京都大学防災研究所年報 第 62 号 B
DPRI Annuals, No. 62 B, 2019

Landslide Hazard Assessment in Western Japan Using Logistic Regression Analysis with Hydrometeorological Factors

Ying-Hsin WU and Eiichi NAKAKITA

Synopsis

We attempt to investigate the efficiency of applying a lithology factor with high-resolution XRAIN observation to accurate landslide hazard estimation. This study presents a model of landslide mapping using logistic regression with geological and high-resolution hydrometeorological factors, and analyzes hazardous conditions of landslide disasters occurred in Kure, Hiroshima during the heavy rainfall event in July of 2018. The same as the practical method of landslide early warning, the hydrometeorological factors are hourly cumulative rainfall and soil-water index calculated by using XRAIN data. The lithology factor is derived from the seamless geological map. As a first trial, the model was simply calibrated using linear logistic regression on a recent landslide inventory composing of 727 events in Chugoku Region after 2012. 85% and 15% of events are used for training and accuracy test, and the calibrated model achieves a high accuracy of 91.3%. To verify, our model was applied to estimate landslide occurrence during H30.07 heavy rainfall in Kure, Hiroshima. The result verified our model can estimate highly accurate occurrence location.

Keywords: landslide, logistic regression, XRAIN, lithology, hydrometeorology, heavy rainfall disasters

1 Introduction

In the last two years, fronts and typhoons brought huge amount of rainfall in Kyushu and Chugoku regions to trigger numerous severe inundation and landslide disasters and to cause fatalities and much property damages. For successful early warning of landslide and debris flow hazards, governmental authorities, e.g., Japan Meteorological Agency (abbreviated as JMA), Ministry of Land, Infrastructure, Transport and Tourism (MLIT) or local governments, assess hazard occurrence by using the famous method of critical line which utilizes high-resolution radar rainfall data (Osanai et al., 2010). The present method can practically forecast landslide occurrence by judging the motion of a temporal snake line on a phase plane composed of long-term soil-water index [mm], short-term hourly cumulative rainfall [mm/hr], and critical lines calibrated by past events using nonlinear regression of radial basis function network (Kuramoto et al., 2001). The method of critical line has been proven very effective and robust for disaster discrimination and prevention (Osanai et al., 2010; Watanabe et al., 2018).

Taking the advantage of the robustness of the critical line method, we mainly attempt to investigate how the estimation efficiency can be acquired by explicitly involving an additional factor representing geological property for hazard prediction, because the influence of geological setting is implicitly considered in present critical lines so far. Also, as a minor concern, assessment using the present method is based on a coarser spatial resolution (1 by 1 km) which may cause estimations lose important information of concentrated and localized rainfall which may facilitate shallow landslide occurrence. So, for criterion determination, one motivation of this study is to utilize high-resolution rainfall data observed by XRAIN, which is the operating eXtended RAdar Information Network for rainfall observation covering almost all Japan regions. XRAIN observation is unique and advantageous because of its wide coverage and high resolution comparing to any other weather radar system in the world. However, XRAIN data is just available since 2012, so a recent landslide inventory may be required to determine new corresponding criterion for hazard prediction.

For the two aforementioned concerns, as a first trial, this study continues to utilize the same two hydrometeorological indexes with an additional parameter of lithology type. Because the lithological factor is a categorical variable, the famous logistic regression is applied for constructing our model of classification, because it is the most preferred method among all statistical models of landslide susceptibility mapping according to the latest literature review (Reichenbach et al., 2018). To explain again, as a first trial, attempting to understand the efficiency of incorporating a lithological factor with high-resolution radar rainfall data for landslide prediction, we will simply apply linear logistic regression (Menard, 2001), which is proven to be still robust in landslide-prone area (Chang et al., 2007). With the help of well-calibrated present critical lines, we then reveal how linear logistic regression with a lithological factor and being calibrated by a recent landslide inventory estimates landslide occurrence. Finally, the latest shallow landslide disasters occurred in Kure, Hiroshima by heavy rainfall in early July of 2018 were used for verification.

