



Estimating landslide susceptibility areas considering the uncertainty inherent in modeling methods

Ho Gul Kim¹ · Dong Kun Lee^{2,3,9} · Chan Park⁴ · Yoonjung Ahn⁵ · Sung-Ho Kil⁶ · Sunyong Sung⁷ · Gregory S. Biging⁸

Published online: 27 September 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Landslides are one of the most dangerous types of natural disasters, and damage due to landslides has been increasing in certain regions of the world because of increased precipitation. Policy decision makers require reliable information that can be used to establish spatial adaptation plans to protect people from landslide hazards. Researchers presently identify areas susceptible to landslides using various spatial distribution models. However, such data are associated with a high amount of uncertainty. This study focuses on quantifying the uncertainty of several spatial distribution models and identifying the effectiveness of various ensemble methods that can be used to provide reliable information to support policy decisions. The area of study was Inje-gun, Republic of Korea. Ten models were selected to assess landslide susceptibility. Moreover, five ensemble methods were selected for the aggregated results of the 10 models. The uncertainty was quantified using the coefficient of variation and the uncertainty map we developed revealed areas with strongly differing values among single models. A matrix map was created using an ensemble map and a coefficient of variation map. Using matrix analysis, we identified the areas that are most susceptible to landslides according to the ensemble model with a low uncertainty. Thus, the ensemble model can be a useful tool for supporting decision makers. The framework of this study can also be employed to support the establishment of landslide adaptation plans in other areas of the Republic of Korea and in other countries.

Keywords Spatial distribution model · Ensemble model · Coefficient of variation · Adaptation plans

1 Introduction

Extreme weather events, such as heavy rainfall, typhoons, heat waves, and cold waves, have increased because of climate change and have caused extensive damage in the Republic of Korea (ROK) (Boo et al. 2006; Sung et al. 2012). Landslides caused by heavy rainfall represent one of the worst types of disasters, and local governmental deci-

sion makers are attempting to establish disaster prevention zones in order to reduce landslide damage. These zones restrict construction activities in areas susceptible to landslides and specify safe separation distances for development (Chiou et al. 2015). In addition, government officials of the ROK are attempting to establish climate adaptation plans that will prevent future losses of life and protect properties from landslides.

✉ Dong Kun Lee
dklee7@snu.ac.kr

¹ Department of Human Environment Design, Major in Landscape Urban Planning, Cheongju University, Cheongju 28503, Republic of Korea

² Research Institute of Agriculture Life Science, Seoul National University, Seoul 08826, Republic of Korea

³ Department of Landscape Architecture and Rural System Engineering, Seoul National University, Seoul 08826, Republic of Korea

⁴ Department of Landscape Architecture, University of Seoul, Seoul 02504, Republic of Korea

⁵ Korea Environment Institute, Seoul, Republic of Korea

⁶ Department of Landscape Architecture, Kangwon National University, Chuncheon, Republic of Korea

⁷ National Institute of Ecology, Seocheon, Republic of Korea

⁸ Department of Environmental Science, Policy, and Management, University of California, Berkeley, USA

⁹ CALS, Seoul National University, Seoul 151-921, Korea

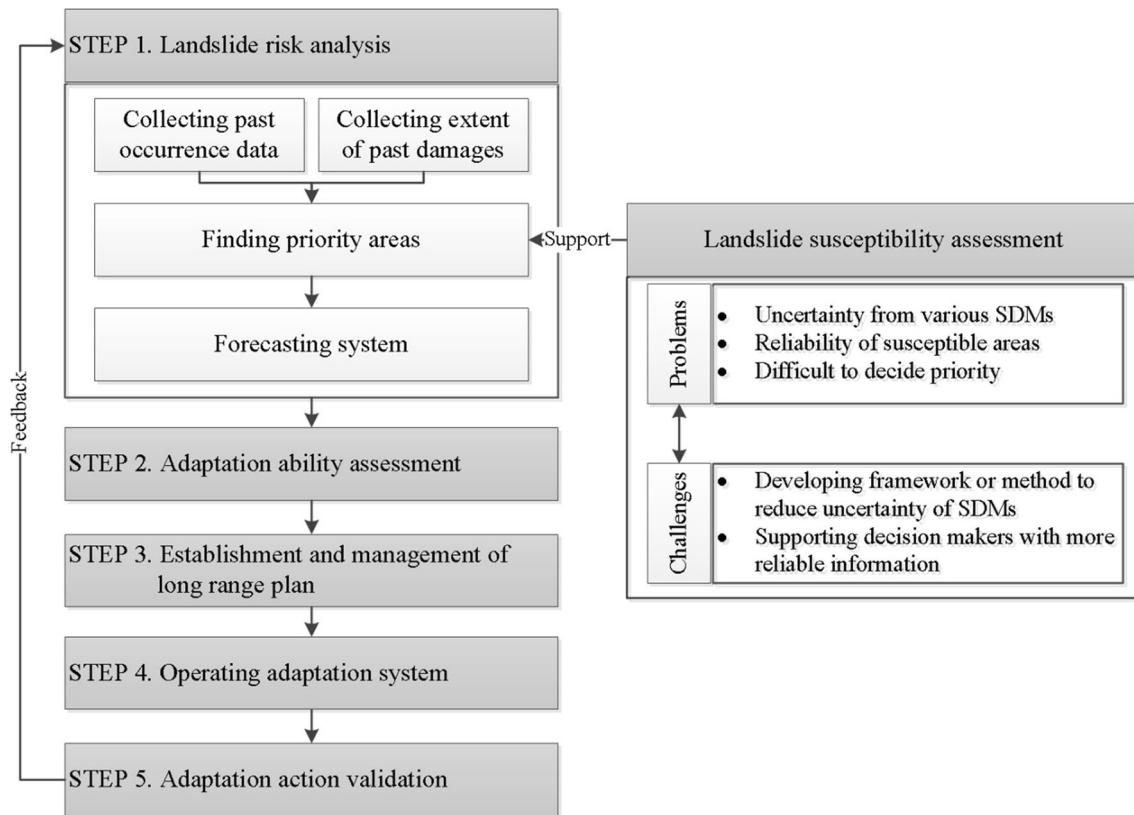


Fig. 1 Challenges of landslide susceptibility assessments used for supporting the decision-making process

Inje-gun in Gangwon-do experienced severe damage due to landslides in 2006 and 2007 (Kim et al. 2011a; Yoo et al. 2012), causing many human injuries and losses of public facilities. Inje-gun consists of 91% mountainous terrain, and various residential and agricultural land cover types (i.e., households, farmlands, and roads) are located adjacent to the forested land (Oh et al. 2009). In this region, alpine agriculture is an important source of income. The local government of Inje-gun is required to develop strategic management plans that will protect lives and properties from landslides.

As a key tool for developing strategic management plans, landslide susceptibility maps are required by Inje-gun's local government (Akgun et al. 2008; Kappes et al. 2012). These maps can identify the areas of forestland that are vulnerable to landslides triggered by extreme rainfall, and they can thus be used to restrict potentially risky human activities in susceptible areas, such as the construction of roads and residential facilities and the planting of crops. Additional development without consideration of the region's landslide susceptibility will leave Inje-gun more vulnerable to future damage. Maps of areas susceptible to landslides are also critical for climate adaptation plans, which decision makers can use to enhance community resilience (Felicísimo et al. 2013).

Uncertainties in landslide susceptibility information can lead to undesirable social costs, such as infringements on private property rights and unnecessary economic investments. With the aim of reducing these uncertainties, this study investigates the amount of uncertainty associated with different types of landslide susceptibility modeling data and works to combine this information in new ways. Figure 1 illustrates the general process that can be used to establish a landslide adaptation plan (Kim 2012). As noted, a landslide susceptibility assessment involves the determination of priority areas in Step 1 in this process, and this is typically accomplished with the use of spatial distribution models (SDMs).

Many previous studies have developed SDMs that can be used to analyze a region's landslide susceptibility (Catani et al. 2005; Yesilnacar and Topal 2005; Akgun et al. 2008; Yilmaz 2009, 2010; Akgun 2012; Torizin 2016). Various models exist for assessing landslide susceptibility. In particular, artificial intelligence methods, such as artificial neural networks (ANNs) and neuro-fuzzy logic, have been applied in recent studies (Bui et al. 2012; Li et al. 2012; Park et al. 2013; Zare et al. 2013; Pradhan 2013; Nourani et al. 2014; Dehnavi et al. 2015; Lee et al. 2015). ANN is computing systems based on the biological neural networks of animal brains. ANN is one of the

machine learning model to calculate spatial distribution of specific event. However, there are two types of uncertainty in landslide susceptibility studies using SDMs. The first type of uncertainty is caused by variables such as topographic factors, soil materials, and vegetation factors (Felicísimo et al. 2013; Lian et al. 2014; Wang et al. 2016b; Tongal and Booij 2017). The second type stems from differences between SDMs (Claessens et al. 2007; Ladle and Hortal 2013; Kwon 2014; Niedzielski and Miziński 2016; Liu et al. 2016).

Ideally, the results of a landslide susceptibility assessment should provide reliable scientific information that can support the identification of priority areas and can be used to forecast susceptible areas (Bonachea et al. 2009; Sudmeier-Rieux et al. 2012). Researchers have proposed a method to minimize uncertainty, which relies on the use of multiple variables and models (Buisson et al. 2010; Miao et al. 2016). However, it is very difficult to establish one optimum model for landslide susceptibility mapping because each SDM has different properties, and their reliability differs due to combinations of variables. In particular, ensemble methods have been employed in remote sensing classification (Clinton et al. 2015; Wang et al. 2016a) and species distribution modeling has been used to minimize the limitations of a single model (Thuiller et al. 2009) and forecast future climate conditions (Krishnamurti et al. 2000; Riebau and Fox 2005; Son et al. 2014; Kovalchuk et al. 2017).

Ensemble methods have also been applied to assess landslide susceptibility. Bartlett and Shawe-Taylor (1999) reduced bias error by applying the theory of large margin classifiers coupled with ensemble techniques. Bühlmann and Yu (2003) developed an ensemble consisting of two different linear regression models. Rokach (2010) reviewed various ensemble methods to achieve improved prediction performances. Recently, Ghosh and Acharya (2011) employed consensus clustering to produce more robust and stable results. Lee and Oh (2012) developed and applied an ensemble method to construct a reliable model by using logistic regression, the frequency ratio, weight of evidence information, and ANN. Althuwaynee et al. (2014) used the bivariate evidential belief function (EBF) as a bivariate to explore the integration validity with an analytic hierarchy process (AHP) and employed logistic regression as a multivariate method for spatial mapping.

In this context, the main objective of this study is to assess the landslide susceptibility in a target region by using ensemble methods based on multiple SDMs and optimum variables. This study addresses three research questions: First, how to select optimal variables for SDMs; second, how to quantify the uncertainty from various SDMs; and third, whether an ensemble model can help decrease the uncertainty of SDMs and effectively support

decision making. These research questions are related to attempts to produce more reliable data concerning areas susceptible to landslides, which can support the decision-making process regarding land development.

2 Methods

2.1 Scope of study

The study site is Inje-gun in Gangwon-do, ROK (Fig. 2). Inje-gun is located in the northeastern region of ROK. The study site experienced severe damage due to landslides in 2006 and 2007. Moreover, alpine agricultural resources, which represent an important industrial output in Inje-gun, have been damaged by landslides every year. Thus, Inje-gun was deemed an appropriate study site to develop landslide susceptibility maps. The spatial resolution of the data for the study area was set to 30 m × 30 m. The temporal scope of the study was set to 2006, considering the reliability of historical landslide occurrence data.

2.2 Landslide occurrence data

Landslide occurrence data are required to establish SDMs in support of the local government of Gangwon-do (Fig. 3). Such data comprise spatial coordinates, the extent of damage in each area, and addresses, all of which were collected by MODIS (the moderate resolution imaging spectroradiometer) satellite imagery, field surveys, and resident reports from July to September in 2006.

This study focuses on landslide areas over 2000 m², in accordance with a previous study on landslide susceptibility models (Kim et al. 2015). In that study, models that considered landslide occurrence areas over 2000 m² exhibited the highest area under the receiver operating characteristic (ROC) curve¹ compared with other models that used different-sized areas. A total of 341 occurrence points were used as input data for landslide susceptibility modeling. Of the landslide occurrence data, 80% were used to train the model, and 20% were utilized to test the model.

¹ An ROC curve is a plot graph that shows the diagnostic ability of a binary classification method considering a threshold value. ROC curves are created by plotting the true positive rate (TPR) against the false positive rate (FPR). The TPR is also called the sensitivity and the FPR is also known as the probability of false alarm, and it can be calculated as (1 – specificity). Thus, the ROC curve represents the sensitivity as a function of FPR. ROC analysis is used as a tool to select optimal models and to discard suboptimal ones. The area under the curve (AUC) is the same as the probability that a classifier will grade a randomly chosen positive case higher than a randomly chosen negative case. The AUC is similar to the Mann–Whitney U, which tests whether positive cases are graded higher than negative cases.

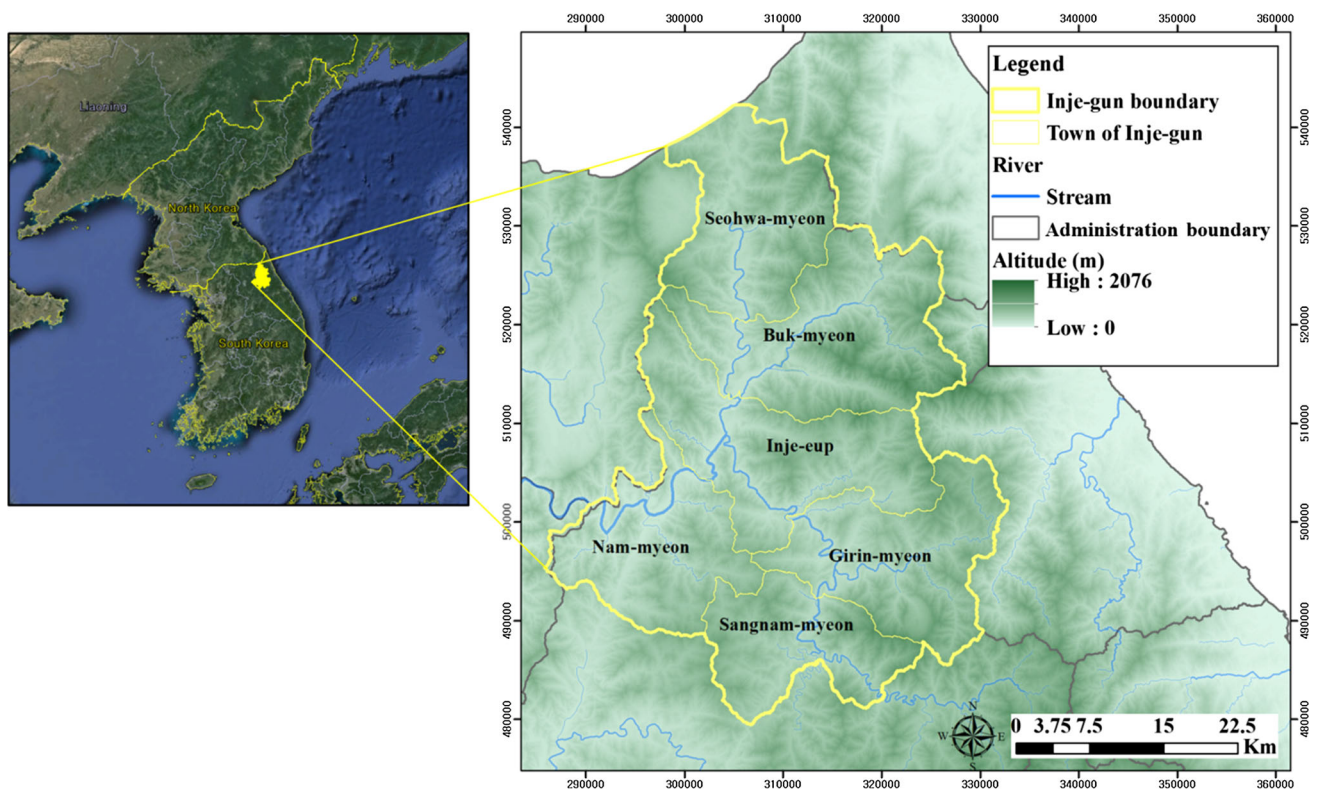


Fig. 2 Study site (Inje-gun, Gangwon-do, ROK)

For the absence data, 700 pseudo-absence points were created by considering a minimum distance of 200 m from occurrence areas. The absence data means that location data of areas with no landslide occurrence. It is required data to run SDMs. However, we do not have data on absence points. Actually, absence points are not important data for the government or relevant research institute. Because they focus on collecting data for landslide occurrence areas or points. However, absence data is required to run SDMs. Thus, we had to utilize pseudo-absence points to run SDMs. The minimum distance between landslide occurrence points and absence points was set to 200 m in consideration of a previous study regarding landslide damage in Inje-gun (Son et al. 2009).

