Identification of Eroded and Erosion Risk Areas Using Remote Sensing and GIS in the Quebrada Seca watershed

Identificación de áreas erosionadas y en riesgo de erosión mediante percepción remota y SIG en la microcuenca Quebrada Seca

Cristopher E. Camargo-Roa, Carlos E. Pacheco-Angulo, Sergio A. Monjardin-Armenta, Roberto López-Falcón and Tatiana Gómez-Orgulloso

ABSTRACT

The aim of this research was to identify eroded areas and areas at risk of erosion (EAER) as indicators of soil degradation by water erosion in a semiarid watershed of the Venezuelan Andes in 2017. To this effect, remote sensing techniques and geographic information systems (GIS) were used, focusing on spectral reflectance data from a satellite image, given the absence of continuous pluviographic information and data on soil properties in developing countries. This methodology involved estimating the potential water erosion risk (PWER) and mapping eroded and erosion risk areas (EAER) based on calculating the spectral Euclidean distance to bare soils and a remote sensing technique, which was selected via linear regression. Receiver operating characteristics (ROC) curves were determined to define classification thresholds, which were validated by means of a supervised classification and associated to PWER values. The main results indicate that EAER1 identified more eroded areas with bare soils (229.77 ha) as opposed to EAER2 (195.57 ha). Similarly, it was evident that the first alternative was more successful that the second (sum of the first three principal components). The PWER analysis, in addition to the erosion mapping developed and other data and criteria, such as minimum area size of interest, could help to consider necessary soil conservation measures.

Keywords: spectral Euclidean distance, vegetation indices, principal components analysis, maximum likelihood

RESUMEN

El objetivo de esta investigación fue identificar áreas erosionadas y en riesgo de erosión (AEER) como indicadores de degradación de suelos por erosión hídrica en una cuenca semiárida de los Andes venezolanos en el año 2017. Para ello, se emplearon técnicas de percepción remota y sistemas de información geográfica (SIG), enfocándose en los datos espectrales de reflectancia de una imagen satelital, dada la ausencia de información pluviográfica continua y datos de propiedades del suelo en países en vías de desarrollo. Esta metodología implicó la estimación del riesgo potencial de erosión hídrica (RPEH) y la generación de cartografía de áreas erosionadas y en riesgo (AEER) a partir del cálculo de distancia espectral euclidiana a suelos desnudos y de una técnica de percepción remota seleccionada mediante regresión lineal. Se determinaron curvas ROC (características operativas del receptor) para definir umbrales de clasificación, los cuales fueron validados mediante una clasificación supervisada y asociados a valores de RPEH. Los resultados principales indican que EAER1 identificó más áreas erosionadas con suelos desnudos (229.77 ha) a diferencia de EAER2 (195.57 ha). De igual modo, se evidenció que la primera alternativa tuvo mayores aciertos en contraste con la segunda (sumatoria de los tres primeros componentes principales). El análisis de RPEH, además de las cartografías de erosión desarrolladas y otros datos y criterios como el tamaño del área mínima de interés, podrían ayudar a considerar medidas necesarias en cuanto a conservación de suelos.

Palabras clave: distancia espectral euclidiana, índices de vegetación, análisis de componentes principales, máxima verosimilitud

Introduction

Water erosion, a soil degradation process also regarded as an environmental hazard (Mohammed et al., 2020; Duguma, 2022), is caused by precipitations falling on vulnerable bare terrain, which, as runoff across the slope, drag the soil along to finally deposit it in low areas or mire and obstruct bodies of water (Ávila and Ávila, 2015; Omuto and Vargas, 2019). This makes it the main soil degradation process, as it quantitatively and qualitatively affects the rootable volume of soils intended for agricultural production (Morales-Pavón et al., 2016) and contributes...
to the decline of many other essential ecosystem services (Chaudhary and Kumar, 2018; FAO, 2019).

A way to observe water erosion is through multispectral images captured via remote sensing. From this perspective, eroded soils are characterized by a spectral response similar to that of bare soils, i.e., much more uniform than that of vegetation, which exhibits a flatter reflectivity curve (Chuvieco, 2016), thus indicating the existence of bare soils or parent material outcrops as an effective indicator of areas subjected to erosion (Beguería, 2006). In light of the above, remote sensing techniques and geographical information system (GIS) procedures become necessary. These allow obtaining the spatial and temporal distribution of the diverse factors involved, along with their classifications (Rosales-Rodriguez, 2021). Thereupon, it is also necessary to perform visual and statistical analyses in order to understand and validate the generated cartography, whose purpose is to obtain a more precise and reliable cartographical indicator.

Some remote sensing techniques focus on visual quality by trying to improve the location of the data for analysis, in such a way that the features of interest are more evident (e.g., contrast expansion, color composition, and filtering) (Lillesand et al., 2015; Chuvieco, 2016). Other techniques aim to generate continuous variables, such as vegetation indices, which have proven to be efficient at evaluating soil degradation—with erosion among them (Ngandam et al., 2016)—by transforming them into Net Primary Productivity (NPP) values (Sartori et al., 2018), which has led them to be more frequently employed as quantitative indicators of ecosystem functioning (Orr et al., 2017). In the same way, they have been used to monitor vegetation in arid and semiarid lands (Najafi et al., 2020), as well as to generate the C-factor for models such as USLE (Universal Soil Loss Equation) (Meinen and Robinson, 2021).

