A Remote Sensing Based Method to Detect Soil Erosion in Forests

Hanqiu Xu 1,*, Xiujuan Hu 1, Huade Guan 2, Bobo Zhang 1, Meiya Wang 1, Shanmu Chen 3 and Minghua Chen 3

1 College of Environment and Resources, Institute of Remote Sensing Information Engineering, Key Laboratory of Spatial Data Mining & Information Sharing of Ministry of Education, Fujian Provincial Key Laboratory of Remote Sensing of Soil Erosion, Fuzhou University, Fuzhou 350116, China; huxiujuan@fzu.edu.cn (X.H.); zbbo@outlook.com (B.Z.); m161110016@fzu.edu.cn (M.W.)
2 National Centre for Groundwater Research and Training, College of Science and Engineering, Flinders University, Adelaide, SA 5001, Australia; huade.guan@flinders.edu.au
3 Fujian Monitoring Station of Water and Soil Reservation, Fuzhou 350001, China; SMchen@outlook.com (S.C.); MHchen@outlook.com (M.C.)

* Correspondence: hxu@fzu.edu.cn; Tel: +86-591-2286-6071

Received: 30 January 2019; Accepted: 26 February 2019; Published: 2 March 2019

Abstract: Rainwater-induced soil erosion occurring in the forest is a special phenomenon of soil erosion in many red soil areas. Detection of such soil erosion is essential for developing land management to reduce soil loss in areas including southern China and other red soil regions of the world. Remotely sensed canopy cover is often used to determine the potential of soil erosion over a large spatial scale, which, however, becomes less useful in forest areas. This study proposes a new remote sensing method to detect soil erosion under forest canopy and presents a case study in a forest area in southern China. Five factors that are closely related to soil erosion in forest were used as discriminators to develop the model. These factors include fractional vegetation coverage, nitrogen reflectance index, yellow leaf index, bare soil index and slope. They quantitatively represent vegetation density, vegetation health status, soil exposure intensity and terrain steepness that are considered relevant to forest soil erosion. These five factors can all be derived from remote sensing imagery based on related thematic indices or algorithms. The five factors were integrated to create the soil erosion under forest model (SEUFM) through Principal Components Analysis (PCA) or a multiplication method. The case study in the forest area in Changting County of southern China with a Landsat 8 image shows that the first principal component-based SEUFM achieves an overall accuracy close to 90%, while the multiplication-based model reaches 81%. The detected locations of soil erosion in forest provide the target areas to be managed from further soil loss. The proposed method provides a tool to understand more about soil erosion in forested areas where soil erosion is usually not considered an issue. Therefore, the method is useful for soil conservation in forest.

Keywords: red-soil erosion; SEUFM; detection model; yellow leaf index; fractional vegetation coverage; vegetation health; principal components analysis

1. Introduction

Soil erosion is a serious problem common in many areas of the world. It causes severe land degradation and affects soil fertility, agriculture productivity, water quality and regional ecosystems and environments [1]. Among various types of soil erosion, rainwater-induced soil erosion under forest canopy is a special phenomenon of soil erosion [2,3]. Such soil erosion occurs in areas with moderate to high forest coverage, where soil erosion is often not considered an issue.
However, severe soil loss in forest areas is common in many red soil regions of southern China [2–5]. This issue has attracted considerable attention and has been an active research topic for the last decade. Liao et al. [6] investigated soil erosion in forests based on 9-year monitoring data of seven sub-catchments in Zhejiang Province, southern China and summarized the processes of soil erosion under forest canopy. Through both artificial rainfall simulation experiments and laboratory experiments, Xu [7] examined the soil-loss process of Masson pine forest and its dynamics in Fujian Province. Zhu et al. [4] investigated soil erosion occurring in the forest of the Taihu watershed in Jiangsu Province and attributed the cause of the erosion to low soil fertility. Lei [3] studied soil loss in forests in a granite region of Xingguo County, Jiangxi Province and provided suggestions to control soil erosion in the forest. He et al. [2] discussed the status and causes of soil erosion in *Pinus massoniana* woodlands in hilly red soil regions of southern China and found that the lack of vegetation mulch under trees and human disturbance were the main factors causing the erosion. Zhang et al. [5] assessed soil erosion from remote sensing data in rehabilitated high-density forests of Hetian, a typical red soil region in southern China and found that the eroded areas were primarily distributed in locations with elevations between 300 to 500 m.

To date, a wide range of methods has been developed to monitor and estimate soil loss at multiple scales. Among them, Revised/Universal Soil Loss Equation (R/USLE) [8,9] is most commonly used. Other models include Aerial Nonpoint Source Watershed Environment Response Simulation (ANSWERS) [10], Water Erosion Prediction Project (WEPP) [11], Soil and Water Assessment Tool (SWAT) model [12], Unit Stream Power-based Erosion Deposition (USPED) [13], European Soil Erosion Model (EUROSEM) [14], Limburg Soil Erosion Model (LISEM) [15,16] and Hillslope Erosion Model (HEM) [17]. These models are different in the number and type of input factors, complexity, processes considered and amount of data required [18]. Estimation of soil erosion using the models is still a challenge due largely to data availability [1].

Remote sensing techniques have been increasingly used to detect the areas at the risk of extensive soil erosion [5,8,9]. The remote sensing based methods provide a cost-effective way to investigate soil erosion whereas the terrains are not accessible or direct field methods are expensive [19,20]. Various remote sensing based methods have been developed and applied to detect features related to soil erosion. Satellite-based spectral indices, for example, Normalized Difference Vegetation Index (NDVI), Normalized Difference Soil Index (NDSI), Tasselled Cap Transformation (TCT), along with Linear Spectral Unmixing Analysis (LSMA) have been frequently employed to estimate soil erosion process [21,22], investigate soil exposure intensity [23], measure soil reflectance [24,25], evaluate soil erosion status [5,26] and assess soil properties and bare soil fractions [27]. In addition, object-oriented analysis has been used to aid image classification and gully erosion identification [28,29]. Recently, machine-learning approaches have been applied to reveal statistical relationships between gully erosion and its controlling factors [30] and to retrieve key factors for soil erosion risk detection [31].

