17th World Conference on Earthquake Engineering, 17WCEE
Sendai, Japan - September 13th to 18th 2020

GIS-BASED STATISTICAL LANDSLIDE HAZARD ANALYSIS: A STUDY ON THE 2018 HOKKAIDO EARTHQUAKE-TRIGGERED LANDSLIDES

P. Ge [1] and K. Meguro [2]

[1] *Ph.D. Student, Department of Civil Engineering, Graduate School of Engineering, The University of Tokyo (UTokyo), gepl@iis.u-tokyo.ac.jp*
[2] *Director/Professor, International Center for Urban Safety Engineering (ICUS), Institute of Industrial Science (IIS), UTokyo, meguro@iis.u-tokyo.ac.jp*

Abstract

The 2018 Hokkaido Eastern Iburi Earthquake struck the eastern region of Iburi with an epicenter of 42.686°N, 141.929°E, a depth of 35 km, and a moment magnitude of 6.6 (Mw) on September 6th, just one day after the powerful Typhoon Jebi passed. This earthquake triggered approximately 6,000 landslides over an area of 400 km^2 near the Atsuma Town region, causing more than 80% fatalities in this disaster. Most triggered landslides are shallow landslides moving down the volcano air-fall pumice layer, with several deep-seated landslides also found in the southeast of the area. The number and total area of the landslides were said to be the largest in Japan since the Meiji Era. A good understanding and analysis of the extensive earthquake-triggered landslides are important for landslide hazard management and mitigation in the future. Therefore, we carried out a landslide hazard analysis using a geographic information system (GIS)-based multivariate statistical approach. Instead of exploring the slope failure mechanisms via physical models, the statistical approach establishes a relationship between independent variables represented by conditioning factors of landslides and dependent variables, in this case: landslides, through proper use of statistic indicators. It is based on the assumption that the past and present are keys to the future, which means the future landslides will occur under similar conditions of the previous landslides. First, the earthquake-induced landslide inventory and relevant conditioning factors were collected, preprocessed, and constructed as a database on a GIS platform as dependent and independent variables for statistical analysis. Considered intrinsic and extrinsic conditioning factors include geology, topography, surface vegetation, rainfall, and ground motion characteristics. Then representative indicators of the conditioning factors, such as the slope gradient and aspect, were calculated and processed in ArcGIS software, and the predictive ability of them was analyzed to find the effective factors for optimizing the model. Next, landslide inventory data were randomly divided into a ratio of 7:3, with 70% of the data used for model training and 30% of the data applied for model validation. All applied spatial data, either in raster or vector types, were converted and resampled into 10 m raster cells for raster calculation to incorporate various layers of information. Finally, a landslide hazard map in the studied area was constructed and validated by statistical analysis using these prepared and selected data. It is expected to provide some reference for future landslide hazard monitoring and management in this area by updating the varying conditioning factors in the model.

Keywords: landslide hazard analysis; statistical approach; Hokkaido earthquake; geographic information system (GIS)

1. Introduction

Landslides are defined as mass movements of rock, earth, or debris down a slope [1]. They can occur when forces acting down-slope (mainly due to gravity) exceed the strength of earth materials that compose the slope [2]. Therefore, both factors that can increase down-slope forces and factors that can decrease earth strength can cause landslides, including natural phenomena (e.g., rainfall and earthquakes) and human activities (e.g., blasting and underground mining). Landslides are considered to be widespread problematic geo-hazards worldwide that can cause casualties, property damages, and economic loss in mountainous areas. Globally, approximately 17% of the fatalities occurred due to landslides with around 66 million people living within high-risk landslide areas [3]. A good understanding of landslide hazard condition is an important task that may help the government, decision-makers, and engineers in land use planning and slope management. Analyses of previous landslides can provide valuable information that may help estimate, manage, and mitigate landslide hazards in the future.

On September 6th, 2018, just one day after the Super Typhoon Jebi passed, an earthquake with a moment magnitude of 6.6 (Mw) struck the Eastern Iburi region in Hokkaido, Japan, and triggered extensive landslides near the towns of Atsuma, Mukawa, and Abira. This earthquake, officially known as the 2018 Hokkaido Estern Iburi Earthquake, occurred at the epicenter of 42.686°N, 141.929°E, with a depth of 35km [4], and caused 41 fatalities, including 36 people killed by the triggered landslides. The number and total area of the triggered landslides were approximately 6,000 and 400km^2, which were said to be the largest in Japan since Meji Era [5]. Most triggered landslides are shallow landslides, several meters depth, moving down the pyroclastic fall deposit layer on the slope with high mobility and long run-out. Several deep-seated ones were also found in the southeast of this area, including one that formed a landslide dam. Moreover, most landslides are spoon and planar types with small to medium sizes, same as the rainfall-induced landslides [6-7]. The earthquake, rainfall, and pyroclastic fall deposit surface soil layer were considered to be important causative factors for the extensive triggered landslides. Following this disaster, a good understanding and analysis of the massive event-based landslides may provide some valuable information for landslide hazard management and mitigation in the future.