Other techniques have also been created, such as i) soil indices (IS) used to estimate soil degradation types (Ngandam et al., 2016; Li and Chen, 2018); ii) principal components analysis (PCA) to discriminate types of bare soils or landslides (Romero et al., 2017; Basu et al., 2020); and iii) spectral mixture analysis (SMA) to map the C-factor or determine bare soils in order to construct vulnerability indicators (Demaria and Aguado, 2013).

In this sense, all these techniques may be used as steps prior to highlighting erosion and later allowing its semi-automatized classification in order to obtain a risk map (Duguma, 2022). These techniques also aim to overcome the tedious task of visually interpreting satellite images, as well as the considerable amount of time required to carry it out (Leal et al., 2018).

By making a special emphasis on erosion risk mapping, which usually indicates the relative probability of it taking place within a certain area as compared to others (Ganasri and Ramesh, 2016; Opeyemi et al., 2019), a distinction can be made between potential, defined as the maximum possible soil loss in the absence of vegetation cover and conservation practices (i.e., only considering the interaction between the physical factors of the soil: soil erodibility, rain erosivity, and topography), and actual, which is determined based on the sum of the land cover/use factor and the previous ones (Plambeck, 2020). The former tends to be substantially higher than the latter (Drzewiecki et al., 2014).

Therefore, the importance of remote sensing techniques to map erosion and its risk has become evident. In the same way, these methods have proven to be valuable for generating a cartography of soil degradation in climate change studies. This has been demonstrated in global products, such as soil degradation assessment (GLADA) (Anderson and Johnson, 2016); in regional ones, with erosion risk modeling in Europe (Panagos et al., 2015); and in national ones, with Malawi’s atlas of soil loss (Omuto and Vargas, 2019). In the same way, some of these techniques have been used to monitor changes in cover or NPP and have been proposed within the analysis of land degradation neutrality (LDN) (Orr et al., 2017), whose sustainable development goal (SDG) for 2030 is to tackle desertification, rehabilitate degraded land and soil, and strive towards a world with “neutral soil degradation” (UNGA, 2015).

Identifying areas that have been eroded or are at risk of erosion also allows government agencies (via cross-referencing with population density maps) to initiate rehabilitation and protection activities, to perform territory planning for environmentally sustainable socioeconomic development, and to determine areas that are susceptible to hillslope processes (Ngandam et al., 2016; Efiong et al., 2021). Additionally, every action taken to address soil degradation may simultaneously contribute to the objectives of the fight against climate change, to the preservation of biological diversity, and to the SDGs (Orr et al., 2017).

In light of the aforementioned ideas, the main objective of this study was to map areas that are eroded and at risk of erosion using remote sensing techniques and GIS procedures, given that many models developed for this type of estimation cannot be adequately executed because data are missing to complete their parameters, e.g., soil erodibility, which requires properties such as structure and permeability, among others, which, in practice, are usually scarce or nonexistent for many parts of the world (Avila and Avila, 2015; United Nations, 2021). All results were obtained at a watershed located in a semiarid environment, a space where soils often stand out for their susceptibility to hydric erosion (Tsegaye et al., 2020).

Materials and methods

Study area: the studied watershed is located between the 72°26'43” - 72°21'23”O and 7°58'29” - 7°54'11”N. Its area is 37,31 km² (Figure 1), and its altitude oscillates between 253 and 1 622 m.
Resources: the resources used for generating EAER and PWER are shown in Table 1.

Table 1. Resources

<table>
<thead>
<tr>
<th>Data input</th>
<th>Denomination</th>
<th>Level of detail</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satellite image</td>
<td>Landsat 8</td>
<td>SR = 30 m</td>
<td>USGS</td>
</tr>
<tr>
<td>Satellite image</td>
<td>Digital Globe</td>
<td>SR = 0.5 m</td>
<td>Digital Globe</td>
</tr>
<tr>
<td>Digital elevation model</td>
<td>ASTER GDEM v2*</td>
<td>SR = 30 m</td>
<td>USGS</td>
</tr>
<tr>
<td>Rain records</td>
<td></td>
<td>Monthly/annual averages</td>
<td>MARNR - MinTIC</td>
</tr>
<tr>
<td>Soil physical properties</td>
<td>Texture and organic mater</td>
<td>Percentage</td>
<td>UNET Bio-environmental Lab</td>
</tr>
<tr>
<td>Soils distribution types map</td>
<td>Venezuelan environmental systems</td>
<td>1:250 000</td>
<td>MARNR</td>
</tr>
<tr>
<td>Roads</td>
<td>Roads</td>
<td>SR = 0.5 m</td>
<td>OSM</td>
</tr>
</tbody>
</table>


Source: Authors

General methodology: a systematic scientific literature review was carried out to identify a deterministic PWER (potential water erosion risk) model, remote sensing techniques, and GIS procedures, with the purpose of identifying eroded areas and areas at risk of erosion (EAER). Four main phases were implemented: 1) developing the PWER model, 2) obtaining the EAER, 3) evaluating the degree of accuracy regarding the EAER with a supervised maximum likelihood classification (MLC) and a comparison of PWER and EAER, and 4) analyzing the results (Figure 2).

