Predicting landslides probabilities along mountain road in Taiwan

Jianye Ching  
National Taiwan University of Science and Technology, Taipei, Taiwan

Hung-Jiun Liao  
National Taiwan University of Science and Technology, Taipei, Taiwan

ABSTRACT: This paper gives a brief introduction on the background and current status of landslides along mountain roads in Taiwan and presents a detailed analysis based on the Gaussian Process model for predicting locations and occurrence times of future landslides using historical data. Based on inherent and man-made features of failed and not-failed slopes, locations of possible future landslides are predicted. Together with historical rainfall data, a rainfall fragility graph is established. The analysis results show that the Gaussian Process model is effective in predicting landslide potentials and probabilities. Comparisons of the Gaussian Process analysis and the discriminant function analysis are made, which show that the former outperforms the latter in many aspects. The results are valuable for predicting where and when landslides would occur in future heavy rainfalls.

1 INTRODUCTION

The island of Taiwan was formed by the collision action of Eurasia plate and Philippine sea plate. It is relatively young in geological age. For the total area of 36000 km$^2$, mountains cover more than two thirds of Taiwan. The area percentages of mountains are about 28.40% for elevation of 0~100m, 39.48% for 100~1000m, 20.22% for 1000~2000m, 10.73% for 2000~3000m, and 1.17% for 3000~4000m, respectively. To accommodate population of more than 23 million, a large portion of the lower mountain area has already been intensively developed for cultivation and tourism. To keep up with the development, an extensive mountain road network has been built over the past decades.

The total length of roads with elevation above 100m in Taiwan is more than 67,000 km. Some of them were built with high engineering standards, but a large number of them were built with low engineering standards. Even worse, some were built along river valley with cut-slope methods and suffered from both river scouring problems on the down-slope and stability problems on the up-slope. Therefore, landslides in different failure types are not unusual along mountain roads when the slopes are experiencing long period of rainfalls or torrential rain accompanied with typhoons.

When a landslide occurs, it may cause traffic interruption, damage vehicles and injure personnel. Therefore, it is desirable to predict where and when landslides may happen to provide a safe traffic condition for the public. Although research results on landslide prediction have been reported, those focusing on landslides along mountain roads are rare. Landslides along mountain roads are somewhat different from other types of landslides: they are affected by numerous features, especially manmade features. Due to this nature, landslide prediction for slopes along mountain roads is by no means a straightforward matter but it is an interesting topic to work with. In this study, Route T-18 in central Taiwan is chosen to demonstrate the suitability of landslide prediction using Gaussian Process model. Two main questions are of concern: (a) Given the historical landslide data along the demonstrative mountain roads in Taiwan, where are the locations along the roads with high landslide potential in the future? (b) Given the historical landslide and rainfall data, what are the landslide probabilities of the slopes along the roads in a future heavy rainfall? The former mainly concerns with the locations of future landslides, while the latter concerns with the time of landslide occurrence in future rainfalls.

Note that the concept of probability is highlighted in this study due to the complexity of slope stability problems: it can never be certain where and when a slope will fail because there are numerous
uncertainties involved. In this study, a full probabilistic analysis based on the Gaussian Process model (MacKay 1998, William and Rasmussen 1996, Neal 1997; Gibbs 1997) is presented to predict the landslide potentials and probabilities. This new model is versatile and flexible. Compared to discriminant function analysis (Fisher 1936), the Gaussian Process model possesses the desirable property that there is no need to assume the functional form for the discriminant function, which is highly nontrivial task for complicated problems. Moreover, the performance of the Gaussian Process analysis is comparable to neural network while the required computational cost is much less.

2 BACKGROUND

2.1 Typhoons
Totally, 52 typhoons had attacked Taiwan between 1994 and 2004 with an average of 4.7 typhoons annually. Among them, Herb (year 1996), Toraji and Nari (year 2001), and Mindulle (year 2004) were the most devastating ones and caused a large amount of landslides in Taiwan. The accumulative rainfalls range from 1994mm for Typhoon Herb to 414mm for Typhoon Nari.