Therefore, in this study, applying the massive event-based landslides, a landslide hazard analysis was carried out by geographic information system (GIS) using the statistical approach. The GIS is a system designed to capture, store, manipulate, analyze, manage, and present all kinds of spatial or geographic data. It has become a popular technology applied in calculating and managing natural hazards including landslide hazards since the middle of 1980s [8]. The statistical approach can construct a model applying the landslide inventory (dependent variable) and causative factors of landslides (independent variables) by the use of proper indicators. It does not require detailed physical parameters, such as soil strength, soil depth, and hydrological parameters, which are difficult to collect in a large area [9], comparing with the deterministic approach that explores the slope failure mechanism by physical models. Moreover, the model constructed by the statistical analysis may be applied for landslide prediction, monitoring, and management in the future by updating changed causative factors, based on the assumption that the future landslides will occur under similar conditions of the previous landslides.

In order to construct the landslide hazard model by statistical analysis, landslide inventory and conditioning factors of the landslides were first collected as dependent and independent variables. Topography, geology, surface vegetation, rainfall, and ground motion characteristics were all collected as intrinsic and extrinsic conditioning factors. Then, all collected data, either in a raster or vector format, were converted into unified format-10m raster cells for further analyses and raster calculation. After that, a non-landslide data set that has the similar size to the landslide cells were randomly selected for further analysis and model construction, as the number of landslide cells accounted for only 1.26% of the total cells, representing a rare event data [10]. Then, several conditioning factors were analyzed by the values of the area under the receiver operating characteristic curves (AUC), Pearson's correlation coefficient, tolerance (TOL), and variance inflation factors (VIF) to check their effectiveness and multicollinearity problems. Finally, a landslide hazard

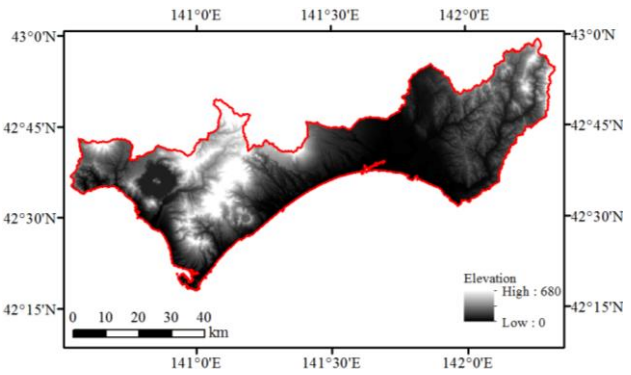
model was constructed by logistic regression using the prepared and selected dependent and independent data, and a landslide hazard map was created in the studied area.

2. Study Area and Data Acquisition

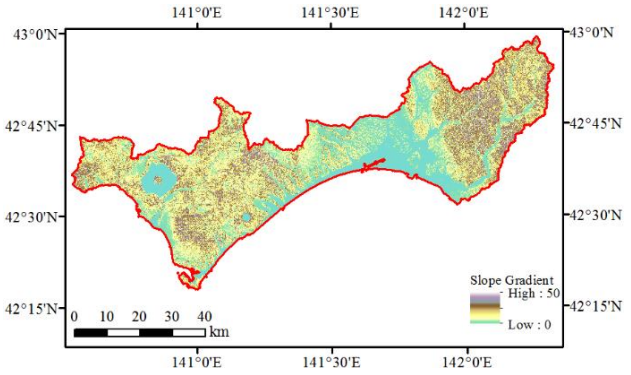
The study area is Iburi, located in south central part of Hokkaido, Japan. It has an elongated shape, lying between latitude 42.30° N ~ 42.99° N and longitude 140.53° E ~ 142.33° E, and stretching 98km from south to north and 152km from west to east. Iburi has a coastline of 218 km to the Pacific Ocean on the south and borders Shiribeshi Subprefecture, Ishikari Subprefecture, and Sorachi Subprefecture in the north, Oshima Subprefecture in the west, and Hidaka Subprefecture in the east. The area and population of Iburi are 3,698km^2 and 42 million, accounting for 4.4% of the total area of Hokkaido, and 7.6% of the total population in Hokkaido, respectively.

To carry out landslide hazard statistical analysis in the study area, dependent (i.e., the event-based landslide inventory) and independent variables (i.e., conditioning factors of the landslides) were first collected. Geospatial Information Authority of Japan (GSI) published a first-hand landslide database several days after the earthquake. Nevertheless, most of the published landslides are composed of several or dozens of landslides. Regarding the unreasonable landslide unit, Zhang et al. [11] carried out a manual segmentation and combination according to high-resolution aerial images and a digital elevation model (DEM), creating a detailed landslide inventory map. This detailed landslide inventory map created by Zhang et al. [11] was applied as the dependent variable for landslide hazard model construction in this study.