2 Data and Processing

One main purpose of this study is to apply high resolution XRAIN observation for regional landslide prediction. Because the high-resolution XRAIN composite observation (250 by 250 m in one minute) is available only after 2012, we extracted 727 real landslide events occurred in Shimane, Yamaguchi, and Hiroshima prefectures in from 2012 to 2014 as a landslide inventory for model calibration. The landslide inventory is mainly composed of landslide disasters occurred in late July of 2013 at the boundary of Shimane and Yamaguchi prefectures (Wang et al., 2014), debris flow disasters in late August of 2014 in northern Hiroshima city (Wang et al., 2015), and some other disasters occurred in Shimane prefecture from 2012 to 2014. All disasters are basically rainfall-triggered shallow landslides. The information of occurrence timing and location of each disaster was collected for calculating corresponding hourly cumulative rainfall and soil-water index. As aiming to apply logistic regression, we randomly extracted other 727 non-occurrence places to collect factors for calibration through ArcGIS bulit-in functions. Locations of all events in landslide inventory are illustrated in Fig. 1. Referring to our target area, we downloaded the precipitation datasets observed by Nogaibara and Oshio radar stations from Data Integration & Analysis System (DIAS). The boundary of XRAIN data we used is shown in Fig. 1. For downloading rainfall data corresponding to the rainfall event of each event in the landslide inventory, the starting timing of computing hourly cumulative rainfall and soil-water index was set to be five days before the occurrence timing.

By following the spatial resolution of XRAIN data, the grid size is also set to be 250 by 250 m. Then, for each event, the computed soil-water index and hourly cumulative rainfall at occurrence timing were extracted for regression analysis, and the distribution and corresponding histograms of all samples are illustrated in Fig. 2. We can easily observe that the most event number appears at 220 mm of soil-water index for landslide occurrence, and the number of non-occurrence samples possesses a decreasing tendency to zero until the soil-water index approaches about 120 mm. However, no concentrated tendency of the factor of hourly cumulative rainfall exists for the occurrence samples even the rainfall intensity approaches less than 20 mm/hr. The scatter diagram of Fig. 2 shows no clear boundary classifying hazard occurrence, so herein we would like to introduce a lithology factor for possible discrimination of hazard occurrence. For this goal, referring to the coordinates of occurrence location of each event, we extracted corresponding lithology information from seamless digital geological map of Japan published by Geological Survey of Japan (retrieved from https://gbank.gsj.jp/seamless/).

