Proceedings of Science and Mathematics
Faculty of Science, Universiti Teknologi Malaysia
Vol. 18, 2023, page 98 – 106

# Groundwater Level Prediction Based on River Water Level Using Stochastic Differential Equation (SDE)

**Nurul Izzah Jefri, Arifah Bahar***
Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia
*Corresponding author: arifah@utm.my

**Abstract**
The objective of this study is to predict groundwater levels based on river water levels using a stochastic differential equation (SDE) and the Geometric Brownian Motion (GBM) method. Groundwater, as a vital natural resource, plays a crucial role in preserving ecosystems and meeting diverse human needs. Accurate predictions of groundwater levels are essential for effective water resource management and sustainable planning. The benchmark data for this study comprised of groundwater level measurements from five test wells located in Loji Rawatan Air Sungai Linggi and Sungai Muda, provided by the National Water Research Institute of Malaysia (NAHRIM) for three days in October and November 2014 and August 2015. Additionally, the river water level data spanning a ten-years period from 2010 to 2020, obtained from the Department of Irrigation and Drainage Malaysia (JPS), were used as secondary data to estimate the parameters required for predicting groundwater fluctuations through SDE. Non-linear regression analysis using Microsoft Excel Solver was employed to assess the benchmark data and establish correlations between groundwater levels and physical parameters such as river water levels and precipitation. The GBM approach was implemented in RSTUDIO version 2023.03.1-446 to simulate groundwater water levels, generate stochastic processes, and calibrate the relevant parameters using historical data. The GBM groundwater equation that was fitted with the parameters from non-linear regression described the relationship between both variables. The accuracy and reliability of the model were evaluated using the root mean square error (RMSE), a statistical measure. The results obtained from the non-linear regression analysis revealed complex relationships and interactions between variables. The GBM method demonstrated its ability to generate stochastic processes for simulating groundwater water levels and predicting the groundwater level. Ultimately, the groundwater levels were calculated, providing insights into hydrological activities and characteristics.

**Keywords:** groundwater level; river water level; non-linear regression; Geometric Brownian Motion (GBM); water level prediction

## 1. Introduction

The presence of water is essential for all life on this earth. It plays an important role in every aspect such as in human body, blood, trees, air, oceans, rivers, and soil. The total amount of water on the planet is about 1.4 × 10^9 km^3 (Fitts, 2002) and distributed among the main reservoir such as oceans, surface water and groundwater. People use a lot of surface water than the groundwater because it is closer and easy to use. Because of that, in Malaysia, the use of groundwater is relatively low at only 3% (Hussin et al, 2020). Occasionally, Malaysia Peninsula will have heavy rain and have a considerable amount of river, groundwater is not their first choice as the water supply.

Groundwater level prediction is crucial in Malaysia to manage geological issues and mitigate flooding risks. Increased groundwater levels can lead to water table rise, resulting in surface water flooding and infrastructure damage. Heavy rainfall, especially during the northeast monsoon season from November to March, contributes to the flooding occurrences in the country.

In December 2021, Malaysia faced severe floods due to heavy rainfall, leading to significant consequences. Additionally, the Batang Kali Landslide Tragedy in December 2022 resulted in numerous casualties and injuries. These incidents highlight the importance of accurate groundwater level predictions for effective disaster management and mitigation efforts. During flood disasters in Malaysia, two stages of SOPs, namely Pusat Kawalan Operasi Bencana (PKOB) and Pos Kawalan Tempat Kejadian (PKTK), are implemented immediately to coordinate rescue operations and manage information centers. Typically, PKTK is led by the Chief District Police as the Operation Commander, with the District Fire Officer as the deputy. In the 2021 flood, 42,265 members from various responder agencies and 1,843 combined assets were involved.

