

# Probabilistic Prediction Method of Erosion Volume and Deposition Area from Rainfall Observation Data

Kazuki Yamanoi<sup>1\*</sup>, Kaori Shikakura<sup>2</sup>, Kenji Kawaike<sup>1</sup>, and Satoru Oishi<sup>3</sup>

<sup>1</sup>Kyoto University, Disaster Prevention Research Institute, 8128235, Yoko-oji Shimomisu Higashinokuchi, Fushimi, Kyoto, Japan

<sup>2</sup>Kyoto University, Graduate School of Engineering, 8128235, Yoko-oji Shimomisu Higashinokuchi, Fushimi, Kyoto, Japan

<sup>3</sup>RIKEN Center for Computational Science, 6500047, Minatojima-Minamimachi, Chuo-ku, Kobe, Japan

**Abstract.** We propose a methodology to estimate the spatial distribution of the probability of sediment deposition due to debris flow from rainfall data by combining the probability prediction of erosion volume based on an ordinal logistic regression and a sediment transport simulation. By using the Receiver Operating Characteristic (ROC) curve and Area Under Curve (AUC) we have selected the best combination of short- and long-term rainfall indices used as explanatory variables in the ordinal logistic model. The results showed that the regression model using 60-minute and 48-hour rainfall indices performed well and that the regression model using three events improved the predictability of local disasters in 2014. Furthermore, we performed Monte Carlo debris-flow simulations using rainfall data from 2014 using the model. We confirmed that the spatial distribution of disaster probability is consistent with the actual damage.

## 1 Backgrounds

Generally, the prediction method of sediment-related disasters can be divided into real-time rainfall-based prediction (alerting) and affected-area prediction by using the empirical or physically-based method (hazard mapping). Both methods are already implemented in society. In Japan, sediment disaster alert estimated from SWI (soil water index) calculated by a tank model having three vertical tanks and 60 min rainfall[1]. The warning is mainly issued to the residents in the vulnerable area specified as the sediment disaster (special) precaution zone. However, the false alert rate is large because sediment movement does not take place in all warning areas when the warning is issued, and there are areas within an area where debris flows are relatively difficult to reach[2], resulting in differences in hazards.

It is still difficult to accurately predict the sediment-disaster occurrence and affected area from the rainfall observation because geological structure is complicated. However, the spatial distribution of the damage possibility can be valuable information to minimize the human damage and risk evaluation. This study proposes a methodology to quantify the damage possibility by combining the statistically-based yield volume prediction and sediment transport simulation based on the stony debris flow model.

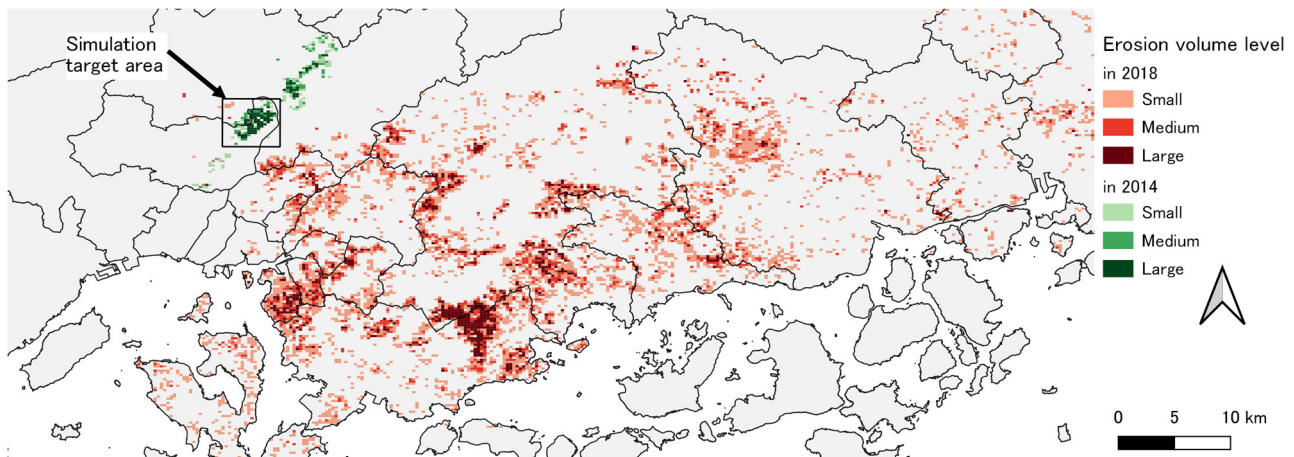
## 2 Statistic prediction of the erosion volume