As for the remote sensing data sources used in soil erosion detection, optical multispectral imagery is most frequently used. Nevertheless, Synthetic Aperture Radar (SAR) data can provide an auxiliary source and the fusion of microwave and optical data can help soil erosion monitoring [32,33]. Recently, data from Unmanned Aerial Vehicle (UAV) imagery have also been used for small-scale monitoring of soil erosion, which provided very-high spatial resolution images for the monitoring [34–36].

There are a number of factors that influence soil erosion and are therefore involved in model construction for soil erosion detection. Of them, rainfall is one of the frequently-considered factors as rainfall-induced soil erosion occurs commonly in various environments. Long-term rainfall data are useful to identify the period of strong soil erosion, when used as the rainfall erosivity factor (R-factor) of RUSLE for investigating rainfall erosivity variability [37]. Rainfall influences land surfaces by its duration, intensity and cumulative amount per event. A recent study in a clayey coastal hilly area in Italy showed that soil erosion was correlated more with cumulative rainfall than with peak rainfall intensity [38]. Rainfall can have interactions with land covers and therefore, the coupling of rainfall and vegetation helped a quick identification of the month that had the highest soil erosion risk [39].
Soil erodibility is also an important factor used in modelling soil erosion. It works as a key factor in the models such as RUSLE (K-factor) and Chinese Soil Loss Equation (CSLE) [40] and is useful for estimating soil erosion rate [41]. Work in Ecuadorian Andes indicated that soil erodibility was the most important factor in the dry season for agricultural areas [42]. A study carried out in Shaanxi province, one of the most serious soil erosion regions in China, showed that soil erodibility played the second most important role among the four factors (rainfall erosivity, slope steepness and slope length being the other three) [40].

Remote sensing derivable factors also play critical roles in soil erosion detection. Fractional vegetation coverage (FVC) is among the most frequently-used factors. It can even be used alone to effectively detect the surfaces prone to soil erosion [43]. Many studies employed FVC to estimate the cover-management factor (C-factor) of the RUSLE model [44–47]. FVC has been shown to explain 57% of the sediment yield variance in a soil erosion modelling [43] and is a useful indicator to assess sediment connectivity between adjacent catchments [48]. Slope is another remote sensing data derivable factor that has been commonly-used in soil erosion monitoring as many of deriving forces to soil erosion have been associated with slope [9,11,15,49,50]. Slope is an important basic input of many soil erosion models currently in operation in the world, such as those of [9,11,14,15,17]. The factor can be represented by slope steepness and slope length [9,11,17], while the latter is not as popular as the former. A recent study in the karst regions of southwestern China found that slope degree influenced soil loss more than slope length [51].

Although various models have been successfully applied for soil erosion detection around the world, there is no suitable model for all soil erosion applications. In particular, the application of the existing models to estimate soil erosion under forest canopy becomes even more challenging because the relevant surface features are hidden under the canopy [5]. To date, few studies have focused on this special soil-erosion phenomenon in forest. Therefore, this study aims to develop a method to quickly detect soil erosion spots occurring under forest canopy solely using remote sensing techniques. A model, referred to as the Soil Erosion Under Forest Model (SEUFM), is proposed and tested. The model incorporates remotely sensed erosion-related cause and/or effect variables, including fractional vegetation cover, tree health status, soil exposure degree and slope. It provides a relatively easy-to-use method to identify the areas in forest that are at risk of soil erosion. The model was applied in a subtropical forest in Changting County of Fujian Province in southern China.

2. Study Area

Changting County is located in western Fujian Province of south-eastern China between 116°00′–116°40′E and 25°18′–26°02′N. The county has long been a severe red soil loss area in southern China. After the treatments to reduce soil erosion over 30 years, the soil loss has been dramatically reduced. The mountainous and hilly areas have been gradually covered with forests [23]. Nevertheless, soil loss is still occurring in many forested areas due to the exposure of topsoil under forest canopy (Figure 1).
Figure 1. Photo showing a site in Changting that is vulnerable to soil erosion due to exposed top-soils under the forest canopy.

Changting covers an area of 3099 km$^2$. It has a subtropical monsoon climate with an annual average temperature of 18.3 °C and annual rainfall ranging from 1500 to 1700 mm (http://www.changting.gov.cn/xxgk/tjxx). Topographically, the county sinks at the central part around Hetian town, which is thus called the Hetian Basin in the region. The elevations of the county range from 210 to 1459 m above mean sea level. Changting has a forested area of 2487 km$^2$ (http://www.changting.gov.cn/xxgk/tjxx). Based on the 2014 Landsat 8 image (Figure 2), most of the county’s area is covered by green forests, except for the area surrounding the Hetian basin, where greyish colour in the image suggests the exposure of surface soil.

