

An ensemble landslide hazard model incorporating rainfall threshold for Mt. Umyeon, South Korea

Ananta Man Singh Pradhan¹ · Hyo-Sub Kang¹ · Ji-Sung Lee¹ · Yun-Tae Kim¹

Received: 20 November 2016 / Accepted: 12 April 2017 / Published online: 16 May 2017
© Springer-Verlag Berlin Heidelberg 2017

Abstract In this study, a new ensemble method was developed to assess landslide hazard models in Mt. Umyeon, South Korea, using the results of a physically based model as a conditioning factor (CF). Hydrological conditions were obtained from the national-scale rainfall threshold. To incorporate rainfall threshold in landslide initiation, national landslide inventory data were used to prepare *I-D* and *C-D* thresholds. A series of factor of safety (*FS*) distribution maps were prepared using a physically based model with a 12-h cumulative rainfall threshold. We created an ensemble model to overcome limitations in the physically based model, which could not incorporate important environmental variables such as hydrology, forest, soil, and geology. To determine the effect of CFs on landslide distribution, spatial data layers of elevation, drainage proximity, soil drainage characters, stream power index, sediment transport index, topographic wetness index, forest type, forest density, tree diameter, soil type geology, and the *FS* distribution map were analyzed in a maximum entropy-based machine learning algorithm. Validation was performed with a receiver operating characteristic curve (ROC). The ROC showed 65.9% accuracy in the physically based model, whereas the ensemble model had higher accuracy (79.6%) and a prediction rate of 89.7%. The ensemble landslide hazard model is a new approach, incorporating the *FS* distribution map into the available independent environmental variables.

Keywords Ensemble model · Landslide susceptibility · Maximum entropy · Physically based model · Rainfall thresholds

Introduction

Rainfall is widely recognized as a major landslide-triggering factor in mountainous landscapes (Iverson 2000). Landslides are dangerous phenomena and it is important to interpret the conditions under which they most frequently occur. Landslides are recurrent and cause damage and casualties; therefore, great efforts are made to set up early warning systems (EWS) that can forecast their occurrence (Rosi et al. 2012). In recent decades, numerous studies have identified rainfall as the top predictive factor for landslide occurrence and have, therefore, endeavored to use rainfall characteristics to construct early warning and evacuation systems (Guzzetti et al. 2007; Wu et al. 2011; Brunetti et al. 2010; Martelloni et al. 2012; Turkington et al. 2014; Melillo et al. 2015, 2016; Petschko et al. 2014). Many researchers have also investigated landslide susceptibility models based on heuristic reasoning or statistical methods (Carrara et al. 1995; Guzzetti et al. 1999; van Westen et al. 2003; Lee and Pradhan 2006; Safaei et al. 2010). Some studies have used fuzzy logic and artificial neural networks for landslide susceptibility mapping (Ermini et al. 2005; Catani et al. 2005; Sezer et al. 2011; Akgün et al. 2012), and several researchers have proposed different physically based approaches based on infinite slope stability models, which are generally coupled with hydrological models (Montgomery and Dietrich 1994; Dietrich et al. 1995; Iverson 2000; Baum and Godt 2010). These methods are either probability- or scenario-based approaches. Although physically based models are easy to understand and have high predictive capabilities, they depend on the spatial distribution

✉ Yun-Tae Kim
yuntkim@pknu.ac.kr

¹ Department of Ocean Engineering, Geosystems Engineering Laboratory, Pukyong National University, 559-1, Daeyeon3-dong, Nam-gu Busan 608-737, South Korea

of various geotechnical data, which are very difficult to obtain. Statistical probability-based methods can include conditioning factors (CFs) that influence slope stability, such as forest, soil, and geology, which are unsuitable for physically based models. Statistical models rely on good landslide inventories of the site.

Shallow landslides are very common in the mountains of South Korea. Since the Korean Peninsula is part of the East Asian monsoon region, more than 60% of the annual precipitation is in the form of heavy rainfall influenced by typhoons and the rainy season (June–September). Consequently, intensive landslides occur during this period. Most historical typhoon events in South Korea have clearly shown that the spatial distribution of shallow regional landslides is focused in areas with high rainfall concentration (Pradhan and Kim 2014; Pradhan et al. 2016). However, despite the importance of rainfall as a factor triggering shallow landslides, few studies have addressed the rainfall threshold in South Korea. Previous studies determined that the rainfall threshold for shallow landslide initiation has been treated as a separate group in landslide susceptibility modeling, and warning models have mainly been based on rainfall return periods.

Against this background, the main objective of this study is to prepare a landslide hazard map incorporating rainfall thresholds. We used a physically based model coupled with a hydrological model to assess safety factors (FSs) in hilly terrain for various rainfall conditions on Mt. Umyeon, south of Seoul. The main difference between the present study and previous approaches is that rainfall threshold values are incorporated into the physically based model to assess four different warning levels. We used an ensemble approach to evaluate the output of a physically based model for a statistical machine learning model in varying hydrological conditions. The proposed ensemble model is an integrated process resulting from two different model approaches: physically based and statistical, and then synthesizing the results into a single score to improve the accuracy of predictive analytics.

Study area

The study area, located in the Seocho district and the southern part of Seoul (37.45°–37.48° N, 126.9°–127.04° E), covers approximately 6.8 km² (Fig. 1). The maximum elevation of the mountain is 312 m above mean sea level. The study area mainly comprises biotite gneiss, although the central part of the mountain is primarily granitic gneiss. In some areas of granitic gneiss formations, faults are reported by of Korean Geotechnical Society (2011) as depicted in Fig. 2. The soil at the site is principally weathered residual soil, covered by poorly sorted sand and gravel with a silty matrix. A significant area in the soil maps contained no data because some parts of the

study area contain a military camp, to which access is restricted for civilians.

The mountain is in the center of a dense residential area, such that landslide hazards have a great societal impact compared with landslides that occur in rural areas. According to Yune et al. (2013), the main cause of landslides in the Mt. Umyeon area is rainfall; the cumulative rainfall for 2 months before the landslide event was 1498.5 mm at Namhyeon and 1105 mm at Seocho station. Antecedent rainfall led to soil saturation on Mt. Umyeon, and several sites had already become vulnerable to slope failure. Additional torrential rainfall produced surface runoff continuously and weakened the ground, eventually resulting landslides which began at 9 A.M. (KST), July 27, and transformed rapidly into fast debris flows, which killed 16 people and damaged ten buildings, leading to economic losses of about USD 15 million. Some glimpses during landslide event are presented in Fig. 3.

On Mt. Umyeon, 163 landslide initiation locations were mapped using 1:5000-scale topographic maps (Fig. 1). A landslide inventory map was prepared by consulting satellite images and web-based digital aerial photographs with 50-cm resolution provided by Naver (www.map.naver.com) along with Global Positioning System (GPS) field surveys. A random partitioning algorithm was used to separate training landslides from validation landslides (Pradhan and Kim 2016a, b). There is no exact mathematical rule to determine the required minimum size of training data sets (Nefeslioglu et al. 2008; Caniani et al. 2008); however, there is a requirement that the training data include all CFs for the study area. Among the 163 landslide locations, 146 (approximately 90%) were selected for the training data set and the remaining 17 (about 10%) were used as validation data.

Methodology

This study was performed in three steps (Fig. 4): (1) rainfall thresholds of different warning level for shallow landslide initiation in South Korea were identified from landslide inventories and rainfall data; (2) factor of safety (*FS*) distribution maps were calculated based on warning levels obtained in the first step; and (3) an ensemble model was designed by incorporating the CFs responsible for landslide initiation. For the ensemble model, 11 CFs were collected from different sources and the *FS* distribution map obtained in the second step was considered as one of the CF.

Rainfall threshold

A threshold curve in the form of $I = \alpha D^{-\beta}$, where α (the intercept) and β (slope of power law curve) are constants; I is intensity and D is duration, following the works of various researchers (Caine 1980; Aleotti

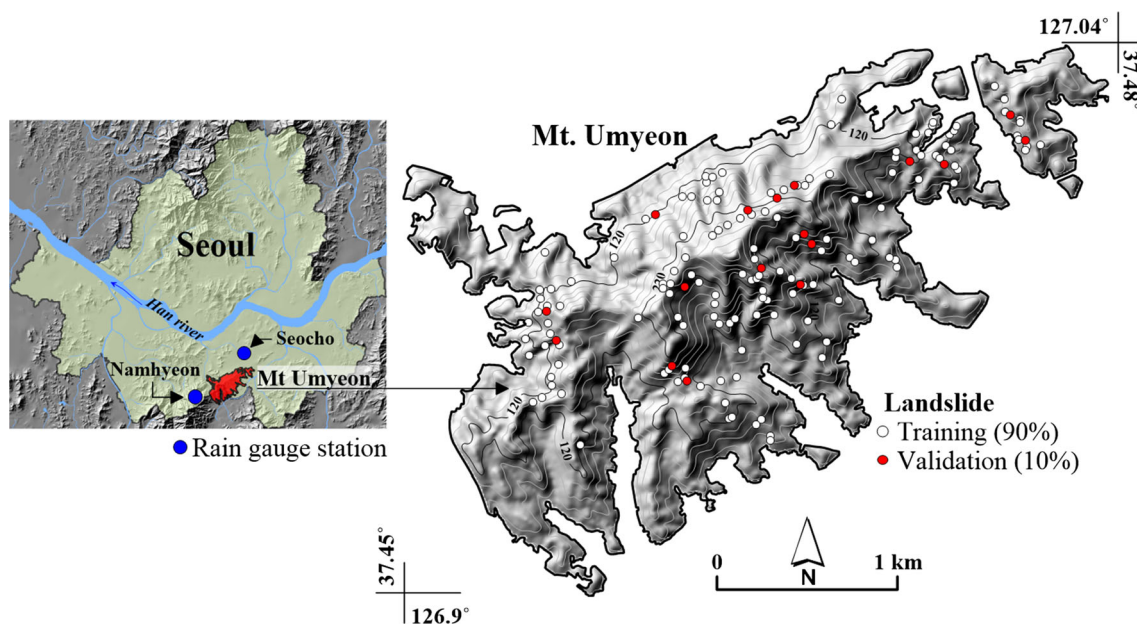


Fig. 1 Location of Mt. Umyeon and landslide distribution

2003; Glade et al. 2000; Jakob and Weatherly 2003; Chen et al. 2005; Hong et al. 2005; Guzzetti et al. 2007; Dahal et al. 2008), were used to determine rainfall thresholds. A frequentist approach was adopted to determine the slope and intercept of the power law curve selected to represent the rainfall threshold. The rainfall data are plotted in a single graph, and the distribution of the rainfall conditions, $\log(I)$ vs. $\log(D)$, that have resulted in landslides is fitted with a linear equation of the type $\log(I) = \log(\alpha) - \beta \log(D)$. For each

rainfall event (D, I) , the difference $\delta(D)$ between the logarithm of the event intensity $\log[I(D)]$ and the corresponding intensity value of the fit $\log[I_f(D)]$ was calculated as $\delta(D) = \log[I(D)] - \log[I_f(D)]$. Then, the probability density function (pdf) of the distribution of $\delta(D)$ is determined through Kernel Density Estimation (Venables and Ripley 2002). This method is based on a frequency analysis of the empirical rainfall conditions that have resulted in known landslides (Brunetti et al. 2010), where the probability density function of the