2.3 Modeling and ensemble methods

The process of modeling is shown in Fig. 4. First, we established landslide occurrence data as the dependent variable and independent variables relevant to landslides. Second, we selected models to analyze landslide susceptibility, and ten SDMs were selected as landslide susceptibility models. Third, we undertook pilot modeling with all 13 collected variables in order to find optimal variables. Fourth, we selected 10 optimal variables, considering the variable importance and coefficients between variables.

Through this process, we ran models with optimal variables and generated landslide susceptibility maps, AUC values, and response curves. Fifth, we aggregated the results of the models by using five ensemble methods. Finally, we quantified the uncertainty of the modeling results by calculating the coefficient of variation.² Analytical programs were used, including ArcMap 10.1 and Python 2.7 of ESRI and the Biomod2 package of R studio.

We now provide more detailed information for each process. The second research question involved quantifying the uncertainty of SDMs. Therefore, we used as many different SDMs as possible. Researchers have also used various SDMs to identify areas susceptible to landslides (Catani et al. 2005; Akgun et al. 2008; Yilmaz 2009, 2010; Bui et al. 2012; Li et al. 2012; Park et al. 2013; Zare et al. 2013; Pradhan 2013; Nourani et al. 2014; Dehnavi et al. 2015; Lee et al. 2015). Through a review, we identified 10 SDMs and decided to use all of them. These 10 SDMs were classified into two categories based on the mechanisms of the models, namely, statistically-based models and machine learning-based models (Appendix 1). Each SDM

² Coefficient of variation is also known as relative standard deviation. It is a standardized value of dispersion of a probability distribution. It is calculated by the ratio of the standard deviation to the mean. In this study, it is used for quantifying uncertainty of modeling results.

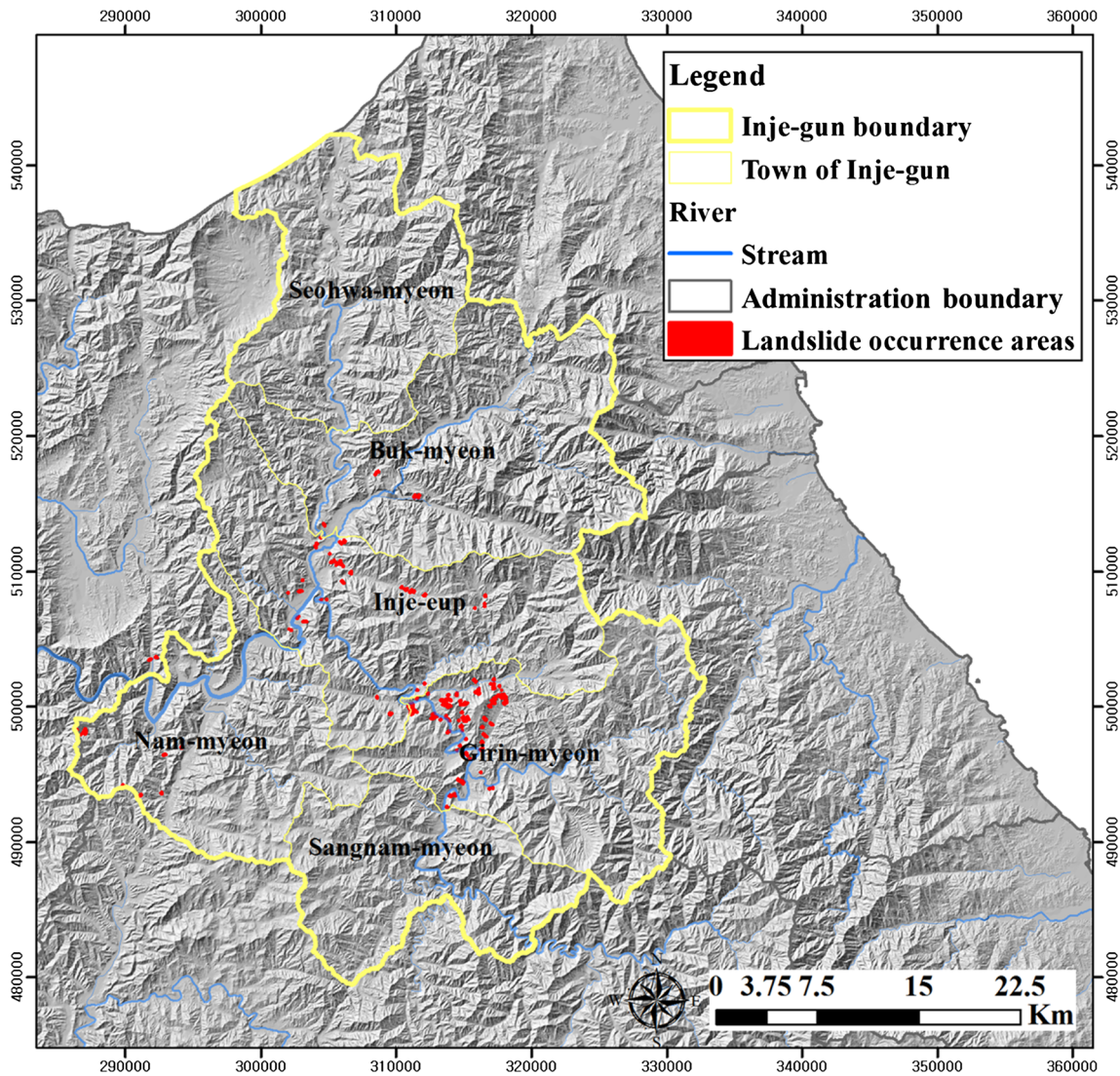


Fig. 3 Landslide occurrence areas in Inje-gun (2006)

had different characteristics in terms of categories, data requirements, and response functions.

We ran a pilot model with 13 relevant variables in order to select optimal variables. The 13 variables were collected by reviewing previous studies (Table 1). Previous studies related to these variables are listed in Table 1. The major factors for landslide models are classified into four categories (climate factors, topography factors, ground material, and vegetation factors).

As a result of running the pilot model to find optimal variables, we obtained the importance of each variable (Table 2). The importance of variables was evaluated using the same principle as the random forest variable importance algorithm. This principle rearranges a single variable of the data and carries out a model prediction using the “rearranged” data set. Then, we calculated the Pearson’s correlation between reference predictions and the

“rearranged” one. The returned value is between 0 and 1. A high value indicates that the variable has a greater influence on the model.

In addition, we analyzed the correlations between variables. Variables that had a high coefficient (0.6) and relatively low importance were excluded from the list. X011_100mm (No. 1) and X012_120mm (No. 2) exhibited high coefficients. We selected X012_120mm by considering the importance of the variables. X021_3days and X022_5days also exhibited high coefficients, and we selected X022_5days because it had a higher importance than X21_3days. Diamclass (No. 13) was excluded because its importance was very low in all SDMs.

Through the above process, we selected 10 optimal variables: X012_120mm (number of days with over 120 mm of rainfall), X022_5days (5 days of maximum rainfall), X060_3day (number of days with over 150 mm

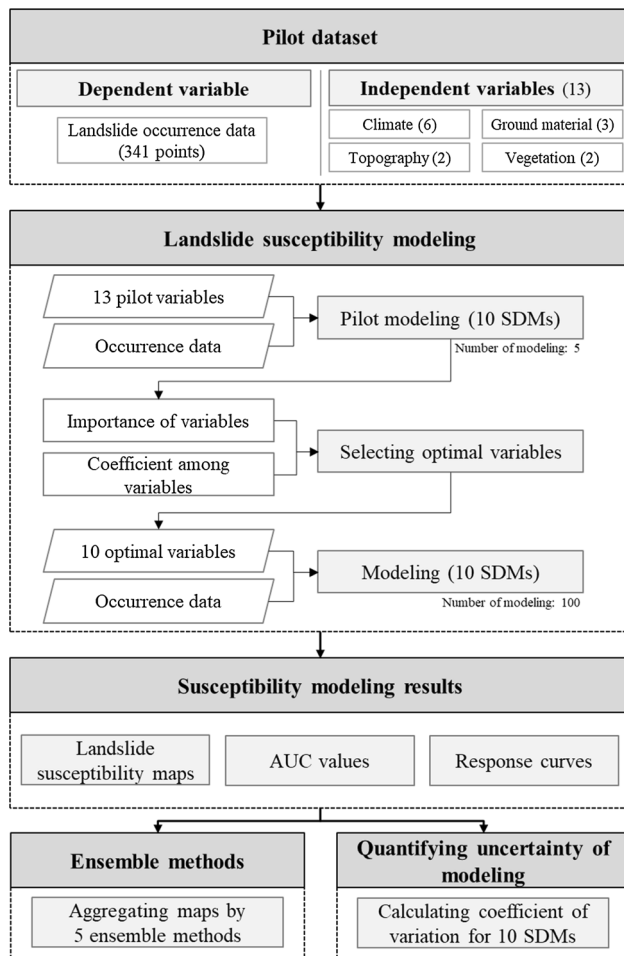


Fig. 4 Flowchart of modeling process

for 3 days of maximum rainfall), X070_dailymax (daily maximum rainfall), altitude, slope, soildepth_km (soil depth), soildrain_km (soil drainage) and soiltype_km (soil type), and ageclass (age of forest). The distribution maps of the variables are presented in Appendix 2.

We ran 10 SDMs with optimal variables. However, SDMs based on machine learning technology do not always produce the same results from the same data. Because machine learning models learn different way with same input data for each run. In this context, the contributions of variables were changed for every run. Thus, we ran each SDM 100 times so that we could consider the differences between the results from the same model. To evaluate the reliability of the results of the SDMs, we selected the ROC method, which is explained in the previous section.

The third research question of this study involved analyzing the effectiveness of ensemble methods. We reviewed various ensemble methods that were based on the concept of minimizing the differences among results of various SDMs (Bartlett and Shawe-Taylor 1999; Bühlmann

and Yu 2003; Dimitriadou et al. 2003; Rokach 2010; Ghosh and Acharya 2011; Lee and Oh 2012; Althuwaynee et al. 2014).

As a result of the review process, five types of representative ensemble methods were selected (Table 3). Each ensemble method utilized different methods to aggregate probabilities or binary values from various SDMs. Specifically, the ensemble methods included (1) PM (mean of probabilities), (2) PCI (confidence interval of the mean of probabilities), (3) PME (median of probabilities), (4) CA (mean of the binary), and (5) PMW (weighted mean based on model performance). Table 3 presents detailed information about these methods with reference to the previous study (Thuiller et al. 2015).

The probabilities of landslide occurrence were calculated for each grid in two dimensions (30 m × 30 m). These probabilities were converted to binary values (0 or 1) by using a cutoff value for each model. If the probability of a certain cell was over the cutoff value, then the cell was considered a landslide-susceptible area.

The first research question, concerning the uncertainties of variables, was considered in order to analyze landslide susceptibility by using the same variables in every SDM. The optimum variables were selected by considering the importance of variables, predictive performance of models (AUC values), and multicollinearity. The selected optimum variables were used to conduct each SDM. We minimized the effect of uncertainty from variables through this process. To quantify the uncertainty between the results of the SDMs, which was the second research question addressed in this work, the CV value of each cell was calculated. The CV value was calculated using the probabilities from the landslide susceptibility maps derived using the 10 SDMs.

3 Results and discussion

3.1 Landslide susceptibility modeling and ensemble methods

As a result of modeling, we obtained the importance of variables, landslide hazard maps, and AUC values. The importance of variables helps understand the contribution of each variable to the model. The hazard maps provide information about dangerous areas susceptible to landslides, and the AUC value represents the standard of a model's reliability.

Table 4 presents each variable's importance for each SDM. We ran each SDM 100 times and obtained 100 importance values for each variable. Thus, we calculated the average importance value for each variable to summarize the results. The "X070_dailymax" (daily maximum rainfall), "altitude," "X022_5 days" (5 days of maximum

Table 1 Input variables for SDMs

Category	Variables	Code name	References	Data references
Climate factor	(1) Number of days with over 100 mm of rainfall	X011_100mm	Kim and Chae (2009), Choi et al. (2011) and Yoo et al. (2012)	Observation data (Korea Meteorological Administration 2011)
	(2) Number of days with over 120 mm of rainfall	X012_120mm		
	(3) 3 days of maximum rainfall (mm)	X021_3days	Guzzetti et al. (2008, Kim and Chae (2009) and Choi et al. (2011)	
	(4) 5 days of maximum rainfall (mm)	X022_5days		
	(5) Number of days with over 150 mm for 3 days of maximum rainfall	X060_3day		
	(6) Daily maximum rainfall (mm)	X070_dailymax		
Topography factor	(7) Slope (degrees)	slope	Ayalew and Yamagishi (2005, Ermini et al. (2005), Kim and Chae (2009), Oh (2010), Choi et al. (2011) and Yeon (2011)	Digital elevation model (DEM) (KME 2008)
	(8) Altitude (m)	altitude		
Ground material	(9) Soil depth	soildepth_km	Pradhan and Lee (2010), Oh (2010) and Yeon (2011)	Soil map (WAMIS 2006)
	(10) Soil drainage	soildrain_km		
	(11) Soil type	soiltype_km		
Vegetation factor	(12) Age of forest	ageclass	Yeon (2011) and Oh (2010)	Map of forest type (KME 2005)
	(13) Class of diameter at breast height	diamclass	Kim et al. (2011a, b) and Yeon (2011)	

Table 2 Importance of variables for selecting optimal variables from pilot modeling

No	Variables	SDMs										Optimal variable ^a
		MAX ENT	CTA	SRE	FDA	MARS	RF	GLM	GBM	GAM	ANN	
1	x011_100mm	0.058	0.057	0.025	0.073	0.058	0.038	0.068	0.041	0.070	0.081	X
2	x012_120mm	0.251	0.117	0.027	0.270	0.239	0.061	0.245	0.048	0.244	0.377	O
3	x021_3days	0.035	0.071	0.036	0.034	0.055	0.033	0.028	0.018	0.029	0.068	X
4	x022_5days	0.138	0.332	0.168	0.153	0.412	0.092	0.141	0.095	0.164	0.115	O
5	x060_3day	0.212	0.376	0.189	0.184	0.134	0.237	0.135	0.084	0.124	0.568	O
6	x070_dailymax	0.477	0.599	0.285	0.643	0.481	0.413	0.606	0.462	0.628	0.632	O
7	slope	0.026	0.015	0.031	0.000	0.000	0.008	0.000	0.001	0.008	0.002	O
8	altitude	0.343	0.307	0.390	0.253	0.303	0.253	0.311	0.311	0.319	0.096	O
9	soildepth_km	0.065	0.102	0.149	0.023	0.000	0.025	0.140	0.051	0.202	0.201	O
10	soildrain_km	0.011	0.000	0.129	0.000	0.000	0.003	0.052	0.001	0.066	0.026	O
11	soiltype_km	0.000	0.000	0.152	0.000	0.000	0.002	0.000	0.000	0.005	0.005	O
12	ageclass	0.017	0.000	0.039	0.000	0.000	0.004	0.000	0.000	0.004	0.013	O
13	diamclass	0.001	0.001	0.002	0.000	0.000	0.000	0.001	0.001	0.002	0.002	X

“O” means that the variable was selected as an input variable. “X” means that the variable was not selected as an input variable

rainfall), and “X060_3day” (number of days with over 150 mm for 3 days of maximum rainfall) variables exhibited higher importance than the other variables according to their average values from the 10 models. The

“ageclass” (age class), “soildrain_km” (soil drainage), and “soiltype_km” (soil type) variables exhibited moderate levels of importance.

Table 3 Ensemble methods to integrate results of SDMs Thuiller et al. (2015)

Abbreviation	Description
PM	Mean of probabilities. The PM ensemble model calculates the mean of probabilities for the selected models
PCI (upper and lower)	Confidence interval. The ensemble model for PCI is the confidence interval for the probability of the mean. This model is a good complement for the probability of means. Two ensemble models are constructed using this model: 1. The upper model (there is less than a $100 \times \text{PCI}/2\%$ chance to obtain probabilities higher than the ones given) 2. The lower model (there is less than a $100 \times \text{PCI}/2\%$ chance to obtain probabilities lower than the ones given)
PME	Median of probabilities. The PME ensemble model is the same as the probability of the median for the selected models. The median is better than the mean for assessing the impacts of outliers
CA	Models committee averaging. The CA ensemble model first transforms the probabilities of selected models into binary values by using the cutoff value of each model. After transformation, the model calculates the average of binary values
PMW	Weighted mean of probabilities. The PMW calculates the relative importance of the weights by using the proportion of evaluation scores. Therefore, the results of “good” models are discriminated from those of “bad” models

Through modeling, 10 binary maps were projected by using the results from the 10 SDMs (Appendix 3). The red colored areas were deemed to be landslide-susceptible areas for every run of each SDM. The 10 SDMs exhibited different spatial patterns for areas susceptible to landslides (Appendix 3). The differences between SDMs were the main reason for the uncertainty in estimating landslide-susceptible areas in the studied area. The SRE method predicted the largest susceptible areas, while RF predicted the smallest susceptible areas among the 10 SDMs. The central area of Inje-gun was identified as a susceptible area in almost all SDMs. However, the detailed locations of susceptible areas were different according to each SDM. Therefore, ensemble methods were required to account for the uncertainty derived from the differences between SDMs.

