

## Original Research Article

# MODELLING GEOCHEMICAL ANOMALIES USING THE BLOCK MAXIMA APPROACH IN EXTREME VALUE THEORY

## Abstracts

Geochemical anomalies are crucial in mineral exploration because they are linked to deposit formation. The generalized extreme value (GEV) distribution is used to fit the block maxima (BM) technique in extreme value theory (EVT) to geochemical anomalies of Au concentration in this paper. According to the shape parameter computation, the Fréchet family of GEV distributions is a good fit for the data. Statistical inference is performed by examining numerous return levels matching to the return periods, with the findings suggesting that the possibility of a higher Au concentration return level is highly unlikely.

**Keywords:** Extreme value theory, generalized extreme value distribution, geochemical anomalies, part per billion.

## 1 INTRODUCTION

Ghana's most important mineral resources are gold, diamonds, manganese, and bauxite. For over two decades, gold has been the most important mineral produced in the country, accounting for about 90 percent of all mining profits. Minerals are essential for every country's economic growth. Because minerals have such a large impact on Ghana's economy, the mineral sector's profits may be maximized, and the company can prevent mineral shortages by accurately calculating the mining process.

Mineral resource forecasts have proven to be a difficult task for academics in recent years. Despite its natural wealth, Ghana is vulnerable since its control generation is dependent on foreign cash. In order to cope with new challenges, it is critical to analyze and construct a model that can accurately anticipate mineral reserves. The major goal of accurate mineral production data forecasting is to make the most effective and optimal use of natural resources in economic growth. Because datasets are nonlinear, nonstationary, and have time-varying characteristics, mineral resource prediction is a critical issue. Many academics are attempting to predict nonlinear datasets with complicated time-varying properties with high accuracy.

When looking at a data set, the goal is generally to figure out what the typical observations from the underlying distribution are. This is commonly accomplished by calculating a measure of central tendency, such as the mean, before creating a confidence interval. Normal distributions and time series methods make up the majority of common statistical techniques. These are all based on symmetric distributions, which fail to account for the tail behavior of fat (heavily) tailed and asymmetric distributions [1];[2]. In geology and mining, long-tailed or heavy-tailed distributions have been identified. High sample values in contrast to the center of the distribution are common in grade distributions for precious metals such as gold [3] and size distributions for oil and gas resources [4].

In the domains of mineral exploration and mineral resource assessment, delineating geo-anomalies to aid in the discovery of mineral resources is a standard procedure. A geo-anomaly is distinguished from its surroundings by major changes in composition, texture, structure, and origin [14]. Although no single theory for describing the mechanisms of the formation of geochemical anomalies in deeply weathered and transported terrains with regolith or other land cover has been accepted (Mann, 2010), several mechanisms assuming the upward migration of ions to the surface have been proposed, including electrochemistry, diffusion, groundwater pumping, convection, capillary rise, and vegetation [15][16]. At the International Statistics Congress in Seoul, Republic of Korea, [5] established the geological anomaly (maximum observations) as an extreme value based on a mathematical basis (as discussed by [6]). According to [7] in 2012, extreme value theory (EVT) is a statistical science that investigates the limiting distribution of the lowest and greatest value and assesses the risk of extreme occurrences. Because most EVT applications are motivated by the need to predict the likelihood of large observations, such as geological anomalies, this thesis will focus on the right tail of the underlying distribution, which entails using the block maxima approach to look at extremely large observations (geochemical anomalies).

## **2 DATA AND METHODOLOGY**

### **2.1 Source of Data**

2750 soil samples were collected on a grid basis from the research region using the soil auger drilling method, with an average depth of 3 m in Ghana's Wassa Amefi West, for the study. The gold

concentration in parts per billion in these samples was determined by the Ghana Geological Survey Department using graphite furnace-atomic absorption spectrometry (GF-AAS) [12,13].

