

*Original Research*

# Multivariate Logistic Regression Model for Soil Erosion Susceptibility Assessment under Static and Dynamic Causative Factors

**Abdulkadir Taofeeq Sholagberu<sup>1,2\*</sup>, Muhammad Raza Ul Mustafa<sup>1,3</sup>,  
Khamaruzaman Wan Yusof<sup>1</sup>, Ahmad Mustafa Hashim<sup>1</sup>, Mumtaz Muhammad Shah<sup>4</sup>,  
Muhammad Waris A. Khan<sup>5</sup>, Mohamed H. Isa<sup>1</sup>**

<sup>1</sup>Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS,  
32610 Seri Iskandar, Perak, Malaysia

<sup>2</sup>Department of Water Resources and Environmental Engineering, University of Ilorin,  
PMB 1515, Ilorin, Kwara State, Nigeria

<sup>3</sup>Centre for Urban Resource Sustainability, Institute of Self-Sustainable Building, Universiti Teknologi PETRONAS,  
32610 Seri Iskandar, Perak, Malaysia

<sup>4</sup>Department of Earth Sciences, Quaid-i-Azam University, 45320 Islamabad, Pakistan

<sup>5</sup>Faculty of Industrial Management, Universiti Malaysia Pahang, 26300 Gambang, Pahang, Malaysia

*Received: 27 May 2018*

*Accepted: 3 June 2018*

## Abstract

Soil erosion is a devastating land degradation process that needs to be spatially analyzed for identification of critical zones for sustainable management. Geospatial prediction through susceptibility analysis assesses the occurrence of soil erosion under a set of causative factors (CFs). Previous studies have considered majorly static CFs for susceptibility analysis, but neglect dynamic CFs. Thus, this study presents an evaluation of erosion susceptibility under the influence of both non-redundant static and dynamic CFs using multivariate logistic regression (MLR), remote sensing and geographic information system. The CFs considered include drainage density, lineament density, length-slope and soil erodibility as static CFs, and land surface temperature, soil moisture index, vegetation index and rainfall erosivity representing the dynamic CFs. These were parameterized to establish geospatial relationships with the occurrence of erosion. The results showed that length-slope had the highest positive impact on the occurrence of erosion, followed by lineament density. During the MLR classification process, predicted accuracies for the eroded and non-eroded locations were 89.1% and 83.6% respectively, with an overall prediction accuracy of 86.6%. The model's performance was satisfactory, with 81.9% accuracy when validated using the area-under-curve method. The output map of this study will assist decision makers in sustainable watershed management to alleviate soil erosion.

**Keywords:** soil erosion, susceptibility, dynamic causative factors, logistic regression

## Introduction

Soil erosion has been identified as one of the most devastating phenomena threatening soil sustainability. It is a natural geomorphic process involving detachment and displacement of soil particles causing adverse impacts on the watershed [1-4]. This is responsible for about 80% of current land degradation on agricultural land worldwide [5]. Soil erosion is a product of complex interactions of some biophysical factors such as climate, topographic nature, soil characteristics and land-use/land-cover management. Among these, Rahman et al. [2] highlighted that soil and topographical characteristics of a particular location strongly influence erosion rates. This is due to diverse hydrological processes, rainfall patterns, land-use and land management practices in different locations [3, 6]. Anthropogenic activities such as urbanization, inappropriate farming activities, mining and deforestation exert great pressure on natural ecosystems, thereby accelerating soil erosion processes [7]. The resultant impacts of extensive soil erosion include depleting agricultural productivity, water quality deterioration, landslide risk on the steep slope and sedimentation of reservoirs, which may result in flood risk and reduced hydropower generation. Sustainable management practices that will dampen these challenges require indices to quantify soil erosion, analyze its spatial distribution and identify critical areas through susceptibility analysis [8]. Erosion susceptibility analysis evaluates the probability of occurrence of erosion events based on its past spatial distribution under the influence of specific causative factors (CFs) [9]. In recent times, remote sensing, geographic information system (GIS) and statistical techniques have been applied for this purpose. The former is often used for the provision of moderate resolution spatial data required for the analyses. Logistic regression is one of the commonly used multivariate (statistical) techniques for susceptibility modeling, which analyzes the interactions among the erosion CFs.

In susceptibility analysis, there exist subjectivity issues in the selection of CFs as various researchers used varying numbers and types of CFs. This is due to the absence of specific framework in the existing literature [10-12]. The variability in the number of CFs considered in previous studies was reported by Abdulkadir, et al. [13]. Magliulo [14] cited other researchers who argued and recommended that the selection of CFs should not have double consequences on the results. In other words, CFs should be “non-redundant” as the inclusion of “redundant” factors could cause overweighting of the results [15]. The researchers further highlighted that simultaneous consideration of some factors such as slope angle with length-slope (LS-factor), plan curvature, stream power index, slope aspect and topographic wetness index as CFs is considered “redundant.” This is because they have mathematical connections with slope angle,

although they might have different geomorphological and hydrological significances. A critical survey of the literature indicated that most of the factors considered in the previous studies were majorly static in nature. These set of factors often remain the same throughout the months of the year and even up to decades despite the change in rainfall cycle. However, some dynamic CFs that respond to rainfall cycles, such as land surface temperature (LST), rainfall erosivity (R-factor) and soil moisture index (SMI), are sometimes neglected due to unavailability of the data [14] or the absence of such in the existing literature. In spite of their significance, their impacts on soil erosion susceptibility are yet to be investigated. Consequently, there is a need to investigate the addition of such dynamic CFs to frequently used non-redundant static CFs in susceptibility assessment. Hence, this study evaluates erosion susceptibility in Cameron Highlands watershed under the influence of both non-redundant static and dynamic CFs using multivariate logistic regression, remote sensing and GIS techniques. Cameron Highlands, located in the Pahang State of Malaysia, is a complex watershed characterized with extensive vegetable cultivation on the hillslope, high rate of urbanization and deforestation resulting in incessant soil erosion risks [16, 17]. Some of the negative impacts of soil erosion currently experienced include deterioration of water quality via increased water turbidity [18, 19], siltation of rivers and built water bodies that could trigger flood risk and reduced hydropower generation [20, 21]. Hence, accurate and acceptable soil erosion susceptibility assessment is crucial for sustainable watershed management.

