Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq

Othman, A. A.; Gloaguen, R.;

Perma-Link to Publication Repository of HZDR:

https://www.hzdr.de/publications/Publ-25508

Accepted Manuscript

Full length article

Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq

Arsalan A. Othman, Richard Gloaguen

Please cite this article as: Othman, A.A., Gloaguen, R., Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq, *Journal of Asian Earth Sciences* (2017), doi: http://dx.doi.org/10.1016/j.jseaes.2017.05.005

Integration of spectral, spatial and morphometric data into lithological mapping: A comparison of different Machine Learning Algorithms in the Kurdistan Region, NE Iraq

Arsalan A. Othman[1,2,3*] Richard Gloaguen[3]

[1]Remote Sensing Group, Institute of Geology, TU Bergakademie Freiberg, B.-von-Cotta-St., 2, D-09599 Freiberg, Germany; E-mail: arsalan-ahmed.othman@student.tu-freiberg.de;

[2]Iraq Geological Survey, Sulaymaniyah Office, Sulaymaniyah, Iraq; E-Mail: arsalan.aljaf@gmail.com;

[3] Helmholtz Zentrum Dresden Rossendorf, Helmholtz Institute Freiberg for Resource Technology, Division of exploration technology, Chemnitzer Str. 40, 09599 Freiberg, Germany; E-Mail: r.gloaguen@hzdr.de;

*Author to whom correspondence should be addressed; e-Mail: arsalan.aljaf@gmail.com, Tel: 017628854718, Fax: +49-373-139-3599.

Abstract

Lithological mapping in mountainous regions is often impeded by limited accessibility due to relief. This study aims to evaluate (1) the performance of different supervised classification approaches using remote sensing data and (2) the use of additional information such as geomorphology. We exemplify the methodology in the Bardi-Zard area in NE Iraq, a part of the Zagros Fold – Thrust Belt, known for its chromite deposits. We highlighted the improvement of remote sensing geological classification by integrating geomorphic features and spatial information in the classification scheme. We performed a Maximum Likelihood (ML) classification method besides two Machine Learning Algorithms (MLA): Support Vector Machine (SVM) and Random Forest (RF) to allow the joint use of geomorphic features, Band Ratio (BR), Principal Component Analysis (PCA), spatial information (spatial coordinates) and multispectral data of the Advanced Space-borne Thermal Emission and Reflection radiometer (ASTER) satellite. The RF algorithm showed reliable results and discriminated serpentinite, talus and terrace deposits, red argillites with conglomerates and limestone, limy conglomerates and limestone conglomerates, tuffites interbedded with basic lavas, limestone and

Metamorphosed limestone and reddish green shales. The best overall accuracy (~80%) was achieved by Random Forest (RF) algorithms in the majority of the sixteen tested combination datasets.

1 Introduction

Remote sensing data allow geologists to make important distinctions between morphological features and rock types, and to aid in prospecting for ores and hydrocarbons (Lillesand et al. 2004; van der Meer et al. 2012). Producing a traditional lithological map in geologically complex areas and poorly accessible terrain is challenging, expensive and time-consuming. Remote sensing data can provide detailed information over large areas, especially in arid and semi-arid regions (Grebby et al. 2010; Grebby et al. 2011; Brown et al., 1998; Pignatti et al., 2009). However, in rugged terrains illuminated under low local solar elevation angles, the topographic effects hinder the remote-sensed classification of the lithological units in mountainous regions (Othman and Gloaguen, 2014, 2013a). In addition, dense vegetation cover such as forest complicates lithological classifications based on reflectance-sensor data (Grebby et al. 2010; Carranza and Hale 2002).

Numerous studies have established the capability of Machine Learning Algorithms (MLAs) to discriminate land cover classes; using multi- or hyper-spectral reflectance data (Foody and Mathur 2004; Ham et al. 2005; Pal and Mather 2005; Lee et al. 2012; Liesenberg and Gloaguen 2013). Several of the classification studies established the potential of Support Vector Machines (SVM) for the discrimination of lithological units (Waske et al. 2009; Liesenberg and Gloaguen 2013; Yu et al. 2012; Othman and Gloaguen 2014; Cracknell and Reading 2014).

On the other hand, morphological features provide the quantitative descriptions of landforms (Goudie and Group, 1990). Several geological applications highlighted the use of morphological indices for the detection and prediction of mass movements (Othman and Gloaguen 2013b; Othman and Gloaguen 2014; Othman and Gloaguen 2015; Borghuis et al. 2007; Chen et al. 2014), neotectonics (Jacques et al. 2014; Grohmann 2004; Garrote et al. 2008; Mahmood and Gloaguen 2012) and lithological mapping (Othman and Gloaguen 2014). In a previous study, Othman and Gloaguen

(2014) used SVMs to combine geomorphic features, textures and multispectral data of the Advanced Space-borne Thermal Emission and Reflection radiometer (ASTER) satellite. They used surface index (Andreani et al., 2014b) to improve the overall accuracy of the lithological classification of the Mawat Ophiolite Complex (in the northern part of the Kurdistan Region, Iraq) from 73% to 79.3% in mountainous regions. Grebby et al. (2010) derived the morphometric variables such as slope, curvature, surface roughness and hypsometric integral from the Digital Terrain Model (DTM) obtained from an airborne Light Detection And Ranging (LiDAR). The derived layers were used to map the lithological units in the upper section of the Troodos Ophiolite (Cyprus) by applying an artificial neural network (ANN) classification method. The resulting lithological map achieved an overall accuracy of 65.4% and provided more details than the existing geological map. Webster et al. (2006a and b) visually identified subtle topographic differences in a LiDAR-derived DEM in conjunction with field observations and utilized the acquired data to map three basaltic flow units of the North Mountain Basalt, in Nova Scotia, eastern Canada. They discriminated these three units based on the contrasts between their morphological characteristics such as slope, smoothness, and relief.

The aim of this study is to compare and assess the efficacy of three classification algorithms (the Maximum Likelihood (ML), the SVM, and the Random Forest (RF)) in order to establish the criteria of choice in terms of input parameters and classifiers for lithological mapping. For this purpose, we selected Bardi-Zard area, located in the Iraqi Zagros Fold – Thrust Belt. We used sixteen dataset combinations obtained from 20 thematic layers to evaluate the algorithms' performances. The thematic layers include ASTER data, Band Ratio (BR), Principal Component Analysis (PCA), spatial information (spatial coordinates), and geomorphic features. The accuracy of the lithological maps is evaluated using in-situ information and an existing lithological map.

2 Location, Climate and Geological Setting

The study area, Bardi-Zard lies within the Kurdistan Region in the northeastern part of Iraq between latitude 36°38'14"N and 36°42'34"N, and between longitude 44°56'11"E and 45°00'42"E. The study area is characterized by rugged

topography with elevations ranging from 1331 to 2554 m (a.s.l.). The study area covers 26.11 km^2, and encompasses part of Erbil Governorate in the Kurdistan Region, NE Iraq (Fig. 1).

According to the climate classification system of Köppen-Geiger (Kottek et al., 2006), the climate of Bardi-Zard area is warm temperate with dry and hot summer (Csa). It is characterized by dry summers and wet winters. The total annual precipitation (776 mm) occurs from October to May. The snowfalls occur during more than 10 days per year in average between November and April. Monthly temperatures range between -0.5 °C in January and 39.8 °C in August (Agro-Meterological, 2006).

The Bardi-Zard area is part of the NW-trending Zagros Fold–Thrust Belt (Fig. 1), which resulted from the collision between the Arabian and the Eurasian plates, and the closure of the Neo-Tethys. The NW–SE shortening of the Zagros Fold–Thrust Belt resulted from the ongoing subduction of the Arabian Plate beneath the Eurasian Plate (Moghadam et al. 2013; McQuarrie et al. 2003; Sissakian 2012). This belt is around 2000 km long, extending from SE of Turkey through Iraq to southern part of Iran. It is a part of the Alpine-Himalayan Mountain Range System (Othman and Gloaguen 2013a; Othman and Gloaguen 2013b; Alavi 1994; Alavi 2004). The Zagros Fold–Thrust Belt in the Iraqi side, consists of two main tectonic zones: (1) the Inner Platform (Stable Shelf); (2) the Outer Platform (Unstable Shelf), which comprises the Mesopotamia Foredeep, the Low Folded Zone, the High Folded Zone, the Imbricated Zone (IZ), and the Zagros Suture Zone (ZSZ) (Jassim and Goff 2006, Agard et al. 2011; Fouad 2012, Othman and Gloaguen 2013a; Lawa and Koyi 2013; Fig. 1). The study area lies within the Zagros Suture Zone (Fig. 1).

The study area consists of different lithostratigraphic units (Fig. 2), which are part of Walash Series. The oldest rock-group dates to Paleocene – Eocene periods and the youngest is Upper Eocene, in addition to overlaying Quaternary sediments. Besides the Quaternary sediments and the intrusive rocks such as serpentinite, the Bardi-Zard area includes three rock-groups.

The lower (calcareous shale) rock-group consists mainly of shales, aphanitic limestones and calcareous conglomerate. At the base of the sequence, thick beds of intercalated gray and gray-green occasionally phyllite-like shales occur. This

4

sequence is followed by interbeddings of gray-green shales and dark-gray laminated fossiliferous limestones, calcareous conglomerates, brownish-yellow marls and calcareous shales. The upper part of this group is composed of thick beds of calcareous conglomerates with subsidiary layers of aphanitic limestones cemented by gray and light gray calcareous substance. An aggregate thickness of the calcareous-shale rock-group exceeds 800 m, the age of the middle and upper parts is Upper Eocene. Shales of the lower part of the sequence appear to be of the Paleocene age (Vasiliev and Pentelkov, 1962).

The middle (volcanic) rock-group consists of basic lavas (basalts), fragmental and volcanic rocks, the motley-colored conglomerate being the most abundant, greywackes, agglomerates, tuffs and tuff conglomerates. At the base of this group, a thick (up to 150–200 m) bed of motley-colored conglomerates occurs, interbedded with greenish-red greywackes, red colored argillites, calcareous polymict sandstones, gray sandy limestones, and rarely, gray nummulitic limestones and brown sideritized dolomites. The volcanic rocks are essentially of basic composition. Tuffaceous rocks have massive, conglomeratic, brecciated structure, psammitic, lithic, crystal-lithic textures. The lavas have massive, vesicular, porphyritic structures, basaltic, microophilitic hyalopilitic textures. The volcanic rocks, which have gray-green and dark olive-green color, suffered an intensive metamorphization grade. The principal rocks are tuffs and basalts. The tuffs consist of fragments of basalts, fragments of angular grains of pyroxenes, and more rarely, feldspars and quartz. In basalts, the fragments are occasionally present as grains of serpentinized olivine. Aggregate thickness of the volcanic rock-group exceeds 1000 m, the age of the middle (volcanic) rock-group is Eocene (Vasiliev and Pentelkov, 1962).

The upper (calcareous-argillite) rock-group includes essentially red colored calcareous argillites and gray massive limestones. Subsidiary brown conglomerates, red siliceous-jasper rocks, reddish-gray sands and light-greenish argillites are also present. At the base of the sequence, a bed of brownish-gray and brown fine-pebbled conglomerates occurs usually, locally, interbedded with red colored siliceous-jasper rocks, highly ferruginous argillites, and more rarely, limestones. This part of the sequence is remarkable for facies changes of rocks and their variable thicknesses (Vasiliev and Pentelkov, 1962).

The proper intrusive rocks are found as inter-formational sheets and sills of altered ultrabasic rocks. The main part of the intrusions occurs as sheets, sills, and more rarely as phacoliths. The age of the ultrabasic intrusions appears to be Eocene. Contacts between serpentinites and country rocks are well defined. The majority of the metamorphosed intrusive rocks are serpentinites. These altered rocks have a complex composition due to the gradational nature of the serpentinization process and heterogeneous composition of primary intrusive rocks (harzburgites, dunites, pyroxenites, etc.). The most abundant variety of the serpentinized rocks form the main part of the massive are granular and aphanitic serpentinites derived from autometasomatic alteration of ultrabasic rocks. The occurrence of chromite veins and lenses is mainly restricted to the sheared varieties of serpentinites. The Quaternary sediments are mainly composed of talus and terrace deposits (Vasiliev and Pentelkov, 1962).

3 Material and methods

3.1 Data Characteristics and Software

ASTER provides more spectral bands in the SWIR than Landsat therefore, increasing its ability to spectrally discriminate rock types on the Earth's surface (Di Tommaso and Rubinstein, 2007; Gad and Kusky, 2007; Guha et al., 2013; Othman and Gloaguen, 2014; Rajendran et al., 2013, 2012). Each scene of ASTER provides 14 bands and covers 60×60 km^2 on land (Abrams et al., 2001). We processed three Visible Near InfraRed (VNIR) bands and six ShortWave InfraRed (SWIR) bands of two ASTER data. These two ASTER level 1A system scenes were cloud-free and were acquired on 27 September 2004. They were provided by the Iraq Geological Survey (GEOSURV-Iraq). The scenes were orthorectified and projected using a WGS84 spheroid and the UTM 38N projection.