Figure 2. Landsat 8 image of Changting County (October 8, 2014) (Red, Green and Blue channel combination: Bands 4, 3 and 2). The rectangle marks the location of the Hetian basin and the field survey sites are shown as yellow dots.
3. Materials and Methods

3.1. Remote Sensing Data and Image Pre-processing

The data source used in this study is mainly from Landsat 8 satellite imagery and ASTER DEM. The satellite image is a Landsat 8 Operational Land Imager (OLI) image of Changting (path 121, Row 42), acquired on October 8, 2014. The Level 1T image was downloaded from the EarthExplorer of U.S. Geological Survey (https://earthexplorer.usgs.gov). The image was radiometrically corrected using the algorithms and coefficients from the official Landsat 8 Data Users Handbook with the addition of the atmospheric correction model of Chavez [52,53]. The main equations are given below:

\[ L_\lambda = M_L Q_{\text{cal}} + A_L \quad (1) \]

where \( L_\lambda \) is the spectral radiance of band \( \lambda \) (W/(m\(^2\) sr µm)), \( M_L \) is the radiance multiplicative scaling factor for band \( \lambda \), \( A_L \) is the radiance additive scaling factor for band \( \lambda \) and \( Q_{\text{cal}} \) is the pixel value in DN of band \( \lambda \).

\[ \rho_\lambda = (M_\rho Q_{\text{cal}} + A_\rho)/\cos\theta \quad (2) \]

where \( \rho_\lambda \) is the top of atmosphere (TOA) reflectance of band \( \lambda \), \( M_\rho \) is the reflectance multiplicative scaling factor for band \( \lambda \), \( A_\rho \) is the reflectance additive scaling factor for band \( \lambda \) and \( \theta \) is the solar elevation angle. The parameters for the calculation with the equations can be found in the metadata of the image.

3.2. Selection of Predictor Factors

This study aims to provide a method solely based on relatively easier obtained remote sensing data in order to make the model more applicable. Accordingly, the factors were selected mainly from those that can be directly derived from remote sensing imagery. In addition, soil erosion under forest canopy is a special phenomenon of soil erosion that is related closely to plant status. Therefore, the selection of predictor factors focuses more on those that are closely related to vegetation.

Studies show that soil erodibility in forest can be closely related to the agents including fractional vegetation coverage, vegetation health status, soil exposure degree and slope steepness of terrain [23,54–56]. According to the Standards for Classification and Gradation of Soil Erosion [57], the soil erosion rate in planar surfaces is measured with two factors: FVC and slope. Accordingly, vegetation coverage and slope are the two important factors for the identification of soil loss. Nevertheless, these two factors are only applicable to common types of soil erosion. Information included in these two factors may not be sufficient to detect soil erosion under forest canopy. Therefore, in addition to the two factors, the health status of vegetation and soil exposure degree are also introduced to the model. The reasons to include these additional two factors are:

1. Healthy trees reduce raindrop-induced soil erosion by forming a dense and multi-storey canopy with plenty of branches and leaves and, therefore, can greatly dissipate the kinetic energy of raindrops before reaching the ground surface. In addition, healthy trees often have more developed root zones, which reduce soil erosion potential. In contrast, poor healthy trees with sparse branches and leaves cannot effectively obstruct the strike of raindrops and hence are unable to reduce the rainfall erosivity and fails to protect the soil from the direct impact of raindrops and throughfall [58]. Yan et al. [59] and Zhu et al. [4] reported that the areas covered with unhealthy trees tended to be more likely to develop soil erosion. On the other hand, a surface with soil erosion does not retain good water and nutrient condition, which may lead to a poor tree health condition. Thus, tree health status can be a good indicator for soil erosion.

2. Our field surveys found that soil erosion in forest usually took place in the areas with highly exposed topsoil due to the absence of sub-storey grass mulch covers (Figure 1). The work of Zhou et al. [60] has shown that the on- and near-ground vegetation layer in forest is the key to preventing soil erosion. He et al. [2] pointed out that soil erosion in forest mostly occurred in the areas
that were devoid of understory grass covers and stated that a high degree of soil exposure under tree canopy was the most important cause for soil erosion in forest. Zhang et al. [5] reported that many soil erosion sites in Changting County had almost no vegetation under the forest canopy and exhibited various degrees of soil exposure. Sun [61] found that a combined vegetation cover of trees and grass reduced surface runoff by 25% and sediment by 90% in comparison with a forest without ground vegetation cover.

Accordingly, the major factors that influence soil erosion in forest are FVC, vegetation health status, soil exposure degree and slope. The vegetation health status can be determined by two factors, yellow leaf index and nitrogen index. The selected factors are essential to identify the spots in forest that are particularly prone to erosion and thus were considered in the SEUFM model development. All five factors, detailed in the next section, can be derived directly from remote sensing data. This provides a quick and easy way to detect soil erosion potential in forest.

3.3. Retrieval of the Selected Factors

3.3.1. FVC

Regional FVC is usually computed using remote sensing based models [62]. Two commonly used models are those proposed by Carlson and Ripley [63] and Gutman and Ignatov [64], expressed as:

Carlson and Ripley:

\[
FVC = \left( \frac{NDVI - NDVIt_0}{NDVI_{v0} - NDVIt_0} \right)^2
\]  

(3)

Gutman and Ignatov:

\[
FVC = \frac{NDVI - NDVIt_0}{NDVI_{v0} - NDVIt_0}
\]  

(4)

where \( NDVI \) is the NDVI value of a pixel, \( NDVIt_0 \) is the NDVI value for bare soil selected from entirely bare soil and \( NDVI_{v0} \) corresponds to the NDVI value of a surface with an FVC of 100%, selected from extremely dense forest.

The difference between Equations (3) and (4) lies in whether the relationship between FVC and NDVI is quadratic or linear. The quadratic FVC (hereafter referred to as “FVC\text{\_quadratic}”) has a lower value than the linear FVC (hereafter referred to as “FVC\text{\_linear}”). As a result, FVC\text{\_quadratic} could avoid the overestimation of FVC in medium- to low-vegetation cover areas but may underestimate FVC in high-vegetation cover areas. The FVC\text{\_linear} performs inversely. Therefore, the selection of an appropriate FVC model is highly dependent on the status of vegetation coverage in the area to be studied [56,65].