## Material and Methods

### Erosion CFs in Susceptibility Assessment

It has been established that the quality of selected CFs and method of the analysis could influence model accuracy [22, 23]. Hence, it is pertinent to consider non-redundant CFs in the analysis. A total of four non-redundant static CFs were considered with four dynamic factors. The static CFs include drainage density, lineament density, K-factors and LS-factor, while dynamic CFs are NDVI, R-factor, LST and SMI. The static CFs have been identified to trigger soil erosion and applied by many researchers in susceptibility studies. For instance, drainage density [8, 24-26], lineament density [8, 27, 28], LS-factor [9, 29-33], K-factor [34-36], and NDVI [25, 28, 37] were implemented in those studies. However, some dynamic CFs such as R-factor, LST and SMI are rarely considered despite their roles in the soil erosion process. R-factor is a climatic factor that strongly impacts soil characteristics through type, intensity and amount of rainfall [14]. It measures the aggressiveness of raindrops to induce soil erosion by actions of rainfall and runoff [27, 38, 39]. LST is another dynamic CF that measures the degree of heat

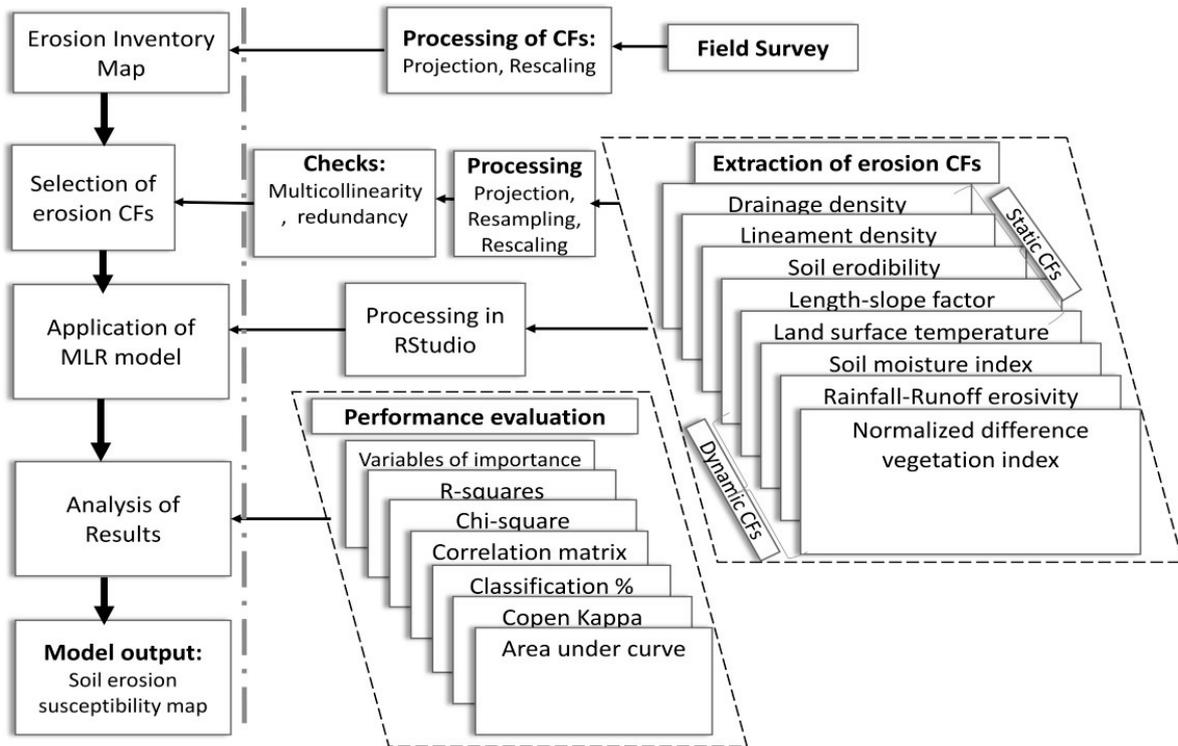


Fig. 1. Methodological flowchart for soil erosion susceptibility mapping.

energy exchange between the earth and atmosphere [38]. Its spatial distribution is affected by some terrain characteristics that impact the ascending solar radiation [41, 42], and it controls the distribution of soil moisture [43]. Removal of vegetation exposes the soil surface to high surface radiant energy that dries up the soil [44]. Such dried soil has the following properties as highlighted by Xue, et al. [45]: “high apparent density (compaction), low porosity, low permeability to air and water, low soil aggregates, and low water storage,” which promote soil erosion. Magliulo [14] highlighted that soil moisture is a significant factor that influences erosion processes. It affects both soil aggregate stability and shear strength. Karnieli and Ben-Asher [46] noted that soil erosion rate is dependent on rainfall intensity and soil moisture during daily runoff modeling in Marl watersheds, USA. Thus, R-factor, LST and SMI (which are scarcely used) were considered in this study in order to evaluate their roles in erosion susceptibility results.

#### Extraction of Erosion CFs

The overall procedures adopted in this study are illustrated in the methodological flowchart in Fig. 1. The erosion CFs were obtained from different sources and prepared in an ArcMap environment. Digital elevation model (DEM) is the digital representation of topography and other ground features on a watershed. Drainage density and LS-factor were extracted from hydrologically corrected DEM of 5m resolution. The stream network of the study area was supplied as the input file in the “Focal Statistics” platform of spatial

analyst tool of ArcMap for the extraction of drainage density map in Fig. 2b). LS-factor comprises slope gradient and slope length factors derivable from DEM. It was evaluated based on Equation (1), provided by Parveen and Kumar [47], and the resulting layer is as shown in Fig. 2c).

$$LS = (flowAccumulation * \frac{resolution}{22.1})^{0.6} * (Sin(Slope) * \frac{0.0174}{0.09})^{1.3} \tag{1}$$

A soil map of Cameron Highlands was extracted from the digital soil map of the world (DSMW). William’s approach in Equation (2) was applied for the computation of K-factor [48]. The extracted soil map and the component factors were processed in GIS environment for development of the K-factor map (Fig. 2d).

$$K_{factor} = f_{csand} * f_{cl-st} * f_{orgc} * f_{hisand} \tag{2}$$

...where  $f_{csand}$  = factor for high coarse-sand content,  $f_{cl-st}$  = factor for high clay-to-silt ratios,  $f_{orgc}$  = factor for high organic carbon content,  $f_{hisand}$  = for soils with extremely high sand content.