To deal with the landslides, the road maintenance sector tends to implement various types of slope stabilization measures, such as tied-down RC grid structure and gravity retaining wall to protect the failed up-slopes. However, the stabilization action cannot be done for all the slopes along the roads due to financial constraints, and new landslides may very likely occur in slopes that never failed and stabilization measures are absent. Therefore, it is desirable to develop techniques to predict the locations and occurrence times of new landslides. This is achieved through analyzing historical data gathered from the slopes along the roads, which will be explained in details in the next section.

2.2 Landslide database
Landslides along the mileage between 21.5km and 83.5km of Route T-18 are documented. In total, 55 failed unprotected slopes along T-18 were extracted from the road maintenance records. Among them, 12 slopes failed during Typhoon Herb, 18 during Toraji, 9 during Nari, and 16 during Mindulle due to heavy rainfalls. To match the number of failed slopes, 54 not-failed unprotected slopes were chosen. The not-failed sites were checked on the field to assure that no evidence of previous slope failure is visible. Note that the not-failed slopes in the database are roughly uniformly distributed over the chosen Route T-18 section.

The data format is as follows: \( D = \{ (x_i, t_i) \colon i = 1, \ldots, 109 \} \), where \( x_i \in \mathbb{R}^{15} \) contains the values of the 15 landslide features of the \( i \)-th slope in the database; \( t_i = 1 \) if that slope failed, otherwise \( t_i = 0 \); \( p \) is the total number of slopes in the database. Totally, 15 landslide features are carefully chosen based on engineering judgments to reflect the status of a slope: these features are believed to have significant influence on the stability of a slope. The landslide features are obtained from digital maps and DTM information as well as from field reconnaissance.

2.3 Landslide features
Fifteen landslide features are categorized into natural features and man-made features. Among them, thirteen are natural features, and two are man-made features. The natural features cover the aspects of topography (4 features: slope direction, slope angle, slope height, and road curvature), geologic conditions (1 feature: outcrop strata age), bedrock structure (2 features: slope and dip direction difference and slope and dip angle difference), weathering & fracturing (2 features: block size and rock volume percentage), vegetation cover (2 features: area percentage of vegetative cover and thickness of canopy cover), drainage condition (1 feature: catchment area size), and seismicity (1 feature: peak ground acceleration induced by the Chi-Chi earthquake). The man-made features (2 features: excavation height at toe and change of slope grade due to toe cutting) quantify the impacts induced by road construction.

3 LANDSLIDE LOCATION ANALYSIS
Recall that one of the goals of this paper is to develop methods to predict the locations of future landslides. In this section, a methodology is proposed for the purpose. Predicting landslide occurrence
times will the subject of a later section.

The key concept proposed here is “landslide potential”. Since the slopes in the database have been classified into two groups (failed and not-failed), it is sensible to study the similarity between the slope of interest and each of the two groups. If the slope of interest is more likely to be in the failed group, its landslide potential is higher, and vice versa. Here a slope with high landslide potential means it is more likely to be a failed slope in the database, i.e. a dangerous slope, rather than a not-failed slope or a safe slope.

A single index is used to capture the similarity: \( P(t=1|x,D) \), i.e. the probability that \( t = 1 \) given \( x \) and \( D \), where \( x \) contains the fifteen features of the slope of interest; \( t \) denotes the membership of the slope, i.e. \( t = 1 \) if it belongs to the dangerous group and \( t = 0 \) for the safe group; \( D \) is the historical data. Note that once \( P(t=1|x,D) \) is known, \( P(t=0|x,D) \) can be readily calculated since \( P(t=0|x,D)=1- P(t=1|x,D) \). Therefore, \( P(t=1|x,D) \) quantifies the landslide potential of the slope of interest.

In this paper, the Gaussian Process analysis (MacKay 1998, William and Rasmussen 1996, Neal 1997, Gibbs 1997) is implemented to estimate the landslide potential, i.e. to compute \( P(t=1|x,D) \). A second purpose of the Gaussian Process analysis is to rank the importance of the fifteen features. The discriminant function analysis (Fisher 1936), which has been implemented by many previous researchers to predict landslide potentials and probabilities, is also implemented to compare with the Gaussian Process analysis. For the discriminant function analysis, the commercially available program SPSS is employed. Due to the limitation of paper length, the details for the discriminant function analysis are not presented here.