### 2.1 Stochastic prediction of the debris-flow erosion volume by employing ordinal logistic regression

We set the target area as Hiroshima prefecture, Japan, where two recent large-scale sediment disasters happened. The first one, which occurred in 2014, includes 107 debris flows and 59 shallow slides, which caused 44 injuries and 74 deaths [3]. The second one caused 624 sediment movements in a wider area in Hiroshima prefecture compared to the first event [4].

Total erosion volume is important in evaluating and simulating debris flow. This study selected it as the regression model's objective variable, as in the previous study [5]. The airborne lidar precisely measure the volume by comparing the surface elevation in two periods, usually before and after the debris flow occurrence. This study mainly uses the surface difference data for the 2018 event. This data was obtained by comparing the 1-m DEM taken before the event and taken two-three weeks after the event. Due to the few debris flows that occurred before the 2018 event and the lack of significant rainfall after the event until the measurement, the negative volumes in this differential data can be considered as the amount of erosion caused by debris flow. However, since data were unavailable for some traces of the 2018 event and all traces of the 2014 event, this study also used a method of estimating erosion volume using only trace data for the area where data were unavailable. In the method, the line at the 12 degrees of surface slope [6] divided the trace into erosion area and deposition area. We assumed

\* Corresponding author: [yamanoi.kazuki.6s@kyoto-u.ac.jp](mailto:yamanoi.kazuki.6s@kyoto-u.ac.jp)



**Fig. 1.** Estimated erosion volume levels.

**Table 1.** AUC value predicting the 2014 event's erosion volume magnitudes.

Magnitude	AUC for "short-term RI" - "long-term RI"								
	10m-24h	10m-48h	10m-72h	30m-24h	30m-48h	30m-72h	60m-24h	60m-48h	60m-72h
Small	<b>0.9875</b>	0.9852	0.9844	0.9625	0.9854	0.9846	0.9482	0.9616	0.9553
Medium	0.9873	0.9857	0.9851	0.9815	0.9865	0.986	0.9866	<b>0.9876</b>	0.9871
Large	0.9947	0.994	0.9936	0.9895	0.9948	0.9945	0.9967	<b>0.9967</b>	0.9966

that the area whose slope is larger than 12 degrees has a 1m erosion depth. To treat the erosion volumes and precipitation variables in same resolution, we summarized the erosion volume in the 250m-resolution cells which are same as the radar-based rainfall observation.

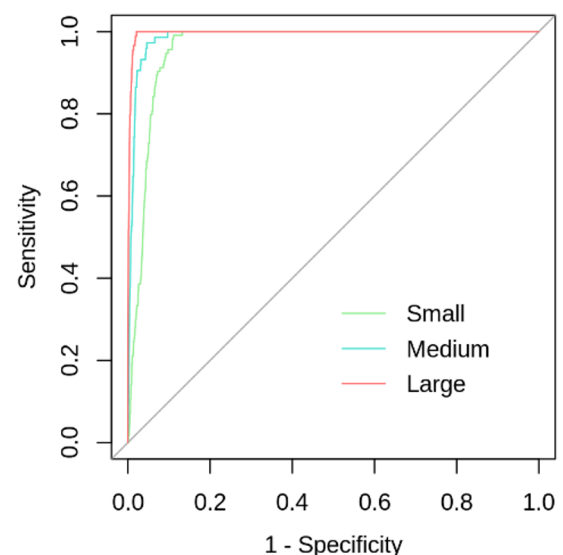
The ordinal logistic regression is a regression method to predict the ordinal value. In this study, we categorized the erosion volume in the mesh system  $V$  into three ranks  $Y$ , small  $Y=1$  ( $0 < V < 980$  [ $m^3$ ]), medium  $Y=2$  ( $980 < V < 2695$  [ $m^3$ ]), and  $Y=3$  large ( $V > 2695$  [ $m^3$ ]), which are the objective variables. The estimated categories for 2018 and 2014 events are shown in Fig. 1. The thresholds were set at the top 10% and 30% of the erosion volumes in the 2018 event.