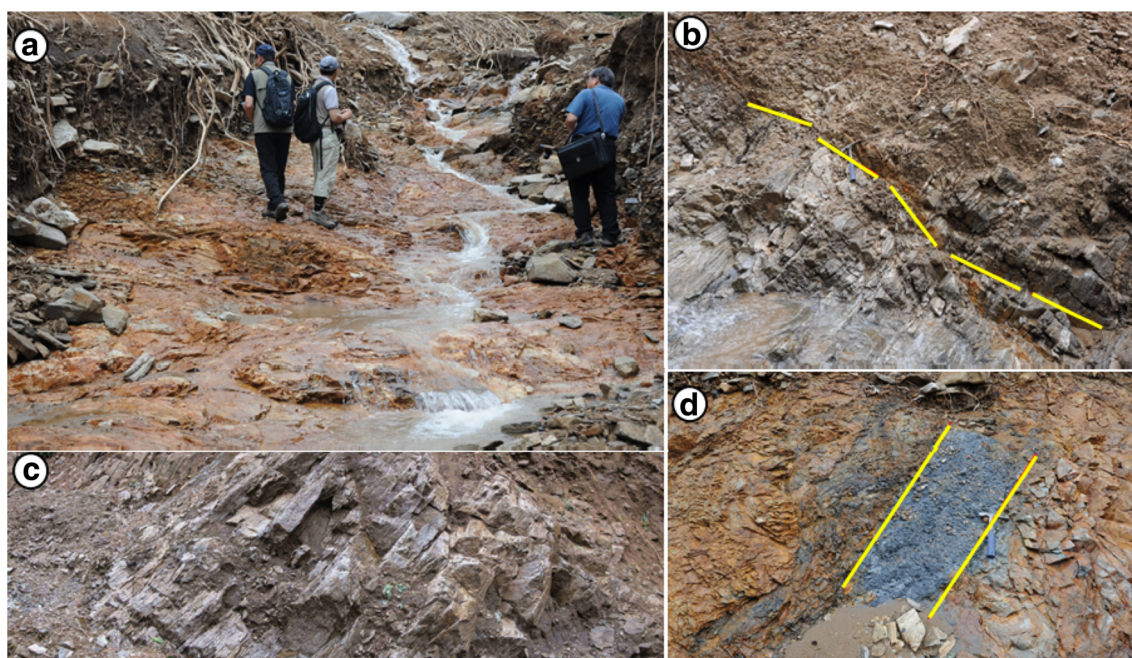


Fig. 2 **a** Landslide initiation area in biotite gneiss; **b** fault observed in biotite gneiss; **c** exposure of biotite gneiss; and **d** fault gouge

Fig. 3 Photographs show the landslide and its damages
(Photographs: Korean Geotechnical Society)



distribution is determined by fitting the result (via the least squares method) with a Gaussian function, as shown in Eq. 1.

$$f(\delta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\delta-\mu)^2}{2\sigma^2}\right) \quad (1)$$

where μ is the mean of the distribution that determines the position of the center of the peak, and σ is the standard deviation, which adjusts the scale of the curve. Based on the modeled distribution of δ , multiple thresholds can be defined, corresponding to different probabilities of exceeding thresholds (Brunetti et al. 2010). If the tested models are coupled with meteorological forecasts, they can be part of early warning systems (EWS) for managing landslide risk (Keefer et al. 1987). In order to achieve this, Safeland Deliverable D1.5 (2012) proposed three thresholds, i.e., four warning stages. Three thresholds encompass the 5th, 20th, and 50th exceedance probabilities of the landslide data set, and the warning levels refer to an increasing probability of the occurrence of landslides when precipitation overcomes

the given threshold. These thresholds can be used for EWS based on a real-time comparison between warning levels and hourly rainfall data recorded from a nearby network of rainfall stations.

Physically based model

The safety factor of infinite slopes can be calculated through the relationship between driving shear stress and resisting shear strength (Iverson 2000). To couple a hydrological model with the infinite slope model, FS was calculated (Brunsden and Prior 1984) as

$$FS = \frac{C'}{\gamma_s H \sin\theta \cos\theta} + \frac{\tan\phi'}{\tan\theta} - \frac{\gamma_w h \cos^2\theta \tan\phi'}{\gamma_s H \sin\theta \cos\theta}, \quad (2)$$

where C' is cohesion (kN/m^2), ϕ' is the internal frictional angle (degree), γ_s is the unit weight of soil (kN/m^3), H is the soil depth (m), h is the depth to the water table (m), θ is the local slope, and γ_w is the unit weight of water (kN/m^3). Figure 5 shows a schematic of an infinite slope-stability model.

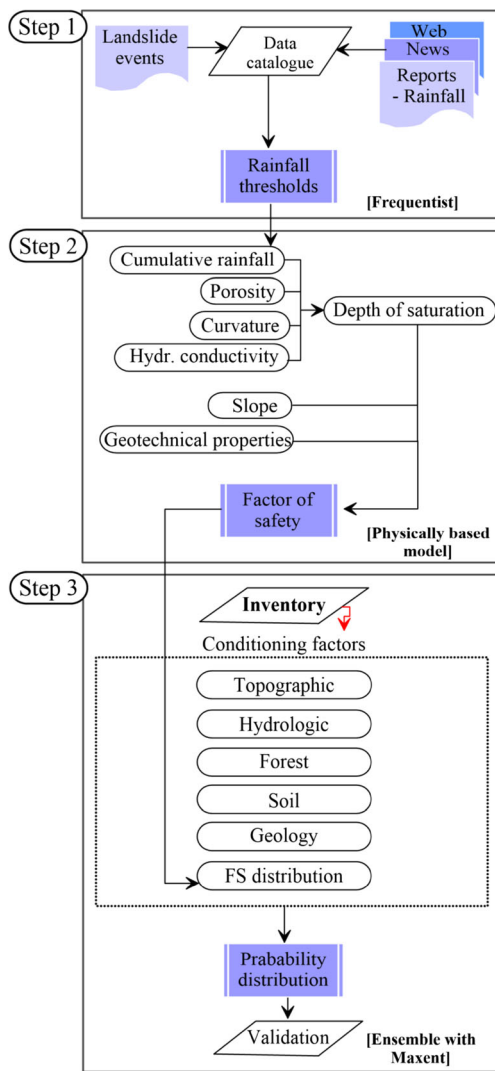
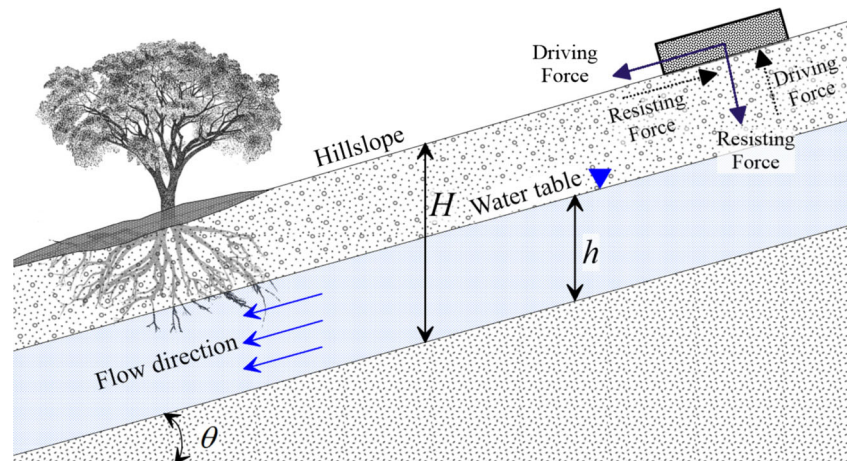


Fig. 4 Flow chart of the study

Proper infiltration models are necessary to perform reasonable estimation of the depth of saturation or infiltration depth, which reduces the shear strength of the soil cover. The amount

Fig. 5 Schematic of the infinite-stability model



of saturated depth through flow can be determined by considering various rainfall intensities. We applied the steady-state hydrological model proposed by Iida (1984) to estimate the saturated depth (h) based on rainfall intensity (R). He assumed that the saturated through flow in the soil layer moves down along a flow line as per Darcy's law, and introduced a horizontal component of flow velocity. The flow depth can be calculated as

$$h = \frac{R}{\mu} \left[t + \left(\frac{\varepsilon}{2} \right) V_s t^2 \right], \tag{3}$$

where t is the time (d), V_s is the horizontal velocity component (m/d), ε is the curvature of a particular terrain cell (m^{-1}), and μ is the effective porosity. V_s can be calculated expressed in terms of saturated hydraulic conductivity and effective porosity as follows:

$$V_s = \frac{k_s}{\mu} \sin\theta \cos\theta \tag{4}$$

where k_s is the saturated hydraulic conductivity (m/d).

