An empirical approach to estimate soil erosion and its correlation with landslide events (a case study from Vazroud watershed, N Iran)

Jalal Zandi, Karim Solaimani and Mahmood Habib Nezhad

Abstract: Soil erosion is one of the most important natural phenomena in the world, and it poses a significant threat in Iran, due to land degradation and desertification. To cope with this problem, essentially, we must determine the erosion-prone areas. Many studies have been carried out and different models and methods have been applied to this problem. In this study, we used an empirical method (RUSLE) to prepare an erosion susceptibility map for the Vazroud watershed, Mazandaran Province, Iran. A landslide location map was generated on the basis of satellite image elements interpretation and field observations were used to validate soil erosion intensity in the study area. A statistical-based frequency ratio analysis and “Receiver Operating Characteristic” (ROC) was conducted in the study area for correlation purposes. The results of the statistical correlation showed a satisfactory agreement between the prepared RUSLE-based soil erosion map and landslide distribution. To calculate the prediction accuracy of the generated erosion susceptibility map, we compared it with a map of landslide event locations. Based on this comparison, the area under curvature (AUC) value was 0.716. These results revealed that the RUSLE model used in this study is able to satisfactorily predict landslides event.

Keywords: Soil erosion, Landslide, ROC curve, RUSLE model, GIS

Introduction
Erosion is the detachment and transportation of material from a surface and it occurs whenever the eroding or driving forces generated by water and wind exceed the resisting forces of rock soils (Morgan, 2005). Water erosion is a major problem because of its social and economic effects. It is responsible both for direct and indirect damages and it can reactivate surficial landslides by locally increasing the steepness of slopes (Conoscenti et al., 2008). The most serious type of water erosion is in mountain areas, especially erosion from unstable and loose geologic material, resulting to gully erosion and the mass movement of soil and rock (e.g., Selby, 1982; Lee, 2004). For these reasons, there is a great interest in producing accurate maps of active water erosion areas and sediment sources, especially at the basin scale, which better corresponds to the needs of environmental decision making (Begueria, 2006). Erosion susceptibility maps are also land resource evaluations. Such maps classify land into areas with similar erosion susceptibility. Predicting the location of highly susceptible areas to erosion is the most important part of erosion prevention, as it allows the identification of the proper location and type of prevention measures needed. In literature, there are many studies presenting erosion susceptibility mapping techniques by different methods (e.g., Conoscenti et al., 2008; Lesschen et al., 2008; Mueller et al., 2005; Begueria, 2006; Evrard et al., 2007; Zandi, 2012). These are empirical and physically based methods. The empirical methods estimate erosion by combining a pre-fixed set of physical parameters, based on certain standardized coefficients or procedures. A good example is the Universal Soil Loss Equation (USLE) method (Wischmeier and Smith, 1965). The physically based methods (e.g., the WEPP model, Nearing et al., 1989) mathematically describe the process of detachment, transportation and deposition of the eroded soil. Whereas the empirical methods can be applied at a basin scale, the physically based methods are not suitable for basin scale studies because they require a large and extremely detailed set of parameters. More recently, sophisticated empirical models for mapping erosion susceptibility by Geographical Information Systems (GIS) can be found in the literature (e.g., Qing et al., 2008; Park et al., 2011; Oliveira et al., 2011; Fernandez and Margarita, 2011). The RUSLE model has been used, validated and calibrated in several parts of humid and semi-humid regions of Iran (Asadi et al., 2010). Soil erosion is the main cause of land sliding in the study area and controlling the erosion becomes essential in order to prevent landslides (Zandi, 2012). By adjusting the factors which are responsible for soil erosion like land use/cover, the rate of soil erosion can be minimized. Thus, in this study, soil erosion spatial assessment is needed. GIS is an efficient tool to integrate various datasets and assess any dynamic system such as soil erosion (Pradhan et al., 2012; Zandi, 2012). Since it is important to study soil erosion (as one of the major environmental threats), it is proposed to assess the soil erosion and correlate it with the other equally important and related threat (landslides) in Vazroud watershed. In particular, the objectives of this study were: (i) Assess the soil erosion; (ii) Correlate with “events landslide locations” in the absence of ground data at ungaged ecosystems; (iii) Investigation on the ability of model to predict occurrence or non-occurrence of landslides by applying ROC curve.