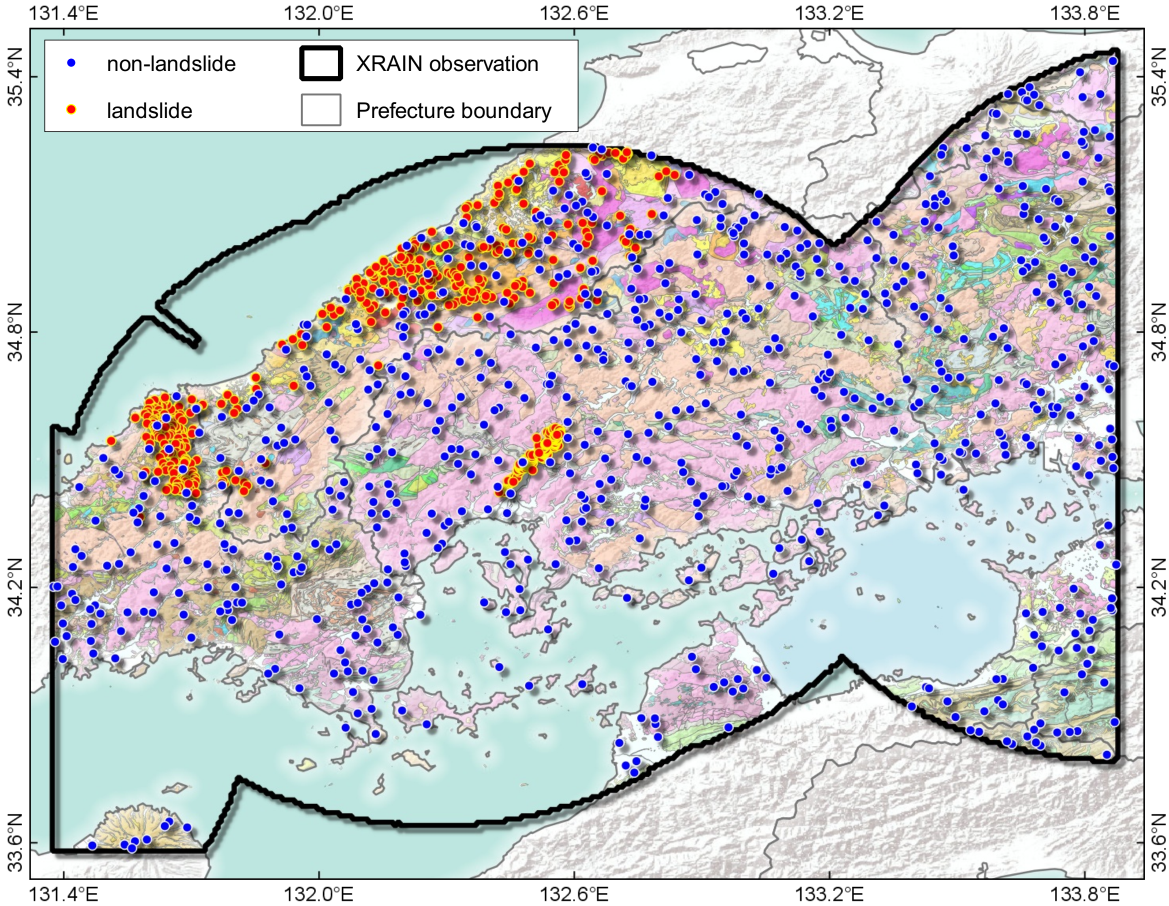

Figure 1: Map of landslide inventory and target area. Black thick line indicates the range of XRAIN data. Solid red and blue circles denote the landslide and non-landslide events both of which are respectively composed of 727 events. The lithological settings are demonstrated using the official legends and colors published by Geological Survey of Japan (retrieved from https://gbank.gsj.jp/seamless/legend_shosai_e.html).

Table 1: Logistic regression coefficients and the significance of explanatory variables. Coefficients of categorical variable GEO for different lithology are tabulated in Table 2.

Explanatory variables	Coefficients	p-value
Soil Water Index (SWI)	0.0147	0.0000
Hourly rainfall (RAIN)	0.0687	0.0000
Constant β_0	-5.5118	0.415

3 Logistic Regression Analysis

3.1 Model Calibration and Test

In this work, the linear logistic regression model considers three variables, i.e., hourly cumulated rainfall (abbreviated as RAIN), soil water index (SWI), and lithology type (GEO). The logit function and probability function respectively read

$$\text{Logit}(Y) = \beta_0+\beta_1\times\text{RAIN}+\beta_2\times\text{SWI}+\text{GEO}, \quad (1)$$

$$P = \frac{1}{1+\exp^{-\text{Logit}(Y)}}, \quad (2)$$

where β_0, β_1, and β_2 are the intercept and coefficients to calibrate, and GEO is a categorical variable representing different lithological types. 85% of the datasets, or says 1,235 events, were used for training, and the rest, i.e., 219 events, for accuracy test. All calibrated coefficients are tabulated in Tables 1 and 2.

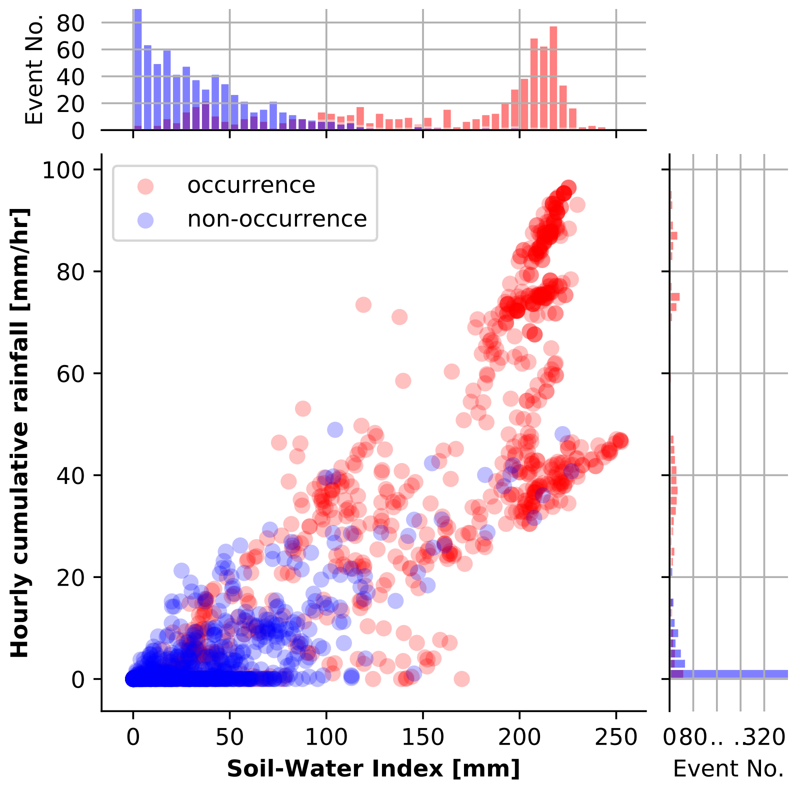

Figure 2: Scatter diagram and histograms of all samples

With p-values approaching 0, the coefficients of RAIN and SWI shows the significance of hydrometeorological factors for classifying landslide occurrence. But as having a larger p-value, β_0 is not significant for classification. However, this condition could be compensated by GEO, the categorical explanatory variable representing the feature of lithology at each grid, as will be explained by using the two most frequent lithology types shortly.

