Monte Carlo method to valuate CAT bonds of flood in Surabaya under Jump Diffusion Process

Hengky Kurniawan¹, Endah RM Putri², Chairul Imron³, Dedy D. Prastyo⁴
¹,²,³ Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
⁴ Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia
E-mail: hengky16@mhs.matematika.its.ac.id

Abstract. The characteristic of natural disasters is that they tend to be rare. However, once they occur, natural disasters can result in enormous losses. In Indonesia, the natural disasters that cause the most losings are floods. Insurance companies will suffer immensely if many policyholders make claims due to losings from this natural disaster. Therefore, a financial instrument is urgently needed to transfer this risk to another party who is better at bearing the risk of loss. This financial instrument is called a CAT bond. The causes of flooding diverge into the hydrological aspect and hydraulic aspect. Maximum rainfall affects the hydrological aspect. In this research, we use a study case of flood in Surabaya. The maximum rainfall is predicted using the Gaussian Process Regression. Furthermore, the losses due to flooding obtained are modeled by the Jump Diffusion Process. The MAPE value obtained by the Jump Diffusion Process is much smaller than the Diffusion Process. From the results of the Monte Carlo simulation, it leads to the conclusion that the trigger value of the CAT bond contract significantly affects the CAT bond price, while the interest rate and the reduction proportion have a slightly significant effect.

1. Introduction
Risk is a danger or consequences that can occur as a result of an ongoing process or a future event [1]. According to the Indonesian National Board for Disaster Management (BNPB), there have been 1,426 natural disasters in 2019 [2]. The number of disasters tends to increase from year to year so that it gives an impact in the form of increased disaster risk. Disaster risk is the potential loss caused by a disaster in an area and a certain time which can be in the form of death, injury, illness, life threatened, loss of security, displacement, damage or loss of property, and disruption of community activities due to a combination of danger, vulnerability, and capacity [3].

In 2019, floods ranked second as a disaster with the highest frequency of occurrence. Tornado was the most frequent disaster with 568 recorded events followed by floods with 385 recorded events [2]. Following BNPB notes, the total loss caused by the flood in 2019 entirely has reached billions of rupiah. The loss is measured from infrastructure damaged by flooding. The impact of flooding is not only in the form of infrastructure damage, but also health problems in the community, disrupting the way of transportation, and even in certain cases fatalities can occur [4].
One of the solutions to overcome the risk of loss is to transfer the risk to other parties who are better prepared to bear the losses incurred. Zong Gang Ma and Cho Qun Ma (2012) state that although natural disasters do occur rarely, the resulting losses can be enormous [5]. Therefore, it is necessary to have an insurance company that guarantees value and transfers disaster risk by buying a reinsurance contract. In carrying out their duties, insurance companies require financial instruments where risks arising from natural disasters are linked to bonds. These bonds is called catastrophe bonds or abbreviated to CAT bonds which works by insuring disaster events to the capital market [6].

Much research has been done on the CAT bonds. One of them, Bodoff (2009) modeled CAT bonds valuation as a linear function of losses incurred to analyze the market price of CAT bonds [7]. Van Son (2013) shows that the CAT bonds, in addition to being affected by disaster risk, are also influenced by exchange rate volatility and its correlation with the interest rate [8]. Siyamah (2019) conducted a comparative analysis of the valuation of CAT bonds influenced by currency exchange rates using the Monte Carlo method and the quasi-Monte Carlo method [9]. In this research, the price of flood CAT bonds is calculated based on runoff discharge and flood loss that is modeled by Jump Diffusion Process with constant interest rates. The runoff discharge is determined based on hydrological discharge and hydraulic discharge.

2. Hydrological discharge

Hydrological discharge is influenced by maximum rainfall and watersheds specification. The maximum rainfall is predicted using the regression method based on data of maximum rainfall from six urban villages in Surabaya. These six urban villages are Keputih, Kedung Cowek, Gubeng, Wono Rejo, Wonokromo, and Gunung Sari. To facilitate the prediction process and make it easier to find out the difference between the predicted results and the actual data, the maximum daily rainfall data needs to be cumulatively summed.