## 2.2 METHODOLOGY

### 2.21 Model Formulation

Suppose  $X_1, X_2, \dots, X_n$  where  $i=1,2,\dots,n$  are independent and identically distributed (i.i.d) random variable(r. v) with  $E(X_i) = \mu$  and  $var(X_i) = \sigma^2$ . From Central limit theorem[10] If *sample mean*  $\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$  then :

$$\lim_{n \rightarrow \infty} \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = N(0,1) \quad (1)$$

According to fisher Tippet theorem [8], as the sample sizes(n) increases the distribution of the extreme maxima  $M_n$  (*geochemical anomalies*) will follow generalized Extreme value (GEV)

The distribution GEV is given:

$$G(x) = \exp\left[-\left(1 + \varepsilon \left(\frac{x - \mu}{\sigma}\right)\right)^{-\frac{1}{\varepsilon}}\right], \text{ where if } 1 + \varepsilon \left(\frac{x - \mu}{\sigma}\right) > 0 \quad (2)$$

GEV is made up of three parameters:

1.  $\varepsilon$  shape parameter (Extreme Value Index),  $-\infty < \varepsilon < \infty$
2.  $\sigma$  - Scale parameter (Dispersion of the  $M_n$ ),  $\sigma > 0$
3.  $\mu$  - location parameter (mean of the  $M_n$ ),  $-\infty < \mu < \infty$

### 2.22 Domain of Attraction of GEVD

The GEV consists of three family type of distribution:

- i. If  $\varepsilon = 0$ , *Gumbel type*

$$\Lambda(x) = e^{-e^{-x}}, -\infty < x < \infty \quad (3)$$

The Gumbel domain is  $(-\infty, +\infty)$

- ii. If  $\varepsilon < 0$ , *Weibull type*

$$\Psi_\alpha(x) = \begin{cases} e^{-(-x)^{-\alpha}}, & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases} \quad (4)$$

The Weibull domain is  $(-\infty, \mu - \frac{\sigma}{\varepsilon})$

iii. if  $\varepsilon < 0$ , Fréchet type

$$\Phi_{\alpha}(x) = \begin{cases} 0 & , \text{ if } x < 0 \\ e^{-x^{-\alpha}} & , \text{ if } x \leq 0 \end{cases} \quad (5)$$

The Fréchet domain is  $(\mu - \frac{\sigma}{\varepsilon}, \infty)$

### 2.23 Estimation of Parameter

Suppose  $x_1, x_2, \dots, x_n$  are i.i.d data on  $x \sim \text{GEV}(\mu, \sigma, \varepsilon)$ .

$$\begin{aligned} L(\mu, \sigma, \varepsilon) &= \prod_{i=1}^n \left\{ \frac{1}{\sigma} \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon} - 1} \right\} \cdot \exp \left\{ - \left( 1 + \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon}} \right) \right\} \\ &= \frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \frac{1}{\sigma} \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon} - 1} \right\} \exp \left\{ - \left( 1 + \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon}} \right) \right\} \end{aligned}$$

$$\log L(\mu, \sigma, \varepsilon) = -n \log \left( \frac{1}{\sigma} + 1 \right) - \sum_{i=1}^n \log \left( 1 + \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon}} \right) - \sum_{i=1}^n \left( 1 + \varepsilon \left( \frac{x_i - \mu}{\sigma} \right)^{\frac{-1}{\varepsilon}} \right) \quad (6)$$

### 2.24 Return Level and Exceedance Period

A return level ( $X_T$ ) with period T (in years) is the level expected to exceeded on average once in every T year and is given by:

$$X_T = \mu + \frac{\sigma}{\varepsilon} \left[ \left( -\log \left( 1 - \frac{1}{T} \right) \right)^{-\varepsilon} - 1 \right], \text{ with exceedance Probability} = \frac{1}{T} \quad (7)$$

## 3 RESULTS AND DISCUSSION

This section's descriptive analysis highlights the most essential aspects of the data collected. The table 3.1 below provides a short overview of all the factors under investigation.

Table 1: A short overview of all the factors under investigation.

Mean	Maximum	Minimum	Std.	Skewness	Kurtosis
52.07	990.0	0.1	78.19	6.25	46.45

In general, the index's maximum and lowest Au contents are considerably distinct. The standard deviation is also high, indicating that Au content varies greatly. Positive skewness is also apparent, showing that the right tail is especially severe, indicating non-symmetric yields for the Au concentration.