Average AUC values were calculated for the 10 models (MAXENT, CTA, SRE, MDA, MARS, RF, GLM, GBM, GAM, and ANN) to evaluate the reliability of each model (Table 5). The RF model exhibited the highest AUC value (0.979) among the 10 models. The AUC values of eight models were over 0.9, while the other two models had

values over 0.8. Therefore, every model exhibited good performance according to the criteria used in previous studies (Hansson et al. 2005; Franklin 2009). The ROC plots for each model are presented in Appendix 4. We observed different AUC response curves according to each SDM. Some response curves appeared unstable, but we identified important sections of each variable by running models repeatedly. We could not remove unstable curves because the reason for instability was related to the properties of machine learning models. However, we decreased the uncertainty by increasing the number of times the models were run.

Five ensemble methods were applied to synthesize the results of the 10 SDMs and account for the uncertainty. The ensemble models were also evaluated using the ROC method (Table 6). The PMW method exhibited the highest AUC value among the five ensemble methods. The PM and PCI (upper model) methods exhibited the second highest AUC values. The other methods also had high AUC values; therefore, every ensemble method achieved high reliability.

Five ensemble maps were derived by projecting the results of each ensemble method (Appendix 5). The cutoff

Table 4 Average importance of variables for 100 evaluations

Variables	Average of a hundred evaluations										
	MAX ENT	CTA	SRE	FDA	MARS	RF	GLM	GBM	GAM	ANN	
X012_120mm	0.198	0.159	0.021	0.155	0.175	0.065	0.197	0.043	0.184	0.103	
X022_5days	0.213	0.173	0.182	0.235	0.208	0.187	0.133	0.131	0.144	0.586	
X060_3day	0.167	0.316	0.163	0.049	0.056	0.266	0.064	0.063	0.064	0.360	
X070_dailymax	0.500	0.516	0.231	0.645	0.590	0.434	0.644	0.450	0.631	0.366	
ageclass	0.038	0.026	0.036	0.001	0.005	0.023	0.008	0.002	0.029	0.027	
altitude	0.345	0.450	0.330	0.411	0.401	0.279	0.401	0.403	0.354	0.417	
slope	0.088	0.023	0.064	0.001	0.001	0.011	0.002	0.002	0.004	0.036	
soildepth_km	0.055	0.002	0.147	0.001	0.002	0.006	0.005	0.002	0.059	0.005	
soildrain_km	0.001	0.002	0.103	0.001	0.003	0.002	0.003	0.000	0.039	0.007	
soiltype_km	0.034	0.002	0.139	0.000	0.000	0.003	0.003	0.000	0.013	0.010	

value for each ensemble method was used to make a binary map. All maps showed similar susceptible areas (violet color) in the central area of Inje-gun. However, the detailed locations of susceptible areas were different according to each ensemble method.

In particular, the CA method showed larger susceptible areas than the other ensemble methods. The CA method's larger estimates resulted from the technique of calculating the average of the binary values for the 10 SDMs and then using that average value to identify susceptible areas. Other ensemble methods calculated the probabilities of the 10 SDMs to determine susceptible areas. The CA method could provide a good ensemble method for decision makers who want more stable data than the data produced by the other methods. In this study, the PMW method was selected as the optimal ensemble method to analyze areas susceptible to landslides after considering the results of the evaluation (i.e., the AUC values).

The optimal ensemble model (PMW) produced a higher AUC value than any other single model. The extent of landslide-susceptible areas with the PMW model was 30,290 ha. Meanwhile, the extent of susceptible areas determined using the SRE model (which had the lowest AUC among the 10 single models) was 48,992 ha, and that of the RF single model (which had the highest AUC among the 10 single models) was 24,359 ha. The single model SRE resulted in large susceptible areas and the single model RF predicted small susceptible areas compared to the optimal ensemble model. Thus, the ensemble model was helpful for decreasing the differences between the single models according to the AUC values and extents of susceptible areas.

Table 5 Results of the evaluation of the landslide models

Model	AUC value (average of 100 run)	Rank
MAXENT	0.861	9
CTA	0.925	6
SRE	0.812	10
FDA	0.929	4
MARS	0.930	3
RF	0.979	1
GLM	0.922	7
GBM	0.959	2
GAM	0.926	5
ANN	0.908	8

3.2 Quantifying the uncertainty of models

The CV map was classified by considering the standard deviation values (Fig. 5). The northern and southern areas of Inje-gun exhibited high CV values; conversely, the central areas of Inje-gun exhibited low CV values. This pattern was predictable, considering that the ensemble maps showed similar susceptible areas in the central study areas.

However, there were also areas that had high CV values in the central areas at more detailed scales. Therefore, decision makers can utilize the CV map to view susceptible areas in more detail. In general, the CV map provided a more robust basis to judge landslide-susceptible areas of Inje-gun along with the ensemble map. Meanwhile, the northern areas of the CV map showed no information because of the off-limit military area adjacent to North Korea. Meanwhile, the CV map could not consider uncertainty resulting from variables. Uncertainties from variables were limited by using the same input variables for each SDM.

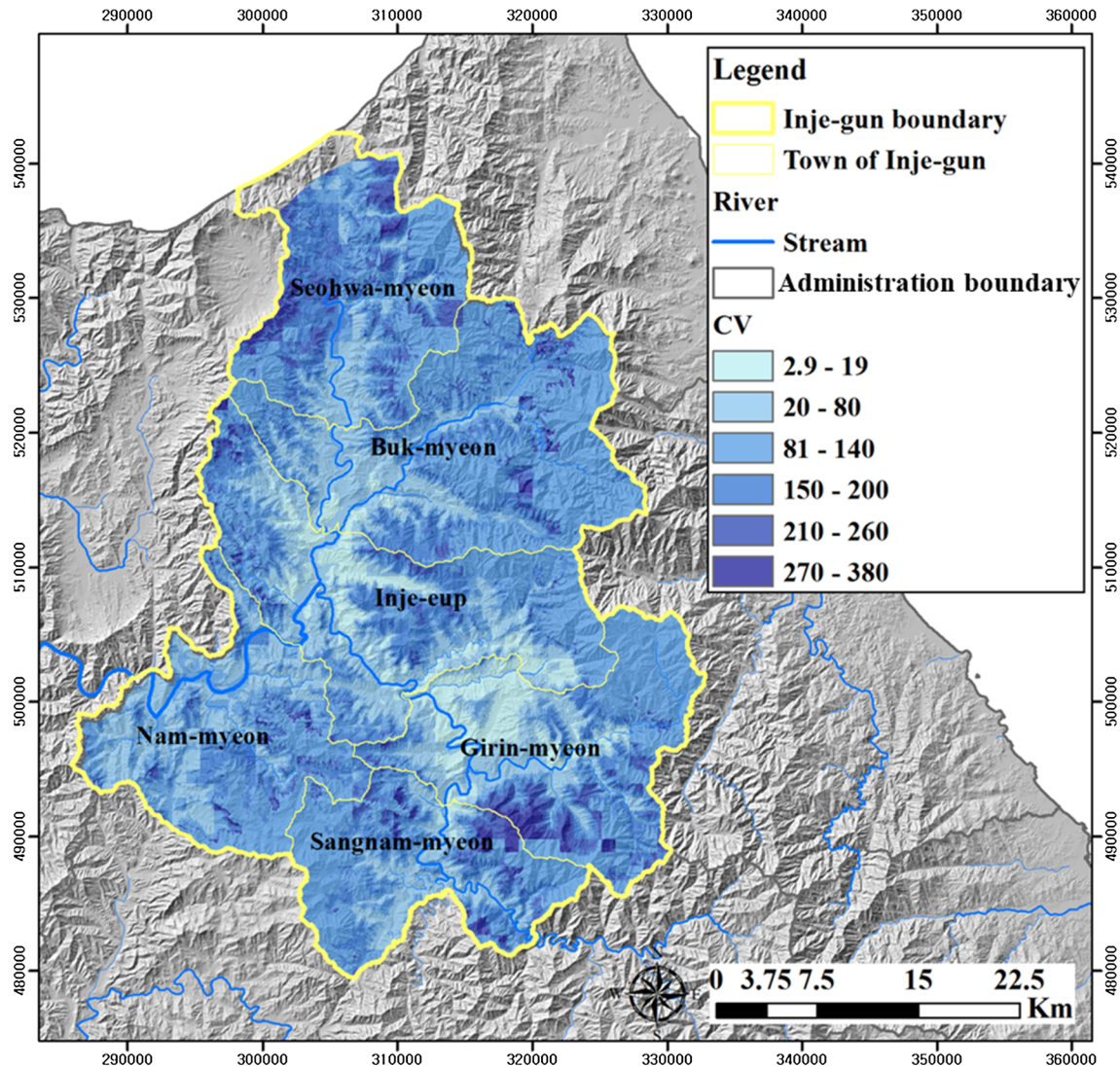
The third research question involved identifying the effectiveness of the ensemble model for reducing the uncertainty from SDMs. This study used two types of matrices to analyze the relationship between the PMW ensemble map (optimal ensemble model) and the CV map (uncertainty map). The first type of matrix was constructed using the probabilities of landslides (PMW ensemble map) and the uncertainty (Fig. 6). The probability of a landslide was classified into five grades (1–5), and the uncertainty was classified into six grades (10–60). The classification method for both maps used the standard deviation. The matrix created a cross table to identify the effectiveness of the ensemble map in terms of reducing the uncertainty from the various SDMs. The values of 15 in the matrix table indicated cells that had a probability of landslide with a low uncertainty. Meanwhile, a value of 61 in the matrix table indicated that a cell had a high uncertainty with a low probability of a landslide.

The map in Fig. 6 shows areas of two key types, namely, areas with high probabilities of a landslide and low uncertainty (14, 15) and areas with low probabilities of a landslide and low uncertainty (21, 22). It was difficult to identify areas with high probabilities for landslides and high uncertainty (65, 55). However, we identified small areas with low probabilities and high uncertainty (51, 61). This can provide a basis for using the ensemble method to help evaluate the uncertainty from various SDMs.

The second type of matrix was constructed using landslide-susceptible areas (binary map of the PMW ensemble model) and the CV map (Fig. 7). Figure 7 also clearly shows the low uncertainty regions of unsusceptible areas

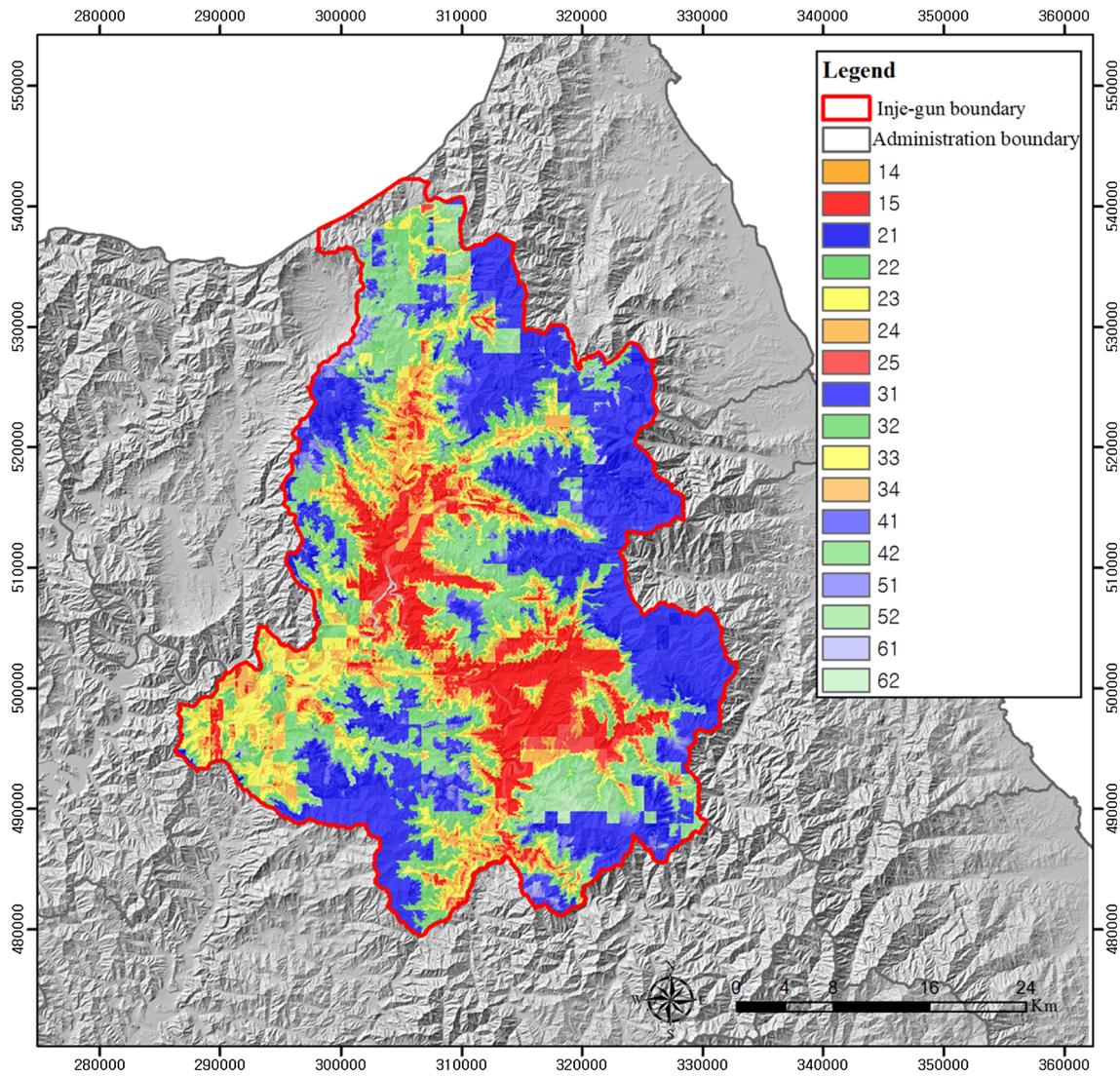
Table 6 Results of the evaluation of the ensemble models

Abbreviation	Cutoff	Sensitivity	Specificity	AUC	Rank
PM	568.500	93.200	96.149	0.986	2
PCI (lower model)	463.500	93.600	95.764	0.985	3
PCI (upper model)	668.000	93.200	96.021	0.986	2
PME	708.000	92.400	96.277	0.982	4
CA	470.000	95.600	92.426	0.985	3
PMW	546.000	94.400	96.149	0.988	1

**Fig. 5** Coefficient of variation (CV) map (uncertainty map). The high CV value mean high uncertainty of modeling results

(11, 21). It was easier to identify landslide hazard areas using this matrix than the first matrix map. This map indicated low uncertainty regions associated with susceptible areas much more clearly than Fig. 6. Thus, this ensemble map can provide decision makers with reliable information regarding landslide hazard areas with low uncertainty.

In general, decision-makers and policy makers need reliable and detailed information to identify priority areas and allocate resources to those areas. In this respect, the map of the relationship between susceptible areas and the uncertainty could be utilized to determine the most urgent areas using credible data. Thus, the ensemble model has strong potential for assessing landslide-susceptible areas.