One drawback of the discriminant function analysis is that it has to be specified the functional form of the discriminant function, or equivalently, the functional form of the separating boundary of the two membership groups in the feature space. In the absence of physical basis for the specification, a usual choice is to take the weighted sum of the features (weights are unknown and to be determined) as the discriminant function. However, it is highly nontrivial to appropriately specify the functional form of the discriminant function if the problem at hand is highly complicated and not well understood such as the landslide problem. Inappropriate specification may lead to large bias.

For complicated problems, it is desirable to adopt an approach that is flexible in the sense that the discriminant function can be potentially arbitrary. The discriminant function analysis is not qualified for this purpose. Nonetheless, neural network fits in this purpose as long as the number of nodes in the network is sufficiently large. However, the required computational cost for neural network is quite high because when implementing neural network, a non-trivial high-dimensional non-convex optimization problem must be solved.

On the other hand, the Gaussian Process analysis is recently developed in the field of artificial intelligence for non-parametric regression and classification. It has the advantage of neural work: the discriminant function can be potentially arbitrary, but its required computational cost is much less than that required by neural network. In fact, Neal (1996) showed that Gaussian Process models are equivalent to neural networks with an infinite number of nodes whose weights are Gaussian random variables. This result is essential: in order to implement infinite-node neural networks, an infinite number of parameters are needed, so the computational cost can be extremely high, but the same task can be achieved by using Gaussian Process models with only a few parameters.

### 3.1 Gaussian Process model for discriminant function

A Gaussian process on \( \mathbb{R}^n \) is a stochastic process \( \{Y(x) : x \in \mathbb{R}^n\} \) such that any finite combination of \( \{Y(x_1), Y(x_2), \ldots, Y(x_m)\} \) is jointly Gaussian. Similar to multivariate Gaussian random variables, a Gaussian Process model is fully defined by the mean, variance, and covariance of the process.

Suppose the discriminant function \( Y \) is known (although in fact it is not), the landslide potential of a slope whose features are given as \( x \) is modeled as:

\[
P(t=1|Y(.),x) = \frac{1}{1 + e^{-Y(x)}} = \text{sig}(Y(x))
\]  

(1)

where \( t \) indicates the group membership, i.e. \( t = 1 \) if the slope belongs to the dangerous group and \( t = 0 \) if the slope belongs to the safe group.
otherwise; the discriminant function $Y$ is, in fact, uncertain and is modeled as a Gaussian Process; $\text{sig}(\cdot)$ denotes the sigmoid function. The uncertain discriminant function $Y$ is to be determined, and our goal is to estimate this function by using the past data $D = \{(x_i,d_i); i = 1,\ldots,p\}$.

Instead of assuming the functional form of the $Y$ function (as done in the discriminant function analysis), in the Gaussian Process analysis, it is only necessary to assume that the $Y$ function is “smooth” in a certain sense. Note that this smoothness constraint is desirable for sensible prediction. In the context of Gaussian Processes, this smoothness requirement can be enforced by letting $Y(x_i)$ and $Y(x_j)$ be highly correlated when $x_i$ and $x_j$ are “similar”. Physically, this smoothness constraint means that if the features of two slopes are similar, then the landslide potentials will be also similar, exemplified by the following covariance model:

$$
\text{Cov}\left[Y(x_i), Y(x_j) \mid H \right] = \theta_1 \cdot e^{-\frac{1}{2} \sum \frac{(x_i^{(k)} - x_j^{(k)})^2}{\sigma^2}} + \theta_2 \equiv C(x_i, x_j)
$$

(2)

where $x_i^{(k)}$ denotes the $k$-th feature in $x_i$; the exponent is the negative one-half of the square of the weighted Euclidean distance between $x_i$ and $x_j$; $\text{Cov}$ denotes covariance. One can verify that if the $x_i$-$x_j$ distance is small (i.e. $x_i$ and $x_j$ are similar), the covariance (or correlation) between $Y(x_i)$ and $Y(x_j)$ will be large, and vice versa; $\{\theta_1, \theta_2, r_1, r_2, \ldots, r_n\} \equiv H$ are the uncertain model parameters, called the hyper-parameters.

It is worthwhile to discuss the significance of the hyper-parameters. Among them, $\theta_1 + \theta_2$ gives the variance of $Y(x)$; $\theta_2$ gives the baseline variance of $Y(x)$ (the variance of the uncertain mean value of $Y(x)$) and $\theta_1$ gives the variance of $Y(x)$ besides the baseline; $r_k$ governs the importance of the k-th feature: if $r_k$ is large, the k-th feature is influential, and vice versa.