Landsat-8 data was acquired on 3 July 2016 from the USGS Earth Resource Observation Systems Data Center for the location (Row/Path: 57/127) on a relatively clear day with about 3.96% cloud cover. Lineaments were extracted from atmospherically and radiometrically

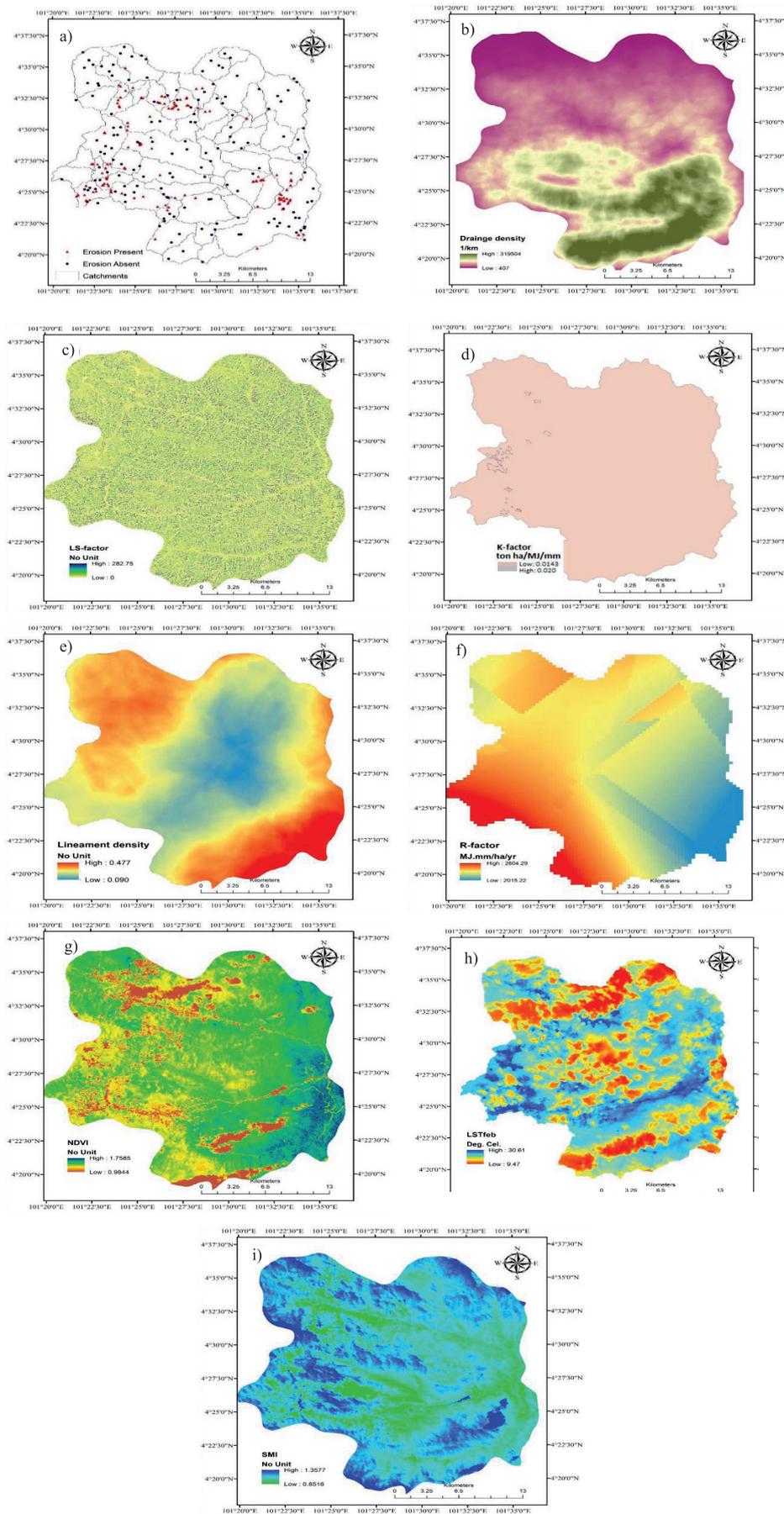


Fig. 2. a) Soil erosion inventory map, b) Drainage density map, c) LS-factor, d) K-factor, e) Lineament density map, f) R-factor, g) NDVI map, h) LST, and i) SMI.

corrected Landsat-8 with the use of ArcMap 10.0, ENVI 5.3 and Geomatica 2016 software. The output shapefile was processed in ArcMap using the split-line model to split the compounded lines into simple lines. Subsequently, a lineament density map (Fig. 2e) was developed using “Line Density” of spatial analyst tool. Furthermore, R-factor was estimated from monthly average rainfall records using the popular expression proposed by Wischmeier and Smith [49] in Equation (3). Rainfall records (2006-2016) for the synoptic weather stations within and around the study area were obtained from the Malaysian Department of Irrigation and Drainage for the evaluation of R-factor map in Fig. 2f). The NDVI map (Fig. 2g) was extracted from bands 4 and 5 of Landsat-8 images. For the avoidance of negative values, Equation (4) was adopted in this study.

$$R = \sum_{i=1}^{12} 1.735 * 10^{1.5 \log_{10} \left[ \frac{P_i}{P} - 0.08188 \right]} \quad (3)$$

...where  $P_i$  = monthly average of rainfall for month  $i$  (mm) and  $P$  = annual average of rainfall (mm)

$$NDVI = \frac{Band\ 5 - Band\ 4}{Band\ 5 + Band\ 4} + 1 \quad (4)$$

Thermal infrared sensors (TIRS; i.e., bands 10 and 11 of Landsat-8) were utilized for the retrieval of LST. The respective specific thermal conversion constants ( $K_1 = 774.89$ ,  $K_2 = 1321.08$ ) and ( $K_1 = 480.89$ ,  $K_2 = 1201.14$ ) for thermal bands 10 and 11 were used for estimating brightness temperature (BT). This, along with land surface emissivity ( $\epsilon$ ) and proportion vegetation ( $P_v$ ), were used for estimating LST in Equation (5) [50]. The procedure for the estimation and development of LST map (Fig. 2h) can be found in Abdulkadir, et al. [51] and Teerawong, et al. [52]. Lastly, an SMI spatial distribution map (Fig. 2i) was extracted from Landsat-8 using an NDVI-LST space technique as described by Sandholt, et al. [53], Sruthi and Aslam, [54] and Potić et al. [55]. After the preparation of CFs, they were resampled to have the same projection (i.e., UTM 47 North, WGS 84), same grid sizes (30x30 m) and the same numbers of columns by rows (1025x1112).

$$LST = \frac{BT}{(1 + w * \frac{BT}{p} * \ln(\epsilon))} \quad (5)$$

...where  $w$  = wavelength of emitted radiance (11.5  $\mu$ m), and  $p$  is defined as  $p = h * c/s$  in which  $h$  = Plank's constant (6.626x10<sup>-34</sup> Js),  $c$  = velocity of light (3.0x10<sup>8</sup> m/s),  $s$  = Boltzmann constant (1.38 \*10<sup>-23</sup> J/K).

