Journal of Hydroscience and Hydraulic Engineering
Vol. 16, No. 2, November, 1998, 63-70

ESTIMATION OF THE OCCURRENCE CONDITION OF A DEBRIS FLOW

By

H. Takabatake, M. Hirano, T. Moriyama and K. Kawahara

Kyushu University, Fukuoka, Japan

SYNOPSIS

The critical-occurrence-condition of a debris flow due to rainfall is described in terms of concentration time of rainfall and cumulative rainfall. Since the occurrence-condition may vary stochastically with time, statistical analysis is required. As there is little information available concerning the distributions of concentration time and critical cumulative rainfall, a nonparametric method is a valuable means of investigation. In this paper, the EM (Expectation-Maximization) algorithm is used in order to obtain a nonparametric estimation of distributions of concentration time and the critical cumulative rainfall. The method is applied to the data of the Mizunashi River near the Unzen Volcano. Demonstrated here is that the distributions of the concentration time and the critical cumulative rainfall can be approximated by lognormal distribution.

INTRODUCTION

Debris flows have frequently occurred in the Mizunashi River since the eruption of the Unzen Volcano in 1990. Recently, the occurrence condition of a debris flow seems to have changed with dormancy of the volcano. The critical condition of the occurrence of a debris flow can be described in terms of concentration time of rainfall and cumulative rainfall(2). In the case of an unsteady occurrence condition of a debris flow, it is difficult to estimate this critical conditions by deterministic method. In this paper, a statistical method for estimating the critical-occurrence-condition of a debris flow is introduced and the probability of the occurrence of a debris flow is estimated. This needs the distribution functions of the critical factors, concentration time of rainfall and cumulative rainfall. But there is little information on them. In the following, a non-parametric method, like the Kaplan-Meier method, is used in order to survey the form of a distribution function which has two variables, the concentration time and the critical cumulative rainfall.

THE OCCURRENCE CONDITION OF A DEBRIS FLOW

The occurrence condition of a debris flow is described by assuming that a debris flow occurs with the collapse of a slope when the depth of subsurface flow exceeds the critical value as(2)

$$l \geq k\ T \sin\theta \ /\lambda \tag{1}$$

and

$$R_c = \frac{\lambda\ H_c}{\cos\ \theta} \leq \int_{t-T}^{t} r(u)\ du \tag{2}$$

where l is the length of slope, θ is the angle of inclination of slope, k is the hydraulic conductivity, λ is the void content, H_c is the critical depth of subsurface flow at which slope failure will begin, r is the rainfall intensity at the time t and T is the concentration time of rainfall.

From Eqs. (1) and (2), one obtains

$$r_c = \frac{H_c\ k}{l} \tan\theta \le r_a = \frac{1}{T}\int_{t-T}^{t} r(u)\ du \tag{3}$$

Concentration time of rainfall on a slope is given by

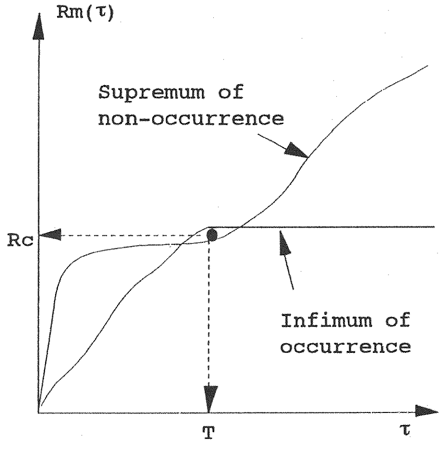
$$T = \frac{\lambda\ l}{k \sin\theta} \tag{4}$$

These equations describe the occurrence criterion of a debris flow as functions of concentration time T and critical cumulative rainfall R_c (or critical rainfall intensity r_c). It is possible to estimate these values by measuring H_c, l , k and θ in the field, but the estimated values will not be accurate enough for practical use due to the large errors in the measurements. It is therefore preferable to estimate T and R_c from rainfall and debris flow data.