Probability of landslide (Ensemble map)	High	5	15	25	35	45	55	65
		4	14	24	34	44	54	64
	Low	3	13	23	33	43	53	63
		2	12	22	32	42	52	62
		1	11	21	31	41	51	61
Matrix		10	20	30	40	50	60	
		Low ↔ High						
		Uncertainty (Coefficient of variation, CV)						

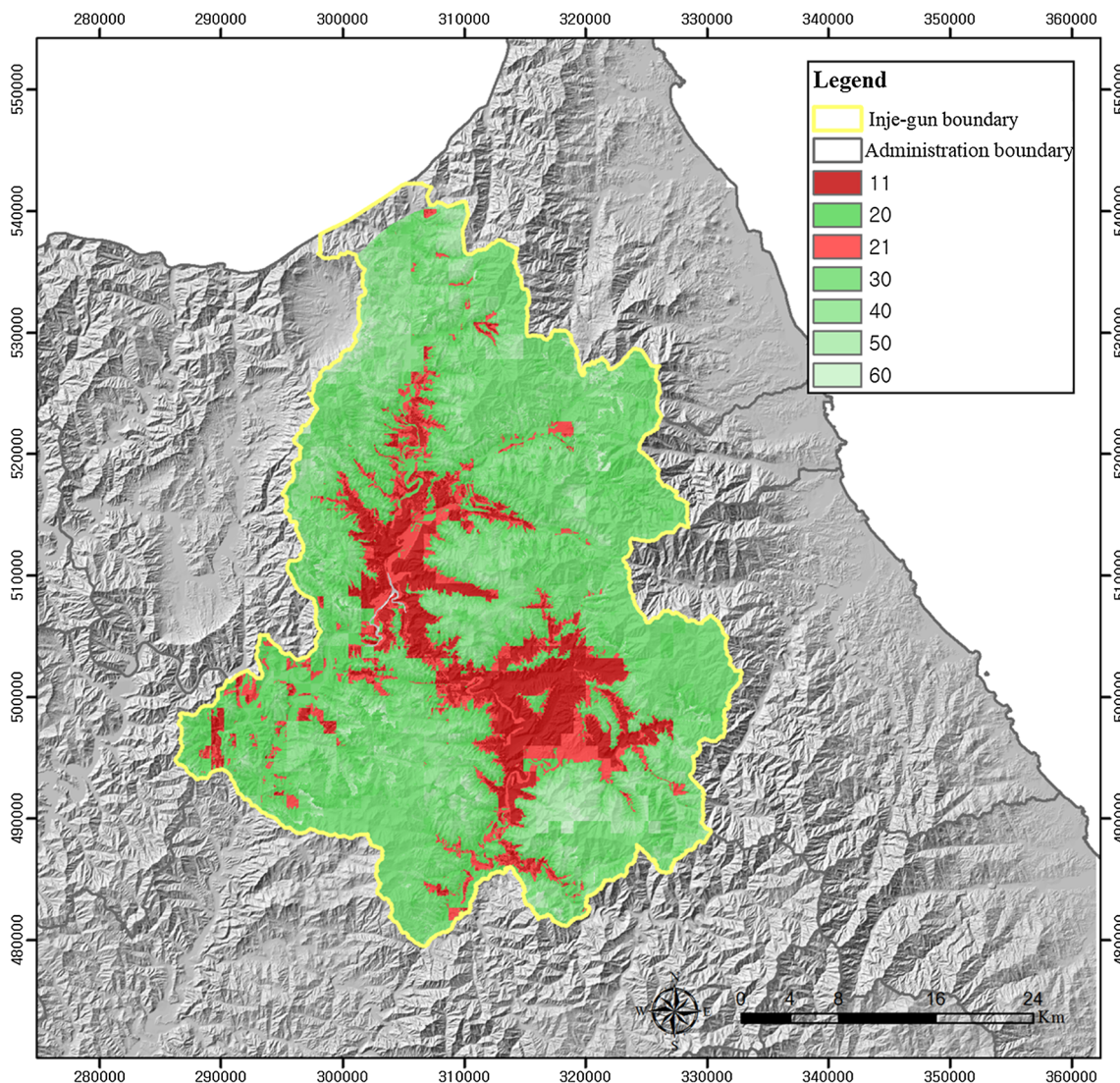
Fig. 6 Relationship between the probability of a landslide from the PMW ensemble model and uncertainty. Areas with high probabilities of a landslide and low uncertainty (14, 15) and areas with low

probabilities of a landslide and low uncertainty (21, 22) are key information for decision making of adaptation plan

In summary, an ensemble map can help minimize uncertainties among SDMs and effectively support decision makers.

4 Conclusion

As a result of landslide susceptibility modeling, susceptible areas for each single model were derived as spatial maps. An interesting finding was that the susceptible areas



Landslide susceptibility (Ensemble map)	Susceptible area	1	11	21	31	41	51	61
	No susceptibility	0	10	20	30	40	50	60
Matrix			10	20	30	40	50	60
			Low ↔ High					
			Uncertainty (Coefficient of variation, CV)					

Fig. 7 Relationship between landslide-susceptible areas from the PMW ensemble model and uncertainty. It shows easier information than Fig. 6 by simple classification of landslide susceptibility

exhibited similar spatial patterns, but detailed patterns differed according to each SDM. This uncertainty from single SDMs causes problems for decision makers.

In ensemble modeling, PMW was selected as the optimal ensemble model among five ensemble methods. Meanwhile, the CV was calculated to quantify the uncertainties among the SDMs. The PMW exhibited better

performance than each single model. In particular, the estimation areas from PMW were moderate compared with those from RF (which had the highest AUC among the single models) and SRE (which had the lowest AUC among the single models). This provides one method of evaluating the reliability of an ensemble model.

The relationship between the landslide susceptibility (ensemble map) and uncertainty (CV map) illustrates the reliability of the ensemble model data. Susceptible areas with low uncertainty provide useful information for protection efforts, and unsusceptible areas with low uncertainty can be prioritized for future safe development by policy makers. In this study, the ensemble model exhibited better performance than any single model. We hope that these results will be useful for local adaptation plans.

However, future work is required to improve variables, considerations of uncertainty, and the reliability of ensemble models. First, we did not select variables by conducting a field survey because the study site was too large to collect field data. Thus, we used variables which were selected by a national institute by considering the reliability of data. It is necessary for future studies to improve our variables by considering the geological and structural characteristics of soils. In addition, we calculated extreme rainfall data by using daily rainfall data from July 2006. If we used hourly data, the accuracy of the model could be improved. Second, landslide time series data are needed to design better models. This study used only one year's worth of data (from 2006). Third, estimations of future landslide-susceptible areas under different climate change scenarios are required. This study focused only on past landslide susceptibilities. However, the extent of susceptibility for an area can change in the future.

A total of 91% of the area in Inje-gun consists of forested lands; therefore, the local authorities of Inje-gun must establish suitable adaptation plans for landslides. Reliable data regarding landslide susceptibilities are critical for these planning efforts. Furthermore, there are many other areas that are vulnerable to landslides in the ROK and other countries, and the framework used here is applicable to these areas as well. In this context, this study illustrates how to estimate more reliable landslide susceptibility data using various SDMs and an ensemble model. The results from this approach can help policy decision makers to better design adaptation plans in order to minimize landslide damage.

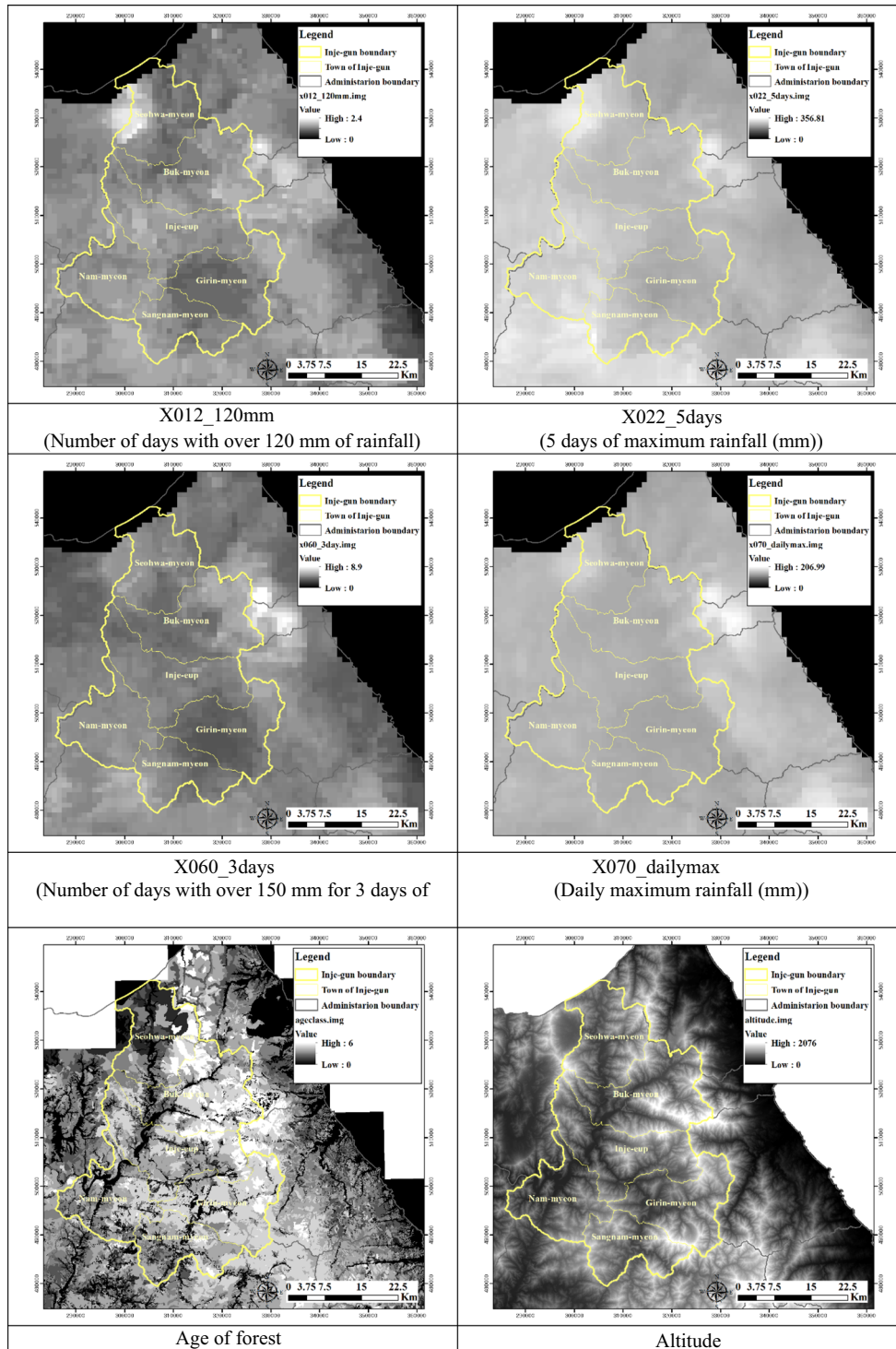
Funding This work was supported by Korea Ministry of Environment (MOE, Project No. 2016000210004) as “Public Technology Program based on Environmental Policy” and the BK 21 Plus Project in 2015 (Seoul National University Interdisciplinary Program in Landscape Architecture, Global Leadership Program toward Innovative Green Infrastructure).

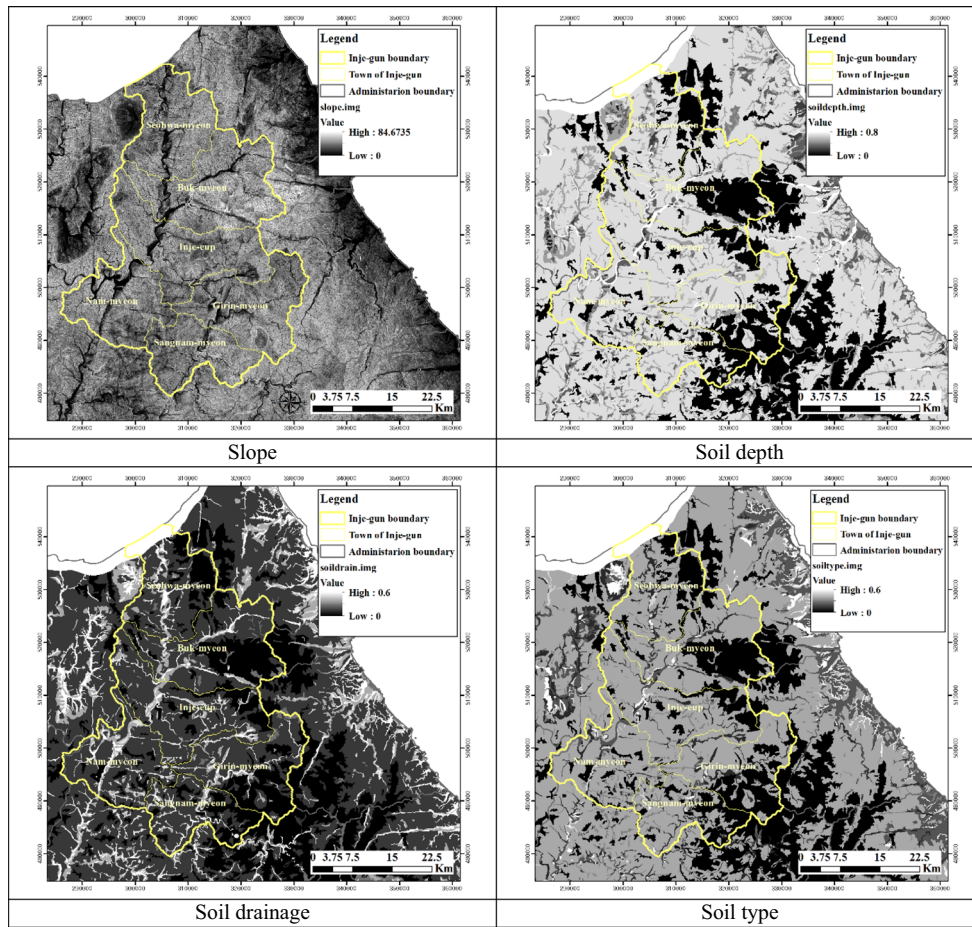
Appendix 1: Key features of the 10 SDMs (Franklin 2009)

Model	Full name of model	Category	Occurrence data required	Response function	Features of the model
GLM	Generalized linear model	Statistically based model	Occurrence/No occurrence	Parametric linear, polynomial, piecewise, interaction terms	GLMs are a representative model among SDMs. GLMs are a generalization of the multiple regression model that uses the link function to accommodate response variables that are distributed normally, namely, the response distributions
GAM	Generalized additive model	Statistically based model	Occurrence/No occurrence	Smoothing function, estimated using local regression, splines or other methods	GAMs in SDMs are suggested as a powerful methodology to detect and describe non-linear response functions. The results of GAMs can be used to build a parametric model
MARS	Multivariate adaptive regression splines	Statistically based model	Occurrence/No occurrence	Adaptive piecewise linear regression combines splines and binary recursive partitioning	MARS can give a type of a generalization of a stepwise linear regression. MARS are suited to analyses with large numbers of variables or a modification of the regression tree method
GBM	Generalized boosted regression model	Machine learning based model	Occurrence/No occurrence	Weighted and unweighted model averaging applied to decision trees	GBMs are similar to weighting variables that consider higher probabilities of selection, instead of weighting equal probabilities for subsequent variables
CTA	Classification tree analysis	Machine learning based model	Occurrence/No occurrence	Divisive, monothetic decision rules (thresholds) from binary recursive partitioning	The goal of CTA is to divide data into homogeneous subgroups. The subgroups consist of variables that have similar values or are in the same class in regard to the ranges of values for the variables

Model	Full name of model	Category	Occurrence data required	Response function	Features of the model
ANN	Artificial neural network	Machine learning based model	Occurrence/ No occurrence	Non-linear decision boundaries in covariate space	ANN can be described as a two-stage classification or regression model. A hidden layer of ANN comprises features that are linear combinations of input variables. The output variable is a weighted combination of features in the hidden layer
SRE	Rectilinear envelope similar to BIOCLIM	Machine learning based model	Occurrence only	Fuzzy classification approach	SRE is a boxcar or parallelepiped classifier that uses BIOCLIM. SRE assesses the potential distribution of the dependent variable by using the multi-dimensional environmental space bounded by the values for all dependent variables
MDA	Mixture discriminant analysis	Machine learning based model	Occurrence/ No occurrence	Linear	MDA is a type of linear discriminant analysis that models the multivariate density of variables by using a mixture of multivariate normal distributions
RF	Random forest	Machine learning based model	Occurrence/ No occurrence	Weighted and unweighted model averaging applied to decision trees	Random forests is a type of bootstrap aggregating method that builds de-correlated trees and averages the trees. Many trees are constructed with subsets of input data. Furthermore, each division of the tree model is also constructed with a random subset of input variables
MAXENT	Maximum entropy algorithm	Machine learning based model	Occurrence only	Non-linear response functions can be described	Maximum entropy is based on statistical mechanics and information theory. MAXENT can analyze the best approximation of an unknown distribution by using the maximum entropy method, which considers the most spread out and closest to uniform values

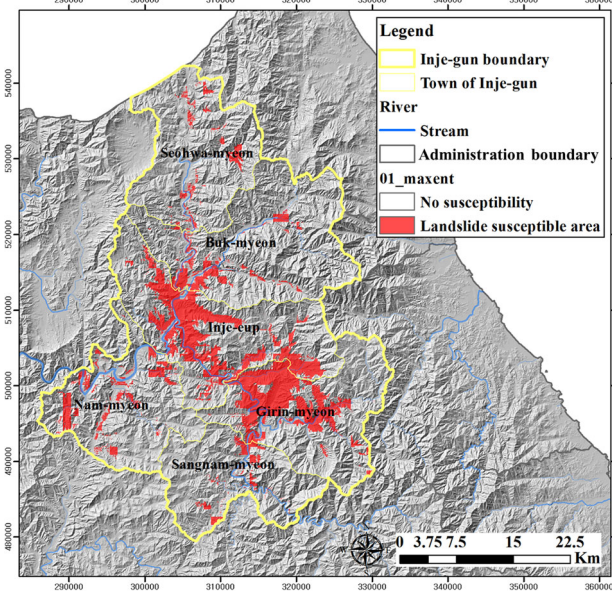
Appendix 2: Maps of variables for landslide susceptibility model