#### Application of Multivariate Logistic Regression Model

Logistic regression is a multivariate technique applied to establish a relationship between the

occurrence of soil erosion and a set of CFs [8, 56]. It involves one or more predictor variables (i.e., CFs) to predict the probability of a binary or categorical response [24, 57]. The response variable (erosion) could be binary (presence or absence) or categorical, and the predictor variables could be binary, continuous, categorical, or a combination [58]. The method can be applied without variables necessarily satisfying the normality, linearity and homogeneity of variance assumptions as in the case of other techniques [58]. The extensive descriptions of multivariate logistic regression (MLR) can be found in Hosmer and Lemeshow [57]. In MLR analysis, an approximate equal number of binary cases (i.e., erosion-present points and erosion-absent points) are recommended [10]. The inventory map in Fig. 2a) shows the spatial distribution of soil erosion, which was prepared from information obtained from documentary sources, model and field campaign. The map contains 159 erosion-present points. Using slope map as a guide, a total of 159 erosion-absent points were randomly created through the use of a “fishnet tool” of ArcMap for building the MLR model. The erosion-present and erosion-absent points totaling 318 data points (response class) were used to extract equivalent values for each of the eight CFs. The information obtained was used to build MLR model in RStudio environment. This evaluated the weights for each CF and produced other relevant statistics for performance evaluation of the model. The weights were applied to evaluate erosion susceptibility for all the locations within the watershed of Cameron Highlands in the GIS environment. The performance of the MLR model was evaluated using receiver operating characteristics (ROC) curve analysis. The area under the curve (AUC) of the ROC describes the performance of classification models. It is a graphical plot that illustrates the diagnostic ability of a binary classifier whose acceptable value ranges between 0.5 and 1.0 [26].

## Results and Discussion

This section presents the results obtained from the evaluation of erosion CFs and susceptibility analysis. Spatial distribution maps for the CFs were prepared in a GIS environment and the resulting maps are presented in Figs 2(b-i). The map in Fig. 2b) indicates that northern and southern regions of the study area are characterized by low and high drainage densities, respectively. This indicates that the southern region has relatively high relief compared to the northern region. The LS-factor map in Fig. 2c) sees its values range from 0 to 282.75 with respective mean and standard deviation of 2.85 and 6.26. Processing the soil map of the study area showed that the mapping units have basically two soil types and the watershed is predominately filled with soil type coded as “Ao90-2/3c.” This can be distinguished as shallow soils with low potassium content and can be found on the steep slopes. The mapping unit is

Table 1. Multicollinearity diagnostic test index for CFs.

Variables	*Multicollinearity statistics	
	Tolerance	VIF
Drain density	0.743	1.346
Lineament density	0.993	1.007
LS-factor	0.823	1.215
LST	0.933	1.072
NDVI	0.528	1.895
R-factor	0.588	1.701
K-factor	0.812	1.231
SMI	0.646	1.549

composed of about 70% mountainous terrain and the available soil type is classified as Orthic Acrisols by the U.S. Food and Agricultural Organization (FAO). Also, about 70% of the dominant soil is made up of 65% and 35% of medium and heavy texture, respectively. The percentage of sand, silt, clay and organic carbon obtained from the DSMW database were 51.6%, 15.8%, 30.6% and 2.25%, respectively. This information was implemented for the computation and development of the K-factor map. The K-factor map obtained (in Fig. 2d) showed that its values ranged between 0.0143 and 0.020 ton ha/MJ/mm. The extracted lineament

density in Fig. 2e) had minimum and maximum values to be 0.0899 and 0.4766, respectively. High and low lineaments were found in the southeast and central regions of the watershed.

The dynamic CFs such as R-factor map was prepared from rainfall data while SMI, NDVI and LST were obtained from atmospherically and radiometrically corrected Landsat-8 data. Estimated R-factor values with known weather station coordinates were imported into ArcMap and interpolated for other areas within the watershed. The minimum, maximum, average and standard deviation for R-factor (in Fig. 2f) were 2015.22, 2604.00, 2285.13 and 111.12 MJ mm/ha/yr, respectively. For vegetation index, the NDVI map in Fig. 2g) had its values between 0.9844 and 1.7585. The distribution across the watershed gives information about the degree of greenness and health of vegetation. Locations (pixels) with the lowest NDVI values are likely to be built-up/urban areas with little or no presence of unhealthy vegetation in the watershed. The NDVI values indicate that the region had fairly abundant vegetation that protect the earth's surface. The spatial LST distribution map in Fig. 2h) saw its values range between 9.47°C and 30.61°C, with an average value of 18.13°C. By comparing LST and NDVI maps, it shows that regions with the lowest NDVI values have the highest LST values. This may be a result of land-use change that exposes the earth's surface in those areas [51, 59]. A spatial distribution map of SMI (in Fig. 2i) with its values ranged between 0.8516 and 1.3577 had

Table 2. Correlations matrix for CFs.

		Drainage density	Lineament density	LS-factor	LST	NDVI	R-factor	K-factor	SMI
Drainage density	Pearson Correlation	1							
	Sig. (1-tailed)								
Lineament density	Pearson Correlation	0.037	1						
	Sig. (1-tailed)	0.245							
LS-factor	Pearson Correlation	0.006	0.093*	1					
	Sig. (1-tailed)	0.454	0.043						
LST	Pearson Correlation	-0.235**	-0.289**	0.096*	1				
	Sig. (1-tailed)	0	0	0.038					
NDVI	Pearson Correlation	0.206**	0.190**	-0.192**	-0.518**	1			
	Sig. (1-tailed)	0	0	0	0				
R-factor	Pearson Correlation	-0.420**	-0.284**	0.104*	0.368**	-0.443**	1		
	Sig. (1-tailed)	0	0	0.028	0	0	0		
K-factor	Pearson Correlation	0.245	0.405	-0.225	0.035	0.145	0.045	1	
	Sig. (1-tailed)	0	0	0.01	0	0	0	0	
SMI	0.408	-0.057	-0.033	0.106	0.237	-0.150	0.224	1.000	
	Sig. (1-tailed)	0	0.01	0	0.02	0	0	0	1

\*\*Correlation is significant at the 0.01 level (1-tailed), \*Correlation is significant at the 0.05 level (1-tailed).

Table 3. Variable of importance in MLR.

