POLISH JOURNAL OF SOIL SCIENCE
VOL. LIII/1 2020 PL ISSN 0079-2985

DOI: 10.17951/pjss/2020.53.1.97

LALEH PARVIZ*

PERFORMANCE EVALUATION OF REMOTE SENSING DATA WITH MACHINE LEARNING TECHNIQUE TO DETERMINE SOIL COLOR

Received: 27.02.2019
Accepted: 17.03.2020

Abstract. The aim of the present research is the determination of soil color by spectral bands and indices obtained from MODIS images. For this purpose, soil samples were collected from East Azerbaijan Province (Iran) and their color and texture were investigated through Munsell color system and hydrometer method, respectively. Stepwise regression, principle component analysis and sensitivity function methods were employed to find the dominant indices and bands using artificial neural network (ANN) as one of the machine learning techniques. The improved indices as the model input had better performance, for example, the calculation of correlation coefficient between indices and hue showed 51.48% increase of correlation coefficient with comparison of the normalized difference vegetation index (NDVI) to modified soil adjustment vegetation index (MSAVI) and 54.54% correlation enhancement of soil adjustment vegetation index (SAVI) compared to MSAVI. Stepwise regression method along with error criteria decline may enhance the performance of soil color model. In comparison with multivariate regression, ANN model exhibited better performance (with a 12.61% mean absolute error [MAE] decline). Temporal variation of modified perpendicular drought index (MPDI) as well as band 31 could justify the Munsell soil color components variations specifically chroma and hue. MPDI and thermal bands could be employed as a precise indicator in soil color analysis. Thus, remote sensing data combined with machine learning technique can clarify the procedure potential for soil color determination.

Keywords: soil color, Munsell chart, stepwise regression, ANN, MPDI

* Faculty of Agriculture, Azarbaijan Shahid Madani University, Tabriz, Iran, e-mail: laleh_parviz@yahoo.com

INTRODUCTION

As one of the major attributes, soil color has been used by the USDA Natural Resources Conservation Service (NRCS) to describe soil morphology (Stiglitz *et al.* 2017). The soil properties and internal processes can be determined with reference to soil color. Moreover, it is a comprehensive indicator of water drainage, aeration, and organic matter content of the soils (Henry 1991, Han *et al.* 2016). In this regard, searching for an efficient method to determine soil color is of crucial importance.

Matching the soil sample with the standard Munsell color charts is a common method to determine soil color, which has three components: Hue, Value and Chroma (HVC). The dominant wavelength can be described by hue, while chroma and value represent the saturation and brightness, respectively (Leone and Escadafal 2001). The condition of the sample and proficiency of the person who performs the matching can influence the precision of the visual Munsell color system (Singh *et al.* 2004). Indeed, this method is highly sensitive to user's skill (Stiglitz *et al.* 2016). The use of Munsell charts for soil color determination over large areas is time-consuming and expensive and it requires large number of samples. But this problem can be resolved by multi-spectral reflectance. In this context, satellite data have shown high potency in presenting, mapping, and forecasting the processes including soil classification, erosion and providing soil maps in the vast areas (Singh *et al.* 2004). For this purpose, the remote sensing technique can be considered as the useful tool because it can resolve the limitations of ground observation (Huesca *et al.* 2014).

The remote sensing has been an appropriate alternative to determine soil color (Singh *et al.* 2004). Satellite images with various spectral bands can be employed to obtain the desired information. Satellite indices can be derived by applying mathematical tools on spectral bands (Raghavendra and Mohammad Aslam 2017). Based on the governing condition, the indices have to be improved; for instance, soil's background conditions can influence partial canopy spectra and calculated indices. Thus, soil's background effect has to be minimized. Therefore, Huete (1988) introduced the vegetation index which employed the soil-adjusted factor depending on the vegetation level. The precision of these indices has been validated in numerous applications, for instance, the crop coefficient and evapotranspiration were calculated with the vegetation index in Northeast Asia by Park *et al.* (2017), Bijaber *et al.* (2018) used the satellite indices for drought monitoring, and other researches related to soil color determination by satellite-based indices are mentioned in the following part.

The correlation between the normalized difference vegetation index (NDVI) and components of Munsell soil color chart has been studied through the use of the National Oceanic and Atmospheric Administration (NOAA) / the Advanced Very High-Resolution Radiometer (AVHRR) data in Brazil. Based on the line-

ar empirical model, good correlations could be observed among NDVI, hue and chroma. The NDVI was more associated with the hue and chroma than with the value. It has been shown that various indices can prove their validation under specific environmental conditions (Singh *et al.* 2004). The components of the Munsell soil color charts determination have been investigated in arid regions using Landsat (TM) data of 20 sample sites based on re-sampling into the size of 3 × 3 pixels. A strong correlation between the first and seventh bands of TM sensor with Munsell notation expressed the dependency of hue, value and chroma to the visible bands of Landsat data. Three Munsell components were modeled and value exhibited the strongest correlation compared to the hue and chroma. However, further studies have been recommended to obtain accurate results (Matinfar *et al.* 2010). A modified attribute agreement analysis was employed to evaluate the soil color uncertainty with the Munsell soil color charts. The accurate results from the Munsell chart were associated with the hue component (Marques-Mateu *et al.* 2018). It should be mentioned that spectrophotometric and colorimetric alternatives instead of the Munsell soil color chart are costly and require external power sources which makes them difficult methods for determination of the soil color in the field (Levin *et al.* 2005, Thompson *et al.* 2013, Stiglitz *et al.* 2017).

Several studies have addressed the relationship between soil texture and the three components of Munsell soil color system. Based on these studies, value is very weakly correlated with clay content and there is also a very weak correlation between the sand and color components. However, all color components are well correlated with the silt content (Gunal *et al.* 2008).

Assimilation of machine learning techniques and remote sensing data has been widely employed for the prediction of soil features (i.e. soil organic matter, cation exchange capacity, magnesium and potassium content, and pH). A combination of satellite data with machine learning algorithms can result in a highly accurate estimation of soil features highlighting their applicability in mapping the soil properties (Khanal *et al.* 2018). As the spectral bands and satellite indices are the modeling tools in remote sensing applications, their comprehensive study is vital for determination of soil color, however, a limited number of researches have addressed this issue. Based on the literature, the determination of the soil color requires a complete assessment of satellite data performance (Singh *et al.* 2004, Matinfar *et al.* 2010).

The goal of the present study is the evaluation of the applicability of remote sensing information for determination of the soil color leading to a precise comparison of the spectral and thermal bands as well as various indices. For this purpose, the samples were collected from East Azerbaijan Province, Iran. Stepwise regression, principal components analysis and sensitivity function methods were applied to find the most important indices or bands. The performance of methods was also assessed using an artificial neural network (ANN) and multivariate regression models.