To validate our model, the accuracy test was performed using 219 events (15 %) of all datasets. The test shows the accurate ratio of prediction reaches 91.3%, and the results of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are respectively tabulated in Tab. 3. Having Confirmed the reliability of classification of our model, the area under the curve (AUC) of receiver operating characteristic (ROC) is 0.968, as is shown in Fig. 3. The test verified both of the accuracy and applicability of our model.

3.2 Model Result and Discussion

Taking examples of the two most frequent lithological types, we can easily observe the capability of

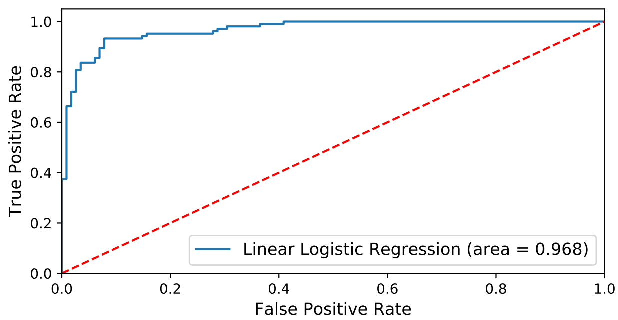

Figure 3: Curve of Receiver Operating Characteristic (ROC). Area under the curve (AUC) is 0.968.

our model for classifying or estimating landslide occurrence. The two frequent types are Late Cretaceous granite and Late Cretaceous non-alkaline felsic volcanic rocks, as are illustrated in Figs. 4 5, respectively. The results show different estimation of occurrence probability for the two types. In the example of the most frequent lithological type of Late Cretaceous granite, the present critical lines can capture most of events, and our model can also discriminate the occurrence with some probability estimation (Please see Fig. 4). But in the example of the sec-

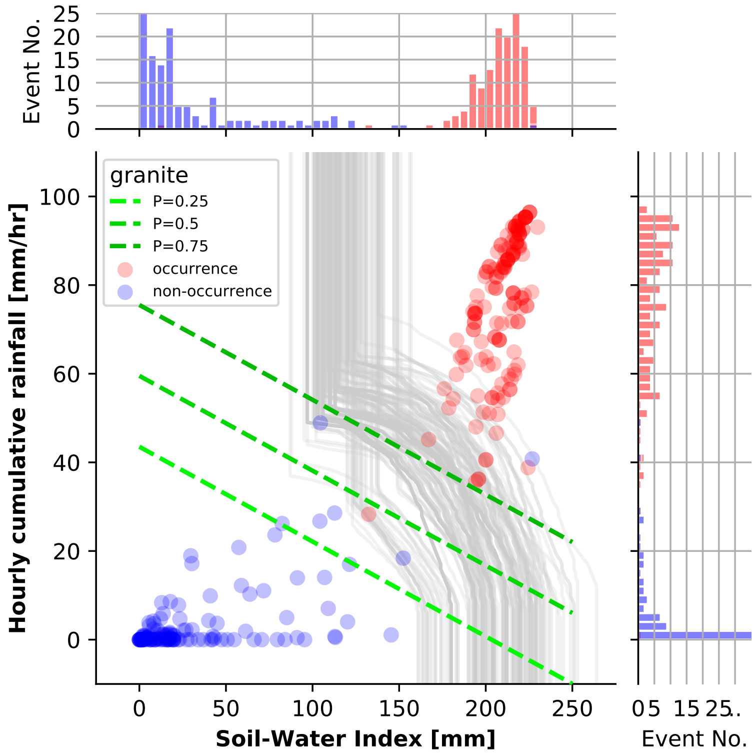

Figure 4: Scatter diagram and histograms of the grid mainly consisted of Late Cretaceous granite. Light to deep green lines respectively denote the estimated probability of 0.25, 0.5, and 0.75 using our calibrated model. Gray lines denote relating critical lines.