	$\alpha$	S.E.	Wald	df	Sig.	Exp( $\alpha$ )
Drainage density	0.101	0.373	0.142	1	0.706	1.151
K-factor	1.876	14.422	0.599	1	0.439	6.527
Lineament density	2.039	3.177	0.497	1	0.481	9.388
LS-factor	2.254	0.016	92.978	1	0.000	1.167
LST	0.331	0.067	24.216	1	0.000	1.393
NDVI	-6.278	1.399	20.140	1	0.000	0.002
R-factor	0.102	0.002	0.968	1	0.325	1.002
SMI	1.189	4.346	4.666	1	0.031	3.284
Constant	-14.178	6.105	5.393	1	0.020	-

$\alpha$  = coefficients, S.E. = standard error, Wald = Wald chi-square values, df = degree of freedom, Sig. = Significance, Exp( $\alpha$ ) = Exponential of coefficients.

average and standard deviation values of 1.0965 and 0.0438, respectively. This range of SMI values showed a relatively high soil moisture in the watershed as the value is a bit far from +1. Most locations within the watershed have SMI values greater than 0.3. This suggests that such locations have favorable soil moisture or no drought conditions as classified by Parida, et al. [60].

In MLR modeling, Ozdemir and Altural [58] reported Hosmer and Lemeshow [57] that model fitting is sensitive to multicollinearities among the predictor variables (i.e., CFs). Thus, erosion CFs were analyzed with respect to tolerance and variance inflation factor (VIF) to check for multicollinearities. According to Menard [61], tolerance value smaller than 0.1 or VIF more than 10 for any CFs signifies a serious multicollinearity issue and such should be excluded from the analysis [58]. Diagnostic analysis for multicollinearity in Table 1 showed that there was no multicollinearity issue. Hence, all the CFs were implemented in building the MLR model. Correlation matrix among erosion CFs was evaluated using Pearson correlation method for 1-tailed test statistics. The results are presented in Table 2 showing their respective correlation at 0.01 and 0.05 confidence intervals. The weights for erosion CFs and some other useful statistical parameters are presented in Table 3.

The weights indicate that all the CFs were significant and positively impacting on erosion except NDVI. Since the Cameron Highlands watershed is majorly occupied by forest, this could be responsible for the higher negative NDVI relationship with the occurrence of soil erosion. This is because the forest protects the soil against erosion [2, 25]. LS-factor had the highest positive effect on the occurrence of soil erosion, followed by lineament density. The LS-factor, K-factor and R-factor also played significant roles in the development of erosion. Heavy rainfall characterized by the study area coupled with its topographic nature increases the

aggressiveness of rainfall to initiate erosion [62]. The lower weight of R-factor obtained in the model suggests relatively little influence on the development of erosion. This may be due to higher NDVI values attributed to the study area that provides protection to many locations within the watershed against water-induced erosion. The presence of lineaments in a watershed has been reported by many researchers to trigger erosion [9, 63]. The result of this study conformed with the description of Cameron Highlands [64]. Analysis of soil characteristics showed the presence of two soil types. The weight of the K-factor indicated a positive correlation with the occurrence of soil erosion, although its lower value suggested that it has a relatively low impact on soil erosion. Furthermore, the results also indicate that LST and SMI were significant in triggering erosion.

The Cox and Snell pseudo- $R^2$  and Nagelkerke- $R^2$  values in Table 4 indicate that the CFs account for 52% and 69.5% probability of occurrence of soil erosion. In most studies, Nagelkerke- $R^2$  value often appears higher than the Cox and Snell pseudo- $R^2$ . To further ascertain the model's performance, classification summary of the observed and predicted occurrence of soil erosion was evaluated using 50% of correct prediction as the cut-off value [56]. The classification results show that the predicted accuracies for the eroded and non-eroded were 89.1% and 83.6%, respectively, while the overall predicted accuracy was 86.6%, as presented in Table 4. Analysis of Cohen's kappa index ( $k$ ), whose formula is presented in Equation (6), was carried out to further assess the model's performance. This measures the inter-rater agreement between two raters as to the presence or absence of soil erosion. In other words, it is a metric that compares an observed accuracy with an expected accuracy (random chance). Some of the criteria used for its estimation are true positive (TP), false positive (FP), true negative (TN) and false negative (FN) [11, 65] with respect to Table 4.

Table 4. Classification accuracy measure.

Observed	Predicted		Total	% Correct
	Absent	Present		
Absent	133	26	159	83.6
Present	20	164	184	89.1
Total	153	190	343	
Overall Percentage				86.6

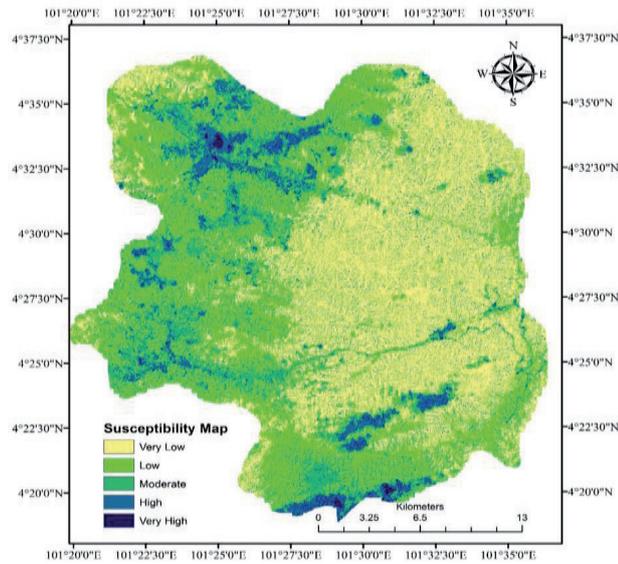


Fig. 3. Soil erosion susceptibility map for Cameron Highlands.

$$k = \frac{OA - AC}{1 - AC} \tag{6}$$

...where observed agreement (OA) = (TP+TN) represents the fraction of pixels/points that are correctly classified as eroded and non-eroded while agreement by

chance (AC) = ((TP+FN)(TP+FP)+(FP+TN)(FN+TN)). This represents the fraction of locations (pixels) for which the agreement is expected by chance [11]. The value of k ranges between 0 and 1, in which the value of 1 indicates a complete agreement and 0 indicates no agreement at all. The analysis of k index yielded a value of 0.73, indicating a substantial agreement between the two raters.

