IRAQI STATISTICIANS JOURNAL

https://isj.edu.iq/index.php/isj

ISSN: 3007-1658 (Online)

Exploring Generalized Goel-Okumoto Process Parameters Estimation Strategies with Application

Najat Ahmed Taha [1], Muthanna Subhi Sulaiman[1]

[1] Department of Statistics and Informatics, College of Computer Science and Mathematics, University of Mosul, Mosul, Iraq

ARTICLE INFO

Article history:
Received 20/11/2024
Revised 20/11/2024
Accepted 14/1/2025
Available online 15/5/2025

Keywords:
Generalized Goel-Okumoto process,
Non-homogeneous Poisson process,
Maximum likelihood estimator,
Shrinkage method,
ABC algorithm

ABSTRACT

Estimating the parameters of stochastic processes is essential to understand behavior of processes and make correct decisions. In this paper, Generalized Goel-Okumoto process (GGOP) was used, which was used flexible frameworks to capture behavior of time-dependent systems, especially in the non-homogeneous Poisson process (NHPP), and its parameters were estimated using important traditional methods, namely the Maximum Likelihood (MLE) method and the Shrinkage (SH) method. An intelligent method represented by the Artificial Bee Colony algorithm (ABC) was proposed. This paper will contribute to a comparison between traditional and intelligent methods to estimate the parameters of the process under study, and toenhance the research results in understanding stochastic processes and supporting better decision-making, a realistic application was conducted on the shutdowns of the Badush Expansion Plant in Nineveh Governorate for the period from 1\1\2024 - 25\8\2024. It was found that the ABC algorithm in estimating the time rate of the GGOP was better than the maximum likelihood estimator and the shrinkage method.

1. Introduction

Stochastic processes are important models in modeling various phenomena in engineering, finance, biology, and other fields. Estimating the parameters of stochastic processes is essential to understand the behavior of processes and make correct decisions [1,2]. The Goel-Okumoto model and Goel model [3]; Musa-Okumoto model [4]; Zhao and Xie model [5] are proposed models based on the non-homogeneous Poisson process (NHPP). It has been proposed as a time-rate of occurrence of the process using some probability distributions. Pham [6] solved a generalized differential equation by which the NHPP based rate of occurrence is controlled. Accordingly, the Generalized Goel-Okumoto process (GGOP) has been widely used to model reliability, survival, and failure data in reliability engineering and biostatistics [7,8]. This process provides flexible frameworks to capture the behavior of time-dependent systems. Accurate parameter estimation in these processes is crucial for effective reliability analysis and risk assessment. However, parameter estimation in stochastic processes is often difficult due to the complexity of the underlying models and the presence of random chaos in the data. Various estimation techniques have been developed to address this challenge, including maximum likelihood estimator (MLE), shrinkage estimator (SH) and an intelligent method has been proposed to estimate the parameters of the process represented by the Artificial Bee Colony algorithm (ABC) [9]. The comparative study will contribute to the development of the latest techniques in parameter estimation for stochastic processes and provide valuable insights to researchers and practitioners in reliability engineering, biostatistics, and related fields. The results of this research enhance our

* Corresponding author.E-**mail** address: muthanna.sulaiman@uomosul.edu.iq
https://doi.org/10.62933/k6wr1g09

understanding of stochastic processes and support better decision-making in various applications.

2. Methodology

A nonhomogeneous Poisson process is a widely used model describing events that take place in time and whose frequency changes with time [10]. The mean value function up to time t is also defined as $m(t) = E[N(t)]$ which quantifies the rate behavior of occurrence of this process and is called the intensity function and is dependent of time t. This process must satisfy the following conditions [11]:

1) $N_0 = 0$, the number of occurrences at the beginning of the process is equal to zero.
2) The process increases with independent and unstable increments.
3) $p\{N_{t+\Delta t} - N_t \geq 2\} = 0(\Delta t)$ (1)
4) $p\{N_{t+\Delta t} - N_t = 1\} = \lambda(t)\Delta t + 0(\Delta t)$ (2)

So, the number of events in NHPP that occur in the period (t) follows the Poisson distribution with probability density function:

$$p\{N_{t+\Delta t} - N_t = n\} = \frac{e^{m(t)}[m(t)]^n}{n!} \quad (3)$$

where $(n = 0,1,2, \ldots)$, and $m(t)$ represent to the average value function, or the cumulative intensity function, which is derivable and always continuous as the following formula:

$$m(t) = E[N(t)] = \int_0^t \lambda(u)du \;, t \geq 0 \quad (4)$$

where $\lambda(t)$ is called the rate of occurrence of events or intensity function and it is a function of time (t) [12].

