Mapping landslide release area using Random Forest Model

M R Darminto ¹, and Hone-Jay Chu ²

¹ International Master Program on Natural Hazards Mitigation and Management, National Cheng Kung University, 1 University Road, Tainan, Taiwan
² Department of Geomatics, National Cheng Kung University, 1 University Road, 701 Tainan, Taiwan

Abstract. Landslides pose threats not only to infrastructure around the world but also to local communities. One particularly susceptible area in Taiwan is in the Zhoukou River basin, Kaoping watershed. Landslides source area plays an important role in landslide occurrence, where the triggering stage initiates the failure. The conditions of landslide source area are assumed to be the same in the future. This study aimed to produce a Random Forest model that accurately predicts future landslide release in this area by validating the predictions against those observed landslide releases in this region. The landslide data were recorded in the year of 2010, a year after typhoon Morakot stroked Taiwan in 2009, triggering huge number of landslides all over the country. This study proposed the new concept to separate landslides area into release as its original source and focuses on using the topographical factors derived from Digital Elevation Model (DEM) as the independent variable in predicting landslides occurrence, including Slope, Aspect, Curvature, Topographic Wetness Index, Average Slope and Distance from the river, and an additional geological map of the study area.

An observed landslide release occurrence layer posed as the dependent variable classifier in the model. First, data sampling strategies applied show an optimal model to be created with the highest Area Under Curve (AUC) value of 0.814. Next, this model identified the most influential factors causing landslides. Aspect, were determined as being most influential factor, where Distance from river, and Slope as second and third most influential.

The concept of release area separation showed a better AUC value model compared to the model using conventional full landslide inventory. The random forest model also showed a reliable result when compared to logistic regression and decision tree using the same data sampling, with the AUC value of 0.814, 0.65, and 0.728 respectively. The results have proven that random forest model is suitable for producing landslide release susceptibility map.

1. Introduction
Mass movements are the movement of certain surface material in an area that is caused by gravity. There are many types example to this very sudden movement such as; landslides, rock falls, debris flows, snow avalanches. The general term of all the mass movements is called landslides. Within the present article the term “landslide” is used in a broad sense, including all relevant types of gravitational mass movements.

Landslides are considered around the world as one of the most disastrous and phenomenal natural hazards [1]. This kind of mass movement occurs mainly because of specific geological formation, steep and rugged land surfaces and extreme climate conditions result in a high degree of instability. Dynamics of hydrological processes and certain elevation patterns due to sharp changes in altitude lead to
substantial differences in environmental characteristics in mountainous area. Additionally, human activities like road construction or deforestation can contribute to this hazard [2][3].

To mitigate landslide impact, it is necessary to assess and manage areas that are susceptible to them. Hence, in recent years, the assessment of landslide hazard and risk assessment has become a major topic of interest. Predicting potential landslides has been a challenge to scientists in the past due to the complex combination of factors, such as physical attributes and climatic conditions. Potential landslide prediction is defined as the propensity of an area to generate landslides with susceptibility represented by assuming that “landslides are most likely to occur in conditions similar to those that have caused past failures”. This assumption will be the underlying basis for generating the landslide susceptibility map [3]. Previous studies produced landslide susceptibility maps to show areas where landslides are more likely to occur according to environmental physical factors. These studies employed various statistical methods; such as logistic regression (LR), decision tree, random forest, etc, to calculate the landslide susceptibility in these areas.

A study by [4] in Wadi Tayyah Basin, Asir Region, Saudi Arabia, analyzed 4 different data mining models to produce landslide susceptibility maps, namely; random forest (RF), boosted regression tree (BRT), classification and regression tree (CART), and general linear (GLM). 11 landslides factors were used, including slope aspect, altitude, distance from faults, lithology, plan curvature, profile curvature, rainfall, distance from streams, distance from roads, slope angle, and land use, as the independent variables. The results were than compared for landslides susceptibility mapping.