We used Environment for Visualizing Images (ENVI) software for data processing (mosaic, subset, algebraic operations, radiometric and atmospheric correction, filtering and classification). The EnMap-Box (Rabe et al., 2014) is freely available and a user-friendly image processing tool. It works as a platform-independent or an add-on in the commercial IDL/ENVI software (Waske et al., 2012). This tool was used within ENVI environment to perform the random forest classification algorithm. Topographic Position Index (TPI) was extracted using the upgraded version of

TecDEM2.2 (Shahzad and Gloaguen, 2011); a MATLAB-based software, which permits the extraction of TPI from DEM (Andreani et al., 2014a). All GIS operations and final map preparations were made using ArcGIS10.1 (ESRI 2011) and QGIS 2.6 (QGIS 2014). Statistical operations were carried out using R-based scripts (RStudio 2016).

3.2 Data Preprocessing

We performed radiometric and atmospheric corrections using the Fast Line-of-Sight Atmospheric Analysis of Spectral Hypercubes (FLAASH) algorithm. The principles of FLAASH algorithm is based on a MODTRAN4 approach to eliminate the molecular and particulate scattering, absorption, and adjacency effects in order to retrieve at-surface reflectance values (ρ) (Felde et al., 2003; Liesenberg and Gloaguen, 2013). The FLAASH algorithm (Abrams and Hook, 2002) was used to extract surface reflectance by implementing Equation (1). Moreover, the Digital Elevation Models (DEMs) of the ASTER data were extracted from ASTER Nadir (N) and backward looking (3B) bands (0.76–0.86 μm). The surface reflectance bands were stacked; the two ASTER scenes were mosaicked and cropped to cover the study area.

$$L_\lambda = (DN - 1) * UCC \tag{1}$$

where, Lλ is the at-sensor radiance measured in W·m−2·sr−1·μm−1, DN is the data digital number, UCC is the unit conversion coefficients

3.3 Principal Component Analysis and Band Ratio

The resulting layers of PCA and BR have been frequently included as additional inputs to aid the spectral separability of lithological units (Sultan et al. 1987; Liberti et al. 2009; Grebby et al. 2011; Cracknell and Reading 2014; Li et al. 2011). We employed these two image enhancement techniques in order to improve lithological discrimination and to increase the accuracy assessment of the classified map. PCA is a multivariate statistical technique that can be applied to remove the redundancy and the irradiance effects of the satellite data (Amer et al., 2010; Gomez et al., 2005). Nine bands of the Principal Components (PCs) were derived from the reflectance of the ASTER data. The PCs are uncorrelated variables and contain each a specific linear combination of ASTER bands (Shokralla et al., 2016). In an effort to improve

lithological discrimination, we selected the first five PCs bands that have ~99% of the total ASTER image variance; and discarded the last four PCs bands. No geological information can be gained from these four components as they represent mostly noise.

Band ratio is an arithmetical operation, where the pixel value of one band is divided by the pixel value of another band (Lillesand and Kiefer, 2004). BR is used to minimize the spectral variation caused by topography or sunlight illumination angle related to grain sizes (Lillesand and Kiefer 2004; Sultan et al. 1987). Three BRs have been performed in this study: (1) ratio of band 1/2 for identifying ferric iron (Fe^{+3}) rocks (Rowan and Mars, 2003); (2) ratio of band 4/3 for detecting ferric oxides rocks (Kanlinowski and Oliver, 2004); and (3) BR 5/4 for recognizing ferrous silicates rocks (Kanlinowski and Oliver, 2004). The purpose of the above mentioned BRs is to help the identification of the rocks rich in ferric iron, ferric oxides and ferrous silicates.

3.4 Geomorphic Indices and Spatial Information

Reflectance bands, PCs and BR data are easy to obtain but in some areas they cannot be used to efficiently discriminate rock units. Vegetation cover and sometimes the similarity in the chemical composition of the rocks make these areas difficult to be mapped using these types of data (Amer et al., 2010). On the other hand, specific rock types have different behavior under erosion, weathering and tectonics. Therefore, we used geomorphic indices as additional layers to enhance the lithological classification and improve the overall mapping performances (Othman and Gloaguen 2014; Kühni and Pfiffner 2001).

The variation in the erosion characteristics for each lithological unit is diversified in surface morphology. Therefore, the lithological units are often correlated with the surface morphology (Grebby et al., 2011; Kühni and Pfiffner, 2001). One of the simple indices that provide information about the surface morphology is the Topographic Position Index (TPI). It measures the variation between elevation at the central pixel (E_c) and the average elevation (E_A) around it, within a predetermined matrix length (M). We calculated the TPI using ASTER DEM by applying Equation (2) (De Reu et al., 2013; Weiss, 2001), which is:

8

$$TPI = E_c - (\frac{1}{nM}\sum_{i \in m} E_i) \qquad (2)$$

Negative TPI values indicate that the central pixel is situated lower than its average surroundings; while positive TPI values indicate that the central pixel is located higher than the average. We implemented a script in the TecDEM toolbox in order to compute the TPI for the studied area. We used a moving window of 130 pixels (~1.95 km) to extract the TPI (Andreani et al., 2014a).

Performance of spatial coordinates as an input layer to improve the lithological units discrimination is an uncommon approach, but it has been applied successfully in a few investigations such as Kovacevic et al. (2009), and Cracknell and Reading (2014). We employed ArcToolBox within ArcGIS 10.1 to create the X and Y coordinates layers as additional layers within the dataset used for lithological classification. First, we converted the DEM band from raster structure to points. Then, we added x and y coordinates to the attribute table of the point shapefile. Finally, we converted the point shapefile to raster structure using x, y coordinates to assign values to the output raster of X, and Y layers, respectively.

We tested sixteen dataset combinations as an input data layers to prepare a lithological map of Bardi-Zard area (Table 1). These dataset combinations contain each a selection of nine surface reflectance bands, five PCAs, three BRs, TPI, and X and Y layers.

3.5 Training Samples Procedure

The previously existing geological map (Vasiliev and Pentelkov, 1962), is used as a source for the training and validating samples selection. It was mapped in detail, at the scale of 1: 20,000. Firstly, the geological map was scanned with 300 dpi and georeferenced to the UTM coordinate system in Zone 38 north. Then, the lithological units of Bardi-Zard area were digitized in GIS. We used the former geological map to select carefully the training samples corresponding to seven lithological units, besides a class for vegetation (i.e. Regions of interest, ROIs). To discriminate the vegetated pixels, we computed the normalized difference vegetation index (NDVI), which is expressed as the difference between the near infrared (NIR) and red bands normalized to the sum of those bands (Rouse et al., 1974).

We selected 2430 pixels as training samples (Fig. 2). It represents ~0.021% of the entire ASTER data. On the final map, the vegetated class was merged with the talus and terrace deposits class, mostly hosting the vegetation. The ROIs were used to calculate the ML and the SVM classification methods directly, while RF required to convert the training samples to class image. We used cross-validation techniques to assess the results.

3.6 Classifiers

In the current study, three supervised classification algorithms have been applied:

3.6.1 Maximum Likelihood (ML)

The ML (Richards, 1961) is a common and traditional multispectral image classification method. Several studies applied the ML classification algorithm to discriminate lithological rock units (Grebby et al., 2011; Rajendran and Nasir, 2014). In the current work, we used the ML algorithm to classify 16 datasets combinations on a per-pixel basis. This method is embedded in the ENVI 4.7 software.

The main deficiencies of this algorithm are that ML has an underlying assumption of a normal distribution of the data within each class and may be biased for unequally sized training classes (Mather et al., 1998). Therefore, testing other approaches, which need no such assumption on data distribution is necessary.

3.6.2 Support Vector Machine (SVM)

The SVM algorithm was proposed by Vapnik (1999). It is one of the suitable techniques, which has been widely used within the remote sensing community (Okujeni et al., 2013; Zhang et al., 2013). It has been successfully used for lithological mapping (Brenning et al., 2012; Heumann, 2011; Huang et al., 2002; Othman and Gloaguen, 2014; Puertas et al., 2013; Yu et al., 2012). This method is embedded in the ENVI 4.7 software. Details of the SVM principle may be found in numerous recent publications such as (Bazi and Melgani, 2006; Yu et al., 2012).

The usual SVM kernel types are Linear, Polynomial, Radial Basis Function, and Sigmoid (Yu et al., 2012). In the current work, we only considered the radial basis function (RBF) kernel. The radial basis function is the most used

kernel type and is versatile when the penalty parameter is at 100 (Yang, 2011). Hence, penalty parameter of 100 and gamma in kernel function 0.05 were used. The penalty parameter is especially significant for non-separable classes.

3.6.3 Random Forest (RF)

RF is a relatively new tool based on transformed high-dimensional spectral feature spaces. It is a supervised classification technique proposed by Breiman (2001). The RF algorithm is used in an increasing number of remote sensing studies (Zhao et al. 2014; Timm and McGarigal 2012; Grinand et al. 2013; Attarchi and Gloaguen 2014; Li et al. 2014), but remains rarely used in lithological classifications (Cracknell and Reading, 2014). Similar to the SVM method, the RF is designed to solve regression and classification estimation problems (Hastie et al., 2009). It also shows great capability of classifying remote sensing data with complex and nonlinear data distributions in high-dimensional spectral feature spaces. The RF algorithm depends on two strategies: bagging or (bootstrap aggregating) and decision tree. The RF uses a bagging algorithm to generate numerous new training data sets, while it uses a decision tree algorithm to identify the probability of each class within each decision tree. In a classification case, RF gets a class vote from each tree, and then classifies them using a majority vote (Hastie et al., 2009; Breiman, 2001). This method is embedded in the EnMap-Box software.

First, we tested the RF algorithm with three numbers of tree types (50, 100 and 200), in addition to two different impurity functions (entropy and Gini coefficient) for the dataset that involves entire layers. Subsequently, the number of trees (100) and impurity function of Gini coefficient were held as constant during the whole dataset combinations tests. Moreover, selected minimum number of samples was 1 in node, the minimum impurity was 0 and the default number of features equaled the square root of the total number of input features for the entire dataset combinations.

3.7 Accuracy Assessment

We estimated the error by calculating the Kappa coefficient (k), and the contingency matrix diagonal. The Kappa coefficient represents the measurement of agreement between the classified set and the true reference data (Cohen, 1960). The contingency matrix delivers the Overall (OA), User's (UA) and Producer's (PA) Accuracies (Congalton,

1991). The overall accuracy is the ratio between the total of correct pixels and the total number of pixels. The user's accuracy is describes the commission errors and the producer's accuracy provide omission errors related to the individual classes. The kappa coefficient considers the entire contingency matrix instead of the contingency matrix diagonal (Grebby et al., 2010; Grinand et al., 2013; Timm and McGarigal, 2012; Yang, 2011). For this purpose, the previously digitized lithological map was converted to raster. Thereafter, 2,000 test samples (ROIs) were selected randomly from a reference (lithological) map to assess the accuracy.

4 Results

4.1 Training Area Statistics

The average values of PCA, BR and TPI as well as ASTER reflectances for each class is shown in Figs. 3 and 4. The average overall reflectance of shale unit (shale, limestone, calcareous conglomerate and slate rocks) is higher than other rock types, while the average overall reflectance of serpentinite rocks is lower than other rock types (Fig. 3A). The shale rocks appear in gray color, while serpentinite rocks appear in greenish blue color in ASTER R3:G2:B1 color composites (Fig. 5A). The average overall reflectance of the talus class (talus and terrace deposits) is higher than the conglomerate class (limy conglomerate and limestone), the limestone class (limestone and metamorphic rocks) and the argillites class (argillites with conglomerates and limestone). Thus, the talus class appears lighter in color than the conglomerate class (limy conglomerates and limestone conglomerates), which appears lighter in color than the limestone (limestone and metamorphic limestone) and the argillites classes (Fig. 5A). The average overall reflectance of the serpentinite rocks is lower than the tuffites interbedded with basic lavas (Fig. 3A). The average overall reflectance of the igneous rocks is lower than the sedimentary rocks. The average overall reflectance of the vegetation class is lower than other classes in the last four SWIR bands. The average overall reflectance of the vegetation class also shows clearly the red edge (Fig. 3A). The vegetation has a very distinguishable red color in ASTER R3:G2:B1 color composites (Fig. 5A). All classes have significant absorptions in band 8 at 2.295-2.365 μm wavelength region of ASTER reflectance data. The serpentinite shows an enhanced absorption in band 6, which is different from other signatures (Fig. 3A). The talus,

12

argillite, and shale classes have high absorptions in band 5 at 2.145-2.185 μm wavelength region of ASTER reflectance data. These absorptions followed by increasing on the reflectance in band 6 (2.185-2.225 μm) for talus, and shale classes. The serpentinite and vegetation classes show significant absorption in band 6 of ASTER reflectance data. The maximum average reflectance registered in the ASTER data for Bardi-Zard area is ~0.34 for the vegetation class at the 0.76-0.86 μm wavelength region of ASTER reflectance data. The minimum average reflectance registered in the ASTER data is ~0.10 for the vegetation class at the 0.63-0.69 μm wavelength region of ASTER reflectance data (Fig. 3A).