2.1 Generalized Goel-Okumoto Distribution

Generalized Goel-Okumoto (GGO) distribution is a mathematical model that is used mainly for analyzing software reliability and for estimating the efforts of managing projects. This model predicts the errors and faults throughout time as the original Goel-Okumoto distribution is modified by adding certain parameters to improve its estimation. This model involves three main parameters: the time rate, the primary rate, and the intensity of the failure rate [13].

The model is based on the idea that the rate of detecting the failures changes with time and this makes it suitable for the various software environments. The theoretical basis of this distribution was developed by Goel and Okumoto in 1979 as they developed a model that relies on a non-homogenous Poisson process, which enables failures in the software lifecycle [14].

The GGO model is used in different fields including the reliability estimation of software in various stages of its development. This model is considered efficient in estimating the number of possible errors and predicting the performance over time, improving the testing and development processes [15]. It can be used as a model of fault growth in complex systems as it helps the speed at which the faults are identified and fixed thus improving the system response to maintenance. In certain applications in other industries, this distribution is used in some fields like analyzing the reliability of industrial tools and estimating the shelf life of products in addition to analyzing data in the areas of public health and the environment. The most prominent characteristic is that it is a flexible distribution that allows the modification of the distribution shape based on a set of variables and this enhances the accuracy of estimations. The model greatly depends on the previous data for estimating the performance and this makes it a strong tool for prediction. It also assists in predicting the number of the predicted faults and the number of hours required to fix them. It can, also, be used to analyze the effect of different. From all that has been mentioned above, the model is used in various fields including the development of software, cost estimation, and performance analysis, which makes it a valuable tool for project management [16].

The density function of GGO distribution is formulated as follows [8]:

$$f(t) = abce^{-bt}t^{c-1}, t \geq 0, a, b > 0, c > 1 \quad (5)$$

where a, b, and c are parameters.

The cumulative function of GGO distribution can be written as follows:

$$F(t) = \alpha\left[1 - e^{-bt^c}\right] \quad (6)$$

2.2 Generalized Goel-Okumoto process

Assuming that the process $\{N(t), t \geq 0\}$ is a non-homogeneous Poisson process, the number of events occurring in the time interval (0, t) follows a Poisson distribution with a rate of occurrence as in Eq. (3) [12].

Therefore, the Generalized Goel-Okumoto process (GGOP) is described by describing $\lambda(t)$ as it distributes under the GGO distribution as follows:

$$\lambda(t) = abce^{-bt}t^{c-1}\,; t \geq 0, a, b > 0, c > 1 \quad (7)$$

where a, b, and c are the parameters of GGO distribution. Parameter a is the total predicted number of failures, and the parameters b and c represent the quality of the test process. It can be observed that rate of occurrence $\lambda(t)$ increases at the beginning of the process and then decreases when $a, b > 0\,, c > 1$.

So, the mean rate of occurrence of the process will be [17]:

$$m(t) = \int_0^t \lambda(u)du$$
$$= \int_0^t \left(abce^{-bu^c}u^{c-1}\right)du$$
$$= \alpha\left[1 - e^{-bt^c}\right] \quad (8)$$

3. Parameters Estimation of GGOP

Using the various methods of estimation to estimate the parameters of the stochastic processes is considered very important in terms of selecting the best estimator for the rate of occurrence, which results in obtaining the best stochastic model that represents better data in the best way as it is characterized with a high accuracy and high speed. Using estimators for the generalized Goel-Okumoto process using various methods to estimate the rate of occurrence and making comparisons between them to obtain a suitable method for estimation and representing the data in the best way.

3.1 Maximum Likelihood Estimation

Maximum Likelihood Estimation (MLE) is considered a method to estimate the parameters of a certain likelihood distribution based on certain data of observation. This can be done by maximizing the likelihood function that the observed data with the most likelihood is under the proposed statistical model. The point, in the space of information, that maximizes the likelihood function is called the maximum likelihood estimation [18]. The logic of the maximum likelihood estimation is characterized by spontaneity and flexibility and therefore, this method became a pioneer means in statistical inference it is characterized by certain traits like uniformity, non-bias, and efficiency and these traits make it suitable to be used with the complex models like Goel-Okumoto model [19].

