Contents lists available at DJM
DESIMAL: JURNAL MATEMATIKA
p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm
http://ejournal.radenintan.ac.id/index.php/desimal/index

Estimating flood hazard rate in parepare using likelihood approach single decrement method

Ahmad Fajri S*, Nurul Fuady Adhalia H, Putri Ayu Maharani, Syahrul Ramadhan Tahir

Institut Teknologi Bacharuddin Jusuf Habibie, Indonesia

ARTICLE INFO

Article History
Received : 12-10-2023
Revised : 30-10-2023
Accepted : 10-11-2023
Published : 30-12-2023

Keywords:
Hazard Rate; Flood; Single Decrement; Likelihood; Waiting Time.

*Correspondence: E-mail: ahmadfajri@ith.ac.id
Doi: 10.24042/djm.v6i3.19145

ABSTRACT

Flooding is an example of a random stochastic process. An important parameter to determine the chance of a flood occuring is the hazard rate. Therefore, a hazard rate estimation model is needed. One of the methods used to estimate the hazard rate at point t_0 was the single decrement method with a likelihood approach that required exit time information, namely the time when a flood occurs and the assumed distribution of waiting times for the next flood to occur. The distribution of waiting times was assumed to be linear and exponential. Hazard rate estimation used flood data that occurred in Parepare. The hazard rate estimator obtained using these two waiting time assumptions was transformed into a parametric model. The parametric model used was a regression model with linear, quadratic, and cubic assumptions. Based on the research results, the best parametric model was a quadratic regression model for the assumed exponential distribution of waiting times based on R Square, Mean Square Error, and real regression tests. The estimated hazard rate value obtained can be applied to estimate the probability of a flood event occurring in the interval $(0, t_0]$. The selected parametric model is expected to be able to estimate the hazard rate value accurately.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

A stochastic process is a process that fits to model phenomena that contain elements of uncertainty. An example of the use of stochastic processes in phenomena that occur is to explain and predict future events through a series of calculations and analyses carried out. For example, to predict natural phenomena such as natural disasters (floods, earthquakes, landslides, and so on).

Flooding is an event that occurs due to the accumulation of water that falls and cannot be accommodated by the ground (Yohana, Griandini, & Muzambeq, 2017). Flooding is one example of a random stochastic process. A stochastic model that can explain unpredictable events in space and time is called a point process, where the time a flood occurs can be viewed as a point in a specific area (Sunusi, Jaya, Islamiyati, & Raupong, 2013). In the last two years, there have been the worst floods ever in Parepare. According to the Regional Disaster Management Agency, the victims affected by the flood were 1,354 heads of families and 5,292 people, of whom 2 died. Other losses included damaged houses, schools, mosques, vehicles, and other parts.

One important parameter to determine the chance of a flood occurring is the hazard rate. Hazard rate estimation used flood data that occurred in Parepare. If the hazard rate is known, then the joint density distribution for the realization of flooding in the interval $(0, T)$ can be known so that the maximum likelihood estimator can be known. Therefore, obtaining a suitable parametric model is essential for estimating the hazard rate.

Research on hazard rate estimation includes estimating the hazard rate using a parametric approach, namely through a point process likelihood equation called the Hazard Rate Likelihood Point Process (HRLPT), where this process is limited to estimating the hazard rate for the observation interval (Ogata, 1999; Vere-Jones, 1995).

Furthermore, a method for analytically estimating the hazard rate of the temporal point process was developed for some cases of waiting time and predicting the hazard rate in the following period (Sunusi, 2010). Another method for estimating the hazard rate for flooding is the Hazard Rate Single Decrement method (Darwis, Sunusi, Gunawan, Mangku, & Wahyuningsih, 2009). This method comes from the estimation method in actuarial studies used in mortality tables. In actuarial science, the single-decrement approach deals with the study in which death is the only random event to which sample members are subject. The approach is used to construct a mortality table to estimate the premium and predict the reserve. If both death and withdrawal are random events, the environment is called a double decrement. Adaptation to flood prediction: the single decrement approach deals with the study in which flood occurrence time is the only random event.

METHOD

The data used in this research is data on the number of flood events in Parepare in 2017–2023 from the Parepare City Regional Disaster Management Agency. The software used to analyze the data is IBM SPSS Statistics Version 29.

Survival analysis is a procedure for analyzing data using the time until an event occurs (Kleinbaum & Klein, 2007). There are two main functions in the Survival Analysis, namely the hazard function and the survival function (Collet, 2003).

