Large scale soil erosion modeling for a mountainous watershed

P. Zhou\textsuperscript{1}, J. Nieminen\textsuperscript{2}, T. Tokola\textsuperscript{2}, O. Luukkanen\textsuperscript{1} & T. Oliver\textsuperscript{3}

\textsuperscript{1}Viikki Tropical Resources Institute, University of Helsinki, Finland
\textsuperscript{2}Department of Forest Resource Management, University of Helsinki, Finland
\textsuperscript{3}University of Barcelona, Spain

Abstract

Soil erosion control requires a quantitative evaluation of potential soil erosion on a specific site. The Revised Universal Soil Loss Equation (RUSLE), Remote Sensing (RS), and Geographic Information System (GIS) were used to model soil erosion intensity for soil conservation and vegetation rehabilitation in an Upper Min River (UMR) watershed, which is in the Upper Yangtze River basin. Data used in this study to generate the soil loss were Landsat Enhanced Thematic Mapper (ETM) images, Digitized Elevation Model (DEM), soil erodibility, rainfall erosivity, and inventory data. The non-parametric $k$-nearest neighbor ($k$-NN) method was used to produce the cover management map by integrating the ETM images and vegetation coverage data measured in the 625 sample plots. The root mean square errors and significance of biases at pixel level were evaluated in order to find optimal parameters. Four raster maps have been produced for the soil erodibility, rainfall erosivity, slope length and steepness, and cover management factor, and the map with different soil loss risks has been produced for soil erosion potential. The result can be beneficial to the erosion control and ecological restoration in the degraded mountainous watershed.

Keywords: soil erosion, RUSLE, DEM, $k$-NN method, Upper Min River Watershed.

1 Introduction

Soil erosion is a worldwide environmental problem that degrades soil productivity and water quality, causes sedimentation and increases the
probability of floods. The 1998 flood in the Upper Yangtze raised public attention to the problems of soil erosion and sedimentation. Soil loss control calls for a quantitative evaluation of potential soil erosion on a specific site. Scientists have studied different methods to assess soil erosion loss by water, for instance, universal soil loss equation (USLE) [1], revised soil loss equation (RUSLE) [2], the $^{137}\text{Cs}$ technique [3], and WEPP Hillslope model [4]. Among them, the USLE and RUSLE provided a convenient tool for soil loss evaluation by taking the climate, geographical terrain, conservation support practice, soil, and vegetation into considerations. The RUSLE was developed to incorporate the considerable amount of erosion information and to address specifically the application of the USLE to land uses other than agriculture [2]. The model can be used to any geographic region by modifying its factors. The factors are rainfall runoff erosivity factor, soil erodibility factor, slope length and steepness factor, cover management factor, and support practice factor [1, 2]. An important reference manual for applying the USLE to disturbed forestlands is “A guide for predicting sheet and rill erosion on forest land” [5]. The USLE and RUSLE have been used widely in evaluating the soil erosion risk in watershed and highland [6, 7, 8].

Slope length and steepness factor, which reflects the terrain on a given site, can be computed from the digital elevation model (DEM) [9, 10]. Rainfall and runoff erosivity factor was calculated based on the storm events and rainfall data in many studies [6, 7, 8, 11]. However, in mountainous watershed, orographic effects caused by mountainous terrain can result in a significant positive correlation between precipitation and elevation [12, 13, 14]. In the UMR watershed, precipitation tends to increase with an increase in elevation because of the orographic effect of mountainous terrain and the foehn effect [15]. In our study area, we examined the relationship between elevation and precipitation for 38 stations, and selected cokriging as a method for estimating average annual precipitation of the whole watershed.

Cover management C-factor in the soil loss equation was defined as the ratio of soil loss from land cropped under specified conditions from the corresponding loss from clean tilled, continuous fallow [1]. However, in large scale UMR watershed, where are not mainly covered by agricultural lands, the cover management factor is not only affected by the agricultural crops. Ma et al. used the proposition of vegetation reflectance in pixel end members to calculate C factor [7]. In our study, we used the non-parametric $k$-nearest neighbour ($k$-NN) multi-source estimation method to estimate coverage data and produce the coverage map by integrating the satellite images and field data with optimal parameters. The $k$-NN method has been widely used in a variety of forest estimation and biomass mapping applications over the years [16, 17, 18, 19], and therefore, can be applied in vegetation cover estimation.