The Social Welfare Department provided daily necessities and 427,340 packs of food to flood victims. The Kementerian Perumahan dan Kerajaan Tempatan reported significant progress in cleaning and damage repair, covering 80.6% of affected areas in five states. In the Batang Kali tragedy, which occurred on December 16, 2022, a landslide took 31 lives. Malaysia's Prime Minister, Dato Seri Anwar Ibrahim, ordered an immediate and systematic search and rescue operation involving government departments.

The Search and Rescue team comprised 250 personnel from the police, army, Fire and Rescue Department, Special Malaysia Disaster Assistance and Rescue Team, civil defence force, and K9 dog tracking units. Therefore, the study for groundwater level prediction is important as we can predict the water level for future planning and reference. Before, the availability of groundwater resource in Selangor, Malaysia have been studied and due to the impact of extreme dry condition showed that the groundwater table declined in most area in Selangor (Hossain et al, 2019)  It also important for this project to identify the physical parameters used for predicting the groundwater level and to construct an appropriate model for groundwater to incorporate the parameter in predicting the groundwater level or head.

The objective of this study is identifying the physical parameters used in the study for predicting the groundwater level. This study also wants to determine appropriate model for river water level in predicting the groundwater fluctuations. Lastly, this study also wants to predict the groundwater level in monitoring wells at Loji Rawatan Air Sungai Linggi and Sungai Muda.

## 2. Literature Review

Groundwater is a vital resource essential for human survival, as it undergoes the hydrological process known as the water cycle. It is replenished through natural recharge, primarily driven by rainfall. Rainwater enters the earth through evaporation and precipitation, lifting the water level and groundwater level. The groundwater fills the underground pores, fractures, and spaces between grains of sand and cracks in rocks, forming an aquifer (Poeter et al. 2020). When the unsaturated zone can no longer hold water, gravity pulls the excess water down to the water table. Groundwater can be accessed by drilling wells and extracting water through pipes. Monitoring wells are also used to track water quality and levels.

Aquifers are underground layers of rocks and sediments that are saturated with groundwater. They are characterized by their porous and permeable nature, which allows water to seep through and accumulate as groundwater. Aquifers play a crucial role in storing and supplying water for various purposes. Aquifers may be categorized based on the kind of rocks or sediments that make up their composition (Brown, 2022). There are two types of aquifers, confined aquifer, and unconfined aquifer.

Under the soil, we will meet with unconfined aquifer then the confined aquifer which has a layer of impenetrable rock or clay above it.

The uppermost point of the saturated zone in the subsurface of the Earth is defined by the water table. It stands for the point at which the ground is completely saturated with water. Depending on geography, temperature, vegetation, and human activity, the water table's location may change. Wells are dug into the ground below the water level to ascertain the depth of the water table. Water will flow into the well, making it possible to gauge and keep track of the height of the water table.

Smith et al (2013) investigated the use of neural networks and polynomial regression as non-linear regression models for predicting groundwater levels. To capture the non-linear correlations between these factors, the researchers gathered long-term hydrological data, including precipitation, temperature, and groundwater levels. They then created non-linear regression models. The outcomes showed how well non-linear regression performed in predicting groundwater levels.

In some of the study in stock prices, geometric Brownian motion is proved to be more accurate than in forecasting the stock prices for two-week period (Augustini et al, 2018). Estember and Marana (2016) show that GBM produces accurate and effective forecast for stock prices compared to the artificial neural network. Hamdan et al (2020) used Geometric Brownian Motion to represent the future price for Malaysian gold prices, Kijang Emas.

## 3. Materials and methods

### 3.1 Data Processing

The data of groundwater level was collected in minutes and has different time increment compared with the river water level data. While river water level was collected daily for a year. Hence, the groundwater level data was aggregated to be in daily. Let the data from Loji Rawatan Air Sungai Linggi be as the example from the report.