### 3.2 Estimation of landslide potential

The final goal of the analysis is to estimate the landslide potential $P(t=1 \mid x,D)$. If the discriminant function $Y$ is known, one can see that

$$
P(t = 1 \mid x,D,Y(.) ) = \frac{1}{1 + e^{-Y(x)}}
$$

i.e. the estimation of $P(t=1 \mid x,D)$ is trivial if $Y$ is given. Therefore, the estimation of $Y$ function is essential for the estimation of $P(t=1 \mid x,D)$. Unfortunately, the estimation of the $Y$ function based on $D$ is highly non-trivial. This is because the $Y$ function is only indirectly observed in the data $D$ through (3): if in the data $t_j = 1$, one can see $Y(x_i)$ is more likely to be positive, and if $t_j = 0$, $Y(x_i)$ is more likely to be negative.

Let us now start from an easier position: in the case when the $Y$ function is directly observed at $\{x_i; i = 1,\ldots,p\}$ and when the hyper-parameters $H$ are also given, the estimation of the $Y$ function is simple. Let us denote $a_i$ as the observed value of $Y(x_i)$. For simplicity of notation, we denote the vector $[a_1, a_2, \ldots, a_p]$ by $a_{1:p}$, $[Y(x_1), Y(x_2) \ldots Y(x_p)]$ by $Y_{1:p}$, and $[t_1, t_2, \ldots, t_p]$ by $t_{1:p}$, respectively. Due to the Gaussian inference theorem, the updated (posterior) probability density function (PDF) of $Y(x)$ given $a_{1:p}$ and $H$ is also Gaussian with the following mean and variance:

$$
E\left[Y(x) \mid H, a_{1:p} \right] = \text{Cov}\left(Y(x), Y_{1:p} \mid H\right) \text{Var}\left(Y_{1:p} \mid H\right)^{-1} a_{1:p} \equiv \mu\left(H, a_{1:p}\right)
$$

$$
\text{Var}\left[Y(x) \mid H, a_{1:p}\right] = \text{Var}\left[Y(x) \mid H\right] - \text{Cov}\left(Y(x), Y_{1:p} \mid H\right) \text{Var}\left(Y_{1:p} \mid H\right)^{-1} \text{Cov}\left(Y(x), Y_{1:p} \mid H\right)^{T} \equiv \tau\left(H, a_{1:p}\right)
$$

(4)

where
\[
Var[Y(x) \mid H] = C(x, x) = \theta_1 + \theta_2 \\
Cov\left(Y(x), Y_{i,p} \mid H\right) = \begin{bmatrix}
C(x, x_1) & \cdots & C(x, x_p) \\
\vdots & \ddots & \vdots \\
C(x_{p-1}, x_1) & \cdots & C(x_{p-1}, x_p)
\end{bmatrix} \\
Var(Y_{i,p} \mid H) = \\
\begin{bmatrix}
\theta_1 + \theta_2 & C(x_1, x_2) & \cdots & C(x_1, x_p) \\
\vdots & \ddots & \vdots & \vdots \\
C(x_{p-1}, x_1) & \cdots & C(x_{p-1}, x_p)
\end{bmatrix}
\]

\text{(5)}

In reality, the \( Y \) function is not directly observed; instead, only \{\( t_i; i = 1, \ldots, p \}\} are observed. Moreover, the hyper-parameters \( H \) are uncertain as well. The goal of the Gaussian Process analysis is to estimate \( P(t=1 \mid x, D) \). According to the Theorem of Total Probability:

\[
P(t=1 \mid x, D) = \int P(t=1 \mid x, H, D, a_{i,p}, Y(x)) f(Y(x) \mid H, D, a_{i,p}) f(a_{i,p}, H \mid D) dY(x) da_{i,p} dH 
\]

\text{(6)}

where \( f(\cdot) \) denotes conditional PDF. Because

\[
P(t=1 \mid x, H, D, a_{i,p}, Y(x)) = P(t=1 \mid H, Y(x)) = \text{sig}(Y(x))
\]

\[
f(Y(x) \mid H, D, a_{i,p}) = f(Y(x) \mid H, a_{i,p})
\]