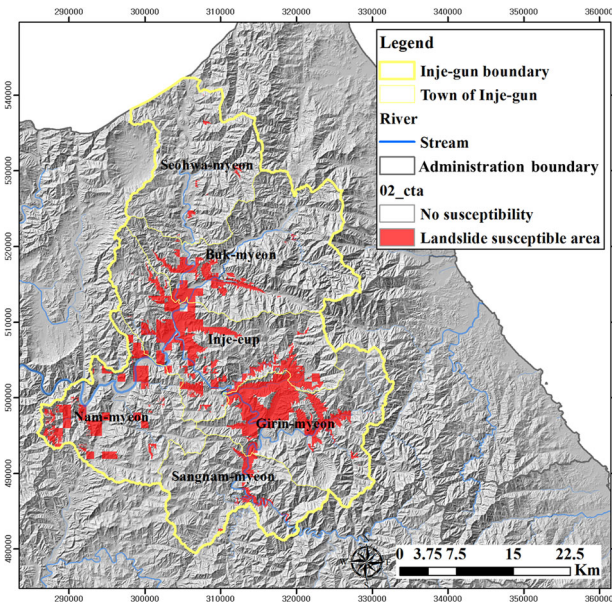


Appendix 3: Landslide projections of the 10 models for present conditions

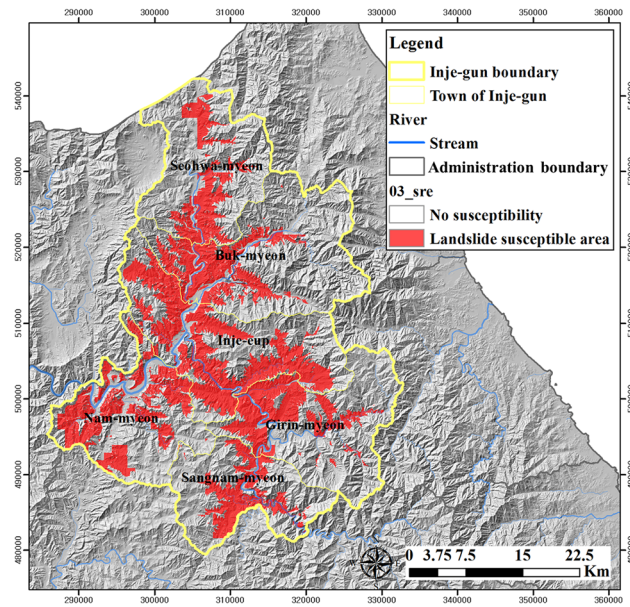
1. MAXENT



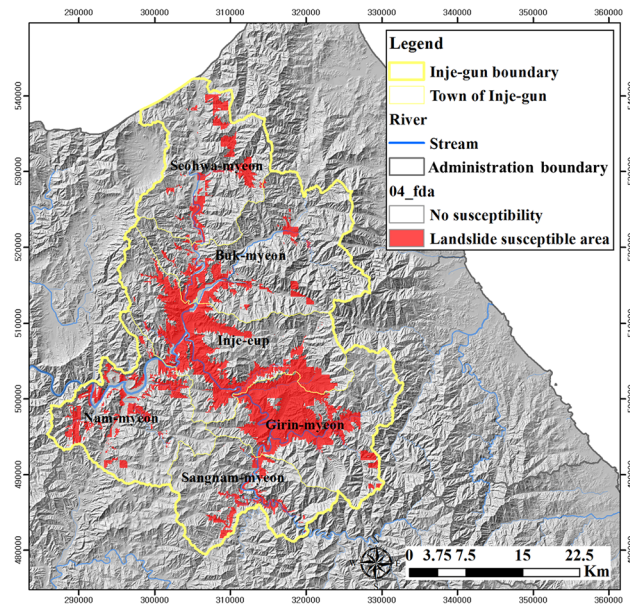
2. CTA



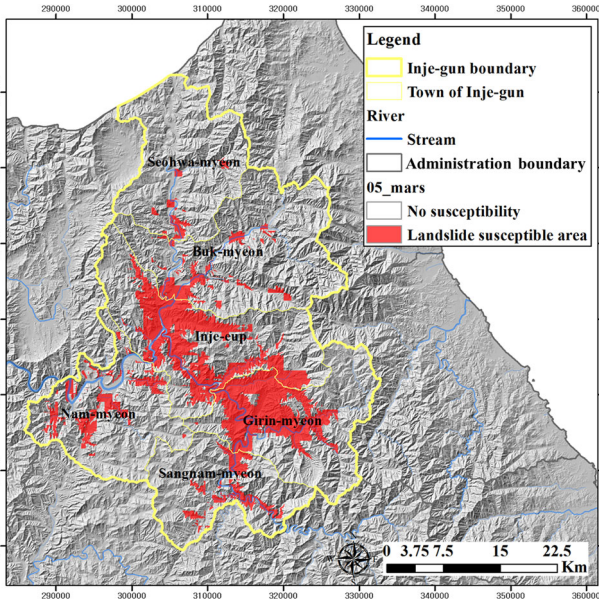
3. SRE



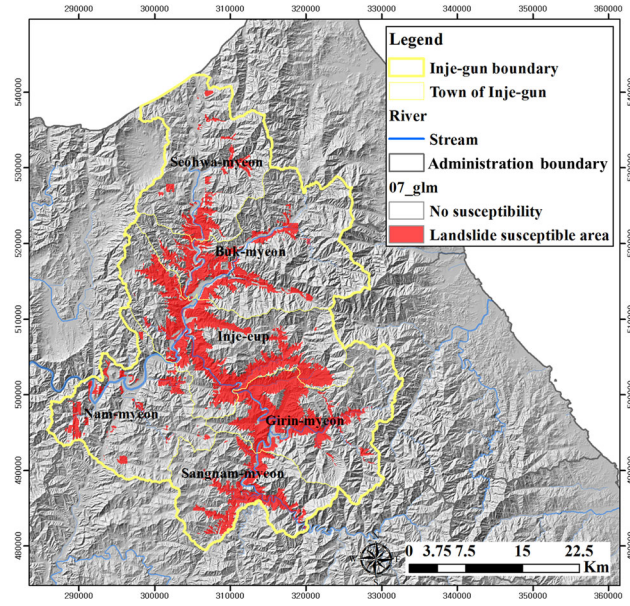
4. FDA



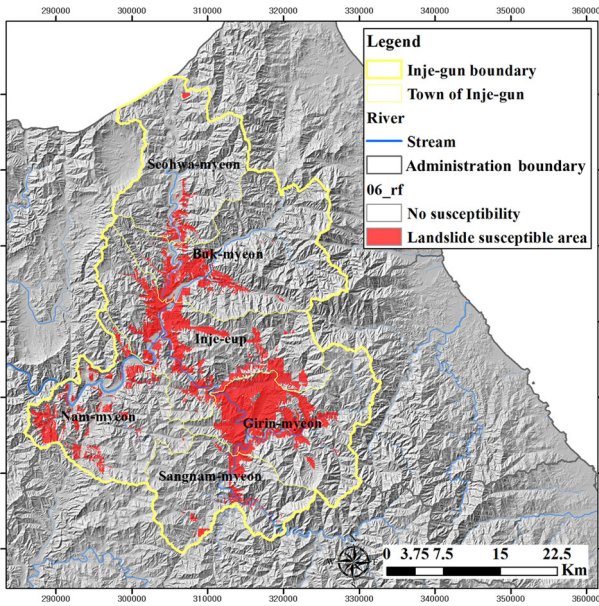
5. MARS



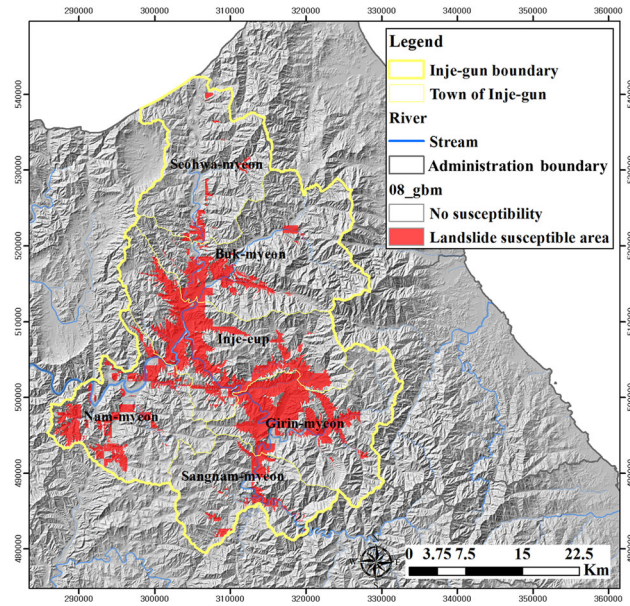
7. GLM



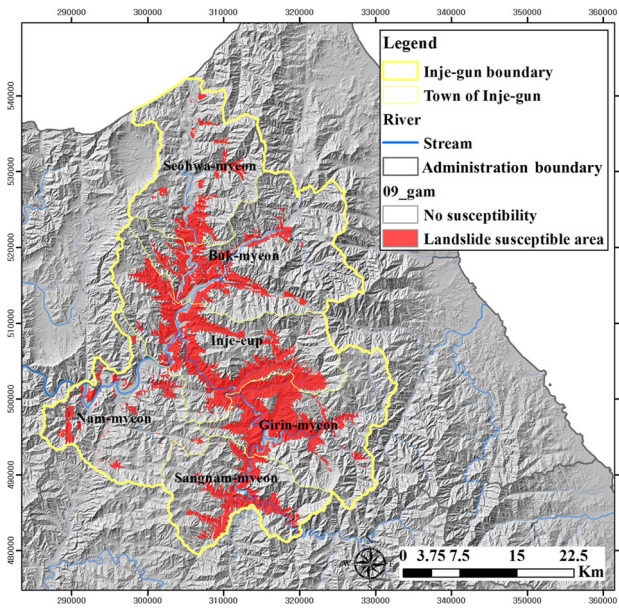
6. RF



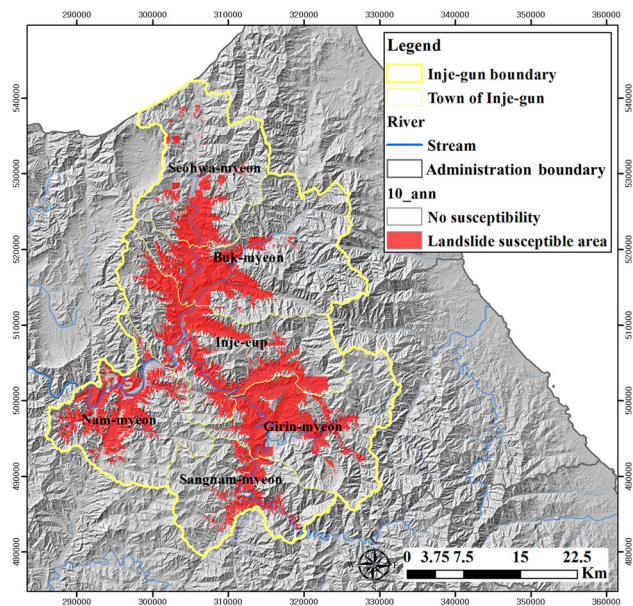
8. GBM



9. GAM



10. ANN

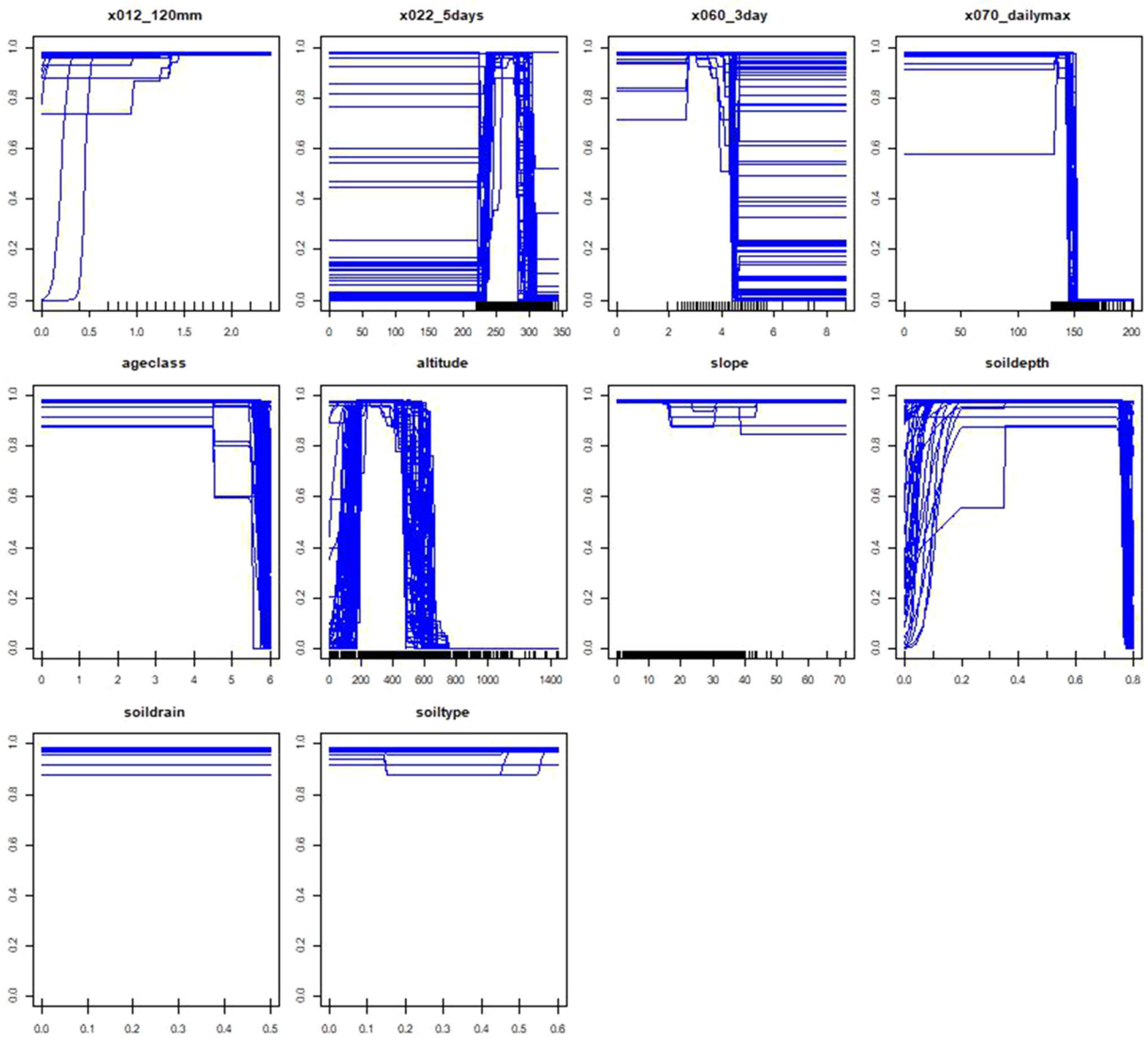


Appendix 4: ROC plots for each model

Axis	Value	Scope
X	Value of each variable	The scope of value varies on variables. Please see the Table 1
Y	Probability of landslide occurrence	0–1