Lu et al. [20] explored the relationships between the soil erosion and land use and land cover distribution, they found that most climax and mature forests are in low erosion risk areas, while agroforestry and pasture are usually associated with medium to high risk areas. A good plant cover is generally capable of preventing surface erosion, and reducing landslides as well. Removal of vegetation can greatly increase runoff and soil erosion particularly in...
mountainous areas [21]. Soil erosion control especially calls for the forest restoration or rehabilitation to reduce the erosion loss and improve soil stability.

The questions we are going to answer in the paper are: (1) How to model the soil erosion loss in this mountainous watershed? and (2) How much is the soil erosion risk in the area?

2 Study area

The Upper Yangtze River Basin is a mountainous region, which has an area of 1.04×10^6 km^2, a mean annual runoff discharge of 4.35×10^8 m^3, a mean sediment yield of 5.17×10^8 t and a population of 1.4×10^8 [22]. The basin is one of the most severely eroded areas in China. Water erosion results in both on-site soil degradation and off-site problems related to downstream sedimentation [23].

The Upper Min River, which is one of the most important tributaries of the Upper Yangtze River, is 341 km long with a drainage area of 23,037 km^2. The watershed is located in Sichuan Province, South West China. The area is governed by the southeast and southwest monsoons. The complex topography, with elevations ranging from 900 m to 5700 m, results in steep gradients of rainfall. The Upper Min river watershed has been divided into five ecozones: the Sub-tropical (1300–2200 m), Temperate (2200–2600 m), Sub-alpine (2600–3200 m), Boreal (3200–3600 m) and Arctic zone (3600–5700 m) [24]. At present, the forest cover is around 21% of the whole watershed area. Our 625 inventory plots were randomly placed in the middle and upper reaches of the UMR watershed, between 31º-34º N, 103º-104º E, with an area of about 7400 km^2, see fig. 1. The vegetation ranges from subtropical evergreen broadleaved forest to the alpine meadows.

3 Method

3.1 Model structure

The soil loss (A) due to water erosion per unit area per year (Mg ha\(^{-1}\) yr\(^{-1}\)) was quantified using RUSLE by the following equation:

\[
A = R \times K \times LS \times C \times P
\]

where \(A\) is the average soil loss due to water erosion, \(R\) the rainfall and runoff erosivity factor (MJ mm ha\(^{-1}\)h\(^{-1}\)yr\(^{-1}\)), \(K\) the soil erodibility factor (Mg h MJ\(^{-1}\) mm\(^{-1}\)), \(L\) the slope length factor, \(S\) the slope steepness factor, \(C\) the cover and management practice factor, and \(P\) the support practice.

3.2 Data and processing

3.2.1 Rainfall and runoff erosivity factor (\(R\))

\(R\) is the long term annual average of the product of event rainfall kinetic energy (E) in MJ ha\(^{-1}\) and the maximum rainfall intensity in 30 minutes (\(I_{30}\)) in mm h\(^{-1}\).
The R values were correlated with annual precipitation \[1, 25\]. We used the following equation to calculate the R factor, which has been adopted for application in the RUSLE model \[2\]:

\[
R = -0.0334P_a + 0.006661P_a^2
\]  

(2)

where \( R \) is rainfall and runoff factor (MJ mm ha\(^{-1}\) h\(^{-1}\) yr\(^{-1}\)), and \( P_a \) the measured annual precipitation in mm.

The average annual precipitation (APP) and elevation data from 38 meteorological stations in the research area were obtained to check the correlation between precipitation and elevation. The APP surface was interpolated with a multivariate geostatistic cokriging model \[26\]. The R factor surface was then calculated by eqn. (2) from the APP surface using the raster calculation in spatial analyst.

Figure 1: Upper Yangtze River watershed and sample plots in the research area. (•) showed the 625 sample plots, (——) showed the drainage network, (■) showed the Upper Yangtze River basin, and (□) showed the detail research area.
3.2.2 Slope length and slope steepness factors (LS)

$L$ factor and $S$ factor, which reflect the topographic erosion susceptibility on a given site, were computed together from the digital elevation model (DEM). The DEM used is based on a digital topographic map, with 100-m elevation contour lines and stream data. In order to achieve a geomorphological realistic surface, it was interpolated to a 25-m cellsize grid with the Topogrid algorithm [27] which generates a hydrologically correct grid DEM using contour lines and stream data. The slope was calculated using the maximum downhill direction method, in which the slope value for each raster cell is obtained from the angle formed between the cell itself and the lowest neighboring cell. The flow direction was calculated with the $D_{\infty}$ (infinite directions) method developed by Tarboton [28], by which dispersed or rilled flow is estimated for each cell from the slopes to the lower neighboring cells. In flat areas of the DEM (no lower neighboring cells), the method approached by [29] was used to calculate flow direction. Flow accumulation, the number of cells contributing with its flow to each cell was calculated from the flow direction raster. The DEM sinks filling the slope angle, the flow direction and the flow accumulation were calculated by Taudem, an ArcGIS 9.0 extension developed by Tarboton. For this project, an approach developed by Moore and Burch [9, 10] was used to compute LS factor:

\[
LS = L \times S \times 1.4
\]

\[
L = (fa \times cs / 22.13)^m
\]

where $LS$ is computed slope length ($L$) and slope steepness ($S$) factor, $fa$ the flow accumulation (contributing area or upslope area), $cs$ the cellsize, and $m$ the slope-length exponent, as explained in the equation:

\[
m = \beta / (1 + \beta)
\]

where $\beta$ is the ratio of rill to the interrill erosion for conditions when the soil is moderately susceptible to both, and is computed by $\theta$ with the following equation [30]:

\[
\beta = (\sin \theta / 0.0896) \left[3.0 \times (\sin \theta)^{0.8} + 0.56\right]
\]

Where $\theta$ is slope angle in degree. Table 1 shows the values for $m$ computed from eqn (6) and (7), and applied to eqn (4) to calculate a raster map for $L$ factor. $S$ is calculated by the following equations:

\[
S = 10 \times \sin \theta + 0.03 \quad \text{If slope < 9 percent}
\]

\[
S = 16.8 \times \sin \theta - 0.05 \quad \text{If slope \geq 9 percent}
\]
3.2.3 Soil erodibility factor ($K$)

K factor is soil erodibility factor, which represents both susceptibility of soil to erosion and the rate of runoff. Specifically, the $k$ factor is a function of particle size distribution, organic matter content, structure and permeability [1, 2, 11]. $K$ was calculated using the equation recommended by Wischmeier and Smith [1]:

$$K = (2.1 \times (vf^2 + vf \times sa)^{1.14} \times 10^{-6} \times (12 - om) + 0.0325 \times (pe - 2) + 0.025 \times (st - 3)) \times 0.1317$$ (10)

Where $K$ is the soil erodibility factor (Mg h MJ$^{-1}$mm$^{-1}$), $vf$ the percentage of very fine sand plus silt, $sa$ the percentage of sand, $om$ the percentage of organic matter, $pe$ the permeability class, and $st$ the structure class. In this study area, eqn (10) was used to calculate $K$ value of each soil type. $K$ values and the map of soil type were used to produce the raster map of $K$ factor.

3.2.4 Cover management factor ($C$)

Vegetation cover at three levels (canopy cover, under canopy cover, surface cover) was recorded from 625 sample plots. The canopy cover was measured by densiometer. The $k$ nearest neighbor ($k$-NN) method was used to produce the canopy cover map and total vegetation cover map by integrating Landsat ETM+ image information and collected vegetation coverage. In this study we used two consecutive Landsat ETM+ scenes (WRS2 130/037 & 130/038, 10th July 2002 [31]) for the cover factor estimation. A set of parameters was chosen for the $k$-NN method in predicting the vegetation coverage map. The parameters were the image features, the weight for each band, the distance, the number of nearest neighbors the value of $k$, and the geographical reference area from which the nearest field plots are selected. The leave-one-out cross validation method is applied to calculate the root mean square errors (RMSE) and the average biases of predictions at the single pixel level for different combination of $k$-NN estimation. The RMSE and biases were used as a measure of reliability of the continuous variables. The cover management factor ($C$) then was calculated from vegetation coverage data using the equation recommended by Renard et al. [2].

$$CC = 1 - F_e \times \exp(-0.03048 \times H)$$ (11)
where $CC$ is the canopy cover subfactor range from 0 to 1, $F_c$ is fraction of land surface covered by canopy, and $H$ (m) is distance that raindrops fall after striking the canopy. The equation was used to calculate both the canopy cover subfactor and under canopy cover subfactor. For the former, we used the average weighted tree height 8.341m; and for the latter, we used the estimated shrub and grass height 0.5 m.

3.2.5 Support practice factor ($P$)

$P$ values range from 0 to 1, value 0 represents a very good manmade erosion resistance facility, and value 1 represents no manmade resistance erosion facility. In the study area, there are some agricultural support practices [7]. However, most of the farmlands in the study area are small and self-managed lands, and the spatial resolution of ETM+ imageries is 30 m, so it is impossible to distinguish the practices in a large scale watershed from the available data. We use $P = 1$ for all the lands.