MATERIALS

Study area

The soil samples were collected from the lands of Azarbaijan Shahid Madani University in Azarshahr located in East Azerbaijan Province (northwest of Iran) as shown in Fig. 1. According to the De Martonne and Torrent-White climate classification index, the predominant climate of region is semiarid. The proximity of the studied area to Urmia Lake was one of the serious challenges

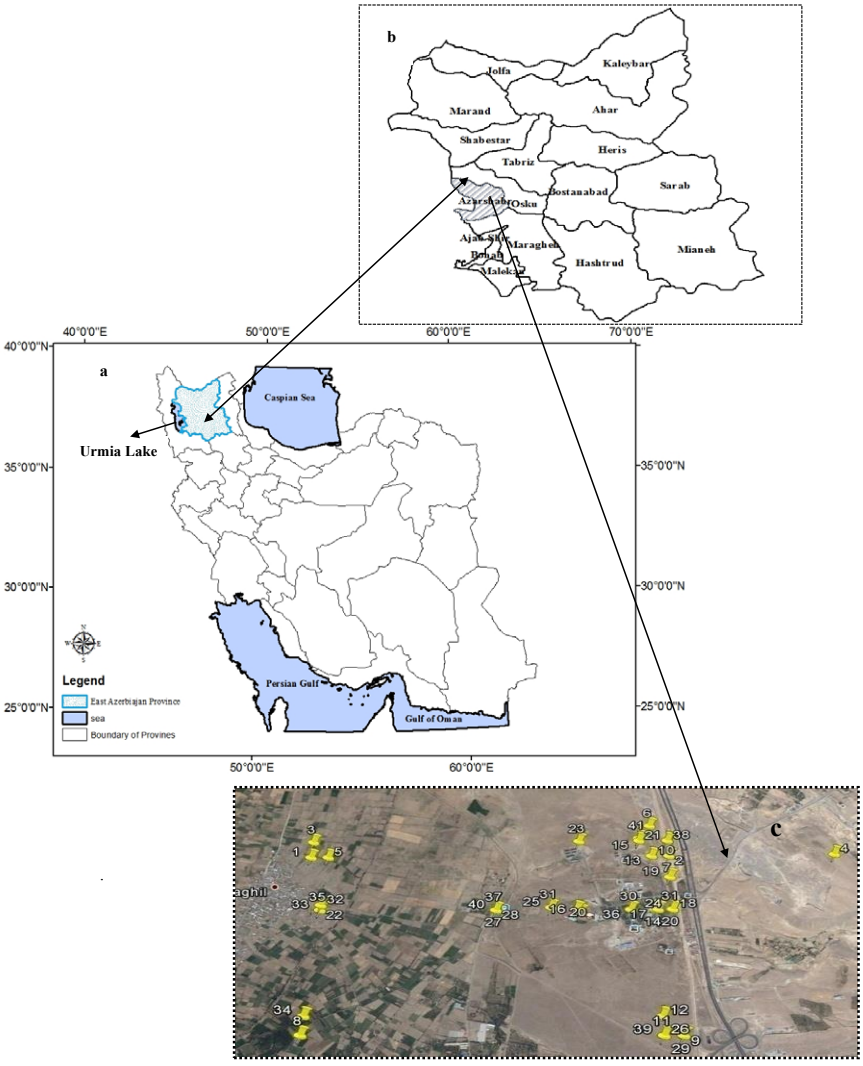


Fig. 1. The study area; East Azerbaijn Province location in Iran (a), cities of Province; highlighted part is Azarshahr (b) and sampling point's situation in Azarshahr (c).

as drying of the lake had an adverse impact on the climate of region (Mahsafar *et al.* 2017). In this context, assessment of the soil parameters (i.e. soil color) in that region is crucial as it may improve the soil condition through measures taken according to these parameters. A total of 41 topsoil samples were obtained where the apparent changes of soil were the selection criteria in samples location. Hydrometer method and Munsell color charts were employed for evaluation of soil texture and color, respectively. Soil texture of the studied area was mainly sand and loamy sand with very low organic matter content.

The basic knowledge of this study are indices of satellite images and reflectance solar bands: band 1 (bandwidth: 620–670 nm), band 2 (841–876 nm), band 3 (459–479 nm blue), band 4 (545–565 nm green), band 31 (10.78–11.28 μm) and band 32 (11.77–12.27 μm). Therefore, it covers the full spectral range consisting of visible (band 3, 4, 1), near-infrared (band 2) and thermal emissive (band 31, 32) bands. The satellite data were attained from Terra MODIS data on some days of April (2015 – 96 and 2015 – 108), May (2017 – 137), June (2017 – 158, 2017 – 172 and 2017 – 175) and October (2015 – 277, 2016 – 282) at the resolution of 1 km.

METHODS

Indices of satellite images

The data derived from satellite images, spectral and thermal bands, were expressed as indices which were calculated based on the ratios or normalized differences of 2 or 3 bands. The used indices are represented in Table 1 (note that land surface temperature (LST) shows the performance of the thermal bands).

Table 1. Indices of satellite images with spectral and thermal bands

Name	Abbrev.	Equation
Normalized Difference Vegetation Index	NDVI	$\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$
Vegetation Condition Index	VCI	$\frac{NDVI - NDVI_{\min}}{NDVI_{\max} - NDVI_{\min}}$
Soil Adjustment Vegetation Index	SAVI	$\frac{(1+L)(\rho_{nir} - \rho_{red})}{\rho_{nir} + \rho_{red} + L}$
Modified Soil Adjustment Vegetation Index	MSAVI	$0.5[2\rho_{nir} + 1 - \sqrt{(2\rho_{nir} + 1)^2 - 8(\rho_{nir} - \rho_{red})}]$
Perpendicular Drought Index	PDI	$\frac{1}{\sqrt{M^2 + 1}}(\rho_{red} + M\rho_{nir})$

Name	Abbrev.	Equation
Modified Perpendicular Drought Index	MPDI	$\frac{\rho_{red} + M\rho_{nir} - f(\rho_{v,red} + M\rho_{v,nir})}{(1-f_v)\sqrt{M^2+1}}$
Temperature Condition Index	TCI	$\frac{LST_{\max} - LST}{LST_{\max} - LST_{\min}}$
Temperature Vegetation Index	TVX	$\frac{LST}{NDVI}$

ρ_{nir} and ρ_{red} represent pixel reflectance in the near-infrared and red bands; $NDVI_{max}$ and $NDVI_{min}$ are maximum and minimum NDVI; L is soil adjustment factor; M represents the soil line slope; f_v is the fraction of ground surface covered by vegetation; $\rho_{v,red}$ and $\rho_{v,nir}$ are the pure vegetation reflectance in the red and near bands; LST, LST_{max} and LST_{min} are land surface temperature, maximum and minimum LST of each pixel (Agapiou *et al.* 2012, Shahabfar *et al.* 2012).