*Table 1 Aggregated Groundwater Level in Loji Rawatan Air Sungai Linggi and River Water level from Sua betong near Sungai Linggi, Negeri Sembilan for three days*

| **Aggregated Groundwater Level Data for Loji Rawatan Air Sungai Linggi** | | | |
|---|---|---|---|
| No | Date | River Water level (m) | Groundwater Level (m) |
| 1 | 5-Nov-14 | 4.96 | 5.3 |
| 2 | 6-Nov-14 | 5.12 | 8.37 |
| 3 | 7-Nov-14 | 5.73 | 8.01 |

Due to aggregated data from minutes to days, the data points were reduced to three data points. The Loji Rawatan Air Sungai Linggi, located in Rantau, Negeri Sembilan, is a water treatment plant that operates within a specific timeframe. The data collected from this plant spans from 5th November 2014 to 8th November 2014, encompassing a total duration of 76 hours, equivalent to 3 days.

### 3.2 Non-Linear Regression (NLR) Analysis

The regression analysis was done to find the correlation between log transform groundwater level and log transform river water level (1) at the location. Non-linear regression was done based on the data and the coefficient was considered as the parameters used in GBM model.

$$Y = \epsilon + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{1}$$

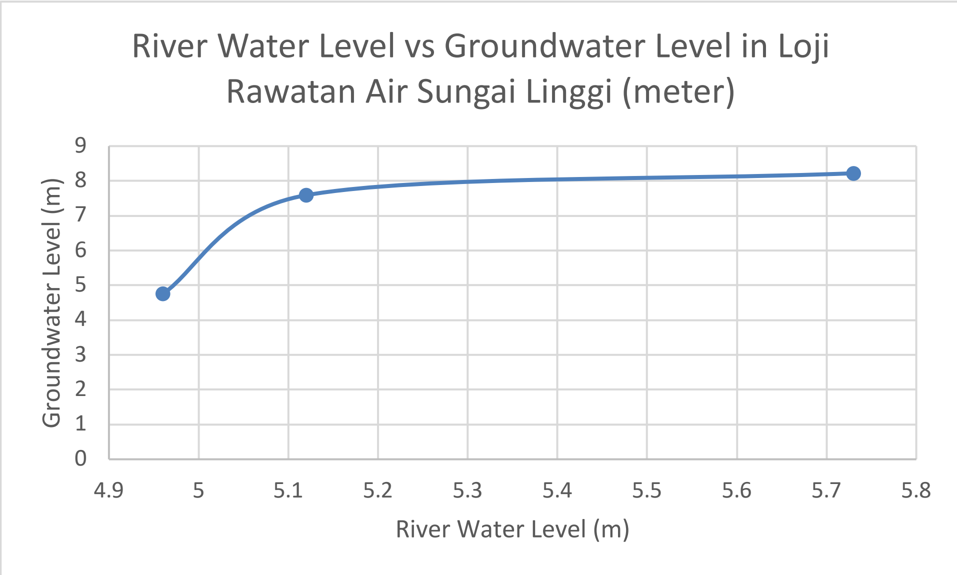


**Figure 1** Scatter plot of Groundwater level over River Water Level in Sg Linggi

Table 2 Excel Output for NLR in Sungai Linggi

| ***Output from The Non-Linear Regression Analysis*** | |
|---|---|
| $R^2$ | *0.344267* |
| ***p-Value =*** | *0.60082* |

From the regression analysis, the $R^2$was 0.344267 explained the correlation between the GW in Loji Rawatan Air Sungai Linggi and RW in Sua Betong. Estimated for 36% could be explain from the relationship and the p-Value = 0.5879 which is larger than alpha, α=0.05. From here, we can accept H_0= no significant correlation between groundwater level and river water level.

### 3.3 Parameter Estimation

By regression analysis that was done before, the mu (μ) and sigma (σ) are the coefficient result in the regression output were taken as the estimated parameters for every location of monitoring well for estimating the drift and diffusion in GBM.