If the process $\{N(t), t \geq 0\}$ represents the GGOP with a rate of occurrence that is limited with the form (7), then the likelihood function of the occurrence times $(t_1, t_2, \dots, t_n)$ that $(0 < t_1 \leq t_2 \leq \cdots \leq t_n \leq 0)$ is defined with the following formula [20]:

$$f_n(t_1, t_2, \dots, t_n) = \pi_{i=1}^{n}\lambda(t_i)e^{-m(t_0)} \quad (9)$$

The likelihood function of GGOP for the period $(0, t]$ with the density function:

$$L = \pi_{i=1}^{n}\left(abce^{-bt_i{}^c}t_i{}^{c-1}\right)e^{-\alpha[1-e^{-bt_i{}^c}]} \quad (10)$$

To estimate the process parameters using the maximum likelihood method, we begin by taking the normal logarithm of the formula (10) to obtain the logarithmic maximum likelihood function that is given by the following equation:

$$\ln L = n\ln a + n\ln b + n\ln c - b\sum_{i=1}^{n} t_i^c + \sum_{i=1}^{n} lnt_i^{c-1} - a\left(1 - e^{-bt_0^c}\right) \quad (11)$$

To estimate the values of the parameters a, b, and c, the likelihood function (11) for each parameter and then we make the resulting derivative equal to zero and this gives the following equations system:

$$\frac{\partial \ln L}{\partial a} = \frac{n}{a} - \left(1 - e^{-bt_0^c}\right) \quad (12)$$

$$\frac{\partial \ln L}{\partial b} = \frac{n}{b} - \sum_{i=1}^{n} t_i^c - ae^{-bt_0^c}\, t_0^c \quad (13)$$

$$\frac{\partial \ln L}{\partial c} = \frac{n}{b} - b\sum_{i=1}^{n} t_i^c \ln t_i + \sum_{i=1}^{n} \ln t_i^c \ln(\ln t_i) - ae^{-bt_0^c}\, bt_0^c\, \ln(\ln t_0) \quad (14)$$

let: $\frac{\partial \ln L}{\partial a} = 0, \frac{\partial \ln L}{\partial b} = 0, \frac{\partial \ln L}{\partial c} = 0$, then:

$$\hat{a} = \frac{n}{1-e^{-\hat{b}t_0^{\hat{c}}}} \quad (15)$$

$$\hat{b} = \frac{n}{\sum_{i=1}^{n} t_i^{\hat{c}} + \hat{a}e^{-\hat{b}t_0^{\hat{c}}}t_o^{\hat{c}}} \quad (16)$$

$$\hat{c} = \frac{n}{\hat{b}\sum_{i=1}^{n} t_i^{\hat{c}} \ln t_i - \sum_{i=1}^{n} lnt_i^{\hat{c}} \ln(lnt_i) + \hat{a}\, e^{-\hat{b}t_0^{\hat{c}}}\, bt_0^{\hat{c}} \ln(\ln t_0)} \quad (17)$$

3.2 Shrinkage Method

This method relies on the previous information, as the idea of the shrinkage estimator (SH) is pivoted on using the available previous information about the parameter to be estimated and then the formula of the shrinkage value is written as follows [21,22]:

$$\hat{\theta}_{SH} = z\hat{\theta} + (1 - z)\theta_0 \quad (18)$$

where:

$\hat{\theta}$: an unbiased primary estimator.

θ_0: previous information about the parameter.

θ_0: is a constant (or estimated) value that reflects prior knowledge or an estimate of an underlying parameter. It does not depend on the current sample; it may come from:

- previous studies or historical data.
- expert knowledge.
- average of several populations or estimators.
- the centre of a reasonable range for the parameters.

z: the amount of shrinkage and it takes the value between (0 to 1) in this research, a value for this amount will be proposed, which is (0.5).

Presupposing the primary values of the estimator that were estimated using the maximum likelihood method is performed as follows:

$$\hat{\theta}_{SH} = z\theta_{MLE} + (1 - z)\theta_0 \quad (19)$$

To estimate the parameters of the model used in this method, the following will be followed:

$$\hat{a}_{SH} = z\hat{a}_{MLE} + (1 - z)a_0 \quad (20)$$

$$\hat{b}_{SH} = z\hat{b}_{MLE} + (1 - z)b_0 \quad (21)$$

$$\hat{c}_{SH} = z\hat{c}_{MLE} + (1 - z)c_0 \quad (22)$$

3.3 Artificial Bee Colony Algorithm

Artificial Bee Colony (ABC) is an optimization algorithm derived from the food-searching behavior of a colony of honey bees. Previously it has been used effectively in solving optimization problems, for example, identifying parameters of complex mathematical models. In some cases, traditional methods may not provide accurate parameter estimates, especially when the sample size is very small or the distribution is complex. Estimating GGOP parameters using the ABC algorithm has been proposed to achieve more accurate and easier estimations than traditional methods. Below, a brief procedure is discussed on how to use the ABC algorithm to estimate three parameters of GGOP [23,24]:

1) Initialization:
- Identify three important estimates that will need to be constructed. To make the calculations and writing less complicated we will call them θ_1, θ_2, and θ_3.
- Set limits for each parameter in the search area; this includes the number of food sources (solutions), the number of bees, both employed and onlooker bees, and the maximum number of iterations.
- After doing that you should specify the range in which each of the values for the defined parameters should be searched, which will define the search space.

2) Generate Initial Population:
- Begin a population of potential food sources (or possible solutions) scattered across the designated area randomly. the design parameter space. Each food source has its paired solution which means a certain θ_1, θ_2, and θ_3 is associated with every food source.
- To find the fitness for each of the present food sources. In parameter estimation, the fitness can be referred to as an objective concerning a function such as Root Mean Squared Error (RMSE) or other errant measures that separate between factual and estimated values. Each employed bee has a food source (solution) and moves in the neighborhood of this solution altering one or more features. It needs to be constructed. We will call them θ_1, θ_2, and θ_3 to simplify calculations and writing.

3) Generate Initial Population:
- Randomly start a population of possible food sources (solutions) within the defined area of parameters. the defined parameter bounds. Every food source corresponds to a potential solution which means a certain θ_1, θ_2, and θ_3.
- Calculate the fitness of each food source. In parameter estimation, the fitness can be rooted in an objective concerning a

function such as RMSE or other error measures that give the difference between factual and estimated values.

4) Employed Bee Phase:
Each employed bee has a food source (solution) and searches in the vicinity of this solution changing one or several parameters. This alignment is typically done in a manner of adding some arbitrary and random non-systematic error to each of the parameter's values.

5) Onlooker Bee Phase:

- To do so, each onlooker bee then continues to search the identified food source by adjusting the parameters slightly by applicable benchmarks and determining the fitness.
- If the new solution found by an onlooker bee is better than the previous one, it memorizes the new solution replacing the previous food source solution.

6) Output:
The last estimate of the three parameters is obtained by using the solution on completion of the iterative process as the best or near-best solution.

Several experiments were conducted and initial values were assumed within the conditions of the algorithm the program prepared for this purpose in MATLAB/R2017b and based on the value next elementary:

1- Number of websites being searched: $N = 1$.
2- Swarm size: $n = 50$.
3- The final results were obtained upon iteration and this iteration was adopted:

$$Iteration = 25$$

4- Abandoned solution: $limit = 25$

4. Root Mean Squared Error

In this paper, RMSE is used to compare each of the MLE, SH and ABC algorithm applied to estimate the rate of occurrence of GGOP; this measure is considered as a fitness function for the ABC algorithm. If m_i represents the actual mean rate of occurrence of the process, and $\widehat{m}_i$ be the estimated mean rate of occurrence, then [8]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{m}_i - m_i)^2}{n}} \qquad (23)$$

5. Results and discussion

In this paragraph, a realistic application was conducted on the shutdown of the Badush Expansion Plant in Nineveh Governorate for the period from 1\1\2024 - 25\8\2024.

The first step includes ensuring that the data used is suitable for the studied operation, and this is done by examining the real data and plotting the distribution of cumulative days with the shutdown on a logarithmic scale. If the majority of these points are arranged in a straight line, the data is considered consistent with the occurrence rate of GGOP. Using MATLAB\R2017b language, the following Figure 1 was obtained:

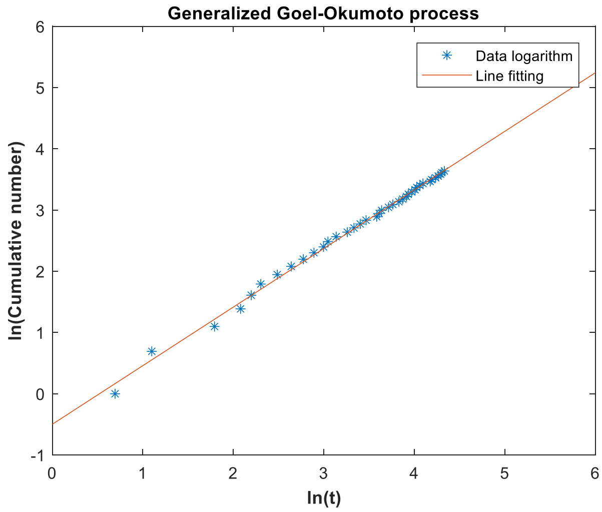

Figure 1. Plot of the test of the suitability of the shutdown data under study for GGOP.