Based on the weights of the CFs, the susceptibility map presenting the probability of occurrence of erosion was produced as shown in Fig. 3. For better presentation and visualization, susceptibility level of the watershed to erosion was classified into five classes using the natural breaks method. The classes are very low, low, moderate, high and very high susceptible zones constituting 3.4%, 71.4%, 24.0%, 1.0% and 0.2% of the total area of Cameron Highlands, respectively. In this study, the soil erosion susceptibility map was validated with the use of the area-under-curve (AUC) method. “This measures the quality of probabilistic model describing its ability to reliably predict the occurrence or non-occurrence of events” [66]. AUC has been widely accepted and adopted by several researchers for evaluating susceptibility mapping accuracy [8, 11, 58, 67, 68]. Its values range between 0 and 1, where its closeness to 1 indicates perfect prediction and closeness to 0.5 indicates that the model is inaccurate [69, 70]. The AUC value for MLR analysis was estimated to be 0.819 with the ROC curve shown in Fig. 4. This indicates that multivariate logistic regression performed excellently in erosion susceptibility assessment under both static and dynamic CFs.

### Conclusion

Adequate soil erosion susceptibility assessment is critical for implementing sustainable watershed management approaches especially in complex watersheds like the Cameron Highlands. The accuracy of erosion susceptibility mapping could be affected by the quality and number of CFs considered in the analysis. Further investigations are thus required in this aspect as many previous studies implemented majorly static CFs and left out dynamic CFs. In order to geospatially predict susceptible locations and develop an accurate susceptibility map, both static and dynamic CFs

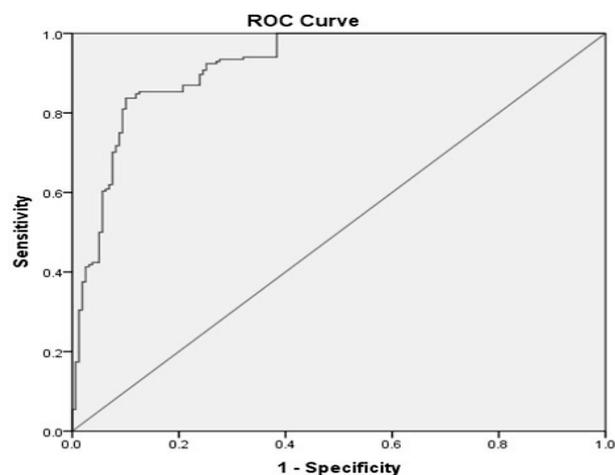


Fig. 4. ROC for validation of the susceptibility map.

were included in the modeling procedure. Redundant CFs were excluded in the analysis as recommended in the literature [14]. Hence drainage density, lineament density, LS-factor and K-factor were considered as static CFs, and LST, SMI, NDVI and R-factor as dynamic CFs. The MLR model, GIS and remote sensing were applied to establish a geospatial relationship between the CFs and the occurrence of soil erosion in the study area. The results showed that the selected CFs were positively impacting on soil erosion with the exception of NDVI. LS-factor had the highest positive effect on the occurrence of soil erosion, followed by lineament density. The values of Cox and Snell pseudo- $R^2$  and Nagelkerke- $R^2$ , respectively, indicated that CFs account for 52% and 69.5% probability of occurrence of soil erosion in the study area. During the classification, predicted accuracies for the eroded and non-eroded points were 89.1% and 83.6%, respectively, with an overall prediction accuracy of 86.6%. This indicates that the model correctly predicted the occurrence of erosion in the study area. Furthermore, analysis of Cohen's kappa index produced a 73% classification agreement between the presence and absence of soil erosion during the modeling, which can be regarded as a substantial agreement. The predictive accuracy of the model was validated by using area-under-curves technique, and the accuracy was found to be 81.9%. This indicates that the medium-scale erosion susceptibility map was accurate and acceptable. The map could be used by watershed planners and decision makers for sustainable development, optimal land-use planning and prevention of erosion.

### Acknowledgements

This research was supported by Universiti Teknologi PETRONAS, Malaysia under the University Research Internal Fund (URIF) 2016 Grant with Cost Centre No: 0153AA-G04. The authors also acknowledge the Malaysian Department of Irrigation and Drainage for providing useful data.

### Conflict of Interest

The authors declare no conflict of interest.

### References

- MARIA K., PANTELIS S., FILIPPOS V. Soil erosion prediction using the revised universal soil loss equation in a GIS framework, Chania, Northwestern Crete, Greece. *Environmental Geology*, **57** (3), 483, **2009**.
- RAHMAN M.R., SHI Z., CHONGFA C. Soil erosion hazard evaluation - an integrated use of remote sensing, GIS and statistical approaches with biophysical parameters towards management strategies. *Ecological Modelling*, **220** (13), 1724, **2009**.
- SUN W., SHAO Q., LIU J., ZHAI J. Assessing the effects of land use and topography on soil erosion on the Loess Plateau in China. *Catena*, **121**, 151, **2014**.
- OCHOA P., FRIES A., MEJIA D., BURNEO J., RUÍZ-SINOJA J., CERDÀ A. Effects of climate, land cover and topography on soil erosion risk in a semiarid basin of the Andes. *Catena*, **140**, 31, **2016**.
- ANGIMA S., STOTT D., O'NEILL M., ONG C., WEESIES G. Soil erosion prediction using RUSLE for central Kenyan highland conditions. *Agriculture, Ecosystems & Environment*, **97** (1), 295, **2003**.
- BRUNNER A., PARK S., RUECKER G., DIKAU R., VLEK P. Catenary soil development influencing erosion susceptibility along a hillslope in Uganda. *Catena*, **58** (1), 01, **2004**.
- VAHABI J., NIKKAMI D. Assessing dominant factors affecting soil erosion using a portable rainfall simulator. *International Journal of Sediment Research*, **23** (4), 376, **2008**.
- AKGÜN A., TÜRK N. Mapping erosion susceptibility by a multivariate statistical method: a case study from the Ayvalik region, NW Turkey. *Computers & Geosciences*, **37** (9), 1515, **2011**.
- LUCÀ F., CONFORTI M., ROBUSTELLI G. Comparison of GIS-based gully susceptibility mapping using bivariate and multivariate statistics: Northern Calabria, South Italy. *Geomorphology*, **134** (3) 297, **2011**.
- AYALEW L., YAMAGISHI H. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology*, **65** (1), 15, **2005**.
- TEHRANY M.S., PRADHAN B., MANSOR S., AHMAD N. Flood susceptibility assessment using GIS-based support vector machine model with different kernel types. *Catena*, **125**, 91, **2015**.
- LEE M.J., KANG J.E., JEON, S. Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS. In *Geoscience and Remote Sensing Symposium (IGARSS)*, 895, **2012**.
- ABDULKADIR T.S., MUHAMMAD M.R., KHAMARUZAMAN W.Y., AHMAD H.M. Geostatistical based susceptibility mapping of soil erosion and optimization of its causative factors: A conceptual framework. *Journal of Engineering Science and Technology*, **12** (11), 2880, **2017**.
- MAGLIULO P. Assessing the susceptibility to water-induced soil erosion using a geomorphological, bivariate statistics-based approach. *Environmental Earth Sciences*, **67** (6), 1801, **2012**.
- CONFORTI M., AUCELLI P.P., ROBUSTELLI G., SCARCIGLIA F. Geomorphology and GIS analysis for mapping gully erosion susceptibility in the Turbolo stream catchment (Northern Calabria, Italy). *Natural hazards*, **56** (3), 881, **2011**.
- WENG T.K., MOKHTAR M.B. Emerging issues towards sustainable river basin management in Cameron Highlands, Malaysia. *Environment and Natural Resources Journal*, **9** (2), 58, **2011**.
- TAMENE L., LE Q.B. Estimating soil erosion in sub-Saharan Africa based on landscape similarity mapping and using the revised universal soil loss equation (RUSLE). *Nutrient Cycling in Agroecosystems*, **102** (1), 17, **2015**.
- JAMIL N.R., RUSLAN M.S., TORIMAN M.E., IDRIS M., RAZAD A.A. Impact of landuse on seasonal water quality at highland lake: A case study of Ringlet Lake,