Martial mad Methods
Description of study area
Vazroud watershed has an area of 14123 ha is located in central Mazandaran province, Iran (Fig. 1). Erosion status study in Vazroud is inevitable owing to specific importance to provide urban water, promenade trait, in the other hand, land use changes in forest to habitat and incorrect use of agricultural lands such as: housing, infrastructure and roads development and mining.
This particular area was selected for frequent destroyer landslide occurrences and soil erosion problems in upstream of watershed, specially.

Altitudes of the area ranges between 270 and 3580 m (asl), slope gradients ranges between 0 and 66 degree with an average of 26.74 degree. Dense vegetation covers the lower altitude and gentler slopes, while vegetation is sparse at steeper slopes and higher altitudes. The mean annual precipitation and temperature are 600 mm and 10.6 °C, respectively. Six meteorological stations located within and near the study area with base statistical common data were used (Joorband, Vaz, Chamestan, Lavij, Takker and Razan) for the period 1987-2007. The stations location is shown in (Fig. 1).

Data and methods

The first phase of the research was mapping soil erosion in study area by means of RUSLE model and then, correlates this map with landslide event locations. The RUSLE model was used to estimate the average annual soil loss, which is expressed as mass per unit area per year (tons ha\(^{-1}\) yr\(^{-1}\)). RUSLE model (Renard et al., 1997), has been applied to assess soil erosion over extended areas and in different contexts, including forests, rangeland, and disturbed areas (Terranova et al., 2009). RUSLE governing equation is a function of 6 input factors involving: rainfall erosivity, soil erodibility, slope length and slope steepness, cover, and management and conservation practices. The RUSLE was applied in a GIS medium to determine the average annual soil loss and its distribution in the study area.

\[ A = R \times K \times L \times S \times C \times P \]  

Where A is the average soil loss per unit area by erosion (t ha\(^{-1}\) year\(^{-1}\)), R is the rainfall erosivity factor (MJ mm ha\(^{-1}\) h\(^{-1}\) year\(^{-1}\)), K is the soil erodibility factor (t ha h MJ\(^{-1}\) ha\(^{-1}\) mm\(^{-1}\)), L is the slope length factor, S is the steepness factor, C is the cover and management practice factor, P is the conservation support practice factor. The L, S, C, and P values are dimensionless.

Rainfall runoff erosive factor (R)

The erosivity factor R (MJ mm ha\(^{-1}\) h\(^{-1}\) year\(^{-1}\)) was calculated according to the equation given by (Zandi, 2012), derived from a spatial regression analysis which obtained from synoptic stations of Mazandaran Province, based on the easily available mean annual rainfall (P in mm). The R-factor is given by a regression equation as:

\[ R = -8.12 + 0.562P \]  

To determine the value of the R-factor in this study, the average of annual historic rainfall event (1987-2007) was collected from six meteorological stations located within and near the study area. Using spatial analyst extension in ArcGIS (Ver. 10, ESRI), Spline interpolation was done to generate an estimated surface from these scattered set of point data (Fig. 2).

Soil erodibility factor (K)

To evaluate this factor, interested network including nested – systematic designed with taking soil samples from 10 cm depth of different type of soil map. The value of the K-factor was calculated using the following formulas (Renard et al. 1997):

\[ K = 7.594\{0.0034 + 0.0405\exp(-1/2(((\log D_g + 1.659))/0.7101)^2)\} \]  

\[ D_g = \exp(0.01\sum f_i \ln m_i) \]

Where \( D_g \) is the geometric mean diameters of soil particle, \( m_i \) is the arithmetic mean of the particle size limits of class \( i \), \( f_i \) is the particle size fraction in percent of class \( i \). Soil information data were put into the attribute database of the soil map. Then the
erodibility K-factor (t ha h MJ⁻¹ ha⁻¹ mm⁻¹) were calculated for each soil mapping unit using formula (3) and (4) in GIS. (Fig. 2) illustrates the spatial distribution of the soil erodibility factor K when interpolate with Kriging method.