NDVI is a highly potent index but suffers from several drawbacks. NDVI has high values in heterogeneously covered lands where the soil, climate and ecosystem are more favorable (Rymuza *et al.* 2012). Improvement and development of SAVI and MSAVI were carried out by iterative and variable L function which minimized the soil background effect (Jiang *et al.* 2007). The relationship of red and near-infrared reflectance could be expressed by soil line through plot characterization of the spectral behavior of a pixel without variations in the vegetation and substantial moisture content. PDI is defined based on the vertical distance from partial cover in the near-infrared and red spectral space to the line dissecting the coordinate origin. Accordingly, parameters such as type of land cover, land use and soil heterogeneity are not taken into account (Ghulam *et al.* 2007). Shahabfar *et al.* (2012) introduced the modified perpendicular drought index (MPDI). LST indicates the energy balance of the processes on earth's surface at both regional and global levels (Ghulam *et al.* 2007). TCI was used to estimate the share of thermal band in condition evaluation (Du *et al.* 2013), while TVX simply showed the ratio of thermal and reflected radiations.

The indices and bands possessing substantial impact on the three components of Munsell color charts (HVC) play a more significant role in soil color determination. In this regard, stepwise regression, principle components analysis and sensitivity function methods are applied to seek for the effective indices or bands. The details of these methods will be discussed in following sections. Also, the association of Munsell color system (HVC) components with soil texture was evaluated in this research.

Stepwise regression

Stepwise regression applies an automatic model selection route and it is particularly useful in cases with a large number of potential explanatory variables (Sharma and Jin 2015). Stepwise regression is one member of linear regression family which uses step-by-step iterative construction. In this method, the independent variables are selected in an automatic manner, then the stepwise regression will be recursively constructed by addition or removal of independent variables in which some matrix tools are employed (Peng *et al.* 2018).

Principle component analysis

The method was known as the most significant multivariate statistical methods. Principle component analysis (PCA) is capable of transforming the large number of variables into some dominant variables (with lower number) called "principle components". The analysis is based on this fact that the total variance of variables can explain the maximum variance of dataset (which have high dependence on primary variables). The primary principle component has the highest possible variance. The second component is computed in such a way that it is orthogonal to the first component with the highest possible inertia. The other components will be calculated in the same way (Abdi and Williams 2010, Rymuza *et al.* 2012). The Kaiser-Meyer-Olkin test shows the suitability in terms of sampling adequacy, and possibility of PCA performance and the minimum value of that for an acceptable PCA was expressed 0.5 according to Wuttichaikitcharoen and Babel (2014).

Sensitivity function

A univariate regression analysis has to be employed to evaluate the relationship between two variables to determine the performance of indices or bands relative to Munsell color system components (HVC). This implies investigation of the indices or bands sensitivity throughout the full range of the Munsell color system components (HVC). In a bivariate regression model, the Munsell color system components (HVC) are considered as independent variables while and satellite data are taken as dependent ones. The error in estimation of dependent variables ($\hat{y}$) can be formulated through standard error of $\hat{y}$. The standard error of $\hat{y}$ in linear regression, curvilinear models and in nonlinear regression can be estimated by Eq. 1 and 2, respectively. The sensitivity function (s) definition is also mentioned in Eq. 3.

$$\sigma_{\hat{y}_i} = \sqrt{\sigma^2 X_i^{'}(X^{'}X)^{-1}X_i} \tag{1}$$

$$\sigma_{\hat{y}_i} = \sqrt{\sigma^2 F_i^{'}(F^{'}F)^{-1}F_i} \tag{2}$$

$$s = \frac{\hat{y}'}{\sigma_{\hat{y}}} = \frac{d\hat{y}/dx}{\sigma_{\hat{y}}} \tag{3}$$

Where:
σ^2 shows MSE,
X represents the independent variables matrix,
X_i denotes the i^{th} row of X, F stands for the derivative matrix and
F_i is the i^{th} row of F matrix (Yang and Guo 2014, Ji and Peters 2007).

The higher the positive value of s led to the larger sensitivity of x to y. $\hat{y}'$ significance can be also explored through use of Student's or z-statistics. Student's t- or z-statistics can be used to test the significance of $\hat{y}'$. The result of the t-test is indicated with the corresponding p-value which either rejects (p-value < α). The s-value itself indicates the t-score (Ji and Peters 2007).

ANN model

ANN can be employed to assess the methods efficiency in accurate selection of bands or indices. ANN was developed as a representation of the biological neural network in mathematics. Using ANN, the complicated associations among input and output data can be described. Similar to the mathematical model calibration, ANN involves learning or training. An activation function is defined as a function employed for transformation of a unit (neuron) activation level into an output signal. The weights of ANN will be determined based on the error minimization in observation and simulation data (Ozgür 2005, Akhtar *et al.* 2009). The feed-forward back propagation network was applied in this study.

Error criteria of model performance evaluation

Equations 4–7 describe the error criteria for assessing ANN performance related to different inputs methods.

Root Mean Square Error $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - S_i)^2} \tag{4}$$

Relative Root Mean Square Error $$RRMSE = \frac{RMSE}{\overline{O}} \tag{5}$$

Mean Absolute Relative Error $$MARE = \frac{\sum_{i=1}^{N}\left|\frac{O_i - S_i}{O_i}\right|}{n} \tag{6}$$

Mean Absolute Error $$MAE = \frac{\sum_{i=1}^{N}|O_i - S|}{n} \tag{7}$$

Where:
O_i are observation data,
S_i are simulation data,
$\overline{O}$ is mean of observed data,
n is total number of data (Moustris Kostas *et al.* 2011).

The criteria with the lowest values reflect the best performance. Fig. 2 shows the research procedures in a flowchart.

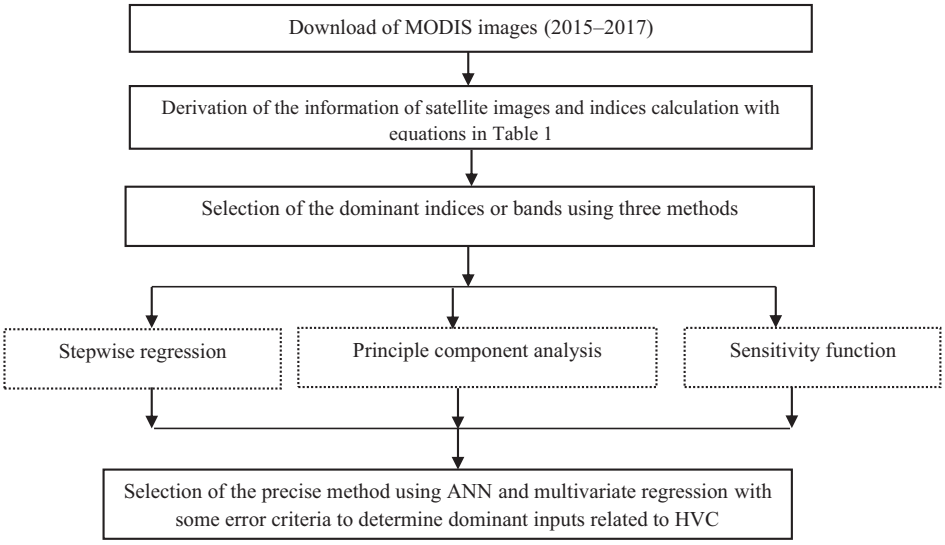


Fig. 2. The flow chart of ANN modeling between HVC and remote sensing data with emphasis on selection dominate inputs.