### 3.1 TEST OF STATIONARITY OF THE SAMPLE DATA

The augmented Dickey-Fuller (ADF) and sequence correlation analysis are the most common approaches employed in stable test zones to analyze the self-correlation of geological data. The ADF may be used to retrieve the sample data's test results (Table 3.2). The null hypothesis that the data is not stationary should be rejected since Au's p-value of 0.067 is within the significant level 5 percent rejection zone. As a result, the Au sequences do not have unit roots; instead, they are stationary sequences that confirm the stationary tendency of the time series.

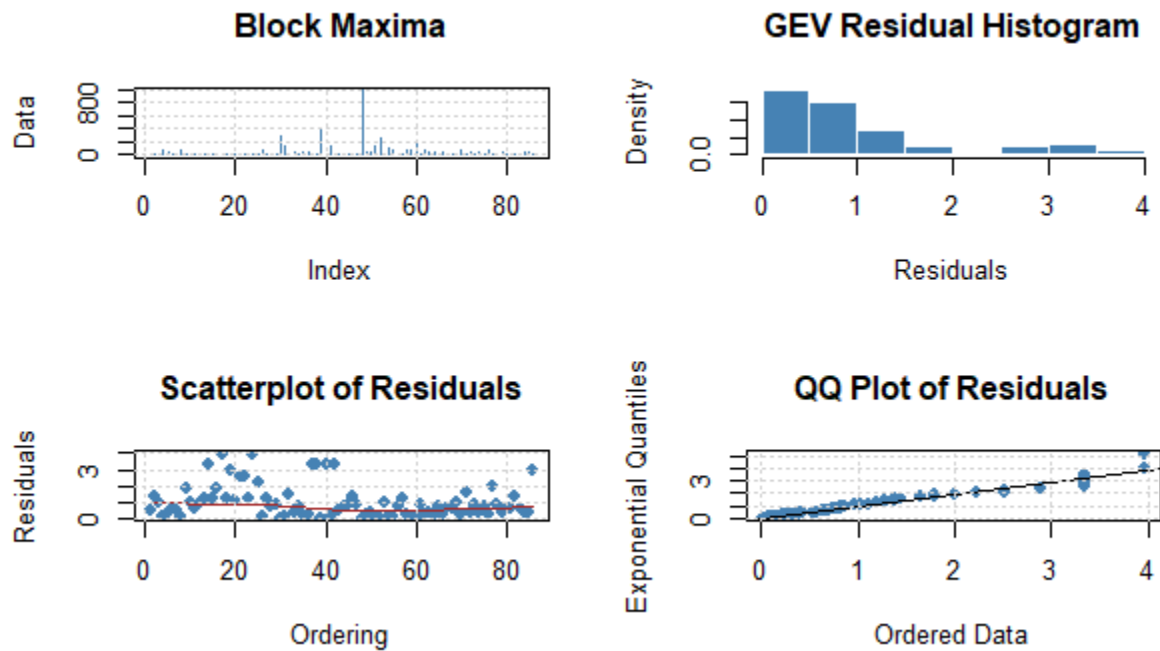
**Table 2: The ADF test of Au Concentration**

Null hypothesis: Data is not Stationary

Element	DF- value	P-Value
Au	-13.32	0.067

### 3.2 Block Maxima Approach (GEVD)

Because the samples are stable, the maximum Au content is modeled using a stationary model. Although stationarity does not ensure independence, we may assume the data are independent and use the traditional EVT to characterize the stationary sequence as an independent sequence because the blocking technique decreases data reliance. The block maxima method is currently used to compute our daily gold (Au) concentration in part per billion (ppb). The correct selection of the periods that define the blocks is critical to the success of this method. The recommended periods are block sizes of 30 observations due to the nature of the gold concentration and also to ensure that there is enough data for the extremal type theorem to hold. The research period's data (i.e., 2750 observations) is divided into 30 non-overlapping sub-samples, with the highest observed value picked. As a consequence, there are 91 observations in all, each of which provides the daily Au concentration.