Landsat 8 OLI/TIRS image processing: a satellite image from October 1, 2017, was downloaded (LC08_L1TP_00705_20171001_20171013_01_T1_sr) with surface reflectance values and L1T corrections (IGAC, 2013), as well as topographic correction performed by Camargo et al. (2021).

Digital Globe image: a natural color image from May 7, 2018, was employed. It had a spatial resolution of 0.5 m (SIGIS, 2019) and was supported by Google Earth, which allows its use in non-profit research (Thenkabail, 2016). Said image aided in the selection of ‘ground-truth’ samples in order to increase risk prediction reliability, based on a qualitative approach in erosion severity classes (Auerswald et al., 2018; Fischer et al., 2018; Batista et al., 2019).

PWER methodology

This methodology is based on two principles. The first one of a fundamental nature, expresses that it is the land that suffers the attack of the forces of climate (climatic aggressiveness) and that it, in turn, offers variable degrees of resistance, which poses a relation that determines the degrees of degradation in a given area. The second one indicates that potential risk assessment is most useful when relatively unstable or non-permanent factors (vegetation and land use) are not included in the calculation (FAO et al., 1980). Recent studies that use these principles include, among others, Guerra et al. (2020), Allafta and Oop (2021), and Al-Mamari et al. (2023). The PWER manifests via Equation (1):

\[ PWER = f(C, S, T) \]  

PWER is expressed in Mg ha\(^{-1}\)year\(^{-1}\); C is the rain erosivity factor; S is the soil factor, estimated by means of texture (st) and erodibility (se) subfactors; and T is the topography factor (FAO et al., 1980; Rosales and García, 2015).

The climate factor (C) was evaluated based on the modification of the Fournier index (1960) (Arnoldus, 1977), which is best correlated with the EI30 value (maximum rain intensity in mm·hr\(^{-1}\) with 30 min duration), which has been verified in several parts of the world and is regarded as valid for Venezuela (Pacheco, 2012). The original Fournier index was not employed, as it does not consider that there are areas whose rainfall regime may have more than a monthly precipitation peak (Muñoz et al., 2014). Then, interpolated monthly precipitation surfaces were generated using IDW (inverse distance weighting) instead of kriging, as not all assumptions for its use are fulfilled (Hämmerly et al., 2019).
The S-factor was estimated based on st and se features associated with a soil type distribution map (MARNR, 1983) given the scarce availability of detailed soil data, which generally leads to considering reconnaissance studies (Quiñonez and Dal Pozzo, 2008). The texture classes of the first ones were conformed (USDA, 2020) and reclassified into three general categories recognized by FAO-UNESCO (1976), in which in turn allowed assigning the valuations necessary for the model as per FAO et al. (1980). As for the second ones, the nomogram of the USLE K factor was employed (Foster et al., 1981), which is between 0 and 0.09 Mg ha h / ha MJ mm, using organic matter and texture percentages, later associated with erodibility values and classes and valuations as per FAO et al. (1980).

Finally, the T-factor was obtained from the ASTER GDEM, in which three slope classes were distinguished: a) flat to mildly undulating (0-8%); b) strongly undulating to hilly (8-30%); and c) strongly eroded to mountainous (>30%) (FAO-UNESCO, 1976), which were rated at (a) 0.35; (b) 3.5, and (c) 11.0 (FAO et al., 1980).

EAER methodology

Two mapping alternatives were developed: the first one was based on the spectral Euclidean distance between the reflectivity of each pixel in the satellite image and the bare soils category, and the second one on diverse remote sensing techniques, with which linear regressions were established. ROC (receiver operating characteristics) curves were applied to both products, which allowed defining classification thresholds and associated uncertainties aimed at detecting similar contiguous spectral zones of eroded and erosion risk areas (Beguería, 2006; Alatorre and Beguería, 2009). For thresholds, 100 independent samples (pixels) were selected, which evidenced <10% of vegetation cover, randomly distributed and defined with the help of the Digital Globe image, as the risk of erosion is considered to be high when this value is low (Wang et al., 2021).

An ROC curve is a graph that incorporates all sensitivity/ specificity pairs resulting from the continuous variation of cutoff points throughout the range of observed results. This offers a global view of diagnostic accuracy by providing significant data on the probability of correctly classifying an individual by means of a determined variable (Ampudia et al., 2017). Their equations are:

\[
\text{sensitivity} = \frac{a}{a+c} \quad (2)
\]

\[
\text{specificity} = \frac{d}{b+d} \quad (3)
\]

where \(a\) are true positives, \(b\) true negatives, \(c\) false positives, and \(d\) false negatives. Sensitivity expresses the proportion of correctly predicted positive pixels, and specificity represents the proportion of correctly predicted negative pixels.