Another study by [5] in Three Gorges Reservoir, China, employed random forest and decision tree methods to generate landslide susceptibility map. 34 predictor variables in total (topographic, geological, hydrological, land cover) were used. The results concluded a higher prediction produced by random forest compared to decision tree in generating landslide susceptibility map.

Many studies explained above, offered various concepts of developing a model to produce landslide susceptibility map. The study of [4] and [5] concluded the capability of random forest to generate a reliable result in predicting future landslides and its ability to determine the important factors. However, the studies only considered the occurrence of landslide disregarding its release (source) area.

According to [6], a systems divided a landslide into failure, post failure and propagation stages based on their distinct kinematic characteristics. Failure and post-failure stages occur inside the landslide source areas, while the propagation stage includes the movement of failed soil mass from the source area to the deposition area where the soil mass stops. During the failure stage inside the source deposits, the increase of pore water pressure can be generated by rainfall that directly infiltrate the slope surface [7], mostly from the high (source) area and propagate in depth through groundwater flow patterns related to the slope setting [8]. A previous study proposing a concept of handling the landslide inventory was done by [9]. In this case the landslides inventory was separated into the release and deposition by using elevation. Release area is considered as the starting point (source) of any landslide propagation until it stops at some deposition area. The separation of the releases and depositions area was based on the elevation of the pixels.

Besides rainfall and pore water pressure, seismic activities (earthquakes) have been long recognized as important factors causing instability of landslide source areas, especially in mountainous area. These triggering factors initiates the potential threat by interfering the material strength, gravitational stress, external forces due to its shaking, and generate a landslide source area as a result [10]. In Taiwan, landslide hazard reaches high levels in mountainous part of the island (such as the present study area) where argillaceous slate lithology, earthquakes, typhoons and heavy rainfall facilitate to favor source area, and followed by landslide [11]. A deep analysis of landslide release (source) area can lead to understanding the triggering mechanism of landslides and to contribute the assessment of hazard [12]. Thus, mapping the topographical and geological characteristics of landslide source area during past events is a good estimation of the potential landslide source likely to impact a specific site in the future [13].

The present study proposed a new approach to consider landslide release area – as a source of landslides occurrence – separated from the transport area and the random forest method to accurately produce landslide release susceptibility map by employing 7 dem-derived topographical factors and
lithology as landslide factors. The result is capable to show areas where potential release areas of landslides are more likely to be generated in the future, and more importantly it can provide valuable information for future urban planning and mitigation management and prevention.

2. Study area and datasets
Zhoukou River basin was chosen as the study area because (1) it is an area in which large number of landslides events occurred, (2) sufficient landslides inventory data is available, and (3) large numbers of landslides occurred all over the area. The headwaters of Kaoping River emerge from the Jade Mountain Range and flow north to south through a series of small, steep sloping basins. The Kaoping River consists of five main tributaries: the Chishan, Laonong, Ailiao, Bulao and Zhoukou Rivers. The Zhoukou River is located on south east part of Kaoping River basin, in Laonong upstream (Figure 1). The boundary of the study area is defined as 22°50’ to 23°00’ north latitude and 120°39’ to 120° 52’ east longitude and covers an area of 242 km2. This covers an elevation ranging from 143 to 2773 meters and an average slope of 42°.

The landslides inventory map used in this study was derived from Aerial Survey Office of the Forest Bureau which annually produced landslide maps by using a semi-automatic approach that delineates the landslide areas with 2 m resolution Formosat-2 imagery. The landslide data were recorded in the year of 2010, a year after typhoon Morakot stroked Taiwan in 2009, triggering huge number of landslides all over entire country. Since the original data of the landslides inventory contain some polygons that separated one landslide into one or more polygons. Corrections were made to merge polygons that represent single landslide in the study area.

The general geological setting of the area is shown in Figure 4d, and the lithological properties are summarized in Table 1. Zhoukou River flows from northeast and southeast to the southwest direction, and crosses several sedimentary and metamorphic formations with different geological ages.

![Figure 1. Study Area of Zhoukou River Basin.](image)
Due to vibrant tectonic activities, series of imbricated structures; such as faults, were formed in northeast – south west direction separating Chaochou formation into two parts, and smaller faults from north to south direction separating the Chaochou formation from Pilushan formation.