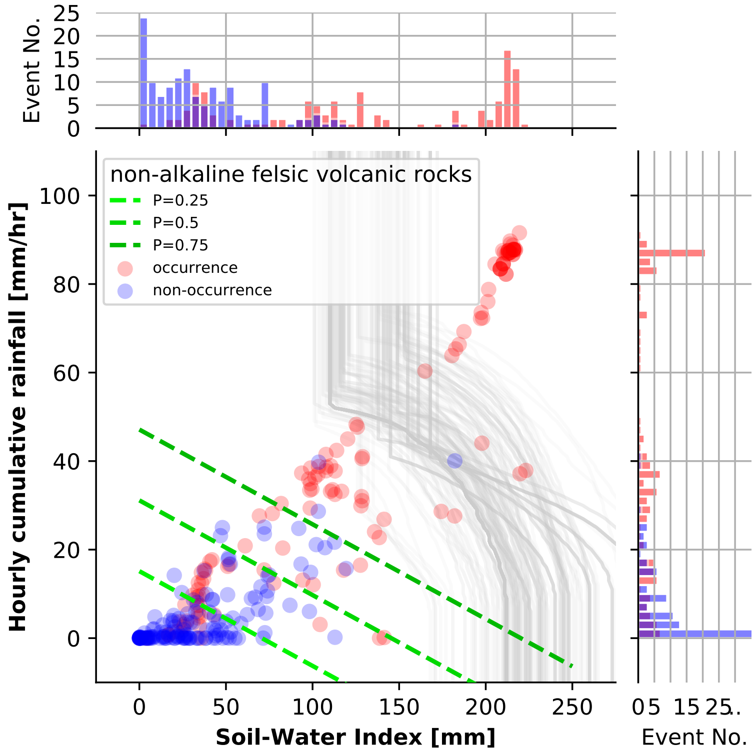

Figure 5: Scatter diagram and histograms of the grid mainly consisted of Late Cretaceous non-alkaline felsic volcanic rocks. Light to deep green lines respectively denote the estimated probability of 0.25, 0.5, and 0.75 using our calibrated model. Gray lines denote relating critical lines.

Table 2: Coefficients of categorical explanatory variable of GEO for logistic regression

Lithology [Index]	Coefficients	Lithology [Index]	Coefficients
LP to H marine and non-marine sediments [0]	3.4124	LP middle terrace [2]	5.8069
LC marine sandy turbidite [4]	0.7751	E to LC granodiorite [5]	0.4326
LC granite [7]	1.4236	mafic schist (Sambagawa) [10]	0.9329
pelitic schist (Sambagawa) [11]	0.6297	ME granite [12]	4.1624
felsic plutonic rocks [13]	2.2165	EM to MM non-alkaline felsic volcanic rocks [14]	7.9244
LC non-alkaline felsic volcanic rocks [16]	3.3738	LC felsic plutonic rocks [18]	0.0090
EM to MM non-alkaline mafic volcanic rocks [24]	4.9610	M to LM non-alkaline mafic volcanic rocks [29]	8.7735
LP to H swamp deposits [31]	6.2777	LP non-alkaline pyroclastic flow volcanic rocks [32]	0.5779
LE to EO marine and non-marine sediments [33]	0.4641	EM to MM marine and non-marine sediments [34]	0.3843
EP non-alkaline pyroclastic flow volcanic rocks [35]	0.9619	MP non-alkaline pyroclastic flow volcanic rocks [36]	0.5420
Sangun-Suo Metamorphic Rocks (pelitic schist) [45]	3.2872	M to LM non-alkaline felsic volcanic rocks [49]	0.1858
ultramafic rocks [53]	-0.6549	EP marine and non-marine sediments [54]	3.3864
H reclaimed land [58]	0.7398	EP volcanic debris [61]	0.6801
Sambagawa Metamorphic Rocks (mafic schist) [63]	0.3588	melange matrix of P1 accretionary complex [64]	0.8529
melange matirx of E to MJ accretionary complex [66]	0.3456	Sambagawa Metamorphic Rocks (pelitic schist) [67]	0.2296
C1 to P1 limestone block of C1 to P1 accretionary complex [73]	-0.784	MP non-alkaline mafic volcanic rocks [74]	0.4018
LC marine muddy turbidite [75]	0.6446	Sangun-Chizu Metamorphic Rocks (pelitic schist) [76]	-0.0509
LC mafic plutonic rocks [77]	2.1995	LC granodiorite [78]	0.9491
P1 marine sedimentary rocks [79]	3.7936	LC non-marine sediments [81]	6.1354
LC non-alkaline mafic volcanic rocks [82]	2.9519	LC marine turbidite [83]	0.6797
P2 to EE granite [85]	3.4242	LP to H fan deposits [86]	2.5137
P2 to EE granodiorite [87]	3.3777	LP lower terrace [88]	4.0953
sandstone of M to LJ accretionary complex [91]	0.2964	Ryoke Metamorphic Rocks (siliceous gneiss) [94]	-0.6517
MP marine and non-marine sediments [100]	0.4197	Q volcanic debris [104]	2.7688
EC non-marine sedimentary rocks [106]	0.6559	LC non-alkaline felsic volcanic intrusive rocks [107]	2.7749
T to MJ chert block of M to LJ accretionary complex [109]	0.6977	P gabbro and diorite in accretionary complex [110]	3.0581
ME non-alkaline felsic volcanic rocks [139]	3.2200	Sambagawa Metamorphic Rocks (mafic schist)[160]	-0.3628
Sambagawa Metamorphic Rocks (siliceous schist) [161]	0.7205	C1 to P1 chert block of C1 to P1 accretionary complex [168]	6.8398
Sangun-Chizu Metamorphic Rocks (mafic schist) [169]	0.2935	E to LC granite [176]	0.5898
Sangun-Suo Metamorphic Rocks (mafic schist) [177]	1.7954	Sambagawa Metamorphic Rocks (psammitic schist) [192]	0.3970
E to LC mafic plutonic rocks [196]	0.9294	late EC non-marine sedimentary rocks [199]	6.2387
M to LT marine and non-marine sedimentary rocks [202]	0.2278	sandstone of P1 accretionary complex [203]	3.1896
Ryoke Metamorphic Rocks (pelitic gneiss) [204]	0.2020	Sangun-Suo Metamorphic Rocks (psammitic schist) [206]	-0.6155
Ryoke Metamorphic Rocks (pelitic gneiss) [207]	0.6130	LC marine conglomerate [208]	-0.0205
LE to EO non-alkaline felsic volcanic rocks [217]	5.6623	EP non-alkaline felsic volcanic rocks [218]	5.4026
Sangun-Suo Metamorphic Rocks (siliceous schist) [219]	2.6703	LE to EO non-alkaline mafic volcanic rocks [220]	3.7878
LE to EO mafic plutonic rocks [221]	4.1112	LE to EO granodiorite [222]	6.5324
LE to EO granite [224]	4.8099	ME non-alkaline mafic volcanic rocks [225]	4.6947
P1 mafic volcanic rocks in accretionary complex [226]	0.6712	ME mafic plutonic rocks [228]	4.2461
P2 to EE non-alkaline felsic volcanic intrusive rocks [232]	0.3066	E to MM mafic plutonic rocks [244]	4.6485
P2 to EE mafic plutonic rocks [303]	4.9080	Sangun-Renge Metamorphic Rocks (schist) [325]	0.2367