Gaussian Process Regression does not require normally distributed data. Gaussian Process Regression utilizes a Bayesian approach in regression. In the Gaussian Process Regression, there are terms of prior distribution and posterior distribution. The prior distribution is a parameter probability based on the observed data. The posterior distribution is a conditional probability obtained from prior distribution information and the dataset. For ease of understanding, let’s say $y = wx + \varepsilon$. The prior distribution is $p(w)$ while the prior distribution is obtained by equation

$$p(w|y, X) = \frac{p(y|X, w)p(w)}{p(y|X)}$$ (1)

where $w$ represents parameter value, $X$ represents the dataset or predictor set, $y$ represents the target, $p(w|y, X)$ represents posterior distribution, $p(y|X, w)$ represents the likelihood, and $p(y|X)$ represents marginal likelihood. Predictive distribution is calculated by weighting all possible results of the prediction with the calculated posterior distribution to obtain the prediction at the point $x^*$. Thus, the predictive distribution is

$$p(f^*|x^*, y, X) = \int_{w} p(f^*|x^*, w)p(w|y, X)dw$$ (2)

Prior and likelihood is often assumed to be normally distributed (Gaussian) to facilitate the integration process. Based on this assumption and the completion of the predictive distribution, a predictive result point can be obtained with the calculated mean and variance [10].

The process of implementing Gaussian Process Regression utilizes the Matlab toolbox named Regression Learner because it can conduct training and testing for the model chosen by the user.
The advantage of using this toolbox is that Matlab can determine the best parameters by itself. Some methods that can be chosen to do hyperparameter optimization are Bayes, Grid Search, or Random Search approaches. In this simulation, hyperparameter optimization is conducted using the Grid Search algorithm.

Grid Search algorithm is also known as hyperparameter optimization. In general, this algorithm uses a \( d \) dimension matrix where \( d \) represents the number of parameters. The elements of this matrix are all possible pairs of parameter values or \((x_1, x_2, \ldots, x_d)\) where \(x_1\) is one of the possible values of the first parameter, \(x_2\) is one of the possible values from the second parameter, and so on. The stages of this algorithm are

(i) Determine the lower and upper limits of the value of each parameter to form the grid.
(ii) For all elements on the grid, do the following.
   (a) Calculate value based on the objective function.
   (b) If the value of this element is better than the value of the previous element (if exist), note the parameter value of this element.

Grid Search algorithm requires enormous memory and long computation time. Even though, this algorithm can discover the best global parameters because this algorithm checks all possible parameter value. So, the Grid Search algorithm is suitable for a few data or simple model cases. This case is a simple model case because there are few parameters.

![Regression result of Keputih](image)

**Figure 1.** Regression result of Keputih

The accuracy of the Gaussian Process Regression in predicting the maximum daily rainfall is enormous. Even the regression curve formed can be said to very coincide with the real curve. Therefore, it is not surprising that the results of the Gaussian Process Regression only have MAPE value of 0.96865\% only. The watersheds specification for each watershed is obtained by randomizing the specification type. The predicted maximum rainfall and randomized watersheds specifications results are processed to obtain rainfall intensity using the Mononobe formula

\[
I_t = \frac{R_{24}}{24} \left( \frac{24}{t_c} \right)^{\frac{2}{3}}
\]

where \(I_t\) represents the maximum intensity (mm/hour) for rain with concentration time of \(t\) and \(R_{24}\) represents the maximum rainfall in 24 hours. Concentration time is obtained by randomizing
the distribution of channel types and implementing the Kerby formula

\[ t_c = 1.44 \left( n_k \times \frac{l}{\sqrt{s}} \right)^{0.467} \]  \tag{4}

where \( n_k \) represents the roughness coefficient, \( l \) represents the length of the furthest water trajectory, and \( s \) represents the average slope of the channel. Furthermore, the calculation of hydrological discharge is calculated using rational method

