Spatial Distribution of Landslide Risk and Exposure Based on CART Model and GIS Technology - Case Study of Panzhihua Basin, Yangtze River, China

Yin Long1, Runsen Zhang2 & Zhaoling Li3

1 Department of Environment Systems, Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8563, Japan
2 Center for Social and Environmental Systems Research, National Institute for Environmental Studies, Tsukuba, Japan
3 Graduate School of Life and Environmental Sciences, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

Correspondence: Yin Long, University of Tokyo, Japan. E-mail: pzhlongyin@gmail.com

Received: September 16, 2017      Accepted: November 7, 2017     Online Published: November 29, 2017
doi:10.5539/jsd.v10n6p124                  URL: https://doi.org/10.5539/jsd.v10n6p124

Abstract
Spatial distribution of natural disaster risk has been taking as a reference in urban planning and disaster management. Especially to the remote mountainous areas located in southwestern China where landslide disaster could cause enormous damage. In those areas, only few exposure researches could reveal disaster risk and exposure distribution which related to landslide occurrence since lack of landslide indicator information. This research aimed at combining landslide risk mapping with highly exposure mapping to reveal distribution of highly landslide exposure in a remote mountainous area: Panzhihua basin, Yangtze River, China. In this research, 203 landslide point data (103 occurrences and 100 absences) were collected based on governmental landslide inventory and fieldwork to build CART (Classification and Regression Tree) model, which was applied to investigate the relationship between landslide disaster and four independent variables (vegetation, elevation, slope and top soil type). Finally, we developed a landslide-prone hazard map based on transportation, threatened population and property distribution to illustrate potential socioeconomic damage of landslide occurrence. The output maps show that landslide-prone area basically follows valley stretch and Yangtze Rivers. Northeastern and center zone in study area are classified with higher risk to generate landslide disaster. Threaten population and property exposure shows similar distribution in northeastern and northwestern areas while detached in center and southeastern areas. To the transportation sector, there are 75.3km of local road, 8.6km of main road, 29.5km of railway and 39.2km of national highway are highly exposed to landslide occurrence respectively.

Keywords: CART model, limited landslide indicator information, risk exposure, remote mountainous area, GIS