### Table 1. Geology formation of study area.

| Class | Code | Formation       | Lithology                                      | Geological Age |
|-------|------|-----------------|------------------------------------------------|----------------|
| 1     | 1510 | Chaochou Formation | Argillite, slate intercalated with sandstone lentils. | Middle Miocene |
| 2     | 1630 | Terrace Gravel  | Mud, sand and gravel                           | Pleistocene - Holocene |
| 3     | 6020 | Pilushan Formation | Slate with meta-sandstone with igneous rock. | Eocene          |
| 4     | 6060 | Alluvium        | Gravel, sand and clay                          | Holocene        |

### 3. Dataset and Methodology

#### 3.1 Landslide release susceptibility mapping

The whole concept of landslide release susceptibility mapping is explained in Figure 2. Each of the data employed is transformed from spatial to tabular, and then used for training and testing in the model. The next step is to choose subsets of the training data, and then to employ the random forest in R software. The output of the model is analyzed and validated by using the testing data afterwards, to determine whether the model is reliable. If the result of the model is considered in poor accuracy, different training dataset was chosen. The step will be repeated until a model with a good accuracy is produced (Figure 2). Then, the last step is to find the potential vulnerable areas based on the best accuracy result.

![Figure 2. The flowchart of this research process.](image-url)
In order to analyze the landslide release susceptibility, the random forest model was implemented in R software. Out of 8,758,211 total points in the study area, 1,356 were used as the training points, while the rest of them as the testing points. All of the maps in the model were set in 5 x 5 meter grid format. The 1,356 training points were distributed in 150 landslide releases from 460 landslide releases in total of the study area. Four data instances were originally applied to generate the model accuracy in order to gain the best accuracy results. The first instance was to set the combination of landslide and no landslide data to each 50% equally, and the second and third combination were applied to combination of landslide and no landslide to 70% - 30% and 80% - 20%, and the fourth combination was set to 90% landslide and 10% no landslide. The accuracy results of the different combinations are explained in chapter 4.

After generating the model with the training data, validation was performed by calculating the Area Under Curve (AUC) of the receiver operating curves (ROC). This method has been widely used as a measure of performance of a predictive rule [14]. According to [15] and [16], correlation of predictive capability and AUC could be quantified as follows: excellent (0.9–1), very good (0.8–0.9), good (0.7–0.8), average (0.6–0.7), and poor (0.5–0.6).

In summary, the landslide release susceptibility map was performed as follows: 1) landslide inventory was separated into release and deposition area, 2) topographical data were collected and extracted for modeling process, 3) landslide susceptibility assessment was conducted by using random forest model, 4) validation of the susceptibility map is performed with testing data for prediction rate.

3.2 Landslide factors
The previous studies offer a variety of landslide factors to choose in order to perform the analysis in this study. However, the datasets used in this study - slope, elevation, geological map, curvature, wetness index, aspect, distance from the river, and average slope from the river - seem to impact the results. The factors explained above are the independent variables, while the dependent variable is whether a landslide release has or has not occurred in the given area.

The digital elevation model used in this study was obtained from the Ministry of the Interior of Taiwan with resolution of 5x5 meter. The digital elevation model was processed to eliminate the no data value contained in the original data, which then directly used as the elevation factors afterwards.

Topographical factors, including slope, curvature, wetness index, aspect, distance from the river, and average slope from the river were all derived from the digital elevation model, by using ArcMap 10.1. The river in the distance from river factors was determined from wetness index by setting the threshold of more than 17.000 to be defined as the river and validated using satellite image.

The rainfall data was not employed in this study since the whole study area was assumed to have the identical precipitation variation along the similar topographical conditions, while the distance from fault to identify the earthquake effects was not employed either due to the absence of major faults in the study area which usually cause the earthquake in the whole area of Taiwan.