Abbreviations of Geological ages: E: Early, M: Middle, L: Late, P: Pleistocene, H: Holocene, C: Cretaceous, E: Eocene, M: Miocene, O: Olligocene, J:Jurassic, C1: Carboniferous, P1: Permian, P2: Paleocene, Q: Quaternary, T: Triassic

Table 3: Confusion matrix of training results. The accuracy test of our logistic regression model is 91.3%. (TP: True positive; FP: False positive; FN: False negative; TP: True negative.)

		Predicted results	
		Positive	Negative
Actual events	Positive	106 (TP)	9 (FN)
	Negative	10 (FP)	94 (TN)

ond frequent lithological type of Late Cretaceous non-alkaline felsic volcanic rocks in Fig. 5, some events are not satisfactorily classified because they occurred in rather low hourly cumulative rainfall or soil-water index. Particularly, Figure 6 illustrates one example event of rainfall-triggered shallow landslide occurred nearby Kakinmoto Shrine in Yamaguchi prefecture (132.236E, 34.840N) at around 9:00 JST on 28 July 2018. The lithology there is just Late Cretaceous non-alkaline felsic volcanic rocks (see Fig. 5). The snake line of this event only evolved in the region of which the soil-water index is less than the value of JMA yellow alert. However, the hazard tendency can be suc-

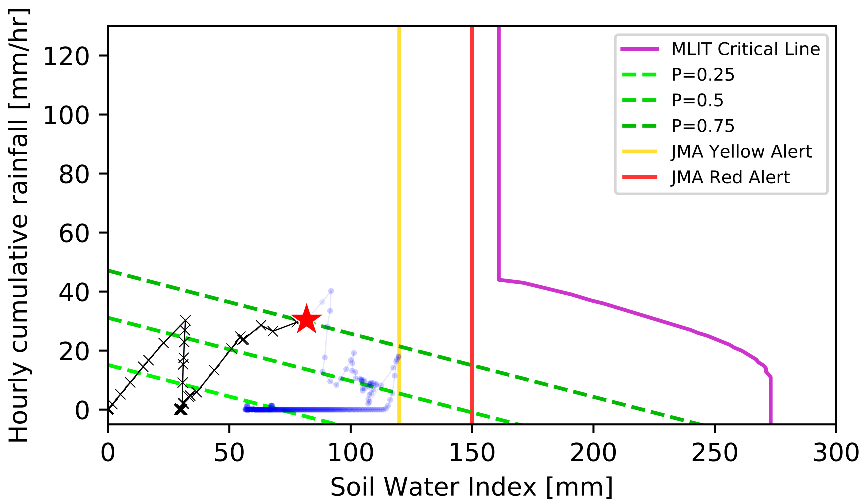

Figure 6: Comparison of critical lines and our model estimation of the event occurred nearby Kakinmoto Shrine in Yamaguchi prefecture (132.236E, 34.840N). Red star denotes the real occurrence timing (around 9:00 JST on 28 July 2013).

cessfully estimated through our calibrated model. Although more verification is necessary, this case could reflect that inclusion of lithological setting can help estimate hazardous conditions, and is suggested to be considered for prediction. Here it also reveals that our model could discriminate hazardous condition for different lithology types, which is just one of the purposes of this study.