In Fig. 3B, the average PCA-1 value represents the difference between all classes within the PCA components. The vegetation class has an average PCA-5 higher than the other classes while the limestones have the lowest average PCA-5 values (Fig. 3B). The vegetation class has average PCA-3 values lower than other classes and average PCA-2 values is higher than the other classes. The vegetation class appears green color in R: PCA-1, G: PCA-2, B: PCA-3 color composites (Fig. 5B). With the exception of vegetation, the average PCA-1 values and the average PCA-3 values are the highest values for all units. The average PCA-1 values of the rock units have the similar behavior as ASTER reflectances. Serpentinite rocks have the lowest average PCA-1 values, while the shale class has average PCA-1 values higher than the other classes followed by talus class (Fig. 3B). The talus, shale, argillite, limestone, and conglomerate classes are higher than the other rock classes in PCA-1, PCA-2 and PCA-3 (Fig. 3B). The average PCA-1 value is higher than the average PCA-2 and PCA-3 values. Moreover, the average PCA-3 value is a bit higher than the average PCA-2 values. Therefore, the talus, shale, argillite, limestone, and conglomerate units appear pink in R: PCA-1, G: PCA-2, B: PCA-3 color composites (Fig. 5B). As a result of the decreasing in the average PCA-1 values and the increasing in the average PCA-3 values for the igneous (serpentinites and tuffites interbedded with basic lavas) rocks (Fig. 3B), these rocks appear violet to blue color in PCA-1, G: PCA-2, B: PCA-3 color composites (Fig. 5B).

Fig. 4 exhibits the average of band ratio of R (b2/b1): G (b4/b3): B (b5/b4), which is computed from the surface reflectance of ASTER data. It shows that the limestone class has average band B2/B1 ratio values higher than the other classes, while the lowest average values of band B2/B1 ratio is vegetation followed by serpentinite classes. In the BR

B4/B3 and B4/B5, the serpentinites show a higher average value than other classes, while the vegetation class has the lowest average values (Fig. 4A). The vegetation class appears in dark color, while shale and talus classes appear in reddish orange color (Fig. 6A).

In Figure 4, the behavior of the argillites and tuffites interbedded with basic lavas classes are very close to each other. They show ramp features (i.e. high in B2/B1, moderate in B4/B3 and low in B4/B5). As a result, the argillites and tuffites interbedded with basic lavas classes appear in coral red color (Fig. 6A). The serpentinite class is high in B4/B3 and moderate in B2/B1 and B4/B5 (Fig. 4A) making the serpentinite class to appear in turquoise color (Fig. 6A). The behavior of the talus, conglomerate, shale and limestone classes are close to each other. They have high value in B2/B1 and in B4/B3, and low value in B4/B5. However, the values of talus, conglomerate, shale and limestone classes in B2/B1 are higher than their values in B4/B3; therefore, in Figure 6A these classes appear in ochre color.

Fig. 6B shows that the TPI ranges between -338.9 and 518.7 m. Fig. 4B shows that all boxplots of classes have upper quartile values of the TPI < 200 m, while the lower quartile values are > -165 m. Fig. 6B and the boxplot interquartile range of TPI in Fig. 4B show that argillite and serpentinite classes are higher than all surrounding classes, while vegetation, talus, limestone, and shale classes are lower than the surrounding classes. The TPI for talus rocks is lower than other lithological classes (Figs. 4B and 6B).

4.2 Lithological Classification of the Bardi-Zard Area

The comparison between the best spatial distribution of the RF, the SVM and the ML classifications is presented in Figure 7, where input combination datasets include ASTER reflectance, PCA, BR, TPI and, X and Y spatial coordinates data. These lithological maps are filtered out of a small amount of noise using a sieve filter. The final classified maps involve seven classes, where the vegetation class is merged with talus and terrace deposits classes for simplification purposes. The largest class in the three classified lithological maps is red argillites with conglomerates and the limestone class with an accumulated surface ranges between 6.074 km^2 of the RF and 7.895 km^2 of the ML (Table 2). Parts of the argillites class (north Krkawz village) is misclassified as talus class for the RF and the SVM maps (Fig. 7B and 7C).

Moreover, the argillites class (east of Krkawz village) is misclassified as the tuffites interbedded with basic lavas class in the SVM map (Fig. 7C). Part of the talus class is misclassified as tuffites north of Kekan village within all classified maps (Fig. 7B, 7C and 7D). In the ML map, the talus class is misclassified as shale class near Rayat and Azadi villages (Fig. 7D). The interbedded tuffite with basic lavas class, southeast of Rayat village is misclassified as argillites class within all classified maps (Fig. 7B, 7C, and 7D). The serpentinite class is well classified within the SVM map (Fig. 7C), where it covers a large area in the RF and the ML maps, especially, near Rayat village (Fig. 7B and 7D). The smallest class is limestone and metamorphic limestone class, which covers between 0.031 km^2 of the ML and 0.258 km^2 of the SVM (Table 2). The limestone class is missing almost in all classified maps (Fig. 7B, 7C, and 7D). The conglomerate class is missing in the southeastern part of the study area within the SVM and the ML maps (Fig. 7C and 7D). All lithological classifications show the major geological units represented by trends and locations of contacts between lithological classes (Fig. 7).

4.3 Classification Accuracy

Figure 8 shows the kappa and overall accuracy for different combination datasets described in Table 1 using the ML, the SVM and the RF methods. The overall accuracy (OA) is evaluated using the contingency error matrix to assess validation. All datasets with the contribution of the RF classifier have the best overall accuracy (~79.66%), while the kappa coefficient is 0.75. The best accuracies for the SVM and the ML are 67.59% and 79.66, respectively, which are obtained by using all datasets (combination). The lowest overall accuracy was achieved for the ML (~53.46%), while the kappa coefficient is ~0.423 (Fig. 8). Kappa coefficient exhibits similar behavior of overall accuracy for all combination datasets. The classifications based on reflectance datasets reached an overall accuracy of 59.22% for the SVM and ~55% for the ML and the RF. The inclusion of the BR to the reflectance dataset resulted in a significant increase in overall accuracy with the RF ~10% and with the SVM and the ML ~5%. Moreover, the addition of spatial coordinates to the reflectance dataset brought a remarkable increase in overall accuracy of the RF ~19% and ML ~9%, while it is modest with the SVM. There is no significant change in overall accuracy for all classifications after combining the TPI, the

15

PCAs and both of the TPI and PCAs layers in reflectance dataset. On the contrary, there is a slight decrease in accuracy for all types of classifier methods. Removing (of) PCA and TPI from all datasets did not affect the overall accuracy of the ML classifier (Fig. 8A), while it decreased the overall accuracy of the SVM from 72.85% to 66.03%, and the overall accuracy of the RF from 79.66% to 76.45%. Removing of spatial coordinates' layers shows the maximum decreasing in all classifiers, followed by BR, then TPI and finally PCAs layers.

Figure 9 illustrates the results of user's (UA) and producer's (PA) classification accuracies from the ML, the SVM and the RF for all datasets' groups that are composed of reflectance, BRs, PCAs, TPI and spatial coordinates. Figure 9A shows that the best producer's accuracy classifications of the ML, the SVM and the RF are 85.44% in argillites class, 85.38% and 91.9% in tuffites and lava classes, respectively. The lowest producer's accuracies of the ML, the SVM and the RF are 7.3%, 17.07% and 46.34% in limestone class, respectively. Figure 9B illustrates that the user's accuracy classification of all datasets' groups reach 100% for the ML within limestone class, 86.67% for the SVM within serpentinite and 91.09% for RF within argillites classes. The conglomerate class shows the lowest user's accuracy for all three classifier methods. It achieved 45.61%, 38.04% and 49.72% for the ML, the SVM and the RF classifier methods, respectively.

5 Discussion

Identifying different lithological units using remote sensing techniques is a difficult task due to the spectral similarity resulted from the analogous chemical compositions of different lithological units. The C-O, hydroxyl and Fe^{+2} forced all classes to strong absorption in band 8 (Abrams et al., 1988; Mars and Rowan, 2010; Othman and Gloaguen, 2014; Rajendran et al., 2013). The trend of shale and serpentinite spectra in bands 7 and 8, which is different than the rest of the spectra, indicates hydroxyl bearing minerals. The talus, argillite and shale classes have high absorptions in band 5 (2.145-2.185 µm) of ASTER reflectance data because they involve quartz (Fig. 3A; Othman and Gloaguen 2014). The shale unit (shale, limestone, calcareous conglomerate and slate rocks) has reflectance higher than other rock types in all bands due to the effect of high reflectance of calcite, dolomite and clay mineral (Fig. 3A). The serpentinite rocks are

darker than other rock types and have reflectance lower than other rock types in visible bands due to olivine minerals content.

Variation in PCA-1 values of the rock units in Fig. 3B is the outcome of the reflectance density of the rock units in Fig. 3A. The low value of BR 2/1 for vegetation is due to the high reflectance of green color. For this reason serpentinite appears with green color, too (Figs. 4A and 6A). The low TPI value of talus and terrace sediments, shale, and conglomerates units corresponds to the regions, which are more easily eroded, while a high TPI value is associated with hard lithology that are slightly eroded, which gives the serpentinite cliff landscape (Figs. 4B and 6B).

Moreover, several band ratios have been established to detect specific rock types, which are suggested by articles for B4+B6/B5 (alunite), B6+B8/B7 (dolomite) (Rowan, and Mars 2003), and for B7+B9/B8 (Carbonate) (Ninomiya 2002). The results were not encouraging. No significant improvement appeared in the accuracy of the produced classified maps. Therefore, the final input dataset did not include these band ratios.

The RF achieved the best accuracy in eight dataset groups, while the SVM achieved best accuracy only in five dataset groups. The ML and the RF were the best classifiers for the rest (three) dataset groups. The best overall accuracy for entire dataset groups goes to the RF classifier (~79.66%), followed by the SVM (~72.85%) then the ML (67.59%). This result coincides with the study of Cracknell and Reading (2014) such that the RF model is better than the SVM and the ML methods for lithological classification.

Generally, spatial neighbors (closer observations) are more related in lithology than those farther (Fischer and Getis, 2009). This issue is clear in the geological units therefore the spatial coordinate layers allowed a significant increase in the discrimination between the lithological units, especially those that have spectral similarities. Moreover, the combination of the spatial coordinates (XY) layers with the different datasets increased the overall accuracy of the RF more than the ML and the SVM algorithms, and the increase in the ML is more than the SVM algorithms. The increase in accuracy for RF is 12.43%, for SVM is 2.79% and for ML is 9.53% when adding spatial coordinates (XY) layer to the different datasets (Table 3). The combination of TPI layer allows a substantial increase in overall accuracy for SVM

with almost all datasets. The overall accuracy of the ML algorithm also increased slightly when we combined the TPI layer in almost all of the dataset combinations (Table 3). There is a wide variation between datasets including TPI classified using RF. The mean of decreasing the accuracy of RF is -2.52%, whereas the mean of decreasing the accuracy of SVM is 1.43% and the ML is 0.41% by adding TPI layer to the different datasets (Table 3). Generally, including TPI to the datasets in conjunction with the RF is inefficient.

The decrease of the user's and producer's accuracies for the conglomerate class is mainly due to the fact that argillite and talus and terrace sediments classes contain gravels and conglomerates. Thus, there is an overlap in the ranges of all dataset groups between these rock units. The decrease in the user's and producer's accuracies for limestone is due to the fact that almost all of the classes contain either carbonate rocks as limestone lenses or cemented by highly calcareous materials. The second reason of mixing limestone with sedimentary rock types is that sedimentary rocks can contain calcareous cement. The Conglomerate and limestone classes consist of several small and elongated patches of outcrops, which contributed in decreasing their accuracy.

This study indicates that combining additional layers (PCA, BR, spatial coordinate, and TPI) has contributed in the increasing of the accuracy of discrimination (Fig. 9). PCA and BR are widely applied as combined layers, while spatial coordinates are very rarely combined, and TPI has never been combined. The most effective combination layers are the spatial coordinates followed by BRs, TPI and then PCAs.