In order to select an appropriate model, both FVC\text{\_quadratic} and FVC\text{\_linear} were tested with the Landsat 8 OLI image of Changting. The retrieved FVC results of the two models were then verified with the Google Earth (GE) high spatial resolution image of 20 December 2014, using the subpixel accuracy assessment method suggested by Wu and Murray [66]. For every sampled FVC pixel, the corresponding high-resolution GE image was digitized and the FVC was calculated. In this way, each sampled FVC pixel was related to the actual percentage of vegetation coverage of the GE high-resolution image within that pixel.

3.3.2. Nitrogen Reflectance Index (NRI)

Nitrogen is one of the most important nutritional elements for plant growth. It plays an important role in the process of the formation of cell, tissue and organs of plants. Soil sampling for nitrogen availability is well documented but requires considerable effort for sample collection and processing. The nitrogen index developed using remote sensing techniques provides a rapid assessment of plant
nitrogen status on a frequent basis [67] and thus was adopted in this study to characterize the vegetation health status. The Nitrogen Reflectance Index (NRI) of Bausch and Duke [67] is expressed as:

$$NRI = \frac{NIR}{Green}$$ \hspace{1cm} (5)

where \textit{Green} and \textit{NIR} are the reflectance of green and near-infrared bands.

3.3.3. Yellow Leaf Index (YLI)

Stress from soil moisture and nutrient can lead to yellow leaves. In the plant growing season, the occurrence of a great number of yellow leaves is an important indicator of unhealthy vegetation [68]. A yellow band of remote sensing image (e.g. WorldView 2 imagery) would be of value in detecting the yellow appearance of vegetation. However, Landsat 8 OLI imagery does not have a yellow band and, therefore, cannot directly observe the status of plant yellow leaves. Nevertheless, a yellow band can be simulated by the addition of green and red bands according to the principle of three primary colours [69] and is computed as:

$$Yellow = \frac{(Green + Red)}{2}$$ \hspace{1cm} (6)

The validation of such simulated yellow band derived from the green and red bands of Landsat 8 imagery was performed by comparison with a WorldView 2 image. The WorldView-2 image has eight bands, including coastal, blue, green, yellow, red, red edge, near-infrared 1 and near-infrared 2 bands. Among them, the real yellow band can be used for the validation of the simulated yellow band based on Landsat 8.

3.3.4. Soil Exposure Index

Soil erosion in forest often occurs in the areas where topsoil is left bare and exposed to raindrop impact and surface runoff due to the absence of understory grass and shrub cover. Information of exposed soils under forest canopy is useful to detect the area prone to erosion. Remote sensing based models could effectively detect bare or exposed soil surface owing to its advances in enhancing soil features from background information [70]. In this study, the Normalized Difference Soil Index (NDSI) of Kearney et al. [71] was adopted to identify exposed soils in forest. The index is expressed as:

$$NDSI = \frac{(MIR1 - NIR)}{(MIR1 + NIR)}$$ \hspace{1cm} (7)

where \textit{MIR1} is the reflectance of the mid-infrared band covering a wavelength between 1570 to 1650 nm.

3.3.5. Slope

The slope factor is represented by slope angle that can be derived from a digital elevation model (DEM) [55]. Current commercial remote sensing software usually has a tool to convert DEM to slope. A 30 m resolution ASTER DEM of Changting was downloaded from the USGS website (https://earthexplorer.usgs.gov) and then converted to a slope angle map.

3.4. Model Development

The soil erosion under forest model (SEUFM) was developed with the factors including FVC, NRI, YLI, NDSI and slope. Due to the difference in unit and data ranges, the five factors have to be normalized within [0, 1] before they can be integrated. The formula used for the normalization is as follows:

$$NI_i = \frac{(I_i - I_{\text{min}})}{(I_{\text{max}} - I_{\text{min}})}$$ \hspace{1cm} (8)

where \textit{NI}_i is the normalized value of pixel \textit{i} of a factor, \textit{I}_i is the value of pixel \textit{i} of the factor, \textit{I}_{\text{min}} and \textit{I}_{\text{max}} are the minimum and maximum values of the factor, respectively.
The five normalized factors were integrated to form SEUFM using two methods, that is, principal components analysis (PCA) and the multiplication approach.

PCA is a statistical technique that converts a set of measurements into a set of values of uncorrelated principal components (PCs) using an orthogonal transformation. PCA automatically weighs the contribution of each variable into each principal component based on the variable’s loading [69]. The integration of the selected five factors in a model involves an evaluation of the contribution of each factor to soil erosion, which is very likely to be different. This suggests that a weight may have to be assigned to each factor. If the soil erosion can be associated with one or more PCs, PCA can be a good method to integrate the selected factors for SEUFM because PCA can automatically quantify the contribution of each factor into each PC according to factor loadings. This can avoid the trouble and uncertainty caused by the prescribed weights [69].

Figure 3 summarizes the main techniques and procedures for the development of SEUFM algorithms addressed above.