The hazard rate at point t_0 is symbolized by μ_{t_0}, as is usually used in the actuarial field. Let $X(t_0) = T - t_0$ denote the waiting time until the next flood occurs if it is known that t_0 is the first time a flood appears and T is the time the next flood appears. Let also μ and S denote the hazard rate and survival function. Hazard rate μ_{t_0} can be expressed as

$$\mu_{t_0} = \lim_{\Delta_{t_0 \to 0}} \frac{P\left(t_0 \le T \le t_0 + \Delta_{t_0} \middle| T > t_0\right)}{\Delta_{t_0}}$$
$$= -\frac{d}{dt_0} \ln S(t_0) = \frac{-\frac{d}{dt_0} S(t_0)}{S(t_0)}.$$

The probability that there will be no flooding until $t_0 + \Delta t_0$ if it is known that no flooding will occur until t_0 is

$$\int_{t_0}^{t_0+\Delta t_0} d\ \ln S(t_0) = -\int_{t_0}^{t_0+\Delta t_0} \mu_{t_0} dt_0$$
$${}_{\Delta t_0}p_{t_0} = e^{-\int_0^{\Delta t_0} \mu(t_0+s)\, ds}.$$

Let $t_0 = 0$, immediately after the flood occurs

$${}_{\Delta t_0}p_{t_0} = S(\Delta t_0) = P(T > \Delta t_0)$$
$$= e^{-\int_0^{\Delta t_0} \mu(s)\, ds}$$

that is the survival function (Le, 1997).

The distribution of the recurrent time T and the waiting time until the next flood occurs $X(t_0)$ are each expressed as follows (Bowers, Gerber, Hickman, Jones, & Nesbitt, 1997).

$$T \sim {}_{\Delta t_0}p_{t_0}\mu_{t_0}$$
$$X(t_0) \sim {}_{\Delta t_0}p_{t_0}\mu_{t_0+\Delta t_0}.$$

The expression ${}_{\Delta t_0}p_{t_0}\mu_{t_0+\Delta t_0}$ states the probability that a flood will occur between t_0 and $t_0 + \Delta t_0$ if it is known that no flood has occurred until t_0 with

$$\int_0^\infty {}_{\Delta t_0}p_{t_0}\mu_{t_0+\Delta t_0}\ dt = 1$$

and

$$\frac{d}{dt}\ {}_{\Delta t_0}p_{t_0} = -\ {}_{\Delta t_0}p_{t_0}\mu_{t_0+\Delta t_0}.$$

Estimating the hazard rate using the single-decrement method uses a maximum likelihood estimation. The maximum likelihood estimation is often used because the procedure is clear and appropriate for determining the parameters of a distribution (Krishnamoorthy, 2015). Let d_{t_0} denotes the number of floods that have occurred in the interval $(t_0, t_0 + 1]$ and $n_{t_0} - d_{t_0}$ denotes the flood that occurred after $t_0 + 1$. Likelihood L for the i-th flood in the interval $(t_i, t_i + 1]$ if it is known that no flood has occurred until t_i is given by

$$\begin{aligned} L_i &= f\big(t_0(i)|T > t_0(i)\big) \\ &= \frac{f\big(t_0(i)\big)}{S(t_0)} \\ &= \frac{S(t_0(i))\mu(t_0(i))}{S(t_0)} \end{aligned}$$

namely the i-th contribution to L.

If $y_i = t_0(i) + t$ is the time when the i-th flood occurs in the interval $(t_0, t_0 + 1]$ with $0 < y_i \le 1$, then

$$L_i = \frac{S(t_0 + y_i)\mu(t_0 + y_i)}{S(t_0)} = {}_{y_i}p_{t_0}\mu_{t_0+y_i}.$$

The contribution of the number of floods d_{t_0} to L is $\prod_{i=1}^{d_{t_0}} {}_{y_i}p_{t_0}\mu_{t_0+y_i}$. The contribution of $n_{t_0} - d_{t_0}$ is $(p_{t_0})^{n_{t_0}-d_{t_0}}$ where n_{t_0} denotes the number of floods that occurred at or after t_0 so that the total likelihood L is

$$L = (p_{t_0})^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{y_i}p_{t_0}\mu_{t_0+y_i}. \quad (1)$$

Equation (1) requires an assumption of the waiting time for a flood to occur. Those assumptions are linear and exponential distributions, denoted in the form q_{t_0}.

First, for linear distribution, it is obtained

$$\mu_{t_0+y} = \frac{q_{t_0}}{1 - {}_y q_{t_0}} \quad (2)$$

so that the total likelihood L becomes

$$\begin{aligned} L &= \big(p_{t_0}\big)^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{y_i}p_{t_0}\mu_{t_0+y_i} \\ &= (1 - q_{t_0})^{n_{t_0}-d_{t_0}} {q_{t_0}}^{d_{t_0}}. \end{aligned}$$

The natural logarithm of L is

$$\ln L = (n_{t_0} - d_{t_0}) \ln\big(1 - q_{t_0}\big) + d_{t_0} \ln q_{t_0}.$$

By using the necessary conditions for the optimality of the first order derivative,

$$\frac{\partial}{\partial q}(\ln L) = 0$$
$$\frac{d_{t_0}}{q_{t_0}} - \frac{(n_{t_0} - d_{t_0})}{1 - q_{t_0}} = 0$$

is obtained. Therefore,

$$\hat{q}_{t_0} = \frac{d_{t_0}}{n_{t_0}}. \quad (3)$$

By using Equations (2) and (3), the obtained estimated hazard rate value is

$$\hat{\mu}_{t_0} = \frac{\hat{q}_{t_0}}{1 - \hat{q}_{t_0}}. \quad (4)$$

Second, for exponential distribution, it is obtained

$$\mu_{t_0+y} = -\ln(p_{t_0}) \text{ and } {}_y p_{t_0} = (p_{t_0})^y$$

so that the total likelihood L becomes

$$\begin{aligned} L &= (p_{t_0})^{n_{t_0}-d_{t_0}} \prod_{i=1}^{d_{t_0}} {}_{y_i}p_{t_0}\mu_{t_0+y_i} \\ &= \mu^{d_{t_0}} \exp{(-\mu)} \left[\big(n_{t_0} - d_{t_0}\big) + \sum_{i=0}^{d_{t_0}} y_i\right]. \end{aligned}$$