6 Conclusions

This study compared the ML, the SVM, and the RF methods to classify lithological units in Bardi-Zard area, NE Iraq. For this task, we used surface reflectance bands, BRs, PCs, spatial coordinate's layers and TPI. This work shows that the RF method is more accurate than the SVM and the ML methods for lithological classification almost in all dataset groups. The best classifier of the RF method is higher than the best classifiers of the SVM and ML by ~7% and ~12%, respectively. Including spatial coordinate (XY) layers significantly increase the kappa coefficient, and overall, producer's and user's accuracies for all tested datasets. The TPI improved the SVM classification as the overall accuracy

increased when we combined TPI layer. The overall accuracy of the ML algorithm also slightly increased by combined TPI layer with almost all tested datasets, while including the TPI to the datasets in conjunction with the RF is inefficient (Shokralla et al., 2016).

In summary, we show that the including of TPI and spatial coordinate (XY) layers together with the other datasets could be more useful than including each of them separately. This is the case in the ML and the SVM, while it is ineffective with the RF.

Acknowledgments

Author Contributions

Arsalan Ahmed Othman prepared and accomplished the study. He wrote the manuscript. Richard Gloaguen outlined the research, and supported the analysis and discussion. He also followed-up writing of the manuscript in all stages. Both authors have checked and revised the manuscript.

References

ESRI, 2011. ArcGIS Desktop: Release 10.

Abrams, M., Hook, S., 2002. ASTER User Handbook , Version 2. ed. Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA 91109.

Abrams, M., Hook, S., Laboratory, J.P., 2001. ASTER user handbook (Version 2) (Report). Pasadena, CA-91109, USA.

Agard, P., Omrani, J., Jolivet, L., Whitechurch, H., Vrielynck, B., Spakman, W., Monié, P., Meyer, B., Wortel, R., 2011. Zagros orogeny: a subduction-dominated process. Geol. Mag. 148, 692–725. doi:10.1017/S001675681100046X

Agro-Meterological, 2006. Climate data (Dataset). Kurdistan-Iraq.

Alavi, M., 2004. Regional stratigraphy of the Zagros fold-thrust belt of Iran and its proforeland evolution. Am. J. Sci. 304, 1–20.

Alavi, M., 1994. Tectonics of the zagros orogenic belt of iran: new data and interpretations. Tectonophysics 229, 211–238. doi:10.1016/0040-1951(94)90030-2

Amer, R., Kusky, T., Ghulam, A., 2010. Lithological mapping in the Central Eastern Desert of Egypt using ASTER data. J. African Earth Sci. 56, 75–82. doi:http://dx.doi.org/10.1016/j.jafrearsci.2009.06.004

Andreani, L., Gloaguen, R., Shahzad, F., 2014a. A new set of MATLAB functions (TecDEM toolbox) to analyze erosional stages in landscapes and base-level changes in river profiles. Geophys. Res. Abstr. 16.

Andreani, L., Stanek, K.P., Gloaguen, R., Krentz, O., Domínguez-González, L., 2014b. DEM-based Analysis of Interactions Between Tectonics and Landscapes in the Ore Mountains and Eger rift (East Germany and NW Czech Republic). Remote Sens. Submitted.

Attarchi, S., Gloaguen, R., 2014. Classifying Complex Mountainous Forests with L-Band SAR and Landsat Data Integration: A Comparison among Different Machine Learning Methods in the Hyrcanian Forest. Remote Sens. 6, 3624–3647.

Bazi, Y., Melgani, F., 2006. Toward an optimal SVM classification system for hyperspectral remote sensing images. IEEE Trans. Geosci. Remote Sens. 44, 3374–3385.

Borghuis, A.M., Chang, K., Lee, H.Y., 2007. Comparison between automated and manual mapping of typhoontriggered landslides from SPOT-5 imagery. Int. J. Remote Sens. 28, 1843–1856. doi:http://dx.doi.org/10.1080/01431160600935638

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brenning, A., Long, S., Fieguth, P., 2012. Detecting rock glacier flow structures using Gabor filters and IKONOS imagery. Remote Sens. Environ. 125, 227–237. doi:http://dx.doi.org/10.1016/j.rse.2012.07.005

Brown, D.G., Lusch, D.P., Duda, K.A., 1998. Supervised classification of types of glaciated landscapes using digital elevation data. Geomorphology 21, 233–250.

Carranza, E.J.M., Hale, M., 2002. Mineral imaging with Landsat Thermatic Mapper data for hydrothermal alteration mapping in heavily vegetated terrane. Int. J. Remote Sens. 23, 4827–4852.

Chen, W., Li, X., Wang, Y., Chen, G., Liu, S., 2014. Forested landslide detection using LiDAR data and the random forest algorithm: A case study of the Three Gorges, China. Remote Sens. Environ. 152, 291–301. doi:http://dx.doi.org/10.1016/j.rse.2014.07.004

Cohen, J., 1960. A coefficient of agreement of nominal scales. Psychol. Meas. 2, 37−46.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sens. Environ. 37, 35–46. doi:http://dx.doi.org/10.1016/0034-4257(91)90048-B

Cracknell, M.J., Reading, A.M., 2014. Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. Comput. Geosci. 63, 22–33. doi:http://dx.doi.org/10.1016/j.cageo.2013.10.008

De Reu, J., Bourgeois, J., Bats, M., Zwertvaegher, A., Gelorini, V., De Smedt, P., Chu, W., Antrop, M., De Maeyer, P., Finke, P., Van Meirvenne, M., Verniers, J., Crombé, P., 2013. Application of the topographic position index to heterogeneous landscapes. Geomorphology 186, 39–49.

Di Tommaso, I., Rubinstein, N., 2007. Hydrothermal alteration mapping using ASTER data in the Infiernillo porphyry deposit, Argentina. Ore Geol. Rev. 32, 275–290. doi:10.1016/j.oregeorev.2006.05.004

Di Tommaso, I., Rubinstein, N., 2007. Hydrothermal alteration mapping using ASTER data in the Infiernillo porphyry deposit, Argentina. Ore Geol. Rev. 32, 275–290. doi:http://dx.doi.org/10.1016/j.oregeorev.2006.05.004

ENVI, 2013. Exelis Visual Information Solutions.

Felde, G.W., Anderson, G.P., Cooley, T.W., Matthew, M.W., adler-Golden, S.M., Berk, A., Lee, J., 2003. Analysis of Hyperion data with the FLAASH atmospheric correction algorithm, in: Geoscience and Remote Sensing Symposium, 2003. IGARSS '03. Proceedings. 2003 IEEE International. pp. 90–92 vol.1. doi:10.1109/IGARSS.2003.1293688

Fischer, M.M., Getis, A., 2009. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer Berlin Heidelberg.

Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1335–1343.

Fouad, S.F., 2010. Tectonic Map of Iraq. Scale 1:1,000,000, 3rd, ed.; GEOSURV: Baghdad, Iraq.

Fouad, S.F.A., Nasir, W.A.A., 2009. Tectonic and structural evolution. In: Geology of Al-Jazira Area. Iraqi Bull. Geol. Min. 3, 33–48.

Gad, S., Kusky, T., 2007. ASTER spectral ratioing for lithological mapping in the Arabian–Nubian shield, the Neoproterozoic Wadi Kid area, Sinai, Egypt. Gondwana Res. 11, 326–335. doi:http://dx.doi.org/10.1016/j.gr.2006.02.010

Garrote, J., Garzón Heydt, G., Cox, R.T., 2008. Multi-stream order analyses in basin asymmetry: A tool to discriminate the influence of neotectonics in fluvial landscape development (Madrid Basin, Central Spain). Geomorphology 102, 130–144. doi:http://dx.doi.org/10.1016/j.geomorph.2007.07.023

Gomez, C., Delacourt, C., Allemand, P., Ledru, P., Wackerle, R., 2005. Using ASTER remote sensing data set for geological mapping, in Namibia. Phys. Chem. Earth, Parts A/B/C 30, 97–108. doi:http://dx.doi.org/10.1016/j.pce.2004.08.042

Goudie, A., Group, B.G.R., 1990. Geomorphological Techniques. Routledge.

Grebby, S., Cunningham, D., Naden, J., Tansey, K., 2010. Lithological mapping of the Troodos ophiolite, Cyprus, using airborne LiDAR topographic data. Remote Sens. Environ. 114, 713–724.

Grebby, S., Naden, J., Cunningham, D., Tansey, K., 2011. Integrating airborne multispectral imagery and airborne LiDAR data for enhanced lithological mapping in vegetated terrain. Remote Sens. Environ. 115, 214–226.

Grinand, C., Rakotomalala, F., Gond, V., Vaudry, R., Bernoux, M., Vieilledent, G., 2013. Estimating deforestation in tropical humid and dry forests in Madagascar from 2000 to 2010 using multi-date Landsat satellite images and the random forests classifier. Remote Sens. Environ. 139, 68–80. doi:http://dx.doi.org/10.1016/j.rse.2013.07.008

Grohmann, C.H., 2004. Morphometric analysis in geographic information systems: applications of free software GRASS and R. Comput. Geosci. 30, 1055–1067. doi:http://dx.doi.org/10.1016/j.cageo.2004.08.002

Guha, A., Singh, V.K., Parveen, R., Kumar, K.V., Jeyaseelan, A.T., Dhanamjaya Rao, E.N., 2013. Analysis of ASTER data for mapping bauxite rich pockets within high altitude lateritic bauxite, Jharkhand, India. Int. J. Appl. Earth Obs. Geoinf. 21, 184–194. doi:http://dx.doi.org/10.1016/j.jag.2012.08.003

Ham, J., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification of hyperspectral data. IEEE Trans. Geosci. Remote Sens. 43, 492–501. doi:10.1109/TGRS.2004.842481

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer.

Heumann, B.W., 2011. An object-based classification of Mangroves using a hybrid decision tree—support vector machine approach. Remote Sens. 3, 2440–2460.

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote Sens. 23, 725–749. doi:10.1080/01431160110040323

Jacques, P.D., Salvador, E.D., Machado, R., Grohmann, C.H., Nummer, A.R., 2014. Application of morphometry in neotectonic studies at the eastern edge of the Paraná Basin, Santa Catarina State, Brazil. Geomorphology 213, 13–23. doi:http://dx.doi.org/10.1016/j.geomorph.2013.12.037

Jassim, S.Z., Goff, J.C., 2006. Geology of Iraq. Geological Society of London.

Kanlinowski, A., Oliver, S., 2004. ASTER Mineral Index Processing (Report). Remote Sensing Application Geoscience

Australia , Australia.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Zeitschrift 15, 259–263. doi:10.1127/0941-2948/2006/0130

Kovacevic, M., Bajat, B., Trivic, B., Pavlovic, R., 2009. Geological Units Classification of Multispectral Images by Using Support Vector Machines, in: Intelligent Networking and Collaborative Systems, 2009. INCOS '09. International Conference on. pp. 267–272. doi:10.1109/INCOS.2009.44

Kühni, A., Pfiffner, O.A., 2001. The relief of the Swiss Alps and adjacent areas and its relation to lithology and structure: topographic analysis from a 250-m DEM. Geomorphology 41, 285–307. doi:http://dx.doi.org/10.1016/S0169-555X(01)00060-5

Lawa, H.K. and Koyi, A.I., 2013. Tectono-stratigraphic evolution of the NW segment of the Zagros Fold-Thrust Belt, Kurdistan, NE Iraq. J. Pet. Geol. 36, 75–96.

Lee, S., Song, K.Y., Oh, H.J., Choi, J., 2012. Detection of landslides using web-based aerial photographs and landslide susceptibility mapping using geospatial analysis. Int. J. Remote Sens. 33, 4937–4966.

Li, N., Frei, M., Altermann, W., 2011. Textural and knowledge-based lithological classification of remote sensing data in Southwestern Prieska sub-basin, Transvaal Supergroup, South Africa. J. African Earth Sci. 60, 237–246. doi:http://dx.doi.org/10.1016/j.jafrearsci.2011.03.002

Liberti, M., Simoniello, T., Carone, M.T., Coppola, R., D'Emilio, M., Macchiato, M., 2009. Mapping badland areas using LANDSAT TM/ETM satellite imagery and morphological data. Geomorphology 106, 333–343. doi:http://dx.doi.org/10.1016/j.geomorph.2008.11.012

Liesenberg, V., Gloaguen, R., 2013. Evaluating SAR polarization modes at L-band for forest classification purposes in Eastern Amazon, Brazil. Int. J. Appl. Earth Obs. Geoinf. 21, 122–135. doi:http://dx.doi.org/10.1016/j.jag.2012.08.016

Lillesand, T.M., Kiefer, R.W., 2004. Remote Sensing and Image Interpretation, Fifth. ed. John Wiley & Sons, Inc.,

United States of America.