The probability-based rainfall threshold for a 12-h duration (5th, 20th, and 50th exceedance probabilities) was used as the rainfall factor and saturation depth was considered for 12-h in Eq. (3) because the effective contributing area over a 12-h drainage time was strongly correlated with landslide occurrence on Mt. Umyeon (Pradhan and Kim 2017). Equations (2)–(4) were used to determine the FS for the slope of the terrain. All analyses were performed in a GIS environment (ArcGIS 10.2 and ILWIS 3.8, with the spatial analyst package). The benefit of a physically based model using GIS is the ability to express a wide range of recalculations of FS with varying rainfall intensities.

Ensemble model with maximum entropy

The ensemble model creates and combines multiple models to improve model results. To create a machine learning model,

we used a maximum entropy model (MaxEnt), which is increasingly employed in various earth science studies (Dyke and Kleidon 2010; Phillips et al. 2006) and has proven to be a very powerful statistical prediction tool (Convertino et al. 2013). MaxEnt represents the conditional density function of covariates π as presence, a random site x from the set X in the study area, and records 1 if the landslide is present at x , and 0 if it is absent. We then incorporated total rainfall using national-scale rainfall threshold levels into a physically based infinite-slope model. The result of the physically based model was then used as an input in MaxEnt to subsume conditional factors that were inapplicable in the physically based model.

Spatial database

A digital elevation model (DEM) is an important component in physically based and statistical landslide susceptibility mapping (Kawabata and Bandibas 2009; Fressard et al. 2014) because it generates attributes such as slope, aspect, and curvature (Shahabi et al. 2015; Tsangaratos and Ilija 2016). For this study, a 10×10 m DEM was prepared from LIDAR data. The LIDAR data was obtained from National Geographic Information Institute (NGII). The terrain slope calculates the slope at any pixel on the surface, i.e., the magnitude of the gradient at any point. From the definition of gradient, slope was derived from the first derivative function of the DEM. In the study area, the slope ranges from 0 to 47° (Fig. 6a). Generally, steeper slopes have a higher probability of experiencing landslides. The second derivative, curvature, corresponds to the convergence or divergence of water during overland flow (Oh and Pradhan 2011). Plan curvature reflects the rate of change of the terrain aspect angle measured in the horizontal plane. Negative values indicate divergent water flow over the surface, and positive values indicate convergent

flow. A value of ± 1.5 indicates that the surface is linear. Plan curvature distribution is presented in Fig. 6b. The slope and curvature rasters were created using the spatial analysis tool and were applied to the infinite-slope model but slope and curvature (suitable for a physically based model) were excluded in final ensemble model to avoid duplication of input parameters.

In general, the occurrence of landslides in an area was governed by various CFs, such as topography, hydrology, forest, soil, and geology (Pradhan et al. 2016). Eleven CFs were considered for the ensemble model, which could not incorporate into the physically based model. The source and significance of these CFs are presented in Table 1. DEM have been used as a CF in several studies (Yalcin 2008; van Westen 2004). Although there is no direct relationship between elevation and landslide occurrence, research has shown an increase in landslide occurrence at higher elevations (Ercanoglu and Gokceoglu 2004). Elevation in our DEM ranged from 20 to 312 m above mean sea level (Fig. 7a).

Hydrologic factors may adversely affect hillslope stability through erosion or saturation (Gökceoglu and Aksoy 1996). The proximity of a slope to a drainage site is important in determining instability because most landslides occur near streams. The Euclidean distance function in ArcGIS was used to find drainage proximity (Fig. 7b). Soil drainage characters (SDC) describe the frequency and duration of soil saturation. SDC determines whether water moves slowly, rapidly, or not at all. The study area contained the SDC classifications very poor, poor, moderate, and well (Fig. 7c). Stream power index (SPI) is a measure of the erosive power of water. As the specific catchment area and slope gradient increase, the amount of water contributed by upslope areas and water flow velocity increase (Moore et al. 1991). Figure 7d shows the spatial distribution of SPI in the study area. Sediment transport index (STI) characterizes the erosion and deposition processes

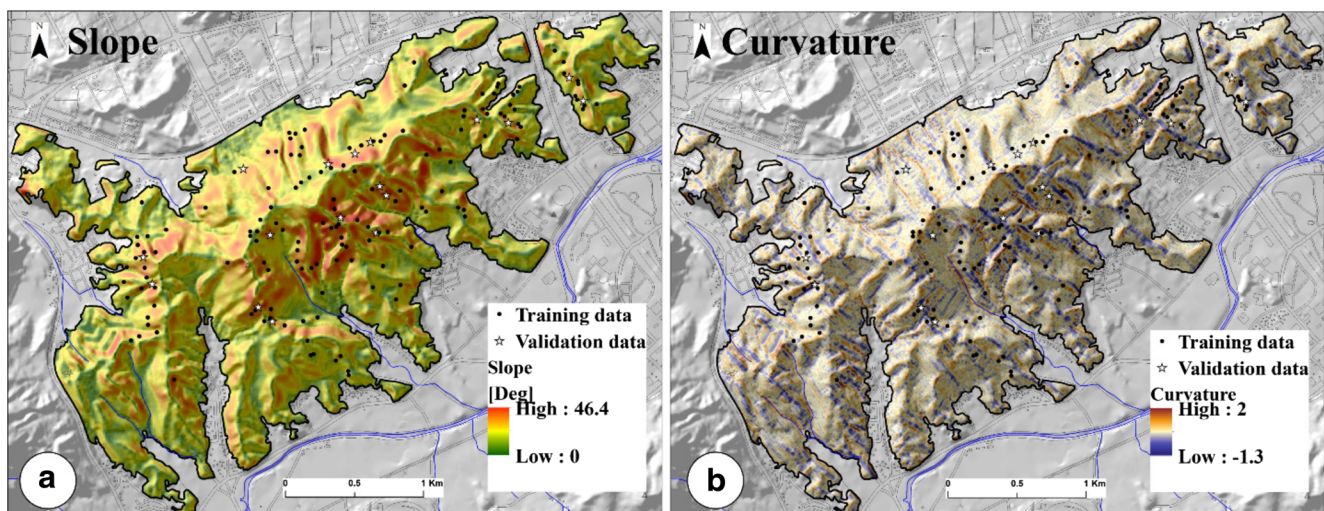


Fig. 6 a Slope distribution and b curvature distribution

Table 1 Eleven conditioning factors considered for ensemble model, their source and significance

Conditioning factors	Source	Significance
Elevation	National Geographic Information Institute (NGII), ArcGIS 10.2	Variation in climate, vegetation, potential energy
Drainage proximity	(NGII), ArcGIS 10.2	Influence of fluvial processes, saturation and erosion
SDC	Korea Forest Service	Frequency and duration of the period when soil is saturated with water.
SPI	(NGII), ArcGIS 10.2	Measures erosive power of flowing water
STI		Net erosion and deposition rates; provides an estimate of transportation capacity and erosion
TWI		Soil moisture conditions, source areas of runoff generation
Forest type	Korea Forest Service	Root strength, root penetration, water holding capacity
Forest density		Interlocking long fibrous binders, maintain soil pore water pressure
Tree diameter		Large diameter more roots
Soil type		Influence the rate of water movement, water holding capacity
Geology	Korea Institute of Geoscience and Mineral Resources (KIGAM)	Lithological properties of a hillslope

(Pradhan and Kim 2016b). A dimensionless STI (Fig. 7e) was calculated by combining the slope factors of length and steepness. The topographic wetness index (TWI) describes the effect of topography on the locations and sizes of saturated source (Fig. 7f) and is indicative of the spatial distribution of soil moisture (Beven and Kirkby 1979).

Vegetation helps to stabilize forested slopes by providing root strength and by modifying the saturated soil water regime (Ziemer 1981). Plant roots can penetrate the soil mass to anchor into bedrock fractures, cross zones of weakness to more stable soil, and provide interlocking long fibrous binders

within a weak soil mass. Forest factors for the present study were adapted from the Korean Forest Institute (KFI). The forest in the study area consists mainly of coniferous, broad-leaf, and mixed-forest types (Fig. 8a). High tree density means more roots and greater capacity to maintain water and soil pore pressure in heavy rain (Lee et al. 2004). KFI has classified the forest density into loose, moderate, and high (Fig. 8b), assuming that large-diameter trees have more roots and thus greater capacity to maintain soil strength (Lee et al. 2004). Tree stem diameter is subjective; KFI has classified the trees in the study site as small- and moderate-diameter ones