4 Estimation of Landslide Disasters Occurred in Kure During H30.07 Heavy Rainfall

4.1 Background

Brought by stationary front and Typhoon No.7, a record-breaking heavy rainfall occurred in western Japan from June, 28 to July, 8 in 2018, and triggered serious floods and massive shallow landslides in many areas. To verify again our calibrated model, we attempt to estimate hazardous conditions of landslides in Kure city in Hiroshima. In our target area, there were 2,934 shallow landslides, published by Geospatial Information Authority of Japan (retrieved from http://www.gsi.go.jp/BOUSAI/H30.taihuu7gou.html), as the distribution is shown in Fig. 8. The period of soil-water index calculation spans from 00:00 JST on 20 June 2018 to 00:00 JST on 10 July 2018. The lithological type of most area at each grid was extracted as the representative one. The, our model was applied to estimate the occurrence probability of time variation everywhere in the target area.

4.2 Result and Discussion

In general practice (Chang et al., 2007), the value 0.5 of logistic regression estimation is used to reflect landslide occurrence. To confirm the performance of our model estimation, we varied the criterion value from 0.5 to 0.99, accumulated all grids having the estimated probability greater than or equal to the criterion, and counted the total number of grids which match the real disaster locations. Then, the successful estimation rate is determined by the fraction of the matched grid number to total disaster locations of 2,934. Figure 7 shows the successful estimation rate under different criterion values. The rate remains 100% until the criterion of 0.56, and decreases as the criterion increases. The rates are 99.2%, 95.5%, and 85.9% while the criterion are set to be 0.7, 0.8 and 0.9, respectively. This means that our model gives highly accurate estimation. However, as is illustrated in Fig. 8, the distribution of estimated disasters under the criterion of 0.9 obviously overestimates disastrous locations.

5 Concluding Remarks

This research elaborated how inclusion of a lithological factor with high-resolution rainfall data provided by XRAIN can influence regional landslide hazard prediction. For applying XRAIN data, a recent landslide inventory consisting of 727 events in Chugoku region was prepared. As a first trial, a simplified linear logistic regression has been successfully ap-

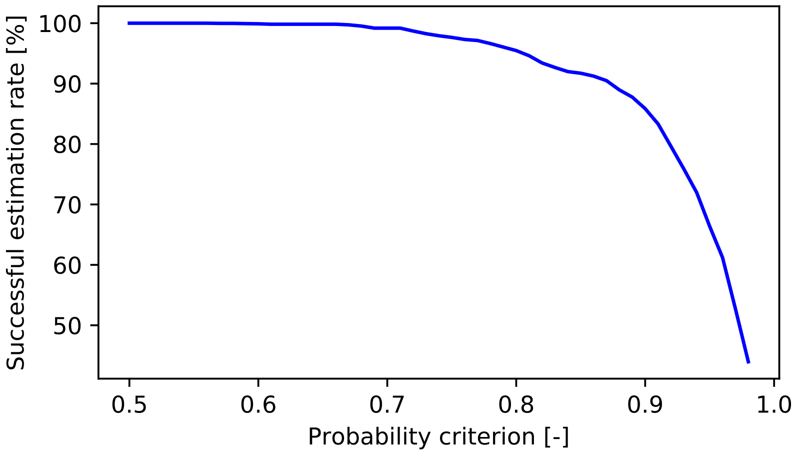

Figure 7: Successful estimation rate of occurrence of landslides in Kure during the H30.07 heavy rainfall.