Figure 1 shows the cumulative number of shutdowns in days with their occurrence times on a logarithmic scale for the data under study. It is noted that the scatter plot shows a linear behavior of the data, hence the possibility of modeling such data using GGOP. To evaluate the performance of GGOP estimation methods, the daily shutdown rate of the station during the study period was estimated using MATLAB/R2017b. The process parameters were estimated using the proposed estimation methods, the expected time rate of shutdowns of the plant was obtained, and the RMSE between the actual and estimated values was calculated. The results of this analysis are shown in Table 1 and Figure 2 below.

Table 1: Estimations of GGOP parameters using three methods with RMSE values for the time rate of shutdowns of plant.

Method	$\hat{a}$	$\hat{b}$	$\hat{c}$	RMSE
ABC	2.0006	4.2308	1.0001	0.0024
MLE	1.8548	0.9713	1.0908	0.0853
SH	1.6774	2.9357	1.2954	0.0566

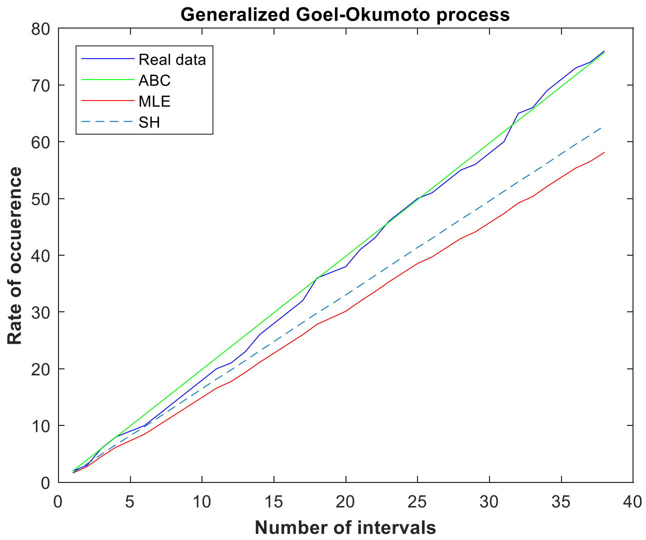

Figure 2. Estimates of the cumulative time rate of shutdowns of plant compared to real data.

Table 1 shows that the RMSE value for the ABC algorithm in estimating the rate of occurrence for the GGOP is lower than that of the MLE and SH methods. This indicates that the proposed ABC algorithm is more efficient in estimating the parameters of GGOP. Figure 2 shows the estimated time rate functions of the GGOP using each of the ABC, MLE, and SH estimation methods, compared to the cumulative real data values representing shutdowns of Badush Expansion Plant in Nineveh Governorate.

6. Conclusions

The present study aims to estimate the parameters of the Generalized Goel-Okumoto process using two types of methods, the first one uses important traditional methods which are the maximum likelihood method and the shrinkage method. The second one is to propose the use of an intelligent method which is the artificial bee colony algorithm. Through an application on real-world data from the Badush Expansion Plant in Nineveh Governorate, it was found that the estimated time rate function using ABC algorithm is closer to the real data than the rest of the methods used in estimation, followed by SH method, then MLE method, based on RMSE criterion. This indicates the accuracy of the results of the proposed intelligent algorithm for use in estimating the parameters of the process under study.

References

[1] Okamura, H., & Dohi, T. (2021). Application of EM algorithm to NHPP-based software reliability assessment with generalized failure count data. *Mathematics*, *9*(9), 985.

[2] Alsultan, F. A., & Sulaiman, M. S. (2024). Bayesian Estimation of Power Law Function in Non-homogeneous Poisson Process Applied in Mosul Gas Power Plant–Iraq. *Iraqi Journal of Science*, 2596-2604.

[3] Goel, A. L., & Okumoto, K. (1979). Time-dependent error-detection rate model for software reliability and other performance measures. *IEEE transactions on Reliability*, *28*(3), 206-211.