**Figure 1: The daily maximum Au conc. of GEV fit.**

A sample plot of block maxima on the left and a histogram depicting the GEV of residuals on the right are the two top plots in Figure 1. (right). At the bottom is a scatter plot of the residuals that shows the time of block maxima. The purpose is to search the data for a possible temporal trend. To aid in judging this, a simple fitted curve (using the fExtremes package in R) is superimposed, and there is evidence of a systematic trend. The solid line is the smoothing of scattered residuals obtained by a spline method. In the bottom right corner, the residuals are presented in a QQ-plot. The QQ plot of the fitted model, which is based on the GEV fitted to all 91 block maxima, stays close to the straight line.

**Table 3: Parameter Estimates for the GEV Model Using MLE**

Parameter	Estimates	Standard Error	Confidence Interval
Scale ( $\sigma$ )	7.1810	0.14276	6.9901 - 7.4608
Location ( $\mu$ )	6.2696	0.11300	6.0481 - 6.4910
Shape ( $\xi$ )	0.9913	0.02088	0.9503 - 1.0321

All estimates provided by both estimators have a positive Extreme Value Index (EVI) called shape parameter ( $\xi$ ), suggesting that the underlying distribution belongs to the Fréchet domain of attraction. Furthermore, the Wald confidence range for the form parameter excludes zero, implying that the Au concentration distribution is beyond the Gumbel zone of attraction.

As illustrated in Figures 2a, 2b, and 2c, the profile log-likelihood delivers increased accuracy for the confidence intervals of estimations, according to [10].