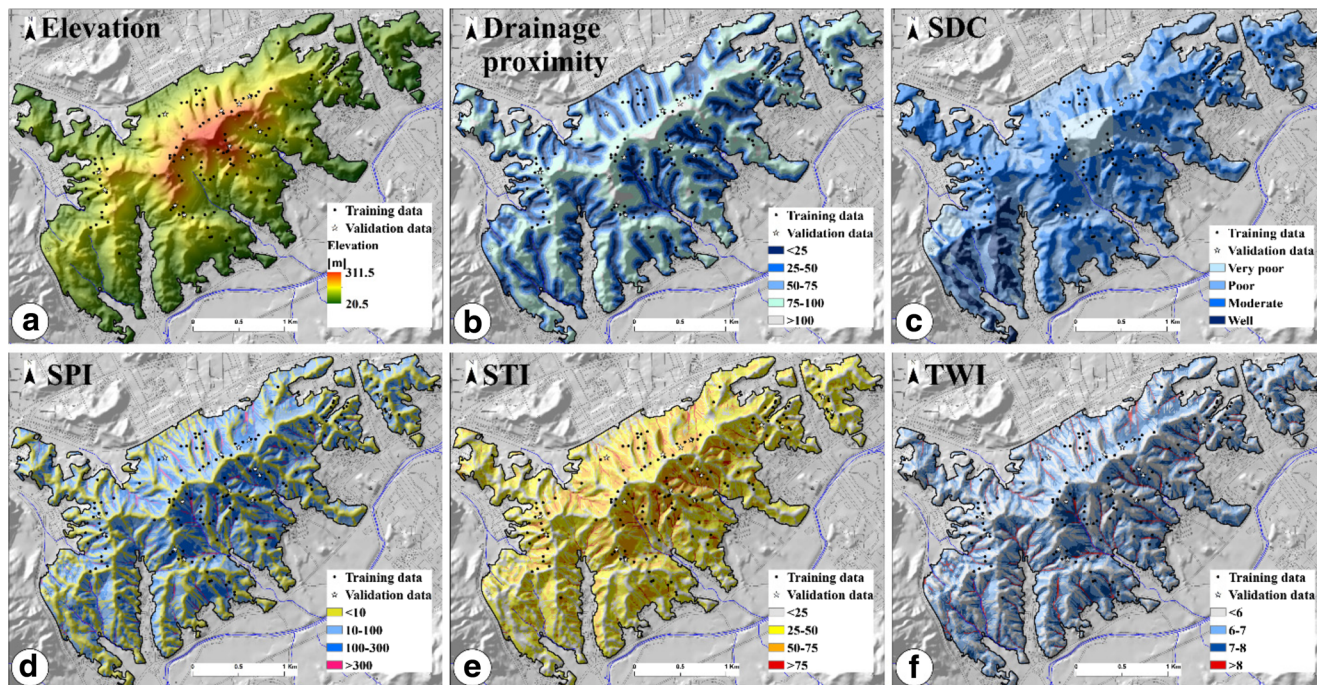


Fig. 7 a Elevation, b drainage proximity, c soil drain characteristics, d stream power index, e sediment transport index, and f topographic wetness index

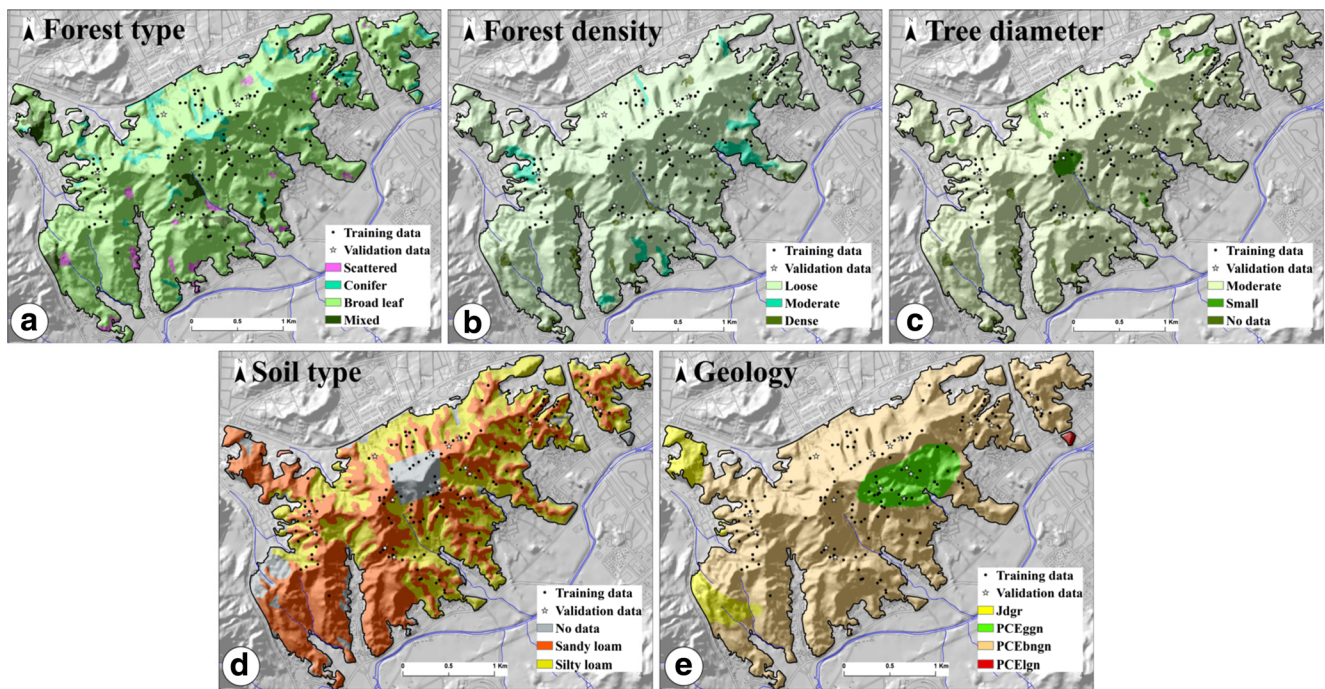


Fig. 8 a Forest type, b forest density, c tree diameter, d soil type, and e) geology

(Fig. 8c). Forest density and tree diameter were extracted from the forest map. The soil type also reveals the properties that influence the rate of water movement and the capacity of the soil to hold water (Sidle et al. 1985); related properties such as particle size and distribution of the soil matrix affect slope stability. The study area mainly comprises sandy loam and silty loam (Fig. 8d).

Different lithological units have different landslide susceptibility values (Carrara et al. 1995; Pachauri et al. 1998). As mentioned in our description of the study site, most of the mountain consists of biotite gneiss, with granitic gneiss at the center of Mt. Umyeon (Fig. 8e).

Results and discussion

Rainfall threshold

For this study, a national landslide inventory was compiled using data from NDMI (National Disaster Management Institute), various reports, and newspapers. Rainfall data were collected from the nearest meteorological observatories from the location of landslides. The search was limited to the period between May 1999 and August 2012. A total of 255 landslide events occurred in 1999–2012; among them, only 198 events were identified as rainfall-triggered landslides in weathered soil in South Korea. In this study, a critical rainfall concept was used as defined by Aleotti (2003). Critical precipitation indicates the amount of rainfall from the time (“zero point”) in

which a sharp increase in rainfall intensity is observed to the triggering time of the (first) landslide.

The critical intensity (I) and duration (D) for each event were plotted on a log-log diagram (Fig. 9a) to define rainfall events with durations between 2 and 77 h. The data were fitted

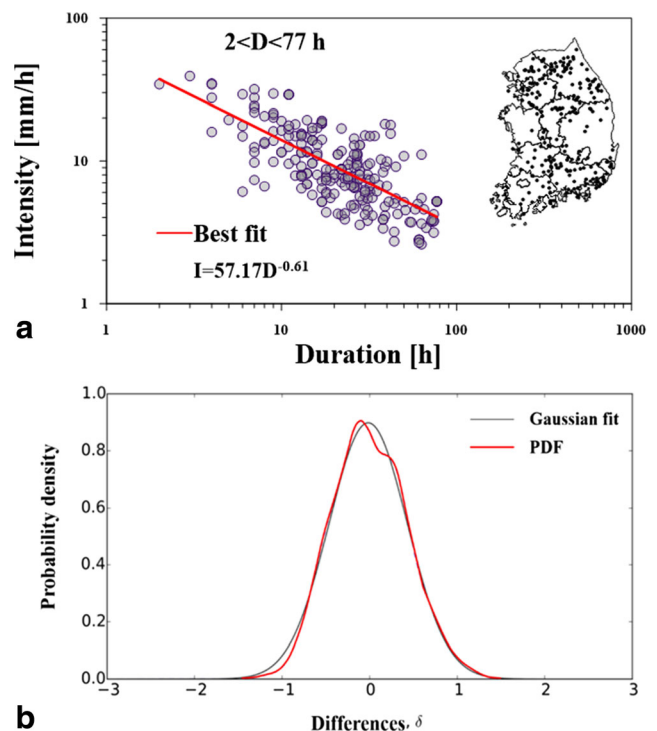


Fig. 9 198 rainfall events that resulted in landslides in South Korea and line of best fit

with the least squares method, yielding an equation of best fit of $I = 57.17D^{-0.61}$. Figure 9b shows the Gaussian fit of the probability density function for the 198 data points (I - D) shown in Fig. 9a. The proposed thresholds were used to propose a prototype EWS, and the limits of these intervals were used to set up four warning levels, namely, the regression lines of the 5th percentile, 20th percentile, and 50th percentile (Fig. 10a). The designated 5% percentile rainfall threshold shows that rainfall intensity greater than 18.56 mm/h is necessary to trigger shallow landslides for rainfall durations shorter than 2 h. For continuous rainfall of longer than 77 h, landslides may be triggered by 2 mm/h rainfall intensity. For validation purposes, the data characterizing 35 landslide events in Busan (in the south of South Korea) during 1999–2014 were superimposed on the I - D plot (Fig. 10a). All landslide events occurring in Busan during that period were above 5% exceedance probability (Fig. 10a). Conversion from I - D thresholds to C - D thresholds may easily be accomplished at each time step (Fig. 10b). The rainfall path observed in the Mt. Umyeon landslide event is also shown in Fig. 10. Warning levels were selected on the basis of the 5th percentile, 20th percentile, and 50th percentile (Table 2). They were classified as “null” (below the 5th percentile), “watch” (5th–20th percentile), “attention” (20th–50th percentile), and “alarm” (above the 50th percentile) warning levels.