The natural logarithm of L is

$$\ln L = d_{t_0} \ln \mu - \mu \left[\big(n_{t_0} - d_{t_0}\big) + \sum_{i=0}^{d_{t_0}} y_i\right].$$

By using the necessary conditions for the optimality of the first-order derivative, it is obtained that

$$\frac{\partial}{\partial \mu}(\ln L) = 0$$
$$\frac{d_{t_0}}{\mu} - \left[\big(n_{t_0} - d_{t_0}\big) + \sum_{i=0}^{d_{t_0}} y_i\right] = 0$$

so that the estimated hazard rate value is

$$\hat{\mu}_{t_0} = \frac{d_{t_0}}{\left[(n_{t_0} - d_{t_0}) + \sum_{i=0}^{d_{t_0}} y_i\right]}. \quad (5)$$

Since q corresponds one-to-one with μ, the estimator of $\hat{q}_{t_0}$ is

$$\hat{q}_{t_0} = 1 - \hat{p}_{t_0} = \big(1 - e^{-\hat{\mu}_{t_0}}\big). \quad (6)$$

Equation (5) is more informative than the linear solution in Equation (4). Equation (5) has information on the time of occurrence of the event, the number of events d_{t_0} in the waiting time interval $(t_0, t_0 + 1]$, and the number of events $(n_{t_0} - d_{t_0})$ after $(t_0, t_0 + 1]$. Equation (5) meets the empirical hazard rate requirements based on historical data. After the hazard rate value for each point is obtained based on the data, the hazard rate equation is estimated using a parametric model, namely the regression model with linear, quadratic, and cubic assumptions. Figure 1 shows the research flow that has been carried out.