RESULTS

Satellite images data and their relationship with Munsell system components

The satellite images on some days of April, May, June and October of 2015, 2016 and 2017 were used. Soil sampling was easier in these months and that was the criteria in months selection. Input variables were band 1, 2, 3, 4, 31, 32 as well as indices mentioned in Table 1 covering visible, near-infrared and thermal bands. Regarding the variety of near-infrared and thermal bands, the resolution was set to 1 km. The data driven from satellite images were on average 9 pixels (3 × 3) centered on the sampling point. LST was computed using split window algorithm as developed by Price (1984).

Moreover, the effect of soil texture on the Munsell color system (HVC) components was evaluated, therefore, the results of soil texture determination of samples are brought in Table 2.

Table 2. Specifications of soil texture related to the soil samples

Soil sample number	% sand	% clay	% silt	Soil texture	Soil sample number	% sand	% clay	% silt	Soil texture
1	92.5	3.7	3.7	sand	22	93.0	4.5	2.5	sand
2	89.5	3.0	7.5	sand	23	89.5	4.5	6.0	sand
3	91.8	2.5	5.6	sand	24	91.5	5.0	3.5	sand
4	94.4	2.2	3.2	sand	25	87.5	4.5	8.0	sand
5	87.4	0.6	11.8	sand	26	93.5	5.0	1.5	sand
6	85.6	3.5	10.8	loamy sand	27	91.0	5.0	4.0	sand
7	90.3	2.5	7.0	sand	28	95.5	4.0	0.5	sand
8	86.1	8.2	5.6	loamy sand	29	91.0	5.0	4.0	sand
9	91.2	4.3	4.3	sand	30	93.0	5.5	1.5	sand
10	91.0	4.5	4.5	sand	31	92.5	5.0	2.5	sand
11	82.5	9.0	8.5	loamy sand	32	87.0	5.0	8.0	loamy sand
12	92.5	3.0	4.5	sand	33	93.5	4.5	2.0	sand
13	96.2	1.2	2.5	sand	34	94.4	4.5	1.1	sand
14	95.9	1.7	2.3	sand	35	94.9	4.0	1.1	sand
15	93.0	3.0	4.0	sand	36	93.0	5.5	1.5	sand
16	94.5	3.5	2.0	sand	37	91.5	4.5	4.0	sand
17	93.0	3.0	4.0	sand	38	87.5	4.5	8.0	sand
18	87.0	5.5	7.5	loamy sand	39	92.0	4.0	4.0	sand
19	90.5	5.5	4.0	sand	40	94.0	4.5	1.5	sand
20	91.5	2.5	6.0	sand	41	94.0	4.5	1.5	sand
21	93.0	5.0	2.0	sand					

The soil texture of studied area was dominantly composed of sand and loamy sand with lower spatial distribution of loamy sand. The correlation of satellite-driven data and soil particle size with three Munsell color charts (HVC) components is tabulated in Table 3. Based on Hurst (1977), hue data were numerically presented as 7.5YR = 17.5, 10YR = 20, 5YR = 15, and 2.5YR = 12.5.

Table 3. Relationship of Munsell soil system components of Munsell color system (HVC) with soil texture particles and satellite information

Name	H	V	C
Sand	0.01	-0.14	0.30
Clay	-0.01	0.00	-0.01
Silt	-0.24	0.09	-0.30*
NDVI	0.10	-0.07	-0.07
LST	0.35*	0.12	0.51**

Name	H	V	C
TCI	0.06	-0.27	0.19
TVX	0.06	-0.04	0.3
VCI	-0.34*	-0.11	-0.42**
SAVI	0.09	-0.13	0.16
MSAVI	0.15	-0.08	0.17
PDI	0.15	-0.06	0.57**
MPDI	0.13	-0.01	0.61**
Band 1	0.11	-0.03	0.59**
Band 2	0.17	-0.07	0.53**
Band 3	-0.15	-0.19	0.22
Band 4	0.00	-0.09	0.48**
Band 31	0.40**	0.05	0.55**
Band 32	0.37*	0.15	0.41**

*and** significant at 0.05 and 0.01 level of probability, respectively. Hue (H), Value (V), Chroma (C).

With regard to the hue, the significant parameters included LST, VCI, band 31 and 32, while the important parameters of chroma were silt, LST, VCI, PDI, MPDI, band 1, 2, 4, 31, and 32. The value component showed no correlation with the other data. Low correlation between soil texture and Munsell components could be attributed to the homogenously distributed soil texture as the soil texture of the studied area possessed poor diversity. The chroma had higher number of the significant correlation coefficients compared to the hue. Hue and chroma showed the strongest correlation with band 31 and MPDI, respectively. The chroma correlation with MPDI was, however, stronger than that of hue. In general, the reflectance and thermal bands were effective in both cases (hue and chroma), the thermal bands influences were, however, more profound for hue. The value exhibited the lowest correlation to clay contents and color. Hue was negatively correlated to clay. Moreover, visible spectral bands were weakly correlated with hue. Furthermore, improved indices increased the correlation coefficient. For instance, in case of hue determination, the correlation coefficient was enhanced by 51.48% from NDVI to MSAVI, from SAVI to MSAVI, it also exhibited 54.54% increase. Concerning chroma determination, we observed a 6.25% improvement from SAVI to MSAVI, while this enhancement was 7.2% from PDI to MPDI. Regarding the variety of studied indices and bands, three methods were applied to find the major inputs of models.

Stepwise regression

Automatic choice of independent variables by stepwise regression analysis had been conducted to find important variables on soil color components. The effective variables of hue were related to band 31, while MPDI and band 31 were for chroma. Band 31 is impressive on two components of Munsell color.

Principle component analysis

Effective factors of the Munsell soil color components (HVC) were determined by PCA. The Kaiser-Meyer-Olkin was calculated as 0.599, showing its suitability in terms of sampling adequacy and possibility of PCA performance. Eigenvalues trend is depicted in Fig. 3 which clearly suggests the descending trend of eigenvalue by increase of the component number. Moreover, the first five components are responsible for 95.25% of the entire variation (their eigenvalues were higher than 1). Table 4a also lists the PCA results. As Table 4b shows, for the first component, PDI, MPDI, band 1, 2, 4 possessed the maximum coefficient values, in case of the second component, maximum coefficients were assigned to band 31, 32; while in the third, fourth and fifth components NDVI, sand and clay exhibited maximal coefficient values. The coefficients values which related to the first component were, however, greater than the others.