The obtained rainfall thresholds were used to compute the FS of the terrain through spatial analysis. Using the slope,

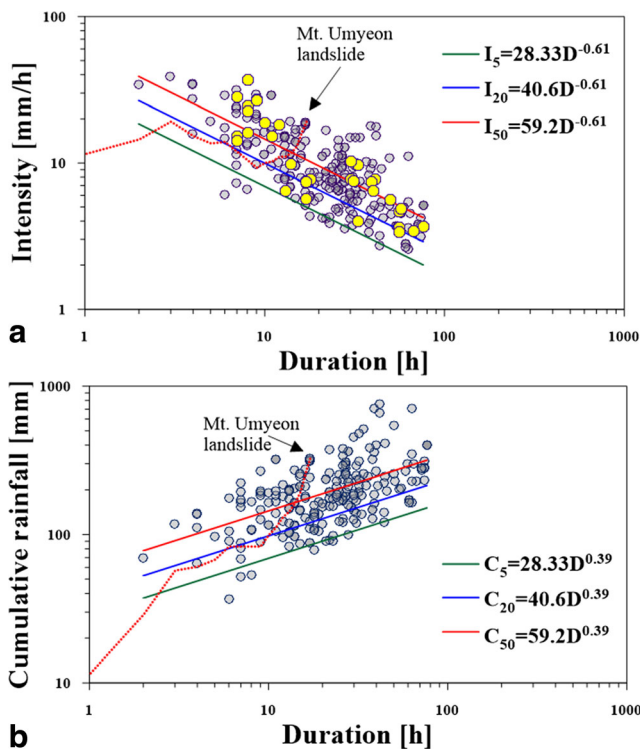


Fig. 10 Rainfall thresholds for different warning levels and rainfall path of Mt. Umyeon landslide: **a)** I - D conditions, yellow circles are Busan validation data and **b)** C - D shallow landslide conditions in South Korea

Table 2 Rainfall thresholds and corresponding cumulative rainfall

Classification	Criteria	Equation	Cumulative rainfall (mm) 12 h
Null	< 5% rainfall (0–5%)	–	–
Watch	5% < rainfall threshold < 20%	$C_{5\%} = 28.3D^{0.39}$	73.7
Attention	20% < rainfall threshold < 50%	$C_{20\%} = 40.6D^{0.39}$	105.7
Alarm	> 50% rainfall threshold	$C_{50\%} = 59.2D^{0.39}$	154.1

effective porosity, and saturated hydraulic conductivity, the subsurface flow velocity (Eq. 4) rasters were computed. Using this as an input, along with the curvature of the terrain, the saturated flow depth (Eq. 3) raster layers were computed for the given hydrological conditions. In this study, a 12-h cumulative rainfall threshold was used for the saturation depth calculation for different warning levels. The 12-h cumulative rainfall was reconstructed using the equations given in Table 2. Soil properties were obtained from geological engineering investigations for landslide hazard restoration works conducted by the Korean Society of Civil Engineering (2012) and Korean Geotechnical Society (2011) and National Forestry Cooperative Federation (2011). The sampling site is presented in Fig. 11. The geotechnical parameters for the study area are listed in Table 3, including cohesion, angle of internal friction, and unit weight of soil, which were rasterized for compatibility in spatial analysis. Soil depth is an important factor in an assessment of hillslope instability (Lanni et al. 2011). Park et al. (2013) stated that the thickness of the

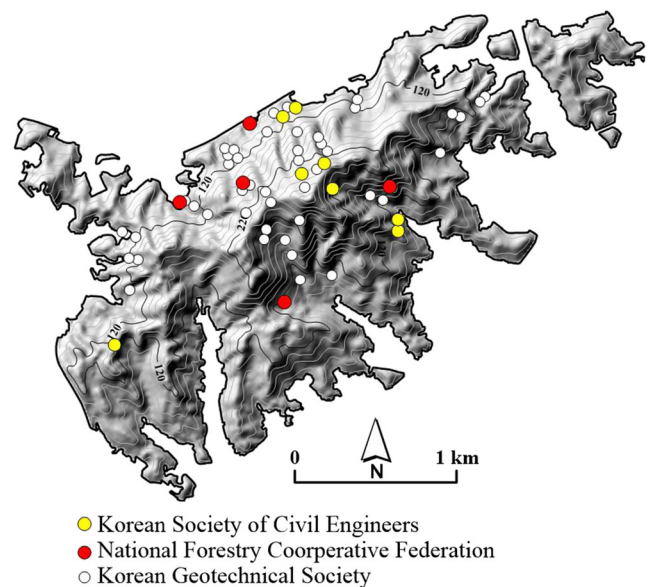


Fig. 11 Location of sampling sites by different agencies

Table 3 Geotechnical parameters of the study area

Properties	Value
Unit weight of soil (kN/m^3)	18.1
Cohesion C' (kPa)	7.7
Friction angle ϕ' ($^\circ$)	28.5
Saturated hydraulic conductivity k_s (ms^{-1})	3.37×10^{-6}
Effective porosity μ	0.45
Average soil depth H (m)	2

colluvium is generally less than 2 m in Seoul because of the relatively shallow depth of the bedrock, and hence shallow landslides are frequent. In this study, soil depth of the main scarp of the shallow landslides were used; i.e., the soil depth was estimated through field observations of the 50 landslide scarps. The value adopted for the entire mountain was considered to be equal to the soil depth at which the slope failure occurred, and the average soil depth was estimated as 2 m.

The next step was to estimate FS values using the saturated flow depth raster layers corresponding to different hydrological conditions (Eq. 2). The FS maps were classified as $FS < 1$, 1–1.2, 1.2–1.5, and > 1.5 . Dry conditions yielded no unstable pixels (Fig. 12a). Using 12-h cumulative rainfall and 5% exceedance probability, the unstable area increased to 0.58%. Under this particular condition, the area with $FS < 1$ can be considered a “watch” warning level (Fig. 12b). With 20% exceedance probability, the unstable area increased to 2.76%, and these unstable areas can be assigned to the “attention” warning level (Fig. 12c). Similarly, at a rainfall threshold of 50%, the unstable area was 7.52% (Fig. 12d); this area is assigned the “alarm” warning level.

The relationship between landslide occurrence and CFs

Figure 13 shows the results of frequency analysis that explore the relationship between CFs and landslide occurrence. It is

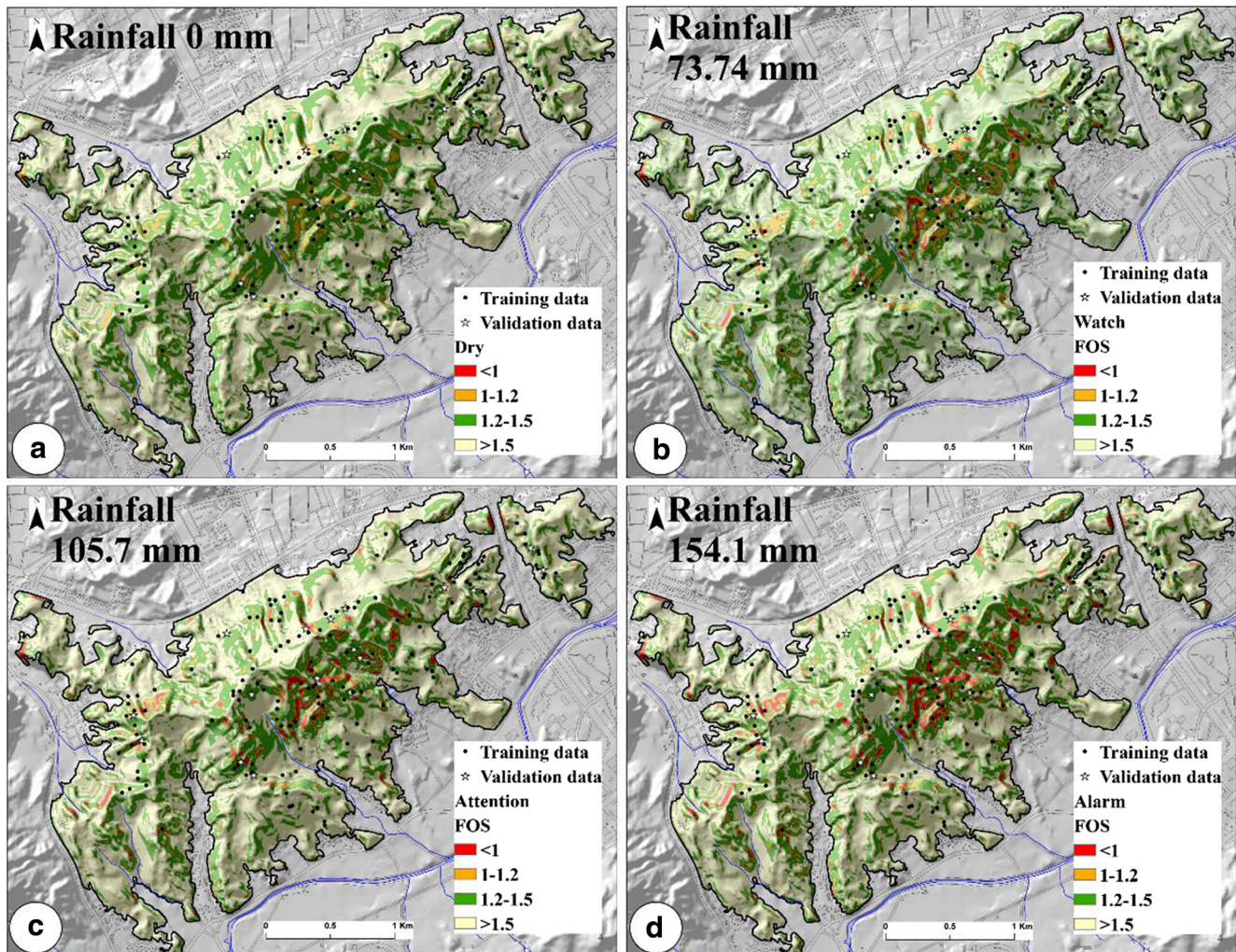


Fig. 12 Distribution of factor of safety (FS) in different rainfall scenarios (red area is unstable): a) dry conditions (\sim null), b) watch, c) attention, and d) alarm