\[ Q_a = 0.278 \times C \times I_t \times A \]  \tag{5}

where \( Q_a \) represents the hydrological discharge, \( C \) represents the drainage coefficient and \( A \) represents the area of drainage (km\(^2\)). The drainage coefficient in an area consisting of several regions with different drainage coefficients can be approached using the combined drainage coefficient value [11]. The combined drainage coefficient is a weighted average of the drainage coefficient. The combined drainage coefficient value can be obtained by the formula

\[ \bar{C} = \frac{\sum_{j=1}^{n} C_j A_j}{\sum_{j=1}^{n} A_j} \]  \tag{6}

where \( \bar{C} \) represents the value of the combined drainage coefficient, \( C_j \) represents the drainage coefficient of type \( j \), \( A_j \) represents the area of channel type \( j \), and \( n \) represents the number of channel types.

3. Hydraulic discharge

This process consisting of drainage channel randomization and implementation of the Manning formula. After randomizing the drainage channel, on each type of channel, a hydraulic radius is calculated based on the Manning equation. The calculation of the hydraulic radius for each U-Ditch channel specification is made using the equation

\[ R_h = \frac{A_{cs}}{P_w} \]  \tag{7}

where \( R_h \) represents the radius of hydraulics, \( A_{cs} \) represents the cross-sectional area of the channel and \( P_w \) indicates the circumference of the affected section. The calculated hydraulic radius then used to obtain the flow rate based on the Manning formula

\[ Q_b = \frac{1}{n_m} \times R_h^{\frac{2}{3}} \times s^{\frac{1}{2}} \times A_{cs} \]  \tag{8}

where \( Q_b \) represents the hydraulic discharge.

The hydraulic discharge is calculated for all each region. In the Manning equation, values are needed \( n_m \) is the Manning roughness coefficient. Because it has been assumed drainage channel are built using U-Ditch, then the Manning roughness coefficient used is the coefficient roughness of concrete. This is obtained based on the main ingredient in making U-Ditch is concrete. The value of the roughness coefficient from concrete is 0.015 [12].

4. Flood and losses

Hydrological discharge and hydraulic discharge are used to predict which urban villages will be flooded. For each exposed village, the number of people who are exposed, and the size of each village is sought. Because there is no data on the number of inhabitants per urban village,
the population of an urban village is sought using a proportion of urban villages with an area of sub-district. This can be done with the assumption that population density is uniform for each sub-district observed.

Factors considered to determine the losses caused by floods are regional factors, life factors, and multipliers. From each area expected to be exposed by flood, the area of the flooded area is the area multiplied by a random number with a value between 0 and 1. When written in an equation, it becomes

$$F_a = \alpha \times C_a$$  \hspace{1cm} (9)

where $F_a$ represents flooded area, $\alpha$ represents the distributed randomness factor $U(0,1)$, and $C_a$ represents candidate area. From each area exposed to flooding, the number of individuals exposed to flooding is the number of lives in the area multiplied by a random number between 0 and 1 which when written becomes

$$F_p = \beta \times C_p$$  \hspace{1cm} (10)

where $F_p$ represents flooded people, $\beta$ represents the distributed randomness factor $U(0,1)$, and $C_p$ represents candidate people. The multiplier factor is a comparison of the hydrological discharge with the hydraulic discharge. It used as material for calculating losses due to the greater runoff discharge caused the higher loss. The equation to calculate the multiplier factor per region for each time is

$$M_f = \frac{Q_a}{Q_b}$$  \hspace{1cm} (11)

where $M_f$ represents multiplier factor or multiplier, $Q_a$ represents the hydrological discharge, and $Q_b$ represents the hydraulic discharge. From these three factors, the formula for calculating losses per region is

$$L_w = (L_p \times F_p + L_a \times F_a) \times M_f$$  \hspace{1cm} (12)

with $L_w$ declaring loss per region, $L_p$ declaring loss per person exposed, and $L_a$ declaring Loss per area or loss per m $^2$. The following simulation use the value of $L_p = 25000$ and the value of $L_a = 10000$. The losses from the six urban villages are added up to get the total losses.