Table 3 Estimated Parameters of Mu and Sigma using NLR

| ***Estimated Parameters*** | | |
|---|---|---|
| | ***Mu (μ)*** | ***Sigma (σ),*** |
| ***Sg Linggi*** | *1.468494* | *0.026513* |
| ***W1*** | *5.303425* | *2.054795* |
| ***W2*** | *4.342289* | *0.409985* |

As the parameters were taken from the regression analysis was for simulating the GWL in all locations, the simulated data will be divided into two part which were Testing Data and Training Data. From there, the training data from the simulated GWL will be used to calculate the the estimated parameters Drift and Diffusion used for predicting the future values of GWL in Geometric Brownian Motions. The estimated parameters will be validated by testing them using the 'Testing Data' from the simulated GWL data.

The Drift equation; $$\hat{\mu} = \frac{\sum_{i=1}^{N} G_i}{N.dt} \tag{2}$$

The Diffusion Equation; $$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N}(G_i - \bar{G})^2}{(N-1).dt}} \tag{3}$$

$\hat{\mu} = \text{Drift}$
$G_i = \text{Groundwater level difference}$
$N = \text{Number of data}$
$dt = \text{time step}$

Table 4 Estimated Parameters of Drift and Diffusion using GBM

| ***Calculated Parameters*** | | |
|---|---|---|
| | ***Drift (μ)*** | ***Diffusion (σ),*** |
| ***Sg Linggi*** | *4.4017381* | *2.101157968* |
| ***W1*** | *-0.0225829* | *0.0112315187* |
| ***W2*** | *-0.38518929* | *0.2574584639* |

The drift and diffusion parameters were calculated based on the estimated μ (mu) and σ (sigma) values obtained from the training dataset. The GBM model was then applied to the testing dataset to evaluate the model's predictive performance and generalization ability. By simulating 100 paths, the model provided a representative sample of potential future trajectories, enabling the analysis of important statistical properties.

### 3.4 Forecasting

The GBM model was then used to forecast and predict the 3 days of groundwater level based on the estimated parameters from each monitoring well in Sungai Linggi and Sungai Muda.

General GBM model; $$W(t) = W_0 \; exp[\mu t + \; \alpha\beta_t] \tag{4}$$

Loji Rawatan Air Sungai Linggi GBM Model:

$$W(t) = W_0 \; exp[\,(4.4017381)t + \;(2.101157968)\,\beta_t\,] \tag{5}$$

Table 5 Comparison of Raw Data and the Predicted GW

| | **Raw data from Aggregated Groundwater Level (m)** | **Predicted Groundwater Level (m)** |
|---|---|---|
| **Loji Rawatan Air Sungai Linggi** | 5.3 | 5.30 |
| | 8.37 | 5.31 |
| | 8.01 | 5.32 |