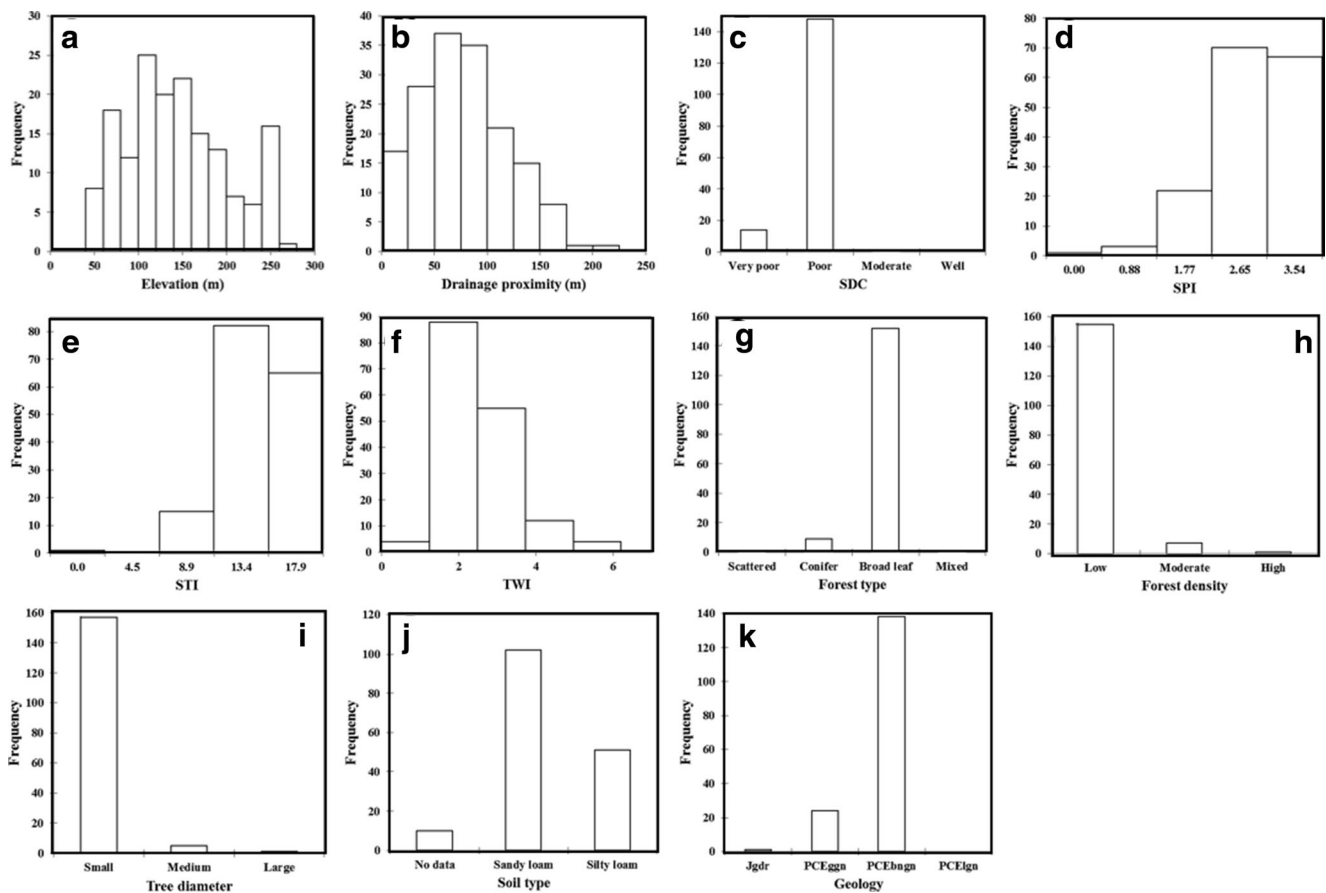


Fig. 13 Correlation between landslide frequency and the CFs

seen that most of the landslides occurs in confined to an elevation range of 60 to 250 m (Fig. 13a). At this intermediate elevation, slopes may be prone to slide due to the cover by thin colluvium. The correlation analysis between drainage proximity and landslide occurrence is shown in Fig. 13b. The maximum distance from the centerline of drainage is about 500 m. The landslide frequencies are high up to a drainage proximity of 120 m, but beyond that landslide frequencies become low. In case of SDC, landslides are concentrated in very poor and poor drainage conditions (Fig. 13c). Drainage problems often arise from lack of large-sized pores. In small-sized pores, water is slow to move and soils easily become waterlogged. For SPI, high SPI values are highly correlated to landslide occurrence (Fig. 13d) because a high SPI value is indicative of water contribution from upslope area and high water velocities (Pradhan et al. 2016). In case of STI, landslide frequency mostly occurred at high value of STI (Fig. 13e), a high STI accounts for the effect of topography on erosion. Regarding to TWI, landslide occurrence is highly correlate with low value of TWI (Fig. 13f). A low TWI values indicates high moisture accumulation. The correlation between landslide and forest type shows that most of the landslide concentrated in the broad leaf tree forest as shown in Fig. 13g. In case of forest density, landslide occurrence is highly correlated with low

forest density as shown in Fig. 13h. The small-diameter trees shows high correlation with landslide occurrence (Fig. 13i). The landslide frequency is high in the slopes covered by sandy

Table 4 Multi-collinearity diagnosis indexes for CFs used in ensemble model

CFs	Multi-collinearity indexes	
	Tolerance	VIF
Elevation	0.7999	1.2501
Drainage proximity	0.8710	1.1481
SDC	0.9164	1.0913
SPI	0.0324	30.9007
STI	0.0888	11.2604
TWI	0.0815	12.2685
Forest type	0.4755	2.1030
Forest density	0.5657	1.7678
Tree diameter	0.3910	2.5575
Soil type	0.8716	1.1473
Geology	0.9336	1.0712
FS	0.6272	1.5943

Bold values are failed during collinearity test

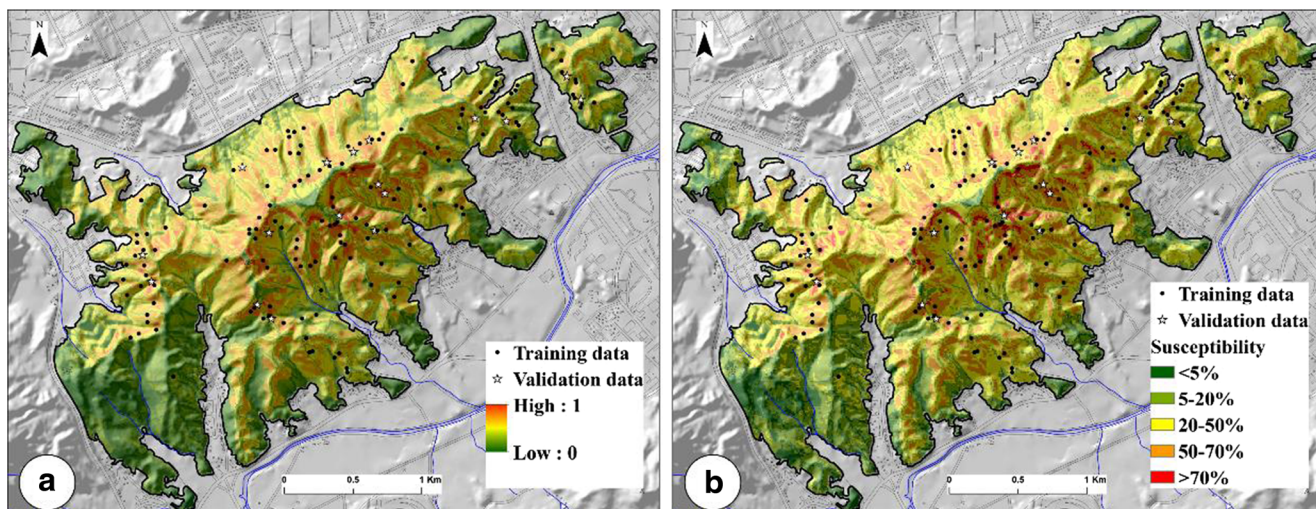


Fig. 14 a Probability distribution of landslide hazard and b classified probability of landslide hazard

loam as shown in Fig. 13j. In case of geology, most of the existing landslides are concentrated in PCEbngn lithological formation as shown in Fig. 13k.

Ensemble warning model

To create the ensemble model, the result of the physically based model for the ‘alarm’ warning level was used as one of several CFs because the *I-D* and *C-D* thresholds observed in the landslide event of Mt. Umyeon were flagged as “alarm” (Fig. 10a, b).