![Figure 2. Summed losses](image)

Loss prediction is done using the Jump Diffusion Process curve that best matches the existing loss curve. To find out how well the curve formed with the existing loss curve is by calculating the MAPE value. Therefore, losses are accumulated to avoid the number of zeros.
Flood losses obtained are used to form a flood loss model. Flood losses are modeled by following Jump Diffusion Process. Merton’s Jump Diffusion Process model can be written as [13]

\[ dL_t = \mu dt + \sigma dW_t + (Y - 1)dN_t \]  

where \( \mu \) represents drift or trend, \( \sigma \) represents volatility, \( W_t \) represents Brownian motion, \( Y \) represents the jump factor normally distributed with an average of \( \mu_J \) and standard deviation of \( \sigma_J \), and \( N_t \) declare a standard Poisson process with the parameter \( \lambda \). The value of \( \mu \) can be found by the average loss, while \( \sigma \) can be found from the standard deviation of losses. If the jumping aspect is omitted or given a value equal to zero, the above model will change to the Diffusion Process model. Because it is random, curves can move up or down. As the cumulative loss curve is a rising monotonous curve, an adjustment procedure is needed for the resulting curve. The adjustment made is if the loss at \( t \) is smaller than the loss at \( t-1 \), then the loss at \( t \) is equal to the loss at \( t-1 \).

Then a jump detection is performed to obtain a \( \lambda \) jump frequency, an average jump of \( \mu_J \), and a jump standard deviation of \( \sigma_J \). Jump detection is performed by the Sequential Average algorithm. The stages of this algorithm are

(i) Specify the length of the first data for which there are no jumps.
(ii) For all data starting with data after the last data in step 1 do the following.
   (a). Calculate local average values up to the current data.
(b). Calculate the local standard deviation to the current data.
(c). If the difference between the current data and the local average divided by the local standard deviation is more than 3, note that a jump occurs in this data.

Figure 5. Matching curve with jump

It can be seen that the best curve of the Jump Diffusion Process model has a smaller MAPE value compared to the best curve of the Diffusion Process model result. In other words, the curve result of Jump Diffusion Process model is better to represent data loss even though the difference between the two is not significant.

5. CAT bonds

In the CAT Bond, if a natural disaster occurs during the CAT Bond contract and the loss due to natural disaster received by the sponsor exceeds $H$ (a threshold specified in the contract), the loss from the disaster will be paid by the issuer to the sponsor. Conversely, if no disaster triggers the CAT Bond, all money deposited by the issuer as part of the CAT Bond will be returned to the investor. The CAT bond price model can be obtained by looking for expectations of the formed cash flow structure.

In the cash flow, there is a trigger time for CAT Bond which will be denoted by $\eta$. The $\eta$ value is the first time the loss value reaches the $H$ threshold or in other words, $\eta$ is the minimum $t$ where the cumulative loss at time $t$ or $L_t$ is greater or equal to $H$. The value of $\eta$ can be written as the following equation.

$$P(t, T) = \mathbb{E}_Q \left[ D(t, T) \cdot \Psi \bigg| F_t \right]$$

where $P(t, T)$ represents the price of CAT Bond at the time of $t$ with the end time of the contract $T$ and $\Psi$ represents the cash flow [8]. Because the price of CAT Bond is for investors, the determination of the price of CAT Bond needs to be made in terms of investors. Cash flow that will occur at the end of the $T$ contract is that the investor will get all his money back if a natural disaster occurs after the contract ends. If a natural disaster occurs before the CAT Bond contract period expires, investors will get their money back after deducting the proportion of $\omega$ where this discount is the amount of money given issuer as part of the compensation to the sponsor. The value of $\omega$ is between 0 and 1 depending on the amount of loss incurred. Such cash flows can be written as the following equation.

$$\Psi = V - \omega V \mathbb{1}_{\eta<T} = V \cdot (1 - \omega \mathbb{1}_{\eta<T})$$

(15)
By utilizing the expectation nature of the multiplication operation and the fundamental nature of the forward martingale measure, the equation above can be written as

\[ P(0, T) = D(0, T) \cdot V \cdot E_Q \left[ (1 - \omega I_{\eta < T}) \bigg| \mathcal{F}_t \right] \]  

(16)

After getting the required model, a simulation is performed using MATLAB R2019b. The Monte Carlo simulation is constructed to obtain cash flow expectations. The principle of the Monte Carlo algorithm is to average \( N \) probabilities of an experiment being randomly generated, so that when it is averaged it will approach the original answer. The steps of this algorithm are

(i) Define the stochastic model.
(ii) Set the number of iteration.
(iii) For every iteration, do the following.
   (a). Generate random number as the input for the stochastic model.
   (b). Compute the result of the stochastic model based on the generated random number.
(iv) Calculate the average of the obtained results.