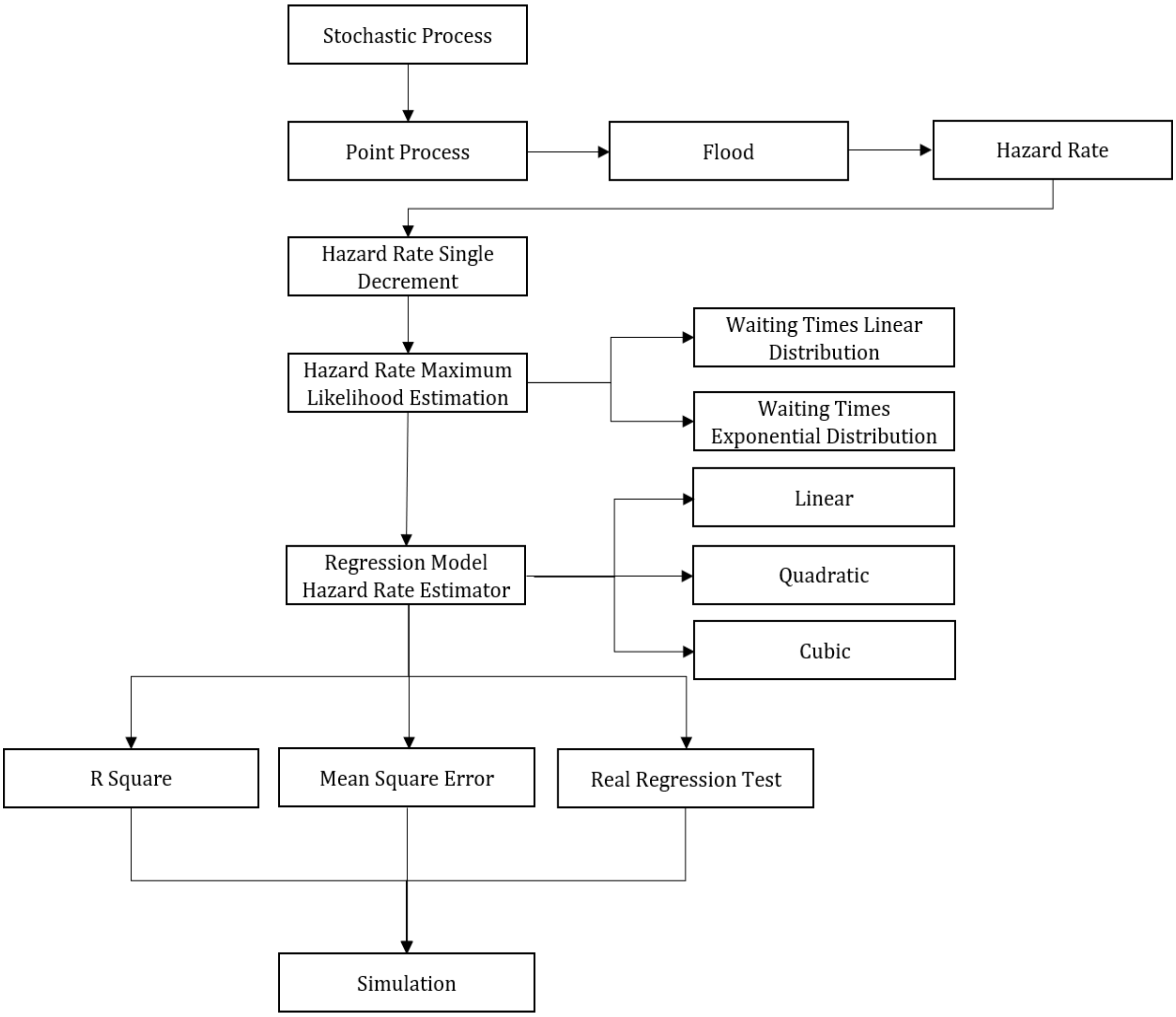

Figure 1. Research Flowchart

RESULTS AND DISCUSSION

Flood data that occurred in Parepare in 2017–2023 is shown in Table 1.

Table 1. Floods occurred in Parepare in 2017–2023

No.	Date	Location
1.	16/12/17	Soreang, Wattang Soreang
2.	16/12/17	Soreang, Bukit Harapan
3.	17/12/17	Soreang, Wattang Soreang
4.	17/12/17	Soreang, Bukit Harapan
5.	18/12/17	Ujung, Lapadde
6.	05/11/18	Bacukiki Barat, Cappa Galung
7.	05/11/18	Bacukiki Barat, Cappa Galung
8.	18/12/20	Bacukiki, Watang Bacukiki
9.	18/12/20	Bacukiki, Lemoe
10.	12/09/21	Ujung, Labukkang
11.	12/09/21	Ujung, Mallusetasi
12.	12/09/21	Soreang, Ujung baru
13.	06/03/22	Bacukiki, Watang Bacukiki
14.	18/11/22	Bacuikiki, Galung Maloang
15.	18/11/22	Ujung, Lapadde
16.	18/11/22	Ujung, Lapadde
17.	18/11/22	Ujung, Lapadde
18.	18/11/22	Bacukiki, Lompoe
19.	18/11/22	Ujung, Lapadde
20.	18/11/22	Ujung, Lapadde
21.	18/11/22	Ujung, Lapadde
22.	18/11/22	Bacukiki Barat, Bumi Harapan
23.	18/11/22	Bacukiki, Watang Bacukiki
24.	18/11/22	Bacukiki, Watang Bacukiki
25.	05/01/23	Bacukiki Barat, Lumpue
26.	01/02/23	Bacukiki, Galung Maloang
27.	01/02/23	Bacukiki, Lemoe
28.	01/02/23	Bacukiki, Lompoe
29.	01/02/23	Bacukiki, Watang Bacukiki
30.	01/02/23	Bacukiki Barat, Bumi Harapan
31.	01/02/23	Bacukiki Barat, Lumpue
32.	01/02/23	Ujung, Lapadde
33.	01/02/23	Soreang, Bukit Harapan
34.	01/02/23	Soreang, Ujung Baru

By using Equations (3), (4), (5), and (6), which have been formulated previously, the hazard rate values for the assumed linear and exponential distribution of waiting times are obtained in Table 2 and Table 3.

Table 2. Estimating the Hazard Rate Single-Decrement Likelihood Approach for Assumed Linear Distribution of Waiting Times

No.	Interval	Year	d_{t_0}	n_{t_0}	q_{t_0}	μ_{t_0}
1.	(0,1]	2017	5	34	0.147	0.172
2.	(1,2]	2018	2	29	0.068	0.074
3.	(2,3]	2019	0	27	0	0
4.	(3,4]	2020	2	27	0.074	0.080
5.	(4,5]	2021	3	25	0.120	0.136
6.	(5,6]	2022	12	22	0.545	1.200

Information:

d_{t_0}: The number of floods that occur in the interval $(t_0, t_0 + 1]$.

n_{t_0}: The number of floods that occurred at or after t_0.

q_{t_0}: The probability of a flood occurring in the interval $(t_0, t_0 + 1]$ if it is known that no flood has occurred until t_0.

μ_{t_0}: Flood hazard rate immediately after t_0.

Table 3. Estimating the Hazard Rate Single-Decrement Likelihood Approach for Assumed Exponential Distribution of Waiting Times

No.	Interval	Year	d_{t_0}	$n_{t_0} - d_{t_0}$	$\sum_{i=1}^{d_{t_0}} y_i$	q_{t_0}	μ_{t_0}
1.	(0,1]	2017	5	29	4.805	0.137	0.147
2.	(1,2]	2018	2	27	1.693	0.067	0.069
3.	(2,3]	2019	0	27	0	0	0
4.	(3,4]	2020	2	25	1.928	0.071	0.074
5.	(4,5]	2021	3	22	2.095	0.117	0.124
6.	(5,6]	2022	12	10	8.821	0.471	0.673

Information:

y_i : Time of appearance of the i-th flood in the interval $(t_0, t_0 + 1]$.

d_{t_0} : The number of floods that occur in the interval $(t_0, t_0 + 1]$.