4 Results

The historical precipitation data and station elevations from 38 meteorological stations were obtained to estimate the average annual precipitation (AAP) over the entire watershed. The AAP showed a significant ($p < 0.01$) correlation of $r = 0.74$ with the station elevation. A multivariate cokriging interpolation method was used in the analysis since it takes into consideration the elevation which significantly affects precipitation. Root mean square errors (RMSE) were calculated to investigate the estimation accuracy. The RMSE by cokriging estimation was 86.88 mm, which was reduced by 28.2% to kriging estimation (121.2 mm). The estimated AAP was used for calculation of rainfall and runoff erosivity R-factor in ArcGIS. The R factor varied from 1288 to 3342 MJ mm ha$^{-1}$ h$^{-1}$ yr$^{-1}$ (fig. 2).

The watershed occupied a raster grid space of 7700 rows by 2736 columns, and elevations ranged from 1261 m to 5537 m. Approximately 94.7 percent of the watershed has slopes steeper than 9 percent. Slope angles ranged from 0 to 77.2 degrees with a mean of 25.9 degrees and standard deviation of 12.0 degrees. As a result of applying Taudem, flow accumulation ranged from 1 to 15496180 m with 98% less than 1000 m, and slope length factor ranged from 0 to 3398, with a mean of 85.9 and 99.2% less than 120. The slope steepness $S$ factor varied between 0.03 to 15.88 with a mean of 6.7 and a standard deviation of 3.1.

The canopy cover map and total vegetation cover map were produced using $k$-NN method. The root mean square errors (RMSE) and the average biases of predictions at the single pixel level were evaluated for each combination of parameters. The value of $k$ (8), the distance (55 km), the bands (1, 2, 3, 4, 5, 7) and their optimal weights were chosen when RMSE and bias were minimal. The $C$ factor was calculated from the produced raster maps by using eqn (5). The cover management $C$ factor ranged from 0.015 to 0.892 (fig. 2).
The soil erodibility K factor was between 0.036 and 0.043 Mg h MJ⁻¹mm⁻¹ (fig. 2). The estimated soil loss for the research area varied from 325 to 83240 Mg ha⁻¹ per year. According to the soil loss amount and field inventory result, we divided them into four ordinal classes: extreme risk (> 10000), high risk (3000 - 10000), moderate risk (1000 - 3000), low risk (< 10000) and No data (fig. 2). No
data values were assigned in two circumstances: firstly, excluded data with flow accumulation values higher than 1200 m (1.7% of the cells), which are coincident with the main stream paths; and secondly, excluded data with LS factor values higher than 1600 (0.0023% of the cells), which happened only in isolated cells with extremely high slopes and contributing areas. Totally 0.7% of the cells had No data value. Table 2 showed the area and proportion of each of the soil erosion potential categories. More than half of the watershed (58.1%) showed moderate, high or extremely high erosion risks.

Table 2: Derivation of the ordinal categories of soil erosion potential and the area and proportion of each category.

<table>
<thead>
<tr>
<th>Numeric range</th>
<th>Erosion potential</th>
<th>Area (ha)</th>
<th>Proportion (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 1000</td>
<td>low</td>
<td>297007.7</td>
<td>40.2</td>
</tr>
<tr>
<td>1000 - 3000</td>
<td>moderate</td>
<td>306909.9</td>
<td>41.6</td>
</tr>
<tr>
<td>3000 - 10000</td>
<td>high</td>
<td>120109</td>
<td>16.3</td>
</tr>
<tr>
<td>&gt;10000</td>
<td>extreme</td>
<td>9202.3</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>no data</td>
<td>4896.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>

5 Discussion

In the UMR watershed, the average annual precipitation was positively correlated with elevation ($r = 0.74$, $p < 0.01$), which supports similar findings in Algarve (Portugal), southern Nevada and southeastern of California ($r = 0.75$, $p < 0.05$) [12, 13, 14]. The effect of elevation on precipitation can be used to improve the geostatistical interpolation. The RMSE by cokriging estimation was reduced by 28.2% to kriging estimation in the UMR watershed, and 54% reduction has been reported in Nevada and southeastern of California [13].

Slope calculations made with a maximum downhill method conserved the variability and the maximum slope values. This method produced no underestimation, since no averaging was used. Flow directions calculated by $D_\infty$ (infinite directions) improved significantly the water flow modelling, by allowing dispersed flow to be modelled over the surface. This method calculated the flow direction from the lowest continuous neighbouring cells and fractionated the water flow between them, simulating dispersed water flow and generating natural looking flow maps. Other studies mostly use the D8 approach method, by O'Callaghan and Mark [32]. However, D8 method produces unrealistically rilled water flow with lots of straight lines in flow accumulation maps, because it can only produce 8 different flow directions, to one of the neighbouring cells (cardinal or diagonal direction).