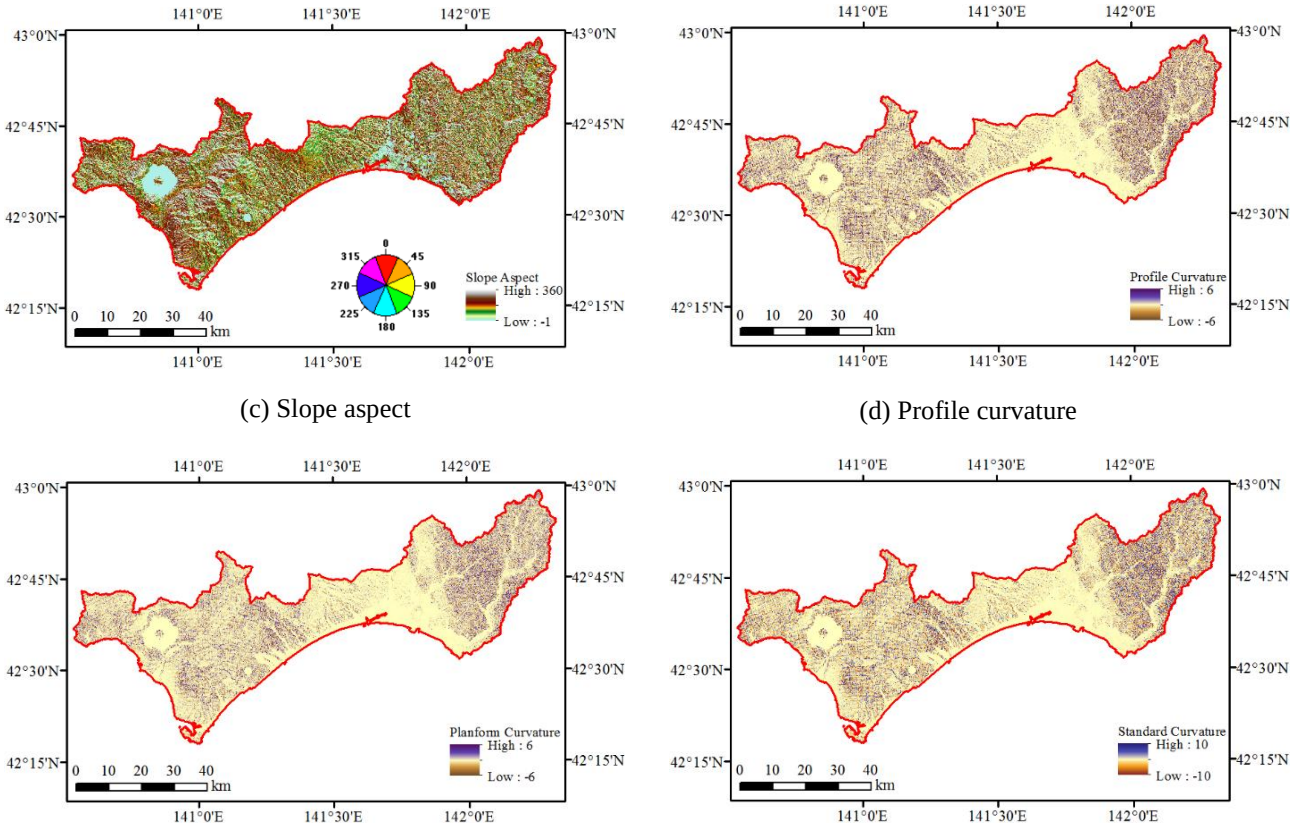
Collected conditioning factors as independent variables included topography, geology, surface vegetation, rainfall, and ground motion characteristics. Topography characteristics were calculated and derived from a 10m resolution DEM provided by GSI. Calculated characteristics included elevation, slope gradient, slope aspect, and planform/profile/standard curvature (Fig.1). Geology characteristics were collected from the Geological Survey of Japan, AIST, which provided a polygon vector geology map with a scale of 1:200,000 (Fig.2). Surface vegetation characteristics were obtained from the Biodiversity Center of Japan, which provided a polygon vector surface vegetation map with a scale of 1:50,000 (Fig.3). Rainfall information was collected and derived from the daily precipitation data measured at 14 valid rainfall stations in Iburi, Hokkaido, which were provided by the Japan Meteorological Agency (JMA). 3-day/1-week/2-week/3-week/1-month cumulative precipitation before the disaster was calculated, using these daily precipitation data at each station. Moreover, interpolation was carried out for the point rainfall data using the inverse distance weighted (IDW) technique to supplement data in areas without stations (Fig.4). Ground motion information was obtained from the U.S. Geology Survey (USGS), which provided vector shakemaps of Modified Mercalli Intensity (MMI), peak ground acceleration (PGA), peak ground velocity (PGV), and 0.3s/1.0s/3.0s peak spectral acceleration (PSA03/10/30) (Fig.5).