Mathematically, this simulation can be written as [14]

\[ \frac{1}{N} \sum_{i=1}^{N} f(x_i) \rightarrow \int_{0}^{1} f(x)dx = E(f(x)) \]  

(17)

for \( N \to \infty \) and \( x_i \) follows uniform distribution \( (x_i \sim U(a,b)) \) with lower bound \( a \) and upper bound \( b \). For example, parameter values are used following Table 1.

| Parameter | Value     |
|-----------|-----------|
| \( T \)   | 1         |
| \( V \)   | \( 10 \times 10^9 \) |
| \( \mu \)  | \( 8.5802 \times 10^6 \) |
| \( \mu_J \) | \( 1.8201 \times 10^7 \) |
| \( \lambda \) | 0.0055 |
| \( \sigma \) | \( 4.8758 \times 10^6 \) |
| \( \sigma_J \) | \( 3.3335 \times 10^6 \) |

Table 1. Parameter value

We use one as the value of \( T \) cause we assumed the contract for the bond is one year. As stated in the introduction, the loss in 2019 is in billions of rupiah. That’s the reason for the face value of the bond is supposed to be in billions. Face value is determined by the bond issuer, but as an example of simulation, 10 billion rupiahs is chosen. The other five parameters are based on the calculation in the Jump Detection process. In addition to the parameters contained in Table 1, there are also parameters of the number of iterations or \( N \) which are input from the user. There are also threshold parameters or \( H \), proportion parameters or \( \omega \) and interest rate parameters or \( r \). For example, set the value of \( N \) equal to 10,000 and note the value of \( H \) equals \( 6 \times 10^8 \), \( \omega \) equals 0.8, and \( r \) equals 0.05. From these four parameters it will be found that the appropriate CAT Bond price is \( 3,37420 \times 10^9 \). Then it is needed to analyze the relationship between CAT Bond prices and the input parameter. As the first step in the analysis, the relationship between CAT Bond prices and \( N \) are tested. In this simulation, the value of \( N \) starts from 1100 to 10000 with a step of 100. In this simulation the value of \( H \) is set at \( 6 \times 10^8 \), the value of \( \omega \) is set at 0.8 , and the value of \( r \) is set at 0.05.
The price of CAT Bond is sufficiently converging for a different $N$ value. The price of CAT Bond formed ranges from $3 \times 10^9$ to $4 \times 10^9$. Therefore, it can be said that $N$ does not really affect the price of CAT Bond. Henceforth, we will look for the effect of the threshold $H$ on the price of CAT Bond. In this simulation the value of $H$ starts from $1 \times 10^8$ to $50 \times 10^8$ with step for $10^8$. In this simulation the value of $N$ is set at the value of 1000, the value of $\omega$ is set at the value of 0.8, and the value of $r$ is set at the value of 0.05.

The price of CAT Bond has changed significantly for different $H$ values. Therefore, it can be said that $H$ greatly influences the price of CAT Bond especially at $H$ between $1 \times 10^8$ to $20 \times 10^8$, whereas for $H$ more than $2 \times 10^8$ CAT Bond prices continue to increase even though they are not significant. In addition, the price of CAT Bond also increased rapidly in the range of threshold $10 \times 10^8$ to $15 \times 10^8$. Henceforth, we will look for the effect of the value of the $\omega$ proxy on the price of CAT Bond. In this simulation the value of $\omega$ starts from 0.01 to 0.9 with step of 0.01. In this simulation the value of $N$ is set at the value of 1000, the value of $H$ is set at the value of $6 \times 10^8$, and the value of $r$ is set at the value of 0.05.
The price of CAT Bond has changed significantly for different $\omega$ values. The greater value of $\omega$ results in a decrease in prices on CAT Bond. Henceforth, we will look for the effect of $r$ on the price of CAT Bond. In this simulation the value of $r$ starts from 0.01 to 0.2 with step for 0.01. In this simulation the value of $N$ is set at the value of 1000, while the value of $H$ is set at the value of $6 \times 10^9$. The results of this simulation presented in the following figure.