1. Introduction
Landslide is one of the most frequent disasters which occurred in mountainous areas. In the up-stream basin of Yangtze River, landslides became one of the major natural hazards because of geological feature and climate characteristics that could make it sensitive to landslide disaster. Up-stream basin of Yangtze River was located in the southwestern China which is famous for mountainous terrain and monsoon climate. There are a lot of mountainous cities established along this river and Panzhihua City is one of them. Panzhihua City was built as industrial base in 1965 for its diversity of mineral resources. From that time, numerical factories were built in the river basin area for resource exploitation. As a riverside city of Yangtze River, Panzhihua city is extremely vulnerable to landslide risk. Landslide occurrences of this city have caused enormous loss of life and property damage in the past years. In 1998, due to the concentrated precipitation in rain season, landslide has caused 700 million (Unit: Yuan) loss in total. Among the loss brought by landslide, loss of destroyed transportation system is one of the serious sectors. Landslide occurrence could cut down both roads and air transportation which were regarded as lifeblood to remote mountainous areas. For example, Panzhihua Airport has attacked by a serious landslide in 2011.7.15th that this airport was forced to suspend until 2013.6.28th which brought various inconveniences and incurred numeric loss to the whole region. Therefore, landslide is considered a serious disaster which needs to be put on spot point of natural disaster management of study area.
To mitigate the negative effects of landslide disaster, proper disaster management was needed in this area. The first step of disaster management is target disaster risk assessment. Landslide disaster assessment composes both risk evaluation itself and its influence on socioeconomic sector. Landslide susceptibility is the first step toward regional landslide management. Risk analysis has been conducted in many disaster risk studies. However, only very few studies produce risk maps, and lack of temporal landslides data is one of the reasons (Vranken et al. 2015). Here, we applied CART (Classification and Regression Tree) to separate potential landslide-prone areas from others to generate landslide risk map and overlap it with socioeconomic distribution in GIS. Then, we could get spatial distribution of landslide exposure of studies area. In the risk calculation sector of previous studies, both deterministic and statistical methods have been utilized in natural landslide risk assessment. Deterministic quantitative methods focus on small scale areas since they need exhaustive data from individual slopes which depending on engineering principles of slope instability expressed in terms of the safety factor (Lulseged Ayalew et al. 2005). However, although deterministic methods have been proved as high accuracy of landslide risk, it cannot reveal regional landslide risk distribution in large scale. On the other side, statistical analyses are wildly applied in landslide sensibility studies, too. This method could link landslide environment indicators to landslide inventory database and generate regional landslide susceptibility map in large scale. Among previous statistical analyses, generalized additive models (Jason N et al.,2011, Park et al. 2008), logistic regression (Saro Lee et al. 2006,2007; Lulseged Ayalew et al. 2005, Gregory C et al. 2003 and Bai et al. 2010) and analytical hierarchic process (Marko Komac et al. 2006; HR Pourghasemi et al. 2012; P. Kayastha et al. 2013; Yoshimatsu et al.,2006; and Sujit Mondal et al. 2012) are representative statistical method which have been applied in landslide risk assessment. However, as a statistical method, CART has not been widely used although it is a new technique for landslide research (Saito, et al. 2009). CART is ideally suited for nonlinear relationships especially in fields of ecology and geology since it can deal with both continuous and discrete data. GLENN et al. (2000) demonstrated the flexibility of CART since it can handle broad range of data including missing values in both response and explanatory variables. Ángel, et al. (2012), Yeon et al. (2010) applied classification tree technique to assess landslide susceptibility and demonstrate this method could be used efficiently for landslide susceptibility analysis and might be used for prediction of various spatial events. Also, CART could distinguish landslide-prone areas from other areas which could have a clear and simple visualization to be the reference to policy makers. On the top of natural risk assessment, integrated risk distribution is still lack of socioeconomic exposure which could provide policy guidance to landslide risk management. The combination of natural disaster risk and exposure is essential to local landslide risk management.

Integrated landslide risk evaluation is used to describe the landslide-related potential adverse effects and total expected economic loss in a region (Chaofeng Liu et al. 2015). To the influence upon socioeconomic sector, exposure can be used to describe the size of local properties and populations which are subject to loss by landslide directly and indirectly. In 2012, the IPCC report gave an explicate explanation of disaster and risk management which emphasized the exposure analysis in disaster; disaster was defined as a hazardous physical event which has an influence on life, property and other environmental factors (IPCC SERX 2012). Socioeconomic effect distribution is in need to illustrate correlation between natural disaster risk and regional exposure. Exposure quantitative analysis needs large amount of information to support it and it is extremely difficult to remount mountainous area specially to developing countries. Therefore, generation of a clear distribution of landslide exposure is extremely needed for current situation.

In this research, we used CART model to reveal the relationship between landslide and environmental indicators to generated landslide-prone distribution map. Then, we extended the landslide-prone distribution to socioeconomic sector for revealing population and property exposure distribution.

The main objectives are to illustrate geographic distribution of landslide risk in study area based on CART model and to only visualize highly exposure distribution against landslide risk of Panzhihua basin, Yangtze River, China. Based on the conclusion, this study aims at providing spatial distribution information to local policy maker which will benefit landslide risk mitigation policy, urban planning and infrastructure construction in a clear way.
1.1 Study Area

![Map of study area](image)