$n_{t_0} - d_{t_0}$: The number of floods that occur after $t_0 + 1$.
q_{t_0} : The probability of a flood occurring in the interval $(t_0, t_0 + 1]$ if it is known that no flood has occurred until t_0.
μ_{t_0} : Flood hazard rate immediately after t_0.

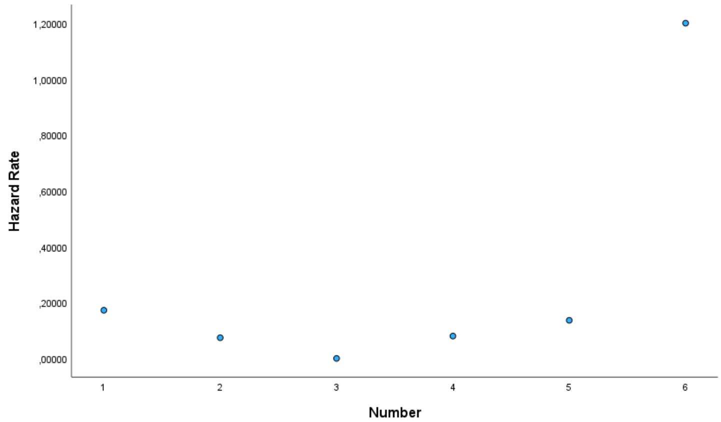

Figure 2. Plot Hazard Rate for Assumed Linear Distribution of Waiting Times

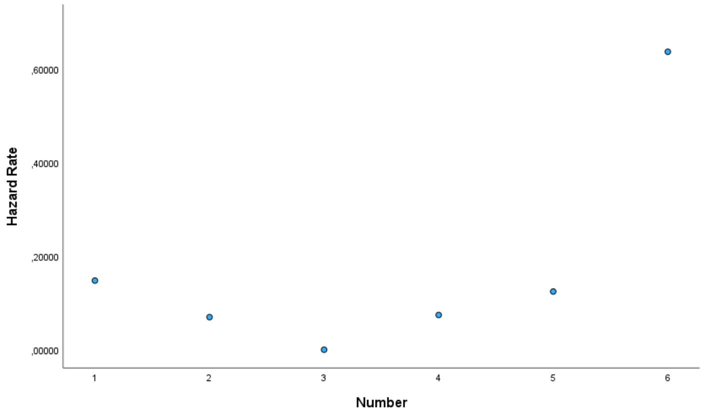

Figure 3. Plot Hazard Rate for Assumed Exponential Distribution of Waiting Times

The results in Table 2 and Table 3 state the hazard rate value for the assumed linear and exponential distribution of waiting times at t_0 with the notation μ_{t_0}. Estimating the regression equation from the hazard rate requires the condition that the hazard rate must be normally distributed. One method to use in order to check the normality of the data is the P-P plot. Therefore, the P-P plot of the hazard rate from Tables 2 and 3 is presented in Figures 4 and 5.

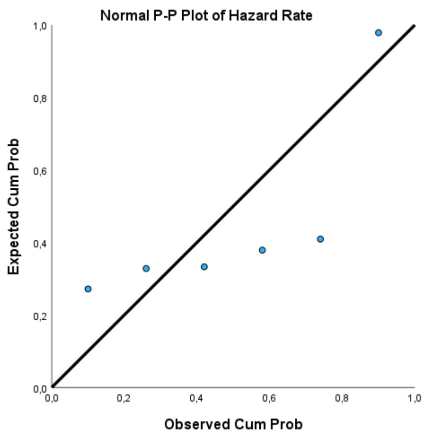

Figure 4. P-P Plot Hazard Rate for Assumed Linear Distribution of Waiting Times

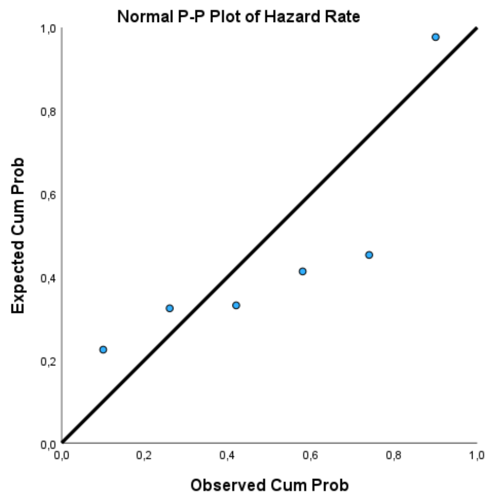

Figure 5. P-P Plot Hazard Rate for Assumed Exponential Distribution of Waiting Times

The hazard rate in Table 2 and Table 3 is not distributed normally based on Figure 4 and Figure 5 because the P-P plot hazard rate does not follow a diagonal line. Therefore, it is necessary to normalize the hazard rate data in Table 2 and Table 3. One way to normalize the data is to use the Box-Cox transformation. The Box-Cox transformation of hazard rate from Table 1 and Table 2 used value $\lambda = -0.5$ which means that the transformation carried out was $\mu_{t_0}^* = \mu_{t_0}{}^{-0.5}$, to obtain the expected value of the desired hazard rate. Table 4 and Table 5 below are the results of the Box-Cox transformation of the hazard rate from Table 2 and Table 3.