Sensitivity and specificity values of 1 represent the likelihood of omission (type II, or false negative) and commission (type I, or false positive) errors (Alatorre and Beguería, 2009). In order to determine eroded areas, a sensitivity value of 0.9 was fixed, which corresponds to a 10% probability of omission errors. For areas at risk, a value of 0.8 was set (20% probability of omission error).

A supervised maximum likelihood classification (MLC) of the covers was carried out, as it is more precise than an unsupervised one because its classes are previously known (Liang and Wang, 2020). This, to later evaluate accuracy at determining bare soil cover with EAER by means of cross-tabulation. These results were updated with highways (OSM) and validated via global precision (Chuvieco, 2016) and kappa statistics (Cohen, 1960).

The remote sensing techniques implemented were vegetation indices based on slope and distance, soil indices (equations in Table 2), and others such as principal components analysis (PCA) (Pearson, 1901), spectral mixture analysis (SMA) (Boardman, 1992) and tasseled cap brightness (T) (Kauth and Thomas, 1976), with the coefficients derived by Baig (2014) for Landsat 8 Oli, as shown in Equation (4):

\[ B = b_2 \times 0.3029 + b_3 \times 0.2786 + b_4 \times 0.4733 + b_5 \times 0.5599 + b_6 \times 0.508 + b_7 \times 0.1872 \]  
(4)

Then, a linear regression analysis was carried out in order to determine the degree of dependence present between two variables (Shobha and Rangaswamy, 2018). To this effect, the bare soil spectral Euclidean distance was considered as an independent variable (x) and each technique as a dependent variable (y). Correlation tests were conducted in order to show the degree of the linear relation. Finally, all p-values generated were lower than 0.001, taking into account that values lower than 0.05 had to be considered in the analysis.

The resulting maps show three categories: (i) eroded areas, understood as those with no vegetation and denoting active erosion; (ii) at risk areas, those with little vegetation and prone to erosion; and (iii) no erosion, areas with good vegetation cover that ‘apparently’ protect against erosion.

Results

PWER

The behavior of the factors (Figure 3) allows defining a C-factor between 78.94 and 106.65. S was defined as fine (0,1) and medium (0,3) in terms of subfactor st, and as light (0,5) in terms of subfactor se (Table 3). T was between 0.35 and 11. Once the variables were defined, they were multiplied in map algebra in order to obtain PWER and later reclassify it to establish the soil erosion risk classes (Figure 4), thus obtaining 542.07 ha for ‘none to light’ (14.59%), 1 805.76 ha for ‘moderate’ (48.60%), and 1 767.37 ha for ‘high’ (36.80%).
Table 2. Indices

<table>
<thead>
<tr>
<th>Author</th>
<th>Vegetation index (slope-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jordan, 1969</td>
<td>$DVI = \frac{NIR - R}{NIR + R}$</td>
</tr>
<tr>
<td>Pearson and Miller, 1972</td>
<td>$RVI = \frac{R}{NIR}$</td>
</tr>
<tr>
<td>Rouse et al., 1973</td>
<td>$NDVI = \frac{NIR - R}{NIR + R}$</td>
</tr>
<tr>
<td>Baret and Guyot, 1991</td>
<td>$NRVI = \frac{RVI - 1}{RVI + 1}$</td>
</tr>
<tr>
<td>Deering et al., 1975</td>
<td>$TVI = \sqrt{NDVI} + 0.5$</td>
</tr>
<tr>
<td>Thiam, 1997</td>
<td>$TTVI = \sqrt{ABS(NDVI + 0.5)}$</td>
</tr>
<tr>
<td>Perry and Lautenschlager, 1984</td>
<td>$CTVI = \frac{NDVI + 0.5}{ABS(NDVI + 0.5)} \times \sqrt{ABS(NDVI + 0.5)}$</td>
</tr>
<tr>
<td>Huete, 1988</td>
<td>$SAVI = \frac{NIR - R}{NIR + R}(1 + L)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Vegetation index (distance-based)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richardson and Wiegand, 1977. Rewritten by Jackson et al., 1983</td>
<td>$PVI_0 = \left(\frac{NIR - a.R - b}{\sqrt{a^2 + 1}}\right)$</td>
</tr>
<tr>
<td>Perry and Lautenschlager, 1984</td>
<td>$PVI_1 = \frac{b.NIR - R + a}{\sqrt{b^2 + 1}}$</td>
</tr>
<tr>
<td>Walther and Shabaani, 1991</td>
<td>$PVI_2 = \frac{(NIR - a)(R + b)}{\sqrt{a^2 + 1}}$</td>
</tr>
<tr>
<td>Qi et al., 1994</td>
<td>$PVI_3 = a.NIR - b.R$</td>
</tr>
<tr>
<td>Baret and Guyot, 1991</td>
<td>$ATSAVI = \frac{a.NIR - a.R - b}{(a.NIR + R - a.b + 0.08(1 + a^2))}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Author</th>
<th>Soil index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nganfdam et al., 2016</td>
<td>$NDBSI = \frac{Swir1 - NIR}{Swir1 + NIR} + 0.001$</td>
</tr>
<tr>
<td>Celik, 2018</td>
<td>$BSI = \frac{(R + Swir1) - (NIR + B)}{(R + Swir1) + (NIR + B)}$</td>
</tr>
<tr>
<td>Li and Chen, 2018</td>
<td>$BI = f(Tasseled Cap(Brightness), NDBaI_2)$</td>
</tr>
<tr>
<td>FAO-UNESCO, 1976</td>
<td>$NDBaI2 = \frac{Swir1 - Tirs1}{Swir1 + Tirs1}$</td>
</tr>
</tbody>
</table>