A new approach of measuring the slope was proposed in this study. The slope angle was measured from the nearest river cells instead of the 8 neighborhood cells that are usually employed in measuring the slope degree. In further explanations, the term “Angle” in this study refers to this new approach of slope angle. The geological map was obtained from Taiwan Central Geological Survey, Ministry of Economic Affairs at the scale of 1:50,000. All the map of the factors is shown in Figure 3.
Figure 3. All eight landslide factors: a) digital elevation model, b) distance from river, c) slope from river, d) geological map, e) slope, f) aspect, g) wetness index, h) curvature.

To summarize the correlation of the factors and landslide release occurrence, Table 2 shows the details of the landslides occurrence distribution in each class of each factors previously described.

Table 2. shows the details of the landslides occurrence distribution in each class of each factors

| Factors | Class | Landslides Release |
|---------|-------|--------------------|
| Aspect  | Flat  | 0 %                |
|         | North | 1.67 %             |
|         | Northeast | 8.73 %        |
|         | East  | 19.3 %             |
| Slope          | Percentage |
|---------------|------------|
| Southeast     | 25.6 %     |
| South         | 23.9 %     |
| Southwest     | 16.4 %     |
| West          | 3.7 %      |
| Northwest     | 0.7 %      |
| < 15°         | 0 %        |
| 15° – 30°     | 22.3 %     |
| 30° – 45°     | 67.2 %     |
| 45° – 60°     | 9.3 %      |
| 60° – 75°     | 0.7 %      |
| 75° – 90°     | 0 %        |

| Distance from river | Percentage |
|---------------------|------------|
| < 500               | 20 %       |
| 500 – 1000          | 37.3 %     |
| 1000 - 1500         | 25 %       |
| 1500 – 2000         | 12.7 %     |
| 2000 – 2500         | 4.6 %      |
| 2500 - 3000         | 0 %        |
| < 15°               | 4.9 %      |
| 15° – 30°           | 53.3 %     |
| 30° – 45°           | 40 %       |
| 45° – 60°           | 1.5 %      |
| 60° – 75°           | 0.1 %      |
| 75° – 90°           | 0 %        |
| < 200               | 0 %        |
| 200 – 400           | 0 %        |
| 400 – 600           | 4.9 %      |
| 600 – 800           | 8.6 %      |
| 800 – 1000          | 12.2 %     |
| 1000 – 1200         | 10.9 %     |

| Geological Map      | Percentage |
|---------------------|------------|
| Pilushan Formation  | 15.8 %     |
| Chaochou Formation  | 84.2 %     |
| Alluvium            | 0 %        |
| Terrace Gravel      | 0 %        |
| < 500               | 20 %       |
| 500 – 1000          | 37.3 %     |
| 1000 - 1500         | 25 %       |
| 1500 – 2000         | 12.7 %     |
| 2000 – 2500         | 4.6 %      |
| 2500 - 3000         | 0 %        |
| < 15°               | 4.9 %      |
| 15° – 30°           | 53.3 %     |
| 30° – 45°           | 40 %       |
| 45° – 60°           | 1.5 %      |
| 60° – 75°           | 0.1 %      |
| 75° – 90°           | 0 %        |
| < 200               | 0 %        |
| 200 – 400           | 0 %        |
| 400 – 600           | 4.9 %      |
| 600 – 800           | 8.6 %      |
| 800 – 1000          | 12.2 %     |
| 1000 – 1200         | 10.9 %     |
3.3 Landslide release area separation

The landslide inventory map handling is important since this modeling approach is based on the assumption that past landslides are the key to the future. Therefore, landslide inventory map with high accuracy and detail information are the first important step to deal with.

The preprocessing of landslide inventories often suffers from a missing or unsatisfactory sub setting into release, transit and deposition areas. The reason for this problem is not necessarily related to deficient mapping efforts, but rather to the impossibility to identify each zone in the field or from remotely sensed data. Appropriate sub setting, however, is required before using the inventory for statistical analyses of landslide release [9]. In this study, the release part of the landslide area – where a release area is considered as the initial start of the propagation and stop at the deposition area – will be used as the dependent variable. A reproducible procedure to separate the release and deposition area from other part of given landslide area is explained in Figure 4.
In the present study, an assumption that a release area is 1/3 of the total area in each landslide is suggested. All observed landslide pixels within the area of 1/3 from the top of maximum elevation of each landslide is set as release pixels.