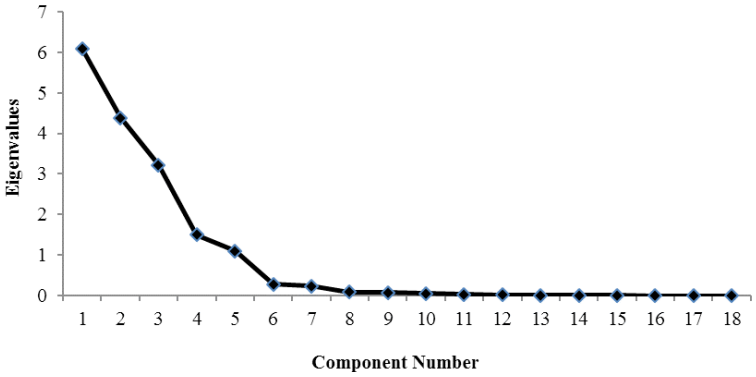


Fig. 3. Screen plot of investigated bands, indices and soil texture particles

Table 4a. PCA eigenvectors of investigated bands, indices and three soil texture particles.

Factor	Sand	Clay	Silt	NDVI	LST	TCI	TVX	VCI	SAVI
PCA 1	0.13	0.19	-0.31	0.04	0.58	0.66	0.43	-0.4	0.39
PCA 2	-0.56	0.17	0.53	0.63	-.65	0.51	-0.4	0.33	0.67
PCA 3	0.26	-0.22	-0.15	0.75	0.27	-0.25	-0.71	-0.67	0.51
PCA 4	-0.74	0.57	0.51	-0.08	0.25	-0.22	-0.16	0.03	-0.08
PCA 5	0.16	0.71	-0.56	0.1	0.03	0.17	-0.01	0.37	0.13

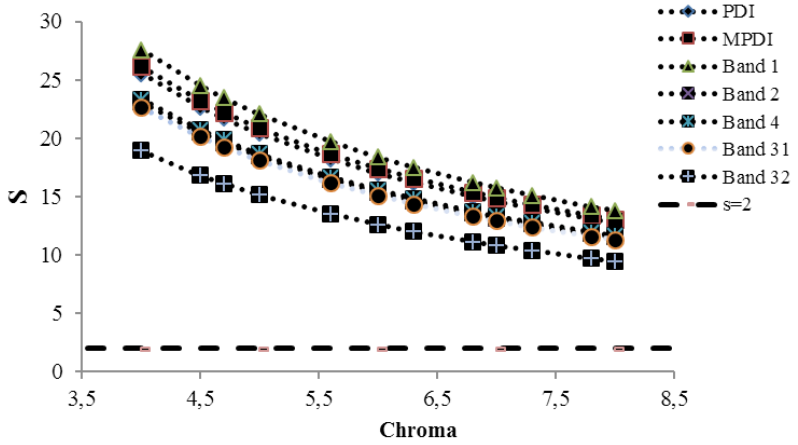
Table 4b. PCA eigenvectors of investigated bands, indices and three soil texture particles.

Factor	MSAVI	PDI	MPDI	Band 1	Band 2	Band 3	Band 4	Band 31	Band 32
PCA 1	0.44	0.98	0.95	0.93	0.94	0.69	0.91	0.45	0.31
PCA 2	0.6	0.13	-0.06	-0.17	0.24	0.5	0.11	-0.73	-0.78
PCA 3	0.64	0.03	-0.07	-0.27	0.18	-0.49	-0.35	0.36	0.34
PCA 4	-0.02	0.01	0.06	0.03	0.002	-0.09	0.004	0.24	0.35
PCA 5	0.06	-0.1	-0.08	-0.13	-0.06	-0.01	-0.08	0.09	0.04

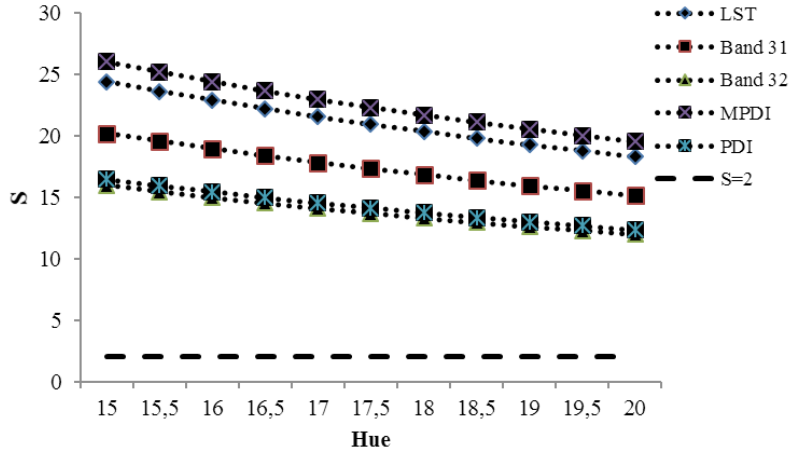
Sensitivity function

The sensitivity function can be calculated by finding the relationship between three Munsell color charts (HVC) components and satellite information. Fig. 4 represents the results concerning sensitivity function and the impact of the three components of Munsell color charts (HVC) on the sensitivity trend of each index or band. For $s > 2$ (t-score of 2 is the critical value), the sensitivity of indices or bands to the Munsell soil color components (HVC) became significant. The chroma exhibited high sensitivity to band 1, MPDI and PDI, while the sensitivity of hue was high to MPDI and LST. In case of value, it exhibited high dependence upon LST and PDI. The sensitivity of SAVI and MPDI remained relatively consistent throughout the value range. The indices, however, showed decreased sensitivity for cases with chroma, hue and value less than 7, 19.5 and 7.5, respectively. MPDI had high sensitivity with regard to two Munsell components.

a



b



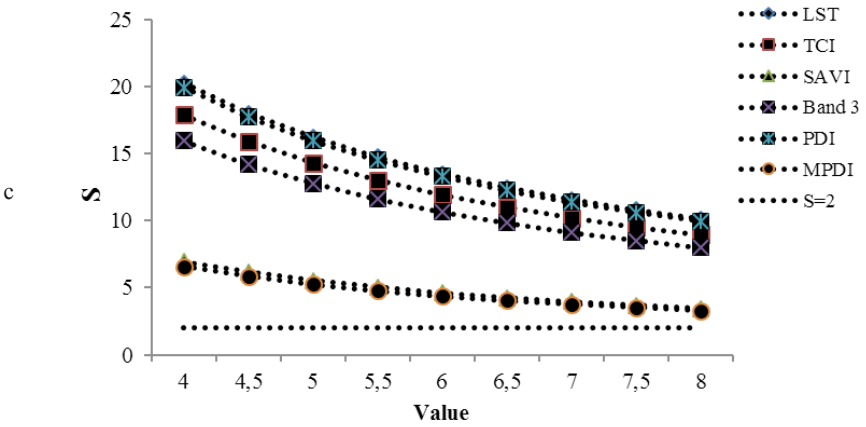


Fig. 4. Sensitivity functions of satellite information; indices and bands, related to chroma (a), hue (b) and value (c).