Figure 1. Location and administrative division of study area

Study area incorporates Yangtze basin area located in Panzhihua City. Firstly, Panzhihua city (108°08′E - 102°15′E and 26°05′N - 27°21′N) is a prefecture-level city of Sichuan Province in southwest of China, at the convergence of Sichuan and Yunnan, where is 749 km from Chengdu in the north and 351 km from Kunming in the south. This area is characterized by long summer, abrupt daily temperature change, indistinct seasons, dry air, concentrated rainfall, rich sunshine, active evaporation and diversified microclimates. Since of monsoon climate, concentrated rainy season usually only lasts from early June to October; while dry season lasts from November to May. Those natural characteristics result in a high frequency of precipitation-induced landslide and caused great economical lose every year. Basin area along Yangtze River which located in Panzhihua city is called Panzhihua basin of Yangtze River In this study, we selected Panzhihua basin of Yangtze River as study area (Fig.1). The main reason of study area selection is that study area gathered most of population and urban infrastructures in Panzhihua City, since riverside area offers the convenient transportation even it also indicates higher vulnerability to landslide in this area.

1.2 Research Flow

Satellite image, fieldwork and government landslide inventory were feed to the CART Modeling. The assessment was approached from three aspects: transportation system, property distribution and population. Fig. 2 presented the methodological flowchart of this study. The processes comprise three main parts: 1) Spatial variable dataset, 2) CART Modeling and Validation, and 3) Exposure visualization.
1.3 Dataset

As a topographical disaster, landslide formation can be caused by different variables that determine landslide hazard. In previous studies, it can be grouped into two categories: 1) intrinsic variables that contribute to landslide susceptibility, such as geology, slope gradient, elevation, geotechnical properties of the slope-forming material, vegetation cover, and drainage patterns; 2) extrinsic variables that tend to trigger landslide in an area of given susceptibility, such as heavy rainfall, earthquakes, and volcanoes (Wu et al. 1995), since limited information of mountainous areas, environmental indicators of landslide risk was assessed by 4 spatial variables; vegetation, slope, elevation, and soil type. Data source was explained in Table 1.

Table 1. Data source and format explanation

<table>
<thead>
<tr>
<th>Category</th>
<th>Environment Predictor</th>
<th>Unit</th>
<th>Platform</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography</td>
<td>Altitude</td>
<td>m</td>
<td>ArcGIS.</td>
<td>SRTM.NASA</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10.0</td>
<td>Resolution:30m</td>
</tr>
<tr>
<td>Topography</td>
<td>Slope</td>
<td>°</td>
<td>ArcGIS.</td>
<td>Resolution:30m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10.0</td>
<td></td>
</tr>
<tr>
<td>Vegetation</td>
<td>NDVI</td>
<td>--</td>
<td>ENVI</td>
<td>ALOS data</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.8</td>
<td>PRISM &amp; AVNIR-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Resolution: 10m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ScS_SceneID=&quot;ALAV2A269143070&quot;</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Observation Date=20110212</td>
</tr>
<tr>
<td>Top soil</td>
<td>Soil Type</td>
<td>--</td>
<td>ArcGIS.</td>
<td>Public Database from East District Government of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10.0</td>
<td>Panzhihua City</td>
</tr>
</tbody>
</table>

1) **Independent variables: spatial data**

Topographical data were derived from the SRTM 90m Digital Elevation Database version 4.1 (NASA) with a resolution of 90 m by 90 m. The elevation layer was corrected with a spatial reference from the Beijing Geodetic Coordinate System 1954. The elevation range of the study area was 936 meters to 3,146 meters.
Gradient ranges from 0 degree to 65.04 degree. Both gradient variable and elevation variable were derived from this dataset.

Top soil type could be the key point in precipitation-induced landslide disaster. In study area, red soil is the most widely distributed. Based on the hydrothermal conditions and soil erosion, red soil can be described as two sub-categories; 1): mountain soil, mainly distributed in the 1300m to 1800m with the color of reddish brown. 2): brown-red soil, mainly distributed between 1700m to 2200m. Lateritic red soil distributed among the range of 937m to 1400m. Based on the soil data provided from Public info-net of Panzhihua, top-soil was understood by a correlation with elevation variance.