3.4 Random forest
Random Forest is a machine learning model that is able to generate a large number of decision trees to interpret the spatial relationships in landslide occurrences. This technique works by producing many trees during the training time, and generate classes of the classification or regression individual trees [17]. For the regression algorithm, the dependent variable estimation is obtained by employing the result averages, while for the classification algorithm, the decision trees is built to output the class. To explain the relationships between the dependent variables and independent variables, random forest model does not require any initial assumptions. Hence, it is a proper way to investigate hierarchical interactions and non-linear in large datasets [18].

Among other models such as decision tree, support vector machine, logistic regression, which mostly used to generate landslide susceptibility map, the random forest technique has only recently been applied [19]. The main advantages of the random forest, that it can rank the factors by importance and addresses high dimensional datasets. This technique also uses unbiased estimation in building the model where the training process is fast and simple to be implemented [20]. The Random Forest algorithm data does not need to be rescaled, transformed, or modified. It has resistance to outliers in predictors, and can handle the missing values automatically [17]. Another benefit of random forest is its resistance to over training and growing a large number of random forest trees where it does not create a risk of over fitting, where each tree is a completely independent random experiment [4].

4. Results
Following the mechanism to separate the landslide release area from the whole landslide inventory previously explained in Chapter 3c, the landslide release map is presented in Figure 5. The landslide release map consists of the data (value 1) landslide release, and (value 0) no-landslide release. The given landslide release map is further employed in the model as the dependent factors in producing the landslide release susceptibility map.
The main parameters of random forest model are mtry and ntree. \textit{mtry} represents square root of factor numbers, while \textit{ntree} is the number required by the user to be constructed by the model. Generally, the optimal parameters are selected by the model based on the highest accuracy. To determine the best value of the parameters, the range of \textit{ntree} values was set from 10 to 2000, with random interval where the best value was compared based on its effect on the AUC value. The results are shown in figure 6.

The highest accuracy based on the results presented in figure 6, are values of 500 trees and 2000 trees. The value of 500 was selected in all further modelling process due to its high accuracy and shorter computation time, compared to value of 2000 tree which shows near identical accuracy with longer computation time.

The relationships among the factors were complicated, and the factors themselves were variously affected the landslide releases (Hong et al., 2017). To define which factors possess the largest impact on the landslide releases occurrence, the importance of the factors in the model was analyzed (Figure 7).

As previously explained in the chapter 3, four different data instances of landslide and no landslide were employed to gain the best accuracy result. The first combination of 50\% equal landslide and no landslide showed fair accuracy, although an improvement was necessary. The second combinations with 70\% - 30\% showed an improvement of accuracy, where the accuracy continue increasing as the third combination was employed with 80\% landslide and 20\% no landslide. However, when the last combination of 90\% - 10\% was employed, the accuracy dropped. To maintain the best accuracy for all the process in the models, the third combination was employed for all the further modeling in this study.

In addition: two concepts of landslide releases and landslide full inventory were both modeled by using random forest to analyse the differences, models with different data resolution were employed to be compared one another, logistic regression and decision tree model were also constructed with the identical datasets to compare the landslide susceptibility map results from each model.
4.1. Comparison of landslide release and landslide inventory

Many studies employing various statistical models to landslides occurrence prediction commonly treat the historical landslides record without regarding any separation in release and deposition as proposed in this study. To analyse the different, a comparison between the two given concepts was applied.

The training data for landslide full inventory was made by employing 4000 points distributed on the same 150 landslides as previously employed for the landslide release concept. More points were applied in the training data, due to more landslides area needed to be covered in the landslides full inventory. The identical 8 factors were also used as the independent factors where the landslides inventory as the dependent factors using the random forest model.