The $K$ values of the soils in our study area ranged from 0.036 to 0.043 Mg h MJ$^{-1}$mm$^{-1}$. Compared to the $K$ values of tested soils in USLE (0.03 - 0.69 tons acre hr/hundreds of acre ft-ton in) [1], which are from 0.004 to 0.09 Mg h MJ$^{-1}$mm$^{-1}$, the soils in the UMR watershed have the moderate erodibility.
To find a suitable C-factor, a canopy cover and total vegetation surface cover were calculated using the $k$-NN technique for their estimation. Several methods were applied for topographic normalization of the imagery [33]. However, the elevation changes in the study area were so great (1254 m - 5527 m) that not all of the shadowing effects were removed from the imageries. To maximize spectral variability bands 1-5 and 7 were included in the analysis. The calibration of the $k$-NN parameters was performed as outlined by several articles dealing with forest estimation using $k$-NN methods [27, 34, 35]. The RMSE and significance of biases at sample plot pixel level were evaluated in order to choose the most optimal parameters, such as numbers of $k$, distance, and bands weights. The value of $k$ (8) was chosen when the variation in bias and RMSE were minimal. The built in cross-validation method of bias and error estimation was applied in all calculations.

Morgan [36] argues that 10 Mg ha$^{-1}$yr$^{-1}$ is an appropriate boundary measure of soil loss over which agriculturists should be concerned. This was identified as the separation of the low and moderate categories in RUSLE [2]. Soil loss in highland conditions in Kenya ranged from 30 to 666 Mg ha$^{-1}$yr$^{-1}$ [8], while the calculated soil loss (325 - 83240 Mg ha$^{-1}$yr$^{-1}$) in our research area is much higher. The complex terrain, with elevations from 1261 m to 5537 m, and slopes from 0 to 77.2 degree could be one reason to compute so high soil loss potential. Van Remortel et al [37] argue that erosion model can be used to derive patterns of erosion, but not necessarily the actual loss of erosion, because of the limitations of the methods used to derive some component factor values. Millward and Mersey [6] found that relative comparisons of soil loss among land areas are more critical than assessing the absolute soil loss in a particular cell. A visual interpretation and validation of the resulting erosion risk map was performed for all the sample clusters. The sites were given a subjective risk scale ranging from No risk – Low – Moderate – High – Extreme bases upon the general site characteristics. The high or extreme high erosion risks mostly occurred on the downhill gullies with long proceeding slope lengths (red areas in the map). Some considerations should be given to the vulnerable areas, where the landslides or mudslides could happen easily according to the soil loss potential.

A good plant cover is generally capable of preventing surface erosion [21]. The cover management factor with a range from 0.015 to 0.892 indicated that the loss of soil erosion can be greatly reduced by a higher vegetation cover. For the large scale soil conservation, little work can be done to reduce rainfall and runoff erosivity, soil erodibility, slope length and slope steepness, so vegetation restoration and support practice would be the way to reduce the soil loss risk.

We have estimated the error from ETM+ images to the canopy cover map and total vegetation coverage map, and calculated the interpolation error of average annual precipitation surface. However, the model is still subjected to errors due to the limitations of the methods to estimate some component factor values, and the lack of possibilities on quantitatively verifying the actual erosivity from our sites. Problems seemed to be mostly concentrated in areas with thick canopy coverage and a high measured ground cover percentage. The erosivity
discrepancies in the forests caused by the modelled rainfall influence from the tree canopies on the ground without taking properly into account the forest floors ground layer. Problematic areas also include rill and valleys where the calculated risk values seemed relatively high when compared with field experiences. This may be explained by the methods used in calculating the LS factor. On the other hand hill tops within the Arctic zone (3600-5400m) were somewhat overestimated since the ETM+ cover of these regions was highly cloud covered. The results could be improved if a cloud free fully topographically normalized image was available, and the C-factor calculation model would take into consideration under canopy ground coverage.

References


[31] GLCF (Global Land Cover Facility) - www.landcover.org as a source of:
- U.S. Geological Survey. 10th July 2002, Landsat ETM+, Scene, WRS-2 Path 130, Row 037, Orthorectified Geocover, Sioux Falls, South Dakota: USGS.
- U.S. Geological Survey. 10th July 2002, Landsat ETM+, Scene, WRS-2 Path 130, Row 038, Orthorectified Geocover, Sioux Falls, South Dakota: USGS