**Source:** Authors
Table 3. Soil subfactors processing

<table>
<thead>
<tr>
<th>Soil Texture</th>
<th>Textural classes(^1) (FAO-UNESCO, 1976)</th>
<th>Valuation textural classes(^2) (FAO et al., 1980)</th>
<th>K-values(^3) (Foster et al., 1981)</th>
<th>Valuation erodibility classes(^4) (FAO et al., 1980)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clay loam</td>
<td>Medium</td>
<td>0,3</td>
<td>0,028</td>
<td>0,5 (light)</td>
</tr>
<tr>
<td>Clay</td>
<td>Fine</td>
<td>0,1</td>
<td>0,013</td>
<td>0,5 (light)</td>
</tr>
<tr>
<td>Silty clay</td>
<td>Fine</td>
<td>0,1</td>
<td>0,022</td>
<td>0,5 (light)</td>
</tr>
<tr>
<td>Sandy loam</td>
<td>Medium</td>
<td>0,3</td>
<td>0,023</td>
<td>0,5 (light)</td>
</tr>
<tr>
<td>Sandy clay loam</td>
<td>Medium</td>
<td>0,3</td>
<td>0,013</td>
<td>0,5 (light)</td>
</tr>
<tr>
<td>Sandy loam</td>
<td>Medium</td>
<td>0,3</td>
<td>0,019</td>
<td>0,5 (light)</td>
</tr>
</tbody>
</table>

1 Soil texture classes: coarse: <18% clay & >65% sand; medium: <35% clay & <65% sand; fine: >35% clay
2 st Classification: coarse: 0,2 medium: 0,3 fine: 0,1 stony phase: 0,5
3 K-value: light: <0,03; moderate: 0,03-0,06; high: 0,06 >.
4 se valuation: light: 0,5; moderate: 1,0; high: 2,0.

Source: Authors

Maximum likelihood classification (MLC)

Six categories were established, and their areas were extracted: clouds (13,95 ha), highways (163,98 ha), high vegetation (1,331,10 ha), low vegetation (1,874,07 ha), bare soils (including rocky outcrops) (2,777,46 ha), and infrastructure (69,36 ha) (Figure 5). The map had a global accuracy of 81,83% and a kappa index of 0,79.

Figure 3. PWER factors and subfactors
Source: Authors

Figure 4. Soil erosion risk map
Source: Authors

Figure 5. Maximum likelihood classification
Source: Authors

EAER1

The ROC curve showed a high sensitivity, i.e., good capabilities to correctly classify positive pixels (bare soils). The 0,9 (U1) and 0,8 (U2) thresholds indicated spectral distances of 0,114 and 0,087. Both represent a balance between sensitivity and specificity. That is to say, an approach to the most ideal conditions (100%), which allowed generating the map (Figure 6) with the following results: eroded areas: 379,44 ha; erosion risk: 260,91 ha; no erosion: 2 911,95 ha (Figure 6).
Cross-referencing with MLC indicated that the eroded areas coincided with 229,77 ha of bare soils. In the same way, this category intercepted ‘high vegetation’ in 0,99 ha, ‘low vegetation’ in 126,99 ha, and ‘infrastructure’ in 21,69 ha. For its part, the erosion risk class mainly intercepted ‘low vegetation’ (236,52 ha), ‘bare soils’ (20,61 ha), ‘infrastructure’ (3,33 ha), and ‘high vegetation’ (0,45 ha) (Table 4).

Cross-referencing with MLC indicated that the surface occupied by eroded areas coincided with 195,57 ha of bare soils and intercepted ‘high vegetation’ in 0,27 ha, ‘low vegetation’ in 31,95 ha, and ‘infrastructure’ in 33,39 ha. For its part, ‘erosion risk’ intercepted ‘low vegetation’ in 15,84 ha, ‘bare soils’ in 9,9 ha, and ‘infrastructure’ in 0,09 ha (Table 6).

Table 4. Cross tabulation: EAER1-MLC (ha)

<table>
<thead>
<tr>
<th>MLC</th>
<th>Eroded areas</th>
<th>Erosion risk</th>
<th>No erosion</th>
</tr>
</thead>
<tbody>
<tr>
<td>High vegetation</td>
<td>0,99</td>
<td>0,45</td>
<td>1 329,66</td>
</tr>
<tr>
<td>Low vegetation</td>
<td>126,99</td>
<td>236,52</td>
<td>1 510,56</td>
</tr>
<tr>
<td>Bare soils</td>
<td>229,77</td>
<td>20,61</td>
<td>27,36</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>21,69</td>
<td>3,33</td>
<td>44,37</td>
</tr>
</tbody>
</table>

