-6](https://doi.org/10.1016/0037-0738(88)90137-6).
- Hayashi, H., Oinuma, K., 1965. Relationship between infrared absorption spectra in the region Of 450–900 CM-and chemical composition of chlorite. *Am. Min.* 50 (3–4), 476–483.
- Hobley, E., Brereton, A.G., Wilson, B., 2017. Soil charcoal prediction using attenuated total reflectance mid-infrared spectroscopy. *Soil Res.* 55 (1), 86–92. <https://doi.org/10.1071/SR16068>.
- Janik, L.J., Skjemstad, J., Shepherd, K., Spouncer, L., 2007. The prediction of soil carbon fractions using mid-infrared-partial least square analysis. *Aust. J. Soil Res.* 45 (2), 73–81. <https://doi.org/10.1071/SR06083>.
- Jeleńska, M., Hasso-Agopowicz, A., Kądziałko-Hofmokl, M., Sukhorada, A., Bondar, K., Matviushina, Z.H., 2008. Magnetic iron oxides occurring in chernozem soil from Ukraine and Poland as indicators of pedogenic processes. *Stud. Geophys. Geod.* 52 (2), 255–270. <https://doi.org/10.1007/s11200-008-0017-z>.
- Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology Macgraw Hill, New York.
- Khresat, S., Rawajfih, Z., Buck, B., Monger, H., 2004. Geomorphic features and soil formation of arid lands in Northeastern Jordan. *Arch. Agron. Soil Sci.* 50 (6), 607–615. <https://doi.org/10.1080/03650340400005572>.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., 2017. The caret package homepage. URL <http://caret.r-forge.r-project.org>.
- Liland, K., 2016. Extended multiplicative signal correction. package “EMSC”, Date 2016-04-24. Repository CRAN. Available online <https://cran.r-project.org/web/packages/EMSC/index.html>.
- Liland, K.H., Mevik, B.H., 2015. baseline: Baseline Correction of Spectra. URL <https://CRAN.R-project.org/package=baseline>. R package version: 1.2-1.
- Madejová, J., 2003. FTIR techniques in clay mineral studies. *Vib. Spectrosc.* 31 (1), 1–10. [https://doi.org/10.1016/S0924-2031\(02\)00065-6](https://doi.org/10.1016/S0924-2031(02)00065-6).
- Martens, H., Stark, E., 1991. Extended multiplicative signal correction and spectral interference subtraction: New preprocessing methods for near infrared spectroscopy. *J. Pharm. Biomed. Anal.* 9 (8), 625–635. [https://doi.org/10.1016/0731-7085\(91\)80188-F](https://doi.org/10.1016/0731-7085(91)80188-F).
- Massawe, B.H.J., Subburayalu, S.K., Kaaya, A.K., Winowiecki, L., Slater, B.K., 2018. Mapping numerically classified soil taxa in Kilombero valley, Tanzania using machine learning. *Geoderma* 311, 143–148. <https://doi.org/10.1016/j.geoderma.2016.11.020>.
- McConnell, J.S., Hossner, L.R., 1989. X-ray diffraction and infrared spectroscopic studies of adsorbed glyphosate. *J. Agric. Food Chem.* 37 (2), 555–560. <https://doi.org/10.1021/jf00086a061>.
- McLennan, S.M., 1993. Weathering and Global Denudation. *J. Geol.* 101 (2), 295–303. <https://doi.org/10.1086/648222>.
- Mohanty, B., Gupta, A., Das, B.S., 2016. Estimation of weathering indices using spectral reflectance over visible to mid-infrared region. *Geoderma* 266, 111–119. <https://doi.org/10.1016/j.geoderma.2015.11.030>.
- Moradi, N., Rasouli Sadeghiani, M., Sepehr, E., Abdollahi Mandoulakani, B., 2012. Effects of low-molecular-weight organic acids on phosphorus sorption characteristics in some calcareous soils. *Turk J. Agric. For.* 36 (4), 459–468. <https://doi.org/10.3906/tar-1106-38>.
- Müller, C.M., Pejčic, B., Esteban, L., Piane, C.D., Raven, M., Mizaihoff, B., 2014. Infrared attenuated total reflectance spectroscopy: An innovative strategy for analyzing mineral components in energy relevant systems. *Sci. Rep.* 4, 1–11. <https://doi.org/10.1038/srep06764>.
- Nadłonek, W., Bojakowska, I., 2018. Variability of chemical weathering indices in modern sediments of the vistula and Odra rivers (Poland). *Appl. Ecol. Env. Res.* 16 (3), 2453–2473. https://doi.org/10.15666/aeer/1603_24532473.
- Nawar, S., Mouazen, A.M., 2017. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena* 151, 118–129. <https://doi.org/10.1016/j.catena.2016.12.014>.
- Nelson, R.E., 1982. Carbonate and gypsum. In: Page, A.L., Miller, R.H., Keeny, R. (Eds.), *Methods of Soil Analysis. Part 2—Chemical and Microbiological Properties*. American Society of Agronomy, Madison, WI, USA, pp. 181–196 (Madison, WI, PP).