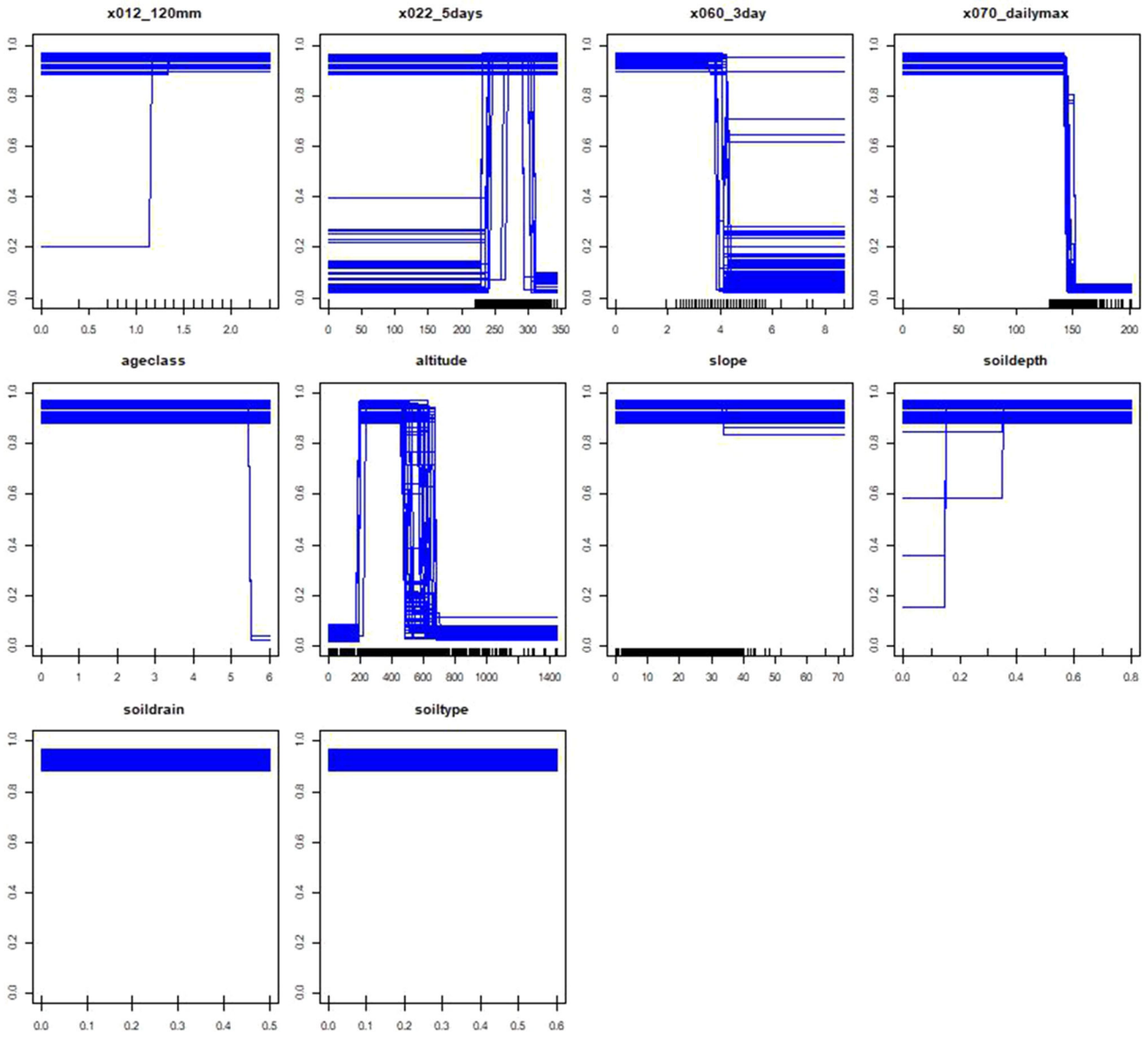
1. MAXENT

Response curves for landslide's MAXENT:Phillips



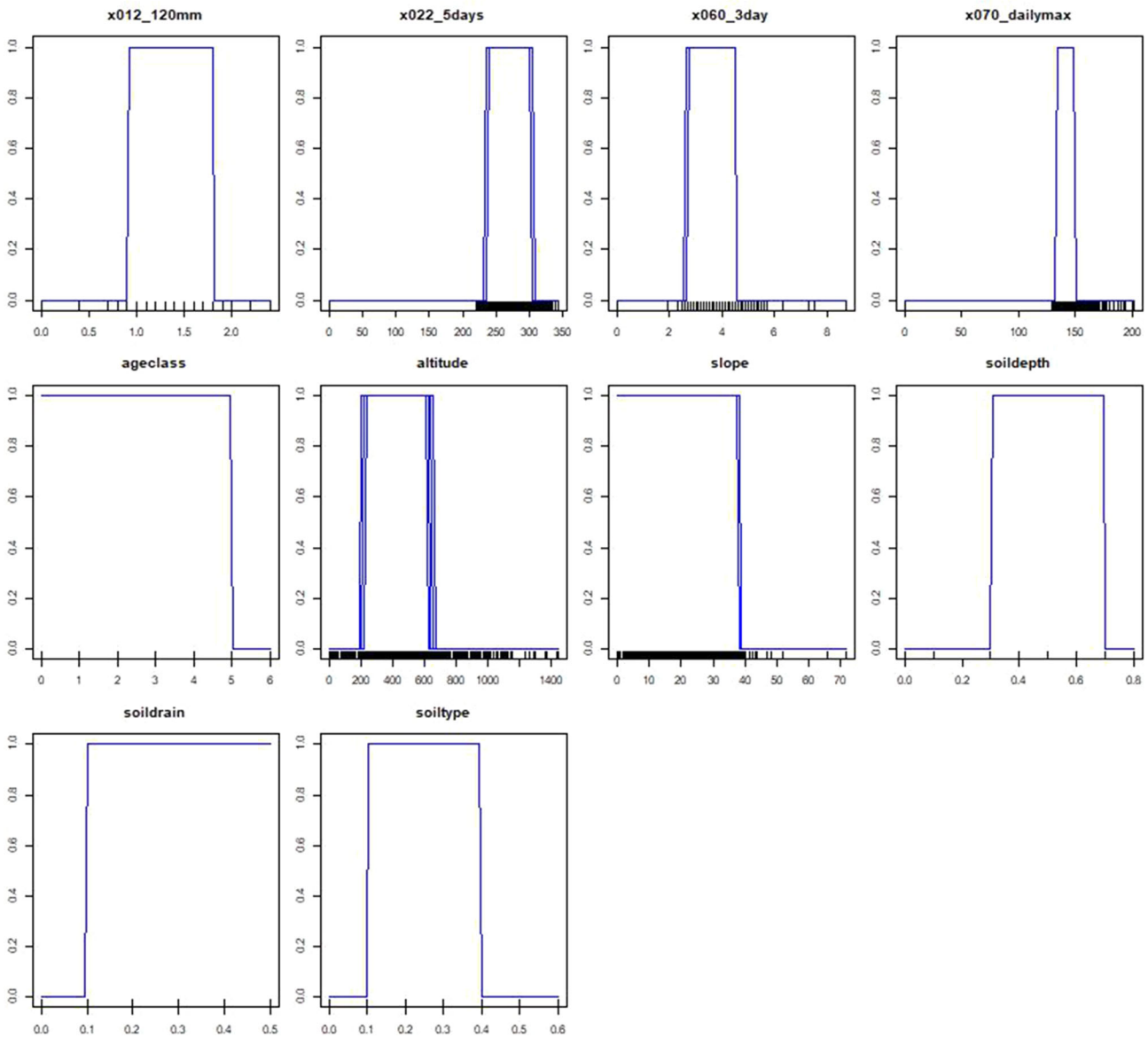
2. CTA

Response curves for landslide's CTA



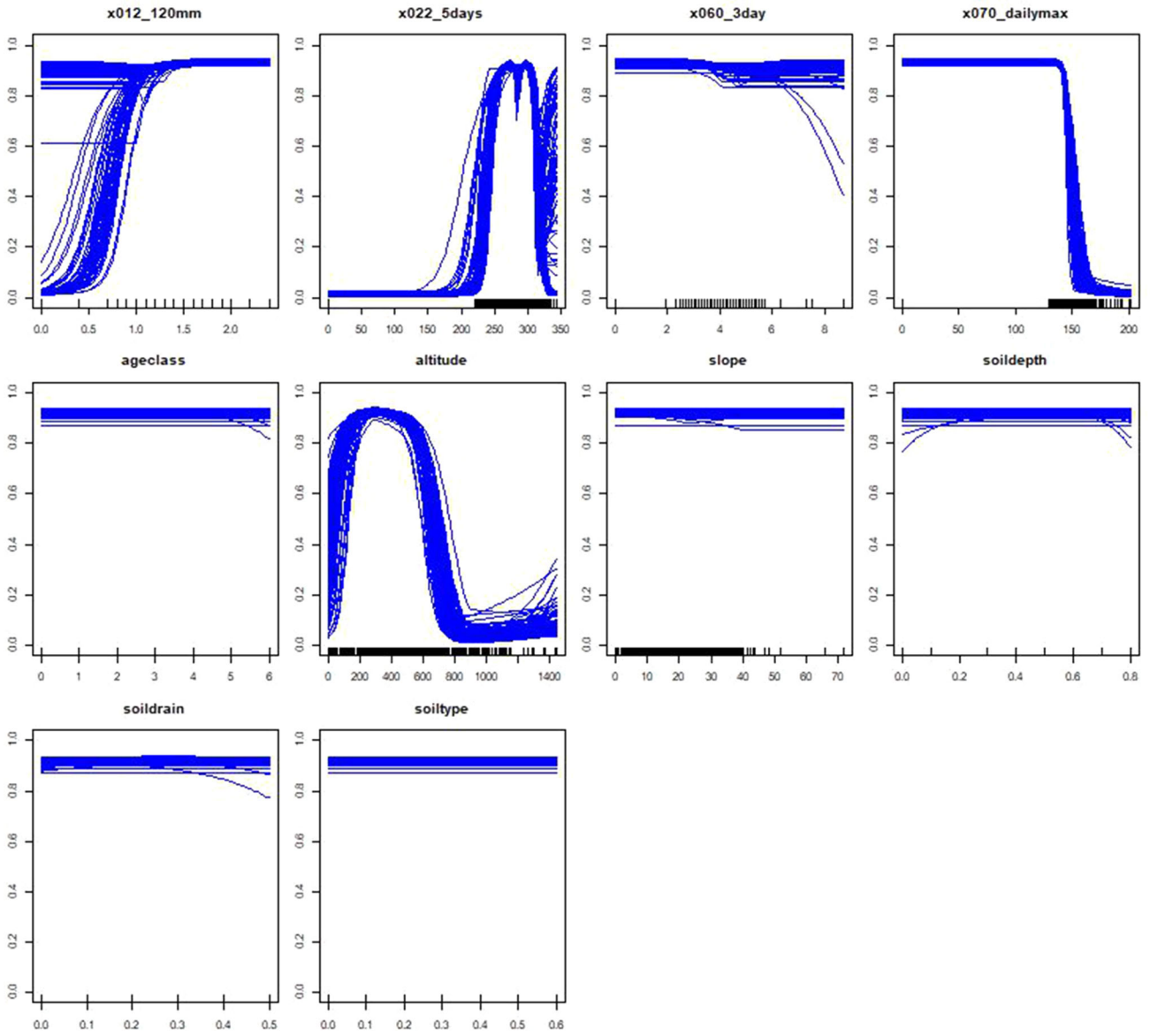
3. SRE

Response curves for landslide's SRE



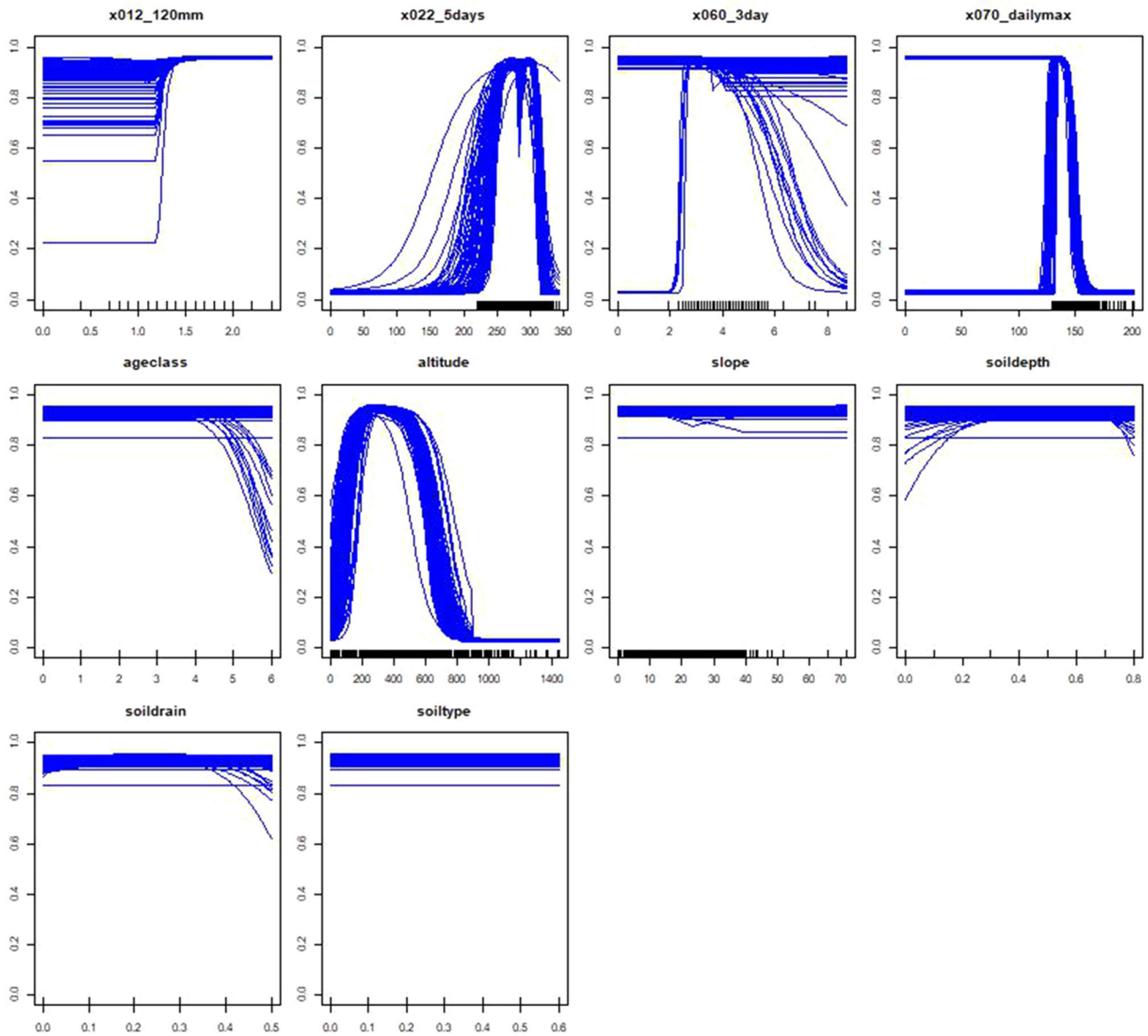
4. FDA

Response curves for landslide's FDA



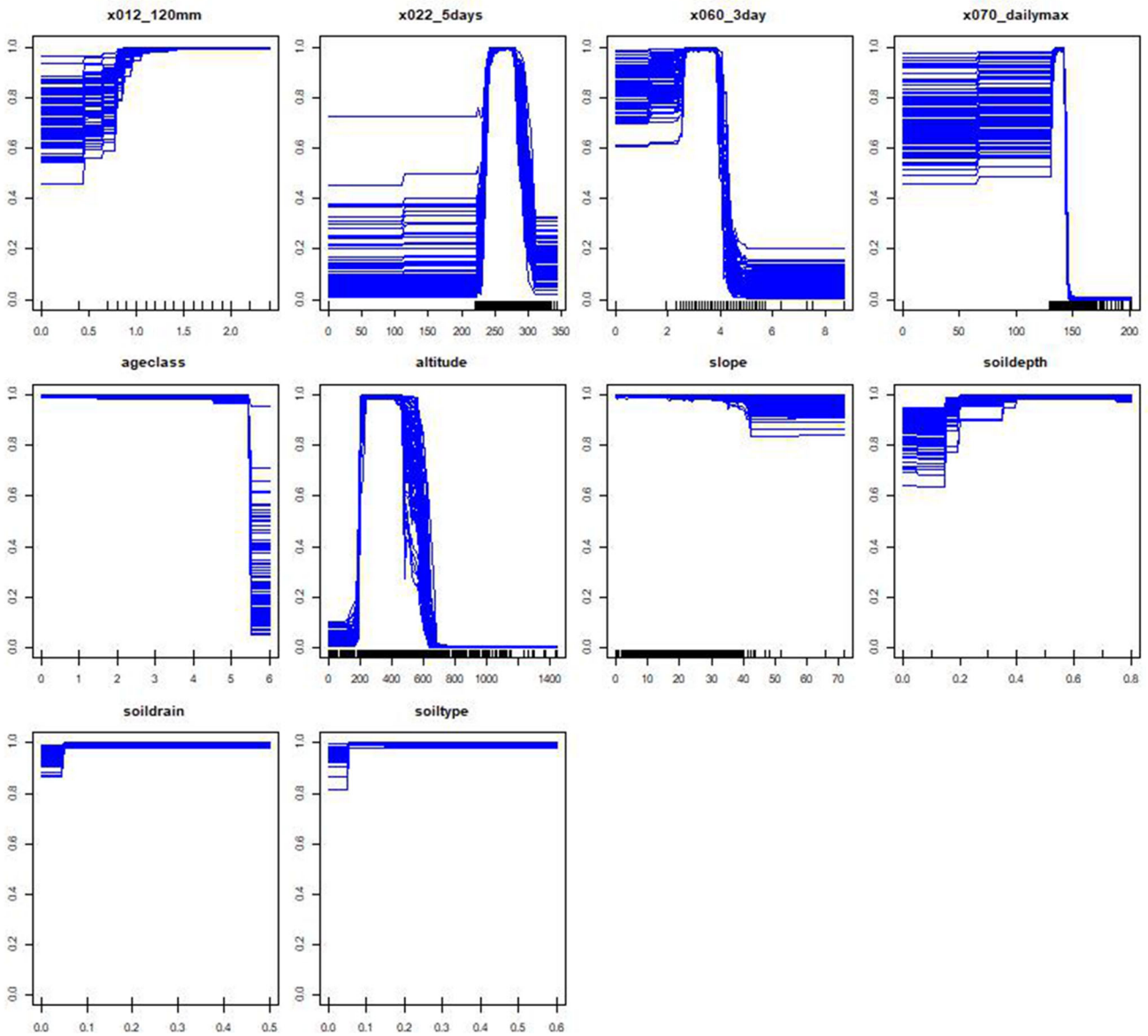
5. MARS

Response curves for landslide's MARS



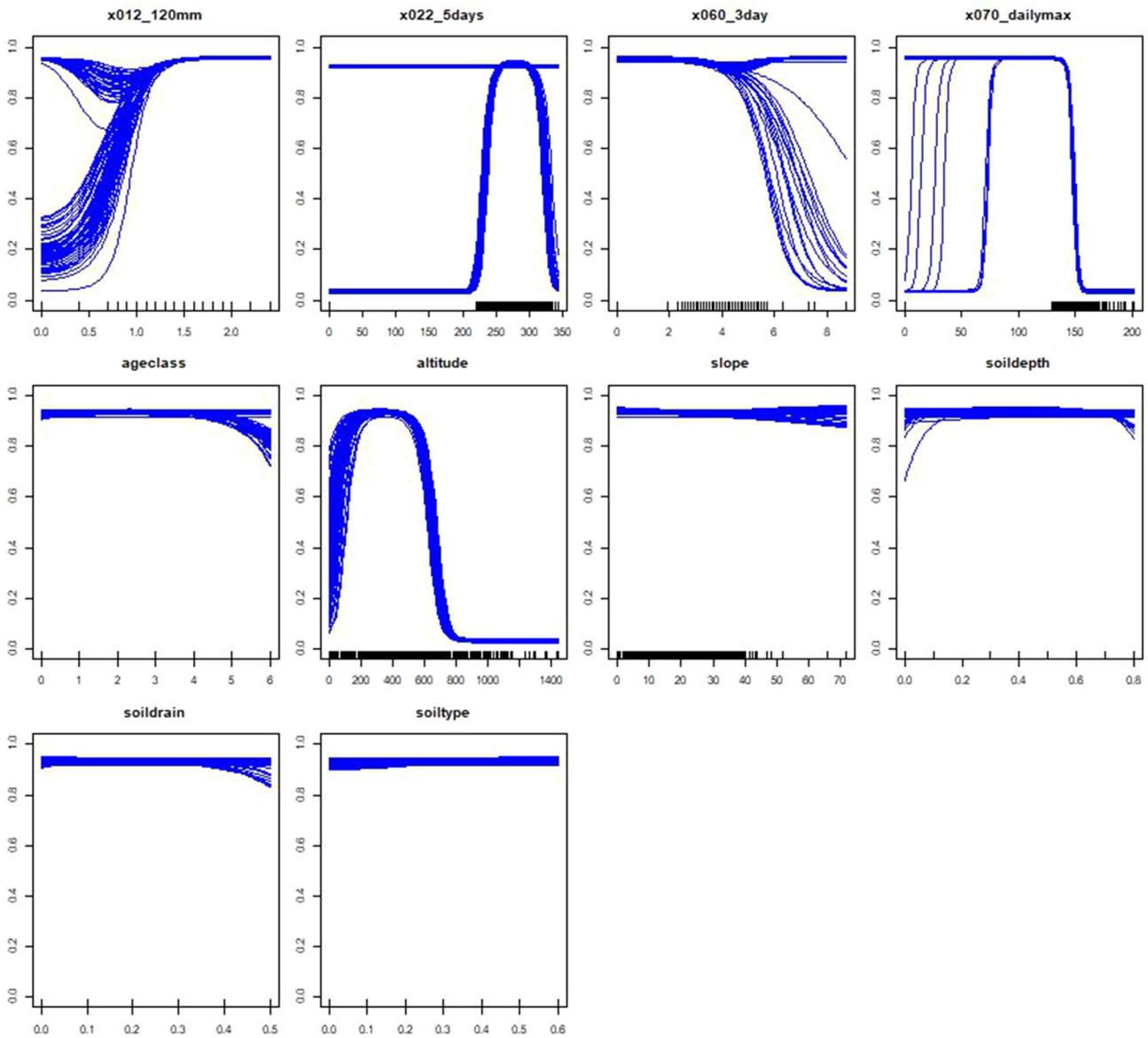
6. RF

Response curves for landslide's RF



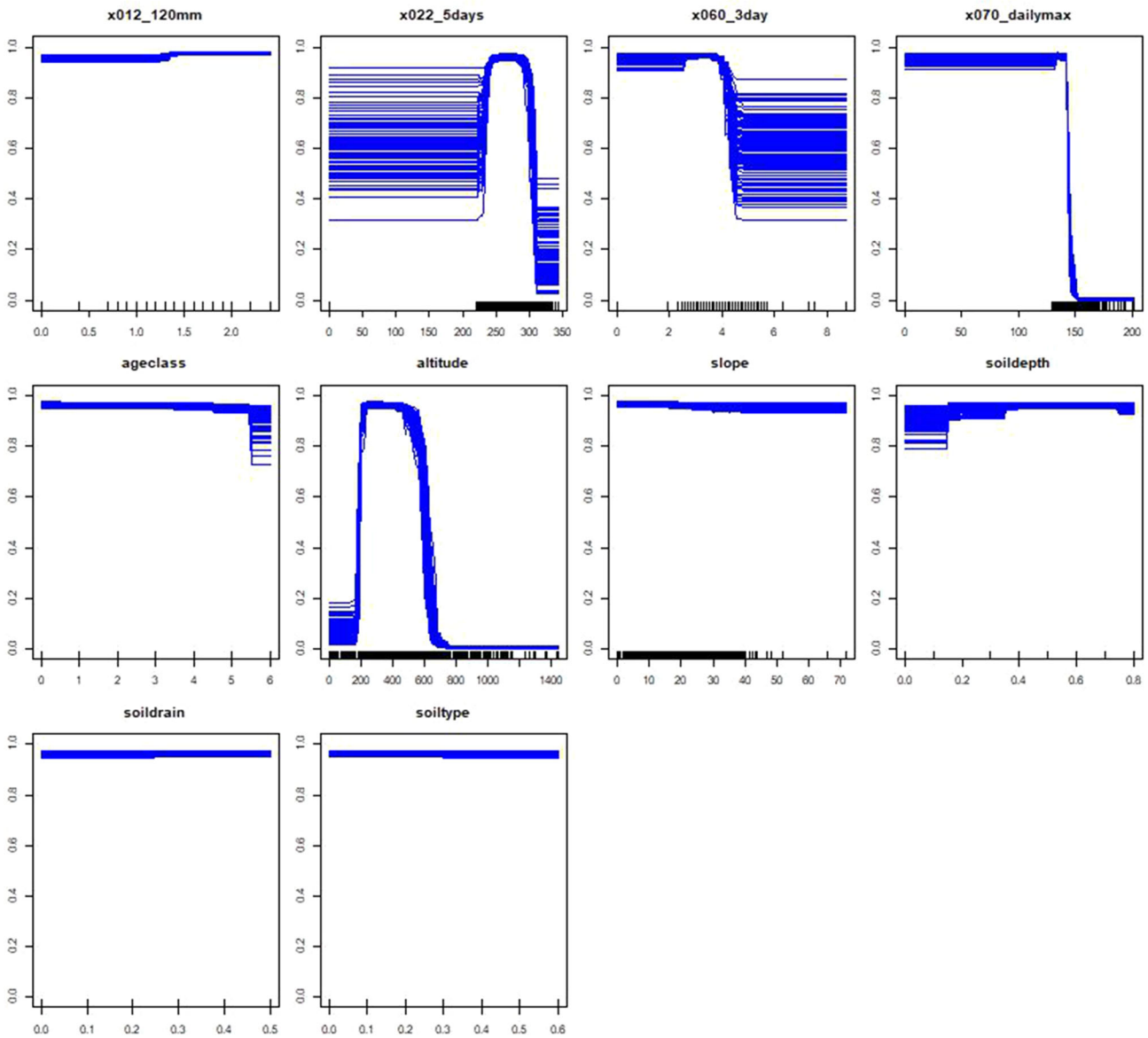
7. GLM

Response curves for landslide's GLM



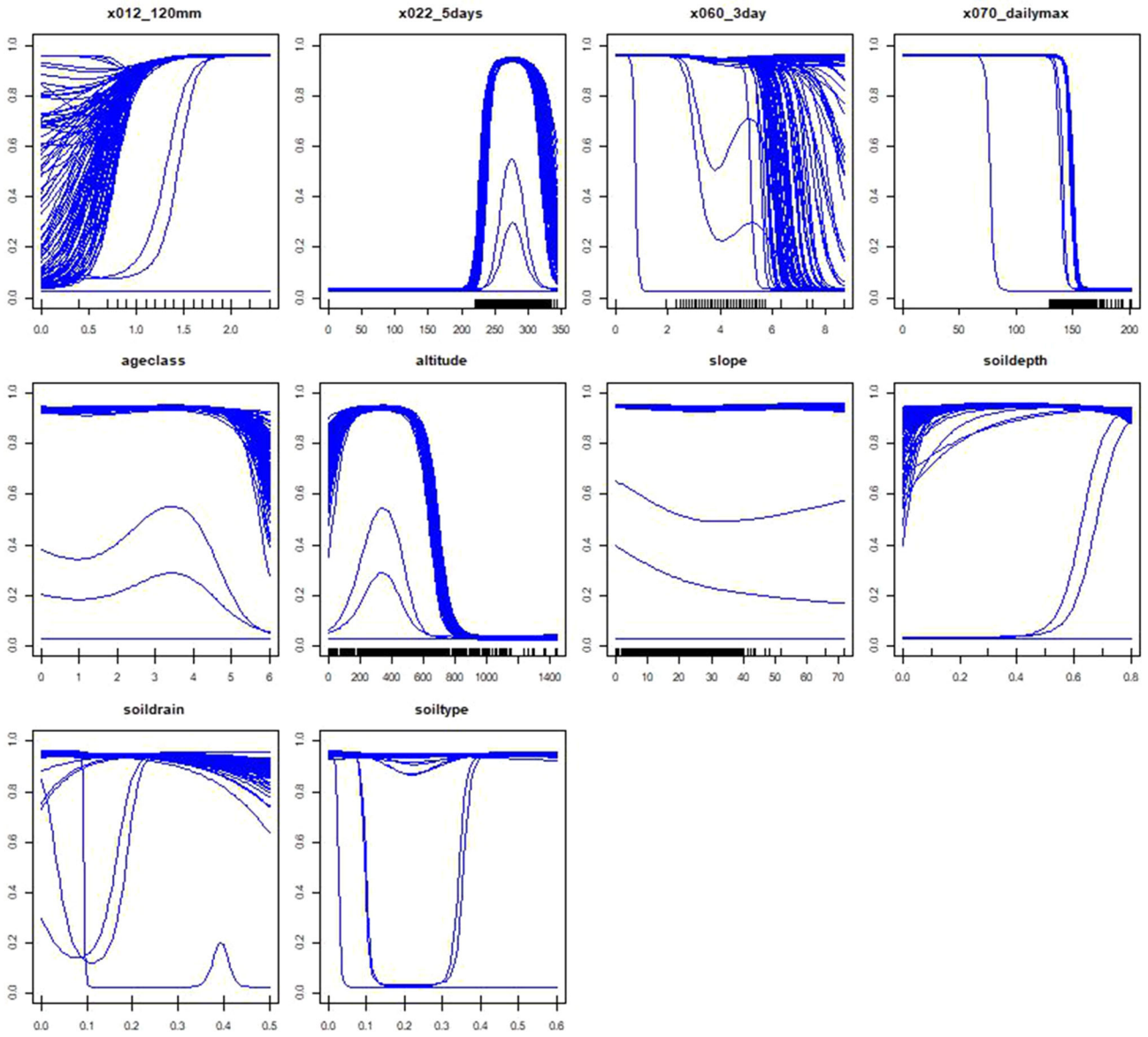
8. GBM

Response curves for landslide's GBM



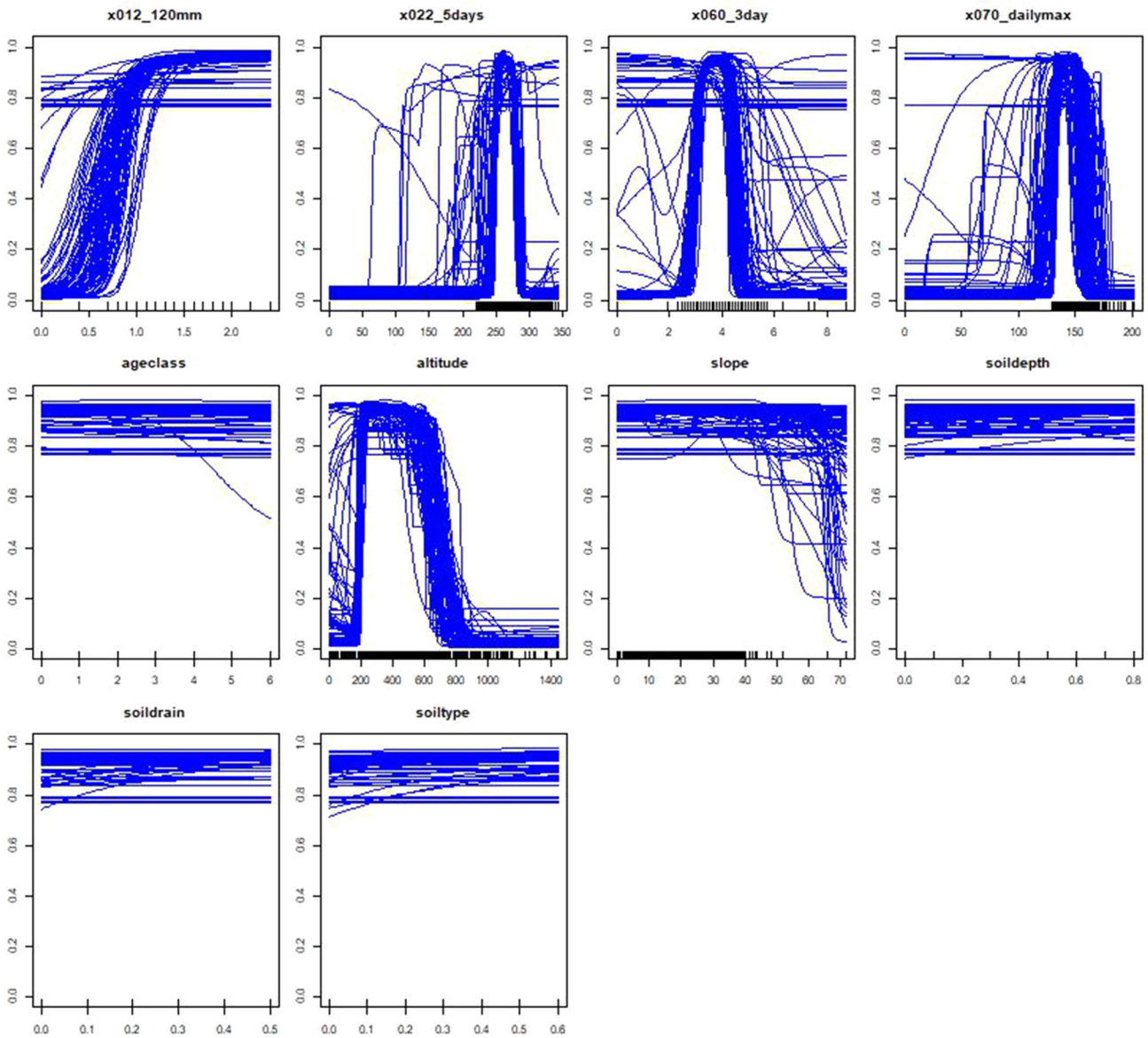
9. GAM

Response curves for landslide's GAM



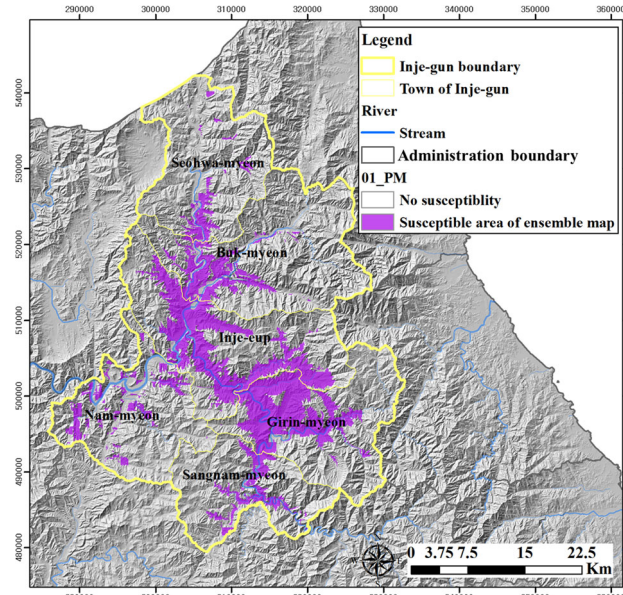
10. ANN

Response curves for landslide's ANN

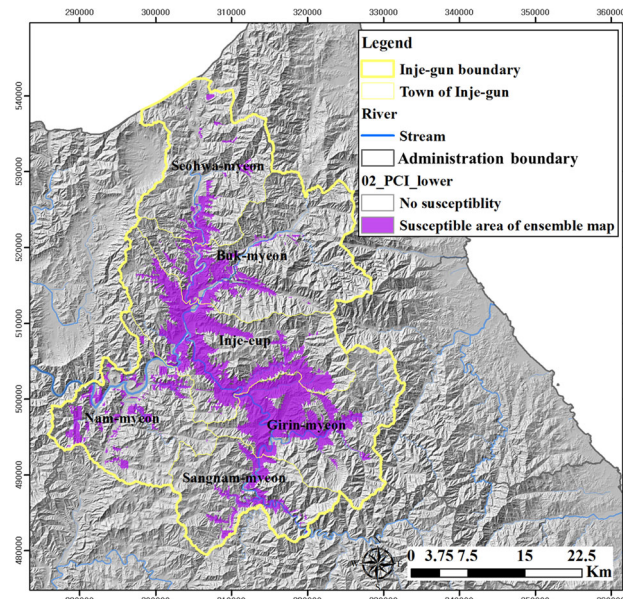


Appendix 5: Results of ensemble models for the present conditions

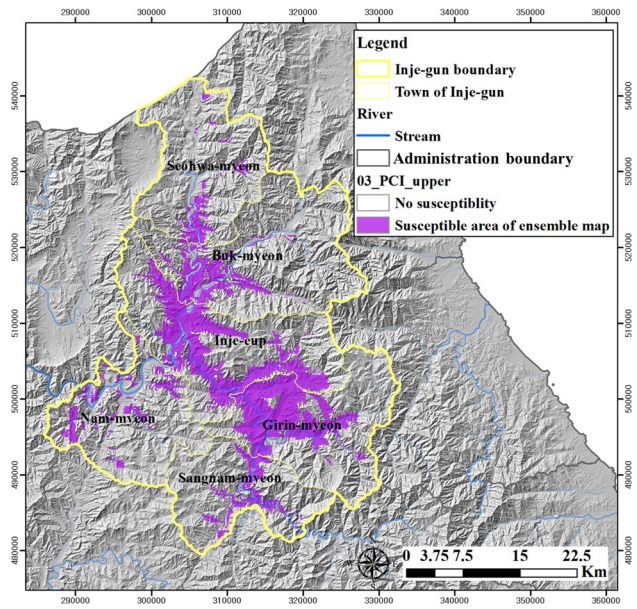
1. PM



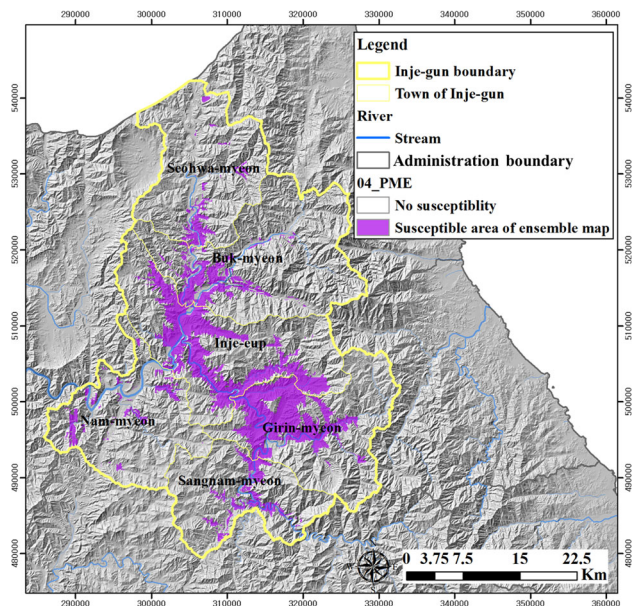
2. PCI lower



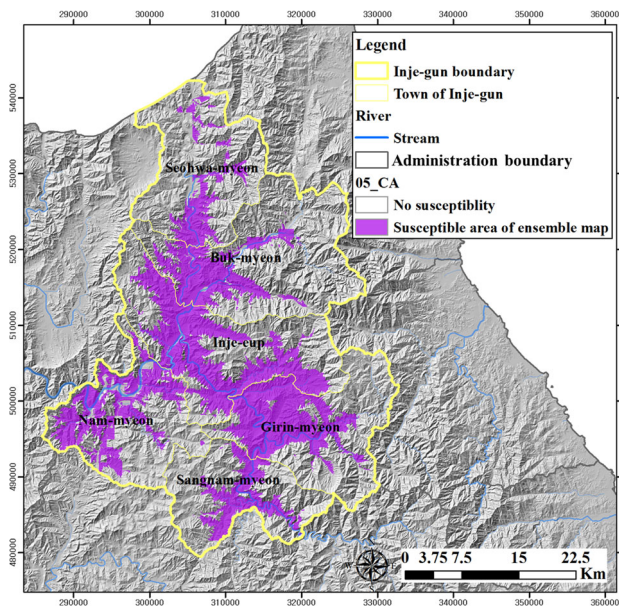
3. PCI upper



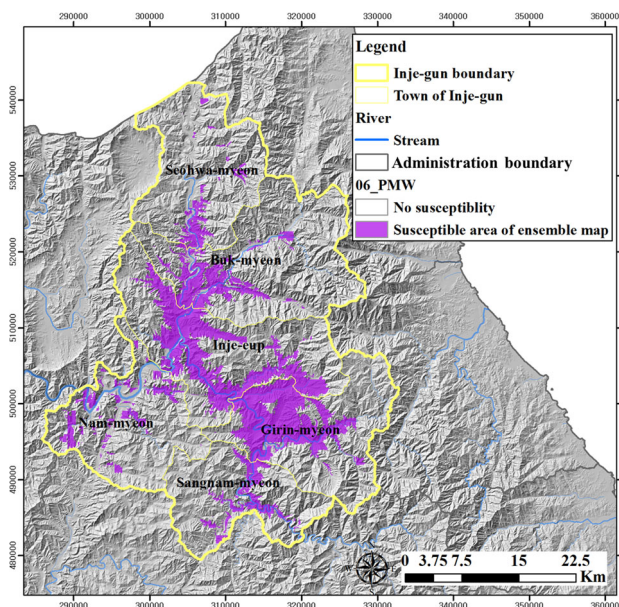
4. PME



5. CA



6. PMW



References

- Akgun A (2012) A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at Izmir, Turkey. *Landslides* 9:93–106
- Akgun A, Dag S, Bulut F (2008) Landslide susceptibility mapping for a landslide-prone area (Findikli, NE of Turkey) by likelihood-frequency ratio and weighted linear combination models. *Environ Geol* 54:1127–1143
- Althuwaynee OF, Pradhan B, Park H-J, Lee JH (2014) A novel ensemble bivariate statistical evidential belief function with knowledge-based analytical hierarchy process and multivariate statistical logistic regression for landslide susceptibility mapping. *CATENA* 114:21–36
- Ayalew L, Yamagishi H (2005) The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology* 65:15–31
- Bartlett P, Shawe-Taylor J (1999) Generalization performance of support vector machines and other pattern classifiers. In: *Advances in Kernel Methods—Support Vector Learning*, pp 43–54
- Bonachea J, Remondo J, Terán D et al (2009) Landslide risk models for decision making. *Risk Anal* 29:1629–1643
- Boo K, Kwon W, Baek H (2006) Change of extreme events of temperature and precipitation over Korea using regional projection of future climate change. *Geophys Res Lett* 33:8–11. <https://doi.org/10.1029/2005GL023378>
- Bühlmann P, Yu B (2003) Boosting with the L 2 loss: regression and classification. *J Am Stat Assoc* 98:324–339
- Bui DT, Pradhan B, Lofman O et al (2012) Landslide susceptibility mapping at Hoa Binh province (Vietnam) using an adaptive neuro-fuzzy inference system and GIS. *Comput Geosci* 45:199–211
- Buisson L, Thuiller W, Casajus N et al (2010) Uncertainty in ensemble forecasting of species distribution. *Glob Change Biol* 16:1145–1157
- Catani F, Casagli N, Ermini L et al (2005) Landslide hazard and risk mapping at catchment scale in the Arno River basin. *Landslides* 2:329–342
- Chiou I-J, Chen C-H, Liu W-L et al (2015) Methodology of disaster risk assessment for debris flows in a river basin. *Stoch Environ Res Risk Assess* 29:775–792. <https://doi.org/10.1007/s00477-014-0932-1>
- Choi JH, Oh JY, Kim YS, Kim HT (2011) Analysis of the controlling factors of an urban-type landslide at Hwangryeong mountain based on tree growth patterns and geomorphology. *J Eng Geol* 21:281–293
- Claessens L, Schoorl JM, Veldkamp A (2007) Modelling the location of shallow landslides and their effects on landscape dynamics in large watersheds: an application for Northern New Zealand. *Geomorphology* 87:16–27
- Clinton N, Yu L, Gong P (2015) Geographic stacking: decision fusion to increase global land cover map accuracy. *ISPRS J Photogramm Remote Sens* 103:57–65. <https://doi.org/10.1016/j.isprsjprs.2015.02.010>
- Dehnavi A, Aghdam IN, Pradhan B, Varzandeh MHM (2015) A new hybrid model using step-wise weight assessment ratio analysis (SWARA) technique and adaptive neuro-fuzzy inference system (ANFIS) for regional landslide hazard assessment in Iran. *CATENA* 135:122–148
- Dimitriadou E, Weingessel A, Hornik K (2003) *A cluster ensembles framework*. IOS Press, Amsterdam