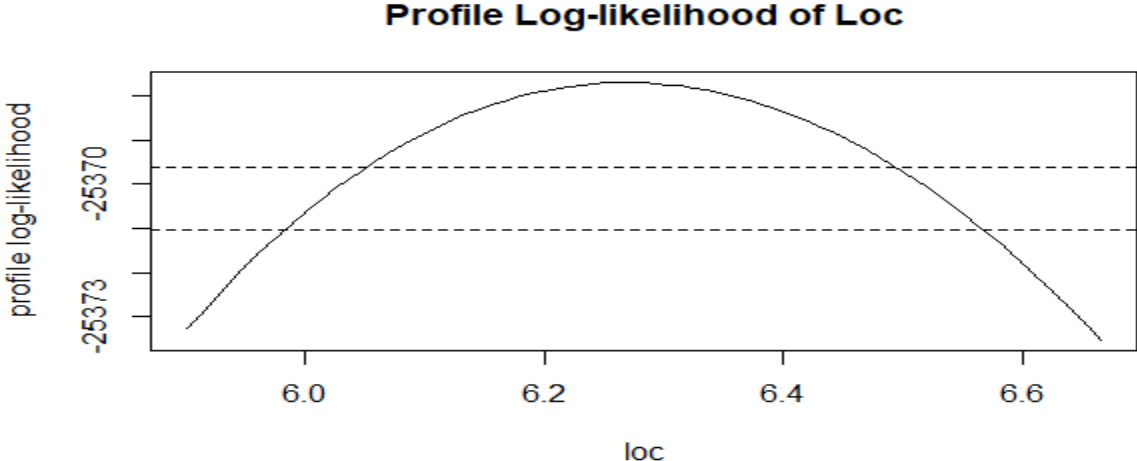


Figure 2a: Profile Likelihood for Location parameter

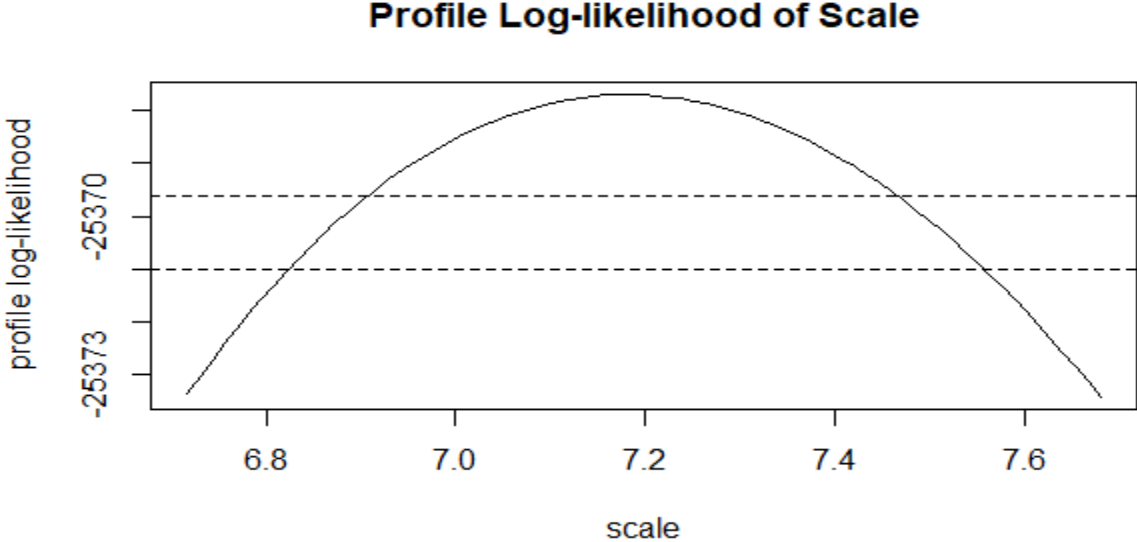
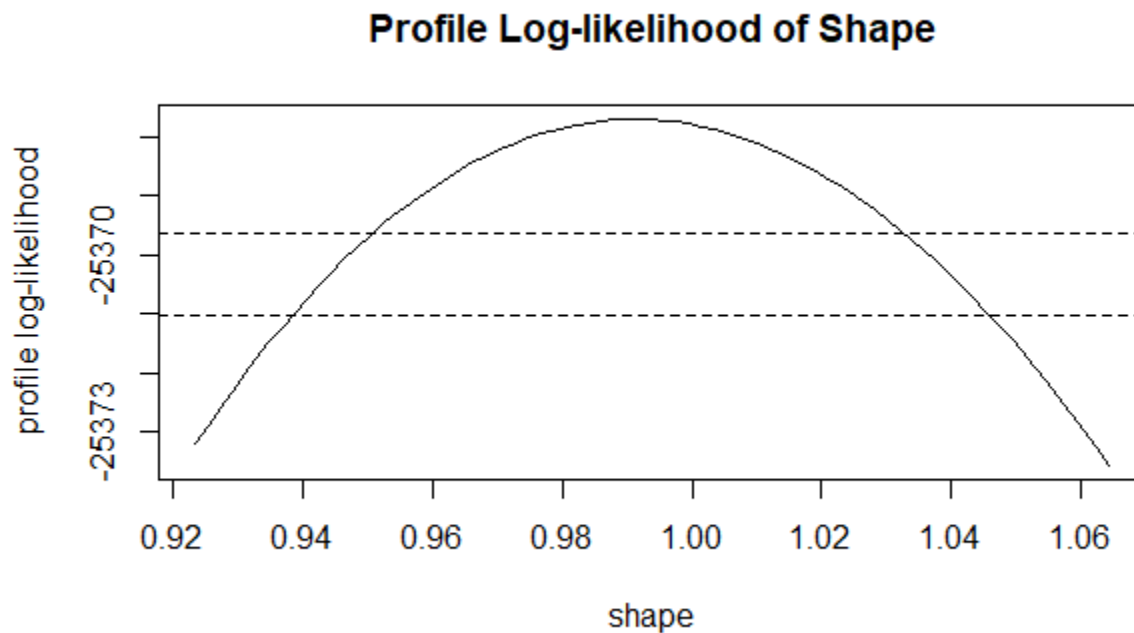