Source: Authors

Figure 6. EAER1
Source: Authors

Table 5. Linear regressions

<table>
<thead>
<tr>
<th>Technique</th>
<th>R</th>
<th>R²</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA sum</td>
<td>0,648</td>
<td>0,42</td>
<td>y = 0,44 + 1,64X</td>
</tr>
<tr>
<td>SAVI</td>
<td>0,592</td>
<td>0,351</td>
<td>y = 0,6 + 2,75X</td>
</tr>
<tr>
<td>NDVI</td>
<td>0,591</td>
<td>0,35</td>
<td>y = 0,40 + 1,83X</td>
</tr>
<tr>
<td>TTVI</td>
<td>0,579</td>
<td>0,335</td>
<td>y = 0,95 + 0,87X</td>
</tr>
<tr>
<td>CTVI</td>
<td>0,566</td>
<td>0,321</td>
<td>y = 0,95 + 0,86X</td>
</tr>
<tr>
<td>PCA₁</td>
<td>0,542</td>
<td>0,294</td>
<td>y = 0,33 + 0,83X</td>
</tr>
<tr>
<td>TVI</td>
<td>0,511</td>
<td>0,261</td>
<td>y = 1,15 + 1,13X</td>
</tr>
<tr>
<td>PCA₂</td>
<td>0,494</td>
<td>0,244</td>
<td>y = 0,09 + 0,71X</td>
</tr>
<tr>
<td>PVI₁</td>
<td>0,46</td>
<td>0,211</td>
<td>y = 0,19 + 0,33X</td>
</tr>
<tr>
<td>PCA₃</td>
<td>0,294</td>
<td>0,087</td>
<td>y = 0,02 + 0,11X</td>
</tr>
<tr>
<td>PVI₃</td>
<td>0,277</td>
<td>0,077</td>
<td>y = 0,03 + 0,09X</td>
</tr>
<tr>
<td>DVI</td>
<td>0,221</td>
<td>0,049</td>
<td>y = 0,17 + 0,27X</td>
</tr>
<tr>
<td>PVI₀</td>
<td>-0,017</td>
<td>0</td>
<td>y = -0,04 - 0,02X</td>
</tr>
<tr>
<td>PVI₂</td>
<td>-0,082</td>
<td>0,007</td>
<td>y = 0,02 - 0,33X</td>
</tr>
<tr>
<td>RVI</td>
<td>-0,547</td>
<td>0,299</td>
<td>y = 0,41 - 1,43X</td>
</tr>
<tr>
<td>Tasseled cap</td>
<td>-0,552</td>
<td>0,306</td>
<td>y = -0,45 - 0,83X</td>
</tr>
<tr>
<td>NRVI</td>
<td>-0,595</td>
<td>0,354</td>
<td>y = -0,40 - 1,84X</td>
</tr>
<tr>
<td>NDBSI</td>
<td>-0,598</td>
<td>0,358</td>
<td>y = 0,1 - 1,64X</td>
</tr>
<tr>
<td>ALME</td>
<td>-0,642</td>
<td>0,413</td>
<td>y = -0,83 - 3,55X</td>
</tr>
<tr>
<td>ATSAVI</td>
<td>-0,675</td>
<td>0,456</td>
<td>y = -1,04 - 6,03X</td>
</tr>
<tr>
<td>BI</td>
<td>-0,68</td>
<td>0,462</td>
<td>y = 8,05E4 - 8,48E4X</td>
</tr>
<tr>
<td>BSI</td>
<td>-0,703</td>
<td>0,494</td>
<td>y = 0,12 - 5,72X</td>
</tr>
<tr>
<td>NDBIa2</td>
<td>-0,742</td>
<td>0,551</td>
<td>y = 4,83E4 - 4,67E4X</td>
</tr>
</tbody>
</table>

y = dependent variable; x = independent variable
Source: Authors

Then, an ROC curve was elaborated based on 100 samples of PCA sum values. The 0,9 (U1) and 0,8 (U2) thresholds indicated values of -0,117 and -0,107 for the PCA sum, respectively, which represent a good balance between sensitivity and specificity, thus allowing to generate the map (Figure 7) with the following results: eroded areas: (261,18 ha); erosion risk: 25,83 ha; no erosion: 3 265,29 ha.
With the purpose of understanding the PWER as an indicator of erosion sensitivity, 14 bare soil pixels (samples) were first analyzed (Figure 8) based on the relationship between the thematic categories defined for the EAER, the classification by FAO et al. (1980), and the PWER values (Table 7). It must be highlighted that, given that the samples of bare and erosion-exposed soils were identified based on a higher-resolution image, they should only be classified as eroded or erosion risk areas in the EAER (‘no erosion’ would be an imprecision generated by the change in spatial resolution or threshold definition). Secondly, the possible soil loss quantities were totaled for eroded and erosion risk areas, with the purpose of evidencing the approach with the most losses.

The results allowed defining a higher success rate in EAER1. However, an imprecision was found in sample 5, which was considered to belong to ‘no erosion’ (Figure 8e), as it is completely uncovered. This was the only incongruence generated in EAER1, since the correct category should have been ‘eroded areas’, i.e., spectral distance values lower than 0.087. For its part, EAER2 showed three contradictions when it categorized samples 1, 2, and 3 as ‘no erosion’, even when they were uncovered (Figures 8a, 8b, and 8c). These pixels should have been lower than the PCA sum value of -0.107 (incongruences refer to values lower than those established by the thresholds in the ROC curves).