- Ermini L, Catani F, Casagli N (2005) Artificial neural networks applied to landslide susceptibility assessment. *Geomorphology* 66:327–343
- Felicísimo ÁM, Cuartero A, Remondo J, Quirós E (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides* 10:175–189
- Franklin J (2009) *Mapping species distributions*. Cambridge University Press, Cambridge

- Ghosh J, Acharya A (2011) Cluster ensembles. *Wiley Interdiscip Rev Data Min Knowl Discov* 1:305–315
- Guzzetti F, Peruccacci S, Rossi M, Stark CP (2008) The rainfall intensity–duration control of shallow landslides and debris flows: an update. *Landslides* 5:3–17
- Hansson SL, Røjvall AS, Rastam M et al (2005) Psychiatric telephone interview with parents for screening of childhood autism–tics, attention-deficit hyperactivity disorder and other comorbidities (A–TAC) Preliminary reliability and validity. *Br J Psychiatry* 187:262–267
- Kappes MS, Papathoma-Köhle M, Keiler M (2012) Assessing physical vulnerability for multi-hazards using an indicator-based methodology. *Appl Geogr* 32:577–590
- Kim J (2012) The analysis of planning methode and case study for Model “Climate Change Adaptation City”. *J Korea Inst Ecol Archit Environ* 12:13–19
- Kim WY, Chae BG (2009) Characteristics of rainfall, geology and failure geometry of the landslide areas on natural terrains, Korea. *J Eng Geol* 19:331–344
- Kim GH, Yune CY, Lee HG, Hwang JS (2011a) Debris flow analysis of landslide area in Inje using GIS. *J Korean Soc Surv Geod Photogramm Cartogr* 29:47–53
- Kim KH, Jung HR, Park JH, Ma HS (2011b) Analysis on rainfall and geographical characteristics of landslides in Gyeongnam Province. *J Korean Environ Restor Technol* 14:33–45
- Kim HG, Lee DK, Park C et al (2015) Evaluating landslide hazards using RCP 4.5 and 8.5 scenarios. *Environ Earth Sci* 73:1385–1400
- Korea Ministry of Environment (KME) (2005) GIS map of forest type for Republic of Korea
- Korea Ministry of Environment (KME) (2008) Digital elevation model for Republic of Korea
- Korea Meteorological Administration (2011) White Paper for Typhoon of Republic of Korea
- Kovalchuk SV, Krikunov AV, Knyazkov KV, Boukhanovsky AV (2017) Classification issues within ensemble-based simulation: application to surge floods forecasting. *Stoch Environ Res Risk Assess* 31:1183–1197. <https://doi.org/10.1007/s00477-016-1324-5>
- Krishnamurti TN, Kishtawal CM, Zhang Z et al (2000) Multimodel ensemble forecasts for weather and seasonal climate. *J Clim* 13:4196–4216
- Kwon HS (2014) Applying ensemble model for identifying uncertainty in the species distribution models. *J Korean Soc Geospat Inf Syst* 29:55:47–52
- Ladle R, Hortal J (2013) Mapping species distributions: living with uncertainty. *Front Biogeogr* 5:4–6
- Lee S, Oh H-J (2012) Ensemble-based landslide susceptibility maps in Jinbu area, Korea. In: Pradhan B, Buchroithner M (eds) *Terrigenous mass movements*. Springer, Berlin, pp 193–220
- Lee M-J, Park I, Lee S (2015) Forecasting and validation of landslide susceptibility using an integration of frequency ratio and neuro-fuzzy models: a case study of Seorak mountain area in Korea. *Environ Earth Sci* 74:413–429
- Li Y, Chen G, Tang C et al (2012) Rainfall and earthquake-induced landslide susceptibility assessment using GIS and artificial neural network. *Nat Hazards Earth Syst Sci* 12:2719–2729
- Lian C, Zeng Z, Yao W, Tang H (2014) Extreme learning machine for the displacement prediction of landslide under rainfall and reservoir level. *Stoch Environ Res Risk Assess* 28:1957–1972. <https://doi.org/10.1007/s00477-014-0875-6>
- Liu K, Yao C, Chen J et al (2016) Comparison of three updating models for real time forecasting: a case study of flood forecasting at the middle reaches of the Huai River in East China. *Stoch Environ Res Risk Assess* 31:1–14. <https://doi.org/10.1007/s00477-016-1267-x>
- Miao F, Wu Y, Xie Y et al (2016) Research on progressive failure process of Baishuihe landslide based on Monte Carlo model. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-016-1224-8>
- Niedzielski T, Miziński B (2016) Real-time hydrograph modelling in the upper Nysa Klodzka river basin (SW Poland): a two-model hydrologic ensemble prediction approach. *Stoch Environ Res Risk Assess*. <https://doi.org/10.1007/s00477-016-1251-5>