(a) Elevation



(b) Slope gradient



(c) Slope aspect

(d) Profile curvature

(e) Planform curvature

(f) Standard curvature

Fig. 1 – Topography maps of the study area

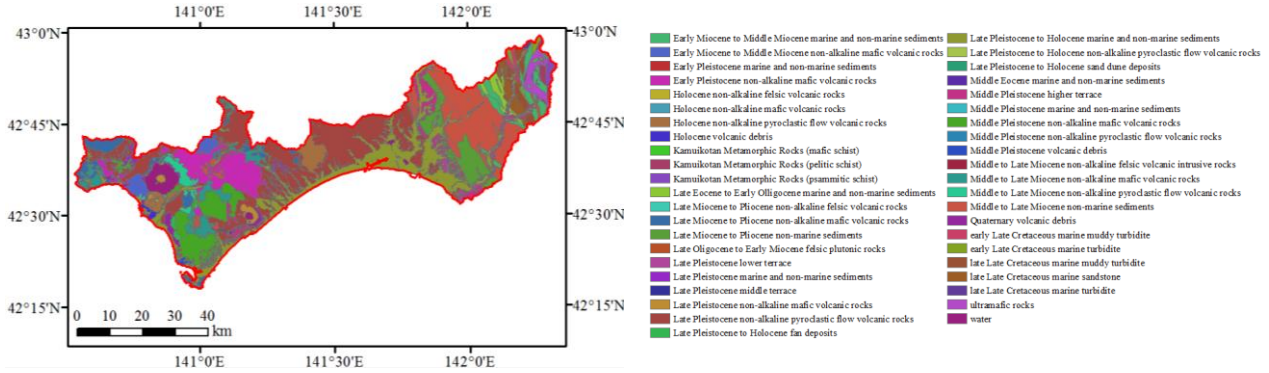


Fig. 2 – Geology map of the study area

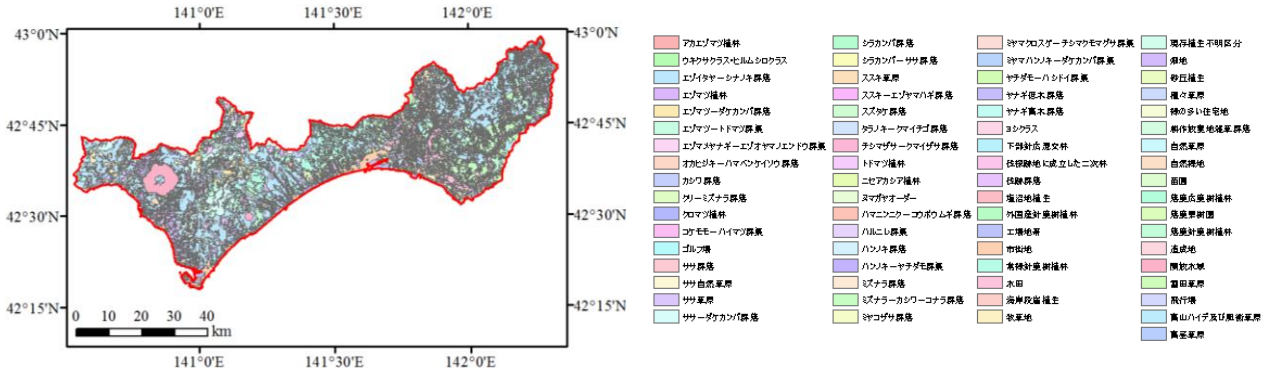
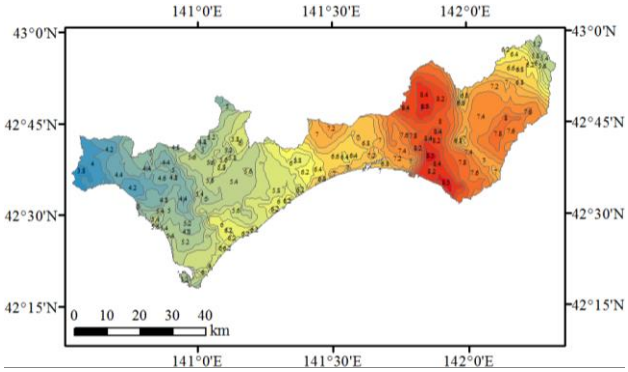
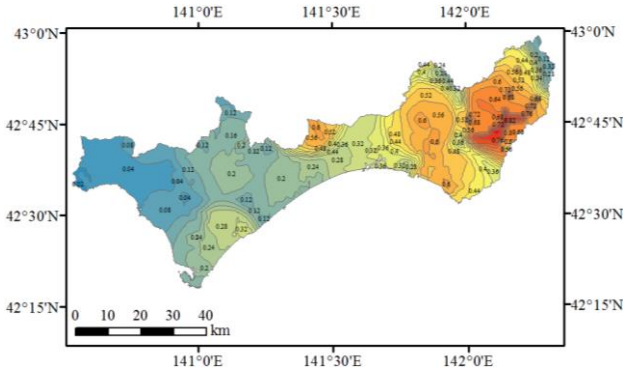