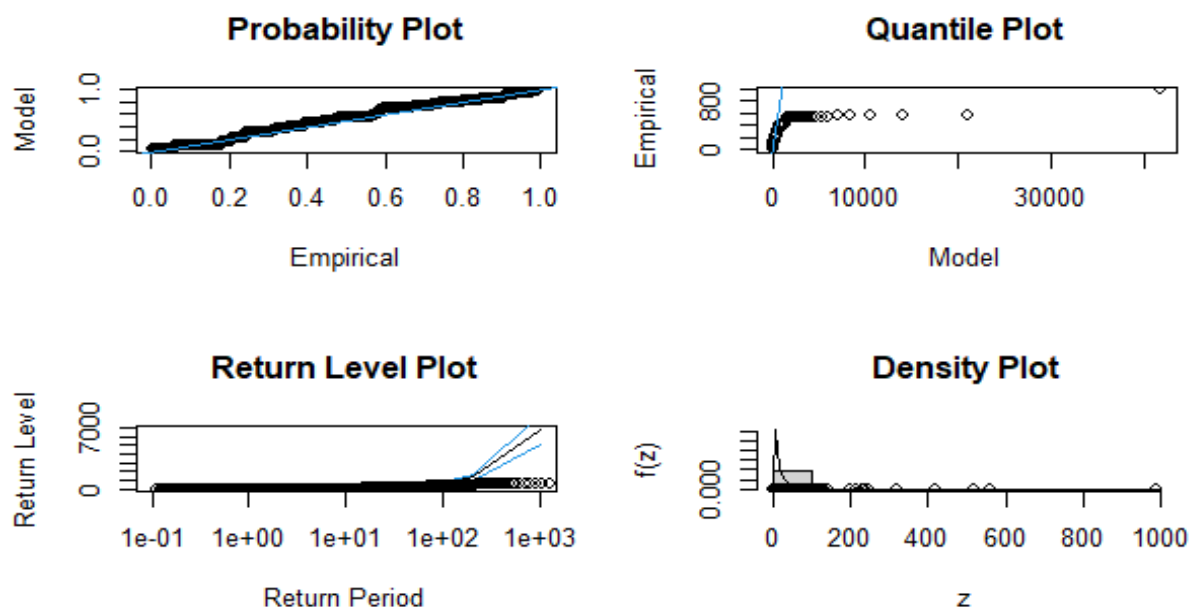
Figure 2b: Profile Likelihood for Scale parameter



**Figure 2c: Profile Likelihood for Shape parameter**

### 3.4 Model Diagnostics

The fitted GEV models for the monthly maximum gold concentration are further evaluated using diagnostic plots. A QQ-plot (a), a PP-plot (b), and a histogram overlay of the fitted GEV's density curves are shown in the pictures (c). The graphs are shown in Figure 3. For the QQ-plot and the PP-plot, a good fit should result in a straight one-to-one line of points. In most cases, the QQ plot is preferred over the probability plot. Because the QQ-plot and the PP-plot are both linear, the model must be right. As a consequence, the GEV model fitted the data well. Based on the histogram, the density appears to be in line with the data points. As a result, we conclude that the diagnostic plots support the fitted model.



**Figure 3. Diagnostic Plots for GEV**

The diagnostic graphic and positive estimate derived from the shape parameter clearly show that the GEV model belongs to the Fréchet domain of attraction; we then estimate the distribution's lower endpoint. The GEV in the Fréchet domain distribution, according to Coles (2001), has an unlimited upper bound. The GEV model is defined as follows:

$$G(x) = \exp\left[-\left(1 + 0.99 \left(\frac{x-6.27}{7.18}\right)^{-1}\right)^{-0.99}\right]$$

Since GEVD has an infinite upper bound and a finite lower bound in the Fréchet domain. The term  $\mu - \sigma/\xi$  represents the lower limit. According to the research, the bottom bound for the left tail is -0.974, which is substantially lower than the lowest gold concentration ever recorded for the time period in question. Table 4 estimates the distribution's tails' return level and exceedance probability.

**Table 4: Estimate of Return level and Exceedance Probability**

Return Period	Return level	Exceedance Probability
2 years	9.4432	0.11
5 years	31.067	0.03
10 years	66.433	0.06
20 years	136.63	0.007
50 years	345.58	0.003
100 years	691.413	0.001

According to our findings in table 4, larger gold concentrations are associated with a lower right-tail exceedance probability, meaning that the possibility of a return level with a higher value is extremely low.