A multi-collinearity test was done on the set of 12 CFs to reduce the dimension of CFs. The variance inflation (VIF) and tolerance (TOL) are widely used indexes of the degree of multi-collinearity (Dou et al. 2015, Kavzoglu et al. 2014). A

VIF value greater than or equal to 5 and a TOL value less than 0.2 indicates a serious multi-collinearity problem (O’Brien 2007, Menard 1995). In this study, both these indexes were calculated which is summarized in Table 4. There is multi-collinearity between TWI, SPI, and STI; therefore, the remaining nine CFs were selected to generate the ensemble using MaxEnt. The probability distribution of the CFs was created using a Gibbs distribution, an exponential distribution parameterized by a vector of feature weights because Gibbs distribution minimizes the relative entropy in the variable space (Dudik et al. 2007). The final map shows the probability distribution of landslide hazard ranges from 0 to 1 (Fig. 14a). To improve visualization, the landslide hazard map used <5%, 5–20%, 20–50%, 50–70%, and >70% classes (Fig. 14b). However, this final map does not predict when or exactly

Fig. 15 Logistic output response curves of the model

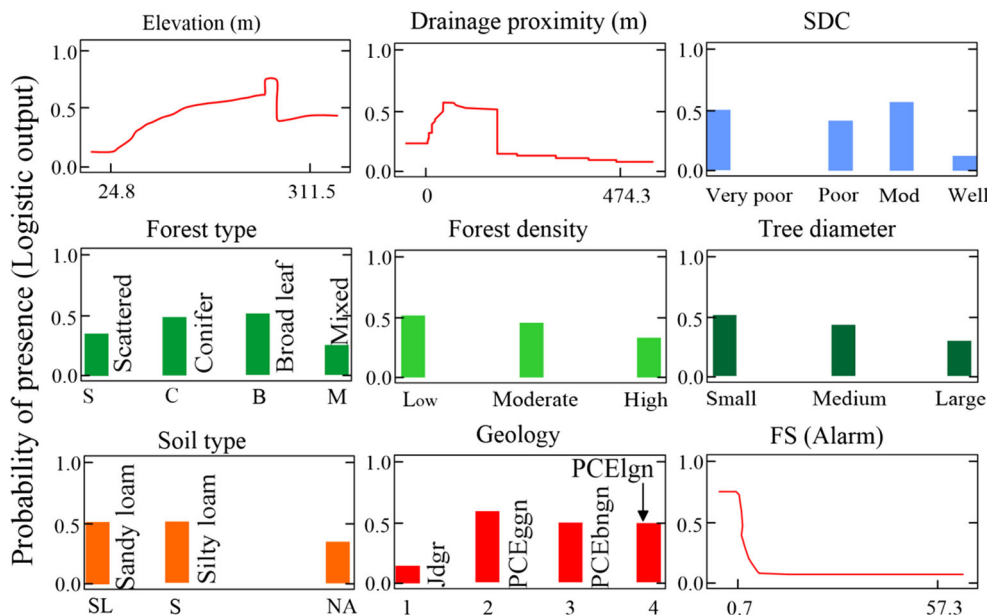
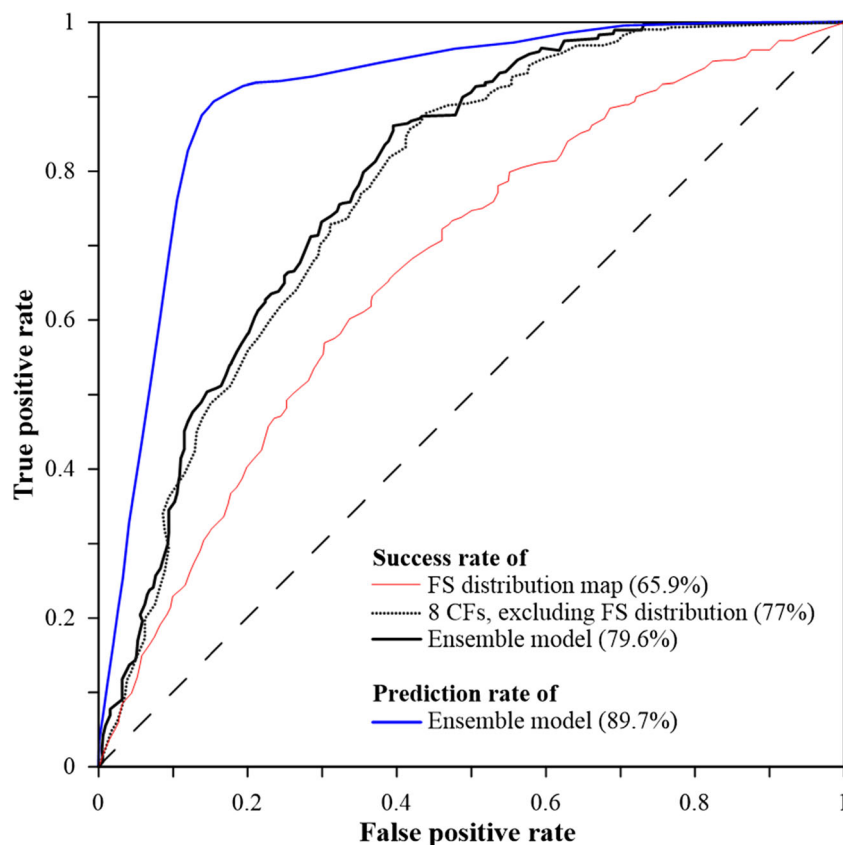


Fig. 16 Receiver operator curve (ROC) for model validation



where landslides will occur during a specific triggering event. Those classified hazard zones simply represent differences in the chance of landslide occurrence that can be expected over the long-term. The zoning shown on hazard map (Fig. 14b) is determined by analysis of landslide distribution in relation to the perceived CFs controlling those landslides on a regional scale, so it is only applicable to regional planning.

The conditional probabilities of the landslide susceptibility model are shown in the resulting response curves (Fig. 15). Response curves show how each CF affects the prediction. In the response curves, the y-axis is predicted probability, as given by the logistic output. As elevation increases, the probability of landslide occurrence also increases. The hillslope that is nearest to the drainage (20–200 m) shows a high probability of failure. This study considered main scarp to be relatively far from the mainstream line. Thus, the hillslope that is very close (i.e., < 20 m) to the mainstream line showed a lower probability of landslide occurrence. In present study, the SDC classes with very poor, poor, and moderate are more susceptible for presence of occurrence of landslide.

Broadleaf and conifer trees provided poor root reinforcement and anchoring, and the root system of broadleaf trees was less extensive in the study area. Broadleaf and conifer trees show a higher probability of landslide occurrence.

Lower tree density resulted in poor root reinforcement and anchoring, such that the probability of landslide occurrence was high for both low and moderate tree densities. Small- to medium-diameter trees resulted in greater probability of landslide occurrence.

The silty loam soil type, followed by sandy loam, showed the highest probability of landslide occurrence. In the silty loam and sandy loam areas with shallow soil depth, water could easily percolate through and create a perched aquifer in the boundary between soil and rock.

The area occupied by gneiss was prone to landslides. The granitic gneiss (PCEggn) and biotite banded gneiss (PCEbngn) lithological formations showed a higher probability of landslides. The PCEggn lithological unit has a small areal extent, but the landslide density of this unit was the highest among the lithological classes.

According to response curve, the FS distribution for the “alarm” warning level shows that areas with low FS values (< 1) are more susceptible to landslide occurrence. As the FS value increases, the probability of landslide decreases, indicating stable conditions for $FS > 1$.

The complexity of landslides just as other natural processes requires a broad based philosophy in understanding it since the behavior of most intrinsic factors affecting these processes is not well known.

Accuracy assessment

Model performance was assessed using a receiver operating characteristic curve (ROC), which is applied in many fields to test model performance. The ROC is an effective tool to represent the character of a forecast system (Swets 1988), and area under curve (AUC) is a useful indicator to validate the prediction performance of the model. Chung and Fabbri (2003) distinguished between success- and prediction-rate curves. The success-rate curve is based on a comparison of the hazard map with the landslides used in modeling (i.e., the training data), and the prediction-rate measurement is carried out with the validation landslide inventory. Figure 16 shows that the AUC value of the success-rate curve (77%) using eight CFs (i.e., elevation, SDC, forest type, forest density, tree diameter, soil type, and geology) is higher than for the “alarm” level *FS* distribution map (65.9%), whereas, the ensemble model (i.e., using nine CFs, including “alarm” *FS* distribution as 9th CF) using the training data showed a 79.6% success rate, which is slightly higher than that excluding “alarm” *FS* distribution map as one of the CFs. Pradhan and Kim (2016a) also reported that the combination of a statistical model and a physically based model can increase model performance. In this study, model performance was not greatly improved, but the result indicates that the ensemble model approach is sound. Additionally, the prediction-rate of the ensemble model was calculated using the validation data, with a result of 89.7%, indicated that the result is excellent.

Conclusion

An ensemble landslide hazard model was proposed and applied at Mt. Umyeon, south of Seoul, in which an *FS* distribution map obtained from a physically based model was integrated into an ensemble model. The physically based model was prepared using rainfall threshold values from national inventory data. The resulting physically based model can incorporate rainfall, steady-state infiltration depth, and *FS* distribution. However, it cannot include certain types of environmental data, such as topography, hydrology, forest type, soil type, and geologic CFs. The ensemble model could integrate several CFs that cannot be included in a physically based model alone. Thus, the obtained *FS* distribution for the “alarm” warning level was used as a CF in the ensemble model. The proposed ensemble model had high landslide prediction accuracy according to ROC validation, much higher than that of the physically based model alone. The accuracy of the physically based

model was 65.9%, whereas the ensemble model had a 79.6% success rate with a prediction rate of 89.7%. The ensemble landslide susceptibility model can be used as an EWS to encourage local authorities and the population to monitor rainfall variation, such that the local population may be alerted to avoid or evacuate threatened areas.

Acknowledgements This research was supported by the Public Welfare and Safety Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning (grant No. 2012M3A2A1050977), a grant (13SCIPS04) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA) and the Brain Korea 21 Plus (BK 21 Plus).