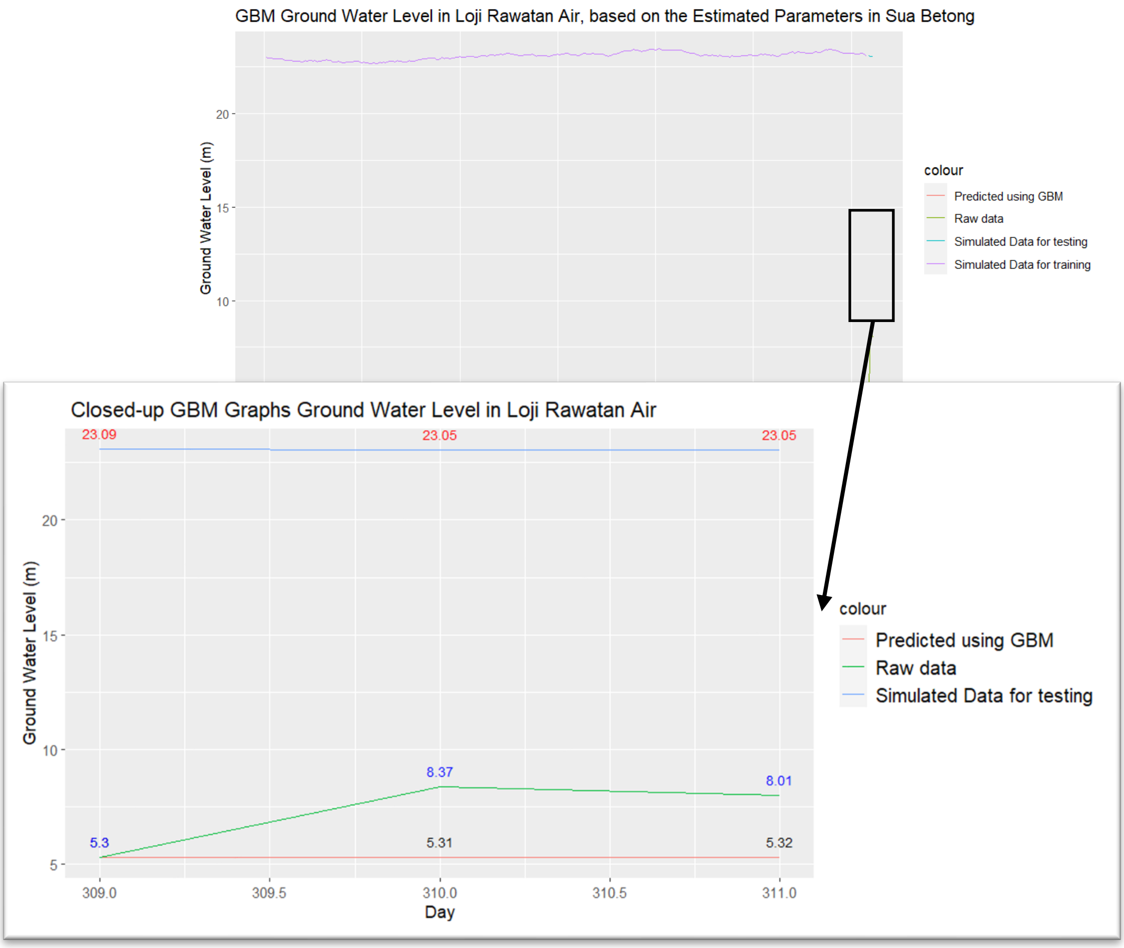


**Figure 2** Closed up GBM Graph on Predicted Values and Raw Data

### 3.5 Forecast Accuracy

Then, the accuracy was calculated using Root Mean Square Error (RMSE). It is a commonly used metric for evaluating the accuracy of a forecasting model or predicting model performance.

RMSE Loji Rawatan Air Sungai Linggi = 2.350713

The RMSE value for W0 which is the monitoring well in Loji Rawatan Air Sungai Linggi showed that the predicted values and actual values are in 1.141903 error. It means that much of error can be found from the predicted values and the actual simulated data and the GBM model can explain the correlation between the river water level collected in Sua Betong Sungai Linggi and groundwater level in Loji Rawatan Air Sungai Muda.

### 3.6 Simulation Analysis

Based on GBM equation, simulation is made for forecasting the groundwater level with 100 trajectories, with every trajectory using the GBM equation $W(t) = W_0 \ exp[\mu t + \alpha\beta_t]$ with 340 iterations. The GBM simulation analysis was done by using the estimated parameters Drift and Diffusion that was calculated in equation (4.3.2.1) and (4.3.2.2) in R studio based on the simulated GWL data.