Mahmood, S.A., Gloaguen, R., 2012. Appraisal of active tectonics in Hindu Kush: Insights from DEM derived geomorphic indices and drainage analysis. Geosci. Front. 3, 407–428.

Mars, J.C., Rowan, L.C., 2010. Spectral assessment of new ASTER SWIR surface reflectance data products for spectroscopic mapping of rocks and minerals. Remote Sens. Environ. 114, 2011–2025. doi:http://dx.doi.org/10.1016/j.rse.2010.04.008

Mather, P.M., Tso, B., Koch, M., 1998. An evaluation of Landsat TM spectral data and SAR-derived textural information for lithological discrimination in the Red Sea Hills, Sudan. Int. J. Remote Sens. 19, 587–604.

McQuarrie, N., Stock, J.M., Verdel, C., Wernicke, B.P., 2003. Cenozoic evolution of Neotethys and implications for the causes of plate motions. Geophys. Res. Lett. 30, 2036. doi:10.1029/2003gl017992

Moghadam, H.S., Stern, R.J., Chiaradia, M., Rahgoshay, M., 2013. Geochemistry and tectonic evolution of the Late Cretaceous Gogher–Baft ophiolite, central Iran. Lithos 168–169, 33–47. doi:http://dx.doi.org/10.1016/j.lithos.2013.01.013

Ninomiya, Y., 2002. Mapping quartz, carbonate minerals, and mafic-ultramafic rocks using remotely sensed multispectral thermal infrared ASTER data. Proceedings of SPIE. The International Society for Optical Engineering Maldague, XXIV.

Okujeni, A., van der Linden, S., Tits, L., Somers, B., Hostert, P., 2013. Support vector regression and synthetically mixed training data for quantifying urban land cover. Remote Sens. Environ. 137, 184–197. doi:http://dx.doi.org/10.1016/j.rse.2013.06.007

Othman, A.A., Gloaguen, R., 2014. Improving lithological mapping by SVM classification of spectral and morphological features: The discovery of a new chromite body in the Mawat ophiolite complex (Kurdistan, NE Iraq). Remote Sens. 6, 6867–6896. doi:10.3390/rs6086867

Othman, A.A., Gloaguen, R., 2013a. Automatic extraction and size distribution of landslides in kurdistan region, NE

Iraq. Remote Sens. 5, 2389–2410. doi:10.3390/rs5052389

Othman, A.A., Gloaguen, R., 2013b. River Courses Affected by Landslides and Implications for Hazard Assessment: A High Resolution Remote Sensing Case Study in NE Iraq–W Iran. Remote Sens. 5, 1024–1044. doi:10.3390/rs5031024

Othman, A., Gloaguen, R., 2015. Landslide susceptibility mapping in Mawat area, Kurdistan Region, NE Iraq: a comparison of different statistical models. Remote Sens. Under revi.

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011. doi:10.1080/01431160512331314083

Pignatti, S., Cavalli, R.M., Cuomo, V., Fusilli, L., Pascucci, S., Poscolieri, M., Santini, F., 2009. Evaluating Hyperion capability for land cover mapping in a fragmented ecosystem: Pollino National Park, Italy. Remote Sens. Environ. 113, 622–634.

Puertas, O.L., Brenning, A., Meza, F.J., 2013. Balancing misclassification errors of land cover classification maps using support vector machines and Landsat imagery in the Maipo river basin (Central Chile, 1975–2010). Remote Sens. Environ. 137, 112–123. doi:http://dx.doi.org/10.1016/j.rse.2013.06.003

QGIS-Team, 2014. Qgis geographic information system.

Rabe, A., Jakimow, B., Held, M., van der Linden, S., Hostert, P., 2014. EnMAP-Box.

Rajendran, S., al-Khirbash, S., Pracejus, B., Nasir, S., Al-Abri, A.H., Kusky, T.M., Ghulam, A., 2012. ASTER detection of chromite bearing mineralized zones in Semail Ophiolite Massifs of the northern Oman Mountains: Exploration strategy. Ore Geol. Rev. 44, 121–135. doi:http://dx.doi.org/10.1016/j.oregeorev.2011.09.010

Rajendran, S., Nasir, S., 2014. Mapping of high pressure metamorphics in the As Sifah region, NE Oman using ASTER data. Adv. Sp. Res. doi:http://dx.doi.org/10.1016/j.asr.2014.11.026

Rajendran, S., Nasir, S., Kusky, T.M., Ghulam, A., Gabr, S., El-Ghali, M.A.K., 2013. Detection of hydrothermal mineralized zones associated with listwaenites in Central Oman using ASTER data. Ore Geol. Rev. 53, 470–488.

doi:http://dx.doi.org/10.1016/j.oregeorev.2013.02.008

Richards, F.S.G., 1961. A Method of Maximum-Likelihood Estimation. J. R. Stat. Soc. Ser. B 23, 469–475. doi:10.2307/2984037

Rouse, J.W., Haas, R.H., Schelle, J.A., Deering, D.W., 1974. Monitoring the vernal advancement or retrogradation of natural vegetation (Report), Type III. Greenbelt, MD.

Rowan, L.C., Mars, J.C., 2003. Lithologic mapping in the Mountain Pass, California area using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) data. Remote Sens. Environ. 84, 350–366. doi:http://dx.doi.org/10.1016/S0034-4257(02)00127-X

RStudio, 2016. Integrated development environment for R (Version 0.99.893) [Computer software]. Boston, MA.

Shahzad, F., Gloaguen, R., 2011. TecDEM: A MATLAB based toolbox for tectonic geomorphology, Part 2: Surface dynamics and basin analysis. Comput. Geosci. 37, 261–271. doi:10.1016/j.cageo.2010.06.009

Shokralla, S., Morelli, J.E., Krause, T.W., 2016. Principal Components Analysis of Multifrequency Eddy Current Data Used to Measure Pressure Tube to Calandria Tube Gap. IEEE Sens. J. 16, 3147–3154. doi:10.1109/JSEN.2016.2529721

Sissakian, V.K., 2013. Geological evolution of the Iraqi Mesopotamia Foredeep, inner platform and near surroundings of the Arabian Plate. J. Asian Earth Sci. 72, 152–163. doi:http://dx.doi.org/10.1016/j.jseaes.2012.09.032

Sissakian, V.K., 2012. Geological evolution of the Iraqi Mesopotamia Foredeep, inner platform and near surroundings of the Arabian Plate. J. Asian Earth Sci.

Sultan, M., Arvidson, R.E., Sturchio, N.C., Guinness, E.A., 1987. Lithologic mapping in arid regions with Landsat thematic mapper data: Meatiq dome, Egypt. Geol. Soc. Am. Bull. 99, 748–762.

Timm, B.C., McGarigal, K., 2012. Fine-scale remotely-sensed cover mapping of coastal dune and salt marsh ecosystems at Cape Cod National Seashore using Random Forests. Remote Sens. Environ. 127, 106–117. doi:http://dx.doi.org/10.1016/j.rse.2012.08.033

van der Meer, F.D., van der Werff, H.M.A., van Ruitenbeek, F.J.A., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., Smeth, J.B. de, Woldai, T., 2012. Multi- and hyperspectral geologic remote sensing: A review. Int. J. Appl. Earth Obs. Geoinf. 14, 112–128. doi:http://dx.doi.org/10.1016/j.jag.2011.08.002

Vapnik, V.N., 1999. The nature of statistical learning theory, Second Edi. ed. Springer, New York, NY, USA.

Vasiliev, M.M., Pentelkov, V.G., 1962. Prospecting-exploration of the Bard-i-Zard chromite occurrence and adjacent areas in 1961. (Report). GEOSURV, Baghdad, Iraq.

Waske, B., Benediktsson, J.A., Árnason, K., Sveinsson, J.R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. Can. J. Remote Sens. 35, S106–S116. doi:10.5589/m09-018

Waske, B., van der Linden, S., Oldenburg, C., Jakimow, B., Rabe, A., Hostert, P., 2012. imageRF – A user-oriented implementation for remote sensing image analysis with Random Forests. Environ. Model. Softw. 35, 192–193. doi:http://dx.doi.org/10.1016/j.envsoft.2012.01.014

Webster, T.L., Murphy, J.B., Gosse, J.C., 2006a. Mapping subtle structures with light detection and ranging (LIDAR): flow units and phreatomagmatic rootless cones in the North Mountain Basalt, Nova Scotia. Can. J. Earth Sci. 43, 157–176. doi:10.1139/e05-099

Webster, T.L., Murphy, J.B., Gosse, J.C., Spooner, I., 2006b. The application of lidar-derived digital elevation model analysis to geological mapping: an example from the Fundy Basin, Nova Scotia, Canada. Can. J. Remote Sens. 32, 173–193. doi:10.5589/m06-017

Weiss, A.D., 2001. Topographic position and landforms analysis, in: ESRI Users Conference. San Diego, CA.

Yang, X., 2011. Parameterizing support vector machines for land cover classification. Photogramm. Eng. Remote Sensing 77, 27–38.

Yu, L., Porwal, A., Holden, E.-J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. Comput. Geosci. 45, 229–239. doi:http://dx.doi.org/10.1016/j.cageo.2011.11.019

Zhang, J., Lin, X., Ning, X., 2013. SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas. Remote Sens. 5, 3749–3775.

Zhao, L., Yang, J., Li, P., Zhang, L., 2014. Seasonal inundation monitoring and vegetation pattern mapping of the Erguna floodplain by means of a RADARSAT-2 fully polarimetric time series. Remote Sens. Environ. 152, 426–440. doi:http://dx.doi.org/10.1016/j.rse.2014.06.026

Figure Captions

Fig. 1 Location map of the study area in the Imbricated Zone (IZ) of the Iraqi part of the Zagros Suture Zone (modified after Jassim and Goff 2006; Sissakian 2013; Fouad 2010).

Fig. 2 Geological map overlaying the DEM from ASTER of the Bardi-Zard area that indicated in Figure 1 (Vasiliev and Pentelkov, 1962).

Fig. 3 Average of (A) surface reflectance; (B) PCA components, based on training samples selected from ASTER data acquired on 27 September 2004 and for seven lithological units and vegetation.

Fig. 4 (A) Average of three BRs computed from surface reflectance of ASTER data; (B) boxplot of the TPI data based on training samples for the eight discriminated classes.

Fig. 5 (A) ASTER data R3:G2:B1, acquired on 27 September 2004; (B) Principal component analysis PC1:PC2:PC3 of the study area.

Fig. 6 (A) Band ratio of ASTER data R band2/band1:G band4/band3:B band5/band4, acquired on 27 September 2004; (B) Topographic Positional Index (TPI) of the study area.

Fig. 7 (A) reference lithological map, (B) RF, (C) SVM and (D) ML lithological discriminated classes from combined layers (i.e., ASTER reflectance, PCA, BR, TPI and, X and Y spatial coordinates data).

Fig. 8 (A) Overall accuracy and (B) kappa for different combination datasets using ML, SVM and RF methods.

Fig. 9 Classification accuracy for all datasets using Ml, SVM and RF methods (A) Producer's accuracy and (B) User's accuracy.

Table Captions

Table 1 Input proposed classification datasets groups.

Table 2 Areas of classes in Km^2 for the ML, SVM and RF classified maps and previous lithological map.