\text{(7)}

\text{(6)} can be written as

\[
P(t=1 \mid x, D) = \int \text{sig}(Y(x)) f(Y(x) \mid H, a_{i,p}) f(a_{i,p}, H \mid D) dY(x) da_{i,p} dH
\]

\text{(8)}

Note that \( f(Y(x) \mid H, a_{i,p}) \) is a Gaussian PDF whose mean and variance are specified in (4). According to MacKay (1992), the integral of the product of a sigmoid function and a Gaussian PDF can be approximated as

\[
\int \text{sig}(Y(x)) f(Y(x) \mid H, a_{i,p}) dY(x) \approx \text{sig}\left(-\mu(H, a_{i,p})/\sqrt{1+\pi \cdot \tau(H, a_{i,p})}/8\right)
\]

\text{(9)}

so (8) becomes

\[
P(t=1 \mid x, D) \approx \int \text{sig}\left(-\mu(H, a_{i,p})/\sqrt{1+\pi \cdot \tau(H, a_{i,p})}/8\right) \cdot f(H, a_{i,p} \mid D) da_{i,p} dH
\]

\text{(10)}

Therefore, the estimation of \( P(t=1 \mid x, D) \) involves the evaluation of the high-dimensional integral in (10), which is a highly non-trivial task.

In this paper, the integral in (10) is estimated through the Law of Large Number:

\[
P(t=1 \mid x, D) \approx \frac{1}{N} \sum_{i=1}^{N} \text{sig}\left(-\mu(H^{(i)*}, a_{i,p}^{(i)*})/\sqrt{1+\pi \cdot \tau(H^{(i)*}, a_{i,p}^{(i)*})}/8\right)
\]

\text{(11)}

where \( a_{i,p}^{(i)*} \) and \( H^{(i)*} \) are the \( i \)-th sample pair drawn from \( f(a_{i,p} \mid H, D) \). A stochastic simulation procedure based on Gibbs sampler (Geman and Geman 1984) and hybrid Monte Carlo (Duane et al. 1987) is presented to efficiently draw samples from \( f(a_{i,p} \mid H, D) \). The estimated integral is exactly the estimated value of the landslide potential of the slope of interest.
3.3 Results of landslide location analysis

Common practice of examining the performance of the adopted model/analysis is to quantify the so-called training errors, which are described as follows: Once the adopted model is trained by the data D, the trained model is used to predict the landslide potential for each slope in the database, i.e. compute \( P(t_i=1|x_i,D) \) for \( i = 1, \ldots, 109 \) in this study. If the landslide potential \( P(t_i=1|x_i,D) \) is larger than 50%, the \( i \)-th slope is predicted as a dangerous slope, otherwise, it is predicted as a safe slope. Compare the prediction with the actual status of the \( i \)-th slope, and the ratio of false prediction is the training error. Unfortunately, this calculation procedure of error is not fair. Because the \( i \)-th slope is already within the training data when its landslide potential is to be predicted, the resulting training error cannot effectively reflect the actual prediction errors on unseen slopes.

For fair calculation of prediction errors, the so-called leave-one-out (LOO) prediction errors of the adopted model are adopted here. The LOO prediction error is an unbiased estimate of prediction error on unseen slopes of the trained model. The basic idea of LOO prediction error is to mimic the prediction process by removing one data point out of the training dataset and use the removed data point for prediction testing. The procedure of computing LOO prediction error is as follows: Remove the \( i \)-th data point \( \{x_i,t_i\} \) from the dataset, call the remaining database \( D_{-i} \). Estimate \( P(t_i=1|x_i,D_{-i}) \). If \( P(t_i=1|x_i,D_{-i}) \) is larger than 50%, the \( i \)-th slope is predicted as a dangerous slope, otherwise, it is predicted as a safe slope. Compare the prediction with the actual status of the \( i \)-th slope. Do so for \( i = 1, \ldots, 109 \), and the ratio of false prediction is exactly the LOO prediction error.