- Cameron Highlands, Pahang. From Sources to Solution: Springer, 409, **2014**.
19. KHALIK W.M.A.W.M., ABDULLAH M.P., AMERUDIN N.A., PADLI N. Physicochemical analysis on water quality status of Bertam River in Cameron Highlands, Malaysia. *Journal of Materials and Environmental Science*, **4** (4), 488, **2013**.
  20. JENSEN L., LARIYAH M.S., MOHAMED N.M.D., PIERRE Y.J. Challenge in running hydropower as source of clean energy: Ringlet reservoir, Cameron Highlands case study. Proceedings National Graduate Conference, Universiti Tenaga Nasional, November 8-10, Malaysia, **2012**.
  21. JANSEN L., LARIYAH M.S., MOHAMED N.M.D., PIERRE Y.J. Hydropower reservoir for flood control: A case study on Ringlet reservoir, Cameron Highlands, Malaysia. *Journal of Flood Engineering*, **4** (1), 87, **2013**.
  22. TEHRANY M.S., PRADHAN B., JEBUR M.N. Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS. *Journal of Hydrology*, **504**, 69, **2013**.
  23. PRADHAN B. A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, **51**, 350, **2013**.
  24. YALCIN A., REIS S., AYDINOGLU A., YOMRALIOGLU T. A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. *Catena*, **85** (3) 274, **2011**.
  25. KAVZOGLU T., SAHIN E.K., COLKESEN I. Selecting optimal conditioning factors in shallow translational landslide susceptibility mapping using genetic algorithm. *Engineering Geology*, **192**, 101, **2015**.
  26. RAHMATI O., HAGHIZADEH A., POURGHASEMI H.R., NOORMOHAMADI F. Gully erosion susceptibility mapping: The role of GIS-based bivariate statistical models and their comparison. *Natural Hazards*, DOI 10.1007/s11069-016-2239-7, **2016**.
  27. PRADEEP G.S., KRISHNAN M.N., VIJITH H. Identification of critical soil erosion prone areas and annual average soil loss in an upland agricultural watershed of Western Ghats, using analytical hierarchy process (AHP) and RUSLE techniques. *Arabian Journal of Geosciences*, **8** (6), 3697, **2015**.
  28. GAYEN A., SAHA S. Application of weights-of-evidence (WoE) and evidential belief function (EBF) models for the delineation of soil erosion vulnerable zones: a study on Pathro river basin, Jharkhand, India. *Modeling Earth Systems and Environment*, **1**, **2017**.
  29. FERNÁNDEZ S., MARQUÍNEZ J., MENÉNDEZ-DUARTE R. A sapping erosion susceptibility model for the southern Cantabrian Range, North Spain. *Geomorphology*, **95** (3), 145, **2008**.
  30. CONOSCENTI C., AGNESI V., ANGILERI S., CAPPADONIA C., ROTIGLIANO E., MARKER M. A GIS-based approach for gully erosion susceptibility modeling: A test in Sicily, Italy. *Environmental Earth Sciences*, **70**, 1179, **2013**.
  31. ANGILERI S.E., CONOSCENTI C., HOCHSCHILD V., MÄRKER M., ROTIGLIANO E., AGNESI V. Water erosion susceptibility mapping by applying Stochastic Gradient Treeboost to the Imera Meridionale River Basin (Sicily, Italy). *Geomorphology*, **262**, 61, **2016**.
  32. SHIT P.K., PAIRA R., BHUNIA G., MAITI R. Modeling of potential gully erosion hazard using geo-spatial technology at Garbheta block, West Bengal in India. *Modeling Earth Systems and Environment*, **1** (2), 01, **2015**.
  33. ALAA M.A.A., ALI K.A.A. Susceptibility mapping of gully erosion using GIS-based statistical bivariate models: a case study from Ali Al-Gharbi District, Maysan Governorate, southern Iraq. *Environmental Earth Sciences*, **77**, 249, **2018**.
  34. WANG B., ZHENG F., RÖMKENS M.J., DARBOUX F. Soil erodibility for water erosion: A perspective and Chinese experiences. *Geomorphology*, **187**, 1, **2013**.
  35. KARTIC K.M., ANNADURAI R., RAVICHANDRAN P.T. Assessment of soil erosion susceptibility in Kothagiri Taluk using revised universal soil loss equation (RUSLE) and geo-spatial technology. *International Journal of Science Research Publications*, **4** (10), 13, **2014**.
  36. TEHRANY M.S., SHABANI F., DYMPHNA N.J., KUMAR L. Soil erosion susceptibility mapping for current and 2100 climate conditions using evidential belief function and frequency ratio. *Geomatics, Natural Hazards and Risk*, **1**, **2017**.
  37. SINGHAL P.K., SHRIVASTAVA P. Challenges in sustainable development: *Anmol.*, **2004**.
  38. LAL R. Soil degradation by erosion. *Land Degradation & Development*, **12** (6), 519, **2001**.
  39. ABDULKADIR T.S., MUSTAFA M.R., KHAMARUZAMAN Y.W., HASHIM A.M. Evaluation of rainfall-runoff erosivity factor for Cameron Highlands, Pahang, Malaysia. *Journal of Ecological Engineering*, **17** (3), 01, **2016**.
  40. GAO J., WU S. Simulated effects of land cover conversion on the surface energy budget in the Southwest of China. *Energies*, **7** (3), 1251, **2014**.
  41. SOHEILA Y.J. The effect of land use on land surface temperature in the Netherlands. M.Sc Thesis Submitted to Lund University Sweden, **2013**.
  42. MAZEN E.A. Assessing MODIS land surface temperature (LST) over Jeddah. *Polish Journal of Environmental Studies*, **26** (4), 1461, **2017**.
  43. BOORI M.S., VOZENILEK V., BALZTER H., CHOUDHARY K. Land surface temperature with land cover classes in ASTER and Landsat data. *Journal of Geophysics and Remote Sensing*, **4** (1), 1, **2015**.
  44. BOORI M.S., VOŽENÍLEK V., BURIAN J. Land cover disturbance due to tourism in Czech Republic. Proceedings of the Fifth International Conference on Innovations in Bio-Inspired Computing and Applications IBICA, Springer, **2014**.
  45. XUE X., LUO Y., ZHOU X., SHERRY R., JIA X. Climate warming increases soil erosion, carbon and nitrogen loss with biofuel feedstock harvest in tallgrass prairie. *Gcb Bioenergy*, **3** (3), 198, **2011**.
  46. KARNIELI A., BEN-ASHER J. A daily runoff simulation in semi-arid watersheds based on soil water deficit calculations. *Journal of hydrology*, **149** (1-4), 9, **1993**.
  47. PARVEEN R., KUMAR U. Integrated approach of universal soil loss equation (USLE) and geographical information system (GIS) for soil loss risk assessment in Upper South Koel Basin, Jharkhand. *Journal of Geographic Information System*, **4** (6), 588, **2012**.
  48. WAWER R., NOWOCIEN E., PODOLSKI B. Eal and calculated Kusle erodibility factor for selected Polish soils. *Polish Journal of Environmental Studies*, **14** (5), 655, **2005**.