![Flowchart of the technical procedures for the development of soil erosion under forest model (SEUFM).](image)

4. Results

4.1. Validation of Two FVC Models and Yellow Leaf Index (YLI)

Figure 4 is the validation result of the two FVC models using the method described in 3.3.1. It shows that FVC\textsubscript{linear} has higher accuracy than FVC\textsubscript{quadratic} as its scatters spread closer to 1:1 line (dotted line in Figure 4) than those of FVC\textsubscript{quadratic} and its RMSE (4.63) is much lower than that of FVC\textsubscript{quadratic} (18.42). The scatters of FVC\textsubscript{quadratic} lie far above 1:1 line, indicating a significant underestimation of FVC\textsubscript{quadratic}. This result is consistent with Zhou et al [60] in which the FVC\textsubscript{linear} is reported out-performing the quadratic FVC\textsubscript{quadratic} for a pinon-juniper woodland. Thus, FVC\textsubscript{linear} is chosen to retrieve FVC for the study area.
The validity of the simulated yellow band using YLI (Equation (6)) was assisted with a WorldView 2 image of Changting, acquired on 13 December 2011. The green and red bands of the WorldView 2 image were mixed to produce a yellow band using Equation (6) and then compared with the real yellow band of the image. Table 1 shows that the simulated yellow band is highly correlated with the real yellow band ($R^2 = 0.984, p < 0.001$) with a nearly identical mean value. This suggests that the yellow band generated using Equation (6) effectively simulates a real yellow band and can be served as a YLI.

Table 1. Comparison between simulated and real yellow bands.

<table>
<thead>
<tr>
<th></th>
<th>WorldView 2 Yellow Band</th>
<th>Simulated Yellow Band (Invariant Patches)</th>
<th>WorldView 2 Yellow Band (Invariant Patches)</th>
<th>Landsat 8 Simulated Yellow Band (Invariant Patches)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.055</td>
<td>0.054</td>
<td>0.055</td>
<td>0.063</td>
</tr>
<tr>
<td>Std Dev</td>
<td>0.034</td>
<td>0.032</td>
<td>0.032</td>
<td>0.030</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.984</td>
<td>0.928</td>
<td></td>
<td>0.928</td>
</tr>
</tbody>
</table>

The actual yellow band of the WorldView 2 image was further compared with the Equation (6)-derived yellow band of the Landsat 8 image. Due to the time difference (three years apart) and seasonal difference (October vs. December) between the two images, only invariant vegetation patches were chosen from both images for the comparison. The comparison yielded an $R^2$ of 0.928 (Table 1 and Figure 5, $p < 0.001$), suggesting that the simulated Landsat 8 yellow band is reliable. The lower agreement than the comparison result of WorldView 2 image (0.984) is probably due to the difference both in season and spatial resolution between the two images though invariant vegetation patches were applied. The seasonal difference between the two images also contributes to a higher mean of the simulated Landsat 8 yellow band, compared with the mean of the real yellow band of the WorldView 2 image, as the Landsat 8 image was acquired in early October while the WorldView 2 image was in December.
4.2. Model Construction

4.2.1. PCA-based Model

The first principal component (PC1) measures the highest variation within the five-factor dataset and each succeeding component has the highest variance in the data which is not already accounted for by previous PCs. Table 2 is a PCA of the five factors, represented by an eigenvector of the covariance matrix in which the contribution of each factor to PCs is given in its loading. Of the five PCs, PC1 has the largest possible variance as its eigenvalue accounts for 80.12% of the variability in the whole data of the study area. PC2 describes the second largest amount of variance in the data that is not already described by PC1 and makes up 13.04% of data variance.

<table>
<thead>
<tr>
<th>Table 2. Eigenvector of covariance matrix and related eigenvalue of each principal component (PC).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PC1</strong></td>
</tr>
<tr>
<td>FVC</td>
</tr>
<tr>
<td>NRI</td>
</tr>
<tr>
<td>NDSI</td>
</tr>
<tr>
<td>YLI</td>
</tr>
<tr>
<td>Slope</td>
</tr>
<tr>
<td>Eigenvalue</td>
</tr>
<tr>
<td>Percent eigenvalue (%)</td>
</tr>
</tbody>
</table>

In PC1 and PC2, the loadings of NDSI and YLI have the same positive sign and thus can be classified in one category. FVC and NRI have the same negative sign for their loadings in both PC1 and PC2 and hence can be assigned to the other group. Based on our understanding, these two groups of factors happen to have contrasting relation with soil erosion. However, the loading of Slope has a negative sign in PC1 but a positive sign in PC2. This may reflect two different mechanisms of the slope condition for soil erosion—(1) a steeper slope reduces human activity and thus decreases the erosion potential and (2) a steeper slope enhances erosion due to water and landslide potential. Thus, it is likely that PC1 and/or PC2 can be used to quantify soil erosion.

The succeeding components from PC3 to PC5 are neglected as they account only for a very small amount of data variance and the sign of the loading of each factor is not interpretable in terms of their possible relationship with soil erosion.
Given that PC1 accounts for 80% variance of the five factors, the first SEUFM option (SEUFM$_1$) is based on PC1:

$$SEUFM_1 = PC1_{(FVC, NRI, YLI, NDSI, Slope)}$$ (9)

The second SEUFM option (SEUFM$_{1&2}$) is based on the sum of PC1 and PC2 in order to consider the positive sign of the slope factor. We did not use PC2 alone as it accounts only for 13% of variance within the whole data.

$$SEUFM_{1&2} = (PC1 + PC2)_{(FVC, NRI, YLI, NDSI, Slope)}$$ (10)

To calculate the SEUFM$_1$ and SEUFM$_{1&2}$, the images of five factors were used as five bands to form a new image. The PCA was then performed on the new five-band image using statistic/remote sensing software. The calculated results were normalized to have a value ranging from 0 to 1. The greater the value, the more likely the occurrence of soil erosion under tree canopy.

4.2.2. Multiplication-based Model

The model was developed by the integration of the five factors through multiplication. As indicated by PCA, the sign of FVC and NRI is negative, whereas the sign of Slope is either negative or positive. Accordingly, the model can be developed in the following two forms:

$$SEUFM_{m-slope} = (1 - FVC) \cdot (1 - NRI) \cdot (1 - Slope) \cdot YLI \cdot NDSI$$ (11)

$$SEUFM_{m+slope} = (1 - FVC) \cdot (1 - NRI) \cdot Slope \cdot YLI \cdot NDSI$$ (12)

The calculated results of SEUFM$_{m-slope}$ and SEUFM$_{m+slope}$ also need to be normalized within [0, 1]. A greater value indicates a more likelihood of the occurrence of soil erosion in forest.