For the explanatory variables, we employed a mean slope gradient and two rainfall-related indexes, the short-term rainfall index (SRI) and the long-term rainfall index (LRI). In this study, we selected the maximum 10min, 30min, and 60 min rainfall as SRI, and 24 h, 48 h, and 72 h rainfall when the SRI reached the maximum as explanatory variables.

In this study, 2018 data trained the logistic model, and 2014 data tested it. The regression employing nine sets of SRI and LRI yielded the nine ordinal logistic models.

## 2.2 Evaluation of the regression model

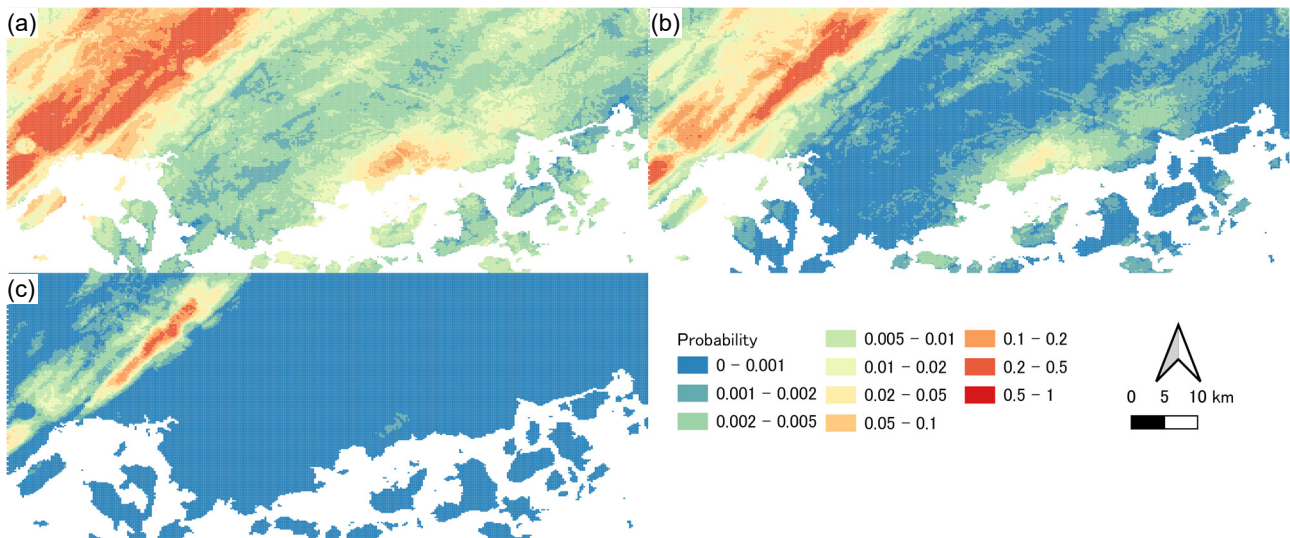
We applied the ROC (Receiver Operating Characteristic) curve and AUC (Area Under Curve) to evaluate the performance of the nine logistic models employing the multiple pair of SRI and LRI. ROC is the relationship curve between the sensitivity (True Positive Rate) and 1-specificity (False Positive Rate) in changing the threshold to divide the positive and negative states. The large AUC, the area under the ROC curve, indicates the high performance of the prediction model, which can