The landslide release susceptibility map and landslide susceptibility map are presented in Figure 8b and 8a. The susceptibility maps were reclassified into three classes: Low (0 – 0.30), Medium (0.3 – 0.63), High (0.63 <). The landslide release susceptibility map predicted 36 % of low susceptible, 31% of medium susceptible, and 33% of high susceptible where 80% of the total landslide release were predicted in high susceptible area, while the landslide susceptibility map by using full inventory predicted 0.14% of low susceptible, 0.35% medium susceptible, and 0.51% of high susceptible, where 83% of the total landslide were predicted in high susceptible area. The accuracy of both models is shown in Figure 9b and 9a. The landslide release susceptibility map shows better AUC of 0.815 compared to landslide susceptibility map with AUC of 0.798. The importance of factors in both models is shown in Figure 7. The key factors that control the landslide occurrence in full inventory model are Aspect, Elevation, and Distance from the river, while in landslide release model, the most important factors are Aspect, Distance to the river, and slope angle from the river.

The main reason Aspect became the most influential factor is strongly correlated to the topographical formation of the study area. The elevation in the study area varies from low in west part to high in the east part of the area. Most landslides occurrence was strongly related to the storms (typhoons) coming from both south west and east of Taiwan, causing aspect facing, south west, south, south east and east to be more susceptible to landslide. The given condition also explains the major different of the importance factors in both concepts. Elevation is considered as second most important factor in full inventory; while in landslide release it is considered as fourth (Figure 7).

Elevation controls the occurrence of landslide in full inventory since most of the landslides in the study area span in large range of elevation, while the landslide release area occurred in certain distance from the river in certain elevation condition. Hence the elevation is considered less important in landslide release concept.

![Figure 7. Important factors in two different concepts (full inventory and landslide release).](image-url)
4.2. Comparison of different data resolution in prediction accuracy

In order to generate landslide susceptibility map, various field data are required. Although there are many agencies, especially in Taiwan that provides field data in various types, common problem regarding to obtain a high-resolution data is that they often available with expensive costs, and not available in many remote areas. To see if the methods used in this study can also be implemented using a coarser data resolution, identical factors used in the previous process were constructed in 30 x 30-meter resolution. The training data were also constructed in 30-meter resolution. The random forest model was employed to generate the map.

The result of the 30m model is shown in Figure 8d. The susceptibility map was reclassified identically as the previous models. The model predicted 36% of low susceptible, 34% of medium susceptible, and 30% of high susceptible, where 73% of the total landslide releases are predicted in high susceptible area. The AUC of the 30m model is presented in Figure 9d, where the accuracy is 0.804.

The comparison of the resultant map of both models seems to have similar outlook in general (Figure 8b and 8d). Although due to its resolution, the red color density (that shows the high susceptible area) in 5-meter resolution is higher as it represented a smaller area in real situation. The capability to predict the landslide release occurrence and the high AUC value concluded that model with 5-meter resolution and 30-meter resolution has both good prediction accuracy and can be evaluated relatively similar to each other.

4.3 Comparison of different models

Logistic regression and decision tree are two models commonly employed in many studies to produce landslide susceptibility map. In order to analyses the results between LR, decision tree and random forest model, comparison of the three models were constructed. The previous 8 independent factors were employed along with the landslide release as the dependent factors to construct the decision tree and logistic regression model. The resulted landslide release susceptibility maps of the two models are shown in Figure 8c and 8e. The maps were also reclassified identically as the previous models with low index (0 – 0.30), medium (0.30 – 0.63), and high (0.63<).

Decision tree model, predicted 48.9% low susceptible, 2.1% medium susceptible, and 49% high susceptible, where 88.4% of the total landslide release were predicted in high susceptible area. The logistic regression model, predicted low susceptible area of 7.1 %, medium of 47.2 %, and high of 45.7% of the total study area which predicted 66 % of the total landslide release in high susceptible area. The accuracy of both models is presented in Figure 9c and 9e, where the AUC of decision tree and logistic regression model are 0.727, and 0.652 respectively.