ANN sensitivity analysis

The selected variables with the three mentioned methods were employed as an ANN input. ANN sensitivity related to the type of activation function; Log Sigmoid, Tangent Sigmoid and Pure Linear; performances were examined through the error criteria. Fig. 5 illustrates RMSE variation for various types of activation functions of ANN model with regard to PCA method. RMSE calculation related to the value and chroma components revealed that minimum RMSE of activation functions can be expected for Tangent Sigmoid and Linear for hidden and output layers concerning value, and for chroma, the lowest RMSE of activation functions were observed in Log Sigmoid and Linear for hidden and output layers. The maximum RMSE was assigned to Tangent Sigmoid and Log Sigmoid of hidden and output layers for value and chroma components.

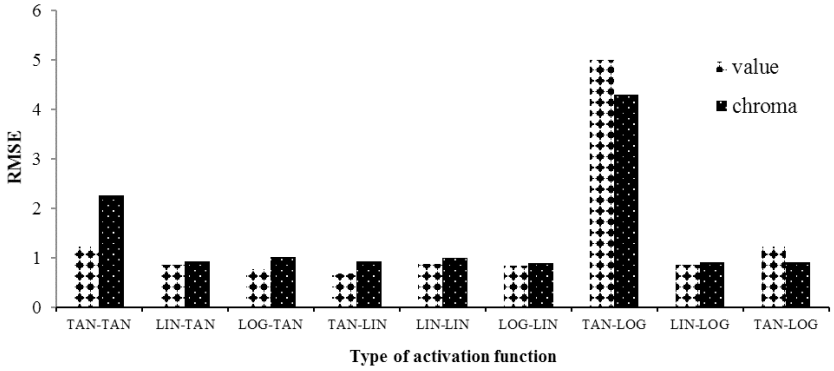


Fig. 5. Sensitivity analysis of ANN to find precise activation function of hidden and output layers (Log Sigmoid [LOG]- Tangent Sigmoid [TAN]; Pure Linear [LIN]) related to Chroma and Value determination.

Evaluation the performance of input selection methods

The performance of stepwise regression, principle and sensitivity function methods using ANN and multivariate regression was assessed as presented in Fig. 6 (the abbreviation of figure are the following HSM – hue stepwise multivariate, HPM – hue principle multivariate, HSFM – hue sensitivity function multivariate, VPM – value principle multivariate, VSFM – value sensitivity function multivariate, CSM – chroma stepwise multivariate, CPM – chroma principle multivariate, CSFM – chroma sensitivity function multivariate, HAS – hue ANN stepwise, HPA – hue principle ANN, HSFA – hue sensitivity function ANN, VPA – value principle ANN, VSFA – value sensitivity function ANN, CSA – chroma stepwise ANN, CPA – chroma principle ANN, CSFA – chroma sensitivity function ANN-RRMSE). In terms of error criteria, the stepwise regression possessed the least error value compared to other methods. For stepwise regression, band 31 was selected for hue, while MPDI and band 31 were selected for chroma. MPDI and thermal bands were selected as the dominant variables for soil color determination. By comparing the statistical methods, it was revealed that ANN model could reduce the errors.

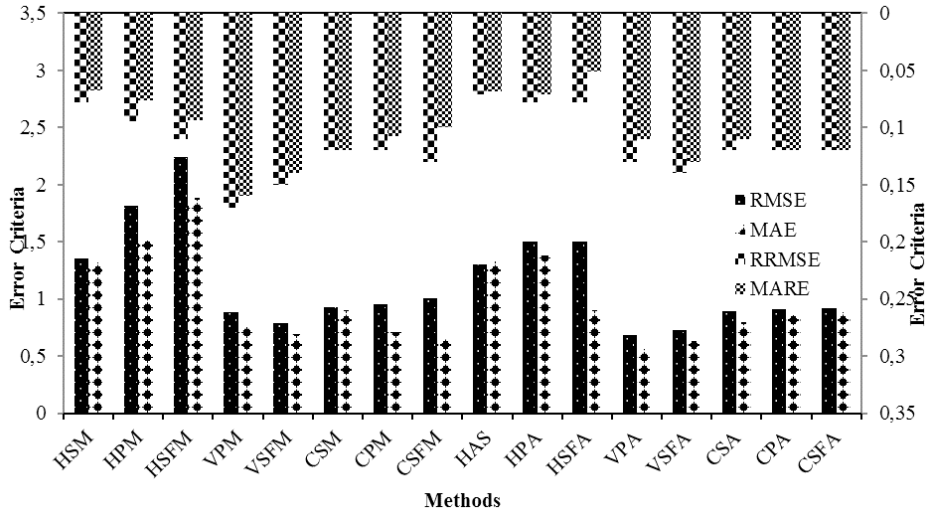


Fig. 6. Comparison of the different input selection methods and different models for three components of Munsell color system (HVC) determination.

The performance of satellite indices in the term of temporal variations was also investigated. June possessed the maximum LST among the addressed months. Temperature enhancement resulted in decline of soil moisture content, while soil reflectance showed an increasing trend by decrease of soil moisture. Increase of soil reflectance resulted in a lighter color. The chroma values ranged from 0 (neu-

tral color) to 8 (maximal color intensity). During June, maximum chroma values were associated with the highest MPDI and LST. Wang *et al.* (2004) expressed that NDVI is negatively correlated to LST. The highest LST values were recorded in June, which did not coincide with the months possessing minimum NDVI. It indicates higher efficiency of thermal bands when compared to NDVI. Low TCI values reflect a serious moisture condition (Du *et al.* 2013), thus minimum values of TCI are expected in June, which is not in line with the mentioned month. Such an inconsistency raises some doubts about TCI efficacy.