Table 1 shows the traditional training error rates and the LOO prediction error rates induced by the Gaussian Process analysis and the discriminant function analysis. Note that both training errors are quite small (one of them is zero), but these are not realistic estimates for the actual prediction errors on unseen slopes. The LOO prediction errors, which more realistically reflect the actual prediction error rates, are always larger than the training error rates. It is also clear that the Gaussian Process analysis results in smaller LOO prediction error rates than the discriminant function analysis, indicating that the performance of the former is superior.

Table 1 Training errors and LOO prediction errors for the landslide location analysis

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Training errors</th>
<th># of False LOO Predictions (out of 109)</th>
<th>LOO Prediction Error Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discriminant Function Analysis</td>
<td>7.3%</td>
<td>13</td>
<td>11.9%</td>
</tr>
<tr>
<td>Gaussian Process Analysis</td>
<td>0%</td>
<td>6</td>
<td>5.5%</td>
</tr>
</tbody>
</table>

Note: LOO = leave-one-out analysis method

Having shown that the Gaussian Process analysis results in consistent prediction of landslide potential, we are now more confident to draw the following conclusion: The landslide potential predicted by the Gaussian Process analysis is satisfactory (the prediction error rate is as low as 5.5%). Figure 1 shows the landslide potential map predicted by the analysis for the entire Route T-18, where dark regions are high landslide potential (dangerous) segments and light regions are low potential (safe) segments. This figure is beneficial for the road maintenance sector to predict the locations of potential landslides along the road.

It is worth highlighting that the result of the analysis in the landslide location analysis is the landslide potential, i.e. the probability that the slope of interest belongs to the dangerous group. The concept of probability is taken in the analysis to accommodate the intrinsic large uncertainties in the problem. With these uncertainties, it is not realistic to give a yes-or-no answer. Instead, a probabilistic approach will be more appropriate. To this aspect, although not shown, the landslide potential predicted by the discriminant function analysis is found to be either very close to 0 or very close to 1, implying that the associated uncertainty is little. This contradicts the reality of the problem at hand: there exist many uncertainties so it is expected that the estimated landslide potential should not be close 1 or 0. Nevertheless, the landslide potentials calculated by the Gaussian Process analysis do not exhibit this undesirable behavior.
3.4 Relative importance among features
As discussed in a previous section, the $r_k$ of Gaussian Process model governs the relative importance of the $k$-th feature, hence the mean values of the $r_k$ samples (obtained from Gibbs sampler and hybrid Monte Carlo) quantify the relative importance of the feature: the smaller the $r_k$ mean value, the more important the $k$-th feature. The analysis shows that slope height, catchment area, height of toe cutting, block size, and change of slope angle are the dominant features, among them are the two man-made features. This result implies that the slope stability along mountain roads is noticeably affected by road construction. Also note that catchment area is among the dominant features. This result is consistent with the sense that slope stability should be sensitive to the amount of seepage and surface water, which is in general proportional to the size of catchment area.

It is of interest to verify if the removal of unimportant features will seriously degrade the performance of the proposed methodology or not. According to the analysis, the LOO prediction error rate of the Route T-18 data is only 1.83% (2 false LOO predictions out of 109) when only the most important five features are taken. It is clear that the performance of the Gaussian Process analysis with only five features is exceptional, implying that five features may be sufficient for accurate predictions. Even so, the number of fifteen features is not reduced and still used in this Gaussian Process analysis.

4 LANDSLIDE OCCURRENCE TIME ANALYSIS

The goal of the landslide location analysis is to identify the locations of the dangerous slopes, i.e. the slopes with high predicted landslide potentials. Besides predicting potential landslide locations, it is desirable to additionally predict “when” (i.e. during which typhoon) the dangerous slopes will fail. In Taiwan, landslides are mostly triggered by heavy rainfalls during typhoons. As a consequence, we propose a second stage of analysis (landslide potential is the first stage) to predict the occurrence times of landslides for the dangerous slopes. In the occurrence time analysis, the size of catchment area and “effective” rainfall amount are the two features studied. The latter is treated as the triggering feature of landslides. The goal here is to develop a methodology that determines the relationship between landslide probability and the two features for a dangerous slope given the past landslide and rainfall data. This relationship is called the rainfall fragility graph. As seen later, the rainfall fragility graph is useful in estimating real-time landslide probability for a dangerous slope. Moreover, if the rainfall amount can be effectively predicted a priori, it is then possible to predict the occurrence times of landslides.
with the rainfall fragility graph.