To make further comparison, visual analysis of the resultant map of the models can provide some differences. All three models were constructed by using the identical data points and instances, with 80% landslides and 20% no landslides. However, the map produced by each model showed different level of susceptibility in the study area. The map produced by logistic regression seemed to show some over fitting result, where almost half of the total area are considered as high susceptible, where the model accuracy is considered the lowest among the three models.

A better result compared to logistic regression was produced by the decision tree model with higher accuracy, although the map also predicted half of the study area as high susceptible, with only a very small part of the area are considered in medium susceptible. The best result is shown in map produced by random forest with the highest accuracy among the three, where only 33% of the study area is predicted as high susceptible and cover 80% of the total observed landslide release area.
Figure 8. a). Random forest model with full inventory, b). Random forest with landslide release, c). Decision tree model landslide release, d). Random forest model of 30m data resolution, e). Logistic regression landslide release.
Figure 9. The AUC curve of each model. a) random forest full inventory, b) random forest of landslide release, c) decision tree landslide release, d) random forest of 30m data resolution, e) logistic regression landslide release.

5. Conclusion
This study shows that with topographical combination of factors, a model that can successfully predict the occurrence of landslides release in this landscape of Zhoukou River basin, Taiwan.

The comparison of logistic regression, decision tree, and random forest by using the same data concluded that random forest model has the best prediction accuracy on predicting landslides, since the results showed that random forest model possess the best AUC value among the three models employed, where the resultant maps has proven the advantages of random forest to resists over fitting result. This study has also proven coarse data resolution can still be used to produce landslide susceptibility map with relatively similar accuracy compared to finer resolution.

The concept of landslide release area proposed earlier, offered an easier alternative way to handle landslides inventory for the source of landslide occurrence prediction with a reliable accuracy result.

A new measurement to calculate the average angle from the river has proven to be more useful when employed as landslide factors to generate a model in random forest with a good result.

The methodology used in this study could be applicable in other study areas to mapping landslide release. Finally, new concept of considering the historical release (source) of landslide occurrence to predict future landslide release occurrence was proposed, which can provide valuable information for
further investigations in the field, urban planning and infrastructures, and suitable hazards mitigation management and prevention strategies, since a potential release area could not only harm the exact location, but also other locations around it.

One of the main challenges on the present study was to define the best data sampling strategies as it strongly affected the prediction accuracy of the model. Further investigations to produce better data sampling strategies in order to gain better prediction accuracy is strongly encouraged.

Landslides employed in this study were all treated the same disregarding the types. Separating types of landslides and its triggering factors could produce more reliable and valuable landslide susceptibility map.

The topographical landslide factors previously employed to predict landslide occurrence were mostly derived from the digital elevation model. Hence, a study employing more landslide factors such as precipitation, geological fault formation, in order to produce a wider landslide susceptibility map with more factors consideration in the future could be made.

Since the present study earlier proposed a prediction of landslide release occurrence, disregarding its potential propagation and impact. One idea is to employ a model to predict the potential impact based on its propagation probability. So that a prediction could be generated not only for a landslide release occurrence, but also the potential impact area based on its predicted potential landslide release.