It appears that the price of CAT Bond changes is not too significant for different values of $r$. The price formed is still around $3 \times 10^9$. Therefore, it can be said if $r$ enough to affect the price of CAT Bond, where for $r$ that will increase will cause the price of CAT Bond continues to decline.

6. Conclusion

The model used to calculate losses per area is the number of individuals who are victims multiplied by the amount of loss per person, plus the area of submerged area multiplied by the amount of loss per square meter, the result of the sum then multiplied by the multiplier factor. The calculated losses then will be modeled. The equation used to model the flood losses due to rainfall obtained by the regression method is

$$dL_t = 8.5802 \times 10^6 dt + 4.8758 \times 10^6 dW_t + (Y - 1)dN_t$$
where \( W_t \) is Brownian motion, \( Y \) is normally distributed jump factor \( N(1.8201 \times 10^7, 3.3335 \times 10^6) \), \( N_t \) is standard Poisson process with \( \lambda = 0.0055 \). Based on this loss model, CAT Bond price for a flood is determined by the expectation of the product of the discount factor multiplied with cash flow.

From the simulation that had already done, the number of iterations or \( N \) does not significantly affect the price of CAT Bond. The amount of the threshold or \( H \) greatly influences the price of CAT Bond especially at \( H \) between \( 1 \times 10^8 \) to \( 20 \times 10^8 \), whereas for \( H \) more than \( 2 \times 10^8 \) the price of CAT Bond remains increased though not significantly. The value of \( \omega \) or the proportion of reduction that is getting bigger results in a price reduction on the CAT Bond. The interest rate or \( r \) is enough to affect the price of CAT Bond, where for an increasing \( r \) will cause the price of CAT Bond continues to decrease.

References

[1] Cline P B 2003 *Wilderness Risk Managers Conference, NOLS*
[2] BNPB 2019 Data Informasi Bencana Indonesia [http://bnpb.cloud/dibi/beranda](http://bnpb.cloud/dibi/beranda)
[3] Indonesia U P 2010 Pusat Pendidikan Mitigasi Bencana [http://p2mb.geografi.upi.edu/Mitigasi\_Bencana.html](http://p2mb.geografi.upi.edu/Mitigasi\_Bencana.html)
[4] Rosyidie A 2013 *Journal of Regional and City Planning* 24 241–249
[5] Ma Z G and Ma C Q 2013 *Insurance: Mathematics and Economics* 52 243–254
[6] Cox S H and Pedersen H W 2000 *North American Actuarial Journal* 4 56–82
[7] Bodoff N and Gan Y 2009 *Casually Actuarial Society E-Forum, Spring 2009*
[8] Lai V S, Parcollet M and Lamond B F 2014 *International Review of Financial Analysis* 33 243 – 252 ISSN 1057-5219
[9] Siyamah I 2019 *Analisis Perbandingan Valuasi Cat Bond dengan Metode Monte Carlo dan Kuasi Monte Carlo* Master’s thesis Institut Teknologi Sepuluh Nopember
[10] Williams C K and Rasmussen C E 2006 *Gaussian processes for machine learning* vol 2 (MIT press Cambridge, MA)
[11] Suripin 2004 *Sistem Drainase Perkotaan yang Berkelanjutan* (Andi)
[12] Wong T S and Lim C 2006 *Journal of Hydrology* 331 205–218
[13] Matsuda K 2004 *Department of Economics. The Graduate Center, The City University of New York*
[14] Gönçü A 2009 *Monte Carlo and Quasi-Monte Carlo Methods in Financial Derivative Pricing* dissertation Florida State University