4.1 Database
The database employed here is based on the database for the landslide location analysis. However, there are some changes in the database for the occurrence time analysis. Firstly, most features, except the size of catchment area, are removed from the database, and one more feature, i.e. effective rainfall amount, is added into the database so that the total number of features is now two. The rationale of doing so is as follows: the purpose of most of the fifteen features in the landslide location analysis is to discriminate dangerous slopes from safe slopes. Once the dangerous slopes are located with the landslide location analysis, in the occurrence time analysis, it is only necessary to consider the amount of water supply to determine the landslide probability. Therefore, only the two features that are related to the amount of water supply, i.e. effective rainfall amount and size of catchment area, are considered in the occurrence time analysis.

Secondly, the slope data for Typhoon Herb (year 1996) is removed from the database since the landslide behaviors of the slopes is believed to change significantly after the Chi-Chi earthquake (year 1999). Finally, all not-failed slope data in the landslide location analysis is removed since the goal here is to predict landslide probabilities only for dangerous slopes.

For the database in the occurrence time analysis, a failed slope is directly copied from the landslide location analysis database (only the catchment size is kept), and the effective rainfall amount of the landslide-causing typhoon is used. Such slopes are called failed dangerous slopes. It should be noted that a failed slope here was a not-failed slope before the landslide-causing typhoon, indicating that each failed slope in the database can be used to create not-failed data points. This is achieved by taking a failed slope in the database and replacing its effective rainfall amount by the effective rainfall amount during each of the typhoons prior to the landslide-causing typhoon. By doing so, each failed slope in the database is duplicated into several not-failed slopes. Such slopes are called not-failed dangerous slopes. The failed dangerous slopes occurred during Typhoons Toraji, Nari, and Mindulle are chosen here. Among them, Typhoon Toraji was the earliest typhoon in date. So it is not used in generating the not-failed dangerous slopes. Following the above procedure, the total numbers of failed dangerous slopes and not-failed dangerous slopes in the new database are 43 and 33, respectively. These data points are shown in Figure 2.

4.2 Effective rainfall amount
The actual hourly rainfall data were obtained from 34 rainfall stations located in the mountainous region of central Taiwan. However, most of the rainfall stations are not adjacent to Route T-18. Therefore, the hourly rainfall amount history (in mm) at a dangerous slope is determined by interpolation with the inverse-distance-weighting (IDW) method built in the ArcGIS program. Once the hourly rainfall amount history at a dangerous slope is obtained, the following procedure is employed to quantify the effective rainfall amount $R_{eff}$ caused by a typhoon: Firstly, the 72-hour average rainfall $A_{72}$ is calculated at each slope for that typhoon. Secondly, a 12-hour moving window is employed to calculate the 12-hour moving average rainfall time history in the 72-hour interval. The maximum 12-hour moving average rainfall $M_{12}$ is therefore the maximum height of this time history. Note that $A_{72}$ is used to quantify the accumulated rainfall in the past 72 hours, while $M_{12}$ is used to quantify the rainfall intensity in the past 72 hours. The effective rainfall amount $R_{eff}$ is defined as the average of $A_{72}$ and $M_{12}$:

$$R_{eff} = \frac{1}{2} \left( A_{72} + M_{12} \right)$$  \hspace{1cm} (12)

The above equation takes into account the accumulative effect of the entire rainfall event and also the short-term impact of the rainfall intensity.

4.3 Results of occurrence time analysis
Instead of concerning with “landslide potential” in the landslide location analysis, here we are con-
cerned with the probability of landslide occurrence of a dangerous slope right after a typhoon. Although the goal here is different from the one in the landslide location analysis, the same Gaussian Process analysis can be taken to compute the landslide occurrence probability \( P(t=1|x,D) \), where \( x \) contains the two features (catchment area and \( R_{\text{eff}} \)) of the target dangerous slope; \( t \) denotes the status of the slope, i.e. \( t = 1 \) if it fails and \( t = 0 \) if it does not fail; \( D \) is the new database.