6. References

[1] Tsangaratos, P., Ilia, I., 2016. Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi Perfection, Greece. Landslides 13, 305–320. doi:10.1007/s10346-015-0565-6.
[2] Wahyudhi, I. Made Ari, I. N. Norken and I Ketut Suputra. 2015. Stakeholder Participation in the Operation and Maintenance of Unda Irrigation Networks. Spectral Journal Vol, 3 No. 2.
[3] Poudyal, C.P., 2013. Landslide susceptibility analysis using decision tree method, Phidim, Eastern Nepal. Bull. Dep. Geol. Vol 15 2012.
[4] Yeon, Y.-K., Han, J.-G., Ryu, K.H., 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol. 116, 274–283. doi: 10.1016/j.enggeo.2010.09.009.
[5] Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M., 2016. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. Landslides 13, 839–856. doi:10.1007/s10346-014-0614-1.
[6] Zhang, K., Wu, X., Niu, R., Yang, K., Zhao, L., 2017. The assessment of landslide susceptibility mapping using random forest and decision tree methods in the Three Gorges Reservoir area, China. Environ. Earth Sci. 76, 405. doi:10.1007/s12665-017-6731-5.
[7] Cascini, L., Cuomo, S., Pastor, M., Sacco, C., 2013. Modelling the post-failure stage of rainfall-induced landslides of the flow type. Can. Geotech. J. 50, 924–934. doi:10.1139/cgj-2012-0375.
[8] Tsaparas, I., Rahardjo, H., Toll, D., Leong, E., 2002. Controlling parameters for rainfall-induced landslides. Comput. Geotech. 29, 1–27. doi:10.1016/S0266-352X(01)00019-2.
[9] Ng, C.W.W., Shi, Q., 1998. Influence of rainfall intensity and duration on slope stability in unsaturated soils. Q. J. Eng. Geol. Hydrogeol. 31, 105. doi:10.1144/GSL.QJEG.1998.031.P2.04.
[10] Mergili, M., Chu, H.-J., 2015. Integrated statistical modelling of spatial landslide probability. Nat Hazards Earth Syst Sci Discuss 2015, 5677–5715. doi:10.5194/nhessd-3-5677-2015.
[11] Tang, C.-L., Hu, J.-C., Lin, M.-L., Angelier, J., Lu, C-Y., Chan, Y.-C., Chu, H.-T., 2009. The Tsaoling landslide triggered by the Chi-Chi earthquake, Taiwan: Insights from a discrete element simulation. Eng. Geol. 106, 1–19. doi:10.1016/j.enggeo.2009.02.011.
[12] HSU, T.L., 1977. Mass movement in the Tsaol-Ling area, Yunlin-Hsien, Taiwan. Proc. Geol. Soc. China 20, 114–118.
[13] Avanzi, G.D., Galanti, Y., Giannecchini, R., Duchi, S., Presti, D.L., Marchetti, D., 2013. DP Test in Geotechnical Characterization of Shallow Landslides Source Area: Results and
Perspectives, in: Margottini, C., Canuti, P., Sassa, K. (Eds.), Landslide Science and Practice: Volume 1: Landslide Inventory and Susceptibility and Hazard Zoning. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 249–255. doi:10.1007/978-3-642-31325-7_33.

[13] M. Silveira, S. Heleno, P. Pina, 2016. Texton-Based Ensemble Classification of Landslide Source and Transport Areas in VHR Imagery. IEEE Geosci. Remote Sens. Lett. 13, 1787–1791. doi:10.1109/LGRS.2016.2610003.

[14] Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng. Geol. 79, 251–266. doi:10.1016/j.enggeo.2005.02.002.

[15] Kantardzic, M., 2002. Data Mining: Concepts, Models, Methods and Algorithms. John Wiley & Sons, Inc., New York, NY, USA.

[16] Tien Bui, D., Tuan, T.A., Hoang, N.-D., Thanh, N.Q., Nguyen, D.B., Van Liem, N., Pradhan, B., 2017. Spatial prediction of rainfall-induced landslides for the Lao Cai area (Vietnam) using a hybrid intelligent approach of least squares support vector machines inference model and artificial bee colony optimization. Landslides 14, 447–458. doi:10.1007/s10346-016-0711-9.

[17] Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. doi:10.1023/A:1010933404324

[18] Olden, J.D., Kennard, M.J., Pusey, B.J., 2008. Species invasions and the changing biogeography of Australian freshwater fishes. Glob. Ecol. Biogeogr. 17, 25–37. doi:10.1111/j.1466-8238.2007.00340.

[19] Immitzer, M., Atzberger, C., Koukal, T., 2012. Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data. Remote Sens. 4. doi:10.3390/rs4092661.

[20] Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J.P., 2012. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J. Photogramm. Remote Sens. 67, 93–104. doi:10.1016/j.isprsjprs.2011.11.002.

Acknowledgements
We would like to thank John F. Harrison, and Prof. Chih-Hua Chang from National Cheng Kung University for their data support and their experts opinion related to this research.