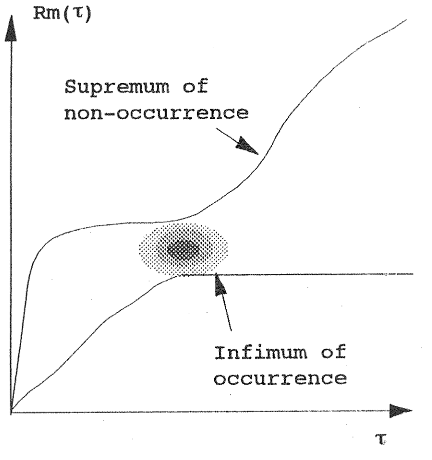
The cumulative rainfall between time $t-\tau$ and t defined below is calculated from time series of rainfall and the maximum value of $R(\tau,t)$, $R_m(\tau)$ is obtained.

$$R(\tau,\ t) = \int_{t-\tau}^{t} r(u)\ du \tag{5}$$

If there are no errors in the data and above assumption about the occurrence criterion of a debris flow is quite true, the plotted line of $R_m(\tau)$ on the coordinate of τ-$R_m(\tau)$ as shown in Fig.1 should exceed the critical point, (T, R_c), when a debris flow occurs, and should not exceed it when a debris flow does not occur. Thus the envelope of the infimum of $R_m(\tau)$ curves with occurrence of debris flows and the supremum of $R_m(\tau)$ curves without occurrence of debris flows will cross near the point (T, R_c) as shown in Fig.1(a). In fact, since the field condition varies with time, the occurrence criterion is not described as a point but should spread on the τ-$R_m(\tau)$ field as shown in Fig.1(b).



(a) (T, Rc) is constant.



(b) (T, Rc) varies with time.

Fig.1 Schematic illustration of supremun and infimum curves for occurrence of debris flows.

STATISTICAL ESTIMATION OF THE OCCURRENCE CONDITION OF A DEBRIS FLOW

Estimating the distribution of the occurrence condition by maximum likelihood method

It is assumed that T and R_c are distributed with a joint probability density function $f(T, R_c; \phi)$, where ϕ is a parameter vector of f. The procedure to estimate ϕ by using the maximum likelihood method is following.

The probability that a debris flow occurs with the i-th rainfall event, P_i is expressed by the probability that (T, R_c) is located in the area under curve of $R_{mi}(t)$ on τ-$R_m(\tau)$ plane.

$$P_i = \int_0^{\infty} \{ \int_0^{R_{mi}(T)} f(T, R_c ; \phi) \; dR_c \} \; dT \tag{6}$$

where $R_{mi}(\tau)$ is $R_m(\tau)$ in the i-th rainfall event.

Loglikelihood function $L(\phi)$ for whole rainfall events is given by

$$L(\phi) = \sum_{i=1}^{n} \{\delta_i \; P_i(\phi) + (1 - \delta_i) [1 - P_i(\phi)]\} \tag{7}$$

where n is the number of rainfall events, δ_i is 1 if a debris flow occurs and 0 if does not in the i-th rainfall event. The maximum likelihood estimate, $\hat{\phi}$ is the value of ϕ which gives the maximum value of $L(\phi)$.

Nonparametric estimation of distribution of the occurrence condition of a debris flow

Since there is little information available about the distribution of (T, R_c) or (T, r_c), the nonparametric method is efficient for estimating the distribution curves. As rainfall data are generally recorded as discrete values, τ-$R_m(\tau)$ plane shown in Fig.1 is divided into J×K discrete points, where J is the number of points along the τ axis and K represents that along the $R_m(\tau)$ axis. For nonparametric estimation, it is assumed that (T, R_c) follows the multinominal distribution that probability that (T, R_c) appears at the point of number (j,k) on discrized τ-$R_m(\tau)$ plane is π_{jk} at once rainfall event. By replacing ϕ with $\pi = \{\pi_{jk} | j=1, 2, \ldots, J; k = 1, 2, \ldots, K\}$ in Eq.(7), the probability P_i that a debris flow occurs with the i-th rainfall event is expressed as

$$P_i = \sum_{j=1}^{J} \sum_{k,\; R_c \le R_{mi}(\tau)} \pi_{jk} \tag{8}$$

The maximum likelihood estimate of π is obtained by equalizing every partial derivative function of $L(\pi)$ with π_{jk} to zero, where $L(\pi)$ is obtained by substituting Eq.(8) into $P_i(\phi)$ of Eq.(7). Since it is difficult to solve this in the explicit form, the maximum likelihood estimate of π is calculated by the iterative method using the EM algorithm(1). The procedure is as follows:

First, if (T, R_c) in every rainfall event was observed completely, log likelihood function $L_0(\pi)$ corresponding to Eq.(7) should be

$$L_0(\pi) = \sum_{j=1}^{J-1}\sum_{k=1}^{K-1} C_{jk} \ln \pi_{jk} + \sum_{k=1}^{K-1} C_{Jk} \ln \pi_{Jk} + \sum_{j=1}^{J-1} C_{jK} \ln \pi_{jK} + C_{JK} \ln \left\{1 - \sum_{j=1}^{J-1}\sum_{k=1}^{K-1} \pi_{jk} - \sum_{j=1}^{J-1} \pi_{jK} - \sum_{k=1}^{K-1} \pi_{Jk}\right\} \tag{9}$$

where C_{jk} is the number of times that (T, R_c) at the discrete point of number (j,k) was observed through whole rainfall events. In Eq.(9), terms with no effect on π_{jk} are omitted, and the following relationship about parameters of multinominal distribution π_{jk} is used.

$$\sum_{j=1}^{J}\sum_{k=1}^{K} \pi_{jk} = 1 \tag{10}$$

In actual observation, data about maximum cumulative rainfall $R_m(\tau)$ and records of the occurrence/non-occurrence of a debris flow are obtained. We define $R_m(\tau)$ and the occurrences/non-occurrences of debris flows through whole events, as data $M = \{(\tau,\ R_{mi}(\tau)),\ \delta_i;\ i = 1,\ 2,\ \ldots,\ n\}$.

Using data M and approximation of π, $\pi^{(p)}$, in expectation step of the EM algorithm, conditional expectation $Q(\pi,\ \pi^{(p)})$ of $L_0(\pi)$ is expressed as

$$\begin{aligned} Q(\pi,\ \pi^{(p)}) &= E\left[L_0(\pi) \mid M,\ \pi^{(p)}\right] \\ &= \sum_{j=1}^{J-1}\sum_{k=1}^{K-1} E\left[C_{jk} \mid M,\ \pi^{(p)}\right] \ln \pi_{jk} + \sum_{k=1}^{K-1} E\left[C_{Jk} \mid M,\ \pi^{(p)}\right] \ln \pi_{Jk} \\ &\quad + \sum_{j=1}^{J-1} E\left[C_{jK} \mid M,\ \pi^{(p)}\right] \ln \pi_{jK} \\ &\quad + E\left[C_{JK} \mid M,\ \pi^{(p)}\right] \ln \left\{1 - \sum_{j=1}^{J-1}\sum_{k=1}^{K-1} \pi_{jk} - \sum_{j=1}^{K-1} \pi_{jK} - \sum_{k=1}^{K-1} \pi_{Jk}\right\} \end{aligned} \tag{11}$$

and $E\left[C_{jk} \mid M,\ \pi^{(p)}\right]$ is expressed by using δ_i in Eq.(7) and P_i in Eq.(8).

$$E\left[C_{jk} \mid M,\ \pi^{(p)}\right] = \sum_{i=1}^{n} \frac{g_i(j,k)\ \pi_{jk}^{(p)}}{\delta_i\ P_i + (1 - \delta_i)(1 - P_i)} \tag{12}$$

where $g_i(j,k)$ is

$$g_i(j,k) = \begin{cases} 1 & R_c \le R_{mi}(\tau),\ \delta_i = 1 \ \text{ or } \ R_c > R_{mi}(\tau),\ \delta_i = 0 \\ 0 & R_c \le R_{mi}(\tau),\ \delta_i = 0 \ \text{ or } \ R_c > R_{mi}(\tau),\ \delta_i = 1 \end{cases} \tag{13}$$

The maximization step on the EM algorithm is to obtain π which gives maximum value of Q. This is obtained through equalizing partial derivative functions of Q by π_{jk} to zero.