![Rainfall Fragility Graph](image.png)

**Figure 2 Rainfall fragility graph for dangerous slopes along Route T-18**

Figure 2 shows the results of the Gaussian Process analysis and is called the rainfall fragility graph. In the figure, the crosses “+” indicate the failed dangerous slopes in the database (43 data points), while the circles “o” are the not-failed dangerous slopes (33 data points). Note that in the lower-left region, i.e. small rainfall and small catchment area, most slopes did not fail, while most slopes failed in the upper-right region, i.e. large rainfall and large catchment area. This observation agrees with our intuition. The contour values indicate the value of the predicted landslide probability \( P(t=1|x,D) \) by the Gaussian Process analysis based on 43 failed dangerous slopes and 33 not-failed dangerous slopes. This rainfall fragility graph can then be used to predict landslide probability of a dangerous slope due to future typhoon.

Note that in Figure 2, the mixture of the failed and not-failed data points is evident, indicating that the differentiation between the failed and not-failed dangerous slopes is difficult. It once again indicates the necessity of adopting the concept of probability when dealing with the landslide problems. There are several possibilities for this difficulty: (a) the effective rainfall and catchment area are not the only controlling features for discriminating the failed cases from not-failed cases, i.e. there are some missing essential features; (b) it is difficult to perfectly quantify the rainfall amount; (c) the IDW method of interpolating the rainfall needs improvement; (d) the union of the above. Understanding the possible causes and improving prediction results are left as future directions.

4.4 Implementing rainfall fragility graph

The rainfall fragility graph can be used to estimate the real-time landslide probability of a dangerous slope, as demonstrated with the following example. Let us consider the rainfall history shown in Figure 3 for a dangerous slope whose catchment area is 27332 m². Since the \( R_{\text{eff}} \) index depends on the past
72-hour rainfall record, the rainfall history can be easily converted in real time to the $R_{eff}$ history, shown in Figure 3. Moreover, at each time instant, one can compute the landslide probability given the real-time $R_{eff}$ and the catchment area size with the aid of the rainfall fragility graph in Figure 2. Therefore, the landslide probability can be determined in a real-time manner, as shown in Figure 3.

If rainfall forecast is available, it is also possible to implement the rainfall fragility graph to predict landslide probability of a dangerous slope. However, the accuracy of the landslide prediction will heavily rely on the accuracy of the rainfall prediction.

![Figure 3 The rainfall, $R_{eff}$ and landslide probability time histories of a demonstrative dangerous slope](image)

5 CONCLUSION

Evaluating landslide potentials and probabilities of slopes along mountain roads is a complicated matter due to large uncertainties. In this paper, a full probabilistic analysis based on the Gaussian Process model is proposed to predict landslide locations and occurrence times along a mountain road in Taiwan. The landslide database for the slopes along the Alishan mountain road, which contains the features of 55 failed and 54 not-failed slope cases, is adopted to demonstrate the analysis. The analysis results show that the prediction error is around 5.5% for the landslide location analysis. The performance of the Gaussian Process analysis is superior to that of the discriminant function analysis, which is easier to operate than the former but less flexible. The relative importance of the chosen features is quantified. It is found that the Gaussian Process analysis based on the most importance five features performs exceptionally well. For the landslide occurrence time analysis, a rainfall fragility graph is proposed. Scattered data in the graph indicates that the uncertainties are significant and the need to adopt the concept of probability on landslide issue.

The research results of this paper include the landslide potential map and the rainfall fragility graph. The former quantifies the landslide potentials of the slopes along the road, so it is valuable for predicting the locations of future landslides, while the latter relates the landslide probability with the amount of rainfall and size of catchment area of a dangerous slope, hence it is valuable for predicting
the occurrence times of future landslides. Note that the implementation of the latter requires accurate prediction of rainfall amount, which is a challenging research by itself and will be left as a future research topic.

Finally, it should be pointed out that the rainfall fragility graph shown in Figure 2 is developed based on the rainfall data and the nature/man-made features of slopes gathered from Route T-18 of Taiwan. Modifications may be needed when applied to other mountain roads in other locations.

ACKNOWLEDGEMENTS

The authors wish to thank the Second and the Fifth Road Maintenance Divisions of the Taiwan Highway Bureau for providing the case history records and all the necessary helps to carry out the field observations along Route T-18. The financial supports provided by the Harbor and Marine Technology Center (Project No. MOTC-IOT-95-H1DB004) and the National Science Council of the Republic of China (Project No. 94-2625-Z-011-002) are also appreciated.

REFERENCES