For i=1;
$$W_1 = W_0 \; exp[\mu t + \alpha\beta_t] \quad (6)$$
$$W_2 = W_1 \; exp[\mu t + \alpha\beta_t] \quad (7)$$

.
.
.

$$W_{340} = W_{339} \; exp[\mu t + \alpha\beta_t] \quad (6)$$

The same was done for 100 trajectories. The graph representation is shown as below

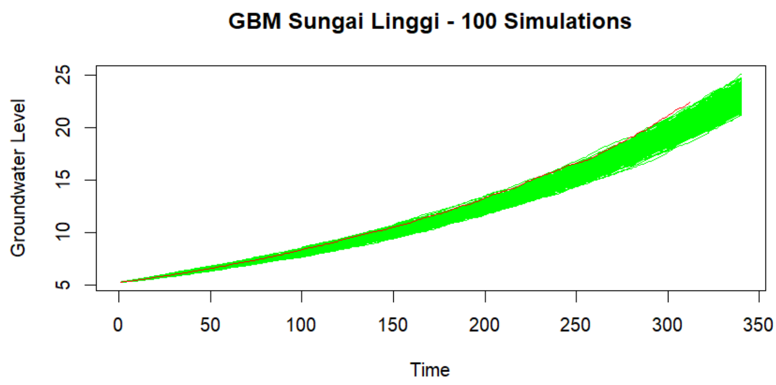


**Figure 3** Simulated Paths of GBM Sungai Linggi

Based on the GBM simulation model with 100 paths, the confidence interval for each GWL was calculated

$$e^{lnW_0+[\mu t+ \alpha\beta_t]} < W(t) < e^{lnW_0+[\mu t+ \alpha\beta_t]} \quad (4.3.6.4)$$

Table 6 Confidence Level for GBM simulated Paths

| **Confidence Interval of Groundwater Level** | | |
|---|---|---|
| **Well** | **Lower CI 95%** | **Upper CI 955** |
| **Loji Rawatan Air Sungai Linggi** | 5.3000 | 5.3254 |

Based on the table above, there were some simulations that location outside of the confidence level using the GBM model. The initial condition (RWL0) of the GBM model affects it significantly. The simulated trajectories might differ significantly over time depending on even small changes in the original value

## Conclusion

The initial NLR analysis revealed a low correlation between groundwater and river water levels. The R^2 value and p-value indicated a lack of significance between the variables, making it challenging to predict groundwater levels solely based on collected river water level data. The $\boldsymbol{R^2}$values for the three wells were 30% in Loji Rawatan Air Sungai Linggi. To increase the precision of groundwater level forecasts, additional elements outside river water levels may need to be taken into account given the observed errors and low correlation. Groundwater dynamics can be significantly influenced by local geological features, precipitation patterns, and hydrological processes, all of which should be considered when modelling and making predictions. Based on the GBM model, the predicted groundwater levels in Sungai Linggi show a lower value compared to the original data. The predicted river water levels are 5.30m, 5.31m, and 5.32m, while the original raw data indicates values of 5.3m, 8.37m, and 8.01m, respectively. The results of the GBM model indicate that the predicted groundwater levels closely align with the raw data. Although the predicted values may not be highly accurate, they still hold value as a reference for future analysis and can provide insights into potential trends in water level dynamics. The relationship between river water and groundwater is influenced by hydrological connections, such as discharge flow, the thickness of the saturated zone, and the physical properties of the aquifer. In cases where hydrogeological disconnect occurs, inconsistencies in water levels between the river and the ground may arise.

## References

[1] Eileen Poeter, Ying Fan, John Cherry, Warren Wood, Doug Mackay, 2020, Groundwater in our water cycle – getting to know Earth's most important fresh water sourcek. The Groundwater Project, Guelph, Ontario, Canada. https://doi.org/10.21083/978-1-7770541-1-3

[2] Fitts C. R. 2002, Groundwater Science, Academic Press, An Imprint of Elsevier Science, ISBN 0-12-257855-4

[3] Hussin N. H., Yusoff I., Raksmey M., 2020, Comparison of Applications to Evaluate Groundwater Recharge at Lower Kelantan River Basin, Malaysia, Department of Geology, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia, Geosciences 2020, 10(8), 289; https://doi.org/10.3390/geosciences10080289