Table 4. Box-Cox Transformation Results Hazard Rate Values for the Assumed Linear Distribution of Waiting Times

No.	Interval	Year	μ_{t_0}	$\mu_{t_0}^*$
1.	(0,1]	2017	0.172	2.408
2.	(1,2]	2018	0.074	3.674
3.	(2,3]	2019	0	-
4.	(3,4]	2020	0.080	3.535
5.	(4,5]	2021	0.136	2.708
6.	(5,6]	2022	1.200	0.912

Information:
μ_{t_0}: Flood hazard rate immediately after t_0.
$\mu_{t_0}^*$: Flood hazard rate immediately after t_0 results of the Box-Cox transformation

Table 5. Box-Cox Transformation Results Hazard Rate Values for Assumed Exponential Distribution of Waiting Times

No.	Interval	Year	μ_{t_0}	$\mu_{t_0}^*$
1.	(0,1]	2017	0.147	2.600
2.	(1,2]	2018	0.069	3.787
3.	(2,3]	2019	0	-
4.	(3,4]	2020	0.075	3.669
5.	(4,5]	2021	0.124	2.834
6.	(5,6]	2022	0.637	1.252

Information:
μ_{t_0}: Flood hazard rate immediately after t_0.
$\mu_{t_0}^*$: Flood hazard rate immediately after t_0 results of the Box-Cox transformation

P-P Hazard rate plots from Table 4 and Table 5 are presented in Figure 6 and Figure 7.

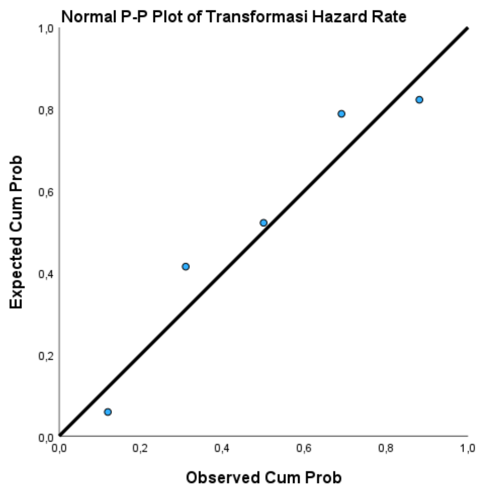

Figure 6. P-P Plot Box-Cox Transformation Results Hazard Rate Values for the Assumed Linear Distribution of Waiting Times

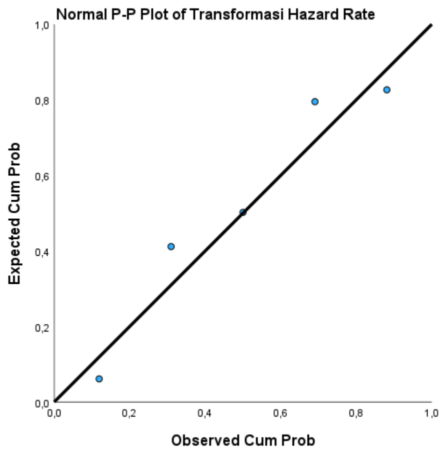

Figure 7. P-P Plot Box-Cox Transformation Results Hazard Rate

Values for the Assumed Exponential Distribution of Waiting Times

Hazard rates in Table 4 and Table 5 are distributed normally based on Figure 6 and Figure 7 because the P-P Plot hazard rate follows a diagonal line. The hazard rates in Table 4 and Table 5 are transformed into a parametric model. The parametric model used is a regression model with linear assumptions $\hat{\mu}^*_{t_0} = \alpha_0 + \alpha_1 t_0 + \varepsilon$, quadratic assumptions $\hat{\mu}^*_{t_0} = \alpha_0 + \alpha_1 t_0 + \alpha_2 {t_0}^2 + \varepsilon$ and cubic assumptions $\hat{\mu}^*_{t_0} = \alpha_0 + \alpha_1 t_0 + \alpha_2 {t_0}^2 + \alpha_2 {t_0}^2 + \varepsilon$. The resulting model for $\hat{\mu}^*_{t_0}$ is shown in Table 6 and Table 7 below.

Table 6. Estimated Hazard Rate Equation for the Assumed Linear Distribution of Waiting Times

Regression Model	Regression Equations	R Square	Mean Square Error (MSE)	Real Regression Test $(\alpha = 0,05)$
Linear	$\hat{\mu}^*_{t_0} = 3.641 - 0.276t_0$	0.266	1.201	Not Real
Quadratic	$\hat{\mu}^*_{t_0} = 0.652 + 2.162t_0 - 0.353{t_0}^2$	0.993	0.017	Real
Cubic	$\hat{\mu}^*_{t_0} = 0.366 + 2.522t_0 - 0.469{t_0}^2 + 0.011{t_0}^3$	0.994	0.027	Not Real

Table 7. Estimated Hazard Rate Equation for the Assumed Exponential Distribution of Waiting Times

Regression Model	Regression Equations	R Square	Mean Square Error (MSE)	Real Regression Test $(\alpha = 0,05)$
Linear	$\hat{\mu}^*_{t_0} = 3.745 - 0.255t_0$	0.268	1.101	Not Real
Quadratic	$\hat{\mu}^*_{t_0} = 0.995 + 1.989t_0 - 0.325{t_0}^2$	0.994	0.012	Real
Cubic	$\hat{\mu}^*_{t_0} = 0.587 + 2.501t_0 - 0.489{t_0}^2 + 0.015{t_0}^3$	0.997	0.011	Not Real

The comparison curve for the estimated hazard rate from Table 5 and Table 6 is shown in Figure 8 and Figure 9.