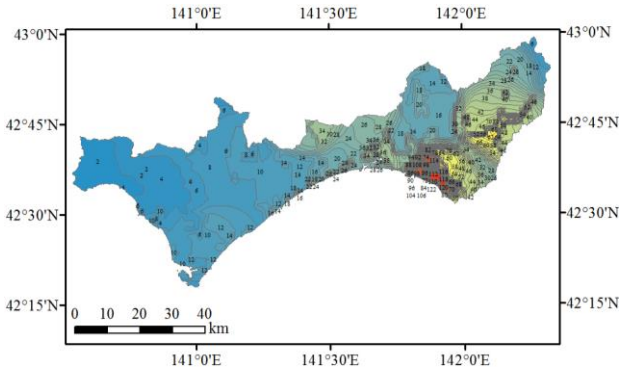
Fig. 3 – Surface vegetation map of the study area



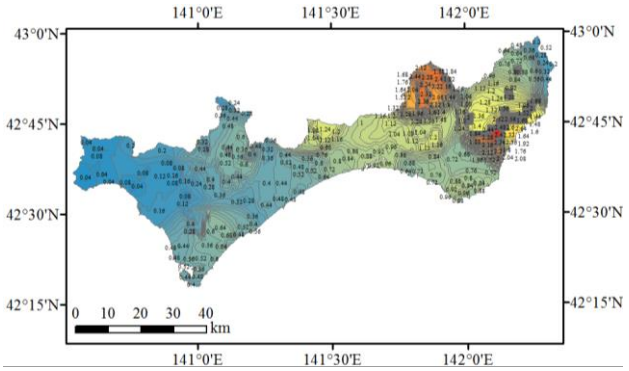
(a) MMI



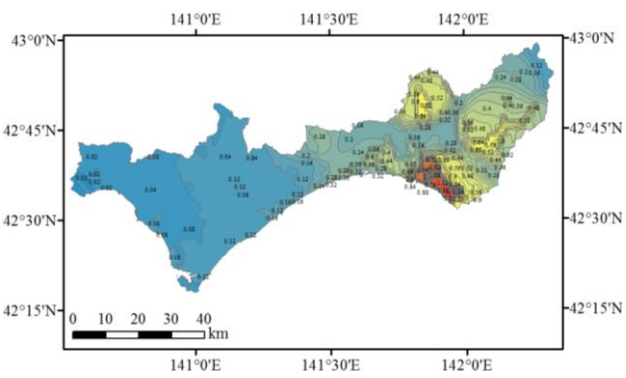
(b) PGA



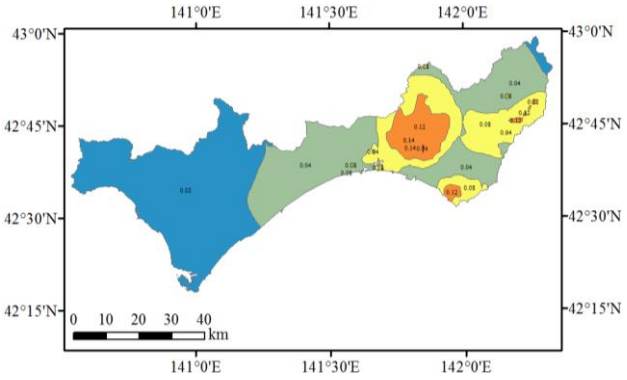
(c) PGV



(d) PSA03

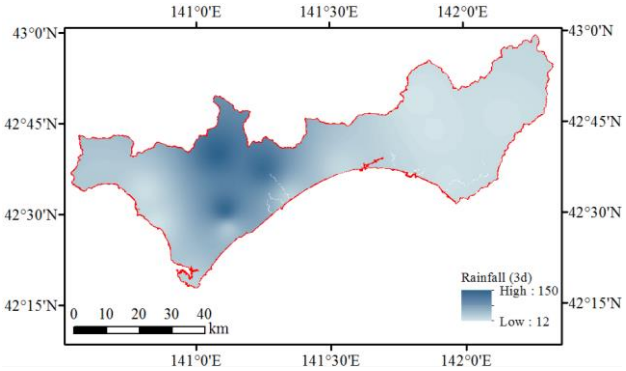


(e) PSA10

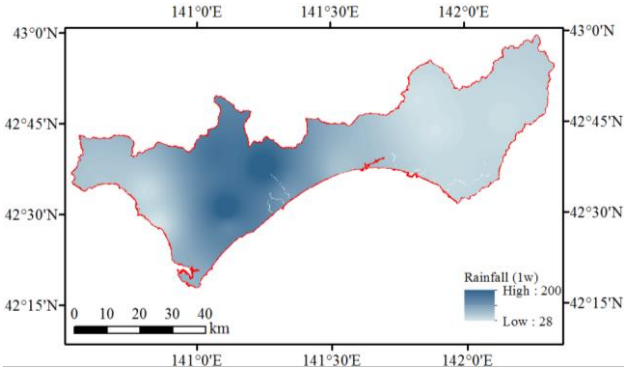


(f) PSA30

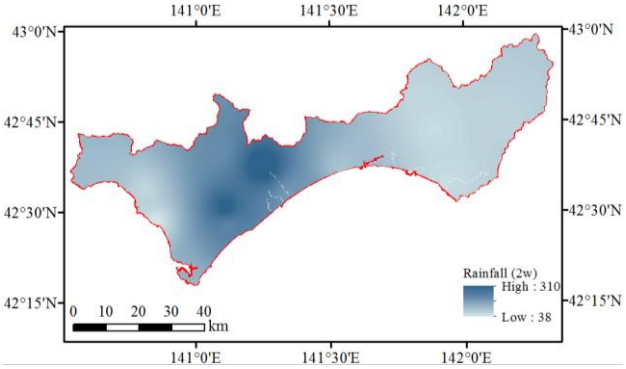
Fig. 4 – Ground motion maps of the study area



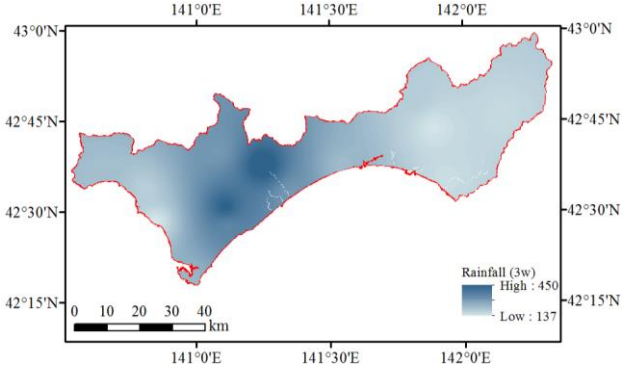
(a) 3-day cumulative precipitation before the disaster



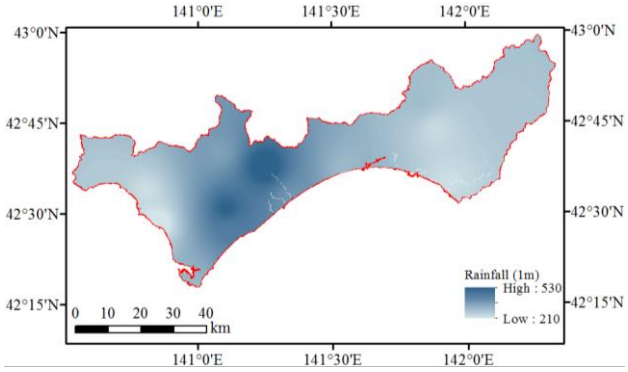
(b) 1-week cumulative precipitation before the disaster



(c) 2-week cumulative precipitation before the disaster



(d) 3-week cumulative precipitation before the disaster



(e) 1-month cumulative precipitation before the disaster

Fig. 5 – Precipitation maps of the study area

3. Data Pre-processing and Preparation

The collected data included both raster and vector types. In order to carry out further analyses for landslide hazard model construction, these different types of data were all converted into 10×10m raster cells. The number of total raster cells is more than 30 million, and the number of raster cells tagged as landslides is around 400 thousand. Landslide cells represent only 1.26% of the total number of cells, and therefore are considered as rare event data [10, 12-13]. The term 'rare event data' was induced in political sciences, which described the binary dependent variables with dozens to thousands of times fewer 1s than 0s. When modeling such conditions by popular statistical procedures (e.g., logistic regression), the occurrence probability of the rare event can be sharply underestimated [10]. It is generally suggested to use the same number of landslide and non-landslide cells for model training [13-15], even though there are also studies applying unequal proportions [16-17]. In this study, an equal proportion of landslide and non-landslide cells was adopted. All landslide cells were applied, and a non-landslide dataset with a similar size to the landslide cells was randomly selected to make the ratio between the landslide presence and absence to be 1.