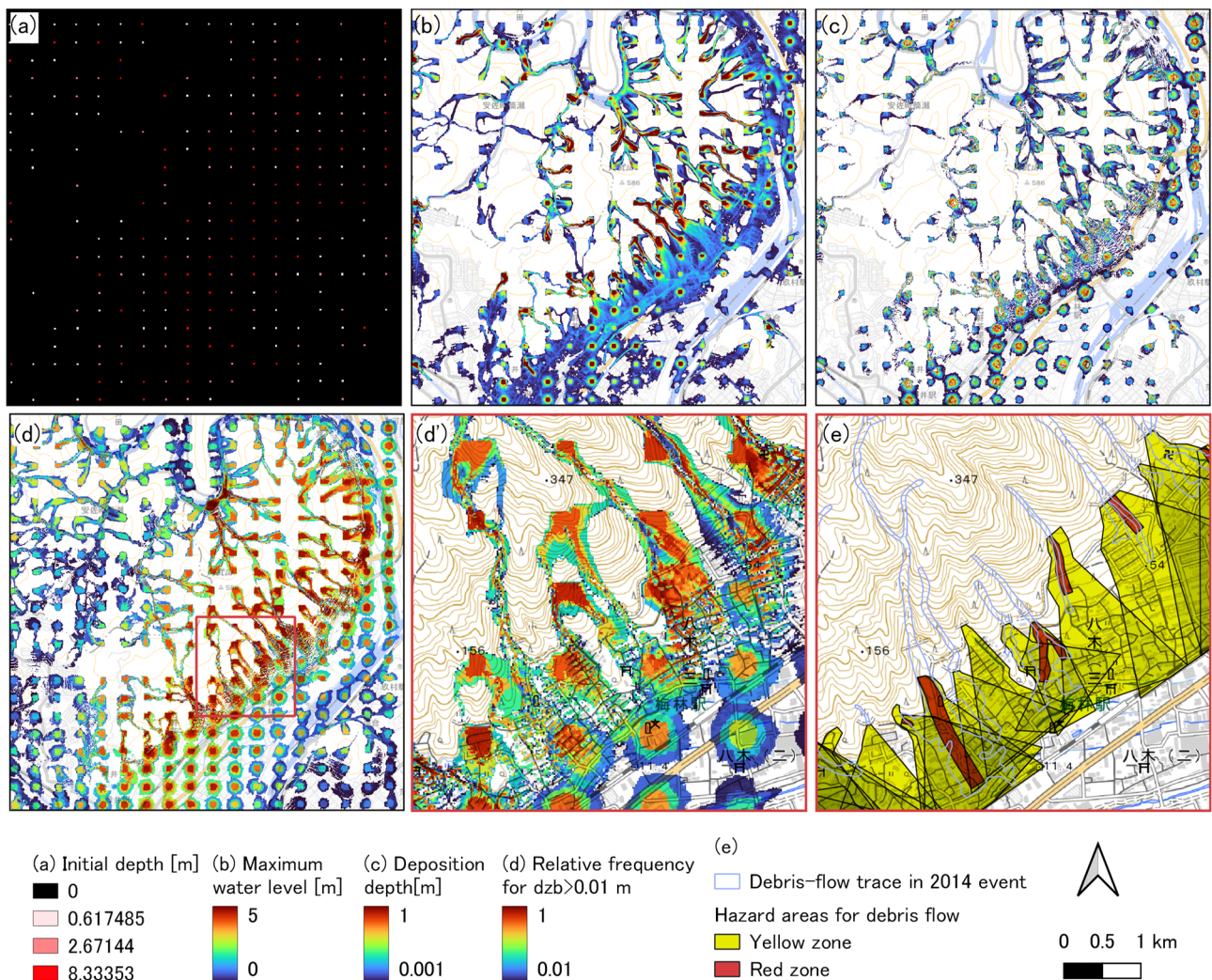
**Fig. 2.** ROC curve for 2014 event by using logistic model employing 60-min and 48-hour rainfall.

accurately divide the positive and negative states. Fig.2 shows the example of the ROC curve of the prediction model created by the 2018 disaster date and applied to the 2014 disaster, which adopts the 60-minute and 48-hour rainfall as SRI and LRI, respectively.

The AUC value is listed in table 1. According to this table, the pair of 60-minute and 48-hour rainfall showed the best score among all pairs predicting medium and large-scale sediment yield. Therefore, we selected this pair to generate the debris-flow simulation input. Note that the precipitation area in 2014 is limited in the target area, thus the FPR can be small with large sensitivity, resulting high AUC value. Also, the all precipitation indexes (i.e. 10 min, 30 min, 60 min, 24 hour, 48 hour, and 72 hour rainfall) around the damaged area were larger than the indexes in the other area. This is the reason why all AUC values exceed 0.94.



**Fig. 3.** Estimated occurrence probability of the small (a), medium (b), and large (c) scale of erosion in the 2014 event employing 60 min and 48-hour rainfall.



**Fig. 4.** Examples of the debris flow simulation input and outputs, generated initial debris flow mass (a), calculated maximum water level (b), and deposition depth (c). (d) and (d') are the preliminary hazard as a relative frequency for the area larger than 0.01 m deposition. (e) is the current practical hazard map for reference.