DISCUSSION

Correlation coefficient analysis was used to investigate the effect of soil texture on the three components of Munsell chart. Positive correlation implies the growth of Munsell components upon enhancement of the soil texture particles. A study related to this matter indicated the lowest coefficients between clay contents and value. Gunal *et al.* (2008) also reported a poor correlation between the value and clay contents. The value was negatively and positively correlated with sand and clay, respectively. The silt and clay possess high visible light reflectance resulting in light-color soils. Furthermore, the specific surface area of sand particles is lower compared to the silt or clay. So, organic matters could easily cover the sand particles giving rise to darker colors (Ibanez-Asensio *et al.* 2013). The hue was positively correlated with sand, while its correlation with clay and silt was negative which is in line with the studies by Kone *et al.* (2009), and Leone and Escadafal's (2001) reports. A positive correlation between the variables and the hue means that they make the soil more yellowish, while a negative correlation implies an increase in red ratio of soil color. The negative correlation between the hue and clay could result from weathered soil particles. Soils with finer texture exhibited more saturated colors which might be attributed to the reddish finely-grained sedimentary rocks developed as a result of elevated iron (in oxide form). Higher chroma values were detected in sandy soils (Ibanez-Asensio *et al.* 2013), which was verified by a positive correlation between chroma and soil contents.

Escadafal *et al.* (1989) have studied the Munsell soil color and soil reflectance by Landsat MSS, they reported a weak correlation between the hue and visible spectral bands. Soil color components always exhibited less than 50% correlation with spectral bands, while their correlation with MPDI, PDI and band 31 always exceeded 50%. Therefore, MPDI and thermal bands were considered as dominant inputs. In terms of sensitivity function, MPDI showed high sensitivity to the two components of the Munsell color system. According to Baret and Guyot (1991), noise can be dramatically decreased by sensitivity analysis with modified indices. Mohamed *et al.* (2018) have recommended visible and

near-infrared spectroscopy as a quick tool for mapping the soil properties. The PDI and MPDI showed similar results. Accordingly, Ghulam *et al.* (2007) have declared that MPDI and PDI will lead to similar results in lands with decreased vegetation cover. Moreover, MPDI is capable of precise identification of water and energy cycle as well as the cover type of the surface (Ghulam *et al.* 2007).

The stepwise procedure showed improved performance as it solves the problem of multicollinearity through declining the total number of variables (Gilabert *et al.* 2002). ANN exhibited better performance compared to multivariate regression, as it can solve the complicated interactive non-linear and multivariate systems (Ozgür 2005). Multivariate regression methods have been compared with ANN in terms of soil parameters determination. For instance, Marashi *et al.* (2017) estimated soil aggregate stability indices, Shabani and Norouzi (2015) also predicted the soil's cation exchange capacity. All these researchers agreed on the superiority of ANN compared to multivariate regression. Thus, combining the satellite images information and machine learning techniques can enhance the precision of results as confirmed in the study by Khanal *et al.* (2018).

CONCLUSIONS

Soil color is one of the significant soil features capable of characterizing the horizons separation and identifying different types of soil. The limited number of Munsell color chips as well as the user-dependent matching process has restricted the application of Munsell color system. Therefore, in this study, it was attempted to apply a different methodology which was a combination of satellite images and machine learning techniques to improve the determination of soil color. The modified indices (i.e. MSAVI and MPDI) improved the values of correlation coefficients. MPDI resulted in the maximum correlation coefficient reflecting the necessity of defining different soil lines for each soil type. Thermal bands could effectively describe soil moisture. The mathematical procedures and parameters of indices are the major factors in modeling the band's behavior as confirmed by MPDI performance. Specific equations should be defined depending on the purpose. The use of stepwise regression to select the dominant variables resulted in the lowest error values. Compared to the multivariate regression, ANN exhibited better performance in the estimation of the soil color as it had lower errors. Thus, it can be concluded that the MPDI and thermal band can be employed as the precise tools for soil color analysis. The satellite data and soil color are highly dependent on the information regarding the soil line and its parameters. Temporal variation of MPDI and band 31 can be explained by the Munsell soil color components. The dominant spectral band and indices are powerful tools in the satellite-based evaluation of soil color, which can be exploited in various fields of soil and water sciences.

ACKNOWLEDGEMENTS

We are grateful for grants (97/D/773) from the Deputy of Research and Technology of Azarbaijan Shahid Madani University.

REFERENCES

[1] Abdi, H., Williams, L.J., 2010. *Principal component analysis*. WIREs: Computational Statistics, 2: 433–459.
[2] Agapiou, A., Hadjimitsis, D.G., Alexakis, D.D., 2012. *Evaluation of broadband and narrowband vegetation indices for the identification of archaeological crop marks.* Remote Sensing, 4(12): 3892–3919.
[3] Akhtar, M.K., Corzo, G.A., Van Ande, S.J., Jonoski, A., 2009. *River flow forecasting with artificial neural networks using satellite observed precipitation pre-processed with flow length and travel time information: case study of the Ganges river basin.* Hydrology and Earth System Sciences, 13: 1607–1618.
[4] Baret, F., Guyot, G., 1991. *Potentials and limits of vegetation indices for LAI and APAR assessment.* Remote Sensing of Environment, 35(2–3): 161–173.
[5] Bijaber, N., El Hadani, D., Saidi, M., Svoboda, M.D., Wardlow, B.D., Hain, C.R., Christian, P., Yessef, M., Rochdi, A., 2018. *Developing a Remotely Sensed Drought Monitoring Indicator for Morocco*. Geosciences, 8(5): 1–18.
[6] Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., Huanga, Y., 2013. *A comprehensive drought monitoring method integrating MODIS and TRMM data*. International Journal of Applied Earth Observation and Geoinformation, 23: 245–253.
[7] Escadafal, R., Girard, M.C., Courault, D., 1989. *Munsell soil color and soil reflectance in the visible spectral bands of landsat MSS and TM data*. Remote Sensing of Environment, 27(1): 37–46.
[8] Ghulam, A., Qin, Q., Teyip, T., Li, Z.L., 2007. *Modified perpendicular drought index (MPDI): A real-time drought monitoring method.* ISPRS Journal of Photogrammetry and Remote Sensing, 62: 150–164.
[9] Gilabert, M.A., González-Piqueras, J., García-Haro, F.J., Melia, J., 2002. *A generalized soil-adjusted vegetation index*. Remote Sensing of Environment, 82: 303–310.
[10] Gunal, H., Ersahin, S., Yetgin, B., Kutlu, T., 2008. *Use of chromameter-measured color parameters in estimating color-related soil variables*. Communications in Soil Science and Plant Analysis, 39: 726–740.
[11] Han, P., Dong, D., Zhao, X., Jiao, L., Lang, Y., 2016. *A smartphone-based soil color sensor: for soil type classification*. Computers and Electronics in Agriculture, 123: 232–241.
[12] Henry, D.F., 1991. *Fundamentals of soil science*, 8th ed. Wiley Publication, p. 354, https://www.wiley.com/en-us/9780471522799.
[13] Huesca, M., Litago, J., Merino-de-Miguel, S., Cicuendez-López-Ocaña, V., Palacios-Orueta, A., 2014. *Modeling and forecasting MODIS-based Fire Potential Index on a pixel basis using time series models*. International Journal of Applied Earth Observation and Geoinformation, 26: 363–376.
[14] Huete, A.R., 1988. *A soil-adjusted vegetation index (SAVI)*. Remote Sensing of Environment, 25: 295–309.
[15] Hurst, V.J., 1977. *Visual estimation of iron in saprolite*. Geological Society of America Bulletin, 88: 174–176.
[16] Ibáñez-Asensio, S., Marqués-Mateu, A., Moreno-Ramón, H., Balasch, S., 2013. *Statistical relationships between soil colour and soil attributes in semiarid areas*. Biosystems Engineering, 116(2): 120–129.