**Topographic factors L and S**

Methodology that applied in this study is a program available which automatically processes the DEM input to compute the LS-factor (van Remortel et al. 2004). The program was originally written in Arc Macro Language (AML) (Hickey, 2000) and has been updated in 2004 with a free compiler of C++ programming language to be more efficient in processing. The command for calculating the L factor is based on (Eq. 5). The C++ executable program computes the cumulative slope lengths and substitutes this value as λ (Fig. 2).

\[
L = \left(\frac{\lambda}{22.13}\right)^m
\]

\[
m = \beta((\beta + 1)) \quad \text{AND} \quad \beta = \frac{(\sin \theta 0.0896)}{(3 \times (\sin \theta )^{0.8 + 0.56})}
\]  

The slope length exponent \( m \) is related to the variable \( \beta \), which is a ratio of rill erosion (defined as erosion caused by overland flow) to interrill erosion (defined as erosion caused by rainfall). The method to calculate the exponent \( m \) is given by equation [6] (McCool, 1987).

**Cover and management practices factor (C)**

In order to estimating \( C \) factor, Sample values were collected from various land cover at 20 locations of watershed area randomly and register with GPS. The \( C \) factor observation values range from nearing 0.35, occur on bare land with little vegetation and high erosion, while low values, of less than 0.1, are found in areas of dense forest or grain cover with low erosion.

With this assumption that there exists a linear correlation between NDVI and \( C \) factor, correlation equation were obtained and used as equation transform (Zandi, 2012). NDVI map derived from the near infrared band and red band of the TM Landsat at 2010/6/4.

\[
\text{NDVI} = \frac{(b4 - b3)}{(b4 + b3)}
\]  

**Conservation practices factor (P)**

The value for \( P \) is taken by using practices analogy since the data is not readily available for study area. Therefore, in the absence of conservation practices data in this study area, P factor was assumed as unity (Asadi et al., 2010; Pradhan et al., 2012).

**Landslide inventory map**

The maps show the locations and properties of landslides that have moved in the past. These slope failures were related to soil erodibility, topographical, and climatic conditions, thus, they can often facilitate the prediction of locations and conditions of future landslides. For this reason, it is important to determine the location and area of the landslide accurately when preparing the landslide susceptibility maps. Landslide inventory mapping is the systematic mapping of existing landslides in a region using different techniques such as field survey, air photo/satellite image interpretation, and literature search for historical landslide records. A landslide inventory map provides the spatial distribution of locations of existing landslides. The landslides in the study area were determined by comprehensive field surveys. The landslides which are currently indefinite in characteristics and boundaries were identified using old dated satellite images. In this study, Landslide inventory map were obtained from field studies, a previous inventory map, and satellite image analyses from IRS P5 (Fig. 5).

**Results**

**Soil erosion mapping**

The five parameters layers were all converted into grid diagrams with 30 m×30 m cell in a uniform coordinate system. Then the GIS input layers were multiplied, as described by the RUSLE, to estimate annual soil loss on a pixel-by-pixel basis, and the spatial distributions of the soil erosion in the study area \( A \) (in t/ha/y) was obtained (Fig. 2). The values of the \( R, K, LS, \) and \( C \) factor are shown in Table 1.
The average annual R factor value varies from 249 to 468 MJ mm h\(^{-1}\) year\(^{-1}\) and the mean value is 382 MJ mm h\(^{-1}\) h\(^{-1}\) year\(^{-1}\) (Fig. 2). The standard deviation is 58.73. According to the map, there is more rainfall erosivity in the north and middle of the watershed than that in the south which has a close relation with the decreasing trend of rainfall from the south to the north and non-uniformity of spatial distribution of rainfall in the Vazroud watershed.