This research applied NDVI (Normalized Difference Vegetation Index) to describe vegetation variable of study area. The NDVI itself thus varies -1.0 to 1.0. Red and NIR stands for the spectral reflectance measurements acquired in the visible (red) and near-infrared band respectively. When applied to an ALOS (Advanced Land Observing Satellite) image, band 4 and band 3 corresponded to NIR and Red band respectively.

The formula is given by:

\[
NDVI = \frac{(NIR - Red)}{(NIR + Red)}
\]

Satellite image of study area was clipped by the scene of an ALOS image with resolution of 10 m by 10 m. This image was acquired by the sensor of AVINIR-2 on Feb. 12th 2011 in GeoTIFF format. The image projection was the UTM and UTM zone was 47N.

2) Dependent variables: landslide data

In fieldwork, we selected two main routes to conduct route census for landslide point data. The directions of selected routes were west-east and north-south respectively. Elevation ranges from 983 m to 2,220 m. Based on fieldwork observation, 105 points (100 no-landslide data and 5 landslide data) were collected in Aug. 2014.

<table>
<thead>
<tr>
<th></th>
<th>Model</th>
<th>Verification</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Landslide</td>
<td>No-landslide</td>
<td>Landslide</td>
</tr>
<tr>
<td>Government</td>
<td>50</td>
<td>-</td>
<td>48</td>
</tr>
<tr>
<td>Fieldwork</td>
<td>-</td>
<td>50</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>53</td>
</tr>
</tbody>
</table>

Local government conducts a geological expedition project every year to illustrate landslide disaster and this research cited the 2011 landslide hazard data. Combining the fieldwork data with government data, landslide point data contained 103 points of landslide data and 100 points of no-landslide data totally which were divided in model and verification database separately.

3) Dataset compilation

The preparation of the spatial independent dataset and landslide point dataset laid the foundation for CART modeling of landslide risk distribution. Point data were divided into a landslide and no-landslide point datasets. We extracted values from spatial variables using geographic locations of all the point data through ArcGIS® and converted the attribute table to a CSV file.

Although the principal trigger factor of landslide here is precipitation, environmental indicator database does not include precipitation since rain volume shows stable and uniform throughout study area. Therefore, precipitation parameter has been eliminated out since its unapparent variance. Hence, landslide environment indicator database was compassed with: top soil, elevation, slope gradient and vegetation (NDVI)

1.4 CART Modeling

Classification and Regression Tree (CART) (Breiman et al., 1984) is a binary recursive partitioning method yielding a class of models called tree-based models (Qian 2009). Classification and regression trees are ideally suited for the analysis of complex ecological data (Glenn et al., 2000). Moreover, CART fits many exploratory studies because it can address continuous and discrete data. In CART, each split is partitioned into two subsets based on one rule. A rule was generated on each split based on the minimum class impurity. The recursive
partitioning rule can be described on a conceptual level as a process of reducing the measure of impurity (Breiman et al., 1984).

A splitting procedure is applied on each split separately. It reaches its minimum (Zero) when all cases in the node fall into a terminal node.

\[
GINI(S) = 1 - \sum_{j=1}^{n} P_j^2
\]  

(2)

The Gini Impurity Index \( GINI(S) \) is given by

where \( P_j \) is the relative frequency of class \( j \) in set \( S \), and \( n \) is the class number of examples in set \( S \). After splitting set \( S \) into two subsets \( N_1 \) and \( N_2 \), the Gini Index of split data is defined as

\[
GINI_{split}(S) = \frac{N_1}{N} GINI(S_1) + \frac{N_2}{N} GINI(S_2)
\]  

(3)

where the smallest \( GINI_{split}(S) \) is adopted to split \( N_1 \) and \( N_2 \), and \( N = N_1 + N_2 \).

Elevation (DEM data), slope, soil type, and Normalized Difference Vegetation Index (NDVI) were used to model a landslide classification tree.