After selecting raster cells for analysis, receiver operating characteristic (ROC) analyses were carried out for the three curvature-related conditioning factors (e.g., planform curvature, profile curvature, and standard curvature), six earthquake-related conditioning factors (e.g., MMI, PGA, PGV, PSA03, PSA10, and PSA30), and five rainfall-related conditioning factors (e.g., 3-day, 1-week, 2-week, 3-week, and 1-month cumulative precipitation before the disaster), to select the effective ones for model construction [18]. The larger the AUC value in ROC analyses, the more effective the factor. Standard curvature, PSA03, and 1-week cumulative precipitation before the disaster showed better performance in the analyses, respectively, and therefore were selected for subsequent analyses. After that, the Pearson's correlation coefficient, TOL, and VIF were calculated for the remaining numerical independent variables, to diagnose the multicollinearities between them. Pearson's correlation coefficient means the covariance of two variables divided by the product of their standard deviations. Its values are between -1 and 1, where -1 means completely negative linear correlation, 1 means completely positive linear correlation, and 0 means no linear correlation. A Pearson's correlation value greater than 0.7 indicates high collinearity [19-20]. VIF is the quotient of the variance in a model with multiple terms by the variance of a model with one term alone. It is the reciprocal of TOL, and measures how much the variance of the estimated coefficient is inflated by multicollinearity. A large VIF value indicates that the associated independent variable is highly collinear with other independent variables. A TOL value smaller than 0.1 and a VIF value greater than 10 means that the multicollinearity is problematic. In this case, the calculated correlation coefficient, TOL, and VIF for the remaining numerical conditioning factors are listed in Tables 1 and 2, from which it can be seen that there are no collinearity problems among these conditioning factors.