The average K value in the study area varies from 0.03 to 0.06 and the mean value is 0.048 t ha h MJ\(^{-1}\) ha\(^{-1}\) mm\(^{-1}\) (Fig. 2). The standard deviation is 0.005. It can be seen from the soil erodibility map that the K factor value is higher in the south and southwest, except for some particular places.

The Vazroud watershed is a mountainous catchment with high elevation variation and characterized by increasing elevation values from north to south, with a maximum drop of 3580 m. The southwest area of the watershed has the highest variability in elevation, the steepest slopes and, as a consequence, the greatest LS values. It can be seen from (Fig. 2) that the LS factor value in the study area varies from 0.001 to 132 and the mean value is 15.03, the majority of the study area has LS values of less than 10. Some areas have LS values of greater than 20; this area does have steep slopes, such as along the river in the middle of the watershed.

The graphs of regression analysis and C factor are given in Fig. 3. R shows the correlation coefficient of regression analysis.
The final C factor spatial distribution map was generated using the regression equation [8] in Spatial Analyst tool of ArcGIS 10 software (Fig. 3).

\[
C \text{ factor} = 0.407 \cdot 0.5953 \times \text{NDVI}
\]

[8]  

NDVI map derived from TM Landsat (eq. 7) and the C-factor value varies from 0 to 0.35 and the mean value is 0.11 (Fig. 3). Owing to the larger area of bare land and rangeland located in the hillside, edge of the valley, the higher C-factor value occurs in that area also. Due to the absence of conservation practices data in this study area, P factor was assumed as 1 in total watershed area.

### Annual soil loss

The average annual soil loss in the Vazroud watershed was computed by overlaying the five factor maps using RUSLE. As seen in (Fig. 2), the average annual soil loss in most of the area is between 15 and 162 t ha\(^{-1}\) year\(^{-1}\) and the mean value is 26 t ha\(^{-1}\) year\(^{-1}\). As regard to the spatial variation, the southwest part of the watershed, some specific area in excess of 200 t ha\(^{-1}\) year\(^{-1}\), has more erosion than the southeast part.

### Table 2 Area and amount of soil loss of each soil erosion risk category

<table>
<thead>
<tr>
<th>Erosion categories</th>
<th>Numeric range (t ha(^{-1}) year(^{-1}))</th>
<th>Area (ha)</th>
<th>Area percentage (%)</th>
<th>Soil loss (×100 t year(^{-1}))</th>
<th>Soil loss Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>&lt;5</td>
<td>3792</td>
<td>26.9</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Low</td>
<td>5 - 10</td>
<td>1436</td>
<td>10.2</td>
<td>7.5</td>
<td>5.9</td>
</tr>
<tr>
<td>Moderate</td>
<td>10 - 20</td>
<td>2527</td>
<td>17.9</td>
<td>14.8</td>
<td>11.6</td>
</tr>
<tr>
<td>High</td>
<td>20 - 40</td>
<td>3275</td>
<td>23.2</td>
<td>28.1</td>
<td>22.1</td>
</tr>
<tr>
<td>Very High</td>
<td>&gt;40</td>
<td>3051</td>
<td>21.7</td>
<td>75.7</td>
<td>59.7</td>
</tr>
</tbody>
</table>

Erosion categories and numeric range were classified according to extensive field observation and collected sample from study area and Google Earth satellite images.