The default first CART model, showing a complexity parameter of 0.001, resulted in an overcomplicated model due to overfitting. Due to the fitting noise, from 20 split xerror values became larger (Fig. 3.). Therefore, in this research, a CP-value of 0.010 was adopted as a threshold value to prune the model. The pruned tree had 7 splits and 8 terminal nodes (Fig. 4.). A terminal node can be read by Y (Yes: Landslide) and N (No: No-landslide). The CART model was created by the programming language R version 3.2.0. We used rpart package to conduct CART modeling.

![Figure 3. Plot of Cross-validation error against CP-values](image)

129
1.5 Landslide Risk Distribution and Its Exposure

Based on classification rules created by R language, landslide risk distribution map of the study area was generated through the ArcGIS® Model Builder. We generated risk exposure distribution maps by clipping transportation, expected properties loss and threatened population distribution out through landslide-prone vector file.

Transportation data (road network and railways) was derived from National Geomatics Center of China⁵. Threaten population and properties were derived from government landslide point database. Based on points data of threaten population and expected properties loss, we applied IDW (Inverse Distance Weighting) interpolation to generate threaten population and expected properties loss of whole study area and clip the landslide-prone area out respectively.

2. Results

Table 3. Landslide-prone area statistics

<table>
<thead>
<tr>
<th></th>
<th>Study Area (Unit: Km²)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Area</td>
<td>3282</td>
<td>100%</td>
</tr>
<tr>
<td>Landslide Area</td>
<td>564.752</td>
<td>17.208%</td>
</tr>
<tr>
<td>No-landslide Area</td>
<td>2717.248</td>
<td>82.792%</td>
</tr>
</tbody>
</table>

Based on CART model, landslide-prone distribution map was generated (Fig.5.). Table 3 shows the ratio of landslide-prone area and no-landslide area separately. Landslide-prone area is up to 564.752 Km² and counts for 17.2% of study area. The remaining 82.792% area is no-landslide area which indicated less vulnerability to landslide. The validation dataset incorporated reminding 50 no-landslide point and 48 landslide points. Accuracies of the validation were 70.4% for no-landslide and 64.8% for landslide.

Finally, we developed a landslide risk map (Fig. 5). Analysis of landslide risk distribution lays a foundation for landslide risk management. This landslide risk distribution map can be used for future land use and infrastructure planning. On the other hand, this landslide risk distribution map can be used as a reference for rescue efforts when landslides occur.

Table 4. Landslide-prone area and transportation exposure (Unit: Km)

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Landslide</th>
<th>No-landslide</th>
<th>LS-Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rail_way</td>
<td>133.439</td>
<td>29.471</td>
<td>103.968</td>
<td>0.220858</td>
</tr>
<tr>
<td>National_highway</td>
<td>161.796</td>
<td>39.270</td>
<td>122.526</td>
<td>0.242713</td>
</tr>
<tr>
<td>Local_road</td>
<td>299.983</td>
<td>75.354</td>
<td>224.629</td>
<td>0.251194</td>
</tr>
<tr>
<td>Main_road</td>
<td>287.949</td>
<td>18.246</td>
<td>269.703</td>
<td>0.063365</td>
</tr>
</tbody>
</table>
The correlation between landslide-prone distribution and elevation variation could be observed from the Fig. 6. According to the exposure map, it could be understood that landslide-prone area basically followed the direction of streams and valleys since water movement strike and significant altitude difference could waken slope stability. Steep gradient increased susceptibility of mountain because of gravity. As a mountainous city, it follows narrow urban layout because terrain characteristic provides convenient transportation and drinking water. Therefore, valley and basin accumulate dense population and property, which makes those areas more sensitive to landslide disaster.