Table 1 – Correlation coefficients between each two conditioning factors

	Elevation	**Slope gradient**	**Slope aspect**	**Standard curvature**	**PSA03**	**Precipitation (1 week)**
Elevation	1.0000	0.1679	0.0575	0.0568	-0.3229	0.5682
Slope gradient	--	1.0000	0.0879	0.006	0.1961	-0.1751
Slope aspect	--	--	1.0000	-0.0067	0.0483	-0.0455
Standard curvature	--	--	--	1.0000	-0.0642	0.0652
PSA03	--	--	--	--	1.0000	-0.5650
Precipitation (1 week)	--	--	--	--	--	1.0000

Table 2 –Tolerance (*TOL*) and variance inflation factor (*VIF*) values for the conditioning factors

Item	**Elevation**	**Slope gradient**	**Slope aspect**	**Standard curvature**	**PSA03**	**Precipitation (1 week)**
TOL	0.8385	0.9767	0.9998	1.0000	0.8901	0.7328
VIF	1.1926	1.0239	1.0002	1.0000	1.1235	1.3646

4. Landslide Hazard Model Construction

After data pre-processing, preparation, and analyses, there are eight landslide conditioning factors left, including six numerical variables (i.e., elevation, slope gradient, slope aspect, standard curvature, PSA03, and 1-week cumulative precipitation before the disaster) and two nominal variables (i.e., geology and surface vegetation). Applying these conditioning factors as independent variables and landslide inventory as the dependent variable, a landslide hazard model was constructed by logistical regression. Logistic regression is a multivariate analysis approach that can be used to model the relationship between a dichotomous dependent variable and a set of independent variables. It is efficient, highly interpretable, and does not require the independent variables to have a normal distribution. Moreover, the independent variables in logistic regression can be either continuous, discrete, or any combinations of them. A logistic function can be written as Eq.1, with $\hat{p}$ representing the probability of an event occurring (the probability of landslide occurrence in this case), $\hat{\alpha}$ representing the intercept, and β_i representing the coefficient for the independent variable x_i.

$$P(Y=1)=\hat{p}=\frac{1}{1+e^{-(\hat{\alpha}+\widehat{\beta_1}x_1+\widehat{\beta_2}x_2+\cdots+\widehat{\beta_n}x_n)}} \quad (1)$$

In order to construct and evaluate the landslide hazard model by logistic regression, the entire data were randomly split into a ratio of 7:3, with 70% of the data applied for model training and 30% of the data applied for model validating. The regressed logistic function is shown in Eq.2, which achieved an accuracy of 92.82% and a Kappa statistic of 0.86 in this case. Applying this logistic function, a landslide hazard map in the studied Iburi area was generated and shown in Fig.6. In the future application, by updating the changed factors, that is, the rainfall and earthquake conditions, it may provide some valuable information for landslide hazard prediction, monitoring, and management in the study area.

$$P(Y=1)=\hat{p}=\frac{1}{1+e^{0.0036\,F1-0.0418\,F2-0.0002\,F3+0.0553\,F4+0.3512\,F5+0.2458\,F6+F7+F8-1.3584}} \quad (2)$$

Here $\hat{p}$ is the probability of landslide occurrence, F_1 is elevation, F_2 is slope gradient, F_3 is slope aspect, F_4 is standard curvature, F_5 is PSA03, F_6 is 1-week cumulative precipitation before the disaster, F_7 is geology, and F_8 the surface vegetation.