### 3 Monte-Carlo debris flow runout simulation inputting the predicted erosion volume

Using the observed rainfall in 2014, the regression model yielded the probability of occurrence of three-scale erosion, as shown in fig. 3. Although the higher area of the small-scale probability is much broader than the occurred area, as shown in fig.1, the area with the higher probability of large-scale erosion is consistent with it.

We converted the probability distribution into the debris flow mass by comparing the pseudo-random numbers generated for each grid and three probabilities calculated by the logistic model. Assuming the mass accumulates in the center of the grid, the mass volume was translated into the depth, shown in Fig.4. (a) as an example. Inputting the depth as the initial condition, we performed the debris-flow simulation [7] based on Takahashi's stony debris flow model [8]. The simulation calculated the maximum water (and sediment) level (b) and deposition depth (c). Since the generated volume includes the erosion volume in the stream channels, we neglected the erosion process in the transportation. Note that the logistic model evaluates the possibility in generating the mass from rainfall and topographical indexes neglecting the land use. Therefore, the strange-shaped debris mass was also generated in the city area as a result. On the contrary, the water level and deposition depth around the mountainous valley outlet, where the actual damage was observed, seems consistent with the actual phenomena.

We iterated the mass generation 100 times and conducted a 100-case Monte-Carlo runout simulation, producing 100 patterns of the deposition depth and water level maps. Although there is no convinced method to determine the degree of hazard from predicted variables such as deposition depth and water level, this study regard the area whose deposition depth is greater than the representative grain size (0.01m), which is the input parameter, as the affected area of debris flow. Thus, we organized the 100 patterns into the relative frequency for the deposition depth greater than 0.01 m. The frequency distribution is shown in fig. 4 (d) and (d'). Assuming that there is no other uncertainty than the location and volume of the initial debris-flow mass, and the 0.01m deposition can be a hazard to the population or buildings, this frequency is equivalent to the hazard possibility. This method can estimate the distribution of the quantitative hazard possibility (i.e. preliminary hazard) inside the practical hazard area, as shown in fig.4 (e). Additionally, since this methodology can be applied by employing real-time rainfall observation, such information can be expected to greatly enhance the existing evaluation system, for example, when selecting evacuation shelters.

### 4 Conclusions

In this study, we developed an ordinal logistic model to estimate the probability distribution for the sediment yield (erosion) scale by changing the pair of SRI and

LRI using the 2018 disaster data. The ROC curve and AUC in the 2014 disaster prediction evaluated the models' performances. The regression model using 60-minute and 48-hour rainfall recorded relatively higher performance than the other models. We generated 100 patterns of sediment production (debris flow mass) data in the 2014 event condition to execute a Monte Carlo debris flow simulation. As a result, the relative frequency of deposition area, which is regarded as the hazard possibility with several assumptions, was obtained. Real time execution of this method estimates the hazard probability in the practical hazard area at any time, contributing to advancing the warning and evacuation systems.

### References

1. N. Osanai, T. Shimizu, K. Kuramoto, S. Kojima, T. Noro, *Landslides*, **7(3)**, 325-338 (2010).
2. K. Nakatani, K. Amou, Y. Fujimoto, Y. Satofuka, M. Fujita, *J. Nat. Disaster Sci.*, **35**, 65-77 (2016)
3. F. Wang, Y. Wu, H. Yang, Y. Tanida, A. Kamei, *Geoenvironmental Disasters*, **2**, 17 (2015)
4. R. Hashimoto, T. Tsuchida, T. Moriwaki, S. Kano, *Soils Found.and Foundations*, **60**, 283-299, (2020)
5. S. Noguchi, K. Yamanoi, K. Kawaike, H. Nakagawa, *J. Jpn. Soc. Civ. Eng. Ser. B1*, **76(2)**, I\_877-I\_882 (2020)
6. J. Kanbara, T. Uchida, Y. Tanaka, T. Matsubara, K. Ikeda, H. Todate, *J. Jpn. Soc. Eros. Control. Eng.*, **70(1)**, 38-45, (2017)
7. K. Yamanoi, S. Oishi, K. Kawaike, H. Nakagawa, *J. Flood Risk Manag.*, **15(2)**, e12776, (2022)
8. T. Takahashi, *Debris flow Mechanics, Prediction and Countermeasures*, (Taylor & Francis, London, 2007)