4.3. Detecting Soil Erosion Potential in Forest

Four thematic images corresponding to the factors of FVC, NRI, YLI and NDSI were then retrieved from the Landsat 8 image of Changting and shown in Figure 6. The croplands were masked out from the thematic images since this study focuses on forest areas. The four thematic factor images coupled with DEM-derived Slope factor image were then used as input data to calculate the SEUFMs using the Equations of (9) to (12).

Independent field investigations, led by the staff members from the Fujian Monitoring Station for Water and Soil Conservation, were carried out in Changting during the rainy seasons of 2014 and 2015 to validate the extracted results. The staff members of the station well know the locations where soil erosion in forest occurs based on the historical monitoring data of the station mentioned above and their local knowledge and experience. A total of 77 accessible sites in forest were hence chosen (Figure 2), which includes 57 sites of soil erosion and 20 sites of non-erosion. Statistics show that the differences in the five factors between the erosion and non-erosion sites are substantial (Table 3).

| Table 3. | Statistical characteristics of the five factors in soil erosion and non-soil erosion sites. |
|----------------|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|              | Soil Erosion                  | Non-Soil Erosion |                |                |                |                |                |                |                |
|               | Min     | Max     | Mean    | Std Dev | Min     | Max     | Mean    | Std Dev | Percent Difference (%) | p-Value |
| Slope         | 0.070   | 0.541   | 0.219   | 0.096   | 0.026   | 0.397   | 0.206   | 0.135   | 6.31            | 0.821   |
| FVC           | 0.454   | 0.785   | 0.651   | 0.074   | 0.721   | 0.897   | 0.804   | 0.064   | -19.03          | 0.000   |
| NRI           | 0.283   | 0.467   | 0.358   | 0.040   | 0.375   | 0.594   | 0.476   | 0.065   | -24.78          | 0.003   |
| NDSI          | 0.069   | 0.524   | 0.283   | 0.084   | 0.110   | 0.275   | 0.184   | 0.049   | 53.80           | 0.001   |
| YLI           | 0.214   | 0.433   | 0.309   | 0.049   | 0.172   | 0.292   | 0.226   | 0.048   | 36.73           | 0.004   |
Figure 6. Thematic images of the five factors. (a) FVC, (b) NRI, (c) NDSI, (d) YLI and (e) Slope. White areas in (a) to (d) are the areas with a slope degree less than 5°.

The validation results show that for the two PCA-based models, the highest accuracy was achieved when the threshold value was adjusted to ~1.045 times of the mean. While for the two multiplication-based models, the optimum threshold was roughly the mean ±10%, respectively (Table 4). Values greater or smaller than the thresholds all resulted in a decrease in accuracy.

Table 4. Mean and threshold values of each SEUFMs.

<table>
<thead>
<tr>
<th>SEUFM</th>
<th>Mean/Initial Threshold</th>
<th>Final Threshold</th>
<th>Threshold to Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEUFM₁</td>
<td>0.355</td>
<td>0.371</td>
<td>1.045</td>
</tr>
<tr>
<td>SEUFM₁+slope</td>
<td>0.350</td>
<td>0.366</td>
<td>1.046</td>
</tr>
<tr>
<td>SEUFMₘ-slope</td>
<td>0.153</td>
<td>0.166</td>
<td>1.089</td>
</tr>
<tr>
<td>SEUFMₘ+slope</td>
<td>0.088</td>
<td>0.079</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Among the four SEUFM models, SEUFM₁, the model based on PC1, achieves the highest accuracy close to 90% and SEUFMₘ-slope, the multiplication-based model, has the second highest accuracy of 81% (Table 5). Thus, SEUFM₁ is chosen as the optimum model to identify soil erosion locations under forest in Changting. The result indicates a soil erosion area of 312.76 km² in the forests of the county (Figure 7), which accounts for 12.58% forest in the county. The detected result (312.76 km²) is within the area (300~350 km²) estimated by the county’s Monitoring Station for Water and Soil Conservation. In addition, Zhang et al. [5] conducted a similar study to identify forest soil erosion in Changting but only concentrated their study area in Hetian town. Using a vegetation-restoration-degree (VRS) model, they found that the eroded forested area in Hetian was 67 km² in 2009. This is comparable with
our result of the same area (64.8 km$^2$). A 2.2-km$^2$ difference could be due to a 5-year time difference between the two studies.

Table 5. Accuracy assessment of the four SEUFMs.