References

- Akgün A, Sezer EA, Nefeslioglu HA, Gokceoglu C, Pradhan B (2012) An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Comput Geosci* 38(1):23–34
- Aleotti P (2003) A warning system for rainfall-induced shallow failures. *Eng Geol* 73(3):247–265
- Baum RL, Godt JW (2010) Early warning of rainfall-induced shallow landslides and debris flows in the USA. *Landslides* 7(3):259–272
- Beven KJ, Kirkby MJ (1979) A physically based, variable contributing area model of basin hydrology. *Hydrol Sci J* 24(1):43–69
- Brunetti MT, Peruccacci S, Rossi M, Luciani S, Valigi D, Guzzetti F (2010) Rainfall thresholds for the possible occurrence of landslides in Italy. *Nat Hazards Earth Syst Sci* 10(3):447–458
- Brunsdon D, Prior DB (1984) Slope stability. Wiley, New York, p 620
- Caine N (1980) The rainfall intensity: duration control of shallow landslides and debris flows. *Geogr Ann Ser B*:23–27
- Caniani D, Pascale S, Sdao F, Sole A (2008) Neural networks and landslide susceptibility: a case study of the urban area of Potenza. *Nat Hazards* 45(1):55–72
- Carrara A, Cardinali M, Guzzetti F, Reichenbach P (1995) GIS technology in mapping landslide hazard. In: *Geographical information systems in assessing natural hazards*. Springer, Netherlands, p 135–175
- Chen CY, Chen TC, Yu FC, Yu WH, Tseng CC (2005) Rainfall duration and debris-flow initiated studies for real-time monitoring. *Environ Geol* 47(5):715–724
- Catani F, Casagli N, Ermini L, Righini G, Menduni G (2005) Landslide hazard and risk mapping at catchment scale in the Arno River basin. *Landslides* 2(4):329–342
- Chung CJF, Fabbri AG (2003) Validation of spatial prediction models for landslide hazard mapping. *Nat Hazards* 30(3):451–472
- Convertino M, Troccoli A, Catani F (2013) Detecting fingerprints of landslide drivers: a MaxEnt model. *J Geophys Res Earth Surf* 118(3):1367–1386
- Dahal RK, Hasegawa S, Nonomura A, Yamanaka M, Masuda T, Nishino K (2008) GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environ Geol* 54(2):311–324
- Dietrich WE, Reiss R, Hsu ML, Montgomery DR (1995) A process-based model for colluvial soil depth and shallow landsliding using digital elevation data. *Hydrol Process* 9(3–4):383–400

- Dou J et al (2015) Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. *PLoS One* 10(7):e0133262
- Dudík M, Phillips SJ, Schapire RE (2007) Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. *J Mach Learn Res* 8:1217–1260
- Dyke J, Kleidon A (2010) The maximum entropy production principle: its theoretical foundations and applications to the earth system. *Entropy* 12(3):613–630
- Ercanoglu M, Gokceoglu C (2004) Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (west Black Sea region, Turkey). *Eng Geol* 75(3):229–250
- Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66(1):327–343
- Fressard M, Thiery Y, Maquaire O (2014) Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the pays d'Auge plateau hillslopes (Normandy, France). *Nat Hazards Earth Syst Sci* 14(3):569–588
- Glade T, Crozier M, Smith P (2000) Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical “antecedent daily rainfall model”. *Pure Appl Geophys* 157(6–8):1059–1079
- Gökceoglu C, Aksoy H (1996) Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Eng Geol* 44(1):147–161
- Guzzetti F, Carrara A, Cardinali M, Reichenbach P (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology* 31(1):181–216
- Guzzetti F, Peruccacci S, Rossi M, Stark CP (2007) Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorol Atmos Phys* 98(3–4):239–267
- Hong Y, Hiura H, Shino K, Sassa K, Suemine A, Fukuoka H, Wang G (2005) The influence of intense rainfall on the activity of large-scale crystalline schist landslides in Shikoku Island, Japan. *Landslides* 2(2):97–105
- Iida A (1984) Hydrologic method of estimation of topographic effect on saturated throughflow. *Trans Jpn Geophys Union* 5(1):1–12
- Iverson RM (2000) Landslide triggering by rain infiltration. *Water Resour Res* 36(7):1897–1910
- Jakob M, Weatherly H (2003) A hydroclimatic threshold for landslide initiation on the north Shore Mountains of Vancouver, British Columbia. *Geomorphology* 54(3):137–156
- Kavzoglu T, Sahin EK, Colkesen I (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides* 11(3):425–439
- Kawabata D, Bandibas J (2009) Landslide susceptibility mapping using geological data, a DEM from ASTER images and an artificial neural network (ANN). *Geomorphology* 113(1):97–109
- Keefer DK et al (1987) Real-time landslide warning during heavy rainfall. *Science* 238(4829):921–926
- Korean Geotechnical Society (2011) Research contract report: addition and complement causes survey of Mt. Woomyeon Landslide. Koran Geotechnical Society, Seoul, 268p
- Korean Society of Civil Engineering (2012) Research contract report: causes survey and restoration work of Mt. Woomyeon Landslide. Korean Society of Civil Engineers, Seoul, 435p
- Lanni C, McDonnell JJ, Rigon R (2011) On the relative role of upslope and downslope topography for describing water flow path and storage dynamics: a theoretical analysis. *Hydrol Proc* 25(25):3909–3923
- Lee S, Pradhan B (2006) Probabilistic landslide hazards and risk mapping on Penang Island, Malaysia. *J Earth Syst Sci* 115(6):661–672
- Lee S, Ryu JH, Won JS, Park HJ (2004) Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. *Eng Geol* 71(3):289–302
- Martelloni G, Segoni S, Fanti R, Catani F (2012) Rainfall thresholds for the forecasting of landslide occurrence at regional scale. *Landslides* 9(4):485–495
- Melillo M, Brunetti MT, Peruccacci S, Gariano SL, Guzzetti F (2015) An algorithm for the objective reconstruction of rainfall events responsible for landslides. *Landslides* 12(2):311–320
- Melillo M, Brunetti MT, Peruccacci S, Gariano SL, Guzzetti F (2016) Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events. *Landslides* 13(1):165–172
- Menard S (1995) Applied logistic regression analysis. Sage, Thousand Oaks
- Montgomery DR, Dietrich WE (1994) A physically based model for the topographic control on shallow landsliding. *Water Resour Res* 30(4):1153–1171
- Moore ID, Grayson RB, Ladson AR (1991) Digital terrain modelling: a review of hydrological, geomorphological, and biological applications. *Hydrol Process* 5(1):3–30
- National Forestry Cooperative Federation (2011) Official archive for restoration work of Mt. Woomyeon landslide
- Nefeslioglu HA, Gokceoglu C, Sonmez H (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Eng Geol* 97(3):171–191
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41(5):673–690
- Oh HJ, Pradhan B (2011) Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. *Comput Geosci* 37(9):1264–1276
- Pachauri AK, Gupta PV, Chander R (1998) Landslide zoning in a part of the Garhwal Himalayas. *Environ Geol* 36(3–4):325–334
- Park DW, Nikhil NV, Lee SR (2013) Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. *Nat Hazards Earth Syst Sci* 13(11):2833–2849
- Petschko H, Brenning A, Bell R, Goetz J, Glade T (2014) Assessing the quality of landslide susceptibility maps—case study Lower Austria. *Nat Hazards Earth Syst Sci* 14(1):95–118
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190(3):231–259
- Pradhan AMS, Kim YT (2014) Relative effect method of landslide susceptibility zonation in weathered granite soil: a case study in Deokjeok-ri Creek, South Korea. *Nat Hazards* 72(2):1189–1217
- Pradhan AMS, Kim YT (2016a) Evaluation of a combined spatial multi-criteria evaluation model and deterministic model for landslide susceptibility mapping. *Catena* 140:125–139
- Pradhan AMS, Kim YT (2016b) Spatial data analysis and application of evidential belief functions to shallow landslide susceptibility mapping at Mt. Umyeon, Seoul, Korea. *Bull Eng Geol Environ* 1–17. Published online first
- Pradhan AMS, Kim YT (2017) GIS-based landslide susceptibility model considering effective contributing area for drainage time. *Geocarto Int*. doi:10.1080/10106049.2017.1303089 online first
- Pradhan AMS, Kang HS, Lee S, Kim YT (2016) Spatial model integration for shallow landslide susceptibility and its runoff using a GIS-based approach in Yongin, Korea. *Geocarto Int* 1–22
- Rosi A, Segoni S, Catani F, Casagli N (2012) Statistical and environmental analyses for the definition of a regional rainfall threshold system for landslide triggering in Tuscany (Italy). *J Geogr Sci* 22(4):617–629

- Safaei M, Omar H, Yousof ZB, Ghiasi V (2010) Applying geospatial technology to landslide susceptibility assessment. *Electron J Geotech Eng* 15:677–696
- Safeland (2012) Statistical and empirical models for prediction of precipitation-induced landslides, Deliverable D1.5
- Sezer EA, Pradhan B, Gokceoglu C (2011) Manifestation of an adaptive neuro-fuzzy model on landslide susceptibility mapping: Klang valley, Malaysia. *Expert Syst Appl* 38(7):8208–8219
- Shahabi H, Hashim M, Ahmad BB (2015) Remote sensing and GIS-based landslide susceptibility mapping using frequency ratio, logistic regression, and fuzzy logic methods at the central Zab basin, Iran. *Environ Earth Sci* 73(12):8647–8668
- Sidle RC, Pearce AJ, O'Loughlin CL (1985) Hillslope stability and land use. American geophysical union, Washington, D.C.
- Swets JA (1988) Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293
- Tsangaratos P, Ilia I (2016) Landslide susceptibility mapping using a modified decision tree classifier in the Xanthi prefecture, Greece. *Landslides* 13(2):305–320
- Turkington T, Ettema J, Van Westen CJ, Breinl K (2014) Empirical atmospheric thresholds for debris flows and flash floods in the southern French alps. *Nat Hazards Earth Syst Sci* 14(6):1517–1530
- Van Westen CJ (2004) Geo-information tools for landslide risk assessment: an overview of recent developments. In: *Landslides: evaluation and stabilization*. CRC Press, Boca Raton, p 39–56
- Van Westen CJ, Rengers N, Soeters R (2003) Use of geomorphological information in indirect landslide susceptibility assessment. *Nat Hazards* 30(3):399–419
- Venables WN, Ripley BD (2002) *Statistics and computing*. Springer, New York
- Wu CH, Chen SC, Chou HT (2011) Geomorphologic characteristics of catastrophic landslides during typhoon Morakot in the Kaoping watershed, Taiwan. *Eng Geol* 123(1):13–21
- Yalcin A (2008) GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. *Catena* 72(1):1–12
- Yune CY, Chae YK, Paik J, Kim G, Lee SW, Seo HS (2013) Debris flow in metropolitan area—2011 Seoul debris flow. *J Mount Sci* 10(2): 199–206
- Ziemer RR (1981) The role of vegetation in the stability of forested slopes. In: *Proceedings of the international union of forest research organisations*. Kyoto, Japan, p 297–308