[4] Musa, J. D., & Okumoto, K. (1984, March). A logarithmic Poisson execution time model for software reliability measurement. In *Proceedings of the 7th international conference on Software engineering* (pp. 230-238).

[5] Zhao, M., & Xie, M. (1996). On maximum likelihood estimation for a general non-homogeneous Poisson process. *Scandinavian journal of statistics*, 597-607.

[6] Pham, H., & Zhang, X. (1997). An NHPP software reliability model and its comparison. *International Journal of Reliability, Quality and Safety Engineering*, *4*(03), 269-282.

[7] Asraful Haque, M., & Ahmad, N. (2021, December). Modified Goel-Okumoto software reliability model considering uncertainty parameter. In *Mathematical Modeling, Computational Intelligence Techniques and Renewable Energy: Proceedings of the Second International Conference, MMCITRE 2021* (pp. 369-379). Singapore: Springer Singapore.

[8] Adel, S. H., Fatah, K. S., & Sulaiman, M. S. (2023). Estimating the Rate of Occurrence of Extreme value process Using Classical and Intelligent Methods with Application: nonhomogeneous Poisson process with intelligent. *Iraqi Journal of Science*, 3054-3065.

[9] Hussain, A. S., Sulaiman, M. S., & Fatah, K. S. (2022, August). Estimation of Rayleigh Process Parameters Using Classical and Intelligent Methods with Application. In *2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM)* (pp. 447-452). IEEE.

[10] Al-Hashimi, M. M., Hayawi, H. A., & Al-Kassab, M. (2024). A comparative study of traditional methods and hybridization for predicting non-stationary sunspot time series. *Computer Science*, *19*(1), 195-203.

[11] Ra, Y. C. (2021). A Comparative Analysis on the Performance of Finite Failure NHPP Software Reliability Model Based on Rayleigh-type Lifetime Distribut ion. *Journal of Theoretical and Applied Information Technology*, *99*(24), 6162-6172.

[12] Yan, T. (2019). *Nonhomogeneous poisson process models with a generalized bathtub intensity function for repairable systems* (Master's thesis, Ohio University).

[13] Erto, P., Giorgio, M., & Lepore, A. (2018). The generalized inflection S-shaped software reliability growth model. *IEEE Transactions on Reliability*, *69*(1), 228-244.

[14] Chichan, A. M. A. (2020). Estimate the parameters of the Generalized Goel-okumoto model using the Maximum likelihood and the shrinkage methods. *Tikrit Journal of Administration and Economics Sciences*, *16*(52 part 3).

[15] Asraful Haque, M. (2022, November). Software reliability models: A brief review and some concerns. In *The International Symposium on Computer Science, Digital Economy and Intelligent Systems* (pp. 152-162). Cham: Springer Nature Switzerland.

[16] SH, A., Fatah, K. S., & Sulaiman, M. S. (2023). Estimating the Rate of Occurrence of Exponential Process Using Intelligence and Classical Methods with Application. *Palestine Journal of Mathematics*, *12*.

[17] Pham, H. (2007). *System software reliability*. Springer Science & Business Media.

[18] Rossi, R. J. (2018). *Mathematical statistics: an introduction to likelihood based inference*. John Wiley & Sons.

[19] Ward, M. D., & Ahlquist, J. S. (2018). *Maximum likelihood for social science: Strategies for analysis*. Cambridge University Press.

[20] Van Calster, B., van Smeden, M., De Cock, B., & Steyerberg, E. W. (2020). Regression shrinkage methods for clinical prediction models do not guarantee improved performance: simulation study. *Statistical methods in medical research*, *29*(11), 3166-3178.

[21] Pels, W. A., Adebanji, A. O., Twumasi-Ankrah, S., & Minkah, R. (2023). Shrinkage Methods for Estimating the Shape Parameter of the Generalized Pareto Distribution. *Journal of Applied Mathematics*, *2023*(1), 9750638.

[22] Sharma, A. B. H. I. S. H. E. K., Sharma, A., Choudhary, S., Pachauri, R. K., Shrivastava, A., & Kumar, D. (2020). A review on artificial bee colony and it’s engineering applications. *Journal of Critical Reviews*, *7*(11), 4097-4107.

[23] Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *Journal of global optimization*, *39*, 459-471.

[24] Lavanya, G., Neeraja, K., Basha, S. A., & Sangeetha, Y. (2017). Parameter estimation of goel-okumoto model by comparing aco with mle method. *International Research Journal of Engineering and Technology*, *4*(3), 1605-1615.