Table 3 variations in overall accuracy of ML, SVM and RF classifications resulted from addition spatial coordinates and/or TPI.
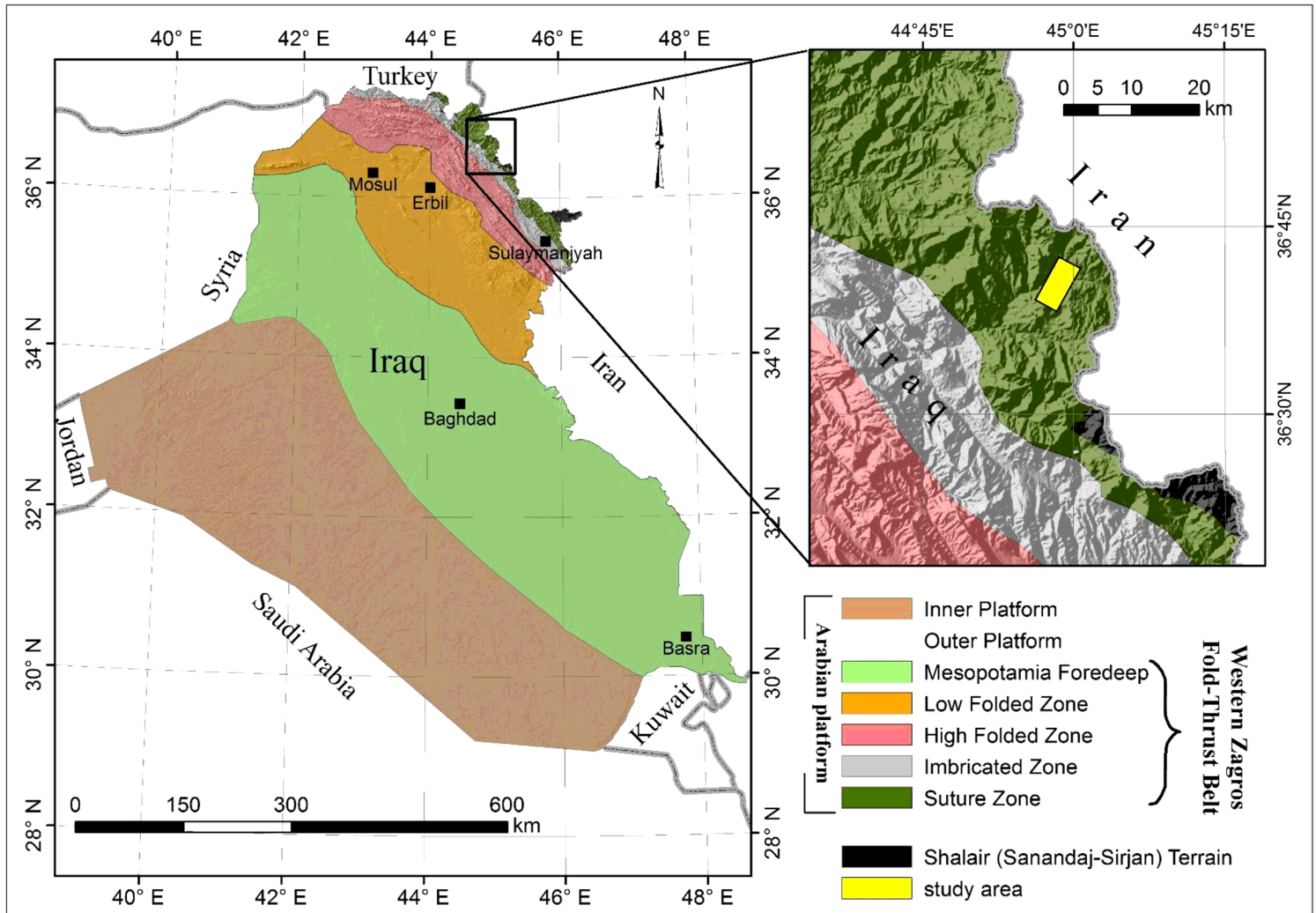
Fig. 1 Location map of the study area in the Imbricated Zone (IZ) of the Iraqi part of the Zagros Suture Zone (modified after Jassim and Goff 2006; Sissakian 2013; Fouad 2010).
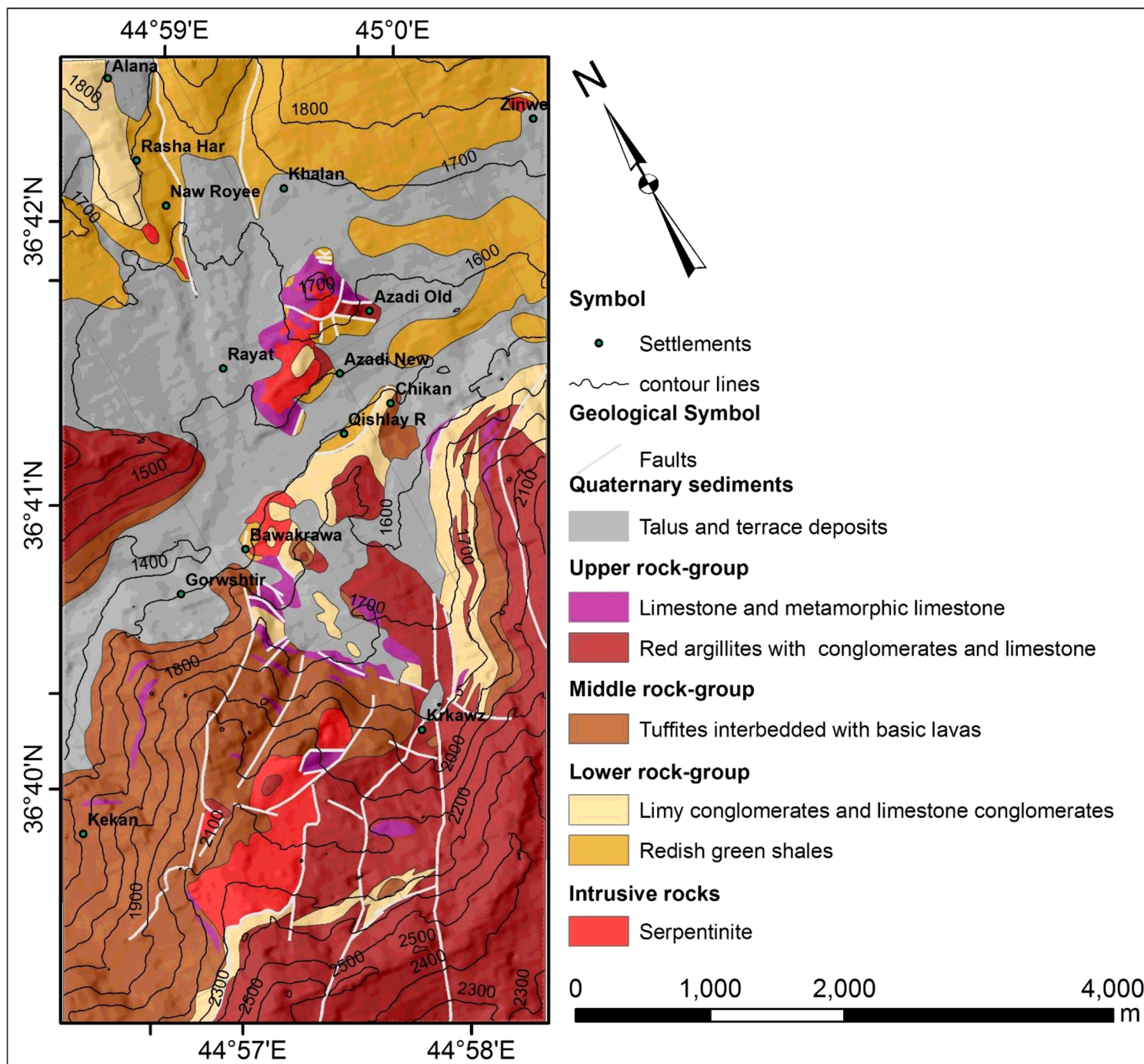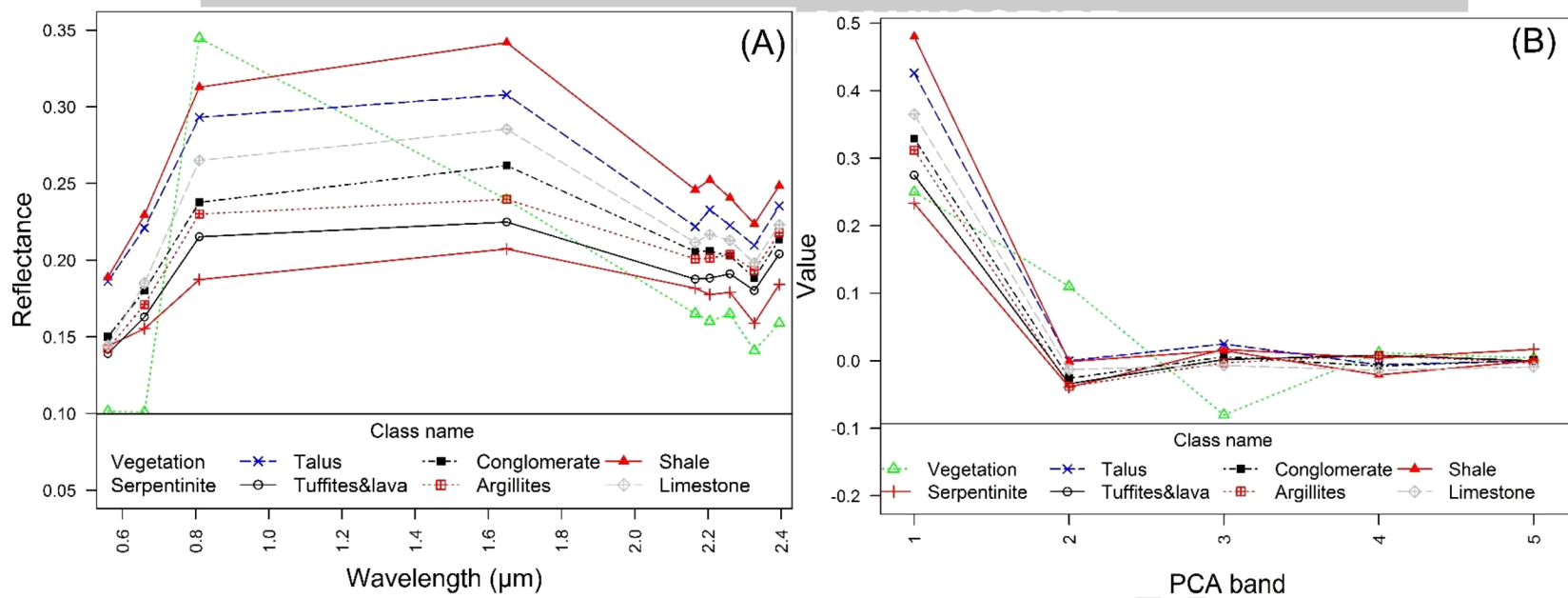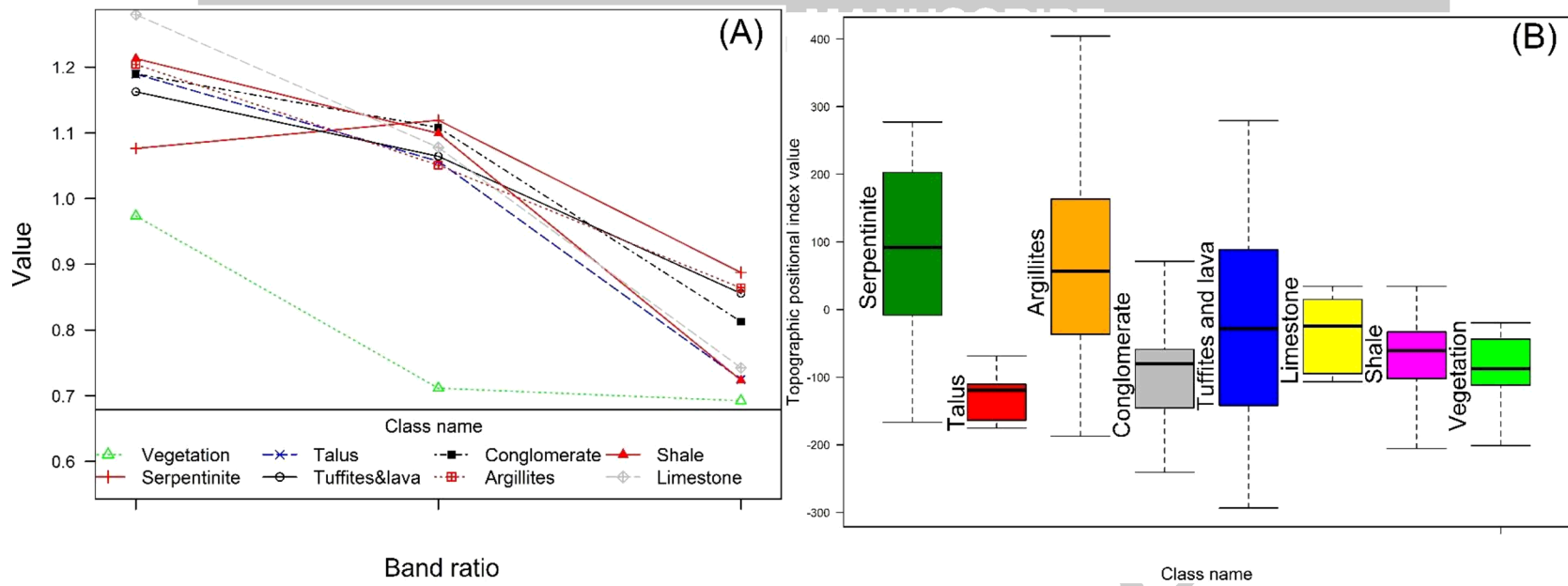
Fig. 2 Geological map overlaying the DEM from ASTER of the Bardi-Zard area that indicated in Figure 1 (Vasiliev and Pentelkov, 1962).

Fig. 3 Average of (A) surface reflectance; (B) PCA components, based on training samples selected from ASTER data acquired on 27 September 2004 and for seven lithological units and vegetation.