<table>
<thead>
<tr>
<th>SEUFM</th>
<th>Erosion</th>
<th>Non-Erosion</th>
<th>Total</th>
<th>User's Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SEUFM_1$ (threshold: 0.371)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion</td>
<td>51</td>
<td>2</td>
<td>53</td>
<td>96.23</td>
</tr>
<tr>
<td>Non-Erosion</td>
<td>6</td>
<td>20</td>
<td>26</td>
<td>76.92</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>22</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Producer’s accuracy (%)</td>
<td>89.47</td>
<td>90.91</td>
<td>Kappa</td>
<td>0.761</td>
</tr>
<tr>
<td>Overall accuracy (%)</td>
<td>89.87</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SEUFM_{1+2}$ (threshold: 0.366)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion</td>
<td>44</td>
<td>6</td>
<td>50</td>
<td>88.00</td>
</tr>
<tr>
<td>Non-Erosion</td>
<td>13</td>
<td>16</td>
<td>29</td>
<td>55.17</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>22</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Producer’s accuracy (%)</td>
<td>77.19</td>
<td>72.73</td>
<td>Kappa</td>
<td>0.455</td>
</tr>
<tr>
<td>Overall accuracy (%)</td>
<td>75.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SEUFM_{m-slope}$ (threshold: 0.166)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion</td>
<td>48</td>
<td>6</td>
<td>54</td>
<td>88.89</td>
</tr>
<tr>
<td>Non-Erosion</td>
<td>9</td>
<td>16</td>
<td>25</td>
<td>64.00</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>22</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Producer’s accuracy (%)</td>
<td>84.21</td>
<td>72.73</td>
<td>Kappa</td>
<td>0.547</td>
</tr>
<tr>
<td>Overall accuracy (%)</td>
<td>81.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$SEUFM_{m+slope}$ (threshold: 0.079)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Erosion</td>
<td>40</td>
<td>10</td>
<td>50</td>
<td>80.00</td>
</tr>
<tr>
<td>Non-Erosion</td>
<td>17</td>
<td>12</td>
<td>29</td>
<td>41.38</td>
</tr>
<tr>
<td>Total</td>
<td>57</td>
<td>22</td>
<td>79</td>
<td></td>
</tr>
<tr>
<td>Producer’s accuracy (%)</td>
<td>70.18</td>
<td>54.55</td>
<td>Kappa</td>
<td>0.225</td>
</tr>
<tr>
<td>Overall accuracy (%)</td>
<td>65.82</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 7. Detected soil erosion areas under forest canopy in Changting County (a) and a 3D perspective of soil erosion areas in forests in the central county (Hetian basin) area (b).

Figure 7 clearly shows that the soil erosion in forests of Changting occurs mainly along the Hetian basin roughly in a north-south direction but is concentrated at the lower part of the hillside with little density of forest and therefore, where the protective effect of canopy is low (Figure 7b). The mapped
erosion areas are consistent with our understanding of soil erosion development in relation to the characteristic of the pedospheric condition, vegetation cover and human activities in the study area.

5. Discussion

This study has revealed that the Hetian basin is the main area susceptible to forest soil erosion in Changting. The basin has long been the most severe soil loss areas in the county due to intensive disturbance by human activities [5,23]. In order to control soil erosion, a three-decade reforestation program has been carried out through restoration of vegetation in the previously deforested lands. Since the early 1980s, the local residents have been planting a great number of Pinus massoniana trees in the areas. However, the red soils developed from parent granitic rocks in the region are acidic and often low in nutrients. The iron oxide-rich soils inhibit the growth of shrub and grass [5]. Due to low soil fertility, a lack of shrub/grass sub-storey covers, as well as frequent human activities, the rehabilitated forests, particularly those developed in the hillslopes surrounding the Hetian basin (Figure 7b), are in poor health status and thus susceptible to soil erosion.

The study results show that the PC1-based model, SEUFM, can detect soil erosion areas in forest with high accuracy. This suggests the effectiveness of the selected five factors in association with soil erosion. Figure 8 indicates the degree of separation of each factor between erosion and non-erosion sites of the 77 field validation locations.

Figure 8. Separability of each factor between erosion and non-erosion sites (Y: Erosion, N: Non-Erosion).

Among the five factors, FVC, NRI, YLI and NDSI can be approximately separated between erosion and non-erosion sites within the range of the mean ±1 standard deviation. Table 3 indicates that compared with the non-erosion sites, the erosion sites have lower mean values in FVC and NRI by 19.0% and 24.8%, respectively and higher mean values in NDSI and YLN by 53.8% and 36.7%, respectively. These differences are tested significantly as their p-values are much less than the significance level of 1% (Table 3). However, the data ranges of Slope between erosion and non-erosion sites are overlapped. Table 3 shows that although the mean of Slope of the erosion sites is higher than that of the non-erosion sites by 6.31%, the very high p-value (0.821) suggests insignificance of the difference. This supports our previous discussion that the slope effect on soil loss in forest has two opposite mechanisms. Generally, high-angle slopes are conducive to high levels of erosion. However, this may not be always the case in forested areas where the slopes are covered with forests with various densities. Areas of low slope angles may suffer more erosion because lower slope inclinations are usually subject to more frequent human activities such as walking, weeding and firewood gathering and thus have no or poorly-developed sub-storey grass/shrub mulch. The study conducted by Zhang et al. [5] found that
the soil erosion in forests in Hetian mainly occurred in areas with slope angles of 8° to 15° and became weaker when the slope was steeper than 15°.

The dual effects of the slope factor explain why the slope factor has a negative loading sign in PC1 but a positive loading sign in PC2. Initially, we considered excluding the slope factor for the SEUFM development due to its dual effects but our experiments showed that a four-factor model without Slope does not perform as well as the one with five factors. This suggests the importance of Slope for the SEUFM.

Two models that consider the erosion-enhancement effect of the slope factor are SEUFM_{1+2} and SEUFM_{m+slope}. Neither of them performs as well as the ones without considering this erosion enhancement effect (Table 5). This suggests that the slope may play a role in reducing human activity more than in enhancing erosion in the studied forest areas.

In this study, the two multiplicative models, SEUFM_{m+slope} and SEUFM_{m+slope}, have lower accuracy than two PCA-based models, SEUFM_{1} and SEUFM_{1+2}, correspondingly (Table 5). This is likely related to the fact that the factors are equally weighted in the models. The multiplicative method assumes that each factor reflects the erosion independently. Hence, they can be multiplied to obtain a lumped and enhanced association with the erosion. However, the correlation matrix (Table 6) indicates that the assumption of independence between the selected factors is not supported in the study area as the selected five factors are more or less related. Thus, there is a need to somehow weight the different information embedded in these erosion-associated but interdependent factors for erosion estimation. While, the PCA-based method provides such a weighing approach.