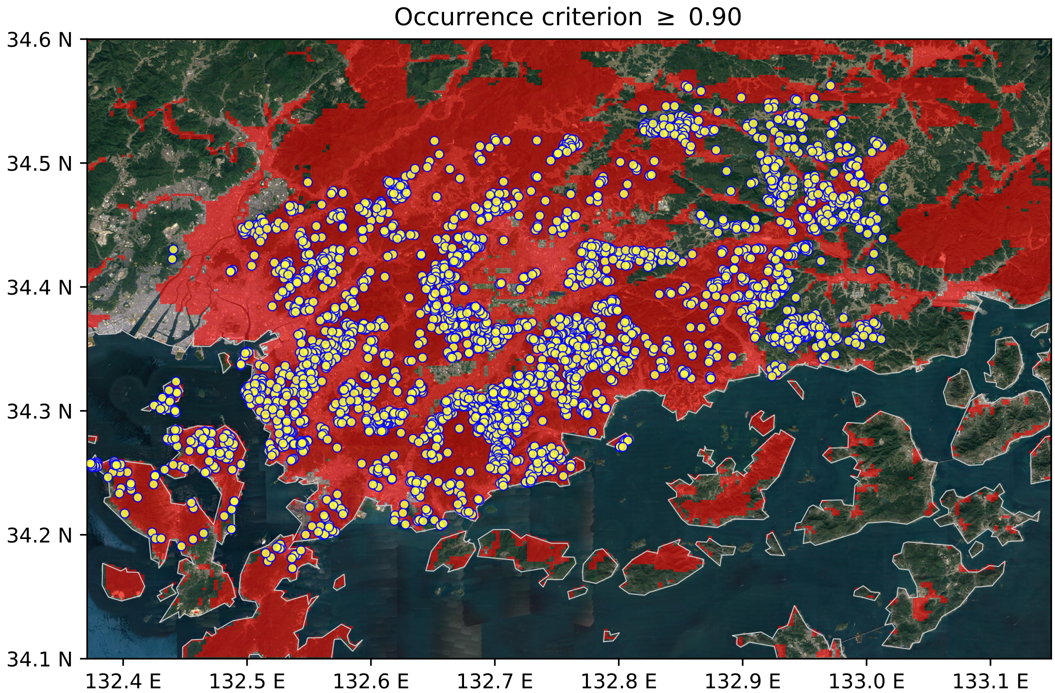

Figure 8: Comparison of estimated results and real events in Kure, Hiroshima while the criterion of landslide occurrence is set to be greater than or equal to 0.90 using our calibrated model (transparent red patches). The successful estimation rate is 85.9%. Solid yellow circles are real disaster places published by Geospatial Information Authority of Japan. The background aerial photo was retrieved from google map.

plied for estimating occurrence probability. Even only linear logistic regression was used for classification, our model clearly demonstrates the capability of landslide prediction considering the geological setting with high-resolution radar rainfall data. As being capable of hazard discrimination, our calibrated model is planned to be used for assessment of landslide hazard under climate change effects. Meantime, as our calibrated linear regression is too simplified, to achieve more accurate estimation, the last and most important thing is to applying nonlinear regression and quantifying uncertainty as our next steps of future improvement, and to performing assessment of the future tendency of landslide hazards.

Acknowledgements

The research is financially supported by *Theme D Integrated Hazard Prediction* in *Integrated Research Program for Advancing Climate Models* (TOUGOU). For the valuable critical line informationThe authors appreciate MLIT and the governments of the prefec-

tures which are covered in our analysis area. Values of yellow and red alerts were from JMA website. XRAIN data was retrieved from *Data Integration & Analysis System* (DIAS). Seamless Digital Digital Geological Map of Japan we used is published by Geological Survey of Japan. Some disaster events can be retrieved from the Center for Natural Disaster Reduction Research and Eduction (https://www.geo.shimane-u.ac.jp/wangfw/database/homepage/) at Shimane University.

References

Chang, K.-T., et al. (2007): Modeling typhoon- and earthquake-induced landslides in a mountainous watershed using logistic regression, *Geomorphology*, Vol. 89, pp. 335-347.

Kuramoto, K., et al. (2001): A study on a method for determining non-linear critical line of slope failures during heavy rainfall based on RBF network, *Doboku Gakkai Ronbunshu.*, Vol. 672, pp. 117-132.

Menard, S. (2001): *Applied Logistic Regression Analysis*, 2nd ed., Sage Publication.

Osanai, N., et al. (2010): Japanese early-warning for debris flows and slope failures using rainfall indices with Radial Basis Function Network, *Landslide*, Vol. 7, pp. 325-338.

Reichenbach P., et al. (2018): A review of statistically-based landslide susceptibility models, *Earth-Sci. Rev.*, Vol. 180, pp. 60-91.

Wang, F., et al. (2014): Landslides triggered by a heavy rainfall at the boundary of yamaguchi and shimane prefectures in july 2013, *J. Jpn. Landslide Soc.,* Vol. 51, No. 2, pp. 66-69.

Wang, F., et al. (2015): Preliminary investigation of the 20 August 2014 debris flows triggered by a severe rainstorm in hiroshima city, japan, *Geoenvironmental Disasters*, Vol.2, No.1, 17.

Watanabe, A., et al. (2018): Estimation of shallow landslide risk in a mountainous area using a single X-band multi-parameter radar and C-band radars, *J. JSNDS*, Vol.37, No.3, pp. 295-311.

(Received June 17, 2019)