Fig. 4 (A) Average of three BRs computed from surface reflectance of ASTER data; (B) boxplot of the TPI data based on training samples for the eight discriminated classes.
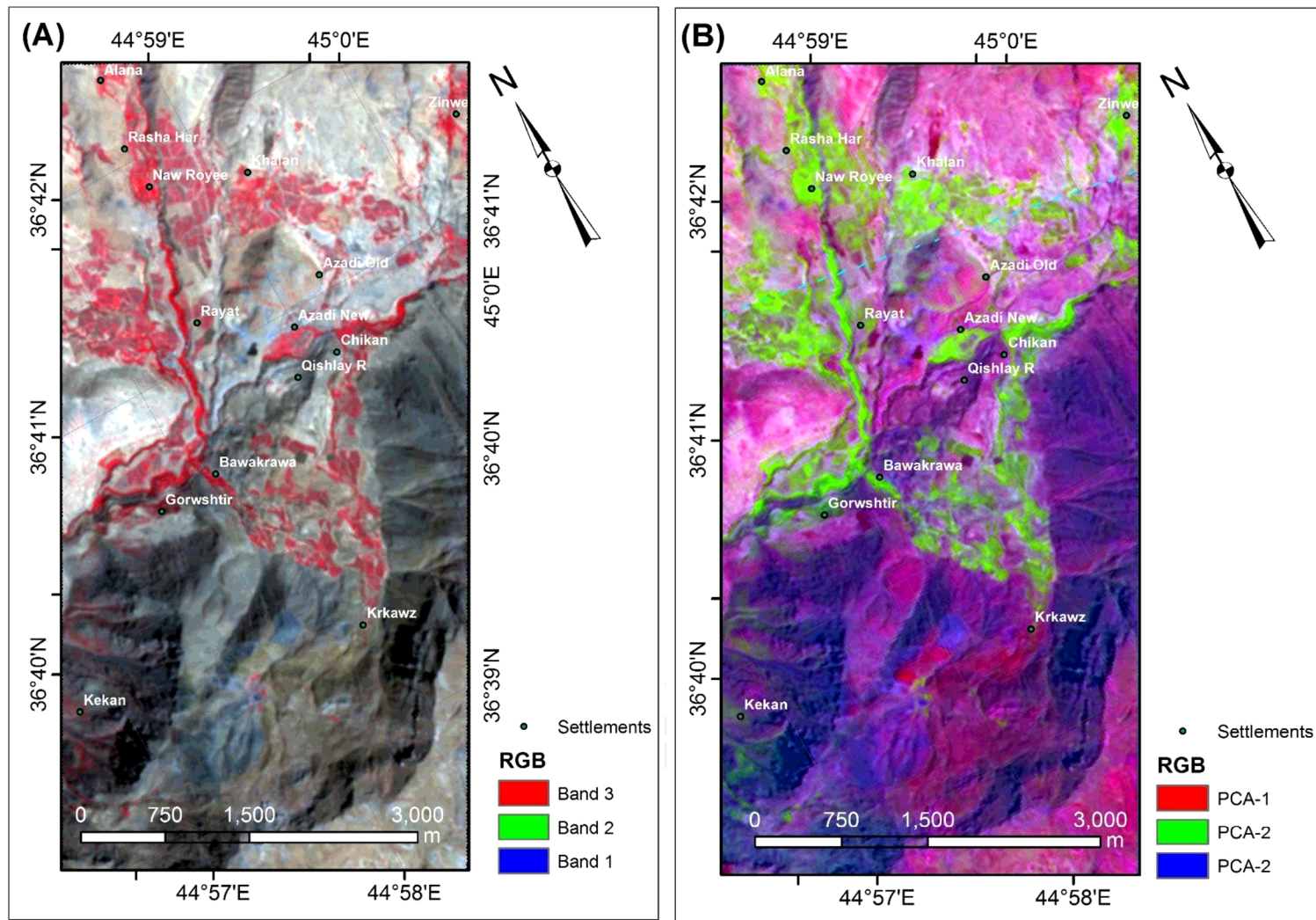
ACCEPTED MANUSCRIPT