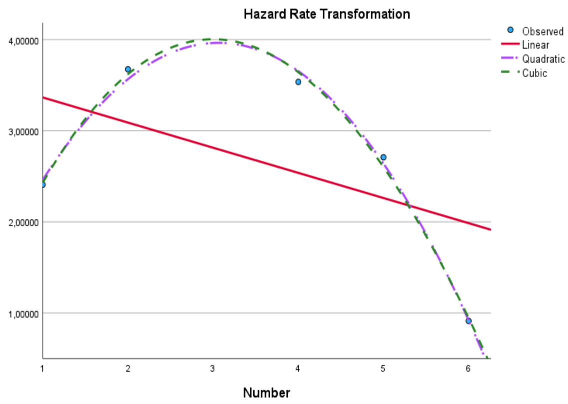

Figure 8. Comparison Curve of the Estimated Hazard Rate for the Assumed Linear Distribution of Waiting Time

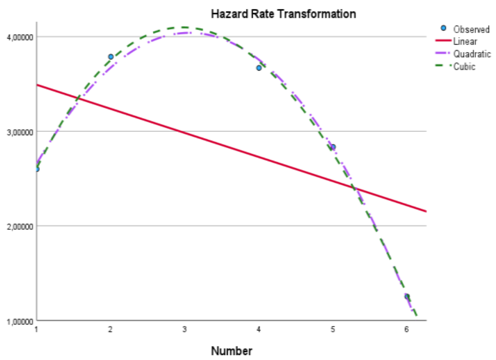

Figure 9. Comparison Curve of the Estimated Hazard Rate for the Assumed Exponential Distribution of Waiting Times

Based on Table 6 and Table 7, the estimated regression equations with linear, quadratic, and cubic models were obtained. At the significance level of $\alpha = 0.05$, the quadratic regression model $\hat{\mu}^*_{t_0} = 0.652 + 2.162t_0 - 0.353{t_0}^2$ for the assumed linear distribution of waiting times and $\hat{\mu}^*_{t_0} = 0.995 + 1.989t_0 - 0.325{t_0}^2$ for the assumed exponential distribution of waiting times were significantly different, while the linear and cubic regression models were not significantly different. Based on Figure 8 and Figure 9, the quadratic and cubic regression curves follow the data pattern. The best model chosen is a quadratic model for the assumed exponential distribution of waiting times based on R Square, Mean Square Error, and real regression tests. If this quadratic model is assumed to be able to estimate the hazard rate value in real life in the future, then this model can be used. By using reverse transformation, the hazard rate equation in Table 3 was obtained for an assumed exponential distribution of waiting times, which is $\hat{\mu}_{t_0} = \frac{1}{{\hat{\mu}^*_{t_0}}^2}$. For example, the value of the flood hazard rate in 2026 is

$$\hat{\mu}_{10} = \frac{1}{{\hat{\mu}^*_{10}}^2}$$

$$= \frac{1}{(0.995+1.989(10)-0.325(10)^2)^2} = 0.00741.$$

This means that the estimated level of flooding in 2026 is 0.00741. This means that the level of flooding in 2026 will be 0.00741.

The hazard rate is related to the estimated chance of an event occurring in a certain area. The probability of an event occurring is $\mu_{t_0}\Delta_{t_0}$ in the interval $(t_0, t_0 + \Delta_{t_0}]$, while the probability of no event occurring in the interval $(t_0, t_0 + n]$ is

$${}_n p_{t_0} = e^{-\int_0^n \mu(t_0+s)\,ds}.$$

In other words,

$${}_{t_0} p_0 = e^{-\int_0^{t_0} \mu(s)\,ds}$$

is the probability of no flood event in the interval $(0, t_0]$. Therefore, the probability of at least one flood event in the interval $(t_0, t_0 + \Delta_{t_0}]$ is

$${}_{t_0} q_0 = 1 - {}_{t_0} p_0 = 1 - e^{-\int_0^{t_0} \mu(s)\,ds}$$

assuming that there are no flooding events in the interval $(0, t_0]$.

The estimation of flood events in the interval $(0, t_0]$ was calculated by using the flood hazard rate value in the interval $(0, t_0]$, determining the probability that there will be no flood events in the interval $(0, t_0]$ and determining the probability that there will be at least one flood event in the interval $(t_0, t_0 + \Delta_{t_0}]$. Estimated flood events were assessed through the hazard rate using a quadratic regression model for exponentially distributed waiting times, namely

$$\hat{\mu}_{t_0} = \frac{1}{{\hat{\mu}^*_{t_0}}^2}$$

$$= \frac{1}{\left(0.995+1.989t_0-0.325{t_0}^2\right)^2}$$

based on R Square, MSE, and real regression tests. The probability of a flood occurring in the interval $\left(t_0, t_0 + \Delta_{t_0}\right]$ using the data obtained is shown in Table 8 and Figure 10.