[17] Ji, L., Peters, A.J., 2007. *Performance evaluation of spectral vegetation indices using a statistical sensitivity function*. Remote Sensing of Environment, 106: 59–65.
[18] Jiang, Z.H., Huete, A.R., Li, J., Qi, J., 2007. *Interpretation of the modified soil-adjusted vegetation index isolines in red-NIR reflectance space*. Journal of Applied Remote Sensing, 1: 1–12.
[19] Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. *Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield*. Computers and Electronic in Agriculture, 153: 213–225.
[20] Kone, B., Yao-Kouame, A., Ettien, J.B., Oikeh, S., Yoro, G., Diatta, S., 2009. *Modelling the relationship between soil color and particle size for soil survey in Ferralsol environments*. Soil & Environment, 28(2): 93–105.
[21] Leone, A.P., Escadafal, R., 2001. *Statistical analysis of soil colour and spectroradiometric data for hyperspectral remote sensing of soil properties (example in a southern Italy Mediterranean ecosystem)*. International Journal of Remote Sensing, 22(12): 2311–2328.
[22] Levin, N., Ben-Dor, E., Singer, A., 2005. *A digital camera as a tool to measure colour indices and related properties of sandy soils in semi-arid environments*. International Journal of Remote Sensing, 26(24): 5475–5492.
[23] Mahsafar, H., Maknoon, R., Saghafian, B., 2017. *The impact of climate change on water level of Urmia Lake*. Research in Marine Sciences, 2(2): 83–94.
[24] Marashi, M., Mohammadi Torkashvand, A., Ahmadi, A., Esfandyari, M., 2017. *Estimation of soil aggregate stability indices using artificial neural network and multiple linear regression models*. Spanish Journal of Soil Science, 7(2): 122–132.
[25] Marqués-Mateu, A., Moreno-Ramón, H., Balasch, S., Ibáñez-Asensio, S., 2018. *Quantifying the uncertainty of soil colour measurements with Munsell charts using a modified attribute agreement analysis*. Catena, 171: 44–53.
[26] Matinfar, H.R., Alavi Panah, S.K., Rafiei Emam, A., 2010. *Remotely sensed data evaluation on soil spectral properties in arid regions*. Iranian Journal of Range and Desert Research, 16(4): 560–573.
[27] Mohamed, E.S., Saleh, A.M., Belal, A.B., Gad, A.A., 2018. *Application of near-infrared reflectance for quantitative assessment of soil properties.* The Egyptian Journal of Remote Sensing and Space Sciences, 21(1): 1–14.
[28] Moustris Kostas, P., Larissi Ioanna, K., Nastos Panagiotis, T., Paliatsos Athanasios, G., 2011. *Precipitation forecast using artificial neural networks in specific regions of Greece*. Water Resources Management, 25: 1979–1993.
[29] Ozgür, K., 2005. *Daily river flow forecasting using artificial neural networks and auto-regressive models*. Turkish Journal of Engineering and Environmental Science, 29: 9–20.
[30] Park, J., Baik, J., Choi, M., 2017. *Satellite-based crop coefficient and evapotranspiration using surface soil moisture and vegetation indices in Northeast Asia*. Catena, 156: 305–314.
[31] Peng, Y., Zhang, Y.J., Liu, D.T., Liu, L.S., 2018. *Degradation estimation using feature increment stepwise linear regression for PMW inverter of electro-mechanical actuator*. Microelectronics Reliability, 88(90): 514–518.
[32] Price, J.C., 1984. *Land surface temperature measurements from the split window channel of the NOAA 7 advanced very high resolution radiometer*. Journal of Geophysical Research, 89: 7231–7237.
[33] Raghavendra, B.R., Mohammad Aslam, M.A., 2017. *Sensitivity of vegetation indices of MODIS data for the monitoring of rice crops in Raichur district, Karnataka, India*. The Egyptian Journal of Remote Sensing and Space Sciences, 20: 187–195.
[34] Rymuza, K., Turska, E., Wielogórska, G., Bombik, A., 2012. *Use of principal component analysis for the assessment of spring wheat characteristics.* Acta Scientiarum Polonorumseria Agricultura, 11(1): 79–90.
[35] Shabani, A., Norouzi, M., 2015. *Predicting cation exchange capacity by artificial neural network and multiple linear regression using terrain and soil characteristics*. Indian Journal of Science and Technology, 8(28): 1–10.

[36] Shahabfar, A., Ghulam, A., Eitzinger, J., 2012. *Drought monitoring in Iran using the perpendicular drought indices*. International Journal of Applied Earth Observation and Geoinformation, 18: 119–127.
[37] Sharma, M.J., Jin, Y.S., 2015. *Stepwise regression data envelopment analysis for variable reduction*. Applied Mathematics and Computation, 253: 126–134.
[38] Singh, D., Herlin, I., Berroir, J.P., Silva, E.F., Simoes Meirelles, M., 2004. *An approach to correlate NDVI with soil colour for erosion process using NOAA/AVHRR data*. Advances in Space Research, 33: 328–332.
[39] Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J., 2016. *Evaluation of an inexpensive sensor to measure soil color*. Computers and Electronics in Agriculture, 121: 141–148.
[40] Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J., Pargas, R., Glover, B., Mooney, J., 2017. *Soil color sensor data collection using a GPS-enabled Smartphone application*. Geoderma, 296: 108–114.
[41] Thompson, J., Pollio, A., Turk, P., 2013. *Comparison of Munsell soil color charts and the GLOBE soil color book*. Soil Science Society of America Journal, 77: 2089–2093.
[42] Wang, Y., Woodcock, C.E., Buermann, W., Stenberg, P., Voipio, P., Smolander, H., Häme, T., Tian, Y., Hu, J., Knyazikhin, Y., Myneni, R.B., 2004. *Evaluation of the MODIS LAI algorithm at a coniferous forest site in Finland.* Remote Sensing of Environment, 91: 114–127.
[43] Wuttichaikitcharoen, P., Babel, M.S., 2014. *Principal component and multiple regression analyses for the estimation of suspended sediment yield in ungauged basins of northern Thailand*. Water, 6: 2412–2435.
[44] Yang, X., Guo, X., 2014. *Quantifying responses of spectral vegetation indices to dead materials in mixed grassland.* Remote Sensing, 2: 4289–4304.