- Nourani V, Pradhan B, Ghaffari H, Sharifi SS (2014) Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. *Nat Hazards* 71:523–547
- Oh HJ (2010) Landslide detection and landslide susceptibility mapping using aerial photos and artificial neural networks. *Korean J Remote Sens* 26:47–57
- Oh CY, Choi CU, Kim KT (2009) Analysis of landslide characteristics of Inje area using SPOT5 images and GIS analysis. *Korean J Remote Sens* 25:445–454
- Park S, Choi C, Kim B, Kim J (2013) Landslide susceptibility mapping using frequency ratio, analytic hierarchy process, logistic regression, and artificial neural network methods at the Inje area, Korea. *Environ Earth Sci* 68:1443–1464
- Pradhan B (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Comput Geosci* 51:350–365
- Pradhan B, Lee S (2010) Delineation of landslide hazard areas on Penang Island, Malaysia, by using frequency ratio, logistic regression, and artificial neural network models. *Environ Earth Sci* 60:1037–1054
- Riebau AR, Fox DG (2005) Damage assessment of agrometeorological relevance from natural disasters: economic and social consequences. In: Sivakumar MV, Motha RP, Das HP (eds) *Natural disasters and extreme events in agriculture*. Springer, Berlin, pp 119–135
- Rokach L (2010) Ensemble-based classifiers. *Artif Intell Rev* 33:1–39
- Son JW, Kim KT, Lee CH, Choi CU (2009) Analysis of landslide in Inje region using aerial photograph and GIS. *J Korean Soc Geospat Inf Syst* 17:61–69. <https://doi.org/10.1017/CBO9781107415324.004>
- Son CY, Kim JS, Il Moon Y, Lee JH (2014) Characteristics of tropical cyclone-induced precipitation over the Korean River basins according to three evolution patterns of the Central-Pacific El Nino. *Stoch Environ Res Risk Assess* 28:1147–1156. <https://doi.org/10.1007/s00477-013-0804-0>
- Sudmeier-Rieux K, Jaquet S, Derron M-H et al (2012) A case study of coping strategies and landslides in two villages of Central-Eastern Nepal. *Appl Geogr* 32:680–690
- Sung JH, Kang H-S, Park S et al (2012) Projection of extreme precipitation at the end of 21st Century over South Korea based on Representative Concentration Pathways (RCP). *Korean Meteorol Soc* 22:221–231
- Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD—a platform for ensemble forecasting of species distributions. *Ecography (Cop)* 32:369–373
- Thuiller W, Georges D, Engler R (2015) *biomod2 package manual*
- Tongal H, Booij MJ (2017) Quantification of parametric uncertainty of ANN models with GLUE method for different streamflow dynamics. *Stoch Environ Res Risk Assess* 31:993–1010. <https://doi.org/10.1007/s00477-017-1408-x>
- Torizin J (2016) Elimination of informational redundancy in the weight of evidence method: an application to landslide susceptibility assessment. *Stoch Environ Res Risk Assess* 30:635–651. <https://doi.org/10.1007/s00477-015-1077-6>

- Wang X, Yang T, Li X et al (2016a) Spatio-temporal changes of precipitation and temperature over the Pearl River basin based on CMIP5 multi-model ensemble. *Stoch Environ Res Risk Assess* 31:1–13. <https://doi.org/10.1007/s00477-016-1286-7>
- Wang YY, Huang GH, Wang S (2016b) CVaR-based factorial stochastic optimization of water resources systems with correlated uncertainties. *Stoch Environ Res Risk Assess* 31:1543–1553. <https://doi.org/10.1007/s00477-016-1276-9>
- Water Resources Management Information System (WAMIS) (2006) GIS map of soil for Republic of Korea
- Yeon YK (2011) Evaluation and analysis of Gwangwon-do landslide susceptibility using logistic regression. *J Korean Assoc Geogr Inf Stud* 14:116–127
- 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
- Yilmaz I (2009) Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat—Turkey). *Comput Geosci* 35:1125–1138
- Yilmaz I (2010) Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine. *Environ Earth Sci* 61:821–836
- Yoo N, Yoon B, Um D et al (2012) Analysis of rainfall characteristics and landslides at the west side area of Gangwon Province. *J Korean Geoenviron Soc* 13:75–82
- Zare M, Pourghasemi HR, Vafakhah M, Pradhan B (2013) Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multi-layer perceptron (MLP) and radial basic function (RBF) algorithms. *Arab J Geosci* 6:2873–2888