Table 6. Correlation matrix of the selected factors.

<table>
<thead>
<tr>
<th></th>
<th>Slope</th>
<th>FVC</th>
<th>NRI</th>
<th>NDSI</th>
<th>YLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slope</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC</td>
<td>0.414</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NRI</td>
<td>0.406</td>
<td>0.869</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NDSI</td>
<td>−0.324</td>
<td>−0.829</td>
<td>−0.705</td>
<td>1.000</td>
<td>0.767</td>
</tr>
<tr>
<td>YLI</td>
<td>−0.421</td>
<td>−0.897</td>
<td>−0.759</td>
<td>0.767</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The threshold values listed in Tables 4 and 5 for separating erosion and non-erosion sites in forest are around the mean values of SEUFM images (within ±10%). This suggests that the mean could be initially used as a threshold but is subject to a manual adjustment to achieve an optimum threshold. This method for selecting threshold is probably specific to the study area and may not be suitable for other areas. Nevertheless, the selected five factors for SEUFM are rooted in the concept of soil erosion conditions in forested areas. Therefore, the method is not area-dependent and can be used as a spatially-explicit tool to evaluate the soil erosion potential in forest.

SEUFM is the model especially designed for detecting soil erosion occurring in forested areas. Therefore, apart from the factors that are usually selected for soil erosion detection like slope and soil factors, SEUFM is enhanced with more vegetation-related factors such as yellow leaf index, nitrogen reflectance index, as well as FVC. This makes it different from other soil erosion related models that usually have only one vegetation-related factor, such as RUSLE, LISEM and WEPP. Zhang et al. [5] also detected soil erosion in forest in Changting. They developed a VRD model in which two vegetation-related factors (FVC and leaf area index) were involved. Their result is in general agreement with this study, in spite of the differences in model factors and model construction between the two methods. The other difference between SEUFM and other models is that SEUFM is derived completely from remote sensing data and is designed to detect sites prone to soil erosion in forest but not to quantitatively measure the amount/intensity of soil loss. Therefore, the result of SEUFM is a binary image. Each pixel of the image is represented by a value of 0 or 1 (indicating erosion or non-erosion state) rather than a value representing the amount of soil loss. Nevertheless, the model may be used to obtain qualitative values by classifying soil erosion into high, moderate and low levels,
as the modelling result is represented by a value between 0 and 1. The greater the value, the more likely the occurrence of soil erosion.

Remote sensing has been served as a measurement surrogate of spatial distributions of soil-erosion status and provides accurate detection of soil erosion sites for decades. Nevertheless, remote sensing based mapping of soil erosion in forest remains a challenge because the methods specially developed for this have seldom been addressed in previous studies. This study has demonstrated an application of a full remote sensing based model (SEUFM) in soil erosion detection in forested areas. The five factors of the model are easily obtained as they are all derived from remote sensing data alone. This will overcome the difficulty in data availability [1]. The integration of the five factors using PCA also avoids the weight allocation that may cause errors in the results due to subjective selections of weights. The method proposed in this paper will facilitate quick and effective detection of soil erosion in forest over a very large area, which makes SEUFM a relatively simple soil erosion monitoring tool for forested lands.

6. Conclusions

Soil erosion in forests is closely related to, and can be reflected in, factors such as forest coverage, forest health status, soil exposure intensity and slope. These are the factors considered in this study for the development of the SEUFMs. Of the two types of SEUFM models developed in this study, the PCA-based method achieved higher accuracy than the multiplication-based method because it handles the five factors according to their contributions (loadings) to each PC. The lower accuracy of the multiplication-based models is due to its assumption of an equal independent association of the five factors with soil erosion. Among the four developed models, SEUFM\(_1\), the PC1-based model, mapped the areas of soil erosion in forest with the highest accuracy.

Whether a model is appropriate for a certain application depends on a number of factors, including data requirement and ease of use. The lack of data is usually the main drawback of model applications. The proposed method in this study is completely based on remote sensing techniques and is relatively easy to use. The required data of each factor for the model can be derived directly from remote sensing imagery. This advantage makes the model more applicable. SEUFM is intended for use as a spatially-explicit tool to detect areas with potential soil erosion in forest. The method-derived results would be useful for decision makers in the development of future soil conservation strategies to reduce soil loss in forest.

**Author Contributions:** H.X. conceived the idea and designed the experiments; H.X., X.H., B.Z. and M.W. performed the experiments; S.C. and M.C. led the ground truth; H.X. analysed the data and wrote the paper; H.G. reviewed and commented on the manuscript. All authors gave final approval for publication.

**Funding:** This research was funded by China’s National Science and Technology Supporting Project, grant number 2013BAC08B01-05 and Fujian Provincial Monitoring Station of Water and Soil Reservation, grant number MSK201704.

**Acknowledgments:** We appreciate the six anonymous reviewers for their helpful comments and suggestions on the manuscript.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

**References**


38. Piacentini, T.; Galli, A.; Marsala, V.; Miccadei, E. Analysis of Soil Erosion Induced by Heavy Rainfall: A Case Study from the NE Abruzzo Hills Area in Central Italy. *Water* 2018, 10, 1314. [CrossRef]


41. Mateos, E.; Miguel, E.J.; Ormaetxea, L. Soil erosion and forests biomass as energy resource in the basin of the Oka River in Biscay, northern Spain. *Fores* 2017, 8, 258. [CrossRef]

42. Ochoa, P.A.; Fries, A.; Mejia, D.; Burneo, J.I.; Ruiz-Sinoga, J.D.; Cerda, A. Effects of climate, land cover and topography on soil erosion risk in a semiarid basin of the Andes. *Catena* 2016, 140, 31–42. [CrossRef]