$$\pi_{jk} = \frac{E\left[C_{jk} \mid M,\ \pi^{(p)}\right]}{n} \tag{14}$$

Due to the characteristics of the EM algorithm(1), π_{jk} is adjusted from the previously given $\pi_{jk}^{(p)}$ to a value closer to the maximum likelihood estimate of π_{jk}, $\hat{\pi}_{jk}$. Repeating the substitution of π_{jk} obtained from Eq.(14) into $\pi_{jk}^{(p)}$ in Eq.(12), π_{jk} converges into $\hat{\pi}_{jk}$. At the first step on the iterative calculation, the values of $\pi_{jk}^{(p)}$ on the points (j,k) where the values of $R_m(\tau)$ do not appear in the rainfall data $\{(\tau,\ R_{mi}(\tau));\ i = 1,\ 2,\ \ldots,n\}$ are given zero and those where the values of $R_m(\tau)$ do appear are given same value satisfying Eq.(10).

THE RESULT OF THE ANALYSIS

Nonparametric joint distribution functions of concentration time T and critical cumulative rainfall R_c as well as T and critical rainfall intensity r_c are estimated using data at the Mizunashi River near the Unzen Volcano. Rainfall data for ten minutes collected at the Unzen Meteorological Observatory from May '91 to September '96 are used. The occurrences of debris flows were recognized using seismographic data for the same period. Because the occurrence condition of a debris flow along the Mizunashi River is considered to have changed significantly since July '93(3), distribution functions before and after July '93 are determined separately.

The results are shown in Figs.2,3 and 4 and Table 1. The figures illustrate the estimated cumulative probabilities of T, R_c and r_c. Straight lines in these figures indicate lognormal distribution with same averages and standard deviations as those of estimated nonparametric joint distribution functions shown in Table 1.

Table 1 Statistics calculated from nonparametric estimated distribution functions.

Variable		Average	Standard deviation	Correlation Coefficient	
				T	ln T
T	before Jul.'93	49.2 min.	36.5 min.		
	after Jul.'93	43.8 min.	35.4 min.		
R_c	before Jul.'93	9.4 mm	5.5 mm	0.33	
	after Jul.'93	26.9 mm	33.8 mm	0.66	
r_c	before Jul.'93	3.2 mm/10min.	2.0 mm/10min.	-0.67	
	after Jul.'93	6.3 mm/10min.	3.6 mm/10min.	-0.03	
ln T	before Jul.'93	3.52 (= ln 33.9)	0.93 (= ln 2.54)		
	after Jul.'93	3.41 (= ln 30.2)	0.89 (= ln 2.44)		
ln R_c	before Jul.'93	2.02 (= ln 7.53)	0.72 (= ln 2.06)		0.50
	after Jul.'93	2.73 (= ln 15.4)	0.94 (= ln 2.55)		0.70
ln r_c	before Jul.'93	0.90 (= ln 2.46)	0.75 (= ln 2.11)		-0.62
	after Jul.'93	1.64 (= ln 5.15)	0.70 (= ln 2.01)		-0.31

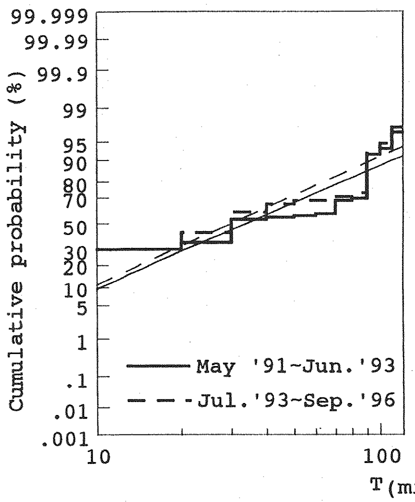


Fig.2 Estimated distribution functions of T (Straight lines are lognormal distribution.)

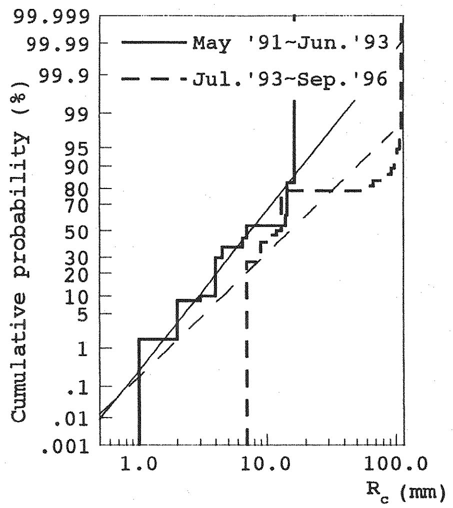


Fig.3 Estimated distribution functions of R_c (Straight lines are lognomal diatribution.)