- Nelson, D.W., Sommers, L.E., 1996. Total carbon, organic carbon, and organic matter. Methods of soil analysis part 3—chemical methods(methodsofsoil3): 961-1010.

- Nesbitt, H.W., Young, G.M., 1982. Early Proterozoic climates and plate motions inferred from major element chemistry of lutites. *Nature* 299 (5885), 715–717. <https://doi.org/10.1038/299715a0>.
- Nesbitt, H.W., Young, G., 1984. Prediction of some weathering trends of plutonic and volcanic rocks based on thermodynamic and kinetic considerations. *Geochim. Cosmochim. Acta*. 48 (7), 1523–1534. [https://doi.org/10.1016/0016-7037\(84\)90408-3](https://doi.org/10.1016/0016-7037(84)90408-3).
- Liu, J., Han, J., Xie, J., Wang, H., Tong, W., Ba, Y., 2020. Assessing heavy metal concentrations in earth-cumulic-orthic-anthrosols soils using Vis-NIR spectroscopy transform coupled with chemometrics. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 226, 117639. <https://doi.org/10.1016/j.saa.2019.117639>.
- Oliva, P., Viers, J., Dupré, B., 2003. Chemical weathering in granitic environments. *Chem. Geol.* 202 (3–4), 225–256. <https://doi.org/10.1016/j.chemgeo.2002.08.001>.
- Osat, M., Heidari, A., Eghbal, M.K., Mahmoodi, S., 2016. Impacts of topographic attributes on Soil Taxonomic Classes and weathering indices in a hilly landscape in Northern Iran. *Geoderma* 281, 90–101. <https://doi.org/10.1016/j.geoderma.2016.06.020>.
- Parker, A., 1970. An index of weathering for silicate rocks. *Geol. Mag.* 107(6), 501–504. <https://doi.org/10.1017/S0016756800058581>.
- Pham, B.T., Qi, C., Ho, L.S., Nguyen-Thoi, T., Al-Ansari, N., Nguyen, M.D., Nguyen, H.D., Ly, H.B., Le, H.V., Prakash, I., 2020. A Novel Hybrid Soft Computing Model Using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil. *Sustainability* 12 (6), 2218. <https://doi.org/10.3390/su12062218>.
- Pham, B.T., Prakash, I., Bui, D.T., 2018. Spatial prediction of landslides using hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. *Geomorphology* 303, 256–270. <https://doi.org/10.1016/j.geomorph.2017.12.008>.
- Price, J.R., Velbel, M.A., 2003. Chemical weathering indices applied to weathering profiles developed on heterogeneous felsic metamorphic parent rocks. *Chem. Geol.* 202 (3–4), 397–416. <https://doi.org/10.1016/j.chemgeo.2002.11.001>.
- Qi, H., Paz-Kagan, T., Karniel, A., Jin, X., Li, S., 2018. Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data. *Soil Till Res.* 175, 267–275. <https://doi.org/10.1016/j.still.2017.09.006>.
- Reeves, J.B., 2010. Near-versus mid-infrared diffuse reflectance spectroscopy for soil analysis emphasizing carbon and laboratory versus on-site analysis: where are we and what needs to be done? *Geoderma* 158 (1–2), 3–14. <https://doi.org/10.1016/j.geoderma.2009.04.005>.
- Rial, M., Cortizas, A.M., Rodríguez-Lado, L., 2016. Mapping soil organic carbon content using spectroscopic and environmental data: A case study in acidic soils from NW Spain. *Sci. Total Environ.* 539, 26–35. <https://doi.org/10.1016/j.scitotenv.2015.08.088>.
- Roaldset, E., 1972. Mineralogy and geochemistry of Quaternary clays in the Numedal area, southern Norway. *Nors Geol Tidsskr.* 52, 335–369.
- Roozitalab, M.H., Siadat, H., Farshad, A., 2018. The Soils of Iran. Springer. <https://doi.org/10.1007/978-3-319-69048-3>.
- Saikia, B.J., Parthasarathy, G., 2010. Fourier transform infrared spectroscopic characterization of kaolinite from Assam and Meghalaya. Northeastern India. *J. Mod. Phys.* 1 (04), 206–210. <https://doi.org/10.4236/jmp.2010.14031>.
- Saikia, B.J., Parthasarathy, G., Borah, R., Borthakur, R., 2016. Raman and FTIR spectroscopic evaluation of clay minerals and estimation of metal contaminations in natural deposition of surface sediments from Brahmaputra river. *Int. J. Earth Sci.* 7 (07), 873–883. <https://doi.org/10.4236/ijg.2016.77064>.
- Shao, J., Yang, S., Li, C., 2012. Chemical indices (CIA and WIP) as proxies for integrated chemical weathering in china: Inferences from analysis of fluvial sediments. *Sediment. Geol.* 265–266, 110–120. <https://doi.org/10.1016/j.sedgeo.2012.03.020>.