49. WISCHMEIER W.H., SMITH D.D. Rainfall energy and its relationship to soil loss. *Eos, Transactions American Geophysical Union*, **39** (2), 285, **1958**.
50. TAN K., LIAO Z., DU P., WU L. Land surface temperature retrieval from Landsat 8 data and validation with geosensor network, *Frontiers of Earth Science*, **11**, 20, **2017**.
51. ABDULKADIR T.S., MUHAMMAD M.R., KHAMARUZAMAN W.Y., AHMAD H.M. Assessing the influence of terrain characteristics on spatial distribution of satellite derived land surface parameters in mountainous areas. 37th IAHR World Congress, 13-18th August, Kuala Lumpur Malaysia, **37** 2897, **2017**.
52. TEERAWONG L., TORSACK G., TANUTDECH R. Application of remote sensing for temperature monitoring: The technique for land surface temperature analysis. *Journal of Ecological Engineering*, **18** (3), 53, **2017**.
53. SANDHOLT I., RASMUSSEN K., ANDERSEN J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of environment*, **79** (2), 213, **2002**.
54. SRUTHI S., ASLAM M.M. Agricultural drought analysis using the NDVI and land surface temperature data; a case study of Raichur district. *Aquatic Procedia*, **4**, 1258, **2015**.
55. POTIĆ I., BUGARSKI M., MATIĆ-Varenica J. Soil moisture determination using remote sensing data for the property protection and increase of agriculture production. *Responsible Land Governance: Towards an Evidence Based Approach*. The World Bank, Washington DC, U.S. **2017**.
56. CONOSCENTI C., ANGILERI S., CAPPADONIA C., ROTIGLIANO E., AGNESI V., MÄRKER M. Gully erosion susceptibility assessment by means of GIS-based logistic regression: a case of Sicily (Italy). *Geomorphology*, **204**, 399, **2014**.
57. HOSMER D., LEMESHOW S. *Applied logistic regression*. 2<sup>nd</sup> ed. New York: John Wiley and Sons, **2000**.
58. OZDEMIR A., ALTURAL T. A comparative study of frequency ratio, weights of evidence and logistic regression methods for landslide susceptibility mapping: Sultan Mountains, SW Turkey. *Journal of Asian Earth Sciences*, **64**, 180, **2013**.
59. DENG Y., WANG S., BAI X., TIAN Y., WU L., XIAO J., CHEN F., QIAN Q. Relationship among land surface temperature and LUCC, NDVI in typical karst area. *Scientific Reports*, **8** (1), 641, **2018**.
60. PARIDA B.R., COLLADO W.B., BORAH R., HAZARIKA M.K., SAMARAKOON L. Detecting drought prone areas of rice agriculture using a MODIS-derived soil moisture index. *GIScience & Remote Sensing*, **45** (1), 109, **2008**.
61. MENARD S. *Logistic Regression: From introductory to advanced concepts and applications*. Thousand Oaks, CA, Sage, **2010**.
62. ABDULLAH J. *Highlands Developments in Malaysia. Globalization and marginalization in mountain regions*. Springer, 147, **2016**.
63. KACHOURI S., ACHOUR H., ABIDA H., BOUAZIZ S. Soil erosion hazard mapping using Analytic Hierarchy Process and logistic regression: a case study of Haffouz watershed, central Tunisia. *Arabian Journal of Geosciences*, **8** (6), 4257, **2015**.
64. BARROW C., NGAI WENG C., MASRON T. Issues and challenges of sustainable agriculture in the Cameron Highlands. *Malaysian Journal of Environmental Management*, **10** (2), 89, **2009**.
65. GUZZETTI F., CARRARA A., CARDINALI M., REICHENBACH P. Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, **31** (1), 181, **1999**.
66. NANDI A., SHAKOOR A. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology*, **110** (1), 11, **2010**.
67. ALTHUWAYNEE O.F., PRADHAN B., LEE S. Application of an evidential belief function model in landslide susceptibility mapping. *Computers & Geosciences*, **44**, 120, **2012**.
68. MAHALINGAM R., OLSEN M.J., O'BANION M.S. Evaluation of landslide susceptibility mapping techniques using lidar-derived conditioning factors (Oregon case study). *Geomatics, Natural Hazards and Risk*, **7** (6), 1884, **2016**.
69. BUI D.T., PRADHAN B., LOFMAN O., REVHAUG I., DICK O.B. Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Computers & Geosciences*, **45**, 199, **2012**.
70. JEBUR M.N., PRADHAN B., TEHRANY M.S. Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment*, **152**, 150, **2014**.