Correlation coefficients between T and R_c, and between T and r_c, are also shown in Table 1.

Figure 2 shows the estimated marginal distribution functions of T as cumulative probabilities, $\Pi_{j\cdot} = \sum_{m=1}^{j} \pi_{m\cdot}$, where $\pi_{j\cdot} = \sum_{k=1}^{K} \pi_{jk}$. Comparing the stepped curves of the nonparametric distribution functions with the straight lines of lognormal distribution, they appear to be approximated by lognormal distribution. No significant change between the distributions of T before and after July '93 is seen in Fig.2 and Table 1.

The estimated marginal distribution functions of R_c are shown in Fig.3 as cumulative probabilities, $\Pi_{\cdot k} = \sum_{m=1}^{k} \pi_{\cdot m}$, where $\pi_{\cdot k} = \sum_{j=1}^{J} \pi_{jk}$. This figure and Table 1 show that the distribution of R_c have changed considerably since July '93. The estimated distribution function of R_c seems to be approximated by lognormal distribution before July '93, but it can not be approximated after July '93.

Figure 4 shows that the distributions of r_c, both before and after July '93, can also be approximated by lognormal distribution.

THE PROBABILITY DISTRIBUTION OF THE OCCURRENCE CONDITION OF A DEBRIS FLOW

The above mentioned analysis, by using the EM algorithm, illustrates that marginal distributions of T, R_c and r_c can be approximated by lognormal distribution. This is consistent with the fact that T, R_c and r_c are defined by the products of many factors, as seen in Eqs.(2),(3) and (4), leads to the conclusion that their distributions are close to lognormal distributions due to the central limit theorem. If the joint distribution of X1 = ln T and X2 = ln R_c is expressed by bivariate normal distribution, joint distribution of (T, R_c) is given as following function of Eq.(15).

If the parameters in this function are given by a certain method, the probability that a debris flow occurs can be calculated by the integration of this function.

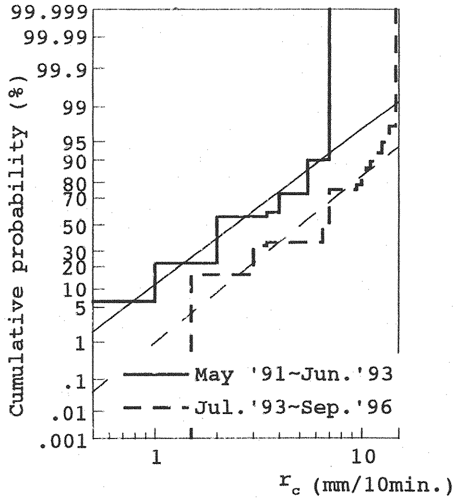


Fig.4 Estimated distribution function of r_c
(Straight lines are lognormal distribution)

$$f(T, R_c) = \frac{1}{2\pi \sigma_{X1} \sigma_{X2} \sqrt{1-\rho^2}\, T R_c} \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(\ln T - \mu_{X1})^2}{\sigma_{X1}^2} - \frac{2\rho(\ln T - \mu_{X1})(\ln R_c - \mu_{X2})}{\sigma_{X1}\sigma_{X2}} + \frac{(\ln R_c - \mu_{X2})^2}{\sigma_{X2}^2}\right]\right\} \tag{15}$$

CONCLUSION

The nonparametric distribution functions of concentration time T and critical cumulative rainfall R_c, which the occurrence condition of a debris flow follows, have been estimated using the EM algorithm. The estimated values of T and R_c are distributed lognormally. At the Mizunashi River, a significant change in the occurrence condition between before and after July '93 is confirmed. The probability that a debris flow occurs can be calculated from rainfall data using the estimated distribution function.

Detailed analysis of the joint distribution of concentration time and critical cumulative rainfall, as well as the introduction of a parametric method, is all required in order to describe the probability that a debris flow occurs.