Considering the high success rate of the samples in both EAER, the PWER condition of the 14 points could be observed. In the case of sample 5 in EAER1, potential degradation is considered to be high, with 132.61 Mg ha⁻¹year⁻¹. On the other hand, sample 3 in EAER2 showed a loss of 44.34 Mg ha⁻¹year⁻¹, which represents a moderate degradation (the remaining samples can be interpreted in the same way). Finally, the soil losses from eroded and erosion risk areas in the first map were much greater than those of the second (by 162.88%) (Table 8).
Discussion

Even though the PWER methodology dates back to 1980, it is still valid because its application is ideal when there is no adequate information on precipitation, evapotranspiration, lithology, and soils, which is required by methodologies such as USLE or the one proposed by Pacheco (2012) or more recently by Nasir et al. (2023). On the contrary, the need for minimal information such as monthly precipitations, free MDE, and reconnaissance-level soil mapping provides this methodology with execution plausibility, simplicity, and scientific robustness. It is also important to note that many of the models currently used require the inclusion of impractical variables (Demirel and Tüzün, 2011), which highlights the need for new approaches in the modeling of erosion processes (Nasir et al., 2023).

A case that illustrates the above is the comparison between the potential erosion results obtained in this study and those of Condori-Tintaya et al. (2022) and Nasir et al. (2023). While the former used soil structure information and calculated a more complex length and slope factor based on equations, the latter employed a multi-criteria decision making (MCDM) approach involving the development of 15 variables and an analytical hierarchy process (AHP).

Access to GIS and satellite images has had great impact on erosion modeling. Models can now be applied with relative ease at a large scale and in a distributed fashion, and they can present results in pixels that allow identifying where they occur and their magnitude, as well as at different temporal and spatial scales (Batista et al., 2019). However, it is also recognized that the predictive capability of large-scale erosion models is not the best. Therefore, Alewell et al. (2019) have argued that the same should not strive for them to make accurate predictions of soil losses, but instead to explore scenarios and focus on understanding relative differences in erosion rates, which would help to identify areas prone to these degradation processes, which is the objective of this study.

In the same way, very high-resolution images may be used to test erosion models, which has not been widely performed by researchers, with Fischer et al. (2018) perhaps being the first to focus completely on its interpretation. They found promising results, such as a high correlation ($R^2 = 0.91$) of visually defined erosion classes with moderate soil losses, thus allowing them to define a semi-quantitative assessment approach, which is much simpler than proposing hypothesis tests (Batista et al., 2019).

Other reasons to support the use of this methodology is that it excludes speculations on the validity of the models’ predictions, and it allows identifying scenarios leading to great or small soil losses (Auerswald et al., 2018). This study did not employ a semi-quantitative approach because the real soil loss values of the samples selected were not calculated independently, so as to be able to determine the $R^2$, which denotes a predominantly qualitative assessment to address comparative analyses.

The key to proceeding with the identification of EAE lies in the accurate extraction of vegetation cover and land use (Wang et al., 2013). In this case, it involved aiming for the highest possible accuracy regarding bare soils, with which the spectral Euclidean distance was obtained to later define thresholds in ROC curves, since the presence of vegetation cover reflects resistance capabilities against erosion or its risk (Wang et al., 2021).

Starting with ‘eroded areas’ and its intersection with bare soils (379,44 ha = 100%), cross-tabulation between ML and the EAER allowed corroborating that there was a higher intersection in EAER1 (60,55%) in comparison with EAER2, which coincided by 51,54%. This result evinces the former’s greater reliability, which is very likely due to the not-so-high correlation and determination coefficients of the PCA sum with respect to the spectral Euclidean distance.

As for the cross-reference between ‘eroded areas’ and ‘low and high vegetation’, EAER1 showed a 3,43% coincidence, unlike EAER2, which reported 0,87%. The intersection between ‘eroded areas’ and ‘infrastructure’ was lower in EAER1, with 0,58%, in contrast with the EAER2’s 0,90%. These results are also explained by the differentiation inability shown by the latter, which is due to the not-so-high correlation and determination of the PCA sum.

These results position EAER1 as a baseline alternative in contrast with the PCA sum. This conclusion is reasonable, as this input was used as an independent variable in linear regressions.

As for the selection of thresholds in ROC curves, given the varying sensitivity/specificity, other thresholds with more or less discrimination power could be selected. Selecting two with a high sensitivity would reciprocally allow achieving a high probability of correctly classifying a pixel whose real situation is defined as positive. The advantage of the ROC curve lies in the fact that it uses all possible cutoff points in the database, through which better thresholds are determined, thus indicating that this test has a very useful and correct classification power (Bermui et al., 2022).