## **CONCLUSION**

In the mining industry, modeling the frequency of occurrence of Au concentration is critical for assessing the effects of maximum Au concentration on ore deposit production. The GEVD is used to model the maximum daily Au concentration in the western part of Ghana. The diagnostic tools, the P-P and Q-Q plots, which are shown in Figures 1 and 3, respectively, are used to establish this. The Fréchet family is a suitable distribution for modeling maximum daily Au concentration in Ghana due to the positive value and asymptotic behavior of the shape parameter of GEVD. The validity of this claim is established by calculating the confidence interval for the shape parameters, which is determined to be within positive ranges. Statistical inference is performed by examining numerous return levels matching to the return periods, with the results indicating that the likelihood of a return level with a greater value is extremely unlikely.

## References

- [1] Gençay, R., Selçuk, F., & Whitcher, B. (2005). Multiscale systematic risk. *Journal of International Money and Finance*, 24(1), 55-70.
- [2] Wentzel, D. C., & Mare, E. (2007). Extreme value theory—An application to the South African equity market. *Investment Analysts Journal*, 36(66), 73-77.
- [3] Sichel, HS\*, Kleingeld, WJ\*\* & Assibey-Bonsu, W. (1992). A comparative study of three frequency-distribution models for use in ore evaluation. *Journal of the Southern African Institute of Mining and Metallurgy*, 92(4), 91-99.
- [4] Seyedghasemipour, S. Javad, and B. B. Bhattacharyya. "The loghyperbolic: An alternative to the lognormal for modeling oil field size distribution." *Mathematical geology* 22.5 (1990): 557-571.
- [5] Yongqing, C., & Pengda, Z. (1998). Zonation in primary halos and geochemical prospecting pattern for the Guilaizhuang gold deposit, eastern China. *Nonrenewable Resources*, 7(1), 37-44.
- [6] Bhole, M. V., Manson, A. L., Seneviratne, S. L., & Misbah, S. A. (2012). IgE-mediated allergy to local anaesthetics: separating fact from perception: a UK perspective. *British journal of anaesthesia*, 108(6), 903-911.
- [7] Chen, M., Suzuki, A., Thakkar, S., Yu, K., Hu, C., & Tong, W. (2016). DILIrank: the largest reference drug list ranked by the risk for developing drug-induced liver injury in humans. *Drug Discov Today*, 21(4), 648-653.
- [8] Moritz, M. A. (1997). Analyzing extreme disturbance events: fire in Los Padres National Forest. *Ecological Applications*, 7(4), 1252-1262.
- [9] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1), 43.
- [10] Matthys, G., Delafosse, E., Guillou, A., & Beirlant, J. (2004). Estimating catastrophic quantile levels for heavy-tailed distributions. *Insurance: Mathematics and Economics*, 34(3), 517-537.
- [11] Qin, F., Liu, B., & Guo, K. (2016). Using EVT for Geological Anomaly Design and Its Application in Identifying Anomalies in Mining Areas. *Mathematical Problems in Engineering*, 2016.
- [12] Xie, X.J.; Wang, X.Q.; Zhang, Q.; Zhou, G.H.; Cheng, H.X.; Liu, D.; Cheng, Z.Z.; Xu, S.F. Multi-scale geochemical mapping in China. *Geochem. Explor. Environ. Anal.* 2008, 8, 333–341. [CrossRef]
- [13] Xie, X.J.; Mu, X.Z.; Ren, T.X. Geochemical mapping in China. *J. Geochem. Explor.* 1997, 60, 99–113
- [14] Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A. and Gocayne, J.D., 2001. The sequence of the human genome. *science*, 291(5507), pp.1304-1351.
- [15] Anand, R.R. and Robertson, I.D., 2012. The role of mineralogy and geochemistry in forming anomalies on interfaces and in areas of deep basin cover: implications for exploration. *Geochemistry: Exploration, Environment, Analysis*, 12(1), pp.45-66.

[16] Kennedy, R. E., Yang, Z., & Cohen, W. B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr—Temporal segmentation algorithms. *Remote Sensing of Environment*, 114(12), 2897-2910.

UNDER PEER REVIEW