### Assessment on soil erosion risk zone

The soil erosion map has a continuous scale of numerical values and there is a need to separate these values into susceptibility classes. There are several mathematical methods for the classification susceptibility degrees (Ayalew et al., 2004; Suzen and Doyuran, 2004). The standard deviation classifier was used in this study after surveying the ground realities, which obtained through field visits and Google Earth Satellite images. Most areas of the watershed fall within the minimal (27%), which are mostly seen in the north and near outlet of the watershed. About 4% of the watersheds are under high to extreme erosion risk, which are mostly found in the south and southwest of the watershed (Fig. 5). As regard to the gross amount of soil loss, about 6% of the total soil loss occurs in the area of minimal to low erosion and nearly 70% occurs in the area of high to extreme erosion (Table 2). So management practices should be adopted in the areas of high to extreme risk erosion in order to reduce soil loss.
Validation of the erosion susceptibility map

Soil erosion is in relation with topography, vegetation cover; soil erodibility, rainfall and land use (Beskow et al., 2009). Each type of erosion is also one stage of the other type of erosion, in the other word, the appearance of each type of erosion helps to other type to be occurring (Refahi, 2008). As there is no facility/infrastructure to validate soil erosion intensity values in the study area, landslides locations maps, generated from: P5 sensor of IRS satellite imagery (2.5m spatial accuracy), previous inventory and extensive field survey, was used to correlate the soil erosion intensity map (Pradhan et al., 2011). Landslide locations occurred during the past 20 years. Ninety-nineteen landslides polygons were digitized. The pixel size of the landslide inventory and all parameter maps was 30 m. Landslides areas are overlaid with the soil erosion map of 2010 and is shown in Fig. 5. Soil erosion map was correlated with the help of frequency ratio-based statistical analysis. Frequency ratio approaches are based on the observed relationships between distribution of landslides and soil erosion intensity to reveal the correlation between these two related phenomena in the study area. The frequency was calculated from the analysis of the relationship between landslides and the attribute factors as given in Table 3. The frequency ratios of each soil erosion ranges were calculated from their relationship with landslide events. In the relationship analysis, the ratio is that of the area where landslides occurred to a particular area of erosion prone zone. A value of 1 is an average value. If the value is greater than 1, it means a higher correlation, and a value lower than 1 means lower correlation of occurring soil erosion (Pradhan et al., 2011).

Frequency ratio computed for Vazroud watershed is shown in Table 3. The relationship between soil erosion and landslides occurrence shows that very high erosion zone has higher probability of landslides. For minimal and low erosion zone, frequency ratio is <0.8, which indicates it has very low probability of landslide occurrences. For very high erosion zones, frequency ratio are found to be >1.8, which indicates they have more probability of landslide occurrences. Fig. 6 shows the graphical representation of frequency ratio to each soil erosion prone zone. Perusal of the graph indicates a linear relationship between frequency ratio and soil erosion zones. This correlation result shows satisfactory agreement between the erosion map and the landslide events/location data.

Correlation of soil erosion map with landslides events

The final RUSLE map was checked by overlaying it with the landslide inventory map. A landslide was considered a “correct” prediction when at least part of it was situated within a high probability value, for which 0.5 was used as the cut-off value (Dai and Lee, 2002). Otherwise, the prediction was considered to be “wrong.” Based on the above criteria, the 891 landslides pixels in the study area were predicted by the model as shown in Table 4. The results indicated that 689 of the 891 (77.33%) observed landslides were correctly predicted, and 202 of the 891 pixels observed landslides were incorrectly predicted. The accuracy of the model was
also evaluated by calculating the Relative Operating Characteristics (ROC) method. In the ROC method, the area under the ROC curve values range from 0.5 to 1.0 and are used to evaluate the accuracy of the model. To apply the ROC method to the study area, a concise and representative dataset was prepared using selected pixels from landslides (891 pixels) and randomly selected non-landslide (891 pixels) locations in the investigated area. The prediction capability of each model is determined by the area under the curve (Fig. 7). The value for the area under the ROC curve was found to be 0.716, with an estimated standard error of 0.013 (Fig. 7). The results of this correlation showed a relatively satisfactory agreement between the soil erosion intensity map and landslide events data.