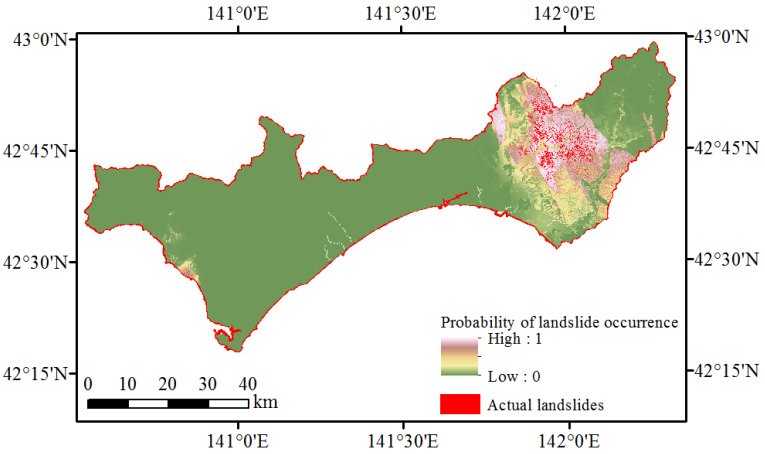


Fig. 6 – Generated landslide hazard map

5. Conclusions

Applying the massive event-based landslides triggered by the 2018 Hokkaido Eastern Iburi Earthquake, a landslide hazard analysis was carried out by the GIS-based multivariate statistical approach. The landslide inventory and corresponding causative factors of the landslides, including topography, geology, surface vegetation, rainfall, and ground motion characteristics, were collected, preprocessed, analyzed, and constructed into a database in the GIS platform, as dependent and independent variables for analysis. A landslide hazard model was constructed by logistic regression using these collected variables, which achieved good results. It is expected to provide some valuable information for the prediction, monitoring, and management of future landslides in the study area by updating the changed causative factors such as the rainfall conditions.

6. Acknowledgements

The authors appreciate the financial support from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT).

7. References

[1] Cruden DM (1991): A simple definition of a landslide. *Bulletin of Engineering Geology and the Environment*, **43** (1), 27-29.

[2] USGS. Available online: https://www.usgs.gov/faqs/what-a-landslide-and-what-causes-one?qt-news_science_products=0#qt- news_science_products (Last accessed on 23 December 2019).

[3] Hong H, Pradhan B, Xu C, Bui DT (2015): Spatial prediction of landslide hazard at the Yihuang area (China) using two-class kernel logistic regression, alternating decision tree and support vector machines. *Catena*, **133**, 266-281.

[4] USGS. Available online: https://earthquake.usgs.gov/earthquakes/eventpage/us2000h8ty/executive#executive (last accessed on 23 December 2019).

[5] Osanai N, Yamada T, Hayashi S, Kastura S, Furuichi T, Yanai S, Murakami Y, Miyazaki T, Tanioka Y, Takiguchi S, Miyazaki M (2019): Characteristics of landslides caused by the 2018 Hokkaido Eastern Iburi Earthquake. *Landslides*, **16** (8), 1517–1528.

[6] Yamagishi H, Yamazaki F (2018): Landslides by the 2018 Hokkaido Iburi-Tobu earthquake on September 6. *Landslides*, **15** (12), 2521–2524.

[7] Ge P, Gokon H, Meguro K, Koshimura S (2019): Study on the Intensity and Coherence Information of High-Resolution ALOS-2 SAR Images for Rapid Massive Landslide Mapping at a Pixel Level. *Remote Sensing*, **11** (23), 2808.

[8] Chau KT, Sze YL, Fung MK, Wong WY, Fong EL, Chan LCP (2004): Landslide hazard analysis for Hong Kong using landslide inventory and GIS. *Computer & Geoscience*, **30** (4), 429-443.

[9] Lee C (2015): Review and perspectives on methodology for landslide hazard analysis. *10th Asian Regional Conference of IAEG*, Kyoto, Japan.

[10] King G, Zeng L (2001): Logistic regression in rare events data. *Political Analysis*, **9** (2), 137-163.

[11] Zhang S, Li R, Wang F, Iio A (2019): Characteristics of landslides triggered by the 2018 Hokkaido Eastern Iburi earthquake, Northern Japan. *Landslides*, **16** (9), 1691-1708.

[12] Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006): Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). *Geomorphology*, **76** (3), 392-410.

[13] Bai S, Wang J, Lü G, Zhou P, Hou S, Xu S (2010): GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. *Geomorphology*, **115** (1-2), 23-31.

[14] Süzen ML, Doyuran V (2004): Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Engineering Geology*, **71** (3-4), 303-321.

[15] Nefeslioglu HA, Duman TY, Durmaz S (2008): Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). *Geomorphology*, **94** (3-4), 401-418.

[16] Ayalew L, Yamagishi H, Ugawa N (2004): Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. *Landslides*, **1** (1), 73-81.

[17] Domínguez-Cuesta MJ, Jiménez-Sánchez M, Berrezueta E (2007): Landslides in the Central Coalfield (Cantabrian Mountains, NW Spain): Geomorphological features, conditioning factors and methodological implications in susceptibility assessment. *Geomorphology*, **89** (3-4), 358-369.

[18] Lee CT (2014): Statistical seismic landslide hazard analysis: An example from Taiwan. *Engineering Geology*, **182**, 201-212.

[19] Bui DT, Tuan TA, Klempe H, Pradhan B, Revhaug I (2016): Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, **13**, 361-378.

[20] Booth GD, Niccolucci MJ, Schuste EG (1994): *Identifying proxy sets in multiple linear regression: an aid to better coefficient interpretation*. U.S. Dept. of Agriculture, Forest Service, Ogden.