- Singh, L.P., Parkash, B., Singhvi, A.K., 1998. Evolution of the lower Gangetic Plain landforms and soils in West Bengal. India. *Catena* 33 (2), 75–104.
- Skogholz, J., Liland, K.H., Indahl, U.G., 2019. Preprocessing of spectral data in the extended multiplicative signal correction framework using multiple reference spectra. *J. Raman Spectrosc.* 50 (3), 407–417. <https://doi.org/10.1002/jrs.v50.3.10.1002/jrs.5520>.
- Sorokina, O., Gysev, M., 2018. Weathering reflected by the chemical composition of alluvial soils from the Zeya and Selendzha river valleys. *Sci. China Earth Sci.* 61 (5), 604–613. <https://doi.org/10.1007/s11430-017-9162-5>.
- Soil Survey Staff, 2014. Keys to soil taxonomy. 11th ed. U. S. Department of Agriculture, Natural Recourses Conservation Service.
- Rasaei, Z., Bogaert, 2019. Spatial filtering and Bayesian data fusion for mapping soil properties: A case study combining legacy and remotely sensed data in Iran. *Geoderma*. 344, 50–62. <https://doi.org/10.1016/j.geoderma.2019.02.031>.
- R Core Team., 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing 2015, Vienna, Austria. ISBN 3-900051-07-0. Available: <http://www.R-project.org/> (1.12. 2015).
- Vaculikova, L., Plevova, E., 2005. Identification of clay minerals and micas in sedimentary rocks. *Acta Geodyn. Geomater.* 2 (2), 167–175.
- Viscarra Rossel, R.A., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. *Geoderma* 158 (1–2), 46–54. <https://doi.org/10.1016/j.geoderma.2009.12.025>.
- Viscarra Rossel, R.A., Walvoort, D.J.J., McBratney, A.B., Janik, L.J., Skjemstad, J.O., 2006. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* 131 (1–2), 59–75. <https://doi.org/10.1016/j.geoderma.2005.03.007>.
- Vogt, T., 1927. Sulitelmefletts geologi og petrografi. *Norg. Geol. Unders.* 121, 1–560 (in Norwegian, with English abstract).
- Voicu, G., Bardoux, M., 2002. Geochemical behavior under tropical weathering of the Barama-Mazarum greenstone belt at Omai gold mine, Guiana Shield. *Appl. Geochim.* 17 (3), 321–336. [https://doi.org/10.1016/S0883-2927\(01\)00085-3](https://doi.org/10.1016/S0883-2927(01)00085-3).
- Wang, H., Liu, C., Deng, L., 2018. Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting. *Sci. Rep.* 8 (1), 14285. <https://doi.org/10.1038/s41598-018-32511-1>.
- Wiesmeier, M., Barthold, F., Blank, B., Kögel-Knabner, I., 2011. Digital mapping of soil organic matter stocks using Random Forest modeling in a semi-arid steppe ecosystem. *Plant Soil.* 340 (1–2), 7–24. <https://doi.org/10.1007/s11104-010-0425-z>.
- Wilson, M., 1999. The origin and formation of clay minerals in soils: past, present and future perspectives. *Clay Miner.* 34 (1), 7. <https://doi.org/10.1180/00985599545957>.
- Wilding, L.G., 1985. Soil spatial variability: Its documentation, accommodation and implication to soil surveys. In: Nielsen, D.R., Bouma, J. (Eds.), *Soil Spatial Variability Proceedings of a Workshop of the ISSS and the SSA*, Las Vegas PUDOC, Wageningen. pp. 166–187.
- Wright, M.N., Ziegler, A., 2015. ranger: A fast implementation of random forests for high dimensional data in C++ and R. arXiv preprint arXiv:1508.04409.
- Xu, X., Du, C., Ma, F., Shen, Y., Wu, K., Liang, D., Zhou, J., 2019. Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques. *Geoderma*. 355 (1), 113905 <https://doi.org/10.1016/j.geoderma.2019.113905>.
- Zannah, T.I., Jusop, S., Ishak, C.F., Roslan, I., 2016. FTIR and XRD Analyses of Highly Weathered Ultisols and Oxisols in Peninsular Malaysia. *Asian J Agric Food Sci.* 4 (04), 191–201.
- Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P., 2019. Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma* 338, 445–452. <https://doi.org/10.1016/j.geoderma.2018.09.006>.
- Zhang, Y., Hartemink, A. E., 2018. Digital mapping a soil profile. *Eur. J. Soil Sci.* 70(1), 27–41. <https://doi.org/10.1111/ejss.12699>.
- Zhang, G.L., Pan, J.H., Huang, C.M., Gong, Z.T., 2007. Geochemical features of a soil chronosequence developed on basalt in Hainan Island. China. *Rev. Mex. Cienc. Geol.* 24 (2), 261–269.