Fig. 5 (A) ASTER data R3:G2:B1, acquired on 27 September 2004; (B) Principal component analysis PC1:PC2:PC3 of the study area.
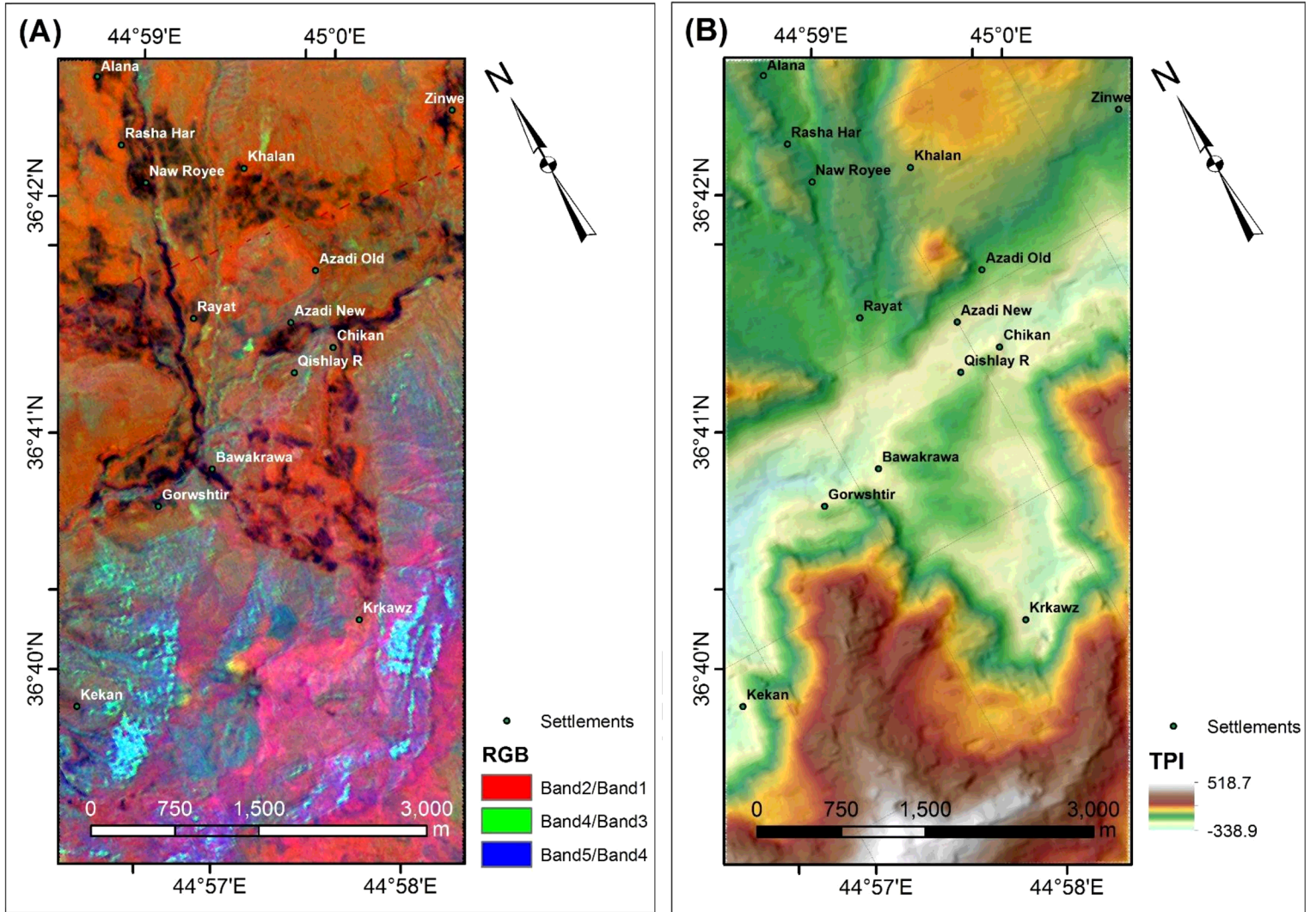
Fig. 6 (A) Band ratio of ASTER data R band2/band1:G band4/band3:B band5/band4, acquired on 27 September 2004; (B) Topographic Positional Index (TPI) of the study area.
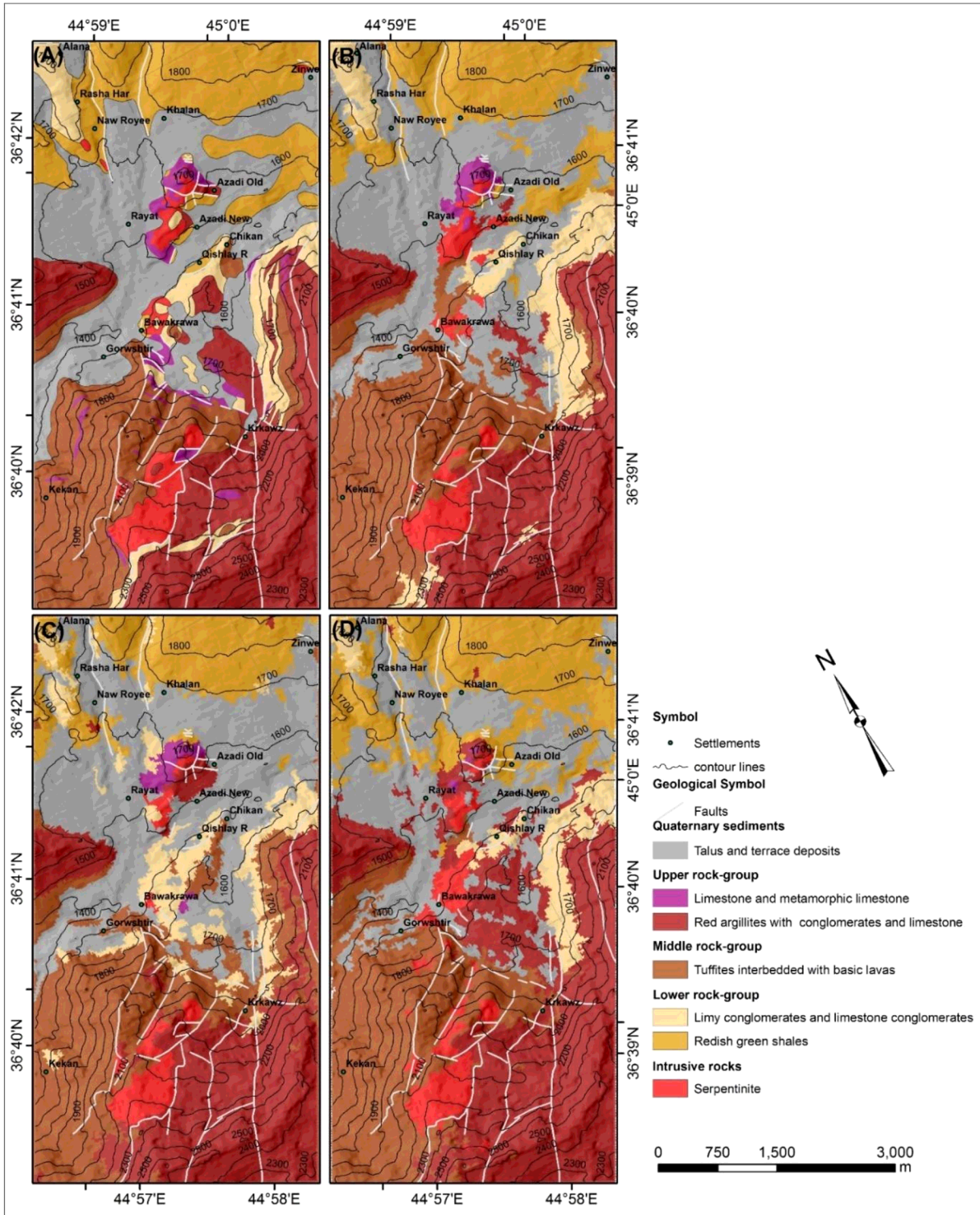
Fig. 7 (A) reference lithological map, (B) RF, (C) SVM and (D) ML lithological discriminated classes from combined layers (i.e., ASTER reflectance, PCA, BR, TPI and, X and Y spatial coordinates data).
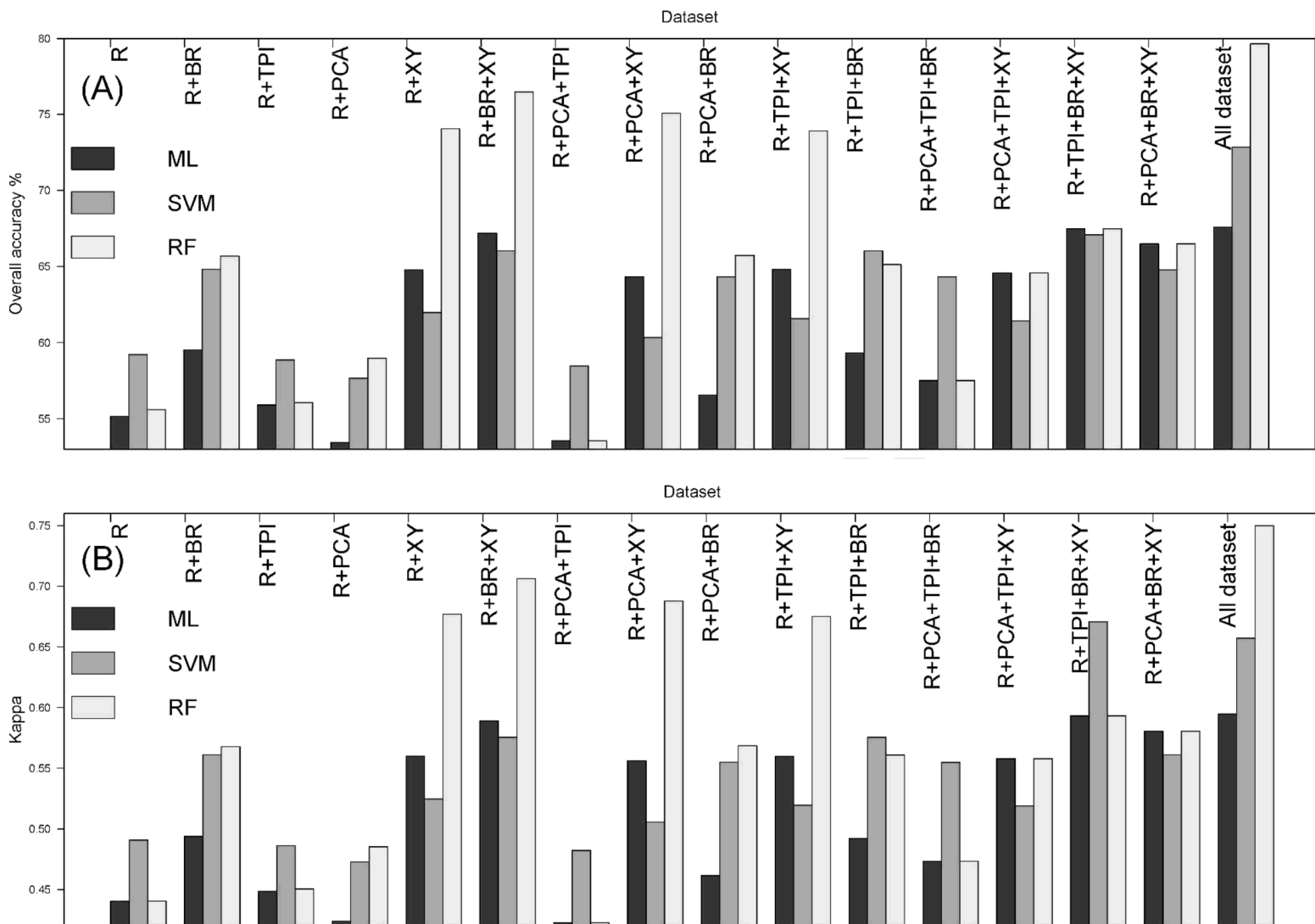
Fig. 8 (A) Overall accuracy and (B) kappa for different combination datasets using ML, SVM and RF methods.
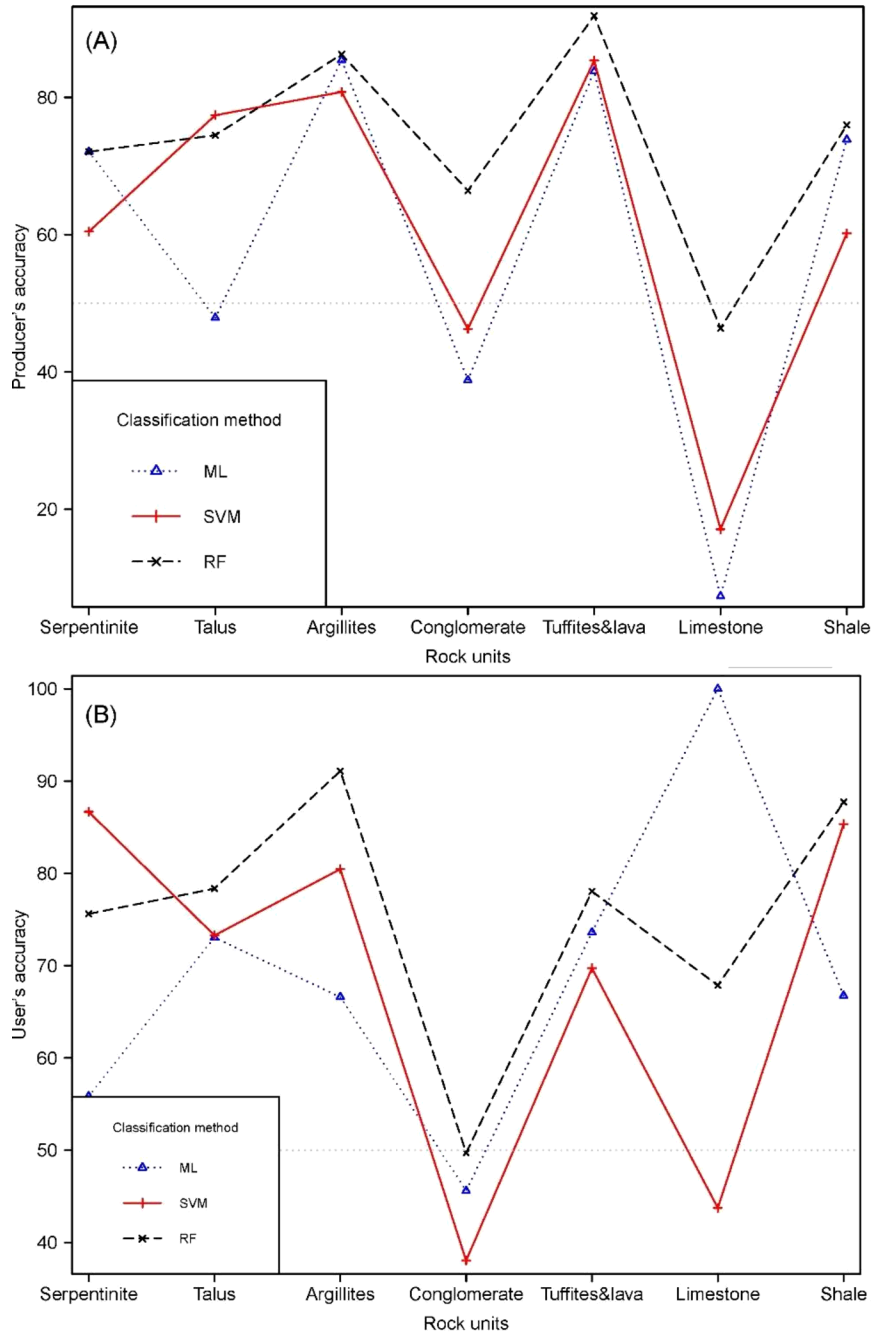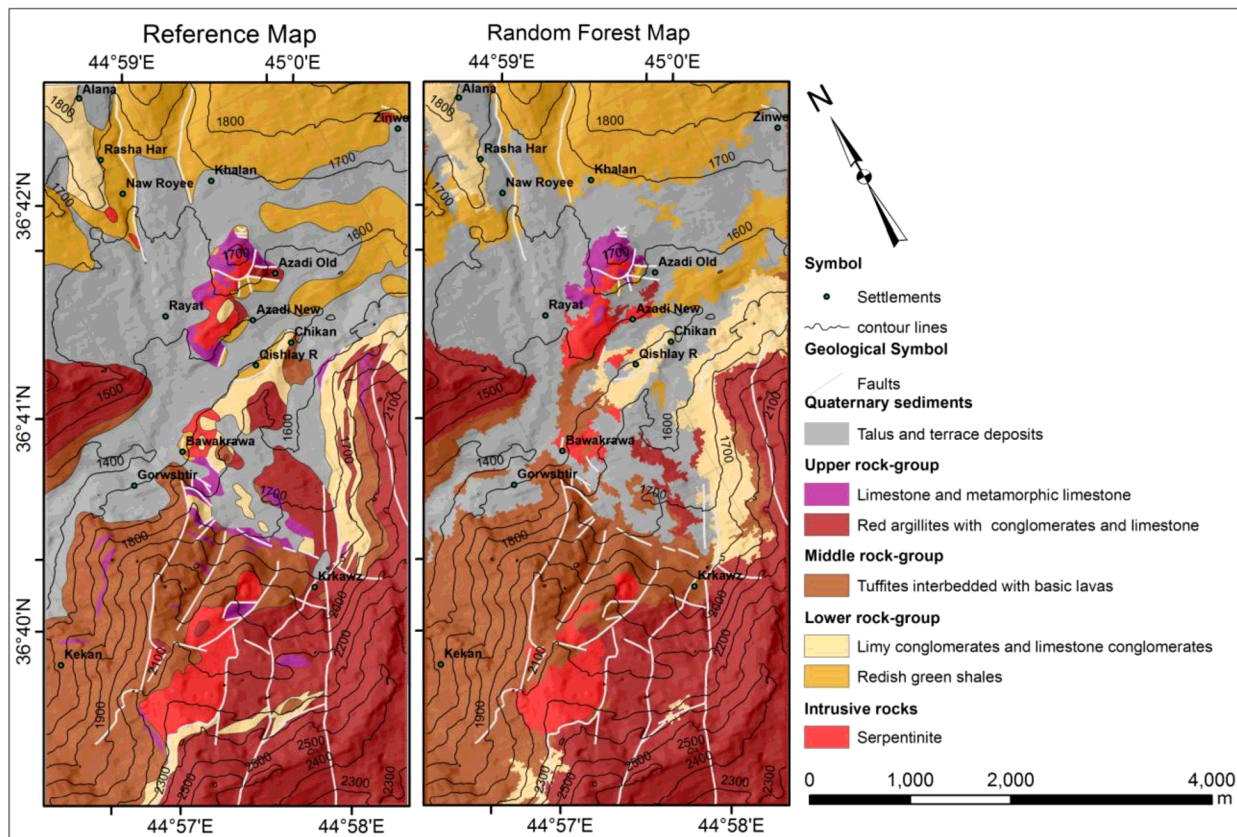
Fig. 9 Classification accuracy for all datasets using Ml, SVM and RF methods (A) Producer's accuracy and (B) User's accuracy.

Group	Designation	Variables
R	reference bands	9
R+BR	Reference bands and band ratios	12
R+PCA	Reference bands and principal components	14
R+TPI	Reference bands and TPI	10
R+XY	Reference bands, and X and Y layers	11
R+BR+PCA	Reference bands, band ratios and principal components	17
R+BR+TPI	Reference bands, band ratios and TPI	13
R+BR+XY	Reference bands, band ratios, X and Y layers	14
R+PCA+TPI	Reference bands, principal components and TPI	15
R+PCA+XY	Reference bands, principal components, and X and Y layers	16
R+TPI+XY	Reference bands, TPI and X and Y layers	12
R+PCA+TPI+XY	Reference bands, principal components, TPI, and X and Y layers	17
R+BR+TPI+XY	Reference bands, band ratios, TPI and X and Y layers	15
R+BR+PCA+XY	Reference bands, band ratios, principal components and X and Y layers	19
R+BR+PCA+TPI	Reference bands, band ratios, principal components, and TPI	18
All dataset	Reference bands, band ratios, principal components, TPI and X and Y layers	20

Class name	Lithological map	ML	SVM	RF
Serpentinite	1.045	1.254	0.762	1.074
Talus and terrace deposits	7.407	5.397	7.760	7.188
Red argillites with conglomerates and limestone	6.327	7.895	6.295	6.074
Limy conglomerates and limestone conglomerates	1.816	1.312	2.254	2.204
Tuffites interbedded with basic lavas	4.787	5.646	5.924	5.839
Limestone and Metamorphic limestone	0.638	0.031	0.258	0.246
Reddish green shales	4.094	4.558	2.861	3.487

Method	Added layer	Max.	Min.	Mean	S.D.
ML	Spatial coordinates	11.02	-0.20	9.53	1.21
SVM	Spatial coordinates	8.52	-0.40	2.79	2.50
RF	Spatial coordinates	22.14	-10.47	12.43	7.71
ML	TPI	1.10	-0.20	0.41	0.47
SVM	TPI	8.07	-0.40	1.43	2.8
RF	TPI	13.18	-10.47	-2.52	7.66

Graphical abstract

Highlights

- Adding geomorphic indices in the dataset increases the classification accuracy.

- The best overall accuracy is achieved by the Random Forest algorithms (~80%).

- Topographic Position Index seems to have higher impact in SVM.