REFERENCES

1. DEMPSTER, A. P., LAIRD, N. M. and RUBIN, D. B. : Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society B(39), pp.1-38, 1977.
2. HIRANO, M., HIKIDA, M. and MORIYAMA, T. : Occurrence Criterion and Discharge Prediction of Debris Flow in the Active Volcanic Area, Proceedings of the Japanese Conference on Hydraulics, JSCE, Vol.30, pp.181-186, 1986(in Japanese).
3. HIRANO, M., HASHIMOTO, H. and KAWAHARA, K. : Characteristics of Debris Flow at the Mizunashi River in Mt. Unzen, Annual Journal of Hydraulic Engineering, JSCE, Vol.40, pp.1027-1032, 1996(in Japanese).

APPENDIX - NOTATION

The following symbols are used in this paper:

C_{jk}	=	the number of times that (T, R_c) of discrete point number (j,k) on τ-$R_m(\tau)$ plane is observed through whole rainfall events;
$f(T, R_c; \phi)$	=	joint probability density function of T and R_c;
$g_i(j,k)$	=	defined by Eq.(13);
H_c	=	critical depth of subsurface flow at which slope failure will begin;
j	=	discrete point number along the τ axis;
J	=	the number of discrete points along the τ axis;
k	=	hydraulic conductivity (used at p.1 and 2);
k	=	discrete point number along the $R_m(\tau)$ axis (used from p.3 to p.5);
K	=	the number of discrete points along the $R_m(\tau)$ axis;
l	=	length of slope;
$L(\phi)$	=	log likelihood function for whole rainfall events;
$L(\pi)$	=	log likelihood function given the data M;
$L_0(\pi)$	=	log likelihood function if (T, R_c) in every rainfall event is observed completely;
m	=	discrete point number along the τ axis or the $R_m(\tau)$ axis;
M	=	data of $R_m(\tau)$ and the occurrences/non-occurrences of debris flows through whole events, $\{(\tau, R_{mi}(\tau)), \delta_i; i = 1, 2, \ldots, n\}$;
n	=	the number of rainfall events;
P_i	=	probability that a debris flow occurs with the i-th rainfall event;
$Q(\pi, \pi^{(P)})$	=	conditional expectation of $L_0(\pi)$ given data M;
r	=	rainfall intensity;
r_a	=	average rainfall intensity;
r_c	=	critical rainfall intensity;
$R(\tau,t)$	=	cumulative rainfall between time $t-\tau$ and t;
R_c	=	critical cumulative rainfall;
$R_m(\tau)$	=	maximum value of $R(\tau,t)$;
$R_{mi}(\tau)$	=	$R_m(\tau)$ in the i-th rainfall event;
t	=	time;
T	=	concentration time of rainfall;
$X1$	=	$\ln T$;
$X2$	=	$\ln R_c$;
δ_i	=	1 if a debris flow occurs and 0 if does not in the i-th rainfall event;
θ	=	angle of inclination of slope;
λ	=	void content;

μ_{X1}	= avarage of X1;
μ_{X2}	= average of X2;
π	= parameter vector of multinominal distribution $\{\pi_{jk} \mid j=1, 2, \ldots, J;\ k = 1, 2, \ldots, K\}$;
π_{jk}	= probability that (T, R_c) appears at the point of number (j,k) on discrized τ-$R_m(\tau)$ plane at once rainfall event;
$\hat{\pi}_{jk}$	= maximum likelihood estimate of π_{jk};
$\pi_{j\cdot}$	= marginal nonparametric probability distribution of T at point of number j;
$\pi_{\cdot k}$	= marginal nonparametric probability distribution of R_c at point of number k;
$\pi^{(p)}$	= approximation of π on the iteration of the EM algorithm;
$\Pi_{j\cdot}$	= marginal nonparametric distribution of T as cumulative probabilities;
$\Pi_{\cdot k}$	= marginal nonparametric distribution of R_c as cumulative probabilities;
ρ	= correlation coefficient between X1 and X2;
σ_{X1}	= standard deviation of X1;
σ_{X2}	= standard deviation of X2;
τ	= cumulative time;
ϕ	= parameter vector of $f(T, R_c; \phi)$; and
$\hat{\phi}$	= maximum likelihood estimate of ϕ.

(Received December 2, 1997; revised August 17, 1998)