Therefore, the proposed method for identifying EAER differs from associating higher loss values (Mg ha$^{-1}$ year$^{-1}$) yielded by USLE and/or similar models as an indirect measure of diverse erosion risk degrees (e.g., Chaudhary and Kumar, 2018; Mohammed et al., 2020), which is supported by the assumption that low values are less vulnerable and high

Table 8. Comparison EAER1 and EAER2

<table>
<thead>
<tr>
<th>Class</th>
<th>Mg ha$^{-1}$ year$^{-1}$</th>
<th>Class</th>
<th>Mg ha$^{-1}$ year$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ea 1</td>
<td>219 045,65</td>
<td>Ea 2</td>
<td>136 147,80</td>
</tr>
<tr>
<td>Er 1</td>
<td>181 670,42</td>
<td>Er 2</td>
<td>16 284,01</td>
</tr>
<tr>
<td>Total</td>
<td>400 716,07</td>
<td>Total</td>
<td>152 431,81</td>
</tr>
</tbody>
</table>

Source: Authors
values refer to a greater erosion sensitivity (Meshesha et al., 2012). This also contrasts with the idea of obtaining them by ranking and later zoning, which results from intersecting diverse slope gradients from an MDE or land uses with soil losses (Meshesha et al., 2012; Khosrokhani and Pradhan, 2013; Wang et al., 2013).

The equivalence between remote sensing techniques obtained by linear regression was manifested by Ngandum et al. (2016), who proposed a vegetation index as the independent variable, which placed an exaggerated reliability in said product as a predictor. This, unlike the design implemented in this study, which was defined based on the spectral distance of bare soils.

The comparison between the qualitative and the quantitative results allows establishing a bidirectional complementary analysis relationship. In the first place, EAER helped to identify areas with erosion or highlight those at risk, in which soil losses can be interpreted. Secondly, it allows establishing the inverse relationship, i.e., the observation of possible high and moderate losses identified by the PWER provides information on what happens in said areas from a qualitative perspective.

For a wider understanding and for determining the possible existence of erosion or of areas that might be subjected to it, it is suggested that eroded and erosion risk areas be identified with both mappings. This consideration is supported by the identification of possible substantial differences at potential points of interest.

Finally, based on the study by Batista et al. (2019) regarding the fact that the current, different erosion models do not systematically surpass each other, this research agrees that calibration is the only mechanism to improve their performance (i.e., a better conceptual understanding of their operation). Therefore, this study rejects the notion that these can be validated (rather than evaluated), stressing the need to define adjustment tests (or evaluating degrees of reliability) based on multiple data sources, which allow for a wide study of the usefulness and consistency of the developed methodologies. This, considering that the more thorough the tests, the more likely it is that deficient performances are found (critical awareness of the methods).

**Conclusions**

Using remote sensing images allows conducting research on erosion both qualitatively and quantitatively. However, these must be used with care, as their unthinking use may lead to over- or underestimation. The use of very high-resolution images must also be considered as a mechanism to evaluate model performance. Resorting to higher resolution imagery or covering larger extensions implies a higher computational cost because it involves larger amounts of data. This could be overcome by using cloud computing platforms such as Google Earth Engine (GEE), which can compute and process large volumes of geospatial data in very short time intervals and have been recently applied to the preparation of the variables required by methods such as USLE and RUSLE (Papaioanndis et al., 2019; Kumar et al., 2022), so their application with the models developed in this study is also plausible.

Although technological advances are evident, it should not be forgotten that erosion models are not necessarily true or free of apparent flaws; recognizing them improves the attitude towards evaluating them and changes the way their performances are characterized and communicated, ultimately leading to a better understanding of soil erosion (Batista et al., 2019).

This study constitutes a contribution to the lack of precipitation and soil data, which is necessary in parametric methodologies. Therefore, it adopted the premise that this lack should be only resolved with free and available digital inputs that help identify potential erosion hot spot areas and simulate erosion responses to land use and climate change, which makes it a solution that could be associated with variables derived from MDE (e.g., humidity indices) or other kinds of categorical or continuous mapping. Nevertheless, regardless of the methods to be employed, an on-field survey of geo-referenced measurements (when possible) regarding erosion characteristics must not be discarded, with the purpose of broadening the model’s evaluation.

Even though it is true that the methods developed for identifying EAER are metric-static and implemented with satellite images from diverse dates, these could help to obtain time series allowing to better understand the dynamics of erosion processes and therefore acquire greater knowledge for soil conservation and ecosystem management. They could also be replicated in other spaces in a semi-automatized fashion, and they could serve as first inputs to define areas where observations should be focused or to map types and degrees of erosion.

**Acknowledgements**

To the doctoral program in Forest and Environmental Sciences (CEFAP – FCFA – ULA, Mérida, Venezuela), to which this document is linked. To the UNET Bio-Environmental Laboratory for their support. We would like to thank the Journal’s Reviewers and Editorial Committee for their contributions and observations.

**Author contributions**

C. E. C-R. conceived the research, did the background research, collected the data, developed the workflow, and performed the assessment. C. E. P-A., R. L-F., S. A. M-A., and T. G-O., supervised the research and provided critical feedback. All authors contributed to writing the manuscript and approved its definitive version for publication.
References


Food and Agriculture Organization of the United Nations (FAO), United Nations Educational, Scientific and Cultural Organization (UNESCO) (1976). *Mapa mundial de suelos 1:5.000.000.000*. FAO.