Table 8. Estimated Probability of Flooding for the Assumed Exponential Distribution of Waiting Times in the interval $\left(t_0, t_0 + \Delta_{t_0}\right]$

No.	Interval	t_0	$\hat{\mu}_{t_0}$	${}_{t_0}p_0$	${}_{t_0}q_0$
1.	(0,1]	1	0.141	0.868	0.132
2.	(1,2]	2	0.074	0.862	0.138
3.	(2,3]	3	0.061	0.831	0.169
4.	(3,4]	4	0.071	0.752	0.248
5.	(4,5]	5	0.126	0.532	0.468
6.	(5,6]	6	0.662	0.018	0.982

Information:
$\hat{\mu}_{t_0}$: Estimator of flood hazard rate at t_0.
${}_{t_0}p_0$: Probability that no flooding will occur in the interval $(0, t_0]$.
${}_{t_0}q_0$: Probability of at least one flood in the interval $(0, t_0]$.

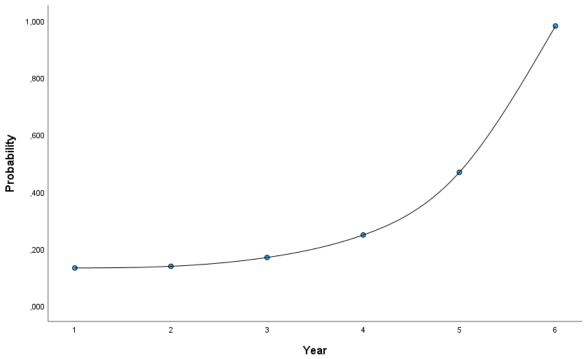

Figure 10. Graph of the Relationship between the Probability of a Flood Occurring and the Time Difference t_0 Since the Last Flood

Based on Table 8, if it is known that a flood has not occurred at $t_0 = 2$, then the probability of at least one flood event in the interval (2,3] is 0.169. Figure 10 shows the probability of a flood occurring after an exponential waiting time. The horizontal axis shows the time interval t_0 since the flood last occurred, and the probability of at least one flood event occurring within a six-year period is 0.982.

CONCLUSIONS AND SUGGESTIONS

One method for determining the estimated hazard rate value is the single-decrement method with a likelihood approach. The hazard rate estimator for the assumed linear distribution of waiting times is $\hat{\mu}_{t_0} = \frac{\hat{q}_{t_0}}{1-\hat{q}_{t_0}}$ with $\hat{q}_{t_0} = \frac{d_{t_0}}{n_{t_0}}$, while the hazard rate estimator for the assumed exponential distribution of waiting times is $\hat{\mu}_{t_0} = \frac{d_{t_0}}{\left[(n_{t_0}-d_{t_0})+\sum_{i=0}^{d_{t_0}} y_i\right]}$. Hazard rate calculations used flood data in Parepare in 2017–2023. Based on the results of the data analysis, the best parametric model for the two waiting time distribution assumptions is the quadratic regression model, and the percentage of this model that can be used is at least 90% based on R Square, Mean Square Error and the real regression test. After obtaining the estimated hazard rate value, the estimated probability of at least one flood event occurring in the interval $\left(t_0, t_0 + \Delta_{t_0}\right]$ is determined.

Suggestions for further research are to estimate the hazard rate using other methods, waiting time assumptions, or parametric models.

REFERENCES

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., & Nesbitt, C. J. (1997). *Actuarial mathematics*. The Society of Actuaries.

Collet, D. (2003). *Modelling survival data in medical research* (2nd ed.). Chapman and Hall.

Darwis, S., Sunusi, N., Gunawan, A. Y., Mangku, I. W., & Wahyuningsih, S. (2009). Single decrement approach for estimating earthquakes hazard rate. *Advances and Applications in Statistics*, *11*(2), 229–237.

Kleinbaum, D. G., & Klein, M. (2007). *Survival Analysis, a self-learning text*. Springer.

Krishnamoorthy, K. (2015). *Handbook of statistical distributions with applications* (2nd ed.). Chapman and Hall/CRC.

Le, C. T. (1997). *Applied survival analysis*. John Willey.

Ogata, Y. (1999). Seismity analysis through point-process modelling: A review. *Pure and Applied Geophysics*, *155*(2–4), 471–507. https://doi.org/10.1007/s000240050275

Sunusi, N. (2010). *Pengembangan estimasi hazard rate proses titik temporal dan aplikasinya pada prakiraan kemunculan gempa*. Institut Teknologi Bandung.

Sunusi, N., Jaya, A. K., Islamiyati, A., & Raupong. (2013). *Studi temporal point process pada analisa prakiraan peluang waktu kemunculan gempa, mitigasi dan manajemen sumber daya alam*. Universitas Hasanuddin.

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, *11*(4), 503–538. https://doi.org/10.1016/0169-2070(95)00621-4

Yohana, C., Griandini, D., & Muzambeq, S. (2017). Penerapan pembuatan teknik lubang biopori resapan sebagai upaya pengendali banjir. *Jurnal Pemberdayaan Masyarakat Madani (JPMM)*, *1*(2). https://doi.org/10.21009/jpmm.001.2.10