[4] Sasmita Sahoo, Madan K. J., 2013, Groundwater-level prediction using Multiple Linear Regression and Artificial Neural Network Techniques: A Comparative Assessment. Springer-Verlag Berlin Heidelberg, Hydrogeology Journal (2013), 21: 1865–1887, DOI 10.1007/s10040-013-1029-5

[5] Heath, Ralph C.,1983, Basic ground-water hydrology: U.S. Geological Survey Water-Supply Paper 2220,86p. ISBN0-607-68973-0

[6] M. Azizah, M. I. Irawan and E. R. M. Putri, 2020, Comparison of stock price prediction using geometric Brownian motion and multilayer perceptron, AIP Conference Proceedings 2242, 030016 (2020); https://doi.org/10.1063/5.0008066

[7] Ross S. M., 2014, Introduction to probability models, Eleventh edition, Elsevier Inc., ISBN 978-0-12-407947-9

[8] Hyndman, R.J., ANthanasopoulos G., 2018, Forecasting: principle and practice. 2nd edition, Melbourne, Australia. https://otexts.com/fpp2/index.html

[9] Mansour & Hughes, 2018, British Geological Survey, Summary of results for national scale recharge modelling under conditions of predicted climate change. http://nora.nerc.ac.uk/id/eprint/521605/

[10] Tyson B., Gabrielli A., Evers J., 2022, Water Tables and Aquifers, National Geographic Society, https://education.nationalgeographic.org/resource/water-tables-and-aquifers

[11] Ritzema H. P, 1994, Frequency and Regression Analysis. In Drainage Principles and Applications, Publ. 16, pp. 175-224, International Institute for Land Reclamation and Improvement (ILRI), Wageningen, The Netherlands. ISBN 90-70754-33-9

[12] Woessner, William W. and Eileen P. Poeter 2020, Hydrogeologic Properties of Earth Materials and Principles of Groundwater Flow. The Groundwater Project, Guelph, Ontario, Canada. ISBN: 978-1-7770541-2-0

[13] Astro Awani, 2021, Trajedi Banjir 2021 https://pulse.astroawani.com/banjir2021

[14] Postal Bencana, Pusat Kawalan Bencana Negara, National Disaster Command Centre (NADMA). https://portalbencana.nadma.gov.my/ms/laporan

[15] Berita Harian, 2021, Bah Besar Penutup 2021, https://www.bharian.com.my/berita/nasional/2021/12/905583/bah-besar-penutup-2021

[16] Bernama, 2022, NADMA: Last Victim of Batang Kali landslide identified, New Strait Times. https://www.nst.com.my/news/nation/2022/12/864352/nadma-last-victim-batang-kali-landslide-identified?utm_source=dable

[17] Trisha N., 2022, Batang Kali landslide: PM orders multiagency SAR op, will head to site at night, The star. https://www.thestar.com.my/news/nation/2022/12/16/batang-kali-landslide-pm-orders-multi-agency-sar-op-will-head-to-site-at-night

[18] Zulkefli N., 2022, Kudos to Batang Kali landslide search and rescue team, The Malaysian Insight. https://www.themalaysianinsight.com/g/419403

[19] Bishal Dey, Bidesh Roy, Subir Datta, Taha S. U., 2022, Forecasting ethanol demand in India to meet future blending targets : A comparison of ARIMA and various regression models, The scientific committee of the 9th International Conference on Power and Energy Systems Engineering (CPESE 2022), https://doi.org/10.1016/j.egyr.2022.11.038

[20] Yildiz U. 2019, Simulating Stock Prices in Phyton using Geometric Brownian Motion, Towards Data Science. https://towardsdatascience.com/simulating-stock-prices-in-python-using-geometric-brownian-motion-8dfd6e8c6b

[21] R. Estember, M. Marana: (2016) Forecasting of stock prices using Brownian motion – Monte Carlo simulation, In proceeding so international conference on industrial engineering And operation management Kuala Lumpur