Road is the lifeline of remote mountainous area. When landslides occur, the road is sensitive to the landslide effect which was shown as Fig. 7. To the mountainous area, any damage to the road network is anticipated to be an obstacle to access of escape routes. In study area, transportation system mainly divided

Landslide occurrence finally affects residents and properties. In the exposure maps of Fig. 9 and Fig. 10, it can be observed that not only the center part of the central area but also Miyi County in the north is highly exposed to landslides as well. In recent years, urbanization in Panzhihua City has attracted more and more residents to settle down in Miyi County. With development, the landslide risk is expected to increase in a synchronized way. Based on the landslide risk and exposure analysis, the current urban design and population distribution in Panzhihua City are found vulnerable to landslide. Miyi County became a new area of high risk exposure because of urbanization and vegetation deterioration from new construction.

Figure 5. Landslide-prone area based on CART model
Figure 6. Landslide-prone area and elevation distribution

Figure 7. Landslide-prone area with transportation distribution
Figure 8. Transportation exposure distribution to landslide-prone area

Figure 9. Property exposure to landslide-prone area
The output shows that property exposure is different from population exposure distribution. Property exposure distributed densely along the center part of Panzhihua basin. However, population was distributed along the southeastern zone of study area. The reason could be understood from historical development of Panzhihua city. This city was built as an industrial base for China in 60s last century so that industrial parks gathered in the center area and distributed along the Yangtze basin. After the development of iron and steel industry, population also accumulated in high speed. Finally, southeastern area of study area became to mature living quarters in recent decades and was settle down across the river from industrial areas.

4. Discussion

In this research, potential landslide distribution was computed using the statistical model which is suit for remote mountainous area with limited environment indicator information. This part is the platform for the assessment of potential negative impact on transportation, population and property from landslide hazard. The output could be applied in optimization resource allocation and increase the efficiency of landslide risk management. Local government sends geological expedition team every year to illustrate landslide disaster. This is a tough work for human force and need large amount of financial aid. Landslide exposure map could contribute to local government for narrowing expedition range and emphasizing key zones which will increase efficiency and accuracy of expedition.

In general, Yangtze River attracts many researches which concerning landslide formation and risk assessment. In previous researches, Zhihua (1999) considered that landslides and debris flows have caused serious damages to Jinsha Basin (Part of Yangtze River basin) and strongly impacted the environment of Panzhihua City. Landslides actually became a big problem since Panzhihua city is developing to the second biggest city with large population in Sichuan province. Landslide hazard map which generated by certain factors has identified Xiaojiang Basin of Jinsha River as a high risk area (Lan et al. 2014). This area is located not far away from the study area in this research. In the Panzhihua city, the landslide which occurred along connecting road in Miyi County has been analyzed in the method of landslide volume and drainage tunnel analysis (Wu et al. 2006) which focus on deterministic quantitative methods. And another field survey was conducted in the airport landslide disaster (East District, Panzhihua City) adopted emergency monitoring and warning models to demonstrate high landslide risk of Panzhihua Airport (Wang et al. 2013). And in this research, we conducted landslide risk and exposure assessment separately and narrows range of landslide risk distribution by local
exposure distribution. Landslide risk and exposure distribution has general directive significance to landslide management. In this research, we found that northeastern areas are highly exposed to landslide risk both on property and; southeastern area is highly exposed to landslide risk on population and central area is highly exposed on properties. Highly property exposure gathered along center part since industrial factories distributed densely and highly population exposure was located on the other side of river. Development process of Panzhihua City takes account for this phenomenon. Panzhihua city was built as an industrial base last century. Most of factories are gathered along the central areas. Then residential areas are located across the river. Also, because of recent urbanization, northeastern area (Miyi County) is highly exposure area both on property and population since new construction and exploitation in recent year.

The integration of landslide risk and exposure to disaster will be essential in adapting landslide disaster. Improving the utilization of landslide exposure map is key ingredients in effective disaster management and rescue plan after landslide. Exposure distribution describes disaster visualization from natural hazard risk to socioeconomic effect. Landslide management needs to review natural landslide risk distribution but also local transportation, population and property distribution. Therefore, it is advisable that landslide risk management considers both landslide risk and exposure distribution from social aspect. Thus, landslide management will enhance the sustainability and resilience to natural disaster especially in remote mountainous area.

References


**Notes**


**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).