**Table 3** Frequency ratio values of landslide occurrences vs. soil erosion intensity map of 2010

<table>
<thead>
<tr>
<th>Soil erosion level</th>
<th>Pixel in domain</th>
<th>% of total area (a)</th>
<th>% of landslide area (b)</th>
<th>Frequency ratio (b/a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>41791</td>
<td>26.7</td>
<td>21.2</td>
<td>0.8</td>
</tr>
<tr>
<td>Low</td>
<td>16557</td>
<td>10.6</td>
<td>4.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Moderate</td>
<td>28347</td>
<td>18.1</td>
<td>12.34</td>
<td>0.7</td>
</tr>
<tr>
<td>High</td>
<td>36440</td>
<td>23.3</td>
<td>22.04</td>
<td>0.9</td>
</tr>
<tr>
<td>Very High</td>
<td>33374</td>
<td>21.3</td>
<td>40.1</td>
<td>1.8</td>
</tr>
<tr>
<td>Total</td>
<td>156509</td>
<td>100.0</td>
<td>100.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

**Fig 6** Frequency ratio analysis of soil erosion map of 2010 with landslides

**Fig 7** ROC curve evaluation for RUSLE model to prediction landslides
The value is 26 t ha$^{-1}$.

Spatial distribution of soil erosion. The average annual soil loss in most of the area is between 15 and 162 t ha$^{-1}$ year$^{-1}$. According to the gross amount of soil loss, about 6% of the total soil loss occurs in the area of minimal to low erosion and nearly 70% occurs in the area of high to extreme erosion. So management practices should be adopted in the areas of high erosion risk in order to reduce soil loss.

In this study, the landslide locations areas mapped for the purpose of the validation stage were used to check the accuracy of the erosion susceptibility map since no map showing the erosion susceptibility of the study area was available. The relationship between soil erosion and landslides occurrence shows that very high erosion zone has higher probability of landslides. The AUC defines the quality of the probabilistic model by describing its ability to reliably predict the occurrence or non-occurrence event. The value for the area under the ROC curve was found to be 0.716, with an estimated standard error of 0.013. These results revealed that the map obtained is acceptable for a medium-scale erosion and landslide susceptibility map, and that this kind of data analysis has applications in both local and regional planning. This result was later agreed by Pradhan et al. (2012). The integrated approach presented is relatively easy, fast, and straightforward, showing good potential for successful application in other areas.

Acknowledgment

Our thanks to Zandi S, for plentiful discussions and help with the graphics and preliminary analyses of this paper, as well as to the natural resources office of Mazandaran for questionnaires that give to us.

Table 4 Area Under Curve Test Result Variable(s): Landslides and Erosion (RUSLE)

<table>
<thead>
<tr>
<th>Area</th>
<th>Std. Error</th>
<th>Asymptotic Sig.</th>
<th>Asymptotic 95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Lower Bound</td>
</tr>
<tr>
<td>0.716</td>
<td>0.013</td>
<td>0.000</td>
<td>0.691</td>
</tr>
</tbody>
</table>

The test result variable(s): Landslide & Erosion has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Conclusion

In the study area, erosion occurs in several forms. The most visible form is landslide erosion. Adopting RUSLE and GIS, this study developed and applied a simple methodology to predict soil loss and soil erosion risk at a large watershed scale and determined spatial distribution of soil erosion. The average annual soil loss in most of the area is between 15 and 162 t ha$^{-1}$ year$^{-1}$ and the mean value is 26 t ha$^{-1}$ year$^{-1}$. According to the gross amount of soil loss, about 6% of the total soil loss occurs in the area of minimal to low erosion and nearly 70% occurs in the area of high to extreme erosion. So management practices should be adopted in the areas of high erosion risk in order to reduce soil loss.

The test result variable(s): Landslide & Erosion has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Acknowledgment

Our thanks to Zandi S, for plentiful discussions and help with the graphics and preliminary analyses of this paper, as well as to the natural resources office of Mazandaran for questionnaires that give to us.

Reference


